Alcibar, J., Aguirre Ortuzar, A., Aizpurua, J.I. (2026). A Probabilistic Physics-Aware Battery Health Management Approach for Inspection Drone Operations. In: Crespo Márquez, A., Seecharan, T., Abdul-Nour, G., Amadi-Echendu, J., Sánchez Herguedas, A., Goti Elordi, A. (eds) Digital Maintenance and Asset Digitalization. Engineering Asset Management Review, vol 5. Springer, Cham. https://doi.org/10.1007/978-3-032-21502-4_11

A Probabilistic Physics-aware Battery Health Management Approach for Inspection Drone Operations

Jokin Alcibar[1,*], **Aitor Aguirre-Ortuzar**[1], **and Jose I. Aizpurua**[2,3]

1 Electronics and Computer Science Department, Mondragon University, Loramendi 4, 20500 Mondragón, Spain

2 Computer Science and Artificial Intelligence Department, University of the Basque Country (UPV/EHU), 20018 Donostia, Spain

3 Ikerbasque, Basque Foundation for Science, 48011 Bilbao, Spain

*Corresponding author: jalcibar@mondragon.edu

Abstract The increasing deployment of inspection drones for monitoring remote and critical infrastructure presents new opportunities and challenges in asset management. These drones operate in demanding environments, where ensuring operational reliability is essential. Among the various subsystems, battery health plays a central role in determining mission success and safety. This chapter presents a physics-aware probabilistic approach for battery health management, integrating data-driven techniques with physics-based models to improve the predictability of battery performance. At the core of this methodology lies a probabilistic machine learning model that provides uncertainty quantification, enabling more informed and robust decision-making. By incorporating this uncertainty-aware perspective into bat--ery discharge forecasting, the approach supports advanced digital maintenance strategies. The methodology is demonstrated to drone-based inspections of offshore wind energy infrastructure, highlighting its contribution to enhancing asset reliability and enabling condition-aware maintenance strategies.

1 Introduction

Unmanned Aerial Vehicles (UAVs) are widely used in civilian operations including surveillance and inspection through their vertical take-off capabilities and precision in remote environments rations (Valavanis and Vachtsevanos, 2015). They significantly enhance safety and data collection efficiency in hazardous conditions (Eleftheroglou et al., 2019). However, UAV reliability remains challenged by battery limitations, payload constraints, and weather conditions that can cause mission failures (Gatti et al., 2015).

Digital maintenance strategies highlight the critical role of Battery Management Systems (BMS) in real-time diagnostics and intelligent decision support (Sierra et al., 2019). These systems enable proactive fault detection and dynamic path replanning to ensure mission continuity. However, accurate end-of-discharge (EOD) voltage prediction faces challenges from environmental uncertainties and measurement errors (Sierra et al., 2018). In addition, reliable estimation of State of Health (SOH) and Remaining Useful Life (RUL) are necessary to quantify battery degradation and anticipate end-of-service conditions during operation. Integrating uncertainty modeling into battery degradation analysis is essential for developing robust predictive maintenance methodologies supporting autonomous UAV operations (Zhang et al., 2022).

Battery health management methods are categorized as physics-based, data-driven, or hybrid (Demirci et al., 2024; Vanem et al., 2021). Physics-based methods use physical models, often updated with algorithms like Particle Filtering (PF) (Djuric et al., 2003) or Unscented Kalman Filter (UKF) (Bai et al., 2023) to estimate battery health. Data-driven approaches instead rely on degradation data to model capacity fade (Mitici et al., 2023). Finally, hybrid methods integrate both strategies, leveraging the accuracy of physical modeling and the adaptability of data-driven techniques (Guo et al., 2015).

Physics-based Approaches

Physics-based approaches to battery degradation have been explored in various studies. Daigle and Kulkarni (2013) proposed a prognostics method grounded in electrochemical modeling, where the system state is updated using the UKF. This approach yielded precise EOD predictions with quantified uncertainty, though it demands the estimation of numerous parameters, complicating optimization. In another contribution, Pola et al. (2015) developed an empirical state-space model for estimating state of charge (SOC) and EOD using a PF. Their model accurately forecasted battery discharge duration and offered confidence intervals. However, it lacked adaptability for multi-cell batteries due to a static open-circuit voltage curve. In this direction, Sierra et al. (2019) designed a prognostics framework for Li-Po batteries that integrates a simplified battery model with evolutionary strategies for parameter estimation. Their method includes a feedback mechanism to correct process noise variance, reducing estimation bias. Nonetheless, such physics-based methods often require intensive computation, limiting real-time deployment.

Data-Driven Approaches

Several data-driven methods have been proposed to overcome the limitations of physics-based models (Ng et al., 2020). Data-driven methods, such as Artificial Neural Networks (ANNs), Gaussian Processes, and Support-Vector Machines, have addressed EOD prediction and degradation inference (Ochella et al., 2022). Deep learning techniques, including Convolutional Neural Networks (CNNs) (Couture & Lin, 2022; Mitici et al., 2023), Recurrent Neural Networks (RNNs) (Guo et al., 2024; Zhao et al., 2023), and hybrid models (Mazzi et al., 2024), excel in extracting temporal features and mapping nonlinear relationships. However, purely data-driven approaches face limitations in robustness for unseen data and uncertainty quantification (Aizpurua et al., 2023; Arias Chao et al., 2022).

To improve uncertainty estimation, advanced methods like encoder-decoder architectures with Monte Carlo (MC) dropout (Biggio et al., 2023), Bayesian techniques (Zhang et al., 2022), and Probabilistic Neural Networks (Che et al., 2024) have been developed. While these enhance prediction accuracy and uncertainty quantification, data-driven models remain largely interpretability-limited, often seen as black-box solutions.

Hybrid Approaches

Hybrid modeling approaches that combine physical principles with data-driven techniques have gained attention for enhancing interpretability in predictions (Fink et al., 2020). For instance, Yeregui et al. (2023) developed a model integrating a Reduced Order Model (ROM) with a Long Short-Term Memory (LSTM) network to estimate the SOC in lithium-ion batteries. Nevertheless, this work did not analyze uncertainty. Similarly, Fernández et al. (2023) introduced a Bayesian neural network guided by physical knowledge to improve predictions in lateral-load test scenarios. Their model uses Approximate Bayesian Computation instead of conventional loss functions, leading to better performance in prediction and uncertainty quantification (UQ).

A related methodology involves Physics-Informed Neural Networks (PINNs), which incorporate governing physical laws into the training process by embedding them in the loss function. PINNs have shown strong results across engineering fields, such as computational mechanics (Li et al., 2021), fluid dynamics (Raissi et al., 2020), material engineering (Zhang and Gu, 2021), and fault diagnosis (Feng et al., 2023; Ni et al., 2023). More recently, Sun et al. (2023) proposed a physically informed LSTM model for battery prognostics, though without addressing uncertainty. Nascimento et al. (2023) used a physics-guided RNN framework with ROMs and neural networks to model voltage discharge and represent model uncertainty.

Table 1. Summary of recent battery health-state estimation methods.

Ref.	**Dataset**	**Model**	**Output**	**UQ**
(Sierra et al., 2019)	Operational	Physics based	*EOD_t, SOC	Yes
(Sierra et al., 2018)	Operational	Physics based	*EOD_t, SOC	None
(Bai et al., 2023)	Laboratory	Physics based	RUL	Yes
(Hong et al., 2020)	Laboratory	Data-Driven	RUL	Yes
(Hsu et al., 2022)	Laboratory	Data-Driven	**EOD_v, EOL	None
(Zhao et al., 2023)	Laboratory	Data-Driven	SOH	None
(Han et al., 2022)	Laboratory	Data-Driven	SOH, RUL	None
(Meng et al., 2023)	Laboratory	Data-Driven	SOH	None
(Biggio et al., 2023)	Laboratory	Data-Driven	**EOD_v	Partial
(Zhang et al., 2022)	Laboratory	Data-Driven	RUL	Yes
(Yeregui et al., 2023)	Laboratory	Hybrid	SOC	None
(Sun et al., 2023)	Laboratory	Hybrid	SOH	None
(Nascimento et al., 2023)	Laboratory	Hybrid	SOC	Yes
Proposed Approach	Operational	Hybrid	**EOD_v	Yes

*Legend: *EOD_t: EOD time. **EOD_v: EOD Voltage.*

Table 1 summarizes recent advancements in battery health-state estimation and emphasizes the key contributions of this research. As shown in the table, only a few works have focused on analyzing drone battery data collected under real-world operating conditions (Sierra et al., 2019, 2018). Most existing research relies on battery data obtained from controlled laboratory discharge processes. Additionally, many analytical approaches fail to explicitly quantify or differentiate between data uncertainty and model uncertainty. Addressing these gaps, this study introduces a physics-aware probabilistic model that leverages real-world operational battery data to quantify uncertainty.

