


Modeling the Direct Synthesis of Dimethyl Ether using Artificial Neural Networks

Nirvana Delgado Otalvaro, Pembe Gül Bilir, Karla Herrera Delgado*, Stephan Pitter, and Jörg Sauer

DOI: 10.1002/cite.202000226

 This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



Supporting Information
available online

Artificial neural networks (ANNs) are designed and implemented to model the direct synthesis of dimethyl ether (DME) from syngas over a commercial catalyst system. The predictive power of the ANNs is assessed by comparison with the predictions of a lumped model parameterized to fit the same data used for ANN training. The ANN training converges much faster than the parameter estimation of the lumped model, and the predictions show a higher degree of accuracy under all conditions. Furthermore, the simulations show that the ANN predictions are also accurate even at some conditions beyond the validity range.

Keywords: Artificial neural network, Dimethyl ether, Kinetics, Modeling

Received: October 28, 2020; *revised:* January 14, 2021; *accepted:* February 05, 2021

1 Introduction

Using artificial neural networks (ANNs) is the most widespread machine learning approach for modeling complex phenomena due to their simple formulation, flexibility and robustness [1, 2]. ANNs have proven to be suitable for creating predictive models for chemical engineering processes and several applications have been subject of research in the last decades such as the evaluation and modeling of complex kinetic data [3–6], catalyst design [7, 8], soft sensing [1, 9], advanced process control [10], and others [11]. Studies regarding the application of ANNs for the synthesis of dimethyl ether (DME) have been reported, e.g., for the screening of additives [7, 8], the optimization of temperature profiles in a temperature gradient reactor [12], and the modeling of the single process steps [13, 14]. Furthermore, ANNs have been used for predicting the performance of the liquid phase direct synthesis of DME over CuO/ZnO/Al₂O₃ and H-ZSM-5 catalysts [9]. In this work, we used ANNs to model the direct synthesis of DME from CO₂-rich synthesis gas over a mixed catalyst bed of commercial CuO/ZnO/Al₂O₃ (CZA) and γ -Al₂O₃ catalysts at high pressure. DME is of general interest due to its potential for chemical energy storage, making it a promising key compound in power to fuel technologies [15–20]. However, the detailed reaction mechanism of this system is still controversial [21]. One of the main difficulties for modeling the direct DME synthesis concerns changes in the catalyst during time on stream. It has been shown that the catalytic active state of CZA

dynamically adjusts to the process conditions [22, 23], particularly at high CO₂ contents in the synthesis gas feed [24]. In addition, water formation not only influences the active centers of CZA, but also those on the solid acid dehydration component (i.e., γ -Al₂O₃) [25, 26]. Morphological and structural changes induced by certain operating conditions or interactions with reactants, intermediates or products make it almost impossible to correlate a vast array of experiments at different working conditions using a simple kinetic model [27].

The ANNs used to model the direct synthesis of DME map the input-output relationships in intrinsic kinetic data taken over a wide range of operating conditions and inlet feed compositions. The ANNs applied are fully connected multi-layer feedforward networks trained by supervised learning. A brief summary of the theoretical background regarding the design and training of ANNs, is provided in the Supporting Information (SI). For the ANN design, several back-propagation training algorithms as well as different activation functions and network architectures have been tested. Additionally, a data partitioning scheme is pre-

Nirvana Delgado Otalvaro, Pembe Gül Bilir,
Dr. Karla Herrera Delgado, Dr. Stephan Pitter,
Prof. Dr.-Ing. Jörg Sauer
karla.herrera@kit.edu
Karlsruhe Institute of Technology, Institute of Catalysis Research
and Technology, Hermann-von-Helmholtz-Platz, 76344 Eggen-
stein-Leopoldshafen, Germany.m

sented, which enables the data division for training and testing in an automated fashion. We conduct simulations within and beyond the model's validity range to shed light on the ANN's predictive ability in both operational windows, and report on the ability of simple ANNs in modeling this system in comparison to that of a lumped kinetic model fitted to the same data.

2 Data and Methodology for the ANN's Design

Shallow feedforward ANNs (ANNs with one hidden layer) were designed and implemented in Matlab[®] software R2018a v9.4.0. The experimental kinetic data used for training and testing were acquired and published in a previous work [28]. The used data set consists of 180 experiments carried out in a fixed bed reactor at 50 bar using a 1:1 mechanical mixture of a commercial CZA catalysts and γ -Al₂O₃. The syngas composition, the temperature (T) and the total gas flow ($\dot{V}_{N,in}$) were varied during the experiments as summarized in Tab. 1, while the hydrogen amount in the feed gas was determined for each experiment according to Eq. (1). The remaining fraction of the feed gas consisted of a mixture of the inert gases argon and nitrogen.

$$y_{H_2,in} = 2.3(y_{CO,in} + y_{CO_2,in}) + y_{CO_2,in} \quad (1)$$

Table 1. Conditions of kinetic data taken from Delgado Otalvaro et al. [28]

