



Experimental investigation of power available in lithium-ion batteries

Adriano Schommer^{*}, Davide Domenico Sciortino, Denise Morrey, Gordana Collier

Faculty of Technology, Design and Environment, Oxford Brookes University, Wheatley Campus, Oxford, OX33 1HX, UK

HIGHLIGHTS

- Temperature and SOC are the most influential factors on power availability.
- Effects of stack pressure (20–60 kPa) were negligible.
- Load history can potentially influence power availability.
- Cell-to-cell variations were more significant than experimental errors.

ARTICLE INFO

Keywords:

State of power
State of charge
Stack pressure
Load history
Temperature
Design of experiments
ANOVA

ABSTRACT

The reliable and cost-effective operation of battery packs relies on state of power (SOP) algorithms to estimate the available power of the system. The challenges in developing these algorithms include the nonlinear behavior of batteries under high-power demands and the impact of temperature, state of charge (SOC), stack pressure and previous load history at high C-rates. This study employs analysis of variance (ANOVA) and design of experiments (DOE) to assess the impact of key factors on the power output of lithium-ion LCO pouch cells. The findings demonstrate that the effect of cell-to-cell variation on power output is more pronounced than degradation and random errors of the experiments. Further analysis shows that temperature and state of charge have a significant influence on power availability (p -value < 0.05), while stack pressure does not show a significant impact within the tested ranges (20–60 kPa). Notably, the load history factor approached the significance threshold with a p -value of 0.06, highlighting its potential importance in highly dynamic load profiles at increased C-rates. This research underscores the critical factors influencing battery performance and emphasizes the necessity of meticulous statistical methods in the development of accurate power estimation methods.

1. Introduction

Enabling the battery to operate at its limits is a key element for high-performance automotive applications. For this purpose, advanced battery management systems employ state of power (SOP) algorithms to estimate the maximum power a battery can source or sink within a time horizon while ensuring the operating limits are not violated. For instance, constraints on state of charge (SOC), voltage, current, and temperature underscore the importance of accurately estimating the available power of a battery [1,2]. However, developing such algorithms is challenging due to the predictive nature of power estimation, and the nonlinear behavior of batteries. Hence, conducting experiments becomes essential to comprehend the impact of various factors on the available power and to provide guidance for validating the algorithms.

Different methods to assess SOP have been extensively reported in [3]. The experiments were classified according to the power pulse employed in the validation tests. Constant current (CC), constant voltage (CV), or constant current constant voltage (CCCV), constant voltage

constant power (CVCP) pulses were reported in previous study, and throughout investigated in [4]. In [5], validation methods were further classified into static and dynamic SOP tests. Static SOP tests employ pulses from a rest condition, such as the Hybrid Pulse Power Characterization (HPPC) test [6]. Even though HPPC tests are useful for parameter identification of battery models [5–7], it computes the power available based on a simple Rint battery model, neglecting changes of SOC during the pulse and diffusion dynamics of the battery [8–10]. On the other hand, dynamic tests attempt to reproduce real-world operating conditions by inserting current pulses at random points of a drive cycle [11–13].

At the system level, the battery pack consists of many cells connected in series and parallel to meet voltage and capacity requirements. However, due to manufacturing tolerances and material differences at the microscale level, cell-to-cell variations exist, limiting the power available of the pack by its weakest cell [14,15]. That is, the cell

^{*} Corresponding author.

E-mail address: adriano.schommer@brookes.ac.uk (A. Schommer).

with the highest voltage during charge or the lowest voltage during discharge.

The cell format factor also plays a role. Pouch-format cells have a flexible aluminum casing, allowing the cell to expand and contract during cycling due to the intercalation of lithium ions within the electrodes. The lack of mechanical support makes pouch cells sensitive to external pressure. Particularly, studies have demonstrated that stack pressure increases the power output by decreasing impedance at the cost of reduced capacity and life span [16–21].

Due to the dynamics of lithium-ion transport, the power available is also affected by the previous load history of the cell. During cycling, diffusion of lithium ions occurs at a slower timescale compared to the instantaneous polarization effects. That is, the terminal voltage deviation from the battery open-circuit voltage under load [22]. Consequently, when a dynamic load is applied (i.e., a drive cycle), the voltage response at a given time instant comprises the instantaneous polarization and previous diffusion processes that are still undergoing [23]. For this reason, dynamic SOP tests validate SOP predictions by taking the load history into account. Finally, temperature and SOC significantly impact the power output of a cell. Higher temperatures increase power capabilities but lead to accelerated degradation over time. Also, for a given load, the power output is higher at higher SOC [24–26].

The effects of stack pressure, load history, SOC, and temperature on a lithium-ion cell have been reported in the literature. However, a comparative study of these factors on power available is still missing. Ranking factors by the order of impact supports efficient development of SOP algorithms: if the parameters with the most impact are known, modeling efforts are better informed. Unfortunately, extensive testing using high current loads may introduce errors to the results because of induced battery degradation. To address this challenge, the experimental framework developed in this paper sheds light on how to test the sensitivity of a battery to factors of interest while mitigating the degradation effects.

Two sets of experiments were conducted using lithium cobalt oxide (LCO) high-power pouch cells. The first experiment investigated whether cell-to-cell variations or battery degradation were the dominant effects in the tests. The second experiment quantified the impact of stack pressure, load history, SOC, and temperature on the power output.

2. Methodology

This section outlines the methods used to evaluate which factors influence the power output of a high-power pouch cell. It begins with the specifications of the tested cells, followed by a description of the experimental setup. Next, a brief introduction to the statistical methods is provided. The final part presents the Design of the Experiments (DOE) approach.

2.1. Cell specifications

The experiments were conducted on five fresh Melasta 6.8 Ah lithium cobalt oxide (LCO) cells. The cell specifications are given in Table 1.

Before the tests, the cells were preconditioned at 25 °C by five cycles at 1C discharge rate and C/2 charge rate.

2.2. Experimental setup

The experimental setup is shown in Fig. 1(a). Two 60 A load channels of an Arbin LBT-21084-HC cell cycler were used in parallel to perform the tests. The voltage measurement precision of the cycler is rated as 0.75 mV. The ambient temperature was controlled by a thermal chamber Binder KB115, in which the temperature fluctuation is expected to be within ± 0.1 °C, according to the manufacturer. Additionally, three type-T thermocouples rated to an accuracy of ± 0.5 °C

Table 1
Cell specifications.

