An Empirical Investigation of V-I Trajectory based Load Signatures for Non-Intrusive Load Monitoring

T. Hassan, F. Javed and N. Arshad, *Member, IEEE*

Abstract-- **Choice of load signature or feature space is one of the most fundamental design choices for non-intrusive load monitoring or energy disaggregation problem. Electrical power quantities, harmonic load characteristics, canonical transient and steady-state waveforms are some of the typical choices of load signature or load signature basis for current research addressing appliance classification and prediction. This paper expands and evaluates appliance load signatures based on V-I trajectory – the mutual locus of instantaneous voltage and current waveforms – for precision and robustness of prediction in classification algorithms used to disaggregate residential overall energy use and predict constituent appliance profiles. We also demonstrate the use of variants of differential evolution as a novel strategy for selection of optimal load models in context of energy disaggregation. A publicly available benchmark dataset REDD is employed for evaluation purposes. Our experimental evaluations indicate that these load signatures, in conjunction with a number of popular classification algorithms, offer better or generally comparable overall precision of prediction, robustness and reliability against dynamic, noisy and highly similar load signatures with reference to electrical power quantities and harmonic content. Herein, wave-shape features are found to be an effective new basis of classification and prediction for semi-automated energy disaggregation and monitoring.**

Index Terms—**Load signature, feedforward neural networks, load monitoring, optimization, smart grids, supervised learning, support vector machines**

I. INTRODUCTION

Non-Intrusive Load Monitoring (NILM) is the disaggregation of overall demand profile of a household into individual signatures of appliances switched on at a particular instant or within a specified time period. This disaggregation is carried out without using intrusive physical sensors on individual appliances. Aggregate demand profile (ADP) or instantaneous power consumption is typically observed via a single energy meter installed at service entry point of the household. Load signatures of individual appliances refer to metrics that characterize their operating state and temporal behavior. NILM is an important provision for diagnostic feedback in residential and commercial energy feedback systems (REFS and CEFS) and for long-term

diagnostics for settings such as microgrids and distribution sectors, particularly as an enabling technology for demand-side management ([1] - [4]). This study evaluates a new kind of two-dimensional load signature for non-intrusive profiling and identification of residential appliances. Load signatures have been employed effectively as bases for applications in profiling, status monitoring and safety assurance of electrical loads ([2]-[3], [5]-[9]). Our design choice for load signatures is wave-shape features (WS) based on the mutual trajectory of instantaneous voltage and current waveforms. Lam and colleagues originally introduced these load signatures as a new basis for establishing and comparing load taxonomies [5]. In this study, in contrast, we expand and evaluate load characteristics in WS for precision and robustness of prediction in NILM as a multi-class classification problem. Our empirical results illustrate that precision of prediction and robustness against dynamic, noisy and highly similar load signatures in NILM with wave-shape features (WS) is better or generally comparable to that of traditional benchmark load signatures over a variety of classification algorithms. To evaluate our algorithms we have used a publically available dataset for benchmarking i.e. Reference Energy Disaggregation Dataset (REDD) [10].

II. RELATED WORK

Contemporary research on implementation of a NILM system typically addresses the following design choices.

Granularity over time of ADP: Granularity over ADP refers to the rate at which the installed meter is able to observe and report the overall instantaneous power consumption. This design choice is usually associated with the distinction between event-based and non-event based operation. In event-based operation, decision frequency for NILM is the frequency at which appliances in the household change their state ([3], [6]-[7]). This choice is appropriate for real-time diagnostic feedback in residential and commercial energy feedback systems and is considered for this paper. Non-event based NILM operates on aggregated consumption data and is useful for long-term or medium term diagnostics on energy consumption of particular households or of larger settings such as microgrids and distribution sectors.

Definitions of load signatures: As introduced earlier, the definitions of load signatures refer to the metrics used to identify operating characteristics of individual devices connected within the setting in question. Various existing implementations of NILM systems primarily differ in choice

of steady-state or transient, fundamental frequency or harmonic frequency signatures. For instance, Hart [1], Cole and Albicki [8] employ steady-state and transient changes of active real power as basis of classification. In contrast, Laughman et. al. [6] and Lee et. al. [7] employ harmonic magnitudes associated with step changes in overall load for establishing and comparing appliance profiles. Liang and colleagues [3] have demonstrated the use of raw single-cycle current, instantaneous power and admittance waveforms for the purpose. A comprehensive review of signature types evaluated for NILM appears in surveys by Zeifman [9] and Ahmad et. al. [11]. Lam et. al. [5] study typical bases for load taxonomies and introduce wave-shape features (WS) as a competitive new basis using hierarchical clustering. According to the study, WS characterize the shape of instantaneous current demand for a particular device, carry engineering meanings and result in larger relative differentiation between appliances with different operating principles. A prior examination of WS for NILM however, does not exist to best of the authors' knowledge and hence the subject of this paper.

Initial acquisition of load signatures: Hart's canonical work refers to NILM's two distinct modes of operation, with different degrees of intrusiveness [1]. The first uses a one-time calibration period where appliance signatures are manually collected for supervised learning algorithms. The mode is referred to as 'Manual-Setup' (MS-NILM) or semi-automated energy disaggregation and is considered for this paper. Hart [1], Liang et. al. [3] and Cole and Albicki [8] are typical implementations of MS-NILM. The second operational mode for NILM, termed 'Automatic-Setup', uses *a priori* information about expected load characteristics and unsupervised machine learning algorithms to automatically disaggregate ADP. Contributions by Parson et. al. [12] and Kim [13] are recent investigations of unsupervised energy disaggregation. MS-NILM is computationally convenient, however, the intrusive collection and labeling of signatures makes it tedious to set up and adapt to new appliances.

Selection of learning and optimization algorithms: This refers to the parameter search and performance optimization of learning algorithms, for instance, artificial neural network or support vector method employed to learn appliance signatures. This choice of algorithms for disaggregation is also fundamentally relative to both the required response time for NILM system and the choice of load signatures. A wide variety of supervised and unsupervised learning algorithms ([3], [8], [12]-[15]) and optimization strategies such as integer programming and metaheuristics ([16]-[19]) have been observed effective in their capacity for load profiling and disaggregation. This capacity is typically distinct relative to load conditions at the time of operation, for instance, number of simultaneously operating appliances, noise levels, electrical interference from neighboring system, etc. [3].

