University of Strathclyde Engineering

INDRAPRASTHA INSTITUTE *of* INFORMATION TECHNOLOGY **DELHI**

Evaluation of Non-intrusive Load Monitoring Algorithms for Appliance-level Anomaly Detection

Haroon Rashid*, Vladimir Stankovic, Lina Stankovic, Pushpendra Singh*

*Indraprastha Institute of Information Technology, Delhi, India
University of Strathclyde, Glasgow, UK

EPSRC Engineering and Physical Sciences Research Council

MARIE CURIE ACTIONS

IEEE ICASSP-2019, Brighton, UK, May 2019

https://sensible.eee.strath.ac.uk

Motivation

- Non-intrusive load monitoring (NILM): Effective method to detect appliance usage and provide energy consumption estimates

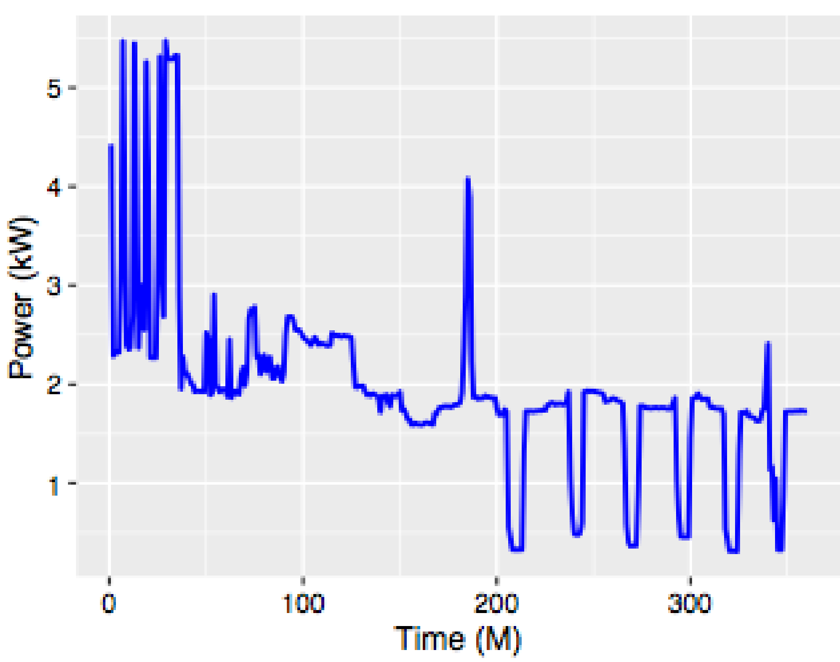
- NILM's suitability has been demonstrated for numerous practical applications inc. appliance modelling, energy feedback, demand-side management, smart home, etc. => thus removing the need of expensive and intrusive submetering

- But can current NILM approaches be leveraged for detecting anomalous operation of an appliance?

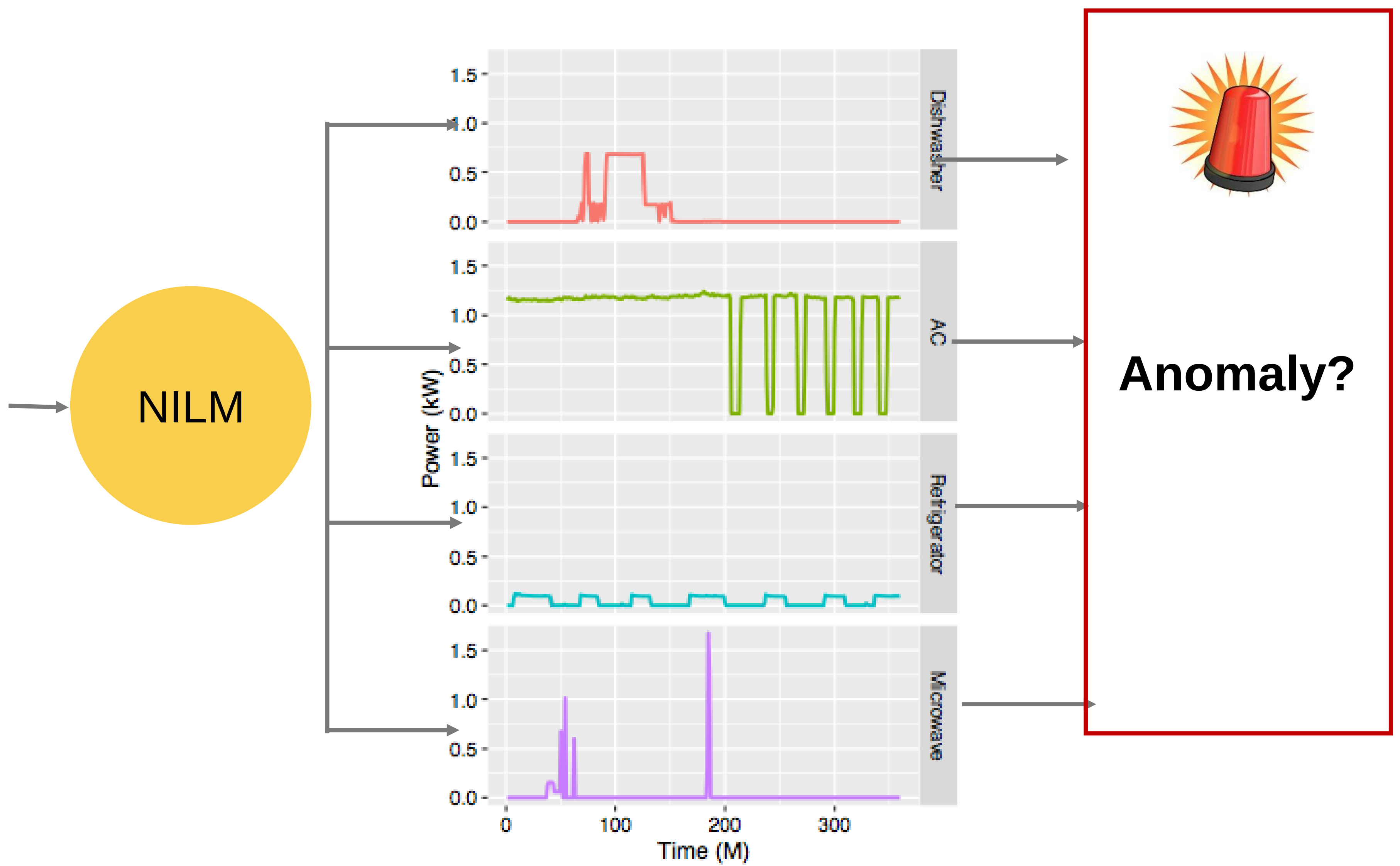
- *Anomalous* implies no 'known' or 'labelled' anomaly signatures to train on or model

