24th International Conference & Exhibition on Electricity Distribution (CIRED)

12-15 June 2017

Session 1: Network components

Data analysis of battery storage systems

ISSN 2515-0855
doi: 10.1049/oap-cired.2017.0657
www.ietdl.org

Merlinda Andoni ✉, *Wenshuo Tang, Valentin Robu, David Flynn*

School of Engineering and Physical Sciences, Smart Systems Group, Heriot-Watt University, UK
✉ *E-mail: ma146@hw.ac.uk*

Abstract: Battery energy storage systems can assist distribution network operators (DNOs) to face the challenges raised by the substantial increase in distributed renewable generation. A challenge is that these resources are intermittent and often 'invisible' to the DNO. If not monitored, the aggregate size of small embedded generation resources can cause thermal wearing of distribution assets and voltage excursions, especially in sunny/windy periods with insufficient local demand. Several developers of energy storage solutions, with technologies such as lithium-ion (Li-ion) batteries, offer their products to address peak shaving, frequency and voltage control needs within the network. Once deployed within the energy network batteries experience capacity degradation with usage, these companies will need to incorporate methods from prognostics and health management (PHM) in order to better manage their products. The main deliverable of this project is validation of data analysis, based on relevance vector machine, to predict the remaining useful life of Li-ion batteries. The accuracy of the predictions for different batteries is all within 10 cycles (within 8.5% relative error). These results confirm the importance of PHM methods within a distribution system operator model, where lifecycle management of critical sub-systems and systems will become increasingly important to network operators.

1 Introduction

Recent years have seen large penetration of distributed energy resources (DERs) in medium-voltage (MV) and low-voltage (LV) distribution grids, as part of decarbonisation agenda of the power generation sector. This growth has been supported by low-carbon initiatives, regulatory changes and provision of financial incentives that promote renewable technologies and reduction of CO_2 emissions.

Further use of renewable generators, which are inherently intermittent and difficult to predict, may offer benefits and opportunities for distribution network operators (DNOs); however it also poses new challenges that need to be addressed. Existing conventional generation was originally designed for efficient operation; however it is now required to provide flexible services, which often leads to exceeding desired economic operation set points. Moreover, power grids were designed for bulk energy generation, transmission and distribution to load centres than interconnection of multiple generating units, which are dispersed in various locations of the grid. Electricity grids are undergoing a massive change, as smaller generators are often connected at MV or LV level leading to two-way flows of electricity.

One key challenge is the large increase of small rooftop PV array installations and wind turbines connected at LV networks, which are not under the control of the DNO. In fact, much of the generation is not monitored and therefore is 'invisible' to the network operator. The impact of the aggregate generation, which is fluctuating and depending on weather conditions, may cause power quality and system stability issues. Controls such as voltage and frequency regulation can become challenging. Typically, frequency control is supported by expensive peaking generators providing 'spinning reserve' and voltage is kept within acceptable technical limits by the operation of capacitor banks or on-load-tap-changers (OLTC). Frequent use of these measures though may reduce the lifetime of this equipment.

Another key problem is voltage violations to feeders with high-photovoltaic (PV) penetration, especially at times when PV generation is high and local demand is low. For instance, the generation from PV arrays is at its highest point at noon, when household consumption is low. Usually, the peak residential demand is observed in the early evening, leading to asynchronous generation and demand. This can lead to temporary overvoltage and reverse power flows which are undesirable for DNOs for various reasons related to technical limitations of the OLTC and protection equipment or network congestion issues [1].

Energy storage systems (ESS) can mitigate some of the challenges faced due to the stochastic nature of renewable generation, such as prevent voltage violations, match demand to supply or make efficient use of renewable generation by minimising renewable curtailment. Curtailment occurs when the power that can be generated cannot be absorbed by the power system, due to network constraints or excess renewable generation [2] In fact, energy storage is recognised as a key technology for electricity grid transition towards the smart grid [3]. Storage can also provide frequency regulation and voltage support and therefore increase the power system reliability and efficiency. Furthermore, ESS can provide load support in cases of power system contingencies or sudden increase of demanded energy, playing the role of reserve power plants. Utilisation of storage devices can lead to network upgrade deferral, enhanced utilisation of distribution assets and lower power supply costs [4].

