Dynamic hybrid model for biopesticides production using *Bacillus thuringiensis* strains

Sergio FIGUEROA CARDONA
Carlos ROBLES RODRIGUEZ
Rim EL-JENI
Luc FILLAUDEAU
Cesar ACEVES LARA

DCAI 2024
June 26th-28th, 2024

International Conference DCAI

Toulouse Biotechnology Institute
Bio & Chemical Engineering

Outline

1. Biopesticides with *Bacillus thuringiensis (Bt)*
2. Proposed hybrid dynamic model
3. Results
4. Conclusions and perspectives

tbi Toulouse Biotechnology Institute

International Conference DCAI

Biopesticides with *Bt*

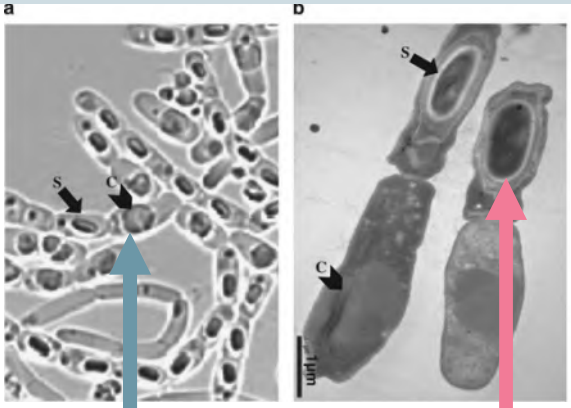
SAFWA

alternative biopesticides for **SAFe** integrated pest and **WAter** management around Mediterranean

10 laboratories from 7 countries

Goal:
Industrial production of biopesticides

Bacillus thuringiensis (Bt)



Cry-proteins/ endotoxins/ proteins

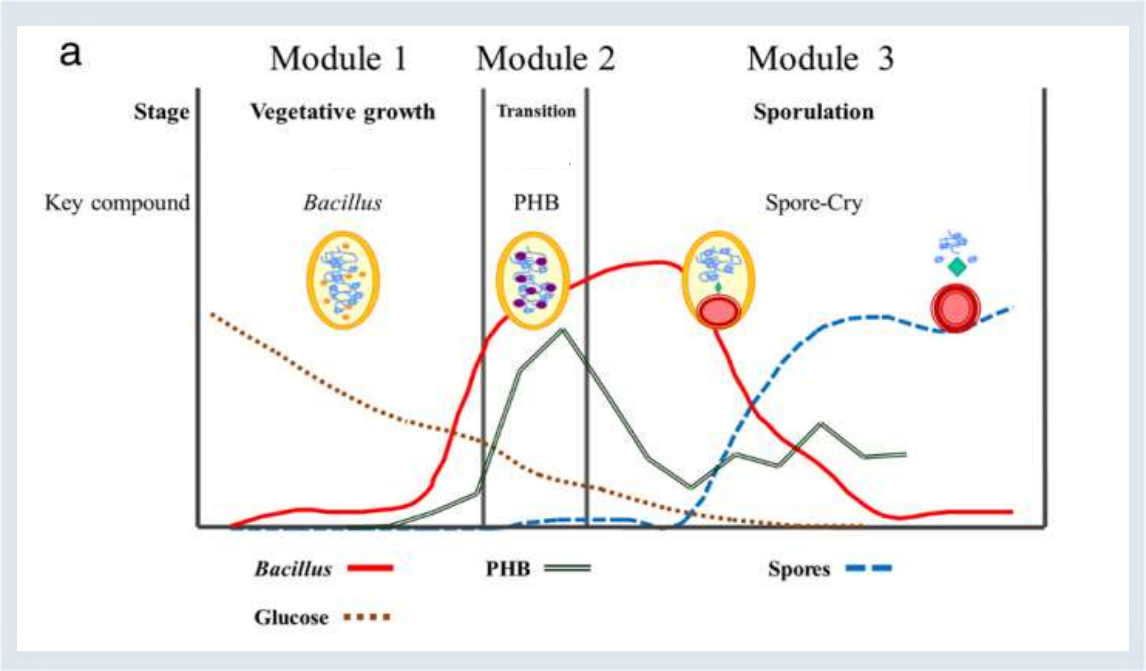
Spores

(Deng *et al.*, 2015)

