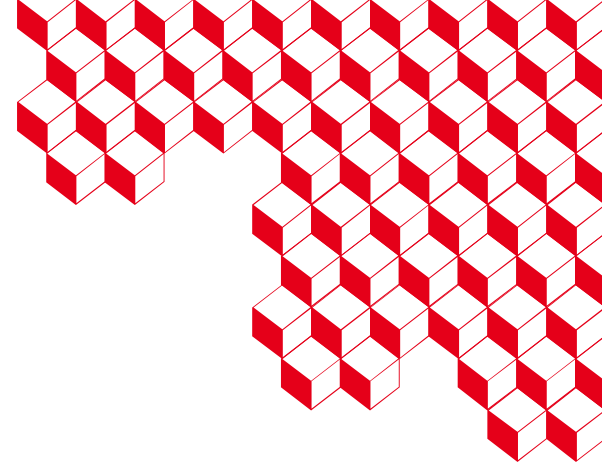




Prediction of physical fields under linear equality constraints

ETICS Annual Research Schools, October the 2nd, 2026

NASSOURADINE MAHAMAT HAMDAN^{1,2}

Supervisors : Clément Gauchy¹, Pierre-Emmanuel Angeli¹

PhD Director : Sébastien Da Veiga²

mahamat-hamdan.nassouradine@cea.fr

CEA Saclay¹ ENSAI CREST²

-

Figure: Scheme of the experiment carried out in Buice and Eaton, 2000.

Motivation

- Verification & Validation of CFD simulators (VVUQ)

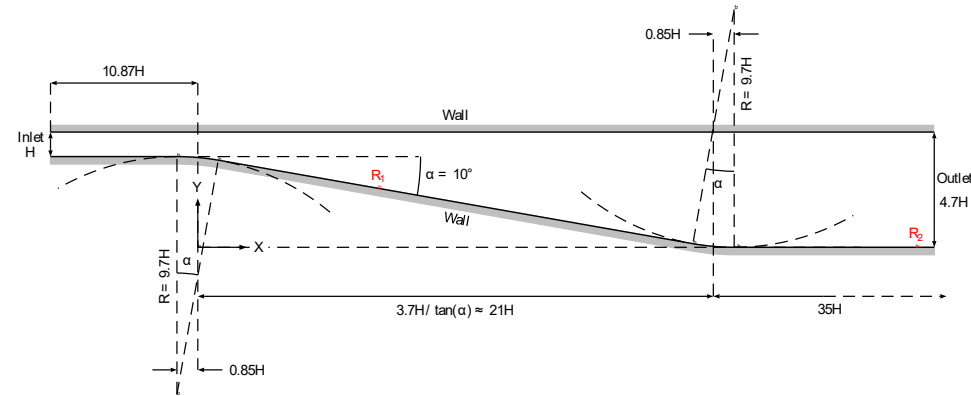


Figure: Scheme of the experiment carried out in Buice and Eaton, 2000.



Computer code

$$\mathcal{M} : \mathcal{X} \rightarrow \mathbb{R}^P$$

(RANS solver with Trio-CFD)

Motivation

- Verification & Validation of CFD simulators (VVUQ)

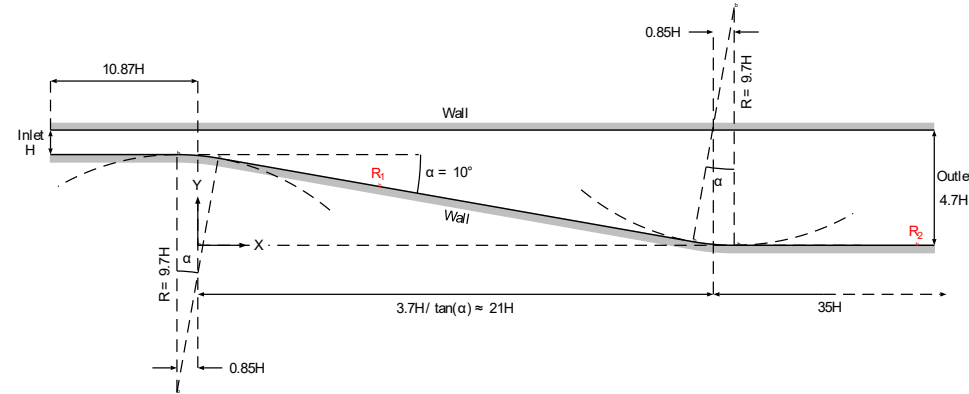


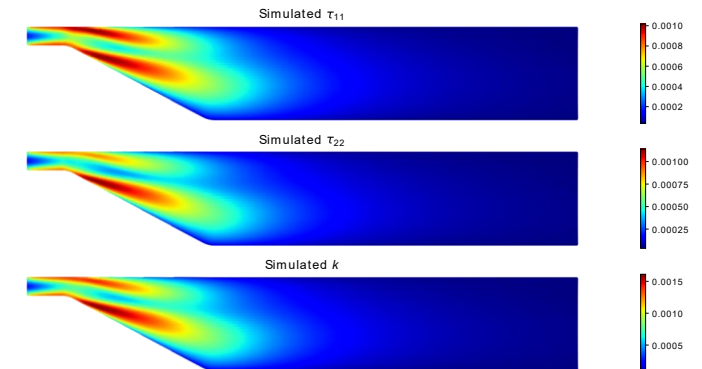
Figure: Scheme of the experiment carried out in Buice and Eaton, 2000.

Computer code

$$\mathcal{M} : \mathcal{X} \rightarrow \mathbb{R}^P$$

(RANS solver with Trio-CFD)

Outputs $\in \mathbb{R}^S$
 $S \approx 140\,000$: mesh size



Motivation

- Verification & Validation of CFD simulators (VVUQ)

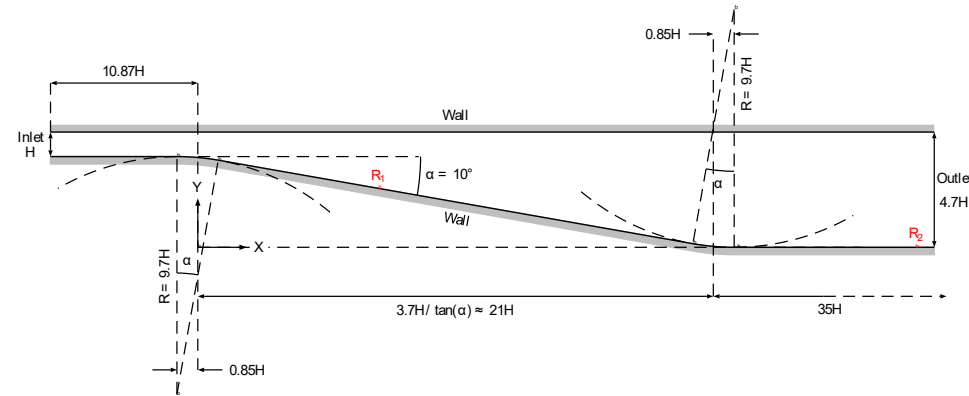


Figure: Scheme of the experiment carried out in Buice and Eaton, 2000.

Input

Closure coefficients
of the **RANS** model:
 $\mathcal{X} = (C_\mu, \sigma_k, \sigma_\varepsilon, C_{\varepsilon_1}, C_{\varepsilon_2})$

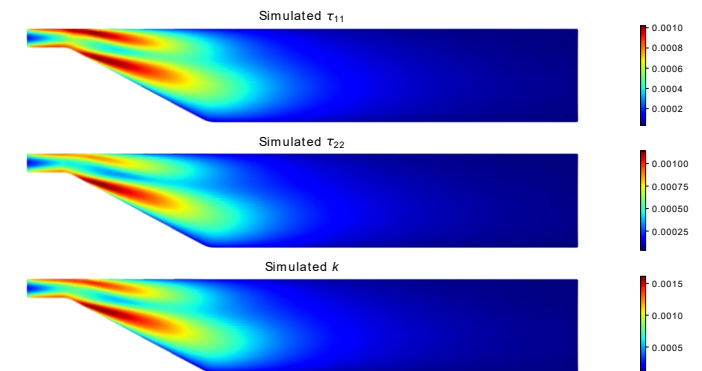


Computer code

$$\mathcal{M} : \mathcal{X} \rightarrow \mathbb{R}^P$$

(**RANS** solver with **Trio-CFD**)

Outputs $\in \mathbb{R}^S$
 $S \approx 140\,000$: mesh size



Motivation

- Verification & Validation of CFD simulators (VVUQ)

➡ Goal: Validation of $\mathcal{M}$

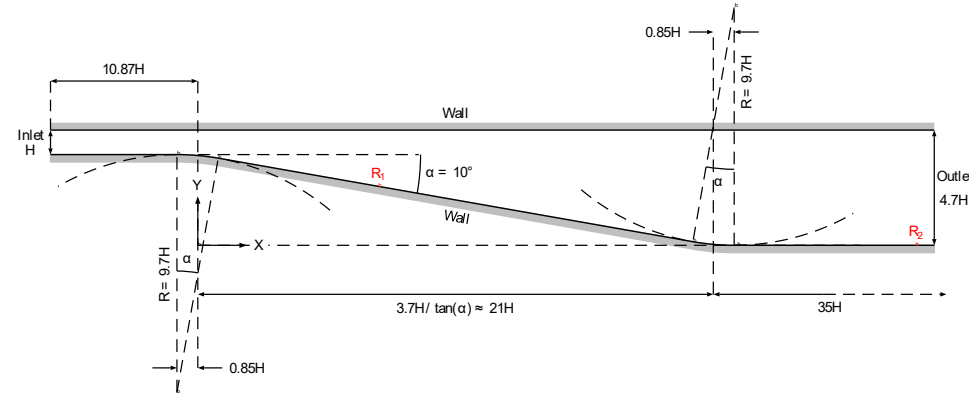


Figure: Scheme of the experiment carried out in Buice and Eaton, 2000.

Input

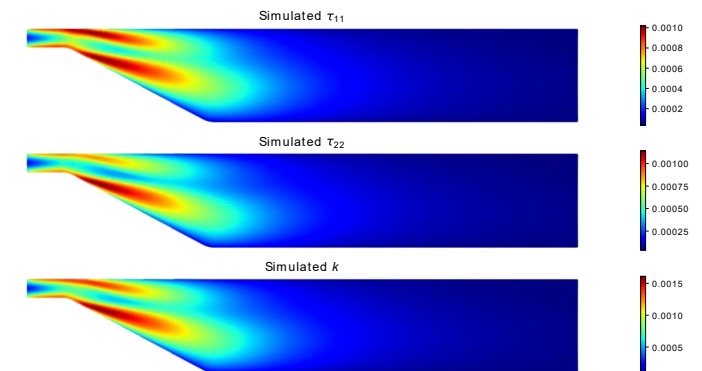
Closure coefficients
of the **RANS** model:
 $\mathcal{X} = (C_\mu, \sigma_k, \sigma_\varepsilon, C_{\varepsilon_1}, C_{\varepsilon_2})$

Computer code

$$\mathcal{M} : \mathcal{X} \rightarrow \mathbb{R}^P$$

(**RANS** solver with **Trio-CFD**)

Outputs $\in \mathbb{R}^S$
 $S \approx 140\,000$: mesh size



Motivation

- Verification & Validation of CFD simulators (VVUQ)

➡ **Goal: Validation of $\mathcal{M}$**

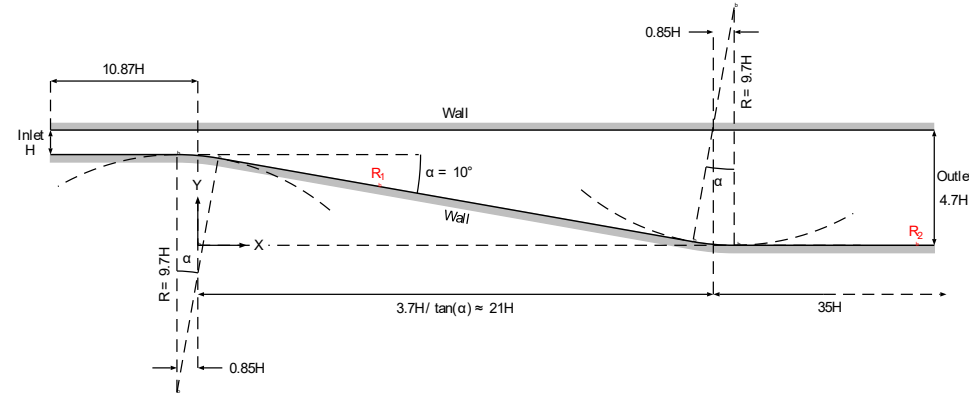


Figure: Scheme of the experiment carried out in Buice and Eaton, 2000.

