

HOW DOES RELATIVE HUMIDITY AFFECT ELECTRICITY DEMAND?

by

Ying Ruan Chen

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Engineering Management

Charlotte

2015

Approved by:

Dr. Tao Hong

Dr. Badrul Chowdhury

Dr. Yesim Sireli

©2015
Ying Ruan Chen
ALL RIGHTS RESERVED

ABSTRACT

YING RUAN CHEN. How does relative humidity affect electricity demand? (Under the direction of DR. TAO HONG)

Weather is a key driving factor of electricity demand. Among all the weather variables included in load forecasting models over the past half a century, temperature is the most commonly used one. Although humidity has also been discussed in the load forecasting literature, it is not as formally studied as temperature. In reality, a large portion of the electricity demand is caused by heating, ventilation, and air conditioning in order to meet people's comfort level, which is primarily determined by temperature and humidity. In this thesis, how relative humidity affects electricity demand will be investigated.

The case study is conducted at North Carolina Electric Membership Corporation, a large generation and transmission cooperative in the United States, for its system total load and the loads of three power supply areas. It is found that relative humidity plays a vital role in driving electricity demand during the warm months (June, July, August and September). This thesis proposes a systematic approach to include relative humidity variables in a regression analysis framework, resulting in the recommendation of a group of relative humidity variables. The proposed models with the recommended addition of relative humidity variables improve the forecast accuracy of Tao's Vanilla Model and its three derivatives in 24-hour ahead, one-week ahead, two-week ahead and one-year ahead ex post load forecasting settings. The improvement obtained from this case study ranges from 4.05% to 9.39% for NCEMC total ex post load forecasting on the test data (holdout sample).

ACKNOWLEDGMENTS

First of all, my sincere thanks are due to my advisor Dr. Tao Hong for consistent instruction and patient help in the process of this current research. His precious suggestions not only have been applied in this research, but also will be my motto in future career.

Second, I'd like to thank the other two committee members, Dr. Badrul Chowdhury and Dr. Yesim Sireli. Their constructive suggestions helped to greatly improve the presentation of this thesis.

Third, I also want to thank group members at the Big Data Energy Analytics Laboratory, Julian Huber, Bidong Liu, and Jingrui Xie, and coworkers at the North Carolina Electric Membership Corporation, Tom Laing, Laura White, and Margaret Yelverton. They provided generous help and support to my research.

Finally, I want to express my deepest love and appreciation to my family: my husband, Mr. Jinze Chen; my parents, Mr. Jinfu Ruan and Ms. Juanfen Ye; and my son, Joel Chen. This research won't be accomplished without their encouragement, tolerance, sacrifice and help.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF ABBREVIATIONS	ix
NOMENCLATURE	x
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: LITERATURE REVIEW	3
2.1 Recent Research in Load Forecasting	3
2.2 Temperature Variables	4
2.3 Other Weather Variables	7
2.4 More about Temperature and Relative Humidity	9
CHAPTER 3: THEORETICAL BACKGROUND AND METHODOLOGY	10
3.1 Theoretical Background	10
3.1.1 Linear Regression Models in Load Forecasting	10
3.1.2 Benchmark Process	10
3.2 Methodology	11
3.2.1 Error Measure	11
3.2.2 Cross Validation	12
3.2.3 Out-of-Sample Test	13
CHAPTER 4: CASE STUDY	14
4.1 Overview	14
4.2 Exploratory Data Analysis	16
4.3 Model Development	30

CHAPTER 5: RESULTS AND DISCUSSION	33
5.1 Cross Validation Results	33
5.2 Recommended Addition of Relative Humidity Variables	34
5.3 Out-of-Sample Test Results	35
5.4 Performance on Supply Area Level	38
CHAPTER 6: CONCLUSION	43
REFERENCES	45

LIST OF TABLES

TABLE 1: References of methods used in load forecasting	3
TABLE 2: Functional forms to represent the relationship between load and temperature	6
TABLE 3: Temperature variables in load forecasting	7
TABLE 4: Weather variables and indices in load forecasting	8
TABLE 5: Correlation coefficient between load and relative humidity by month in 2011	21
TABLE 5: Cross validation combinations	31
TABLE 6: MAPEs(%) of the models with additional relative humidity terms	33
TABLE 7: Tested model groups	36

LIST OF FIGURES

FIGURE 1: A typical scatter plot between hourly load and temperature in North Carolina (2009-2012)	5
FIGURE 2: Service areas of NCEMC and the locations of 27 weather stations	15
FIGURE 3: Time series plot of load at NCEMC from 2009 to 2012	17
FIGURE 4: Time series plot of relative humidity at NCEMC from 2009 to 2012	18
FIGURE 5: Load and relative humidity scatter plot (2011)	20
FIGURE 6: Load and relative humidity scatter plots by month (2011)	22
FIGURE 7: Load and relative humidity scatter plot in summer (2011)	24
FIGURE 8: Load and relative humidity scatter plots in summer by hour (2011)	25
FIGURE 9: Relative humidity and temperature scatter plot (2011)	27
FIGURE 10: Relative humidity and temperature scatter plots by month (2011)	28
FIGURE 11: Load and heat index scatter plot (2009-2012)	29
FIGURE 12: MAPEs of base models and proposed with added relative humidity terms	34
FIGURE 13: MAPEs of models tested with NCEMC total load	37
FIGURE 14: Summer and non-summer MAPEs of models tested with NCEMC total load using B_I	38
FIGURE 15: MAPEs of models tested with (a) SA1, (b) SA2, and (c) SA3 load, respectively	40
FIGURE 16: Summer and non-summer MAPEs of models tested with supply areas' load using B_I	42

LIST OF ABBREVIATIONS

AI	artificial intelligence
ANN	Artificial Neural Networks
CDD	cooling degree days
HDD	heating degree days
HI	heat index
LTLF	long term load forecasting
LOO	Leave-one-out
LPO	Leave-p-out
MAPE	Mean Absolute Percentage Error
MTLF	medium term load forecasting
NCEMC	North Carolina Electric Membership Corporation
NOAA	National Oceanic and Atmospheric Administration
NWS	National Weather Service
STLF	short term load forecasting
SVM	support vector machine
THI	temperature-humidity index
VFCV	V-fold cross-validation
VSTLF	very short term load forecasting
WCI	wind chill index

NOMENCLATURE

H_t	Hour, class variable, 24 hours of a day.
M_t	Month, class variable, 12 months of the year.
RH_t	Current hour relative humidity.
RHS_t	$RH_t \times S$, cross effect of relative humidity and summer.
RHS_t^2	$RH_t \times RH_t \times S$, cross effect of relative humidity and summer.
S_t	Summer, dummy variable. June, July, August and September is defined as summer.
T_a	Average temperature of the past 24 hours.
$Trend_t$	A linear trend variable.
T_t	Current hour temperature.
T_{t-k}	Temperature of the previous k-th hour.
W_t	Weekday, class variable, 7 days of a week.

CHAPTER 1: INTRODUCTION

The electricity grid has been labelled as the most complicated machine human beings have ever created (Schewe & Brennan, 2008). The lack of ability to massively store electricity leads to the requirement that the supply and demand must be balanced all the time. Thus, making accurate load forecasts is of great importance to power systems planning and operations (Hong, 2014).

Based on different forecasting horizons, the load forecasting can be categorized into four classes: very short term load forecasting (VSTLF), short term load forecasting (STLF), medium term load forecasting (MTLF), and long term load forecasting (LTLF) (Hong, 2010). The cut-off points are 1 day, 2 weeks and 3 years, respectively (Hong and Shahidehpour, 2015).

Over the past five decades, load forecasting has been extensively studied in the literature (Hong, 2014; Weron, 2006). Some of them focused on STLF (Hippert, Pedreira, & Souza, 2001; Hong, 2010), while others were on LTLF (Hong & Shahidehpour, 2015; Willis, 2002;)

Due to extensive use of electricity-powered equipment and appliances, human activities and weather patterns are the main driving factors of electricity demand (Hong, 2014). More specifically, part of the electricity consumption is used to keep the environment meeting people's comfort level, which is primarily determined by temperature and humidity. While temperature (dry bulb temperature) is the most deep-

rooted weather variable in the load forecasting literature, other weather variables, such as humidity, have not been formally studied as much as temperature.

The objective of this thesis is to investigate methods to add relative humidity information to temperature base linear regression models through a case study from North Carolina Electric Membership Corporation (NCEMC). First, the relationship between electricity demand and relative humidity will be explored through exploratory data analysis. Then, how to include humidity variables in the linear regression load forecasting models to improve their load forecasts will be investigated. The expected outcome is a linear regression benchmark model that includes both temperature and relative humidity variables. This would be an extension and continuation of the benchmarking effort taken in Hong (2010), where the author proposed a popular benchmark model (a.k.a. Tao's Vanilla Benchmark Model).

The structure of this thesis is as follows: Chapter 2 reviews the relevant literature; Chapter 3 presents the theoretical background and methodology used in this thesis; Chapter 4 introduces the case study, including the exploratory data analysis and the model development process; Chapter 5 presents the case study results and proposes the recommended addition of relative humidity variables; Chapter 6 concludes this thesis with the discussion of possible future research directions.

CHAPTER 2: LITERATURE REVIEW

2.1 Recent Research in Load Forecasting

Numerous techniques have been used for load forecasting. Most of them belong to two categories: statistical techniques and artificial intelligence (AI) techniques.

TABLE 1 is a summary of the representative techniques that have been used in load forecasting models together with the corresponding references.

TABLE 1: References of methods used in load forecasting

Category	Methods	References
Statistical techniques	Time series models	(Hagan & Behr, 1987); (Weron, 2006)
	Linear regression models	(Papalexopoulos & Hesterberg, 1990); (Hong, 2010)
	Semi-parametric additive models	(S. Fan & Hyndman, 2012); (Goude, Nedellec, & Kong, 2014)
AI techniques	Artificial Neural Networks (ANN)	(Khotanzad, Afkhami-Rohani, & Maratukulam, 1998); (Hippert et al., 2001);
	Fuzzy logic and fuzzy regression	(Ranaweera, Hubele, & Karady, 1996); (K.-B. Song, Baek, Hong, & Jang, 2005); (Hong & Wang, 2014)
	Support vector machine (SVM)	(Chen, Chang, & Lin, 2004); (S. Fan, Chen, & Lee, 2008)

In the last decade, the research in load forecasting has gone beyond trials of various techniques. In STLF at aggregated level, Hong (2010) proposed a systematic

method for variable selection. In hierarchical load forecasting, Fan, Methaprayoon, and Lee (2009) proposed a multiregion load forecasting method. The IEEE Working Group on Energy Forecasting organized the Global Energy Forecasting Competition 2012 (Hong, Pinson, & Fan, 2014) with one of the tracks on hierarchical electric load forecasting. In LTLF, Xie, Hong, and Stroud (2015) proposed a long term retail energy forecasting solution with consideration of customer attrition.

Comparing with conventional point (or single-valued) load forecasting, probabilistic load forecasting provides more information about the uncertainty of the future. Hyndman and Fan (2010) proposed a density forecasting method which provides probabilistic distributions for annual and weekly peak load forecast. This method has been deployed at Australia Energy Market in 2007. Hong, Wilson, and Xie (2014) proposed a long term probabilistic load forecasting and normalization method, which has been deployed in many utilities worldwide. Hong and Fan (2015) offered a tutorial review on probabilistic load forecasting.

Since this thesis is devoted to the investigation of humidity in load forecasting, the usage of weather variables in the coming sections will be reviewed. Section 2.2 is dedicated to temperature variables. Section 2.3 covers other weather variables. Section 2.4 provides a deep-dive into the temperature and humidity variables.

2.2 Temperature Variables

Various weather variables and weather related indices have been used in load forecasting. Temperature is the most popular one. In summer, load increases as temperature increases, in response to cooling needs. In winter, load increases as

temperature decreases to meet heating needs. This relationship between load and temperature results in a “hockey stick” shape as shown in FIGURE 1.

