Hybrid physics-informed neural networks for lithium-ion battery modeling and prognosis

Renato G. Nascimento\textsuperscript{a,b,}\textsuperscript{*}, Matteo Corbetta\textsuperscript{a}, Chetan S. Kulkarni\textsuperscript{a}, Felipe A.C. Viana\textsuperscript{b}

\textsuperscript{a} KBR LLC, NASA Ames Research Center, Moffett Field 94035 CA, USA
\textsuperscript{b} Department of Mechanical and Aerospace Engineering, University of Central Florida, Orlando, FL 32816-8030, USA

\textbf{HIGHLIGHTS}

- Physics-informed neural network implements physics into a deep learning framework.
- Model capable of forecasting end-of-discharge while accounting for battery aging.
- Aging prognostic uses information from a fleet of batteries via an ensemble model.

\textbf{ABSTRACT}

Lithium-ion batteries are commonly used to power unmanned aircraft vehicles (UAVs). The ability to model and forecast remaining useful life of these batteries enables UAV reliability assurance. Building principled accurate models is challenging due to the complex electrochemistry that governs battery operation. Alternatively, reduced order models have the advantage of capturing the overall behavior of battery discharge, although they suffer from simplifications and residual discrepancy. This paper presents a hybrid modeling approach that directly implements physics within deep neural networks. While most of the input-output relationship is captured by reduced-order models, data-driven kernels reduce the gap between predictions and observations. A reduced-order model based on Nernst and Butler–Volmer equations represents the overall battery discharge, and a multilayer perceptron models the battery non-ideal voltage. Battery aging is characterized by time-dependent internal resistance and the amount of available Li-ions, which are modeled through an ensemble of variational Bayesian multilayer perceptrons. The approach is validated using data publicly available through the NASA Prognostics Center of Excellence website. Results showed that our hybrid battery prognosis model can be successfully calibrated, even with a limited number of observations. Moreover, the model can help optimizing battery operation by offering long-term forecast of battery capacity.

1. Introduction

Electric and hybrid systems are at the foundation of the current transportation transformation, in which small and large vehicles are expected to use electric or hybrid power-trains as a source of propulsion. This includes both ground vehicles, like hybrid and electric cars driven by thousands of people every day, and future air vehicles [1–3]. Electric power-trains start with the energy storage unit, which most of the time is composed of Lithium-ion (Li-ion) or Lithium-polymer (Li-Po) battery packs. Although new concepts for energy storage are under consideration, like Lithium-air batteries [4] and nanotechnology applied to energy storage [5,6], Li-ion and Li-Po batteries remain the most common and consolidated way to provide energy to electric power-trains.

To achieve aircraft propulsion electrification, experts suggest a power-to-weight ratio of about 300 kW/kg by 2021 for 3–4 passenger Urban Air Mobility (UAM) type vehicles [7], which is expected to be achieved over the next years of technological development according to some qualitative estimates [7,8]. Power over weight is not, however, the only requirement for a successful application of electric energy to aviation. Reliable tools to monitor the battery in real-time and perform diagnosis and prognosis are also equally important, for the obvious reason that aircraft loss of propulsion has more serious consequences than that of a ground vehicle.
A number of recent research efforts on battery management systems included works on diagnosis and prognosis to monitor the state of charge (SOC) of the battery and predict the remaining time to discharge or the aging associated with multiple charge-discharge cycles. Physics-based, empirical and data-driven models have been tested in multiple scenarios, with their advantages and drawbacks. The need of knowledge of the physical phenomena at some degree of granularity and long–short term memory layers, but has no physics in the model.

One may identify similarities between our proposed hybrid physics-informed neural network approach and other methods that combine physics-based models and statistical learning. For example, Kalman filtering and its variants often use reduced-order models that are constantly updated as observed data becomes available such that changes in model parameters can be tracked. Within the battery literature, the interested reader is referred to [26–28]. Our approach differs in which we use the neural network machinery to allow for (a) explicit hybridization of the model, such that model-form uncertainty can be captured by data-driven layers; (b) nonlinearity in the model which can be directly handled by backpropagation; and (c) tracking of model parameters through variational inference. While the resulting model is fast to compute and can be used for predictions in real time, the training and update can be computationally intensive (few hours in today’s computers). This shortcoming might not be there in some implementations of the Kalman filtering and we can see this as being the compromise due to the added capabilities of our hybrid approach.

Hybrid methodologies combining both worlds are now emerging in many fields, given recent advancements in software architectures to easily train data-driven algorithms and deep learning models, and the challenge in collecting useful data for critical systems (even with the advancement of the big data era, useful engineering data in many application areas are still small). The work presented in [29] goes in that direction, by extracting features related to the physics of charge of the battery and combining them with Gaussian process regression. Another example of hybrid approach is proposed in [30]. However, those examples show the implementation of a hybrid framework where
data-driven and physics-based models are used in synergy, but they remain distinct in their definitions.

As a matter of fact, blending physics-based and data-driven techniques in a true sole model has its own challenges related to the identification of the merging point. Open research questions are "where does the flexibility of learning models outweighs the accuracy of physics?" "can data-driven models compensate for the lack of knowledge of physical parameters?" or "can physics guide and/or improve optimization of data-driven models (avoiding overfitting)?" Identifying parameters or a portion of a physics-based model that can be substituted by a data-driven algorithm is a challenging task. The latter can also cause over-fitting and reduce the extrapolation capability of the hybrid model. For that reason, careful initialization of the model parameters and constraints resembling physics in the estimation algorithm are paramount to ensure good fitting, but not obvious to define at times (as a matter of fact, the careful initialization of the parameters is also important in deep learning models for reasons related to gradient-based optimization and saturation of activation functions). However, in real-time applications, learning models can compensate the natural inter-specimen variability typically observed in fleets of the same (or similar) systems, thus aiding uncertainty quantification, and advances in deep learning training systems in the last decade made the adoption of large data-driven models very appealing, even though the risk of over-fitting in a limited-data scenario is high. Relying too heavily on deep learning models requires a vast amount of data that are rarely available in the area of complex engineering systems and in the field of fault detection and failure prediction. Moreover, deep learning models obfuscate the understanding of the model itself, reducing its interpretability. This might also hinder the model adoption in the field, as users may be reluctant to adopt a model they cannot interpret, as suggested by the latest research initiatives towards explainable and interpretable artificial intelligence in engineering [31,32] and other fields [33,34].

