Hybrid Learning of Transport Equations with Differentiable Neural Solvers from Experimental Data

Arthur Jessop, Mohammed Alsubeihi, Ashwin Kumar Rajagopalan
Department of Chemical Engineering
The University of Manchester, Manchester, M13 9PL
[arthur.jessop, mohammed.alsubeihi, a.rajagopalan]@manchester.ac.uk

Ben Moseley
Department of Earth Science and Engineering & I-X Centre for AI in Science
Imperial College London, South Kensington, SW7 2BX
b.moseley@imperial.ac.uk

Abstract

Transport equations are used across chemical engineering to model dynamic systems. However, these often rely on hand picked empirical equations to describe underlying physical laws, with no guarantee that these will accurately fit data - resulting in trial and error. We demonstrate an end-to-end differentiable solver that is used to learn underlying physical laws from real experimental data by integrating hybrid components into a mechanistic model. Compared to an industry standard benchmark model and solver, we show that our hybrid modelling approach achieves better model predictions without the need for additional data, and drastically outperforms the benchmark in both forward simulation and training. This approach can lead to improvements in the efficient design, operation, and control of processes in the fine chemicals and pharmaceuticals industry.

1 Introduction

Transport equations are essential for modelling systems across chemical engineering, for example, describing rheology in the formulation of consumer goods [1], population balances for pharmaceutical manufacturing [2], and adsorption/chromatography for applications such as CO_2 capture [3]. These models crucially rely on empirical equations to describe system dynamics accurately. Often, these equations must be hand picked and fitted to experimental data through trial and error, with no guarantee that a proposed functional form will adequately describe the system dynamics. This is especially problematic for complex systems or industries requiring rapid model development across varying conditions - such as pharmaceutical manufacturing, where time and material constraints prohibit extensive empirical testing for each new compound. In this case, material might be restricted to a few grams for designing an early stage process. Here, computational approaches can help to reduce the experimental, and therefore time and material burden.

Recent advances in scientific machine learning (SciML) offer an alternative: integrating neural networks (NNs) directly into mechanistic models [4, 5], facilitating a data-driven approach to learning underlying physical laws. Practical development of such approaches has been limited in the pharmaceutical and fine chemicals industry, often due to a lack of available data rich enough to describe process dynamics [6], and computational constraints that see many turn away from modelling.

1st Workshop on Differentiable Systems and Scientific Machine Learning @ EurIPS 2025.

In this paper, we 1) present a SciML framework for empirical law discovery for transport equations, and 2) demonstrate and release a curated dataset appropriate for SciML and analysis on data suitability for learning empirical laws. Key enablers of our framework are JAX and automatic differentiation. Through these tools we learn physical laws by leveraging an end-to-end differentiable PBE solver [7] which allows us to optimise neural networks within a mechanistic model. We find that our framework can discover empirical laws effectively and gain key insights into data requirements. By exploiting our models differentiability, we are able to gain further insight into its behaviour by performing sensitivity analysis on our initial conditions for example, something that is often overlooked in industry. This demonstrates a major step towards streamlining model fitting and prediction, opening the door towards Industry 4.0/5.0 [8], and end-to-end manufacturing goals.

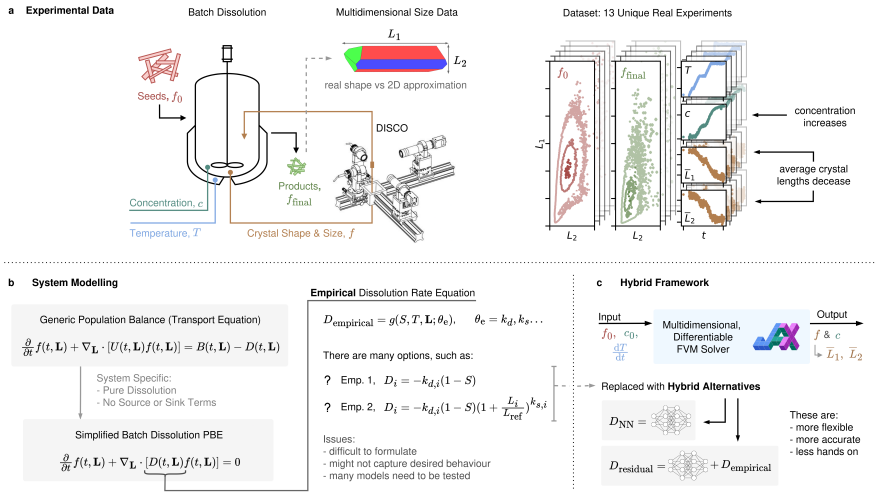

Figure 1: (a) an illustration of the experimental data collection and the dataset, (b) the general and specific transport equations used to model this system, along with example empirical equations for the underlying physical law, and (c) representations of our solver and the two hybrid components.

