



On parameter interpretability of phenomenological-based semiphsical models in biology

Laura Lema-Perez, Rafael Munoz Tamayo, Jose Garcia-Tirado, Hernan
Alvarez

► To cite this version:

Laura Lema-Perez, Rafael Munoz Tamayo, Jose Garcia-Tirado, Hernan Alvarez. On parameter interpretability of phenomenological-based semiphsical models in biology. Informatics in Medicine Unlocked, 2019, 15, pp.100158. 10.1016/j.imu.2019.02.002 . hal-02192507

HAL Id: hal-02192507

<https://hal.science/hal-02192507>

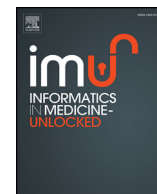
Submitted on 23 Jul 2019

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 - NonCommercial - NoDerivatives 4.0
International License



On parameter interpretability of phenomenological-based semiphysical models in biology

Laura Lema-Perez^{a,*}, Rafael Muñoz-Tamayo^b, Jose Garcia-Tirado^c, Hernan Alvarez^a

^a Universidad Nacional de Colombia, Facultad de Minas, Escuela de Procesos y Energía, Kalman Research Group, Cra 80 No 65-223, 050041, Medellín, Colombia

^b UMR Modélisation Systémique Appliquée aux Ruminants, INRA, AgroParisTech, Université Paris-Saclay, 75005, Paris, France

^c Center for Diabetes Technology, University of Virginia, Charlottesville, VA, USA

ARTICLE INFO

Keywords:

Biological systems
Identifiability
Mechanistic models
Parameter interpretability
Phenomenological based semi-physical model (PBSM)

ABSTRACT

Empirical and phenomenological-based models are used to represent biological and physiological processes. Phenomenological models are derived from the knowledge of the mechanisms that underlie the behavior of the system under study, while empirical models are derived from data analysis to quantify relationships between variables of interest. For studying biological systems, the phenomenological modeling approach offers the great advantage of having a structure with variables and parameters with physical meaning that enhance the interpretability of the model and its further use for decision making. The interpretability property of models, however, remains a vague concept. In this study, we addressed the interpretability property for parameters of phenomenological-based models. To our knowledge, this property has not been deeply discussed, perhaps by the implicit assumption that interpretability is inherent to phenomenological-based models. We propose a conceptual framework to address parameter interpretability and its implications for parameter identifiability, using a simple but relevant model representing the enzymatic degradation of β -casein by a *Lactococcus lactis* bacterium. We illustrated the usefulness of integrating parameter interpretability in the process of construction and exploitation of phenomenological models.

1. Introduction

How can we assess the capability of a mathematical model to provide mechanistic insight on the system under study? That is, how does the mathematical structure of the model translate and capture the phenomena taking place in the system? To what extent can we mechanistically interpret our model? In biotechnology, biology, and biomedical fields two main approaches exist to model processes of interest, namely empirical and phenomenological-based modeling. Empirical-based models are derived from data, while phenomenological-based models are derived from knowledge about the process. In biomedical fields, phenomenological-based models are more relevant than empirical-based models since, in addition to prediction, their parameters and variables provide information that can be used to perform diagnoses, discriminate clinical risk groups and guide treatment by stratifying patients by disease severity [1,2], as we have discussed in our model developments of glucose metabolism with an application perspective for controlling diabetes [3]. In spite of this, in the fields mentioned before many models have been developed from an empirical point of view by using black box modeling approaches like machine

learning and fuzzy models. Machine learning models, for example, are increasingly used in the field of medicine and healthcare but there is still an inability by humans to understand how these models work and the meaning their parameters have. Some approaches have been proposed to improve the level of explanation and interpretability of such empirical models, that is to open the black box [4] (the definition of interpretability is not direct in black box models [5,6]). The deployment of empirical-based models encounters its first hurdle in the difficulty of formalising the definition of central concepts such as transparency, explanation, and interpretability. For example, in Miller [7] the author argues the need to explain artificial intelligence models to provide more transparency to the algorithms from the social sciences. For black box models, interpretability has been regarded as a means to enhance trust in model predictions and to reach features as close as possible to humans regarding decision making [1,8–10]. In this context, Lou, Y. et al. [1] evaluated a method for rule-based learning [11] and applied generalized additive models [1,2,12] to real healthcare problems to get *intelligible* and accurate models, in order to predict risk prior to hospitalizations, to take more informed decision about hospitalization, and to reduce healthcare costs by reducing hospital

* Corresponding author.

E-mail addresses: llemap@unal.edu.co (L. Lema-Perez), hidalvarez@unal.edu.co (H. Alvarez).

<https://doi.org/10.1016/j.imu.2019.02.002>

Received 6 November 2018; Received in revised form 9 February 2019; Accepted 11 February 2019

Available online 16 February 2019

2352-9148/ © 2019 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

admissions [1]. In the same line, Lou et al. [1,12] use *intelligible* models to refer to those models that can be easily interpreted by users. For decision models, the interpretability concept has been ascribed to (i) the ability to make decisions similarly to humans being will do [13,14], and (ii) the ability to be understood [13,15]. Since, decision-making is favoured by the understanding of how the model works. In that sense, optimal decision-based models are those that provide a trade-off between the predictive accuracy and interpretability [16]. For neural networks models, interpretation and explanation of model predictions are challenging aspects as introduced in Ref. [17]. Recent works have been undertaking the development of approaches for improving model interpretability [18] including applications on palliative care service [19]. Similar efforts on improving interpretability are found for fuzzy rule-based systems [20]. In this work, we focus on the interpretability concept as applied to phenomenological-based models and its link with parameter identifiability. While parameter identifiability has been applied to a certain extent to phenomenological-based models [21,22], interpretability has been rarely addressed. One of the few literature studies illustrated how the integration of two mathematical models could enhance the parameter interpretability and validity of a physiological model representing the muscular binding of calcium ions [23].