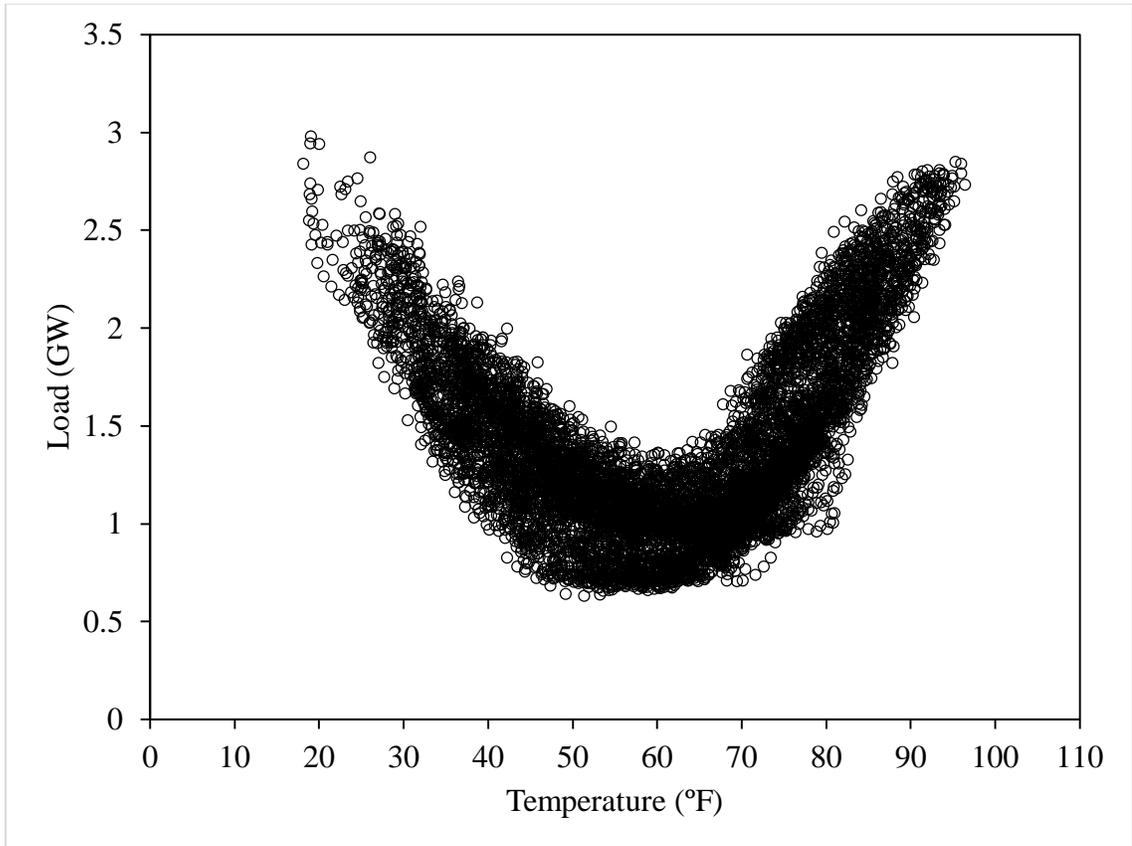


FIGURE 1: A typical scatter plot between hourly load and temperature in North Carolina (2011)

There are many ways to represent the aforementioned relationship in load forecasting models, such as piecewise linear regression models and second order polynomials. TABLE 2 below shows several functional forms that have been used in load forecasting models together with the corresponding references.

TABLE 2: Functional forms to represent the relationship between load and temperature

Functional Forms	References
Piecewise linear regression models	(S. Fan et al., 2008); (S. Fan et al., 2009)
Second order polynomials	(Abou-Hussien, Kandlil, Tantawy, & Farghal, 1981); (Y.-H. Song & Wang, 2003)
Third order polynomials	(Hagan & Behr, 1987); (Hong, 2010)

Furthermore, there are many other ways to include temperature information, such as current hour temperature, previous hour temperature, and maximum (or minimum) temperature. TABLE 3 summarizes several popular temperature variables that have been used in load forecasting models together with the corresponding references.

TABLE 3: Temperature variables in load forecasting

Temperature Forms	References
Current hour temperature	(S. Fan et al., 2009); (Hong, 2010)
Previous hour temperature	(S. Fan et al., 2009); (Hong, 2010)
Daily maximum/minimum temperature	(Park, El-Sharkawi, Marks, Atlas, & Damborg, 1991); (Papalexopoulos, Hao, & Peng, 1994); (Kiartzis, Bakirtzis, & Petridis, 1995)
Average temperature during a defined past time period	(Douglas, Breipohl, Lee, & Adapa, 1998); (Hong, Liu, & Wang, 2015)
Cooling/heating degree days (CDD/HDD)	(Papalexopoulos et al., 1994); (Hor, Watson, & Majithia, 2005)
Dew point temperature, wet bulb temperature	(Saifur Rahman, 1990); (Saini, 2008); (Raza & Khosravi, 2015)

2.3 Other Weather Variables

Other weather variables and indices have also been used by the researchers and practitioners, such as relative humidity, wind speed, heat index (HI) or temperature-humidity index (THI), and wind chill index (WCI). TABLE 4 lists a few frequently used ones in load forecasting models together with the corresponding references.

TABLE 4: Weather variables and indices in load forecasting

Weather variables or weather indices	References
Relative humidity	(Saifur Rahman, 1990); (J. Y. Fan & Mcdonald, 1994); (Mirasgedis et al., 2006)
Wind speed	(J. Y. Fan & Mcdonald, 1994); (McSharry, Bouwman, & Bloemhof, 2005); (PJM, 2015)
HI/ THI	(Saifur Rahman, 1990); (S. Rahman & Hazim, 1996)
WCI	(S. Rahman & Hazim, 1996); (PJM, 2015)

Both HI and WCI can be interpreted as adjusted temperatures. National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS) uses HI to represent the human-perceived equivalent temperature in warm weather (Rothfusz, 1990; Steadman, 1979). The HI equation is as follows. It includes both temperature and relative humidity variables.

$$\begin{aligned}
 HI = & -42.379 + 2.04901523 \times T + 10.14333127 \times RH - 0.22475541 \times T \times RH - \\
 & 6.83783 \times 10^{-3} \times T^2 - 5.481717 \times 10^{-2} \times RH^2 + 1.22874 \times 10^{-3} \times T^2 \times RH + \\
 & 8.5282 \times 10^{-4} \times T \times RH^2 - 1.99 \times 10^{-6} \times T^2 \times RH^2
 \end{aligned} \tag{1}$$

where T stands for temperature, RH stands for relative humidity. The pre-defined coefficients calculated based on multiple regression analysis. There will be some further adjustments when relative humidity is low but temperature is high or when relative humidity is high but temperature is low.

During the cold weather, when wind increases heat loss, people feel colder (Bluestein, 2015). The WCI by NOAA NWS measures the human-perceived equivalent

temperature in cold weather. Both temperature and wind speed are accounted in its calculation.

2.4 More about Temperature and Relative Humidity

Recall that electricity demand is mainly driven by people's comfortable level. Human beings may feel comfortable when temperature is high and relative humidity is low, or when temperature is low and relative humidity is high. Because of the combined impact of relative humidity and temperature on human comfort level, it is reasonable to consider both of them in load forecasting.

Relative humidity appeared in a few load forecasting papers. Saifur Rahman (1990) considered the effect of relative humidity for late spring, summer and early autumn in load forecasting. THI is used to replace temperature in his model when the forecasted day was in April to September and the temperature was between 76°F and 91°F. Hor et al. (2005) found that including relative humidity in the model can improve monthly load forecast accuracy during the summer months in UK.

Nevertheless, the load forecasting literature on relative humidity represents only a small fraction of what has been done on temperature. Since relative humidity has not yet received the attention it deserves in the load forecasting area, I would like to start a formal and systematic investigation on it in this thesis.

CHAPTER 3: THEORETICAL BACKGROUND AND METHODOLOGY

3.1 Theoretical Background

3.1.1 Linear Regression Models in Load Forecasting

Linear regression is one of the most widely deployed load forecasting methods. Independent variables (weather, calendar variables, etc.) are fed to the models to predict the dependent variable (load). The equation of a multiple linear regression with p independent variables X_1, \dots, X_p is (Kutner, Nachtsheim, Neter, & Li, 2004):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{i,p} + \varepsilon_i \quad (2)$$

The prediction is made based on parameter estimated (β_0 to β_p) from the historical data. A linear model can include both main effects (independent variables and their polynomials) and cross effects (interactions between the independent variables). Hong (2010) proposed a linear regression based approach to STLF, where the linear models can be augmented for VSTLF, MTLF and LTLF. Hong's methodology has been used by many utilities (Hong, Wilson, et al., 2014). In this thesis, investigation based on linear models as the continuation of the work done in (Hong, 2010) are conducted.

3.1.2 Benchmark Process

Benchmarking is an essential component of load forecasting. It helps to set the standard of modeling process, evaluate the improvement of new models, and make comparison within or between utilities. Hong (2010) pointed out that a good benchmark model should be simple, widely applicable, reproducible, and accurate. The

characteristics of linear regression make it an ideal technique for creating a benchmark model. Following these criteria, Hong (2010) proposed a multiple linear regression benchmark model with temperature variables for STLF (a.k.a. Tao's Vanilla Benchmark Model):

$$y_t = \beta_0 + \beta_1 Trend_t + \beta_2 M_t + \beta_3 W_t + \beta_4 H_t + \beta_5 W_t H_t + f(T_t) \quad (3)$$

where, \hat{y}_t is the forecasted load, $Trend$ stands for a linear trend, M_t , W_t , and H_t are class variables month, weekday, and hour, T_t is current hour temperature, β_i are the coefficients estimated using historical data, and

$$f(T_t) = \alpha_1 T_t + \alpha_2 T_t^2 + \alpha_3 T_t^3 + \alpha_4 T_t M_t + \alpha_5 T_t^2 M_t + \alpha_6 T_t^3 M_t + \alpha_7 T_t H_t + \alpha_8 T_t^2 H_t + \alpha_9 T_t^3 H_t \quad ($$

The above model was used as the benchmark model in the Global Energy Forecasting Competition 2012 (Hong, Pinson, et al., 2014). Among over 100 teams in the load forecasting track of the competition, this model finally ranked top 25%. This thesis focuses on improving this model by adding relative humidity variables.

3.2 Methodology

3.2.1 Error Measure

In this thesis, the Mean Absolute Percentage Error (MAPE) is used for model evaluation. It is one of the most popular error statistics in business forecasting including load forecasting. The definition of the MAPE is as follows:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (5)$$

where A_t stands for actual value (observation). F_t stands for forecasted value. The MAPE measures the size of the prediction's error in percentage terms compared with the observation. The smaller the MAPE value is, the more accurate the model is.

This thesis focuses on the model accuracy evaluation, which the utilities are most interested in. Other aspects such as autocorrelation of the residuals, multicollinearity of the dependent variables are not tested.

3.2.2 Cross Validation

The data is split into training data and validation data when building and selecting models. To overcome the potential overfitting issues, cross-validation is used (Arlot & Celisse, 2010). The basic idea of cross validation is to split the data into several pieces and use some pieces of data to predict the others. The simple average performance of all the combinations is then used for model validation.

There are several popular cross validation methods, such as Leave-one-out (LOO), Leave-p-out (LPO), and V-fold cross-validation (VFCV). For a dataset having N data points, LOO stands for using $(N-1)$ data points as training data for parameter estimation, and the one left data point as validation data for variable selection. Because every data point has to be used as validation data, the process above will be repeated a total of N times. The average performance of these N combinations are used for model validation.

Similarly, LPO corresponds to using $(N-p)$ data points as training data, and the p left data points as validation data. There are $\binom{N}{p}$ possible validation datasets. Thus, the parameter estimation and variable selection process needs to be repeated $\binom{N}{p}$ times. When the sample size N is very large, the computational time is long using these two methods.

