



MODULE 1 CHAPTER 5. ANALYSIS OF DATA CONSISTENCY



Horizon 2020 European Union Funding for Research & Innovation This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847097



Contents

ABB	REVIATIONS
SHO	PRT SUMMARY
1	INTRODUCTION
2	FIRST STUDY OF THE SIMULATION ENVIRONMENT7
3	DATA ANALYSIS AND ERROR CORRECTION
3.1.	Definition of the format in which data is received8
3.2.	Establishment of an operation schedule8
3.3.	Adaptation and standardization of the frequency of monitoring data9
3.4.	Quality preliminary analysis of the data9
3.5.	Data consistency analysis10
3.6.	Error correction11
4	ADAPTATION TO APPROPRIATE FORMATS12
5	ESTIMATION OF END USERS DEMAND
4.1.	LoadProfileGenerator Tool (LPG)12
4.2. 4.2.1	City Energy Analyst demand profile generator
6	REFERENCES16





List of Figures

Figure 1: SO WHAT data monitoring methodology main steps	6
Figure 2: phases for general data analysis and error correction process [1]	6
Figure 3: tools to generate demand profiles	7
Figure 4: a) Typical load profile for residential load. b) Typical load profile for Industrial-one sh	nift
workload. c) Typical load profile for Industrial—two shift workload. d) Typical load profile t	for
commercial load [2]	8
Figure 5: variable with low variability (recorded with lower frequency)	9
Figure 6: variable with high variability (recorded with higher frequency)	9
Figure 7: use of color conditional formatting to highlight information	10
Figure 8: example of heat exchanger with waste heat for steam formation [3]	.11
Figure 9: example of human behaviour simulator that can be modelled by LPG [5]	.13
Figure 10: load profile generated in LPG [5]	.13
Figure 11: buildings studied (left), surroundings (middle) and streets (right). Example in C	ΕA
Dashboard	14
Figure 12: load curve for district (monthly) generated in CEA	.15
Figure 13: load curve for one of the buildings (Hourly) generated in CEA	.15





Abbreviations

CEA: City Energy Analyst DH&CN: District Heating & Cooling Network LPG: Load Profile Generator RES: Renewable Energy Sources SCADA: Supervisory Control and Data Acquisition WH/C: Waste Heat/Cold





Short summary

In this module, it will be given a detail explanation of the methodology followed to analyse the consistency and availability of monitoring data in order to simulate possible waste heat recovery scenarios. The main objective of this module as part of SOWHAT project is the identification of strategies and procedures to collect data for mapping and quantifying the potential for waste heat and cold recovery and valorisation and for integration of renewables in industrial contexts.

The main inputs at this document come from the study conducted within SOWHAT project in WP1, WP2 and WP5 and have a relationship with the next activities in the project:

- **Task 1.1/ D.1.1** Report on Industrial Site Demo Assessment. This deliverable is associated to the assessment on waste heat potential and evaluate which are the main algorithms to model industrial energy processes and those ones related to the evaluation of potential local waste heat and cold. This quantification of the WH/C streams at the industrial demo sites are very useful for the demo sites interpretation and to evaluate the data received from the 3 industries.
- **Task 1.2.** Overcoming barriers in data collection and data format required, whose aim is to identify strategies and procedures to overcome barriers and criticalities in the collection of relevant data for mapping and qualifying the potential for WH/C recovery and RES integration.
- **Relation with the D1.4 and D1.5,** where importance of Energy Audit is explained. The deliverable **D1.4** focuses on requirements for data for SO WHAT tool, which globally coincide with data that are required for an energy audit of an industrial plant. It is therefore useful to provide an overview, in this chapter of the energy audit process and of the required input data and **D1.5** introduce this kind of document as a main in the information of every Industrial site.
- **Task 5.1/D5.2** Report on the instrumentation installation and commissioning. It is the deliverable that present the accuracy of the tool and its impact on the demo sites. Deliverable of report format and confidential only for members of the consortium.
- This document is related as well to the **work package 2** (Specification of the tool and common IT framework), where a definition of technical requirements for the SO WHAT Tool is studied.