Model interpretability is a term used in various works but without an explicit definition [13,23,24]. The meaning of that term is not direct because the model as a whole is a complex piece of knowledge. Therefore, model interpretability is not an on-off property; rather, its evaluation requires grading a model on a scale of interpretability. Obviously, that scale requires a metric to generate the value of interpretability for a given model. This metric is the major problem of establishing an interpretability scale. As example, let us consider two models of 30 and 3 parameters, respectively. If each model has only one noninterpretable parameter, an on-off approach would indicate that both models are not interpretable. But if an interpretability index (II) is stated as: $II = 1 - \frac{NP_{NoI}}{N_{TotP}}$, with NP_{NoI} the number of noninterpretable parameters and N_{TotP} the total number of parameters, the II for the first model will be $1 - \frac{1}{30} = 0.9666$ and for the second one will be $1 - \frac{1}{3} = 0.6666$. Does this proposed II give useful information about model size or complexity? Due to this unsolved item, in the current work, interpretability will be only evaluated in terms of individual parameters. Interpretability of model parameters is the result of multiple factors including the level of detail or specification [25] associated to model granularity [26]. Due to the lack of formalism about interpretability as a property of the parameters in a model, there is no consensus about quantifying or measuring such a property. The approach we want to elaborate in this article consists in referring the interpretability of a model to its parameters and the degree to which those parameters have physical meaning. In that regard, the focus of this work is on Phenomenological-Based Semi-physical Models (PBSMs) [27], of which, to the best of our knowledge, the concept of interpretability has not been deeply discussed, perhaps due to the implicit assumption that interpretability is inherent to the PBSM since they are derived from a phenomenological representation of the system under study. In this work, we propose a conceptual framework that can facilitate the incorporation of interpretability for model construction. Such a conceptual framework can provide useful information to develop further a mathematical formalism to characterise parameter interpretability. To facilitate our developments, we use a simple model as an example to elaborate our developments. The paper is organized as follows. In Section 2, we present a summary of the steps of a modeling methodology proposed by Ref. [28] to build PBSMs. In Section 3, a conceptual framework for interpretability analysis is established using a simple mathematical model of the dynamics of enzymatic hydrolysis of β -casein by a *Lactococcus lactis* bacterium. Finally, in Section 4 we discuss the potential links between interpretability and identifiability. Concluding remarks are provided in Section 5.

2. The process of PBSM construction

The construction of a model may be compared to a form of art. This subjective character explains the existence of several methodologies for building PBSMs [29–34]. In our research group (KALMAN, Universidad Nacional de Colombia), several studies have been developed [27,28,35] to propose the following methodology, described by 10 steps, which are summarized here in the interest of completeness.

- 1. Process description and model aim:** a verbal description of the process taking place is performed including a process flow diagram as graphical representation. Also, the model aim is set by the question that is expected to be answered by the model.
- 2. Model hypothesis and level of detail:** a hypothesis or analogy about the behavior of the real process is proposed. Although the present methodology was originally intended for process engineering systems, it can be extended to any type of process by means of a model hypothesis. A model hypothesis is a feasible analogy for the unknown phenomena in terms of known and well studied phenomena. If the modeled process is located in a specific area of engineering in which the phenomena of the process are known, the hypothesis is the description of those phenomena and an analogy is not necessary. Otherwise, the process must be related to a known process, an analogy is required, and a set of assumptions is fixed. The level of detail is determined by the model objective, that is, the question that will be solved by the model.
- 3. Definition of the process systems:** a process system is an abstraction of a part of the process under study [31]. Each process system (PS) is a part of the real process, and this part should be as real as possible in terms of physical distinctions and changes in phases or characteristics showing spatial variations in the process of interest.
- 4. Application of the conservation law:** the conservation law is applied to every PS defined in step 3. Typically, mass, energy, and momentum are accounted for. The equations obtained are described by either a set of ordinary differential equations in lumped models or a set of partial differential equations in distributed models; these form the basic structure of the model.
- 5. Determination of the basic structure of the model:** the set of equations needed to describe the model objective is selected. Those equations with trial information are discarded.
- 6. Definition of the variables, structural parameters and constants:** A list of variables, structural parameters, and constants is made. Variables are quantities whose values result from the solution of the model equations forming the basic structure. Parameters are values that need to be defined beforehand to solve the model. They can be known values or must be identified. Finally, the constants are fixed values either because of its universality (e.g., the gravity constant) or because of choices by the modeler choice (e.g., setting a parameter with a known value from the literature).
- 7. Definition of constitutive and assessment equations and functional parameters:** constitutive and assessment equations are proposed to calculate the largest number of unknown parameters of each process system. The set of constitutive and assessment equations are selected according to the modeler's knowledge and criteria.
- 8. Verification of the degrees of freedom (DoF):** the DoF are the difference between the number of unknowns and the number of equations.
- 9. Construction of the computational model:** the solution of the mathematical model is carried out by a computational program able to solve the set of differential and algebraic equations forming the model.
- 10. Model validation:** the model's domain of validity with respect to available experimental data and other validated models is verified.

Table 1
Definition of concepts used in this study.

