SoftwareX 24 (2023) 101584

Contents lists available at ScienceDirect

SoftwareX

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

Original software publication

DATTES: Data analysis tools for tests on energy storage

Eduardo Redondo-Iglesias [a,c,*], Marwan Hassini [a,b,c], Pascal Venet [b,c], Serge Pelissier [a,c]

[a] Univ Lyon, Univ Eiffel, ENTPE, LICIT-ECO7 Lab, 69500 Bron, France
[b] Univ Lyon, Université Claude Bernard Lyon 1, INSA Lyon, Ecole Centrale de Lyon, CNRS, Ampère, UMR5005, 69622 Villeurbanne, France
[c] ERC GEST (Eco7/Ampère Joint Research Team for Energy Management and Storage for Transport), 69500 Bron, France

ARTICLE INFO

Keywords:
Energy storage
Experiments
Matlab
GNU octave
Data analysis
Open science
FAIR

ABSTRACT

Experiments are essential to understand the behaviour and performance of energy storage systems. In this field, a considerable amount of experimental data is generated and data processing is a tedious task. To date, research teams working in the field of energy storage tend to focus on developing their own analysis tools rather than using existing open source software. This strategy can be detrimental to the quality and reproducibility of the research. This paper presents DATTES, a free and open source software for analysing experimental battery data. The software provides a comprehensive and customizable toolkit for extracting, analysing and visualizing experimental data. It also creates gateways to other open software and tools. In this way, DATTES enables users to get the most out of their experimental data and engage in open and reproducible science.

Code metadata

Current code version	DATTES 23.05
Permanent link to code/repository used for this code version	https://github.com/ElsevierSoftwareX/SOFTX-D-23-00363
Code Ocean compute capsule	
Legal Code License	GNU GPL v3
Code versioning system used	git
Software code languages, tools, and services used	Matlab/GNU Octave
Compilation requirements, operating environments & dependencies	Matlab or GNU Octave on Linux, OSX or Windows.
If available Link to developer documentation/manual	https://dattes.gitlab.io/
Support email for questions	dattes@univ-eiffel.fr

1. Motivation and significance

Data processing is playing an increasingly important role in scientific studies. This time-consuming activity typically consumes 30 percent or more of a scientist's research time [1]. Software such as MATLAB, Python or R is commonly used, and according to a study by Hettrick et al. 70% of scientists admit that their research would not even be possible without it [2].

In the field of energy storage, there are several data processing limitations to overcome. Firstly, a wide variety of tools and methods are used to generate experimental data. As a result, it is difficult to compare the experimental data generated. In addition, most research teams have developed their own data processing tools adapted to their specific needs and data sets. Although this organization can be explained by the researchers' need to process the data from their projects quickly, such a development strategy limits the reproducibility of the work and tends to significantly increase the time spent by the researchers on programming. Furthermore, as the vast majority of researchers are self-taught developers, they may lack experience of good software development practices. As a result, the quality of the research may be compromised. Comparison between studies from two different laboratories is also made difficult because each team is likely to use its own methodology for generating and processing data. This lack of common data processing tools is a major barrier to sharing and comparing experimental results [3].

Open source software has recently gained increasing interest in the field as it allows for the sharing of programming efforts, best practices, and analysis methodologies. In the current landscape of open source software for energy storage systems, Python dominates.

* Corresponding author at: Univ Lyon, Univ Eiffel, ENTPE, LICIT-ECO7 Lab, 69500 Bron, France.
E-mail address: eduardo.redondo@univ-eiffel.fr (Eduardo Redondo-Iglesias).

https://doi.org/10.1016/j.softx.2023.101584
Received 6 June 2023; Received in revised form 13 September 2023; Accepted 10 November 2023
Available online 20 November 2023

BEEP, cellpy, impedance.py or Pybamm are some examples of popular software in the field of batteries [4–6]. They are all written in Python. This monopoly can be explained by the qualities of this versatile programming language. Python is free and open source, it emphasizes the readability of the code and it comes with a very complete library of tools for data analysis, such as Scipy, Numpy or Matplotlib. However, although this language is slowly gaining interest in the energy storage community, MATLAB remains by far the most popular language. For example, Google Scholar has indexed more than 200,000 publications that mention the keywords "battery" and "MATLAB". This is three times more than for "Python". MATLAB's popularity is due to the variety of applications it offers. The use of matrices and arrays makes it easy to use in scientific problem solving. Finally, MATLAB is the basis for Simulink, a block diagram environment for modelling, simulation, and analysis of multiphysics dynamic systems.

