Shipboard Fault Detection Through Nonintrusive Load Monitoring: A Case Study

Peter A. Lindahl, Daisy H. Green, Gregory Bredariol, Andre Aboulian, John S. Donnal, and Steven B. Leeb

***Abstract*— As crew sizes aboard maritime vessels shrink in efforts to reduce operational costs, ship operators increasingly rely on advanced monitoring systems to ensure proper operation of shipboard equipment. The nonintrusive load monitor (NILM) is an inexpensive, robust, and easy to install system useful for this task. NILMs measure power data at centralized locations in ship electric grids and disaggregate power draws of individual electric loads. This data contains information related to the health of shipboard equipment. We present a NILM-based framework for performing fault detection and isolation, with a particular emphasis on systems employing closed-loop hysteresis control. Such controllers can mask component faults, eventually leading to damaging system failure. The NILM system uses a neural network for load disaggregation and calculates operational metrics related to machinery health. We demonstrate the framework's effectiveness using data collected from two NILMs installed aboard a U.S. Coast Guard cutter. The NILMs accurately disaggregate loads, and the diagnostic metrics provide easy distinction of several faults in the gray water disposal system. Early detection of such faults prevents costly wear and avoids catastrophic failures.**

***Index Terms*— Monitoring, sensors, power measurement, load modeling, fault detection, fault diagnosis, condition monitoring.**

I. Introduction

OVER the past several decades, the US Navy, US Coast Guard (USCG), and the commercial maritime industry have all reduced crew sizes on ships in an effort to decrease operational costs [1], [2]. As such, ships increasingly rely on advanced monitoring systems to ensure proper operation of mechanical and electrical systems [3], [4]. These systems range from individual machine sensors measuring motor or generator shaft torque, speed, and vibration [5]–[7], to bulk electrical system measurements for power system state estimation and fault detection [8], [9].

Bulk electrical power data can also provide information on individual loads through nonintrusive load monitoring (NILM) techniques. NILM is the process of disaggregating individual electric loads, e.g., a motor or heater, from electrical measurements taken at a centralized location. This sort of monitoring can work in a number of ways, most notably by either solving for a combination of loads that best fits the aggregate measured power (non-event based) or by detecting characteristic changes in the total electric power delivered (event based). Numerous optimization and machine learning approaches have been applied for this disaggregation task, and several review papers (e.g. [10]–[12]) detail the evolution of NILM systems and current approaches. Generally, the disaggregation algorithms are supervised learning techniques often using event based methods or unsupervised learning techniques often using non-event based methods.

Unsupervised approaches attempt to disaggregate bulk electrical measurements into individual load contributions without (or at least with minimal) labelled training data. This is desirable as collecting training data can be expensive, laborious, and impractical. Suzuki *et al.* [13] used an integer programming approach based on single cycles of current waveforms from each load to determine which loads are present in the aggregate current waveform at a given moment. Bhotto *et al.* [14] improved upon this approach by incorporating linear-programming methods and correcting algorithm outputs based on a state diagram. Other unsupervised approaches to the load disaggregation task include the use of Factorial Hidden Markov Models (FHMMs) [15], [16], and recently, Koutitas and Tassiulas proposed an unsupervised but event-based method for load disaggregation using a fuzzy logic approach that can perform well on low-frequency data typical of low-cost smart meters [17].

Supervised approaches typically outperform the unsupervised approaches in terms of accuracy, but come with the disadvantage of requiring labelled data. Still, the benefit of high accuracy can outweigh the data collection cost if the load monitoring is used to improve a facility's operations, e.g., monitoring for equipment or system faults. Event-based supervised learning approaches require the extraction of transient features from the power stream ([18] provides a fairly comprehensive list of proposed features). Low-frequency meters are typically limited to steady-state changes in current or power. As the frequency (and cost) of the meter increases into

Manuscript received August 2, 2018; accepted August 30, 2018. Date of publication September 6, 2018; date of current version October 10, 2018. This work was supported in part by The Grainger Foundation, in part by the MIT-KUST Program, in part by the Office of Naval Research NEPTUNE Program, and in part by the Cooperative Agreement between the Masdar Institute of Science and Technology, Abu Dhabi, United Arab Emirates and the Massachusetts Institute of Technology, Cambridge, MA, USA—Reference 02/MI/MIT/CP/11/07633/GEN/G/00. The associate editor coordinating the review of this paper and approving it for publication was Prof. Boby George. *(Corresponding author: Peter A. Lindahl.)*

P. A. Lindahl, D. H. Green, A. Aboulian, and S. B. Leeb are with the Research Laboratory of Electronics (RLE) and the Department of Electrical Engineering and Computer Science (EECS), Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: lindahl@mit.edu).

G. Bredariol is with the United States Coast Guard, Saint Petersburg, FL 33701 USA.

J. S. Donnal is with the U.S. Naval Academy, Annapolis, MD 21402-5000 USA.

Digital Object Identifier 10.1109/JSEN.2018.2869115

the kHz range however, additional features such as transient shape metrics [19], harmonic power components [20], [21], and voltage-current trajectories [22], [23] become available. Such additional features increase the information contained in each signature for use by the disaggregation algorithm. A multitude of supervised learning techniques have been applied for load disaggregation such as support vector machines (SVMs) [24], [25] and both shallow [26] and deep neural networks (NNs) [27]. Researchers have also investigated combinations of algorithms including a committee decision mechanism combining NNs with unsupervised optimization techniques [28].

In this paper, we present a full NILM framework featuring a two hidden-layer NN for load disaggregation, with the NN approach chosen due to its general flexibility and track-record of good performance for disaggregation [10]. The NILM system features high frequency (8 kHz) smart meter hardware to extract transient shape metrics from the real and reactive power streams for input into the NN. From the disaggregated load information, the NILM system calculates operation metrics correlated to system level faults of the ship's on-off controlled systems. These controllers can mask component faults which do not result in total system failure [29]. These faults, however, can increase energy consumption, place wear on machinery, and reduce system performance. Data collected from two NILMs installed on engine-room subpanels aboard the USCG Famous class cutter SPENCER showcase the NILM's ability to accurately disaggregate load information and reveal strong correlations between the calculated metrics and faults observed in ship systems. Thus, the major contributions of this paper are the presentation of a full-NILM framework capable of performing fault detection and isolation (FDI) via NN-based disaggregation, and the demonstration of the system's performance in a real-world military ship setting.

This paper expands upon the preliminary work originally presented in [30]. Namely, we replace the previously presented correlation method for load identification with the NN-based method which significantly improves accuracy. Additionally, we analyze the NILM performance over a longer time period allowing the detection of additional machinery faults beyond the single fault presented in [30]. Finally, we present three additional operational metrics important for fault detection and identification. We demonstrate their usefulness through histogram comparisons of normal and faulty machinery operation and their empirically derived likelihoods of missing faults or incorrectly reporting faults.

