Non-Intrusive Load Monitoring by Multi-constraints Elastic-Net Regression

Yu Ran[1], Zhixin Zhang[2*]
[1]*Basic Department, Southwest Jiaotong University Hope College, China*
[2]*School of Big Data and Statistics, Sichuan Tourism University, China*
**Corresponding Author*

Abstract: Non-intrusive load monitoring (NILM) is a critical technology in smart energy management, enabling detailed appliance-level energy disaggregation from a single aggregated power signal without the need for individual sensors. In this paper, we introduce a novel regression-based approach for NILM that incorporates multiple physical constraints—sparsity, smoothness, and finite-state consistency—into a unified optimization framework. The proposed model enhances the classical Elastic Net method by integrating domain-specific knowledge to better represent the operating characteristics of household appliances. Specifically, the added constraints help to promote solutions that are not only sparse and stable but also physically plausible, leading to more accurate identification of appliance states and power consumption patterns. Evaluated on the widely-used AMPds and UK-DALE dataset, our method demonstrates a marked improvement over Elastic Net, Lasso, GSP and HMM, with an average increase in R2 score performance of approximately 5%-7% across a diverse set of appliances in AMPds and 2%-6% in UK-DALE. Additionally, some complex appliances will get significant improved results by our method. The results under- score the importance of incorporating structural prior knowledge into regression models for NILM and highlight the potential of the method for practical applications in energy analytic and residential load monitoring.

Keywords: Non-Intrusive Load Monitoring Multi-Constraint Regression; Energy Disaggregation; Elastic Net; Appliance Identification

1. Introduction

The push for carbon neutrality and peak emissions is driving a significant societal and economic transformation, with energy system restructuring at its core. A key strategy for achieving these dual-carbon objectives is enhancing energy efficiency. Buildings are a major contributor to energy consumption, accounting for 36-40% of national energy use in countries like Germany and Russia, with households responsible for 23% of this total ([25]–[8]). Research indicates that intelligent energy management, particularly real-time monitoring of appliance-level power and automated efficiency audits, can reduce household electricity consumption by 12-35% through personalized behavioral prompts. As a result, Non-Intrusive Load Monitoring (NILM) underscore the importance in the smart grid [36]. A significant advantage of NILM is its minimal invasiveness; it only requires sensors or collectors at the user's electricity meter to gather power data, avoiding direct access to specific electrical behavior details. This approach respects user privacy. When combined with block-chain technology, private data security can be effectively managed [10]. Furthermore, by separating user identity from power data, NILM can integrate with privacy protection and secure data sharing techniques [4]. This ensures that even if data is acquired by third parties, it cannot be traced back to specific users, thereby safeguarding privacy.

In summary, NILM offers cross-disciplinary applications in the following areas:

1) Cryptographic approaches: Homomorphic encryption and secure multi-party computation enable NILM calculations on encrypted data without exposing original consumption information or model parameters [26]. These methods allow mathematical operations on ciphertexts, maintaining confidentiality. A hybrid encryption scheme combining AES and RSA efficiently encrypts bulk data while securing key exchange and digital signatures for

authentication [28], balancing efficiency with strong security.
2) Differential privacy: This rigorous framework provides quantifiable privacy guarantees in NILM systems by injecting controlled noise into smart meter readings or processed data [5, 33]. Research has established theoretical bounds between privacy parameters and NILM performance, optimizing the trade-of between privacy and data utility [32]. Recent developments include DP2-NILM, a distributed framework integrating local and global differential privacy within federated learning architectures [5].
3) Federated learning architectures: Federated learning facilitates decentralized model training across multiple endpoints without centralizing raw consumption data. This preserves privacy by keeping sensitive data at its source and sharing only model updates. Strategies like FedAvg and FedProx have been adapted for NILM, enabling aggregated learning across households while minimizing privacy exposure [5].
4) Integration with electricity markets: NILM’s application in electricity markets supports dynamic billing and enhances demand-response programs [14]. Privacy-preserving load forecasting allows real-time monitoring and dynamic pricing without compromising individual household data. These advancements enable system operators to evaluate load forecasting models using hybrid deep learning and compute consumer bills based on dynamic pricing, all without accessing raw consumption data [14].
Overall, the integration of privacy-preserving techniques with NILM technology facilitates valuable smart grid applications while addressing critical consumer privacy concerns. Ongoing research aims to refine the balance between data utility and privacy, with emerging techniques focusing on lightweight cryptographic solutions, improved federated learning strategies, and more precise differential privacy implementations. These advancements support the responsible deployment of NILM in electricity markets, enabling dynamic pricing, demand-response initiatives, and grid optimization without compromising consumer privacy.
The current landscape of Non-Intrusive Load Monitoring (NILM) research is primarily structured around two distinct methodological paradigms: non-event-based and event-based approaches, each with its own philosophical underpinnings and technical implementations [24, 49].
Non-event-based methods, often characterized as “direct” or “state-based” disaggregation, treat the aggregate power signal as a continuous-time series generated by a combination of underlying appliance states. These approaches aim to model the temporal evolution of these states directly, without explicitly detecting discrete switching moments. A cornerstone technique in this category is the Hidden Markov Model (HMM) and its variants [20], which model each appliance as a Markov process with hidden states (e.g., 'off,' 'low,' 'high') and observable power emissions. The disaggregation task becomes one of inferring the most likely sequence of hidden states for all appliances given the observed total power. Similarly, Bayesian methods [47] provide a probabilistic framework to infer appliance states by combining prior knowledge (e.g., typical power ratings, usage patterns) with the observed aggregate data. These methods are particularly powerful for modeling appliances with multi-state operation or continuously variable power draws, as they reason about the entire load composition holistically over time.
Conversely, event-based methods adopt a “detect-then-classify” philosophy. This paradigm is predicated on the observation that most appliance operations are initiated by a discrete switch-on or state-change event [3], which manifests as a step-change or characteristic transient in the aggregate power signal. The process is typically two-staged: first, a detection stage identifies the precise points in time where significant power changes occur. Techniques here include margin-based window approaches [2], which analyze power differences within sliding windows to flag potential events, and statistical change-point detection [11], which uses statistical tests to identify points where the underlying data distribution shifts. Second, a classification stage analyzes the features of each detected event (e.g., magnitude of real/reactive power change, transient shape) to identify the specific appliance responsible. This approach is often more intuitive and computationally efficient for appliances with distinct, infrequent switching events.
In the area of NILM, two classical perspectives exist: classification and regression. We summarize their characteristics below.

