Projet CEMRACS 2021
Metamodeling for metabolic models: application to a PDE model of Salmonella infection

Simon Labarthe[1,2], Clémence Frioux[1], and David Sherman[1]

[1]Equipe Pléiade, Inria/INRAE, Inria Bordeaux-Sud-Ouest, 33400 Talence
[2]INRAE, Univ. Bordeaux, BIOGECO, F-33610 Cestas

April 2021

1 Introduction : modeling in microbial ecology

Microbial ecology focuses on the study of microbial communities, called a microbiota, interacting with their environment, regulated by the microbiota host [9, 2]. A well known example of such a symbiotic ecosystem is given by the gut microbiota, a community of hundreds of bacterial species living in the large intestine lumen and regulated by the epithelial cells of the colon of the host. The main drivers of the dynamics are the metabolism of each microbial specie, the interactions between micro-organisms and their spatio-temporal interactions with the host. Mathematical and numerical models of this system aim at providing mechanistic interpretations of biological observations, predicting the evolution of these ecosystems (for example in pathological situations) or defining controlling actions to lead them towards a targeted state.

A lot of tools have been developed to model the bacterial metabolism. This project aims at coupling a metabolic model describing the microbial metabolism to a PDE description of the gut environment in order to represent the infection of an enteric pathogen : *Salmonella enterica* Typhimurium. The metabolic model will be the source function of PDE population dynamics equations modeling the microbial density and the nutritional environment. Classical metabolic models being based on an optimization problem, numerical issues arise when solving the PDE : an optimization problem must be solved at each time and space step, leading to intractable computations. We then want to substitute the optimization problem by an approximate model, built with a RKHS metamodeling method. The RKHS metamodel is a machine learning approach : an approximation of the model image is built from the model evaluation in a sample of the definition space (i.e. a learning database). This metamodel will be

used to predict the model response for new points outside the learning database, with a faster computation than the original optimization problem, allowing for PDE computation.

The work program of this project will consist in :

1. Implement and explore numerically an ODE model of Salmonella Thyphimurium infection of an healthy gut. This ODE system will include a complex metabolic model of the different bacteria based on optimizations. The aim is twofold : a) understanding the biological context and the infection dynamics by studying this toy model ; b) building a gold standard for further assessment of the RKHS metamodeling approximation.
2. Develop a strategy of implementation of the RKHS learning database. The main difficulty is to provide a sound sampling of the definition space, by placing a limited budget of N points in areas of interest for the ODE dynamics.
3. implement a python implementation of the RKHS metamodel computation. Namely, it consists in resolving a high-dimensional optimization problem with appropriate regularization criteria.
4. Assess the RKHS metamodel accuracy.
5. Implement a spatialized version of the ODE model with a PDE model of the gut environment.

Existing tools will be provided for the main parts of the project. The main implementation effort will be put on step 3).

We now provide further informations on the modeling frameworks, the RKHS metamodeling techniques and the biological context. In Section 2, the Flux Balance Analysis (FBA) metabolic modeling framework is introduced. In section 3, the FBA metabolic model is coupled with a population dynamics model of microbial population. The FBA approximation with a metamodel is then introduced in Section 4. The biological context of Salmonella colonization will be introduced. Finally, the work program is presented in a more detailled version in Section 7.

2 FBA model of microbial metabolism

2.1 Microbial metabolism described as an optimization problem

A classical modelling framework to represent the microbial metabolism is Flux Balance Analysis (FBA) [6, 8]. These models are built upon the microorganism genome : the genes are annotated to identify the enzymes they code for, and the biochemical reactions they are involved in. Then, the whole set of reactions embedded in the genome are combined in a genome-scale metabolic network that connects the substrate metabolites the microorganism is able to

metabolize to the synthesized biomass and end-product metabolites produced by the microbe.

Namely, if we note $(m_i)_{0\leq i<N_m}$ the set of the N_m metabolites that can be found in a micro-organism, and $(r_j)_{0\leq j<N_r}$ the set of the N_r reactions coded in the genome, then mass conservation equations can be written on the internal concentration of the metabolites :

$$\partial_t[m_i] = \sum_{l\in R(m_i)} \theta_{m_i,l}\nu_l \tag{1}$$

In this equation, $R(m_i)$ is the subset of reactions involving the metabolite m_i, $\theta_{m_i,l}$ is the stoechiometric coefficient of the metabolite m_i in the reaction l (negative for consumption reaction, and positive for production reaction) and ν_l is the reaction flux, i.e. the quantity of metabolite involved in the reaction by time and microbial biomass units (the flux unit is $\mu mol.t^{-1}.g^{-1}$, and the sign of the flux depends on the side of the reaction).

