Energy and AI 5 (2021) 100081

Contents lists available at ScienceDirect

Energy and AI

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

Review

Lithium-ion battery data and where to find it

Gonçalo dos Reis [a,b,*], Calum Strange [a], Mohit Yadav [c], Shawn Li [a]

[a] School of Mathematics, University of Edinburgh, The King's Buildings, Edinburgh EH9 3FD, UK
[b] Centro de Matemática e Aplicações (CMA), FCT, UNL, Quinta da Torre, 2829-516 Caparica, Portugal
[c] Indian Institute of Technology, Kanpur, India

HIGHLIGHTS

ARTICLE INFO

Article history:
Received 17 February 2021
Received in revised form 21 April 2021
Accepted 22 April 2021
Available online 30 April 2021

Keywords:
Lithium battery
Public data
Battery data
Battery tests
Machine learning

ABSTRACT

Lithium-ion batteries are fuelling the advancing renewable-energy based world. At the core of transformational developments in battery design, modelling and management is data. In this work, the datasets associated with lithium batteries in the public domain are summarised. We review the data by mode of experimental testing, giving particular attention to test variables and data provided. Alongside highlighted tools and platforms, over 30 datasets are reviewed.

1. Introduction

Lithium batteries currently dominate the battery market and the associated research environment. They display favourable properties when compared to other existing battery types: high energy efficiency, low memory effects and proper energy density for large scale energy storage systems and for battery/hybrid electric vehicles (HEV) [1]. Given these facts, lithium production has been expanding rapidly and the use of lithium batteries is wide spread and increasing [2].

From design and sale to deployment and management, and across the value chain [3], data plays a key role informing decisions at all stages of a battery's life. During design, data-informed approaches have been used to accelerate slower discovery processes such as component development and production optimisation (for electrodes, electrolytes, additives and formation) [4,5]. At sale, they can classify batteries based on expected lifetime [6,7]. At deployment, data on the expected lifetime and performance of batteries – for a range of chemistries, geometries, capacities and manufacturers – can help to determine the best battery for a given application: under different ageing stresses such as various charge/discharge currents [6,8,9], operating temperatures [10–12], depth of discharges (DODs) [13,14] and periods of disuse [15,16]. In use, the battery management system (BMS), controlling the battery's operation, relies heavily on data both for its own design and for the training and calibration of the models it uses.

Data driven approaches are showing great promise and proof of this is the growing body of literature exploring the interplay between data-driven techniques and battery applications [17–20]. The approach has been deployed in the design of new models for the estimation of state of health (SOH) [21–24], state of charge (SOC) [25,26] and internal resistance (IR) [27,28]; the prediction of remaining useful life (RUL) under cycling degradation [6,7,29], calendar ageing [30] and from electrochemical impedance spectroscopy (EIS) data [31]; the identification and prediction of phase change-points in capacity fade curves (knees) [7] and IR rise curves (elbows) [32]; new general online estimation methods for advanced BMSs [33]. Moreover, the data-driven paradigm has been used to improve fault detection [34–36], charge management [37,38], thermal management [39] and so much more: from materials development based on atomistic principles [40] to techno-economic analysis [41–44] and approaches to recycling [45].

Batteries are subjected to a wide range of operating conditions in turn influencing their performance, and thus, data covering these conditions is fundamental to the design and validation of accurate models. Physics-based and empirical models, often used in the BMS or 'in the cloud' with new 'digital twin' approaches [18,46], require careful calibration of model parameters; and, machine learning and statistical

* Corresponding author at: School of Mathematics, University of Edinburgh, The King's Buildings, Edinburgh EH9 3FD, UK.
E-mail address: g.dosreis@ed.ac.uk (G. dos Reis).

https://doi.org/10.1016/j.egyai.2021.100081

Nomenclature

HEV	hybrid electric vehicles
BMS	battery management system
SOC	state of charge
SOH	state of health
RUL	remaining useful life
EIS	electrochemical impedance spectroscopy
NCA	lithium nickel cobalt aluminium oxide ($LiNiCoAlO_2$)
LFP	lithium iron phosphate ($LiFePO_4$)
NMC	lithium nickel manganese cobalt oxide ($LiNiMnCoO_2$)
HPPC	hybrid power pulse characterisation
NASA	national aeronautics and space administration
DOD	depth of discharge
CALCE	centre for advanced life cycle engineering
LCO	lithium cobalt oxide ($LiCoO_2$)
CV	constant-voltage
CC-CV	constant-current constant-voltage
CC	constant-current
EOL	end of life
IR	internal resistance
RPT	reference performance tests
OCV	open-circuit voltage
FUDS	federal urban driving schedule
DST	dynamic stress test
HWFET	highway fuel economy driving schedule
UDDS	urban dynamometer driving schedule
LMO	lithium ion manganese oxide ($LiMn_2O_4$)
ECM	equivalent circuit model
eVTOL	electric vertical takeoff and landing
EV	electric vehicle
DOE OE	U.S. department of energy's office of electricity

based approaches require large amounts of data for training and perform poorly when predicting 'out of distribution' (in circumstances which differ greatly from those present in the data used to train the models). Within their vast scope of deployment, batteries undergo application specific degradation: the demands placed on an electric vehicle (EV) battery – periods of high, varying, load followed by extended rest – are quite different from those placed on powertools, laptops, cellphones, stationary energy storage, aeroplanes or satellites. For this reason, application specific data is needed and we bring attention to this in our discussion.

Well formatted and easily accessible public datasets will bring 'fresh eyes' to problems. Not everyone has access to a Lab to run experiments or the funds required to purchase data. Data that remains local to its generating lab can be leveraged only by a tiny fraction of a wide community of experts. The benefits of public data are numerous: researchers performing experiments gain a reference for their design and new insights into their data as other researchers with cross-domain expertise employ it; modellers and industry profit greatly from the ability to validate results and speed up discovery on public data; and, the barrier to entry is lowered for those new to an area. More data means more research and research is essential for economic growth, job creation and societal progress [47].

The **main contribution** of this work is to provide an actionable summary of publicly available lithium-ion battery data, giving particular attention to explored test variables and provided data. With this information, we hope to inform future research and experimental design, and encourage the sharing of new, accessible and well formatted datasets. To assist the reader, at the end of the main sections we provide tables summarising the presented datasets by cell, test variables, given data and number of cells with hyperlinks.

This work is organised as follows. The accessible testing data is categorised in Section 2 according to type and includes datasets available on request. Tools, libraries, platforms and a perspective on current limitations are covered in Section 3. Section 4 contains the conclusion of this review work and is followed by a nomenclature listing.

Links to data: All web links have been verified (at final submission). The links are given with bibliographical number and direct hyperlinks attached to the word 'URL'.

License: Datasets are provided under certain license attributions mainly according to Creative Commons [48, URL], the Open Database License [49, URL] and the Database Contents License [50, URL]. We refer to the supplementary material Section 1 for a summary description of the shorthand nomenclature.

Reference for 18650 type cells: Where full cell descriptions for a dataset were not given by the generating authors we refer to the resource [51, URL] which provides an extensive reference for the identification of 18650 type batteries.

2. Where is the data?

Historically, interest in different cell chemistries, testing conditions and procedures evolved reflecting the technological improvements batteries underwent. The first significant public battery dataset can be traced back to 2008 published by NASA [52]. As new battery chemistries appeared, the interest shifted from lithium iron phosphate (LFP) to lithium nickel manganese cobalt oxide (NMC) and lithium nickel cobalt aluminium oxide (NCA) batteries. Both NMC and NCA chemistries are better suited for power tools, e-bikes and other electric powertrains as they offer higher specific energy, reasonably good specific power and long lifespan. In Fig. 1, a hierarchical architecture of existing battery datasets across time is given. The number of cells tested and the variety of testing variables explored has increased with growing interest in data-driven techniques and a desire to understand more complex interactions.

Cell chemistry, number of tested cells and testing conditions are key to determine the usefulness of a specific battery dataset. We provide a comprehensive examination of the available datasets, in particular, highlighting these three elements.

2.1. Cycle ageing data

The generation of cycling data from the beginning to the end of a battery's life requires a significant investment of time and resources spanning many months or years. Experiments are run to investigate the influence of in-cycle factors (charging current, discharging current, temperature and DOD) on the capacity retention and (sometimes) rise in the internal resistance of batteries. Typically, cycle ageing datasets include in-cycle measurements of current, voltage and temperature, and per-cycle measurements of capacity and IR or impedance. Models are then developed according to the recorded cycling dataset to, among other things, predict future capacity retention, internal resistance growth and other health metrics. An overview of the typical recorded data and modelling pipeline for cycling (in particular, high-throughput) degradation datasets is illustrated in Fig. 2.

We prioritise in this section datasets with multiple cells, frequent in-cycle measurements and labs with multiple datasets. Smaller datasets (with only a few cells) and datasets without any in-cycle measurements are left to the end of this section (Section 2.1.8). The reader is invited to consult Table 2, at the end of the section, for an overview of the datasets discussed here.

2.1.1. National aeronautics and space administration

NASA hosts two high-throughput battery datasets on their website [53, URL] totalling 62 cells. We provide here a brief description of the datasets, for a full cell-by-cell experimental description see the 'ReadMe' file accompanying the datasets.

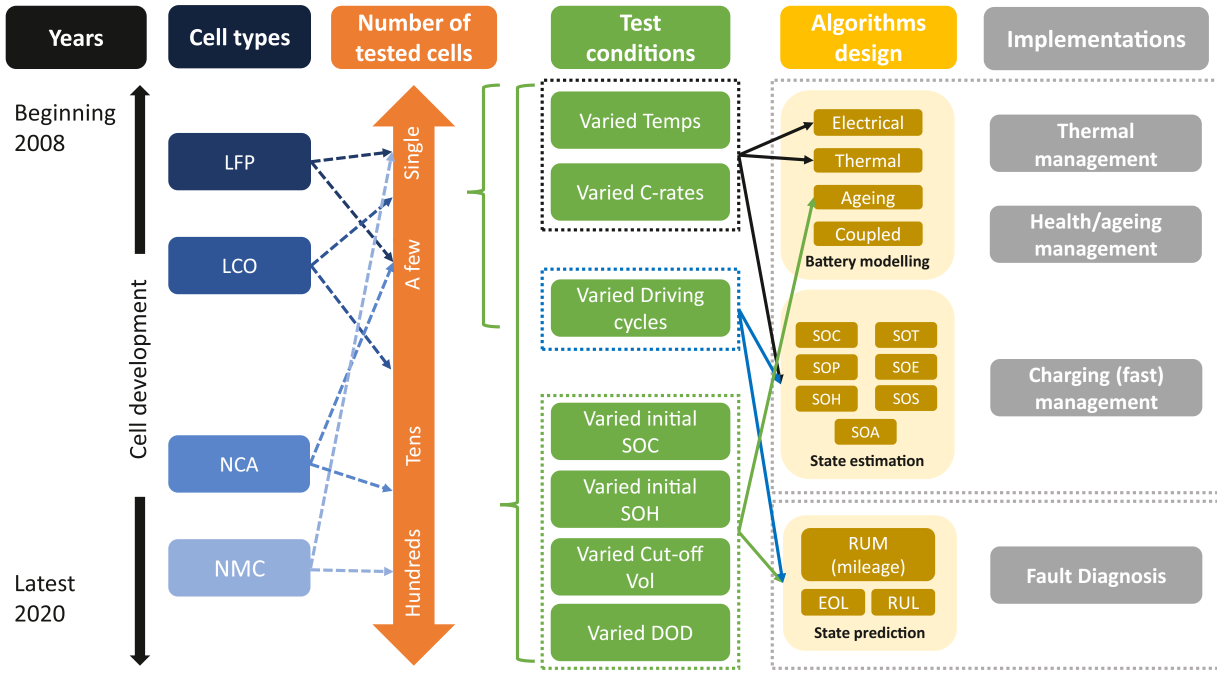

Fig. 1. Hierarchical architecture of the existing battery datasets from an historical point of view.