Parameter	Specification
Cell	Melasta SLPBB142124
Cathode material	Lithium cobalt oxide (LiCoO ₂)
Form factor	Pouch
Dimensions	42 × 125 × 10.7 mm (W × H × T)
Weight	124 g
Nominal capacity	6.8 Ah
Maximum continuous charge	2 C
Maximum continuous discharge	15 C
Voltage operating range	3 V–4.2 V
Operating temperature	–20–60 °C

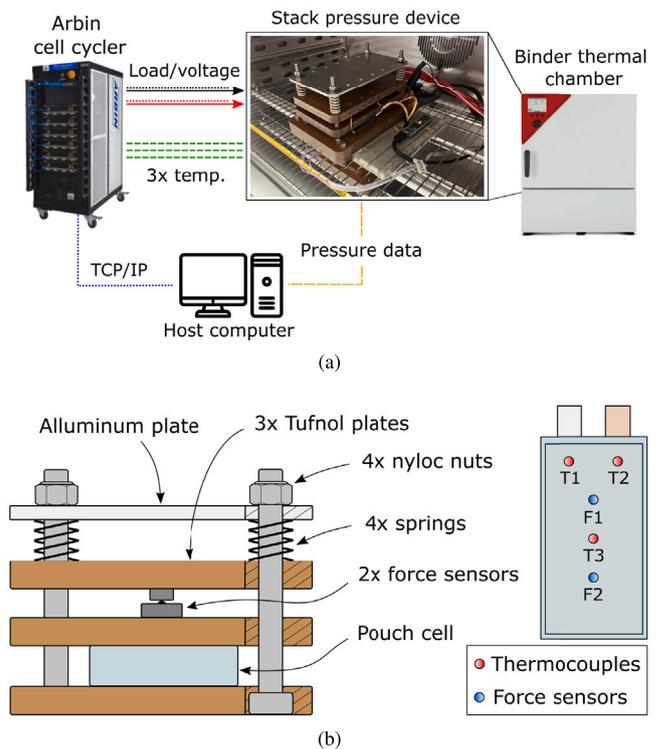


Fig. 1. (a) Experimental setup for the power tests. (b) Schematic of the stack pressure device developed to pressurize the pouch cell (left), location of thermocouples and force sensors (right).

were used to measure the cell surface temperature at the center, and near the negative and positive terminals.

Stack pressure was applied to the cell using the device illustrated in Fig. 1(b), developed based on previous work of [20,27,28]. It uses springs to compress the cell against two plates made of a Tufnol, a synthetic resin bonded laminated composite material with high mechanical and electrical insulation properties. The pressure data was computed using measurements of two TE FX29 force sensors connected to a Teensy microcontroller. Fig. 1(b) also illustrates the location of thermocouples and force sensors.

To improve the repeatability of the experiments, the load cables and voltage sensing cables setup was not changed throughout the tests [29]. The design of the pressure device allows swapping cells without disassembling the load and voltage sensing cables.

2.3. Test protocol

The goal of the experiments developed in this study was to evaluate the effects of SOC, temperature, stack pressure, and load history on power output. Therefore, the power output was the response variable

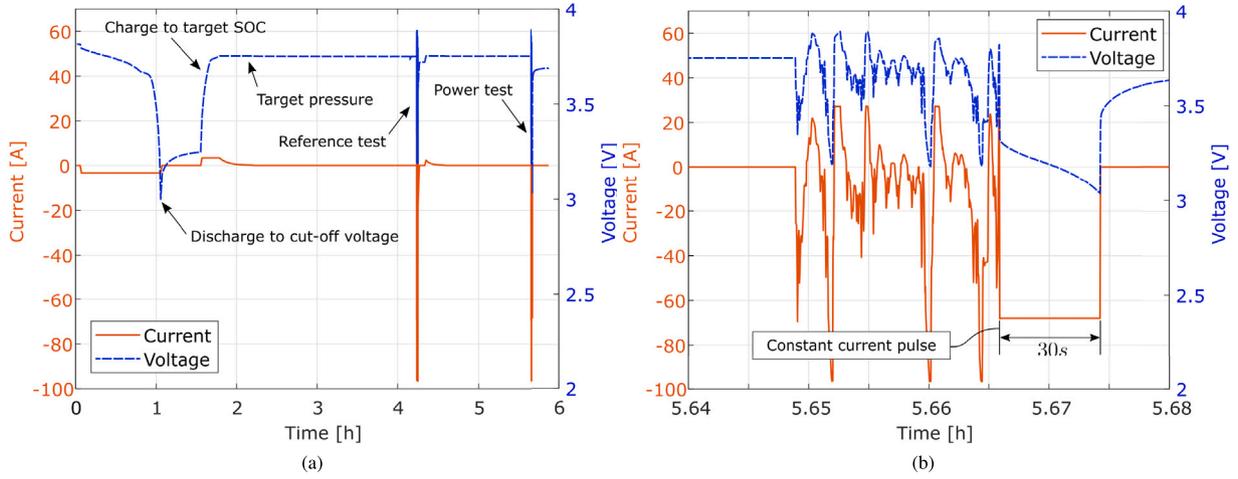


Fig. 2. Experiment to assess the power available in a cell. (a) Overview of the test schedule. (b) Zoom into the power test comprised of a drive cycle followed by a 30-second constant current pulse. Charge current is positive (regenerative braking), discharge is negative.

of the experiments and defined through Eq. (1) as the arithmetic mean power output from a constant current pulse.

$$P_{mean} = \frac{1}{n} \sum_{i=1}^n v_i \cdot I \quad (1)$$

Specifically, P_{mean} is the arithmetic mean power, n is the total number of measurements, v_i is the i th terminal voltage measurement and I is the current.

To account for the load history of the battery, a constant current pulse was inserted into a drive cycle. Because this research focuses on high-performance automotive applications, the drive cycle selected was representative of a four-wheel-drive (4WD) electric Formula Student vehicle completing one lap of an endurance event.¹ Hereafter, the combination of the drive cycle and the constant current pulse will be called the *power test*. Fig. 2 provides an overview of the test schedule (a), and a closer view of the power test (b).

The battery was discharged to the cut-off voltage before a constant-current constant-voltage (CCCV) charge to the target SOC based on a target voltage. A rest period was included to allow the open-circuit voltage (OCV) to stabilize, allow the cell to soak into the test temperature, and to calibrate the stack pressure device to the target pressure. Then, a direct current internal resistance (DCIR) test was performed to quantify the power fade of the battery [30]. Even though the internal resistance could be computed using the power pulse, a method proposed by Arbin Instruments was used to improve the accuracy of the measurements [31]. This method employs ten small amplitude charge and discharge pulses, computing the internal resistance by taking the average of the resistance measurements of each pulse. In addition, a reference load profile was added to the reference performance test (RPT) using the same drive cycle shown in Fig. 2(b). The purpose of the reference load was to establish the baseline measurements of the voltage response before the power test. The battery was then recharged to the target SOC and allowed to rest once more before the power test.

Published literature suggests that a pulse of 30 s is long enough for the polarization overvoltages caused by ionic mass transport, solid-state lithium diffusion, and charge-transfer kinetics to become significant [23,30]. Therefore, the 68 A pulse (10C) was applied for 30 s, designed to amplify the polarization overvoltages and the effect of diffusion limitation of the cell.

¹ The endurance is an autocross style racing course of the Formula Student competition. The load profile was developed by the 2022 and 2023 Oxford Brookes Racing Formula Student team using the lap time simulation software AVL VSM™ and data from racetrack testing.

2.4. Statistical methods

The analysis of variance (ANOVA) approach [32] was employed to evaluate the impact of cell-to-cell variations and cell degradation on mean power by investigating theoretically identical cells. Subsequently, the same approach was used to understand the significance of specific factors affecting mean power. The advantage of ANOVA is that it determines the statistical significance of experimental results by comparing variances. Specifically, the variance due to experimental errors (e.g., random errors and nuisance factors) is compared to the variance due to parameter changes in the experiment. In the field of statistical inference, these parameter changes are referred to as treatments. Therefore, ANOVA determines whether the observed differences in results are due to parameter changes or fall within the range of experimental errors.

The basic idea is that each test is performed multiple times (replicates) to assess the variance due to experimental errors. Then, the variance from different tests (using different experimental parameters) is computed and compared against the variance due to errors. ANOVA datasets are often presented as box plots to visualize these variances. For example, if five cells are tested to identify cell-to-cell variations, there will be five box plots, one for each cell. The range of each box plot illustrates the variance due to experimental errors, while the differences in means between the box plots indicate the variance due to differences between cells.

