

**P. Perdikaris, et al.,
“Nonlinear information fusion algorithms for
data-efficient multi-fidelity modeling”,
Proc. R. Soc. A 473, 20160751 (2017).**

前山伸也

メタ階層ダイナミクスユニット, 核融合科学研究所

フュージョン・インフォマティクス勉強会

2024年11月28日 15:00 – 16:30, 場所：研究I期棟601セミナー室/オンライン

今回の狙い

最近、マルチフィデリティモデリングの観点からプラズマ乱流輸送モデルに取り組んでいる。

- 勉強会で理解を共有し、転移学習アプローチとの差異についての共同研究に発展？
- 乱流輸送モデルに限らない、マルチフィデリティモデリングのプラズマ・核融合応用？

文献1. 非線形自己回帰ガウス過程回帰の原著論文 ←今回紹介する論文

P. Perdikaris, et al., "Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling", Proc. R. Soc. A 473, 20160751 (2017). <https://doi.org/10.1098/rspa.2016.0751>

文献2. 私たちが用いているプラズマ乱流輸送モデリングの事例

S. Maeyama, et al., "Multi-Fidelity Information Fusion for Turbulent Transport Modeling in Magnetic Fusion Plasma", accepted in Sci. Rep. (2024). <https://doi.org/10.21203/rs.3.rs-4791765/v1>

文献3. "multi-fidelity", "transfer learning" の検索語で被引用数の高いもの

S. Chakraborty, "Transfer learning based multi-fidelity physics informed deep neural network", J. Comput. Phys. 426, 109942 (2021). <https://doi.org/10.1016/j.jcp.2020.109942>

Abstract

マルチフィデリティモデリングは、低精度だが多数のLow-fidelityデータと高精度だが少数のHigh-fidelityデータを組み合わせて推論の精度を向上させる手法。特にlow-とhigh-fidelityモデルの間に強い相関がある場合に効果的。

しかし、多くの場合low-fidelityモデルは特定の入力パラメータの範囲でのみ良い推論となるが、それ以外の範囲では誤った傾向を与えうる。

この論文では、Gaussian過程回帰と非線形自己回帰に基づく確率論的フレームワークを提案する。複数のモデル間の複雑な非線形・パラメータ依存の相互相関の学習を可能とし、Low-fidelityモデルが与える誤った予測に対する保護機能としても働く。

この手法は、新しいクラスのマルチフィデリティ情報融合アルゴリズムであり、計算量の増大を防ぎつつ、既存の線形自己回帰手法を拡張したものである。

提案手法の性能をいくつかのベンチマーク問題でテストした。

1. Introduction

Low-fidelityモデル・・・単純で低計算コストだが、潜在的に不正確なモデル（データ多数）

High-fidelityモデル・・・正確だが、複雑で高計算コストのモデル（データ小数）

High-fidelityモデルだけでは計算コストが高すぎて（データが足りなすぎて）調べられない問題を、Low-fidelityモデルと組み合わせることでどうにか対処したい。

→ **マルチフィデリティモデリング**

多くのマルチフィデリティモデリングは、ガウス過程(Gaussian process, GP)回帰に基づく。

E.g., Kennedy & O'Hagan [5]: GP+線形自己回帰

しかし、単純な線形自己回帰では不十分

写像の合成により非線形かつパラメータ位置依存の相互相関を学習できる手法を開発する。

2. Methods – (a)

(a) Regression with Gaussian processes

- 観測された入力/出力データセット $D = \{x_i, y_i\} = (x, y)$ がある。
- このデータが未知の関数関係 $y = f(x)$ を持ち、その関数 f は平均関数 $m=0$ 、共分散関数 k を持つガウス過程確率分布により生成されると仮定する。共分散関数を与える関数 k は適切なカーネル関数で与える。(例えば、 $k(x, x'; \theta) = \exp(-\theta|x - x'|)$ など)
- データセットを利用して構築される事後分布 f_* は、

$$p(f_* | \mathbf{y}, \mathbf{X}, x_*) = \mathcal{N}(f_* | \mu_*(x_*), \sigma_*^2(x_*)), \quad (2.3)$$

$$\mu_*(x_*) = k_{*n} \mathbf{K}^{-1} \mathbf{y} \quad (2.4)$$

and

$$\sigma_*^2(x_*) = k_{**} - k_{*n} \mathbf{K}^{-1} k_{*n}^\top, \quad (2.5)$$

where $k_{*n} = [k(x_*, x_1), \dots, k(x_*, x_n)]$ and $k_{**} = k(x_*, x_*)$. Predictions are computed using the posterior mean μ_* , while uncertainty associated with these predictions is quantified through the posterior variance σ_*^2 .

ある x に対して、 $y=f(x)$ の予測値は平均 μ 、予測の不確かさは標準偏差 σ として推定できる。

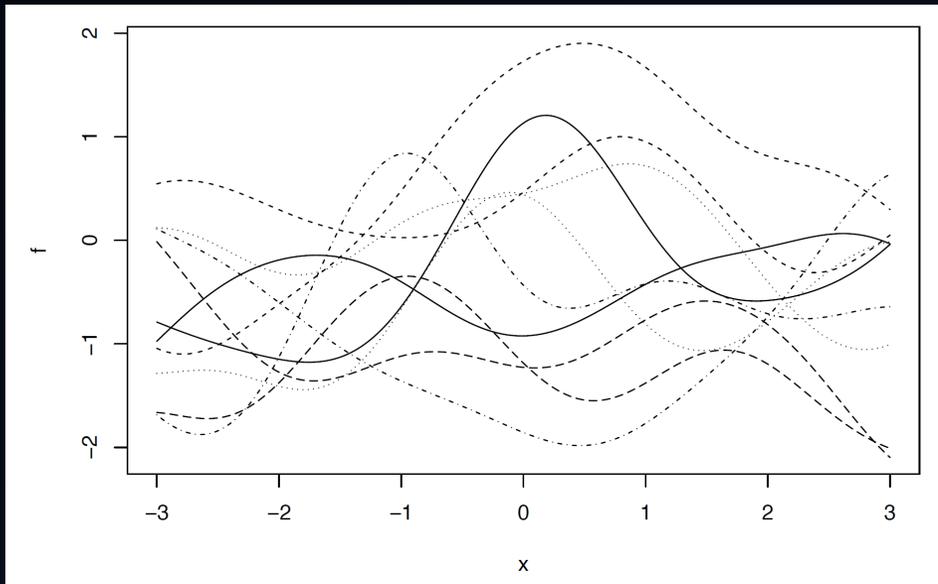
- 回帰モデルを特徴づけるハイパーパラメータ θ は、周辺対数尤度 (モデルからデータセットが観測される確率) を最大化するように選ぶ

$$\log p(\mathbf{y} | \mathbf{x}, \theta) = -\frac{1}{2} \log |\mathbf{K}| - \frac{1}{2} \mathbf{y}^\top \mathbf{K}^{-1} \mathbf{y} - \frac{n}{2} \log 2\pi. \quad (2.2)$$

補足：Gauss過程 $\mathcal{GP}(f|\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$

確率関数 $f(\mathbf{x})$ が、平均関数 $\mu(\mathbf{x})$ と共分散関数 $k(\mathbf{x}, \mathbf{x}')$ で特徴づけられる(\mathbf{x} が連続なので無限次元) 正規分布をとる。