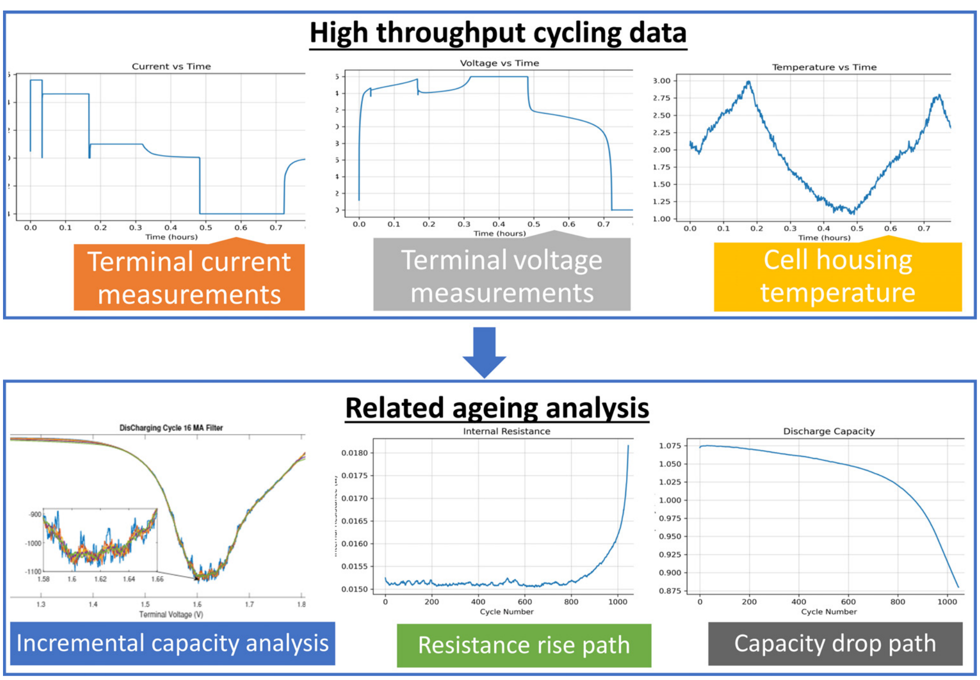

Fig. 2. The typical plots of a high-throughput cycling dataset encompassing measured terminal current, voltage and temperature variations. Capacity, IR, voltage and temperature can then be used for the ageing analysis.

The first of these datasets 'Battery Data Set' [10] contains data for 34 Li-ion 18650 cells with a nominal capacity of 2 Ah (we were unable to confirm the chemistry of these cells). This dataset was also the first publicly available battery dataset and has had a profound impact on the field; Table 1 summarises representative research work drawing on this dataset, giving a glimpse at its influence. Cells were cycled in a range of ambient temperatures (4 °C, 24 °C, 43 °C), charged with a common CC-CV protocol and with different discharging regimes. The dataset includes in-cycle measurements of terminal current, voltage and cell temperature, and cycle-to-cycle measurements of discharge capac-

Table 1
NASA dataset repository: Related papers and the corresponding research conducted. (See additionally Supplementary material Table 3 for full details.).

Category	SOH estimation and RUL prediction	Health prognostics and fault diagnostics	Battery modelling	Algorithms introduction and comparison
Ref	[54–60]	[61–63]	[64,65]	[66]

Table 2
Overview of cycle ageing datasets. 'gr' stands for 'graphite', 'Cal' denotes calendar ageing, 'Chrg' charge protocol and 'Dhrg' discharge, 'E' denotes 'energy'. Here, we use 'IR' to denote both internal resistance and impedance. No 'test variables' indicates that all cells in the experiment were cycled in the same way.

Location with weblink	Paper ref	Cell (form size chemistry)	Test variables	Data given	No. of cells
NASA [53, URL]	[10]	18650 2 Ah (?)	Dhrg, T	Q, IR, V, I, T	34
	[9]	18650 2.2 Ah LCO	Chrg, Dhrg, T	Q, IR, V, I, T	28
CALCE [67, URL]	[68,70]	prismatic 1.1 Ah LCO	Chrg, Dhrg	Q, IR, E, V, I, T	15
	[68,70]	prismatic 1.35 Ah LCO	Chrg, Dhrg, T	Q, IR, E, V, I, T	12
	[13]	pouch 1.5 Ah LCO	Chrg, DOD	Q, V, I	16
TRI [71, URL]	[6]	18650 1.1 Ah LFP/gr	Chrg	Q, IR, V, I, T	124
	[72]		Chrg	Q, V, I, T	233
Sandia [74, URL]	[11]	18650 multiple	Dhrg, DOD, T	Q, E, V, I, T	86
Oxford [75, URL]	[16]	18650 3 Ah NCA/gr	Chrg, Cal	Q, E, V, I, T	28
HNEI [80, URL]	[79]	18650 2.8 Ah NMC-LCO/gr	–	Q, E, V, I	15
EVERLASTING [84, URL] [85, URL]	[83]	18650 3.5 Ah NCA/gr	Chrg, Dhrg, T	Q, E, V, I	28
KIT [86, URL]	[8]	— 40 Ah NMC/gr	Chrg, Dhrg	V, I, T	44
UCL [87, URL]	[88]	18650 3.5 Ah NCA/gr	–	Q, V, T	1
Berkeley [89, URL]	–	18650 2.6 Ah LCO/gr	Chrg	Q, V, I, T	1
Xi'an Jiaotong [92, URL]	[90,91]	pouch 27 Ah NMC/gr	–	Q, E, V, I	2
Diao et al.[93, URL]	[12]	pouch 3.36 Ah LCO/gr	Chrg, Dhrg, T	Q	192
Poznan [94, URL]	[14]	18650 2.6 Ah NMC/carbon	Chrg, Dhrg, DOD, T	Q, I, T	28

ity and EIS impedance readings. The dataset is provided in '.mat' format under a double-attribution license.[1] The experiments were ended when cell capacity fell below 30% or 20% of nominal capacity.

The second dataset hosted by NASA, the 'Randomised Battery Usage Data Set' [9], contains data for 28 lithium cobalt oxide (LCO) 18650 cells with a nominal capacity of ∼2.2 Ah. The cells in this dataset were continuously operated. The dataset consists of 7 groups of 4 cells each group cycled at a set ambient temperature (room temp, 40° C); for 5 of these groups the cells were CC-charged to a fixed voltage and then discharged with currents selected at random from the group's discharge distribution table (7 different regimes). The other two groups were randomly charged and discharged. The dataset includes in-cycle measurements of terminal current, voltage and cell temperature, and measurements of discharging capacity and EIS impedance readings at 50 cycle intervals. The dataset is provided in a '.mat' format and measurements appear to have been taken until the cells reached between 80% to 50% SOH.

2.1.2. Centre for advanced life cycle engineering

The Centre for Advanced Life Cycle Engineering (CALCE) battery group has carried out substantive cycling tests for a diverse range of LCO/graphite cells. These datasets are hosted on their website [67, URL] – publications using the data should cite the corresponding CALCE article(s). Data is grouped by cell specification and not all data for a given specification comes from the same publication. We provide here a brief description of the datasets, for a full experimental description see the description on the website and the associated papers.

CALCE hosts data for 15 LCO prismatic CS2 cells grouped by experimental conditions (and publication) into 'Type-1' to 'Type-6'. 'Type-1' and 'Type-2' accompany one paper [68] and 'Type-3' to 'Type-6' another [69]. Type-1' consists of four 0.9 Ah cells, 'Type-2' of four 1.1 Ah cells and 'Type-3' to 'Type-6' each contain between one and two 1.1 Ah cells. The cells appear to have been cycled at room temperature (23 °C) and the experiments investigate different depths and ranges of partial charge and discharging, with a variety of C-rates. The dataset provides the cell cycler logs in Excel or '.txt' format containing measurements of current, voltage, discharge/charge capacity and energy, internal resistance and impedance. For each cell there are multiple files each containing the data for multiple cycles; the files are named according to the date at which they were recorded and, in our opinion, a significant amount of pre-processing is required to use this dataset. The data was recorded until batteries had (at least) passed their end of life (EOL), 80% SOH, with less than 200 cycles of data for the 'Type-1' batteries and approximately 800 cycles for the other cells.

The second set of cells tested by CALCE are 12 LCO prismatic CX2 cells with a rated capacity of 1.35 Ah. Which, similarly to the CS2 cells, are grouped into 'Type-1' to 'Type-6'. 'Type-1' and 'Type-2' (four cells each) were cycled in the same way as 'Type-1' of the CS2 cells [70]. The other four groups each have a single cell cycled with a variety of charge/discharge protocols; one of the cells was cycled at a range of temperatures (25 °C, 35 °C, 45 °C, 55 °C). The datasets are provided in the same format as the CS2 data with the same measurements.

In subsequent battery experiments [13], the group examined the influence of different depths of discharge (DOD) and discharging current stresses on the ageing of pouch cells: testing 16 LCO 1.5 Ah pouch cells in a 'semi-temperature controlled' room (25 ± 2 °C) [13]. The dataset is grouped by DOD and discharging protocol, provided in '.mat' format, containing cycler voltage, current and charge/discharge capacity data for between 400 and 800 'equivalent cycles'.

2.1.3. Toyota research institute in partnership with MIT and Stanford

In partnership with MIT and Stanford, the Toyota Research Institute (TRI) has published two substantial and easy to use high-throughput cycling datasets. Combined, these datasets contain data for 357 ($= 124 + 233$) commercial LFP/graphite cells manufactured by A123 Systems (APR18650M1A) with a rated capacity of 1.1 Ah. These two datasets are hosted online [71, URL], with accompanying experimental descriptions, under 'CC BY 4.0'.[2] The datasets are provided in '.csv', MATLAB struct and (second dataset only) JSON struct formats and a link to a GitHub repository with initial scripts is provided with the data. We point to the file structure of these datasets as a reference for future work: organised by cell → cycle → recorded data. Papers utilising the data should cite the appropriate publication.

The first of these datasets [6] (124 cells) was designed to explore the influence of fast charging protocols on cell ageing. Each cell was cycled with one from a range (72 different profiles) of one or two step fast charging protocols and a common CC-discharge protocol. The cells were cycled in a temperature controlled environment (30 °C). Data was logged from cycle 2 until a cell reached its EOL (80% SOH) – between 150 to 2300 cycles. The dataset contains in-cycle measurements of temperature, current, voltage, charge and discharge capacity, as well as per-

[1] As per the NASA description: 'Publications making use of databases obtained from this [the NASA] repository are requested to acknowledge both the assistance received by using this repository and the donors of the data.'

[2] To avoid confusion with Constant Current (CC), we add quotation marks when referring to a Creative Commons License.

cycle measurements of capacity, internal resistance and charge time. The data is split into three batches corresponding to three blocks of experiments carried out separately. In the accompanying paper [6] a feature based model is built on data from the first 100 cycles to predict the EOL. Since the dataset's release, numerous other papers have been published working with this data.

The second of these datasets [72] (233 cells) builds on the first: designing an approach to quickly optimise fast charging protocols. Again, cells were cycled in a temperature controlled environment (30 °C) with a common discharging protocol. The dataset is split into five batches of between 45 and 48 cells each; these batches were tested sequentially: for the first batch one of 224 different six-step charging protocols was chosen at random for each cell, the cells were tested for 100 cycles and then a model (trained on previously collected data) was used to predict the EOL based on this data. This prediction was used to inform the selection of charging protocols for the next batch of cells. This was repeated with the first four batches; the final batch was then tested until past the EOL comparing the selected optimal charging protocols with several other protocols. The dataset contains the same readings as the first dataset of 124 cells [6] except for the exclusion of IR readings. An attempt has been made to recover this missing data [32] where the IR has been predicted with a CNN model trained on the first dataset; this predicted IR data can be found online [73, URL].

2.1.4. Sandia national lab

The Sandia National Lab has performed testing for three chemistries of 18650 form cells: 'LFP from A123 Systems (APR18650M1A, 1.1 Ah), NCA from Panasonic (NCR18650B, 3.2 Ah), and NMC from LG Chem (18650HG2, 3 Ah)' [11]. In total there are 86 cells (30 LFP, 24 NCA and 32 NMC). The data from this study has been made available on the Battery Archive website [74, URL] – see Section 3.1 below. The data is shared under a double attribution license and on the website is denoted by the 'SNL' keyword. The experimental description is available on the Battery Archive page and in the relevant publication [11]. The cells were cycled at a range of temperatures (15 °C, 25 °C and 35 °C) with different DODs (0–100%, 20–80% and 40–60%) and discharge currents (0.5C, 1C, 2C and 3C); at least 2 cells from each chemistry were cycled in each combination of temperature, DOD and discharge current (12 groups) apart from the 3C discharge for the NCA cells. All cells were charged with a fixed rate of 0.5C. The cells were cycled until reaching their EOL (80% SOH) – at the time of publication cycling was still ongoing. The dataset contains in-cycle measurements of current, voltage, temperature and energy (Wh), and per-cycle measurements of charged/discharged capacity (bottom to top of DOD range) and other summary statistics. Periodically (roughly every 3% capacity loss), EIS measurements were taken measuring the full capacity of the cell. The data is provided in '.csv' format

2.1.5. Battery intelligence lab at the university of Oxford

The Battery Intelligence Lab at the University of Oxford hosts several battery degradation datasets on their homepage [75, URL]. We review here the 'Path dependence battery degradation dataset' [16] which is made of three parts. The files are provided under '.mat' format and all are licensed under Open Data Commons' ODbL v1.0 & DbCL v1.0 license. The dataset parts can be found as follows: Part 1 [75, URL] or [76, URL]; Part 2 [77, URL]; and Part 3 [78, URL].

The 3-year long project [16], spanning 2017–2020, studied ageing 'path dependence' of Li-ion cells by subjecting them to combined load profiles comprising fixed periods of calendar and cyclic ageing. The path dependence phenomena reflects the ageing sensitivity of cells to the order and periodicity of calendar ageing and cyclic ageing. The study analysed 28 commercial 3 Ah 18650 NCA/graphite cells (NCR18650BD). The dataset is provided in 3 parts (Part 1, 2 & 3) with the 28 cells split among ten groups (9 groups of 3 cells; 1 group of 1 cell), all tested at 24 °C. We provide a small breakdown for reference and point to the informative 'ReadMe' files. The data provided includes time, current, voltage, capacity and temperature, and the RPT and EIS testing data.