- Anomaly detection problem => are anomaly detection approaches designed for submetering appliance signatures transferable to NILM signatures?

What would we like to achieve?



Aggregate smart meter data
at household level

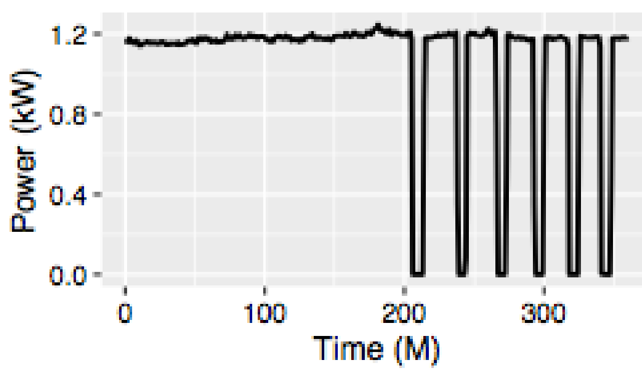


Disaggregated appliances’ data

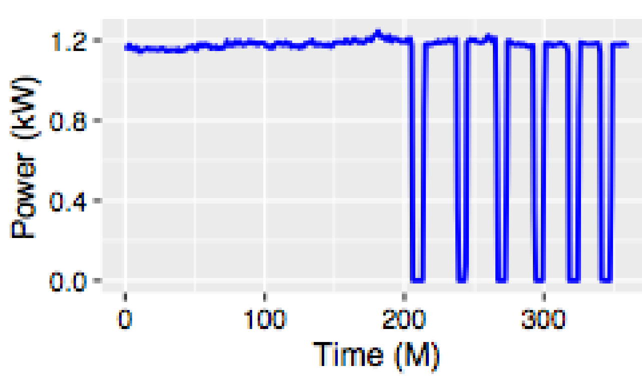
NILM's suitability for anomaly detection

Research Question: Can **current NILM approaches** be leveraged for detecting anomalous operation of an appliance with **current anomaly detection approaches**?

NILM



*



* source: https://en.wikipedia.org/wiki/Wattmeter

ANOMALY DETECTION

Accuracy 1

Accuracy 2

Is Accuracy 1 ≈ Accuracy 2 ?

NILM can be used for anomaly detection

State-of-the-art w.r.t appliance anomaly detection

- Anomaly detection in aggregate meter readings well studied topic
- Appliance-level anomaly detection approaches currently designed for submetering data, based on appliance-specific energy modelling
- With improvement in NILM performance, authors were first to evaluate in depth suitability of NILM outputs for anomaly detection using current NILM approaches and traditional rule-based anomaly detection designed around submetered data and artificially injected anomalies[1]
  - Findings[2]: Appliance-level signature generated by NILM is not of sufficient fidelity to accurately detect anomalous behaviour because NILM tends to generate signatures that it has "learnt", and thus while NILM can detect appliance events accurately it cannot separate anomalous from non-anomalous signatures and consequently reproduces the "learnt" signature

[1] Residential electrical loads measurements with simulated anomalies in air conditioner and refrigerator dataset, 2019, DOI: 10.15129/d712ccac-21a1-40d2-8456-41217b62a6d5
[2] H. Rashid, P. Singh, V. Stankovic and L. Stankovic, "Can non-intrusive load monitoring be used for identifying an appliance's anomalous behaviour?" *Applied Energy*, 2019

Contributions

- An anomaly detection approach that works for both NILM and submetered data
  - Testing on submetered data reports its baseline performance whereas testing on NILM data shows the usability of NILM for identifying anomalous appliances.
- A post-processing algorithm for improving anomaly detection capability of traditional NILM.
- Evaluation on actual (not synthetic) anomalies within the REFIT data
- Release of first publicly annotated real appliance-anomaly dataset[3] from the REFIT dataset .

[3]Annotated load anomalies from the REFIT dataset, 2019: DOI: 10.15129/9729a2a0-11ce-4cce-b0d0-144c483fcb33

Methodology

- Detect anomalies within the 2-year REFIT dataset focusing on appliances with 2 cycles, e.g., fridge-freezer, electrical heater
- Characterise anomalies of these cyclical appliances
- Develop rule-based anomaly detection algorithm for cyclical appliances
- Design post-processing algorithm to solve NILM’s inability to reproduce anomalous signature
- Evaluate proposed anomaly detection algorithm on multiple NILM algorithms outputs with and without post-processing

