

Physics-Informed Feature Space Evaluation for Diagnostic Power Monitoring

Daisy H. Green, Aaron W. Langham, Rebecca A. Agustin, Devin W. Quinn, Steven B. Leeb

Abstract—Sensing solutions provide a rich feature set for electromechanical load monitoring and diagnostics. For example, qualities that describe the operation of an electromechanical load can include measurements of power, torque, vibration, electrical current demand, and electrical harmonic content. If properly interpreted, these measurements can be utilized for energy management, condition-based maintenance, and fault detection and diagnostics. When monitoring several loads from an aggregate data stream, a well posed feature space will permit not only load identification, but also the characterization of faults and gradual changes in the health of an individual machine. Many feature selection methods assume static and generalizable data, without consideration of concept drift and evolving behavior over time. This paper presents a method for evaluating load separability in a feature space prior to the application of a pattern classifier, while accounting for changing operating conditions and load variability. A four-year load dataset is used to validate the method.

NOMENCLATURE

CPP	Controllable pitch propeller.
FOP	Fuel oil purifier.
MPDE	Main propulsion diesel engine.
NILM	Nonintrusive load monitor.
P, Q, S	Real, reactive, and apparent powers.
P_{ss}, Q_{ss}	Real and reactive steady-state powers.
SSDG	Ship service diesel generator.

I. INTRODUCTION

Increasingly abundant sensing technologies for electromechanical loads provide access to measurements of power, torque, vibration, electrical current demand, and electrical harmonic content [1], [2]. These measurements provide an abundant feature set for electromechanical load monitoring and diagnostics. Condition-based maintenance (i.e., maintenance performed as necessary to maintain load availability and performance), requires the identification of the gradual degradation of equipment performance, or “soft faults” [3]. Efficiently scheduled service prevents “hard faults” that cause complete equipment failures. For convenience and lower costs, sensors are often installed at a centralized location. As a result, individual loads need to be disaggregated from the

This work was supported in part by the Office of Naval Research NEPTUNE Program and in part by The Grainger Foundation.

Daisy H. Green is with the Department of Architecture, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (email: dhgreen@mit.edu).

Aaron W. Langham, Rebecca A. Agustin, and Steven B. Leeb are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

Devin W. Quinn is with the United States Coast Guard, Alameda, CA 94501 USA.

aggregate stream. However, the challenge in applying pattern classifiers to load datasets is ensuring correct results even in fault scenarios or as operating conditions change. That is, the applicability and utility of pattern classifiers for fault detection and diagnostics relies on having a well-chosen or “informed” feature space with predictable characteristics for recognizing observations that arise outside of initial training data.

One subset of features provided by electromechanical loads can be obtained using a nonintrusive load monitor (NILM). A typical NILM scenario involves monitoring the power service to a collection of loads, e.g. from the main feeder to an electrical service panel or from the output of a source or generator [4]. A great deal of nonintrusive load monitoring research assumes training data is forever representative of new data, without regard to changing operating conditions or faulty behavior [5]. Unfortunately, real data is not static [6]. Load behavior may evolve over time, referred to in many domains as concept drift [7]–[9]. Open-access datasets for nonintrusive load monitoring are generally restricted to healthy residential appliances [10], [11]. These datasets cannot generalize to the raft of loads and different operating conditions and fault conditions possible in residential, industrial, and commercial sites. As a result, the training data for a practical nonintrusive load monitoring classifier likely needs to be collected from the specific system of interest over a short period of time. Such a limited training dataset may not fully characterize the long-term distributions of loads in a given feature space. The accuracy of any classifier depends on the separability of the data. In this context, a separable feature space is one that permits classification of load events even with future load variability and drift. In evaluating feature spaces for separability, the question becomes: is the data uniquely recognizable, and are there, or will there be, unresolvable overlaps in the feature space now or as equipment ages and operating conditions change?

Several methods exist to evaluate feature importance in a given feature space. Neighborhood component analysis and minimum redundancy maximum relevance have been used in the context of nonintrusive load monitoring, due to their physical interpretability [12], [13]. Other methods such as random forest variable importance analysis, recursive feature elimination, and Boruta [14] are effective in ranking feature importance. Techniques for evaluating class separability based on distance measures include interclass distance, intraclass distance, the Fisher ratio, and scatter matrices [15], [16]. However, these methods do not take into account the possibility that a load’s behavior in the feature space will adapt over time. Automated diagnostics in rotating machines, specifically for

bearing faults in different operating environments is detailed in [17]–[19], using transfer learning and domain generalization techniques. However, these papers only consider a single machine, and do not perform disaggregation of multiple loads.

This work's contribution is a method to address the problem of feature space evaluation in a multi-load environment in the presence of physics-related concept drift. The drift is assumed to be gradual, due to either normal operation or load degradation. To do so, this work provides a check for electromechanical load separability to inform the decision of adding features to the feature space. The separability check seeks to determine if a classifier trained on a dataset limited in time will be reliable in the future. The separability check does not assume that the data is independent and identically distributed (i.i.d.) or that classes are balanced. The separability check is independent of the type of classifier used, and does not prescribe any particular classification methodology. Using every feature available can increase data separability, but can result in overly fit, non-generalizable decision boundaries due to the sparsity of training data [20]. In addition, adding extra features may incur extra costs if they require additional sensors or signal processing development. Thus, it is desirable to start with a compact set of features and only add features when necessary.

The proposed separability check is based on geometric overlap using hyperellipsoidal regions. An initial feature space is selected based on known electromechanical load behaviors. This feature space is then assessed by applying the load separability check to a limited set of labeled data, which defines regions of load behavior within the feature space. The presence of overlapping regions indicates current or future regions of uncertainty for a potential NILM classifier. As more labeled data is obtained, this check can be periodically run to re-evaluate the feature space for potential conflicts. The proposed load separability check is demonstrated with a four-year dataset of loads collected from a US Coast Guard (USCG) vessel.

This paper is organized as follows: Section II gives a review of the variability and drift that exists in load operation. Section III introduces a method to check for load separability in a given feature space. Section IV shows a demonstration of these methods with four years of shipboard electromechanical data.

II. LOAD VARIABILITY

There exists an abundance of physical features that reveal information about load behavior and are valuable for load identification. Commonly used features include the fundamental real power (P) and reactive power (Q) changes in steady-state and inrush characteristics, such as peak, duration, and shape [2], [4]. P and Q are computed using the Sinefit algorithm [21] by computing the short-term average of the fundamental in-phase and quadrature components of the sampled current over every ac line cycle and assuming a stiff sinusoidal system voltage. For instance, Fig. 1 and Fig. 2 show two examples of measured current and the computed real and reactive power streams from data collected by a NILM installed on

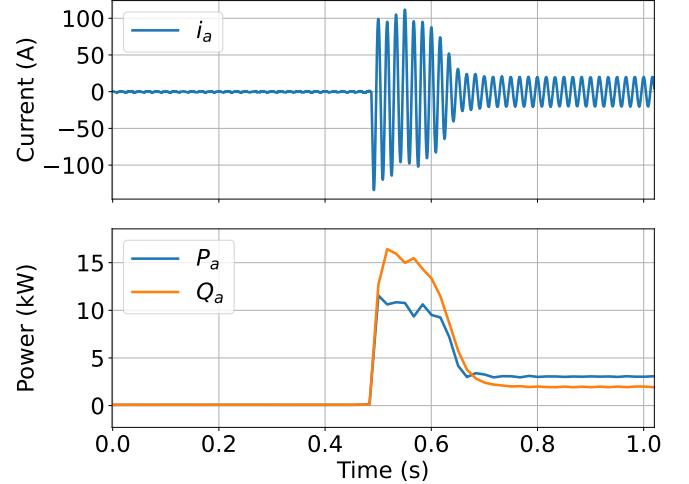


Fig. 1: Phase- a current and real and reactive powers for a turn-on transient of a controllable pitch propeller pump.