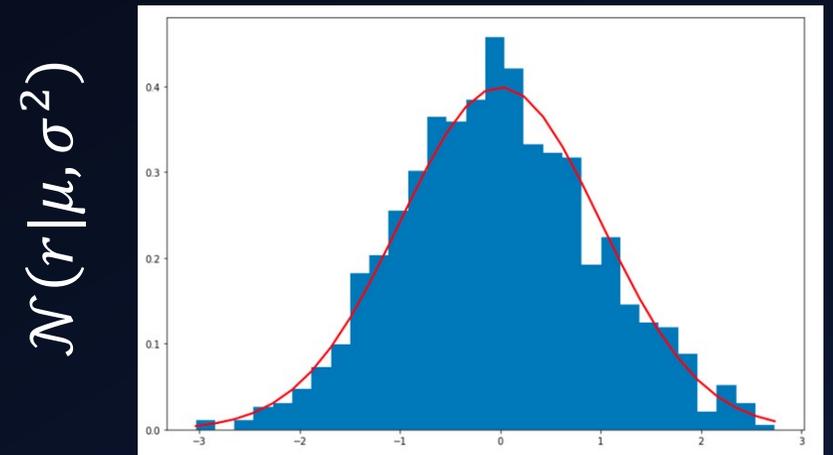
$$p(f|\mathbf{x}) = \mathcal{N}(f|\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}))$$
$$\langle f(\mathbf{x}) \rangle = \mu(\mathbf{x})$$
$$\langle (f(\mathbf{x}) - \mu(\mathbf{x}))(f(\mathbf{x}') - \mu(\mathbf{x}')) \rangle = k(\mathbf{x}, \mathbf{x}')$$



Cf. Gauss確率分布 $\mathcal{N}(r|\mu, \sigma^2)$

確率変数 $r \in \mathbb{R}$ の確率分布が、平均 μ と分散 σ^2 で特徴づけられる正規分布をとる

$$p(r) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(r - \mu)^2}{2\sigma^2}\right)$$



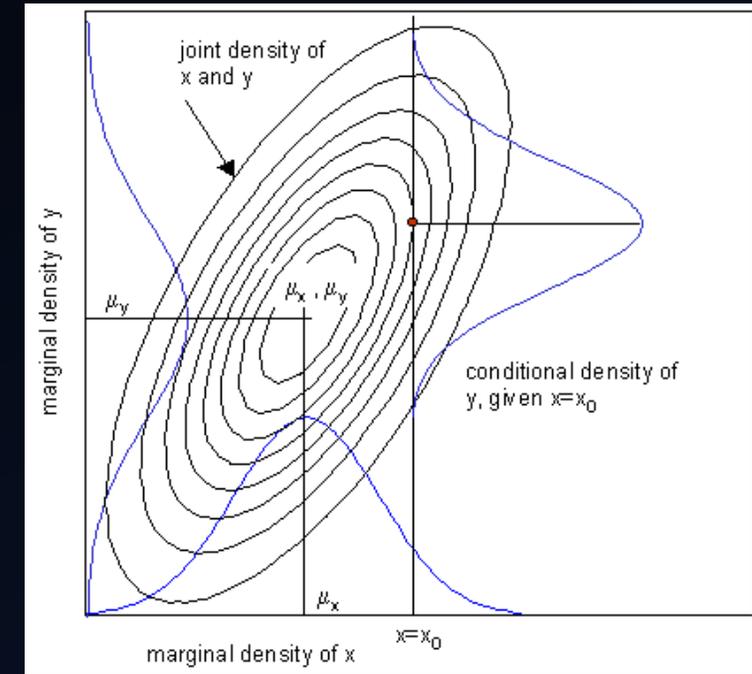
$$\langle r \rangle = \mu, \quad \langle (r - \mu)^2 \rangle = \sigma^2$$

補足：確率記号の読み方

$p(x, y)$. . . x, y の同時確率分布

$p(y|x)$. . . x が与えられた下での y の条件付確率分布

$p(y) = \int p(y|x)p(x)dx = \int p(x, y)dx$. . . y の周辺確率分布



$p(f|y, X)$. . . データセット y, x を利用して推定された関数 f の確率分布
(f は関数であって、値ではない)

$p(f_*|y, X, x_*)$. . . データセット y, x を利用して推定された関数に入力 x_* を入れたときの出力値 f_* の確率分布 (x_* はある一つの値と定めても、出力される f_* の値がばらつく。 f_* の値はどんな分布を取るか。)

$\mathcal{N}(f_*|\mu_*(x_*), \sigma_*^2(x_*))$. . . 出力される値 f_* は平均 μ_* 、分散 σ_*^2 の正規分布 (ただし、平均や分散の値は位置 x_* に依る)

2. Methods – (b)

(b) Multi-fidelity modelling with recursive Gaussian processes

Lowest (1st)-fidelity data: $D_1 = \{x_1, y_1\}$

2nd-fidelity data: $D_2 = \{x_2, y_2\}, \dots$

の様に多段のマルチフィデリティデータがあるとする。

従来の線形自己回帰モデル[Ref. 6, Kennedy & O'Hagan]は

$$f_t(x) = \rho f_{t-1}(x) + \delta_t(x), \quad (2.6)$$

ρ ↑ $f_{t-1}(x)$ ↑ $\delta_t(x)$ ↑
t段目のガウス過程回帰 (平均 μ_t ・共分散カーネル k_t)
一つ下のfidelityのデータに対するガウス過程回帰モデル
 y_t と y_{t-1} の間のスケーリング定数

このモデルでは、条件付き共分散について以下の関係が成り立つ。[6,9]

$$\text{cov}\{f_t(x), f_{t-1}(x') \mid f_{t-1}(x)\} = 0, \quad \forall x \neq x', \quad (2.7)$$

$f_{t-1}(x)$ ↑ $f_{t-1}(x')$ ↑
t-1段目のモデルについて $f_{t-1}(x)$ の情報が与えられた下で、
 $f_t(x)$ の推定に対して、異なる位置での $f_{t-1}(x')$ から得られる情報はもはやない。

2. Methods – (c)

(c) Nonlinear information fusion algorithms

(i) General formulation

自己回帰マルチフィデリティモデルを非線形化

$$f_t(x) = z_{t-1}(f_{t-1}(x)) + \delta_t(x), \quad (2.10)$$

z_{t-1} もGP事前分布とする。(多段のGP事前分布によるdeep GPとして一般化)

しかし、このままでは解析的に計算できず、計算コストも非常に高い。もう少し機能を制限しても、多少は解析的に扱えて計算量も標準的なGP回帰程度のアルゴリズムを考える。

(2.10)の様に多段のGP事前分布を設定して一度にdeep GPとして学習するのはやめて、(2.10)の f_{t-1} というGP事前分布の代わりに、 f_{*t-1} という前段のGP事後分布を用いたスキームを考える。つまり、

$$f_t(x) = g_t(x, f_{*t-1}(x)), \quad (2.11)$$

$$g_t \sim \mathcal{GP}(f_t | \mathbf{0}, k_t((x, f_{*t-1}(x)), (x', f_{*t-1}(x'))); \theta_t).$$

2. Methods – (c)

式(2.11)はたたき台としてはよさそうだが、1つ目の気づきは x_t と $f_{*t-1}(x_t)$ という異なる空間にあるデータの間での共分散関数をどのように選ぶべきかということ。

式(2.6)を参考に

$$k_{t_g} = k_{t_p}(x, x'; \theta_{t_p}) \cdot k_{t_f}(f_{*t-1}(x), f_{*t-1}(x'); \theta_{t_f}) + k_{t_\delta}(x, x'; \theta_{t_\delta}), \quad (2.12)$$

という形にすることにして、カーネル関数の関数形は以下で与えよう。

$$k_t(x, x'; \theta_t) = \sigma_t^2 \exp \left(-\frac{1}{2} \sum_{i=1}^d w_{i,t} (x_i - x'_i)^2 \right), \quad (2.13)$$

↑ 入力パラメータ空間 x の異方性を考慮したハイパーパラメータ (ARD weights)
↑ 分散の大きさを決めるハイパーパラメータ

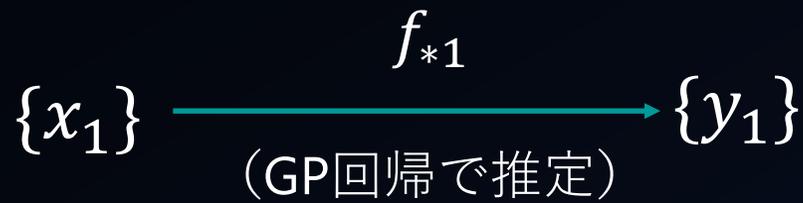
※決定しなければならないパラメータ数は、(k_{t_p} より $d+1$ 個) + (k_{t_f} より1個(σ^2 は k_{t_p} との積で共通化し、入力は f の1次元のみ)) + (k_{t_δ} より $d+1$ 個) = $2d+3$ 個

