Large Scale Heat-pump Non-Intrusive Load Monitoring

Ben Frazer*

August 14, 2022

Contents

---

*2704250F@student.gla.ac.uk

1   Introduction

1.1   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. National Grid hopes that the newly introduced role of Distribution System Operator (DSO) will facilitate competing services such as flexibility and reinforcement on the distribution system to drive down costs.

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 achieved though intrusive means such as surveying however developments in the field of Non Intrusive Load Monitoring may also be relevant to this problem.

1.2   NILM

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 1991 paper, recent developments in artificial neural networks (ANN) have lead to substantial improvements in this area.

Typical Household NILM problem, Adapted from Hart 1991
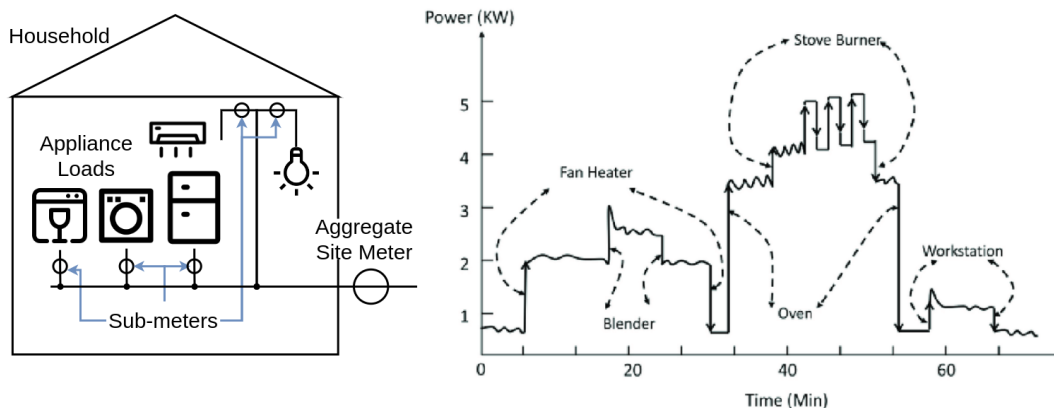
1.3   Aims

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 DSO's might face is:

1. Much lower sample period typically on the order of 15mins from existing smart meters
2. Higher power aggregation level (i.e. more households) simulating a feeder level measurement
3. Multiple devices of the same class downstream of a given measurement
4. No prior knowledge of the type of loads under a given measurement point

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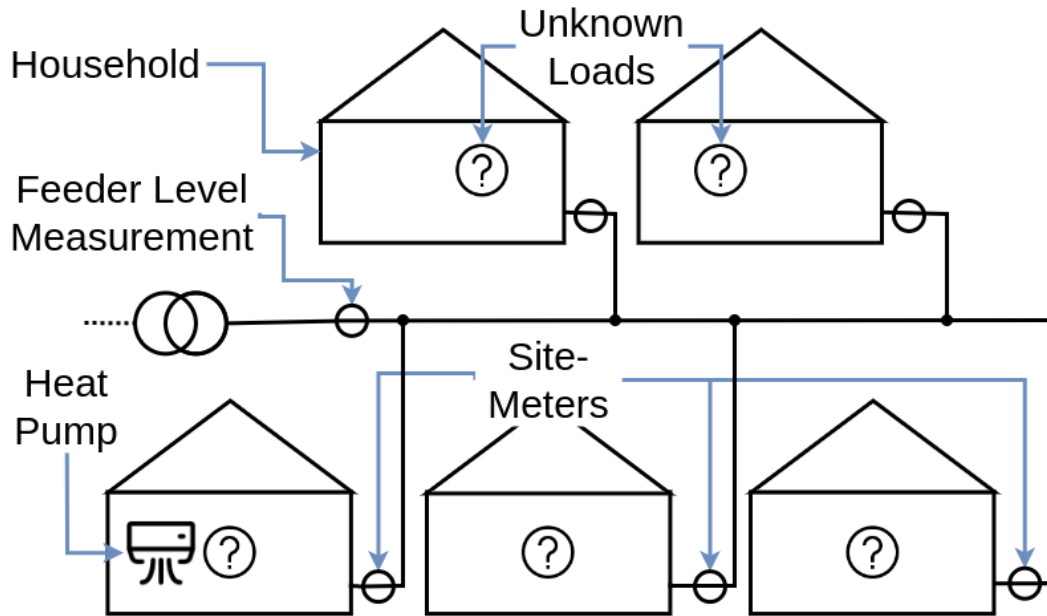

