Advances in Applied Energy 4 (2021) 100065

Contents lists available at ScienceDirect

Advances in Applied Energy

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

Design and validation of synthetic duty cycles for grid energy storage dispatch using lithium-ion batteries

Kevin Moy[a], Seong Beom Lee[a], Stephen Harris[b], Simona Onori[a,*]

[a] Department of Energy Resources Engineering, Stanford University, 367 Panama Street, Stanford, CA 94305, United States
[b] Energy Storage Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States

ARTICLE INFO

Keywords:
Energy storage
Duty cycle
K-means clustering
Principal component analysis
Lithium-ion battery
Grid application

ABSTRACT

Energy storage systems (ESSs) are a critical component of the electric grid, dispatching (charging and discharging) to performing grid applications such as frequency regulation, energy arbitrage, and peak shaving. Today, lithium-ion batteries are considered the best electrochemical ESSs in grid applications, and various cathode chemistries have been developed. On the electric grid, batteries can provide up to 13 ESS services, and different combinations of grid services and chemistries produce different battery aging and life performance under the given dispatch. Therefore, the characterization of each grid application dispatch can give an insight into optimal participation strategies for lithium-ion chemistries for each grid service. In this paper, an efficient algorithm is presented which uses a dispatch interval matrix to capture metrics in the ESS dispatch relevant to lithium-ion battery aging and performance, and implements unsupervised learning and dimensionality reduction on this matrix to produce characteristic duty cycles of the dispatch, from which synthetic duty cycles are produced that are suitable for laboratory testing and fast simulation. The algorithm is demonstrated for the dispatch under the grid application of peak shaving. Finally, an electrochemical-aging model is used to simulate a lithium-ion battery under both the original power dispatch and the synthetic duty cycle to validate the effectiveness of the method proposed in this paper to retain the operating stress factors.

1. Introduction

In 2019, worldwide renewable generation capacity reached over 2 terawatts, with the fastest-growing resources of solar photovoltaic and wind power accounting for 90% of all net renewable capacity additions [21]. These resources can comprise a substantial amount of the grid generation power. Variable renewable generation (i.e. wind and solar power) constituted over 22% of the total energy generated on the California Independent System Operator (CAISO) grid in 2018 [45]. The introduction of these variable generation resources poses challenges to conventional methods for planning the daily operation of the electric grid. For example, diurnal availability of solar generation can aggravate generation ramping problems when load power consumption increases as solar power production decreases. This behavior, termed the “duck curve”, impedes further deployment of variable renewable generation at scale [2].

Energy storage systems (ESSs) are considered as a solution to address the aforementioned drawbacks of variable renewable generation. ESSs connected to the electric grid can participate in grid applications, such as peak shaving, frequency regulation, solar firming, and voltage support, offsetting the variability of renewable generation and maintaining grid stability. Among ESSs, electrochemical energy storage is the main technology deployed today for grid services, of which over 90% is provided by lithium-ion batteries (LIBs) [47]. The modular design of LIBs allows for flexibility in construction and installation, and during operation, LIBs can respond quickly to meet power requests from various grid services [6].

However, the degradation resulting from participation in grid applications has a strong impact on profitable LIB operation [3,48]. This degradation results from both cycle aging and calendar aging. Cycle aging occurs when batteries are actively dispatched at non-zero power. In LIBs, capacity fade caused by cycling occurs due to unwanted side reactions, such as solid-electrolyte interphase (SEI) layer formation and growth, intercalation-induced stresses, and for graphite-anode LIBs, lithium plating at the negative electrode [37]. These side reactions are dependent on the operating temperature, state of charge, and C-rate, defined as the ratio of power dispatch to nominal energy. On the other hand, calendar aging accounts for all degradation processes that result independent of dispatch, and therefore occurs when the batteries are at rest. They are exacerbated by high temperatures and high state of charge at which the battery is stored. Calendar aging is an important factor in

* Corresponding author.
E-mail addresses: kmoy14@stanford.edu (K. Moy), sblee84@stanford.edu (S.B. Lee), sjharris@lbl.gov (S. Harris), sonori@stanford.edu (S. Onori).

https://doi.org/10.1016/j.adapen.2021.100065
Received 10 August 2021; Received in revised form 22 August 2021; Accepted 22 August 2021
Available online 26 August 2021

Nomenclature

Symbol	*quantity*
X	dataset matrix, dimension $\mathbb{R}^{m\times n}$
$(X)_{q,*}$	q th row of X
m	number of observations of dataset matrix X
n	number of variables of dataset matrix X
μ_i	mean of column i of X, dimension $\mathbb{R}^{n\times 1}$
A	column mean-centered matrix of X, dimension $\mathbb{R}^{m\times n}$
C	sample variance matrix of X, dimension $\mathbb{R}^{n\times n}$
σ_i	singular value i of C, in order of decreasing value
v_i	principal component i of X, corresponding to σ_i, dimension $\mathbb{R}^{p\times 1}$
V	matrix of principal components, dimension $\mathbb{R}^{n\times n}$
Σ	diagonal singular value matrix of C. dimension $\mathbb{R}^{n\times n}$
X_p	p-dimension matrix representation of X, dimension $\mathbb{R}^{m\times p}$
F_p	fraction of original variation of X retained by X_p
$F_{\min}$	minimum fraction of retained variation
p^*	optimal number of principal components for dimensionality reduction
P	reduced-dimension matrix representation of X, dimension $\mathbb{R}^{m\times p^*}$
k	number of clusters used in *k-means* clustering
x_i	observation vector i, dimension $\mathbb{R}^m$
c_j	cluster centroid j, dimension $\mathbb{R}^m$
N_c	optimal number of clusters
$dist_{within}(k)$	maximum cluster spread, as function of k
$dist_{between}(k)$	minimum pair-wise distance between cluster centroids, as function of k
$k_{\max}$	maximum number of clusters used to determine N_c
$g[b]$	sequence of values in time indexed by b, of length B
$G[c]$	discrete Fourier Transform of $g[b]$ indexed by c, of length B
P_G	periodogram of $G[c]$, of length B
f	frequency resolution of $g[b]$
W_{nom}	rated power of the ESS
E_{nom}	rated energy capacity of the ESS
$W(t)$	power dispatch for the ESS dataset, as a function of time t
$SOE(t)$	state of energy for the ESS dataset, as a function of time t
$T(t)$	outdoor air temperature for the ESS dataset, as a function of time t
d	interval used for characterization algorithm
N_d	number of non-zero dispatch intervals
M	interval metrics matrix, dimension dimension $\mathbb{R}^{N_d\times 12}$
w_j	power dispatch for representative interval j
R_Q	RMSE between capacity trajectories
$Q_{syn}(t)$	capacity trajectory under synthetic duty cycle cycling, as a function of time t
$Q_{disp}(t)$	capacity trajectory under power dispatch cycling, as a function of time t

grid applications when the resting period is substantial or not negligible compared to dispatch periods [42].

There are two main factors that affect the degradation trajectory in LIBs. The first one is the charging/discharging pattern of duty cycles typical of grid applications. A duty cycle is a power or current profile representing the battery charge and discharge operation in response to the grid application dispatch. Each grid application imposes different duty cycles on LIBs, leading to different capacity fade trends which can significantly impact their durability and operation lifetime [9]. For example, for a given battery chemistry, batteries that have been routinely charged with a relatively low C-rate will undergo a less detrimental degradation trajectory with less cycle aging, and therefore will last longer than LIBs which have experienced frequent fast-charging. Fast-charging can cause various degradation modes triggered by side reactions, including thermal effects and graphite-anode lithium-plating phenomena.

The second factor determining degradation comes from the cathode chemistry. Nickel-manganese-cobalt oxide (NMC), Lithium-iron phosphate (LFP), and nickel-cobalt-aluminum oxide (NCA) are examples of cathode chemistries which can be adopted for grid applications. For the same charging/discharging pattern, NCA and LFP typically have higher specific power/energy and maintain a longer lifespan compared to NMC [16,43]. However, NCA is known to be less safe due to thermal runway effects triggered by the exacerbated heat release of oxygen and overcharging, and is more predisposed to short circuiting than other chemistries [9]. Overall, LIBs would experience different degradation trajectories depending on combinations of the charging/discharging pattern and lithium-ion chemistries.

This paper focuses on the characterization of the grid-specific duty cycles, which addresses LIB degradation in two aspects. First, the characterization of grid applications systematically provides dispatch information affecting LIB performance and lifetime performance. Second, synthetic duty cycles to mimic grid-specific LIB dynamic behaviors can be created based on the characterization of grid duty cycles. The generation of synthetic duty cycles has the potential to facilitate laboratory testing to identify battery aging trajectories while maintaining the same characteristics of grid duty cycles. In addition, understanding the degradation mechanisms triggered by characteristic grid-specific duty cycles is key to developing predictive LIB models that can be integrated into cost/benefit analyses and real-time control strategies, in order to maximize profit and minimize lost capacity for LIBs.

A wealth of literature has been produced on this topic for automotive batteries, including the production and application of synthetic duty cycles for laboratory testing of lithium-ion cells used in hybrid vehicles [28], as well as the characterization of battery aging in plug-in hybrid vehicles using physics-based models [14]. However, current studies in renewable grid applications show the lack of a systematic approach to characterize grid-specific duty cycles. Sandia National Laboratory has implemented a testing protocol to assess the performance of ESSs using duty cycles obtained for various grid applications, including peak shaving, frequency regulation, photovoltaic (PV) smoothing, and solar firming [7]. For PV smoothing, ESS duty cycles were generated directly from existing PV generation profiles, without identification of characteristic duty cycles [41]. For frequency regulation, duty cycling was analyzed via the Fourier transform, and “aggressive” and “average” days were extracted from the dispatch and used as characteristic duty cycles [39]. The *ad hoc*, subjective construction of these duty cycles limits their general application beyond such a test-bed study. In other studies, performance and life of LIBs were investigated, by modelling LIB degradation and performance from simulated grid application dispatch. In [44], the aging of LIBs was modelled under frequency regulation given different dispatch methodologies (e.g. re-establishing LIB state of charge between duty cycles), and in [48], the effects of different degradation models on the optimal energy arbitrage revenue were demonstrated. However, the analysis of the impact of dispatch on degradation is limited by the lack of rigorous characterization of the dispatch. The literature depends on arbitrarily-selected days of dispatch to demonstrate modelling results, which do not adequately capture features of the dispatch that explain the LIB degradation trajectories.

Our previous work on characterizing duty cycles for grid ESSs formulated a systematic approach using *k-means* clustering and spectral analysis on the power dispatch, and included a preliminary method of creating synthetic duty cycles [31,32]. This paper follows the previous approach by including additional state of energy and ambient temperature information along with the power dispatch of the ESS into metrics representing features in grid service operation that are relevant to LIB

degradation. The proposed characterization algorithm presented in this paper uses principal component analysis (PCA) to improve the performance of *k-means* clustering and a more rigorous method to determine the number of clusters for *k-means* clustering to produce characteristic duty cycles, and subsequently synthetic duty cycles, for grid ESS dispatch. This paper also presents the validation of the synthetic duty cycles using an electrochemical aging model, demonstrating that the simulated capacity fade trajectory of a cell cycled with the synthetic duty cycles is the same as that of a cell cycled with the real dispatch profile. To the best of our knowledge, this paper is the first to present such validation.

