Expert Systems With Applications 115 (2019) 709–723

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

An integrated imputation-prediction scheme for prognostics of battery data with missing observations

Roozbeh Razavi-Far[a,*], Shiladitya Chakrabarti[a], Mehrdad Saif[a], Enrico Zio[b,c]

[a] Department of Electrical and Computer Engineering, University of Windsor, Windsor, Ontario N9B 3P4, Canada
[b] Ecole CentraleSupélec, Université Paris Saclay, France
[c] Department of Energy, Politecnico di Milano, Milano, Italy

ARTICLE INFO

Article history:
Received 10 June 2017
Revised 15 August 2018
Accepted 16 August 2018
Available online 22 August 2018

Keywords:
Lithium-ion batteries
Prognostics and health management
Remaining useful life
Incomplete scenarios
Missing data imputation
Extreme learning machines

ABSTRACT

This paper focuses on the development of a prognostic scheme for estimating the remaining useful life (RUL) of Lithium-ion batteries with missing observations. The scheme has two main modules based on extreme learning machines: pre-processing and prediction. The pre-processing module uses novel single and multiple imputation techniques to estimate the missing observations. The prediction module aims to obtain precise predictions even in the presence of missing observations and with the related uncertainty. The pre-processing module sends imputed subsets of samples to the prediction module, which makes use of extreme learning machines for one-step and multi-steps predictions. The prediction module contains various multi-steps prediction strategies including iterative, direct and DirRec, which use the constant-current experimental capacity data for the long-term prediction of the remaining useful life. Accurate prediction of RUL requires continuity in the time-series dataset. The proposed scheme is designed to build an intelligent prediction system with the ability to handle time-series data containing missing values and is robust enough to generate a complete time-series dataset and, then, make short or long term predictions. The experimental results confirm that the proposed framework can be beneficial for intelligent diagnostic and prognostic systems related to battery as well as other wide range of applications. The main focus of the paper is the development of an integrated imputation-prediction scheme and not the evaluation of individual performances of the imputation or prediction techniques.

1. Introduction

The lithium-ion batteries are a vital element for many machines, computers and mobile devices (Goebel, Saha, Saxena, Celaya, & Christophersen, 2008). The growing popularity of these batteries can be attributed to their long lifetime, high energy density, high efficiency, high charging and discharging rates, light weight, high galvanic potential, wide temperature range, lower self-discharge rate and no memory effect (Charkhgard & Farrokhi, 2010; Razavi-Far, Chakrabarti, & Saif, 2016a). Failures caused by fast degradation and depletion can lead to performance reduction, unpredicted halting problem and even adverse consequences. Thus, the prognostics and health management (PHM) of lithium-ion batteries has been become of primary importance in the electronics industry (Goebel et al., 2008; Widodo, Shim, Caesarendra, & Yang, 2011; Zhang & Lee, 2011).

Contemporary PHM systems usually combine measurements, models, observations in computer programs for early detecting faults, identifying system states and predicting fault progression. A typical PHM framework is generally composed of various modules containing data processing, condition monitoring and health estimation, prediction of the remaining useful life (RUL), condition-based maintenance (CBM) and intelligent decision-making (Vachtsevanos, Lewis, Roemer, Hess, & Wu, 2006).

Prognostic is the prediction of the system lifetime, based on existing conditions and previous operation profiles. If accurate, it can result in appropriate health management through a set of protective and maintenance actions (Zio, 2013). The prediction of the remaining useful life of a system can be considered as the decisive step in the PHM process (Bai & Wang, 2016).

Quick and accurate estimation of the remaining useful life can assist reducing the failure rate of the batteries (Vachtsevanos et al., 2006). The RUL information is valuable for estimating the feasible time for battery replacement. The RUL of a battery is a func-

* Corresponding author.
E-mail addresses: roozbeh@uwindsor.ca (R. Razavi-Far), chakrab@uwindsor.ca (S. Chakrabarti), msaif@uwindsor.ca (M. Saif), enrico.zio@ecp.fr, enrico.zio@supelec.fr, enrico.zio@polimi.it (E. Zio).

https://doi.org/10.1016/j.eswa.2018.08.033
0957-4174/

tion of the battery degradation level and operational conditions. The nonlinear and uncertain degradation profiles of batteries reveal the need for accurate and nonlinear predictive models. Monitoring a battery reliability during its operational usage is mainly based on two indicators, namely state of charge (SOC) and state of health (SOH) (Liu, Pang, Zhou, Peng, & Pecht, 2013; Widodo et al., 2011). SOC is a measure of remaining available stored energy and SOH is a measure of battery degradation. SOC estimation techniques have been widely used in the electronics industry. In recent years, there has been an increasing interest on SOH estimation (Chao & Chen, 2011) and RUL prediction to prevent possible accidents, as a result of the growing demand for lithium-ion batteries (Acuna, Orchard, Silva, & Perez, 2015). The RUL prediction task can be performed through a wide range of prognostic methods, that are mainly classified under two major categories of model-based and data-driven techniques (Liao & Kottig, 2014; Saha, Goebel, Poll, & Christophersen, 2009; Xu, Li, & Chen, 2016; Zhao, Quan, & Cai, 2014).

Model-based approaches account for a majority of the modern RUL estimation methods; however, in many cases it is difficult to achieve a precise and well-established model (Liao & Kottig, 2014; Xu et al., 2016; Zaidan, Relan, Mills, & Harrison, 2015). All parameters of the prediction models need to be initialized and tuned. Moreover, these parameters cannot be easily updated during the prediction phase for emerging operational conditions (Razavi-Far et al., 2016a). Besides, rapid changes at the final steps of the lifecycle increase the difficulty to learn the final steps of the degradation profiles (Sbarufatti, Corbetta, Giglio, & Cadini, 2017). On the other hand, data-driven approaches can help to reduce these issues for the RUL prediction (Barre, Suard, Gerard, Montaru, & Riu, 2014).

Data-driven approaches utilize existing and past observations of the batteriesâ;; degradation profiles for the prediction of future values (Lee, Cui, Rezvanizaniani, & Ni, 2012; Razavi-Far, Farajzadeh-Zanjani, Chakrabarti, & Saif, 2016b). Informative features, like voltage, capacity, current and impedance can be extracted and used for the RUL prediction. RUL can be predicted by learning the relationships among these features and following the degradation trends (Baraldi, Maio, Al-Dahidi, Zio, & Mangili, 2017; Razavi-Far et al., 2016a). Capacity is a commonly used health indicator for Li-ion batteries as they are easier to extract and provide a simple mechanism to predict the end-of-life (EOL) of the batteries. Evaluating the knee points of the voltage in the discharge cycles is another mechanism used for EOL prediction, but they require more complex computational algorithm. The voltage and current are commonly used as health indicator for single-use or non-rechargeable batteries. The observations and informative features play a vital role for achieving an accurate prediction model (Javed, Gouriveau, Zerhouni, & Hissel, 2016). Moreover, the data-driven approaches are faster and easier to implement than model-based approaches (Saxena et al., 2012), even though they may require a large number of observations in the training phase and the attained predictive model is non-transparent.

This work mainly focuses on data-driven approaches to predict the remaining useful life of batteries under incomplete data scenarios. This allows evaluating the sensitivity of the data-driven prognostic techniques w.r.t. the quantity and quality of the battery data with missing observations.

Contemporary intelligent data-driven prediction techniques are not equipped to handle missing data. Most of the intelligent and expert systems have existing data collected over time. Many batch of data are considered incomplete due to the presence of missing features. The inability of prediction systems to process such incomplete data renders the batch useless, thus losing valuable knowledge present in the complete features. The proposed scheme aims to process such incomplete batches of observations, impute them to form complete datasets and utilize the entire knowledge in the prediction module. The proposed framework can be beneficial for intelligent diagnostic and prognostic systems, recommenders, decision making and expert systems related to battery prognostics over short and long horizons as well as other wide range of applications.

Contemporary prognostic systems should have the ability to perform short-term estimates, long-term estimates and predictions with missing observations. One-step-ahead prediction (OSP) techniques are used to predict short-term batteriesâ;; conditions (Razavi-Far et al., 2016a). Multi-steps-ahead prediction (MSP) techniques are used to predict long-term batteriesâ;; conditions (Lee et al., 2012).

Prediction of the RUL heavily depends on the presence of real-time observations. Any loss of information due to transmission or sensor errors can have a negative impact (Acuna et al., 2015). Missing observations can be avoided prior to prediction by simply discarding incomplete observations, which results in loss of information and possibly unrealistic estimations and predictions (Razavi-Far, Zio, & Palade, 2014). Missing observations can be replaced with the mean and median of the target, but this can bias the estimations and predictions. It is necessary, therefore, to use single or multiple imputation techniques for improving the accuracy of the estimations.

This paper focuses on the development of an integrated prognostics framework for one-step-ahead and multi-steps-ahead predictions of Li-ion batteries RUL, under incomplete data scenarios. The proposed prognostics scheme contains two major modules: pre-processing and prediction (see Fig. 1). The pre-processing module contains two novel techniques, which make use of extreme learning machines for single imputation (ELMSI) and multiple imputation (ELMMI) of the missing observations. It also normalizes the input data a priori, in order to improve the prediction accuracy. The effectiveness of the proposed imputation techniques are studied and compared to other state-of-the-art single and multiple imputation techniques (hereafter called competitors) in providing complete sets of observations to the prediction module.

This paper also aims to develop several predictors based on extreme learning machines (ELMs), for estimation of the RUL of the batteries. These ELM-based predictors are studied and compared to other state-of-the-art data-driven prediction techniques (hereafter called competitors), including neuro-fuzzy networks (NFs) (Jang, 1993), group method of data handling (GMDH), which is a combinatorial multi-layer learning scheme (Ivakhnenko, 1978), and random forests (RF). RF is an ensemble of decision trees, where each individual tree is trained on a random subspace by means of bootstrapping (Breiman, 2001). Here, three Neuro-Fuzzy (NF) predictors, so-called NF-GP, NF-FCM and NF-SC, are constructed by resorting to different rule selection strategies, i.e., grid partitioning, fuzzy-c-means, and subtractive clustering (Takagi & Hayashi, 1991). These predictors are formed in the nonlinear autoregressive (NAR) structures for one-step-ahead prediction and used along with several MSP structures, namely iterative, direct and DirRec, for the long-term prediction. All these strategies make use of current and past observations of the batteries and train several models to predict the future battery capacity values over short and long horizons.