From a classification standpoint, feature extraction is crucial. Most such studies fall under event-based NILM, which offers several key advantages over non-event-based methods:

1) Computational efficiency: Inference occurs only upon detected events in the aggregate signal, significantly reducing computational demands.

2) Robustness to new appliances: The introduction of previously unmodeled appliances does not impact the classification accuracy of known appliance events.

3) Easier unsupervised implementation: The sparse nature of load events simplifies unsupervised modeling of appliances.

In NILM studies, many researchers have focused on identifying appliance characteristics such as current, power, or various statistical properties derived from measurements. However, classification accuracy has not yet reached desired levels [9]. To improve the extraction of representative load features, the concept of an appliance signature has been introduced to encapsulate features, allowing the application of feature extraction techniques from other fields, including deep learning methods used in computer vision [21]. The V-I trajectory, first introduced by Lam et al. [17], serves as a distinctive load signature, providing deep insights into voltage and current data for the monitored appliance. Numerous studies have expanded on this, introducing various shape features, as documented in the work by Wang et al. [32]. Additionally,Du et al. [7] has transformed the V-I trajectory into different structured visual representations, including grayscale and color-encoded V-I trajectories. The V-I trajectory can be viewed as a geometric object containing information from both voltage and current signals, which can distinguish between various appliance types by its shape, making it a feasible approach for extracting appliance characteristics in NILM tasks.

Deep learning techniques also offer alternative feature extraction methods. For instance, conditional autoencoders are used for feature extraction in the work by Yu et al. [38]; deep convolutional neural networks (DCNNs) for trajectory classification in the work by Yin et al. [37]; hash dictionary conversion methods in the work by Li et al. [18]; and advanced mathematical representations using recurrent graphs and Gramian matrices in the work by Zhang et al. [42]. Furthermore, for image representation in NILM, many deep learning-based feature extraction methods can be employed. Examples include Zhang et al. [41]'s recurrent progressive fusion-based learning for image classification; Xiao and Xia [34]'s image enhancement method based on image decomposition and deep learning; Song et al. [29]'s wavelet-guided difusion deep learning model; Huang et al. [13]'s deep intrinsic image decomposition via physics-aware neural networks; Huang et al. [12]'s DecoupleNet methodology for low-light images; and Jiang et al. [15]'s knowledge-based global-local attention network for image classification. Wu et al. [35] proposed a hybrid approach integrating Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) to capture image featuresof appliances, achieving efficient household load identification. Yuan et al. [40] integrated Gated Recurrent Units (GRU) with Temporal Convolutional Networks (TCN) to extract potential features from training data. Zhao et al. [44] proposed Mean Teacher method which employs a teacher-student framework to effectively everage unlabeled data. Zhou et al. [48] introduced local feature extraction based on an efficient deep learning model. Some other promising approach to extract valuable information from mislabeled data for robust models training can be refered to works by Zhao et al. [45] and Zhao et al. [46]. While these methods are highly beneficial for image representation in NILM, they often suffer from a “black box” nature, lacking clear mathematical interpretability.

Therefore, we aim for a method that can interpret NILM with explicit mathematical meaning. Based on this motivation, regression provides another viable solution for NILM. Unlike classification-based approaches that identify appliance states, regression-based NILM treats energy disaggregation as a sequence-to-sequence regression problem, aiming to reconstruct continuous power waveforms of individual appliances [33]. This perspective preserves detailed power fluctuations and is more adaptable to complex multi-state devices (e.g., washing machines with variable cycles).

Some classical achievements in regression models for NILM can be summarized as follows:

1) Long Short-Term Memory (LSTM) Networks: LSTMs effectively capture long-term temporal dependencies in power sequences, making them suitable for learning patterns in high-frequency smart meter data. For example, Bayesian-

optimized LSTM models with non-causal structures have been proposed to handle multi-state devices and improve regression accuracy [16].

2) Sequence-to-Sequence (Seq2Seq) and Sequence-to-point(Seq2Point) Frameworks: Encoder-decoder architectures (e.g., combining Bi-directional GRUs with attention mechanisms) map aggregated power sequences to appliance-level power values by focusing on critical temporal features [27]. Dai et al. [6] employees sequence-to-point models and transfer learning to disaggregate total household energy consumption into individual appliance usage. Residual connections further enhance training stability and feature extraction for such methods.

3) Feature engineering: Regression-based NILM utilizes both low-frequency (e.g., active/reactive power) and high-frequency features (e.g., voltage-current waveforms) [30]. High-frequency features improve the tracking of rapid power changes but increase computational complexity.

4) Optimization frameworks: Bayesian parameter optimization is employed to fine-tune model hyperparameters (e.g., layer depth, learning rate), addressing issues like overfitting and enhancing generalization across households [22].

Regression-based NILM offers a granular, continuous representation of appliance energy consumption, advancing smart grid analytic for energy efficiency and demand response. While deep learning models (e.g., LSTMs and Seq2Seq) have improved regression performance, NILM research must address scalability and generalization barriers for industrial and residential deployment. Therefore, establishing and analyzing models from a mechanistic perspective is necessary. In this paper, a regression model based on multi-condition constraints is proposed for NILM. More precisely, the proposed method can be viewed as an improved version of Elastic Net. Testing is carried out on the AMPds dataset, and a comprehensive discussion comparing the proposed method and Elastic Net is provided to explain its effectiveness.

The main contributions of this paper can be listed as following:

1) A NILM model based on multi-constraints regression is proposed

2) The proposed method is tested on two public datasets AMPds and UK-DALE with satisfactory results.

