Long Short-Term Memory Networks for the Prediction of Fuel Cell Voltage and Efficiency

Martin Angerbauer, Michael Grill and André Casal Kulzer

October 26, 2022

Article

Long Short-Term Memory Networks for the prediction of Fuel Cell Voltage and Efficiency

Martin Angerbauer [1,*], Michael Grill [1,] and André Casal Kulzer [1,]

[1] FKFS Research Institute for Automotive Engineering and Vehicle Engines Powertrain Systems Stuttgart; www.fkfs.de
* Correspondence: martin.angerbauer@fkfs.de

Abstract: Fuel cells are once again experiencing an upswing, as they are a possibility for climate-friendly mobility. Nevertheless, the aim is to operate them at the highest efficiency, which is highly dependent on the operating condition. In order to operate fuel cells at the highest possible efficiency continuously, fast-calculating and reliable predictions are essential. One approach to provide these predictions is artificial neural networks (ANN), which are significantly faster compared to phenomenological models. In this work, recurrent neural networks (RNN) are trained with dynamic data of a proton exchange membrane fuel cell (PEMFC). Due to the different time scales of the processes that occur during the operation of the fuel cell, the latest operating state is not sufficient for a precise prediction.
Since, for example, the absorption and release of water take place slowly, earlier states are required in order to consider these processes. Therefore, the choice fell on RNN with long short-term memory cells (LSTM), which are trained using time series of various dynamic operating cycles. Thus, all time scales are regarded in one combined model that offers fast prediction.

Keywords: LSTM; Fuel Cell; Fast Running Model

check for updates

Citation: Angerbauer, M.; Grill, M.; Kulzer, A. Long Short-Term Memory Networks for the prediction of Fuel Cell Voltage and Efficiency. *Preprints* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

Fuel cells (FC) are one of the propulsion technologies that is mentioned consistently when talking about climate-neutral vehicles. Due to the low operating temperature, low operating pressure, compact size, and high power density, proton-exchange membrane (PEM) FC are the most suitable FC technology for individual transportation applications [7]. Since the only product of the reaction of FC is water, they do not emit any pollutants in contrast to internal combustion engines. The great benefit of FC electric vehicles (FCEV) over battery electric vehicles (BEV) is the shorter refueling time. However, the cost of FC are still high and the hydrogen infrastructure is not well developed yet. Moreover, hydrogen is still too expensive to make FCEV a serious competitor for BEV as future energy storage technology [10].
To maximize the range of FCEV, the highest operating efficiency is desired. Two energy management strategies are distinguished: online control strategy and offline control strategy. The offline control strategy is a rule-based strategy that optimizes a cost function. For this approach, it is necessary to know the entire driving cycle in advance. By means of the entire cycle, an optimization is conducted. In contrast to this approach, online control strategies are based on real-time controllers and do not require prior knowledge of the cycle. However, it is not ensured that they achieve the global optimum of a cycle. Pereira et al. [9] presented an online energy management system for an FCEV that is based on a recurrent neural network (RNN) that is able to predict the nonlinear dynamics of the FC and achieves higher efficiencies than a heuristic approach. The combination of the nonlinear model predictive control combined with an RNN uses operating conditions like temperature and load current as well as internal conditions as electrode flooding and membrane drying as input values.
A non-linear autoregressive neural network for an online energy management strategy is proposed by Zhou et al. [12]. They predict the velocity of an FCEV based on time-series analysis by means of the moving window method. The predicted velocity is in the next