A formal definition of the approach is based on the decomposition of the total sum of squares SS_{Total} (total variability) into the sum of the squares $SS_{Treatments}$ due to treatments (variability between groups) and the sum of the squares SS_{Error} due to random error (variability within groups) as follows

$$SS_{Total} = SS_{Treatments} + SS_{Error} \quad (2)$$

The sum of squares (SS) measures the deviation of the data from the mean. That is, taking the arithmetic mean from all experiments of all tests, the SS represents the deviation of within groups (experimental error) or between groups (treatment) with respect to the mean. The SS provides the foundation to compute the F-ratio, which is the core of the ANOVA. The F-ratio is defined as the ratio between the mean squares due to treatments and the random error:

$$F = \frac{MS_{Treatments}}{MS_{Error}} \quad (3)$$

where the mean square due to treatments is defined as

$$MS_{Treatments} = \frac{SS_{Treatments}}{(a-1)} \quad (4)$$

Table 2
Design matrix of the experiments.

Parameter	Type	Description
Pulse arithmetic mean power	Response variable	The metric being observed in response to the manipulated design factors in the experiment.
Pulse C-rate Pulse duration	Held constant factor	Factors that are intentionally kept constant throughout the experiment to eliminate their influence.
State of charge Temperature Stack pressures Load history	Design factor	Factors deliberately manipulated in the experiment to observe their effects on the response variable.
Cell-to-cell variations Degradation	Nuisance factor	Factors that are not of primary interest but might still impact the results and need to be controlled or noted.

and the mean square due to the error as

$$MS_{Error} = \frac{SS_{Error}}{(N - a)} \quad (5)$$

where a is the number of groups, and N is the number of test replicates. The degree of freedom of treatments is $(a - 1)$, and the degree of freedom of the error is $(N - a)$.

Considering a null hypothesis, H_0 , that the means μ of different groups are the same, and an alternative hypothesis, H_1 , that at least one of the means μ is significantly different, the ANOVA tests the null hypothesis (Eq. (6)) against the alternative hypothesis (Eq. (7)).

$$H_0 : \mu_1 = \mu_2 = \mu_i \quad (6)$$

$$H_1 : \mu_1 \neq \mu_2 \neq \mu_i \quad (7)$$

Given an alpha level² (critical p -value) for accepting or rejecting the null hypothesis, if the F-ratio is significant, it indicates that at least one group has a mean significantly different from others. The F ratio (Eq. (3)) is used to calculate the p -value based on a F-distribution [32].

The p -value indicates the degree of compatibility between the statistical model and the observed data. It indicates how well the observed data matches the statistical model's predictions, not the probability of the hypothesis being true. For instance, a p -value of 0.1 shows insufficient evidence against the null hypothesis at a 0.05 significance level.

While the one-way ANOVA described so far performs well in revealing the impact of a single change of levels, it fails to provide significant information when multiple levels are tested because of the family-wise error rates. That is, the increase in the probability of incorrectly rejecting the null hypothesis (type I error, false positive), due to the compound effect of applying the alpha level in multiple comparisons³ [33,34]. To adjust the alpha levels in multiple comparisons, this work uses the Tukey–Kramer method [33] which provides a balanced approach between types I and II errors (false positives and false negatives, respectively). It performs a pairwise test of a null hypothesis that two tests have the same means, against the alternative hypothesis that the means are different.

The confidence level of hypothesis testing is influenced by the sample size (i.e., experiment repetitions), making it one of the most crucial aspects of an experiment [32]. However, the sample size is not predetermined because it depends on the system's sensitivity to a given input. For instance, if a design factor has a minor impact on the response variable, detecting its effects is more difficult, requiring more replicates to achieve the same confidence level compared to a factor

² The alpha level α is the accepted probability of erroneously rejecting the null hypothesis (type I error).

³ The family-wise error is given by the equation $1 - (1 - \alpha)^n$, where α is the alpha level (critical value) of the accepted type I error in the experiment, and n is the number of comparisons. If $\alpha = 0.05$ and $n = 10$ comparisons are made, the family-wise error is not 0.05, but $1 - (1 - 0.05)^{10} \approx 0.4$. This means the probability of false positives increases from 5% to 40%.

with a larger impact. The threshold of a small or large effect is called the minimum detectable effect (MDE). The magnitude of the MDE is inversely proportional to the sample size required for a given *power level* of the hypothesis test. That is, the probability of detecting an effect that actually exists [35]. If the MDE is small, more tests are required to detect an effect.

2.5. Design of experiments

In the field of DOE, factors that might influence the performance of the system are called *potential design factors* or *nuisance factors* [32]. In this study, SOC, ambient temperature, load history, and stack pressure were selected as *design factors*, while the magnitude and the duration of the pulse were *held constant* throughout the test matrix. Additionally, the experiments were designed to minimize degradation and the impact of cell-to-cell variations, the *nuisance factors*. Table 2 summarizes the design matrix of the experiments.

The cell degradation cannot be neglected nor controlled during the experiments. However, if the effect of degradation is small compared to other *nuisance factors* such as cell-to-cell variations, it can be treated as an *allowed-to-vary* factor, in which case the variability of performance could be balanced in the test matrix by the randomization of the test sequence [32].

The range and levels at which factors are varied or held constant during tests are defined by the aims of the experiment and by the design limitations of the cell. The SOC and ambient temperature are well defined by cell manufacturers, in contrast to stack pressure which often lacks information. According to Melasta [36], improving the performance of this cell by applying stack pressure is challenging due to trade-off with degradation. This information is in agreement with published literature, where capacity fade increased when higher levels of mechanical pressure were applied to the cell [37]. Furthermore, studies have been reported in experiments using a wide range of pressure, from 8 to 1200 kPa [17,20]. Therefore, the stack pressure levels tested in this study were chosen based on the intended application for a Formula Student vehicle. The aim was to understand the sensitivity of the cell to a small amount of pressure, here defined as <60 kPa.

The levels of SOC were selected with a bias towards the lower region of SOC, where the state of power estimation is more challenging due to the proximity to the OCV “knee-point” and the related highly nonlinear behavior of the polarization voltage response on that operating region [38]. The 10C constant current pulse discharges approximately 8% of SOC in 30 s. Considering an initial SOC of 20%, the pulse brings the SOC down to 12%, which is close enough for the overvoltages to reach the knee-point of the OCV curve. From this point, the voltage response drops abruptly towards the minimum voltage. Therefore, considering the levels of C-rate and duration of the pulse proposed, 20% SOC was the closer the experiments could get before over-discharging the cell.

The effects of load history were tested by inserting the 30-s pulse into three points (initial conditions) of the drive cycle, as presented in Fig. 3. The first initial condition (40 s) occurs at a high discharge point, the second (50 s) at a neutral point where the current is close to zero,

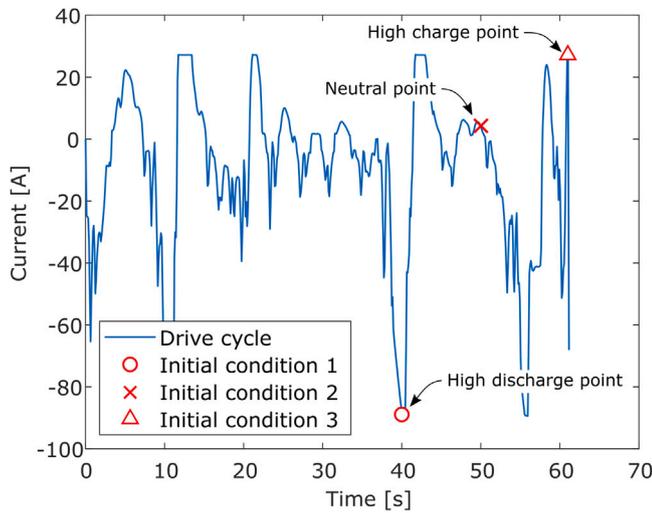


Fig. 3. Current-time plot of the Formula Student drive cycle shows the initial conditions chosen to test the influence of load history on the power available.

