# SustDataED: A Public Dataset for Electric Energy Disaggregation Research

Miguel Ribeiro, Lucas Pereira, Filipe Quintal, Nuno J. Nunes

Madeira Interactive Technologies Institute
Laboratory for Robotics and Engineering Systems
Funchal, Portugal
{jose.ribeiro, lucas.pereira, filipe.quintal, nuno.nunes}@m-iti.org

## I. INTRODUCTION

In this paper we present the SustDataED (SustData for Energy Disaggregation) dataset. The dataset is an extension to the original SustData dataset [1], and consists of electric energy consumption and room occupancy measurements taken from a single-family residence in Portugal composed of four householders. Current and voltage waveforms were sampled from the mains at 12.8 kHz for the duration of 10 days. The dataset also contains the individual consumption for 17 individual loads, measured at 0.5 Hz complemented with individual labels for the state transitions of those loads. Lastly, the dataset was enriched with household occupancy information at the room level that was collected using the Estimote Beacon platform. With this dataset we aim to motivate the development and evaluation of novel energy disaggregation algorithms, not only at the individual appliance level but also of household activities (e.g., doing the laundry or preparing breakfast).

## II. DATA COLLECTION SETUP

### A. Whole-house Energy Consumption

In order to measure the whole-house consumption data, we use the LabJack U6, a data acquisition board (DAQ)[1], with 14 analog input ports and 16-bit resolution for a maximum sampling rate of 50 kHz. In the current version of SustDataED we are sampling current and voltage waveforms at 12.8 kHz and storing the resulting measurements in the bespoke SURF file format [2]. The resulting files are later post-processed using the built-in programming interface (SURF-PI) in order to generate power measurements that are relevant for energy disaggregation (e.g., current RMS, voltage RMS, real power and reactive power).

### B. Individual Appliance Consumption

For the appliance-level data collection, we use the Plugwise[2] system, which is a commercially available, distributed sub-metering platform. To this end we extended an open-source Python package[3] that allows direct access to the raw measurements of each individual appliance being measured. More precisely, we changed the original scripts such that it was possible to capture plug-level measurements every two seconds instead of the default 10 seconds. The whole system is controlled by a netbook running Ubuntu 14, to which we added a web application to facilitate the configuration of the plugs during the deployment and maintenance phases and to provide the householders with a user interface where they can check the individual consumption of the monitored appliances.

### C. Room Occupancy

Room occupancy data was collected using the Estimote[4] beacons platform. This allowed us to associate electricity consumption with the house occupants. In the current version of the datasets we used 6 beacons, each one in a different room of the house, namely: three bedrooms, living room, kitchen and laundry room.

To collect the information from the beacons, we developed an Android *smart-watch* application to analyze those beacons with the strongest signal every 5 seconds. A record was stored every time there was a change in room, and there were at least two consecutive readings on the new room, thus eliminating erroneous readings from signal interference.

Upon deploying the beacons, the signal strength was tested in each room with an android test application in order to guarantee the best placement of each beacon, i.e., the placement that does a better job in identifying the rooms the residents are currently in.

## III. DATASET CONTENTS

SustDataED is a publicly available dataset and can be downloaded from: *http://aveiro.m-iti.org/data/sustdata*. The following files are available:

- SURF File Format
  - Current and Voltage at 12.8 kHz and 50 Hz
  - Real and Reactive Power at 50 Hz
- SQLite Database:
  - Plugwise measurements at 0.5 Hz
  - Plugwise power events
  - Room leaps

Table 1 bellow summarizes the current version of the dataset and Fig. 1 shows the relationship between the three data collections that compose SustDataED.

---

[1] LabJack, https://labjack.com/u6
[2] Plugwise, https://plugwise.com
[3] Plugwise Python, https://github.com/SevenW/Plugwise-2-py
[4] Estimote, *www.estimote.com*

## IV. FUTURE WORK

Datasets like this one are always work in progress as the authors strive to continuously add new data. In our particular case we are currently collecting data from a second household. Furthermore, we are currently attempting to integrate specific sensors for different appliances (e.g., ceiling lights and fast switching appliances like washers and driers).

Lastly, we are also looking at ways to improve the human activity detection, either by means of additional sensing technology or human-computer interaction techniques like the daily reconstruction methods through diary studies.

## REFERENCES

[1] L. Pereira, F. Quintal, R. Gonçalves, and N. J. Nunes, "SustData: A Public Dataset for ICT4S Electric Energy Research," in *ICT4S 2014*, Sweeden, 2014.