step used to perform offline optimization strategies.
Other researchers train RNN for predicting the degradation of FCs. Zheng et al.[2] use an RNN with long short-term memory (LSTM) cells for the performance prediction of a PEMFC under dynamic conditions. Besides the polarization curve, the LSTM network predicts the performance degradation of the FC. A novel model called navigation sequence driven LSTM is proposed by Wang et al. [11], which is an advancement of standard LSTM in order to break the historical degradation data limitations. A comparison of different neural networks for the prediction of the remaining useful life of an FC is conducted by Long et al. [8] A gated recurrent unit network, a back propagation neural network, and an LSTM network deliver an effective prediction, as the understanding of the failure mechanism is not required. Similarly, an LSTM network is trained by Gu et al. [1] in order to set up a flooding fault diagnosis. The LSTM network models the fuel cell water state and is compared to a simple ANN without a memory. Another degradation model was introduced by Zuo et al. [13]. They predict the FC output voltage degradation based on long-term dynamic loading cycle durability test data.
In this work, LSTM neural networks are used to predict the voltage and efficiency of a 6 kW FC during dynamic operation. The development of this methodology is part of the research project "Development Platform 4.0" ("Entwicklungsplattform 4.0"). In this project, a wide and universal development platform is developed that is supported by diverse artificial intelligence (AI) techniques. The basis of the platform is a data management system that connects measurement with simulation tools and in-use data. This connection between real and virtual test scenarios is elementary for the generation of a wide and reliable data basis, which is required for the training of neural networks for predicting complex transient systems. The universal approach is applied to the mentioned PEMFC for which the test-bench measurements and simulation data are in the form of time series. The developed LSTM neural networks are multiple times faster than the phenomenological models that were used for the data generation. Furthermore, no licenses are required, which allows a high parallelization and the prediction of many possible operating points. This enabled the deployment of the models in many different applications during the entire life cycle of an FCEV: the verification of measured data from the test bench with a quick comparison with the neural network model that is implemented in the development platform. Additionally, the model supports the dimensioning and examination of single components of the FC system. Many variations are calculated within a short time.
The short calculation time is also beneficial for the offline optimization of the FC control system because many possible system applications are evaluated quickly in order to maximize the system's efficiency. It is also possible to implement the LSTM model directly on the fuel cell control unit (FCCU) and perform an online optimization. The buffer battery that is required for abrupt load changes involves another degree of freedom for the entire FC system, which makes its optimization more challenging. Dependent on the desired purpose, the architecture and inputs of the LSTM networks are adopted.
The remainder of this article is structured as follows. Firstly, the process of the data generation of the time series for the training of the LSTM network is described. In the next paragraphs, the data preparation and pre-processing are explained before the focus is set to the model set-up. The results are presented and discussed in the consecutive paragraphs and the paper ends with a short summary and brief outlook of future work within this project.

2. Fundamental approach of the methodology

To measure steady state operating points at the test bench, it requires time until the entire system is in steady state. In particular, thermal inertia has large time scales. Given that the measurement devices record the signals also between the changes of steady state operating points, a large date volume of time series is obtained. The steady state operating points are used to calibrate a FC simulation model within a multi-physics 0d/1d simulation. The measured data serves also as validation data for the calibrated simulation model. By means of the simulation, the scope of the boundary conditions is widened more easily and faster than on the test bench. Additionally, the ability to conduct several simulations at the same time accelerates the data creation process and extreme conditions are possible, e.g., very high or low ambient temperature. The test bench data and simulation data are collected and connected in the Development Platform 4.0. It is essential, that the position of the sensors on the test bench correspond accurately that of the virtual sensors of the simulation model to ensure the transferability of the data. The connection of the data facilitates the utilization of both data roots for the training of an ANN. The schematic process of the data generation and utilization is sketched in Figure 1.