Modeling approach

Modeling of biopesticides by *Bt*

Bt goes through **three main stages** in a **reactor**:



(Navarro-Mtz & Pérez-Guevara, 2014)

There are previous **kinetic** models

Advantages:

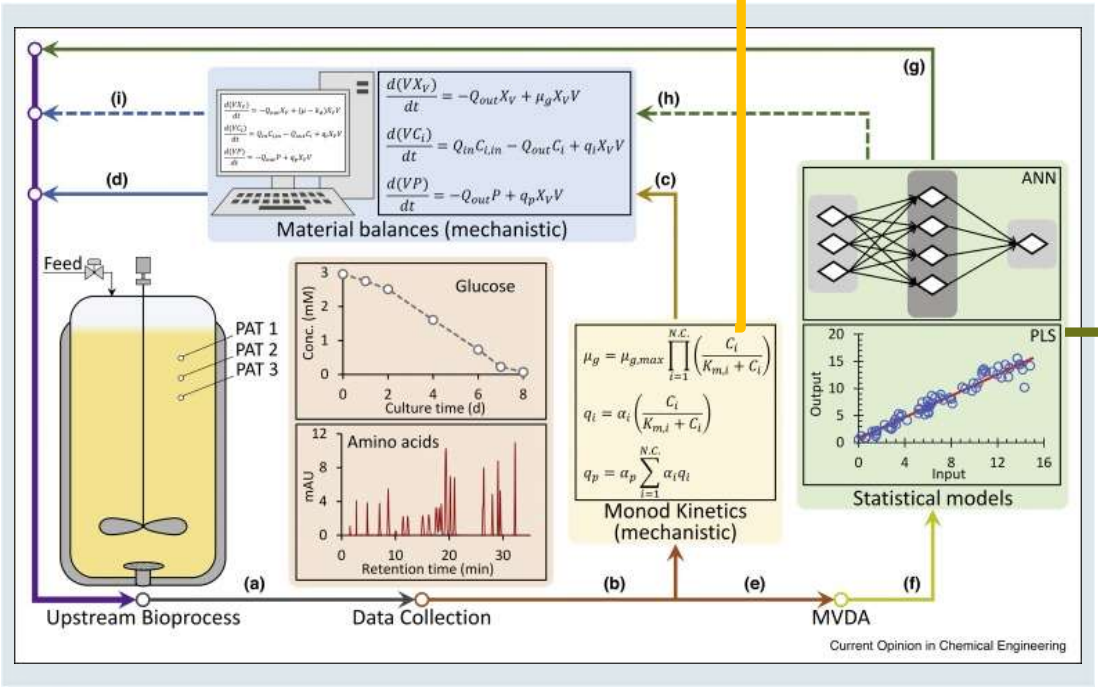
- Presented an accurate description of the **biomass production**
- Considered **oxygen** and **multiple substrates** influence
- Determined that **sporulation and protein** production are **related**

Disadvantages:

- The protein and spore production **has not been properly** simulated

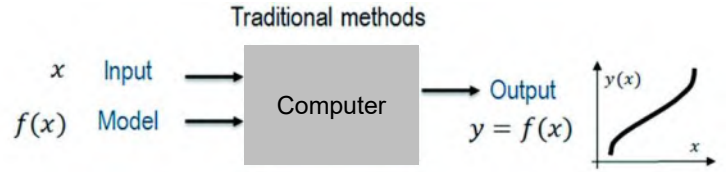
We propose a **hybrid modeling** approach

