Bedřichov 2018

Smart Infrastructures & Big Data Research

Bruno Rossi

brossi@mail.muni.cz
Department of Computer Systems and Communications,
Lasaris (Lab of Software Architectures and Information Systems)
Masaryk University, Brno

lasaris
www.lasaris.cz

Context of Research

Smart Infrastructures

Multidisciplinary research on complex cyberphysical systems

Tomáš Pitner
Head of the lab

Jan Rosecký

Bruno Rossi

Radka Lomitzká

Katarína Hrabovská

Tomáš Havelka

Martin Schvarcbacher

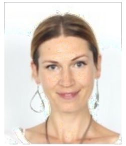
David Kvapil

Filip Procházka

Jan Herman

Big Data Analysis

Building knowledge in Big Data Analysis in different domains (e.g., Smart Grids, Bioinformatics, Cybercrime)



Barbora Bühnová
Head of the Big Data Analysis group

Jan Herman

Bruno Rossi

Bangui Hind

Tomáš Rebok

Martin Macák

Mouzhi Ge

Research Goals

Smart Infrastructures: **RG:** support the SG infrastructure by means of tests/simulations/experiments:

→ we reviewed different testing process frameworks that could be applied to the context of SG (e.g. ISO/IEC/IEEE 29119);

→ we looked at different aspects of simulations (co-simulations) by means of several frameworks that can be adopted;

→ we started working on a software platform that could allow management of SG tests/simulation/experiments (including both hardware and virtualized devices)

Big Data Analysis: **RG:** support SG data analytics by means of large-scale infrastructure. Main focus is on anomaly detection:

→ we reviewed the whole area of SG data analysis to get an overall view and started collaborations;

→ we gained experience with the Metacentrum infrastructure;

→ we are looking at alternative big data architectures for anomaly detection in SG data;

Focus of the presentation

I will focus this presentation on two articles:

→ M. Schvarcbacher, K. Hrabovská, B. Rossi, T. Pitner (2018). *“SGTMP: Smart Grid Testing Management Platform”* (submitted to journal) **(Smart Infrastructures)**

→ B. Rossi, S. Chren (2018). *“Smart Grids Data Analysis: A Systematic Mapping Study”* (submitted to journal, pre-print: https://arxiv.org/abs/1808.00156) **(Big Data Analysis)**


EXPERIENCE REPORT

SGTMP: Smart Grid Testing Management Platform

Martin Schvarcbacher[2] | Katarína Hrabovská[1] | Bruno Rossi[1] | Tomáš Pitner[1]

[1]Faculty of Informatics, Masaryk University, 602 00 Brno, Czech Republic

[2]Faculty of Science, University of Amsterdam, Spui 21, 1012 WX Amsterdam, The Netherlands

Correspondence
Bruno Rossi, Faculty of Informatics, Masaryk University, 602 00 Brno, Czech Republic
Email: brossi@mail.muni.cz

Funding information

The Smart Grid (SG) is nowadays an essential part of the modern society, providing two-way energy flow and smart services between providers and customers. The main drawback is the SG complexity, with a SG composed of multiple layers, with devices and components that have to communicate, integrate and cooperate as a unified system. Such complexity brings challenges for ensuring proper reliability, resilience, availability, integration, security of the overall infrastructure.

In this paper, we introduce a new Smart Grid Testing Management Platform (SGTMP) for executing real-time hardware-in-the-loop SG tests and experiments that can simplify the testing process in the context of interconnected SG devices. We discuss the context of usage, the system architecture, the interactive web-based interface, the provided API, and the integration with co-simulations frameworks to provide virtualized environments for testing. Furthermore, we provide one main scenario about SG devices stress testing that can showcase the applicability of the platform.

KEYWORDS
Smart Grid Testing Platform, Smart Meter, ISO/IEC/IEEE 29119 Software Testing Standard, Hardware in the Loop, Co-Simulation Frameworks

1 | INTRODUCTION

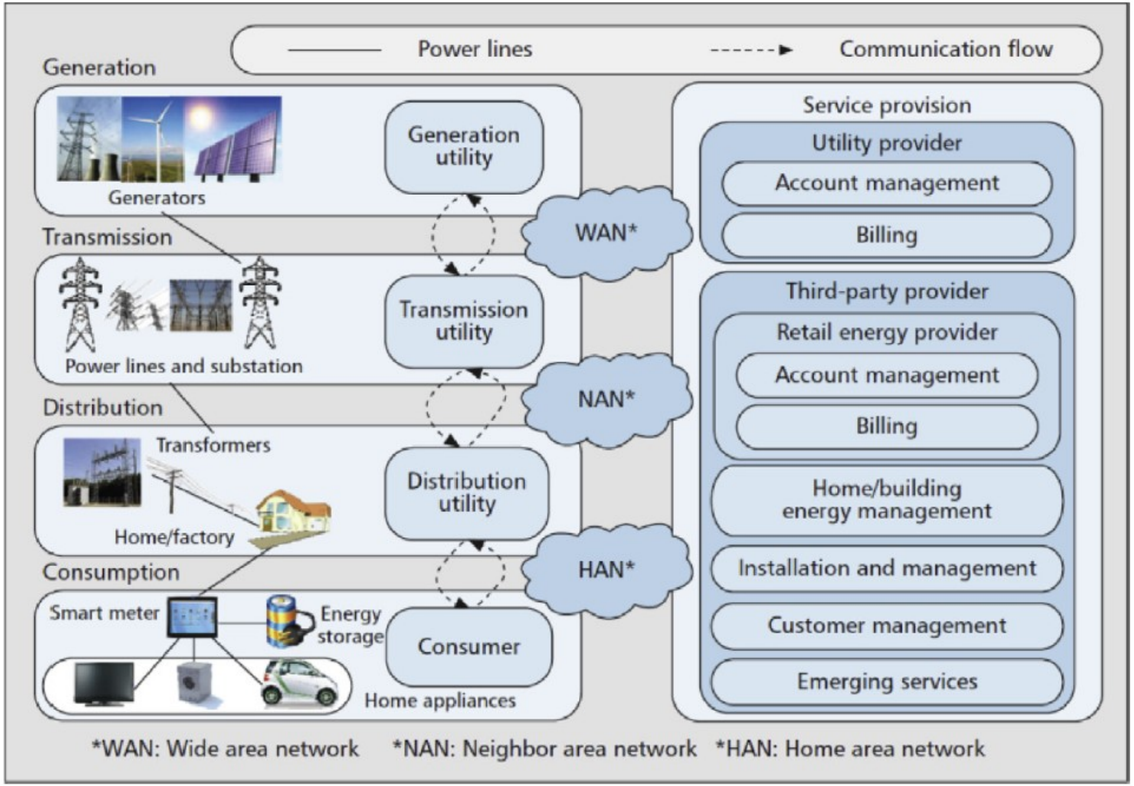
The traditional power grids distributed and managed energy from a centralized system in one direction, from the energy provider to the customers[1]. With growing demands for reliability and efficiency of the grid operation, there has been a necessity to develop a more interactive, interconnected and dynamic grid model[2]. As a result, Smart Grids (SGs) improve electricity management by involving advanced two-way communications and operational capabilities, which

Abbreviations: SG, Smart Grid; SGTMP, Smart Grid Testing Management Platform; SM, Smart Meter; SMDC, Smart Meter Data Concentrator; AMM, Advanced Metering Management; HIL, Hardware-In-the-Loop.

SGTMP: Smart Grid Testing Management Platform

(Smart Infrastructures)

Smart Grid



A Smart Grid can be seen from multiple points of view:

From one side, it can be seen as the integration of different devices, communication and IT infrastructure to provide an economically **efficient, sustainable power system**

From another side, it can be seen as **Collaborative Services Ecosystem**

Li, Weilin, and Xiaobin Zhang. "Simulation of the smart grid communications: Challenges, techniques, and future trends." Computers & Electrical Engineering 40.1 (2014): 270-288.

SG Testing/Simulation needs

SGTMP fullfills the requirements of the **provision of a configurable**, **GUI-supported environment to allow cyber-physical systems testing and simulations***

Further, it is meant to support **common needs in SG testing/simulations** and the support of **ISO/IEC/IEEE 29119 testing standard**:

Reference	Criteria	Support Requirements
Kok et al. [20]	Power flow	Real (1:1, scaled); simulated
Kok et al. [20]	Data flows	Power grid only; information grid only; combined
Kok et al. [20], Karnouskos and Holanda [21]	Interaction capture	RT capture&monitoring; large data volume; simulation playback
Karnouskos and Holanda [21], Wang et al. [22]	Topological changes	Before test; at simulation start; during runtime, multiple changes
Karnouskos and Holanda [21]	Multi-agent systems	One entity; breakdown into components
Karnouskos and Holanda [21]	Simulator integration	Well defined API; extensibility
Karnouskos and Holanda [21], Hahn et al. [23]	Entity classification	Power producer/consumer/transporter; state reporter; network intruder; SCADA
Hahn et al. [23]	Network requirements	Network analysis; packet injection; expose to simulated intruders
Wang et al. [22]	Topology generation	Automatic; determine if model generalizes; model future SG deployments
Wang et al. [22]	Testing platform	Support different SG topologies

* Steinbrink C, Schlögl F, BabazadehD, Lehnhoff S, Rohjans S, Narayan A. Future perspectives of co-simulation in the smart grid domain. In: 2018 IEEE International Energy Conference (ENERGYCON) IEEE; 2018.