II. US Coast Guard Cutter SPENCER

The USCG cutter (USCGC) SPENCER (Fig. 1) is a Famous class, 270 ft. (82 m), medium endurance vessel based in Boston, MA. The ship maintains a 100-person crew and an operational tempo of 185 days deployed per year. Typical patrols require one to two months at sea with operation purposes including environmental stewardship, law enforcement, fisheries protection, and national security.

When at sea, two 475 kW V12 Caterpillar ship-service diesel generator (SSDG) sets provide power to the ship's microgrid, a delta-configured 60 Hz, 254/440V system. These generator sets, along with the twin ALCO V18 main diesel engines (MDE) propelling the ship, are located in the ship's engine room. Each diesel machine requires auxiliary electrical equipment, e.g., fluid pumps and heaters, to maintain operational readiness when in standby mode. Two electrical subpanels (port subpanel and starboard subpanel) power these loads along with several other engine room loads (see Table I) critical for ship operation. Many of these loads operate under closed-loop hysteresis control and in tandem with other loads as part of a larger controlled system, e.g., the keep-warm systems of the diesel machines. We installed NILM meters on the two subpanels to track the operation of these engine room systems and to develop and test methods for condition monitoring.

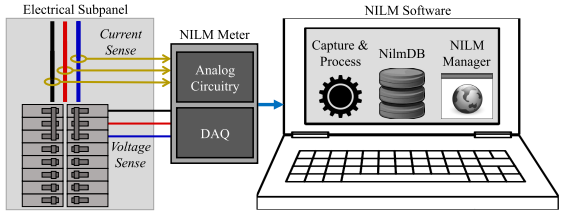
Fig. 1. The USCGC SPENCER at sea [31].

TABLE I
Engine Room Loads Monitored Through the Two NILM Systems

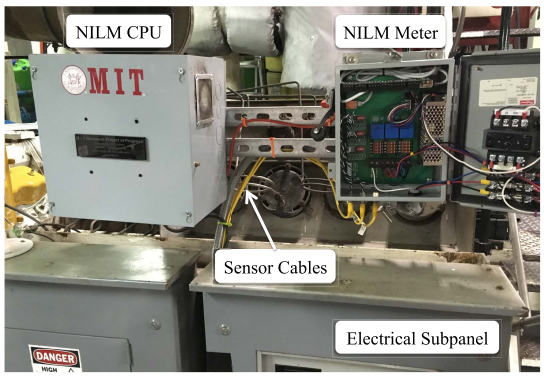
Load	Power Rating	Delta Phases	Power Factor	Port Panel	Stbd Panel
Main diesel engine (MDE) keep-warm system					
Lube oil (LO) heater	12 kW	3ϕ	1.0	x	x
Jacket water (JW) heater	9.0 kW	3ϕ	1.0	x	x
Prelube pump	2.2 kW	3ϕ	0.82	x	x
Ship service diesel generator (SSDG) keep-warm system					
Jacket water (JW) heater	7.5 kW	3ϕ	1.0	x	x
Lube oil (LO) heater	1.3 kW	1ϕ	1.0	x	x
Diesel oil (DO) purifier 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 (CPP) hydraulic pump	7.5 kW	3ϕ	0.82	x	x
Auxiliary saltwater (ASW) cooling pump	7.5 kW	3ϕ	0.85	x	
Gray water pumps	3.7 kW	3ϕ	0.85	x	
Bilge and ballast pump	11 kW	3ϕ	0.85		x
Shaft turning gear motor	1.5 kW	3ϕ	0.76		x
Main diesel engine fuel tank drain pump	1.0 kW	3ϕ	0.85		x
Oily water separator	6.7 kW	3ϕ	0.90		x

III. Non Intrusive Load Monitoring

Fig. 2 shows a conceptual overview of the NILM system along with an image of the port subpanel NILM installed



(a)



(b)

Fig. 2. (a) Overview of the Nonintrusive Load Monitor platform and (b) the port subpanel NILM installation on the USCGC SPENCER.

on the SPENCER. The NILM meter measures the 3-phase voltages and currents supplied to the subpanel. A data acquisition unit (DAQ) samples the sensor outputs and transmits the data via ethernet to a host computer. This computer processes the current and voltage data into power streams, which are then stored in a NILM-optimized database (NilmDB [32]) and made available to the NILM Manager [33], a platform for creating and operating custom load identification and condition monitoring algorithms.

A. Data Acquisition and Preprocessing

The DAQ samples sensor outputs at 8 kHz per channel with 16-bit resolution. This high sample rate and high resolution allows the precise measurement of power transients useful for identifying individual loads changing states. To reduce memory requirements while maintaining transient information, the host PC preprocesses the raw sensor data using the Sinefit algorithm [20], which extracts spectral envelopes for each phase current over each measured voltage line cycle. That is, in-phase and quadrature current components are calculated as,

$$a_k[m] = \frac{2}{N}\sum_{n=0}^{N-1} i\,[n]\cos\left(k\frac{2\pi n}{N}\right) \tag{1}$$

$$b_k[m] = \frac{2}{N}\sum_{n=0}^{N-1} i\,[n]\sin\left(k\frac{2\pi n}{N}\right), \tag{2}$$

respectively. Here, m represents the time-index for a particular line-cycle (thus $a_k[m]$ and $b_k[m]$ are calculated at a rate of 60 Hz), N represents the number of data points collected over that line cycle, and k is the line-frequency harmonic (e.g. $k = 1$ corresponds to 60 Hz, and $k = 7$ to the 420 Hz component). With the ship's phase voltages well approximated as having fixed amplitudes, the real and reactive power components can be calculated as,

$$P_k[m] = V_{ph}a_k[m] \tag{3}$$

$$Q_k[m] = V_{ph}b_k[m], \tag{4}$$

respectively. Here, V_{ph} is the phase voltage amplitude (254 V_{rms}).

This spectral conversion has two major benefits. First, it converts the time-dependent amplitudes and phases of the currents into data streams useful in disaggregating individual loads in the power network. When a load turns on or off, there is a corresponding transient step in $P_1[m]$ and $Q_1[m]$. The size of the transient corresponds to the size of the load and the direction of the step in $P_1[m]$, positive or negative, indicates if the load turned on or off. For non-linear loads, e.g., lighting ballasts, variable speed motor drives, and even some conventional motors [34], transient steps in the harmonic streams ($k > 1$) accompany those in the fundamental streams ($k = 1$).

