

Large Scale Heat-pump Non-Intrusive Load Monitoring

Ben Frazer*

August 12, 2022

Contents

1	Introduction	2
1.1	Context	2
1.2	NILM	2
1.3	Aims	2
2	State of the art	4
2.1	Algorithms	4
2.1.1	Mean Disaggregation (Mean)	4
2.1.2	Factorial Hidden Markov Model (FHMM)	4
2.1.3	Combinatorial Optimization (CO)	5
2.1.4	Sequence to Sequence (Seq2Seq)	5
2.1.5	Sequence to Point (Seq2Point)	6
2.1.6	De-noising Auto-encoder (DAE)	7
2.2	Tools	7
2.2.1	NILMTK	7
2.2.2	Python	7
3	Implementation	8
3.1	Dataset synthesis	8
3.1.1	Mixed Dataset structure	8
3.1.2	Increasing power aggregation dataset structure	9
3.2	Algorithm Training and Evaluation	9
3.2.1	Increasing Temporal Aggregation	9
3.2.2	Increasing Power Aggregation	9
3.2.3	Performance Metric - Mean Normalized RMSE (MNRMSE)	9
4	Results	10
4.1	Discussion & conclusions	11
4.1.1	CO & FHMM	11
4.1.2	Neural Network based Algorithms	11
4.1.3	Conclusions	11
5	Links	11

*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.

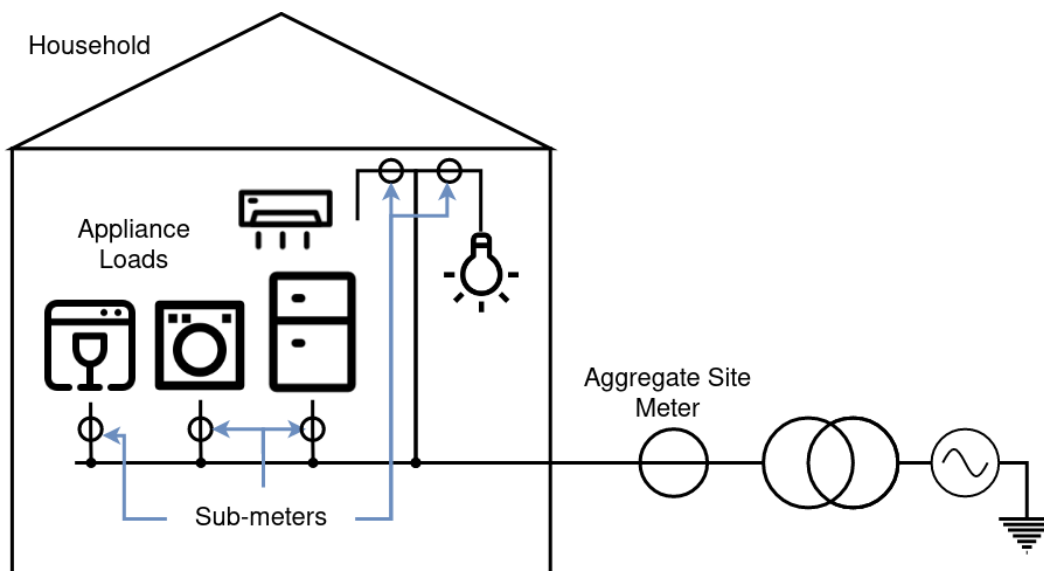
Balancing investment in flexibility services such as demand side response (DSR) network reinforcement requires an understanding of the potential DSR capacity of a given region of network. This information could of course be achieved through intrusive means such as surveying however developments in the 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. Non intrusive - because there are no potentially inconvenient sub-meters on every appliance of interest, rather an unintrusive single external measurement.

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.

