



Large-Scale Heat Pump Energy Disaggregation

01. Context

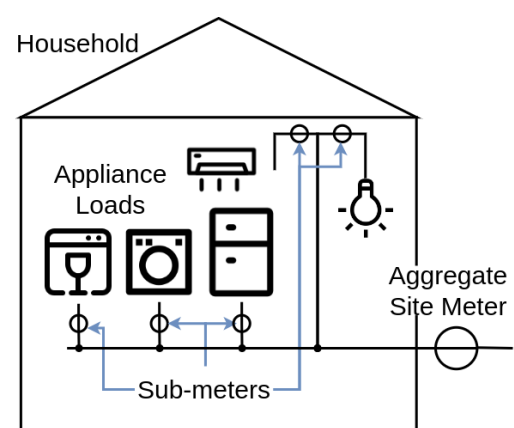
The transition to a lower carbon economy will put pressure on the electrical power infrastructure in the following two key ways:

- Intermittent generation means that providing an interrupted service at low cost will be challenging
- More loads such as electric vehicles (EVs) and heat-pumps will cause increasing congestion and voltage excursions on the network.

Incentivise changes to consumer behaviour through a flexibility market would allow demand to shape dynamically around network constraints and energy availability. Balancing investment in flexibility services such as Demand Side Response (DSR) network reinforcement requires an understanding of the potential DSR capacity of a given region of network. This information could, of course, be gained through means such as surveying - however, developments in the field of Non Intrusive Load Monitoring may also be relevant to this problem.

02. Non Intrusive Load Monitoring

Non Intrusive Load Monitoring (NILM) or non-intrusive appliance load monitoring is the process of estimating load profiles from an aggregated measurement of current voltage data.



Though the idea of NILM has been in the literature since Hart's seminal 1981 paper, recent developments in artificial neural networks (ANN) have led to substantial improvements in this area.

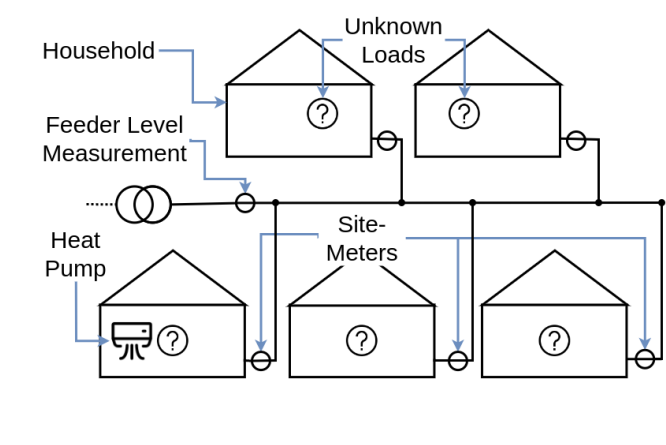
Figure 1. Typical Household NILM problem, Adapted from Hart 1991

03. Aims

NILM has traditionally been targeted towards disaggregating single appliances from single households at sample periods ranging between ~1s-3min. The key distinction between this type of problem and what future system operators might face is: A) Lower sample period on the order of 15mins smart meters B) Additional household's load at the feeder level C) Multiple devices of the same class downstream of a given measurement D) No prior knowledge of the type of loads under a given measurement point This work aims to investigate points A) and B) for heat-pumps by testing the performance of a subset of NILM algorithms presented in the literature under:

- Increasing numbers of homes in aggregate (henceforth increasing power aggregation)
- Decreasing sample rate (henceforth increasing temporal aggregation)

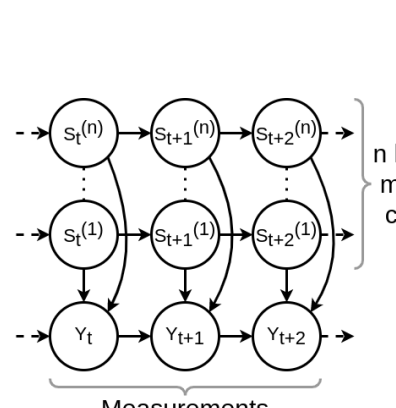
Figure 2. Example of the "Large Scale" NILM that arises from the measurement points that the system operator might have access to on a typical feeder system.