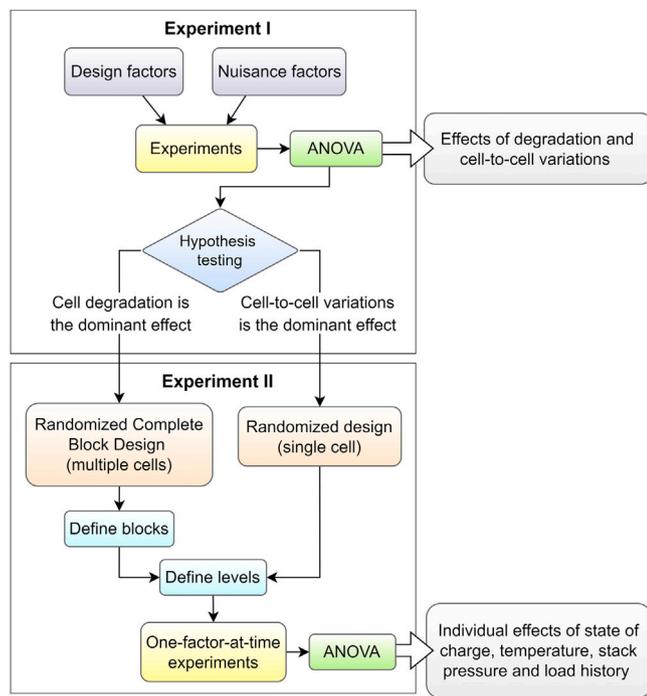


Fig. 4. Flowchart of the experiments.

and the last initial condition (61 s) contrasts with the first at a high charge point of the drive cycle.

Leveraging the sequential experimentation proposed by [32], the workflow was divided into two experiments iteratively employing small-scale experiments to test assumptions, methods, assess the influence of factors and calibrate levels at which factors are tested in the experiments. Fig. 4 provides an overview of the experiments that can be described as follows:

Experiment I – The first experiment was designed based on an ANOVA to test the hypothesis that the effects of degradation would be more significant than cell-to-cell variations. Depending on whether degradation or cell-to-cell variations is the dominant effect, different treatments are needed to mitigate the nuisance factors. If the effects of cell degradation are dominant, cell-to-cell variations are treated by a Randomized Complete Block Design (RCBD). This approach uses

different cells (blocks) for the experiments. The reasoning is that, by spreading the test matrix into multiple fresh cells, fewer tests are conducted on each cell, thus, decreasing the degradation of each cell. On the other hand, if the effects of cell-to-cell variations are dominant, the test matrix is executed using a single cell. The degradation is then treated as an allow-to-vary factor, balanced by the randomization of the test matrix [32].

Five different cells were used and the tests were repeated five times for a total of 25 tests. Different cells have slightly different capacities, which leads to slight differences of the initial SOC before each power test. However, the capacity variations were less significant than the differences in internal resistance of the cells. Therefore, the errors caused by the small variations in the initial SOC were considered as an allow-to-vary factor alongside small differences due to temperature fluctuations of the thermal chamber.

As discussed previously in Section 2.4, computing the sample size (i.e., number of test repetitions) is challenging because the standard deviation of the experiments is not known in advance. Therefore, this experiment was used as a pilot test to compute the standard deviation in order to adjust the sample size required for the second experiment.

Experiment II – Designed to assess the impact of individual factors on the power available. A one-factor-at-time (OFAT) approach was used to increase or decrease the levels of factors of each test, one at a time. This experiment was designed based on the outcomes of the first experiments, from which it was learned that the cell-to-cell variations were the dominant *nuisance factor*. Therefore, the test matrix was executed in a single cell using a randomized run order. Four factors were tested at three different levels. Based on the assessment of the sample size used in the first experiment, each test was repeated three times. Therefore, the test matrix comprised 36 tests from which 27 were unique, and nine had overlap levels with other tests in the matrix. The number of unique tests is calculated as $n = [k(L - 1) + 1]r$, where k is the number of factors, L is the number of levels, and r is the number of repetitions.

The experiments were designed in stages, adjusting the levels of factors and testing procedures based on earlier findings. The experimental data is available in [39]. Experiments I and II are presented in Sections 3 and 4, respectively. Each section will present the levels of factors in a table format.

3. Experiment I: effects of degradation, cell-to-cell variations, and sample size

This experiment aimed to evaluate the impact of cell degradation and cell-to-cell variations in the tests. Because of the high C-rates applied continuously during the power pulse (10 C, 30 s), it was hypothesized that the degradation would be the dominant nuisance factor. Therefore, the null hypothesis, H_0 , was that, given the variability of the experiments due to cell degradation, the impact of cell-to-cell variations would not be statistically relevant. Thus, the alternative hypothesis, H_1 , was that the variance introduced by using different cells would exceed the effects of degradation.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 \quad (8)$$

$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \quad (9)$$

where, μ_{1-5} are the arithmetic mean power of the pulse observed for each test repeated for each of the five cells.

3.1. Test procedure

Five identical test replicates were conducted in five fresh cells. A reference performance test (RPT) was conducted on each cell to measure cell capacity at 25 °C. The internal resistance was measured in every test according to the test protocol (Section 2.3). To assess the repeatability of the stack pressure setup, the device was completely

Table 3
Baseline level of factors for the tests.

T	P	Load history	SOC
25 °C	40 kPa	50 s (neutral point)	20%

released by unscrewing the lock nuts and re-calibrating to the target pressure before each test.

Repeatability of the tests and hence the variance within groups is affected by experiment setup errors and cell degradation. The latter might be neglected due the limited number of current loads to which each cell is exposed. However, because of the high C-rate during the power pulse, the degradation was also included in this ANOVA. Regarding the variance between groups, the effect is attributed to cell-to-cell variation.

The level of factors of the experiments are presented in Table 3. To account for the nonlinear behavior of the terminal voltage at the low SOC region, the baseline level of SOC was set to 20%. Moreover, temperature and stack pressure were set to the mid-range of values expected for the application. Similarly, the initial condition was set to the neutral point to minimize its effects (please refer to Fig. 3). In doing so, the observed response variable was focused on the effects of low SOC instead of a convoluted effect of SOC with low and high temperatures, stack pressures, or high charge and discharge load history.

3.2. Sample size

The sample size was computed based on the highest standard deviation observed in the pilot experiments ($\sigma = 0.15$ W, from cell 01). Fig. 5 illustrates the effects of the sample size and MDE on the power level of the hypothesis test. The data annotations show the MDE used to compute each curve, both in watts, and in terms of the standard deviation of the experiment. The higher the MDE, the smaller is the sample size necessary to detect the effect. For a sample size of three replicates, and a minimum detectable effect of 0.75 W, the power level of the hypothesis test was above 90%. For the sake of comparison, a variation of 0.5% of the maximum power output observed in the pilot experiments would result in ≈ 1 W. Therefore, computing the sample size based on a minimum detectable effect of 0.75 W was considered reasonable.