The remaining of the paper is organized as follows. Section 2 gives an overview on the reduced-order battery discharge model used to build the hybrid model. Section 3 details our proposed physics-informed neural network model. Section 4 describes in detail our main results including model fitting and validation, random-loading discharge predictions, and battery aging modeling and forecast. Finally, Section 5 concludes the paper by summarizing significant remarks, and providing insight on potential future studies.

2. Reduced-order physics-based model

The model adopted in this paper was first proposed in [18] and then improved in [35]. The purpose of the model was to work in real-time or quasi-real-time applications, and was designed with ordinary rather than partial differential equations, at the cost of some precision and accuracy loss. The use of ordinary differential equations facilitates the embedding of the model in fast prediction frameworks. Because of such simplifications, the model uses physics-based equations as well as some empirical derivations. As equations and descriptions are based on [18], the interested reader is referred to that article for more details and a thorough description of the model.

2.1. Equilibrium potential

The model consists of the Nernst’s equation for the equilibrium potential of a Li-ion battery:

\[
V_{eq,i} = U_0 + \frac{RT}{mF} \ln \frac{1-x_i}{x_i} + V_{nl,i}.
\]

where the electrode (negative or positive) is indicated by the subscript \(i = \{n, p\}; U_0\) is the reference potential; \(R\) is the universal gas constant; \(T\) is the electrode temperature; \(m\) is the number of electrons transferred in the reaction; \(F\) is the Faraday constant; \(x\) is the mole fraction for the Lithium-intercalated host material; and \(V_{nl,i}\) is the non-ideal voltage and activity correction term, null in ideal conditions. Details about \(V_{nl}\) will be provided later in the section. The mole fraction is computed as the ratio between the amount of Li-ion \(q\) in electrode \(i\), and the amount of available (moving) Li-ions \(q_{max}\):

\[
x_i = q_i/q_{max} \quad \text{and} \quad q_{max} = q_n + q_p. \tag{2}
\]

In order to accommodate the concentration gradient at the surface of the electrode, the total volume of the battery is split into two control volumes, bulk and surface, and the concentrations of Li-ions are calculated accordingly:

\[
c_{i,n} = \frac{q_{i,n}}{v_{i,n}} \quad \text{and} \quad q_{max} = q_{i,n} + q_{i,p} + q_{i,s} + q_{i,a}. \tag{3}
\]

The diffusion rate from the bulk to the surface is:

\[
q_{i,s,n} = D \left( c_{i,n} - c_{i,s,n} \right). \tag{4}
\]

where \(D\) is the diffusion constant. The rates of change of the charges \(q\) are calculated using the diffusion rate and the applied electric current \(i_{app}\):

\[
\dot{q}_{i,n} = i_{app} + q_{i,p} \quad \text{and} \quad \dot{q}_{i,s} = -i_{app} - q_{i,n}. \tag{5}
\]

The concentration overpotential is calculated using the Nernst’s equation for the surface, which is by substituting \(x_i\) with \(x_{i,s}\):

\[
x_{i,s} = \frac{q_{i,s,n}}{q_{max}} \quad \text{and} \quad q_{max} = q_{i,s,n} v_{i,s}/v_{i}. \tag{6}
\]

The solid-phase Ohmic resistance, electrolyte Ohmic resistance, and current collector resistance can be lumped together into \(R_0\) to calculate the voltage drop: \(V_0 = i_{app} R_0\), where \(i_{app}\) is the required current [18].

The surface overpotential is described by the Butler-Volmer equation [18], and the voltage \(V_{eq}\) can be expressed as:

\[
V_{eq,i} = \frac{RT}{nF} \arcsinh \left( \frac{J_i}{2F \rho} \right) \tag{7}
\]

This way, the battery output voltage is defined by:

\[
V = V_{eq,n} - V_{eq,p} - V_0 - V_{nl,n} - V_{nl,p}, \tag{8}
\]

which will serve as output of the physics-informed model.

2.2. Non-ideal voltage, internal resistance, and amount of available Li-ions

As part of the simplifications used to obtain the described reduced-order physics-based model, the non-ideal voltage, battery resistance, and amount of available Li-ions are elements of the model fitted using experimental data. As already mentioned, the non-ideal voltage \(V_{nl}\) is the activity correction term needed to compute \(V_{eq}\). The model in [18,19] fits the experimental data using the Redlich–Kister expansion

\[
V_{nl,i}(x_i; A_{ij}) = \frac{1}{mF} \sum_{k=0}^{N_i} A_{ij} \left[ (2x_i - 1)^{k+1} - 2x_i k (1 - x_i) \right] / (2x_i - 1)^k \tag{9}
\]

The mole fraction \(x_i\) is the independent variable, the coefficients \(A_{ij}\) are identified through data-fitting, and the number of elements in the sum \(N_i\) is empirically-derived as well. For the batteries tested in this work, the reference papers [18,19] kept \(N_i = 12\) and \(N_p = 0\), thus using a constant non-ideal voltage for the negative side of the electrode.