The distribution of the transient in $P_1[m]$ and $Q_1[m]$ provides information about the load type. Loads that are dominantly resistive, such as heaters, cause transients in $P_1[m]$ but not in $Q_1[m]$, while loads that have an inductive component, e.g. induction motors, cause transients in both. Further, the transient shapes of these step changes also correspond to the physical operation of the loads. The ship's heaters have resistivities that are stable with temperature so transients are flat steps with no overshoot. The ship's motors however draw substantial in-rush currents during start-up leading to transients with large peaks followed by decays to steady state. Thus, monitoring these power streams for these transient characteristics allows for better load identification.

The second benefit for the spectral conversion process is that it effectively compresses the 8 kHz current and voltage data into 60 Hz power streams with a compression factor,

$$C_r = \frac{f_s}{N_k f_l} \tag{5}$$

where f_s is the sample rate, N_k is the number of harmonic envelopes calculated including the fundamental, and f_l is the line frequency, 60 Hz. The NILM systems installed on the SPENCER calculate the real and reactive power components for each phase at the fundamental ($k = 1$) and 3rd, 5th, and 7th harmonics ($k = 3, 5, 7$). Thus, $N_k = 4$, and the effective compression ratio (C_r) is 33, but the spectral streams still maintain the important amplitude, phase, and harmonic information contained in the 8 kHz current waveforms.

This sample frequency was chosen to accommodate accurate spectral streams for all harmonics. For these NILM systems monitoring up to the 7th harmonic (420 Hz), the minimum sample rate required according to the Nyquist Theorem is 840 Hz, however in practice an often used rule of thumb is to increase the sample rate by an approximate factor of 10 beyond the Nyquist rate. In addition to accommodating for any higher frequency content in the line currents, increasing the sample rate also benefits measurements of the lower frequency content. Considering (1) and (2), $a_k[m]$ and $b_k[m]$

are the averaged product of the measured current and the corresponding sinusoid across all sampled data points during a line cycle. The effects of noise in this current measurement therefore diminish as the sample rate increases, leading to an increase in the signal-to-noise ratio of $P_k[m]$ and $Q_k[m]$.

Further, increasing the sample rate also reduces the effects of spectral leakage caused by non-integer ratio relationships between the ship's electrical frequency and the sample rate. Spectral leakage causes consecutive line-cycle estimates of $P_k[m]$ and $Q_k[m]$ to exhibit step-changes when the number of samples in a given line-cycle increases or decreases by one from the previous line-cycle. These steps decrease in size, and thus look less like loads turning on or off, as the sampling rate increases. While perfectly matching the sample rate to be an integer multiple of the nominal electrical system frequency in theory eliminates this effect, in practice fluctuations in the ship's electrical system frequency (measured to be up to $\pm$ 0.5 Hz around the nominal 60 Hz on the SPENCER) make this impractical. Ultimately, the sample rate of 8 kHz produces sufficiently authentic $P_k[m]$ and $Q_k[m]$ streams from which transient features (described below) are extracted with high repeatability.

B. Transient Detection and Feature Extraction

The first step towards identifying load state changes is to detect transients in the phase-B power stream as this phase powers all monitored loads including the single-phase lube oil load (Table I). For this task the NILM employs a step-detector. First, it filters the real power ($P_1[m]$) stream through a 101-point moving median filter to smooth the data, and then it applies a first order difference filter with the form,

$$y[m] = x[m] - x[m-1], \tag{6}$$

where x is the real power after passing through the median filter. If $y[m]$ is greater than a positive threshold ΔP_{th} (set to 300 W on the SPENCER), the NILM identifies $t_{on} = t[m]$ as the location of a turn on event. If $y[m]$ is less than the negative of this threshold, the NILM identifies $t_{off} = t[m]$ as the time of a turn off event. To reduce the number of false detections, the NILM imparts a minimum time between consecutive turn on events or turn off events.

For each detected transient, the NILM system calculates a set of features on each phase. Turn on event features include the changes in steady state real and reactive power, the transient peaks of both the real and reactive powers, and the transient duration. Fig. 3a provides a reference diagram of these feature metrics for a conceptual turn on transient.

To calculate the duration of the start-up transient (Δt_{tran}) the NILM again applies the difference filter of (6), but in this case x is the unfiltered real power. Then the NILM calculates the absolute value of a three-point moving average,

$$|\bar{y}[m]| = \frac{1}{3}|y[m-1] + y[m] + y[m+1]| \tag{7}$$

The end of the start-up transient, $t_{end} = t[m]$, is defined as the time when $|\bar{y}[m]|$ is less than a set threshold, ΔP_{tran} (set to 10 W on the SPENCER). With t_{end} and t_{on}, the NILM calculates the transient duration as,

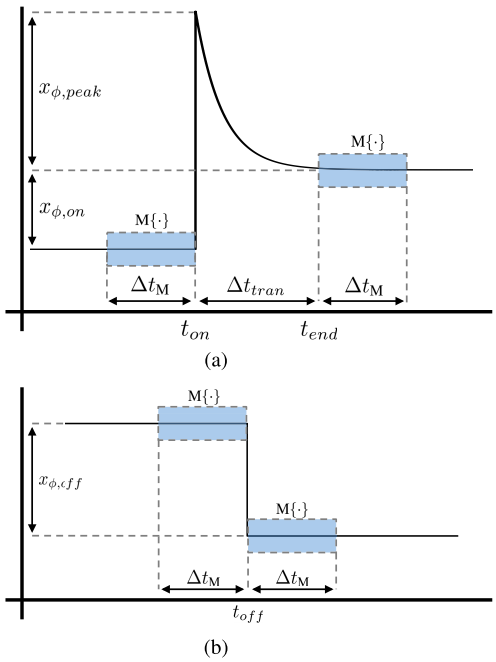


Fig. 3. Signature features extracted from (a) turn on transient and (b) turn off transient.

$$\Delta t_{tran} = t_{end} - t_{on}. \tag{8}$$

This metric effectively determines the time it takes for the load to reach steady-state. Therefore, changes in steady state real and reactive powers after a load turns on are calculated as the difference between their median values over windows Δt_M seconds in length before and after this transient, i.e.,

$$x_{\phi,on} = \mathrm{M}\{x_\phi(t_{end} < t \leq t_{end} + \Delta t_{\mathrm{M}})\} - \mathrm{M}\{x_\phi(t_{on} - \Delta t_{\mathrm{M}} \leq t < t_{on})\}. \tag{9}$$

Here, M{·} represents the median function, x can be either the real or reactive power streams, and ϕ represents the phase (A, B, or C). These windows are shown in blue in Fig. 3a, and are set to $\Delta t_M = 1$s for this study. The transient real and reactive peak values are calculated as,

$$x_{\phi,peak} = \max\{x_\phi(t_{on} \leq t < t_{end})\} - \mathrm{M}\{x_\phi(t_{end} < t \leq t_{end} + \Delta t_M)\}. \tag{10}$$

These values correspond to the maximum in-rush currents of the load as it turns on.

