ISEC 2024 – 3rd International Sustainable Energy Conference
Positive Energy Buildings and Districts
https://doi.org/10.52825/isec.v1i.1137

Published: 22 Apr. 2024

How Can Non-Intrusive Load Monitoring Contribute to the Assessment of the Smart Readiness Indicator?

Felix Rehmann[*1,2], Siling Chen[1,2], Falk Cudok[1,2], and Rita Streblow[1,2]

[1] Digital Networking of Buildings, Energy Supply Systems and Users, Einstein Center Digital Future

[2]Technische Universität Berlin

*Correspondence: Felix Rehmann, rehmann@tu-berlin.de

Abstract. The Smart Readiness Indicator (SRI) is a framework introduced by the EU in 2018 to assess smart buildings in various aspects. However, the SRI has been criticized for several limitations, including its ambiguous service definitions. This paper proposes the application of Non-Intrusive-Load Monitoring (NILM) technology to enhance SRI evaluation on the example of SRI service E-12. NILM can be used to disaggregate energy consumption data to end use levels and allows for granular non-intrusive energy consumption measurement. The study involves a rigorous methodology using open sensor data and NILM algorithms to evaluate device-specific energy consumption We evaluate the IDEAL dataset and three different frequencies (5s, 15min, 1h), three different algorithms (CO, RNN, Seq2Point) and one data imputation strategies (forward filling). The results show that with a higher frequency, the performance metrics (F-score, normalized absolute error) increase. Regarding further considerations, we identify a trade-off between resource and energy efficiency, as well as privacy considerations with increasing measurement frequency. To achieve its aims for awareness, the SRI needs to consider interoperability and appropriate aggregations (frequency and spatial).

Keywords: Smart Readiness Indicator, Non-Intrusive-Load Monitoring, Smart Meter, Monitoring, Energy Efficiency

1. Introduction

The European Union introduced the concept of the Smart Readiness Indicator (SRI) in 2018 to raise awareness for three key functionalities of smart buildings: occupant-centric control, grid flexibility and energy optimization [1]. The SRI is a framework that covers nine technical domains and standardizes the evaluation of a buildings' readiness to achieve the core objectives. Electricity, as one of the domains, is evaluated based on criteria such as monitoring of energy consumption or application of sensors. The SRI has been the subject of various critiques in academic and policy literature, mainly because it (a) insufficiently considers building typologies and system types [2], [3] (b) lacks integration of the building and it's technology into districts and cities [3], [4] (c) contains abstract service level definitions [5] and (d) focuses on the design outcomes rather than the real impact of the technology installed on the energy consumption in a building [6]. Critique is paramount considering the fact that Information and Communication Technology (ICT) that could reduce the energy consumption also requires resources such as electricity [7], [8].

Admittedly, "smarter" buildings do not automatically lead to better buildings considering energy optimization, grid flexibility or occupant-centric control. In addition, the SRI does not provide sufficient guidelines on technical requirements for each specific level and the gained benefits. Besides hardware, software can often also be used to achieve significant increases of SRI scores [9]. Hence, the evaluation of the SRI should consider the specific implementations (e.g., number of sensors, effect of software) and its impact in comparison to the anticipated benefits.

Addressing that, this research delves into complementing the SRI evaluation criteria, by providing ecological, data governance and technical perspectives for consideration in implementation of a SRI service. By applying the technique of Non-Intrusive-Load Monitoring (NILM), we provide an example to investigate the challenges associated with data collection and its availability for the SRI, emphasizing the importance of clear data requirements. The approach highlights the need for standardized data reporting and integration mechanisms, promoting data transparency and reliability. For this the SRI Service E-12 "Feedback – Reporting Information - Reporting information regarding electricity consumption" which has the precondition "Always to be assessed" is considered.

NILM is a technology that enables non-intrusive measurement of individual end uses of a measurement point through algorithms. This allows for the identification and tracking of device usage, without individual measurements. As a result, device-specific feedback on usage behavior and consumption can be provided. Such direct feedback can lead to up to 12 percent energy savings compared to no feedback on electricity consumption [10]. Considering the efficacy of NILM and the current limitation of SRI, we explore how NILM can contribute to the assessment of the SRI and help define more precise requirements. We achieve this by discussing an SRI service in detail and providing arguments and alternatives within a broader scope, considering arguments and requirements from different perspectives such ecological, system integration and data privacy. For this, we discuss several options on implementing the SRI and demonstrate the achieved potential effects by using Non-Intrusive-Load Monitoring (NILM).

2. Methodology

Following the research framework introduced above, this section describes the case study and the workflow of this study. Figure 1 provides a visual illustration of this, in addition to the introduced research framework.

