SHORT TERM PEAK TIMING FORECASTING

by

Shreyashi Shukla

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

2023

Approved by:

_____

Dr. Tao Hong

_____

Dr. Churlzu Lim

_____

Dr. Pu Wang

_____

Dr. Jay Wu

ABSTRACT

SHREYASHI SHUKLA. Short Term Peak Timing Forecasting. (Under the direction of DR. TAO HONG)

Peak load forecasting is crucial for reliable and effective grid operation. The day-to-day operation of the power grid requires load scheduling and dispatches of different energy resources including Energy Storage System and Demand Side Management programs. An effective implementation of these peak-shaving strategies relies heavily on when the peak demand occurs. Hence, forecasting the timing of peak load is as important as forecasting its magnitude.

A review of relevant literature indicates that there is no inclusive study on the topic of peak timing forecasting. This research aims to bridge the gap between industry requirements and academic research by addressing some key questions. First, the study defines the different forms of peak timing problems that can benefit grid operation. Next, we investigate the problem of how we measure the peak timing forecast errors. The research critically reviews error measures used in the literature for peak timing forecasting. Based on the findings five application-specific error measures are proposed. The research then focuses on one of the manifestations of the peak timing problem, that is, forecasting daily peak hours.

We analyzed the accuracy of peak hour forecasts from a state-of-the-art hourly load forecasting model and set it as the benchmark. The model selection process using different peak timing errors and load shape errors is investigated. Furthermore, two novel peak-hour forecasting frameworks are developed and their effectiveness is empirically demonstrated in two case studies. The first case study comes from an anonymous medium-sized utility in the U.S., while the second one is publicly accessible data from ISO New England. The proposed models demonstrate improved forecast results on the benchmark model by 12-16% in the test years of the two

case studies. Additionally, when the models are only evaluated on the critical days with very high demands, they outperform the benchmark by 25-53%. Findings from this study emphasize the importance of developing explicit models for peak hour forecasting by analyzing the key determinants that vary with geographical location and regional factors.

DEDICATION

To my family

ACKNOWLEDGEMENTS

I would like to express heartfelt gratitude to my advisor, Dr. Tao Hong, whose unwavering guidance and expertise have been instrumental throughout the duration of my research. During my Master's program, I was inspired by one of the courses taught by Dr. Tao Hong, on Energy Analytics, which ignited my passion and set me on the path to delve deeper into the realm of load forecasting. These past five years have been a remarkable period of growth, largely attributed to the invaluable insights and constructive feedback generously provided by my advisor.

I am also deeply thankful to Dr. Pu Wang for her steadfast guidance that significantly influenced the trajectory of my research endeavors. Her suggestions and critical evaluations enriched my learning and gave me new perspectives. I also take this opportunity to convey my sincere thankfulness to Dr. Jay Wu and Dr. Churlzu Lim for their insightful suggestions, and critical evaluation of my presentation.

I extend my heartfelt appreciation to my fellow research group members for their collaborative spirit, stimulating discussions, and assistance in various aspects of this endeavor. Furthermore, I would like to acknowledge the financial support provided by the GASP to carry out this research with the necessary resources and tools. Lastly, I owe a special debt of gratitude to my advisor, who not only guided me intellectually but also played a crucial role in facilitating grants and contracts that supported my academic pursuits.

In conclusion, my journey would not have been possible without the support of my family. I am genuinely thankful to my spouse, Kunal, and my daughter, Aanya, for their unconditional love and words of motivation.

TABLE OF CONTENTS

LIST OF TABLES xi

LIST OF FIGURES xiii

1 INTRODUCTION 1

1.1 Peak Load Forecasting 1

1.2 Business Needs and Peak Timing Forecasting 2

1.3 Dissertation Outline 6

2 LITERATURE REVIEW 9

2.1 Peak Load Forecasting 11

2.1.1 Methodologies 11

2.1.2 Techniques 21

2.2 Peak Timing Forecasting 31

3 CASE STUDIES 34

3.1 Case Study 1: MeSU 34

3.2 Case Study 2: ISONE 36

4 THEORETICAL BACKGROUND 40

4.1 Multiple Linear Regression 40

4.2 Recency Effect 40

4.3 Logistic Regression 42

4.4 Forecast Evaluation 43

4.4.1 Error Measures for Load Forecasting Problems 43

4.4.2 Error Measures for Classification Problems 43

5 ERROR MEASURES FOR PEAK TIMING FORECAST 45

5.1 Literature Review 45

5.2 Taxonomy of Peak Timing Problems 51
5.2.1 Peak Load Day 51
5.2.2 Peak Hour 53
5.2.3 On-peak Period 54
5.2.4 Peak Shape 56
5.3 Evaluating Peak Timing Forecasts with Traditional Error Measures . 58
5.4 Proposed Error Measures for Peak Timing Forecasting 64
5.4.1 Skill Score 64
5.4.2 Displacement Score 65
5.4.3 Weighted Displacement Error 66
5.4.4 Balanced Accuracy 67
5.4.5 Peak Shape Error 68

6 PEAK HOUR FORECASTING BASED ON HOURLY MODELS 70
6.1 Methodology 70
6.1.1 Benchmark 70
6.1.2 Forecasting Framework 71
6.2 Evaluation of Peak Hour Forecasts 73
6.3 Model Selection for Peak Hour Forecasting 76
6.3.1 Model Selection with Peak Timing Error Measures 76
6.3.2 Model Selection based on Ex-ante Forecast Errors 86

7 PROPOSED MODELS FOR PEAK HOUR FORECASTING 91
7.1 Peak Hour Forecasting: a Classification Approach 91
7.1.1 Modeling for MeSU 92
7.1.2 Modeling for ISONE 99
7.2 A Two-Stage Peak Hour Forecasting Model 105
7.3 Results and Discussion 108

8 CONCLUSION 116

REFERENCES 119

LIST OF TABLES

TABLE 3.1 Summary statistics of load data (in MW) for MeSU 34

TABLE 3.2 Summary statistics of load data (in MW) for ISONE. 38

TABLE 5.1 Demonstration: time series shape error. 47

TABLE 5.2 Summary of error measures used in literature for peak timing forecasting. 48

TABLE 5.3 Demonstration: different scenarios of TP and FP. 61

TABLE 5.4 Demonstration: different class ratios. 64

TABLE 5.5 HSS for different cases with 10 true PLDs in a month. 65

TABLE 5.6 DS values for different values of *de* when $l = 5$ 66

TABLE 5.7 BA scores for scenarios with three hours on-peak period. . . . 68

TABLE 5.8 Summary of proposed peak timing error measures. 69

TABLE 6.1 Peak hour displacement errors of the benchmark model. 74

TABLE 6.2 Peak hour displacement errors of the benchmark model for PLDs. 75

TABLE 6.3 MAPE and wDE values for *top level* (MeSU) on the validation data (year 2016). 78

TABLE 6.4 MAPE and wDE values for ISONE on the validation data (year 2016). 79

TABLE 6.5 wDE values from M0, and their comparison with those from the seven tracks on the validation data (year 2016) of MeSU. . . . 81

TABLE 6.6 Performance of models selected from the seven tracks for ex-post forecasting on test data for MeSU. 82

TABLE 6.7 Performance of models selected from the seven tracks for ex-ante forecasting on test data of MeSU. 83

TABLE 6.8 Performance of models selected from the seven tracks for ex-ante forecasting on test data of ISONE. 84

TABLE 6.9 Performance of models selected based on temperature forecast on test data of MeSU. 87
TABLE 6.10 Performance of models selected based on temperature forecast on test data of ISONE. 88
TABLE 6.11 *d*-*h* pairs selected from the two methods for the validation year 2016. 88
TABLE 7.1 M2 base model validation results, MeSU. 94
TABLE 7.2 M2 validation results based on models trained on probable peak hours only, MeSU. 96
TABLE 7.3 M2 validation results with daily temperature features added, MeSU 97
TABLE 7.4 M2 validation results with the indicator of daily maximum temperature hour added, MeSU. 99
TABLE 7.5 M2 base model validation results, ISONE. 100
TABLE 7.6 M2 validation results based on models trained on probable peak hours only, ISONE. 102
TABLE 7.7 M2 validation results with daily temperature features added, ISONE. 103
TABLE 7.8 M3 validation results, MeSU. 107
TABLE 7.9 M3 validation results, ISONE. 108
TABLE 7.10 wDE values for ex-ante forecasts on test data of MeSU. 109
TABLE 7.11 wDE values for ex-ante forecasts on test data of ISONE. 110
TABLE 7.12 Comparing the distribution of *de* on PLDs from the three models. 113
TABLE 7.13 Sum of wDE values from ex-post forecasting, and sum of wDE values from ex-ante peak hour forecasting for the test data of the two case studies. 114

LIST OF FIGURES

FIGURE 1.1 Levels of daily electricity demand. 1
FIGURE 1.2 Peak load forecasting use cases. 3
FIGURE 1.3 Demand side management programs. 4
FIGURE 1.4 Demonstration: DR deployment. 5
FIGURE 1.5 Demonstration: charging and discharging scheduling of ESS. . 5
FIGURE 1.6 Outline of dissertation. 8
FIGURE 2.1 Number of publications on "peak electric load" or "peak demand forecast" in WoS. 10
FIGURE 2.2 Publication compiled over the forecasting horizon. 11
FIGURE 2.4 Publications compiled over techniques used across different forecasting horizons. 22
FIGURE 3.1 Load history of Top Level under MeSU. 35
FIGURE 3.2 Temperature series from 28 weather stations under *top level* of MeSU. 35
FIGURE 3.3 Peak hour distribution for *top level* of MeSU in different months of the year 2016. 36
FIGURE 3.4 Load zones under ISONE. 37
FIGURE 3.5 Peak hour distribution for ISONE in different months of the year 2016. 39
FIGURE 5.1 Demonstration: scenarios with different peak timing errors but same d-RMSE. 47
FIGURE 5.2 Load duration curve. 52
FIGURE 5.3 Number of PLDs in the *top level* of the two case studies. . . . 53
FIGURE 5.4 Peak hours on PLDs in different zones of ISONE 54

FIGURE 5.5 DR-deployment demonstration: a) event called earlier than the actual peak period, b) event called later than the actual peak period. 55
FIGURE 5.6 On-peak period demonstration. 56
FIGURE 5.7 Demonstration: load shape forecast for generating ESS discharging profile. 58
FIGURE 5.8 Daily MAPE vs. TPR. 59
FIGURE 5.9 Examples when daily MAPE and peak hour forecast accuracy are contrary. 60
FIGURE 5.10 Comparing BA, HSS, and TSS for different combinations of TPR and TNR. 63
FIGURE 5.11 Comparing wDE values with DS values for $l = 5$. 67
FIGURE 5.12 Demonstration: BA for on-peak period. 68
FIGURE 6.1 Rolling window for three test years. 71
FIGURE 6.2 Framework for day-ahead load forecasting. 72
FIGURE 6.3 wDE values from ex-ante forecasts vs. wDE values from ex-post forecasts vs. theoretical best wDE values from Recency effect framework. 75
FIGURE 6.4 Comparing *d-h* pairs selected with MAPE and *d-h* pair selected with wDE. 77
FIGURE 6.5 *d-h* pairs selected for each year for *top level* of MeSU by different tracks. 85
FIGURE 6.6 Vertical displacement errors and horizontal displacement error on a day under *top level* of MeSU (1/13/2017). 86
FIGURE 6.7 Boxplot of temperature forecast errors for (a) MeSU, and (b) ISONE. 89

FIGURE 6.8 Time series of temperature forecast errors for (a) MeSU, and (b) ISONE. 90
FIGURE 7.1 Heatmap showing the peak hour distribution over a year, with darker shade showing the high demand hours of a day (*top level* of MeSU, year 2016). 92
FIGURE 7.2 Probable peak hours of *top level* of MeSU (year 2016) 95
FIGURE 7.3 Heatmap showing the peak hours for PLDS for the three years of *top level* of MeSU. 95
FIGURE 7.4 Scatterplots of a) timing of maximum temperature of day vs. peak hours in summer, b) timing of minimum temperature vs. peak hours in winter (*top level* of MeSU, year 2016) 98
FIGURE 7.5 Scatterplot of daily average temperature vs. peak hour for ISONE (year 2017). 100
FIGURE 7.6 Probable peak hours of ISONE (year 2017). 101
FIGURE 7.7 Scatterplots of a) timing of maximum temperature of day vs. peak hours in summer, b) timing of minimum temperature vs. peak hours in winters for ISONE (year 2017). 103
FIGURE 7.8 Framework of the model M3. 106
FIGURE 7.9 Scatterplots of daily maximum temperature vs. daily minimum temperature grouped by peak hours for (a) PLDs and (b) non-PLDs (*top level* of MeSU, year 2016). 112
FIGURE 7.10 Scatterplot of load vs. temperature for a) MeSU, b) ISONE (year 2016). 115

LIST OF ABBREVIATIONS

AR-Recency	Auto-Regressive Recency effect Model
BA	Balanced Accuracy
de	Displacement Error
DER	Distributed Energy Resources
DR	Demand Response
DS	Displacement Score
DSM	Demand Side Management
ESS	Energy Storage System
GEFCom2012	Global Energy Forecasting Competition 2012
HSS	Heidke Skill Score
MAPE	Mean Absolute Percentage Error
MLR	Multiple Linear Regression
PF	Peak Load Forecasting
PLD	Peak Load Day
SAE	Shape Absolute Error
TPR	True Positive Rate
wDE	Weighted Displacement Error

CHAPTER 1: INTRODUCTION

1.1 Peak Load Forecasting

Peak electric demand forecasting is one of the most critical tasks in electricity supply and demand management. Generating electricity to meet demand happens very instantaneously. Utilities face the challenge of meeting electricity demands while maintaining efficient operations and preventing system failures. A range of generator types helps keep the supply in balance with demand at various demand levels (Figure 1.1). Base demand is met by low-cost generators such as nuclear plants, which have limited ramping capabilities. On the other hand, peak demand must be met by generators with rapid ramping capabilities to respond quickly to significant variations in demand. Nevertheless, peak load generators have a very high marginal cost [1]. Thus, accurate peak load forecasting is vital to maintain a reliable and stable electricity supply, optimize resource allocation, and prevent blackouts or brownouts.

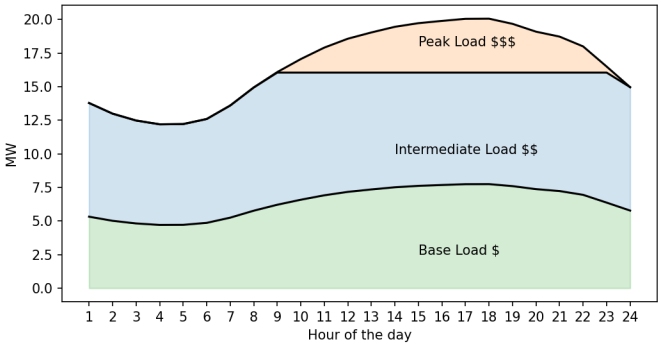


Figure 1.1: Levels of daily electricity demand.

Peak load represents the highest levels of electricity demand experienced by a power system, typically occurring during a specific time of the day, week, month, or year [2]. The definition of peak load is determined by the data sampling rate and data

aggregation level. For example, daily peak load is defined as the maximum demand of a day for an hourly interval or sub-hourly interval. Annual peak refers to the maximum demand in a year at a specified interval, which could be monthly, hourly, or sub-hourly. In other words, the peak load of a system, be it an appliance, home, building, distribution, or transmission network, is its maximum interval demand for a given period.

1.2 Business Needs and Peak Timing Forecasting

Peak load forecast is leveraged by different stakeholders, including utilities, regulatory commissions, and trading firms, for various solutions [3]. The annual peak forecast is needed one to several years ahead for future investment decisions and resource planning. Overestimation of the peak demand results in excessively high reserves. On the other hand, under-forecasting may result in blackouts and brownouts. The risk cost of oversizing or undersizing a power plant is quite high. One study estimated that a 1% forecast accuracy for a 1-gigawatt peak plant can be valued at $500,000 per year [4]. Several regulatory authorities also review these long-term forecasts [5]. Utilities require daily peak demand forecasts a year in advance for maintenance scheduling, grid planning, and tariff structuring. A seasonal peak load forecast at the transmission system is necessary for a seasonal assessment of the appropriateness of the resources. Day-ahead peak forecasts are a critical input for daily grid operation. The task of grid operators is to meet the demand at a minimum cost while satisfying operation constraints. Furthermore, the day-ahead peak forecast is an essential tool for electricity pricing and market trading to optimize hedging strategies.

These different applications call for different time frames over which forecasts are to be made. Using the categorization of the forecast horizon mentioned by [6], the peak load forecast can be categorized into short-term, medium-term, and long-term. A forecast horizon of up to two weeks ahead can be categorized as short-term peak load forecasting, one month to up to 3 years ahead can be classified as medium-

term peak load forecasting, and several years ahead can be called long-term peak load forecasting. Figure 1.2 outlines the peak load forecasting categorization and its applications at different spatial and data sampling levels.

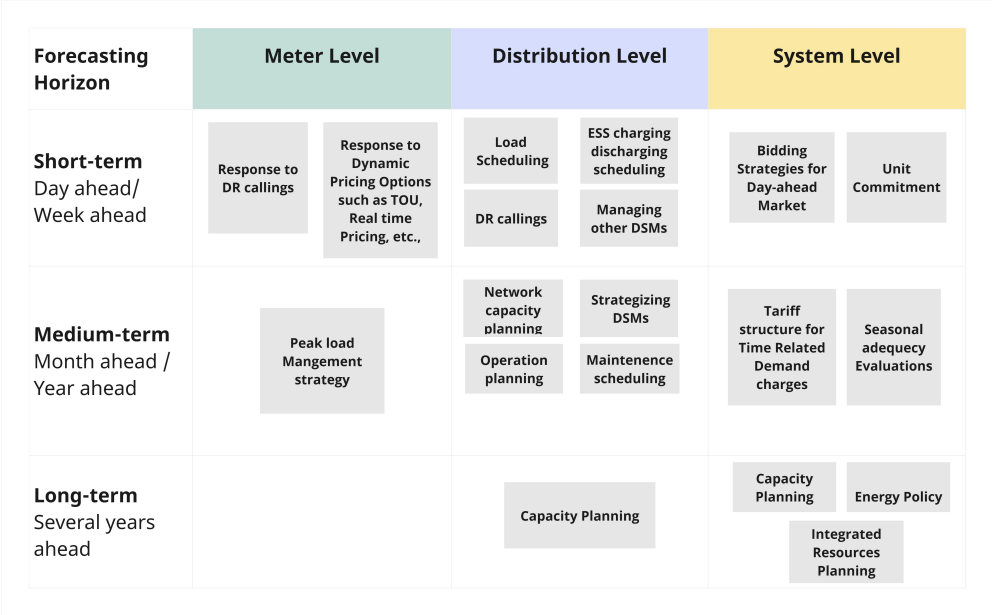


Figure 1.2: Peak load forecasting use cases.

In recent years, peak load forecasting has become increasingly critical due to the evolving grid. Factors such as the integration of distributed energy resources (DER) [7], the drive toward heating electrification [8], and the increasing adoption of electric vehicles [9] have added complexity to the task of managing peak demands. Moreover, the industry is aiming to improve the load factor by achieving an even use of electricity over time. Three key solutions have been initiated: demand-side management (DSM), energy storage system (ESS) integration [10], and vehicle-to-grid (V2G) integration [11]. Since the early 1980s, DSM tools have been deployed to accomplish different load-shaping objectives, such as peak shaving, load shifting, and valley filling. [12]. At the same time, ESS and V2G integration provide auxiliary sources to balance the grid.

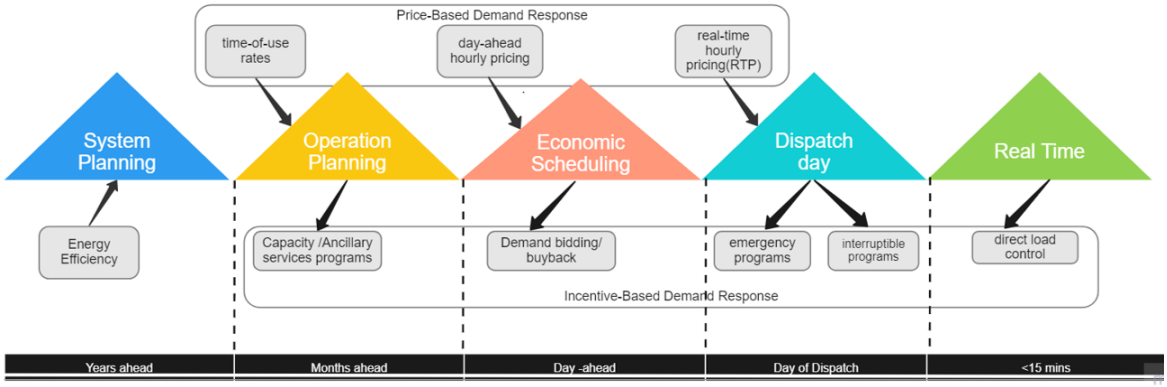


Figure 1.3: Demand side management programs.

Some of the primary tools of DSM are energy efficiency programs, Price-based demand response, and Incentive-based demand response [13] as depicted in Figure 1.3. The Demand Response (DR) programs work by load curtailment and load shifting, as illustrated in Figure 1.4. The DR events are called on days when the demand is expected to be very high. Some DR programs are dispatchable, where the operators can control the appliances for a customer in the event of extreme peak demand. Other DR programs are time-sensitive pricing, and incentive-based event calls that require customer response. Events are called at specific times of day when peak demand is likely to occur. Hence, for this purpose, the peak load is not confined to a single point but spans a period of instances. A suboptimal schedule for calling the event could lead to peak rebounds due to a shift in demand to the actual peak hours. An underestimated peak load forecast may under-credit the DR load-reduction benefits [14]. Therefore, the success of DR deployment depends on forecasting the days of expected high demand and also the time when it would occur on the day.

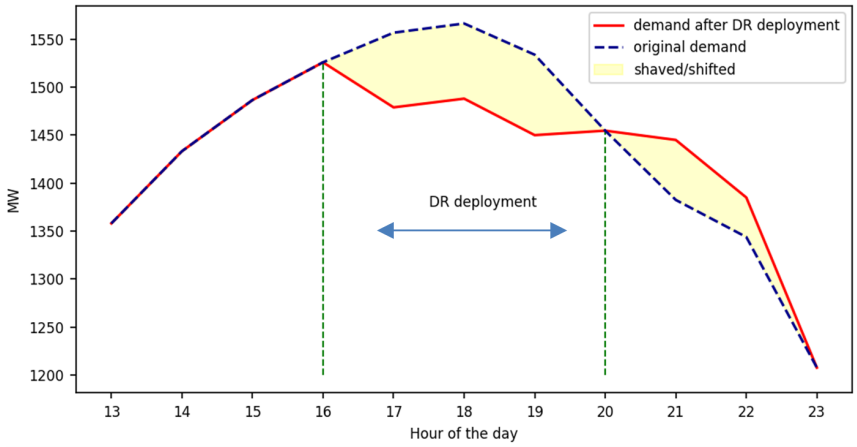


Figure 1.4: Demonstration: DR deployment.

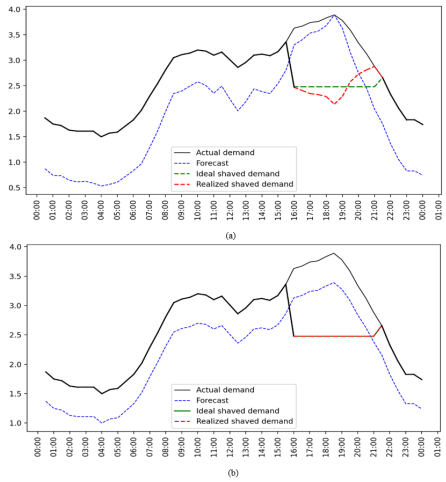


Figure 1.5: Demonstration: charging and discharging scheduling of ESS.

With the integration of renewables such as photovoltaic (PV) systems, matching dynamic electricity demand with intermittent energy production is a major challenge. Deploying ESS is one solution. However, scheduling of charging and discharging profiles for ESS necessitates forecasting the peak timing as well as the shape of the peak period, as shown in Figure 1.5. A poor forecast will undermine the peak shaving opportunity. It may also lead to substantial loss in battery storage capacity, which decreases their lifecycle and increases energy costs [15].

Peak load forecasting problem is multi-dimensional. The errors of peak load forecasts can be decomposed into magnitude, displacement, duration, and shape errors.

The magnitude error reflects vertical deviation in the forecast of peak load level. The displacement error reflects the horizontal deviation in the timing of the peak forecast. Instead of confining the peak to one point, some business applications require identifying the period of peak demand. Hence, the third component is the duration error which refers to the error in the length of the peak period forecast. Peak shape error refers to deviation in the shape of the predicted load curve during the peak period. Unlike magnitude error, the shape error is independent of the level of peak load forecasts. These four dimensions can provide a holistic view to optimize the grid scheduling in the modern grid.

The peak timing forecasts can be deterministic or probabilistic. A deterministic forecast would give the hour of day when the peak demand is most likely to occur. In contrast, a probabilistic forecast would quantify the uncertainties of the forecast by giving the probability of peak occurrence for each hour or day. Decisions for applications such as grid scheduling, DSM, ESS scheduling, and day-ahead market hedging methods are often made a few hours to a few days in advance. Therefore, for a short forecasting horizon, the forecast should be usable and deterministic, so that it can be consumed. Nevertheless, probabilistic peak time forecasts are useful for long-term applications such as tariff designing, maintenance scheduling, and planning for peak load control programs. This research focuses on the deterministic forecast of peak timing for the short-term forecasting horizon.

1.3 Dissertation Outline

Figure 1.6 depicts an outline of the dissertation. First, in Chapter 2, we thoroughly review the literature to understand the work on peak load and timing forecasting. Chapter 3 introduces the case studies used in this research and conducts some exploratory data analysis. Chapter 4 introduces some fundamental concepts on techniques used and forecast evaluations. All forecasts require a proper evaluation. Therefore, we also explore the literature for error measures used in peak timing

forecasting in Chapter 5. In this chapter, we also establish the taxonomy of peak timing problems. Motivated by our findings from the study, we propose some new error measures for evaluating peak timing forecasts in the conclusions of the chapter. In Chapter 6, we present our work on model selection for peak hour forecasting. In Chapter 7, we propose the two new models we developed for peak hour forecasting, discussing their test results and evaluating their strengths and weaknesses. Chapter 8 concludes this research.

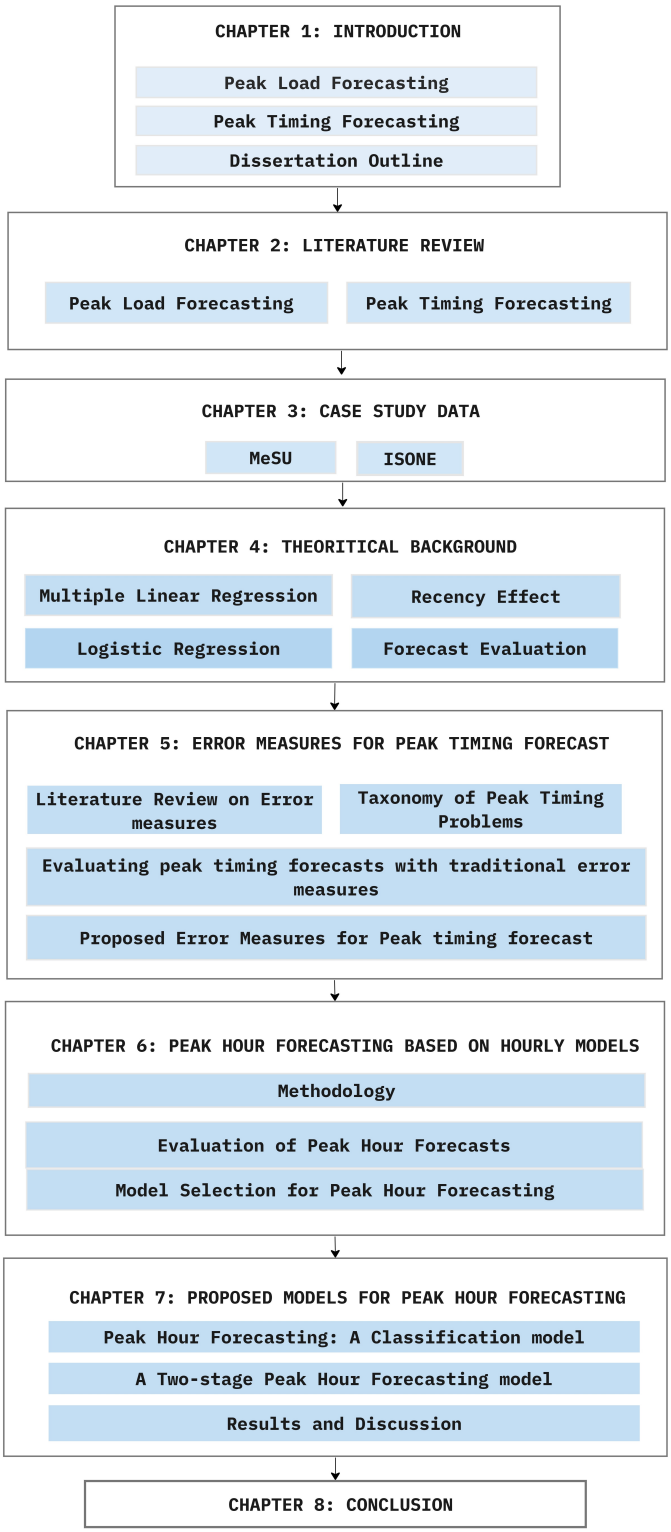


Figure 1.6: Outline of dissertation.

