HSC Research Report

HSC/13/15

Energy forecasting: Past, present and future

Tao Hong

University of North Carolina at Charlotte, USA

Hugo Steinhaus Center
Wrocław University of Technology
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
http://www.im.pwr.wroc.pl/~hugo/

Energy Forecasting: Past, Present and Future

Tao Hong

Abstract

When turning on the switch, people expect the light would be on. However, the business to keep the lights on is not that straightforward. Dr. Tao Hong offers a practical overview of energy forecasting, an important task that electric utilities have been doing every day for over a century.

Key Words: Energy forecasting; load forecasting; smart grid.

Key Points

- Energy forecasting covers a wide range of forecasting subjects in the utility industry, such as short term load forecasting, long term load forecasting, spatial load forecasting, electricity price forecasting, demand response forecasting and renewable generation forecasting.
- The storage limitation and societal necessity of electricity lead to several interesting features of energy forecasting, such as the complex seasonal patterns, 24/7 data collection across the grid, and the need to be extremely accurate.
- As a fundamental business problem, energy forecasting practices have gone through several important stages, such as the engineering approach with charts and tables in the pre-PC era, and the computer based methods in late 20th century.
- Smart grid investment and technologies brings challenges to the energy forecasting field, such as demand response forecasting and renewable generation forecasting. The century-old energy forecasting finds its new life in the smart grid era.
- Advancement of energy forecasting relies on following rigorous out-of-sample tests, understanding business needs, and learning from many disciplines such as statistics, electrical engineering, meteorological science, etc.

Hong, T. (2014). Energy forecasting: past, present and future. *Foresight: The International Journal of Applied Forecasting*, (32), 43-48.

1. Introduction

Our electric power systems are the most complex man-made object ever, producing and delivering electricity to 5.6 billion people on the planet. Similar to airlines, consumer package goods, and oil & gas industries, the electric power industry needs forecasts of supply, demand and price, so called energy forecasts, to plan and operate the grid. Figure 1a illustrates a typical short-term load forecasting problem while Figure 1b depicts a typical long term load forecasting problem.

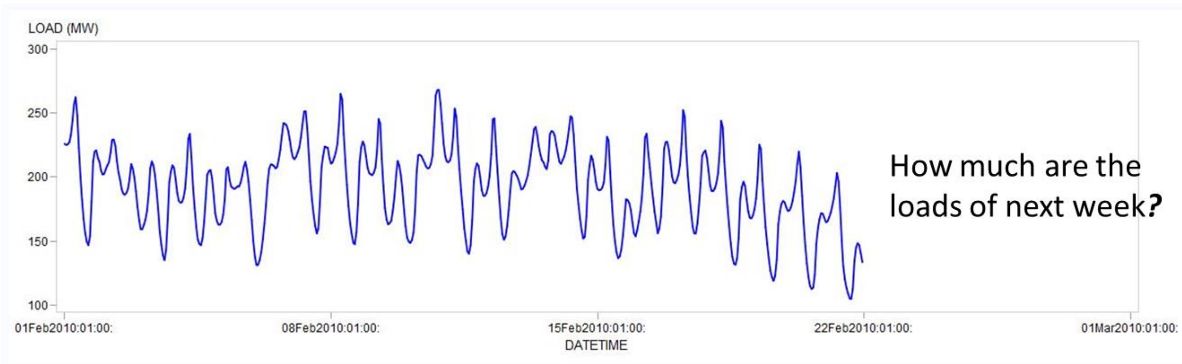


Fig. 1a Three weeks of hourly loads of a small utility. A typical short term forecasting problem is to predict the hourly loads for the next few days.