Group 1–4, 3 cells per group, were aged through cycling at a low C rate (C/2 and C/4) followed by 5 or 10 days of calendar ageing with RPTs run every 48 cycles. The first 18-months of experimental data is presented in 'Part 1' with months 19–36 presented in 'Part 2'. Additional to cell Groups 1–4, in Part 2 one finds Group 5 & 6 as control experiments. The cells of Group 5 are exposed to continuous C/2 cycling while Group 6 is exposed only to calendar ageing (at 90% SOC). Group 7–10 are presented in the dataset's 'Part 3' and parallels Group 1–4. Here, each group is cycled with CC-CV profiles then 5 or 10 days of calendar ageing. Reference performance tests (RPT) and EIS tests are used periodically to characterise the cells to differentiate the influence of different storage times and C-rates on battery degradation.

2.1.6. Hawaii natural energy institute

Researchers from the Hawaii Natural Energy Institute (HNEI) investigated the variability of cell degradation across 51 cells through cycling [79]. Data for 15 of these cells is shared on the Battery Archive website [80, URL] (denoted by 'HNEI' dataset). These 15 cells are commercial 2.8 Ah NMC-LCO/graphite 18650 cells (LG Chem, model 'ICR18650 C2'). The cells were cycled with fixed 1.5C discharge and C/2 charge protocols at 25 °C for ~1000 cycles. The dataset contains in-cycle measurements of current, voltage and charged/discharged capacity and energy, and per cycle measurements of charge/discharge capacity. Roughly every 100 cycles RPTs were run which are also present in the data. Files are in '.csv' format and shared under 'CC BY 4.0' plus 'source attribution' to Battery Archive. Additional experimental details and cell summary statistics (e.g. initial cell weight and received SOC) can be found in the accompanying paper [79].

2.1.7. EVERLASTING project

The recent European Commission funded project 'Electric Vehicle Enhanced Range, Lifetime And Safety Through INGenious battery management' (EVERLASTING) [81, URL] has published several battery related datasets on the '4TU.ResearchData' website [82, URL]. Of particular interest are the three datasets connected with the technical report produced by the project [83]. The report explores ageing from three different angles: drive cycle, calendar and CC-CV ageing at a range of temperatures; the datasets are described in the relevant sections of our paper.

Of these datasets, one experiment 'Lifecycle ageing' was carried out to investigate the interactions between temperature, charge/discharge C-rates and capacity loss. These experiments were performed on 28 Li-ion 18650 3.5 Ah commercial cells for a range of temperatures (0 °C, 10 °C, 25 °C and 45 °C), discharge C-rates (0.5C, 3C) and charge C-rates (0.5C, 1C). Two cells were tested at each possible pairing of temperature/charge-rate and temperature/discharge rate (except for 0 °C discharge). All 'charge' ('discharge') experiments had a common discharge (charge) profile. The data is hosted separately grouped by temperature (0 °C and 10 °C) [84, URL] and (25 °C and 45 °C) [85, URL]. The provided data is in '.csv' format with cycler logs (including voltage, current, charge/discharge capacity and energy) from characteristic cycles run roughly every two months – it is unclear if the data is complete.

2.1.8. Others

The Karlsruhe Institute of Technology (KIT) provides cycling data for 4 battery packs each consisting of 11 NMC/graphite 40 Ah cells on their website [86, URL] (under 'CC BY 4.0'). The batteries were cycled, at room temperature, in series with a range of charge/discharge profiles (detailed in the relevant paper [8]). The dataset provides high frequency (cell-by-cell and battery wise) measurements of voltage, temperature and inverter current/voltage for each of the tested charge/discharge profiles. The dataset is provided in well structured folders with '.csv' files and a starter MATLAB script.

Provided on the University College London (UCL) data website [87, URL] is cycling data for a single 3.5 Ah LG Chem NCA INR18650 MJ1 cell, given under 'CC0 1.0'. The cell was cycled according to the manufacturers recommendations in a fixed ambient temperature (24 °C) for 400 cycles [88]. The dataset provides in-cycle measurements of temperature, voltage and capacity, and per-cycle measurements of charge/discharge capacity, given in '.csv' format.

Berkeley provides data from a single Sanyo 18650 3.7V 2.6 Ah LCO/graphite cell on the Dryad Data website [89, URL] (under 'CC BY 4.0'). The cell was cycled with a variety of non-standard fast charging protocols. The dataset contains in-cycle measurements of voltage, current, temperature and charge/discharge capacity for 46 consecutive cycles and is provided in '.csv' format.

Researchers from Xi'an Jiaotong University [90,91] deploy the Coulomb counting method in combination with data-driven techniques to propose methods for SOC calibration and estimation. Both works use the same cycling data for battery cells under a regime of *fast capacity degradation*. The cells full physical description is found in the relevant paper [90, Table 1]. To summarise, two lithium-ion pouch cells with chemistry NMC/graphite and nominal capacity of 27 Ah were cycled from new until reaching 80% capacity. The cells were cycled with a CC-CV charge and CC discharge followed by a 30min relaxation period between cycles; the chamber's temperature was fixed at 40 °C and a total of about 400 cycles is recorded. The full cycling data and description file can be found in the paper's supplementary material [90] and it is unclear under which sharing license it is offered. However, provided separately [92, URL] under 'CC0 1.0' are the first 100 cycles of data (in '.xlsx' format). The experimental data recorded is: battery voltage, current, charging/discharging capacity and energy.

Diao et al. [12] provide a dataset considering the influence of ambient temperature, discharging current stress and cut-off points of charging current for CC-CV cell ageing. This dataset is hosted on the Mendeley data platform [93, URL] and shared under 'CC BY-NC 3.0'. In the experiment, 192 LCO/graphite pouch-type 3.36 Ah cells were tested using the above three stress factors. The dataset contains capacity measurements taken at 50 cycle intervals, is given in '.mat' format and only 182 of the 192 cells appear to be listed.

Researchers at Poznan University of Technology provide data for 28 Samsung NMC/carbon 2.6 Ah 18650 cells on the Mendeley data platform [94, URL] (under 'CC BY 4.0'). The cells were cycled at a variety of temperatures, DODs and charging/discharging currents until reaching 80% SOH. The dataset consists of 'learning data' from 28 cells containing summary measurements of ambient temperature, discharging current, DOD, average charging current and number of equivalent cycles for cells at a range of SOH values (9 measurements from 100% to 80% SOH), given in '.xlsx' format. This data was used in the paper [14] to train several models to predict the cell's current SOH.

Lastly, we mention data, shared by researchers at the University of Oviedo under 'CC BY 4.0', for two LFP pouch cells [95, URL]. The cells were tested at room temperature (23 °C) for a single full charge/discharge cycle at a constant current rate of C/25 [96]. The dataset contains voltage, current and temperature readings, from the charge/discharge cycle, sampled every 2 s for a total experimental time of 60 h (details can be found in the associated paper [96, Section 5]).

2.2. Drive cycle data

Energy is required to propel an automobile. With a conventional internal combustion engine the combustion of fossil fuels, converted to mechanical energy, drives the vehicle forward. However, with global concern surrounding greenhouse gas levels there is an urgent push for the automobile industry to reduce carbon emissions. For this reason, standardised testing procedures capturing the dynamic power demands of driving are indispensable: allowing the relative efficiency and performance of engines to be compared. These standard test procedures are referred to as driving cycles.

A driving cycle is a standardised dynamic vehicle drive schedule encoded by a velocity-time table/profile. The velocity and acceleration are pre-scheduled per time step, and thus the required mechanical power is a function of time. The integral of mechanical power over the duration of the driving schedule represents the total energy required for a specific driving cycle. For electric vehicles the battery system generates this required mechanical energy. Datasets collected by cycling batteries according to the drive schedules can be used to compare the efficiency of EVs with traditional vehicles and to test the performance of derived battery models and SOC estimation algorithms under realistic conditions.

The globally recognised driving cycle tables can be divided into three groups: European driving cycles, US driving cycles and Asian (Japanese, Chinese -Beijing) driving cycles [97,98]. For example, the Urban Dynamometer Driving Schedule (UDDS) [99] is commonly used for 'city-based EV driving cycle tests' representing light-duty city driving conditions. US06 represents an aggressive driving cycle with high engine loads. The european drive cycle ARTEMIS [100] contains 12 driving cycles that range across several driving conditions: congested urban, free-flow urban, secondary roads, main roads and motorways. The Highway Fuel Economy Driving Schedule (HWFET) is used to describe cars cruising under 60mph on a highway. And, the Air Resources Board LA92 dynamometer driving profile was developed to depict a driving cycle with higher top and average speed, lower idle time, fewer stops per mile and a higher maximum rate of acceleration when compared with UDDS.

An overview of driving cycle data reviewed in this section can be found in Table 5.

2.2.1. University of Wisconsin-Madison & McMaster university

The battery research group at the University of Wisconsin-Madison offers a battery testing dataset covering four typical driving cycles: US06, HWFET, UDDS and LA92. The dataset, published on the Mendeley data website [101, URL] (under 'CC BY 4.0'), contains data from a single 2.9 Ah NCA Panasonic 18650PF cell. The cell was cycled according to the above driving cycles and an additional 'neural network driving cycle' systematically through a range of temperatures (25 °C, 10 °C, 0 °C, 10 °C, and 20 °C, in that order). A full experimental description can be found in the accompanying 'ReadMe' file. The dataset includes characterisation data from Hybrid Power Pulse Characterisation (HPPC) and EIS tests, and in-cycle measurements from the driving cycles including voltage, current, capacity, energy and temperature. The data is presented in '.mat' and '.csv' files with a well structured format sorted by temperature, test type and drive cycle.

The same group, but operating at McMaster University, provides another driving schedule test dataset for a series of battery tests carried out for a single 3 Ah LG Chem INR18650HG2 NMC cell [102, URL]. The cell was cycled at six different ambient temperatures (40 °C, 25 °C, 10 °C, 0 °C, 10 °C, and 20 °C) according to the same mix of drive cycles as the Panasonic cell. The dataset contains the same data as for the Panasonic cell (in a similar format) with the addition of 'prepared data' which has been processed in order to train and test a provided SOC estimator.

The above two driving cycle datasets, hosted by University of Wisconsin-Madison and McMaster University, provide a benchmark for driving cycle tests and are at the heart of crucial contributions in the development of SOC estimation algorithms and battery models; some of these works are reviewed in Tables 3 and 4.

2.2.2. Others

Researchers from the University of Science and Technology of China (USTC) have explored the co-estimation of model parameters and SOC for batteries and ultra-capacitors [109]. The data accompanying this research has been shared in the journal 'Data in Brief' [110, URL] along with an experimental description, under 'CC BY 4.0'. A lithium battery pack (LFP-1665130-10 Ah, produced by Fujian Brother Electric CO., LTD of China – 4 prismatic cells in series) and an ultra-capacitor (BCAP3000 P270 2.7V/3.0Wh, produced by Maxwell Technologies, Inc.) were each cycled once according to two different driving

Table 3
Panasonic 18650PF Li-ion Battery Data: Related paper and the corresponding research conducted.

Category	Ref	Detail
SOC estimation	[103]	This paper introduces a data-driven approach for State of Charge (SOC) estimation of Li-ion batteries using a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM).
	[104]	This paper proposed a stacked bidirectional LSTM neural network for SOC estimation of lithium-ion batteries.
Battery modelling	[105]	In this paper, the potential of applying advanced machine learning techniques to model lithium-ion batteries is explored. Rather than using the more common ECM and physics-based models, a data-driven approach is used to build battery models.
	[106]	In this paper, ECMs of lithium-ion batteries are built to capture various the electrochemical properties of the battery. The ECMs are validated by a series of five automotive drive cycles performed at temperatures ranging from -20 °C to 25 °C.

Table 4
LG 18650HG2 Li-ion Battery Data: Related paper and the corresponding research conducted.

Category	Ref	Detail
SOC estimation	[107]	This paper introduces a data-driven approach for State of Charge (SOC) estimation of Li-ion batteries using a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM).
	[108]	This paper proposed a stacked bidirectional LSTM neural network for SOC estimation of lithium-ion batteries.

Table 5
Overview of Driving cycle data. ‘cycle’ here denotes the use of different drive cycle profiles, ‘E’ denotes ‘energy’. No ‘test variables’ indicates that all cells in the experiment were cycled in the same way.