There are seven sections in this paper: Peak Shaving, Methods, Data Description, Dispatch Interval Matrix, Algorithm for Duty Cycle Characterization, Results and Discussion, and Conclusions. In the Peak Shaving section, we discuss the motivation for peak shaving, and how LIBs are used to support this grid application. In the Methods section, we describe PCA, *k-means* clustering, and the fast Fourier transform (FFT), and their relevance to duty cycle characterization. In the Data Description section, we present the dispatch data used for the characterization and synthesis of duty cycles. In the Dispatch Interval Matrix section, we describe the construction of a matrix consisting of metrics relevant to lithium-ion battery calendar and cycle aging from the dispatch data. In the Algorithm for Duty Cycle Characterization section, we present the method for duty cycle characterization using the dispatch interval matrix. In the Results and Discussion section, we apply the algorithm to different dispatches, compare the resulting characteristic duty cycles, present synthetic duty cycles as constructed from characteristic duty cycles, validate the aging characteristics of the synthetic duty cycles against those of the original dispatch, and discuss the applicability of this approach to general grid ESS dispatch. Concluding remarks are found in the Conclusion section.

2. Peak shaving

In the stationary grid market, LIBs can provide up to 13 different services, where each of them would require different energy and power requirements [19]. In a back-up power application, for example, the battery system must provide energy quickly upon users' need in emergency situations, which would call for batteries with high specific power; on the other hand, in the daily charging and discharging residential photovoltaic/ESS application, the LIB must provide a more stable charging and discharging performance, requiring a battery with high cycle durability [43].

Among these grid services, peak shaving is most commonly performed by residential (household) LIBs, which are rapidly growing on the grid. Between 2014 and 2018, annual installations of residential LIBs in the United States increased from 2.25 MWh to 185 MWh, and are projected to exceed 2900 MWh by 2032 [18]. The basic concept of peak shaving is rooted in the conventional electric grid operation, whereby the electric utility supplies electricity via the utility distribution grid to meet the load power demand of the consumer, e.g., house, office, or factory. The consumer is then billed monthly by the utility for this supplied electricity. The utility tariff rate determines this bill based on electricity costs such as time-of-use energy charges and power demand charges, which are applied to the electric energy and peak (maximum) power supplied by the grid. These quantities are measured by the flow of energy and power through the utility meter.

Peak shaving is used to lower the monthly peak power supplied by the grid ("shaving" the peak), while still keeping time-of-use energy charges low. This reduces the power flow measured by the utility meter, and therefore lowers the utility bill for the consumer. Different strategies for peak shaving have been demonstrated, including reducing peak consumption of flexible facility loads, managing charging of electric vehicles, and dispatching battery ESSs [4,46].

For ESSs, peak shaving is accomplished by managing the dispatch of the ESS according to both the load and the tariff rate: discharging when the load is large, and charging from the grid when electricity is cheap [17]. This grid application has a large market size, with approximately 5 million potential residential ESS peak shaving systems in the United States alone [30]. Real-time operation of an ESS for peak shaving can involve simple control loops to discharge or charge the ESS based on power flow from the grid and state of charge of the battery [25], or include an optimization algorithm to compute optimal peak shaving dispatch given additional constraints, such as the power flow through grid infrastructure [36] or the utilization of local solar PV generation [33]. In the system shown in Fig. 1, the ESS can perform peak shaving by dispatching to offset the load demand.

3. Methods

This section presents the fundamental methods used in the duty cycle characterization. PCA is used to reduce the dimension of the dataset of metrics computed for the ESS dispatch, which is then clustered according to *k-means* clustering. In addition, the Fast Fourier Transform is used to include frequency information from the dispatch into the metrics.

3.1. Principal component analysis

PCA is used to reduce the dimensionality of a dataset consisting of a large number of interrelated variables by first obtaining orthogonal vectors representing new, uncorrelated variables for the dataset, which are termed principal components. Dimensionality reduction then takes place by projecting the dataset along a subset of these principal components, while retaining as much as possible of the variation (i.e. the spread of the data within each variable) originally present in the dataset [23]. By representing the dataset of m observations with n variables as a real-valued matrix $X \in \mathbb{R}^{m\times n}$, PCA produces the principal components $v_1, v_2, \dots v_n$ as vectors in $\mathbb{R}^n$, with corresponding singular values $\sigma_1, \sigma_2, \dots \sigma_n$ ordered in decreasing magnitude (i.e. $\sigma_1 > \sigma_2 > \dots > \sigma_n$). The details of this process can be found in Appendix A.

Next, X is projected along a subset of its principal components $v_1, \dots, v_p$, $p < n$, reducing X from dimension n to dimension p. This is performed by multiplying the matrix X by the matrix of vectors $v_1, \dots, v_p$ to produce the PCA subspace matrix X_p

$$X_p = X[v_1, \dots, v_p], \quad X_p \in \mathbb{R}^{m\times p} \tag{1}$$

The optimal number of principal components p^* is determined using the retained variation of X present in X_p [23]. F_p is defined as the variation of X_p over the original variation of X

$$F_p = \frac{\sum_{i=1}^{p} \sigma_i}{\sum_{j=1}^{n} \sigma_j}, \tag{2}$$

Using the singular values $\sigma_1, \dots \sigma_p$ corresponding to the principal components used to construct X_p to represent the variation of X_p, and all singular values to represent the original variation of X. p^* is then defined as the minimum number of components needed for F_p to exceed a minimum threshold $F_{\min}$ by ensuring that a majority of the variation of X is retained in X_p with a minimal number of principal components

$$p^* = \arg\min_p \{F_p > F_{\min}\}, \quad p = 1, 2, \dots, n \tag{3}$$

The value for $F_{\min}$ is dependent on the application and the resulting trade-off between accuracy and computational effort. In this paper, $F_{\min} = 0.9$. P is then the projection of X onto the first p^* principal components

$$P = X[v_1, \dots, v_{p^*}], \quad P \in \mathbb{R}^{m\times p^*}, \tag{4}$$

Which represents the optimal PCA subspace for X, and is used as the input for *k-means* clustering in this paper.

3.2. K-means *clustering*

Clustering is a class of unsupervised learning methods that organize and partition data into groups, for which members of a group share some features [22]. Clustering has been widely employed for characterizing

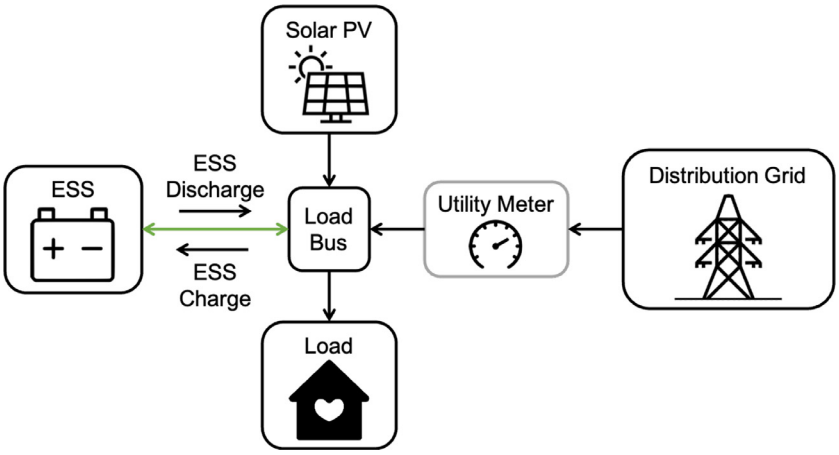


Fig. 1. Power flow diagram for a system that can use an ESS for peak shaving. A load bus electrically connects the utility distribution grid, the solar PV array, the load, and the ESS, allowing power to flow between them. Arrows indicate direction of power flow. Power flows into the load bus from the utility distribution grid, and then flows out to the load, and the direction is unchanged by the dispatch of the ESS. The bidirectional dispatch of the ESS is highlighted in green, with direction of ESS charge and ESS discharge also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

time-series data [27]. In [10], *k-means* clustering was applied to a yearly solar generation profile to provide a reduced set of daily power profiles for further voltage stability analysis, and in Xu and Zhang [52] wind speed and insolation data clusters were used to construct multiple scenarios of renewable energy resources for a combined concentrated solar power/wind turbine plant. In [20], clusters were used to generate representative electricity load demand profiles for further sensitivity analyses, and in Rhodes et al. [38], seasonal trends were demonstrated within electricity usage clusters, which were then used to drive a regression model between consumers and electricity consumption. In [13], *k-means* clustering was adopted as a technique to study driving cycles for electric vehicles, comparing clusters of charge/discharge pulses to the road operating conditions during these pulses.

In *k-means* clustering, k is defined as the number of clusters to be constructed from m observations of data. Each observation is assigned a cluster by the closest Euclidean distance to the cluster centroid, defined as the mean of all observations in the cluster. Given k desired clusters, and dataset of observation vectors $x_1, \dots, x_m$, *k-means* clustering determines k cluster centroids as a solution $\phi^* = \left[c_1, \dots, c_k\right]^T$ that minimizes the cost function J [1]

$$J = \sum_{i=1}^{m} \min_{j=1,\dots,k} \left\| x_i - c_j \right\|^2 \tag{5}$$

Along with ϕ^*, the results of *k-means* clustering are the cluster assignments of each observation. Each cluster j consists of its centroid c_j and its observation members. Additionally, m_j is defined as the number of observation members for cluster j, where $\sum_i^k m_j = m$ (as every observation is assigned to a cluster).

Among all choices for k, selecting the optimal number of clusters N_c is an important step in the application of *k-means* clustering, as it entails the generation of clusters that are distinct from each other. The position of the cluster centroids and their respective observation members within this subspace can thus be used to assess the output of *k-means* clustering, and ultimately obtain N_c.

First, $dist_{within}(k)$ is defined as the maximum average distance between the cluster centroid c_j and all m_j of the cluster members

$$dist_{within}(k) = \max \left\{ \frac{1}{m_j} \sum_{x_i \in \text{cluster } j} \left\| x_i - c_j \right\|^2, \quad j = 1, 2, \dots k, \right\} \tag{6}$$

Representing the maximum spread among all clusters. Clusters that are tightly-packed together, where all cluster members are close to their respective cluster centroids, will produce a small $dist_{within}(k)$, while clusters that are spread out, where cluster members are far from their respective cluster centroids will produce a large $dist_{within}(k)$.

However, the relative positioning of each cluster must be taken into consideration. Because the cluster centroids represent the average position of each cluster, $dist_{between}(k)$ is then defined as the minimum pairwise distance between each pair of cluster centroids

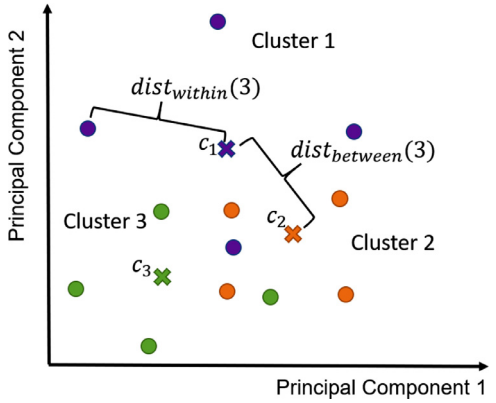


Fig. 2. Illustrative example of clustering on observations in a PCA subspace of two principal components, with number of clusters $k = 3$. Circles represent observations (i.e. rows of X_r), while the cluster centroids c_i are labeled and denoted with crosses. Color is used to represent membership to a given cluster (e.g. all green circles are observations that have been assigned to the cluster with cluster centroid denoted by the green cross). Because $dist_{between}(3) < dist_{within}(3)$, the clusters are indistinguishable from each other. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$dist_{between}(k) = \min \left\{ \left\| c_a - c_b \right\|^2, \quad a \in 1, \dots, k, b \in 1, \dots, k, a \neq b \right\} \tag{7}$$

Clusters which are close to each other will produce a small $dist_{between}(k)$, while clusters which are far apart will produce a large $dist_{between}(k)$. Note that if $dist_{between}(k) < dist_{within}(k)$, clusters are more spread out than they are distanced from each other, and they can overlap and become indistinguishable from each other. An example of this effect is shown in Fig. 2.

