Enhanced Validation Methods and Benchmark

J. Kamsamrong, E. Widl, J. S. Schwarz, K. Heussen, P. Raussi, G. Arnold, A. De Paola, Z. Feng, O. Werth, M. C. Pham, and Q. T. Tran

Abstract This chapter introduces enhanced validation methods and benchmarks for cyber-physical energy systems, focusing on reproducibility and uncertainty man-

---

J. Kamsamrong (✉) · J. S. Schwarz · O. Werth
OFFIS – Institute for Information Technology, Oldenburg, Germany
e-mail: jirapa.kamsamrong@offis.de

J. S. Schwarz
e-mail: jan.soeren.schwarz@offis.de

O. Werth
e-mail: oliver.werth@offis.de

E. Widl
AIT Austrian Institute of Technology, Vienna, Austria
e-mail: edmund.widl@ait.ac.at

K. Heussen
Technical University of Denmark, Kongens Lyngby, Denmark
e-mail: kheu@dtu.dk

P. Raussi
VTT Technical Research Centre of Finland, Espoo, Finland
e-mail: petra.raussi@vtt.fi

G. Arnold
Fraunhofer Institute for Energy Economics and Energy System Technology IEE, Kassel, Germany
e-mail: gunter.arnold@iee.fraunhofer.de

A. De Paola
Joint Research Centre of the European Commission, Ispra, Italy
e-mail: antonio.de-paola@ec.europa.eu

Z. Feng
University of Strathclyde, Glasgow, UK
e-mail: zhiwang.feng@strath.ac.uk

M. C. Pham · Q. T. Tran
French Alternative Energies and Atomic Energy Commission, Paris, France
e-mail: minh-cong.pham@cea.fr

Q. T. Tran
e-mail: quoctuan.tran@cea.fr

T. I. Strasser et al. (eds.), *European Guide to Smart Energy System Testing*,
SpringerBriefs in Energy, https://doi.org/10.1007/978-3-031-99451-7_3

agement. It presents benchmarks for electrical networks, multi-energy networks, and ICT-enhanced power systems, facilitating comprehensive testing and validation. Additionally, it highlights the contributions of the ERIGrid 2.0 project in developing these benchmarks and methods to support open science and improve system-level validation.

1 Introduction

CPES are complex, involving multiple domains and dependencies between different systems. This complexity poses significant challenges in managing the variability for testing and validation. Benchmarks serve as valuable tools for students, researchers, and industry practitioners by facilitating the testing of CPES. Existing reference models and lab setups work for specific small-scale experiments but face challenges for testing system-level hypotheses, for upscaling to real systems, or for extending to analyses to other domains (e.g., from the electricity grid to heat or communications). Another common challenge in validation is the recognition and declaration of uncertainties, which often hinders result reproducibility. A systematic approach for analyzing and accounting of uncertainty factors would mediate this challenge.

This chapter introduces the benchmark models developed in ERIGrid 2.0 and methods toward reproducible systems validation for CPES. The developed benchmarks and methods facilitate the methodical design and documentation of experiments, supporting open science and the further development of open-source paradigms [7].

2 Cyber-Physical Energy Systems Benchmarks

2.1 Concept and Approach

Modern energy systems have become increasingly complex, integrating multiple domains and introducing interdependencies. The shift from a centralized to a decentralized structure has underscored the need for ICT integration to enable real-time monitoring, enhanced control, and efficient data processing. This is crucial not only within a single sector but also across interconnected systems (e.g., Power-to-Heat) ensuring seamless energy flow management across domains. This advancement in energy systems creates the need for improving existing benchmarks and developing new benchmarks representing CPES for testing and validation.

The first question that may arise is: What are the differences between a model and a benchmark? Several power system models are published and available, such as the SIMBENCH, CIGRE, and IEEE families of reference models. *A model* refers to a mathematical or computational representation of a system (sometimes already implemented as source code or in binary form) for simulation and analysis. *A benchmark* in this context is a reference or standardized model used for performance evaluation and comparison on a certain application context, while *Benchmarking* refers to the process of comparing solutions on the benchmark or reference model. Thus, benchmarks can be valuable tools for knowledge consolidation through comparative analysis, as well as for testing and validation in CPES.

