French Identification group

Gaussian process modeling of bioprocesses: application to chinese ovary hamster cells cultivated in bioreactors

K. Colin, M. Mäkinen, H. Schwarz, V. Chotteau, E.W. Jacobsen, H. Hjalmarsson

KTH Royal Institute of Technology

November 25, 2021

Context of the study

- Monoclonal antibodies (mAb): inhibition of viral infection.

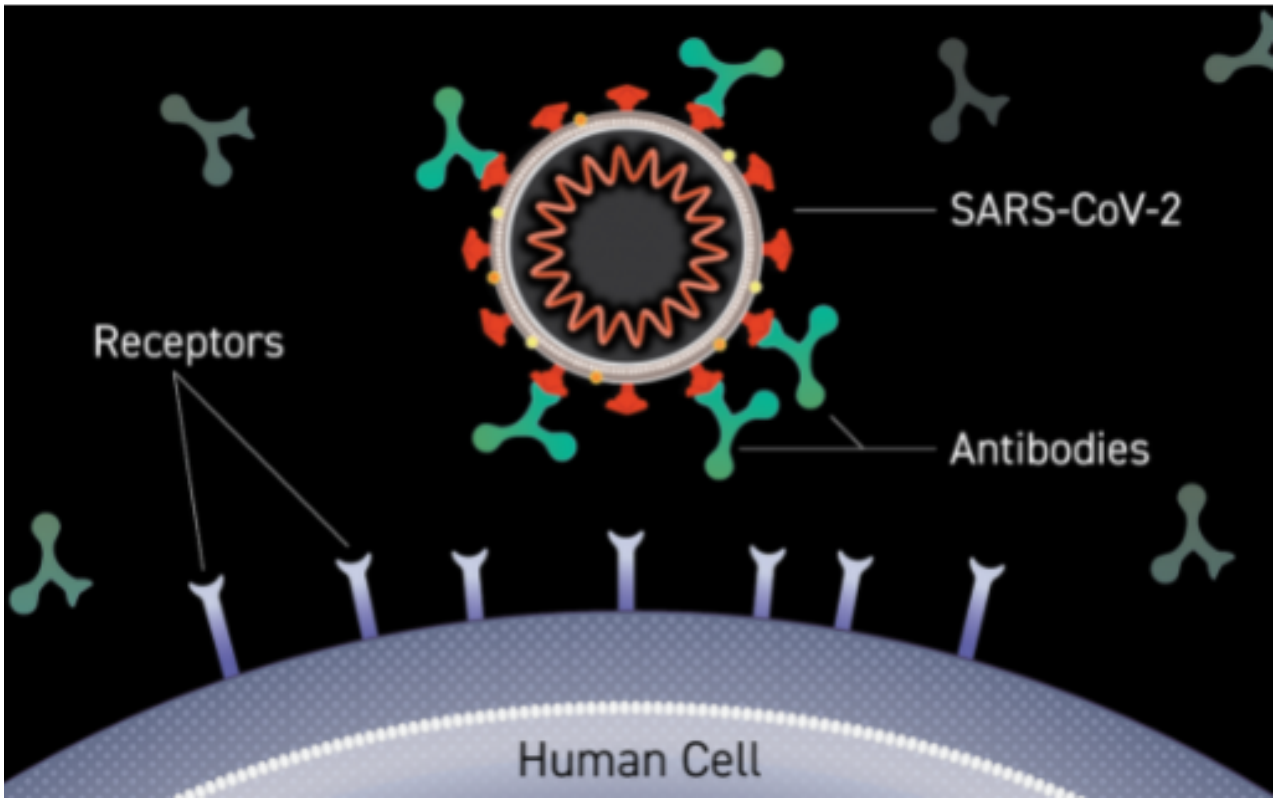
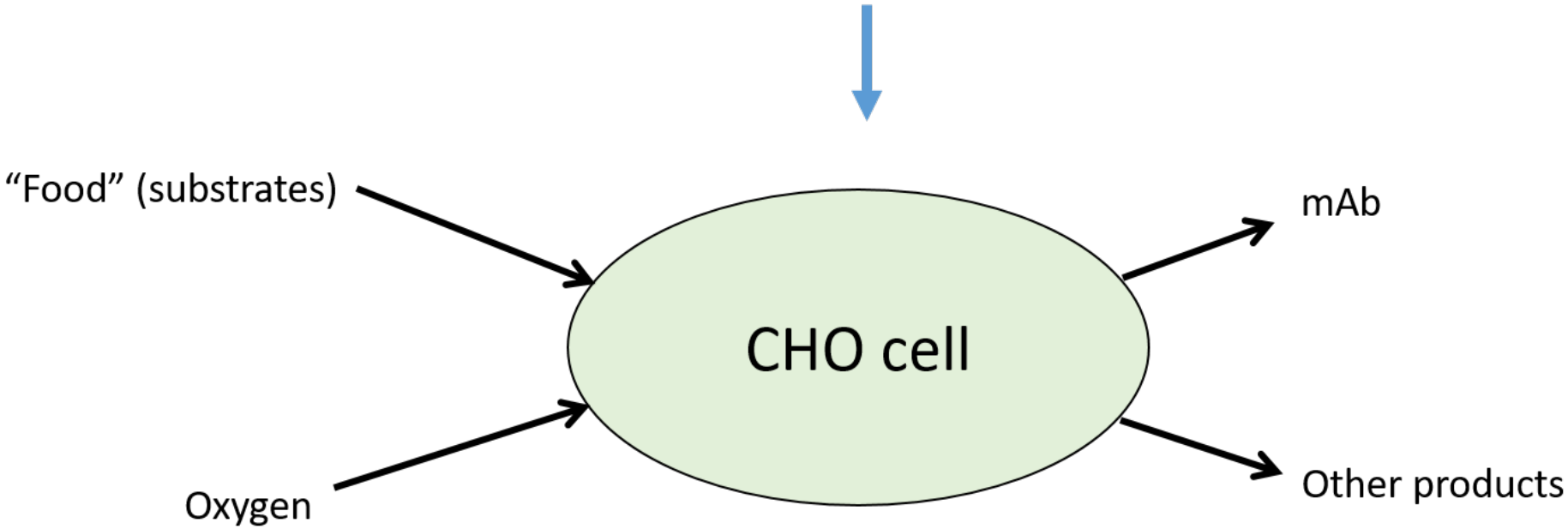


Figure: Scheme of mAb against COVID-19[1]

[1] Source: website of National Institutes of Health (https://www.nih.gov/)

Context of the study

- In industry, mAb is produced by Chinese Hamster Ovary (CHO) cells.



Context of the study

- Chinese Hamster Ovary (CHO) cells cultivated in large-scale bioreactors.

Figure: Picture of a bioreactor[2]

