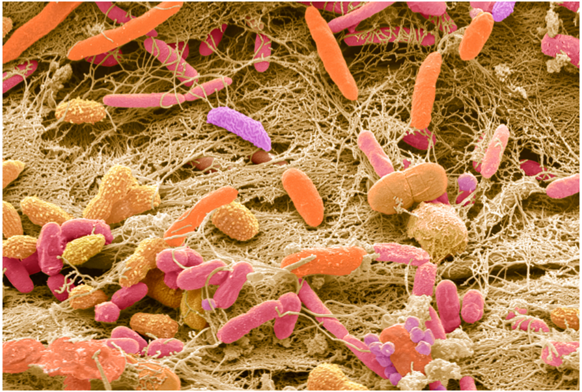
Metamodeling for metabolic models: application to a PDE model of Salmonella infection

Simon Labarthe, Clémence Frioux,
David Sherman, Thibault Malou,
Julien Martinelli

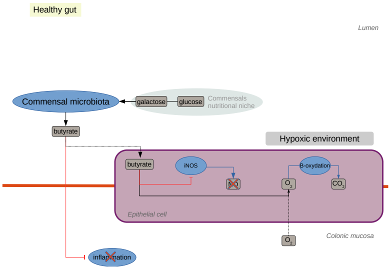
Biological context: human microbiomes

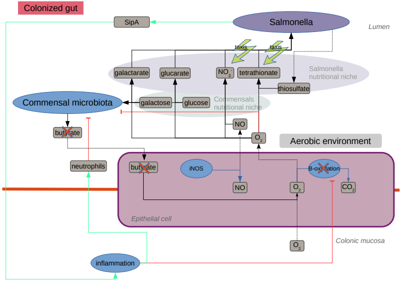
- 10^{13} bacterias in human host
- **Hundreds of different species**
- Bacterial metagenome 150 times bigger than human genome
- From a bacterial genome we can infer the **metabolic functions** of the organism



Case study: food poisoning with Salmonella

- Complex ecological interactions
- Gut microbiome involved in health (commensal bacteria) and disease (salmonella)
- Spatiotemporal mechanisms perturbed by infection





Model the infection: track the space/time evolution of bacteria and metabolites

Population dynamics

In the gut, metabolites $(m_i)_{0 \leq i \leq N_m}$ and bacteria $(b_i)_{0 \leq i \leq N_b}$ interact with each other.
(compounds, $N_c = N_m + N_b$)
A spatialized population dynamics model (for one bacteria):

$$\partial_t b + \mathscr{T}_b(b) = \nu_b(m) b \tag{1}$$

$$\partial_t m_i + \mathscr{T}_m(m) = \nu_i(m) b \tag{2}$$

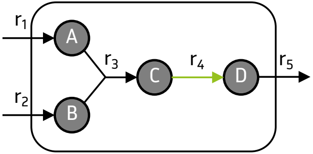
where $\mathscr{T}_b$ and $\mathscr{T}_m$ are transport terms depending on fluid dynamics model and chemotactic speed field.
A global population dynamic model:

$$\partial_t b = \nu_b(m) b \tag{3}$$

$$\partial_t m_i = \nu_i(m) b \tag{4}$$

ν_l: mass flux of compound l induced by the chemical reaction to produce the bacteria b, source term of both models.

Flux Balance Analysis (FBA)



$0 \leq \nu \leq 2$

$$\begin{array}{cccccc} r_1 & r_2 & r_3 & r_4 & r_5 & \\ \left(\begin{array}{ccccc} +1 & 0 & -1 & 0 & 0 \\ 0 & +1 & -1 & 0 & 0 \\ 0 & 0 & +1 & -1 & 0 \\ 0 & 0 & 0 & +1 & -1 \end{array}\right) & & & & & \begin{array}{c} A \\ B \\ C \\ D \end{array} \end{array}$$

Objective

$$\nu^* = \underset{\substack{S\nu=0 \\ 0=c_{\min}\leqslant\nu\leqslant c_{\max}=2}}{\operatorname{argmax}} \quad \nu_4$$

(can maximize any function of the form $\beta^T \nu$)

Steady state assumption $S\nu = 0$

$$\frac{dA}{dt} = 0 \iff \nu_1 - \nu_3 = 0$$

$$\frac{dB}{dt} = 0 \iff \nu_2 - \nu_3 = 0$$

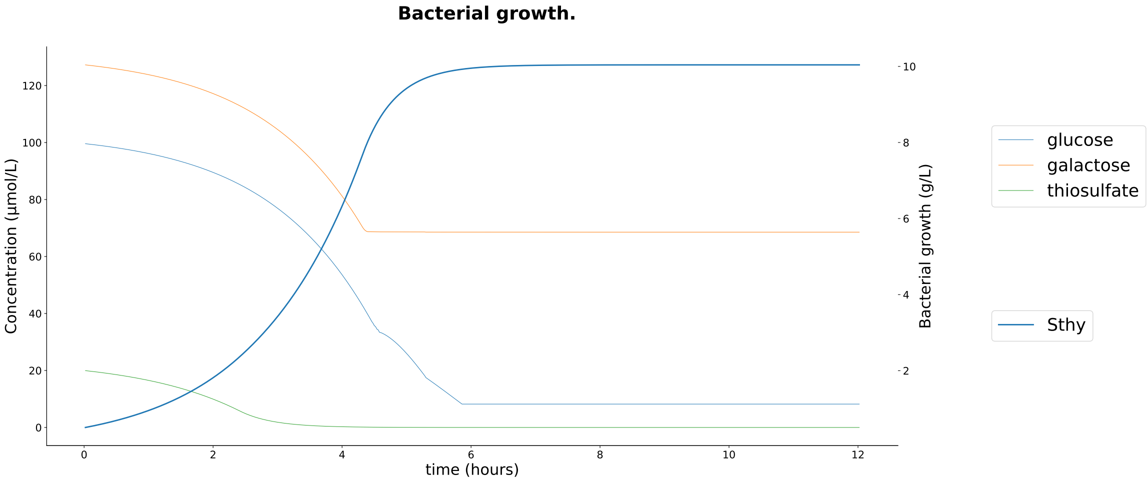
$$\frac{dC}{dt} = 0 \iff \nu_3 - \nu_4 = 0$$

