Duke Energy’s Residential Energy Research Project: Developing Optimization Models and Analytics at the Household

David Mulder, Leidos Engineering, Cincinnati, OH
David Lawrence, Duke Energy, Charlotte, NC
Chris Bogart, Ph.D., Leidos Engineering, Charlotte, NC

ABSTRACT

Duke Energy is developing and testing algorithms for use in its operations as part of a distribution grid optimization initiative. Optimization is about increasing utility distribution grid operational efficiency, which includes: 1) providing stable and quality power to its customers; 2) increasing reliability through less outages and faster restoral times; and 3) reducing energy losses from more efficient operations. A key optimization opportunity lies with choreographing customer load for those operating on the same phase or feeder section. The objective of the two year Residential Energy Research Project (RERP) is to collect information in conjunction with other technologies installed (meters, line sensors, communication nodes, weather stations, etc.) to understand how granular customer information – when synchronized with other smart grid information – could provide grid operational value to Duke Energy.

Overview

McAlpine is a test area in Charlotte, N.C., where Duke Energy's Emerging Technologies Organization (ETO) installs, tests and evaluates new technologies and applications. The area consists of one substation, six distribution circuits and approximately 15,000 customers. Technologies installed to date include smart meters, distribution line sensors, weather stations, smart appliances, PEV chargers and energy storage devices.

At Duke Energy, ETO is responsible for the identification, testing, development and evaluation of new transmission, distribution, information technology and telecommunications technologies. During one of these projects, ETO wanted to understand the impact of consumer appliance usage and weather data on residential electric usage patterns. To do this analysis, ETO deployed three new technologies to collect granular environmental information: (1) residential home circuit-level dataloggers, (2) pole-mounted weather stations and (3) power quality meters. The goal of this work was to determine the unique appliance usage patterns that can not only be applied to interval data but also used to deliver enhanced operations and new products and services for consumers. Table 1 below provides a summary of the sensors used for the project.

.

Table 1: Technology Summary

Technology Project	Sensor & Data Rate	Purpose
Residential Circuit-Level Datalogger	Sensors installed on 24 different circuits within the participant's home. Data collected at one-minute intervals.	Track how customers use energy across all circuits with focus on the major appliances.
Power Quality Meter	Four power quality meters installed in the geographic service area; 222 variables at one-minute intervals. Distribution grid operations variables (voltage, current, harmonics, frequency, loading (kW), etc.).	Track operating parameters of electricity as it moves through the distribution grid. Evaluate value of these parameters at the "edge" of the distribution grid. Understand distribution grid operations on a granular level, and how it may be influenced by customer usage and weather.
Local Weather Station	Seven tower-mounted weather stations across the test bed area to collect granular weather information.	Understand impact of weather on residential usage and power quality parameters. Which weather variables are the best predictors of energy usage? Which weather variables are the best predictors of energy generation? Is there a relationship between weather and grid operations?

Figure 1 below shows where the residential dataloggers (in red) and weather stations (in blue) are installed.