Parameter	Value
T [K]	493, 503, 513, 523, 533
$\dot{V}_{N,in}$ [slpm] ^{a)}	0.2, 0.3, 0.4, 0.5, 0.6, 0.7
$y_{CO_2,in}$ [%]	1, 3
$y_{CO,in}$ [%]	4, 8, 15

a) Standard liter per minute, $T = 0^\circ\text{C}$ and $p = 1.01325$ bar

The ANNs were trained to predict the mole fraction of the main species (CO, CO₂, H₂ and DME) in the product gas based on the composition of the syngas ($y_{CO,in}$, $y_{CO_2,in}$, $y_{H_2,in}$) and the varied operating conditions. Hence, the input vector (x) and target vector (y) are summarized as follows:

$$x^T = [y_{CO,in}, y_{CO_2,in}, y_{H_2,in}, T, \dot{V}_{N,in}] \quad (2)$$

and

$$y^T = [y_{CO,out}, y_{CO_2,out}, y_{H_2,out}, y_{DME,out}] \quad (3)$$

For the design of ANNs the network architecture, i.e., the number of neurons in the hidden layer, as well as a suitable activation function of these neurons and a training algorithm must be determined. Since there is no generally

accepted theoretical basis to address these questions, answers are obtained empirically. For this purpose, various network architectures and multiple functions were screened and analyzed concerning the resulting accuracy and convergence time (refer to SI for further details on the evaluated algorithms). The assessment was carried out in regard to the mean squared error (MSE) and the convergence time. For this initial screening, the experimental input data were divided randomly into three data subsets: training, validation and test data containing 70 %, 15 % and 15 % of the experimental data, respectively. The validation subset was used for training to improve generalization through early stopping, except in the case of Bayesian regularization where generalization is achieved by regularization and no validation subset is required [29, 30]. The randomized data classification was constant for all trials conducted in this initial screening to ensure that the same samples were used in all cases, thus, excluding any influence of the data division from the preliminary results.

After determining the most appropriate functions several networks were trained using the pseudo-random two-stage data-partitioning scheme presented in Sect. 3.1. The error function on the test data was considered the determining factor for selecting the best network. Since this data set is completely independent of the training routine, the error on these data is a sufficient indicator of both the prediction accuracy and the generalization of the network.

Posterior to the training and network selection, simulations were performed with the selected network. The responses of the ANN were evaluated in comparison to a lumped kinetic model parametrized to the same experimental data used for the ANN training. The parameters of the lumped model were fitted to kinetic data measured in the absence of transport limitations. The assumptions of steady state, isothermal and isobaric operation, negligible gradients in radial direction and negligible backmixing effects apply. Therefore, only the effects of chemical reaction and thermodynamic equilibria are included in this model. However, since the lumped kinetic model is based on balance equations and partially on knowledge of the reaction mechanism, it is expected to deliver better predictions than the ANN when extrapolated.

The adjusted coefficient of determination R_{adj}^2 was computed as a measure of the goodness of fit (Eq. (4)). Different from the coefficient of determination R^2 (Eq. (5)), R_{adj}^2 takes the number of degrees of freedom of each model into consideration, hence, providing an unbiased basis for the comparison of two different model structures.

$$R_{adj}^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p} \quad (4)$$

$$R^2 = 1 - \frac{\sum_{n=1}^N (y_{n,out} - \hat{y}_{n,out})^2}{\sum_{n=1}^N (y_{n,out} - \bar{y}_{out})^2} \quad (5)$$

In Eqs. (4) and (5), N is the total number of experiments, and p is the number of model parameters. $\hat{y}_{n,out}$ and $y_{n,out}$ are the predicted and measured mole fraction of an arbitrary component in the product gas for experiment n , and \bar{y}_{out} is the mean value of the measured mole fraction over all experiments.

3 Results and Discussion

3.1 Network Design and Training

During the initial screening the activation and training functions available in Matlab[®] were changed systematically in order to find the most suitable function for the available data. The screening showed that the piecewise linear functions (ReLU, satlin, tribas and satlins) perform poorly compared to the nonlinear functions (radbasn, elliotsig, tansig, softmax and logsig). The best performance was obtained with the widely used logarithmic sigmoid function logsig (refer to Fig. S5 in the SI). When evaluating the training functions no convergence was achieved in any of the run trials with the algorithms Gradient Descent (gd) and Gradient Descent with Momentum (gdm). On the other hand, the Jacobian backpropagation methods Levenberg-Marquardt (lm) and Bayesian regularization (br) provide more accurate predictions than the gradient descent algorithms (cgp, scg, rp, bfg, cgb and cgf). Between lm and br, the lowest MSE and fastest convergence was achieved with br (Fig. S6). This Matlab[®] training function is based on the Bayesian interpolation frame proposed by MacKay [31] which is advantageous for problems where the data set is limited since no validation subset is required [29]. Furthermore, Bayesian regularization calculates and trains only the number of parameters necessary to minimize the target function (effective number of parameters) [32,33]. As a result, fewer parameters are used than are available reducing the model sensitivity to the network architecture, as long as the minimum number of neurons is provided. Based on these advantages and the empirically obtained results, Bayesian regularization was selected for the network design.