Initial prototype

Smart Grids Co-Simulations with Low-Cost Hardware

Martin Schvarcbacher and Bruno Rossi
Faculty of Informatics
Masaryk University, Brno, Czech Republic
Email: {schvarc,brossi}@mail.muni.cz

***Abstract*—Smart Grids have nowadays gained wide diffusion and relevance. Due to the complexity of the grid, many Smart Grids laboratories have emerged over the years to provide partially virtualized environments for testing and co-simulation testbeds for the modern grid. However, the costs for setting-up Smart Grids laboratories are substantial, representing a barrier for newcomers and for educational purposes. In this paper, we propose an hardware-in-the-loop (HIL) architectural solution based on Arduino and Raspberry PI boards, supported by the Mosaik framework to simulate different Smart Grids scenarios on a small and cost-effective scale. We highlight the educational benefits that the solution can bring for understanding simulations and HIL in an affordable & effective way in an easy-to-deploy environment.**

***Keywords*-Smart Grids, Smart Meters, Hardware in the Loop, Co-Simulations, Cyber-Physical Systems.**

I. Introduction

A Smart Grid has been defined as a form of electricity network that enables "intelligent" integration of all the actions and behaviours of the connected actors, to efficiently deliver sustainable, economic and secure electricity supplies [1], [2].

While modern Smart Grids have ambitious aims, they also pose several challenges, mainly due to the multidisciplinary nature of the area, ranging from power equipment to needs in terms of data analysis to increase the "smartness" of the power grid. As such, communication between the different involved roles is fundamental, to the point that the education of students to several aspects of the grid is seen as one of the main challenges in the area [3]. For this reason, a recent trend is the emergence of Smart Grids laboratories that can serve not only to test Smart Grids software and devices, but also to educate students to the real needs of large-scale Smart Grids in a controlled environment [4].

However, average costs of setting-up a Smart Grid laboratory are in the order of €2M, reaching €30M for larger laboratories [4]. Such values represents a serious barrier for setting-up new laboratories for educational purposes. For this reason, in this paper we propose a virtualized and low-cost environment that students can use to test and validate different Smart Grids scenarios. Such environment can be a first step for looking into hardware-in-the-loop (HIL) and co-simulation environments—environments that are focused on orchestrating several simulations running on different devices, combining also software simulations [3], [5]. The importance of simulations is relevant in the area of Smart Grids due to the complexity of the different layers and sub-systems involved. Simulations can help in tackling away some of the complexity, by having simulation models that run in their own runtime environment.

The proposed solution is based on the Raspberry Pi and Arduino platforms that can be used to test a co-simulation environment for Smart Grids. We focus in particular on two scenarios that can be relevant for the presented prototype: i) sunlight level scenario, to simulate sunlight levels, and ii) a load scenario to predict power usage over time.

The paper is structured as follows. Section II presents the related works, in terms of other low-cost hardware-software solutions that have been proposed to simulate the Smart Grids infrastructure. Section III presents the details of building the prototype integrating hardware devices (Raspberry Pi and Arduino), software components (Mosaik) and the proposed architecture. Section IV proposes examples of usages with two main scenarios. Section V presents evaluation and discussion and Section VI concludes the paper.

II. Related Work

The usage of low-cost hardware/software for Smart Grids testing and validation has acquired more interest in recent years, mainly due to the availability of cheap devices that can be used for the purpose.

Commodity Hardware for Smart Grids—typically Raspberry Pis and Arduinos—has been used for a wide range of applications, for example to enable Smart Meters readings (voltage and users' power consumption) to be remotely transmitted [6], to test self-healing capabilities for multi-agent based approaches [7], or for network reconfiguration in secondary substations [8].

Aurilio et al. (2014) deployed a Raspberry Pi as data concentrator in a low-cost solution for the management and control of a power network based on power meters, monitoring connected loads by communicating with a data concentrator (Raspberry Pi) via Power Line bus [9].

However, while commodity hardware has been used in different parts of the Smart Grids infrastructure, our work is more focused in the area of co-simulations.

Armendariz et al. (2014) developed a platform for co-simulation based on a real-time power system simulator (Opal-RT), and a communication network emulator (OPNET). Rasp-

---

Smart Grids Co-Simulations with Low-Cost Hardware

lasaris
Lab of Software Architectures and Information Systems FI MU

Martin Schvarcbacher and Bruno Rossi

FACULTAS INFORMATICAE UNIVERSITAS MASARYKIANA

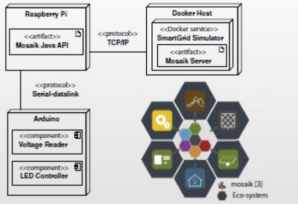
Project Context

We showcase a low-cost environment that students can use to test and validate different Smart Grids scenarios. Such environment can be a first step for looking into hardware-in-the-loop (HIL) and co-simulation environments - environments that are focused on orchestrating several simulations running on different devices, combining also software simulations. This solution can be used for understanding simulations and HIL in an affordable and effective way in an easy to deploy environment [1].

Goals

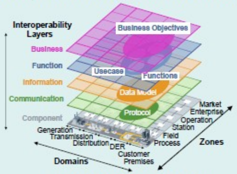
- ★ Education of students in co-simulation concepts using easily accessible and hands-on training in Smart Grids technologies
- ★ Creating ways for cheaper hardware prototyping of Smart Grids by having low-cost simulation nodes

Architecture



Smart Grids & Lasaris

- ★ The Smart Grid can be regarded as an electricity network that benefits both from two-way cyber-secure communication technologies and computational intelligence for electricity generation, transmission, substations integration and consumption to reach the goals of a safe, secure, reliable, resilient, efficient, and sustainable infrastructure [4].
- ★ Lasaris is involved in research on Smart Grids with industrial partners:
  - ➢ Supporting Smart Grids testing/simulation infastructure
  - ➢ Data analysis for Smart Grids (load control, anomalies detection)



Smart Grid Architecture Model [5, 6]

Scenarios

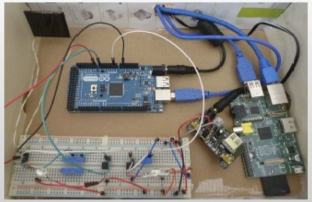
a) Sunlight Levels for a Location

- ★ Uses past weather data to estimate sunlight levels
- ★ Allows evaluating different PV panel deployments
- ★ Each node represents a PV power station in the grid

b) Power Grid Load

- ★ Knowing whether the power grid can meet the current or near future requirements becomes necessary as more intermittently available renewable resources are added to the power grid [2]
- ★ The amount of power produced is compared to expected grid load to determine power deficiencies when using only renewable resources
- ★ Used to determine when power plants need to be switched on to supplement renewable energy sources
- ★ Predicting the required production capacity can be beneficial for the Smart Grid stability

Simulation Node



Hardware Node Components

- ★ Photo-voltaic (PV) panel:
  - ➢ Produces power proportional to the illumination levels
- ★ LED array:
  - ➢ Generates multiple illumination levels
- ★ Arduino Mega:
  - ➢ Controls the LED array and reads the voltage level from a PV panel
  - ➢ Sends measured data and receives control commands
- ★ Raspberry Pi:
  - ➢ Data collection and network communications

Project Results

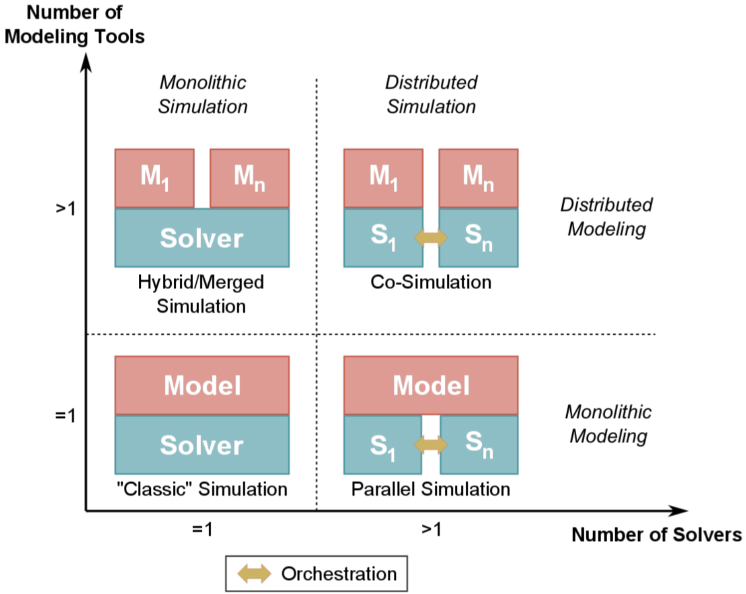
- ★ We created a platform for Smart Grid deployment prototyping
- ★ Students can easily setup their own Smart Grid environments and test them under various changing conditions
- ★ Test cases include: Smart Grid deployment, interoperability, stability
- ★ Our future goal is a full power grid simulation using only commodity hardware

References

[1] M. Schvarcbacher and B. Rossi, "Smart Grids Co-Simulations with Low-Cost Hardware," 43rd Euromicro Conference on Software Engineering and Advanced Applications, Vienna, 2017 (to appear).
[2] S. Chren, B. Rossi and T. Pitner, "Smart grids deployments within EU projects: The role of smart meters," 2016 Smart Cities Symposium Prague (SCSP), Prague, 2016, pp. 1-5.
[3] O. Nannen et al., "Low-cost integration of hardware components into co-simulation for future power and energy systems," IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society, Yokohama, 2015, pp. 5304-5309.
[4] B. Rossi, S. Chren, B. Buhnova and T. Pitner, "Anomaly detection in Smart Grid data: An experience report," 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, 2016, pp. 2313-2318.
[5] CEN-CENELEC-ETSI, Smart Grid Coordination Group, "Smart Grid Reference Architecture," 2012.
[6] M. Uslar et al, "Standardization in smart grids: Introduction to IT-related methodologies, architectures and standards," Springer Science & Business Media, 2012.

What are co-simulations?

