# Setting the Benchmark for Non-Intrusive Load Monitoring: A Comprehensive Assessment of AMI-based Load Disaggregation

**Prepared By**

PECAN STREET

**Prepared For**

EEme

**January 2015**

## Foreword

The idea of using data collected from single point sensing at mains to infer appliance-level electricity consumption is not new. But technology to facilitate the collection of data required for such analytics has only recently started to become cost-effective and pervasive owing largely to the adoption of Advanced Metering Infrastructure (AMI), also known as smart meters, by utilities. The Edison Institute reported in its September 2014 study that approximately 43% of US homes have a smart meter installed[1]. Given the increasing availability and ease of access to smart meter data, AMI-based demand-side analytics is emerging as a scalable value proposition.

The academic community has been exploring the problem of energy disaggregation, also known as Non-Intrusive Load Monitoring, or NILM, for more than two decades. A clear consensus on the correct features, methods, and algorithms for solving NILM is yet to be reached. What complicates the issue further is the lack of standard datasets with extensive ground truth in which the algorithms can be validated. Reviews of the field have currently placed the accuracy of disaggregation algorithms on low frequency data - the kind provided by smart meters - to be around 0.55[2]. The reader is advised to see the Greentechmedia article that summarizes such studies[3]. These validation-oriented reviews were subject to two limitations, namely, small sample sizes of a handful of homes and short time spans only covering a couple of months at most. Unfortunately, the utility smart grid ecosystem has not tapped AMI's full value as extensive validations are required before implementing new technology solutions – these reviews have failed to build a reliable business case for many stakeholders given their narrow scale and scope. For this reason, EEme, LLC decided to conduct a comprehensive 3rd party validation study based on a large sample size

[1] The Edison Foundation. September 2014. *Utility-Scale Smart Meter Deployments.*
[2] Carrie Armel et al. 2013. Energy Policy. *Is disaggregation the holy grail of energy efficiency?*
[3] Greentechmedia. November 2013. *Putting Energy Disaggregation Tech to the Test.*

of year-round ground truth residential load data to build the foundation for business cases that can leverage AMI analytics and load disaggregation.

With this validation study administered by Pecan Street, EEme has charted the boundaries of reliability and accuracy of load disaggregation. NILM solutions and vendors have historically suffered from lack of extensive 3rd party validation that is perceived as the foundation of technology adoption in the utility industry.

To our knowledge, this is the first time a commercial load disaggregation algorithm has been validated by a 3rd party so extensively and on such a large scale. We believe that testing the algorithm against a robust database of actual loads is the only viable method of converging on an accurate and reliable solution. To our knowledge, this is also the first time a commercial vendor has come public with its accuracy metrics for disaggregation algorithms, and, hence, set the benchmark for the industry. Given the scope of the validation, from both temporal and sample size standpoints, we are proud to state that EEme's algorithms are at the forefront of the disaggregation industry.

We are confident that the accuracy figures generated in this study can pave a reliable path for smart-meter- based analytics for customer segmentation and engagement in energy efficiency and demand response for utilities and demand-side management (DSM) program administrators. This in turn can help reduce capital and operational costs that pertain to the design, implementation and evaluation of DSM programs.

Enes Hosgor, Ph.D.
*CEO & Founder*
*EEme, LLC*

PECAN STREET

# A Comprehensive Assessment of AMI-based Load Disaggregation from EEme

January 2015

by
Bert Haskell, CTO
Grant Fisher, Data Group Director
April Hersey, Database Administrator

Pecan Street conducted this study to evaluate the load disaggregation algorithm developed by EEme, LLC, an energy analytics company spun out of Carnegie Mellon University.

## Sample and Methodology

Pecan Street tested the accuracy of EEme's disaggregation algorithm by applying the algorithm to a 264-home data sample. The algorithm was tested on a year of 15-minute interval whole-home electricity use from the 264 homes.

For all 264 homes used in the sample, Pecan Street had measured whole-home electricity use at one-minute intervals. It had also measured one-minute interval electricity use from 12 to 24 individual circuits in each home using current transformers (CT's) attached to the circuits.

Pecan Street testing staff used the circuit-label measurements as ground-truth data to calculate the effectiveness of EEme's disaggregation algorithm in identifying and measuring electricity use from individuals loads.

## Summary of Observations

When applied to a year of 15-minute whole home electricity use from 264 homes, EEme's algorithm had the following median monthly error ratios:

- -0.31 for heating, ventilation and air conditioning (HVAC) electricity use
- -0.28 for refrigerator electricity use
- -0.45 for clothes dryer
- 0.33 for dishwasher

## Background

A number of commercial offerings claim the ability to identify the presence, and measure the magnitude, of electricity use from individual end uses such as appliances through access only to whole-home smart meter data. These offerings typically consist of a proprietary algorithm that attempts to disaggregate this whole-home electricity use and provide measurements for select end uses.

Having accurate measurements of specific end uses can be helpful to utility planners developing and evaluating the performance of energy efficiency and behavior programs such as appliance rebates, bill insert information and demand response programs.