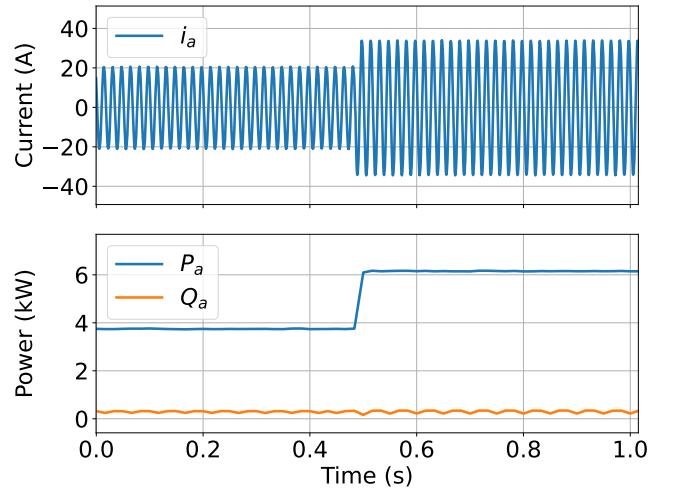


Fig. 2: Phase- a current and real and reactive powers for a turn-on transient of a ship service diesel generator jacket water heater.

a shipboard subpanel. In Fig. 1, a controllable pitch propeller (CPP) pump inrush transient occurs at 0.5 s. In Fig. 2, a ship service diesel generator (SSDG) jacket water heater turns on at 0.5 s, on top of a base load. From the P and Q streams, an event detector extracts inrush and steady-state features for each detected load event. The physical task of some loads may also create higher order current harmonics. For example, power electronics, such as rectifiers without power factor correction, may contribute higher order harmonics at multiples of the line frequency [22]. Direct voltage and current measurements can also serve as additional streams of data for extracting relevant features [23]. Most nonintrusive load monitoring research assumes that these physical features will not change over time. In practice, these features will change with changing operating conditions, machinery aging and wear, or abnormal load behavior. In addition, imbalanced three-phase loads can result from faulty operation, but are outside of the scope of

this work.

The identification of loads, even in drifting states, is necessary for fault detection and diagnostics using nonintrusive load monitoring. Little work has been done on the use of nonintrusive load monitoring for fault detection and diagnostics. Existing research that does consider fault detection mostly deals with load timing faults, such as short-cycling or elongated duty cycle, rather than changing electrical characteristics, such as real and reactive power and higher-harmonic current consumption [24], [25]. Changing electrical characteristics can complicate load identification, but are also useful diagnostic indicators. In addition, drifting load operation can take months or years to manifest in electrical data. This can complicate long-term load disaggregation when using a static classifier. This work demonstrates new feature space evaluation techniques using a four-year dataset collected during our field studies onboard a US Coast Guard vessel. The electrical loads represented in this dataset include three-phase loads with approximately linear models at their operating points, including resistive heaters and induction motors.

Since many classification techniques are not readily interpretable, their classification decision is unclear and potentially inaccurate as more data is collected. This is compounded when the underlying distribution of new data is different than the distribution of the training data, known as concept drift. A load may drift to an entirely different region of the feature space than the classifier expected, and may become non-separable from other loads. It is difficult to determine the optimal separator for future data because there are likely many large-margin, low-density separators that can accurately classify the training dataset [26]. That is, concept drift can cause performance degradation regardless of classifier choice. With physically informed assumptions made on the character of the concept drift, the feature space should be assessed to find the most meaningful set of electrical characteristics that provide adequate separability between classes. The scope of this work assumes the drift in electrical behavior is gradual, analogous to incremental concept drift [9].

When there is unpredictable concept drift in an uninterpretable feature space, little can be done to characterize this unknown drift. For example, if a class drifts inconsistently in either shape or direction, it will be difficult to provide a geometric characterization that can strike a balance between being too loose and too compact. If a feature space is physically informed, the drift can be interpreted in terms of the load's operation.

Domain knowledge about how loads drift in a physically informed feature space is necessary to guide the characterization of this drift for separability testing. Two shipboard loads are presented as examples for characterizing load behavior, a fuel oil purifier (FOP) centrifugal motor and a controllable pitch propeller (CPP) pump. For illustrative purposes, consider steady-state real power (P_{ss}) and steady-state reactive power (Q_{ss}), where the steady-state power was extracted from load on-events, calculated for 0.5 seconds after each event. In a Q_{ss} versus P_{ss} feature space, the drift is likely to occur along a single major axis. Fig. 3 shows the normalized histograms of the FOP centrifugal motor's P_{ss} for five different periods

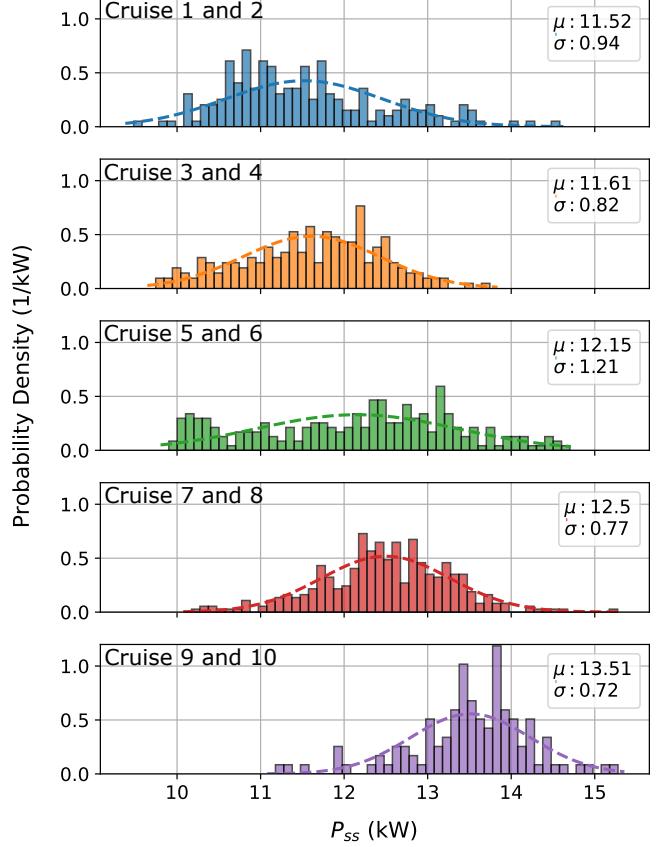


Fig. 3: Normalized histograms of the FOP centrifugal motor's steady-state real power (P_{ss}) for two cruises at a time. Gaussians are fitted with labeled means and standard deviations.

in time. The data is divided into “cruises,” where each cruise represents a period of one to three months in which the vessel is mostly underway at-sea. There is a break in data collection during the vessel’s in-port periods after each cruise. Each plot shows two cruises of data, e.g., the first row is the first and second cruise, the second row is the third and fourth cruise, and so on. A Gaussian probability density function was fit to each dataset with the mean (denoted as μ) and standard deviation (denoted as σ) from the data, as shown in the labels on the plots. The mean of the FOP centrifugal motor's steady-state real power increases as time progresses. Although this example shows only one dimension having a Gaussian distribution, a similar trend can be seen in the steady-state reactive power. In Fig. 4, the Gaussian-like distribution of P_{ss} and Q_{ss} over the entire dataset of ten cruises can be observed for the FOP centrifugal motor. The increase in power can likely be attributed to load aging and wear.