Abnormal behavior of an appliance -> anomaly

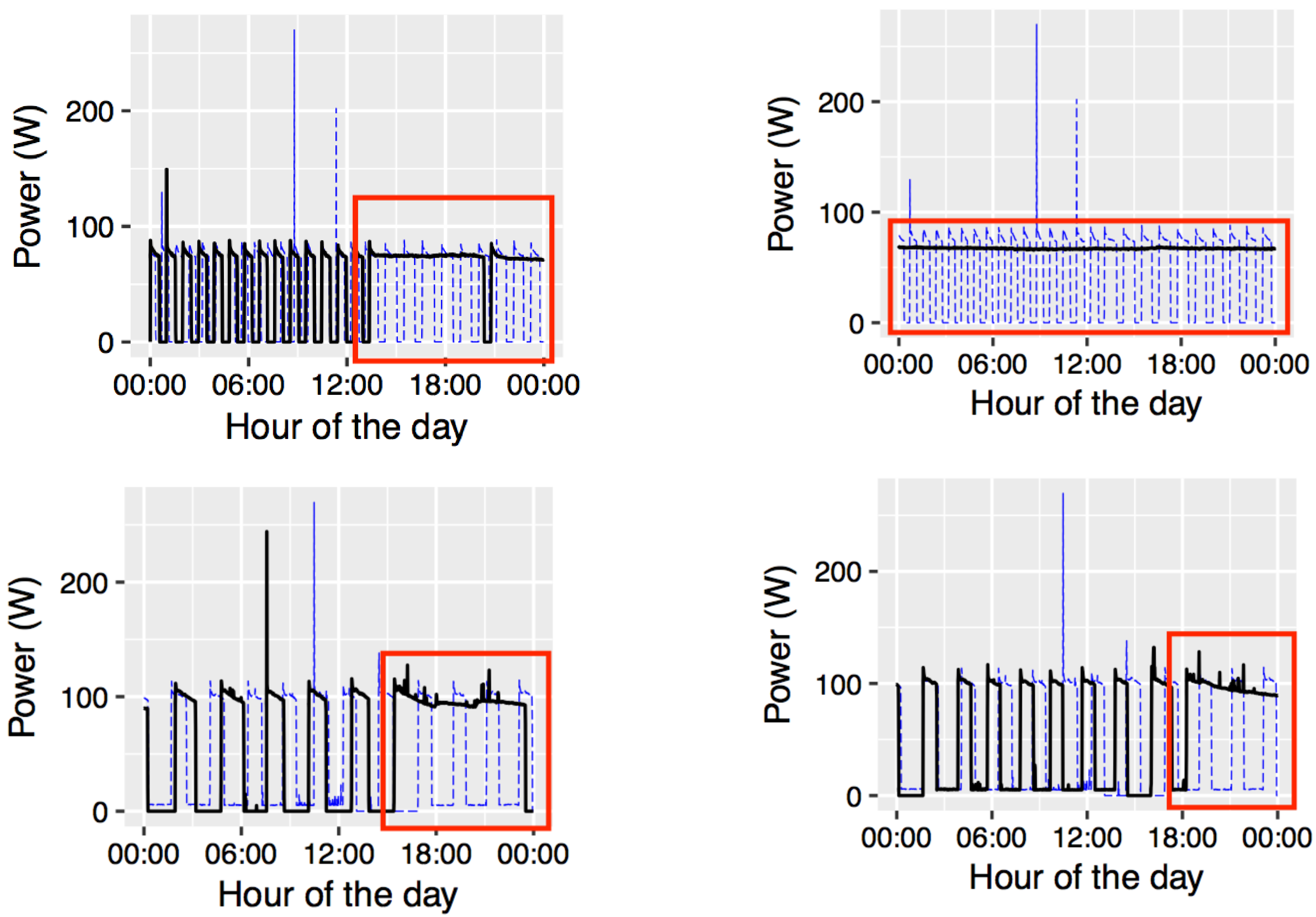


Fig. Thin dashed lines show appliances normal consumption pattern and thick solid lines show consumption pattern on an anomalous day