$$\frac{dD}{dt} = 0 \iff \nu_4 - \nu_5 = 0$$

Maximazing ν_4 yields

$$\nu = \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 2 \end{pmatrix}$$

Example - global population dynamic model



Metamodeling: approximating FBA

From now on: $[c_{\min}, c_{\max}] = [c, 0]$

$$v = (v_1, \dots, v_{N_c})^T = (\mathscr{F}_{FBA}(c)_1, \dots, \mathscr{F}_{FBA}(c)_{N_c})^T$$

$v = \mathscr{F}_{FBA}(c)$ is expensive to compute: intractable numerical PDE solve.
Build a metamodel to approximate $c \mapsto v_i$ for each compound

We also want to investigate constraint interactions in the estimation of v.

Dataset: $(C, \mathscr{V})$ with $C = (c_i^j)_{\substack{1 \leqslant j \leqslant N_{obs} \\ 1 \leqslant i \leqslant N_m}}$ and for a given compound $\mathscr{V} = (v^j)_{1 \leqslant j \leqslant N_{obs}}$.

Hoeffding decomposition

Let $X = (X_1, \dots, X_d)$ be independent variables with law P_X and $h : \mathscr{X} \to \mathbb{R}$ s.t. $h(X) \in L^2(P_X)$.

Theorem *[Hoeffding, 1948]*
$\exists!$ expansion of h of the form

$$h(X) = h_0 + \sum_{p \in \mathscr{P}} h_p(X_p) \tag{5}$$

with:

$$\begin{aligned} h_0 &= \mathbb{E}[h(X)] \\ h_p(X_p) &= \mathbb{E}[h(X)|X_p] - \sum_{w \subsetneq p} h_w(X_w) \qquad \forall\, p \in \mathscr{P} \end{aligned}$$

All the h_p are centered and orthogonal with respect of $L^2(\mathscr{X}, P_X)$.

From which space should the estimator be picked?

Reproducing Kernel Hilbert Space (RKHS)

Let $\mathscr{H}$ be a Hilbert space of functions $f : \mathscr{X} \subset \mathbb{R}^d \to \mathbb{R}$ with the inner product $\langle \cdot, \cdot \rangle$.
$\mathscr{H}$ is a RKHS if $\forall\, x \in \mathscr{X}$ the functional

$$
\begin{aligned}
L_x : \quad \mathscr{H} &\longrightarrow \mathbb{R} \\
f &\longmapsto f(x)
\end{aligned}
$$

is continuous.

From the Riesz theorem, $\exists!\, k_x \in \mathscr{H}$ such that $\forall\, x \in \mathscr{X}, \forall\, f \in \mathscr{H}, f(x) = L_x(f) = \langle f, k_x \rangle$.
The reproducing kernel of $\mathscr{H}$ is

$$
\begin{aligned}
k : \quad \mathscr{X} \times \mathscr{X} &\longrightarrow \mathbb{R} \\
(x, x') &\longmapsto k_{x'}(x) = \langle k_x, k_{x'} \rangle
\end{aligned}
$$

Representer theorem

Let k be a positive definite kernel with corresponding RKHS $\mathscr{H}$.
Let $\mathscr{G} : \mathbb{R} \to \mathbb{R}$ be a stricly increasing function.
Given training points $(x^1, \dots, x^n)$ in $\mathscr{X}$, consider the cost function
$J(f) = L(f(x^1), \dots, f(x^n)) + \mathscr{G}(||f||^2_{\mathscr{H}})$

Theorem *[Kimeldorf & Wahba, 1970]*
If $\hat{f}$ is a function such that $J(\hat{f}) = \inf_{f \in \mathscr{H}} J(f)$, then f admits a representation of the form

$$\hat{f}(x) = \sum_{j=1}^{N_{obs}} \alpha^j k(x^j, x) \tag{6}$$

with $\alpha^j \in \mathbb{R} \; \forall j$.

Linear combination of $k(x^j, \cdot) \implies$ parametric optimization problem

In practice, $L = \sum_{j=1}^{N_{obs}} (y^j - \hat{f}(x^j))^2$.
In the case where $\mathscr{G}(||f||_{\mathscr{H}}) = \lambda ||f||^2_{\mathscr{H}}$ and L is convex, the solution exists and is unique.