"**Co-simulation** is defined as the coordinated execution of two or more simulation models that differ in their representation as well as in their runtime environment"*



Common terminology

→ **Emulation** (integrated or co-simulated): emulated component mimics the the real world hardware counterpart

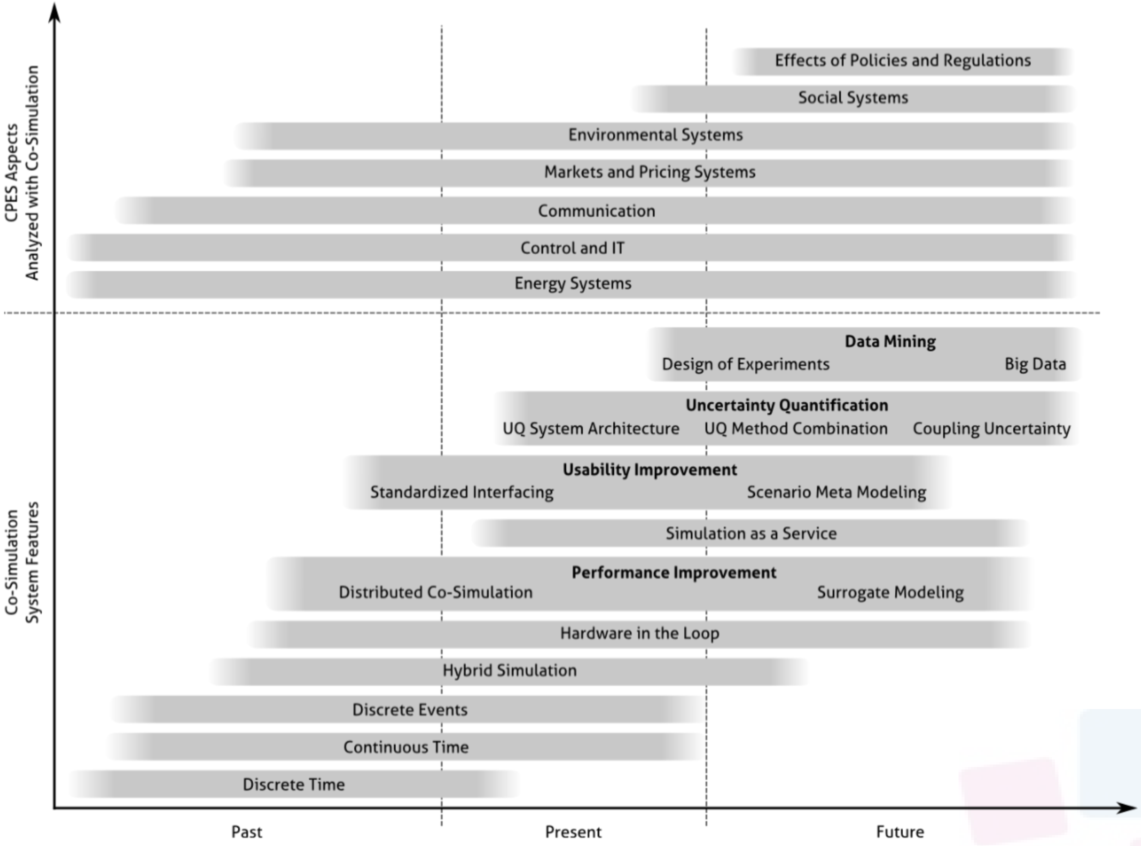
→ **Co-simulation**: orchestrate simulations running by different means

→ **Real-time simulations**: the real time expectation that the simulator needs to fulfill to interact with external components (hardware or software)

→ **Hardware in the loop (HiL)**: used to develop complex real-time embedded systems in which some components are real hardware, whereas others are simulated

* Steinbrink C, Schlögl F, BabazadehD, Lehnhoff S, Rohjans S, Narayan A. Future perspectives of co-simulation in the smart grid domain. In: 2018 IEEE International Energy Conference (ENERGYCON) IEEE; 2018.

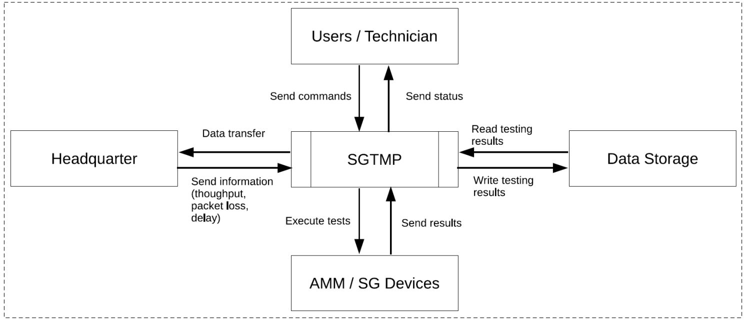
Co-simulations research needs

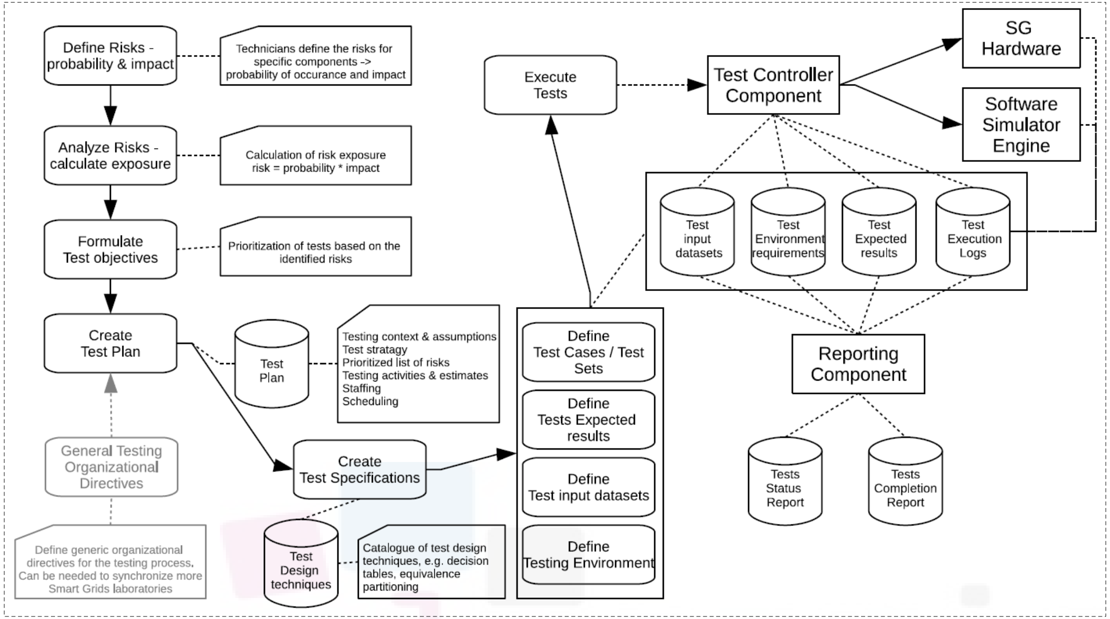


* Steinbrink C, Schlögl F, BabazadehD, Lehnhoff S, Rohjans S, Narayan A. Future perspectives of co-simulation in the smart grid domain. In: 2018 IEEE International Energy Conference (ENERGYCON) IEEE; 2018.

SGTMP Context

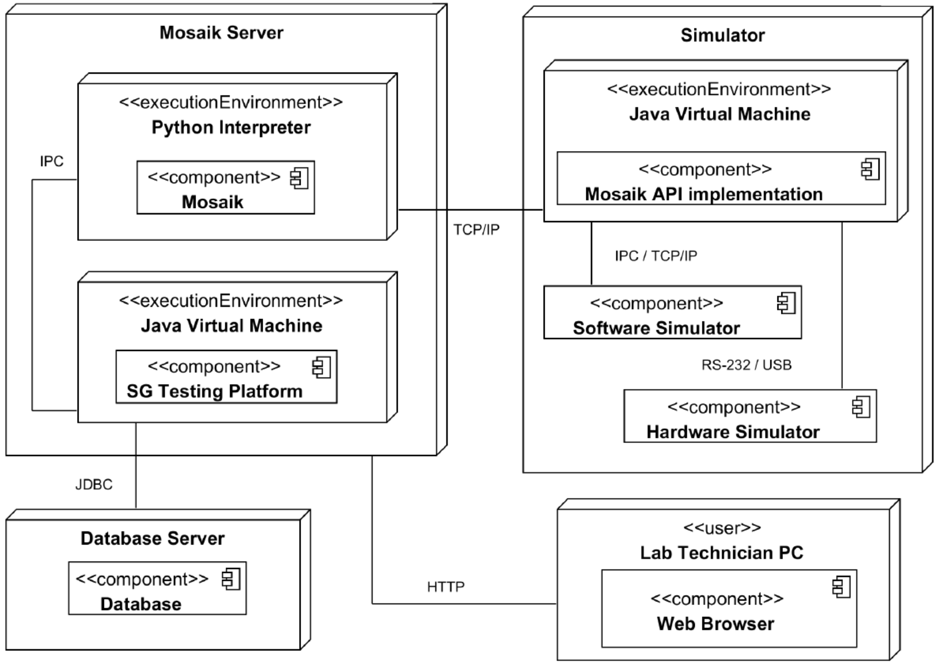
The Smart Grid Testing Management Platform (SGTMP) must allow the **execution of real-time hardware-in-the-loop SG tests** and **experiments** that can **simplify the testing process** in the context of **interconnected SG devices, supporting co-simulations**





SGTMP Architecture

→ Java + REST API + HTTP GUI server
→ Supporting Mosaik via Inter-Process Communication (IPC)
→ Hardware simulators via a hardware interface (e.g., RS-232, Ethernet, USB)