Location with URL	Cell (form size chemistry)	Test variables	Data recorded	No. of cells
Madison [101, URL]	18650 2.9 Ah NCA	cycle, T	Q,V, I, E, T, EIS	1
McMaster [102, URL]	18650 3 Ah NMC	cycle, T	Q,V, I, E, T, EIS	1
USTC [110, URL]	prismatic 10 Ah LFP	cycle	V, I	1
EVERLASTING	18650 3.5 Ah NCA/gr			
[111, URL]	16 cell modules	T	Q,V, I, T	2
[84, URL] [85, URL]	18650 3.5 Ah NCA/gr	cycle, DOD, T	Q,V, I, E, T	18
Oxford [75, URL]	pouch 0.74 Ah —	–	Q,V, T	8
Aachen [113, URL]	18650 3.4 Ah NCA/C+Si	–	Q,OCV, V, I, T	28

Table 6
Overview of Calendar Ageing degradation data. Here, ‘time’ denotes the frequency at which ageing was interrupted to take measurements.

Location with URL	Cell (form size chemistry)	Test variables	Data recorded	No. of cells
CALCE [67, URL]	pouch 1.5 Ah LCO/gr	SOC, T, time	Q, IR, E, V, I	144
Oxford [127, URL]	18650 3 Ah NCA/gr	–	Q, E, V, I, T	1
EVERLASTING [128, URL] [129, URL]	18650 3.5 Ah NCA/gr	SOC, T	Q, E, V, I	8

cycles (DST and UDDS) at room temperature. The dataset, provided in ‘.xlsx’ format, contains per second measurements of current and voltage for the battery and ultracapacitor during the two drive cycle profiles.

The EVERLASTING project provides two drive cycling datasets both shared under ‘CC BY-NC 4.0’. The first of these datasets [111, URL] contains data for two battery modules each built from 16 NCA/graphite 3.5 Ah LG Chem INR18650 MJ1 cells. The modules were cycled at a variety of temperatures according to an ‘adapted real driving profile’. The dataset contains in-cycle measurements of pack voltage, current, charge/discharge capacity, ambient temperature and per-cell temperature. The second of these datasets, described in the EVERLASTING report [83], contains data for 16 NCA 3.5 Ah LG Chem INR18650 MJ1 cells cycled according to a recorded city drive profile for two DOD ranges (70-90% and 10-90%) and at a variety of temperatures (0 °C, 10 °C, 25 °C and 45 °C) – 2 cells per combination. In addition, 2 cells were cycled according to a recorded highway drive profile at 25 °C (10–90% DOD). This dataset is stored in two locations according to temperature: (10 °C and 0 °C) [84, URL] and (45 °C and 25 °C) [85, URL]. The datasets are both in ‘.csv’ format but with different information depending on the temperature. The cells cycled at 25 °C and 45 °C include measurements of voltage, current and charge/discharge capacity and energy; whereas, the data for the cells at 0 °C and 10 °C has a different file structure and additionally includes temperature readings.

The Oxford Battery Intelligence Laboratory provides the ‘Battery Degradation Dataset 1’ [112] on their website [75, URL], licensed under Open Data Commons’ ODbL & DbCL. This dataset contains data for eight 740 mAh lithium-ion pouch cells manufactured by Kokam (part number SLPB533459H4). The cells were cycled at a constant ambient temperature (40 °C) using a CC-CV charging regime and the ARTEMIS [100] driving cycle discharging profiles until their EOL (80% SOH). The dataset is provided in ‘.mat’ format containing voltage, temperature and discharge capacity (mAh) measurements. These measurements (taken at 10 minute intervals) were recorded during characterisation tests performed every 100 cycles. A full experimental description can be found in the PhD thesis of C. Birkl [112, Chapter 5.2].

The Institute for Power Electronics and Electrical Drives at Aachen University hosts drive cycling data for 28 Samsung 18650 NCA/carbon + silicon cells with a nominal capacity of 3.4 Ah on their website [113, URL] (under ‘CC-BY-4.0’). The cells were cycled at a fixed ambient temperature (25 °C) with a CC-CV charging regime and a recorded drive cycling discharge profile. The dataset contains in-cycle measurements of voltage, current and temperature, and checkup tests

Table 7
Non-publicly available Battery Data: Related paper and the corresponding research conducted. (See additionally Supplementary material Section 3.).

Applications	Ref	Data features
SOC/SOH estimation	[144]	Real-word EV data; bulk datasets (300 EV & 400 HEV); battery pack health; NMC batteries; long-term test (over 12 months);used for big data analysis and machine learning method
	[145]	Single cell tests (3 cells); SOC/SOH; statistical data-driven model fusion; 18650 LCO; DST and capacity tests
Battery modelling	[146]	A small batch of cells tested (51 cells/20 cells); cylindrical and pouch cells; NMC and NMC-LMO batteries; varied temperatures and accelerated ageing tests; electrical/thermal/ageing modelling
	[147]	A few cells tested (27 cells); calendar ageing test; long-term tests (over one year); NCA batteries; varied temperatures and initial SOC; calendar ageing modelling
Fault identification	[36]	31 NMC cells; charge profiles (rate ranged from 4C to 9C); data driven method; Li-Plating; 450 cycles; 30 °C test temperature; capacity, end-of-charge rest voltage (EOCV), open circuit voltage (OCV), and Coulombic efficiency (CE) were recorded
Capacity related early heath prognostics/ RUL prediction	[148]	A small batch of cells tested (35 cells); NMC batteries; early fault detection; real data collected on production lot samples; data-driven methods
	[149]	A few of cells tested; varied temperatures; incremental capacity analysis; LFP, NCM capacity data

every 30 cycles with capacity, quasi open-circuit voltage (OCV) and pulse tests (at 80%, 50% and 20% SOC). The dataset is provided in '.csv' format and a detailed experimental description can be found in the accompanying 'MetaData' file.

For completeness, we mention a dataset available to 'IEEE DataPort'[3] subscribers [114, URL], under 'CC BY 4.0', containing data from simulated driving cycles composed according to the Federal Test Procedure repository. We point the reader to the data description given on their website.

2.3. Characterisation data for cell modelling

The cycling performance of different lithium battery chemistries is varying and highly dependent on operating conditions (temperature, current load, age). To evaluate the viability of lithium batteries to a given application, features of battery performance such as the OCV-SOC table, impedance and IR are necessary. These cycling features can then be used to model the electrical dynamics and cycling performance of a battery. The experimental data collected for this purpose mainly targets the short-term responses of current and voltage, and focuses on the impedance variance at different battery SOC levels and temperatures.

2.3.1. In-cycle battery data

The CALCE battery group has piloted the research in terms of battery modelling and internal state estimation providing a series of baseline in-cycle datasets for cells from different types of lithium batteries [67, URL] (under 'attribution' license). Two experiments, namely, low-current OCV and incremental-current OCV, have been deployed to collect OCV for commercial INR 18650-20R 2 Ah NMC/graphite cells. The OCV dataset includes different OCV-SOC tables achieved at three ambient temperatures (0 °C, 25 °C and 45 °C). Voltage responses under different dynamic current profiles, such as, the Dynamic Stress Test (DST), Federal Urban Driving Schedule (FUDS), US06 Highway Driving Schedule and Beijing Dynamic Stress Test, are provided to test the accuracy of the proposed SOC estimation algorithms [115,116] and analyse the dependence of SOC estimation on OCV variations due to temperature [117]. Since the OCV-SOC table is temperature-sensitive, further investigation has been conducted by CALCE battery group providing a temperature-dependent OCV-SOC dataset for 2.23 Ah A123 LFP/graphite cells. The temperature-dependent OCV-SOC dataset is collected from low current OCV tests for a wide range of temperatures spanning from 10 °C to 50 °C with an interval of 10 °C. Additionally, the experimental data of DST and FUDS tests performed in the corresponding ambient temperatures is available. All data is given in '.xlsx' format and provided is the data from the OCV tests and in-cycle data from the drive cycles (including voltage, current, charge/discharge capacity and energy, IR and impedance).

In order to develop an advanced model which reproduces the thermal and electrical dynamics of the battery, Planella et al. [118] at Warwick University tested the cycling behaviours of commercial 5 Ah LG INR21700 M50 NMC cells with a range of ambient temperatures (0 °C, 10 °C and 25 °C) and C-rates (0.1C, 0.5C, 1C and 2C). In their experiments, four cells are tested per specific C-rate and temperature. The dataset, along with experimental description and additional scripts, is hosted on Github [119, URL] (under BSD 3-Clause License). The data is the cycler logs given in '.csv' format containing voltage, current, capacity, energy and temperature readings from cycles run to compare the derived model with real data.

Few works have been conducted to test the discharging power behaviour of cells. One publicly available dataset [120, URL], under 'CC BY-NC 3.0', has investigated the behaviour of 4 types of 18650 Li-ion cells, produced by 4 different cell manufacturers (LG 18650-HB6 1.5 Ah NMC, Panasonic NCR18650B 3.35 Ah NCA, Shenzhen IFR18650 1.5 Ah LFP and Efest IMR18650 3.1 Ah LMO – between 8 and 13 cells per manufacturer), at different constant power discharge rates [121] and a constant ambient temperature (25 °C). In particular, the experiments were designed to capture the available power response of cells at high (out of specification) current loads. The provided data, given in '.csv' and '.mat' format, appears to contain cycler logs for each cell (spanning ~6 h) with voltage, current, temperature, charge/discharge capacity and power measurements, however, no column headings or 'ReadMe' file are given.

2.3.2. Impedance spectroscopy data

Applying electrochemical impedance spectroscopy (EIS) to measure the impedance of lithium batteries is widely accepted in battery research [122]. EIS can separate the dependence of different components by varying the frequency of applied AC currents. Allowing the contribution of the solution resistance, charge transfer/polarisation resistance, double layer capacity, wire inductance etc., to be interpreted from the measured responses. The EIS provides a tool to understand and model the complicated non-linear electrochemical process occurring inside a battery. A typical characterisation process for a lithium battery, using EIS measurements according to the frequency domain analysis and modelling, can be found [123]; the frequency setting of EIS inputs are standard for most systems: ranging from 20 mHz to 10 kHz. In general, high-frequency EIS responses are considered indicative of inductive behaviour and low-frequency responses indicative of capacitive behaviour.

As given in Fig. 3, the Nyquist representation of an impedance spectrum (acquired from the lithium battery test) is used to fit an equivalent circuit model (ECM) – an ECM provides a simplified battery model as

[3] This dataset is not 'public' but we are aware that many readers may have IEEE memberships. We have not verified the contents of this dataset.

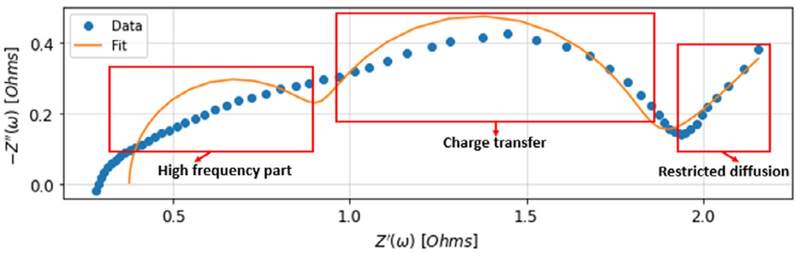

Fig. 3. A typical Nyquist plot: Battery characterisation using EIS measurements.

a circuit of standard components whose parameters are fitted to approximately replicate the measured response. The lithium Nyquist plot can, in general, be divided into three parts based on the frequency responses. In the high-frequency segment the inductive behaviour of wires is dominant, contributed to by the cables impedance, fittings, connectors, cell tabs and current collectors. Over the middle-frequency segment the shape of the Nyquist plot behaves like a depressed semi-circle, representing the charge transfer resistance and double-layer capacitance, its shape affected by the temperature. For the low-frequency segment the slower ion-diffusion process dominates the cell dynamics and measuring the resistance response here takes longer to perform. In addition, this process can be influenced by many factors, such as electrode material, porosity, operating temperatures, SOC and voltages. Typically, cell capacitance has a very steep slope (around 90 degrees). As illustrated in Fig. 3, at points the frequency responses behave closely to those of a capacitor.

One of the largest broad-scale datasets of EIS measurements has been shared on the 'zenodo' platform [124, URL] (under 'CC BY 4.0') containing over 20,000 EIS readings collected from 12 Eunicell LR2032 45 mAh LCO/graphite cells. The cells were cycled at a range of different temperatures (25 °C, 35 °C and 45 °C) with multiple frequencies of EIS measurements taken at different SOC levels. The experiment was stopped when cells reached their EOL (80% SOH). The dataset was recorded to accompany research exploring the prediction of RUL and SOH from EIS data [31]. The provided data contains the EIS measurements (resistance, impedance and phase at a range of frequencies) and, separately, measurements of capacity in '.txt' format.

In Section 2.7 below, we refer to another EIS dataset [125] available upon request.

2.4. Calendar ageing

Calendar ageing 'comprises all ageing processes that lead to the degradation of a battery cell independent of charge-discharge cycling' [15]. Such ageing is most pronounced in applications where periods of idleness are longer than operation, such as with electric vehicles. It is argued that calendar ageing may also play a role in cycle ageing studies where cycle depths and current rates are low [126]. In this section we overview datasets dedicated to calendar ageing see Table 6 for an overview of the datasets.