Hybrid dynamic model

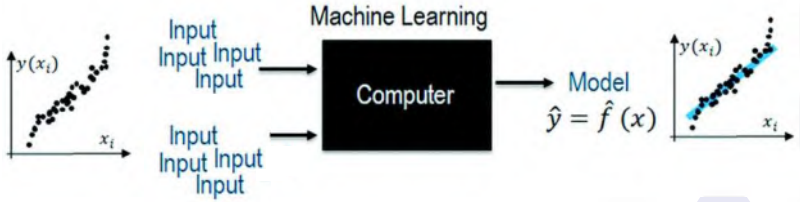


(Tsopanoglou & Jiménez del Val, 2021)

Mechanistic modeling



Data-driven modeling



(Mowbray *et al.*, 2022)

Better for simulating unknown **bioprocess** mechanisms (e.g. kinetics)

Can consider **different operating** conditions

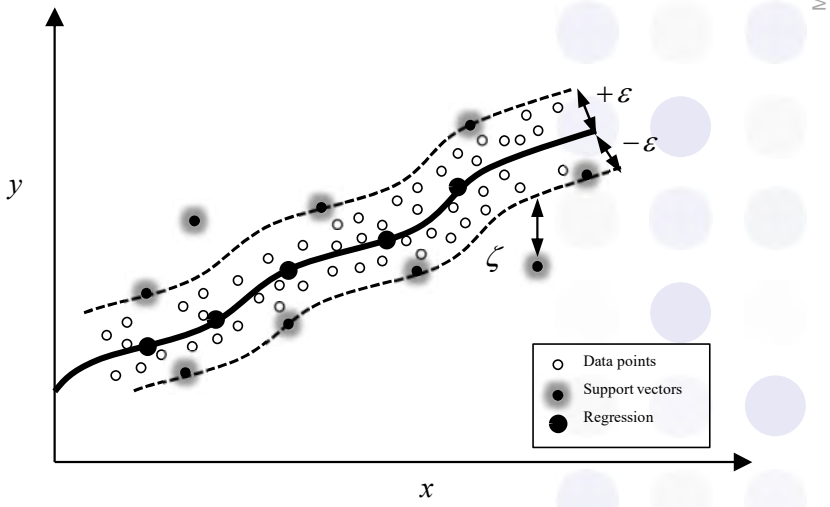
Hybrid dynamic model

Mechanistic model + **Data driven model**

- **Simplified**: few parameters required
- **Versatile**: small set of data

Support Vector Machine (SVM)

$$y = w^T \varphi(x) + b$$



Solves these practical issues:

- **Lack of knowledge** of biochemical mechanisms
- **High costs** of frequent sampling and **online** measurements
- Challenges in pre-determining **set-points**

Is **limited by** the data quantity and quality

Proposed hybrid dynamic model

Mechanistic model

(Monroy *et al.*, 2021)

$$\frac{dX}{dt} = (\mu - k_d) \cdot X$$

$$\frac{dS}{dt} = -\frac{\mu \cdot X}{Y_{XS}} = r_S$$

$$\frac{dPro}{dt} = X \cdot k_{pro} = r_{pro}$$

$$\frac{dSpo}{dt} = X \cdot k_{spo} = r_{spo}$$

$$\frac{\mu_{max} \cdot S}{(K_c \cdot X) + S} = \mu$$

Inputs

Data driven model

SVM regression – Gaussian Kernel

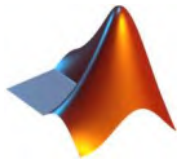
$$r^*_{pro} = f(strain, \mu, r_S, r_{pro}, r_{spo})$$

$$r^*_{spo} = f(strain, \mu, r_S, r_{pro}, r_{spo}, r^*_{pro})$$

$$\frac{dSpo}{dt} = r^*_{spo}$$

$$\frac{dPro}{dt} = r^*_{pro}$$

Recalculation



Data available

Table 1. *Inventory of experimental data from lab tests in the project.*

	Strain 1 (BLB1)		Strain 2 (HD1)		Strain 3 (Lip)	
	Batch	**Data per variable**	**Batch**	**Data per variable**	**Batch**	**Data per variable**
Training	**1**	10	**4**	10	**7**	26
	2	10	**5**	7	**8**	8
Validation	**3**	11	**6**	26	**9**	10