This module will address the topics related to general aspects of the above-mentioned pillars and it will provide technical inputs to the trainee and a solid basis for the prosecution of the training activities. The handbook structure is here reported:

- 1) Chapter 1: Description of the waste H/C: how it is generated and where it can be seen most in the industrial plants.
- 2) Chapter 2: Analysis of WH/C technologies.
- 3) Chapter 3: Insights into the data needed to be collected to run the simulations and the related formats.
- 4) Chapter 4: Inputs related to the mapping, the tool's needs to map local RES and municipality or an industrial plant feature.
- 5) Chapter 5: High-level information and details linked to the simulation environment, including possible checks and evaluation of the consistency of data.
- 6) Chapter 6: The cause-effect relation between input data and outputs will be introduced in terms of qualitative and quantitative information.





1 Introduction

The following training module presents high-level information and details linked to the simulation environment of industrial waste heat and cold recovery, including possible checks and evaluation of the consistency of data. In this module the first half of the document will be structured in three sections (Figure 1). After that, a guide based on modelling different demand profiles will be presented.

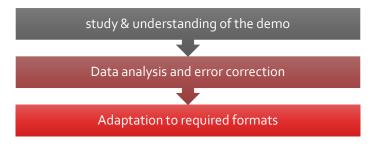


Figure 1: SO WHAT data monitoring methodology main steps

The *first step: study and understanding of the industry* is the industry study (from whom the monitoring data has been obtained) and understanding exercise is going to be required. For this training, this section will stay in the background, giving more importance to the second step.

The *next step: data analysis and error correction* the main part of all the chapter and it can be divided into more phases.

A data analyst's job is to understand the objective and its goals in enough depth that they can frame the problem the right way. After defining the aim of what it want to be solved with that data, it will be continued with the different steps of adaptation of the monitoring data and study of its consistency.



Figure 2: phases for general data analysis and error correction process [1]

This second phase (see Figure 1) is crucial for making sure that the data used is high-quality for its application. Finally, a correct representation of the data can show a global view to end the analysis process. Fixing a strategy for the complete process can help in order to replicate the procedure followed for different types of data. Extrapolating this generic method to this module simulation environment, data will be the adecuate to analyse heat potential sources in a particular industry and indeed, to be able to generate a real quality thermal production profile based on that waste heat monitoring data.

And the *last step: adaptation to required formats* depending on the objective behind the monitoring data use or where that dataset is going (SCADA systems, data analysis and visualization tools...), adaptation of the format takes place, in order to facilitate its application.





In this part of the process, type of consumers (residential, services, industrial...) who will benefit from this waste heat should be known so their temporal demand profiles (hourly, daily...) can be predicted, in the event that real profile demands were not given. To complete this guide for future applications in the same environment, examples of *demand forecasting tools* are recommended in this chapter.





LoadProfileGenerator

Among current demand forecasting tools, these

Figure 3: demand forecasting tools

These two tools have been selected because they encompass ways of defining demand from two perspectives: *LoadProfileGenerator (LPG)* is based on the demand of people behaviour of the household and *CityEnergyAnalyst (CEA)* focuses more on a detailed building model.

2 First study of the simulation environment

In the field of industry, any type of process with heat excess can be a potential source of heat recovery. Although it may seems obvious, it is wise to note that before proceeding to analyze the data and correct errors, a demo study and understanding exercise is going to be required. It will be difficult to perform a good fault detection on the received monitoring data if the physical system associated with the received monitoring data is not fully understood.

This actions are developed in complex industrial plants with much distinction between its applications. Therefore, there is no easy replicability in the location of possible heat/ cold recovery points within the process. As with the whole process in general, this study and understanding process must be customized for each specific case depending on multiple factors such as:

- the quality and quantity of the explanatory documentation provided together with the monitoring data;
- the level of complexity of the industrial settles;
- the standardization of production process of the demo;
- the professional background of the people in charge of the study.

For this reason, a few methods and sources of information are proposed for this study that will help tu understand the process and to proceed with the data analysis such as: *energy audits* where its energy situation is directly connected to the identification of potential heat/cold recovery sources, *flow diagrams* which is an effective procedure that can be helful to understand an industrial process with its visual representation and *Sankey diagram elaboration* for representing an entire input and output energy flow that allows the location of potential sources of residual heat precisely and easily.