Approximation of the Hoeffding decomposition in a RKHS

(Durrande *et al.*, 2013), (Huet & Taupin, 2017)

Construction of $\mathcal{H}$ from direct sums of given $\mathcal{H}_a$:

$$\mathcal{H}_{0a} = \{f_a \in \mathcal{H}_a,\ \mathbb{E}_{X_a}[f_a(X_a)] = 0\}\ \forall a \in \{1, \dots, d\}$$

$$k_{0a}(X_a, X'_a) = k_a(X_a, X'_a) - \frac{\mathbb{E}_{U\sim P_a}[k_a(X_a, U)]\mathbb{E}_{U\sim P_a}[k_a(X'_a, U)]}{\mathbb{E}_{(U,V)\sim P_a \otimes P_a}[k_a(U, V)]}$$

$$k(X, X') = \left(\prod_{a=1}^{d}(1 + k_{0a}(X_a, X'_a)\right) = 1 + \sum_{p \in \mathscr{P}} k_p(X_p, X'_p) \tag{7}$$

with $k_p(X_p, X'_p) = \prod_{a \in p} k_{0a}(X_a, X'_a)$ and $\mathcal{H}_p$ its associated RKHS. Then,

$$\mathcal{H} = \left(\prod_{a=1}^{d} \mathbb{1} \overset{\perp}{\oplus} \mathcal{H}_{0a}\right) = \mathbb{1} + \sum_{p \in \mathscr{P}} \mathcal{H}_p \tag{8}$$

Then, the approximation $\hat{f}$ of h also admits a Hoeffding decomposition, each $\hat{f}_p$ approximates h_p.

Metamodel optimization problem

From the representer theorem and the previously constructed RKHS

$$\hat{\theta}_0, (\hat{\theta}_p)_{p\in\mathscr{P}} := \underset{\substack{\theta_0 \in \mathbb{R} \\ \theta_p = (\theta_p^j)_{1\leqslant j\leqslant N_{obs}} \forall p \in \mathscr{P}}}{\operatorname{argmin}} \|\mathscr{V} - (\theta_0 \mathbb{1}^T + \sum_{p\in\mathscr{P}} K_p\theta_p)\|_2^2 + \mathscr{G}(W, \theta_p) \quad (9)$$

with $K_p \in \mathbb{R}^{N_{obs}\times N_{obs}}$ the Gram matrix such that $(K_{p\ j_1,j_2})_{1\leq j_1,j_2\leq N_{obs}} = k_p(c^{j_1}, c^{j_2})$.

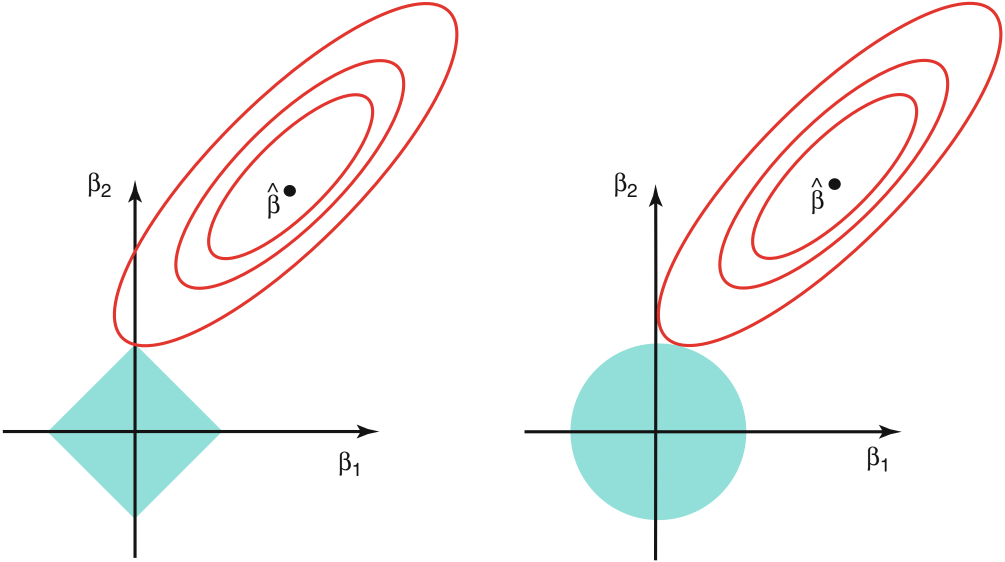
$\mathscr{G}$ is a regularization term. In this context, the regularization term is a group LASSO term:

$$\mathscr{G}(W, \theta_p) = N_{obs}\mu \sum_{p\in\mathscr{P}} \|W\theta_p\|_2$$

with μ an hyperparameter and W some weight matrix. Ideally, $W = K_p^{1/2}$. For now, $W = Id$.

L1 penaly enforces sparsity

(Figure from *An introduction to statistical learning*, Tibshirani *et al.*)



Resolution

Let $J(\theta_0, \theta) = \|\mathscr{V} - \theta_0 \mathbb{1}^T - \sum_{p\in\mathscr{P}} K_p\theta_p\|^2 + \sum_{p\in\mathscr{P}} \|W\theta_p\|_2$

Algorithm 1: RKHS group lasso

Set $\theta = [0]_{|\mathscr{P}|\times N_{obs}}$

repeat

  Compute $\theta_0 = \operatorname{argmin}_{\theta_0} J(\theta_0, \theta) = \frac{1}{N_{obs}} \sum_{j=1}^{N_{obs}} \left(\nu^j - \sum_{p\in\mathscr{P}} (K_p\theta_p)_j\right)$

  for $p \in \mathscr{P}$ **do**

    Compute $R_p = \mathscr{V} - \theta_0 \mathbb{1}^T - \sum_{p\neq w} K_w\theta_w$

    if $\|\frac{2}{\sqrt{N_{obs}}} W^{1/2} R_p\| \leqslant \mu$ **then**

      $\theta_p \leftarrow 0$; `// Soft thresholding`

    else

      $\theta_p \leftarrow \operatorname{argmin}_{\theta_p} J(\theta_0, \theta)$; `// one step of gradient descent`

until convergence.

