PNNL-29124

Physics-Informed Machine Learning with Application to Solid Oxide Fuel Cell System Modeling and Optimization

September 2019

Jie Bao
Chao Wang
Zhijie Xu
Brian J. Koeppel

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights**. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <https://www.ntis.gov/about>
Online ordering: http://www.ntis.gov

PNNL- 29124

Physics-Informed Machine Learning with Application to Solid Oxide Fuel Cell System Modeling and Optimization

September 2019

Jie Bao
Chao Wang
Zhijie Xu
Brian J. Koeppel

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Summary

Traditional data-driven machine learning (ML) techniques were combined with the physics-based SOFC-MP model toward improving SOFC system-level performance prediction. Four different physics-informed ML methods were proposed to demonstrate better prediction accuracy compared with traditional ML methods. First, it was preferred that the existing Kriging-based surrogate model not be entirely abandoned, so errors between Kriging regression prediction and true solution are learned using ML. The goal was to find the pattern of the error distribution as a function of model input so discrepancies between the Kriging prediction and true solution could be reduced by adding the additional ML-trained error term to the existing prediction. Second, deep neural networks (DNN) coupled with the mass balance model (MBM) have been developed to significantly decrease the reduced order model (ROM) prediction error with fewer training data compared to the traditional Kriging-based ROM and traditional DNN regression approaches. Third, DNN regression coupled with a neural networks (NN) classifier and MBM was developed. It provides superior performance on classifying the physically operational conditions for natural gas fuel cell (NGFC) than traditional classification approaches. Finally, two kinds of pre-trained DNN transfers were studied and tested. One was transferring a pre-trained DNN for low-fidelity model data to a high-fidelity model. The second one was transferring a pre-trained DNN from state-of-the-art NGFC to advanced NGFC. These physic-informed ML frameworks can serve as a solid foundation for estimating U.S. Department of Energy-targeted fuel cell performance and exhibit great potential to be employed to other engineering applications.

Acknowledgments

This research is supported by FY2019 PNNL EED LDRD.

Acronyms and Abbreviations

CCS	carbon capture and storage
CH_4	methane
CO_2	carbon dioxide
DNN	deep neural networks
DT	decision tree
MBM	mass balance model
ML	machine learning
MWe	megawatt electric
NGFC	natural gas fuel cell
NN	neural networks
O/C	oxygen to carbon ratio
O_2	oxygen
PNNL	Pacific Northwest National Laboratory
RF	random forest
RMSE	root-mean-square-error
ROM	reduced order model
SOA	state-of-the-art
SOFC	solid oxide fuel cell
SOFC-MP	SOFC Multi-Physics
SVM	support vector machine

Contents

Summary ii
Acknowledgments ii
Acronyms and Abbreviations ii
1.0 Introduction 1
2.0 Fuel Cell Model for Physics-Informed ML Integration 2
3.0 Modeling Results and Discussion 4
3.1 Physics-Informed ML Based on Existing Surrogate Model 4
3.2 DNN Coupled with Mass Balance Model 6
3.3 DNN Regression Coupled with NN Classifier and Mass Balance Model 7
3.4 Transferred DNN 8
3.4.1 Transferring DNN for Different Fidelity Model Data 8
3.4.2 Transferring DNN from State-of-the-Art NGFC to Advanced NGFC 9
4.0 Conclusions 10
5.0 References 10

Figures

Figure 1. NGFC system schematic chart 2
Figure 2. Flow chart to compare three different surrogate models. 4
Figure 3. Comparison between three different model predictions using a different number of training datasets. 5
Figure 4. Probability density function for three model predictions and true solutions 6
Figure 5. Results for physics-informed ML tool to construct enhanced fidelity ROM for NGFC. 7
Figure 6. Comparison between different classifiers for determining if the fuel cell can be physically operational given a set of input parameters. 8
Figure 7. Prediction error for MBM pre-trained DNN and comparison to other approaches. 9
Figure 8. Cell voltage RMSE comparison between fresh and transferred model DNN for advanced NGFC system using different numbers of training data 10

Tables

Table 1. Model input parameters, ranges, and descriptions. 3
Table 2. Model output parameters that can be used for system designs. 3

1.0 Introduction

Machine learning (ML) is rapidly developing and very popular, but successful examples for solving practical science, engineering, or industrial problems are rare. Therefore, we proposed to develop an approach of applying ML to address such problems and used SOFC system modeling and optimization as a demonstration. PNNL developed the SOFC-MP (SOFC Multi-Physics) software (Koeppel et al. 2011; Lai et al. 2011) for the U.S. Department of Energy (DOE) fuel cell program to provide high-fidelity estimates for fuel cell stack performance, but it is not computationally efficient enough when integrated into power system-scale models. There are over 50 input parameters in the stack model that potentially impact the fuel cell operation and reliability, and there are over 40 performance outputs to be considered for power system model interactions. Identification of the influence of key parameters and parameter combinations has been unsuccessful using manual or traditional design–of-experiments techniques and regression models. In addition, the current Kriging-based surrogate model for SOFC performance prediction encountered some shortcomings (Pan et al. 2013). The discrepancy between the surrogate model prediction and true solution from SOFC-MP usually decreases with increased training set size. However, for some outputs, this discrepancy in the current SOFC surrogate cannot decrease below ~5%–10%, even with a further increase in the training set size. Currently, there is no well-developed method to further decrease this discrepancy.