3 Data analysis and error correction

A Data Analyst collects, treats and analyzes large data sets. Their role is to discover how data can be used to answer questions and solve problems. They are expected to collect and assemble data and process it to discover trends and insights useful to the business.





The data analysis and error correction step is so critical. It must be taken into account that there is not a valid format for all the industries (different hardware, software, language, methodology). The procedure followed can be structure in the following sections.

3.1. Definition of the format in which data is received

Data is originally generated in some type of Supervisory Control and Data Acquisition (SCADA) characteristic of each industry or software programs in the same line. The process usually follows the same steps: data is collected in some Data Base (DB) and then it is exported from there in a manual or automatic way. In some cases the exportation is done directly from this SCADA and this can cause problems such as: the ease of understanding the data written, the name and number of variables and the frequency in which they are recorded, this can lead to a very large file size for some software (special attention to the selection of the program used for the analysis), computer file format (potential data in DB format).

3.2. Establishment of an operation schedule

Before entering into the monitoring data interpretation is so important to establish the particular industry operating schedule in order to easily locate not only annual suspensions for maintenance and upkeep stops, but also work regime (shifts) and stops that occurred during the period of time associated with the data like, Holidays or vacations.

It is also crucial to establish the type of industry, from where data is monitorized, to analyse the data received. For example, it is not the same an industry with two shift work days, than one with one shift work or the ones with 24/7 schedule. It will be easier to interpret the changes in the load with this previous information.

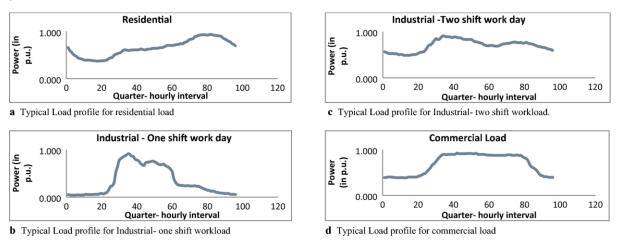


Figure 4: a) Typical load profile for residential load. b) Typical load profile for Industrial-one shift workload. c) Typical load profile for Industrial—two shift workload. d) Typical load profile for commercial load [2]

In the analysis a qualitative definition of seasonality must be included and obviously the data must respond accordingly to it. For example the demand for hot water for space heating should be much higher in the winter months or the cooling demand of a refrigeration process should be much higher during the summer.

This gives weight to the previous documentation section (preliminary knowledge of the operation calendar) where useful information should be gather before. However, if this information is not available, monitoring data will have to be interpreted because they will reflect these operation schedules. In other words, the study of the monitoring data will allow stablishing the operation





calendar, but in some cases it will be difficult to distinguish between unexpected stop situations and lack of data situations.

3.3. Adaptation and standardization of the frequency of monitoring data

Depending on the sensor, technology or equipment you required to monitor, the frequency of the recorded data would be higher or lower and even more, not all variables will need to be recorded with the same sampling frequency: variables with high variability are usually recorded more frequently compared to slower variables.



Figure 5: variable with low variability (recorded with lower frequency)





For example, the input of fuel (into the boilers for steam production), such as municipal waste, are very large loads that do not require very high data collection frequencies (Figure 5), that is why the data has a total daily record, while the value of steam that is introduced into a turbine to produce electricity in a power generation plant will require a higher registration (Figure 6).

An adaptation of the data to a homogeneous frequency will favour its analysis, otherwise the sections of quality preliminary analysis and data consistency will be more difficult to develop.

3.4. Quality preliminary analysis of the data

Before checking if the data are consistent and valid to be used in the tool, it is of vital importance to carry out a preliminary quality study for a first check of the monitored data.

It is important to develop a first variability check of each of the variables, recording the maximums and minimums to check for significant imbalances. This would show if the data have different orders





of magnitude (that do not coincide with the expected response, taking into account the seasonal time and operational schedule) and negative values where it does not make sense.

Talking about an analysis of numerical values, another tool that is very useful for a quickly analysing of variability of the data (apart from maximums and minimums) is the application of color scales to long list of different variables values. Color conditional formatting allows to highlight relation between variables, sudden changes in the list, values that are not in appropriate range, etc.