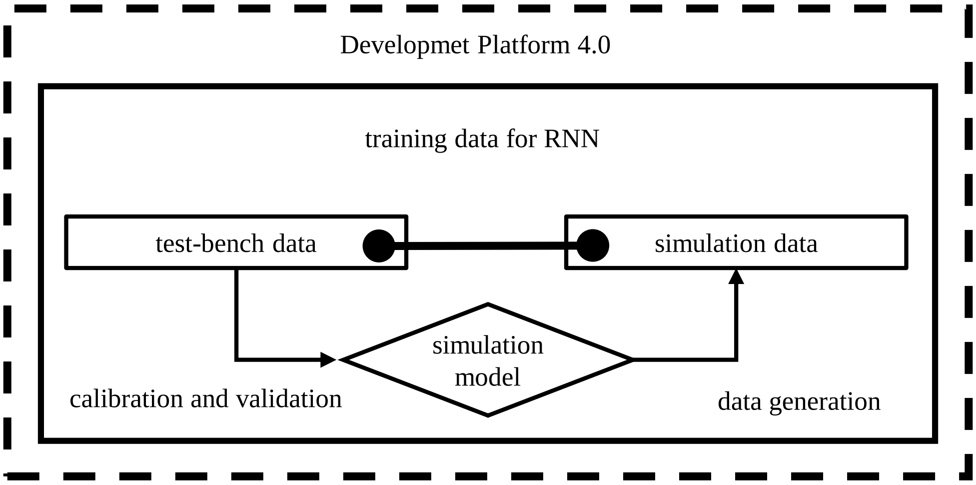

Figure 1. Schematic chart of the data process

The PEMFC efficiency depends on various controllable factors but also some that are not changeable as for example the ambient temperature. In different simulations, the parameters are varied corresponding to Table 1. Different load profiles are generated automatically, that define the current over the time. Some of the load profiles focus the low load, others on high load operation to cover a wide operating area. All simulations are started at ambient temperature to implement the warm-up and gather enough data of the thermal behaviour of the FC.

Table 1. Variated parameters and their range

Parameter	Min. value	Max. value
Ambient temperature in °C	-20	50
Load (Current) in A	0	400
Anode and cathode stoichiometric ratio in fraction	1	5
Anode and cathode pressure in bar	1	3
Target membrane relative humidity in %	60	90
Ambient relative humidity in %	20	80

2.1. Data preparation

this work concentrated on the training with data generated by the simulation. Approximately 35 h of data was gathered and analysed in approximately 160 simulation runs. The export of the single simulations is standardized and always follows the same pattern to automatize the following steps of the data preparation and pre-processing.

To guarantee a sufficient data quality the training data is prepared before the data is brought in form during the pre-processing. For all parameters an upper and lower boundary is defined within which their value must lie. This avoids, that unplausible data points are taken for the following steps.

The fixed values for the upper and lower values are necessary because classic methods of outlier detection did not achieve the desired outcome. Many correct values are detected as outliers; exceptionally often, the values of the stoichiometric ratios are falsely detected as outliers, because their value reaches very high or low values for a short time at the event of an abrupt load change. The fix boundary of this value needs to be set appropriately.

2.2. Data pre-processing

If the frequency of the raw data is higher than the desired frequency of the time series for the training of the LSTM network, the raw data is averaged over the desired time interval. In this manner, the time step between two data points is normalised for all test bench measurements and simulation runs. If the time step between two raw data points is too long, or data points were removed in the data preparation process, a na value is set to indicate the gap for the following process.
The so prepared time series are split into windows with a certain length and a specified overlap. Finding the optimal window length is part of the investigation. If a window contents a na value it is not used, but the next window starts with the following data point. The idea of the mapping of one output value to windows of three input values is exemplary outlined in Figure 2. The output is always the last time point of the data window.