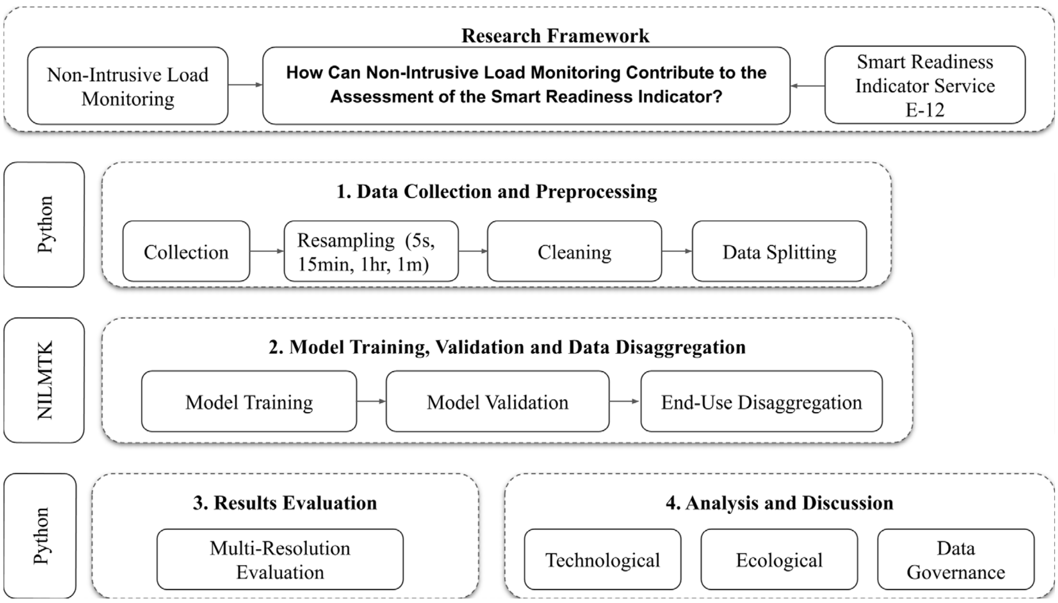


Figure 1. *The proposed study design.*

2.1 SRI-Service

The SRI encompasses 57 pre-defined services. Evaluating them all in this paper is not feasible. Instead, we focus on discussing a specific service and giving broader criteria and methodology for evaluation of another service. As an example, we take the SRI service E-12: Feedback – Reporting Information Reporting information regarding electricity consumption. The service and its functionalities levels are stated in Table 1. The service and its definition are taken from the SRI sheet version 4.4. We consider this service to be central, as the collection and processing of information on energy consumption are essential to enable other services such as control. Furthermore, electrification through renewable energies is considered one of the main strategies for decarbonizing the energy consumption of the existing building stock [11].

Table 1. *The SRI service E-12 and its functionality levels [12]*

Smart ready service	**Functionality level 0 (as non-smart default)**	**Functionality level 1**	**Functionality level 2**	**Functionality level 3**	**Functionality level 4**
Reporting information regarding electricity consumption	None	Reporting on current electricity consumption on building level	Real-time feedback or benchmarking on building level	Real-time feedback or benchmarking on appliance level	Real-time feedback or benchmarking on appliance level with automated personalized recommendations

2.2 Case Study

There are a variety of datasets which are commonly used in NILM e.g. [13], [14]. For further analysis, we chose the IDEAL dataset, which contains metered electricity data from 255 homes in the UK and Scotland. Data was recorded at 1 Hz up to 23 months per household [15]. Due to various reason described by the authors sometimes data is lacking (e.g., transmission errors) or was not recorded during a specific time (e.g. household was not yet part of the study). Household 73 is one of households with the most recorded data and was hence selected.

2.3 Data Processing and Result Analysis

As displayed in Figure 1, the workflow of our work consists of four parts: (1) data collection and pre-processing; (2) NILMTK model training and validation; (3) result evaluation and (4) analysis. Each part contains multiple steps. First, data is collected and pre-processed. For the selected household (73), we extract the metered electricity consumption data collected from January 1st, 2018 till June 14th, 2018 (164 days) and resample the data into three different frequencies: 5 seconds, 15 minutes, and 1 hour. Another frequency, 1 month, is initially planned but not implemented. This is because a training period exceeding 3 months yields more convincing disaggregation results for monthly frequency, but demands significantly greater computational effort due to the large number of data points involved in 5-second time series (e.g., around 1,067,040 data points within a 3-month training period). Due to computational restriction this experiment was not feasible. The resampled data are subsequently being cleaned, where measurements that are consecutively missing for more than 5 time steps from either any device or the total meter are removed. Missing data that have a missing period of shorter than 5 time steps can be supplemented using forward filling (i.e., missing values are

filled with the last valid observation forward). In the next step, the cleaned data are divided into training and testing datasets (with a division ratio of 0.7 for training).
Second, several NILM algorithms from the open-source NILMTK[1] Library [14], [16] are applied to identify energy consumption of devices, including CO, RNN and Seq2Point. In this process, end use disaggregation models are trained and validated for each device. Eventually, the total electricity consumption data is disaggregated using the trained models. The process iteratively runs for all investigated frequencies, with slight variation of model settings as indicated in Table 2.

Table 2. *Overview of model settings for different resolutions. The setting is based on previous experimental runs using the models and the data, considering the covered time span (that includes daily or weekly patterns) and the required computational time.*

	Number of epochs	Batch size
5 seconds	50	32768
15 minutes		1024
1 hour		256

Third, evaluation metrics are defined and used to access the efficacy of disaggregation performances across different resample frequencies. We applied a commonly used metric, the F-score, which the qualitative identification of end-uses and are calculated based on True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The calculation of F-score is described as follows (Batra et al., 2014) [16]:

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F - score = 2 \times \frac{precision \times recall}{precision + recall}$$

In addition, a metric to assess qualitative performance of end-use disaggregation is used: normalized absolute error (NAE). NAE is calculated by the sum of absolute differences between the prediction and measurement of each time step (yt and xt, for the time span of 1 to n), normalized by the sum of ground truth data, which is expressed as:

$$normalizedabsoluteerrorfordevicea, NAE_a = \frac{\sum_{t=0}^{n} \left| y_{a,t} - x_{a,t} \right|}{\sum_{t=0}^{n} x_{a,t}}$$