This work aims to investigate points A) and B) for heat-pumps by testing the performance of a subset of NILM algorithms presented in the literature under:

- Increasing power aggregation levels
- Increasing temporal aggregation levels i.e. decreasing sample rate

2   State of the art

Sample of Disaggregation algorithms estimating a heat-pump load profile at a variety of power aggregation levels, the true appliance ground truth is plotted in black, whilst the lowest plot shows the total feeder level aggregation at each "Level"
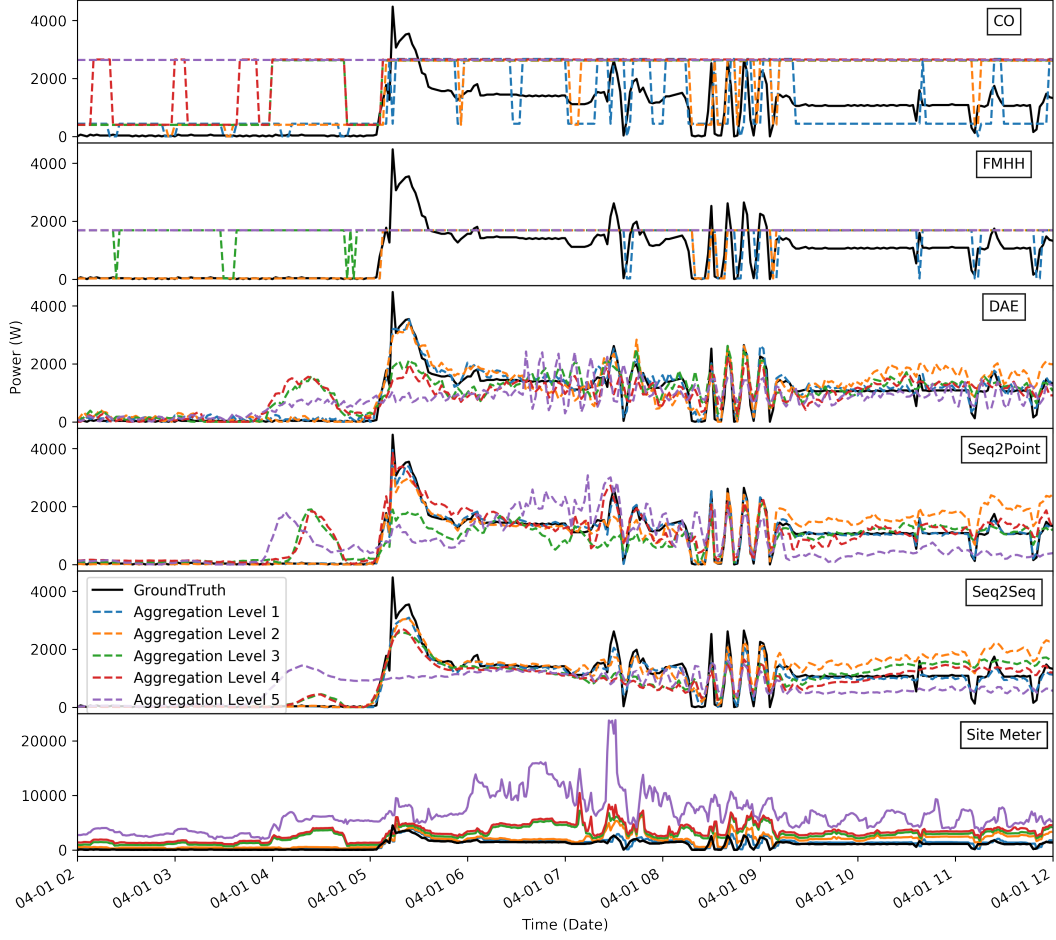
2.1   Algorithms

