A MultiScale Framework for Nonintrusive Load Identification

Daisy H. Green, Steven R. Shaw, Peter Lindahl, Thomas J. Kane, John S. Donnal, and Steven B. Leeb

***Abstract*—Nonintrusive load monitoring, i.e., the process of identifying individual load information from aggregate electrical measurements, is useful for a variety of smart grid applications including energy scorekeeping, condition monitoring, and activity tracking. Numerous load disaggregation algorithms have been used for nonintrusive monitoring. Many of these perform well only on certain datasets or load types, because transient electrical events can occur on vastly different time-scales and operating schedules with significantly different regularities. This paper presents a nonintrusive load monitoring framework that allows multiple algorithms to be used across multiple time-scales, with their outputs combined to enhance load recognition. Results are demonstrated with power system data from a United States Coast Guard Cutter (USCGC), demonstrating the utility of the framework for developing applications for condition-based maintenance, among other applications.**

***Index Terms*—Condition-based maintenance (CBM), energy efficiency, fault detection, nonintrusive load monitoring.**

I. Introduction

RESEARCH on nonintrusive load monitoring tends to focus on disaggregation techniques aimed at incrementally improving various accuracy metrics, often a percentage of load operation or power consumption identified in an aggregate power stream [1], [2]. Fundamentally, nonintrusive monitoring is *ad hoc* and conjectural. A nonintrusive load monitor (NILM) will not be perfect under all conditions because load classification algorithms depend on the type and number of target loads in the data set, and the features associated with the load transients [3], [4]. Additionally, nonintrusive monitoring scenarios present uncertainties in load composition over time due to changing load mix, abnormal load behavior, or other disturbances. No single load disaggregation algorithm will best serve every application. There is also no "correct" way to assess accuracy that is significant for all applications [4].

The application matters. Different approaches for nonintrusive disaggregation allow a trade-off between computational complexity in monitoring, accuracy in determining necessary information for a given application, and flexibility in dealing with changing load compositions. For example, energy scorekeeping in residential homes differs substantially from system diagnostics in an industrial manufacturing center. Accurate identification of a subset of loads with a reasonable computational complexity may be more important than total consumption characterization [5], [6].

This paper presents a multiscale, multialgorithmic framework for organizing the signal processing for nonintrusive monitoring. Our goal in this paper is **not** to compare methods on a particular set or sets of power data. The utility of the framework presented here has been assessed with field demonstrations and results for various applications, including on US Coast Guard ships [7]–[9], at a school [10], and on a US Army microgrid [11]. The framework presented and reviewed in this paper represents the distilled effort of years of field testing of different signal processing approaches for nonintrusive load monitoring. The way a load consumes energy is a reflection of the physical task the load is performing. Thus, the framework appropriates the signal processing tools for finding these features in the observed data. The framework is an example and a guide that permits orderly application of computing resources for a nonintrusive load monitoring problem. It can be tailored to any specific application; it easily incorporates different disaggregation techniques while remaining focused on fundamental physical features of the energy consumption of each load. Do transients vary over time and duration? Are they randomly distributed or repeatable? Do they present harmonic content to the utility? These features and others form the foundation for nonintrusive load identification. New statistical analysis techniques are introduced here as part of this framework, to improve identification of loads which do not have a repeatable load transient.

The details and results of this framework are demonstrated on a 270 ft (82 m) United States Coast Guard Cutter (USCGC), SPENCER. In environments such as the microgrid of a ship,

Manuscript received April 1, 2019; revised May 13, 2019; accepted June 11, 2019. Date of publication June 17, 2019; date of current version January 14, 2020. This work was supported in part by the Office of Naval Research NEPTUNE Program, in part by the Grainger Foundation, in part by the MITei–Exxonmobil Collaboration, and in part by the Cooperative Agreement between the Masdar Institute of Science and Technology and Massachusetts Institute of Technology (02/MI/MIT/CP/11/07633/GEN/G/00). Paper no. TII-19-1234. *(Corresponding author: Daisy H. Green.)*

D. Green, P. Lindahl, T. Kane, and S. Leeb are with the Department of Electrical Engineering and Computer Science (EECS), Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: dhgreen@mit.edu; lindahl@mit.edu; tjkane@mit.edu; sbleeb@mit.edu).

S. Shaw is with the Department of Electrical and Computer Engineering, Montana State University, Bozeman, MT 59717 USA (e-mail: sshaw@montana.edu).

J. Donnal is with the Department of Weapons and Systems Engineering, U.S. Naval Academy, Annapolis, MD 21402 USA (e-mail: donnal@usna.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2019.2923236

signs of impending failure in mission-critical equipment are often visible in the electrical system weeks before an abrupt fault occurs [12]. Thus, information provided by a NILM system can provide invaluable feedback for aiding in optimizing operations and reducing equipment failures [7], [8]. Disaggregated load events are used in conjunction with the NILM dashboard [13] graphical platform, to deliver a timeline of load activity and diagnostic metrics to Coast Guard personnel for real-time monitoring and diagnostic focused analytics.

The rest of this paper is organized as follows. Following a literature review in Section II, Section III presents the signal processing framework for organizing nonintrusive event detection, organizing a full suite of deterministic and statistical tools. Section IV describes the use of the framework and presents results as applied to a case study on SPENCER. Section V concludes this paper.

II. Nonintrusive Load Monitoring

In an electrical system there are four main behaviors of loads: ON/OFF, *Finite State Machine (FSM)*, *continuously variable*, and *continuously ON* [3]. Both ON/OFF loads and *FSM* loads have sequences of changes of state that are clear step changes in power, and consume a constant amount of power at each state or operating mode. An ON/OFF load has only two states, ON or OFF, while a *FSM* load has several operating states due to its complex operation. The load transient acts as a load signature allowing NILM algorithms to identify the state of the loads. A *continuously ON* load consumes a constant power draw for days, or even weeks at a time. These loads can be identified if the initial ON event is observed or with the use of optimization techniques [14]. Finally, a *continuously variable* load has a variable power draw, and does not have a fixed number of states. This means there may not be repeatability in the power draw characteristics and it will be difficult to detect these loads using identification approaches that rely on steady state. One subset of *continuously variable* loads are those controlled by power electronics, including variable speed drives (VSDs) and dimmable lighting. The use of power electronics enables many loads to operate over a wide variable power range; thus there may not be a unique power consumption pattern. However, the power electronics contribute to significant higher order harmonics and a waveform based estimator can be used to estimate the real and reactive power consumed by variable power loads [15]. In contrast, another subset of *continuously variable* loads are not controlled by power electronics and do not have significant higher order harmonics, but have a power draw that appears practically stochastic. These loads may reach a steady state, but the time it takes to reach the steady state is highly variable and the load may not remain in steady state for the full operation. In order to correctly identify this subset of loads, referred to here on as "statistical loads," a statistical based method for load identification is presented in this paper as part of the identification framework.

A. Previously Reported Disaggregation Techniques

There have been many disaggregation techniques proposed in nonintrusive monitoring research. Disaggregation techniques can be broadly categorized into event based or nonevent based approaches, depending on whether load signatures are extracted from the power signals or not. Nonevent based techniques include optimization, such as linear programming, which attempts to minimize the error between an extracted feature vector and a database of known loads [14]. For event-based approaches, some commonly used load features include steady state step changes in power [16], harmonic frequency content [17], transient shapes [18], and voltage–current trajectories [19]. The load signatures that can be extracted from power data depend on the frequency of the sensors. Low frequency meters are lower cost, but typically limit features to steady state signatures. High frequency meters, in the range of kHz, allow for increased resolution.