Term	Definition
Variable	Quantity to be solved by the model.
Basic structure	Set of equations obtained after applying the conservation law. At this level, the functions that represent the phenomena that take place in the study object are not detailed mathematically.
Extended structure	Set of equations allowing the specification of the parameters represented by mathematical functions. The extended structure results from defining the mathematical equations of the parameters contained in the model basic structure. Some of these equations, called assessment equations, are trivial, i.e., they imply only the assignment of a numerical value to a parameter.
Model structure	Set of equations consisting in the union of the basic and extended structures.
Constitutive and assessment equations	Equations inside the extended structure of the model acting as a mathematical specification of a parameter.
Structural parameter	Parameter inside the basic structure of the model. The structural parameter represents either a quantity that varies in time or a scalar.
Functional parameter	Parameter inside any constitutive or assessment equation. It is categorized as coupled parameter, no-coupled parameter or scalar parameter. These parameters result from the extended structure, once the mathematical equations of the structural parameters are specified.
Scalar functional parameter	Parameter with numerical value (datum) independent of time. This type of parameter can be known <i>a priori</i> or determined by parameter estimation.
Non-coupled functional parameter	Parameter associated with a mathematical function that does not depend on any variable of the model.
Coupled functional parameter	Parameter that depends on at least one variable of the model.
Parameter interpretability	Given a model structure for a system, a parameter p_i is interpretable if it has physical meaning in the real object. In a specific knowledge context, the symbol of the interpretable parameter provides additional information or knowledge about the phenomena under consideration compared to a simple numerical value. The interpretability of a parameter as a property depends on the model structure. Also, the parameter position in the model structure helps to provide interpretability for the parameter being defined.
Contextualized interpretability	Physical meaning of a parameter valid only in a specific mathematical model. The meaning is dependent on the considerations and hypothesis used to deduce the mathematical model within a given context.
General interpretability	Inherent physical meaning of the parameter within a model in a specific scientific domain, i.e., its interpretation is independent on assumptions used to deduce the basic model structure.
Noninterpretability	The parameter has no physical meaning within the model. Noninterpretable parameters must be then represented by a symbol without an interpretable property in the knowledge domain of the process.

3. Setting a conceptual framework for interpretability analysis

In this section, we propose a conceptual framework for parameter interpretability analysis. The concepts that constitute the proposed framework to analyse parameter interpretability are defined and summarized in Table 1. For the sake of clarity, the conceptual framework is studied using a simple mathematical model that describes the dynamics of enzymatic hydrolysis of β -casein by a *Lactococcus lactis* bacterium in a batch system [36]. The **basic structure** of the model is obtained by applying a component mass balance, which results in the following unique differential equation:

$$\frac{dx}{dt} = -r(\cdot) \quad (1)$$

where x (in μM) is the concentration of the substrate and $r(\cdot)$ ($\mu\text{M}/\text{min}$) is the reaction rate, using the symbol (\cdot) to indicate the dependency of this structural parameter with respect to time and any other variable or parameter of the model. It is worth pointing out that global mass balance is worthless in this type of process since no continuous inflow or outflow occurs. From Tables 1 and x is the **variable** whose dynamic trajectory is obtained by solving the model and $r(\cdot)$ is the unique **structural parameter**. Note that at this level of detail, the mathematical equation that represents $r(\cdot)$ is not yet defined. This fact suggests that for this example, Equation (1) is a unique representation of the phenomena of interest (i.e., the hydrolysis of β -casein).

The mathematical definition of the structural parameter $r(\cdot)$ is the key element for the construction of the complete **model structure**, that is, for the set of equations that define the model in its basic and extended form. Multiple mathematical functions exist to define $r(\cdot)$ and describe the hydrolysis rate of the intact β -casein. In the study analyzed here [36], the authors evaluate four kinetic candidate functions to determine the best function for the $r(\cdot)$ parameter in terms of the goodness of fit:

- First-order kinetics:
- $$r(\cdot) = k_1 Ex \quad (2)$$
- n th-order kinetics:

$$r(\cdot) = k_n Ex^n \quad (3)$$

- Michaelis-Menten kinetics:

$$r(\cdot) = k_c E \frac{x}{K_m + x} \quad (4)$$

- Competitive inhibition kinetics:

$$r(\cdot) = k_c E \frac{x}{K_m \left(1 + \frac{I}{K_i}\right) + x} \quad (5)$$

- with $I = x_0 - x$. This expression can be further manipulated to reduce the number of its parameters to:

$$r(\cdot) = b_1 E \frac{x}{b_2 - x} \quad (6)$$

with

$$b_1 = \frac{k_c K_i}{K_m - K_i} \quad (7)$$

$$b_2 = \frac{K_m(K_i + x_0)}{K_m - K_i} \quad (8)$$

where E is the enzyme concentration, measured in optical density units (OD_{600}). The parameter k_1 ($1/\text{OD}_{600} \text{ min}$) is the hydrolysis rate constant for the first-order kinetics, and k_n ($1/\mu\text{M}^{n-1} \text{OD}_{600} \text{ min}$) is the rate constant for the kinetics of order n . For the Michaelis-Menten equation, k_c ($\mu\text{M}/\text{OD}_{600} \text{ min}$) denotes the catalytic rate constant and K_m (μM) the substrate affinity constant. For the inhibition kinetics, K_i (μM) is the inhibition constant. The concentration of the inhibitor I (μM) is considered to be equal to the concentration of β -casein that has been hydrolyzed ($x_0 - x$), with x_0 the initial protein concentration.