The targeted DOE sponsors, such as Office of Fossil Energy (FE) and Office of Energy Efficiency & Renewable energy (EERE), have shown strong interests in having highly accurate models for fuel cell system design and optimization. ML could achieve this vision, but currently there are no existing ML approaches for such applications. DOE targets (e.g., ~10% reduced cost for energy generation or ~5% increased fuel utilization) are hard to achieve using the current surrogate model with a maximum error of ~5%–10%. Because the stack model is the core piece for the industrial system-scale model, the error in the stack model is usually amplified significantly while propagating through the system model. Therefore, the estimated fuel cell system error must be decreased to <1% to provide reliable stack model prediction.

To resolve all the aforementioned issues, we proposed to integrate data-driven ML techniques with the physics-based SOFC-MP model for more efficient and accurate surrogate model prediction. The outcome of this high-fidelity surrogate model will be employed for DOE high-visibility 'pathway studies' demonstrating how programmatic cost-of-energy targets can be achieved and how research thrusts should be prioritized to realize the needed technology advancements. Four main ML approaches have been developed for demonstration of the applicability of SOFC. (1) It was preferred that the existing Kriging-based surrogate model was not entirely abandoned, so errors between Kriging regression prediction and true solution can be learned using ML. The goal was to find the pattern of the error distribution as a function of model input so the discrepancy between Kriging prediction and true solution can be reduced by adding the additional ML-trained error term to the existing prediction. (2) Deep neural networks (DNN) coupled with the mass balance model (MBM) was developed to significantly decrease the reduced order model (ROM) prediction error with fewer training data compared to the traditional Kriging-based ROM and traditional DNN regression approach. (3) DNN regression coupled with a neural network (NN) classifier and MBM is developed. It provides superior performance on classifying the physically operational conditions for the natural gas fuel cell (NGFC) than traditional classification approaches. (4) Two kinds of pre-trained DNN transferring were studied and tested. One is transferring DNN for low-fidelity model data to a high-fidelity model. The second one is the transferring DNN from state-of-the-art NGFC to advanced NGFC.

The advanced NGFC refer to the fuel cell with next generation electrodes materials, whose activation loss and Ohmic loss is 50% of the SOA NGFC's electrodes. The proposed physics-informed ML framework will not only be used in SOFC modeling, but also will serve similar functions in other science and engineering applications.

2.0 Fuel Cell Model for Physics-Informed ML Integration

A system model of a 550 MWe NGFC system with carbon capture and storage (CCS) developed using the Aspen Plus® platform is considered in the present study to test the ROM integration. Noteworthy system features include an atmospheric SOFC stack, an oxygen driven auto-thermal natural gas reformer, an oxy-combustor to recover unutilized anode-exhaust chemical energy, a steam bottoming cycle to recover stack exhaust heat energy, and a purification unit to provide a carbon dioxide (CO_2) stream suitable for sequestration. The auto-thermal reformer provides a prescribed degree of external reformation of the pipeline natural gas, whose composition is primarily methane (CH_4) (~93% molar fraction in the present case), to modulate the amount of endothermic on-cell reforming. An anode gas recirculation loop is included to maintain a suitable oxygen (O_2) to carbon ratio (O/C) to prevent carbon formation in addition to providing the steam required for internal reformation. Cathode exhaust recirculation is provided to not only limit the fresh air flow to increase overall system efficiency but also to lower the cathode side recuperator (Cathode HTX in Figure 1) size and costs. Further details about the system and modeling approach can be found in Iyengar et al. (2014).

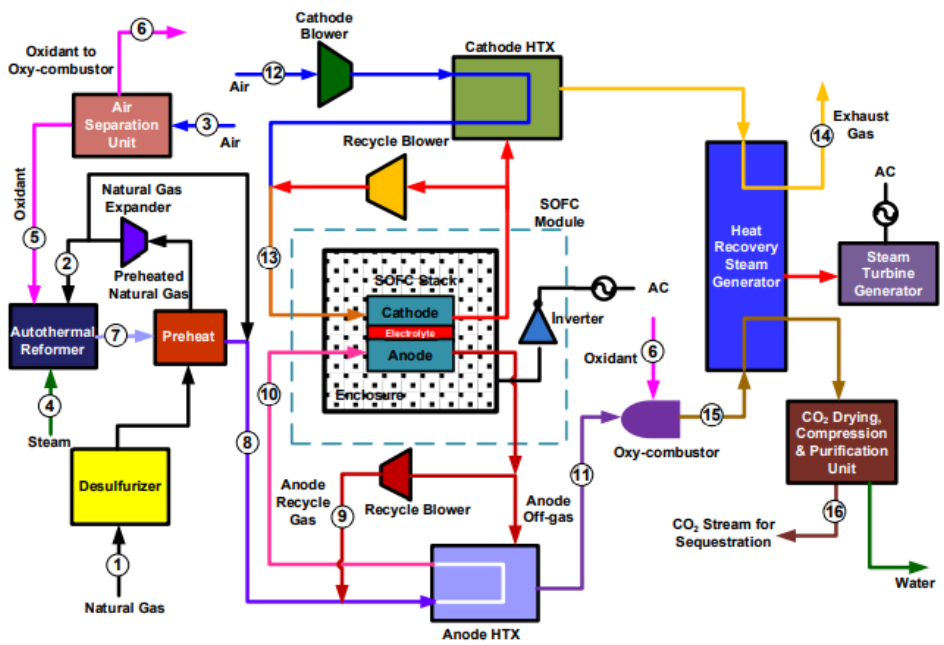


Figure 1. NGFC system schematic chart (Source DOE/NETL).