In the FBA models, an additional fictitious biochemical reaction is considered : the biomass reaction r_b, which is the key of the FBA models. This reaction connects the biomass precursors to the biomass b with the chemical equation

$$\sum_{i\in R(b)} \theta_{m_i,r_b} m_i \to b$$

where m_i are all the constituents of the biomass, i.e. the metabolites needed by the microorganism for growth (to duplicate the genomic material, the metabolism machinery, the cellular membrane, etc...). The biomass reaction flux ν_b is then the amount of microbial biomass produced by time and biomass unit, with unit ($g.t^{-1}.g^{-1}$ by convention, or t^{-1}).

The FBA models aims at predicting this growth rate ν_b while observing biological constraints such as the mass conservation equations (1). To achieve this prediction, the FBA framework makes important assumptions :

— **Steady-state assumption.** All internal metabolites are assumed to be at steady-state in the cell, so that the mass conservation equation (1) reduces to

$$\sum_{l\in R(m_i)} \theta_{m_i,l}\nu_l = 0,$$

resulting in a linear system on the flux vector $\nu := (\nu_l)_{0\leq l<N_r}$

$$A \cdot \nu = 0 \tag{2}$$

where A is the reaction matrix, i.e. the matrix of dimension $N_m \times N_r$ with $A_{ij} := \theta_{m_i,l}$ the stoechiometric coefficient of metabolite i in the reaction j, gathering the whole set of conservation equations for the metabolites and reactions involved in the metabolic network.

— **Biomass maximization.** The microbes are assumed to be instantaneously maximizing the biomass production in a given nutritional context.

— **Flux constraints.** Every flux are constrained by intrinsic limits, related for example to metabolite transporter capacities, or known enzymatic efficiency. This limits are noted c_{min} and c_{max} so that

$$c_{min} \leq \nu \leq c_{max}$$

Hence, the biomass production and all the metabolic fluxes in the microbial machinery can be predicted with the constrained optimization FBA problem

$$\text{find } \nu^* \in \mathbb{R}^{N_r}, \text{such that } \nu^* := \begin{array}{c} \arg\max \\ \nu \in \mathbb{R}^{N_r} \\ A \cdot \nu = 0 \\ c_{min} \leq \nu \leq c_{max} \end{array} \quad \nu_b \tag{3}$$

This problem searches for the optimal growth rate obtained by the system under mass-balance and flux constraints. Mathematically speaking, this optimization problem belongs to the class of linear programming problems : very efficient solvers exist for such a problem, even for high dimensional problems like this one, where N_r is classically around several thousands. A classical FBA toolbox is the cobra toolbox (in Matlab environment) or its python equivalent cobrapy [3]. Problems arise for PDE computations when this optimization problem must be solved millions of times for time and space integration.

2.2 Nutritional environment described as constraints on uptake fluxes

Important FBA model parameters are constraints on substrate uptake from the extra-cellular compartment into the intra-cellular compartment, i.e. the first reactions at the entrance of the metabolic network. This constraints represent the possible uptake for the microorganism, hence representing a proxy of the microbe nutritional environment, i.e. the available nutrients for the microbial species.

The uptake reactions are exchange reactions, i.e. reactions at the interface between the intra and extracellular media, with negative fluxes. Indeed, by construction, exchange reactions are reactions

$$m_i \longrightarrow \boldsymbol{m}_i$$

between the extracellular pool $\boldsymbol{m}_i$, i.e. the nutritional environment, and the intracellular pool m_i of the corresponding metabolite.

If we note $(c^{(up)}, 0)$ the constraints on the uptake fluxes of the N^{up} metabolites in the extra-cellular environment, we get a mapping $\mathcal{F}_{FBA}$ between $c^{(up)}$ and the FBA solution

$$\begin{array}{rcl} \mathcal{F}_{FBA}: & \mathbb{R}^{N^{up}} & \longrightarrow \mathbb{R}^{N_r} \\ & c^{(up)} & \mapsto \nu^* \end{array}$$

where ν^* is the FBA solution with the constraints $(c^{(up)}, 0)$. This mapping allows to tune the uptake constrains to adapt the FBA prediction to a specific nutritional environment context.