The novelty of the proposed scheme is that it uses the same algorithm in both the imputation and prediction modules. Handling missing data in time-series prediction have been proposed before using AR models (Anava, Hazan, & Zeevi, 2015; Wei, Wang, Shi, Gao, & Li, 2017) and nonlinear filters (Wu, Wang, Mao, Dua, & Lia, 2014). However, such methods do not impute the missing data. This makes the system computationally expensive as the missing features have to be handled for every prediction process. The strength of the proposed scheme is that it imputes the missing features in the pre-processing module and this complete data

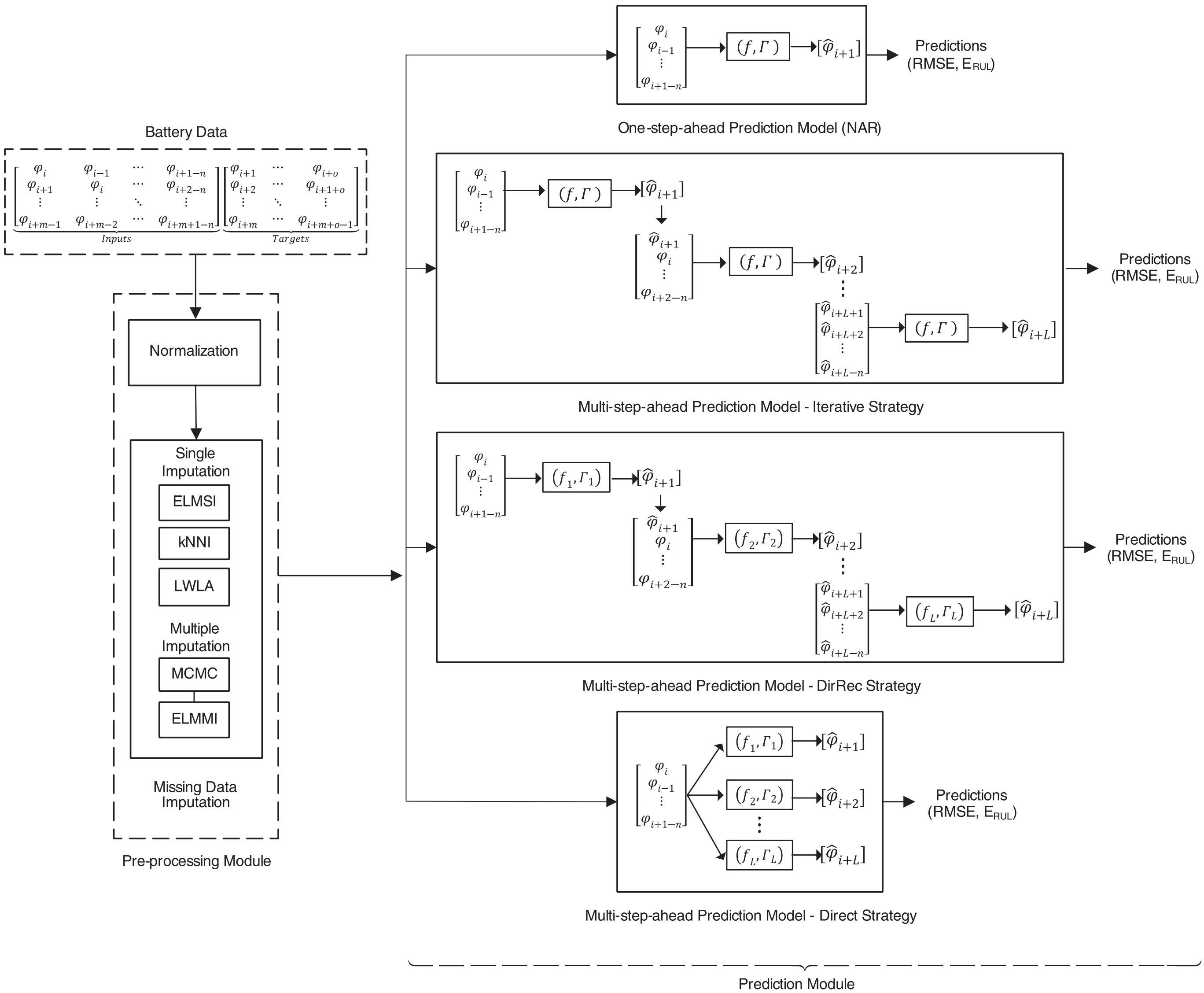

Fig. 1. General diagram of the prognostic system. Battery data goes through the pre-processing module and, then, feed into the prediction module, which contains various prediction strategies.

can then be utilized by multiple prediction techniques. Imputation of the data prevents the need of data manipulation for future predictions, thus making the process fast and efficient. The imputed data can be saved as part of the knowledge system and utilized later for training. There are few existing approaches like Regularized extreme learning machine for regression with missing data (Yu et al., 2013) and Extreme learning machine for missing data using multiple imputation (Sovilj et al., 2016) that uses ELM for handling missing data. Even though these approaches have their own merit in handling missing data, they were found to be either too complex, computationally extensive or fail to impute the missing data completely. Hence, new techniques were developed for the pre-processing module in the proposed prognostic framework, where missing observations are replaced with imputed ones. The mechanisms are customized for time-series datasets, low in complexity and provides integration with the prediction module. The proposed scheme, is however, not without some drawbacks. A large sized time-series dataset often contains significant change in regression model between different intervals of time cycles. Using data from different intervals can often lead to over-fitting of data and skewed results. Further studies in tuning the data horizon used for training as well as optimization of the hidden neurons can prevent over-fitting of the data for both imputation and prediction modules.

The novelty of the proposed prognostic framework lies in (1) the development of ELM-based techniques for imputing missing observations, i.e., ELMSI and ELMMI, (2) the use of various ELM networks assembled in different structures, for prediction over short-term and long-term horizons, (3) the integration of the ELMSI and ELMMI with ELM-based predictors for prognostics with missing observations, and (4) the use of ELMMI in the pre-processing module for multiple imputation of missing observations to estimate the uncertainty and confidence of estimations.

The paper does not claim that the proposed imputation techniques are significantly superior to the other methods used for comparison. The performance evaluation of the individual imputation or the prediction techniques in comparison to other state-of-the-art techniques are beyond the scope of this paper and would require a larger number of datasets for comparison. Instead, the paper wishes to highlight on the simple and effective proposed prognostic scheme formed using an integrated ELM-based imputation-prediction framework.

The remainder of the paper is structured as follows. Section II of the paper briefly presents the case study and the problem statement. Section III presents the techniques proposed to impute the missing values and predict the remaining useful life of the Lithium-ion batteries. Section IV discusses and compares the experimental results; this comparison considers various state-of-the-art data-driven imputation and prediction techniques, to predict the RUL of batteries. Section V presents the conclusions.

2. Problem statement

The primary goal of the paper is to analyze the degradation in Li-ion batteries and predict their RUL over both short and long horizons, under incomplete data scenarios. Three different Li-ion batteries datasets, namely B0005, B0006 and B0007, are used for the estimation of the RUL in terms of remaining battery capacity. These datasets were collected at the Idaho National Lab and can be obtained from the NASA Prognostic Center of Excellence Data Repository (Saha & Goebel, 2007). The data observations were collected from various operating conditions (at a constant temperature), including charge, discharge and impedance, to represent the rapid aging of Li-ion batteries over continuous charge and discharge cycles. The batteries have been charged at the constant rate of 1.5*A*, till the voltage reaches 4.2*V* and, then, the voltage is sustained at a constant level, till the current reduces to 20mA. In discharge cycles, the batteries maintain a constant current level till the voltage drops below a pre-set value (Zhao et al., 2014). Batteries B0005, B0006 and B0007 are considered to be fully discharged when their voltage drops to 2.7V, 2.5V and 2.2V, respectively. This charge and discharge cycles are continued until the rated capacity drops from 2Ahr to 1.4Ahr, a drop of 30%, which is a criteria for the battery end-of-life (Saxena et al., 2012). Capacity of a battery, *C (Ahr)*, is a common index to predict its remaining useful life (Lee et al., 2012) and, hence, in this paper, capacities from 168 discharge cycles are extracted and utilized to create the required inputs for the predictors, in order to estimate the RUL of the batteries.

The following prediction strategies can be considered to construct various predictors and estimate the useful life of the Li-ion batteries for short and long-term horizons:

2.1. One-step-ahead prediction (OSP) strategy

One-step-ahead prediction (OSP) is used to estimate the RUL of a battery for a short horizon. The method uses the nonlinear autoregressive (NAR) structure (Bai & Wang, 2016) to construct predictive models:

$$\hat{\varphi}_{i+1} = f(\varphi_i, \varphi_{i-1}, \varphi_{i-2}, \ldots, \varphi_{i+1-n}, \Gamma) + \mathcal{E}_{i+1} \tag{1}$$

where φ denotes the battery capacity at discharge cycles, i denotes the cycle number, $\mathcal{E}$ denotes the prediction error for each cycle, $n-1$ denotes the number of lags, f denotes the approximation function generated by the predictor during the training phase and Γ stands for a set of parameters of the estimated function.

2.2. Multi-step-ahead prediction (MSP) strategies

Multi-step-ahead prediction (MSP) strategies are used to estimate the RUL of batteries for a long horizon. These prediction strategies can be classified under three main categories: iterative, DirRec and direct (Sorjamaa, Hao, Reyhani, Ji, & Lendasse, 2007; Taieb, Sorjamaa, & Bontempi, 2010). These strategies can be used for MSP of battery capacity.

(*a*) *Iterative strategy* is a very common technique to perform multi-steps-ahead prediction (Sorjamaa et al., 2007). This technique utilizes a one-step-ahead prediction to predict the future value and, then, uses iteratively the predicted value as a known input to predict the subsequent L future values. The first future value can be estimated as follows:

$$\hat{\varphi}_{i+1} = f(\varphi_i, \ldots, \varphi_{i+2-n}, \varphi_{i+1-n}, \Gamma) \tag{2}$$

The next future value can be estimated using the same prediction model, by resorting to the recent predicted value $\hat{\varphi}_{i+1}$ as follows $\hat{\varphi}_{i+2} = f(\hat{\varphi}_{i+1}, \varphi_i, \ldots, \varphi_{i+3-n}, \varphi_{i+2-n}, \Gamma)$. Prediction of the L subsequent values using this technique can, then, be formulated as follows:

$$\hat{\varphi}_{i+l} = \begin{cases} f(\varphi_i, \ldots, \varphi_{i+1-n}, \Gamma) & \text{if } l = 1 \\ f(\hat{\varphi}_{i+l-1}, \ldots, \hat{\varphi}_{i+1}, \varphi_i, \ldots, \varphi_{i+l-n}, \Gamma) & \text{if } l \in \{2, \ldots, n\} \\ f(\hat{\varphi}_{i+l-1}, \ldots, \hat{\varphi}_{i+l-n}, \Gamma) & \text{if } l \in \{n+1, \ldots, L\} \end{cases} \tag{3}$$

where f denotes the one-step prediction model, $\hat{\varphi}$ denotes the predicted battery capacity and L denotes the total number of predictions. The equation shows that for $l > n$, the estimations are only based on the predicted values, which can result in low accuracy due to the propagation of the error accumulated at each step of the subsequent estimations (Taieb et al., 2010; Tran, Yang, & Tan, 2009).

(*b*) *DirRec strategy* predicts multi-steps-ahead values similarly to the iterative strategy. It, however, generates a new prediction model after each prediction step, using the recently predicted values. The older observations are discarded upon emergence of the newly predicted values in the training subset, preserving a constant number of observations in the training subset at each iteration (Tran et al., 2009). Prediction of the L subsequent values using this technique can be formulated as follows:

$$\hat{\varphi}_{i+l} = \begin{cases} f_l(\varphi_i, \ldots, \varphi_{i+1-n}, \Gamma_l) & \text{if } l = 1 \\ f_l(\hat{\varphi}_{i+l-1}, \ldots, \hat{\varphi}_{i+1}, \varphi_i, \ldots, \varphi_{i+l-n}, \Gamma_l) & \text{if } l \in \{2, \ldots, n\} \\ f_l(\hat{\varphi}_{i+l-1}, \ldots, \hat{\varphi}_{i+l-n}, \Gamma_l) & \text{if } l \in \{n+1, \ldots, L\} \end{cases} \tag{4}$$

where f_l denotes the predictive model constructed at the lth prediction step. Re-training using new values can improve the estimation accuracy, but error propagation is still an issue in this strategy (Gouriveau & Zerhouni, 2012).