Input

Closure coefficients
of the **RANS** model:
 $\mathcal{X} = (C_\mu, \sigma_k, \sigma_\varepsilon, C_{\varepsilon_1}, C_{\varepsilon_2})$

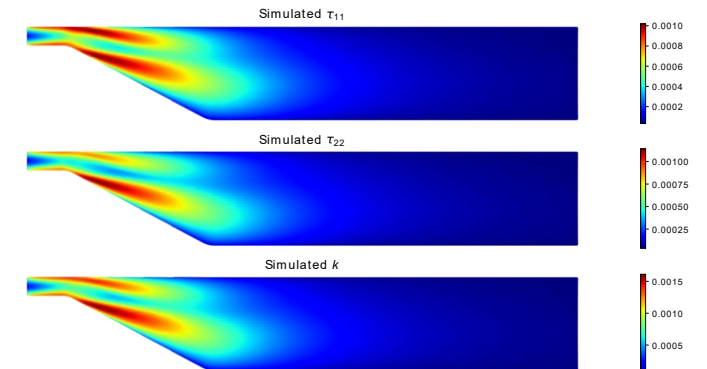
Computer code

$$\mathcal{M} : \mathcal{X} \rightarrow \mathbb{R}^P$$

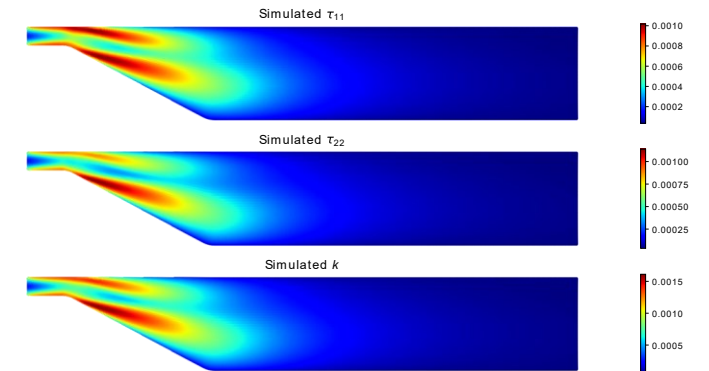
(**RANS** solver with **Trio-CFD**)

× **Computationally expensive !**

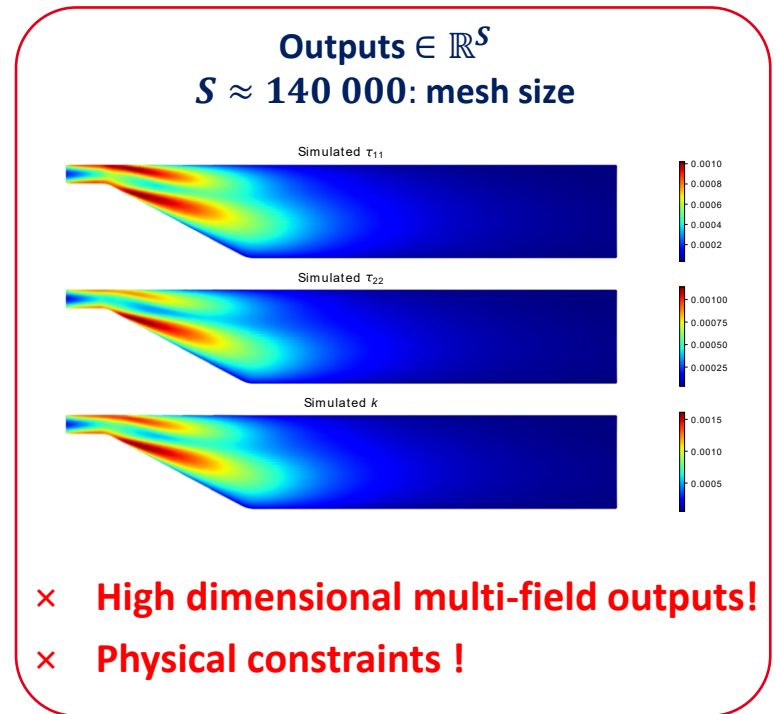
Outputs $\in \mathbb{R}^S$
 $S \approx 140\,000$: mesh size



- 👉 **Goal: Validation of $\mathcal{M}$**



- 👉 **Goal: Validation of $\mathcal{M}$**



Physics constraints for incompressible flows

Many physics problems are governed by **linear constraints**, for example:

- The law of conservation of mass implies **linear relationships** between the concentrations or proportions of different chemical species in a given solution

Physics constraints for incompressible flows

Many physics problems are governed by **linear constraints**, for example:

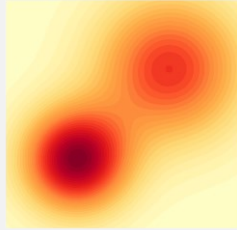
- The law of conservation of mass implies **linear relationships** between the concentrations or proportions of different chemical species in a given solution
- The incompressibility condition in fluid dynamics leads **to linear constraints** on the velocity field or, equivalently, a zero-trace stress or strain tensor

$$\begin{array}{c} \mathbf{y}(\mathbf{x}) = [\boldsymbol{\tau}_{11}(\mathbf{x}), \boldsymbol{\tau}_{22}(\mathbf{x}), k(\mathbf{x})] \\ P = 3 \times S \end{array} \quad \left| \quad \begin{array}{cccc} \begin{array}{c} \text{Heatmap of } \tau_{11}(\mathbf{x}) \end{array} & + & \begin{array}{c} \text{Heatmap of } \tau_{22}(\mathbf{x}) \end{array} & - \frac{4}{3} \begin{array}{c} \text{Heatmap of } k(\mathbf{x}) \end{array} & = & \begin{array}{c} \text{Heatmap of } \mathbf{0}_S \end{array} \\ \boldsymbol{\tau}_{11}(\mathbf{x}) \in \mathbb{R}^S & & \boldsymbol{\tau}_{22}(\mathbf{x}) \in \mathbb{R}^S & & k(\mathbf{x}) \in \mathbb{R}^S & & \mathbf{0}_S \in \mathbb{R}^S \end{array}$$

Reynolds stress tensor $\tau_{ij} = \overline{u'_i u'_j}$ and the turbulent kinetic energy k

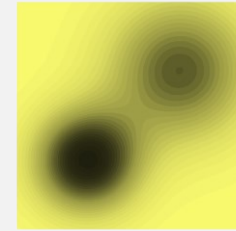
Linear constraints from physics

$$y(x) = [y^1(x), \dots, y^Q(x)]$$
$$P = Q \times S$$



$$y^1(x) \in \mathbb{R}^S$$

...



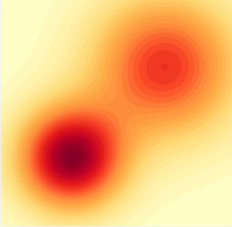
$$y^Q(x) \in \mathbb{R}^S$$

In this work, we are considering:

- An observed **Q-dimensional** vector field output on a fixed mesh of size **S** : $y_i = (y_i^1, \dots, y_i^Q) \in (\mathbb{R}^S)^Q$
- The vector output regression function to estimate: $f = (f^1, \dots, f^Q): \mathcal{X} \rightarrow (\mathbb{R}^S)^Q$

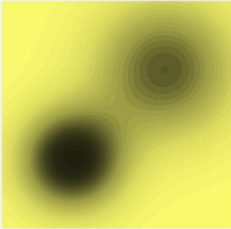
Linear constraints from physics

$$\mathbf{y}(x) = [\mathbf{y}^1(x), \dots, \mathbf{y}^Q(x)] \quad P = Q \times S$$

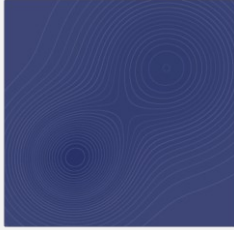
$\alpha^Q(x)$  $\mathbf{y}^1(x) \in \mathbb{R}^S$

$+$

$\dots$

$+$ $\alpha^Q(x)$  $\mathbf{y}^Q(x) \in \mathbb{R}^S$

$=$

 $\mathbf{c}(x) \in \mathbb{R}^S$

In this work, we are considering:

- An observed **Q-dimensional** vector field output on a fixed mesh of size **S** : $\mathbf{y}_i = (y_i^1, \dots, y_i^Q) \in (\mathbb{R}^S)^Q$
- The vector output regression function to estimate: $f = (f^1, \dots, f^Q): \mathcal{X} \rightarrow (\mathbb{R}^S)^Q$
- The class of linear constraints defined through the sum of the Q scalar fields $f^j(.)$:

$$\mathcal{F}[f(x)] = \sum_{j=1}^Q \alpha^j(x) f^j(x) = c(x) \quad (1)$$

where $\alpha^j(.)$ are input-dependent scalar weights, $f^j(.)$ and $c(.)$ are scalar fields. We assume α^j and **c are known**.

Problem statement

Simultaneous modeling of linearly constrained fields

We are considering the multi-output regression task of finding $f = (f_1, \dots, f_Q): \mathcal{X} \rightarrow \mathcal{Y}$, such that:

$$\left\{ \begin{array}{l} y = f(x), \quad y \text{ is a high dimensional tensor, i.e. } P \approx 10^6 \\ \mathcal{F}[f(x)] = \sum_{j=1}^Q \alpha^j(x) f^j(x) = c(x) \end{array} \right.$$

Problem statement

Simultaneous modeling of linearly constrained fields

We are considering the multi-output regression task of finding $f = (f_1, \dots, f_Q): \mathcal{X} \rightarrow \mathcal{Y}$, such that:

$$\begin{cases} y = f(x), & y \text{ is a high dimensional tensor, i.e. } P \approx 10^6 \\ \mathcal{F}[f(x)] = \sum_{j=1}^Q \alpha^j(x) f^j(x) = c(x) \end{cases}$$

Main mathematical tools:

- **Multi-output Gaussian process regression^[1]**
 - ⇒ Suited for small data context
 - ⇒ Provides naturally uncertainty quantification with the prediction
 - ⇒ Can deal with constrained problems
 - ✗ Not suited for high dimensional output
- **Principal component analysis**
 - ⇒ Linear and explicit reconstruction error

Handling linear equality constraints

$$\sum_{j=1}^Q \alpha^j(x) f^j(x) = c(x)$$

Handling linear equality constraints

$$\sum_{j=1}^Q \alpha^j(x) f^j(x) = c(x)$$

Principle of deductive approach

- Select **arbitrary** one output f^l
- Build **unconstrained surrogates** ($\hat{f}^j, j \neq l$) on the remaining Q-1 outputs (independent or joint modeling)

- Deduce f^l prediction using constraint equation:

$$\hat{f}^l = \frac{1}{\alpha^l(x)} \left(c(x) - \sum_{j \neq l} \alpha^j(x) \hat{f}^j(x) \right)$$

Handling linear equality constraints

$$\sum_{j=1}^Q \alpha^j(x) f^j(x) = c(x)$$

Principle of deductive approach

- Select **arbitrary** one output f^l
- Build **unconstrained surrogates** ($\hat{f}^j, j \neq l$) on the remaining $Q-1$ outputs (independent or joint modeling)

- Deduce f^l prediction using constraint equation:

$$\hat{f}^l = \frac{1}{\alpha^l(x)} \left(c(x) - \sum_{j \neq l} \alpha^j(x) \hat{f}^j(x) \right)$$

Parametrization approach^[2] : Constrained GP

- Rewrite the linear constraint on f : $\mathcal{L}[f] = 0$, $\mathcal{L}$ linear operator
- Suppose that: $f(x) = \mathcal{G}_x[g]$
- Then we impose the constraint on the linear operator $\mathcal{G}_x$:

$$\mathcal{L}[\mathcal{G}_x] = 0$$

- Gaussian processes are **closed** under linear transformation

$$g \sim \mathcal{GP}(\mu_g(x), k_g(x, x'))$$

$$\mu_f(x) = \mathcal{G}_x[\mu_g], \quad k_f = \mathcal{G}_x k_g \mathcal{G}_x^T$$

[2] Jidling, C., Wahlström, N., Wills, A., & Schön, T. B. (2017)

Constrained Gaussian Process regression

Given a linear constraint on f : $\mathcal{L}[f] = 0$, $\mathcal{L}$ linear operator

Parametrization approach^[2]:

- Suppose that: $f(x) = \mathcal{G}_x[g]$
- Then we impose the constraint on the operator $\mathcal{G}_x$:

$$\mathcal{L}[\mathcal{G}_x] = 0$$

- Gaussian processes are **closed** under linear transformation

$$g \sim \mathcal{GP}(\mu_g(x), k_g(x, x'))$$

$$\mu_f(x) = \mathcal{G}_x[\mu_g], \quad k_f = \mathcal{G}_x k_g \mathcal{G}_x^T$$

[2] Jidling, C., Wahlström, N., Wills, A., & Schön, T. B. (2017)

Constrained Gaussian Process regression

Given a linear constraint on f : $\mathcal{L}[f] = 0$, $\mathcal{L}$ linear operator

Parametrization approach^[2]:

- Suppose that: $f(x) = \mathcal{G}_x[g]$
- Then we impose the constraint on the operator $\mathcal{G}_x$:

$$\mathcal{L}[\mathcal{G}_x] = 0$$

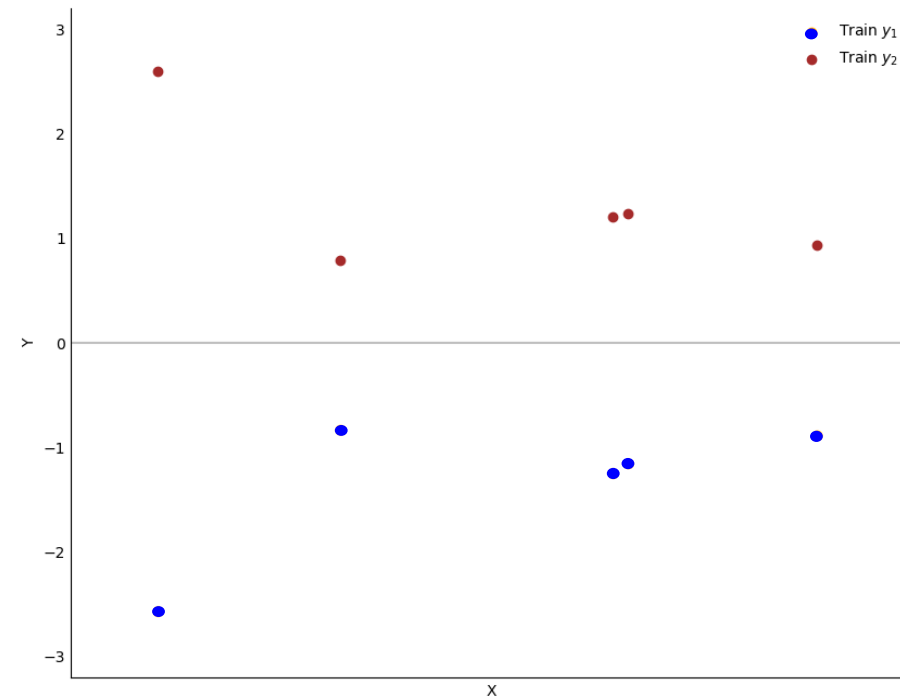
- Gaussian processes are **closed** under linear transformation

$$g \sim \mathcal{GP}(\mu_g(x), k_g(x, x'))$$

$$\mu_f(x) = \mathcal{G}_x[\mu_g], \quad k_f = \mathcal{G}_x k_g \mathcal{G}_x^T$$

