



**HAL**  
open science

# cMFA for multi-omics data integration in microbial community models

Sthyve Junior Tatho Djeanou, Simon Labarthe, Valentina Baldazzi

## ► To cite this version:

Sthyve Junior Tatho Djeanou, Simon Labarthe, Valentina Baldazzi. cMFA for multi-omics data integration in microbial community models. Journées scientifiques 2025 Agroécologie et Numérique, Jan 2025, Dijon, France. hal-04939529

**HAL Id: hal-04939529**

**<https://inria.hal.science/hal-04939529v1>**

Submitted on 11 Feb 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

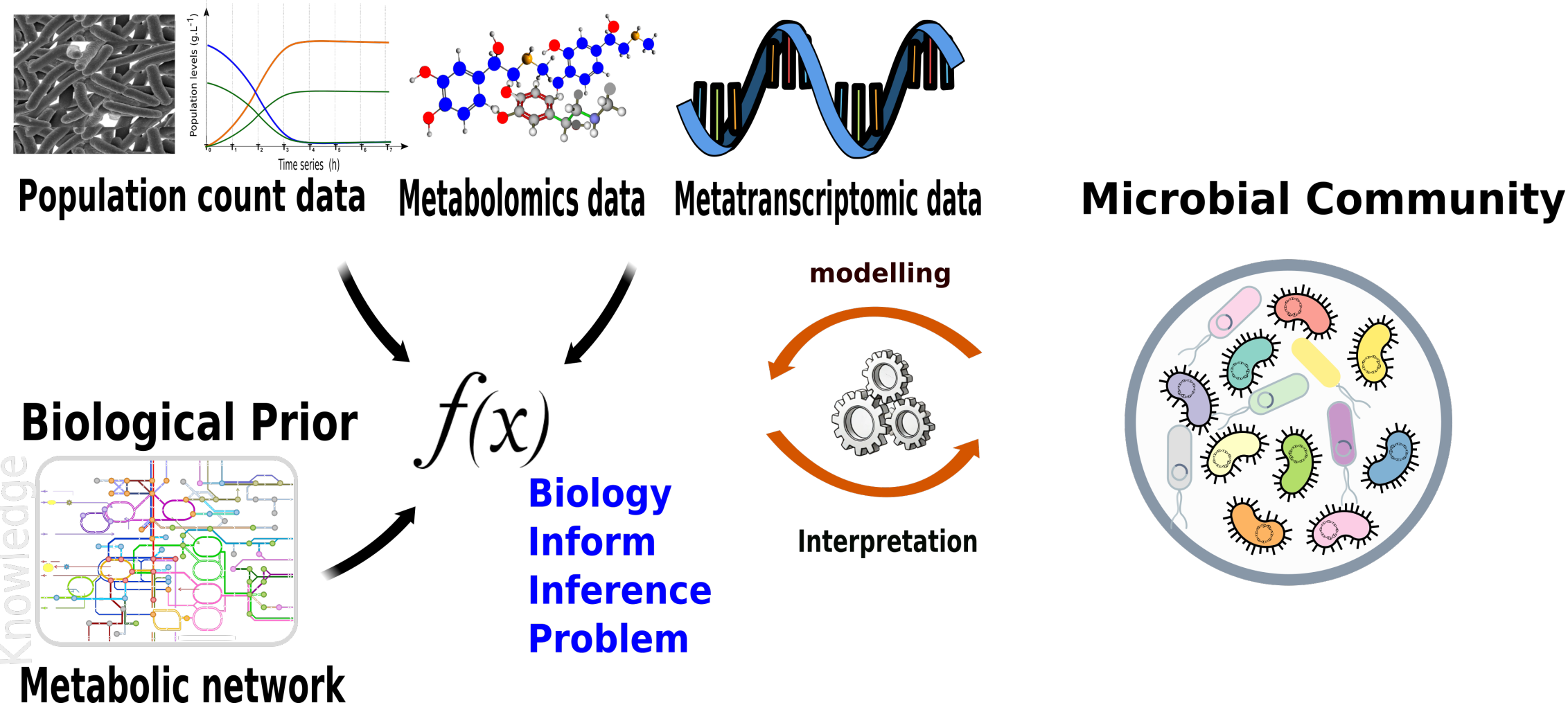


Distributed under a Creative Commons Attribution 4.0 International License



## INTRODUCTION

### Biology-informed inference from multi-omics data



Understanding natural microbial communities presents significant challenges due to the complex interactions and metabolite exchanges between microorganisms. To gain deeper insights into these dynamics, we are focusing on:

- Developing mathematical models based on metabolic networks.
- Inferring metabolic fluxes using biology-informed high-dimensional analysis methods to integrate multi-omics data.

### Scientific Question:

How can mathematical models help us better understand the functioning of microbial communities ?

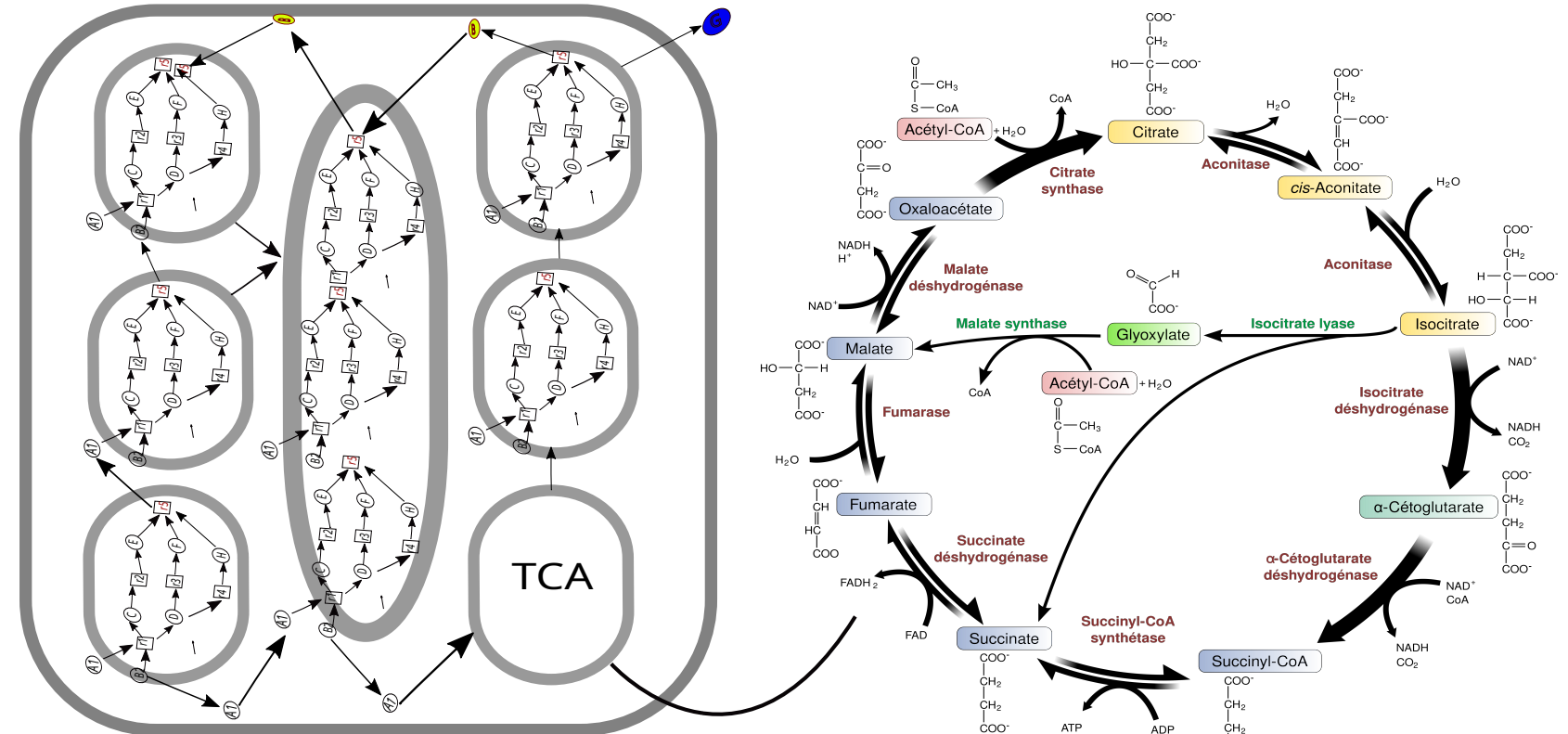
- Internal flux reconstruction based on some external reaction rates.
- Estimating the contribution of each microorganism to the community.