Disaggregation techniques can also generally be divided into supervised or unsupervised techniques, based on whether or not a training process with labeled data is required. Supervised learning often uses pattern recognition to map detected events to a specific load. Some examples include artificial and deep neural networks (NN) [20], [21], and support vector machines [22]. In contrast, unsupervised approaches do not require labeled training data, such as independent component analysis [23]. Recently, approaches to combine classification techniques have been proposed, such as the use of committee decision mechanisms, a technique that incorporates both optimization and pattern recognition [24].

Most NILM research has been focused on absolute accuracy of total load disaggregation, but not deployment cost, usability, or concrete applications [25]. Our experience is that most facilities managers and operators are primarily interested in subsets of information that do not necessarily require flawless disaggregation and a maximum computation effort. The framework presented here distills successful methods we have deployed in the field for over a decade, and assists in guiding the organization of signal processing algorithms suitable for any particular nonintrusive monitoring application.

III. Multiscale, Multialgorithmic Framework

The multiscale, multialgorithmic framework organizes the application of useful feature extraction and signal processing for load identification, targeted for specific applications. It is structured around three foundations of nonintrusive load identification: data acquisition, scale, and variability of power consumption. First, time-series electrical signals are collected that give insight into the physics of the loads. Second, scales of transient events are examined, since transient events can occur on different time-scales. Finally, transients and steady-state power consumption can be variable, ranging from almost deterministic in their predictability to practically stochastic, which leads to differing approaches for load detection.

Fig. 1 outlines the framework for load identification, using example inputs and features. The main stages of the identification process are labeled, and include:

1) **Data Acquisition and Preprocessing** extracts the physical characteristics of energy consumption.
2) **Event Scale Separation** determines the scales for events.
3) **Event Detection** determines load "signatures" for detected events.

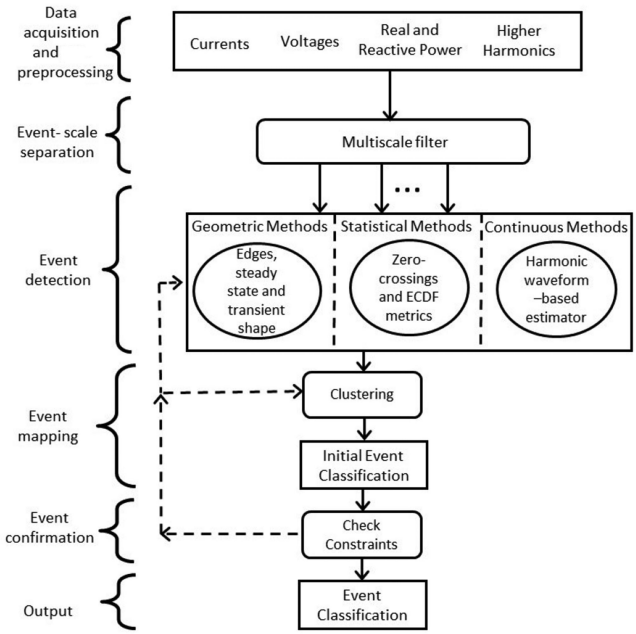


Fig. 1. Diagram for multiscale, multialgorithmic framework.

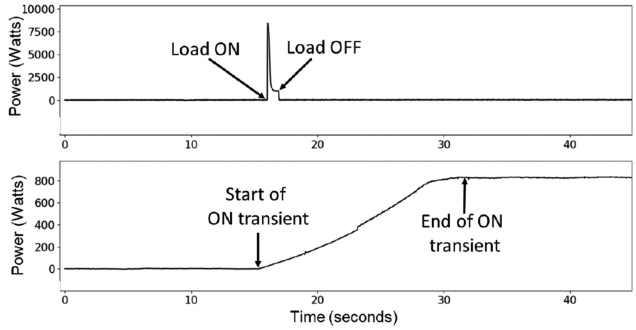


Fig. 2. Pump (top) and condenser fan (bottom) real power.

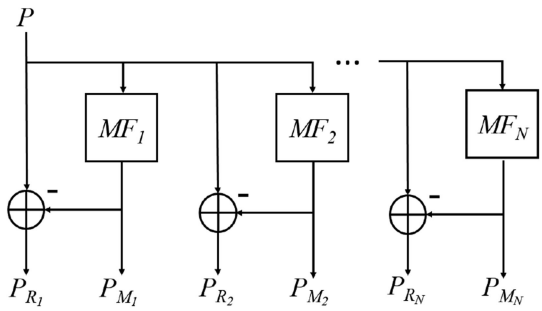


Fig. 3. Diagram for separating short-scale versus long-scale loads, for N different length median filters, MF. P_R is the residual stream and P_M is the median filtered stream.

4) **Event Mapping** matches events to loads.
5) **Event Confirmation** checks constraints between load events.

A. Data Acquisition and Preprocessing

Any data stream that captures the electrical nature of the loads and reflects how the load consumes energy can be used as input to the framework. Examples include voltages and currents, the the real (P) and reactive (Q) power, higher order harmonics, or even impedances. For our applications, the raw voltages and currents waveforms are typically sampled at 3 or 8 kHz. However, there are applications that may only need a low sampling rate (i.e., 120 Hz), or a higher sampling rate, in the order of MHz [3]. Any useful form can be an input to the framework, such as raw voltage and current waveforms, or the rms voltage and current. The voltage and current can be further processed into real (P) and reactive (Q) power, using the Sinefit algorithm [26]. P and Q correspond to the envelopes of in-phase and quadrature current drawn by the load relative to the voltage [17]. Sinefit effectively compresses the high-rate raw current and voltage data into real, reactive, and harmonic power components (third, fifth, and seventh order harmonics in our case) at a rate congruent with the power system line frequency (60 Hz). This promotes space-efficiency while maintaining the richness of the original signal. In [26], the accuracy of the Sinefit algorithm is demonstrated. It is also compared with other methods, such as the Kalman filter, for extracting envelopes of real, reactive, and harmonic content. The example presented in this paper uses P and Q as inputs.

B. Event Scale Separation

Following data preprocessing, a multiscale filter bank separates out loads that operate on different scales, e.g., in time. For instance, Fig. 2 shows an observation of real power (P) of a shipboard graywater pump and the turn-ON transient of a commercial condenser fan with a soft-start variable frequency drive. In this comparison example, the graywater pump is short-cycling, so it turns off a few seconds after turning on. It is necessary that the turn-ON and turn-OFF are properly detected in order to diagnose the fault condition. For comparison, the turn-ON transient of the condenser fan takes about 15 s while it spins up from a slow rpm to a fast rpm. The condenser fan has a single-phase variable speed motor drive, which ramps up when the set point changes in order to avoid power spikes. It would be difficult to detect both the pump event and the condenser fan event using an event detector on a single time-scale. It is very likely that the observation of the longer transient will be interrupted by other load events. To detect both events, a rolling median filter is used on the power stream. A median filter eliminates small fluctuations while preserving sharp edges [27]. Thus, a long median filter preserves the longer transient, but removes the events that occur at a smaller time-scale, such as the graywater pump. Subtracting the medianed stream (P_M) from the original data stream (P) results in the residual stream (P_R),

$$P_R = P - P_M. \tag{1}$$

Now the medianed stream contains the longer-scale events and the residual stream contains the smaller-scale events. Various length median filters can by employed if there are multiple time-scales present in a data stream. At each time-scale, there are median streams and residual streams. This decomposition is represented in Fig. 3.