2.1.1   Mean Disaggregation (Mean)

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

2.1.2   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 repents a single output power, as such a FHMM cannot represent a continuously variable device.
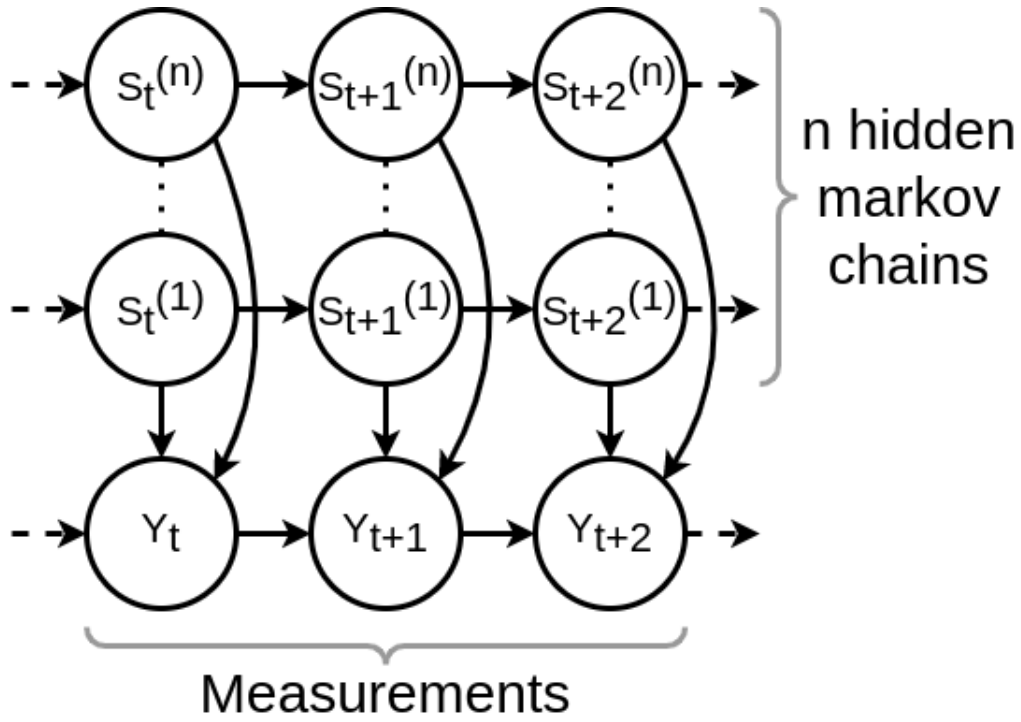
- measurements written wrong

2.1.3 Combinatorial Optimization (CO)

With Combinatorial Optimization (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.

### 2.1.4 Sequence to Sequence (Seq2Seq)

In sequence to sequence, a neural network maps a sliding window of the aggregate site meter to a window of the target appliance.
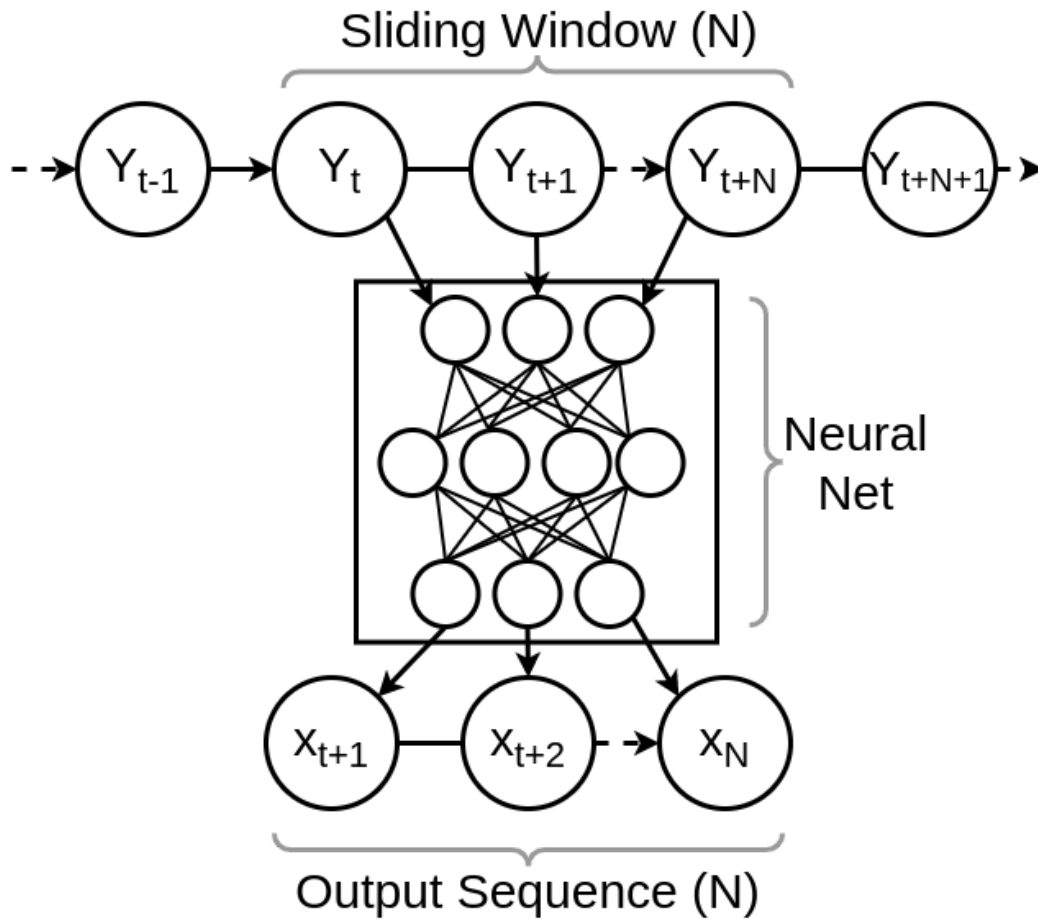
2.1.5 Sequence to Point (Seq2Point)