Lumped internal resistance \(R_0\) and maximum number of available Li-ions \(q_{max}\) drastically affect the prediction performance of the model. Such influence appears at both the inter- and the intra-specimen level; those two parameters change as the battery ages. The loss in active material can be represented by a drop in \(q_{max}\), while an increase in \(R_0\) is representative of a constant Ohmic drop, which causes an
increase in total resistance independent of the battery charge state [35]. Such effects should be accounted for when training on multiple discharge curves of the same battery type. Moreover, different samples of the same fleet of batteries also have slightly different $R_n$ and $q_{max}$, thus varying from one battery to another and contributing to inter-specimen variability. That adds to the variability that can be found in small changes or imperfections of the manufacturing process and raw materials.

3. Physics-informed neural network model

This section details our approach to create a hybrid model that embeds the physics-based reduced order models in a neural network. By design, our proposed hybrid model leverages the physics expressed by the reduced-order models within a recurrent neural network. The hybrid model predictions are expected to capture major trends and could agree with experimental data at least in parts. In order to close any remaining systematic gap, we propose hybridizing the reduced-order model with data-driven layers, such as a multi-layer perceptron. The role of the data-driven layers is to characterize the model-form uncertainty in the reduced-order physics-based models. For a survey about physics-informed neural networks, the reader is referred to [36]. Here, emphasis goes onto the estimation of the non-ideal voltage $V_{ni,n}$, which, in the original model, was estimated with Eq. (9), as well as the total resistance $R_n$ (ohms) and the maximum charge $q_{max}$ (coulombs), which values are driven by the aging of the battery.

3.1. Battery model as a recurrent neural network

A RNN extends the traditional feed forward networks to model time-dependent responses [37–39], as shown in Fig. 1(a). RNNs repeatedly apply transformations to given states in a sequence such that

$$\begin{align*}
 y_t, h_t^T &= f(u_{t-1}, h_{t-1}), \\
(10)
\end{align*}$$

where the subscript $t$ represents the time discretization, $y \in \mathbb{R}^n$ are the observable states, $h \in \mathbb{R}^n$ are the internal states, $u \in \mathbb{R}^m$ are input variables, and $f(\cdot)$ defines the transition between time steps (function of input variables and previous states). In the recurrent neuronal network terminology, different implementations of $f(\cdot)$ are referred to as cells.

Since the surrogate physics model describes the discrete state-space representation of the governing set of ordinary differential equations, it allows the definition of a state update per time step. The approach follows the work presented in [20]; since RNNs naturally operate on discretized temporal sequences, they are suited to perform numerical integration. We proceed by designing a RNN that embeds the battery model, as illustrated in Fig. 1(b). This design corresponds to the implementation of the physics-based model previously described as a directed graph model [40, Ch. 2] coded in the form of a RNN. The RNN cell task is to take the state vector at the previous time step, $h_{t-1}$ and the input at the current time step, $u_t$, and update the state vector $h_t$ and the output $y_t$. The RNN cell produces a one-step-ahead prediction of the entire state-space model

$$
\begin{align*}
 h_t &= g_1(h_{t-1}) \\
y_t &= g_2(h_t)
\end{align*}
(11)$$

where the state, input, and output vectors are

$$
\begin{align*}
 h &= [V, V_n, V_{ni,n}, r_{ap}, q_{ap}, q_{hp}, q_{ip}]^T, \\
u &= u_{app}, \quad y &= V.
\end{align*}
(12)$$

In Fig. 1(b), the cell receives the state vector at the previous time step $(h_{t-1})$ and the input value at the current time step $(u_t)$, and returns the state vector and the output voltage $(h_t$ and $V_t$). The blue boxes are pure physics-blocks, which perform the same calculations of the physics-based model. The changes in the internal voltage states ($\Delta V, \Delta V_n, \Delta V_{ni,n}$) are obtained using an empirical time constant $\tau$ [18]. The dashed-rounded white boxes show variable selectors. The green boxes are representative of the two data-driven models used to estimate the non-ideal voltages, while the yellow circles represents the adjustable physical model parameters, $q_{max}$ and $R_n$. The summation at the top of the cell is representative of Eq. (8).

3.2. Hybrid physics-informed data-driven model

The physics-based model previously discussed has two parameters and two data-driven components that need to be empirically adjusted based on observed data. These are the internal resistance $R_n$, the amount of available Li-ions $q_{max}$, and the non-ideal voltages $V_{ni,n}$, respectively. For example, in [35], authors used test results from specific reference discharge cycles to estimate $R_n$ and $q_{max}$ as model parameters, as well as the coefficients of the Redlich–Kister expansion for $V_{ni,n} \sim \text{i.e., } A_{ki}$ in Eq. (9) – and coefficients of a linear regression model for $V_{ni,n}$.

Therefore, $R_n$ and $q_{max}$ are kept as hybrid model parameters. Multi-layer perceptrons (MLPs) take the role of the Redlich–Kister expansion and linear regression models to capture the dynamic of the non-ideal voltage

$$
V_{ni,n} = MLP_1(x_i, w_i, b_i) \quad \text{and} \quad V_{ni,p} = MLP_2(x_p, w_p, b_p) \quad (13)
$$

where $w_i, b_i, w_p, \text{ and } b_p$ are the MLP parameters. Vectors $w$ represent the collection of weights of the MLP layers, while vectors $b$ represent the collection of biases. Subscripts $i$ and $p$ refer to positive and negative electrode, respectively. Following typical MLP architectures, each node of the networks has its own weight and bias. The MLP is not restricted by the equation form of the Redlich–Kister expansion, and so it allows more flexibility in learning battery dynamics that are not captured by the physics-model. This advantage comes with the need for careful initialization of the MLP parameters, analysis of the model training during development to ensure the MLP is learning what is supposed to (the variability of the non-ideal voltage due to the activity coefficients), and not compensating for uncertainty or variability of other model parameters.

With the introduction of the MLPs in Fig. 1(b), the model becomes a hybrid RNN with nodes that are physics-informed (e.g., Nernst and Butler–Volmer equations) and nodes that are data-driven (MLPs for non-ideal voltage).