A plot of normalized histograms over five periods in time for the CPP pump is shown in Fig. 5. These are the same time periods as for the FOP centrifugal motor histograms in Fig. 3. It can be observed that unlike the FOP centrifugal motor for which P_{ss} generally trends in one direction over time, the CPP pump's power first drifts higher before drifting back to a similar power level as its original state. Intuition

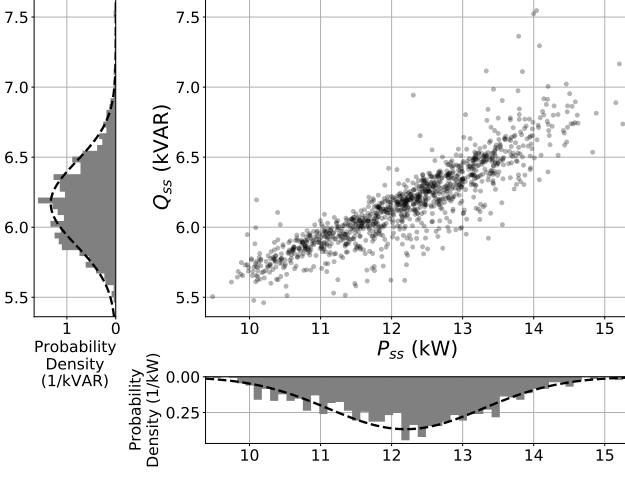


Fig. 4: Q_{ss} versus P_{ss} for the FOP centrifugal motor after ten cruises of data plotted with the normalized probability density for each axis.

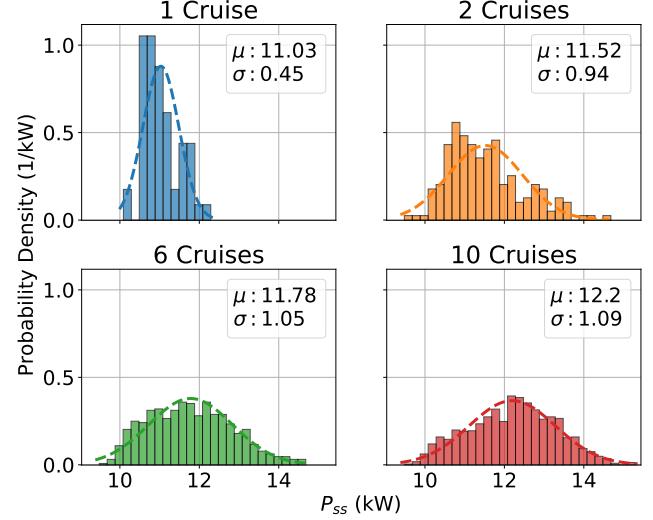


Fig. 6: Normalized histograms of P_{ss} for the FOP centrifugal motor at various points in time.

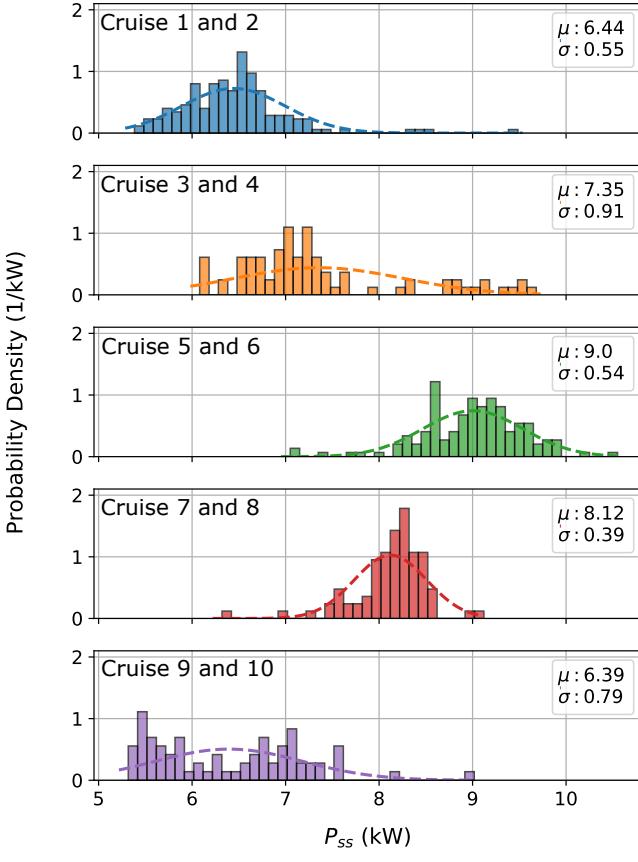


Fig. 5: Normalized histograms of the CPP pump's steady-state real power (P_{ss}) for two cruises at a time. Gaussians are fitted with labeled means and standard deviations.

for the drift of the CPP pump is provided by examining the change in operating condition. The monitored pump is part of the CPP system, which is operated underway to provide the vessel greater maneuverability. The CPP pump is an electric

hydraulic pump, supplementing a separate gear driven pump to provide pressurized hydraulic oil to the CPP system and maintain hydraulic control pressure at the propeller. The drift in the pump's steady-state real power and reactive power consumption over the monitored four-year period correlates with the operating fluid pressure normalized by temperature, as obtained from the ship's logs. There was a large increase in both the power draw and the normalized operating pressure in February 2018, after the replacement of the hydraulic control valves. Then, as the normalized operating pressure decreased over time, the power draw also subsequently decreased.

These two examples demonstrate that with a limited dataset in time it is difficult to determine how much of the load's total distribution is accounted for. Accordingly, the NILM feature selection process should take all known data into account when evaluating the suitability of a feature space. For instance, Fig. 6 and Fig. 7 show the normalized histograms for the FOP centrifugal motor and CPP pump, respectively, after various numbers of cruises (e.g., 1 cruise contains the first cruise of data and 10 cruises contains all 10 cruises of data). For the FOP centrifugal motor, the variance generally increases as the number of cruises increases, due to the load drift. As operating conditions change or faults occur and are fixed, a load can return to a previously observed region in the feature space. For the CPP pump, the variance increases through the fifth cruise due to load drift, after which it slightly decreases as the load drifts back to its original state. Although not all loads appear perfectly Gaussian, a Gaussian approximation still can provide a sufficient characterization while being easy to compute and describe. With this assumption of load characterization, the following section presents the hyperellipsoid characterization for checking load separability.

III. FEATURE SPACE EVALUATION

The existence of concept drift in load behavior leads to the question: how can feature spaces be evaluated for their ability

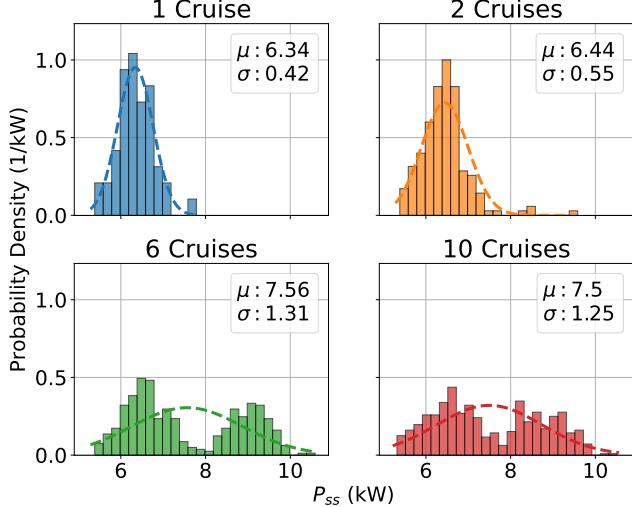


Fig. 7: Normalized histograms of P_{ss} for the CPP pump at various points in time.

to provide separability as loads undergo faults or changing operating conditions? With an understanding of load behavior based on prior observations and knowledge of specific operating patterns, a preliminary set of features can be selected. However, prior to selecting and training a classifier and establishing a diagnostic process within this feature space, the selected features should be evaluated. Even if data is linearly separable, the proposed separability check will identify loads that may become ambiguous in the future if drift occurs. This can inform feature space modification. The separability check is given as an overlap test of geometric characterization regions for each load. An ideal characterization region must be able to represent trends in the load data and anticipated drift, but still be physically reasonable and fairly compact to the region occupied by the load events. Finally, there must be a well-defined and computationally tractable test for overlap between two characterization regions. For illustrative purposes in this section, this method is shown with two-dimensional feature spaces of steady-state real power (P_{ss}) and steady-state reactive power (Q_{ss}). However, this method can be used to assess feature spaces of arbitrary dimension [27].