Outside the battery cycling data, the CALCE group has also studied calendar ageing and a dataset appears on their website [67, URL] ('Pouch Cells: Storage Data and Test Description'): 144 LCO/graphite 1.5 Ah pouch-type cells with three different initial SOC values (0%, 50% and 100%) were calendar aged at four different storage temperatures (40 °C, 5 °C, 25 °C and 50 °C). There are three testing groups, 48 cells per group, with capacity and impedance measurements taken every 3 weeks, 3 months and 6 months, respectively. The dataset is provided in '.xls' format and contains the cycler data (current, voltage, charge/discharge capacity and energy, internal resistance and impedance) from the periodic characterisation cycles.

Group 6 in 'Part 2' [127, URL] of the Oxford 'Path dependence battery degradation dataset' [16] contains the data of one single cell exposed to continuous calendar ageing at 90% SOC. We do not provide further details here and refer the reader to the description given already in Section 2.1.5.

As part of the EVERLASTING project [81] (see Section 2.1.7) calendar ageing was performed on several NCA/graphite 18650 3.5 Ah LG Chem cells (model INR18650 MJ1). The testing was carried out for a range of temperatures (0 °C, 10 °C, 25 °C and 45 °C) and the cells were stored at OCV with different initial SOC levels (10%, 70% and 90%). The data shared online does not appear to be complete; however, data for 2 cells stored at 25 °C [128, URL], 3 cells stored at 0 °C and 3 cells stored at 10 °C [129, URL] is available. The provided data is in '.csv' format, shared under license 'CC BY-NC 4.0', and contains cycler data (voltage, current, capacity and energy) from characterisation tests performed periodically.

Lastly, we refer the reader to Section 2.7 regarding 'Data on demand'. We mention the dataset made available by Dr Dhammika Widanalage (Warwick Manufacturing Group, Warwick University (UK)) which contains many cells tested under calendar ageing.

2.5. Aeroplanes, satellites and energy storage

Beyond traditional cycling, calendar and drive cycle ageing, there are a few public datasets containing battery cycle data from more specialist applications. We review here four datasets relating to usage in aeroplanes, satellites and energy storage.

2.5.1. Aeroplane usage battery data

Another NASA Ames Prognostics Data Repository [53, URL] dataset is the 'HIRF Battery DataSet' [130]. It contains usage data from one single battery powering a small electric unmanned aerial vehicle [54]. The data in provided in '.mat' format under a double attribution license (see Section 2.1.1) a 'reference document' is provided on the NASA website explaining the file structure and experimental details.

Researchers from Carnegie Mellon University provide the 'eVTOL Battery Dataset' [131, URL] (shared under 'CC BY-NC-SA 4.0'). The dataset consists of discharge data from 22 Sony-Murata 18650 3 Ah VTC-6 cells cycled with simulated Electric Vertical Takeoff and Landing (eVTOL) duties [132]. The data is provided in '.csv' format and includes voltage, temperature, current and charge/discharge capacity and energy measurements. The provided 'ReadMe' file and corresponding paper [132] give a full experimental description. This dataset is the first of its kind: providing public eVTOL data.

2.5.2. Simulated satellite operation profile battery data

The last dataset hosted by NASA [53, URL] that we report on is the 'Small Satellite Power Simulation Data Set' [133]. The dataset is provided in '.mat' format under a double attribution license (see Section 2.1.1). It contains data for two BP930 batteries (off-the-shelf 18650 Li-ion cells rated at 2.1 Ah) run 'continuously with a simulated satellite operation profile completion for a single cycle' – experimental description in the corresponding paper [134, Section IV]. Additional details and data descriptions can be found by consulting the 'reference document' provided on the NASA website.

2.5.3. Stationary energy storage

Researchers from the University of Oxford and 'EnergyVill', with data provided on the Oxford Research Archive [135, URL], built a battery ageing model to serve a techno-economic analysis for grid-connected batteries [41–43]. Six Kokam 16 Ah lithium polymer cells (model SLPB-78205130H) were aged following profiles corresponding to optimal trading strategies for stationary batteries in the Belgian day-ahead market of 2014. Experimental details can be found in the mentioned references and the Ph.D. thesis of Reniers [42]. The cycling ageing tests were performed for up to one year to record the entire battery degradation process from the beginning of life to EOL. This dataset contains measured current, voltage and operating temperatures at ~200 s intervals, and monthly capacity measurements (details provided in 'ReadMe' file). Files are given in '.csv' format and the database is shared under both the ODbL v1.0 and DbCL v1.0 license.

2.6. Synthetic data

Data driven approaches require data; thus, a lack of data is a significant barrier to their use. The obvious solution of collecting more data, covering a wide range of operating conditions, is expensive and time-consuming. Another approach is to use the available data to generate more data. This can be achieved by perturbing the data (data augmentation) or by generating artificial data. In this subsection we will review examples of the latter: producing so-called *synthetic data* for battery cells. Synthetic data can enhance existing datasets improving the performance of trained models and allowing for interpolation between cycling conditions not included in the experimental data. This interpolation step may be particularly important to data driven approaches enabling prediction 'outside the distribution' of the experimental data.

Here, we briefly describe one approach [136] to generate synthetic current and voltage data. For the generation of current curves a Markov chain approach can be used: transition probability matrices are constructed from real EV cycling data and then by iterating through the matrices (Markov chain propagation) synthetic current data can be obtained. From the generated current profile 'voltage cluster centroids' (the average value of temporally local voltage clusters obtained via k-means clustering) can then be predicted by a neural network trained on real data. These clusters have been shown to provide an effective feature for the prediction of SOC [137].

A comprehensive synthetic diagnostic dataset containing more than 500,000 individual voltage vs. capacity curves has been generated alongside a prognostic dataset with more than 130,000 individual degradation paths for a commercial graphite based LFP battery [138]. The diagnostic datasets [139, URL] and the prognostic dataset [140, URL] are both available on the Mendeley data website under a 'CC BY 4.0' license. The data is given in '.mat' format [141].

2.7. Data by request

Research projects are often subject to restrictions on the public release of generated datasets, however, upon publication some authors make their data 'available upon request'. This section briefs on such works and the corresponding datasets.

We mention here research carried out at the University of Warwick (UK). Dr Dhammika Widanalage, the principal investigator for the project, has provided us with the following description: 'Warwick University (UK) has been conducting thorough ageing tests on a batch of commercial LG M50 21700 cells (graphite/Si-NMC811). These tests consider two types of cell ageing: calendar and cycling. The calendar ageing tests cover four different ambient temperatures (0 °C, 25 °C, 45 °C, and 60 °C) and thirteen different initial SOC settings. Three cells were tested for each combination of ambient temperature and initial SOC. The cycling ageing tests consist of cells cycled at a variety of current C-rates for two low ambient temperatures (0 °C and 10 °C); the cells were immersed in an oil bath for thermal management. For all experiments (calendar and cycling) RPT were performed periodically to measure capacity losses, IR growth, and to log the pseudo-OCV values. In detail, first the discharge capacity was measured by the CC discharging protocol then the resistance at five different SOC levels (100%, 80%, 50%, 20%, 5%) was measured using pulse charge/discharge HPPC tests. The RPT procedures were run every four weeks for the calendar ageing tests and approximately every two weeks for the cycling ageing tests.' The above described experiments provide a comprehensive ageing dataset and set a benchmark for future data collection. The ageing data can be used for the analysis, modelling, prediction and tracing of ageing trajectories. Unfortunately, an external link to freely access the data cannot be offered. However, the datasets and on-going research progress (corresponding experimental cell data) are available for academic use, on request. If interested, please contact Dr Dhammika Widanalage via email Dhammika.Widanalage@warwick.ac.uk.

Other researchers at the University of Warwick have performed an ageing investigation based on EIS measurements for four NCA 18650 cells [125,142]. Their EIS dataset has been deposited onto the university data repository [143, URL] and is accessible by request. Additional research manuscripts making use of datasets that remain private but whose authors point that the research datasets are available by request are listed in Table 7. A fuller description of their data and experimental work is detailed in the works themselves and summarised in Supplementary material Section 3 (expanding Table 7).

3. Data governance, repositories, tools and future outlook

Lithium batteries have been widely deployed and a vast quantity of battery data is generated daily from end-users, battery manufacturers, BMS providers and other original equipment manufacturers. Two elements are key in enabling the value of data: accessibility and ease of use. If no one can find or understand a public dataset it has no value. And, much of the time spent pre-processing data could be saved given a widely used standard publication format.

In this section, we review data platforms and online repositories that can be used to host data; tools for data validation and processing; and community maintained living reviews

3.1. Data repositories and platforms

Data storage platforms provide a common and easily navigable location to find and (possibly) share data. They also promote standardisation in data format and descriptions. We point here to several repositories hosting public battery data.

- **Scholarly usable, citable and freely accessible**
  1. **Battery Archive** [80, URL]: Battery archive, developed at the City University of New York Energy Institute, provides a free repository of battery testing data which is easily searchable by cell chemistry, form, capacity and test variables. Different datasets, shared by various institutions, have common file formats and the website provides easy access to the data. We highlight their 'rules for metadata' section proposing a common nomenclature to use for descriptions of cells and cycling conditions.
  2. **DOE OE** [150, URL]: The U.S. Department of Energy's Office of Electricity (DOE OE) has collaborated with two national labs, *Sandia National Laboratories* and *Pacific Northwest National Lab*, to carry out battery research addressing energy storage risk assessment and mitigation. Their website provides free access to the resulting research data including abuse tests, cycling tests and EIS measurements.
  3. **NREL** [151, URL]: The National Renewable Energy Laboratory is a national laboratory of the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy, operated by the Alliance for Sustainable Energy, providing free battery datasets

to aid in the development of cell models and tools to facilitate the deployment of renewable energy. Regarding their battery research, well-rounded testing data encompassing the failure data collected from hundreds of abuse tests (nail penetration, thermal abusing, and internal short-circuiting), ageing cycling data, driving cycle data and other commercial oriented battery operating data (collected from EV operation) has been provided.

- **Public Digital data repositories**
  There are several curated data platforms that make research data discoverable, freely reusable and citable. A non-exhaustive list of the publicly accessible data repositories where battery data has been deposited is outlined as follows.
  1. Dryad [152, URL]
  2. Zenodo [153, URL]
  3. European federation of data driven innovation hubs [154, URL]
  4. Mendeley data centre [155, URL]
  5. 4TU.ResearchData [82, URL]
  6. Google Database [156, URL]

3.2. Community maintained reviews and standards

There are a few community maintained online resources listing publicly available battery datasets. The approaches taken to curate such lists differ but represent a critical initial step from the community to make public datasets more accessible and understandable. This review includes, at the time of publication, the datasets in these referred community maintained resources and several other datasets with corresponding descriptions. Researchers with knowledge on where to find battery datasets are heartily invited to contribute to the living reviews listed below.

- Community databases of publicly available battery datasets maintained by
  1. Dr. Valentin Sulzer (University of Michigan): [157, URL]
     (by way of private communication, this resource is no longer maintained)
  2. Dr. Bolfazl Shahrooei (Iranian Space Research Center): [158, URL]
     (community maintained and active)
- Standards and identification references
  1. BatteryStandards.info [159, URL]: Website containing information on around 400 standards for rechargeable batteries including: battery test standards across categories such as characterisation tests, safety tests, performance tests and requirements.
  2. An extensive identification reference for lithium-ion Battery of size-type 18650 covering brand, model, capacity, chemistry, max charge/discharge and link to product specification datasheet is presented in: [51, URL].

3.3. Data processing and validation tools

Battery cycling data is highly complex. Different cycling protocols, cycler manufacturers and experimental configurations make it difficult to compare datasets and validate models. As a result, several high quality open source packages have been created to perform data processing, parsing and validation. We provide a non-exhaustive summary of available tools.

- **Tools for data management and validation**[4]
  1. **BEEP (Battery Evaluation and Early Prediction)** [161, URL]: a package for parsing and featurizing battery cycling data specifically geared towards cycle life prediction [162].
  2. **cellpy** [163, URL]: a package which parses Arbin cycler data and enables manipulation of cycling data using pandas dataframes. In addition, it enables incremental capacity (dQ/dV) analysis and the extraction of open circuit relaxation points.
  3. **impedance.py** [164, URL]: a package for the analysis of electrochemical impedance spectroscopy (EIS) data. Core functionality includes plotting experimental impedance spectra, fitting impedance spectra to equivalent circuit models, computing and plotting the impedance spectra of equivalent circuit models and validation of impedance spectra using the Kramers-Kronig relations.
  4. **Bayesian Hilbert transform** [165, URL]: Python implementation of [166] providing validation of EIS data via the Kramers-Kronig relation recast under a Bayesian framework.

Lastly, we point to means of extracting numerical data from data visualisations, for instance, the open-source software *WebPlotDigitizer* [167, URL]. Given an image of a plot the raw data points are identified in a semi-automated manner. Numerical data is extracted based on the identified data points and user-defined calibration points marked on the plot. Such an approach has been used in Richardson et al. [23] (with MATLAB's GRABIT tool) to extract capacity fade curve data from published work.