The proposed data division and training procedure is illustrated in Fig. 1. In the first stage of data division, the samples were randomly assigned to two subsets: “Design Data” and “Test Data A”. In the second stage, the “Design Data” subset containing 90 % of the samples was divided into “Train Data”, which is used to calculate weights and biases, and “Test Data B” used to compare different models within the framework of Bayesian regularization (without a validation subset). Afterwards, the multi-start strategy was applied by restarting the training procedure from different initial parameter values 100 times. This procedure, labeled as (1) in Fig. 1, screens the parameter space in order to generate different solutions of the optimization problem, and thus, to overcome possible local optimality. After completion, the second stage of data partitioning is repeated to

train the networks based on a different data division (label (2) in Fig. 1). All trained networks and training records were stored in a 100 by 100 array for the subsequent network selection. Finally, the “Test Data A” subset, which contains 10 % of the original samples, was used to provide an unbiased assessment of the network performance on separate data, and thus, of its generalization ability. Thereby, the ANN with the lowest error on these data exhibits the best generalization to the independent data set and was chosen as the most suitable network. A random division is advantageous for the problem at hand considering the multidimensionality of the input space. The presented scheme allows data partitioning in an automated fashion and increases the adaptability of the proposed modeling routine to new data sets of different structures. Furthermore, the model requirements, i.e., high accuracy, fast convergence and good generalization, are fulfilled.

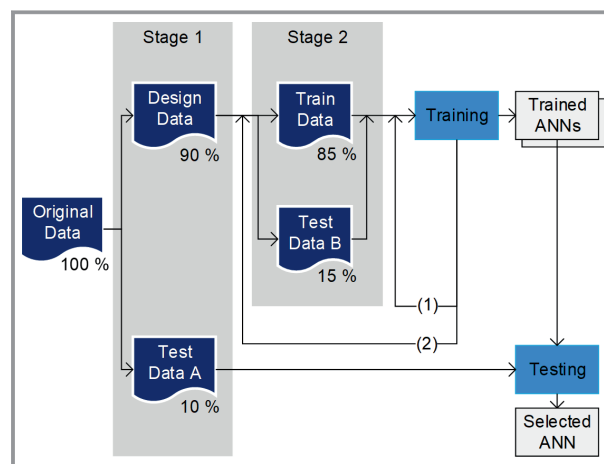


Figure 1. Data division scheme and training strategy.

The training strategy was conducted for networks with up to 15 neurons in the hidden layer. This screening showed that five hidden neurons provide enough complexity for the network to adapt sufficiently to the available data set. Therefore, the network with a 5-5-4 architecture (5 input, 5 hidden and 4 output neurons, Fig. S8) was selected. This structure ensures a sufficient number of parameters to avoid underfitting, while the problem of overfitting is prevented by training the network with Bayesian regularization. The resulting ANN is shown schematically in the SI, where also the parameters of the ANN and further training results are given.

The proposed approach is applicable when modeling with ANNs due to their remarkably fast convergence. For the chosen architecture, the time elapsed after the training of 10 000 networks was 7.9 min (refer to the SI). In contrast, the parametrization of the lumped kinetic model to the same data takes approximately 3.5 h using the same CPU (on windows 10 Pro (64-bit) operating system with i5 processor and 8 GB RAM).

3.2 Evaluation of the Selected ANN

Simulations were performed to evaluate the predictive ability of the selected network. Fig. 2 shows parity plots itemized for the main components in the system displaying the agreement between the measured and predicted concentrations in the product gas for all experiments. Clearly, the model is capable of simulating the observed trends accurately. For all components, the simulated points are evenly distributed around the $y = x$ line, indicating that there are no pronounced systematic deviations between model predictions and experimental data.

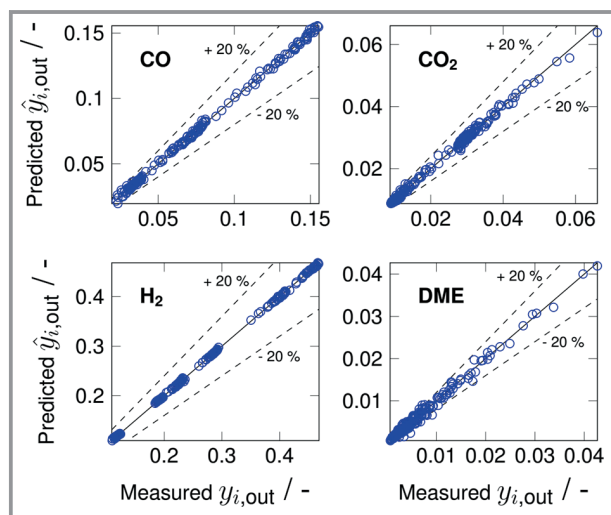


Figure 2. Parity plots for concentrations in the product gas.

We attribute the observed goodness of fit to the fact that appropriate activation and training functions were chosen as well as a network architecture that provides sufficient model complexity and flexibility for modeling. Additionally, the proposed data partitioning scheme proved to be effective in enabling the model to gain insight into the underlying phenomena with the available data.