04. Algorithms Tested

Mean Disaggregation (Mean)

This algorithm predicts by taking the mean of the power in the training set over all time and simply predicting a consistent on state at this magnitude. Though this is an exceedingly primitive approach it is included as a usefully benchmark for other algorithms.



Factorial Hidden Markov Model (FHMM)

In a Factorial Hidden Markov Model (FHMM) a set of observations (Y_t) are dependent on a set of independent Markov chains representing the state of each appliance. Each state in the model represents a single output power, as such a FHMM cannot represent a continuously variable device.

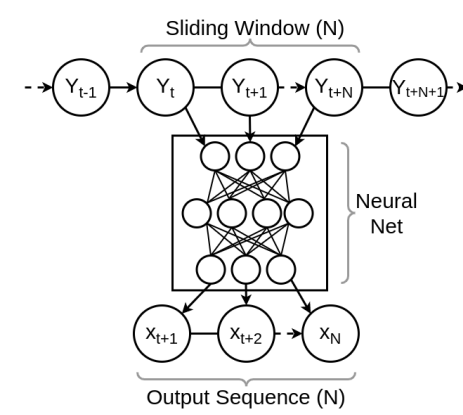
Combinatorial Optimisation (CO)

With Combinatorial Optimisation (CO) the disaggregation problem is formulated as a set of appliances each with a finite set of states corresponding to constant power levels. The algorithm aims to find the combination of these discrete states that minimizes the difference between the household aggregate and the sum of the power for each appliance.

De-noising Auto-encoder (DAE)

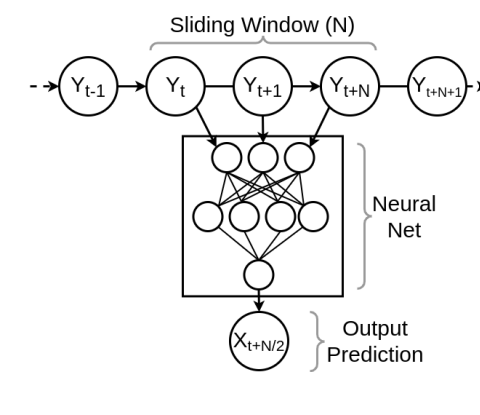
The de-noising auto-encoder is a class of neural network which attempts to reconstruct its input given an assumed source of corruption. In the case of NILM, a fixed window of the aggregate measurement is assumed to be

the corrupted signal and is then while the target appliance load profile is the desired uncorrupted signal.



Sequence to Sequence (Seq2Seq)

In sequence to sequence, a neural network maps a sliding window of the aggregate site meter to a sequence representing the target appliance load profile.



Sequence to Point (Seq2Point)

As before, a neural network operates on a sliding window of the aggregate site meter, this time outputting a single point prediction of the target appliance load profile in the centre of the time window.

05. Tools

NILMTK

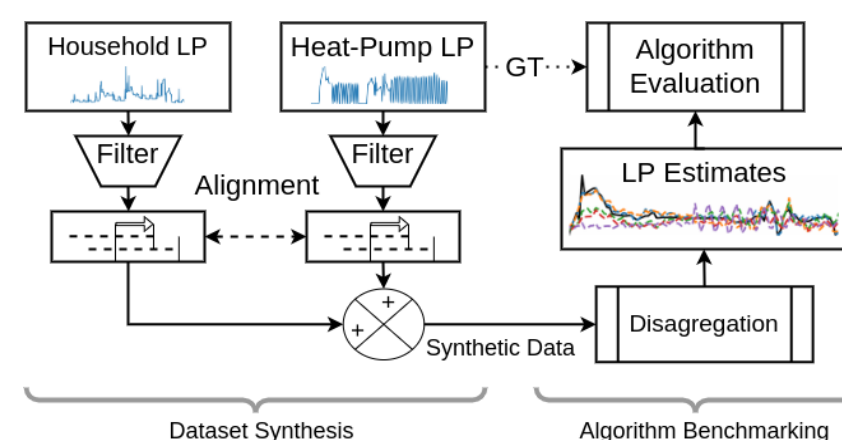
The non intrusive monitoring toolkit (NILMTK) is an open source software project aimed at providing a set of tools, test data and unified framework for benchmarking NILM algorithms. All NILM algorithms tested in this work were made available by the respective researcher in the NILMTK.