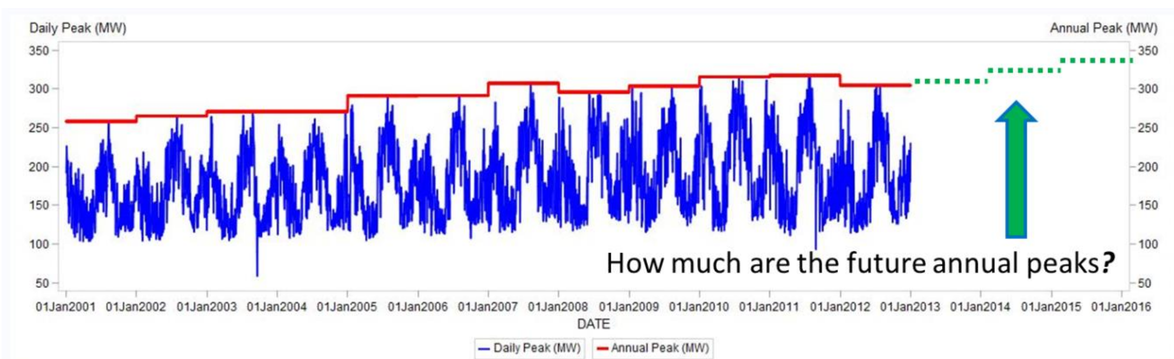


Fig. 1b Twelve years of daily peak load of a small utility. A typical long term forecasting problem is to predict the annual peaks for the next few years.

While many other industries have some form of inventory to store and buffer their products and services, those of the electric power industry, electricity, cannot be massively stored using today's technologies. As a result, electricity has to be generated and delivered as soon as it is consumed. In other words, the utilities have to balance the supply and demand **every moment**. The storage limitation and societal necessity of electricity lead to several interesting features of energy forecasting, such as the complex seasonal patterns, 24/7 data collection across the grid, and the need to be extremely accurate.

This paper discusses the evolution of energy forecasting practices in a chronological order starting with short and long term load forecasting in the pre-PC era. It then summarizes the computer based methods for spatial load forecasting, short term load forecasting and electricity price forecasting, and then moves to the emerging topics in the smart grid era, such as demand response forecasting, renewable generation forecasting, and Global Energy Forecasting Competition. Finally, the paper is concluded with three lessons learned from one hundred plus years of practices.

Hong, T. (2014). Energy forecasting: past, present and future. *Foresight: The International Journal of Applied Forecasting*, (32), 43-48.

2. The origin

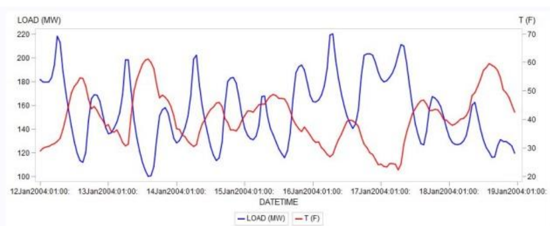
A. Counting the light bulbs

When Thomas Edison and his company developed the Pearl Street Station in 1882, his motivation was to promote the sales of light bulbs. This first steam-powered station initially served about 3000 lamps for 59 customers. When lighting was the sole end use of electricity, energy forecasting was straightforward. The power companies could just count how many light bulbs they installed and planned to install. Then they roughly knew the level of load in the evening. This ancient method is still being used in today's power systems planning, for forecasting the load of street lights.

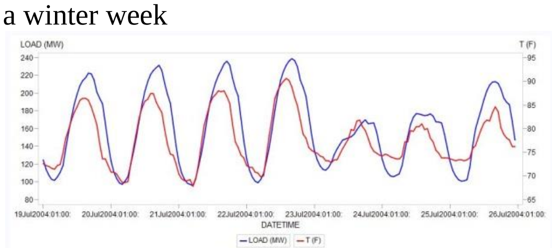
B. The engineering approach

As more and more electric appliances, such as electric iron, radio and electric washer, were being invented and became popular, the forecasting problem gradually turned to be non-trivial. Some special events, such as a presidential speech, may cause a spike in the load curve, because millions of people were listening to the radio at the same time.

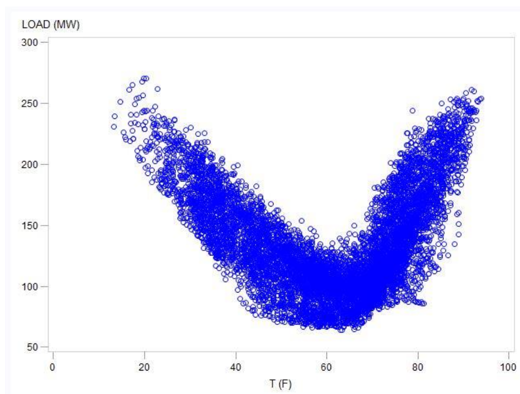
In 1940s, people found that electricity demand was highly affected by weather, principally due to high penetration of air conditioners. Figure 2 shows the relationship between load and temperature via two line plots and a scatter plot. In the winter, load and temperature are negatively correlated primarily due to space heating needs. In the summer, the correlation is positive primarily due to the space cooling needs. Since then, weather variables, such as temperature and humidity, have been widely used to forecast load.



a. Line plots of hourly load and temperature in a winter week



b. Line plots of hourly load and temperature in a winter week



c. A scatter plot of hourly load and temperature. The scatter shows an asymmetric “U” shape. On the left, the lower the temperatures are, the higher the loads are, which is primarily due to heating needs. The opposite happens on the right, which is primarily due to space cooling needs.