Typical Household NILM problem

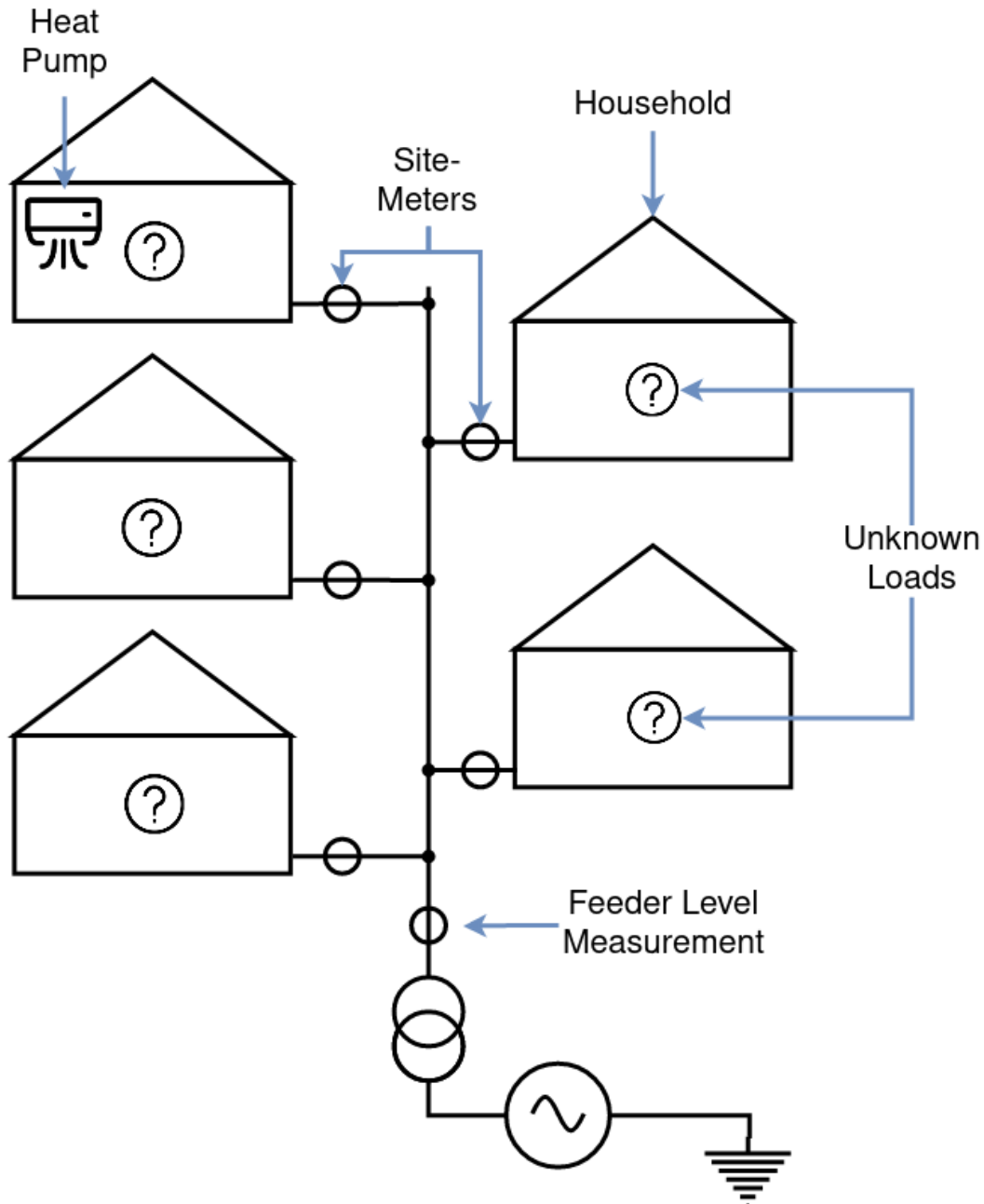


1.3 Aims

NILM has traditionally been targeted towards disaggregating single appliances from single households at sample periods ranging between $\sim 1s-3min$. 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