This paper presents a novel hybrid probabilistic approach for predicting EOD voltage in Li-Po batteries. The method combines MC dropout for epistemic uncertainty and Negative Log Likelihood (NLL) loss for aleatoric uncertainty within a residual architecture. The approach integrates a physics-based discharge model with a probabilistic error-correction component, reducing computational costs and data requirements. The physics-based model extracts battery behaviour features without complex electrochemical details, enhancing efficiency and scalability for complex scenarios. This strategy offers a robust, efficient, and data-aware solution for battery voltage prediction under uncertainty.

Subsequently, probabilistic error correction is implemented by explicitly integrating uncertainty via CNN and MC dropout. A benchmarking analysis is conducted to assess the accuracy and uncertainty of various probabilistic models.

The proposed approach effectively addresses various battery parameter errors, such as calibration issues, enabling accurate EOD voltage prediction across diverse drones. Its effectiveness is demonstrated using an industrial dataset from real drone inspections involving varied manoeuvres and loading conditions, offering practical value to drone fleet operators.

2 Understanding Predictive Uncertainty

Battery EOD voltage predictions can be evaluated based on both accuracy and uncertainty. Uncertainty reflects the model's confidence in its output (Heo et al., 2018), but low uncertainty values alone do not guarantee reliability. This reliability depends on calibration, meaning that the predicted uncertainty must align with the true likelihood of error (Rathnakumar et al., 2023).

To build robust predictive models, it is essential to understand and represent the sources of uncertainty present in the modeling process. In general, uncertainty in machine learning is categorized into two main types: **aleatoric** and **epistemic** uncertainty (Der Kiureghian and Ditlevsen, 2009). Capturing these uncertainties often involves moving beyond deterministic point estimates and instead inferring full probability density function (PDF) over predicted outcomes (Hüllermeier and Waegeman, 2021).

Aleatoric uncertainty arises from inherent randomness in the data-generating process. This form of uncertainty is considered irreducible, as it stems from factors such as measurement noise or environmental variability (Kendall and Gal, 2017). It reflects the stochastic nature of the world, and even with perfect knowledge of the system, such variability cannot be eliminated. In predictive modeling, quantifying aleatoric uncertainty requires characterizing the conditional distribution of the output given the input, rather than just its mean (Tagasovska and Lopez-Paz, 2019).

$$Var(\epsilon|X) = \frac{1}{T}\sum_{t=1}^{T} \hat{\sigma}_t^2\,(X) \qquad Equation\ 1$$

here $\sigma^2(X)$ indicates that the variance of residuals is dependent on X.

In contrast, **epistemic uncertainty** reflects the model's ignorance or lack of knowledge. Unlike aleatoric uncertainty, epistemic uncertainty is reducible: it can be decreased by collecting more data, particularly in regions of the input space that were underrepresented during training (Tagasovska and Lopez-Paz, 2019). This type of uncertainty is especially relevant evaluating model robustness. One common way to capture epistemic uncertainty is to examine the variability of predictions across multiple models trained under different assumptions or initializations (Lakshminarayanan et al., 2017). Such ensemble-based approaches reveal the uncertainty due to limited information, such as ambiguity in model parameters or architecture choices (Hüllermeier and Waegeman, 2021). Epistemic uncertainty is rep--esented as (Kendall and Gal 2017):

$$Var(y) = \frac{1}{T}\sum_{t=1}^{T} f\,(x,\theta_t)^2 - \left(\frac{1}{T}\sum_{t=1}^{T} f\,(x,\theta_t)\right)^2 \qquad Equation\ 2$$

where $f(x,\theta_t)$ denotes the output of the model for the t-th sample, x signifies the sample in question and θ represents the parameters of the model.

The **total uncertainty** (TU), which integrates both epistemic uncertainty (EU) and aleatoric uncertainty (AU), can be expressed as follows (Abdar et al., 2021):

$$TU = \mathrm{EU} + \mathrm{AU} \qquad Equation\ 3$$

This formulation ensures consistent units among the uncertainty terms and adheres to the law of total variance, which governs uncertainty propagation in Bayesian predictive models.

3 Uncertainty-Aware Battery Health Management Approach

Figure 1 illustrates the proposed uncertainty-aware battery health management approach, which consists of three main stages: (i) EOD voltage prediction, (ii) post-processing UQ and (iii) health state diagnosis.

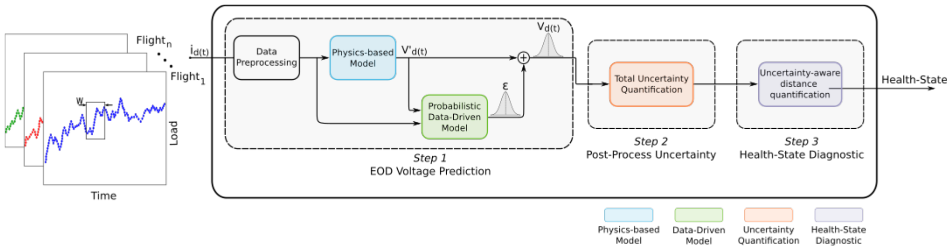

Figure 1. Uncertainty-aware battery health management approach.

3.1 EOD Voltage Prediction

This section details Step 1 of Figure 1. The EOD voltage prediction combines a physics-based model with a data-driven model in an error-correction configuration. Before these steps, a preprocessing stage is performed to clean the monitored data by removing empty variables, duplicate records, and erroneous sensor readings. Missing values are imputed using average values to maintain data consistency.

The physics-based model captures the discharge behaviour of the battery. Given the load at instant *t*, denoted as $i_{d(t)}$, it estimates the terminal voltage $v'_{d(t)}$. This estimate is then passed to the probabilistic data-driven model, which predicts the error. Finally, the last stage applies dynamic error correction to estimate the discharge voltage $v_{d(t)}$.

Physics-based discharge model

The physics-based discharge model is electrochemistry-based battery discharge model is inspired from Daigle and Kulkarni (2013) and implemented through the Prognostics Model Library (Teubert et al., 2023) from NASA Ames Research Center. This model operates using a set of ordinary differential equations to determine the total battery voltage, *V(t)*, which is defined as the difference in potential between the positive and negative current collectors, accounting for the resistance losses come across these current collectors. The state-space of the battery model, *x(t)*, input vector, *u(t)*, and output vector, *y(t)* are defined by the next equations (Salinas-Camus et al., 2023):

$$x(t) = \left[q_{s,p}, q_{b,p}, q_{s,n}, q_{b,n}, V', V'_{n,p}, V'_{n,n}\right]^T \qquad \text{Equation 4}$$

$$u(t) = \left[i_{app}\right] \qquad Equation\ 5$$

$$y(t) = [V] \qquad Equation\ 6$$

The state-space includes q and V', corresponding to the charge on the electrodes and the voltage differentials, respectively. The notations p and n indicate the positive and negative electrodes, which are divided into two volumes representing the surface layer (s) and the bulk layer (b). The input vector, indicates the electric current applied to the battery and the output vector denotes the discharge voltage of the battery.

Probabilistic Data-driven

A probabilistic CNN was developed for the error-correction stage, leveraging its strengths in feature extraction and uncertainty estimation. In drone operations, identifying uncertainty is critical. Aleatoric uncertainty arises from inherent system randomness, such as wind or sensor noise, while epistemic uncertainty results from gaps in knowledge, including modeling errors. Distinguishing between these types of uncertainties enables better decision-making and mitigate risks associated with uncertain information.

CNNs are deep learning models designed to exploit local spatial or temporal patterns, allowing them to capture meaningful structured features in high-dimensional data (Xu et al., 2022). Applying Monte Carlo dropout during both training and testing enhances their robustness and provides a more comprehensive probabilistic interpretation (Choubineh et al., 2023; Mitici et al., 2023). Figure 2 shows the structure of the proposed CNN-based error-correction model, with each layer described below.

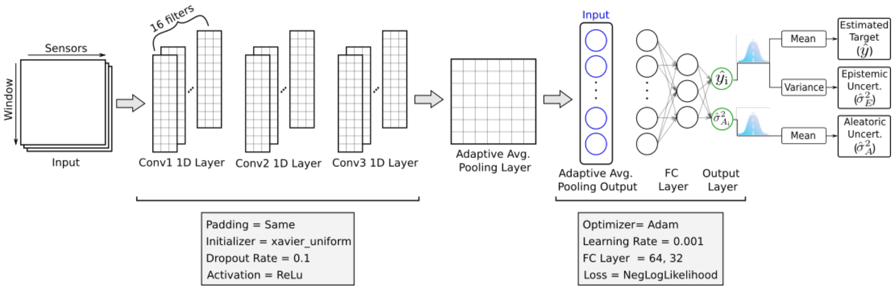

Figure 2. Structure of the proposed uncertainty-aware CNN architecture.