2. Methods – (c)

(ii) Prediction and propagation of uncertainty

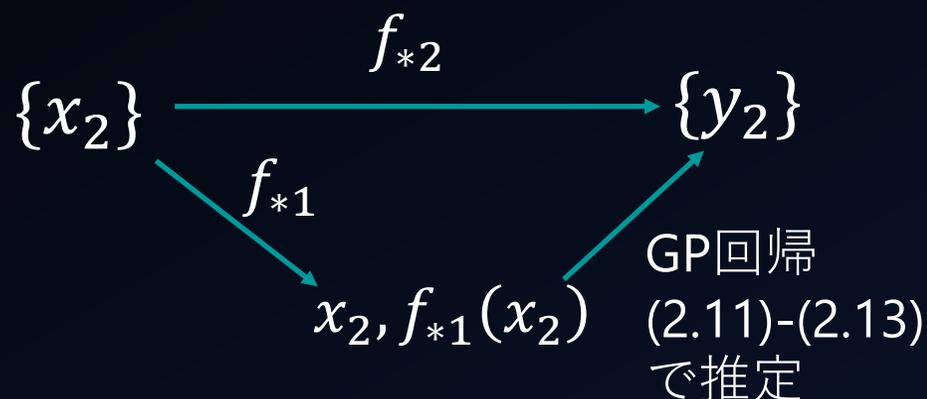
式(2.11)-(2.13)が提案アルゴリズム。

まず第1段では最低次fidelityデータに対する通常のGP回帰を実施。



推定される $f_{*1}(x)$ は式(2.4)の平均、式(2.5)の共分散を持つガウス分布となる。

第2段以降の推定 $f_{*t \geq 2}(x)$ はガウス分布とならないことに注意が必要。



ある1つの値 x を入れても、途中の $f_{*1}(x)$ はガウス分布ではらつくので、最終的な $f_{*2}(x)$ の確率分布を計算するには、ばらつき (uncertainty) の伝播を考慮する必要がある。

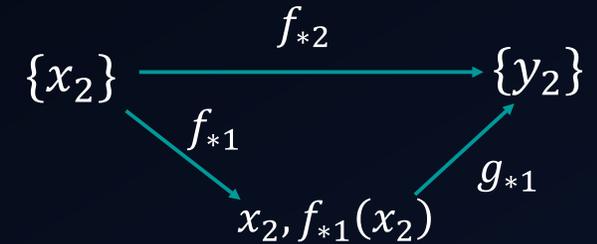
$$\begin{aligned} p(f_{*t}(x_*)) &:= p(f_t(x_*, f_{*t-1}(x_*)) | f_{*t-1}, x_*, y_t, x_t) \\ &= \int p(f_t(x_*, f_{*t-1}(x_*)) | y_t, x_t, x_*) p(f_{*t-1}(x_*)) dx_* \end{aligned} \quad (2.14)$$

第2段以降の平均や分散は式(2.14)に基づきMonte Carlo積分する。

注. おそらく式(2.14)は誤り

$$\begin{aligned}
 p(f_{*t}(x_*) &:= p(f_t(x_*, f_{*t-1}(x_*)) | f_{*t-1}, x_*, y_t, x_t) \\
 &= \int p(f_t(x_*, f_{*t-1}(x_*)) | y_t, x_t, x_*) p(f_{*t-1}(x_*)) dx_*,
 \end{aligned}
 \tag{2.14}$$

与えられた x_* に対し、低フィデリティ予測 $f_{*t-1}(x_*)$ が確率的に値を取るため、すべての取りうる f_{*t-1} について $g_t(x_*, f_{*t-1})$ の確率密度分布を計算して、その上で f_{*t-1} について積分する（周辺化する）ことで x_* のみの依存性として表すことが実際に行っている計算。



$$p(f_{*,t} | x_*) = \int p(g_t(x_*, f_{*,t-1}) | f_{*,t-1}, x_*) p(f_{*,t-1} | x_*) df_{*,t-1}$$

- $p(f_{*t-1} | x_*)$: 低フィデリティ予測 f_{*t-1} がとりうる値について、与えられた x_* の下での条件付き確率密度分布
- $p(g_t(x_*, f_{*t-1}) | f_{*t-1}, x_*)$: 拡張された入力空間で表された高フィデリティ予測関数 $g_t(x_*, f_{*t-1})$ がとりうる値について、与えられた x_*, f_{*t-1} の下での条件付き確率密度分布
- $p(f_{*t} | x_*)$: 高フィデリティ予測 f_{*t} がとりうる値について、与えられた x_* の下での条件付き確率密度分布。右辺の積分により低フィデリティ予測値 f_{*t-1} についての依存性が除去されている。

2. Methods – (c)

(iii) Workflow and computational cost

ステップ1.

最低次Fidelityデータ $\{x_1, y_1\}$ に対するGP回帰

$O(n_1^3)$ 程度の計算量で、推定関数 f_{*1} をGauss型確率分布として得る。

ステップ2.

2段目以降は式(2.11)に従い、 $\{\text{入力}(x_t, f_{*t-1}(x_t)), \text{出力}y_t\}$ というデータセットに対するGP回帰を行う。これも $O(n_1^3)$ 程度の計算量で、推定関数 $f_{*t}(x_*) = g_t(x_*, f_{*t-1}(x_*))$ を与える確率分布を得る。

ステップ3.

平均値や分散（もはや第2段以降は非Gauss分布なので、分散だけでは統計量を記述する上で不十分だが）を推定するには、式(2.14)に従いMonte Carlo積分する。

→多段の誤差の伝搬を十分な精度で計算しようとする、段数が増大するにしたがって計算量が指数的に増大してしまう。計算量を節約しなければならない場合は、2段目以降のすべての推定分布をGauss分布として近似して、t-1段とt段の間の誤差の伝搬だけ計算するようにする。

3. Results

3つの例題に取り組む。

- それぞれの例題では提案手法に対する異なる見方をハイライトする。
- 広い科学コミュニティに対して検証可能なベンチマークを提供する。

また、以下の3種類の回帰手法を比較する

- 標準的なGP回帰(GPと表記)
- 提案された非線形自己回帰マルチフィデリティGP回帰(NARGP)
- 古典的な線形マルチフィデリティGP回帰(AR1と表記, Ref. 6 by Kennedy & O'Hagan)

PythonのオープンソースライブラリGPyを利用して実装した。

19. GPy. 2002 GPy: A Gaussian process framework in python. See <http://github.com/SheffieldML/GPy>.

ちなみに、この論文で用いたコードも公開されている。

Data accessibility. The code and datasets supporting this article are available at <https://github.com/paraklas/NARGP>.

3. Results – (a)

(a) A pedagogical example

- Low-fidelity データ
正弦波, データ点数50点

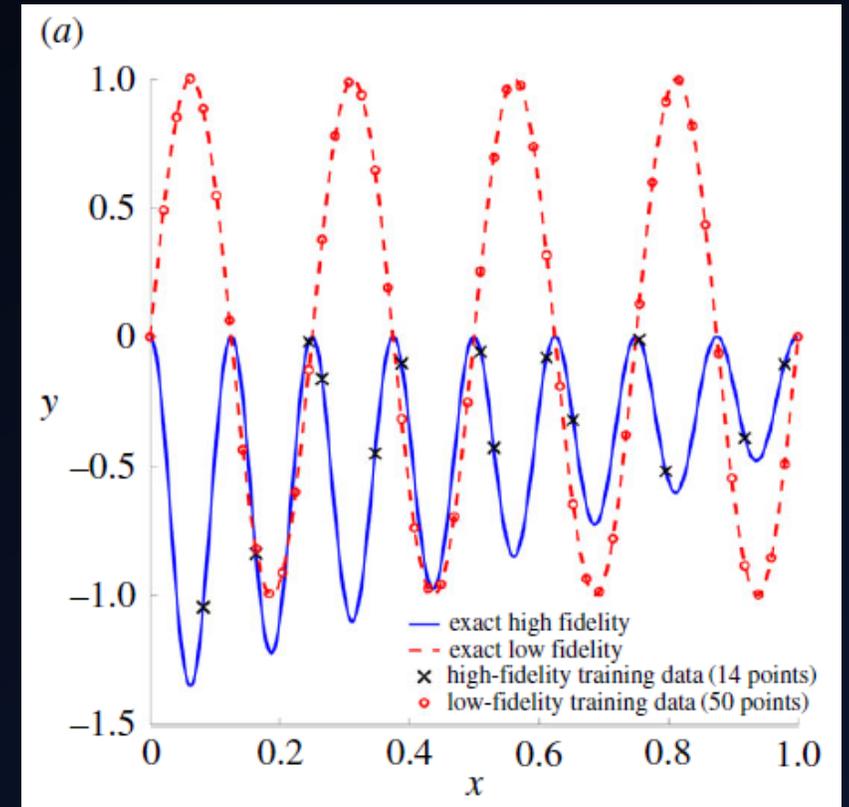
$$f_l(x) = \sin(8\pi x) \quad (3.1)$$

- High-fidelity データ
Low-fidelity関数 f_l を非一様(x依存)かつ非線形(2乗)に変換。データ点数14点 (low-fidelityの50か所の評価位置xの内から14か所)