3) A comprehensive discussion on the reason why multi-constraints regression can improve the result of appliance identification is given.

2. Methodology

2.1 Problem of NILM

In this section, we review the reformulation of NILM from the view of regression.

As well known, the total power consumption is the aggregation of different appliance, i.e.

$$p(t) = \sum p_i(t) + \varepsilon(t) \tag{1}$$

where ϵ(t) is the noise in aggregated signal, Pi (t) is the power consumption of i-th appliance.

Moreover, if we pay more attention to real world, most appliance has various of states–at least, it has on and of states. It will yield significant change of power under different state. So NILM method must consider such case. In this sense, Equation (1) can be reformulated to following formula:

$$p(t) = \sum c_i(t) s_i(t) + \varepsilon(t) \tag{2}$$

where ci (t) is the power consumption corresponding to different mode, si (t) is percentage of corresponding mode in aggregated power consumption.

In general, it will be more accuracy to use sequence to sequence method, rather than sequence to point, as pointed by Zhao et al. [46]. That's to say, We are more tend to give disaggregation result of a time series with length n. In such case, Equation (2) has evoluted into following formula of matrix.

$$\mathrm{P} = \mathrm{CS} + \mathrm{E} \tag{3}$$

where P is the matrix of aggregated signal in d days with size n × d, C is the matrix of power consumption corresponding to different mode with size n × L which is determined by training set, S is the matrix of percentage for corresponding mode in aggregated power consumption with size L × d.

To explain Equation (3) more explicitly, we give following example to illustrate it.

Consider two appliances ap1 and ap2, typical power consumption mode of ap1 and ap2 is collected in three days, given by following matrixes:

$$\mathrm{C}_1 = \begin{pmatrix} 100 & 80 & 90 \\ 120 & 110 & 130 \\ 80 & 90 & 70 \\ 60 & 70 & 50 \end{pmatrix} \quad \mathrm{C}_2 = \begin{pmatrix} 150 & 140 & 160 \\ 50 & 60 & 40 \\ 160 & 150 & 170 \\ 60 & 50 & 30 \end{pmatrix}$$

then matrix C is:

$$C = \begin{pmatrix} 100 & 80 & 90 & 150 & 140 & 160 \\ 120 & 110 & 130 & 50 & 60 & 40 \\ 80 & 90 & 70 & 160 & 150 & 170 \\ 60 & 70 & 50 & 40 & 50 & 30 \end{pmatrix}$$

each column is curve of power consumption of different mode corresponding to a typical day, first three column are corresponding to ap1, last three column are corresponding to ap2. If in a predicted day, mode corresponding to first column takes 0.4 in total consumption of ap1, mode corresponding to three column takes 0.6 in total consumption of ap1, mode corresponding to four column takes 0.5 in total consumption of ap2, mode corresponding to last column takes 0.5 in total consumption of ap2. So the matrix S corresponding to ground truth is $S = \begin{pmatrix} 0.4 & 0 & 0.6 & 0.5 & 0 & 0.5 \end{pmatrix}^T$ and the real power consumption is determined by following multiplication of matrixes:

$$P = \begin{pmatrix} 100 & 80 & 90 & 150 & 140 & 160 \\ 120 & 110 & 130 & 50 & 60 & 40 \\ 80 & 90 & 70 & 160 & 150 & 170 \\ 60 & 70 & 50 & 40 & 50 & 30 \end{pmatrix} \begin{pmatrix} 0.4 \\ 0 \\ 0.6 \\ 0.5 \\ 0 \\ 0.5 \end{pmatrix} = \begin{pmatrix} 249 \\ 171 \\ 239 \\ 89 \end{pmatrix}$$

So the problem of NILM can be summarized to a problem of regression to solve the vector S in real scenario. In this sense, the framework of NILM can be formulated as following optimization problem:

$$\min_S \left\| P - CS \right\|^2 \tag{4}$$

2.2 Constraints of proposed model

Our proposed NILM model integrates several constraints to enhance its accuracy and interpretability, as summarized in Figure 1. These constraints address the inherent characteristics of appliance operation, moving beyond a simple least squares minimization.

2.2.1. Sparse condition

In practical NILM scenarios, the ideal disaggregation represented by a basic least squares formulation is insufficient. It is not ideal as displayed in Equation (4). A fundamental characteristic of appliance usage is that, at any given time, only a subset of appliances are actively consuming power. This implies a “finite state” constraint, where the number of active appliances or their operational states is relatively small. To mathematically enforce this, we introduce an L1-norm regularization term into the optimization objective:

$$\min_S \left\| P - CS \right\|^2 + \beta \left\| S \right\|_1 \tag{5}$$

Here, the L1 norm, defined as the sum of the absolute values of the elements in S, promotes sparsity. By penalizing non-zero elements, the optimization process is encouraged to yield a solution S where many appliance states are zero, effectively identifying which appliances are inactive. This aligns with the real-world observation that only a few appliances operate simultaneously. While the L1 term provides a suitable framework for promoting sparsity, it often leads to solutions that are not sufficiently smooth, necessitating further regularization.

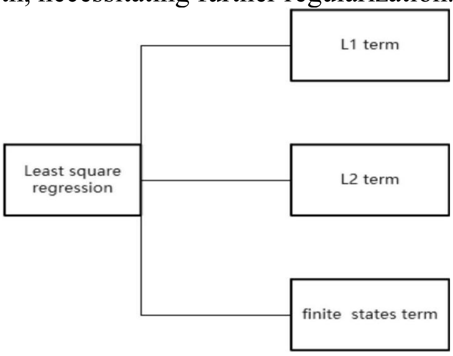

Figure 1. Proposed NILM Model

2.2.2. Smooth condition

As discussed, relying solely on the L1-norm for sparsity can result in solutions that exhibit undesirable abrupt changes or oscillations in the estimated appliance power consumption. To address this, we incorporate an L2-norm regularization term into the optimization objective. This leads to a formulation akin to Elastic Net regression:

$$\min_S \left\| P - CS \right\|^2 + \beta_1 \left\| S \right\|_1 + \beta_2 \left\| S \right\|^2 \tag{6}$$

The L2 norm, specifically the Frobenius norm $\left\| S \right\|^2 = \sum_{i,j} S_{i,j}^2$ is differentiable and penalizes large values in S. Its inclusion encourages a smoother distribution of power consumption across time and appliances. This helps to mitigate the “spiky” or erratic behavior that can arise from pure L1 regularization, leading to more physically plausible and stable disaggregation results. The combination of L1 and L2 regularization provides a balance between sparsity and smoothness, which is crucial for robust NILM.

