Bayesian Hybrid Models for Simulation of Microbial Biohydrogen Photo-Production Processes

Shraman Pal,a Bovinille Anye Cho,b Anotonio Del Rio Chanona,c Dongda Zhang,b Max Mowbray c\*

*aDepartment of Chemical Engineering, IIT Kharagpur, Kharagpur, West Bengal, India*

bThe University of Manchester, Engineering Building A, Manchester, M13 9PL, UK

cImperial College London, South Kensington Campus, London, SW7 2AZ, UK

\*email of the corresponding author: mmowbray@ic.ac.uk

**Abstract**

Biohydrogen is widely regarded as the renewable energy of the future due to the qualities of zero emissions under combustion, high calorific value, and ambient synthesis. However, biohydrogen microbial photo-production systems are complex, providing challenge to simulate with mechanistic models and necessitating hybrid modelling. In this work, we present an effective methodology that leverages adjoint sensitivity analysis to identify a Bayesian neural network (BNN) hybrid model for prediction of a microbial photo-production process. The use of the Bayesian paradigm provides natural quantification of parametric uncertainties. We demonstrate that the scheme identifies an accurate model (3-11% mean percentage error) with well calibrated uncertainty estimates without mechanistic assumption on kinetics. In future, we will develop methods to enable the model transfer across scales.

**Keywords**: Bayesian Hybrid Modelling, Variational Inference, Biohydrogen.

1. Introduction

1.1 Dynamic modelling of biohydrogen photo-production process

The large-scale combustion of fossil fuels to meet the world’s energy is of increasing concern due to global climate change and energy security challenges arising from dwindling fossil fuel supplies. For these reasons, biohydrogen has been coined as “the energy of the future” due to its high calorific value and other appealing properties (Teke et al, 2023). Although biosynthesised by a plethora of microorganisms, *Rhodopseudomonas palustris* (*R. palustris*) has demonstrated the continuous production of biohydrogen during all growth phases, including the stationary phase, lasting significantly longer than observed with other microbial species (Anye Cho et al, 2021). As a result, *R. palustris’s* biohydrogen production has been extensively researched with first principles mechanistic models to facilitate experimental design and bioprocess upscaling to industrial scale from the usual laboratory scales and pilot plant scales (Anye Cho et al, 2021). However, the complexity of bioprocess dynamics results in high investment burden for development and validation of mechanistic model structures. Yet, reported mechanistic model predictions observe high uncertainty (up to 20% variation), often attributed to the difficulty in describing thousands of intracellular metabolite transformations with a few lumped biokinetics parameters. This has led to interest in the development of data-driven approaches.

* 1. Data-driven approaches and uncertainty estimation

The use of data driven approaches has been reviewed with use of Artificial Neural Network (ANN) reported widely (Teke *et al*., 2023). However, most of these studies predicted biohydrogen yield and/or biohydrogen productivity at the final fermentation time step, as a function of fixed process variables providing time invariant predictions. Such simplifications fall short of the underlying bioprocess description involving thousands of metabolite reactions, typically time varying over the course of fermentation. Thus, these approaches are suboptimal when utilized for biohydrogen bioprocess optimization. The few dynamic models reviewed in Teke *et al.* (2023) reported satisfactory overall biohydrogen production prediction (i.e., ), but poor prediction performance in the lag phase, thus requiring improved training approaches. For instance, algorithms predicting multiple steps ahead when provided with only the initial conditions, could alleviate prediction errors for the various growth phases (i.e., lag phase) (Del Rio Chanona et al, 2017).

However, none of the reported algorithms considered modelling uncertainties. Additionally, hybridising these algorithms with mechanistic models could benefit both predictive extrapolation and interpolation. However, this has not been addressed in the literature for biohydrogen production to the best of our knowledge. Therefore, this work aims to identify a Bayesian neural network (BNN) hybrid model for prediction of *R. palustris’s* biohydrogen production process. The BNN’s posterior predictive distributions enable the expression of parameter uncertainty, and direct quantification of the hybrid model uncertainty as a result. However, no literature investigation has reported this so far. Additionally, we highlight the versatility of adjoint sensitivity analysis for estimating the gradients of models with relatively high (i.e., >50) numbers of data driven parameters. Previously, hybrid models have typically been estimated using either (i) two-steps consisting of nonlinear programming and then supervised learning (Rogers et al, 2023), or (ii) via the forward sensitivities approach (von Stosch et al, 2014). Hence, we interrogate the proposed hybrid BNN’s performance and provide in-depth discussions.