## COMMUNITY METABOLIC FLUX ANALYSIS (cMFA)

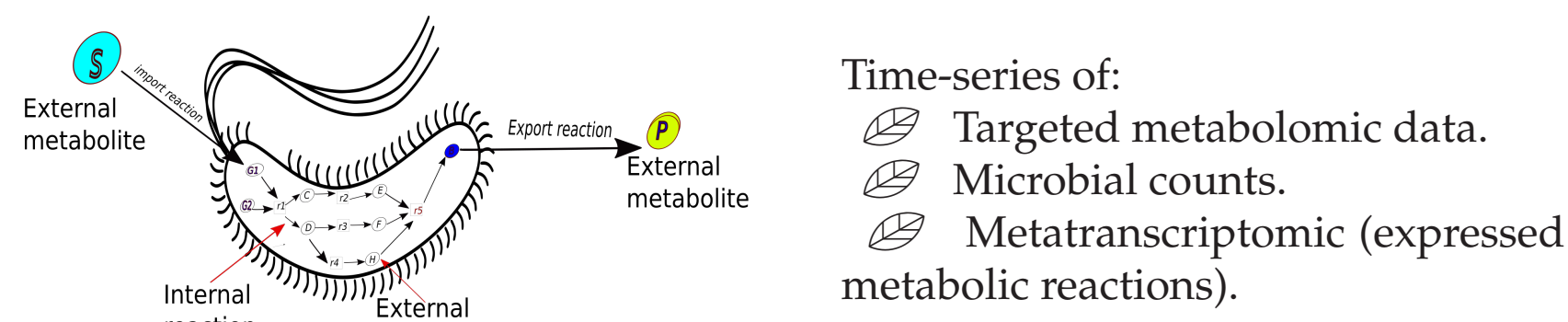
## RESULTS

### METABOLIC NETWORK

A metabolic network refers to the interconnected set of biochemical reactions and pathways that occur within a cell or organism.



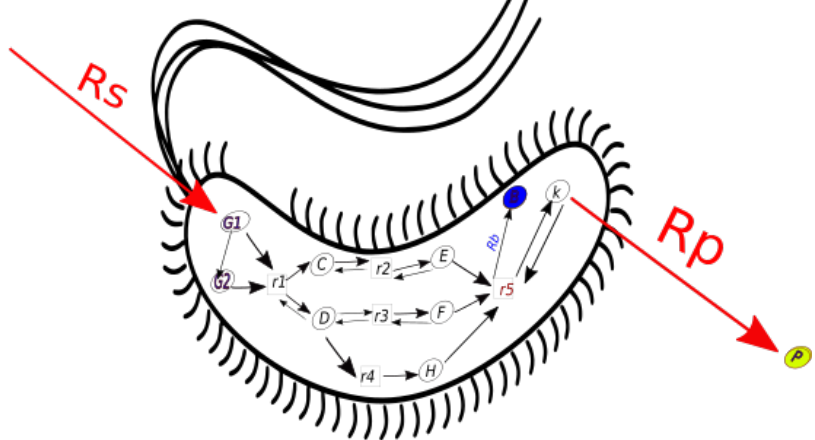
### INPUT DATA DESCRIPTION



Time-series of:  
 • Targeted metabolomic data.  
 • Microbial counts.  
 • Metatranscriptomic (expressed metabolic reactions).