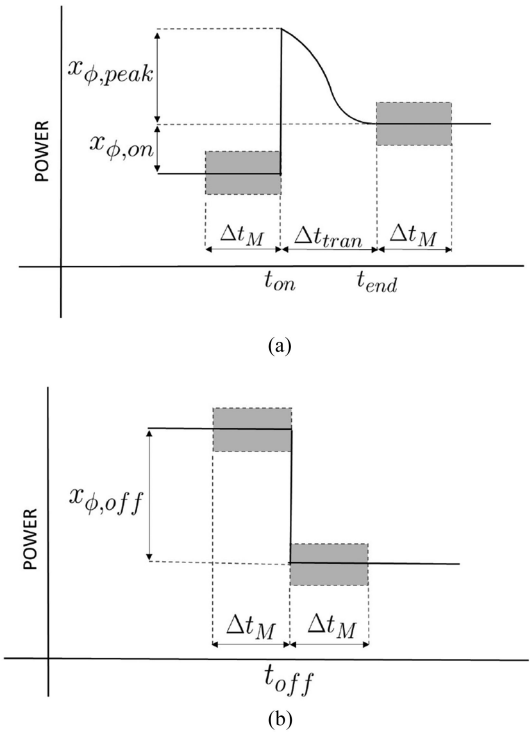


Fig. 4. Signature features extracted from a (a) turn-ON transient and a (b) turn-OFF transient.

C. Event Detection

1) Geometric Methods: Geometric features include some commonly used load characteristics, such as steady-state power levels and transient shapes for fundamental and higher order harmonics. An edge detector or change of mean detector [28] can detect load ON and OFF events. A median filter can be applied to the power streams prior to edge detection in order to preserve sharp edge events, while removing noise [9], [27]. The edge detector applied in this research uses the apparent power (S) stream. From the P and Q streams, apparent power (S) is calculated as, $S = \sqrt{P^2 + Q^2}$. Converting to apparent power simplifies load detection to a single data stream. This stream is convolved against the Laplacian of a Gaussian [29] kernel to compute the smoothed second derivative. This effectively maps step changes in apparent power to zero-crossings (ZC) for easier detection. An empirically-determined threshold is set to remove zero-crossings that are due to small variations of the resulting convolution. A zero-crossing detector is then used to find the location of the steps.

After an ON or OFF event is detected, the P and Q streams for each phase are examined to calculate a set of features. An ON event produces a change in steady state, and a transient which can be characterized by its peak, duration, and shape. The peak is due to the in-rush current as a load turns ON. Fig. 4(a) provides a reference diagram for these features for a conceptual turn-ON transient. The duration of the transient, or the time it takes for the load to reach steady state, is defined as

$$\Delta t_{\text{tran}} = t_{\text{end}} - t_{on} \tag{2}$$

where t_{on} is the time the load turns ON and t_{end} is the end of the start-up transient. Changes in steady-state real and reactive powers after a load turns ON are defined as the difference between the median values over Δt_M length windows, before and after the transient. The length of Δt_M is determined empirically based on the rate of event generation at the site to establish a reasonable steady-state time. These windows are shaded in Fig. 4. The changes in steady state are calculated as

$$\begin{aligned} x_{\phi,on} = {} & \texttt{median}(x_\phi[t_{\text{end}} < t \le t_{\text{end}} + \Delta t_M]) \\ & - \texttt{median}(x_\phi[t_{on} - \Delta t_M \le t < t_{on}]). \end{aligned} \tag{3}$$

Here, x can represent either the real or reactive power streams, and ϕ represents the phase (A,B,C). The transient real and reactive peak values are calculated as

$$\begin{aligned} x_{\phi,\text{peak}} = {} & \texttt{max}(x_\phi[t_{on} \le t < t_{\text{end}}]) \\ & - \texttt{median}(x_\phi[t_{\text{end}} < t \le t_{\text{end}} + \Delta t_M]). \end{aligned} \tag{4}$$

Because there is no transient peak when a load turns OFF, an OFF event is only characterized by its change in steady state. Fig. 4(b) provides a reference diagram for the features of a conceptual OFF event. Changes in steady-state real and reactive powers after a load turns OFF are defined as the difference between the median values over Δt_M length windows, before and after the OFF event. The steady-state changes in real and reactive power for an OFF event are calculated as,

$$\begin{aligned} x_{\phi,\text{off}} = {} & \texttt{median}(x_\phi[t_{\text{off}} < t \le t_{\text{off}} + \Delta t_M]) \\ & - \texttt{median}(x_\phi[t_{\text{off}} - \Delta t_M \le t < t_{\text{off}}]). \end{aligned} \tag{5}$$

In addition to using (3)–(5) to calculate the features illustrated in Fig. 4, the ON events and OFF events can be characterized using a correlation algorithm. The correlation algorithm matches the shape of the input data to known exemplars, and is fully described in [9], [30]. Consider two sampled waveforms f and g, where f is an observation or input signal and g is a load exemplar or example waveform. The correlation metric, M, is

$$M = \frac{(f - \bar{f}) \cdot (g - \bar{g})}{|g - \bar{g}|^2} \tag{6}$$

where $\bar{f}$ and $\bar{g}$ are the mean of f and g, respectively, and are subtracted from the original signals, f and g in order to remove the dc offsets. When M approaches one, this indicates that the exemplar and observation match in both shape and amplitude. Fig. 5 shows example observation data, an exemplar, and the resulting correlation metric as the exemplar window slides across the observation data.

2) Statistical Methods: For many, but not all loads, a high classification accuracy can be achieved by the previously described geometric features. However, statistical loads, which do not have consistent behavior during turn-ON or turn-OFF, are hard to identify with geometric methods. Unlike ON/OFF loads with clear step changes in steady state, the time it takes the load to reach steady state is highly variable and it may deviate from steady state during operation. For example, the ballast pump (BP) from the SPENCER often does not draw its rated power when the pump starts up or shuts down. This behavior is

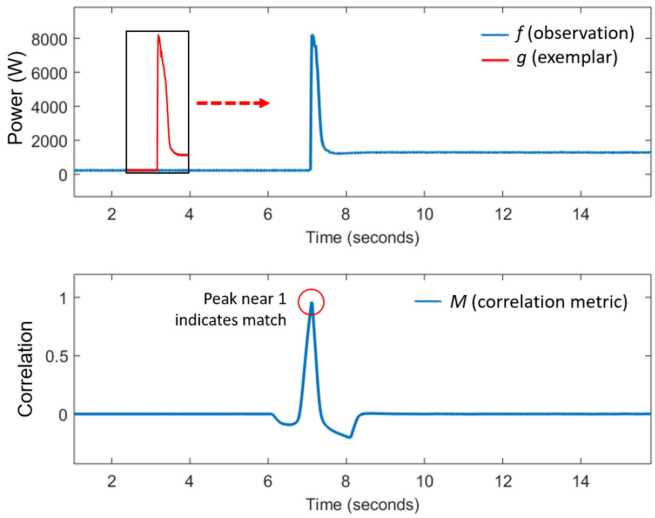


Fig. 5. Correlation metric for transient matching.