The next question then emerges: Which benchmark is suitable for testing in CPES? Existing models address primarily the power system, which may be insufficient when addressing the interdependencies for the CPES. In the context of CPES, functional scenarios serve as an umbrella that describes system functionalities, motivations, use cases, TCs, experimental setups, and their relevance. These scenarios provide a foundation for defining high-level requirements necessary for testing and simulation. High-level use cases can help to define motivation and requirements, while scenarios help outline possible system behaviour and interactions. Several test cases can be developed following the six Functional Scenarios introduced in chapter “Holistic Smart Energy System Validation”, representing key technological areas for CPES. Although it is not feasible to have benchmarks that directly present every TC, relevant benchmarks can potentially be adapted to relevant TCs. Considering the development of CPES, which integrates ICT, power electronics, renewable energy sources, and sector coupling, three major benchmark categories are defined:

1. *Electrical Network*, representing traditional and modern power grids, including distributed energy integration [1].
2. *Multi-Energy Networks*, covering sector coupling, such as power-to-gas, power-to-heat, and integrated energy systems [2].
3. *ICT-Enhanced Power Systems*, addressing the role of communication networks, cybersecurity, and data exchange in modern power grids [3].

These benchmarks provide a structured foundation for testing and validation across different domains, ensuring comprehensive assessment and adaptability to emerging CPES challenges.

2.2 Electrical Network Benchmark

The Electrical Network Benchmark has been designed to provide a versatile testing platform for low-voltage grids in modern power system studies with high penetration of power electronics and renewable energy resources. Several iterations of the model have been considered to maximize the model’s flexibility and strike an optimal trade-off between complexity and simulation speed. The benchmark is derived from several functional scenarios to address, especially electrical issues associated with converters operation, integration of renewables, and control schemes. Differently from the other proposed model, this benchmark focuses on implementations on a single software platform. The functional scenarios relevant for this benchmark are FS1 related to DER ancillary services, FS2 related to microgrids and energy communities, FS4 related

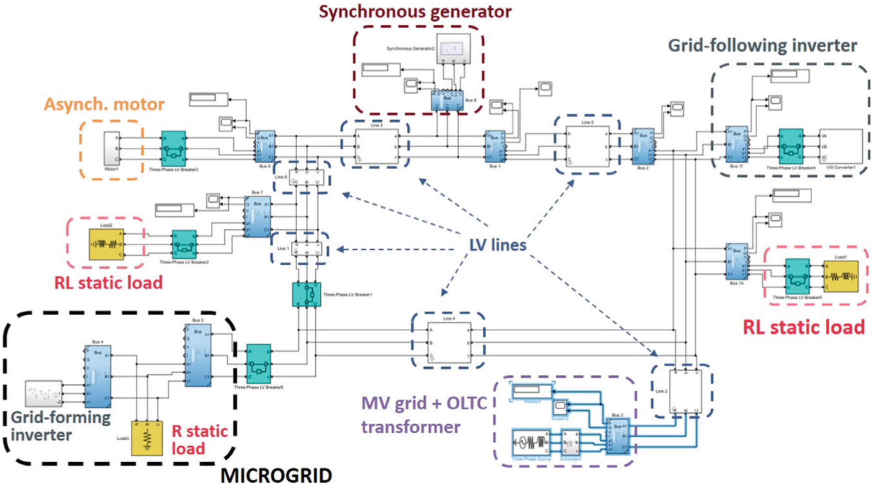


Fig. 1 Overview of the electrical network benchmark

to frequency and voltage stability, and FS5 related to aggregation and flexibility management (see also chapter "Holistic Smart Energy System Validation").

The benchmark includes a sufficiently complex network with an adequate number of buses and lines, circuit breakers for analyzing disconnections and topology changes, and inverters as controllable voltage sources with control schemes. The latter can also be used to represent a microgrid and assess the islanding capabilities of realistic systems. Additionally, at least two distinct DER types, such as PV systems and batteries, must be included to evaluate their interactions with synchronous generation. A distribution MV/LV transformer with an on-load tap-changer should also be incorporated to analyze its performance under varying system conditions. This has resulted in the inclusion of the following components in the benchmark, as shown in Fig. 1.

Examples of how this benchmark can be used include, but are not limited to:

- Testing the grid-forming capabilities of microgrid inverters.
- Voltage control with on-load tap changer controller.

The interested reader can find a detailed description of the model design, documentation and testing in [6].

2.3 Multi-energy Network Benchmark

The Multi-Energy Networks benchmark is a reference setup for sector coupling, where a power-to-heat facility connects a low-voltage electrical grid to a local heating network. An overview of the system configuration is presented in Fig. 2.

