processes

MDPI

Article

Modeling of Continuous PHA Production by a Hybrid Approach Based on First Principles and Machine Learning

Martin F. Luna [1], Andrea M. Ochsner [2], Véronique Amstutz [2], Damian von Blarer [3], Michael Sokolov [4], Paolo Arosio [1] and Manfred Zinn [2,*]

[1] Department of Chemistry and Applied Biosciences, ETH Zurich, 8093 Zurich, Switzerland; lunam@ethz.ch (M.F.L.); paolo.arosio@chem.ethz.ch (P.A.)
[2] Institute of Life Technologies, School of Engineering, University of Applied Sciences and Arts Western Switzerland, 1950 Sion, Switzerland; andrea.ochsner@hevs.ch (A.M.O.); veronique.amstutz@hevs.ch (V.A.)
[3] Infors AG, 4103 Bottmingen, Switzerland; d.vonblarer@infors-ht.com
[4] Data How AG, 8600 Dübendorf, Switzerland; m.sokolov@datahow.ch
* Correspondence: manfred.zinn@hevs.ch

Abstract: Polyhydroxyalkanoates (PHA) are renewable alternatives to traditional oil-derived polymers. PHA can be produced by different microorganisms in continuous culture under specific media composition, which makes the production process both promising and challenging. In order to achieve large productivities while maintaining high yield and efficiency, the continuous culture needs to be operated in the so-called dual nutrient limitation condition, where both the nitrogen and carbon sources are kept at very low concentrations. Mathematical models can greatly assist both design and operation of the bioprocess, but are challenged by the complexity of the system, in particular by the dual nutrient-limited growth phenomenon, where the cells undergo a metabolic shift that abruptly changes their behavior. Traditional, non-structured mechanistic models based on Monod uptake kinetics can be used to describe the bioreactor operation under specific process conditions. However, in the absence of a model description of the metabolic phenomena inside the cell, the extrapolation to a broader operation domain (e.g., different feeding concentrations and dilution rates) may present mismatches between the predictions and the actual process outcomes. Such detailed models may require almost perfect knowledge of the cell metabolism and omic-level measurements, hampering their development. On the other hand, purely data-driven models that learn correlations from experimental data do not require any prior knowledge of the process and are therefore unbiased and flexible. However, many more data are required for their development and their extrapolation ability is limited to conditions that are similar to the ones used for training. An attractive alternative is the combination of the extrapolation power of first principles knowledge with the flexibility of machine learning methods. This approach results in a hybrid model for the growth and uptake rates that can be used to predict the dynamic operation of the bioreactor. Here we develop a hybrid model to describe the continuous production of PHA by *Pseudomonas putida* GPo1 culture. After training, the model with experimental data gained under different dilution rates and medium compositions, we demonstrate how the model can describe the process in a wide range of operating conditions, including both single and dual nutrient-limited growth.

Keywords: artificial intelligence; bioprocess modelling; hybrid models; machine learning; PHA production

Citation: Luna, M.F.; Ochsner, A.M.; Amstutz, V.; von Blarer, D.; Sokolov, M.; Arosio, P.; Zinn, M. Modeling of Continuous PHA Production by a Hybrid Approach Based on First Principles and Machine Learning. *Processes* **2021**, *9*, 1560. https://doi.org/10.3390/pr9091560

Academic Editors: Giannis Penloglou and Alexandros Kiparissides

Received: 20 May 2021
Accepted: 15 July 2021
Published: 1 September 2021

1. Introduction

Plastic production has grown steadily in the last 70 years and is expected to increase even further as global population increases [1]. This puts a significant stress on the environment for several reasons: first, plastics are oil-derived polymers, which depend on a very extractive activity that is not only non-renewable but also potentially harmful to the environment [2]. Secondly, many traditional plastics are non-degradable under natural

conditions and accumulate in the environment [3]. These materials have to be either recycled (which requires not only energy but also a lot of workforce required for classification, processing and administration), burned for energy recycling (contributing to CO_2 increase in the atmosphere) or stored (which requires an increasing amount of space) [4]. As a result, plastics can be found polluting all sort of biomes, the oceans being a particularly concerning one [5].

Under this scenario, biopolymers have been proposed as a green alternative to traditional plastics: they are both renewable (can be produced from natural feedstocks, such as sugars and organic acids) and biodegradable (they can be decomposed by microorganisms) [6]. Polyhydroxyalkanoates (PHA) are biopolyesters produced by microorganisms that serve them as carbon and energy source [7]. Under certain (stressful) conditions, PHA is produced and stored inside the cells to be later recycled when needed. A typical case is nitrogen limitation, where a low concentration of this nutrient in the medium will induce PHA production. Thus, cells growing inside a bioreactor can be manipulated to produce biopolymers that, in turn, can be purified and processed to serve as replacement to synthetic plastics [8,9]. However, even though PHA production has been known for many years, very few industrial processes using this technology are in operation. The main reason for this is mostly economic: plastic production from oil and gas derivatives is cheaper than its green alternative. If oil and gas prices remain low, biopolymer production would remain challenging unless production costs are reduced. Even if their production is promoted using tax incentives, the increasing demand for plastics would require any bio-based production to have a high productivity.

During the last decades, great effort has been put to extend the field of Process System Engineering (PSE) to the bioprocess industry [10]. PSE is a field of research that aims to apply mathematical tools to solve scientific and engineering problems of complex production systems [11]. Design, optimization and control are key activities in any process operation that may be fundamental to achieve economic feasibility. In fact, the application of PSE is ubiquitous in the chemical and petrochemical industry. In the context of biopolymers production, PSE may be a valuable tool to achieve the same levels of productivity and production costs of their synthetic counterpart. However, biological processes are more complex than chemical ones due to several reasons: cell metabolism involves thousands of chemical reactions with internal regulation; there is a great deal of intrinsic variability in production runs and the number of available measurements is very limited compared to the number of variables involved. Thus, the implementation of PSE principles to biomanufacturing is considerably more challenging in comparison to the traditional chemical industry.

Important advances have been done in this direction [12,13]. Mathematical models, which lay at the core of this paradigm, have benefited a lot from the combination of several disciplines, ranging from biotechnology, chemical engineering and data science [14,15]. Mechanistic models that describe the metabolic behavior of the microorganism are available for several PHA producing strains [16–20]. Unfortunately, these models involve systems of equations with several parameters that may require almost complete knowledge of the metabolic network and complex analytical methods (e.g., omic measurement) to estimate them. Furthermore, the extrapolation from predictions about the cell metabolism to the behavior of the cells inside of a bioreactor may be challenging. Non-structured macroscopic models based only on the available measurements in the bioreactor medium (nutrients, biomass, PHA content, etc.) may be easier to develop and even more robust within the experimental region where they were derived [21–23]. However, finding a proper structure (i.e., equations representing the physicochemical and biological principles) that covers a wide range of operating conditions is hard to achieve. On the other hand, data-driven models do not need to assume any principles, since they are purely based on correlations found on data. Thus, they are not biased by prior knowledge [24]. However, the amount of data needed to build these models is usually larger and, more important, extrapolations of the models to previously unseen operating conditions is usually not accurate. In between

these approaches, hybrid models combine the fundamental knowledge of mechanistic models together with the flexibility of data-driven models [25]. The term "hybrid" is generally used with different meanings, and there is not a unique formulation of hybrid models but several versions. In this work, the type of hybrid models considered are the ones that include a traditional formulation of the mass balances of the several species, replacing one or more terms of the balances with machine learning models trained from experimental data [26]. This is a promising approach that has been around for several decades, but it gained momentum in the last years due to the developments in Artificial Intelligence (AI), Data Science, and computational power [27]. Hybrid models for PHA production with different structures have been tested with success for batch and fed-batch systems [28].