- *Input data*: the input to the CNN is organized as a tensor. Each row represents a data sample obtained through windowing. Each column corresponds to a sensor measurement from the inspection drone, such as the current load or the estimated voltage computed using a physics-based model.
- *Convolutional layers* (Conv1 to Conv3 in Figure 2): These layers consist of Conv1D layers, which apply a series of convolutional filters to extract features from the input data. The filters scan the input sequence to identify distinctive patterns related to the target prediction variable. In the context of battery behavior, these filters specifically detect temporal patterns such as voltage-current discharge characteristics, capacity degradation trends across cycles, and variations in discharge profiles that indicate early-stage degradation signatures. Each convolutional layer uses 16 filters to generate the feature maps.
- *Adaptive Average Pooling layer*: In this layer, the average of all activations is computed for each feature map, producing a single scalar value per map. This operation reduces the spatial dimensions and provides a compact summary of each feature map. This layer provides a fixed-size representation independent of the input sequence length and offers a smoother, more stable summary than max-pooling, which is advantageous given the noisy nature of battery signals.
- *Fully Connected (FC) layers:* these layers transform the extracted features into predictions using activation functions. Two fully connected (FC) layers are used, allowing the model to learn complex relationships and patterns from the features. This enables accurate predictions based on the input data.
- *Output layer:* this layer generates the final output based on the inputs and the learned weights from the previous layers. It includes two neurons, one representing the mean and the other the variance, to capture both the expected value and its associated uncertainty. To ensure the variance remains positive, an exponential activation function is applied to that neuron.

Table 2. Architectural and Optimization Hyperparameters of CNN.

Category	Hyperparameter	Value
Architecture	Window-size	10
	Convolutional layers	3
	Number of filters	16
	Kernel size	3
	Padding	Same
	Initializer	Xavier Uniform
	Activation	Rectified Linear Unit
	Number of fully connected layers	2
	Number of nodes in fully connected layers	64,32

Optimization	Loss Function	NLL
	Optimizer	Adam
	Learning Rate	0.001
	Number of epochs	130
	Monte Carlo dropout rate	0.1
	Number of Monte Carlo inferences	100

Dropout layers are strategically placed between the fully connected layers during training to prevent overfitting by randomly deactivating neurons in the network. At inference time, the MC dropout technique is applied to approximate a predictive distribution in a computationally efficient manner (Gal and Ghahramani, 2016). As shown in Figure 3, during training, each distinct dropout mask effectively simulates a different draw from the approximate posterior distribution over the model's parameters. This stochastic sampling enables the estimation of epistemic uncertainty by computing the variance (σ_E^2 in Figure 2) across multiple forward passes improving the model's robustness and predictive accuracy.

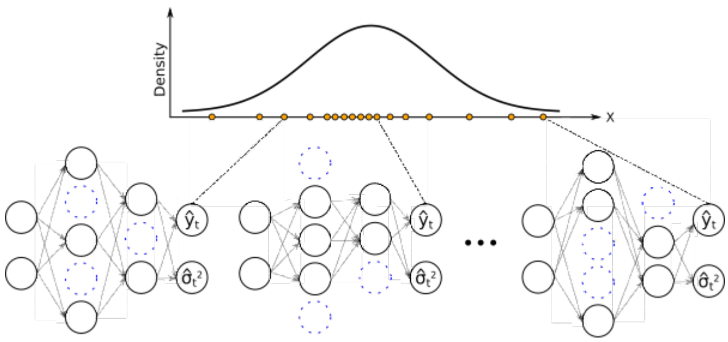

Figure 3. Use of dropout at test stage with MC sampling.

Finally, to capture aleatoric uncertainty, assuming it follows a Gaussian distribution $N(\mu, \sigma^2)$, the CNN architecture is modified to include an additional output for the variance σ^2, as shown in Figure 2. Combined with a negative log-likelihood (NLL) loss function, this allows the model to learn uncertainty directly from the data (Kendall and Gal, 2017).

$$\mathcal{L}_{NN}(y_{true}|\mu, \sigma^2) = \frac{1}{N}\sum_{i=1}^{N}\frac{(y_i - \mu(x_i))^2}{2\sigma(x_i)^2} + \frac{1}{2}log\sigma(x_i)^2 \qquad Equation\ 7$$

where y_i is the true value for the i-th element in the batch, $\mu(x_i)$ is the predicted mean, and $\sigma(x_i)^2$ represents the variance of the distribution.

When the CNN predicts the target variable accurately, the estimated variance is low, minimizing the MSE loss in Equation 7. In high data uncertainty scenarios, the CNN increases the variance to reflect the noise. Thus, the variance (i) quantifies prediction uncertainty and (ii) balances capturing data variability with avoiding overly uncertain predictions.

This approach improves the ability of the CNN to provide accurate predictions while capturing data uncertainty, making it a valuable asset for decision-making in inspection drones.

3.2 Post-Process Uncertainty

This section describes Step 2 of Figure 1, which corresponds to the post-processing stage. This stage aims to infer the total uncertainty to enhance decision-making under uncertain conditions.

Total Uncertainty Quantification

Integrating aleatoric and epistemic uncertainty is essential for capturing and interpreting the total uncertainty within a system, process, or model. To address this, Algorithm 1 outlines the complete procedure for quantifying total uncertainty.

Algorithm 1 Total Uncertainty Inference

Input: data X_{test}, prediction model $h(\cdot)$, number of iterations N
Output: total uncertainty σ_{TU}
for each i $\in N$ **do**
  $[y_i, \sigma^2_{A_i}] = h(X_{test_i})$ ▷ cf.Figure 2
  $\vec{y}[i] = y_i$ ▷ Store mean results
  $\vec{\sigma_A}^2[i] = \sigma^2_{A_i}$ ▷ Store variance results
end for
$\hat{y} = mean(\vec{y})$ ▷ Infer mean pred. values
$\sigma^2_E = var(\vec{y})$ ▷ Infer Ep. Uncertainty
$\sigma^2_A = mean(\vec{\sigma^2_A})$ ▷ Infer Al. Uncertainty
$\sigma_{TU} = \sigma^2_E + \sigma^2_A$
return σ_{TU}

Summing the squares of individual uncertainties emphasizes the contribution of larger ones, as they have a greater influence on the overall measure. This approach ensures that significant sources of uncertainty are not underestimated. Finally, taking the square root of the sum restores the total uncertainty to the original scale, providing a interpretable measure of overall uncertainty.

3.3 Battery Health-State Diagnostics

This section describes Step 3 of Figure 1. Drone health-management decisions are made based on available data, observations, and evidence. In this context, UQ plays a key role by accounting for model and data-related uncertainty, supporting decision-making under uncertain conditions.

In the proposed methodology, uncertainty analysis is used to estimate the battery's health state. A prediction interval, learned from healthy battery data, is defined based on total uncertainty. When new flight data from the inspection drone is available, (i) EOD voltage predictions are made considering this uncertainty, and (ii) the result is assessed to determine whether the observed voltage falls within the predicted range.

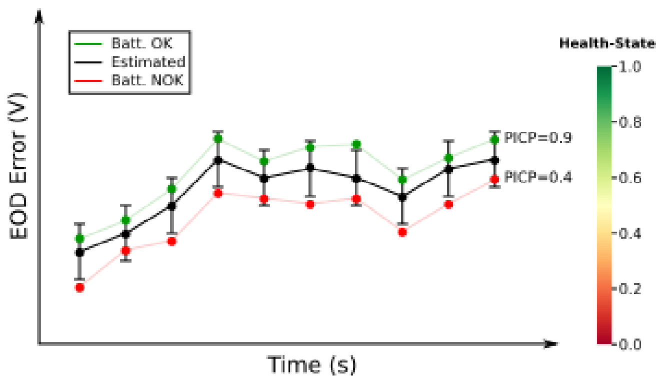

Figure 4. Health-state inference example based on prediction interval quantification.

The Prediction Interval Coverage Probability (PICP) metric quantifies the proportion of true values that fall within the predicted interval defined by total uncertainty (González-Sopeña et al., 2021). Figure 4 illustrates how PICP is used to infer battery health. The green line shows cases where most true values fall within the confidence intervals, indicating a healthy battery (Batt. OK). In contrast, the red line highlights cases where most values lie outside the intervals, suggesting poor battery health (Batt. NOK).