During turn off however, there is no peak in the power streams (Fig. 3b) as loads are simply disconnected via contactors, so the salient features are the real and reactive power steady-state step changes. These are calculated as,

$$x_{\phi,off} = \mathrm{M}\{x_\phi(t_{off} < t \leq t_{off} + \Delta t_M)\} - \mathrm{M}\{x_\phi(t_{off} - \Delta t_M \leq t < t_{off})\}. \tag{11}$$

As a final prevention against falsely detected transients, the NILM ignores any transient without at least one phase showing a steady-state change in real or reactive power larger in magnitude than 300 W / VAR.

C. Load Identification

The NILM implements two fully-connected neural networks (NNs) with two 10-node hidden layers for load identification from the extracted signature features. Each NN was trained off-line using back-propagation and stochastic gradient descent (SGD) [35]. Separate NNs are used for the turn on events and turn off events on each of the two panels, with the input layer receiving the signature features described above. Thus, for the turn on NNs, the input layers have 13 nodes, while the turn off NNs have six nodes.

Each hidden layer node uses a rectified linear unit (ReLU) activation function,

$$f(z) = \max(0, z) \tag{12}$$

where z is the node's input. The output layer for each NN is a softmax layer, with each node producing a value,

$$f(z)_j = \frac{\exp(z_j)}{\sum_{k=1}^{K} \exp(z_k)} \quad \text{for } j = 1 \ldots K \tag{13}$$

where z_j is the input to the jth node, while K is the total number of nodes and corresponds to the number of potential classifications. The choice of a softmax output layer enables multi-class classification. The output is a vector of probabilities that sums up to 1, with the prediction made by selecting the classification that has the highest probability.

As indicated in Table I, many of the monitored loads comprise a larger system and frequently turn on or off simultaneously. This is particularly true for the MDE keep-warm system. To improve classification accuracy, we created additional load classes for combinations of this system's loads. The DO purifier is also comprised of multiple electrical or electromechanical loads, two of which are monitored (notably, a variable-power electric heater is powered from a separate panel). The total power drawn by this system varies during the oil purification cycle as the volume of oil and water varies in the separation chamber and impurities are periodically discharged. Most significantly though, a flushing sequence near the end of the cycle causes a detectable step-up in power. However, the sequence ends with a slow decay in power rather than a step down. Thus, we include an output class for this flushing sequence in the turn on NN, but not the turn off NN. In total, the port panel turn on NN has 14 output classes, while the turn off NN has 12 output classes (the turn off NN has no class for the flushing sequence and no class for all three MDE auxiliary loads turning off simultaneously).

D. Training and Performance Evaluation

To train and test the performance of the NNs, approximately one month of data in-port (Dec. 2016) and one month of data at sea (latter half of Sept. 2016 and latter half of Jan. 2017) was manually labelled based on crew operational logs and physical load parameters [36]. This data was then randomly split into three sets: training, validation, and testing, while ensuring equal distributions of each labeled load class. The validation data set's prediction error was used as a stopping criterion during training [37], while the testing data set was used to ensure the model did not over-fit on the training and validation data. This process was repeated 10 times (Monte Carlo cross-validation) to generate statistical metrics of the NNs performance and variability.

Due to the wide disparities in usage schedules between loads, there is a large imbalance in the number of samples for each class. For example, over the course of a month, gray water pumps turn on and off hundreds of times, while other loads such as the ASW Pump only turn on and off a few times. To prevent the NNs from simply predicting the most common class, the training data was "balanced" by both over-sampling the minority class and under-sampling the majority class. For minority classes, the algorithm applies the synthetic minority over-sampling technique (SMOTE) [38]. In SMOTE, rather than over-sampling with replacement, the minority class is over-sampled by creating synthetic examples along the line segments that join the feature values of the nearest neighbors in the minority class.

The accuracy of the neural networks can be evaluated across each load class by considering two commonly used metrics for classification algorithms, precision and recall [39]. In terms of load identification, recall states the likelihood that the NILM will report when a load turns on (or off). Precision states the likelihood that a reported turn on (or off) event actually happened. More plainly, *recall* answers the question, "*do true events get reported?*", while precision answers the question, "*are reported events true?*"

To calculate these metrics, we tally-up the number of true positive (TP), false positive (FP), and false negative (FN) events for each load class of each NN. For turn on events, these metrics are defined as:

- TP (True positive): The NILM correctly reports that the load turned on.
- FP (False positive): The NILM reports that the load turned on, but it was either a different load that turned on or no load turned on.
- FN (False negative): The NILM does not report that the load turned on when it did. The NILM may have misclassified the turn on event as a different load or missed the event entirely.

Analogous metrics can be defined for turn off events. Precision and recall are then calculated, respectively, as,

$$\text{Pr} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{14}$$

and

$$\text{Re} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{15}$$

Table II reports the average and standard deviation (σ) in the number of TPs, the precision, and the recall of the port panel NILM's transient identification across the 10 randomly selected test data sets. Similar results were achieved on the starboard subpanel. In this table, the performance of each NN's classification accuracy of MDE components are reported by individual load. If the NN correctly identified simultaneous load transients, then each individual load tallied a TP. If the NN misidentified the simultaneous load transient, then any subset of individual loads correctly identified tallied a TP,

TABLE II
ACCURACY OF CLASSIFYING ON-EVENTS

Engine Room System	Load Profile	*Turn On Events*				*Turn Off Events*			
		Total	**TP (σ_{TP})**	**Pr (σ_{Pr})**	**Re (σ_{Re})**	**Total**	**TP (σ_{TP})**	**Pr (σ_{Pr})**	**Re (σ_{Re})**
MDE keep-warm system	LO heater	26	25.7 (0.95)	97% (3.8%)	99% (3.6%)	27	27 (0)	98% (1.9%)	100% (0%)
	JW heater	20	19.7 (0.95)	100% (0%)	99% (4.7%)	21	20.9 (0.32)	100% (0%)	100% (1.5%)
	Prelube pump	24	23.8 (0.42)	93% (3.9%)	99% (1.8%)	26	26 (0)	98% (5.0%)	100% (0%)
SSDG keep-warm system	JW heater	219	219 (0)	100% (0%)	100% (0%)	219	219 (0)	100% (0.2%)	100% (0%)
	LO heater	28	28 (0)	77% (4.5%)	100% (0%)	26	26 (0)	74% (3.4%)	100% (0%)
DO purifier system	Separator motor	19	17.8 (1.1)	96% (3.4%)	94% (6.0%)	19	18.8 (0.42)	99% (3.0%)	99% (2.2%)
	Feed pump	15	15 (0)	99% (2.0%)	100% (0%)	14	14 (0)	94% (4.6%)	100% (0%)
	Flush sequence	19	17.7 (1.1)	95% (5.3%)	93% (5.6%)	N/A	N/A	N/A	N/A
Additional engine room loads	CPP pump	19	18.9 (0.32)	96% (7.0%)	99% (1.7%)	21	20.8 (0.63)	97% (3.1%)	99% (3.0%)
	ASW pump	2	1.8 (0.63)	93% (15%)	90% (32%)	3	3.0 (0)	89% (15%)	100% (0%)
	Gray water pump	409	407.2 (1.0)	99% (0.28%)	100% (0.4%)	417	415.9 (1.4)	99% (0.3%)	100% (0.3%)