The model input parameters and ranges used for this work are listed in Table 1. Based on the selected current density, pressure, internal reforming, recirculation fractions, and utilizations, a material flow balance calculation is performed for the system configuration in Figure 1 to determine the fuel and oxidant compositions at the stack inlet. These compositions, combined with the sampled inlet temperature values, thereby define the individual case to be solved for the fuel cell stack. It is assumed that the cell geometry is fixed, and only operational conditions

are changed, but the approach is flexible enough that other geometric parameters (e.g., the number of cells in the stack) could be input variables for a design study if desired. Various model output parameters were selected for the study including the voltage, fuel/oxidant exhaust temperatures, cell maximum temperatures, cell temperature difference, and fuel/oxidant species. Voltage is one of the important outputs because it determines the stack electrical power in the system model based on the current density input variable. The maximum cell temperature is another output of interest because it defines the allowable thermal operating range to avoid cell material degradation. Table 2 lists all the modeling output parameters that can be used for the system designs.

Table 1. Model input parameters, ranges, and descriptions.

Variables	Ranges	Description
Average Current Density	2000–6000 A/m^2	Cell operating current
Fuel Temperature	15–60°C	Fuel temperature at stack inlet
Internal Reforming	0–1	Prescribes the degree of external reformation
Oxidant Temperature	550–800°C	Oxidant temperature at stack inlet
Oxidant Recirculation	0.0–-0.8	Fraction of cathode exhaust recirculated
Oxygen-to-Carbon Ratio	1.5–3.0	Defines fuel exhaust recirculation needed to achieve the desired O/C ratio at stack inlet
Stack Fuel Utilization	0.4–0.95	Overall fuel utilization
Stack Oxidant Utilization	0.125–0.833	Overall oxidant utilization
Pressure	1-5 atm	Cell operating pressure

Table 2. Model output parameters that can be used for system designs.

Output variables	
Stack Voltage	Outlet Fuel CH_4
Avg Cell Voltage	Outlet Fuel N_2
Max Current Density	Outlet Air Flowrate
Min Current Density	Total Power
Avg Cell Temperature	Air Enthalpy Change
Max Cell Temperature	Fuel Enthalpy Change
Min Cell Temperature	Electrical Efficiency
Delta Cell Temperature	Stack Efficiency
Outlet Fuel Temperature	Fuel Stack Inlet Temperature
Delta Fuel Temperature	Fuel Stack Inlet Flowrate
Outlet Air Temperature	Fuel Stack Inlet H_2 Molar Fraction
Delta Air Temperature	Fuel Stack Inlet H_2O Molar Fraction
Air Heat Exchanger Effectiveness	Fuel Stack Inlet CO Molar Fraction
Fuel Utilization	Fuel Stack Inlet CO_2 Molar Fraction
Outlet Fuel Flowrate	Fuel Stack Inlet CH_4 Molar Fraction
Outlet Fuel H_2	Fuel Stack Inlet N_2 Molar Fraction
Outlet Fuel H_2O	Fuel Temperature after Mix
Outlet Fuel CO	Fuel Temperature before Gibbs Reactor
Outlet Fuel CO_2	Fuel Heat Exchanger Effectiveness

3.0 Modeling Results and Discussion

3.1 Physics-Informed ML Based on Existing Surrogate Model

Pacific Northwest National Laboratory (PNNL) has spent tremendous efforts to develop a Kriging-based surrogate model for SOFC stack performance prediction and integration into higher level system models. Therefore, ML should be used to correct/improve existing models, not to entirely replace them. Here, we propose to learn the surrogate model prediction and true solution discrepancy, not the surrogate model output directly. The step-by-step illustration of the physics-informed ML framework is listed below.

1. Attain prediction from Kriging-based surrogate model
2. Get the error between the true solution and the prediction from Kriging
3. Train the error from Step 2 using ML and find the pattern
4. Physics-informed ML prediction = Kriging prediction + ML trained error

By doing this, we expect the new prediction to be closer to the true solution. After this physics-informed ML is developed, comparisons can be made among three different surrogates, i.e., Kriging, traditional ML, and physics-informed ML. The flow chart of the comparison is shown in Figure 2.