4. Experiment II: individual effects of state of charge, temperature, stack pressure and load history

This experiment quantifies the individual contribution of factors to power output. The experiment was based on a one-factor-at-time approach consisting of a randomized sequence of experiments, where a single factor was varied at a time regarding a baseline set of levels.

The null hypothesis, H_0 , was that the mean power output of the cell was insensitive to different levels of factors. Thus, the alternative hypothesis, H_1 , was that the variance observed was due to the changes of levels.

$$H_0 : \mu_{1,1} = \mu_{1,2} = \mu_{1,3} \quad (10)$$

$$H_1 : \mu_{1,1} \neq \mu_{1,2} \neq \mu_{1,3} \quad (11)$$

where, the subscript of the mean power $\mu_{f,l}$ represents the factor, f , and the level, l , tested.

The test procedure is presented in Section 4.1, and the results in Section 5.2.

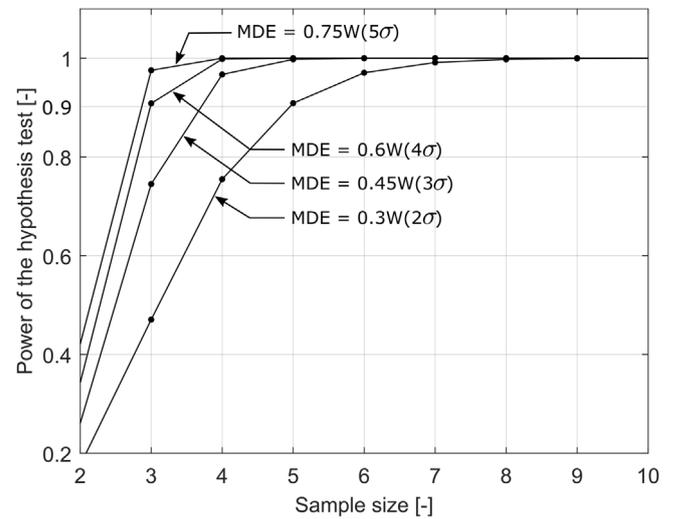


Fig. 5. Power level of the hypothesis test.

4.1. Test procedure

The pool of tests was assigned to a random run order to balance random errors and cell degradation across the test matrix [32]. The pool of tests consisted of three levels for each of the four factors. Each test was replicated three times, resulting in 27 unique tests.

The experiment started with an RPT at 25 °C to measure cell capacity. Additionally, the internal resistance was measured in every test according to the test protocol presented in Section 2.3. The internal resistance was measured at the test temperature, target SOC and stack pressure to detect changes caused by the factors. The levels tested are presented in Table 4.

Similarly to experiment I, the baseline levels and ranges of factors were designed with a bias towards the low SOC range. The range of ambient temperature was selected based on the expected range of temperatures of the battery pack. However, the experiments conducted at 15 °C resulted in voltages close to the minimum voltage limit of the cell. Therefore, temperatures below 15 °C would limit either the C-rate or the duration of the power pulse to avoid over-discharging the cell. The levels of stack pressure were selected to test the sensitivity of the cell to small variations in pressure.

5. Results

5.1. Experiment I

Fig. 6(a) shows a time-series plot of the power output of the five cells tested during the 25 power tests. This plot shows the drive cycle followed by the 30-s power pulse and provides a zoomed view of the power pulse, from where the cell-to-cell variations can be noted. The discharge power is negative, thus cell 05 has the highest power output among the cells tested. The results suggest that the effects of the cell-to-cell variations are not significant during the drive cycle, but are amplified by the relatively long duration and high C-rate of the power pulse. The zoomed part of the plot also suggests that the power output is impacted by individual characteristics of different cells, even though the discharge behavior is similar, monotonically decaying in a nonlinear fashion throughout the pulse.

The power pulse is further explored in Fig. 6(b). The plot shows the mean power output of the five cells tested during the 30-s power pulse. Each pulse was repeated five times on each cell. Therefore, each boxplot contains five data points of the mean power. The median is the

Table 4
Levels of factors for the one-factor-at-time experiments.

Factor	Levels		
	Low	Baseline	High
SOC	20%	40%	60%
T	15 °C	30 °C	45 °C
P (Force ^a)	20 kPa (105 N)	40 kPa (210 N)	60 kPa (315 N)
Load history	40 s (high discharge point)	50 s (neutral point)	61 s (high charge point)

^a Total force applied to set the stack pressure.

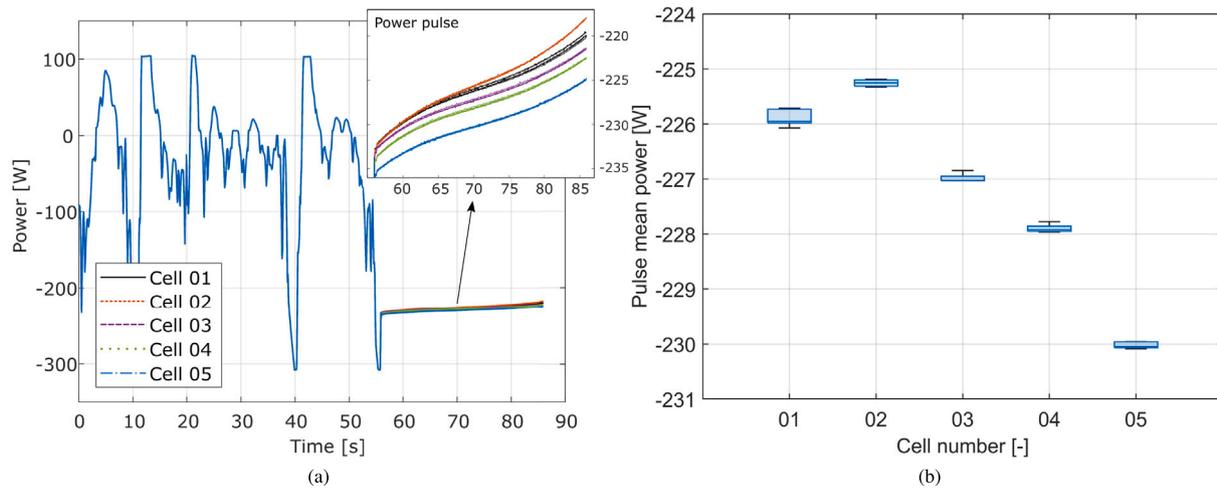


Fig. 6. (a) Time-series plot of the power test. The charge power is positive, discharge is negative. (b) Boxplot of the mean power of the pulse.

response variable of the experiment, and the upper and lower quartile represent the 0.75 quantile and 0.25 quantile, respectively. Based on the criteria of values 1.5x greater than the upper or lower interquartile range (IQR), no outliers were detected [40].

When comparing all the cells, the power output changed as much as 2.1%, where cell 05 produced the highest mean power output (−230 W), and cell 02 the lowest (−225.2 W). The range of the boxplots, identified by the distance between the maximum and minimum values, is relatively constant for tests conducted in the same cell, even though the tests performed on cell 01 had a variability 2.3 times higher than other tests, attributed to random experimental errors. The dispersion of the data can also be described in terms of the standard deviation. Tests on cell 01 resulted in a standard deviation of $\sigma = 0.15$ W, 2.2 times higher than the mean standard deviation observed on the other tests ($\sigma = 0.07$ W). The boxplot also provides insights about the distribution of the data. The power output of the cells is evenly distributed around the median (i.e., the median is centered regarding the IQRs). When comparing different cells, even the highest skewness (cell 01, $s = -0.085$) is close to zero, which indicates the data is symmetrical around the median.