any load incorrectly identified tallied an FP, and any load missed tallied an FN. The total tallies of TPs, FPs, and FNs were used to calculate the precision and recall in accordance with (14) and (15).

The high average accuracy metrics across most load classes combined with the generally small standard deviations indicate that the NNs are not overfitting the training and validation data and that there is good consistency between iterations. The primary exception to this is the relatively low precision of the SSDG LO heater. This load is unique from the others in that its the salient single-phase load powered from the panel. During training, the NN tends to use this load as a "catch-all" for incorrectly detected transients, i.e., those detected by the step-detector of (6) not corresponding to a labeled event. On average, the step-detector incorrectly detects 11 transients during the turn on testing data set with an average of eight incorrectly categorized as an SSDG LO heater; during the turn off testing data set, the step-detector incorrectly detects 14 transients with an average of nine categorized as an SSDG LO heater. The NILMs are configured to identify loads as the class with the maximum probability as reported by the NNs, even when no class is significantly more probable than the others. Thus, these added FPs for the SSDG LO heater could potentially be reduced by relaxing this restriction and allowing erroneously detected transients to be classified as "no known load."

Comparing the recall performance of the turn on NN with that of the turn off NN reveals a slight trend for the turn on NN to perform worse than the turn off NN. While these differences are statistically insignificant according to a two-sample Kolmogorov-Smirnov test for 5% significance, the general trend does correspond to slightly more FNs across the loads for the turn on NN than the turn off NN. These FNs often correspond to FPs for the wrong loads (the NN identifies the wrong load, but the transient is not missed by the event detector). For example, while missed turn on events are very rare for the gray water pumps (99% recall on average), when they are missed they are often misidentified as MDE prelube pump turn on events. This in turn leads to the lower precision score for the MDE prelube pump turn on events. While the turn on NN has more feature inputs useful for load identification, these inputs have more variance across events than do the those of the turn off events due to the in rush currents of the pumps and the need for automated detection of the transient time-length in (8). With these two loads of similar rated powers, there are occasional overlaps in feature values. Still, for the purposes of this paper the performance of the NNs are sufficient for application towards shipboard fault detection.

IV. DIAGNOSTIC INDICATORS FOR FAULT DETECTION

Many of the engine room loads on the SPENCER operate under closed-loop hysteresis control, regulating machine system operating points such as the temperature of the SSDG lube oil and the water level of the gray water holding tank. While these control systems free crew members from manually monitoring and controlling equipment, closed-loop control can mask underlying electromechanical faults [29], [40]. Without perceivable changes to the operating points, subtle problems that do not result in complete system failure often go undetected. These problems, however, can increase energy consumption, impose excessive wear on electromechanical systems, and eventually result in catastrophic failure. For these systems, monitoring run-time and run-frequency metrics, in addition to those used for load disaggregation, provides a useful diagnostic framework for fault detection and identification.

Fig. 4 shows the simulated closed-loop operation of an SSDG JW heater when the generator is in standby. During this time, the heater works in tandem with a circulation pump (powered from an unmonitored panel) to ensure the generator remains at a temperature conducive for startup, between 90°F (32°C) and 120°F (49°C). As shown in the figure, when the circulating water temperature drops to this lower limit, the heater turns on until the water reaches its high temperature limit. At this time, the heater turns off and the temperature drops again.

The outputs of the NNs provide these turn on and turn off times for each load, making it easy to calculate two additional metrics, the time duration that a load is on, T_d, and the period between runs, T_p. These metrics have been shown to be useful diagnostic indicators for detecting pressure leaks in vacuum pumps and compressed-air systems under closed-loop

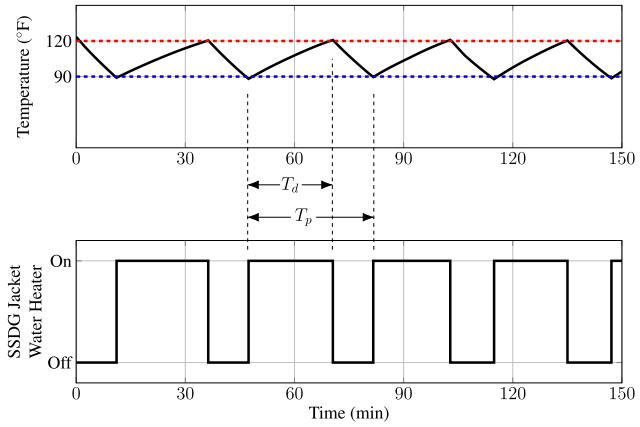


Fig. 4. Simulated operation of a ship service diesel generator jacket water heater under closed-loop hysteresis control.

hysteresis control [29]. These times are labeled for the SSDG jacket water heater example in Fig. 4.

Further, the effective duty cycle,

$$D_e = \frac{T_d}{T_p} \tag{16}$$

is a metric that's both intuitive for engineers and represents a nonlinear mapping, $f\left(T_d, T_p\right)$, providing an additional diagnostic dimension. The total number of runs per day for each load, N_{rpd}, is a metric intuitively related to short-cycling failure modes. Finally, the equipment's steady-state real power, P_{ss}, derivatively relates to the total work it performs over a run cycle. Using pumps as an example, an increase in power draw may indicate aging of the pumps due to inefficiencies introduced by degradation. Collectively, these metrics,

$$\left(N_{rpd}\ T_d\ T_p\ D_e\ P_{ss}\right), \tag{17}$$

provide a robust reference frame for diagnosing hysteresis-controlled system faults. Here, P_{ss} is the three-phase real power feature associated with the corresponding off transient, i.e.,

$$P_{ss} = P_{A,ss} + P_{B,ss} + P_{C,ss}. \tag{18}$$

System Application: Since installing the NILM meters on the SPENCER, we have observed multiple faults in the gray water disposal system. In assessing the diagnostic framework's usefulness in disaggregating faulty system operation from normal operation, we consider the amount of overlap between metric histograms during individual fault conditions and those during normal operation. More concretely, we report the likelihood of Type I and Type II errors (Table III) assuming the histograms are true representations of metric distributions, and that given a metric measurement, a fault detection algorithm will classify the system's operation as normal or faulty by the histogram with the higher probability score for the corresponding bin. Here Type I errors refer to incorrectly reported faults (analogous to FP errors by the NNs), while Type II errors refer to missed faults (analogous to FN errors).

