

1 **TITLE**

2 The effects of nondimensionalization in Physics-Informed Neural Networks
3 applied to the Kinetics of Biological Reactions: A case study with *Lactobacillus casei*

4
5 **AUTHOR LIST**

6 Renan Mendes Frota^a, Ivanildo José da Silva Junior^{a,*}, Amaro Gomes Barreto Junior^b.

7
8 ^a *Universidade Federal do Ceará, Departamento de Engenharia Química, Campus*
9 *do Pici, Bloco 709, 60455-760 Fortaleza, CE, Brazil*

10

11 ^b *Chemical and Biochemical Engineering Processes (EPQB), School of Chemistry—*
12 *EQ/UFRJ, Universidade Federal do Rio de Janeiro (UFRJ), Cidade Universitária, Ilha*
13 *do Fundão, Rio de Janeiro 21949-900, Brazil*

14

15 *Corresponding author.

16 Email address: ivanildo@ufc.br

17 Full postal address: Universidade Federal do Ceará, Departamento de Engenharia
18 Química, Campus do Pici, Bloco 709, 60455-760 Fortaleza, CE, Brazil

19 **KEYWORDS**

20 Physics-Informed Neural Network

21 Nondimensionalization

22 Simulation

23 Lactic Acid

24 Batch reactor

25

26 **ABSTRACT**

27 Physics-Informed Neural Network (PINN) is a relatively new technique
28 capable of simulating physics-based systems with Neural Networks (NN) using
29 mathematical models equations to generate the necessary quantity of data for NN
30 optimization. Bioreactors pose a challenge for PINNs because of their intrinsic
31 complexity, and thus strategies that improve loss reduction and solution convergence
32 may be employed. The optimal layer size, loss weights and nondimensionalization
33 factors were determined for a batch reactor production of lactic acid by *Lactobacillus*
34 *casei*. The nondimensionalization of time increased loss in all cases. PINNs were
35 able to reproduce with great fidelity numerical results and experimental data in
36 constant volume (batch reactor, with loss $< 10^{-4}$), but showed poor performance in
37 models with volume variation (CSTR and fed-batch). CSTR and fed-batch
38 performance improved when X_M (inhibitory biomass concentration) was used as a
39 nondimensionalization scaler of biomass and reactor maximum volume as a
40 nondimensionalization scaler of volume. Biomass concentration and reactor liquid
41 content volume were the most relevant variables to reduce loss values.

42 1 INTRODUCTION

43

44 Lactic acid (LA) is a molecule of great industrial and economical interest. LA
45 and its derivatives are used in pharmaceutical, cosmetics and food industries [1].
46 Since it is an organic molecule, occurs naturally is and perceived as “green” or
47 environmentally friendly by many consumers [2], its perceived value comes not only
48 from the molecule chemical or technological features, but also from a marketing
49 standing point. Another point that drove interest in LA growing since the beginning of
50 the 21th century was its application as a raw material for the production of PLA (poly-
51 lactic acid), an environmentally friendly alternative to plastics and that has found
52 many applications [3,4]. The industry of Lactic Acid will grow to 160 kt by 2025 [5],
53 with a revenue forecast of almost USD 9 billion [6].

54 LA has two optical active forms, which can vary greatly in application. The
55 L(+) form is preferred for some cosmetical, food and pharmaceutical applications
56 since D(-)-Lactic Acid can be a harmful enantiomer ([7]). The chemical synthesis
57 yields a racemic mixture [8]. The fermentative production is usually carried out by
58 LAB (Lactic Acid Bacteria) [9] and many bacteria can produce mainly one of the two
59 enantiomers depending on the process conditions. The optimum conditions for the
60 production of LA by many LAB include slightly acid pH (between 4.5 and 7), the use
61 of a nitrogen source such as yeast extract and a temperature between 27 and 40°C
62 [10,11]. Carbohydrates such as glucose and lactose are the most frequent used
63 substrate in many studies.

64 Batch fermentation is the usual method used to the production of LA in
65 industries around the world. Since many LAB growth and LA production is heavily
66 influenced by lactic acid, substrate and biomass concentrations (due to inhibitory
67 effects), batch mode may be less efficient than continuous or fed-batch reactors [1].

68 The simulation of bioreactions by classical numerical methods is often used
69 and show appropriate predictions and errors. However, computer numerical methods
70 can be very sensitive to the discretization method used and can require considerable
71 time and expertise to implement. Because of this, is not infrequent the use of third-
72 part and/or proprietary software such as Matlab, Aspen Plus or COMSOL
73 Multiphysics [12,13]. Depending on the number of parameters necessary to be
74 estimated, the experiment data production cost be considerably high and the
75 complexity of the model can make the simulation almost prohibitive [14].

76 Artificial Neural Networks (ANN) are universal approximators and can be used
77 to represent many of these models, requiring in some cases less parameters.
78 However, they are computationally and data intensive, thus requiring a large amount
79 of data to be properly trained [15]. One advantage of ANN is that, while they require
80 time to be trained (from seconds to hours), their use is simple and have a relatively
81 low computational cost.

82 Physics-Informed Neural Network (PINN) is an approach to employ deep
83 neural networks capability of universal function approximators to solve complex
84 numerical problems, such as stochastic and high order partial differential equations
85 (PDEs) [16–18]. The technique can be applied to many cases where the numeric
86 solution is highly complex, but also for simpler cases where it provides a robust
87 framework for simulating system with well defined mathematical models. One of the

greatest advantages of PINNs is the possibility to use the mathematical models themselves as a data source instead of raw experimental data, so it is possible to work with *small data*. This is an important advantage because obtaining experimental data is very time-consuming and expensive, and models usually need a great number of experimental points to be validated [19]. Biological reactions kinetics pose a natural challenge for mathematical solution because, while many models may have a relatively simple mathematical description, it often involves multiple derivatives referencing each other, limiting the methods of solving it, or even requiring greater computer processing power and more sophisticated integration methods. Therefore, it is necessary to evaluate PINNs performance and optimization since they were shown to be capable of producing accurate simulation models but in some studies were not able not simulate appropriately one or more variables [13,14].

The great economical and scientific interest in LA is demonstrated in Academia, with many studies conducted on these kind of microorganisms and their LA production using different carbon sources or reactor operation mode were evaluated [10,20,20–24] [10,11,20–24]. Thus, the large amount of scientific literature available and relevance made Lactic Acid production by *Lactobacillus casei* [10] the ideal case study for the application of Physics-Informed Neural Networks in biological reactions and reactors simulation.

