

Weather Variation and Climate Change Impacts on Power System

Po-Chen Chen, *Student Member, IEEE*, and Mladen Kezunovic, *Fellow, IEEE*

Department of Electrical and Computer Engineering, Texas A&M University
College Station, TX 77843-3128, U.S.A.
pchen01@tamu.edu and kezunov@ece.tamu.edu

Abstract—Erratic weather patterns are the vital challenge to the existing power system infrastructure. It is critical to understand how to mitigate the weather impacts on mid- and long-term planning efforts. Yet, it is not typical for utility industry to incorporate weather data to perform the analytics and correlation to determine the impact of weather variation and climate change. In this paper, we analyze maximum temperature as an example to demonstrate the weather variation and climate change impact on the summer peak load in Harris County, Texas, USA.

Index Terms—big data, data analysis, meteorology, power demand, power system planning.

I. INTRODUCTION

Almost all the disturbances to power system operation are the results of weather events. The analysis of weather and climate change impacts is one of the most complex tasks for variety of power system applications such as oscillations and stability detection [1], operation control and coordination [2], cyber and physical infrastructure security [3], cascading failures and blackouts [4], outage management and system restoration [5], and electricity market, demand response and generation forecasting [6].

To maintain a resilient power system, it is important to plan ahead and to prevent negative impacts on the grid. One of the most common weather impacts on power system is the temperature impact on the peak load. Number of papers studied such issues. Asbury [7] discusses the weather load modeling for demand response and forecasting. Suzara [8] utilizes historical monthly peak load data to specify a multiple regression model to predict the impact of temperature. Li *et al.* [9] study the coordination of thermostatically controlled loads (TCL) using a mathematical formulation. Qela *et al.* [10] present an approach to curtailment of peak load with fuzzy approach. Hassan *et al.* [11] suggest constant deviation plans (CDPs) and proportional deviation plans (PDPs) for customer engagement regarding peak load reduction. Nannery *et al.* [12] describe a different load management technique of having customers responding to a utility signal instead of time of day rates. Chen *et al.* [13] discuss about optimization of demand response scheduling

which is govern by the utility cost and probability distributions of exogenous information. Chen *et al.* [14] establish a systematic approach by considering the effects of temperature variation on load demand by using the typical load patterns of customer classes. Oliver *et al.* [15] provides a comprehensive summary of how the impact of climate change may impact the electricity infrastructure in Queensland. McSharry *et al.* [16] uses the probabilistic forecasts for assessing the uncertainty in the peak demand forecasting which provides better decision making and improves risk management. Erdinc *et al.* [17] implements an approach for HVAC demand considering consumer comfort violation minimization, allocation of comfort violation among consumers, and impact of humidity on ambient temperature.

The aforementioned literature survey has shown many techniques to manage the peak load issue from the utility perspective. Nevertheless, one of the most important aspects is missing: utilizing the weather data to continuously improve the decision-making process. Typically, weather application products from commercial vendor [18] are used by utilities. However, to assess the impact thoroughly, the additional weather data analytics and correlation are needed.

Another missing key aspect is the impact of climate change. It has been shown in lots of papers that there will be a dramatic climate change in the next couple decades [19]. From a long-term planning perspective, it is critical to plan ahead for a hardened grid infrastructure to face such challenges.

The focus of this paper is to utilize new weather data analytics to study the impact of historical temperature data from 1930 to illustrate the climate change impact. The data correlation utilizing the historical temperature and summer peak load data from 2008 to 2015 is employed to perform a weather variation impact study. In general, historical weather analytics and climate change studies are done in a relatively larger geographical scale (e.g. Southern Plains or a state of USA). Yet, for a given utility company, the study needs to be done within the target service territory (e.g. at the size of a city). Therefore, weather analytics done for a geographical area much larger than that may not be as useful. In this study, we choose Harris County as the geographical network under study. This

county is the most populous county in the State of Texas, and the third most populous county in the USA, based on the data from U.S. Census Bureau [20]. It may be noted that Harris County is a relatively small area comparing with the entire State of Texas.

