PROBABILISTIC ELECTRIC LOAD FORECASTING

by

Jingrui Xie

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Infrastructure and Environmental Systems

Charlotte

2016

Approved by:

______________________________
Dr. Tao Hong

______________________________
Dr. Badrul Chowdhury

______________________________
Dr. Churlzu Lim

______________________________
Dr. Pu Wang

______________________________
Dr. Jy Wu

# ABSTRACT


JINGRUI XIE. Probabilistic electric load forecasting. (Under the direction of DR. TAO HONG)

Traditionally, utilities have been conservative regarding the infrastructure upgrade. This makes electric load forecasts critical for guiding electric utilities' operation, planning, and maintenance decision-making. For a long while, utilities have relied on the point load forecast that provides a single expected value about the future to guide their decision process. However, when the load forecasting is conducted for the middle or long term, things become more uncertain. For example, while the weather for the next week may be predictable, the forecast of it for the next month(s) or year(s) become unreliable. An unreliable weather forecast is less likely to help on the load forecasting practices. Furthermore, the modernization of the grid has brought many changes to the electric utility industry. Changes such as the increasing penetration of renewable energy, the emerging distribution generation, and the bi-directional communication between the supplier and the end-users have brought much more uncertainties for the utilities' load forecasting practices. The single-valued forecast or point forecast that gives a deterministic forecast about the future load does not provide any information on such uncertainties. In contrast, a probabilistic forecast that estimates the respective probabilities for all the possible future outcomes of a random variable provides opinions on the uncertainties.

Although probabilistic forecasting have been studied for decades and researchers have tried to apply those techniques for probabilistic load forecasting (PLF) for the past several years, there are still many challenging issues in the PLF field, such as lack of quantitative evaluations on the PLF methods, ad-hoc selection of input scenarios for PLF, and the lack

of practical guides for PLF. This dissertation dissects the PLF problem into three key components including the input scenario simulation, the modeling techniques, and the residual analysis. From the input scenario simulation perspective, this dissertation first raises a critical yet never answered question about the lack of methodological foundation for practicing probabilistic load forecasting through input simulation. Such lack of methodological foundation typically results in ad-hoc, judgmental and indefensible choice during the scenario generation step. This dissertation then investigates a framework to evaluate the effectiveness of three different temperature scenario generation techniques, namely the fixed-date method, the shifted-date method, and the bootstrap method, from which an empirical rule-of-thumb is developed to guide the temperature scenario generation practice for PLF. The establishment of this evaluation framework helps to lay a solid methodological foundation for practicing probabilistic load forecasting through temperature scenario simulation. The proposed framework can also be extended to evaluate and guide the practices on generating other input scenarios. The modeling techniques will still rely on the representative ones developed for point load forecasting but the focus will be on how to convert point forecasting results to probabilistic ones. From the residual simulation perspective, studying residual series itself is not anything new in load forecasting and its utility applications. Back to 1970s, for example, researchers were using mean and standard deviation to characterize uncertainties around electric load forecasts for probabilistic load flow analysis. However, most papers in the literature that modeled load forecast residuals assumed normality for the residual distribution. Such normality assumption has rarely been verified through any formal statistical test. This dissertation conducts a comprehensive study regarding the normality assumption of the residuals. It not only studies the residuals from load forecasting as a whole but further considers the potential impact of multiple seasonality existing in electricity demand on the normality assumption of residuals. Moreover, it comprehensively studies whether simulating residuals with the normality assumption improves the probabilistic forecasts which has never been studied before.

Two case studies are used in this dissertation: (1) the first and primary case study is based on the system total demand of North Carolina Electric Membership Cooperation (NCEMC). NCEMC serves 93 out of the 100 counties of North Carolina. The weather condition varies quite a bit within NCEMC's service territory; (2) the second case study is anonymous data from the load forecasting track of Global Energy Forecasting Competition 2014 (GEFCom2014). The data is public available which allows others to reproduce the results presented in this dissertation. Although only two case studies are presented in this dissertation to demonstrate the implementation and comparisons of the different PLF techniques, the PLF techniques discussed in this dissertation have outperformed the ones developed by several other PLF groups worldwide.

From research perspective, this dissertation raises and answers questions regarding the methodological foundation for practicing temperature scenario generation and residual simulation techniques for PLF. The study in this dissertation also points out directions for future research in the PLF field. For example, how to generate and evaluate other weather scenarios such as relative humidity. From practices perspective, the findings from this dissertation study offer multiple practical options for utilities' probabilistic load forecasting practices. Some of the findings presented in this dissertation have been in production use by utilities since 2012 with outstanding performance.

# DEDICATION

To My Family

## ACKNOWLEDGMENTS

Dr. Tao Hong was the first person who introduced me to the energy forecasting field, then became the advisor of my master study in the engineering management program and my Ph.D. study in the infrastructure and environmental systems program. Dr. Hong guided me step-by-step in learning the analytical skills in energy forecasting, building up critical thinking in solving problems, and establishing the attitude and continuous learning habits a researcher should have. His passionate in teaching and researching has inspired me for all the time we have worked together. Dr. Hong is always supportive as an advisor and a friend that encouraged me to continuously challenge myself for improvements. I would like to express my great appreciation for his mentoring and supports in my academic and professional career growth.

Dr. Badrul Chowdhury was the first one who systematically introduced me to the fields of the energy system and energy market. Thanks to Dr. Chowdhury for helping me build the fundamental background of the energy system and market.

Dr. Churlzu Lim opened the door of mathematical optimization for me. He introduced the fundamental and advanced optimization concepts and techniques that have helped me to better understand the modeling techniques that I have used for my energy forecasting research.

Dr. Pu Wang's research on point load forecasting has impacted many of my probabilistic load forecasting implementations. She is also a great co-worker and excellent researcher who provided insightful opinions for my research.

Dr. Jy Wu as the director of the INES program has provided tremendous supports to my study and research. I would like to give my great thanks to Dr. Jy Wu.

Lastly, but the most importantly, I would like to express my deepest appreciation to my husband, Yue Chai, who has unconditionally loved and encouraged me all the way along.

TABLE OF CONTENTS


LIST OF TABLES xii

LIST OF FIGURES xiii

CHAPTER 1: INTRODUCTION 1

1.1 Electric Load Forecasting 1

1.2 Classification of Load Forecasting Problems 3

1.3 Probabilistic Electric Load Forecasting 5

CHAPTER 2: LITERATURE REVIEW 8

2.1 Overview 8

2.2 Load Forecasting Techniques 10

2.2.1 Regression Analysis 10

2.2.2 Artificial Neural Networks 13

2.2.3 Others 14

2.3 Explanatory Variables 16

2.3.1 Weather 16

2.3.2 Calendar 18

2.3.3 Economy 20

2.4 Probabilistic Load Forecasting Methods 22

CHAPTER 3: THEORETICAL BACKGROUND 25

3.1 Multiple Linear Regression 25

3.2 Artificial Neural Networks 28

3.3 Forecast Evaluation 30

3.3.1 Point Forecast Evaluation 30

3.3.2 Probabilistic Forecast Evaluation 32

3.3.3 Sliding Simulation 36

CHAPTER 4: CASE STUDY DATA 38

4.1 North Carolina Electric Membership Corporation 38

4.2 Global Energy Forecasting Competition 2014 41

CHAPTER 5: PROBABILISTIC LOAD FORECASTING 43

5.1 Point Forecasting Models 43

5.1.1 Linear Regression Model 43

5.1.2 Artificial Neural Networks 45

5.2 Input Scenario Generation 46

5.2.1 Temperature Scenario based Probabilistic Load Forecasting 46

5.2.2 Fixed-date Method 48

5.2.3 Shifted-date Method 51

5.2.4 Bootstrap Method 55

5.2.5 Discussion on Empirical Rule-of-thumb 58

5.2.6 Results from Nonlinear Models 61

5.2.7 Results from GEFCom2014-L Data 62

5.2.8 Pros and Cons 64

5.3 Residual Simulation 65

5.3.1 Normal Distribution 66

5.3.2 Group Analysis 68

5.3.3 Post-Processing Probabilistic Load Forecasts with Simulated Residuals 69

5.3.4 Additional Simulation 72

5.3.5 Results from Nonlinear Models 75

5.3.6 Results from GEFCom2014-L Data 77

CHAPTER 6: CONCLUSION AND FUTURE WORK 78

REFERENCES 80

# LIST OF TABLES

TABLE 1: Lower and upper bounds of *n* for different *k* 60

TABLE 2: Quantile scores of year 2011 62

TABLE 3: Quantile scores of year 2011 by shifting the *k* years historical temperature series *n* days around 63

TABLE 4: Pros and cons of three temperature scenario generation techniques 64

TABLE 5: Normality test passing rates (in %) of ex-post point forecast residuals with various grouping methods 68

TABLE 6: Quantile scores for probabilistic forecasts based on various residual simulation methods 70

TABLE 7: Quantile scores for probabilistic forecasts based on various residual simulation 75 methods 75

TABLE 8: Normality test passing rates and quantile scores 77

# LIST OF FIGURES

FIGURE 1: Organization of the dissertation 7

FIGURE 2: Scatter plot of hourly load and temperature during different periods of a day 27

FIGURE 3: A single-layer feeding-forward neural network 29

FIGURE 4: Hourly load and temperature of NCEMC 39

FIGURE 5: Evaluation using sliding simulation for NCEMC 40

FIGURE 6: Hourly load and temperate of GEFCom2014-L data 42

FIGURE 7: Illustration of generating *j* point load forecasts from *j* temperature scenarios 46

FIGURE 8: Illustration of generating *k* temperature scenarios using fixed-date method 48

FIGURE 9: Quantile scores by year 49

FIGURE 10: First-order difference of adjacent average quantile scores from using *k*-1 and *k* years of historical temperature series 50

FIGURE 12: Quantile scores from using temperature series simulated by the shifted-date method 52

FIGURE 13: Quantile forecasts of a summer peak week (Jul 26 – Aug 01, 2011) 54

FIGURE 14: Generating additional temperature scenarios by bootstrapping 55

FIGURE 15: Scatter plot of quantile scores from using fixed and simulated temperature series 56

FIGURE 16: Hourly temperature from the simulated series during the summer peak week of 2006 57

FIGURE 17: Quantile score of shifting the *k* years of historical temperature series *n* days around 59

FIGURE 18: Quantile score from three different temperature generation techniques 61

FIGURE 19: Normality test on ex-post point forecast residuals of linear models 67

FIGURE 20: Normality test passing rate vs. quantile score 71

FIGURE 21: Quantile scores from different number of residual simulations 73

FIGURE 22: Probabilistic monthly peak forecasts with different number of residual simulations 74

FIGURE 23: Normality test passing rate vs. quantile score (Case = NCEMC, nonlinear models) 76

# CHAPTER 1: INTRODUCTION

## 1.1 Electric Load Forecasting

Electric load forecasting, which will be used interchangeably with the term load forecasting in this paper, is the process to predict the electric demand at certain geographical and/or temporal interval into the future. Load forecasts have been used in utilities' operational and planning process since the first day of the central electric power system, while the importance and complicated level of load forecasting problems have evolved. During the very early age of the central electric power system when central stations only served small nearby areas, load forecasting problems were relatively simple and not critical. Counting the number of light bulbs installed on the system might be sufficient to determinate the system capacity. With the advent of alternating current (AC) power system in the late 1890s, electricity can be generated and transmitted to multiple end uses and a broader service area. The inventions in the 20$^{th}$ century such as air conditioner, electric washing machines, electronic television, personal computers, etc. further enriched the variety of electricity end uses and boosted the demand. Such changes made load forecasts more and more important to help maintain the system and financial health of a utility. Load forecasting problems also became more complicated with diversified use patterns that were mainly driven by the interaction of weather conditions and human activities. Since then, utilities have relied on expected values of the load for their generation, distribution, transmission, and service decision makings. Various statistical and artificial intelligence load forecasting techniques have been implemented to provide single-value outputs (point forecasts) of the electric demand into every step of the future. The modernization of the grid in the 21st such as the bidirectional communication between electricity suppliers and end users, the emerging distributed grid, and the increasing penetration of renewable energy brought greater uncertainties to both of the supply and the demand side. Utilities also face more intensive competitions in the electricity market than ever before. As a result, comparing to traditional point forecasts, probabilistic load forecasts have become increasingly useful tools given their success in quantifying the uncertainties in electricity demand.

## 1.2 Classification of Load Forecasting Problems

Load forecasting problems can be categorized from various perspectives. Based on the forecasting horizon, the problem can be classified into short term load forecasting which provides forecasts for the next few hours to few days, and medium or long term load forecasting which yields forecasts from two weeks and up to decades [1]. Alternatively, based on the output formats, the problem can be classified into point forecasting and probabilistic forecasting. Point load forecasts provide a single value for each step into the forecast horizon, while probabilistic load forecasts that are in the form of quantiles, interval, or density function can provide more comprehensive information about the future [2]. Based on the resolution of the data, the problem can also be categorized into hourly, daily, monthly, or annual load forecasting [3]. These aforementioned categories can also interact with each other to further categorize the load forecasting problems such as long term probabilistic load forecasting with hourly information [4].

For the past several decades, point load forecasting has been the focus of academia and industry: point weather forecasts are fed into load forecasting models to provide short term point load forecast [1], [5]–[7], while normalized weather is typically used to generate medium or long term point load forecasts [8]–[10]. During the recent few years, the increased uncertainties on both power supply and demand sides have pushed the electric power industry to take innovative approaches to forecasting. Since probabilistic forecasts offer much more comprehensive information about the future than point forecasts, they are quite helpful to help make informed decisions in the dynamic environment. Therefore, probabilistic load forecasts are becoming more and more attractive to the load forecasting professionals. For example, for long-term planning purpose, utilities may use the median of the probabilistic load forecasts to describe the normal load, while the high percentiles of it (say 90$^{th}$ percentile) can be used to describe the demand under server conditions such as an abnormally hot summer or cold winter day.

In addition, some traditional load forecasting methods or practices especially for long term load forecasting have relied on low-resolution data at annual or monthly interval [11]–[13]. The low-resolution data often offers a limited amount of observations for modeling, which can hardly support enough explanatory variables in a model to capture all the salient features in the electricity demand series. The best practice today is to take advantage of the high-resolution data, such as hourly or sub-hourly load and weather data, to build load forecasting models [3], [4].

In this dissertation, the probabilistic load forecasting problem will be studied with hourly information.

## 1.3 Probabilistic Electric Load Forecasting

Probabilistic forecasts estimate the probability of all the possible outcomes of a random event. It provides forecast on the results of a future event as well as opinions on the uncertainties associated with them. Recent years, probabilistic forecasting methodologies have been studied for electric load forecasting problem, but there are still many challenging issues such as lack of quantitative evaluations on the probabilistic load forecasting methods, ad-hoc selection of input scenarios, and the lack of practical guides. This dissertation dissects the probabilistic electric load forecasting problem into three components including input scenario simulation, modeling techniques, and output analysis. For the input scenarios simulation, a framework will be proposed to evaluate three temperature scenario generation techniques namely the fixed-date method, the shifted-date method, and the bootstrap. The modeling techniques will still rely on the representative ones including multiple linear regression models and artificial neural networks while the focus will be on how to convert point forecasting results to the probabilistic ones. The residuals from the point forecasting models will be analyzed to address the question whether residuals simulated from normal distribution will help improve the quality of probabilistic forecasts for the first time.

The organization of this dissertation is presented in Figure 1. Chapter 2 first provides an overview of the review papers on load forecasting. It then reviews the literature of load forecasting from three aspects, that is the load forecasting techniques, the explanatory variables, and the probabilistic load forecasting techniques. Chapter 3 provides the background information for the main forecasting techniques that will be implemented in this study. That includes the multiple linear regression technique, the artificial neural networks, and the forecast evaluation technique. Chapter 4 introduces the case study data. Chapter 5 presents the process to generate probabilistic forecasts from point forecasting models. Two techniques will be discussed, namely, input scenario generation and residual simulation. The dissertation is then concluded in Chapter 6 with the discussion of the possible extension of the study.

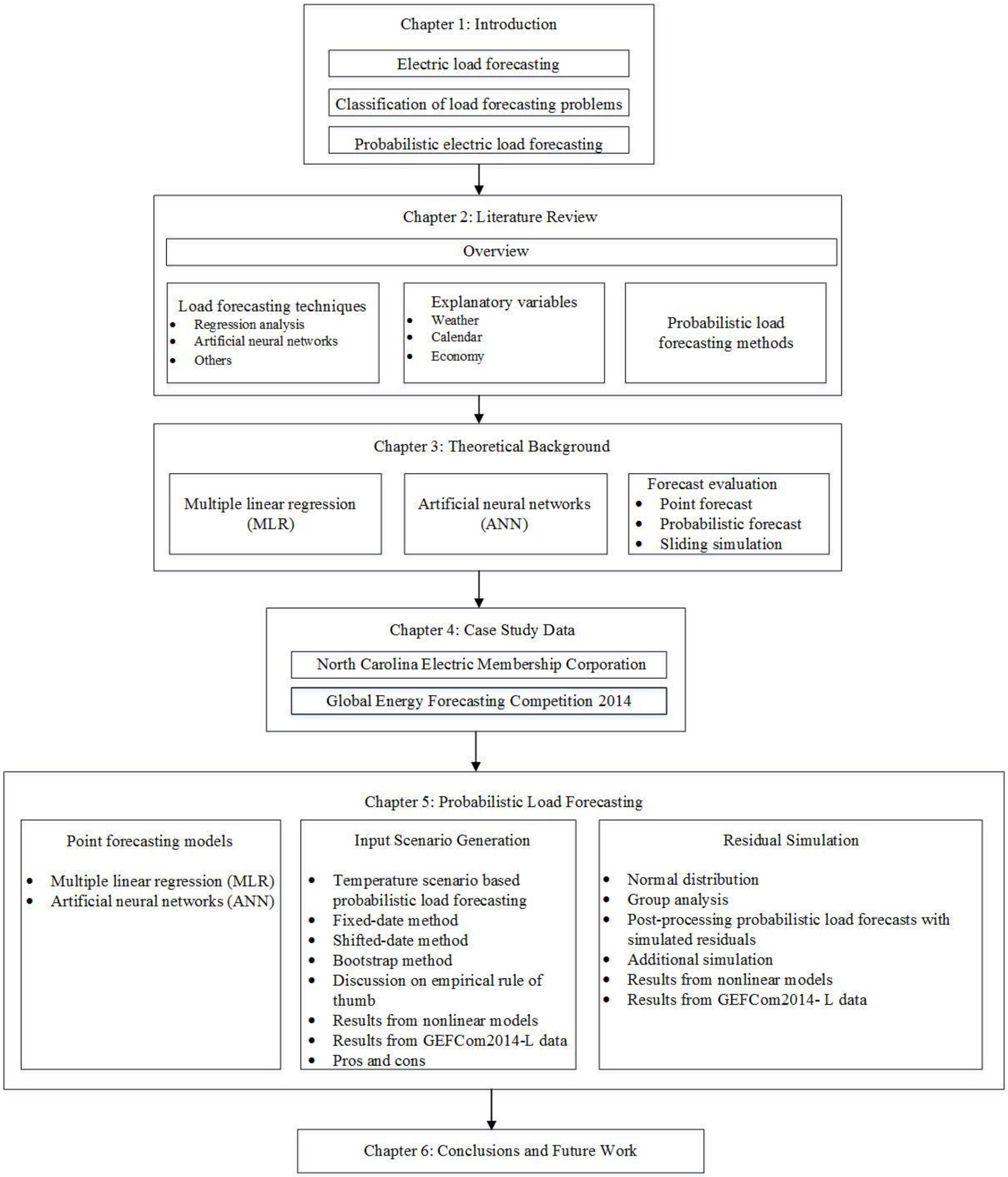



Figure 1: Organization of the dissertation

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Overview

The literature review will start with some review papers in electric load forecasting. Then, it will be devoted to reviewing the following three aspects of load forecasting: (1) representative literature on point load forecasting techniques; (2) representative literature that investigate explanatory variables for load forecasting models; and (3) probabilistic load forecasting techniques to date.

Generally, load forecasting techniques can be classified into two categories: statistical approaches such as regression analysis and time series analysis, and artificial intelligence methods such as artificial neural networks and support vector machine.

Hippert et al. [7] reviewed the representative papers published in the 1990s that reported the application of artificial neural networks (ANN) on short term load forecasting. It aimed to clarify the skepticism on the advantages of applying ANN for forecasting. The paper

highlighted two issues of the existing literature: (1) over-parameterization, and (2) lack of systematic tests.