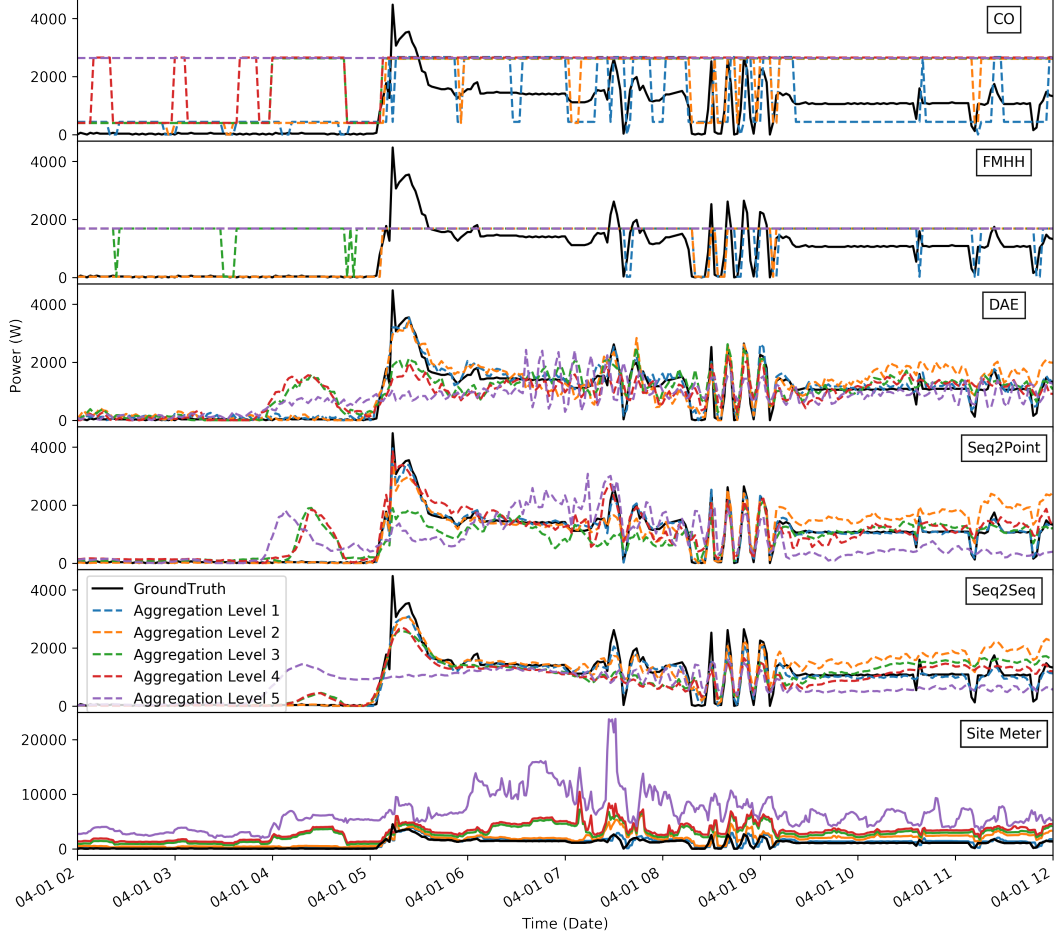


This work aims to investigate points A) and B) for heat-pumps by testing the performance of a subset of NILM algorithms presented in the literature under:

- 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”