In order to produce clusters that are distinct from each other *k-means* clustering will generate clusters that are tightly-packed and far apart from each other. The optimal number of clusters N_c is therefore the value of k which maximizes the difference between $dist_{between}(k)$ and $dist_{within}(k)$

$$N_c = \arg\max_k \left\{ dist_{between}(k) - dist_{within}(k) \right\}, \quad k = 1, 2, \dots, k_{\max}, \tag{8}$$

where the optimal number of clusters N_c are determined up to a maximum number of clusters $k_{\max}$. In this paper, $k_{\max} = 30$.

An example of ideal clustering compared to non-ideal clustering is shown in Fig. 3, in the case where a dataset is projected onto two principal components ($r = 2$), and shown for $k = 3$ and $k = 2$. Because the out-

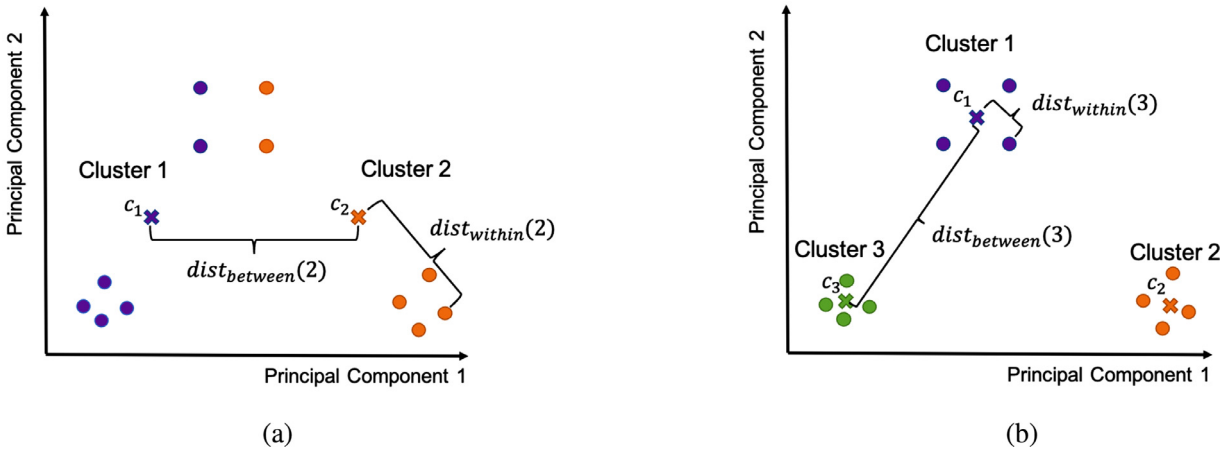


Fig. 3. Illustrative example of clustering on observations in a PCA subspace of two principal components. Circles represent observations (i.e. rows of X_r), while the cluster centroids c_i are labeled and denoted with crosses. Color is used to represent membership to a given cluster (e.g. all green circles are observations that have been assigned to the cluster with cluster centroid denoted by the green cross). (a) For $k = 2$, $dist_{between,k} - dist_{within,k}$ is small, i.e. clusters are loosely spread but close to each other. (b) For $k = 3$, $dist_{between,k} - dist_{within,k}$ is large, i.e. clusters are tightly-packed and far apart from each other. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

put for three clusters produces the greatest difference between $dist_{between}$ and $dist_{within}$, $N_c = 3$ for this dataset in the PCA subspace.

3.3. Fast Fourier transform

In this paper, the fast Fourier transform is used to incorporate frequency information in the power dispatch as metrics for the duty cycle characterization. The FFT is used to compute the Discrete Fourier Transform of a signal represented by a time-series sequence of values $g[b], b = 0, 1, 2, \ldots, B-1$ of length B, in time. This sequence $g[b]$ is then decomposed into a sequence of sinusoidal components $G[c], c = 0, 1, 2, \ldots, B-1$ of length B, of different frequencies bc/B [8]

$$G[c] = \sum_{b=0}^{B-1} g[b]^{-\frac{i2\pi bc}{B}}, \quad c = 0, 1, \ldots B-1 \tag{9}$$

From $G[c]$, one can obtain the corresponding periodogram P_G, which is an estimate of the spectral density of $g[b]$, i.e. the distribution of power among the frequency components of the signal. P_G is defined as follows

$$P_G = \frac{1}{Bf}\left|G[c]\right|^2, \tag{10}$$

With f as the known frequency resolution of $g[b]$. Peaks in the periodogram correspond to the dominant (high-power) frequencies within the signal.

The degradation of a battery is dependent on different mechanisms for charging and discharging. For example, SEI layer growth is one of the main causes of capacity fade and impedance increase in the LIB, and occurs predominantly on the anode during charging [35]. Therefore, separating dispatch into charge and discharge profiles could yield useful insights on potential modes of degradation upon usage. However, charging and discharging are strictly negative and strictly positive, respectively. As the FFT decomposes a signal along zero-mean sinusoidal components, simply separating the charge and discharge and directly taking the FFT of these profiles would yield a bias in the periodogram due to the non-zero mean of each separated profile, appearing as low-frequency components not present in the original dispatch.

Mean centering is applied to ensure that the resulting separated charge and discharge profiles have zero mean, while preserving the frequency components of these profiles. This procedure has previously been applied to analyze duty cycles in hybrid automotive applications [29], and is described for obtaining mean-centered discharge (or charge) profiles, below.

1. Construct a new profile consisting of only non-negative (non-positive) dispatch.
2. Find each individual discharge (charge) instance within this new profile, where a discharge (charge) instance is defined as the battery starting at 0 dispatch, positively (negatively) dispatching, then returning to 0 dispatch.
3. Construct a new mean-centered discharge (charge) profile by concatenating each discharge (charge) instance with a copy of reversed sign (i.e. the same sequence of values as the original instance, where positive values are exchanged for negative ones of the same magnitude, and vice versa).

4. Data description

In this work, three datasets are considered, each representing the operation of an ESS for three different facility loads. These datasets are labelled by the load profiles used to generate them: namely, LargeOfficeNew, SuperMarketNew, and SecondarySchoolNew. Each dataset contains a power dispatch, in units of kW; the state of energy (SOE) of the energy storage, normalized; and the outdoor air temperature, in degrees Celsius. The three profiles are sampled at hourly intervals over one year.

4.1. Power dispatch

For each yearly dataset, the power dispatch $W(t)$ represents ESS operation, where positive values indicate ESS discharge, and negative values indicate ESS charge. The power dispatch is obtained from QuESt, an open-source, freely-available simulation tool for ESS developed by Sandia National Labs [40]. QuESt allows users to select facility load profiles (the hourly consumption of power by the facility), ESS parameters such as rated power and energy, and the tariff rate structure, in order to simulate the ESS operation for each month over one year by optimizing the dispatch for peak shaving.

The power dispatches considered in this paper are obtained from QuESt by simulating the operation of an ESS in San Jose, California with rated power $W_{nom} = 200$ kW and rated energy capacity $E_{nom} = 400$ kWh, on each load profile. The tariff rates were selected appropriate to the peak load, i.e., ensuring that the peak facility load does not exceed the maximum load power specified in the tariff. The power dispatches for peak shaving on each load profile are shown in Fig. 4.

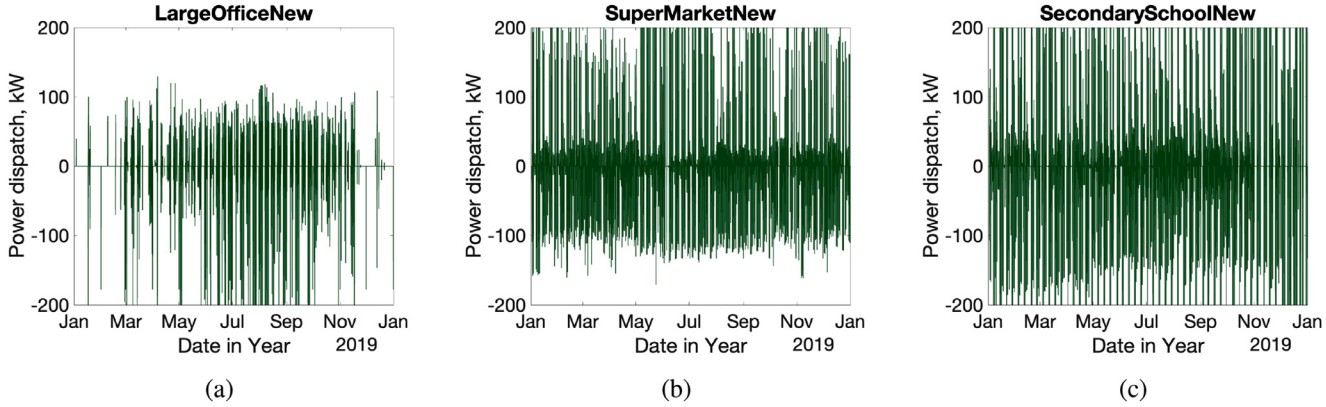


Fig. 4. Power dispatch $W(t)$ of a 200 kW, 400 kWh ESS on: (a) the LargeOfficeNew, with tariff rate E19; (b) the SuperMarketNew load, with tariff rate A10; and (c) the SecondarySchoolNew load, with tariff rate E19. Positive values indicate discharging, while negative values indicate charging. Load profile data and subsequent calculated ESS power dispatch data were obtained from QuESt [40]. Tariff rate data were obtained from Pacific Gas & Electric tariff [34], as accessed via QuESt. Temperature data were obtained from TMY3 weather station data [50].

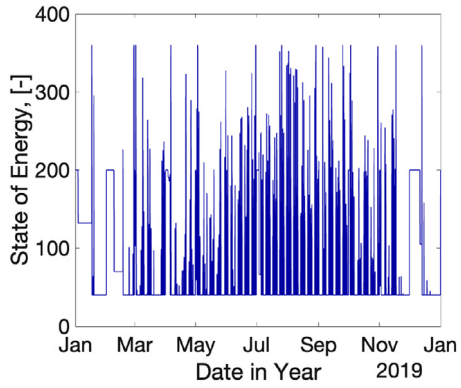


Fig. 5. State of energy $SOE(t)$ of a 200 kW, 400 kWh ESS on the LargeOfficeNew load. This SOE is obtained from the power dispatch $W(t)$ in Fig. 4a, discretized hour by hour.

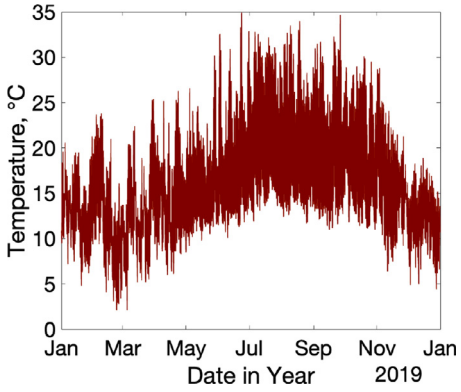


Fig. 6. Outdoor air temperature $T(t)$ obtained from San Jose airport TMY3 weather data, used as the temperature for the dispatch over the year 2019, discretized hour by hour, for all datasets in this paper.