Table 1 details the design of each MLP used in the hybrid model. Similarly to the pure physics model, a linear model is used for the non-ideal voltage of the negative side, which is a MLP with one layer, one hidden unit and no activation function. The non-ideal voltage curve for the positive side is more complex, and therefore a MLP with 3 layers (8 hidden units in the first layer and 4 in the second hidden layers) with hyperbolic tangents as activation functions is used. All hidden units and output units have biases. Both MLPs take the mole fractions $x_i, x_p$, as inputs, and return the corresponding non-ideal voltages. The larger MLP for the estimate of the positive-side non-ideal voltage has 57 trainable parameters, while the MLP for the negative-side non-ideal voltage has only two trainable parameters (slope and bias of the linear model).

It is important to highlight that the hybrid physics-informed neural network models the time-dependent discharge of the battery. As such, the input and output variables are the time histories of current and
battery voltage. These are the only two observable quantities. All states used to describe the physics, including $h = \{T, V_{n_0}, V_{n_0,p}, q_{n_0}, q_{n_0,p}, q_{n_0,p}^r\}^T$, $V_{n_0,p}$ and $V_{n_0,p}^r$ are hidden variables. This makes the training of our hybrid physics-informed neural network very interesting, as $V_{n_0,p}$ and $V_{n_0,p}^r$ are never observed directly and $q_{n_0}$ and $R_{n_0}$ are very deep in the graph (see Fig. 1(b)).

4. Results and discussions

Our implementation is all done in TensorFlow\(^1\) using the Python application programming interface (version 2.3). The uncertainty quantification and probabilistic modeling is performed using the TensorFlow Probability library\(^2\) (version 0.11.0). The next subsections will detail how the model is fitted and validated as well as its use for random discharge prediction and modeling of battery aging. The fitted hybrid physics-informed neural network model was validated against the physics-based model, which is publicly available in the NASA’s Prognostics Model Library\[^{[41]}\].

4.1. Model fitting and validation

The first step consisted into training our hybrid physics-informed neural network, which simultaneously optimizes the values of $q_{n_0}$, $R_{n_0}$, and the MLP parameters ($w_p$, $b_p$, $w_r$, and $b_r$). We chose to do so using the constant-loading data\[^{[24]}\]. The first 3 constant-loading discharge curves are extracted from each of the 12 batteries in the data set\[^{[24]}\]. Each curve was generated with a current draw of 1 A from the fully charged condition of 4.2 V down to a value of 3.2 V, when the tests stopped. Then, the mean squared error is used as loss function $A$:

$$A = \frac{1}{N} (\mathbf{v} - \bar{\mathbf{v}})^T (\mathbf{v} - \bar{\mathbf{v}}),$$

where $N$ is the number of observations, $\mathbf{v}$ are the observed battery voltages over time, and $\bar{\mathbf{v}}$ are the predicted battery voltages by the physics-informed neural network. The model is trained with the Adam optimizer\[^{[42]}\] set with a learning rate of $5 \times 10^{-3}$ and 3000 epochs. All other Adam parameters were kept equal to their default values.

The physics-based portion of the model is initialized with the same physical and empirical parameters suggested for the physics-based surrogate model by original paper\[^{[18]}\]. Battery parameters $R_{n_0}$ and $q_{n_0}$ were initialized to $1.5 \times 10^{-2}$ Ω and $1.4 \times 10^{-3}$ C, respectively, which are reasonable values for the specific type of battery. We introduced further prior knowledge of the system by initializing the MLP for the positive electrode as a piece-wise linear curve representative of the decreasing non-ideal voltage over time. Such an initialization constitute a fundamental step in the design of the hybrid framework. Thus, the local optima the MLP converges to generates a curve resembling the one obtained by fitting experimental data with more traditional procedures, like the Redlich–Kister expansion in (9). In our experience, a random initialization of the MLP for $V_{n_0,p}$ can pose problems such as poor convergence of model parameter optimization, overfitting of data, and lacking of generalization. One might argue this is not a problem only for neural networks (or physics-informed neural networks for that matter), but for nonlinear regression in general. It is important to keep the range of $V_{n_0,p}$ between $-1$ and $1$ (as MLPs are normalized). However, the exact value of $V_{n_0,p}$ for $x_p = 0$ and $x_p = 1$ are not as important (as long as they are close to 1 and $-1$, respectively). Likewise, the exact location of the knee of the bilinear curve is not as important. Even these simple incorporation of physically guided initialization was enough to improve optimization of MLP parameters. Once the MLP is initialized, the training process using a few discharge cycle curves helps fine tuning both MLPs to capture the underlying dynamics of the decreasing non-ideal voltage. The MLP for the negative electrode is the simplest linear model (slope and bias), and so the default Keras initialization parameters is used.

A total of 36 discharge curves were used to simultaneously optimize the values of both MLP parameters as well as $q_{n_0}$ and $R_{n_0}$. This is an important feature of our physics-informed neural network. The estimation of physical model parameters and the optimization of data-driven model parameters is performed simultaneously using the time histories directly. This poses an advantage in many engineering applications, as the need for large data sets is reduced and compensated by the use of physics-based models.