Weron [5] reviewed a wide range of statistical methods for electric load forecasting including similar-day method, regression methods, and time series methods such as exponential smoothing and autoregressive moving average model (ARMA). A number of case studies and implementations of different techniques in MATLAB were also provided to help researchers and analysts in the load forecasting field.

Taylor and McSharry [14] evaluated five different methods including autoregressive integrated moving average (ARIMA) model, periodic AR model, double seasonal Holt-Winters exponential smoothing, an intraday cycle exponential smoothing model, and a principal component analysis based method for short term load forecasting (i.e. up to 24 hours ahead).

Hong [1] provides the most comprehensive review so far for short term load forecasting of the past 50 years. The review was dissected into three parts: (1) a wide range of the load forecasting techniques including statistical approaches and artificial intelligence techniques (2) the explanatory variables for load forecasting models and (3) the most representative publications. It also discussed several issues with the existing literature such as lack of benchmark model to compare with, and the lag in applying modern statistics to short term load forecasting.

Hong and Fan [2] presents the first review paper that was devoted to probabilistic load forecasting across all forecast horizons. It covered the probabilistic literature from two perspectives: (1) on the application side, researches use probabilistic load forecasting as inputs for the decision-making process (2) on the technical and methodological development side, enhancing the quality of the forecast was the main focus.

## 2.2 Load Forecasting Techniques

### 2.2.1 Regression Analysis

Regression analysis is one of the most widely used statistical techniques for electric load forecasting. Regression methods are usually used to model the relationship between the electricity consumption and other variables including weather variables such as temperature, calendar variables such as day type, socioeconomic variables such as macroeconomic index, or the interactions of them.

Papalexopoulos and Hesterberg [15] provided one of the first few early papers that fully studied applying regression analysis for short term load forecasting. Several aspects including the temperature and holiday effects on load consumption, and the robust parameter estimation using regression-based approach were investigated. The proposed method was tested against real data from Pacific Gas and Electric's service territory for short term peak load and hourly load forecasting. Although there exist some issues with the details of the paper, for example, the proposing of using weighted least square method for handling outliers is not very convincing, this paper provides a comprehensive fundamental work for applying regression analysis for short term load forecasting. Later papers have extended on different aspects of the multiple linear regression method for load forecasting [16], [17].

A more recent and more comprehensive study on applying regression analysis for short term load forecasting was provided by Hong [1]. In the paper, an integrated forecasting framework with a short term load forecasting engine was investigated. Most of the paper was devoted to the development of the short term load forecasting engine while some discussions were provided for extending the engine to long term load forecasting. A regression based benchmark model with considerations of temperature and calendar effects was first established. The recency effect, the weekend effect, and the holiday effect were then introduced to elaborate the benchmark model. The recency effect describes the impact of lag temperature on the current load consumption. The weekend effect groups similar weekdays together for modeling to reduce the complexity of the model. The holiday effect models holidays as alternative weekday or weekend days to tackle the problem of short of historical data and changing consumption patterns for holiday load forecasting. In addition, an exponentially weighted least squares approach with higher weights assigned to the most recent observations was proposed in the paper to emphasize the recent status of the system. A systematic and data-driven variable selection process was proposed to include these aforementioned effects and approach. Many aspects of this paper have been reproduced or extended in later literature for both of short term and middle/long term load forecasting: the benchmark model have been reproduced by many scholars [18], [19] and used as the benchmark model in the Global Energy Forecasting Competition 2012 (GEFCom2012) [20]. It was also further elaborated to consider other weather variables such as humidity [21]. The recency effect was further studied for point load forecasting by Wang *et al.* [22] and for probabilistic load forecasting by Xie and Hong [23]. Xie and Chien [24] recently investigated the holiday effect for holiday load forecasting. The proposed variable selection process has been implemented by several later literature for selecting the proper load forecasting model [4], [25]–[27].

A refined parametric model for short term load forecasting that was applied for GEFCom2012 was reported by Charlton and Singleton [28]. The proposed approach first models load as a function of the temperature and day of the data, then refines the model by using temperature information of combined weather stations, adding day-of-season terms, changing the number of seasons, conducting local average of the results, removing outliers, treating public holidays, and using a smoother temperature forecast.

Hong, Wilson et al. [4] developed a multiple linear regression model for long term load forecasting by augmenting the regression-based short term load forecasting model with a macroeconomic indicator. This hourly based model showed better forecasting accuracy than models based on monthly or daily data. The developed model was then applied to different temperature and economy scenarios to yield long term probabilistic load forecast.

Xie and Hong [25] extended the regression-based short term load forecasting model in [1] by introducing a two-stage process with the second stage modeling the residuals from the regression based model. Similar to what was done in [4], multiple temperature scenarios were then applied to the two-stage model for generating probabilistic load forecasting.

### 2.2.2 Artificial Neural Networks

The use of artificial neural networks (ANN) for load forecasting goes back to the early 1990s [29], [30]. Depending on the number of output nodes, the ANN-based load forecasting system can be classified into two groups. The first group is the ones having a single output node, which are usually used to forecast the load of next hour or the peak or total load of next day. The load of next hour and the peak and total load of next day were forecasted using a three-layer feed-forward ANN with temperature information and lag load as inputs in [29]. A partially connected ANN was proposed in [31] to forecast the hourly loads for next week. The proposed method was able to more accurately capture the impact of temperature, day of week, hour of day, and load information on the future load consumption within a shorter training time. An ANN-based approach with considerations of accurate temperature modeling and accurate modeling of special events was implemented in the Pacific Gas & Electric company to forecast the peak and hourly loads [32].

The second group includes the ones having multiple output nodes, which are usually used to forecast a sequence of hourly loads such as the load of each of the next 24 hours. Lee et al. [30] divided a given day into three periods and proposed an ANN-based model with three output nodes to forecast each of the three periods. A notable study on ANN-based approach was reported in a series of publications by Khotanzad et al. in the late 1990s [33]–[35]. The proposed method has multiple output nodes to forecast for each of the 24 hours of a day. It showed improved forecasting accuracy and has been implemented by many utilities in their load forecasting practices.

### 2.2.3 Others

Other than the aforementioned techniques, a larger variety of statistical and artificial intelligence techniques have also been developed for load forecasting:

**Similar-day approach** searches the historical data for day(s) with similar characteristics as the forecasted day. The characteristics being considered include weather information such as temperature, humidity, etc., and day type information such as month of year, day of week, etc. The load of a single similar day or a combination (usually simple or weighted average) of the load of several similar days will be used as the forecasted load. The pure similar-day approach has been either replaced by the dynamic models, such as time series models [36], or has been used in combination with other methods [37].

**Time series approach** assumes that the time series data have some characteristics such as autocorrelation, trend, or seasonal variation so that this approach can be used to detect and model the characteristics. Discussions on applying time series approach for load forecasting can be traced back to 1980s [38]. Taylor discussed a series of seasonal exponential smoothing models for short term load forecasting [39]–[41]. A recent book by Weron [5] offered detailed discussions of applying time series approach for load forecasting and provides corresponding case studies for utilities' implementations.

**Support vector machines (SVM)** performs linear or nonlinear classification to output an optimal hyperplane that categories data points. Compared to traditional least-square regression technique, it can map input data into higher-dimensional spaces and consider both of training loss and regularization term in its objective function. A SVM based approach which is a time series based, winter data only, and without temperature information model became the winning entry in the EUNITE mid-term load forecasting competition [42]. Other applications of the SVM for load forecasting were also reported by more recent literature [43], [44].

**Expert-based systems** were usually built based on the rules and procedures that are defined by human experts in the load forecasting field. Ho et al. [45] proposed a knowledge-based expert system for short term load forecasting of Taiwan power system. Kandil et al. [12], [46] proposed a knowledge-based expert system for long term load forecasting for fast developing utility to tackle the problem of fast changing consumption patterns. However, the application of expert-based systems for load forecasting did not go very far due to their high dependence on the inputs from operators.

**Fuzzy logic** was first investigated to tackle the problem of the high dependence on the inputs from the operator in expert-based load forecasting systems. Ranaweera et al. [47] investigated the fuzzy logic model for short term load forecasting by obtaining fuzzy rules from historical data using a learning algorithm. The inputs of the model were selected based on a combination of engineering judgments and statistical analysis. A recent study focusing on the practical application of fuzzy regression on short term load forecasting was provided by Hong and Wang [48].

## 2.3 Explanatory Variables

### 2.3.1 Weather

Due to the extensive use of electric equipment and appliances, human activities and weather are the main driving factors of electricity demand [49]. A significant part of the electricity consumption is to keep the environment at people's comfort level, which is primarily driven by temperature and humidity.

Temperature, or more specifically, dry bulb temperature, is the most deeply rooted weather variable in the load forecasting literature. In summer, load increases as temperature increases, in response to cooling needs. In winter, load increases as temperature decreases to meet heating needs. Various functional forms of temperature have been used to model the relationship between load and temperature, such as piecewise linear models [44], second order polynomials [50], third order polynomials [1] and high order regression splines [51]. Furthermore, there are various ways to include temperature information, such as temperatures of the current and preceding hours [1], [44], daily maximum (or minimum) temperature [52], average temperature of a defined period [22], [53], and cooling (or heating) degree days [53]. Other than dry bulb temperature, dew point temperature and wet bulb temperature have also been studied in some load forecasting models [54] [55].

Although humidity information can be directly used in load forecasting models [18], [55], it is usually embedded in Heat Index (HI) or Temperature-Humidity Index (THI). For instance, the effect of humidity is considered in load forecasting models for late spring, summer and early autumn in [56], where THI is used to replace temperature in the model during the forecasted period of April to September when the temperature was between 76°F and 91°F. Relative humidity is used to calculate HI in [57] for load forecasting models in a case study in Taiwan. It was found in [58] that including Relative Humidity (RH) in the model can improve monthly load forecast accuracy during the summer months in the UK. Overall, the load forecasting literature on humidity variables is far less common and thorough than that on temperature variables until a recent formal and systematic investigation to identify better humidity variables than HI for load forecasting models [21], where a data-driven approach was proposed to select proper humidity variables for load forecasting.

The use of weather variable(s) for load forecasting depends on the availability of the weather history and forecasts, the meteorological condition of the service territory and/or the season being forecasted. For example, relative humidity is not as predictable as temperature, so although adding humidity information helps with an ex post forecast accuracy, it may not improve the ex ante one [21]. This is one of the reasons that relative humidity is not as popular as temperature for load forecasting. This dissertation studies the temperature scenarios for probabilistic load forecasting and proposes a framework to evaluate temperature scenarios generation techniques, while other weather variables can also be studied following a similar framework.

### 2.3.2 Calendar

The electricity consumption behavior is also largely driven by calendar variables such as month of year, day of week, hour of day, and holidays or special events. Grouping the twelve months into seasons and/or transition seasons such as four seasons of a year were reported to be used in load forecasting models [59]–[61], although the definition of season(s) varies from region to region depending on the local climate. To better capture the distinguish the unique of the transition period, the 12 calendar months were also reported being used for load forecasting [1].

The load consumption on different days of a week may also vary due to the fact that patterns of human behavior are different from one day of week to another. Load forecasting models can be built based on the seven days of a week [34]. But, in most cases, the load consumption pattern during some days of a week could be similar, many literature has reported various methods to group similar days of a week for reducing the complicity level of a load forecasting model [1], [47], [59], [61], [62].

Hour of a day is another driving calendar factor of load consumption. For example, the consumption during night may be more stable than that during the day due to the fact that most of the various human activities occur during the daytime. Many literature reported use hour of day information in their load forecasting model either by grouping hours with similar consumption patterns [63], by using the 24 hours of a day directly [1] [34], or by developing unique forecast models for each hour [18].

In addition, special days such as holidays can significantly impact the electricity consumption patterns. In most cases, holidays were modeled simply as different weekday groups or as weekend days. Song et al. [62] applied fuzzy linear regression on holiday load forecasting, which grouped holidays based on the weekday they fell on. Holidays were treated as weekends in [1], [64], [65], although different modeling techniques were applied. In some other cases, researchers followed more complicated modeling steps for modeling holidays. For example, Khotanzad and Afkhami-Rohani [34] used a two-stage method for holiday electric demand forecasting. In the first stage, the holiday electric demand was forecasted according to the weekday it fell on and the peak load of the holiday was forecasted by using demand series with a similar temperature profile. In the second stage, the holiday electric demand forecasts were reshaped based on the peak load forecasts. Xie and Hong [25] also followed a two-stage approach for electric demand forecasting including holidays, where the second stage modeled the residuals. A recent study by Xie and Chien [24] compared several different techniques to model holiday load through a case study of ISO New England data. It concluded that the selection of holiday modeling technique not only depends on the unique holiday load consumption pattern but also depends on the availability of historical data for modeling.

### 2.3.3 Economy

The general health of the economy has traditionally driven long term electricity consumption, although the relationship between the economy and electric load growth has been changing across much of the industry. In practice, depending upon the drivers of the load, economy indicators such as Gross State Product (GSP), housing stock, employment rate, number of jobs and their combinations are often used for middle/long term load forecasting [1], [4]. For retail electricity providers who provide services in a deregulated environment, the total loads are highly impacted by customer churn [27]. In such cases, customer count is often used as the macroeconomic indicator. In this dissertation, GSP of the utility's service territory is used as the macroeconomic indicator for the first case study. It is selected for two reasons: (1) the service territory of the utility of the case study covers most of a specific state, which likely makes GSP a good indicator of the long term load, and (2) GSP is easy to access and understand. If the utility's territory covers one or a few counties or cities, GDP (Gross Domestic Product) by county or GMP (Gross Metropolitan Product) can be used as the macroeconomic indicator. The second case study uses anonymous data without any economy indicator available, so no economy indicator but a linear trend variable is used.

There are mainly three ways to use GSP in a load forecasting model [4]:

(1) Using GSP as a trend. There is an inherent assumption in this approach: the load's sensitivity to weather and calendar remain constant in the profile over time, while there is part of a base load that grows (or declines) linearly in proportion to the economic growth (or decline). If the forecasting horizon is within a few years, this approach can be a good approximation in practice. As the horizon becomes longer, there can be significant changes

in the number of customers. Consequently, the weather and calendar sensitive loads should grow as well.

(2) Dividing load by GSP. The inherent assumption for this approach is that the load is growing (or declining) at exactly the same rate as the economic growth (or decline). In other words, there is no base load that stays constant while the economy is changing. Take a residential community as a counterexample. Before everyone moves in, the feeders, transformers and street lights are already placed in the community, which leads to a small base load, including no-load loss of transformers, street lighting load, etc. As people are moving in during the next a few years, the total load of this system is growing. However, the small base load stays almost the same since day one. Several ways to extend this approach are to take the natural log or square root of the load or macroeconomic indicator, or both in some combination before performing the division, which allows load to grow faster or slower than the economy.

(3) Using GSP as a trend and interacting GSP with other main and cross effects in the model. This approach assumes end-users' behavior changes as the economic environment changes. Since a significant amount of variables are being added through the additional interaction effects, the resulting model may be over-parameterized. Depending upon the forecasting horizon and the electricity usage pattern, this approach may not provide forecast results that are as accurate as the first two options.

In this dissertation, GSP is used as a trend in the load forecasting model because the forecasting horizon being used is relatively short (i.e. one year) where the customer profile is pretty stable overtime.

## 2.4 Probabilistic Load Forecasting Methods

Probabilistic load forecasts can be generated from one of the three components or a combination of any two or three of them [2]: 1) simulating predictors to generate multiple input scenarios [4], [8], [26], [66]–[71], 2) employing probabilistic forecasting models, such as quantile regression models [72]–[74], and 3) converting point load forecasts to probabilistic forecasts through residual simulation[26], [71], or forecast combination [72].

Among these various approaches to generating probabilistic load forecasts, simulating predictors, particularly temperature scenario generation, is being commonly accepted in practice for its simplicity and interpretability. Although simulating joint temperature and economy scenarios has been reported by Hong et al. [4]. Many different temperature scenario generation methods have been reported in the probabilistic load forecasting literature. In the order from simple to sophisticated, these methods can be categorized into four groups:

1) The fixed-date method picks the temperature profile of a past year and assigns the historical temperatures date by date to a future year to obtain a scenario based forecast. The probabilistic forecast is from $k$ scenarios with equal probability, where $k$ is the total number of years for the temperature history. The method was introduced in [4] and then implemented in [25], [26], [72].

2) The shifted-date method picks the temperature profile of a past year, shifts the profile forward and backward by one or more days, and then assigns each shifted profile date by date to a future year to obtain a scenario based forecast. The probabilistic forecast is formed by $(2n+1)k$ scenarios with equal probability, where $n$ is the number of days the original temperature profile being shifted in each direction. This method has been used by

PJM in their long term load forecasting practices for many years. For example, it's reported that PJM used the weather data from year 1973 to 2013 (i.e. $k$ = 41) to generate 533 scenarios (i.e. $n$ = 6) in its 2015 load forecasting [8].

3) The bootstrap method first segments the temperature profile of each historical year into equal length of time blocks, and then randomly picks the blocks with replacement from any of the historical years to form a new temperature profile. This method was adopted by Australia Electric Market Operator (AEMO) for long term peak load forecasting [66].

4) The surrogate method simulates new temperature series through shuffling and taking the Fourier transform of the original time series to maintain the distribution and autocorrelation of the original temperature series. It was first proposed in [71] for one year ahead load forecasting.

Among these four methods, the first three are practical and widely used in the industry. Nevertheless, the methodological foundation for practicing them is not yet solid, which typically result in ad-hoc, judgmental and sometimes hard to be defended choices during the scenario generation step. For instance, it has never been clear to the industry how many years of weather history would be sufficient to adopt any of these methods. This dissertation will utilize a quantile score to evaluate the first three temperature scenario simulation methods and try to propose a guideline for temperature scenario generation practices.

Probabilistic forecasting techniques can be categorized into two groups based on whether their original purpose was for point forecasting or probabilistic forecasting. For example, Hyndman et al. [75] extend the exponential smoothing methods to enable the computation of prediction intervals. While models such as quantile regression [72], Bayesian models [76] were originally developed for probabilistic forecasting. None of these probabilistic forecasting techniques have been extensively studied for probabilistic load forecasting. Some of the reasons could be compared to point forecasting techniques, these techniques are more computationally intensive and the accuracy of them could be hard to evaluate due to the lack of proper evaluation criteria. This dissertation will not focus on applying probabilistic forecasting techniques but rather to focus on developing probabilistic forecasts from applying point forecasting techniques and scenario generation / residual simulation methods.

For the residual simulation method, in fact, studying residual series itself is not anything new in load forecasting and its utility applications. Back to 1970s, for example, researchers were using mean and standard deviation to characterize uncertainties around electric load forecasts for probabilistic load flow analysis [77]–[79]. However, most papers in the literature that modeled load forecast residuals assumed normality for the residual distribution. In other words, normal distributions were used to model the residuals. Such normality assumption, however, has rarely been verified through any formal statistical test. This dissertation will report the comprehensive study on whether simulating residuals with the normality assumption improves the original probabilistic forecasts or not [26].

# CHAPTER 3: THEORETICAL BACKGROUND

Among the various load forecasting techniques discussed in Section 2.2, multiple linear regression is one of the oldest yet most widely applied forecasting techniques and artificial neural network is one of the most popular techniques for load forecasting since the 1990s. In this chapter, the theoretical background of these two most common load forecasting techniques will be introduced. Later on, they will be implemented for the probabilistic load forecasting case studies. The goal of this chapter is not to provide comprehensive introductions to these techniques but to cover the most relevant information to build and evaluate the load forecasting methods in this dissertation.