Illustration: $\mathcal{L}[f] = f_1 + f_2 = 0$

Constrained GP illustration



[2] Jidling, C., Wahlström, N., Wills, A., & Schön, T. B. (2017)

Constrained Gaussian Process regression

Given a linear constraint on f : $\mathcal{L}[f] = 0$, $\mathcal{L}$ linear operator

Parametrization approach^[2]:

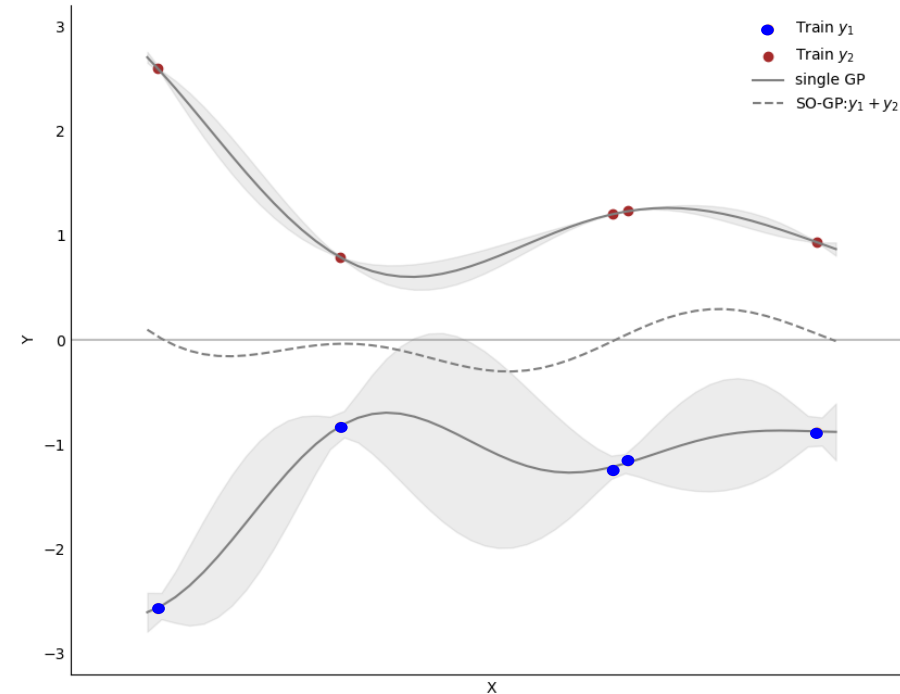
- Suppose that: $f(x) = \mathcal{G}_x[g]$
- Then we impose the constraint on the operator $\mathcal{G}_x$:
$$\mathcal{L}[\mathcal{G}_x] = 0$$
- Gaussian processes are **closed** under linear transformation

$$g \sim \mathcal{GP}(\mu_g(x), k_g(x, x'))$$

$$\mu_f(x) = \mathcal{G}_x[\mu_g], \quad k_f = \mathcal{G}_x k_g \mathcal{G}_x^T$$

Illustration: $\mathcal{L}[f] = f_1 + f_2 = 0$

Constrained GP illustration



[2] Jidling, C., Wahlström, N., Wills, A., & Schön, T. B. (2017)

Constrained Gaussian Process regression

Given a linear constraint on f : $\mathcal{L}[f] = 0$, $\mathcal{L}$ linear operator

Parametrization approach^[2]:

- Suppose that: $f(x) = \mathcal{G}_x[g]$
 - Then we impose the constraint on the operator $\mathcal{G}_x$:
- $$\mathcal{L}[\mathcal{G}_x] = 0$$
- Gaussian processes are **closed** under linear transformation

$$g \sim \mathcal{GP}(\mu_g(x), k_g(x, x'))$$

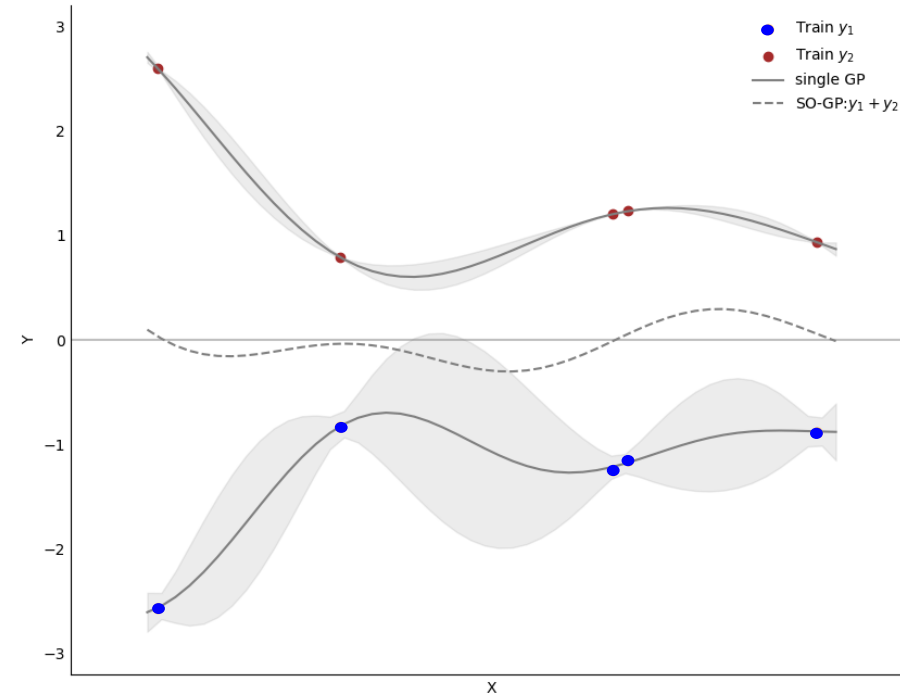
$$\mu_f(x) = \mathcal{G}_x[\mu_g], \quad k_f = \mathcal{G}_x k_g \mathcal{G}_x^T$$

Illustration: $\mathcal{L}[f] = f_1 + f_2 = 0$

- $\mathcal{G}_x = [1, -1]$
- $\mu_f(x) = [\mu_g, -\mu_g]^T$, $k_f = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} k_g$

[2] Jidling, C., Wahlström, N., Wills, A., & Schön, T. B. (2017)

Constrained GP illustration



Constrained Gaussian Process regression

Given a linear constraint on f : $\mathcal{L}[f] = 0$, $\mathcal{L}$ linear operator

Parametrization approach^[2]:

- Suppose that: $f(x) = \mathcal{G}_x[g]$
- Then we impose the constraint on the operator $\mathcal{G}_x$:

$\mathcal{L}[\mathcal{G}_x] = 0$
- Gaussian processes are **closed** under linear transformation

$$g \sim \mathcal{GP}(\mu_g(x), k_g(x, x'))$$

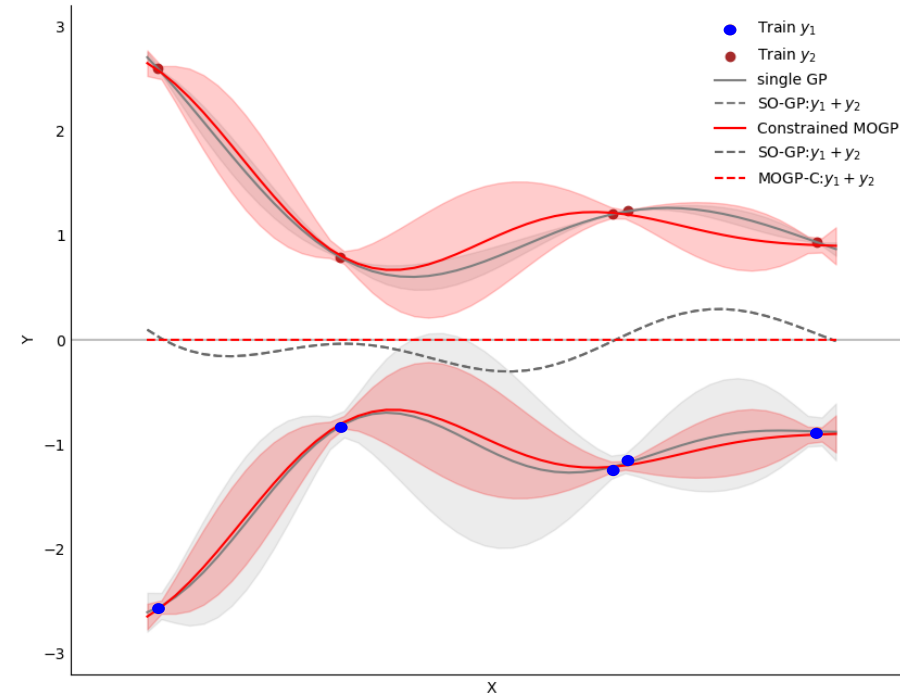
$$\mu_f(x) = \mathcal{G}_x[\mu_g], \quad k_f = \mathcal{G}_x k_g \mathcal{G}_x^T$$

Illustration: $\mathcal{L}[f] = f_1 + f_2 = 0$

- $\mathcal{G}_x = [1, -1]$
- $\mu_f(x) = [\mu_g, -\mu_g]^T$, $k_f = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} k_g$

[2] Jidling, C., Wahlström, N., Wills, A., & Schön, T. B. (2017)

Constrained GP illustration



Comparison of deductive approaches and Constrained MOGP

Deductive approaches: LCM, Independent GP

Joint approach: Constrained MOGP

Analytical benchmark:

$$D = 3, S = 1, Q = 3, P = 3$$

- Normalized inputs $x \in \mathbb{R}^3$
- Normalized outputs

f^1 = Ishigami function

f^2 = Branin function

$$f^3 = -f^1 - f^2$$

Main findings

- ✗ **Heterogenous performances** in deduction-based methods, independent from the size of training set N
- ✓ **Constrained MOGP** achieves **better** or **comparable** predictive accuracy against best scenario of deductive approaches

Comparison of deductive approaches and Constrained MOGP

Deductive approaches: LCM, Independent GP

Joint approach: Constrained MOGP

Analytical benchmark:

$$D = 3, S = 1, Q = 3, P = 3$$

- Normalized inputs $x \in \mathbb{R}^3$
- Normalized outputs

f^1 = Ishigami function

f^2 = Branin function

$$f^3 = -f^1 - f^2$$

Main findings

- ✗ **Heterogenous performances** in deduction-based methods, independent from the size of training set N
- ✓ **Constrained MOGP** achieves **better** or **comparable** predictive accuracy against best scenario of deductive approaches

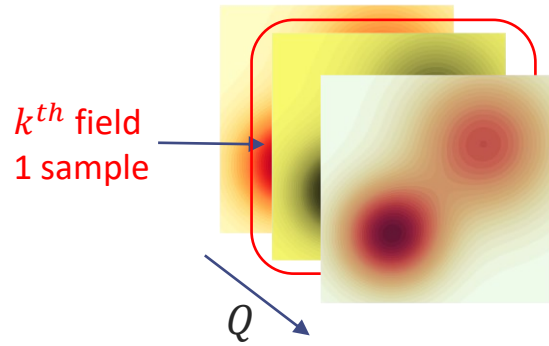
Problem: $\mathcal{O}(N^3 P^3)$ complexity !

Need to be combined with a dimensionality reduction techniques,

but **how**, and **which techniques** ?