In recent years, there has been a growth in interest of battery technologies for grid operation. Historically, high costs have limited the extensive use of battery storage devices for such applications, however different factors such as financial incentives, technology advances and economy of scale expected by the growth of electric vehicles (EVs), are anticipated to be game-changing. EVs can be seen as large volumes of distributed storage dispersed and embedded in the LV network, where DNOs can make use of vehicle-to-grid operation and undertake the batteries management.

On top of decreasing costs, batteries provide several advantages such as high efficiency, fast response time, scalability and have no geographical limitations. Battery technologies are coupled with bidirectional four-quadrant power converters, which can provide solutions to DNOs, such as active and reactive power control [5], voltage support and matching demand to supply. Storage devices can operate as a buffer harvesting and storing excess energy from renewable generators at low demand periods and injecting the energy back to the grid instantly, when demand is high. As a result: 'spinning reserve' requirement by expensive stand-by units is avoided and reverse power flows are minimised. Furthermore, by 'peak shaving', distribution assets such as power cables remain under their thermal limits and their useful lifetime is improved. Moreover, battery storage can reduce transmission and distribution losses [6]. One such system can produce or consume both active and reactive power according to system needs. As a result, battery systems are appropriate for reducing and mitigating voltage sags, swells and flicker [1].

Different types of batteries have been used for grid storage applications (lead acid, NaS) [7], but in the recent years, lithium-ion (Li-ion) batteries have received the power system's community attention, as they are rechargeable, have high energy density and efficiency and are expected to have considerably lower cost in the future (currently higher cost than \$500/kWh) [8]. In [9], Li-ion batteries are used to provide peak load shaving, power curve smoothing and voltage regulation services of a distribution transformer. In [10], three-phase battery storage in residential LV networks is discussed. It can provide peak shaving, valley filling, load balancing and better utilisation of renewable generation.

Battery storage banks can be placed in various parts of the grid such as at substations, at MV [11] or LV level [10], at community or household level. For example, the fluctuating power output of PV generators can be mitigated by small storage capacity placed locally. In fact, the installation of distributed energy storage in the LV part of the network may be cost-effective, especially as there is high penetration of PV, operation close to technical rating limits or unbalanced loads [10]. Additionally, batteries can be placed to support heavily loaded distribution feeders, by reducing the peak demand of the distribution transformer and thus improving their useful lifetime [9].

Depending on location, purpose of installation or service provided, battery storage may follow different business models. A traditional approach is network operator ownership and operation of the storage system. In this case, batteries could be placed centrally at a substation level or strategically at suitable locations in the feeders. However, network operators are mostly interested in the services that energy storage providers can offer, such as the energy or reserve capacity provided. Therefore, it is possible, in the context of the deregulated energy market, for independent storage providers to offer services such as peak shaving, load levelling, frequency and voltage control. The UK has an abundance of independent storage providers; however the business case for ownership of storage systems needs further exploration as the benefits of storage and value added is spans from consumers, to DNOs and the system regulator.

Crucially, if batteries are deployed, there is a requirement for improved prognostics and health management (PHI) to predict the remaining useful life (RUL) and improve operational efficiency. Batteries experience reduced capacity with operation, due to 'cell ageing' and irreversible chemical reactions taking place during usage. Complex factors affect the RUL of batteries such as the depth of discharge and the ambient temperature during operation. As the physical mechanisms that lead to 'cell ageing' are complex to understand, we need to develop tools that can monitor the asset's health and predict failure, in order to increase the reliability and resilience of the overall system. This drives us to develop a more integrated and holistic energy storage management system, using historical data and machine learning techniques. Unlike model-based approach, data-driven approach does not rely on the physical modelling of cell degradation. It uses historical data and battery metrics (such as current, voltage, battery and ambient temperature in our study) to derive a non-parametric model and develop trends that can predict future asset behaviour. Other research works use machine learning methods for asset health management of batteries, as in [12–14].