This work aims to determine the relevance of the deepness (number of hidden layers) or wideness (number of neurons per layer) and the influence of nondimensionalization and loss weights in the PINN simulation of bioreactions. For this, we use as a case study the production of Lactic Acid by *Lactobacillus casei* as

111 described in the literature [10] in CSTR (Continuous Stirred Tank), batch and fed-
112 batch reactors.

113 2 MATERIALS AND METHODS

114 2.1 Nondimensionalization

115 Nondimensionalization is a process in which variables are converted to
116 nondimensional variables using nondimensional factors, in this work called
117 nondimensional scalers. These scalers can substantially improve or worsen the error
118 of the system of equations being solved, thus are an important for giving insights on
119 how to improve the model [25]. The reduced dimension equation may need less
120 computational resources, produce better results or be unable to attain a feasible
121 solution. Since PINNs can be very sensitive to the value of the derivatives they are
122 optimizing, the equations were nondimensionalized to evaluate the procedure impact
123 on final results

124 We define each variable, represented by N , as an nondimensional variable.
125 The variable is equals to the nondimensional variable multiplied by a coefficient:

126

$$N = N_A * N_S \quad (1)$$

127

128 where N is the variable itself, N_A is the nondimensional variable and N_S is the
129 nondimensional scaler. The use of the subscript A represents the nondimensional
130 variable version of N , and S represents the nondimensional scaler of N .

131

132 2.2 Reaction Kinetics Mathematical Model

133 The model of the production of Lactic Acid by *Lactobacillus casei* using whey
134 lactose was proposed and validated by [10] using experimental data of a batch
135 reactor. The second experiment variant (with starting lactose concentration 21.4 g/L
136 and duration of 9 hours) is used in this work. *L. casei* production of Lactic Acid is
137 known to be controlled by inhibitory effects and can be represented by the classical
138 Monod equation adjusted to represent product inhibition effect (represented by $(1 -$
139 $P/P_M)$) and biomass inhibition effect (represented by $(1 - X/X_M)$), where X is the
140 biomass concentration, P is the product concentration, P_M is the inhibitory product
141 concentration, and X_M is the inhibitory biomass concentration. The value of all
142 parameters mentioned are available in the original paper. Thus, the biomass reaction
143 rate model includes product, biomass and substrate inhibitory effects. The biomass
144 concentration derivative over time is given by:

$$145 \quad r_X = \frac{\mu_{max} S}{K_S + S} X \left(1 - \frac{X}{X_M}\right)^f \left(1 - \frac{P}{P_M}\right)^h \quad (2)$$

146
147 where r_X is the reaction rate of biomass, μ_{max} is the maximum possible growth rate, S
148 is the substrate (whey lactose) concentration, K_S is the Monod constant and f and h
149 are factors that represent the toxicity or inhibitory potential adjustments.

150 The product formation rate is given by Luedeking-Piret kinetics, depending on
151 the growth or decrease of biomass linearly (represented by dX/dt) and also on the
152 biomass concentration itself:

$$r_P = \alpha r_X + \beta X \quad (3)$$

153

154 where α is the growth-associated product formation coefficient and β is the non-
155 growth associated product formation coefficient.

156

The substrate consumption kinetics is given by a relationship including the
157 substrate converted to product and the substrate used for biomass maintenance:

$$r_s = \frac{-1}{Y_{PS}} r_p - m_s X \quad (4)$$

158

where Y_{PS} is the product yield coefficient and m_s is the maintenance coefficient.

159

160

Equations 2 - 4 can be nondimensionalized as:

$$\frac{X_S}{t_S} r_{X_A} = \frac{\mu_{max} S_A S_S}{K_S + S_A S_S} X \left(1 - \frac{X_A X_S}{X_M} \right)^f \left(1 - \frac{P_A P_S}{P_M} \right)^h \quad (5)$$

161

$$\frac{P_S}{t_S} r_{P_A} = \alpha \frac{X_S}{t_S} r_{X_A} + \beta X_A X_S \quad (6)$$

162

$$\frac{S_S}{t_S} r_{S_A} = \frac{-1}{Y_{PS}} \frac{P_S}{t_S} r_{P_A} - m_s X_A X_S \quad (7)$$

163

164 2.3 Reactor Mathematical Model

165

The volume of liquid in the reactor is given by:

$$\frac{dV}{dt} = f_{in} - f_{out} \quad (8)$$

166

167

Which can be nondimensionalized as:

$$\frac{V_S}{t_S} \frac{dV_A}{dt_A} = f_{in} - f_{out} \quad (9)$$

168

169 The nondimensional concentration of a generic substance or biomass, N, can
170 be given by:

$$V_A V_S \frac{N_S}{t_S} \frac{dN_A}{dt_A} = V_A V_S \frac{N_S}{t_S} r_{N_A} + f_{in} N_{in} - f_{out} N_A N_S \quad (10)$$

171

172 where N is the concentration of each substance or biomass inside the reactor, V is
173 the volume of liquid inside the reactor (L), t is the time (h), f_{in} is the inlet flow rate (L/h)
174 in the reactor, N_{in} is the concentration of N in the inlet flow, f_{out} is the outlet flow rate
175 (L/h) of the reactor, and N_{out} is the concentration of N in the outlet flow. The subscripts
176 S and A were defined in section 2.1.

177

178 This model can be used for representing both batch, fed-batch and CSTR
179 models. For the batch case, both f_{in} and f_{out} are zero. For the fed-batch case, only f_{out}
180 is zero. For CSTR, f_{in} and f_{out} are greater than zero and $f_{in} = f_{out}$ at steady state.

181 **2.4 Simulation using Physics Informed Neural networks**

182 Physics-Informed Neural Network (PINN) was introduced to the scientific
183 community in 2019 [17]. One of the most important points of the technique is to allow
184 scientists to simulate physical, chemical or biological systems using differential
185 equations as a source of data source for the optimization of Neural Networks (NN)
186 created specifically to solve those problems. This is possible because NN are

187 intrinsically universal approximators [18]. The error or deviation from the correct
 188 values can be evaluated using a loss function. The overall loss function is defined as
 189 the weighted sum of the L^2 residuals [26]:

$$L = \sum w_N \left\| \left(\left. \frac{dN}{dt} \right|_{\text{calculated}} - \left. \frac{dN}{dt} \right|_{\text{predicted}} \right) \right\|^2 \quad (11)$$

190 where N represents X, S, P or V, and w_N is the weight of loss for N.