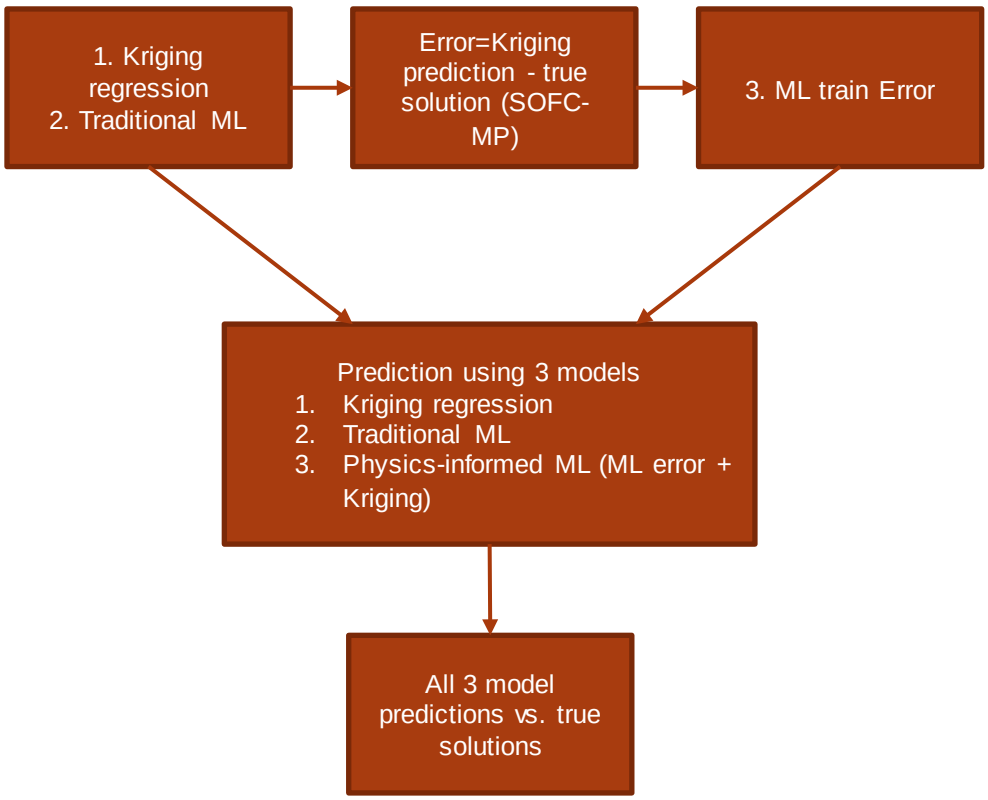


Figure 2. Flow chart to compare three different surrogate models.

For Kriging regression, details can be found in Koeppel et al. (2017). For traditional ML, a DNN is built to predict SOFC performance. The DNN contains a total of four layers and the number of units in each layer are 32, 200, 200, and 256, respectively. There are 9 total input parameters (same as listed in Table 1) and 48 output parameters, which are identical to the input/output parameters in the Kriging-based surrogate model. For physics-informed ML prediction, another DNN is built containing four layers and the number of units in each layer are 32, 200, 200, and 256, respectively. There are nine total input parameters (same as listed in Table 1) and only one output; which is the difference between the Kriging prediction and true solution. Figure 3 shows the comparison results for three different methods. The prediction parameter is maximum cell temperature. The x-axis shows the number of different training data, while the y-axis is the L1 Norm, which is the absolute error between predictions and true solutions. It is obvious that Kriging predictions always exhibit the largest errors, which reflects a worst-prediction accuracy. In addition, our physics-informed ML shows better prediction accuracy compared with traditional ML, and the improvement on accuracy increases with the number of training datasets.

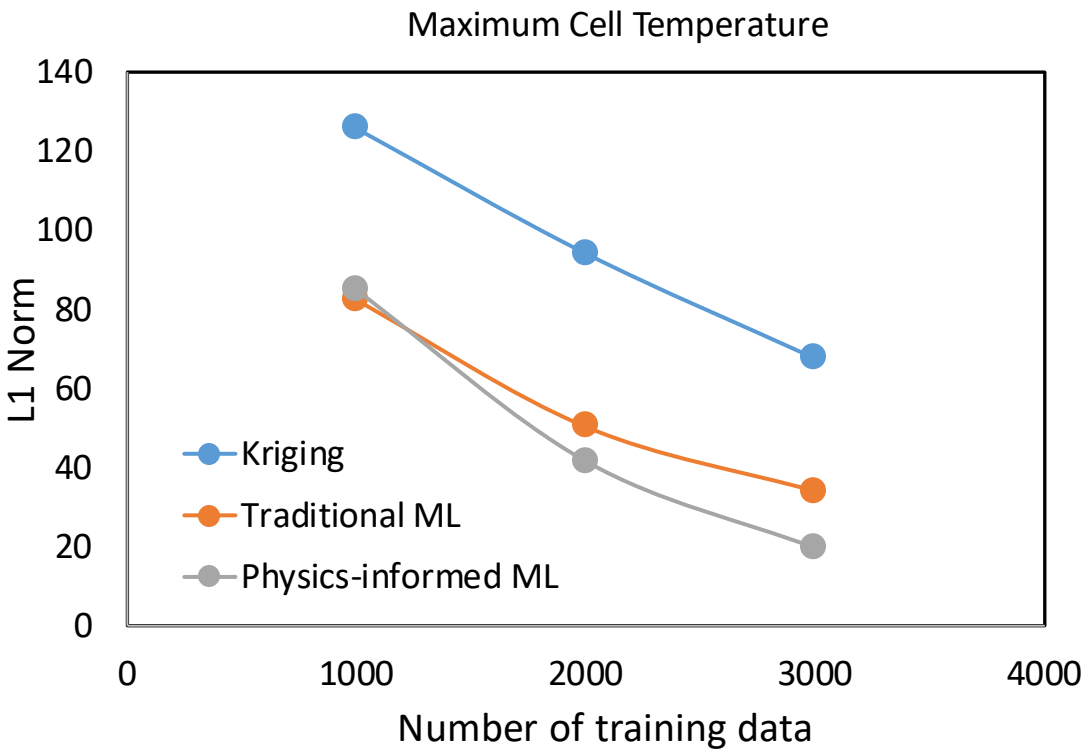


Figure 3. Comparison between three different model predictions using a different number of training datasets.