Figure1: Location of technology equipment within McAlpine Test Area

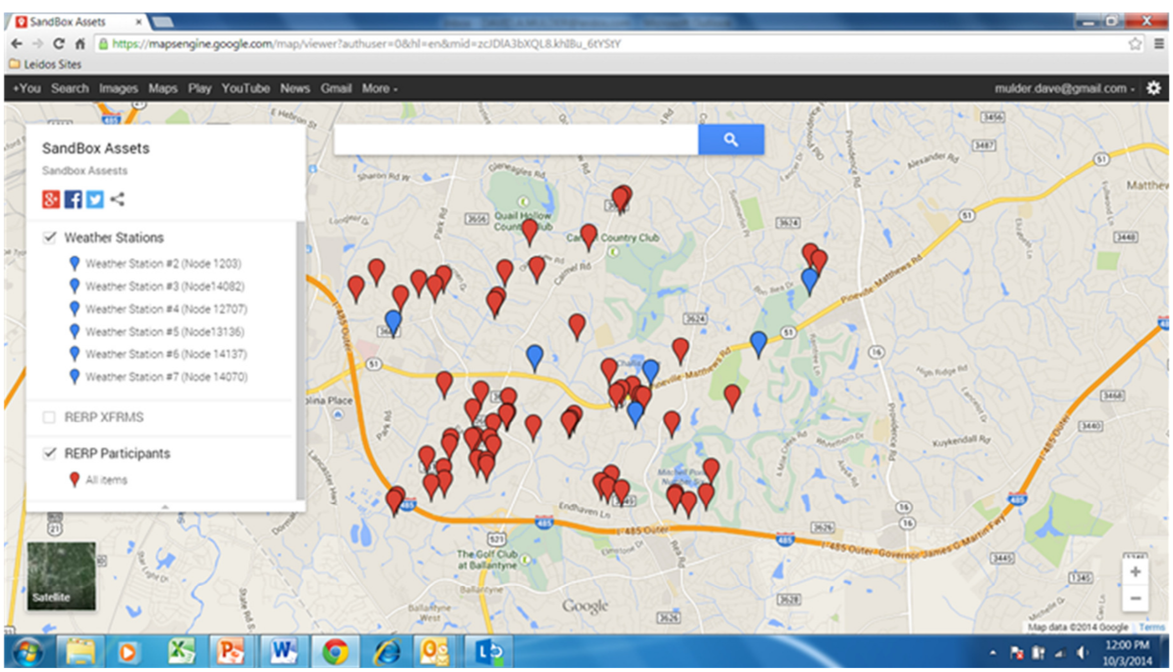
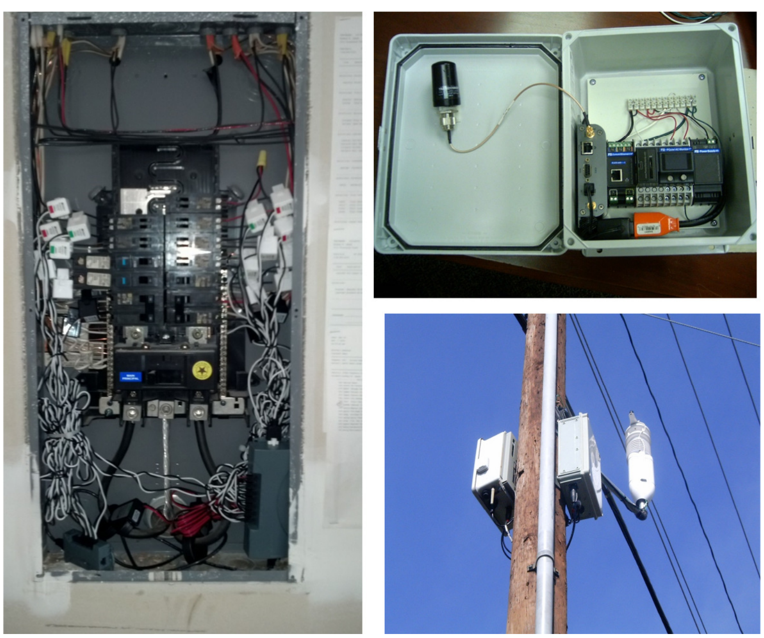


Figure 2 below provides a pictorial overview of each of the technologies that have been deployed. The equipment in the picture at the left is the datalogger system installed in the customer's circuit panel. At the bottom right of the picture is the datalogger device. This collects the data from each circuit and transmits it wirelessly to a gateway that is connected to the residential customer's broadband router. This data is then sent to the datalogger host system, where it is used to populate a customer energy dashboard. This data is also sent daily to a "big data" platform installed in Duke Energy's Charlotte data center, where the analytics are performed.

The equipment in the picture at the top right of Figure 2 is the power quality meter and cellular modem. This is installed at several homes and at a transformer in McAlpine. Data is collected and sent daily via e-mail to Duke Energy. Configurable alerts and alarms are also sent in real time via e-mail.