Despite the popularity of MATLAB, no open software for processing experimental data has been offered in this language. This article presents DATTES (Data Analysis Tools for Tests on Energy Storage), which, to the best of the author's knowledge, is the first open software written in MATLAB for processing experimental battery data. To date and at our best knowledge, DATTES is novel for several reasons:

- It is the first MATLAB software to support the entire data processing chain for battery experiments.
- The number and variety of input data formats that DATTES can handle is very large.
- It is the open source experimental data processing software that offers the largest number of analysis tools.
- It considers data traceability through used of metadata.
- It is designed to be modular and easily coupled with other existing software.

This paper presents the main characteristics of DATTES. A full detailed information on methods, requirements and assumptions could be found in the documentation [7].

2. Software description

DATTES is an open source software, written in MATLAB code and compatible with GNU Octave, that aims to facilitate data analysis for energy storage systems. As these programming languages are very popular in the field, the software can enable a large part of the energy storage community to use an open data processing tool. This section describes the main functions and features of the software.

2.1. DATTES workflow

DATTES aims to transform the raw experimental data into valuable processed results that can be used for visualization, characterization and modelling. DATTES have been designed as a modular framework which architecture can be divided into four steps: *dattes_import*, *dattes_structure*, *dattes_configure* and *dattes_analyse*.

2.1.1. Dattes import

Energy storage research suffers from data quality and format heterogeneity. The first step in this process is to standardize the experimental data. The function *dattes_import* is used to convert the raw data from a battery cycler into a standard format, the XML format following the structure defined in the open source VEHLIB [8] library.

2.1.2. Dattes structure

The second step is to structure the information and to enrich the experimental data with metadata. Metadata allow the enrichment of databases, for example during ageing campaigns, with data about each experiment (experimenter, equipment used, etc.) or about the battery (nominal characteristics, cell identifier, etc.). During this step, the function *dattes_structure* allows to divide the experiment into characteristic phases according to the different modes of battery stress: rest, constant current (CC), constant voltage (CV), etc. This function also makes it possible to determine the state of charge (SoC) of the battery at any moment of the experiment.

2.1.3. Dattes configure

In the third step of data processing, the user can adapt the analysis methodology to its needs. The function *dattes_configure* allows to modify the analysis configuration defined by default in DATTES. The default configuration consists in identifying and determining the characteristics of the batteries wherever possible. Configuring the impedance analysis method allows the user to choose to work only with current pulses of a desired duration or to select the type of equivalent circuit model to identify. In the article [9], the different equivalent circuit models that can be used in DATTES are presented. To date, the user may choose between two circuit topologies: R+RC+RC and R+CPE (Constant Phase Element).

2.1.4. Dattes analyse

Finally, the fourth step is the data analysis step. By using the function *dattes_analysis*, the user can analyse the characteristic magnitudes of a battery. The capacity can be determined in charge or discharge on a constant current phase associated or not with a constant voltage phase. The resistance of the battery can be calculated on charging or discharging current pulses with different times (R_{2sec}, R_{10sec}, etc.). DATTES enable impedance analysis by a current pulse (time-domain model identification) or by the frequency method of Electrochemical Impedance Spectroscopy (EIS). It also makes possible to analyse the behaviour of the battery at low current, either by identifying the Open Circuit Voltage (OCV) through partial charge/discharge (OCV by points) or through low current cycles (pseudo-OCV). The low current cycles can also be used for Incremental Capacity Analysis (ICA) and Differential Voltage Analysis (DVA).

2.2. DATTES auxiliary tools

2.2.1. Dattes save and dattes load

The results of the DATTES workflow can be saved to **.mat* files using the *dattes_save* function, or by adding the *'s'* execution parameter to previous DATTES operations (*dattes_structure*, *dattes_configure*, *dattes_analysis*). This allows the user to interrupt the DATTES workflow at any time and resume it later by loading the result stored in the **.mat* file using the *dattes_load* function.

2.2.2. Dattes plot

The function *dattes_plot* allows user to visualize the results of the different stages of the DATTES workflow.