CHAPTER 2: LITERATURE REVIEW

In this chapter, we will review the representative academic papers on peak load forecasting providing a conceptual review of different techniques and methods that have been used. Peak load forecasting is a subdomain of load forecasting with extensive literature, although not as rich as hourly load forecasting. To avoid a verbose presentation, it is to be noted that the rest of this chapter uses the term "PF" to refer to "Peak electric load forecasting" or "Peak electric load forecast".

A bibliometric study was performed on Sep 23, 2023, to get an overview of existing research in peak load forecasting using the well-established and recognized Web of Science (WoS) database. The query used to retrieve the records is as follows: "TS=("peak load forecast" "peak demand forecast") and Article (Document Types)". Figure 2.1 depicts the number of journal articles on peak load topics from 1990 onwards. There have been 326 publications since 1990. Hence, PF has been a topic of study for at least half a century, and yet no literature review focusing on the subject has been published yet.

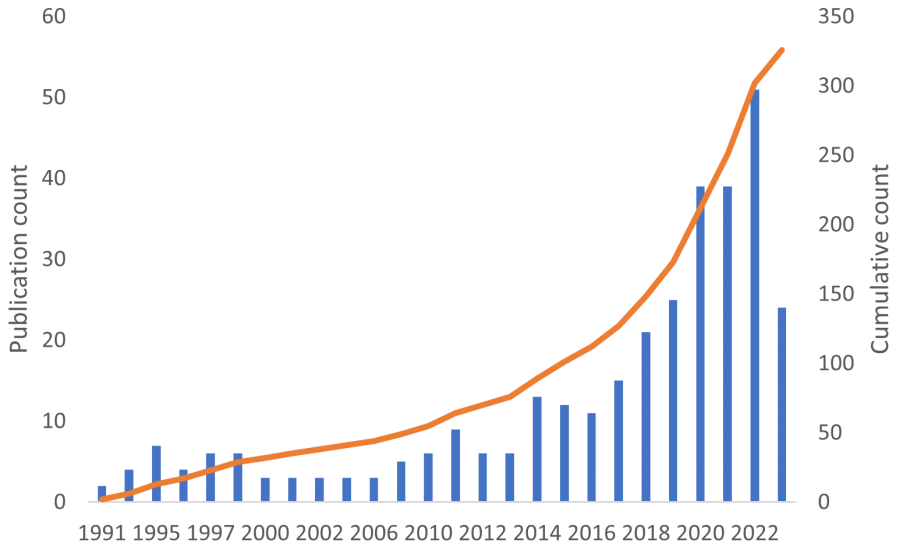


Figure 2.1

Figure 2.2 shows the publications in the fields of PF compiled according to the forecasting horizon. Short-term PF has been the focus of most published work with an increasing trend. However, long-term and medium-term forecasting has also been stepped up in recent decades. Most of the work on long-term peak forecasting is probabilistic, presenting PF in density, quantiles, or intervals.

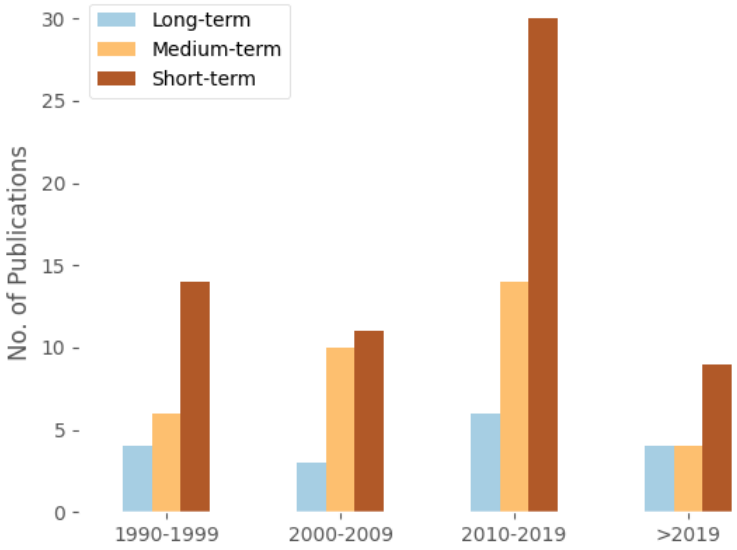


Figure 2.2: Publication compiled over the forecasting horizon.

2.1 Peak Load Forecasting

2.1.1 Methodologies

From the literature review, we find that PF methods are broadly divided into two categories. From the literature, we find that PF methods are broadly divided into two categories. The first method is "Derived Peak Forecasting", which involves forecasting the demand for each time step (hourly/sub-hourly) within a period and then identifying the highest value among all these forecasts as the peak forecast. For example, the daily peak forecast can be derived from the 24 hourly load forecasts. This method can be used to forecast the magnitude as well as the time of peak demand from the same model. The second approach is "Direct Peak Forecasting". It is a time-independent method, where a model directly forecasts the magnitude without specifying the exact hour when the peak will occur. It can also forecast peak timing without determining the magnitude of peak demand. Such models may only require low-resolution data, such as daily frequency samples. Figure 2.3 shows an

example demonstrating the two forecasting frameworks. A similar classification of peak forecasting methodology was adopted by [16].

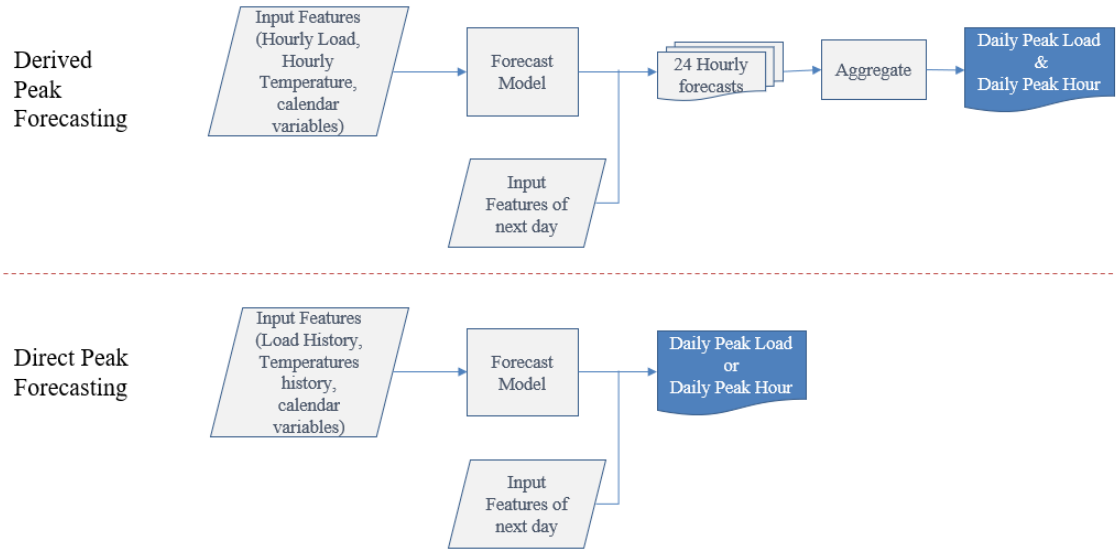


Figure 2.3: An example of the two peak load forecasting methods.

Most publications on peak load forecasting are based on the Direct forecasting of peak load. However, in practice, PF are mostly generated from hourly load forecasts. Using the higher resolution data to derive the peak forecasts has some advantages. There is no loss of information arising from temporal aggregation. Using low-resolution data to forecast annual or monthly peaks would result in a small sample size and limited information. Therefore, the bottom-up approach is a popular way to get peak forecast distributions [2], [17], [18], [19], [20]. However, since the hourly load data is noisy, this approach has some drawbacks too. A study was carried out to understand how the maximum of hourly forecast compared to the peak load forecast from direct models [21]. The paper concluded that the aggregated peak forecasts from the hourly forecasts were negatively biased, and the peak load forecasts obtained directly were closer to the actual peak. This is one of the few studies that have compared the peak load forecasts obtained from the derived model with the one

obtained from the direct model.

Not all the papers that have reported peak forecasting accuracy, were devoted to peak load forecasting. In such studies, the proposed hourly load forecasting model was additionally evaluated for peak forecast by comparing the maximum forecast value for a period with actual maximum value [22], [23], [24]. Assuming the peak is typically observed at a certain time, some studies have evaluated the accuracy of forecasts at that time [25]. A Neural Network model was developed to forecast next-hour load using historical temperature and historic load from the Wisconsin Electric Power Company. The peak load was evaluated at a specific hour (7 PM) when the actual peak was usually observed. In another study, the proposed wavelet decomposition-based model was evaluated for a five hour period representative of peak and off-peak hours [24]. In another study, the models were evaluated at the time of actual peak occurrence [26].

Irrespective of the focus of the study, this research reviews the academic papers that have evaluated peak hour or peak timing forecasts. In the following subsections, we will review representative literature on PF to understand how temperature and other explanatory variables have been used for derived as well as direct PF.

2.1.1.1 Temperature Effect

Of all meteorological factors, the temperature has the strongest impact on electricity demand [27] and [28]. Besides, it is also the most predictable and easily available data at hourly intervals, of all the weather variables. Most studies that forecast the daily peak load using direct PF models employ either of the three aggregated forms of hourly temperatures: daily average temperature, daily maximum temperature, and daily minimum temperature. Table 2.1 presents a summary of the temperature features used by some representative papers. The most widely used among these is the daily maximum temperature together with the daily minimum temperature. Using only the daily average or daily maximum temperature is another common approach.

Hence, the effects of temperature on the daily peak load are often captured using the statistics of the daily temperature curve. In practice, weather forecasts beyond two weeks are typically not very accurate [29]. Hence, temperature features used in medium-long forecasting horizons are modeled by generating scenarios. We shall review the methods adopted to model temperature effects grouped by forecasting horizon.

Table 2.1: Summary of temperature variables used in direct peak load forecasting in literature.

	Daily Average Temperature	Daily Maximum Temperature	Daily Minimum Temperature	References
1	Y			[25], [30], [31], [32], [33], [34], [35], [36], [37], [38],[39]
2		Y		[40], [41], [42], [43], [44], [45], [46], [47]
3	Y	Y		[48], [49], [50], [51]
4		Y	Y	[52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64]
5	Y	Y	Y	[65], [66], [67], [68]

(A) Short-term forecasting

Some earlier research reported using Cooling Degree Days (CDD) and Heating Degree days (HDD) [21], [30]. CDD and HDD are obtained by calculating the departures of daily average temperature values from the baseline temperature.The baseline temperature for CDD and HDD can be determined by analyzing historical load-temperature correlations and physical considerations [21]. The statistical analysis of various regression-based models concluded that adding the CDD and HDD significantly reduces the bias and standard error of the daily peak forecasting model

[16].

In another approach, new temperature variables were developed to represent the non-linear relationship between temperature and peak load. A transformation function was proposed to model the temperature to treat heating and cooling loads on the same regression plane [30]. The function was designed to factor in the seasonal load changes as well as the latest weather-load characteristics. The new transformed variables were then used in the regression framework to forecast the daily peak load.

The accumulation effect of temperature due to sustained high temperatures on hot summer days was modeled by feature engineering [31]. The modified temperature variable was formed as a function of the daily maximum temperature of the current day, the maximum temperature of the previous three days, and the average daily maximum temperature of the last 30 days. The function parameters were optimized through a Genetic Algorithm (GA) to maximize the co-variance of the daily peak load and the modified temperature.

Temperature effect can be modeled by splitting the temperature range into intervals and developing a different local model for each interval [32]. Hence, based on the maximum temperature of the forecast day, the model is chosen. This approach could help deal with the uncertainty of temperature forecasts because it is not necessary to know the exact next day's temperature. However, the impact of a temperature forecast that is off by more than one interval from the actual temperature range was not examined.

The effect of hourly temperature on daily peak has not been investigated so far in any studies. The typical approach is using the daily temperature features in the Direct PF. The hourly temperatures preceding the peak demand can be key determinants. However, not knowing the timing of peak demand, it is not realistic to get these features. Exploring the relationship between the timing of peak demand

and the timing of the day's maximum or minimum temperature could help in understanding how can we leverage the sequence of hourly temperatures preceding the peak demand in the forecasting model. An alternative approach could be a multi-resolution approach [33]. The proposed framework included the vector of half-hourly temperatures as model input to forecast the daily electrical load peak magnitude and timing directly. Hence, instead of explicit feature engineering of hourly temperature, they are treated via functional smooth effects in the Generalized Additive Model. This approach could avoid losing any information and yet retain the parsimony of the low-resolution approach.

In practice, the actual observed temperature is not available for the future. The models estimated with actual temperatures do not account for temperature forecast errors. Some studies propose techniques to reduce the model's sensitivity to temperature forecast errors. These include adding a noise component to the actual temperature series to simulate temperature forecasts [21] or forecasting the day-ahead temperature by fitting low-order autoregression to the daily temperature series [16].

(B) Medium-term & Long-term Forecasting

As weather forecasts for more than two weeks ahead are often unavailable and quite uncertain [29], the approach of using temperature for medium-term and long-term forecasting differs. Some studies have also presented approaches that do not include any weather variables in the modeling. A Support Vector Machine (SVM) model was developed to forecast daily peak load for a month ahead [34]. The proposed methodology was the winning solution of the EUNITE 2001 competition, where temperature data for the forecast period was not provided. Instead of estimating the temperature, the time series modeling scheme was used. In a similar approach an Artificial Neural Network (ANN) based model is approached to forecast month-ahead daily peak load using lagged values of daily peak load and calendar variables as input variables [35].

In another approach annual maximum temperature was used to forecast the annual peak [36]. The study used the historical annual data from 1981 to 2000 to forecast the annual peak load of Taiwan for the years 1997 to 2000.

Most studies on long-term and medium-term peak load forecasting, however, use simulated temperatures for the forecasting period and generate probabilistic forecasts. A probabilistic model was proposed to generate one year ahead daily peak load distribution for the winter season in Central England and Wales [37]. Weather forecasts were generated for the next year by fitting a Bayesian model based on multivariate normal distribution generated from 17 years of historical weather.

A novel seasonal bootstrapping method was proposed to generate temperature simulations [2]. The proposed methodology was empirically demonstrated using a case study from the South Australian region to generate probability distributions of half-hourly peak demand for up to ten years. The distributions of annual and weekly peak demand forecasts were then derived. Regression splines were also used to model the impact of temperature in the proposed semi-parametric model.

Temperature scenarios were generated by using data from 30 years of history to generate the probabilistic forecast of monthly peak load [5]. The base, aggressive, and conservative macroeconomic scenarios were assumed for the test year generating 90 scenarios. The median, 10th, and 90th percentile of the monthly peak forecast were obtained and evaluated against the actuals. The proposed method was verified using a case study from one of the largest electric generation cooperatives in the U.S.

Simulated weather data was used to generate probabilistic forecasts of both the timing and magnitude of the annual peak demand in a province of the Netherlands [19]. Simulated weather data were produced by constructing surrogates of the original temperature, wind speed, and luminosity using Fourier transformation. A 1000 simulation runs generated the distribution of electricity demand forecasts for each day

between 1995 and 2000. The forecast errors were compared with the ex-post forecast errors.

The uncertainties of future daily maximum temperature were modeled by calibrating the physics-based climate models [38]. The data collected from this climate model projection were re-adjusted to account for regional changes by employing the generalized extreme value (GEV) distribution. The methodology was proposed for long-term daily peak load forecasting for the region of Texas. The model was validated by comparing the load densities forecast with actual histograms for the test data.

A probabilistic annual peak demand forecasting methodology based on summer months data from South Australia was proposed [39]. The proposed half-hourly demand model incorporated both deterministic and stochastic components. The deterministic component included modeling the cyclic features of time series, and the relationships with current temperature. The stochastic part included the error component, which was further modeled with autoregressive moving average (ARMA) processes. Annual peaks were generated by taking the highest half-hourly value from each of the 500 generated summers. The empirical CDF of this variable gave the distribution of the annual peak.

2.1.1.2 Other Explanatory Variables

Even while daily energy use may not change much, peak demand varies depending on when the energy is being used. Hence, human activities substantially impact peak load, resulting in yearly, monthly, and daily cycles. A majority of research has modeled these patterns by employing dummy variables for hours of the day, days of the week, months of the year, weekends, and public holidays. Besides, some studies have used the calendar variables to subgroup the data. A smooth function estimated with cubic regression spline was used to model the summertime effect [2]. A separate

model was fitted for the data from each half-hourly period. Different models for different days of the week in every season were developed [40]. Data resampling based on different weekdays was done to train five different models [41].

Various socioeconomic information, such as a forecast of GDP and demographics, have also been incorporated for long-term forecasting. Economic and demographic variables such as GSP chain volume estimates and average electricity price were considered in the long-term forecasting model [2]. The authors modeled these effects as a linear term with the same coefficient for each period since economic relationships are typically weak and change slowly. These annual effects were later combined with the half-hourly models. The long-term GSP forecasts and pricing scenarios were assumed. The model generated the density forecast for yearly and weekly peaks for up to 10 years using simulated temperatures, known calendar effects, and simulated residuals.

Socioeconomic variables that affect the future demand for electricity, such as GVA (Gross Value Added) by sector and the number of households were modeled to generate long-term forecasts of hourly electricity demand [17]. Macroeconomic scenarios were created and then converted into annual growth possibilities for electricity demand. These annual values were then transformed into daily values using the Boot-Feibes and Lisman (BFL) disaggregation method. Temperature simulations were added to the model along with calendar effects. Annual peak loads, peak loads in winter and summer, and monthly peak loads were recorded for each scenario to create density forecasts.

Long-term forecasting problems caused by any influence or disruption of peak load value, such as peak reduction tariff system and load transfers were addressed in a study [42]. The proposed methodology included weather normalization and modeling peak load reduction by S-curve. With the increasing adoption of various peak-shaving measures, backcasting curtailed load is another challenge added to the load forecasting

problem.

2.1.1.3 Feature Selection

With the vast pool of variables to select for modeling, many authors have reportedly used feature extraction and feature selection algorithms. Feature extraction is the process of transforming high dimensional data to lower dimension by changing the representation of the original data in a way that maximizes the variance of the data in the new dimensional representation. The process yields a new set of dimensions called principal components, removing the redundancy of data but also losing interpretability. Feature extraction on 28 variables, including weather variables and calendar variables, was performed using the Principal Component Analysis to extract 11 factors [40].

In contrast to dimension reduction methods, the feature selection process reduces the number of features, by choosing only relevant features and identifying redundant or non-relevant features. A cross-validation approach was used for feature selection [2]. After separating the data into training and validation sets, the input variables are chosen based on the features giving minimum prediction errors on the validation period. The Pearson correlation index was used to select input features that were highly correlated to peak load and independent of one another [43]. Various feature selection methods were examined on different models [44]. Some of them included utilizing the lasso penalty and ridge penalty to penalize the highly correlated predictors, mutual information-based feature selection method, and inference on different permutations of features.

The joint distribution of load and exogenous variables were analyzed to determine the upper tail dependence [45]. The peak load indicative variables were obtained from the bivariate distribution to increase the peak load forecasting. The Random forests algorithm was used to examine the candidate parameters in terms of their influence on the prediction response using the method [22]. The model was trained with dif-

ferent permutations of variables and analyzed for errors to assess the importance of variables in terms of their impacts on the prediction response. The selected features were the time indicator, the HVAC operation schedule, outdoor air temperature, and relative humidity. In another study, the backward feature elimination algorithm was used to perform the input variable selection [46]. The model was first trained using all candidate variables to compute the variable importance for all the variables. Subsequently, the variable with the lowest rank was removed. This was done until the model performance stopped improving. A fuzzy-based feature selection was used to select only relevant features [47]. Gaussian membership function was chosen for the proposed fuzzy-based feature selection method.

Changes in lifestyle, such as the adoption of EVs, up-gradation of household appliances, changes in the HVAC system, etc., change the magnitude and timing of peak demand. Although daily energy use may stay relatively similiar, peak demand fluctuates depending on when the energy is being used. The changes in load profile at lower-level aggregates impact the load profiles at higher levels. These trends are difficult to model due to the scarcity of training data.

2.1.2 Techniques

With time, various techniques and their combinations have been tried out for peak load forecasting, broadly categorized into (a) statistical methods, including Multiple Linear Regression (MLR), time series models, distribution fitting, and Bayesian estimation; (b) Artificial Intelligence (AI) models including neural networks (NN), fuzzy systems, SVM, deep learning networks; and (c) hybrid methods. Figure 2.4 shows the numbers of some notable journal papers using different techniques in different forecasting horizons. Most of the existing research in this area has primarily focused on short-term load forecasting using AI. For long-term forecasting, however, most of the representative papers have used statistical methods such as distribution fitting and MLR. This could be because long-term forecasts are often derived in the form

of probabilities by generating numerous scenarios. Training so many AI models is computationally difficult.

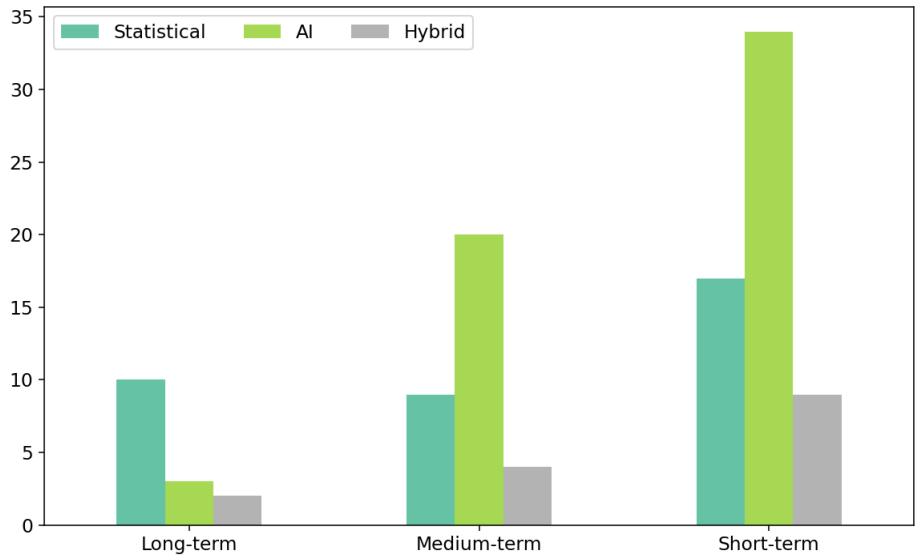


Figure 2.4: Publications compiled over techniques used across different forecasting horizons.

2.1.2.1 Statistical

In statistical models, the relationship between dependent and independent variables is estimated. These models broadly fall into two categories: univariate models (such as exponential smoothing, ARMA, ARIMA, etc.) and multivariate models (such as MLR, additive models, etc.). Univariate models use past values of time series and, sometimes, historical values of the error term to forecast future values. With the multivariate models, the relationship between the dependent variable and other independent variables is estimated. These techniques are data-driven, easier to interpret, and typically less computationally expensive than AI.

MLR analysis is one of the earliest and most widely applied techniques for short-term load forecasting [48]. A short-term forecasting model based on a linear regression-based model, implemented in the Pacific Gas and Electric Company was detailed out

[21]. Temperature effects, calendar factors, and holiday effects were modeled in the proposed regression model with weighted least square estimation. The weights were calculated using the forecast errors of training data to make the model robust against outliers. The paper also described the direct peak load forecast model based on linear regression.

A statistical diagnosis of various regression models to forecast the daily peak load applied was presented [16]. The variables were systematically added to the model, and the diagnostic tests were examined to evaluate the effectiveness of different variables using a case study from a utility in Michigan. An MLR-based model developed for a real-time forecasting competition organized by Puget Sound Power and Light Company was described [49]. They developed hour by hour model to forecast hourly load. The morning peak forecasts and evening peak forecasts were derived by aggregation. The regressors used in their model were calendar variables, holiday effect variables, temperature variables, and the previous day's load variables. They further tested an adaptive version of their model, where the forecasts were exponentially smoothed based on the previous day's forecast errors.

A transformation function of temperature to model the non-linear relationship between maximum temperature and peak load was proposed using MLR technique [50]. The model captured the baseload growth and weather-sensitive load growth for forecasting daily peak load. The proposed method was tested using the Tokyo Electric Power Company's actual load data. [30] proposed two methods based on MLR for same-day and next-day peak load forecasting. Both approaches used the training data from a similar period in previous years. The first approach models the relationship between the daily peak load and the morning reference load for same-day forecasting and removes the trend. The de-trended load is regressed against the weather variable to get the forecast. In the second approach, the trend part was explicitly modeled in an MLR model. The variable transformation was used to estimate changes in load

characteristics for past years relative to the forecast year.

A functional linear regression model was proposed using a case study from Spain for heating demand [51]. The proposed short-term PF model used the load curve of the previous day as the regressor to forecast the daily peak load. Furthermore, the training data was clustered into groups using the k-means algorithm. Thus, a family of functional regression models was obtained.

Time series analysis is also commonly used for PF, as peak load is sequential data with correlated values. An ARIMA model was proposed for the daily peak load time series [20]. The author extended the ARIMA model to incorporate exogenous variables such as temperature and an initial estimate from the operator. A time series model was proposed using a case study from the Great Britain National Grid [52]. The weekly cycle of daily peak load time series was mapped into the polar coordinates to form the elliptic-orbit model. Many studies have also applied Time Series decomposition. The load series was decomposed to get the non-linear trend of the hourly loads via polynomial fit [20]. The single Value Decomposition (SVD) method was applied to the de-trended series to extract the cyclic and the random components. The latter was fitted to a normal distribution of zero means.

Generating peak load probabilistic forecasts by fitting a distribution is another statistical approach in the literature. Exponential distribution was assumed to generate peak load probability density function [53]. Later, the study was extended by assuming multiple distributions such as univariate gamma, beta, and triangular distributions [18]. In another work, the lognormal distribution probability density function was used to abnormal peak load [54] to develop a probabilistic load forecasting model for a building. Normal peak load and abnormal peak load were identified. Probabilistic peak forecasts for normal peak load were generated by considering uncertainties in weather forecasts. Long-term peak load growth of a distribution network was represented by the S-curve [55]. The historical data was fitted to an S-curve at

different spatial levels. The hierarchical forecasting modules reconciled the forecasts at different levels using bottom-up and top-down allocation.

A short-term electrical load forecasting scheme based on Bayesian estimation was described [56]. The model relied on the recursive use of Bayes' rule to generate a forecast. A Bayesian approach was also presented for estimating half-hourly load with separate regressions for each half-hour [24]. The Markov Chain Monte Carlo (MCMC) algorithm was developed to obtain the forecast distribution of hourly load and daily peak load from the posterior distributions of the parameters. Using a radial basis, they used the functional form of a bivariate temperature and humidity effect. The trend was captured using a quadratic trend variable. They also modeled the error using a first-order vector auto-regression exploring the intra-day autocorrelation structure of the errors.

2.1.2.2 Artificial Intelligence

Unlike statistical models, AI models, such as ANN and SVM, do not require explicit specification of the relationship between input and output variables. These models are learning algorithms that try to replicate the human brain. Hence, instead of exploring causal inference, these models learn by minimizing a cost function. As a result, the models are more expensive to compute as they require a longer learning history. Furthermore, these "black-box" models are difficult to comprehend, and tuning the architecture and hyperparameters of the model is not an easy task.

Nevertheless, since the 1990s, ANN models have been widely used in load forecasting. The literature survey for peak load forecasting also reveals that this method is most frequently used. Various algorithms have been proposed to enhance the learning method of ANN architecture. The performance of ANN models with different variations of the conjugate gradient backpropagation (CGBP) learning method was discussed [57]. In the conjugate gradient algorithms, a search was performed along conjugate directions, producing faster convergence than the steepest descent directions.