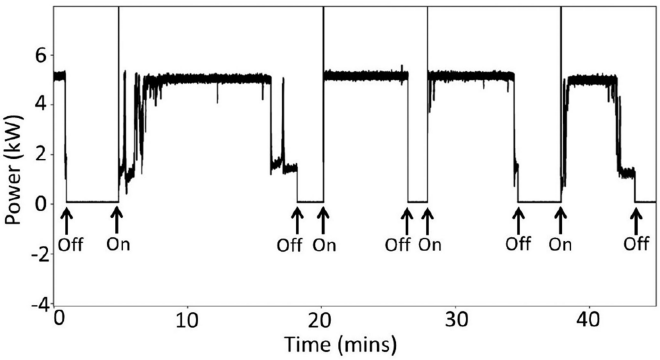


Fig. 6. Several runs of a ballast pump demonstrate inconsistent turn-ON and turn-OFF behavior.

depicted in Fig. 6. In order to correctly classify the BP, statistical methods are added to the load identification framework. Statistical loads are distinguished from loads with variable power due to power electronics, as described in the next section. If the variability is due to power electronics there will be substantial higher order harmonics due to the nonsinusoidal current waveforms. In contrast, statistical loads do not exhibit significant higher order harmonics, and the variability in power draw are consequences of the physical task the load is performing. Two statistical measurements are used for the analysis. The first is a zero-crossing metric, and the second uses the empirical cumulative distribution function (ECDF) to measure distribution. First a relatively wide window rolling median filter preserves the edges due to load ON or OFF events, but removes variations that occur on smaller time-scales. Since the smaller time-scale variations are of interest for the statistical analysis, the residual stream (P_R) is computed as the difference of the observed power minus the medianed filtered stream (1). Multiple size median filters can help distinguish activity that occurs on different time-scales (See Fig. 3). Fig. 7 shows P and P_R for two loads: the controllable pitch propeller pump (CPP) and BP.

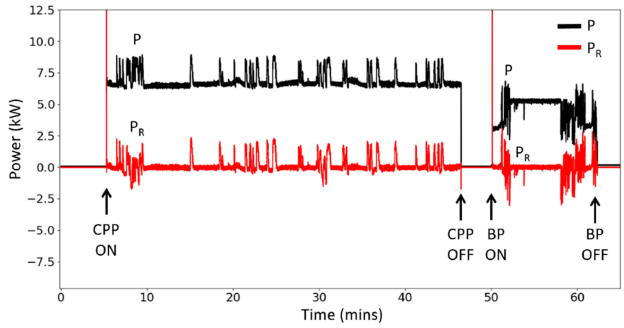


Fig. 7. Two loads that exhibit behavior in the residual stream at this time-scale.

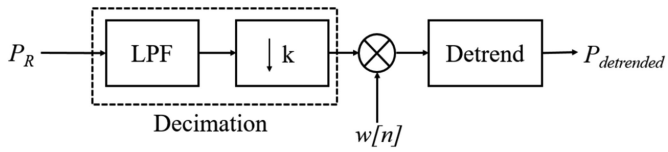


Fig. 8. Block diagram for preprocessing before ECDF and ZC analysis.

The residual stream is then processed as shown in Fig. 8. First, P_R is decimated and windowed. The windowed stream is detrended by subtracting out the mean, which allows for analysis on the variations around zero. The zero-crossing metric is the number of zero-crossings for each detrended waveform normalized by the length of the signal N, i.e.,

$$ZC = \frac{\#\text{ of zero-crossings}}{N}. \tag{7}$$

For the ECDF statistic, two nonparametric, distribution-free tests to measure the equality of one-dimensional probability distributions are considered: 1)The Kolmogorov–Smirnov (KS) test; and 2) The Cramer-von Mises (CvM) test [31]. Both tests compare the distance between the ECDF of a sample and reference ECDF distribution. An ECDF can be denoted as $\hat{F}_n(x) = \hat{P}_n(X \leq x)$, where n is the number of data samples, or length of the signal. This is given by

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x) \tag{8}$$

where I is the indicator function, given by

$$I(x_i \leq x) = \begin{cases} 1, & \text{if } x_i \leq x \\ 0, & \text{if } x_i > x \end{cases}. \tag{9}$$

For each data window, the ECDF is estimated by creating a histogram of the data values, and then applying a cumulative sum. As an example, the histogram and ECDF for the $P_{\text{detrended}}$ stream are shown two different loads in Fig. 9. The histogram and ECDF both show that $P_{\text{detrended}}$ for Load 2 has a larger spread of values then Load 1. The KS test statistic, D_n, represents the least upper bound (or maximum) of the point wise difference between the sample distribution function, $F_n(x)$ and the known

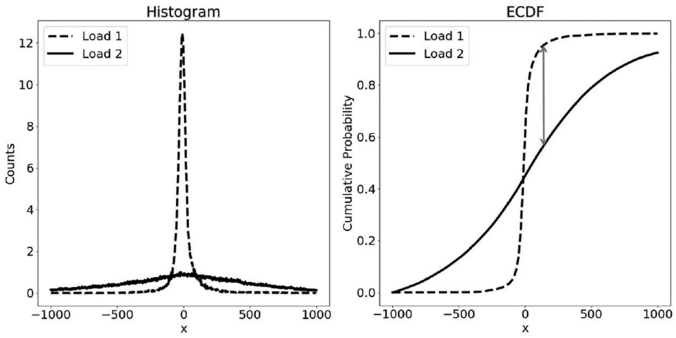


Fig. 9. Histogram and ECDF for two example loads, and the KS statistic.

exemplar distribution function, $F_0(x)$

$$D_n = \max_x \left| \hat{F}_n(x) - F_0(x) \right|. \tag{10}$$

Fig. 9 shows D_n with an arrow. Alternatively, the CvM criterion, w^2, uses the integration of the squared value of the difference between $F_n(x)$ and $F_0(x)$

$$w^2 = \int_{-\infty}^{\infty} (\hat{F}_n(x) - F_0(x))^2 dF_0(x). \tag{11}$$

The ZC and ECDF metrics are used together to identify loads such as the ballast pump, an example of which will be explained in Section IV.

3) Continuous Methods: Tracking the operation of continuously variable loads that demand ever-changing power arises for VSDs, light dimmers, and other loads controlled by power electronics. These loads can not be strictly identified by turn-ON and turn-OFF transients because they do not always have a repeatable power consumption pattern. However, the current waveforms of these loads consist of structural features that can be identified in both the time and frequency domains [15]. The power electronics contribute to nonsinusoidal current waveforms, and thus higher order harmonics for these loads. A waveform-based estimator described in [15] uses the fundamental and higher harmonic current waveforms to disaggregate the power consumption of variable power loads with harmonic signatures.

D. Event Mapping and Event Confirmation

The features for each event are then mapped to a specific load, using one of various pattern recognition approaches. This can be one of the many approaches described in Section II, including supervised or unsupervised techniques. After an initial mapping, constraints need to be checked to ensure a correct output. The first check is to ensure that two consecutive ON events or OFF events are not outputted for a given load. This would indicate that either an event was missed, an event was misclassified, or a nonevent was classified as an event. In the case of misclassification, the process goes back to the event mapping and reclassifies or determines that it is not an event. This is represented by the feedback arrows in Fig. 1. In the case of a missed event, a decision is made about which event to display. Depending on the application, the decision can be made to either reduce the possibility of incorrectly displaying that a load is energized or to reduce the possibility of incorrectly displaying that a load is secured. Constraints are also checked to be used for procedural oversight. The operation of *FSM* loads and interdependent loads are checked for standard operating procedures (SOP), ensuring that equipment or systems that go through multiple stages of operation are sequenced properly. In addition, if the number of loads and their expected power draw is known, optimization methods such as linear integer programming [14] can be integrated to aid in ensuring the correct output.