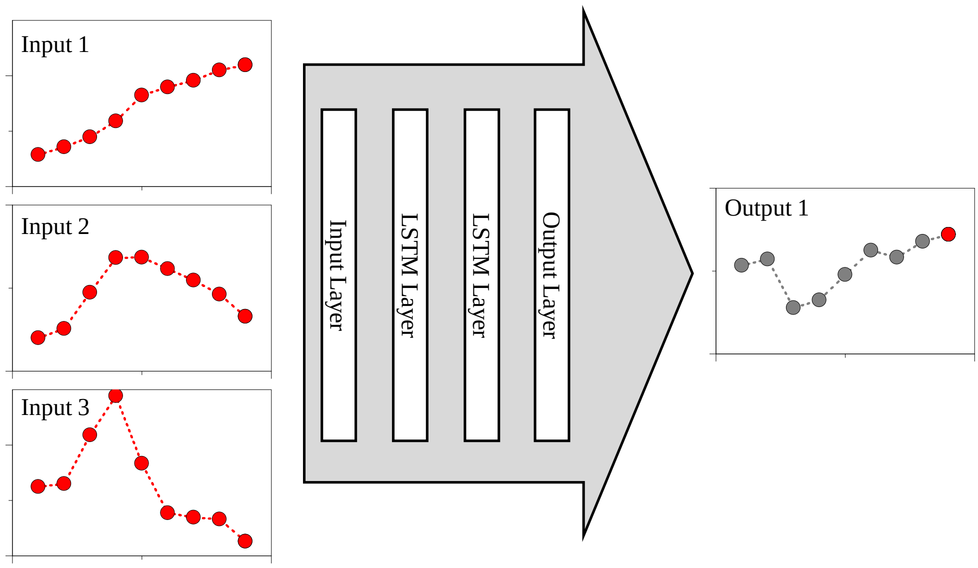

Figure 2. Shape of the input windows and output single value

Generally, the data pre-processing is distinguished between two LSTM training methods: in the stateless training the internal state of the LSTM cells is reset after each batch, while the stateful method keeps the internal state of the cell, until it is reset externally. This influences the pattern of the required input data of the time series. For the utilization of a stateful LSTM network, the single windows must stay in the same order and have the maximum overlap, that two consecutive windows are one time step apart. The stateless LSTM network is trained with the windows only. This means that the order of the windows is independent of the entire simulation run from which they originate. The training was improved, when all windows were shuffled and made independent from their original simulation run.
The same distinction is made for the split between training, validation, and test data. When a stateful LSTM network is applied, the split is made between entire simulation runs. For the training of the stateless network, the windows are allocated to the training, validation, or test data independently of the original simulation run. The total amount of all time series is split into 50 % training, 30 % validation and 20 % test data. The input and output data are scaled in a manner that all values are between -1 and 1. This prevents that different

input values have a large difference in their dimensions and leads to more accurate results. According to [6] the most important influences for the FC operation are current density, membrane humidity, stack temperature, average pressure at the cathode, and stoichiometric ratio at the cathode. These are complemented by the anode pressure for the presented results of the FC prediction.

2.3. Model setup and training

The architecture of the neural network depends strongly on the field of application of the predicting network. For the prediction of the voltage and efficiency of the FC, a LSTM neural network is set up. The LSTM cells were first proposed by Hochreiter and Schmidhuber [3] and are a conjunction of a RNN with a gradient-based learning algorithm. The neural network has a input layer, two or three LSTM layers and an fully connected output layer, as displayed in Figure 2.
In the previous section the differences between LSTM stateful and stateless in the Tensorflow environment are described. This also influences the training process: with stateful LSTM the internal states are reset after each entire time series. The reset is directed as well for the validation and evaluation process. The Huber loss function [4] is calculated after each training epoch to determine the current state of the training. If the validation loss is not decreasing for a certain number of training epochs, the learning rate is reduced. If for another certain number of training epochs no further improvements are achieved, the training is stopped, and the evaluation is conducted. The root mean squared error (RMSE) is calculated for the test data at the end of the training in order to get a value for the comparison of the network's performance. The influence of the hyperparameter, input parameters and other settings of the training are compared in this way. During the training, the network does not see any of the test data, only the training data for the tuning of the weights and the validation data for the assessment of the current state after each epoch. Thus, the evaluation with the test data shows the behaviour of the network with unknown data sets. The architectures of the most accurate LSTM networks are summarized in Table 2. Additionally, a LSTM network with the window length of one is trained to show the differences between pointwise and time series input data. For the optimization of the weights, an algorithm for first-order gradient-based optimization, based on adaptive estimates of lower-order moments is used, also known as ADAM optimizer [5] .

Table 2. Parameters of the LSTM networks that are compared with each other

Name	Number of LSTM layers	Size of LSTM layers	Window length
Stateful LSTM 8	2	50	8
Stateless LSTM 16	3	50	16
Stateless LSTM 1	3	50	1

3. Results

The results of different structures of the LSTM networks are demonstrated by means of a dynamic load for the current that was generated during a Worldwide harmonized Light vehicles Test Cycle (WLTC). This standardized cycle was developed to assess the fuel consumption of vehicles and approaches a characteristic driving cycle over a large range of velocities. It is consequently significant for the evaluation of the LSTM networks. To avoid, that the networks are trained explicitly on this cycle, the data of the WLTC was excluded from the training data. Figure 3 illustrates the predicted voltage over the time of the LSTM networks and the target value calculated by the detailed simulation model with the dashed red line. Additionally, the input of the current density is plotted. The dependence between voltage and current is clearly visible. However, the voltage depends also on other parameters. This which becomes visible in the constant current sections, during which the voltage is initially lower than at the end of the cycle, because the stack temperature has risen.