The features presented in the examples in this section use units of power. However, for features that do not have common units, it is necessary to apply some form of feature scaling such as standardization or normalization [28]. When using the separability check, it is imperative that the dataset in consideration be correctly labeled, such that even points that appear to be outliers can be assumed to represent actual load behavior and not incorrect labeling.

A. Separability Check

For loads that do not drift in the feature space, such as a heater, their distributions in the feature space typically can be characterized as multivariate Gaussians with constant mean and variance. The probability density function (PDF) of

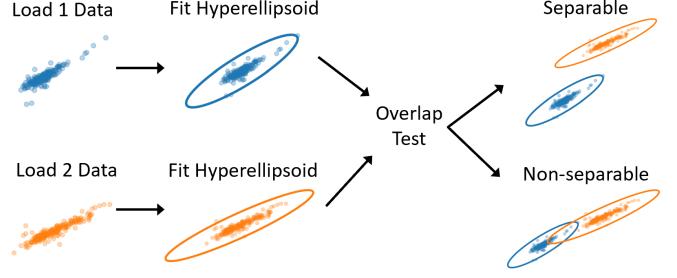


Fig. 8: Diagram illustrating load separability check.

a multivariate Gaussian distribution with an N -dimensional random vector \mathbf{X} , is given as:

$$f_{\mathbf{X}}(x_1, \dots, x_N) = \frac{1}{\sqrt{(2\pi)^N |\Sigma|}} \cdot \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right), \quad (1)$$

where Σ is the covariance matrix of \mathbf{X} , $|\Sigma|$ is the determinant of the covariance matrix, and $\boldsymbol{\mu}$ is the N -dimensional mean vector of \mathbf{X} [29]. The covariance matrix and mean can be approximated with the covariance and mean calculated from collected data. When a load exhibits drift, it often resembles a multivariate Gaussian distribution with non-constant mean and covariance. That is, for some subset of data in time, the data can be approximated as Gaussian. However, over time the distribution may change. With this assumption, the load separability check is performed by first fitting an N -dimensional hyperellipsoid to each load's data. N -dimensional hyperellipsoids are able to represent variance in several orthogonal axes and are the equiprobability contours of a multivariate Gaussian distribution. A hyperellipsoid region of arbitrary dimension can be represented by:

$$(\mathbf{x} - \mathbf{m})^T \mathbf{E}(\mathbf{x} - \mathbf{m}) \leq 1 \quad (2)$$

where \mathbf{m} is an $N \times 1$ vector representing the centroid of the hyperellipsoid and \mathbf{E} is a real symmetric positive-definite $N \times N$ matrix representing the shape and orientation of the hyperellipsoid. Any point \mathbf{x} , that satisfies the inequality in Eq. (2) is either inside or on the surface of the hyperellipsoid [30].

Principal component analysis (PCA) is used to generate the hyperellipsoid regions. Although other techniques for fitting hyperellipsoids exist, PCA is used here since the new set of axes and variances can be understood as fitting an N -variate Gaussian distribution to the data. PCA expresses the most variance in the data possible in each generated principal component axis [31]. The variance (denoted as σ^2) of the data in each of the principal component axes is computed. The standard deviation (denoted as σ) of the data in each principal component axis is obtained by taking the positive square root of these variances. PCA can be computed using eigendecomposition of the covariance matrix or with singular value decomposition [31]. A component matrix is yielded with rows equal to the principal component axes, which can be used as a "transformation matrix" from the feature axes to the principal component axes. The N -dimensional hyperellipsoid's radii are obtained by multiplying the principal

component standard deviations, sorted in descending order, by a user-defined parameter, α . The choice of α is described below. The hyperellipsoid's centroid is obtained by taking the mean of the data points in the feature space. Finally, the hyperellipsoid's rotation is represented using the transformation matrix obtained from PCA. Hyperellipsoidal regions are created in this way for each load in a chosen feature space. The physics of load drift is embedded in these regions in the size and rotation of the hyperellipsoids, obtained using PCA. These regions are then subjected to a test for overlap, as will be later described. Overlapping regions indicate that a feature space that will not be robust at separating the given loads, given load drift. Fig. 8 summarizes the separability check process.

One strategy for choosing the hyperellipsoid size involves taking a constant number of standard deviations in each principal component direction. For example, three standard deviations will cover 99.7% of all hypothetical data, assuming a normal distribution. A disadvantage of this approach, however, is that it is not responsive to the number of load events obtained. Instead, the number of standard deviations for each hyperellipsoid, denoted as α , should be selected based on the amount of confidence in the underlying class distribution. As the number of load events M for a class increases, the confidence that the data is representative of the underlying distribution also generally increases. There is more uncertainty in potential load drift for smaller M . By adjusting α based on class size, this method is able to handle class sizes that are different between loads. Classes that do not have many load events have more uncertainty in their underlying distribution and as a result are assigned a looser hyperellipsoid. The proposed method for selecting α adds an exponentially decaying term to shrink the region size as more load events are collected:

$$\alpha = 3 \left(1 + e^{-M/K} \right), \quad (3)$$

where K is a tunable parameter. For any choice of K , the number of standard deviations α , is bounded between three and six. When M is zero, then α is six. When M approaches infinity, the exponential term goes to zero and α approaches three. When α is close to three there is high confidence that the load has been well-characterized and will no longer drift in the feature space.

By adjusting the parameter K , the rate at which the exponential decays can be adjusted. As K increases, the rate of decay of the exponential decreases, and α will converge to three slower. One consideration for the choice of K is the availability of extra features. If there are abundant features that can be added to the feature space, a larger hyperellipsoid (and thus larger K) can be used in order to give a greater chance of detecting problematic drift. There will also be a greater chance of overlapping regions of non-drifting loads. If additional features are easily obtained, these overlaps can be resolved by adding them to the feature space. However, if additional useful features are not available or are expensive to obtain, overlap detection from non-drifting loads should be avoided in order to avoid unnecessarily adding features. Thus, smaller hyperellipsoids (and thus smaller K) should be

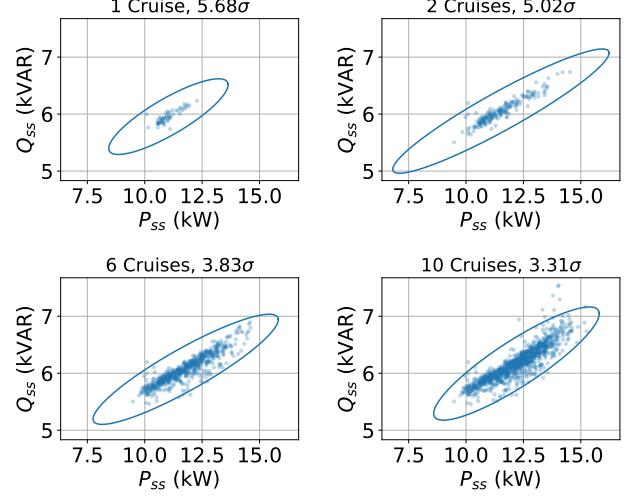


Fig. 9: Hyperellipsoids in the Q_{ss} versus P_{ss} feature space for the FOP centrifugal motor at various points in time.