It is up to the modeler to decide which kinetic function to use to represent the hydrolysis rate of β -casein. Once, the kinetic function is defined by a new equation in addition to the basic structure, we obtain the **extended structure** of the model. The selected kinetic function is a **constitutive equation** of the model that allows the determination of $r(\cdot)$. For example, if we select the first-order kinetic function $r(\cdot) = k_1 Ex$, we say that $r(\cdot)$ is a **structural coupled parameter** that

depends on the **variable** x and two **functional parameters**: k_1 and E . In this case, both functional parameters have physical meaning and are thus considered to be **interpretable**. While the enzyme concentration E is a known numerical value imposed by the experimental protocol, k_1 is a rate constant that needs to be determined via parameter estimation.

Following the case when $r(\cdot)$ is specified by the first-order kinetic rate as in (2)), we analyse parameter interpretability (the analysis also applies to other candidate kinetic functions, bearing in mind that the Michaelis-Menten equation is derived from a biological hypothesis on the enzyme action and thus its parameters have a stronger level of interpretability than those of the kinetic of order n). By analyzing different experimental conditions, it was found that the hydrolysis rate of β -casein was dependent on the initial protein concentration x_0 [36]. That is, the kinetic rate was slower at higher initial protein concentrations. To account for the dependency of the kinetic rate on the initial β -casein concentration, the authors performed a regression analysis with the estimated parameter values obtained for each experimental condition. After regression, the parameter k_1 was further expressed as a power function of the initial β -casein concentration

$$k_1 = \frac{c_1}{x_0^{m_1}} \quad (9)$$

Equation (9) is referred to as a **constitutive equation**, defined by two new **functional parameters**: c_1 and m_1 . These **scalar parameters** are numerical values identified by regression analysis. Table 2 shows a classification of the components of the β -casein model according to the conceptual framework presented in Table 1 and considering that $r(\cdot)$ is defined by the first-order kinetic rate in Equation (2). It is important to note that for the other kinetics options (Equations (3) - (5)) this classification is also applicable. That is, the basic structure or zero specification level is preserved, but the extended structure changes according to the chosen kinetic constitutive equation. The extended structure begins with the first specification level while the basic structure is the zero specification level and is the only one with inherent interpretability in a PBSM. In Table 2, the interpretability is provided as a qualitative categorical characteristic for the individual parameters. However, the whole interpretability of the model cannot be determined yet using the current conceptual framework.

With respect to the **parameter interpretability** of this simple model, it can be said that the **structural parameter** $r(\cdot)$ has **general interpretability** because in the specific scientific domain of chemical and process engineering, the symbol $r(\cdot)$ denotes a reaction rate. The reaction rate determines the dynamics at which reactants are converted into products, i.e., it is the number of moles of substance reacting per time unit within the reaction. The **functional parameter** k_1 has **contextualized interpretability** and refers to the kinetic rate constant derived from the assumption that the hydrolysis rate follows a first-order kinetics. The **functional parameter** E has also **contextualized interpretability** representing the concentration of the enzyme.

Table 2

Classification of the β -casein model components when using the first-order kinetic rate to represent β -casein hydrolysis.

Symbol	Type	Equation	Interpretability
Basic structure and basic specification or zero specification level			
x	Variable	$\frac{dx}{dt} = -r(\cdot)$	Non required ^a
r	Structural parameter	$r(\cdot) = k_1 E x$	General
1st specification level			
k_1	Non-coupled functional parameter	$k_1 = \frac{c_1}{x_0^{m_1}}$	Contextualized
E	Scalar functional parameter	$E = \text{known}$	Contextualized
2nd specification level			
c_1	Scalar functional parameter	$c_1 = \text{known}$	Noninterpretable
m_1	Scalar functional parameter	$m_1 = \text{known}$	Noninterpretable
x_0	Scalar functional parameter	$x_0 = \text{known}$	General

^a Any model variable has inherent interpretability.

Contextualized means that these symbols, k_1 and E , in other contexts can be used for representing other physical properties of the process.

When k_1 is further defined by the constitutive equation (9) with the scalar functional parameters c_1 and m_1 , they are **not interpretable**, since c_1 and m_1 are empirical parameters without physical meaning. However, the parameter k_1 is still interpretable in spite of being expressed as a function of noninterpretable parameters. The interpretability of a parameter is not dependent on the constitutive equation that defines it in a lower specification level.

In this example, we can appreciate the peculiarity of the basic structure of a model and the dependency on the modeler's choices to define the extended structure. One basic structure can lead to multiple extended structures. This extended structure results from the mathematical specification of the structural parameters. Additionally, this example highlights how the parameter interpretability of the model can be affected when the specification levels appear, that is when the structural and functional parameters must be defined through further parametrization. A graphical explanation of the concepts applied in the example is shown in Fig. 1.

4. Links between parameter interpretability and identifiability

In this section, we discuss possible relations between the concepts of interpretability and identifiability.

4.1. Brief recall on parameter identifiability

Identifiability is a structural property of the model that refers to the ability to find a unique best value of the model parameters from available measurements [37,38]. Under the assumption that the model perfectly represents the system, model identifiability is tested in the hypothetical scenario set by continuous noise-free data and experimental conditions that provide a sufficient excitation of the model response. The structural identifiability is independent of real experimental data. Identifiability is a necessary condition for the parameter identification problem to be well posed. Identifiability testing is of great relevance for models where the parameters are biologically meaningful (as is the case for PBSMs) and we may wish to identify them uniquely [39]. Identifiability testing can be helpful to provide guidelines for how to deal with non-identifiability, either providing hints on how to simplify the model structure or indicating when more information (measured data) is needed for the specific experiment [21].