## 3.1 Multiple Linear Regression

Multiple linear regression (MLR) has been widely used in the forecasting field including electric load forecasting. A comprehensive introduction to MLR can be found in [80]. A general matrix form of the MLR model can be defined as (1) where $Y$ is a $n\times 1$ column vector of observations on the response variable, $X$ is a $n\times (p+1)$ matrix representing the one column of ones and $p$ columns of the observations on the explanatory variables, $\beta$ is a $(p+1)$ row vector of model parameters to be estimated and $\varepsilon$ is a $n\times 1$ column vector of random errors.

$$Y = X\beta + \varepsilon \tag{1}$$

In this dissertation, the response variable is the hourly electric demand, the explanatory variables include quantitative variables, qualitative variables, and their interactions. An example of the quantitative variable is the hourly temperature: in the summer, as the temperature increases, the electric demand may also increase; in the winter, as the temperature increase, the electric demand may decrease. However, this temperature and load relationship may not be linear but nonlinear/curvilinear instead. As a result, the polynomial of the independent variables may be used. Qualitative variables such as month, weekday or hour of day can also be used as independent variables by introducing dummy variables into the model. For example, if month is used as an independent variable in the model, then each of the 12 months will have a dummy variable to represent whether the electric demand falls into that month as shown in (2):

$$\begin{bmatrix} X_{1i} \\ X_{2i} \\ \ldots \\ X_{12i} \end{bmatrix} = \begin{bmatrix} \begin{cases} 1 \ if \ \ January \\ 0 \ else \end{cases} \\ \begin{cases} 1 \ if \ \ Feburary \\ 0 \ else \end{cases} \\ \ldots \\ \begin{cases} 1 \ if \ \ December \\ 0 \ else \end{cases} \end{bmatrix} \tag{2}$$

The interactions between variables should be introduced into the MRL model when the impact of one explanatory variable depends on the level of another explanatory variable. For example, Figure 2 shows the different impact of the temperature on the electric demand during different time of a day. In this case, the 24 hours of a day are categorized into four groups as morning, afternoon, evening, and night. The demand during the morning and the night time is relatively lower than that during the afternoon and the evening when people have more activities. Also, the variation of the demand during the night is smaller than that during the rest of the day. This indicates the impact of temperature on load depends on the level of temperature itself as well as the time of a day. The interaction of two explanatory variables can be represented by multiplying these two variables to generate a new explanatory variable. In this case represented by Figure 2, the quantitative variable temperature can be multiplied by the qualitative variable hour of a day to reflect the combined effect.

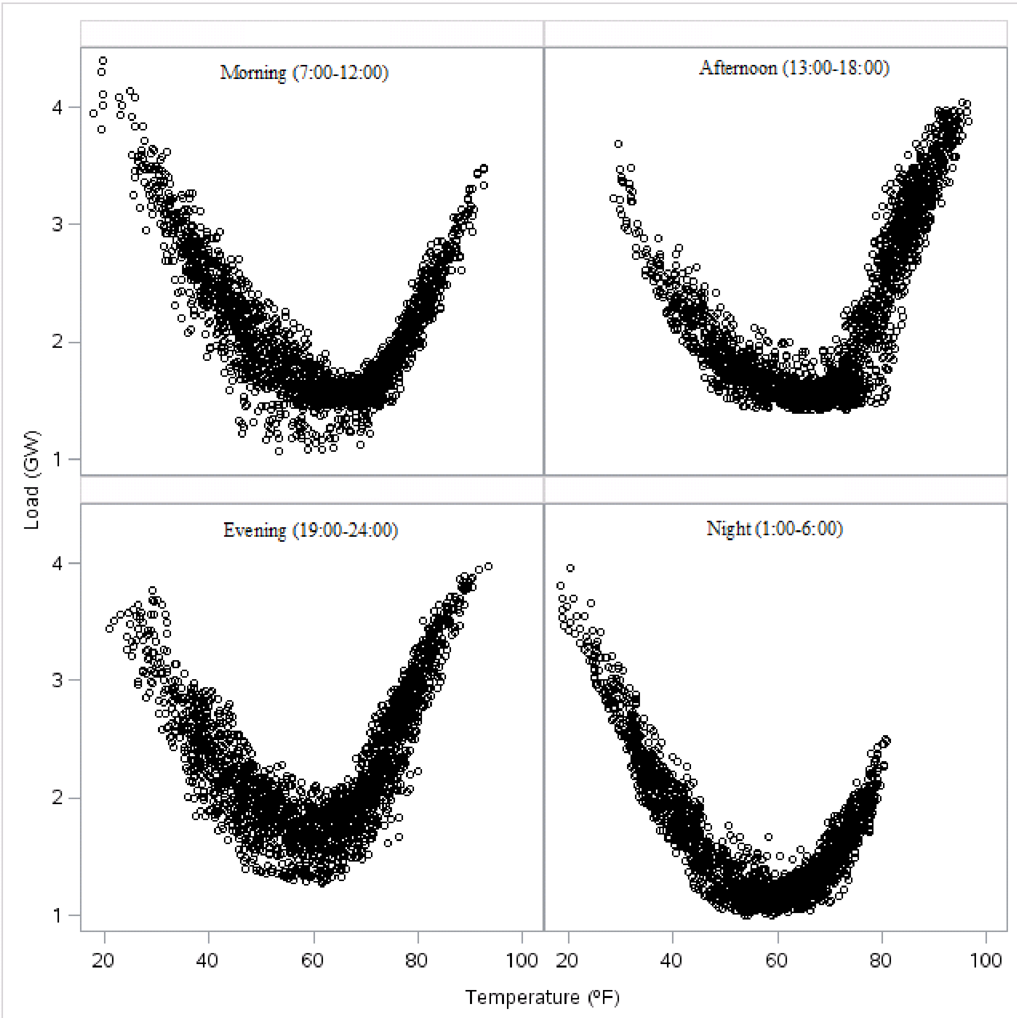



Figure 2: Scatter plot of hourly load and temperature during different periods of a day

## 3.2 Artificial Neural Networks

Artificial neural networks (ANN) models were originally inspired by the biological neural networks, especially brain, to estimate the unknown relationships between the output and a large number of inputs. A comprehensive introduction of ANN is provided in [81], [82]. The types of ANN could vary from single- or multiple-layer single direction logic to complex multi-directional loops, with the later one mainly used for image processing and language recognition. Figure 3 demonstrates the structure of a single-layer feeding-forward ANN which is also the type of ANN that will be implemented in this dissertation. It consists of three parts: (1) the input nodes are the explanatory variables (2) the neurons receive information from the weighted combinations of the input nodes and process them to generate the response. In this study, the neurons are organized in the single hidden layer. The number of neurons does not have to be the same as the number of input nodes. A rule of thumb in practice usually suggests having the number of neurons range between 0.5 to 1.5 times of the number of input nodes. When the neurons process the information, a non-linear activation function is used which makes ANN models able to handle the non-linear relationship between input and output variables. (3) the output node(s) could be our target variable. In this dissertation, only one output node is used for the ANN, which is hourly electric demand.

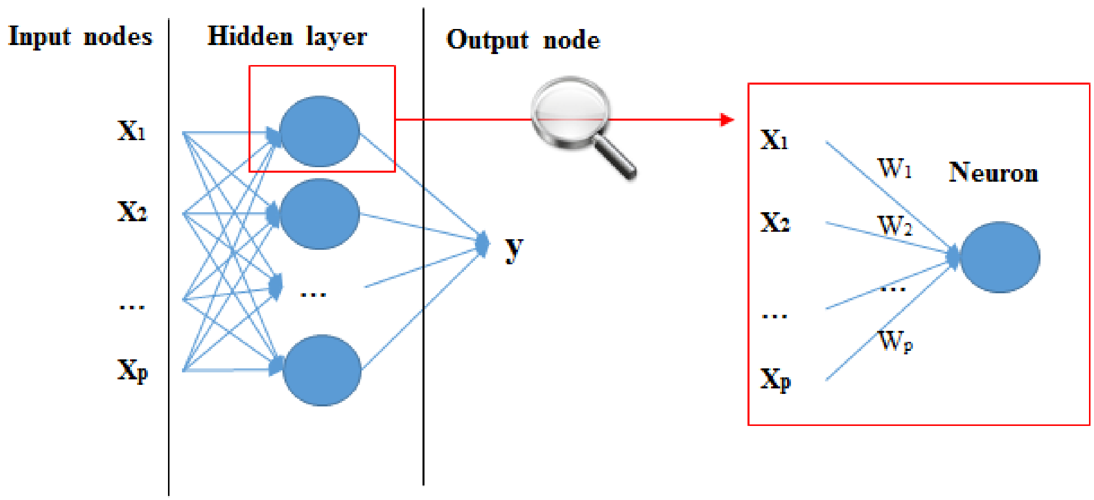



Figure 3: A single-layer feeding-forward neural network

## 3.3 Forecast Evaluation

### 3.3.1 Point Forecast Evaluation

Point forecast evaluation measurements can be categories into the following three groups:

(1) Scale-dependent measures such as root mean square error (RMSE) and mean absolute error (MAE) are error measurements that measure the size of error in units. They can be useful for comparing different methods applied to the same data, but should not be used across different data sets with various scales because they are scale-dependent. In this dissertation, multiple models and data sets with different scales will be compared such that scale-dependent measures will not be useful.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Actual_t - \Pr edict_t)^2} \qquad (3)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}| Actual_t - \Pr edict_t | \qquad (4)$$

(2) Percentage error measures are in the form of $\frac{Actual_t - \Pr edict_t}{Actual_t}$ are scale-free. The most commonly used percentage error measures include mean absolute percentage error (MAPE).

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\frac{| Actual_t - \Pr edict_t |}{| Actual_t |} \qquad (5)$$

The defect of such percentage error measures is that they can be infinite or undefined when the actual value is zero or a very small number such as electric demand at the household level. But, the data being used in this dissertation are all aggregated data at the system or zonal level so this issue is not a concern. In this dissertation, MAPE will be used for evaluating the point forecasts due to its simplicity, interpretability, and being widely adopted by load forecasting practices.

(3) Relative error measures are dividing the error obtained from the proposed method by the one from a standard or benchmark model. However, defining a proper benchmark model for point load forecasting is out of the scope of this dissertation. As a result, relative error measures will not be used for this dissertation

A series of MAPE values could be of interest to a load forecaster:

(1) Hourly load of a specific period which is usually a year depending on the availability of the data and the purpose of the forecasting.

(2) Daily, monthly or annual energy which can be calculated by aggregating the hourly load to daily, monthly or annual level.

(3) Daily, monthly or annual peak load which is the maximum load of a day, a month or a year.

In this dissertation, the hourly MAPE is used to evaluate the load forecasting model because the probabilistic forecasts will also be evaluated for hourly forecasts.

### 3.3.2 Probabilistic Forecast Evaluation

The most commonly used attributes for probabilistic forecast evaluation are reliability, sharpness, and resolution. They were first covered by Pinson, et al. [83] for evaluating probabilistic wind power forecasting. Hong and Fan [2] then extended the discussion on these three attributes for selecting probabilistic load forecasting measures. Reliability evaluates how close the predicted distribution is to the actual one. For example, the 50% prediction intervals should cover 50% of the actual observations. Reliability describes the unconditional coverage of the predicted distribution. Sharpness measures how tightly the predicted distribution covers the actual one. For instance, in the case that two 50% prediction intervals that both cover the 50% of the actual observations but with different width in the boundaries. We will say the one with the narrower boundary width is better regarding the sharpness criteria. Resolution measures how much the predicted distribution varies over time. For example, we will expect the prediction interval for residential load forecast is wider during the daytime when people are more active while expect it to be narrower during the nighttime when the weather condition is smoother and people are less active. If the prediction interval does not vary overtime, we will say it has no resolution. Sharpness and resolution together describe the independence of the target variable (i.e. load) on other factors (such as weather and human activities). These three criteria together describe the conditional coverage of the prediction interval.

There exist several probabilistic forecasting criteria and each of them may address one or several of the aforementioned three evaluation criteria.

(1) Ranked probability score (RPS) measures the mean square error of differences between the cumulative probabilities of the forecasts and that of the observed one. The

continuous ranked probability score (CRPS) is an extension of RPS for continuous probability forecasts (e.g. density function) which measures the integrated squared difference between the cumulative distribution function of the forecasts and that of the observed one. An extension of the RPS and CRPS is the ranked probability skill score (RPSS) or continuous ranked probability skill score (CRPSS). RPSS and CRPSS relate the RPS or CRPS of the forecasts to that of a reference forecast. It can be difficult to establish a reference forecast while one of the most commonly used reference forecasts is the persistent forecast. The RPSS / CRPSS ranges from -∞ to 1 with a negative number indicating the forecasts is worse in accuracy than the reference one. The more negative the number is the worse the forecast to be.

(2) Kolmogorov-Smirnov (KS) statistic is the simplest measure of the unconditional coverage (i.e. the reliability criteria) of the forecasted distribution. It measures the maximum vertical distance between the cumulative distribution functions of the two samples to decide whether they belong to the same distribution. The smaller the number indicates a better forecasted distribution. KS statistic has been used by Magnano and Boland [84] to compare the real and simulated half-hourly load profiles. The drawback of this measure is it is only based on the maximum vertical distance of two cumulative distribution functions but overlooks the distance between them.

(3) Pinball loss function is an error measure for quantile forecasts that penalizes for observations lying far from a given quantile. The use of pinball loss can be tracked back to 1970s when Koenker and Bassett [85] used it as the loss function for estimating the parameter of the regression quantiles. Let $y_t$ be the actual load of time $t$, $\hat{y}_{t,q}$ be the

$q^{th}$ ($q$ = 1, 2, … , 99) quantile of the predicted load of time $t$ and let $p = {}^{q}\!/_{100}$. The pinball loss for the $q^{\text{th}}$ quantile can be defined as

$$Pinball(\hat{y}_{t,q}, y_t, q) = \begin{cases} (1-p)(\hat{y}_{t,q} - y_t) & y_t < \hat{y}_{t,q} \\ p(y_t - \hat{y}_{t,q}) & y_t \geq \hat{y}_{t,q} \end{cases} \tag{6}$$

The mean of the pinball losses across all the quantiles and all the forecasted hours is then taken as the quantile score for the probabilistic forecast. A lower quantile score indicates a better probabilistic forecast. Quantile score considers both of the sharpness and resolution in the evaluation by assessing the forecast at each quantile and has been used for evaluating probabilistic load forecasting since Global Energy Forecasting Competition 2014 [3], [23], [25], [26], [72], [86], [87]. In this dissertation, pinball score will be used to evaluate the probabilistic load forecasts in the form of quantiles due to its popularity and compressive regarding the probabilistic forecasting evaluation criteria.

(4) Winkler score [88] allows a joint assessment of the unconditional coverage (i.e. reliability) and interval width (i.e. sharpness) of a probabilistic forecast. For a central (1-$\alpha$) * 100% prediction interval, let $y_t$ be the actual load of time $t$, $L_t$ and $U_t$ are the lower and upper bounds of the prediction interval, $\delta_t = U_t - L_t$ is the interval width. Winkler score is defined as

$$Winkler(y_t, L_t, U_t, \alpha) = \begin{cases} \delta_t, & L_t \leq y_t < U_t \\ \delta_t + 2(L_t - y_t)/\alpha, & y_t < L_t \\ \delta_t + 2(y_t - U_t)/\alpha, & y_t > U_t \end{cases} \tag{7}$$

It penalizes observations lying far from a given prediction interval and rewards a forecast with a narrower prediction interval. A lower Winkler score indicates a better prediction interval. It was firstly used for evaluating probabilistic load forecasts in [72].

But, as demonstrated [72], the quantile score and the Winkler score generally agree with each other.

### 3.3.3 Sliding Simulation

In the forecasting evaluation practice, the data is generally dissected into two parts: (1) the fit period is used to train the model and the error measurements calculated from this period is to measure the goodness-of-fit of the forecasting method, and (2) the test period is used for the out-of-sample evaluation of forecasting accuracy. Forecasters usually assess the accuracy of a forecasting method based on its out-of-sample performance rather than goodness-of-fit due to two main reasons: Firstly, in-sample errors are likely to understate the forecasting errors because the models are calibrated to fit the in-sample data. Secondly, methods selected by best in-sample fit may not best predict out-sample data.

Among the several out-of-sample evaluation techniques, cross validation and sliding simulation are popular due to their simplicity and universality. A survey of cross-validation procedures for model selection was provided by Arlot and Celisse [89] to review the different cross-validation techniques with an emphasis on the model selection theory behind it. A guideline for choosing the best cross-validation procedures according to the problem in mind was proposed at the end of the paper. Cross-validation techniques split data once or several times and conduct out-of-sample evaluation on one out-of-sample piece with the rest as the training data. The average of the forecast performances of the out-of-sample pieces across the several splits is used for model selection to avoid overfitting as well as the risk of one model may perform extremely well out of luck. Based on the data splitting schema, cross-validation techniques can be grouped into two categories: (1) cross-validation with exhaustive data splitting schema such as leave-one-out [90], [91] and leave-*p*-out [92], in which each one or *p* data points are left out for validation with the rest as training; and (2) cross-validation with partial data splitting such as V-fold cross-validation

[91], in which the data is partitioned into *V* subsamples with equal size and each of them is used for validation with the rest as training.

Tashman [93] offered a comprehensive review of out-of-sample tests in which the sliding simulation technique is reviewed in details. In sliding simulation, both of the historical length and the forecast horizon are fixed while the forecast origin rolls forward to conduct several rounds of out-of-sample forecast evaluation. The average of the forecast performance of the several rounds of out-of-sample tests is used for model selection.

In this dissertation, sliding simulation technique is selected for the following reasons among others. First, comparing with cross-validation, sliding simulation is much less computational intensive yet effective for model selection. Second, the electricity demand pattern of several years ago may not be as informative as those from most recent few years. Sliding simulation technique allows the forecaster to limit the forecasting process to uses most recent historical data for model training. Third, when the historical data is sufficient, compared to cross-validation, sliding simulation is more similar to the real forecasting environment in which the forecaster only have data prior to the forecast origin.

# CHAPTER 4: CASE STUDY DATA

## 4.1 North Carolina Electric Membership Corporation

The primary case study in this dissertation is from North Carolina Electric Membership Corporation (NCEMC). NCEMC is one of the largest electric generation cooperatives in the U.S. and is comprised of a family of corporations formed to support 26 of North Carolina's electric distribution cooperatives. These cooperatives provide energy and related services to more than 950 000 households and businesses in 93 of the 100 North Carolina counties. At NCEMC, load forecasts serve as important inputs to the power supply group to support decisions on electricity purchase contracts.

NCEMC data has also been used as the case study data by several load forecasting literature [3], [4], [21], [26]. It includes 9 years (2003-2011) of hourly load at the system level, the Gross State Product of the state of North Carolina of the corresponding years, and 36 years of hourly temperature data from 1976 to 2011, of which the first 30 years of temperature series is used to generate the temperature scenarios for probabilistic load forecasting. Figure 4 shows the hourly load and temperature of first three training years (2003 – 2005).

The most recent six years of the load data (2006 – 2011) are used for the out-of-sample test on a rolling basis with the length of the rolling estimation window fixed at three years as demonstrated in Figure 5. For example, we first forecast the load for 2006 with the data from 2003 – 2005 as the training data. Then, we roll the forecast origin forward to forecast the year 2007 with the data from 2004 – 2006 as the training data. We repeat this process until we generate forecasts for all these six years from 2006 to 2011.