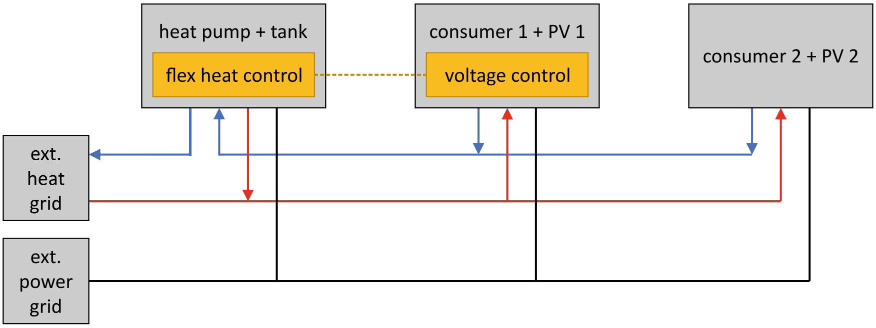


Fig. 2 Overview of the multi-energy networks benchmark [4]

Excess PV generation is used to improve grid stability and support thermal supply. From the perspective of the functional scenarios, this benchmark clusters together all sector coupling-related topics and specifically considers co-simulation for the assessment of the system. The relevant functional scenarios for this benchmark include FS2 related to microgrids and energy communities and FS3 related to sector coupling (see also chapter "Holistic Smart Energy System Validation").

The motivation of this benchmark is to promote research and development in thermal-electrical sector coupling by offering a not-too-complex yet practical reference model. It encourages co-simulation for the analysis of multi-domain systems (power, heat, control) and provides two implementation approaches. Unlike traditional simulation benchmarks, it does not provide numerical validation but serves as a conceptual guide.

Examples of how this benchmark can be used include, but are not limited to:

- Characterizing the availability of power-to-heat services and assessing their impact on the networks.
- Verifying the enhanced self-consumption of renewable energy sources (RES) in a coupled heat and power network through the use of power-to-heat technology.
- Scaling analysis of the heat storage considering sector-coupling effects [9].

2.4 ICT-Enhanced Power Systems Benchmark

The traditional power system has been transformed into a CPES, introducing interdependencies between electric power and communication systems. The performance of one system can directly impact the other. For example, a robust communication network can enhance data exchange in the power grid, while risks can also propagate between them, such as cascading communication failures leading to wide-area blackouts. As fast, reliable bidirectional communication is mandatory for real-time monitoring and control, this benchmark serves as a reference for a CPES setup consisting of the power system with ICT-based communication. It represents the impact of ICT

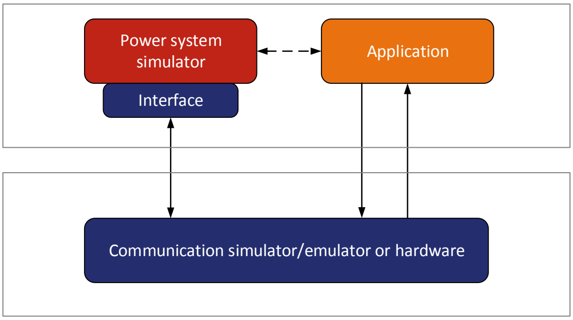


Fig. 3 General system architecture of the ICT-Enhanced power system benchmark

on power systems, focusing on communication networks that could hinder the control algorithms of power system components. The benchmark is aligned with functional scenario FS6 (see chapter "Holistic Smart Energy System Validation") on digitalisation and also encompasses other functional scenarios where ICT infrastructure or communication networks are essential for electric power system operation, enabling the coordination of distributed assets, substation automation, interoperability, and cybersecurity.

A general simulation setup for this benchmark is shown in Fig. 3, which consists of an electric power system simulator, communication simulator, tool/platform for hosting the application, co-simulation tool for information exchange and data connectivity between power system simulator, communication simulator and the application platform.

The use case for this benchmark involves but is not limited to:

- Real-time co-simulation models for power system and ICT interactions such as coordination of distributed assets, voltage control and protection.
- Assessing the reliability and latency of communication networks in CPES.
- Evaluation of communication performance for new CPES components and associated communication technologies.