Python

The NILMTK is written exclusively in Python, and beyond its other advantages it makes this the obvious choice for tasks such as data processing and analysis and presentation.

06. Implementation

Figure 3. High level flow-chart of the algorithm evaluation methodology, "Load Profile" simplified to LP.



Dataset Synthesis

For both the training and benchmarking of these algorithms, test data is needed which includes both the aggregate 'site-meter' measurement as well as a sub-meter measurement just upstream of the heat pump. No suitable dataset was found from the existing NILM datasets, thus data has been synthesized by combining heat-pump and domestic load-profiles.

In total five datasets have been created each made up of five synthetic load profiles at a particular power aggregation level, i.e. number of households in aggregate. A given synthetic load profile is generated by aligning and element-wise summing the heat pump load profile (P) with N household load profiles (H) where N defines the power aggregation level. ($y_t \in Y$) is the synthetic 'site-meter' aggregate measurement at time (t), ($h_{n,t} \in H_n$) is the household measurement, and ($p_t \in P$) is the heat pump measurement at time t.

$$Y_{t,N} = \sum_{n=1}^{n=N} H_{n,t} + P_t$$

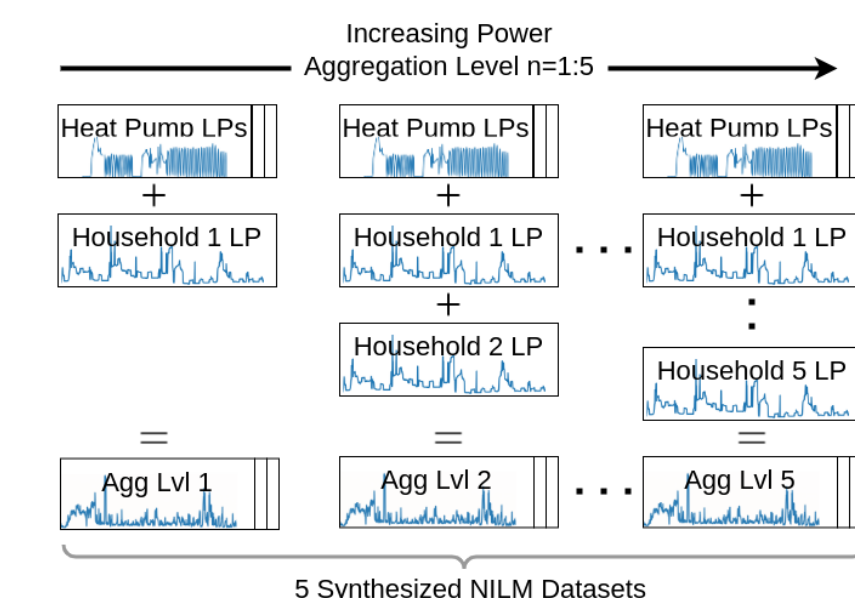


Figure 4. Structure of synthesized NILM datasets

Algorithm Training & Evaluation

Training

The algorithms were trained on a three month span of January-March and benchmarked on the entirety of April.

Increasing Temporal Aggregation

Sweeps with increasing temporal aggregation are generated using the down-sampling feature of the NILMTK with sample periods from 1 to 15 minutes.

Increasing Power Aggregation

Increasing power aggregation is evaluated by running once on each of the five aggregation level datasets at a fixed sample period of 2 mins.