2.2.3. Dattes export

The function *dattes_export* allows to store the analysis results into various files formats. This facilitate the DATTES interoperability with other existing software in the field. By default, the DATTES structure is saved as **.mat* file but other standard file formats are possible such as **.csv* or **.json* formats.

3. Illustrative examples

This section illustrates the DATTES workflow over a sample experimental dataset. The code corresponding to this example is available by running the function *demo_dattes*. Data in this dataset correspond to a characterization experiment described in preceding works [9].

The experimental raw data is output from a Biologic cycler. The function *dattes_import* allows to convert these raw data into a standard format, i.e. XML.

Then the function *dattes_structure* allows to extract the characteristic values of the battery (time, current, voltage, etc.) and to structure the

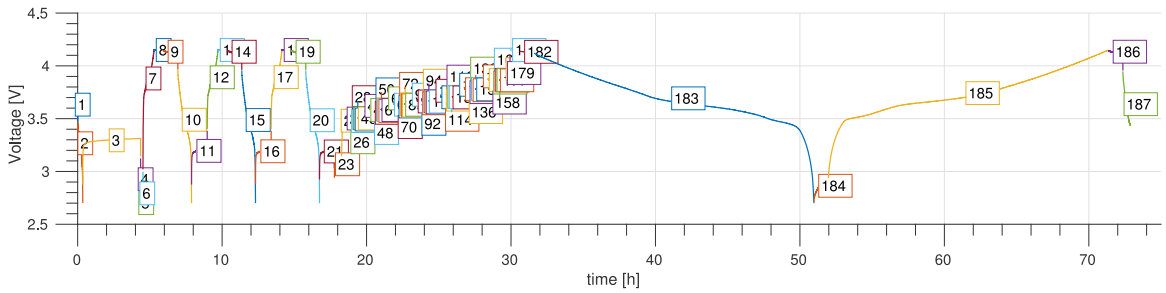

Fig. 1. Experiment decomposition into phases according to the cycler mode.

experiment in phases depending on the cycler mode (CC, CV, rest, etc.). At this level, metadata can be included in the experimental data using JSON files (**.meta* extension). Fig. 1 shows each phase number during the experiment from the previously identified modes. Dividing the experiment into phases allows the user to adapt the analysis method(s) to each part of the experiment.

In this characterization experiment, five parts were defined. Different analysis methods were applied to each part. The first part of the experiment includes phases 1 to 6. It consists of several partial discharges to ensure that the following part starts at a state of charge of 0%. The second part (phases 7 to 21) corresponds to several charge/discharge cycles to measure the battery capacity. The third part covers phases 22 to 182 and consists of a sequence of current pulses at different SoC levels (10, 20, 30, …, 90%). This part can be used to measure the battery impedance across the different SoC levels. The fourth part (phases 183 to 185) is a cycle at a low current rate. Finally, the fifth part (phases 186 and 187) is a battery partial discharge to avoid storing the battery at full charge after the experiment.

In this example, full charges are performed with the standard CC+CV protocol and full discharges are performed with CC to the minimum cut-off voltage U_{min}. Phases 7 to 20 and 183 to 185 can be used to measure battery capacity at 1C and C/20. DATTES will detect these phases in the step *dattes_configure* and this will be used in the next step *dattes_analyse* with the parameter 'C' (analyse capacity).

DATTES also allows the analysis of the equivalent resistance over a constant current pulse, when providing 'R' parameter to *dattes_analyse*.

The Eq. (1) shows that the equivalent resistance ($R_{\Delta t}$) is calculated as the ratio of the voltage drop and the current pulse amplitude at a given time after the pulse start (Δt). By default, the resistance is calculated with Δt equal to 2 and 10 s. If necessary, the function *dattes_configure* allows the user to define the value(s) of Δt of his choice.

$$R_{\Delta t} = \frac{V(t_{pulse} + \Delta t) - V(t_{pulse})}{I_{pulse}} \tag{1}$$

with

- $V(t)$: cell voltage (V),
- I_{pulse}: pulse current amplitude (A),
- $R_{\Delta t}$: equivalent resistance (Ω).

Fig. 2 shows the result of the resistance analysis. The blue and red markers show the 2-second and 10-second resistances, respectively, while the up and down triangles indicate the resistance measured during the charge and discharge pulses, respectively.