[2] L. Pereira, N. Nunes, and M. Bergés, "SURF and SURF-PI: A File Format and API for Non-Intrusive Load Monitoring Datasets," *e-Energy 2014*, Cambridge, UK, 2014.

TABLE I.  ROOM OCCUPANCY AND INDIVIDUAL APPLIANCE CONSUMPTION MEASUREMENTS

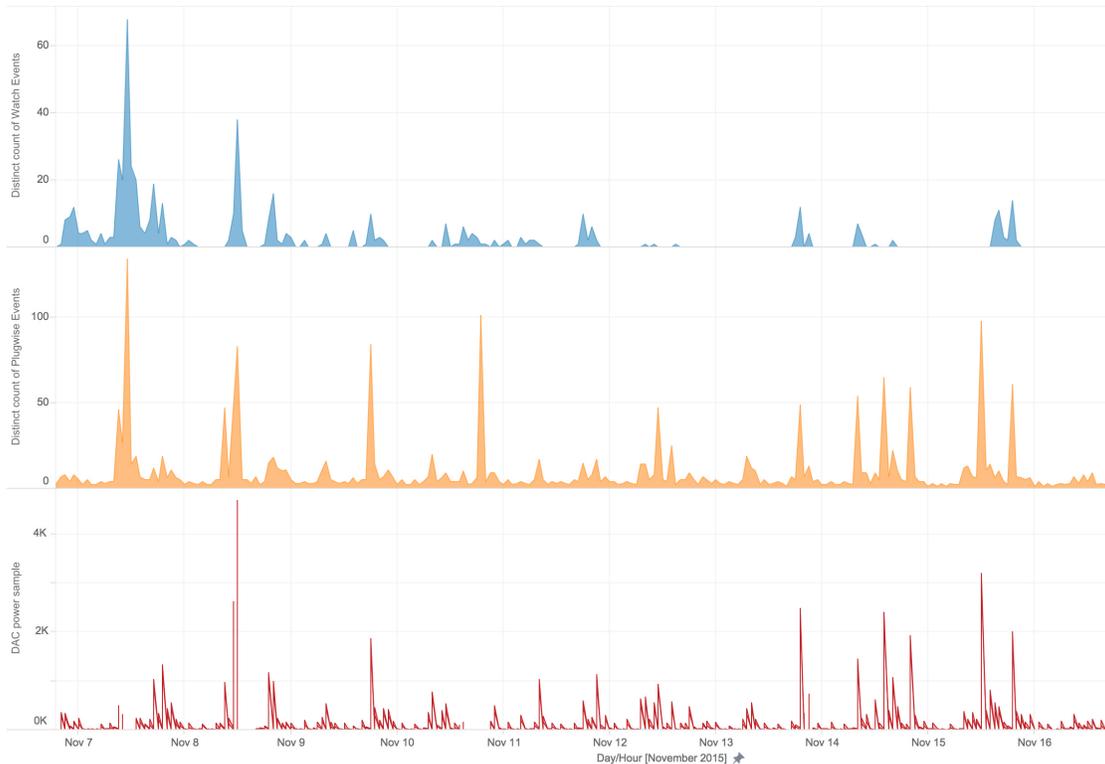| Room | Beacon Inbound Leaps | Appliance | | |
|---|---|---|---|---|
| | | *Name* | *Avg. Power (Watts)* | *Number of Events* |
| Bedroom 1 | 1329 | TV 1 | 47 | 60 |
| Bedroom 2 | 691 | TV 2 | 33 | 22 |
| | | Laptop 1 | 20 | 8 |
| Bedroom 3 | 1006 | TV 3 | 15 | 6 |
| Living room | 1173 | TV 4 | 184 | 44 |
| | | PlayStation | 47 | 20 |
| | | Laptop 2 | 59 | 22 |
| Kitchen | 472 | Stove | 725 | 821 |
| | | Washing Machine | 784 | 20 |
| | | Oven | 77 | 525 |
| | | Refrigerator | 484 | 99 |
| | | Kettle | 993 | 66 |
| | | Microwave | 658 | 102 |
| | | Coffee Machine | 523 | 61 |
| | | Freezer | 84 | 1 |
| | | Dishwasher | 233 | 50 |
| | | Water Heater | 1223 | 338 |
| --- | *4199* | --- | --- | *2265* |



Fig. 1.  Relationship between room occupancy (top), individual appliance power events (center) and whole-house consumption (bottom) aggregated by hour and day.