Performance Metric - Mean Normalised RMSE (MNRMSE)

Each algorithm is evaluated by calculating the Root Means Squared Error (RMSE) between the prediction and ground truth for each algorithm. RMSE is normalised against

the RMSE of the Mean algorithm, yielding a valid comparison metric between appliances. With MNRMSE=1 as the maximum attainable value, 0 being the same as Mean and, negative values represent poorer performance than Mean.

$$MNRMSE_{Alg} = \frac{RMSE_{Mean} - RMSE_{Alg}}{RMSE_{Mean}}$$

07. Results

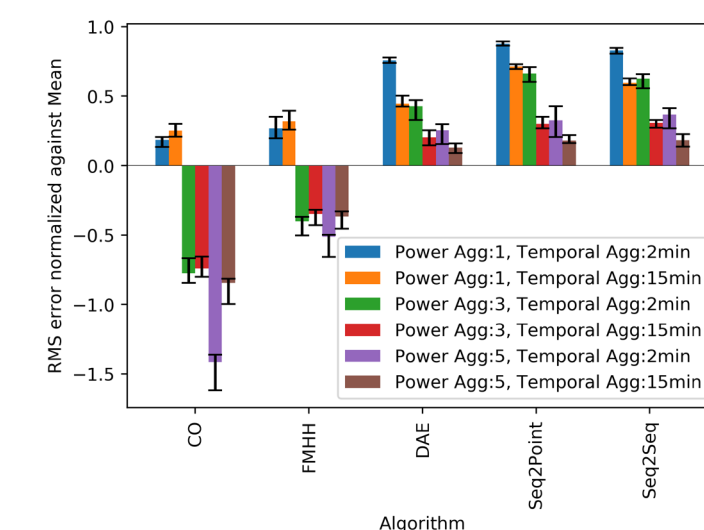
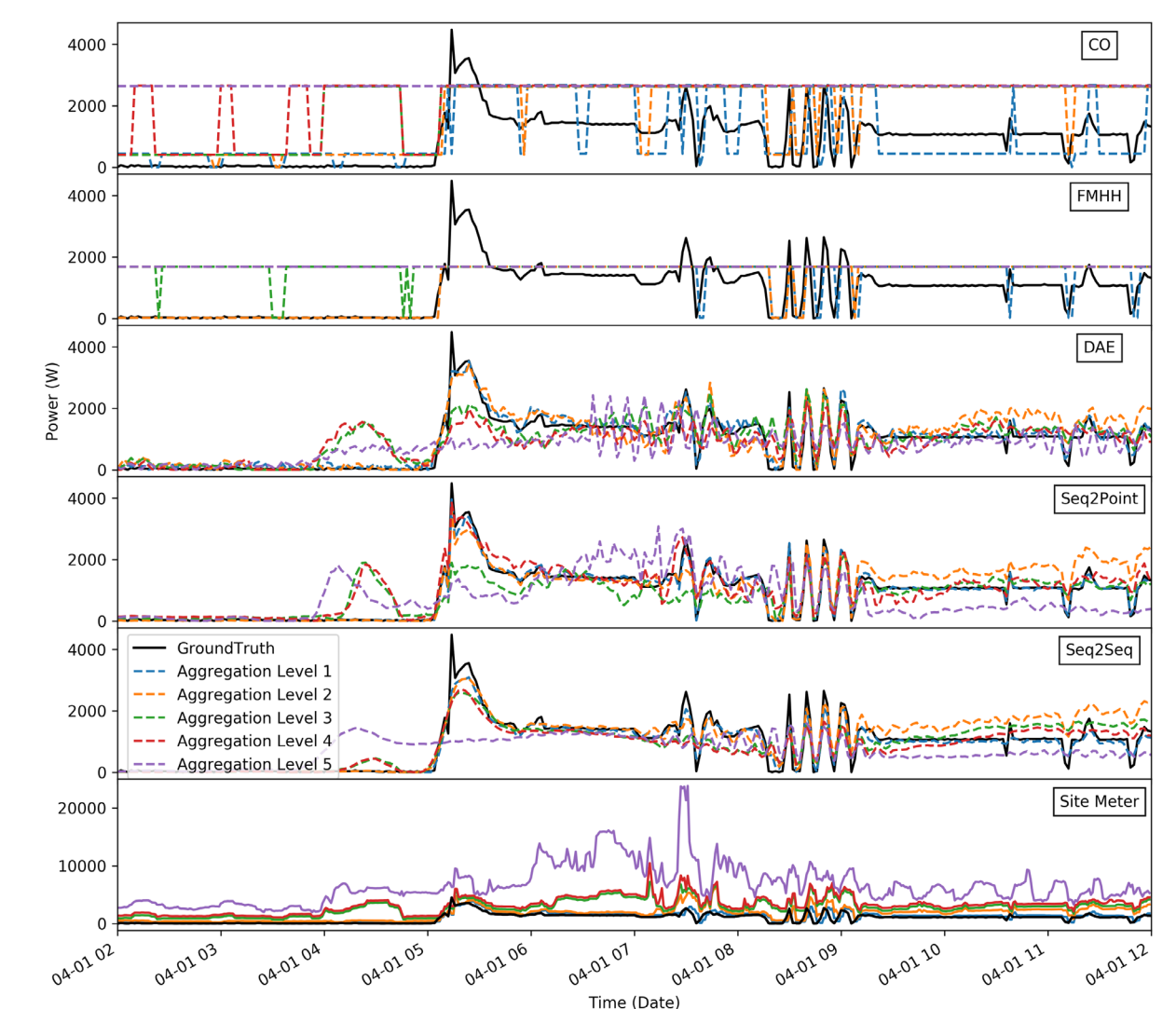