Phases 183 to 185 have been identified as relevant for characterizing the low-current behaviour of the battery. DATTES allows the analysis of the open circuit voltage (OCV), as shown in Fig. 3 (left plot). In this figure, two cycles with a current regime of 1C and C/20 have been taken into account. Thanks to the *dattes_configure* function, the user can limit the analysis of low current phases by defining a maximum current threshold above which the phases are not taken into account. Fig. 3 (right plot) shows the results of an incremental capacity analysis (ICA). For this type of analysis, the filter type is an example of a user-definable parameter [7].

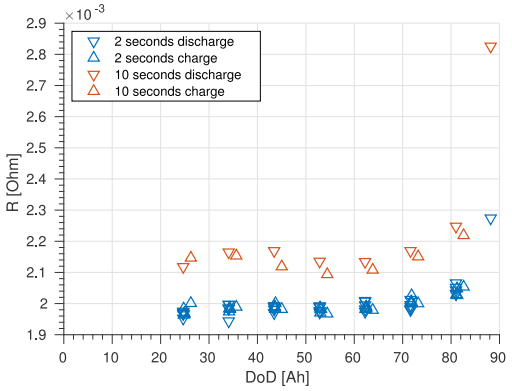

Fig. 2. Equivalent resistance at 2 and 10 s versus depth of discharge (*DoD*).

4. Impact

Several key features of DATTES are significant contributions to existing research in the field. DATTES is notable as the initial MATLAB software supporting the complete data processing chain for battery experimentation.

DATTES can handle an extensive range of input data formats, including data generated by Arbin, Bitrode, Biologic, Digatron, and Neware cyclers, while other data formats are conveniently adaptable. DATTES also converts raw experimental data into a standard format (XML), following VEHLIB library's defined structure. As a result, DATTES is natively compatible with VEHLIB, the energy management software [8].

To the best of the authors' knowledge, DATTES is an open-source software for experimental data processing that provides the highest number of analysis tools. One reason for the capability to handle various experiments is the segmentation method of experiments based on battery usage phases. This simplifies the analysis of relevant patterns for each analysis type (e.g. impedance, pseudo-OCV or ICA, etc.).

Special attention has also been paid to data traceability. Raw data are first structured and enriched with metadata using **.meta* files (*dattes_structure*), e.g. nominal cell properties, experiment date/time/equipment, etc.. After that, each step of the DATTES workflow (*dattes_configure*, *dattes_analyse*) is traceable. The result delivered by DATTES contains information about each analysis method (current levels, phase selection conditions, signal filtering parameters, etc.). The DATTES result structure contains raw experiment profiles (date, time, current, voltage, etc.), metadata, analysis results, and analysis configuration. This data structure is fully documented [7].

Finally, DATTES facilitates the reuse of the analysis data produced. The function *dattes_export* allows to convert the analysis results into different formats such as **.csv*, **.json* or **.m*.

This software has been developed according to Free and Open Source Software (FOSS) and FAIR [10] principles, and considerable effort has been made to write clear and complete documentation . The

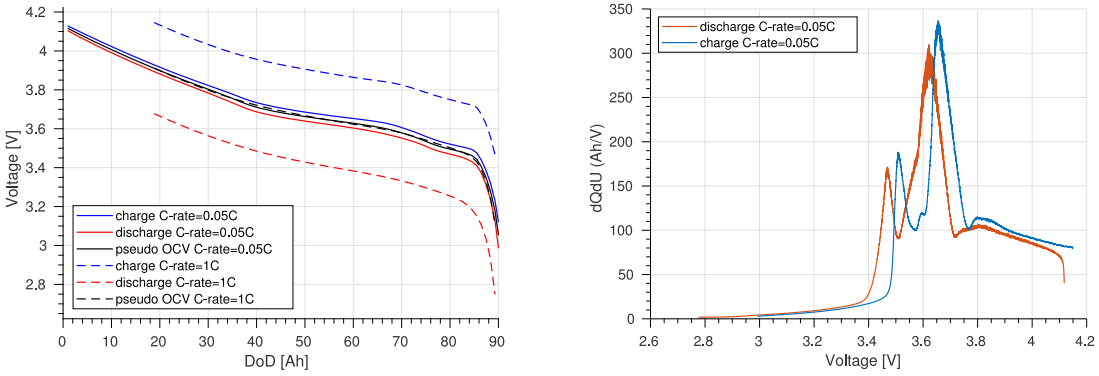

Fig. 3. Pseudo OCV (left) and ICA (right) plots. In pseudo OCV plot line colours are: red for discharge half cycle, blue for charge half cycle, black for pseudo OCV; continuous lines for C/20 cycle, dashed lines for 1C cycle. In ICA plot line colours are: red for discharge half cycle, blue for charge half cycle. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

developers hope that the energy storage community will take advantage of the software, as new users are seen as potential reviewers who can catch bugs, ask for new features, and accelerate the rise of computing standards.