Finally, the equipment in the picture at the bottom right of Figure 2 is the weather station and associated communications infrastructure. The tall cylindrical object is the weather station. The box next to it contains the datalogger and weather station communications equipment. This is connected via Ethernet to a Duke Energy Communications Node, which is the box on the left of the pole. The Communications Node is the core component of Duke Energy's grid modernization infrastructure in McAlpine. It is also used to collect and route data from smart meters, distribution line sensors and distribution line conditioning equipment.

Figure 2: Technologies Installed for Project



Whereas Duke Energy can use smart meter interval data to gain insight as to how an entire home may use energy, it still does not know how it is broken out by the major appliances. The datalogger solution collects information on up to 24 different circuits within the home. Duke

Energy has standardized its circuit-labeling convention to allow it to compare appliance usage characteristics across different households. Finally, information is being collected at 1-minute intervals, for a more granular view of energy usage activity.

With this information, Duke intends to learn a variety of things:

- Customer Usage Characteristics by Appliance
  - Which appliance uses the most energy?
  - What is the most energy used throughout the day?
  - What influences energy use?
  - Which appliances run most consistently throughout the day (defined by the industry as "base load" usage)?
  - Which appliances run during Duke Energy's peak period (defined by the industry as "peak load" usage)?
  - When does the customer's energy usage peak each day?
  - Can Duke Energy demand response, pricing and energy efficiency programs influence customer usage?
- Customer-to-Customer Comparisons
  - Can we do a cross-sectional analysis of similar appliances, end uses or circuits?
  - Can we compare the different home profiles for usage differences?
  - What are the impacts of customer energy usage on transformer loading throughout the day? Which appliances impact this the most?
- Weather Effects on Residential Usage
  - How does weather affect the different appliances?
  - Which weather variables are the best predictors for the different appliances?
  - Is there a lag effect between weather and appliance usage? How long is this effect if there is one?

Residential Energy Research Project (RERP) Participant Selection Process

There are more than 15,000 customers located on the six circuits within McAlpine. Also, Duke Energy has deployed a variety of other technologies, applications and customer programs. The goal of this project was to get as wide a representation as possible of the different activities underway. Therefore the selection process included the following criteria in Table 2 and Table 3 below. Prospective participants did not have to meet ***all*** the criteria; however, the goal was to have representation of all the criteria within the final sample set of customers selected to participate in the project. (For example, research showed that only 5 people owned Plug-In Electric Vehicles (PEVs) in the McAlpine area; therefore, all were sent an invitation to participate.) Out of the approximately 15,000 customers in McAlpine, around 5,300 met at least some of the criteria below and had verifiable contact information. Duke Energy's Customer Prototype Lab is managing the RERP for the ETO, and sent out e-mails and/or direct mail to the 5,300 customers inviting them to participate. Participation required the installation of the datalogger equipment at the home (a three-hour appointment), use of the customer's broadband service to send data to the host data center daily, and commitment to be involved in the project for two years. Furthermore, at the end of the project, Duke Energy would remove the equipment. For participation, the customer would receive $100 at the beginning and end of the project, and free use of a customer energy portal application to review and export his/her data.

Table 2: Participant Selection Criteria

Home size and age diversity
Household size
Duke Energy program(s) participant (See Table 3 for a list of programs)
Heating fuel type
Appliance diversity
PEV owners
More than one on the same transformer
Participants on all six distribution circuits in McAlpine
Control Group: customers not participating in any Duke Energy programs

Table 3: Duke Energy Program Participation by Participant

Duke Energy Program	# Participants
Bill Payment Program	36
Power Manager Peak Reduction Program	25
North Carolina Green Power Program	3
Home Energy Manager Program	4
Smart Thermostat Program	4
Battery Energy Storage Program	2
Smart Appliance Program	2
Control Group (Not Involved in any Program)	17

To the delight of the Customer Prototype Lab and ETO, more than 800 customers responded within the first week indicating interest in participation. 61 participants were ultimately selected to cover the criteria above. Installations were completed during 4Q 2013, and data collection efforts have been underway since late 2013. The project study period is 2014-2015. Equipment removal will occur 1Q 2016. Table 4 below provides summary demographics information on the participants.