In this work, a hybrid model of a continuous bioreactor for the production of PHA is proposed. *Pseudomonas putida* strain GPo1 (ATCC 29347 formerly known as *Pseudomonas oleovorans* GPo1) efficiently converts octanoate into PHA under dual nutrient-limited growth conditions: both nitrogen and carbon are consumed up to limiting concentrations in the medium [9]. This operation mode is very important in order to achieve economic feasibility of the production process but is also very challenging from the modeling side because nutrient concentrations in the dual limitation region are difficult to measure, as they are present only in traces. This complicates the identification of the correct deterministic functions for growth and uptake rates in a wide range of concentrations and operation conditions. However, the mechanistic backbone of the hybrid model coupled with an AI method to learn the growth and uptake kinetics is successful in reproducing the bioreactor operation at several process conditions inside and outside the dual nutrient-limited growth regime.

2. Materials and Methods

2.1. Experimental Setup

Two different datasets available in the literature were used in this work to train and test the proposed hybrid model. The experimental setups are very similar: they consist of a continuous bioreactor with a feeding stream containing a concentrated carbon source (sodium octanoate) and a nitrogen source (ammonium sulfate) being part of the remaining medium components. The carbon to nitrogen ratio in the medium feed (C_{in}/N_{in}) is modulated by setting different flow rates of the two feed pumps. Concomitantly, a harvest stream, where culture broth is continuously removed together with residual nutrients, bacteria and PHA (see Figure 1) keeps the culture volume in the bioreactor constant. Samples were taken periodically, and pH, temperature, dissolved oxygen and volume were controlled and kept at reference values. The same strain, *P. putida* GPo1 ATCC 29347 is used in both cases.

The first dataset (DS1) was taken from a series of chemostat experiments [29]. Five different dilution rates (D = 0.05, 0.1, 0.2, 0.3 and 0.4 h^{-1}), were tested with several values of the feeding composition C_{in}/N_{in}, in order to cover three growth regimes: nitrogen limitation, carbon limitation, and dual limitation. The remaining experimental conditions were identical among all the runs. The working volume of the system was 2 L and it was operated as a chemostat. The temperature and pH setpoints were 30 °C and 7.1, respectively. For each dilution rate, different steady state conditions were established before samples were taken. These samples were collected in ice-cooled 50 mL Falcon tubes and centrifugated. The supernatant was frozen at −20 °C until further analysis, whereas the biomass pellet was freeze-dried for PHA analysis. The cell dry weight was determined with a filter method as described in the original publication [29]. This method retrieved the total biomass that included both the PHA-free biomass as well as the PHA mass. The cellular PHA content was determined by gas chromatography (GC) according to the method proposed by Lageveen et al. [30], and the difference between the total biomass and the PHA mass yields was the PHA-free (residual) biomass X_r. The supernatant was

analyzed to determine the concentrations of ammonium, with the indophenol method described by Scheiner [31], and octanoate using GC [29].

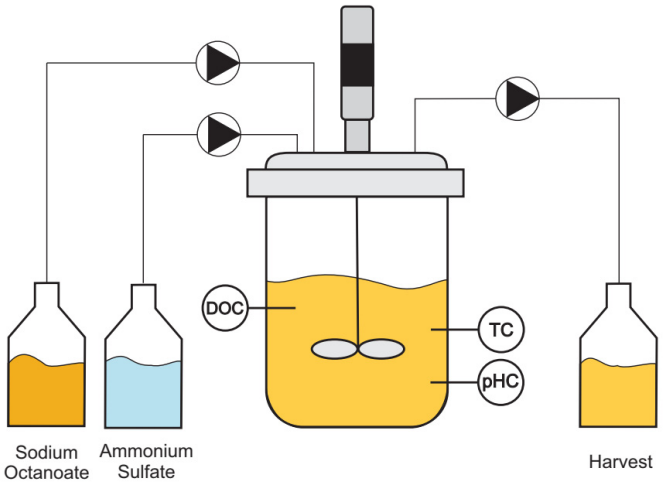

Figure 1. Scheme of the bioreactor setup used in both experimental datasets. Sodium octanoate and ammonium sulfate were used as the only carbon and nitrogen sources, respectively. A harvest stream removed culture broth to keep the liquid volume constant. Dissolved oxygen (DO), temperature (T), and pH were controlled at a setpoint.

In the second dataset (DS2) [32], the bioreactor was operated in transient mode: The dilution rate was fixed ($D = 0.3\ h^{-1}$) and the C_{in}/N_{in} ratio of the feeding stream was changed over time without letting the culture achieve steady state. No other experimental conditions were changed between experiments. The working volume of this bioreactor was 1.5 L, the temperature setpoint was 30 °C and the pH setpoint was 7.0 (0.1 units below DS1). The sampling method and determinations were similar to the ones in DS1, but ammonium and octanoate were measured using on-line analyzers based on those techniques.

Details about the setups, protocols and methods can be found in the original publications [29,32].

2.2. Process Model

Four main species concentrations are used to describe the state of the system: PHA-free biomass of bacteria X_r, poly(3-hydroxyalkanoate) P, carbon C, and nitrogen N (as the equivalent amount of carbon and nitrogen being part of the fed nutrients). The concentrations are expressed in g/L unless stated otherwise. As it was previously stated, the bioreactor can be operated in different regimes depending on the system state and process conditions

A simplified reaction network for PHA production by *P. putida* GPo1 using ammonia and octanoate as nutrients is presented in Figure 2. The cell takes up carbon and nitrogen from the culture broth in order to produce more X_r and PHA. Nitrogen uptake q_n is directly related to biomass growth while carbon can undergo three different pathways, each one represented by an uptake rate: growth q_g, accumulation (PHA formation) q_c and maintenance q_m. Finally, the PHA production rate q_p is proportional to the carbon accumulation rate q_c. Based on this network, the mass balances for the 4 species are formulated in Equations (1)–(4):

$$\frac{dX_r}{dt} = (r_g - D)\ X_r \tag{1}$$

$$\frac{dC}{dt} = (C_{in} - C)\ D - \left(q_C\ f_n + q_m + q_g\right) X_r \tag{2}$$

$$\frac{dN}{dt} = (N_{in} - N)\ D - q_N\ X_r \tag{3}$$

$$\frac{dP}{dt} = q_p \, f_n \, X_r - P \, D \tag{4}$$

While some assumptions about the reaction network had to be made to formulate the mass balances, the mathematical expression for the growth, uptake and production rates are yet to be defined. A traditional Monod-type formalism can be used as a first approximation:

$$r_g = \mu_g \frac{C}{C + K_{C1}} \frac{N}{N + K_N} \tag{5}$$

$$q_C = \mu_C \frac{C}{C + K_{C2}} \tag{6}$$

The specific growth rates μ_g and μ_C are the maximal values used for the X_r and PHA formation, respectively. A differentiation was made between the Monod constants for growth (K_{C1}) and PHA formation (K_{C2}), whereas K_N is used only for the nitrogen affinity of PHA-free biomass. C and N represent the carbon and nitrogen concentrations in the culture broth, respectively. The term f_n accounts for the inhibition of the accumulation and PHA production in the presence of nitrogen in the culture broth. The PHA inhibition constant k_{fn} represents the threshold concentration of nitrogen required to trigger PHA accumulation:

$$f_n = \frac{k_{fn}}{N + k_{fn}} \tag{7}$$

The remaining rates can be calculated using the stoichiometric biomass formation yields $Y_{x/c}$ and $Y_{x/n}$, the PHA formation yield $Y_{pha/c}$, as well as the maintenance coefficient (m_c) as follows:

$$q_g = \frac{r_g}{Y_{x/c}} \tag{8}$$

$$q_N = \frac{r_g}{Y_{x/n}} \tag{9}$$

$$q_p = q_C \, Y_{pha/c} \tag{10}$$

$$q_m = m_C \tag{11}$$

The maintenance rate q_m was considered a constant, and the PHA formation yield $Y_{pha/c}$ was split into two limiting values, $Y^g_{pha/c}$ and $Y^{lim}_{pha/c}$, in order to give q_p enough flexibility to fit the small PHA production present during the growth phase:

$$Y_{pha/c} = Y^g_{pha/c} + \left(Y^g_{pha/c} - Y^{lim}_{pha/c} \right) f_n \tag{12}$$

The system of equations presented in Equations (1)–(12) includes a set θ of 11 parameters that need to be fitted from experimental data. The fitting problem involves the minimization of the Mean Square Error (MSE) function, presented in Equation (12):

$$MSE = \sum_n \sum_j \sum_i \left(\frac{y^{exp}_{i,j,n} - y^{mod}_{i,j,n \ (\theta)}}{\sigma_{j,n}} \right)^2 \tag{13}$$

where $y^{exp}_{i,j,n}$ is the experimental measurement of species j at time t_i in experiment n, while $y^{mod}_{i,j,n}$ is the corresponding model prediction that depends on the model parametrization. The standard deviation $\sigma_{j,n}$ for the experimental measurements in each experiment is used for normalization.

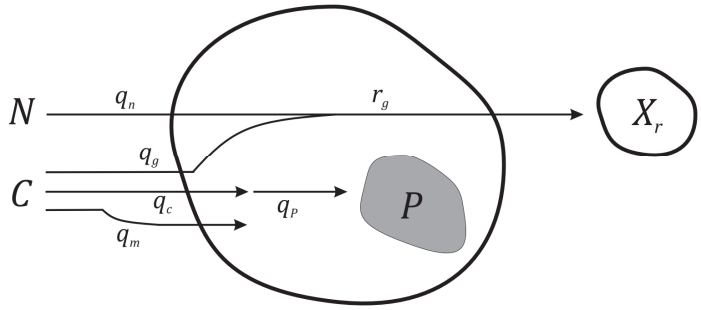

Figure 2. Simplified reaction network for biomass growth and polyhydroxyalkanoates (PHA) production of *P. putida* GPo1. Abbreviations (large caps mean concentrations, in g/L, in culture broth and small caps uptake and reaction rates): N: available nitrogen, C: available carbon, X_r: PHA-free (residual) biomass, P: PHA in culture broth, q_n: nitrogen uptake, q_g: carbon uptake for growth of X_r, r_g: specific growth rate of PHA-free biomass, q_c: carbon uptake for PHA formation, q_m: carbon uptake for maintenance energy, q_p: PHA production.

2.3. Hybrid Model Algorithm

Even when Monod-type kinetic models are good approximations for bioreactor operation in certain conditions, these simple representations do not include the cell's inner regulatory mechanisms and its complex metabolic reactions. Thus, the model presented in Section 2.2. is not complex enough to capture the behavior of the system under a wide range of process conditions. Different approaches can be taken, e.g., using a detailed model for the growth and uptake rates. This would require more knowledge about the metabolic network of the system and how the cells interact with the environment. A more pragmatic approach is the hybrid modeling approach. Instead of assuming explicit expressions for the growth, carbon uptake, and production rates, neural networks are able to derive them from data:

$$r_g = NN_1 \tag{14}$$

$$q_C = NN_2 \tag{15}$$

$$q_p = NN_3 \tag{16}$$

Here, NN_k is the k-th output of a single neural network. The mass balances and the remaining rates are equal to the ones previously presented in Section 2.2. Thus, the hybrid model can be formulated as:

$$\frac{dX_r}{dt} = (NN_1 - D)\, X_r \tag{17}$$

$$\frac{dC}{dt} = (C_{in} - C)\, D - \left(NN_2 \frac{k_{fn}}{N + k_{fn}} + m_S + \frac{NN_1}{Y_C} \right) X \tag{18}$$

$$\frac{dN}{dt} = (N_{in} - N)\, D - \frac{NN_1}{Y_N}\, X_r \tag{19}$$

$$\frac{dP}{dt} = NN_3 \frac{k_{fn}}{N + k_{fn}}\, X_r - P\, D \tag{20}$$

This model includes a multi-output neural network as well as some constant model parameters. The training of the neural network is not straightforward, as the inputs and outputs are not easily available from data. The values of C and N during dual nutrient limitation are below the detection limit of the analytical methods. Furthermore, without these values, the rates cannot be calculated directly using the mass balances. An alternative method like the one presented in Narayanan et al. [33] can be used to calculate the rates indirectly, training the neural network's weights and biases explicitly (i.e., by minimizing Equation (13)). However, the number of parameters increases rapidly with the number of

layers and neurons, and the method can be computationally too intense and thus expensive if a big architecture is required. An alternative is presented here.

As it was stated in Section 2.2., the mechanistic model is not able to fit the whole range of experimental conditions with a unique set of parameters θ (the Monod-type expressions are simple approximations that do not consider complex metabolic phenomena that may arise under a wide range of nutrient concentrations). However, the model performs well when fittings are made only for individual experiments. The regression of that experiment returns not only the specific set of parameters but also the calculated values for the concentrations of all the species, even during dual nutrient limitation. By doing so, the growth, uptake, and production rates can be calculated using Equations (5), (6) and (9) with the specific set of parameters θ_n and the calculated concentrations of the nutrients. The remaining parameters can be fitted from the complete dataset. The algorithm is presented in Table 1.

Table 1. Hybrid model algorithm.