The mean relative error (RE) over all inlet compositions is shown in Fig. 3 against the temperature and the inlet volume flow. Clearly, the ANN shows a higher predictive accuracy than the lumped kinetic model for all species in the entire experimentally covered operating window. This is caused by the flexibility and higher dimensionality of the ANN and its superior capacity to adapt to the data. The RE of CO, CO₂ and H₂ over all data lies below 3% (2%, 2.9% and 0.4%, respectively), while the RE of DME amounts to 11%. Both response surfaces for DME follow the same trend, with the prediction error decreasing with increasing temperature. At low temperatures, the low reaction rates lead to overall low conversion and yield. Hence, resulting in small DME amounts in the product gas and thus in a reduced measuring accuracy [28]. Therefore, the deviations of both models can be mainly attributed to experimental measurement uncertainties. Additionally, the fact that the ANN

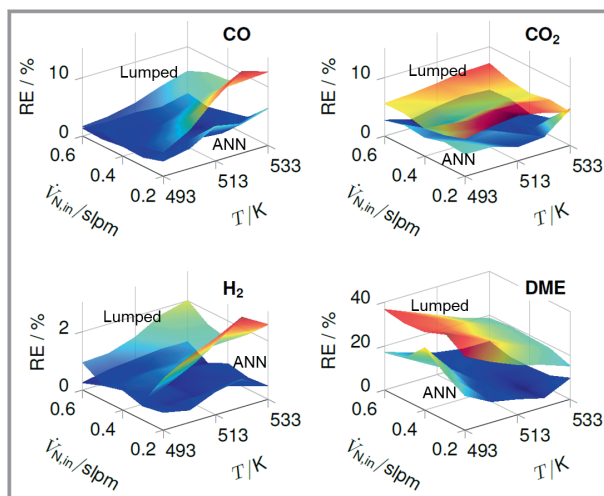


Figure 3. Mean relative error (RE) of prediction for the lumped model and ANN over all data.

did not adapt to the measured values, although the network has sufficient flexibility, is an indication that overfitting was successfully avoided and the data were not simply stored by the network, but the input-output relationships were effectively identified. The adjusted coefficients of determination reported in Tab. 2 highlight the suitability of both models and confirm the better adjustment of the ANN to the experimental data especially for the fractions of DME and CO₂.

Table 2. Adjusted coefficients of determination.

R_{adj}^2	$\hat{y}_{H_2,out}$	$\hat{y}_{CO,out}$	$\hat{y}_{CO_2,out}$	$\hat{y}_{DME,out}$
ANN model	0.999	0.998	0.994	0.984
Lumped model	0.998	0.992	0.984	0.943

In order to determine if the trained ANN is suitable as a non-linear regression tool, the ANN's generalization ability and its suitability to make predictions on unseen data have to be tested. For this purpose, additional simulations were performed for unobserved data within and beyond the model's validity range. The lumped model published in our previous work [28] is employed for a comparative analysis of the ANN's predictions. Since both models were fitted to the same experimental data, these are valid in the same range of conditions, thus, providing a sufficient basis for comparison. In the following, representative results are presented that illustrate and compare the responses of both models. Additional simulation results are given in the SI.

In Fig. 4, the experimental values are sorted arbitrarily in ascending order and depicted along with the superimposed confidence intervals of both fits at a significance level of 95%. It can be observed that the confidence intervals of the ANN predictions are narrower than the confidence interval of the lumped model. It is obvious, in particular for the fractions of DME and CO₂, that the respective confidence

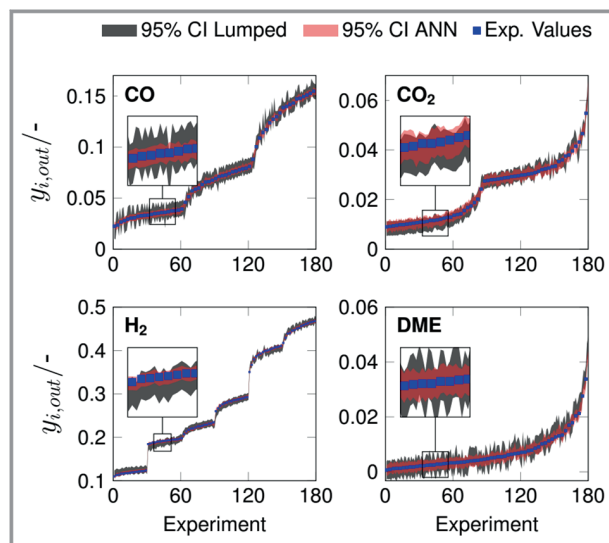


Figure 4. Measured concentrations in the product gas and 95 % confidence intervals (CI) of the ANN and lumped model. For clarity, only every third experimental data point is shown.

intervals of both models are wider in the low concentration range. This is in accordance with the presumption made before in this section that low concentrations of DME are subject to an increased measurement uncertainty, which also explains why this effect is not observable for the fractions of CO or H₂ where the confidence intervals appear to be of the same order of magnitude in the entire operating window.