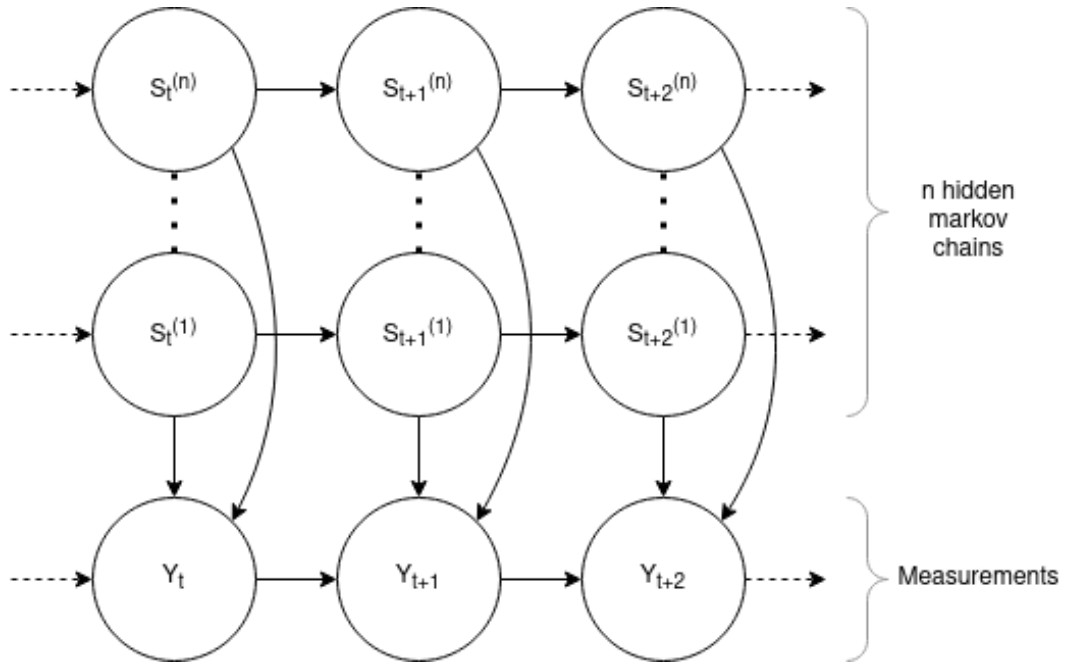
2.1 Algorithms

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 represents a single output power, as such a FHMM cannot represent a continuously variable device.



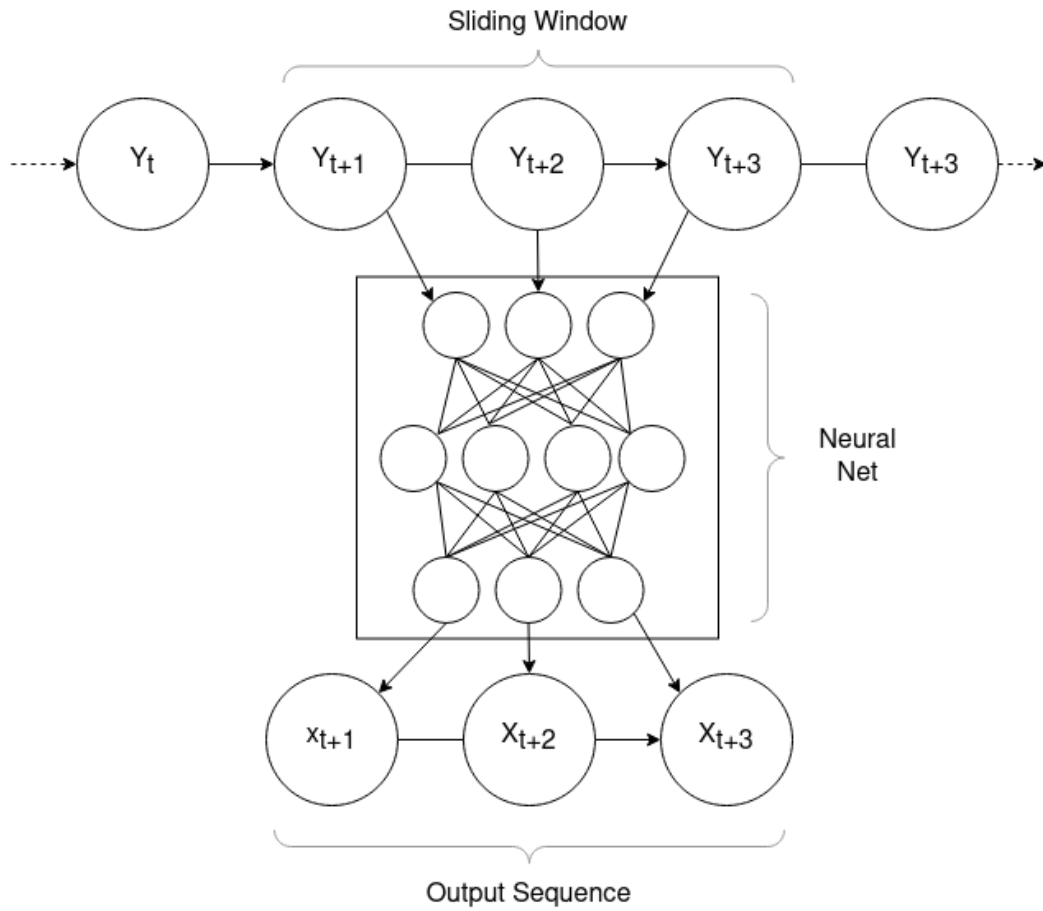
- measurements written wrong

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.

