Data in the Wild

Scott May, Luke Fishback, Daniel Arneman, John Cunningham, Mandar Jog
{smay,lfishback,darneman,jcunningham,mjog}@plotwatt.com
PlotWatt
Durham, North Carolina

ABSTRACT

PlotWatt provides electricity disaggregation to residential users, bringing appliance-level insight into energy consumption for a large number of users. We have encountered diverse challenges working with user data "in the wild" that make Non-Instrusive Load Monitoring (NILM) even more challenging.

We observe a wide range in the quality of our user's data, and consequently have discovered a variety of problems that make practical large-scale electricity disaggregation more difficult. In this talk we highlight some of the common problems with: meter hardware, lack of true knowledge of the appliances in a home, and the challenges of working with complex data even when the data is high quality.

1. INTRODUCTION

Figure 1 shows a single day of electricity consumption for a PlotWatt user in Durham, North Carolina. Measurements were made with a TED1000 at 10-second intervals. Visual inspection of the trace indicates a refrigerator, some cooking and other loads in the morning, and likely air conditioner operation through the afternoon and evening.

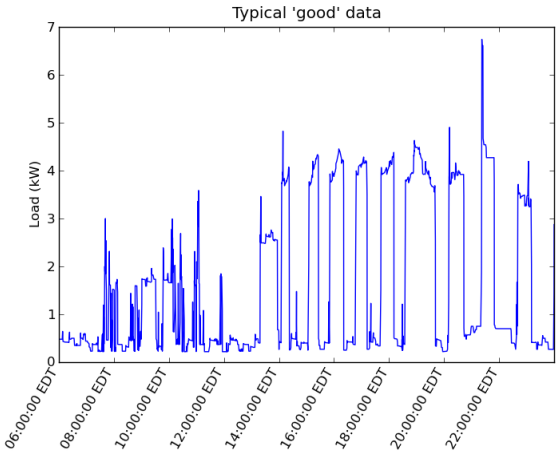

Figure 1: Whole-house energy use for a PlotWatt user.

Unfortunately, in PlotWatt's experience, the quality of the data shown in Figure 1 is exceptional. Most data we collect is not as clean, which makes disaggregation much more difficult. Typical problems and challenges include:

- Measurement limitations, such as a minimum reported reading level and minimum resolution for a device.
- Incorrect installation and setup of the device. For example, a current transformer may be flipped.
- Invalid measurements and data "spikes".
- Intermittent Internet connectivity.
- Lack of a "truth" model for the appliances in a house.
- Houses with large, complex loads.
- Low data rates. Sampling intervals of 30 seconds to 1 minute are typical. Longer intervals are possible, and much more challenging to disaggregate.

2. HARDWARE CHALLENGES

PlotWatt has been purposely agnostic with respect to hardware used for electricity monitoring. We support all the major consumer power meters, as well as data import from sources such as Green Button or Google PowerMeter. Figure 2 shows one problem we encounter with some optical-based power monitoring devices. These devices have a lower limit on the load it can detect during a 30-second measurement interval. Rather than reporting no load during the interval, this device will report the minimum reading of 400 W. Hence, we lose insight into the baseload as well as the ability to disaggregate usage during those periods when the consumption is below the floor.

Figure 3 shows low-resolution data from a current-transformer-based meter. This data improves following a firmware upgrade. The meter reported the load with a resolution of 60 W before the firmware update at 19:00 on the day shown. Following the update, the resolution improves to approximately 1 W. The noise spikes that are apparent in the pre-update data are also eliminated post update.

3. APPLIANCE TRUTH

PlotWatt rarely has access to the appliance "truth" for a home. While we ask our users to complete an appliance

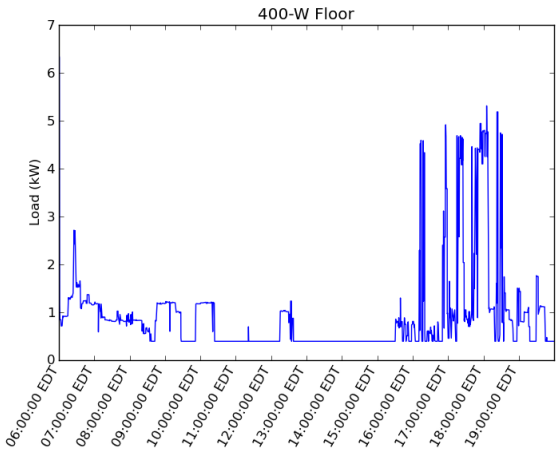

Figure 2: Readings with a 400-W floor on the reported values.