3 FBA model as a source function of a population dynamics model

The population dynamics model will first consider one microbial population, N_{up} substrate metabolites uptaken by the population in the nutritional environment, and N_{rel} metabolites produced and released by the microbes in the environment.

For the set of constrains $c^{(up)} := (c_l^{(up)})_{0 \leq l < N_{up}}$ corresponding to the uptake metabolites, a FBA solution can be computed, predicting the uptake, released and microbial growth flux

$$\nu \in \mathbb{R}^{N_{up}+N_{rel}+1}, \qquad \nu = \mathcal{F}_{FBA}(c^{(up)}). \tag{4}$$

The growth flux $\nu_b := \mathcal{F}_{FBA}(c^{(up)})_b$ represents the growth rate of a population dynamics equation on the microbial population

$$\partial_t b = \mathcal{F}_{FBA}(c^{(up)})_b b \tag{5}$$

In the same way, the metabolite uptake fluxes can be used to model the dynamics of extracellular metabolites $\boldsymbol{m}_i^{(up)}$ representing the nutritional environment.

$$\partial_t \boldsymbol{m}_i^{(up)} = \mathcal{F}_{FBA}(c^{(up)})_i b \tag{6}$$

Additional equations can be added on the N_{rel} released metabolites $\boldsymbol{m}_i^{(rel)}$ that represent metabolites that are produced by the microorganism as a end-product of the metabolic network, and available for the host (or other organisms)

$$\partial_t \boldsymbol{m}_j^{(rel)} = \mathcal{F}_{FBA}(c^{(up)})_j b \tag{7}$$

Conversely, the extracellular concentration of substrate can be used to tailor constraints on the substrate metabolite uptake in the FBA model

$$c_i^{(up)} = -\lambda_i \frac{\boldsymbol{m}_i^{(up)}}{K_i + \boldsymbol{m}_i^{(up)}}, \quad \text{for } 0 \leq i < N_{up} \tag{8}$$

Equation (8) prescribes a Michaelis-Menten dynamics on the uptake capacity of the metabolite.

The dynamic system (4)-(8) is termed dynamic FBA (or d-FBA). The population dynamic model is solved with an integration scheme (for example a semi-implicit Euler scheme), so that a FBA model is solved at each time step.

The same approach can be used to build a spatialized population dynamic model

$$\partial_t b + \mathcal{T}_b(b) = \mathcal{F}_{FBA}(c^{(up)})_b b \tag{9}$$

$$\partial_t \boldsymbol{m}_i + \mathcal{T}_m(\boldsymbol{m}) = \mathcal{F}_{FBA}(c^{(up)})_i b \tag{10}$$

$$c_i = -\lambda_i \frac{\boldsymbol{m}_i}{K_i + \boldsymbol{m}_i} \tag{11}$$

where $\mathcal{T}_b$ and $\mathcal{T}_m$ are transport terms modeling for example diffusion, active motility or transport processes :

$$\begin{aligned} \mathcal{T}_b(b) &= div(\sigma_b \nabla b) + div(b u_b(b,m)) + div(b\chi(m)) \\ \mathcal{T}_m(m) &= div(\sigma_m \nabla m) + div(m u_m(b,m)) \\ u_b := \mathcal{M}_b(b,m), &\qquad u_m := \mathcal{M}_m(b,m) \end{aligned}$$

where $\mathcal{M}_b$ and $\mathcal{M}_m$ are fluid dynamics models, and $\chi(m)$ is a chemotactic speed field [5].

In this case, a FBA optimization must be performed by time step, but also by spatial mesh cell, leading to a numerical bottle-neck : too many optimizations are needed, forbidding PDE resolution in a reasonable time.

4 RKHS metamodel of a FBA metabolic model

Given N points $(c_k^{(up)})_{0\leq k<N}$ in the input subdomain $\mathbb{R}^{-N_{up}}$ forming the learning database, we build $N_{up} + N_{rel} + 1$ vectors $Y_{FBA,i}$ by evaluating the FBA model at the N points of the database

$$Y_{FBA,i} := \mathcal{F}_{FBA}(c_k^{(up)})_i, \quad 0 \leq k < N, \quad 0 \leq i \leq N_{up} + N_{rel} + 1$$

The $N_{up} + N_{rel} + 1$ correspond to the uptake, release and biomass fluxes computed by the FBA model.