Our contributions in this paper can be summarized as follows.

1) We expand the set of load characteristics in WS and present a comprehensive empirical evaluation of WS as basis for profiling, and prediction of appliances in manual-setup, event-based non-intrusive load monitoring. We identify and compare (a) the capacity of load disaggregation and (b) its robustness against dynamic and highly similar load signatures. Our experiments also take into account ambient variation in load signatures due to electrical noise and interference typical of household ADP.

2) We select optimal values of adjustable parameters for employed learning algorithms using population-based global search. The algorithm is based on differential evolution (DE) [20], which is a novel technique in the context of NILM.

3) We present an overview of disaggregation capacity of employed learning algorithms as a function of aforementioned conditions, utilizing reference energy disaggregation dataset (REDD) [10] as benchmark dataset for evaluation.

III. DESCRIPTION OF DESIGN METHODOLOGY

A system-level depiction of typical manual-setup NILM is presented in figure 1. 'Manual-setup' refers to a one-time calibration period to learn the appliances signatures and store them in a database. Once the system learns these signatures it is able to identify the appliances in the system whenever a switching event takes place.

There are usually two forms of signatures used in NILM i.e. snapshot-form signatures and delta-form signatures [21]. Snapshot form refers to signatures that are the aggregate power consumption of all appliances as observed via energy meter installed for NILM. On the other hand delta-form signatures express load behavior in brief windows of time containing only a single switching event rather than a large number of events to facilitate event-based NILM. For the scope of this paper only delta-form signatures are considered.
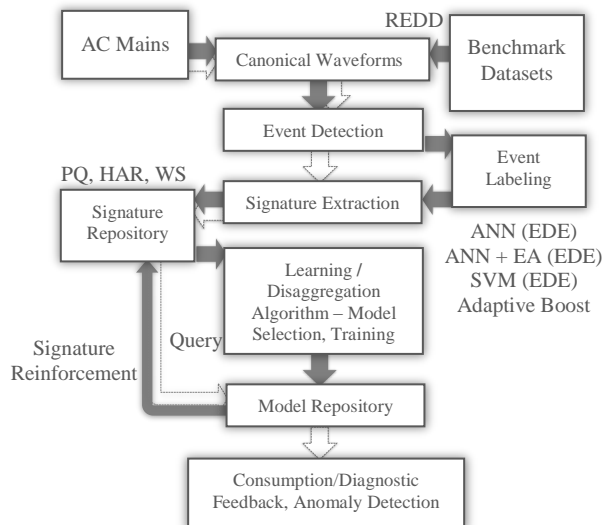

Fig. 1. System-level depiction of manual-setup non-intrusive load disaggregation. Dashed trajectory refers to NILM query; solid refers to NILM training and optimization.

A. Pre-processing Energy Consumption Data

Fig. 1 expresses the need for energy consumption data for the setting where NILM is deployed. For high frequency approaches to NILM such as this study, acquisition of
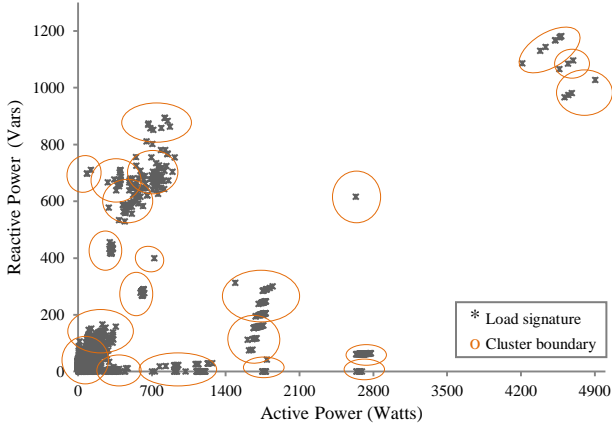
Fig. 2. Consumption values from a subset of switching events from REDD – traditional power metrics (P, Q). Clusters identified via K-means are illustrated on the $P - Q$ plane.

Fig. 3. Harmonic content (HAR) for various appliances.

Fig. 4. V-I trajectories for six different appliances from REDD household # 3.

instantaneous voltage and current waveforms is set up, typically at a rate of more than 100 samples per cycle. This data is used to have the system learn the load signatures of various appliances. Wave-shape features evaluated in the following sections use the mutual trajectory of voltage and current waveforms to deduce information about electrical appliances, hence the high granularity. For the course of this study, benchmark dataset REDD [10] is used as a source of these raw waveforms. Detail on denominations in REDD is mentioned later in the study.

Pre-processing of these raw waveforms includes separation of switching events, which are time instances at which a particular appliance in the household changes state. Fig. 1 expresses this step as 'Event Detection'. Once those instances are determined, cycle-by-cycle snapshots of both voltage and current waveforms are extracted, otherwise known as delta-form signatures [21]. These switching events are 'unlabeled' as they have not been attributed to individual appliances. K-means algorithm outlined as follows, groups these M waveforms in N dimensions into K cohesive groups or 'clusters' with unique appliance labels such that the within-cluster 'distance' is minimized [22]. The step is referred to as 'Event-labeling'.

1) K cluster 'centroids' μ₁, μ₂ ... μₖ $\epsilon \mathbb{R}^N$ are randomly initialized. Waveforms are denoted as $X \epsilon \mathbb{R}^M \times \mathbb{R}^N$.
2) Each of the M waveforms x_i is assigned to cluster k' where $k' : \min_{k'} \|x_i - \mu_{k'}\|^2$.
3) New position of each centroid is calculated by averaging the b waveforms assigned to its corresponding cluster in step 2. For iteration t, $\mu(t+1) = \frac{1}{b}\sum_{j=1}^{b} x_j$.
4) Unless maximum number of iterations is reached, step 2 is repeated.