The same methodology is also implemented to predict cell voltage, and the results are shown in Figure 4. In this case, a fixed 2000 case training dataset is employed, and voltage prediction is plotted in a form of a probability density function. In Figure 4, blue represents true solution, red represents traditional ML prediction, yellow represents Kriging prediction, and purple represents physics-informed ML prediction. It is observable that physics-informed ML predictions are very close to, and almost overlap, the true solution demonstrating its best prediction accuracy compared with other two prediction methods. Again, Kriging prediction shows the worst prediction accuracy.

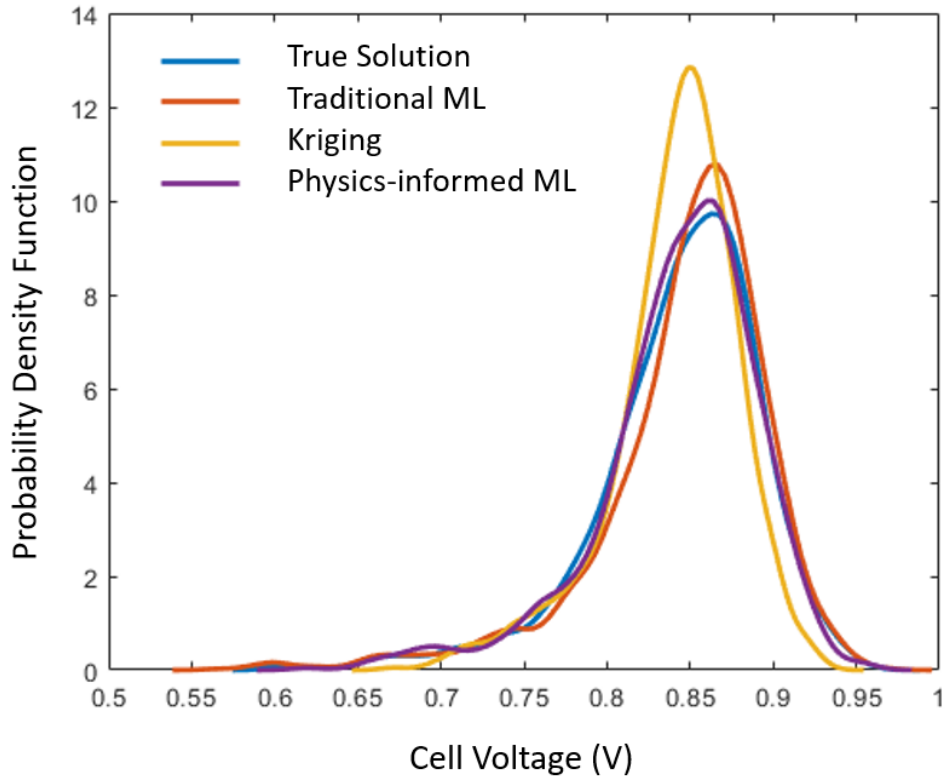


Figure 4. Probability density function for three model predictions and true solution.

3.2 DNN Coupled with Mass Balance Model

A physics-informed ML tool to construct an enhanced fidelity ROM for the NGFC has been developed. It couples the DNN and a simple MBM to significantly decrease the ROM prediction error with fewer training data compared to the traditional Kriging-based ROM construction. The MBM calculates the actual fuel and oxidant consumption in the fuel cell based on the requested current. Because of methane reforming within the NGFC stack, the MBM cannot accurately estimate the species fraction. With the assumption that methane is always consumed first in the cell, MBM can provide a good approximation almost instantly, which can then be used to improve the DNN regression model accuracy. Figure 5(a) and (b) show the comparison between the SOFC-MP simulation and the MBM approximation for the H_2 and H_2O fraction at the fuel cell anode outlet. The MBM can capture the main trend of the simulation results but has considerable deviations. Figure 5(c) and (d) show the comparison between different ROMs. 'MBM+DNN' refers to the ROM using DNN to learn the deviations between MBM and SOFC-MP results for different input parameters. 'DNN' refers to the ROM directly using DNN to learn the relationship between input and output parameters. It shows that both DNN and MBM+DNN can provide much lower prediction error than the traditional Kriging method. MBM+DNN can further decrease the prediction error more than DNN with much fewer training data.