PCA for multi-field data^[3]

Computer
code output:

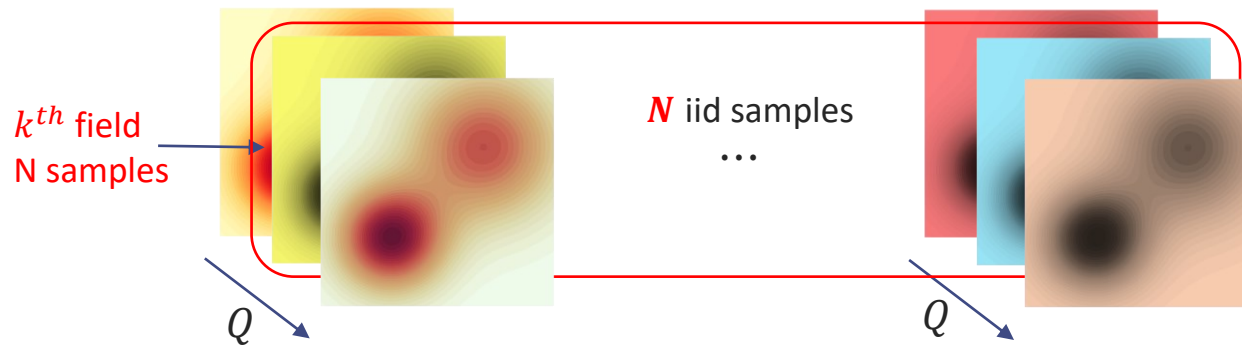


matrix
representation $\rightarrow$

$$Y_k = \begin{bmatrix} y_{1,1}^k & y_{1,s}^k \\ \vdots & \vdots \end{bmatrix}$$

PCA for multi-field data^[3]

Computer
code output:

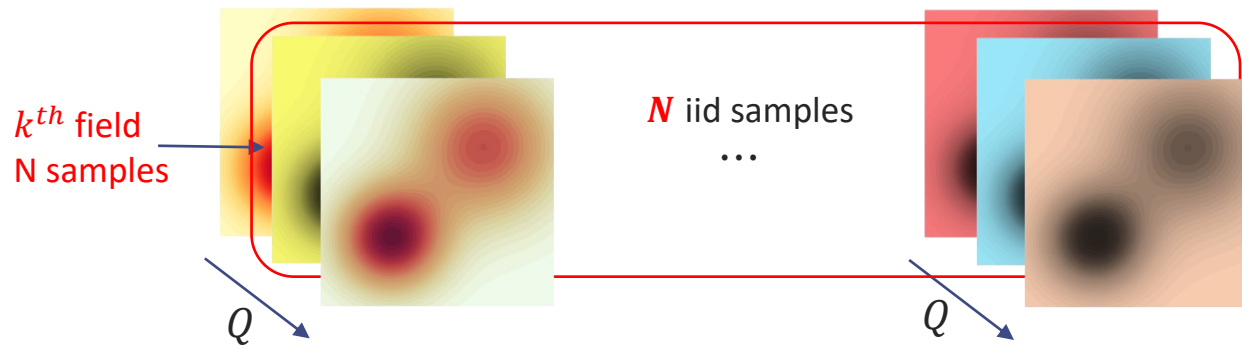


matrix
representation
→

$$\mathbf{Y}_k = \begin{bmatrix} y_{1,1}^k & y_{1,S}^k \\ \vdots & \dots & \vdots \\ y_{N,1}^k & y_{N,S}^k \end{bmatrix}$$

PCA for multi-field data^[3]

Computer
code output:



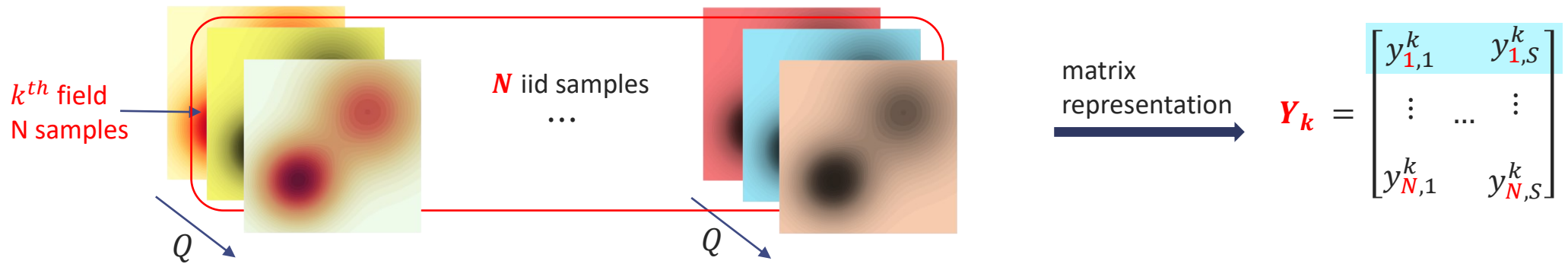
matrix
representation
→

$$Y_k = \begin{bmatrix} y_{1,1}^k & y_{1,S}^k \\ \vdots & \dots & \vdots \\ y_{N,1}^k & y_{N,S}^k \end{bmatrix}$$

- There is Q data matrices for Q physical fields: $Y_1, \dots, Y_Q \Rightarrow$ How can we reduce their dimension S ?

PCA for multi-field data^[3]

Computer
code output:

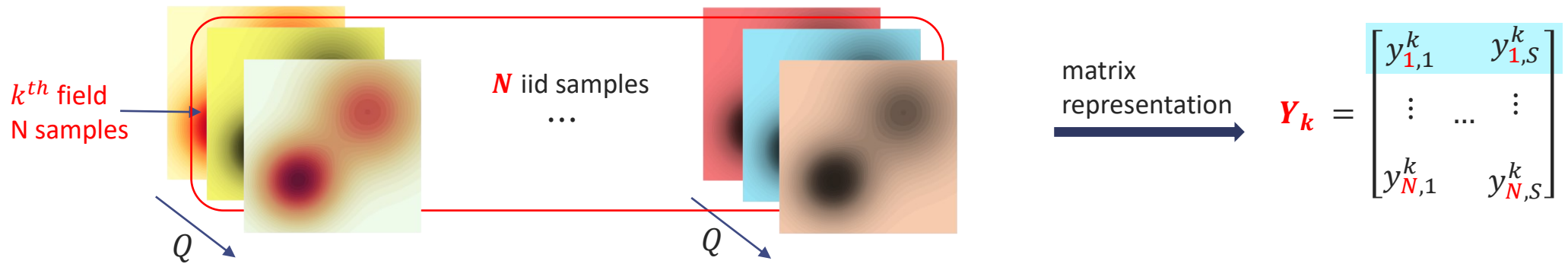


- There is Q data matrices for Q physical fields: $Y_1, \dots, Y_Q \Rightarrow$ How can we **reduce** their **dimension S** ?

	Data matrix	SVD	Cov matrix	Rank m Projector	Reconstruction error
Field-wise PCA	$Y_k \in \mathbb{R}^{N \times S}$	$Y_k = U_k D_k V_k^T$	$\Sigma_k = \frac{1}{N} Y_k^T Y_k$ $K_k = \frac{1}{S} Y_k Y_k^T$	$P_{k,m} = V_{k,m} V_{k,m}^T$	$\mathcal{E}_{k,m}^{field} = \ Y_k - Y_k P_{k,m}\ _F$
Col-wise PCA	$Y_{col} = [Y_1, \dots, Y_Q] \in \mathbb{R}^{N \times QS}$	$Y_{col} = U_{col} D_{col} V_{col}^T$	$K_{col} = \frac{1}{N} Y_{col} Y_{col}^T$	$\Pi_{col,m} = U_{col} U_{col}^T$	$\mathcal{E}_{m,k}^{col} = \ Y_k - \Pi_{col,m} Y_k\ _F$
Row-wise PCA	$Y_{row} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_k \\ \vdots \\ Y_Q \end{bmatrix} \in \mathbb{R}^{QN \times S}$	$Y_{row} = U_{row} D_{row} V_{row}^T$	$\Sigma_{row} = \frac{1}{QN} Y_{row}^T Y_{row}$	$P_{row,m} = V_{row,m} V_{row,m}^T$	$\mathcal{E}_{m,k}^{row} = \ Y_k - Y_k P_{row,m}\ _F$

PCA for multi-field data^[3]

Computer code output:



- There is Q data matrices for Q physical fields: $Y_1, \dots, Y_Q \Rightarrow$ How can we reduce their dimension S ?

	Data matrix	SVD	Cov matrix	Rank m Projector	Reconstruction error
Field-wise PCA	$Y_k \in \mathbb{R}^{N \times S}$	$Y_k = U_k D_k V_k^T$	$\Sigma_k = \frac{1}{N} Y_k^T Y_k$ $K_k = \frac{1}{S} Y_k Y_k^T$	$P_{k,m} = V_{k,m} V_{k,m}^T$	$\mathcal{E}_{k,m}^{field} = \ Y_k - Y_k P_{k,m}\ _F$
Col-wise PCA	$Y_{col} = [Y_1, \dots, Y_Q] \in \mathbb{R}^{N \times QS}$	$Y_{col} = U_{col} D_{col} V_{col}^T$	$K_{col} = \frac{1}{N} Y_{col} Y_{col}^T$	$\Pi_{col,m} = U_{col} U_{col}^T$	$\mathcal{E}_{m,k}^{col} = \ Y_k - \Pi_{col,m} Y_k\ _F$
Row-wise PCA	$Y_{row} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_k \\ \vdots \\ Y_Q \end{bmatrix} \in \mathbb{R}^{QN \times S}$	$Y_{row} = U_{row} D_{row} V_{row}^T$	$\Sigma_{row} = \frac{1}{QN} Y_{row}^T Y_{row}$	$P_{row,m} = V_{row,m} V_{row,m}^T$	$\mathcal{E}_{m,k}^{row} = \ Y_k - Y_k P_{row,m}\ _F$

☞ Which approach is optimal and well suited for a combination with Constrained MOGP?

Theoretical analysis and comparison

- Row-wise PCA vs field-wise PCA

Theorem 1: (Row-wise excess error bound)

Let $\delta_k = \lambda_m^k - \lambda_{m+1}^k$, where λ_m^k denotes the m -th eigenvalue of Σ_k in decreasing order. If $\delta_k > 0$, then the row-wise excess error $\Delta \mathcal{E}_{k,m}^{row} = \mathcal{E}_{k,m}^{field} - \mathcal{E}_{k,m}^{row}$ of the k -th field is directly bounded by

$$\Delta \mathcal{E}_{k,m}^{row} \leq \frac{2n\sqrt{2} \|\Sigma_k\|_F}{\delta_k} \min \left(\frac{\sqrt{m}}{Q} \sum_{j=1}^Q \|\Sigma_j - \Sigma_k\|_{op}, \frac{1}{Q} \sum_{j=1}^Q \|\Sigma_j - \Sigma_k\|_F \right)$$

- Col-wise PCA vs field-wise PCA

Theorem 2: (Col-wise excess error bound)

Let $\delta_k = \lambda_m^k - \lambda_{m+1}^k$. If $\delta_k > 0$, then the col-wise excess error $\Delta \mathcal{E}_{k,m}^{col} = \mathcal{E}_{k,m}^{field} - \mathcal{E}_{k,m}^{col}$ of the k -th field is directly bounded by the distance between Gram matrices:

$$\Delta \mathcal{E}_{k,m}^{col} \leq \frac{2S\sqrt{2} \|K_k\|_F}{\delta_k} \min \left(\frac{\sqrt{m}}{Q} \sum_{j=1}^Q \|K_j - K_k\|_{op}, \frac{1}{Q} \sum_{j=1}^Q \|K_j - K_k\|_F \right)$$



Theorem 1 & 2 bounds guarantee bounded excess error whenever fields exhibit **structural similarities**, providing a theoretical justification of using **multi-field PCA** over **field-wise** when these similarities hold

Theoretical analysis and comparison

- Row-wise PCA vs col-wise PCA

Theorem 3: (Col-wise vs row-wise performance)

Let $\Lambda_{row,m} = nQ \sum_{j=1}^m \lambda_m^{\Sigma_{row}}$, $\Lambda_{col,m} = QS \sum_{j=1}^m \lambda_m^{K_{col}}$. The difference between the reconstruction error of the column-wise and row-wise approaches at rank m is exactly :

$$\mathcal{E}_{k,m}^{col} - \mathcal{E}_{k,m}^{row} = \text{Tr}(D_{row,m}^2) - \text{Tr}(D_{col,m}^2) = \Lambda_{row,m} - \Lambda_{col,m}$$



Neither strategy is intrinsically superior between row-wise and col-wise and the theorem gives an exact dominance criterion

□ Constrained fields case:

- ✗ Field wise and column wise approaches fail to guarantee constraint satisfaction in reconstruction
- ✗ To satisfy linear constraints, they must be used with deductive approaches principle
- ✓ **Row-wise approach preserves linear constraints both in latent space, and in the reconstructed fields**
 - ✎ We propose to combine **row-wise PCA + Constrained MOGP** to solve our initial problem
- ✓ We have **numerically validated** our theoretical results on an analytical benchmark based on **Lotka-Volterra** model

Metamodel for the turbulent Reynolds stress tensor

- **Context:** Uncertainty propagation study in RANS modeling
- Turbulence closure model: standard $k - \varepsilon$
- **Inputs:** closure parameters of the standard $k - \varepsilon$ turbulence model

$$\mathbf{x} = (C_\mu, \sigma_k, \sigma_\varepsilon, C_{\varepsilon_1}, C_{\varepsilon_2})$$

- **Output:** Reynolds stress tensor $\tau_{ij} = \overline{u'_i u'_j}$ and the turbulent kinetic energy k

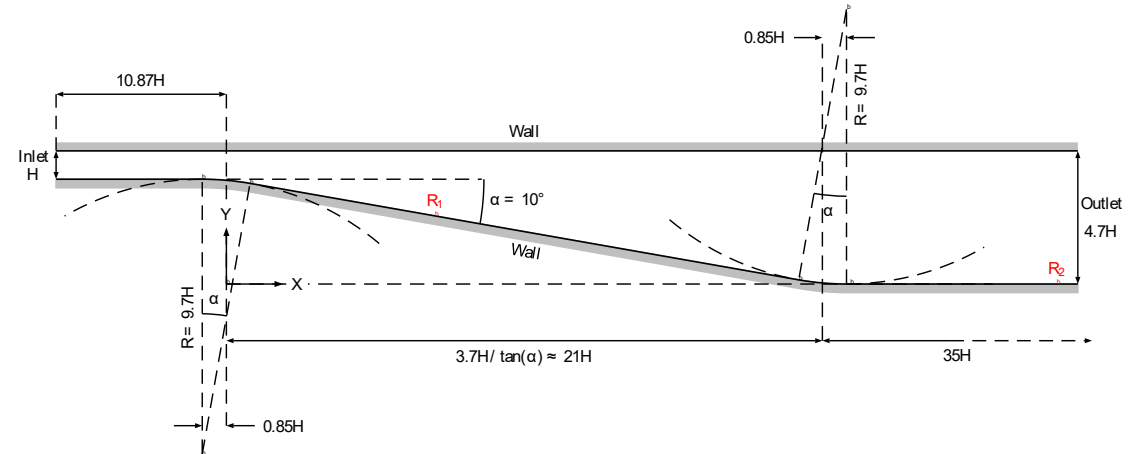
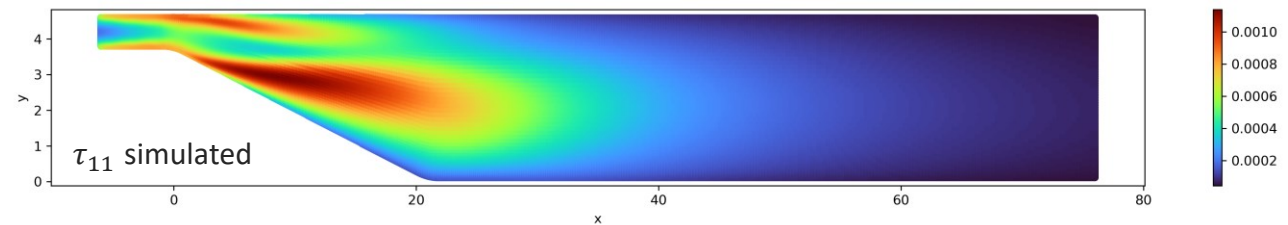