Number of clusters K is the total number of appliances present in REDD database for a sufficiently large value of M, typically 12000 – 20000 in our evaluation. K-means algorithm is preferred in our evaluation for its relative efficiency compared to hierarchical clustering algorithms [22] employed by Lam et. al. [5] in their canonical introduction to WS.
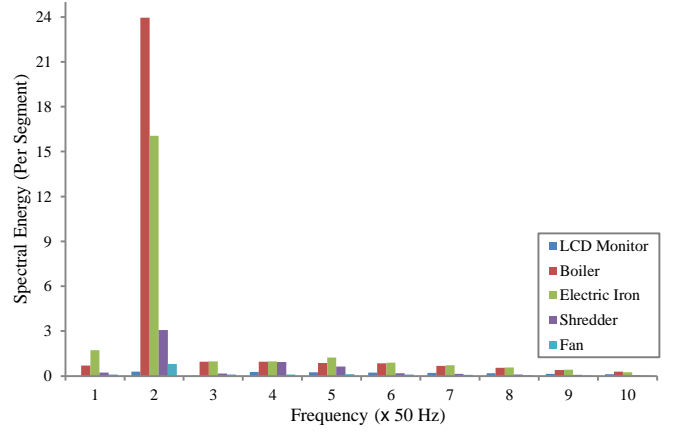
B. Standard Benchmark Load Signatures

Signature extraction follows event-labeling in NILM training phase, as expressed in fig. 1. The first benchmark used in our evaluation is traditional power metrics (PQ) [1]. PQ refers to real and reactive power consumption of a particular appliance (P, Q) and total odd and even harmonic distortion of current waveform (THD_o, THD_e).

The second benchmark used is harmonic content (HAR) of the current waveforms, determined by the spectral energy in contiguous segments of Fourier Transform (FT) of the current waveform. Fig. 2 represents the power consumption of various appliances from one particular household in our example dataset REDD on a $P - Q$ plane for 5000 consecutive switching events. Note the abundance of low-power (<200W) appliance signatures in close proximity. Fig. 3 illustrates HAR consumption values for five common household appliances and the differences in harmonic content thereof.

C. Wave-shape Features (WS)

Lam and colleagues introduced metrics based on voltage and current wave-shape for establishing taxonomy of appliance signatures [5]. Lam's work refers to the mutual locus of instantaneous voltage and current waveform as the V-I trajectory. Fig. 4 illustrates the shape of V-I trajectory for six different appliances from one of REDD households. For
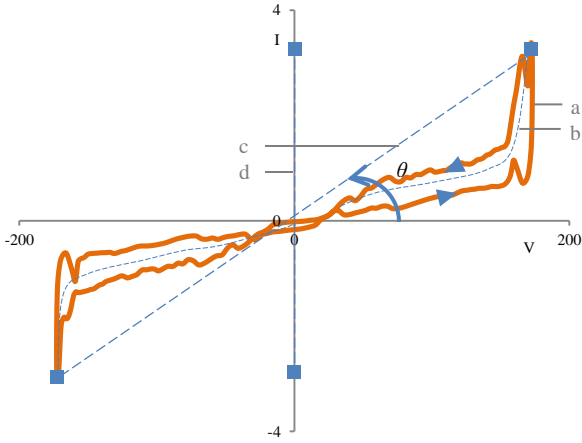
Fig. 5. A graphical illustration of wave-shape metrics: (*a*) V-I trajectory (*b*) Mean curve (*c*) Reference line joining points of highest and lowest I-coordinate in the V-I plane.

LS	areaRL, P_{rms}	curveML, TeHD	numIntersec, ToHD	numIntersec, areaEnc	span, P_{rms}
Coeff.	0.5096	0.6038	0.6861	-0.5215	0.96

appliances with different working principles (for instance, resistive, motor-driven or power-electronic), the V-I trajectory exhibits unique characteristics captured by wave-shape (WS) features. Since operating state of a particular appliance corresponds to shape of instantaneous current waveform of that particular appliance, hierarchical clustering with WS results in larger separation between characteristically dissimilar appliances compared to PQ [5]. An extrapolation of this observation is to verify that WS would allow a multi-class classifier to generalize better to unknown examples. This would prove the effectiveness of WS for load disaggregation over a variety of off-the-shelf learning algorithms and benchmark load signatures.

WS are briefly reviewed as follows; detailed explanation appears in [5].

Looping Direction: 'Looping direction' refers to the anti-clockwise or clockwise curvature of V-I trajectory. It corresponds to the sign of the phase angle difference between voltage and current waveforms, represented in fig. 5 by the direction of curvature of locus *a*. A clockwise curvature refers to an overall capacitive load behavior (current leads voltage in phase) while a counter-clockwise curvature refers to an overall inductive load characteristic (voltage leads current in phase).

Area Enclosed: 'Area Enclosed' refers to the area enclosed by the boundary of the V-I trajectory. It is proportional to the magnitude of phase difference between voltage and current waveforms. Area enclosed by line segment *a* in fig. 5, represents this metric.

Non-linearity of Mean Curve: Mean curve of the V-I trajectory, represented in fig. 5 by line segment *b*, divides it into two identical halves. Degree of distortion of the mean-curve from a straight line is indicative of the non-linearity of electrical behavior of a particular appliance load. An estimate of this metric is the area enclosed under mean curve *b* when viewed with reference to *c*.

Number of Self-Intersections: Number of self-intersections for a V-I trajectory also corresponds to presence of higher order harmonics in current waveform.

Slope of Middle Segment: A near-zero slope of middle segment of the mean curve *b* is often characteristic of power-

electronic loads and serves to differentiate them from other kinds of loads. In figure 5, tangent of the angle θ represents this metric.

Area of Right and Left Segments: This refers to area enclosed by near-vertical segments towards the edges of the V-I trajectory.

We also propose a new shape feature of the V-I trajectory we refer to as *span* of the V-I trajectory. It represents the vertical distance between highest and lowest I-coordinate, equal to length of line segment d in fig. 5. It has a correspondence with fundamental active power *P* as it increases proportional to amplitude of instantaneous current. Its inclusion improves the precision of prediction in NILM by as much as 10% for employed learning algorithms over evaluated switching events from REDD.

Using this shape difference we are able to differentiate between switching events from various appliances. Table I expresses linear correlation coefficients for some of the aforementioned WS and PQ quantities. A magnitude close to one expresses strong linear dependence between two variables while a magnitude close to zero implies minimal linear relationship between the two. Coefficients in table I have been estimated using unique combinations from a set of about fifty thousand switching events belonging to REDD. A strong linear proportionality between curvature of mean line and harmonic content in current waveform expressed by *TeHD* is particularly obvious from the table. An increase in area enclosed by V-I trajectory is typically accompanied by a decrease in the number of self-intersections, illustrated by a negative value for respective coefficient in table I.