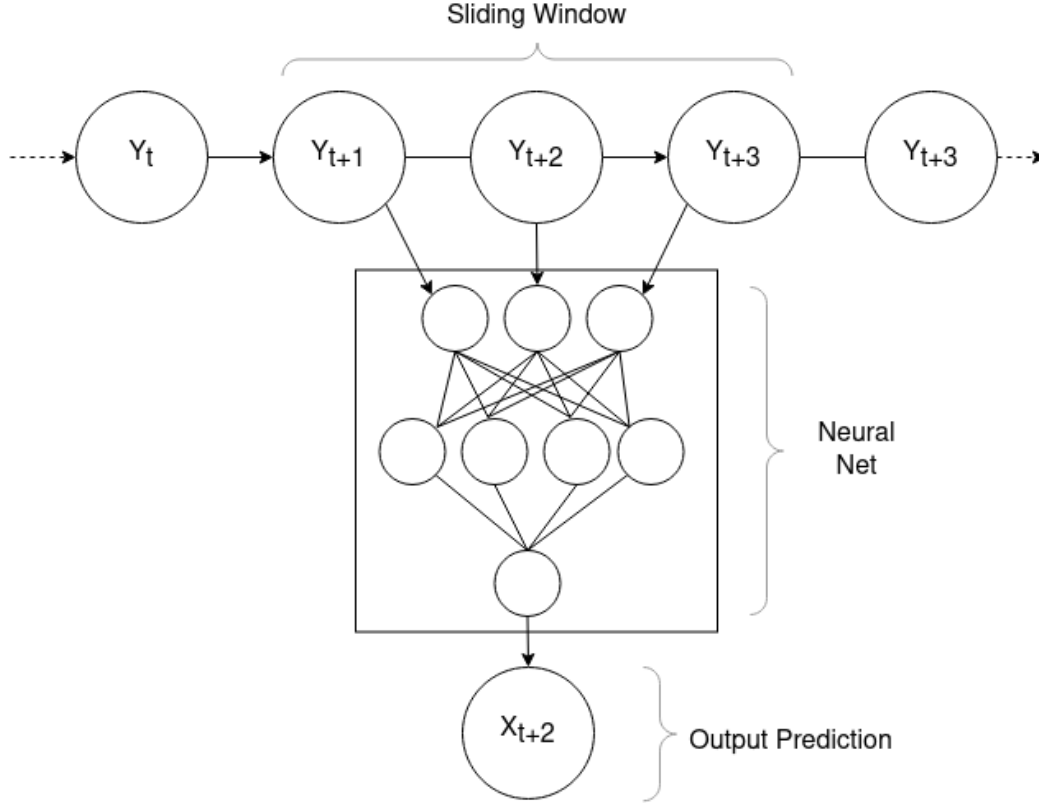
2.1.4 Sequence to Sequence (Seq2Seq)

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.