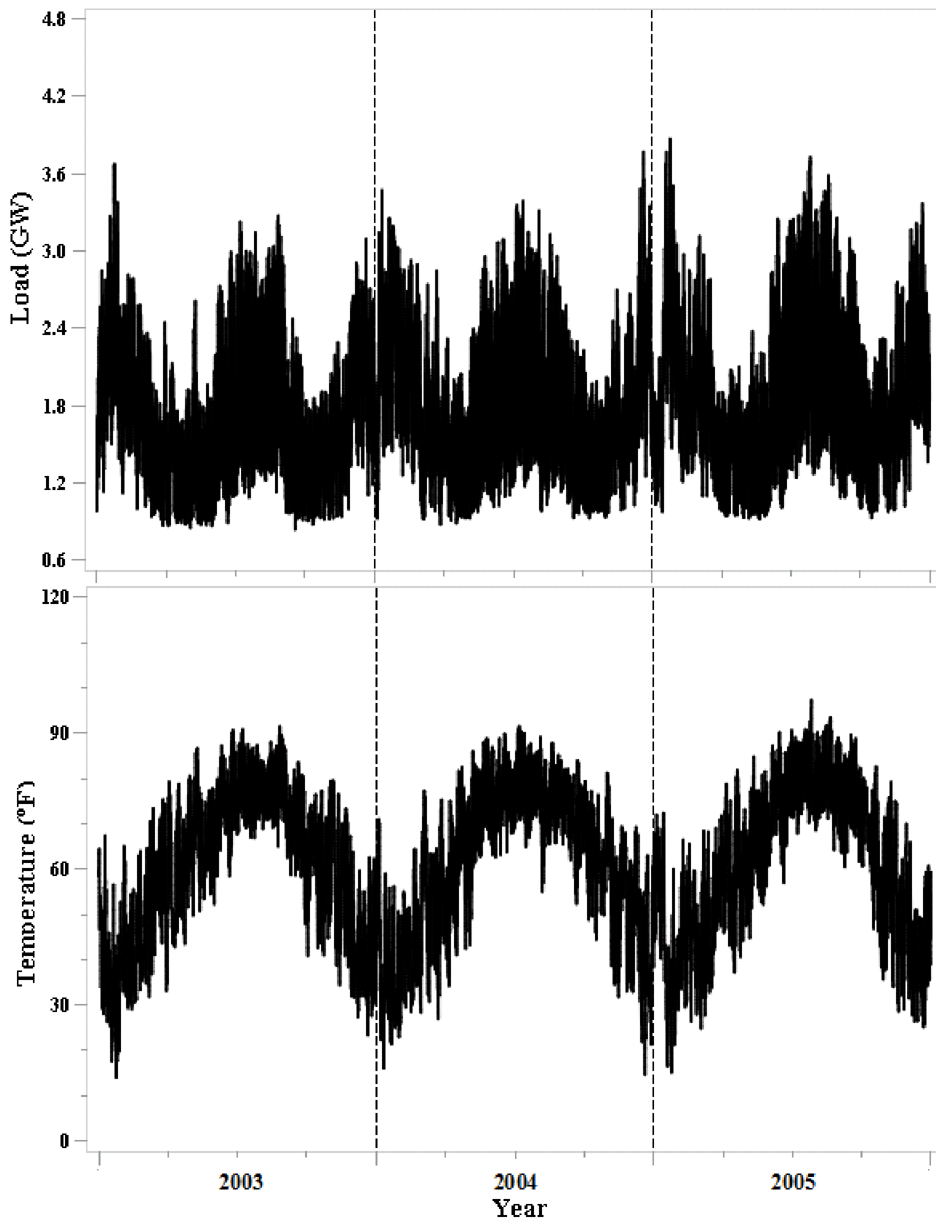



Figure 4: Hourly load and temperature of NCEMC

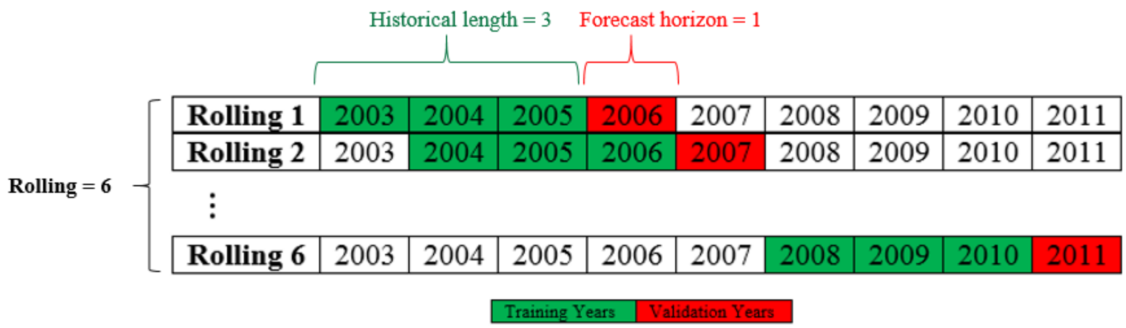



Figure 5: Evaluation using sliding simulation for NCEMC

## 4.2 Global Energy Forecasting Competition 2014

The Global Energy Forecasting Competition 2014 (GEFCom2014) was a probabilistic energy forecasting with four tracks on load, price, wind, and solar forecasting [87]. The competition data has been made public available by the organizer to promote reproducible research and to allow researchers to compare their models and methods using the same datasets. The dataset of the load forecasting track of GEFCom2014 (GEFCom2014-L data) includes seven years of hourly load from 2005 to 2011 and 10 years of hourly temperature data from 2001 to 2010. Due to the limitation in the availability of historical temperature data, only year 2011 will be used for out-of-sample test for the input scenario simulation case study. For the residual simulation case study, since temperature scenarios are not the main focus, the restriction has been relaxed to use three years (i.e. 2009 to 2011) for out-of-sample test. Figure 6 presents the time series plots of hourly load and temperature from 2005 to 2010.

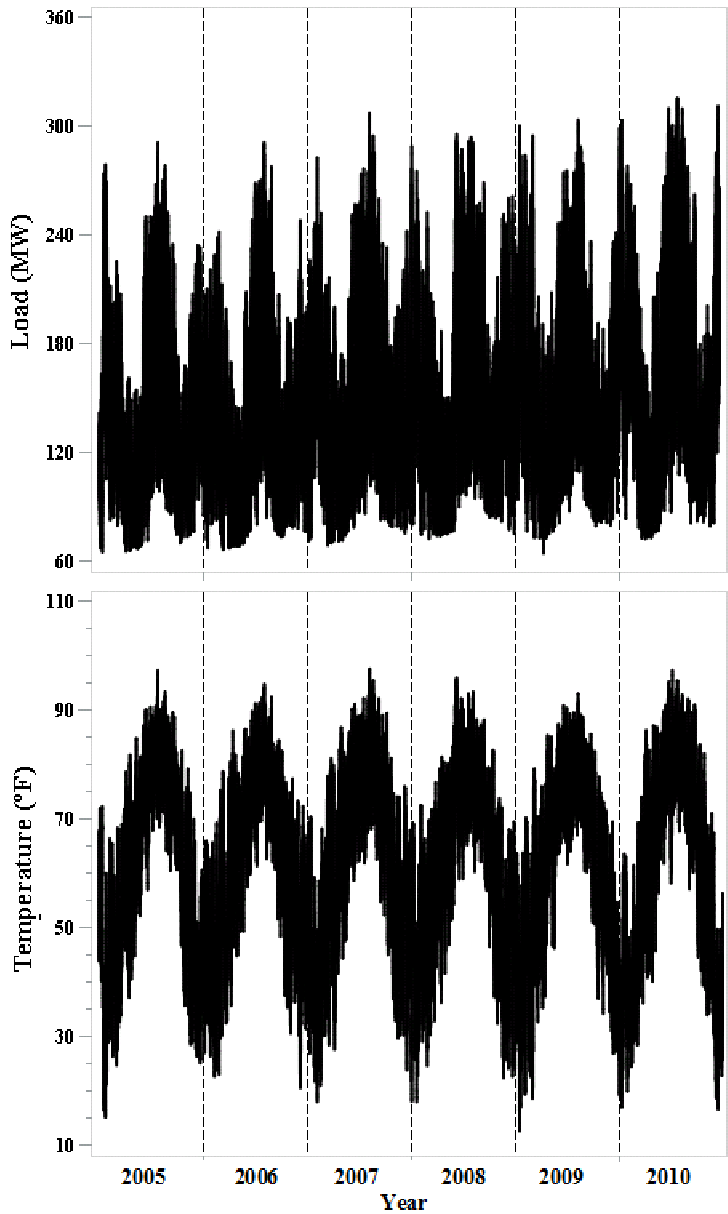



Figure 6: Hourly load and temperate of GEFCom2014-L data

# CHAPTER 5: PROBABILISTIC LOAD FORECASTING

This chapter will discuss two methods to generate probabilistic forecasts: the first one generates probabilistic forecasts by simulating the temperature scenarios for point forecasting models and the second one generates probabilistic forecasts by simulating the residuals from point forecasting models. The chapter will start with introducing the point forecasting models and follow with introducing the two methods.

## 5.1 Point Forecasting Models

### 5.1.1 Linear Regression Model

Three linear regression models with different levels of predictive power are developed for this dissertation. The first model is named as *T-cube* model. It includes *GSP* and the 3rd order polynomials of the current hour temperature as defined in (8) where $GSP_t$ and $T_t$ represent the Gross State Product of the state of North Carolina and the temperature at time *t*, respectively.

$$\hat{y}_t = \beta_0 + \beta_1 GSP_t + \beta_2 T_t + \beta_3 T_t^2 + \beta_4 T_t^3 \tag{8}$$

The second model is Tao's Vanilla Benchmark model, abbreviated as *Vanilla* model in this dissertation. It was first proposed in [1], then used in GEFCom2012 as the benchmark model for the load forecasting track [20], and later reproduced by other scholars [18], [19], [72]. Compared with the *T-cube* model, the *Vanilla* model includes calendar variables (*Month*, *Weekday* and *Hour*) and their interactions with the polynomials of current hour temperature. It is represented by (9)

$$\hat{y}_t = \beta_0 + \beta_1 GSP_t + \beta_2 M_t + \beta_3 W_t + \beta_4 H_t + \beta_5 W_t H_t + \beta_6 T_t + \beta_7 T_t^2 + \beta_8 T_t^3 + \beta_9 T_t M_t + \beta_{10} T_t^2 M_t + \beta_{11} T_t^3 M_t + \beta_{12} T_t H_t + \beta_{13} T_t^2 H_t + \beta_{14} T_t^3 H_t \tag{9}$$

The third model is developed following the model selection process introduced in [1] and implemented in [4], [25]. Compared to the *Vanilla* model, it takes into account of the *Recency Effect* that is meant to capture the impacts of temperature of preceding hours on the current hour's electric demand [1], [22], *Weekend Effect* that is meant to group weekdays with similar electric demand patterns together in order to reduce the complexity of the model [1], and *Holiday Effect* that is used to model holiday as weekend days to address the limited historical data available and the changing of load use patterns issue in forecasting holiday electric demand [1] [24]. This model is named as the *Hong-2014* model in this dissertation and is presented by (10) with the values of $h$ ($h$ = 1, 2, 3) and $d$ ($d$ = 0, 1) varying from year to year depending on the validation dataset that is used for the model selection process.

$$\begin{aligned}&\hat{y}_t = \beta_0 + \beta_1 GSP_t + \beta_2 M_t + \beta_3 W_t + \beta_4 H_t + \beta_5 W_t H_t + \textstyle\sum_d f(\tilde{T}_{t,d}) + \sum_h f(T_{t-h}),\\ &\text{where } f(T_t) = \beta_6 T_t + \beta_7 T_t^2 + \beta_8 T_t^3 + \beta_9 T_t M_t + \beta_{10} T_t^2 M_t + \beta_{11} T_t^3 M_t + \beta_{12} T_t H_t + \beta_{13} T_t^2 H_t + \beta_{14} T_t^3 H_t\\ &\text{and } \tilde{T}_{t,d} = \frac{1}{24}\textstyle\sum_{h=24d-23}^{24d} T_{t-h}\end{aligned} \tag{10}$$

### 5.1.2 Artificial Neural Networks

To further investigate whether the conclusions from this study on probabilistic load forecasting stay the same when the underlying models are nonlinear, this dissertation will be extended to implement three ANN models for the main case study (i.e. NCEMC). The structure of each ANN model is three-layer feed forward. The input variables of each ANN model are assigned based on the independent variables of the corresponding regression model introduced in the previous section with the high order terms and cross effects removed, so that the nonlinear modeling capacity of ANN is used to capture the relationship between load and the input variables.

In details, the first ANN model (*ANN-T-cube*) is comparable to the linear *T-cube* model. It has *GSP* and the current hour temperature as the interval variables. The second ANN model (*ANN-Vanilla*) has all variables from the *ANN-T-cube* model plus *Month*, *Weekday* and *Hour* as the class variables. It is comparable to the linear *Vanilla* model. The third model is comparable to the linear *Hong-2014* model. It has *Month*, *Weekday*, and *Hour* as class variables, temperatures of the current hour and previous three hours, and the average temperatures of the previous 24 hours as interval variables. The polynomials of the temperature terms have been removed to allow the ANN model to identify the nonlinear relationships. The number of hidden neurons is determined via cross validation. Each ANN model is re-diagnosed when forecasting the next year, so the number of hidden neurons may vary from one year to another.

## 5.2 Input Scenario Generation

### 5.2.1 Temperature Scenario based Probabilistic Load Forecasting

For temperature scenario based probabilistic load forecasting, *j* temperature scenarios are first created as the input for the point forecasting models. The techniques to create the temperature scenarios will be discussed in details the following sections. Figure 7 illustrates how to use the generated temperature scenarios to create *j* point load forecasts: by feeding these *j* temperature scenarios into the point load forecasting model, *j* point forecasts can be generated for each hour, where each of the *j* point forecasts comes from one of the *j* temperature scenarios.

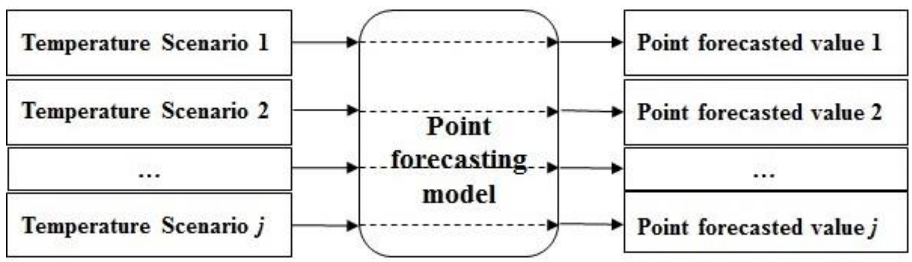



Figure 7: Illustration of generating *j* point load forecasts from *j* temperature scenarios

The temperature scenario based probabilistic load forecast in the form of quantiles is generated from these *j* point forecasts for each hour into the forecast horizon by calculating the 1st to the 99th percentile of the *j* point forecasts. The 1st to the 99th percentile are calculated using the empirical distribution function with averaging. This method will still produce the percentiles in the case that *j* is a small number, say 2. However, when the value of *j* is small, many of the adjacent percentiles may share the same value. In details, let $N$ be the number of non-missing values in the *j* point forecasts at time *t*, and let $\hat{x}_{t,1}, \hat{x}_{t,2}, ..., \hat{x}_{t,N}$ represent the ordered values of them. The $q^{th}$ ($q$ = 1, 2, …, 99) percentile of the predicted load of time *t* is represented by $\hat{y}_{t,q}$. Let $p = q/100$, we define

$$\begin{cases} \text{int} = \text{floor}(N * p) \\ g = N * p - \text{int} \end{cases} \tag{11}$$

The value of $\hat{y}_{t,q}$ is calculated as shown in (12):

$$\hat{y}_{t,q} = \begin{cases} 1/2(\hat{x}_{t,\text{int}} + \hat{x}_{t,\text{int}+1}) & \textit{if} \;\; g = 0 \\ \hat{x}_{t,\text{int}+1} & \textit{if} \;\; g > 0 \end{cases} \tag{12}$$

### 5.2.2 Fixed-date Method

The fixed-date method picks the temperature profile of a past year and assigns the temperatures date by date to a future year to obtain a scenario based forecast. The probabilistic forecast is from $k$ scenarios with equal probability, where $k$ is the total number of years for the temperature history. The method was introduced in [4] and then implemented in [3], [23], [25], [26]. Figure 8 uses the NCEMC case as an example to demonstrate this process: when forecasting the year of 2011, $k$ years of historical temperature series will be assigned to the year 2011 date by date to create k temperature scenarios for the year 2011. When $k$ = 1, only the historical temperature series of year 2005 will be used. When $k$ = 2, the historical temperature of year 2004 and 2005 will be used, and so forth. When $k$ = 30, the historical temperature from 1976 to 2005 will be used.

| Historical Series | | Target Series |
|---|---|---|
| Jan01, 2005 01:00 | -> | Jan01, 2011 01:00 |
| Jan01, 2005 02:00 | -> | Jan01, 2011 02:00 |
| … | | … |
| Dec31, 2005, 24:00 | -> | Dec31, 2011 24:00 |

$k$ =1

| Historical Series | | Target Series |
|---|---|---|
| Jan01, 2004 01:00 | -> | Jan01, 2011 01:00 |
| Jan01, 2004 02:00 | -> | Jan01, 2011 02:00 |
| … | | … |
| Dec31, 2004, 24:00 | -> | Dec31, 2011 24:00 |

$k$ =2

⋮

…

| Historical Series | | Target Series |
|---|---|---|
| Jan01, 1976 01:00 | -> | Jan01, 2011 01:00 |
| Jan01, 1976 02:00 | -> | Jan01, 2011 02:00 |
| … | | … |
| Dec31, 1976, 24:00 | -> | Dec31, 2011 24:00 |

$k$ =30

Figure 8: Illustration of generating $k$ temperature scenarios using the fixed-date method

These $k$ temperature scenarios are then fed into each point forecasting model to create $k$ point forecasts for each hour. The probabilistic load forecasts are then derived from taking the 1$^{st}$ to the 99$^{th}$ percentile of the $k$ point forecasts as introduced in Section 5.2.1.

Figure 9 presents the quantile scores of the Hong-2014 model for each of the years from 2006 to 2011. For each year, the quantile scores show a decreasing trend as the length of temperature history increases. This pattern also appears when the Hong-2014 model is replaced by the other two MLR models. To keep the presentation concise, the average of the quantile scores mainly from the Hong-2014 model will be presented for the rest of this dissertation.

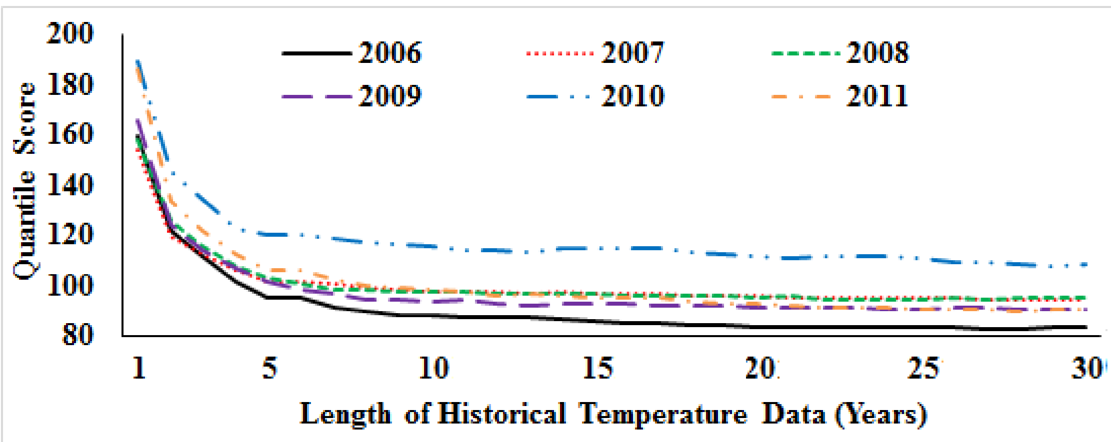



Figure 9: Quantile scores by year
(Case = NCEMC, Model = Hong-2014)

Figure 10 shows the first-order difference of the adjacent average quantile scores from using $k$-1 and $k$ years of historical temperature series. The fact that all bars are above or close to zero further confirms the decreasing trend of average quantile score for each model. Moreover, the decreasing magnitude of the bars in each panel indicates the diminishing improvement as the length of temperature history increases. For all three models regardless of the predictive power, the improvement from increasing length of temperature history is negligible when the history is beyond 15 years or so.

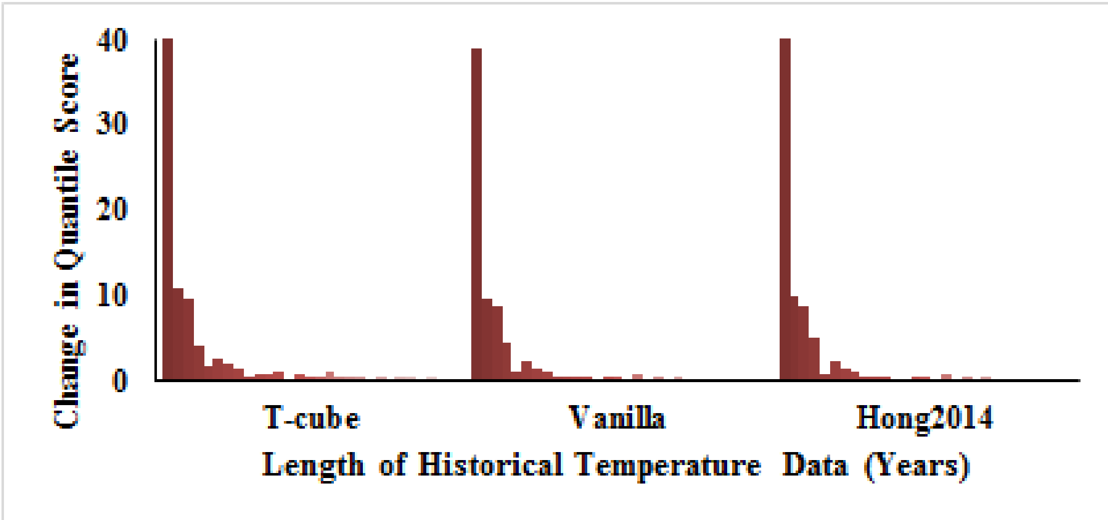



Figure 10: First-order difference of adjacent average quantile scores from using *k*-1 and *k* years of historical temperature series
(Case = NCEMC)

### 5.2.3 Shifted-date Method

The shifted-date method generates more temperature scenarios in additional to those generated from the fixed-date method by shifting the temperature series forward and backward. The rationale behind the shifted-date method is twofold:

1) A season may come in a few weeks sooner or later in different years. For example, the weather of August in 2003 may be similar to the weather of July (or September) in 2006.