The **validation batches** were chosen based on the values of its predictors

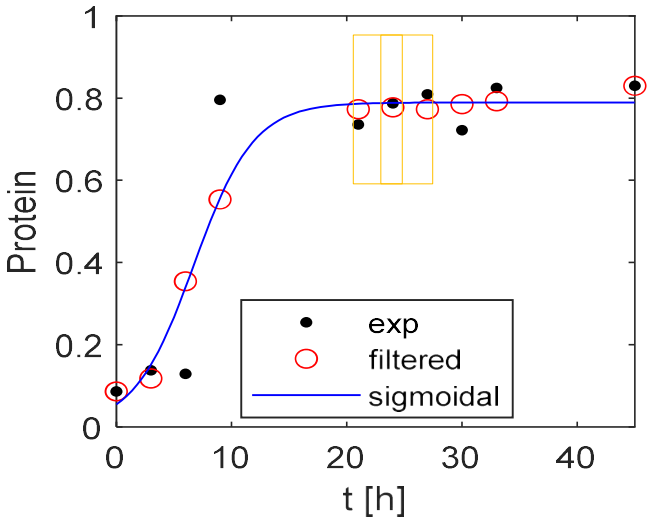
Data filter and sigmoidal function

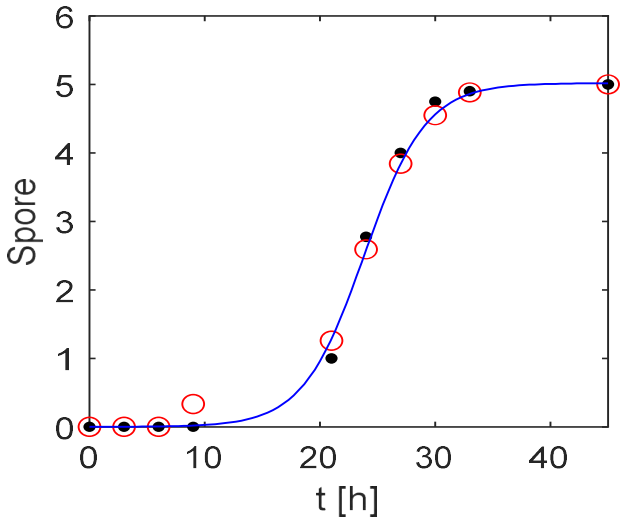
Window median filter

$$y_n = \left(\frac{1}{2 \cdot N + 1}\right) \cdot \sum_{i=n-1}^{n+1} X_i$$

Sigmoidal function

$$f(t) = \frac{1}{a + e^{b+c \cdot t}}$$





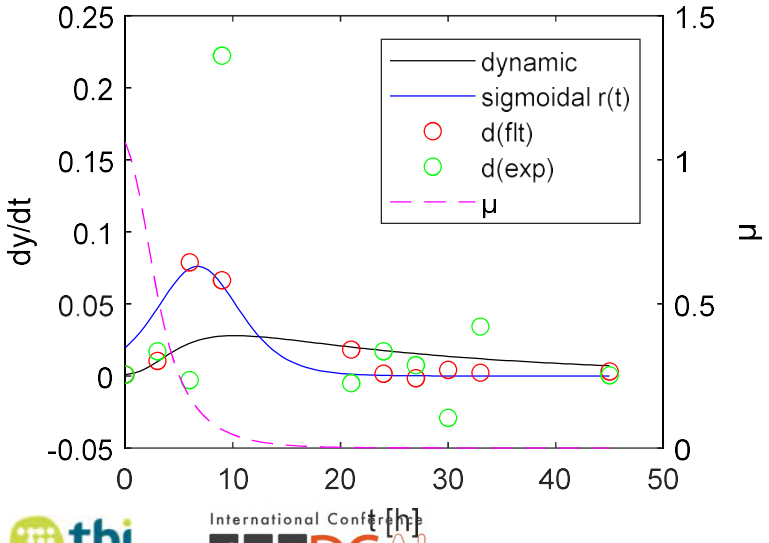
100 data points were generated using the sigmoidal function for each **training batch**