[2] https://www.engr.colostate.edu/CBE101/topics/bioreactors.html

Context of the study

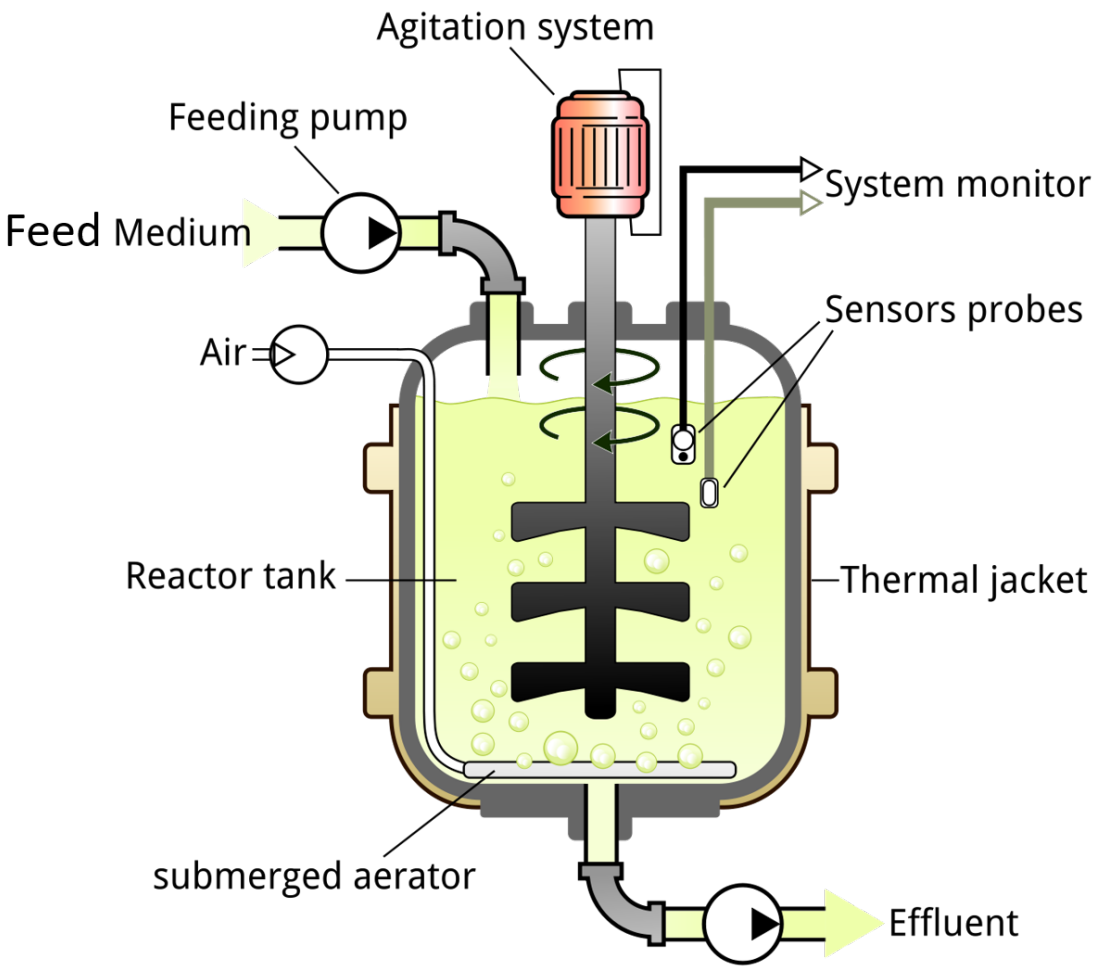


Figure: Scheme of a bioreactor[3]

[3] https://en.wikipedia.org/wiki/Bioreactor/media/File:Bioreactor_principle.svg

Context of the study

- Producing mAb is expensive (and so are the treatments)!
- Ronapreve approved by EU commission against COVID-19 at 1.700 EUR[4]
- **Goal**: maximization of the yield of mAb produced by CHO cells.
- Which variable should we change for the maximization?

[4] https://www.leparisien.fr/societe/sante/ronapreve-regkirona-ce-que-lon-sait-des-deux-traitements-approuves-par-lema-11-11-2021-U64NC7CAQJE2XGVSVW2QF7FIBM.php

Context of the study

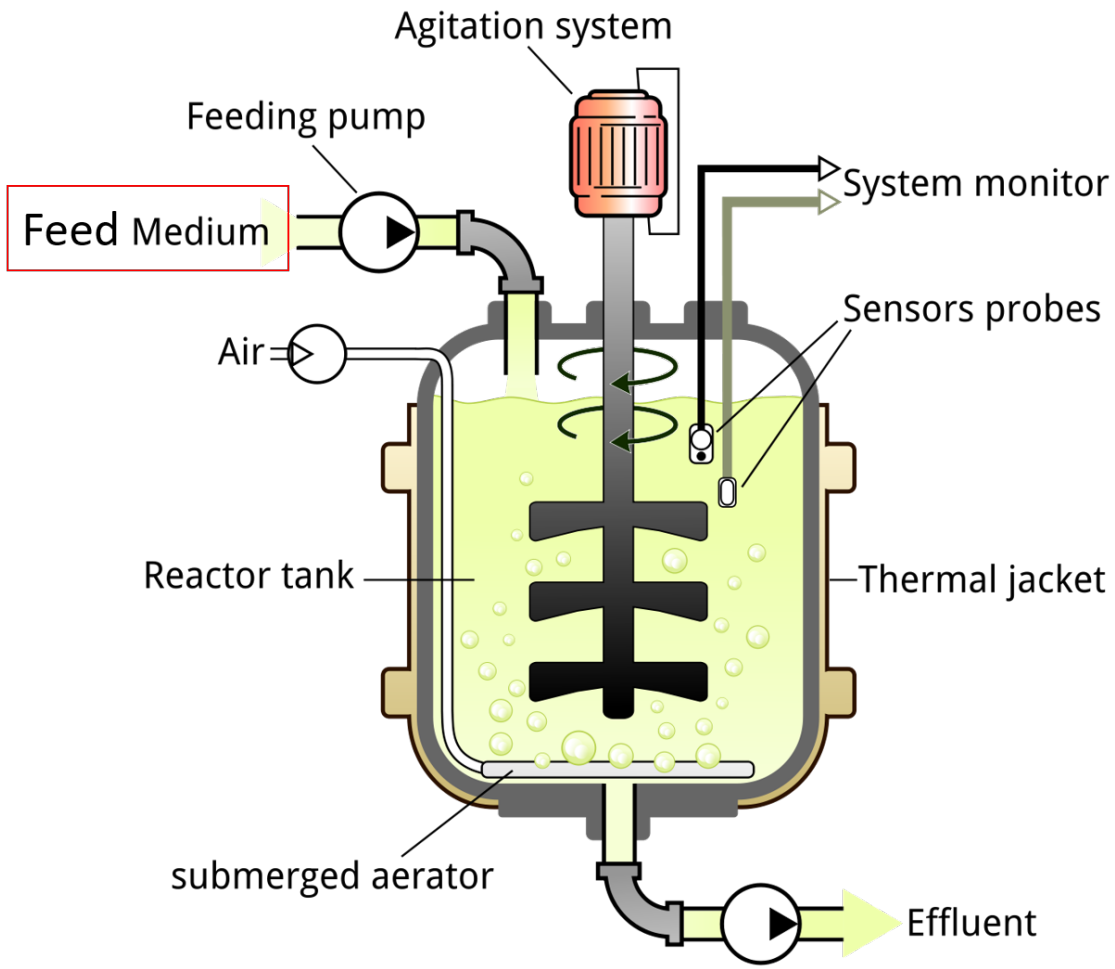
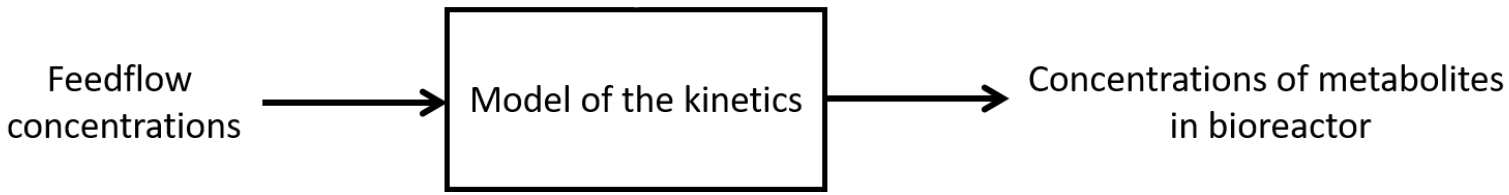


Figure: Scheme of a bioreactor

How to optimize the feed medium?

- Optimize experimentally? No, too long and expensive!
- Solution: model-based optimization.
- What to model? Kinetic evolution of the concentrations in the bioreactor w.r.t. variations of the feed medium concentrations.



Some first principles?

- The model should satisfy some first principles such as mass-balance equation

$$\begin{array}{c}\text{Evolution}\\ \text{of concentration}\end{array} = \begin{array}{c}\text{What the cells}\\ \text{consume/produce}\end{array} + \begin{array}{c}\text{What the feed}\\ \text{medium brings}\end{array} - \begin{array}{c}\text{What the effluent}\\ \text{takes out}\end{array}$$

- With mathematics (x = concentration vector)

$$\frac{dx}{dt}(t) = q(x(t)) + F_{in}(t)x_{feed}(t) - F_{out}(t)x(t)$$

where

- q: uptake/secretion rate (quantity of metabolites consumed/produced by the cells during one day)
- F_{in}: flow rate of feed medium.
- F_{out}: flow rate of effluent.
- x_{feed}: concentration of metabolites in the feed medium.