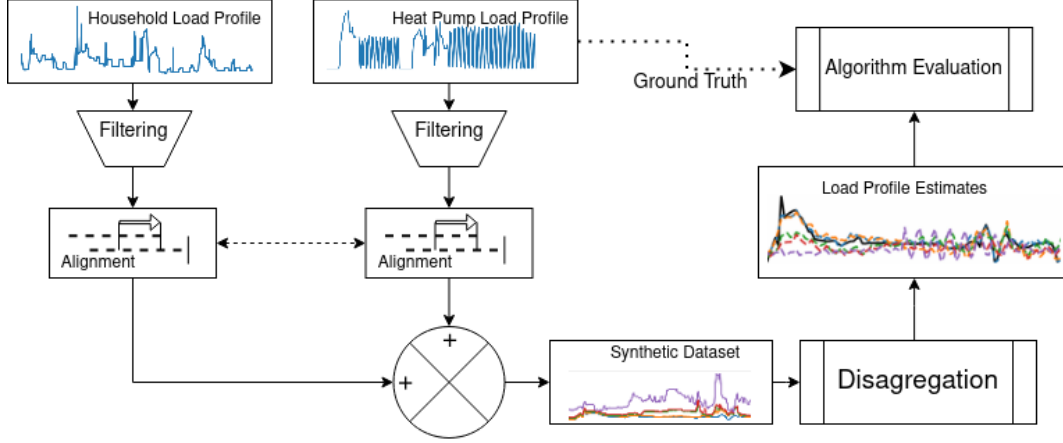
2.2.2 Python

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

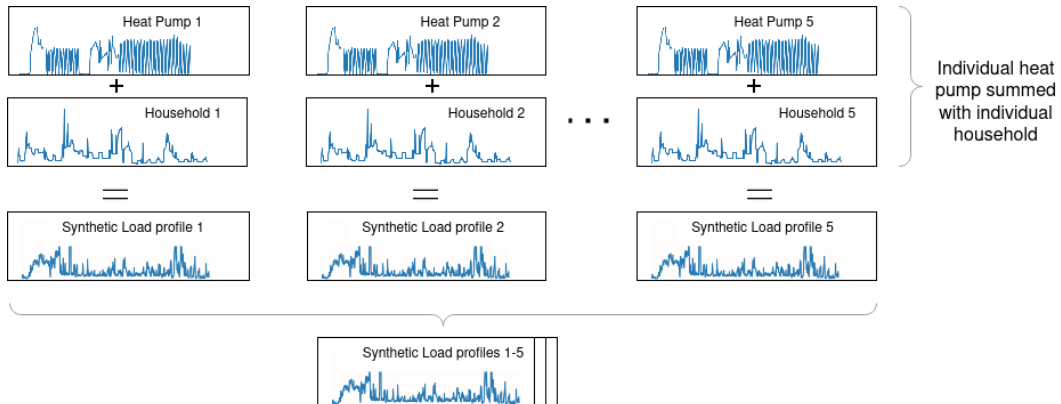
For both the training and benchmarking of these algorithms, test data is needed which includes both the aggregate 'site-meter' measurement as well as a 'sub-meter' measurement just upstream of the heat pump. Through a comprehensive study of the many existing NILM datasets, only few sites with heat-pumps were identified. Data for five individual households with heat pumps has instead been synthesized by combining heat-pump and domestic load-profiles. This data was created following a schema to allow for compatibility with NILMTK which specifies how data/metadata is structured and stored.

$$Y_{t,N} = \sum_{n=1}^{n=N} H_{n,t} + P_t \quad (3.1)$$

For each of the five datasets $Y_{t,N}$ denotes an aggregate measurement of aggregation level N at time t , $H_{n,t}$ denotes the n_{th} household load profile at time step t and P_t denotes an instantaneous measurement of the heat pump load profile.

3.1.1 Mixed Dataset structure

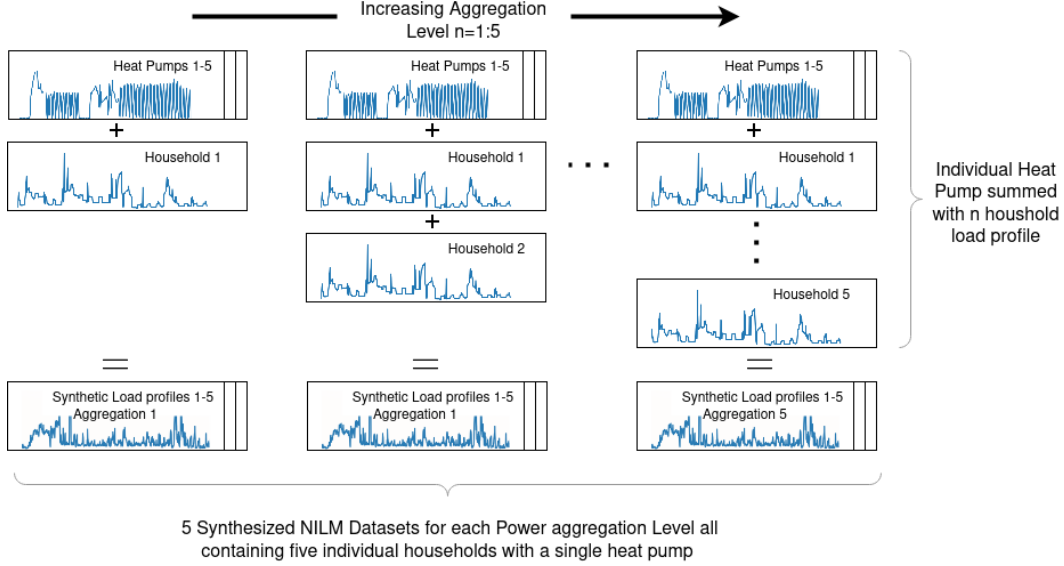
The mixed dataset attempts to represents a random collection of five household load profiles each with a single heat pump. Thus each household within the set represents an aggregation "level" of one.