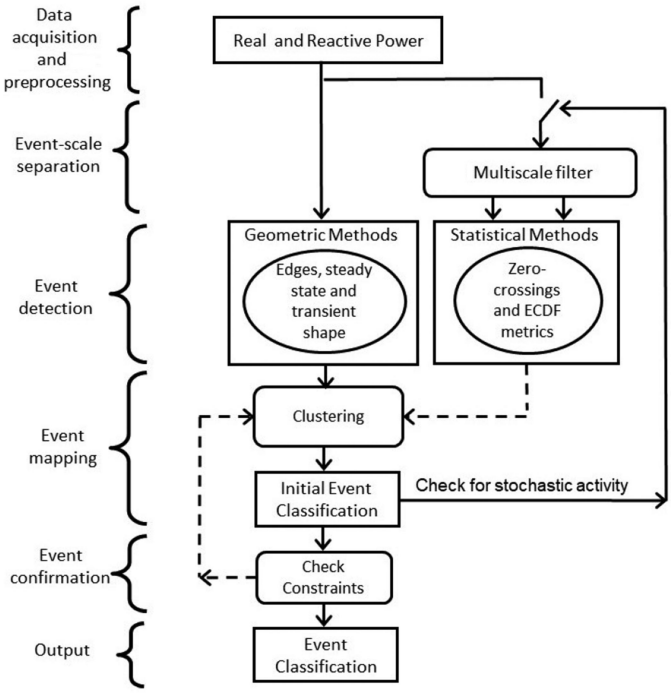


Fig. 10. Modified multiscale, multialgorithmic framework for shipboard application.

IV. Shipboard Case Study

Typically, "customers" for nonintrusive monitoring seek answers in one or more of three categories of information: energy scorekeeping, operator activity tracking, and inputs for condition-based maintenance (CBM). Based on the load mix at a site and the customer's monitoring needs, we use the framework as a guide to prepare specific NILM installations for different applications. An installation on the USCGC SPENCER serves as an example here. The SPENCER crew cares primarily about fault detection (CBM) and activity tracking. The specific needs of the SPENCER crew and the characteristics of the known set of loads lead to specific modifications of the overall framework to provide monitoring with a minimal computational burden. This modified framework is shown in Fig. 10.

For example, none of the loads in the SPENCER engineering space exhibit a slow spin-up compared to the other loads in the space. Thus, a single time-scale provides an adequate time

TABLE I
MONITORED LOADS IN ENGINE ROOM

Load	Power Rating	Delta Phases	PF	Port Panel	Stbd Panel
Main diesel engine (MDE) keep-warm system					
Lube oil (LO) heater	12 kW	3ϕ	1.0	x	x
Jacket water (JW) heater	9.0 kW	3ϕ	1.0	x	x
Prelube (PL) Pump	2.2 kW	3ϕ	0.82	x	x
Ship service diesel generator (SSDG) keep-warm system					
Jacket water (JW) heater	7.5 kW	3ϕ	1.0	x	x
Lube oil (LO) heater	1.3 kW	1ϕ	1.0	x	x
Diesel oil purification (DOP) system					
Separation chamber motor	9.5 kW	3ϕ	0.89	x	
Feed pump	2.6 kW	3ϕ	0.80	x	
Additional engine room loads					
Controllable pitch propeller hydraulic pump (CPP)	7.5 kW	3ϕ	0.82	x	x
Graywater pumps	3.7 kW	3ϕ	0.85	x	
Auxillary saltwater cooling pump (ASW)	7.5 kW	3ϕ	0.85	x	
Ballast pump (BP)	6.7 kW	3ϕ	0.90		x

series for geometric event detection, and the computational burden of computing multiple time-scale streams can be deferred to a later stage of the framework. The SPENCER contains no continuously variable loads from power electronics, thus avoiding the need for continuous load tracking. Two SPENCER loads of interest (CPP and BP) require statistical methods employed on a fine time-scale for complete identification. These two loads exhibit geometric initial transients. Hence, the median-filtered fine time-scale is only computed when necessary, after observing a startup transient that potentially indicates the need, as indicated by the switch in Fig. 10.

The ship has two NILM meters monitoring two electrical subpanels [PORT and STARBOARD (STBD)] in the engine room. The current and voltages are sampled at 8 kHz, and converted to 60 Hz P and Q using Sinefit as described in Section III. The 60 Hz P and Q are used as input to the load identification framework for this case study. Data is stored in a high-speed time-series database, NilmDB, allowing for high-speed and low-bandwidth access to the data [30]. All of the monitored engine room loads are listed in Table I. These monitored subpanels are crucial to the proper operation of ship propulsion, power generation, and auxiliary services. For classification, additional load classes are created for combinations of the main diesel engine (MDE) system loads that frequently actuate or secure together. For instance, the MDE lube oil heater, jacket water heater, and prelube pump often actuate together, creating a repeatable on-transient. Additionally, the diesel oil purifier (DOP) is a *FSM* load so it has a class for each distinct state.

A. Geometric Methods

For the geometric features, all events were detected on a single time-scale. First, 101-point median filters were used on the 60 Hz input P and Q streams, in order to remove small fluctuations. The size of the filter was determined empirically so that all the desired events would be preserved. A NN with two hidden layers was used for classification. The inputs to the NNs were the steady state, transient peak, and the correlation algorithm output for ON events and the steady state and correlation algorithm output for OFF events. For calculating features, Δt_M for (3)–(5) was chosen to be 0.5 s. For the correlation algorithm, each known load exemplar was matched to the input stream. As input to the NN, the metric used was $1 - |M|$ (see Fig. 5). Using this matching algorithm assumes that all loads are known, thus the length of the input vector depends on the number of load classes.

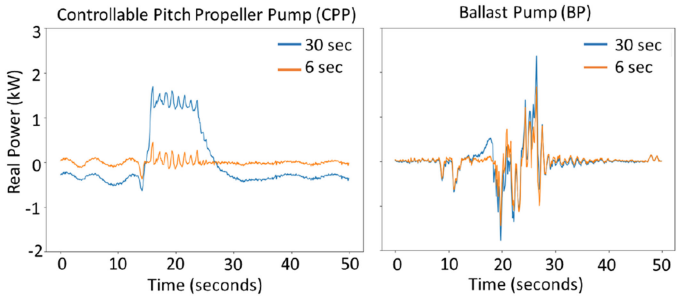


Fig. 11. $P_{\text{detrended},6}$ and $P_{\text{detrended},30}$ show activity on the CPP and the BP on different time-scales.

B. Statistical Methods

The distinctive transient peak when the BP turns ON permits using geometric methods alone to classify BP turn-ON events. However, due to the inconsistent turn-OFF behavior as shown in Fig. 6, it is difficult to correctly classify the BP turn-OFFs using geometric features. The use of statistical metrics aids in the correct identification of BP OFF events. However, since the STBD subpanel has two loads with significant activity in the residual stream, CPP and BP, the two must be distinguished from each other. To do this, $P_{\text{detrended}}$, as shown in Fig. 11 was calculated on streams separated by 6 and 30 s median filters, represented as $P_{\text{detrended},6}$ and $P_{\text{detrended},30}$, respectively. $Q_{\text{detrended},30}$ is also calculated. As indicated in Fig. 8, the power streams, P_R and Q_R, were decimated from a 60 Hz to 10 Hz frequency, windowed into intervals of 50 s, and detrended prior to calculating statistical metrics. As the window moves across the data stream,

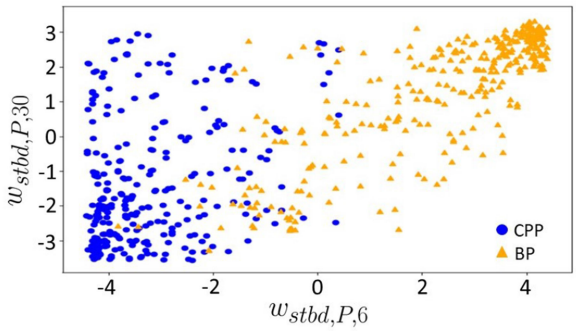


Fig. 12. CvM metrics ($w_{\text{stbd},P,6}$ and $w_{\text{stbd},P,30}$) for CPP and BP.