Some first principles?

- The model should satisfy some first principles such as mass-balance equation

$$\begin{array}{c}\text{Evolution}\\ \text{of concentration}\end{array} = \begin{array}{c}\text{What the cells}\\ \text{consume/produce}\end{array} + \begin{array}{c}\text{What the feed}\\ \text{medium brings}\end{array} - \begin{array}{c}\text{What the effluent}\\ \text{takes out}\end{array}$$

- With mathematics (x = concentration vector)

$$\frac{dx}{dt}(t) = q(x(t)) + F_{in}(t)x_{feed}(t) - F_{out}(t)x(t)$$

where

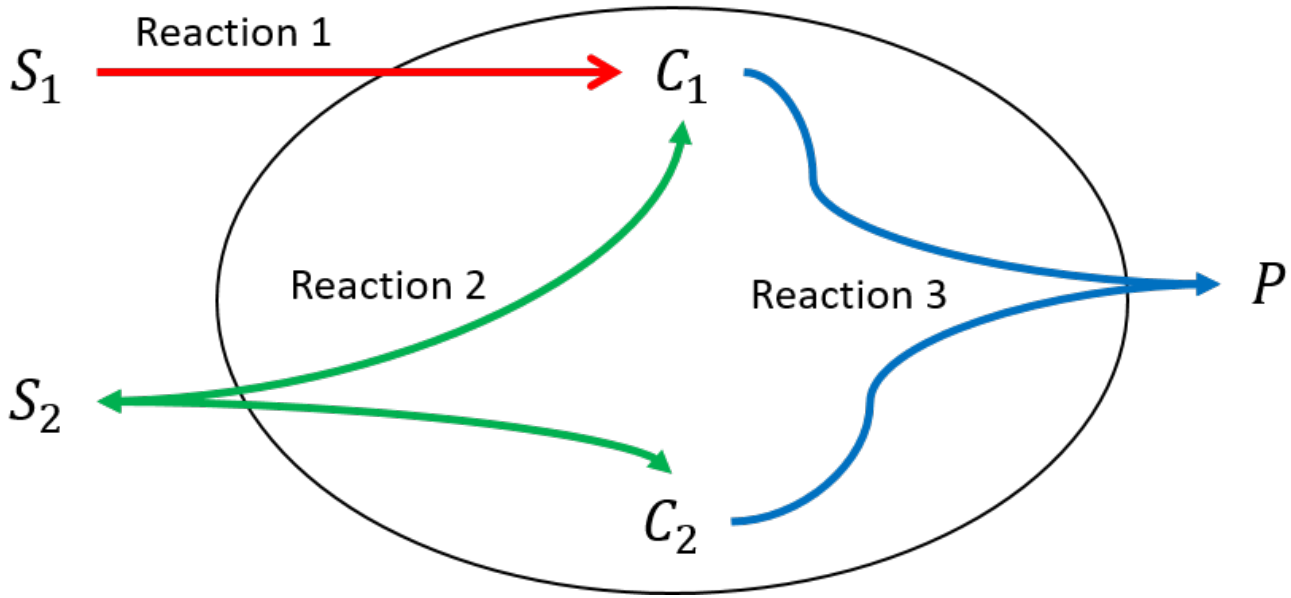
- q: uptake/secretion rate (quantity of metabolites consumed/produced by the cells during one day)
- F_{in}: flow rate of feed medium.
- F_{out}: flow rate of effluent.
- x_{feed}: concentration of metabolites in the medium feed.

How to model $q(x(t))$?

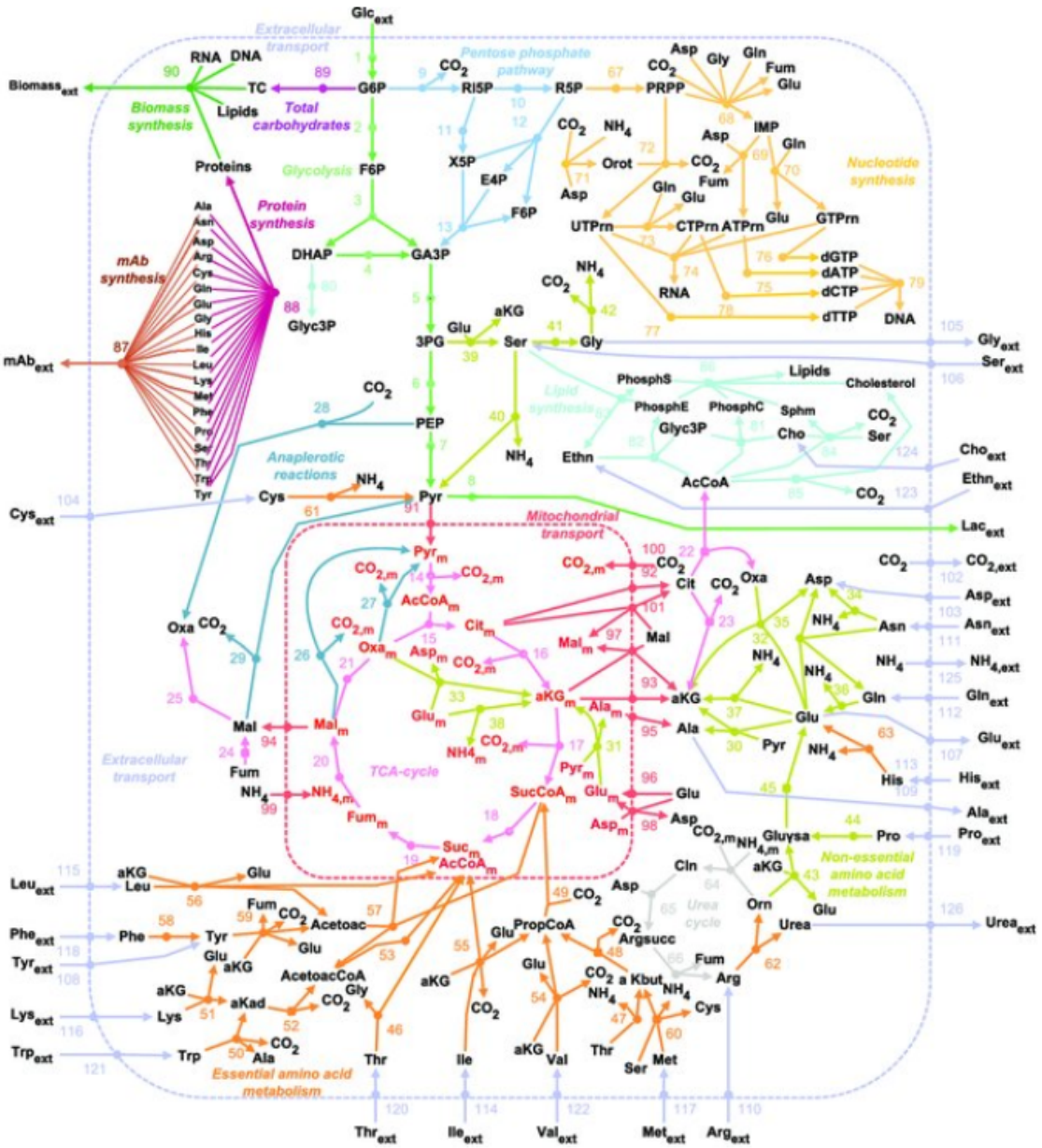
- The vector $q(x(t))$ describes the kinetics of the chemical reactions inside the cells.

- To model $q(x(t))$, we need 3 ingredients
  - Connections between the metabolites (metabolic network).
  - Stoichiometric coefficients of the reactions (stoichiometric matrix)
  - Rate (speed) of all the chemical reactions (kinetic expression)

Ingredient 1: metabolic network



REAL metabolic network



Ingredient 2: Stoichiometric matrix

- Stoichiometric coefficients:
  - Reaction 1: $S_1 \rightarrow 2C_1$
  - Reaction 2: $S_2 \rightarrow C_1 + C_2$
  - Reaction 3: $C_1 + 3C_2 \rightarrow P$

$$\mathbf{A} = \left[\begin{array}{c|ccc} & \text{Reaction 1} & \text{Reaction 2} & \text{Reaction 3} \\ \hline S_1 & -1 & 0 & 0 \\ S_2 & 0 & -1 & 0 \\ C_1 & 2 & 1 & -1 \\ C_2 & 0 & 1 & -3 \\ P & 0 & 0 & 1 \end{array}\right]$$

- Then,

$$q(x) = \mathbf{A}y(x)$$

Ingredient 3: Kinetic expression of the rates

- **Parametric** expression for the rate vector $y(x)$.

