Mathematical Modeling of Microbial Community Dynamics

Ana Paredes^{*,***} Eva Balsa-Canto^{**} Julio R. Banga^{*}

* Computational Biology Lab, MBG-CSIC (Spanish National Research Council), Pontevedra, Spain (j.r.banga@csic.es)
** Biosystems and Bioprocess Engineering, IIM-CSIC (Spanish National Research Council), Vigo, Spain (ebalsa@iim.csic.es)
*** Applied Mathematics Dept., Universidade Santiago de Compostela, Santiago de Compostela, Spain

1. INTRODUCTION

Microorganisms rarely exist in isolation. Instead, they form complex networks of ecological interactions, known as microbial communities. These communities, composed of bacteria, fungi, viruses, and other microorganisms, are ubiquitous across diverse environments, from soil and water to extreme habitats such as hot springs and acidic mines. Microbial communities also thrive within and on plants and animals, including humans, where they play essential roles, particularly in the gut microbiome. Systems biology provides a quantitative framework to study these communities through mathematical models, enabling a structured and nuanced understanding of their dynamics.

2. MODEL CALIBRATION

Our research focuses on the temporal dynamics of microbial communities, for which Ordinary Differential Equations (ODEs) provide an appropriate modeling framework. A critical component of ODE-based modeling is the calibration process, also known as parameter estimation. Calibration involves identifying unknown or non-measurable parameters by adjusting the model to fit experimental data. Typically, this is an iterative process that encompasses several steps (Balsa-Canto et al., 2010; Villaverde et al., 2021).

However, this process is full of possible pitfalls and challenges due to the potential non-uniqueness, ill-conditioning and non-convexity of the estimation problem. Here, we focus on issues related to (i) lack of identifiability and (ii) convergence difficulties during the parameter estimation (under- and over-fitting).

We investigate a set of canonical models with increasing complexity that represent the most common frameworks in microbial ecology, from the most classical and simple ecological models (such as *Generalized Lotka-Volterra* (GLV) models), to more complex models accounting for nutrients dynamics (such as food web models), and coarse-grained

models incorporating different regulatory mechanisms and nutrients' dynamics.

2.1 Structural Identifiability Analysis

Structural identifiability (SI) in the context of ODE-based dynamic models refers to the theoretical possibility of uniquely determining parameter values from ideal model outputs. This assumes perfect, noise-free, and continuous measurements, allowing for an assessment of whether the model structure itself permits unique parameter estimation, independent of data quality or experimental conditions. This concept is crucial because if a model is not structurally identifiable, it means that there could be multiple sets of parameter values that produce the same output, making it impossible to accurately estimate those parameters. Although crucial, structural identifiability analysis (SIA) has been the focus of only a few studies (Balsa-Canto et al., 2020; Remien et al., 2021; Díaz-Seoane et al., 2023). This concept is extremely important because if a model is not structurally identifiable, it means that there could be multiple sets of parameter values that produce the same output, making it impossible to accurately estimate them.

SIA classifies unknown parameters into three groups: globally identifiable, locally identifiable, and non-identifiable. If, after performing the analysis, some parameters are classified as non-identifiable, possible solutions are a reformulation of the model, fixing the non-identifiable parameters to realistic values, or planning additional experiments (if possible). These new experiments could include new observables, experimental conditions, or initial conditions.

The general question of *SIA* for arbitrary non-linear dynamic models described by *ODEs* remains an open and unresolved matter. Nevertheless, significant progress has been made over the past two decades, leading to the development of several promising software tools (Rey Barreiro and Villaverde, 2023). In this study, we focus on three state-of-the-art tools: Structural Identifiability, GenSSI2 and SIAN. After testing these software with the selected case studies, our analysis indicate that the most efficient and robust tool is Structural Identifiability, while both SIAN and GenSSI2 are still reasonable options. When dealing with the most complex models, these tools encountered several difficulties.

^{*} This work was supported by grant PID2020-117271RB-C22 (BIO-DYNAMICS) funded by MCIN/AEI/10.13039/501100011033, grant PID2023-146275NB-C22 (DYNAMO-bio) funded by MICIU/AEI/ 10.13039/501100011033 and ERDF/EU, grant IN607B 2023/04 funded by Xunta de Galicia and grant CSIC PIE 202470E108 (LARGO).