Table 4: Demographics Summary

	Age of Home	Size of Home (sqft)	Number in Home
Min	10	1,000	1
Mean	26	2,607	3
Median	25	2,500	3
Max	43	7,000	6

Big Data Collection Activities

From the outset of this project, Duke Energy utilized its existing technology infrastructure and expertise for selecting the technology, installing the equipment and backhauling data utilizing its communication infrastructure. However, it lacked the big data analytics environment necessary to jump-start the data analysis and strategy development envisioned for this work. Consideration of big data analytics software products currently available highlighted the technical benefits of increased scalability, flexibility and performance by moving towards non-structured data storage. This would allow collection and easy integration of existing data streams from smart meters with new sensor data streams in formats not bound by preconceived/static data relationships. This ability to remove the legacy bias of existing data stores and discover new unknown relationships would provide the analytics power needed to revolutionize the use of utility data to discover new optimization benefits, revenue streams and operational efficiencies.

Duke Energy selected Leidos Engineering to implement its big data analytics platform, Scale2Insight (S2i), providing the underlying data analytics for the RERP. A general-purpose predictive analytics solution, S2i, was selected due to its ability to reduce the exploration time in dealing with new data sets (where the correlations and relationships between data elements were not understood at the start of the project). The primary benefit of using a general-purpose predictive analytics solution, such as S2i, is that it allows data consumers and data analysts to broadly and rapidly work through the cycle of data modeling, visualization and deployment on almost any imaginable use case – today or in the future – with minimal time invested and a modest up-front infrastructure investment. Additionally, S2i was selected because it utilized an unstructured database format that would provide the company with experience in several new technologies related to analytics, data storage and data visualization. This experience was warranted, as the data collection efforts have expanded to include new datasets, including smart meter interval data and additional socioeconomic and demographic data. Table 5 below summarizes the amount of data that has been collected through September 2014.

Table 5: Device Data Set Summary

Device	Residential Datalogger	Weather Station	PQ Meter
# Devices in Project	61	7	4
# Data Elements/Device	149	66	222
# of Records	14.1 Million	1.4 Million	1.6 Million
Total Data Elements Collected	2.1 Billion	91 Million	345 Million
Size of Database	50.3 GB	2.1 GB	8.3 GB

The team initially developed descriptive statistical models to perform simple data validation and outlier analysis of the sensor technologies. Beginning with data mining models, the team identified key statistics about the data sets and developed an understanding of the available information and structure. This understanding allowed the team to use the other model types to obtain more fine-grained and predictive knowledge of relationships in the data. Each turn of the predictive analytics cycle resulted in more enriched workspaces and collaborative models where team members could leverage the work of their teammates and develop more evolved use cases. This flexibility was crucial to the cycle of use case development and analysis as it allowed new questions to arise and be tackled without having to re-formulate each new problem to fit a single algorithm.

Use Case Development and Insights

Once data had been ingested into the big data platform, Leidos worked with Duke Energy personnel to brainstorm and construct use cases that could be used as the foundation for model development. A sample selection of these use cases is described in Table 6 below. Duke Energy will continue to develop use cases through 2015 as new datasets are added and more historical data is collected.