$$f_h(x) = (x - \sqrt{2}) f_l^2(x). \quad (3.2)$$

目標としては、これらのデータからなるべく正確にHigh-fidelity関数（右図青線）を推定したい。

Figure 1a



3. Results – (a)

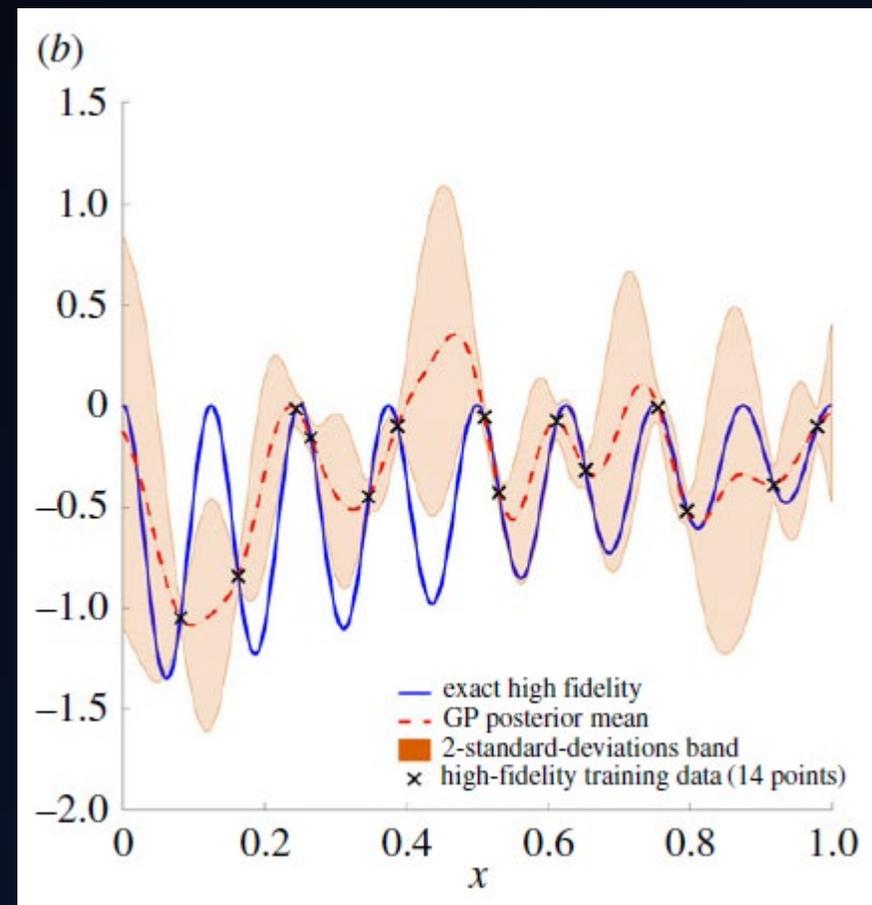
まずは、High-fidelityデータに対して通常のGP回帰を適用する。

十分データ数が多ければこれでも良く回帰できるはずだが、僅か14点(×)では推定したい関数形（青線—）を表わすには粗すぎる。

→ 通常のGP回帰での推定（赤破線--）は全然ダメ。

※ただし、GP回帰のカーネル関数を上手く選ぶと改善されうる。この例では線形+周期関数カーネル。

Figure 1b



3. Results – (a)

Figure 2a は、提案手法である NARGPでの推定。

- High-fidelityデータが無いところもよく推定できている。
- Low-fidelityデータと推定したい真の関数（青線）が異なるトレンドを持っているような領域(例えば $0.25 < x < 0.35$)でさえも良い推定。

NARGPのカーネル式(2.12)により、非線形かつ空間依存のlow-とhigh-fidelityデータ間の相関についても適切に学習できているため(Figure 3a)。

→ Low-fidelityによる誤ったトレンドを防ぎつつも、Low-fidelityの内の役立つ情報は抜き出す。

Figure 2a

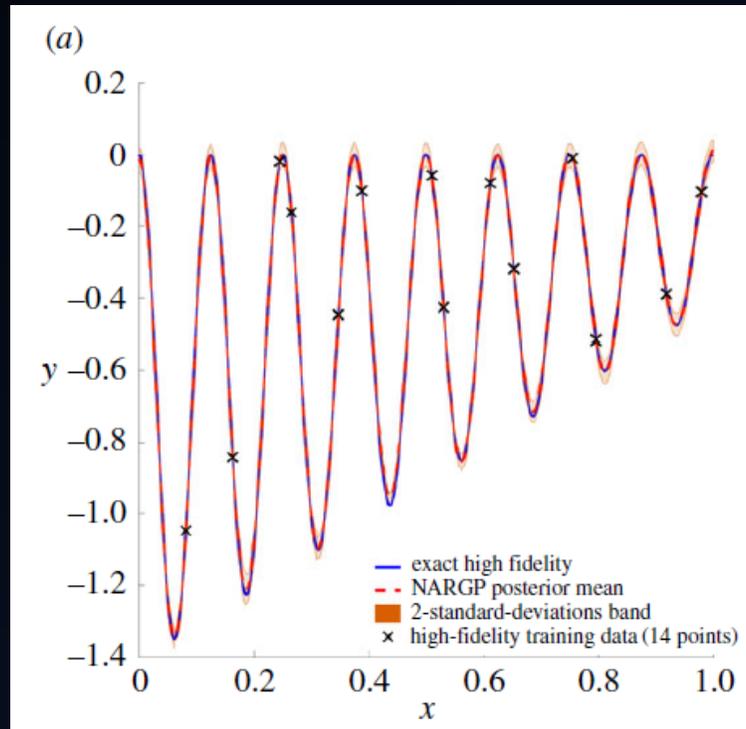
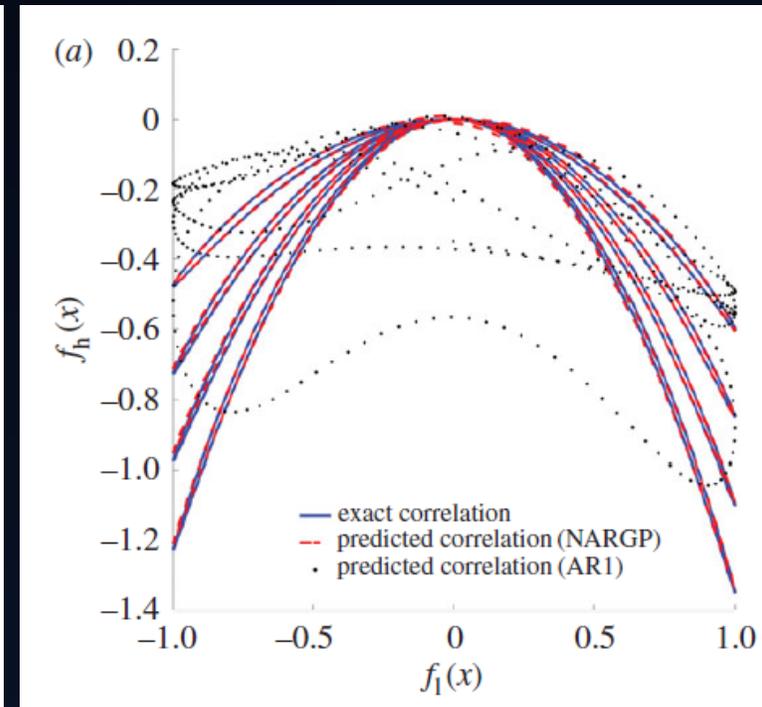