### METABOLIC FLUX ANALYSIS (MFA)

Experimental technique used to estimate intracellular metabolic fluxes by analyzing metabolite production and consumption rates in a biological system, providing insights into the cell's core metabolism.



### COMMUNITY MFA (cMFA)

Find, for  $t \geq 0$ , for a set  $\mathcal{M}$  of  $N_m$  metabolites and for a set  $\mathcal{B}_m$  of  $N_b$  micro-organisms the flux vector

$\hat{\nu}(t) = (\nu^{(\hat{b}_1)}(t), \dots, \nu^{(\hat{b}_{N_b})}(t))$  of dimension  $N_r$  such that

$$\hat{\nu}(t) := \arg \min \sum_{m \in \mathcal{M}} \left\| R_{m,t} - \sum_{b \in \mathcal{B}_m} \chi_t^{(b)} \nu_{m,t}^{(b)} \right\|_2^2 + \lambda \mathcal{R}(\nu)$$

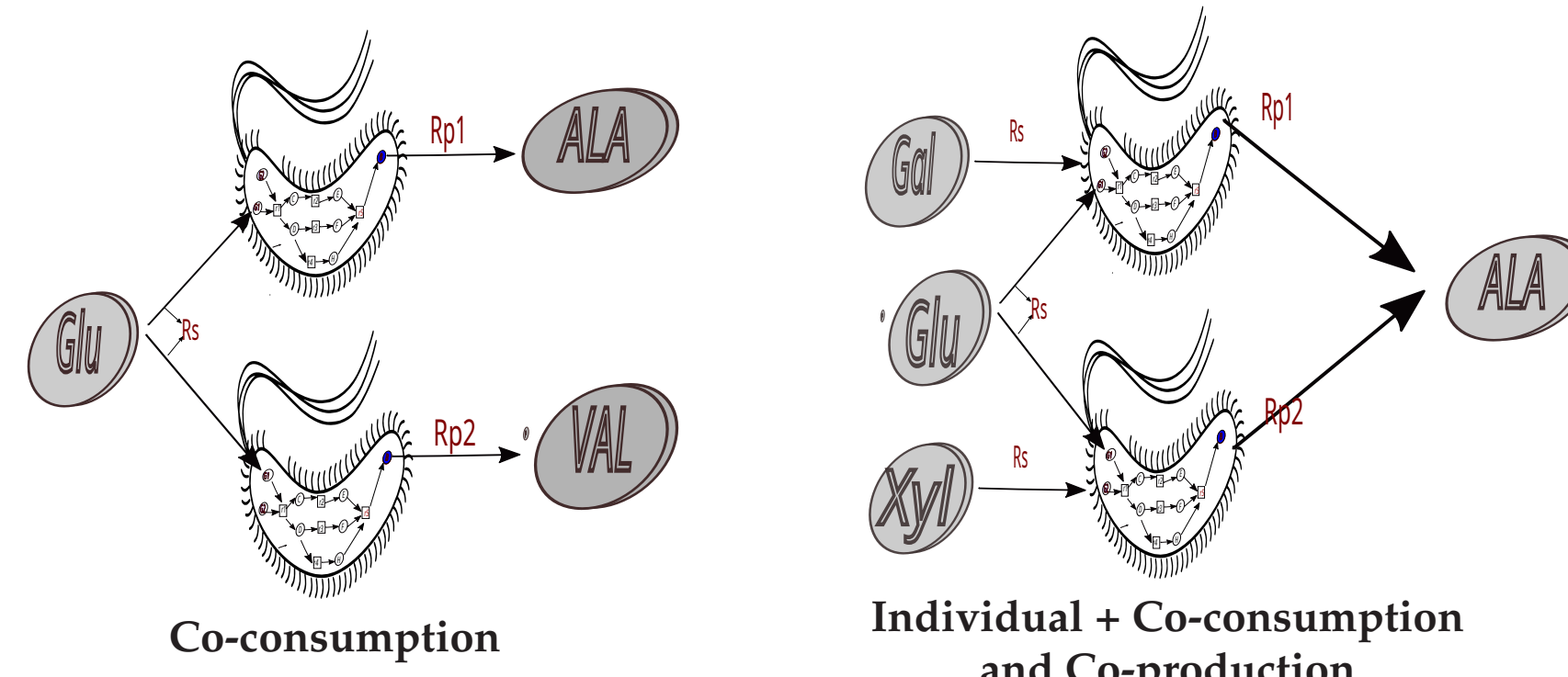
$$S.T. : \begin{cases} S^{(b)} \cdot \nu^{(b)} = 0, \text{ for } b \in \mathcal{B}_m \\ C_{min}^{(b)} \leq \nu^{(b)} \leq C_{max}^{(b)}, \text{ for } b \in \mathcal{B}_m \end{cases}$$