3 Toward Reproducible Systems Validation for CPES

Two fundamental challenges to reproducibility are complexity and uncertainty. Complexity in CPES increases in several ways. For one, domain-extension introduces new cross-domain interaction phenomena that are not possible to reproduce and represent without including all relevant dynamics across several domains of expertise. Further, complexity results from a trend of scaling up by larger numbers of devices and subsystems (scaling out), which challenges the controllability of systems but also challenges the validation methods and tooling. Important practices for achieving reproducibility have been introduced above: the cross-domain benchmarks, the documentation of testing procedures using the HTD (see chapter “Holistic Smart Energy System Validation” and [11]), PreCISE for documentation of model details [5, 8], along with making models available as open-source implementations. Due to the complexity of CPES, another crucial topic for improving reproducibility is the handling of uncertainty in testing and validation, which needs to be addressed by specific methods.

In the following, the repertoire of methods and outline of a framework to tackle variability, sensitivities, scalability, and domain extension, which is described in more depth in [9, 15], is extended.

3.1 Uncertainty Concepts in Testing

Different types of uncertainty occur in the context of experiments. Those can result from a lack of knowledge (epistemic uncertainty) or from intrinsic fluctuations (aleatory uncertainty) and appear at different stages, for example, associated with input data, models, measurements, or the laboratory equipment. An approach to tackle such uncertainties can be described by three steps [15]. First, in the *uncertainty analysis* the uncertainties have to be identified and classified. Part of this first step, one can asses relative importance of uncertainties by quantifying the sensitivity of output measures to the uncertain factors. Second, in the *uncertainty representation*, a suitable approach for mathematically describing the uncertainty has to be found. Third, in the *Uncertainty Quantification (UQ)* the effect of uncertainties on the output of a system is calculated based on different approaches.

For smaller systems, an *analytical approach* might be suitable for UQ. A mathematical description of the system is needed, which can be used to propagate uncertainty through the system. For complex CPES, this is usually not possible, and *sampling approaches* might be more suitable. Those are repeating an experiment multiple times with changing input values and can be used for UQ and *Sensitivity Analysis (SA)*. SA here refers to a versatile set of methods, including local and global approaches, which can also be used for scaling analysis, as well as for creating an importance ranking of factors to identify the most impactful ones for more detailed investigation. More accurate UQ can be performed by explicitly propagating the uncertainty through a system. But this approach requires more knowledge about the uncertainties as described in [16] for stationary simulation models in the context of the co-simulation framework `mosaik`.

An extension of the HTD approach (see chapter “Holistic Smart Energy System Validation” and [11]) was developed to more directly integrate uncertainty analysis (see Sect. 2.3 and [10, 12]). This extension contains the Uncertainty Structure Analysis Tool (USAT), which is an Excel template with 4 structured sheets supporting the different steps of identification and prioritization of uncertainties within an experiment. The USAT integrates features to support the prioritization of factors for both uncertainty and scaling analysis purposes. Additionally, the template was extended in the following aspects:

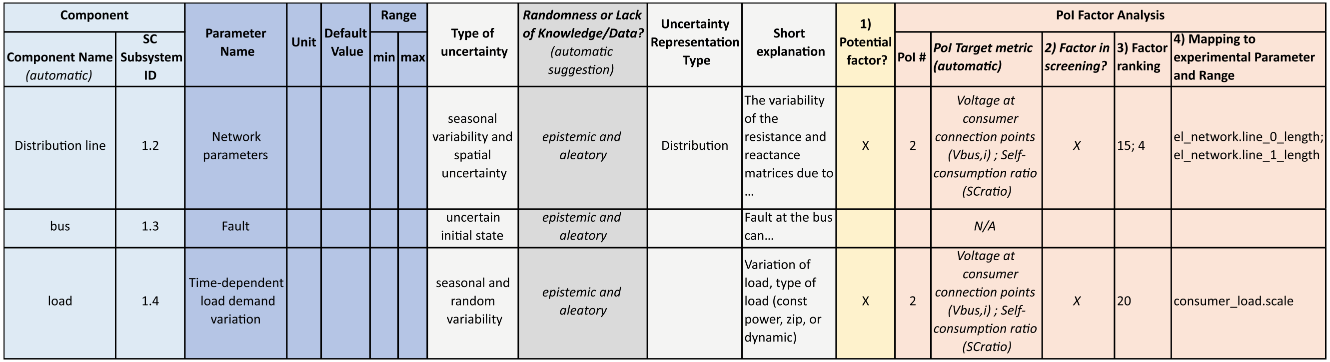
- *Test Case (TC)*: New field for detailed Purpose of Investigation (PoI) and factor analysis alongside existing variability and quality attributes.
- *Qualification Strategy (QS)*: Includes uncertainty identification and management.
- *Test Specification (TS)*: Merges input/output parameters with sources of uncertainty, linking to the detailed USAT.
- *Experiment Realisation*: Links to USAT to assess uncertainty trade-offs.
- *Experimental Specification (ES)*: New fields for experimental setup uncertainties, precision of equipment, measurement uncertainty, and uncertainty management, and is also linking to USAT.

