Article

Benchmarking Optimization-Based Energy Disaggregation Algorithms

Oladayo S. Ajani ¹ , Abhishek Kumar ¹ , Rammohan Mallipeddi ^{1,*} , Swagatam Das ²
and Ponnuthurai Nagarathnam Suganthan ³

¹ Department of Artificial Intelligence, Kyungpook National University, Daegu 37224, Korea; oladayosolomon@gmail.com (O.S.A.); abhishek.kumar.eee13@itbhu.ac.in (A.K.)

² Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata 700108, India; swagatamdas19@yahoo.co.in

³ School of Electrical and Electronic Engineering, Nanyang Technical University, Singapore 639798, Singapore; epnsugan@ntu.edu.sg

* Correspondence: mallipeddi.ram@gmail.com

Abstract: Energy disaggregation (ED), with minimal infrastructure, can create energy awareness and thus promote energy efficiency by providing appliance-level consumption information. However, ED is highly ill-posed and gets complicated with increase in number and type of devices, similarity between devices, measurement errors, etc. To design, test, and benchmark ED algorithms, the availability of open-access energy consumption datasets is crucial. Most datasets in the literature suit data-intensive pattern-based ED algorithms. Recently, optimization-based ED algorithms that only require information regarding the operational states of the devices are being developed. However, the lack of standard datasets and appropriate evaluation metrics is hindering the development of reproducible state-of-the-art optimization-based ED algorithms. Therefore, in this paper, we propose a dataset with multiple instances that are representative of the different challenges posed by ED in practice. Performance indicators to empirically evaluate different optimization-based ED algorithms are summarized. In addition, baseline simulation results of the state-of-the-art optimization-based ED algorithms are presented. The developed dataset, summarization of different metrics, and baseline results are expected to provide a platform for researchers to develop novel optimization-based frameworks, in general, and evolutionary computation-based frameworks in particular to solve ED.

Keywords: non-intrusive load monitoring; optimization-based energy disaggregation; benchmarking; evaluation metrics

Citation: Ajani, O.S.; Kumar, A.; Mallipeddi, R.; Das, S.; Suganthan, P.N. Benchmarking Optimization-Based Energy Disaggregation Algorithms. *Energies* **2022**, *15*, 1600. <https://doi.org/10.3390/en15051600>

Academic Editor: Anastasios Dounis

Received: 13 January 2022

Accepted: 10 February 2022

Published: 22 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the residential sector, which accounts for 30% of global energy consumption [1], providing appliance-level consumption feedback is expected to result in 12% annual energy savings [2] compared to the traditional indirect feedback such as monthly bills. In addition, appliance load monitoring (ALM) can facilitate—(a) identification of faulty and/or energy-inefficient devices [2]; and (b) participate in demand response [3]. Although energy management through the use of computational intelligence methods such as neural network, fuzzy logic, etc., [4,5] as well as transition into sustainable energy sources [6,7] are being explored. The simultaneous advancements in artificial intelligence and smart meters have provided the much necessary impetus to the exponential growth of ALM, which can be intrusive (IALM) or non-intrusive (NIALM) [1]. IALM is more accurate but expensive since it requires that one or more sensors be installed per appliance. However, in NIALM or energy disaggregation (ED) measurements corresponding to the whole house are made through a single sensor, and appliance-level information is obtained through artificial intelligence-based techniques. Recently, ED garnered huge attention from both the research community as well as the industry [2,8,9] because of its capability to promote energy awareness with minimal infrastructure.

ED involves the estimation of the energy consumption, $y_i(t)$ of individual appliances, $i \in 1, 2, \dots, n$, from the aggregated measurements, $y(t)$ obtained at a single point [1,2] over time $t = 1, 2, \dots, T$, such that

$$y(t) = \sum_{i=1}^n y_i(t) + \sigma(t), \quad (1)$$

where $\sigma(t)$ represents noise during measurement. $y(t)$ can be any power feature such as voltage (V), current (I), active power (P) reactive power (Q) and power factor (pf), etc. [10].

Techniques developed to handle the over-parameterized and highly ill-posed ED can be (a) unsupervised, or (b) supervised. Each methodology has its advantages and disadvantages in disaggregating signals [1,2,11] as ED gets complicated due to factors such as an increase in the number and type of devices [1], the similarity between several devices, concurrent switching of multiple devices, and measurement errors [2] etc.

Unsupervised ED approaches such as clustering algorithms [12], Factorial Hidden Markov Models [13] and hybrid algorithms based on sparse signal approximation and Gaussian Mixture Model [14] emphasize unsupervised and generic learning features, but fail when the network contains devices with similar ratings and when the power rating of one device is a linear combination of two or more devices [15]. In Supervised ED methods, representative labeled data are required to train the different associated modules. The type and volume of the required training data depend on the modules present in the algorithm. Based on the modules present and adopted process, the supervised approaches can be categorized as either event-less or event-based approaches [11]. Event-based approaches [16,17] comprise sub-systems that facilitate event detection and event classification. They detect and label the appliance transitions or power events in the aggregated signal [11] using pre-trained systems via supervised or semi-supervised learning. Hence, event-based approaches require labeled training data that includes several power events that occur due to the different appliance transitions. As a result, most event-based approaches are machine learning-based approaches.

On the other hand, event-less approaches [18] attempt to match the aggregated power at each time instance with the consumption of a combination of different devices with the aid of methods such as probabilistic (e.g., Hidden Markov Models), statistical (e.g., Bayesian models), machine-learning, and optimization-based methods. Hence, they require minimal training data compared to the event-based approaches.

Most ED approaches in the literature are machine-learning-based approaches [15,19,20] that face challenges such as:

1. The exponential increase in training data requirement for feature extraction and model construction as the number of appliances increase.
2. Depending on the features employed, data collection needs to be done at high sampling rates for better feature extraction.
3. Since every household is unique concerning the combination of devices present and user-specific usage patterns, the training process must be undertaken separately for each house or fine-tuned with the respective training data.
4. Rare operation or infrequent operation of devices such as coffee makers can create an imbalance in the training data.
5. The performance of the trained model degrades, when there is a slight change in the supply frequency, due to the mismatch in appliance profiles [1].
6. Lack of unified load signatures to model the operation characteristics of various appliance categories.
7. To accommodate new devices into the existing network, the processes of data gathering and training has to be performed again making it ill-suited for real-world implementation.

Optimization-based ED approaches employ information related to appliances such as modes of operation and their rated power to perform ED. Given the information about the different modes of operation and their ratings corresponding to each device, optimization-based ED approaches aim to minimize the squared error between the measured feature vector from the smart meter (say real power) to that of a single device or combinations of devices from the pool of devices in the network [21]. In other words, optimization-based ED alleviates the data-intensive and computationally expensive training process for feature extraction. In addition, new devices can be easily accommodated into the network by extending the pool with information related to appliance states and their rating. In [9,21–23], attempts have been made to solve the ED problem using approaches such as integer programming and evolutionary algorithms. However, research on optimization-based ED is still in preliminary stages due to the lack of a proper standard dataset which is representative of different challenges posed by ED in practice [21]. In addition, the metrics employed for evaluating optimization-based ED algorithms differ significantly even in the limited numbers of works that are reported [9,21–24]. The lack of a standard dataset and appropriate evaluation metrics is hindering the knowledge transfer between different research groups and between academia and industry practitioners.

The goal of this paper is to develop a dataset with multiple instances where each instance poses a challenge corresponding to ED. The dataset will enable a researcher to develop novel problem formulations and algorithmic configurations to address the challenges. In addition, we provide a set of metrics that can evaluate the performance of optimization-based ED algorithms on different aspects. The proposed dataset in combination with evaluation metrics can facilitate sound benchmarking of optimization-based ED algorithms for better reproduction. In addition, the simulation results corresponding to the state-of-the-art optimization-based ED frameworks on the developed dataset with respect to the evaluation metrics considered are summarized.