3.4. Current limitations

Effective energy storage is critical. Improvements in safety, density and longevity mean more reliable devices, vehicles requiring less frequent charging and replacement, and efficient and long lasting stationary energy solutions. Currently, the communication of data between end-users, manufacturers, distributors and providers is weak. Greater transparency in this aspect would accelerate scientific progress in all areas. Fig. 4 illustrates the wide ranging deployment of batteries across industries.

Regarding the data reviewed in the manuscript, we failed to find many examples of 'on field data' where the varying conditions of battery usage can be seen. Examples of such data would be: data regarding aerospace applications either from the perspective of aeroplane electrification or simply from satellite usage where batteries are a mission critical element; battery usage data for energy storage systems (either at home-owner level or at the electric grid level); data regarding electric heavy-duty vehicles (e.g., firetrucks or buses); data linking material science data to cycling data or data connecting manufacturing to degradation; data that can be used to optimise the cell selection process for the purpose of battery pack formation. Moreover, all the data reviewed in this manuscript is from first life applications where the battery was tested from new. We have not found any data on the so-called *battery 2nd life* where the battery, say, was redeployed from a EV into a stationary application like grid energy storage. Lastly, left out of this study was a review of data relating to *abuse testing* and data *containing mechanical measurements*. A representative of the latter would be datasets that include mechanical measurements, e.g., cell dilation or weight.

Battery testing is costly and lengthy, and this is unlikely to change: how can one understand the life-cycle of a battery that lives for 10 years without carrying out 10 years of testing? A sub-problem in this context is the sparsity of recorded data, for example, cells are usually tested within a (dotted) range of fixed temperatures and with fixed cycling conditions. These conditions do not reflect the variability of real use. And, many approaches fall short when interpolating between recorded data: how does one predict cell degradation at 25 °C from data recorded at 40 °C and 10 °C. Methodologies addressing this problem are beginning to emerge [29,168].

From a holistic point of view, the publicly available datasets come in all shapes and sizes. Files appear in '.mat', '.txt', '.csv' or 'Excel' format ('.mat' and '.csv' being the more common) with wildly varying file structures: from raw cycler data – split by cycle, week, month or not

[4] The authors kindly thank Koeller [160] for his assistance in developing this list.

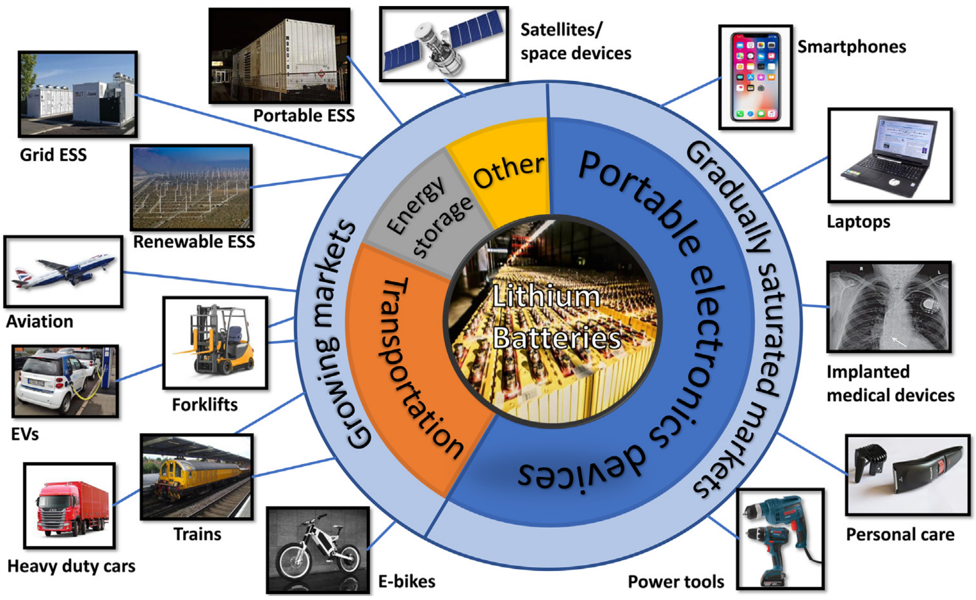

Fig. 4. Lithium battery sample applications.

at all – to structured data with explanatory scripts and text to assist the user. From our understanding, there is a general lack of consensus on the way to present data. For instance, different brands of cell cycling machines output data in different ways including varying nomenclature for the same quantities. In this regard, we highlight again the open-source Python-based framework BEEP (Battery Evaluation and Early Prediction) [161] for the management and processing of high-throughput battery cycling data and the Battery Archive's 'Rules for Metadata' section [80] proposing a common nomenclature for the descriptions of cells and cycling conditions. In the Author's opinion, exemplary datasets for file format and description are those provided by P. Kollmeyer in Section 2.2.1 and the Toyota Research Institute data in Section 2.1.3. We leave a suggestion for any group sharing data: to provide a basic accompanying script (MATLAB or Python) that plots the uploaded data (time series/EIS or capacity/resistance change) and text explaining file structure. This, on its own, would expedite the understanding of datasets; however, there is a clear and greater benefit which could be gained from researchers adopting a uniform file format.

4. Conclusions

Comprehensive battery datasets play a critical role for battery research both in academia and industry. However, publicly available datasets are distributed sporadically as battery testing is costly and lengthy. In this work, a review of the existing battery datasets in the public domain is provided with a category-type break-down covering the testing regimes, cell specifications and provided data. This informs a long view on the available datasets hinting at gaps in the experimental space which in itself presents an opportunity for further work. Lastly, high-quality open source packages for a variety of battery-related tools are also reviewed.

With this work we wish to convey two further messages,

1. the academic community is starved for research data, and
2. we strongly encourage any person or group (academic or industrial) to share their data.

Funding

This project was funded by an industry-academia collaborative grant *EPSRC EP/R511687/1* awarded by *EPSRC & University of Edinburgh* program *Impact Acceleration Account (IAA)*.

G. dos Reis acknowledges support from the *Fundação* para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UIDB/00297/2020 (Centro de Matemática e Aplicações CMA/FCT/UNL).

Declaration of Competing Interest

Authors declare that they have no conflict of interest.

CRediT authorship contribution statement

Gonçalo dos Reis: Writing - review & editing, Supervision. **Calum Strange:** Writing - review & editing. **Mohit Yadav:** Writing - review & editing. **Shawn Li:** Writing - review & editing.

Acknowledgements

All authors thank Dhammika Widanalage (Warwick University, Warwick Manufacturing Group group), Philipp Dechent (RWTH Aachen University), the several members of the 'Battery Bits' community for the helpful discussions, in particular, Jason Köller.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.egyai.2021.100081.

References

[1] Marinaro M, Bresser D, Beyer E, Faguy P, Hosoi K, Li H, et al. Bringing forward the development of battery cells for automotive applications: perspective of R&D activities in China, Japan, the EU and the USA. J Power Sources 2020;459:228073. doi:10.1016/j.jpowsour.2020.228073. https://www.sciencedirect.com/science/article/pii/S0378775320303761.

[2] Ding Y-L, Cano Z, Yu A, Lu J, Chen Z. Automotive Li-ion batteries: current status and future perspectives. Electrochem Energy Rev 2019;2:1–28. doi:10.1007/s41918-018-0022-z.

[3] Howey DA, Roberts SA, Viswanathan V, Mistry A, Beuse M, Khoo E, et al. Free radicals: making a case for battery modeling. Electrochem Soc Interface 2020;29(4):30.

[4] Liu K, Wei Z, Yang Z, Li K. Mass load prediction for lithium-ion battery electrode clean production: a machine learning approach. J Clean Prod 2021;289:125159.

[5] Liu K, Hu X, Zhou H, Tong L, Widanalage D, Macro J. Feature analyses and modelling of lithium-ion batteries manufacturing based on random forest classification. IEEE/ASME Trans Mechatron 2021:1.

[6] Severson K, Attia P, Jin N, Perkins N, Jiang B, Yang Z, et al. Data-driven prediction of battery cycle life before capacity degradation. Nat Energy 2019;4:1–9. doi:10.1038/s41560-019-0356-8.