Fig. 2 Relationship between load and temperature.

Because there were no statistical software packages at that time, an engineering approach was developed to manually forecast the future load using charts and tables. Some of those elements, such as heating/cooling degree days, temperature-humidity index, and wind-chill factor, are inherited by today's load forecasting models. The similar day method, which derives a future load profile using the historical days with similar temperature profiles and day type (e.g., day of the week and holiday), is still used in many utilities' operations centers.

Hong, T. (2014). Energy forecasting: past, present and future. *Foresight: The International Journal of Applied Forecasting*, (32), 43-48.

3. Golden years

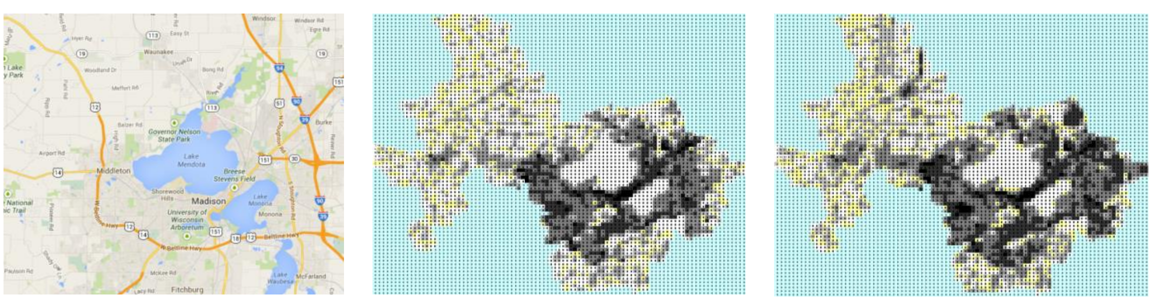
A. Spatial load forecasting

In 1980s, computer applications ramped up. A significant amount of research was devoted to long term spatial load forecasting about when, where and how much load growth will occur (Willis 2002). The forecasting horizon ranges from several years to several decades. These forecasts have been widely used in transmission and distribution planning. Most of them fall in three categories, trending, simulation and hybrid methods.

Trending methods look for some function to fit the past load growth patterns and estimate the future load. The most common trending method is to apply a polynomial regression model to load history data. The advantages of the trending method include ease of use, simplicity, and a short-range response to recent trends of load growth. However, the method often fails to provide a useful estimate of the long-range load, due to over-fitting or extrapolation of high ordered polynomials.

Simulation methods attempted to model the load growth process to reproduce the load history, as well as to identify the temporal, spatial, and magnitude information of the future load growth. They model an urban development process based on land-use information from government, customer rate classes from utilities, and load curve models of consumption patterns. Depending on the quality of data, this approach has had fair to very good short-range accuracy and good to excellent long-range usefulness for planning. Its drawback is expensive development and training cost.

Hybrid methods combine the favorable features of trending and simulation. An ideal hybrid should respond to the recent trend of load history in the short-range, keep the long-range defensibility provided by the simulation methods, and all this without requiring much skill and interaction from the user. A modern hybrid method was originally developed by Tao Hong (2008). The method divides the entire territory into thousands of 50-acre small areas. The small areas are then used to build a hierarchy with multiple levels. Load growth at each small area or region is assumed to be an S curve. The input information includes historical load at the 50-acre small area level, long term load forecast at the corporate level, and land use development plans. The parameters are estimated by minimizing the errors in historical fit and the difference between sum of lower level forecasts and the corresponding high level forecast. This method has been commercialized and deployed to many utilities in North America. Figure 3 shows the results from a case study of a medium size US utility, where the maps of the actual load and load forecast are plotted in MS Excel.



A medium size city of U.S. Actual load today Forecasted load 20 years later

Fig. 3 Long term spatial load forecasting

Hong, T. (2014). Energy forecasting: past, present and future. *Foresight: The International Journal of Applied Forecasting*, (32), 43-48.