Table 6: Examples of Use Cases Developed and Modeled

Use Case Name	Use Case Description	Technologies Used
Data Validation and Outlier Analysis	Three separate use cases, one for each technology, were created to analyze and validate that the data being collected is complete and accurate. Analysis included calculation of descriptive statistics (min, max, average and median), and outlier analysis (# of times values fall outside pre-determined thresholds).	Circuit data logger Power quality monitor Weather station
Customer Peak Analysis	This use case analyzed circuit data logger participants' usage for one week during the summer. The customer peak usage interval was identified, along with the number of times that the total house usage reached 95 percent or greater of the customer peak. Each customer peak and timestamp was documented. Also, the customer peak was broken out by the major circuits that contributed to it.	Circuit data logger
Transformer Coincident Peak Analysis	For the same week as above, the total load of each customer was summed together. The highest values across all customers were the transformer coincident peak. Once the peak was found, Leidos Engineering and Duke Energy determined the number of times that the total load was within 95 percent of coincident peak.	Circuit data logger
Predicting Total Customer Peak Load based on Appliance Ownership	Appliances drive total customer load. Therefore, by knowing what appliances are in the home, Duke Energy may be able to predict both total customer energy use, and their peak usage amount. Predictive	Circuit data logger

Use Case Name	Use Case Description	Technologies Used
	models were developed that focused on specific appliances as indicators. These included air conditioners, electric range, electric dryer, refrigerators and PEVs.	
Predicting Customer Energy Usage Using Weather Variables	Utilities know that weather can influence total customer loads and peak period usage. Historically, Duke Energy has had limited weather data; moreover, this was hourly data from the closest weather station. Duke Energy now has extensive variables to choose from. Therefore, Leidos Engineering and Duke Energy should be able to develop better models to predict customer load intervals by utilizing weather information.	Circuit data logger Weather station

Below are summary results of several of the use cases. The models, algorithms and results of these use cases will continue to be refined as data is collected throughout 2015.

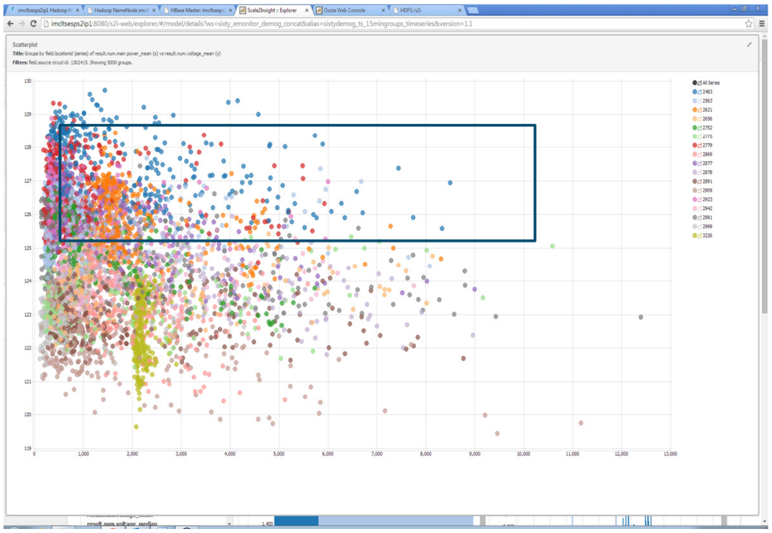
Use Case: Data Validation and Outlier Analysis

Description: Analyze and validate that data being collected is complete and accurate. Analysis includes calculation of descriptive statistics (min, max, average and median) and outlier analysis (# of times values fall outside pre-determined thresholds).

Insights: The objective of this use case was to get a quick descriptive view of the datalogger dataset, and to determine whether the data elements were valid according to user-developed thresholds. A key finding from this analysis focused on the names of the data elements themselves. For the original datalogger pilot, the data elements essentially were the names of the circuits being monitored. These were typically developed and inputted by the installer. Unfortunately, misspelled and misnamed data elements made comparative analysis of these first five homes difficult. As a result, Duke Energy developed a standardized list of labels that were used for the RERP.