2.2.3. Finite state condition

Even with L1 and L2 regularization, the model does not explicitly enforce the fundamental principle of energy conservation or the finite nature of appliance states in a more structured way. For an individual appliance, its total power consumption at any given moment is the sum of its power consumption across its various possible operational states. To mathematically capture this, we introduce an additional constraint that ensures the sum of the disaggregated power for each appliance corresponds to its total power consumption, effectively representing an energy conservation principle at the appliance level.

$$S_1 = R S \tag{7}$$

where S_1 is a matrix whose all elements are one with size k × d (where k is the number of appliances and d is the number of time steps) with all elements equal to one, representing the total power consumption for each appliance across all its states. R is a binary matrix (elements are 0 or 1) that maps the individual states within S to the total power consumption for each appliance. This mapping satisfies the condition:

$$e_1 = R\, e_0 \tag{8}$$

where e_1 is the column vector with dimension k, e0 is the column vector with dimension d. This ensures that the sum of the states for each appliance, as defined by RS , equals the total power for that appliance, represented by S_1.

Roughly, Equation (7) can be viewed as a mathematical expression of energy conservation. When such condition is considered, optimization objective is updated as following:

$$\min_S \|P - CS\|^2 + \beta_1 \|S\|_1 + \beta_2 \|S\|^2 + \beta_3 \|S_1 - RS\|^2 \tag{9}$$

3. Experiments

3.1 Metrics and comparative method

In this study, we employ Graph Signal Processing (GSP), Elastic Net, Lasso and Hidden Markov Model (HMM) as comparative methodologies. GSP, a novel approach, has demonstrated promising performance on the REDD dataset [18] and has garnered significant research interest. HMM is a classical method which have been widely used in NILM [20]. The Elastic Net, a regularized linear regression algorithm introduced by Zou and Hastie [50], can be considered an advancement over Lasso. Its inception was driven by the limitations of both Ridge Regression and Lasso Regression when confronted with intricate datasets. Consequently, it has found extensive application across diverse fields such as bioinformatics [23], financial economics [39], computer vision and signal processing [43], and high-dimensional data scenarios [1]. The fundamental principle of the Elastic Net lies in its fusion of L1 and L2 regularization techniques, thereby harnessing the benefits of both. It concurrently applies L1 and L2 penalties via a convex combination, integrating the advantages of each: it facilitates variable selection, akin to Lasso, to yield sparse models, while also fostering a grouping effect, similar to Ridge, to manage highly correlated features. Concisely, Lasso is defined by Equation (5), and Elastic Net by Equation (6). It is evident that our proposed method represents an enhancement of the Elastic Net. Therefore, a comparison with Elastic Net and Lasso is well-justified. The relationship between our proposed method and Elastic Net and Lasso will be elaborated upon subsequently.

In this implementation, we use R2 evaluation metrics. R2 is a frequently used nomailzed metric in regression, which is calculated by Equation (10). It provides a comparison between predicted value and ground truth: when the model predict the results better, R2 of model is more tend to 1. Parameters are automatically determined by grid search on parameter list as Table 1.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{10}$$

Table 1. Parameter list

β_1	β_2	β_3
0.3	0.3	0.4
0.2	0.4	0.4
0.2	0.3	0.5
0.4	0.2	0.4
0.3	0.2	0.5
0.3	0.4	0.3

3.2 Case I: AMPds

3.2.1. Dataset

In the first case, we employ AMPds (The Annotated Meters for Power disaggregation Series) dataset to test proposed method. AMPds is a highly regarded and essential public dataset in the field of NILM research. Unlike many datasets where the main power signal is synthesized by summing sub-meter data, AMPds provides a directly measured main meter signal. This more accurately reflects the real-world NILM scenario (measurement only at the main panel), includes real-world noise like line losses, and makes experiments more convincing. More information can be refered to Tables 2 and 3.

Table 2. Basic Information

Information	Detail
Collection Period	A full year from April 1, 2012, to March 31, 2013.
Sampling Frequency	1 sample per minute
Measured Parameters	voltage,current, real power, active power, reactive power

Table 3. Measured Appliances and Description

Appliance/Circuit Type	Description
Main	The main electrical entrance to the house; the input signal for NILM algorithms.
Heat Pump	The primary heating device for the house; high-power with strong seasonal patterns.
A/C Fan & Thermostat	Cooling-related equipment used in the summer.
Kitchen Fridge	A cyclically operating device; a classic target for NILM.
Clothes Dryer	A high-power, staged resistive load.
Clothes Washer	A motor load with complex operational patterns.
Dishwasher	A mixed load (heater, motor) with a distinct operational cycle.
Instant Hot Water	A high-power load that activates instantaneously with water use.
Entertainment	An aggregated circuit for TV, amplifier, PVR, etc.; significant standby power.
Security Equipment	A continuously operating low-power load.
Various Plug & Lighting Circuits	(Bedrooms, Basement, Dining Room) Represent dispersed, aggregated general loads.
Garage, Utility Room	Include intermittently used tools and devices.

3.2.2. Results

As can be observed from Table 4 and Figure 2, proposed method has the better results than Elastic Net, Lasso and GSP. Globally, proposed method has about 5% advantage on R2 than Elastic Net, 7% advantage on R2 than Lasso, 4% advantage on R2 than GSP and 7% advantage on R2 than HMM. Moreover, this advantage will be larger in some specific appliance: we can see that in some complex scenario like Entertainment and Instant Hot Water, such advantage can be about 8%, which prove superior of proposed method.