1. Methodology

2.1 Hybrid models for dynamic modelling of biohydrogen photo-production systems

In this work, we assume the availability of a dataset ,where describes the evolution of the system state, , observed at discrete points over a time horizon, , and describes operational conditions that are fixed for a trajectory , over the course of the horizon. Specifically, in this work, we assume that we have measurements of the state representative of the biomass concentration (), , glycerol concentration (), , and biohydrogen collection (), within a small scale photobioreactor with volume, . The operational conditions, ,denote the incident light intensity, (. We assume that the photobioreactor (PBR) is a Schott bottle based PBR and is well-mixed by a magnetic stirrer. We wish to identify an approximating systems model as:

|  |  |
| --- | --- |
|  | (1) |

where represents a general description of the continuous time state dynamics; and we assume represent parameters of a data-driven model, which is embedded within the model structure provided by to predict time-varying parameters as a function of state. In this work, we specify this data-driven element as , which defines specific production or consumption rates of the state components via a neural network model structure. This means we simply leverage structure from the mass balance. We assume that estimation of the parameters, , is characterised by uncertainty, and so are interested in the use of Bayesian estimation practice for the identification of . This leads to discussion of BNN modelling.

2.2 Bayesian Neural Networks

Bayesian estimation practice is underpinned through use of Bayes’ rule for the inference of a posterior parameter distribution, , given definition of the likelihood of the data under the parameters, and choice of a prior distribution, ). Estimation of the posterior for the general case of nonlinear models is challenging and usually dependent on the use of sampling schemes. However, sampling observes an exponential increase in expense with parameter dimensionality, , which leaves its application dependent on relatively low dimensional parameter spaces (i.e. ).

Variational inference is an approach that poses Bayesian estimation as optimisation. The aim is to identify the best approximation of the posterior, , from a variational family, . The optimum minimises the Kullback-Liebler (KL) divergence to the true posterior (Mowbray et al, 2022). Typical parameterisation of the variational family is the joint of independent Gaussian distributions as inherited through the mean-field approximation, which describes the variational distribution as , where represents the mean, and represents a diagonal covariance matrix. The predictive distribution of the model is then expressed by sampling the parameters from the variational distribution. It is worth noting that this means that BNNs typically have twice the number of parameters than their frequentist counterparts have for a given structure. Estimation of can be approached by maximising a surrogate objective, known as the evidence lower bound (ELBO), which when evaluated empirically with samples, from takes the form:

|  |  |
| --- | --- |
|  | (2) |

The ELBO identifies the parameters of that maximise the expected sum of the likelihood of the data under the parameters, and a regularisation term that penalises deviations of the variational posterior approximation from the prior defined. Here, we estimate the moments of the approximation by use of the reparameterization trick, and by making assumption such that the likelihood reduces to the negative sum of square errors.

2.3 Integrating Bayesian neural networks and hybrid models

Within the scope of our hybrid model, we seek means to integrate BNNs for modelling biohydrogen production in continuous time. We formalise the problem as follows:

|  |  |
| --- | --- |
|  | (3) |

where evaluates the squared error to the data at discrete points in time, describes a set of samples with and is the identity matrix. The second set of constraints describe the reparameterization trick, which provides description of the parameter samples from the variational approximation deterministically in terms of its moments. We can estimate the moments of the variational approximation as usual through use of approximate second order optimisation schemes, which leverage first-order gradient information only. We can define the gradient of our ELBO objective with respect to parameters, , as follows:

|  |  |
| --- | --- |
|  | (4) |

where is the gradient of our least squares objective with respect to the variational parameters, and is the gradient of the Bayesian regularization term. Estimation of leverages automatic differentiation. Estimation of is proceeds via adjoint sensitivity analysis.

2.4 Adjoint sensitivity analysis

Estimation of via adjoint sensitivity analysis is appealing because it scales according to the number of BNN parameters only (i.e., rather than as is the case with the forward sensitivities). The procedure follows. Firstly, given a sample of the network parameters, the state dynamics are integrated forward from an initial condition, . The solution, , is stored in memory at checkpoints over the horizon. The cost is also evaluated at the discrete points in the time horizon for which one has measurements, and stored in memory. This enables definition of the adjoint ODE system, the solution of which, is identified by solving a terminal value problem, starting from the end of the horizon, , and proceeding backwards towards the initial time, . The terminal condition is defined by the partial derivatives of the terminal cost with respect to the state, . Due to the discrete nature of the objective, discontinuities exist in the adjoint solution, which is enforced via a jump condition. With the availability of both the state and adjoint solutions, as well as evaluations of the cost function, we can calculate the partial derivative of the least squares objective with respect to the variational parameter as (Chachuat, 2007):

|  |  |
| --- | --- |
|  | (5) |

where is the partial derivative of the least squares cost; is the partial derivatives of the state dynamics; and finally defines the partial derivatives of the initial state components, all of which are defined with respect to the variational parameter, and may be estimated through automatic differentiation. Essentially, we see that the partial derivative is comprised of three terms describing the influence of the parameters on the cost, the dynamics, and the initial conditions, respectively (Chachuat, 2007). In practice, several implementations exist to perform equivalent procedure to adjoint sensitivity analysis. In this work, we used the adjoint procedure that is defined within the JAX-based library, DIFFRAX (Kidger, 2022). We estimate gradients over mini batches of samples drawn from the variational distribution, with updates of the variational parameters at each iteration provided by the ADAM optimizer.