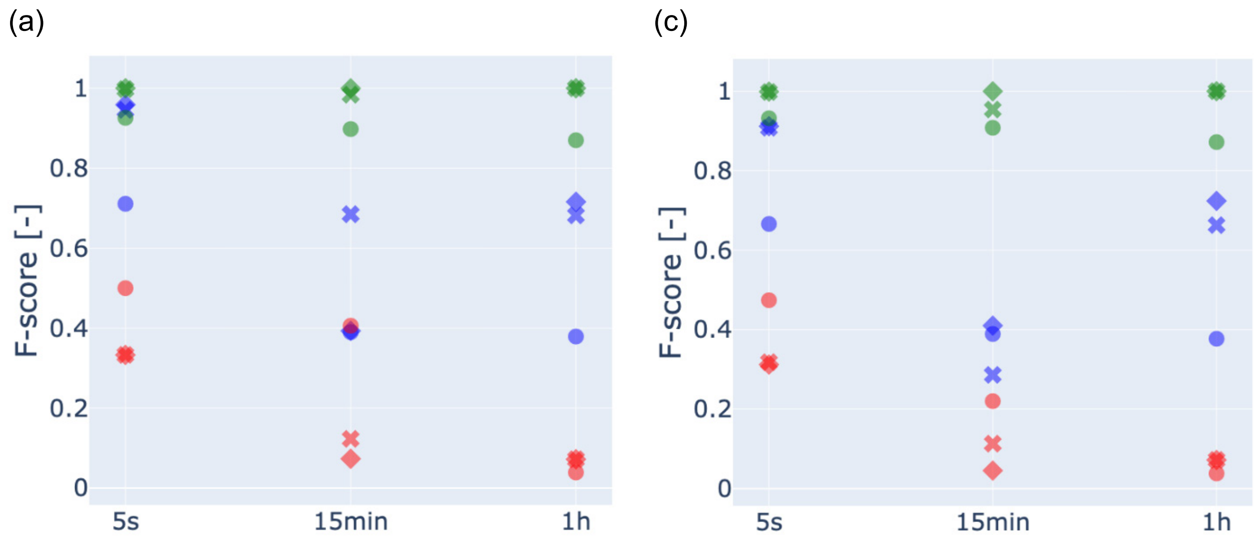
The reason for selecting NAE that normalizes the accumulative errors with total consumption data (instead of other parameters, e.g., Mean Absolute Errors, as frequently used in other literature) is that with normalization, the influence of device-specific energy consumption data can be eliminated. This makes prediction performance easily comparable across various devices. The above-mentioned process, including model training, validation, and result analysis, was conducted using Python 3.8.

3. Results and discussion

3.1 End-use Prediction Performances across Sampling Frequencies, Devices, and Methods

As previously introduced, in this study we use three different algorithms to train and predict end uses based on total electricity consumption time series. The same workflow is applied for three different resampling frequencies. The performance metrics provide insights on the trade-offs between disaggregation performance and metering frequency, which impacts the effort for data acquisition, processing, and storage. The results are shown in Figure 2, where subplots (a) and (b) are results with input dataset where consecutive missing data are removed, while (c) and (d) are from the dataset with an additional missing data filling process.

First, comparing performance metrics with the same input data across different data sampling frequencies, both qualitative and quantitative identification of end uses are better with higher data resolutions. More specifically, F-score, which ideally would approximate 1, is of the highest average value when resampled with a 5s frequency (cf. e.g., Figure 2 (a)). On the contrary, given an optimal disaggregation performance, the normalized absolute error which reflects the deviation between prediction and measurement, would be close to 0. For this metric, the highest resampling frequency (5s) also obtains the best results (cf. e.g., Figure 2 (b)). With a decrease in sampling frequency, the normalized absolute error increases, indicating a worsen disaggregation performance.