2 Methods

Task We learn a physical law governing a crystallisation process, one analogous to that found in the production of pharmaceuticals, where the size and shape distribution of crystals is a key performance attribute. These systems are governed by a range of different sub-processes that describe how crystals grow, dissolve, nucleate (appear, or are born), or agglomerate (stick together) - all of which need to be modelled. Specifically, we learn how crystals dissolve (shrink) upon heating, where molecules move from the solid phase to the liquid phase. An illustration of this process is shown in Figure 1(a).

Modelling Population balance equations (PBEs) are used to model the particle dynamics. These are transport equations that describe the time evolution of the size and shape distribution of a crystal population subject to these underlying phenomena. They can also model biomolecular processes, cell division and bubble dynamics [2]. The generic PBE and the simplified batch dissolution system used in this study are shown in Figure 1(b), both are partial differential equations. The law we learn is D in this simplified model. We solve a fully discrete system, using the finite volume method for spatial coordinates and forward Euler with a CFL condition for temporal coordinates.

Experimental Data Our dataset is high quality and, importantly, contains crystal size and shape information in multiple dimensions (both crystal length and width). Obtaining this level of insight is non-trivial and not available in industry. It is only possible with a small number of research devices, such as the DISCO, developed by Rajagopalan et al. [9, 10], which is used to curate our SciML ready dataset. The dataset contains 13 dissolution experiments where a number of different temperature ramps are used to demonstrate the dissolution behaviour of a crystal population over a range of different conditions, for which the underlying dissolution law is learnt. An illustration of the experimental setup, the DISCO, and a sample of the dataset are shown in Figure 1(a).

Hybrid Components Two NN architectures are proposed to replace the traditional empirical expressions. The first, *Pure NN*, fully replaces the empirical equation with a NN. The second, *Residual NN*, predicts the rate as the sum of a NN and an empirical equation. Here, the NN learns to correct a physics-based prior - in this case the chosen equation of Bötschi et al. [11], 'Emp. 1' in Figure 1(b). Both architectures are illustrated in Figure 1(c). All NNs are of the fully connected MLP type, and use `tanh` activation on the hidden layers. To ensure that the NNs give physically consistent outputs, a negative of the `softplus` activation is used on the output layer to ensure that the rates predicted are strictly negative, as the physics would dictate.

Loss Function & Training Three predictors are used for training: the concentration of the molecules dissolving into the liquid phase and the size of the particles (the length and width of the particles dissolving). Examples from the experimental dataset are shown in Figure 1(a). These are used in a mean squared error loss function where each predictor is equally weighted:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{3} \sum_{j=1}^{N} (y_{i,j} - \hat{y}_{i,j})^2. \tag{1}$$

Both the data and model results for each predictor are standardized using their mean and standard deviation taken from the experimental data (available pre-training) when calculating the loss. An Adam optimizer from `optax` is used to minimise the loss function with a fixed learning rate of 10^{-3}. Each hybrid model is trained for up to 500 epochs, with the final model parameters taken from the epoch yielding the smallest validation loss. Early stopping is also implemented.

Baseline We use the work of Bötschi et al. [11] as our baseline. They followed a traditional approach, fitting a number of empirical equations to the same dataset with a solver written in MATLAB, landing on the 'Emp. 1' formulation as seen in Figure 1(b). For a crystallisation specific model, this represents a state-of-the-art approach, both in terms of the underlying solver, and in terms of the model training - learning multidimensional crystal kinetics in both the length *and* width, a task that is seldom performed due to the availability of suitable data. We compare our two hybrid models to the original empirical results from the baseline when evaluating the predictive performance of the models. For simulation and training performance, we compare the two empirical models from Figure 1(b) in both solvers and demonstrate the performance of hybrid models in our differentiable solver.

3 Results

Simulation & Training Performance The results and comparison to our baseline for simulation and training performance are shown in Figure 2(b). The forward simulation time is the average time taken to simulate one experiment, and the training time is reported as the time required to fit the final model (with early stopping in JAX workflow). The time taken to perform a fixed 500 training epochs is also reported for the JAX models.

Our JAX solver is approximately $13\times$ faster for forward simulations than the baseline solver. These are pure benefits from JAX which provides `jit` compilation and access to GPU acceleration which is not available in the `.mex` compiled MATLAB code. The solver itself is the main bottleneck here, with the dissolution model having little impact on the simulation time - even for the NN with 2×10^5 parameters.