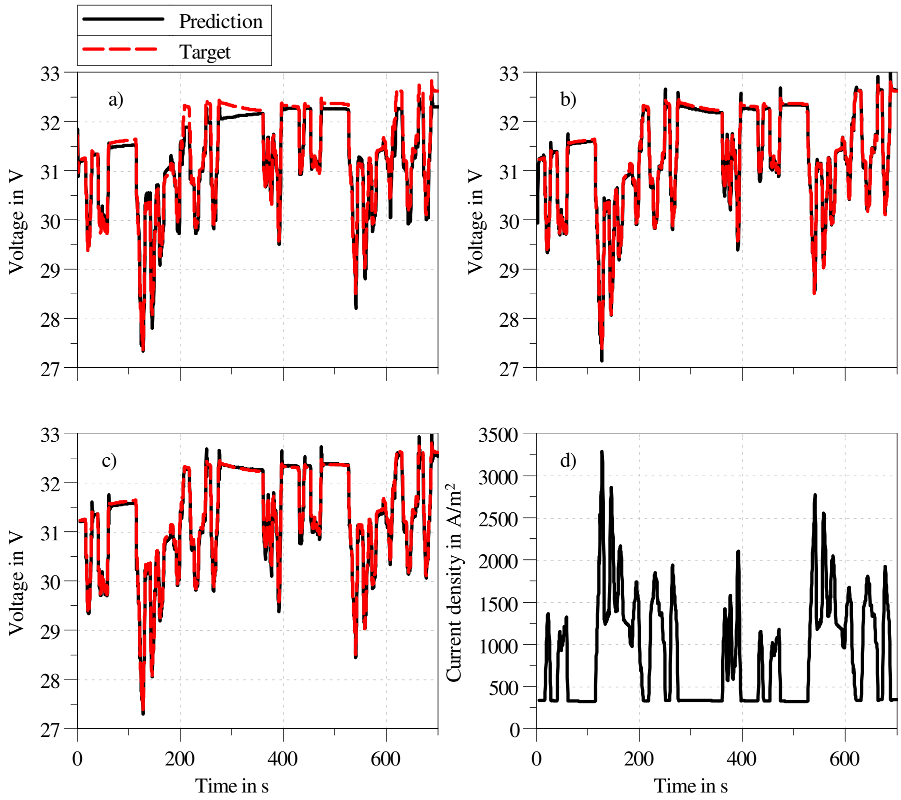

Figure 3. Predicted and targeted voltage of the examined LSTM networks of the test cycle (**a**) Stateless LSTM 1 (**b**) Stateful LSTM 16 (**c**) Stateless LSTM 16 (**d**) current density

The deviations between the predicted and targeted values are highest for a window size of one. This highlights that the evaluation of the current state at a single time step is inaccurate during dynamic operation. The largest deviations to the target occur at the beginning of the constant current density parts and at the current peaks, respectively voltage lowest point. The reason for this is the delay of the oxygen concentration at the cathode. When the current is increased, more oxygen is required. However, from the previous low load part, an oxygen excess is apparent that is consumed, before the voltage drops further. This phenomenon is more accurately considered when using longer windows. All networks overpredict the voltage when the current is decreasing to a minimum. The reason for this behaviour was not found.
Since the graphs only show a small section of the entire test data, the networks are analysed by means of the entire test data set. This evaluation is summarized in Table 3, where the mean absolute error, standard deviation and 99 % respectively 99.9 % percentile is listed for all variants.

Table 3. Performance comparison of the networks for all test data: absolute error for the voltage prediction

Error	Stateful LSTM 8	Stateless LSTM 16	Stateless LSTM 1
Mean	0.166	0.125	0.187
Standard	0.260	0.375	0.440
99 % percentile	1.264	1.143	1.780
99.9 % percentile	2.786	3.901	4.234

The stateless networks with a window size of 16 predicts the voltage more accurately than the stateful network. However, the differences are small and the stateful LSTM network shows better results with a shorter window size. This indicates that the time scales of the essential physical phenomena of the FC are relatively short, and that the long-term memory does only improve the results when the window size is short. The methodology is conducted identically for the efficiency prediction. Here, the comparison of the networks shows the same result.

4. Discussion

The operation of a FC is characterised by various physical effects with different time scales: The electrolyte relaxation occurs within microseconds, the passing of electrons from the electrode to the active species in the electrolyte within milliseconds and the diffusion processes are slow with time scales of seconds up to minutes [6]. The diffusion processes are mainly covered by the networks that analyse a various data points from the past. The mentioned fast processes occur during less than one time step and are thus included within small window sizes.