Key aspects

→ Allows testing of Smart Grid components

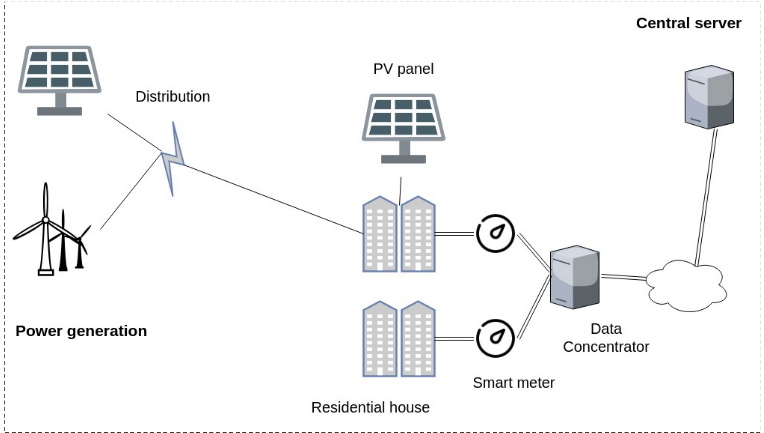
→ User defines Smart Grid topologies, simulator data flows, initial configuration

→ Java based API for simulators (can be HiL/RT/software) – Mosaik takes care of synchronization

→ Users have to implement interfaces to hardware simulators and translate requests between hardware and Mosaik

→ UI allows to manage the definition of testing scenarios to run and collect reports

Application Scenario (1/2)



Sample scenario for components stress testing

Topology

- Energy sources generating electricity (can be virtualized, e.g. with Arduino)
- Electric distribution lines: transmission losses can be simulated
- Several houses with Smart Meters
- Smart Meter Data Concentrators
- Main server collecting data from data concentrators for data analysis

Goals

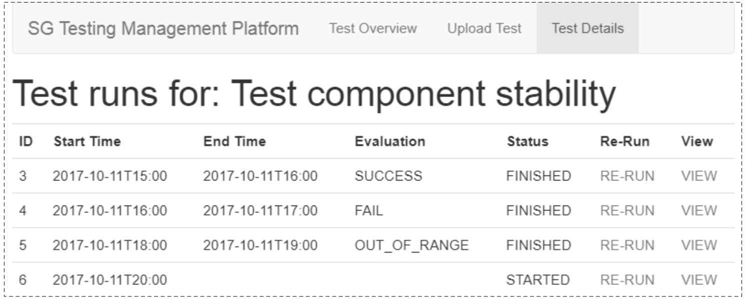
- SMDC need to be able to handle the data collection from multiple SMs at once without losing data or crashing due to data overloading
- SMDC must forward the received data to a central collection server in periodic intervals

Role of SGTMP

- Each SM in the network is instructed to send their observed data by an attached SGTMP node to their SMDC
- performance of the SMDC is observed and evaluated using criteria including: data loss, maximum responses processed per unit of time, accuracy and percentage of data sent from SMDC to the central collection server

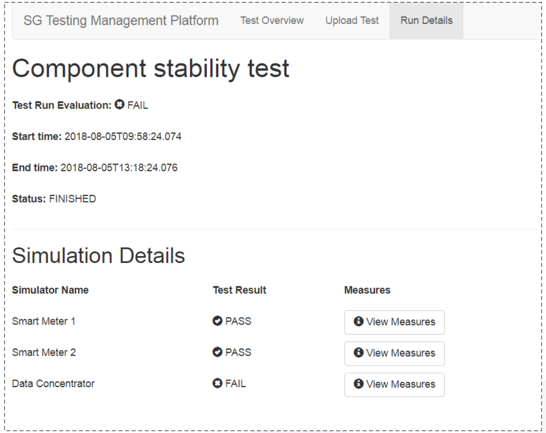
Application Scenario (2/2)

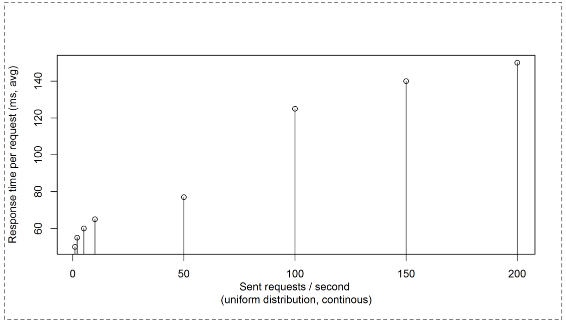
→ Creating and editing tests along with defining their test pass criteria and the SG topology
→ All test definitions can be viewed and modified
→ Individual tests can be started and their progress examined


SG Testing Management Platform Test Overview Upload Test Test Details

Test runs for: Test component stability

ID	Start Time	End Time	Evaluation	Status	Re-Run	View
3	2017-10-11T15:00	2017-10-11T16:00	SUCCESS	FINISHED	RE-RUN	VIEW
4	2017-10-11T16:00	2017-10-11T17:00	FAIL	FINISHED	RE-RUN	VIEW
5	2017-10-11T18:00	2017-10-11T19:00	OUT_OF_RANGE	FINISHED	RE-RUN	VIEW
6	2017-10-11T20:00			STARTED	RE-RUN	VIEW





Executed tests pass/fail → Detailed view with JSON logs → Simulation visual results

Future Works

- → Implementation of visual topology definition (at the moment is xml-based)
- → Implementation of a Domain Specific Language (DSL) to ease the configuration
- → Improvement of visual reporting and UI
- → Further ISO/IEC/IEEE 29119 standard support (e.g. risk-based testing)
- → Pilot studies
- → Supporting other simulation frameworks (?)
- → Release the platform as open source with proper documentation

SGTMP Team

Bruno Rossi

Martin Schvarcbacher

Lubomír Jahn

Katarína Hrabovská


IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS 1

Smart Grids Data Analysis: A Systematic Mapping Study

Abstract—**Data analytics and data science play a significant role in nowadays society. In the context of Smart Grids (SG), the collection of vast amounts of data has seen the emergence of a plethora of data analysis approaches. In this paper, we conduct a Systematic Mapping Study (SMS) aimed at getting insights about different facets of SG data analysis: application sub-domains (e.g., power load control), aspects covered (e.g., forecasting), used techniques (e.g., clustering), tool-support, research methods (e.g., experiments/simulations), replicability/reproducibility of research. The final goal is to provide a view of the current status of research. Overall, we found that each sub-domain has its peculiarities in terms of techniques, approaches and research methodologies applied. Simulations and experiments play a crucial role in many areas. The replicability of studies is limited concerning the provided implemented algorithms, and to a lower extent due to the usage of private datasets.**

Index Terms—**Smart Grids, Cyber-Physical Systems, Data Analytics, Literature Survey, Systematic Mapping Study.**

I. INTRODUCTION

THE Smart Grid (SG) is a two-way cyber-physical system utilizing information to provide safe, secure, reliable, resilient, efficient, and sustainable electricity to end-users [12], [20]. The SG plays nowadays a major role in the integration of the Smart Cities concept by putting into effect the Smart Energy conceptual element: smart electrical energy systems that interconnect utilities and end-users by means of a Smart Infrastructure [8], [19], [24]. The SG is a key enabler, enhancing the decision making process, providing self-healing and automation of the energy grid, and integration of renewable energy sources [24].

There are several definitions of a Smart Grid [13]. The European definition emphasizes the fact that SGs are electricity networks intelligently integrating the behaviour of all actors to reach sustainable, economic and secure energy supply [15]. The United States Department of Energy (USDoE) definition focuses more on the security and safety threats to be addressed with resilient and self-healing mechanisms, providing opportunities for new services and markets [30].

SGs pose several challenges that derive mainly from the integration of the physical infrastructure with information and communication technologies [10]. All these challenges need to be addressed with a holistic view taking into consideration all the different layers that form the SG ecosystem [10]. Some of the main challenges are the increase of importance for availability of communication network over traditional confidentiality and integrity aspects in traditional networks [4], the importance of customers' privacy and security of the infrastructure [27], the relevance of ways to integrate renewable energy sources in a reliable way [39], and the usage of information/data available for self-healing and self-monitoring purposes [2], [3], [21].

This article is focused on this last challenge, as SGs gave rise to large amount of opportunities in terms of data analytics initiatives: the large availability of data from the smart infrastructure allows many decision support initiatives, but also the implementation of predictive algorithms to improve the provided services [35]. Typical examples involve power load forecasting predicting the possible load curve that represents the electricity consumed by customers over time [M132], or Demand Response (DR) representing load balancing of energy supply and demand during peak hours [M86]

The goal of this paper is to provide an overview of data analysis in SGs with a focus on sub-areas, aspects of research, techniques, research methods, tool support, and replicability/reproducibility concerns. We aim at complementing previous survey research like the one from Alahakoon et al. [1], with a more systematic review by aggregating fine-grained knowledge from published articles in different sub-domains of SG data analysis. We have the following main contributions:

- a large Systematic Mapping Study (SMS) [5], [33] on Smart Grids data analytics, including 267 papers. To our knowledge, this is the largest review in the SG data analysis domain in terms of included articles;
- a categorization of different SG data analysis sub-domains, with cross-cutting aspects such as techniques used, research methodologies, aspects investigated, replicability concerns with the availability of source code and datasets. While an SMS cannot ensure that all research is covered [9], it provides a systematic sampling mechanism to look for research aspects holistically;

The article is structured as follows. In section II, we provide an overview of Smart Grids concepts relevant for this article: mainly architecture and components that compose the SG infrastructure. In section III, we define the goals, needs, process and research questions for the Systematic Mapping Study on data analysis in the SG context. In section IV, we present the results from the overall SMS process, divided into answers to the main research questions set. In Section V, we discuss the main threats to validity of the current study. Section VI provides the final conclusions.

II. SMART GRIDS

A SG consists of diverse hardware and software systems with a complex communication infrastructure. To fully understand the smart operations supported by the infrastructure, it is necessary to map the infrastructure to the provided services, usage scenarios, and stakeholders.