Given that $q_{n_0}$ and $R_{n_0}$ are estimated based on experimental data, slight variation in estimated values due to sample-to-sample variation is expected, even if data comes from the same battery type. Under such an assumption, the pair $(q_{n_0}, R_{n_0})$ can be used as surrogate to learn and represent the inter-specimen variability observed in specimens of the same battery type. Thus, a value pair of $(q_{n_0}, R_{n_0})$ is representative of a particular battery and the values in the battery RNN cell are vectorized, creating a pair $(q_{n_0}, R_{n_0})$ for each discharge curve. By so doing, after training, a distribution of the pair $(q_{n_0}, R_{n_0})$ is obtained while the MLPs

---

\(^{1}\) www.tensorflow.org

\(^{2}\) www.tensorflow.org/probability
Here only the output curves of the MLP for the positive side of the electrode. We and new parameters of the MLPs with different folds. Each fold produces a value of $q_{max}$ dispersion of the MLPs' output is quantified when the model is trained obtained a MLP curve, one distribution for $q_{max}$ as well as its prediction capabilities outside the training set. Further the robustness of the model with regards to the training data is assessed this process is repeated and the subsets of (30-6) curves are shuffled, k-fold cross validation the 6 curves left out are not used in the training validation curves at each fold. It is important to point out that in this subsets (30-6) were shuffled to allow new combinations of training and were split into 30 curves for training, and 6 for cross validation. The 36 samples at constant loading originally used for the first model fitting ability of the model over constant-loading discharge curves. The 36 cross validation approach to assess the robustness and generalization confidence bands obtained using the Wilson interval. We chose a k-fold output (from the sets left out for validation) is within the expected of the voltage output for different training sets and assure the voltage prediction the end of discharge can be assessed by looking at the time confidence intervals approximately equal to discharge cycle. The number of voltage data points of the remaining predictions appear to be well-aligned with the true values, with a root mean square error of 27.5 s, in a range of 6700 to 7600 s. We hypothesize that the non-ideal voltage models (MLP_p and MLP_n) can be shared across different batteries and do no change over time (i.e., no aging effect). In order to test this hypothesis, we carry out a cross validation study. In this process, we compare the distribution of the voltage output for different training sets and assure the voltage output (from the sets left out for validation) is within the expected confidence bands obtained using the Wilson interval. We chose a k-fold cross validation approach to assess the robustness and generalization ability of the model over constant-loading discharge curves. The 36 samples at constant loading originally used for the first model fitting were split into 30 curves for training, and 6 for cross validation. The subsets (30-6) were shuffled to allow new combinations of training and validation curves at each fold. It is important to point out that in this k-fold cross validation the 6 curves left out are not used in the training of the model, which will only use the respective set of 30 curves. As this process is repeated and the subsets of (30-6) curves are shuffled, the robustness of the model with regards to the training data is assessed as well as its prediction capabilities outside the training set. Further details about cross validation can be found in [43,44]. For reference, the model is trained using all 36 sample curves and obtained a MLP curve, one distribution for $q_{max}$ and one for $R_0$. The dispersion of the MLP's output is quantified when the model is trained with different folds. Each fold produces a value of $q_{max}$ and one of $R_0$, and new parameters of the MLPs. By comparing the MLP predictions for each fold, the robustness of the model against different combinations of curves in the training set is assessed. If that were not the case, then our approach of using one MLP model for the population build upon similar batteries would be invalid. The MLP outputs appear very similar to each other and close to the reference MLP, suggesting that using one MLP model for the population (multiple batteries with similar characteristics) can be considered a good approximation. The average root mean square error between the 6-fold-learned MLPs and the reference output curve is $1.22 \times 10^{-3}$, and the average standard deviation of the six curves is $8.94 \times 10^{-3}$ — i.e., 0.45% of the non-ideal voltage normalized range ($\approx 2V$). Table 2 shows the mean square error between the predicted output voltage curve, and the observed output voltage curve, averaged over all the 6 validation curves. We proceeded to further validate the output voltage predictions at constant loading by examining the probabilistic coverage of the predicted curves. The 30 sets of trainable elements (MLP parameters, $q_{max}$, $R_0$) obtained with the k-fold cross validation are then used to predict the confidence interval of the output voltage over time, thus producing a range of plausible voltage values at each step of the discharge cycle. The number of voltage data points of the remaining 6 validation samples that fall outside those confidence intervals is counted. Assuming a confidence intervals $1 - a$, it is expected that the percentage of points outside the interval approximately equal to Table 2

<table>
<thead>
<tr>
<th>k-fold</th>
<th>Root mean square error \times 10^{-3} [V]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.5</td>
</tr>
<tr>
<td>2</td>
<td>7.7</td>
</tr>
<tr>
<td>3</td>
<td>6.9</td>
</tr>
<tr>
<td>4</td>
<td>7.3</td>
</tr>
<tr>
<td>5</td>
<td>7.9</td>
</tr>
<tr>
<td>6</td>
<td>7.9</td>
</tr>
<tr>
<td>All data in training</td>
<td>6.8</td>
</tr>
</tbody>
</table>

We did not encounter any challenge when training the linear model used for the negative side, MLP_n, and the variability observed in the MLP output does not influence the Battery model output nearly as much as the MLP used for the positive side of the electrode.
for the same current draw, and increases in resistance $q$ adhering to the trends observed in the rest of the population. The root mean square error between the random-loading discharge curves, after the batteries were tested in constant-loading conditions. The predictive capabilities of our hybrid model are confirmed by this cross-validation analysis.

### 4.2. Predictions of random-loading discharge

The model over random-loading discharge curves is validated using 8 batteries from the same data set. The parameter pair $(q^{\text{max}}, R_0)$ comes from the reference discharge curve for the same battery (at constant-loading), and it is used to estimate the voltage drops under random-loading. Therefore, the first random-loading test curves carried out right after the constant-loading tests are selected. By so doing, the effect of aging over the parameter pair $(q, R)$ is minimized.

Predictions for all 8 batteries under random-loading conditions are evaluated individually. Fig. 4(a) shows the prediction of the first random-loading discharge curves, after the batteries were tested in constant-loading conditions. The root mean square error between the observed voltage values (thick black line) and the predicted curve (dashed-black line) is always below $5.5 \times 10^{-2}$ V and below 0.088% of the true voltage, except for battery #1, where larger discrepancies between model and real data were observed. We did not identify the reason for such a larger error in that particular battery, but we noticed that battery #1 behaves slightly differently from the others and not adhering to the trends observed in the rest of the population.