### Regression problem is:

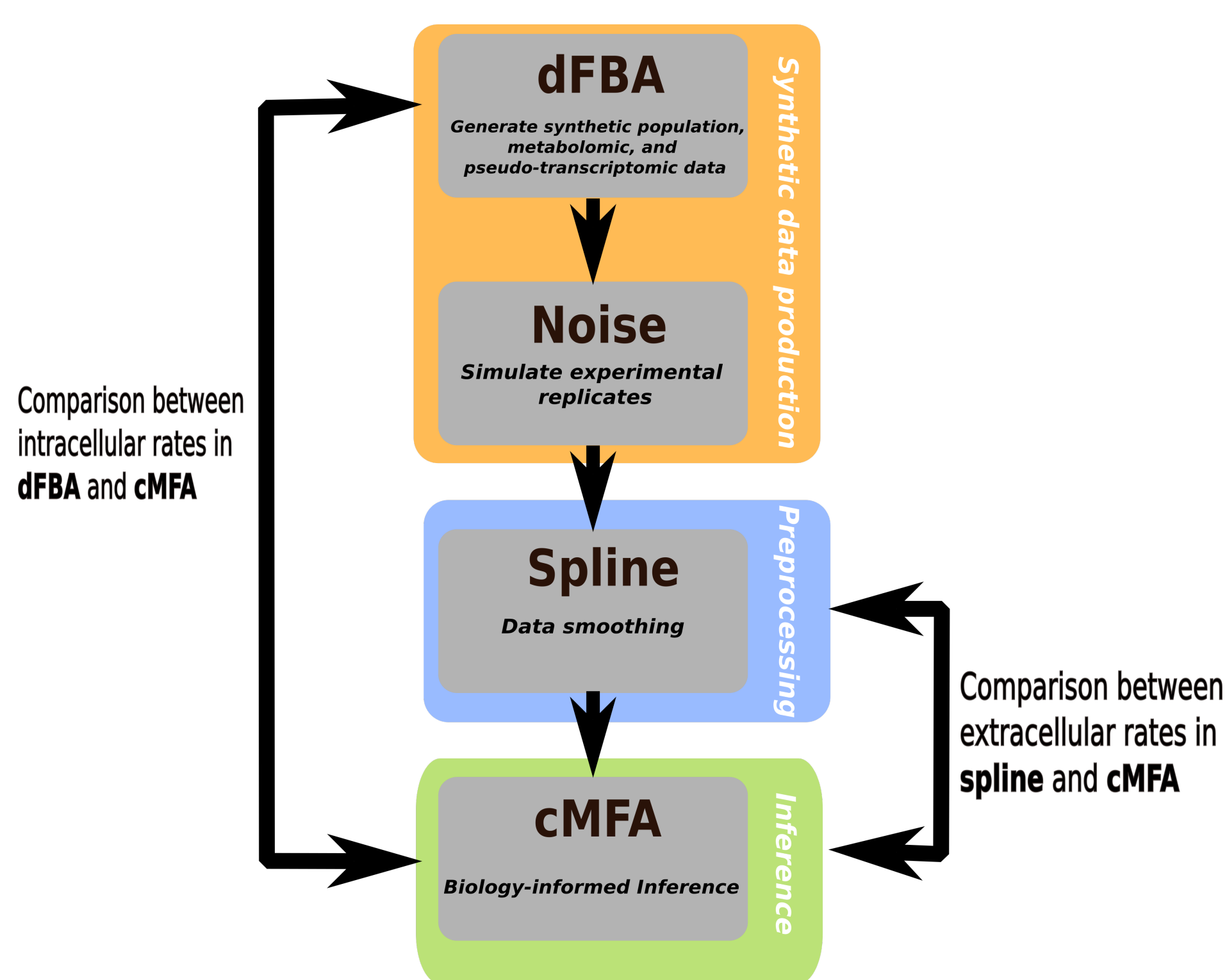
- High-dimensional ( $N_r \sim 5000$ )
- Highly-underdetermined ( $N_r \gg N_m$ )
- Biology-informed ( $S^{(b)}$ )
- Constrained by metatranscriptomics (where  $C_{min}^{(b)} = C_{max}^{(b)} = 0$ )
- Regularized by ridge ( $\mathcal{R}(\nu) = \|\nu\|_2^2$ ) or lasso ( $\mathcal{R}(\nu) = \|\nu\|_1$ ) penalty