On the other hand, VFCV is performed by dividing the data into V subsets of approximately equal size (N/V) , and use $(V-1)$ subsets as training data, the one left subset

as validation data. There are V possible validation datasets. Thus, the parameter estimation and variable selection process will be repeated only V times. Obviously, VFCV has higher computational efficiency compared with the other two methods. Thus, VFCV is used in this case study.

3.2.3 Out-of-Sample Test

Out-of-sample test is used in this thesis to test model accuracy (Tashman, 2000). Calibration windows of four lengths (1-day, 1-week, 2-week, or 1-year) are tested within a rolling scheme. For example, 1-day ahead ex post forecast of year 2012 is compared using the previous two years (2010-2011) data for parameter estimation. The moving window is two years of data. Each time, the calibration window is moved forward one day at a time. Model parameters are re-estimated every day in 2012. By doing so, the model stability over time for various forecasting horizons can be evaluated.

CHAPTER 4: CASE STUDY

4.1 Overview

NCEMC is one of the largest generation and transmission cooperatives in United States. It supports 26 cooperatives to provide service to more than 950,000 customers in North Carolina. Its service area covers 93 of North Carolina's 100 counties.

The data used in the case study is 4 years of hourly load from NCEMC from 2009 to 2012. The weather data is hourly temperature and relative humidity from 27 weather stations in North Carolina for the same four years. The weather data and NCEMC total load data are clean data. The service areas of NCEMC and locations of 27 weather stations are shown in FIGURE 2. The colored areas are the service areas, and the black stars are the locations of the weather stations. Note that one of the stations (KORF) is located in the State of Virginia.

The simple averages of temperature and relative humidity at the 27 weather stations are calculated and used to represent the weather conditions near the NCEMC service territory. To avoid distraction from the scope of this research, the advanced weather station selection methodology proposed by Hong, Wang, and White (2015) is not adopted. Nevertheless, it is believed that weather station selection with humidity input is a promising future research direction of this work.

4.2 Exploratory Data Analysis

FIGURE 3 shows the time series plot of hourly load from 2009 to 2012. The first three years of data (2009-2011) are used for parameter estimation (a.k.a., training) and variable selection (a.k.a., validation). The fourth year data (2012) is used as the hold-out sample for accuracy confirmation (a.k.a., test). The typical seasonal patterns of the load profile can be found in this time series plot. For instance, the load level of winter and summer is higher than that of spring and fall. In NCEMC, the summer peaks are usually higher than winter peaks. During some extremely cold years, however, the annual peak load may occur in the winter.

FIGURE 4 shows the time series plot of relative humidity for the same location during the same period as in FIGURE 3. During the winter months, relative humidity varies within a wider range from 20% to 100%. The variation in the summer months is relatively narrower, mostly from 50% to 100%. This is largely due to the fact that higher temperatures in the summer turn more water into water vapor in the air.

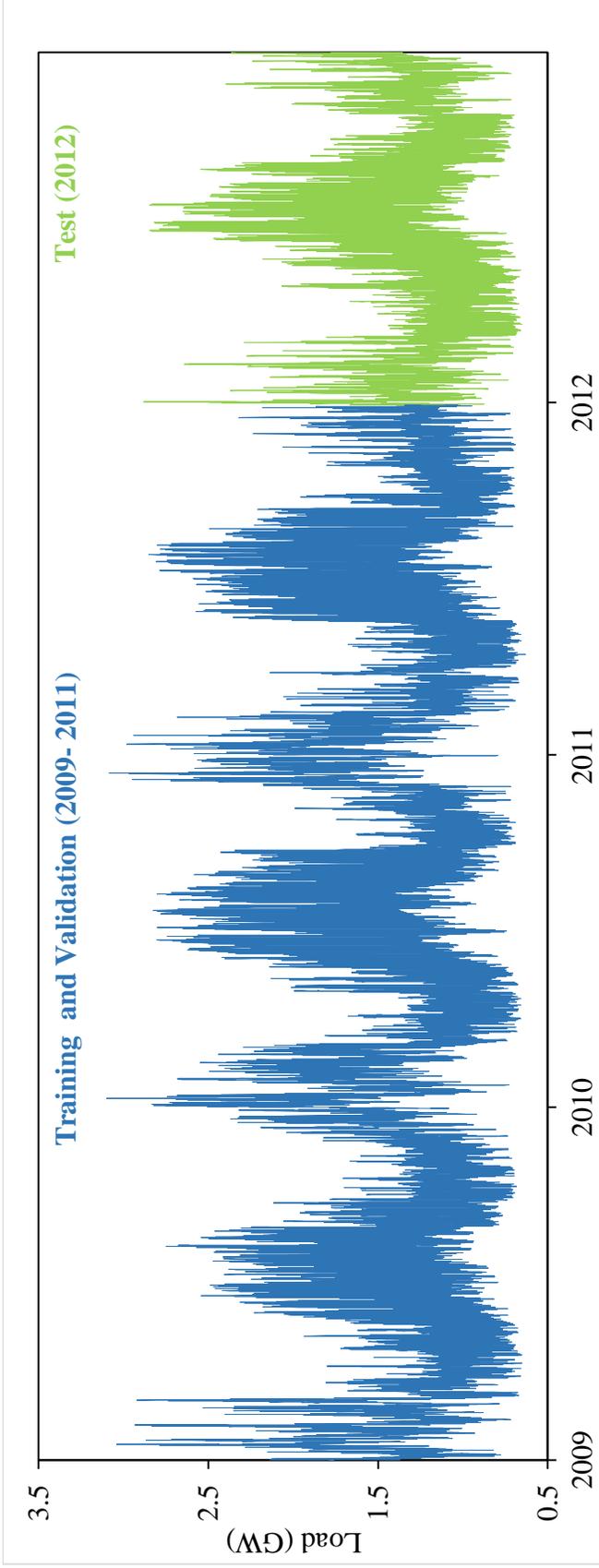


FIGURE 3: Time series plot of load at NCEMC from 2009 to 2012

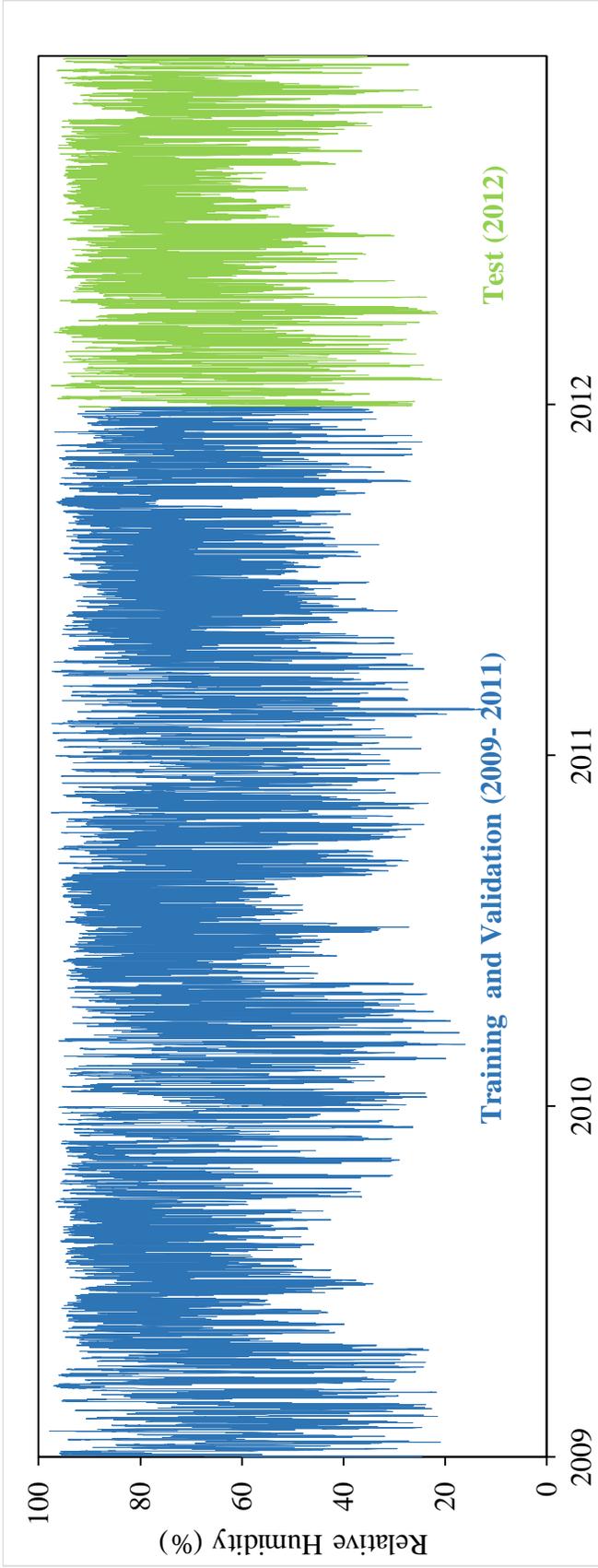


FIGURE 4: Time series plot of relative humidity at NCEMC from 2009 to 2012

FIGURE 5 shows the scatter plot of hourly load and relative humidity in 2011. The correlation coefficient between the two is -0.25. Similarly, the trend line suggests load and relative humidity are negatively related to each other. On the other hand, it is well known that electricity demand has seasonal, monthly, weekly, and diurnal cycles (Hong, 2010). These are caused by diverse human activity patterns during different time periods. In summer, for instance, energy is needed for cooling, in the winter for heating. Besides, people tend to have different energy consumption patterns at weekdays compared to weekends. In addition, the demand is higher during the daytime when people are awake than during the nighttime when people are asleep. Therefore, group analysis by calendar variables is performed to further investigate the relationship between load and relative humidity.

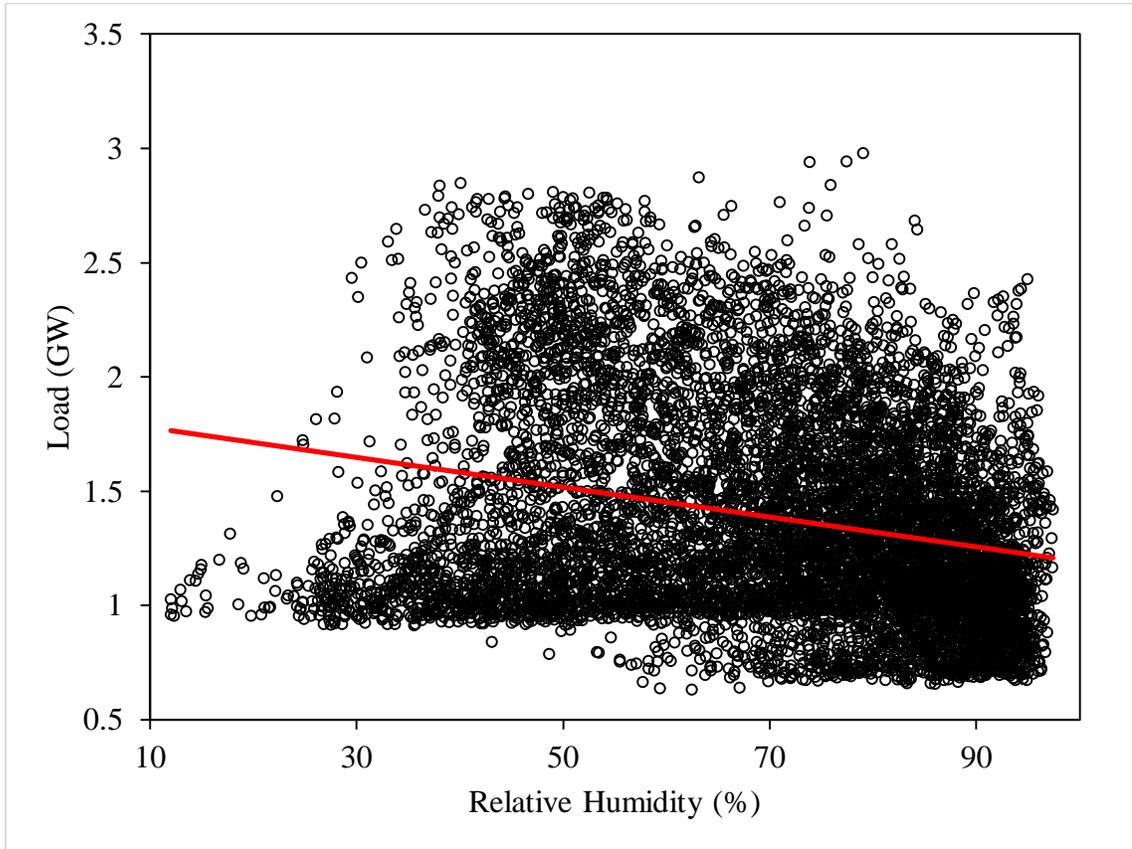


FIGURE 5: Load and relative humidity scatter plot (2011)

FIGURE 6 shows the scatter plots of load and relative humidity by month in 2011. Trend lines are also added to demonstrate the relationship between the two by month. The correlation coefficients between load and relative humidity by month in 2011 are given in TABLE 5. It is observed that the correlation in June to September is stronger than that in other months as highlight with bold in TABLE 5. During these four months, load tends to decrease when relative humidity increases. In the other eight months, the relationship between the two is weak (February, March, April, and May) or uncorrelated (January, October, November and December).