- Inputs: Data set DS, Dilution rates D, Monod-type kinetic mechanistic model
- Fit the mechanistic model to the complete dataset DS to get the nominal set of parameters θ_0
- For each experiment with a given dilution rate D:
  - Re-fit the model parameters related to the growth, uptake, and production rates (use nominal values from θ_0 for the remaining parameters) to get a specific set of parameters θ_n
  - Simulate the experiment, obtaining the nutrient concentrations C and N
  - Calculate r_g, q_c, and q_p using Equations (5), (6) and (10) with the specific set of parameters θ_n and the simulated values of C and N
  - Compile all the values [D, C, N, r_g, q_c, q_p]
- Train a Neural Network (NN) that maps [D, C, N] to [r_g, q_c, q_p]
- Output: Hybrid model, Nominal parameters θ_0, Specific parameters θ_n

Since hybrid models rely on learning from experimental data, usually they require a larger dataset than purely mechanistic models. When integrating the differential equations, the system reaches a state for which no data in the training set do exist and, therefore, the neural network may return considerable errors in the predicted rates, which will, in turn, have an effect on the mathematical integration step. If the system deviates further from the expected behavior, it may reach new unseen conditions, continuing in a feedback loop. This is usually not the case for mechanistic models, where the structures of the kinetic expressions ensure a stable integration under any process conditions. The performance of the hybrid model can be significantly improved by data augmentation. Artificial data can be created using the mechanistic model to fill the state space for which there is no experimental data available. This way, if the system reaches one of these states during integration, the neural network will make its predictions based on the artificial data, making the process more robust. The algorithm presented in Table 2 shows how to combine the individual models for the existing experiments in order to create artificial data for augmentation of the existing dataset.

Table 2. Model-based augmentation algorithm.

- Inputs: Dataset DS, Dilution rates D, Monod-type kinetic mechanistic model
- Fit the mechanistic model to the complete dataset DS to get the nominal set of parameters θ_0
- For each experiment with a given dilution rate D:
  - Re-fit the model parameters related to the growth, uptake and production rates (use nominal values from θ_0 for the remaining parameters) to get a specific set of parameters θ_n
- For any new dilution rate D^* not contained in DS:
  - Interpolate the specific parameters between dilution rates:

$$\frac{(\theta_{n+1}-\theta_{n-1})}{D_{n+1}-D_{n-1}}(D^* - D_{n-1}) \; + \; \theta_{n-1}$$

  - Simulate the experiment using the interpolated parameters θ_{n*} and process conditions D^*, N_{in}, and C_{in}
  - Augment DS with the simulated values of t, X_r, C, N, P and process conditions D^*, N_{in}, and C_{in}
- Output: Augmented Dataset DS*

Here, the augmentation of the dataset was performed by simulating new experiments with different dilution rates to mimic the original experimental design: the dilution rate was kept constant while the C_{in}/N_{in} ratio changed. A different augmentation approach can be used to generate more artificial data, as long as it covers the state space efficiently.

2.4. Dual Limitation Region Modeling

The dual limitation region is usually presented in the literature as C_{in}/N_{in} vs. D plots, indicating the combination of variables that, after achieving steady state, will reach dual nutrient limitation [34]. In principle, it can be described by both types of mathematical models. It is worth noting that the values of both N and C never reach zero (otherwise, the growth rates would have a value of zero as well, and there would be a wash out of the bioreactor), so the limiting values that describe the boundary of the growth regimes have to be fixed arbitrarily at a small residual value. If the culture results in values for C and N equal or below both of these limiting values, then culture conditions can be considered as being part of the dual nutrient-limited growth regime.

For the case of a continuous culture operated under steady state (chemostat) conditions, the left terms of Equations (1)–(3) would be equal to zero. For carbon limited growth, the carbon concentration would reach C_{lim} and by consideration of the feed concentration of nitrogen N_{in}, as well as Equations (1)–(3), (7)–(9) and (11) the following expressions are derived:

$$r_g = D \tag{21}$$

$$(C_{in} - C_{lim})\, D = \left(q_C \frac{k_{fn}}{N + k_{fn}} + m_C + \frac{D}{Y_{x/c}} \right) X_r \tag{22}$$

$$(N_{in} - N)\, Y_{x/n} = X_r \tag{23}$$

From there, the carbon limitation regime can be defined in terms boundary C_{in}/N_{in} values of the process conditions:

$$\frac{C_{in}}{N_{in}} = \frac{C_{lim}}{N_{in}} + \left(q_C \frac{k_{fn}}{N + k_{fn}} + m_S + \frac{D}{Y_{x/c}} \right) \frac{(N_{in} - N) Y_{x/n}}{N_{in}\, D} \tag{24}$$

To solve Equation (23), Equation (20) should be solved first to obtain the value of N, fixing the carbon concentration as C_{lim}. So far, no expressions have been given for r_g and q_c. Either Equations (5) and (6) can be chosen for the mechanistic model, or Equations (14) and (15) for the hybrid model. In the case of the mechanistic model an explicit expression for Equation (21) can be found, but for the hybrid it has to be solved implicitly.

The same procedure can be applied to the boundary of the nitrogen-limited growth regime:

$$\frac{C_{in}}{N_{in}} = \frac{C}{N_{in}} + \left(q_C \frac{k_{fn}}{N_{lim} + k_{fn}} + m_S + \frac{D}{Y_{x/c}} \right) \frac{(N_{in} - N_{lim}) Y_{x/n}}{N_{in}\, D} \tag{25}$$

In this case, solving Equation (21) yields the value of C, fixing the nitrogen concentration as N_{lim}. Again, both models can be used for the growth and uptake rates.

Finally, it is worth noting that the value of N_{in} is assumed as constant in Equations (24) and (25). This is done in concordance with the bibliography, where these curves are shown as two-dimensional plots for experiments where N_{in} is constant always at the same value. As can be seen from the equations, the relationship of the ratio C_{in}/N_{in} with the dilution rate D depends on the nitrogen concentration in the feed N_{in}.

2.5. Model Implementation

All models and algorithms were implemented in Matlab 2020b in an Intel Core i5-9500 CPU. The system of differential equations was solved using function "*ode15s*". Mechanistic models were fitted by minimizing Equation (11) with function "*patternsearch*", initializing

multiple times from different starting points to avoid local minima (also known as a multiple shooting method). The Neural Networks were fitted using function "*feedforwardnet*".

3. Results and Discussion

Following the method proposed in Section 2, the first dataset DS1 was used to fit the mechanistic model nominal parameters θ_0. The experimental data for all experiments together with the mechanistic model predictions are presented in Figure 3. As can be seen, the model performs well for most of the runs and species, but it struggles to get the carbon concentration close to the carbon-limited growth regime. This is crucial to describe dual nutrient limitation, which is one of the purposes of the model due to its importance for an optimal bioreactor operation.