Fig. 5 displays simulation and experimental results in the temperature range between 453 K and 573 K. The range where both models are formally valid (between 493 K and 533 K) is marked in gray for better visualization.

The predictions of the ANN within the model's validity range are slightly closer to the experimental values than the predictions obtained with the lumped kinetic model, consistent with the previous discussion. Since the phenomena in this range are dominated by reaction kinetics, the effects observed under these conditions can be explained by the temperature dependence of the reaction rate, described by the Arrhenius equation. With increasing temperature, the fraction of DME and CO₂ in the effluent increases, while the fraction of CO and H₂ decreases. The fact that CO₂ behaves as a product can be attributed to the water-gas shift reaction, which is promoted by the CZA catalyst and, in the evaluated range, is faster than the CO₂ hydrogenation. With regards to the total gas flow, it is observed that at decreasing values, the fraction of CO and H₂ at the reactor outlet decreases as well, while the fraction of DME and CO₂ increases. These results can be explained by the inverse relationship between the total gas flow with residence time and gas load, that lead to higher conversion and product yield. Furthermore, the consistency of this effect throughout the entire investigated gas flow range can be attributed to a constant selectivity towards DME. A detailed description of the

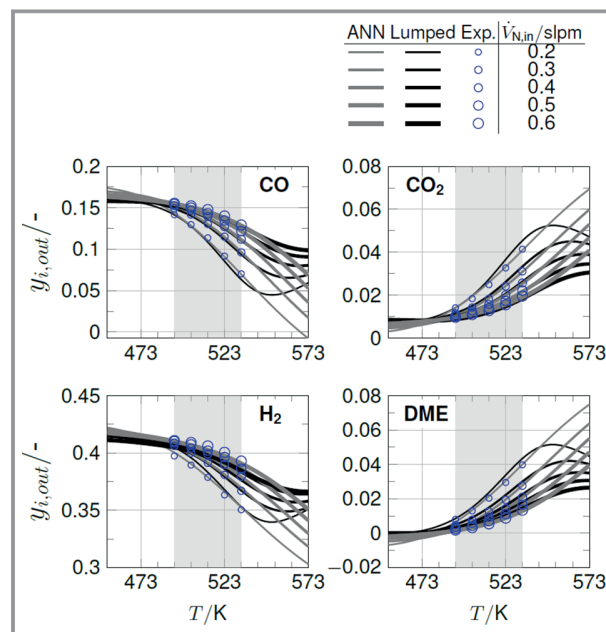


Figure 5. Components mole fraction in the product gas. Gray area marks the range covered experimentally. Feed: 16.1 % CO, 0.8 % CO₂, 42.3 % H₂, 40.8 % inert gas (Ar and N₂). $p = 50$ bar.

observed phenomena can be found elsewhere [28]. Model predictions in this operational range demonstrate the high level of agreement between the simulated and measured values, also showing a smooth mapping and the ANN's ability to generalize and make predictions for unseen data within the model's validity range.

Unexpectedly, the predictions of the lumped model and the ANN at temperatures below 493 K are similar although the ANN was not trained in this temperature range. Both models indicate that at low temperatures the reaction rates are too low to achieve high conversion. Hence, the concentrations of all components are close to the respective values in the feed gas. There are no additional constraints in the ANN's structure that prevent negative concentrations to be computed (in the lumped kinetic model, this effect is prevented inherently by the balance equations). Thus, at low temperatures some negative values are predicted. However, for DME and CO₂, progressions do not decrease steeply into the negative quadrant with decreasing temperatures. Instead, all values in this temperature range are close to zero. Similar good prediction accuracy despite extrapolation was observed for most but not all feed compositions and components (refer to SI, Fig. S9 to S13). Therefore, although the underlying model is able to extrapolate accurately for most conditions in this range, the quality of the predictions cannot be guaranteed in all cases. The predictions for temperatures above 533 K provide valuable insights into the phenomena comprised by the models. As the main chemical reactions involved in the DME synthesis are exothermic, high temperatures are kinetically favorable, but thermodynamically unfavorable. This trade-off of

exothermic reactions is reflected by a change in the slope of the concentration profile and is taken into account in the lumped kinetic model by the equilibrium constants in the rate expressions. However, since the kinetic data were measured at conditions at which the influence of the equilibrium is minor (kinetic regime), the ANN has no information about the characteristics of this phenomenon, causing the predictions of both models to diverge at high temperatures. With increasing temperature, the concentration profiles predicted with ANN follow the observed trend in the experimentally covered range, i.e., increasing for DME and CO₂, and decreasing for H₂ and CO, while the concentration profiles computed with the lumped kinetic model exhibit the expected points of inflection. Similarly, predictions for low flow rates at which mass transport limitation occurs can be expected to be inaccurate because the model was parameterized to fit intrinsic kinetic data, i.e., in an operating range with negligible influence of mass and heat transport.