RKHS metamodeling consists in defining a reproducing kernel $\mathcal{K}(x_1, x_2)$, for $x_1, x_2 \in \mathbb{R}^2$, from which we can define a functional family $f_{\nu, c_j^{(up)}}(x)$ indexed by $(c_j^{(up)})_{0\leq j<N}$, and $0 \leq \nu < N_\nu$ where $N_\nu = \sum_{p=1}^{N_m} \binom{N_{up}}{p}$. In other words, ν indexes the N_ν parts of $\chi := \{c_{j,1}^{(up)}, \cdots, c_{j,N_{up}}^{(up)}\}$, the set of input variables, containing up to N_m elements. We then consider interactions between input variables up to order N_m. By abuse of notation, we can also design by ν the part of χ that it indexes. The family $f_{\nu, c_j^{(up)}}(x)$ have analytical expression that reads

$$\begin{aligned} &\mathbb{R} \longrightarrow \mathbb{R} \\ &x \mapsto f_{\nu, c_j^{(up)}}(x) := \prod_{c_{j,k}^{(up)} \in \nu} \mathcal{K}(c_{j,k}^{(up)}, x) \end{aligned}$$

The method consists in approximating the vectors $Y_{FBA,i}$ in the functional space spanned by the functional family $f_{\nu,c_j^{(up)}}(x)$. For $i \in 0, \cdots, N_{up}+N_{rel}+1$, we then aim at finding $(\hat{\theta}_{0,i}, \hat{\theta}_{\nu,i})$ such that

$$(\hat{\theta}_{0,i}, \hat{\theta}_{nu,i}) := \underset{\substack{\theta_{0,i} \in \mathbb{R} \\ \theta_{\nu,i} \in \mathbb{R}^{N_\nu * N}}}{\arg\min} \|Y_{FBA,i} - (\theta_{0,i} + \sum_{\nu=1}^{N_\nu} K_\nu \theta_{\nu,i})\|_2^2 + \mathcal{G}(K_\nu, \theta_{\nu,i}) \tag{12}$$

where the matrices K_ν are N dimensional Gram matrices

$$K_\nu := \left(f_{\nu,c_{j_1}^{(up)}}(c_{j_2}^{(up)}) \right)_{0 \leq j_1, j_2 < N}.$$

The matrix K_ν is then the evaluation of the functions $f_{\nu,c_j^{(up)}}$ in the points of the learning database, so that $K_\nu \theta_{\nu,i}$ is the evaluation of the function

$$c \mapsto \sum_{j=1}^{N} \theta_{\nu,i_j} f_{\nu,c_j^{(up)}}(c)$$

in the N points of the database.

The operator $\mathcal{G}$ is a regularization operator. In practive, we will use the RKHS ridge group sparse regularization criteria defined in [4] :

$$\mathcal{G}(K_\nu, \theta_{\nu,i}) := \gamma \sqrt{n} \sum_{\nu=1}^{N_\nu} \|K_\nu \theta_{\nu,i}\| + n\mu \sum_{\nu=1}^{N_\nu} \|K_\nu^{1/2} \theta_{\nu,i}\| \tag{13}$$

where γ and μ are hyper-parameters to be estimated with suitable *ad-hoc* methodology (for example cross-validation). In this criteria, the first term $\sqrt{n} \sum_{\nu=1}^{N_\nu} \|K_\nu \theta_{\nu,i}\| = n \sum_{\nu=1}^{N_\nu} \frac{1}{\sqrt{n}} \|K_\nu \theta_{\nu,i}\|$ is a composite term seeking for regularity (by minimizing the L-2 norms of the functions $c \mapsto \sum_{j=1}^{N} \theta_{\nu,i_j} f_{\nu,c_j^{(up)}}(c)$) together with selecting the most regularizing K_ν matrix. The second term is a RKHS group lasso term selecting the functions $c \mapsto \sum_{j=1}^{N} \theta_{\nu,i_j} f_{\nu,c_j^{(up)}}(c)$ in the RKHS norm.