Figure: Scheme of the experiment carried out in Buice and Eaton, 2000. The wall of the channel are defined by straight lines and circular arcs of centers O_1 and O_2



Metamodel for the turbulent Reynolds stress tensor

- **Context:** Uncertainty propagation study in RANS modeling
- Turbulence closure model: standard $k - \varepsilon$
- **Inputs:** closure parameters of the standard $k - \varepsilon$ turbulence model

$$\mathbf{x} = (C_\mu, \sigma_k, \sigma_\varepsilon, C_{\varepsilon_1}, C_{\varepsilon_2})$$

- **Output:** Reynolds stress tensor $\tau_{ij} = \overline{u'_i u'_j}$ and the turbulent kinetic energy k
- The incompressibility condition is translated by:

$$\tau_{11} + \tau_{22} - \frac{4}{3}k = 0$$

Datasets generation details:

- Mesh composed of 200 000 triangular elements
- CEA/DES CFD solver TrioCFD
- Reynolds number = 18000
- $N_{train} = 50$ and $N_{test} = 100$

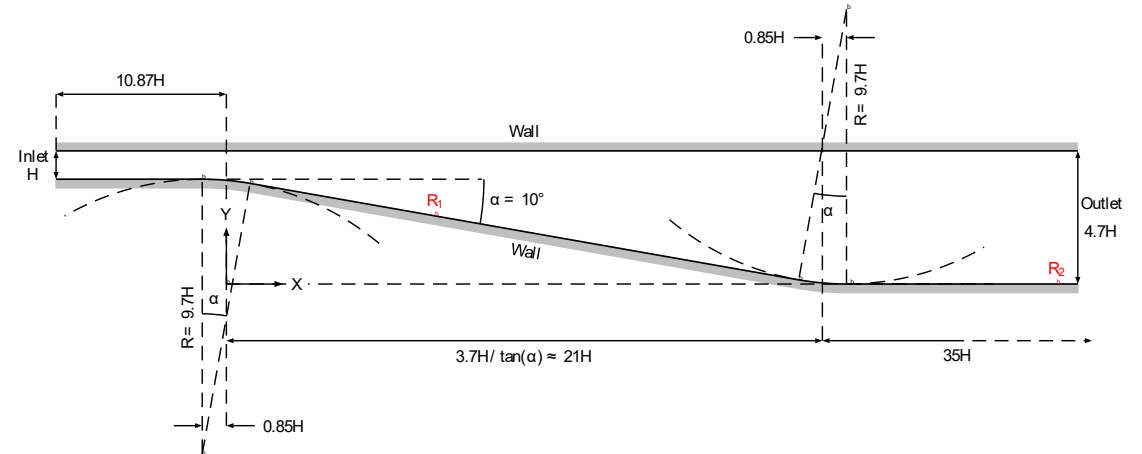
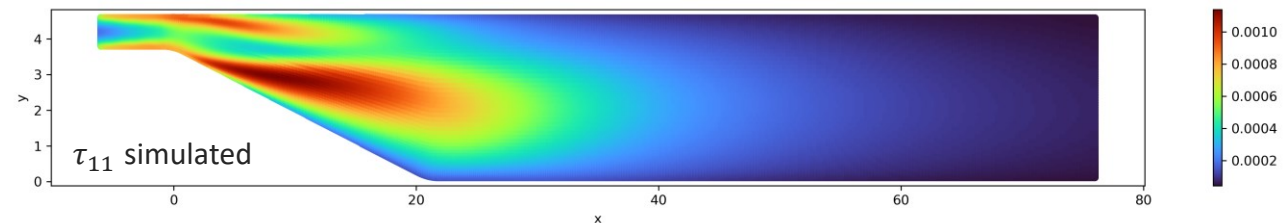


Figure: Scheme of the experiment carried out in Buice and Eaton, 2000. The wall of the channel are defined by straight lines and circular arcs of centers O_1 and O_2



RRMSE result on the LEVM Reynolds stress tensor

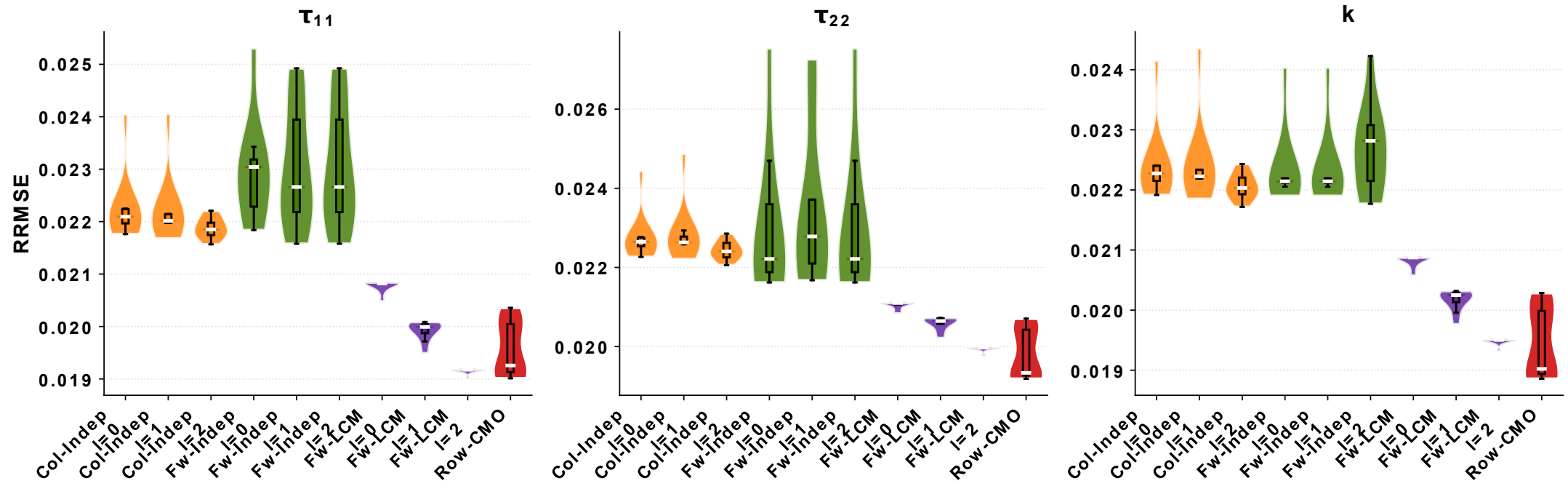


Fig: Final RRMSE (median $\pm$ std) with $m=8$ latent dimension for each method. Our method **Row-CMO** against 3 deductive alternatives

RRMSE result on the LEVM Reynolds stress tensor

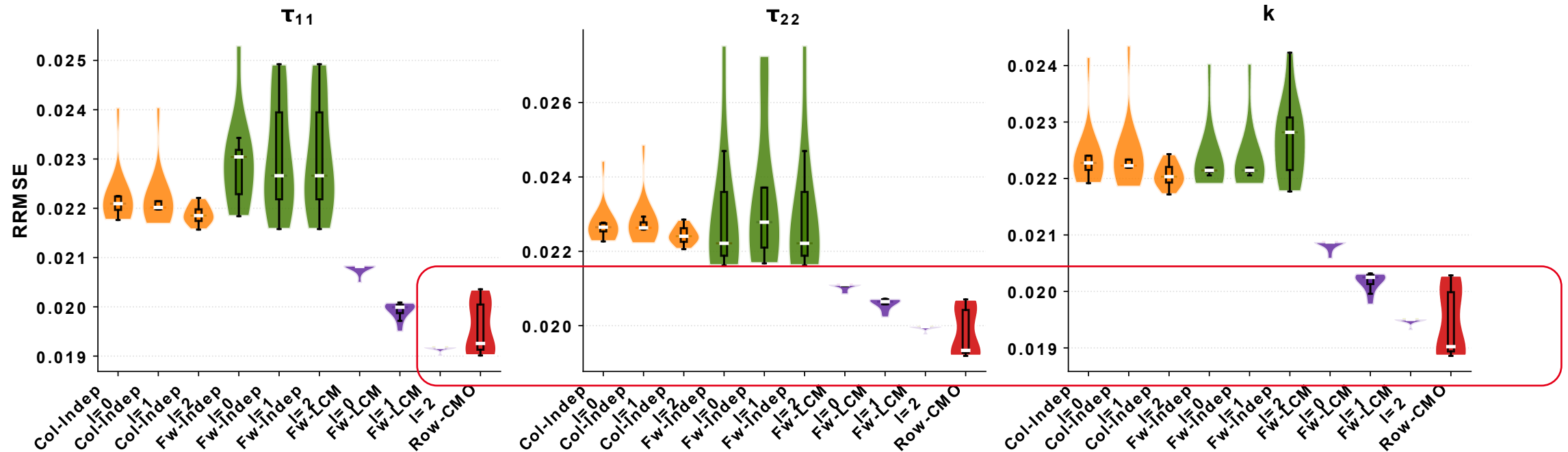


Fig: Final RRMSE (median $\pm$ std) with $m=8$ latent dimension for each method. Our method **Row-CMO** against 3 deductive alternatives

- Final RRMSE reveals that our **method** ends up outperforming alternatives and achieves the lowest RRMSE on all three outputs.
- Deductive approaches yield varying results across variants: this confirms their inherent instability

Prediction and uncertainty illustration

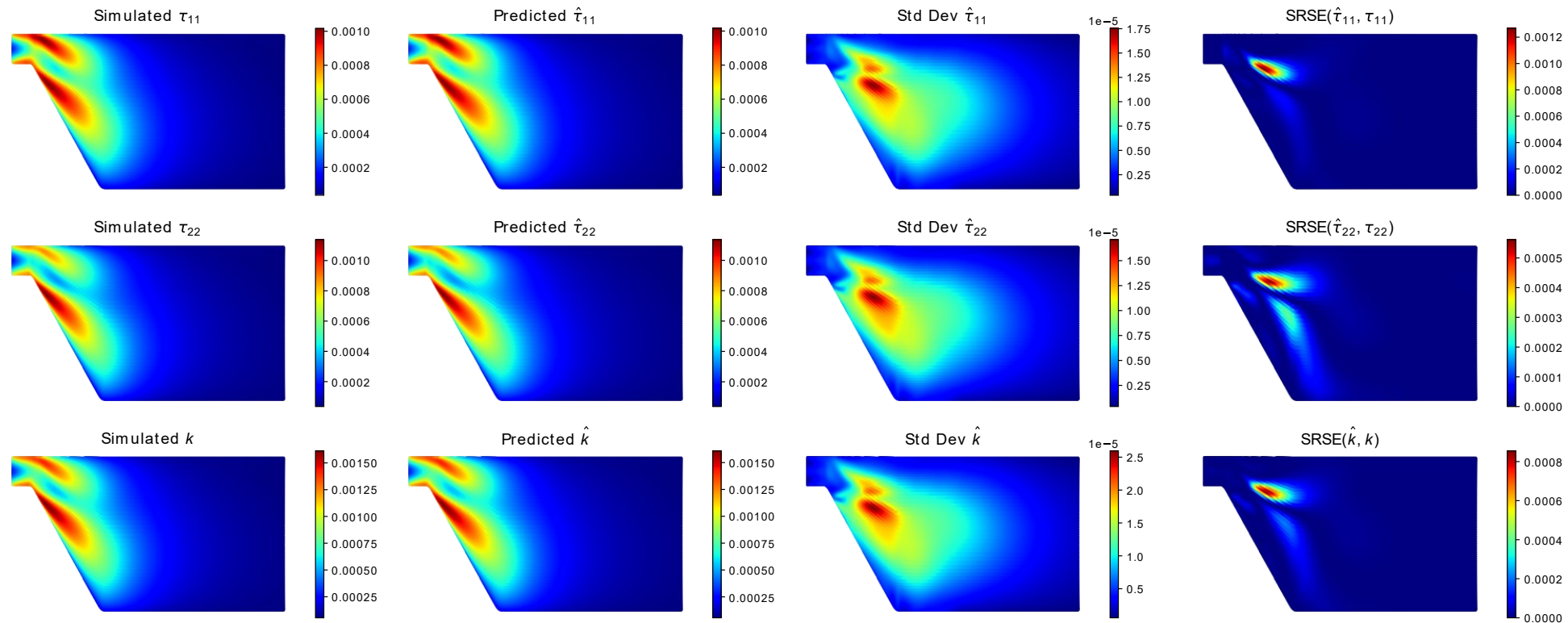


Fig: Row-wise PCA-Constrained MOGP predictions of the CFD test case for a representative test input (with $m=8$)

The model accurately captures the spatial structure. Error and uncertainty are concentrated in the recirculation zone of the diffuser.