Based on PICP results, a health-state diagnosis is computed for each flight. This metric is essential for evaluating predictive performance and serves as a key indicator in the battery health diagnostic system.

4 Case Study

The proposed approach is evaluated using data from multiple flights conducted with inspection drones. These drones are used for the inspection of defects in offshore wind turbines, e.g. cracks, which are processed through on-boarding processing software. All drones are equipped with Li-Po batteries. In this study, a single cell extracted from a 6S Li-Po battery pack is analysed, where 6S denotes six cells connected in series. The full battery pack has a nominal capacity of 30,000 mAh.

During each flight, several parameters are recorded, including load and discharge voltage. Current is measured via an ACS758 Hall effect sensor placed on the main positive wire to reduce the risk of short circuits. Voltage is monitored using a filtered voltage divider circuit, helping to prevent false return-to-launch events, especially under strong winds that cause sudden motor speed increases. While external factors such as ambient temperature and pressure also influence battery discharge, their effects are manifested through changes in current consumption. Accordingly, it is assumed that the current variable implicitly captures the environmental variability, and the estimation therefore relies on it to accurately predict the EOD voltage.

Although ambient temperature and pressure can influence battery discharge, their effects are largely manifested through changes in current consumption. Because our model directly incorporates the load current, we assume that this variable implicitly captures most of the environmental variability. For this reason, temperature and pressure were excluded from the analysis. We note this as a limitation and identify the explicit integration of these factors as future work.

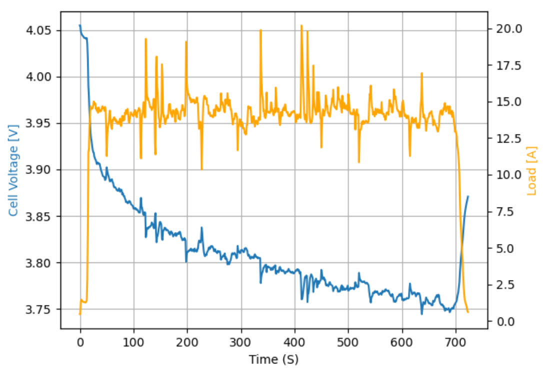

Figure 5. Available datasets for the flight #73.

Figure 5 presents a sample of the variables collected per flight in this study. The dataset comprises 26,156 samples gathered from 33 separate flight sessions. The number of samples used for validation varies depending on the duration of each flight. To process this large volume of data, a laptop with an Intel® Core™ i5-10210U processor (1.60 GHz) and 16 GB of RAM was used, leveraging the PyTorch framework for deep learning analysis.

Table 3 shows the distribution of how flight data has been split into training and testing subsets over a three-year period. It also presents the total number of recorded flights for each year.

Table 3. Train-test split on the flights within the dataset.

	2021	**2022**	**2023**	**Total**
Train	18	3	3	24(72.72%)
Test	7	1	1	9 (27.27%)
Total	25	4	4	33 (100%)

The CNN hyperparameters shown in Table 2 were tuned based on the training dataset. Optimal performance was achieved using the Adam optimizer with a learning rate of 0.001, applying 16 filters, and configuring the two fully connected layers with 64 and 32 neurons, respectively.

5 Results

The validation of the proposed method is carried out in two stages. The first phase involves benchmarking the accuracy and robustness of EOD voltage prediction models to identify the most effective approach. The second phase addresses battery health diagnostics, where flight data from different batteries is analyzed to estimate uncertainty and evaluate their health status.

5.1 EOD Voltage Prediction

The predictive accuracy of the proposed CNN model has been compared with other probabilistic models. Namely, Quantile Linear Regression (QLR) (Chung, Neiswanger, et al. 2021), Quantile Regression Forest (QRF) (Meinshausen 2006), and Quantile Gradient Boosting (QGB) (Friedman 2001) have been selected as these probabilistic models exhibit high performance in analogous scenarios, demonstrating robustness and accurate predictions across varied applications (Song et al. 2022; Bracale et al. 2023; Jiang and Meng 2023).

Table 4 shows the Continuous Ranked Probability Score (CRPS) metric to evaluate the accuracy and calibration of the probabilistic forecasts (Zamo and Naveau, 2018), along with the corresponding standard deviations (std) tested on real flight data (Table 3).

Table 4. CRPS (std) of different probabilistic models tested across real flights.

Flight (#)	Sample (%)	QLR	QRF	QGB	CNN MC dropout
11	14.14	0.023 (0.007)	0.028 (0.021)	0.024 (0.019)	**0.022 (0.003)**
30	15.92	0.037 (0.013)	0.037 (0.024)	0.039 (0.033)	**0.023 (0.003)**
39	15.51	0.017 (0.005)	0.027 (0.023)	0.021 (0.027)	**0.016 (0.006)**
44	3.99	0.031 (0.009)	0.036 (0.021)	0.033 (0.019)	**0.026 (0.003)**
69	13.83	0.029 (0.011)	0.034 (0.022)	0.035 (0.031)	**0.023 (0.002)**
73	14.71	0.031 (0.009)	0.036 (0.024)	0.038 (0.034)	**0.023 (0.003)**
85	13.41	0.045 (0.014)	0.066 (0.037)	0.049 (0.056)	**0.034 (0.008)**
88	8.49	0.027 (0.007)	0.032 (0.019)	0.035 (0.031)	**0.022 (0.002)**
TOTAL	100	0.027 (0.009)	0.037 (0.023)	0.034 (0.031)	**0.023 (0.003)**

CRPS metric is computed for each test sample within every flight, resulting in a collection of CRPS values per flight. This collection is then modeled using a Gaussian probability distribution, defined by its mean and standard deviation. A smaller standard deviation indicates higher consistency between the predicted distribution and the actual outcomes. It is important to clarify that this standard deviation refers to the variability of the CRPS values across samples, not the dispersion of the predictive distribution itself.

Results indicate that the CNN MC dropout achieves superior performance compared to the other probabilistic approaches. It yields both the lowest prediction error and the smallest standard deviation, suggesting that its forecasts are not only more accurate but also more stable and reliable.

Figure 6 provides an overview of error prediction for a single flight using various data-driven error-correction methods. The results show that (i) all physics-aware models effectively capture voltage discharge error dynamics, and (ii) each approach incorporates different sources of uncertainty through probabilistic error estimates.

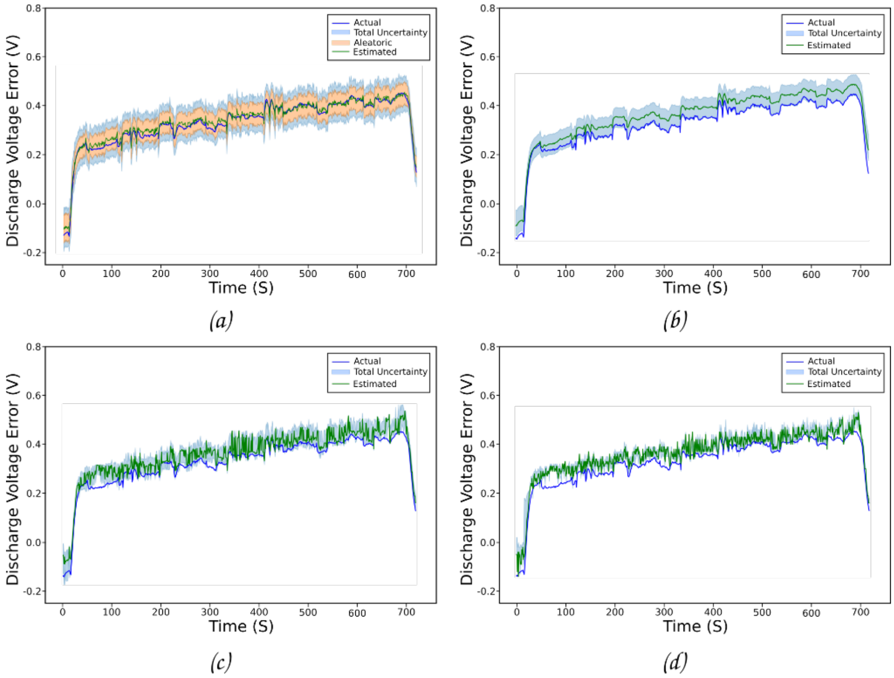

Figure 6. Error correction performance across different models for flight #73 (a) Error correction with CNN Dropout (b) Error correction with QLR (c) Error correction with QRF (d) Error correction with QGB.