Prediction

Once the metamodel is trained, we can use it to predict the flux on an unseen constraint sample c^{new}:

$$\mathscr{F}_{MM}(c^{new}) = \theta_0 + \sum_{j=1}^{N_{obs}} \sum_{p \in \mathscr{P}} \theta_p^j k_p(c^j, c^{new}) \tag{10}$$

Prediction

Once the metamodel is trained, we can use it to predict the flux on an unseen constraint sample c^{new}:

$$\mathscr{F}_{MM}(c^{new}) = \theta_0 + \sum_{j=1}^{N_{obs}} \sum_{p \in \mathscr{P}} \theta_p^j k_p(c^j, c^{new}) \tag{10}$$

Dataset generation

1. Simulate multiple ODE trajectories with random initial conditions y_0
2. Subsample the obtained couples (c, v) and remove duplicates.
3. Apply perturbation $c^{pert} = c\varepsilon,\ \varepsilon \sim \mathscr{N}(1, \sigma)$ and compute $v^{pert} = \mathscr{F}_{FBA}(c^{pert})$

Prediction

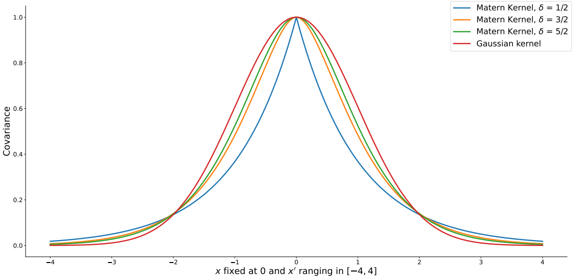
Once the metamodel is trained, we can use it to predict the flux on an unseen constraint sample c^{new}:

$$\mathscr{F}_{MM}(c^{new}) = \theta_0 + \sum_{j=1}^{N_{obs}} \sum_{p \in \mathscr{P}} \theta_p^j k_p(c^j, c^{new}) \tag{10}$$

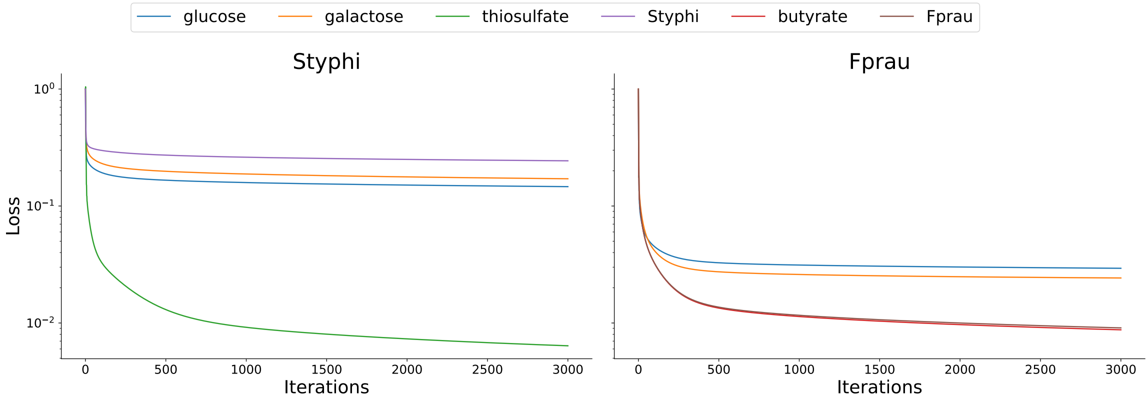
Dataset generation

1. Simulate multiple ODE trajectories with random initial conditions y_0
2. Subsample the obtained couples (c, v) and remove duplicates.
3. Apply perturbation $c^{pert} = c\varepsilon,\ \varepsilon \sim \mathcal{N}(1, \sigma)$ and compute $v^{pert} = \mathscr{F}_{FBA}(c^{pert})$