4.2. SOE of energy storage

The state of energy $SOE(t)$ for each yearly dataset is also a direct output from QuESt, and is derived from the power dispatch. The SOE provides a normalized metric for the available dispatchable energy stored by the ESS. Given the initial SOE, $SOE(0)$, $SOE(t)$ can be calculated as:

$$SOE(t) = SOE(0) - \frac{1}{E_{nom}} \int_0^t W(\tau) d\tau \tag{11}$$

In QuESt, for all datasets, the initial SOE is configured as $SOE(0) = 0.5$ (i.e. 50% of rated energy capacity, or 200 kWh). Furthermore, the dispatch is configured such that the battery is returned to $SOE(0)$ at the beginning of each month. This is the default constraint for SOE as supplied by QuESt, and is preserved in this analysis to ensure that the battery maintains safe operating conditions and does not over-charge or over-discharge across the dispatch. As an example, the state of energy for the power dispatch on the LargeOfficeNew load shown in Fig. 4a is shown in Fig. 5.

4.3. Outdoor air temperature

The outdoor air temperature data $T(t)$ were taken from TMY3 atmospheric data, which in turn was derived from weather measuring stations in the United States, as described in Wilcox and Marion [50]. This paper uses outdoor air temperature data from the San Jose Airport measuring station, shown in Fig. 6.

Key assumptions In QuESt, each power dispatch is the result of simulating ESS dispatch with full knowledge of each load profile, i.e. a deterministic optimized dispatch. The load profile data implemented in QuESt were constructed from simulated building data according to the U.S. Department of Energy Commercial Reference Buildings [12]. It is assumed that the ESS exactly performs this dispatch. Lastly, the outdoor air temperature is assumed to be the same for all facilities.

5. Dispatch interval matrix

Each dataset is pre-processed into the dispatch interval matrix M before the rest of the duty cycle characterization. First, the year is divided into intervals (e.g. one hour, one day, two weeks, etc.). Then, twelve metrics relating to the power, SOE, and temperature are computed for each non-zero power dispatch interval (i.e. when the ESS is actively dispatching), and then normalized. These metrics are listed in Table 1 and were chosen to address characteristics of the dispatch relevant to lithium-ion battery calendar and cycle aging by including the power and frequency contents of the dispatch, as well as the operating conditions (i.e. SOE and temperature) during dispatch. For non-zero dispatch days without discharge or charge, the respective metrics are set to NaN. In this paper, the interval is one day (24 h), starting from midnight (00:00) on January 1st.

The metrics in Table 1 are now described in detail and demonstrated for an example day of dispatch in the LargeOfficeNew dataset. Fig. 7 shows the power dispatch of this day, with charging periods high-

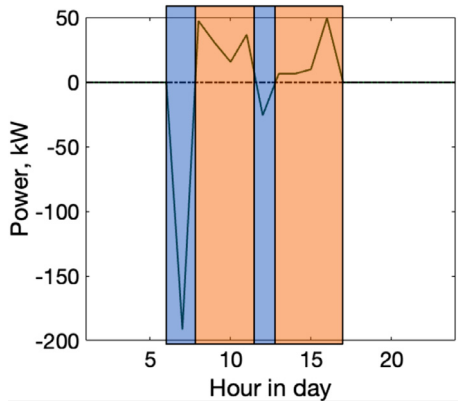


Fig. 7. Example day of power dispatch taken from the LargeOfficeNew peak shaving dispatch shown in Fig. 4a. The charging events are highlighted in blue, while the discharging events are highlighted in orange. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

lighted in blue and discharging periods highlighted in orange, which are used in the calculation of the following metrics. The actual values of this example day of dispatch are shown as the third column of Table 1.

Metrics 1 and 3 count the number of discharge/charge events, defined as the ESS dispatch starting at zero power, dispatching positively/negatively, and then returning to zero power. Metrics 3 and 4 record the absolute maximum positive/negative power dispatch of the ESS during all discharge/charge events, highlighted in orange/blue in Fig. 7. Similarly, metrics 5 and 6 record the absolute average positive/negative power dispatch of the ESS during all discharge/charge events.

For metrics 7–10, the corresponding day of SOE and temperature are also considered during the same charge and discharge events. These are shown in Fig. 8, with the discharge/charge periods highlighted in blue and orange across the power dispatch, SOE, and temperature for the example day. Metrics 7 and 8 record the average SOE of the ESS during all discharge/charge events, highlighted in orange/blue in Fig. 8. Similarly, metrics 9 and 10 record the average outdoor air temperature of the ESS during all discharge/charge events.

The process to determine the peak discharging/charging frequency in metrics 11 and 12 is demonstrated below for the peak discharging frequency for the example day of dispatch. First, only the discharge segments (highlighted in orange in Fig. 7) are retained. These segments are then concatenated and mean-centered. This mean-centered discharge profile is shown in Fig. 9a. Next, the mean-centered discharge profile is concatenated 100 times, as shown in Fig. 9b, to produce a longer signal as a input for FFT. The resulting periodogram is shown in Fig. 9c. The frequency corresponding to the peak of this periodogram is used as the peak discharging frequency.

6. Algorithm for duty cycle characterization

The characterization algorithm shown in Algorithm 1 clusters the dispatch interval matrix M to produce characteristic duty cycles. A similar approach was used in Dembski et al. [11], where vehicle driving cycles were clustered given statistical metrics of the velocity and acceleration of the driving cycles. The clustering is performed exclusively in the PCA subspace of M. PCA has previously been shown to improve the performance of *k-means* clustering, as principal components have been shown to be equivalent to the continuous solutions to the discrete cluster membership indicators [15].

The algorithm is divided into four main sections. First, lines 1–4 of the algorithm obtain the corresponding dispatch interval matrix M, as described in Section 5. Next, lines 5–7 of the algorithm perform PCA and

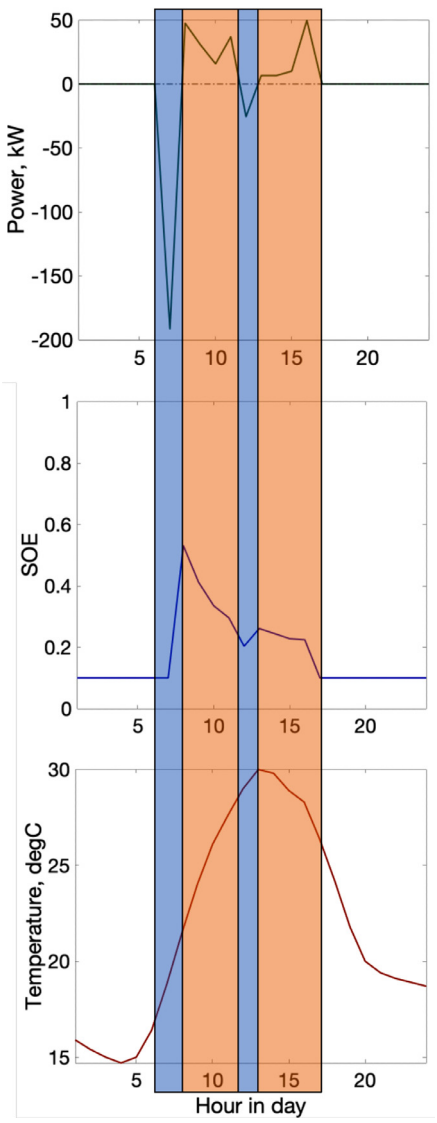


Fig. 8. Example day of dispatch, SOE, and temperature, with charging events highlighted in blue and discharging events highlighted in orange. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

dimensionality reduction as described in Section 3.1 on M. As there are 12 metrics, there are also 12 singular values and principal components of M. The PCA subspace matrix P is obtained using Eq. (4), with p^* obtained using Eqs. (2) and (3). Then, lines 8–9 of the algorithm perform *k-means* clustering on P, including the determination and application of the optimal number of clusters N_c, as described in Section 3.2. This produces the N_c cluster centroids $c_1, \ldots, c_{N_c}$ of P. *k-means* clustering on the PCA subspace matrix P, and selection of the characteristic duty cycles. Finally, lines 10–14 of the algorithm obtain and return the characteristic duty cycles from the power dispatch W of the dataset. In line 11, the representative interval d_j of cluster j is determined using P, by finding the row of P which minimizes the Euclidean distance to the cluster centroid c_j

$$d_j = \arg\min_m \left\{ \left\| (P)_{m,*} - c_j \right\| \right\}, \quad m = 1, 2, \ldots, N_d \tag{12}$$

In line 12, the corresponding row of W is used as the representative power interval dispatch w_j, which then represents the jth characteristic duty cycle. This is repeated for all cluster centroids $c_1, \ldots, c_{N_c}$ of P to obtain the N_c characteristic duty cycles of the dataset.

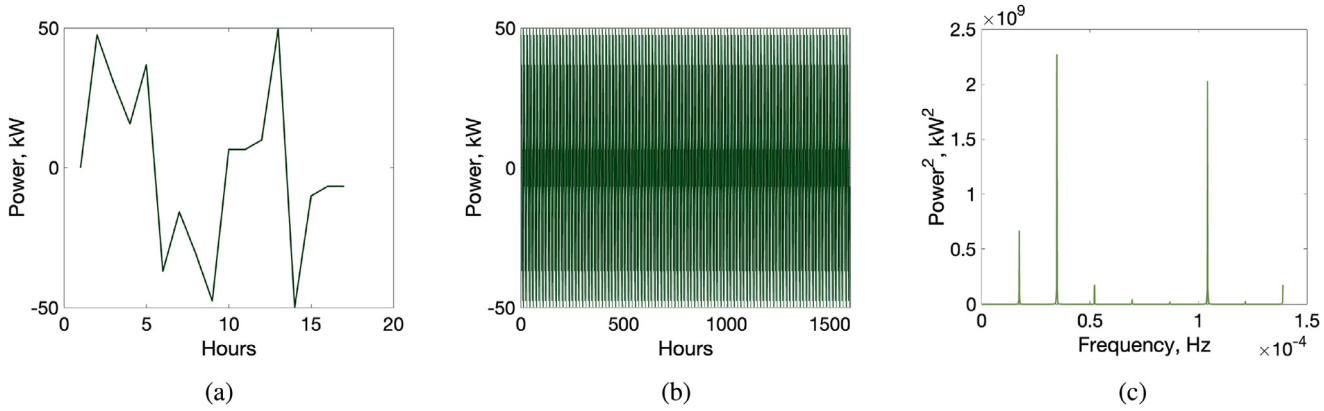


Fig. 9. Process to determine the peak discharging frequency in the example day of dispatch shown in Fig. 7. (a): The discharge segments highlighted in Fig. 7 are concatenated and mean-centered. (b): The mean-centered discharge in Figure 9a is concatenated 100 times. (c): The periodogram of Figureb. The frequency corresponding the peak of this periodogram is the peak discharging frequency.

Algorithm 1 Characterization algorithm for ESS grid application dispatch.

Require: Arrays of: power dispatch W, state of energy SOE, and outdoor air temperature T

for $d = 1, 2, \ldots, N_d$ for N_d non-zero dispatch intervals in the year **do**
  Compute the 12 metrics for interval d using $(W)_{d,*}$, $(SOE)_{d,*}$, and $(T)_{d,*}$
  Store normalized metrics for interval d in the dispatch interval matrix M
end for
Obtain principal components $v_1, \ldots, v_{12}$ and singular value $\sigma_1, \ldots, \sigma_{12}$ of M via PCA
Select number of principal components p^* using Equations 2 and~3 with $p = 1, \ldots, 12$
Obtain PCA subspace matrix P of M using Equation 4
Select optimal number of clusters, N_c using Equation 6,~7, and~8, with rows $(P)_{d,*}$ of P as the observations
Obtain cluster centroids $c_1, c_2, \ldots, c_{N_c}$ of P with number of clusters N_c via *k-means* clustering
for $j = 1, 2, \ldots N_c$ **do**
  Select representative interval d_j of cluster j using Equation 12
  Obtain representative power interval dispatch $w_j = (W)_{d_j,*}$
end for
return $w_1, w_2, \ldots, w_{N_c}$ as the N_c characteristic duty cycles

7. Results and discussion

The characterization algorithm is demonstrated in detail for the LargeOfficeNew peak shaving load. The characteristic duty cycles are then presented and discussed for all three datasets in this paper, followed by the design and formation of synthetic duty cycles for laboratory testing of the calendar and cycle aging of batteries. The synthetic duty cycles are then validated against their respective power dispatches using a physics-based electrochemical aging model. This section concludes with a brief discussion of the applicability of the characterization algorithm to additional grid services.