Figure 5 (above). Sample of disaggregation algorithm predictions at a variety of power aggregation levels, ground truth is plotted in black, whilst the final plot shows total feeder aggregate measurement at each aggregation level.

Figure 6 (left). Average normalized performance for algorithms at selected temporal & power aggregation levels, Aggregation shortened to Agg.

Algorithm performance over all aggregation scenarios is summarized in Figure 5 with a subset of results at various levels of both power and temporal aggregation.

Sweeps over increasing aggregation level are shown in Figure 6 with the aggregation plotted in SI units as the dependant variable while the other type is kept constant (CO and FHMM omitted due to poor performance).

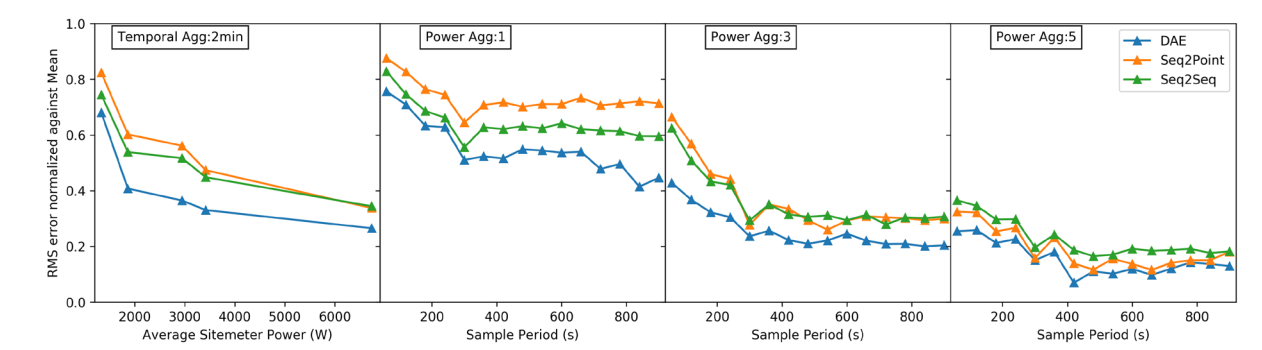


Figure 7. MNRMSE algorithm performance at increasing aggregation levels.

08. Discussion & Conclusions

CO & FHMM

Since CO and FHMM rely on a finite set of states, poor performance with a continuously variable load partially expected. It is noted that these algorithms appear more sensitive to power aggregation than temporal aggregation. Both algorithms quickly drop below 0 implying that they offer no predicting power under these conditions.

Neural Network based Algorithms

The neural network based algorithms performed collectively substantial better, seq2point performing on average the best. In both cases, an initial rapid drop in performance is observed before settling. At the highest power aggregation, the sensitivity to temporal aggregation, seems to be much lower.

Conclusions

- Neural Networks based disaggregation algorithms, specifically seq2seq/seq2point seems to hold the greatest promise for heat pump disaggregation for both high temporal and power aggregation
- High performance is observed for seq2point/seq2seq in a scenario replicating that of a typical domestic smart meter (high temporal, low power aggregation)
- As expected, poorest performance is observed at the combination of high temporal and power aggregation
- Performance is substantially more sensitive to power than temporal aggregation