191

192 To test many different PINN configurations, a custom Grid Search and
 193 repetition loop were created. The PINN evaluation (without the repetition loop)
 194 scheme adopted is shown in Figure 1. Both hyperparameters and equations
 195 parameters are fixed before each loop. Then, the Neural Network is fed with the data
 196 from the equation system. The outputs (labeled as OUT in Figure 1) are the variables
 197 of interest (X, P, S and V) and their respective derivatives according to time. Then,
 198 the loss function is calculated comparing the given derivatives at step “n” compared
 199 with what their actual value should be. This can be achieved simply by making the
 200 difference (subtraction) of the derivative calculated (as show in equations 5-10) and
 201 the ones predicted by the NN.

202 In the original paper that introduced PINN [17] questions such as how deep
 203 and wide the neural network must be to appropriately represent the mathematical
 204 models, the relevance of normalization and/or nondimensionalization and how the
 205 weights of loss functions can impact on the model loss and accuracy were raised.
 206 Because this subject still being debated and more studies need to be carried out [18],
 207 these points will be evaluated in the next sections. We compare traditional numeric
 208 methods (Euler Forward Method), experimental [10] and PINN generated solutions
 209 for the simulation of biological reactions, in specific the production of LA by *L. casei*,

210 and strategies to improve performance and reduce loss in different reactors
211 regiments (batch, fed-batch and CSTR). Therefore, is out of the scope of this work to
212 discuss meticulously the implementation of PINNs.

213

214 **2.5 Simulation configurations**

215 The main questions to be answered are:

- 216 • Is it possible to approximate simple, ordinary differential equations (ODEs),
217 with PINNs using lower numbers of neurons and layers? The simple Burgers
218 equation required at least 20 neurons per layer and a few layers in another
219 study [26];
- 220 • Do the weights of the loss function of each variable impact significantly on the
221 accuracy and number of steps required for acceptable results?
- 222 • What impacts does the independent variable (time) nondimensionalization
223 cause to the loss?
- 224 • How does the dependent variables (X, P, S, V) nondimensionalization affect
225 the system? Are some of them more relevant than the others?
- 226 • Since the same mathematical model of reactor is being used for the three
227 cases (batch, fed-batch and CSTR), can the best configuration (group of
228 parameters and hyperparameters, such as number of neurons per layer and
229 nondimensional scalers) for the batch reactor also represent accurately the
230 other two?

- Is it possible to find a pattern of relationship between hyperparameters, loss weights and nondimensionalization scalers in order to improve the traditional Random Grid Search, replacing it with a more rational and predictable approach?

Each reactor operation mode starts in a different state. Batch reactor simulation starts with 5 L of solution and Fed-batch and CSTR starts with 1 L, all with initial concentrations of Biomass, LA and whey lactose as the same used in the experiment (case 2 of [10]). CSTR is simulated on a greater time period (96 h) while batch and fed-batch are both simulated for 10.6 h (slightly above the experimental time of 9 h). The batch reactor, by definition, has closed inlets and outlets. The fed-batch reactor volume is allowed to increase indefinitely. CSTR starts in transient state and is expected to converge to stationary state after some time, with its volume being regulated by a function in order to open the outlet as the volume capacity (5L) is approached. The inlet flow of fed-batch is set to 2L/h, while the inlet flow of CSTR is set to 1 L/h. The outlet function for the CSTR is given by:

$$f_{in} * (V/V_{max})^7 \quad (12)$$

where V_{max} is the maximum liquid volume in the reactor, set as the same of the maximum volume of the reactor used in the original experiment (5 L).

All simulations were run with all weights of loss function (w_v, w_x, w_p, w_s) = 1, all nondimensionalization scalers (t_s, X_s, P_s, S_s, V_s) = 1, number of training points = 800, number of testing points = 1000, layer size = 22 neurons in 3 layers, epochs (iterations of Adam algorithm [27]) = 30,000, and learning rate = 1×10^{-3} , except when

254 explicitly stated otherwise. The PINN model was solved using DeepXDE [26], and
255 hyperbolic tangent activation was used to to enhance convergence speed.

256 3 RESULTS AND DISCUSSION

257 In order to try to find patterns between many of the hyperparameters and
258 process parameters, tests were conducted in parallel instead of conducting a
259 traditional Random Grid Search, where a myriad of hyperparameters is set, every
260 possible combination is iterated and the result with the smallest total loss chosen.
261 Table 1 summaries the tests. Table 2 shows which parameters were tested in each
262 test and case.

263 Nondimensionalization effects can greatly improve CSTR and fed-batch
264 performance, but easily made batch reactor simulation yield greater losses. The
265 variation of t_s did not only not improve the simulation of batch reactor error, but also
266 resulted in greater loss (Figure 2 - a). From all the values tested for the independent
267 variable (t) nondimensionalization, the best t_s was found to be one. In other words,
268 the best model is found to be the one which does not nondimensionalize the time.
269 The nondimensionalization of time, as a rule, increase the loss of the model and, in
270 many cases, made the NN stagnate at points next to loss = 1. Thus, the
271 nondimensionalization of time was ruled out as noneffective. It must be noted,
272 however, that the 6 cases tested for time were greater or equal 1. Since the loss
273 increased with the increase of t_s , values of $t_s \leq 1$ may improve model performance in
274 systems with multiple dependency (P and S depend directly on X , and X , P and S
275 depend on V), such as the presented.

276 The nondimensionalization (test "non_dim") of X , P , S and V was conducted in
277 sequence, with 6 different combinations. The layer size was tested as 5 layers with
278 32 neurons and 3 layers with 22 neurons. CSTR loss was reduced greatly using the

279 smaller layer. The best cases for CSTR were n1 and n6. The best cases for fed-batch
280 was n4, with loss $< 10^{-2}$. Despite of that, none of these models were capable of
281 representing adequately any of the reactors, even with more than 80,000 epochs.