One of the first attempts to formalize the overall structure of a SG was done by the US National Institute of Standards and Technologies (NIST) which developed a conceptual model of

Smart Grids Data Analysis: A Systematic Mapping Study

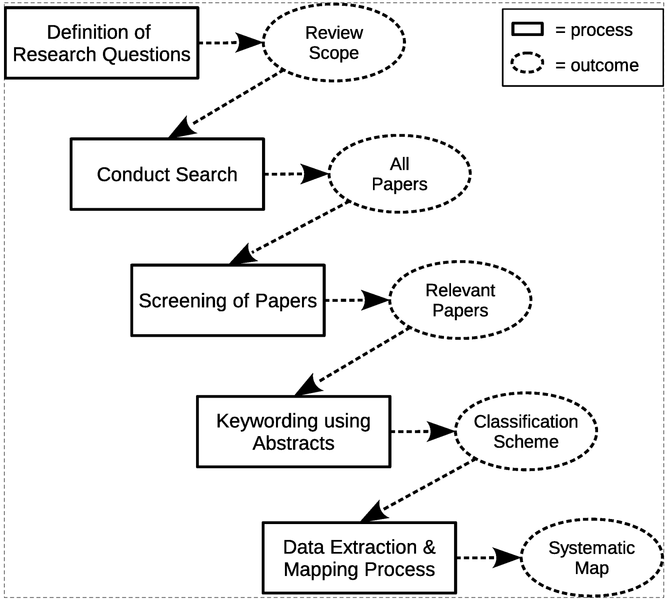
(Big Data Analysis)

Mapping Smart Grids Data Analysis Research

Smart Grids data analytics has gathered lots of attention in recent years

Our goal was to map existing research to understand areas, techniques, approaches used

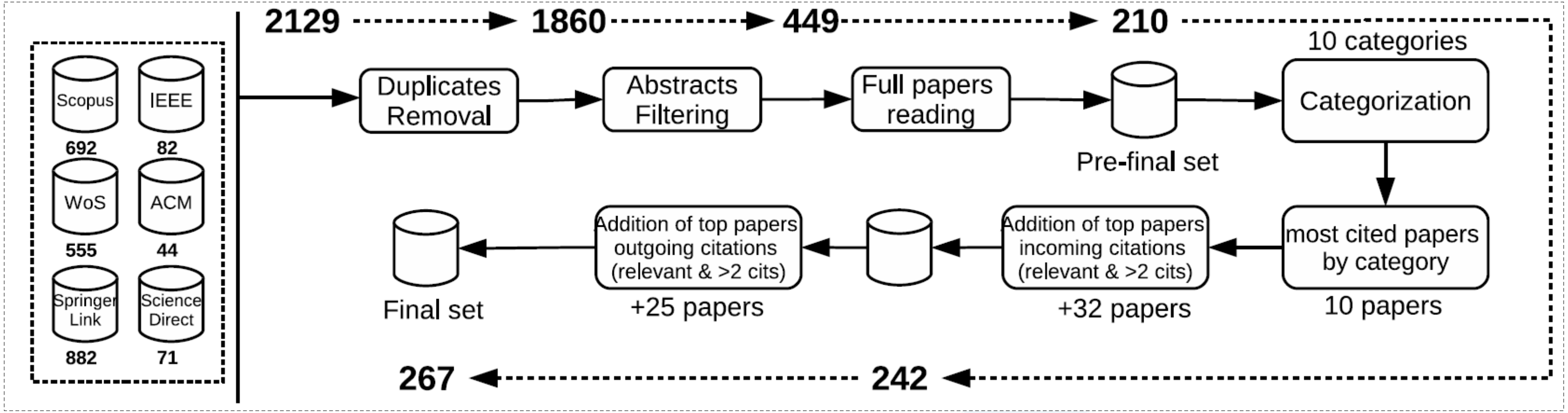
We performed a Systematic Mapping Study (SMS)



SMS: Process

Overall, we used **six digital repositories**

267 articles were included in the final review



SMS: Research Questions

Seven Main RQs:

RQ1. Which SG **application sub-domains** are **more popular** in terms of **research** and **their trends**?

RQ2. What are **common aspects** that are discussed in the identified sub-domains?

RQ3. What are **popular terms** that characterize each sub-domain?

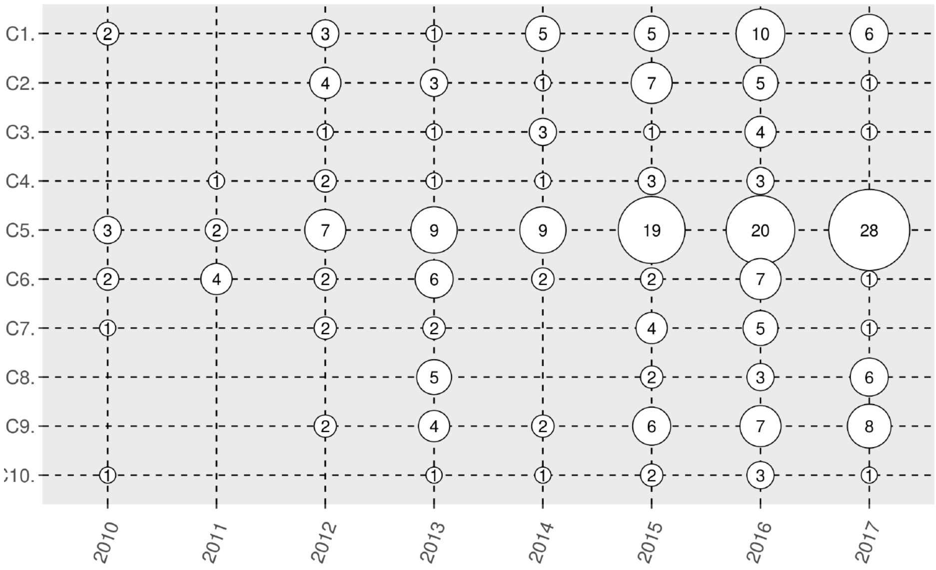
RQ4. Which are the reported **most used techniques** in the identified sub-domains?

RQ5. Which are the **most used software tools / development environments** used for data analysis in the identified sub-domains?

RQ6. What are the reported most used **quantitative research methods** used in the identified sub-domains?

RQ7. What is the status of **replicability / reproducibility** of the studies in terms of datasets used and **availability of implemented algorithms**?

SMS: Main Aspects (RQ1)



C1. Customer Profiling: Classification/clustering of users in common classes according to common characteristics (e.g. usage of appliances);

C2. Energy output forecasts: Prediction of energy output from renewable energy resources;

C3. Events analysis: Analysis of logs/events generated at different levels of the Smart Grids infrastructure (e.g. to detect anomalies);

C4. Load segregation: Disaggregating information about energy consumption on an appliance-by-appliance basis;

C5. Power loads/consumption analysis: Predicting the power consumption with the ultimate goal of reaching balance of supply and demand in the power market;

C6. Power quality: Power disturbance classification and algorithms for countermeasures and data compression;

C7. Pricing: Dynamics of forecasting electricity price and demand;

C8. Privacy: Data anonymization algorithms and other concerns related to disclosing private information about consumers;

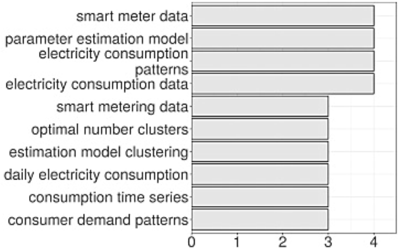
C9. Security: Algorithms dealing with countermeasures/prevention of attacks to the smart grids infrastructure;

C10. Smart Grid Failures: Aspects of SG failures, faults, and countermeasures;

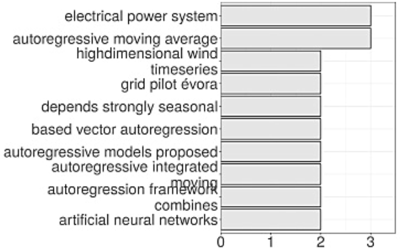
SMS: Main Aspects (RQ2)

Category	Aspects
C1. Customer Profiling	Load profile clustering ([M2], [M3], [M26], [M40], [M75], [M84], [M94], [M112], [M129], [M160], [M167], [M173], [M204], [M212], [M214], [M232], [M233], [M239], [M245]), power consumption pattern recognition ([M7], [M44], [M49], [M64], [M67], [M131], [M150], [M153], [M183], [M265], [M266]), power load forecasting ([M56], [M250]), events/tasks extraction ([M50])
C2. Energy output forecast	forecast renewable power sources ([M28], [M29], [M47], [M60], [M68], [M82], [M83], [M93], [M108], [M113], [M120], [M145], [M196], [M202], [M208], [M247], [M252], [M261], [M262]), power indicator forecasts ([M128])
C3. Events analysis	Data stream processing ([M48], [M58], [M96], [M180]), clustering events ([M138], [M139]), critical events analysis ([M87], [M248]), anomaly detection ([M154], [M179]), recommendation for energy utilization ([M114]), smart meters grouping ([M123])
C4. Load segregation	Non-intrusive appliance load monitoring [M35], [M66], [M92], [M130], [M147], [M207], [M219], disaggregate smart home sensor data [M135], [M144], [M260], [M267]
C5. Power loads / consumption analysis	consumption clustering ([M109], [M136], [M137], [M182], [M190], [M198], [M199], [M206], [M210], [M221], [M259], [M263]), consumption prediction ([M6], [M10]–[M12], [M19], [M23], [M24], [M32], [M36], [M42], [M45], [M53], [M57], [M65], [M69], [M70], [M78], [M79], [M81], [M86], [M88], [M90], [M98], [M101]–[M103], [M106], [M110], [M111], [M115]–[M117], [M122], [M124], [M125], [M132], [M146], [M152], [M156]–[M159], [M161], [M162], [M164]–[M166], [M171], [M172], [M174], [M175], [M178], [M187], [M193], [M194], [M203], [M211], [M213], [M218], [M220], [M226]–[M228], [M237], [M240], [M242], [M253], [M255], [M264]), consumption data analysis and modelling ([M14], [M20], [M25], [M30], [M43], [M51], [M80], [M118], [M201], [M256])
C6. Power quality	power quality disturbances classification ([M22], [M33], [M34], [M37], [M63], [M73], [M104], [M121], [M155], [M170], [M215]), power data compression ([M55], [M59], [M71], [M133], [M134], [M140], [M181], [M192], [M231], [M244], [M246]), meter placement for quality estimation ([M1], [M9]), energy losses detection ([M38]), missing data imputation [M177], [M205]
C7. Pricing	pricing forecasting ([M5], [M186], [M188], [M189], [M200], [M222]–[M225], [M229], [M243], [M249]), pricing impact on customer behaviour ([M27], [M241]), pricing for demand-side management ([M91], [M107])
C8. Privacy	privacy preserving data aggregation ([M4], [M21], [M95], [M97], [M105], [M141], [M217], [M234], [M238]), data re-identification ([M39], [M235]), appliance data obfuscation ([M72], [M85], [M89]), privacy in theft detection ([M216]), privacy preserving customer profiling ([M236])
C9. Security	Intrusion detection ([M8], [M15], [M52], [M74], [M76], [M77], [M143], [M148], [M209]),false data injection attacks ([M13], [M16], [M17], [M31], [M100], [M119], [M151], [M163], [M168], [M184], [M185], [M230], [M251], [M257]), energy theft ([M54], [M99], [M169], [M191], [M254], [M258]), distinguishing cyber-attacks from physical faults ([M18])
C10. SG failures	fault status detection [M41], [M46], [M61], [M62], [M126], [M127], [M142], [M176], fault type classification [M197], power distribution reliability [M149], [M195]

