Embedding Physics into Neural ODEs to learn Kinetics from Integral Reactors

Tim Kirchera, Felix A. Döppela, Martin Votsmeiera,b,\*

aTechnische Universität Darmstadt, Peter-Grünberg-Straße 8, 64287 Darmstadt, Germany

bUmicore AG & Co. KG, Rodenbacher Chaussee 4, 63457 Hanau, Germany

martin.votsmeier@tu-darmstadt.de

Abstract

While the digitalization of chemical research and industry is vastly increasing the amount of data for developing kinetic models, model parametrization is not keeping up. To take advantage of the full potential of this data, machine learning tools are required that autonomously learn kinetics from reactor data. Previously, we introduced Global Reaction Neural Networks with embedded stoichiometry and thermodynamics for kinetic modelling. When trained as a neural ordinary differential equation (neural ODE), they discovered kinetics from integral reactor measurements of an equilibrium limited steam reforming reactor whereas conventional neural ODEs failed. We now extend their application to another industrially relevant case of reactors operating at full conversion. Using the preferential oxidation of CO in H2 rich streams as an example, we show that the physics-embedded neural network discovers kinetics from stiff systems containing cases of both full conversion and equilibrium limitation using integral reactor data.

**Keywords**: Neural ODE, Digitalization, Kinetic Modelling

* 1. Introduction

The design and control of chemical reactors requires accurate kinetic models. While the wealth of data for parameterizing kinetic models is increasing due to the emergence of big data frameworks (Wulf et. al, 2021), model parametrization is not keeping up. Machine learning models that autonomously learn kinetics from reactor data are required to tackle this challenge. Neural ordinary differential equations (neural ODEs) emerged as the state-of-the-art for learning system dynamics from time-series data, such as integral reactor measurements, with neural networks (Chen et al., 2018). Still, neural ODEs face several challenges in the case of kinetic modelling. The limited data availability in a laboratory setting requires neural network architectures that generalize with little training data. Also, chemical source terms frequently vary over many orders of magnitude and conventional neural networks struggle to model such data distributions (Döppel and Votsmeier, 2022). Recently, Kircher et al. (2023) proposed a Global Reaction Neural Network for efficiently modelling chemical source terms. This physics-embedded neural network was used together with neural ODEs and accurately learned kinetics from reactor data of a steam reforming reactor operating close to the equilibrium. In this paper we extend the application of the physics‑embedded neural networks to stiff chemical systems operating close to or at full conversion. We use the preferential oxidation of CO in H2 rich streams as an example, which contains both equilibrium limitations (water-gas shift equilibrium) and full conversion (oxidation reactions) and concentrations and source terms ranging over many orders of magnitude.

* + 1. Global Reaction Neural Networks

Global Reaction Neural Networks with embedded stoichiometry and thermodynamics (GRNN) use prior knowledge of the reaction system to map reaction conditions to the corresponding chemical source terms (Kircher et. al, 2023). When coupled with reactor physics, they can be trained as neural ODEs on integral reactor data (Figure 1).

Ein Bild, das Text, Diagramm, Schrift, Entwurf enthält.

Automatisch generierte Beschreibung

Figure 1: Training a GRNN coupled with reactor physics on integral reactor data.

The neural network layers of the GRNN with trainable weights map the molar flows and temperature to latent representations of the forward reaction rates :

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

The De Donder relation (Dumesic, 1999) is used to compute thermodynamically consistent net reaction rates from the forward rates using tabulated thermochemistry. These net rates approach zero if reactants are depleted and predict the correct equilibrium:

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

With the equilibrium constant *K*, mole fraction , stoichiometric coefficient , species index i and global reaction index j. Then, chemical source terms are computed using the embedded stoichiometry. These source terms strictly conserve the atom balance:

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

The embedded balance equations are then used to calculate the change in mole flow.

* 1. Numerical Experiments and Simulations
     1. Microkinetic Model

We consider the same reaction mechanism as used in our previous work for modeling the detailed surface kinetics of the preferential oxidation of CO (Döppel and Votsmeier, 2022). The mechanism describes the CO oxidation, H2 oxidation, water-gas shift reaction as well as the preferential oxidation of CO and the promoting role of H2O on CO oxidation on platinum. The mechanism is described by a microkinetic model with 36 reactions and 9 surface species. We use the kinetic parameters provided by Hauptmann et. al (2010). Elementary reaction rates are parametrized by Arrhenius kinetics. Deviating from the original implementation, we enforce thermodynamic consistency by calculating the rate constants of the backward reactions from the kinetic constant of the forward reaction and the equilibrium constant. A detailed description of this procedure as well as all kinetic and thermodynamic parameters are provided by Kircher et. al (2023).