As above a neural network operates on a sliding window of the aggregate site meter, this time outputting a single point predictions of the target appliance in the centre of the time window.
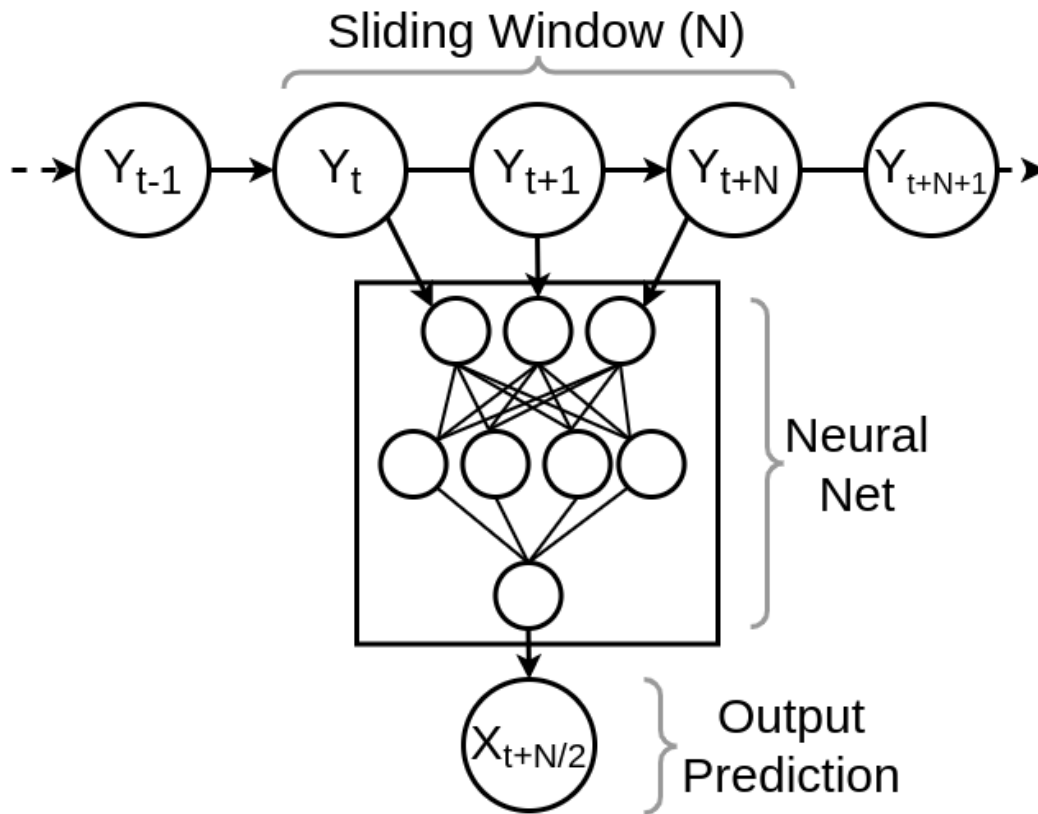
2.1.6   De-noising Auto-encoder (DAE)

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.