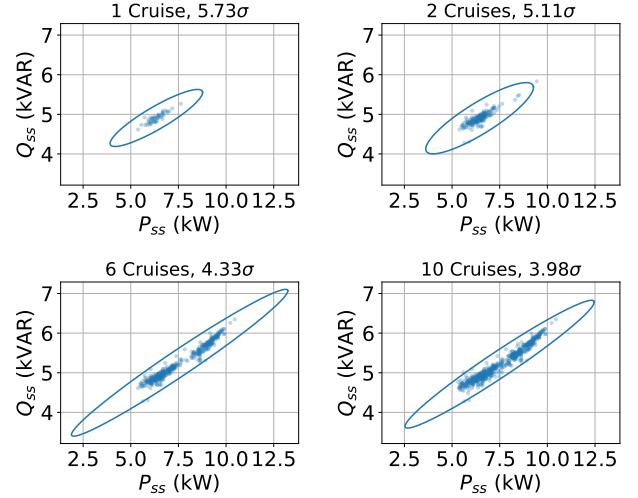


Fig. 10: Hyperellipsoids in the Q_{ss} versus P_{ss} feature space for the CPP pump at various points in time.

used. With smaller hyperellipsoids, problematic drift will be detected later than with larger hyperellipsoids.

The two presented shipboard loads from Section II are used as an example. Using $K = 500$ in the Q_{ss} versus P_{ss} feature space, the fuel oil purifier (FOP) centrifugal motor and controllable pitch propeller (CPP) pump data points and hyperellipsoids are shown in Fig. 9 and Fig. 10, respectively. The hyperellipsoids are drawn after various number of cruises to show the evolution of the hyperellipsoid characterization regions as the number of load events increases and as α subsequently decreases. The FOP centrifugal motor starts with 57 labeled load events after the first cruise. Its hyperellipsoid region is created with 5.68 standard deviations (noted as 5.68σ in the figure), which is on the upper end of possible values for α . After ten cruises there are 1131 load events. The new hyperellipsoid region uses 3.31 standard deviations, which is now on the lower end of possible values for α . For the

CPP pump, after one and ten cruises, there were 48 and 561 labeled load events, respectively. Thus, the hyperellipsoid regions were created with 5.73 and 3.98 standard deviations, respectively. For both loads, the rotation of the hyperellipsoid is approximately the same after one cruise as it is after ten cruises, meaning the drift has continued along approximately the same major axis.

B. Test for Overlap

To check for overlap of the hyperellipsoid regions, the approach of [32] is used, in which the overlap condition is developed as a root-counting problem of a convex polynomial $K(\lambda)$. This polynomial represents the hypothetical circumscribing hyperellipsoid appearing in the region between the two hyperellipsoids under consideration. $K(\lambda)$ is generated using the shape and center of both hyperellipsoids, where \mathbf{E} and \mathbf{m} from Eq. (2) represent the shape and center, respectively. For two hyperellipsoids with shapes \mathbf{A} and \mathbf{B} , and centers \mathbf{c} and \mathbf{d} , respectively, every point, \mathbf{x} , in the union of both hyperellipsoids must satisfy

$$\lambda(\mathbf{x} - \mathbf{c})^T \mathbf{A}(\mathbf{x} - \mathbf{c}) + (1 - \lambda)(\mathbf{x} - \mathbf{d})^T \mathbf{B}(\mathbf{x} - \mathbf{d}) \leq 1, \quad (4)$$

where $\lambda \in [0, 1]$. This inequality is then transformed into the following representation of a hypothetical hyperellipsoid with shape \mathbf{E}_λ and center \mathbf{m}_λ that circumscribes the region of intersection of the two hyperellipsoids:

$$(\mathbf{x} - \mathbf{m}_\lambda)^T \mathbf{E}_\lambda (\mathbf{x} - \mathbf{m}_\lambda) \leq K(\lambda), \quad (5)$$

where

$$\mathbf{E}_\lambda = \lambda \mathbf{A} + (1 - \lambda) \mathbf{B}. \quad (6)$$

The convex polynomial $K(\lambda)$ is given as:

$$K(\lambda) = 1 - (\mathbf{d} - \mathbf{c})^T \left(\frac{1}{1 - \lambda} \mathbf{B}^{-1} + \frac{1}{\lambda} \mathbf{A}^{-1} \right)^{-1} (\mathbf{d} - \mathbf{c}). \quad (7)$$

A necessary and sufficient condition for the two hyperellipsoids to not overlap or touch is for there to be a value of λ in $(0, 1)$ such that $K(\lambda) < 0$. Since $K(\lambda)$ is convex, all that is required is to count the roots of $K(\lambda)$ on $(0, 1)$. Rather than using symbolic math to solve for the roots of $K(\lambda)$, [32] presents an algorithm to compute $P(\lambda) = \det(\mathbf{E}_\lambda) \cdot K(\lambda)$, which is also convex and has the same roots in $(0, 1)$ as $K(\lambda)$. Using Sturm's theorem as described in [33], the number of roots of this polynomial in $(0, 1)$ can be efficiently computed. Sturm's theorem is used to generate a "Sturm sequence" of polynomials using polynomial differentiation and division. By counting the number of sign alternations in these sequences, the number of polynomial roots in the given interval can be computed. If there are two roots, the hyperellipsoids do not overlap and the two loads are considered "separable." Otherwise, the hyperellipsoids either are tangent or overlap, and the two loads are considered "non-separable."

This method works as long as $\mathbf{A} - \mathbf{B}$ is invertible. It is assumed that this will be the case since the radii of two given hyperellipsoids are very unlikely to be exactly the same. With this overlap test defined for two hyperellipsoids, several load characterizations can be checked for overlap two at a time.

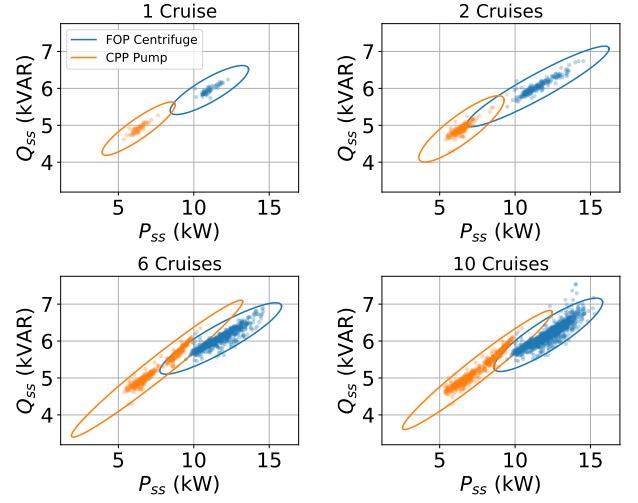


Fig. 11: Hyperellipsoids in the Q_{ss} versus P_{ss} feature space for the FOP centrifugal motor and CPP pump at various points in time.

TABLE I: Shipboard loads in dataset.