282 The layer size test was performed solely for the batch reactor case and is
283 show in 2 - b. There were tested 12 different layer deepness and wideness. With a
284 constant number of neurons of 22, experiments t_lay7, t_lay8 and t_lay9 varied only
285 the number of layers to 3, 4 and 5, respectively. While there is a great improvement
286 from 3 to 4 layers, the loss of 5 layers is about the same, for 22 more neurons to
287 optimize. We can infer that, while increasing the number of layers may improve loss
288 reduction, it can also need many more epochs to optimize, reducing the loss for a
289 same given number of epochs. Experiments t_lay4, t_lay5 and t_lay6 were done for
290 3, 4 and 8 layers respectively, and show that when using smaller number of neurons
291 more deepness (number o layers) is necessary to reduce loss and get out of local
292 minimums. From t_lay 1 to 3, we can infer that ratios of neuron/layers (N/L) < 1 can
293 yield better results than ratios > 1 for this range. From t_lay 4 to 6, we can infer that
294 an $N/L > 4$ can yield better results increasing the number of layers, with N/L varying
295 from 5.33 to 2.67. From t_lay 7 to 9, the loss reduces as N/L goes from 7.33 to 4.4.
296 The loss is also smaller comparing t_lay11 ($N/L = 10.67$) to t_lay12 ($N/L = 6.4$). While
297 there is a general trend in this data that $N/L = 3$ or less can generate better models, it
298 was not possible to determine a general rule for the given system of equations. The
299 best performance (smaller loss) was achieved in t_lay12, the model with more
300 neurons per layer (32 neurons in 5 layers, $N/L = 6.4$).

301 Loss weights tests were performed in different layer sizes (6 layers with 8
302 neurons, 3 layers with 22 neurons, 5 layers with 32 neurons, and 3 layers with 90

neurons) (data not published). It is clear that the simulation for the batch reactor showed a good performance for any group of parameters, with the exception of W3, where $w_p = 3$. Fed-batch is the model with the worst performance in almost every group of weights, but performs better at W2, with $w_x = 3$. Surprisingly, fed-batch loss was smaller in the deeper and narrower layer size (8 neurons in 6 layers), with an error of 8.03×10^{-1} after 120,000 epochs. Batch performed better in the most equilibrated layer size (32 neurons in 5 layers), with an error of 1.37×10^{-5} after 45,000 epochs. The only configuration where CSTR performed well (loss $< 10^{-2}$) was also the only one where it got smaller losses than the other two reactors, with layer size of 90 neurons in 3 layers and 120,000 epochs. This can't be determined as an optimal point, though, since both batch performed slightly worse than the other variations and fed-batch loss was greater than 10. This indicates that the CSTR model needs more effort to be solved, probably due to one more variable (f_{out}) that the others don't have, as their outlet is closed. The batch and fed-batch models, though, not only benefit from simpler models because of less computational cost, but also can achieve slower losses in fewer epochs in simpler Neural Networks. Using 5 layers with 32 neurons in combination with case W3 was the only configuration where CSTR loss was smaller than the one of the batch simulation. Excluding this case, almost every combination of layer deepness and wideness and weight factors resulted in acceptable results for batch (loss $< 10^{-3}$), a greater loss for CSTR and an even greater loss for fed-batch.

For validation of the model proposed, the system of equations was solved for the batch reactor case using the traditional Euler Method (with 240 discretization points in time) and the result is plotted, alongside PINN (60,000 epochs and layer size of 32 neurons and 5 layers) and experimental data (from [10]) in Figure 2 - c. We

327 can conclude that the given PINN configuration can represent with great accuracy
328 the biological reaction. The graphic doesn't answer, however, if the same model is
329 capable of solve both the reaction and the volume variation, since in batch $dV/dt = 0$.
330 The data (not published) show that the most probable culprit to the deviation when
331 simulating CSTR and Fed-batch models is the volume. Despite having one inlet and
332 one outlet, the CSTR eventually approaches a constant internal liquid volume, while
333 fed-batch increases infinitely. In weight test, it was observed a trend where the
334 existence of volume variation made the solution very unstable and of difficulty
335 convergence. Since X depends on V and P and S depend on both X and V , they
336 were significantly affected when the model could not predict appropriately the values
337 of V and dV/dt . X , P and S values always present high losses when the volume
338 predict loss is too high. Thus, it was expected that a greater w_v would make the
339 solver more sensitive to errors in the volume variable and, consequently, be able to
340 produce more precise results. Since this could not be validated in the loss weight
341 tests, it is possible that this specific problem requires a greater amount of neurons
342 (>90) and layers (>5) than the ones that were tested.

343 None of the combination of above configurations (layer size, number o
344 neurons per layer, number of iterations/epochs, loss weights or nondimensional
345 scalers) was capable of representing the fed-batch and CSTR reactors appropriately.
346 A final test was tried, applying alongside the Adam algorithm, L-BFGS [28] as a
347 previous and post-processing step, combined with loss weights and
348 nondimensionalization found to be the best fit on previous attempts. The result ins
349 show in Figure 3 - a. The name of each subplot is constructed as "Name of the
350 reactor : Name of the variable". So "Batch : V " for instance, represents the volume in

351 the batch reactor over time. While the batch reactor model presents a persistent
352 fluctuation of volume prediction in comparison to the correct value, it is irrelevant
353 since more than 3 orders of magnitude smaller than the total volume. In another
354 study with multiple equations, PINNs behaved similarly, being able to predict
355 accurately two variables while one showed noise or considerable deviation [29]. The
356 CSTR model results are alike the ones for many other configurations and the same
357 reactor: only one of the four variables values could be more or less acceptably
358 predicted, while the others are clearly deviated. Fed-batch has the greatest loss of
359 the three models in almost any configuration. Also, both fed-batch and CSTR errors
360 are clearly stagnated in values of order of magnitude of about 10^0 .

361 Another relevant point is that there are expressions raised to exponents in the
362 inhibitory terms of equation 5. Ideally X and P would never be higher than X_m and P_m ,
363 respectively, but this can't be assumed from a Neural Network that is learning and
364 updating itself based on errors. Because of this, fluctuation of errors on volume or
365 biomass (which were determined to be the most sensible features, probably because
366 they are directly involved in the equations of P and S) can rapidly destabilize the
367 system, predict values of X and P that are greater than their respective maximum
368 values and generate negative numbers raised to non integer numbers (h and f
369 parameters of equation 5) that generate errors that can not be computed or are
370 infinite. Many tests were prematurely interrupted, since the error was infinite, and the
371 system was incapable of recover from these deviations. While this may induce us to
372 think that the losses values being too high were entirely related to the biomass and
373 product inhibitory effects, the batch model results (Figures 2 - c and 3 - a) prove that
374 they are not the cause – at least not the single cause. The volume presents some

375 fluctuations even though X, P and S are appropriately in accordance with the
376 reference (solution obtained using Euler Method) in the batch reactor simulation,
377 while all variables show a great deviation from reference in CSTR and fed-batch.
378 Therefore, we conclude that the main challenges faced by the system while trying to
379 minimize the loss function were the volumetric balance over time, cells mass balance
380 over time, or both. This is confirmed by the reduced loss of CSTR and fed-batch
381 when using nondimensionalization factors for volume and biomass. However, the
382 mere use of loss weights and nondimensionalization were not effective enough to
383 reduce the loss to acceptable levels ($<10^{-3}$) and enable the same set of
384 hyperparameters to represent batch, fed-batch and CSTR reactor models.