Output and Predictors

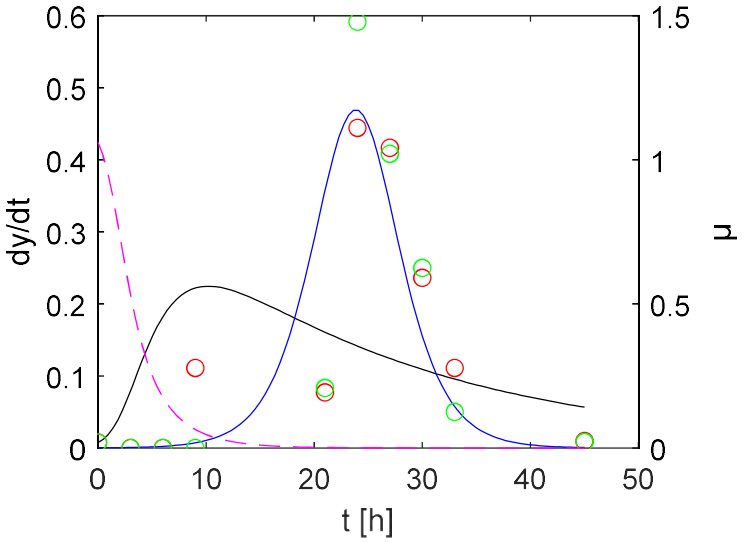
Sigmoidal function derivative

$$\frac{d(f(t))}{dt} = \frac{(-b * e^{b+c \cdot t})}{(a + e^{b+c \cdot t})^2}$$

$$\frac{dPro}{dt} = r^*_{pro}$$



$$\frac{dSpo}{dt} = r^*_{spo}$$



Inputs (Predictors)