Finally, DATTES is designed to be modular and easily coupled with other existing software. It can enable the energy storage community to design better batteries, build robust models, and make more accurate predictions with standardized, peer-reviewed methodologies.

5. Future work

DATTES already offers a comprehensive platform for processing battery data from experimentation to analysis and visualization. However, it is also continually evolving and future releases may introduce new features.

One potential enhancement is reliability analysis. By evaluating the quality of raw data and applied analysis methods, the reliability of resultant output could be assessed and reported within the DATTES data structure.

Another possible feature could be the linking of results from multiple tests on a single or batch of cells, for example monitoring performance over temperature or time. In the current iteration of DATTES, each processed test generates a DATTES data structure that contains raw data, metadata, and results specific to that particular test. By utilizing the metadata included in each DATTES data structure, results can be categorized based on date, cell unique identifier, or test temperature.

6. Conclusions

Processing experimental data is a central task for energy storage researchers on a daily basis. Until now, most research teams have focused on proprietary or laboratory-made tools, which severely limits the reproducibility of research. The lack of open tools available for MATLAB, the most popular language in the field, has been a significant obstacle to the widespread use of such tools. This paper introduces DATTES, the first experimental data processing software for batteries written in MATLAB code and distributed under an open licence.

DATTES contributes to existing open tools by being the first MATLAB software to support the entire data processing chain for battery experiments. This software provides a comprehensive and configurable toolkit for extracting, analysing and visualizing experimental data. The number and variety of input data formats that DATTES can handle is very large, as is the number of analysis tools it provides. DATTES is also one of the first software packages to consider data traceability through the use of metadata. The software also provides gateways to other open software and tools.

DATTES thus enables users to make good use of their experimental data and to participate in open and reproducible science.

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.1

Data availability

Data to reproduce the illustrative example is shared by link in the code.

References

[1] Hannay JE, MacLeod C, Singer J, Langtangen HP, Pfahl D, Wilson G. How do scientists develop and use scientific software? In: 2009 ICSE workshop on software engineering for computational science and engineering. IEEE; 2009, p. 1–8.

[2] Hettrick S. 2014: Software in research survey. 2018, http://dx.doi.org/10.5281/zenodo.1183562.

[3] Hassini M, Redondo-Iglesias E, Venet P. Lithium-ion battery data: From production to prediction. Batteries 2023;9:385. http://dx.doi.org/10.3390/batteries9070385.

[4] Herring P, Gopal CB, Aykol M, Montoya JH, Anapolsky A, Attia PM, et al. Beep: A python library for battery evaluation and early prediction. SoftwareX 2020;11:100506. http://dx.doi.org/10.1016/j.softx.2020.100506.

[5] Sulzer V, Marquis SG, Timms R, Robinson M, Chapman SJ. Python battery mathematical modelling (pybamm). J Open Res Softw 2021;9(1). http://dx.doi.org/10.5334/JORS.309.

[6] Murbach MD, Gerwe B, Dawson-Elli N, Tsui L-k. Impedance. py: A python package for electrochemical impedance analysis. J Open Source Softw 2020;5(52):2349. http://dx.doi.org/10.21105/joss.02349.

[7] DATTES: Data analysis tools for tests on energy storage. URL https://gitlab.com/dattes/dattes/.

[8] VEHLIB. URL https://gitlab.univ-eiffel.fr/eco7/vehlib.

[9] Hassini M, Redondo-Iglesias E, Venet P. Second-life batteries modeling for performance tracking in a mobile charging station. World Electr Veh J 2023;14(4):94. http://dx.doi.org/10.3390/wevj14040094.

[10] Jiménez RC, Kuzak M, Alhamdoosh M, Barker M, Batut B, Borg M, et al. Four simple recommendations to encourage best practices in research software. F1000Research 2017;6. http://dx.doi.org/10.12688/f1000research.11407.1.