TABLE 5: Correlation coefficient between load and relative humidity by month in 2011

Month	Correlation Coefficient
1	-0.01
2	0.30
3	0.27
4	-0.27
5	-0.59
6	-0.78
7	-0.79
8	-0.76
9	-0.71
10	-0.15
11	-0.04
12	0.11

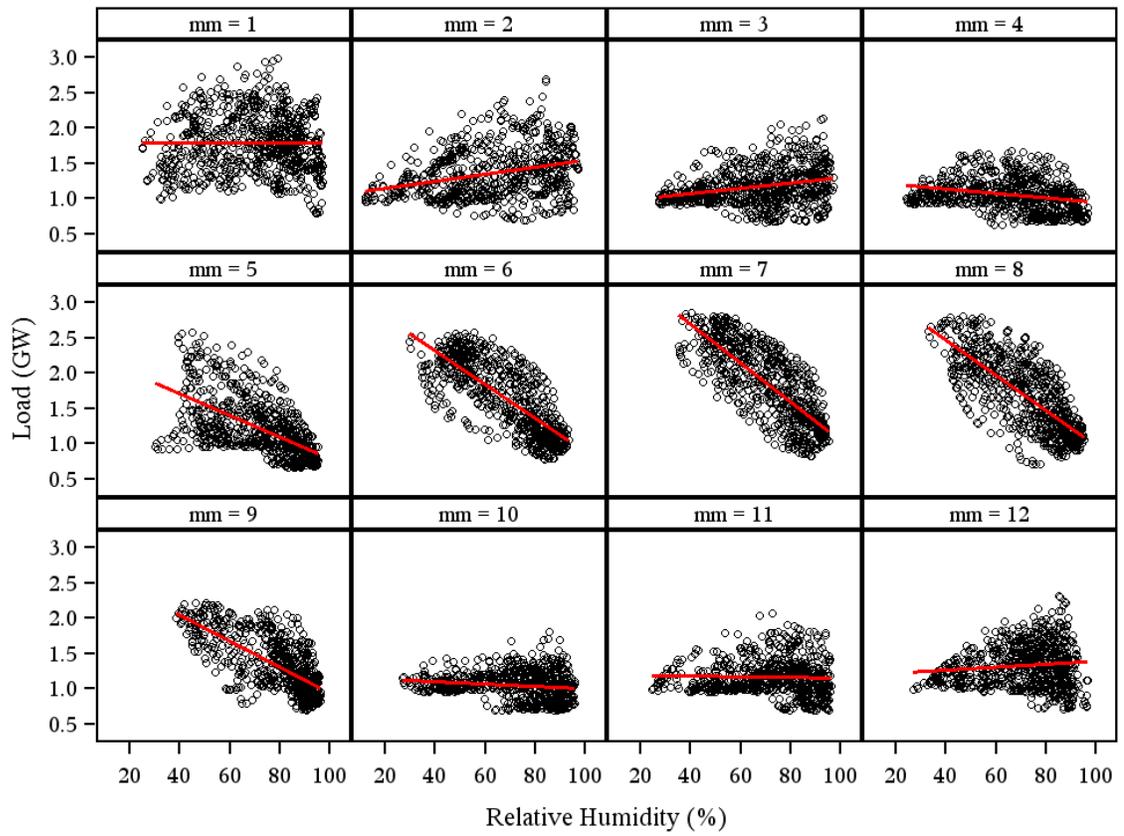


FIGURE 6: Load and relative humidity scatter plots by month (2011)

Recall that the human comfort level is determined by both humidity and temperature. The scatter plots by month indicate that the influence of relative humidity on load is stronger during warmer months (June, July, August and September). In this case study, June, July, August and September is named as summer months. FIGURE 7 shows the scatter plots of load and relative humidity in the summer of 2011, which present a sharper pattern than the one in FIGURE 5. The correlation coefficient between the two variables are -0.76 in summer. Overall, load decreases as relative humidity increases over the summer.

FIGURE 8 shows the scatter plots of load and relative humidity by hour in the summer of 2011. It is noted that the relationship between the two variables varies at different hours. During nighttime hours (Hours 1 to 9), load has a tendency to be low when the relative humidity is high. The correlation coefficients between the two during these nighttime hours are larger than -0.20 . These suggest weak negative correlation. During the daytime (Hours 10 to 24), load increases when relative humidity decreases. The correlation coefficients between the two are between -0.60 to -0.20 . These stand for stronger negative correlation. Therefore, there is hourly difference on the interactions between load and relative humidity.

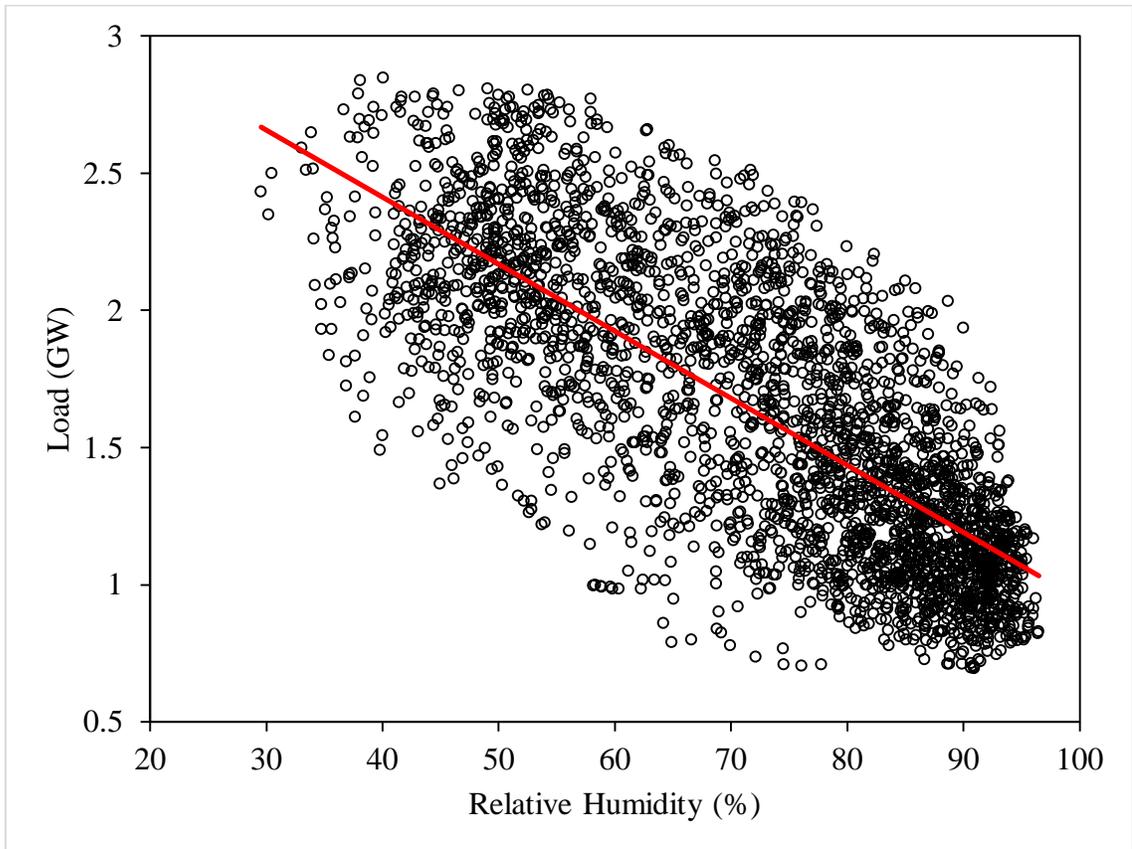


FIGURE 7: Load and relative humidity scatter plot in summer (2011)

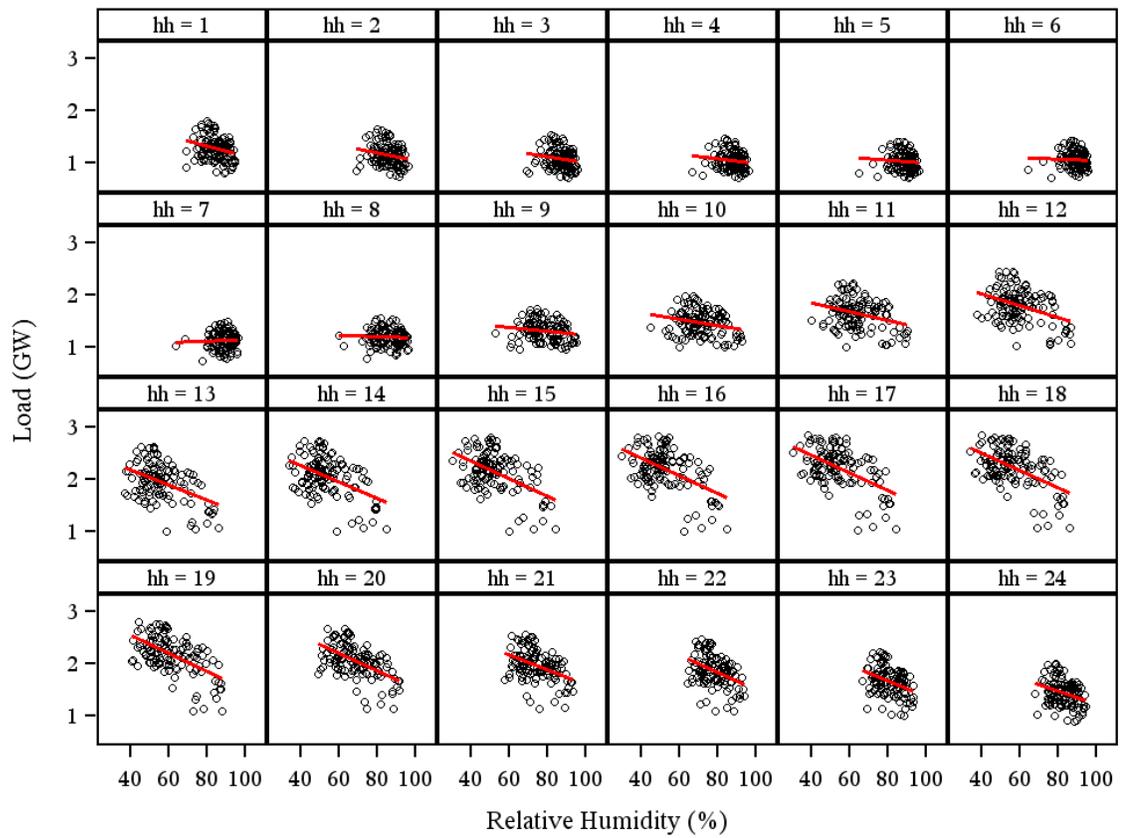


FIGURE 8: Load and relative humidity scatter plots in summer by hour (2011)

FIGURE 9 shows the scatter plot of relative humidity and temperature using data of 2011. Their correlation coefficient is -0.19. It is also noted from the scatter plot that relative humidity decreases when temperature increases. FIGURE 10 provides the scatter plots of relative humidity and temperature by month in 2011. It is noticed that they are strongly related in summer months (June to September) that relative humidity decreases quickly when temperature increases. The correlation coefficients between the two during these four months are -0.83, -0.83, -0.83 and -0.68, respectively. Those of the rest months are mostly range from -0.54 to -0.02. As implied in the time series plot of relative humidity in FIGURE 4, warmer temperature increases the atmosphere's ability to hold water vapor. Thus, the variation of relative humidity is narrower in summer. This results in a stronger relationship between temperature and relative humidity. Besides, FIGURE 1 shows that load increases when temperature increases in summer. All these factors lead to load increases when relative humidity decreases in summer as shown in FIGURE 7.