### CO-CULTURE DEFINITION

Biological situation of E. coli to produce amine acids and ALA super producer



### BENCHMARK METHODOLOGY



### GENERATING DATA: dFBA

Dynamic Flux Balance Analysis (dFBA) is an extension of traditional Flux Balance Analysis (FBA), which is a mathematical modeling approach used in systems biology to predict the flow of metabolites through metabolic networks.

$$\begin{aligned} \partial_t b_i &= B \nu_i(M, b) b_i \\ \partial_t M_j &= \sum_{i \in C} \nu_{i,j}(M, b) b_i \end{aligned}$$

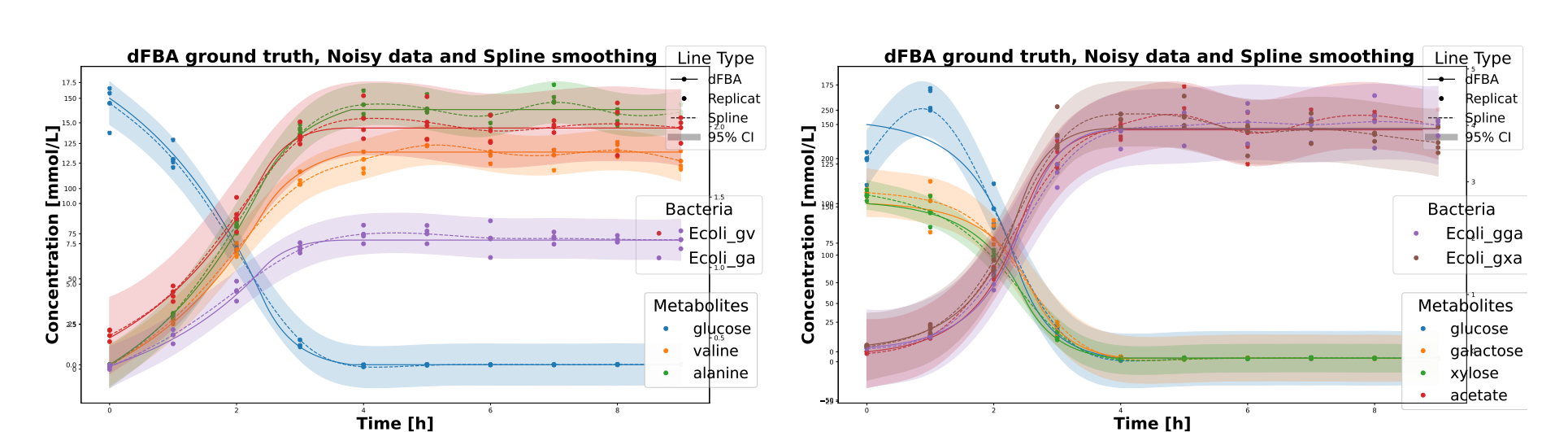
$$\nu_i = \arg \max B \nu_i$$