2.2   Tools

2.2.1   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.

2.2.2   Python

The NILMTK is written exclusively in Python, and beyond it's other advantages it makes this the obvious choice for tasks such as data processing and analysis and presentation. In particular heavy use was made of the following non core python libraries:

- Pandas - Data processing, filtering and synthesis
- Numpy - as above
- Matplotlib - Data Presentation (static figures)
- Plotly - interactive plots on github

3   Implementation

High level flow-chart of the algorithm evaluation methodology
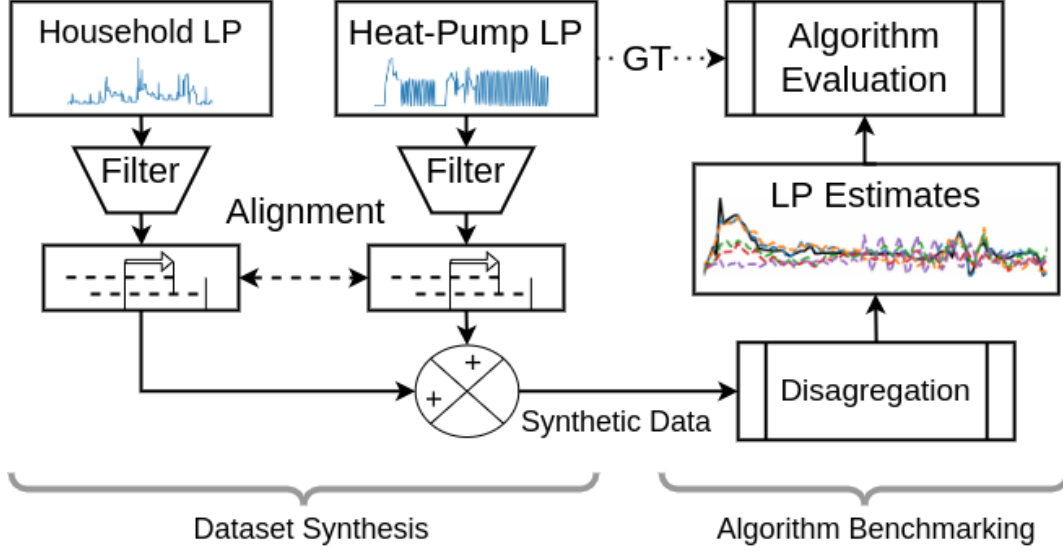

3.1   Dataset synthesis

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 containing five synthetic load profiles each with a singe heat pump at a perticular same power aggregation level. The data for a given synthetic load profile is generated by aligning and element-wise summing the heat pump load profile set P with N household load profiles H where N defines the power aggregation level.

$$y_{t,N} = \sum_{n=1}^{n=N} h_{n,t} + p_t \tag{3.1}$$

$y_t \in Y$ 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.