B. Short term load forecasting

Computers not only helped improve the practices of spatial load forecasting, but also those of short term load forecasting (STLF). Late in the last century, the power industry went through a major structural change, which made accurate short term load forecasts even more critical. People first tried to apply statistical techniques, such as regression analysis and time series analysis, to STLF. Then Artificial Intelligence (AI) became one of the hottest terms in the scientific community, resulting in hundreds of papers reporting AI based approach to STLF. In 1990s, Electric Power Research Institute (EPRI) sponsored a project that developed several artificial neural networks based short term load forecasters (Khotanzad et al. 1998). Some research generated from this EPRI project was later commercialized and became a popular STLF service provider in today's industry.

The models based on AI techniques, such as ANN, fuzzy logic and support vector machine, were *black-box models* that appealed to organizations unwilling to build an in-house team of forecasting analysts. Many utilities were still not comfortable with black box approaches and instead developed forecasts using the classical method such as the similar day method, and statistical techniques such as multiple linear regression. Some utilities with an in-house forecasting team also build black-box models or purchase forecasts for comparison purposes. The most recent and comprehensive research on regression-based short term load forecasting was done by Hong (2010). Hong's methodology was soon adopted by many utilities, retailers and trading firms worldwide and became part of the engine of a commercial energy forecasting solution. Table 1 summarizes the key attributes of the different approaches to short term load forecasting.

Table 1: Comparison among Short Term Load Forecasting techniques

	Pros	Cons
Multiple Regression	Interpretable; easy to implement, update and automate; good accuracy	Relies on explanatory variables; needs explanatory variables, a designated functional form and at least two years of history
ARIMA	Few parameters to estimate; does not require a long history; good accuracy in (very) short term	Low accuracy in longer term; high cost to implement, update and automate; difficulty to interpret moving average
ANN	Minimum statistical or domain knowledge required; good accuracy during normal days	Heavy computation; over-parameterization; difficult to interpret; low accuracy during extreme weather conditions

C. Electricity price forecasting

The inception of electricity markets brought a new challenge to the industry: electricity price forecasting. While the load forecasts tell the utilities how much power supply they need to balance the demand, the price forecasts help them figure out how much energy they should buy or sell.

Three categories of price forecasting methods have been employed: simulation, statistical, and artificial intelligence.

Simulation methods require a mathematical model of the electricity market, load forecasts, outage information, and bids from the market participants.. Conducting a simulation takes very specialized skills and knowledge of power systems as well as sophisticated software packages Price forecasting accuracy is highly dependent upon the quality of the input information; the load forecasts at each node of the market is a driver of the electricity price forecast.

Statistical and AI based methods do not require the comprehensive knowledge of the market operations. They use historical prices, weather, outages, and loads to forecast future prices. Weron (2006) discussed applications of several statistical techniques for load and price forecasting. Their implementation cost is less than those of the simulation methods. However, they often find difficulties in forecasting *price spikes* as illustrated in Figure 4. These spikes are mostly due to congestions in the transmission network, which can sometimes be picked up through simulation methods.

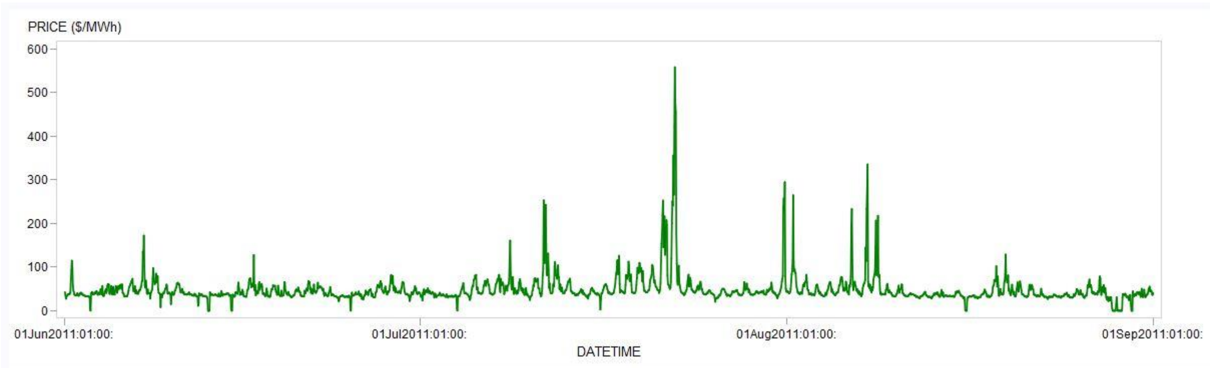


Fig. 4 Electricity is the most volatile commodity in the world. Price spikes are hard to predict.