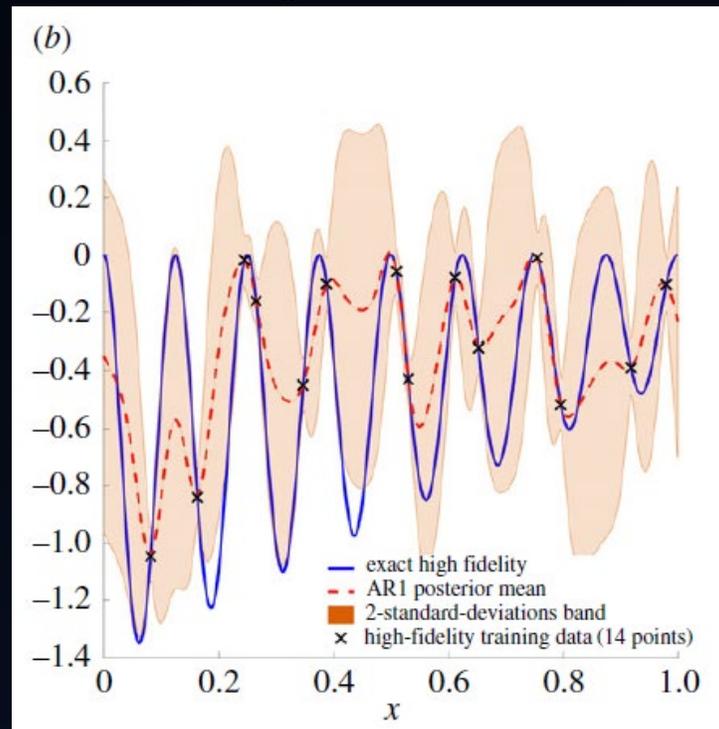
Figure 3a



3. Results – (a)

Figure 2b の古典的線形AR1では、あまり良い推定ができていない。AR1ではLow-とHigh-fidelityデータ間相関も正しく抽出できていない(Figure 3a)。

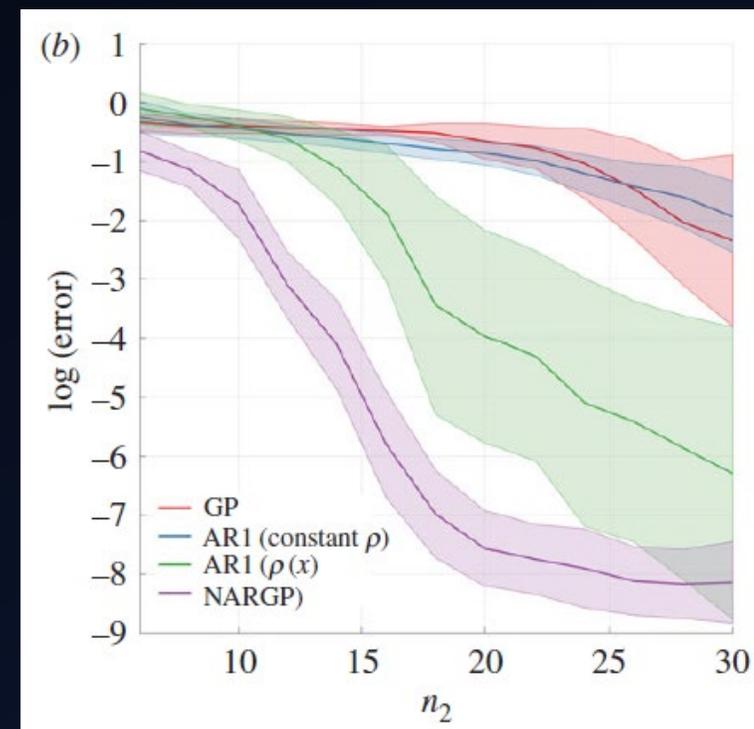
Figure 2b



High-fidelityデータ点数に対する収束性を見ると、

- NARGPが最も早く収束している様子が分かる。
- AR1の係数に空間依存性を持たせたAR1- $\rho(x)$ の場合は、先ほどのAR1-constant ρ より改善。
- 通常のGP回帰の収束はまだ先。

Figure 3a



3. Results – (a)

NARGPでは

$$f_h(x) = g_h(x, f_l(x))$$

として、

High-fidelity データの関係 $x \rightarrow f_h(x)$ や Low-と High-fidelity データの相関 $f_l(x) \rightarrow f_h(x)$ を別々に扱うのではなく、