Compressor-based home appliances: types of anomaly

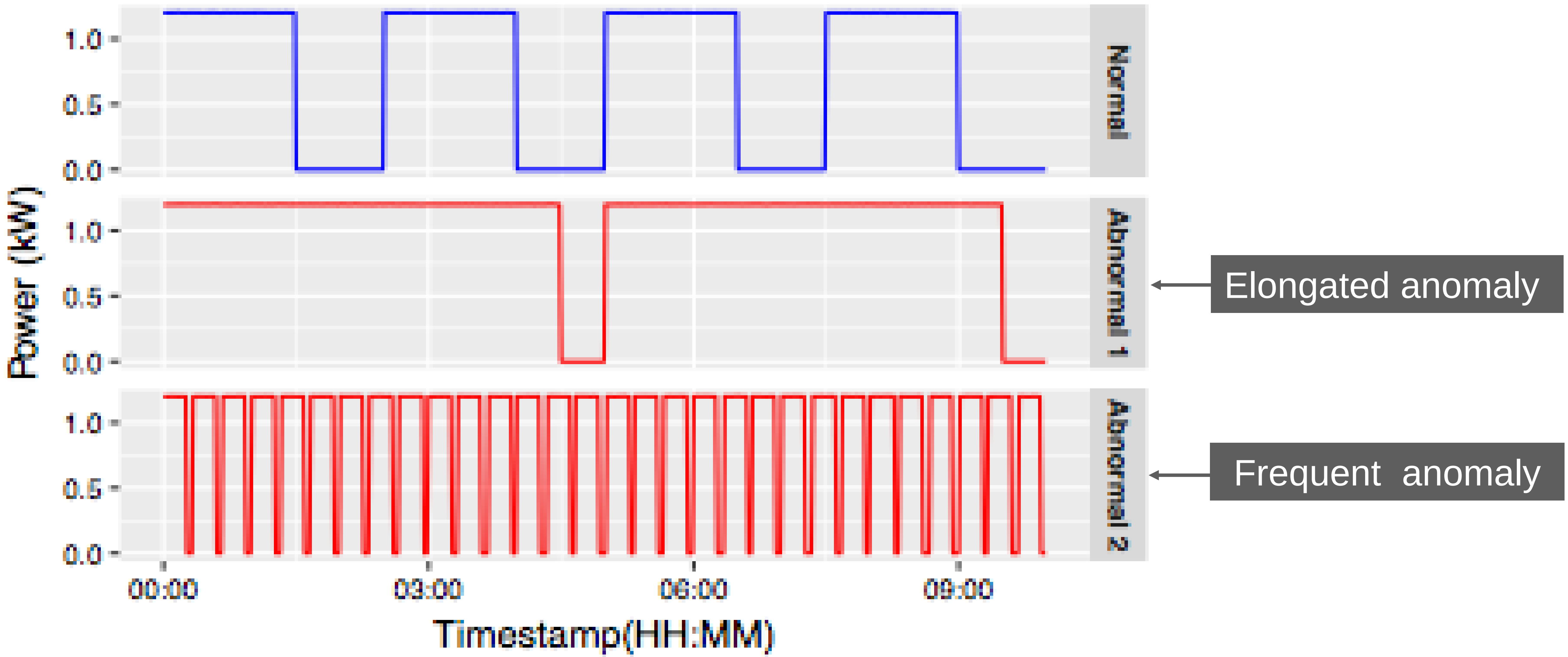


Fig. Power consumption signature of air conditioner (AC) in three different modes

Rule-based Anomaly Detection Algorithm

- Training Phase:
  - Collect power readings of an appliance for D normal days
  - For each day, count the number of cycles (*c*) in the appliance signature and calculate the energy (*e*) consumption of each cycle as

$$C_{train} = mean(c_i), i \in \{1, \ldots, D\} \tag{1}$$

$$\sigma^{C}_{train} = std(c_i), i \in \{1, \ldots, D\}, \tag{2}$$

$$E_{train} = mean(\mathbf{e}_i), i \in \{1, \ldots, D\}, \tag{3}$$

$$\sigma^{E}_{train} = std(\mathbf{e}_i), i \in \{1, \ldots, D\}, \tag{4}$$

Rule-based Anomaly Detection Algorithm

- Testing Phase:
  - For each test day, compute again the number of cycles and the energy consumption of each cycle.
  - Flag anomaly, if:
    - the average energy consumption of test day cycles is significantly greater than train day cycles -> *Elongated anomaly* $E_{test} > \alpha * (E_{train} + n * \sigma^{E}_{train})$
    - the number of cycles taken by an appliance is significantly greater than the train day cycles -> *Frequent anomaly*
      $$C_{test} > C_{train} + n * \sigma^{C}_{train}$$

Dataset

- Publicly available REFIT dataset
  - Two year dataset of 20 UK homes
  - Both aggregate and appliance-level data
  - Downsampled from eight sec. to one minute
  - Selected five homes (1, 10, 16, 18, 20) having anomalies in heater and freezer usage
  - Used four months of data from each of these homes

Ground truth Establishment

- Appliance’s consumption found significantly different than historical consumption
  - Flagged as anomalous and marked as S (sure)
  - Noted time-duration of the anomaly
- Appliance’s consumption found significantly different from its historical consumption, but anomalous duration seems due to measuring meter malfunctioning (meter stuck case)
  - Flagged as anomalous and marked as NS (not sure)
  - Noted time-duration of the anomaly
- Appliance’s ON cycle duration found significantly longer than its historical consumption and the predecessor OFF cycle also found longer
  - Not marked as anomaly as it is normal

Annotated Anomaly File

House_No	Appliance	Time_Duration	Status	Explanation	Comments
1	Freezer_1	"2014-06-18 23:00:00 ; 2014-06-19 04:00:00"	S	Continuous ON state	
1	Freezer_1	"2014-08-01 18:00:00 ; 2014-08-02 13:00:00"	NS	Sensor seems stuck	
1	Freezer_2	"2014-08-01 18:00:00 ; 2014-08-02 13:00:00"	NS	Sensor seems stuck	
1	Freezer_2	"2014-09-16 15:30:00 ; 2014-09-16 19:20:00"	S	Continuous ON state	

Fig. Screenshot of the annotated file
S= Sure - Certain anomaly
NS= Not Sure - Possible anomaly

First such annotated anomaly dataset:

https://doi.org/10.15129/9729a2a0-11ce-4cce-b0d0-144c483fcb33

University of Strathclyde Engineering

NILM Algorithms

- Super State Hidden Markov Model (SSHMM) [TSG ‘16]
- Unsupervised Graph Signal Processing (GSP) [IEEE Access ’16]
- Latent Bayesian Melding (LBM) [NIPS ’15]
- Factorial Hidden Markov Model (FHMM) [AISTATS '12]
- Combinatorial Optimization (CO) [Proceedings of the IEEE ‘92]

Publicly available implementations were used to get disaggregation results.

Experimental settings

- Focus on houses within the dataset whose appliance signatures displayed anomalous behaviour
- For supervised NILM algorithms, one month of data was used for training and remaining three months for testing.
- Algorithms were tested in a sliding-window manner with a window size of one day.
- 'Noisiness' due to unknown appliances established to fairly assess performance

House #	1	10	16	18	20
Noise[4] %	68	58	63	45	55

$$Noise\ percentage = \frac{\sum_{t=1}^{T} |Y_t - \sum_{i=1}^{N} y_t^i|}{\sum_{t=1}^{T} Y_t} * 100,$$

[4] S. Makonin, F. Popowich, I. V. Bajić, B. Gill, and L. Bartram, "Exploiting HMM Sparsity to Perform Online Real-Time Nonintrusive Load Monitoring," IEEE Transactions on Smart Grid, vol. 7, pp. 2575-2585, Nov 2016.