FIGURE 11 shows the scatter plot of load and HI using 2011 data. It is observed the strong similarity between this scatter plot and the load-temperature scatter plot shown in FIGURE 1. Both of them show the typical "hockey stick" shape. On the left arm, load increases when temperature decreases in winter because of heating. On the right arm, load increases when temperature increases in the summer for cooling. Evidently, the left arms of the "hockey stick" shape in the two figures are identical. The two arms are more separate in FIGURE 11 than they are in FIGURE 1. This is because HI is an adjusted temperature. When the dry-bulb temperature is high, the influence of relative humidity makes HI even higher.

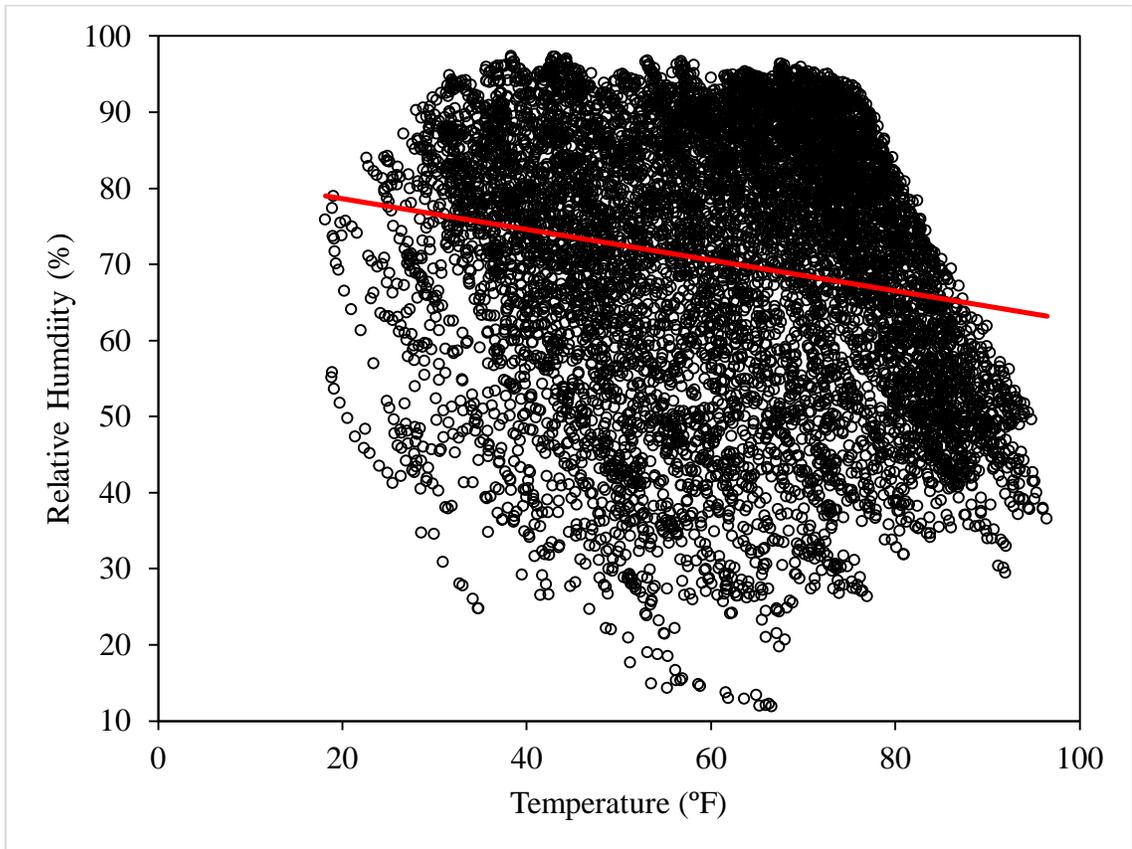


FIGURE 9: Relative humidity and temperature scatter plot (2011)

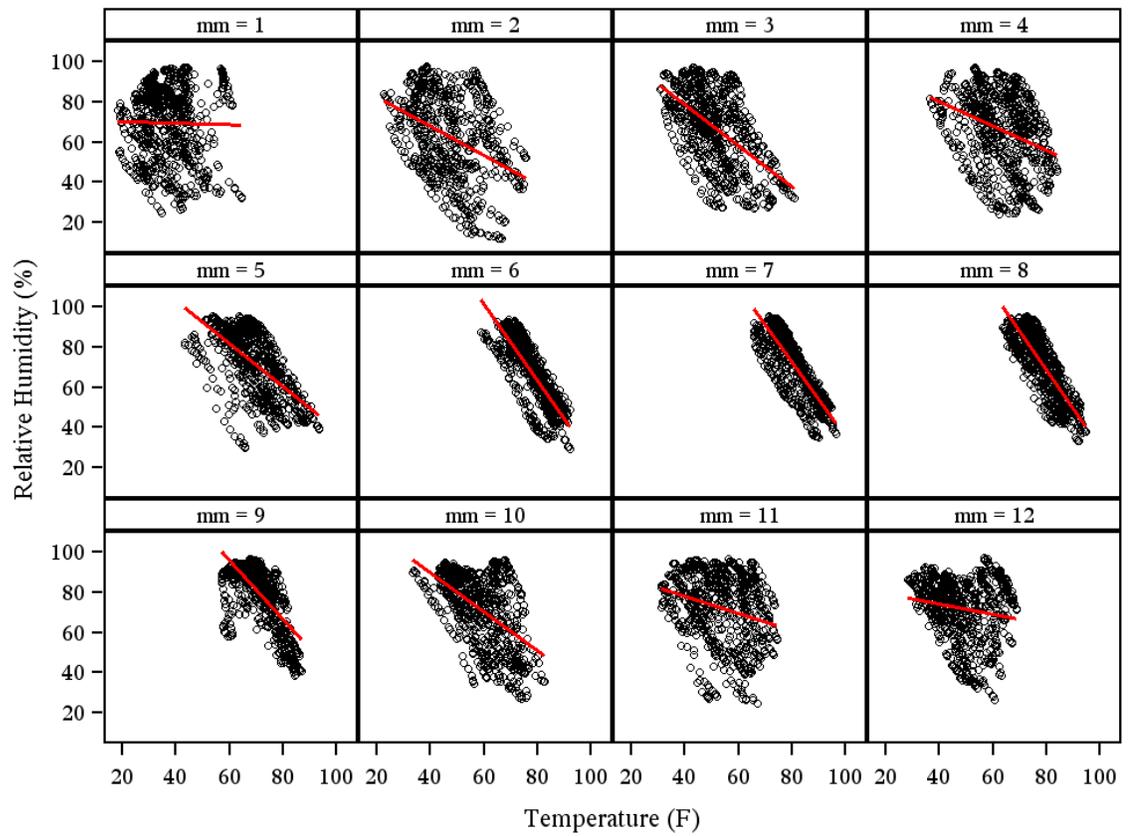


FIGURE 10: Relative humidity and temperature scatter plots by month (2011)

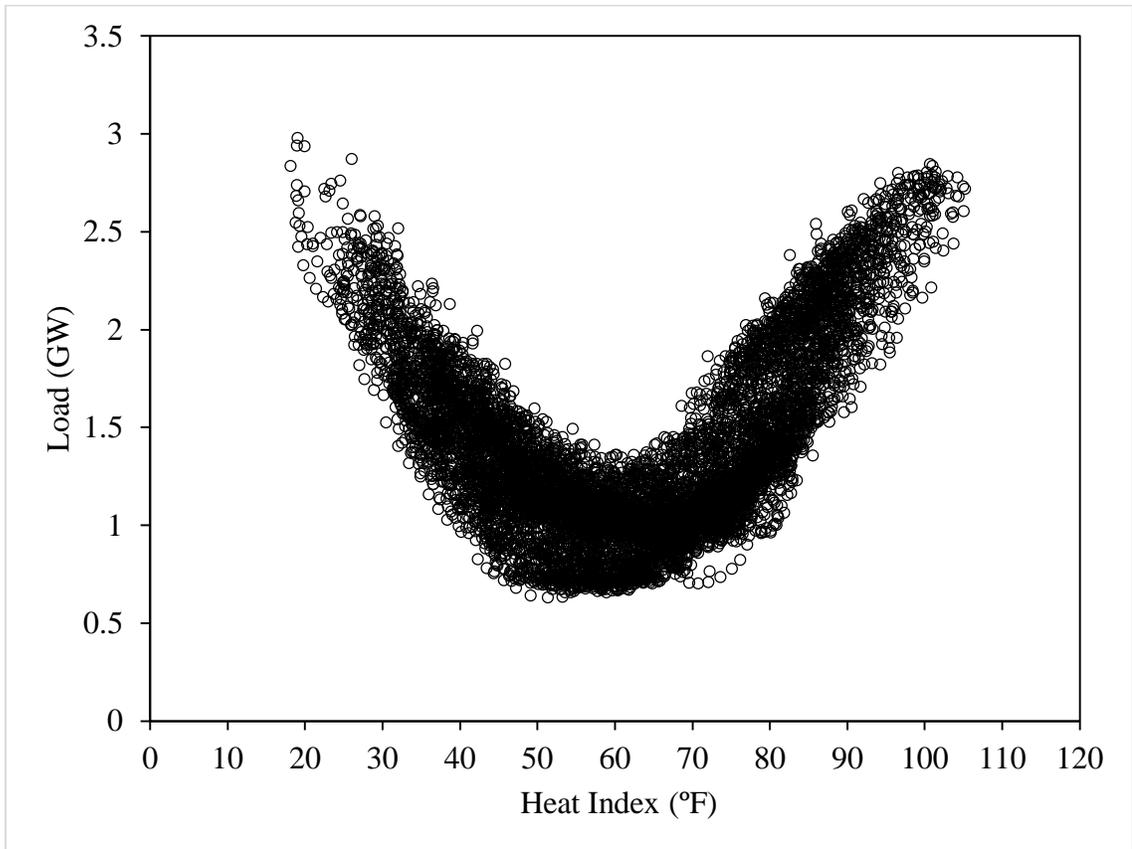


FIGURE 11: Load and heat index scatter plot (2011)

To sum up, scatter plots of load and relative humidity by months given in FIGURE 6 suggest there is a relationship between load and relative humidity during June, July, August and September. In the rest of this thesis, these four months are defined as summer months. The influence of relative humidity during summer will be investigated. The scatter plots of load and relative humidity by hour shown in FIGURE 8 depict the diurnal feature of interactions between load and relative humidity. Scatter plots of load and temperature in FIGURE 1 and load and HI in FIGURE 11 show the similarity between temperature and HI with respect to influencing electricity demand.

4.3 Model Development

This thesis focuses on investigating the influence of relative humidity on load during summer months. Summer (S_t) is defined as a dummy variable. RHS_t , $RH_t \times S$ denotes current hour relative humidity in summer. Similarly, RHS_t^2 , which stands for $RH_t \times RH_t \times S$, is the second order polynomial of relative humidity in summer.

The process of model development in this case study starts with four base models without humidity variables. All the base models are linear regression models. The first base model (B_1) is Tao's Vanilla Benchmark Model as shown in Eqs. (3) and (4). The weather variables used in B_1 are the current hour temperatures (T_t).