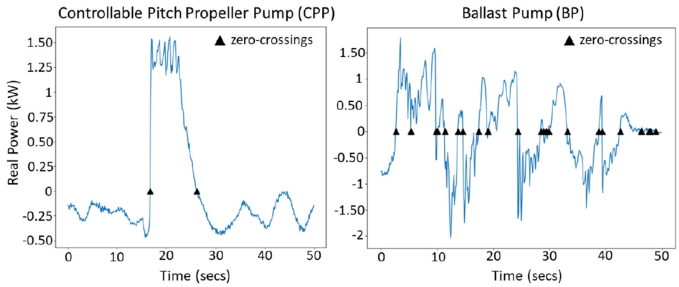


Fig. 13. Windows of $P_{\text{detrended}}$ for CPP and BP showing zero-crossings.

a 80% window overlap is used. The statistical features were only calculated for a window if the standard deviation of $P_{\text{detrended},30}$ was greater than a set threshold, because that indicates significant activity in the residual stream. CPP data is from October 2017 and BP data is from October 2016 to November 2017 (excluding January 20 to February 2017, which was part of the testing dataset). For CPP and BP, there were 329 and 319 windows that displayed significant activity in the residual stream, respectively.

For this application, the CvM criterion was chosen over the KS criterion because it considers the entire distribution within the specified range. Contrarily, the KS criterion ignores everything except the maximum difference, making it more sensitive in the center of the distribution. To represent F_0 from (11), average ECDF curves of past load actuations were used to create ECDF models for each of the loads. Thus, for the STBD subpanel, ECDF curves F_{cpp} and F_{bp} are generated. The integration was estimated by the Riemann left hand sum

$$w^2 = \sum_{i=0}^{D-1} (\hat{F_n}[x_i] - F_0[x_i])^2 \Delta x \tag{12}$$

where D is the number of points in the ECDF and Δx is the distance between x values. In this implementation, $D = 1000$, and x ranged from 2000 to 2000 for the 30 s time-scale and from 1000 to 1000 for the 6 s time-scale. The metric used for classification was the difference in w values for the two possible loads, given as, $w_{\text{stbd}} = w_{\text{cpp}} - w_{bp}$. Fig. 12 shows a scatter plot of the w_{stbd} metric for real power on the 30 and 6 s time-scales, showing a clear distinction between the CPP and BP metric values. These variations that occur at different time-scales are a result of the physical task the load is performing. The oscillations in power for the BP can likely be attributed to air pockets within the bilge and ballast pumping system. When pumping out bilges and ballast tanks, operators will try to get the tanks and bilges to the lowest level possible, causing the pump to take in a mixture of air and water at the inlet to the pump. When the pump is then secured and suction is shifted to a new tank, this air remains in the system and leads to a prolonged start sequence and rapid variations in power draw. Contrarily, the oscillations in power for the CPP are due to the manual operation of the pitch of the propeller blades as the operator is steering the ship. The CPP allows the operator to adjust the amount of thrust generated by a propulsor while maintaining a fixed rotational speed. These mechanical operations of the loads result in the BP having much more activity occurring at the shorter time-scale compared to the CPP.

For ZC, only the 30 s time-scale was considered, since the 6 s time-scale does not preserve the slow variations of the CPP. The average ZC metrics for CPP and BP were 0.013 and 0.066, respectively. Fig. 13 shows example windows of $P_{\text{detrended}}$ and detected zero-crossings for the CPP and BP. On average the BP has more zero-crossings than the CPP, due to the rapid power variations of the BP.

The CvM metrics and ZC metric were put in a feature vector to be used for classification

$$(w_{\text{stbd},P,30}\ w_{\text{stbd},Q,30}\ w_{\text{stbd},P,6}\ ZC_{P,30}). \tag{13}$$

For testing of the statistical methods for the STBD subpanel, new data points of the form of (13) were labeled by using a k-nearest neighbor (k-NN) classifier using $k = 3$. To save computation, the statistical metrics were only calculated after a BP ON event was detected. When the statistical metrics indicated BP activity, the algorithm looked for an OFF event that may have been misclassified as another load due to the nonrepeatable OFF transient and relabeled it as the BP OFF event. It was not the case in this case study, but if the two loads being compared had similar geometric metrics, the statistical metrics could be used for distinguishing both the ON events and OFF events of both loads.

C. Procedural Oversight

After disaggregating load events, constraints were checked to monitor the operation of *FSM* loads such as the DOP and of interdependent loads, such as the loads that make up the MDE system. The DOP has multiple states, but requires only the push of a button for a watchstander to operate. Thus a NILM can monitor the DOP for improper operation. Fig. 14 shows a depiction of the power stream of the DOP and its finite states. For the MDE system, normally when the MDE is started, two distinct events should occur: (1) the LO and JW heaters turning OFF once the "start" button is pressed; and (2) the prelube pump turning off 1–3 s later, once the engine has reached 150 RPM. The top and bottom plots in Fig. 15 show examples of proper and improper operation of the MDE system, respectively. The NILM can alert operators to deviation from SOP.

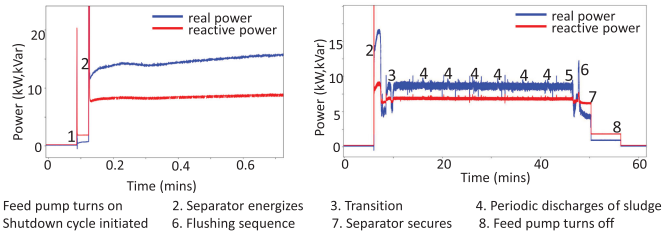


Fig. 14. DOP operating sequence.

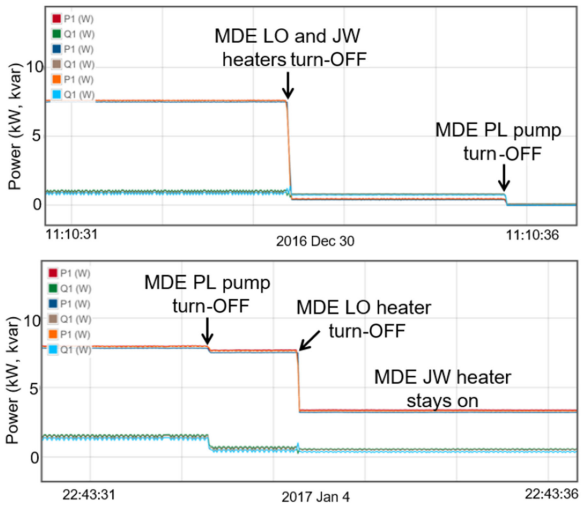


Fig. 15. Correct (top) and incorrect (bottom) operation of MDE system loads.