Table 4. Experiment Results on AMPds

Appliance	Proposed	Elastic Net	Lasso	GSP	HMM
Heat Pump	0.95	0.91	0.88	0.92	0.90
A/C Fan & Thermostat	0.96	0.91	0.89	0.94	0.91
Kitchen Fridge	0.95	0.89	0.85	0.95	0.92
Clothes Dryer	0.99	0.94	0.90	0.96	0.95
Clothes Washer	0.94	0.86	0.84	0.95	0.95
Dishwasher	0.98	0.94	0.91	0.95	0.93
Instant Hot Water	0.99	0.89	0.86	0.97	0.94
Entertainment	0.96	0.88	0.84	0.96	0.95
Security Equipment	0.97	0.91	0.87	0.93	0.92
Various Plug & Lighting Circuits	0.94	0.90	0.90	0.89	0.88
Garage, Utility Room	0.97	0.93	0.89	0.81	0.82

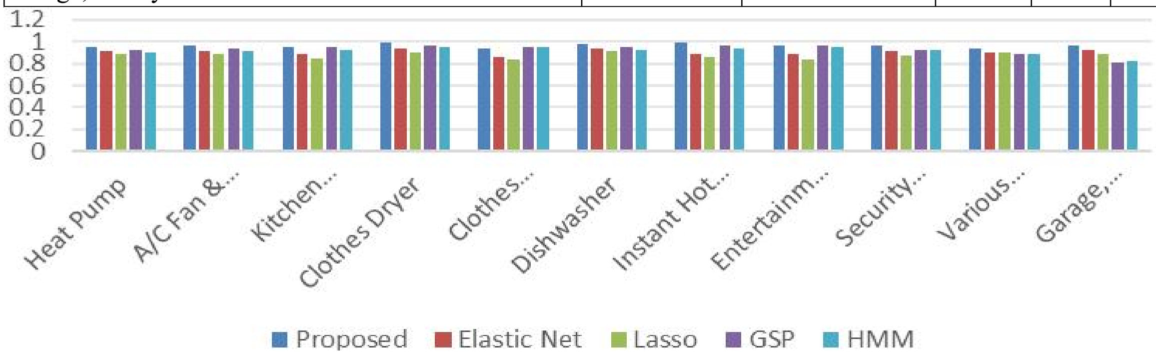

Figure 2. Visualization of Experiment Results on AMPds

3.3. Case II: UK-DALE

3.3.1. Dataset

UK-DALE is a publicly available dataset from the United Kingdom that captures Domestic Appliance-Level Electricity data for the development of disaggregation algorithms. Unlike datasets that only provide whole-house electricity consumption, UK-DALE includes ground-truth measurements for individual appliances. The data were collected between November 2012 and April 2017. The whole-household current and voltage waveforms were recorded at a high sampling rate of 16 kHz, while individual appliance power was sampled at 1/6 Hz (i.e., once every six seconds). For houses 1, 2, and 5, both apparent and real power of the whole household were recorded at 1-second intervals, in addition to the 16 kHz measurements. The dataset also includes a per-channel button press log that timestamps the ON/OFF events of individual appliances. In this paper, we use the six- second appliance-level data from House 1, focusing on a subset of five appliances for the purpose of energy disaggregation.

3.3.2. Results

Table 5. Experiment Results on UK-DALE

Appliance	Proposed	Elastic Net	Lasso	GSP	HMM
washing machine	0.90	0.88	0.83	0.88	0.85
dishwasher	0.92	0.90	0.84	0.92	0.91
kettle	0.88	0.85	0.81	0.87	0.86
fridge	0.90	0.84	0.80	0.86	0.84
microwave	0.81	0.79	0.77	0.87	0.85

As can be observed from Table 5 and Figure 3, proposed method has the better results than Elastic Net, Lasso and GSP. Globally, proposed method has about 4% advantage on R2 than Elastic Net, 7% advantage on R2 than Lasso, 2% advantage on R2 than GSP and 4% advantage on R2 than HMM. Although task in UK-DALE is more difficult in generel, our method still performs well. We will discuss the reason why our method can get better performance than Elastic Net and Lasso from a deep insight in later discussion section.

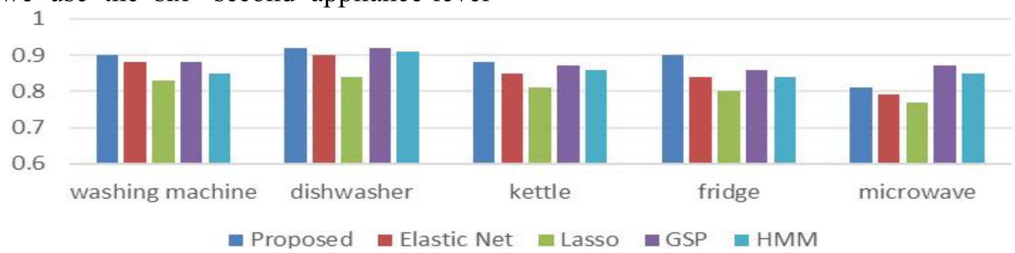

Figure 3. Visualization of Experiment Results on UK-DALE

4. Discussion

In this subsection, we present a comprehensive comparative discussion of the proposed method against two foundational regularization techniques: the Least Absolute Shrinkage and Selection Operator (Lasso) and the Elastic Net. This analysis is situated within the context of high-dimensional statistical learning, where models must be constructed from data characterized by a vast number of features relative to the number of observations—a common challenge in NILM.

The Lasso, introduced by Tibshirani [31], marked a paradigm shift in regression analysis by integrating variable selection with parameter estimation through the imposition of an L1-norm penalty. This penalty has the effect of driving the coefficients of less informative features to exactly zero, thereby producing sparse, interpretable models. Its computational efficiency and automatic feature selection capability offered a significant advantage over traditional and computationally intensive subset selection methods.