For example, a peak demand reduction message will generally prove more impactful if it provides specific, targeted suggestions on use of individual appliances that are relevant to the customer as opposed to general messages such as "Use less energy."

"*Consider setting your pool pump so it does not run between 3 pm and 7 pm*," can be an effective and high-value demand response message — but only for homes with pool pumps. And assessing whether the customer responded to a certain pool pump message requires determining whether any changes in electric use during a particular time frame were from the pool pump as opposed to other large loads with similar profiles such as air conditioners, electric clothes dryers, electric ovens or electric vehicles.

Alternatives to software-based disaggregation are systems that install additional equipment to directly measure individual appliances or circuits. (Major end uses in homes frequently have a dedicated circuit.) Hardware-based options include CT-based systems and plug load monitors.

Algorithm-based disaggregation products offer the promise of relatively low-cost disaggregation because they typically do not require any additional hardware for homes and businesses equipped with smart meters. However, demonstrating the accuracy of its claimed measurements is a challenge for any software-based disaggregation product.

Best practices for developing software-based disaggregation products consist of the following:

- Testing the product's disaggregation algorithm against a robust database of real-world load measurements where not only the whole premises but also individual end uses have been directly measured through hardware-based systems
- Validation of the disaggregation algorithm by a professional third party that
    - has access to a significant database of time-stamped, high resolution electricity use from homes (or businesses, where applicable) as well as time-stamped end use electricity use measurements for these same homes that was recorded using hardware-based measurement systems
    - tests the disaggregation product on a statistically significant sample of whole-premises electricity use data and compares the accuracy of the algorithm's results with the actual end use measurements for these same homes
    - produces an independent report on the results of its testing and specifically on the accuracy of the algorithm product

## Testing process

With the whole home and circuit-level measurements of anonymized homes from this library, Pecan Street evaluated the accuracy of EEme's electricity use estimates for four residential end uses compared to actual values measured in 264 homes.

Pecan Street has developed and maintains what appears to be the world's largest database of disaggregated residential electricity use, with over 1,000 home years of appliance-level energy consumption measurements from actual homes. Most measurements are recorded at one-minute intervals. (Nearly 75 homes have appliance-level electricity use recorded at one-second intervals.)

To carry out this evaluation, Pecan Street provided 15-minute interval whole home data from the 264 homes to EEme's software.

Pecan Street selected 15-minute input data because this tracks with the performance scenario for a software-based disaggregation product, which typically perform their calculations by applying the software's statistical processes to a time-stamped, whole-home energy data stream, such as from a smart meter.

The interval for this data stream varies depending on the resolution setting of the energy measurement device. Utility smart meters typically record electricity use at intervals of one hour or 15 minutes.

For the validation study reported here, Pecan Street provided two sets of input data to EEme: 1) whole-home electricity-use data for 12 months collected at 15-minute intervals for 264 single-family houses in Texas; and 2) historical weather data corresponding to the 12-month period. The average household in this sample consumed 11,132 kWh over the course of the 12-month period.

EEme processed the 15-minute interval data using its proprietary algorithms developed at Carnegie Mellon University, and disaggregated four end uses:

- HVAC
- refrigerator
- dishwasher
- clothes dryer

EEme provided its calculations to Pecan Street. Pecan Street then compared the disaggregation results against the ground truth circuit-level data for the four end uses.

### Absolute error

Absolute error is a useful metric for determining how accurate a disaggregation algorithm's measurements are compared to the actual recorded use.

$$\text{Absolute Error} = \frac{\text{Inferred appliance use} - \text{actual use}}{\text{Actual use}}$$

Pecan Street used the following formula to measure the absolute error of EEme's reported appliance use to actual appliance use. It calculated absolute error for individual months, and absolute error was also calculated for each household and each of the four tested end uses.

## Relative error

Relative error speaks to the reality that the real world consequences of an absolute error are heavily influenced by the significance of the activity for which the error was measured. Larger percentage errors on relatively small uses, for instance, may have less impact than smaller errors for large uses.

Relative error calculations thus can help normalize the contribution of any particular appliance to overall house consumption and may be helpful in home energy management applications. It is calculated using the following formula:

$$\text{Relative Error} = \frac{\text{Inferred appliance use} - \text{actual use}}{\text{Total home use}}$$

Because this metric normalizes the error for total home use, it puts the uncertainty in perspective from an end-user standpoint. End-users who ultimately use such disaggregation-based energy insights may be more likely to evaluate these insights relative to their total bill, e.g., “How much can a new refrigerator reduce my total electricity bill?”

In such a framework, a 20 percent absolute error that led to overestimating savings by 10 cents would be less consequential than a three percent absolute error that led to overestimating savings by $10.

# Results

EEme’s algorithm had the median monthly error ratios over a full year depicted in Table 1.

Negative values mean the algorithm underestimated actual use, while positive values reflect an overestimation.

| | Absolute error | Relative error |
|---|---|---|
| HVAC | -0.31 | -0.11 |
| Refrigerator | -0.28 | -0.02 |
| Clothes Dryer | -0.45 | -0.02 |
| Dishwasher | 0.33 | 0.003 |

Table 1