The analysis also found multiple instances where key variables (e.g. total house load, voltage) fell outside of the defined thresholds. In one case, total load was negative; we found that this home has a solar PV system that produced load on the system. In other case, voltages in the spring months were above 130 in many instances. Figure 3 provides voltage information for a subset of customers in the RERP. The original focus of the analysis was to determine whether certain variables (in this instance, voltage) fell outside of pre-determined thresholds. The original model results indicated that there were thousands of instances where this occurred; specifically, they were above the upper threshold. Further analysis indicated that these were occurring for a certain subset of customers, all on the same circuit. As a result, Duke Energy operations will investigate this to determine what may be causing this event.

Figure 3: Voltage Analysis for Select Participants on a Common Feeder



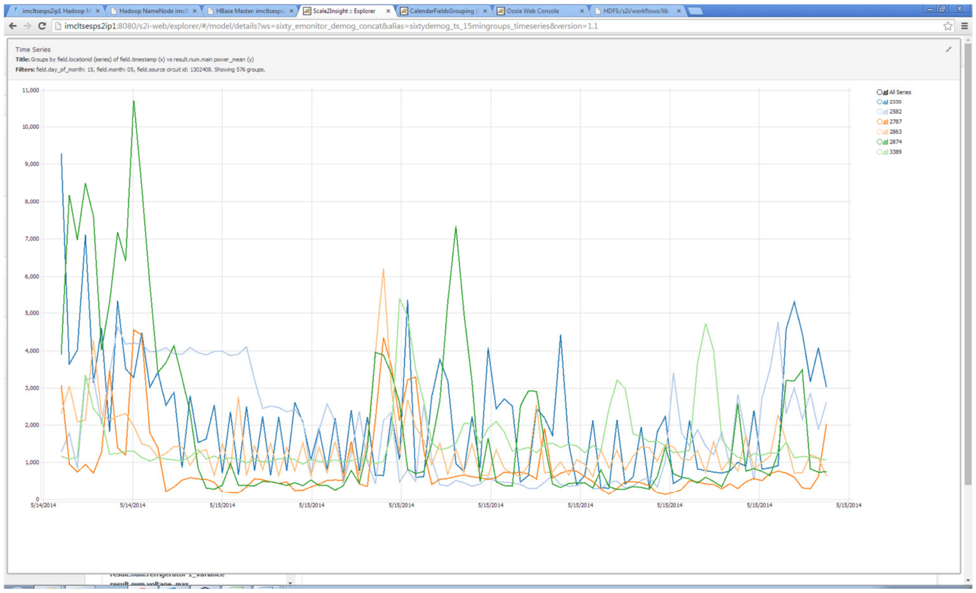
Use Case: Customer Peak Analysis

Description: This use case analyzed the participants' usage for one week during the summer. The customer peak usage interval was identified, along with the number of times that the total house usage reached 95 percent or greater of the customer peak. Each customer peak and timestamp was documented. Also, the customer peak was broken out by the major circuits that contributed to it.

Insights: The biggest contributors to customer peaks were the AC, dryer, electric heat pump, electric water heater and the oven. However, a surprising result was how large a contributor the electric dryer was. In more than 30 percent of the days recorded, the electric dryer represented 50 percent or more of the total customer load at the time of the ***customer*** peak. Whereas this may not always synchronize with the utility system peak, this may come into play from a transformer loading and distribution grid optimization perspective.

Figure 4 below plots whole house loads for select customers, on a particular day during the week. In one instance, whole house loads were more than 11 kilowatts (kW). Further analysis broke down the whole house loads into the different circuits. As expected, central air conditioning represented a significant load at approximately 3 kW during the peak time. More surprising was the electric dryer that used more than 6 kW of load when operating. Depending on the time of day, this can contribute significantly to the system peak or transformer loading. Furthermore, electric dyer use is a year-round issue.

Figure 4: Whole House Load, Select Participants, May 15, 2014



Use case: Transformer Coincident Peak Analysis

Description: This use case analyzed the customers served by the same transformer. First, the total load of each customer was calculated and summed together. Next, the highest values of the sum of customers were determined; this was the transformer coincident peak. Once the peak was found, it was then possible to determine the number of times that the total load was within 95 percent of the coincident peak.