7.1. Detailed results of the characterization algorithm

The results of applying the characterization algorithm are presented in this subsection for the LargeOfficeNew dataset. The interval matrix M is obtained and shown in Fig. 10, where the interval is chosen as one day.

Table 1
Summary of interval metrics used in the characterization algorithm (Algorithm 1), with example pre-normalized values for each metric as demonstrated by the daily dispatch in Figs. 7 and 8.

Number	Name	Example metric value
1	Number of discharge events	2
2	Number of charge events	2
3	Peak discharge power	49.8 kW
4	Peak charge power	191.5 kW
5	Average discharge power	25.5 kW
6	Average charge power	108.6 kW
7	Average discharge SOE	0.256
8	Average charge SOE	0.396
9	Average discharge temperature	27.0 °C
10	Average charge temperature	23.9 deg C
11	Peak discharging frequency	3.5e−5 Hz
12	Peak charging frequency	1.4e−4 Hz

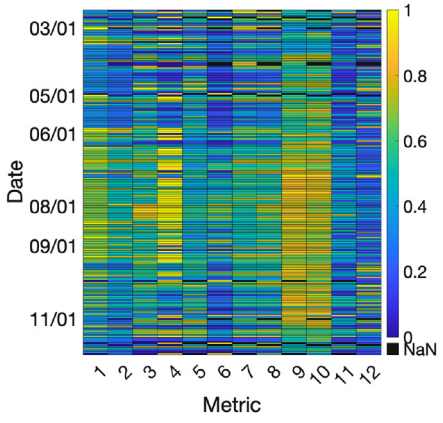


Fig. 10. Interval metrics matrix M for the LargeOfficeNew dispatch, calculated using the metrics in Table 1 for all intervals with non-zero dispatch, with the interval set to one day.

The optimal number of principal components p^* for M is then determined. Fig. 11 plots the fraction of retained variation F_p as a function of number of principal components p, as in Eq. (2). From Eq. (3), with $F_{\min} = 0.9$, the optimal number of principal components is obtained as $p^* = 5$.

The interval matrix M is then projected onto the 5 principal components $v_1, \ldots, v_5$ shown in Fig. 12a, to form the PCA subspace matrix P shown in Fig. 12b.

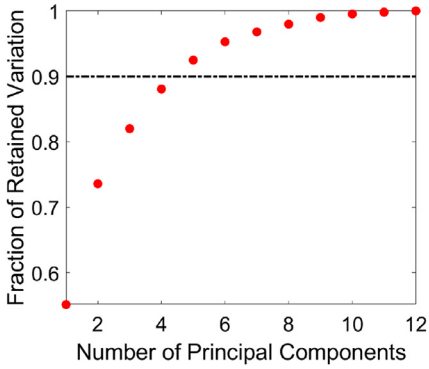


Fig. 11. Fraction of retained variation, F_p, as a function of number of principal components, p, for the interval matrix M in Fig. 10. The minimum fraction of retained variance, $F_{\min} = 0.9$, is plotted as a black dashed line. The optimal number of principal components, p^*, is the minimum number of principal components needed to exceed $F_{\min}$ ($p^* = 5$).

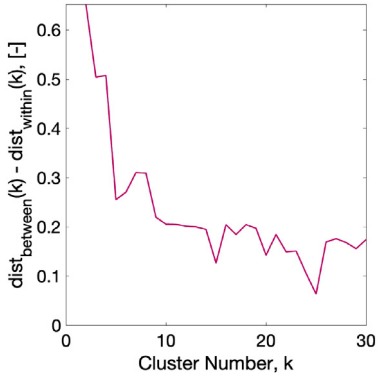


Fig. 13. Determination of optimal number of clusters, N_c, as in Eq. (8), by applying *k-means* clustering to P for different number of clusters k, up to $k_{\max} = 30$. From this plot, $N_c = 2$.

Next, the optimal number of clusters N_c is obtained by applying Eq. (8) on the PCA subspace matrix P. From Fig. 13, $N_c = 2$.

k-means clustering is then applied to P with two clusters, resulting in the two cluster centroids shown in Fig. 14a. The representative intervals of the two cluster centroids obtained are the days of September 15th and May 15th, respectively, and the power dispatch of those days is shown in Fig. 14b and c, which are the characteristic duty cycles w_1 and w_2 of the LargeOfficeNew dataset.

7.2. Characteristic duty cycles

The characterization algorithm is applied to each yearly dataset. The resulting characteristic duty cycles for each dataset are shown in Fig. 15a–c.

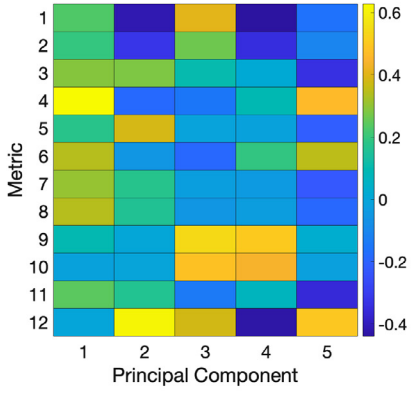
The characteristic duty cycles are indicative of the general behavior of ESSs performing peak shaving. In particular, the characteristic duty cycles show that the ESS immediately charges before the necessary discharge needed to shave that day's load consumption peaks. This is an artifact of the QuESt dispatch optimization for peak shaving, which only seeks to minimize the electric bill, without considering management of battery degradation. This presents an opportunity for a managed charging protocol for LIB in peak shaving: If the necessary discharge for shaving the load peaks can be accurately predicted, then the charge could be spread out throughout the inactive dispatch periods at a lower power, thus reducing C-rate-related aging stress on the battery. On the other hand, the characteristic duty cycles show that active dispatch occurs mainly during daytime hours, leaving the nighttime hours at the beginning and end of each day fairly inactive. These inactive periods present additional opportunity for participation in other grid applications, such as energy arbitrage, when the battery ESS is not used for peak shaving.

The characteristic duty cycles are distinct within each dataset. For example, in Fig. 15b, Duty Cycle 1 represents days with a high-power charge to support a high energy, lower power discharge throughout the day, Duty Cycle 2 represents days with both a high power charge and high power discharge, and Duty Cycle 3 represents a relatively inactive day, with a much smaller charge and discharge for the day. Each duty cycle captures different dispatching behaviors within the dataset.

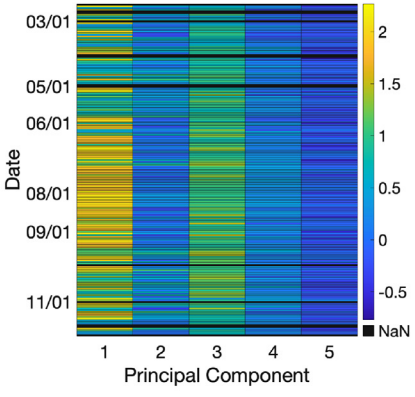
However, the characteristic duty cycles also are unique to each dataset. Those of LargeOfficeNew are different from those of SuperMarketNew, in turn different than those of SecondarySchoolNew. These differences result from the different load profile and tariff rate in each dataset, providing a venue of further study to analyze their effects on ESS dispatch.

Along with the representative interval power dispatch, which are used to form the characteristic duty cycles, the representative interval temperature can also be found. This "characteristic operating temperature" during the characteristic duty cycles for each dataset is shown in Fig. 16.

The characteristic operating temperature provides further insights towards the degradation of the battery. Comparing Figs. 15 and 16 shows that when the ESS is not actively dispatching and calendar aging dominates the degradation process, the temperature is between 10and15 degrees Celsius. This information can be used for calendar aging laboratory testing of batteries stored in this temperature range, and develop temperature-dependent calendar aging models for LIBs.



(a)



(b)

Fig. 12. (a): First p^* principal components of the interval metrics matrix M in Fig. 10, where $p^* = 5$ is the optimal number of principal components for M. (b): The PCA subspace matrix P of the interval metrics matrix M, found by projecting M onto the p^* principal components in 12 a.

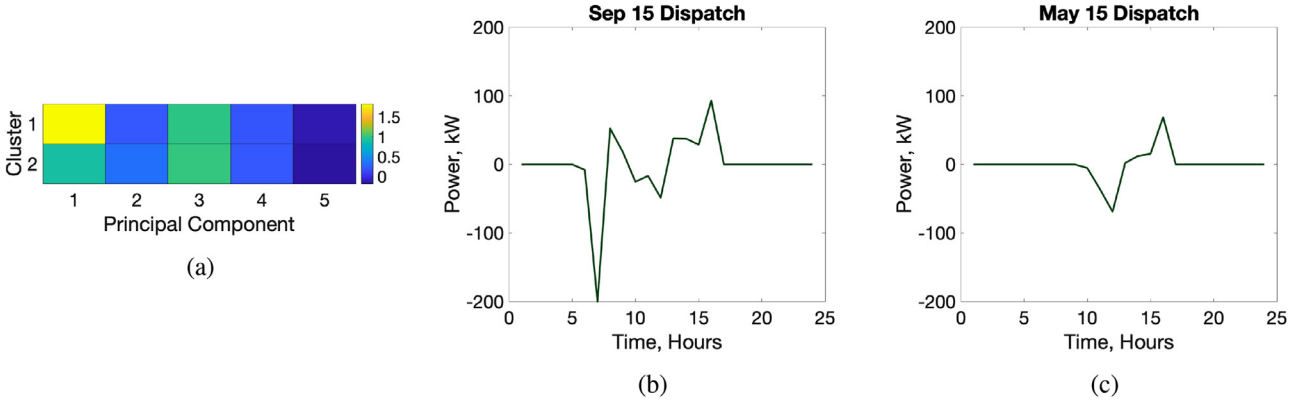


Fig. 14. (a): Cluster centroids obtained for P in Fig. 12b, with optimal cluster number $N_c = 2$. Cluster centroid 1 corresponds to the row of September 15, and cluster centroid 2 to the row of May 15, in P. (b) and (c): The characteristic duty cycles found from the cluster centroids at left, found using the power dispatch of the corresponding representative days of each cluster centroid. The characteristic duty cycles are labeled according to their representative days.