Let us consider $\mathbf{M}(\mathbf{p})$ a fixed model structure with a set of parameters \mathbf{p} describing the input-output behavior of the system under study. The structural identifiability of the parameter p_i is determined from the following equality

$$\mathbf{M}(\mathbf{p}) = \mathbf{M}(\mathbf{p}^*) \Rightarrow p_i = p_i^* \quad (10)$$

If the equality (10) holds for a unique value of the parameter p_i , the parameter is structurally globally identifiable. If there are a finite number of values for p_i that hold the equality (10), the parameter is structurally locally identifiable. If infinite solutions exist for p_i , the parameter is nonidentifiable. A model is structurally globally (or locally) identifiable if all its parameters are structurally globally (or locally) identifiable. A model is non-identifiable if at least one of its parameters is non-identifiable. Different methods have been proposed to test identifiability of linear and nonlinear models. The interested reader is referred to dedicated literature [37,40,41]. To facilitate identifiability testing, software tools such as DAISY (Differential Algebra for Identifiability of SYstems) [21] and GenSSI have been developed [22]. DAISY is implemented in the symbolic language REDUCE and GenSSI is implemented in Matlab. Both of them are freely available. We made use of both toolboxes for our analysis.

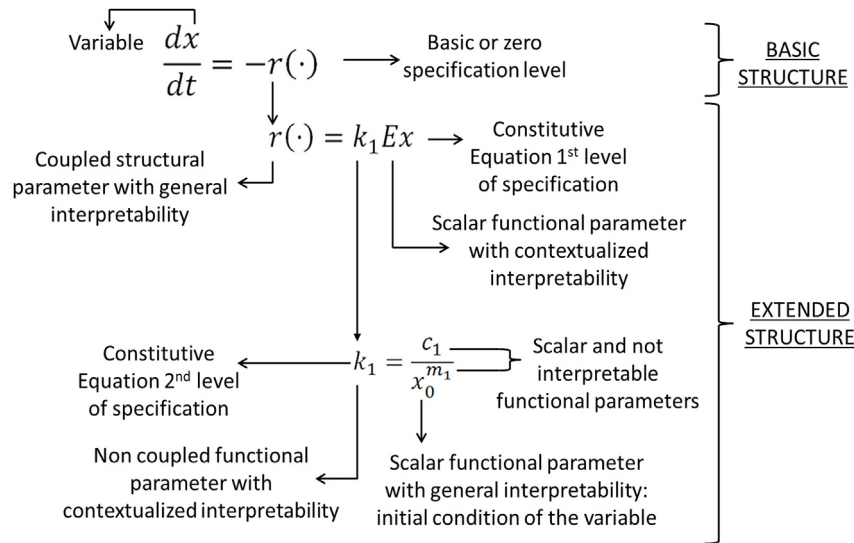


Fig. 1. Concepts applied in a simple model of β -casein hydrolysis by a *Lactococcus lactis* bacterium.

Table 3

A comparison between identifiability and interpretability analysis in β -casein model.

Mathematical expression	Unknown parameters	Identifiability	Interpretability
Basic structure and basic specification or zero specification level $\frac{dx}{dt} = -r$	r	Identifiability does not apply at this level	General
Extended structure - 1st specification level $r = k_1 Ex$	k_1	Globally identifiable ^a	Contextualized
$r = k_n Ex^n$	k_n, n	Locally identifiable ^b	Contextualized
$r = k_c E \frac{x}{K_m + x}$	k_c, K_m	Globally identifiable	Contextualized
$r = k_c E \frac{x}{K_m \left(1 + \frac{I}{K_i}\right) + x}$	$k_c K_m, K_i$	Nonidentifiable	Contextualized
$r = b_1 E \frac{x}{b_2 - x}$	b_1, b_2	Globally identifiable	Nointerpretable
2nd specification level $k_1 = \frac{c_1}{x_0^{m_1}}$	c_1, m_1	Locally identifiable	Nointerpretable

^a Global analysis is performed by using DAISY [21].

^b Local analysis is performed by using GenSSI [22].

4.2. Interpretability vs. identifiability

In our conceptual framework, interpretability is defined as the ability to find the physical meaning of a parameter when the model structure (basic plus extended) and some knowledge of the real process are given. Interpretability is the property of the model parameters, inherited from the model structure, assigning a physical meaning to a parameter within the context where the model is constructed. When the parameter has a physical meaning, it is possible to find from available knowledge a span of numerical values to make its identification easier.

The main role of parameter interpretability for parameter identification is to narrow the search space/domain of the cost function where the identification procedure operates, constraining the values of feasible parameters to match with the existing body of knowledge. On the other hand, structural identifiability is considered a theoretical property. In practice, however, model structure misspecification and noise data can affect the identifiability of the parameters of the model [21] and therefore an accurate identification of the model parameters is not guaranteed. Practical identifiability is then subjected to the quality of

available data. Interpretability can be of help in parameter identification [42] by adding prior knowledge that can be used to constraint parameter estimation. For instance, if a parameter is interpretable, it is possible to know the threshold at which it should be placed. Also, the threshold could be restricted to improve practical identification. A parameter can be non-identifiable, but if it is interpretable, then prior information can be used to facilitate its practical identifiability.

Identifiability and interpretability are relevant properties of PBSMs constructed to gain mechanistic insight of the system under study. A PBSM has a basic structure that is universal and interpretable, that is, all its structural parameters are interpretable. However, it is often necessary to specify the structural parameters in the extended structure, though maintaining the interpretability of a model become more challenging.