4 Summary and Conclusions

In this paper ANNs were used to model the direct synthesis of DME from syngas over a commercial dual catalyst system at high pressure. The exact mechanism of this process is not yet fully understood, and modeling has so far only been possible in limited operating windows. The networks used in this study are shallow, feedforward and fully connected. It was demonstrated that the logarithmic sigmoid function is most applicable for the problem at hand, and that a higher accuracy is obtained when applying training algorithms that use Jacobian backpropagation, particularly Bayesian regularization. A pseudo-random data division scheme allowing data partitioning in an automated fashion was presented. The training was conducted for ANNs of different structures and five hidden neurons proved to provide sufficient model complexity to map the available data. The network with the best performance on unseen data was selected and its predictive ability was assessed by comparison with experimental data and with predictions of a lumped kinetic model parametrized to fit the same database. In summary, it was observed that the ANNs are remarkably fast, very flexible and exhibit a superior adaptability to the experimental data than the lumped kinetic model while still providing a comparable interpolation ability.

Moreover, accuracy of the model predictions outside the experimentally covered parameter range was also evaluated. When the model was extrapolated towards lower reaction rates, i.e., lower temperatures and higher flow rates, the ANN was able to deliver accurate predictions and to describe the single-stage DME synthesis systemically for most components and inlet feeds. This indicates that extrapolations of the model may be admissible for operating conditions at which the phenomena covered by the underlying model takes place. However, it is not possible to predict

deviations prior to training. Extrapolations of the ANN towards higher reaction rates, on the other hand, lead as expected to divergent predictions, as overlapping effects occur (e.g., thermodynamic limitation of exothermic reactions at high temperatures) which, at the current stage of development, cannot be reflected by the ANN that was trained to fit data taken in the operational window dominated only by reaction kinetics.

Our findings underline the suitability of the ANN to act as a predictive tool for Brownfield applications such as soft sensing, real-time optimization, online control, predictive maintenance and others, where models with high flexibility and adaptability, the capacity to map complex nonlinear relationships as well as fast convergence and low computational cost are required. Furthermore, we conclude that ANNs have the potential to be used for modeling the direct DME synthesis in an even wider range of operation where the relationship between input and output variables is ambiguous and modeling under mechanistic assumptions was not yet possible. The presented data partitioning and training methodology can be applied for this purpose with simple requirements: the input-output relationships to be modeled must be measurable and enough data must be available for parameter discrimination, i.e., for the training of the network. One possible application is the modeling of catalyst deactivation as a function of the time on stream and/or the conditions to which the catalyst system is exposed to. Regardless of the catalyst system, most kinetic studies of the direct DME synthesis are carried out under steady state conditions, due to the highly dynamic behavior of the catalysts which makes the mechanistic modeling in a wide range of conditions very challenging. However, if the required data are available, the modeling with the proposed methodology can be easily adapted to new state variables that need to be considered.

Supporting Information

Supporting Information for this article can be found under DOI: <https://doi.org/10.1002/cite.202000226>.

This research was funded by the Helmholtz Association, Research Programme “Storage and Cross-linked Infrastructures”, Topic “Synthetic Hydrocarbons”. Open access funding enabled and organized by Projekt DEAL.

Symbols used

N	[-]	total number of experiments
p	[bar]	pressure
R^2	[-]	adjusted coefficient of determination
R^{2adj}	[-]	coefficient of determination
T	[K]	temperature

$\dot{V}_{N,in}$	[slpm]	total gas flow at standard conditions, standard liter per minute, $T = 0\text{ }^{\circ}\text{C}$ and $p = 1.01325\text{ bar}$
y	[-]	measured mole fraction
\hat{y}	[-]	predicted mole fraction

Sub- and Superscripts

in	quantity at reactor inlet
n	experiment index
out	quantity at reactor outlet
T	transpose of a matrix or vector

Abbreviations

ANN	artificial neural network
br	Bayesian regularization
CI	confidence interval
CZA	CuO/ZnO/Al ₂ O ₃
DME	dimethyl ether
lm	Levenberg-Marquardt
MSE	mean squared error
RE	relative error