2) Electricity demand shows a significant day-of-week pattern. Consequently, if the same temperature profile occurs on a different day of a week, or in the earlier or later of a month than usual, the load profile may be different.

Therefore, instead of fixing the date for a temperature year, it may be worthwhile to create additional temperature scenarios using this shifted-date method. It preserves the autocorrelation of the temperature series naturally while creating additional scenarios to enhance the efficacy of the probabilistic load forecasts. By shifting the temperature series forward and backward by *n* days, this method will result in 2*n*+1 temperature scenarios for each historical year. Suppose we use *k* years of historical temperature data for this simulation process, we will generate with (2*n*+1)*k* temperature scenarios for the load forecasting models. Figure 11 illustrates how to shift a temperature series one-day forward and one-day backward to create additional two temperature scenarios.

| Day of Year | 365 | 1 | 2 | … | 364 | 365 | 1 |
|---|---|---|---|---|---|---|---|
| Original | $T_{D365,Y(i-1)}$ | $T_{D1,Yi}$ | $T_{D2,Yi}$ | … | $T_{D364,Yi}$ | $T_{D365,Yi}$ | $T_{D1,Yi}$ |
| 1-day forward | $T_{D364,Y(i-1)}$ | $T_{D365,Y(i-1)}$ | $T_{D1,Yi}$ | … | $T_{D363,Yi}$ | $T_{D364,Yi}$ | $T_{D365,Yi}$ |
| 1-day backward | $T_{D1,Yi}$ | $T_{D2,Yi}$ | $T_{D3,Yi}$ | … | $T_{D365,Yi}$ | $T_{D1,Y(i+1)}$ | $T_{D2,Y(i+1)}$ |

Figure 11: Generating additional temperature scenarios by shifting dates

The line plot in Figure 12 shows the average quantile scores of the Hong-2014 model for year 2006-2011 by shifting the temperature series from 0 to 6 days forward and backward. There are three observations from it: 1) shifting the temperature series forward and backward to create additional temperature scenarios helps improve the quantile score of the fixed-date method; 2) the improvement is diminishing as the number of shifted days increases and the length of temperature history increases; and 3) the improvement from the shifted-date method is negligible when $n > 4$ and $k > 5$.

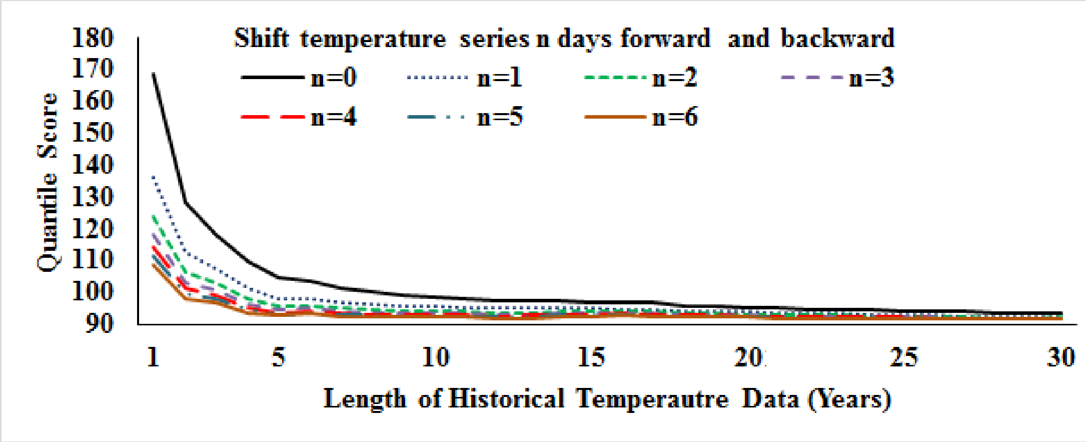



Figure 12: Quantile scores from using temperature series simulated by the shifted-date method (Case = NCEMC, Model = Hong-2014)

Figure 13 presents the actual load (blue dot) of the summer peak week of 2011 and the corresponding quantile forecasts (lines) from the Hong-2014 model. The black dash lines in the plots show the 1st to the 99th quantile forecasts except for the median which is highlighted in red. Each subtitle for the plot is in the form of *k_n* indicating using *k* years of historical temperature series and shifting the temperature series *n* days around to create additional temperature scenarios. For examples, 15_6 represents using 15 years (1991 – 2005) of historical temperature series and shifting the historical temperature series 6 days around to create 180 additional scenarios. We only present the quantile forecasts generated from using 1, 15 and 30 years of historical temperature series and shifting the historical series 0 and 6 days around to avoid verbose presentation. Nevertheless, Figure 13 confirms the findings above. When the historical temperature data is limited, additional temperature scenarios generated from shifting the historical temperature series help improve the efficacy of the probabilistic forecast. As additional years of historical temperature data is being used to generate scenarios, the shifted-date method does not result in significant improvement much further.

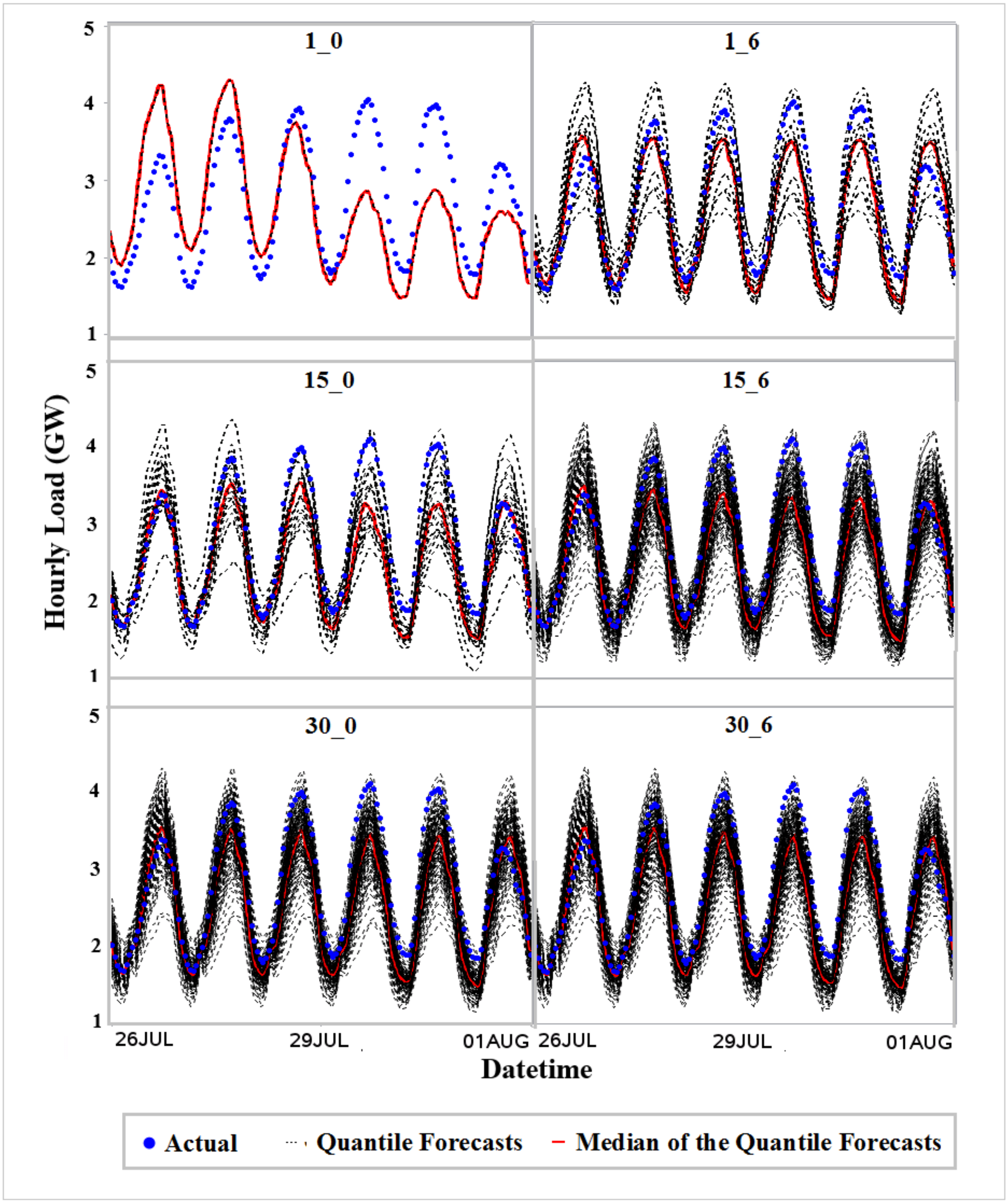



Figure 13: Quantile forecasts of a summer peak week (Jul 26 – Aug 01, 2011)
(Case = NCEMC, Model = Hong-2014)

### 5.2.4 Bootstrap Method

The double season block bootstrap temperature scenario simulation technique first segments each temperature series into blocks with equal length of $m$ days. Let $m = 9$ in this study to be consistent with what was done in [66]. For example, a year with 365 days will have the first 9 days of the year as the first block, the second 9 days of the year as the second block and so on. Each of the first 40 blocks will have equal days (i.e. 9 days), while the last block will only have 5 days. If $k$ years of historical temperature data are used, each of the 41 blocks will be randomly picked from one of the $k$ years to form a new temperature series. Figure 14 illustrates the method with one example: the first block of scenario 1 may come from the first block of year 2001; the second block may come from the second block of year 1973; and so on. If this random pick $l$ times with replacement is done $l$ times, $l$ temperature scenarios will be simulated. In this dissertation, $l$ is set to 100 for demonstration purpose without loss of generality. The cases with $l$ = 500 and 1000 have also been tested, respectively, but only tiny difference after the first decimal place of the quantile scores were observed. Thus, only results from $l$ = 100 are presented in this dissertation to avoid verbose presentation. This simulation process will result in $k$+$l$ temperature scenarios for the PLF.

| | | | | | |
|---|---|---|---|---|---|
| **Scenario 1** | Block1: 2001 | Block 2: 1973 | Block 3: 1998 | … | Block 41: 2004 |
| **Scenario 2** | Block1: 1993 | Block 2: 1997 | Block 3: 2001 | … | Block 41: 1975 |
| ⋮ | | | | | |
| **Scenario *l*** | Block1: 1986 | Block 2: 2004 | Block 3: 1981 | … | Block 41: 1995 |

Figure 14: Generating additional temperature scenarios by bootstrapping

Figure 15 shows the scatter plot of the average quantile scores between the fixed-date method (horizontal axis) and the bootstrap method (vertical axis). The points all lie on or above the diagonal line (i.e. the black line) in this scatter plot, which indicates that the additional temperature scenarios simulated by the bootstrap method does not improve the quantile score. When the lagged temperatures are used in the model (e.g. Hong-2014), the quantile scores from the bootstrap method becomes worse than those of the fixed-date method. This is likely due to the fact that two adjacent blocks that come from different years may create an unrealistic gap on the border.

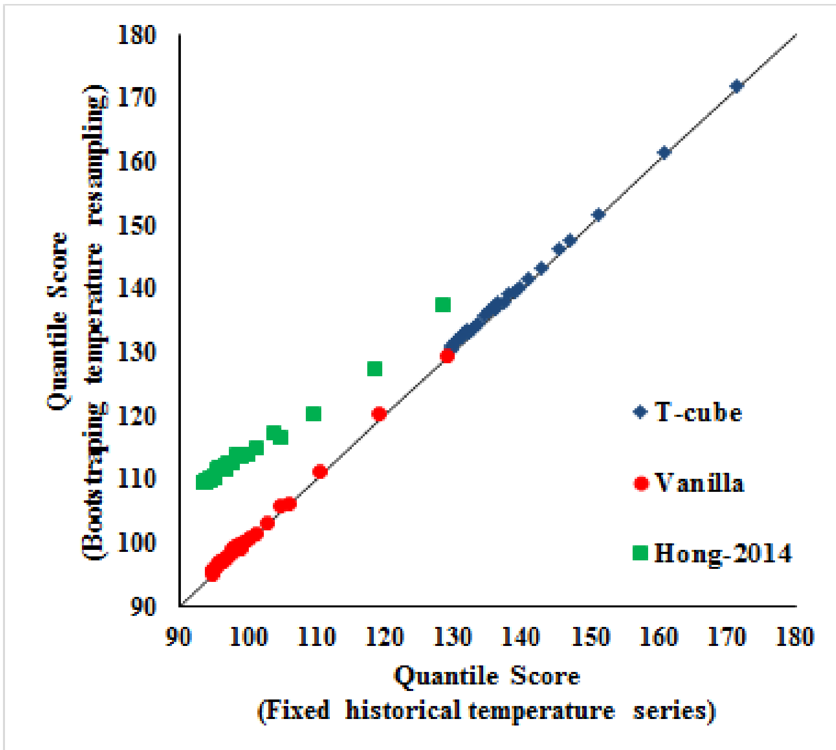



Figure 15: Scatter plot of quantile scores from using fixed and simulated temperature series (Case = NCEMC)

Figure 16 shows an example of the temperature scenario simulated from using 30 years of historical temperature for the summer peak week of 2006 (Aug 01 – Aug 10), with the peak occurring on Aug 4, 2006. The temperature of the first four days were drawn from the corresponding block of the year 1998, while the last six days were drawn from year 1994. This example shows that the simulated temperature series may have significant discontinuity on the border between the two blocks. While this may not have much impact on the peak demand forecasts as it was first proposed in [66], it appears to harm the quantile scores when the forecasts at all the hours are involved. One way to alleviate this issue is to widen each block. As the block size goes up to one year, this bootstrap method becomes identical to the fixed-date method.

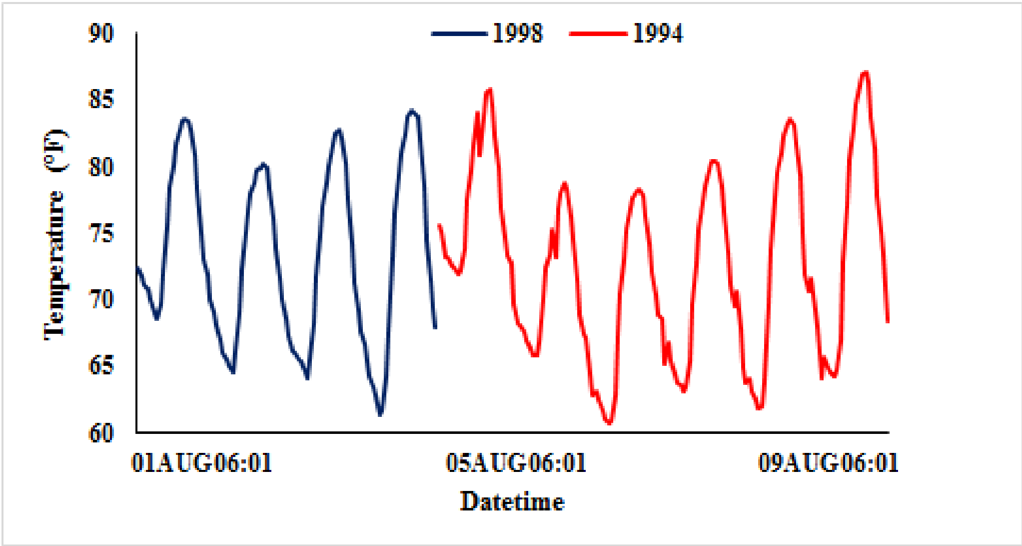



Figure 16: Hourly temperature from the simulated series during the summer peak week of 2006 (Case = NCEMC)

### 5.2.5 Discussion on Empirical Rule-of-thumb

In sum, based on the results from the above three temperature scenario generation methods, the findings are 1) the quantile score of each method shows diminishing improvement as the length of available temperature history increases; 2) comparing with the fixed-date method, the shifted-date method helps improve the quantile score with the improvement diminishing as the number of shifted days increases; and 3) the bootstrap method offers the capability of generating more comprehensive scenarios but does not improve the quantile score of the fixed-date method.

When taking the weather scenario generation approach, one of the first questions is how many years of history are good enough. The variant of this question can be as follows:

1) Have we collected enough historical data to perform probabilistic load forecasting?

2) Do we have enough history so that we can stay with the simplest method, i.e., fixed-date method?

3) With the length of history we have today, which method shall we choose?

Based on the observations made in Section 5.2.4, the bootstrap method can be excluded here to focus on the quantile score improvement. Since the conclusions from the three models are consistent on the fixed-date and shifted-date methods, here we use the Hong-2014 model for illustration purpose.

Figure 17 shows the quantile score from shifting the *k* years of historical temperature n days around. The horizontal axis is the number of days (*n*) being shifted forward and backward. The vertical axis is the quantile score. Each line represents the quantile scores under a certain length of history *k*. As the value of n increases, the quantile score first decreases but will eventually increase. This indicates that shifting the weather history within a range can help improve the quantile score, though shifting too many days can aggravate the quantile score. The rationale is that a season may come in a few weeks sooner or later, but the weather in August should not be treated as typical in November.

When only one year of historical temperature is used, the quantile score is 168.91. By using different combinations of *k* and *n*, the global lowest quantile score we can get is 91.21 by using four years of historical temperature series and shifting it by five days around. Table 1 presents the *n* value to achieve the lowest quantile score when *k* is fixed, and the upper and lower boundary of the value of *n* to achieve at least 95% of the global best improvement (i.e. reducing the quantile score from 168.91 to 95.09). For example, when five years of history is used, we need to shift the five-year historical temperature series at least three days forward and backward to bring the quantile score lower than 95.09. The more years of historical temperature series is used, the smaller the number of shifts that is needed to achieve a desirable quantile score. However, the shift of more than 39 days around will result in a quantile score higher than 95.09. Overall, regardless of the value of *k*, we should be cautious about shifting the historical temperatures series too many days.

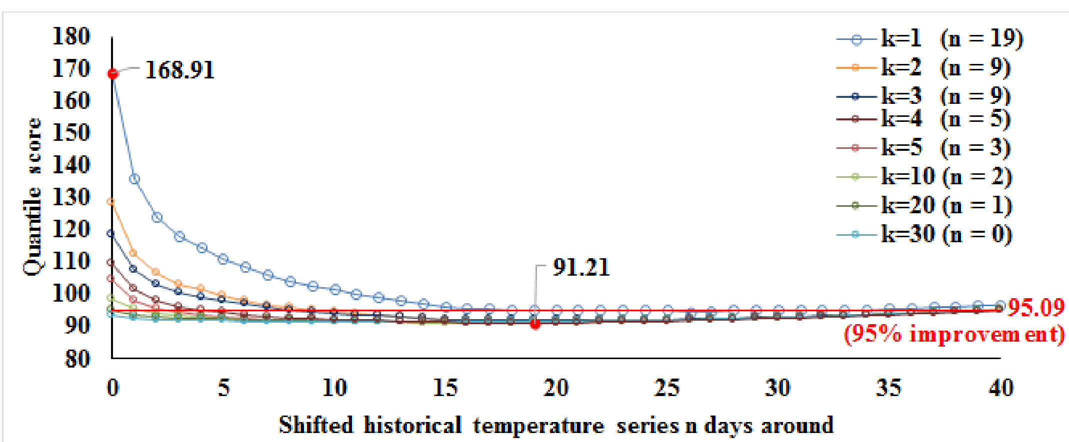



Figure 17: Quantile score of shifting the k years of historical temperature series n days around (Case = NCEMC, Model = Hong-2014)

Table 1: Lower and upper bounds of *n* for different *k*
(Case = NCEMC, Model = *Hong-2014*)

| *k* / *n* | 1 | 2 | 3 | 4 | 5 | 10 | 20 | 30 |
|---|---|---|---|---|---|---|---|---|
| 95%, LO | 19 | 9 | 9 | 5 | 3 | 2 | 1 | 0 |
| Min | 26 | 19 | 19 | 19 | 16 | 17 | 16 | 14 |
| 95%, UP | 32 | 39 | 39 | >40 | 39 | 38 | 39 | 39 |

The following rule of thumb is proposed to offer a practical guide for forecasters to select the effective temperature scenario generation method with the appropriate parameters:

$$k(n+1) > 30 \tag{13}$$

For instance, if 30 or more years of temperature history is available, there would be no significant improvement in quantile score by taking the shifted-date method. If 4 years of temperature history is available, we should shift about 7 days forward and backward to gain most improvement in quantile score.