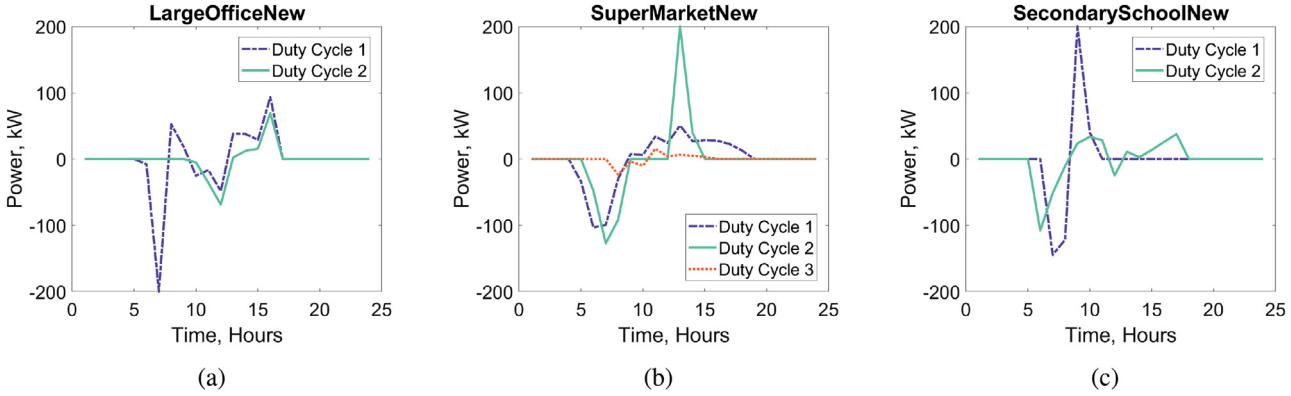


Fig. 15. Characteristic duty cycles for peak shaving dispatch on the (a) LargeOfficeNew, (b) SuperMarketNew, and (c) SecondarySchoolNew load profiles.

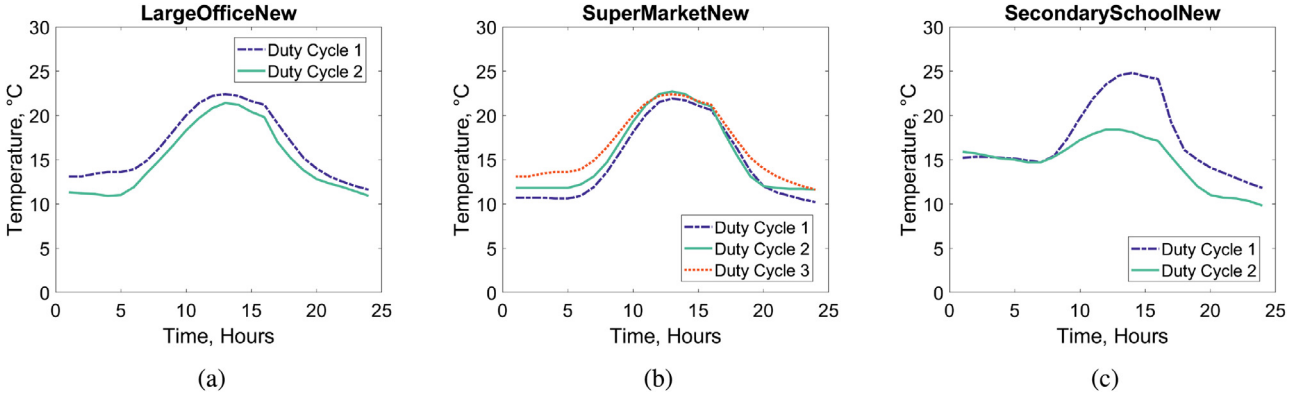


Fig. 16. Characteristic operating temperature, defined as the outdoor air temperature during the representative days used to create the characteristic duty cycles, for peak shaving dispatch on the (a) LargeOfficeNew, (b) SuperMarketNew, and (c) SecondarySchoolNew load profiles.

7.3. Synthetic duty cycles

Accurately estimating capacity fade of LIBs is key to understanding how to maximize their usability and lifetime in stationary grid applications. Testing the capacity fade of an LIB under actual grid application dispatch is time-consuming, as battery degradation occurs over multiple years. However, identification of characteristic grid duty cycles enables creation of synthetic duty cycles, which can be implemented in a laboratory environment in reasonable time.

In order to account for both the cycle aging and calendar aging processes, two different synthetic duty cycles are constructed for each peak shaving dispatch, taken by concatenating their characteristic duty cycles into a single signal. First, the characteristic duty cycles can be concatenated in full to form the calendar/cycle life synthetic duty cycle. Alternatively, the zero-dispatch segments can be removed to form the cycle life synthetic duty cycle. These are shown in Figs. 17–19, as constructed from Fig. 15a–c, respectively. These synthetic duty cycles are much shorter than the yearly dispatch profiles, allowing for the laboratory testing on the full 8760 h in one yearly dispatch with only two to three days (48–72 h) of cycling data.

7.4. Validation of synthetic duty cycle aging characteristics

A high-fidelity electrochemical aging model was used to validate capacity fade trajectory resulting from cycling a cell under synthetic duty cycles against the capacity fade trajectory resulting from cycling a

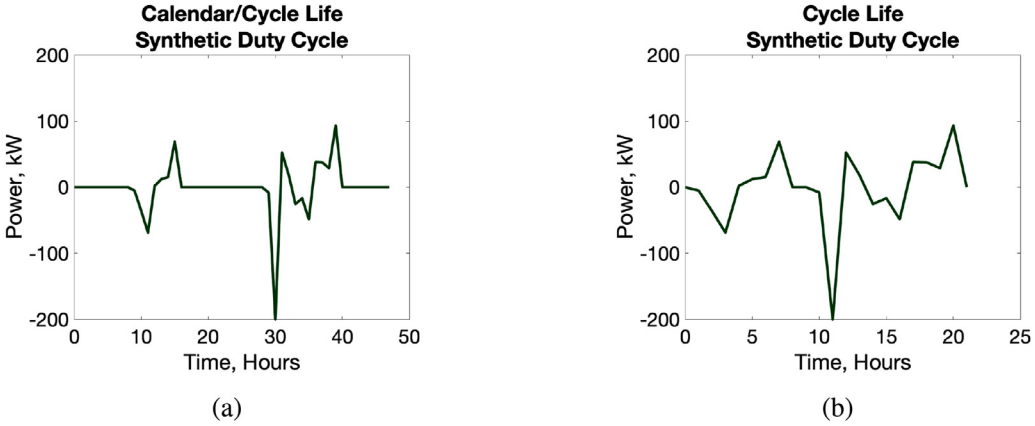


Fig. 17. (a) Calendar/cycle life synthetic duty cycle and (b) Cycle life synthetic duty cycle for the LargeOfficeNew peak shaving dispatch.

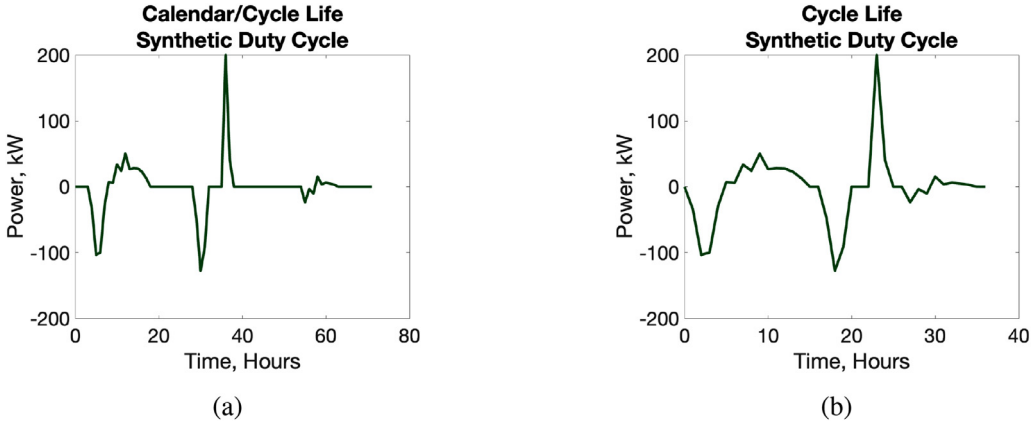


Fig. 18. (a) Calendar/cycle life synthetic duty cycle and (b) Cycle life synthetic duty cycle for the SuperMarketNew peak shaving dispatch.

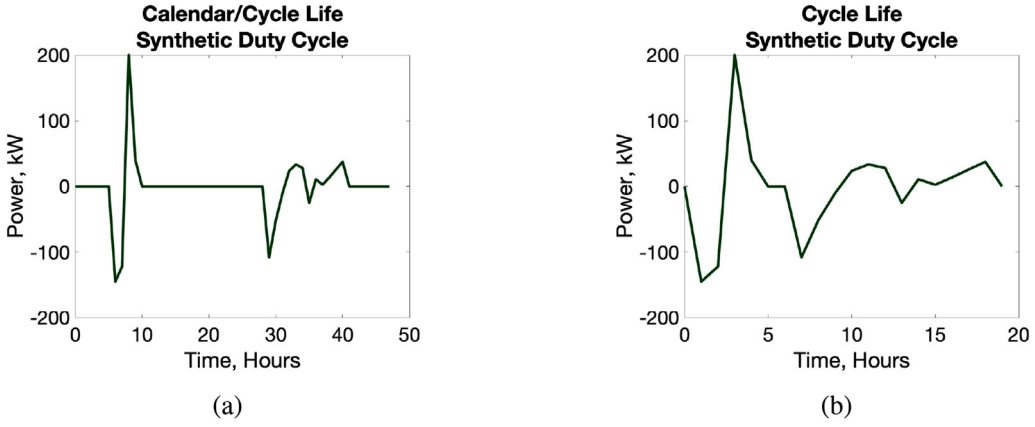


Fig. 19. (a) Calendar/cycle life synthetic duty cycle and (b) Cycle life synthetic duty cycle for the SecondarySchoolNew peak shaving dispatch.

cell under their respective power dispatches. The model is an enhanced single-particle model (ESPM) which captures the calendar- and cycle-aging effects of SEI layer growth and subsequent capacity loss [24,49]. The ESPM uses current applied to a 2 Ah cylindrical NMC cell as an input to a system of coupled partial and algebraic differential equations describing the electrochemical, aging, and thermal dynamics of the cell. Therefore, in order to use this model for validation, the following modifications were made to accommodate the system-level power duty cycle and dispatch profiles generated in this paper. First, the power profiles were downscaled from a maximum power of $W_{nom} = 200$ kW to 3 W, which allows for the same dispatch behavior on the 2 Ah cell as for the 400 kWh LIB. Next, power is incorporated as an input to the electrochemical aging model as demonstrated in Lee et al. [26]. Additionally, the synthetic duty cycles presented in Section 7.3 are charge-increasing. A cell repeatedly cycled with these duty cycles could increase in state of charge until the cell overcharges. To prevent such behavior, a discharge was included at the end of each synthetic duty cycle to ensure that they were charge-neutral. Aside from downscaling, the power dispatches were left unmodified.

The capacity fade is then simulated using the ESPM for two cases: a fresh, unused cell cycled under the power dispatch, and an identical fresh cell cycled under the corresponding synthetic duty cycle. In the

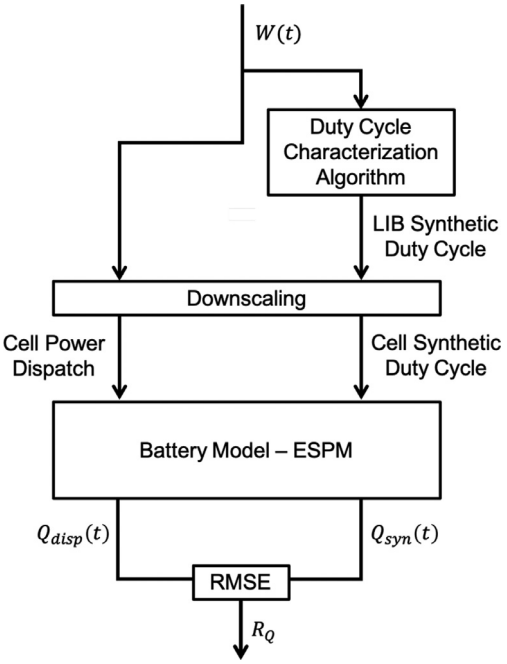


Fig. 20. Flowchart for the validation process, starting with the power dispatch $W(t)$ and producing the RMSE R_Q between the ESPM-simulated capacity fade trajectory resulting from cycling under the power dispatch $Q_{disp}(t)$ and the ESPM-simulated capacity fade trajectory resulting from cycling under the synthetic duty cycle $Q_{syn}(t)$.

first case, the ESPM simulated the cell cycling under the first 300 h of operation on the power dispatch (i.e. starting from January 1st), while in the second case, the ESPM simulated the cell cycling under synthetic duty cycles that were repeated and concatenated to achieve the same operating duration. Finally, the root-mean-squared error (RMSE) R_Q between the capacity trajectory of the cell cycled under the power dispatch $Q_{disp}(t)$ and the capacity trajectory of the cell cycled under the synthetic duty cycle $Q_{syn}(t)$ is calculated as

$$R_Q = \sqrt{\frac{1}{t_{end}} \sum_{t=1}^{t_{end}} \left| Q_{disp}(t) - Q_{syn}(t) \right|^2} \tag{13}$$

where t_{end} is the simulated operation duration (300 h). The validation process is shown in Fig. 20.