The paper is organized as follows. Section II discusses the background. Section III demonstrates the climate impact on the peak temperature. Section IV evaluates the impact of temperature on peak load. Section V contains conclusions.

II. BACKGROUND

A. Weather Data Sources, Types, and Utilization

In general, there are three sources of weather measurements: weather station, (i.e. meteorological station or observation site), satellite, and radar. Weather stations worldwide routinely make the basic meteorological measurements. The standards for weather data collection framework may be found in [21], [22] including:

- Sensor: the group of available sensors, the specifications for detection sensor measuring range, accuracy, and resolution, the requirements regarding the sensor exposure, and placement and station instrumentations.
- Data exchange: the standard formats for weather data exchange among the automated weather information system, and the configuration of network data flow.

The satellite meteorological detection is a passive remote sensing, whereas the radar meteorological detection is an active remote sensing. Radars can emit radio or microwave waves and receive the back-scattering signals from a convective system.

The key issue for the utilities is how to select the most appropriated weather parameters among all possibilities. For specific type of weather events, the weather data analytics require the most applicable weather parameter as inputs. For a practical example regarding tropical cyclones, satellites can observe a tropical cyclone once it forms in the ocean and radar can detect its inner structure as it moves near the continent and eventually lands on.

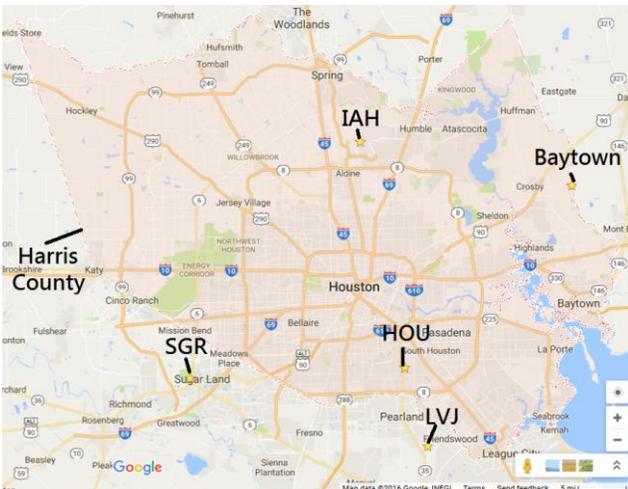


Fig. 1. 5 weather stations used in studies [22].

More details regarding weather data sources can be found in [23]. In the weather data collection and analysis framework, useful weather data may be the original raw data, interpolated data (raw data being processed by assimilation and mathematical model for numerical weather forecast), and weather application products (interpolated data being processed by statistical and meteorologist interpretation).

B. P-Value of Statistical Hypothesis Testing

In statistical hypothesis testing, Null Hypothesis (H_0) is defined as typically the ‘no difference’ or ‘no association’ hypothesis to be tested (usually by means of a significance test) against an Alternative Hypothesis (H_1) (i.e. Research Hypothesis) that postulates non-zero difference or association [24]. It should be noted that H_0 is not always the opposite of H_1 due to Type I and Type II errors [25]. Also, if there is a flaw within H_0 , the results could show H_0 is not supported, but neither H_1 nor the experiments become invalid. The reason why hypothesis testing is applied is to examine two statements H_0 and H_1 which are mutually exclusive, as demonstrated when we look at the impact of temperature data to climate change and peak load data.

R^2 statistic (or coefficient of determination) is defined as the square of the correlation coefficient between two variables [25]. It is to show how close the data fits the regression line, where the higher the value, the better the model fits the data.

F test is a test for the equality of the variances of two populations having normal distributions, based on the ratio of the variances of a sample of observations taken from each [25], which is known as F statistic. Under H_0 , F statistic is approximately to be 1 meaning that two populations are expected to be equal.