They found that the Powell-Beale CGBP (PBCGBP) method performed the best, while the Fletcher-Reeves CGBP (FRCGBP) performed the worst. An improvement of the ANN architecture by tuning the weights obtained from Levenberg-Marquardt (LM) learning algorithm was proposed using an evolutionary algorithm [35]. A uniform search obtained a new generation of weights. The weights were updated if the child had less validation error than its parent. Instead of using a fixed learning rate, [58] discussed the adaptive learning rate backpropagation method, where the learning rate varies per the performance improvement in each iteration.

Many different search algorithms have been tried to address the key challenge of finding the optimum parameters of networks. Clonal Selection Algorithm (CSA) was introduced to find optimal weights and bias coefficients of ANN architecture [47]. A stochastic search technique for the ANN model was proposed that combined Differential Evolution, Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Simulated Annealing algorithms [35]. In addition, the overfitting problem has been addressed using the Bayesian regularization of parameters. The Bayesian framework constrains the convergence of parameters to a set of weights and biases with smaller values by considering a probability distribution over the weight space.

A recurrent ANN (RNN) to estimate four weeks ahead daily peak load [59] . In the preprocessing step, load values of special days and weeks were replaced with artificially processed normal data. Since the proposed algorithm recurrently used the daily peak forecast as input data for the next day, it was prone to accumulation errors. However, the author reported that the weekly pattern learned by the RNN helped to prevent the increase of accumulated errors. The performances of the proposed RNN were verified on load data of South Korea.

A Neural Network Weighted Least Squares (NNWLS) model was developed with a weighted least squares procedure for training [60]. The objective function of the network was to minimize the weighted least square errors. The weights reflected the

cost of error, which was actually the marginal cost of the hour. The hourly load forecast was evaluated for the on-peak period.

SVM is another common AI technique. The winning entry solution to the EUNITE 2001 competition, based on SVM, was described [34]. The results of this work helped elevate SVM to prominence in load forecasting. Like ANN, one of the key challenges of SVM is tuning the hyperparameters such as cost of error, the width of the tube, etc. The authors in this paper found the parameters based on the evaluation of the validation period.

Several research studies have used soft computing algorithms to search for the optimal parameters of the SVM. A chaotic genetic algorithm-simulated annealing algorithm (CGA-SA) was proposed to determine the hyperparameters for SVR-based monthly peak load forecasting model [61]. The CGA part evaluated hyperparameters' initialized population to find the best individual. The next step was to use the chaotic mutation procedure to find the better parameters in the neighborhood. [50] used the Grey Wolf Optimization algorithm to optimize the regularization parameter and radial basis function parameter of the SVM.

A two-step hybrid optimization algorithm to determine the best parameters of SVR was proposed [26] . In the first step, a grid traverse algorithm (GTA) was used to narrow the parameter search area. In the second step, PSO was used to determine the best parameters in the local parameter space using results from the first step. In another approach, peak load series and meteorological variable series were decomposed using the MEMD algorithm [62]. The extracted inputs were fed to an SVR model, and the hyperparameters were optimized using the PSO.

As computer speed and processing power increase, Deep Learning (DL) techniques are becoming more common. The use of DL in peak load forecasting has seen meteoric growth in the number of published works in recent years. Deep belief networks were used for the day-ahead hourly load forecasting problem [63]. The methodology was

tested on two distribution networks from Macedonia and evaluated for daily peak load forecast.

Long short-term memory (LSTM) recurrent neural was used to decompose the load series into five views to identify trends and variations of load series [64]. The LSTM neural network is a gated recurrent neural network that memorizes the new information selectively. The decomposed features and calendar variables were then used to train the radial basis function neural networks (RBFNN).

[65] also proposed a forecast model based on LSTM. A dual-attention-based encoder-decoder was used to enhance the LSTM. The feature attention mechanism selected the relevant features, while the temporal attention mechanism integrated the features from different time nodes. [83] proposed a hybrid convolutional neural network (CNN) to generate the week-ahead daily peak load forecast. The hyperparameters such as kernel size, size of the pooling layer, number of kernels, and dropout ratio of the CNN were optimized by GA.

A three-layer gated recurrent neural network was used [66]. In another study, a CNN model with an attention network was used for residential short-term load forecasting at the meter level. The variables used in the model were calendar variables, voltage levels, current intensity, reactive power, and energy sensor measurements for different household appliances. The paper empirically demonstrated that the proposed model was able to capture the peak-time load better when compared to the other models.

Another critical issue with AI techniques is over-fitting, leading to poor out-of-sample results. Many studies have addressed this problem by employing clustering to create smaller subsets of similar time series from training data. A widely used clustering algorithm for peak load forecasting is the Self-Organizing Map (SOM) network. It is an unsupervised learning approach in which the clustering centers (called neurons in this context) are initially given random weights. The weight of the

'winning neuron' is then adjusted in each iteration after an input sample is examined for proximity to the neurons. Other neurons' weights are changed based on their proximity to the winning neuron. Each clustered subset is then trained using an AI approach, yielding a family of trained models. The model for test data is chosen based on some criteria, such as the distance of the test sample from the cluster centers.

The SOM network was applied in the first stage to cluster the input training dataset into the regular and anomalous segments by analyzing the dynamic features of the time series [30]. SVR models were fitted on the two subsets to forecast day-ahead daily peak load. The algorithm was tested on the ISO New York dataset. Another study used the SOM network to cluster the training data into two subsets for one month ahead of daily peak load forecasting on EUNITE dataset [43]. Each clustered training subset was fitted to an SVR model. Since the input vector for the SOM comprised of the past seven day's peak load, the forecasted peak load values were used in distance evaluation. This approach has the potential problem of error accumulation that has not been addressed in the paper. In a similar approach, SOM network was used to cluster the training data into five subsets [67]. Then, a feed-forward neural network (FFNN) model was developed for each data cluster to forecast the daily peak load.

Other algorithms have also been used to find similar days in history. A k-nearest neighbor (kNN) with a 1-norm distance was used on the temperature variable to find similar days in a 2-week neighborhood around the target day [23]. A typical load curve for a commercial building was generated using the hierarchical agglomerative clustering technique [68]. The load curve was separated into four phases, and the mean power demand for each period was utilized as the clustering algorithm's input. The peak load of the day was forecasted using the estimated load curve, and the actual demand at 6 AM.

Similar days of the week were grouped by using correlation factor resulting in three clusters for each season [69]. Another method used Dynamic Time Warping

(DTW) to establish a similar daily peak load sequence from the historical data [66]. DTW dynamically allows optimal mapping between two time series to represent the similarity between the two series. Peak load segments (a sequence of daily peak loads of the last seven days) were obtained from autocovariance analysis of the daily peak load series, which was then matched to the segments in the historical data using DTW. The rule was that if the two segments were similar, the following corresponding segments would be similar too. The matched segment was then fed into the Gated Recurrent Unit (GRU) along with calendar and temperature variables. The EUNITE dataset was used to evaluate performance.

2.1.2.3 Hybrid Models

Integrating statistical analysis and machine learning in a model is a common approach. These hybrid models tend to perform better since different models are able to capture diverse aspects of load patterns. Typically the load series is broken into different components, where each component is modeled separately. Wavelet decomposition of clustered daily load data was done to break it down into high-frequency and low-frequency series [69]. The day-ahead low and high-frequency peak loads were forecasted using the two ANN models. Finally, the forecasted peak load signal was obtained by reconstructing the predicted low and high-frequency components.

A hybrid model using fuzzy logic (FL) and ANN for the day-ahead peak load forecast was proposed [70]. The FL model yielded a value representing the expected change in the peak load from the previous day's peak load. The output of the FL module was used as an input to the ANN module, along with calendar and temperature variables. Fuzzy Neural network (FNN) exploits the learning capability of ANN to solve the problem represented by fuzzy logic [41]. The FNN-based model was proposed to forecast the day ahead peak load. FNN can be viewed as a three-layer feed-forward network, with an input layer where inputs are classified into fuzzy memberships (fuzzification), a hidden layer corresponding to the individual fuzzy rules

modified adaptively, and a final fuzzy output layer of load forecasts (defuzzification). In another application, the FNN model was proposed to forecast noon peaks and evening peaks [71]. The fuzzy model used here was rule-based, consisting of if-then rules. The output of each rule was a function of independent variables. The Recursive Least Squares Estimation method was employed to calculate the rule of consequent parameters. The model was trained using a GA.

Instead of integrating different approaches into one model, a forecast combination brings together the forecasts from individual models from different techniques. The core idea is to leverage the information underlying individual forecasts. Combining forecasting approaches empirically provides more robust and accurate forecasts than individuals [72]. A weighted averaging approach was proposed, where the two forecast series were combined to generate the final forecasts [45]. One approach was based on decomposing time series into sub-series, where each subseries was modeled using a Deep Belief Network (DBN). The second approach used aggregation of multiple outputs from DBN network developed on peak load indicative variable for each exogenous variable. built An ensemble model by combining eight base models using different techniques(MLR, SVR, RF, MLP, BT, MARS, ARIMA, and kNN) was built [38]. GA was used to optimize the weights of eight base models in the final ensemble model. A hybrid forecasting model involving models based on ARIMA, Logistic Regression (LR), and ANN, was developed to predict whether the next day would observe the peak demand of the month [73].

2.2 Peak Timing Forecasting

The literature focusing on peak timing forecasting is scarce. One of the first studies to examine load forecasting models for peak timing accuracy was conducted by [19]. The paper proposed their methodology to determine the probability of the occurrence of peak demand. The case study was from a province in the Netherlands for years between 1995 and 2000. The analysis showed that cold temperatures and the holiday

season led to peak day occurrence in the weeks leading up to Christmas. Using weather simulations, the model estimated the probability of the peak load day on different days in December.

"PELD" (Peak Electric Load Day) was defined as the day when the high demand occurs [73]. The z-score of the distribution of the daily peak load of a month was computed to define if a day is PELD or not. Month-ahead PELD forecasting models were developed from two approaches. One of the tracks was a two-step process with monthly-threshold-based forecasting models. First, a monthly threshold value was obtained from a daily peak forecasting model using an elastic net regression analysis. Next, day-ahead load forecasts at 30-minute intervals were generated and the monthly threshold was used to determine whether or not a day qualified as PELD. The second approach used a binary classification forecasting model to directly classify a day as PELD or non-PELD. Finally, a hybrid model was proposed wherein the forecasts from the two were combined. The study was extended to evaluate the effect of "Behind the meter Renewable Energy Generation" on the performance of the proposed algorithm [74].

A Bayesian approach to generate the forecast distribution of half-hourly load for long-term and short-term horizons [24]. The proposed method generated a predictive distribution of the time of the daily peak. The study also noted that seasonal factors contribute strongly to the distribution of daily peak hours. Another study evaluated the performance of their proposed short-term peak load forecasting model for magnitude, and also the ability of the model to forecast the occurrence of daily peak hours [75]. They compared the distribution of actual peak hours and forecasted peak hours by the Kolmogorov-Smirnov test. The results showed that larger time displacements between the forecasted and actual values of the peak hour were less frequent than smaller displacements, as most of the displacements were within three hours. The case study used in the experiment came from the Colorado State University campus.

The forecast of timing of daily peak load for commercial and industrial customers was analyzed in [76]. Contrary to residential loads, industrial loads do not exhibit any short-term seasonality. Hence, a model using stochastic analysis of the demand was developed. The peak time was forecasted by comparing the logarithmic rate of change between the data and the historical references. The methodology was tested on three case studies with different sampling rates of one hour or 15 minutes. The results were analyzed for accuracy, and the study concluded that the methodology could not distinguish spikes from actual peaks in higher sampling rate data. In another study, abnormal peak load was defined as the hourly load with significant and positive deviation from its neighboring loads in magnitude [54]. They used the historical probability of each hour to observe the peak demand to forecast the occurrence of abnormal peak load for every month.

Evidently, the study on the timing aspect of peak forecasting is very limited in the literature. The primary focus of PF has been on magnitude to date. The peak timing aspect has received little attention. Due to a lack of comprehensive research on the topic, there are some open questions:

1. How do we measure the peak timing forecast accuracy?

2. How do the load forecasting models developed for the entire load profile perform for peak timing forecasting?

3. Are models specifically developed for peak timing forecasting more accurate?

CHAPTER 3: CASE STUDIES

3.1 Case Study 1: MeSU

Our first case study is based on data from an anonymous medium-sized U.S. power utility, that we call MeSU. The load time series consists of 8 years (2012-2019) of hourly load data from three adjoining supply areas within the utility's service territory (3.1) and the aggregate zone. The aggregated level is called the *top level*, which is an aggregate of three supply areas named SA1, SA2, and SA3, and a small-sized supplier. However, we only used the load data from the three largest supplier areas as separate load zones for the experiment. Table 3.1 provides an overview of the statistics of load data for each of the four zones for the eight years. Data shows that SA1 is the largest supply area, while SA3 is the smallest.

Table 3.1: Summary statistics of load data (in MW) for MeSU

	Mean	Std.	Min	Max
Top Level	1534.8	552.6	429.7	4712.8
SA1	1118.1	370.4	507.6	3204
SA2	233.1	82.7	101.2	553.1
SA3	147.6	47	58	375.6

Weather data is available from 28 weather stations that are located within or near the three supply areas. The weather data consists of historical observed hourly temperature data and day-ahead hourly temperature forecast data for the period of the given load history. The temperature forecast is released at 7 a.m. and covers the following 40 hours. The hourly temperature records from the 28 weather stations for

a day from 2016 (06/14/2016) are plotted in Figure 3.2. We see a lot of variations in the readings, and hence selecting weather stations for each zone is an essential task.

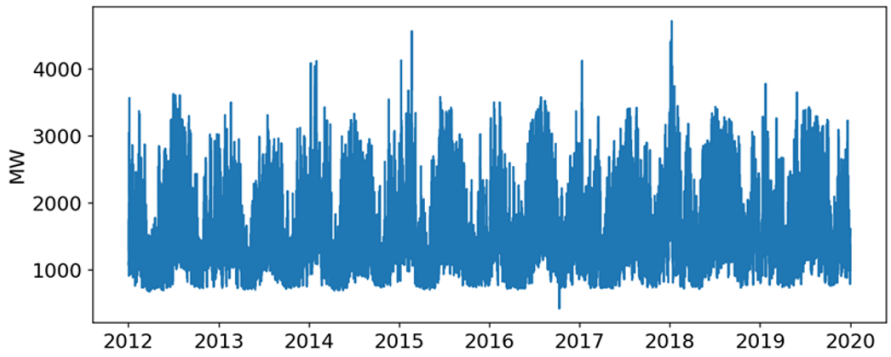


Figure 3.1: Load history of Top Level under MeSU.

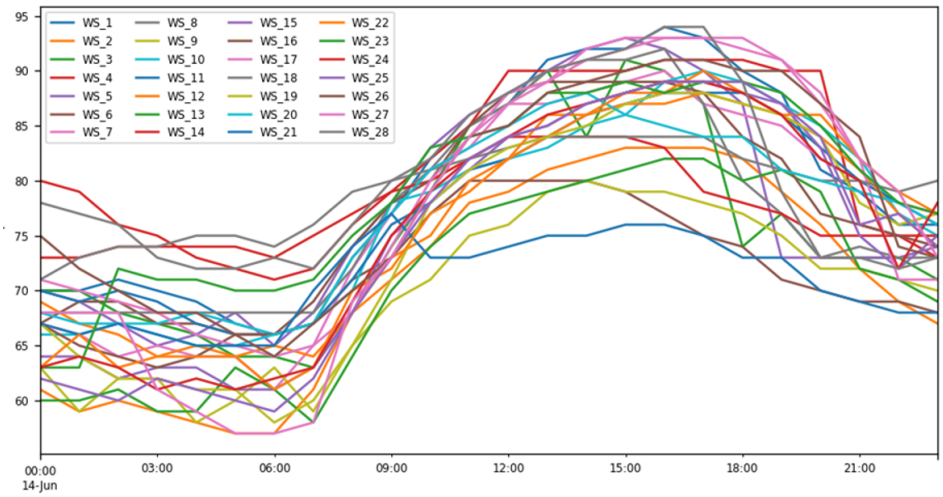


Figure 3.2: Temperature series from 28 weather stations under *top level* of MeSU.

Figure 3.3 shows the histogram of daily peak hours for each month of the year 2016. We see seasonality in the timing of peak demand. Summer months see extended long periods of peak demand in the evening. Months with mild weather such as April and October, the peak hours are distributed between morning and evening peaks. In winter months, days with morning peaks dominate over days with evening peaks.

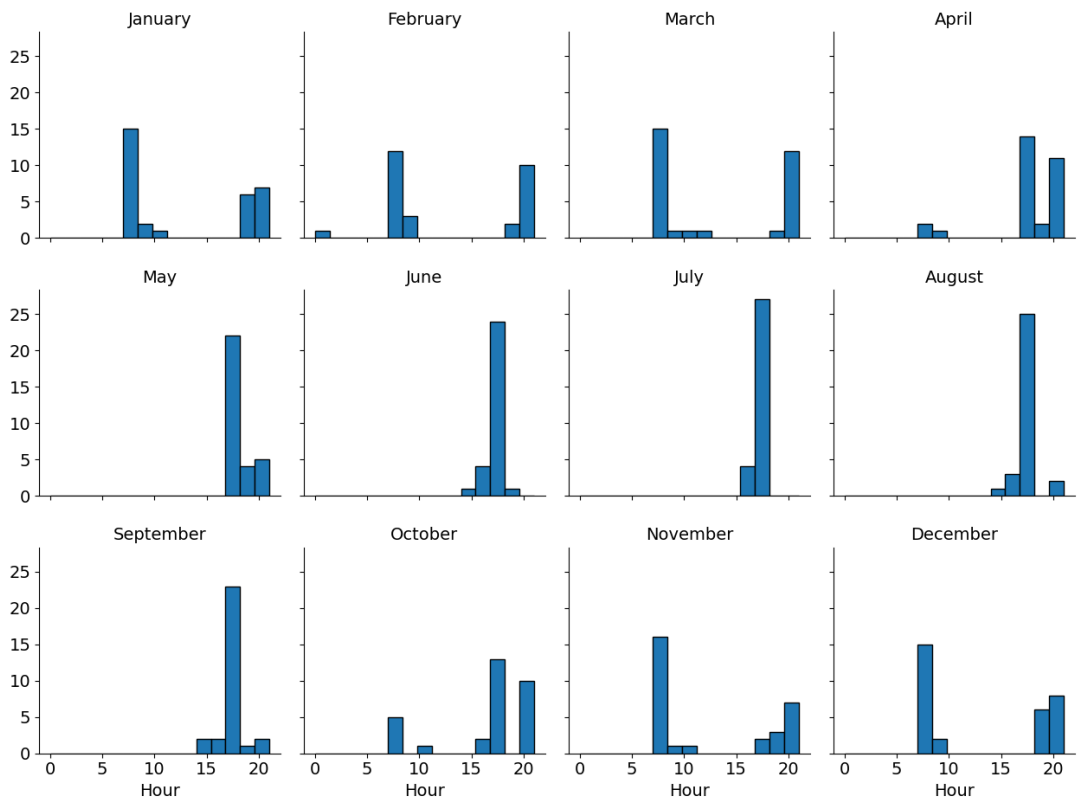


Figure 3.3: Peak hour distribution for *top level* of MeSU in different months of the year 2016.

3.2 Case Study 2: ISONE

The second case study is publicly available data from ISO New England, which serves the six states in the northeast U.S. including Connecticut (CT), Massachusetts (MA), Maine (ME), New Hampshire (NH), Rhode Island (RI), and Vermont (VT). MA is dissected into three load zones, namely Northeastern MA and Boston (NEMASS), Southeastern MA (SEMASS), and Western Central MA (WCMASS). Each of the other five states is considered a load zone. The aggregated zone is named ISONE. Figure 3.4 shows the eight load zones of ISO New England [77]. The summary statistics of the nine zones are shown in Table 3.2, where zones are arranged in the order of the average load statistics.

In this paper, we use six years of hourly load and temperature data published by

ISO New England from 2014 to 2019 to conduct the case study. The observed temperature data from weather stations corresponding to each load zone is also given. However, the historical temperature forecast is not archived. Recently, the European Centre for Medium-Range Weather Forecasts (ECMWF) published the historical numerical weather prediction (NWP) [78]. This dataset contains four years of historical weather forecast history (2017-2020), and 14 weather forecast variables at a spatial resolution of $0.5^{\circ} \times 0.5^{\circ}$grid, covering most of North America. We leverage this data to get the real temperature forecast for all the zones by mapping the coordinates of the weather station corresponding to each load zone, to the closest grid point. Hence, we get the hourly temperature forecasts for three years (2017-2019) for this case study.

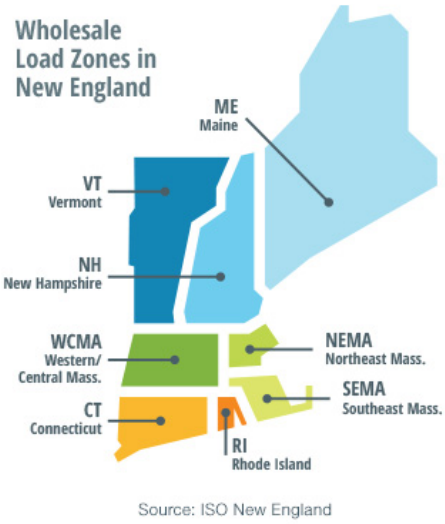


Figure 3.4: Load zones under ISONE.

Table 3.2: Summary statistics of load data (in MW) for ISONE.

	Mean	Std.	Min	Max
ISONE	14211.7	2828.9	7892.0	27707.0
CT	3409.5	762.9	1372.0	7219.0
NEMASSBOST	2830.1	563.8	1665.8	5658.0
WCMASS	1922.1	372.8	739.0	3650.0
SEMASS	1659.1	385.2	875.0	3645.0
NH	1316.7	262.1	521.0	2433.0
ME	1303.3	205.9	711.6	2135.0
RI	916.3	204.3	365.0	1967.0
VT	627.5	110.7	178.1	985.0

The peak hour distribution across months for ISONE is depicted in Figure 3.5. Unlike the first case study, the peak hours in summer are scattered between morning and evening, while in winter, the system dominantly peaks in the evening only.

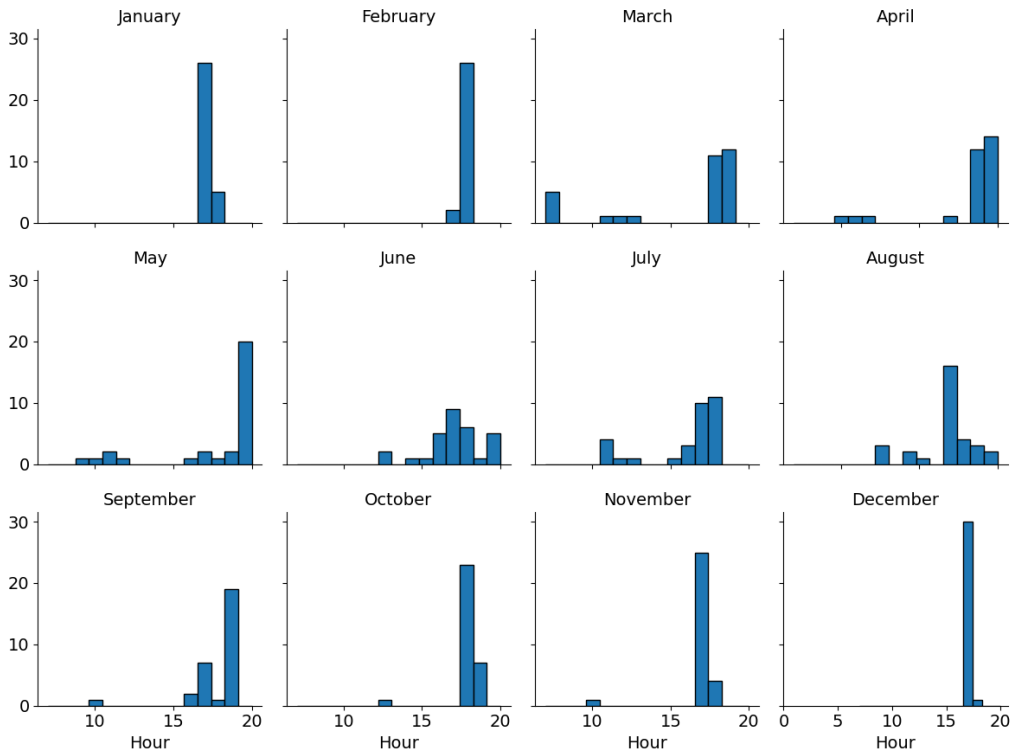


Figure 3.5: Peak hour distribution for ISONE in different months of the year 2016.

CHAPTER 4: THEORETICAL BACKGROUND

4.1 Multiple Linear Regression

MLR is a traditional statistical approach that models the relationship between the independent variables and the response variable. The model can be presented as:

$$Y_t = \beta_0 + \beta_1 \ X_{1,t} + \beta_2 \ X_{2,t} + \beta_3 \ X_{3,t} + \ldots \beta_k \ X_{k,t} + \varepsilon_t \tag{4.1}$$

The most commonly used estimator is the Ordinary Least Squares (OLS), which works on the basic principle of minimizing the Sum of Square Error (SSE). For a sample size of N, SSE term can be presented as:

$$SSE = \sum_{t=1}^{N} \left(Y_t - \ \beta_0 - \beta_1 X_{1,t} - \beta_2 X_{2,t} - \beta_3 X_{3,t} \ldots \beta_k X_{k,t}\right)^2 \tag{4.2}$$

The model coefficients can be estimated by solving the closed-form matrices or using gradient descent techniques, the latter is beneficial when dealing with huge datasets.

4.2 Recency Effect

An MLR-based model, popularly named Tao's Vanilla Benchmark model, was proposed [48]. This model is represented by Equation 4.3.

$$\widehat{L_t} = \beta_0 + \beta_1 \ Trend_t + \beta_2 \ D_t + \ \beta_3 \ H_t + \beta_4 \ M_t + \beta_5 H_t D_t + \ f\left(T_t\right) \tag{4.3}$$

The observed load at time t is represented as as L_t, the load forecast is presented as $\widehat{L_t}$. The $Trend$ is an increasing natural number to represent the linear trend of the load history. Calendar variables are modeled as class variables. The hour of day is represented by H_t with 24 levels. Week of the day is represented by D_t with

seven levels, and the calendar month is represented by M_t with 12 levels. The non-linear relationship between the temperature and load is captured by using up to third order temperature polynomials in the model. The seasonal change in temperature is modeled by the interaction of temperature terms with the months of a year. To model the human activities related cycles, temperature terms interact with the hours of a day, and the hours of a day interact with the days of a week.

$$\begin{aligned} f\left(T_t\right) = \ & \beta_6 T_t + \beta_7 {T_t}^2 \ + \beta_8 {T_t}^3 + \beta_9 H_t T_t + \beta_{10} H_t {T_t}^2 \ + \beta_{11} H_t {T_t}^3 + \\ & \beta_{12} M_t T_t + \beta_{13} M_t {T_t}^2 \ + \beta_{14} M_t {T_t}^3 \end{aligned} \tag{4.4}$$

The impact of lag temperatures, or the temperatures from the previous hours, on the current hour's demand is referred to as the Recency effect [48]. Further, [79] investigated the number of lagged hourly temperatures and lagged daily moving average temperatures that yield the best forecast accuracy. The approach deploys lagged hourly temperatures (T_{t-h}, h = 1, 2, . . ., 72) and daily moving average temperatures of previous d days. The daily moving average temperature of each d^{th} day can be represented as:

$$\widetilde{T_{t,d}} = \frac{1}{24} \sum_{h=24d-23}^{24d} T_{t-h} \tag{4.5}$$

, where d= 1, 2, . . .,7. Then, Tao's Vanilla benchmark model can be extended to accommodate the Recency effect as:

$$\begin{aligned} \widehat{L}_t = & \beta_0 + \beta_1 \ Trend_t + \beta_2 \ D_t + \ \beta_3 \ H_t + \beta_4 \ M_t + \beta_5 H_t D_t \\ & + \ f\left(T_t\right) \ + \ \sum_d f(\widetilde{T_{t,d}}) + \ \sum_h f(T_{t-h}) \end{aligned} \tag{4.6}$$

This model is called the Recency Effect model. The modeling approach requires finding the best *d*-*h* pair on validation data. For the GEFCom2012 dataset, this model has demonstrated superiority above Tao's Vanilla benchmark model(Hong, 2010) by an average of 12 to 15 percent [79].