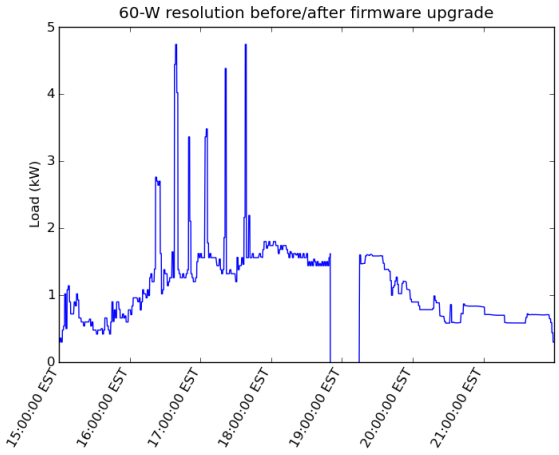

Figure 3: Readings before and after firmware update at 19:00 to improve resolution.

questionnaire when they register, the answers are neither complete nor infallible. Users generally know if they have a refrigerator or air conditioner. However, their accuracy is much worse on questions of electric versus gas water heater, electric or gas dryer, or even source of heat.

For example, Figure 4 shows a trace from one user's house on a day where the temperature ranged from 22°F to 43°F. This user claimed to have an electric heat pump in their questionnaire, yet there is no evidence for a heat pump on this day or any other.

4. COMPLEX DATA

Figure 5 shows a house with a large and complex load that often exceeds 100 kWh per day. With a high baseload, multiple large (3-5 kW) appliances, and numerous smaller ones, this house is particularly difficult to disaggregate. Yet, high-use homes have the largest potential reductions in consumption based on the insight gained from energy disaggregation.

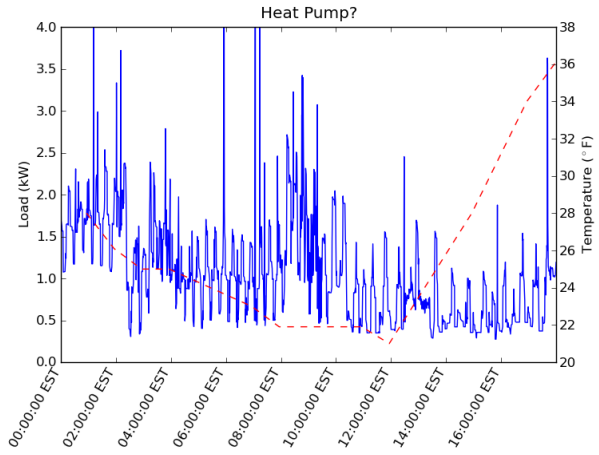

Figure 4: 25°F outside, yet no evidence of an electric heat pump.

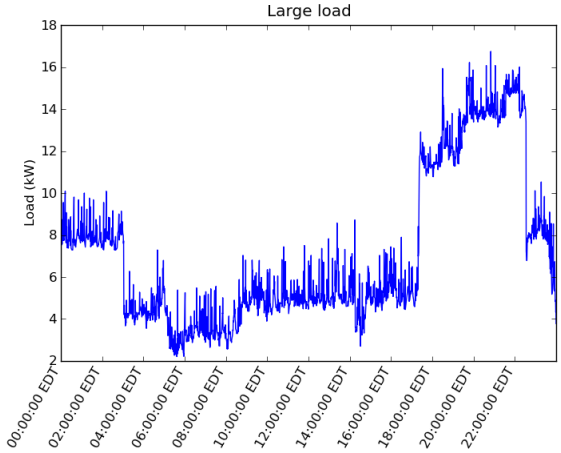

Figure 5: Readings for a house consuming 100 kWh per day.

5. CONCLUSIONS

Designing practical NILM algorithms that can operate on the very real problems of data in the wild is very challenging. We encourage the research community to embrace the problems outlined above:

- Design power meters that are affordable, accurate, reliable, easy to install, and support data rates up to 1 Hz.
- Develop techniques to identify the appliance "truth".
- Disaggregate electricity for high-use, complex homes.
- Construct algorithms that perform well at low data rates.

PlotWatt is prepared to open our dataset to support researchers to solve these challenges. Please contact us for further details.