P value is defined as the probability of the observed data (or data showing a more extreme departure from H_0) when H_0 is true, and α stands for the significance level [25]. When linear regression is performed in our study, the confidence level of the slope is

$$(1 - \alpha) \times 100\% \quad (1)$$

If P value is smaller than α , then the slope is statistically significant. The value of α usually depends on the nature of hypothesis studies and traditionally it may be 0.1, 0.05, or 0.01 but can be some other values. In our study α is set to be 0.1 since the level and the factor we consider may not be easily predictable.

III. CLIMATE IMPACT ON PEAK TEMPERATURE

In this section, we will analysis the temperature data to demonstrate its impact to climate change. The term ‘‘annual hot days’’ is defined as the days in a year where the maximum temperature (T_{max}) is greater than a threshold. The null hypothesis H_0 is defined as the number of annual hot days does not increases as time goes on, where if the P value is smaller than α then the H_0 is rejected. α is set to be 0.1 so the confidence interval is at 90%.

The historical temperature data from Daily Summaries data of Climate Data Online (CDO) in National Centers for

TABLE I. STATISTICAL RESULTS OF DATA SETS A, B, AND C

Data Set	T_{max} (°F) Threshold	R^2 statistic	F Statistic	P Value	Estimate of MSE	Confidence Interval	Regression Line
A	95	0.024	2.030	0.158	201.032	[-0.015, 0.194]	$y = -160.256 + 0.089x$
	96	0.021	1.813	0.182	123.571	[-0.016, 0.148]	$y = -120.639 + 0.066x$
	97	0.020	1.688	0.197	66.021	[-0.013, 0.106]	$y = -86.358 + 0.047x$
	98	0.035	3.042	0.085	30.338	[0.002, 0.083]	$y = -80.527 + 0.042x$
	99	0.024	1.998	0.161	11.220	[-0.004, 0.046]	$y = -39.460 + 0.021x$
	100	0.036	3.123	0.081	4.156	[0.001, 0.031]	$y = -30.448 + 0.016x$
B	95	0.166	8.744	0.005	373.022	[0.274, 0.995]	$y = -1233.887 + 0.634x$
	96	0.138	7.033	0.011	306.285	[0.189, 0.842]	$y = -1006.205 + 0.515x$
	97	0.123	6.179	0.017	224.217	[0.134, 0.693]	$y = -809.643 + 0.413x$
	98	0.113	5.579	0.023	133.680	[0.088, 0.519]	$y = -595.628 + 0.303x$
	99	0.074	3.542	0.066	81.362	[0.020, 0.357]	$y = -370.318 + 0.189x$
	100	0.056	2.630	0.112	42.912	[-0.004, 0.240]	$y = -232.091 + 0.118x$
	101	0.044	2.032	0.161	21.412	[-0.013, 0.160]	$y = -144.068 + 0.073x$

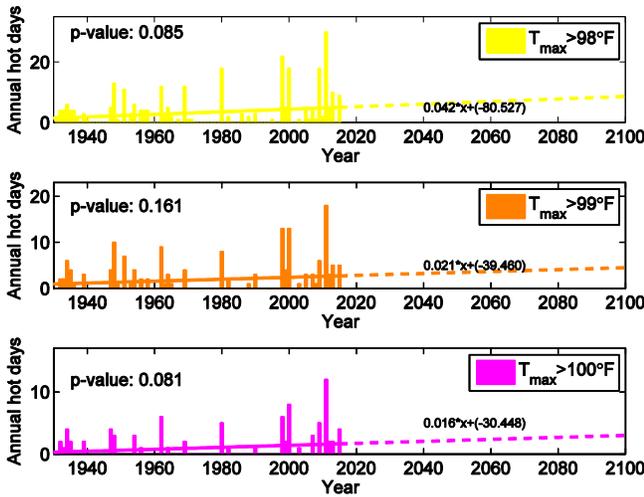


Fig. 2. Linear regression for Data Set A for T_{max} threshold 98°F to 100°F.