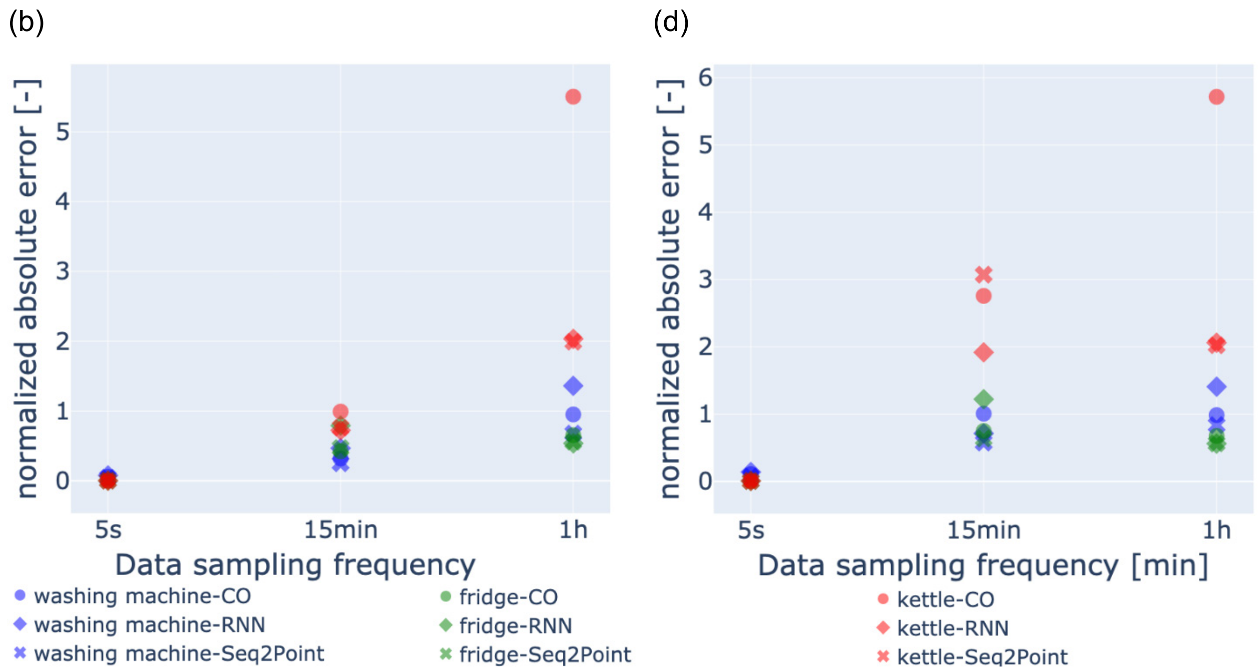


Figure 2. *Illustration of prediction performance metrics, F-score and normalized absolute error, for prediction of different end uses (washing machine, fridge, and kettle) with the experimented data resolutions (5s, 15min, 1h) and applied algorithms (CO, RNN, Seq2Point). Subplots (a) and (b) are results based on the dataset with consecutive missing data removed, while subplots (c) and (d) are based on the dataset with additional forward filling of none-consecutive missing data.*

Second, even for the same sampling frequency, disaggregation performance varies among different devices, which are illustrated in different colors. In general, fridge has the highest F-scores compared to the other devices, while kettle are less correctly identified, resulting in the much lower F-scores values. For normalized absolute error, while the identification of kettle performs slightly better for the 5s resolution, when resampled into a larger time step (15min, 1h). the algorithms cannot quantitatively identify the usage of kettle anymore. Notably, with the current experiment settings, no applied algorithm can be identified as the perfect method, as their performances differ between devices and resolutions. What should be also noted is that the variation between devices stretches when the resampling frequency is increased. While fridge might still remain identifiable and could be relatively quantitatively well predicted, the other devices (especially kettle) become less possible to identify – both qualitatively and quantitatively. The differences of prediction performance between devices can indicate the impact of energy consumption profiles. The energy consumption profile of a kettle consists of multiple small time steps with a relative low energy usage, making it difficult to predict and especially sensitive to sampling frequencies. On the other hand, fridge is of a rather constant energy consumption pattern and hence the highest predictable.

Third, no improvement is observed in filling NAN with forward filling. Although the input time series is prolonged, no or even contra improvement is obtained. However, this does not exclude any other filling method – further methods can be tested while considering the computational efforts.

3.2 Impact of Data Quantity and Quality

In this study, we resample measurement data with a total duration of 164 days in out experiment. In an optimal case where data can be collected, transmitted, and stored correctly at all time steps, the 49 days testing days (30% of the total duration) would equal to 846720, 4704, and 1176 data points for 5s, 15min, and 1h resampling frequencies, respectively. As our results in Section 4.1 show, better prediction performances can be achieved with a higher resampling frequency. This is not only because a higher frequency enables better records of the energy consumption profiles (e.g., time series of kettle usage with shorter and smaller fluctuations can be recorded), but also because of the potentially higher density of data used for trainings of the data-driven disaggregation algorithms. Hence both data quantity and quality are essential for disaggregating end uses from the total electricity consumption data.

As an example, Figure 3 visualizes the variation of disaggregation performances depending on the input data frequency and the selected algorithm. With a higher frequency (5s), the disaggregated energy consumption time series (a washing machine from home 73) using both algorithm RNN and algorithm Seq2Point can fit the fluctuation and magnitude of the resampled ground truth time series better, while CO interprets much smaller fluctuations from the total consumption time series. When data are resampled with a lower frequency, 15min, the time series is described by fewer data points. While the algorithms especially Seq2Point can disaggregate the time series relatively well, the conveyed information is much reduced. For instance, neither can peak load nor numbers of energy consumption peaks can be interpreted from the 15min-resolution time series. While resampled to the 1h resolution, the profile becomes a single point, giving even less information. Accordingly, the peak consumption is reduced, as during the resampling process, average values are taken from the closest time steps.