The rest of the manuscript is structured as follows. Section 2 details a review of the existing literature on the different optimization-based ED frameworks. Section 3 highlights the issues with the existing datasets regarding their suitability for optimization-based ED. Section 4 provides details regarding the design of the new benchmark dataset, while Section 5 summarizes the different metrics suitable for the evaluation of optimization-based ED algorithms. In Section 6, the simulation results of the state-of-the-art optimization-based ED algorithms are summarized and analyzed. Section 7 presents conclusions and some future prospects.

2. Optimization-Based Energy Disaggregation: Literature Review

Generally, electrical appliances tend to operate at different modes, where each mode is characterized by an approximate power rating as shown in Table 1. Given the number of devices (n), their operational states with power rating, the ED problem can be formulated as an optimization problem with single or multi-objective functions [21].

An appliance i with a maximum of l_i non-OFF modes operating at $P_i = [p_i^{(1)}, \dots, p_i^{(l_i)}]^T$ can be represented with l_i virtual two-state devices with ON/OFF (1/0). For n devices each with l_i non-OFF states, the power rating corresponding to the $m = \sum_{i=1}^n l_i$ virtual devices is given by an $(m \times 1)$ vector $P = [P_1, P_2, \dots, P_i, \dots, P_n]^T$. At time t , if the status corresponding to the m virtual 2 state devices is given by

$$S(t) = [s_1^{(1)}(t), \dots, s_1^{(l_1)}(t), \dots, s_i^{(1)}(t), \dots, s_i^{(l_i)}(t), \dots, s_n^{(1)}(t), \dots, s_n^{(l_n)}(t)]^T, \quad (2)$$

where $s^{(j)}(t) \in \{0, 1\}$ for $j = 1, 2, \dots, m$. Here, the objective of ED is to find an appropriate set of devices at each time instance ($S(t)$) to the best fit the measured aggregated signal ($y(t)$). Therefore, the least square error between the estimated and measured aggregate signals can be considered as shown below [22,25,26].

$$\text{minimize } f = \sum_{t=1}^T (y_i(t) - \hat{y}_i(t))^2, \quad (3)$$

where $\hat{y}(t) = S(t)'P$.

A thorough enumeration of all possible combinations of S to obtain the optima to the binary optimization problem given by (3) is prohibitive due to the exponential increase in the number of possible combinations as the number of devices in the network becomes large. In literature, optimization frameworks based on integer programming [23], mixed integer programming [25], evolutionary algorithms [9,21,22,24], etc., have been explored to handle the ED problem formulation given by (3).

The simple ED formulation given by (3) is over-parameterized and therefore the solutions obtained may fail to represent the true appliance operation characteristics. The common empirical issues not factored into the basic formulation in (3) are summarized below.

- (a) **Issue 1:** Appliance i with l_i non-OFF operating modes, represented as l_i virtual devices, may operate in more than one of the possible modes at a given time which is impractical.
- (b) **Issue 2:** Devices designed for continuous operation such as smoke alarms most probably on “stand-by” and rarely switch to high-power states. In such cases, for better performance, it is essential to constrain that at least one of the virtual appliances (l_i) corresponding to continuous appliance i is ON at any given time.
- (c) **Issue 3:** The power rating of one virtual device can be similar to others or can be represented as a linear combination of multiple devices resulting in multiple possible solutions for a given aggregate value obtained from the smart meter.
- (d) **Issue 4:** As illustrated in Table 1, discrete values are employed to represent the power ratings of appliance operational modes in optimization-based ED. And, at time t , the power consumption $y_i(t)$ of the i -th device can be expressed as

$$y_i(t) = \begin{bmatrix} s_i^{(1)}(t) \\ \vdots \\ s_i^{(l_i)}(t) \end{bmatrix} \odot \begin{bmatrix} p_i^{(1)} \\ \vdots \\ p_i^{(l_i)} \end{bmatrix} + \sigma_i(t), \quad (4)$$

where $[s_i^{(1)}(t); \dots; s_i^{(l_i)}(t)]$ denotes the status of l_i non-OFF operational modes of appliance i at time t , $\odot$ represents the element-wise product. However, when an appliance switches states, relative to the smart meter’s sampling rate, the measured power may differ from the discrete power values considered. The unaccounted power corresponding to device i due to the discrete approximation is given by $\sigma_i(t)$, the magnitude of which depends on the operational characteristics of the device, number of modes considered, power rating of the device, etc. From Table 1, it can be observed that substantial power deviations ($\odot$) are attributed to modes with high power ratings. Therefore, $\sigma_i(t)$ is bound to be higher at time t where appliance i consumes high power. In addition, if the power deviation matches with the rated power of a device operating at low power then the formulation given by (3) may consider switching the low power device as well, even though it is not in operation.

- (e) **Issue 5:** Modern smart meters can provide data sampled at high-frequency where successive measurements are possible at extremely short intervals (say 10 s). At such a high sampling rate, the ON/OFF switching events will be sparse because in practice a device switched ON/OFF is expected to be in the same state for a certain period which is much higher than the sampling rate. However, at each time instance, minimization of least square error independently as given in (3) combined with **Issues 3 and 4**, results in frequent appliance switching (ON/OFF). In other words, the formulation in (3) fails to enforce temporal sparsity and therefore the recovered signal may fail

to represent the practical operation of the appliance. The temporal sparsity can be achieved by

$$\text{minimize } TSE(\Delta S) = \sum_{j=1}^m \sum_{t=1}^T |\Delta S^{(j)}(t)|, \quad (5)$$

where $S = [S(1), \dots, S(i), \dots, S(T)]$ is $(m \times T)$ matrix, $TSE(\cdot)$ denotes the total switching events in ΔS given by

$$\Delta S = S \cdot D,$$

where differential matrix (D) of size $T \times (T - 1)$ is given by:

$$D = \begin{bmatrix} -1 & & & \\ 1 & -1 & & \\ & 1 & \ddots & \\ & & \ddots & -1 \\ & & & 1 \end{bmatrix}.$$

In [25], **Issue 1** was addressed by an inequality constraint that enforces the device to operate only in one of the l_i states or completely switching OFF all the l_i virtual devices. **Issue 2** was handled with an equality constraint [25] to enforce that continuous operating devices are operating in at least one of the l_i non-OFF states.

To address the issue of linear combination of devices highlighted as **Issue 3**, it is necessary to resolve the ties when the power rating of a state is similar to the linear combination(s) of multiple power states of other devices. In [25], it has been empirically proven that satisfactory results can be obtained by choosing a combination of appliances where the least number of devices are ON at a given time.

To address the issues pertaining to the continuity of device operation (as a result of **Issues 3, 4, 5**), in [25], the solution obtained by optimizing (3) is further processed by performing—(1) state correction based on available state transition diagram, (2) median filtering, and (3) linear programming-based refinement. In addition, to address issues related to temporal sparsity, two different works referred to as Sparse optimization (Sopt) [27] and Sparse Switching Event Recovering (SSER) [8,28] were proposed.

In Sopt [27], the over-parameterized formulation in (3) is modified by adding regularization terms as shown

$$\begin{aligned} \text{minimize } f &= \sum_{t=1}^T (y(t) - \hat{y}(t))^2 + \\ &\Rightarrow \lambda_1 \sum_{i=1}^n \sum_{t=1}^T \left\| \begin{bmatrix} w_i^{(1)}(t) \\ \vdots \\ w_i^{(l_i)}(t) \end{bmatrix} \odot \begin{bmatrix} s_i^{(1)}(t) \\ \vdots \\ s_i^{(l_i)}(t) \end{bmatrix} \right\|_1 + \\ &\Rightarrow \lambda_1 \sum_{i=1}^n \sum_{t=1}^T k_i \left\| \begin{bmatrix} s_i^{(1)}(t) - s_i^{(1)}(t-1) \\ \vdots \\ s_i^{(l_i)}(t) - s_i^{(l_i)}(t-1) \end{bmatrix} \right\|_\infty, \end{aligned} \quad (6)$$

subject to $\sum_{j=1}^{l_i} s_i^{(j)}(t) = 1, i = 1, \dots, n$, and $t = 1, \dots, T$.