Given the physics model structure, battery aging can be captured in part by $q^{\text{max}}$ and $R_0$. Reductions in $q^{\text{max}}$ accelerate the rate of discharge for the same current draw, and increases in resistance $R_0$ lead to drop in voltage [18]. As shown in Fig. 4(b), shortly after a reference discharge cycle (left panel), the values of $q^{\text{max}}$ and $R_0$ are updated and predictions given by the model are in good agreement with observed data (middle panel). As the battery accumulates hours of operation, i.e., it ages, the model prediction starts disagreeing with the observed data (right panel).

The aging effect does not limit the capability of the model trained with the proposed approach. In practice, we propose that one tracks the model prediction versus observed data. Then, the latest discharge curves can then be used to update the pair $(q^{\text{max}}, R_0)$ by fitting the model, optimizing only the battery parameters pair while keeping the non-ideal voltage $V_n$ weights fixed. By doing so, the updated model predictions will account for the aging effect introduced by the latest loading curve. This continuous learning procedure is suitable for one-step-ahead prediction of aging. However, it does not enable far-ahead forecast of the battery aging, represented by a reduction in maximum capacity. We address how to forecast aging by using a fleet-based approach and ensemble modeling in the next section.

### 4.3. Aging effect

#### 4.3.1. Aging parameters and cumulative energy

Battery aging is manifested in the variations of the parameter pair $(q^{\text{max}}, R_0)$ as loading cycles accumulate in the battery life. Therefore, we propose using the pair as a proxy indicator for aging. In terms of modeling, this implies that $(q^{\text{max}}, R_0)$ are updated throughout the battery life, while the parameters for the two MLPs (which characterize interval voltage) are kept the same.

The pair $(q^{\text{max}}, R_0)$ is updated using the sequential, constant-loading discharge curves available in the data set for each battery. These reference-discharge curves are followed by a number of random-loading discharge, after which a new reference-discharge is performed. This information allowed us to characterize the changes in $q^{\text{max}}$ and $R_0$ as the batteries accumulated loading cycles. Fig. 5(a) shows that, as expected, $q^{\text{max}}$ decreased over time (although not shown, as expected, $R_0$ increased over time) The rate of decay is visually similar for all the batteries since they have same characteristics and experienced similar loading conditions.

Although interesting, curves of $q^{\text{max}}$ and $R_0$ as a function of time might not be as useful in real life, as they do not capture the dependency with regards to required power per cycle. Instead, we suggest using cumulative energy as independent variable. With that, operators will be able to perform analysis and trade-off studies regarding mission dispatch and remaining useful life. Fig. 5(b) shows battery capacity and $q^{\text{max}}$ as a function of the cumulative energy. Circles refer to the capacity values, while full dots refer to $q^{\text{max}}$ values, and their magnitude is shown in the double-y axis of the panel. The figure shows data for all batteries, emphasizing the similarities among them. The clear

---

**Fig. 3.** Assessment of the model generalization by evaluating how many points of a experimental discharge curve fall outside the confidence intervals computed by the interrogation of the hybrid model with 30 different sets of model parameters. The two panels show results obtained with the data set subdivision of 30 reference curves and 6 validation curves.

\[ a \cdot 100\%. \] The percentage points that should fall within the confidence interval is obtained using the Wilson interval \([45,46]\):

\[
\text{CI} = \hat{p} \pm z_{a/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{N}},
\]

where $\hat{p}$ is the approximated target probability; $z_{a/2}$ is the $a/2$ critical value for the standard normal distribution, which denotes the $(1 - a) \times 100\%$ percentile; and $N$ is the number of observations.

In this case, since 6 curves out of 36 for cross validation are used, $N = 36$ and $\hat{p} = 100 \times 6/36 \approx 16.7\%$. Using the Wilson interval, $\hat{p}$ is expected to be in $16.7\% \pm 14.6\%$ interval (considering one standard deviation). Fig. 3 shows two examples of this validation process for the best case (points outside the confidence interval are $3.1\%$ of the entire curve) and the worst case ($32.1\%$ outside the CI) among the 6-fold validation. On average, 16.99\% of the output voltage data points fall outside the CI, with a large standard deviation of $10.74\%$. Therefore, the Wilson interval still defines the result, 16.99\% $\pm$ 10.74\%, as acceptable. The apparently large standard deviation is just an artifact of the small sample size (6-curves) and it does not relate to our model predictions. The predictive capabilities of our hybrid model are confirmed by this cross-validation analysis.
correlation can, in principle, allow us to model aging by modeling the decay of its two proxies ($q_{\text{max}}$ and $R_0$) as a function of the past energy drawn. Fig. 5(b) also suggests that up to 1 kWh, all batteries show a very narrow dispersion of the capacity and $q_{\text{max}}$ drop, manifesting low inter-specimen variability. The curves then start diverging and the inter-specimen variability increases. Nevertheless, battery capacity and $q_{\text{max}}$ trends remain highly correlated.

As previously shown, once $q_{\text{max}}$ and $R_0$ are tracked, predictions for random discharge cycles are expected to improve. Fig. 5(c) illustrates the adjustment in the same random discharge cycle shown in Fig. 4(b) (right panel). The bias in the model prediction due to the misestimation of $q_{\text{max}}$ and $R_0$ is removed once these parameters are updated.

Assuming that $q_{\text{max}}$ and $R_0$ can be tracked for an existing set of batteries, we propose using an ensemble model that incorporates information from the set batteries in the prediction of the battery that is currently being monitored. The ensemble model combines the estimates from recorded $q_{\text{max}}$ and $R_0$ curves with the data points available for the current operating battery, and produce the distribution a-posteriori of the future battery capacity. The underlying hypothesis is that the model has access to data from batteries (of the same type) that were already being used and therefore they are in a later aging state. Early in the history of the battery usage, this approach has the advantage of leveraging the observed behavior past 1 kWh from other batteries, and therefore capturing nonlinearities. Details of the developed ensemble is discussed next.