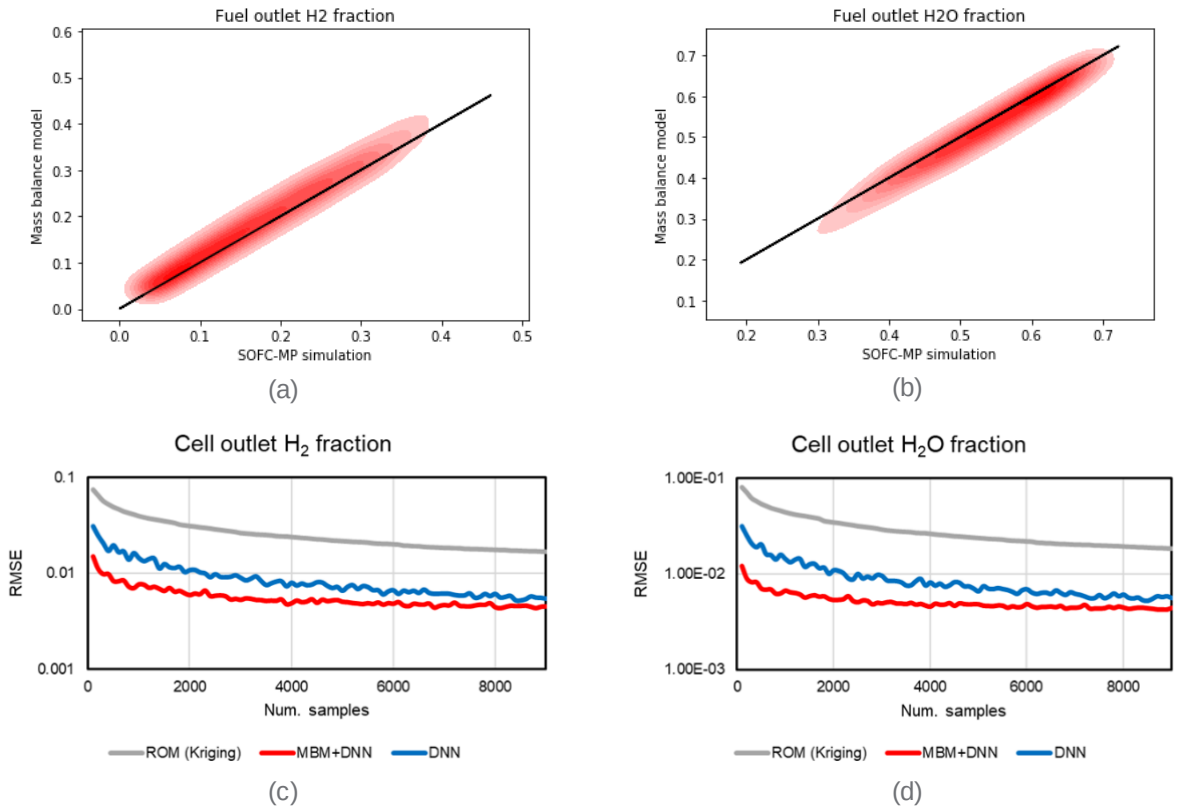
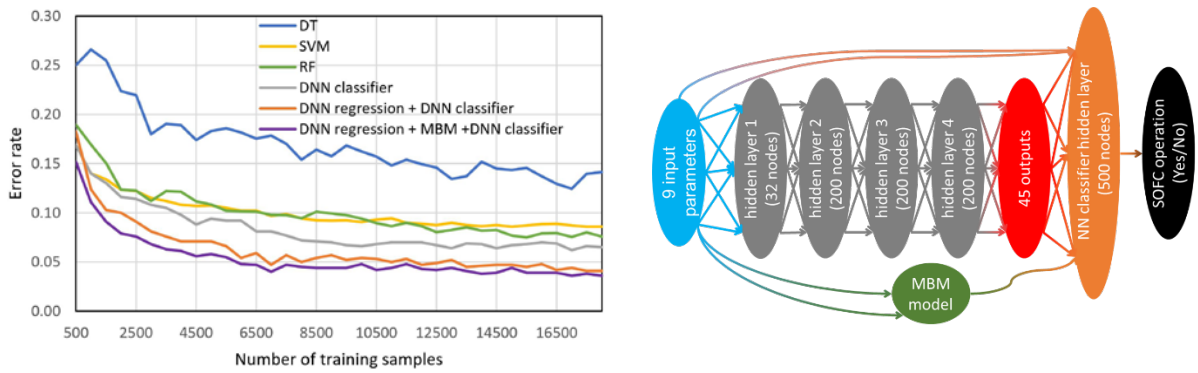


Figure 5. Results for physics-informed ML tool to construct enhanced fidelity ROM for NGFC.

3.3 DNN Regression Coupled with NN Classifier and Mass Balance Model

It was noticed that not all input parameter combinations can make the fuel cell physically operational. Therefore, before using the ROM to predict the outputs, it is necessary to use a classifier to assess the feasibility of successful fuel cell operation. Figure 6(a) shows the comparisons between different classifiers. They are decision tree (DT), random forest (RF), support vector machine (SVM), DNN, DNN regression + DNN, and DNN regression + MBM + DNN. Clearly, DNN provides better accuracy than DT, RF, and SVM. The DNN regression provides more inputs for the classifier, and it is noticed that DNN regression for a non-physically operational case deviates more from the physical constraints, like mass conservation, than when a physically operational case is explored. Thus DNN regression + DNN classifier can stably provide better accuracy than the traditional DNN classifier. The MBM model provides an additional mass balance check for the classifier further improving the accuracy. Figure 6(b) shows a flow chart of the DNN regression + MBM + DNN classifier, which has been shown to most accurately predict whether or not a fuel cell can be physically operational given a set of input parameters.