- For biochemical reactions: Monod kinetics

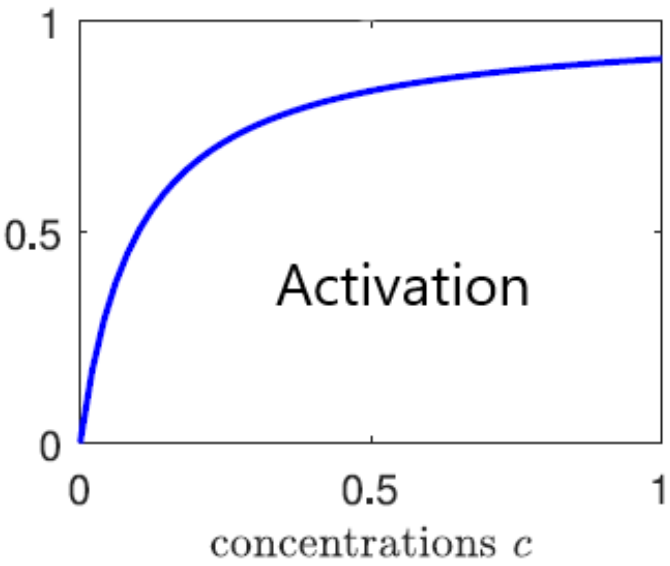
$$\text{Rate of Reaction i} = \prod_{j=1}^{n} \text{Modulation function}(x_j, \eta_{ij})$$

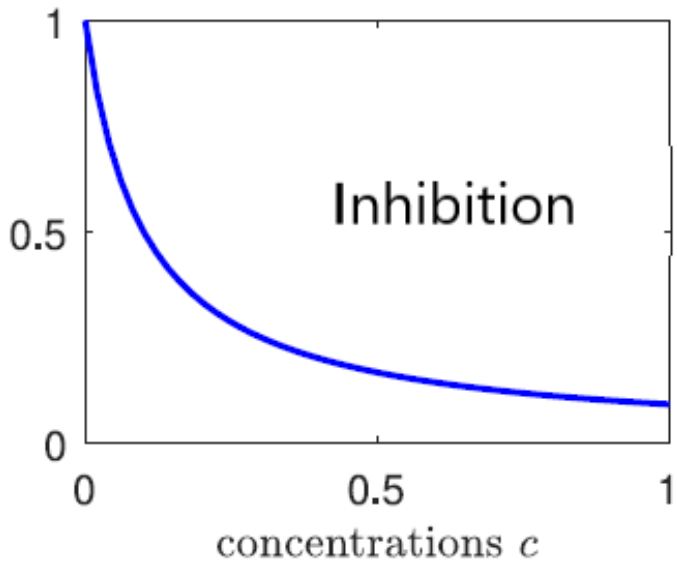
- 4 types of modulation functions:

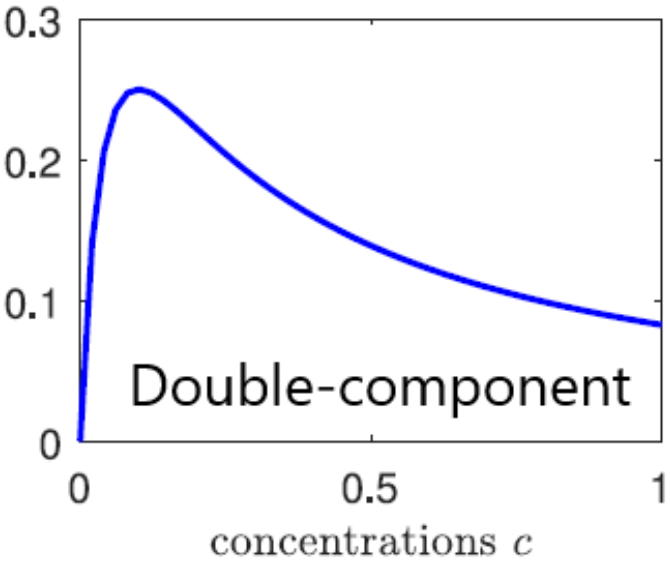
$$\text{Activation :} \quad \frac{x_j}{x_j + \theta_{ij}} \qquad \text{Inhibition :} \quad \frac{1}{1 + \mu_{ij} x_j}$$

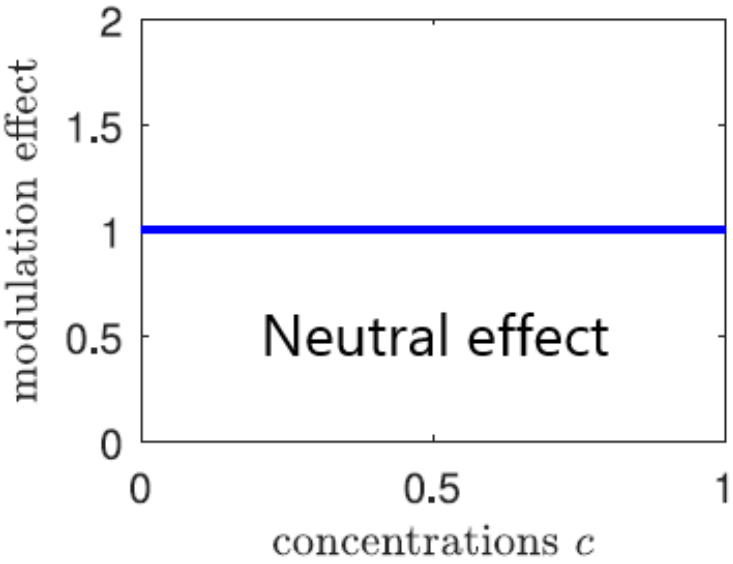
$$\text{Double-component :} \quad \frac{x_j}{x_j + \theta_{ij}} \frac{1}{1 + \mu_{ij} x_j} \qquad \text{Neutral effect :} \quad 1$$

Ingredient 3: Kinetic expression of the rates









Final parametrized dynamic model

- Final mass balance-equation with kinetics

$$\frac{dx}{dt}(t) = \mathbf{A}y(x(t), \eta) + F_{in}(t)x_{feed}(t) - F_{out}(t)x(t)$$

- Goal: identify η.

- Everything is measured/known except $y(x(t), \eta)$. We can estimate the data of the rates!

From dynamic to static map modeling

- Data: $\{y(t), x(t)\}_{t=1}^{N}$.

- The identification problem for the i-th rate is

$$\eta_i^* = \text{arg min} \sum_{t=1}^{N} ||y_i(t) - y_{m,i}(x(t), \eta_i)||^2$$

where

$$y_{m,i}(x(t), \eta_i) = \prod_{j=1}^{n} \frac{x_j(t)}{(x_j(t) + \theta_{ij})(1 + \mu_{ij} x_j(t))}$$

- We have transformed the nonlinear dynamic identification problem into a nonlinear static map identification.

Design of a kernel for Gausian process regression of Monod functions

The identification problem

- Focus on one rate.

- The identification problem is

$$\eta^* = \text{arg min} \sum_{t=1}^{N} ||y(t) - y_m(x(t), \eta)||^2$$

where

$$y_m(x(t), \eta) = \prod_{j=1}^{n} \frac{x_j(t)}{(x_j(t) + \theta_j)(1 + \mu_j x_j(t))}$$

The identification problem

- Local minima: we need a good initialization of the parameters.

- Main idea: perform a GP regression of **each** modulation function as intermediate modeling method for the initialization.

Step 1: Gaussian process modeling

- Assume additive noise:

$$y(t) = \prod_{j=1}^{n} h_j(x_j(t), \eta_j) + \epsilon(t)$$

where ϵ is a white Gaussian noise of variance σ_e^2.