Hong, T. (2014). Energy forecasting: past, present and future. *Foresight: The International Journal of Applied Forecasting*, (32), 43-48.

4. The smart grid era

A. Demand response forecasting

In the past decade, the electric power industry has undergone a grid-modernization process, installing millions of smart meters, sensors, and communication devices. Smart grid technologies bring great potential for a greener and more reliable grid at reduced cost. One way to achieve these goals is through demand response (DR), which is defined by US Federal Energy Regulatory Commission as "changes in electric usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized."

To effectively design and implement the DR programs, utilities have to perform a series of analytical tasks, such as forecasting electricity price, forecasting load at household level with or without DR programs, and after-the-fact estimation of the consumption pattern changes of due to various demand response programs. A major challenge in this area is to estimate the normal consumption pattern, or baseline. Because time is irreversible, there is no way for a utility to conduct an experiment to get the actual value of normal consumption pattern for an event day (the day when DR programs are triggered). There are over a hundred baseline estimation methods. Most of them are based on a simple average of the load profiles of several days prior to the date when DR programs are triggered. These methods are shown to be insufficient at household level. There is not yet a proven solution to this problem.

B. Renewable generation forecasting

With the generation from wind turbines, rooftop PV panels and solar farms, the century-old energy forecasting problem finds a new life. Volatility of renewable generation brings a big challenge to the system operators and energy traders. Figure 5 shows one week of solar generation at 5-min interval. While many large and medium utilities today operate their systems with the error of one day ahead load forecasts at 3% or lower, they can hardly achieve the similar accuracy in forecasts of their wind or solar generation.

Major advances are being made in renewable-generation forecasting by both meteorologists and forecasters. The meteorologists sample the state of the fluid at a given time and apply the equations of fluid dynamics and thermodynamics to estimate the state of the fluid at some time in the future. This numerical weather prediction approach requires significant computing resources as well as working knowledge of meteorological science. Moreover, further analysis is needed to translate wind and solar forecasts into renewable generation forecasts. The statistical alternative uses meteorological forecasts as one of its modeling inputs, among others, such as lagged wind power and calendar variables. A recent review of the state-of-the-art is by Giebel et al. (2011).

