DOCKETED

Docket Number:	16-IEPR-05
Project Title:	Electricity Demand Forecast
TN #:	212361
Document Title:	Small Versus Big Data
Description:	N/A
Filer:	System
Organization:	Michael Cockayne
Submitter Role:	Applicant Consultant
Submission Date:	7/20/2016 9:19:06 AM
Docketed Date:	7/20/2016

Comment Received From: Michael Cockayne
Submitted On: 7/20/2016
Docket Number: 16-IEPR-05

Small Versus Big Data

Additional submitted attachment is included below.

Michael Cockayne
Post Office Box 4014
Redondo Beach, CA 90277
(310)598-9722
Mike@LoadForecast.net

LoadForecast.net

July 20, 2016

Comment on Joint Agency IER Workshop on Energy Demand Forecast and Doubling of Energy Efficiency – Data and Analytical Needs

Role in the Proceeding

I am the Principal at LoadForecast.net. LoadForecast.net is currently under contract to LADWP to provide services related to Load Forecasting on a part-time basis. The opinions that follow are my own.

Connection between Forecast and Energy Efficiency Data Needs?

Listening to the Workshop over the phone, it was not clear to me that anyone detailed the connections and differences between data needs for Load Forecast and the Doubling of Energy Efficiency problem. What follows is my view of the nexus between the two.

Small Data versus Big Data

For purposes of this discussion, I am using rule of thumb type definitions for Small Data and Big Data found on the World Wide Web.

Small Data is structured data stored on a single server that is readily comprehensible by humans.

Big Data is messy data stored on multiple servers better approached by machines learning techniques.

Load Forecast Data is Small Data

All the data necessary to create long-term forecasts for all the IOUs and POUs in the State of California could be stored on a single instance of one of the web services providers. (Computer Scientists may suggest better ways to optimize access and design the database.)

Energy data at the aggregate level is already abundantly available on the Web. Aggregate energy data is already determined to a public good.

Michael Cockayne LoadForecast.net
Post Office Box 4014
Redondo Beach, CA 90277
(310)598-9722
Mike@LoadForecast.net

Although the aggregate energy data is available, the purpose of the Web publication of the data has never been for Load Forecast purposes. Much of the data is organized by revenue class for financial analysis. Currently, energy data on the Web tends to be stored as snapshots in time rather than in the form of time series. With the Commission's proposed move from annual forecasts to 8760 forecasts, it suggests the State is moving beyond forecasting events and is now interested in forecasting energy flows. I suggest that the State make available structured datasets of 8760 energy flows on a single server as a starting place.

I suggest making multiple views of the data. An obvious view would be Title 20 class by entity but you could also break it down into climate zone or receiving station. The Title 20 classes might be disaggregated into building types.

In addition, the server should have 8760 shapes for Demand Response, Energy Efficiency, Distributed Generation and Electric Vehicles.

In addition to the energy data, I would add weather, demographic and economic data to server all of which is currently publically available.

Obviously, I cannot design a Load Forecast Data Server in this comment but hopefully the concept is clear So much time is spent in data collection for Load Forecast at different entities around the State that it seems like you could raise efficiency by centralizing the process. If now is not the time for State Load Forecast data server, I think it will be in the near future.

Load Forecast Data is a client of Big Data

Most utilities in the state are at some stage of building Advanced Meter infrastructure (AMI). Without performing a utility survey, I would guess that the current AMI systems are mostly used as billing solutions. AS AMI is built out through the State of California, the opportunity for creating Big Data solutions and performing analyses grows.

I think the long-run view is that Big Data collection will replace the current load research which is based on sampling. If you ask a current AMI manager if it is feasible to replace the load research samples by populations, the answer is likely to be "not feasible". The two main issues being missing reads and bad reads. But it is not difficult to imagine a machine-driven sampling process that is closer to populations than we are today. Maybe you could do sequential dynamic sampling using machine-driven process. A search of the Web for "dynamic sampling" shows that there are tools available waiting to be applied to the energy sector. Big Data could have a huge impact on making data collection more efficient which is the area where a huge amount of the time in the load forecasting process is spent now.

Michael Cockayne
Post Office Box 4014
Redondo Beach, CA 90277
(310)598-9722
Mike@LoadForecast.net

LoadForecast.net

Load Forecast data is one of many clients of Big Data. From Big Data, well-structured datasets will be developed as inputs to the Load Forecast.

To answer the question of the workshop, I believe you will need Big Data to support the State's goal of doubling energy efficiency by 2030. From Big Data, similar to what is done with Load Research, you will be able to identify problems that will generate ideas for emerging technologies and aid in the design of TOU and Demand Response programs. Also Big Data will support the excellent creative work that is being done as demonstrated in Data Analytics portion of the Workshop.

Will Big Data Replace the Need for a Load Forecast?

Maybe. Eventually? From my training, machine-processed forecasting systems tend to be better at short-term forecasts than long-term forecasts. I think this tendency is due regime change which causes stationarity problems. The very goal of doubling energy efficiency by 2030 is call to change regimes. Human judgement will be required to bridge these changes in a long-term forecast.

Biggest Hurdle is Personnel

POUs are hard-pressed to keep up with field work in maintaining AMI. Personnel is not on-hand or current personnel would have to be re-trained to manage a Big Data component to the AMI. Big Data processing is a function currently under-budgeted or perhaps not budgeted at all. Young professionals who could be trained in Big Data think it is dead end career under-valued with no path to higher-salaried management jobs. Civil Service rules need to be changed to hire entry-level data scientists.