This paper presents an important first step in achieving this vision, namely the design of a data-driven prognostic approach for the Li-ion battery. We present this concept with preliminary experimental results using the state-of-art machine learning technique, relevance vector machine (RVM), to predict the RUL of Li-ion batteries. We adopt our proposed model with the open-source, life cycle test data set from the National Aeronautics and Space Administration (NASA) battery repository. Preliminary experimental results show the proposed algorithm is able to generate sufficiently accurate prediction results. Specifically, our experiments show the predicted RUL for four different battery packs in the NASA dataset are all within 10 cycles of the actual RUL at the inspection starting points. Due to its predictive capability, the proposed algorithm is expected to be used in further work in the asset management system for large scale energy storage networks.

2 Algorithm and data source

2.1 Brief introduction of relevance vector machine

RVM is a technique in supervised learning, first developed by Tipping in 2000 [15]. The basic idea of RVM is a Bayesian treatment of support vector machine (SVM). The Bayesian treatment leads to probabilistic predictions, and allows arbitrary kernel functions to be utilised. The following section briefly explains the principles of RVM.

A training data set is composed of input vectors $\{X_n\}_{n=1}^N$ and corresponding targets $\{T_n\}_{n=1}^N$, which can be either values or classification labels, depending on application.

In supervised learning, predictions are often made based on a model $y(x)$, which is the sum of M linearly weighted basis functions. Formally, the model is defined as

$$y(\boldsymbol{x};\boldsymbol{w}) = \sum_{i=1}^{M} \omega_i \phi_i(X) = \boldsymbol{\omega}^{\mathrm{T}} \boldsymbol{\phi}(x)$$

where $\boldsymbol{\phi}(X) = (\phi_1(x), \phi_2(x), \ldots, \phi_M(x))^{\mathrm{T}}$ represents the basis function for each x, while $\boldsymbol{w} = (\omega_1, \omega_2, \ldots, \omega_m)^{\mathrm{T}}$ represent the adjustable weights associated with each basis function. In SVM, the basis functions used is kernel functions. For SVM, $y(\boldsymbol{x})$ can be re-written as

$$y(\boldsymbol{x};\boldsymbol{w}) = \sum_{i=1}^{M} \omega_i K\left(x, x_i\right) + w_0 = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{\phi}(x)$$

In RVM, kernel functions are used as in SVM, and the target values $\{T_n\}_{n=1}^N$ are assumed to be samples from the above model with additive noise, expressed as

$$t_n = y\left(x_n; \boldsymbol{w}\right) + \varepsilon_n$$

In this equation, ε_n is the noise factor which is assumed to be normally distributed with mean zero and variance σ^2, furthermore, assuming independence of t_n, the likelihood of the entire dataset can be written as

$$p\left(\boldsymbol{t}|\boldsymbol{w}, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\left\|\boldsymbol{t} - \phi w\right\|^2\right\}$$

where $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)$; and $\boldsymbol{\phi} = (\phi(x_1), \phi(x_2), \ldots, \phi(x_n))^{\mathrm{T}}$.

The above maximum likelihood function is likely to result in severe over-fitting because the number of parameters equals the number of input examples. The RVM utilises a Bayesian perspective to resolve this issue. Specifically, the RVM introduces additional constraints on weight parameters $\boldsymbol{w}$. Specifically, a zero-mean Gaussian prior distribution is chosen over $\boldsymbol{w}$

$$p(w|a) = \prod_{i=0}^{N} \left(N\left(\omega_i|0, a_i^{-1}\right)\right)$$

where α is the vector of $N+1$ hyperparameters, these hyperparameters are associated independently with every weight. The prior is defined as Γ distribution of the following form:

$$p(w|a) = \prod_{i=0}^{N} \Gamma(a_i|a, b), \quad p(\beta) = \Gamma(\beta|c, d), \quad \beta = \sigma^2$$

the priors a, b, c, and d are set to zero to be non-informative, hence these hyperpriors are scale invariance, making the predictions independent of linear scaling of either t or $\phi(x)$. In order to arrive at sparsely distributed weight parameters ω, RVM uses Bayes' rule, and it is possible to derive the posterior (conditional probability) distribution over the weights parameter w:

$$p(\boldsymbol{\omega}|\boldsymbol{t}, \boldsymbol{a}, \sigma^2) = \frac{p(\boldsymbol{\omega}|\boldsymbol{t}, \boldsymbol{a}, \sigma^2)p(\boldsymbol{\omega}|\boldsymbol{a})}{p(\boldsymbol{t}|\boldsymbol{a}, \sigma^2)}$$
$$= (2\pi)^{(-(N+1)/2)} \left|\sum\right|^{(-1/2)} \exp -\frac{1}{2}\left(w - \mu^{\mathrm{T}} \sum -1(w - \mu)\right)$$

where σ and μ are the posterior covariance and mean, respectively. Typically, the RVM requires computation and optimization of the hyper parameters α, however, as the training dataset increases, the range of α may increase to infinity. In this case, the matrix Σ does not have an inverse, making relevance vectors impossible to derive. Moreover, as the amount of data increases, computation efficiency is also reduced. Therefore, an iterative expectation–maximization (EM) [16] algorithm will be used for RVM training in this paper. Using this particular algorithm can directly avoid the step of optimizing hyperparameters.

With the established EM algorithms, predictions of battery RUL are obtained through the following procedure. First, the available capacity degradation data are used as the inputs for the developed method. Then, RVM is performed on the data for regression analysis. Next, the obtained relevance vectors are fed into the iterative EM algorithm to generate capacity prediction of the next cycle. This new predictive information is put together with the original relevance vectors to form a new training set for the RVM model. This process will terminate at the iterative kth cycle once the $(k+1)$th prediction satisfies the prediction error criteria. In this study, battery failure can is considered to occur once the maximum capacity of the battery drops to 70% of its nominal value. At this stage, the obtained output is valued against the failure threshold criteria to determine the RUL of the battery.

2.2 Data source and battery capacity degradation model

The battery data used to conduct the prognostic experiment were obtained from the open-source, life cycle test data repository of the NASA Ames Prognostics Center of Excellence. In this dataset, 34 Li-ion battery packs (four batteries in one pack) were used to run the life cycle test in different experimental conditions. Each battery pack was run repeatedly through charge and discharge operations. A typical charge and discharge process is regarded as a valuable cycle, which is the key measurement of the RUL of the batteries in our prognostic model. Specifically, in the charging process, batteries are charged at a constant current of 1.5 A until the battery voltage reached 4.2 V, then batteries were continued to be charged at a constant voltage until the charge current dropped to 20 mA. Discharging was carried out at constant current at 2 A until the battery voltage dropped from 4.2 V to a cut-off voltage.

The experiment was conducted in two different room temperatures (25 and 4°C) and four different cut-off discharge voltage (2.7, 2.5, 2.2 and 2.5 V) for different packs. Furthermore, the experiments were terminated when batteries reach their end-of-life criteria of 30% fade in initial rated capacity (70% remaining capacity). Beyond this point, the batteries are no longer considered as reliable power generating assets. During the cycle test, all key physical measurement was recorded through sensors. The data format is shown in Fig. 1.

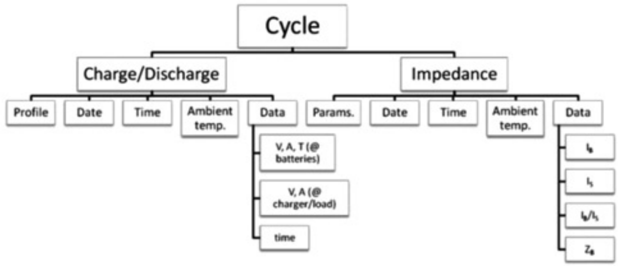

Fig. 1 *Data structure*

3 Experiment and result analysis

We adopt the RVM model to NASA battery dataset to evaluate the proposed battery RUL prediction algorithm.

To measure the error of the predict RUL of the battery, we define the absolute error (AE) and relative error (RE) as $AE = |R - \ddot{R}|$, and $RE = (|R - \ddot{R}|/R)$, where R is the actual RUL value and $\ddot{R}$ is the predicted RUL value.