The second base model B_2 is an extension to B_1 , which adds a set of effects related to the average temperature of the last 24 hours T_a :

$$y_t = \beta_0 + \beta_1 Trend_t + \beta_2 M_t + \beta_3 W_t + \beta_4 H_t + \beta_5 W_t H_t + f(T_t) + f(T_a) \quad (6)$$

The third base model B_3 is an extension to B_2 , which adds a set of effects related to the temperature of the previous hour T_{t-1} :

$$y_t = \beta_0 + \beta_1 Trend_t + \beta_2 M_t + \beta_3 W_t + \beta_4 H_t + \beta_5 W_t H_t + f(T_t) + f(T_a) + f(T_{t-1}) \quad (7)$$

Likewise, appending a set of effects related to the temperature of the previous two hours T_{t-2} to B_3 produces the last base model, B_4 :

$$y_t = \beta_0 + \beta_1 Trend_t + \beta_2 M_t + \beta_3 W_t + \beta_4 H_t + \beta_5 W_t H_t + f(T_t) + f(T_a) + f(T_{t-1}) + f(T_{t-2}) \quad (8)$$

The primary motivation of this practice is to ensure that the resulting recommendation of relative humidity variables can help improve the temperature base models with various complexities.

Preliminary tests have been conducted to examine the influence of different relative humidity terms and cross effects associated with them. In these tests, three-fold cross validation is used, which split the three-year data (2009-2011) to three combinations as shown in TABLE 6. All of them use two-year data as training data for parameter estimation, and the other one-year data as validation data for variable selection. The simple average of the MAPEs from the three combinations is used for model comparison. The relative humidity related terms are added to the base models one at a time, and checked whether they can reduce the average MAPE by 0.01% in absolute value. The relative humidity variables tested are up to the third order. The temperature variables that interact with relative humidity variables are the current hour temperature up to the third order.

TABLE 6: Cross validation combinations

Combination number	Training data	Validation data
Combination 1	2009, 2010	2011
Combination 2	2009, 2011	2010
Combination 3	2010, 2011	2009

In this thesis, terms that show positive influence in at least three base models are chosen. If a higher order term is selected, the corresponding lower order terms. At the end will also be included. The following candidates are selected: RHS_t , RHS_t^2 , $T_t \times RHS_t$, $T_t^2 \times RHS_t$, $T_t \times RHS_t^2$, $T_t^2 \times RHS_t^2$, $RHS_t \times H_t$, $RHS_t^2 \times H_t$.

In the process of model development, the following two steps are conducted repeatedly for each of the four base models:

- 1) Add relative humidity variables to a base model.

For each base model, candidates mentioned above will be added in the following sequence, grouped based on their features:

- a. RHS_t , RHS_t^2 ;
 - b. $T_t \times RHS_t$, $T_t^2 \times RHS_t$;
 - c. $T_t \times RHS_t^2$, $T_t^2 \times RHS_t^2$;
 - d. $RHS_t \times H_t$, $RHS_t^2 \times H_t$.
- 2) Calculate the MAPE values in a three-fold cross validation setting, and compare the simple average of the MAPEs from the three combinations.

CHAPTER 5: RESULTS AND DISCUSSION

5.1 Cross Validation Results

The MAPEs of the models based on the four base models are listed in TABLE 7. The *None* column corresponds to the MAPEs of the four base models. The columns after it are appending one set of relative humidity terms at a time until all the main and cross effects are included. In total, five models are compared for each base model. It is observed that including relative humidity terms gradually reduces the error on validation data (a.k.a., post-sample fit data) for all four base models. The incremental contribution of the two main effects RHS_t and RHS_t^2 is very minor. Nevertheless, due to the improvement from related cross effects, these two main effects are kept in the model.

TABLE 7: MAPEs(%) of the models with additional relative humidity terms

Base model	<i>None</i>	RHS_t RHS_t^2 $T_t \times RHS_t$ $T_t^2 \times RHS_t$				RHS_t RHS_t^2 $T_t \times RHS_t$ $T_t^2 \times RHS_t$	
		RHS_t RHS_t^2	$T_t \times RHS_t$ $T_t^2 \times RHS_t$	$RHS_t \times H_t$ $RHS_t^2 \times H_t$			
B_1	5.21	5.20	5.08	5.03	4.91		
B_2	4.10	4.10	4.03	3.98	3.85		
B_3	3.87	3.87	3.82	3.76	3.70		
B_4	3.79	3.80	3.73	3.68	3.62		

FIGURE 12 is a summary about the MAPEs of the four temperature base models and those proposed with added relative humidity terms. Overall, the MAPE decreases from 5.21% to 4.91%, from 4.10% to 3.85%, from 3.87% to 3.70%, and from 3.79% to 3.62% for B_1 , B_2 , B_3 and B_4 , respectively. The improvement from adding relative humidity variables appears to be stable and consistent in all base models. The ex post forecast accuracy improvement ranges from 4.39% to 5.76%.

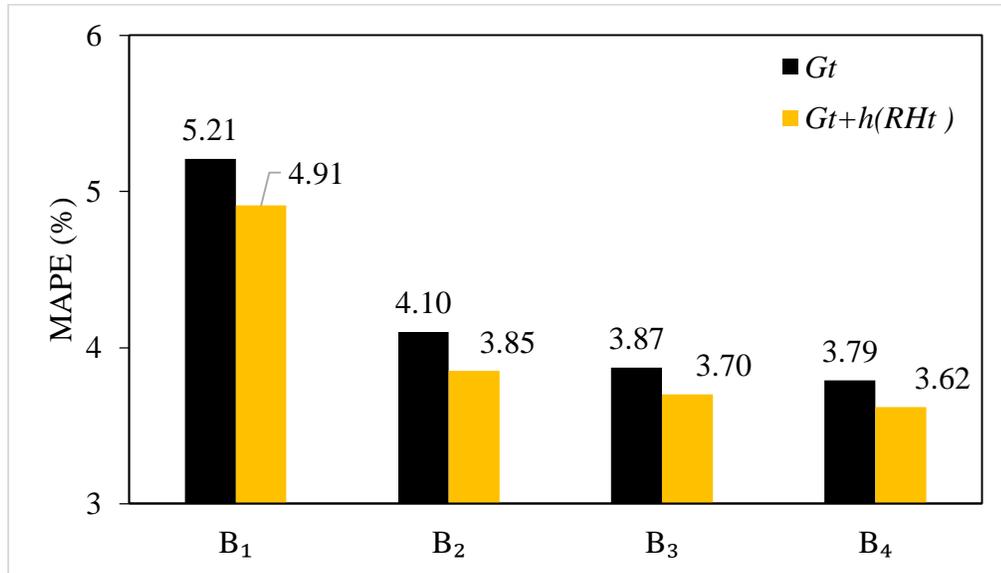


FIGURE 12: MAPEs of base models and proposed with added relative humidity terms

5.2 Recommended Addition of Relative Humidity Variables

To sum up, from the results of adding relative humidity effects to B_1 to B_4 models, it is evident that including relative humidity effects will steadily improve the forecast accuracy. As a result, it is proposed to add the eight effects (RHS_t , RHS_t^2 , $T_t \times RHS_t$, $T_t^2 \times RHS_t$, $T_t \times RHS_t^2$, $T_t^2 \times RHS_t^2$, $RHS_t \times H_t$, $RHS_t^2 \times H_t$) to the temperature

dependent base models to make benchmark models including both temperature and humidity effects.

The proposed model can be presented by this equation:

$$y_t = G_t + h(RH_t) \quad (10)$$

where G_t represents a base model depending upon temperature variables, such as the ones shown in Eqs. (3), (7), (8), and (9).

$$h(RH_t) = \gamma_1 RHS_t + \gamma_2 RHS_t^2 + \gamma_3 T_t \times RHS_t + \gamma_4 T_t^2 \times RHS_t + \gamma_5 T_t \times RHS_t^2 + \gamma_6 T_t^2 \times RHS_t^2 + \gamma_7 RHS_t \times H_t + \gamma_8 RHS_t^2 \times H_t \quad (11)$$

Comparing Eq (11) with the NWS' formula for HI Eq (1), it is evident that Eq (11) extends the HI equation by adding interactions between relative humidity and hour. Another key difference between the two formulae is that the parameters in the proposed model will be estimated based on the data set, while those in the HI formula are pre-defined constants. In other words, the proposed modeling methodology in this thesis offers more flexibility than NOAA's HI formula.

5.3 Out-of-Sample Test Results

An out-of-sample rolling test is used to test the forecast accuracy of the proposed models. Hourly data of the year 2012 has been hidden from parameter estimation or model selection. Here it is used as the hold-out sample to test the model performance. The rolling window embraces two years of data. The following four different forecasting horizons are tested to evaluate the models' accuracy: one-day, one-week, two-week, and one-year.

A model including HI variables is examined to compare the proposed model with conventional practice. Since HI can be seen as an adjustment of the temperature, following the example of Eq (6), $f(HI_t)$ is designed as:

$$f(HI_t) = \alpha_1 HI_t + \alpha_2 HI_t^2 + \alpha_3 HI_t^3 + \alpha_4 HI_t \times M_t + \alpha_5 HI_t^2 \times M_t + \alpha_6 HI_t^3 \times M_t + \alpha_7 HI_t \times H_t + \alpha_8 HI_t^2 \times H_t + \alpha_9 HI_t^3 \times H_t \quad (12)$$

The models tested are listed in TABLE 8. G_t represents a base model as mentioned before. The average MAPEs of three model groups, the base models, the base models with the addition of HI variables, and the base model combined with the addition of recommended RH variables.

TABLE 8: Tested model groups

Model Group	Model equation
TM_1	G_t
TM_2	$G_t + f(HI_t)$
TM_3	$G_t + h(RH_t)$

FIGURE 13 shows the MAPEs of models tested with NCEMC total load data using the four base models. The most salient feature is that the proposed model can significantly improve forecast accuracy for all the forecasting horizons, for all the four base models. It is also noted that the improvement from the HI variables is around half as that of the proposed model for all four base models.

As mentioned before, HI is an adjustment to the temperature. The base models already cover most of the information from temperature. The proposed model, on the other hand, is independent from temperature variables. That is a reason why its room for

improvement is larger. Another reason is that the proposed model offers more flexibility than the HI models. Overall, the models with proposed RH variables outperform the counterpart. The improvement to the base models ranges from 4.05% to 9.39%.

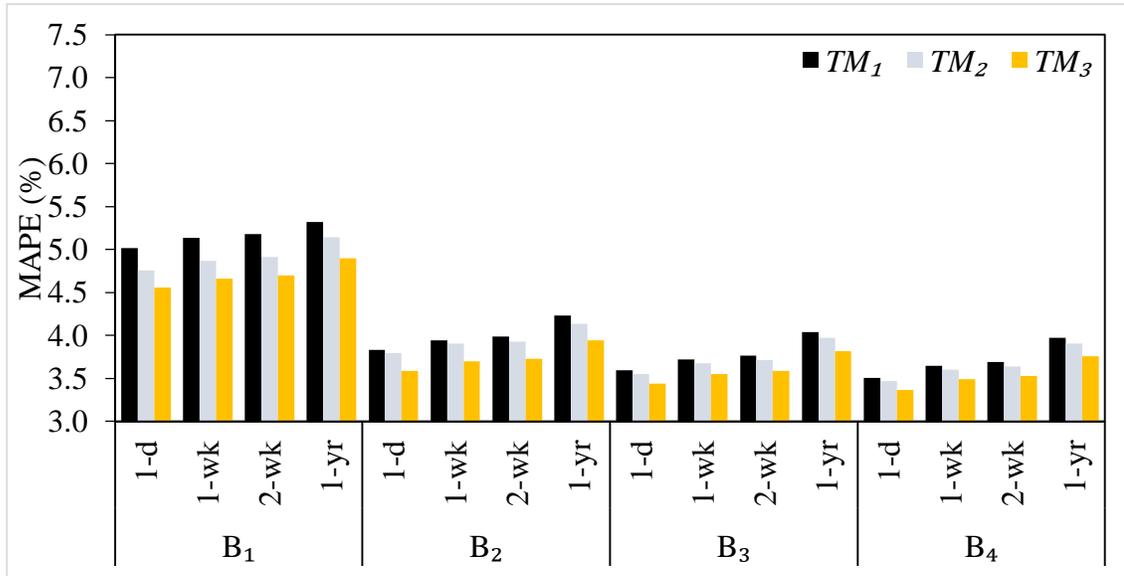


FIGURE 13: MAPEs of models tested with NCEMC total load

It is also interesting to investigate whether the model improvement has seasonal difference. To do so, the average MAPEs of summer and non-summer are calculated for every forecasting horizon and every base model. Most of the prominent characteristics are persistent among different base models. The summer and non-summer MAPEs with the NCEMC total load data using B_1 are presented in FIGURE 14. There are a number of remarkable attributes. First of all, forecasts in summers are better than those in non-summers. Secondly, the proposed model can substantially improve model accuracy in both summer and non-summer.