The capacity fade trajectories are shown in Fig. 21 for ESPM-simulated cell cycling under the SuperMarketNew and SecondarySchoolNew power dispatches and their respective synthetic duty cycles shown in Figs. 18a and 19 a. The LargeOfficeNew dataset was excluded from this analysis, as the power dispatch for this dataset (shown in Fig. 4a) is relatively inactive at the start of the year. Therefore, a battery cycled under this dispatch will exhibit calendar-aging-dominant capacity fade, which the ESPM cannot currently address.

For the SuperMarketNew and SecondarySchoolNew datasets, the RMSEs were $R_Q = 0.143\%$ and $R_Q = 0.072\%$, respectively. This validates that the synthetic duty cycles are able to reproduce the same aging trajectories present in the entire yearly power dispatch.

7.5. Applicability to general grid ESS dispatch

In this paper, the characterization algorithm is demonstrated in depth on the dispatch of energy storage for the grid application of peak shaving. However, as discussed in Section 2, there are 13 different grid services, each of which require a different ESS dispatch. The applicability of the methods used in this paper are not limited to peak shaving and can be applied to any grid ESS dispatch, with some modifications as discussed below.

It is first worth noting what portions of this paper would not need to be modified from the approach presented for peak shaving. The general framework of Algorithm 1 and its supporting methods presented in Section 3 do not assume any particular grid service, only requiring the resulting ESS dispatch. Similarly, once the characteristic duty cycles are produced from Algorithm 1, the creation and validation of synthetic duty cycles will be the same for any grid dispatch, regardless of the grid service performed.

However, grid services can operate on different time scales than the peak shaving presented in this paper, and under different operating conditions. For example, the grid service of frequency regulation requires the ESS to dispatch by following a signal that changes every two seconds [5]. Furthermore, as an electricity market service, frequency regulation allows the ESS owner to bid and reserve a variable power capacity allotted for frequency regulation signal-following [51], with the opportunity to recharge the ESS with electricity purchased from the day-ahead market [3]. Compared to ESS dispatch for peak shaving, both the time scale and operating conditions are different in ESS dispatch for frequency regulation.

Within the context of this paper, changing from peak shaving to frequency regulation would require a change to the construction of the dispatch interval matrix presented in Section 5. Frequency regulation would require an interval much smaller than the 24-hour day used in this paper. Previous studies have used 2-hour intervals for frequency regulation ESS dispatch [39]. Once the interval is defined, the process

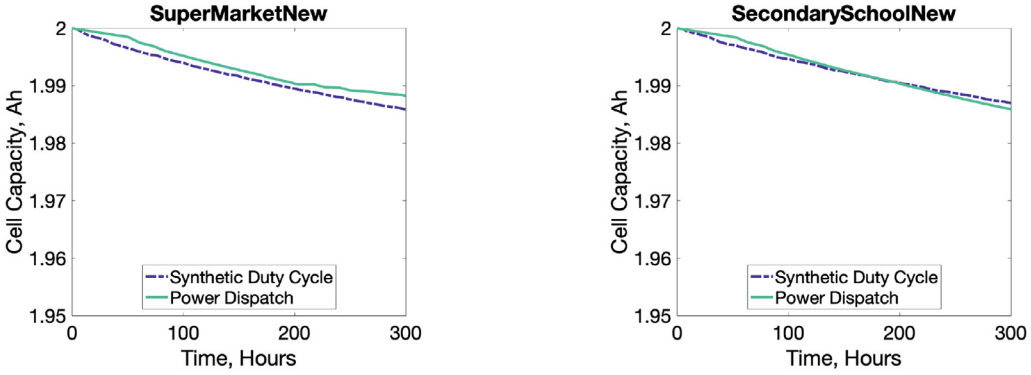


Fig. 21. Comparison of the capacity fade trajectory resulting from cycling under the power dispatch and the capacity fade trajectory resulting from cycling under the synthetic duty cycle for the (a) SuperMarketNew, and (b) SecondarySchoolNew yearly datasets.

of creating synthetic duty cycles would be the same as is presented in this paper.

8. Conclusions

This paper presented a framework for characterizing the duty cycle of ESS dispatch for grid applications using dispatch metrics, principal component analysis, and *k-means* clustering. The algorithm was applied to three different peak shaving dispatch profiles, and characteristic duty cycles were produced for each. As a result, lab-prone synthetic duty cycles were developed and applied to dispatch profiles for peak shaving. The developed synthetic duty cycles were validated against their respective original dispatch profiles using an electrochemical aging model, demonstrating that the aging characteristics were preserved while reducing the testing signal from thousands to tens of hours, providing tractable testing signals for laboratory testing of LIBs. Synthetic duty cycles presented in this paper can be used to cycle fresh batteries (irrespective of their cathode chemistry) and/or second-life batteries retired from used electric vehicles.

Furthermore, the research presented in this paper will lead to laboratory-developed models that encode the characteristics of grid application dispatch, leading to more accurate simulation tools for grid-scale LIBs. The measured capacity and internal impedance of batteries driven by synthetic duty cycles can be used, for example, for parameter identification of models for LIBs which include degradation from grid service dispatch. Such models can be used to build degradation-aware control frameworks and simulation tools for grid-scale LIBs, which could lead to better-informed strategies for the deployment and real-time operation of this grid storage technology.

Declaration of Competing Interest

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

Acknowledgments

The research presented within this paper is supported by the Bits and Watts Initiative within the Precourt Institute for Energy at Stanford University. S.J.H. was supported by the Assistant Secretary for Energy Efficiency, Vehicle Technologies Office of the U.S. Department of Energy under the Advanced Battery Materials Research Program.

Appendix A. Principal component analysis

The principal components and singular values of a real-valued matrix $X \in \mathbb{R}^{m\times n}$ can be obtained by the following steps:

1. Compute the vector $\boldsymbol{\mu} \in \mathbb{R}^{n\times 1}$, with the jth entry μ_j as the mean of the jth column of X

$$\mu_j = \frac{1}{m}\sum_{i=1}^{m} X_{ij}, \quad \forall j \in 1, \dots, n \tag{14}$$

2. Subtract the column mean from all columns in X to obtain the column mean-centered matrix, $A \in \mathbb{R}^{m\times n}$, with $\mathbf{1} \in \mathbb{R}^{m\times 1}$ as the vector of all ones of length m

$$A = X - \mathbf{1}\boldsymbol{\mu}^T \tag{15}$$

3. Compute the sample covariance matrix $C \in \mathbb{R}^{n\times n}$

$$C = \frac{1}{m-1} A^T A \tag{16}$$

4. Use singular value decomposition to obtain the eigenvalues $\sigma_1, \sigma_2, \dots \sigma_n$ and the eigenvectors $v_1, v_2, \dots v_n$ of C

$$C = V \Sigma V^T, \tag{17}$$

where matrix $V \in \mathbb{R}^{n\times n}$ and the diagonal matrix $\Sigma \in \mathbb{R}^{n\times n}$ are defined as the following

$$V = [v_1, v_2, \dots, v_n] \tag{18}$$

$$\Sigma = \mathbf{diag}(\sigma_1, \sigma_2, \dots, \sigma_n), \tag{19}$$

where the eigenvalues are ordered by decreasing magnitude (i.e. $\sigma_1 > \sigma_2 > \dots > \sigma_n$).

The columns of V are the principal components of X. The eigenvalues $\sigma_1, \sigma_2, \dots \sigma_n$ of C are the singular values of X. The principal components are orthogonal, i.e. $VV^T = V^T V = I$, where I is the identity matrix.

References

[1] Arthur D, Vassilvitskii S. *K*-means++: the advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms. USA: Society for Industrial and Applied Mathematics; 2007. p. 1027–35.
[2] California Independent System Operator. What the Duck Curve Tells us About Managing a Green Grid. Tech. Rep.. California ISO; 2013. https://www.caiso.com/Documents/FlexibleResourcesHelpRenewables_FastFacts.pdf
[3] Cao Y, Lee SB, Subramanian VR, Zavala VM. Multiscale model predictive control of battery systems for frequency regulation markets using physics-based models. J Process Control 2020;90:46–55. doi:10.1016/j.jprocont.2020.04.001.
[4] Caprino D, Della Vedova ML, Facchinetti T. Peak shaving through real-time scheduling of household appliances. Energy Build 2014;75:133–48. doi:10.1016/j.enbuild.2014.02.013.
[5] Carroll R. PJM manual 12: balancing operations. PJM; 2020. https://www.pjm.com/~/media/documents/manuals/m12.ashx
[6] Chen T, Jin Y, Lv H, Yang A, Liu M, Chen B, Xie Y, Chen Q. Applications of lithium-ion batteries in grid-scale energy storage systems. Trans Tianjin Univ 2020;26:208–17.
[7] Conover DR, Crawford AJ, Fuller JC, Gourisetti SNG, Viswanathan VV, Ferreira S, Schoenwald D, Rosewater D. Protocol for Uniformly Measuring and Expressing the Performance of Energy Storage Systems. Tech. Rep.. Albuquerque, NM, USA: Sandia National Laboratories; 2016. doi:102172/1249270.
[8] Cooley JW, Lewis PAW, Welch PD. The fast Fourier transform and its applications. IEEE Trans Educ 1969;12(1):27–34. doi:10.1109/TE.1969.4320436.
[9] Crawford AJ, Huang Q, Kintner-Meyer MC, Zhang J-G, Reed DM, Sprenkle VL, et al. Lifecycle comparison of selected Li-ion battery chemistries under grid and electric vehicle duty cycle combinations. J Power Sources 2018;380:185–93. doi:10.1016/j.jpowsour.2018.01.080.
[10] Deeba SR, Sharma R, Saha TK, Chakraborty D. A tool to estimate maximum arbitrage from battery energy storage by maintaining voltage limits in an LV network. In: 2015 IEEE PES asia-pacific power and energy engineering conference (APPEEC); 2015. p. 1–5. doi:10.1109/APPEEC.2015.7380894.
[11] Dembski N, Guezennec Y, Soliman A. Analysis and experimental refinement of real-world driving cycles. SAE Trans 2002;111:322–33. http://www.jstor.org/stable/44743063
[12] Deru M, Field K, Studer D, Benne K, Griffith B, Torcellini P, et al. U.S. Department of Energy Commercial Reference Building Models of the National Building Stock Tech. Rep.. Golden, CO: National Renewable Energy Laboratory; 2011. doi:102172/1009264.
[13] Devie A., Montaru M., Pelissier S., Venet P.. Classification of duty pulses affecting energy storage systems in vehicular applications. In: 2010 IEEE vehicle power and propulsion conference. p. 1–6.. 10.1109/VPPC.2010.5729137
[14] Di Filippi A, Stockar S, Onori S, Canova M, Guezennec Y. Model-based life estimation of Li-ion batteries in PHEVs using large scale vehicle simulations: an introductory study. In: 2010 IEEE vehicle power and propulsion conference; 2010. p. 1–6. doi:10.1109/VPPC.2010.5729020.
[15] Ding C, He X. *K*-means clustering via principal component analysis. In: Proceedings of the twenty-first international conference on machine learning. In: ICML '04. New York, NY, USA: Association for Computing Machinery; 2004. p. 29. doi:10.1145/1015330.1015408.
[16] Dinger A, Martin R, Mosquet X, Rabl M, Rizoulis D, Russo M, Sticher G. Batteries for Electric Cars. Tech. Rep.. The Boston Consulting Group; 2010.
[17] Energy Storage Association. End-user bill management. 2013. https://energystorage.org/end-user-bill-management/.
[18] Finkelstein J., Kane S., Rogers M.. How residential energy storage could help support the power grid. https://www.mckinsey.com/industries/electric-power-and-natural-gas/our-insights/how-residential-energy-storage-could-help-support-the-power-grid; 2019.
[19] Fitzgerald G, Mandel J, Morris J, Touati H. The Economics of Battery Energy Storage. Tech. Rep.. Rocky Mountain Institute; 2015.
[20] Green R, Staffell I, Vasilakos N. Divide and conquer? *K*-means clustering of demand data allows rapid and accurate simulations of the british electricity system. IEEE Trans Eng Manag 2014;61(2):251–60. doi:10.1109/TEM.2013.2284386.
[21] International Renewable Energy Agency. Renewable Capacity Highlights. Tech. Rep.. IRENA; 2020.
[22] Jain AK, Murty MN, Flynn PJ. Data clustering: a review. ACM Comput Surv 1999;31(3):264–323. doi:10.1145/331499.331504.
[23] Jolliffe IT. Principal component analysis. New York: Springer; 2002.