Identifiability analysis applies only to scalar parameters (see the definition of scalar parameters in Table 1). In the β -casein model, the structural parameter $r(\cdot)$ is a time variant quantity and thus identifiability testing is not relevant. The quantity $r(\cdot)$ is interpretable and we might wonder if it is possible to estimate it from the available measurements (x). The reconstruction of $r(\cdot)$ belongs to another subject, namely observability, which is not detailed here.

A structural identifiability analysis was performed for the β -casein model by using both the DAISY software tool [21] and GenSSI-Matlab [22], to evaluate how the identifiability properties of the model change with respect to the level of specification or granularity and the candidate constitutive equations. Table 3 summarizes the identifiability and Interpretability analysis. It can be noted that the basic structure of the model is interpretable but its identifiability cannot be tested because $r(\cdot)$ is not scalar. However, its identifiability analysis is later applied and is affected when the structural parameter $r(\cdot)$ is defined by different kinetics. When $r(\cdot)$ is replaced by the first-order kinetic, the model is still identifiable. But, when k_1 is further defined by a mathematical expression dependent on the initial concentration of the protein (located in the second specification level), its identifiability is modified. In the same way, for the second form of competitive inhibition kinetics, where functional parameters b_1 and b_2 are not replaced, the model is globally identifiable, but once b_1 and b_2 are defined and replaced at the next level of specification, the identifiability of the model is affected. Parameters k_1 , k_n , k_c , K_m , and K_i are interpretable from Michaelis-Menten kinetics, but parameters b_1 and b_2 are not interpretable. When the mathematical expression of Michaelis-Menten is changed for the expression with parameters b_1 and b_2 to make its identification easier, interpretability is affected.

We deduce that a PBSM can have an extended structure to identify its parameters and an extended structure to interpret the model parameters. In the case of the β -casein model, two extended structures of the model can be considered depending on the goal: if the goal is to perform parameter identification, the mathematical expression containing parameters b_1 and b_2 is more convenient. On the other hand, if the goal is to exploit the descriptive ability of the model, the mathematical expression with interpretable parameters is then selected. Note that to perform an identifiability analysis of the whole model, all parameters must be replaced by the mathematical expression defining them, whilst interpretability analysis does not require replacing the constitutive equations in the upper specification levels.

4.3. How useful are interpretability and identifiability for constructing PBSMs?

To elaborate further on the usefulness of integrating parameter interpretability at different levels of granularity when dealing with phenomenological modeling of biological systems, we will analyse the two following examples.

1. Let us recall that the Michaelis-Menten kinetics is derived from the following reaction scheme



where E , S , ES , and P are the concentrations of enzyme, substrate, an enzyme-substrate complex, and product, respectively. The parameters k_1 and k_2 are the forward and reverse reaction rates of the reversible reaction that form the complex ES . The parameter k_3 is the reaction rate constant of the product formation reaction. Under certain assumptions, the rate of product formation v is given by

$$v = v_{max} \frac{S}{K_s + S} \quad (12)$$

with the parameters v_{max} and K_s the maximum rate constant and the substrate affinity constant, respectively. It should be noted that Equation (12) is an aggregated representation of Equation (11). While in the two equations, the parameters are interpretable, Equation (12) has less number of parameters. Based on our identifiability studies [?], our educated guess is that the practical identifiability of Equation (12) is superior than that of Equation (11). Accordingly, Equation (12) offers a perfect compromise between parsimony and interpretability.

2. Let us now consider the following kinetic model [43] that incorporates thermodynamic principles on the mathematical representation of microbial growth rate μ

$$\mu = \mu_{max} \exp\left(-\frac{E_M + E_{dis}}{S v_{harv} E_{cat}}\right) \quad (13)$$

with S (mol/L) the energy-limiting substrate. The parameter μ_{max} (h^{-1}) is the maximum specific growth rate constant. The parameters E_{dis} (kJ/mol) and E_M (kJ/mol) are the dissipated exergy and stored exergy during growth, respectively. E_{cat} (kJ/mol) is the catabolic exergy of one molecule of energy-limiting substrate and v_{harv} is the volume (L) at which the microbe can harvest the chemical energy in the form of substrate molecules. Equation (13) can be expressed as

$$\mu = \mu_{max} \exp\left(-\frac{K}{S}\right) \quad (14)$$

with K (mol/L) the affinity constant derived from

$$K = \frac{E_M + E_{dis}}{v_{harv} E_{cat}} \quad (15)$$

As occurs with the Michaelis-Menten kinetics, Equation (14) has better practical identifiability properties than Equation (13). We have

recently incorporated the kinetic Equation (14) into a dynamical mathematical model describing the metabolism of methanogenic archaea [39]. We showed that the parameters of our resulting model are identifiable. The model represented satisfactorily in vitro data of three methanogens. If we would have been interested only on data fitting, knowledge on the derivation of the parameter K and its interpretability would have been unnecessary. However, our interest in our modeling endeavour was to provide biological explanations about observed kinetic and energetic differences between the methanogens. By taking into account the interpretability property of the model parameters, we could provide a biological explanation of the factors responsible of observed energetic differences among the three microbes.

5. Conclusion

Due to the lack of a formal definition of the interpretability concept in the literature and the fact that this topic is just emerging, we propose a conceptual framework for parameter interpretability. We discussed the links between parameter interpretability and identifiability. We illustrated how interpretability and identifiability are instrumental properties to take into consideration in the construction and exploitation of PBSMs.

The concepts here described provide a useful framework for undertaking the construction of models of biological/biomedical systems where the physical meaning of the model structure is a desired property. These concepts are of particular use for modeling systems that are poorly studied and thus facilitate further exploitation of *in silico* simulation. PBSMs offer great advantages for representing biological systems as they allow the enhancement of model capabilities in a sequential way, the integration of multiscale information into the same model, and a guarantee of direct interpretability of model basic structure. In addition, endowing a parameter of a PBSM with interpretability is an easier task than doing the same for empirical models.