Following the previous analysis, an ANOVA was conducted using MATLAB® software (version R2023b), where the variance within groups could be compared to the variance between groups. The results are summarized in Table 5, known as the ANOVA table [32]. The mean squared error (MS column) represents the error of the model fitted to the data. Therefore, the table shows that the mean squared error due to cell-to-cell variations (error between groups) is four orders of magnitude higher than the error of the model fitted to the data within groups (cell degradation and the experimental random errors). The lack of fitness of the model leads to the rejection of the null hypothesis that the power output of different cells is the same, supporting the assumption that cell-to-cell variations have a greater impact on power output than the combined effects of cell degradation and random errors of the experiment. This is demonstrated by the low p-value (<0.05) presented in Table 5.

Table 5
ANOVA results of pilot tests.

Source of variation	SS	DOF	MS	F-ratio	p-value
Cell-to-cell variations	70.03	4	17.50	2032.14	8.71e−26
Error ^a	0.17	20	0.0086		
Total	70.20	24			

^a Combined effects of cell degradation and experimental random errors.

5.2. Experiment II

The ANOVA in the second experiment was performed using Tukey's pairwise comparisons to correct the increase in the family error rate of the multiple levels tested. Table 6 summarizes the statistical analysis of the experiment. Columns "Level A" and "Level B" show the levels of the pairwise comparison. The column "A-B" provides the difference between the mean power output of Levels A and B. The lower and upper limits of the confidence intervals (CI) are given in the third and fifth columns, respectively. The tables also contain the mean power value between the levels and the resulting p-value of the one-way ANOVA.

The SOC and temperature were identified as the dominant factors influencing the cell power output. The variation of these factors (from level A to B, column "A-B") resulted in a notable change in power, e.g., with temperature variation from 15 °C to 45 °C the power change was up to 22.5 W. Moreover, the difference of the means between 20% and 40% SOC was higher than the difference between 40% and 60%. This is explained by the increased internal resistance and the lower electrical potential of the cell at lower SOC values. This finding is statistically significant, as evidenced by the rejection of the null hypothesis ($p\text{-value} < 0.05$).

In contrast, pressure and load history exhibited power variations (column "A-B") that were generally more than two orders of magnitude lower compared to the dominant factors. The p-value indicated a failure to reject the null hypothesis, suggesting that changes in these factors did not result in significant mean power variations. Additionally, the

Table 6

Pairwise comparison of the effect of factors on power output.

Factor	Level A	Level B	CI lower limit [W]	A-B [W]	CI upper limit [W]	Mean [W]	p-value
SOC	20%	40%	5.7	6.4	7.1	-232.0	3.2e-07
	20%	60%	10.0	10.8	11.5	-238.5	4.4e-09
	40%	60%	3.6	4.3	5.1	-242.8	3.7e-06
T	15 °C	30 °C	14.7	15.8	16.8	-216.3	4.7e-09
	15 °C	45 °C	21.4	22.5	23.5	-232.1	3.3e-11
	30 °C	45 °C	5.6	6.7	7.7	-238.8	2.8e-06
P	20 kPa	40 kPa	-0.52	0.24	1.0	-232.0	0.61
	20 kPa	60 kPa	-0.47	0.29	1.1	-232.3	0.51
	40 kPa	60 kPa	-0.72	0.05	0.81	-232.4	0.98
Load hist.	40 s	50 s	-1.2	-0.02	1.2	-233.2	0.99
	40 s	61 s	-2.3	-1.1	0.04	-233.2	0.06
	50 s	61 s	-2.3	-1.1	0.06	-232.1	0.06

confidence interval (CI) ranges included both positive and negative values, implying that they encompassed a null value and hence supporting the conclusion that there is no meaningful difference between the levels tested.

5.2.1. Effects of state of charge

The effect of SOC on power output is further explored in Fig. 8(a). In the zoomed-in view of the plot, it is possible to observe the higher impact of SOC at lower SOC values. The power fade at 20% SOC is also accentuated after 10 s of the pulse. Nevertheless, the power output at 40% and 60% SOC follows a similar shape, albeit the accelerated power fade at the end of the pulse is not noticeable due to the higher OCV of the cell at that SOC region.

5.2.2. Effects of temperature

Fig. 8(b) demonstrates the effect of temperature on the power output. Because of the large impact on the internal resistance, temperature was expected to contribute the most to the power output. Noting that the axis scale of the figures is the same, it is possible to observe that the impact $\pm 50\%$ of temperature changes is even higher than the changes in SOC. Furthermore, the power output decreases significantly at the lowest temperature. This phenomenon is also observed in Table 6, where the difference between 15 °C and 30 °C (column A-B) is higher than the difference between 30° and 45 °C. In addition, at the lowest temperature, the power fade is accelerated. This is noticeable in the last third of the pulse duration.

As expected, the contribution of temperature has the highest impact among the factors tested. According to Table 6, testing the cell 15 °C below the baseline temperature of 30 °C has an effect more than two times higher than if tested 15° above the baseline.

5.2.3. Effects of stack pressure

The results for stack pressure variation are presented in Fig. 8(c). The stack pressure applied does not seem to have a significant impact on the power output of the cell. The zoomed-in detail reveals a subtle trend of power increasing with the increase in stack pressure. This is aligned with the ANOVA (Table 6) where the highest variation in power output was observed by changing the pressure from 20 to 60 kPa.

5.2.4. Effects of load history

The effect of load history is presented in Fig. 8(d). To enhance the readability of the graph, the datasets are aligned by the start of the 30-s pulse. The impact of load history on the power output is negligible. The zoomed-in plot shows the marginal differences between applying the pulse after the high discharge point and a neutral point. However, the results from the second and third pairwise comparisons (Table 6) reveal an effect close to the minimum detectable effect of 1 W. In fact, the p-values are slightly above the 0.05 threshold suggesting further work is needed to better understand the load history effects in highly dynamic load profiles at higher C-rates.

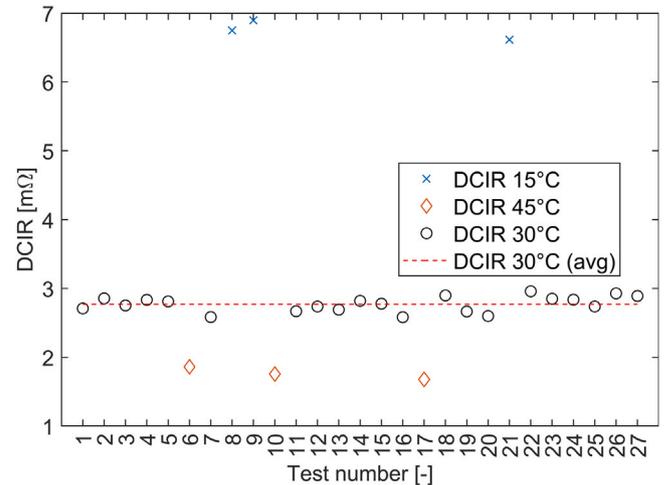


Fig. 7. Internal resistance measurements during Experiment II (matrix composed of 27 tests).

5.2.5. Internal resistance measurements

The ANOVA of Experiment II was performed based on the findings of the first experiment, which indicated that the cell-to-cell variations were more significant than the experimental random errors and cell degradation. However, because the test matrix of Experiment II was significantly larger than Experiment I, the power fade due to the increase in the internal resistance was monitored throughout the tests. Fig. 7 provides the measurements of the DCIR for Experiment II matrix (27 tests). The power fade would be characterized by an increase in the internal resistance along the tests, which was not observed. For enhanced visualization of a potential trend, the dotted line represents the average internal resistance of all tests at 30 °C. It can be noted that the data points are not trending upward, suggesting the capacity fade due to degradation was negligible throughout the test matrix.