[7] Fermín P, McTurk E, Allerhand M, Medina-Lopez E, Anjos MF, Sylvester J, et al. Identification and machine learning prediction of knee-point and knee-onset in capacity degradation curves of lithium-ion cells. Energy AI 2020;1:100006.
[8] Steinbuß G, Rzepka B, Bischof S, Blank T, Böhm K. FOBSS: monitoring data from a modular battery system. In: Proceedings of the tenth ACM international conference on future energy systems; 2019. p. 456–9.
[9] Bole B, Kulkarni C, Daigle M. Randomized battery usage data set. NASA Ames Progn Res Center 2009.
[10] Saha B, Goebel K. Battery data set. NASA Ames Progn Res Center 2007.
[11] Preger Y, Barkholtz HM, Fresquez A, Campbell DL, Juba BW, Romàn-Kustas J, et al. Degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. J Electrochem Soc 2020;167(12):120532.
[12] Diao W, Saxena S, Pecht M. Accelerated cycle life testing and capacity degradation modeling of $LiCoO_2$-graphite cells. J Power Sources 2019;435:226830.
[13] Saxena S, Hendricks C, Pecht M. Cycle life testing and modeling of graphite/$LiCoO_2$ cells under different state of charge ranges. J Power Sources 2016;327:394–400.
[14] Burzyński D, Kasprzyk L. A novel method for the modeling of the state of health of lithium-ion cells using machine learning for practical applications. Knowledge-Based Syst 2021;219:106900.
[15] Keil P, Schuster SF, Wilhelm J, Travi J, Hauser A, Karl RC, et al. Calendar aging of lithium-ion batteries. J Electrochem Soc 2016;163(9):A1872.
[16] Raj T, Wang AA, Monroe CW, Howey DA. Investigation of path-dependent degradation in lithium-ion batteries. Batteries Supercaps 2020;3(12):1377–85.
[17] Hu C, Ye H, Jain G, Schmidt C. Remaining useful life assessment of lithium-ion batteries in implantable medical devices. J Power Sources 2018;375:118–30.
[18] Wu B, Widanage WD, Yang S, Liu X. Battery digital twins: perspectives on the fusion of models, data and artificial intelligence for smart battery management systems. Energy AI 2020:100016.
[19] Ng M-F, Zhao J, Yan Q, Conduit GJ, Seh ZW. Predicting the state of charge and health of batteries using data-driven machine learning. Nat Mach Intell 2020:1–10.
[20] Aykol M, Gopal CB, Anapolsky A, Herring PK, van Vlijmen B, Berliner MD, et al. Perspective–combining physics and machine learning to predict battery lifetime. J Electrochem Soc 2021;168(3):030525.
[21] Qin T, Zeng S, Guo J. Robust prognostics for state of health estimation of lithium-ion batteries based on an improved PSO–SVR model. Microelectron Reliab 2015;55(9-10):1280–4.
[22] Kim J, Yu J, Kim M, Kim K, Han S. Estimation of Li-ion battery state of health based on multilayer perceptron: as an EV application. IFAC-PapersOnLine 2018;51(28):392–7.
[23] Richardson RR, Osborne MA, Howey DA. Gaussian process regression for forecasting battery state of health. J Power Sources 2017;357:209–19.
[24] Roman D, Saxena S, Robu V, Pecht M, Flynn D. Machine learning pipeline for battery state of health estimation. Nat Mach Intell 2021. doi:10.1038/s42256-021-00312-3.
[25] Chang W-Y. The state of charge estimating methods for battery: areview. Int Sch Res Not 2013;2013:7.
[26] Chemali E, Kollmeyer PJ, Preindl M, Emadi A. State-of-charge estimation of Li-ion batteries using deep neural networks: a machine learning approach. J Power Sources 2018;400:242–55.
[27] Remmlinger J, Buchholz M, Meiler M, Bernreuter P, Dietmayer K. State-of-health monitoring of lithium-ion batteries in electric vehicles by on-board internal resistance estimation. J Power Sources 2011;196(12):5357–63.
[28] Liang K, Zhang Z, Liu P, Wang Z, Jiang S. Data-driven Ohmic resistance estimation of battery packs for electric vehicles. Energies 2019;12(24):4772.
[29] Tang X, Liu K, Wang X, Gao F, Macro J, Widanage WD. Model migration neural network for predicting battery aging trajectories. IEEE Trans Transp Electrif 2020;6(2):363–74.
[30] Liu K, Hu X, Wei Z, Li Y, Jiang Y. Modified Gaussian process regression models for cyclic capacity prediction of lithium-ion batteries. IEEE Trans Transp Electrif 2019;5(4):1225–36.
[31] Zhang Y, Tang Q, Zhang Y, Wang J, Stimming U, Lee AA. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. Nat Commun 2020;11(1):1–6.
[32] Strange C, Li S, Gilchrist R, dos Reis G. Elbows of internal resistance rise curves in Li-ion cells. Energies 2021;14(4):1206.
[33] Pan H, Lü Z, Wang H, Wei H, Chen L. Novel battery state-of-health online estimation method using multiple health indicators and an extreme learning machine. Energy 2018;160:466–77.
[34] Pan Y, Feng X, Zhang M, Han X, Lu L, Ouyang M. Internal short circuit detection for lithium-ion battery pack with parallel-series hybrid connections. J Clean Prod 2020;255:120277.
[35] Yang R, Xiong R, He H, Chen Z. A fractional-order model-based battery external short circuit fault diagnosis approach for all-climate electric vehicles application. J Clean Prod 2018;187:950–9.
[36] Chen B-R, Kunz MR, Tanim TR, Dufek EJ. A machine learning framework for early detection of lithium plating combining multiple physics-based electrochemical signatures. Cell Rep Phys Sci 2021;2(3):100352.
[37] Maia LK, Drünert L, La Mantia F, Zondervan E. Expanding the lifetime of Li-ion batteries through optimization of charging profiles. J Clean Prod 2019;225:928–38.
[38] Liu K, Zou C, Li K, Wik T. Charging pattern optimization for lithium-ion batteries with an electrothermal-aging model. IEEE Trans Ind Inform 2018;14(12):5463–74.
[39] Shang Y, Liu K, Cui N, Wang N, Li K, Zhang C. A compact resonant switched-capacitor heater for lithium-ion battery self-heating at low temperatures. IEEE Trans Power Electron 2019;35(7):7134–44.
[40] Deringer VL. Modelling and understanding battery materials with machine-learning-driven atomistic simulations. J Phys 2020;2(4):041003.
[41] Reniers JM, Mulder G, Ober-Blöbaum S, Howey DA. Improving optimal control of grid-connected lithium-ion batteries through more accurate battery and degradation modelling. J Power Sources 2018;379:91–102.
[42] Reniers JM. Degradation-Aware Optimal Control Of grid-Connected Lithium-ion Batteries. University of Oxford; 2019. Ph.D. thesis.
[43] Reniers JM, Mulder G, Howey DA. Unlocking extra value from grid batteries using advanced models. J Power Sources 2021;487:229355.
[44] Thorgeirsson AT, Scheubner S, Fünfgeld S, Gauterin F. Probabilistic prediction of energy demand and driving range for electric vehicles with federated learning. IEEE Open J Veh Technol 2021;2:151–61. doi:10.1109/OJVT.2021.3065529.
[45] Zhou Z, Duan B, Kang Y, Shang Y, Cui N, Chang L, et al. An efficient screening method for retired lithium-ion batteries based on support vector machine. J Clean Prod 2020;267. doi:10.1016/j.jclepro.2020.121882.
[46] Li W, Rentemeister M, Badeda J, Jöst D, Schulte D, Sauer DU. Digital twin for battery systems: cloud battery management system with online state-of-charge and state-of-health estimation. J Energy Storage 2020;30:101557.
[47] European data portal. A european strategy for data. https://digital-strategy.ec.europa.eu/en/policies/strategy-data.
[48] Creative Common. Homepage of Creative Commons licenses. URL https://creativecommons.org/licenses.
[49] Open Knowledge Foundation. Homepage of open data commons open database license (ODbL). URL https://opendatacommons.org/licenses/odbl/.
[50] Open Knowledge Foundation. Homepage of database contents license (DbCL) v1.0. URL http://opendatacommons.org/licenses/dbcl/1.0/.
[51] Online identification reference spreadsheet for 18650 Li-ion cells. URL https://docs.google.com/spreadsheets/u/1/d/1fYjDxxCJXfm2wdpGWCaOUGq8V8TOEgsnplHQa4YQpRQ.
[52] Dashlink - Li-ion battery aging datasets. URL https://c3.nasa.gov/dashlink/resources/133/.
[53] Prognostics center of excellence - data repository. NASA Ames Progn Res CenterURL https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository.
[54] Hogge EF, Bole BM, Vazquez SL, Celaya JR, Strom TH, Hill BL, et al. Verification of a remaining flying time prediction system for small electric aircraft Annual conference of the PHM society (Vol.7, No.1; 2015.
[55] Bole B, Kulkarni CS, Daigle M. Adaptation of an Electrochemistry-based Li-ion Battery Model to Account for Deterioration Observed Under Randomized Use. Tech. Rep.. SGT, Inc Moffett Field United States; 2014.
[56] Olivares BE, Munoz MAC, Orchard ME, Silva JF. Particle-filtering-based prognosis framework for energy storage devices with a statistical characterization of state-of-health regeneration phenomena. IEEE Trans Instrum Meas 2012;62(2):364–76.
[57] Penna JAM, Nascimento CL, Rodrigues LR. Health monitoring and remaining useful life estimation of lithium-ion aeronautical batteries. In: 2012 IEEE aerospace conference. IEEE; 2012. p. 1–12.
[58] Zhou J, Liu D, Peng Y, Peng X. Dynamic battery remaining useful life estimation: an on-line data-driven approach. In: 2012 IEEE international instrumentation and measurement technology conference proceedings. IEEE; 2012. p. 2196–9.
[59] Saha B, Goebel K. Model adaptation for prognostics in a particle filtering framework. Int J Progn Health Manag 2011;2:61.
[60] Saha B, Goebel K. Modeling Li-ion battery capacity depletion in a particle filtering framework. In: Proceedings of the annual conference of the prognostics and health management society. San Diego, CA; 2009. p. 2909–24.
[61] Eker ÖF, Camci F, Jennions IK. Major challenges in prognostics: study on benchmarking prognostic datasets. In: Proceedings of the 1st European conference of the prognostics and health management society. Dresden, Germany, 3-5 July 2012. PHM Society; 2012. p. 148–55.
[62] Liu D, Pang J, Zhou J, Peng Y. Data-driven prognostics for lithium-ion battery based on Gaussian process regression. In: Proceedings of the IEEE 2012 prognostics and system health management conference (PHM-2012 Beijing). IEEE; 2012. p. 1–5.
[63] Tang L, Zhang B, DeCastro J, Hettler E. An integrated health and contingency management case study on an autonomous ground robot. In: 2011 9th IEEE international conference on control and automation (ICCA). IEEE; 2011. p. 584–9.
[64] Luo W, Lv C, Wang L, Liu C. Study on impedance model of Li-ion battery. In: 2011 6th IEEE conference on industrial electronics and applications. IEEE; 2011. p. 1943–7.
[65] Xiangyang G, Jun Z, Ning N. Transient behavior modeling and physical meaning analysis for battery. In: 2010 international conference on computer application and system modeling (ICCASM 2010), 2. IEEE; 2010. V2–383.
[66] Saxena A, Celaya JR, Roychoudhury I, Saha S, Saha B, Goebel K. Designing data–driven battery prognostic approaches for variable loading profiles: Some lessons learned. In: European conference of prognostics and health management society; 2012. p. 72–732.
[67] CALCE battery research group homepage. URL https://web.calce.umd.edu/batteries/data.htm.
[68] He W, Williard N, Osterman M, Pecht M. Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method. J Power Sources 2011;196(23):10314–21.
[69] Williard N, He W, Osterman M, Pecht M. Comparative analysis of features for determining state of health in lithium-ion batteries. Int J Progn Health Manag 2013;4(1):14–20.
[70] Xing Y, Ma EW, Tsui K-L, Pecht M. An ensemble model for predicting the remaining useful performance of lithium-ion batteries. Microelectron Reliab 2013;53(6):811–20.
[71] Toyota Research Institute. Experimental data platform. 2021. URL https://data.matr.io/1/.
[72] Attia PM, Grover A, Jin N, Severson KA, Markov TM, Liao Y-H, et al. Closed-loop