D. Results

The described framework was tested on busy intervals of data from SPENCER, using the Joule [32] data processing framework. Joule is a robust tool that models the data pipeline as a series of processing "modules" with "streams" of information passing between them. Data flows between modules without needing to access the database as an intermediary, allowing for efficient, real-time monitoring of load events. Different applications require different accuracy metrics. For example, an interface that gives users their appliance breakdown in real-time is significantly different than determining which appliances are using the most energy over a longer time-scale. The former values instantaneous, real-time accuracy, while the latter permits off-line analysis for greater accuracy [4]. For the given application, since the crew aboard SPENCER cares about the breakdown of loads being used at a given moment, it is crucial that individual loads are accurately detected in real-time. Thus, the accuracy for each class was evaluated by considering the following parameters [33]:

1) True Positive (TP): a load event occurred and was correctly identified.
2) False Positive (FP): a load event was classified, but that event did not occur.
3) False Negative (FN): a load event occurred but that event was not classified.

TABLE II
ACCURACY OF CLASSIFYING STBD AND PORT PANEL EVENTS

Load	TP		precision		recall	
	ON	OFF	ON	OFF	ON	OFF
Main diesel engine (MDE) keep-warm system - STBD						
JW Heater	44	44	1	1	1	1
LO Heater	44	44	1	0.978	1	0.978
Prelube Pump	40	40	1	0.976	0.930	0.909
Ship service diesel generator (SSDG) keep-warm system - STBD						
JW Heater	18	18	1	1	1	1
LO Heater	353	353	0.997	1	0.981	0.986
Diesel Oil Purification (DOP) System						
Separation Chamber Motor	55	55	0.965	0.965	0.965	0.965
Feed Pump	41	41	0.976	0.976	1	1
Additional engine room loads						
Graywater Pump	1645	1645	0.999	0.999	0.999	0.999
CPP Pump - STBD	74	74	1	1	1	1
Ballast Pump (only geometric)	22	22	1	1	0.629	0.629
Ballast Pump (geometric and statistical)	32	32	1	1	0.914	0.914

These parameters were used to determine the classifier's recall and precision, which answers two fundamental questions:

1) What is the likelihood that a load event is reported? ($recall$)
2) What is the likelihood that a reported event is correct? ($precision$)

$$recall = \frac{TP}{TP + FN}, \quad precision = \frac{TP}{TP + FP}. \tag{14}$$

The results of running the identification algorithm on one month of SPENCER data from January 20 to February 20, 2017 are presented in Table II. For training using the geometric features, data from was October 1 to 20, 2016 (at sea) and December 14 to 18, 2016 (in-port) were used to train a NN. For the statistical features, the models described in Section IV-B were used. To reduce the possibility of incorrectly displaying that a load was energized, if two consecutive ON events were detected, the first was ignored. Likewise, if two consecutive OFF events were detected, the second was ignored.

To accurately classify the majority of the loads, use of geometric methods was sufficient. The precision and recall values are close to or equal to one, which indicates near perfect performance in identifying a specific class. Because each load is performing a different task and is operated through different controls, there is a wide variation in the number of TP. For instance, the graywater pump is operated automatically by a controller that uses conductivity sensors to detect water levels and provide feedback for pump control. In contrast, the DOP system is run manually by an operator when diesel oil needs to be cleaned. As a result, under normal operation the graywater pump runs everyday up to a few times an hour, while the DOP does not run everyday and only up to a few times a day. Because the MDE system consists of interdependent loads, there are a similar number of actuations for each part of its system, including the lubeoil heater, jacketwater heater, and prelube pump. The few FN and FP across various loads are due to variations in the power data, load failures, and because loads sometimes actuate or secure

near each other. The only load that does not have good performance with geometric methods is the ballast pump. For the BP, geometric methods only accurately classified 63% of the OFF events, leading to the ON events being ignored. With the addition of statistical methods, the classification accuracy increased to 91%. Using geometric methods alone was unable to correctly classify the BP, because the BP often turned OFF while it was drawing less than the rated power. For some applications, such as energy scorekeeping, the addition of the statistical method would be unnecessary. However, for shipboard watchstanding and maintenance applications a single missed event can lead to a missed detection of failure in mission critical equipment. The addition of the statistical method greatly improves the deployability of NILM in this case. These results are an example of an application-centric use of NILM, tailoring both the identification algorithms and assessments of their accuracy for a specific application. This demonstration of the utility of the framework is one of our most recent applications taken from over a decade of field data testing the framework. Additional similarly successful results and applications include onboard US Coast Guard ships [7]–[9], at a public school [10], and on a US Army microgrid energized by diesel generators [11].

V. Conclusion

In this paper, the application of the load identification framework was demonstrated with power system data from the USCG cutter SPENCER. The results demonstrate the applicability of the framework for a variety of load types on the cutter, including ON/OFF, *FSM*, and statistical loads. The geometric methods were used for all of the ON/OFF loads and demonstrated near perfect performance in identifying loads, with precision and recall values close to one. Procedural oversight aided in quantifying the results of *FSM* and interdependent systems. Finally, the statistical methods and multiple time-scales were utilized for the statistical loads.

Nonintrusive power monitoring enables access to information that was previously unattainable. With flexible sensor solutions such as noncontact sensors [34] and graphical platforms such as the NILM dashboard [13], new doors are opened to apply nonintrusive monitoring for energy scorekeeping, activity tracking, fault detection, and CBM. Even with these new possibilities, nonintrusive monitoring will only be successful if signal processing approaches are tailored to focus on customer needs. For example, an energy scorekeeping application may not require 100% identification of all loads in the system, just the critical equipment. The framework presented in this paper is adaptable to any application. The framework's applicability has been demonstrated on various field demonstrations. By excluding or including algorithms and scales, the multiscale, multialgorithmic framework is a guide for allocating the tools to balance load disaggregation with computing effort for the application at hand.

Acknowledgment

The authors gratefully acknowledge the support and dedication of the US Coast Guard, and, in particular, the spectacular crew of USCGC SPENCER.

References

[1] S. M. Tabatabaei, S. Dick, and W. Xu, "Toward non-intrusive load monitoring via multi-label classification," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 26–40, Jan. 2017.

[2] K. Basu, V. Debusschere, S. Bacha, U. Maulik, and S. Bondyopadhyay, "Nonintrusive load monitoring: A temporal multilabel classification approach," *IEEE Trans. Ind. Inform.*, vol. 11, no. 1, pp. 262–270, Feb. 2015.

[3] A. Zoha, A. Gluhak, M. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, p. 1683816866, Dec. 2012.

[4] S. K. Barker, S. Kalra, D. Irwin, and P. D. Shenoy, "Nilm redux: The case for emphasizing applications over accuracy," in *Proc. 2nd Int. NILM Workshop*, Jun. 2014.

[5] E. T. Mayhorn, G. P. Sullivan, R. S. Butner, H. Hao, and M. C. Baechler, "Characteristics and performance of existing load disaggregation technologies," Pacific Northwest National Laboratory, Richland, WA, USA, Tech. Rep., Apr. 2015.

[6] D. He, W. Lin, N. Liu, R. G. Harley, and T. G. Habetler, "Incorporating non-intrusive load monitoring into building level demand response," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1870–1877, Dec. 2013.

[7] G. Bredariol, J. Donnal, W. Cotta, and S. Leeb, "Automatic watchstander through nilm monitoring," presented at ASNE Day 2016. American Society of Naval Engineers, 2016.