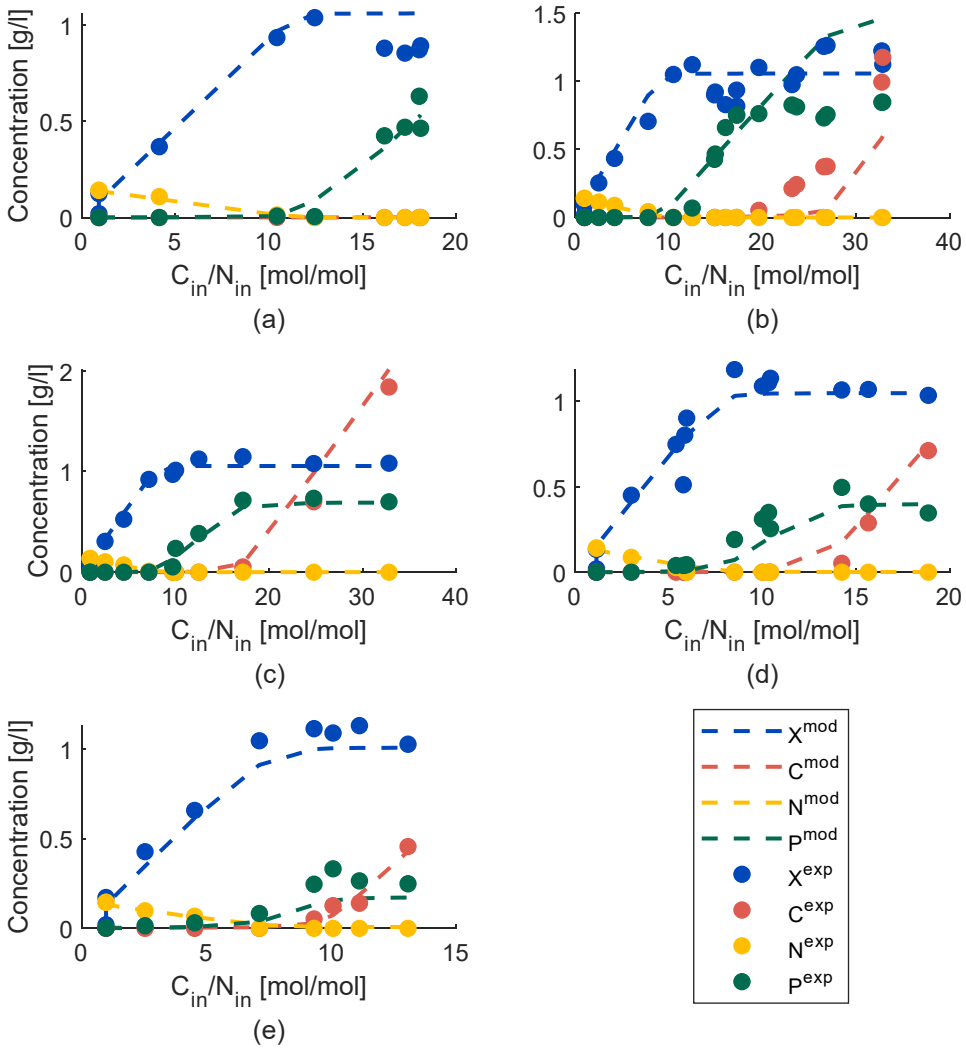

Figure 3. Mechanistic model predictions and experimental data for the five experiments of DS1. *P. putida* GPo1 cultures achieved steady states at different C_{in}/N_{in} ratios. Each experiment corresponds to a set dilution rate: (**a**) $D = 0.05\ h^{-1}$ (**b**) $D = 0.1\ h^{-1}$ (**c**) $D = 0.2\ h^{-1}$ (**d**) $D = 0.3\ h^{-1}$ (**e**) $D = 0.4\ h^{-1}$.

The nominal model parameters together with an estimation of their 95% confidence intervals are presented in Table 3. The parameter distributions are not particularly widespread. However, there are certain experimental conditions that the model fails to simulate. This is likely due to limitation in the flexibility of the model, i.e., it is a problem of the model structure rather than a problem of the accuracy of the parameter estimation.

Table 3. Nominal model parameters and their estimated 95% confidence intervals.

Parameter	Value	CI	Units
μ_g	4.499×10^{-1}	$[4.483; 4.700] \times 10^{-1}$	h^{-1}
μ_C	2.413×10^{-1}	$[2.175; 3007] \times 10^{-1}$	h^{-1}
K_{C1}	4.278×10^{-4}	$[3.864; 4.675] \times 10^{-4}$	g/L
K_{C2}	6.382×10^{-3}	$[0.578; 6.948] \times 10^{-2}$	g/L
K_N	9.372×10^{-4}	$[0.848; 1.024] \times 10^{-3}$	g/L
m_C	2.536×10^{-2}	$[2.300; 2.770] \times 10^{-2}$	h^{-1}
k_{fn}	9.656×10^{-3}	$[0.877; 1.000] \times 10^{-2}$	g/L
$Y_{x/c}$	1.183	[1.102; 1.335]	g/g
$Y_{x/n}$	7.074	[6.501; 7.097]	g/g
$Y^{g}_{pha/c}$	3.992×10^{-1}	$[3.671; 4.000] \times 10^{-1}$	g/g
$Y^{lim}_{pha/c}$	6.022×10^{-1}	$[4.355; 6.027] \times 10^{-1}$	g/g

Next, the hybrid model approach was applied to the same dataset. Each individual experiment was fitted and the unique set of parameters θ_n was obtained. The model-based augmentation algorithm (presented in Table 2) was run to obtain artificial datapoints for dilution rates ranging from 0.025 h^{-1} to 0.445 h^{-1}, with a step increase of 0.001 h^{-1}. The augmented dataset was then used to train the hybrid model with the algorithm presented in Table 1. The selected neural network was a feedforward net with 4 layers of 5 nodes. The dataset split for training/validation/test was 90/5/5%. The features of the network were C, N and D, with outputs being r_g, q_c and q_p. The hybrid model predictions are shown in Figure 4. It clearly outperforms the mechanistic model. Especially, it can be seen that the carbon concentration predictions in the supernatant follow the real data much closer in the case of the hybrid model. This can be seen particularly clear in the second experiment (D = 0.1 h^{-1}), presented in Figures 3b and 4b.

A comparison of both models is presented in Table 4. The absolute root mean squared error (RMSE) and the R^2 coefficient are presented for each species and each model. It can be seen that while for X_r and N the performances are similar, for C and PHA the Hybrid model has smaller RMSE and R^2 closer to 1, which indicates a better performance, in agreement with the visual comparison of the fittings in Figures 3 and 4.

The performance of the hybrid model was then tested with an independent dataset, DS2. None of the experiments in this dataset was used to fit the model, so it could be used for external validation. Instead of operating at steady state, the feeding profiles in these sets of experiments change dynamically in time. Furthermore, the nitrogen concentration in the feed was kept at 0.15 g/L, the same as in DS1, but it changes in two of the three experiments of DS2. Despite these differences in the bioreactor operation, the hybrid model manages to predict the outcome of the experiments accurately. The predictions of the model and the experimental data are plotted in Figure 5.

Even though the model performed well in general, there were some discrepancies between the predictions and the experimental data. In experiments 2 and 3 (Figure 5b,c), there was an offset between the predicted and the measured free biomass. The reasons may be related to the data used for training the model: The experiments were performed under a unique concentration of nitrogen in the feed. Since the PHA-free biomass concentration strongly depends on the nitrogen fed to the bioreactor (as can be seen for example in Equation (23)), the lack of different feeding conditions in the training set may have hampered the extrapolation capabilities of the model, especially for biomass.