4.3 Logistic Regression

Logistic Regression is a specific regression analysis approach applied to predict an outcome in probability, and hence, is often used for classification problems. Logistic regression, like linear regression, is a parametric model in which the coefficient values are estimated in order to predict the output [80]. In contrast to linear regression, logistic regression transforms the output using a logistic function to predict the probability of the default class. The logistic regression model is expressed as follows:

$$\log\left(\frac{p_t}{1-p_t}\right) = \theta_0 + \theta_1\ X_{t,1} + \theta_2\ X_{t,2} + \ldots\ \theta_k\ X_{t,k} + \varepsilon_t \tag{4.7}$$

, where p_t denotes the conditional probability of success, given the explanatory variable $[X_{t,1}, X_{t,2} \ldots X_{t,k}]$.

Maximum Likelihood Estimator (MLE) is used to estimate the parameters of Logistic regression. MLE gives the values of parameters that, out of all the possible values for θ, maximize the likelihood of the observed data. If y_t represent the true label, the cost function is given as follows:

$$\sum y_t log(p_t)\ +\ (1-y_t)log(1-p_t) \tag{4.8}$$

The peak timing forecasting problem can be formulated as a binary or a multi-level classification problem. With binary classification, the instance of the occurrence of peak demand can be labeled as an event, indicated by 1. The peak hour forecasting problem can also be formulated as a multi-level classification problem. The different probable peak hours can be the nominal classes of the model. Once the event state is established, the actual or forecast value of peak demand is no longer relevant; only the event status is used to calculate metrics.

4.4 Forecast Evaluation

4.4.1 Error Measures for Load Forecasting Problems

Different types of error measures are used to evaluate hourly load forecasts. Due to its interpretability and scale independence, mean absolute percentage error (MAPE) is one of the most frequently used error metrics in the domain of load forecasting. However, a significant disadvantage of employing MAPE is that it gives a heavier penalty when the actual value is smaller, and hence it will become invalid for actual values of zero or close to zero. MAE and RMSE are scale-dependent, and hence, they cannot be used across different data sets. In addition, RMSE is sensitive to outliers as it gives a higher penalty for large errors.

$$Mean\ Absolute\ Percentage\ Error\ (MAPE) = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{L_t - \hat{L}_t}{L_t}\right| \tag{4.9}$$

$$Mean\ Absolute\ Error\ (MAE)\ = \frac{1}{N}\sum_{t=1}^{N}\left|L_t - \ \hat{L}_t\right| \tag{4.10}$$

$$Root\ Mean\ Square\ Error\ (RMSE) = \frac{1}{N}\sqrt{\sum_{t=1}^{N}\left(L_t - \ \hat{L}_t\right)^2} \tag{4.11}$$

In our case studies, we hardly encounter loads that are close to zero. Hence, in this study, we use MAPE as the error measure to evaluate hourly load forecasts.

4.4.2 Error Measures for Classification Problems

Classification problem with binary outcomes is best described using a confusion matrix as in Fig. 9. True Positive (TP) refers to the count of true events detected, and False Positive (FP) refers to the false events detected. True Negative (TN) is the count of true non-events detected, while False Negative (FN) is the count of false non-events detected.

	Predicted True	Predicted False
Actual True	TP	FN
Actual False	FP	TN

$$Sensitivity\ or\ Recall\ or\ \ True\ Positive\ Rate(TPR) \ = \frac{TP}{TP+FN} \tag{4.12}$$

$$True\ Negative\ Rate(TNR) \ = \frac{TN}{TN+FP} \tag{4.13}$$

$$Precision\ or\ Positive\ Predictive\ Value(PPV) \ = \frac{TP}{TP+FP} \tag{4.14}$$

$$Balanced\ accuracy(BA) \ = \frac{TPR+TNR}{2} \tag{4.15}$$

$$F_1\ score \ = \frac{2 \ \times TP}{(2 \times TP)+FP+FN} \tag{4.16}$$

$$CSI \ = \frac{TP}{TP+FP+FN} \tag{4.17}$$

True Positive Rate (TPR), defined in Equation 4.13, gives the ratio of the number of the events identified to the number of actual events. Precision, defined in Equation 4.14, gives a ratio of the number of events identified to the total number of events identified. This metric accounts for FP and a higher FP would lower the Precision. However, the metric ignores the missed events, that is, the FN. Another metric given by Equation 4.16, known as the F1 score, combines TPR and Precision. Equation 4.15 gives the Balanced Accuracy, which is the average of TPR and TNR. Another common metric is the Critical Success Index (CSI), given by Equation 4.17, also known as the threat score. The formulation of CSI is very similar to F1 score, with TN removed from both the numerator and denominator. This could be useful for the evaluation when the number of true negatives is dominantly large.

CHAPTER 5: ERROR MEASURES FOR PEAK TIMING FORECAST

5.1 Literature Review

The literature on peak timing forecasting mainly gives two problem sets - forecasting daily peak load instances and forecasting the days with peak demands. In this section, we will look at the error measures used by the prior studies for measuring the forecast errors for the two problems.

Peak timing forecasts have been evaluated by comparing the distribution of peak hour forecast with the actual distribution [54], [75]. This method provides a visual indication of the goodness of the forecast models. However, it does not quantify the errors. [76] formulated peak timing forecasting problem as a binary classification problem, and employed classification problem error measures such as TPR, FPR, Sensitivity, and F1-score. Balance Accuracy (BA) was used to evaluate PELD forecasts [73] and [74]. The probability of peak demand occurrence on different days in a month was generated using forecasts generated by different long-term temperature scenarios [19]. The paper simply compared if the actual day of peak demand falls on a date with a relatively high probability of peak demand.

A new error metric was developed to evaluate sub-hourly forecasts at the meter level [81]. A set of forecasts was created by shifting the forecast series by a few intervals. Then each set was evaluated with actual load series using the absolute p-norm with $p = 4$. The minimum error over the set of all permutations was taken as the final error to allow for some flexibility in the timing of the forecast. Using the 4-norm errors, the authors tried to associate large errors with missed peaks, which might not be true as a large error could also correspond to non-peak hours.

Some studies have innovated with application-specific error measures. Domain Bias

Percentage Error (DBPE) based on Granger's 'linlin' function was proposed in [82]. It allows different penalties for over-prediction and under-prediction. Further, they configured the constraint to favor either under-prediction or over-prediction. In most applications related to a peak demand of smart grid, under-prediction of forecasts is more deleterious. Although the metric is intended to address peak demand error, the proposed error measure does not specifically measure the peak timing error.

An error metric was introduced to evaluate load forecasts used for optimizing the charging and discharging schedule [83]. The metric only evaluated evening hours and was named time series shape error. The 6 PM load was used as the reference hour load, and the difference between the forecast at time t and at 6 p.m. and the difference between the actual load at time t and at 6 p.m. were then evaluated with RMSE formulation. The formulation is given in equation 5.1.

$$\begin{aligned} \varepsilon = & \left(\sum_{t} Err \right)^{\frac{1}{2}}, where \\ Err & = [(F_t - F_{6PM}) - (L_t - L_{6PM})]^2 \end{aligned} \tag{5.1}$$

, where $t \in$ ESS discharge hours (i.e., 5 p.m. to 8 p.m.), F_t is the load forecast of hour t, F_{6PM} is the load forecast at 6PM, L_t is the actual load of hour t, L_{6PM} is the actual load at 6 p.m. This weighted sum of the time series shape error and the mean square error was used as the loss function for forecasting load. The proposed method was implemented in the winning entry of the Western Power Distribution (WPD) Presumed Open Data (POD) competition, where the task was to generate the charging and discharging profile of ESS. Although the metric is implementable, it is biased towards the error of the reference hour load, as shown in Table 5.1. Case 1 and Case 2 have the same absolute errors in adjoining hours of the 6 p.m. and ε is 9 for both cases. In Case 3, the reference hour load has the same error, however, this time ε is quite high at 37.

Table 5.1: Demonstration: time series shape error.

	Case 1			Case 2			Case 3		
	A	F	Err	A	F	Err	A	F	Err
16:00	8	11	9	8	8	0	8	8	9
17:00	12	12	0	12	12	0	12	12	9
18:00	14	14	0	14	14	0	14	17	0
19:00	13	13	0	13	10	9	13	15	1
20:00.	12	12	0	12	12	0	12	12	9
21:00	11	11	0	11	11	0	11	11	9
ε			**9**			**9**			**37**

Another recent study proposed Relaxed accuracy for measuring the daily peak timing for half-hourly data [33]. The measure is similar to TPR but with some flexibility. It gets the value of 1 if the peak load instance is within two instants away of the actual peak, and 0 otherwise.

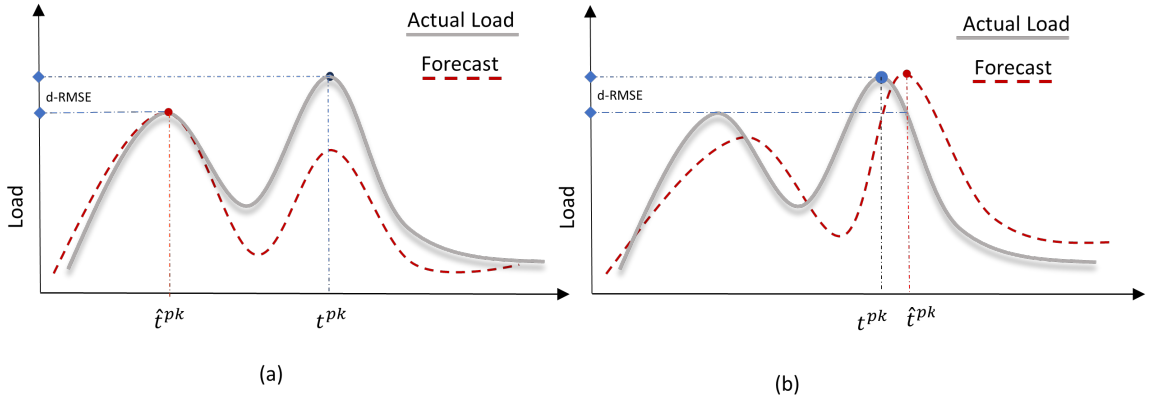


Figure 5.1: Demonstration: scenarios with different peak timing errors but same d-RMSE.

The paper also proposed a new error measure d-RMSE which is based on the difference of actual load at the predicted peak time $L_{\hat{pt}}$ and actual load at actual peak time L_{pt}, in other words, actual peak load.

$$d\text{-}RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(L_{PH} - L_{\widehat{PH}}\right)^2} \tag{5.2}$$

The authors argued that if the two are very close, then the peak timing error has no effect operationally. This may not hold true as demonstrated in Figure 5.1. In case (a), the DR event would be triggered in the incorrect time window, resulting in a missed peak-shaving opportunity. In case (b), where the timing error for the peak is small, the time of day during which the DR event is triggered is close to the true peak hour. However, the *d-RMSE* is the same in both cases. Therefore this error measure is not able to differentiate the two cases and does not help in measuring the displacement error of peak timing.

Table 5.2: Summary of error measures used in literature for peak timing forecasting.

	Error Measure	Reference
Peak Load Days	TPR, BA	[73],[74]
Peak Hours	Comparing distributions, Displacement error, TPR, R-accuracy, *d-RMSE*	[54],[75],[76],[33]
on-Peak period	Timeseries shape error	[84]

Table 5.2 summarizes the error measures used in prior studies pertaining to peak load timing. As we have seen, there has been little research on peak timing, and there is no standard method or error metric to quantify how well a model performs in forecasting peak timing. Therefore, we looked into error measures used in some other fields for the event detection problem. In the discipline of meteorology, several studies have been done on the metrics for evaluating event detection models. [85] examined verification approaches that have been investigated and established in the meteorological realm. One of the sections emphasized the issues related to the verification of extreme events. In terms of binary choices, "extreme" is not as black-and-white as it may seem. The extreme events can be defined by selecting values exceeding a pre-defined threshold, also known as peaks over the threshold, or selecting maximum

value over a long period, such as annual maxima. For extreme events, it is desirable to generate probabilistic forecasts [86]. Brier Score is a popular score adopted to assess probabilistic forecasts of binary outcomes [87]. It measures the Euclidean distance between the actual outcome (o_t) and the forecasted probability (f_t) assigned to the outcome for each observation as shown in equation 5.3.

$$\frac{\sum (f_t - o_t)^2}{N} \tag{5.3}$$

[88] discussed common metrics used to verify event detection models such as Accuracy, CSI, F1 score, and also, skill scores such as Heidke skill score (HSS), Peirce skill score (PSS), and Gilbert skill score (GSS) in the research of magnetospheric physics. Skill scores compare performance against a reference model.

$$Skill\ Score = \frac{metric_{model} - metric_{ref}}{metric_{perfect} - metric_{ref}} \tag{5.4}$$

A skill score of 0 indicates that the new model has the same skills as the reference model, and a skill score of -1 indicates that the new model has inferior skills to the reference model. The reference model could be a persistent model or even a random chance.

Some studies have developed error measures for a specific application. An optical flow-based verification method was first published in [89]. The method breaks down the forecast error into components: displacement, amplitude, and residual errors. A similar idea was coined by [90] to quantify spatial differences between the forecast and observation fields using an image-matching algorithm. The authors developed Displacement Error (DIS) and amplitude error (AMP) to quantify how accurately features are predicted in position and amplitude. These two errors are normalized and summed to get a displacement-amplitude score (DAS). Such forecast verification methods can be relevant for the verification of peak timing problems.

In the lack of adequate research on the peak timing forecasting problem, there are

some fundamental questions that remain unanswered.

1. What are the different application-specific forms of peak timing problems?

2. Are the traditional load forecasting error measures and the classification problem statistics suitable for the operational requirement of peak timing forecasting?

3. Finally, can we leverage error measures used in other domains to verify peak timing forecasting?

In the next sections, we will explore these questions by first defining different forms of peak timing problems, critically examining traditional metrics, and finally proposing error measures for peak timing problems.

5.2 Taxonomy of Peak Timing Problems

5.2.1 Peak Load Day

The first problem of Peak timing forecasting is identifying the days when the peak demand of a month or a season is likely to occur, which we will call Peak Load Day (PLD). Customers under a demand charge structure are charged for the peak load consumed during their billing period [91]. Monthly PLDs are the days when the demand is considerably higher than the average demand for the month. Customers can effectively plan and implement demand response activities to reduce peaks and avoid high demand charges if they have information on the PLDs in advance. Seasonal PLDs are the days when the system load is considerably higher than the average demand in a year. The operators need to know the seasonal PLDs in advance to plan for DSMs such as calling DR events and maintenance activities. Because the number of DR events that may be called in a season is limited, calling PLD on the wrong day is disadvantageous. For the purpose of this study, we only analyze seasonal PLDs, and henceforth seasonal PLDs will be referred to as PLDs. The number of PLDs in a year is determined by the threshold or limiting load that is used to classify a day as PLD. One way to determine the PLDs is from the load duration curve. A load duration curve is simply sorted load data ordered from highest to lowest values. The load duration curve helps visualize the number of hours during which a given load has prevailed (Figure 5.2). The vertical dash lines indicate the load levels from 95% of the annual peak to 5% of the annual peak. For example, demand up to 90% of the annual peak, which is represented by the second vertical line, occurs only for 2% of 8760 hours in the year. Demand less than 30% of annual peak demand is observed at all hours and hence, it is the base load of this system.

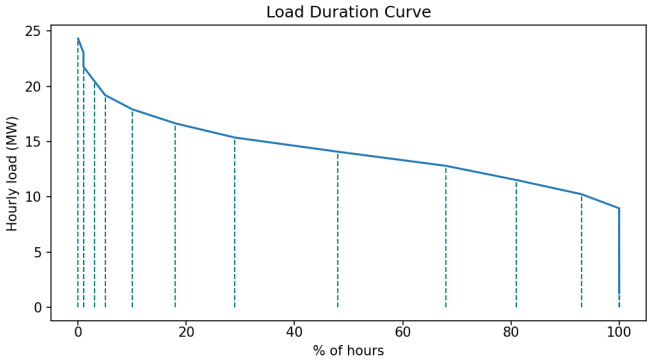


Figure 5.2: Load duration curve.

We can define a threshold for PLD by choosing the percentage of load hours from the Load duration curve. However, this definition would require knowledge of the generation capacity of different power plants and other operation details. Therefore in this research, we have used a threshold based on the statistics of hourly load distribution over a year. If μ is the mean of hourly loads in a year, and σ is the standard deviation of hourly loads in the year, then the z-score of daily peak load (P_t) is Z_t as in equation 5.5. The z-score of daily peak load can be used against a threshold to classify a day PLD or not. The threshold can be chosen from the operational view or electricity market view of terming 'abnormal demand'. For instance, a threshold of 2 was used to define electricity price spike [92], and to classify PELD [73].

The probability of obtaining a z-score of 2 or larger for a normally distributed test statistic is $p \leq 0.023(1 - 0.9773)$. Z_t greater than 2, implies that the P_t is more than two standard deviations above the mean (μ). Using a z-score of 2 as the threshold would yield between 8 and 201 PLDs in a year. If we define a binary variable PD to indicate if a day is a PLD or not, then PD is defined in the 5.6.

$$Z_t = \frac{P_t - \mu}{\sigma} \tag{5.5}$$

$$\text{PD} = \begin{cases} 1, & \text{if } Z_t \geq 2 \\ 0, & \text{otherwise} \end{cases} \tag{5.6}$$

Figure 5.3 visualizes the count of PLDs in the two case studies. The number of PLDs ranges between 30 to 90 days per year.

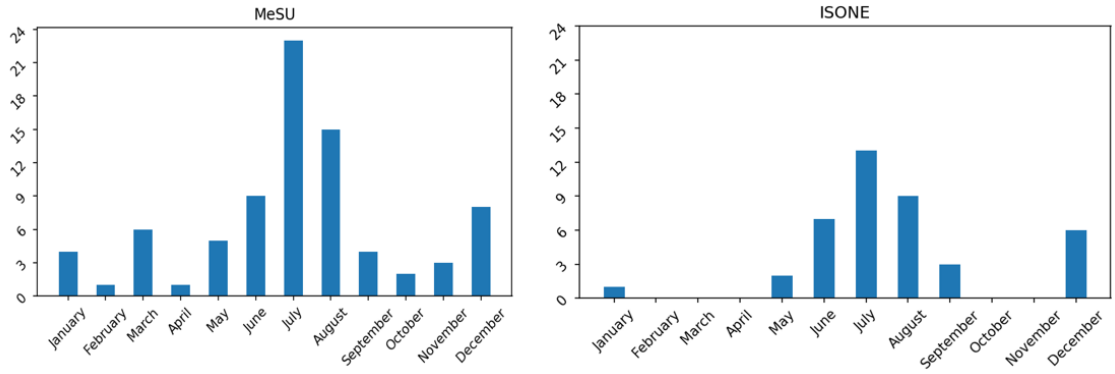


Figure 5.3: Number of PLDs in the *top level* of the two case studies.

5.2.2 Peak Hour

Peak hour is the instant the highest demand is observed during a period. The definition also depends on the frequency of the data (hourly or sub-hourly). In this research, we are working on the hourly data and hence peak hour is the hour when the peak demand of a day occurs. Daily peak hours of PLDs for ISONE have been plotted for the nine zones of ISONE in Figure 5.4.

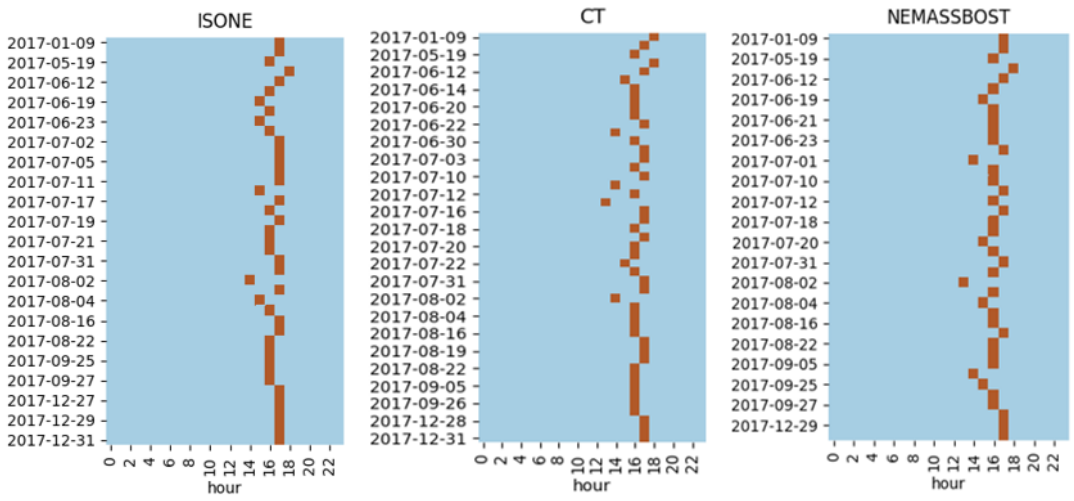
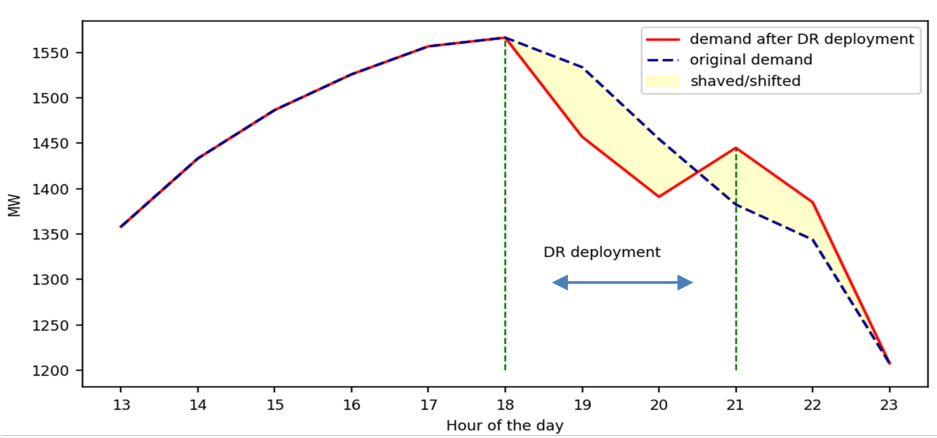


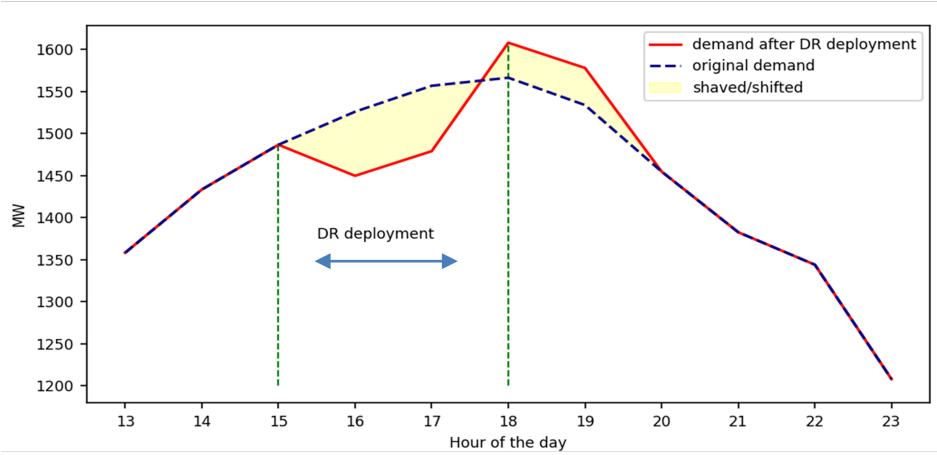
Figure 5.4: Peak hours on PLDs in different zones of ISONE .

5.2.3 On-peak Period

Peak shaving strategies are deployed in a time window on PLDs. For example, DR events are called for a few hours on the anticipated PLD. Missing the actual peak hours by calling DR during the wrong window would jeopardize the peak shaving opportunity as shown in Figure 5.5 (a), where the DR is deployed after the actual peak hour. On the other hand, calling DR much earlier than the actual peak hour would shift the demand to the true peak hours, raising the true peak demand even more (Figure 5.5 (b)).



(a)



(b)

Figure 5.5: DR-deployment demonstration: a) event called earlier than the actual peak period, b) event called later than the actual peak period.

Similarly, for the use case of scheduling charging and discharging ESS, it is very important to know the best discharging hours for the optimum peak shaving. Hence, it is imperative to know the "on-peak period", which can be defined in different ways.

Most of the utilities fix the on-peak period during the evening time to schedule the discharge of the PV-powered ESS [93]. Utilities also choose on-peak periods from the load profile to design tariff structures. Because the peak occurs at different times

during the summer and winter, this is done separately for each season. Overall, in practice, a broad window of hours where the peak is most likely to occur is selected as the on-peak period. As long as peak demand is within the defined window, the cost of misclassification is trivial.

In this research, we define the on-peak period (Q) as the time window extending 2 hours before and 2 hours after the peak hour, where hourly load exceeds 90% of the daily peak load (P_t), as shown in Figure 5.6. Hence, the on-peak period can span between one hour and five hours. This definition of "on-peak period" focuses on a specific time of day that varies every day. It may be complicated for seasonal tariff designing, but it is crucial for the implementation of dispatchable peak-shaving techniques and real-time pricing strategies.

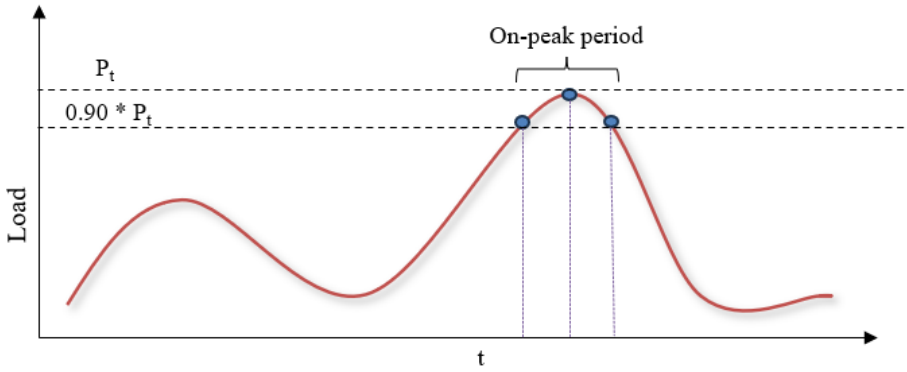


Figure 5.6: On-peak period demonstration.

5.2.4 Peak Shape

Regardless of magnitude, the shape of the load profile during the on-peak period is especially relevant for the efficient deployment of DR programs and scheduling ESS discharging profile. We demonstrate this by an example of scheduling the discharge of an ESS that has a limited capacity. The ESS discharge profile defines how much energy is to be discharged in each time interval of the discharging period. The optimum peak shaving is achieved when the discharge profile follows the load profile

exactly. The challenge is to create a discharge schedule without knowing the future demand.

In Figure 5.7, we show two situations (a) and (b). The evening peak of actual demand is to be reduced between 16:00 and 21:00 hours by discharging an ESS. The green line is the ideal shaved load profile that can be achieved if the load forecast perfectly matches the actual load. In scenario (a), the peak hour forecast is consistent with the actual peak hour, but the load shape forecast during the evening hours differs significantly from the actual shape. The realized shaving load, shown by the red line, is obtained by optimizing the ESS discharge schedule using load forecast. We can see that due to forecast error, the full potential of peak shaving is not realized. In scenario (b), the demand is underestimated, but the shape forecast fits the actual load shape. Even though the forecast error is large in this case, we are still able to achieve optimal peak shaving. As a result, while selecting a load forecasting model for this problem, it is essential to select a model that closely tracks the load shape rather than a model with a lower MAPE or perfect peak hour forecast.