* + 1. Generating Synthetic Experimental Data

The input range was chosen to cover typical operating conditions met in a reactor for the removal of CO impurities from high concentration H2 streams by preferential oxidation of CO with small amounts of added O2. Table 1 defines ranges of temperatures and mole fractions that capture this process (Manasilp & Gulari, 2002). Inlet conditions with a total molar fraction over 1 are discarded and resampled. Instead of sampling the O2 mole fractions, a factor is sampled. The inlet O2 mole fraction is calculated by multiplying this factor with the inlet CO mole fraction:

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

Table 1: Range of sampled inlet temperature, mole fractions and oxygen factor.

|  |  |  |
| --- | --- | --- |
| **Inlet Condition** | **Minimum** | **Maximum** |
| T / K | 400 | 440 |
|  | 0.400 | 0.750 |
|  | 0.050 | 0.150 |
|  | 0.005 | 0.020 |
|  | 0.100 | 0.200 |
|  | 0.5 | 1.5 |

The reactor is modeled as an isothermal, isobaric steady state 1-D plug flow reactor:

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

with cross-section area (1 m2), molar flow (mols-1), axial reactor coordinate (m), catalyst concentration (26.3 mol m-3) and chemical source term (s-1). The reactor has a length of 1 m, and the flow velocity is 1 m s-1. For a given inlet condition, axial reactor trajectories are calculated by numerical integration of the balance equation by ode15s in MATLABR2021a. Synthetic reactor experiments use molar flows measurements from six evenly distributed locations along the reactor length, the first at the inlet and the last at the outlet, and the temperature. The molar flows are perturbed by multiplicative (ϵ) and additive (η) gaussian noise with a mean of 0 and a standard deviation of 1. The minimum molar flow is defined as 10-6 mols-1, and all molar flows below are set to this value.

|  |  |
| --- | --- |
|  | (6) |

* + 1. Neural Network Hyperparameters

We train both GRNN and conventional neural networks. The models take the thermo-chemical state consisting of the temperature and molar flow of reactants as inputs and map them to the chemical source terms of all reactants. The inputs are transformed by mapping the temperature to the inverse temperature and the mole flows to their logarithm. Subsequently, the features are normalized by min-max scaling to a range of -1 and 1. The transformed and normalized features are input to the single hidden layer of the neural network with 50 nodes. We use hyperbolic tangent activation functions in the hidden layer. The weights are initialized by Xavier initialization. For the conventional neural network, the weights and biases of the last hidden layer are initialized as zeros such that the model is initially an identity mapping to stabilize training and improve performance.

* + 1. Neural ODE Training

Training is performed in full batch using the torch.optim.LBFGS quasi-newton optimizer implementation with strong wolfe line search, a learning rate of 1, absolute and relative tolerance of 10-50 and a history of size 100. Backpropagation is performed through the numerical solution of the differentiable ODE. Ten models with different seeds were trained. For training, we use a two-phase training method. First, the models are trained by a single step euler solver for 2500 epochs. During this phase, each reactor experiment is segmented into input-output pairs of adjacent measurements, such that the models approximate the gradient between molar flow measurements. For each seed, the model with the lowest validation error during training was retained. Second, the solver of the retained models is changed to RK4 with 1000 steps, and they are trained for 25 epochs. During this phase the models are given inlet molar flows of each reactor experiment, are integrated for the full length of the reactor, and the solution is evaluated where measurement data is available. For each seed, the model with the lowest validation error during training is saved. Of the ten saved models, the model with the lowest validation error is chosen for evaluation. L2 regularization is used with a regularization parameter of 10‑4. The models are trained to minimize the root mean squared error loss of predicted logarithmic molar flows and noisy molar flows of n=5 species over the m=5 measurement positions along the reactor and experiments:

|  |  |
| --- | --- |
|  | (7) |

Training and inference of neural networks is performed using pytorch 1.12.1. Differentiable ODE-solvers by torchdiffeq 0.2.3 are used for training the neural ODEs. The datasets were created in MATLABR2021a. Training is performed on an Intel® Xeon® Platinum 9242 with 96 cores and 2x Intel® AVX-512 units per core.

* 1. Results and Discussion

We demonstrate that our Global Reaction Neural Network with embedded stoichiometry and thermodynamics (GRNN) autonomously learns kinetics from multiple noisy integral reactor measurements. As a first step, we defined feed conditions where we required our machine learned kinetic model to work (Table 1). Note that the machine learned kinetic models need to cover a wider range, since molar flows change along the reactor. In a second step we created artificial reactor data by randomly sampling inlet conditions from the input range and running the reactor model for these conditions. In this manner we generated a training- (80 experiments), a validation- (20 experiments) and a test dataset (100 experiments). In a third step, the reactor experiments were discretized to five measurements locations per trajectory to match the limited data obtainable in a laboratory reactor experiment. Finally, the data was perturbed by gaussian noise. This yielded data that is typically used for parametrizing kinetic models. A neural ODE was then set up and trained on this data. Figure 2 shows an example of the ground truth (dots), the noisy training data (diamonds), and the predictions of the trained GRNN (lines) for the O2 and CO molar flows (left) and the O2 source terms (right). The global reactions governing the system are also shown on the right. The model is only trained on the noisy training data and does not have access to information on the ground truth molar flows or source terms.