- μ
- rS
- rSpo
- rPro
- strain

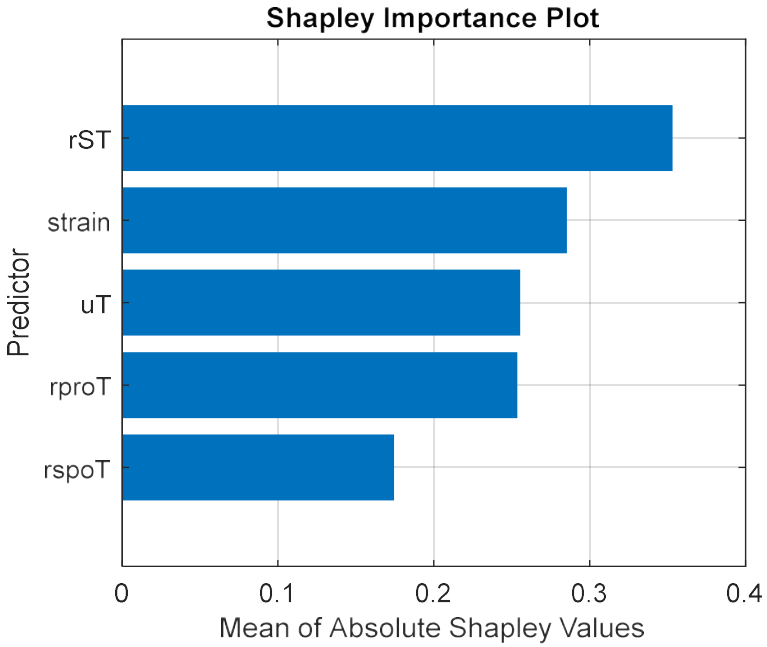
SVM

Outputs

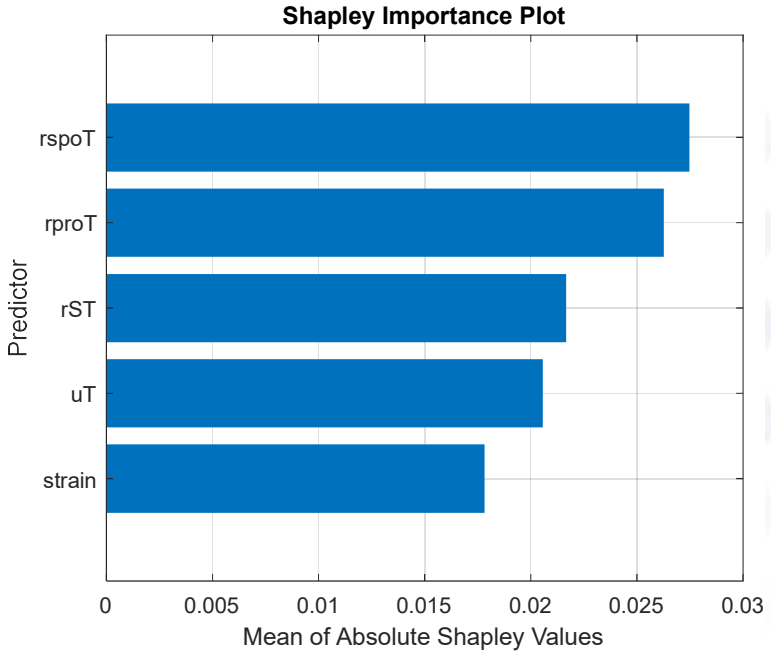
- rpro
- rspo

Shapley values

r_{pro}*



r_{spo}*



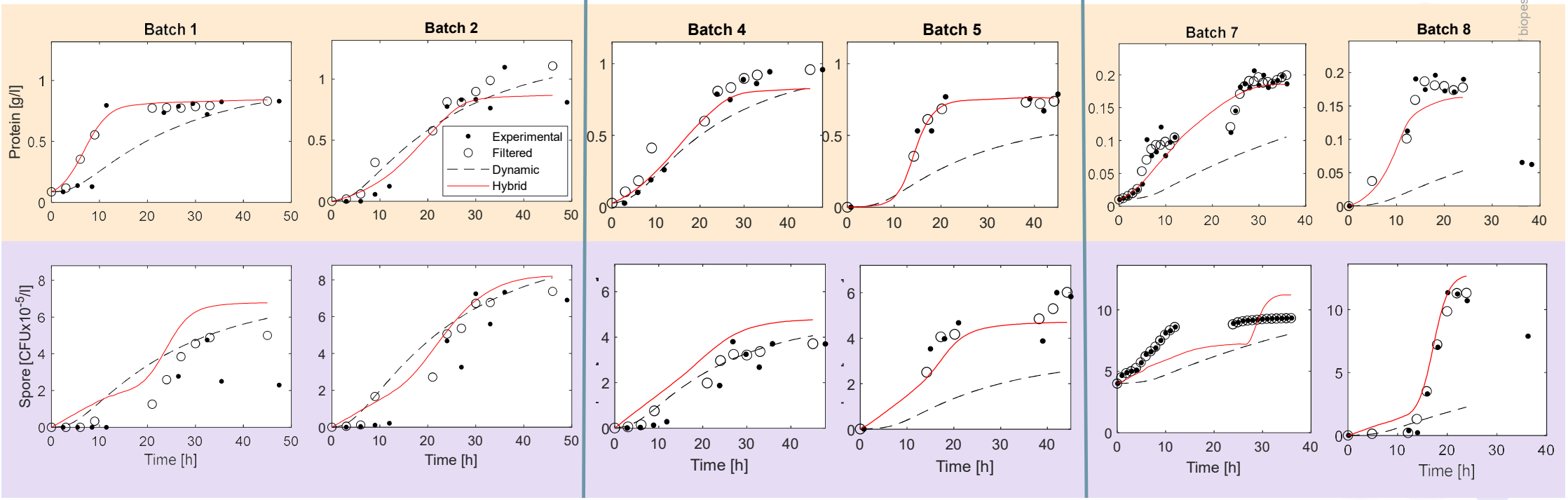
All predictors have **significant influence** over the model output