Both the mechanistic and the hybrid model were then used to describe the dual nutrient limitation regime in a C_{in}/N_{in} vs. D plot with the method described in Section 2.4. The values of N_{in}, N_{lim} and C_{lim} had to be fixed arbitrarily. They were chosen to be N_{in} = 0.15 g/L, N_{lim} = 0.02 g/L and C_{lim} = 0.03 g/L in order to be comparable with results shown in [34] and the analytical methods used therein. The predictions of the mechanistic and the hybrid models for DS1 are shown in Figure 6a,b, respectively. The dual nutrient limitation regime is presented as the gray area, while the black dots represent the experimental conditions

where dual limitation starts and terminates. The mechanistic model predicted smoother curves, with the expected shape for a Monod-type kinetic model. However, it did not perform well for some of the dilution rates (especially $D = 0.1\ h^{-1}$), predicting a wider dual *C*, *N* limitation regime. The hybrid model, on the other hand, performed very well for all the dilution rates, as could be expected since one of its main objectives was to accurately fit the carbon and nitrogen concentrations. However, it is worth noting that the shape of the model for unseen conditions (those that have not been tested experimentally) was found to not be as smooth as with the mechanistic model. This effect was mainly due to the imputation algorithm used to generate artificial data, which linearly combines different parametrizations to simulate the additional experimental conditions. If the amount of data for different conditions increases, the hybrid model will improve its performance, resembling more and more the experimental results. This is an important feature of the hybrid model: it is flexible enough to add new data by retraining the neural network, which would correct the predictions in the regions near the new experiments. The mechanistic model, on the other hand, will have to find a compromise between all the experiments with a unique set of parameters that is already having troubles predicting the existing ones.

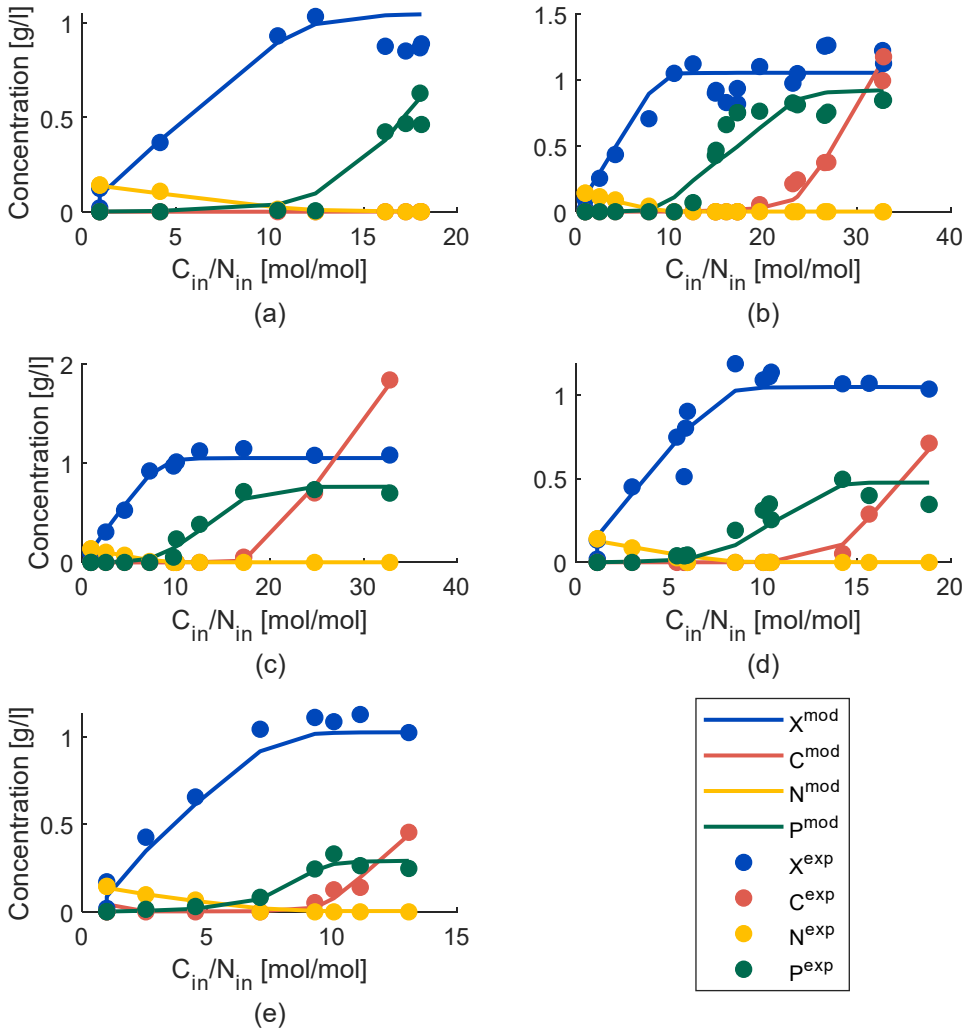

Figure 4. Hybrid model predictions and experimental data for the five experiments of DS1. Each experiment corresponds to a constant dilution rate: (**a**) $D = 0.05\ h^{-1}$ (**b**) $D = 0.1\ h^{-1}$ (**c**) $D = 0.2\ h^{-1}$ (**d**) $D = 0.3\ h^{-1}$ (**e**) $D = 0.4\ h^{-1}$.

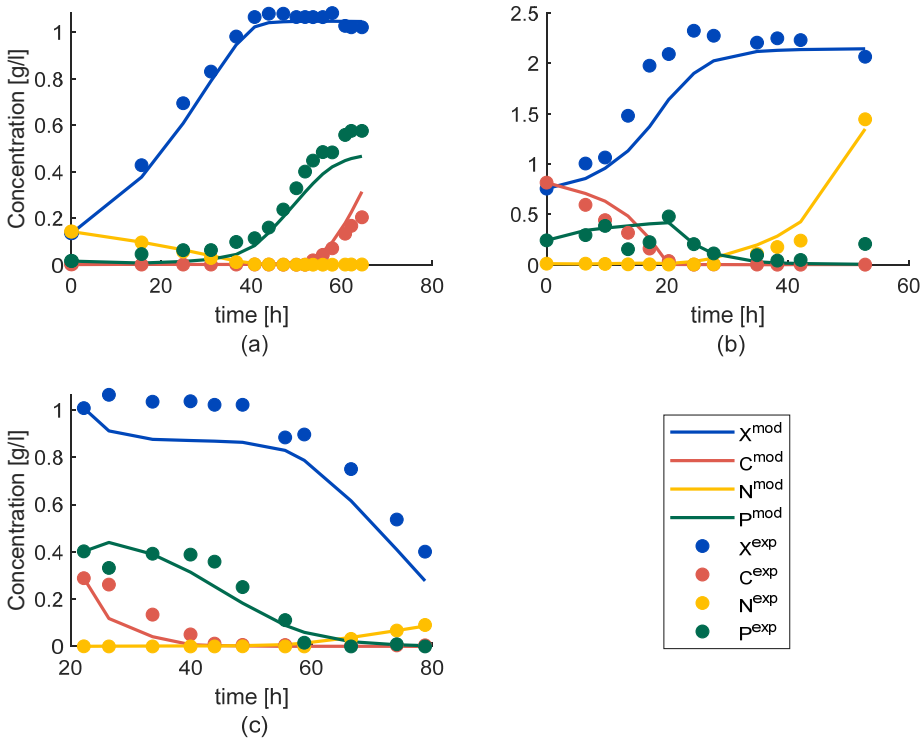

Figure 5. Hybrid model predictions and experimental data for the three transient experiments of DS2. All experiments operated at $D = 0.3\ h^{-1}$ but each one, (**a**) Experiment 1, (**b**) Experiment 2, and (**c**) Experiment 3, had a different C_{in}/N_{in} dynamic profile (see the original publication for details [32]).