Conclusion

- ✓ We introduce a constrained GP framework for for high dimensional fields that combines row-wise PCA with joint modeling to efficiently handle linearly constrained physical fields

Conclusion

- ✓ We introduce a constrained GP framework for high dimensional fields that combines row-wise PCA with joint modeling to efficiently handle linearly constrained physical fields
- ✓ We provide a theoretical analysis that quantifies when this PCA strategy remains competitive with standard approaches, beyond its key advantage of full compatibility with constrained MOGP

Conclusion

- ✓ We introduce a constrained GP framework for high dimensional fields that combines row-wise PCA with joint modeling to efficiently handle linearly constrained physical fields
- ✓ We provide a theoretical analysis that quantifies when this PCA strategy remains competitive with standard approaches, beyond its key advantage of full compatibility with constrained MOGP
- ✓ Through benchmark, we demonstrate the robustness of joint modeling compared to deductive approaches which can exhibit heterogeneous performances

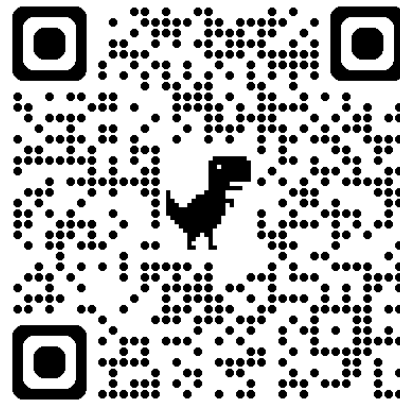
Conclusion

- ✓ We introduce a constrained GP framework for for high dimensional fields that combines row-wise PCA with joint modeling to efficiently handle linearly constrained physical fields
- ✓ We provide a theoretical analysis that quantifies when this PCA strategy remains competitive with standard approaches, beyond its key advantage of full compatibility with constrained MOGP
- ✓ Through benchmark, we demonstrate the robustness of joint modeling compared to deductive approaches which can exhibit heterogenous performances
- ✓ Finally, we validate the practical relevance of our approach on an industrial CFD case, highlighting clear benefits in data-scarce settings

Conclusion

- ✓ We introduce a constrained GP framework for high dimensional fields that combines row-wise PCA with joint modeling to efficiently handle linearly constrained physical fields
- ✓ We provide a theoretical analysis that quantifies when this PCA strategy remains competitive with standard approaches, beyond its key advantage of full compatibility with constrained MOGP
- ✓ Through benchmark, we demonstrate the robustness of joint modeling compared to deductive approaches which can exhibit heterogeneous performances
- ✓ Finally, we validate the practical relevance of our approach on an industrial CFD case, highlighting clear benefits in data-scarce settings

Preprint available here =>



References

- [1] Alvarez, M. A., Rosasco, L., & Lawrence, N. D. (2012). Kernels for vector-valued functions: A review. Foundations and Trends® in Machine Learning.
- [2] Jidling, C., Wahlström, N., Wills, A., & Schön, T. B. (2017). Linearly constrained Gaussian processes. Advances in neural information processing systems.
- [3] R. W. Preisendorfer and C. D. Mobley. (1988). Principal component analysis in meteorology and oceanography. Elsevier.
- [4] Carl U. Buice and John K. Eaton. Experimental investigation of flow through an asymmetric plane diffuser: (data bank contribution)1. Journal of fluids Engineering, 122(2):433-435, 01 200. doi: 10.1115/1.483278.

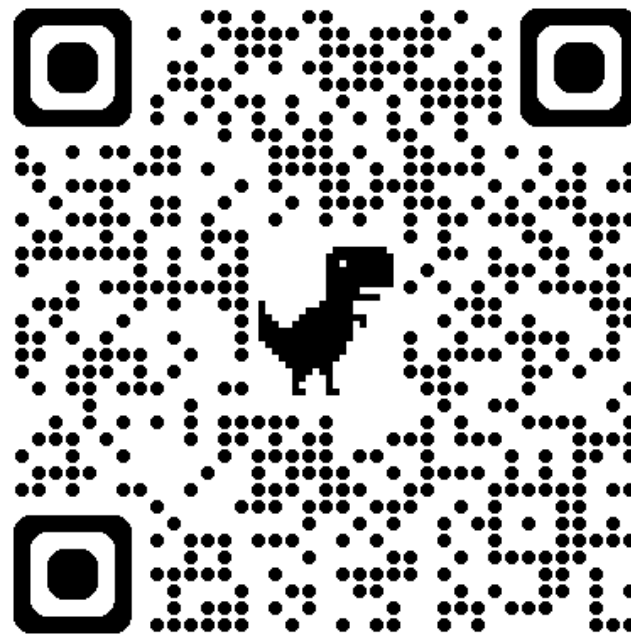
Acknowledgments:

This work has been funded by the French *Agence Nationale de la Recherche* ExaMA project, under the France 2030 initiative, with reference ANR-22-EXNU-0002.



Thanks !

Any questions ?



Preprint link !



PROGRAMME
DE RECHERCHE
NUMÉRIQUE
POUR L'EXASCALE

Comparison of deductive approaches and Constrained MOGP

Analytical benchmark:

$$D = 3, S = 1, Q = 3, P = 3$$

- Normalized inputs $x \in \mathbb{R}^3$
- Normalized outputs

$$f^1 = \text{Ishigami function}$$

$$f^2 = \text{Branin function}$$

$$f^3 = -f^1 - f^2$$

Deductive approaches: LCM, Independent GP

Joint approach: Constrained MOGP

Comparison of deductive approaches and Constrained MOGP

RMSE of the predictive distribution of f^2 : Violin plots across 200 train/test splits.

Analytical benchmark:

$$D = 3, S = 1, Q = 3, P = 3$$

- Normalized inputs $x \in \mathbb{R}^3$
- Normalized outputs

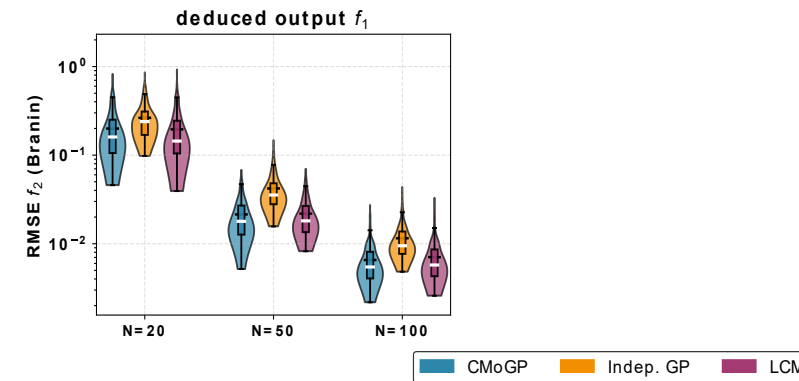
f^1 = Ishigami function

f^2 = Branin function

$f^3 = -f^1 - f^2$

Deductive approaches: LCM, Independent GP

Joint approach: Constrained MOGP



Comparison of deductive approaches and Constrained MOGP

RMSE of the predictive distribution of f^2 : Violin plots across 200 train/test splits.

Analytical benchmark:

$$D = 3, S = 1, Q = 3, P = 3$$

- Normalized inputs $x \in \mathbb{R}^3$
- Normalized outputs

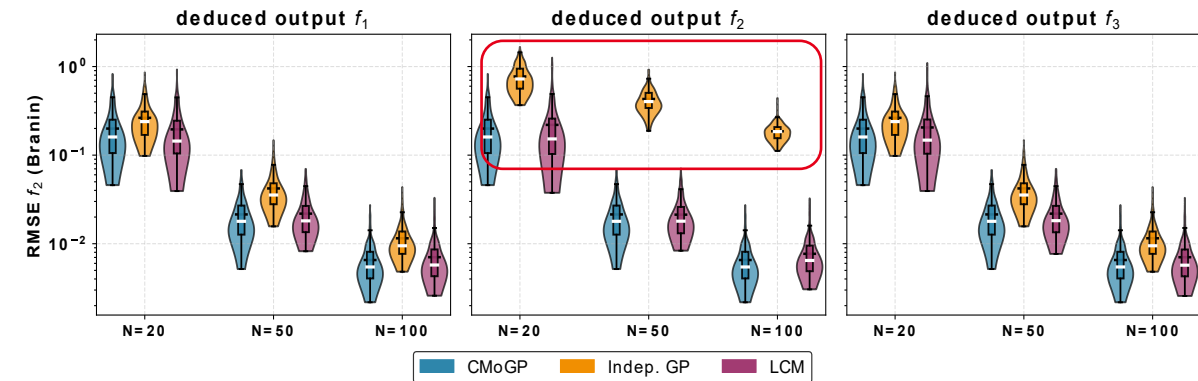
f^1 = Ishigami function

f^2 = Branin function

$f^3 = -f^1 - f^2$

Deductive approaches: LCM, Independent GP

Joint approach: Constrained MOGP



Comparison of deductive approaches and Constrained MOGP

Analytical benchmark:

$$D = 3, S = 1, Q = 3, P = 3$$

- Normalized inputs $x \in \mathbb{R}^3$
- Normalized outputs

f^1 = Ishigami function

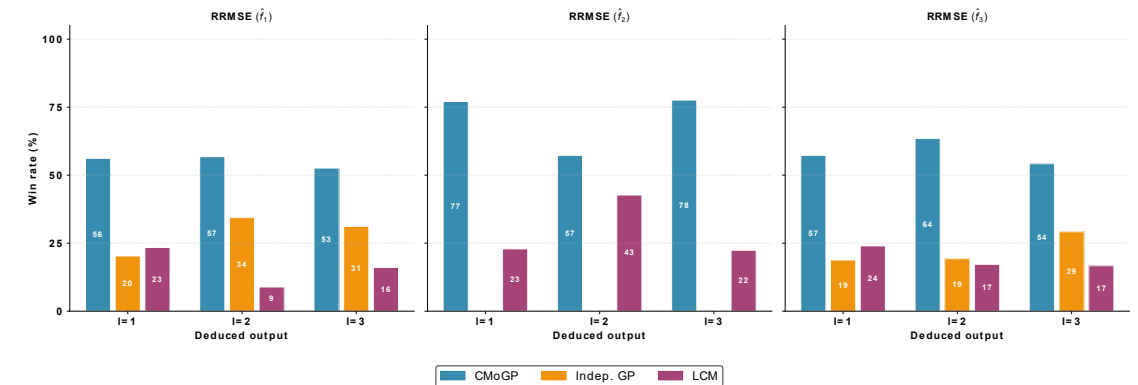
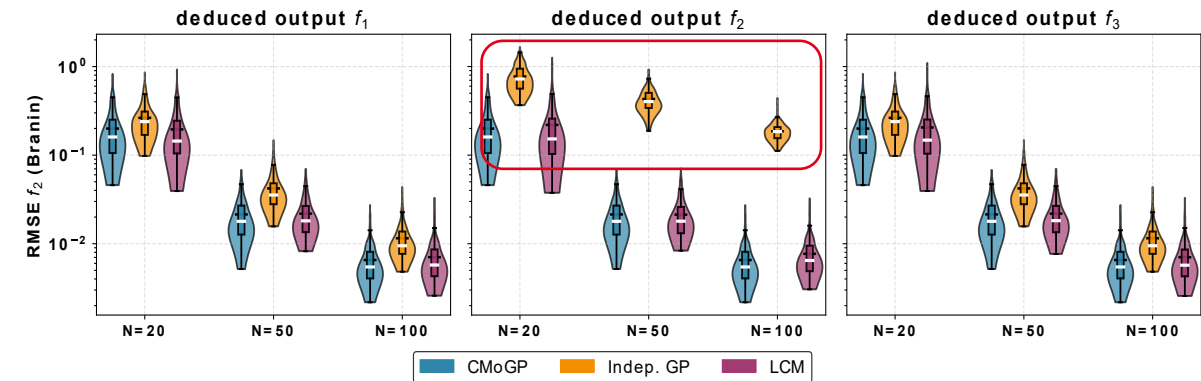
f^2 = Branin function

$f^3 = -f^1 - f^2$

Deductive approaches: LCM, Independent GP

Joint approach: Constrained MOGP

RMSE of the predictive distribution of f^2 : Violin plots across 200 train/test splits.



Win rate (%) of each method on the RMSE of each output (f^1 , f^2 , f^3) across all deduction scenarios with $N=50$.

Comparison of deductive approaches and Constrained MOGP

RMSE of the predictive distribution of f^2 : Violin plots across 200 train/test splits.

Analytical benchmark:

$$D = 3, S = 1, Q = 3, P = 3$$

- Normalized inputs $x \in \mathbb{R}^3$
- Normalized outputs

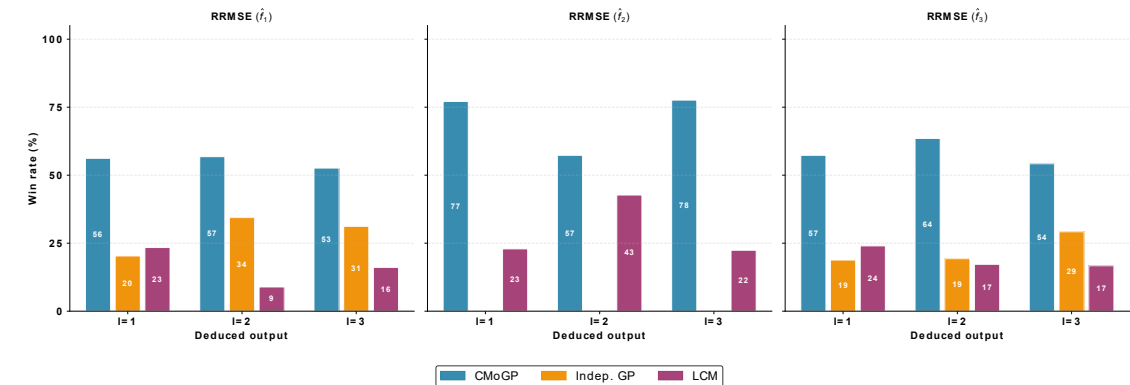
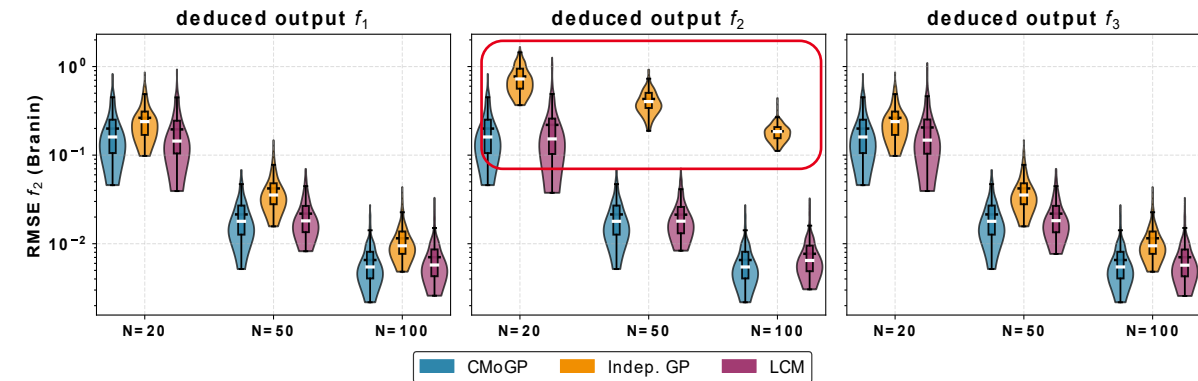
f^1 = Ishigami function

f^2 = Branin function

$f^3 = -f^1 - f^2$

Deductive approaches: LCM, Independent GP

Joint approach: Constrained MOGP



Win rate (%) of each method on the RMSE of each output (f^1 , f^2 , f^3) across all deduction scenarios with $N=50$.