- Each function h_j is modeled as a zero-mean Gaussian process

$$h_j \sim \mathcal{GP}(0, k_j)$$

- We need an appropriate kernel function for the modeling of the 4 types of Monod functions.

Kernel in the literature

- Kernel proposed in the literature[5]

$$k_j(x_j(t), x_j(t'), \Theta_j) = \gamma_j \left(\frac{x_j(t)}{x_j(t')} \right)^{-\delta_j \log\left(\frac{x_j(t)}{x_j(t')} \right)}$$

[5] Wang, M., Risuleo, R. S., Jacobsen, E. W., Chotteau, V., Hjalmarsson, H. (2020). Identification of nonlinear kinetics of macroscopic bio-reactions using multilinear Gaussian processes. Computers Chemical Engineering, 133, 106671

Kernel in the literature

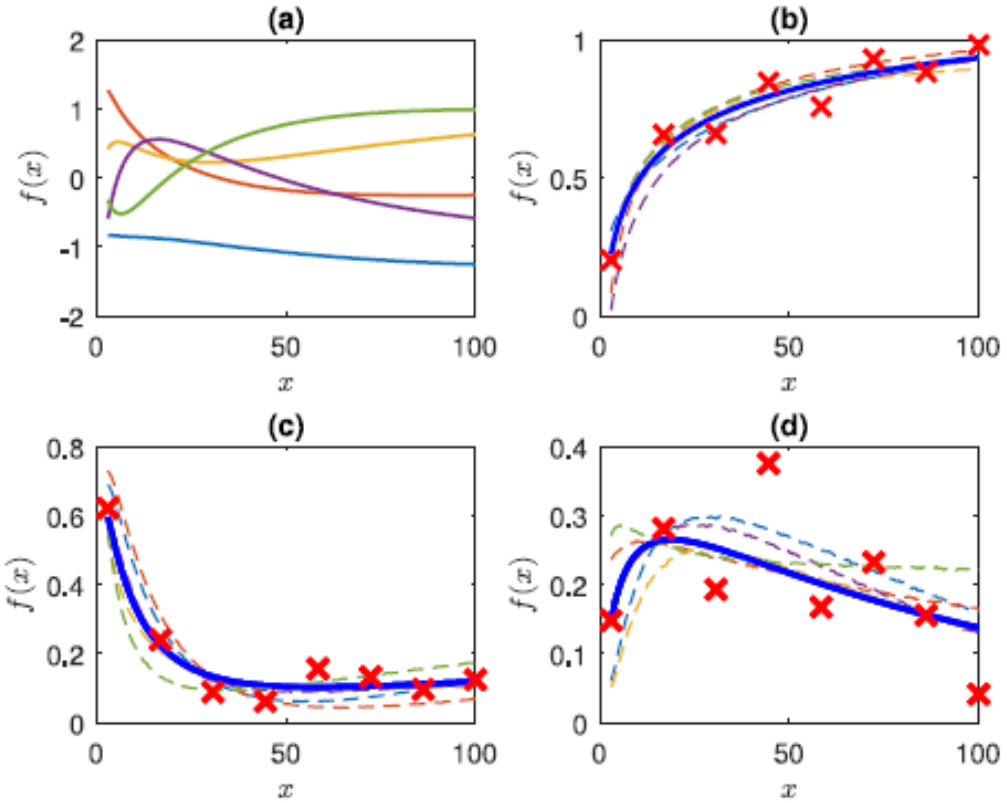


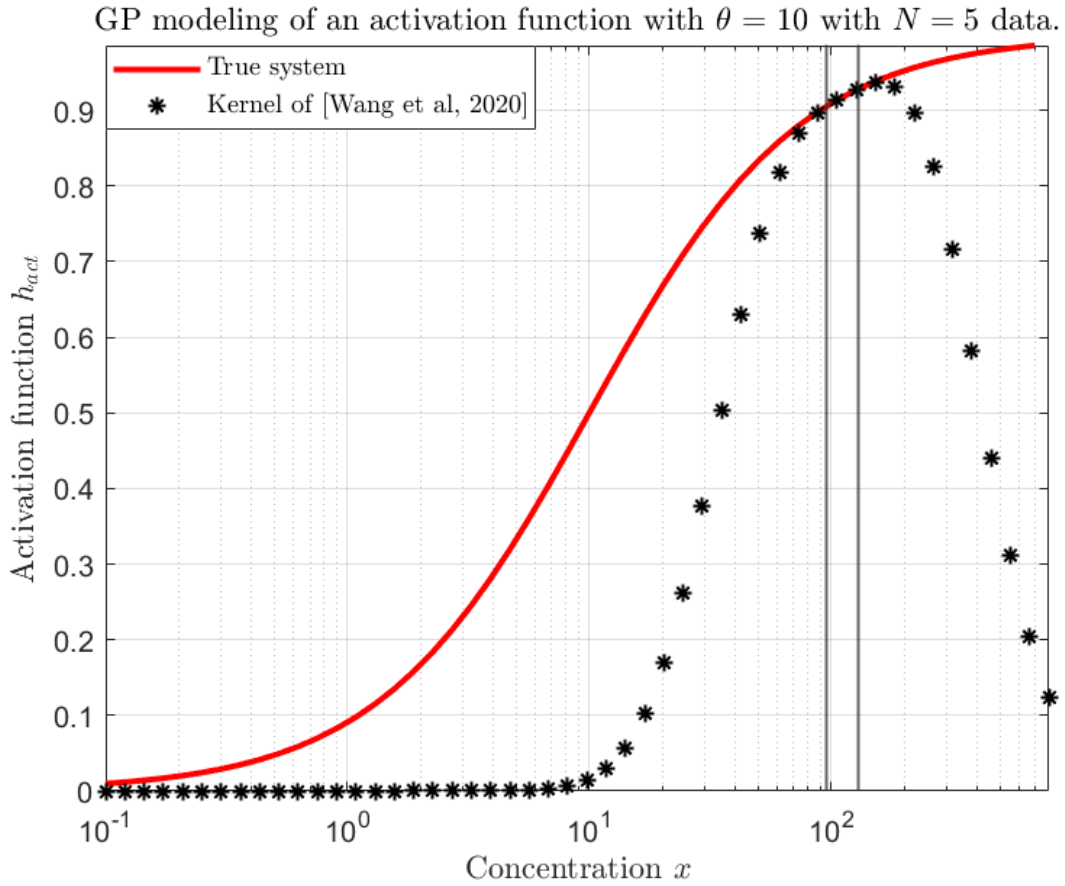
Figure: Some examples of posterior means[6]

[6]Wang, M., Risuleo, R. S., Jacobsen, E. W., Chotteau, V., Hjalmarsson, H. (2020). Identification of nonlinear kinetics of macroscopic bio-reactions using multilinear Gaussian processes. Computers Chemical Engineering, 133, 106671

Kernel in the literature

• However, it can yield inaccurate estimates for small data set and/or poorly distributed in the input space ($\approx$ bioreactor data).

• Example with $y = x/(x+10)$ with $N = 5$ noiseless data randomly chosen in the interval $[80, 130]$



Research problem

- What to do?

- Idea 1: experiment design (but it is costly!).

- Idea 2: incorporate priors in the kernel design.

Research problem

Design better-tailored kernel function for the modeling of the Monod functions h_j.

- Idea: incorporate the **structure** of the Monod functions in the design.

Design of a better kernel: activation

- Consider again a simple activation function for the true system

$$y(x(t)) = h_{act}(x(t), \theta) + \epsilon$$

where $h_{act}(x(t)) = x(t)/(x(t) + \theta)$.

- The covariance between two output data $y(x(t))$ and $y(x(t'))$ is equal to

$$E[y(x(t))y(x(t'))] = h_{act}(x(t), \theta)h_{act}(x(t'), \theta) + E[\epsilon(t)\epsilon(t')]$$

- Ideal kernel function for modeling of h_{act} is then

$$k_{act,ideal}(x(t), x(t'), \theta) = h_{act}(x(t), \theta)h_{act}(x(t'), \theta)$$

- But (i) this kernel is invalid (covariance matrix never positive definite) and (ii) it depends on the unknown θ

Design of a better kernel: activation

- Idea: integrate w.r.t. θ between two bounds θ^- and θ^+!

$$k_{act}(x, x') = \int_{\theta^-}^{\theta^+} h_{act}(x, \theta) h_{act}(x', \theta) d\theta$$

$$= \begin{cases} x^2 \left(\frac{1}{\theta^- + x} - \frac{1}{\theta^+ + x} \right) & \text{if } x = x' \\ \frac{xx'}{x - x'} \log \left(\frac{\theta^+ + x'}{\theta^+ + x} . \frac{\theta^- + x}{\theta^- + x'} \right) & \text{elsewhere} \end{cases}$$

- We obtain a valid kernel constructed directly from activation function (prior added in the design)!