Following a similar approach presented in this paper, practitioners may come up with different rules of thumb based on the practical considerations. For instance, the threshold may be chosen as 90% or 99%. Alternatively, the rule of thumb may be targeting the optimal value instead of a threshold.

### 5.2.6 Results from Nonlinear Models

Figure 18 presents the average quantile score of the ANN-Hong-2014 model with the temperature scenarios generated from the three aforementioned temperature scenario generation techniques: (1) the fixed-date method, (2) the shifted-date method with n = 6, and (3) then bootstrap method. The results from the other two models are not presented here because they show a similar pattern. The conclusions from using the linear models do not change when we switch to nonlinear models. Shifting historical temperature series around to create additional temperature scenarios helps significantly improve the PLF only when few years of temperature history are used. Additional scenarios generated by bootstrapping resampling do not contribute to the improvement of the quantile score.

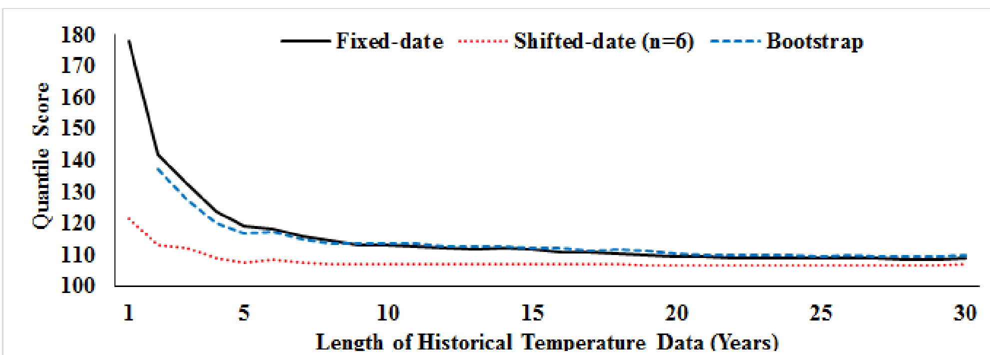



Figure 18: Quantile score from using three different temperature generation techniques (Case = NCEMC, Model = ANN-Hong-2014)

### 5.2.7 Results from GEFCom2014-L Data

Table 2 lists the quantile scores from the three linear models using the three aforementioned temperature scenario generation techniques, namely fixed-date (FD), shifted-date with n =6 (SD) and bootstrap (B). The lowest quantile score for each model group is highlighted in bold. The three conclusions still hold when we extend the study to this new set of data.

Table 2: Quantile scores of year 2011

| | T-cube | | | Vanilla | | | Hong-2014 | | |
|---|---|---|---|---|---|---|---|---|---|
| k | FD | SD | B | FD | SD | B | FD | SD | B |
| 1 | 18.02 | **12.09** | n/a | 15.35 | **10.11** | n/a | 15.49 | **10.02** | n/a |
| 2 | 13.52 | **10.78** | 13.59 | 11.28 | **8.74** | 11.37 | 11.32 | **8.59** | 13.20 |
| 3 | 12.14 | **10.33** | 12.23 | 9.85 | **8.17** | 9.92 | 9.75 | **8.00** | 11.72 |
| 4 | 11.56 | **10.25** | 11.61 | 9.36 | **8.11** | 9.40 | 9.22 | **7.95** | 11.22 |
| 5 | 11.16 | **10.20** | 11.20 | 8.89 | **7.97** | 8.90 | 8.77 | **7.82** | 10.62 |
| 6 | 11.12 | **10.25** | 11.15 | 8.90 | **8.08** | 8.96 | 8.80 | **7.94** | 10.64 |
| 7 | 10.94 | **10.19** | 10.98 | 8.66 | **7.96** | 8.74 | 8.54 | **7.80** | 10.48 |
| 8 | 10.86 | **10.20** | 10.84 | 8.62 | **7.98** | 8.57 | 8.52 | **7.84** | 10.23 |
| 9 | 10.79 | **10.15** | 10.83 | 8.55 | **7.91** | 8.56 | 8.44 | **7.77** | 10.23 |
| 10 | 10.71 | **10.14** | 10.79 | 8.42 | **7.85** | 8.51 | 8.29 | **7.70** | 10.19 |

Table 3 shows the quantile score of the Hong-2014 model for year 2011 with the $k$ years of historical temperature series being shifted $n$ days around. A cooler color in the background of the cell indicates the lower quantile score. For each column, the bolded quantile score is the lowest value for each fixed $k$. The global lowest quantile score is 7.16 that is achieved by using 10 years of historical temperature series and shifting the temperature series 16 days around. The quantile scores we get for each $k$ and $n$ combination by following the proposed rule of thumb are underscored. The 95% threshold in this case is 15.49 – 0.95(15.49 – 7.61) = 8.00. By following the proposed rule of thumb, most of the cases reach the 95% threshold. The only two exceptions are $k$ = 1 and $k$ = 2, where the lowest quantile scores (8.42 and 8.01 respectively) are still larger than the 95% threshold.

Table 3: Quantile scores of year 2011 by shifting the *k* years historical temperature series *n* days around (Model = *Hong-2014)*

| n \ k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 15.49 | 11.32 | 9.75 | 9.22 | 8.77 | 8.80 | 8.54 | 8.52 | 8.44 | 8.29 |
| 1 | 12.71 | 9.92 | 8.79 | 8.53 | 8.26 | 8.38 | 8.16 | 8.20 | 8.14 | 8.03 |
| 2 | 11.57 | 9.33 | 8.42 | 8.27 | 8.07 | 8.22 | 8.02 | 8.06 | 8.01 | 7.91 |
| 3 | 10.99 | 9.08 | 8.30 | 8.19 | 8.01 | 8.17 | 7.96 | 8.00 | 7.94 | 7.85 |
| 4 | 10.58 | 8.92 | 8.21 | 8.13 | 7.96 | 8.11 | 7.92 | 7.95 | 7.88 | 7.80 |
| 5 | 10.25 | 8.75 | 8.09 | 8.04 | 7.89 | 8.02 | 7.86 | 7.89 | 7.81 | 7.75 |
| 6 | 10.02 | 8.59 | 8.00 | 7.95 | 7.82 | 7.94 | 7.80 | 7.84 | 7.77 | 7.70 |
| 7 | 9.85 | 8.48 | 7.93 | 7.88 | 7.77 | 7.88 | 7.76 | 7.80 | 7.73 | 7.67 |
| 8 | 9.71 | 8.42 | 7.88 | 7.84 | 7.73 | 7.84 | 7.73 | 7.76 | 7.70 | 7.64 |
| 9 | 9.57 | 8.35 | 7.85 | 7.83 | 7.71 | 7.81 | 7.71 | 7.75 | 7.68 | 7.63 |
| 10 | 9.47 | 8.30 | 7.82 | 7.81 | 7.70 | 7.80 | 7.71 | 7.74 | 7.68 | 7.63 |
| 11 | 9.36 | 8.26 | 7.80 | 7.80 | 7.69 | 7.79 | 7.70 | 7.73 | 7.68 | 7.63 |
| 12 | 9.26 | 8.21 | 7.77 | 7.78 | 7.68 | 7.77 | 7.69 | 7.71 | 7.67 | 7.62 |
| 13 | 9.17 | 8.18 | 7.75 | 7.76 | **7.68** | 7.75 | 7.67 | 7.69 | 7.66 | 7.61 |
| 14 | 9.11 | 8.16 | **7.75** | **7.76** | 7.68 | **7.74** | 7.66 | 7.68 | 7.65 | 7.61 |
| 15 | 9.05 | 8.17 | 7.76 | 7.76 | 7.69 | 7.74 | 7.66 | 7.67 | 7.64 | 7.61 |
| 16 | 8.99 | 8.16 | 7.77 | 7.77 | 7.70 | 7.75 | 7.65 | 7.66 | 7.64 | 7.61 |
| 17 | 8.95 | 8.15 | 7.77 | 7.77 | 7.71 | 7.75 | **7.65** | 7.65 | **7.64** | 7.61 |
| 18 | 8.90 | 8.13 | 7.77 | 7.77 | 7.71 | 7.75 | 7.65 | **7.65** | 7.64 | **7.61** |
| 19 | 8.85 | 8.11 | 7.77 | 7.77 | 7.72 | 7.76 | 7.66 | 7.66 | 7.65 | 7.61 |
| 20 | 8.80 | 8.09 | 7.77 | 7.78 | 7.73 | 7.76 | 7.66 | 7.66 | 7.65 | 7.62 |
| 21 | 8.77 | 8.08 | 7.77 | 7.79 | 7.74 | 7.76 | 7.67 | 7.68 | 7.66 | 7.63 |
| 22 | 8.74 | 8.08 | 7.78 | 7.80 | 7.75 | 7.77 | 7.68 | 7.69 | 7.68 | 7.64 |
| 23 | 8.71 | 8.07 | 7.80 | 7.81 | 7.76 | 7.78 | 7.69 | 7.70 | 7.69 | 7.66 |
| 24 | 8.69 | 8.07 | 7.82 | 7.83 | 7.77 | 7.79 | 7.71 | 7.71 | 7.70 | 7.67 |
| 25 | 8.66 | 8.06 | 7.84 | 7.84 | 7.79 | 7.80 | 7.72 | 7.73 | 7.72 | 7.69 |
| 26 | 8.63 | 8.05 | 7.85 | 7.85 | 7.80 | 7.81 | 7.74 | 7.74 | 7.73 | 7.71 |
| 27 | 8.60 | 8.04 | 7.86 | 7.86 | 7.81 | 7.83 | 7.75 | 7.76 | 7.75 | 7.72 |
| 28 | 8.56 | 8.02 | 7.87 | 7.88 | 7.83 | 7.84 | 7.77 | 7.77 | 7.76 | 7.74 |
| 29 | 8.53 | 8.01 | 7.88 | 7.89 | 7.84 | 7.85 | 7.78 | 7.79 | 7.78 | 7.76 |
| 30 | 8.50 | 8.01 | 7.89 | 7.90 | 7.86 | 7.87 | 7.80 | 7.80 | 7.79 | 7.78 |
| 31 | 8.48 | **8.01** | 7.90 | 7.91 | 7.87 | 7.88 | 7.81 | 7.81 | 7.81 | 7.79 |
| 37 | **8.42** | 8.03 | 7.99 | 8.00 | 7.95 | 7.97 | 7.91 | 7.91 | 7.91 | 7.91 |

### 5.2.8 Pros and Cons

Table 4 summarizes the pros and cons of the aforementioned temperature scenario generation methods from the following three perspectives: 1) the number of temperature scenarios a method can potentially produce, 2) the complexity involved for implementation, and 3) the quantile score of the resulting probabilistic forecasts. The bootstrap method has the capability of generating very large number of scenarios. The fixed-date method is easiest to implement among the three. The empirical studies in this paper have shown that the shifted-date method can lead to the lowest quantile score among the three. In sum, there is no single method that outperforms all others based on these comparison matrix, in practice, forecasters should choose the one that best suitable for their jurisdiction.

Table 4: Pros and cons of three temperature scenario generation techniques

| | **# of scenarios** | **Complication** | **Quantile score** |
|---|---|---|---|
| **Fixed-date** | Least | Least | Moderate |
| **Shifted-date** | Moderate | Moderate | Lowest |
| **Bootstrap** | Most | Most | Highest |

## 5.3 Residual Simulation

Prediction interval estimates the interval that future observations will fall into with a certain probability. For example, one can estimate that with a 90% possibility, tomorrow's peak load will fall into the range between 120MW and 180MW. Prediction interval is a widely-adopted way to describe the uncertainties in the forecast. A simple and commonly used prediction interval for the estimated value $\hat{y}$ can be calculated as shown in (14), where $\alpha$ is the multiplier (such as 1.96 for a 95% prediction interval) and $\hat{\sigma}$ is the standard deviation of the forecast distribution.

$$\hat{y} \pm \alpha\hat{\sigma} \tag{14}$$

This calculation of prediction interval assumes that the error term or the residuals are normally distributed and uncorrected. The violation of such normality assumption likely makes the confidence interval to be too wide or too narrow and not an ideal estimate of the probability that a given forecast will exceed some threshold in a particular direction. In practices, such normality assumption has rarely been validated through any formal statistical test. This section will start with looking into the distribution of the residuals from the point load forecasting models. Then, the process to simulate the residuals will be introduced for improving the probabilistic load forecast. Although the residuals are simulated from normal distribution, this residual simulation method does not rely on the validity of the normality assumption to improve probabilistic load forecasts.

### 5.3.1 Normal Distribution

Residuals are defined as the difference between the actual value and the forecasted one. The validity of the normality assumption on the residuals from the three representative underlying linear regression models will be first investigated graphically and by the Kolmogorov-Smirnov (K-S) test. The K-S test is a nonparametric test that can be used to compare a sample with a reference probability distribution which is the normal distribution in this dissertation. It quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. The null hypothesis of the K-S test is the sample is drawn from the reference distribution.

Figure 19 shows the histograms of the residuals from the three ex post forecasts, the normal distribution, and the Kolmogorov-Smirnov test results. The blue histograms represent the empirical distributions, while the red lines represent the normal distributions. The axes of the three sub-figures are in the same scale. None of the residual series passes the K-S test based on the significance level of 0.05 with the critical value as 0.0095 [94]. In other words, there is not enough evidence showing that the residuals are normally distributed.

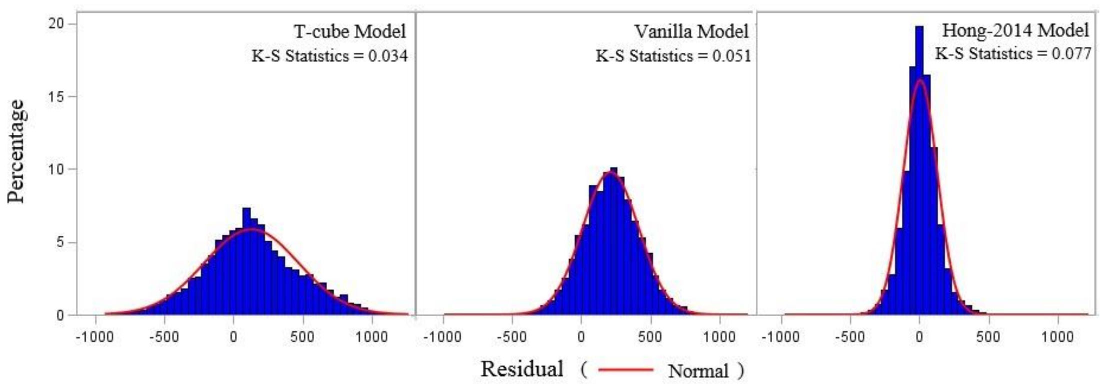



Figure 19: Normality test on ex-post point forecast residuals of linear models

### 5.3.2 Group Analysis

Since electricity demand series has multiple seasonal patterns, the residuals may behave differently by hour of the day, day of the week or month of the year. Instead of performing K-S test on all the residuals as one group, the residuals can be grouped by different calendar variables (month, weekday and hour) and their combinations. Table 5 lists the results of K-S test performed on each group. Using the significance level of 0.05, the results show that performing group analysis on the residuals helps increase the passing rate of K-S test. On average, grouping the residuals by month and hour yields the highest passing rate across all three models.

Table 5: Normality test passing rates (in %) of ex-post point forecast residuals with various grouping methods

| | **Grouping Method** | **Number of Levels** | **2005** | **2006** | **2007** | **2008** | **2009** | **2010** | **AVG.** |
|---|---|---|---|---|---|---|---|---|---|
| *T-cube* | No Grouping | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Weekday | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Month | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Hour | 24 | 33.3 | 25 | 29.2 | 16.7 | 25 | 12.5 | 23.6 |
| | Month + Weekday | 78 | 21.4 | 16.7 | 14.3 | 19.0 | 17.9 | 26.9 | 19.4 |
| | Weekday + Hour | 168 | 69.0 | 62.5 | 46.3 | 57.1 | 64.9 | 67.3 | 61.2 |
| | Month + Hour | 288 | **76.7** | **76.4** | **76.7** | **84.0** | **81.3** | **75.3** | **78.4** |
| *Vanilla* | No Grouping | 1 | 0 | 0 | **100** | 0 | 0 | 0 | 16.7 |
| | Weekday | 7 | 14.3 | 0 | 0 | 0 | 0 | 0 | 2.4 |
| | Month | 12 | 8.3 | 0 | 16.7 | 16.7 | 0 | 0 | 7.0 |
| | Hour | 24 | 29.2 | 8.3 | 25 | 25 | 0 | 8.3 | 16.0 |
| | Month + Weekday | 78 | 40.5 | 36.9 | 51.2 | 41.7 | 35.7 | 32.1 | 39.7 |
| | Weekday + Hour | 168 | **83.9** | 65.5 | 79.2 | **72.0** | 45.2 | 72.6 | 69.7 |
| | Month + Hour | 288 | 72.6 | **72.2** | 72.9 | 70.1 | **70.5** | **78.1** | **72.7** |
| *HONG-2014* | No Grouping | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Weekday | 7 | 12.5 | 0 | 0 | 0 | 0 | 0 | 2.1 |
| | Month | 12 | 8.3 | 8.3 | 0 | 0 | 0 | 8.3 | 4.2 |
| | Hour | 24 | 16.7 | 8.3 | 12.5 | 8.3 | 0 | 4.2 | 8.3 |
| | Month + Weekday | 78 | 34.8 | 37.7 | 26.0 | 41.1 | 43.0 | 36.1 | 36.5 |
| | Weekday + Hour | 168 | 63.5 | 62.5 | **80.4** | 56.3 | 38.7 | 43.8 | 57.5 |
| | Month + Hour | 288 | **77.4** | **68.6** | 74.7 | **73.3** | **76.4** | **78.5** | **74.8** |

#### 5.3.3 Post-Processing Probabilistic Load Forecasts with Simulated Residuals

To investigate whether residual simulations help further improve the probabilistic load forecasts, the original point forecast results from applying temperature scenarios generated from the fixed-date method are post-processed by adding 100 simulated residuals to the forecasted load of each hour. The 100 simulated residuals are random numbers generated from a normal distribution with parameters derived from the empirical distribution of the point forecasts residuals of the previous year (i.e. the validation year).

Table 6 lists the quantile scores of the three models with different group analysis options. The "best group" method is to use the grouping method that returns the lowest quantile score of previous year as the grouping method of the current year. Considering *Vanilla* as an example, year 2006 is the first year that has probabilistic forecasts. Since the previous year to use as the reference to select the best group is not available, "N/A" is listed in Table 5. The lowest quantile score for *Vanilla* in year 2006 is 83.65, corresponding to "no grouping" method. Therefore, for year 2007, the selected "best group" strategy is "no grouping", which returns the quantile score 94.35 in 2007. On the other hand, the best group strategy in hindsight for 2007 is to group by hour, of which the quantile score is 94.30.

Overall, the following observations can be drawn from the results:

(1) The differences between the hindsight best average quantile score and the original score for *T-cube*, *Vanilla*, and *Hong-2014* are 33.2, 0.9, and 0.2 respectively. In other words, adding residuals simulated from normal distribution helps improve the probabilistic forecasts of deficient underlying models. The improvement is diminishing as the underlying model is being improved.

(2) From the perspective of forecasting improvement, how to improve the forecasts from the models with high predictive power is of the interest. Based on the results listed in Table 5, when the underlying model is very comprehensive (i.e., Hong-2014), the improvement is negligible (i.e., 0.2 in this case study).