Single "Mixed" synthetic data-set with 5 households each with a single heat-pump

3.1.2 Increasing power aggregation dataset structure

The five increasing power aggregation datasets respectively represent a feeder with a single household containing a heat pump with N other non heat pump households. Thus each dataset represents one particular power aggregation “level” between 1-5. It should be noted that each dataset itself contains five household.



3.2 Algorithm Training and Evaluation

For consistency all synthesized data at minimum covers the same four month span between January 1st to April 1st. The algorithms are all allocated a three month span of January-March for training, and then bench-marked on the entirety of April. All testing was conducted using the NILMTK’s experimentation API which allows for easy definition of individual experiments and extraction of metrics such as overall Root Mean Squared Error (RMSE).

3.2.1 Increasing Temporal Aggregation

This condition is tested by using the down-sampling feature of the NILMTK experimentation API on the mixed dataset. Sample periods of 1 to 15 minutes are tested in 1 minute intervals.

3.2.2 Increasing Power Aggregation

Increasing power aggregation is tested with the experimentation API with a fixed sample period of 2 mins but with each of the five power aggregation level datasets.

3.2.3 Performance Metric - Mean Normalized RMSE (MNRMSE)

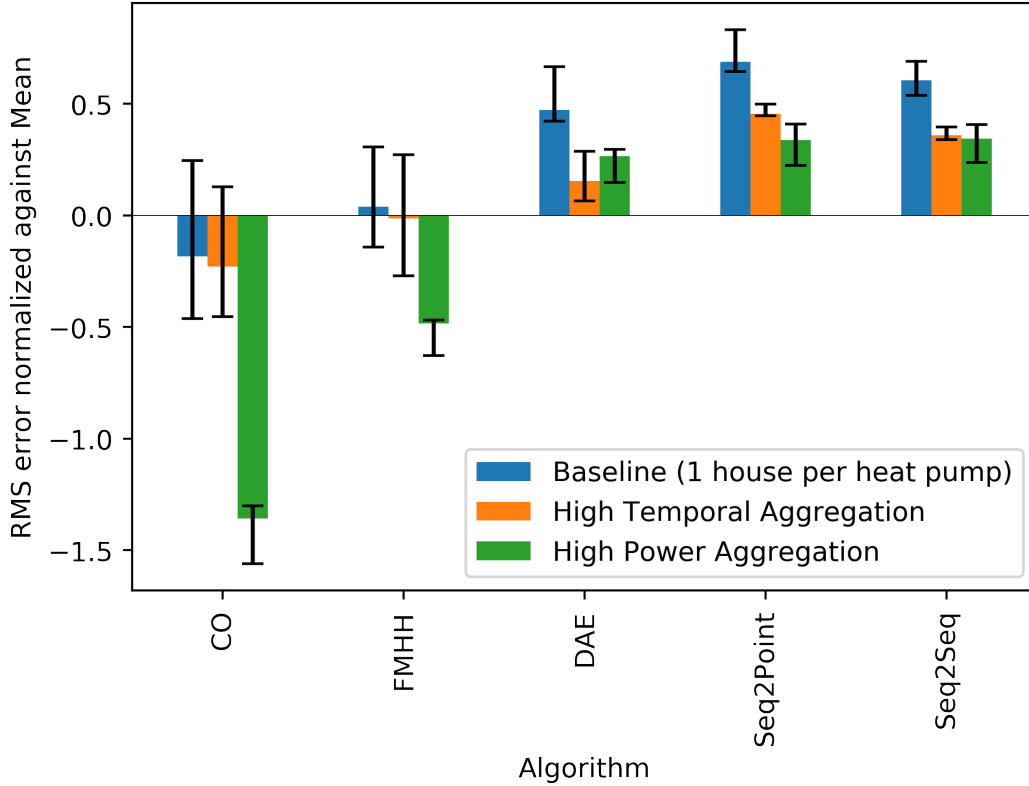
Performance of each algorithm is evaluated by calculating first the RMSE between the prediction vs the heat-pump ground truth for each algorithm. Since RMSE will be a function of the average power of a given heat-pump, this metric is normalized against the RMSE incurred by the Mean algorithm for that same household, allowing for valid comparisons between households. This also has the benefit of providing an intuitive range of values, where 1 is the maximum possible value, 0 is the same performance as Mean, and negative values represent a serious failure.