Acknowledgment

Laura Lema-Perez is funded by a doctoral scholarship from Colciencias (call 647).

References

- [1] Lou Y, Caruana R, Gehrke J, Hooker G. Accurate intelligible models with pairwise interactions. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13. 2013. p. 623–31. [\let\tempa\bibinfo@X@doi10.1145/2487575.2487579](#).
- [2] Hastie T, Tibshirani R. Generalized additive models: some applications. J Am Stat Assoc 1987;82(398):371–86. [\let\tempa\bibinfo@X@doi10.1080/01621459.1987.10478440](#).
- [3] Lema-Perez L, Garcia-Tirado J, Builes-Montaña C, Alvarez H. Phenomenological-Based model of human stomach and its role in glucose metabolism. J Theor Biol 2019;460:88–100. ISSN 00225193. [\let\tempa\bibinfo@X@doi10.1016/j.jtbi.2018.10.024](#).
- [4] Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A Survey of Methods for Explaining Black Box Models, ACM Computing Surveys (CSUR) 51(5). 2018. p. 42. [\let\tempa\bibinfo@X@doi10.1145/3236009](#).
- [5] Li Y. Deep Reinforcement Learning: An Overview. 2018. p. 1–85.
- [6] Wu M, Hughes MC, Parbhoo S, Zazzi M, Roth V, Doshi-velez F. Beyond sparsity : tree regularization of deep models for interpretability. The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18). 2018. p. 1670–8.
- [7] T. Miller, Explanation in Artificial Intelligence: Insights from the Social Sciences doi: [\let\tempa\bibinfo@X@doiXiv:1706.07269v1](#).
- [8] Kim B. Interactive and Interpretable Machine Learning Models for Human Machine Collaboration. Massachusetts Institute of Technology; 2015.
- [9] Molnar C. Interpretable Machine Learning A Guide for Making Black Box Models Explainable, Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. 2018.
- [10] Ridgeway G, Madigan D, Richardson T, O'Kane J. Interpretable boosted Naïve Bayes classification. the 4th International Conference on Knowledge Discovery and Data Mining (KDD-1998). 1998. p. 101–4.
- [11] Ambrosino R, Buchanan BG, Cooper GF, Fine MJ. The use of misclassification costs to learn rule-based decision support models for cost-effective hospital admission strategies. Proceedings of the Annual Symposium on Computer Application in Medical Care. 1995. 304–8, ISSN 0195–4210.