Structure of synthesized NILM datasets

Increasing Aggregation
Level n=1:5

5 Synthesized NILM Datasets
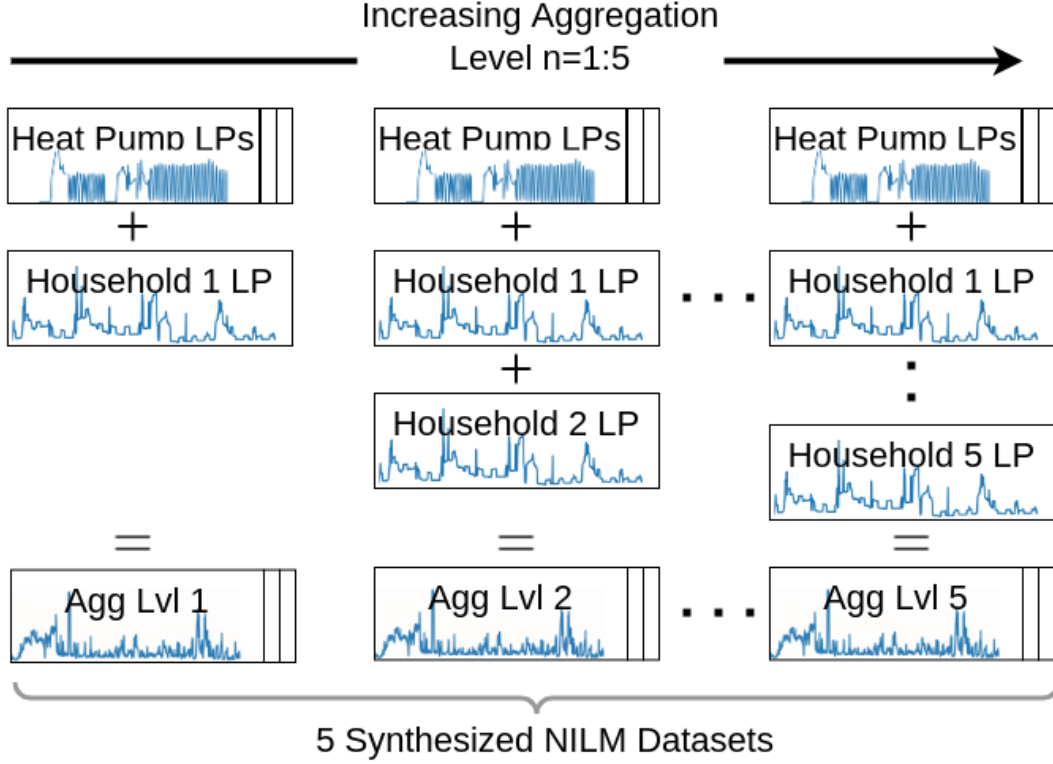
3.2   Algorithm Training and Evaluation

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

3.2.1   Increasing Temporal Aggregation

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

3.2.2   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.

### 3.2.3   Performance Metric - Mean Normalized RMSE (MNRMSE)

Each algorithm is evaluated by calculating the RMSE of the prediction vs the heat-pump ground truth for each algorithm. RMSE is normalized against the RMSE of by the Mean algorithm for that same household, yielding a valid comparisons metric between appliances. With MNRMSE 1 as the maximum attainable value, 0 being the same performance as Mean and, negative values represent poorer performance than Mean.

$$\text{MNRMSE}_{alg} = \frac{\text{RMSE}_{Mean} - \text{RMSE}_{alg}}{\text{RMSE}_{Mean}} \tag{3.2}$$

4 Results

Algorithm performance is summarized in figure 4 by comparing a subset of results at various levels extremes of aggregation. Each bar represents an algorithms average MNRMSE for a given temporal and power aggregation with error bars display the variability between heat-pumps.