385 While PINNs are very interesting and open the doors of Machine Learning to
386 many applications in fields of science that don't have a multitude of data available,
387 we believe a better approach would be to use PINNs to model only the reaction given
388 a starting point, and to solve the physics-involved variables, which are much more
389 straightforward, by a traditional numeric method. The system would be, then, a hybrid
390 of PINN and traditional numeric integrator or solver system. We also could not find a
391 clear relation between hyperparameters, nondimensionalization and loss weights that
392 could serve as a guide for future works and substitute or rationalize the Grid Search
393 other than the loss increasing with t_s increasing.

394 4 CONCLUSIONS

395 Simpler Neural Networks, with less neurons and layers, were capable of
396 producing acceptable losses for the reaction kinetics (batch reactor) only, but
397 performed poorly on balancing the reactions and the volume variation simultaneously.
398 The optimum performance and lowest loss of the batch simulation was achieved with
399 32 neurons in 5 layers, 30,000 epochs and no nondimensionalization strategy. Almost
400 all configurations produced high losses when dV/dt was not expected to be zero. The
401 nondimensionalization with $X_S=X_M$ and $V_S=V_{\text{reactor}}$ in these cases was capable of
402 reducing the error in more than one order of magnitude in some cases, but the total
403 loss was still much superior than the one of the batch reactor ($<10^{-3}$). Since the PINN
404 performance for the batch reactor was exceptionally good and both CSTR and fed-
405 batch models required a large amount of epochs to yield a still high loss, in cases
406 where both the volume of the reactor is varying considerably and there are multiple
407 reactions that reference each other, a better approach may be combine the PINN for
408 solving the biological or chemical set of equations (group of reactions), which is
409 intrinsically more complex, and use a regular numeric method, like Euler or Runge-
410 Kutta, to solve the physics part of the system (volumetric and mass balance) which is
411 usually more well-known and established.

412 Physics-Informed Neural Networks were shown to represent with adequate
413 accuracy biological reactions with complex inhibition terms and multiple derivative
414 dependency. However, all combinations tested failed to represent adequately both
415 the concentrations of X, S, and P and liquid volume variation in the reactor. We
416 recommend future works to try employing nondimensionalization scalars for time (t_s)

417 of values less than 1, even if the proposed t_s must be arbitrary and not based on
418 actual constants from the equations. To explore new strategies to deal with these
419 kind of physical and biochemical complex systems and to propose another ways to
420 rationalize the process of creating a PINN model, we also recommend paying special
421 attention to the independent (in this case, t_s) and quasi-independent (V) variables,
422 since they can influence all other variables and be responsible for the success of
423 failure of the model.

424 **5 ACKNOWLEDGMENTS**

425 The authors thank CAPES (Coordenadoria de Aperfeiçoamento de Pessoal de
426 Nível Superior) for the financial support provided and an award of a scholarship.

- [1] A. Nancib, N. Nancib, A. Boubendir, J. Boudrant, The use of date waste for lactic acid production by a fed-batch culture using *Lactobacillus casei* subsp. *rhamnosus*, *Braz. J. Microbiol.* 46 (2015) 893–902. <https://doi.org/10.1590/S1517-838246320131067>.
- [2] P.M. de Oliveira, L.P. Santos, L.F. Coelho, P.M. Avila Neto, D.C. Sass, J. Contiero, Production of L (+) Lactic Acid by *Lactobacillus casei* Ke11: Fed Batch Fermentation Strategies, *Fermentation*. 7 (2021) 151. <https://doi.org/10.3390/fermentation7030151>.
- [3] A. Komesu, M.R.W. Maciel, R.M. Filho, Lactic Acid Production to Purification: A Review, (2017) 20.
- [4] Z. Li, S. Ding, Z. Li, T. Tan, L-Lactic acid production by *Lactobacillus casei* fermentation with corn steep liquor-supplemented acid-hydrolysate of soybean meal, *Biotechnol. J.* 1 (2006) 1453–1458. <https://doi.org/10.1002/biot.200600099>.
- [5] J.P. López-Gómez, M. Alexandri, R. Schneider, J. Venus, A review on the current developments in continuous lactic acid fermentations and case studies utilising inexpensive raw materials, *Process Biochemistry*. 79 (2019) 1–10. <https://doi.org/10.1016/j.procbio.2018.12.012>.
- [6] N.A.S. Din, S.J. Lim, M.Y. Maskat, S.A. Mutalib, N.A.M. Zaini, Lactic acid separation and recovery from fermentation broth by ion-exchange resin: A review, *Bioresour. Bioprocess.* 8 (2021) 31. <https://doi.org/10.1186/s40643-021-00384-4>.
- [7] M. Pohanka, D-Lactic Acid as a Metabolite: Toxicology, Diagnosis, and Detection, *BioMed Research International*. 2020 (2020) 1–9. <https://doi.org/10.1155/2020/3419034>.
- [8] A.O. Büyükkileci, S. Harsa, Batch production of L(+) lactic acid from whey by *Lactobacillus casei* (NRRL B-441), *J. Chem. Technol. Biotechnol.* 79 (2004) 1036–1040. <https://doi.org/10.1002/jctb.1094>.
- [9] Y. Sun, Z. Xu, Y. Zheng, J. Zhou, Z. Xiu, Efficient production of lactic acid from sugarcane molasses by a newly microbial consortium CEE-DL15, *Process Biochemistry*. 81 (2019) 132–138. <https://doi.org/10.1016/j.procbio.2019.03.022>.
- [10] D. Altıok, F. Tokatlı, Ş. Harsa, Kinetic modelling of lactic acid production from whey by *Lactobacillus casei* (NRRL B-441), *J. Chem. Technol. Biotechnol.* 81 (2006) 1190–1197. <https://doi.org/10.1002/jctb.1512>.
- [11] W. Sriphochanart, W. Skolpap, Temperature shift and feeding strategies for improving L-lactic acid production by *Lactiplantibacillus plantarum* in batch and fed-batch cultures, *Process Biochemistry*. 113 (2022) 11–21. <https://doi.org/10.1016/j.procbio.2021.12.006>.
- [12] E. Gencturk, K.O. Ulgen, Understanding HMF inhibition on yeast growth coupled with ethanol production for the improvement of bio-based industrial processes, *Process Biochemistry*. 121 (2022) 425–438. <https://doi.org/10.1016/j.procbio.2022.07.015>.
- [13] Y. Li, J. Xu, A PDF discretization scheme in wavenumber-frequency joint spectrum for simulating multivariate random fluctuating wind fields, *Probabilistic*