References

- J. Mohd Ali, M. A. Hussain, M. O. Tade, J. Zhang, *Expert Syst. Appl.* **2015**, *42* (14), 5915–5931. DOI: <https://doi.org/10.1016/j.eswa.2015.03.023>
- M. Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems*, Pearson education, London **2005**.
- V. Goncalves, K. M. Honório, A. B. F. da Silv, in *Artificial Neural Networks – Architectures and Applications* (Ed: K. Suzuki), IntechOpen, London **2013**, 203–223. DOI: <https://doi.org/10.5772/51275>
- F. Amato, J. L. González-Hernández, J. Havel, *Talanta* **2012**, *93*, 72–78. DOI: <https://doi.org/10.1016/j.talanta.2012.01.044>
- M. Shiva, H. Atashi, F. Tabrizi, A. Mirzaei, A. Zare, *Fuel Process. Technol.* **2013**, *106*, 631–640. DOI: <https://doi.org/10.1016/j.fuproc.2012.09.056>
- S. Sunphorka, B. Chalermisinsuwan, P. Piumsomboon, *Fuel* **2017**, *193*, 142–158. DOI: <https://doi.org/10.1016/j.fuel.2016.12.046>
- K. Omata, Sutarto, M. Hashimoto, G. Ishiguro, Y. Watanabe, T. Umegaki, M. Yamada, *Ind. Eng. Chem. Res.* **2006**, *45* (14), 4905–4910. DOI: <https://doi.org/10.1021/ie050640g>
- K. Omata, Sutarto, M. Hashimoto, G. Ishiguro, Y. Watanabe, T. Umegaki, M. Yamada, *Ind. Eng. Chem. Res.* **2005**, *44* (2), 296–301. DOI: <https://doi.org/10.1021/ie049302q>
- G. R. Moradi, F. Parvizian, *Can. J. Chem. Eng.* **2011**, *89* (5), 1266–1273. DOI: <https://doi.org/10.1002/cjce.20558>
- J. Appel, C. Colombo, U. Dätwyler, Y. Chen, N. Kerimoglu, *Chimia* **2016**, *70* (9), 621–627. DOI: <https://doi.org/10.2533/chimia.2016.621>
- H. Z. Li, Z. Liu, Z., *Catalysts* **2017**, *7* (10), 306. DOI: <https://doi.org/10.3390/catal7100306>
- K. Omata, T. Ozaki, T. Umegaki, Y. Watanabe, N. Nukui, M. Yamada, *Energy Fuels* **2003**, *17* (4), 836–841. DOI: <https://doi.org/10.1021/ef0202438>
- P. Valeh-e-Sheyda, F. Yaripour, G. Moradi, M. Saber, *Ind. Eng. Chem. Res.* **2010**, *49* (10), 4620–4626. DOI: <https://doi.org/10.1021/ie9020705>
- S. Alamolhoda, M. Kazemeini, A. Zaherian, M. R. Zakerinasab, *J. Ind. Eng. Chem.* **2012**, *18* (6), 2059–2068. DOI: <https://doi.org/10.1016/j.jiec.2012.05.027>
- M. Stiefel, R. Ahmad, U. Arnold, M. Döring, *Fuel Process. Technol.* **2011**, *92* (8), 1466–1474. DOI: <https://doi.org/10.1016/j.fuproc.2011.03.007>
- E. Catizzzone, G. Bonura, M. Migliori, F. Frusteri, G. Giordano, *Molecules* **2017**, *23* (1), 1–28. DOI: <https://doi.org/10.3390/molecules23010031>
- M. Gentzen, W. Habicht, D. E. Doronkin, J.-D. Grunwaldt, J. Sauer, S. Behrens, *Catal. Sci. Technol.* **2016**, *6*, 1054–1063. DOI: <https://doi.org/10.1039/C5CY01043H>
- N. Dahmen, U. Arnold, N. Djordjevic, T. Henrich, T. Kolb, H. Leibold, J. Sauer, *J. Supercrit. Fluids* **2015**, *96*, 124–132. DOI: <https://doi.org/10.1016/j.supflu.2014.09.031>
- U. Mondal, G. D. Yadav, *J. CO₂ Util.* **2019**, *32*, 299–320. DOI: <https://doi.org/10.1016/j.jcou.2019.02.003>
- U. Mondal, G. D. Yadav, *J. CO₂ Util.* **2019**, *32*, 321–338. DOI: <https://doi.org/10.1016/j.jcou.2019.02.006>
- J. Park, H. S. Kim, W. B. Lee, M.-J. Park, *Catalysts* **2020**, *10* (10), 655. DOI: <https://doi.org/10.3390/catal10060655>
- R. Schlögl, *Angew. Chem., Int. Ed.* **2015**, *54* (11), 3465–3520. DOI: <https://doi.org/10.1002/anie.201410738>
- S. Kuld, M. Thorhauge, H. Falsig, C. F. Elkjær, S. Helveg, I. Chorkendorff, J. Sehested, *Science* **2016**, *352* (6288), 969–974.
- J.-D. Grunwaldt, A. M. Molenbroek, N. Y. Topsøe, H. Topsøe, B. S. Clausen, *J. Catal.* **2000**, *194* (2), 452–460. DOI: <https://doi.org/10.1006/jcat.2000.2930>
- F. Frusteri, M. Migliori, C. Cannilla, L. Frusteri, E. Catizzzone, A. Aloise, G. Giordano, G. Bonura, *J. CO₂ Util.* **2017**, *18*, 353–361.
- A. I. Osman, J. K. Abu-Dahrieh, D. W. Rooney, S. A. Halawy, M. A. Mohamed, A. Abdelkader, *Appl. Catal., B* **2012**, *127*, 307–315. DOI: <https://doi.org/10.1016/j.apcatb.2012.08.033>
- K. F. Kalz, R. Kraehnert, M. Dvoyashkin, R. Dittmeyer, R. Gläser, U. Krewer, K. Reuter, J.-D. Grunwaldt, *ChemCatChem* **2017**, *9*, 17–29. DOI: <https://doi.org/10.1002/cctc.201600996>
- N. Delgado Otalvaro, M. Kaiser, K. Herrera Delgado, S. Wild, J. Sauer, H. Freund, *React. Chem. Eng.* **2020**, *5*, 949–960. DOI: <https://doi.org/10.1039/d0re00041h>
- Z. S. H. Chan, H. W. Ngan, A. B. Rad, A. K. David, N. Kasabov, *Neurocomputing* **2006**, *70* (1), 409–419. DOI: <https://doi.org/10.1016/j.neucom.2005.12.131>
- D. J. C. Mackay, *Network: Comput. Neural Syst.* **1995**, *6*, 469–505. DOI: https://doi.org/10.1088/0954-898X_6_3_011
- D. J. C. MacKay, *Neural Comput.* **1992**, *4*, 415–447. DOI: https://doi.org/10.1007/978-94-017-2219-3_3
- F. Burden, D. Winkler, in *Artificial Neural Networks: Methods and Applications* (Ed: D. J. Livingstone), Humana Press, Totowa, NJ **2009**. DOI: <https://doi.org/10.1007/978-1-60327-101-1>
- H. Okut, Bayesian Regularized Neural Networks for Small n Big p Data, in *Artificial Neural Networks – Models and Applications* (Ed: J. L. Garcia Rosa), IntechOpen, London **2016**. DOI: <https://doi.org/10.5772/63256>
- S. Zendeheboudi, N. Rezaei, A. Lohi, *Appl. Energy* **2018**, *228*, 2539–2566. DOI: <https://doi.org/10.1016/j.apenergy.2018.06.051>
- I. A. Basheer, M. Hajmeer, *J. Microbiol. Methods* **2000**, *43* (1), 3–31. DOI: [https://doi.org/10.1016/S0167-7012\(00\)00201-3](https://doi.org/10.1016/S0167-7012(00)00201-3)
- K. Hornik, M. Stinchcombe, H. White, *Neural Networks* **1989**, *2* (5), 359–366. DOI: [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)