$x, f_l(x) \rightarrow f_h(x)$ という対応関係を非局所(x依存)・非線形関数として表現。

Figure 3aのような複雑なLow-とHigh-fidelityデータの相関をほどこき、滑らかな局面 \mathcal{G} として写像を推定する。

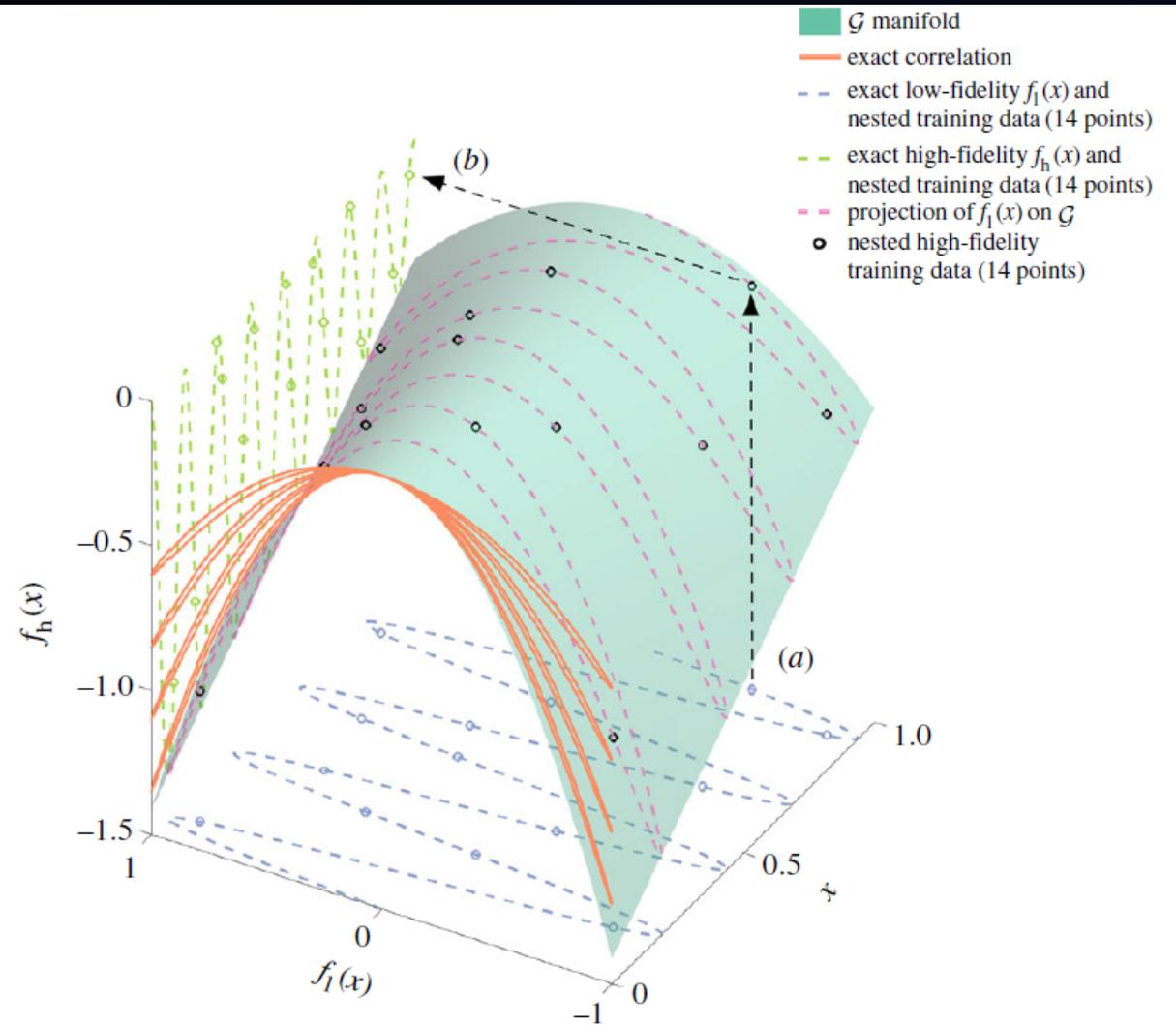


Figure 4. A pedagogical example: the NARGP algorithm can capture complex nonlinear, non-functional and space-dependent cross-correlations by inferring the nonlinear latent manifold \mathcal{G} that governs the functional relation between the inputs x and the outputs of the low- and high-fidelity models $f_l(x)$ and $f_h(x)$, respectively. (a) The low-fidelity model is projected onto the nonlinear latent manifold \mathcal{G} that is inferred using the deep non-parametric representation of equation (2.11). (b) The high-fidelity function $f_h(x)$ is recovered by a smooth mapping from the \mathcal{G} manifold to the high-fidelity data. (Online version in colour.)

3. Results – (b)

(b) Multi-fidelity approximation of the Branin function

入力xが2次元かつ、フィデリティも3段階としたテスト問題。

High-fidelity (Branin関数)

$$f_3(x) = \left(\frac{-1.275x_1^2}{\pi^2} + \frac{5x_1}{\pi} + x_2 - 6 \right)^2 + \left(10 - \frac{5}{4\pi} \right) \cos(x_1) + 10, \quad (3.3)$$

Medium-fidelity

$$f_2(x) = 10\sqrt{f_3(x-2)} + 2(x_1 - 0.5) - 3(3x_2 - 1) - 1 \quad (3.4)$$

Low-fidelity

$$f_1(x) = f_2(1.2(x+2)) - 3x_2 + 1. \quad (3.5)$$

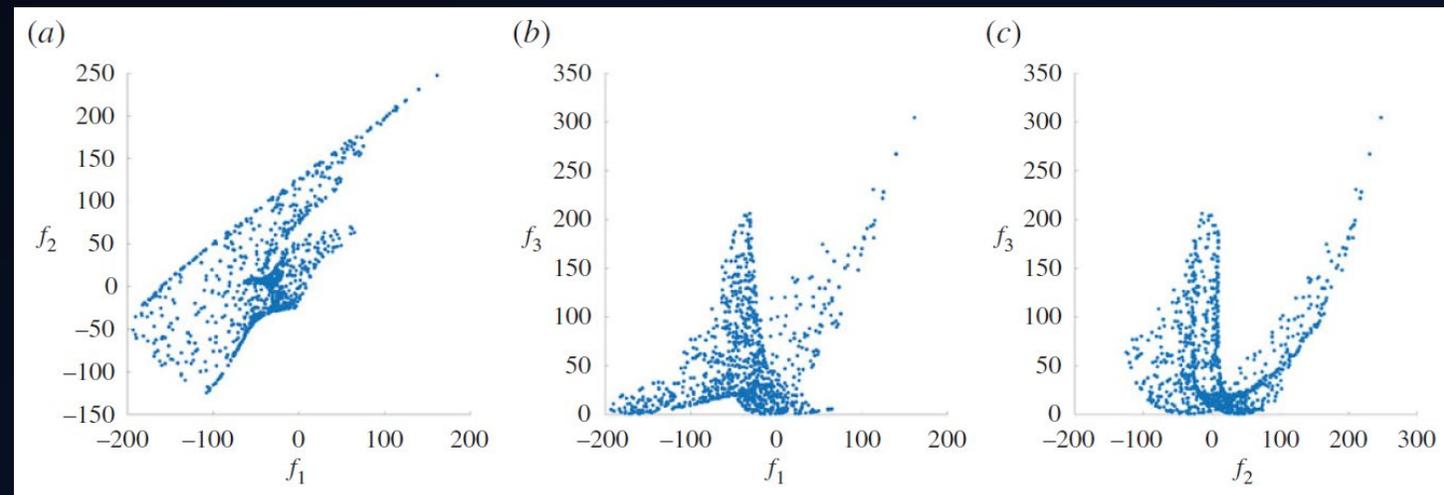
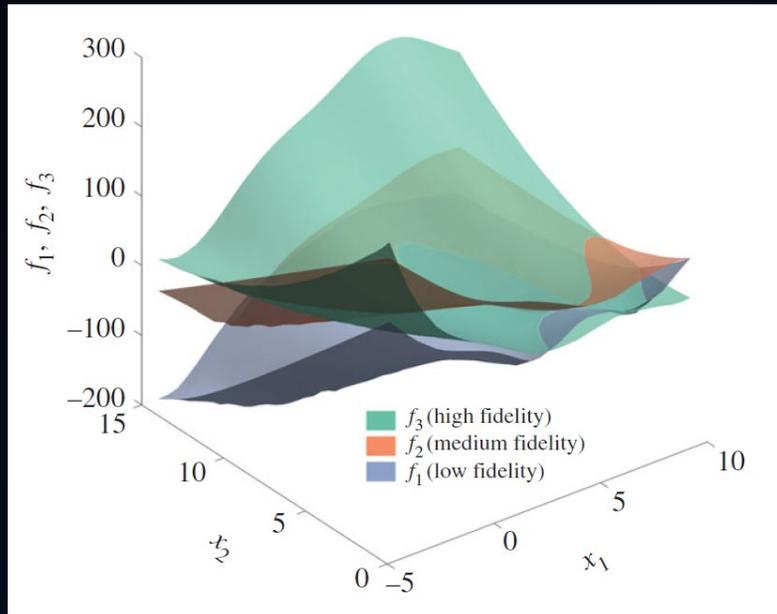


Figure 6 各関数間の相関

3. Results – (b)

ランダムにサンプリングしたLow-fidelityデータ $n_1=80$ 点, Medium-fidelityデータ $n_2=40$ 点, High-fidelityデータ $n_3=20$ 点を作成。

目標としては、これらを用いてなるべく正確にHigh-fidelity (Branin)関数を推定したい。
→ NARGPによる推定はFigure 7。Branin関数をよく推定している(a)し、推定の標準偏差(b)は、誤差の大きいところかどこか(c)おおよそ示している。