[8] G. Bredariol, D. Green, A. Aboulian, J. C. Nation, P. Lindahl, and S. B. Leeb, "NILM: A smarter tactical decision aid," presented at Technol., Syst. Ships Day, and American Society of Naval Engineers, 2017.

[9] J. Paris *et al.*, "Scalability of non-intrusive load monitoring for shipboard applications," presented at ASNE Day 2009. American Society of Naval Engineers, 2009.

[10] M. D. Gillman *et al.*, "Energy accountability using nonintrusive load monitoring," *IEEE Sensors J.*, vol. 14, no. 6, pp. 1923–1931, Jun. 2014.

[11] M. Gilman *et al.*, "Accounting for every kilowatt," Defense Acquisition Univ., Fort Belvoir, VA, USA, Tech. Rep., Oct. 2014.

[12] J. C. Nation *et al.*, "Nonintrusive monitoring for shipboard fault detection," presented at IEEE Sensors Appl. Symp., Glassboro, NJ, USA, Mar. 2017.

[13] A. Aboulian *et al.*, "Nilm dashboard: A power system monitor for electromechanical equipment diagnostics," *IEEE Trans. Ind. Inform.*, vol. 15, no. 3, pp. 1405–1414, Mar. 2019.

[14] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Improving nonintrusive load monitoring efficiency via a hybrid programing method," *IEEE Trans. Ind. Inform.*, vol. 12, no. 6, pp. 2148–2157, Dec. 2016.

[15] W. Wichakool, Z. Remscrim, U. A. Orji, and S. B. Leeb, "Smart metering of variable power loads," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 189–198, Jan. 2015.

[16] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

[17] C. Laughman *et al.*, "Power signature analysis," *IEEE Power Energy Mag.*, vol. 1, no. 2, pp. 56–63, Mar. 2003.

[18] S. B. Leeb and J. L. Kirtley, "A multiscale transient event detector for nonintrusive load monitoring," in *Proc. 19th Annu. Conf. IEEE Ind. Electron.*, Nov. 1993, pp. 354–359.

[19] T. Hasan, F. Javed, and N. Arshad, "An empirical investigation of V-I trajectory based load signatures for non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 870–878, Mar. 2014.

[20] H.-T. Yang, H.-H. Chang, and C.-L. Lin, "Design a neural network for features selection in non-intrusive monitoring of industrial electrical loads," presented at 11th Int. Conf. Computer Supported Cooperative Work in Des. Conf., Melbourne, VIC., Australia, Apr. 2007.

[21] J. Kelly and W. Knottenbelt, "Neural nilm: Deep neural networks applied to energy disaggregation," presented at ACM BuildSys 15, Seoul, South Korea, 2015.

[22] L. Du, Y. Yang, D. He, R. G. Harley, T. G. Habetler, and B. Lu, "Support vector machine based methods for non-intrusive identification of miscellaneous electric loads," presented at 38th Annu. Conf. IEEE Ind. Electron. Soc., Montreal, QC, Canada, Oct. 2012.

[23] A. A. Munshi and Y. A.-R. I. Mohamed, "Unsupervised nonintrusive extraction of electrical vehicle charging load patterns," *IEEE Trans. Ind. Inform.*, vol. 15, no. 1, pp. 266–279, Jan. 2019.

[24] J. Liang, S. Ng, G. Kendall, and J. Cheng, "Load signature studypart ii: Disaggregation framework, simulation, and applications," *IEEE Trans. Power Del.*, vol. 25, no. 2, pp. 561–569, Apr. 2010.

[25] M. Kahl, C. Goebel, A. U. Haq, T. Kriechbaumer, and H.-A. Jacobsen, "Nofare: A non-intrusive facility resource monitoring system," in *Proc. 4th D-A-CH Conf. Energy Informat.*, vol.9424, Karlsruhe, Germany, 2015, pp. 59–68.

[26] J. Paris, J. S. Donnal, Z. Remscrim, S. B. Leeb, and S. R. Shaw, "The sinefit spectral envelope preprocessor," *IEEE Sensors J.*, vol. 14, no. 12, pp. 4385–4394, Dec. 2014.
[27] W. Karl, S. Leeb, L. Jones, J. Kirtley, and G. Verghese, "Applications of rank-based median filters in power electronics," *IEEE Trans. Power Electron.*, vol. 7, no. 3, pp. 437–443, Jul. 1992.
[28] R. Jain, R. Kasturi, and B. G. Schunck, *Machine Vision*, New York, NY, USA: McGraw-Hill, Inc., 1995.
[29] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London B: Biol. Sci.*, vol. 207, no. 1167, 1980, pp. 187–217.
[30] J. Paris, J. S. Donnal, and S. B. Leeb, "NilmDB: The non-intrusive load monitor database," *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2459–2467, Sep. 2014.
[31] D. A. Darling, "The Kolmogorov-Smirnov, Cramer-von Mises tests," *Ann. Math. Statist.*, vol. 28, no. 4, pp. 823–838, 1957.
[32] J. Donnal, "Joule: A real time framework for decentralized sensor networks," *IEEE Int. Things J.*, vol. 5, no. 5, pp. 3615–3623, Oct. 2018.
[33] P. Klein, J. Merckle, D. Benyoucef, and T. Bier, "Test bench and quality measures for non-intrusive load monitoring algorithms," presented at 39th Annu. Conf. IEEE Ind. Electron. Soc., Vienna, Austria, Nov. 2013.
[34] J. S. Donnal and S. B. Leeb, "Noncontact power meter," *IEEE Sensors J.*, vol. 15, no. 2, pp. 1161–1169, Feb. 2015.

Daisy H. Green received the B.S. degree in electrical engineering from the University of Hawaii, Manoa, Hawaii, USA, in 2015 and the M.S. degree in electrical engineering in 2018 from the Massachusetts Institute of Technology, Cambridge, MA, USA, where she is currently working toward the Ph.D. degree in electrical engineering.

Steven R. Shaw received the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2000.

He is currently a Professor with the Department of Electrical and Computer Engineering, Montana State University, Bozeman, MT, USA. His current research interests include system identification and controls, energy, and fuel cell systems.

Peter A. Lindahl received the Ph.D. degree in engineering from Montana State University, Bozeman, MT, USA, in 2013.

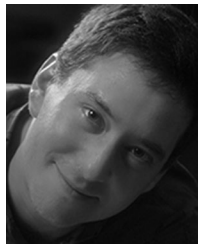
He was a Postdoctoral Associate in the Research Laboratory of Electronics at the Massachusetts Institute of Technology, Cambridge, MA, USA, and currently works at Exponent. His research interests include sensors and instrumentation for energy and power systems, renewable energy generation, and energy policy.

Thomas J. Kane received the M.S. degree in mechanical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2019.

He was previously stationed as Damage Control Assistant aboard USCGC MELLON and as a Port Engineer for the National Security cutter fleet. He is currently a Lieutenant with the U.S. Coast Guard. His next duty station will be onboard USCGC CAMPBELL as the Engineer Officer.

John S. Donnal received the B.S. degree from Princeton University, Princeton, NJ, USA, in 2007, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2013 and 2016, respectively, all in electrical engineering.

He is currently working as a Faculty Member with the U.S. Naval Academy in Weapons and Systems Engineering. His research interests include nonintrusive load monitoring synthesis, energy harvesting, and communications systems.



Steven B. Leeb received the doctoral degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1993.

Since 1993, he has been working as a member with the faculty of the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. He also holds a joint appointment in Massachusetts Institute of Technology's Department of Mechanical Engineering. He is concerned with the development of signal processing algorithms for energy and real-time control applications.