(*c*) *Direct strategy* is an alternative for long-term prediction that constructs L distinct models f_l to predict L future values, as follows (Gouriveau & Zerhouni, 2012):

$$\hat{\varphi}_{i+l} = f_l(\varphi_i, \varphi_{i-1}, \ldots, \varphi_{i+1-n}, \Gamma_l) \quad l \in [1, L] \tag{5}$$

This MSP strategy delivers a higher prediction accuracy, since prediction errors are not propagated through the subsequent prediction cycles. However, training a large number of models increases the complexity of the process and its execution time (Sorjamaa et al., 2007). This strategy also induces a conditional independence of the L predictive models, while ignoring complex dependencies among the features, which may lead to biased estimates (Taieb et al., 2010). It should be noted that the direct strategy can potentially skew the prediction over a large horizon. The training model is based on the dependencies of the initial observations and any changes in features dependencies for subsequent observations are not reflected.

2.3. Incomplete scenario

This work aims to develop predictive models for OSP and MSP of Li-ion batteries capacity values in order to predict the RUL of the batteries. These OS and MS predictors require a complete sequence of observations and their prediction accuracies depend on the availability and quality of real-time observations. On the other hand, missing data is a very common problem in industrial applications, which is usually caused by equipment failures and/or data transmission errors. OS and MS predictions are very challenging tasks in the presence of missing observations, where capacity values at some discharge cycles are not measured or properly collected.

There are three major missing data mechanisms, namely Missing Completely at Random (MCAR), Missing at Random (MAR) and Missing Not at Random (MNAR) (Scheffer, 2002). MCAR describes missing data where the missingness mechanism does not depend on the feature of interest, or any other observed feature (Scheffer, 2002). MAR describes missing data where the missingness mechanism depends on the observed data. MNAR describes missing data where the missingness mechanism depends on the value of the missing feature itself (Scheffer, 2002). The first two missing mechanisms MCAR and MAR are ignorable. In other words, they can be predicted through observed features. However, the MNAR mechanism is not ignorable and the estimations and predictions are not reliable. In this work, missing observations arising from the MCAR mechanism are considered, i.e., they are due to sensor errors or changes in data collection.

There are usually two different ways to handle missing observations, imputation or discarding. The latter ignores missing observations, which is the easiest process but it may lose important dependencies and information as well. This can lead to inaccurate predictions. On the other hand, imputation is the process where any missing observation is replaced with an estimated value (Scheffer, 2002). Missing data imputation can be classified under single imputation and multiple imputation (Junior, do Carmo Nicoletti, & Zhao, 2016; Scheffer, 2002). This work develops two novel ELM-based techniques to impute missing observations. These ELM-based imputation techniques have been devised in the prognostics scheme as a part of the pre-processing module, which aims to impute missing values and feed complete sets of observations to the data-driven predictor. A novel ELM-based single imputation technique (ELMSI) has been developed and devised in the pre-processing module. Two different contemporary techniques for single imputation are also implemented along with ELMSI in the pre-processing module. These advanced state-of-the-art imputation techniques are used for comparison: kNN imputation (regress on the k-most similar features (Batista & Monard, 2002)) and LWLA imputation, which effectively integrates local structure with global distribution (Liu, Dai, & Yan, 2010).

In general, uncertainty in estimations and data can result in poor predictions (Sankararaman, Daigle, & Goebel, 2014). To represent uncertainty, multiple imputation techniques have been devised also in the prognostic scheme. Multiple imputation techniques are beneficial due to their capability to reveal uncertainty and facilitate precise predictions even with missing observations. The pre-processing module, then, contains a novel ELM-based multiple imputation (ELMMI) technique, which adds the factor of uncertainty in the predictions. A Markov Chain Monte Carlo (MCMC) technique for multiple imputation (Rubin, 2009) is also implemented, along with ELMMI in the pre-processing module, which makes use of a Bayesian network in order to learn from the data and estimate the missing observations, iteratively.

3. Proposed prognostics scheme

This paper proposes an effective prognostic scheme (see Fig. 1), which makes use of ELMs for both imputation and prediction tasks. ELM can predict accurately for both short-term and long-term horizons, train faster than other popular methods due to the fact that its hidden nodes are randomly generated and can also be used to impute the missing observations. In this work, three different variants of ELM, including classical ELM (ELM), kernelized ELM (KELM) and online sequential ELM (OS-ELM), are adopted as different prediction strategies and used in the prediction module. However, the pre-processing module merely uses the ELM algorithm for single and multiple imputation. This section aims to briefly introduce the selected state-of-the-art ELM techniques and, then, proposes the novel ELM-based single and multiple imputation techniques, respectively.

3.1. Extreme learning machines

ELM is a generalized single hidden layer feedforward network (SLFN) (Bequao & Lessmann, 2017; Huang, Zhou, Ding, & Zhang, 2012), which can provide a tradeoff between the prediction accuracy and the computational cost. The pseudo-code of ELM is presented in Algorithm 1 .

Algorithm 1: ELM (Huang et al., 2012).

INPUTS:
Set of observations $\mathbb{S} = \left\{ (x_j, y_j) | x_j \in \mathbb{R}^n, y_j \in \mathbb{R}^o \right\}$
p stands for the # of hidden nodes
DEFINITIONS:
m is the # of observations
n is the # of features or input nodes
o is the # of output nodes
x and y are the input and target vectors
ω_i is the weight linking the ith hidden node to inputs
b_i denotes the threshold of the ith hidden node
1. Randomly generate ω_i and b_i for each hidden node
2. Compute each hidden node h_i as:
$h_i(x_j) = f(\omega_i x_j + b_i), \quad \omega_i \in \mathbb{R}^p, \quad b_i \in \mathbb{R}^p$
where f stands for the sigmoid activation function:

$$f(\omega_i x_j + b_i) = \frac{1}{1 + exp(-(\omega_i x_j + b_i))} \qquad (6)$$

3. Form the randomized hidden layer matrix H:

$$H = \begin{vmatrix} h(x_1) \\ \vdots \\ h(x_m) \end{vmatrix} = \begin{vmatrix} f(\omega_1 x_1 + b_1) & \dots & f(\omega_p x_1 + b_p) \\ \vdots & \ddots & \vdots \\ f(\omega_1 x_m + b_1) & \dots & f(\omega_p x_m + b_p) \end{vmatrix}_{m \times p} \qquad (7)$$

4. Find the optimal output weights $\beta = [\beta_1, \dots, \beta_p]^T_{p \times o}$:

$$\beta = H^{\dagger} Y \qquad (8)$$

where Y stands for the target matrix:

$$Y = \begin{vmatrix} y_1^T \\ \vdots \\ y_m^T \end{vmatrix} = \begin{vmatrix} y_{11} & \dots & y_{1o} \\ \vdots & \ddots & \vdots \\ y_{m1} & \dots & y_{mo} \end{vmatrix}_{m \times o}, \qquad (9)$$

$H^{\dagger}$ is the Moore–Penrose generalized inverse of H:

$$H^{\dagger} = \begin{cases} (H^T H)^{-1} H^T & \text{if } H^T H \text{ is non-singular} \\ H^T (H H^T)^{-1} & \text{if } H H^T \text{ is non-singular} \end{cases} \qquad (10)$$

5. Compute the final output for $x_j, \; j = 1, \dots, m$:
$f_{ELM} = H\beta = \sum_{i=1}^{p} \beta_i h_i(x_j)$

Given a set $\mathbb{S}$ of m observations, ELM randomly assigns the input weights and hidden layer biases (step 1), in order to decrease computational cost while preserving overall prediction accuracy. ELM analytically tunes the output weights to attain the following output function:

$$f_{ELM} = \sum_{i=1}^{p} \beta_i h_i(x_j) = \sum_{i=1}^{p} \beta_i f(\omega_i x_j + b_i) \qquad (11)$$

where β stands for the output weights, h_i stands for the nonlinear feature mapping, ω_i stands for the weight linking the ith hidden node to the input nodes, and b_i is the threshold of the ith hidden node. During the training session, ELM minimizes the prediction errors $||H\beta - Y||^2$ and the norm of the output weights $||\beta||$ as follows:

$$\begin{cases} \text{Minimize:} & L_{ELM} = \frac{1}{2}||\beta||^2 + \frac{\varsigma}{2}\sum_{j=1}^{m}||\varepsilon_j||^2 \\ \text{Subject to:} & h(x_j)\beta = y_j^T + \varepsilon_j^T, \quad j = 1, \dots, m \end{cases} \tag{12}$$

where $\varepsilon_j = \left\{\epsilon_{j1}, \dots, \epsilon_{jo}\right\}^T = \sum_{i=1}^{p}\left\|f_{ELM}(x_j) - y_j\right\|$ denotes the error vector for the o output nodes w.r.t. x_j and ς stands for the regularization parameter, which aims to balance the above optimization terms. To solve this optimization problem, one can find the minimal norm least-square solution of the following linear system:

$$H\beta = Y \tag{13}$$

The attained least-square solution $\beta = H^{\dagger}Y$ (step 4) yields the good generalization capability, minimized error and fast convergence. $H^{\dagger}$ is obtained a priori, by resorting to the orthogonal projection technique (see Eq. 10) (Bequao & Lessmann, 2017; Huang et al., 2012).

3.2. Kernelized extreme learning machine (KELM)

KELM uses an ELM kernel matrix, which can be described by means of Mercer condition (Huang, 2014) as stated below:

$$\Theta_{KELM} = HH^T : \quad K_{ELM}(x_i, x_j) = h(x_i).h(x_j) \tag{14}$$

The output of KELM can be reformulated as follows:

$$f_{KELM}(x) = \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_m) \end{bmatrix}^T \left(\frac{I}{\varsigma} + \Theta_{KELM}\right)^{-1} Y \tag{15}$$

where I and $K(.)$ stand for an identity matrix and kernel function, respectively, and

$$\Theta_{KELM} = HH^T = \begin{bmatrix} K(x_1, x_1) & \dots & K(x_1, x_m) \\ \vdots & \ddots & \vdots \\ K(x_m, x_1) & \dots & K(x_m, x_m) \end{bmatrix}. \tag{16}$$

Variations of kernel functions, which satisfy Mercer condition (Huang, 2014), can be used to train KELM. In this work, the wavelet kernel function and the radial basis function (RBF) have been used to train the Kernelized ELMs for prediction:

$$K_{WAV}(x_i, x_j) = cos(\tau\frac{||x_i - x_j||}{\upsilon})exp\left(-\frac{||x_i - x_j||^2}{\zeta}\right) \tag{17}$$

$$K_{RBF}(x_i, x_j) = exp\left(-\frac{||x_i - x_j||^2}{\xi}\right) \tag{18}$$

where $\tau, \upsilon, \zeta, \xi$ are the model parameters that are tuned during the training session.

3.3. Online sequential extreme learning machine (OS-ELM)

OS-ELM is an extension of ELM, which can be trained in a sequential manner (Liang, Huang, Saratchandran, & Sundararaja, 2006; Singh, Kumar, & Singla, 2015). This is an extremely accurate and fast technique for online learning, which receives observations batch by batch (Liang et al., 2006; Singh et al., 2015). OS-ELM is a recursive learning procedure, which contains two phases for initialization and sequential learning. In this work, OS-ELM is used as a predictor. The pseudo-code of the OS-ELM algorithm and its two main phases are not presented here for the sake of conciseness: the reader can refer to Liang et al. (2006) for a more detailed explanation.