- Engineering Mechanics. (2023) 103422.
<https://doi.org/10.1016/j.pro bengmech.2023.103422>.
- [14] A. Janoska, J. Buijs, W.M. van Gulik, Predicting the influence of combined oxygen and glucose gradients based on scale-down and modelling approaches for the scale-up of penicillin fermentations, *Process Biochemistry*. 124 (2023) 100–112. <https://doi.org/10.1016/j.procbio.2022.11.006>.
 - [15] T.Y. Pai, S.H. Chuang, H.H. Ho, L.F. Yu, H.C. Su, H.C. Hu, Predicting performance of grey and neural network in industrial effluent using online monitoring parameters, *Process Biochemistry*. 43 (2008) 199–205. <https://doi.org/10.1016/j.procbio.2007.10.003>.
 - [16] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, *Neural Networks*. 2 (1989) 359–366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8).
 - [17] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational Physics*. 378 (2019) 686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>.
 - [18] V.V. Santana, M.S. Gama, J.M. Loureiro, A.E. Rodrigues, A.M. Ribeiro, F.W. Tavares, A.G. Barreto, I.B.R. Nogueira, A First Approach towards Adsorption-Oriented Physics-Informed Neural Networks: Monoclonal Antibody Adsorption Performance on an Ion-Exchange Column as a Case Study, *ChemEngineering*. 6 (2022) 21. <https://doi.org/10.3390/chemengineering6020021>.
 - [19] F.A.N. Fernandes, S. Rodrigues, Optimization of panose production by enzymatic synthesis using neural networks, *Process Biochemistry*. 41 (2006) 1090–1096. <https://doi.org/10.1016/j.procbio.2005.11.020>.
 - [20] Md. Altaf, B.J. Naveena, G. Reddy, Use of inexpensive nitrogen sources and starch for l(+) lactic acid production in anaerobic submerged fermentation, *Bioresource Technology*. 98 (2007) 498–503. <https://doi.org/10.1016/j.biortech.2006.02.013>.
 - [21] P. Dey, P. Pal, Modelling and simulation of continuous L (+) lactic acid production from sugarcane juice in membrane integrated hybrid-reactor system, *Biochemical Engineering Journal*. 79 (2013) 15–24. <https://doi.org/10.1016/j.bej.2013.06.014>.
 - [22] F. Rezvani, F. Ardestani, G. Najafpour, Growth kinetic models of five species of Lactobacilli and lactose consumption in batch submerged culture, *Brazilian Journal of Microbiology*. 48 (2017) 251–258. <https://doi.org/10.1016/j.bjm.2016.12.007>.
 - [23] A. Thakur, P.S. Panesar, M.S. Saini, Optimization of process parameters and estimation of kinetic parameters for lactic acid production by *Lactobacillus casei* MTCC 1423, *Biomass Conv. Bioref.* 9 (2019) 253–266. <https://doi.org/10.1007/s13399-018-0347-1>.
 - [24] M.Y. Vera-Peña, H. Hernández-García, F.E. Valencia-García, Kinetic modeling of lactic acid production, co-substrate consumptions and growth in *Lactiplantibacillus plantarum* 60-1, *Revista DYNA*. 89 (2022) 51–59.
 - [25] I. Alhama Manteca, A. Soto Meca, F. Alhama, Mathematical characterization of scenarios of fluid flow and solute transport in porous media by discriminated nondimensionalization, *International Journal of Engineering Science*. 50 (2012) 1–9. <https://doi.org/10.1016/j.ijengsci.2011.07.004>.

- [26] L. Lu, X. Meng, Z. Mao, G.E. Karniadakis, DeepXDE: A Deep Learning Library for Solving Differential Equations, *SIAM Rev.* 63 (2021) 208–228. <https://doi.org/10.1137/19M1274067>.
- [27] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, (2017). <http://arxiv.org/abs/1412.6980> (accessed January 23, 2023).
- [28] R.H. Byrd, P. Lu, J. Nocedal, C. Zhu, A Limited Memory Algorithm for Bound Constrained Optimization, *SIAM J. Sci. Comput.* 16 (1995) 1190–1208. <https://doi.org/10.1137/0916069>.
- [29] Z. Mao, A.D. Jagtap, G.E. Karniadakis, Physics-informed neural networks for high-speed flows, *Computer Methods in Applied Mechanics and Engineering*. 360 (2020) 112789. <https://doi.org/10.1016/j.cma.2019.112789>.

429 7 LEGENDS

430 Figure 1: Both the parameters and hyperparameters are selected beforehand,
431 iterated and each evaluated using the respective NN. The PINN updates the NN
432 according to the loss in each epoch of iteration

433 Figure 2: (a) Loss of test data over 30,000 epochs for nondimensional time
434 test (t_S), (b) Loss of wideness and deepness of layer test ($layer_size$) over 30,000
435 epochs and (c) comparison between experimental data, numerical and PINN results
436 for the simulation of a batch reactor.

437 Figure 3: (a) The profile of biomass (X), product (P) and substrate (S)
438 concentrations and volume (V) for each reactor operation mode (batch, fed-batch
439 and CSTR) simulated with both Euler method and PINN (b) The total loss for the
440 simulation of the batch reactor over the number of epochs specified.

441 Table 2: for all tests were the values are not specified, loss weights (w_V , w_X ,
442 w_P , w_S) and nondimensionalization scalars (V_S , X_S , P_S , S_S) were set to 1.