6. Conclusion

This study presented a two-phase experimental framework that integrates ANOVA and DOE methodologies to statistically analyze the influence of SOC, temperature, stack pressure, and load history on the power output of lithium-ion pouch cells. Initial investigations aimed to mitigate nuisance variables such as cell-to-cell variability and the combined effects of degradation and random errors. The effects were evaluated using the average power output data from a 30-s pulse (10C rate) following a drive cycle for motorsport applications.

In the first experiment, five theoretically identical cells were tested to discern the relative impacts of nuisance factors on power output. A series of 25 experiments demonstrated that cell-to-cell variation

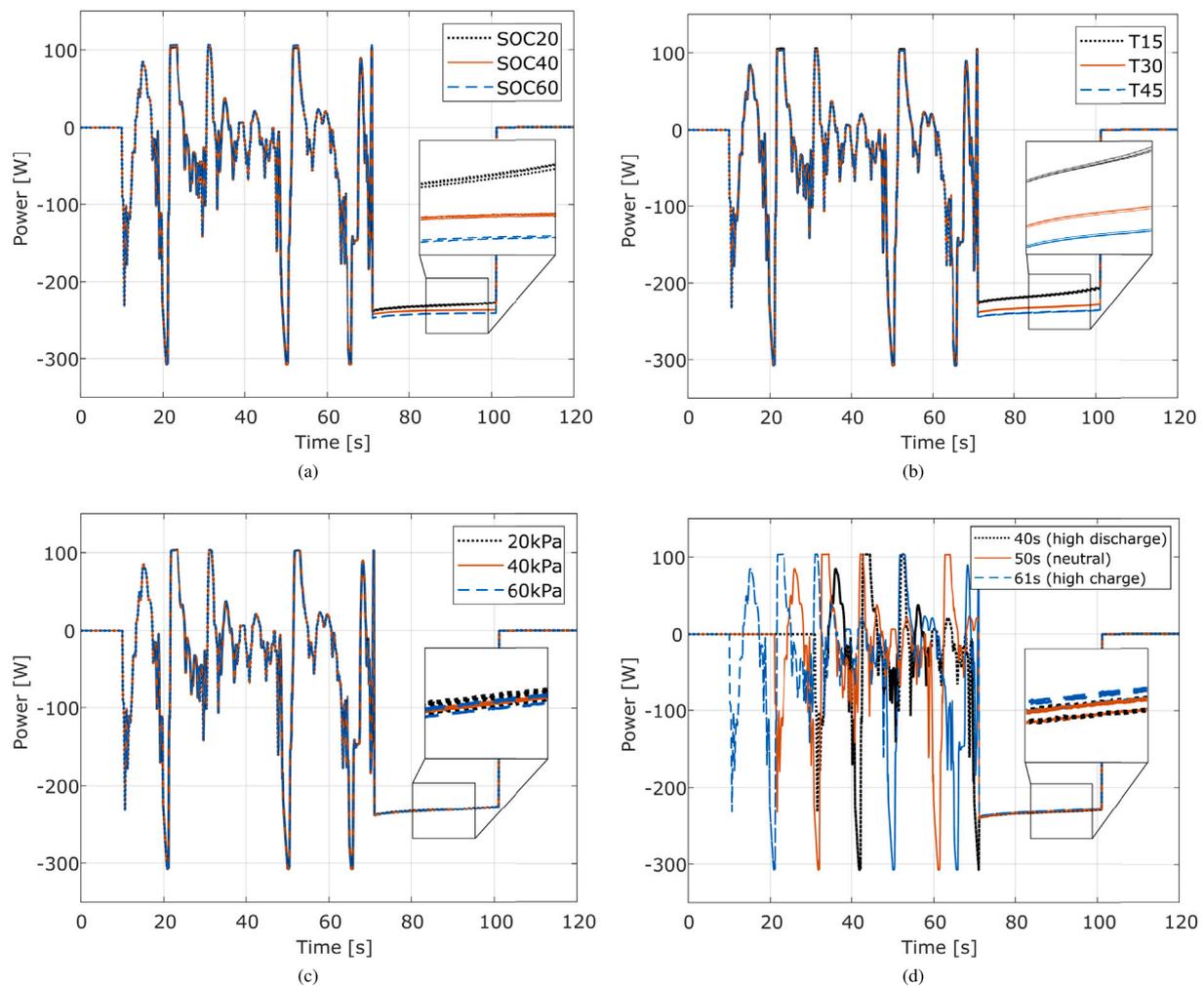


Fig. 8. (a) Effects of SOC on the power output at 30 °C and 40 kPa of stack pressure. (b) Effects of temperature on the power output with 40 kPa of stack pressure and 40% initial SOC. (c) Effects of stack pressure on the power output at 30 °C and 40% initial SOC. (d) Effects of load history on the power output at 30 °C, 40 kPa of stack pressure and 40% initial SOC.

significantly affects the power output, with statistical tests rejecting the null hypothesis of identical mean power outputs across cells.

Subsequently, the second experiment focused on a single cell, utilizing a randomized design space to control the experimental variability. This phase included three replicates for each test setting to achieve a 90% power level in hypothesis testing. Notably, the maximum standard deviation observed was 0.15 W, which is minimal relative to the overall power output range of 225 W to 230 W, suggesting a high degree of repeatability. An OFAT approach was employed, varying one of the four factors while holding others constant, to isolate the effects of each factor.

Temperature and SOC emerged as the most critical factors affecting the power output, with their variations leading to significant power discrepancies, often two orders of magnitude greater than those caused by less influential factors. Although the effect of load history was less significant, its impact on scenarios involving abrupt power shifts and high C-rates needs further exploration. Despite a p -value close to the conventional significance threshold of 0.05, further investigation into load history is recommended, particularly at higher C-rates. The stack pressure, conversely, demonstrated negligible influence within the tested ranges.

The findings underscore the importance of individual factors on the power performance of lithium-ion cells. Future works will employ a comprehensive factorial design to thoroughly explore the interdependencies between pressure, temperature, SOC, and load history. The

versatility of the proposed methodology suggests its applicability across various physical systems and their phenomenological models, providing a robust framework for exploring how calibration factors can influence model responses under diverse operational conditions.

CRediT authorship contribution statement

Adriano Schommer: Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Daide Domenico Sciortino:** Writing – review & editing, Validation. **Denise Morrey:** Resources, Project administration. **Gordana Collier:** Resources, Project administration.

Declaration of competing interest

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

Data availability

The data is available at Schommer and Domenico Sciortino (2021) [39].

Acknowledgments

The authors wish to thank the Oxford Brookes Racing Formula Student team for their support throughout this research, and the Oxford Brookes University, UK for financial and technical support.

Funding

This research received no external funding.