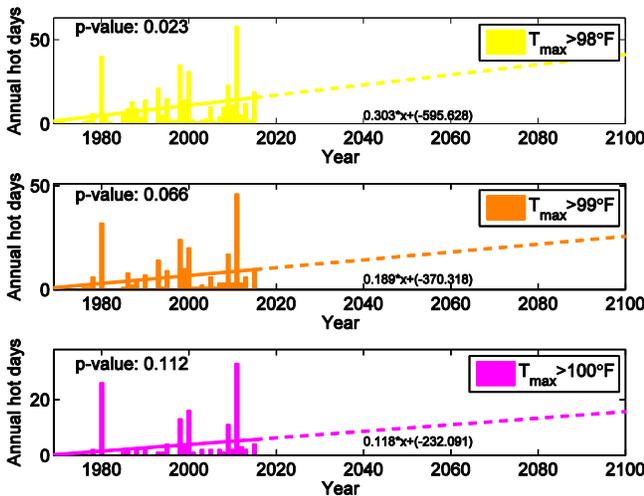


Fig. 3. Linear regression for Data Set B for T_{max} threshold 98°F to 100°F.

Environmental Information (NCEI) [26] is used. We look at two different sets of input data from two weather stations listed below. The geographical locations are shown in Fig. 1. The

reasons for the selection are due to the completeness of data and the geographical locations.

Data Set A: Houston William P. Hobby Airport (ID: GHCND:USW0012918) from Aug. 1st, 1930 to Oct. 9th, 2016.

Data Set B: Houston Intercontinental Airport (ID: GHCND:USW0012960) from Jun. 1st, 1969 to Oct. 9th, 2016.

Figs. 2 and 3 show the results of linear regression for Data Sets A and B, respectively when T_{max} is 98°F to 100°F. Table I provides the R^2 statistic, F statistic and its P -value, and an estimate of the estimate of mean square error (MSE) for Data Set A, B, and C when T_{max} is 95°F to 101°F, respectively. Note data points available are more in Data Set A than in Data Set B.

While looking at R^2 statistic in Table I, due to other anomalies in climate (e.g. El Nino and La Nina), it is expected that the values will be low. In Data Set A, the values are higher when T_{max} is higher, which is opposite in Data Set B. In Figs 2 and 3, some years have 0 hot days which is bad estimate samples for R^2 statistic.

While looking at F statistic, the values are higher when T_{max} is higher, which is opposite in Data Set B. In order to reject H_0 , F statistic need to be higher (i.e. the ratio is further away from 1).

While looking at confidence interval (CI), the negative lower confidence numbers suggest that the sample size may not be sufficiently large enough. In this case, they may be replaced with very small positive numbers for practical purposes.

The slope for all regress lines is positive. For the slope to be statistically significant within the given confidence interval at the 90% confidence level, the P value needs to be smaller than α . In Data Set A, it satisfies the condition when T_{max} values are 98°F and 100°F; in Data Set B, it satisfies the condition when T_{max} values are 95°F to 99°F. Note even though the R^2 statistic values are generally low, we still draw statistically significant predictors, which demonstrates how variations in the predicted response values could be associated with variations in

TABLE II. PEAK LOAD

Peak Load Date	Peak Load Consumption (GW)
7/21/2008	13.35719
6/25/2009	13.82766
8/23/2010	14.11831
8/18/2011	14.86684
6/26/2012	14.48539
8/13/2013	14.49347
8/25/2014	14.44517
8/11/2015	15.61568

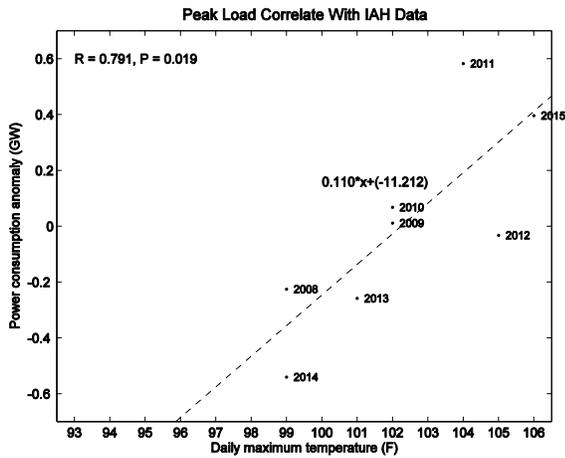


Fig. 4. Linear regression studies between T_{max} and peak load anomalies using T_{max} data from IAH.