This estimation problem is numerically expensive, but it can be done off-line once for all. Then, the function $\mathcal{F}_{FBAi}$ can be estimated in a new point $\tilde{c}^{(up)}$ in the input parameter space with the formula

$$\mathcal{F}_{FBA_i}(\tilde{c}^{(up)}) \simeq \hat{\mathcal{F}}_{FBA_i}(\tilde{c}^{(up)}) := \hat{\theta}_{0,i} + \sum_{\nu=1}^{N_\nu} F_\nu(\tilde{c}^{(up)}) \cdot \hat{\theta}_{\nu,i}) \tag{14}$$

where $F_\nu(\tilde{c}^{(up)})$ is the N dimensional vector

$$F_\nu(\tilde{c}^{(up)}) := \left(f_{\nu,c_{j_1}^{(up)}}(\tilde{c}^{(up)}) \right)_{0 \leq j_1 < N}$$

i.e., the evaluation of the functions of the function family at the new point $\tilde{c}^{(up)}$.

This analytical formula is much faster to compute than the original optimization problem.

5 Biological context of Salmonella infection

This project will focus on the colonization of the gut microbiota by an enteric pathogen : Salmonella Thyphimurium, which uses a very complex mechanism to invade the gut.

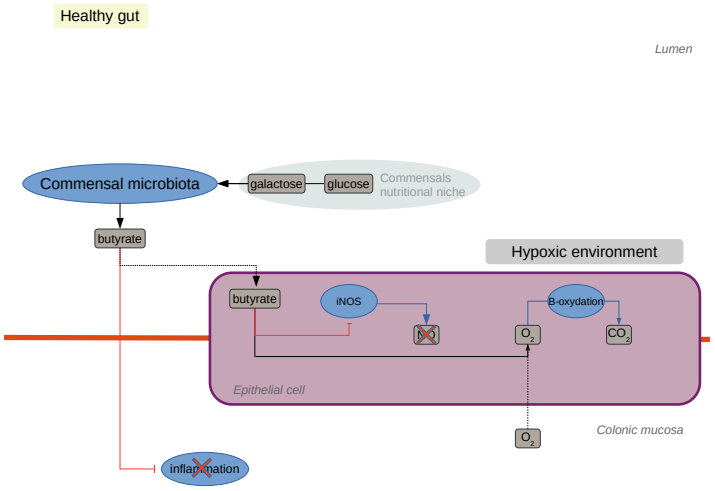
Healthy gut The environment of an healthy gut is anaerobic : the commensal microbiota are then specialized microbes relying on anaerobic metabolism to grow without oxygen. Actually, a main part of the gut microbiota are strictly anaerobic, meaning that oxygen is harmfull to them. With this anaerobic metabolism, the commensal microbiota consumes fibre-derived sugars (e.g.. glucose and galactose) and produces short-chain fatty acids (SCFA) – mainly butyrate, acetate and propionate – that are absorbed by the host for its own metabolism : butyrate being the main energetic source of the intestinal cells, that are respired with the oxygen brought to the intestine by the blood system. A virtuous cycle is then set up (see Figure 1a) : the commensal microbiota produces butyrate that is metabolized by the host with oxygen ; consequently, this oxygen does not diffuse to the lumen ensuring hypoxia and a favorable habitat for the butyrate-producing anaerobes. Salmonella is not very efficient in an anaerobic environment : the pathogen will have to hack this regulation mechanism, to create a favorable niche and invade the gut. [1, 7]

Colonized gut When arriving in the gut lumen, the pathogen releases a virulence factor (sipA) that triggers an inflammation in the epithelial cells (see Figure 1b). The host cells produce neutrophils, immunitary cells sent into the gut lumen where they trap the bacteria they encounter (pathogenic bacteria but also SCFA-producing symbionts). Then, the production of butyrate drops, and this metabolite is no longer available for the epithelial cells : the oxygen reaching the cells is no longer metabolized and start flowing in the gut lumen. This oxygen will be harmful for the butyrate-producing anaerobes, which initiates a vicious circle. The oxygen will also oxydize nutrients present in the gut, providing very efficient energetic sources for the pathogen alone, allowing it to take over from the commensal bacteria. Namely, galactose, glucose and thiosulfate will be oxydized to galactarate, glucarate end tetrathionate. In the mean time, inflammation induces the production of nitric oxyde, which is oxydated in nitrate, also very favorable for the pathogen.[1, 7]