D. Disaggregation Algorithms

Choice of learning algorithm is a function of the characterization of the learning problem and the load signature used. The inputs to each learning algorithm are a specific number of 'training examples'; each example is a feature vector representing particular load signatures coupled with respective appliance label, $(\mathbf{x_i}, y)$, $\mathbf{x_i} \in \mathbb{R}^F$ from previous 'signature extraction' and 'event labeling' steps in fig. 1. For instance, HAR training example for each switching event contains 77 real numbers representing spectral power for concurrent segments of frequency for FT of current waveform alongside an appliance label. These 77 real numbers are extracted from FT for a range of 0-4kHz. WS contains 7 real numbers, PQ contains 4 real numbers in addition to appliance labels. These feature vectors are divided into three sets namely training, cross-validation and test sets. Training and cross-validation examples are used by learning algorithm to produce generalizable load models subsequently used to predict labels for test set feature vectors.

In our evaluation we have used four learning algorithms to evaluate our approach. These four algorithms represent popular choices of off-the-shelf learning algorithms proven

effective for a wide variety of classification problems across domains ([23]-[24]).

First of these algorithms is a feed-forward artificial neural network (ANN) trained using Levenberg-Marquardt (LM) method. A popular non-linear learning algorithm, ANN topologies are detailed in [25].

The second algorithm is a hybrid learning algorithm (ANN + EA), composed of an artificial neural network coupled with an evolutionary algorithm (EA with momentum) block for local search. Once the weights and parameters W_n for the ANN part are determined by LM or any other training algorithm [25], the EA part conducts a local search around the final solution for better precision of prediction. The principle for 'evolution' from W_n to the next 'generation' of parameters W_{n+1} is expressed by the following equations [23].

$$\Delta W_{n+1} = m\,\Delta W_n + (1 - m)gW_n \qquad (1)$$

$$W_{n+1} = W_n + \Delta W_{n+1} \qquad (2)$$

In 1 and 2, momentum constant m is typically a real number from the range (0,1) that describes the degree of 'presence' of W_n in the next generation W_{n+1}. Method to ascertain value of momentum constant m using stochastic global search is described in the next subsection. The value of g, a constant real number, is selected from the range $(0, 0.1)$ for all generations of EA [23].

The third algorithm evaluated for the course of this study is support vector method (SVM) with a Gaussian kernel function. Intuitively, the kernel function describes a distance measure that suggests weightage for various training examples in order to construct a hypothesis that determines the class label for each training instance. Self-selection of empirical constants associated with SVM kernel (C, σ) is addressed in the next subsection. SVM theory and application to MS-NILM for a relatively small set of appliance feature instances is discussed in more relative detail by Onoda et. al. [26].

The fourth and final algorithm is Adaptive Boost (AdaBoost) that uses decision stumps as weak classifiers. The premise of AdaBoost is to iteratively boost the prediction performance of the weak classifier over the training examples. This is accomplished by maintaining a distribution of weights that is updated in successive iterations so as to 'concentrate' on instances that are misclassified in the previous iterations [26].

Liang et. al. [3] and Onoda et. al. [26] have demonstrated the use of ANN, SVM and Adaptive Boost in context of NILM under separate system configurations. No systematic examination of all four with WS over benchmark datasets exists so far, hence these have been employed for this study.

E. Model Selection and Performance Optimization

Prior to training of all the aforementioned learning algorithms, setup parameters (number of neurons in each hidden layer N_h in case of ANN, momentum constant m in case of EA, kernel parameters C and σ in case of SVM) have to be specified. This step is typically referred to as 'model selection' [28]. In the course of this study, an enhanced variant of differential evolution (EDE) is used to automatically search and select these setup parameters. Fig. 6 illustrates the model
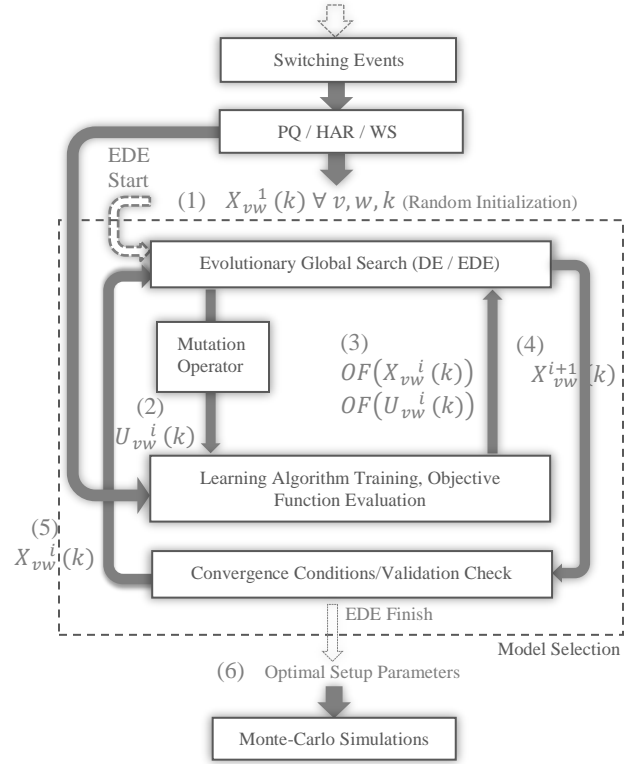

Fig. 6. Global search using enhanced differential evolution for model selection in ANN, ANN + EA and SVM.

selection step from fig. 1 in detail. A specific number of randomly chosen setup parameters or a candidate 'population' of parameters is chosen to be evaluated for prediction accuracy. The population is altered or 'mutated' and a new population is selected by selectively combining the initial and altered population according to specific rules (describing the 'fitness' of altered population relative to that of original population) so as to gradually maximize the prediction accuracy or minimize prediction error in the new population as this cycle is repeated. The term 'objective function' refers to prediction error in the following passage.

Differential Evolution (DE) is a heuristic, population-based global search strategy that offers more relative certainty and efficiency of convergence for minimization problem for non-linear continuous space functions [20]. Following passage reviews the algorithm used for parameter search [29], with special reference to DE and EDE.