- [12] Lou Y, Caruana R, Gehrke J. Intelligible models for classification and regression. KDD '12: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. 2012. p. 150–8. [\let\@tempa\bibinfo@X@doi10.1145/2339530.2339556](https://doi.org/10.1145/2339530.2339556).
- [13] Z. C. Lipton, The Mythos of Model Interpretability, CoRR abs/1606.03490.
- [14] Doshi-Velez F, Kim B. Towards A Rigorous Science of Interpretable Machine Learning. 2017. p. 1–13.
- [15] Biran O, Cotton C. Explanation and justification in machine learning : a survey. International Joint Conference on Artificial Intelligence Workshop on Explainable Artificial Intelligence (IJCAI-XAI). 2017. p. 8–13.
- [16] Huysmans J, Dejaeger K, Mues C, Vanthienen J, Baesens B. An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. Decis Support Syst 2011;51(1):141–54. doi:\let\@tempa\bibinfo@X@doihttps://doi.org/10.1016/j.dss.2010.12.003.
- [17] Montavon G, Samek W, Müller K-R. Methods for interpreting and understanding deep neural networks. Digit Signal Process 2018;73:1–15. ISSN 1051–2004, doi:\let\@tempa\bibinfo@X@doihttps://doi.org/10.1016/j.dsp.2017.10.011.
- [18] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Batra, D. Parikh, Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization .
- [19] Avati A, Jung K, Harman S, Downing L, Ng A, Shah NH. Improving palliative care with deep learning. BMC Med Inf Decis Mak 2018;18(4):122. ISSN 1472–6947 \let\@tempa\bibinfo@X@doi10.1186/s12911-018-0677-8.
- [20] Castiello C, Maria A, Lucarelli M, Mencar C. Interpretable fuzzy partitioning of classified data with variable granularity. Appl Soft Comput J 2019;74:567–82. ISSN 1568–4946 \let\@tempa\bibinfo@X@doi10.1016/j.asoc.2018.10.040.
- [21] Bellu G, Saccomani MP, Audoly S, D'Angiò L. DAISY: A new software tool to test global identifiability of biological and physiological systems. Comput Methods Progr Biomed 2007;88(1):52–61. ISSN 01692607 \let\@tempa\bibinfo@X@doi10.1016/j.cmpb.2007.07.002.
- [22] Chis OT, Banga JR, Balsa-Canto E. GenSSI: A software toolbox for structural identifiability analysis of biological models. Bioinformatics 2011;27(18):2610–1. ISSN 13674803 \let\@tempa\bibinfo@X@doi10.1093/bioinformatics/btr431.
- [23] Rockenfeller R, Günther M. Hill equation and Hatze's muscle activation dynamics complement each other: enhanced pharmacological and physiological interpretability of modelled activity-pCa curves. J Theor Biol 2017;431:11–24. ISSN 0022–5193 \let\@tempa\bibinfo@X@doi10.1016/j.jtbi.2017.07.023.
- [24] Juang CF, Chen CY. Data-driven interval type-2 neural fuzzy system with high learning accuracy and improved model interpretability. IEEE Trans Cybern 2013;43(6):1781–95. ISSN 21682267 \let\@tempa\bibinfo@X@doi10.1109/TSMCB.2012.2230253.
- [25] Poursabzi-Sangdeh F, Goldstein DG, Hofman JM, Wortman Vaughan J, Wallach H. Manipulating and measuring model interpretability. 31st Conference on Neural Information Processing Systems - NIPS. 2018. p. 89–93.
- [26] Maier JF, Eckert CM, Clarkson PJ. Model granularity and related concepts. Proceedings of International Design Conference, DESIGN DS, 84. 2016. p. 1327–36. ISSN 18479073.
- [27] Alvarez H, Vega P, Revollar S, Lamanna R. Metodología para la Obtención de Modelos Semifísicos de Base Fenomenológica Aplicada a una Sulfatadora de Jugo de Caña de Azúcar. Journal: Revista Iberoamericana de Automática e Informática Industrial RIAI 2009;6(3):10–20. ISSN 16977912 \let\@tempa\bibinfo@X@doi10.1016/S1697-7912(09). 70260–70262.
- [28] Hoyos E, López D, Alvarez H. A phenomenologically based material flow model for friction stir welding. Mater Des 2016;111:321–30. ISSN 0264–1275 \let\@tempa\bibinfo@X@doi10.1016/j.matdes.2016.09.009.
- [29] Basmadjian D, Farnood R. The Art of Modeling in Science and Engineering with Mathematics. second ed. 2006. ISBN 978–1584884606.
- [30] Dobre TG, Marciano JGS. Chemical Engineering: Modelling, Simulation and Similitude, WileyVCH. Alemania: Verlag GmbH & Co. KGaA; 2007. ISBN 9783527611096 \let\@tempa\bibinfo@X@doi10.1002/9783527611096.
- [31] K. Hangos, I. T. Cameron, Process Modelling and Model Analysis, vol. 4, Academic Press, ISBN 9780874216561, doi:\let\@tempa\bibinfo@X@doi10.1007/s13398-014-0173-7.2, 2001.
- [32] Luyben WL. Process modeling, simulation, and control for chemical engineers. 2 edn McGraw-Hill; 1990.
- [33] Marlin TE. Process control: designing process and systems for dynamic performance. second ed. USA: McGraw-Hill, Inc; 1995. ISBN 0-07-040491-7.
- [34] Woods RL, Lawrence KL. Modeling and Simulation of Dynamic Systems. first ed. USA: Pearson; 1997. ISBN 978–0133373790.
- [35] Ortega Quintana FA, Alvarez Zapata H, Botero Castro HA. Enfrentando el modelado de bioprocesos: una revisión de las metodologías de modelado. Facing bioprocess modeling: a review of the methodologies of modeling. Enfrentando a modelagem de bioprocesos: uma revisão de metodologias de modelagem. ION, Investigación, Optimización y Nuevos procesos en Ingeniería 2017;30(1):74 \let\@tempa\bibinfo@X@doi10.18273/revion.v30n1-2017006.
- [36] Muñoz-Tamayo R, de Groot J, Bakx E, Wierenga PA, Gruppen H, Zwietering MH, Sijtsma L. Hydrolysis of β -casein by the cell-envelope-located PI-type protease of *Lactococcus lactis*: A modelling approach. Int Dairy J 2011;21(10):755–62. ISSN 09586946 \let\@tempa\bibinfo@X@doi10.1016/j.idairyj.2011.03.012.
- [37] Walter E, Pronzalo L. Identification of parametric models from experimental data, Springer Berlin Heidelberg. 19973540761195. Paris, Milan, Barcelona.
- [38] Chin SV, Chappell MJ. Structural identifiability and indistinguishability analyses of the Minimal Model and a Euglycemic Hyperinsulinemic Clamp model for glucose-insulin dynamics. Comput Methods Progr Biomed 2011;104(2):120–34. ISSN 01692607 \let\@tempa\bibinfo@X@doi10.1016/j.cmpb.2010.08.012.
- [39] Muñoz-Tamayo R, Puillet L, Daniel JB, Sauvart D, Martin O, Taghipoor M, Blavy P. Review: To be or not to be an identifiable model. Is this a relevant question in animal science modelling? Animal 2018;12(4):701–12 <https://doi.org/10.1017/S1751731117002774>.
- [40] Pohjanpalo H. System identifiability based on the power series expansion of the solution. Mathematical Bioscience 1978;41(1):21–33. ISSN 0025–5564 [https://doi.org/10.1016/0025-5564\(78\).90063-90069](https://doi.org/10.1016/0025-5564(78).90063-90069).
- [41] O. T. Chis, J. R. Banga, E. Balsa-Canto, Structural identifiability of systems biology models: A critical comparison of methods, PLoS One 6(11), ISSN 19326203, doi:\let\@tempa\bibinfo@X@doi10.1371/journal.pone.0027755.
- [42] Contreras Montes J, Misa Llorca R, Urutera Vivanco L. Algoritmos para identificación de modelos difusos interpretables. IEEE Latin America Transactions 2007;5(5):346–51. ISSN 15480992 \let\@tempa\bibinfo@X@doi10.1109/TLA.2007.4378527.
- [43] Desmond-Le Quémener Elie, Bouchez Théodore. A thermodynamic theory of microbial growth. ISME J. Multidiscip. J. Microb. Ecol. 2014;8(8). <https://doi.org/10.1038/ismej.2014.7>.