However, Lasso's reliance solely on L1 regularization exposes several well-documented limitations. First, when predictors are highly correlated, it tends to arbitrarily select one variable from the group while ignoring the others, leading to unstable and potentially non-robust models. Second, its capacity for variable selection is fundamentally constrained: in the canonical $p > n$ scenario, it can select at most n variables, which may be restrictive for complex, high-dimensional problems. These shortcomings catalyzed the development of the Elastic Net by Zou and Hastie [50], which synergistically combines L1 and L2 (ridge) regularization penalties. This hybrid approach was designed to retain Lasso's sparsity-inducing property while mitigating its key weaknesses. The advantages of Elastic Net over Lasso can be summarized as follows:

1) The grouping effect: The incorporated L2 penalty encourages a grouping property,

whereby highly correlated predictors are assigned comparable coefficient magnitudes. This leads to more stable and robust models, as correlated features tend to be either included or excluded from the model together, reflecting their collective relationship with the response variable.
2) Dual functionality: It effectively performs automatic variable selection (via the L1 component) while also managing multicollinearity and improving conditioning of the problem (via the L2 component).
3) Overcoming the dimensionality limit: The combined regularization allows the Elastic Net to select more than *n* features in high-dimensional settings, often yielding superior predictive performance compared to Lasso when the true underlying model involves numerous correlated predictors.
Consequently, Elastic Net presents itself as a viable and enhanced alternative to Lasso for regression problems inherent to NILM. Nonetheless, it is not without its own disadvantages. A primary critique, particularly relevant for applied domains like NILM, is that its grouping effect is driven purely by statistical correlations within the data, without incorporation of domain-specific prior knowledge. This can be a significant drawback. In the context of NILM, for example, the ideal grouping should align with physical appliances; basis vectors (features) corresponding to the same electrical device should be selected jointly. However, Elastic Net may incorrectly group basis vectors from different appliances that coincidentally exhibit similar power consumption patterns, simply due to statistical correlation, thereby compromising the physical interpretability and accuracy of the disaggregation result.
Building upon this analysis, our proposed method is designed as a targeted improvement over the Elastic Net framework. The core innovation lies in explicitly embedding domain knowledge to concretely define grouping structures—specifically, enforcing that basis vectors known a priori to belong to an individual physical appliance form an inseparable group during the model fitting process. This structured sparsity approach ensures that the selection mechanism operates on a per-appliance basis, rather than on statistically correlated but physically meaningless feature clusters. As a result, the method effectively reduces erroneous grouping across appliances with superficially similar consumption signatures, leading to a model that is both more accurate and more aligned with the physical reality of the electrical load. The superior performance of our method, as quantitatively demonstrated in Table 3, stems directly from this principled integration of domain structure, allowing it to significantly outperform both the standard Elastic Net and Lasso in the NILM regression task.

5. Conclusion

This work introduces a novel methodology for NILM based on a multi-constraint regression framework. Our proposed technique constitutes a significant advancement over the standard Elastic Net regression, which, while powerful, groups features based solely on statistical correlation. The core innovation of our approach lies in the explicit integration of domain-specific structural knowledge as prior constraints during the model fitting process. By enforcing that basis vectors corresponding to the same physical appliance are treated as a cohesive group, our method fundamentally redirects the learning objective from pure statistical fitting to physically meaningful disaggregation. This ensures the model's sparsity pattern and coefficient estimates are not only data-driven but also semantically aligned with the real-world configuration of electrical appliances, leading to enhanced interpretability and accuracy.
The efficacy of this method was rigorously evaluated using the publicly available AMPds and UK-DALE datasets. Experimental results robustly demonstrate that NILM based on our constrained regression framework is not only feasible but also markedly superior to conventional regularized regression baselines, including Lasso and the standard Elastic Net. The proposed method effectively mitigates the issue of erroneous cross-appliance grouping, a key limitation of its predecessors, thereby achieving more precise identification of individual load signatures. A comprehensive discussion of the results, including comparative performance metrics and ablation studies, substantiates the effectiveness and advantages of our model's design philosophy.
Main limitations of proposed method can be summarized as following:
1) Data scarcity: Supervised regression requires labeled data (appliance-level power), which is costly to acquire

2) Generalization: Models trained on one dataset may underperform in new environments due to varying appliance characteristics

So in future work, we will make more efforts on following aspects:

1) Semi-supervised and unsupervised learning: Leveraging unlabeled data via graph-based semi-supervised learning or clustering reduces reliance on annotations

2) Transfer learning: Pre-training models on large datasets (e.g., REFIT) before fine-tuning for specific households

3) Real-time applications: Lightweight models for edge computing in smart meters

In summary, this research presents a principled, knowledge-enhanced regression framework that advances the state-of-the-art in NILM by ensuring model estimates are both statistically sound and physically plausible, providing a solid foundation for more reliable and interpretable energy disaggregation.

References

[1] Araveeporn, A. (2021). The higher-order of adaptive Lasso and elastic net methods for classification on high dimensional data. Mathematics, 9(10), 1091. https://doi.org/10.3390/math9101091

[2] Azzini, H. A. D., Torquato, R., & da Silva, L. C. P. (2014). Event detection methods for nonintrusive load monitoring. In IEEE PES General Meeting Conference & Exposition, 1-5. https://doi.org/10.1109/PESGM.2014.6939797

[3] Breyer J, Alizai M H, Samant S, et al. Advanced Filtering of Unknown Devices in Event-Based NILM. ACM SIGENERGY Energy Informatics Review, 2024, 4(5): 8-16. https://doi.org/10.1145/3727200.3727204

[4] Cui, M., Han, D., Liu, H., Li, K.-C., Tang, M., Chang, C.-C., ..., & Guan, Y. L. (2024). Secure data sharing for consortium blockchain-enabled vehicular social networks. IEEE Transactions on Vehicular Technology, 73(12), 19682-19695. https://doi.org/10.1109/TVT.2024.3448207

[5] Dai, S., Meng, F., Wang, Q., & Chen, X. (2024). DP2-NILM: A distributed and privacy-preserving framework for non-intrusive load monitoring. Renewable and Sustainable Energy Reviews, 191, 114091. https://doi.org/10.1016/j.rser.2023.114091