Based on the generic domain-independent method of description of System Configurations (SCs) introduced with the original HTD, the USAT is also well-suited to analyse multi-domain configurations. The scaling-up analysis is demonstrated in [9, 10].

3.2 Tool Chain

For the facilitation of the approaches described in the previous section, tools were developed in ERIGrid 2.0. Especially, the extension of the HTD for handling uncertainty aspects as described in Sect. 2.3 and a Design of Experiments (DoE) toolbox [14] can play together to implement a holistic tool chain for handling uncertainties in experiments. The DoE toolbox aims to support users in UQ and SA of experiments by providing a structure for the object-oriented description of the parameterization and variations and performing sample generation based on this. It, therefore, provides a complete parameterization for the recommended experiment runs and also supports in analysis of the results and the effects of the variations on the inputs. The usage of the USAT is explained on the multi-energy networks benchmark (see Sect. 2.3). In Fig. 4, the SC sheet is shown, which contains a list of parameters from the multi-energy benchmark.

For each parameter, information regarding its uncertainty can be collected, and the impact of the different parameters can be ranked with the help of the DoE toolbox, as described in [13]. Exemplarily shown are the distribution lines and the load, which were chosen as potential factors for the factor analysis (i.e., yellow column “(1) Potential factor?”), mapped to concrete experimental parameters in the implementation of the multi-energy networks benchmark (i.e., red column “(4) Mapping to experimental parameter and range”), and the results of a screening analysis was inserted (i.e., red column “(3) Factor ranking”). With those steps, the DoE toolbox can be used together with the HTD process. The USAT and DoE toolbox have the potential to be directly integrated in the future so that the parameterization for the DoE toolbox can be directly exported from the USAT. Such automation would also support the potential systematic investigation of quantitatively more complex configurations that arise, especially for scaling out test scenarios.



Component		Parameter Name	Unit	Default Value	Range		Type of uncertainty	***Randomness or Lack of Knowledge/Data?*** *(automatic suggestion)*	Uncertainty Representation Type	Short explanation	1) Potential factor?	PoI Factor Analysis				
Component Name *(automatic)*	**SC Subsystem ID**				**min**	**max**						**PoI #**	***PoI Target metric (automatic)***	***2) Factor in screening?***	**3) Factor ranking**	**4) Mapping to experimental Parameter and Range**
Distribution line	1.2	Network parameters					seasonal variability and spatial uncertainty	*epistemic and aleatory*	Distribution	The variability of the resistance and reactance matrices due to …	X	2	*Voltage at consumer connection points (Vbus,i) ; Self-consumption ratio (SCratio)*	*X*	15; 4	el_network.line_0_length; el_network.line_1_length
bus	1.3	Fault					uncertain initial state	*epistemic and aleatory*		Fault at the bus can…			*N/A*			
load	1.4	Time-dependent load demand variation					seasonal and random variability	*epistemic and aleatory*		Variation of load, type of load (const power, zip, or dynamic)	X	2	*Voltage at consumer connection points (Vbus,i) ; Self-consumption ratio (SCratio)*	*X*	20	consumer_load.scale

Fig. 4 Uncertainty structure analysis tool (USAT) SC parameter

4 Way Forward

Researchers use laboratory experiments and computer simulations to test hypotheses, but such results often lack reproducibility. The ERIGrid 2.0 project addressed these challenges by providing guidelines on handling uncertainties and scalability to enhance testing and validation for CPES. The developed methods and toolchain have been demonstrated and offer opportunities for further integration and automation. To support the analysis of scaling-out cases, for example, configurations will need to be created using algorithmic approaches. The HTD, USAT, and generic system configuration approach offer a useful semantic basis for developing interoperable and scalable toolchains for test specification, experimentation and analysis.

The three benchmarks and validation methods were developed in ERIGrid 2.0 following the FAIR principles ensuring findability, accessibility, interoperability, and reusability [17]. However, there remains room for improvement, particularly in enhancing interoperability. Future work should focus on advancing semantic interoperability and benchmark compatibility with multiple simulation platforms towards seamless integration across different research environments. These principles are not only supporting standardization efforts, but they are also fostering long-term sustainability and collaboration within the energy research community. Ongoing initiatives towards FAIR energy systems research will provide services and guidelines for the community that can serve as an orientation for benchmark compatibility.