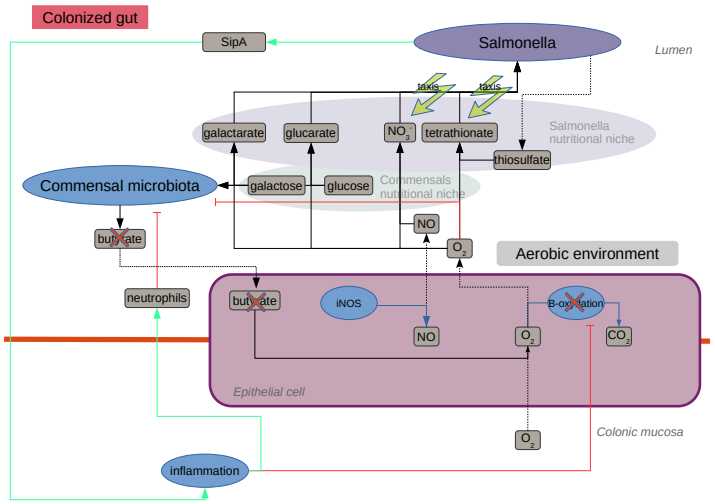
Figures sketching these situations can be found in Fig. (1a-1b).

6 Population dynamics model of Salmonella infection, including host inflammatory response

We will first build a population dynamics model of Salmonella infection. The commensal microbiota will be represented by a unique strain of butyrate-producing bacteria : *Faecalibacterium Prauznitzii*. This bacteria belongs to one



(a) Healthy gut at homeostasis : the colon lumen is hypoxic, so that commensal microbiota produces butyrate from sugars, which is consumed by the host with the blood-stream oxygen, regulating anaerobia.



(b) Salmonella colonization process : the pathogen triggers inflammation, decreasing commensals levels. Butyrate production drops down, reducing availability for the host. Epithelial cell metabolism switches from aerobic to anaerobic : blood-stream oxygen is no longer consumed and starts flowing in the gut lumen creating an aerobic niche for the pathogen.

FIGURE 1 – Figure recapitulating the regulation in an healthy gut, and *S. Thyphymurium* colonization mechanisms.

of the dominant genus in the gut microbiota, and is very often used as a probiotic.

The model will be a compartment model : a first compartment will describe the gut lumen while the second will stand for the epithelial cells.

Luminal compartment Noting S_{th} and F_{prau} the *Salmonella Thyphymurium* and *Faecalibacterium Prauznitsii* populations, m a vector gathering all the nutrient concentrations of interest to describe the nutritional environment and n the neutrophils population, we set

$$\partial_t S_{th} = \mathcal{F}_{S_{th}}(c^{(up)}_{S_{th}})S_{th} - \rho n \tag{15}$$

$$\partial_t F_{prau} = \mathcal{F}_{F_{prau}}(c^{(up)}_{F_{prau}})F_{prau} - \rho n - \alpha \frac{o_2}{K_o + o_2} \tag{16}$$

$$\tag{17}$$

where $\mathcal{F}_{S_{th}}$ (resp. $\mathcal{F}_{F_{prau}}$) is the FBA metabolic model of the pathogen (resp. the commensal). The parameter ρ represents the trapping by the neutrophils n. The term $\alpha \frac{o_2}{K_o + o_2}$ models the deleterious effect of the oxygen level o_2 on the obligate anaerobe F_{prau}, with a Michaelis-Menten dynamics and a tuning parameter α.

The vector m is indexed by $i \in \{Gal, Gluc, NO, GalO, GlucO, NO_3, thio, tet, o_2, but\}$ standing for, respectively, galactose, glucose, nitric oxyde, galactarate (i.e. oxydized galactose), glucarate (i.e. oxydized glucose), nitrate, thiosulfate, tetrationate (i.e. oxydized thiosulfate), oxygen and butyrate. We note $S = \{Gal, Gluc\}$ the set of fibre-derived sugars, which is the nutritional niche of F_{prau} and $O = \{GalO, GlucO, NO_3, tet\}$ the set of oxydized nutrients. The set $O \cup \{thio\}$ is the nutritional niche of S_{th}.

The FBA model constraints, specific to each bacteria, are computed with the equations

$$c^{(up)}_{F_{prau},i} = -\lambda_i \frac{m_i}{K_i + m_i}, \text{ for } i \in S \tag{18}$$

$$c^{(up)}_{S_{th},i} = -\lambda_i \frac{m_i}{K_i + m_i}, \text{ for } i \in O \cup \{thio\}. \tag{19}$$

$$\tag{20}$$

We set

$$\partial_t m_{o_2} = \gamma(e_{o_2} - m_{o_2}) - \sum_{s \in S \cup \{NO, thio\}} \beta_s m_{o_2} m_s \tag{21}$$

$$\partial_t m_{NO} = \gamma(e_{NO} - m_{NO}) - \beta_{NO} m_{o_2} m_{NO} \tag{22}$$

$$\tag{23}$$

In these equations, the first term is a diffusive term describing the flow of nitric oxyde and oxygen from the epithelial cell compartment into the lumen. The second one represents oxydative processes.