[6] Dai, S., Wang, Q., & Meng, F. (2021). A telehealth framework for dementia care: An ADLs patterns recognition model for patients based on NILM. In International Joint Conference on Neural Networks, 1-8. https://doi.org/10.1109/IJCNN52387.2021.9534058

[7] Du, Z., Yin, B., Zhu, Y. et al. A NILM load identification method based on structured V-I mapping. Sci Rep 13, 21276 (2023). https://doi.org/10.1038/s41598-023-48736-8

[8] Ghosh S, Chatterjee A, Chatterjee D. Extraction of statistical features for type-2 fuzzy NILM with IoT enabled control in a smart home. Expert Systems with Applications, 2023, 212: 118750.https://doi.org/10.1016/j.eswa.2022.118750

[9] Ghosh S, Mitra A, Chakrabarti S, et al. Decomposition-transformation assisted optimized heterogeneous classification strategy in NILM. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-11.https://doi.org/10.1109/TIM.2023.3291008

[10] Han, D., Zhu, Y., Li, D., Liang, W., Souri, A., & Li, K.-C. (2022). A blockchain-based auditable access control system for private data in service-centric IoT environments. IEEE Transactions on Industrial Informatics, 18(5), 3530-3540. https://doi.org/10.1109/TII.2021.3114621

[11] Hao P, Yan Z, Liang J, et al. Self-Supervised Event Detection and Labeling Method for Non-Intrusive Load Monitoring. IEEE Transactions on Instrumentation and Measurement, 2025.https://doi.org/10.1109/TIM.2025.3613888

[12] Huang, P., Chen, X., Feng, X., Wang, Q., Zhang, D., Han, L., & Han, J. (2026). DecoupleNet: Domain-specific task decoupling network for low-light image enhancement. Pattern Recognition, 171, 112203. https://doi.org/10.1016/j.patcog.2025.112203

[13] Huang, Y., Liu, K., Chen, T., Xu, Y., & Ji, H. (2026). Deep intrinsic image decomposition via physics-aware neural networks. Pattern Recognition, 172, 112299. https://doi.org/10.1016/j.patcog.2025.112299

[14] Ibrahem, M. I., & Fouda, M. M. (2024). A lightweight privacy-preserving load forecasting and monitoring scheme supporting dynamic billing for smart grids: No KDC required. IEEE Internet of Things Journal, 11(19), 32160-32171. https://doi.org/10.1109/JIOT.2024.3426486

[15] Jiang, T., Li, Y., Li, Y., Xing, W., Yu, M., Xie, F., & Ta, D. (2026). A segmentation knowledge-based global-local attention network for tumor classification in breast ultrasound images. Pattern Recognition, 171, 112152. https://doi.org/10.1016/j.patcog.2025.112152

[16] Kaselimi, M., Doulamis, N., Doulamis, A., Voulodimos, A., & Protopapadakis, E. (2019). Bayesian-optimized bidirectional LSTM regression model for non-intrusive load monitoring. In 2019 IEEE International Conference on Acoustics, Speech and Signal Processing, 2747-2751. https://doi.org/10.1109/ICASSP.2019.8683110

[17] Lam, H. Y., Fung, G., & Lee, W. (2007). A novel method to construct taxonomy electrical appliances based on load signaturesof. IEEE Transactions on Consumer Electronics, 53(2), 653-660. https://doi.org/10.1109/TCE.2007.381742

[18] Li, X., Zhao, B., Luan, W., & Liu, B. (2022). A training-free non-intrusive load monitoring approach for high-frequency measurements based on graph signal processing. In Asia Conference on Power and Electrical Engineering, 859-863. https://doi.org/10.1109/ACPEE53904.2022.9783715

[19] Lu, J., Yu, Z., Zhang, J., Zhao, R., Li, K., & Zeng, K. (2023). Identification of appliance V-I trajectory based on hash image retrieval method. In Power System and Green Energy Conference, 1146-1151. https://doi.org/10.1109/PSGEC58411.2023.10255957

[20] Luan, W., Yang, F., Zhao, B., & Liu, B. (2022). Industrial load disaggregation based on Hidden Markov Models. Electric Power Systems Research, 210, 108086. https://doi.org/10.1016/j.epsr.2022.108086

[21] Ma, R., Kang, W., Sha, J., Liu, J., & Zhu, D. (2024). NILM load identification based on CNN and V-I trajectory. In International Conference on Power Science and Technology, 2079-2084. https://doi.org/10.1109/ICPST61417.2024.10601881

[22] Mitra, A., Ghosh, S., Mohapatra, A., & Chakrabarti, S. (2024). Appliance identification via combinatorial fusion analysis-assisted Bayesian-optimized classifier. IEEE Transactions on Smart Grid, 15(1), 1165-1176. https://doi.org/10.1109/TSG.2023.3286949

[23] Musib, L., Coletti, R., Lopes, M. B., Mouriño, H., & Carrasquinha, E. (2024). Priority-Elastic net for binary disease outcome prediction based on multi-omics data. BioData Mining, 17(1), 45. https://doi.org/10.1186/s13040-024-00401-0

[24] Pereira, L., & Nunes, N. (2018). Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—A review. WIREs Data Mining and Knowledge Discovery, 8(6), e1265. https://doi.org/10.1002/widm.1265

[25] Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. Energy and Buildings, 40(3), 394-398. https://doi.org/10.1016/j.enbuild.2007.03.007

[26] Pötter, H., Lee, S., & Mossé, D. (2021). Towards privacy-preserving framework for non-intrusive load monitoring. In Proceedings of the Twelfth ACM International Conference on Future Energy Systems, 259-263. https://doi.org/10.1145/3447555.3464873

[27] Regan, J., Saffari, M., & Khodayar, M. (2022). Deep attention and generative neural networks for nonintrusive load monitoring. The Electricity Journal, 35(5), 107127. https://doi.org/10.1016/j.tej.2022.107127