These outcomes provide another proof that the proposed model is superior to the HI model. Overall, the improvement of the proposed model ranges from 9.44% to 12.36% in relative MAPE reduction in summer. The improvement is smaller in non-summer, from 7.32% to 8.15%. This difference is caused by the relationship between load and relative humidity being stronger in summer as implied in the scatter plots of load and relative humidity by month in FIGURE 6.

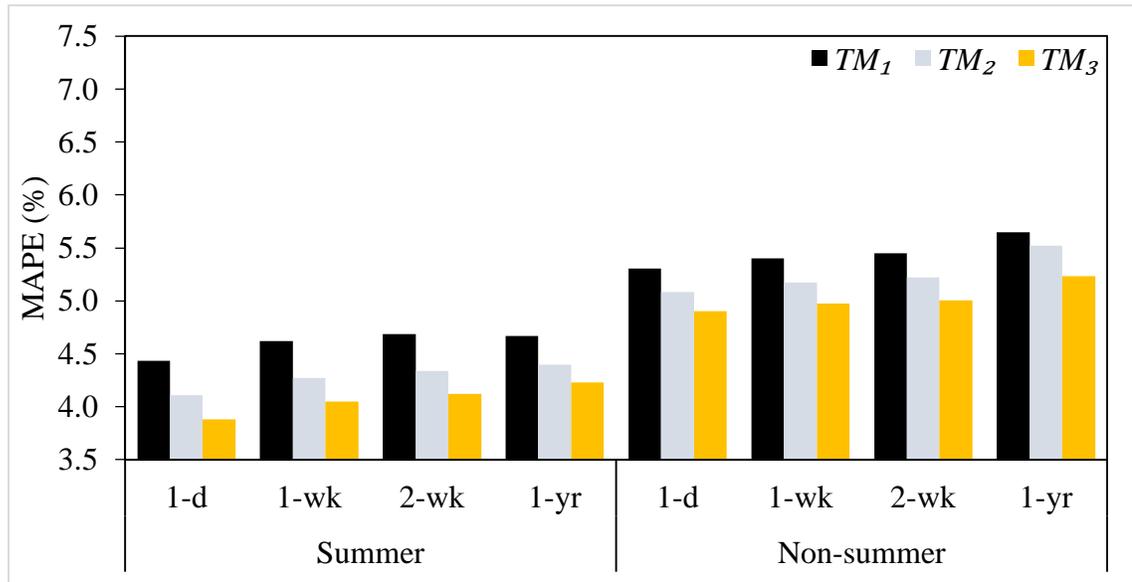


FIGURE 14: Summer and non-summer MAPEs of models tested with NCEMC total load using B_I

5.4 Performance on Supply Area Level

The service territory of NCEMC can be divided into three supply areas. Specifically, SA_1 stands for Progress Energy, SA_2 represents Duke Energy Carolinas, and SA_3 denotes Dominion NC Power. Hence, the stability of the proposed model is also examined by forecasting load of these three supply areas. FIGURE 15 shows the MAPEs of the models tested with SA_1 to SA_3 load, respectively.

The proposed model can notably improve forecast accuracy of all the four base models for all the three supply areas. This is consistent with the results of NCEMC total load. These enhancements are considerably larger than those by adding HI terms, too. The contribution of the proposed model is significant. Nevertheless, those of HI model become marginal when the base models become more complex. The relative MAPE reduction of the proposed model ranges from 2.39% to 7.67%, 3.26% to 8.89%, and 2.10% to 6.43% for SA_1 , SA_2 , and SA_3 , respectively.

The HI models' performances, on the other hand, are not very steady. They are not always improving the models' accuracy, especially when the base model is complicated. In Figure 14b, the MAPEs of TM_2 model are slightly larger than that of TM_1 model when forecasting one-week or two-week ahead SA_2 load using B_3 and B_4 as base model.

The SA_3 forecasts show the narrowest gap between the HI model and the proposed model. This is due to the numerous outliers in the dataset. They deteriorate the overall model accuracy, too. In spite of the data quality issue, the proposed model still outperforms the HI models constantly.

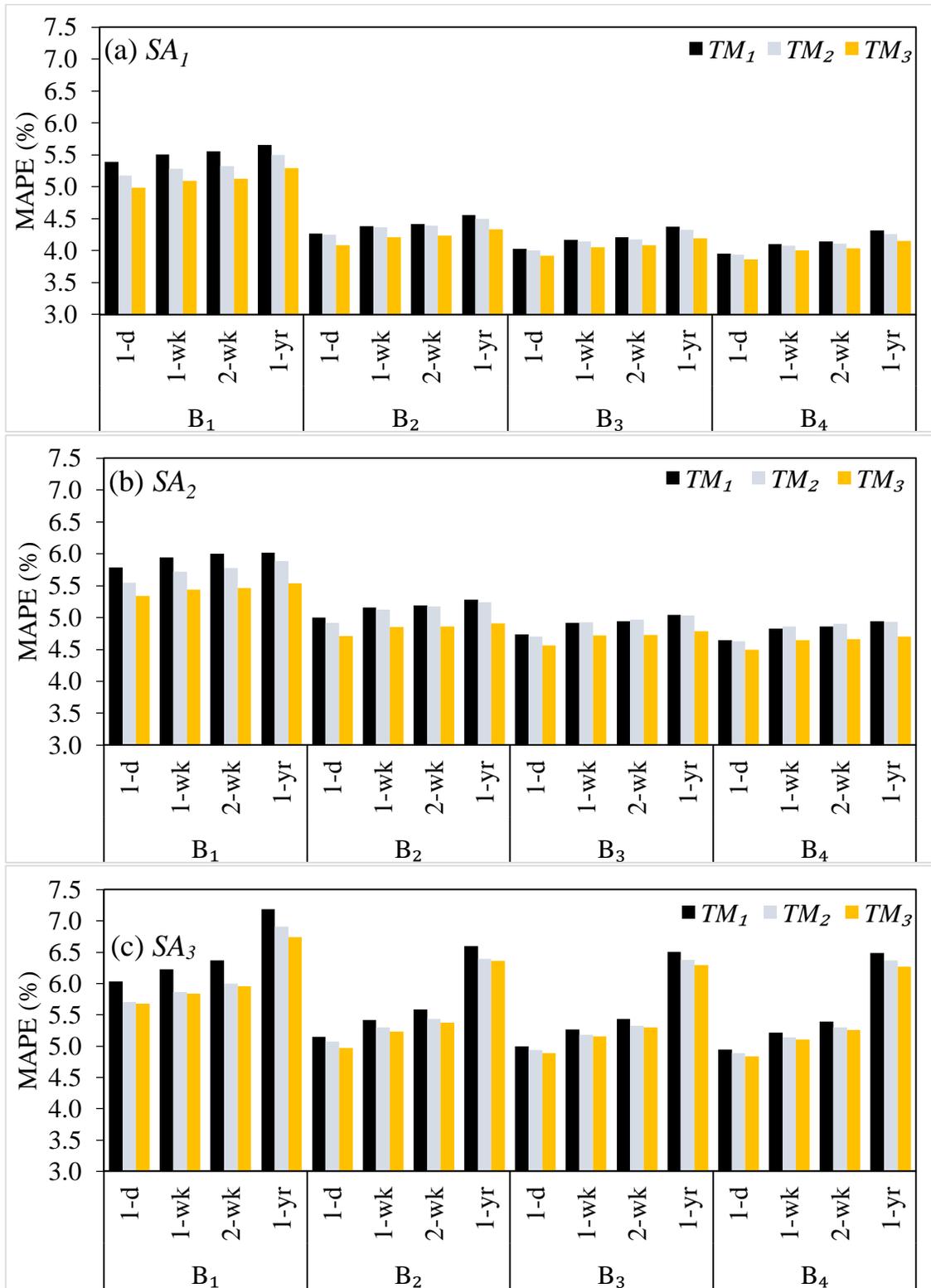


FIGURE 15: MAPEs of models tested with (a) SA_1 , (b) SA_2 , and (c) SA_3 load, respectively

FIGURE 16 shows the seasonal differences of supply areas' load forecast. For consistence, only results of B_1 models are provided. In all the scenarios, the proposed model outdoes the base model and the HI model. In FIGURE 16 a and b, it is noticed that the summer forecasts significantly beat the non-summer forecasts in SA_1 and SA_2 load. The improvement in summer for SA_2 load forecast ranges from 12.22% to 14.26% in relative MAPE reduction. The improvement is smaller for SA_1 load and SA_3 load forecast, where it ranges from 5.17% to 9.07% and from 8.25% to 8.65%, respectively. On the other hand, the improvement in non-summer is smaller, ranges from 6.77% to 7.32%, 5.50% to 6.16%, and 4.63% to 5.13% for the three supply areas, respectively. In SA_3 , the non-summer forecasts outperform the summer forecasts. This is probably due to the data quality problem mentioned in Section 5.4. In general, the proposed model improves forecast accuracy of the base models.

In summary, model tests for various forecasting horizons using different base models prove the stability and superior performance of the proposed model. Its performance is stable when appended to temperature base models of various complexities. It is a robust model that it can improve ex post load forecast accuracy in 24-hour ahead, one-week ahead, two-week ahead and one-year ahead settings and for all three supply areas and the total load of NCEMC.