1. Case Study

We assume that we have a small dataset generated from a ground truth nonlinear kinetic model descriptive of our biohydrogen photo-production system, as presented in [2].We assume isothermal operation, such that we drop terms expressing temperature dependence from the original work:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|

|  |  |  |
| --- | --- | --- |
|  |  |  |

 | (6) |

where , , ,and and describe the effects of incident light intensity on the system. These terms approximate the spatial heterogeneity of the effects of light intensity, by using the average of light intensity within the reactor under assumption on its geometry and the use of the lambert beer law to describe its spatial decay from the PBR surface. Please see the original paper [2]for moreinformation. The ground truth model is sampled under different incident light intensities, , to generate a dataset of different trajectories under the model dynamics, starting from the same initial condition of . Each trajectory is associated with measurements over a horizon of hours. The trajectory corresponding to the 3rd condition of was used for validation, the others for training. Through the framework proposed in Section 2, we aim to identify a Bayesian hybrid model using the training data available. We define a neural network structure consisting of 4 hidden layers with rectified linear unit activation functions, and a hyperbolic tangent over the final hidden layer. This enables us to scale the predictions of the data-driven element to a pre-defined range. The input data were standardised to have a variance equal to 1, based on the mean and standard deviation of each state component. Dormand Prince's 8/7 method was utilized for forward numerical integration of the system state, as well as the backward adjoint integration. The prior on the parameters was defined as for all parameters in the network.

The model is assessed via the mean absolute percentage error (MAPE), and the coverage probability (CP). The CP defines the probability with which a certain number, , of standard deviations of the model’s prediction, will cover the residual between the data, and the mean prediction. Please see (Mowbray et al, 2022) for definition of these metrics. Table 1 details the model’s performance on the training and validation datasets.

Table 1: Summary of results for BNN hybrid model in training and validation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predictive task | State component | MAPE % | CP  | CP  |
| Training |  | 5.91 | 1.00 | 1.00 |
|  | 7.63 | 1.00 | 1.00 |
|  | 10.68 | 1.00 | 0.95 |
| Validation |  | 5.76 | 1.00 | 1.00 |
|  | 3.23 | 1.00 | 1.00 |
|  | 10.61 | 1.00 | 1.00 |

The results presented in Table 1 demonstrate a highly accurate predictive model in the mean, both in training and validation. This is highlighted by MAPE of 3-11% across both training and validation tasks for all state components. Additionally, the model performance is consistent across predictive tasks. Analysis of the CP indicates that uncertainty may not be perfectly calibrated, given that we should expect a reduction in the CP when . Ideally, in this case CP should be 0.95, instead of 0.997 as is the case for . This is only observed for the state component in training. The validation performance is visualized by Fig. 1, which shows the model’s prediction from the initial state over the horizon relative to the data available. Analysis of the figure highlights the predictive accuracy in the mean and emphasizes the calibration of the model uncertainty.



Figure 1: Visualization of model predictive performance. Left - biohydrogen; right - substrate. Blue scatter points represent measurements; orange line plot represents mean prediction; blue shaded region represents the 75th and 25th percentiles of the model’s predictive distribution.

1. Conclusions

In this paper, we demonstrated the construction of a Bayesian hybrid model using mean-field variational inference and adjoint sensitivity analysis. The model identified made limited mechanistic assumptions beyond the use of mass balance. However, it was able to identify accurate predictive performance in the region of 3-11% mean percentage error. As a result, this framework, provides avenue to mitigate dependence on mechanistic model construction practice, which can be highly time consuming. In future, we will investigate the effect of integrating our approach with additional mechanistic assumptions on the kinetics of the system. Furthermore, we will explore use of the model in the context of transfer learning to enable adaptation to different scales and geometries of biohydrogen photo-production processes.

References

Teke GM, Anye Cho B, Bosman CE, Mapholi Z, Zhang D, Pott RWM. Towards industrial biological hydrogen production: a review. World Journal of Microbiology 2023.

Anye Cho B, Ross BS, du Toit JP, Pott RWMC, del Río Chanona‬‬‬‬ EA, Zhang D. Dynamic

modelling of Rhodopseudomonas palustris biohydrogen production: Perturbation analysis and

photobioreactor upscaling. Int J Hydrogen Energy 2021;46:36696–708.

Del Rio-Chanona EA, Fiorelli F, Zhang D, Ahmed NR, Jing K, Shah N. An Efficient model

Construction Strategy to Simulate Microalgal Lutein Photo-Production Dynamic Process.

Biotechnol Bioeng 2017;114:2518–27.

Rogers AW, Song Z, Ramon FV, Jing K, Zhang D. Investigating ‘greyness’ of hybrid

model for bioprocess predictive modelling. Biochem Eng J 2023;190.

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.

Mowbray, M., Kay, H., Kay, S., Caetano, P. C., Hicks, A., Mendoza, C., ... & Zhang, D. (2022).

Probabilistic machine learning based soft-sensors for product quality prediction in batch

processes. Chemometrics and Intelligent Laboratory Systems, 228, 104616.

Chachuat, B. (2007). Nonlinear and dynamic optimization: From theory to practice.

Kidger, P. (2022). On neural differential equations. arXiv preprint arXiv:2202.02435.