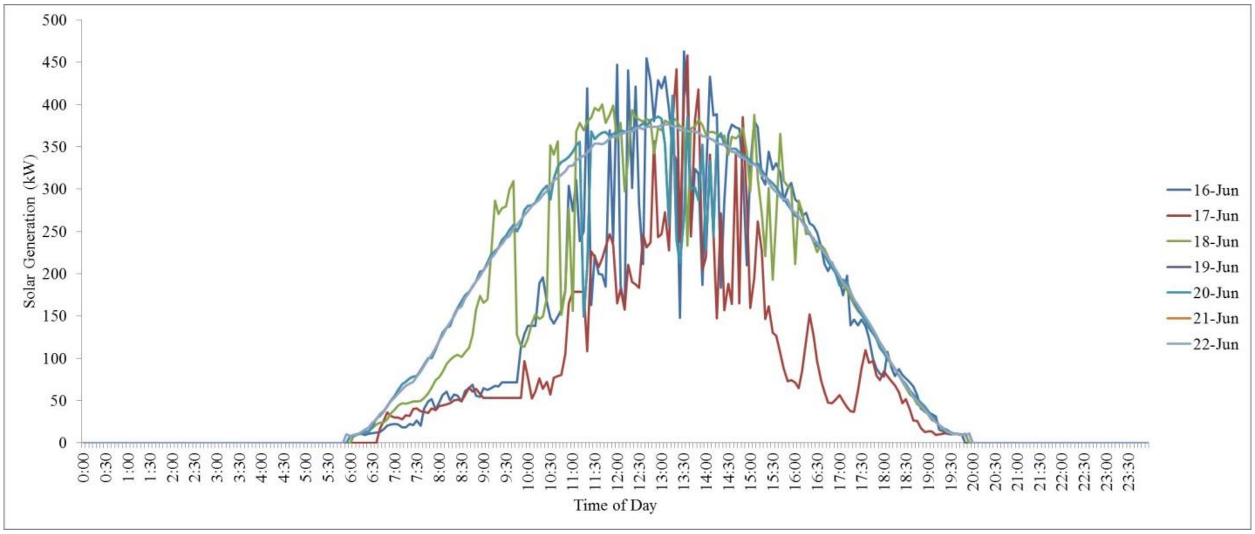


Fig. 5 One week of solar power generation at 5-minute interval. The volatility is depending upon the mood of the cloud. The generation profile is a nice bell curve if it is a sunny day. In a cloudy or partially cloudy day, the generation is quite unpredictable.

C. Global Energy Forecasting Competition

To tackle the emerging challenges in energy forecasting, IEEE Working Group on Energy Forecasting organized the Global Energy Forecasting Competition 2012 (GEFCom2012), which brings together many new ideas to the energy forecasting field from data scientists in many different industries. The competition consists of two tracks, hierarchical load forecasting and wind power forecasting. There are 2000+ entries submitted by 200+ teams. Finally 8 teams formulated by people from 8 countries are recognized as the winners of GEFCom2012. Multiple linear regression and gradient boosting machine show up among the top 4 winning entries of each track. A more comprehensive introduction of GEFCom2012 is by Hong, Pinson and Fan (2014).

Difficulties in accurately predicting price, renewable generation and long-term load have prompted interest in probabilistic forecasts. Probabilistic forecasts offer a more comprehensive description of the future and are popular for use in risk management. There are still many research challenges with probabilistic energy forecasting, such as visualization, interpretation, and evaluation of the results, incorporating probabilistic inputs, and integration with business decision making processes. With the success of GEFCom2012, the competition organizers will launch an upgraded version in 2014. The theme of GEFCom2014 is probabilistic energy forecasting. There will be four tracks covering load, price, wind and solar forecasting.

5. Lessons learned

There are many lessons learned as the field is being advanced over the past century, three major ones are the follows:

A. Out-of-sample tests

Many energy forecasting papers reported amazingly low errors but failed miserably in practice. A primary reason is lack of rigorous out-of-sample tests. For instance, an obvious mistake is to use a model with 1000+ parameters to fit a dataset with a few hundred observations. Most mistakes made in the literature are more difficult to identify. For instance, when developing a regression model, the authors slice the data to two pieces. One piece is for parameter estimation, the other piece as validation to calculate Mean Absolute Percentage Error (MAPE). After they see the MAPE value too high, they will change the model and re-calculate MAPE until the MAPE value is low enough to be impressive. When submitting the paper, they report the model with the lowest MAPE. Although this validate data was never used for model fit, its information is used for variable selection when building the model. In real life, we will never get a chance to use tomorrow's actual load to build the model. Peeking the future can produce nice results on a paper, but not in practice. Tashman (2006) offered a comprehensive review of out-of-sample tests, of which sliding simulation is the most appropriate and effective one for energy forecasting.

B. Understanding of business needs

There are many reasons that utilities do things the way they do today. Some can be improved, some cannot. For instance, the long term forecasting methodology used for rate case have to be published and archived. Many parties, such as managers in the utility, regulatory commission and shareholders (if the utility is investor-owned) may review the document even years after the case. Therefore, the long term load forecasting methodology has to be interpretable and transparent for ease of communication and defense. Many utilities are simply not allowed using black box models for long term forecasts for rate case. Despite how fancy the models are, they are not useful for the business.

C. It takes a village

In today's dynamic environment, virtually all types of energy forecasts are connected. A short term load forecasting model can be augmented to a long term model by adding macroeconomic indicators (Hong, Wilson and Xie, 2013). Electricity price is no longer driven by load only. The volatile renewable generation of wind and solar farms also affects the price significantly. Price signals trigger demand response programs, which in turn affect the loads. A best practice today for utilities is to build an in-house analytics center of excellence, where statisticians, data miners, meteorologists, business liaisons, IT analysts, and software developers can work together to tackle the emerging challenges of energy forecasting.

Energy forecasting is an interdisciplinary field. To further advance the knowledge, we have to involve various communities, such as statistical forecasting, artificial intelligence, meteorological science, and electrical engineering. To make the research meaningful to the industry, it is necessary to take inputs from utilities. It took a village to get to where we are today. And it will take a village to be where we want to be in the future.