3.4. ELM-based single imputation (ELMSI)

The proposed ELMSI is a novel single imputation technique, which initially splits the dataset into two different complete X_k and incomplete X_u subsets and, then, for any observation x_t in X_u, it makes use of all complete observations in X_k to train an ELM model, by taking incomplete features of x_t as the target features and the remaining features as inputs. The constructed ELM model can, then, be used to estimate the missing features of x_t. The pseudo-code of ELMSI is presented in Algorithm 2 .

Algorithm 2: ELM-Based Single Imputation (ELMSI).

INPUTS:
$X_{m\times n}$ is an incomplete dataset
m and n are the # of observations and features
Γ is a set of custom parameters for the ELM model
Γ includes the # of hidden nodes and activation function
DEFINITIONS:
n^{mis} is the # of missing features for an observation
n^{obs} is the # of observed features for an observation
m^{mis} is the # of incomplete observations
m^{obs} is the # of complete observations
X_k is the complete subsets of X
X_u is the incomplete subsets of X
1. SPLIT X into X_k and X_u
2. SORT X_u w.r.t. n^{mis} from smallest to largest values
3. NORMALIZE both X_k and X_u
for $\forall t \in X_u (t = 1, \dots, m^{mis})$ **do**
  4. TRAIN an ELM model (Algorithm 1) using X_k
  $[X_k^{mis}, H, \hat{\beta}] = f_{ELM}(X_k^{obs}, \Gamma)$
  where H is the hidden node matrix
  and $\hat{\beta}$ is the output weight matrix
  5. PREDICT incomplete scores of $x_t \in X_u$ by f_{ELM}
  $\hat{x}_t^{mis} = f_{ELM}(x_t^{obs}, H, \hat{\beta})$
  The completed observation is, then, $\hat{x}_t = (x_t^{obs}, \hat{x}_t^{mis})$
  6. APPEND the completed observation $\hat{x}_t$ in X_k
end

3.5. ELM-based multiple imputation (ELMMI)

The proposed ELMMI is a novel multiple imputation technique, whose pseudo-code in presented in Algorithm 3 . ELMMI initially splits the dataset into two different complete X_k and incomplete X_u subsets. It creates $\alpha \in [1, 5]$ number of imputed sets by looping through each observation x_t of X_u. Each α generates a separate fully imputed dataset of size $m \times n$. In each loop, it makes use of all complete observations in X_k to train an ELM model, using all incomplete features in x_t as targets and the remaining features as inputs. The constructed ELM model is, then, used to estimate the missing features of x_t. A similarity function λ (see *GRC* in Algorithm 3) is, then, used to calculate the similarity between the recently imputed observation $\hat{x}_t$ and every observation in X_k (Zhang, 2012). The similarity function uses a weighting coefficient γ with a range of 0.1 to 0.9, to define a set of neighbours. The weighting coefficient controls the selection of neighbours and is updated for each loop. ELMMI searches in X_k, then, selects the s nearest neighbours of $\hat{x}_t$ to form X_s. The number of nearest neighbours s is not a fixed value, but rather a fraction of the total number of observations (here it is set to 10%). It, then, makes use of

Algorithm 3: ELM-Based Multiple Imputation (ELMMI).

INPUTS:
$X_{m\times n}$ is an incomplete dataset
Γ is a set of ELM parameters
DEFINITIONS:
See ELMSI for $n^{mis}, n^{obs}, m^{mis}, m^{obs}, X_k, X_u$
X_{imp} is the imputed set with complete observations
$\lambda_{(x_t,x_c)}$ is the similarity between x_t and x_c observations
$\lambda_{(x_t,x_c)} = (1/n)\sum_{j=1}^{n} GRC(x_{tj}, x_{cj})$
GRC is the gray relational coefficient (Zhang, 2012)
$GRC(x_{tj}, x_{cj}) = 0.5/(|x_{tj} - x_{cj}| + 0.5)$
1. SPLIT X into X_k and X_u
2. SORT X_u w.r.t. n^{mis} from smallest to largest values
3. NORMALIZE both X_k and X_u
for $\alpha = (1:1:5)$ **do**
  for $\forall t \in X_u (t = 1, \dots, m^{mis})$ **do**
    4. TRAIN an ELM model using X_k w.r.t. x_t:
    $[\hat{X}_k^{mis}, H_k, \hat{\beta}_k] = f_k(X_k^{obs}, \Gamma)$
    where H is the hidden node matrix
    and $\hat{\beta}$ is the output weight matrix
    5. PREDICT missing scores of $x_t \in X_u$ by f_k:
    $\hat{x}_t^{mis} = f_k(x_t^{obs}, H_k, \hat{\beta}_k)$
    The completed observation is $\hat{x}_t = (x_t^{obs}, \hat{x}_t^{mis})$
    for $\forall c \in X_k (c = 1, \dots, m^{obs})$ **do**
      6. CALCULATE the weighting coefficient
      $\gamma = (2\alpha - 1)/10$
      7. CALCULATE the similarity function
      $\lambda_{(\hat{x}_t,x_c)} = (1-\gamma)\lambda_{(x_t^{obs},x_c^{obs})} + \gamma\lambda_{(\hat{x}_t^{mis},x_c^{mis})}$
      where γ is the weighting coefficient
      γ adjusts λ w.r.t. x_t^{obs} and $\hat{x}_t^{mis}$
    end
    8. SORT X_k w.r.t. λ values in descending order
    9. SELECT s observations with largest λ values
    $X_s = \{x_i \in X_k\}_{i=1}^{|s|=0.1\times m}$
    10. TRAIN an ELM model using
    X_s: $[\hat{X}_s^{mis}, H_s, \hat{\beta}_s] = f_s(X_s^{obs}, \Gamma)$
    11. PREDICT missing scores of $x_t \in X_u$ by f_s:
    $\hat{x}_t^{mis} = f_s(x_t^{obs}, H_s, \hat{\beta}_s)$
    The completed observation is $\hat{x}_t = (x_t^{obs}, \hat{x}_t^{mis})$
    12. REPLACE the completed $\hat{x}_t$ in X_k
  end
  13. RETURN the imputed set $X_{imp}(\alpha) = X_k$
end

all the observations in X_s to train an ELM model, using all corresponding incomplete features in x_t as targets and the rest of the features as inputs. The ELM model newly trained, by means of the closest neighbours, is, then, used to estimate the missing values of x_t. It finally adds the complete observation $\hat{x}_t$ to X_k.

4. Experimental results

The proposed prognostic scheme makes use of ELM-based prediction techniques (ELM, OS-ELM, KELM-WAV and KELM-RBF) and other state-of-the-art techniques, i.e., GMDH, RF, NF-FCM, NF-SC and NF-GP, in various experiments for one-step and multi-steps predictions, under incomplete scenarios. It also uses single imputation (i.e., ELMSI, kNNI and LWLA) and multiple imputation (i.e., ELMMI and MCMC) techniques to impute the missing observations prior to prediction.

Prediction results are compared in terms of prediction error (E_{RUL}) and root mean square error (RMSE). RMSE computes the difference between the predicted values and the target. E_{RUL} stands for the error in the prediction of the end of life that can be calculated as the difference between the actual and predicted cycle numbers at which the battery reaches the end of life, when the capacity value crosses a preset threshold. For these sets of experiments, the thresholds for batteries B0005, B0006 and B0007 are 1.299, 1.424 and 1.41, respectively.

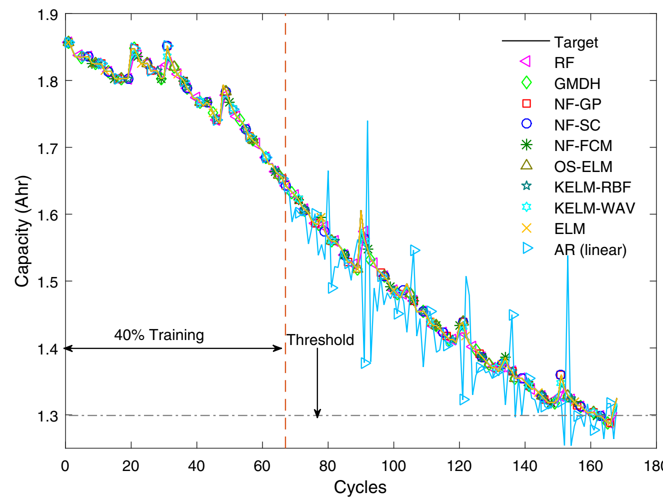

Fig. 2. One-step ahead prediction results for all non-linear prediction techniques compared to linear regression model, using the first 40% of the observations for training.

The prediction module relies on the trained predictive model using capacity data, which depends on data collection, transmission and storage. The collected sequence of observations usually has missing values, which may significantly impact on the prediction accuracy. To evaluate the performance of a predictive model in presence of missing observations, an incomplete data scenario has been created by inducing a set of missing observations in various cycles of the battery aging datasets. In some experiments related to multiple imputation techniques, the missingness has been randomly induced to different cycles, where a higher probability has been given to the last cycles of the available sequence used for training. The purpose was to evaluate the performance of the imputation techniques in a situation of greater uncertainty. The proposed prognostic scheme is intended for MCAR data, however, it is observed to perform on MAR data with high accuracy as well.

K-fold, stratified, nested cross-validation (CV) is used to evaluate the model performance in the experiments. Training, validation and test subsets to carry out the experiments for selecting proper parameters. Consequently, we repeated the experiments using nested cross-validation technique. The nested CV contains two stages. The inner stage is used to tune the model parameters and then the outer stage is used to evaluate the performance of the attained model from the inner stage. In the process, each dataset is first split into k_1 outer sets of time-series data. Each of the k_1 splits are used as the test subset while the remaining k_1 -1 splits are used as training subset. The inner stage further splits each of the training subset into k_2 splits. It iteratively holds out one of the k_2 splits for validation and a training model is generated using the k_2 -1 splits. This stage is used to tweak and optimize the hyperparameters to lower the RMSE of the predicted result. Once optimized, the parameters are used to re-train a prediction model in the outer stage and then evaluated on the test subset to get an unbiased estimate.

Fig. 2 depicts the one-step ahead prediction results for the linear and non-linear prediction models generated by each predictor along with the target, on the B0005 data using first 40% of the observations for training. The linear prediction result was gener-

Table 1
Compressed column representation of the missing dataset prior to imputation.

Count	φ_i	φ_{i-1}	φ_{i-2}	φ_{i-3}
88				
15				
15				
2				
16				
2				
2				
16				
2				
2				
2				
2				
164	23	23	24	24

ated using a linear autoregressive (AR) model and compared with the various non-linear predictors used in the experiments. The Figure clearly shows that even when using a small number of observations for training, non-linear prediction techniques generate models that are much closer to the target than a linear regression model. This shows the advantage of using a non-linear model in battery prognostics, which can use limited number of observations in order to generate an efficient prediction model.

The prediction module relies on the trained predictive model using capacity data, which depends on data collection, transmission and storage. The collected sequence of observations usually has missing values, which may significantly impact on the prediction accuracy. To evaluate the performance of a predictive model in presence of missing observations, an incomplete data scenario has been created by inducing a set of missing observations in various cycles of the battery aging datasets. The missingness has been randomly induced to different cycles.