- ✗ Heterogenous performances in deduction-based methods, independent from the size of training set N
- ✓ Constrained MOGP achieves **better** or **comparable** predictive accuracy against best scenario of deductive approaches

Comparison of deductive approaches and Constrained MOGP

Analytical benchmark:

$$D = 3, S = 1, Q = 3, P = 3$$

- Normalized inputs $x \in \mathbb{R}^3$
- Normalized outputs

f^1 = Ishigami function

f^2 = Branin function

$f^3 = -f^1 - f^2$

Deductive approaches: LCM, Independent GP

Joint approach: Constrained MOGP

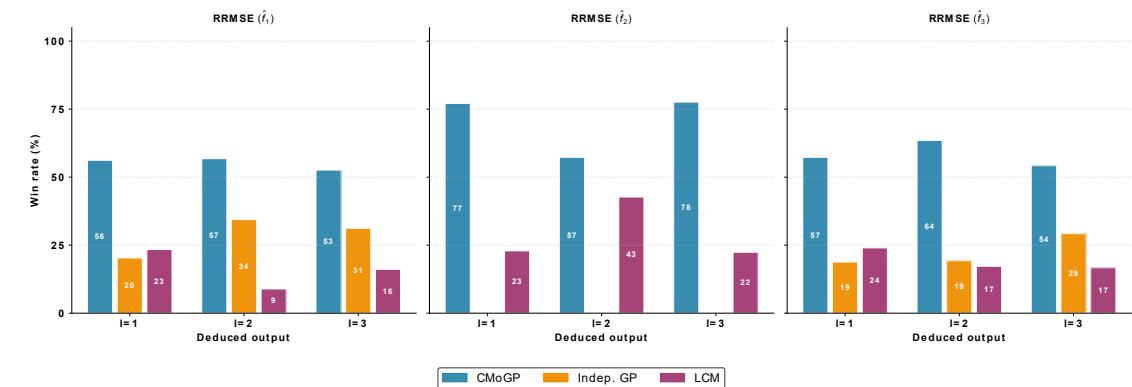
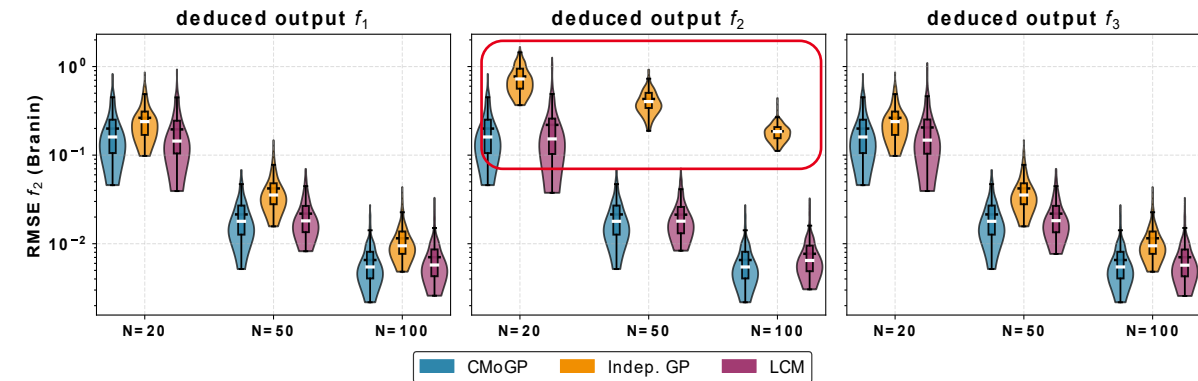
- ✗ **Heterogenous performances** in deduction-based methods, independent from the size of training set N
- ✓ **Constrained MOGP** achieves **better or comparable** predictive accuracy against best scenario of deductive approaches

Problem: $\mathcal{O}(N^3 P^3)$ complexity !

Need to be combined with a dimensionality reduction techniques,

but **how**, and **which techniques** ?

RMSE of the predictive distribution of f^2 : Violin plots across 200 train/test splits.



Win rate (%) of each method on the RMSE of each output (f^1 , f^2 , f^3) across all deduction scenarios with N=50.

Metamodel for population dynamics

- **Lotka-Volterra equations:**

$$\begin{aligned} p: & \text{prey density population} \\ q: & \text{predator density population} \end{aligned} \quad \begin{cases} p'(t) = ap(t) - bp(t)q(t) \\ q'(t) = -cq(t) + dp(t)q(t) \end{cases} \quad (1)$$

$r(t) = \log(q(t)), s(t) = \log(p(t))$

- **Analytical benchmark configuration:**

$D = 2, S = 400, Q = 4, N_{test} = 100, b \in [0.37, 0.40], d \in [0.0, 0.06]$

- **Metamodeling problem:**

Inputs: $x = (b, d)$,

Outputs: $p(t), q(t), r(t), s(t)$

Constraint equation:

$$dp(t) - cs(t) + bq(t) - ar(t) = C_0$$

Metamodel for population dynamics

▪ Lotka-Volterra equations:

$$\begin{aligned} p: & \text{prey density population} \\ q: & \text{predator density population} \\ r(t) &= \log(q(t)), \quad s(t) = \log(p(t)) \end{aligned} \quad \begin{cases} p'(t) = ap(t) - bp(t)q(t) \\ q'(t) = -cq(t) + dp(t)q(t) \end{cases} \quad (1)$$

▪ Analytical benchmark configuration:

$D = 2, S = 400, Q = 4, N_{test} = 100, b \in [0.37, 0.40], d \in [0.0, 0.06]$

▪ Metamodeling problem:

Inputs: $x = (b, d)$,

Outputs: $p(t), q(t), r(t), s(t)$

Constraint equation:

$$dp(t) - cs(t) + bq(t) - ar(t) = C_0$$

Our model:

- Row-CMO = Row-wise PCA
+ CMOGP

Deductive approaches:

- Col-Indep = Column-wise PCA
+ Independent GP
- Fw-Indep = Field-wise PCA
+ Independent GP
- Fw-LCM: Field-wise PCA
+ LCM

l : Deduced output indices

Metamodel for population dynamics

Lotka-Volterra equations:

$$\begin{aligned} p: & \text{prey density population} \\ q: & \text{predator density population} \\ r(t) &= \log(q(t)), \quad s(t) = \log(p(t)) \end{aligned} \quad \begin{cases} p'(t) = ap(t) - bp(t)q(t) \\ q'(t) = -cq(t) + dp(t)q(t) \end{cases} \quad (1)$$

Analytical benchmark configuration:

$$D = 2, S = 400, Q = 4, N_{test} = 100, b \in [0.37, 0.40], d \in [0.0, 0.06]$$

Metamodeling problem:

Inputs: $x = (b, d)$,

Outputs: $p(t), q(t), r(t), s(t)$

Constraint equation:

$$dp(t) - cs(t) + bq(t) - ar(t) = C_0$$

Our model:

- Row-CMO = Row-wise PCA + CMOGP

Deductive approaches:

- Col-Indep = Column-wise PCA + Independent GP
- Fw-Indep = Field-wise PCA + Independent GP
- Fw-LCM: Field-wise PCA + LCM

l : Deduced output indices

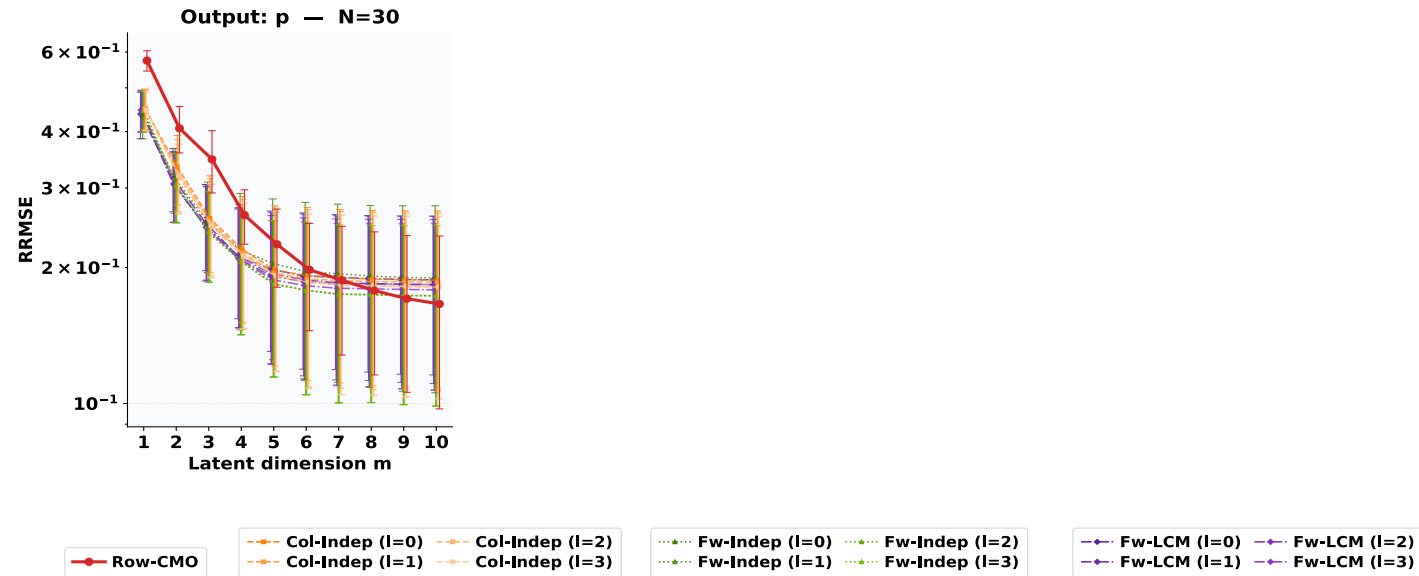


Fig: RRMSE (mean $\pm$ std) on output p for all methods. **Beyond $m=4$, Row-CMO surpasses all alternative at higher latent dimensions**

Metamodel for population dynamics

Lotka-Volterra equations:

$$\begin{aligned} p: & \text{prey density population} \\ q: & \text{predator density population} \\ r(t) &= \log(q(t)), \quad s(t) = \log(p(t)) \end{aligned} \quad \begin{cases} p'(t) = ap(t) - bp(t)q(t) \\ q'(t) = -cq(t) + dp(t)q(t) \end{cases} \quad (1)$$

Analytical benchmark configuration:

$D = 2, S = 400, Q = 4, N_{test} = 100, b \in [0.37, 0.40], d \in [0.0, 0.06]$

Metamodeling problem:

Inputs: $x = (b, d)$,

Outputs: $p(t), q(t), r(t), s(t)$

Constraint equation:

$$dp(t) - cs(t) + bq(t) - ar(t) = C_0$$

Our model:

- Row-CMO = Row-wise PCA + CMOGP

Deductive approaches:

- Col-Indep = Column-wise PCA + Independent GP
- Fw-Indep = Field-wise PCA + Independent GP
- Fw-LCM: Field-wise PCA + LCM

l : Deduced output indices

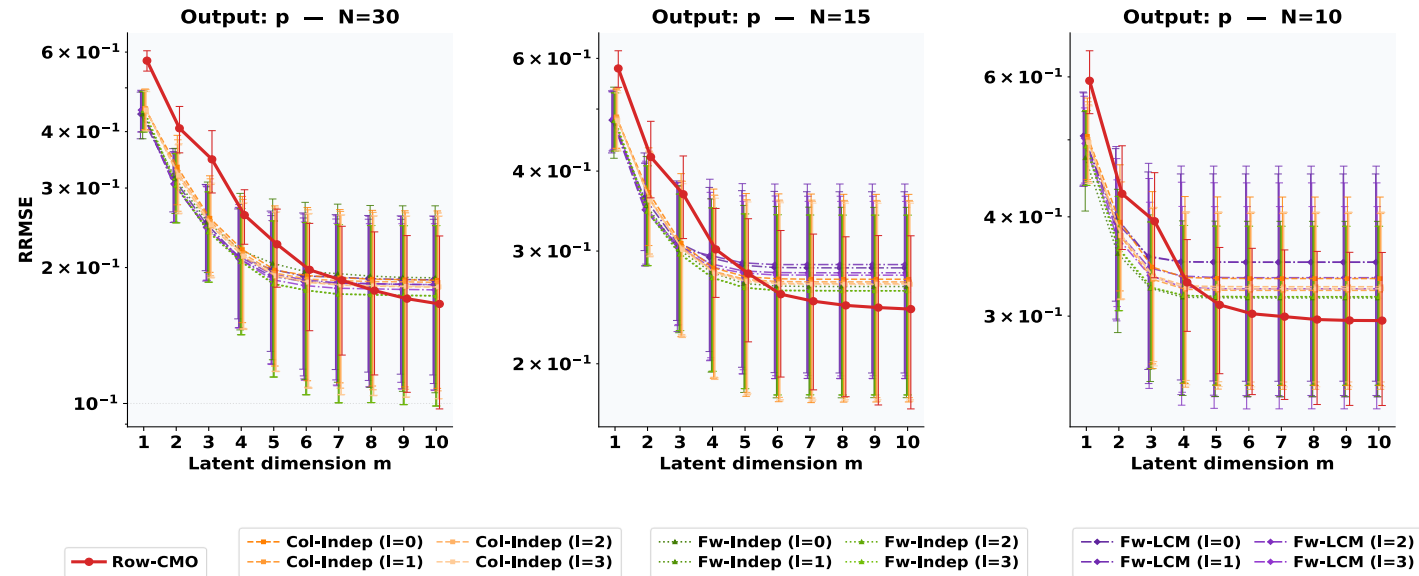


Fig: RRMSE (mean $\pm$ std) on output p for all methods. Beyond $m=4$, Row-CMO surpasses all alternative at higher latent dimensions