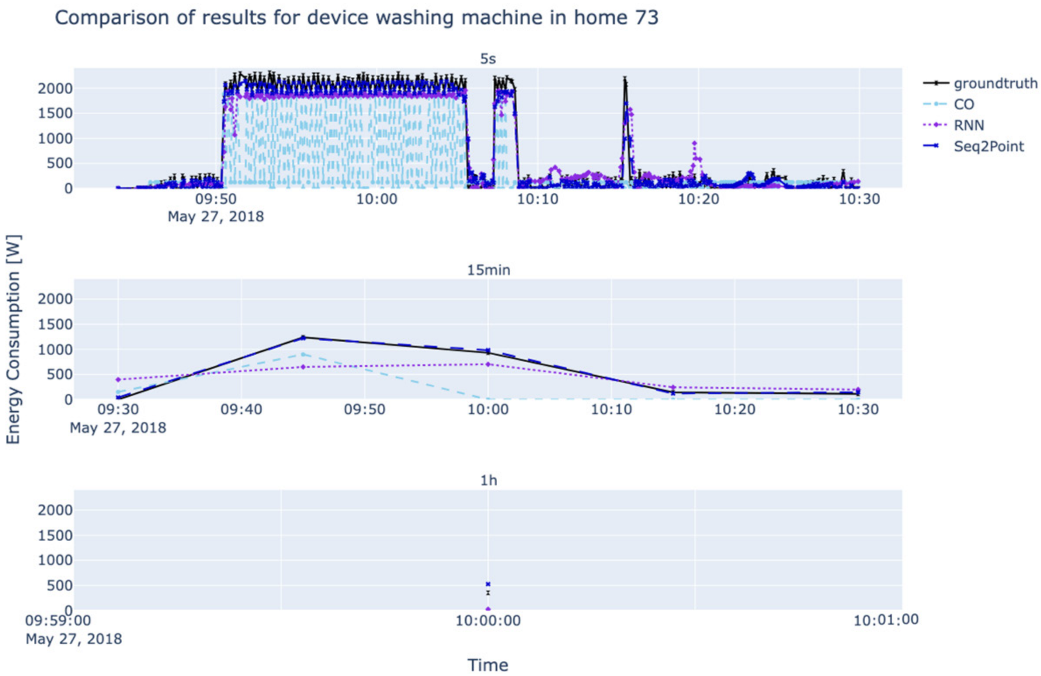


Figure 3. *Comparison of different disaggregation results using three resampled time series of a washing machine from home 73 in the IDEAL dataset (zoomed to 9:30-10:30, May 27, 2018). Time series of the ground truth data (measurement data, resampled to 5s, 15 minutes and 1 hour) are shown in solid black color, whereas the disaggregated results using resampled total home consumption data are displayed in blue series dash or dot lines with different markers: CO – circle markers in sky blue, RNN – diamond marker in blue violet, and Seq2Point – x markers in medium blue. Different frequencies are considered as "real-time".*

However, having data of a higher sampling frequency does not necessarily ensure good disaggregation performances. Although not influencing the overall evaluation shown in Figure 2, there are time steps with abundant missing values within the experimented dataset, for which the disaggregation algorithms cannot properly perform. An example is shown in Figure 4, where more missing data exists when resampled with 5s than with the 15min frequency. Consequently, the 5s time series presents less temporal fluctuation and also generates worse disaggregation outcomes.

To examine the quantity and partially the quality of the data, Figure 5 is created to visualize the number of available data points in the testing dataset. Although 5s resolution could provide data of a higher resolution, more data lost is observed in the investigated time frame. Especially little data can be collected and used to train the algorithms to identify usage of kettle. Although forward filling is not proved effective in our study, filling of missing data could magnificently enrich the input dataset. With missing data filling, we almost obtain a complete dataset for the 15min frequency (around 94% data points can be collected). Because not much missing data exist in the resampled 1h-resolution data, an implementation of data filling is not essential in our case (available data increase from 94% to 98% after the filling process). With this analysis, we conclude that measuring with a higher frequency can lead to more resource exhaustive data handling processes, i.e., more data points are generated and should be transmitted and processed, which may increase the chances of missing data. Although missing data filling strategies can be carried out to make up for the data lost, attention in planning, implementing, and impact evaluating of sampling with high frequencies should be paid.

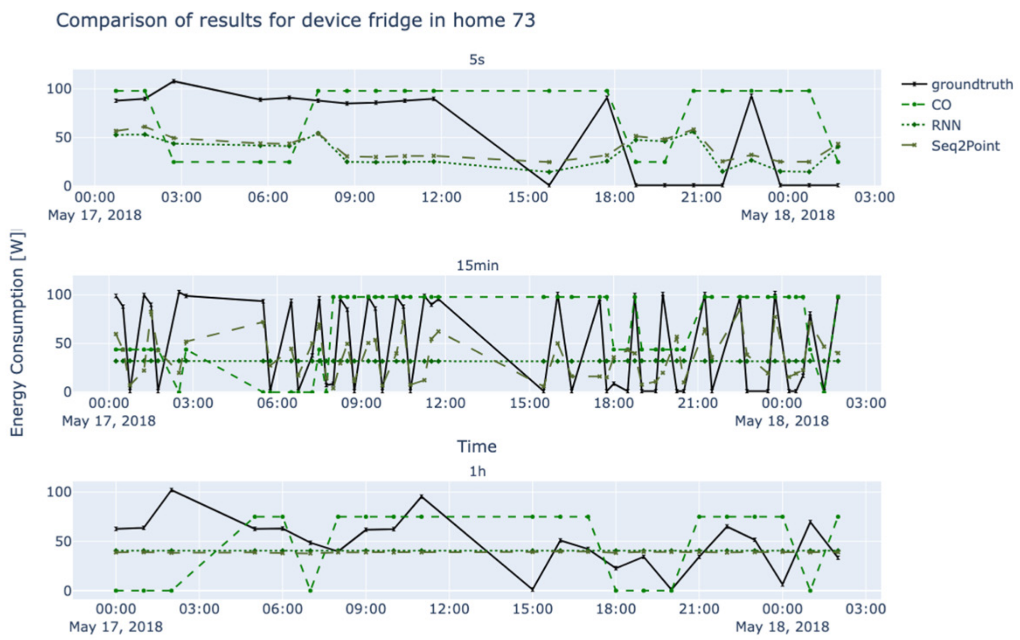


Figure 4. *Comparison of different disaggregation results using three resampled time series of a washing machine from home 73 in the IDEAL dataset (zoomed to 00:00 on May 17, 2018, till 02:00 on May 18, 2018). Time series of the ground truth data (measurement data, resampled to 5s, 15min and 1 hour) are shown in solid black color, whereas the disaggregated results using resampled total home consumption data are displayed in blue series dash or dot lines with different markers: CO – circle markers in green, RNN – diamond marker in dark green, and Seq2Point – x markers in dark olive green. Different frequencies are considered as "real-time".*