Table 1 shows the missing data pattern in a collapsed column format. The grey solid squares indicate the missing values. The scenario includes 164 observations (the left most column in the last row of Table 1). This collapsed column Table shows the missing observations into the minimum number of rows possible. The lowermost row of Table 1 reports the number of missing observations in each input vector (φ_i, φ_{i-1}, φ_{i-2} and φ_{i-3}) in the scenario. For instance, φ_i has been missed in 23 observations. The number of the complete observations is 88 (see the second row of the Table 1). The rest of the rows represent incomplete observations. For instance, the fifth row shows an observation where φ_i and φ_{i-1} are missing together, and this type of missingness occurs two times in the scenario (see the left most column).

This incomplete set of observations including three input lags is fed to the pre-processing module, where missing values are imputed a priori and, then, fed to the prediction module.

Several OS and MS predictions are performed based on all battery cells data by changing the number of input lags between two and five. Moreover, the number of training observations is also changed to 40%, 60% and 80% of the total number of observations. Each of the battery datasets were treated as a separate problem. However, the graphical representation of the results from all three datasets are combined in order to keep the size and format of the section more concise. The plots depicting the RMSE and the E_{RUL} of the prediction results, show the overall performance of each imputation and prediction technique for all three datasets. Each of the dataset show a similar trend in performance of each technique without any bias.

4.1. Single imputation

The pre-processing module contains ELMSI (see Algorithm 2), which is compared with other state-of-the-art imputation techniques, i.e., kNNI and LWLA. The pre-processing module, then, imputes missing values of the incomplete scenario by means of different imputation techniques (i.e., kNNI, LWLA and ELMSI), where each imputation technique provides a complete subset, which is used by the prediction module to estimate the RUL.

4.2. Multiple imputation

The pre-processing module can also be equipped with multiple imputation techniques, which can reveal uncertainty in estimations. Here, Markov Chain Monte Carlo (MCMC) and ELMMI techniques have been devised in the pre-processing module for multiple imputation of the missing values. Each multiple imputation technique generates five completed subsets. Each of these completed subset is, then, fed to the prediction module. Multiple imputation using MCMC was generated with Fully conditional specification or chained equations imputation method. This is an iterative MCMC method that can be used when the pattern of missing data is arbitrary. Variables with missing values are imputed sequentially in each iteration. The imputed variables from one step are then also used as predictors for the subsequent steps. The dataset in the experiments were categorized as scale variables and hence, linear regression was selected for the model along with 10 iterations for multiple imputation of the missing data.

4.2.1. One-step-ahead prediction

One-step prediction can be performed through nine OS predictors, which receive three complete sequences of observations through different single and multiple imputation techniques.

For the incomplete scenario containing 15% missing values, Fig. 3 illustrates the RMSE values for each OS predictor applied on the subsets imputed by single and multiple imputation techniques. This incomplete data scenario contains three input lags and the first 60% of the observations are used for training. The solid squares represent the means of the RMSE values obtained by each OS predictor.

For prediction results generated using single imputation techniques, OS-ELM produces the lowest RMSE value when combined with ELMSI (as illustrated in Table 2). Different lines depict the mean of the RMSE values obtained by each imputation technique. From Fig. 3 we see that ELMSI performs well as an imputation technique when combined with most of the predictors. On the other hand, the RF and KELM-RBF predictors work better with kNNI. Overall, the attained results show that the proposed strategy of integrating ELMSI with several ELM- and NF-based predictors as pairs of imputation and prediction techniques, is able to achieve accurate and stable estimations.

For prediction results generated using multiple imputation techniques, OS-ELM produces the lowest RMSE value when combined with ELMMI (as illustrated in Table 2). The two different

Table 2
Mean RMSE values ($\times 10^{-3}$) calculated by each OS predictor on the completed subsets imputed by different missing data imputation techniques for all performed tests.

Imputation Technique	RF	GMDH	NF-GP	NF-SC	NF-FCM	OS-ELM	KELM-RBF	KELM-WAV	ELM
ELMSI	6.314	4.105	1.962	1.910	1.925	**1.865**	15.124	5.489	1.930
kNNI	6.116	4.631	2.316	2.310	2.331	2.285	14.846	5.349	2.323
LWLA	7.258	6.147	3.256	3.245	3.245	3.265	18.412	6.689	3.304
ELMMI	8.126	8.764	7.521	7.506	7.512	**7.487**	12.215	7.782	7.530
MCMC	9.295	8.892	8.845	8.902	8.860	8.853	12.105	8.958	8.428

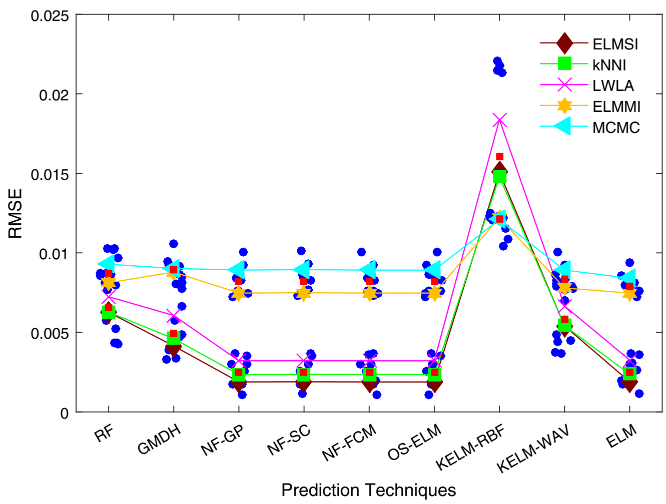

Fig. 3. RMSE values calculated by each OS predictor on the completed subsets imputed by different missing data imputation techniques for all performed tests. The lines in the plot represent the mean RMSE value for each imputation technique with each OS predictor. The blue dots represent the RMSE value for each individual experiment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

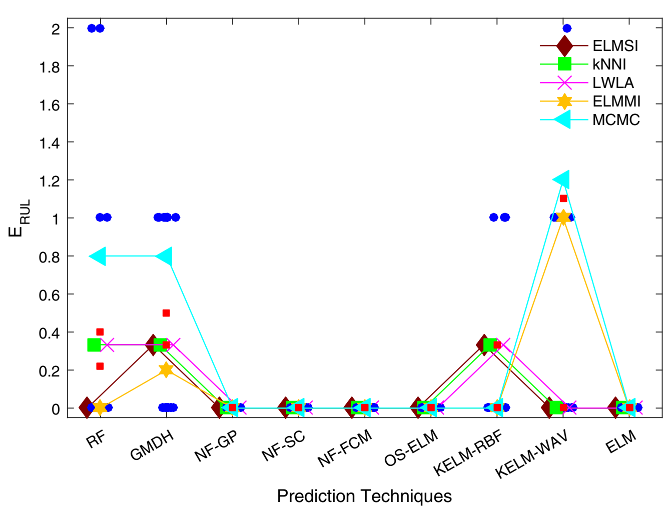

Fig. 4. E_{RUL} values calculated by each OS predictor on the completed subsets imputed by different single and multiple imputation techniques for all performed tests. The lines in the plot represent the mean E_{RUL} value for each imputation technique with each OS predictor. The blue dots represent the E_{RUL} value for each individual experiment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

lines depict the means of the RMSE values obtained by each multiple imputation technique. This Figure also shows that most of the OS predictors can provide a better prediction profile when combined with ELMMI, except for KELM-RBF.

From Fig. 4, the E_{RUL} values are plotted for each OS predictor along with each imputation strategy for all experiments performed on all the battery cells. A positive (negative) value for E_{RUL}, in terms of cycle, denotes a late (early) estimation, which indicates that the estimated trend crosses the threshold after (before) the target. An early prediction is often preferred over a late prediction. The individual lines represent mean of the E_{RUL} values for each imputation strategy.

For prediction results generated using single imputation techniques, this Figure indicates that OS-ELM, ELM, NFs and KELM-WAV can accurately predict the cross-threshold cycle using the imputed subsets generated. The attained OSP results for E_{RUL} (as illustrated in Table 3) also show that most of the OS predictors can provide less prediction error, when combined with ELMSI.

For prediction results generated using single imputation techniques, this Figure indicates that OS-ELM, ELM, NFs and KELM-WAV can accurately predict the cross-threshold cycle using the imputed subsets generated. The attained MSP results for E_{RUL} (as illustrated in Table 3) also show that most of the OS predictors can provide less prediction error, when combined with ELMMI.

4.2.2. Multi-steps prediction

Here, the pre-processing module has the same structure as in one-step prediction. However, the prediction module makes use of all of the long-term prediction techniques along with three different MSP strategies, i.e., iterative, DirRec and direct.

Fig. 5 depicts the prediction results for the multi-steps prediction models generated by each predictor along with the target, for the test $60\% - 3$ on the B0005 data imputed by ELMSI technique. This indicates that the input contains three lags and the first 60% of the observations are used for training. This Figure illustrates that the direct strategy, for the multi-steps prediction models, achieves lower prediction errors compared to the iterative and DirRec strategies, except for the NF predictors. ELM and KELM predictors show the least deviation from the target in estimated trends between the three different strategies.

Experimental results (not presented due to space constraint) also indicate that addition of more lags does not increase the accuracy of the prediction techniques.

In Fig. 6, the RMSE values are plotted for each predictor along with each imputation technique. The three lines in the figure represent the means of the RMSE values for each single imputation technique. This Figure also shows that ELMSI achieves good performance for missing data imputation, except for the combination with the NF and GMDH predictors in the direct prediction strategy, for which kNNI and LWLA provide slightly better results.

Table 4 shows that most predictors perform better for the iterative MSP strategy when combined with single imputation techniques. The direct MSP strategy produces better performance for prediction results generated with multiple imputation techniques. ELM and OS-ELM achieve good performances in all of the three MSP strategies. Besides, Fig. 6 shows that ELM can provide the most accurate and stable predictions among all MSP profiles based on the imputed datasets, while KELM-RBF provides the largest range of variation for all three strategies. It also shows the pre-