- But, two additional hyperparameters (θ^- and θ^+)!

Step 2: kinetic modeling

- How to choose θ^- and θ^+?

- If no prior on kinetic parameters θ, we choose "wide bounds" but still biologically realistic (e.g., $\theta^- = 0.01$ and $\theta^+ = 100$).

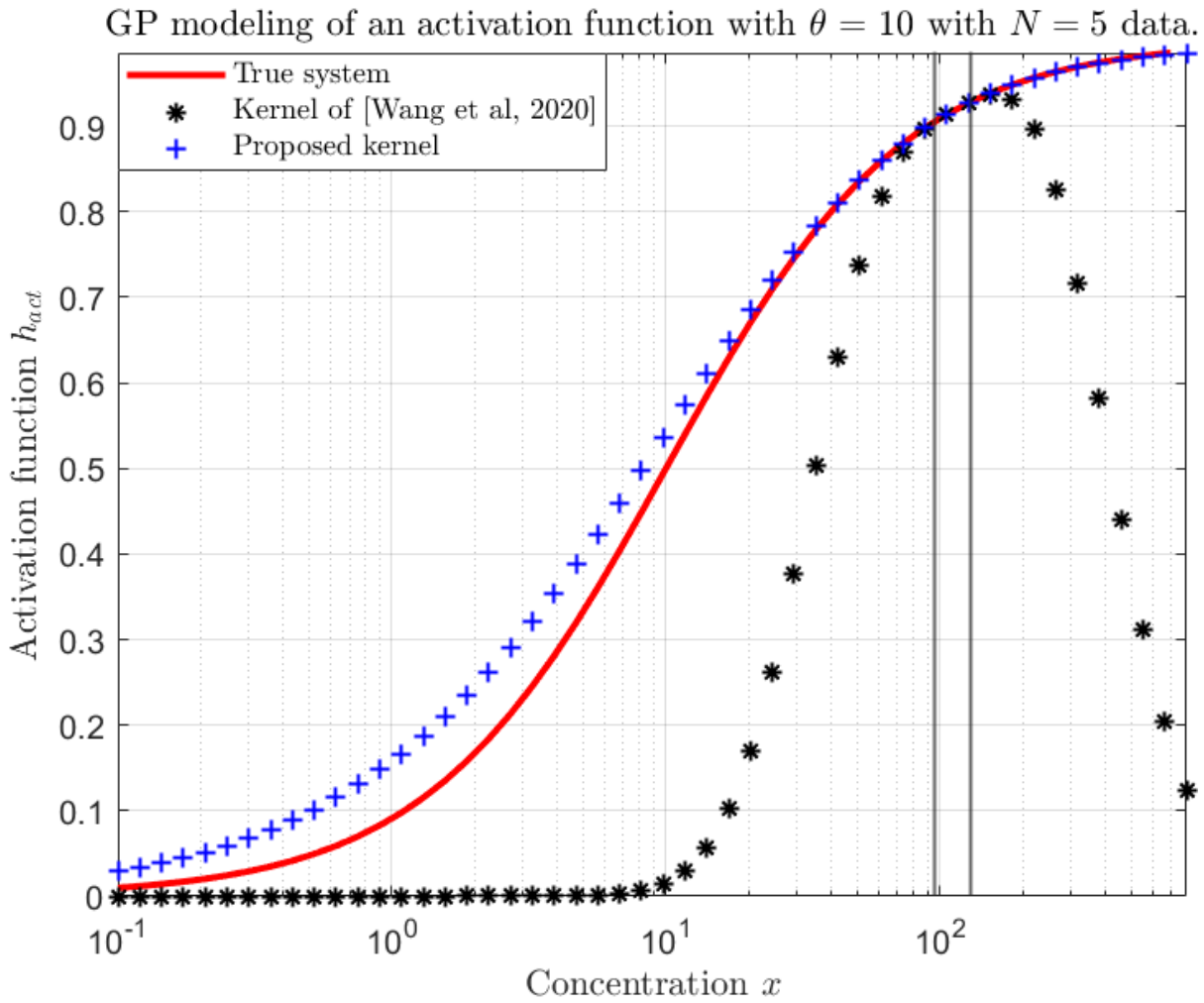
- However, if we have priors on θ in the form of uncertainty intervals

$$\theta \in [\Theta^-, \Theta^+]$$

then we choose $\theta^- = \Theta^-$ and $\theta^+ = \Theta^+$.

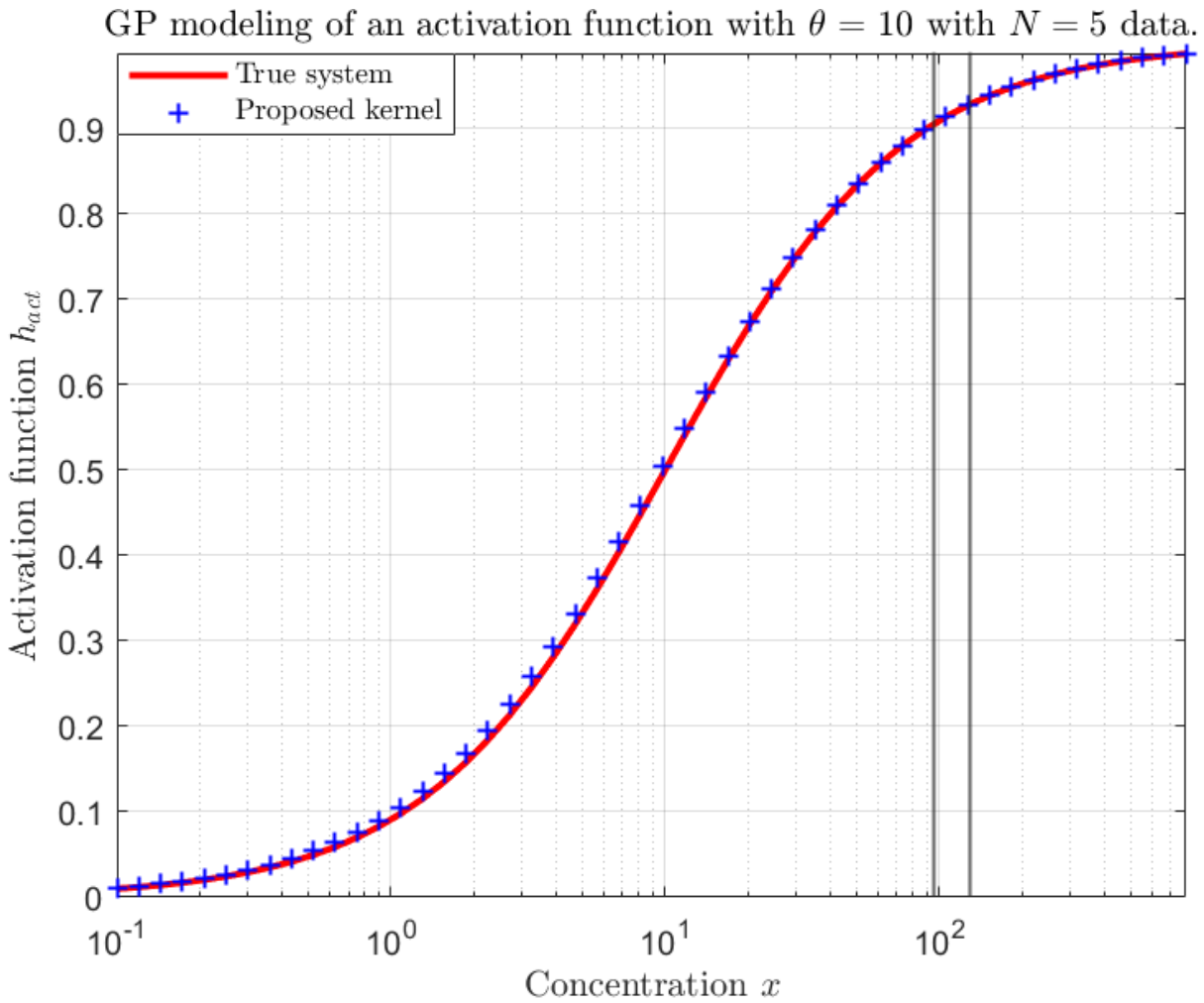
Design of a better kernel: activation

- Back to the example with bounds $\theta^- = 0.01$ and $\theta^+ = 100$.