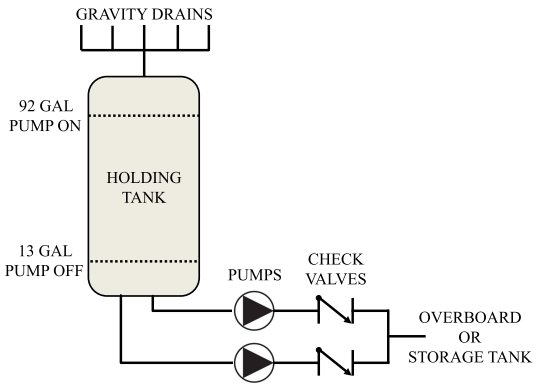


Fig. 5. Conceptual diagram of the gray water system.

V. Gray Water Disposal System

The gray water disposal system collects and transfers or disposes the relatively clean water from showers, sinks, washing machines, and other appliances. As depicted in Fig. 5, gravity drains transfer the water from individual receptacles to a 138 gallon (522 L) holding tank. When the tank is full, a pump discharges the water from the tank either overboard or to a larger storage tank depending on the vessel's location and applicable regulations. Two identical pumps (for redundancy) alternate each cycle to discharge the tank. Conductivity sensors detect water levels and provide feedback for pump control. When water reaches the high sensor (92-gallon mark), a pump turns on and begins discharging. The pump then turns off when water reaches the low sensor (13-gallon mark). Thus, the pump expels 79 gallons (300 L) during each normal pump run.

During monitoring, the gray water system experienced two types of faults, the first a failed high-level tank sensor and the second likely a failed check valve. Both types of faults went unnoticed by the crew for extended periods as the gray water disposal system still performed its job collecting and disposing gray water. However, these faults caused significantly more pump runs and in some cases short-cycling runs, both which place undue stress on the system.

Fig. 6 shows the effects of these faults in the feature space defined in (17) as compared to normal operation. Under normal conditions, approximately 15 total pump runs occur each day. Pump runs typically have a duration (T_d) ranging between 60 seconds to 80 seconds, though the time period between runs (T_p) varies significantly as pump run frequency is highly dependent on crew schedules. With often relatively long times between runs, the effective duty cycle of the system ranges from near 0% to about 6%.

A. Faulty High-Level Sensor

In December 2015, the high-level sensor failed due to solid residue shorting its electrodes. This caused premature high-level readings, which in turn resulted in frequent pump runs with an effective period on the order of just a couple minutes. These pump runs were also very short, typically only one to four seconds, as the tank filled very little between runs. This fault went unnoticed by the crew, but resulted in almost

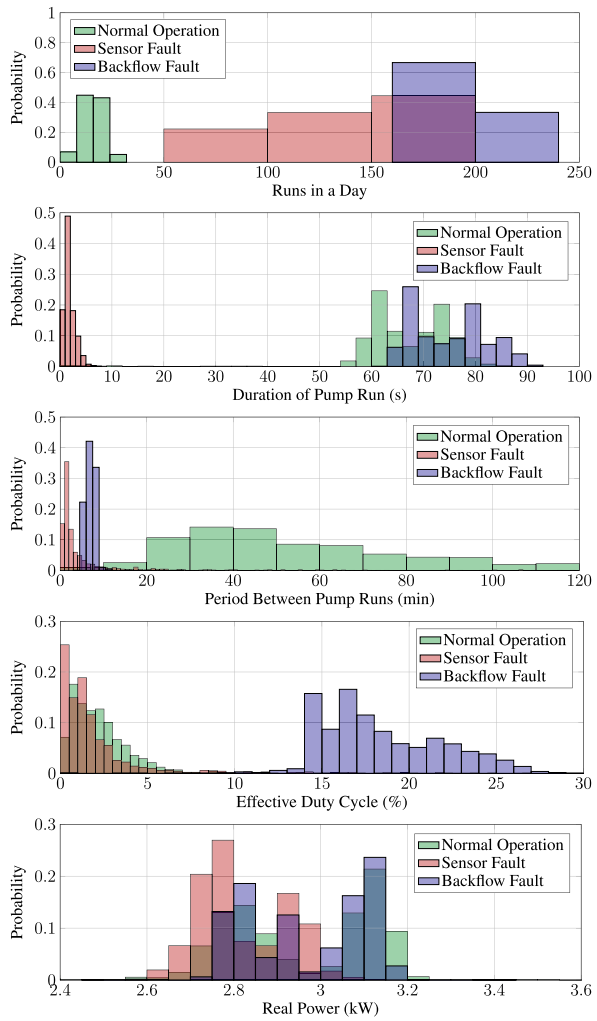


Fig. 6. Histograms showing the effects of two faults on the fault detection feature space.

10× more *recorded* pump run events than normal. In fact, the *actual* number of pump runs was certainly higher, as in some cases the runs were too short for the load to reach steady-state, causing the NILM to ignore the detected transient.

The histograms in Fig. 6 show that monitoring for runs per day, pump duration, and pump period provide a highly-effective reference frame for detecting this type of fault. Even including outliers caused by on/off events missed by the NILM or breaks in recorded data, the normal and sensor fault runs per day and pump duration histograms have virtually no overlap (<1% of either Type I or Type II errors). Similarly, only 1% of the normal period histogram is covered by the sensor fault's histogram (Type I error), though 15% of the sensor fault histogram is covered by the normal's (Type II error). Both features show superior separation to that of the real power features used by the NILM for identifying loads.

B. Storage Tank Backflow Failure

During the months of January to March 2017, the gray water disposal system exhibited a second fault type. This fault occurred each time the crew redirected the gray water system output from overboard to a 6000 gallon (22.7 kL) storage tank in preparation for docking in a foreign port. After approximately 10 normal pump runs, the pumps proceeded to run at a very regular interval, $T_p \approx 7$ mins, or roughly 13× more frequently than normal. Simultaneously, the pump durations increased by approximately 10 seconds resulting in an effective duty cycle range of approximately 15% to 28%. The histograms of Fig. 6 and overlap metrics of Table III show that the runs per day, periods between pump runs, and the corresponding duty cycle histograms contain very little overlap between the pumps' normal operation and during the backflow fault.