Ein Bild, das Text, Diagramm, Reihe, Schrift enthält.

Automatisch generierte Beschreibung

Figure 2: Profile of O2 and CO molar flow (left) and O2 source terms (right) predicted by the trained GRNN (line), the noisy training data (diamonds) and the ground truth (circles).

The GRNN recovers the true molar flows with a higher accuracy than the perturbed data used for training it (Figure 2, left). The relative error of all model predictions to the ground truth mole flows, ranging over 8 orders of magnitude, is 5.9 %. This is lower than the relative deviation of the noisy training data towards the ground truth of 16.5 %. While this might seem unexpected, it is important to note that the neural ODE is highly biased towards the true solution due to the embedded information on stoichiometry and equilibria. Additionally, the model is improved by training on multiple experiments. We also show that the GRNN accurately recovers the non-monotonic dynamics ranging over multiple orders of magnitude (Figure 2, right). We then evaluated the GRNN on the previously unseen test dataset. We also trained and evaluated a conventional neural network in a same manner using the same data. Figure 3 shows parity plots of the selected ground truth molar flows and total mass flow and predictions by the trained GRNN (left) and conventional neural network (right).

|  |  |
| --- | --- |
|  |  |
| 1. GRNN | 1. Conventional Neural Network |

Figure 3: Parity plots of selected test set ground truth molar flows as well as the total mass flow and the predictions of the trained a) GRNN and b) conventional neural network.

Although only 100 reactor experiments with considerable noise were used for training and validation, the GRNN generalizes very well to unseen data for a system that contains species that are fully converted (O2) and species which are affected by equilibria (CO and H2). H2O and CO2 molar flows are predicted as accurately as the H2 molar flows. The mass balance, a fundamental aspect of any kinetic model, is exactly fulfilled by the GRNN due to the embedded stoichiometry. On the contrary, the conventional neural network does not capture the dynamics of the system and violates the mass balance.

* 1. Conclusion and Outlook

Kinetic models play a crucial role in model-based engineering of chemical reactors and the increasing availability of integral reactor data demands a general-purpose solution for automated kinetic modelling. We showed that embedding thermodynamics and stoichiometry into neural ODEs is a crucial step towards this goal as it ensures the models are physically sound and robust to noise. Therefore, they generalize well to unseen data, accurately recover the ground truth chemical source terms, and obey the mass balance by design. As this approach has proven successful for the two most important edge cases of industrial applications, namely equilibrium limitation and high conversion, we anticipate that it will be applicable to a wide range of systems and thus contribute to the acceleration of kinetic model development.

* 1. Acknowledgements

The authors gratefully acknowledge the computing time provided on the high-performance computer Lichtenberg at the NHR Centers NHR4CES at TU Darmstadt. This is funded by the Federal Ministry of Education and Research, and the state governments participating on the basis of the resolutions of the GWK for national high-performance computing at universities.

References

R.T.Q. Chen, Y. Rubanova, J. Bettencourt, D.K Duvenaud, 2018, Neural ordinary differential equations, Advances in neural information processing systems, Vol. 31.

J.A. Dumesic, 1999, Analyses of Reaction Schemes Using De Donder Relations, Journal of Catalysis 185, 496–505.

F.A. Döppel, M. Votsmeier, 2022, Efficient machine learning based surrogate models for surface kinetics by approximating the rates of the rate-determining steps, Chemical Engineering Science 262, 117964.

W. Hauptmann, M. Votsmeier, H. Vogel, D.G. Vlachos, 2011, Modeling the simultaneous oxidation of CO and H2 on Pt – Promoting effect of H2 on the CO-light-off, Applied Catalysis A: General 397, 174–182.

T. Kircher, F.A. Döppel, M. Votsmeier, 2023, A neural network with embedded stoichiometry and thermodynamics for learning kinetics from reactor data, ChemRxiv preprint: 10.26434/chemrxiv-2023-rpr35.

A. Manasilp, E. Gulari, 2002, Selective CO oxidation over Pt/alumina catalysts for fuel cell applications, Applied Catalysis B: Environmental 37, 17–25.

C. Wulf, M. Beller, T. Boenisch, O. Deutschmann, S. Hanf, N. Kockmann, R. Krähnert, M. Oezaslan, S. Palkovits, S. Schimmler, S.A. Schunk, K. Wagemann, D. Linke, 2021, A Unified Research Data Infrastructure for Catalysis Research – Challenges and Concepts, ChemCatChem 13, 3223–3236.