In practice, incorporating the aging effect in the battery physics-informed neural network cell presented in Fig. 1(b) implies in incorporating the dependency of $q_{\text{max}}$ and $R_0$ with respect to the cumulative energy drawn from the battery. Moreover, it is highly desirable to include uncertainty estimates in the forecasted values of the two parameters as a function of cumulative energy.

In order to capture battery-to-battery variation due to aging, we use variational multi-layer perceptrons to build the models for $q_{\text{max}}$ and $R_0$ as a function of cumulative energy. The variational models capture trends and dispersion of $q_{\text{max}}$ and $R_0$ as a function of cumulative energy similarly to nonlinear regression models, in which parametric models describe the expected value, and discrepancy is modeled through Gaussian error with zero mean and unknown variance. The details for these models are given in Table 3. The Tensorflow Probability library is used for the ensemble model, allowing consistency with the model used to predict the voltage drop, and providing a unique solution capable of handling voltage drop predictions and aging at the same time. Fig. 6 illustrates a comparison between the predictions with variational multi-layer perceptrons, each neural network parameter follows a Gaussian distribution, as opposed to being a deterministic value. In the optimization of parameters, the variational model uses a loss function that is the sum of negative log-likelihood and the Kullback–Leibler divergence, which estimates the difference between a reference distribution and the posterior distribution.
Fig. 5. Variation of $q_{\text{max}}$ as batteries accumulate cycles and its effects in the model voltage prediction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3

<table>
<thead>
<tr>
<th>Layer</th>
<th># neurons</th>
<th>Activation</th>
<th># neurons</th>
<th>Activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>4</td>
<td>elu</td>
<td>1</td>
<td>linear</td>
</tr>
<tr>
<td>#2</td>
<td>2</td>
<td>elu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#3</td>
<td>1</td>
<td>linear</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

from the variational models and the $q_{\text{max}}$ used for training (similar performance was observed for $R_0$).

4.3.2. Aging and voltage drop forecast with an ensemble model

Once $q_{\text{max}}$ and $R_0$ starts being collected for a specific battery, one can build the proposed variational MLPs and use them to forecast $q_{\text{max}}$ and $R_0$ even past the current cumulative energy for the battery. Nevertheless, the process is prone to large error if only few $q_{\text{max}}$ and $R_0$ data points are available.

Alternatively, predictions for a new battery can be improved if sub-population of batteries have already been used to collect a set of $q_{\text{max}}$ and $R_0$ in the available set. In addition, in order to benefit even further from the available set of models, we propose creating an ensemble model as a weighted average. For example, for $q_{\text{max}}$:

$$\hat{q}_{\text{max}}(E) = \sum \omega_i q_{\text{max}}^i(e)$$

where $\Sigma_j = e_j e_j^T$ and $e_j$ is a vector with a measure of error. In this paper, the prediction error between the $i$th model in the ensemble and
Fig. 7. Example of $q_{\text{max}}$ forecast using the ensemble of variational MLPs. Orange and green represent priors and posteriors, respectively. Black and gray dots are available and unknown (for the sake of the example) $q_{\text{max}}$ values for the new battery, respectively. Shaded areas and dashed lines denote the 95% prediction interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(a) $q_{\text{max}}$ and individual prior and posterior variational MLP predictions. Information about batteries #1, #5, and #8 is shown in orange. Dashed lines denote the 95% prediction interval.

(b) Ensemble weights. Battery #4 is the one left out for validation.

(c) Ensemble forecast for battery #4. From left to right, 6, 16, and 26 values of $q_{\text{max}}$ were used in the Bayesian update and ensemble weight computation.

the observed data points is used for the battery that the ensemble is been built for.

The ensemble weights can be computed based on the predictions coming from the sub-population priors. Nevertheless, there are benefits in using ensemble weights out of the posterior models (particularly as the collected set of $q_{\text{max}}$ and $R_0$ for the new batteries starts capturing its specific behavior). We argue that, over time, enough $q_{\text{max}}$ and $R_0$ pairs will be collected and the effect of the sub-population information is reduced.

In order to validate the approach, an ensemble model with 7 out of the 8 batteries subject to random loading is used to predict the aging of the battery left out (battery #4 in this case). It is assumed that 6 data points in the ($E, q_{\text{max}}$) plane have been collected, which means that battery #4 has already been discharged 6 times. Then Eq. (16) is used to generate weights for the ensemble model. For illustration purposes, ensembles were built using both prior and posterior variational MLPs and show them hereafter. Fig. 7 summarizes the weight-distributed and the forecasted $q_{\text{max}}$ for battery #4 using this approach. Fig. 7(a) shows in orange the $q_{\text{max}}$ and respective prior variational MLP predictions for three batteries in the sub-population. It also shows the $q_{\text{max}}$ used to obtain the posterior variational MLP predictions (in black and green, respectively). The future points of the battery being monitored (gray dots) are shown there for illustration only. Fig. 7(b) shows the weights associated with each variational MLP coming from the sub-population of batteries. The weights indicate the similarity between each model and the current data points from battery #4 (left out for testing). Finally, Fig. 7(c) illustrates the forecasts before (orange lines) and after (green line and colored area) the Bayesian update. From left to right, the number of available points in the Bayesian update increases (from 6 to 16 to 26 points). As expected, reduction in forecast uncertainty is observed. This is true particularly when moving from 6 to 16 points. The ensemble model forecast varied less when moving from 16 to 26 points (indicating model convergence). The blue thick line and shaded area show predictions of the aging parameter $q_{\text{max}}$ in a pure data-driven regression fashion, ignoring data from other batteries of the same fleet. The “observation only” approach suggests the variability of the results when ignoring prior information from the fleet.