We show that the main benefit of the JAX workflow is the training performance. For comparable models (Emp 1. & Emp. 2.), the JAX workflow is substantially faster than the traditional approach - $3.4\times$ and $13.8\times$ respectively. Moreover, the MATLAB workflow scales poorly with parameter count: increasing from 2 to 4 parameters (Emp. 1 to Emp. 2) raises fitting time from 15 to 64 min. This is due to the `fmincon` optimizer relying on numerical gradients, requiring tens to hundreds of model evaluations per iteration. Such scaling drastically limits the complexity of models that can be practically fitted. In contrast, the JAX workflow scales very efficiently. For the same two models, the fitting time rises by only 12 s. Even for the NNs, increasing from 162 to $204\,602$ parameters adds just 42 s (15% increase). This efficiency arises from automatic differentiation, which provides exact gradients and handles large parameter sets with minimal overhead. By comparison, fitting a 4-parameter MATLAB model takes 64 minutes; extending this to even the smallest NN shown here (162 parameters) would be computationally infeasible. Flexible frameworks, like our hybrid models, can be used for autonomous equation discovery and model refinement to name a few.

Model Predictive Performance Figure 2(a) shows the model predictions for the validation dataset across each experimental predictor detailed in Section 2. Both hybrid approaches fit the data more closely than the empirical model, achieving lower validation losses of 1.02 and 1.12 for the *Pure NN* and *Residual NN*, respectively, compared to 1.23 for the baseline model. It is important to note that the hybrid models were trained using the same dataset (and split) as the baseline, demonstrating that improved performance is achieved without additional data. This shows how hybrid components can be readily integrated as an alternative to the empirical approach without sacrificing predictive performance or performing additional experiments to obtain more data, which in many cases would be a concern. Despite lower losses, all models show comparable qualitative predictive behaviour, a limitation attributed to the dataset rather than the model architecture. A key benefit of the hybrid approach is that only a single model needs to be trained, whereas many empirical equations might have to be trialled - with each of these requiring expertise to define.

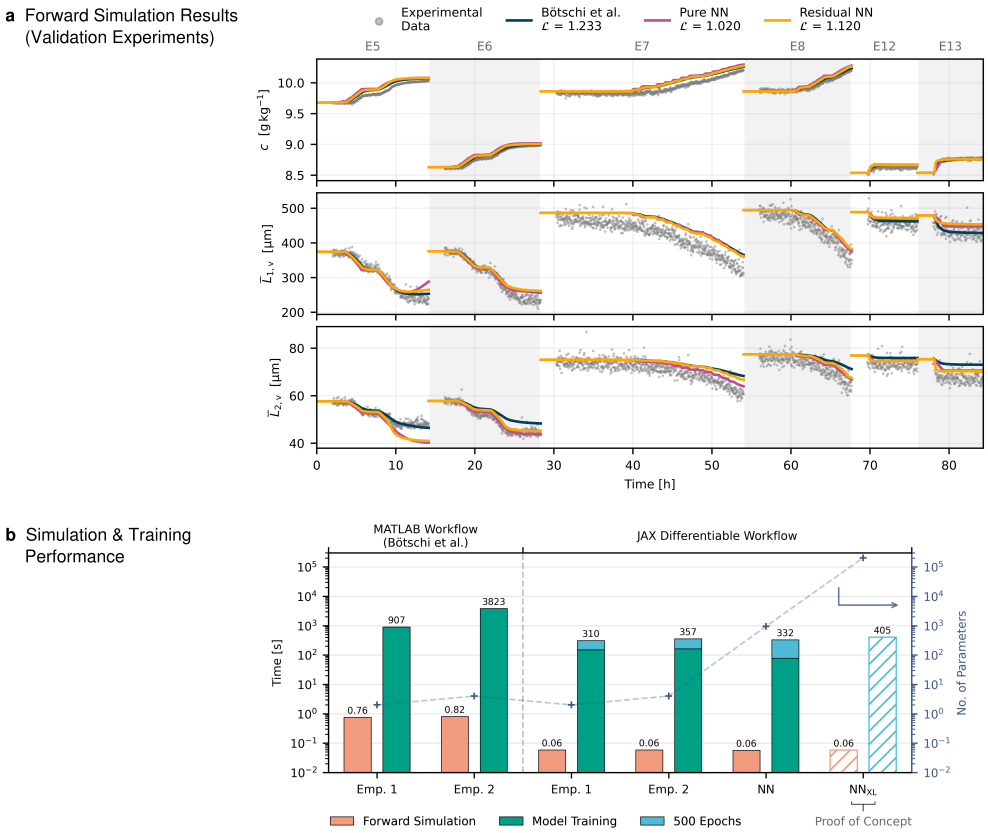

Figure 2: Comparison of (a) the predictive performance of our hybrid models against the benchmark, and (b) the computational performance of our fully differentiable workflow vs. the benchmark [11].