In (6), the second term in combination with equality constraint promotes sparsity in $[s_i^{(1)}(t); \dots; s_i^{(l_i)}(t)]$ and ensures that at least one element is non-zero to address **Issue 2**. In (6), the third term is expected to promote temporal sparsity in $s_i^j(t)$. The non-negative weight vector $[w_i^{(1)}(t), \dots, w_i^{(l_i)}(t)]^T$, hyperparameters (λ_1 and λ_2) and $k_i (i = 1, \dots, n)$ are fine-tuned by cross validation. Specifically, an adequate amount of data is required for training the parameters.

In SSER [8,28], ED is expressed as a constrained single-objective problem, where minimizing the total number of ON/OFF switchings (5) or maximizing the temporal sparsity subject to power limit constraints given by (7) are employed. Given the approximate power deviation variation ($\Theta = [\Theta_1, \Theta_2, \dots, \Theta_m]^T$) corresponding to each power state ($P = [P_1, P_2, \dots, P_m]^T$). At time t , the prospective state vector of appliance $S(t)$ is expected to satisfy

$$S'(t)(P - \Theta) \leq y(t) \leq S'(t)(P + \Theta). \quad (7)$$

In other words, corresponding to each operational mode, in addition to the power rating, deviation from the rated power is expected to be available. This is because the effectiveness of SSER heavily relies on the power variation Θ at any given time instance [8]. However, it is quite challenging to obtain the required value of power variation (Θ) corresponding to every operational mode of each appliance.

In [28], two multi-objective formulations of the ED problem were considered. In the first formulation, considering real power, the two objectives employed are

$$\text{minimize} \begin{cases} f_1 = |y(t) - \hat{y}(t)| \\ f_2 = \phi_o d_o(s(t), s(t-1)) + \phi_s d_s(s(t), s(t-1)), \end{cases} \quad (8)$$

where $d_s(s(t), s(t-1))$ accounts for the number of modes changes between events, while $d_o(s(t), s(t-1))$ accounts for the number of ON/OFF changes. The multi-objective problem provides a set of trade-off solutions, from which an optimal solution is chosen based on an a-priori decision-maker (DM) function defined as

$$DM = f_1(s(t)) + \left[(1 + f_2(s(t))) \sqrt{|f_1(s(t)) - f_1(s(t-1))|} \right] \quad (9)$$

In the second formulation, two objectives one each corresponding to energy features - real power (P) and reactive power (Q) are formulated as

$$\begin{cases} f_1(s(t)) = |y(t) - \hat{y}(t)| y = \text{real power} \\ f_2(s(t)) = |x(t) - \hat{x}(t)| x = \text{reactive power}. \end{cases} \quad (10)$$

The decision-maker (DM) function employed in this formulation is given by

$$DM = \frac{1}{2} \sum_{i=1} f_i(s(t)). \quad (11)$$

The work in [29] follows the same principle and extends the number of features considering current (I), reactive power (Q), real power (P), apparent power (S), and harmonic (H).

From the review, it is evident that different formulations including objective(s) and/or constraints are proposed as they play a very important role in the performance improvement of optimization-based ED algorithms. In addition, it can be observed that as the number of devices and the number of states per device (the size of the database) increases, the algorithmic performance degrades. In other words, the similarity between the power states of devices in the network, the possible representation of a high power state by a combination of two or more low power states, and inherent noise affect the performance of the algorithm. Therefore, effective formulations combined with effective algorithms are needed to solve the ED problem. In other words, due to the nature of the ED problem, the possibility of addressing ED problem as a single or multi objective formulations with or without constraints which can be addressed as a binary optimization problem is opened

up. Moreover, performing ED for a large number of devices results into a large scale optimization problem. Hence, this provides basis for the evolutionary research community to develop algorithms which can efficiently be applied to solve the several possible formulations of the ED problems [30]. Progressively, [31] combined the use of least square error as well as temporal sparsity to realize a multi objective framework.

Table 1. Operational modes with corresponding power ratings and power deviation for appliances in the network [9].

No. of Appliances	Appliance	Maximum No of Modes	Power Rating (p)			Power Deviation (Θ)		
n		l_i	p_i^1	p_i^2	p_i^3	Θ_i^1	Θ_i^2	Θ_i^3
D1	LCD-Dell	1	25	-	-	5	-	-
D2	LCD-LG	1	22	-	-	5	-	-
D3	Coffee Make	3	700	900	1100	100	100	100
D4	iMac	2	35	50	-	5	10	0
D5	Desktop	2	40	50	-	15	20	-
D6	Server	1	130	-	-	20	-	-
D7	Water Cooler	3	65	380	450	5	10	10
D8	Laptop	3	15	30	70	5	10	10
D9	Microwave	3	1000	1200	1700	100	100	100
D10	Printer	3	400	700	900	50	80	100
D11	Refrigerator	2	115	350	-	15	10	-

3. Issues with Existing Datasets for Optimization-Based ED

Taking into account the different formulations shown in Section 2, for proper benchmarking of optimization-based ED algorithms, the dataset should contain

1. A measured aggregate signal that is inherently noisy due to measurement error.
2. Ground truth power consumption information corresponding to each device in the network.
3. The information regarding the number of operational modes and their rated power for each appliance.

For ED, in general, numerous datasets [32–46] were collected. The characteristics datasets are summarised in [11,35] based on the information present in them such as features, sampling rate, number of days, etc. From the summary, it is evident that

1. None of the datasets [32–46] provide information related to the number of modes and their ratings for corresponding devices, which is essential for optimization-based ED. However, the information regarding modes and their rating as shown in Table 1, can be approximated if a significant amount of data related to the device operation is available [8].
2. In datasets such as [32–34], the aggregate signal is not provided. In works such as [25], the aggregate is constructed by adding up the power consumption of individual appliances. However, the employment of constructed aggregate signal may not continue the noisy characteristics that are inherent to ED.

3. In datasets such as [42–46], the aggregate, and individual device power consumption are measured but the measurement is not synchronized. Due to the lack of synchronization between the aggregate and appliance level measurement information, the performance evaluation which is a crucial step in benchmarking becomes difficult.
4. In datasets such as [35–41], for some of the devices, the appliance-specific ground truth regarding power consumption is not measured or provided. The unavailability of the appliance-specific consumption information makes it difficult to—(a) approximate the modes and ratings corresponding to devices and (b) evaluate the performance of algorithms.

In literature, UVIC [9] is the only dataset that contains the information required for benchmarking optimization-based ED algorithms. UVIC [9] dataset contains synchronized measurements corresponding to aggregate and appliance-level consumption for seven consecutive days. In addition, the number of modes and their corresponding power ratings with power deviation in each mode are approximated as shown in Table 1.

UVIC dataset contains the required information for benchmarking optimization-based ED algorithms. However, in the dataset, the information regarding the seven different days is presented. However, the performance characteristics of the algorithms that a particular aggregate signal is expected to test are not highlighted. In other words, a good benchmark dataset should contain several different instances where each instance should represent one of the issues that prop-up due to the varying device usage characteristics

4. Benchmark Dataset Design

ED optimization problem is separable along the time axis. Therefore, to evaluate optimization-based ED algorithms, it is not necessary to have instances with aggregate signals that are too long such as a day when sampled at every 10 s. In other words, instances with shorter time periods such as an hour-long would suffice. However, the instances should be diverse in terms of the challenges they pose to the optimization algorithms. Therefore, using the information present in the UVIC dataset [9], we developed multiple (18) instances of 1-h, sampled every 10 s. The instances were developed considering the following criteria, where each criterion corresponds to one or more issues related to the optimization-based ED problem highlighted in Section 2.