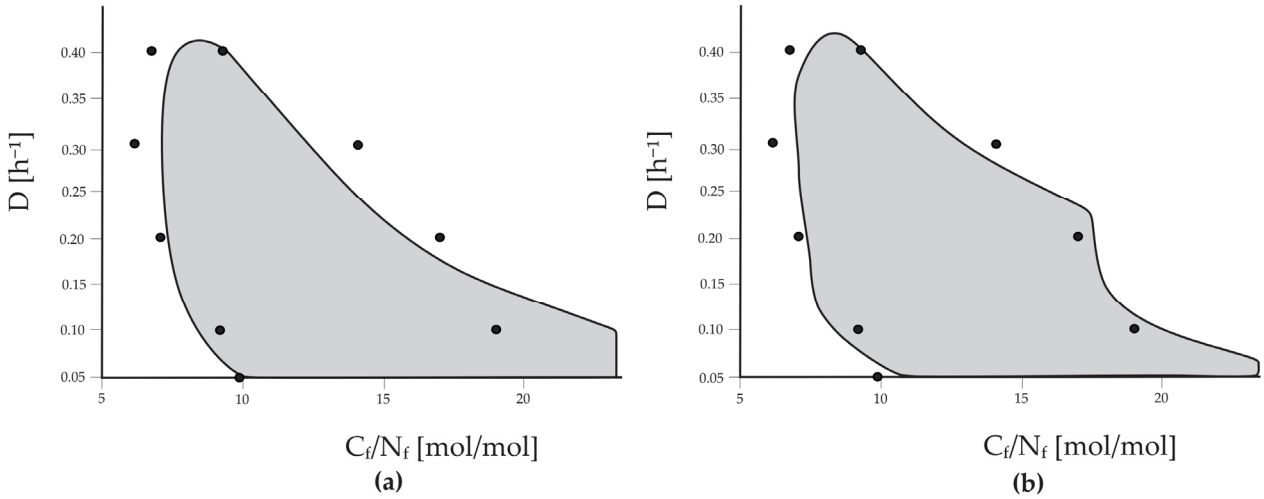

Figure 6. Dual nutrient limitation region (in gray) calculated using the (**a**) mechanistic model and (**b**) the hybrid model, together with experimental data estimated from DS1 (depicted with •).

Table 4. Nominal model parameters and their confidence intervals.

	Mechanistic Model		**Hybrid Model**	
	RMSE [g/L]	**R^2**	**RMSE [g/L]**	**R^2**
X_r	0.1060	0.9255	0.1037	0.9288
C	0.1243	0.8520	0.0367	0.9871
N	0.0080	0.9790	0.0076	0.9810
PHA	0.1737	0.6628	0.0857	0.9179

4. Conclusions

A hybrid model was proposed for the operation of a bioreactor containing the PHA producing *P. putida* strain GPo1. The model uses a neural network to account for the growth, uptake, and production rates. Since these quantities cannot be measured directly, Monod-type kinetic expressions are used as support to calculate them. Once the neural network is trained, it is used together with mass balances for the main species in order to describe the dynamics of the process under a wide range of operating conditions. The hybrid model performed very well, especially in describing the carbon and nitrogen concentrations, which is key for the description of the dual nutrient-limited growth regime.

The hybrid model presented in this work includes several interesting features and advantages over traditional mechanistic models. The flexibility of the neural network together with the robustness provided by first principles (mass balances and stoichiometric transformations) allows the hybrid model to fit experiments with different process conditions without overfitting, as is shown by the good predictions in the external dataset. Fundamental knowledge about the system can be embedded through the kinetic models used to calculate the rates in the individual experiments (then generalized by the neural network). Furthermore, the assistance of the individual mechanistic models also allows for the calculation of hidden states that cannot be easily measured, like the nutrient concentrations for the dual-limited growth regime.

The hybrid model used here is focused on the dependency of the growth and uptake rates with regards to the feeding conditions and nutrient concentrations. However, different process variables like temperature, pH, or other medium components can be added to the neural networks to study their impact on growth and PHA production (given that the experimental data contains enough variations in their values). Furthermore, the influence of the type of carbon source (e.g., amount of C atoms per molecule) on the polymer structure or properties can be built into the model, provided that the data is available. Of course, more complex models will be more challenging to develop, but probably the production of the experimental data will be more limiting than the modeling effort.

Finally, it is worth mentioning that other machine learning models can be used instead of Neural Network to learn the rates from data. Neural networks present some very useful features (they are very flexible, allow for multi-output responses, are easy to retrain with new data) and are perhaps the most popular method used in hybrid modeling. However, advances are being done in hybrid models with other methods (such and Gaussian Process or Supported Vector Machines) that may present interesting alternatives that should be investigated.

Process System Engineering principles and methods are expected to play a bigger role in bioprocessing in the near future. Mathematical models like the ones presented in this work are important tools that will help to render environmentally friendly technologies more productive, robust, and economically sustainable.

Author Contributions: Methodology: M.F.L.; validation: A.M.O. and V.A.; resources: D.v.B.; formal analysis: M.S.; supervision: P.A. and M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Swiss Innovation Agency—Innosuisse under the grant program Impulse (grant number IP-LS 37635.1).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Geyer, R.; Jambeck, J.R.; Law, K.L. Production, use, and fate of all plastics ever made. *Sci. Adv.* **2017**, *3*, 25–29. [CrossRef]
2. Heidbreder, L.M.; Bablok, I.; Drews, S.; Menzel, C. Tackling the plastic problem: A review on perceptions, behaviors, and interventions. *Sci. Total Environ.* **2019**, *668*, 1077–1093. [CrossRef]
3. Hanik, N.; Amstutz, V.; Zinn, M. Microplastics-From anthropogenic to natural. *Chimia* **2019**, *73*, 841–843. [CrossRef]
4. Dijkstra, H.; van Beukering, P.; Brouwer, R. Business models and sustainable plastic management: A systematic review of the literature. *J. Clean. Prod.* **2020**, *258*, 120967. [CrossRef]