TABLE III

LIKELIHOOD OF TYPE I AND TYPE II ERRORS BASED ON HISTOGRAM DISTRIBUTIONS

Metric	Sensor Fault		Backflow Fault	
	Type I	Type II	Type I	Type II
Runs per day	<1%	<1%	<1%	<1%
Duration	<1%	<1%	10%	32%
Period	7.8%	15%	3.3%	1.7%
Duty	22%	49%	<1%	<1%
Real Power	27%	17%	70%	8.0%

While a 13× increase in water use could also explain these shifts in feature metrics, the gray water system exhibited these characteristics the entirety of the "inport" status. Over these times, that much water use would have exceeded the storage tank's capacity several times over, though the ship's crew did not report emptying this storage tank in their logs, nor did they mention any overflows. Instead, the system likely experienced a backflow fault due to a failed check valve. The gray water expelled from the holding tank into the storage tank (which sits higher on the boat) flowed back through the broken check valve after the gray water pump turned off causing the holding tank to refill and the process to repeat.

C. Signs of Future Pump Failure

The period of normal system operation used in this analysis, Aug. 6th, 2016 to Nov. 3rd, 2016, also revealed data potentially corresponding to an aging pump. Fig. 7 depicts a scatter plot of normal pump runs in the P_{ss} vs. T_d (real power vs. run duration) reference frame. The plot reveals two distinct clusters, the first centered at approximately 62 seconds and 2.8kW, and the second centered at 73 seconds and 3.1kW. Using the 2-means clustering algorithm to categorize this data and crosschecking timestamps to ensure an alternating pattern of pump operation, we can assign the two gray water pumps to the two clusters as shown in the figure. Doing so suggests that pump 2 draws approximately 10% more power and operates 11 seconds longer than pump 1. This corresponds to pump 2 requiring nearly 30% more energy for each pump run than pump 1.

Unfortunately, the crew does not have detailed records of pump replacements and repairs, though the ship's chief mechanic reported that one of the pumps had been replaced within the last few years. Thus, its reasonable to speculate that the less efficient pump operates in this manor due to aging,

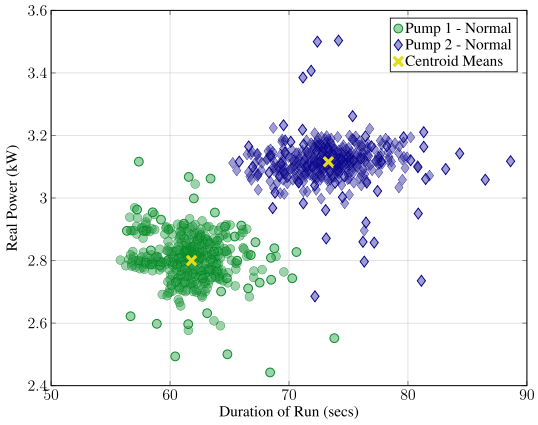


Fig. 7. The feature space of pump duration and pump real power draw shows two distinct clusters corresponding to the two pumps.

and the longer durations and higher power draws might be an indication of impending failure.

VI. Conclusion

As maritime crew sizes decrease, ships increasingly rely on automated sensor and instrumentation systems to ensure optimal operation of equipment. In particular, closed-loop controlled machines benefit from advanced monitoring as controllers can obscure machinery faults. In the best-case scenario, the faults cause inefficiencies incurring additional resource costs. In worst-case scenarios, faults eventually lead to catastrophic failure resulting in extended system downtime and significant capital expenditures for machinery replacement.

This paper presented a framework for using nonintrusive load monitoring (NILM) for fault detection and isolation (FDI). Two NILMs were installed in the engine room of the USCGC SPENCER collecting bulk power data on 19 ship loads. Using a neural network classification algorithm, the NILM disaggregates this power information by load, in turn providing several metrics for fault diagnostics. We explored this metric space in relation to faults observed in the ship's gray water system and showed these metrics provide information sufficient for effective detection and identification of fault types. Ultimately, these auto-generated metrics could be combined with advanced clustering algorithms [41] for fully automated fault detection and identification [42], [43].

Acknowledgment

The authors thank the crew of the U.S. Coast Guard Cutter SPENCER for their commitment to duty and their generosity and technical support in accessing and understanding ship systems and operations.

References

[1] R. Powers, "Automation as a manpower reduction strategy in navy ships," M.S. thesis, Dept. Mech. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 2016.

[2] M. Ljung and M. Lützhöft, "Functions, performances and perceptions of work on ships," *WMU J. Maritime Affairs*, vol. 13, no. 2, pp. 231–250, 2014.

[3] K. P. Logan, "Intelligent diagnostic requirements of future all-electric ship integrated power system," *IEEE Trans. Ind. Appl.*, vol. 43, no. 1, pp. 139–149, Jan. 2007.

[4] A. S. Eruguz, T. Tan, and G.-J. van Houtum, "A survey of maintenance and service logistics management: Classification and research agenda from a maritime sector perspective," *Comput. Oper. Res.*, vol. 85, pp. 184–205, Sep. 2017.

[5] C. Schantz *et al.*, "Retrofittable machine condition and structural excitation monitoring from the terminal box," *IEEE Sensors J.*, vol. 16, no. 5, pp. 1224–1232, Mar. 2016.

[6] R. Zachar, P. Lindahl, J. Donnal, W. Cotta, C. Schantz, and S. B. Leeb, "Utilizing spin-down transients for vibration-based diagnostics of resiliently mounted machines," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 7, pp. 1641–1650, Jul. 2016.

[7] H. Dzapo, Z. Stare, and N. Bobanac, "Digital measuring system for monitoring motor shaft parameters on ships," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 10, pp. 3702–3712, Oct. 2009.

[8] A. Benigni, G. D'Antona, U. Ghisla, A. Monti, and F. Ponci, "A decentralized observer for ship power system applications: Implementation and experimental validation," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 2, pp. 440–449, Feb. 2010.

[9] W. Li, A. Monti, and F. Ponci, "Fault detection and classification in medium voltage DC shipboard power systems with wavelets and artificial neural networks," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 11, pp. 2651–2665, Nov. 2014.

[10] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16838–16866, 2012.

[11] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Trans. Consum. Electron.*, vol. 57, no. 1, pp. 76–84, Feb. 2011.

[12] N. F. Esa, M. P. Abdullah, and M. Y. Hassan, "A review disaggregation method in non-intrusive appliance load monitoring," *Renew. Sustain. Energy Rev.*, vol. 66, pp. 163–173, Dec. 2016.