- C1: Most devices are in operation (**Issues 1, 2, 3, 4, 5**)
- C2: Only continuous devices are in operation (**Issues 1, 2**)
- C3: Only low power devices operation (**Issues 1, 3, 5**)
- C4: Only high power devices operation (**Issues 3, 4, 5**)
- C5: Continuously operating devices are switched off (**Issues 1, 5**)
- C6: Power rating of one device is a linear combination of one or more devices (**Issues 1, 3**)
- C7: Devices with similarity in power ratings corresponding to states are in operation (**Issues 1, 3, 5**)
- C8: Devices where power deviation corresponding to the states of one device matches power rating of other devices are in operation (**Issues 1, 4, 5**)
- C9: Concurrent switching of devices with similarity in states or one as a linear combination of the others (**Issues 3, 5**)

Table 2 summaries (18) instances in terms of the criteria used to design them. In addition, we also present the amount of operation time corresponding to each device as a percentage of the whole time (360 time steps). The developed dataset which contains the aggregate signal, appliance-specific ground truth consumption, and information regarding modes of operation with corresponding power rating are provided in the <https://github.com/RammohanMallipeddi/NILM-bench-suite.git>, accessed on 5 August 2021.

Table 2. Data Instances and Different aspects of ED Problems.

Instances	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	Criteria
I_1	1	0.642	0	1	1	0.956	1	1	0	0.053	0.497	c1
I_2	1	0.428	0.139	1	1	0.992	1	1	0	0	0.619	c1
I_3	1	0	0.047	1	1	0.975	1	0.006	0.008	0	0.692	c2
I_4	0	0	0	1	1	1	1	0	0	0	0.5	c2
I_5	1	0.592	0	1	1	0.961	0.872	1	0	0.014	0.306	c3
I_6	1	0.425	0.136	1	1	0.992	1	1	0	0	0.536	c3
I_7	1	0	0.169	1	1	0.997	1	0	0.022	0.047	1	c4
I_8	1	0	0.044	1	1	1	1	1	0.114	0	0.564	c4
I_9	0	0	0	1	1	1	0.386	0.003	0	0	0.378	c5
I_{10}	1	0	0.014	1	1	1	0.331	1	0	0.042	0.461	c5
I_{11}	0.494	0	0	1	1	1	1	1	0	0.064	0.519	c6
I_{12}	0.919	0	0	1	1	1	0.583	0.319	0	0.003	0.9	c6
I_{13}	1	0	0.031	1	1	1	0.528	0.275	0.072	0	0.544	c7
I_{14}	0	0	0	1	1	1	1	0.008	0	0	0.508	c7
I_{15}	0	0	0	1	1	0.983	0.992	1	0	0	0.336	c8
I_{16}	0	0	0	1	1	1	0.997	0.003	0	0.1	0.389	c8
I_{17}	0	0	0.044	0.997	1	0.992	1	0.003	0	0.031	0.578	c9
I_{18}	1	0.086	0	1	1	0.967	0.583	1	0	0.017	0.517	c9

5. Performance Metrics for Optimization-Based ED

In literature, inconsistency in measuring and evaluating the performance of ED algorithms is evident [8,25,27]. To empirically evaluate and fairly compare the performance of different ED algorithms, it is essential to define and employ standard performance metrics.

Performance metrics should be able to evaluate the overall performance of the algorithm as well as each component in the algorithm. For instance, in event-based approaches, the common components are—event detection, event classification, and energy estimation. Depending on the distinct components of each algorithms, there are measures to quantitatively assess each algorithmic component as well as the overall performance. Hence, the measures can be categorized as— event classification metrics, event detection metrics, energy estimation metrics, as well as overall metrics. A comprehensive summary of the various performance metrics is presented in [11]. However, optimization-based event-less approaches do not depend on event detection and classification. On the contrary, attempt to disaggregate the total load in different time slices. Therefore, metrics such as F-measure that are generally employed to measure the classification accuracy of algorithms will not be appropriate. Consequently, event-less approaches only require metrics to evaluate the final energy estimation. In this section, we summarize the different metrics with respect to which optimization-based event-less ED algorithms can be compared.

In literature, it has been agreed that the performance evaluation of ED algorithms from the different perspectives cannot be done effectively with a single metric. In other words, a set of different metrics each evaluating the different performance aspects is used. The metrics are grouped into appliance-specific performance indicators and overall performance indicators.

The evaluation metrics used are outlined as follows, where $y_i(t)$ is the actual power consumption of the i -th device at t , and $\hat{y}_i(t)$ is the estimate provided by the ED algorithm.

5.1. Appliance-Specific Performance Indicators

1. Per-appliance accuracy (AC_i) [47] of appliance i -th is given by

$$AC_i = 1 - \frac{\sum_{t=1}^T |y_i(t) - \hat{y}_i(t)|}{2 \sum_{t=1}^T |y_i(t)|}. \quad (12)$$

Per-appliance accuracy gives a measure of the ability of the algorithm in estimating the device level power consumption in the entire time horizon.

2. Estimated Energy Fraction Index (EEFI) ($\hat{h}_i$) is the ratio between the estimated energy corresponding to the i -th appliance and the recovered aggregated signal given by

$$\hat{h}_i = \frac{\sum_{t=1}^T \hat{y}_i(t)}{\sum_{i=1}^N \sum_{t=1}^T \hat{y}_i(t)}. \quad (13)$$

To analyze the performance of an algorithm, EEFI needs to be compared with the Actual Energy Fraction Index (AEFI) (h_i), which indicates the portion of the actual energy consumption by the i -th appliance with respect to the measured aggregated signal. Here, h_i can be defined as

$$h_i = \frac{\sum_{t=1}^T y_i(t)}{\sum_{i=1}^N \sum_{t=1}^T y_i(t)}. \quad (14)$$

For every device i in the network, the closeness of $\hat{h}_i$ to h_i is an indication of the algorithm's effectiveness.

3. Relative Squared Error (RSE_i) is a normalized metric that measures the error between the measured and the estimated power consumption for each appliance. It indicates the ability of the algorithms in estimating each device power consumption profile over time relative to the actual consumption. RSE of i -th device expressed as

$$RSE_i = \frac{\sum_{t=1}^T (y_i(t) - \hat{y}_i(t))^2}{\sum_{t=1}^T y_i^2(t)}. \quad (15)$$

An effective ED algorithm is expected to provide higher per-appliance accuracies combined with lower RSE values. Furthermore, the EEFI ($\hat{h}_i$) should be as close as possible to the AEFI (h_i).

However, due to the discrete approximation of the power ratings corresponding to the operating modes

- (a) AC_i corresponding to appliances operating at high power states are expected to be lower
- (b) RSE_i corresponding to appliances operating at high power states is expected to be higher.

The effect gets magnified (decrease of AC_i and increase of RSE_i) as the devices operate at high power states for longer. However, AC_i and RSE_i can be employed to compare the performance of ED algorithms, because an algorithm performing better is expected to have larger AC_i and lower RSE_i for devices operating at high power modes.

5.2. Overall Performance Metrics

1. Overall Accuracy (ACC) [47] indicates the effectiveness of the algorithm in estimating the aggregated consumption over the whole time interval and is given by

$$ACC = 1 - \frac{\sum_{t=1}^T \sum_{i=1}^n |y_i(t) - \hat{y}_i(t)|}{2 \sum_{t=1}^T \sum_{i=1}^n |y_i(t)|}. \quad (16)$$

2. Overall State Prediction Accuracy (SPA) is given by

$$\text{SPA} = 1 - \frac{\sum_{n=1}^N \|S^* - \hat{S}\|_1}{N.T}, \quad (17)$$

where S^* is the estimated state matrix, $\hat{S}$ is the actual state matrix, n is the number of devices and T is the total time interval (360 samples) over which the disaggregation is performed.