For $s \in S$, we have

$$\partial_t m_s = -\mathcal{F}_{F_{prau},s}(c^{(up)}_{F_{prau}})F_{prau} + D_s \tag{24}$$

where $\mathcal{F}_{F_{prau},s}$ is the consumption fluxes predicted by the FBA model, and D_s is a source term representing the dietary influx of sugar.

For thiosulfate, we also have

$$\partial_t m_{thio} = -\mathcal{F}_{S_{th},thio}(c^{(up)}_{S_{th}})S_{th} + D_{thio} \tag{25}$$

where D_{thio} represents a source term for thiosulfate, coming from the detoxyfication of H_2S by the host.

For $s \in O$, we also set

$$\partial_t m_s = -\mathcal{F}_{S_{th},s}(c^{(up)}_{S_{th}})S_{th} + \beta_s m_{o_2} m_s \tag{26}$$

to represent the consumption flux and the oxydative process.

Finally, butyrate population is represented by

$$\partial_t m_{but} = \mathcal{F}_{F_{prau},but}F_{prau} + \gamma(e_{but} - m_{but}) \tag{27}$$

where the first term is the metabolic production of butyrate given by the FBA, and the second one is a diffusive flux representing butyrate absorption by the epithelium.

The population of neutrophils is described by the equation

$$\partial_t n = -\gamma_n(e_n - n) - d_n n \tag{28}$$

where the first term of the right hand side is a diffusive flux of neutrophils through the epithelium, and the second term is a death rate.

Epithelial compartment The epithelial cells populations are gathered in the vector e indexed by $i \in \{NO, o_2, but, n\}$. Their respective dynamics are

$$\partial_t e_{NO} = K^\nu(1 - \frac{e^\nu_{but}}{K^\nu + e^\nu_{but}})e_{NO}(L_{NO} - e_{NO}) - d_{NO}e_{NO} + \gamma(m_{NO} - e_{NO}) \tag{29}$$

where the first term represents nitric oxide production regulated by butyrate, and the second term stands for the natural degradation of NO in the cells. The parameter ν tunes the slope of the logistic functions : when $e_{but} \gg K$, the logistic term tends to 1, so that the production term vanishes, whereas when e_{but} is small, the logistic term vanishes and NO production is pulled towards the maximal concentration L_{NO}. The last term is a diffusive term.

Oxygen and neutrophils dynamics are represented in a similar way :

$$\partial_t e_{O_2} = K^\nu(1 - \frac{e^\nu_{but}}{K^\nu + e^\nu_{but}})e_{O_2}(L_{O_2} - e_{O_2}) - d_{O_2}e_{O_2} + \gamma(m_{O_2} - e_{O_2}) \tag{30}$$

$$\partial_t e_n = K^\nu(1 - \frac{e^\nu_{but}}{K^\nu + e^\nu_{but}})e_n(L_n - e_n) - d_n e_n + \gamma(m_n - e_n) \tag{31}$$

7 Work program of the CEMRACS project

The working program of the CEMRACS will be the following.

7.1 Python implementation of the ODE model of Salmonella infection

An implementation of equations (15) to (26) will be developed. A semi-implicite Euler scheme will be proposed, in order to preserve the positivity of the solutions. The FBA model will be solved with the Cobrapy package.

7.2 Learning database definition

A learning database will be assembled with the following methodology :

1. A set of ODE system solutions will be computed for N_i different initial conditions. For initial conditions $T_i^* := (S_{th,i}^*, F_{prau,i}^*, m_i^*)$, $i = 1, ..., N_i$ at $t = 0$, let us note $(S_{th}, F_{prau,i}, m)(t|T_i^*))$ the ODE solution for $t > 0$.
2. Noting Δt the time step, we will assemble the array $X_{N_i,N_T} := (c_{S_{th}}^{(up)}(n\Delta), c_{F_{prau}}^{(up)}(n\Delta))$ for $n = 1, \cdots, N_T$ the number of time step.
3. The array will be supplemented by little stochastic perturbations around the points X_{N_i,N_T}. Noting

$$\varepsilon \simeq \mathcal{N}(N_u p, N_T * n_s)$$

where n_s is the number of additional stochastic points per time step, we define

$$\tilde{X} := X_{N_i,N_T} + \varepsilon$$

4. The vectors $Y_{FBA,F_{prau},i}$, $Y_{FBA,S_{th},i}$ will be built for each metabolic model by computing the FBA models on $\tilde{X}$.