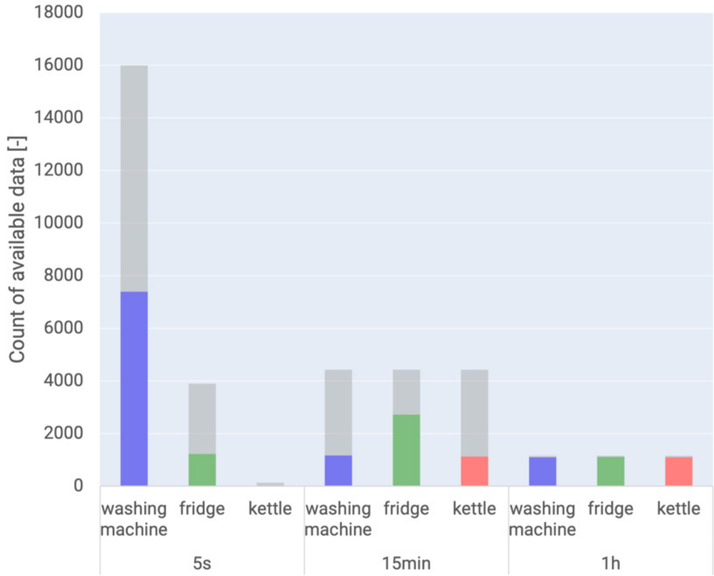


Figure 5. *An overview of available data for various devices when resampled to different frequencies. For each device, the lower bar with a specific color indicates the number of available data points after consecutive missing data are removed, while together with the grey color bar above, the column illustrates the number of data points obtained after filling missing data with forward filling.*

3.3 A technical perspective

There is a variety of ways in which a digital application pursuing a specific goal can be described or developed [17]. This makes a general evaluation of digital applications unattainable. However, a few general aspects should be considered within the implementation of digital applications and hence should be integrated in the SRI governance framework. These are interoperability, privacy, and robust processes, whenever necessary. A often criticized aspect in the context of building control technologies is the lack of commonly applied interfaces and standards [18]. Furthermore, the data and systems should be interoperable. For example, it is conceivable that closed systems may meet the SRI criteria requirements but lead to increased efforts if (partial) system changes occur. In [19], criteria for the interoperability of systems across domains are defined. These include for example, considering technical interoperability through established standards or joint definitions of criteria for security and privacy. Reflecting on the service E-12 and the results of our analysis, three key questions arise.

1. What to and how to report data effectively?

If the reporting is not in relation to a person and their actions, or not comprehensible to them, it does not bring added value to the service. Benchmarks and feedback should be comparable and understandable [20]. If the frequency or spatial aggregation is not appropriate, e.g., in case of large buildings, or the given statements are wrong, e.g., in case of false or mismatching feedback and benchmarks, the service might not fulfill its duties. Hence, the scale and comparison need to be designed, such that benchmarking and recommendations are considerable (e.g., comparing flats of same size or orientation) and understandable. One should also consider the kind of equipment that is monitored, as there is a huge difference between a kettle and a fridge, and how their electricity demands are flexible and variable.

2. What is the necessary frequency to gather sufficient data to provide valuable information?

As previously stated, frequency and spatial aggregation should be appropriate depending on the purpose. Considering energy consumption and our study settings, we recommend 15 minutes as a good trade-off between accuracy and required resources (measurement, computing, for instance). Further improvement can be achieved with fine tunings of disaggregation model parameters. As discussed in Section 3.2, we see variety of both data quantity and quality in further analysis and reporting, hence it is also essential to ensure the quality of data collection, indicated by different criteria such as data consistency, accuracy, precision, and completeness.

3. How should the infrastructure be set up to provide the functionality levels?

Considering the increased energetic and resource efforts, to record, store, process, and visualize data at higher frequency and for more devices, as well as the increased effort for monitoring system integration, one should aim to confirm integration. Software-based approaches, as we show in this paper are a sufficient alternative to hardware-based monitoring, once the models have been trained and validated.

3.4 An ecological perspective

From an ecological perspective, data should only be used measured, processed, stored, and analyzed in which it supports a specific use. As the SRI evaluates buildings regarding the capability to perform 3 key tasks, optimization of energy efficiency, and in-use performance, adoption to occupant needs, adoption to signals from the grid. Evaluating the ecological benefits of the SRI should consider these three arguments. Hence, data should only be gathered and processed in the amount which is needed to fulfill the task.