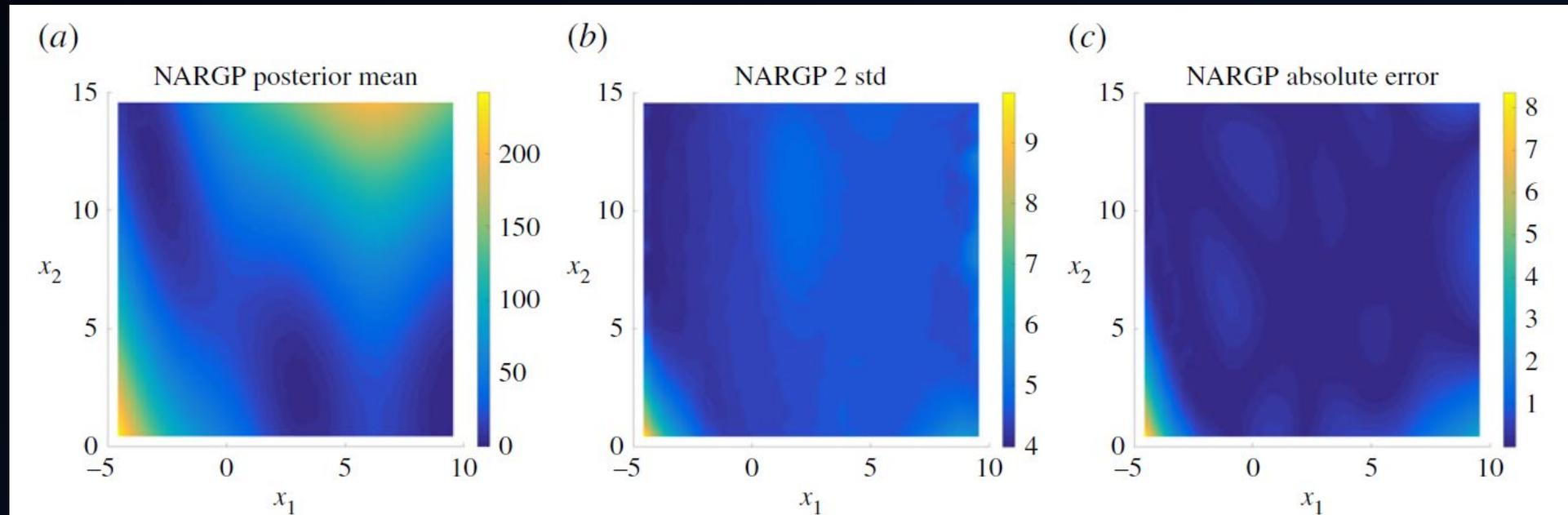
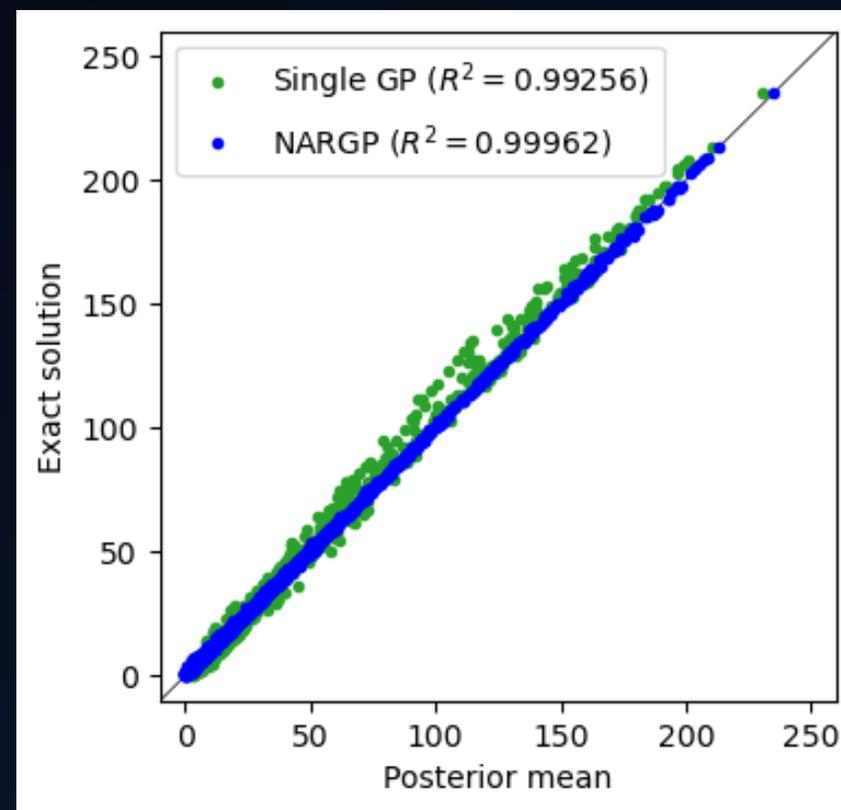
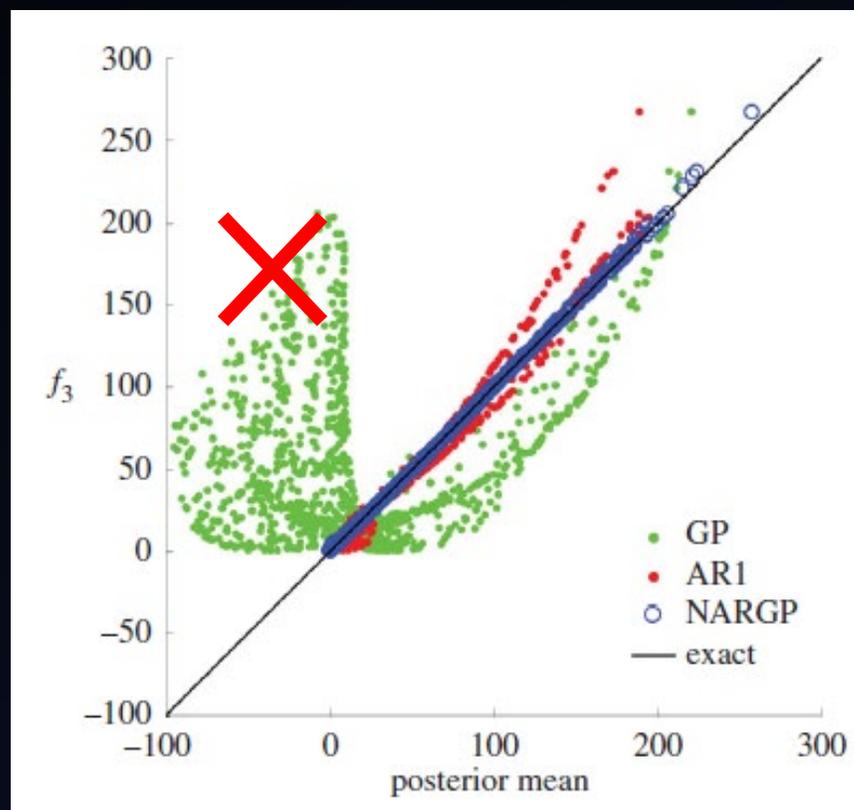


Figure 7. Multi-fidelity approximation of the Branin function: (a) NARGP posterior mean, (b) 2 standard deviations and (c) absolute point-wise error with respect to the exact solution f_3 . (Online version in colour.)

3. Results – (b)

Figure 8で回帰プロットを示しているが、この単一GPの結果としてプロットされている緑の点は、論文の結果が誤っている（コード中のバグ）。正しくやってみたところ、右図のようになる。

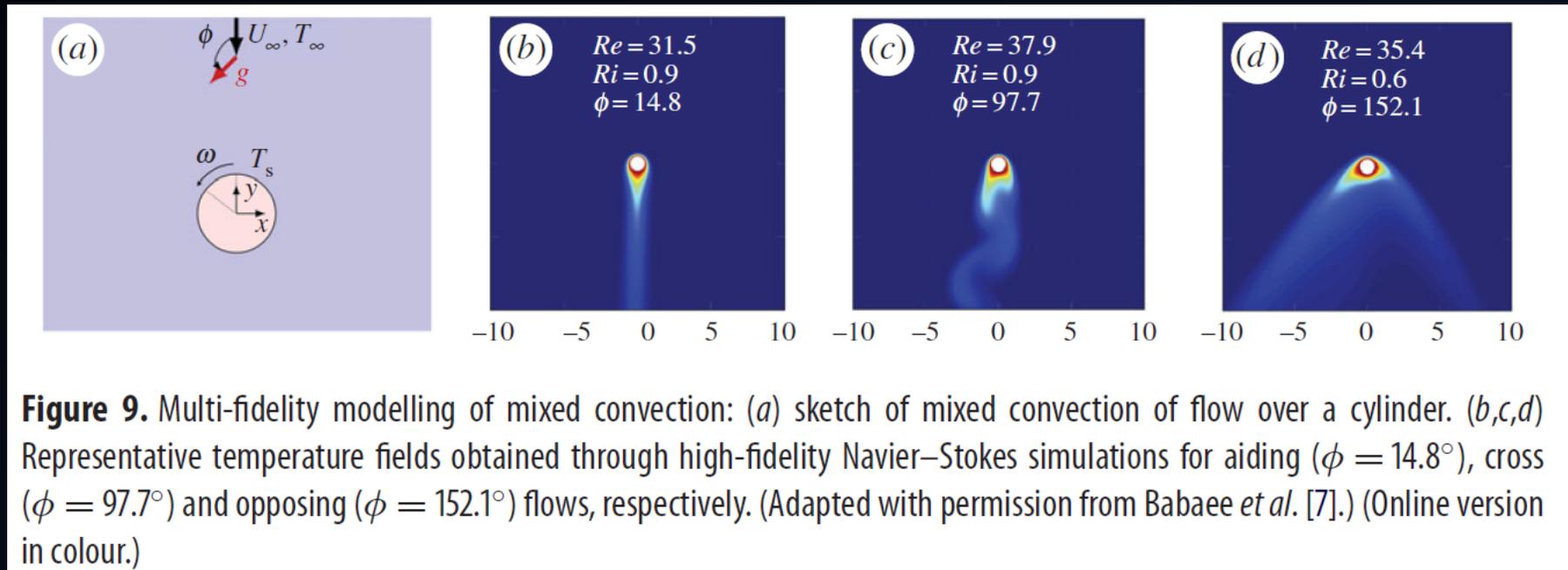
Figure 8（誤り）



3. Results – (c)

(c) Multi-fidelity modelling of mixed convection based on experimental correlations and numerical simulations