[24] Lam F, Allam A, Joe WT, Choi Y, Onori S. Offline multiobjective optimization for fast charging and reduced degradation in lithium-ion battery cells using electrochemical dynamics. IEEE Control Syst Lett 2021;5(6):2066–71. doi:10.1109/LCSYS.2020.3046378.
[25] Leadbetter J, Swan L. Battery storage system for residential electricity peak demand shaving. Energy Build 2012;55:685–92. doi:10.1016/j.enbuild.2012.09.035.
[26] Lee SB, Pathak C, Ramadesigan V, Gao W, Subramanian V. Direct, efficient, and real-time simulation of physics-based battery models for stand-alone PV-battery microgrids. J Electrochem Soc 2017;164:E3026–34. doi:10.1149/2.0031711jes.
[27] Liao TW. Clustering of time series data—A survey. Pattern Recognit 2005;38(11):1857–74. doi:10.1016/j.patcog.2005.01.025.
[28] Liu Z, Ivanco A, Onori S. Aging characterization and modeling of nickel-manganese-cobalt lithium-ion batteries for 48 V mild hybrid electric vehicle applications. J Energy Storage 2019;21:519–27. doi:10.1016/j.est.2018.11.016.
[29] Liu Z, Onori S, Ivanco A. Synthesis and experimental validation of battery aging test profiles based on real-world duty cycles for 48-V mild hybrid vehicles. IEEE Trans Veh Technol 2017;66(10):8702–9. doi:10.1109/TVT.2017.2717187.
[30] McLaren JA, Gagnon PJ, Mullendore S. Identifying Potential Markets for Behind-the-Meter Battery Energy Storage: A Survey of U.S. Demand Charges. Tech. Rep.. Golden, CO: National Renewable Energy Laboratory; 2017. doi:102172/928611.
[31] Moy K, Lee S, Onori S. Characterization of duty cycles for the peak shaving electric grid energy storage application. In: Proceedings of the 2020 ASME dynamic systems and control conference, Pittsburgh, Pennsylvania, USA; 2020.
[32] Moy K, Lee SB, Onori S. Characterization and synthesis of duty cycles for battery energy storage used in peak shaving dispatch. ASME Lett Dyn Syst Control 2021;1(4). doi:10.1115/1.4050192.
[33] Narimani MR, Asghari B, Sharma R. Energy storage control methods for demand charge reduction and PV utilization improvement. In: 2017 IEEE PES asia-pacific power and energy engineering conference (APPEEC); 2017. p. 1–5. doi:10.1109/APPEEC.2017.8308929.
[34] Pacific Gas & Electric. Tariffs. 2020. [Online; accessed 16-December-2020], https://www.pge.com/tariffs/index.page.
[35] Ramadass P, Haran B, Gomadam PM, White R, Popov BN. Development of first principles capacity fade model for Li-ion cells. J Electrochem Soc 2004;151(2):A196. doi:10.1149/1.1634273.
[36] Reihani E, Sepasi S, Roose LR, Matsuura M. Energy management at the distribution grid using a battery energy storage system (BESS). Int J Electr Power Energy Syst 2016;77:337–44. doi:10.1016/j.ijepes.2015.11.035.
[37] Reniers J, Mulder G, Howey D. Review and performance comparison of mechanical-chemical degradation models for lithium-ion batteries. J Electrochem Soc 2019;166:A3189–200. doi:10.1149/2.0281914jes.
[38] Rhodes JD, Cole WJ, Upshaw CR, Edgar TF, Webber ME. Clustering analysis of residential electricity demand profiles. Appl Energy 2014;135:461–71. doi:10.1016/j.apenergy.2014.08.111.
[39] Rosewater D, Ferreira S. Development of a frequency regulation duty-cycle for standardized energy storage performance testing. J Energy Storage 2016;7:286–94. doi:10.1016/j.est.2016.04.004.
[40] Sandia National Labs. QuESt: An Energy Storage Evaluation Application Suite. https://www.sandia.gov/ess-ssl/tools/quest; 2018.
[41] Schoenwald DA, Ellison J. Determination of Duty Cycle for Energy Storage Systems in a PV Smoothing Application. Tech. Rep.. Albuquerque, NM, USA: Sandia National Laboratories; 2016. doi:102172/1331494.
[42] Smith K, Saxon A, Keyser M, Lundstrom B, Ziwei Cao, Roc A. Life prediction model for grid-connected Li-ionbattery energy storage system. In: 2017 American control conference (ACC); 2017. p. 4062–8. doi:10.23919/ACC.2017.7963578.
[43] Stan A, Świerczyński M, Stroe D, Teodorescu R, Andreasen SJ. Lithium ion battery chemistries from renewable energy storage to automotive and back-up power applications - an overview. In: 2014 international conference on optimization of electrical and electronic equipment (OPTIM); 2014. p. 713–20. doi:10.1109/OPTIM.2014.6850936.
[44] Stroe D, Knap V, Swierczynski M, Stroe A, Teodorescu R. Operation of a grid-connected lithium-ion battery energy storage system for primary frequency regulation: a battery lifetime perspective. IEEE Trans Ind Appl 2017;53(1):430–8. doi:10.1109/TIA.2016.2616319.
[45] Sun Y, Wachche SV, Mills A, Ma O. 2018 Renewable Energy Grid Integration Data Book. Tech. Rep.. Golden, CO: National Renewable Energy Laboratory; 2020.
[46] Uddin M, Romlie MF, Abdullah MF, Abd Halim S, Abu Bakar AH, Chia Kwang T. A review on peak load shaving strategies. Renew Sustain Energy Rev 2018;82:3323–32. doi:10.1016/j.rser.2017.10.056.
[47] U.S. Energy Information Administration. Battery Storage in the United States: An Update on Market Trends. Tech. Rep.. EIA; 2020. https://www.eia.gov/analysis/studies/electricity/batterystorage/
[48] Wankmüller F, Thimmapuram PR, Gallagher KG, Botterud A. Impact of battery degradation on energy arbitrage revenue of grid-level energy storage. J Energy Storage 2017;10:56–66. doi:10.1016/j.est.2016.12.004.
[49] Weaver T, Allam A, Onori S. A novel lithium-ion battery pack modeling framework - series-connected case study. In: 2020 American control conference (ACC); 2020. p. 365–72. doi:10.23919/ACC45564.2020.9147546.
[50] Wilcox S, Marion W. Users manual for TMY3 data sets (Revised). National Renewable Energy Laboratory; 2008. doi:102172/928611.
[51] Xu B, Shi Y, Kirschen DS, Zhang B. Optimal battery participation in frequency regulation markets. IEEE Trans Power Syst 2018;33(6):6715–25. doi:10.1109/TPWRS.2018.2846774.
[52] Xu T, Zhang N. Coordinated operation of concentrated solar power and wind resources for the provision of energy and reserve services. IEEE Trans Power Syst 2017;32(2):1260–71. doi:10.1109/TPWRS.2016.2571561.

Kevin Moy is a second year Ph.D. candidate in the Energy Resources Engineering department at Stanford University. Previously, he received the bachelor's degree in Engineering Physics with a concentration in Renewable Energy and a coterminal master's degree in Civil and Environmental Engineering (Atmosphere and Energy), both from Stanford University.

Seong Beom Lee is currently a senior battery modeling researcher at Romeo Power, Inc. Prior to joining Romeo Power, he was a postdoctoral research fellow in the Department of Energy Resources Engineering at Stanford University. Seong Beom received a B.E. in Chemical Engineering and a Master's in Chemistry, from Sogang University, Seoul, South Korea. He also earned a Ph.D. in Chemical Engineering from the University of Washington, Seattle. His research topics are computational modeling of electrochemical battery systems, battery/grid integration, and data-driven approaches. Several of his research publications were among the most-read articles in the journals. Seong Beom is also passionate about developing new engineering courses. One of his research papers is being adopted as a curriculum for undergraduate laboratory courses at research-oriented universities in the United States.

Stephen Harris received a BS degree in chemistry from UCLA and a Ph.D. in physical chemistry from Harvard University. After a Miller Post-Doctoral Fellowship at UC Berkeley, he began his career at the General Motors Research Labs. Apart from a stint at the Ford Scientific Research Labs, Steve worked at GM until 2011, when he was awarded a Miller Visiting Professorship in the UC Berkeley Chemistry Department. Since then he has worked in the Materials Science and Energy Storage Divisions at Lawrence Berkeley Lab. During that time he has been a Visiting Scholar in the Materials Science and Engineering Department at Stanford, has consulted for battery companies and venture capital companies, and is on the advisory board for several battery startups.

Simona Onori holds a Laurea Degree in CSE, an M.S. in ECE, and a Ph.D. in Control Engineering from University of Rome 'Tor Vergata', University of New Mexico, and University of Rome 'Tor Vergata', respectively. She is an Assistant Professor in the Energy Resources Engineering Department at Stanford University where she also holds a courtesy appointment in EE. Her research is in sustainable transportation, clean energy and secondary life battery areas with a focus on hybrid and electric propulsion systems,electrochemical and electrical energy storage devices, emission mitigation devices and grid-storage systems. She serves as the Editor-in-Chief of the SAE International Journal of Electrified Vehicles since 2020 and she is a Distinguished Lecturer of the IEEE Vehicular Technology Society. She is the recipient of the 2020 U.S. DOE C3E Award in the research category, the 2019 Board of Trustees Award for Excellence, Clemson University, 2018 Global Innovation Contest Award, LG Chem, 2018 SAE Ralph R. Teetor Educational Award, 2017 NSF CAREER award.