Shipboard System	Load
<i>Fuel Oil Purifier (FOP) System</i>	FOP centrifugal motor FOP feed pump
<i>Main Propulsion Diesel Engine (MPDE) System</i>	MPDE jacket water (JW) heater MPDE lube oil (LO) heater MPDE prelube (PL) pump
<i>Ship Service Diesel Generator (SSDG) System</i>	SSDG jacket water (JW) heater
<i>Additional Engine Room Loads</i>	Controllable pitch propeller (CPP) pump Graywater pump Bilge and ballast pump

Detecting overlap between a set of loads will require at most $L(L - 1)/2$ overlap checks, where L is the number of loads. An advantage of the hyperellipsoid region is that its overlap check does not depend on the number of load observations. That is, the overlap test is computed with the N -dimensional hyperellipsoid parameters, not with individual points. As a result, the computational power required for this check does not increase as the number of load events increases.

Fig. 11 shows the hyperellipsoid characterization regions for the FOP centrifugal motor and CPP pump from Fig. 9 and Fig. 10, respectively. After one cruise, even though the two loads are clearly linearly separable, the hyperellipsoid characterization regions are overlapping. The FOP centrifugal motor and CPP pump data points have high enough variability for the separability check to anticipate future misclassification of the CPP pump and FOP centrifugal motor. The hyperellipsoid regions continue to overlap for subsequent cruises.

IV. METHOD DEMONSTRATION

To validate the utility of the separability check on a larger dataset of loads, nine loads from a USCG vessel are presented. The nine evaluated loads are listed in Table I. A dataset was assembled from the first two cruises of the vessel after installation of the NILM, from August 2016 to March 2017. The hyperellipsoid boundaries are plotted in Fig. 12 and

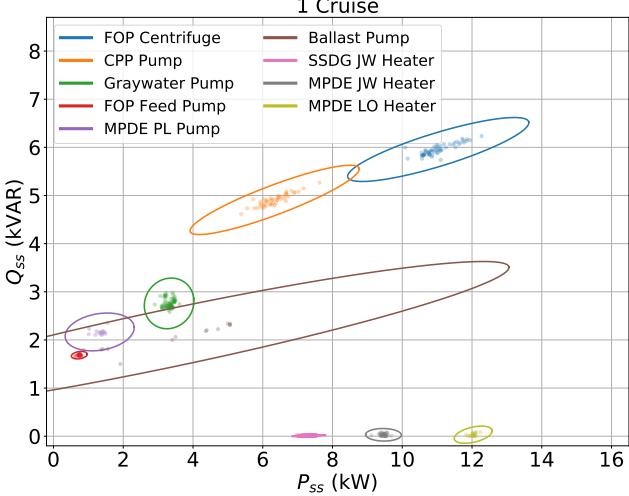


Fig. 12: Hyperellipsoids in the Q_{ss} versus P_{ss} feature space for nine loads after one cruise of data collection.

TABLE II: Table of pairwise load separability in Q_{ss} versus P_{ss} feature space after two cruises of data collection. “S” represents separable pairs of loads; “N” represents non-separable pairs of loads.

Load	Ballast	FOP Feed	MPDE PL	Graywater	CPP
FOP Cent	S	S	S	S	N
CPP	S	S	S	S	
Graywater	N	S	N		
MPDE PL	N	S			
FOP Feed	N				

Fig. 13 for the nine loads after the first cruise and second cruise of data collection, respectively. After only one cruise, the CPP pump overlaps with the FOP centrifugal motor region, as was previously described. Also after only one cruise, the bilge and ballast pump region overlaps with the graywater pump, MPDE prelube pump, and FOP feed pump regions. This shows that the output of a given classifier is highly unpredictable for these four loads. The drifts of the MPDE prelube pump and graywater pump are not significant enough after one cruise to fail the overlap test. However, after two cruises the ensuing load drift results in the regions overlapping. The three heater loads have less variance in Q_{ss} and P_{ss} , and thus have tighter hyperellipsoids (hence why the SSDG JW’s hyperellipsoid is barely visible at the plot’s zoom level). The hyperellipsoids for these heater loads do not overlap with any other loads, indicating that the feature space provides sufficient separability for these three loads. Table II shows the separability check results for each pair of loads in the Q_{ss} versus P_{ss} feature space after two cruises of data collection. Only the six loads that were non-separable with at least one other load are included.

Two classifiers were trained with the dataset from the first two cruises to validate the problematic overlap identified by the separability check. A linear support vector machine (SVM) and deep neural network (DNN) were chosen as examples to demonstrate how concept drift can cause performance degradation for both linear and non-linear classifiers. Different

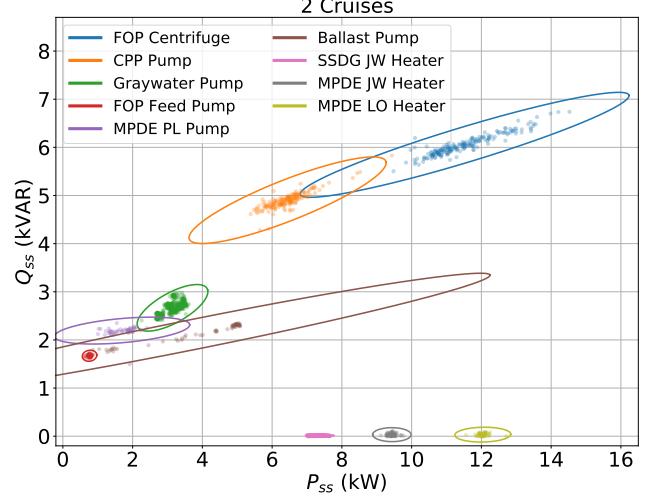


Fig. 13: Hyperellipsoids in the Q_{ss} versus P_{ss} feature space for nine loads after two cruises of data collection.

classifiers may have slightly different results. The data was split into 80% training and 20% validation, with data stratification to allocate load events evenly based on class frequency. The trained models had perfect classification accuracy for all load events in the validation dataset (which is a subset of the data from the first two cruises) for both classifiers. That is, without any consideration for load drift, this would appear to be a sufficient feature space. However, when the model was then tested on the data collected on the USCG vessel for the eight later cruises (from April 2017 to September 2020), drift of several loads in the feature space resulted in significantly reduced accuracy. Fig. 14 shows the decision boundaries (from the models trained on the first two cruises) and the data points from the last eight cruises. Fig. 15 shows the normalized confusion matrices for all loads that did not have an perfect accuracy, where for the DNN the training and testing were performed ten times and results averaged. There are many separators that can accurately classify the training data, such as the decision boundaries shown in Fig. 14. However, these decision boundaries assume a static dataset. Thus, when the loads drift, the misclassification they cause is similar for both classifiers, as shown by the similar confusion matrices. For both classifiers, as anticipated by the separability check, the CPP pump is often misclassified as the FOP centrifugal motor, the MPDE prelube pump is often classified as the graywater pump, and the bilge and ballast pump is often misclassified as the FOP feed pump. Also as expected from the separability check, the three loads that do not show any degradation in classification performance are the SSDG JW heater, MPDE JW heater, and MPDE LO heater.