We use as kernel the Matern kernel for k_a

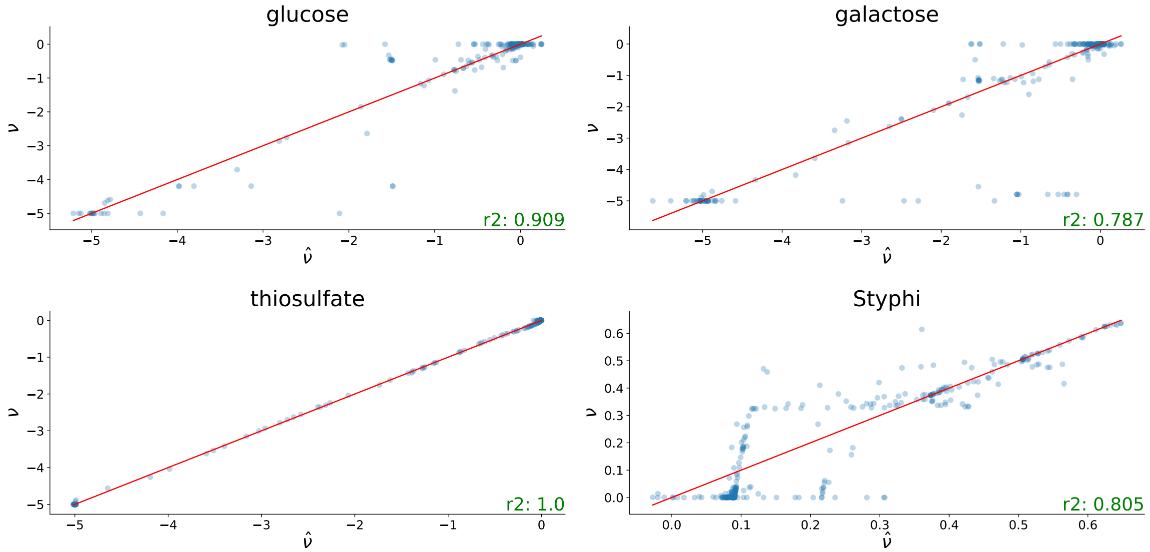


Optimization results ($N_{obs} = 2100, \mu = 0$)