Regarding energy efficiency, multi-family houses, functionality levels 1 and 2 offer no added value as they do not allow for individual household consumption analysis, making it impossible to identify specific high-consumption units. However, for single-family houses, these levels can identify unusually high consumption. Functionality levels 3 and 4, which provide device-level consumption details and personalized recommendations, respectively, are more helpful in identifying devices that consume high amounts of energy and are thus more conducive to achieving goals. Regarding the ability to meet user needs, whether the digital application meets a goal of providing relevant information to a person's actions, the ability is met at all service levels for single family homes. However, at the building level in multi-family houses or large nonresidential buildings this goal is typically not sufficiently met. For effective network interaction, in addition to reporting, a signal regarding the network or supply situation must be transmitted. If this signal is directed at individual households, it lays the groundwork for network interaction, hence we consider a positive rating for the higher functionality levels.

3.5 A data protection perspective

Considering data protection as a legal assessment of the three goals of the SRI, or the functional levels contained therein, the evaluation must (like any data protection evaluation) be measured against the adherence to the underlying fundamental assessment principles. The various functional levels can be examined to a certain extent abstractly in terms of the inherent data protection risks they pose. Thus, there is a tendency for the risk potential to increase with higher functional levels, and therefore the effort required to ensure compliance with processing principles (e.g., data security) increases as well. This assumption is based on the fact that with higher data resolution and availability, the quality and quantity of information that can be derived from a dataset also increase, as does the likelihood that various datasets can be linked together to derive further information.

There are two central limitations to the above statement. Firstly, a higher data protection relevance does not necessarily mean that a functional level is per se "worse" than one with lower relevance. It simply means that the technical, organizational, and legal requirements that must be placed on the system to ensure effective protection of the rights of the affected individuals increase. From the perspective of the responsible parties, this may indeed be associated with higher resource expenditure. Secondly, neither the existence of the tendency described above nor the existence of the principles presented above should obscure the fact that the data protection legal assessment of an application depends significantly on specific circumstances, or the context of data processing and the technical-organizational framework used for data processing.

To enhance the informative value of general statements about the data protection risk of a functional level, for this reason, as is already within the scope of data sufficiency, a distinction can be made according to (at least two) building types: single-family homes and multi-family homes or large non-residential buildings. This is based on a further assumption: the smaller the reference object (i.e., the building to which the data relates) of the data processing is, or the fewer the number of people who are present in the reference object the more information can be derived about the person.

4. Conclusion and outlook

Digitalization of energy systems should evaluate the trade-offs between technical investments, added value, data protection at an appropriate scale, and ecological net-benefits of the system. As discussed in this paper with the example of the SRI Service E-12, increasing the SRI functionalities generates more benefits regarding optimizing energy efficiency, enhancing user needs and system integration through grid signals. For each single service, a variety of technology, digital applications and frameworks can be used to achieve functionality levels.

The benefits and drawbacks of each level often depend on the scale and the frequency of the service. From a data protection perspective, even slightly collecting data is worse than collecting no data at all. From an ecological perspective, installing equipment or software which cannot be used to achieve the proposed benefits has affects the ecological impact negatively. We consider it from this perspective to be good to directly aim for interoperability at the highest service level.

The NILM-based analysis provided a perspective, on how governance frameworks such as the SRI should consider specific implementations to sustain their net positive benefits. As our results show and especially Figure 4 terms such as real-time should be evaluated considering the expected benefits. If for example, feedback on a personal level is not provided and interoperability is not designed, but data is recorded at a high frequency with respective costs, the proposed SRI system benefits are not achieved. It is up to governmental agencies, industry, and research to define technologies, and methods that can fulfill possible goals without wasting resources and energy.

For further analysis, we suggest considering the SRI as a framework which can be used to map digital applications in the operation phase of the building. Further studies could for example aim at mapping digital applications within the research sector towards these services and their functionality levels. It provides use-cases, which are not standardized but still are comparable within a given area, as we show within this paper. Additionally, state-of-the-art imputation studies can be applied to the preprocessing and the criticality of data quality and quantity can be further explored.

Data availability statement

Data supporting the statements is obtained from the described repositories and can be accessed from there.

Author contributions

Felix Rehmann (Conceptualization, Data curation, Formal Analysis, Visualization, Writing – original draft, Writing – review & editing), Siling Chen (Conceptualization, Data curation, Formal Analysis, Software, Visualization, Writing – original draft, Writing – review & editing), Falk Cudok (Conceptualization, Writing – Review & Editing), Rita Streblow (Conceptualization, Funding acquisition, Project administration, Supervision, Writing – Review & Editing)

Competing interests

The authors declare that they have no conflict of interest.

Funding

We gratefully acknowledge the financial support by the Federal Ministry for Economic Affairs and Climate Action (BMWK), promotional reference: 03EWB004A.

Acknowledgement

We gratefully acknowledge the support of Astrid Aretz, Jan Kegel and Swantje Gährs for their perspective on the ecological impact of the SRI. We also show our gratitude to Max von Grafenstein and Valentin Rupp for their perspective on data protection and data governance and their impact on the SRI.

Further Remarks

The data protection criteria catalogue will be published at: https://wissen-digital-ewb.de/pages/criteriaCatalog

References

[1] *Directive (EU) 2018/ of the European Parliament and of the Council of 30 May 2018 amending Directive 2010/31/EU on the energy performance of buildings and Directive 2012/27/EU on energy efficiency*, vol. 2018/844. 2018. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018L0844

[2] V. Apostolopoulos, P. Giourka, G. Martinopoulos, K. Angelakoglou, K. Kourtzanidis, and N. Nikolopoulos, "Smart readiness indicator evaluation and cost estimation of smart retrofitting scenarios - A comparative case-study in European residential buildings," *Sustain. Cities Soc.*, vol. 82, p. 103921, Jul. 2022, doi: 10.1016/j.scs.2022.103921.