$$S.T. \begin{cases} S_i \nu_i = 0 \\ C_M \leq \nu_i^s \leq 0 \\ 0 \leq \nu_i^p \leq C_{max} \\ -C_{max} \leq \nu_i^{in} \leq C_{max} \end{cases} \quad S = \begin{pmatrix} m_1 & r_1 & r_2 & r_3 \\ m_2 & -1 & 0 & 0 \\ m_3 & -1 & 0 & 0 \\ m_4 & 1 & -1 & 0 \\ m_5 & 1 & 0 & -1 \\ m_6 & 0 & 1 & 0 \\ m_7 & 0 & 0 & 1 \\ m_7 & 0 & 0 & 0 \end{pmatrix}$$

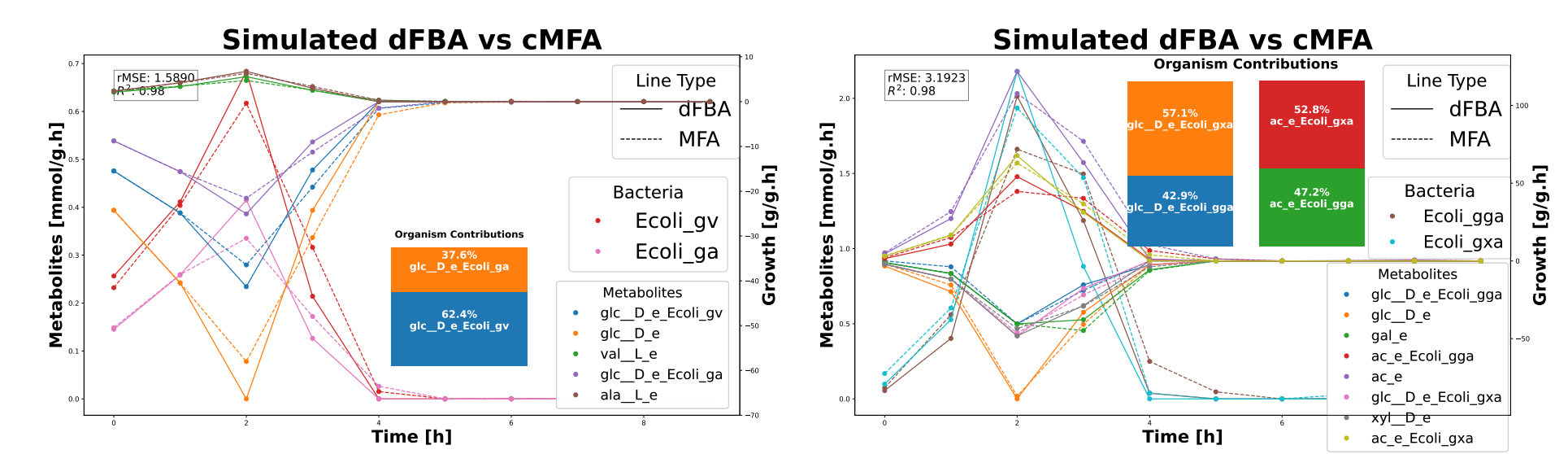
### IMPLEMENTATION

- NumPy and Pandas for data handling and structure.
- PyGAM for spline smoothing.
- Matplotlib for visual representations.
- COBRApy for constraint-based modeling (dFBA).
- OSQP for quadratic programming problems.