3. Fraction of Total Energy Assigned Correctly (FTEAC) is defined as the overlap between the two indices referred to as EEFI and AEFI by each appliance over all the appliances in the network. Mathematically, it can be defined as

$$\text{FTEAC} = \sum_{i=1}^n \min\{h_i, \hat{h}_i\}. \quad (18)$$

The largest possible value of FTEAC is one when the fraction of power consumption corresponding to each device in the measured and estimated aggregate signals perfectly match with one another. When the power consumption of some devices in the network is underestimated/overestimated then the value of FTEAC decreases.

ACC indicates the capability of the algorithm in allocating the right power values to corresponding appliances at each time instance, while SPA measures the ability of the algorithm in estimating the states of appliances [9]. From the definition of the metrics, it can be observed that an algorithm with a higher SPA can have a low ACC if the operation corresponding to low duration, the high power device is not properly estimated. In optimization ED algorithms, ACC of 100% does not conclude the superiority of the algorithm due to the discrete values employed to represent the operational states of appliances. As a result, even though an algorithm can accurately predict the operation characteristics of the device over the entire time interval, the error between the estimated and measured aggregate signals is expected. Furthermore, the error is proportional to the power rating of the operational states at any given time. This can be attributed to the high power variation associated with high power modes (Table 1). Therefore, a single measure cannot completely indicate that one algorithm is better than the other. Hence, for better analysis of ED algorithms, it is essential to compare the algorithms on a set of metrics instead of a single metric.

6. Simulation Results and Analysis

To provide the baseline results, the existing optimization-based ED algorithms are simulated on all 18 instances and the results are summarized with respect to the performance metrics described. The complete results for all the 18 instances are present in the Supplementary Materials. The state-of-the-art algorithms are analyzed using 4 (I_1 , I_2 , I_{10} and I_{12}) of the 18 instances developed. The results on the 4 instances are presented in Tables 3–6. The best values corresponding to each indicator, except ACC, are highlighted for better visualization. As mentioned earlier, due to the discrete value representation of the power states, the value of energy lost and the optimal value of ACC cannot be estimated. The codes for IP, ALIP, and MONILM are obtained from the authors of the original publications, while SSER and Sopt are reproduced with the help of the information presented in the respective publications. For all the algorithms, the parameters are fine-tuned to the best of our knowledge. However, the same set of parameters corresponding to an algorithm are employed for all the instances in the dataset. All the simulations were performed in MATLAB 2020a with 64-bit Windows 10, 3.30GHz. CPU and 24GB RAM. For better analysis, we categorize the 5 state-of-the-art algorithms as—(a) algorithms based on only least square error (as shown in Equation (3)) such as IP and ALIP, and (b) algorithms based on both least square error (as shown in Equation (3)) and temporal sparsity (as shown in Equation (5)).

From Tables 3–6, it is evident that ALIP outperforms IP on instances I_1 , I_2 and I_{12} in terms of per-appliance and overall indicators. The Superior performance of ALIP can be associated with the additional constraints (inequality constraint to address **Issue 1** and equality constraint to address **Issue 2**) enforced. However, in instance I_{10} (Table 5), the performance of IP is better than ALIP, evident from overall indicators (SPA and FTEAC). Because, in I_{10} were continuously operating devices are OFF, the use of equality constraint forces algorithm to estimate the aggregate signal by switching ON the continuously operating devices. This results in the overestimation of power consumption related to continuously operating devices such as Water-cooler (D7) and Refrigerator (D11), as indicated by the corresponding $\hat{h}_i$ values in Table 5.

From Tables 3 and 4, it is evident that the performance of SSER is not always but consistently better than Sopt and MONILM. This is due to the significantly large number of parameters, in MONILM and Sopt compared to SSER, which are sensitive to the type and number of devices in operation. In addition, in SSER, the total variation corresponding to the whole time period (T) is minimized, while in MONILM and Sopt only the variation in the states between the current and previous time instances are considered. Furthermore, the performance of MONILM also depends on the DM function employed.

In instance I_1 , where most of the devices are operating, D3 (Coffeemaker) and D9 (Microwave) do not operate at all, while the operation of D10 (Printer) is significantly low. In Addition, most of the power is consumed by continuously operating devices (see h_i values). Therefore, enforcing temporal sparsity is expected to result in an elevated performance which is evident from the results in Table 5 where SSER outperforms ALIP. However, in instance I_2 which is similar to I_1 , but D10 (Printer) has a significant presence in terms of operation. In this case, the performance of SSER decreases drastically, because enforcing temporal sparsity results in the continuous operation of some devices (say D9 and D10) while continuously switching OFF (D3 and D11). This is expected because of similarity in the power ratings of states corresponding to the devices (D3, D9, D10 and D11).

In instance I_{10} , where all the continuously operating devices are switched OFF, the overall performance of IP seems better than all the other state-of-the-art algorithms. Because, as most of the continuously operating devices are not operating, enforcing sparsity does not help which is evident from the performance of SSER, Sopt, and MONILM. In instance I_{12} , where the power rating corresponding to a device is a linear combination of other devices, the performance of ALIP is better than all the other algorithms because ALIP is the only algorithm that has a provision to handle **Issue 3**.

From the results, it is clear that the proposed dataset is posing a challenge to the different frameworks in literature. Motivated by the results, where none of the frameworks were able to significantly outperform the others on all instances, the current dataset can be used for benchmarking future optimization-based ED frameworks. In addition, the set of metrics summarized helps analyze the performance of the optimization-based ED algorithms from multiple perspectives.

Table 3. Baseline results for instance I_1 .

No of Appliances		AC_i				h_i		$\hat{h}_i$				RSE_i				
N	LP	ALIP	MONILM	SSER	S_{OPT}	Actual energy	IP	ALIP	MONILM	SSER	S_{OPT}	IP	ALIP	MONILM	SSER	S_{OPT}
1	0.6967	0.6648	0.8722	0.5111	0.8403	0.05	0.0201	0.0169	0.0381	0.0011	0.0348	0.6066	0.6704	0.2556	0.9778	0.3194
2	0.6790	0.6162	0.8410	0.9171	0.9400	0.0328	0.0216	0.0144	0.0359	0.0308	0.0450	0.8638	0.9208	0.5983	0.1561	0.4469
3	1.00	1.00	1.0000	1.0000	1.0000	0	0.0953	0.0221	0.0131	0.0040	0.0233	-	-	-	-	-
4	0.6871	0.6612	0.8093	0.8780	0.8073	0.0728	0.0488	0.0387	0.0714	0.0878	0.1023	0.5391	0.6140	0.2651	0.1003	0.1778
5	0.7026	0.6663	0.8039	0.8499	0.8226	0.0734	0.0476	0.0403	0.0842	0.0941	0.1023	0.5073	0.6019	0.2493	0.1452	0.2055
6	0.7374	0.9358	0.8605	0.9345	0.5012	0.2267	0.1527	0.2663	0.2113	0.2653	0.0007	0.5128	0.0764	0.2175	0.0795	0.9972
7	0.6888	0.7923	0.7633	0.9411	0.9194	0.3025	0.2022	0.2307	0.1895	0.2964	0.2860	0.7700	0.3687	0.4554	0.0769	0.1109
8	0.4338	0.5763	0.4203	0.6961	0.4363	0.0489	0.0723	0.0398	0.0950	0.0578	0.0898	1.8482	1.0479	2.1133	0.8115	1.9037
9	1.0	1.00	1.0000	1.0000	1.0000	0	0	0.0057	0	0	0	-	-	-	-	-
10	0.6362	0.6593	0.7096	0.8178	0.6640	0.0759	0.1288	0.0550	0.0307	0.0466	0.0358	1.6428	0.8208	0.4959	0.2537	0.6601
11	0.45520	0.9656	0.8919	0.9375	0.9767	0.1158	0.2105	0.2729	0.2320	0.1178	0.2420	2.659	2.2818	2.5971	0.2854	1.2418
Overall metrics																
							IP	ALIP			MONILM			SSER		S_{OPT}
Overall Energy Disaggregation Accuracy (ACC (%))							98.7396	99.8051			99.6126			99.673		96.5757
State Prediction Accuracy (SPA (%))							51.3131	60.0758			49.8990			73.2576		42.2475
Fraction of Total Energy assigned correctly (FTEAC)							0.7337	0.7785			0.7769			0.9125		0.7011

Table 4. Baseline results for instance I_2 .