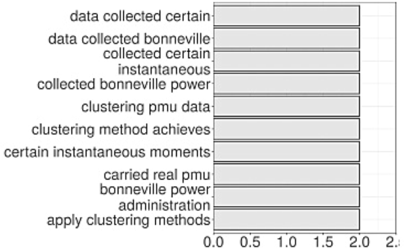
SMS: Popular Terms (RQ3)



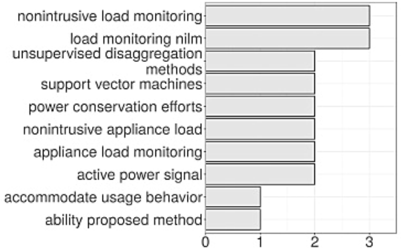
(C1) Customer Profiling



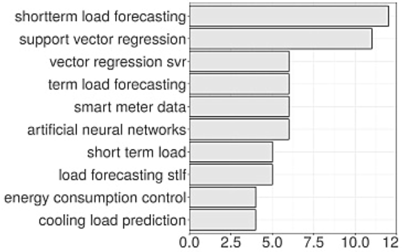
(C2) Energy output forecast



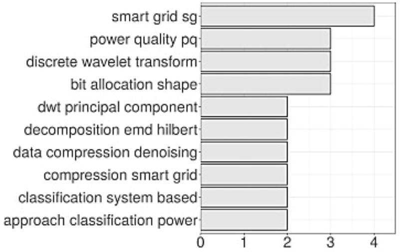
(C3) Events Analysis



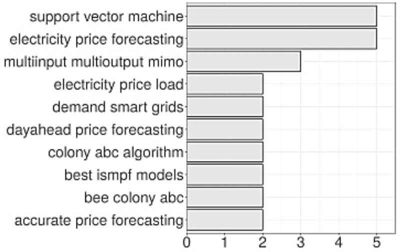
(C4) Load Segregation



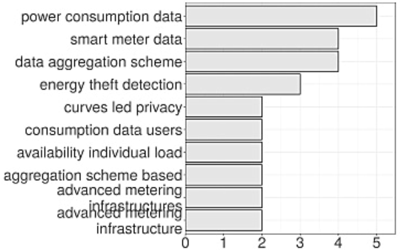
(C5) Power loads/consumption



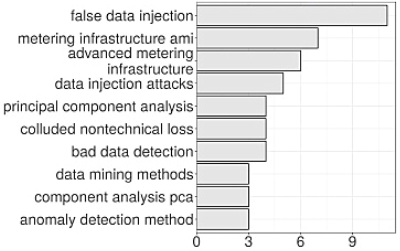
(C6) Power quality



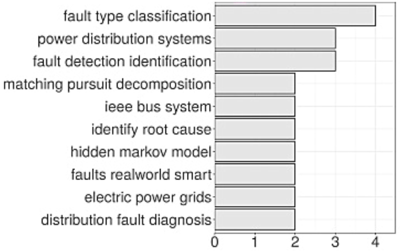
(C7) Pricing



(C8) Privacy



(C9) Security



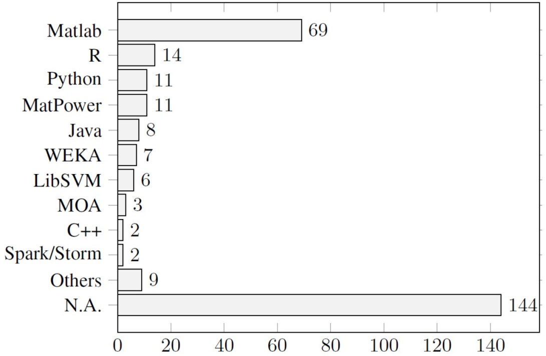
(C10) SG failures

SMS: Main Techniques (RQ4)