DE prescribes an empirical constant RR (recombination rate) that decides the threshold for combining altered and original populations of parameters. Choice of this constant is a function of observer's experience and expertise. EDE is an enhanced variant of DE proposed in [29] whereby instead of an empirical recombination rate (RR), a 'fitness function' is described that weighs the fitness of mutant population relative to fitness of original population,

$$\text{Fitness of Mutant Population} = \frac{\frac{1}{OF(U)}}{\frac{1}{OF(X)} + \frac{1}{OF(U)}} \qquad (3)$$

$$\text{Fitness of Original Population} = \frac{\frac{1}{OF(X)}}{\frac{1}{OF(X)} + \frac{1}{OF(U)}} \qquad (4)$$

In eqs. 3 and 4, 'U' represents the mutant population of training parameters, 'X', the original population, 'OF', the objective function. So, a constant RR is replaced by a dynamic RR that weighs how much the altered population excels the original population in prediction accuracy.

A brief summary of the overall proposed strategy is described as follows. Fig. 6 illustrates the sequence of steps involved in the EDE-based global search algorithm, labeled as follows.

1) Populations of system variables (number of hidden layer neurons in case of ANN, kernel parameters in case of SVM or momentum constant in EA) or 'genes' for kth trainer and ith iteration, $X_{vw}^{i}(k), k = 1, 2, \ldots, N$ are randomly initialized in the beginning, a total of M individuals ($v = 1, 2, \ldots, M$) and G genes per individual for each trainer ($w = 1, 2, \ldots, G$).

2) Each individual in the mutant population $U_{vw}^{i}(k)$ is determined by a linear combination of genes from three randomly chosen individuals in the original population.

$$U_{vw}^{i}(k) = X_{aw}^{i}(k) + F \times \left(X_{bw}^{i}(k) - X_{cw}^{i}(k)\right)$$
$$a \neq b \neq c \neq v, \qquad F > 0$$

3) Objective function values for mutant and original population are determined by evaluating prediction error for the multi-class classifier in question (for a specific feature space and a specific learning algorithm). Fitness functions of mutant and original populations are determined from 1 and 2.

4) Each gene in $X_{vw}^{i+1}(k)$ is determined as follows:

$$x_{vw}^{i+1}(k) = \begin{cases} u_{vw}^{i}(k) \; if \; rand < Fitness\ Function\ (U) \\ x_{vw}^{i}(k) \; if \; rand \geq Fitness\ Function\ (U) \end{cases}$$

5) Step 2 is repeated with $X_{vw}^{i+1}(k)$ in place of $X_{vw}^{i}(k)$ until one of the stopping conditions is met, that is, either the objective function is minimized beyond a specific threshold (a real number) or it stays constant for a specific number of consecutive iterations (resulting in a 'validation stop') or the maximum number of iterations is reached.

6) After one of the stopping conditions is met, the best out of M individuals in the population $X_{vw}^{i+1}(k)$ – the one with the least value for objective function – is chosen to be the optimal set of setup parameters for learning algorithm in question.

For a multi-class classifier, performance optimization of each class can be carried out in parallelized fashion (requiring N replicas of above algorithm) such that for selected values of k, $\left|OF_k - \frac{1}{N}\right| < \varepsilon$. An alternative choice would be to use a single objective function following the above rule; simplest instances would be for selected values of k, $\sum_k \left|OF_k - \frac{1}{N}\right|$ or $\sum_k (OF_k - \frac{1}{N})^2$, depending on the penalty required as error grows and whether OF can adapt to degree of convergence of parameters like with the basic case of $\max_k OF_k$. However, design of objective functions that exhibit both dynamic adaptation and selective minimization is outside the scope of this study and will be addressed in subsequent research.

IV. Experimental Evaluation

This section presents numerical evaluation of benchmark signatures and learning algorithms for precision of prediction of appliance labels for switching events extracted from REDD. We have evaluated load signatures corresponding to these switching events for WS, PQ, HAR using the algorithms reviewed earlier: artificial feed-forward neural network (ANN), hybrid neural network (ANN + EA), support vector machine (SVM) and adaptive boost (AdaBoost) for one-dimensional decision stumps.

Precision of prediction $\eta \in [0, 1]$, used as accuracy metric, is determined as the number of correctly predicted classes for extracted switching events in the database, weighted by the total number of switching events. This study assumes perfect recall, that is, all true switching events are detected and reported to learning algorithms. In our evaluations, this assumption is conveniently justified since REDD reports the instances of change in instantaneous voltage and current waveforms. Due to sufficient acquisition rate, effectively the only false positives in the event detection phase are the instances where the change-event is split into two, each registering one half of the current response representative of respective appliance. The redundant events need to be programmatically eliminated. In the case where NILM is aided by sensors on individual appliances, false positives in event detection are dependent on frequency response of the sensor, noise and interference from neighboring appliances. Accuracy metrics that account for these false positives, such as F-score are recommended in this case [3].

A. Description of Dataset and Initial Conditions

Reference Energy Disaggregation Dataset (REDD) is a publicly available dataset containing detailed energy usage information of several homes over extended periods and in two granularities [20]. The low granularity data is average real power consumption of multiple households (both the mains and individual circuits) at a frequency of approximately 1Hz for mains and 0.33Hz for individual circuits. High granularity data is AC voltage and current waveform data from household mains acquired using commercial load monitors at a frequency of 16.5 kHz.

TABLE II
CIRCUIT LABELS FOR REDD HOUSEHOLD # 3 – 22 CHANNELS

Cct. Label	Appliance	Cct. Label	Appliance	Cct. Label	Appliance
1, 2	mains	9	dishwasher	16	microwave
3	unknown	10	furnace	17	lighting
4	unknown	11	lighting	18	Smoke alarms
5	lighting	12	unknown	19	lighting
6	electronics	13	Washer-dryer	20	Bathroom gfi
7	refrigerator	14	Washer-dryer	21	Kitchen outlets
8	disposal	15	lighting	22	Kitchen outlets

REDD high frequency data considered for this evaluation contains energy usage data worth about twenty days for two houses. In case of house # 3, for instance, a total of 22 channels of voltage and current waveform data are present. Prior to clustering the delta-form signatures, only appliances with $P > 50W$ are considered.