Performance Metrics

- Root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{T}}$$

- Pearson correlation coefficient:

$$\rho_{s,p} = \frac{cov(s,p)}{\sigma_s \sigma_p}$$

- F1 score:

$$F1\ score = 2 * \frac{precision * recall}{precision + recall}$$

Disaggregation Performance

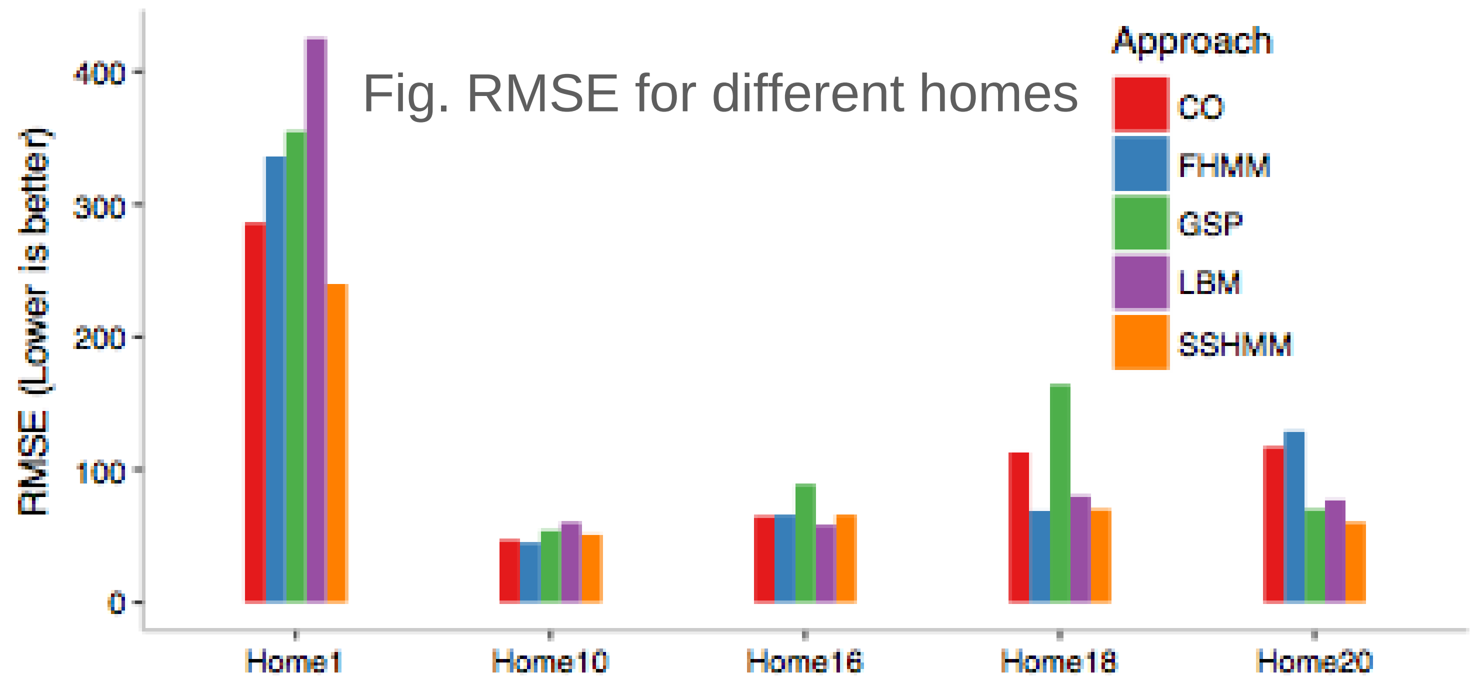


Fig. RMSE for different homes

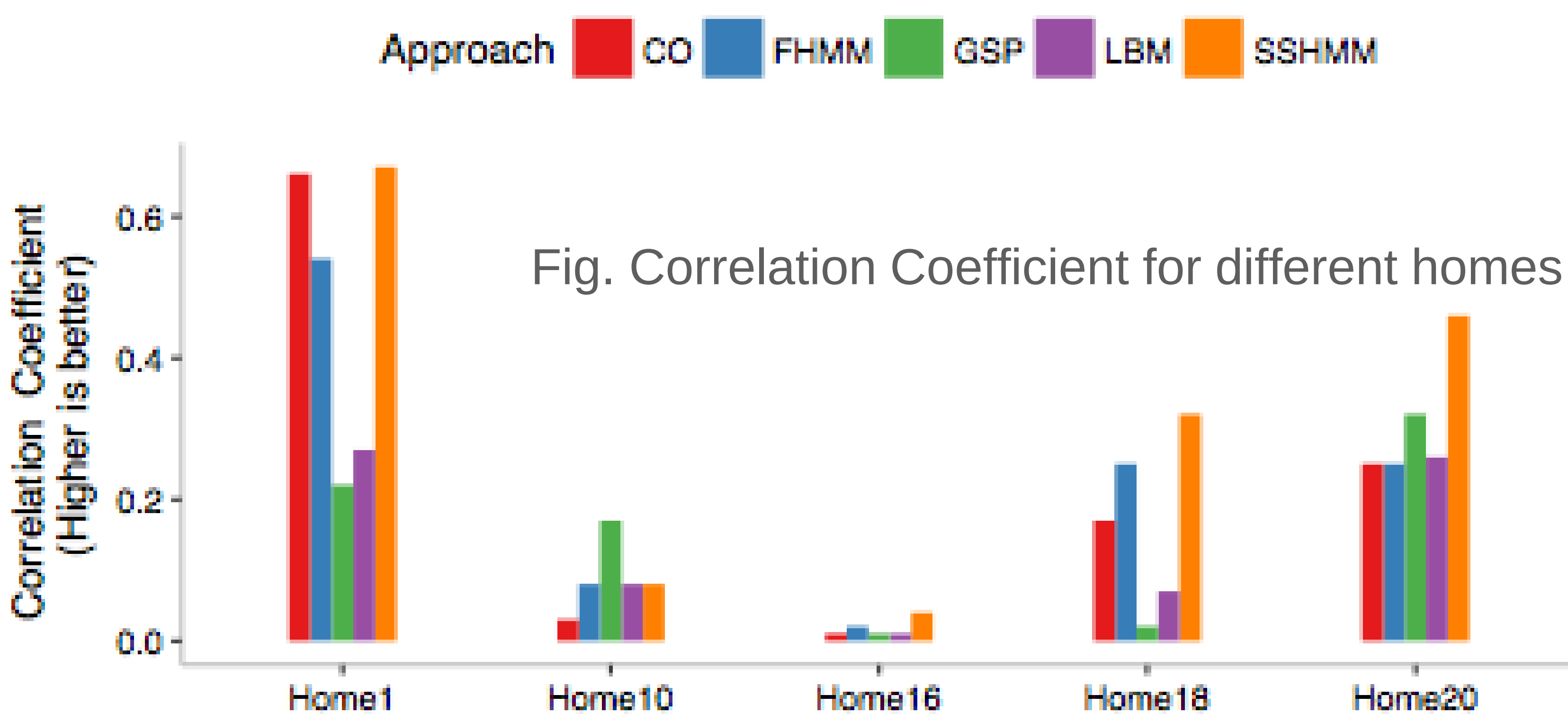
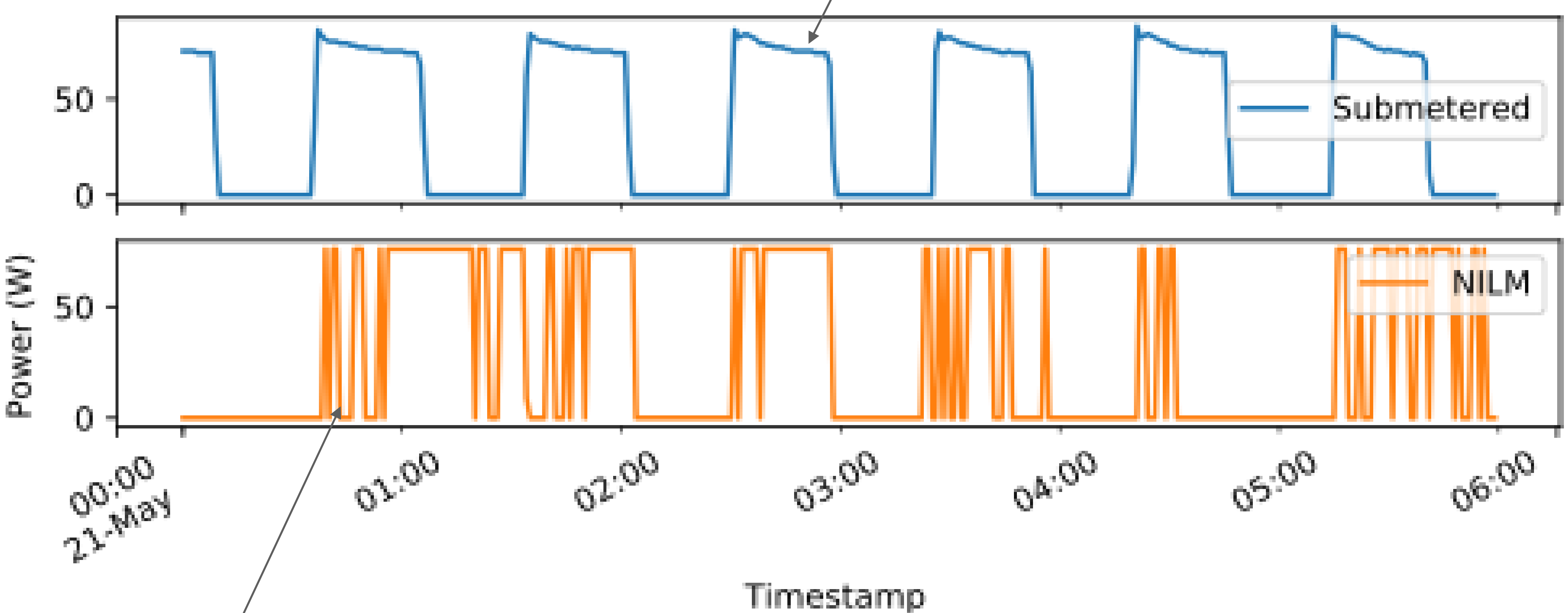