Specifically, GenSSI2 (a fully symbolic tool) experienced a significant surge in memory consumption and computation time, and sometimes failed to guarantee the uniqueness of the solution.

2.2 Practical Identifiability Analysis

Model calibration involves finding the optimal parameter values that best align the model outputs with experimental data. This process is formulated as a non-linear optimization problem. The objective is to estimate the parameters which minimize a cost function that quantifies the discrepancy between model predictions and observed data. This optimization is conducted subject to the constraints imposed by the system of ordinary differential equations (ODEs) that define the model, as well as any additional algebraic constraints that may apply.

Even with structural identifiability, practical identifiability can be compromised by insufficient data or noise, affecting parameter uniqueness and model reliability. There are several methods to assess Practical Identifiability, including the use of the Fisher Information Matrix (FIM), Profile Likelihoods or Bayesian sampling based procedures, for example.

To perform a Practical Identifiability Analysis (PIA), here we employ the AMIGO2 (Advanced Model Identification using Global Optimization) toolbox for Matlab, which facilitates parameter estimation using global optimization, followed by sensitivity analyses and FIM-based PIA (Balsa-Canto et al., 2016).



Fig. 1. Two solutions from the calibration of a *GLV* model considering two-species: local optimum (left), and global optimum (right).

Underfitting (Figure 1) occurs when the estimation algorithm converges to a local optimum. As a consequence, the calibrated model fails to capture the underlying dynamics of the data, leading to inaccurate parameter estimates. Utilizing global optimizers in AMIGO2 allowed us to sidestep these local solutions.

Due to the flexibility and oscillatory nature of several of the models considered, we also observed that their calibration can result in overfitting, i.e. fitting the noise instead of the signal (see the example in Figure 2). In other words, the fit is very good, but the predictive power of the calibrated model is very poor. To surmount this common pitfall, at least two strategies are possible: (i) simplifying the model (sensitivity analyses can help to select the parameters to be fixed or removed); (ii) use regularization techniques (to reduce the ill-conditioning of the problem).



Fig. 2. Overfitting in a GLV model for a three species system, showing spurious oscillations in the dynamics.

3. CONCLUSION

In this study, we have addressed several key issues involved in the mathematical modeling of the dynamics of microbial communities. In particular, we considered the calibration of dynamic models composed of deterministic nonlinear ordinary differential equations. First, we illustrated why Structural Identifiability Analysis (*SIA*) is a critical step in model calibration. After testing the latest available software tools, our results indicate that **Structural Identifiability** is the most robust and efficient. Second, we also illustrated two other potential pitfalls during parameter estimation, underfitting and overfitting, which can compromise the calibrated model accuracy. Addressing these challenges strengthens the predictive power of the model, facilitating more effective applications in microbial ecosystem management.

REFERENCES

- Balsa-Canto, E., Alonso, A.A., and Banga, J.R. (2010). An iterative identification procedure for dynamic modeling of biochemical networks. *BMC systems biology*, 4, 1–18.
- Balsa-Canto, E., Alonso-Del-Real, J., and Querol, A. (2020). Mixed growth curve data do not suffice to fully characterize the dynamics of mixed cultures. *Proc. Natl. Acad. Sci. U. S. A.*, 117(2), 811–813.
- Balsa-Canto, E., Henriques, D., Gábor, A., and Banga, J.R. (2016). AMIGO2, a toolbox for dynamic modeling, optimization and control in systems biology. *Bioinformatics*, 32(21), 3357–3359.
- Díaz-Seoane, S., Sellán, E., and Villaverde, A.F. (2023). Structural identifiability and observability of microbial community models. *Bioengineering (Basel)*, 10(4).
- Remien, C.H., Eckwright, M.J., and Ridenhour, B.J. (2021). Structural identifiability of the generalized lotka–volterra model for microbiome studies. *Royal Society Open Science*, 8(7), 201378.
- Rey Barreiro, X. and Villaverde, A.F. (2023). Benchmarking tools for a priori identifiability analysis. *Bioinformatics*, 39(2), btad065.
- Villaverde, A.F., Pathirana, D., Fröhlich, F., Hasenauer, J., and Banga, J.R. (2021). A protocol for dynamic model calibration. *Briefings in Bioinformatics*, 23(1).