Subsequent to signature extraction, relative sizes of 0.45, 0.1 and 0.45 are utilized for training, cross-validation and test sets in training of learning algorithms: a base case for further comparison in next section. EDE (typically 30 individuals, 50 iterations) is utilized to assert the best possible training parameters. In our evaluation, $F = 0.5$ is generally a good choice for EDE step 2 that provides a reasonable balance between time required for EDE convergence and precision of models with suggested parameters. Subsequently, Monte Carlo simulations are conducted with these parameters to evaluate η.

B. Monte-Carlo Simulations

A Monte-Carlo simulation evaluates the precision of employed learning algorithms a large number of times under an unaltered computational capacity for the duration of the simulation. For instance, with ANN as learning algorithm, Levenberg-Marquardt (LM) algorithm is invoked in every iteration of the simulation to search for ANN weights (using the number of neurons in each hidden layer from the previous model-selection step). The large number of simulations accounts for variability in precision of prediction of trained ANN due to random initialization of weights in LM algorithm.

Available degrees of freedom while pre-processing the datasets in order to extract switching events are:

- Relative sizes of training and test sets,
- threshold on active appliance power consumption (P_{min}) and purity of clustering done during pre-processing [22]

Unique combinations of these criteria can be modified to generate a unique combination of switching events from the dataset and a unique Monte-Carlo simulation.

Relative sizes of training and test sets compare the robustness of performance for various algorithms over PQ, HAR and WS, particularly during early operational periods in an event-based NILM system where training examples are too few for the learning algorithm to account for all possible load scenarios and consequently to yield generalizable appliance models.

Purity of clustering [22] during the pre-processing module is related to the optimal number of appliance classes selected in order to cluster the extracted waveforms. This optimal number of classes is a function of threshold on power consumption of appliances, P_{min}, that are included in a particular Monte-Carlo simulation. A large number of appliances with values of load signatures in close proximity (high 'similarity'), noisy and/or highly non-sinusoidal conditions might lead to a larger number of anomalous training examples ('outliers') that might not represent the most frequent operating state for the appliance in question, which might render the learning algorithm more susceptible to incorrect classification, as will be demonstrated in the next section.

It is important to indicate here that these two function as so-called *extrinsic* criteria for generating unique load scenarios, in the sense that these are modifiable while retaining the original sequence of occurrence of switching events and do not represent intrinsic load characteristics. In section V, this original sequence of switching events from REDD is modified. A large number of unique load scenarios are generated using appliance profiles (feature vectors) from REDD. This allows us to simulate the effects of dynamic and noisy load signatures on capacity of classification for various algorithms and load signatures. Research by Liang and colleagues [3] presents convenient criteria for load dynamics, appliance similarity and electrical noise that are consulted for the purpose of this study. Definitions for these criteria *intrinsic* to composite load are mentioned in the next section.

C. Numerical Results

Table III lists the median overall, training set and test-set precision of prediction for the three benchmark load signatures (an Intel Core i7 machine, CPU clock 3.1 GHz, 8GB of RAM). It is evident from the numerical figures in table III that wave-shape metrics (WS) outperform or generally compare with both PQ and HAR in precision of prediction. It is important to indicate here that since HAR is represented by 77 real numbers expressing spectral power for various segments of the Fourier transform while WS is represented by 7 real numbers for metrics described in the previous section so for orders of magnitude less number of features, WS outperforms or at a minimum, offers comparable prediction accuracy with HAR and hence is relatively more robust for event-based operation. It is also evident that including EA with momentum alongside ANN as learning algorithm does not provide a substantial improvement in precision over ANN.

TABLE III
MEDIAN OVERALL/TEST SET/TRAINING SET PRECISION (X 100%) FOR VARIOUS
CHOICES OF LOAD SIGNATURE FROM MONTE-CARLO SIMULATIONS

LS	ANN	ANN + EA	SVM	AdaBoost
PQ	88.5/88.3/88.8	88.8/88.6/88.9	97.2/96.9/97.5	99.3/98.8/98.8
HAR	82.1/80.4/83.4	82.8/81.4/84.4	98.7/98.0/99.3	98.8/97.4/99.8
WS	92.0/91.5/92.5	91.5/90.9/92.1	98.1/97.1/98.9	99.1/98.7/99.1

Table IV lists the maximum, mean and median precision for ANN, ANN+EA, SVM and AdaBoost with PQ, HAR and WS.

TABLE IV
OVERALL MAXIMUM/MEAN/MEDIAN PRECISION (X 100%) FOR VARIOUS
CHOICES OF LOAD SIGNATURE FROM MONTE-CARLO SIMULATIONS

LS	ANN	ANN + EA	SVM	AdaBoost
PQ	91.1/88.4/88.5	91.7/88.9/88.8	97.4/97.3/97.2	99.4/99.3/99.3
HAR	87.1/81.1/82.1	88.0/82.3/82.8	98.9/98.7/98.7	98.9/98.8/98.8
WS	94.3/91.4/92.0	96.2/90.8/91.5	98.3/98.1/98.1	99.2/99.1/99.1

Fig. 7 indicates an almost comparable resilience to increased proportion of test set data for PQ, HAR and WS with AdaBoost and ANN. Fig. 8 indicates the drop in precision of prediction with smaller values of P_{min}, for ANN as learning algorithm. This generally corresponds to a large number of low-powered load signatures in close proximity, as discussed earlier.
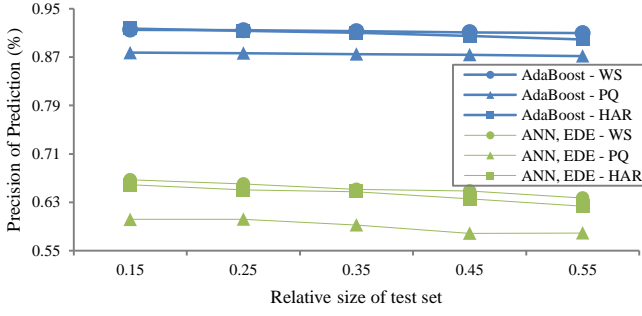
Fig. 7. Trends of precision of prediction for learning algorithms as a function of relative size of training, test and cross-validation sets.
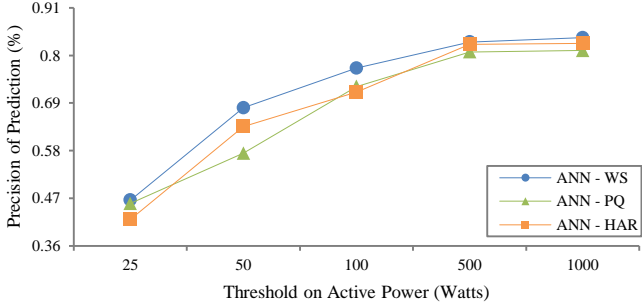