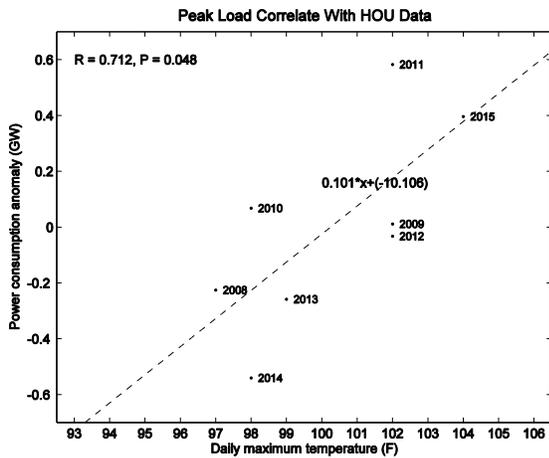


Fig. 5. Linear regression studies between T_{max} and peak load anomalies using T_{max} data from HOU.

the mean response values. In this case, we can still draw the conclusion, we might expect that increased hot days per year in the Harris County will be more likely to occur as climate continues to become warmer in the next a few decades.

IV. CORRELATION BETWEEN PEAK LOAD AND TEMPERATURE

In this section, we will analyze the correlation between the peak load data and the temperature. One geographical area of peak load consumption data at northeast of Harris County is chosen for the study. We will use temperature data from 5

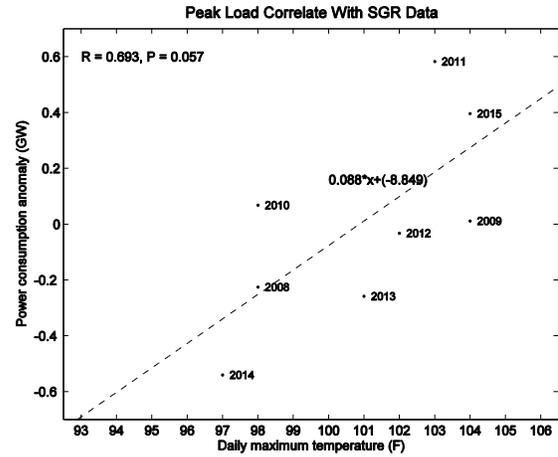


Fig. 6. Linear regression studies between T_{max} and peak load anomalies using T_{max} data from IAH.

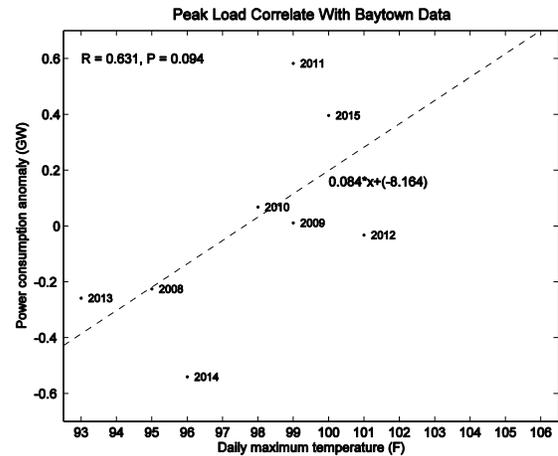


Fig. 7. Linear regression studies between T_{max} and peak load anomalies using T_{max} data from Baytown.