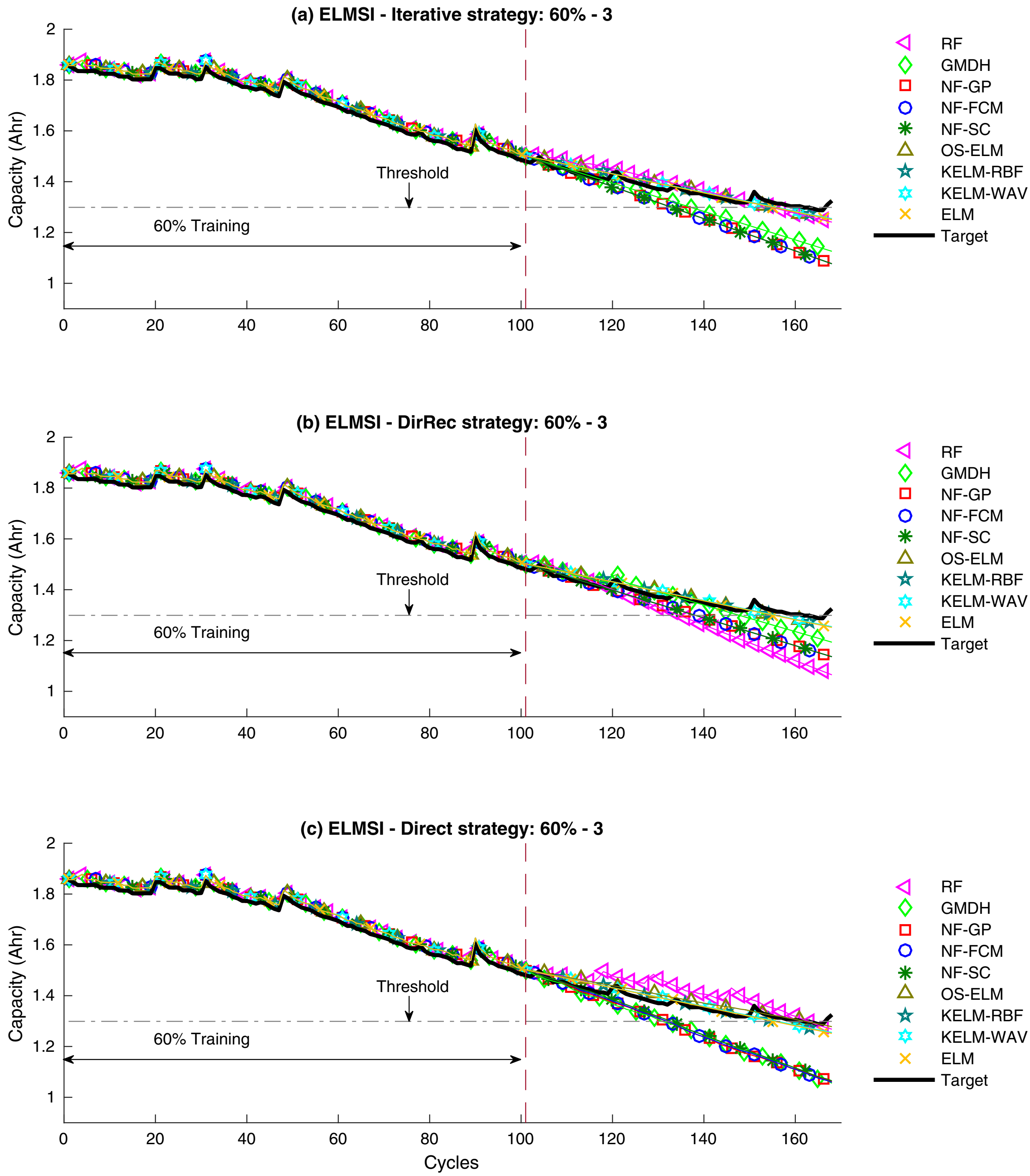

Fig. 5. Prediction results for the MSP strategies along with ELMSI technique, using three input lags and the first 60% of the observations for training.

Table 3
Mean E_{RUL} values calculated by each OS predictor on the completed subsets imputed by different missing data imputation techniques for all performed tests. Prediction results generated by multiple imputation techniques that failed to reach the threshold were ignored.

Imputation Technique	RF	GMDH	NF-GP	NF-SC	NF-FCM	OS-ELM	KELM-RBF	KELM-WAV	ELM
ELMSI	0	0.4	0	0	0	0	0.4	0	0
kNNI	0.4	0.4	0	0	0	0	0.4	0	0
LWLA	0.4	0.4	0	0	0	0	0.4	0	0
ELMMI	0	0.2	0	0	0	0	1	0	0
MCMC	0.8	0.8	0	0	0	0	1.2	0	0

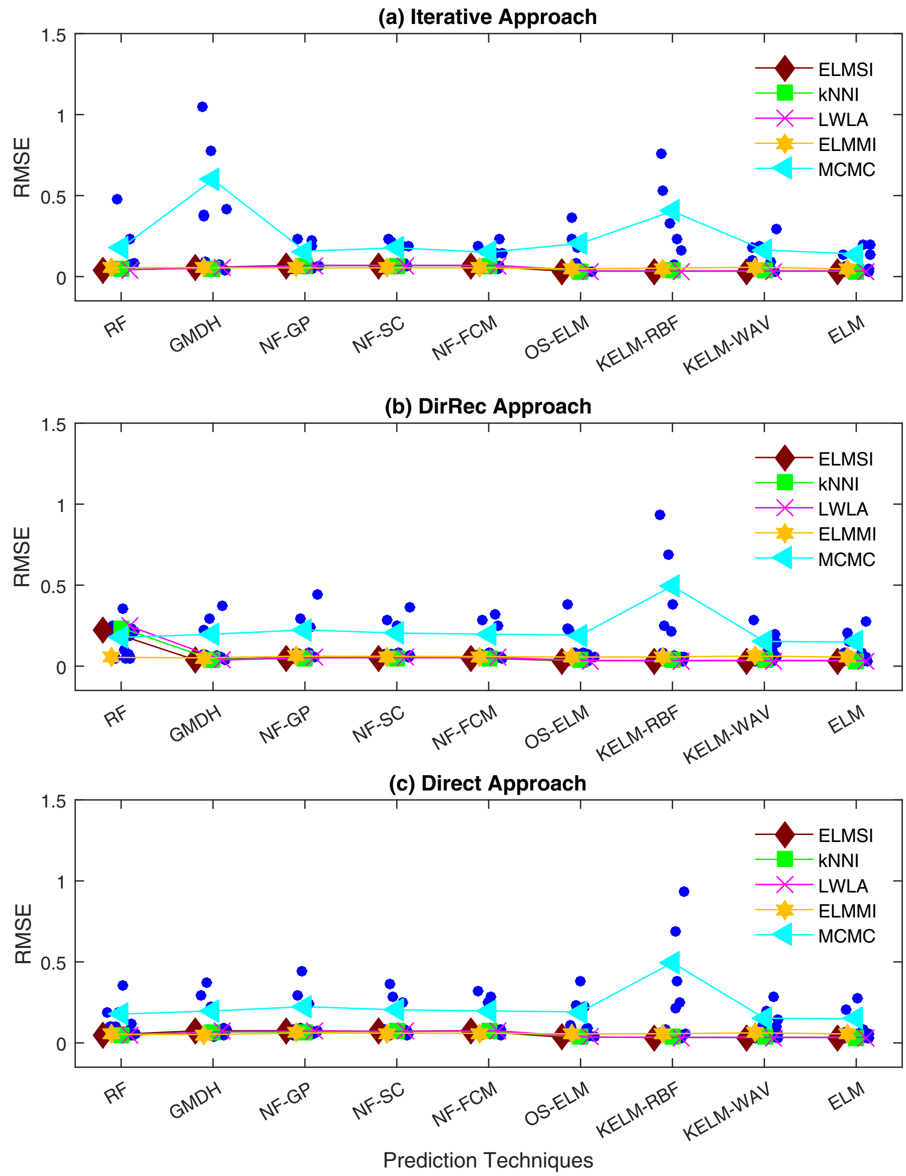

Fig. 6. RMSE values calculated by each predictor with different MSP strategies along with each single and multiple imputation technique for all performed tests. The lines in the plot represent the mean RMSE value for each imputation technique with each OS predictor. The blue dots represent the RMSE value for each individual experiment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Mean RMSE values ($\times 10^{-2}$) calculated by each MS predictor on the completed subsets imputed by different missing data imputation techniques for all performed tests.

	Imputation technique	RF	GMDH	NF-GP	NF-SC	NF-FCM	OS-ELM	KELM-RBF	KELM-WAV	ELM
Iterative	ELMSI	3.8102	5.4234	6.8306	6.8414	6.8421	3.4226	3.4291	3.4285	**3.3985**
	kNNI	4.5602	5.0020	6.8216	6.8522	6.8384	3.4320	3.4326	3.4326	3.4316
	LWLA	4.2520	5.5745	6.8625	6.8641	6.8632	3.4124	3.4227	3.4119	3.4110
	ELMMI	5.6784	5.6512	6.1352	6.0405	5.9256	5.6003	5.6714	6.2141	**5.5512**
	MCMC	17.7778	59.7878	22.3874	20.4014	19.7478	20.6541	40.3284	16.3932	14.8787
DirRec	ELMSI	22.6792	3.9581	5.1542	5.1547	5.1530	3.4245	3.4185	3.4180	3.4115
	kNNI	23.6141	3.9987	5.1641	5.1682	5.1623	3.4314	3.4387	3.4310	3.4315
	LWLA	24.5853	3.8482	5.2352	5.2366	5.2313	**3.4110**	3.4118	3.4118	3.4112
	ELMMI	5.3892	5.1346	6.1304	6.0504	5.9261	**5.6071**	5.6798	6.2105	5.6114
	MCMC	17.7777	19.7525	22.3898	20.4040	19.7452	19.1616	49.5200	15.2446	14.8725
Direct	ELMSI	4.8225	7.4154	7.5555	7.2489	7.5050	3.4989	3.5215	3.5210	**3.4985**
	kNNI	4.9562	6.7220	6.9210	7.1514	7.2385	3.5540	3.5343	3.5305	3.5290
	LWLA	4.7777	7.0140	7.3232	7.1144	7.4245	3.5778	3.5214	3.5220	4.3405
	ELMMI	5.3854	5.1313	5.3254	5.3782	5.3409	4.6810	5.1620	5.7543	**4.6185**
	MCMC	17.4152	19.7575	15.6954	17.7100	15.0045	19.1652	49.5857	15.2414	13.9551

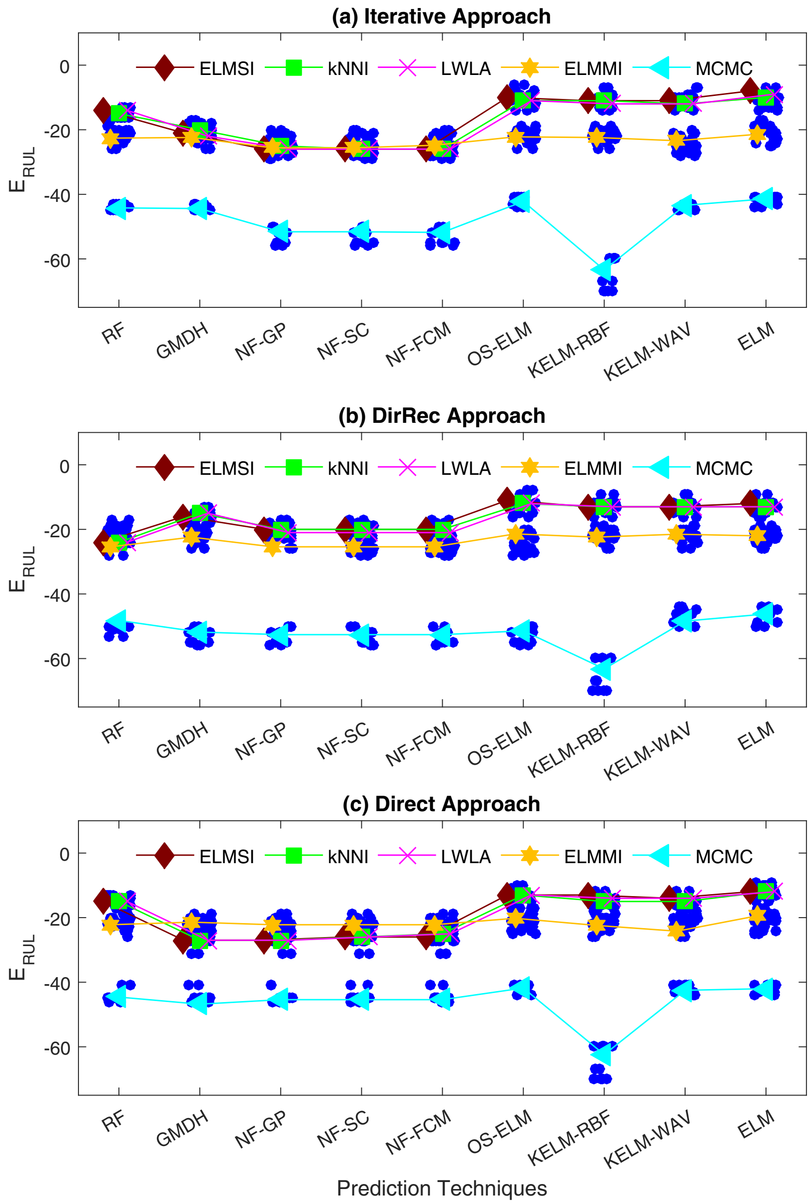

Fig. 7. E_{RUL} values calculated by each MS predictor on the completed subsets imputed by different single and multiple imputation techniques for all performed tests. The lines in the plot represent the mean E_{RUL} value for each imputation technique with each MS predictor. The blue dots represent the E_{RUL} value for each individual experiment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

dictions through ELMMI are more stable than predictions through MCMC over all three MSP strategies.