Result training sets

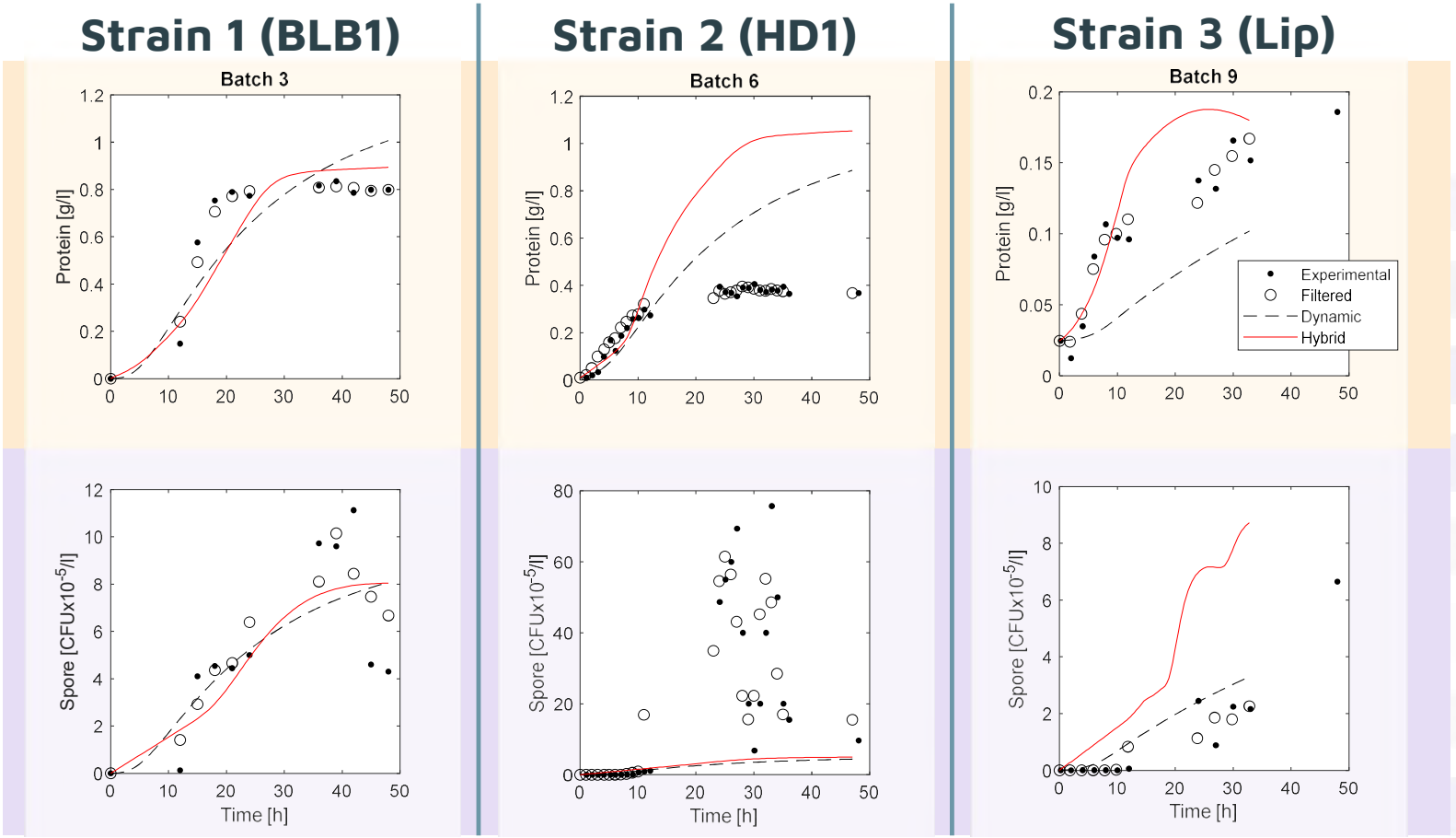
Strain 1 (BLB1)