First, we implement the RUL estimation with battery No. 5 in this dataset, in which different starting points are selected. These starting points are selected, namely the 40th, 60th, and the 80th cycles. The RUL prediction results are shown in Table 1.

In Table 1, we can see that for battery no. 5 all of the predicted RUL errors are less than 10 cycles at different starting points. Moreover, all the actual RUL values are almost located in confidence intervals (CI). Figs. 2 and 3 show the plot of real data and the predictive point estimates for battery No. 5. Fig. 2 shows the plot with starting point at the 40th cycle, while Fig. 3 shows the plot starting at the 80th cycle.

Our results show that the RVM estimation has a good performance in the long term prediction on forecasting the battery RUL. In particular, the latter the starting cycle is (cycle 80), the more

Table 1 Quantitative results for RUL predictions using our RVM-based algorithms

Battery no.	Starting point	True RUL	Predicted RUL	AE	RE%
5	40	124	117	7	5.6
5	60	124	120	4	3.2
5	80	124	121	4	3.2
6	40	112	103	9	8
6	60	112	102.5	9.5	8.4
6	80	112	107	5	4.4
7	40	166	158	8	4.8
7	60	166	159	7	4.2
7	80	166	159	7	4.2
18	40	132	122	10	7.5
18	60	132	124	8	6.1
18	80	132	126	6	4.5

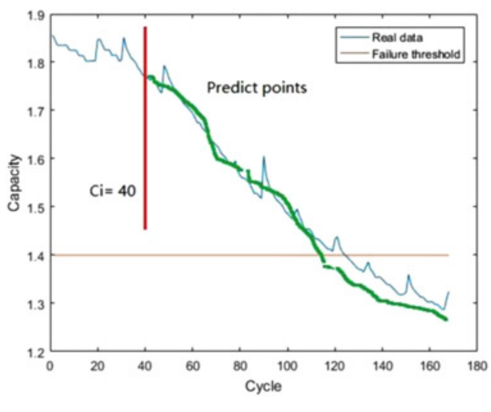

Fig. 2 *RUL perdition curve for battery No. 5 at 40th cycle*

CIRED, Open Access Proc. J., 2017, Vol. 2017, Iss. 1, pp. 96–99

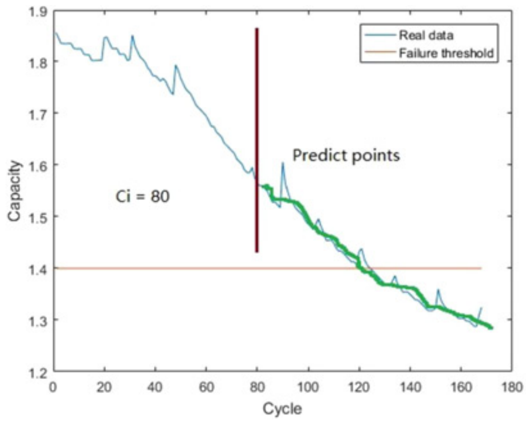

Fig. 3 *RUL perdition curve for battery No. 5 at 80th cycle*

accurate the resulting prediction of RUL will be. To verify and evaluate the adaptability of the proposed method, we also implemented the RUL prediction experiment using other batteries. The experimental results with battery No. 6, No. 7 and No. 18 are shown in Table 1. Similar to that of battery No. 5, the prediction precision measurements AE and RE are satisfactory. The prediction precision proves that the proposed method has a good performance for the application we consider.

The experiments presented above use NASA battery data to test the predictive capacity of the proposed algorithm. The results for four different battery packs shows that regardless of the starting point, whether 40th, 60th or 80th cycle, the proposed algorithm and prediction procedure can generate RUL prediction that lie within 10 cycles of the true battery RUL. Our prediction tracks closely the real test cycle data, and performance measurement RE lies within 8% for all 4 packs.