[13] K. Suzuki, S. Inagaki, T. Suzuki, H. Nakamura, and K. Ito, "Nonintrusive appliance load monitoring based on integer programming," in *Proc. SICE Annu. Conf.*, Aug. 2008, pp. 2742–2747.

[14] M. Z. A. Bhotto, S. Makonin, and I. V. Bajić, "Load disaggregation based on aided linear integer programming," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 64, no. 7, pp. 792–796, Jul. 2017.

[15] J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Proc. Workshop Data Mining Appl. Sustainability (SIGKDD)*, San Diego, CA, USA, vol. 25, 2011, pp. 59–62.

[16] M. Aiad and P. H. Lee, "Unsupervised approach for load disaggregation with devices interactions," *Energy Buildings*, vol. 116, pp. 96–103, Mar. 2016.

[17] G. C. Koutitas and L. Tassiulas, "Low cost disaggregation of smart meter sensor data," *IEEE Sensors J.*, vol. 16, no. 6, pp. 1665–1673, Mar. 2016.

[18] N. Sadeghianpourhamami, J. Ruyssinck, D. Deschrijver, T. Dhaene, and C. Develder, "Comprehensive feature selection for appliance classification in NILM," *Energy Buildings*, vol. 151, pp. 98–106, Sep. 2017.

[19] L. K. Norford and S. B. Leeb, "Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms," *Energy Buildings*, vol. 24, no. 1, pp. 51–64, 1996.

[20] 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.

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

[22] T. Hassan, 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.

[23] L. Du, D. He, R. Harley, and T. G. Habetler, "Electric load classification by binary voltage–current trajectory mapping," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 358–365, Jan. 2016.

[24] Y. H. Lin and M. S. Tsai, "Applications of hierarchical support vector machines for identifying load operation in nonintrusive load monitoring systems," in *Proc. 9th World Congr. Intell. Control Automat.*, Jun. 2011, pp. 688–693.

[25] 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.

[26] Y.-H. Lin and M.-S. Tsai, "A novel feature extraction method for the development of nonintrusive load monitoring system based on BP-ANN," in *Proc. Int. Symp. Comput., Commun., Control Automat. (3CA)*, vol. 2, May 2010, pp. 215–218.
[27] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environ.*, 2015, pp. 55–64.
[28] J. Liang, S. K. K. Ng, G. Kendall, and J. W. M. Cheng, "Load signature study—Part II: Disaggregation framework, simulation, and applications," *IEEE Trans. Power Del.*, vol. 25, no. 2, pp. 561–569, Apr. 2010.
[29] J. Paris, J. S. Donnal, R. Cox, and S. B. Leeb, "Hunting cyclic energy wasters," *IEEE Trans. Smart Grid*, vol. 5, no. 6, pp. 2777–2786, Nov. 2014.
[30] J. C. Nation *et al.*, "Nonintrusive monitoring for shipboard fault detection," in *Proc. IEEE Sensors Appl. Symp. (SAS)*, Mar. 2017, pp. 1–5.
[31] United States Coast Guard. *Welcome Aboard the USCGC Spencer (WMEC 905)*. Accessed: Jul. 24, 2017. [Online]. Available: https://www.uscg.mil/lantarea/cgcSpencer/default.asp
[32] 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.
[33] J. S. Donnal, J. Paris, and S. B. Leeb, "Energy applications for an energy box," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 787–795, Oct. 2016.
[34] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.
[35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[36] P. Lindahl, G. Bredariol, J. Donnal, and S. B. Leeb, "Noncontact electrical system monitoring on a U.S. Coast Guard cutter," *IEEE Instrum. Meas. Mag.*, vol. 20, no. 4, pp. 11–20, Aug. 2017.
[37] L. Prechelt, "Early stopping—But when?" in *Neural Networks: Tricks of the Trade* (Lecture Notes in Computer Science), vol. 1524. Berlin, Germany: Springer-Verlag, 1997, ch. 2, pp. 55–69.
[38] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.
[39] S. Makonin and F. Popowich, "Nonintrusive load monitoring (NILM) performance evaluation," *Energy Efficiency*, vol. 8, pp. 809–814, Jul. 2015.
[40] T. DeNucci *et al.*, "Diagnostic indicators for shipboard systems using non-intrusive load monitoring," in *Proc. IEEE Electr. Ship Technol. Symp.*, Jul. 2005, pp. 413–420.
[41] R. Xu and D. Wunsch, II, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, May 2005.
[42] A. Bellini, F. Filippetti, C. Tassoni, and G.-A. Capolino, "Advances in diagnostic techniques for induction machines," *IEEE Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4109–4126, Dec. 2008.
[43] J. Kwak, T. Lee, and C. O. Kim, "An incremental clustering-based fault detection algorithm for class-imbalanced process data," *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 3, pp. 318–328, Aug. 2015.

Peter A. Lindahl received the Ph.D. degree in engineering from Montana State University in 2013. He is currently a Post-Doctoral Associate with the Research Laboratory of Electronics, Massachusetts Institute of Technology. His research interests include sensors and instrumentation for energy and power systems, sustainable energy systems, and energy economics, analytics, and policy.

Daisy H. Green received the B.S. degree from the University of Hawaii at Mānoa in 2015. She is currently pursuing the M.S. and Ph.D. degrees in electrical engineering with the Research Laboratory of Electronics, Massachusetts Institute of Technology.

Gregory Bredariol received the B.S. degree in mechanical engineering from the U.S. Coast Guard Academy in 2011. He is currently pursuing the M.S. degree with the Massachusetts Institute of Technology. He was a Student Engineer and a Damage Control Assistant aboard USCGC DILIGENCE and an Assistant Engineer Officer aboard USCGC BERTHOLF. He is currently on active duty in the U.S. Coast Guard. His research interests include maritime applications of nonintrusive load monitoring.

Andre Aboulian received the S.B. degree in electrical engineering and computer science from the Massachusetts Institute of Technology in 2016, where he is currently pursuing the master's degree with the Research Laboratory of Electronics. His interests include embedded systems, distributed software systems, power electronics, computer architecture, and human–computer interaction.

John S. Donnal received the B.S. degree from Princeton University, Princeton, in 2007, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology in 2013 and 2016, respectively, all in electrical engineering. He is currently a Faculty Member of 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 Ph.D. degree from the Massachusetts Institute of Technology (MIT) in 1993. He has been a member of the MIT Faculty, Department of Electrical Engineering and Computer Science, since 1993. He also holds a joint appointment with the Department of Mechanical Engineering, MIT. He is concerned with the development of signal processing algorithms for energy and real-time control applications.