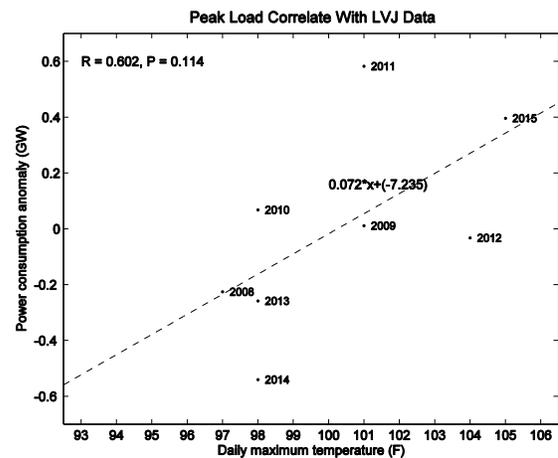


Fig. 8. Linear regression studies between T_{max} and peak load anomalies using T_{max} data from LVJ.

weather stations for the network under study: Houston William P. Hobby Airport (HOU), Sugar Land Regional Airport (SGR), Houston Intercontinental Airport (IAH), Baytown, and

Pearland Regional Airport (LVJ). Note the IAH is the only airport which is within the studied network and the other 3 are not, but they are around. HOU and IAH weather station details are given in Section III. The geographical locations are shown in Fig. 1. The IDs for SGR and Baytown are GHCND:USW00012977 and GHCND:USC00410586, respectively.

The peak load data is the day having the most power consumption of 2008 to 2015 as shown in Table II. For each data point, we obtain T_{\max} from all 4 weather stations and perform linear regression studies. Due to very small sample size here (only 8 data points) it may become inconsistent to attempt to predict the future trend of anomalies.

The peak load data is detrended to analyze if the anomalies in the increasing rate are correlated to the change of T_{\max} . By subtracting the data with the fit line, the fluctuations about the trend within the data points can be better analyzed.

Figs. 4, 5, 6, 7, and 8 show the analysis for the peak load consumption anomalies versus IAH, HOU, SGR, Baytown, and LVJ data respectively. α is set to be 0.1 so the confidence interval is at 90%. In each sub-figure, the y-axis is the detrended data of peak load consumption, and the pairwise linear correlation coefficient R and the P value are indicated. The P value corresponds to R , is the testing for null hypothesis H_0 (no correlation) against alternative hypothesis H_1 (correlated).

In Fig. 4, all the results show relatively high correlation between the peak load anomalies and T_{\max} . The null hypothesis is rejected from the given P values except the data of LVJ. The comparison of R and P shows that the IAH data has the highest correlation with the peak load data anomalies then the other 4 airports. The IAH data presents relatively higher R and relatively much lower P values comparing to other 4 sets of data. The explanation may be that IAH location is the only one within the studied geographical network but not the other 4.

V. CONCLUSIONS

This paper makes several contributions:

- We apply hypothesis tests for data analytics and data correlation utilizing the historical temperature and summer peak load data from 2008 to 2015 to illustrate a weather variation impact study.
- Even though the R^2 statistic values are generally low, we reveal statistically significant predictors indicating that an increase in the number of hot days per year in the Harris County will be more likely to occur as climate continues to become warmer in the next a few decades.
- The peak load data is shown to have correlation with the change in the maximum temperature, where the IAH data has strong correlation coefficient, and the reason may be that it is the only weather station within the geographical area within studied network.

VI. ACKNOWLEDGMENT

The authors would like to acknowledge CenterPoint Energy for providing the historical peak load data.