After two cruises of data, although the trained classifiers showed that the data was classifiable at the time, the classifiers were not reliable as more data was collected. The problematic drift is anticipated by the separability check. In fact, the hyperellipsoids for two cruises of data and ten cruises of data are relatively similar even with further load drift and with a

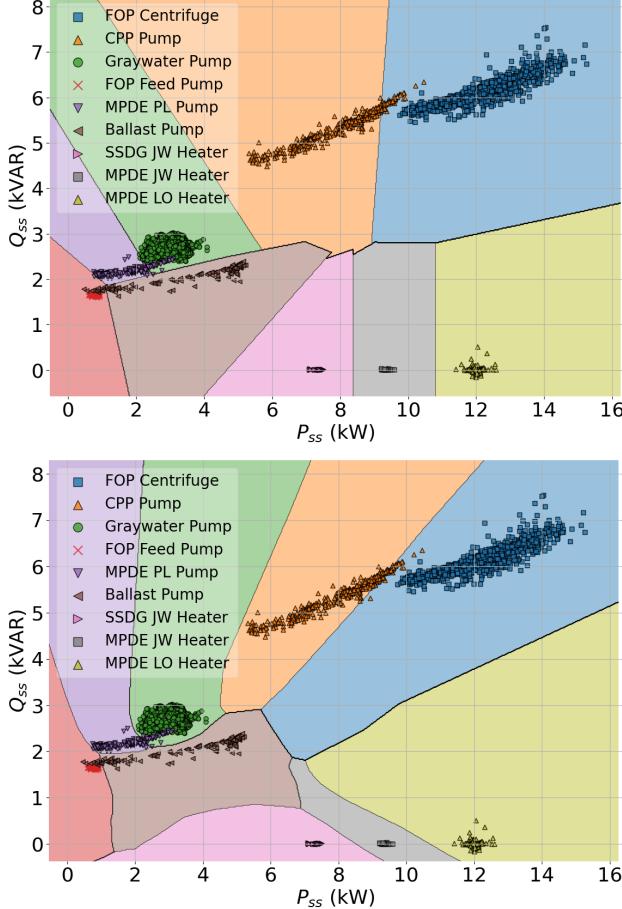


Fig. 14: Q_{ss} versus P_{ss} boundaries, trained with a linear SVM (top) and DNN (bottom) using the first two cruises of data collection. The data points for eight later cruises are plotted.

		SVM						DNN					
		FOP Cent	CPP	Graywater	FOP Feed	MPDE PL	Ballast	FOP Cent	CPP	Graywater	FOP Feed	MPDE PL	Ballast
Actual	FOP Cent	1	0	0	0	0	0	1	0	0	0	0	0
	CPP	.158	.842	0	0	0	0	.137	.863	0	0	0	0
	Graywater	0	0	1	0	0	0	0	0	1	0	0	0
	FOP Feed	0	0	0	1	0	0	0	0	0	1	0	0
	MPDE PL	0	0	.151	.005	.844	0	0	0	.156	.015	.826	.003
	Ballast	0	0	0	.099	0	.901	0	0	0	.082	.001	.917

Fig. 15: Normalized confusion matrices for linear SVM and DNN classifiers trained on the first two cruises of data and tested on the eight later cruises using a two-dimensional feature space.

large difference in number of load events. For example, Fig. 16 shows the hyperellipsoids (as dashed lines) and data points after ten cruises of data for the six loads that did not have perfect accuracy. Also plotted are the hyperellipsoids (in solid lines) after two cruises of data, which were previously shown in Fig. 13. As shown, the drift is relatively well characterized

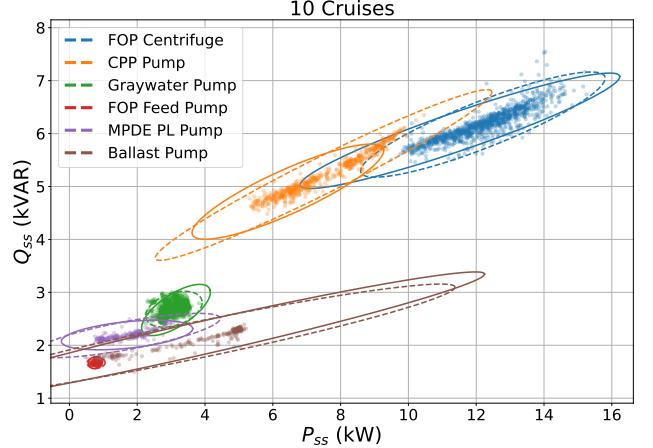


Fig. 16: Hyperellipsoids (dashed) and data points in the Q_{ss} versus P_{ss} feature space for six loads after ten cruises of data collection. Hyperellipsoids (solid) are also drawn from the first two cruises of data.

after only two cruises since the hyperellipsoids are drawn with a higher number of standard deviations for fewer load events.

Due to the identification of the problematic drift, additional features should be added to the feature space. To demonstrate, the maximum apparent power at inrush (S_{peak}) and the transient time are added to the feature space, since they are not redundant with steady-state power. The overlap test is run after two cruises of data collection using this four-dimensional feature space (P_{ss} , Q_{ss} , S_{peak} , transient time). Since the features do not have common units, min-max normalization is applied, such that for each feature axis the range of data is transformed into [0,1]. This scaling is applied for both the overlap test and for training classifiers. Only the CPP pump and FOP centrifugal motor hyperellipsoid regions are overlapping in the separability check, indicating that this is likely a more suitable feature space than the demonstrated two-dimensional feature space. Both a DNN and SVM (using a radial basis function kernel) were trained and tested to validate the reduced problematic overlap. Similar to before, the first two cruises of data were split into 80% training and 20% validation, with data stratification. The model was then tested on the eight later cruises. For the DNN the training and testing were performed ten times and the results averaged. The confusion matrices are shown in Fig. 17, for the same six loads that are shown in Fig. 15. The results have significantly improved, as predicted by the separability check.

V. CONCLUSION

The hyperellipsoid region shown is only one example of possible regions that can be fitted to available load data for the purposes of gauging the separability of a particular feature space. In different domains with different types of drift, alternate feature space characterizations could be conceived, such as N -dimensional rectangular prisms or single N -dimensional hyperspheres, as long as a method exists to test for overlap of the geometric regions that are produced.

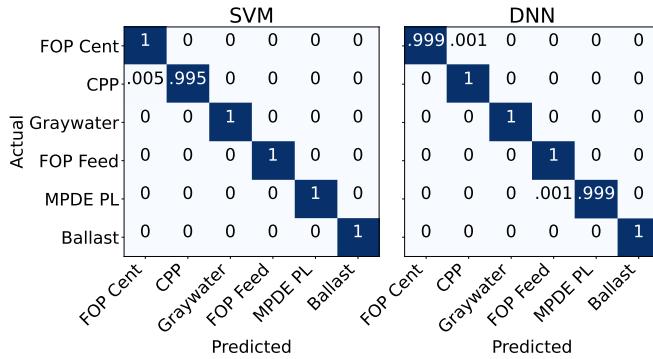


Fig. 17: Normalized confusion matrices for radial basis function kernel SVM and DNN classifiers trained on the first two cruises of data and tested on the eight later cruises using a four-dimensional feature space.

Electromechanical load behavior may change due to normal variation, such as inherent mechanical variability or a change in operating conditions. Changes in load behavior could also be indicative of an underlying fault or degradation of various internal mechanisms that may require repair or replacement. For a nonintrusive load monitoring classifier, this manifests as concept drift that can reduce classification performance on future data or even make the load completely indistinguishable. Understanding the physics and time-dependency behind changing load behavior can inform feature space selection improvements, enabling the applicability of nonintrusive monitoring for equipment health monitoring and diagnostics. The presented separability check ensures the selection of a physically informed feature space that allows for load disaggregation, even when loads drift over time. This is especially important for isolated, microgrid, and generation-constrained systems, where system reliability is critical.

VI. ACKNOWLEDGMENTS

The authors gratefully acknowledge the support and dedication of the US Coast Guard and in particular the crew of USCGC SPENCER for granting access to their ship.