実問題への応用として、対流混合の問題（加熱円柱による自然対流中に、角度 ϕ で強制対流を加える。円柱から流体への熱の伝わり方はどのように変わるか？）に適用してみよう。



※自然対流 ... 重力 g 下の静止流体中に高温の物体を置くと、物体に接する流体が加熱されて浮力が生じて発生する上昇流のこと。

3. Results – (c)

$$\text{Nusselt数 } Nu = \frac{\text{対流による熱伝達率 (どのような流れができるかに依存)}}{\text{熱伝導率}}$$

対流混合問題では、Reynolds数 Re とRichardson数 Ri と強制対流の角度 Φ に依存 $Nu = f(Re, Ri, \phi)$

Ref. [7]では、Low-fidelityデータとして経験則 f_l を、High-fidelityデータとしてNavier-Stokes乱流の直接数値シミュレーションデータ f_h を用いて、従来の線形AR1-constant ρ モデルによるマルチフィデリティモデリングを実施。

→Cross-flow ($\Phi \sim 90^\circ$)あたりで推定が悪く、Ref.[7]の著者は、「このパラメータ領域ではLow-fidelityデータは有用ではなく、High-fidelityデータが不可欠である」と結論付けた。

動機：上記の推定の悪さは、用いたAR1モデルの推定能力の限界によるもので、NARGPを用いれば適切なLow-とHigh-fidelityデータ相関の情報を抽出できるのでは？

※Reynolds数 Re ... 粘性散逸に対する慣性（対流）の比 $Re = UL/\nu$

※Richardson数 Ri ... 強制対流（慣性）に対する自然対流（浮力）の比 $Ri = g\beta\Delta TL/U^2$

3. Results – (c)

Ref. [7]と同じデータセットに対し、GP, AR1, NARGPによる推定をそれぞれ適用

全データセット：Low-fidelityデータ1000点、High-fidelityデータ300点

学習に用いたデータ数（ランダムサンプル）：

Low-fidelityデータ200点、High-fidelityデータ18点から36点

学習に用いなかったHigh-fidelityデータは推定性能の検証用に利用。

→ NARGPはGP, AR1モデルよりも、少ないデータ数でよい推定。

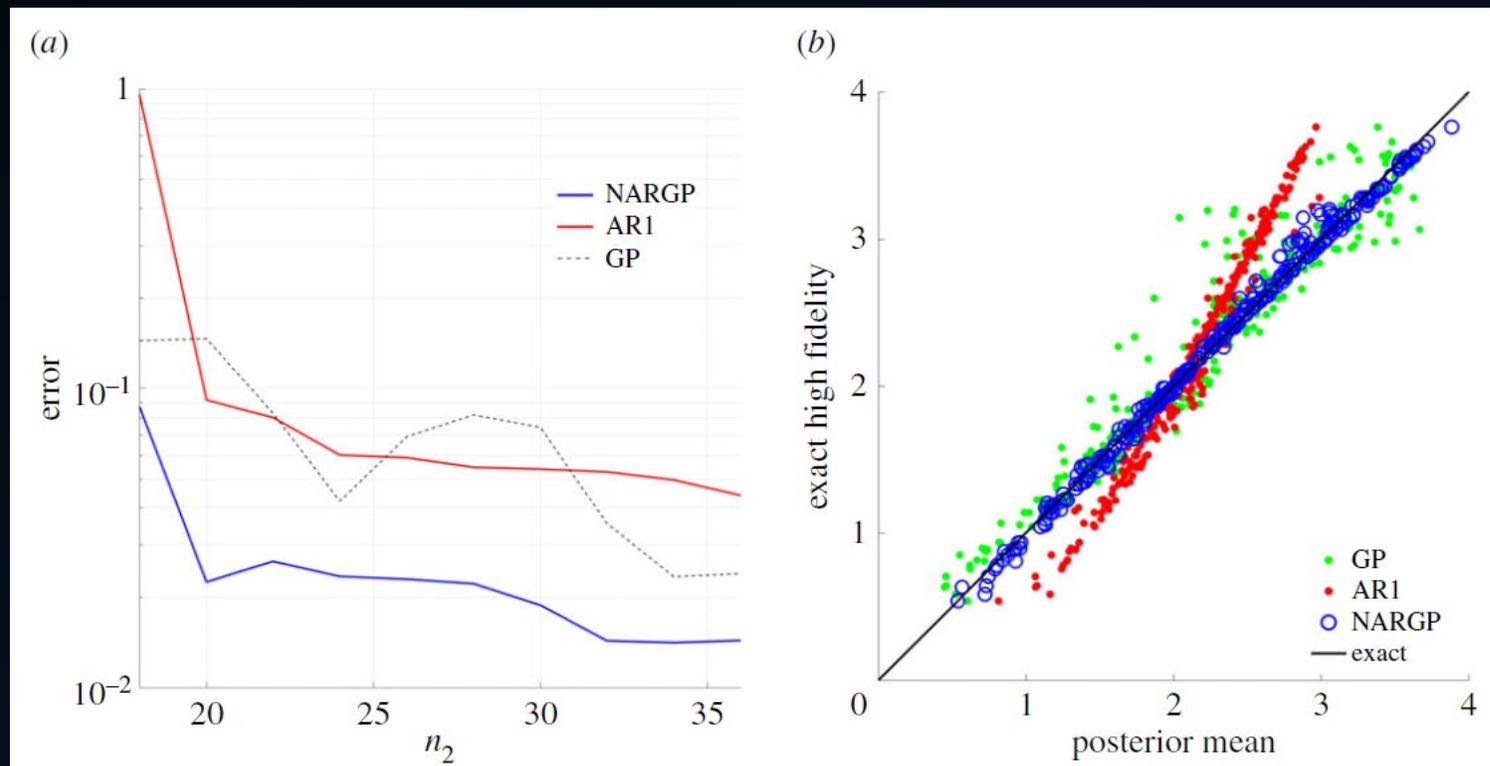


Figure 10

4. Conclusion

非線形自己回帰Gauss過程回帰を提案した。

(Nonlinear Auto-Regressive Gaussian Process regression, NARGP)

低精度だが多数のLow-fidelityデータと高精度だが少数のHigh-fidelityデータ（またはさらに多段のMulti-fidelityデータ）を融合して、求めたい高精度関数を推定するマルチフィデリティモデリング手法。

各段での学習を分離した式(2.11)の形にすることで計算コストを通常のGP回帰と同程度に抑える。入力 x とLow-fidelity出力 f_l という異なるデータを扱うカーネル関数として式(2.12)を提案し、それにより入力 x に対する空間非一様性もLow-とHigh-fidelityデータ間の非線形相関も扱えるという拡張を可能とした。

今回のテストでは学習データにノイズのない場合のみに限定した（それでも十分実用的）学習データにノイズがある場合に対しても、個々の段における学習をノイズを考慮したGP回帰（heteroscedastic GP, stochastic variational inference GP, warped GP）などに置き換えることで拡張できるだろう。