Fig. 8. Trends of precision of prediction for learning algorithms as a function of minimum threshold on appliance power consumption.
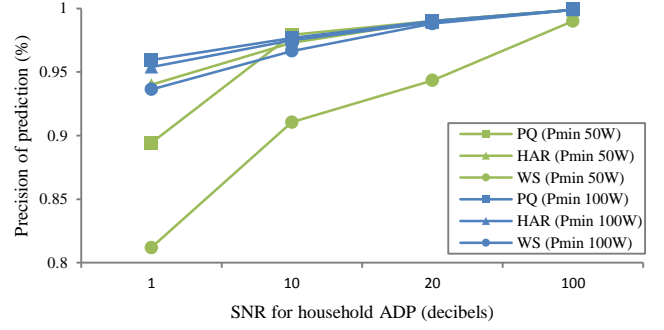

Fig. 9. Precision of prediction as a function of signal-to-noise ratio for household ADP using Adaptive Boost as classification algorithm.
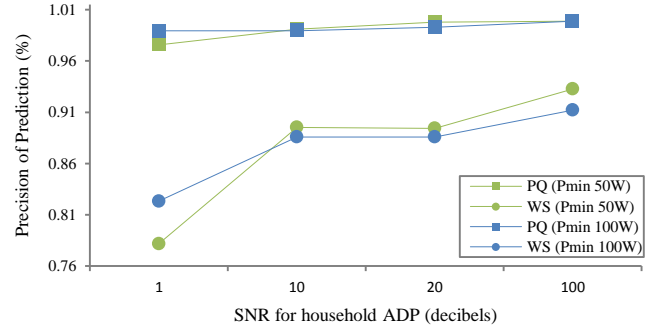

Fig. 10. Precision of prediction as a function of signal-to-noise ratio for household ADP using Support Vector Method as classification algorithm.
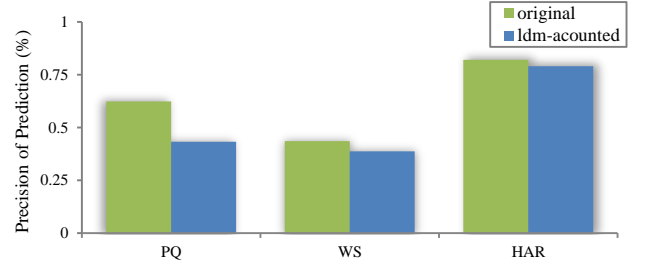

Fig. 11. Disparity in precision of prediction for various load signatures and classification algorithms with inclusion of load dynamics

V. EXTENDED SIMULATIONS

In the next portion of our empirical study, appliance load profiles (feature vectors) derived from REDD are used to generate simulations for thousands of different load scenarios. Each load scenario contains a sequence of switching events for a defined period of time, typically ranging from a day to a month. These scenarios are generated by sequential addition/subtraction of individual signatures to/from the composite load. The signatures from our database contain as many as 40 distinct appliance signatures for different respective operating states. The simulation accounts for dynamic loading of various ubiquitous loads (particularly household electric motors and air conditioners) and electrical noise. A complete description of implemented NILM simulator is outside the scope of this paper and will be detailed in a future study. However, the relevant results of experiments pertaining to the degrees of freedom affecting the precision of prediction for various algorithms with WS are detailed.

A. Sensitivity towards Noise in Load Signatures

Effect of noise is simulated by adding zero mean white Gaussian noise to ADP of generated load scenarios. For this particular set of Monte-Carlo simulations (P_{min} = 50W, 100W, 14 days, 15 switching events per hour with a normally distributed frequency), all employed algorithms classify more than 90% of training examples of load signatures for SNR > = 20. However, precision of prediction drops with decrease in SNR, most considerably for WS which suffers more drastic drop in precision compared to PQ and HAR for all employed algorithms. Fig. 9 and 10 illustrate the case for Adaptive Boost and Support Vector Method. About 20% drop in case of

adaptive boost (with P_{min} = 50W which corresponds to about 12000 training examples) is apparent from fig. 10 (a). This type of response is due to drastic shape distortion of V-I trajectories of respective appliances for small values of SNR.

B. Sensitivity towards Dynamic Load Signatures

Dynamic behavior of appliance loads is recognized as inherent variability of shape for instantaneous current waveform due to of dynamic loading. Similarity refers to the proximity of numerical values of load signatures. Discrete Fourier transform ($X(f_k) = k_r + jk_i$) of various single-cycle snapshots (CW) of an appliance load is calculated using FFT algorithm. Range of values for k_r and k_i in the database expresses the degree of variability in CW for respective appliances. Inverse Fourier transform is used to reconstruct CW for appliance signatures keeping in account the updated k_r and k_i values weighted by LDM. A modification of the relationship between reconstructed and original CW proposed by Liang et. al. [3] is employed for this study.

$$CW_i(t) = \frac{1}{2} \sum_{c=0}^{N-1} \left[\left(sel(k_{r,c}) \right) + j \left(sel(k_{i,c}) \right) \right] e^{\frac{j2\pi ct}{N}} \qquad (8)$$

In eq. 8, $sel(k_{r,c})$ and $sel(k_{i,c})$ refer to a uniform sampling function that randomly selects any one of the available k_r and k_i in the database for a particular appliance, in order to produce all possible variations of the instantaneous current waveform with a uniform probability in the updated CW used for subsequent Monte-Carlo simulations. Fig. 11 highlights the disparity in precision of prediction for original and adjusted CW, AdaBoost and PQ, WS and HAR over a large number of simulated load scenarios. PQ generally suffers a slightly larger disparity in precision compared to HAR and WS for this set of Monte-Carlo simulations (P_{min} = 50W, 20 days, SNR = 10, 15 switching events per hour with a normally distributed frequency). A larger separation between signature clusters in case of WS and HAR might be associated with this particular observation, so as to allow for a relatively robust precision of prediction. However, this difference in precision for WS and PQ diminishes for smaller values of SNR where WS suffers an explicit disadvantage obvious from the previous subsection.