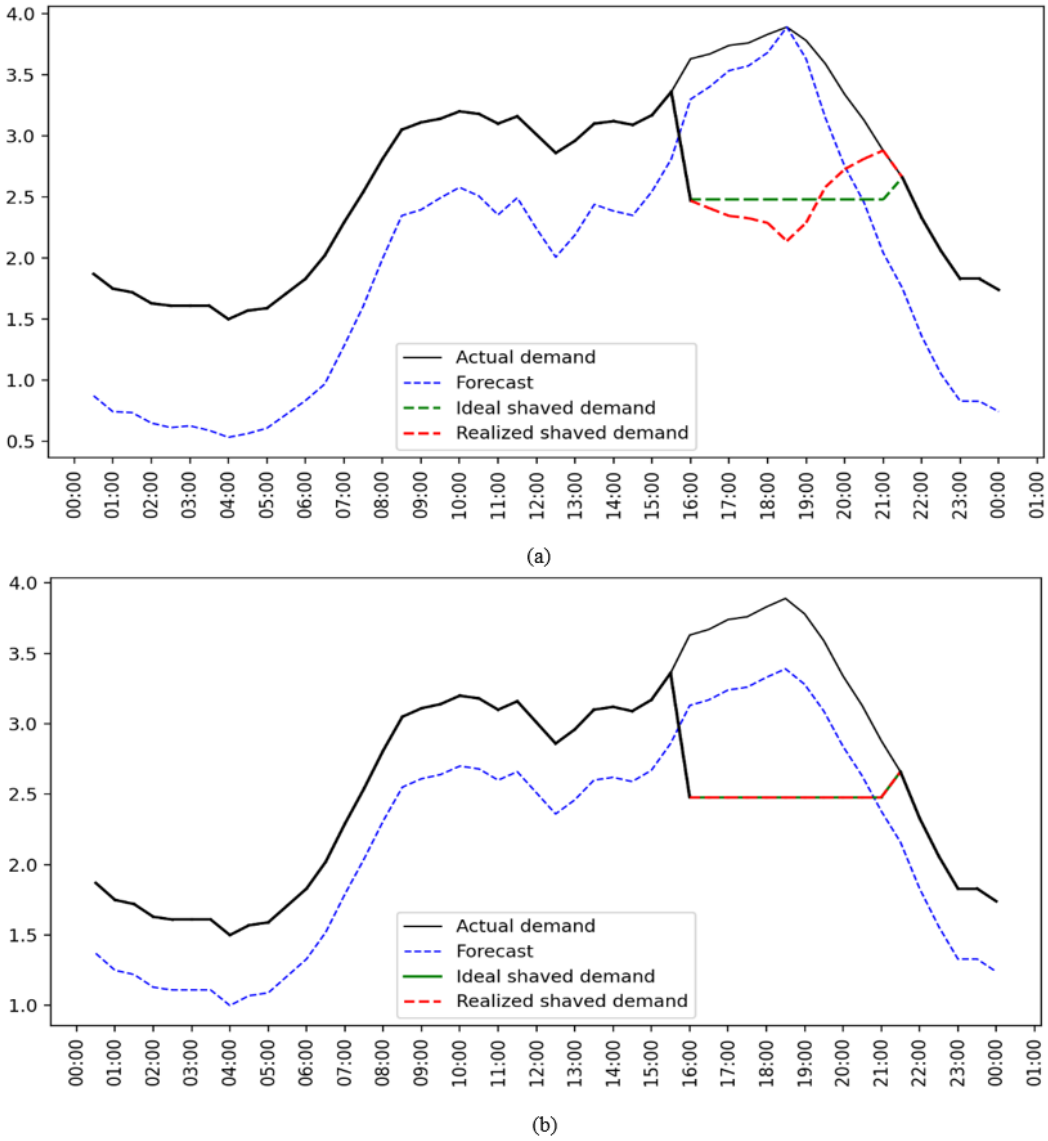


Figure 5.7: Demonstration: load shape forecast for generating ESS discharging profile.

5.3 Evaluating Peak Timing Forecasts with Traditional Error Measures

This section will employ traditional error measures to evaluate the peak timing forecasts. Peak timing forecasts can be derived from hourly load forecasts. In Chapter 3, we learned that the hourly load forecasts are commonly evaluated using MAPE. The measure evaluates how closely hourly forecasts match up with the actual load. The obvious question is whether a lower MAPE also translates to a superior peak timing forecast. We ran an experiment on the aggregated level of both case studies

to investigate if days with a lower MAPE correspond to a higher TPR and vice versa.

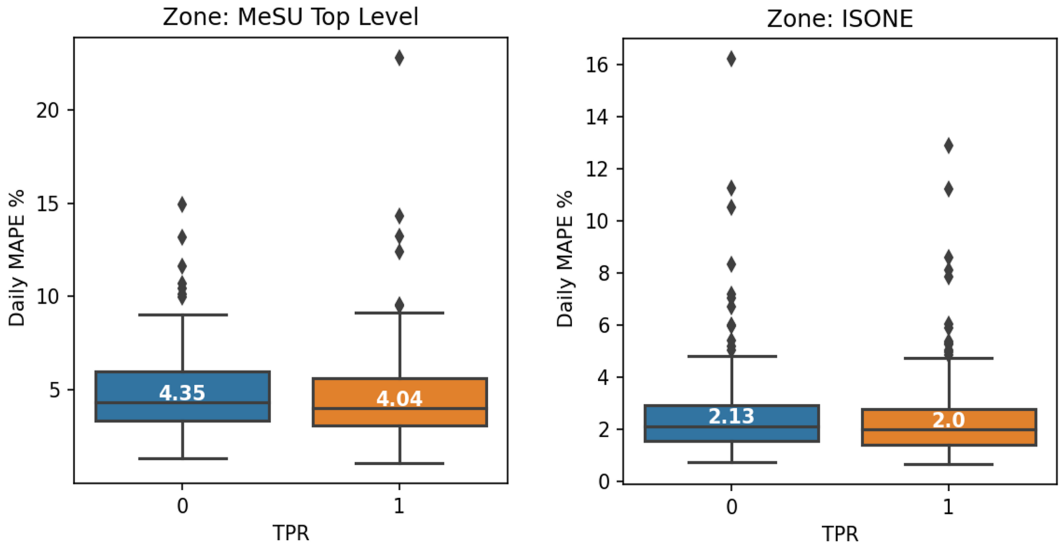


Figure 5.8: Daily MAPE vs. TPR.

In the experiment, we deployed the MLR-based Recency Effect model with d=1 and h=1, thereon called $R_{1,1}$. We evaluated the hourly forecasts from $R_{1,1}$ by daily MAPE, which is obtained by aggregating the absolute percentage error (APE) for each day in the validation year. Peak timing forecasts from $R_{1,1}$ are evaluated with daily TPR. Figure 5.8 presents the distribution of MAPE when TPR = 1 and TPR = 0. Although the median of daily MAPE is slightly lower on days when TPR = 1 than when TPR = 0, the two groups are not statistically different. The p-value from t-test was found at 0.97 and 0.72 for MeSU and ISONE case studies, respectively. There are several days with low MAPE and zero TPR, and several days when MAPE is quite high but TPR is 1. The plots in the first two rows of Figure 5.9 demonstrate some days when daily MAPE was considerably high, yet the peak hour forecast was matched with the actual. The third and fourth rows are the plots with examples when MAPE is low, yet the peak hour forecast is mismatched by more than two hours.

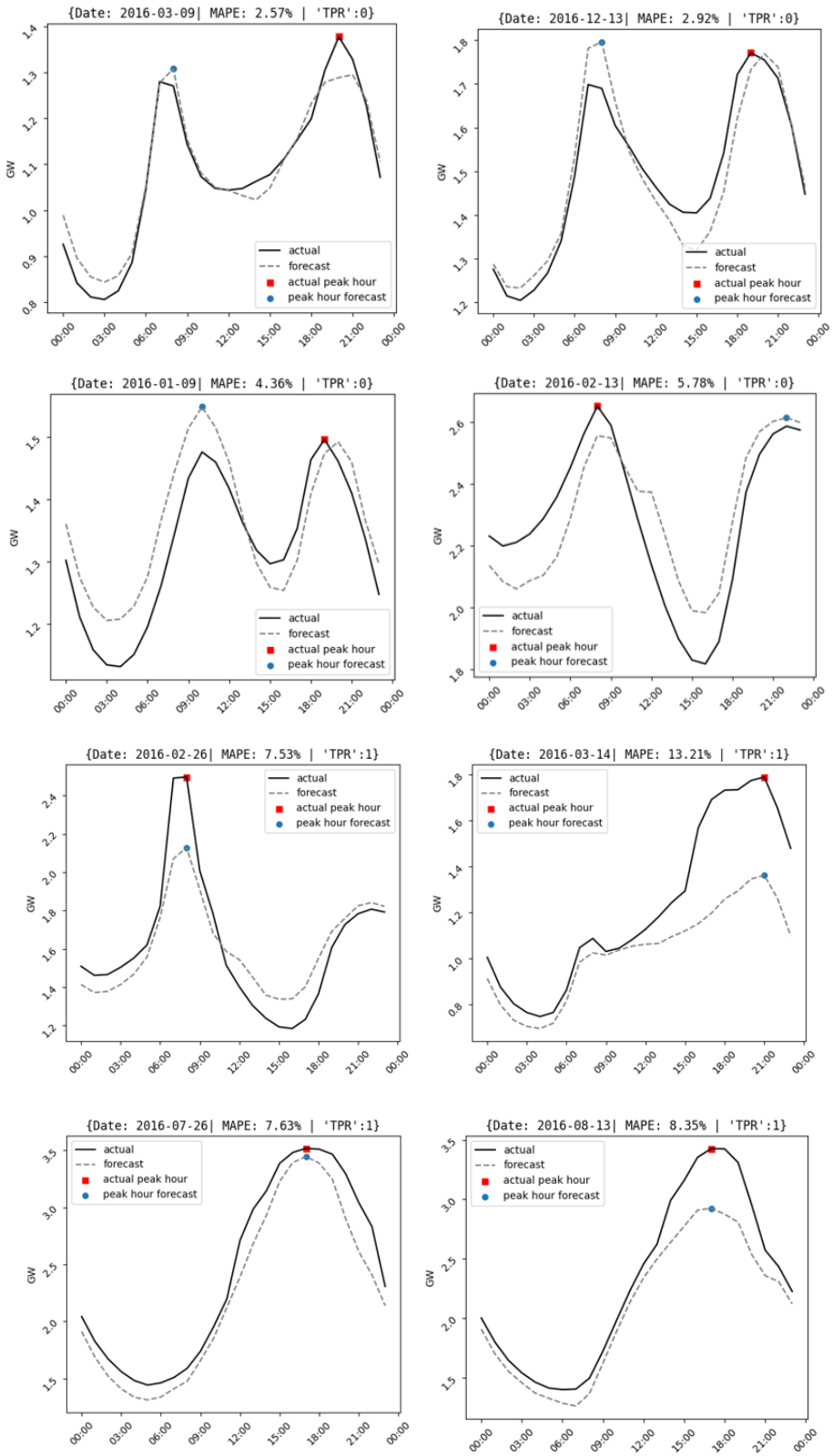


Figure 5.9: Examples when daily MAPE and peak hour forecast accuracy are contrary.

The traditional error measures of load forecasts, such as MAPE, penalize mismatch of the peak load magnitude. However, they also measure off-peak period forecast error. Since these measures are aggregated over a day or a month or a year, the measures do not give a true picture of the peak period. Therefore, the hourly forecast error measures do not explicitly quantify peak load characteristics.

We then investigated evaluating peak timing error using some basic classification error measures (TPR and PPV), and some other common statistics such as F1-score, CSI, and BA. We chose a case with a class ratio of 5:25, which is a suitable example for peak hour forecasting as well as PLD forecasting problems. We also employed two skill scores, Heidke Skill Score (HSS) and Pierce Skill Score (PSS), also known as Hanssen-Kuipers or True Skill Statistic (TSS), as calculated by Equations 5.7 and 5.8 respectively.

$$HSS = \frac{2 \quad (ad - bc)}{(a+c)(c+d) + (a+b)(b+d)} \tag{5.7}$$

$$TSS = \frac{(ad - bc)}{(a+c)(b+d)}, \tag{5.8}$$

where a= TP, b = FP, c= FN, d= TN.

Table 5.3: Demonstration: different scenarios of TP and FP.

Case	Truth Event	Truth non-event	TP	FP	FN	TN	TPR	PPV	F_1	CSI	BA	HSS	TSS
1A	5	25	5	0	0	25	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1B			5	3	0	22	1.00	0.63	0.77	0.63	0.94	0.71	0.88
1C			5	5	0	20	1.00	0.5	0.67	0.50	0.90	0.57	0.80
1D			3	0	2	25	0.60	1.00	0.75	0.60	0.80	0.71	0.60
1E			3	5	2	20	0.60	0.38	0.46	0.30	0.70	0.32	0.40
1F			0	5	5	20	0.00	0.00	0.00	0.00	0.40	-0.20	-0.20
1G			0	10	5	15	0.00	0.00	0.00	0.00	0.30	-0.29	-0.40

Table 5.3 shows the seven scenarios with different true positives and false positives. In the first three cases, the TP = 5, that is all 5 events have been detected, while FP varies between 0 and 5. TPR ignores any effect of FP. All other six error measures aptly differentiate the cases giving heavier penalties when FP is high. In cases 1D and 1E, only 3 of the 5 events have been detected. PPV still gives a perfect score when FP = 0, while the TPR score is the same for both cases ignoring the effect of FP. Hence, both TPR and PPV do not tell the complete story and are not suitable for the task. Cases 1F and 1G both have a TP of 0, but case 1G has a higher FP. The two skill scores and BA respond appropriately, however, F1 and CSI could not distinguish between high and low FP when TP is 0. As per the definition of PLDs, there could be months with no PLD, hence F1 score and CSI scores are not applicable for this application.

Next, we compared how the two skill scores and BA change with different ratios of TPR and TNR. Figure 5.10 shows how the scores change with different values of TNR under a fixed TPR. They all converge to 1 when TP = 1 and FP = 0, but the BA is the least responsive to increases in FP of the three. For example, when TPR = 1, the BA ranges from 0.8 to 1, with different TNR, whereas the HSS ranges from 0.3 to 1. When TP is held constant, we find that both TSS and BA change linearly, while HSS changes non-linearly.

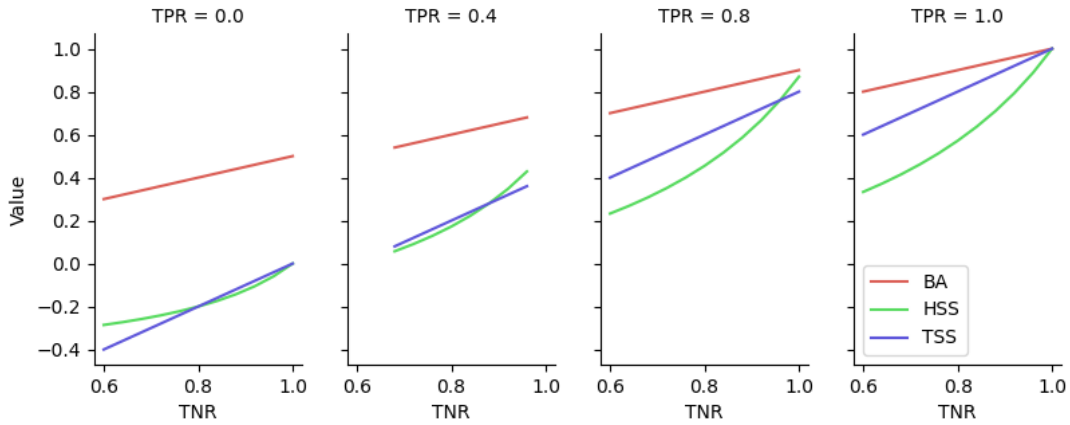


Figure 5.10: Comparing BA, HSS, and TSS for different combinations of TPR and TNR.

Since the number of PLDs in a month may range from 0-10 days, we also tried to understand how the class ratio can influence the evaluation using these three scores. A demonstration with three cases is shown in Table 5.4; case 2 has a class ratio of 10:20, case 3 has a class ratio of 5:25, and case 4 has a class ratio of 2:28. For all the cases the TPR is 1, however, TNR is different. The same number of days have been misclassified in each case. Intuitively we expect that scores would penalize heavier when the perfect TPR is achieved at the cost of misclassifying more events than the true events. However, both BA and TSS ignore the cost of misclassification, awarding higher scores to cases 3 and 4. As a result, BA and TSS are skewed toward the class ratio. A biased error measure would risk the model selection process when validating a model on different months with different class ratios. Of the three, HSS seems to be the most suitable error to differentiate classification accuracy under varying class ratios.

Table 5.4: Demonstration: different class ratios.

Case	Truth Event	Truth non-event	TP	FP	FN	TN	TPR	TNR	BA	HSS	TSS
2	10	20	10	5	0	15	1.00	0.75	0.88	0.67	0.75
3	5	25	5	5	0	20	1.00	0.80	0.90	0.57	0.80
4	2	28	2	5	0	23	1.00	0.82	0.91	0.38	0.82

We have thus analyzed the traditional error measures used in load forecasting and binary event classification problems. Error measurements that have hitherto been used to evaluate load forecasts were found insufficient for the peak timing problem. Common error measures for classification problems have their own limitations, and therefore cannot be applied directly to peak timing errors.

5.4 Proposed Error Measures for Peak Timing Forecasting

In this section, we propose five error measures for evaluating the four different aspects of the peak timing problems; Peak Load Day, Peak Hour, On-peak period, and Peak Shape. We seek application-specific error measurements that can be easily generalized.

5.4.1 Skill Score

The forecast of PLDs is generally required a few weeks or a month ahead. The count of PLDs varies from month to month. There could be months with 5-10 PLDs or months with no PLD. It is crucial to penalize misclassification of days as PLDs, that is, false positives, since there is a limit on calling the DR events. DR programs are labor-intensive and frequently involve actions that may affect a customer's level of comfort, such as modifying thermostat settings. As discussed in the previous section, the skill score HSS seems to be the most reliable statistic for a problem such as this. Theoretically, the score ranges from $-\infty$ to 1, where a negative score means that the

forecast is worse than a random guess, while 1 indicates a perfect score. However, for this application, the score will not go beyond -0.20. Table 5.5 presents some scenarios and corresponding HSS scores with 10 true PLDs in a month of 30 days. However, it should be noted that HSS, unlike BA, is not intuitive and is hard to infer easily from the confusion matrix.

Table 5.5: HSS for different cases with 10 true PLDs in a month.

Scenarios	TP	FP	FN	TN	HSS
1	10	0	0	90	1.00
2	8	2	2	88	0.78
3	5	5	5	85	0.44
4	10	20	0	60	0.40
5	5	10	5	70	0.31
6	0	10	10	80	-0.11
7	0	20	10	70	-0.15

5.4.2 Displacement Score

Instead of categorizing the detection of the peak hour as black or white, we can quantify the accuracy of a peak timing forecasting model in terms of the displacement of the peak hour forecast from the actual peak hour. We will call this peak displacement error. The peak displacement error of a day is denoted by de as in equation 5.9, where PH is the actual peak hour and $\widehat{PH}$ is the peak hour forecast of the day.

$$de = \left|PH - \widehat{PH}\right| \tag{5.9}$$

The displacement error defined above is the horizontal displacement measure of peak load. However, it may be irrelevant to measure if the displacement error extends beyond a few hours. For example, a displacement error of 10 hours is equally detrimental for DR deployment as a displacement error of 5 hours. Moreover, a high

displacement error from a few days could skew the overall metric when aggregated for some days. Therefore, we can choose to cap the displacement error if de is more than a certain number of hours. From the perspective of this reasoning, we define a new score, the Displacement Score (DS). If l is the length of the displacement that we intent to quantify, then DS is represented by equation 5.10

$$DS(de, l) = \max\left\{0,\ \frac{l - de}{l}\right\} \tag{5.10}$$

If we consider $l = 5$, then the values DS vary from 1 to 0 with steps of 0.20 as shown in Table 5.6. The DS gives step scores with each step equally spaced. It is an interpretable score that is, scale-free, and adaptable. A higher score indicates a better forecast.

Table 5.6: DS values for different values of de when $l = 5$

de	DS
0	1
1	0.80
2	0.60
3	0.40
4	0.20
>=5	0

5.4.3 Weighted Displacement Error

DS imposes a linear penalty on error increments. However, we might consider applying a more stringent penalty to higher displacements, rather than just reducing the score linearly. Hence, we propose another error measure named Weighted

Displacement Error (wDE), which is defined as:

$$wDE(de, l) = \min\left(l, \frac{de^2}{l}\right) \tag{5.11}$$

, where de is the displacement error defined by equation 5.9. The larger displacements are punished severely because of the exponential of de. The error measure is capped to the l number of displacements that is intended to be quantified. Beyond the displacement of l hours, the error is capped to a value of l. Figure 5.11 compares the DS and wDE for $l = 5$. Unlike DS, a lower wDE indicates a better forecast. wDE values for $l = 5$ vary nonlinearly between 0 and 5, getting steeper with increasing de. The larger span of the error distinguishes different displacements evidently. On the other hand, as discussed earlier, DS varies linearly between 0 and 1.

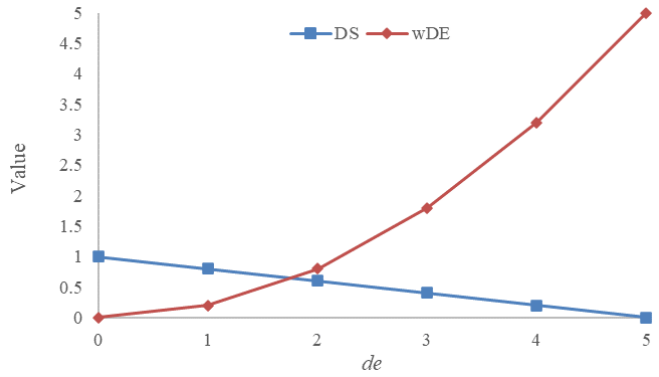


Figure 5.11: Comparing wDE values with DS values for $l = 5$.

5.4.4 Balanced Accuracy

Forecasting of the on-peak period (Q) is crucial for the effective implementation of peak-shaving strategies. It is important to identify all on-peak hours, even at the risk of misclassification. In other words, missing an hour is worse than detecting it at the cost of misclassification. Furthermore, by definition, the maximum number of false positives allowed is 5. Since the class ratio variation for this problem is not huge, we propose to use BA as the error measure for the on-peak period detection problem.

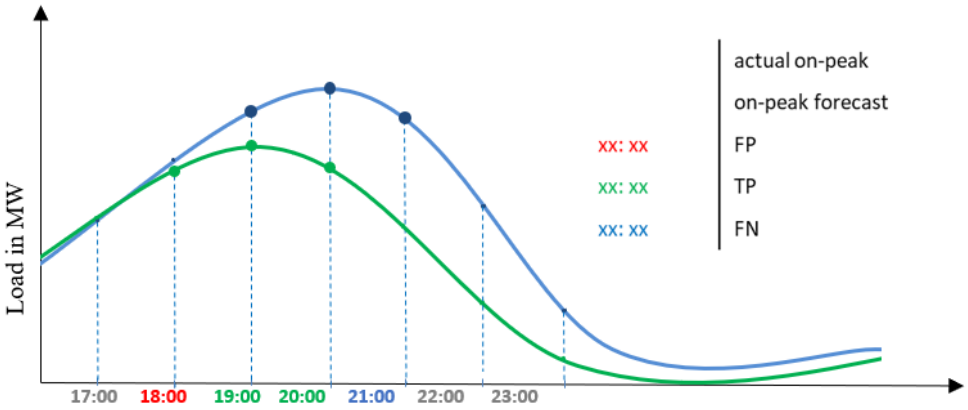


Figure 5.12: Demonstration: BA for on-peak period.

Table 7 shows some of the scenarios with 5 actual on-peak hours and their respective RA and BA scores. For the maximum of 5 FP, BA would range from 0.37 to 1.0.

Table 5.7: BA scores for scenarios with three hours on-peak period.

Scenarios	on-peak period	TP	FP	FN	TPR	TNR	BA
1	3	2	0	1	0.7	1	0.83
2	3	3	2	0	1	0.9	0.95
3	3	3	1	0	1	0.95	0.98
4	3	3	0	0	1	1	1
5	3	0	5	16	0	0.76	0.38

5.4.5 Peak Shape Error

To measure the shape forecast of the on-peak period, we first define the shape of the load profile. Load shape is obtained by normalizing the hourly load (L_t) of the on-peak period (Q), with respect to the daily peak load (P_t). Similarly, load shape $\hat{S}_t$ is obtained by dividing the hourly load forecasts $\hat{L}_t$ with the daily peak forecast $\hat{P}_t$.

$$S_t = \frac{L_t}{P_t} \tag{5.12}$$

$$\hat{S}_t = \frac{\hat{L}_t}{\hat{P}_t} \tag{5.13}$$

Then, the peak shape error (PSE) of a day is the sum of the absolute error of the shape forecast for the on-peak period, S.

$$PSE = \sum_{t \,\epsilon\, Q} \left|(S_t - \hat{S}_t)\right| \tag{5.14}$$

To summarize, we have proposed five error measures for the four forms of peak timing problems as shown in Table 5.8. Two of these errors are adapted from other domains, while Peak hour and Peak Shape error measures are novel in concept and formulation.

Table 5.8: Summary of proposed peak timing error measures.

	Error Measure	Definition
Peak Load Day	Skill Score (SS)	$\text{HSS} = \frac{2\ (\text{ad} - \text{bc})}{(\text{a}+\text{c})(\text{c}+\text{d}) + (\text{a}+\text{b})(\text{b}+\text{d})}$ where, a= TP, b = FP, c= FN, d= TN
Peak Hour	Displacement Score (DS)	$DS = max\left\{0, \frac{l - de}{l}\right\}$
	Weighted Displacement Error (wDE)	$wDE = min\left\{l, \frac{de^2}{l}\right\}$
On-Peak period	Balanced Accuracy (BA)	$BA = \frac{TPR + TNR}{2}$
Peak Shape	Peak Shape Error (PSE)	$PSE = \sum_{t \,\epsilon\, Q} \left\|(S_t - \hat{S}_t)\right\|$

CHAPTER 6: PEAK HOUR FORECASTING BASED ON HOURLY MODELS

Peak hour forecasting is normally a day-ahead task. Traditionally, peak timing forecasts are derived from hourly load forecasts. In this chapter, we will evaluate the peak timing forecasts obtained from the state-of-the-art hourly load forecasting model. The following section will introduce the framework of the forecasting process adopted in this research. We would then describe the benchmark model and investigate some new model selection methodologies relevant to peak timing forecasting.

6.1 Methodology

6.1.1 Benchmark

In this study, the Recency effect model proposed in [79] serves as the benchmark model. The complexity of the model depends on values of the parameters h and d which are determined by investigating the forecasting accuracy of the validation year. The model with the lowest MAPE values in the validation year is selected to forecast the test period. We will explore 100 different models by varying h and d from 0 to 24, and 0 to 3, respectively. For day-ahead forecasting, we augment the model by introducing a lagged load variable. Since in practice, the forecast for the next day is generated in the morning of the current day, the lagged 24 hours of load data is not available for all hours of the next day. Hence, we would use the actual load information 48 hours before the forecast hours. The adopted forecasting framework is similar to the one setup in [94]. By adding the lagged dependent variable (L_{t-48}), the Recency Effect model turns into an Auto-Regressive Recency Effect model (AR-Recency). Equation 6.1 represents the AR-Recency model. We will refer to this model from now on as M0.

$$\begin{aligned}\widehat{L_t} =& \beta_0 + \beta_1\ Trend_t + \beta_2\ D_t + \ \beta_3\ H_t + \beta_4\ M_t + \beta_5 H_t D_t \\ &+\ f\left(T_t\right)\ + \beta_{15}\ L_{t-48} + \ \sum_d f(\widetilde{T_{t,d}}) + \ \sum_h f(T_{t-h})\end{aligned} \tag{6.1}$$

, where $f\left(T_t\right)$ and $\widetilde{T_{t,d}}$ have been described in equations 4.4 and 4.5 respectively. From the load forecast values, we derive the peak hour of the day. If $\hat{L}_{t,j}$ is the daily peak load forecast of day j, then the peak hour forecast of the day, $\widehat{PH_{t,j}}$, is given by Equation 6.2.

$$\widehat{PH_{t,j}} = \left\{H_{t,j} \ \mid \ \hat{L}_{t,j} = \ \hat{P}_{t,j}\right\} \tag{6.2}$$

6.1.2 Forecasting Framework

In the first step, load and temperature data are cleansed and weather stations are selected for each load zone. The process of weather station selection has been followed in accordance with [101]. The next step is data partitioning. The recent three years of the data (2017-2019) are held out for the out-of-sample test for both case studies. The year preceding the test year is designated as the validation year for model selection. For example, for the test year 2018, the year 2017 would be the validation period. The training data length is a fixed two-year period. The model selected from the validation period is retrained on the data from the two years prior to the test year and forecasts generated from the test year. The process is rolled over until we generate forecasts for all three years from 2017 to 2019.

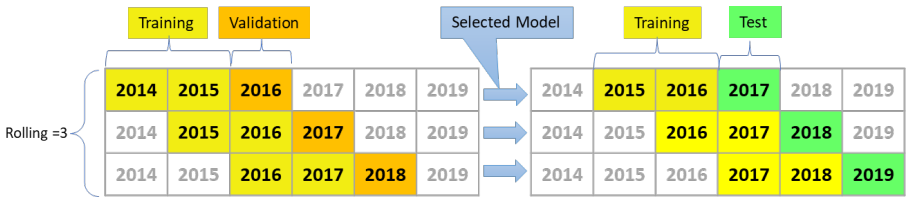


Figure 6.1: Rolling window for three test years.