(a) (b)

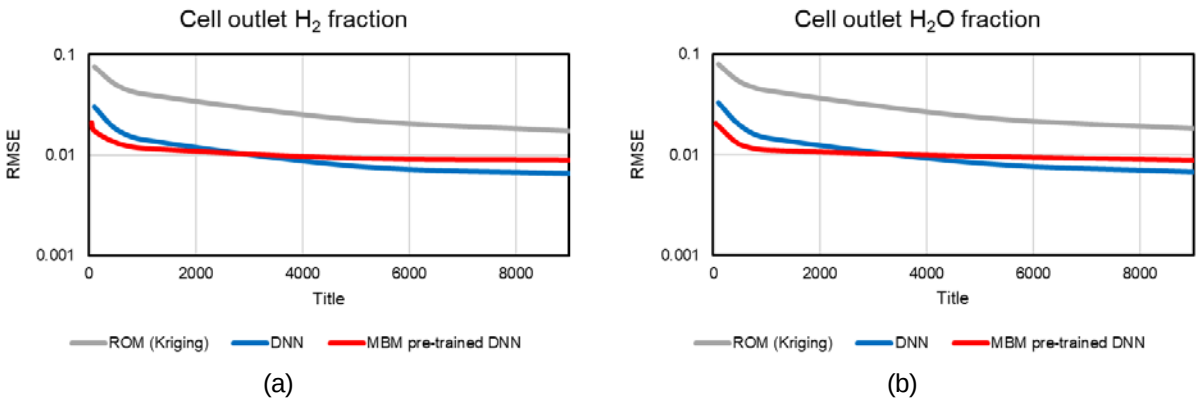
Figure 6. Comparison between different classifiers for determining if the fuel cell can be physically operational given a set of input parameters.

3.4 Transferred DNN

The creation of a ROM surrogate provides detailed performance predictions for the system under a given set of assumptions (e.g., given electrochemical performance, given system flowsheet). It was also investigated if a ROM developed under those conditions could still provide beneficial knowledge for cases with different assumptions for the system flowsheet or performance. In other words, could the existing pre-trained neural networks developed for one condition be transferred to another condition and beneficially help enhance the rate of learning? To answer this question, two kinds of DNN transferring were studied and tested. One was transferring a pre-trained DNN for low-fidelity model data to a high-fidelity model. The second one was transferring a pre-trained DNN for SOA NGFC to advanced NGFC.

3.4.1 Transferring DNN for Different Fidelity Model Data

As mentioned in Sections 3.2 and 3.3, a simple MBM model can provide the approximations of molar fractions for the species in NGFC systems almost instantly. This means that almost infinite training data can be generated with nearly no cost. Therefore, 40,000 MBM prediction data were used to pre-train a DNN regression model. The pre-trained DNN was then fine-tuned by small amounts of the true simulation data from SOFC-MP. Figure 7 shows the model prediction error for different numbers of samples from SOFC-MP simulations. The transferred DNN can provide better prediction than the traditional DNN regression with a small amount of SOFC-MP simulation data. However, the advantages of the transferred DNN decrease with an increased number of training samples. With more than about 3500 true simulation data from SOFC-MP, the transferred DNN provides a higher error rate than the traditional DNN regression approach. This transferred DNN approach is still valuable for when the true simulation data is limited and difficult to acquire.



(a) (b)

Figure 7. Prediction error for MBM pre-trained DNN and comparison to other approaches.

3.4.2 Transferring DNN from State-of-the-Art NGFC to Advanced NGFC

Another implementation of the transferred DNN method is to transfer an established DNN framework from state-of-the-art (SOA) NGFC to advanced NGFC. The advanced NGFC in this work assumes a 50% reduction for both activation and ohmic over potential compared with SOA NGFC. First, the SOA NGFC DNN framework is established using the maximum available number of SOA NGFC SOFC-MP data. The goal is to make the SOA NGFC DNN framework as accurate as possible. Then, the entire DNN framework including all the weights and bias values is transferred and ready to be implemented by using the advanced NGFC SOFC-MP data. In the meantime, another DNN framework (fresh DNN) with the same number of layers and neurons is generated without any pre-existing knowledge of either SOA or advanced NGFC information. In order to make an apples-to-apples comparison, both DNN frameworks are trained using the same number of training datasets from the advanced NGFC SOFC-MP simulations. The cell voltage root-mean-square-error (RMSE) for the same testing dataset is plotted in Figure 8. It can be seen that RMSE decreases with an increasing number of training dataset for both DNN frameworks. However, it is obvious that the transferred model DNN (in orange) exhibits less RMSE compared with the fresh DNN (in blue), especially when the training datasets are small. The results are expected because with an initial knowledge of SOA NGFC, the transferred model DNN may bring some underlying physics and assist the prediction of advanced NGFC performance, thereby yielding less RMSE compared with fresh DNN when the training dataset is not sufficient. However, once more training datasets are fed into the fresh DNN framework, the advantages of transfer model DNN are not that significant and the RMSE between those two DNNs becomes comparable.