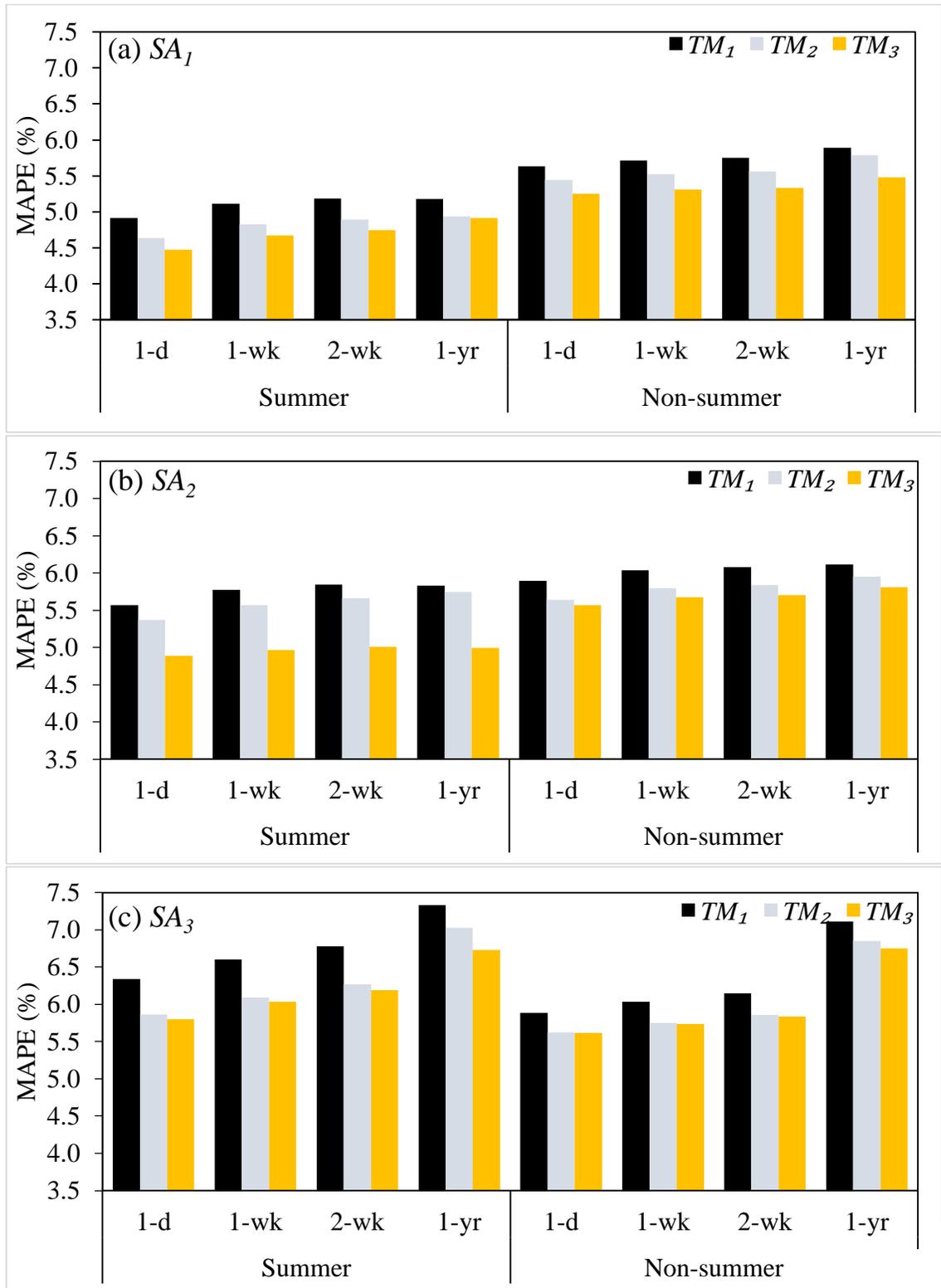


FIGURE 16: Summer and non-summer MAPEs of models tested with supply areas' load using B_1

CHAPTER 6: CONCLUSION

The requirement of matching supply and demand in electric power system makes accurate load forecasts necessity in the electric industry. Weather is one of the key driving factors of load forecasting since numerous electric appliances became popular. Various weather variables have been deployed in load forecasting. Temperature is the most commonly used one. Humidity has been used in some research and some utility forecasting models. However, humidity variables have not been well studied. When humidity variables are used in load forecasting models, they are usually embedded in the form of HI.

This thesis investigated method to add relative humidity information to temperature base models through a case study from NCEMC. In order to better understand the influence of relative humidity on load, we performed extensive exploratory data analysis. Scatter plots of load and relative humidity by months suggest a strong relationship between the two in warmer months from June to September. Therefore, the period from June to September is defined as summer. The differences among the relationship between load and relative humidity by hour in summer are notable, too.

Four base models are used for model development. The first one is Tao's Vanilla Benchmark Model, a multiple linear regression model. Other base models are derived from it by adding lagged and average temperature variables. Adding relative humidity

variables to those base models can improve forecast accuracy. We propose to add eight effects (RHS_t , RHS_t^2 , $T_t \times RHS_t$, $T_t^2 \times RHS_t$, $T_t \times RHS_t^2$, $T_t^2 \times RHS_t^2$, $RHS_t \times H_t$, $RHS_t^2 \times H_t$) to the temperature based models. The ex post forecast accuracy improvement ranges from 4.39% to 5.76% on validation (post-sample fit) data.

Model test is done for four different forecasting horizons on rolling basis, one-day, one-week, two-week, and one-year. The service areas of NECMC can be divided into three major supply areas. The proposed model is tested through forecasting load of both NECMC total load and the three supply areas. The various model test results prove the superiority of the proposed model to not only the base models, but also the HI model. The improvement on the test data (holdout sample) ranges from 4.05% to 9.39% for NCEMC total ex post load forecasting with the four forecasting horizons. In conclusion, the proposed model meets the criteria of being simple, widely applicable, reproducible, and accurate. It can be deployed in utilities for various forecasting horizons and at different hierarchies.

Relative humidity is a factor influencing the load during warmer weather, while wind is important for load during cold weather load. Therefore, we recognize a few possible future research directions. First, we can investigate the models with both temperature and wind speed variables. Second, we can investigate the models with temperature, relative humidity and wind speed variables.

REFERENCES

- Abou-Hussien, M. S., Kandlil, M. S., Tantawy, M. A., & Farghal, S. A. (1981). An Accurate Model for Short-Term Load Forecasting. *IEEE Transactions on Power Apparatus and Systems, PAS-100*(9), 4158–4165.
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4*, 40–79.
- Bluestein, M. (2015). Basic atmospheric structure and concepts | Wind chill. In *Encyclopedia of Atmospheric Sciences (Second Edition)* (pp. 7–11).
- Chen, B.-J., Chang, M.-W., & Lin, C.-J. (2004). Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001. *IEEE Transactions on Power Systems, 19*(4), 1821–1830.
- Douglas, A. P., Breipohl, A. M., Lee, F. N., & Adapa, R. (1998). The impacts of temperature forecast uncertainty on bayesian load forecasting. *IEEE Transactions on Power Systems, 13*(4), 1507–1513.
- Fan, J. Y., & McDonald, J. D. (1994). A real-time implementation of short-term load forecasting for distribution power systems. *IEEE Transactions on Power Systems, 9*(2), 988–994.
- Fan, S., Chen, L., & Lee, W.-J. (2008). Machine learning based switching model for electricity load forecasting. *Energy Conversion and Management, 49*(6), 1331–1344.
- Fan, S., & Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems, 27*(1), 134–141.
- Fan, S., Methaprayoon, K., & Lee, W. J. (2009). Multiregion load forecasting for system with large geographical area. *IEEE Transactions on Industry Applications, 45*(4), 1452–1459.
- Goude, Y., Nedellec, R., & Kong, N. (2014). Local short and middle term electricity load forecasting with semi-parametric additive models. *IEEE Transactions on Smart Grid, 5*(1), 440–446.
- Hagan, M. T., & Behr, S. M. (1987). The Time Series Approach to Short Term Load Forecasting. *IEEE Transactions on Power Systems, 2*(3), 785–791.
- Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems, 16*(1), 44–55.

- Hong, T. (2010). *Short Term Electric Load Forecasting*. North Carolina State University.
- Hong, T. (2014). Energy Forecasting : Past , Present , and Future. *The International Journal of Applied Forecasting*, (32), 43–48.
- Hong, T., & Fan, S. (2015). Probabilistic electric load forecasting: a tutorial review. *International Journal of Forecasting*, to appear.
- Hong, T., Liu, B., & Wang, P. (2015). Electric Load Forecasting with Recency Effect: a Big Data Approach. *International Journal of Forecasting*, to appear.
- Hong, T., Pinson, P., & Fan, S. (2014). Global energy forecasting competition 2012. *International Journal of Forecasting*, 30(2), 357–363.
- Hong, T., & Shahidehpour, M. (2015). *Load forecasting case study*. National Association of Regulatory Utility Commissioners.
- Hong, T., & Wang, P. (2014). Fuzzy interaction regression for short term load forecasting. *Fuzzy Optimization and Decision Making*, 13(1), 91–103.
- Hong, T., Wang, P., & White, L. (2015). Weather station selection for electric load forecasting. *International Journal of Forecasting*, 31(2), 286–295.
- Hong, T., Wilson, J., & Xie, J. (2014). Long term probabilistic load forecasting and normalization with hourly information. *IEEE Transactions on Smart Grid*, 5(1), 456–462.
- Hor, C. L., Watson, S. J., & Majithia, S. (2005). Analyzing the impact of weather variables on monthly electricity demand. *IEEE Transactions on Power Systems*, 20(4), 2078–2085.
- Hyndman, R. J., & Fan, S. (2010). Density forecasting for long-term peak electricity demand. *IEEE Transactions on Power Systems*, 25(2), 1142–1153.
- Khotanzad, A., Afkhami-Rohani, R., & Maratukulam, D. (1998). ANNSTLF - Artificial neural network short-term load forecaster - generation three. *IEEE Transactions on Power Systems*, 13(4), 1413–1422.
- Kiartzis, S. J., Bakirtzis, A. G., & Petridis, V. (1995). Short-term load forecasting using neural networks. *Electric Power Systems Research*, (33), 1–6.
- Kutner, H., Nachtsheim, J., Neter, J., & Li, W. (2004). *Applied Linear Statistical Models*, 5th ed. McGraw-Hill/Irwin.

- McSharry, P. E., Bouwman, S., & Bloemhof, G. (2005). Probabilistic forecasts of the magnitude and timing of peak electricity demand. *IEEE Transactions on Power Systems*, 20(2), 1166–1172.
- Mirasgedis, S., Sarafidis, Y., Georgopoulou, E., Lalas, D. P., Moschovits, M., Karagiannis, F., & Papakonstantinou, D. (2006). Models for mid-term electricity demand forecasting incorporating weather influences. *Energy*, 31(2-3), 208–227.
- Papalexopoulos, A. D., Hao, S., & Peng, T.-M. (1994). An implementation of a neural network based load forecasting model for the EMS. *IEEE Transactions on Power Systems*, 9(4), 1956–1962.
- Papalexopoulos, A. D., & Hesterberg, T. C. (1990). A regression-based approach to short-term system load forecasting. *IEEE Transactions on Power Systems*, 5(4), 1535–1547.
- Park, D. C., El-Sharkawi, M. a., Marks, R. J., Atlas, L. E., & Damborg, M. J. (1991). Electric load forecasting using an artificial neural network. *IEEE Transactions on Power Systems*, 6(2), 442–449.
- PJM. (2015). *Load Forecasting and Analysis*. Retrieved from <http://www.pjm.com/contributions/pjm-manuals/manuals.html>
- Rahman, S. (1990). Formulation and analysis of a rule-based short-term load forecasting algorithm. *Proceedings of the IEEE*, 78(5), 805–816.
- Rahman, S., & Hazim, O. (1996). Load forecasting for multiple sites: Development of an expert system-based technique. *Electric Power Systems Research*, 39(3), 161–169.
- Ranaweera, D. K., Hubele, N. F., & Karady, G. G. (1996). Fuzzy logic for short term load forecasting. *International Journal of Electrical Power & Energy Systems*, 18(4), 215–222.
- Raza, M. Q., & Khosravi, A. (2015). A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews*, 50, 1352–1372.
- Rothfus, L. P. (1990). *The heat index equation (or, more than you ever wanted to know about heat index)*. Fort Worth, Texas.
- Saini, L. M. (2008). Peak load forecasting using Bayesian regularization, Resilient and adaptive backpropagation learning based artificial neural networks. *Electric Power Systems Research*, 78(7), 1302–1310.
- Schewe, P. F., & Brennan, T. J. (2008). *The grid: A journey through the heart of our electrified world*. *Optical Engineering* (Vol. 47). The National Academies Press.

- Song, K.-B., Baek, Y.-S., Hong, D. H., & Jang, G. (2005). Short-term load forecasting for the holidays using fuzzy linear regression method. In *IEEE Power Engineering Society General Meeting, 2005*.
- Song, Y.-H., & Wang, X.-F. (2003). *Operation of Market-oriented Power Systems*. Springer-Verlag London.
- Steadman, R. G. (1979). The Assessment of Sultriness. Part I: A Temperature-Humidity Index Based on Human Physiology and Clothing Science. *Journal of Applied Meteorology*, 18(7), 861–873.
- 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*. Wiley.
- Xie, J., Hong, T., & Stroud, J. (2015). Long-Term Retail Energy Forecasting With Consideration of Residential Customer Attrition. *IEEE Transactions on Smart Grid*, 6(5), 2245–2252.