4 Conclusions

We have demonstrated how our end-to-end differentiable workflow can drastically accelerate traditional particle dynamics modelling. Systems can be simulated faster, and larger (more complex) models can be fitted in a fraction of the time compared to legacy implementations, with no extra data requirements. For conventional modelling, this framework enables rapid testing of multiple empirical equations in a fraction of the previous fitting time. For hybrid or SciML-based modelling, it provides a scalable foundation that integrates neural networks directly into mechanistic equations, removing the need to manually propose and test multiple empirical models.

These benefits extend beyond the case studied here. The same framework can be applied to learn any kinetic expression within the PBE or even other transport equations entirely, such as those governing adsorption/chromatography [3, 12].

References

[1] Howard Brenner. *Interfacial Transport Processes and Rheology*. Elsevier. ISBN 978-1-4832-9227-4.

[2] Doraiswami Ramkrishna. *Population Balances : Theory and Applications to Particulate Systems in Engineering*. Academic Press. ISBN 0-12-576970-9.

[3] Ashwin Kumar Rajagopalan, Adolfo M. Avila, and Arvind Rajendran. Do adsorbent screening metrics predict process performance? A process optimisation based study for post-combustion capture of CO2. 46:76–85, . ISSN 1750-5836. doi: 10.1016/j.ijggc.2015.12.033. URL https://www.sciencedirect.com/science/article/pii/S175058361530181X.

[4] Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal Differential Equations for Scientific Machine Learning. URL http://arxiv.org/abs/2001.04385.

[5] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. 3(6):422–440. ISSN 2522-5820. doi: 10.1038/s42254-021-00314-5. URL https://www.nature.com/articles/s42254-021-00314-5.

[6] Ian de Albuquerque, Marco Mazzotti, David R. Ochsenbein, and Manfred Morari. Effect of needle-like crystal shape on measured particle size distributions. 62(9):2974–2985. ISSN 1547-5905. doi: 10.1002/aic.15270. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/aic.15270.

[7] Mohammed Alsubeihi, Arthur Jessop, Ben Moseley, Cláudio P. Fonte, and Ashwin Kumar Rajagopalan. Modern, Efficient, and Differentiable Transport Equation Models Using JAX: Applications to Population Balance Equations. ISSN 0888-5885. doi: 10.1021/acs.iecr.4c04208. URL https://doi.org/10.1021/acs.iecr.4c04208.

[8] Aditya Akundi, Daniel Euresti, Sergio Luna, Wilma Ankobiah, Amit Lopes, and Immanuel Edinbarough. State of Industry 5.0—Analysis and Identification of Current Research Trends. 5(1): 27. ISSN 2571-5577. doi: 10.3390/asi5010027. URL https://www.mdpi.com/2571-5577/5/1/27.

[9] Ashwin Kumar Rajagopalan, Janik Schneeberger, Fabio Salvatori, Stefan Bötschi, David R. Ochsenbein, Martin R. Oswald, Marc Pollefeys, and Marco Mazzotti. A comprehensive shape analysis pipeline for stereoscopic measurements of particulate populations in suspension. 321: 479–493, . ISSN 1873328X. doi: 10.1016/j.powtec.2017.08.044.

[10] Daniel Biri, Anna Jaeggi, Oleksandr Prykhodko, William Hicks, Giulio Perini, Marco Mazzotti, and Ashwin Kumar Rajagopalan. Is There a "Right" Particle Size and Shape Measurement Tool?: A Comparative Study of Seven Online and Offline Devices on Nine Different Particle Populations. URL https://chemrxiv.org/engage/chemrxiv/article-details/68400722c1cb1ecda01ff4ea.

[11] Stefan Bötschi, Ashwin Kumar Rajagopalan, Manfred Morari, and Marco Mazzotti. Feedback Control for the Size and Shape Evolution of Needle-like Crystals in Suspension. IV. Modeling and Control of Dissolution. 19(7):4029–4043. ISSN 15287505. doi: 10.1021/acs.cgd.9b00445.

[12] Hassan Azzan, Ashwin Kumar Rajagopalan, Anouk L'Hermitte, Ronny Pini, and Camille Petit. Simultaneous Estimation of Gas Adsorption Equilibria and Kinetics of Individual Shaped Adsorbents. 34(15):6671–6686. ISSN 0897-4756. doi: 10.1021/acs.chemmater.2c01567. URL https://doi.org/10.1021/acs.chemmater.2c01567.