REFERENCES

- [1] V. Madani and D. Novosel, "Taming the Power Grid," *IEEE Spectrum: The Smarter Grid*, Jan. 2005.
- [2] J. Taft, *et al.*, "Ultra-Large-Scale Power System Control and Coordination Architecture," *Pacific Northwest National Laboratory (PNNL)*, Jun. 2014.
- [3] U. S. Department of Energy, "Cyber and Physical Security Chapter 3: Technology Assessments," *Quadrennial Technology Review 2015*.
- [4] North American Electric Reliability Council (NERC), "Technical Analysis of the August 14, 2003, Blackout: What Happened, Why, and What Did We Learn," Jul. 2004.
- [5] Executive Office of the President, "Economic Benefits of Increasing Electric Grid Resilience to Weather Outages," Aug. 2013.
- [6] P. Cappers, *et al.*, "Mass Market Demand Response and Variable Generation Integration Issues: A Scoping Study," Ernest Orlando Lawrence Berkeley National Laboratory (LBNL), Oct. 2011.
- [7] C. E. Asbury, "Weather load model for electric demand and energy forecasting," *IEEE Trans. Power Apparatus and Syst.*, vol. 94, no. 4, pp. 1111-1116, Jul. 1975.
- [8] H. Suzara, "Modeling the Impact of Temperature on Peak Electricity Demand in California," University of California Berkeley, May. 2008.
- [9] S. Li, *et al.*, "Market-Based Coordination of Thermostatically Controlled Loads—Part I: A Mechanism Design Formulation," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp 1170-1179, Mar. 2016.
- [10] B. Qela and H. T. Mouftah, "Peak Load Curtailment in a Smart Grid Via Fuzzy System Approach," *IEEE. Trans. Smart Grid*, vol. 5, no. 2, Mar. 2014.
- [11] N. U. Hassan, *et al.*, "Customer Engagement Plans for Peak Load Reduction in Residential Smart Grids," *IEEE. Trans. Smart Grid*, vol. 6, no. 6, Nov. 2015.
- [12] P. R. Nannery and M. S. Foltin, "The Application Of Peak Activated Rated A Technique To Load Management," *IEEE Trans. Power Apparatus and Syst.*, vol. PAS-104, no. 11, Nov. 1985.
- [13] W. Chen, *et al.*, "Optimal Scheduling of Demand Response Events for Electric Utilities," *IEEE. Trans. Smart Grid*, vol. 4, no. 4, Dec. 2013.
- [14] C. S. Chen, *et al.*, "Temperature Effect to Distribution System Load Profiles and Feeder Losses," *IEEE Trans. Power Syst.*, vol. 16, no. 4, Nov. 2001.
- [15] E. Oliver, *et al.*, "How is climate change likely to affect queensland electricity infrastructure into the future?" *IEEE/PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, Nov. 2015.
- [16] P.E. McSharry, S. Bouwman, and G. Bloemhof, "Probabilistic forecasts of the magnitude and timing of peak electricity demand," *IEEE. Trans. Power Syst.* vol. 20, no. 2, May. 2005.
- [17] O. Erdinc, *et al.*, "End-User Comfort Oriented Day-Ahead Planning for Responsive Residential HVAC Demand Aggregation Considering Weather Forecasts," *IEEE. Trans. Smart Grid*, in press.
- [18] Commercial weather vendor web sites serving the U.S. [Online] Available. <http://www.nws.noaa.gov/im/more.htm>
- [19] D. Acke, "Zeroing In: Lessons from the European Climate Foundation's Roadmap 2050 Project," *IEEE Power and Energy Magazine*, vol. 12, no. 5, Sep.-Oct. 2014.
- [20] U.S. Census Bureau. [Online] Available. <https://www.census.gov/>
- [21] Federal Meteorological Handbook No. 1 - Surface Weather Observations and Reports, FCM-H1-2005, NOAA, Sep. 2005.
- [22] Standard Formats for Weather Data Exchange Among Automated Weather Information Systems, FCM-S2-1994, NOAA, Nov. 1994.
- [23] P.-C. Chen and M. Kezunovic, "Fuzzy Logic Approach to Predictive Risk Analysis in Distribution Outage Management," *IEEE Trans. Smart Grid*, in press.
- [24] B. S. Everitt and A. Skrondal, *The Cambridge Dictionary of Statistics* 4th ed, Cambridge University Press, New York, 2010.
- [25] E. L. Lehmann and J. P. Romano, *Testing statistical hypotheses* 3rd ed, Springer-Verlag, New York, 2005.
- [26] Climate Data Online (CDO), National Centers for Environmental Information (NCEI). [Online] Available. <https://www.ncdc.noaa.gov/cdo-web/>