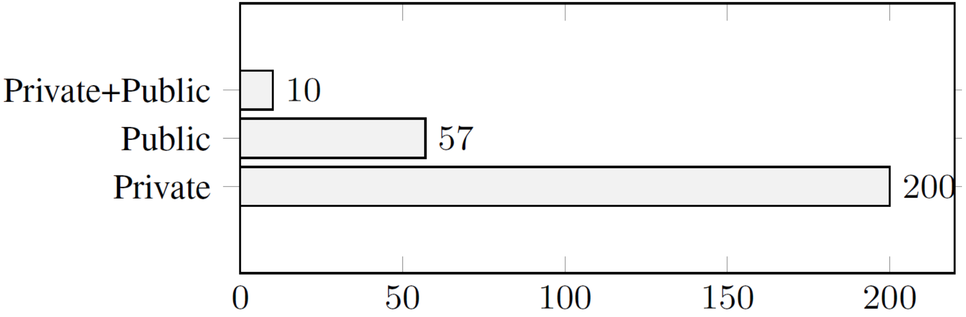
Category	Techniques
C1. Customer Profiling	**k-means clustering (16)** ([M2], [M3], [M40], [M44], [M49], [M64], [M67], [M84], [M94], [M129], [M153], [M160], [M204], [M212], [M214], [M245]), **fuzzy c-means clustering (7)** ([M49], [M64], [M173], [M204], [M245], [M265], [M266]), **Hierarchical Clustering (HAC) (7)** ([M44], [M56], [M64], [M94], [M204], [M212], [M232]), **Support Vector Machine (SVM) (6)** [M3], [M112], [M150], [M204], [M239], [M250], Self-Organising Map (SOM) (4) [M2], [M64], [M167], [M212], Multi Layer Perceptron (MLP) ANN (3) [M40], [M150], [M232], t-means clustering [M183], k-Nearest Neighbour (kNN) [M112], [M204], Random Forest [M7], [M64], [M204], Decision Trees [M3], [M150], Principal Component Analysis (PCA) [M44], [M173], multi-resolution clustering (MRC) [M160], Wavelet-based clustering [M56], ART Neural Network [M131], dynamic clustering with Haussdorff similarity distance [M26], Ensemble Methods [M3], Genetic Algorithms [M3], Discriminant Analysis [M3], Discrete Wavelet Transform [M129], Deep Learning CNNs [M233], statistical-based [M75], Takagi-Sugeno fuzzy models (FM) [M239], binary regression analysis [M214], Logistic Regression [M67], [M67], Dirichlet Process Mixture Model (DPMM) [M94], Apriori algorithm [M7]
C2. Energy output forecast	**autoregressive forecasting model (7)** ([M28], [M29], [M68], [M83], [M120], [M252], [M261]), **autoregressive integrated moving average (ARIMA) (3)** ([M82], [M196], [M202]), **Support Vector Regression (SVR) (3)** ([M60], [M108], [M145]), **k-Nearest Neighbour (k-NN) (3)** ([M82], [M108], [M202]), k-means clustering [M113], Linear Regression [M113], [M208], Logistic Regression [M68], Partial Least Squares Regression (PLSR) [M262], local linear model for time series [M247], time-series clustering with Quality Threshold (QT) algorithm [M208], Support Vector Machine (SVM) [M196], ANN:Self-Organising Map (SOM) [M145], ANN:Multilayer Perceptrons (MLP) [M82], [M93], Feedforward Neural Network (FFNN) [M202], Backpropagation Neural Network (BNN) [M128], Genetic Algorithms [M202], Random Forests [M108], Ensemble Methods [M113], Ordinary Least-Squares Fitting [M28], generalized autoregressive conditional Heteroscedasticity GARCH model [M47], Gaussian Conditional Random Fields (GCRF) [M261], Particle Swarm Optimization (PSO) [M128]
C3. Events analysis	**k-means clustering (5)** ([M114], [M138], [M139], [M154], [M179]), **hierarchical clustering algorithm (HAC) (3)** ([M87], [M138], [M139]), Decision Trees [M48], x-means clustering [M154], DBSCAN (Density Based Spatial Clustering of Applications with Noise) [M138], [M139], Mean-Shift Clustering (MSC) [M114], Support Vector Machines (SVM) [M96], [M139], k-Nearest Neighbour (k-NN) [M248], d-stream - time-series clustering algorithm [M180], time series clustering with Dynamic Time Warping (DTW) [M139], Hoeffding Adaptive Tree (HAT) [M58], ADaptive sliding WINdow (ADWIN) [M58], Piecewise Aggregate Approximation (PAA) [M139], Singular Value Decomposition (SVD) [M154], [M248], exponential smoothing forecasting method [M123], Independent Component Analysis [M248], Kernel Ridge Regression [M248], Parzen Density Estimator (PDE) [M154], Monte Carlo simulations [M123]
C4. Load segregation	**Hidden Markov Model (HMM) (3)** ([M35], [M135], [M260]), **Support Vector Machines (SVM) (2)** ([M147], [M267]), **k-Nearest Neighbour (k-NN) (2)** ([M144], [M147]), **ANN:Multilayer Perceptrons (MLP)** (2) ([M219], [M267]), Principal Component Analysis (PCA) [M219], Regression Models [M130], Bayes classifier [M219], Ensemble Methods [M147], Dynamic Time Warping (DTW) [M35], Karhunen Loeve (KL) expansion [M66], Ant colony optimization [M92], ZIP Model-phaselet [M207]
C5. Power loads / consumption	**Multi Layer Perceptron (MLP) ANN (23)** ([M14], [M25], [M42], [M45], [M78], [M81], [M101], [M110], [M122], [M124], [M132], [M156], [M164], [M171], [M178], [M187], [M210], [M211], [M213], [M228], [M237], [M242], [M256]), **Support Vector Machines (SVM) (21)** ([M36], [M53], [M57], [M65], [M78], [M79], [M81], [M106], [M115], [M117], [M122], [M157], [M159], [M166], [M187], [M193], [M203], [M227], [M240], [M253], [M256]), **autoregressive integrated moving average (ARIMA) (13)** ([M6], [M19], [M32], [M42], [M53], [M78], [M90], [M103], [M116], [M117], [M178], [M190], [M255]), **k-means clustering (11)** ([M12], [M14], [M109], [M136], [M137], [M157], [M162], [M190], [M193], [M201], [M210], [M228]), **Linear Regression Analysis (11)** ([M11], [M45], [M57], [M78], [M79], [M111], [M117], [M118], [M122], [M165], [M198], [M211]), **Genetic Algorithms (GA) (7)** ([M30], [M42], [M78], [M116], [M124], [M194], [M227]), **fuzzy c-means clustering (6)** ([M136], [M137], [M174], [M218], [M221], [M263]), **Fuzzy Logic (6)** ([M102], [M125], [M178], [M220], [M226], [M242]), g-means clustering [M182], DBSCAN clustering [M190], [M221], **Hierarchical Agglomerative Clustering (HAC) (6)** ([M136], [M137], [M199], [M210], [M218], [M259]), fuzzy subtractive clustering method [M206], **k-Nearest Neighbors (k-NNs) (6)** ([M69], [M78], [M90], [M221], [M228], [M264]), Self-Organizing-Maps(SOM) [M109], [M171], [M193], [M201], [M213], deep neural network (DNN) [M79], Radial Basis Function neural network model (RBF-PCA-WFCM) [M174], Wavelet Neural Networks [M178], [M264], Decision Trees [M70], [M90], [M157], [M253], [M264], Random Forests (RF) [M11], [M78], [M79], [M90], [M159], Wavelet Transform [M116], [M182], [M211], Bayesian Networks [M23], Power Factor Analysis [M198], Principal Component Analysis (PCA) [M24], [M32], [M174], [M187], Kalman Filter [M152], Simulated Annealing [M124], MultiVariate Gaussian Distribution Function (MVGDF) [M10], Montecarlo Simulations [M10], [M111], Pattern Sequence-based Forecasting (PSF) [M12], FA (Factor Analysis) [M187], linear discriminate analysis (LDA) [M161], Markov Models [M14], [M98], Particle Swarm Optimization (PSO) [M125], [M164], [M166], Apriori algorithm [M157], [M166], evolutionary local kernel regression [M146], GARCH [M115], Difference Auto-Regressive (DAR) [M172], Online Sequential Extreme Learning Machine (OS-ELM) [M162], Kernel Ridge Regression (KRR) [M90], [M264], Lyapunov optimization technique [M158], Ensemble Models [M78]
C6. Power quality	**wavelet transform (7)** ([M55], [M73], [M133], [M134], [M192], [M244], [M246]), **S-Transform algorithm (4)** ([M33], [M34], [M104], [M121]), **Principal Component Analysis (PCA) (4)** [M59], [M181], [M244], [M246], k-means clustering [M244], Bayesian classifier [M22], Bayesian Network [M9], rule-based classification [M215], ANN: Multi-Layered Perceptrons (MLPs) [M37], [M155], Support Vector Machines (SVM) [M73], [M244], probabilistic neural network (PNN) [M121], [M155], k-Nearest Neighbour (k-NN) [M155], [M177], decision trees [M37], [M104], Self-Organizing Maps (SOM) [M55], fuzzy decision tree (FDT)-based classifier [M34], balanced neural tree [M33] Fuzzy-ARTMAP neural network [M63], Fourier Transform [M38], [M104], Hilbert transform (HT) [M33], piecewise compression technique [M71], nonlinear autoregressive model with exogenous inputs [M170], Kalman Filter [M170], SZIP algorithm [M231], Singular Value Decomposition (SVD) [M244], [M246], Piecewise Aggregate Approximation (PAA) [M244], Slack-Referenced Encoding (SRE) [M140], Linear Interpolation Imputation [M205], weighted least square method [M1]
C7. Pricing	**Support Vector Machine (SVM) (6)** ([M91], [M222]–[M224], [M243], [M249]), **Wavelet Transform (5)** ([M91], [M222]–[M224], [M229]), **ARIMA (3)** ([M91], [M222], [M229]), **Artificial Bee Colony (ABC) (3)** ([M91], [M223], [M224]), Fuzzy Inference Net (FIN) [M188], Fuzzy Self Organising Maps (SOM) [M188], fuzzy c-means clustering [M241], ANN:Extreme Learning Machine (ELM) [M225], Principle Component Analysis (PCA) [M243], [M249], ANN: Multi-Layered Perceptrons (MLPs) [M200], Grey Correlation Analysis (GCA) [M243], Relevance Vector Machines (RVMs) [M5], Linear Regression [M5], Autoregressive Moving Average [M5], Ensemble Models [M5], reference models for price estimation (RMPE) [M186], generalized autoregressive conditional Heteroscedasticity (GARCH) [M229], data association mining (DAM) algorithms [M189], Apriori algorithm [M189], Gravitational Search Algorithm (GSA) [M222], Markov Decision Process [M27], Reinforcement Learning [M27], Montecarlo simulation [M107]
C8. Privacy	Fuzzy c-means clustering [M85], Random Gaussian Noise [M105], Colored Noise [M217], Symmetric Geometric Noise [M21], Secret Sharing Scheme [M95],Elliptic Curve Based Data Aggregation (ECBDA) [M238], Collaborative Anonymity Set Formation (CASF) [M4], Wavelet-based Multi-resolution Analysis (MRA) [M141], adversarial strategy algorithm [M234], differentially private aggregated sums [M97], Integer Linear Optimization (ILP) [M39], Tolerable Deviation algorithm [M89], partitioning algorithm [M235], Haar Wavelet transform [M72], Profile Matching protocol using Hamming distance [M236], Kalman filter [M216]
C9. Security	**Principal Component Analysis (PCA) (8)** ([M16]–[M18], [M74], [M100], [M148], [M185], [M257]), **Support Vector Machines (SVM) (5)** ([M18], [M74], [M151], [M185], [M258]), k-means clustering [M13], [M251], fuzzy c-means clustering [M185], DBSCAN clustering [M148], [M185], Decision Trees [M52], [M54], Feedforward Neural Network (FFNN) [M143], Multi-layer Perceptron (MLP) [M185], k-Nearest Neighbor (kNN) [M185], clustering-based anomaly detection [M209], Ensemble Model [M143], Exponential Smoothing [M15], Hoeffding Tree [M76], [M77], Markov Chains [M8] , generalized likelihood ratio test (GLRT) [M230], particle swarm optimization (PSO) [M251], random FDA (random false data attack detection) [M163], Weighted Residual Error Method [M119], Chi-Square Test [M17], [M119], Kullback-Leibler divergence [M31], generalized linear model (GLM) [M31], Cascade Potential Ranking [M184], Two Stage Branching Algorithm [M184], Colored Petri Nets [M168], Recursive Least Squares [M99], Linear Regression [M254], Technical Loss Model [M191], autoregressive model [M169]
C10. SG failures	**k-means clustering (2)** ([M62], [M127]), Basic Sequential Algorithm Scheme (BSAS) [M62], Genetic Algorithms (GA) [M62], Support Vector Machines (SVM) [M61], General Regression Neural Networks [M46], Hidden Markov Model [M127], Ordered Weighted Averaging (OWA) [M142], Radial basis functions (RBF) ANN [M142], Logistic Regression [M41], k-Nearest Neighbour (k-NN) [M127], Wavelet Transform (WT) [M142], Principal Component Analysis (PCA) [M61], dynamic optimal synchrophasor measurement devices selection algorithm (OSMDSA) [M126], Wavelet Packet Decomposition [M46], Least Square Phasor Estimation [M197], Reliability Indexes (SAIDI, SAIFI,...) Thresholds [M149], Multivariate analysis of variance (MANOVA) [M195]

SMS: Main tools & datasets (RQ5 & RQ7)

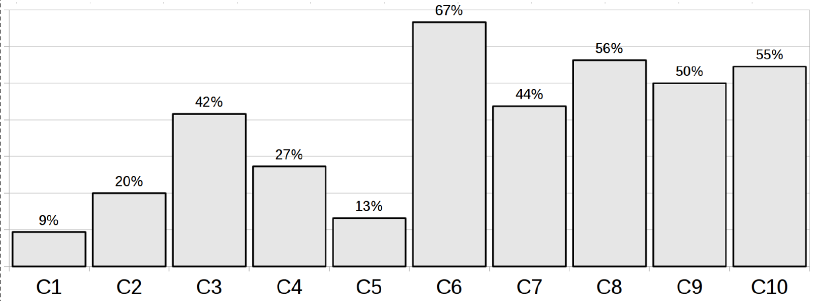
Tools used



Datasets used



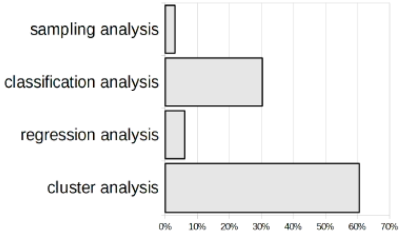
Simulations usage:



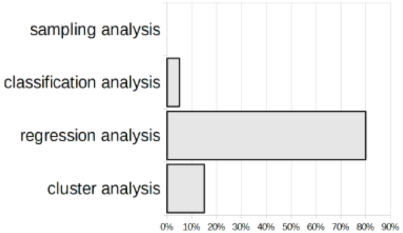
Public software (only 4/267)

Article	Public Repository
Bianchi et al. (2015) [M32]	https://bitbucket.org/ispamm/distributed-esn
Natividad et al. (2017) [M190]	https://github.com/powertac/powertac-tools
Hoeverstad et al. (2015) [M116]	https://github.com/axeltidemann/load_forecasting
Bonfigli et al. (2015) [M35]	https://nilmtk.github.io

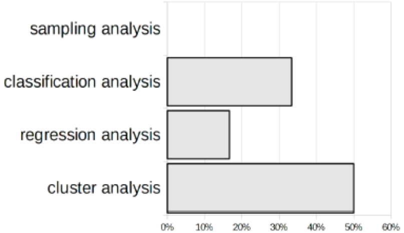
SMS: Research Methods(RQ6a)



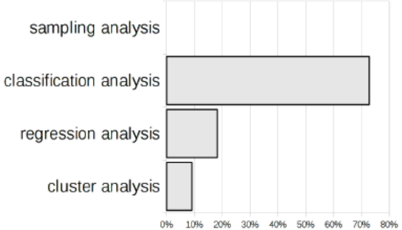
(C1) Customer Profiling



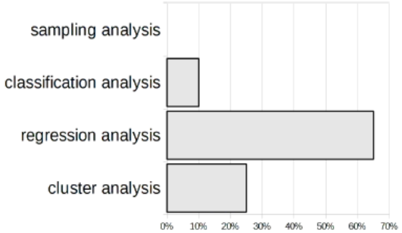
(C2) Energy output forecast



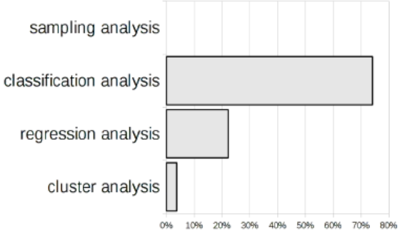
(C3) Events Analysis



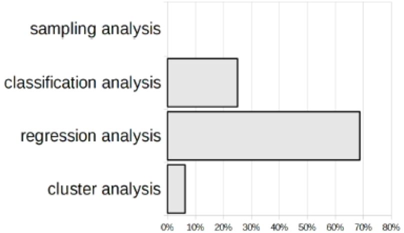
(C4) Load Segregation



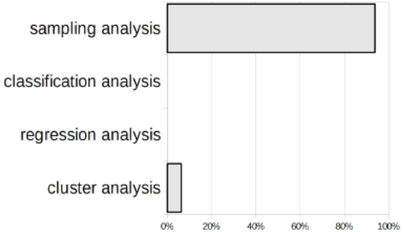
(C5) Power load/consumption



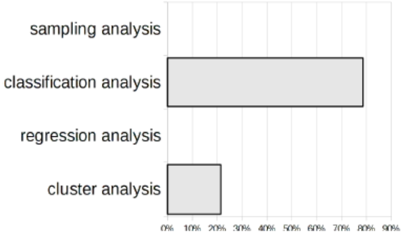
(C6) Power quality



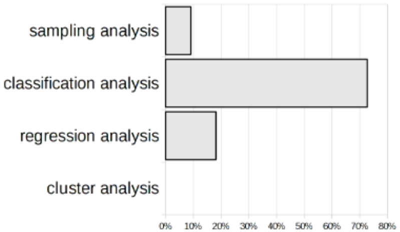
(C7) Pricing



(C8) Privacy

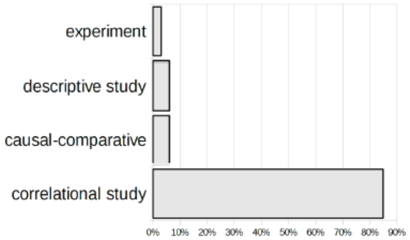


(C9) Security

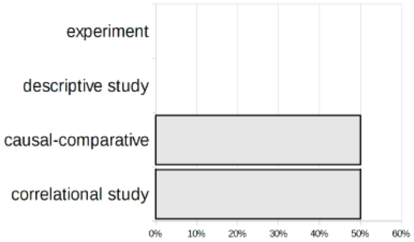


(C10) SG failures

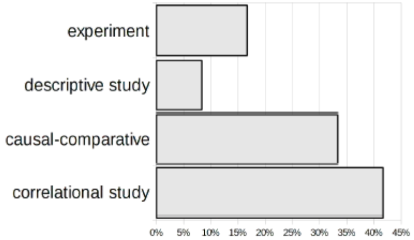
SMS: Research Methods(RQ6b)



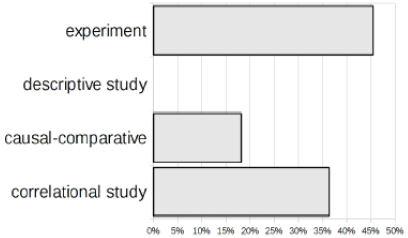
(C1) Customer Profiling



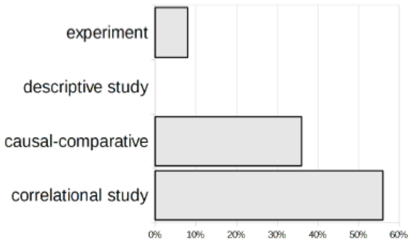
(C2) Energy output forecast



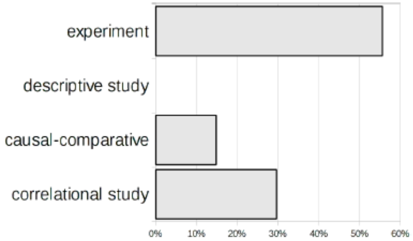
(C3) Events Analysis



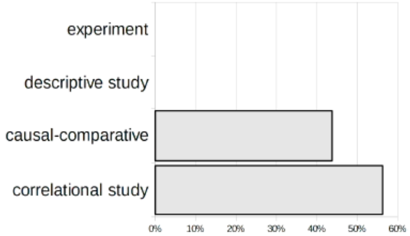
(C4) Load Segregation



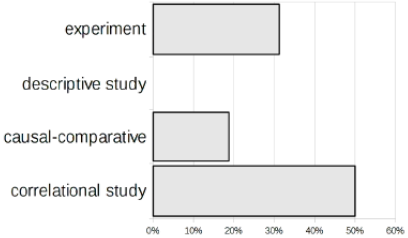
(C5) Power load/consumption



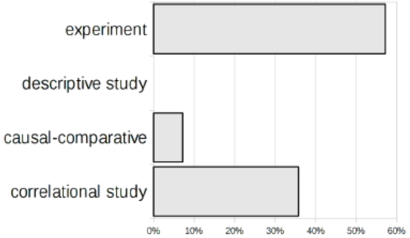
(C6) Power quality



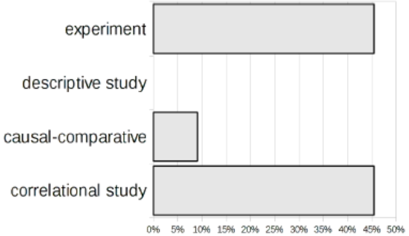
(C7) Pricing



(C8) Privacy



(C9) Security



(C10) SG failures

Outcomes from the SMS

→ Clear picture about the situation of different research areas (however, an SMS can never be fully complete)

→ Understanding of usage of tools and datasets available

→ Publishing posting the pre-print allowed to get contact for collaboration

→ Future work: big data platform for Smart Grids data analysis

Main published material

→ M. Schvarcbacher, K. Hrabovská, B. Rossi, T. Pitner (2018). *"SGTMP: Smart Grid Testing Management Platform"* (submitted to journal, not yet available)

→ B. Rossi, S. Chren (2018). *"Smart Grids Data Analysis: A Systematic Mapping Study"* (submitted to journal, pre-print: https://arxiv.org/abs/1808.00156)

→ Chren, V., Rossi, B., Bühnová, B., Pitner, T. (2018). *"Reliability Data for Smart Grids: Where the Real Data Can be Found"*, in the 4th IEEE Smart Cities Symposium Prague (SCSP) 2018, IEEE. [download]

→ Schvarcbacher, M., Rossi, B. (2017). *"Smart Grids Co-Simulations with Low-Cost Hardware"*, in 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA) 2017, IEEE, DOI: 10.1109/SEAA.2017.43. [download]

→ Rossi, B., Chren, S., Bühnová, B., Pitner, T. (2016). *"Anomaly Detection in Smart Grid Data: An Experience Report"*, in IEEE International Conference on Systems, Man, and Cybernetics (SMC2016), IEEE. [download]

→ Chren, V., Rossi, B., Pitner, T. (2016). *"Smart Grids Deployments within EU Projects: The Role of Smart Meters"*, in the 2nd IEEE Smart Cities Symposium Prague (SCSP) 2016, IEEE. ISBN:978-1-5090-1116-2, DOI: 10.1109/SCSP.2016.7501033. [download]

Theses:

→Schvarcbacher M. (2018). *"Smart Grid Testing Management Platform"*, BSc Thesis Masaryk University, Brno [download]

→ Hrabovská K. (2017). *"Supporting a Smart Grids Laboratory: Testing Management for Cyber-Physical Systems"*, MSc Thesis Masaryk University, Brno [download]