Table 1 - Conclusions for each test

| Test Name | Objectives | Conclusion |
|-----------------|--|--|
| t_S | Evaluate the effect of nondimensionalization scaler of time (t_s) | In general, the greater t_s the greater the loss |
| non_dim | Evaluate the effect of nondimensionalization scalars of X_s , P_s , S_s , and V_s | Batch, fed-batch and CSTR optimizations vary greatly, and there is no general rule |
| layer_size | 1) Evaluate the effect of layer size in performance 2) Check if it is possible to rationalize the ratio neuron/layers (N/L) in function of the loss | 1) Complex NNs (more than 4 layers or more than 60 neurons per layer) took much more epochs to minimize loss, but it was still not sufficient for achieving acceptable simulation results, while some simple NN (3 layers with 32 neurons and 4 layers with 22 neurons) were generally acceptable for batch reactor simulation; 2) It was not possible to find any clear relationship between loss, N/L and total number of neurons |
| weight | Evaluate the impact of loss weight in all cases studied | There is not set of weights that reduce the loss for all cases. Each case has a specific set of weight that minimizes loss. CSTR |
| multiple_config | Evaluate if the same configuration can appropriately represent batch, fed-batch and CSTR | Almost all configurations tested can produce agreeable results simulating the batch reactor, and almost none can represent fed-batch and CSTR cases |
| pinn_numeric_xp | Validate the proposed batch model, comparing with the solution using Euler Method and Experimental Data | The model was successfully validated |

Table 2 - Tests Parameters

| Test Name | Default Parameters | Case | Parameters |
|----------------|--|----------|--|
| t _S | 3 layers with 22 neurons 30,000 epochs | case t_1 | t _S = experiment time (9 h) |
| | | case t_2 | $t_S = 1 / (\mu_{\max} * S_o / (K_S + S_o))$ |
| | | case t_3 | $t_S = \alpha * S_o * (K_S + S_o) / \mu_{\max}$ |
| | | case t_4 | $t_S = (1/Y_{PS}) * \alpha * (K_S + S_o) / \mu_{\max}$ |
| | | case t_5 | t _S = 1 |
| | | case t_6 | $t_S = 1 / \mu_{\max}$ |
| non_dim | 3 layers with 22 neurons 80,000 epochs | n1 | V _S = 1, X _S = 1, P _S = 1, S _S = 1 |
| | | n2 | V _S = f _{in} *1h, X _S = 1, P _S = 1, S _S = 1 |
| | | n3 | V _S = V _{max} (5 L), X _S = X _o , P _S = P _o , S _S = S _o |
| | | n4 | V _S = V _{max} (5 L), X _S = X _M , P _S = P _M , S _S = S _o |
| | | n5 | V _S = 1, X _S = 1, P _S = P _M , S _S = S _o |
| | | n6 | V _S = V _{max} (5 L), X _S = X _M , P _S = 1, S _S = 1 |
| layer_size | t _s = 1 30,000 epochs | t_lay1 | 6 layers with 8 neurons |
| | | t_lay2 | 10 layers with 8 neurons |
| | | t_lay3 | 14 layers with 8 neurons |
| | | t_lay4 | 3 layers with 16 neurons |
| | | t_lay5 | 4 layers with 16 neurons |
| | | t_lay6 | 6 layers with 16 neurons |
| | | t_lay7 | 3 layers with 22 neurons |
| | | t_lay8 | 4 layers with 22 neurons |
| | | t_lay9 | 5 layers with 22 neurons |
| | | t_lay10 | 3 layers with 32 neurons |
| | | t_lay11 | 4 layers with 32 neurons |
| | | t_lay12 | 5 layers with 32 neurons |
| weight | 5 layers with 32 neurons 45,000 epochs w _V , w _X , w _P , w _S = 1 | W1 | w _V , w _X , w _P , w _S = 1 |
| | | W2 | w _X = 3 |
| | | W3 | w _P = 3 |
| | | W4 | w _S = 3 |
| | | W5 | w _V = 3 |

| | | | |
|----------------------|-----------------------------------|-------------------|---|
| | | W6 | $w_X = 3, w_V = 3$ |
| | | W7 | $w_P = 3, w_S = 3$ |
| | | W8 | $w_X = 2, w_S = 3, w_V = 5$ |
| | | W9 | $w_X = 5, w_V = 10$ |
| multiple_ config | 5 layers with 32 neurons | Batch | |
| | 80,000 epochs | Fed-Batch | $V_S = V_{\max} (5 L), X_S = X_M, P_S = P_M, S_S = S_o$ |
| | Pre and post L-BFG-S optimization | CSTR | $w_V = 3$ $V_S = V_{\max} (5 L), X_S = X_M$ |
| pinn_number ic_xp | Not applicable | Experimental Data | Not applicable |
| | | Euler | Euler simulation with 240 discretization points |
| | | PINN | 5 layers with 32 neurons 30,000 epochs |

445

Figure 1: Schematic of the PINN model

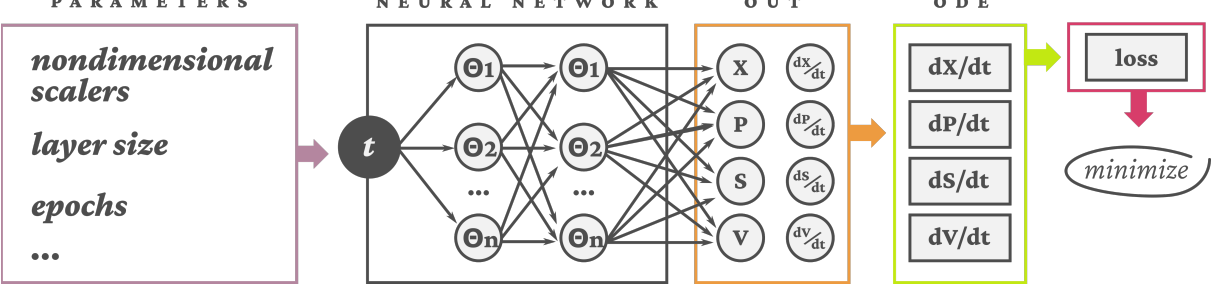


Figure 2: Nondimensional test, layer test and comparison between experimental data and PINN results