For prediction results generated using the proposed imputation techniques, the combination of ELM-based predictor models and ELMSI (ELMMI) for the single (multiple) imputation technique gives good results with low computational complexity and estimation errors.

Fig. 7 illustrates the distribution of the E_{RUL} values for each MS predictor along with each imputation strategy for all experiments performed on all the battery cells. A positive (negative) value for E_{RUL}, in terms of cycle, denotes a late (early) estimation, which indicates that the estimated trend crosses the threshold after (before) the target. An early prediction is often preferred over a late prediction. The individual lines represent mean of the E_{RUL} values for each imputation strategy.

The attained MSP results for E_{RUL} using single imputation techniques, which is illustrated in Table 5, show that the predictive models generated by ELMSI and the ELM-based predictors provide good performances with low prediction errors in calculating the remaining useful life of the battery.

The attained MSP results for E_{RUL} using multiple imputation techniques, which is illustrated in Table 5, show that all MS predictors can provide less prediction error, when combined with ELMMI. The prediction model generated by ELMMI-ELM using the direct strategy produce the least error in prediction of the battery remaining useful life.

Fig. 8 shows the highest profile, the lowest profile and the mean of all profiles for each predictor on all MSP strategies. The Figure illustrates the range of predictions (grey region) by each prediction technique, to reveal the impact of the uncertainty regarding the multiple imputed subsets on the prediction profiles. This Figure shows that, in general, predictions through ELMMI have lower range of variations compared to MCMC. Panel (b) in the Fig. 8 shows that the range of ELMMI predictions (see the grey regions) are not covering the target. However, predictions through

Table 5
Mean E_{RUL} values calculated by each MS predictor on the completed subsets imputed by different missing data imputation techniques for all performed tests. Prediction results generated by multiple imputation techniques that failed to reach the threshold were ignored.

	Imputation Technique	RF	GMDH	NF-GP	NF-SC	NF-FCM	OS-ELM	KELM-RBF	KELM-WAV	ELM
Iterative	ELMSI	−14	−21	−26	−26	−26	−10	−11	−11	**−8**
	kNNI	−15	−20	−25	−26	−26	−11	−11	−12	−10
	LWLA	−14	−22	−26	−26	−26	−11	−12	−12	−9
	ELMMI	−22.6	−22.4	−25.6	−25.6	−24.8	−22.2	−22.4	−23.4	**−21.4**
	MCMC	−44.2	−44.4	−51.6	−51.6	−51.8	−42.2	−63.4	−43.4	−41.4
DirRec	ELMSI	−24	−16	−20	−20	−20	**−11**	−13	−13	−12
	kNNI	−24	−15	−20	−20	−20	−12	−13	−13	−13
	LWLA	−24	−15	−21	−21	−21	−12	−13	−13	−13
	ELMMI	−25.4	−22.4	−25.4	−25.4	−25.4	**−21.4**	−22.4	−21.5	−22.0
	MCMC	−48.2	−51.8	−52.6	−52.6	−52.6	−51.4	−63.2	−48.4	−46.2
Direct	ELMSI	−15	−27	−27	−26	−26	−13	−13	−14	**−12**
	kNNI	−15	−27	−27	−26	−25	−13	−15	−15	−12
	LWLA	−15	−27	−27	−26	−25	−13	−14	−14	−12
	ELMMI	−22.2	−21.4	−22.2	−22.2	−22.2	−20.2	−22.4	−24.2	**−19.4**
	MCMC	−44.6	−46.8	−45.4	−45.4	−45.4	−41.8	−62.4	−42.5	−42.0

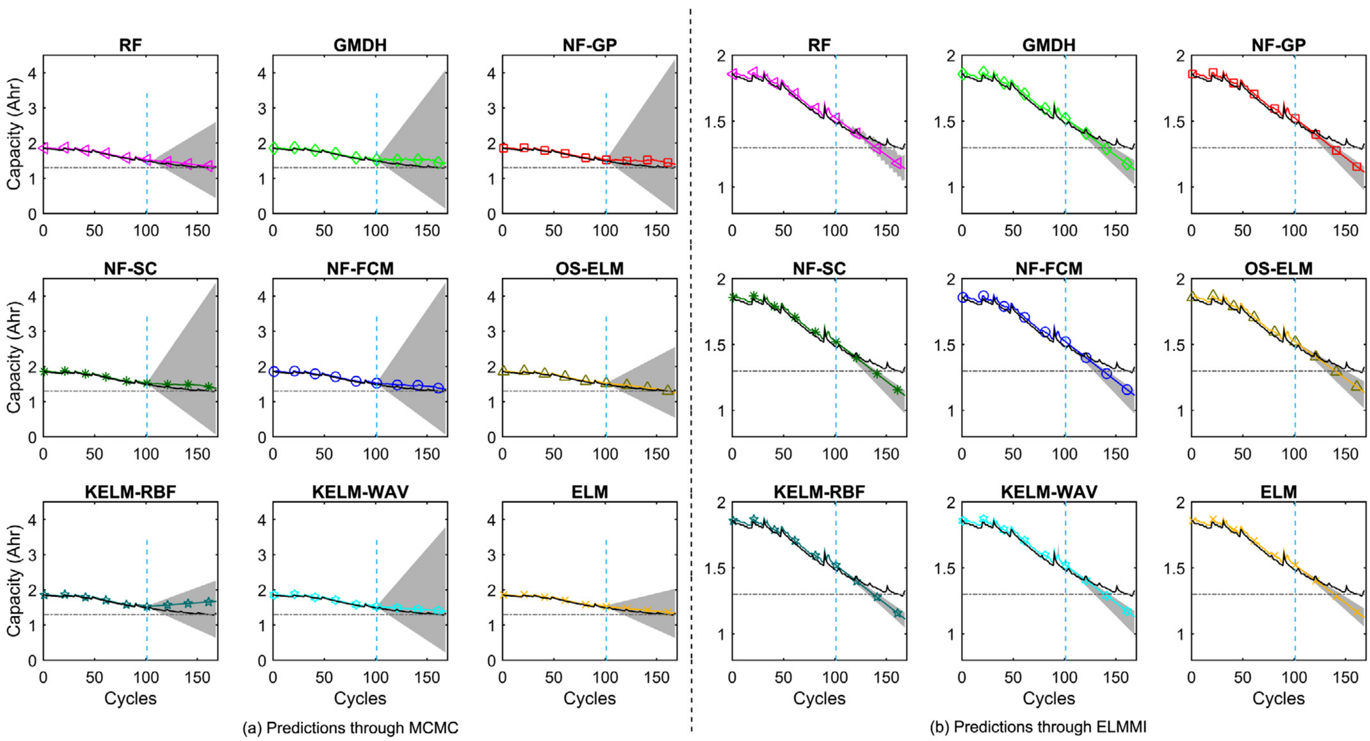

Fig. 8. Range of predictions (grey region) by each MS prediction technique based on the multiple imputed subsets through MCMC and ELMMI on the battery B0005 data.

ELMMI result in small negative E_{RUL} values, which are preferred to large negative or positive E_{RUL} values generated through MCMC.

To conclude, in the one-step prediction, OS-ELM and ELM sprovide good performances as predictor models in terms of RMSE and E_{RUL}, comparable to other state-of-the-art predictors like the neuro-fuzzy models. In the multi-steps prediction, all the ELM-based predictors perform well when combined with all the imputation techniques in terms of RMSE and E_{RUL}.

In terms of imputation, ELMSI and ELMMI perform well in most test cases. In one-step prediction, the proposed scheme reveals the strength of ELMSI and ELMMI as imputation techniques when combined with all the predictors, both in terms of RMSE and E_{RUL}. The proposed imputation techniques provide predictions with low estimation errors, comparable to other popular techniques. In the multi-steps prediction as well, both ELMSI and ELMMI display their strength as imputation techniques when combined with most of the predictors in terms of RMSE and E_{RUL}. ELMSI and ELMMI result in reasonable predictions with low estimation errors for all of the three MSP strategies. ELMMI, as a multiple imputation technique, generates a stable set of estimations with the lowest variance.

The experimental results show that ELM is an excellent tool for the proposed scheme of an integrated imputation-prediction model for the prediction of the short- and long-term remaining useful life of batteries. The algorithm provides a unique feature to generate a fast and accurate missing data imputation model, for both single and multiple imputations, as well as a prediction model. Even though NF predictors generate good prediction models in the one-step ahead predictions, their performance drops considerably in the multi-step ahead predictions. Both NF and RF-based predictors require high computational time, making them less suitable for online applications. For one-step-ahead predictions, the attained results show that the combination of OS-ELM and ELMSI (ELMMI) performs better than other competitors. For multi-step-ahead predictions, the attained results show that the combination of ELM and ELMSI (ELMMI) performs better than other competitors. Three strategies, iterative, DirRec and direct, are used in the multi-step ahead prediction methods. DirRec is highly complex and time consuming, making it not an ideal choice when speed of calculation is an important factor. An iterative strategy is the ideal choice for a longer prediction horizon. Direct strategy, despite the complexity, is the fastest of all the strategies, it provides the best results for short prediction horizons and it is the only strategy able to predict a non-continuous future value.

5. Conclusions

This paper proposes an efficient prognostic scheme for one-step-ahead and multi-steps-ahead predictions of the remaining useful life of Lithium-ion batteries, under incomplete scenarios. The proposed scheme contains two modules, one for pre-processing and one for prediction. In the former, novel techniques, called ELMSI and ELMMI, for single and multiple imputation of missing data have been developed, based on extreme learning machines. In the latter, several extreme learning machine-based prediction techniques have been devised along with various one-step and multi-steps prediction strategies to estimate the RUL of Lithium-ion batteries.

The primary goal of the work was to propose an effective integrated ELM-based framework, in order to handle both imputation and prediction tasks in situations of missing observations. In the pre-processing module, ELMSI (ELMMI) has been compared to some of the state-of-the-art single (multiple) imputation techniques including kNNI and LWLA (MCMC). The attained results show the effectiveness of the proposed imputation techniques in imputing the missing values and providing proper inputs for the prediction module. In the prediction module, different variants of ELM are used for prediction and, then, compared with various state-of-the-art prediction techniques. The attained results show that the ELM-based prediction techniques generate very accurate predictions compared to other competitors. Multiple imputation techniques have also been devised in the pre-processing module, in order to account for the uncertainty in the estimations. The attained results show that ELMMI can lead to predictions with less uncertainty, i.e., more confidence, compared to MCMC. In other words, the proposed prognostic scheme can impute missing observations, if necessary, and produce accurate predictions.