$$\text{MNRMSE}_{alg} = \frac{\text{RMSE}_{Mean} - \text{RMSE}_{alg}}{\text{RMSE}_{Mean}} \quad (3.2)$$

4 Results

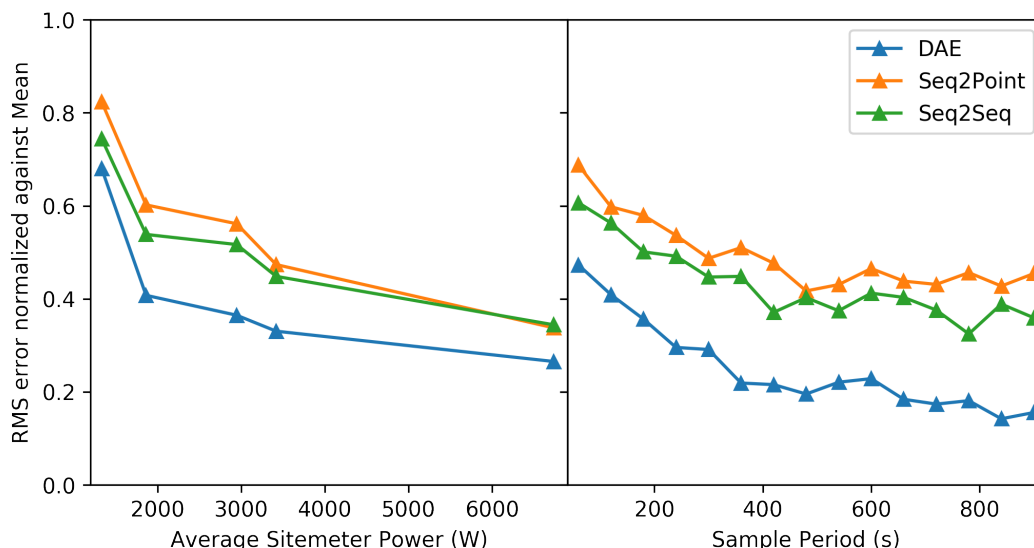
Algorithm performance for both increasing temporal and power aggregation levels is summarized in figure 4 by comparing the two extremes of aggregation. Each bar represents an algorithms average MNRMSE over all households in the set. Positive values thus represent an improvement relative to mean disaggregation, with the maximum possible value being unity, conversely a negative score states than predicting the mean value at all times is less erroneous than this algorithm and represent a serious failure of the algorithm in question. The error bars represent the upper and lower quartiles for each household in the set.

Average normalized performance for all algorithms at the highest temporal and power aggregation levels



The effect of increasing aggregation levels is shown in figs ... by plotting temporal/power aggregation level as the dependent variable against the normalized performance for each algorithm. Note CO and FMHH are omitted due to poor performance. Note that rather than number of household, power aggregation level is plotted as site-meter average as a rough approximation of the fact that each additional household will add a different “amount” of aggregation.

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 fundamentally rely on a finite set of states the poor performance with a continuously variable load is to some degree expected, however in both cases the bulk of the error for these algorithms relates to classification of state. It is also noted that these algorithms seem to be much more sensitive to power aggregation than temporal aggregation however in both cases these algorithm quickly drop to performance worse than Mean implying that they yield no real predictive power.

4.1.2 Neural Network based Algorithms

The neural network based algorithms performed substantial better, seq2point performing on average the best, however being overtaken by seq2seq at the extreme. In both cases, though more extreme for power, the algorithms display a relatively rapid drop in performance before setting. The fact that the temporal algorithm settles, contradicts the intuition that, with less information predictive power should get steadily worse and finally settle at the same performance as Mean. One possibility is that the neural net is relying on a sparse heuristic similar yet better than mean disaggregation such as perhaps predicting the same load profile every day, though this is contradicted by the fact that all algorithms seem to be settling at different levels.

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

5 Links

NILMTK <https://github.com/nilmtnk/nilmtnk>

Fork Of NILMTK <https://github.com/BenjaminFrazer/nilmtnk>

Code and Data https://github.com/BenjaminFrazer/thesis_tools

Emacs 28.1 (Org mode 9.6)