The performance results in Table 4 are consistent with the flight results in Figure 6. Specifically, the CNN with MC dropout achieves superior probabilistic accuracy and reliability regarding voltage estimation errors and prediction consistency compared to other evaluated models [see Figure 6(a)]. This is further supported by the low standard deviation of 0.009 volts in Table 4. In contrast, Figures 6(c) and 6(d) show the largest prediction errors and highest variance, which aligns with the data for Flight #73 (QRF, QGB) in Table 4. Based on these errors, Figure 7 illustrates the overall voltage drop of a single-cell battery alongside its uncertainty.

Evaluating the PDF is essential for quantifying uncertainty. To this end, calibration and sharpness are evaluated. Calibration reflects the statistical consistency between predicted distributions and actual observations, representing a joint property of forecasts and empirical data (Jung et al., 2022). Sharpness indicates how concentrated the predictive distributions are, indicating how tightly confidence intervals focus around a single value (Kuleshov et al., 2018).

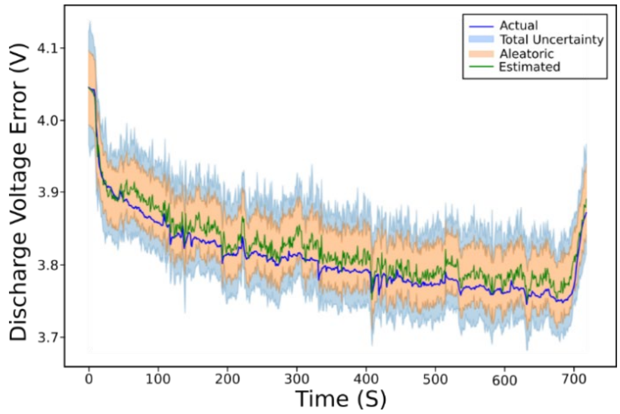

Figure 7. EOD Voltage estimation of a battery cell in Flight #73 using CNN MC dropout.

Figure 8 shows the calibration plots of the developed models, where the X-axis represents predicted probabilities and the Y-axis shows observed outcomes. Ideally, a well-calibrated model aligns along the diagonal, indicating close agreement between predictions and actual results. Calibration quality, illustrated by solid blue lines, is assessed by the area between the calibration curve and the diagonal, referred to as the miscalibration area, represented in blue shaded area. Smaller areas imply better calibration (Tran et al., 2020). The CNN MC dropout model [Figure 8(a)] displays the lowest miscalibration area of 0.04. In contrast, the QRF and QGB models have higher miscalibration areas of 0.21 and 0.23, indicating poorer calibration. The QLR model falls in between, with a miscalibration area of 0.15, showing acceptable performance compared to the other quantile models.

At the same time, sharpness is assessed as a measure of how concentrated the predictive distributions are (Tran et al., 2020). Since the model outputs are Gaussian distributions, sharpness is typically quantified using the predicted variance. A lower variance indicates a sharper distribution. However, it is important to note that there is often a trade-off between how well a model is calibrated and the sharpness of its predictions (Gneiting, Balabdaoui, & Raftery, 2007). Figure 8 illustrates the predictive uncertainty distributions for all probabilistic models, with vertical lines representing their corresponding sharpness levels.

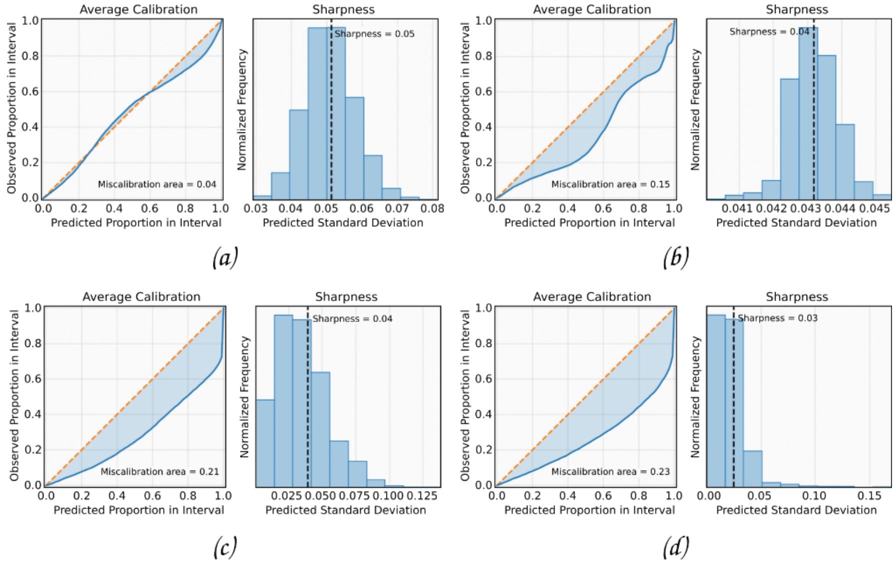

Figure 8. Calibration and sharpness metrics for all UQ methods used in this study. (a) CNN MC dropout model (b) QLR model (c) QRF model (d) QGB model.

As illustrated in Figure 8(d), QGB presents the lowest sharpness, indicating minimal variability in its predictions. However, this model also exhibits inferior calibration, aligning with earlier findings. This highlights the necessity of balancing calibration and sharpness. Within this framework, QLR stands out by achieving an effective compromise between these two aspects, with a sharpness value of 0.04V, as shown in Figure 8(b).

On the other hand, although the CNN MC dropout model exhibits the highest sharpness, the difference compared with the other models is minimal [Figure 8(a)]. It achieves a sharpness of 0.05 V, only 0.01 V higher than that of QLR. This small margin suggests that the increased sharpness of the CNN MC dropout model does not significantly compromise its calibration.

The results demonstrate that the proposed probabilistic physics-aware model outperforms alternatives in terms of CRPS, calibration, and sharpness. This superior performance validates the contribution of the methodology and highlights its effectiveness in predicting EOD voltage.

5.2 Health-State Diagnostic

Table 5 presents the estimated health-states derived from eight different drone flights, each fitted with distinct distinct Li-Po batteries sharing identical capacity specifications. Specifically, each drone flight in the testing dataset employs unique batteries with matching capacity characteristics. Battery degradation was qualitatively evaluated by experts, who classified battery health into categories of good, medium, and bad based on their operational history and usage conditions. Nevertheless, this qualitative classification does not offer precise quantitative metrics. To achieve quantitative degradation measurement, total predictive uncertainty is employed, represented numerically as illustrated in Figure 4, where 0 corresponds to degraded battery health and 1 signifies optimal battery condition.

Table 5. Health-state estimation of different models for different real flights equipped with batteries in different health conditions.

Flight (#)	**QLR**	**QRF**	**QGB**	**CNN MC dropout**	**Ground Truth**
11	0.956	0.806	0.601	**0.998**	Good
30	0.602	0.471	0.191	**0.993**	Good
39	**1.0**	0.841	0.605	0.997	Good
44	0.938	0.903	0.541	**1.0**	Good
69	0.834	0.659	0.306	**1.0**	Good
73	0.763	0.506	0.157	**1.0**	Good
85	0.283	0.184	0.068	**0.948**	Good
88	0.961	0.727	0.241	**1.0**	Good

A detailed analysis of the results presented in Table 4 indicates that the UQ modeling based on the QGB method is less accurate in determining the true health-state of the batteries across different flights. In comparison, the QRF method achieves superior performance, with significantly better results in identifying the batteries corresponding to flights #11, #39, #44, and #88.

On the other hand, the QLR method demonstrates robust health-state index modeling for most batteries, except for flights #30 and #85. In these particular cases, a substantial number of observations fall outside the uncertainty interval, pointing to a suboptimal battery health condition.

Finally, the CNN MC dropout demonstrates the most effective uncertainty modeling, successfully estimating health-states across various batteries, as validated by domain experts. By integrating both aleatoric and epistemic uncertainties, the model offers detailed and accurate insights, allowing the probabilistic CNN to produce a well-calibrated, uncertainty-aware health indicator (Abdar et al., 2021). This method of uncertainty quantification supports the final contribution of the work, showing that jointly addressing aleatoric and epistemic uncertainty enhances the overall quality of decision-making.

6 Discussion

This research presents a robust probabilistic physics-aware battery health management approach for inspection drones. However, before reaching conclusions, it is important to discuss key aspects of the developed methodology in more detail.

6.1 Uncertainty Decomposition

The proposed CNN MC dropout method demonstrates enhanced performance in both accuracy and uncertainty quantification (cf. Section 5.1). Notably, this approach quantifies aleatoric and epistemic uncertainty. As shown in Figure 9, the blue region indicates data-related uncertainty, typically caused by sensor noise (aleatoric), while the purple-shaded region reflects model uncertainty (epistemic).