- [37] G. Cybenko, *Math. Control Signals Syst.* **1989**, *2*, 303–314. DOI: <https://doi.org/10.1007/BF02551274>
- [38] K. Hornik, *Neural Networks* **1991**, *4* (2), 251–257. DOI: [https://doi.org/10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T)
- [39] M. H. Beale, M. T. Hagan, H. B. Demuth, *Neural Network Toolbox TM User's Guide R2013b*, The Mathworks Inc, 7th ed., The MathWorks, Inc. **2013**.
- [40] A. A. Hameed, B. Karlik, M. S. Salman, *Knowledge-Based Syst.* **2016**, *114*, 79–87. DOI: <https://doi.org/10.1016/j.knosys.2016.10.001>
- [41] B. Sharma, P. Venugopalan, *IOSR J. Comput. Eng.* **2014**, *16* (1), 31–35. DOI: <https://doi.org/10.9790/0661-16123135>
- [42] H. Yonaba, F. Anctil, V. Fortin, *J. Hydrol. Eng.* **2010**, *15* (4), 275–283. DOI: [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000188](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000188)
- [43] <https://uk.mathworks.com/help/deeplearning/ug/choose-a-multilayer-neural-network-training-function.html>, (Accessed on April 24, 2020)
- [44] J. E. Dennis, R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Society for Industrial and Applied Mathematics, Philadelphia, PA **1996**. DOI: <https://doi.org/10.1137/1.9781611971200>
- [45] P. E. Gill, W. Murray, M. H. Wright, *Practical Optimization*, Academic Press, London **1981**. DOI: <https://doi.org/10.1002/nme.1620180612>
- [46] M. J. D. Powell, *Math. Program.* **1977**, *12*, 241–254. DOI: <https://doi.org/10.1007/BF01593790>
- [47] E. M. L. Beale, in *Numerical Methods for Nonlinear Optimization* (Ed: F. A. Lootsma), Academic Press, London **1972**, 39–43.
- [48] L. E. Scales, *Introduction to Non-Linear Optimization*, Macmillan International Higher Education in Computer Science Series, Palgrave Macmillan, London **1985**.
- [49] D. W. Marquardt, *J. Soc. Ind. Appl. Math.* **1963**, *11* (2), 431–441. DOI: <https://doi.org/10.1137/0111030>
- [50] M. T. Hagan, M. B. Menhaj, *IEEE Trans. Neural Networks* **1994**, *5* (6), 989–993. DOI: <https://doi.org/10.1109/72.329697>
- [51] R. Battiti, *Neural Comput.* **1992**, *4* (2), 141–166. DOI: <https://doi.org/10.1162/neco.1992.4.2.141>
- [52] M. Riedmiller, H. Braun, in *IEEE International Conference on Neural Networks*, San Francisco, USA, **1993**, *4*, 586–591. DOI: <https://doi.org/10.1109/ICNN.1993.298623>
- [53] M. F. Møller, *Neural Networks* **1993**, *6* (4), 525–533. DOI: [https://doi.org/10.1016/S0893-6080\(05\)80056-5](https://doi.org/10.1016/S0893-6080(05)80056-5)