4 Conclusions

Energy storage systems are expected to play a key role in energy systems and may help DNOs address the challenges introduced by the development of renewable resources. Li-ion batteries have emerged as a promising technology for future grids, where independent storage providers may offer DNOs several grid services. PHI is required for successful battery management and efficient operation. This work utilises a machine learning technique, RVM, to accurately predict the RUL of Li-ion batteries. The algorithm developed is a data-driven method tested on a NASA battery data set. The experiment shown in this work can predict the RUL of the battery with great accuracy, sufficient for power grid operation. In the future we plan to develop similar methods for other types of battery storage and explore different machine learning techniques and data analytics to improve the battery management and useful lifetime.

5 Acknowledgments

The authors would like to acknowledge the support of the InnovateUK HyFES project [102432]. We also thank NASA for providing access to the battery dataset.

6 References

1 Wade, N.S., Taylor, P.C., Lang, P.D., *et al.*: 'Evaluating the benefits of an electrical energy storage system in a future smart grid', *Energy Policy*, 2010, **38**, (11), pp. 7180–7188
2 Andoni, M., Robu, V., Fruh, W.-G.: 'Game-theoretic modeling of curtailment rules and their effect on transmission line investments'. IEEE PES Innovative Smart Grid Technologies Europe 2016, Ljubljana, 2016
3 2009 HM Government: 'The UK low carbon transition plan: national strategy for climate and energy' (London, 2009)
4 Han, X., Ji, T., Zhao, Z., *et al.*: 'Economic evaluation of batteries planning in energy storage power stations for load shifting', *Renew. Energy*, 2015, **78**, (78), pp. 643–647
5 Mehr, T.H., Masoum, M.A., Jabalameli, N.: 'Grid-connected Lithium-ion battery energy storage system for load leveling and peak shaving'. Power Engineering Conf. (AUPEC), 2013
6 Nourai, A., Kogan, V.I., Schafer, C.M.: 'Load leveling reduces T & D line losses', *IEEE Trans. Power Deliv.*, 2008, **23**, (4), pp. 2168–2173
7 Sparancino, A.R., Reed, G.F., Kerestes, R.J., *et al.*: 'Sparacino AR, reed GF, KerestesSurvey of battery energy storage systems and modeling techniques'. IEEE Power and Energy Society General Meeting, 2012
8 Cho, J., Jeong, S., Kim, Y.: 'Commercial and research battery technologies for electrical energy storage applications', *Progr Energy Combust. Sci.*, 2015, **48**, pp. 84–101
9 Ehsan, R., Sepasi, S., Roose, L.R., *et al.*: 'Energy management at the distribution grid using a Battery Energy Storage System (BESS)', *Int. J. Electr. Power Energy Syst.*, 2016, **77**, pp. 337–344
10 Bennett, C.J., Stewart, R.A., Lu, J.W.: 'Development of a three-phase battery energy storage scheduling and operation system for low voltage distribution networks', *Appl. Energy*, 2015, **146**, (146), pp. 122–134
11 Ildbi, B., von Appen, J., Kneiske, T., *et al.*: 'Cost-benefit analysis of battery storage system for voltage compliance in distribution grids with high distributed generation'. Energy Proc. 10th Int. Renewable Energy Storage Conf. IRES, Dusseldorf, 2016
12 Si, X.-S., Wang, W., Hu, C.-H., *et al.*: 'Remaining useful life estimation–A review on the statistical data driven approaches', *Eur. J. Oper. Res.*, 2011, **213**, (1), pp. 1–14
13 Chen, C., Pecht, M.: 'Prognostics of lithium-Ion batteries using model-based and data-driven methods'. IEEE Prognostics and System Health Management (PHM), 2012
14 Nuhic, A., Terzimehic, T., Soczka-Guth, T., *et al.*: 'Health diagnosis and remaining useful life prognostics of Lithium-Ion batteries using data-driven methods', *J. Power Sources*, 2013, **239**, pp. 680–688
15 Tipping, M.E.: 'Sparse bayesian learning and the relevance vector machine', *J. Mach. Learn. Res.*, 2001, **1**, pp. 211–244
16 Tang, W., Flynn, D., Robu, V.: 'Applying machine learning to predict asset health in hybrid maritime vessels'. 13th Int. Conf. on condition monitoring and machinery failure prevention technologies, Paris, 2016