Table 6: Quantile scores for probabilistic forecasts based on various residual simulation methods

| | **Grouping Method** | **2006** | **2007** | **2008** | **2009** | **2010** | **2011** | **AVG.** |
|---|---|---|---|---|---|---|---|---|
| *T-cube* | (original) | 116.13 | 129.45 | 128.90 | 128.57 | 146.82 | 125.15 | 129.2 |
| | No Grouping | 110.51 | 122.47 | 122.38 | 121.36 | 140.07 | 121.80 | 123.1 |
| | Weekday | 110.73 | 122.38 | 122.57 | 121.69 | 140.22 | 121.15 | 123.1 |
| | Month | 109.79 | 122.18 | 121.29 | 118.71 | 129.72 | 118.99 | 120.1 |
| | Hour | 88.85 | 99.79 | 102.53 | 98.57 | 119.79 | 99.27 | 101.5 |
| | Month + Weekday | 109.94 | 122.63 | 121.45 | 119.32 | 129.99 | 119.11 | 120.4 |
| | Weekday + Hour | 87.77 | 98.37 | 101.37 | 97.51 | 118.97 | 98.12 | 100.4 |
| | Month + Hour | **87.18** | **97.82** | **98.79** | **93.30** | **106.13** | **93.02** | **96.0** |
| | Best Group | N/A | **97.82** | **98.79** | **93.30** | **106.13** | **93.02** | 97.8 |
| *Vanilla* | (original) | 84.67 | 95.68 | 96.02 | 90.82 | 106.65 | 92.46 | 94.4 |
| | No Grouping | **83.65** | 94.35 | **95.51** | 89.75 | 107.12 | 92.63 | 93.8 |
| | Weekday | 84.00 | 94.52 | 96.93 | 89.89 | 107.06 | 90.88 | 93.9 |
| | Month | 83.72 | 95.18 | 96.16 | 90.37 | **104.98** | 90.79 | **93.5** |
| | Hour | 83.79 | **94.30** | 95.96 | **89.66** | 106.93 | 90.86 | 93.6 |
| | Month + Weekday | 83.92 | 96.21 | 96.39 | 90.97 | 105.20 | 90.97 | 93.9 |
| | Weekday + Hour | 84.19 | 94.63 | 96.26 | 89.93 | 106.92 | 90.86 | 93.8 |
| | Month + Hour | 84.42 | 95.95 | 96.49 | 90.81 | 105.53 | **90.69** | 94.0 |
| | Best Group | N/A | 94.35 | 95.96 | 89.75 | 106.93 | 90.79 | 95.6 |
| *Hong-2014* | (original) | 82.89 | 95.35 | 94.58 | 90.13 | **108.46** | 90.39 | 93.6 |
| | No Grouping | **82.19** | 94.36 | 94.03 | 89.56 | 110.60 | 90.09 | 93.5 |
| | Weekday | 82.40 | 94.51 | 94.20 | 89.63 | 110.58 | 90.20 | 93.6 |
| | Month | 82.25 | 95.61 | 94.85 | 89.81 | 109.57 | 91.30 | 93.9 |
| | Hour | 82.21 | **94.12** | **93.99** | **89.48** | 110.30 | **90.07** | **93.4** |
| | Month + Weekday | 82.54 | 96.31 | 94.85 | 90.13 | 109.73 | 91.38 | 94.2 |
| | Weekday + Hour | 82.52 | 94.23 | 94.05 | 89.57 | 110.39 | 90.09 | 93.5 |
| | Month + Hour | 83.07 | 95.68 | 94.51 | 90.08 | 109.95 | 91.40 | 94.1 |
| | Best Group | N/A | 94.36 | **93.99** | **89.48** | 110.30 | 90.39 | 95.7 |

To investigate whether a higher normality test passing rate indicates a better quantile score when modeling the residuals with normal distributions, the normality passing rate and the quantile score from each model and each grouping method is plotted in Figure 20. There exist no obvious relationship between the normality passing rate and the quantile score for the Vanilla and the Hong-2014 model. In other words, for this two models, a higher normality passing rate does not necessarily lead to a better quantile score. But, for the T-cube model, there is a downward trend in quantile score when residuals are grouped by seasonal variables such as month, weekday, and hour of a day for residual simulation. This is because the T-cube model doesn't consider the seasonality of electricity demand series in the model and the information are left in the residuals. When the residuals are grouped analyzed by such seasonality variables, it can significantly help improve the forecasts.

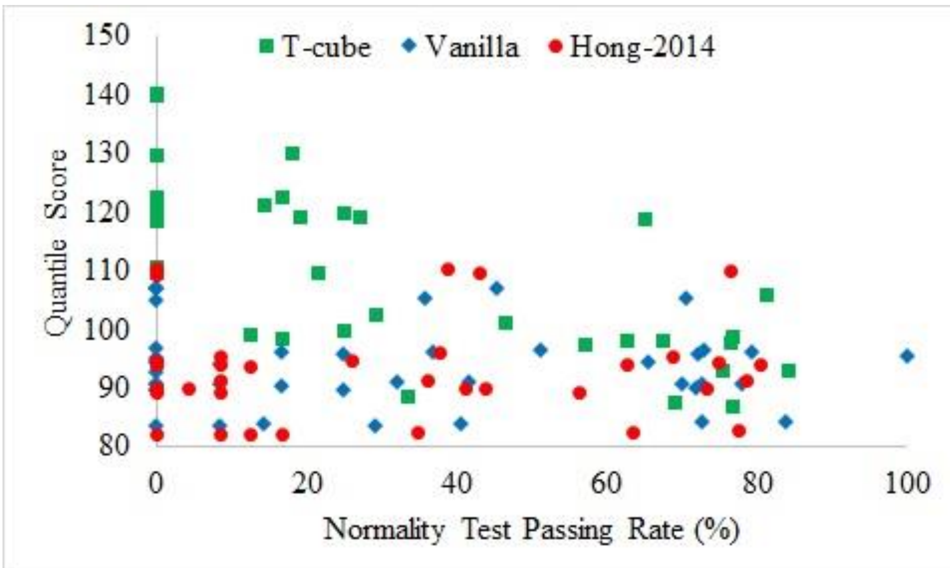
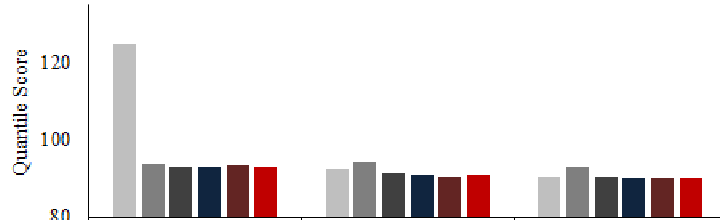



Figure 20: Normality test passing rate vs. quantile score
(Case = NCEMC, Linear models)

### 5.3.4 Additional Simulation

The results presented so far are all based on adding 100 simulated residuals to each hour through the forecast horizon. Further investigation on whether the number of simulations affects the quantile score have also been done. Figure 21 lists the quantile scores of the three regression models under a different number of simulations. Since the results are similar across all years for all grouping strategies, only the results from the year of 2011 with the best grouping strategy of each model for both data sets are shown to avoid verbose presentation. According to the results, when the number of simulations is 50 or above, the quantile scores are quite similar. While 10 simulations do not seem to be good enough, too many simulations beyond 200 do not add much value either.

Figure 22 shows the probabilistic monthly energy forecast of *Hong-2014* model for NCEMC case with the number of simulations as 0, 100 and 500 for year 2011, respectively. The appearance of the probabilistic forecasts also agrees with the numbers in Table 5. Note that the load variation in the winter months is much wider than that of the summer months, which is a common phenomenon in North Carolina's cooperatives due to the wide adoption of electricity heating systems and climate diversity in the winter. Reflecting such phenomenon through probabilistic forecasts is quite helpful to power supply planning.




| | T-Cube | Vanilla | Hong-2014 |
|---|---|---|---|
| 0 | 125.152 | 92.456 | 90.385 |
| 10 | 93.605 | 94.388 | 92.916 |
| 50 | 92.871 | 91.172 | 90.415 |
| 100 | 93.023 | 90.692 | 90.072 |
| 200 | 93.281 | 90.35 | 89.857 |
| 500 | 93.014 | 90.778 | 90.133 |

*Figure 21: Quantile scores from different number of residual simulations*

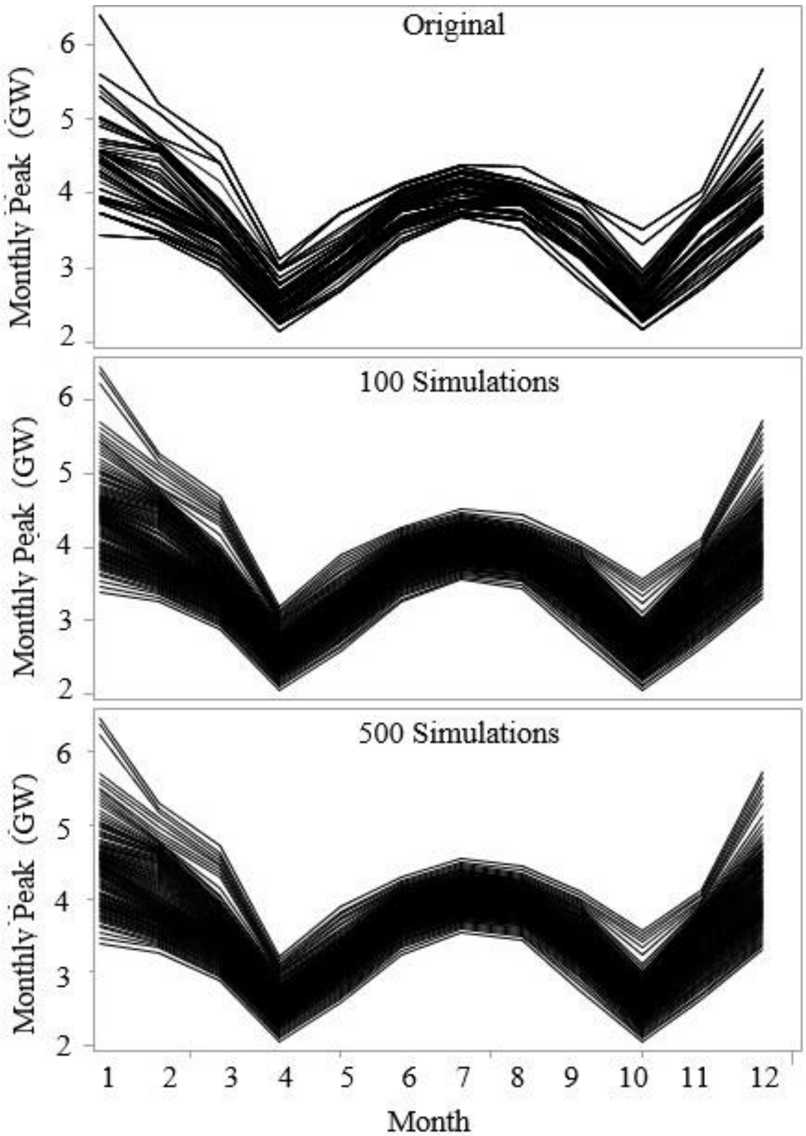



Figure 22: Probabilistic monthly peak forecasts with different number of residual simulations (Case = NCEMC, Model = Hong-2014)

### 5.3.5 Results from Nonlinear Models

Table 7 lists the specifications of the three ANN models, together with the quantile scores from the NCEMC case study. The scatter plot in Figure 23 shows the relationship between normality passing rate and quantile score. The results reaffirm the conclusions drawn earlier. To avoid verbose presentation, only the results from NCEMC case study are presented in this dissertation.

Table 7: Quantile scores for probabilistic forecasts based on various residual simulation methods

| | **Input Variables** | | **2006** | **2007** | **2008** | **2009** | **2010** | **2011** | **AVG.** |
|---|---|---|---|---|---|---|---|---|---|
| *T-cube* | $GSP_t$, $T_t$, (Number of input neurons: 2) | Hidden Neurons | 2 | 3 | 4 | 4 | 4 | 4 | |
| | | (original) | 136.24 | 132.41 | 124.88 | 124.65 | 142.90 | 128.42 | 131.58 |
| | | No Grouping | 127.50 | 115.33 | 120.56 | 120.21 | 137.41 | 124.18 | 124.20 |
| | | Weekday | 127.15 | 116.08 | 120.64 | 120.53 | 137.37 | 124.16 | 124.32 |
| | | Month | 124.96 | 114.36 | 119.78 | 116.94 | 128.91 | 121.66 | 121.10 |
| | | Hour | 105.49 | 95.95 | 102.83 | 99.81 | 117.28 | 104.11 | 104.25 |
| | | Month + Weekday | 125.38 | 114.52 | 119.85 | 117.46 | 129.22 | 122.32 | 121.46 |
| | | Weekday + Hour | 104.02 | 95.10 | 101.68 | 98.78 | 116.45 | 103.14 | 103.20 |
| | | Month + Hour | **100.97** | **92.10** | **99.65** | **94.03** | **105.72** | **97.998** | 98.41 |
| | | Best Group | N/A | **92.10** | **99.65** | **94.03** | **105.72** | **97.998** | **97.09** |
| *Vanilla* | $GSP_t$, $T_t$, $M_t$, $W_t$, $H_t$ (Number of input neurons: 42) | Hidden Neurons | 40 | 20 | 20 | 20 | 20 | 20 | |
| | | (original) | 111.03 | 112.36 | 98.08 | 91.77 | 114.81 | **100.62** | 104.78 |
| | | No Grouping | 101.07 | 101.97 | 96.16 | 91.29 | 111.53 | 123.39 | 104.24 |
| | | Weekday | 100.98 | 99.47 | 120.64 | 91.65 | 111.32 | 124.43 | 108.08 |
| | | Month | 98.25 | 97.87 | 97.19 | 90.45 | 108.82 | 119.59 | 102.03 |
| | | Hour | 100.77 | 101.14 | 96.27 | 91.24 | 111.34 | 124.88 | 104.27 |
| | | Month + Weekday | 99.21 | 94.84 | 97.03 | 91.18 | **108.59** | 120.88 | 101.96 |
| | | Weekday + Hour | 100.75 | 98.20 | **95.99** | 91.53 | 111.10 | 125.27 | 103.81 |
| | | Month + Hour | **98.65** | **94.36** | 96.57 | **90.08** | 108.63 | 120.59 | **101.48** |
| | | Best Group | N/A | **94.36** | 96.75 | 91.53 | 108.63 | **120.88** | 102.43 |
| *Hong-2014* | $GSP_t$, $T_t$, $T_{t-1}$, $T_{t-2}$, $T_{t-3}$, $T_a$, $M_t$, $D_t$, $H_t$ (Number of input neurons: 46) | Hidden Neurons | 25 | 40 | 25 | 25 | 35 | 55 | |
| | | (original) | 101.41 | 101.48 | 96.32 | 90.25 | 113.26 | 98.62 | 100.22 |
| | | No Grouping | 98.86 | 99.36 | 95.43 | 89.70 | 111.05 | 96.80 | 98.53 |
| | | Weekday | 98.90 | 99.19 | 95.44 | 89.88 | 111.04 | 96.79 | 98.54 |
| | | Month | 98.25 | 93.81 | 96.48 | 89.13 | 108.65 | **94.58** | 96.82 |
| | | Hour | 98.00 | 98.56 | 95.18 | 89.53 | 110.92 | 96.956 | 98.19 |
| | | Month + Weekday | 98.55 | 93.64 | 96.47 | 89.21 | 108.71 | 95.26 | 96.97 |
| | | Weekday + Hour | 97.91 | 98.42 | **95.14** | 89.69 | 110.85 | 96.957 | 98.16 |
| | | Month + Hour | **96.57** | **92.86** | 95.64 | **88.96** | **108.49** | 94.72 | **96.21** |
| | | Best Group | N/A | **92.86** | 95.64 | 89.69 | 108.49 | 94.72 | 96.28 |

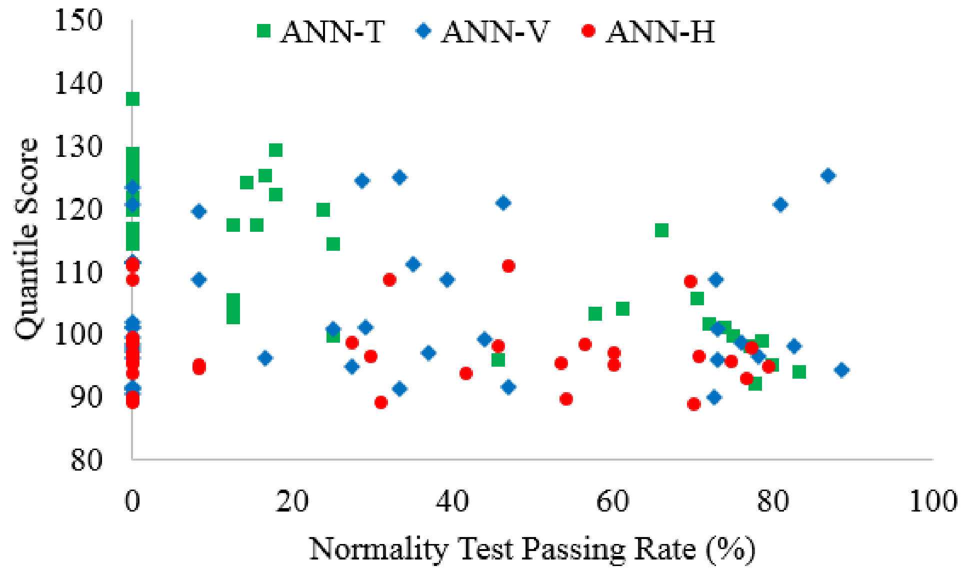



Figure 23: Normality test passing rate vs. quantile score
(Case = NCEMC, ANN models)

### 5.3.6 Results from GEFCom2014-L Data

The same process to simulate residuals and then add them back to the point forecast results is then repeated for the GEFCom2014-L data. Table 8 lists the passing rate of the normality tests against the residuals and the corresponding quantile scores. Although there are only three years of testing period, we have the same observations as from the NCEMC case.

Table 8: Normality test passing rates and quantile scores

| | | Passing Rate (%) | | | | | Quantile Score | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Grouping Method** | **2008** | **2009** | **2010** | **AVG.** | | **2009** | **2010** | **2011** | **AVG.** |
| *T-cube* | (original) | N/A | N/A | N/A | N/A | | 10.88 | 12.18 | 10.78 | 11.28 |
| | None | 0 | 0 | 0 | 0 | | 9.79 | 11.55 | 9.94 | 10.43 |
| | Weekday | 0 | 0 | 0 | 0 | | 9.86 | 11.54 | 9.94 | 10.45 |
| | Month | 0 | 0 | 0 | 0 | | 9.72 | 10.89 | 8.76 | 9.79 |
| | Hour | 16.7 | 20.8 | 16.7 | 18.1 | | 8.72 | 10.55 | 8.95 | 9.41 |
| | Month + Weekday | 23.8 | 28.6 | 26.2 | 26.2 | | 9.80 | 10.90 | 9.79 | 10.16 |
| | Weekday + Hour | 52.4 | 65.5 | 63.7 | 60.5 | | 8.70 | 10.53 | 8.93 | 9.39 |
| | Month + Hour | **84.4** | **85.8** | **77.1** | **82.4** | | **8.34** | **9.57** | **8.52** | **8.81** |
| | Best Group | N/A | **85.8** | **77.1** | 81.4 | | N/A | **9.57** | **8.52** | 9.05 |
| *Vanilla* | (original) | N/A | N/A | N/A | N/A | | 8.29 | **9.41** | 8.39 | 8.70 |
| | None | 0 | 0 | 0 | 0 | | **7.93** | 9.45 | **8.08** | **8.49** |
| | Weekday | 0 | 0 | 0 | 0 | | 7.96 | 9.44 | 8.09 | 8.50 |
| | Month | 8.3 | 0 | 0 | 2.8 | | 8.13 | 9.50 | 8.20 | 8.61 |
| | Hour | 16.7 | 20.8 | 12.5 | 16.7 | | 7.94 | 9.44 | 8.09 | 8.49 |
| | Month + Weekday | 34.5 | 31.0 | 31.0 | 32.2 | | 8.17 | 9.47 | 8.25 | 8.63 |
| | Weekday + Hour | 74.4 | 74.4 | 64.9 | 71.2 | | 7.95 | 9.44 | 8.10 | 8.50 |
| | Month + Hour | **78.5** | **77.4** | **83.0** | **79.6** | | 8.08 | 9.49 | 8.24 | 8.60 |
| | Best Group | N/A | 0 | N/A | 0 | | N/A | 9.45 | 8.39 | 8.92 |
| *Hong-2014* | (original) | N/A | N/A | N/A | N/A | | 8.20 | **9.39** | 8.27 | 8.62 |
| | None | 0 | 0 | 0 | 0 | | **7.99** | 9.55 | **8.04** | **8.527** |
| | Weekday | 0 | 0 | 0 | 0 | | 8.03 | 9.55 | 8.07 | 8.55 |
| | Month | 0 | 8.3 | 8.3 | 5.5 | | 8.14 | 9.59 | 8.14 | 8.62 |
| | Hour | 8.3 | 16.7 | 4.20 | 9.7 | | 8.00 | 9.55 | 8.05 | 8.53 |
| | Month Weekday | 26.7 | 29.2 | 36.7 | 30.9 | | 8.18 | 9.59 | 8.18 | 8.65 |
| | Weekday + Hour | 60.0 | 72.2 | 29.2 | 53.8 | | 8.02 | 9.56 | 8.08 | 8.55 |
| | Month + Hour | **80.9** | **71.18** | **75.0** | **75.7** | | 8.11 | 9.60 | 8.16 | 8.62 |
| | Best Group | N/A | 0 | N/A | 0 | | N/A | 9.55 | 8.27 | 8.91 |

# CHAPTER 6: CONCLUSION AND FUTURE WORK

This dissertation presents a formal study of probabilistic electric load forecasting. The study dissects the probabilistic load forecasting problem into three components: (1) input scenario simulation, (2) probabilistic modeling, and (3) residual simulation. From the input scenario perspective, it proposes a framework to evaluate different temperature scenario generation techniques. The proposed framework is applied to evaluate three temperature scenario generation methods that have all been implemented in the field, namely the fixed-date method, the shifted-date method, and the bootstrap method. The evaluation from the proposed framework help to build the methodological foundation for practicing the temperature scenario generation methodologies. From the output perspective, a comprehensive study has been conducted on the normality assumption of residuals from point forecast models. This comprehensive study on residuals considers the multiple seasonality in electricity demand series and group the residuals by seasonal factors for analysis. A residual simulation method has been proposed to improve the probabilistic load forecasts. The proposed methods offer multiple practical options for utilities to generate their probabilistic load forecasts based on their unique situations such as availability of historical data and comprehensive level of load forecasting models. Some of these proposed methods have been implemented in utilities' probabilistic load forecasting practices. Some further extensions on this topic can be: from the input scenario simulation perspective, although temperature is the most significant weather variable that drives electricity demand, other variables such as relative humidity, wind chill, and etc. may also impact the demand. The proposed framework can be applied to evaluate the scenario generation methods for other weather variables. In addition to weather scenarios, economics scenario can also be generated for creating probabilistic load forecasts. Furthermore, the interactions of scenarios may lead to many more scenarios. The use of economic scenarios and the cross scenarios of temperature scenarios and economic scenarios have been studied by Hong et al. [4] but a more formal and systematical study is needed. From the residual simulation perspective, residual simulated from other distributions or other methods such as bootstrap can also be further analyzed.