No of Appliances		AC_i				$\mathbf{h}$		$\hat{\mathbf{h}}_i$				RSE_i				
N	LP	ALIP	MONILM	SSER	S_{OPT}	Actual energy	IP	ALIP	MONILM	SSER	S_{OPT}	IP	ALIP	MONILM	SSER	S_{OPT}
1	0.7486	0.7413	0.9042	0.6611	0.7736	0.0408	0.0203	0.0198	0.0330	0.0353	0.0223	0.5028	0.5139	0.1917	0.6778	0.4528
2	0.7171	0.7514	0.9171	0.9314	0.5200	0.0175	0.0188	0.0167	0.0320	0.0095	0.0011	1.0768	0.8341	0.9456	1.0544	0.9753
3	0.7748	0.7469	0.7005	0.5189	0.7894	0.1988	0.1447	0.1025	0.0803	0.0693	0.1180	0.5052	0.4501	0.5745	0.9621	0.3815
4	0.6928	0.6683	0.7959	0.8466	0.8973	0.0532	0.0445	0.0414	0.0715	0.0733	0.0636	0.5018	0.5552	0.2234	0.1488	0.0877
5	0.6944	0.7108	0.8423	0.8592	0.8792	0.0581	0.0426	0.0448	0.0761	0.2070	0.0649	0.5194	0.4838	0.1589	0.1247	0.0791
6	0.6825	0.9412	0.8909	0.9324	0.5048	0.1886	0.0867	0.2123	0.1870	0.1263	0.0024	0.5973	0.0256	0.1268	0.0382	0.9892
7	0.6904	0.7776	0.8347	0.6163	0.7840	0.2364	0.2264	0.1856	0.1681	0.0430	0.1119	0.9751	0.4198	0.3149	0.7144	0.4031
8	0.7386	0.7257	0.7194	0.7161	0.7831	0.0913	0.0595	0.0536	0.0632	0	0.0697	0.3915	0.4083	0.4084	0.3522	0.2387
9	1	1	1.0000	1.0000	1.0000	0	0.0236	0.0254	0.0463	0.2931	0	-	-	-	-	-
10	1	1	1.0000	1.0000	1.0000	0	0.1697	0.0880	0.0671	0.1307	0.0776	-	-	-	-	-
11	0.4404	0.9324	0.9762	0.9255	0.0167	0.1152	0.1633	0.2102	0.1754	0.0132	0.4497	2.2178	1.2060	1.2553	0.2540	6.0574
Overall metrics																
							IP		ALIP		MONILM		SSER		S_{OPT}	
Overall Energy Disaggregation Accuracy (ACC (%))							98.65		99.94		99.7671		99.6369		97.4875	
State Prediction Accuracy (SPA (%))							42.7778		54.5960		55.7828		57.0960		53.4848	
Fraction of Total Energy assigned correctly (FTEAC)							0.7575		0.7683		0.7817		0.4079		0.5520	

Table 5. Baseline results for instance I_{10} .

No of Appliances		AC_i				h_i		$\hat{h}_i$				RSE_i				
N	LP	ALIP	MONILM	SSER	S_{OPT}	Actual energy	IP	ALIP	MONILM	SSER	S_{OPT}	IP	ALIP	MONILM	SSER	S_{OPT}
1	0.7514	0.6139	0.9444	0.9917	0.8986	0.0649	0.0326	0.0148	0.0577	0.0638	0.0518	0.4972	0.7722	0.1111	0.0167	0.2028
2	1	1	1.0000	1	1.0000	0	0.0286	0.0144	0.0408	0.0467	0.0565	-	-	-	-	-
3	0.7101	0.5908	0.5901	0.6897	0.8215	0.0298	0.0685	0.0051	0.0152	0.0332	0.0253	2.1854	0.7787	1.0658	1.1639	0.3955
4	0.7205	0.5924	0.7482	0.7871	0.9540	0.0845	0.0814	0.0230	0.0925	0.1204	0.0922	0.4300	0.7885	0.3561	0.2220	0.0157
5	0.7748	0.5653	0.8213	0.8610	0.5000	0.0956	0.0884	0.0160	0.1073	0.1164	0	0.3297	0.8459	0.2264	0.1209	1.0000
6	0.8224	0.9489	0.6649	0.5358	0.9474	0.3072	0.2438	0.3376	0.1219	0.0272	0.3367	0.2896	0.0115	0.6381	0.9209	0.0147
7	0.7675	0.7761	0.9315	0.6452	0.8128	0.1119	0.1382	0.2345	0.2386	0.1331	0.2322	0.8793	1.2719	0.7940	0.9713	0.9104
8	0.6514	0.5705	0.7362	0.7017	0.6729	0.1014	0.0807	0.0276	0.1141	0.0679	0.1359	0.7105	0.8171	0.5271	0.4285	0.7876
9	1	1	1.0000	1	1.0000	0	0.0072	0	0	0	0	-	-	-	-	-
10	0.5418	0.5895	0.5870	0.6424	0.7452	0.0668	0.0447	0.0281	0.0267	0.0859	0.0534	1.3111	0.9346	1.0802	1.2394	0.6870
11	0.6299	0.9777	0.8687	0.9838	0.5000	0.1380	0.1860	0.3292	0.1894	0.3063	0.0008	1.7664	2.0289	1.1164	1.4049	1.0060
Overall metrics																
							IP		ALIP		MONILM		SSER		S_{OPT}	
Overall Energy Disaggregation Accuracy (ACC (%))							98.7602		99.0772		99.3612		99.5509		96.9979	
State Prediction Accuracy (SPA (%))							57.9545		49.9545		49.7727		54.0909		53.0303	
Fraction of Total Energy assigned correctly (FTEAC)							0.8514		0.6716		0.8013		0.6855		0.7362	

Table 6. Baseline results for instance I_{12} .