The stateless LSTM network that was trained with a window size of 16 includes these effects most accurately and shows the lowest errors. It is probable that with the inclusion of the FC degradation, the stateful LSTM network will show benefits like in [2] [13]. The 99 % percentile of stateless LSTM 16 underline, that there are only few test points, for which the prediction deviates far from the targeted value. The 99.9 % percentile shows, that the differences between Steteful LSTM 8 and Stateless LSTM 16 are only small, since the stateful network achieves a better value. However, the deviation of more than 2.7 V for 0.1 % of the test data is significant. Additionally, it has to be mentioned, that even the simulation data can still contain false and unplausible data points that need to be detected and filtered out. Hence, a good pre-processing is essential in order to guarantee a high data quality for the training. Besides the quality, the number of data points of the parameters should be distributed evenly to get better results. Figure 4 illustrates the distribution of four input values. It is clearly visible, that for the presented parameters, the number of points vary strongly. Most data points lie in the area of the typical operation conditions, but the aim is to create a universal model for all boundary conditions. For example, most operating points have a membrane humidity between 80 and 95 %. Outside this range, more data probably lead to a more precise prediction.

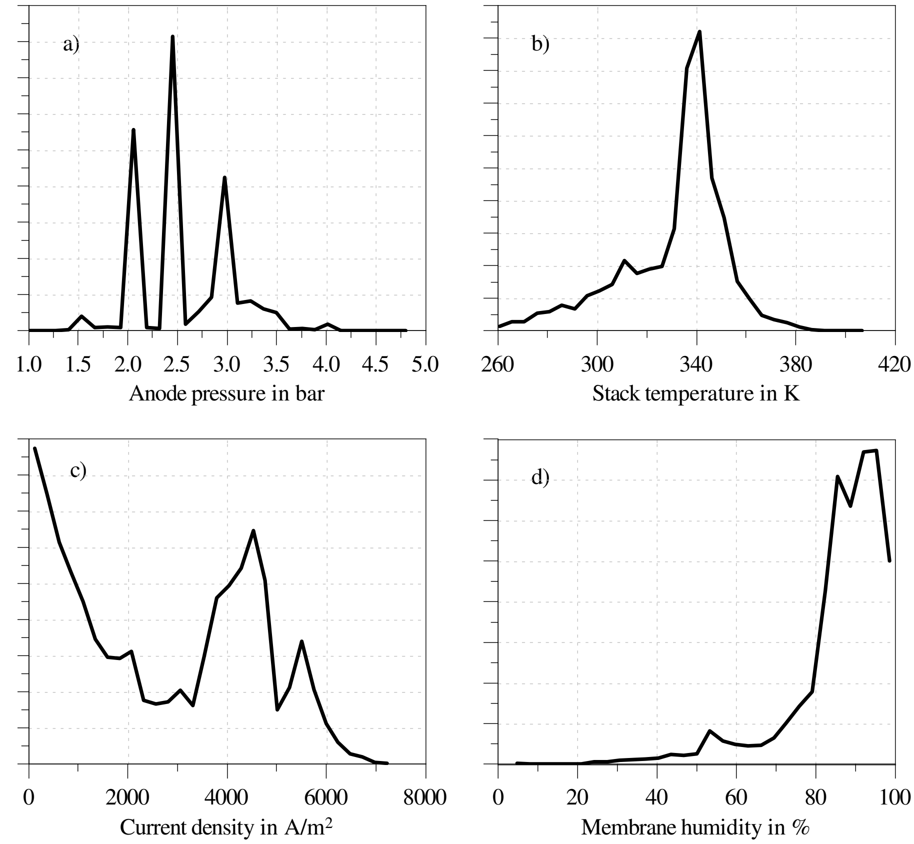

Figure 4. Distribution of selected input parameters (**a**) Anode pressure (**b**) Stack temperature (**c**) Current density (**d**) Membrane humidity

Since the training data is based on a simulation model, the network can only achieve the same accuracy as the simulation model and not all physical effects are covered, for example effects that occur during the excessive formation of water.
However, the calculation time is reduced to a minimum: for the entire cycle with a time step of 0.25 s the calculation time is less than 3 s on one CPU core. This corresponds to a real-time factor of approximately 0.005 and demonstrates the celerity of the LSTM model. Since the simulation model includes the entire FC system, a comparison regarding the simulation time is not suitable.