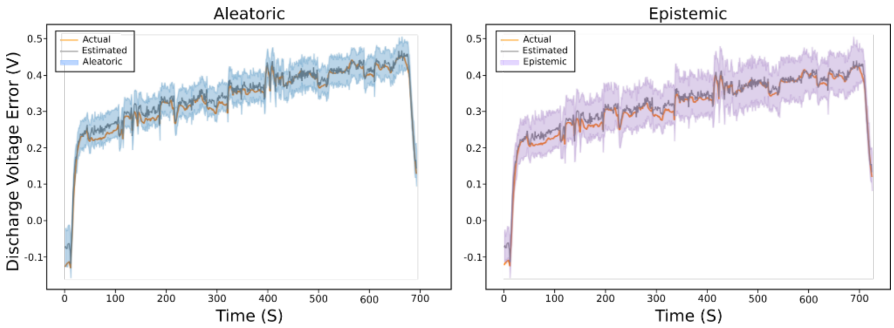

Figure 9. Estimation of EOD Voltage Correction in Flight #73 using CNN MC dropout for the Discrimination of Aleatoric and Epistemic Uncertainty.

The analysis shows epistemic uncertainty is greater than aleatoric uncertainty, indicating that methods to reduce epistemic uncertainty like improving model calibration or expanding training data could address the primary uncertainty source. Distinguishing between uncertainty types is a crucial benefit of this method. The proposed UQ approach offers important insights that enhance our understanding of uncertainty compared to alternative probabilistic approaches (cf. Section 5).

6.2 Sensitivity analysis of dropout rate

To assess how the dropout rate affects the outcomes, this section examines its impact on various performance measures. Using flight #73 as a benchmark, the analysis evaluates how changes in the dropout rate influence calibration, sharpness, and CRPS metrics. The resulting findings are presented in Table 6.

Table 6. Sensitivity analysis of dropout rate for the flight #73.

Dropout Rate	Calibration	Sharpness	CRPS
0.01	0.06	0.05	0.0178
0.05	0.09	0.09	0.0267
0.1	0.24	0.18	0.0482
0.15	0.28	0.21	0.0503
0.2	0.29	0.23	0.0561

The results show that the dropout rate has a significant impact on calibration, sharpness, and CRPS. In particular, increasing the dropout rate from 0.01 to 0.2 leads to a deterioration in these metrics. Calibration shifts from 0.06 to 0.29, sharpness increases from 0.05 to 0.23, and CRPS rises from 0.0178 to 0.0561. Therefore, the choice of a dropout rate of ρ=0.1 in the experiments represents a balanced trade-off between overfitting and high model uncertainty.

6.3 Limitations

Despite its contributions, this research faces several limitations. First, implementing real-time processing on drones is challenging due to the computational cost of CNNs and MC dropout. Second, the dataset may not fully capture the diversity of operating scenarios, potentially excluding irregular or extreme environmental conditions. It also exhibits a temporal imbalance, with most data concentrated in a single year, which may hinder generalization. Third, the ground-truth health labels rely on expert assessments and may introduce subjective bias. Fourth, all cells originate from a single battery manufacturer, limiting the applicability of the findings to other chemistries or designs. Finally, the test set is relatively small and cross-validation was not performed, reducing statistical confidence.

7 Conclusion

This research introduced a physics-aware drone battery health assessment approach combining physics-based models with probabilistic data-driven error prediction. The methodology demonstrated robust uncertainty performance in predicting battery discharge voltage while effectively quantifying data and model related uncertainties

Empirical evaluation on EOD voltage data from real offshore wind turbine inspection missions showed that the probabilistic CNN achieved 14.8% better accuracy than QLR, with 37.8% and 32.3% improvements over QRF and QGB respectively. The framework maintained high prediction performance even when applying a Li-Ion physics-based model to Li-Po batteries, suggesting potential for cross-chemistry transfer, pending validation on other battery chemistries.

Future work will focus on computational efficiency using techniques such as quantization (Cheng et al. 2018), tensor factorization (Kim et al. 2016), and low-rank approximation (Tai et al. 2016). The integration of PINNs will also be explored to incorporate physical constraints directly into the learning process, and benchmarked against the current physics-aware approach. Alternatively, expanding the dataset to include diverse operational scenarios, enriched with environmental sensor data and controlled charge/discharge cycles, will improve model robustness.

Moreover, prediction accuracy could potentially be improved by applying calibration techniques, such as isotonic regression (Kuleshov, Fenner, and Ermon 2018), and by comparing performance with models like Bayesian Neural Networks to combine the strengths of different UQ methods (Alcibar et al. 2024). Finally, while the framework provides decision support, it does not implement a complete maintenance strategy or formal operational decision-making process.

References

Abdar, Moloud, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, et al. 2021. "A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges." *Information Fusion* 76 (December): 243–97. https://doi.org/10.1016/j.inffus.2021.05.008.

Aizpurua, Jose Ignacio, Ibai Ramirez, Iker Lasa, Luis del Rio, Alvaro Ortiz, and Brian G. Stewart. 2023. "Hybrid Transformer Prognostics Framework for Enhanced Probabilistic Predictions in Renewable Energy Applications." *IEEE Trans. Power Delivery* 38 (1): 599–609. https://doi.org/10.1109/TPWRD.2022.3203873.

Alcibar, Jokin, Jose I. Aizpurua, and Ekhi Zugasti. 2024. "Towards a Probabilistic Fusion Approach for Robust Battery Prognostics." In *PHM Society European Conference*, 8:13. https://doi.org/10.36001/phme.2024.v8i1.3981.

Arias Chao, Manuel, Chetan Kulkarni, Kai Goebel, and Olga Fink. 2022. "Fusing Physics-Based and Deep Learning Models for Prognostics." *Reliability Engineering & System Safety* 217 (January): 107961. https://doi.org/10.1016/j.ress.2021.107961.

Bai, Guangxing, Yunsheng Su, Maliha Maisha Rahman, and Zequn Wang. 2023. "Prognostics of Lithium-Ion Batteries Using Knowledge-Constrained Machine Learning and Kalman Filtering." *Reliability Engineering & System Safety* 231 (March): 108944. https://doi.org/10.1016/j.ress.2022.108944.

Biggio, Luca, Tommaso Bendinelli, Chetan Kulkarni, and Olga Fink. 2023. "Ageing-Aware Battery Discharge Prediction with Deep Learning." *Applied Energy* 346 (September): 121229. https://doi.org/10.1016/j.apenergy.2023.121229.

Bracale, Antonio, Pasquale De Falco, Luigi Pio Di Noia, and Renato Rizzo. 2023. "Probabilistic State of Health and Remaining Useful Life Prediction for Li-Ion Batteries." *IEEE Transactions on Industry Applications* 59 (1): 578–90. https://doi.org/10.1109/TIA.2022.3210081.

Che, Yunhong, Yusheng Zheng, Florent Evariste Forest, Xin Sui, Xiaosong Hu, and Remus Teodorescu. 2024. "Predictive Health Assessment for Lithium-Ion Batteries with Probabilistic Degradation Prediction and Accelerating Aging Detection." *Reliability Engineering & System Safety* 241 (January): 109603. https://doi.org/10.1016/j.ress.2023.109603.

Cheng, Jian, Jiaxiang Wu, Cong Leng, Yuhang Wang, and Qinghao Hu. 2018. "Quantized CNN: A Unified Approach to Accelerate and Compress Convolutional Networks." *IEEE Transactions on Neural Networks and Learning Systems* 29 (10): 4730–43. https://doi.org/10.1109/TNNLS.2017.2774288.

Choubineh, Abouzar, Jie Chen, Frans Coenen, and Fei Ma. 2023. "Applying Monte Carlo Dropout to Quantify the Uncertainty of Skip Connection-Based Convolutional Neural Networks Optimized by Big Data." *Electronics* 12 (6): 1453. https://doi.org/10.3390/electronics12061453.

Chung, Youngseog, Willie Neiswanger, Ian Char, and Jeff Schneider. 2021. "Beyond Pinball Loss: Quantile Methods for Calibrated Uncertainty Quantification." *Advances in Neural Information Processing Systems* 34: 10971–84.

Couture, Jonathan, and Xianke Lin. 2022. "Image- and Health Indicator-Based Transfer Learning Hybridization for Battery RUL Prediction." *Engineering Applications of Artificial Intelligence* 114: 105120. https://doi.org/10.1016/j.engappai.2022.105120.