7.3 Python implementation of RKHS modeling

In order to 1) select the K_ν (i.e. the relevant input variables and input interactions), 2) compute the corresponding $\theta_{\nu,i}$ coefficients, we will implement a python solver of the minimization problem (12) with regularization (13). We will use a stochastic gradient descent method, for example by interfacing the tensorflow stochastic gradient descent solver (SGD function in the "optimizers" part of the "Keras" module of tensor flow –documentation can be found here–).

Then, $(\hat{\theta}_{0,i}, \hat{\theta}_{nu,i})$ will be inferred on the learning database.

This section is the main goal of the CEMRACS project : for the other steps, existing numerical tools will be provided. But this solver will be implemented from scratch.

7.4 Assesment of the RKHS metamodels

We will assess the accuracy of the RKHS metamodel by

— Comparing $\mathcal{F}_{S_{th}}$ and $\hat{\mathcal{F}}_{S_{th}}$ for the points in the learning database, and for testing samples obtained by uniform sampling of the FBA constraint ranges. Same assessment will be conducted on $\hat{\mathcal{F}}_{F_{prau}}$.

— Replacing $\mathcal{F}_{S_th}$ and $\mathcal{F}_{F_{prau}}$ by their respective metamodel $\hat{\mathcal{F}}_{S_{th}}$ and $\hat{\mathcal{F}}_{F_{prau}}$ in the system (15)-(30). Noting $(\hat{S}_{th}, \hat{F}_{prau}, \hat{m})(t|T_i^*)$ the solution obtained with this new system, we will quantify the approximation error :

$$e := \left(\sum_{i=1}^{N_i} \|(\hat{S}_{th} - S_{th})(t|T_i^*)\|_2^2 + \|(\hat{F}_{prau} - F_{prau})(t|T_i^*)\|_2^2 + \|(\hat{m} - m)(t|T_i^*)\|_2^2 \right)^{1/2}$$

In this equation, $\|\cdot\|_2$ denotes the temporal L_2 norm.

7.5 PDE computation

A spatialized version of System (15)-(30) will be developped by changing the source function and the boundary conditions of a previous model of gut microbiota in its intestinal environment [5].

Références

[1] Andreas J Bäumler et Vanessa Sperandio. « Interactions between the microbiota and pathogenic bacteria in the gut ». In : *Nature* 535.7610 (2016), p. 85-93.

[2] Seth R Bordenstein et Kevin R Theis. « Host biology in light of the microbiome : ten principles of holobionts and hologenomes ». In : *PLoS Biol* 13.8 (2015), e1002226.

[3] Ali Ebrahim et al. « COBRApy : constraints-based reconstruction and analysis for python ». In : *BMC systems biology* 7.1 (2013), p. 1-6.

[4] Halaleh Kamari, Sylvie Huet et Marie-Luce Taupin. « RKHSMetaMod : An R package to estimate the Hoeffding decomposition of an unknown function by solving RKHS ridge group sparse optimization problem ». In : *arXiv preprint arXiv :1905.13695* (2019).

[5] Simon Labarthe et al. « A mathematical model to investigate the key drivers of the biogeography of the colon microbiota ». In : *Journal of theoretical biology* 462 (2019), p. 552-581.

[6] Jeffrey D Orth, Ines Thiele et Bernhard Ø Palsson. « What is flux balance analysis ? » In : *Nature Biotechnology* 28.3 (2010), p. 245-248. issn : 1087-0156. doi : 10.1038/nbt.1614.

[7] Fabian Rivera-Chávez et Andreas J Bäumler. « The pyromaniac inside you : Salmonella metabolism in the host gut ». In : *Annual review of microbiology* 69 (2015), p. 31-48.

[8] Jan Schellenberger et al. « Quantitative prediction of cellular metabolism with constraint-based models : the COBRA Toolbox v2. 0 ». In : *Nature protocols* 6.9 (2011), p. 1290.

[9] Jean-Christophe Simon et al. « Host-microbiota interactions : from holobiont theory to analysis ». In : *Microbiome* 7.1 (2019), p. 1-5.