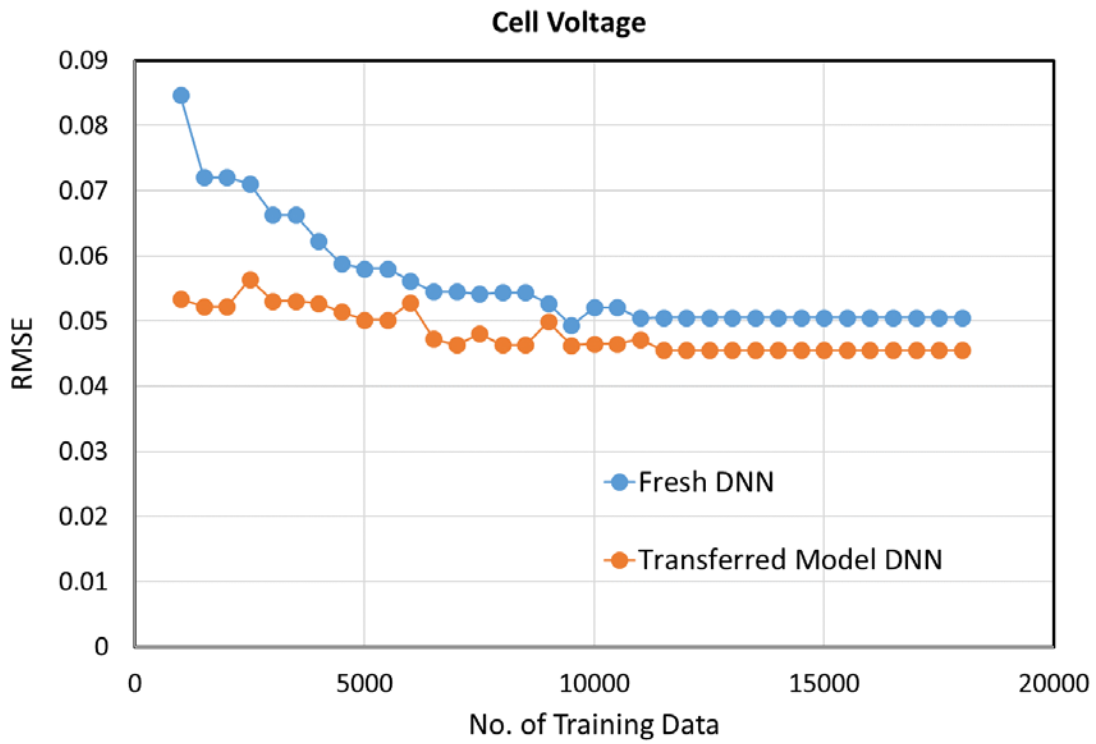


Figure 8. Cell voltage RMSE comparison between fresh and transferred model DNN for advanced NGFC system using different numbers of training data

4.0 Conclusions

In this LDRD project, several applications of ML were explored for SOFC modeling. (1) The errors between traditional Kriging regression prediction and true solution can be learned using ML. The ML-trained pattern of the error distribution can reduce the error of the Kriging based ROM predictions. (2) DNN coupled with the MBM was developed to significantly decrease the ROM prediction error with fewer training data compared to the traditional Kriging-based ROM and traditional DNN regression approach. (3) DNN regression coupled with a neural network (NN) classifier and MBM was developed. It provides superior performance on classifying the physically operational conditions for the natural gas fuel cell (NGFC) than traditional classification approaches. (4) Two kinds of pre-trained DNN transferring were studied and tested. One is transferring DNN for low-fidelity model data to a high-fidelity model. The second one is the transferring DNN from state-of-the-art NGFC to advanced NGFC. The capabilities and tools developed in this project will enhance the future studies of SOFC, and increase the probability of participating into DOE's new ML and/or SOFC related funding opportunities.

5.0 References

Iyengar A, D Newby, D Keairns, BA Hamilton, W Shelton and K Gerdes. 2014. "IGFC and NGFC Pathway Studies – Estimation of Stack Degradation Costs and Salient Results." Presented at *15th Annual SECA Workshop*, July 22, 2014. National Energy Technology Laboratory (NETL). DOE/NETL-2016/1766. Available at https://www.netl.doe.gov/projects/files/IGFCandNGFCPathwayStudiesEstimationofStackDegradCostsandSalientResultsSECAWorkshopPres_072214.pdf.

Koeppel BJ, C Lai, AKS Iyengar, Z Xu, C Wang and GA Hackett. 2017. "Use of a Reduced Order Model (ROM) to Simulate SOFC Performance in System Models." *ECS Transactions* 78(1):2595-2605. DOI: 10.1149/07801.2595ecst.

Koeppel BJ, K Lai and MA Khaleel. 2011. *SOFC-MP 2D User Manual*. PNNL-20433. Richland, WA: Pacific Northwest National Laboratory.

Lai K, BJ Koeppel, KS Choi, KP Recknagle, X Sun, LA Chick, V Korolev and M Khaleel. 2011. "A Quasi-Two-Dimensional Electrochemistry Modeling Tool for Planar Solid Oxide Fuel Cell Stacks." *Journal of Power Sources* 196(6):3204-3222. DOI: 10.1016/j.jpowsour.2010.11.123.

Pan W, J Bao, C Lo, K Lai, K Agarwal, BJ Koeppel and M Khaleel. 2013. "A General Approach to Develop Reduced Order Models for Simulation of Solid Oxide Fuel Cell Stacks." *Journal of Power Sources* 232:139-151. DOI: 10.1016/j.jpowsour.2013.01.057.

PNNL- 29124

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)

www.pnnl.gov