Insights: The coincident peaks of weekdays seem to have a smaller variance as opposed to the coincident peaks of weekends. Concentrating on just Monday vs. Sunday saw an even greater difference between the coincident peaks. The coincident peaks for Monday have a very small stationary variance and show a clear trend over time. The coincident peaks for Sunday have a large variance, and the trend is difficult to spot. Due to the difference between weekdays and weekends, it would be recommended to have at least two models for predicting coincident peak. It would even be worth exploring seven different models (one for each day of the week).

Use Case: Predicting Total Customer Peak Load based on Appliance Ownership

Description: Appliances drive total customer load. Therefore, by knowing what appliances are in the home, Duke Energy may be able to predict both total customer energy use and their peak usage amount. This use case focused on specific appliances as indicators. These include AC, electric range, electric dryer, refrigerators and PEVs.

Insights: A variety of models were run in an attempt to develop predictive models under a variety of appliance scenarios. The models attempted to predict one-minute demand intervals out an hour. The random forecast model in particular does well following the main power movement but always misses low when forecasting spikes in main power. Forecasting these spikes is a major task and should be further examined.

There seems to be evidence to suggest that main power load can be forecasted using appliance loads as predictors. However, more houses are needed with overlapping key appliances to verify that these forecasting models can be used for different houses. Additionally, the current research only used an arbitrary day for analysis and future work should verify that these models work across all days. As the project expands in the coming year, the additional data will drive the model to higher accuracy and reliability.

Use Case: Predicting Customer Energy Usage Using Weather Variables

Description: Utilities know that weather can influence total customer loads and peak period usage. Historically, Duke Energy has had limited weather data and most models used Total Degree Days (TDD). Moreover, this was hourly data from the closest weather station. Duke now has extensive variables to choose from the local weather stations it installed. Therefore, Duke Energy should be able to develop better models to predict customer load intervals by utilizing weather information.

Insights: This use case utilized data from all three of the new sensor technologies. The tower-mounted weather station, residential circuit-level data logger and the power quality monitor. The variables from the tower-mounted weather stations included average solar radiance, average air temperature, average relative humidity and average wind speed. Data is being analyzed at one-minute, 15-minute, 30-minute and 60-minute intervals. By combining this data with the residential circuit-level data logger information and the power quality output, Duke Energy could better understand what weather variables are most reliable for predicting customer usage and generation patterns based on the "appliance fingerprint" of the home. They could also understand if there is a relationship between weather and grid operations through the power quality meter.

Temperature, solar radiation and wind speed were examined as possible predictors for whole house energy usage. Various multiple linear regression models were used to evaluate prediction of usage over different time periods (one-day, four-hour and one-hour intervals). The multiple regression models did not capture short-term variations that may occur over the one- or four-hour periods, but they did successfully capture the general trend in power usage over a day. Results of the modeling found that temperature and humidity were better predictors of main power usage than solar radiance or wind speed. A key insight from the analysis was that in order to refine the predictions of energy usage from weather variables, it would be necessary to apply nonlinear predictive methods going forward.

Next Steps

Duke Energy will be collecting information from these customers and other equipment in its McAlpine test bed through December 2016. Other operational data is being collected from distribution line sensors, AMI data, transformer monitors, smart appliances and PEVs. Additional activities that are planned or underway include:

1. Adding socio-economic variables to the dataset and incorporating the information to refine the original models developed.
2. Additional participant data collection: one set of data not collected during the installation process was the ages of main appliances. A survey will be sent early next year requesting this information.
3. Asking customers if they have reviewed or utilized their energy information via access to through their customized energy portal.
4. Finally, all the customers in the project have smart meters. Models will be developed to compare the circuit data with the meter interval data to determine if there are disaggregation algorithms that can be developed using the circuit data and how it apply it to other meters without the granular data.