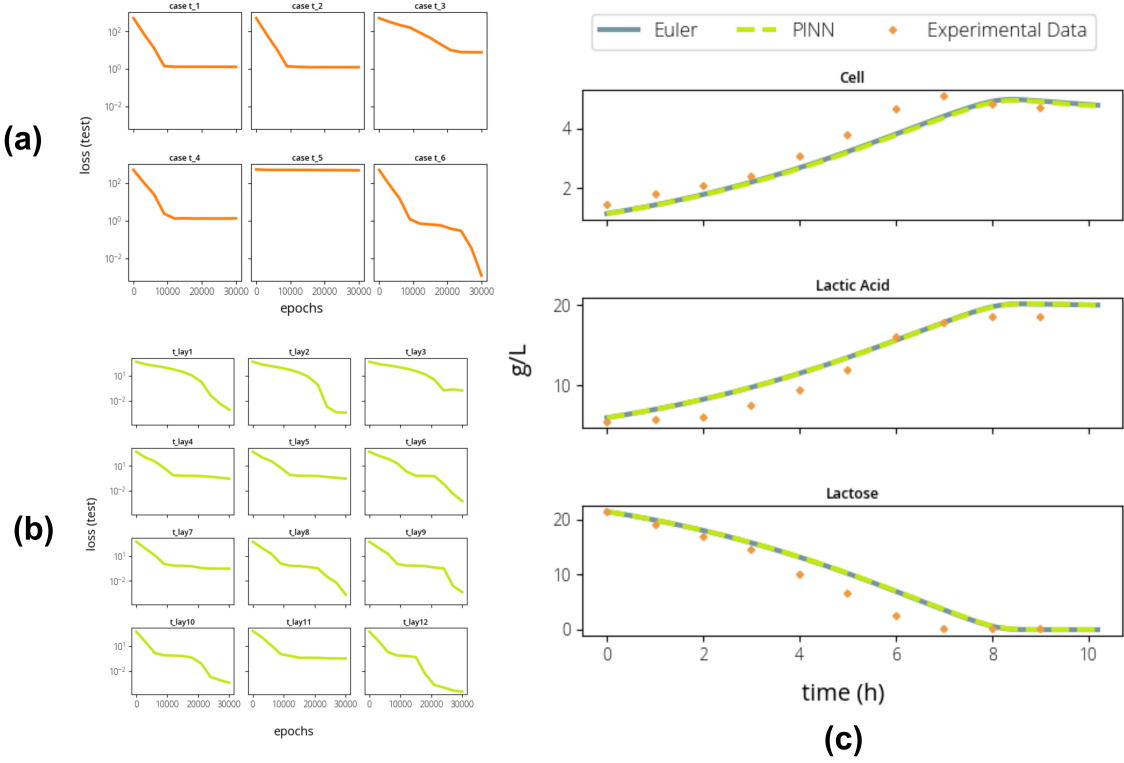
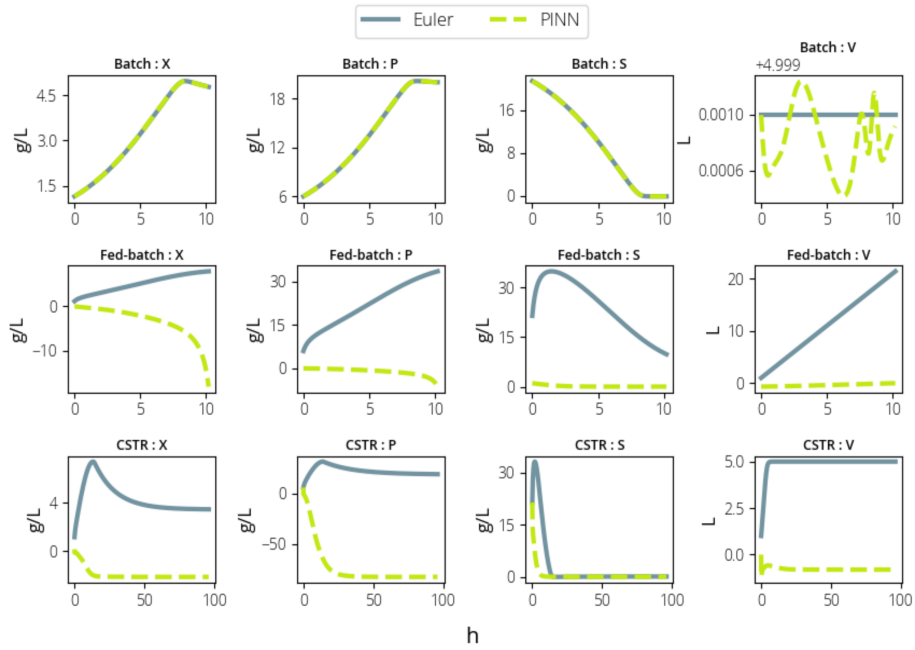
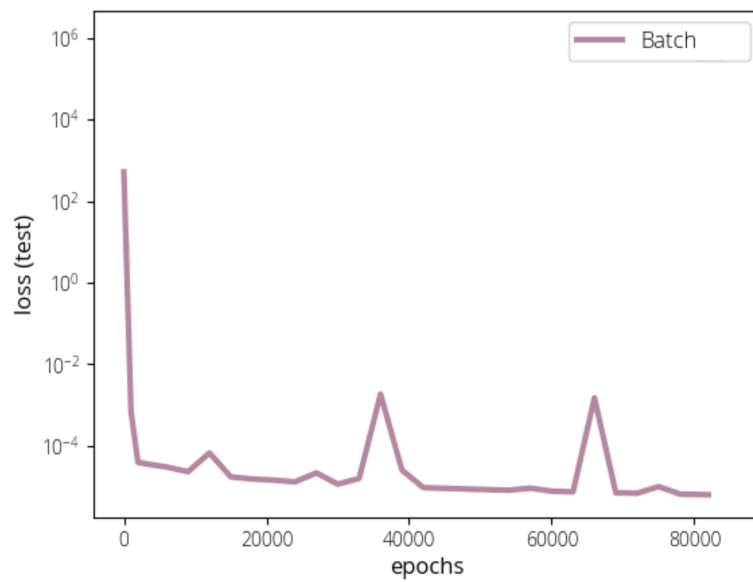


Figure 3: PINN results comparison for Batch, Fed-batch and CSTR reactors



(a)



(b)

451 **10 DECLARATION OF INTEREST**

452 Declarations of interest: none

453 11 HIGHLIGHTS

- 454 1) Volume and biomass concentration are the more relevant variables for loss
455 minimization
- 456 2) The batch reactor was successfully simulated using PINN
- 457 3) PINNs were not capable of appropriately simulate reactors with volume
458 variation
- 459 4) Strategies for simulating reactors with volume in transient state are suggested