Daigle, Matthew, and Chetan S Kulkarni. 2013. "Electrochemistry-Based Battery Modeling for Prognostics." In *Annual Conference of the PHM Society*, 5:13.

Demirci, Osman, Sezai Taskin, Erik Schaltz, and Burcu Acar Demirci. 2024. “Review of Battery State Estimation Methods for Electric Vehicles - Part I: SOC Estimation.” *Journal of Energy Storage* 87 (May): 111435. https://doi.org/10.1016/j.est.2024.111435.

Kiureghian, A.D. and Ditlevsen, O. (2009) ‘Aleatory or epistemic? Does it matter?’, *Structural Safety*, 31(2), pp. 105–112. https://doi.org/10.1016/j.strusafe.2008.06.020.

Djuric, P. M., J. H. Kotecha, Jianqui Zhang, Yufei Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez. 2003. “Particle Filtering.” *IEEE Signal Processing Magazine* 20 (5): 19–38. https://doi.org/10.1109/MSP.2003.1236770.

Eleftheroglou, Nick, Sina Sharif Mansouri, Theodoros Loutas, Petros Karvelis, George Georgoulas, George Nikolakopoulos, and Dimitrios Zarouchas. 2019. “Intelligent Data-Driven Prognostic Methodologies for the Real-Time Remaining Useful Life Until the End-of-Discharge Estimation of the Lithium-Polymer Batteries of Unmanned Aerial Vehicles with Uncertainty Quantification.” *Applied Energy* 254 (November): 113677. https://doi.org/10.1016/j.apenergy.2019.113677.

Feng, Ke, Yadong Xu, Yulin Wang, Sheng Li, Qiubo Jiang, Beibei Sun, Jinde Zheng, and Qing Ni. 2023. “Digital Twin Enabled Domain Adversarial Graph Networks for Bearing Fault Diagnosis.” *IEEE Transactions on Industrial Cyber-Physical Systems* 1: 113–22. https://doi.org/10.1109/TICPS.2023.3298879.

Fernández, Juan, Juan Chiachío, Manuel Chiachío, José Barros, and Matteo Corbetta. 2023. “Physics-Guided Bayesian Neural Networks by ABC-SS: Application to Reinforced Concrete Columns.” *Engineering Applications of Artificial Intelligence* 119: 105790. https://doi.org/10.1016/j.engappai.2022.105790.

Fink, Olga, Qin Wang, Markus Svensén, Pierre Dersin, Wan-Jui Lee, and Melanie Ducoffe. 2020. “Potential, Challenges and Future Directions for Deep Learning in Prognostics and Health Management Applications.” *Engineering Applications of Artificial Intelligence* 92: 103678. https://doi.org/10.1016/j.engappai.2020.103678.

Friedman, Jerome H. 2001. “Greedy Function Approximation: A Gradient Boosting Machine.” *Annals of Statistics*, 1189–1232.

Gal, Yarin, and Zoubin Ghahramani. 2016. “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning.” In *International Conference on Machine Learning*, 1050–59. PMLR.

Gatti, Mauro, Fabrizio Giulietti, and Matteo Turci. 2015. "Maximum Endurance for Battery-Powered Rotary-Wing Aircraft." *Aerospace Science and Technology* 45: 174–79. https://doi.org/10.1016/j.ast.2015.05.009.

Gneiting, Tilmann, Fadoua Balabdaoui, and Adrian E. Raftery. 2007. "Probabilistic Forecasts, Calibration and Sharpness." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 69 (2): 243–68. https://doi.org/10.1111/j.1467-9868.2007.00587.x.

González-Sopeña, J. M., V. Pakrashi, and B. Ghosh. 2021. "An Overview of Performance Evaluation Metrics for Short-Term Statistical Wind Power Forecasting." *Renewable and Sustainable Energy Reviews* 138: 110515. https://doi.org/10.1016/j.rser.2020.110515.

Guo, Jian, Zhaojun Li, and Michael Pecht. 2015. "A Bayesian Approach for Li-Ion Battery Capacity Fade Modeling and Cycles to Failure Prognostics." *Journal of Power Sources* 281 (May): 173–84. https://doi.org/10.1016/j.jpowsour.2015.01.164.

Guo, Li, Hongwei He, Yiran Ren, Runze Li, Bin Jiang, and Jianye Gong. 2024. "Prognostics of Lithium-Ion Batteries Health State Based on Adaptive Mode Decomposition and Long Short-Term Memory Neural Network." *Engineering Applications of Artificial Intelligence* 127: 107317. https://doi.org/10.1016/j.engappai.2023.107317.

Han, Te, Zhe Wang, and Huixing Meng. 2022. "End-to-End Capacity Estimation of Lithium-ion Batteries with an Enhanced Long Short-Term Memory Network Considering Domain Adaptation." *Journal of Power Sources* 520: 230823. https://doi.org/10.1016/j.jpowsour.2021.230823.

Heo, Jay, Hae Beom Lee, Saehoon Kim, Juho Lee, Kwang Joon Kim, Eunho Yang, and Sung Ju Hwang. 2018. "Uncertainty-Aware Attention for Reliable Interpretation and Prediction." In *Advances in Neural Information Processing Systems*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, 31:1–10. Curran Associates, Inc.

Hong, Joonki, Dongheon Lee, Eui-Rim Jeong, and Yung Yi. 2020. "Towards the Swift Prediction of the Remaining Useful Life of Lithium-Ion Batteries with End-to-End Deep Learning." *Applied Energy* 278: 115646. https://doi.org/10.1016/j.apenergy.2020.115646.

Hsu, Chia-Wei, Rui Xiong, Nan-Yow Chen, Ju Li, and Nien-Ti Tsou. 2022. "Deep Neural Network Battery Life and Voltage Prediction by Using Data of One

Cycle Only." *Applied Energy* 306: 118134. https://doi.org/10.1016/j.apenergy.2021.118134.

Hüllermeier, Eyke, and Willem Waegeman. 2021. "Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction to Concepts and Methods." *Machine Learning* 110 (3): 457–506. https://doi.org/10.1007/s10994-021-05946-3.

Jiang, Yan, and Xin Meng. 2023. "A Battery Capacity Estimation Method Based on the Equivalent Circuit Model and Quantile Regression Using Vehicle Real-World Operation Data." *Energy* 284: 129126. https://doi.org/10.1016/j.energy.2023.129126.

Jung, Yongsu, Hwisang Jo, Jeonghwan Choo, and Ikjin Lee. 2022. "Statistical Model Calibration and Design Optimization Under Aleatory and Epistemic Uncertainty." *Reliability Engineering & System Safety* 222 (June): 108428. https://doi.org/10.1016/j.ress.2022.108428.

Kendall, Alex, and Yarin Gal. 2017. "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?" In *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, 30:11. Curran Associates, Inc.

Kim, Donghyun, Chanyoung Park, Jinoh Oh, Sungyoung Lee, and Hwanjo Yu. 2016. "Convolutional Matrix Factorization for Document Context-Aware Recommendation." In *Proceedings of the 10th ACM Conference on Recommender Systems*, 233–40. RecSys '16. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2959100.2959165.

Kuleshov, Volodymyr, Nathan Fenner, and Stefano Ermon. 2018. "Accurate Uncertainties for Deep Learning Using Calibrated Regression." In *Proceedings of the 35th International Conference on Machine Learning*, edited by Jennifer Dy and Andreas Krause, 80:2796–2804. Proceedings of Machine Learning Research. PMLR.

Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell. 2017. "Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles." In *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, 30:12. Curran Associates, Inc.

Li, Wei, Martin Z. Bazant, and Juner Zhu. 2021. "A Physics-Guided Neural Network Framework for Elastic Plates: Comparison of Governing Equations-Based and

Energy-Based Approaches." *Computer Methods in Applied Mechanics and Engineering* 383: 113933. https://doi.org/10.1016/j.cma.2021.113933.

Mazzi, Yahia, Hicham Ben Sassi, and Fatima Errahimi. 2024. "Lithium-Ion Battery State of Health Estimation Using a Hybrid Model Based on a Convolutional Neural Network and Bidirectional Gated Recurrent Unit." *Engineering Applications of Artificial Intelligence* 127: 107199. https://doi.org/10.1016/j.engappai.2023.107199.

Meinshausen, N. (2006) 'Quantile Regression Forests', *Journal of Machine Learning Research*, 7(35), pp. 983–999.

Meng, Huixing, Mengyao Geng, and Te Han. 2023. "Long Short-Term Memory Network with Bayesian Optimization for Health Prognostics of Lithium-Ion Batteries Based on Partial Incremental Capacity Analysis." *Reliability Engineering & System Safety* 236 (August): 109288. https://doi.org/10.1016/j.ress.2023.109288.