optimization of fast-charging protocols for batteries with machine learning. Nature 2020;578(7795):397–402.
[73] Strange C., Li S., Gilchrist R., dos Reis G.. Synthetic IR data for the Attia et al. (2020) battery dataset. https://datashare.is.ed.ac.uk/handle/10283/3798,Accessed: 2020-11-30; 2020.. 10.7488/ds/2957.
[74] Sandia National Lab. Data for degradation of commercial lithium-ion cells as a function of chemistry and cycling conditions. 2020. URL https://www.batteryarchive.org/snl_study.html.
[75] Howey D.. Oxford battery team data and code. 2011. URL http://howey.eng.ox.ac.uk/data-and-code/.
[76] Raj T.. Path dependent battery degradation dataset part 1.. URL https://ora.ox.ac.uk/objects/uuid:de62b5d2-6154-426d-bcbb-30253ddb7d1e. 10.5287/bodleian:v0ervBv6p.
[77] Raj T.. Path dependent battery degradation dataset Part 2. 2021a.. URL https://ora.ox.ac.uk/objects/uuid:be3d304e-51fd-4b37-a818-b6fa1ac2ba9d. 10.5287/bodleian:2zvyknyRg.
[78] Raj T.. Path dependent battery degradation dataset Part 3. 2021b.. URL https://ora.ox.ac.uk/objects/uuid:78f66fa8-deb9-468a-86f3-63983a7391a9. 10.5287/bodleian:j1a2eD7ow.
[79] Devie A, Baure G, Dubarry M. Intrinsic variability in the degradation of a batch of commercial 18650 lithium-ion cells. Energies 2018;11(5):1031.
[80] Battery Archive. Homepage of Battery Archive. URL https://www.batteryarchive.org/study_summaries.html.
[81] Everlasting project homepage. 2021. URL https://everlasting-project.eu/.
[82] Homepage of 4tu.researchdata repository. URL https://data.4tu.nl/.
[83] Trad K. D2.3 - Report Containing Aging Test Profiles and Test Results. Tech. Rep.. EVERLASTING; 2020. URL https://everlasting-project.eu/wp-content/uploads/2020/03/EVERLASTING_D2.3_final_20200228.pdf.
[84] Govindarajan J.. Lifecycle ageing tests on commercial 18650 Li ion cell @ 10 °C and 0 °C. 2021a.. 10.4121/14377295.
[85] Trad K.. Lifecycle ageing tests on commercial 18650 Li ion cell @ 25 °C and 45 °C. 2021a.. 10.4121/13739296.v1.
[86] Steinbuß G., Rzepka B., Bischof S., Blank T., Böhm K.. Frequent observations from a battery system with subunits. 2019b.. 37.01.03; LK 01; 10.5445/IR/1000094469.
[87] Heenan T., Jnawali A., Kok M., Tranter T., Tan C., Dimitrijevic A., et al. Lithium-ion Battery INR18650 MJ1 Data: 400 Electrochemical Cycles (EIL-015)2020a;. 10.5522/04/12159462.v1.
[88] Heenan T, Jnawali A, Kok M, Tranter T, Tan C, Dimitrijevic A, et al. An advanced microstructural and electrochemical datasheet on 18650 Li-ion batteries with nickel-rich NMC811 cathodes and graphite-silicon anodes. J Electrochem Soc 2020;167(14):140530.
[89] Gun D., Perez H., Moura S.. Berkeley: eCal fast charging test data. 2015. URL https://datadryad.org/stash/dataset/doi:10.6078/D1MS3X.
[90] Zhang S, Guo X, Dou X, Zhang X. A rapid online calculation method for state of health of lithium-ion battery based on coulomb counting method and differential voltage analysis. J Power Sources 2020;479:228740. doi:10.1016/j.jpowsour.2020.228740.
[91] Zhang S, Guo X, Dou X, Zhang X. A data-driven Coulomb counting method for state of charge calibration and estimation of lithium-ion battery. Sustain Energy Technol Assess 2020;40:100752.
[92] Zhang S.. Data for: A data-driven Coulomb counting method for state of charge calibration and estimation of lithium-ion battery, version 1. 2018. URL https://data.mendeley.com/datasets/c5dxwn6w92/1.
[93] Diao W.. Data for: Accelerated cycle life testing and capacity degradation modeling of $LiCoO_2$-graphite cells. 2021. 10.17632/c35zbmn7j8.1.
[94] Burzyński D., Kasprzyk L.. Nmc cell 2600 mah cyclic aging data v1.. 10.17632/k6v83s2xdm.1.
[95] Echevarría Cartaya Y., Sanchez L., Blanco Viejo C.. Li-ion battery charge/discharge benchmark. 2017. 10.17632/r4n22f4jfk.1.
[96] Echevarría Y, Blanco C, Sánchez L. Learning human-understandable models for the health assessment of Li-ion batteries via multi-objective genetic programming. Eng Appl Artif Intell 2019;86:1–10.
[97] Brandt D. Driving cycle testing of electric vehicle batteries and systems. J Power Sources 1992;40(1-2):73–9.
[98] EPA, U. S. E. P. A. Criteria air pollutants. 2020. URL https://www.epa.gov/criteria-air-pollutants.
[99] Cole G. A Simplified Battery Discharge Profile Based Upon the Federal Urban Driving Schedule. Tech. Rep.. EG and G Idaho, Inc, Idaho Falls, ID (USA); 1988.
[100] André M. The ARTEMIS european driving cycles for measuring car pollutant emissions. Sci Total Environ 2004;334:73–84.
[101] Kollmeyer P.. Panasonic 18650PF Li-ion battery data. 2018. 10.17632/wykht8y7tg.1.
[102] Kollmeyer P, Vidal C, Naguib M, Skells M. Lg 18650HG2 Li-ion battery data and example deep neural network xEV SOC estimator script. Mendeley Data 2020;3. doi:10.17632/cp3473x7xv.3.
[103] Chemali E, Kollmeyer PJ, Preindl M, Ahmed R, Emadi A. Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries. IEEE Trans Ind Electron 2018;65(8):6730–9. doi:10.1109/TIE.2017.2787586.
[104] Bian C, He H, Yang S. Stacked bidirectional long short-term memory networks for state-of-charge estimation of lithium-ion batteries. Energy 2020;191:116538. doi:10.1016/j.energy.2019.116538. URL http://www.sciencedirect.com/science/article/pii/S0360544219322339.
[105] Zhao R, Kollmeyer PJ, Lorenz RD, Jahns TM. A compact methodology via a recurrent neural network for accurate equivalent circuit type modeling of lithium-ion batteries. IEEE Trans Ind Appl 2019;55(2):1922–31. doi:10.1109/TIA.2018.2874588.
[106] Kollmeyer P, Hackl A, Emadi A. Li-ion battery model performance for automotive drive cycles with current pulse and EIS parameterization. In: 2017 IEEE transportation electrification conference and expo (ITEC); 2017. p. 486–92. doi:10.1109/ITEC.2017.7993319.
[107] Vidal C, Kollmeyer P, Naguib M, Malysz P, Gross O, Emadi A. Robust xEV battery state-of-charge estimator design using a feedforward deep neural network. SAE Int J Adv Curr Prac Mobility 2020;2:2872–80. doi:10.4271/2020-01-1181.
[108] Vidal C, Kollmeyer P, Chemali E, Emadi A. Li-ion battery state of charge estimation using long short-term memory recurrent neural network with transfer learning. In: 2019 IEEE transportation electrification conference and expo (ITEC); 2019. p. 1–6. doi:10.1109/ITEC.2019.8790543.
[109] Wang Y, Liu C, Pan R, Chen Z. Modeling and state-of-charge prediction of lithium-ion battery and ultracapacitor hybrids with a co-estimator. Energy 2017;121:739–50. doi:10.1016/j.energy.2017.01.044. URL https://www.sciencedirect.com/science/article/pii/S0360544217300452.
[110] Wang Y, Liu C, Pan R, Chen Z. Experimental data of lithium-ion battery and ultracapacitor under DST and UDDS profiles at room temperature. Data Brief 2017;12:161–3. doi:10.1016/j.dib.2017.01.019. URL http://www.sciencedirect.com/science/article/pii/S2352340917300197.
[111] De Craemer K., Trad K.T.. Cyclic ageing with driving profile of a lithium ion battery module2021;. 10.4121/14096567.
[112] Birkl C. Diagnosis and Prognosis of Degradation in Lithium-ion Batteries. University of Oxford; 2017. Ph.D. thesis.
[113] Jöst D., Ringbeck F., Blömeke A., Sauer D.U.. Timeseries data of a drive cycle aging test of 28 high energy NCA/C+Si round cells of type 18650. 2021.. URL https://publications.rwth-aachen.de/record/815749. 10.18154/RWTH-2021-02814.
[114] Luzi M.. Automotive Li-ion cell usage data set. 2018.. 10.21227/ce9q-jr19.
[115] He W, Williard N, Chen C, Pecht M. State of charge estimation for Li-ion batteries using neural network modeling and unscented Kalman filter-based error cancellation. Int J Electr Power Energy Syst 2014;62:783–91.
[116] Xing Y, He W, Pecht M, Tsui KL. State of charge estimation of lithium-ion batteries using the open-circuit voltage at various ambient temperatures. Appl Energy 2014;113:106–15.
[117] Zheng F, Xing Y, Jiang J, Sun B, Kim J, Pecht M. Influence of different open circuit voltage tests on state of charge online estimation for lithium-ion batteries. Appl Energy 2016;183:513–25.
[118] Planella F.B., Sheikh M., Widanage W.D. Systematic derivation and validation of a reduced thermalelectrochemical model for lithium-ion batteries using asymptotic methods. arXiv:2011.01611.
[119] Ferran Brosa Planella M.S., Dhammika W.. Systematic derivation and validation of a reduced thermal-electrochemical model for lithium-ion batteries using asymptotic methods. 2020. URL https://github.com/brosaplanella/TEC-reduced-model.
[120] Nebl C., Kotzur F.-O., Koch D., Schweiger H.-G.. Data for: prediction of constant power delivery of lithium-ion cells at high loads. 2020a.. 10.17632/ptxpzt876r.1.
[121] Nebl C, Kotzur F-O, Koch D, Schweiger H-G. Prediction of constant power delivery of lithium-ion cells at high loads. J Energy Storage 2020;30:101552.
[122] Vadhva P, Hu J, Johnson M, Stocker R, Braglia M, Brett D, et al. Electrochemical impedance spectroscopy for all-solid-state batteries: theory, methods and future outlook. ChemElectroChem 2021. doi:10.1002/celc.202100108. n/a.
[123] Firouz Y, Omar N, Timmermans J-M, Van den Bossche P, Van Mierlo J. Lithium-ion capacitor–characterization and development of new electrical model. Energy 2015;83:597–613.
[124] Zhang Y., Tang Q., Zhang Y., Wang J., Stimming U., Lee A.A.. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. 2020d.. 10.5281/zenodo.3633835.
[125] Pastor-Fernández C, Uddin K, Chouchelamane GH, Widanage WD, Marco J. A comparison between electrochemical impedance spectroscopy and incremental capacity-differential voltage as Li-ion diagnostic techniques to identify and quantify the effects of degradation modes within battery management systems. J Power Sources 2017;360:301–18.
[126] Smith A, Dahn HM, Burns J, Dahn J. Long-term low-rate cycling of $LiCoO_2$/graphite Li-ion cells at 55 °C. J Electrochem Soc 2012;159(6):A705.
[127] Raj T., Howey D.A,. Path dependent battery degradation dataset part 2. 2020. URL https://ora.ox.ac.uk/objects/uuid:be3d304e-51fd-4b37-a818-b6fa1ac2ba9d.
[128] Trad K. Calendar ageing test results on commercial 18650 Li ion cell @ 25 °C and 45 °C2021b;. 10.4121/13804304.v1.
[129] Govindarajan J.. Calendar ageing test results on commercial 18650 Li ion cell @ 10 °C and 0 °C2021b;. 10.4121/14377184.v1.
[130] Kulkarni C, Hogge E, Quach C, Goebel K. HIRF battery data set. NASA Ames Progn Res Center 2015.
[131] Bills A., Viswanathan V., Sripad S., Frank E., Charles D., Fredericks W.L.. eVTOL battery dataset. 2021.. 10.1184/R1/14226830.v1.
[132] Bills A., Sripad S., Fredericks W.L., Guttenberg M., Charles D., Frank E. et al. Universal battery performance and degradation model for electric aircraft. arXiv:2008.01527.
[133] Kulkarni C, Guarneros A. Small satellite power simulation data set. NASA Ames Progn Res Center 2015.
[134] Cameron Z, Kulkarni CS, Luna AG, Goebel K, Poll S. A battery certification testbed for small satellite missions. In: 2015 IEEE AUTOTESTCON. IEEE; 2015. p. 162–8.
[135] Reniers J.M., Mulder G., Howey D.A.. Oxford energy trading battery degradation dataset. 2020.. 10.5287/bodleian:gJPdDzvP4.
[136] Pyne M, Yurkovich BJ, Yurkovich S. Generation of synthetic battery data with capacity variation. In: 2019 IEEE conference on control technology and applications (CCTA). IEEE; 2019. p. 476–80.
[137] Pyne M, Yurkovich BJ, Yurkovich S. Capacity fade estimation using supervised

learning. In: 2017 IEEE conference on control technology and applications (CCTA). IEEE; 2017. p. 872–8.
[138] Dubarry M, Beck D. Big data training data for artificial intelligence-based Li-ion diagnosis and prognosis. J Power Sources 2020;479:228806.
[139] Dubarry M.. Graphite//LFP synthetic training diagnosis dataset. 2020a. URL https://data.mendeley.com/datasets/bs2j56pn7y/1.
[140] Dubarry M.. Graphite//LFP synthetic training prognosis dataset. 2020b. URL https://data.mendeley.com/datasets/6s6ph9n8zg/1.
[141] Dubarry M, Beck D. Analysis of synthetic voltage vs. capacity datasets for big data Li-ion diagnosis and prognosis. Energies 2021;14(9):2371. doi:10.3390/en14092371.
[142] Pastor-Fernández C, Bruen T, Widanage W, Gama-Valdez M, Marco J. A study of cell-to-cell interactions and degradation in parallel strings: implications for the battery management system. J Power Sources 2016;329:574–85.
[143] Pastor-Fernandez C.. Data for a comparison between electrochemical impedance spectroscopy and incremental capacity-differential voltage as li-ion diagnostic techniques to identify and quantify the effects of degradation modes within battery management systems. 2016. URL http://wrap.warwick.ac.uk/87247/.
[144] Song L, Zhang K, Liang T, Han X, Zhang Y. Intelligent state of health estimation for lithium-ion battery pack based on big data analysis. J Energy Storage 2020;32:101836.
[145] Song Y, Liu D, Liao H, Peng Y. A hybrid statistical data-driven method for on-line joint state estimation of lithium-ion batteries. Appl Energy 2020;261:114408.
[146] Rogall M, Barai A, Brucoli M, Luk P, Bhagat R, Greenwood D. Dremus: a data-restricted multi-physics simulation model for lithium-ion battery storage. J Energy Storage 2020;32:102051.
[147] Liu K, Ashwin T, Hu X, Lucu M, Widanage WD. An evaluation study of different modelling techniques for calendar ageing prediction of lithium-ion batteries. Renew Sustain Energy Rev 2020;131:110017.
[148] Diao W, Naqvi IH, Pecht M. Early detection of anomalous degradation behavior in lithium-ion batteries. J Energy Storage 2020;32:101710.
[149] Pei P, Zhou Q, Wu L, Wu Z, Hua J, Fan H. Capacity estimation for lithium-ion battery using experimental feature interval approach. Energy 2020;203:117778.
[150] DOE OE. Homepage of DOE OE. 2020. URL https://www.sandia.gov/energystoragesafety-ssl/research-development/research-data-repository/.
[151] NREL. Homepage of the national renewable energy laboratory of the u.s. department of energy. 2020. URL https://www.nrel.gov/research/data-tools.html.
[152] dryad. Homepage of dryad. URL https://datadryad.org/stash.
[153] zenodo. Homepage of zenodo. URL https://zenodo.org/.
[154] euhubs4data. Homepage of euhubs4data. URL https://euhubs4data.eu/datasets/.
[155] mendeley. Homepage of mendeley. URL https://data.mendeley.com/.
[156] google. Homepage of the google database. URL https://blog.google/products/search/discovering-millions-datasets-web/.
[157] V. Sulzer's online spreadsheet of battery datasets. URL https://docs.google.com/spreadsheets/d/183uKKd0JTV46tGFsfvM-OetvHHSELlL26Cetm6bJDDw.
[158] B. Shahrooei's online spreadsheet of battery datasets. URL https://docs.google.com/spreadsheets/d/10w5yXdQtlQjTTS3BxPP233CiiBScIXecUp2OQuvJ_JI.
[159] batterystandards.info. Homepage of batterystandards.info. 2021. URL https://www.batterystandards.info/intro.
[160] Battery Bits. The battery software open source landscape. 2021. URL https://medium.com/batterybits/the-battery-software-open-source-landscape-933b88957ef5.
[161] BEEP. Beep github repository. 2020. URL https://github.com/tri-amdd/beep.
[162] Herring P, Balaji Gopal C, Aykol M, Montoya JH, Anapolsky A, Attia PM, et al. Beep: a python library for battery evaluation and early prediction. SoftwareX 2020;11:100506. doi:10.1016/j.softx.2020.100506. URL https://www.sciencedirect.com/science/article/pii/S2352711020300492.
[163] cellpy. cellpy github repository. 2020. URL https://github.com/jepegit/cellpy.
[164] impedance.py. impedance.py githhub repository. 2020. URL https://github.com/ECSHackWeek/impedance.py.
[165] GitHub B.H.T.. Bayesian Hilbert transform GitHub. 2021. URL https://github.com/ciuccislab/BHT.
[166] Liu J, Wan TH, Ciucci F. A Bayesian view on the Hilbert transform and the Kramers–Kronig transform of electrochemical impedance data: Probabilistic estimates and quality scores. Electrochim Acta 2020;357:136864.
[167] Rohatgi A.. Webplotdigitizer: Version 4.4. 2020. URL https://automeris.io/WebPlotDigitizer.
[168] Qin Y., Adams S., Yuen C. A transfer learning-based state of charge estimation for lithium-ion battery at varying ambient temperatures. arXiv:2101.03704.