Strain 2 (HD1)

Strain 3 (Lip)



Result validation sets



NRMSE calculation

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{y_{max} - y_{min}}$$

Table 2. *NRMSE comparison.*

Variable	Data classification	Experimental data	Mechanistic model	Hybrid dynamic model	Improvement
Protein	Training	Original	0.4231	0.0691	
		Filtered	0.4156	0.0540	
	Validation	Original	0.4637	0.4049	**12%**
		Filtered	0.4713	0.4144	
Spore	Training	Original	0.4373	0.1143	
		Filtered	0.4340	0.1065	
	Validation	Original	0.3067	0.2879	**7%**
		Filtered	0.3514	0.3262	

Conclusions

The hybrid dynamic model **has improved the prediction** of the initial dynamic model

The model requires **multiple predictors** and they have the same influence

The **oxygen influence** in the media should be considered in the model

Perspectives

- The model **could be enhanced** by using **more data** from other experiments
- The same approach will be consider to simulate **real substrate** conditions
- The **oxygen influence** could be included in the model

Thanks for your attention!

Any questions?

Dynamic hybrid model for biopesticides production using *Bacillus thuringiensis* strains

Sergio FIGUEROA CARDONA
Carlos E. ROBLES RODRIGUEZ
Rim EL-JENI
Luc FILLAUDEAU
Cesar A. ACEVES LARA

Acknowledgments. This work was supported by the 'Alternative Biopesticides For Safe Integrated Pest And Water Management Around Mediterranean (SAFWA)' project, which has received funding from the European Union's Horizon 2020 research and innovation program under the grant agreement No2022/Section 2. and the French National Agency for Research Investissements d'Avenir ANR-18-EURE-0021.

Dynamic model for three strains

(Monroy *et al.*, 2021)

$$\frac{dX}{dt} = (\mu - k_d) \cdot X = r_X$$

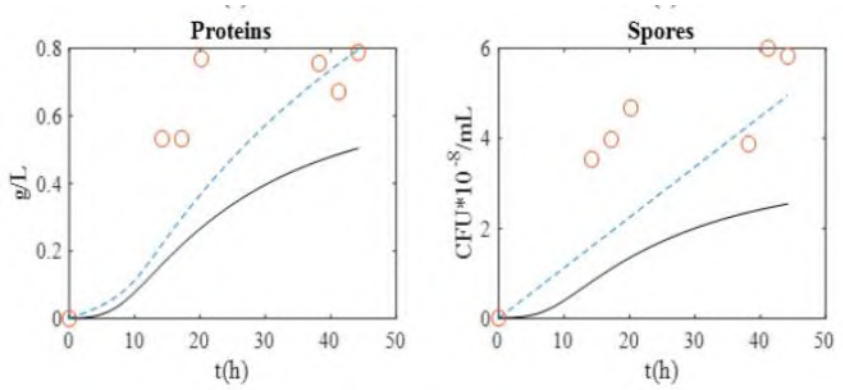
$$\mu = \frac{\mu_{max} \cdot S}{(K_c \cdot X) + S}$$

$$\frac{dS}{dt} = -\frac{\mu \cdot X}{Y_{XS}} = r_S$$

$$\frac{dPro}{dt} = X \cdot k_{pro} = r_{pro}$$

$$\frac{dSpo}{dt} = X \cdot k_{spo} = r_{spo}$$

Training Batch – Strain 02 (HD1)



Validation Batch – Strain 02 (HD1)