Mitici, Mihaela, Ingeborg De Pater, Anne Barros, and Zhiguo Zeng. 2023. "Dynamic Predictive Maintenance for Multiple Components Using Data-Driven Probabilistic RUL Prognostics: The Case of Turbofan Engines." *Reliability Engineering & System Safety* 234 (June): 109199. https://doi.org/10.1016/j.ress.2023.109199.

Nascimento, Renato G., Felipe A. C. Viana, Matteo Corbetta, and Chetan S. Kulkarni. 2023. "A Framework for Li-ion Battery Prognosis Based on Hybrid Bayesian Physics-Informed Neural Networks." *Scientific Reports* 13 (1): 13856. https://doi.org/10.1038/s41598-023-33018-0.

Ng, Man-Fai, Jin Zhao, Qingyu Yan, Gareth J. Conduit, and Zhi Wei Seh. 2020. "Predicting the State of Charge and Health of Batteries Using Data-Driven Machine Learning." *Nature Machine Intelligence* 2 (3): 161–70. https://doi.org/10.1038/s42256-020-0156-7.

Ni, Qing, J. C. Ji, Benjamin Halkon, Ke Feng, and Asoke K. Nandi. 2023. "Physics-Informed Residual Network (PIResNet) for Rolling Element Bearing Fault Diagnostics." *Mechanical Systems and Signal Processing* 200: 110544. https://doi.org/10.1016/j.ymssp.2023.110544.

Ochella, Sunday, Mahmood Shafiee, and Fateme Dinmohammadi. 2022. "Artificial Intelligence in Prognostics and Health Management of Engineering Systems." *Engineering Applications of Artificial Intelligence* 108: 104552. https://doi.org/10.1016/j.engappai.2021.104552.

Pola, Daniel A., Hugo F. Navarrete, Marcos E. Orchard, Ricardo S. Rabié, Matías A. Cerda, Benjamín E. Olivares, Jorge F. Silva, Pablo A. Espinoza, and Aramis Pérez. 2015. "Particle-Filtering-Based Discharge Time Prognosis for Lithium-Ion Batteries with a Statistical Characterization of Use Profiles." *IEEE Transactions on Reliability* 64 (2): 710–20. https://doi.org/10.1109/TR.2014.2385069.

Raissi, Maziar, Alireza Yazdani, and George Em Karniadakis. 2020. "Hidden Fluid Mechanics: Learning Velocity and Pressure Fields from Flow Visualizations." *Science* 367 (6481): 1026–30. https://doi.org/10.1126/science.aaw4741.

Rathnakumar, Rahul, Yutian Pang, and Yongming Liu. 2023. "Epistemic and Aleatoric Uncertainty Quantification for Crack Detection Using a Bayesian Boundary Aware Convolutional Network." *Reliability Engineering & System Safety* 240 (December): 109547. https://doi.org/10.1016/j.ress.2023.109547.

Salinas-Camus, Mariana, Chetan Kulkarni, and Marcos Orchard. 2023. "Battery State-of-Health Aware Path Planning for a Mars Rover." *Annual Conference of the PHM Society* 15 (1). https://doi.org/10.36001/phmconf.2023.v15i1.3511.

Sierra, Gina, Marcos Orchard, Chetan Kulkarni, and Kai Goebel. 2018. "A Hybrid Battery Model for Prognostics in Small-size Electric UAVs." In *Annual Conference of the Prognostics and Health Management Society 2018*, 11.

Sierra, G., M. Orchard, K. Goebel, and C. Kulkarni. 2019. "Battery Health Management for Small-Size Rotary-Wing Electric Unmanned Aerial Vehicles: An Efficient Approach for Constrained Computing Platforms." *Reliability Engineering & System Safety* 182 (February): 166–78. https://doi.org/10.1016/j.ress.2018.04.030.

Song, Bing, Zhipeng Zhang, Yong Qin, Xiang Liu, and Hao Hu. 2022. "Quantitative Analysis of Freight Train Derailment Severity with Structured and Unstructured Data." *Reliability Engineering & System Safety* 224 (August): 108563. https://doi.org/10.1016/j.ress.2022.108563.

Sun, Bo, Junlin Pan, Zeyu Wu, Quan Xia, Zili Wang, Yi Ren, Dezhen Yang, Xing Guo, and Qiang Feng. 2023. "Adaptive Evolution Enhanced Physics-Informed Neural Networks for Time-Variant Health Prognosis of Lithium-Ion Batteries." *Journal of Power Sources* 556 (February): 232432. https://doi.org/10.1016/j.jpowsour.2022.232432.

Tagasovska, Natasa, and David Lopez-Paz. 2019. "Single-Model Uncertainties for Deep Learning." In *Advances in Neural Information Processing Systems*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, 32:12. Curran Associates, Inc.

Tai, Cheng, Tong Xiao, Yi Zhang, Xiaogang Wang, and Weinan E. 2016. "Convolutional Neural Networks with Low-Rank Regularization." https://arxiv.org/abs/1511.06067.

Teubert, Christopher, Katelyn Jarvis Griffith, Matteo Corbetta, Chetan Kulkarni, Portia Banerjee, and Matthew Daigle. 2023. "ProgPy Python Prognostics Packages." https://doi.org/10.5281/ZENODO.8097013.

Tran, Kevin, Willie Neiswanger, Junwoong Yoon, Qingyang Zhang, Eric Xing, and Zachary W Ulissi. 2020. "Methods for Comparing Uncertainty Quantifications for Material Property Predictions." *Machine Learning: Science and Technology* 1 (2): 025006. https://doi.org/10.1088/2632-2153/ab7e1a.

Valavanis, Kimon P, and George J Vachtsevanos. 2015. *Handbook of Unmanned Aerial Vehicles*. Vol. 1. Springer.

Vanem, Erik, Clara Bertinelli Salucci, Azzeddine Bakdi, and Øystein Å Sheim Alnes. 2021. "Data-Driven State of Health Modelling Review of State of the Art and Reflections on Applications for Maritime Battery Systems." *Journal of Energy Storage* 43 (November): 103158. https://doi.org/10.1016/j.est.2021.103158.

Xu, Yadong, Xiaoan Yan, Beibei Sun, and Zheng Liu. 2022. "Global Contextual Residual Convolutional Neural Networks for Motor Fault Diagnosis Under Variable-Speed Conditions." *Reliability Engineering & System Safety* 225 (September): 108618. https://doi.org/10.1016/j.ress.2022.108618.

Yeregui, J., L. Oca, I. Lopetegi, E. Garayalde, M. Aizpurua, and U. Iraola. 2023. "State of Charge Estimation Combining Physics-Based and Artificial Intelligence Models for Lithium-ion Batteries." *Journal of Energy Storage* 73: 108883. https://doi.org/10.1016/j.est.2023.108883.

Zamo, Michaël, and Philippe Naveau. 2018. "Estimation of the Continuous Ranked Probability Score with Limited Information and Applications to Ensemble Weather Forecasts." *Mathematical Geosciences* 50 (2): 209–34. https://doi.org/10.1007/s11004-017-9709-7.

Zhan, Dawei, and Huanlai Xing. 2020. "Expected Improvement for Expensive Optimization: A Review." *Journal of Global Optimization* 78 (3): 507–44.

Zhang, Shuxin, Zhitao Liu, and Hongye Su. 2022. "A Bayesian Mixture Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries." *IEEE Transactions on Transportation Electrification* 8 (4): 4708–21. https://doi.org/10.1109/TTE.2022.3161140.

Zhang, Zhizhou, and Grace X Gu. 2021. “Physics-Informed Deep Learning for Digital Materials.” *Theoretical and Applied Mechanics Letters* 11 (1): 100220. https://doi.org/10.1016/j.taml.2021.100220.

Zhao, Hongqian, Zheng Chen, Xing Shu, Jiangwei Shen, Zhenzhen Lei, and Yuanjian Zhang. 2023. “State of Health Estimation for Lithium-Ion Batteries Based on Hybrid Attention and Deep Learning.” *Reliability Engineering & System Safety* 232: 109066. https://doi.org/10.1016/j.ress.2022.109066.

GLOSARIO:

End-of-discharge (EOD)	Voltaje de fin de descarga
Particle Filtering (PF)	Filtrado de Partículas
Unscented Kalman Filter (UKF)	Filtro de Kalman Sin Aroma
Reduced Order Model (ROM)	Modelo de orden reducido
Long Short-Term Memory (LSTM)	Termino de Memoria a Medio (largo-corto) Plazo (LSTM)