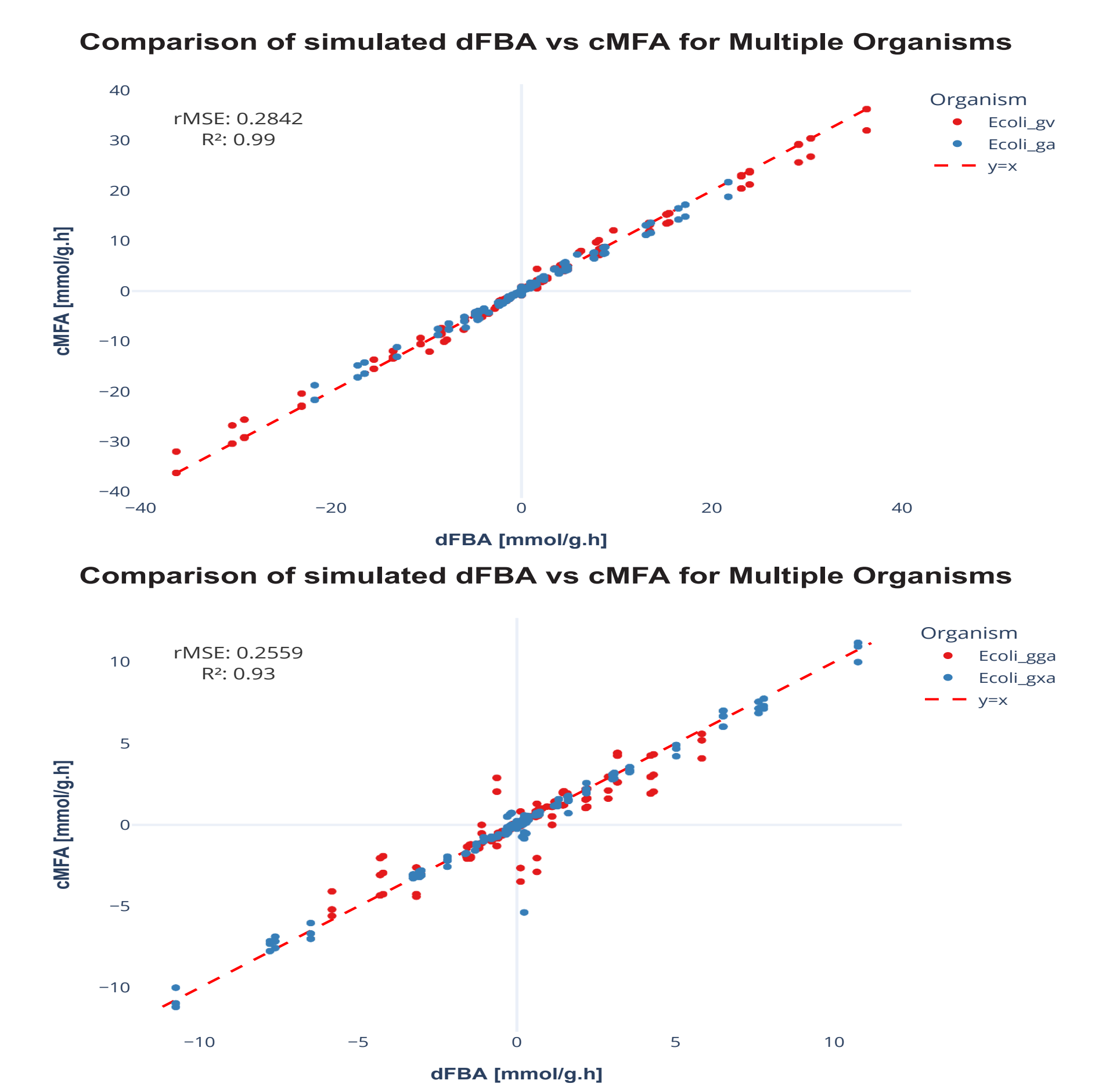
### DATA SMOOTHING WITH SPLINE



### INFERRING EXCHANGE FLUXES: dFBA vs cMFA



### INFERRING INTERNAL FLUXES



## PERSPECTIVES

- Benchmark on real data.
- Expand the community model to incorporate a diverse set of more than 10 to 20 microorganisms.
- The extension of MFA and the inclusion of metatranscriptomic data

## REFERENCES

- Brunner et al., 2023, *PLoS Comput. Biol.* 19, e1011661
- Gerken Starepravo et al., 2022, *Digit. Chem. Eng.* 2, 100011
- Lecomte et al., 2023, *Metab. Eng.* doi.org/10.1016/j.ymben.2024.02.014
- Mahadevan et al., 2002, *Biophys. J.* 83, 1331-1340
- Pansu, 2004, Courbes b-splines