References

1. Erigrid2/benchmark1 (2025). https://github.com/ERIGrid2/benchmark-model-electrical-network. Accessed 20 Mar 2025
2. Erigrid2/benchmark2 (2025). https://github.com/ERIGrid2/benchmark-model-multi-energy-networks. Accessed 20 Mar 2025
3. Erigrid2/benchmark3 (2025). https://github.com/ERIGrid2/benchmark-model-electrical-ict. Accessed 20 Mar 2025
4. CES A, Widl E (2021) ERIGrid2/benchmark-model-multi-energy-networks: v1.0. https://doi.org/10.5281/zenodo.5735005
5. D'Arco S, De Paola A, Widl E, Rajkumar VS, Kamsamrong J, Raussi P, Arnold G, Thomas D, Marinopoulos A, Wild CW, Heussen K, Rikos E, Hoang TT, CortÃs-Borray AF (2022) D-jra1.1 benchmark scenarios. Technical report. https://doi.org/10.5281/zenodo.4032691
6. De Paola A, Thomas D, Kotsakis E, Marinopoulos A, Masera M, Paspatis A, Kontou A, Kotsampopoulos P, Hatziargyriou N (2022) Benchmark models for low-voltage networks: a novel open-source approach. In: 2022 open source modelling and simulation of energy systems (OSMSES), pp 1–6. https://doi.org/10.1109/OSMSES54027.2022.9769097
7. De Paola A, Thomas D, Paspatis A, Widl E, Marinopoulos A, Kotsakis E, Kontou A, Kotsampopoulos P, Hatziargyriou N (2023) Testing-oriented development and open-source documentation of interoperable benchmark models for energy systems. IEEE Open J Ind Electron Soc 4:42–51. https://doi.org/10.1109/OJIES.2023.3234698
8. Widl E (2020) An overview of the precise approach used for the definition of simulation and optimization workflows. Technical report, AIT. https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5cd070ea9&appId=PPGMS

9. Heussen K, Schwarz JS, Cortés A, Feng Z, Nikoletatos J, Pham MC, Tran QT, Adrinanesis P, Kamsamrong J (2024) D-JRA1.3 methods for test upscaling and domain extension. https://doi.org/10.5281/zenodo.8081457
10. Heussen K, Schwarz JS, Schulte E, Feng Z, Ramos Perez LE, Nikoletatos J, Pröstl Andrén F (2025) Uncertainty annotations for holistic test description of cyber-physical energy systems. In: PowerTech 2025
11. Heussen K, Steinbrink C, Abdulhadi IF, Nguyen VH, Degefa MZ, Merino J, Jensen TV, Guo H, Gehrke O, Bondy DEM, Babazadeh D, Pröstl Andrén F, Strasser TI (2019) Erigrid holistic test description for validating cyber-physical energy systems. Energies 12(14):1–31.https://doi.org/10.3390/en12142722
12. Raussi P, Pröstl Andrén F, Heussen K, Edmund W (2024) D-N4.3 extended test description method. https://doi.org/10.5281/zenodo.14221590
13. Schwarz JS, Pham MC, Tran QT, Heussen K (2023) Scaling analysis in a multi-energy system. In: 2023 Asia meeting on environment and electrical engineering (EEE-AM), pp 01–06. https://doi.org/10.1109/EEE-AM58328.2023.10395068
14. Schwarz JS, Ramos Perez LE, Pham MC, Heussen K, Tran QT (2024) A toolbox for design of experiments for energy systems in co-simulation and hardware tests. In: 2024 open source modelling and simulation of energy systems (OSMSES), pp 1–7. https://doi.org/10.1109/OSMSES62085.2024.10668967
15. Schwarz JS, Schulte E, Heussen K, Nikoletatos J, Ramos L, Feng Z (2024) D-JRA1.2 methods for holistic test reproducibility. https://doi.org/10.5281/zenodo.8081442
16. Steinbrink C (2017) A non-intrusive uncertainty quantification system for modular smart grid co-simulation. Carl von Ossietzky Universität Oldenburg, Ph.D
17. Wilkinson MD, Dumontier M, Aalbersberg IJ et al (2016) The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3(1):160018. https://doi.org/10.1038/sdata.2016.18