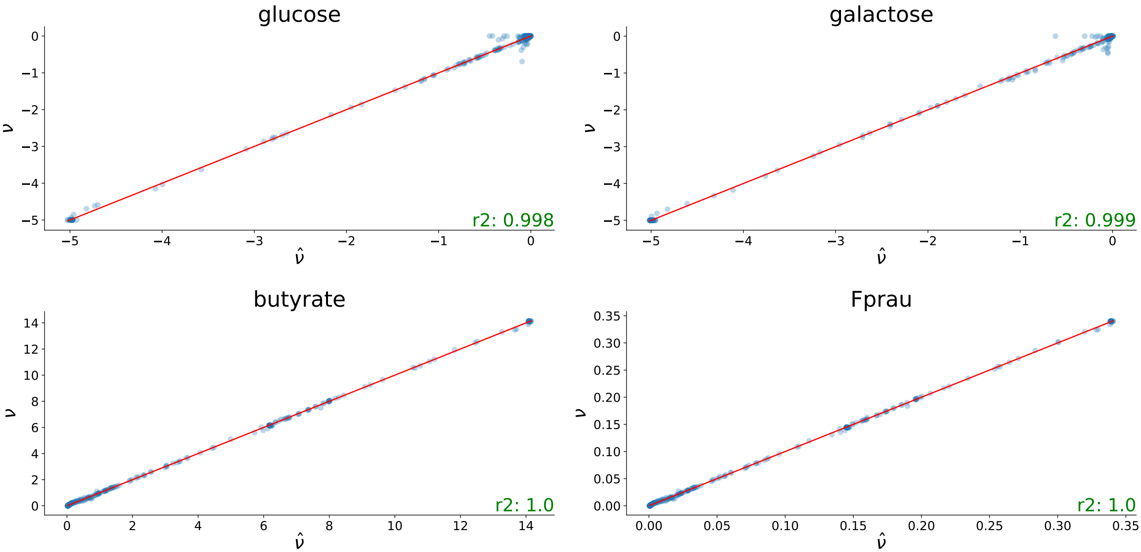


$$\text{Loss: } \frac{\|\hat{\nu}_i - \nu_i\|_2}{\|\nu_i\|_2}$$

QQ plots on 300 unseen (constraints, fluxes) couples

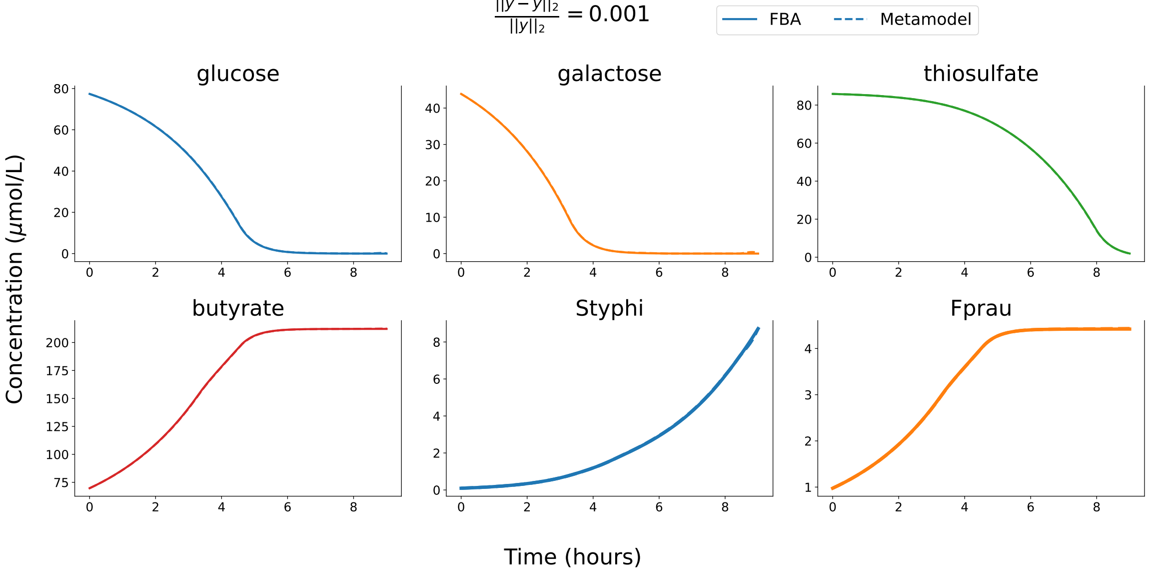


QQ plots on 300 unseen (constraints, fluxes) couples

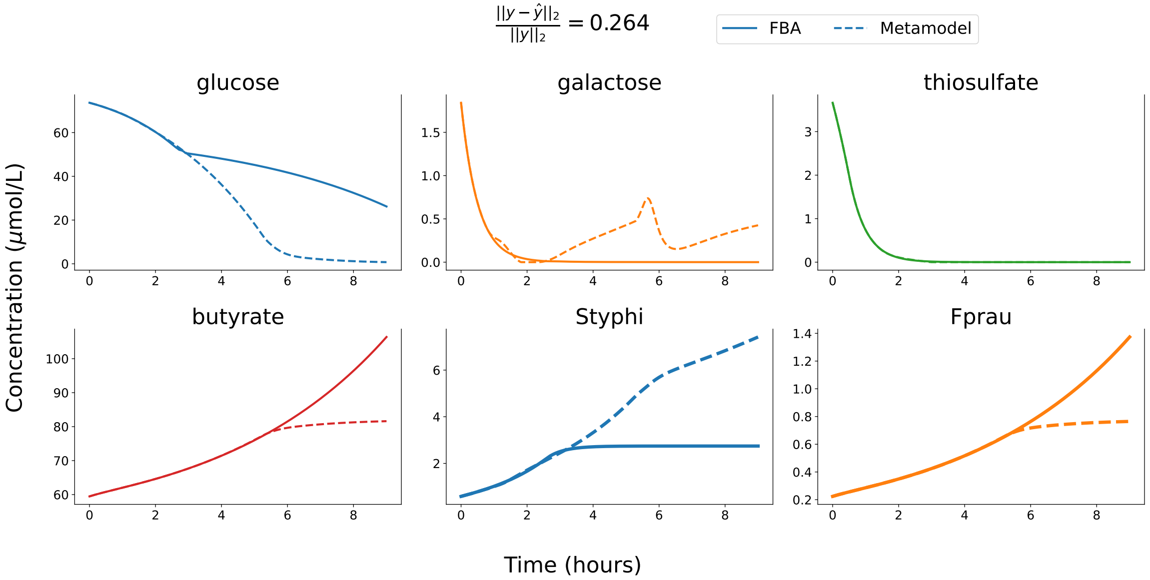


Simulated trajectory - perfect approximation

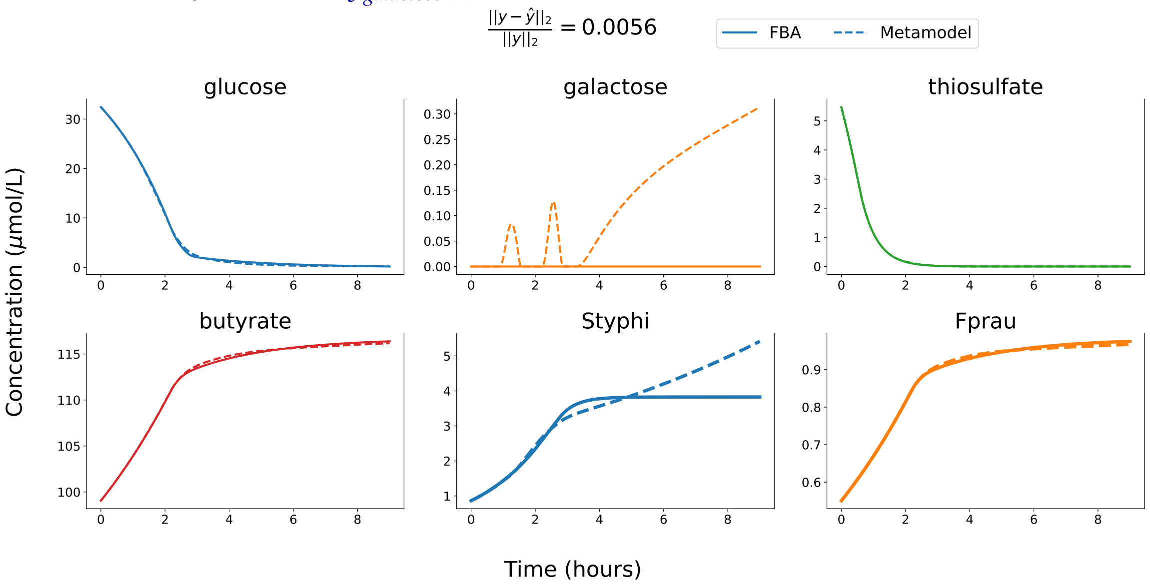
$$\frac{||y - \hat{y}||_2}{||y||_2} = 0.001$$