References

Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., & Draxl, C. (2011). *The state-of-the-art in short-term prediction of wind power: a literature overview*. ANEMOS. plus.
Hong, T. (2008). *Spatial load forecasting using human machine co-construct intelligence framework*. North Carolina State University, Raleigh, NC, USA.
Hong, T. (2010). *Short term electric load forecasting*. North Carolina State University, Raleigh, NC, USA.
Hong, T., Pinson, P., & Fan, S. (2014). Global energy forecasting competition 2012. *International Journal of Forecasting*, *30*(2), 357-363.
Hong, T., Wilson, J., & Xie, J. (2014). Long term probabilistic load forecasting and normalization with hourly information. *Smart Grid, IEEE Transactions on*,*5*(1), 456-462.
Khotanzad, A., Afkhami-Rohani, R., & Maratukulam, D. (1998). ANNSTLF-artificial neural network short-term load forecaster generation three. *Power Systems, IEEE Transactions on*, *13*(4), 1413-1422.
Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, *16*(4), 437-450.
Weron, R. (2006). *Modeling and forecasting electricity loads and prices: A statistical approach* (Vol. 403). John Wiley & Sons.
Willis, H. L. (2002). *Spatial electric load forecasting*. CRC Press.

Biography

Dr. Tao Hong is the Director of BigDEAL (Big Data Energy Analytics Laboratory), NCEMC Faculty Fellow of Energy Analytics, Graduate Program Director and EPIC Assistant Professor of Systems Engineering and Engineering Management at University of North Carolina at Charlotte. He is the Founding Chair of IEEE Working Group on Energy Forecasting, General Chair of Global Energy Forecasting Competition, lead author of the online book Electric Load Forecasting: Fundamentals and Best Practices, and author of the blog Energy Forecasting. Dr. Hong received his B.Eng. in Automation from Tsinghua University in Beijing and his PhD with co-majors in Operations Research and Electrical Engineering from North Carolina State University.

HSC Research Report Series 2013

For a complete list please visit http://ideas.repec.org/s/wuu/wpaper.html

01 *Forecasting of daily electricity spot prices by incorporating intra-day relationships: Evidence form the UK power market* by Katarzyna Maciejowska and Rafał Weron
02 *Modeling and forecasting of the long-term seasonal component of the EEX and Nord Pool spot prices* by Jakub Nowotarski, Jakub Tomczyk and Rafał Weron
03 *A review of optimization methods for evaluation of placement of distributed generation into distribution networks* by Anna Kowalska-Pyzalska
04 *Diffusion of innovation within an agent-based model: Spinsons, independence and advertising* by Piotr Przybyła, Katarzyna Sznajd-Weron and Rafał Weron
05 *Going green: Agent-based modeling of the diffusion of dynamic electricity tariffs* by Anna Kowalska-Pyzalska, Katarzyna Maciejowska, Katarzyna Sznajd-Weron and Rafał Weron
06 *Relationship between spot and futures prices in electricity markets: Pitfalls of regression analysis* by Michał Zator
07 *An empirical comparison of alternate schemes for combining electricity spot price forecasts* by Jakub Nowotarski, Eran Raviv, Stefan Trueck and Rafał Weron
08 *Revisiting the relationship between spot and futures prices in the Nord Pool electricity market* by Rafał Weron and Michał Zator
09 *Rewiring the network. What helps an innovation to diffuse?* by Katarzyna Sznajd-Weron, Janusz Szwabiński, Rafał Weron and Tomasz Weron
10 *Going green: Agent-based modeling of the diffusion of dynamic electricity tariffs* by Anna Kowalska-Pyzalska, Katarzyna Maciejowska, Katarzyna Sznajd-Weron, Karol Suszczyński and Rafał Weron
11 *Forecasting of daily electricity prices with factor models: Utilizing intra-day and inter-zone relationships* by Katarzyna Maciejowska and Rafał Weron
12 *Computing electricity spot price prediction intervals using quantile regression and forecast averaging* by Jakub Nowotarski and Rafał Weron
13 *Long term probabilistic load forecasting and normalization with hourly information* by Tao Hong, Jason Wilson and Jingrui Xie
14 *Fuzzy interaction regression for short term load forecasting* by Tao Hong and Pu Wang
15 *Energy forecasting: Past, present and future* by Tao Hong