Average normalized performance for all algorithms at various temporal & power aggregation levels
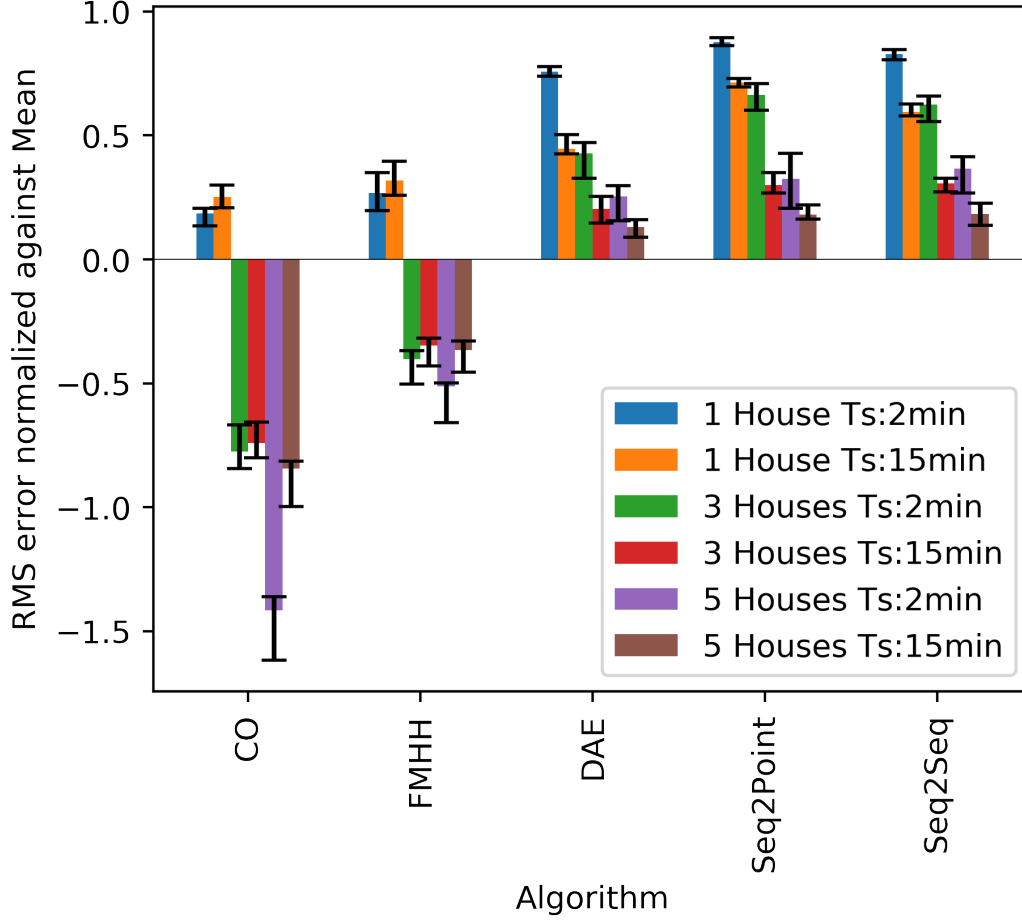
Sweeps of aggregation levels are shown in fig with a given aggregation type plotted as the dependant variable while the other type is kept at a constant value (CO and FHMM omitted due to poor performance).

Average MNRMSE performance per algorithm at (left) increasing power (right) increasing Temporal aggregation levels
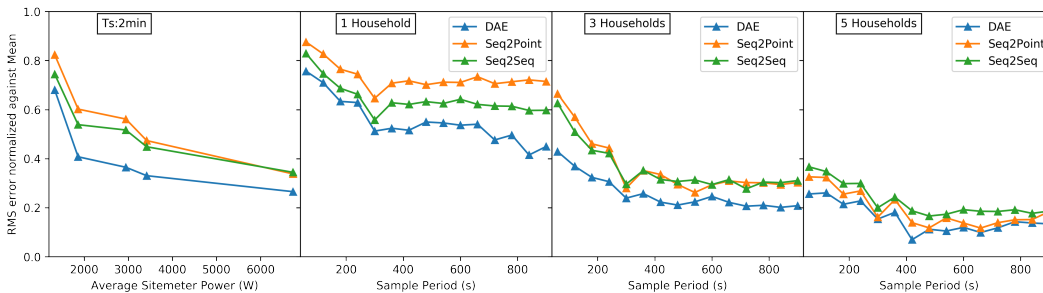
4.1 Discussion & conclusions

4.1.1 CO & FHMM

Since CO and FHMM rely on a finite set of states the poor performance with a continuously variable load partially expected, however in both cases the dominant errors arise from miss-classification of state. It is also noted that these algorithms appear more sensitive to power aggregation than temporal aggregation though in both cases both quickly drop below 0 implying that they offer no predicting power.

4.1.2 Neural Network based Algorithms

The neural network based algorithms performed collectively substantial better, seq2point performing on average the best, however being overtaken by seq2seq at the extreme. In both cases, an initial rapid drop in performance is observed before setting. At the highest power aggregation, the effect of increasing temporal aggregation, seems to be much lower and settles rapidly slightly above mean disaggregation implying there.

4.1.3 Conclusions

- Neural Networks based disaggregation algorithms specifically seq2seq/seq2point seems to hold the greatest promise for large-scale heat pump disaggregation
- For disaggregation in the case of a small feeder power aggregation has the greatest influence on performance, though temporal aggregation still has a substantial impact, especially at higher power aggregation
- Worst performance is observed as

5 Links

NILMTK `https://github.com/nilmtk/nilmtk`
Fork Of NILMTK `https://github.com/BenjaminFrazer/nilmtk`
Code and Data `https://github.com/BenjaminFrazer/thesis_tools`

Emacs 28.1 (Org mode 9.6)