It is worth noting that we fixed the length as well as the origin of the training

period for each rolling window. That is, the model is trained once in each rolling window to generate day-ahead forecasts (6.1). However, we do leverage the recent load information as per the AR-Recency modeling approach explained in the previous section. Figure 6.2 displays the day ahead load forecasting framework that is deployed in this research.

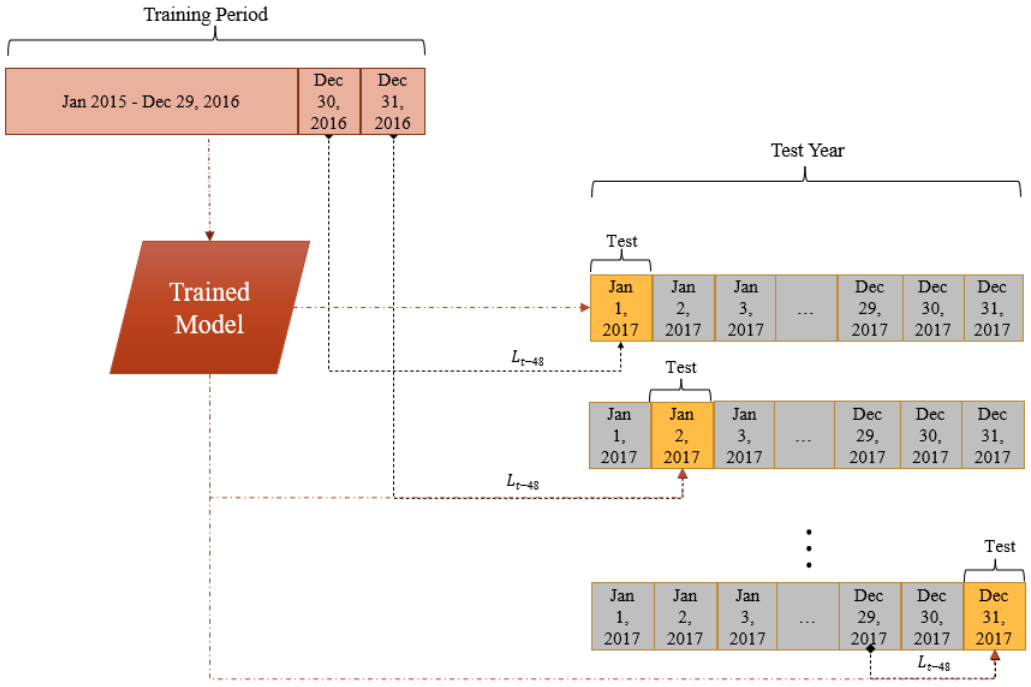


Figure 6.2: Framework for day-ahead load forecasting.

Load forecasting can be ex-ante forecasting or ex-post forecasting, depending upon the kind of information used in the modeling. Ex-post forecasting uses the actual temperature of the test period, which would not be available for the future. Ex-ante forecasting, on the other hand, mimics the real forecasting scenario by using the forecasted temperature for the test period. Thus, ex-post forecasting is a retrospective analysis while ex-ante forecasting predicts future outcomes. The state-of-art approach in the academic literature is based on ex-post forecasting as the temperature forecast data for load zones are not easily available. In this research, however, not only do we have observed readings from weather stations but also the temperature forecast data

for MeSU. Based on weather station selection, the day-ahead temperature forecast from selected weather stations is used for ex-ante load forecasting. The day-ahead temperature forecast data has minimal missing values (less than 0.005%). The actual temperatures are used to fill in these missing values. For ISONE, we have mapped each load zone to the weather forecast available in the data archive maintained by the European Centre for Medium-Range Weather Forecasts (ECMWF)[102]. We were able to get the temperature forecasts for all nine zones of ISONE for three years (2017-2019). Therefore, unlike most prior studies, we will be able to analyze the outcomes of peak timing forecasting using both ex-post and ex-ante methodologies.

6.2 Evaluation of Peak Hour Forecasts

We begin with a thorough analysis of the peak timing forecast outcomes from the benchmark model across both case studies. The forecasts for a test year are generated a) with actual weather data (ex-post forecasting), and b) with the temperature forecast (ex-ante forecasting).

The hourly forecasts from M0 are used to get the daily peak hour forecasts. We analyzed the displacement error (de) of the forecasts from M0. Table 6.1 presents the number of days with $de \leq 1$ and $de \geq 5$ for the ex-post forecasting. For MeSU, we find that 86-89% of days, M0 forecasts are accurate to within an hour. In the case of ISONE, M0 forecasts deliver 69% to 83% of days with de less than one hour. However, the true business case depends on ex-ante forecasting performance. As demonstrated in the tables, the percentage of days with de within an hour declined significantly for ex-ante forecasting.

Ex-ante forecasting results only on PLDs are shown in Table 6.2. These are the critical days when the peak hour forecast accuracy could help make strategic decisions in the DR program implementation and save millions. As we see there are very few PLDs in the year when the M0 peak hour forecast is displaced by more than five hours from the actual. However, a substantial number of PLDs see a displacement of

2-4 hours which could be deleterious. Deploying DR in the wrong timeframe could lead to rebounds in peak demand that are even higher than the original peak.

Figure 6.3 displays the M0 wDE values for ex-post and ex-ante forecasting along with the best wDE values that may be attained theoretically using the global optimum Recency effect model. This further demonstrates the scope for improvement of peak timing forecasts. In this research, we will use wDE as the error measure to evaluate the peak hour forecasts.

Table 6.1: Peak hour displacement errors of the benchmark model.

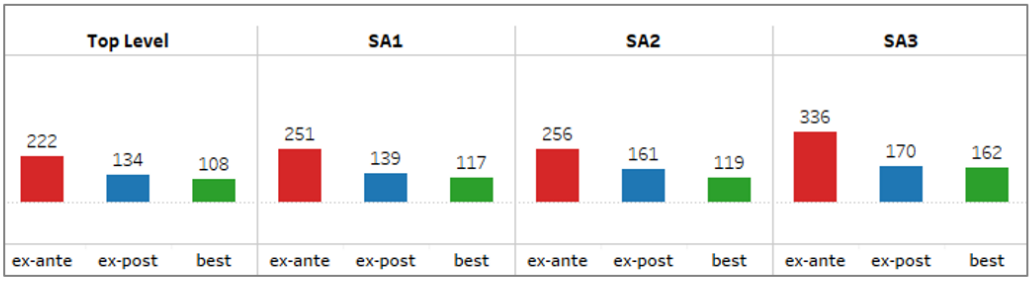
	MeSU ex-post			MeSU ex-ante		
Year 2016	# Days with $de \geq 5$	# Days with $de \leq 1$	% Days with $de \leq 1$	# Days with $de \geq 5$	# Days with $de \leq 1$	% Days with $de \leq 1$
Top Level	17	326	89%	24	277	76%
SA1	17	325	89%	22	242	66%
SA2	21	313	86%	32	274	75%
SA3	20	314	86%	35	228	62%

	ISONE ex-post			ISONE ex-ante		
Year 2017	# Days with $de \geq 5$	# Days with $de \leq 1$	% Days with $de \leq 1$	# Days with $de \geq 5$	# Days with $de \leq 1$	% Days with $de \leq 1$
CT	20	296	81%	27	281	77%
ISONE	21	302	83%	25	284	78%
ME	70	252	69%	67	249	68%
NEMASSBOST	42	263	72%	49	248	68%
NH	41	283	78%	41	268	73%
RI	33	293	80%	42	265	73%
SEMASS	13	321	88%	16	326	89%
VT	36	301	82%	35	303	83%
WCMASS	25	295	81%	24	283	78%

Table 6.2: Peak hour displacement errors of the benchmark model for PLDs.

	MeSU			
Year 2017	#PLD	# Days with $de \geq 5$	# Days with $4 \leq de \leq 2$	# Days with $de \leq 1$
Top	81	2%	17%	80%
SA1	61	2%	52%	46%
SA2	86	1%	24%	74%
SA3	58	3%	52%	45%

	ISONE			
Year 2017	#PLD	# Days with $de \geq 5$	# Days with $4 \leq de \leq 2$	# Days with $de \leq 1$
CT	47	0%	40%	60%
ISONE	41	0%	44%	56%
ME	27	19%	19%	63%
NEMASS	40	3%	33%	65%
NH	36	3%	33%	64%
RI	53	0%	43%	57%
SEMASS	55	0%	11%	89%
VT	48	4%	2%	94%
WCMASS	39	3%	23%	74%



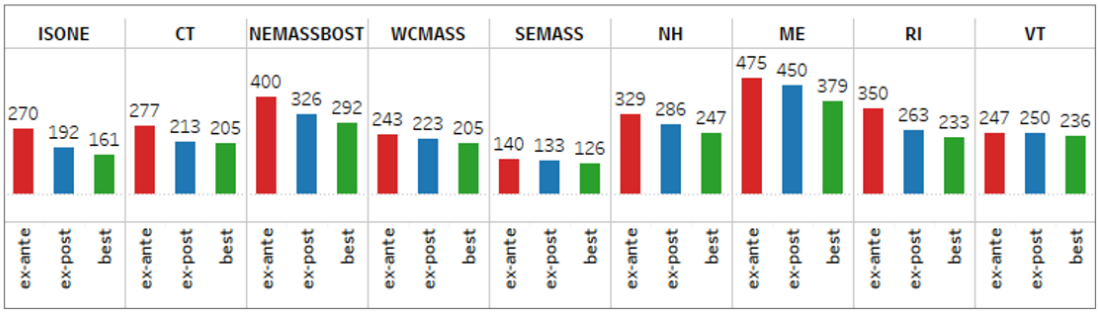


Figure 6.3: wDE values from ex-ante forecasts vs. wDE values from ex-post forecasts vs. theoretical best wDE values from Recency effect framework.

6.3 Model Selection for Peak Hour Forecasting

In this section, we will examine model selection methods for peak timing forecasting using the AR-Recency model. By varying the *d*-*h* pair, we have 100 different models to choose from. In the first subsection, we will use different peak timing error measures to select a model. The test results are evaluated for ex-post as well as ex-ante forecasting by comparing them with the test results of M0. The analysis additionally evaluates the test results exclusively on PLDs. In the second subsection, we further explore selecting a model based on performance on PLDs only and selecting a model on ex-ante forecast analysis.

6.3.1 Model Selection with Peak Timing Error Measures

Traditionally, model selection is done by evaluating the forecasts of the validation period using MAPE of the hourly load forecast. For peak hour forecasting models, we may try to select models using the peak hour error measure, i.e., wDE. Hence, we will generate forecasts from two tracks. Track 1 is the traditional way adopted for the benchmark M0, where *d*-*h* pair is selected with the lowest MAPE of all hours. Track 2 selects *d*-*h* pair based on the lowest wDE values for the validation period.

Figure 6.4 displays how *d*-*h* pair selected on peak timing error measure differs from *d*-*h* pair selected on MAPE for the year 2016. Models chosen based on MAPE are simpler than those chosen based on peak timing errors in both case studies.

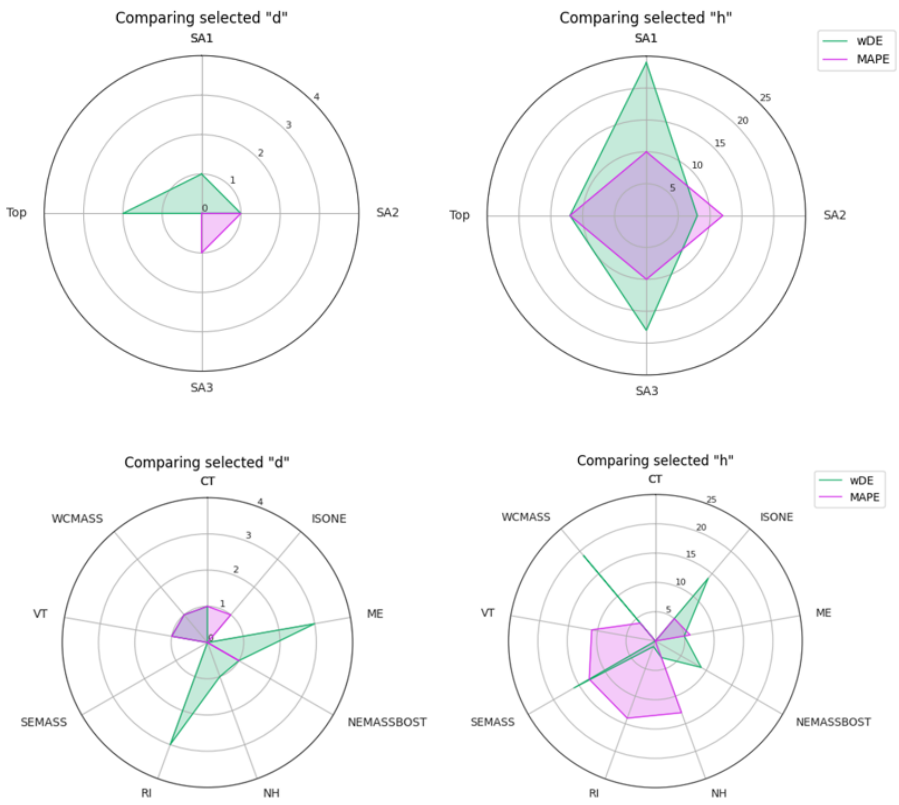


Figure 6.4: Comparing *d*-*h* pairs selected with MAPE and *d*-*h* pair selected with wDE.

Table 6.3 shows the wDE and MAPE values from different AR-Recency models for 2016 of *top level*. A cooler color in the background of the cell indicates a lower error. Similarly, Table 6.4 shows the wDE and MAPE of values from different AR-Recency effect models for 2016 for ISONE. In both case studies, the convergence of wDE value to the local optimum value is not smooth, as the difference in wDE values between two neighboring models is abrupt. For example, wDE value for *d*-*h* pair $(3, 7)$ is 140, which increases to 151 with $(3, 8)$ and again reduces to 130 with $(3, 9)$. Due to the nature of the problem, wDE is a discrete score. Therefore, we do not see a progressive convergence to a local minimum for peak timing scores, unlike MAPE, where the convergence to local minima is apparent.

Table 6.3: MAPE and wDE values for *top level* (MeSU) on the validation data (year 2016).

MAPE

h/ d	0	1	2	3
0	5.40	4.97	5.06	5.29
1	5.14	4.91	5.01	5.28
2	5.01	4.85	4.97	5.23
3	4.94	4.82	4.95	5.21
4	4.89	4.82	4.95	5.20
5	4.85	4.81	4.95	5.21
6	4.83	4.81	4.96	5.21
7	4.82	4.81	4.98	5.22
8	4.81	4.82	5.00	5.25
9	4.80	4.82	5.01	5.26
10	4.79	4.83	5.02	5.28
11	4.78	4.84	5.03	5.29
12	4.77	4.85	5.04	5.30
13	4.78	4.86	5.06	5.32
14	4.79	4.88	5.08	5.35
15	4.80	4.90	5.10	5.38
16	4.81	4.91	5.12	5.40
17	4.81	4.92	5.13	5.41
18	4.82	4.93	5.14	5.43
19	4.84	4.94	5.15	5.44
20	4.87	4.94	5.15	5.43
21	4.90	4.96	5.15	5.45
22	4.93	4.97	5.17	5.48
23	4.98	5.00	5.20	5.51
24	5.01	5.00	5.21	5.54

wDE

h/ d	0	1	2	3
0	184	168	148	151
1	172	156	159	158
2	171	152	140	129
3	166	146	146	139
4	164	153	136	132
5	164	153	144	138
6	166	164	143	143
7	159	158	149	140
8	162	179	162	151
9	160	163	149	130
10	164	165	139	134
11	159	142	131	118
12	161	145	124	118
13	164	167	158	144
14	162	160	145	143
15	153	175	152	147
16	150	157	155	145
17	157	159	137	147
18	169	154	144	148
19	157	151	144	141
20	155	157	152	145
21	157	165	160	157
22	137	177	161	158
23	154	154	170	169
24	166	170	170	159

Table 6.4: MAPE and wDE values for ISONE on the validation data (year 2016).

MAPE

h/d	0	1	2	3
0	2.93	2.55	2.62	3.01
1	2.81	2.52	2.60	2.97
2	2.73	2.50	2.59	2.95
3	2.66	2.48	2.58	2.93
4	2.60	2.47	2.57	2.91
5	2.56	2.47	2.57	2.91
6	2.54	2.47	2.59	2.92
7	2.53	2.49	2.60	2.94
8	2.53	2.50	2.62	2.94
9	2.53	2.51	2.62	2.94
10	2.53	2.52	2.64	2.95
11	2.53	2.53	2.65	2.97
12	2.52	2.54	2.66	2.97
13	2.52	2.56	2.67	2.98
14	2.52	2.58	2.69	2.99
15	2.53	2.60	2.70	2.99
16	2.53	2.61	2.72	3.01
17	2.54	2.63	2.74	3.01
18	2.56	2.65	2.75	3.02
19	2.58	2.67	2.77	3.04
20	2.60	2.69	2.79	3.06
21	2.63	2.71	2.82	3.08
22	2.65	2.73	2.84	3.10
23	2.67	2.75	2.86	3.12
24	2.69	2.76	2.87	3.13

wDE

h/d	0	1	2	3
0	158	158	153	172
1	195	182	202	211
2	201	195	184	186
3	226	179	203	180
4	236	205	214	210
5	220	194	206	196
6	205	190	194	184
7	222	169	178	174
8	203	163	178	175
9	193	158	167	162
10	184	152	161	168
11	188	160	159	168
12	189	167	163	166
13	165	156	164	168
14	148	149	166	162
15	169	154	162	167
16	156	160	153	168
17	155	172	164	157
18	151	170	167	177
19	174	183	178	189
20	173	174	194	194
21	169	178	183	192
22	181	184	184	182
23	178	188	186	177
24	190	194	187	190

Since peak timing error measures did not seem robust for model selection, we additionally explored some other error measures that are continuous and relevant to peak timing. The following seven tracks were created using different error measures for selection described below:

- *d*-*h* pair selected with the lowest MAPE of daily peak load forecast (Track 3),
- *d*-*h* pair selected with the lowest MAPE of load forecast for peak hour +/-2 hours (Track 4),
- *d*-*h* pair selected with the lowest MAPE of load forecast at the actual peak hour and forecasted peak hour (Track 5),
- *d*-*h* pair selected with the lowest Shape Absolute Error (SAE) of shape forecast for peak hour +/-2 hours (Track 6),

- *d-h* pair selected with the lowest SAE of shape forecast for actual peak hours (Track 7a),
- *d-h* pair selected with the lowest SAE of shape forecast for the actual and forecasted peak hours (Track 7b),
- *d-h* pair selected with exponential SAE of shape forecast for all hours (Track 8).

Tracks 3, 4, and 5 are based on the MAPE of the load forecasts at the peak hours. Tracks 6, 7a, and 7b are based on the shape forecast errors at the peak hours. Track 8 is based on the shape error of the entire load profile raised exponentially. The idea is to methodically undervalue the off-peak hours and evaluate only peak hours. Since the shape error values belong to interval [0,1], the exponential of a number between 0 and 1 diminishes lower values while amplifying values close to 1. In this experiment, we have used 16^{th} exponent.

Along with Track 1 and 2, we get nine models selected for each zone. However, the models selected from Track 7a and Track 7b are the same for all zones in both case studies. Therefore, we present their results as Track 7. Table 6.5 presents the wDE values from Track 0, which is also M0.

The results from the seven tracks are shown as percentage changes in wDE values in reference to M0's wDE values for each zone. The % values are color-coded for easy reading with green font indicating a reduction in wDE errors by $\geq 5\%$, which denotes significantly improved results. The red font indicates that the wDE values increased by more than 5%, that is, the forecasts worsened significantly. While the yellow font indicates a neutral effect where the difference in wDE values is between $\pm 5\%$. We will be following a similar format of depicting results henceforth in this chapter.

Table 6.5: wDE values from M0, and their comparison with those from the seven tracks on the validation data (year 2016) of MeSU.

Ex-post forecasting

Zone	M0 (wDE)	Track[2]	Track[3]	Track[4]	Track[5]	Track[6]	Track[7]	Track[8]
Top Level	161	-23%	-3%	-3%	-3%	-8%	-19%	-10%
SA1	140	-14%	-10%	-10%	-10%	-10%	-14%	-5%
SA2	138	-9%	0%	-1%	-9%	6%	-1%	0%
SA3	156	-11%	14%	14%	0%	9%	1%	5%

Ex-ante forecasting

Zone	M0	Track[2]	Track[3]	Track[4]	Track[5]	Track[6]	Track[7]	Track[8]
Top Level	276	-10%	6%	6%	6%	5%	-10%	6%
SA1	283	2%	4%	4%	4%	6%	2%	3%
SA2	330	-14%	19%	18%	-14%	-28%	22%	-27%
SA3	347	11%	0%	0%	0%	1%	13%	11%

Table 6.5 presents the validation results of MeSU. The ex-post forecasting results from Track 2 are the theoretically best outcomes because the track selects the optimal *d*-*h* pairs based on wDE values. The results from Track7 and Track5 look promising. However, the ex-ante forecasting validation results are contradictory. Even with Track 2, the peak timing errors do not show any improvement on the M0 except for a small improvement in the SA1 and SA3 zones.

6.3.1.1 Results for MeSU

Table 6.6 displays the test results for ex-post forecasting for MeSU. We also analyzed the performance of models on PLDs exclusively as these are the days when the accuracy of forecasting peak timing matters the most. Overall, the model selection approach using the different tracks does not outperform the M0 for ex-post forecasting. When analyzing results for PLDs only, we see significant improvement in the test year 2017 for *top level* and SA2 with shaped error-based Tracks. We also see improvement in the year 2019 for SA1 with Track2. However, none of the tracks demonstrate consistent improvement year on year for any zone. The ex-ante forecast-

ing results also have similar conclusions as shown in Table 6.7. For *top level* PLDs, we see consistent improvements from some of the tracks such as Track 2, Track 3, and Track 7 on all test years. However, no track shows consistent improvement in all test years and for all zones.

Table 6.6: Performance of models selected from the seven tracks for ex-post forecasting on test data for MeSU.

Ex-post forecasting									
Zone	Test year	M0 (wDE)	Track^2	Track^3	Track^4	Track^5	Track^6	Track^7	Track^8
Top Level	2017	134	10%	15%	15%	15%	10%	9%	15%
	2018	125	2%	12%	4%	0%	1%	2%	4%
	2019	151	-7%	-9%	-8%	-9%	17%	2%	-8%
SA1	2017	139	13%	15%	15%	15%	14%	13%	2%
	2018	133	0%	17%	7%	20%	2%	0%	14%
	2019	145	11%	4%	11%	4%	-5%	2%	3%
SA2	2017	161	-13%	-2%	1%	-13%	-18%	-6%	-14%
	2018	131	4%	-5%	-11%	-5%	-7%	3%	15%
	2019	243	3%	0%	-2%	3%	-2%	3%	5%
SA3	2017	170	10%	5%	5%	0%	6%	15%	10%
	2018	189	-4%	6%	2%	6%	2%	0%	-4%
	2019	140	6%	6%	4%	3%	32%	8%	-3%
Ex-post forecasting for PLDs only									
Zone	Test Year	M0 (wDE)	Track^2	Track^3	Track^4	Track^5	Track^6	Track^7	Track^8
Top Level	2017	22	38%	4%	4%	4%	16%	-6%	4%
	2018	8	-21%	9%	8%	-8%	4%	0%	8%
	2019	8	-14%	-2%	-12%	-2%	-18%	-12%	-12%
SA1	2017	6	-35%	-30%	-30%	-30%	-35%	0%	-33%
	2018	13	0%	6%	-2%	3%	-3%	0%	-2%
	2019	10	27%	12%	12%	12%	24%	10%	22%
SA2	2017	13	47%	-2%	7%	58%	-7%	0%	-16%
	2018	12	18%	12%	61%	12%	10%	4%	67%
	2019	17	-11%	3%	3%	11%	11%	0%	3%
SA3	2017	18	11%	-4%	-4%	11%	-1%	39%	7%
	2018	21	35%	18%	41%	18%	41%	32%	0%
	2019	9	-4%	0%	-9%	-9%	2%	56%	40%

Table 6.7: Performance of models selected from the seven tracks for ex-ante forecasting on test data of MeSU.

Ex-ante forecasting									
Zone	Test Year	M0 (wDE)	Track^2	Track^3	Track^4	Track^5	Track^6	Track^7	Track^8
Top Level	2017	222	4%	8%	8%	8%	5%	3%	8%
	2018	207	5%	-21%	9%	8%	9%	5%	9%
	2019	214	-10%	-4%	1%	-4%	-24%	-20%	1%
SA1	2017	251	13%	9%	9%	9%	1%	13%	2%
	2018	302	0%	5%	4%	-8%	5%	0%	6%
	2019	302	-4%	3%	-4%	3%	2%	-2%	2%
SA2	2017	256	2%	22%	16%	2%	-12%	23%	-3%
	2018	248	-1%	5%	21%	5%	16%	2%	13%
	2019	358	-4%	1%	0%	-10%	-8%	-4%	-1%
SA3	2017	336	0%	-11%	-11%	0%	2%	14%	9%
	2018	366	4%	-4%	-5%	-4%	-5%	6%	4%
	2019	332	-21%	-16%	-22%	-22%	-2%	-30%	-16%
Ex-ante forecasting for PLDs only									
Zone	Year	M0	Track^2	Track^3	Track^4	Track^5	Track^6	Track^7	Track^8
Top Level	2017	33	-7%	-13%	-13%	-13%	34%	5%	-13%
	2018	40	-13%	-38%	11%	12%	19%	-13%	11%
	2019	39	-33%	-1%	-41%	-1%	-58%	-55%	-41%
SA1	2017	45	15%	5%	5%	5%	12%	15%	7%
	2018	35	0%	61%	9%	-3%	27%	0%	40%
	2019	51	-28%	6%	-28%	6%	6%	-27%	8%
SA2	2017	37	10%	24%	22%	10%	-9%	42%	-3%
	2018	43	0%	13%	40%	13%	36%	-7%	30%
	2019	47	-26%	10%	6%	-26%	-19%	-26%	14%
SA3	2017	52	7%	-26%	-26%	0%	9%	2%	2%
	2018	54	4%	5%	14%	5%	14%	41%	4%
	2019	54	-4%	10%	-20%	-28%	-12%	-24%	-19%

6.3.1.2 Results for ISONE

Based on the model selection on the validation period, Table 6.8 shows the wDE values on the test data (2018-2019) for ISONE. The improvement in wDE values for the M0 model is colored green while red color indicates that the wDE values have worsened. The conclusion is similar to the MeSU case study, we do not find any track giving consistent improvement.

Table 6.8: Performance of models selected from the seven tracks for ex-ante forecasting on test data of ISONE.

Ex-ante forecasting									
Zone	Year	M0 (wDE)	Track^2	Track^3	Track^4	Track^5	Track^6	Track^7	Track^8
ISONE	2018	188	11%	-7%	0%	-7%	0%	9%	0%
	2019	142	10%	14%	-5%	5%	-2%	-9%	5%
CT	2018	244	-3%	-2%	-2%	-8%	-4%	-1%	5%
	2019	191	7%	7%	4%	8%	0%	7%	-7%
ME	2018	383	-3%	11%	10%	-1%	2%	-3%	5%
	2019	399	-4%	-11%	0%	-7%	-4%	-4%	2%
NH	2018	297	-1%	5%	-1%	-1%	0%	0%	0%
	2019	194	32%	18%	6%	20%	10%	18%	18%
RI	2018	333	-12%	3%	0%	-6%	-6%	1%	-6%
	2019	234	-4%	-2%	-2%	5%	-1%	-6%	-6%
VT	2018	211	0%	0%	-1%	0%	-1%	7%	-6%
	2019	240	-8%	-12%	-7%	-12%	8%	-7%	-4%
NEMASS	2018	140	-14%	0%	-10%	-3%	-16%	-18%	2%
	2019	86	18%	-1%	0%	0%	6%	-3%	-6%
SEMASS	2018	285	12%	12%	8%	1%	6%	-2%	-3%
	2019	252	-1%	10%	10%	8%	16%	8%	8%
WCMASS	2018	254	-11%	-5%	-5%	2%	-2%	-8%	-2%
	2019	190	5%	5%	8%	5%	11%	11%	3%
Ex-ante forecasting for PLDs only									
Zone	Year	M0 (wDE)	Track^2	Track^3	Track^4	Track^5	Track^6	Track^7	Track^8
ISONE	2018	13.2	35%	-14%	0%	-14%	50%	35%	50%
	2019	7.2	47%	8%	-8%	25%	-3%	-17%	14%
CT	2018	17.4	-9%	1%	-2%	-31%	-1%	1%	-24%
	2019	28	-19%	1%	0%	-1%	-3%	-19%	-4%
ME	2018	27.2	52%	43%	21%	28%	43%	-12%	29%
	2019	24.8	23%	-2%	0%	32%	7%	23%	18%
NH	2018	19.4	-29%	-26%	-22%	-25%	-15%	-15%	-15%
	2019	20.4	16%	15%	5%	12%	5%	15%	11%
RI	2018	26.2	24%	1%	0%	50%	25%	13%	25%
	2019	35.8	-4%	-17%	-17%	-4%	6%	-6%	-6%
VT	2018	35.2	26%	-1%	-2%	0%	11%	13%	-3%
	2019	19.2	1%	21%	9%	21%	106%	25%	2%
NEMASS	2018	17.4	22%	0%	47%	-8%	32%	26%	29%
	2019	9.8	29%	37%	0%	0%	96%	45%	-10%
SEMASS	2018	27.6	13%	4%	-7%	-23%	-14%	-37%	-27%
	2019	13.8	30%	7%	7%	16%	3%	16%	16%
WCMASS	2018	31.4	-38%	3%	3%	-10%	12%	-26%	12%
	2019	17.6	25%	33%	-13%	33%	5%	5%	13%

6.3.1.3 Discussion

As we have seen none of the model selection methods using peak timing error measures outperform the benchmark model M0. Figure 6.5 plots the best *d*-*h* pair found for the year with MAPE, wDE, Peak MAPE and SAE. While *d*-*h* pairs found on MAPE stay close by over the years, the *d*-*h* pairs on wDE, Peak MAPE and SAE are spread out over the test years. By nature of the problem, all these peak timing error measures work by selecting the errors of specific hours. The selection of hours is a thresholding process. Using a threshold that changes for each day makes the selection process fragile.