5. Conclusion

A methodology is developed for creating dynamic training data for the training of LSTM networks. The data is prepared to improve its plausibility and cut into windows of a specific length. These windows are the input for the LSTM networks to predict the voltage and efficiency of a fuel cell under dynamic operation. The results of three different layers are presented and compared to each other. Furthermore, the physical effects that induce a delay are explained briefly. These effects are represented more precisely by a time series analysis compared to single points. The LSTM networks show great potential regarding the short calculation time. However, the variation of hyperparameters is not completed yet and more accurate results may achievable while keeping computational time short.
The presented methodology enables the complementation of the simulation data by test-bench data. Furthermore, a combined model for the entire PEMFC system is conceivable, that consists of and combines LSTM networks for each component of the system.

Funding: This research was funded by the German Federal Ministry for Economic Affairs and Climate Action grant number 19I21020C.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ADAM	Adaptive estimates of lower-order moments
AI	Artificial intelligence
ANN	Artificial neural network
BEV	Battery electric vehicles
FC	Fuel cell
FCCU	Fuel cell control unit
FCEV	Fuel cell electric vehicle
LSTM	Long short-term memory
MSE	Mean squared error
PEM	Proton exchange membrane
PEMFC	Proton exchange membrane fuel cell
PLM	Product lifecycle management
RMSE	Root mean squared error
RNN	Recurrent neural network
WLTC	Worldwide harmonized Light vehicles Test Cycle

References

1. Gu, X.; Hou, Z.; Cai, J. Data-based flooding fault diagnosis of proton exchange membrane fuel cell systems using LSTM networks. *Energy and AI* **2021**, *4*, 142–149.
2. Zheng, L.; Hou, Y.; Zhang, T.; Pan, X. Performance prediction of fuel cells using long short–term memory recurrent neural network. *International Journal of Energy Research* **2021**, *6*, 9141–9161.
3. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural computation* **1997**, *8*, 1735–1780.
4. Huber, P. J. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics* **1964**, *35*, 73–101.
5. Kingma, D. P., Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2015**, *35*, 73–101.
6. Kurzweil, P. *Brennstoffzellentechnik*, 3rd ed.; Publisher: Springer Fachmedien Wiesbaden, Germany, 2016; pp. 98–101.
7. Larminie, J.; Dicks, A. *Fuel cell systems explained*, 2nd ed.; Publisher: John Wiley & Sons Ltd Chichester, England, 2003; pp. 68-69.
8. Long, B.; Wu, K.; Li, P.; Li, M. A Novel Remaining Useful Life Prediction Method for Hydrogen Fuel Cells Based on the Gated Recurrent Unit Neural Network. *Applied Sciences* **2022**, *12*, 432.
9. Pereira, D. F.; Da Lopes, F. C.; Watanabe, E. H. Nonlinear Model Predictive Control for the Energy Management of Fuel Cell Hybrid Electric Vehicles in Real Time. *IEEE Transactions on Industrial Electronics* **2021**, *68*, 3213–3223.
10. Sulaiman, N.; Hannan, M. A.; Mohamed, A.; Majlan, E. H.; Wan Daud, W. R. A review on energy management system for fuel cell hybrid electric vehicle: Issues and challenges. *Renewable and Sustainable Energy Reviews* **2015**, *52*, 802–814.
11. Wang, C.; Li, Z.; Outbib, R.; Dou, M.; Zhao, D. A novel long short-term memory networks-based data-driven prognostic strategy for proton exchange membrane fuel cells. *International Journal of Hydrogen Energy* **2022**, *47*, 10395–10408.
12. Zhou, D.; Gao, F.; Ravey, A.; Al-Durra, A.; Simoes, M. G. Online energy management strategy of fuel cell hybrid electric vehicles based on time series prediction. In *2017 IEEE Transportation Electrification Conference and Expo (ITEC)*; 2017; pp. 113–118.
13. Zuo, J.; Lv, H.; Zhou, D.; Xue, Q.; Jin, L.; Zhou, W.; Yang, D.; Zhang, C. Deep learning based prognostic framework towards proton exchange membrane fuel cell for automotive application. *Applied Energy* **2021**, *281*, 115937.