

EMBED: A Dataset for Energy Monitoring through Building Electricity Disaggregation

<http://embed-dataset.org/>

POC: Farrokh Jazizadeh, Ph.D.

jazizade@vt.edu

EMBED is a publicly available *fully-labeled dataset* that contains data from three apartment units in Los Angeles, California, collected for at least two weeks (up to four weeks). Aggregate power files, fully labeled appliance event timestamps, and plug load consumption for a variety of appliances have been provided.

The characteristics of the dataset are as follows:

- High resolution (60Hz) power data includes real and reactive power up to 9th harmonics
- Three apartment units with 14, 21, and 27 days of data collection duration
- **Fully labeled events:**
 - In addition to labels for each appliance, each state transition has a sub-label
 - E.g. a refrigerator could include compressor on/off, defrost on/off, and light on/off labels
 - The high-fidelity labels could facilitate the **appliance modeling**
- **Ground truth power data** at appliance level has been provided:
 - For appliances that could be connected to a receptacle
 - Plug load sampling rate 1-2Hz
- Raw current and voltage data @12KHz is available for sharing

Summary of dataset characteristics

Feature	Value
Current sampling rate	12 KHz
Voltage sampling rate	12 KHz
Power resolution	60 Hz
Plug load sampling rate	1-2 Hz
Power metrics	Real (P) Reactive (Q)
Calculated power harmonic content	Up to 9 th harmonics
1 Duration, # of occupants	14 days, 1 occupant
2 Duration, # of occupants	21 days, 1-2 occupants
3 Duration, # of occupants	27 days, 1-2 occupants

Reference:

Jazizadeh, F., Afzalan, M., Becerik-Gerber, B., and Soibelman, L. 2018. EMBED: A Dataset for Energy Monitoring through Building Electricity Disaggregation. In Proceedings of the Ninth International Conference on Future Energy Systems (e-Energy '18). ACM, New York, NY, USA, 230-235. DOI: <https://doi.org/10.1145/3208903.3208939>