No of Appliances		AC_i				h		$\hat{h}_i$				RSE_i				
N	LP	ALIP	MONILM	SSER	S_{OPT}	Actual energy	IP	ALIP	MONILM	SSER	S_{OPT}	IP	ALIP	MONILM	SSER	S_{OPT}
1	0.7296	0.6949	0.9018	0.6556	0.5000	0.0501	0.0254	0.0197	0.0434	0.0169	0	0.5891	0.6133	0.2598	0.7160	1.0000
2	1	1	1.0000	1.0000	1.0000	0	0.0234	0.0160	0.0356	0.0373	0.0290	-	-	-	-	-
3	1	1	1.0000	1.0000	1.0000	0	0.0212	0.0042	0	0	0	-	-	-	-	-
4	0.7126	0.6332	0.8008	0.6536	0.8330	0.0706	0.0606	0.0364	0.0861	0.0438	0.0942	0.4666	0.6694	0.2413	0.6142	0.1654
5	0.7088	0.6292	0.8163	0.8598	0.8243	0.0797	0.0600	0.0347	0.0933	0.0995	0.1087	0.4861	0.6795	0.2202	0.1334	0.1911
6	0.7545	0.9465	0.8697	0.8932	0.9127	0.2567	0.1629	0.2833	0.2337	0.2487	0.2613	0.4358	0.0126	0.1836	0.1314	0.0878
7	0.7282	0.8121	0.7788	0.9060	0.5000	0.2779	0.2278	0.2571	0.2289	0.2391	0	0.6390	0.3326	0.3994	0.1525	1.0000
8	0.6065	0.5574	0.6015	0.6717	0.5270	0.0204	0.0664	0.0416	0.0849	0.0682	0.0887	5.3982	2.5110	6.8094	3.3893	6.1865
9	1	1	1.0000	1.0000	1.0000	0	0	0	0.0061	0.0073	0.0061	-	-	-	-	-
10	0.7000	0.5000	0.5000	0.5000	0.5000	0.0061	0.1277	0.0218	0.0460	0.0145	0	11.3700	2.4400	4.0400	1.9600	1.0000
11	0.5938	0.8915	0.6184	0.9382	0.7128	0.2387	0.2246	0.2961	0.1436	0.2249	0.3900	1.1594	0.5494	0.9286	0.1165	1.2979
Overall metrics																
							IP		ALIP		MONILM		SSER		S_{OPT}	
Overall Energy Disaggregation Accuracy (ACC (%))							98.7795		99.6766		99.5509		99.6588		97.2710	
State Prediction Accuracy (SPA (%))							59.8485		65.8081		47.0202		59.0404		54.2929	
Fraction of Total Energy assigned correctly (FTEAC)							0.7878		0.8697		0.8561		0.8796		0.6660	

7. Conclusions and Future Work

In this work, we summarized the different formulations related to optimization-based ED and highlighted the related issues. The characteristics of the dataset required for optimization-based ED are highlighted and the drawbacks of the available datasets in view of the requirements are summarized. Based on different criteria, a dataset with a diverse set of 18 instances is created. To evaluate the performance of the optimization-based ED frameworks, a set of diverse metrics from the literature are summarized. Finally, the performance of the state-of-the-art optimization-based ED algorithms is compared on the developed instances with respect to the set of evaluation metrics considered. From the results, it is evident that the set of evaluation metrics are capable of evaluating the performance to the optimization-based ED algorithms from different aspects. In addition, the instances are capable of posing a variety of challenges to the existing optimization-based ED frameworks, which is evident from the simulation results. Therefore, the proposed dataset combined with a summarized set of evaluation metrics is expected to help better benchmarking optimization-based ED algorithms.

In the current work, the dataset developed contains information related to a single energy feature (i.e., real power). In the future, developing a dataset by incorporating other energy features such as reactive power, current, voltage, etc., would help develop diverse optimization-based ED frameworks such as multi-objective.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/en15051600/s1>.

Author Contributions: Conceptualization, O.S.A. and R.M.; methodology, O.S.A. and R.M.; formal analysis, O.S.A., A.K., R.M., S.D. and P.N.S.; data curation, O.S.A. and R.M.; writing—original draft preparation, O.S.A., A.K. and R.M.; writing—review and editing, O.S.A., A.K., R.M., S.D. and P.N.S.; supervision, R.M.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used for this work can be obtained at <https://github.com/RammohanMallipeddi/NILM-bench-suite.git>, accessed on 5 August 2021.

Acknowledgments: This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R111A3049810).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zoha, A.; Gluhak, A.; Imran, M.A.; Rajasegarar, S. Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey. *Sensors* **2012**, *12*, 16838–16866. [\[CrossRef\]](#)
2. Zeifman, M.; Roth, K. Nonintrusive appliance load monitoring: Review and outlook. *IEEE Trans. Consum. Electron.* **2011**, *57*, 76–84. [\[CrossRef\]](#)
3. Pamulapati, T.; Mallipeddi, R.; Lee, M. Multi-objective home appliance scheduling with implicit and interactive user satisfaction modelling. *Appl. Energy* **2020**, *267*, 114690. [\[CrossRef\]](#)
4. Cao, Y.; Mohammadzadeh, A.; Tavooosi, J.; Mobayen, S.; Safdar, R.; Fekih, A. A new predictive energy management system: Deep learned type-2 fuzzy system based on singular value decomposition. *Energy Rep.* **2022**, *8*, 722–734. [\[CrossRef\]](#)
5. Tavooosi, J. Intelligent Model Predictive Control for Boiler Temperature. *Autom. Control Comput. Sci.* **2021**, *55*, 16–25. [\[CrossRef\]](#)
6. D’Adamo, I.; Falcone, P.M.; Martin, M.; Rosa, P. A Sustainable Revolution: Let’s Go Sustainable to Get Our Globe Cleaner. *Sustainability* **2020**, *12*, 4387. [\[CrossRef\]](#)
7. Falcone, P.M.; Imbert, E.; Sica, E.; Morone, P. Towards a bioenergy transition in Italy? Exploring regional stakeholder perspectives towards the Gela and Porto Marghera biorefineries. *Energy Res. Soc. Sci.* **2021**, *80*, 102238. <https://doi.org/10.1016/j.erss.2021.102238>.
8. Tang, G.; Wu, K.; Lei, J.; Tang, J. Plug and play! A simple, universal model for energy disaggregation. *arXiv* **2014**, arXiv:1404.1884.