REFERENCES

- [1] A. M. da Silva, R. J. Povinelli, and N. A. O. Demerdash, "Rotor bar fault monitoring method based on analysis of air-gap torques of induction motors," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2274–2283, 2013.
- [2] J. Zhang, X. Chen, W. W. Y. Ng, C. S. Lai, and L. L. Lai, "New appliance detection for nonintrusive load monitoring," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4819–4829, 2019.
- [3] H. M. Hashemian and W. C. Bean, "State-of-the-art predictive maintenance techniques," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 10, pp. 3480–3492, Oct 2011.
- [4] J. Paris, J. S. Donnal, and S. B. Leeb, "NilmDB: The non-intrusive load monitor database," *IEEE Transactions on Smart Grid*, vol. 5, no. 5, pp. 2459–2467, 2014.
- [5] M. Kaselimi, N. Doulamis, A. Voulodimos, E. Protopapadakis, and A. Doulamis, "Context aware energy disaggregation using adaptive bidirectional LSTM models," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3054–3067, 2020.
- [6] B. Celik and J. Vanschoren, "Adaptation strategies for automated machine learning on evolving data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 9, pp. 3067–3078, 2021.
- [7] D. H. Green, A. W. Langham, R. A. Agustin, D. W. Quinn, and S. B. Leeb, "Adaptation for automated drift detection in electromechanical machine monitoring," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2022.
- [8] W. W. Y. Ng, J. Zhang, C. S. Lai, W. Pedrycz, L. L. Lai, and X. Wang, "Cost-sensitive weighting and imbalance-reversed bagging for streaming imbalanced and concept drifting in electricity pricing classification," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1588–1597, 2019.
- [9] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under concept drift: A review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 12, pp. 2346–2363, 2019.
- [10] J. Z. Kolter and M. J. Johnson, "REDD: A public data set for energy disaggregation research," in *Proceedings of the SustKDD workshop on Data Mining Applications in Sustainability*, 2011.
- [11] J. Kelly and W. Knottenbelt, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Scientific Data*, vol. 2, no. 150007, 2015.
- [12] B. Cannas, S. Carcangiu, D. Carta, A. Fanni, and C. Muscas, "Selection of features based on electric power quantities for non-intrusive load monitoring," *Applied Sciences*, vol. 11, no. 2, 2021.
- [13] S. Bao, L. Zhang, X. Han, W. Li, D. Sun, Y. Ren, N. Liu, M. Yang, and B. Zhang, "Feature selection method for nonintrusive load monitoring with balanced redundancy and relevancy," *IEEE Transactions on Industry Applications*, vol. 58, no. 1, pp. 163–172, 2022.
- [14] R. Chen, C. Dewi, S. Huang, and R. Caraka, "Selecting critical features for data classification based on machine learning methods," *Journal Of Big Data*, vol. 7, p. 26, 07 2020.
- [15] K. Mao, "RBF neural network center selection based on Fisher ratio class separability measure," *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 1211–1217, 2002.
- [16] L. Wang, "Feature selection with kernel class separability," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 9, pp. 1534–1546, 2008.
- [17] V. Singh and N. K. Verma, "Intelligent condition-based monitoring techniques for bearing fault diagnosis," *IEEE Sensors Journal*, vol. 21, no. 14, pp. 15448–15457, 2021.
- [18] J. Li, C. Shen, L. Kong, D. Wang, M. Xia, and Z. Zhu, "A new adversarial domain generalization network based on class boundary feature detection for bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–9, 2022.
- [19] T. Han, Y.-F. Li, and M. Qian, "A hybrid generalization network for intelligent fault diagnosis of rotating machinery under unseen working conditions," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.
- [20] A. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.
- [21] J. Paris, J. S. Donnal, Z. Remscrim, S. B. Leeb, and S. R. Shaw, "The sinefit spectral envelope preprocessor," *IEEE Sensors Journal*, vol. 14, no. 12, pp. 4385–4394, 2014.
- [22] W. Wichakool, Z. Remscrim, U. A. Orji, and S. B. Leeb, "Smart Metering of Variable Power Loads," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 189–198, 2015.
- [23] L. Du, D. He, R. G. Harley, and T. G. Habetler, "Electric load classification by binary voltage-current trajectory mapping," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 358–365, 2016.
- [24] H. Rashid, P. Singh, V. Stankovic, and L. Stankovic, "Can non-intrusive load monitoring be used for identifying an appliance's anomalous behaviour?" *Applied Energy*, vol. 238, pp. 796 – 805, 2019.
- [25] P. A. Lindahl, D. H. Green, G. Bredariol, A. Aboulian, J. S. Donnal, and S. B. Leeb, "Shipboard fault detection through nonintrusive load monitoring: A case study," *IEEE Sensors Journal*, vol. 18, no. 21, pp. 8986–8995, 2018.
- [26] Y.-F. Li and Z.-H. Zhou, "Towards making unlabeled data never hurt," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 1, pp. 175–188, 2015.
- [27] R. Agustin, "A Load Identification and Diagnostic Framework for Aggregate Power Monitoring," Master's thesis, Massachusetts Institute of Technology, 2021.
- [28] M. M. Suarez-Alvarez, D.-T. Pham, M. Y. Prostov, and Y. I. Prostov, "Statistical approach to normalization of feature vectors and clustering of mixed datasets," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 468, no. 2145, pp. 2630–2651, 2012.

- [29] B. Wang, W. Shi, and Z. Miao, "Confidence analysis of standard deviational ellipse and its extension into higher dimensional Euclidean space," *PLOS ONE*, vol. 10, no. 3, pp. 1–17, 03 2015.
- [30] J. M. Shapiro, G. B. Lamont, and G. L. Peterson, "An evolutionary algorithm to generate hyper-ellipsoid detectors for negative selection," in *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation*, ser. GECCO '05. New York, NY, USA: Association for Computing Machinery, 2005, p. 337–344.
- [31] I. T. Jolliffe, *Principal Component Analysis and Factor Analysis*. New York, NY: Springer New York, 1986, pp. 115–128.
- [32] I. Gilitschenski and U. D. Hanebeck, "A robust computational test for overlap of two arbitrary-dimensional ellipsoids in fault-detection of Kalman filters," in *2012 15th International Conference on Information Fusion*, 2012, pp. 396–401.
- [33] ———, "A direct method for checking overlap of two hyperellipsoids," in *2014 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 2014, pp. 1–6.



Steven B. Leeb (Fellow, IEEE) received the Ph.D. degree from the Massachusetts Institute of Technology, in 1993. Since 1993, he has been a member on the MIT Faculty with the Department of Electrical Engineering and Computer Science. 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.



Daisy H. Green (Member, IEEE) received the B.S. degree in electrical engineering from the University of Hawai'i at Mānoa, Honolulu, HI, USA, in 2015, and the M.S. degree and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2018 and 2022, respectively. She is currently a Postdoctoral Associate with the Department of Architecture, MIT. Her research interests include the development of signal processing algorithms for energy management and condition monitoring.



Aaron W. Langham (Graduate Student Member, IEEE) received the B.E.E. degree in electrical engineering from Auburn University, Auburn, AL, USA, in 2018, and the M.S. degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2022, where he is currently pursuing the Ph.D. degree in electrical engineering and computer science. His research interests include signal processing and machine learning for energy management.



Rebecca A. Agustin received the B.S. and M.Eng. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2019 and 2021, respectively.



Devin W. Quinn received the M.S. degree in mechanical engineering from the Massachusetts Institute of Technology in 2022. He was previously stationed as a Damage Control Officer aboard USCGC DILIGENCE and Engineer Officer onboard USCGC ESCANABA. He is currently a Lieutenant Commander with the United States Coast Guard, stationed in California.