Figure 7: use of color conditional formatting to highlight information [3]

Another common problem that should be controlled in the data received could be the corresponding to systems that use cumulative type counters, this data will have to be transformed into average values corresponding to the period between two consecutive values of the original counter type variable. These adaptation processes should be included in this preliminary analysis. This cumulative type counters could encounter problems such as:

- Periods without reading with constant accumulated value, time ranges in which the sensor were not recording so that monitoring data does not exist in the data base until the next measured point;
- Periods without reading with variation of the accumulated value, in this case the sensor was reading correctly the process but the system has not registered these intermediate values;
- Zero crossing of the counters. Upon reaching the maximum value in the accumulation register, the counter is reset and begins to increase from the zero value.

There are several examples that could happen in any industrial process, for which this step (quality preliminary analysis), before data consistency analysis is necessary.

3.5. Data consistency analysis

Once the data analyser has reached this section, data will be found that, taking into account all the previous considerations, are technically correct. Due to errors in the computerized recording of the monitoring data (or human factor), data can be presented as formally correct registers but unfortunately without consistency with each other (e.g. corresponding wrong column of the data, duplication of data etc). This inconsistency can appear permanently, in which case it will be easy to detect. This situation could be a reflection of faulty sensor or a bad implemented software registration. On the other hand, this can appear in certain time periods of data and then, it will be more difficult to detect. This issue is usually due to an error in the data dump.





Another issues to take into account in this data review process is the homogenization of the measurement units of the monitored data. Carefully review that the recorded values correspond to the correct units of measure, in other words, the detection of variables that could have been measured in different units because it has been used different sensors. All these situations should be considered, knowing that the monitoring data comes from real industrial sites and day to day problems could appear. Thanks to the detailed study of the demo, the data could be analyse with an industrial point of view, and interpret the data related to the industrial processes:

• Study of the consistency in the data between actions in the process that occur simultaneously or actions that are dependent on others (e.g. if a boiler burns more waste, the production of steam will be greater and consequently it will also be the electrical energy produced in the turbine).

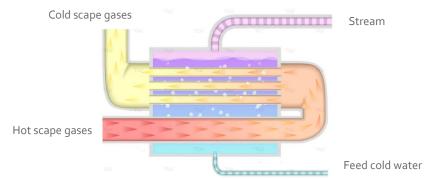


Figure 8: example of heat exchanger with waste heat for steam formation [4]

- Consistency in the inlets and outlets flows;
- Understand and pay attention to the trends of related variables (e.g. energy produced and consumed);
- Special attention to the seasonality (summer vs. winter) this is important to understand if the process and the monitoring data are valid for the aim of the project.

3.6. Error correction

Knowing the possible errors that can occur when analysing the monitored data of a system, the errors will be corrected in order to continue with the use of the tool or the procedure followed to solve the problems presented.

- For those errors that have been due to the manual process of exporting the database, the first step will be to re-request a dump of that data, because it could be an error when taking or writing that data in the sent file. Another option would be to ask for another period of time if it is available. In the end, the objective is usually to work with a year of monitoring data and it is not necessary for it to be from January to December of that same year. The important key is to have all the data for every month or the frequency required;
- In the hypothetical situation of missing data of a period of time, the approximated correction could be solved filling in empty data with data that is available. For example, if a week is missing (for whatever reason) it can be copied the data from another week (as close as possible) to occupy that blank space;
- When two variables are related, one with correct data and the other with erroneous data during a section, the missing information can be obtained with the equation of relation between both variables. The trend of the correct data of one variable against another (not





including those that are not correct) will be obtained and an equation will be obtained that can be used in the section for which adequate information is not available;

• **Construction of simulated data from other variables** to fill in gaps. Sometimes, as mentioned before, there are not enough sensors to calculate all the variables in order to calculate the potential sources of heat recovery, but there will be other variables that, by means of mathematical methods, consistent assumptions or thermodynamic and physical equations, will be able to find these lost variables.

For example, the heat generated is not available, but the electrical energy produced or the amount of fuel (municipal waste, biomass, gas) consumed is available. Another example could be that the number of pieces produced in that industry is available and the average energy cost per piece can be extracted (from other periods) in order to do an estimation of the consumption from the production data.