9. Tang, G.; Wu, K.; Lei, J.; Tang, J. A simple model-driven approach to energy disaggregation. In Proceedings of the 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), Venice, Italy, 3–6 November 2014; pp. 566–571. [\[CrossRef\]](#)
10. Laughman, C.; Lee, K.; Cox, R.; Shaw, S.; Leeb, S.; Norford, L.; Armstrong, P. Power signature analysis. *IEEE Power Energy Mag.* **2003**, *1*, 56–63. [\[CrossRef\]](#)
11. Pereira, L.; Nunes, N. Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—A review. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1265.
12. Gonçalves, H.; Ocneanu, A.; Bergés, M. Unsupervised disaggregation of appliances using aggregated consumption data. In *The 1st KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*; ACM: San Diego, CA, USA, 2011.
13. Johnson, M.J.; Willsky, A.S. Bayesian Nonparametric Hidden Semi-Markov Models. *J. Mach. Learn. Res.* **2013**, *14*, 673–701.
14. Winkler, P.; Le Ray, G.; Pinson, P. Unsupervised Energy Disaggregation: From Sparse Signal Approximation to Community Detection. *IEEE Trans. Smart Grid* **2019**, 1–8. Available online: <http://pierrepinson.com/docs/Lerayetal2019-unsupnilm.pdf> (accessed on 19 June 2021). [\[CrossRef\]](#)
15. Hart, G. Nonintrusive appliance load monitoring. *Proc. IEEE* **1992**, *80*, 1870–1891. doi:10.1109/5.192069.
16. Parson, O.; Ghosh, S.; Weal, M.; Rogers, A. Non-Intrusive Load Monitoring Using Prior Models of General Appliance Types. *Proc. AAAI Conf. Artif. Intell.* **2012**, *26*, 356–362.
17. Kolter, J.Z.; Batra, S.; Ng, A.Y. Energy Disaggregation via Discriminative Sparse Coding. In Proceedings of the 23rd International Conference on Neural Information Processing Systems, NIPS’10, Vancouver, BC, Canada, 6–9 December 2010; Curran Associates Inc.: New York, NY, USA, 2010; Volume 1, pp. 1153–1161. [\[CrossRef\]](#)
18. Rahimpour, A.; Qi, H.; Fugate, D.; Kuruganti, T. Non-Intrusive Energy Disaggregation Using Non-Negative Matrix Factorization With Sum-to-k Constraint. *IEEE Trans. Power Syst.* **2017**, *32*, 4430–4441. [\[CrossRef\]](#)
19. Srinivasan, D.; Ng, W.; Liew, A. Neural-network-based signature recognition for harmonic source identification. *IEEE Trans. Power Deliv.* **2006**, *21*, 398–405. [\[CrossRef\]](#)
20. Tsai, M.S.; Lin, Y.H. Modern development of an Adaptive Non-Intrusive Appliance Load Monitoring system in electricity energy conservation. *Appl. Energy* **2012**, *96*, 55–73. [\[CrossRef\]](#)
21. Ghorbanpour, S.; Pamulapati, T.; Mallipeddi, R. Swarm and evolutionary algorithms for energy disaggregation: challenges and prospects. *Int. J. Bio Inspired Comput.* **2021**, *17*, 215–226.
22. Egarter, D.; Elmenreich, W. EvoNILM: Evolutionary appliance detection for miscellaneous household appliances. In Proceedings of the 15th annual conference companion on Genetic and evolutionary computation, GECCO ’13 Companion, Amsterdam, The Netherlands, 6–10 July 2013.
23. Suzuki, K.; Inagaki, S.; Suzuki, T.; Nakamura, H.; Ito, K. Nonintrusive appliance load monitoring based on integer programming. In Proceedings of the 2008 SICE Annual Conference, Tokyo, Japan, 20–22 August 2008; pp. 2742–2747. [\[CrossRef\]](#)
24. Egarter, D.; Sobe, A.; Elmenreich, W. Evolving Non-Intrusive Load Monitoring. In Proceedings of the 16th European Conference on the Applications of Evolutionary Computation, Vienna, Austria, 3–5 April 2013. Springer: Berlin/Heidelberg, Germany, 2013. [\[CrossRef\]](#)
25. Bhotto, M.Z.A.; Makonin, S.; Bajić, I.V. Load Disaggregation Based on Aided Linear Integer Programming. *IEEE Trans. Circuits Syst. II Express Briefs* **2017**, *64*, 792–796. [\[CrossRef\]](#)
26. Shen, Q.; Wang, X. An analysis of the optimization disaggregation algorithm in the estimation related to energy consumption of appliances in buildings. *Appl. Math. Comput.* **2014**, *234*, 506–519. [\[CrossRef\]](#)
27. Piga, D.; Cominola, A.; Giuliani, M.; Castelletti, A.; Rizzoli, A.E. Sparse Optimization for Automated Energy End Use Disaggregation. *IEEE Trans. Control Syst. Technol.* **2016**, *24*, 1044–1051. [\[CrossRef\]](#)
28. Machlev, R.; Belikov, J.; Beck, Y.; Levron, Y. MO-NILM: A multi-objective evolutionary algorithm for NILM classification. *Energy Build.* **2019**, *199*, 134–144. [\[CrossRef\]](#)
29. Li, L.; Yang, L.; Chen, H.; Li, M.; Zhang, C. Multi-objective evolutionary algorithms applied to non-intrusive load monitoring. *Electr. Power Syst. Res.* **2019**, *177*, 105961. [\[CrossRef\]](#)
30. Osaba, E.; Villar-Rodriguez, E.; Del Ser, J.; Nebro, A.J.; Molina, D.; LaTorre, A.; Suganthan, P.N.; Coello Coello, C.A.; Herrera, F. A Tutorial On the design, experimentation and application of metaheuristic algorithms to real-World optimization problems. *Swarm Evol. Comput.* **2021**, *64*, 100888.
31. Ghorbanpour, S.; Pamulapati, T.; Mallipeddi, R.; Lee, M. Energy disaggregation considering least square error and temporal sparsity: A multi-objective evolutionary approach. *Swarm Evol. Comput.* **2021**, *64*, 100909.
32. Reinhardt, A.; Baumann, P.; Burgstahler, D.; Hollick, M.; Chonov, H.; Werner, M.; Steinmetz, R. On the accuracy of appliance identification based on distributed load metering data. In Proceedings of the 2012 Sustainable Internet and ICT for Sustainability (SustainIT), Pisa, Italy, 4–5 October 2012; pp. 1–9.
33. Gao, J.; Giri, S.; Kara, E.; Bergés, M. PLAID: A public dataset of high-resolution electrical appliance measurements for load identification research: Demo abstract. In Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, Memphis, TN, USA, 3–6 November 2014. [\[CrossRef\]](#)
34. Picon, T.; Meziane, M.N.; Ravier, P.; Lamarque, G.; Novello, C.; Bunetel, J.L.; Raingeaud, Y. COOLL: Controlled On/Off Loads Library, a Public Dataset of High-Sampled Electrical Signals for Appliance Identification. *arXiv* **2016**, arXiv:1611.05803. [\[CrossRef\]](#)

35. Shin, C.; Lee, E.; Han, J.; Yim, J.; Rhee, W.; Lee, H. The ENERTALK dataset, 15 Hz electricity consumption data from 22 houses in Korea. *Sci. Data* **2019**, *6*, 193.
36. Kelly, J.; Knottenbelt, W. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Sci. Data* **2015**, *2*, 150007.
37. Uttama Nambi, S.N.A.; Lua, A.R.; Prasad, R.V. LocED: Location-aware Energy Disaggregation Framework. In Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, Seoul, Korea, 4–5 November 2015. [\[CrossRef\]](#)
38. Beckel, C.; Kleiminger, W.; Cicchetti, R.; Staake, T.; Santini, S. The ECO data set and the performance of non-intrusive load monitoring algorithms. In Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, Memphis, TN, USA, 3–6 November 2014.
39. Monacchi, A.; Egarter, D.; Elmenreich, W.; D’Alessandro, S.; Tonello, A.M. GREEND: An energy consumption dataset of households in Italy and Austria. In Proceedings of the 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm), Venice, Italy, 3–6 November 2014; pp. 511–516. [\[CrossRef\]](#)
40. Anderson, K.; Ocleanu, A.; Carlson, D.R.; Rowe, A.G.; Bergés, M. In *BLUED: A Fully Labeled Public Dataset for Event-Based Non-Intrusive Load Monitoring Research*; ACM: New York, NY, USA, 2012.
41. Rashid, H.; Singh, P.; Singh, A. I-BLEND, a campus-scale commercial and residential buildings electrical energy dataset. *Sci. Data* **2019**, *6*, 190015. [\[CrossRef\]](#)
42. Kolter, J.Z.; Johnson, M.J. REDD: A Public Data Set for Energy Disaggregation Research. *Artif. Intell.* **2011**, *25*, 1–6. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Parson, O.; Fisher, G.; Hersey, A.; Batra, N.; Kelly, J.; Singh, A.; Knottenbelt, W.; Rogers, A. Dataport and NILMTK: A building data set designed for non-intrusive load monitoring. In Proceedings of the 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL, USA, 14–16 December 2015; pp. 210–214. doi:10.1109/GlobalSIP.2015.7418187.
44. Murray, D.; Stanković, L.; Stanković, V. An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. *Sci. Data* **2017**, *4*, 160122.
45. Batra, N.; Singh, A.; Singh, P.; Dutta, H.; Sarangan, V.; Srivastava, M. Data Driven Energy Efficiency in Buildings. *arXiv* **2014**, arXiv:1404.7227. [\[CrossRef\]](#)
46. Mishra, A.; Cecchet, E.; Shenoy, P.; Albrecht, J.R. Smart *: An Open Data Set and Tools for Enabling Research in Sustainable Homes. In Proceedings of the ACM SustKDD’12, Beijing, China, 12 August 2012; Volume 111, p. 108.
47. Makonin, S.; Popowich, F. Nonintrusive load monitoring (NILM) performance evaluation. *Energy Effic.* **2015**, *8*, 809–814.