VI. Conclusions and Future Work

As was empirically demonstrated in earlier sections, wave-shape metrics (WS) are a competitive and robust basis of load classification and disaggregation and have a direct correspondence to operating characteristics of appliances as contained in current wave-shape. This affords NILM applications additional load characteristics for challenging scenarios such as multiple near-identical appliances, dynamic loading of electric motors etc. It therefore establishes a promising direction for research in unsupervised energy disaggregation and fault monitoring of appliances using a combination of different load metrics as one comprehensive basis. Another important research problem is the selection of optimal load models that achieve a balance between memory constraints on their training and overall precision achieved for NILM. Applications of metaheuristic optimization and online learning in NILM can potentially prove instrumental in selection and self-calibration of household load models which would improve the reliability and autonomy of diagnostic feedback applications based on NILM.

References

[1] G.W. Hart, "Non-intrusive load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870-1891, Dec. 1992.

[2] S. Shaw, S. Leeb, L.K. Norford and R. Cox, "Nonintrusive load monitoring and diagnostics in power systems," *IEEE Trans. Instrum. Meas.,* vol.57, no.7, pp. 1445-1454, Jul. 2008.

[3] 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 Delivery,* vol.25, no.2, pp.561-569, Apr. 2010.

[4] C. W. Gellings, "The concept of demand-side management for electric utilities," *Proc. IEEE*, vol. 73, no. 10, pp. 1468-1470, 1985.

[5] H.Y. Lam, G.S.K. Fung, and W.K. Lee, "A novel method to construct taxonomy of electrical appliances based on load signatures," *IEEE Trans. Consumer Elec.,* vol.53, no.2, pp.653-660, May 2007.

[6] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong. "Power signature analysis," *IEEE Power Energy Mag.*, vol. 1, no. 2, pp. 56-63, 2003.

[7] K. D. Lee, , S. B. Leeb, L. K. Norford, P. R. Armstrong, J. Holloway, and S. R. Shaw. "Estimation of variable-speed-drive power consumption from harmonic content," *IEEE Trans. Energy Convers.*, vol. 20, no. 3, pp. 566-574, 2005.

[8] A. I. Cole and A. Albicki, "Non-intrusive identification of electrical loads in a three-phase environment based on harmonic content," in *Proc. 2000 IEEE Instrumentation and Measurement Technology Conf.*, pp. 24-29.

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

[10] J. Zico Kolter and Matthew J. Johnson, "REDD, A public dataset for energy disaggregation research," in *Proc. 2011 SustKDD Workshop Data Mining Applications in Sustainability*, pp. 1-6.

[11] Z. Ahmed, 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.

[12] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *Proc. 2012 Conf. Artificial Intelligence*, pp. 356-362.

[13] H. S. Kim, "Unsupervised disaggregation of low frequency power measurements," M.S. thesis, Dept. Comp. Sci., Univ. Illinois, Urbana-Champaign, IL, 2012.

[14] K.D. Lee, "Electric load information system based on non-intrusive power monitoring," Ph.D. dissertation, Dept. Mech. Eng., MIT, Cambridge, MA, 2003.

[15] S. Leeb, "A conjoint pattern recognition approach to non-intrusive load monitoring," Ph.D. dissertation, Dept. Elec. Eng. & Comp. Sci., MIT, Cambridge, MA, 1993.

[16] H. W. Lai, G.S.K. Fung, H.Y. Lam, and W.K. Lee, "Disaggregate loads by particle swarm optimization method for non-intrusive load monitoring," presented at the 13th Int. Conf. Electrical Engineering, Hong Kong, 2007.

[17] K. Suzuki, S. Inagaki, T. Suzuki, H. Nakamura, and K. Ito, "Nonintrusive appliance load monitoring based on integer programming," in *Proc. 2008 SICE Annual Conference*, pp.2742-2747.

[18] Y. H. Lin, M. S. Tsai, and C. S. Chen. "Applications of fuzzy classification with fuzzy c-means clustering and optimization strategies for load identification in NILM systems," in *Proc. 2011 IEEE Int. Conf. Fuzzy Systems*, pp. 859-866, 2011.

[19] T. Hassan, "Bi-level characterization of manual setup residential non-intrusive demand disaggregation using enhanced differential evolution," presented at the 1st Int. Workshop on Non-Intrusive Load Monitoring, Pittsburgh, PA, USA, 2012.

[20] R. Storn and K. Price, "Differential Evolution - A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.,* vol. 11, no. 4, pp. 341-359, Dec. 1997.

[21] J. Liang, S.K.K. Ng, G. Kendall and J.W.M. Cheng, "Load signature study part 1: basic concept, structure and methodology," *IEEE Trans. Power Del.,* vol. 25, no. 2, pp. 551-560, Apr. 2010.

[22] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge, UK: Cambridge Univ. Press, 2008.

[23] N. Amjady and F. Keynia, "Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm," *IEEE Trans. Power Syst.,* vol. 24, no. 1, pp. 306-318, Feb. 2009.

[24] I. Guyon. (2013, January 10). *SVM Application List* [Online]. Available: http://www.clopinet.com/isabelle/Projects/SVM/applist.html

[25] M. Beale and H. Demuth, *Neural Networks Toolbox for Use with MATLAB: User's Guide*, vol. 3. Massachusetts: The MathWorks, 1998.

[26] T. Onoda, G. Rätsch, and K. R. Müller. "Applying support vector machines and boosting to a non-intrusive monitoring system for household electric appliances with inverters," in *Proc. 2000 Int. ICSC Symposium on Neural Computation*.

[27] Y. Freund, R. Schapire, and N. Abe. "A short introduction to boosting," *J. Japanese Soc. Artificial Intelligence*, vol. 14, pp. 771-780, 1999.

[28] U. Anders and O. Korn. "Model selection in neural networks," *Neural Networks*, vol. 12, no. 2, pp. 309-323, 1999.

[29] N. Amjady, F. Keynia, and H. Zareipour, "Short-term load forecast of microgrids by a new bilevel prediction strategy," *IEEE Trans. Smart Grid*, vol.1, no.3, pp.286-294, Dec. 2010.