4 Adaptation to appropriate formats

Once it has been analysed that the values of the monitoring data have coherence it has to be translated into the correct format and prepare them to be the data imported to the SO WHAT Tool. The adaptation of the data must be presented in the form:

- Correct monitoring data frequency (daily, hourly, minute, etc.);
- Transformation to the universal system or the one needed in the tool;
- With all the errors fixed;
- With correct format of decimal or thousand separators;
- Valid name of the variables in order to be read by the tool.

5 Estimation of end users demand

Though waste heat/ cold has its origin in different industrial environment with a wide variety of generation profiles, consumers could be any user connected to the district heating. Flows from the DH&CN will follow demand curves depending on the use of the building, so that's why tools that can create flexible profiles are required. The ones selected to help with this module are *LoadProfileGenerator (LPG)* and *City Energy Analyst (CEA)*. Both works in a different way depending on the approach the user wants to give to the demand model, LPG is based on people behaviour of the household and CEA focuses more in a detailed building model.

4.1. LoadProfileGenerator Tool (LPG)

LoadProfileGenerator [5] creates load profiles based on a behaviour simulation of the people in a household. This means that the progamme estimates what people are doing at each point in time, based on that it calculates device usage and t he resulting energy usage. Is a tool for MS Windows to generate load profiles for households for electricity, gas, hot water and cold water.

In the LPG there are more than 100 devices pre-configured based on market research. These devices are arranged together with load curves into actions. From this energy load curves are calculated by adding up the energy use of each device at each point in time. The profiles are generated for hot





water, cold water, electricity and gas. For heating and cooling demand there is a simple degree day model to help generate a rough estimate of the consumption.

This is not a building simulator though, and if you need a detailed building model, other software might be the better choice. The focus of the LPG is the human behavior.

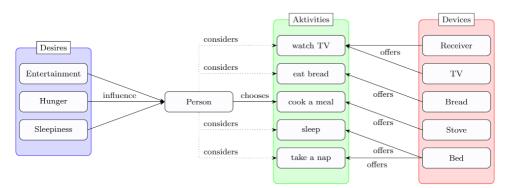
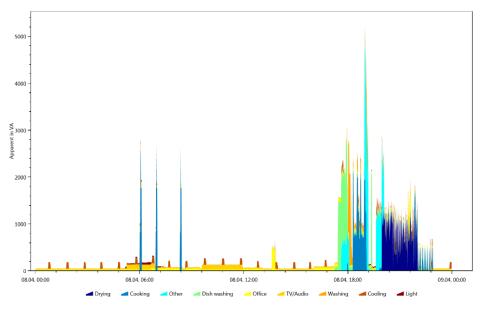


Figure 9: example of human behaviour simulator that can be modelled by LPG [6]

The limitation of this tool is that it can only generate residential profiles, nonetheless it offers a detail level that are not possible in other tools. An example of demand profile is presented in the next figure.





4.2. City Energy Analyst demand profile generator

City Energy Analyst is an urban building simulation platform started in 2013 that combines knowledge of urban planning and energy systems engineering in an integrated simulation platform. This allows to study of the effects, trade-offs and synergies of urban design options and energy infrastructure plans. CEA consists in a collection of tools for the analysis of urban energy systems, one of the tools useful for this training module es the **dynamic demand forecasts**. This can be obtain following steps of the building specification that will be explained below.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847097



4.2.1. Inputs

During the definition of the buildings selected a **set of inputs** are needed to be obtain before being able to run the demand. In every step of the modelling there is an option to add more advanced parameters, depending on the level of detail desired from the simulation.

DATA MANAGEMENT

Archetypes Mapper: level of detail of the assign building properties and systems types from the archetypes database

Zone weather: sets the weather file to use for simulation for a scenario, the user can insert a weather file or use name of the wather files contained within the CEA database

Surrondings helper: query geometry of surrounding buildings from Open Street Maps. Users can define the perimeter buffer length to the surrounding area, the average height of buildings withing the surrondings and average number of floors above ground of all buildings within the surrondings. The default data comes from Open Street Maps.