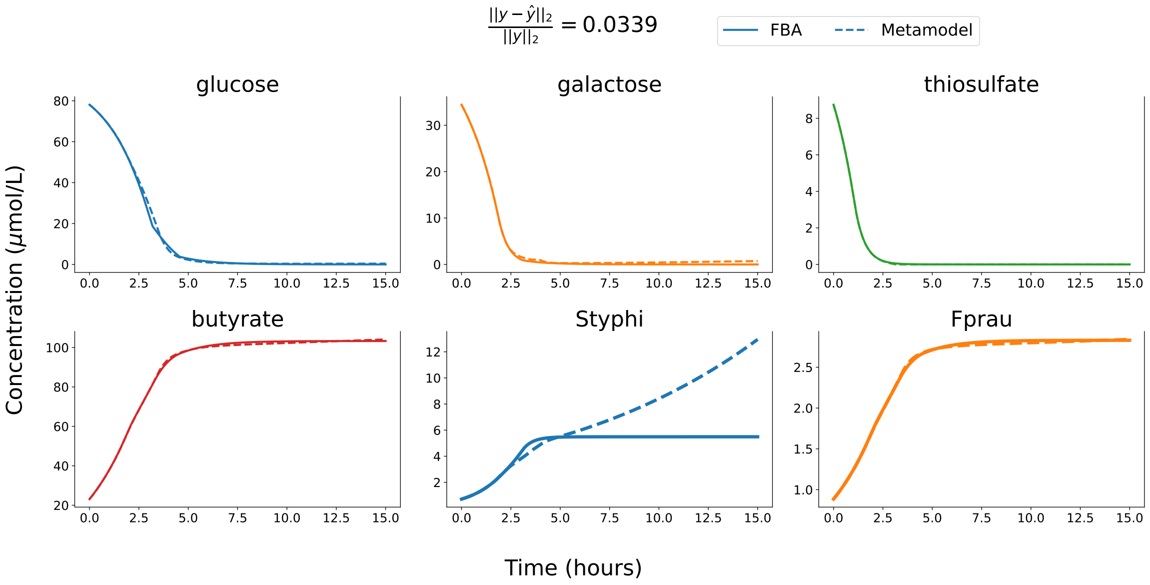
Simulated trajectory - poor approximation



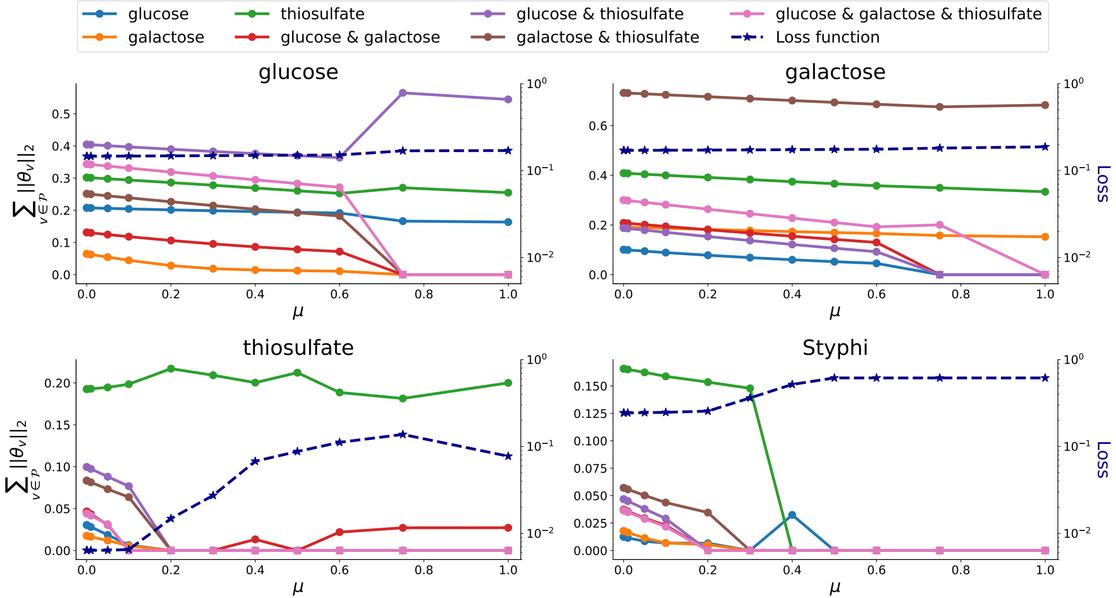
Simulated trajectories - $y_{galactose}(0) = 0$