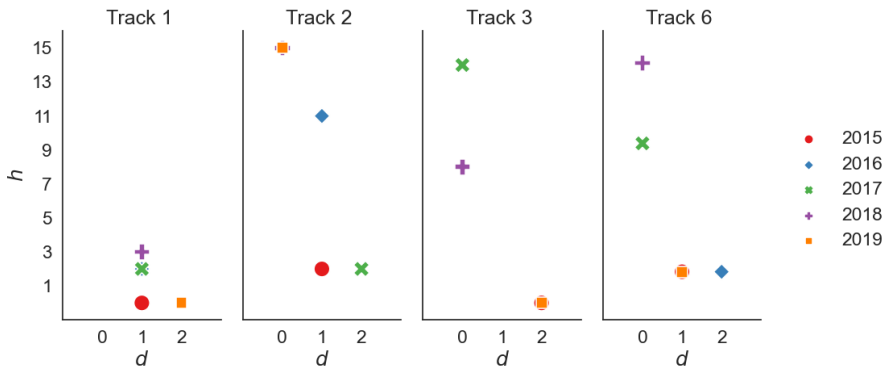


Figure 6.5: *d*-*h* pairs selected for each year for *top level* of MeSU by different tracks.

Secondly, the peak timing problem is a horizontal displacement problem, whereas the goal of any curve-fitting model is to minimize the vertical displacements between the actuals and predictions. If the vertical displacements are minimized, the horizontal displacement error will be lowered as well (Figure 6.6). However, it does not hold true otherwise. When we try to select a model based on peak timing error, we select a model that gives the minimum horizontal displacement. Since the curve fitting models work on vertical displacement, the model selection does not guarantee that the selection would work best for out-sample horizontal errors. Therefore, choosing the model that minimizes vertical displacements, i.e., choosing based on MAPE, is the pragmatic approach.

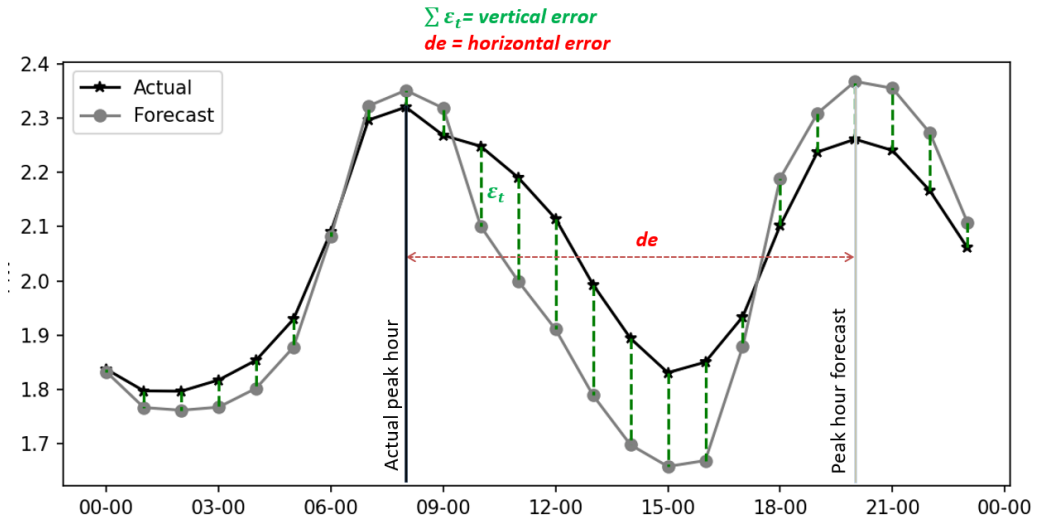


Figure 6.6: Vertical displacement errors and horizontal displacement error on a day under *top level* of MeSU (1/13/2017).

6.3.2 Model Selection based on Ex-ante Forecast Errors

Practitioners use ex-ante forecasts to make decisions [94]. Hence, we will try another approach of model selection based on the ex-ante forecasts analysis of the validation period. In this framework, we will use the temperature forecasts for the validation year instead of the actual temperatures. The idea here is to eliminate any obscured effect of temperature forecasting errors on the model selection. It is important to note that we do not use temperature forecasts for model fitting since the noise in the temperature forecast may lead to biased parameter estimation. The topic is reserved for future investigation and is beyond the scope of this thesis.

Method A is the state-of-the-art approach using ex-post forecast for model selection, the benchmark model M0 uses the same method in this research. Method B uses the temperature forecast for the validation period.

1. Method A: Selecting Model on forecasting accuracy of all days of validation year using actual temperature (M0).

2. Method B: Selecting Model on forecasting accuracy of all days of validation year

using temperature forecast.

6.3.2.1 Results and Discussion

The evaluation of peak hour ex-ante forecasts from the two methods is done on the three test years (2017-2019) for MeSU. Since, in the case of ISONE the temperature forecast data is only available for the period 2017-2019, Method B can only be validated for two years. Hence, for ISONE we have two test years (2018-2019). Table 6.9 displays the outcomes of the three test years from MeSU. Method B performed better than Method A in all zones except for SA1 in 2018. However, overall the forecasts are not very different leading to an average improvement of 5%. The test results of ISONE as shown in Table 6.10. Similar to MeSU, the forecasts from the two methods are similar for most of the zones, but overall, the peak hour forecasts from Method B are slightly worse than those of Method A.

Table 6.9: Performance of models selected based on temperature forecast on test data of MeSU.

Zone	Test Year	Method A ~ M0 (wDE)	Method B
Top Level	2017	222	-16%
	2018	207	-10%
	2019	214	-16%
SA1	2017	251	-5%
	2018	302	6%
	2019	302	-6%
SA2	2017	256	-4%
	2018	248	-3%
	2019	358	-4%
SA3	2017	336	-3%
	2018	366	-5%
	2019	332	-4%
Total		3394	-5%

Table 6.11 compares the Recency effect parameters from the two methods for the Top Level of MeSU and ISONE. For MESU, the parameters for Method A are very similar to those of Method B in the validation years 2017 and 2018. The models

selected by method B are simpler and the year-on-year variation is also minimal. However, for ISONE, the parameters chosen from Method B are very different from those chosen from Method A.

Table 6.10: Performance of models selected based on temperature forecast on test data of ISONE.

Zone	Test Year	Method A ~ M0 (wDE)	Method B
CT	2018	244	4%
	2019	191	-9%
ISONE	2018	188	0%
	2019	142	10%
ME	2018	383	5%
	2019	399	-8%
NEMASS	2018	285	4%
	2019	252	-2%
NH	2018	297	1%
	2019	194	22%
RI	2018	333	0%
	2019	234	10%
SEMASS	2018	140	-3%
	2019	86	9%
VT	2018	211	-1%
	2019	240	-4%
WCMASS	2018	254	0%
	2019	190	13%
Total		4263	4%

Table 6.11: *d*-*h* pairs selected from the two methods for the validation year 2016.

Top Level (MeSU)			
Method A		Method B	
d	*h*	*d*	*h*
0	12	1	2
1	3	1	2
1	4	1	3

ISONE			
Method A		Method B	
d	*h*	*d*	*h*
0	11	-	-
1	5	0	11
2	13	0	12

We further investigated the temperature forecast errors, to understand the difference in two case studies. Figure 6.7 plots the temperature forecast errors (T_{err}) of

the two case studies for each of the three years. We find that temperatures are underestimated for ISONE, resulting in a bias in the T_{err}. While, for MeSU, the median T_{err} stays close to zero. Moreover, T_{err} for MeSU appears to be a stationary series, while for ISONE, T_{err} shows seasonality (Figure 6.8). Because of the inherent bias in the T_{err} in the case study of ISONE, parameters selected from Method B are very different when compared to those selected from Method A. Hence, model selection for the Recency effect framework using ex-ante forecasts analysis on the validation period can lead to better peak hour hours if the temperature forecast errors are not biased.

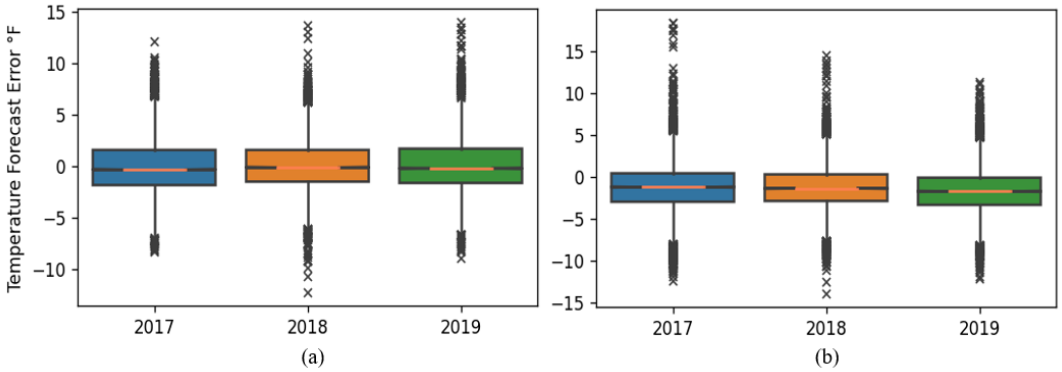


Figure 6.7: Boxplot of temperature forecast errors for (a) MeSU, and (b) ISONE.

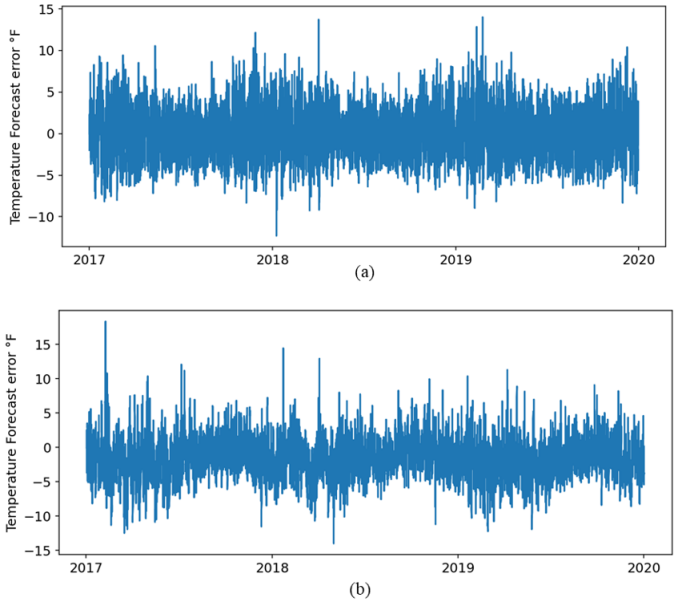


Figure 6.8: Time series of temperature forecast errors for (a) MeSU, and (b) ISONE.

CHAPTER 7: PROPOSED MODELS FOR PEAK HOUR FORECASTING

This chapter explores two different frameworks we propose for peak hour forecasting. The first framework is based on Direct Peak hour forecasting using a binomial classification model. The second framework is a two-stage process built on the hourly load forecasting model. We will only be focusing on short-term ex-ante peak hour forecasting. Furthermore, temperature forecast data will be used for the validation year to select a model. We will use Weighted Displacement Error (wDE) as the error measurement for peak hour forecasts, which is described in section 5.4.3. Since, for ISONE, we have temperature forecast data only for three years (2017-2019), data from 2017 will serve as the validation period for the case study. For MeSU, models will be selected by validating on the year 2016. In the third section, we will analyze how the two proposed models perform in comparison to the benchmark model M0. Test results from 2017-2019 will be analyzed for MeSU, while results from 2018-2019 will be analyzed for ISONE.

7.1 Peak Hour Forecasting: a Classification Approach

The Peak hour forecasting problem can be formulated as a binary classification problem. The event is the occurrence of daily peak demand. As a result, there are two classes: peak hour and non-peak hour. The Logistics Regression model described in Section 3.3 will be used as the classifier. To implement this model, a new binary variable PK_t is generated that takes the value of 1 if the hour t of a day is peak hour otherwise, it is set to 0. PK_t is the target variable of the model. It is also worth noting that the model is not constrained to find one event per day. Therefore, there could be several hours with a high probability of observing peak load in a day, or a

very low probability for all hours of a day. To address this issue, we would take the hour of the day with maximum predicted probability as the peak hour forecast. This model is hereafter referred to as M2.

7.1.1 Modeling for MeSU

(A) Base Model

Figure 7.1 shows the heatmap of the distribution of peak hours over a year for the *top level* of MeSU. The darker shade indicates hours with higher demand in a day. The peak hours are observed at different times of day in different seasons. In the summer, the peak hours occur between 4 p.m. and 7 p.m. In the winter, some days the peak hour is observed in the morning and the other days it is observed in the late evening. During the spring and fall seasons, the occurrence of daily peak hours is random. The timing of the daily peak load is, thus, seasonal.

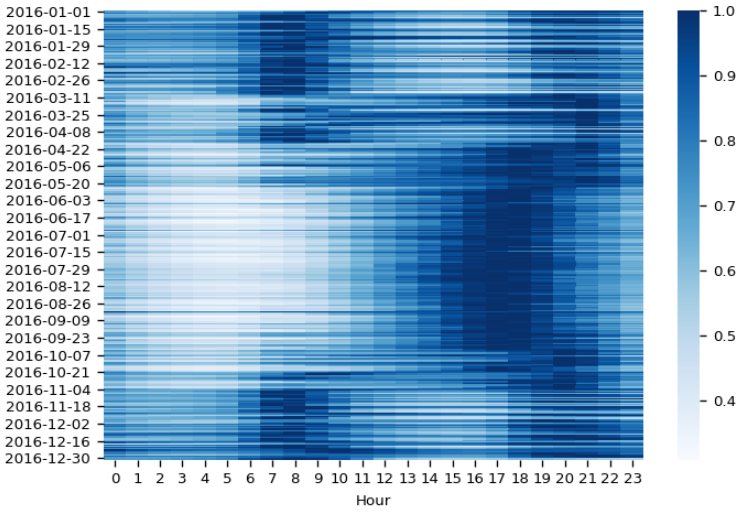


Figure 7.1: Heatmap showing the peak hour distribution over a year, with darker shade showing the high demand hours of a day (*top level* of MeSU, year 2016).

The Vanilla Benchmark model, which can model this seasonality, inspired the baseline model (LR-$T3$) given by equation 7.1. Here, p_t is the probability of an hour to observe peak load, i.e., $p_t = Pr(PK_t = 1)$. The hour of the day is denoted by H_t, representing the Hour of day, the week of the day is represented by D_t, and the calendar

month is represented by M_t. T_t represents the temperature of the hour t.

$$\log\left(\frac{p_t}{1-p_t}\right) = \theta_0 + \theta_1\ D_t + \ \theta_2\ H_t + \theta_3\ H_t D_t + \theta_4\ M_t + \theta_5\ T_t + \theta_6\ T_t M_t + \theta_7\ T_t H_t + \theta_8\ {T_t}^2 + \theta_9\ {T_t}^2 H_t + \theta_{10}\ {T_t}^2 M_t + \theta_{11}\ {T_t}^3 + \theta_{12}\ {T_t}^3 H_t + \theta_{13}\ {T_t}^3 M_t \tag{7.1}$$

It is worth noting that because the target value is a binary variable rather than a continuous variable like *Load*, adding polynomials of temperature may be unnecessary. However, including temperature polynomials helps in modeling the extreme temperature effect. To test this hypothesis, we also tried the model with linear temperature variables (LR-$T1$), and the model with up to second order of temperature polynomials (LR-$T2$), as given by equations 7.2 and 7.3 respectively.

$$\log\left(\frac{p_t}{1-p_t}\right) = \theta_0 + \theta_1\ D_t + \ \theta_2\ H_t + \theta_3\ H_t D_t + \theta_4\ M_t + \theta_5\ T_t + \theta_6\ T_t M_t + \theta_7\ T_t H_t \tag{7.2}$$

$$\log\left(\frac{p_t}{1-p_t}\right) = \theta_0 + \theta_1\ D_t + \ \theta_2\ H_t + \theta_3\ H_t D_t + \theta_4\ M_t + \theta_5\ T_t + \theta_6\ T_t M_t + \theta_7\ T_t H_t + \theta_8\ {T_t}^2 + \theta_9\ {T_t}^2 M_t + \theta_{10}\ {T_t}^2 H_t \tag{7.3}$$

The validation results comparing the three models are given in Table 7.1. LR-$T3$ works better for PLDs that usually are seen on extreme weather days. However, overall, the model LR-$T2$ gives slightly better results. Since PLDs are critical and the overall improvement from LR-$T1$ is slight, we selected LR-$T3$ as our base model. This base model would hereby be called $LR0$.

Table 7.1: M2 base model validation results, MeSU.

	All Days			PLDs		
	LR-$T1$	LR-$T2$	LR-$T3$	LR-$T1$	LR-$T2$	LR-$T3$
Top Level	340	**301**	343	37	27.6	**22.4**
SA1	391	**358**	367	29.2	**29.6**	30
SA2	339	**332**	339	20.2	**18.4**	**18.4**
SA3	**403**	420	434	36	35.8	**30.8**
Total	1473	**1411**	1483	122.4	111.4	**101.6**

(B) Training on Probable Peak hours

The analysis of the hours that are most likely to observe the peak load of the day is shown in Figure 7.2. These hours are analyzed for all the zones of MeSU for 2016. The hours that observe greater than 1% of peak hours in the year are marked as the probable peak hours U1, denoted by Equation 7.4. Similarly, the hours that observe greater than 2% of peak hours in the year are marked U2, denoted by Equation 7.5.

$$U1 = \{H_t : H_t \ \in \ [7, 8, 9, 10, 15, 16, 17, 18, 19, 20, 21, 22]\} \tag{7.4}$$

$$U2 = \{H_t : H_t \ \in \ [7, 8, 9, 16, 17, 18, 19, 20, 21]\} \tag{7.5}$$

Figure 7.3 investigates the peak hours for just PLDs for three years in *top level*. We observe a similar seasonal pattern of peak hours every year. This indicates that we may confine training the model only on the probable peak hours, which might enhance the training process by removing redundant data. Table 7.2 shows the results of the model trained only on hours in U1 and U2, which are called $LR0_{U1}$ and $LR0_{U2}$, respectively. When compared with results from LR0, $LR0_{U1}$ brings significant improvement overall as well as on PLDs. Hence, we choose the model $LR0_{U1}$.

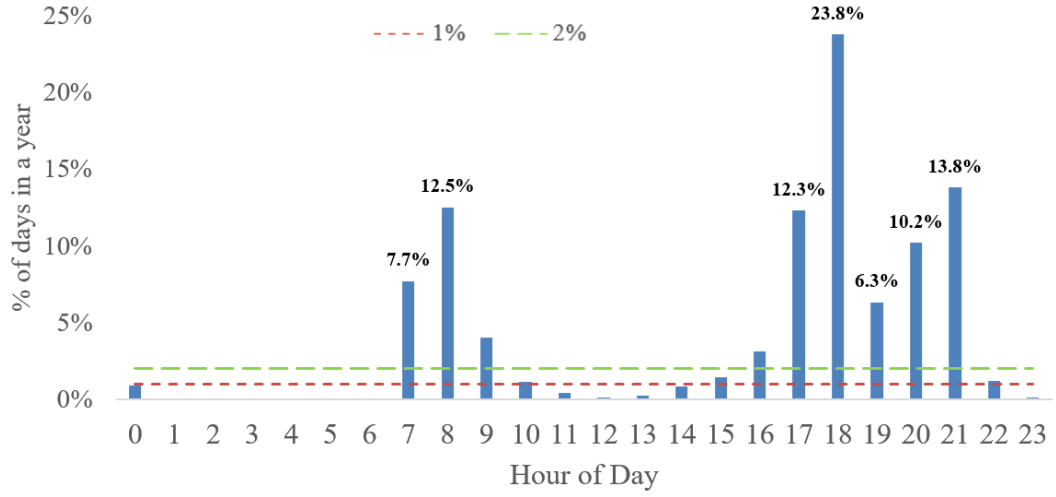


Figure 7.2: Probable peak hours of *top level* of MeSU (year 2016)

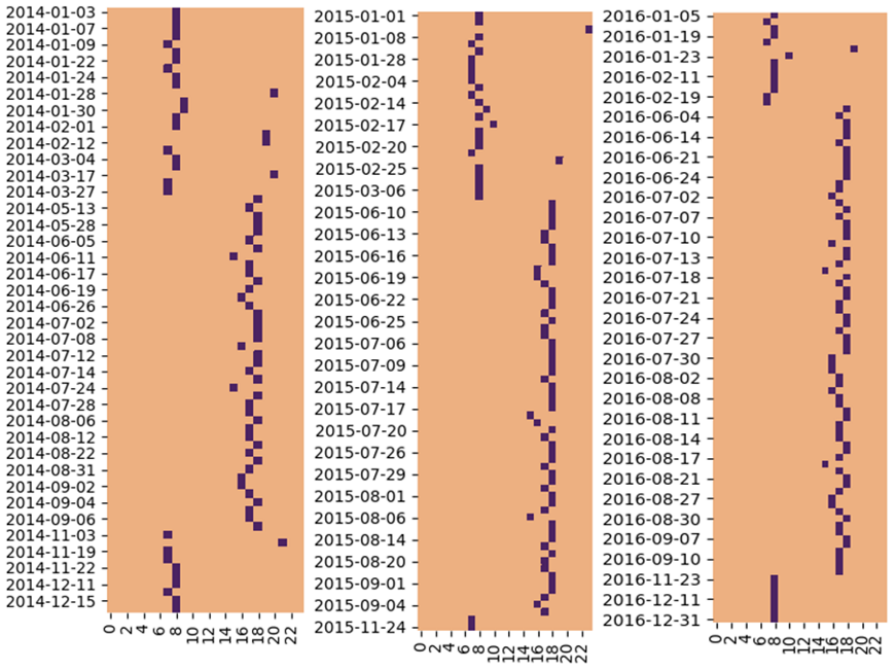


Figure 7.3: Heatmap showing the peak hours for PLDS for the three years of *top level* of MeSU.

Table 7.2: M2 validation results based on models trained on probable peak hours only, MeSU.

	All days			PLDs		
Zone	$LR0$	$LR0_{U1}$	$LR0_{U2}$	$LR0$	$LR0_{U1}$	$LR0_{U2}$
Top Level	343	313	328	22.4	22.4	22.4
SA1	367	343	350	30	20	19.6
SA2	339	327	329	18.4	18.4	18.4
SA3	434	428	438	30.8	30.8	30.8
TOTAL	1483	1411	1445	101.6	91.6	91.2

(C) Add-on: Daily temperature features

We also explored adding low-resolution temperature features. The effects of daily average temperature (Ta), daily maximum temperature (Tx), and daily minimum temperature (Tn) were investigated by adding the following three functions to the baseline model $LR0_{U1}$:

(a) $g\,(Ta)$

(b) $g\,(Tx)$

(c) $g\,(Tx) + \ g\,(Tn)$

where,

$$g\,(T_t) = \ \theta_{14}\ T_t + \theta_{15}\ T_t M_t + \theta_{16}\ T_t H_t \tag{7.6}$$

The validation year results for the three models developed with daily temperature features are shown in Table 7.3. Clearly, $LR0_{U1}$+g(Ta) gives the best results of the three and also improves $LR0_{U1}$ results. We will this model as $LR1$ hereon.

Table 7.3: M2 validation results with daily temperature features added, MeSU .

	All days					PLDs				
	Top	SA1	SA2	SA3	Total	Top	SA1	SA2	SA3	Total
$LR0_{U1}$	313	343	327	428	1411	22.4	20	18.4	30.8	91.6
$LR0_{U1}$+(a)	**280**	**300**	**251**	**394**	**1224**	22	**15.2**	18.2	**26**	**81**
$LR0_{U1}$+(b)	320	367	298	466	1451	18.6	19.2	**15.4**	49.4	103
$LR0_{U1}$+(c)	293	355	305	409	1362	**20.8**	28.2	15.8	42.4	107

(D) Add-on: Timing of Daily Maximum & Minimum Temperature

We also investigated if the timing of highs and lows of the diurnal temperature curve are the driving factors in determining the peak hour of a day. Figure 7.4 (a) plots the hour of daily maximum temperature vs. peak hours for summer months (Ta>75). The hours with daily maximum temperature and peak hours are not co-incident. However, on most days, the peak hour occurs after the occurrence of the day's maximum temperature. Hence, the effect of the timing of daily maximum temperature cannot be overlooked.

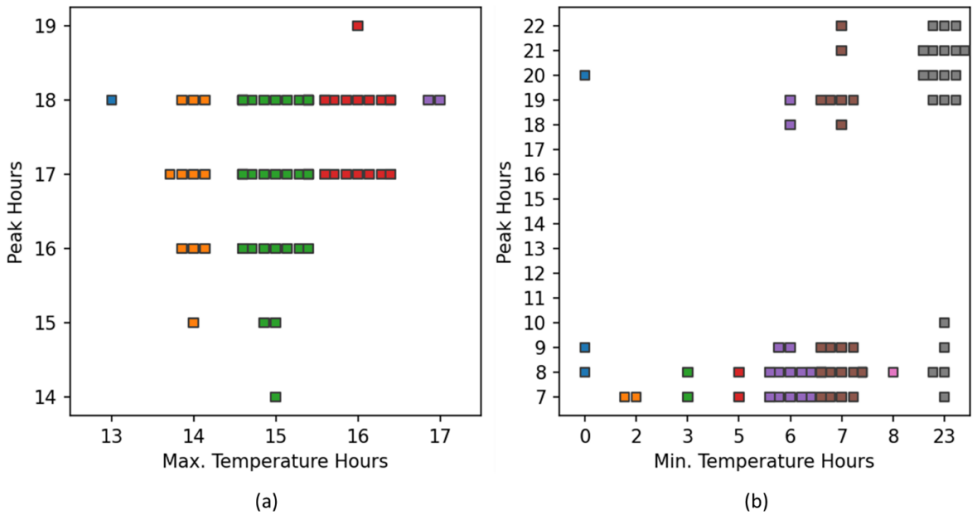


Figure 7.4: Scatterplots of a) timing of maximum temperature of day vs. peak hours in summer, b) timing of minimum temperature vs. peak hours in winter (*top level* of MeSU, year 2016)

Figure 7.4(b) plots the hours of daily minimum temperature vs. peak hours for winter months ($Ta < 50$). Again, the hours with daily minimum temperature and peak hours are not coincident. The peak hour occurrence in winter is a combined effect of increased human activities with a drop in temperature during dusk and dawn. Peak hours do not necessarily follow the minimum temperature hours. Hence, the effect of the timing of the day's minimum temperature cannot be substantiated.