[3] G. Plienaitis, M. Daukšys, E. Demetriou, B. Ioannou, P. A. Fokaides, and L. Seduikyte, "Evaluation of the Smart Readiness Indicator for Educational Buildings," *Buildings*, vol. 13, no. 4, p. 888, Mar. 2023, doi: 10.3390/buildings13040888.

[4] T. Märzinger and D. Österreicher, "Extending the Application of the Smart Readiness Indicator—A Methodology for the Quantitative Assessment of the Load Shifting Potential of Smart Districts," *Energies*, vol. 13, no. 13, p. 3507, Jul. 2020, doi: 10.3390/en13133507.

[5] A. Ożadowicz, "A Hybrid Approach in Design of Building Energy Management System with Smart Readiness Indicator and Building as a Service Concept," *Energies*, vol. 15, no. 4, p. 1432, Feb. 2022, doi: 10.3390/en15041432.

[6] M. Offermann *et al.*, “Anpassung der SRI-Systematik für eine Einführung in Deutschland.” 2022. Accessed: Oct. 25, 2023. [Online]. Available: https://www.bmwk.de/Redaktion/DE/Publikationen/Energie/bericht-zur-angepassten-sri-systematik-fur-deutschland.pdf?__blob=publicationFile&v=4

[7] A. Aretz, N. Ouanes, H. Stange, C. Lenk, R. Holzner, and L.-A. Brischke, “Evaluation of the energy saving potential through systematic data collection of the electricity consumption and heating system operation in the building sector,” in *Eceee Summer Study proceedings*, 2022, pp. 1165–1177.

[8] S. Lange, J. Pohl, and T. Santarius, “Digitalization and energy consumption. Does ICT reduce energy demand?,” *Ecol. Econ.*, vol. 176, p. 106760, Oct. 2020, doi: 10.1016/j.ecolecon.2020.106760.

[9] V. Varsami and E. Burman, “An Evaluation of the Smart Readiness Indicator proposed for Buildings,” presented at the 2021 Building Simulation Conference, Sep. 2021. doi: 10.26868/25222708.2021.30880.

[10] K. Carrie Armel, A. Gupta, G. Shrimali, and A. Albert, “Is disaggregation the holy grail of energy efficiency? The case of electricity,” *Energy Policy*, vol. 52, pp. 213–234, Jan. 2013, doi: 10.1016/j.enpol.2012.08.062.

[11] G. Luderer, C. Kost, and D. Sörgel, “Deutschland auf dem Weg zur Klimaneutralität 2045 - Szenarien und Pfade im Modellvergleich,” p. 359 pages, 2021, doi: 10.48485/PIK.2021.006.

[12] Y. Ma and S. Verbeke, “Smart Readiness Indicator (SRI) ASSESSMENT PACKAGE: PRACTICAL GUIDE SRI CALCULATION FRAMEWORK v4.4.” Jan. 18, 2022.

[13] J. Z. Kolter and M. J. Johnson, “REDD: A Public Data Set for Energy Disaggregation Research,” in SustKDD workshop on Data Mining Applications in Sustainability. 2011. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=d85a51e2978f4563ee74bf9a09d3219e03799819

[14] H. Shastri and N. Batra, “Neural network approaches and dataset parser for NILM toolkit,” in *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, Coimbra Portugal: ACM, Nov. 2021, pp. 172–175. doi: 10.1145/3486611.3486652.

[15] M. Pullinger *et al.*, “The IDEAL household energy dataset, electricity, gas, contextual sensor data and survey data for 255 UK homes,” *Sci. Data*, vol. 8, no. 1, p. 146, May 2021, doi: 10.1038/s41597-021-00921-y.

[16] N. Batra *et al.*, “NILMTK: an open source toolkit for non-intrusive load monitoring,” in *Proceedings of the 5th international conference on Future energy systems*, Cambridge United Kingdom: ACM, Jun. 2014, pp. 265–276. doi: 10.1145/2602044.2602051.

[17] F. Rehmann, F. Cudok, and R. Streblow, “Methods for comparing digital applications in buildings and districts,” *Environ. Res. Infrastruct. Sustain.*, vol. 2, no. 4, p. 045010, Dec. 2022, doi: 10.1088/2634-4505/ac9710.

[18] M. Pritoni *et al.*, “Metadata Schemas and Ontologies for Building Energy Applications: A Critical Review and Use Case Analysis,” *Energies*, vol. 14, no. 7, p. 2024, Apr. 2021, doi: 10.3390/en14072024.

[19] S. Engelsgaard, E. K. Alexandersen, J. Dallaire, and M. Jradi, “IBACSA: An interactive tool for building automation and control systems auditing and smartness evaluation,” *Build. Environ.*, vol. 184, p. 107240, Oct. 2020, doi: 10.1016/j.buildenv.2020.107240.

[20] F. Rehmann, F. Cudok, V. Rupp, J. Kegel, and R. Streblow, “Dashboards: Ziele, Strategien und Umsetzung zur nutzendenfreundlichen Datenaufbereitung,” presented at the 11. Projektleitungstreffen - ENERGIEWENDEBAUEN, pp. 64–74, 2022. [Online]. Available:

https://www.energiewendebauen.de/lw_resource/datapool/systemfiles/agent/ewbpublications/E8CAFC7B833A1757E0537E695E86E4B7/live/document/Dokumentation_11.PL-Treffen.pdf