The paper does not claim that any individual imputation or prediction technique holds a significant advantage over the competitors. The proposed single and multiple imputation techniques show strong performances compared to other state-of-the-art imputation techniques. However, the evaluation of the individual techniques are beyond the scope of this paper and would require further research on a variety of datasets to properly evaluate their performances. The proposed integrated framework predicts the remaining useful life of batteries using the proposed techniques and also the competitors. This shows the effectiveness and versatility of the proposed framework, which can incorporate and interchange a variety of techniques and still produce fast and effective predictions.

The paper aims at improving the short- and long-term prediction performance of the remaining useful life of lithium-ion batteries under missing observations scenarios. A major drawback in any prediction model is over-fitting of data. Observations at different time cycles of a battery life do not follow the same regression model. Hence, using the entire observation for the training of the prediction models is not always recommended. The ability to optimize the training horizon both for imputation and prediction will be a great addition to the study in this paper. Handling of missing data in battery prognostics is bound to be an essential topic in future research in intelligent application systems. Most of the contemporary data-driven predictors are not suited to work with missing observations. Usually missing data scenarios are handled by eliminating the missing observations and only relying on complete observations. However, in time-series based applications such as those here considered of Lithium -ion battery prediction, a break in the time cycle data can lead to erroneous predictions. Therefore, there is a definite need for further research and implementation of fast and accurate missing data imputation techniques in prediction schemes for prognostics of battery data. Unfortunately, missing data imputation techniques are more accurate when there is a large pool of training data. As it was beyond the scope of this paper, the analysis of missing data imputation using fleet-based prognostic scheme would provide an excellent study for such a scenario. The proposed scheme uses a standard batch environment for the prediction of remaining useful life. Converting this to an online just-in-time environment, where the observations are fed into the system one by one or in batches and, then, generating a live prediction model for the remaining useful life of a battery, would be an interesting study. ELM-based prediction techniques are capable of producing multi-output predictions. It would be ideal to utilize a multi-output multi-step ahead prediction technique (e.g. Parallel technique) to reduce the computational time considerably. All these ideas will provide an excellent research platform for improvement in handling of missing data scenarios in the field of battery prognostics.

References

Acuna, D. E., Orchard, M. E., Silva, J. F., & Perez, A. (2015). Multiple-imputation–particle-filtering for uncertainty characterization in battery state-of-charge estimation problems with missing measurement data: Performance analysis and impact on prognostic algorithms. *International Journal of Prognostics and Health Management, 6*, 1–12.

Anava, O., Hazan, E., & Zeevi, A. (2015). Online time series prediction with missing data. *Proceedings of the 32nd International Conference on Machine Learning, PMLR, 37*, 2191–2199.

Bai, G., & Wang, P. (2016). Prognostics using an adaptive self-cognizant dynamic system approach. *IEEE Transactions on Reliability, 65*(3), 1427–1437.

Baraldi, P., Maio, F. D., Al-Dahidi, S., Zio, E., & Mangili, F. (2017). Prediction of industrial equipment remaining useful life by fuzzy similarity and belief function theory. *Expert Systems with Applications, 83*, 226–241.

Barre, A., Suard, F., Gerard, M., Montaru, M., & Riu, D. (2014). Statistical analysis for understanding and predicting battery degradations in real-life electric vehicle use. *Journal of Power Sources, 245*, 846–856.

Batista, G., & Monard, M. (2002). A study of k-nearest neighbour as an imputation method. In *Hybrid Intelligent Systems, Ser-Front Artificial Intelligence and Applications 87, IOS Press* (pp. 251–260).

Bequao, A., & Lessmann, S. (2017). Extreme learning machines for credit scoring: An empirical evaluation. *Expert Systems with Applications, 86*, 42–53.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Chao, K.-H., & Chen, J.-W. (2011). State-of-health estimator based-on extension theory with a learning mechanism for lead-acid batteries. *Expert Systems with Applications, 38*(12), 15183–15193.

Charkhgard, M., & Farrokhi, M. (2010). State-of-charge estimation for lithium-ion batteries using neural networks and EKF. *IEEE Transaction on Industrial Electronics, 57*, 4178–4187.

Goebel, K., Saha, B., Saxena, A., Celaya, J. R., & Christophersen, J. P. (2008). Prognostics in battery health management. *Instrumentation & Measurement Magazine, 8*, 33–40.

Gouriveau, R., & Zerhouni, N. (2012). Connexionist-systems-based long term prediction approaches for prognostics. *IEEE Transactions on Reliability, 61*(4), 909–920.

Huang, G. (2014). An insight into extreme learning machines: Random neurons, random features and kernels. *Cognitive Computation, 6*, 376–390.

Huang, G. B., Zhou, H., Ding, X., & Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics-PART B: Cybernetics, 45*(2), 513–529.

Ivakhnenko, A. (1978). The group method of data handling in long-range forecasting. *Technological Forecasting and Social Change, 12*(2–3), 213–217.

Jang, J.-S. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics, 23*(3), 665–685.

Javed, K., Gouriveau, R., Zerhouni, N., & Hissel, D. (2016). Prognostics of proton exchange membrane fuel cells stack using an ensemble of constraints based connectionist networks. *Journal of Power Sources, 324*, 745–757.

Junior, J. R. B., do Carmo Nicoletti, M., & Zhao, L. (2016). An embedded imputation method via attribute-based decision graphs. *Expert Systems with Applications, 57*, 159–177.

Lee, S., Cui, H., Rezvanizaniani, M., & Ni, J. (2012). Battery prognostics: SOC and SOH prediction. In *Proceedings of the ASME International Manufacturing Science and Engineering Conference* (pp. 1–7).

Liang, N., Huang, G., Saratchandran, P., & Sundararaja, N. (2006). A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Transactions on Neural Networks, 17*(6), 1411–1423.

Liao, L., & Kottig, F. (2014). Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. *IEEE Transactions on Reliability, 63*(1), 191–207.

Liu, C., Dai, D., & Yan, H. (2010). The theoretic framework of local weighted approximation for microarray missing value estimation. *Pattern Recongnition, 43*, 2993–3002.

Liu, D., Pang, J., Zhou, J., Peng, Y., & Pecht, M. (2013). Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression. *Microelectronics Reliability, 53*(6), 832–839.

Razavi-Far, R., Chakrabarti, S., & Saif, M. (2016a). Multi-step-ahead prediction techniques for Lithium-ion batteries condition prognosis. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 004675–004680).

Razavi-Far, R., Farajzadeh-Zanjani, M., Chakrabarti, S., & Saif, M. (2016b). Data-driven prognostic techniques for estimation of the remaining useful life of lithium-ion batteries. In *Proceedings of the IEEE International Conference on Prognostics and Health Management (ICPHM)* (pp. 1–8).

Razavi-Far, R., Zio, E., & Palade, V. (2014). Efficient residuals pre-processing for diagnosing multi-class faults in a doubly fed induction generator, under missing data scenarios. *Expert Systems with Applications, 41*(14), 6386–6399.

Rubin, D. (2009). Multiple imputation for nonresponse in surveys. Wiley Series in Probability and Statistics.

Saha, B., & Goebel, K. (2007). *Battery data set: NASA Ames prognostics data repository*. In http://ti.arc.nasa.gov/project/prognostic-data-repository. NASA Ames, Moffett Field, CA.

Saha, B., Goebel, K., Poll, S., & Christophersen, J. (2009). Prognostics methods for battery health monitoring using a Bayesian framework. *IEEE Transaction on Instrumentation and Measurement, 58*, 291–296.

Sankararaman, S., Daigle, M. J., & Goebel, K. (2014). Uncertainty quantification in remaining useful life prediction using first-order reliability methods. *IEEE Transactions on Reliability, 63*(2), 603–619.

Saxena, A., Celaya, J., Roychoudhury, I., Saha, S., Saha, B., & Goebel, K. (2012). Designing data-driven battery prognostic approaches for variable loading profiles: Some lessons learned. In *Proceedings of the European Conference of Prognostics & Health Management* (pp. 1–11).

Sbarufatti, C., Corbetta, M., Giglio, M., & Cadini, F. (2017). Adaptive prognosis of lithium-ion batteries based on the combination of particle filters and radial basis function neural networks. *Journal of Power Sources, 344*, 128–140.

Scheffer, J. (2002). Dealing with missing data. *Research Letters in the Information and Mathematical Sciences, 3*, 153–160.

Singh, R., Kumar, H., & Singla, R. (2015). An intrusion detection system using network traffic profiling and online sequential extreme learning machine. *Expert Systems with Applications, 42*(22), 8609–8624.

Sorjamaa, A., Hao, J., Reyhani, N., Ji, Y., & Lendasse, A. (2007). Methodology for long-term prediction of time series. *Neurocomputing, 70*(16–18), 2861–2869.

Sovilj, D., Eirola, E., Miche, Y., Bjrk, K., Nian, R., Akusok, A., & Lendasse, A. (2016). Extreme learning machine for missing data using multiple imputation. *Neurocomputing, 174*(A), 220–231.

Taieb, S., Sorjamaa, A., & Bontempi, G. (2010). Multiple-output modeling for multi--step-ahead time series forecasting. *Neurocomputing, 73*(10–12), 1950–1957.

Takagi, H., & Hayashi, I. (1991). Artificial neural network driven fuzzy reasoning. *International Journal of Approximate Reasoning*, 191–212.

Tran, V. T., Yang, B.-S., & Tan, A. C. C. (2009). Multi-step ahead direct prediction for the machine condition prognosis using regression trees and neuro-fuzzy systems. *Expert Systems with Applications, 36*(5), 9378–9387.

Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A., & Wu, B. (2006). Intelligent fault diagnosis and prognosis for engineering systems. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Wei, Y., Wang, H., Shi, S., Gao, H., & Li, J. (2017). Any-time methods for time-series prediction with missing observations. *Proceedings of the IEEE 6th International Congress on Big Data*, (pp. 427–430).

Widodo, A., Shim, M.-C., Caesarendra, W., & Yang, B.-S. (2011). Intelligent prognostics for battery health monitoring based on sample entropy. *Expert Systems with Applications, 38*(9), 11763–11769.

Wu, X., Wang, Y., Mao, J., Dua, Z., & Lia, C. (2014). Multi-step prediction of time series with random missing data. *Applied Mathematical Modelling, 38*(14), 3512–3522.

Xu, X., Li, Z., & Chen, N. (2016). A hierarchical model for lithium-ion battery degradation prediction. *IEEE Transactions on Reliability, 65*(1), 310–325.

Yu, Q., Miche, Y., Eirola, E., Heeswijk, M. V., Severin, E., & Lendasse, A. (2013). Regularized extreme learning machine for regression with missing data. *Neurocomputing, 102*, 45–51.

Zaidan, M., Relan, R., Mills, A., & Harrison, R. (2015). Prognostics of gas turbine engine: An integrated approach. *Expert Systems with Applications, 42*(22), 8472–8483.

Zhang, J., & Lee, J. (2011). A review on prognostics and health monitoring of li-ion battery. *Journal of Power Sources, 196*, 6007–6014.

Zhang, S. (2012). Nearest neighbor selection for iteratively kNN imputation. *Journal of Systems and Software, 85*(11), 2541–2552.

Zhao, Z., Quan, Q., & Cai, K.-Y. (2014). A profust reliability based approach to prognostics and health management. *IEEE Transactions on Reliability, 63*(1), 26–41.

Zio, E. (2013). *Diagnostics and prognostics of engineering systems: methods and techniques* (pp. 333–356). Engineering Science Reference.