Terrain helper: query topography with a fixed elevation

Steets helper: query streets geometry from Open Street Maps



Figure 11: buildings studied (left), surroundings (middle) and streets (right). Example in CEA Dashboardn DEMAND FORECASTING

Building Solar Radiation: CEA uses DaySim to calculate solar radiation for a scenario. Daysim is a validated daylighting analysis software that calculates the annual daylight availability in arbitrary buildings based on the radiance backward raytracer, it uses the lightswitch occupant behaviour model to mimic occupant use of personal controls and to predict energy savings from automated lighting controls. User can modify the simulation parameters.

Building schedules: use CEA models and input schedules to estimate the occupancy profile of buildings. The next example represents a typical Hospital schedule, in CEA there are different types of uses profiles such as: *coolroom, residential, foodstore, gym, hospital, hotel, industrial, laboratory, library, museum, office, parking, restaurant, school, swimming* and *university*. In addition, new use-types can be created if it is needed. An example of the schedule for the hospital use type is presented in Table 1 and Table 2.





Table 1: default month	ly/ yearly	multiplier schedule	(Hospital use-type)

HOSPITAL	1	2	3	4	5	6	7	8	9	10	11	12
Monthly multiplier	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8

HOSPITAL OCCUPANCY	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Weekday	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.6	0.8	1	1	0.9	0.7	0.8	1	0.9	0.8	0.6	0.4	0.4	0.4	0.4	0.4	0.4
Saturday	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.6	0.8	1	1	0.9	0.7	0.8	1	0.9	0.8	0.6	0.4	0.4	0.4	0.4	0.4	0.4
Sunday	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.6	0.8	1	1	0.9	0.7	0.8	1	0.9	0.8	0.6	0.4	0.4	0.4	0.4	0.4	0.4

Not only occupancy is scheduled for the 3 types os profile (weekday, saturday and sunday), but also the following thermal and electrical loads are replied: *appliances, lighting, water, heating, cooling, processes, servers* and *electromobility*. Furthermore, it includes the necessary parameters to take into account the indoor comfort and the internal loads depending on the use profile. After following all the steps mentiones and linking every building with its corresponding schedule according to its particular use, the simulation is ready to calculate the demand. First the specifications for the processing, the different variables and finally the time step resolution output are selected. Figure 12 and Figure 13 show two examples of plots generated in CEA, however, there are many representation possibilities.

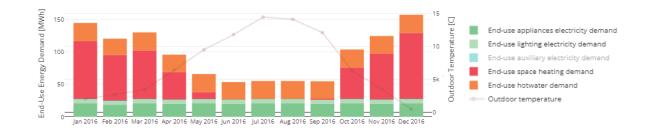


Figure 12: load curve for district (monthly) generated in CEA

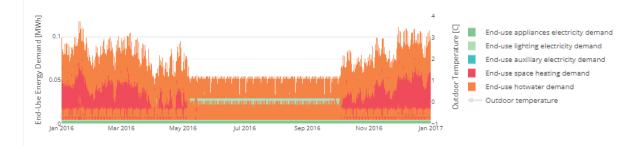


Figure 13: load curve for one of the buildings (Hourly) generated in CEA



Horizon 2020 European Union Funding for Research & Innovation This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847097



6 References

- [1] W. Hillier, "Step-by-Step Guide to the Data Analysis Process," *February*, 2022. https://careerfoundry.com/en/blog/data-analytics/the-data-analysis-process-step-by-step/ (accessed Jul. 21, 2022).
- [2] A. Jain, A. Mani, and A. S. Siddiqui, "Network architecture for demand response implementation in smart grid," Int. J. Syst. Assur. Eng. Manag., vol. 10, 2019, [Online]. Available: https://link.springer.com/article/10.1007/s13198-019-00891-w/figures/5.
- [3] I. Peña and F. Morentín, "Monitoring Management System', SO WHAT H2020 Project, Deliverable 5.1," 2022. [Online]. Available: www.sowhatproject.eu.
- [4] TLV, "Typical examples of waste heat recovery." https://www.tlv.com/global/LA/steamtheory/waste-heat-recovery.html (accessed Jul. 01, 2022).
- [5] N. Pflugradt, "Load Profile Generator (FAQ)," 2020. https://www.loadprofilegenerator.de/faq/ (accessed Jul. 02, 2022).
- [6] "LoadProfileGenerator official page," 2020. https://www.loadprofilegenerator.de/ (accessed Jul. 02, 2022).