# REFERENCES


[1] T. Hong, "Short term electric load forecasting," Ph.D. Dissertation, Graduate Program of Operation Research and Dept. of Electrical and Computer Engineering, North Carolina State University, 2010.

[2] T. Hong and S. Fan, "Probabilistic electric load forecasting: a tutorial review," *Int. J. Forecast.*, vol. 32, no. 3, pp. 914–938, 2016.

[3] J. Xie, T. Hong, and C. Kang, "From high-resolution data to high-resolution probabilistic load forecasts," in *2016 IEEE PES Transmission and Distribution Conference and Exposition, Dallas, TX, May 2-5*, 2016.

[4] T. Hong, J. Wilson, and J. Xie, "Long term probabilistic load forecasting and normalization with hourly information," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 456–462, 2014.

[5] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A statistical Approach*. New York: Wiley Finance, 2006.

[6] S. Fan and R. Hyndman, "Short-term load forecasting based on a semi-parametric additive model," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 134–141, 2012.

[7] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, 2001.

[8] PJM, "PJM Load Forecast Report January 2015 Prepared by PJM Resource Adequacy Planning Department," 2015. [Online]. Available: https://www.pjm.com/~/media/documents/reports/2015-load-forecast-report.ashx. [Accessed: 08-Mar-2016].

[9] National Oceanic and Atmospheric Administration, "1981-2010 U.S. Climate Normals," *National Oceanic and Atmospheric Administration*, 2010. [Online]. Available: https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climate-normals/1981-2010-normals-data. [Accessed: 10-Feb-2016].

[10] E. Feinberg and D. Genethliou, "Load Forecasting," *Appl. Math. Restructured Electr. Power Syst. Optim. Control. Comput. Intell.*, pp. 269–285, 2006.

[11] C. E. Asbury, "Weather load model for electric demand and energy forecasting," *IEEE Trans. Power Appar. Syst.*, vol. 94, no. 4, pp. 1111–1116, 1975.

[12] M. S. Kandil, S. M. El-Debeiky, and N. E. Hasanien, "Long-term load forecasting for fast developing utility using a knowledge-based expert system," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 491–496, 2002.

[13] T. Hong, S. M. Hsiang, and L. Xu, "Human-machine co-construct intelligence on horizon year load in long term spatial load forecasting," *2009 IEEE Power Energy Soc. Gen. Meet. Calgary, Alberta, Canada*, 2009.

[14] J. W. Taylor and P. E. McSharry, "Short-term load forecasting methods: An evaluation based on European data," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2213–2219, 2007.

[15] A. D. Papalexopoulos and T. C. Hesterberg, "A Regression-based approach to short-term system load forecasting," *IEEE Trans. Power Syst.*, vol. 5, no. 4, pp. 1535–1547, 1990.

[16] T. Haida and S. Muto, "Regression based peak load forecasting using a transformation

technique," *Power Syst. IEEE Trans.*, vol. 9, no. 4, pp. 1788–1794, 1994.

[17] O. Hyde and P. F. Hodnett, "An adaptable automated procedure for short-term electricity load forecasting," *IEEE Trans. Power Syst.*, vol. 12, no. 1, pp. 84–94, 1997.

[18] J. D. Black and W. L. W. Henson, "Hierarchical load hindcasting using reanalysis weather," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 447–455, 2014.

[19] B. A. Hoverstad, A. Tidemann, H. Langseth, and P. Ozturk, "Short-term load forecasting with seasonal decomposition using evolution for parameter tuning," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1904–1913, 2015.

[20] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *Int. J. Forecast.*, vol. 30, no. 2, pp. 357–363, Aug. 2014.

[21] J. Xie, Y. Chen, T. Hong, and T. D. Laing, "Relative humidity for load forecasting models," *IEEE Trans. Smart Grid*, vol. in press, 2016.

[22] P. Wang, B. Liu, and T. Hong, "Electric load forecasting with recency effect: a big data approach," *Int. J. Forecast.*, vol. 32, no. 3, pp. 585–597, 2016.

[23] J. Xie and T. Hong, "Comparing two model selection frameworks for probabilistic load forecasting," in *2016 International Conference on Probabilistic Methods Applied to Power Systems, Beijing, China, Oct 16-20*, 2016.

[24] J. Xie and A. Chien, "Holiday demand forecasting in the electric utility industry," in *2016 SAS Global Forum, Las Vegas, NV, Apr 18-21*, 2016.

[25] J. Xie and T. Hong, "GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation," *Int. J. Forecast.*, vol. 32, no. 3, pp. 1012–1016, Dec. 2015.

[26] J. Xie, T. Hong, T. Laing, and C. Kang, "On normality assumption in residual simulation for probabilistic load forecasting," *IEEE Trans. Smart Grid*, vol. in press, 2016.

[27] J. Xie, T. Hong, and J. Stroud, "Long-term retail energy forecasting with consideration of residential customer attrition," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2245 – 2252, 2015.

[28] N. Charlton and C. Singleton, "A refined parametric model for short term load forecasting," *Int. J. Forecast.*, vol. 30, no. 2, pp. 364–368, Apr. 2014.

[29] D. C. Park, M. A. El-Sharkawi, R. J. Marks, L. E. Atlas, and M. J. Damborg, "Electric load forecasting using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 6, no. 2, pp. 442–449, 1991.

[30] K. Lee, Y. Cha, and J. Park, "Short-term load forecasting using an artificial neural network," *Power Syst. IEEE Trans.*, vol. 7, no. 1, pp. 124–132, 1992.

[31] S.-T. Chen, D. C. Yu, and A. R. Moghaddamjo, "Weather sensitive short-term load forecasting using nonfully connected artificial neural network," *Trans. Power Syst.*, vol. 7, no. 3, pp. 1098–1105, 1992.

[32] A. D. Papalexopoulos, S. Hao, and T. Peng, "An Implementation of a neuronal network based load forecasting model for the EMS," *IEEE Trans. Power Syst.*, vol. 9, no. 4, pp. 1956–1962, 1994.

[33] A. Khotanzad, R. C. Hwang, A. Abaye, and D. Maratukulam, "Adaptive modular artificial neural network hourly load forecaster and its implementation at electric utilities," *IEEE*

*Trans. Power Syst.*, vol. 10, no. 3, pp. 1716–1722, 1995.

[34] A. Khotanzad and R. Afkhami-Rohani, "ANNSTLF - Artificial neural network short-term load forecaster - generation three," *IEEE Trans. Power Syst.*, vol. 13, no. 4, pp. 1413–1422, 1998.

[35] A. Khotanzad, R. Afkhami-Rohani, T. L. Lu, A. Abaye, M. Davis, and D. J. Maratukulam, "ANNSTLF - A neural-network-based electric load forecasting system," *IEEE Trans. Neural Networks*, vol. 8, no. 4, pp. 835–846, 1997.

[36] G. Gross and F. D. Galiana, "Short-term load forecasting," *Proc. IEEE*, vol. 75, no. 12, pp. 1558–1573, 1987.

[37] Ying Chen, P. B. Luh, Che Guan, Yige Zhao, L. D. Michel, M. a. Coolbeth, P. B. Friedland, and S. J. Rourke, "Short-term load forecasting: similar day-based wavelet neural networks," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 322–330, 2010.

[38] M. T. Hagan and S. M. Behr, "The time Series approach to short term load forecasting," *IEEE Trans. Power Syst.*, vol. 2, no. 3, pp. 785–791, 1987.

[39] J. Taylor, "Short-term load forecasting with exponentially weighted methods," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 458–464, 2012.

[40] J. W. Taylor and J. W. Taylor, "Triple seasonal methods for short-term electricity demand forecasting," *Eur. J. Oper. Res.*, vol. 44, no. 0, pp. 139–152, 2010.

[41] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *J. Oper. Res. Soc.*, vol. 54, no. 8, pp. 799–805, 2003.

[42] B. Chen, M. Chang, and C. Lin, "Load forecasting using support vector machines: a study on EUNITE competition 2001," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1821–1830, 2004.

[43] S. Fan and L. Chen, "Short-term load forecasting based on an adaptive hybrid method," *Power Syst. IEEE Trans.*, vol. 21, no. 1, pp. 392–401, 2006.

[44] S. Fan, K. Methaprayoon, and W.-J. Lee, "Multiregion load forecasting for system with large geographical area," *IEEE Trans. Ind. Appl.*, vol. 45, no. 4, pp. 1452–1459, Jul. 2009.

[45] K. L. Ho, Y. Y. Hsu, C. F. Chen, T. E. Lee, C. C. Liang, T. S. Lai, and K. K. Chen, "Short term load forecasting of taiwan power system using a knowledge-based expert system," *IEEE Trans. Power Syst.*, vol. 5, no. 4, pp. 1214–1221, 1990.

[46] M. S. Kandil, S. M. El-Debeiky, and N. E. Hasanien, "The implementation of long-term forecasting strategies using a knowledge-based expert system: part-II," *Electr. Power Syst. Res.*, vol. 58, no. 1, pp. 19–25, May 2001.

[47] D. K. Ranaweera, N. F. Hubele, and G. G. Karady, "Fuzzy logic for short term load forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 18, no. 4, pp. 215–222, 1996.

[48] T. Hong and P. Wang, "Fuzzy interaction regression for short term load forecasting," *Fuzzy Optim. Decis. Mak.*, vol. 13, no. 1, pp. 91–103, Sep. 2013.

[49] T. Hong, "Electric load forecasting: fundamentals and best practice." [Online]. Available: https://support.sas.com/edu/schedules.html?id=1326. [Accessed: 13-Jun-2016].

[50] M. S. Abou-Hussien, M. S. Kandlil, M. A. Tantawy, and S. A. Farghal, "An accurate model for short-term load forecasting," *IEEE Trans. Power Appar. Syst.*, vol. PAS-100, no. 9, pp.

4158–4165, 1981.

[51] R. Nedellec, J. Cugliari, and Y. Goude, “GEFCom2012: Electric load forecasting and backcasting with semi-parametric models,” *Int. J. Forecast.*, vol. 30, no. 2, pp. 375–381, Apr. 2014.

[52] D. Alex and S. Hao, “An implementation of a neural network based load forecasting model for the EMS,” *IEEE Trans. Power Syst.*, vol. 9, no. 4, pp. 1956–1962, 1994.

[53] A. P. Douglas, A. M. Breipohl, F. N. Lee, and R. Adapa, “The impacts of temperature forecast uncertainty on bayesian load forecasting,” *IEEE Trans. Power Syst.*, vol. 13, no. 4, pp. 1507–1513, 1998.

[54] S. Rahman, “Formulation and analysis of a rule-based short-term load forecasting algorithm,” *Proc. IEEE*, vol. 78, no. 5, pp. 805–816, 1990.

[55] Z. Wang, M. Begovic, and J. Wang, “Analysis of conservation voltage reduction effects based on multistage SVR and stochastic process,” *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 431–439, 2014.

[56] S. Rahman, “Formulation and analysis of a rule-based short-term load forecasting algorithm,” *Proc. IEEE*, vol. 78, no. 5, pp. 805–816, 1990.

[57] W.-C. Chu, Y.-P. Chen, Z.-W. Xu, and W.-J. Lee, “Multiregion short-term load forecasting in consideration of HI and load/weather diversity,” *IEEE Trans. Ind. Appl.*, vol. 47, no. 1, pp. 232–237, 2011.

[58] C. L. Hor, S. J. Watson, and S. Majithia, “Analyzing the impact of weather variables on monthly electricity demand,” *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 2078–2085, 2005.

[59] S. E. Papadakis, J. B. Theocharis, S. J. Kiartzis, and A. G. Bakirtzis, “A novel approach to short-term load forecasting using fuzzy neuronal networks,” *IEEE Trans. Power Syst.*, vol. 13, no. 2, pp. 480–492, 1998.

[60] K.-B. Song, S.-K. Ha, J.-W. Park, D.-J. Kweon, and K.-H. Kim, “Hybrid Load Forecasting Method With Analysis of Temperature Sensitivities,” *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 869–876, 2006.

[61] H. Chao-Ming, H. Chi-Jen, and W. Ming-Li, “A particle swarm optimization to identifying the ARMAX model for short-term load forecasting,” *Power Syst. IEEE Trans.*, vol. 20, no. 2, pp. 1126–1133, 2005.

[62] K. Song, Y. Baek, D. H. Hong, and G. Jang, “Short-term load forecasting for the holidays using fuzzy linear regression method,” *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 96–101, 2005.

[63] I. Moghram and S. Rahman, “Analysis and evaluation of five short-term load forecasting techniques,” *IEEE Trans. Power Syst.*, vol. 4, no. 4, pp. 1484–1491, 1989.

[64] S. Rahman and R. Bhatnagar, “An expert system based algorithm for short term load forecast,” *IEEE Trans. Power Syst.*, vol. 3, no. 2, pp. 392–399, 1988.

[65] B. Krogh, E. S. de Llinas, and D. Lesser, “Design and Implementation of an on-line Load Forecasting Algorithm,” *IEEE Trans. Power Appar. Syst.*, vol. PAS-101, no. 9, pp. 3284–3289, 1982.

[66] R. Hyndman and S. Fan, “Density forecasting for long-term peak electricity demand,” *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 1142–1153, 2010.

[67] V. Dordonnat, A. Pichavant, and A. Pierrot, “GEFCom2014 probabilistic electric load forecasting using time series and semi-parametric regression models,” *Int. J. Forecast.*, vol. 32, no. 3, pp. 1005–1011, 2016.

[68] J. Xie and T. Hong, “GEFCom2014: Probabilistic electric load forecasting: An integrated framework with forecast combinatin and residual simulation,” *Int. J. Forecast.*, vol. 32, no. 3, pp. 1012–1016, 2016.

[69] P. Gaillard, Y. Goude, and R. Nedellec, “Semi-parametric models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting,” *Int. J. Forecast.*, vol. 32, no. 3, pp. 1038–1050, 2016.

[70] P. Kou and F. Gao, “A sparse heteroscedastic model for the probabilistic load forecasting in energy-intensive enterprises,” *Int. J. Electr. Power Energy Syst.*, vol. 55, pp. 144–154, 2014.

[71] 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, pp. 1166–1172, 2005.

[72] B. Liu, J. Nowotarski, T. Hong, and R. Weron, “Probabilistic load forecasting via quantile regression averaging on sister forecasts,” *IEEE Trans. Smart Grid*, vol. in press, 2016.

[73] S. Haben and G. Giasemidis, “A hybrid model of kernel density estimation and quantile regression for GEFCom2014 probabilistic load forecasting,” *Int. J. Forecast.*, vol. 32, no. 3, pp. 1017–1022, 2016.

[74] S. Bentaieb, R. Huser, R. J. Hyndman, and M. G. Genton, “Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression,” *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2448–2455, 2016.

[75] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, “A state space framework for automatic forecasting using exponential smoothing methods,” *Int. J. Forecast.*, vol. 18, no. 3, pp. 439–454, 2002.

[76] A. Bracale, P. Caramia, G. Carpinelli, A. R. Di Fazio, and P. Varilone, “A bayesian-based approach for a short-term steady-state forecast of a smart grid,” *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1760–1771, 2013.

[77] R. N. Allan, B. Borkowska, and C. Grigg, “Probabilistic analysis of power flows,” *Proc. IEEE*, vol. 121, pp. 1551–1556, 1974.

[78] B. Borkowska, “Probabilistic load flow,” *IEEE Trans. Power Appar. Syst.*, vol. PAS-93, no. 3, pp. 752–759, 1974.

[79] R. N. Allan, a. M. L. Da Silva, and R. C. Burchett, “Evaluation methods and accuracy in probabilistic load flow solutions,” *IEEE Trans. Power Appar. Syst.*, vol. PAS-100, no. 5, pp. 2539–2546, 1981.

[80] J. O. Rawlings, S. G. Pantula, and D. A. Dickey, *Applied Regression Analysis A Regression Tool*. Springer, 2001.

[81] S. Haykin, *Neural networks and learning machines*. Pearson, 2008.

[82] S. Haykin, *Neural networks: a comprehensive foundation*. Mac Millan Publishing Company, 1994.

[83] P. Pinson, H. A. Nielsen, J. K. Moller, H. Madsen, and G. N. Kariniotakis, “Non-parametric

probabilistic forecasts of wind power: Required properties and evaluation," *Wind Energy*, vol. 10, no. 6, pp. 497–516, 2007.

[84] L. Magnano and J. W. Boland, "Generation of synthetic sequences of electricity demand: application in South Australia," *Energy*, vol. 32, no. 11, pp. 2230–2243, 2007.

[85] R. W. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.

[86] J. Xie and T. Hong, "Temperature scenario generation for probabilistic load forecasting," *IEEE Trans. Smart Grid*, 2016.

[87] T. Hong, Pierre Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," *Int. J. Forecast.*, vol. 32, no. 3, pp. 896–913, 2016.

[88] A. H. Myrphy and R. L. Winkler, "Probability forecasting in meteorology," *J. Am. Stat. Assoc.*, vol. 79, no. 387, pp. 489–500, 1984.

[89] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Stat. Surverys*, vol. 4, pp. 40–79, 2010.

[90] M. Stone, "Cross-Validatory Choice and Assessment of Statistical Predictions," *J. R. Stat. Soc.*, vol. 36, no. 2, pp. 111–147, 1974.

[91] S. Geisser, "The predictive sample reuse method with applications," *J. Am. Stat. Assoc.*, vol. 70, no. 350, pp. 320–328, 1974.

[92] J. Shao, "Linear model selection by cross-validation," *J. Am. Stat. Assoc.*, vol. 88, no. 422486–494, 1993.

[93] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: an analysis and review," *Int. J. Forecast.*, vol. 16, no. 4, pp. 437–450, 2000.

[94] H. W. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *J. Am. Stat. Assoc.*, vol. 62, no. 318, pp. 399–402, 1967.