[28] Reka, S. S., Dragicevic, T., Venugopal, P., Ravi, V., & Rajagopal, M. K. (2024). Big data analytics and artificial intelligence aspects for privacy and security concerns for demand response modelling in smart grid: A futuristic approach. Heliyon, 10(15), e35683. https://doi.org/10.1016/j.heliyon.2024.e35683

[29] Song, D., Yan, Q., Wang, H., Yang, J., & Luo, X. (2026). Wavelet-guided diffusion enhancement network with directional learning for single-pixel imaging. Pattern Recognition, 172, 112322. https://doi.org/10.1016/j.patcog.2025.11232

2

[30] Tanoni, G., Principi, E., & Squartini, S. (2024). Non-intrusive load monitoring in industrial settings: A systematic review. Renewable and Sustainable Energy Reviews, 202, 114703. https://doi.org/10.1016/j.rser.2024.114703

[31] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[32] Wang, A. L., Chen, B. X., Wang, C. G., & Hua, D. (2018). Non-intrusive load monitoring algorithm based on features of V–I trajectory. Electric Power Systems Research, 157, 134-144. https://doi.org/10.1016/j.epsr.2017.12.012

[33] Wang, H., Zhang, J., Lu, C., & Wu, C. (2021). Privacy preserving in non-intrusive load monitoring: A differential privacy perspective. IEEE Transactions on Smart Grid, 12(3), 2529-2543. https://doi.org/10.1109/TSG.2020.3038757

[34] Xiao, Y., & Xia, Y. (2026). A novel image enhancement method based on image decomposition and deep neural networks. Pattern Recognition, 172, 112371. https://doi.org/10.1016/j.patcog.2025.112371

[35] Wu, X., Guo, Y., Yan, M., Li, X., Yao, L., & Gong, G. (2023). Non-intrusive load monitoring using identity library based on structured feature graph and group decision classifier. IEEE Transactions on Smart Grid, 14(3), 1958-1973. https://doi.org/10.1109/TSG.2022.3209213

[36] Deng, X.-P., Zhang, G.-Q., Wei, Q.-L., Peng, W., & Li, C.-D. (2022). Fēi qīnrù shì fùhè jiāncè zòngshù [A survey on the non-intrusive load monitoring]. Acta Automatica Sinica, 48(3), 644-663. https://doi.org/10.16383/j.aas.c200270

[37] Yin, H., Zhou, K., & Yang, S. (2023). Non-intrusive load monitoring by load trajectory and multi-feature based on DCNN. IEEE Transactions on Industrial Informatics, 19(10), 10388-10400. https://doi.org/10.1109/TII.2023.3240924

[38] Yu, Z., Zhao, R., Jia, R., Zheng, W., Zhang, J., & Li, S. (2023). Appliance V-I trajectory detection for NILM based on conditional autoencoder. In IEEE Sustainable Power and Energy Conference, 1-5. https://doi.org/10.1109/iSPEC58282.2023.10402911

[39] Yu, Z., Yao, X., & Liu, H. (2025). Bayesian Elastic Net variable selection and application for spatial quantile panel autoregressive model. Communications in Statistics - Simulation and Computation. Advance online publication. https://doi.org/10.1080/03610918.2025.2450698

[40] Yuan, J., Jin, R., Wang, L., & Wang, T. (2024). A nonintrusive load identification method based on dual-branch attention GRU fusion network. IEEE Transactions on Instrumentation and Measurement, 73, 1-10. https://doi.org/10.1109/TIM.2024.3413200

[41] Zhang, X., Zhu, H., Li, X., Hou, B., Zhao, W., Yi, X., ..., & Jiao, L. (2026). Recurrent progressive fusion-based learning for multi-source remote sensing image classification. Pattern Recognition, 171, 112284. https://doi.org/10.1016/j.patcog.2025.112284

[42] Zhang, Y., Wu, H., Ma, Q., Yang, Q., & Wang, Y. (2023). A learnable image-based load signature construction approach in NILM for appliances identification. IEEE Transactions on Smart Grid, 14(5), 3841-3849. https://doi.org/10.1109/TSG.2023.3239598

[43] Zhao, J., Han, M., Ma, A., Jiang, F., Chen, R., Dong, Y., ..., & Chen, Y. (2024). A machine vision-assisted Argonaute-mediated fluorescence biosensor for the detection of viable Salmonella in food without convoluted DNA extraction and amplification procedures. Journal of Hazardous Materials, 466, 133648. https://doi.org/10.1016/j.jhazmat.2024.133648

[44] Zhao, F., Zhou, J., Ren, Y., Liu, H., Wang, Y., & Xia, W. (2023). Mean teacher for semi-supervised non-intrusive load monitoring based on V-I trajectory. In Asia Conference on Power and Electrical Engineering, 1623-1627. https://doi.org/10.1109/ACPEE56931.2023.10135645

[45] Zhao, R., Lu, J., Liu, B., Yu, Z., Ren, Y., & Zheng, W. (2023). Non-intrusive load identification method based on self-supervised regularization. IEEE Access, 11, 144696-144704.

https://doi.org/10.1109/ACCESS.2023.3337385

[46] Zhao, R., Shi, B., Ruan, J., Pan, T., & Dong, B. (2024). Estimating noisy class posterior with part-level labels for noisy label learning. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, 22809-22819. https://doi.org/10.1109/CVPR52733.2024.02152

[47] Zhong, M., Goddard, N., & Sutton, C. (2015). Latent Bayesian melding for integrating individual and population models. In International Conference on Neural Information Processing Systems, 2, 3618-3626.

[48] Zhou, K., Zhang, Z., & Lu, X. (2024). Non-intrusive load monitoring based on an efficient deep learning model with local feature extraction. IEEE Transactions on Industrial Informatics, 20(7), 9497-9507. https://doi.org/10.1109/TII.2024.3383521

[49] Zhu, Z., Wei, Z., Yin, B., Zhang, S., & Wang, X. (2017). A novel approach for event detection in non-intrusive load monitoring. In 2017 IEEE Conference on Energy Internet and Energy System Integration, 1-5. https://doi.org/10.1109/EI2.2017.8245715

[50] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology, 67(2), 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x