The ensemble model provides probabilistic estimates of the aging proxies at any given point in the future, and such values can be used to predict future voltage discharge cycles far-ahead in the battery life. This
approach allows us to obtain predictions with uncertainty estimates of what the future discharge curve will look like after the battery reaches a certain aging level. As done previously, battery #4 is used for testing the method. After collecting data points of \((E, q^{\text{max}})}\) and \((E, R_0)\), the posterior ensemble model is used to forecast the values of \(q^{\text{max}})\), \(R_0\) and their uncertainty from the last collected cumulative energy all the way to \(E = 2.5\) kWh. With that, the voltage discharge curve is predict assuming that the loading profile is known. Fig. 8 illustrate some of the results. Fig. 8(a) shows forecasted reference discharge cycles, which means that current held constant at 1 A. Fig. 8(b) shows forecasted random discharge cycles (see top-right panel of Fig. 4(b) for current profile). All forecasts are made for \(E = 2.5\) kWh. From the left to the right panel, data was collected up to when \(E = 0.5\) kWh, \(E = 1.5\) kWh, and \(E = 2.25\) kWh. Results obtained with the variational model using observations from battery #4 only are compared against results obtained with the ensemble model using fleet priors. As expected, the predictions from the model that use only observations are considerably worst when compared to the ensemble model. In both cases (with or without fleet priors), the prediction of the voltage drop changed as more data points are collected from the aging of the battery; and therefore, voltage discharge forecasts improved. The expected value of the discharge curve predicted using fleet priors is relatively accurate even with only 6 data points were collected. The uncertainty on future values of \(q^{\text{max}})\) and \(R_0\) propagates into large uncertainty bands on the future values of \(V\), as expected (see Fig. 7 for an illustration of how uncertainty in \(q^{\text{max}})\) varies with the number of data points).

5. Conclusions

In this paper, a hybrid physics-informed neural network approach to Li-ion battery prognosis is being proposed. The hybrid model combined known physics of discharge (such as Nernst and Butler–Volmer equations) with flexible machine learning for further uncertainty quantification and improved agreement with experimental data. We accounted for aging of the battery by modeling variation of \(q^{\text{max}})\) and \(R_0\) as a function of cumulative energy draw from the battery, also using multi-layer perceptrons trained with variational inference. Our hybrid model framework has multiple advantages. The hybrid model was capable of predicting voltage discharge curves as well as end-of-discharge of batteries subject to constant or random loading conditions. Our choice to track \((q^{\text{max}})\), \(R_0)\) for each discharge curve enables tailoring of the model to become battery-specific. Therefore, the aging effect caused by a charge–discharge cycles can be captured by updating the parameter pair. The hybrid model can be easily tuned and interpreted, as most computations within the recurrent neural network cell are driven by the physics-based portion of the model. Yet, it takes advantage of the high-flexibility of multi-layer perceptrons to correct for missing physics and to account for uncertainty in some critical model parameters.

The aging model, built upon two variational multi-layer perceptrons used to estimate \(q^{\text{max}})\) and \(R_0\) as a function of the cumulative energy draw from the battery, enables far-ahead prediction of voltage discharge curves by forecasting the values of \(q^{\text{max}})\) and \(R_0)\) far in the future. With the forecasted distributions of \(q^{\text{max}})\) and \(R_0)\), important operational outputs such as voltage drop and end of discharge can be predicted. In order to improve aging model forecasts, we proposed an ensemble model performing a weighted average of existing aging curves from older batteries. The fundamental hypothesis behind the approach is that the battery that is currently being monitored is a member of a fleet of similar batteries in a later aging state. Therefore, information from older batteries to forecast the capacity drop (through the proxies \(q^{\text{max}})\) and \(R_0)\) can be leveraged.

Uncertainty quantification using variational layers also successfully aid the aging forecast process. In all tested cases, predictions embraced the true discharge curves within their prediction intervals (see for example Figs. 7 and 8).

Among the limitations of this study, we can mention that the model was tested using a homogeneous data set, containing a series of discharge curves from similar batteries. All batteries used for testing were subject to similar loading conditions; either constant loading at 1 A or piece-wise random loading between 1 and 4 A. The aging
behavior might change slightly in case of substantially different loading conditions. Our ensemble-model prognosis was successful under the assumption that existing aging data from similar batteries are available. If data from older batteries were not available, the ensemble model output could suffer of the same volatility observed in the predicted discharge curves in Fig. 8, showed in light blue, where only data from the battery under testing was used. 

As future research, we expect to test the model under realistic loading profiles from small unmanned aerial vehicle flights and to extend the physics-informed neural network model to more powertrain components as well as to build a full hybrid powertrain model. A hybrid powertrain model would benefit not only end-of-discharge predictions, but also fault detection and isolation within the powertrain system.

CRediT authorship contribution statement

Renato G. Nascimento: Conceptualization of this study, Methodology, Formal analysis, Software. Matteo Corbetta: Conceptualization, Methodology, Writing, Supervision. Chetan S. Kulkarni: Conceptualization, Physics-based Modeling, Supervision. Felipe A.C. Viana: Methodology, Formal analysis, Writing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data set related to this article is open source, and can be found at the NASA Ames Prognostics Data Repository (Link), under set number 5, “Battery Data Set” and number 11, “Randomized Battery Usage Data Set”, (Saha and Goebel, 2007; Bole et al., 2014 [25]).

Acknowledgments

The contributions of Renato G. Nascimento, Matteo Corbetta, and Chetan Kulkarni were performed under NASA Ames Research Center, Contract No. 80ARC020D0010.

References

[34] B. Bole, C.S. Kulkarni, M. Daigle, Adaptation of an electrochemistry-based Li-ion battery model to account for deterioration observed under randomized use, in: Annual Conference of the Prognostics and Health Management Society, PHM Society, 2014, 9 pages.