Fig. Correlation Coefficient for different homes

Signature reconstruction: SSHMM Example

Actual power consumption of the appliance

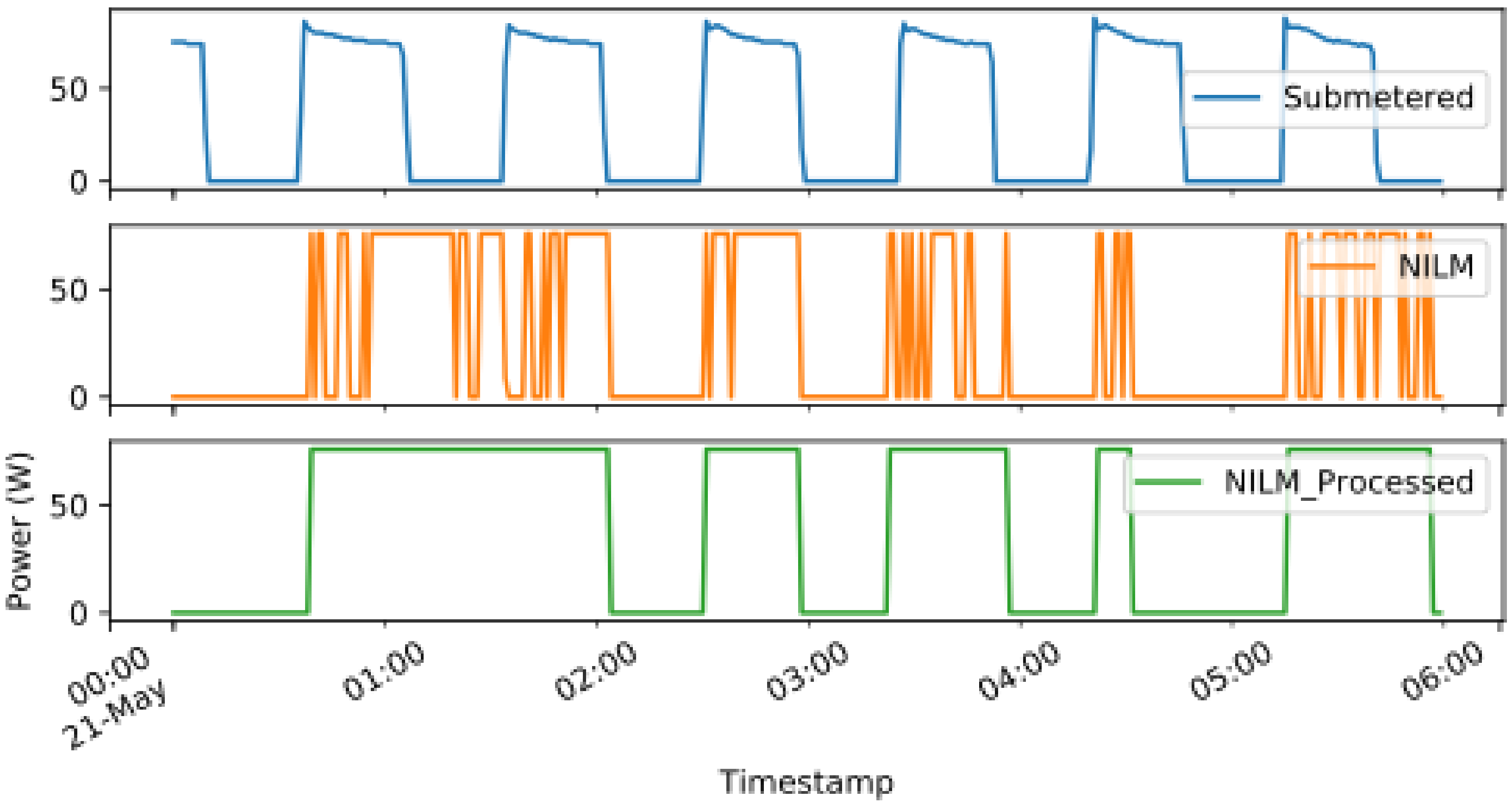


NILM inferred power consumption of the appliance

ON cycles are not predicted accurately

Post-processed NILM

Combine those cycles which had off duration smaller than the average off duration of the cycles found in the actual power consumption trace of the appliance



Post-processed NILM data is better than the unprocessed NILM data

Performance

F-score results without post processing NILM

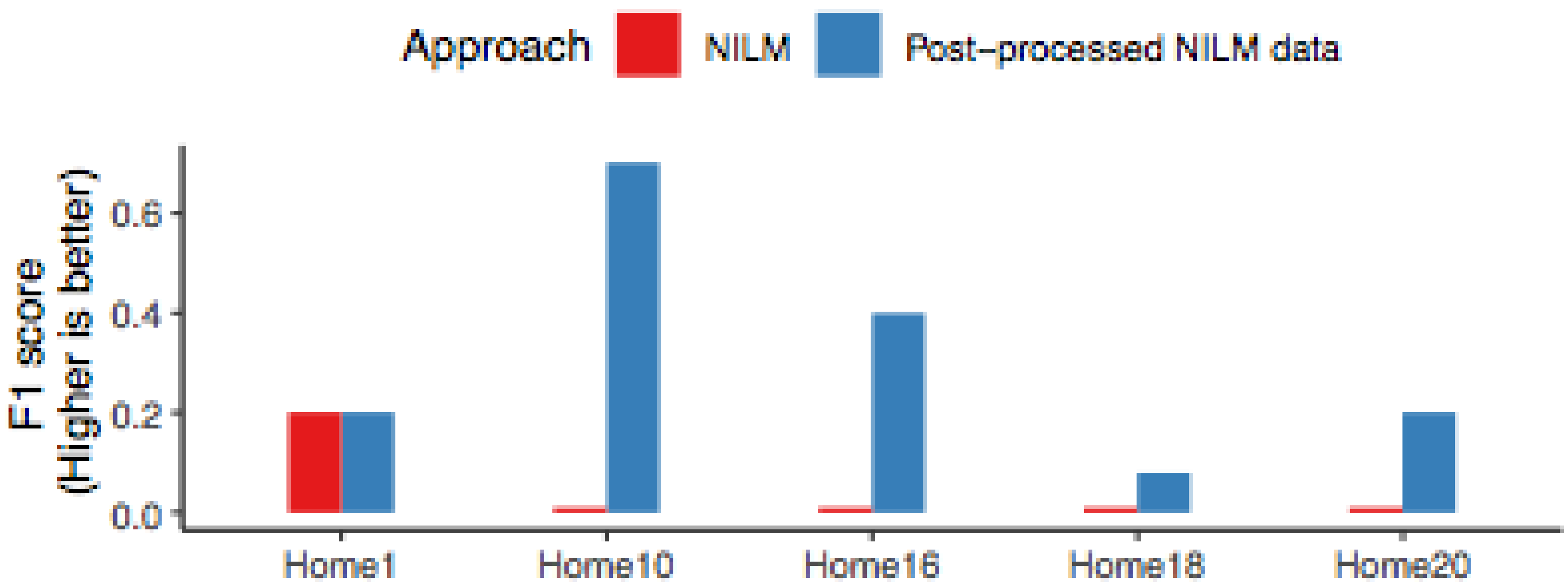
	Home 1						Home 10						Home 16						Home 18						Home 20					
	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP
Precision	0.7	0.1	0.1	0.08	0.1	0.14	1	0	0.6	0	0	0	0.9	0	0	0.23	0	0.24	1	0	0	0.06	0	0.04	1	0	0.33	0.13	0	0.33
Recall	1.0	1.0	1.0	1.00	1.0	1	1	0	0.27	0	0	0	0.9	0	0	0.43	0	0.48	0.7	0	0	0.33	0	1	0.6	0	0.3	0.4	0	0.1
Fscore	0.8	0.2	0.2	0.15	0.2	0.25	1	0	0.37	0	0	0	0.9	0	0	0.3	0	0.32	0.8	0	0	0.1	0	0.08	0.7	0	0.31	0.2	0	0.15