Metamodel for the turbulent Reynolds stress tensor

- **Context:** Uncertainty propagation study in RANS modeling
- Turbulence closure model: standard $k - \varepsilon$
- **Inputs:** closure parameters of the standard $k - \varepsilon$ turbulence model

$$\mathbf{x} = (C_\mu, \sigma_k, \sigma_\varepsilon, C_{\varepsilon_1}, C_{\varepsilon_2})$$

- **Output:** Reynolds stress tensor $\tau_{ij} = \overline{u'_i u'_j}$ and the turbulent kinetic energy k
- The incompressibility condition is translated by:

$$\tau_{11} + \tau_{22} - \frac{4}{3}k = 0$$

Datasets generation details:

- Mesh composed of 200 000 triangular elements
- CEA/DES CFD solver TrioCFD
- Reynolds number = 18000
- $N_{train} = 50$ and $N_{test} = 100$

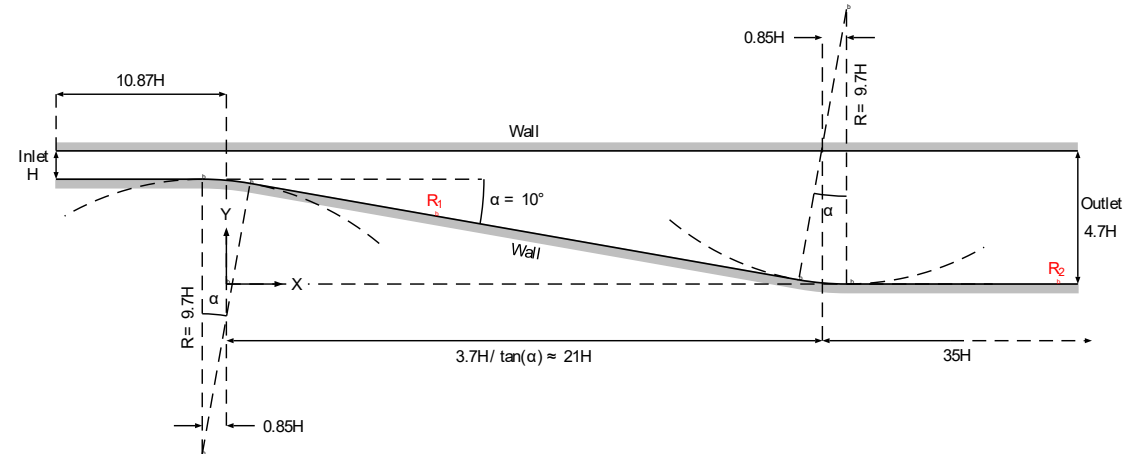
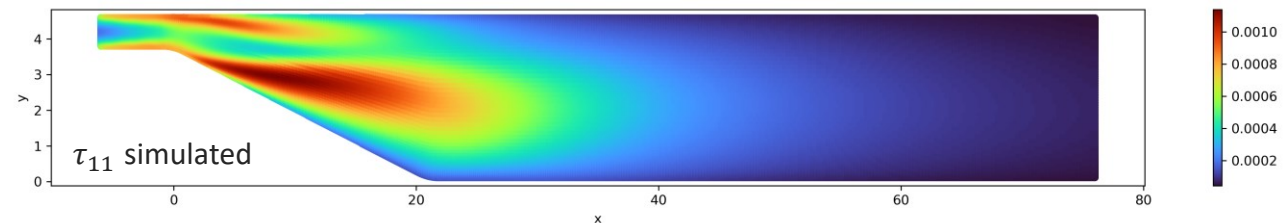


Figure: Scheme of the experiment carried out in Buice and Eaton, 2000. The wall of the channel are defined by straight lines and circular arcs of centers O_1 and O_2



Results on the LEVM Reynolds stress tensor

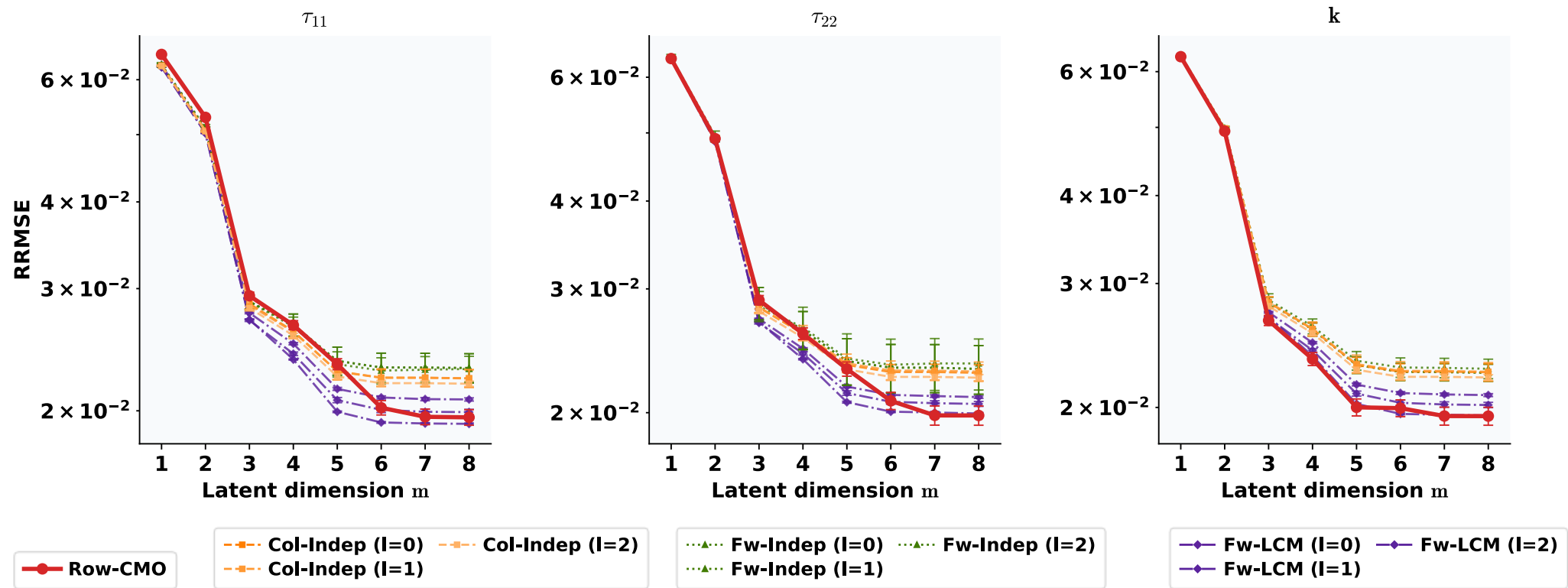


Figure: RRMSE (mean $\pm$ std) on all outputs for all methods.

Appendix : Gaussian process regression

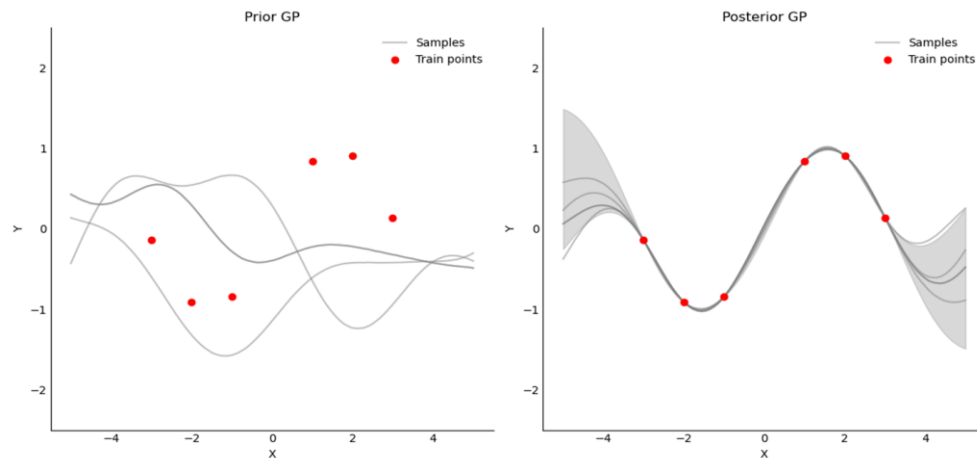
▪ Single-output GP regression^[2]

$$(x_i, y_i)_{i=1, \dots, N} \in \mathcal{X} \times \mathbb{R}, \quad y_i = f(x_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Prior: $f(x) \sim \mathcal{GP}(0, k(x, x'))$, **Kernel:** $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$

Posterior: $f(x_*)|y \sim \mathcal{N}(f_*, \text{cov}(f_*))$

- $f_* = k_*^T (k(X, X) + \sigma^2 I)^{-1}$
- $\text{cov}(f_*) = k(x_*, x_*) - k_*^T (k(X, X) + \sigma^2 I)^{-1} k_*$



▪ Multi-Output GP (MOGP)^[3]

$$(x_i, \mathbf{y}_i)_{i=1, \dots, N} \in \mathcal{X} \times \mathbb{R}^P$$

$$\mathbf{y}_i = \mathbf{f}(x_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \Sigma)$$

Prior: $\mathbf{f}(x) \sim \mathcal{GP}(0, \mathbf{k}(x, x'))$

Kernel: $\mathbf{k}(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^P \times \mathbb{R}^P$

→ Matrix valued kernel

→ Symmetric positive definite

Kernel models:

- Intrinsic coregionalization model

$$\mathbf{k}(X, X) = \mathbf{B} \otimes k(X, X)$$

- Linear model of coregionalization

$$\mathbf{k}(X, X) = \sum_{q=1}^Q B_q \otimes k_q(X, X),$$

$$B_q \in \mathbb{R}^{P \times P}, : k_q(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$$

[2] Rasmussen, C. E. (2003)

[3] Alvarez, M. A., Rosasco, L., & Lawrence, N. D. (2012)

Appendix B

Lotka-Volterra Constraint :

- Constraint given by the Hamiltonian structure^[5] of (1):
$$H(p, q) = dp - c \log(p) + bq - a \log(q) = C_0$$
- Reparametrized Hamiltonian structure:
$$dp - cs + bq - ar = C_0$$
with $s = \log(p), r = \log(q)$

Constraint evaluation metric:

- $h(t) = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} h_i(t),$
- $h_i(t) = d\hat{p}_i(t) - c\hat{s}_i(t) + b\hat{q}_i(t) - a\hat{r}_i(t) - c_i(t)$

RRMSE metric :

$$RRMSE^2 = \frac{1}{N} \sum_{i=1}^N RRMSE^2(y_{sim}^i, y_{pred}^i)$$
$$RRMSE^2(y_{sim}^i, y_{pred}^i) = \frac{\|y_{sim}^i - y_{pred}^i\|_2^2}{S \|y_{sim}^i\|_\infty^2}$$

Appendix B

Uncertainty propagation

Parameters of the $k - \varepsilon$ standard turbulence model calibrated on experimental data :

- C_μ : constant
- σ_ε : The Prandtl turbulent dissipation number
- C_{ε_1} : The source term coefficient
- C_{ε_2} : The well term coefficient
- σ_k :

C_μ	σ_ε	σ_k	C_{ε_1}	C_{ε_2}
0.09	1.30	1.00	1.44	1.92

Table: The usual values of the calibrated parameters

PCA for multi-field data

We provided 3 theoretical results in order to compare these three approaches:

- ❑ **Row-wise PCA vs field-wise PCA:** The excess of row-wise PCA error for a field k is guaranteed to be bounded by a small number when the **covariance matrices** of the other fields are close to that of field k
- ❑ **Col-wise PCA vs field-wise PCA:** A similar excess-error bound holds for col-wise PCA, with similarity measured through the **Gram matrices**

➡ These 2 results provide theoretical justification of using row-wise PCA and col-wise PCA over field wise when these similarities hold

- ❑ **Row-wise PCA vs col-wise PCA:** We give an exact dominance criterion which state that neither strategy is intrinsically superior

Numerical Applications

Lotka–Volterra equations — Population Dynamics

Comparison between our proposed framework row-wise PCA-Constrained MOGP and deductive alternatives based on:

- field-wise PCA
- column-wise PCA

Main findings

- Strong agreement between our proposed theoretical criterion and observed RRMSE rankings
 - It can therefore be helpful in practice to evaluate how much we can loose by relying on row-wise PCA
- In data-scarce regimes, the proposed framework consistently achieves the lowest RRMSE as the latent dimension increases.