Simulated trajectory - increased time horizon

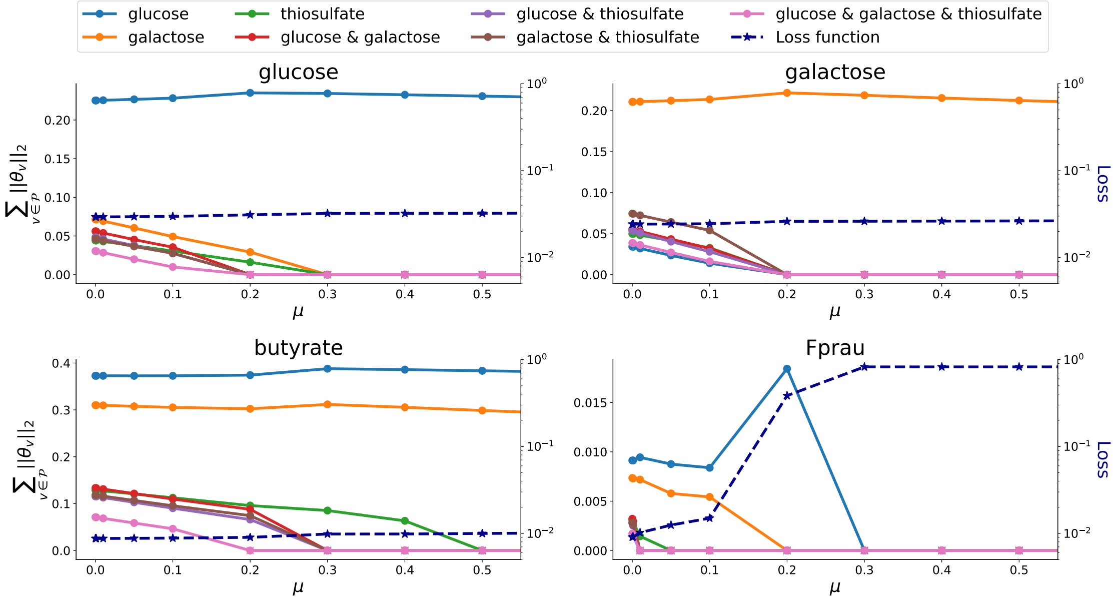


Lasso paths



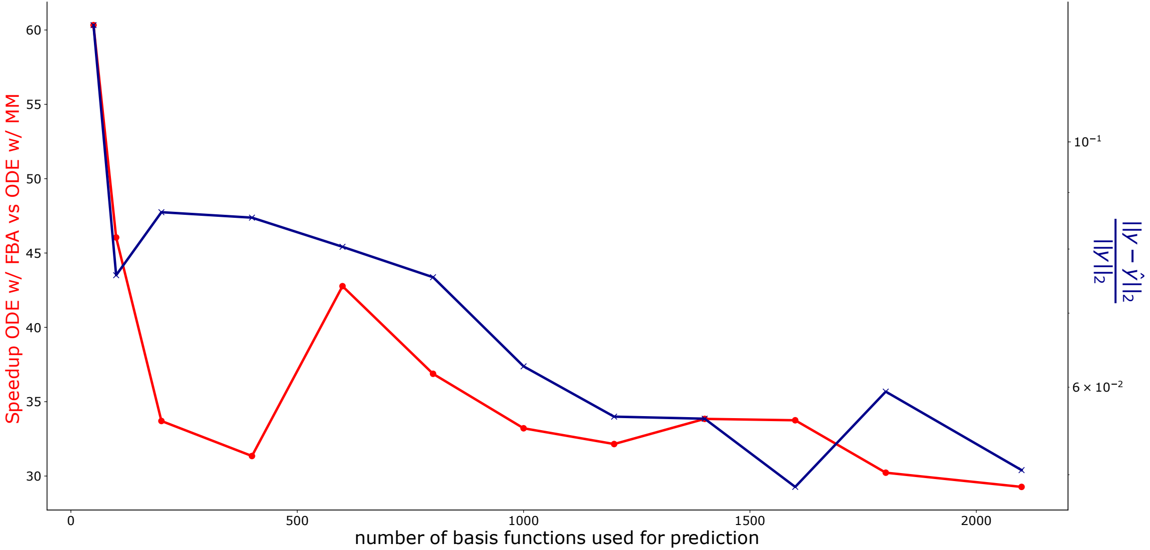
glucose
galactose
thiosulfate
glucose & galactose
glucose & thiosulfate
galactose & thiosulfate
glucose & galactose & thiosulfate
Loss function
glucose
galactose
thiosulfate
Styphi
$\sum_{v \in \mathcal{P}} ||\theta_v||_2$
μ
Loss

Lasso paths



glucose
galactose
thiosulfate
glucose & galactose
glucose & thiosulfate
galactose & thiosulfate
glucose & galactose & thiosulfate
Loss function
glucose
galactose
butyrate
Fprau
$\sum_{v \in \mathcal{P}} ||\theta_v||_2$
μ
Loss

Speedup: ODE using FBA versus ODE using metamodel



Conclusion

- Python library for:
  - Generating and simulating different population dynamics experiments
  - Learning metamodel of FBA

Conclusion

- Python library for:
  - Generating and simulating different population dynamics experiments
  - Learning metamodel of FBA

- Metamodel build to approximate the Hoeffding decomposition
  - Enables fast computations
  - Allows to investigate the crucial interactions $p \in \mathscr{P}$ of a model.

Conclusion

- Python library for:
  - Generating and simulating different population dynamics experiments
  - Learning metamodel of FBA

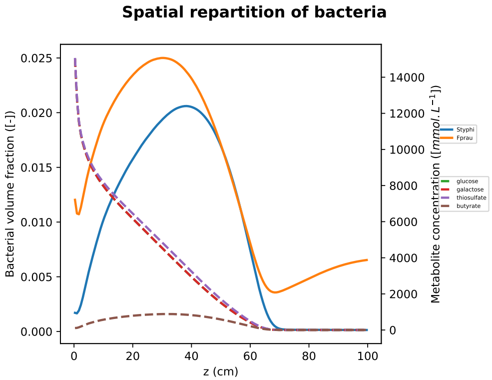
- Metamodel build to approximate the Hoeffding decomposition
  - Enables fast computations
  - Allows to investigate the crucial interactions $p \in \mathscr{P}$ of a model.

- First results convincing, but some dynamics are not well caught
  - Lack of goodness of fit when fluxes are 0 can be corrected.

Perspectives

- Exploring solutions to preserve null fluxes (e.g. another kernel?)
- Test different regularization terms
- Regularization on observations rather than on subsets
- Try other optimization methods
- Integration of the results into the PDE model

PDE Simulation



- 2D Finite volume resolution of a system of population dynamics equation with a fluid dynamics model of the gut content (CEMRACS 2015 tribute)
- 150×20 space cells $\times \sim 560$ time integration steps $\Rightarrow$ 1 M model evaluations.
- Longitudinal distribution of radial averaged concentrations.
- Proof of concept: not physiological $\Rightarrow$ further improvements needed.