References

- [1] W. Han, F. Altaf, C. Zou, T. Wik, State of power prediction for battery systems with parallel-connected units, *IEEE Trans. Transp. Electrif.* 8 (1) (2021) 925–935.
- [2] W. Waag, C. Fleischer, D.U. Sauer, Adaptive on-line prediction of the available power of lithium-ion batteries, *J. Power Sources* 242 (2013) 548–559.
- [3] R. Guo, W. Shen, A review of equivalent circuit model based online state of power estimation for lithium-ion batteries in electric vehicles, *Vehicles* 4 (1) (2021) 1–29.
- [4] R. Guo, C. Hu, W. Shen, Battery peak power assessment under various operational scenarios: A comparative study, *IEEE Trans. Transp. Electrif.* (2024).
- [5] R. Guo, W. Shen, An enhanced multi-constraint state of power estimation algorithm for lithium-ion batteries in electric vehicles, *J. Energy Storage* 50 (2022) 104628.
- [6] J.P. Christophersen, *Battery Test Manual for Electric Vehicles*, Revision 3, Tech. rep., Idaho National Lab.(INL), Idaho Falls, ID (United States), 2015.
- [7] P. Malysz, J. Ye, R. Gu, H. Yang, A. Emadi, Battery state-of-power peak current calculation and verification using an asymmetric parameter equivalent circuit model, *IEEE Trans. Veh. Technol.* 65 (6) (2015) 4512–4522.
- [8] G.L. Plett, High-performance battery-pack power estimation using a dynamic cell model, *IEEE Trans. Veh. Technol.* 53 (5) (2004) 1586–1593.
- [9] L. Pei, C. Zhu, T. Wang, R. Lu, C. Chan, Online peak power prediction based on a parameter and state estimator for lithium-ion batteries in electric vehicles, *Energy* 66 (2014) 766–778.
- [10] A. Farmann, D.U. Sauer, A comprehensive review of on-board state-of-available-power prediction techniques for lithium-ion batteries in electric vehicles, *J. Power Sources* 329 (2016) 123–137.
- [11] S. Wang, M. Verbrugge, J.S. Wang, P. Liu, Power prediction from a battery state estimator that incorporates diffusion resistance, *J. Power Sources* 214 (2012) 399–406.
- [12] M. Esfandyari, V. Esfahanian, M.H. Yazdi, H. Nehzati, O. Shekoofa, A new approach to consider the influence of aging state on lithium-ion battery state of power estimation for hybrid electric vehicle, *Energy* 176 (2019) 505–520.
- [13] R. Guo, W. Shen, Toward accurate online state of power estimation for lithium-ion batteries in electric vehicles, in: 2023 IEEE 6th International Electrical and Energy Conference, CIEEC, IEEE, 2023, pp. 402–407.
- [14] C. Dangwal, D. Zhang, L.D. Couto, P. Gill, B. Sebastien, W. Zeng, S.J. Moura, Pack level state-of-power prediction for heterogeneous cells, in: 2022 American Control Conference, ACC, IEEE, 2022, pp. 1066–1073.
- [15] M. Naguib, P. Kollmeyer, A. Emadi, Lithium-ion battery pack robust state of charge estimation, cell inconsistency, and balancing, *IEEE Access* 9 (2021) 50570–50582.
- [16] J. Cannarella, C.B. Arnold, State of health and charge measurements in lithium-ion batteries using mechanical stress, *J. Power Sources* 269 (2014) 7–14.
- [17] L. Aiello, P. Ruchti, S. Vitzthum, F. Coren, Influence of pressure, temperature and discharge rate on the electrical performances of a commercial pouch li-ion battery, *Batteries* 10 (3) (2024) 72.
- [18] Y.C. Zhang, O. Briat, J.-Y. Deletage, C. Martin, G. Gager, J.-M. Vinassa, Characterization of external pressure effects on lithium-ion pouch cell, in: 2018 IEEE International Conference on Industrial Technology, ICIT, IEEE, 2018, pp. 2055–2059.
- [19] R. Li, W. Li, A. Singh, D. Ren, Z. Hou, M. Ouyang, Effect of external pressure and internal stress on battery performance and lifespan, *Energy Storage Mater.* (2022).
- [20] V. Müller, R.-G. Scurtu, K. Richter, T. Waldmann, M. Memm, M.A. Danzer, M. Wohlfahrt-Mehrens, Effects of mechanical compression on the aging and the expansion behavior of Si/C-composite| NMC811 in different lithium-ion battery cell formats, *J. Electrochem. Soc.* 166 (15) (2019) A3796–A3805.
- [21] X. Cai, C. Zhang, Z. Chen, L. Zhang, D.U. Sauer, W. Li, Characterization and quantification of multi-field coupling in lithium-ion batteries under mechanical constraints, *J. Energy Chem.* (2024).
- [22] G.L. Plett, *Battery Management Systems*, Volume I: Battery Modeling, Vol. 1, Artech House, 2015.
- [23] T. DuBeshter, J. Jorne, Pulse polarization for Li-ion battery under constant state of charge: Part I. Pulse discharge experiments, *J. Electrochem. Soc.* 164 (11) (2017) E3539.
- [24] S. Ma, M. Jiang, P. Tao, C. Song, J. Wu, J. Wang, T. Deng, W. Shang, Temperature effect and thermal impact in lithium-ion batteries: A review, *Prog. Nat. Sci.: Mater. Int.* 28 (6) (2018) 653–666.
- [25] Z. Lu, X. Yu, L. Wei, F. Cao, L. Zhang, X. Meng, L. Jin, A comprehensive experimental study on temperature-dependent performance of lithium-ion battery, *Appl. Therm. Eng.* 158 (2019) 113800.
- [26] F. Zheng, J. Jiang, B. Sun, W. Zhang, M. Pecht, Temperature dependent power capability estimation of lithium-ion batteries for hybrid electric vehicles, *Energy* 113 (2016) 64–75.
- [27] K. Lukow, B. Planden, Modular battery pressure fixture, 2021, URL <https://github.com/katielukow/MBPF>.
- [28] A. Leonard, B. Planden, K. Lukow, D. Morrey, Investigation of constant stack pressure on lithium-ion battery performance, *J. Energy Storage* 72 (2023) 108422.
- [29] J. Taylor, A. Barai, T. Ashwin, Y. Guo, M. Amor-Segan, J. Marco, An insight into the errors and uncertainty of the lithium-ion battery characterisation experiments, *J. Energy Storage* 24 (2019) 100761.
- [30] A. Barai, K. Uddin, M. Dubarry, L. Somerville, A. McGordon, P. Jennings, I. Bloom, A comparison of methodologies for the non-invasive characterisation of commercial Li-ion cells, *Prog. Energy Combust. Sci.* 72 (2019) 1–31.
- [31] A. Instruments, *MitsPro8.0 user's manual*, 2019.
- [32] D.C. Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons, 2017.
- [33] S. Lee, D.K. Lee, What is the proper way to apply the multiple comparison test? *Korean J. Anesthesiol.* 71 (5) (2018) 353.
- [34] G.M. Sullivan, R.S. Feinn, *Facts and fictions about handling multiple comparisons*, 2021.
- [35] P. Bruce, A. Bruce, P. Gedeck, *Practical statistics*, 2017.
- [36] L. Melasta, Shenzhen Battery Co., Email to nicole liu, 29 august, 2022, 2022.
- [37] J. Cannarella, C.B. Arnold, Stress evolution and capacity fade in constrained lithium-ion pouch cells, *J. Power Sources* 245 (2014) 745–751.
- [38] R. Guo, W. Shen, A data-model fusion method for online state of power estimation of lithium-ion batteries at high discharge rate in electric vehicles, *Energy* (2022) 124270.
- [39] A. Schommer, D. Domenico Sciortino, Dataset: Experimental investigation of power available in lithium-ion batteries, 2024, URL <https://doi.org/10.5281/zenodo.12543440>.
- [40] M. Frigge, D.C. Hoaglin, B. Iglewicz, Some implementations of the boxplot, *Amer. Statist.* 43 (1) (1989) 50–54.