Design of a better kernel: activation

- Assume that we know that $\theta \in [5, 20]$. Chosen bounds: $\theta^- = 5$, $\theta^+ = 20$.



Design of a better kernel: inhibition

- Inhibition function

$$h_{inh}(x) = 1/(\mu x + 1)$$

- We can similarly construct a kernel with integration:

$$k_{inh}(x, x') = \int_{\mu^-}^{\mu^+} h_{inh}(x, \mu) h_{inh}(x', \mu) d\mu$$

$$= \begin{cases} \dfrac{1}{x}\left(\dfrac{1}{\mu^- x + 1} - \dfrac{1}{\mu^+ x + 1}\right) & \text{if } x = x' \\ \dfrac{1}{x - x'} \log\left(\dfrac{\mu^+ x + 1}{\mu^+ x' + 1} . \dfrac{\mu^- x + 1}{\mu^- x' + 1}\right) & \text{elsewhere} \end{cases}$$

Design of a better kernel: double-component

- Double-component

$$h_{dc}(x) = h_{act}(x, \theta) h_{inh}(x, \mu)$$

- New kernel:

$$k_{dc}(x, x') = k_{act}(x, x') k_{inh}(x, x')$$

Design of a better kernel: neutral effect

- Neutral effect

$$h_{ne}(x) = 1$$

- We should consider a mean function and a zero covariance.

- Instead of this, we approximate for neutral function by

$$h_{ne}(x) = 1 + \varepsilon x$$

with $\varepsilon << 1$ (10^{-3}).

- Then, $k_{ne}(x, x') = (1 + \varepsilon x)(1 + \varepsilon x')$.

Final kernel

• We have a kernel for each type of kinetics. For each modulation function, we have to select the best kernel.

• BUT, we do not know the type of kinetic beforehand.

• Final kernel:

$$k_j(x_j, x_j') = \beta_{act,j} k_{act}(x_j, x_j') + \beta_{inh,j} k_{inh}(x_j, x_j') + \beta_{dc,j} k_{dc}(x_j, x_j') + \beta_{ne,j} k_{ne}(x_j, x_j')$$

• By tuning the hyperparameters $\beta_j = (\beta_{act,j}, \beta_{inh,j}, \beta_{dc,j}, \beta_{ne,j})^T$, we can select different types of kinetics.

• With $\beta_j = (1, 0, 0, 0)^T$, we will consider activation functions for h_j.

Tuning the hyperparameters with Empirical Bayes

Hyperparameter estimation

- With n different metabolites, we have n hyperparameter vectors β_j to estimate.

- We have data

$$\boldsymbol{y} = \begin{pmatrix} y(1) \\ \vdots \\ y(N) \end{pmatrix}$$

- Empirical Bayes: we want to estimate the hyperparameter vector $\beta = (\beta_1^T, \cdots, \beta_n)^T$ such that we maximize the likelihood $p(\boldsymbol{y}|\beta)$.

- BUT

$$y(t) = \prod_{j=1}^{n} h_j(x_j(t)) + \epsilon(t)$$

$$h_j \sim \mathcal{GP}(0, k_j(\beta_j))$$

$\rightarrow$ Likelihood intractable!

Hyperparameter estimation

• How to circumvent the problem?

• Solution: introduce latent variables $\boldsymbol{h}_j = \{h_j(t)\}_{t=1}^N$ for all $j = 1, \cdots, n$ ($\rightarrow$ $N \times n$ latent variables).

• How to estimate? Iterative estimation of the hyperparameters and latent variables as follows

- From $\beta^{(k)}$, we sample the latent variables $\boldsymbol{h}_j$ from the posterior $p(\boldsymbol{h}_1, \cdots, \boldsymbol{h}_n | \boldsymbol{y}, \beta^{(k)}) \Longrightarrow \hat{\boldsymbol{h}}_j^{(k)}$.
- From $\hat{\boldsymbol{h}}_j^{(k)}$, we estimate β which maximizes $p(\boldsymbol{h}^{(k)}|\beta) \Longrightarrow \beta^{(k+1)}$
- $k \rightarrow k+1$

• Expectation maximization + sampling (heuristic convergence to global optimum for β).

Hyperparameter estimation

- How to sample from joint posterior $p(\boldsymbol{h}_1, \cdots, \boldsymbol{h}_n | \boldsymbol{y}, \beta^{(k)})$?

- Solution: Gibbs sampling, i.e., iterative sampling of the conditional posterior distributions.

$$\widehat{\boldsymbol{h}}_1 \rightarrow \widehat{\boldsymbol{h}}_2 \rightarrow \cdots \rightarrow \widehat{\boldsymbol{h}}_n \rightarrow \widehat{\boldsymbol{h}}_1 \rightarrow \cdots$$

Hyperparameter estimation

- Hyperparameter optimization?

- Solve

$$\beta_j^* = \underset{\beta_j}{\arg\max} \; -\log\det(K(\beta_j)) - \hat{\boldsymbol{h}}_j^T K^{-1}(\beta_j)\hat{\boldsymbol{h}}_j \tag{1}$$

where

$$K(\beta_j) = \beta_{act,j}K_{act} + \beta_{inh,j}K_{inh} + \beta_{dc,j}K_{dc} + \beta_{ne,j}K_{ne} \tag{2}$$

- Nonconvex optimization:
  - Bruteforce optimization.
  - Combinatorial optomization.

Numerical example

Numerical example

- Toy example with $n = 6$ metabolites and $\sigma_\epsilon^2 = 10^{-3}$.

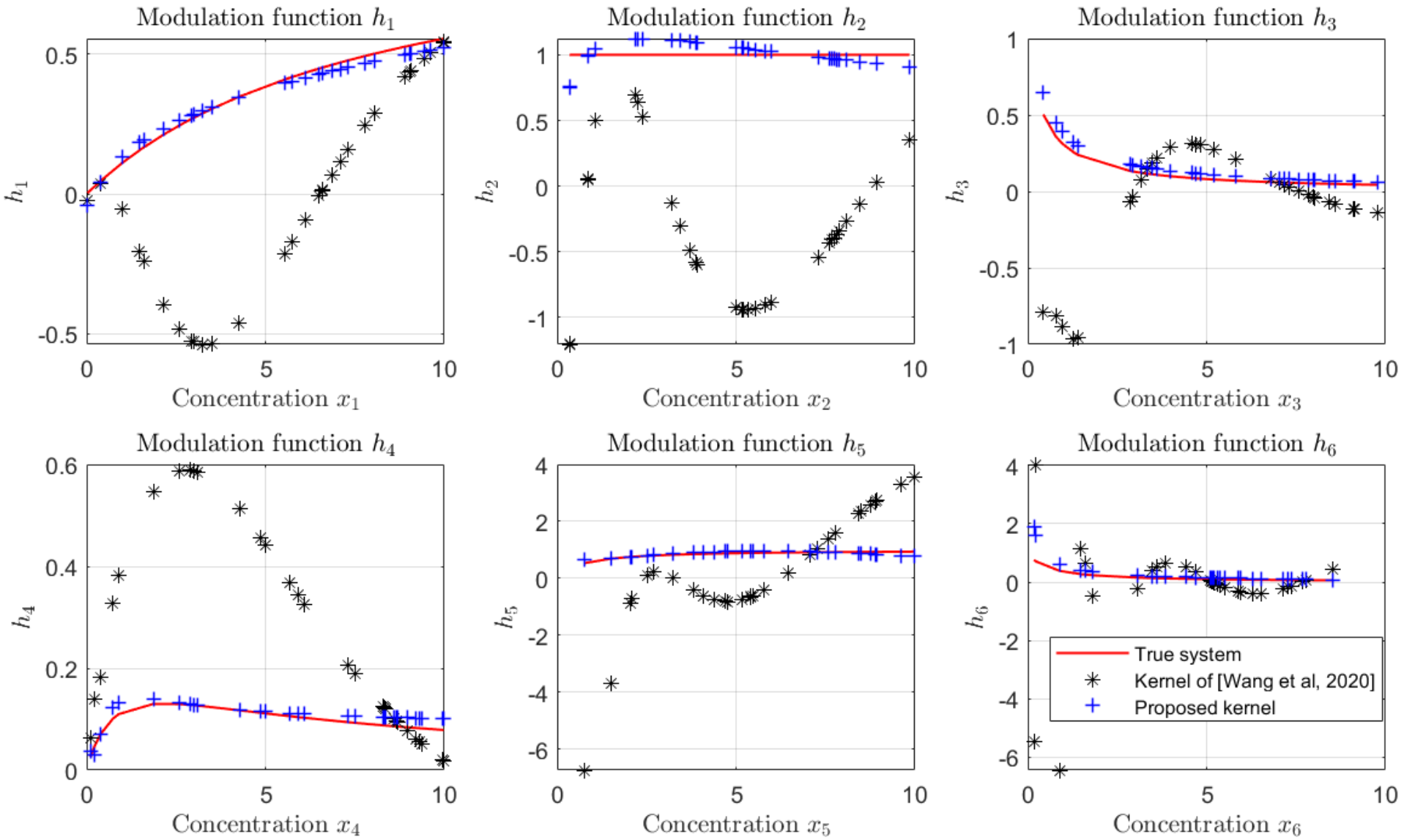
$$y = h_1 \times h_2 \times h_3 \times h_4 \times h_5 \times h_6 + \epsilon$$