F-score results with post processing NILM

	Home 1						Home 10						Home 16						Home 18						Home 20					
	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP
Precis.	0.7	0.1	0.1	0.01	0.1	0.14	1	0	0.5	0	0.5	0.5	0.9	0	0.25	0.2	0.3	0.2	1	0	0.04	0.07	0.04	0.04	1	0.5	0.3	0.14	0.11	0.13
Recall	1.0	1.0	1.0	1.00	1.0	1	1	0	0.9	0	1	0.6	0.9	0	1	0.5	1	0.6	0.7	0	1	0.67	1	1	0.6	0.1	0.5	0.5	1	0.6
Fscore	0.8	0.2	0.2	0.15	0.2	0.25	1	0	0.7	0	0.7	0.5	0.9	0	0.4	0.3	0.4	0.3	0.8	0	0.08	0.13	0.08	0.08	0.7	0.17	0.3	0.22	0.2	0.21

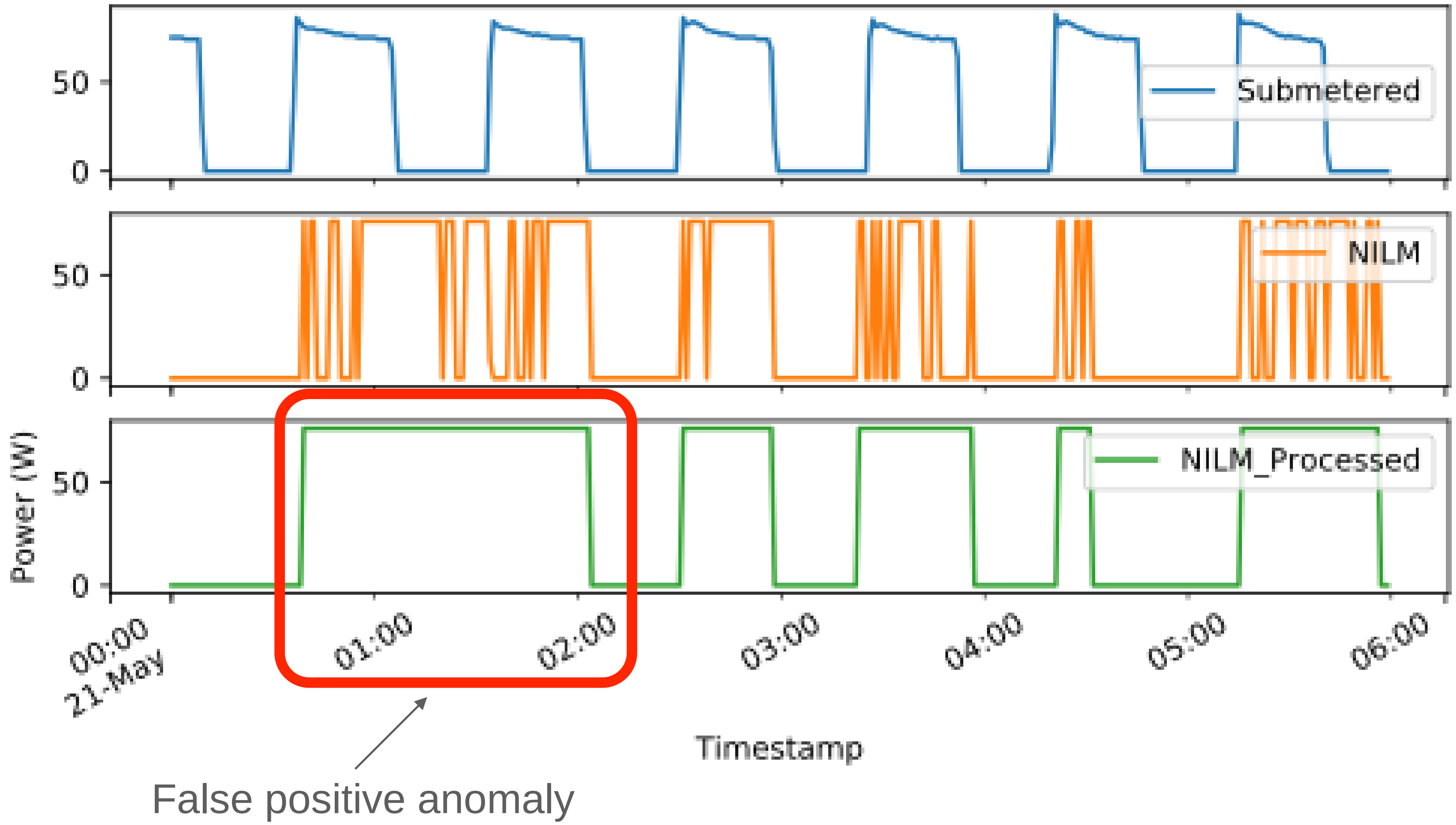
Improvement of F1 score with post-processing, enabling AEM to flag up all anomalies

F1 score improvement



F1 score on post-processed NILM data is still low

False positives



Post-processing results in false positive anomalies

Conclusion & Future work

- Appliance-level anomalies cannot be detected by using state-of-the-art off-the-shelf NILM directly nor using rule-based anomaly detection algorithms designed for submetered appliance signatures

- Research challenges:

  - Propose new anomaly-aware NILM techniques

  - Develop novel anomaly-detection rules suitable for NILM-based detection