5. Derraik, J.G.B. The pollution of the marine environment by plastic debris: A review. *Mar. Pollut. Bull.* **2002**, *44*, 842–852. [CrossRef]
6. Pietrini, M.; Roes, L.; Patel, M.K.; Chiellini, E. Comparative life cycle studies on poly(3-hydroxybutyrate)-based composites as potential replacement for conventional petrochemical plastics. *Biomacromolecules* **2007**, *8*, 2210–2218. [CrossRef] [PubMed]
7. Mozejko-Ciesielska, J.; Szacherska, K.; Marciniak, P. *Pseudomonas* species as producers of eco-friendly polyhydroxyalkanoates. *J. Polym. Environ.* **2019**, *27*, 1151–1166. [CrossRef]
8. Masood, F. *Polyhydroxyalkanoates in the Food Packaging Industry*; Elsevier Inc.: Amsterdam, The Netherlands, 2017; ISBN 9780128119433.
9. Amstutz, V.; Hanik, N.; Pott, J.; Utsunomia, C.; Zinn, M. Tailored biosynthesis of polyhydroxyalkanoates in chemostat cultures. *Methods Enzymol.* **2019**, *627*, 99–123.
10. Koutinas, M.; Kiparissides, A.; Pistikopoulos, E.N.; Mantalaris, A. Bioprocess systems engineering: Transferring traditional process engineering principles to industrial biotechnology. *Comput. Struct. Biotechnol. J.* **2013**, *3*, e201210022. [CrossRef] [PubMed]
11. Pistikopoulos, E.N.; Barbosa-Povoa, A.; Lee, J.H.; Misener, R.; Mitsos, A.; Reklaitis, G.V.; Venkatasubramanian, V.; You, F.; Gani, R. Process systems engineering–The generation next? *Comput. Chem. Eng.* **2021**, *147*, 107252. [CrossRef]
12. Narayanan, H.; Luna, M.F.; von Stosch, M.; Cruz Bournazou, M.N.; Polotti, G.; Morbidelli, M.; Butté, A.; Sokolov, M. Bioprocessing in the digital age: The role of process models. *Biotechnol. J.* **2020**, *15*, 1900172. [CrossRef]
13. Martínez, E.C.; Cristaldi, M.D.; Grau, R.J. Dynamic optimization of bioreactors using probabilistic tendency models and Bayesian active learning. *Comput. Chem. Eng.* **2013**, *49*, 37–49. [CrossRef]
14. Novak, M.; Koller, M.; Braunegg, G.; Horvat, P. Mathematical modelling as a tool for optimized PHA production. *Chem. Biochem. Eng. Q.* **2015**, *29*, 183–220. [CrossRef]
15. Zinn, M.; Witholt, B.; Egli, T. Dual nutrient limited growth: Models, experimental observations, and applications. *J. Biotechnol.* **2004**, *113*, 263–279. [CrossRef]
16. Gadkar, K.G.; Doyle, F.J.; Crowley, T.J.; Varner, J.D. Cybernetic model predictive control of a continuous bioreactor with cell recycle. *Biotechnol. Prog.* **2003**, *19*, 1487–1497. [CrossRef]
17. Franz, A.; Song, H.S.; Ramkrishna, D.; Kienle, A. Experimental and theoretical analysis of poly(β-hydroxybutyrate) formation and consumption in *Ralstonia eutropha*. *Biochem. Eng. J.* **2011**, *55*, 49–58. [CrossRef]
18. Duvigneau, S.; Dürr, R.; Carius, L.; Kienle, A. Hybrid cybernetic modeling of polyhydroxyalkanoate Production in *Cupriavidus necator* using fructose and acetate as substrates. *IFAC PapersOnLine* **2020**, *53*, 16872–16877. [CrossRef]
19. Riascos, C.A.M.; Gombert, A.K.; Silva, L.F.; Taciro, M.K.; Gomez, J.G.C.; Le Roux, G.A.C. Metabolic pathways analysis in PHAs production by *Pseudomonas* with ^{13}C-labeling experiments. *Comput. Aided Chem. Eng.* **2013**, *32*, 121–126.
20. Dias, J.M.L.; Oehmen, A.; Serafim, L.S.; Lemos, P.C.; Reis, M.A.M.; Oliveira, R. Metabolic modelling of polyhydroxyalkanoate copolymers production by mixed microbial cultures. *BMC Syst. Biol.* **2008**, *2*, 59. [CrossRef] [PubMed]
21. Beyenal, H.; Chen, S.N.; Lewandowski, Z. The double substrate growth kinetics of *Pseudomonas aeruginosa*. *Enzym. Microb. Technol.* **2003**, *32*, 92–98. [CrossRef]
22. Annuar, M.S.M.; Tan, I.K.P.; Ibrahim, S.; Ramachandran, K.B. A kinetic model for growth and biosynthesis of medium-chain-length poly-(3-hydroxyalkanoates) in *Pseudomonas putida*. *Braz. J. Chem. Eng.* **2008**, *25*, 217–228. [CrossRef]
23. Gumel, A.M. Growth kinetics, effect of carbon substrate in biosynthesis of mcl-PHA by *Pseudomonas putida* Bet001. *Braz. J. Microbiol.* **2016**, *438*, 427–438.
24. Patnaik, P.R. Neural network designs for poly-β-hydroxybutyrate production optimization under simulated industrial conditions. *Biotechnol. Lett.* **2005**, *27*, 409–415. [CrossRef] [PubMed]
25. Feyo De Azevedo, S.; Dahm, B.; Oliveira, F.R. Hybrid modelling of biochemical processes: A comparison with the conventional approach. *Comput. Chem. Eng.* **1997**, *21*, S751–S756. [CrossRef]
26. Von Stosch, M.; Oliveira, R.; Peres, J.; Feyo de Azevedo, S. Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Comput. Chem. Eng.* **2014**, *60*, 86–101. [CrossRef]
27. Venkatasubramanian, V. The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE J.* **2019**, *65*, 466–478. [CrossRef]
28. Lopes Dias, J.M.; Lemos, P.; Serafim, L.; Oehmen, A.; Reis, M.A.M.; Oliveira, R. Development and implementation of a nonparametric/metabolic model in the process optimisation of PHA production by mixed microbial cultures. *Comput. Aided Chem. Eng.* **2007**, *24*, 995–1000.
29. Durner, R.; Witholt, B.; Egli, T. Accumulation of poly[(R)-3-hydroxyalkanoates] in *Pseudomonas oleovorans* during growth with octanoate in continuous culture at different dilution rates. *Appl. Environ. Microbiol.* **2000**, *66*, 3408–3414. [CrossRef] [PubMed]
30. Lageveen, R.G.; Huisman, G.W.; Preusting, H.; Ketelaar, P.; Eggink, G.; Witholt, B. Formation of polyesters by *Pseudomonas oleovorans*: Effect of substrates on formation and composition of poly-(R)-3-hydroxyalkanoates and poly-(R)-3-hydroxyalkenoates. *Appl. Environ. Microbiol.* **1988**, *54*, 2924–2932. [CrossRef] [PubMed]
31. Scheiner, D. Determination of ammonia and Kjeldahl nitrogen by indophenol method. *Water Res.* **1976**, *10*, 31–36. [CrossRef]
32. Zinn, M.; Durner, R.; Zinn, H.; Ren, Q.; Egli, T.; Witholt, B. Growth and accumulation dynamics of poly(3-hydroxyalkanoate) (PHA) in *Pseudomonas putida* GPo1 cultivated in continuous culture under transient feed conditions. *Biotechnol. J.* **2011**, *6*, 1240–1252. [CrossRef] [PubMed]

33. Narayanan, H.; Sokolov, M.; Morbidelli, M.; Butté, A. A new generation of predictive models–the added value of hybrid models for manufacturing processes of therapeutic proteins. *Biotechnol. Bioeng.* **2019**, *116*, 2540–2549. [CrossRef] [PubMed]
34. Egli, T.; Zinn, M. The concept of multiple-nutrient-limited growth of microorganisms and its application in biotechnological processes. *Biotechnol. Adv.* **2003**, *22*, 35–43. [CrossRef] [PubMed]