Modulation function	Type of kinetic	θ_j	μ_j
h_1	Activation	8.01	-
h_2	Neutral	-	-
h_3	Inhibition	-	2.27
h_4	Double Component	6.81	0.82
h_5	Activation	0.67	–
h_6	Inhibition	–	1.8

Table: True type of kinetics and parameters for the 6 modulation functions h_i.

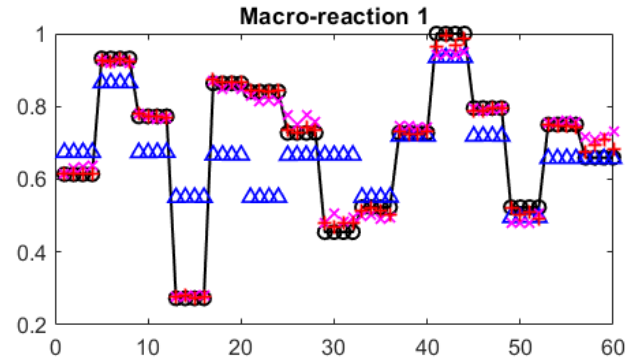
- $N = 30$ concentration data chosen randomly in interval $[0, 10]$.

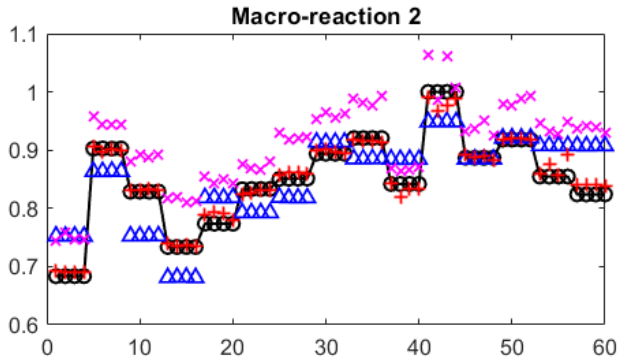
Numerical example

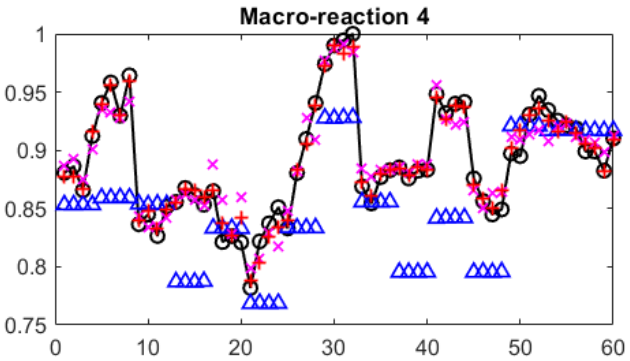


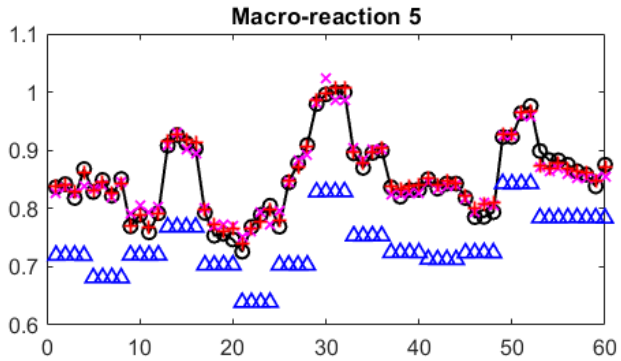
Real-life data with 4 metabolites in kinetic modeling

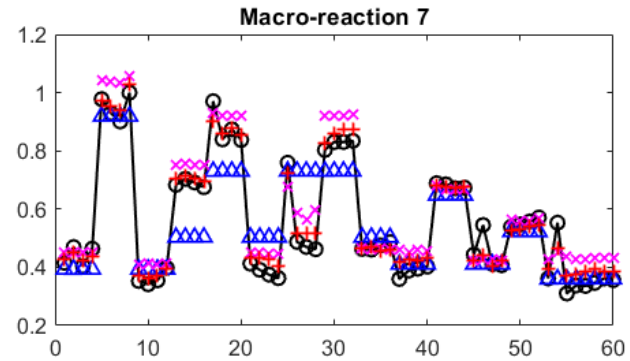
- Black: data, Blue: kernel from literature, Red and magenta: new kernel

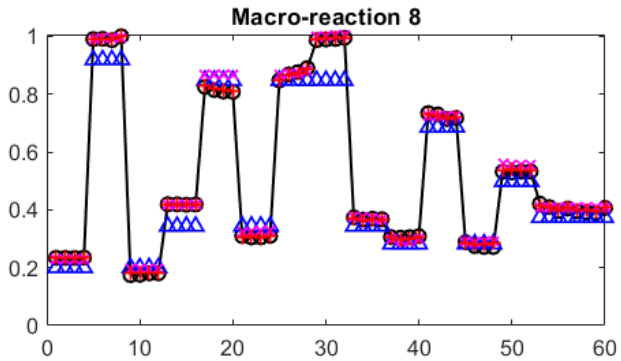












Conclusion and possible extension

Conclusion

- Design of a better-tailored kernel for the modeling of Monod functions.

- Better results than kernel in the literature.

- Main idea: incorporate the structure of the functions to be modeled in the kernel design.

Future work and possible extension

- Test on real-life data.

- First possible generalization of the study

$$\int_{\theta^-}^{\theta^+} h_{act}(x,\theta)h_{act}(x',\theta)d\theta \rightarrow \int_{-\infty}^{\infty} h_{act}(x,\theta)h_{act}(x',\theta)p(\theta)d\theta$$

Future work and possible extension

• Second possible extension: assume that we have a static map to be modeled of this form

$$y(x) = \sum_{j=1}^{n} \phi_j(x, \theta_j) + \epsilon$$

where $\phi_j(x, \theta_j)$ are parametrized nonlinear basis functions (structure known, parameter θ_j unknown).

• Nonlinear least-square optimization $\rightarrow$ local minimum.

• Idea: GP regression with a kernel equal to a linear combination of kernels designed as

$$k_j(x, x') = \int_{-\infty}^{\infty} \phi_j(x, \theta_j)\phi_j(x', \theta_j)p(\theta_j)d\theta_j$$

Thank you for your attention!