To model the timing of daily maximum temperature hours, we defined a binary variable, MX_t. The value of 1 denotes that the hour observes the maximum temperature of the day and the value of 0 denotes vice-versa. We then tried the following three add-ons. If MX_{t-1} denote the lagged variable of MX_t by one hour, MX_{t-2} denote the lagged variable of MX_t by two hours, then the three candidate models are denoted in 7.7.

$$\begin{aligned} LR_{H0} &= LR1 + \ \theta_{17}\ MX_t \\ LR_{H1} &= LR_{H0} + \ \theta_{18}\ MX_{t-1} \\ LR_{H2} &= LR_{H1} + \ \theta_{19}\ MX_{t-2} \end{aligned} \tag{7.7}$$

Table 7.4: M2 validation results with the indicator of daily maximum temperature hour added, MeSU.

	Sum of wDE values from all 4 zones for all days in the year 2016	Sum of wDE values from all 4 zones for PLDs in the year 2016
$LR1$	1218	81
LR_{H_t0}	**1216**	**78.8**
LR_{H_t1}	1214	87
LR_{H_t2}	1262	84

Table 7.4 displays the validation results for MeSU for the three models. The model LR_{H0} improves the $LR1$ results slightly for PLDs as well as for all days. Hence, the final M2 model selected for MeSU is presented in the equation 7.8.

$$\begin{aligned}\log\left(\frac{p_t}{1-p_t}\right) =& \theta_0 + \theta_1\ D_t + \ \theta_2\ H_t + \theta_3\ H_t D_t + \theta_4\ M_t + \theta_5\ T_t + \theta_6\ T_t M_t + \theta_7\ T_t H_t \\ &+ \theta_8\ {T_t}^2 + \theta_9\ {T_t}^2 H_t + \theta_{10}\ {T_t}^2 M_t + \theta_{11}\ {T_t}^3 + \theta_{12}\ {T_t}^3 H_t + \theta_{13}\ {T_t}^3 M_t \\ &+ g\left(Ta_t\right) + \theta_{17}\ MX_t\end{aligned} \tag{7.8}$$

7.1.2 Modeling for ISONE

For ISONE, the daily average temperatures in the region are unlikely to go beyond 80°F as shown in Figure 7.5. The peak hours occur mostly between 5 p.m. and 8 p.m. Even though the region sees frigid temperatures, morning peaks are very rare. This is because the primary fuel of space heating in ISONE is non-electric. The recent heating electrification share is only 12-13% [95]. Although the timing of peak demand seems to be less dependent on the temperature conditions, on really hot and cold days, it occurs primarily at 5 p.m. We followed a similar step-by-step validation approach as for MeSU.

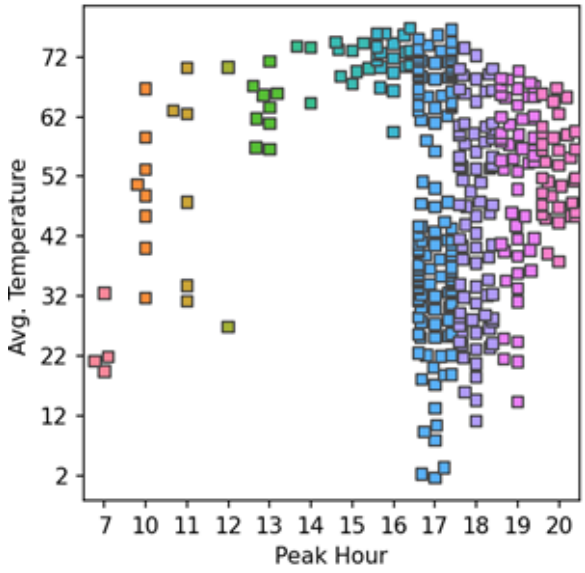


Figure 7.5: Scatterplot of daily average temperature vs. peak hour for ISONE (year 2017).

(A) Base Model

The three base models; *LR-T*1, *LR-T*2, and *LR-T*3 were evaluated. For PLDs, *LR-T*2 gives the best results of the three, while for all days *LR-T*3 performs best as shown in Table 7.6. Since the accuracy of PLDs is more important, we chose LR-T2 as the base model, hereon called LR0.

Table 7.5: M2 base model validation results, ISONE.

	All Days			PLDs		
Region	*LR-T*1	LR-T2	LR-T3	*LR-T*1	*LR-T*2	*LR-T*3
CT	290	**252**	259	10	**9.6**	**9.6**
ISONE	266	259	**239**	7	**6.8**	6.4
ME	499	481	**470**	**13.8**	**13.8**	14
NEMASSBOST	348	355	**343**	7.4	**6.4**	8.4
NH	311	**270**	271	9.8	**8.4**	8.4
RI	363	345	**307**	38.6	**28.2**	31
SEMASS	214	**168**	172	**8.6**	9.2	12.4
VT	**246**	256	261	**14.6**	16	16.4
WCMASS	287	**248**	253	7.8	**6.6**	8.8
Total	2824	2634	**2575**	117.6	**105**	115.4

(B) Training on Probable Peak Hours

We also analyzed the probable peak hours for all the zones of ISONE for 2017 Figure 7.6. The probable peak hours that observe greater than 1% of peak hours in a year are marked as a set, referred to as N1, and denoted by equation 7.9. Similarly, Set N2 for hours with greater than 2% of peak hours in the year is denoted by equation 7.10.

$$N1 = \{H_t : \ H_t \in \ [7, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]\} \tag{7.9}$$

$$N2 = \{H_t : \ H_t \in \ [10, 11, 13, 16, 17, 18, 19, 20]\} \tag{7.10}$$

Models $LR0_{N1}$ and $LR0_{N2}$ are created by training the LR0 on hours from N1 and N2, respectively . The validation results for both $LR0_{N1}$ and $LR0_{N2}$ are not compelling as seen in Table 7.6. Therefore, we chose to train base model LR0 using all hours for training.

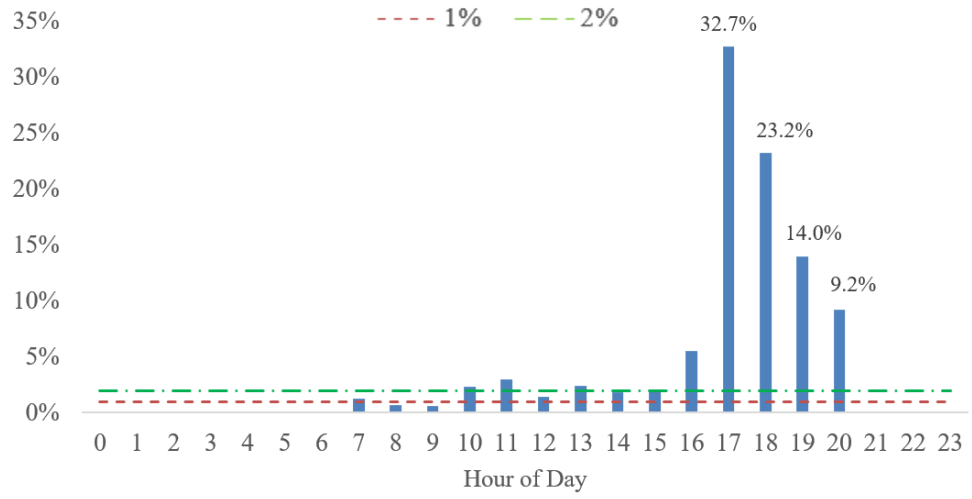


Figure 7.6: Probable peak hours of ISONE (year 2017).

Table 7.6: M2 validation results based on models trained on probable peak hours only, ISONE.

	All days			PLDs		
Zone	$LR0$	$LR0_{N1}$	$LR0_{N2}$	$LR0$	$LR0_{N1}$	$LR0_{N2}$
CT	252	250	254	9.6	9.6	9.6
ISONE	259	260	251	6.8	6.8	6.8
ME	481	485	490	13.8	13.8	13.8
NEMASS	355	355	341	6.4	6.4	6.4
NH	270	267	272	8.4	8.4	7.6
RI	345	355	346	28.2	28.4	31.4
SEMASS	168	167	168	9.2	9.2	9.2
VT	256	254	265	16	16	16
WCMASS	248	255	258	6.6	6.6	6.6
TOTAL	2634	2650	2646	105	105.2	107.4

(C) Add-on: Daily Temperature Features

The effect of daily average temperature (Ta), daily maximum temperature (Tx), and daily minimum temperature (Tn) were investigated. Table 7.7 presents the sum of wDE values from all nine zones of ISONE from the three candidate models for the validation period. As in MeSU, adding the term $g\,(Ta)$ gives the best improvement over the base model.

Table 7.7: M2 validation results with daily temperature features added, ISONE.

	All days	PLDs
$LR0$	2634	105
$LR0 + g(Ta)$	2523	**100.4**
$LR0 + g(Tx)$	**2520**	128
$LR0 + g(Tn)$	2576	129.4

(D) Add-on: Timing of Daily Maximum & Minimum Temperature

We also investigated the effect of the timing of daily maximum and minimum temperature on the timing of peak hour in Figure 7.7. Evidently, the effect of timing of daily minimum temperature is non-determinant as the heating load is majorly non-electric. Also, there is no strong influence of the timing of the daily maximum temperature on the daily peak hour. It seems that human activities are the primary drivers of peak timing.

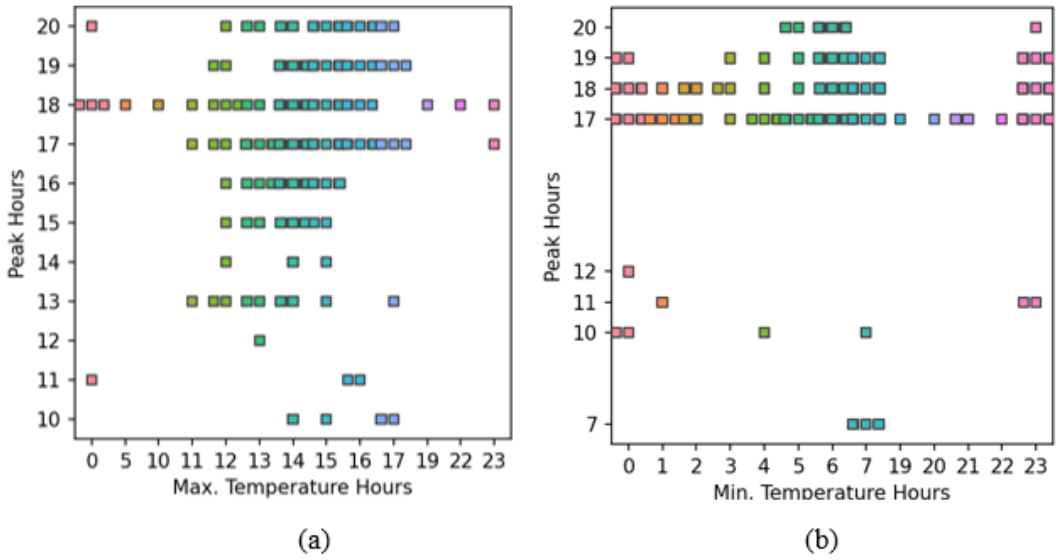


Figure 7.7: Scatterplots of a) timing of maximum temperature of day vs. peak hours in summer, b) timing of minimum temperature vs. peak hours in winters for ISONE (year 2017).

Therefore, the final M2 model selected for ISONE is given by the equation 7.11.

$$\begin{aligned}\log\left(\frac{p_t}{1-p_t}\right) =& \theta_0 + \theta_1\ D_t + \ \theta_2\ H_t + \theta_3\ H_t D_t + \theta_4\ M_t + \\ & \theta_5\ {T_t}^2 + \theta_6\ {T_t}^2 M_t + \theta_7\ {T_t}^2 H_t + g\left(Ta_t\right)\end{aligned} \tag{7.11}$$

7.2 A Two-Stage Peak Hour Forecasting Model

The second proposed modeling framework is a two-stage modeling process, built upon an hourly load forecasting model with a secondary post-processing stage. The first stage (Stage 1) employs an hourly load forecasting model that outputs the forecasts of the daily load profile, or what we call daily shape forecasts. The second stage (Stage 2) is a post-processing phase where the model learns the relationship between the daily shape forecasts from Stage 1 and the actual daily peak hour to give the final peak hour forecast. This model is hereafter referred to as M3.

In Stage 1, a state-of-the-art hourly load forecasting model is used to get the hourly load forecasts. In this research, we will use the M0 model for the first stage. We also need to generate the in-sample hourly load forecast, that is, the load forecasts for the training period. From the hourly load forecast, we derive the shape forecast for the training and test periods. The shape forecast $\hat{S}_t$ is given in the equation 7.12 :

$$\widehat{S}_t = \frac{\hat{L}_t}{\widehat{P}_t} \tag{7.12}$$

, where $\widehat{P}_t$ is the daily peak load forecast and $\hat{L}_t$ is the hourly load forecast. In Stage 2, a classification model is developed to forecast the daily peak hour using the shape forecast, calendar variables, and daily temperature variables as predictors. The process flow is demonstrated in Figure 7.8.

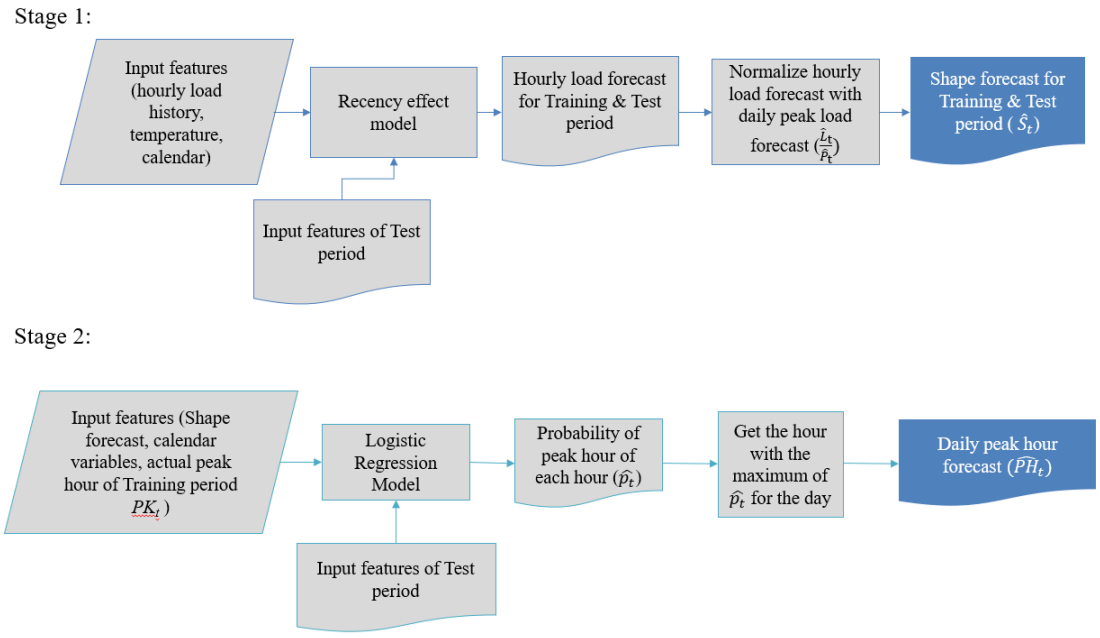


Figure 7.8: Framework of the model M3.

Stage 2 is a post-processing stage where the model learns the pattern of shape forecast and its relationship with the calendar variables to get the target peak hour of a day. Since the shape forecast value ranges between 0 to 1 for every day, it is easier for the classifier to learn the chronology of daily peak timing.

We have used Logistic Regression as the classifier for Stage 2, hereafter called TS. Our experiment will use model M0 to get the shape forecast from Stage 1. Similar to the implementation approach of M2, a new binary variable PK_t is generated where $PK_t \in 0{,}1$. PK_t is the target variable of TS. The hour of the day with the maximum predicted probability from TS will be taken as the peak hour forecast.

It is easy to see that the shape forecast and hour of day are major determinants in the TS model. In addition, since the effect of shape forecast is not independent of the effect of the hour of day, we will add their interaction term also. Since the pattern of peak hours varies every month, we also add the term. $H_t M_t$ to model the effect of the hour that is not independent of the effect of the month. Hence, the formulation of TS is given by the equation below :

$$\log\left(\frac{p_t}{1-p_t}\right) = \alpha_0 + \alpha_1\ H_t\ + \alpha_2\ M_t\ + \alpha_3\ \hat{S}_t\ \ + \alpha_4 H_t\ \hat{S}_t + \alpha_5 H_t\ M_t \qquad (7.13)$$

We further explored training the model only for probable peak hours, as discussed in section 7.1. The two variants of the model TS, obtained by training the model on U1 and U2 (denoted by equations 7.4 and 7.5) are referred to as $TS1$ and $TS2$. The validation results for the MeSU are shown in Table 7.8.

Table 7.8: M3 validation results, MeSU.

	wDE for all Days				wDE for PLDs			
Zone	M0	TS	$TS1$	$TS2$	M0	TS	$TS1$	$TS2$
Top Level	268	273	254	**246**	59	52.6	51.4	**39.4**
SA1	300	319	308	**280**	47.6	52.8	52.8	**39.2**
SA2	343	329	320	**316**	41.6	25.6	18.2	**16.6**
SA3	355	343	333	**289**	51	44.8	38.6	**24**
Total	1267	1263	1215	**1131**	199	175.8	161	**119**

The Model $TS2$ clearly performs better and gives a unanimous improvement over the Stage 1 wDE values. Hence, $TS2$ is chosen for Stage 2 for MeSU. The selected model for the proposed methodology is hereby called M3.

A similar model selection approach is adopted for ISONE. $TS1$ and $TS2$ are the models trained on N1 and N2, respectively. The validation results from TS, $TS1$, and $TS2$ are shown in Table 7.9. The model $TS1$ improves M0 results for PLDs as well as for all days in a year. $TS2$ slightly gives better for PLDs, but worse for all days when compared with $TS1$. Hence, we chose $TS1$ for Stage 2 of ISONE as the final M3 model.

Table 7.9: M3 validation results, ISONE.

	wDE for all Days				*wDE for PLDs*			
Zone	M0	*TS*	TS1	TS2	M0	*TS*	TS1	TS2
ISONE	236.2	203	190	189	33.6	25.4	25.4	19.4
VT	231.8	219	199	209	15.2	12.4	12.4	12.4
ME	439.8	288	295	302	34.6	14.6	14.4	14.4
CT	248	220	215	221	28.6	16.2	16.2	14
RI	313.8	264	266	274	36.4	37.2	36	45.4
NH	303.8	235	235	238	24.6	19.2	19.2	20
SEMASS	131	119	114	113	9.4	10.2	10.2	9.2
WCMASS	223	194	192	201	20.2	11.8	11.8	7.4
NEMASS	370.6	281	282	287	29	6.4	6.4	6.4
TOTAL	2498	2023	**1988**	2034	232	153	152	**149**

7.3 Results and Discussion

The test results from the two proposed peak timing forecasting models as discussed in the preceding sections of this chapter are presented in this section. The M0 model is the benchmark model in this research. The model M2 is the classification peak hour forecasting model described in section 7.1. The M3 model is the two-stage model described in section 7.2. All three models are evaluated for all days of the test years as well as explicitly for PLDs for the two case studies.

Table 7.10: wDE values for ex-ante forecasts on test data of MeSU.

	All Days			
Year	Zone	M0	M2	M3
2017	Top Level	222	286	**207**
	SA1	251	307	**200**
	SA2	256	249	**234**
	SA3	336	300	**283**
2018	Top	207	233	**165**
	SA1	302	251	**249**
	SA2	248	237	**209**
	SA3	366	**254**	279
2019	Top Level	214	238	**189**
	SA1	302	258	**254**
	SA2	358	352	**344**
	SA3	332	329	**230**
Total		3394	3296	**2845**

	PLDs only			
Year	Zone	M0	M2	M3
2017	Top Level	33	**23.2**	23.6
	SA1	45.4	**18**	22.4
	SA2	37	17.6	**17**
	SA3	51.8	24.8	**26**
2018	Top	39.6	**23**	27.4
	SA1	34.8	**20.2**	**23**
	SA2	42.8	**12.6**	21.4
	SA3	54	**10**	28
2019	Top Level	38.6	**12.2**	21.4
	SA1	50.8	**28.4**	33.4
	SA2	47	23.2	**22**
	SA3	54	25.4	**20**
Total		529	**239**	286

Table 7.10 lists the wDE values obtained from the three models for the four zones of MeSU. M3 significantly improves peak hour forecasts over M0 for all days of the test year and PLDs. M3 gives an improvement of 16% on all days of the year and 46% on PLDs. However, overall, the wDE values of M2 are trivially better than those of M0. But on PLDs, M2 improves the forecast by 55%, outperforming M3 also. Furthermore, both M2 and M3 show improvement in all zones for all three test years for PLDs.

Table 7.11: wDE values for ex-ante forecasts on test data of ISONE.

	Test results : All days					
	2018			2019		
Zone	M0	M2	M3	M0	M2	M3
CT	244	266	**213**	191	201	**166**
ISONE	188	221	**165**	142	154	**121**
ME	**383**	429	419	399	335	**314**
NEMASS	285	294	**265**	252	268	**231**
NH	297	271	**265**	**194**	259	196
RI	333	320	**287**	234	22	**164**
SEMASS	140	157	**113**	86	125	**83**
VT	211	191	**172**	240	234	**196**
WCMASS	254	269	**204**	190	207	**164**
Total	2336	2419	**2103**	1930	2008	**1635**

Test Results : PLDs

	2018			2019		
Zone	M0	M2	M3	M0	M2	M3
CT	17.4	12.6	**8.4**	28	**9.6**	14
ISONE	13.2	12.4	**5**	7.2	13.6	**3.8**
ME	**27.2**	30.8	40	**24.8**	**23.6**	25
NEMASS	27.6	**18**	**18**	13.8	**9.8**	12.6
NH	19.4	**11.4**	13.2	20.4	17.8	**12.4**
RI	26.2	**10.2**	26	35.8	**16.8**	21.8
SEMASS	**17.4**	20.2	25	9.8	9.2	9.4
VT	35.2	**17.6**	30	19.2	17.8	**12.2**
WCMASS	31.4	16.4	**14.4**	**17.6**	24.8	18.4
Total	215	**150**	180	177	143	**130**

Table 7.11 shows the wDE values for ISONE test years from the two proposed models and M0. M3 outperforms M0 on all days of the year, giving an overall improvement of 12% on both test years. M3 also improves M0 peak timing forecasts for PLDs by 21%. Although M2 apparently does not give any improvement over M0 for all days, it performs best for PLDs, giving an overall improvement of 25% over M0 for the two test years.

Thus, it is evident that M2 is able to forecast the peak timing of PLDs better than M0 and even M3. However, due to worse performance on non-PLDs, the overall performance of M2 is worse than M0. We compared the temperature profiles of PLDs and non-PLDs to figure out why M2 is effective on PLDs but not on all days of the year. Figure 7.9 plots the maximum and minimum temperatures of day vs. peak hour for the *top level*. For PLDs, we notice that when the minimum temperature of day falls below 40°F, there is a high probability of a morning peak at 7 a.m. or 8 a.m. Additionally, for these days if the day's maximum temperature is below 40°F, we observe a late evening peak from 7 p.m. to 9 p.m. During the summer, the system mostly peaks between 4 and 6 p.m. However, for non-PLDs, it is not easy to draw a definite pattern. On cold days, the system may peak in the evening even when the temperature falls below 40°F. Besides, the peak hour distribution in summer spreads wider between 3 to 9 p.m. Therefore, without hourly load information, M2 is not able to reliably predict the peak hour for non-PLDs.

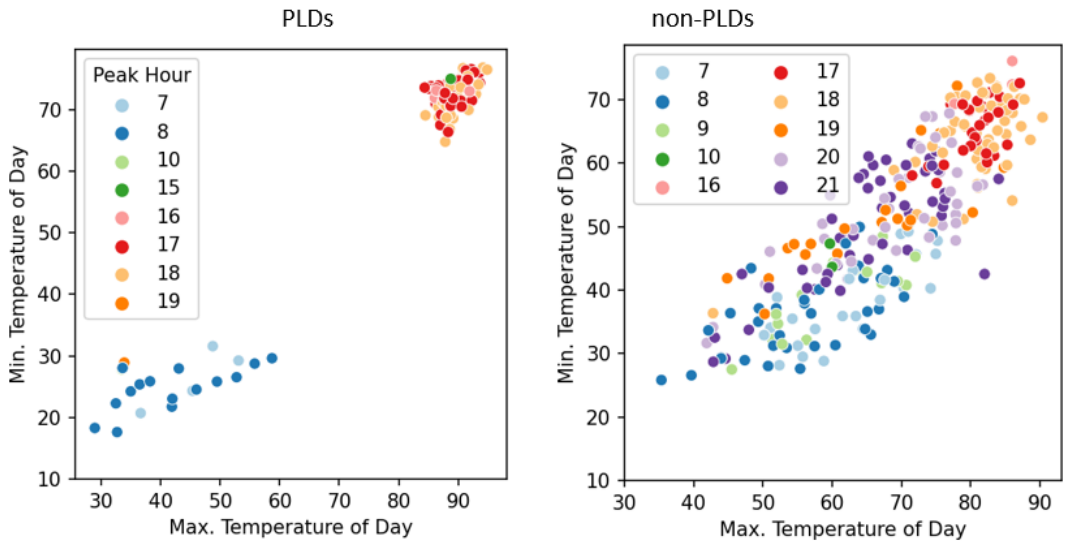


Figure 7.9: Scatterplots of daily maximum temperature vs. daily minimum temperature grouped by peak hours for (a) PLDs and (b) non-PLDs (*top level* of MeSU, year 2016).

It is important to understand how the peak timing forecasts from M2 and M3 differ from M0 forecasts. Table 7.12 displays the distribution of displacement error (*de*) from the three models for PLDs. For each case study, the values shown are the percentage days of PLDs from all zones with the respective *de*. For instance, with M0 model, 21% of PLDs from all 4 zones in the three test years of MeSU had zero *de*.

Table 7.12: Comparing the distribution of *de* on PLDs from the three models.

MeSU PLDs				ISONE PLDs			
	M0	**M2**	**M3**		**M0**	**M2**	**M3**
$de = 0$	21%	57%	39%	$de = 0$	42%	53%	53%
$de = 1$	39%	32%	41%	$de = 1$	34%	31%	29%
$de = 2$	30%	6%	17%	$de = 2$	13%	7%	9%
$de = 3$	9%	1%	1%	$de = 3$	7%	4%	4%
$de = 4$	1%	1%	1%	$de = 4$	2%	2%	3%
$de \geq 5$	1%	3%	0%	$de \geq 5$	2%	2%	2%

M2 and M3 both bring drastic improvement in peak timing forecasts by increasing the matched peak hours and reducing the number of days with displacement in peak hour forecasts. For example, the days with zero displacement error for MeSU increased from 21% with M0 to 57% with M2 and 39% with the M3 model. The percentage of days with $de \geq 3$, lowers to 5% with M2 and 2% with M3. A similar trend is evident in ISONE, where, with M2 and M3, the count of days with matched peak hours increases to 53%. The count of days with $1 \leq de \leq 3$ reduces to 42% from 54% of M0. The count of days with $de \geq 4$ from M2 and M3 does not show any improvement. It should be noted that displacement errors of up to three hours are decisive in jeopardizing DR implementation. Hence, this is a significant improvement.

We also compared the ex-post forecasting results of the models M2 and M3 with M0. Table 7.13 (Top) displays the sum of wDE values for all test years across all zones for MeSU for ex-post and ex-ante forecasting. For ex-post forecasting, neither M2 nor M3 outperformed M0. Hence, with the actual temperature for the forecast period, M0 forecasts are quite robust. However, M0 forecast accuracy deteriorates significantly from ex-post to ex-ante forecasting. On the other hand, ex-ante forecasts for M2 remain quite close to M2 ex-post forecasts. For example, wDE values from

Table 7.13: Sum of wDE values from ex-post forecasting, and sum of wDE values from ex-ante peak hour forecasting for the test data of the two case studies.

Sum of wDE values for all test years of all zones - MeSU

	All Days			PLD		
	M0	M2	M3	M0	M2	M3
Ex-post	1862	2808	2008	158	244	161
Ex-ante	3394	3296	2845	529	239	286

Sum of wDE values for all test years of all zones - ISONE

	All Days			PLD		
	M0	M2	M3	M0	M2	M3
Ex-post	4189	4950	3590	320	286	263
Ex-ante	4263	4427	3739	392	293	310

M0 increased by 82% and 235% for all days and PLDs respectively. While for M2 the values increased by only 17% and -2% for all days and PLDs respectively. For M3 the wDE value increase is 36% and 79% respectively. This suggests that M2 forecasts are less sensitive to temperature forecast errors than M0 and M3.

Table 7.13 (Bottom) displays results for ISONE. M3 provides better results than M0 even for ex-post forecasting, while M2 forecasts are worse for ex-post. When we compare the ex-ante peak timing forecasts with the ex-post forecasts for ISONE, we find that the difference in wDE values from the same model is not as large as it is for MeSU. For example, the wDE values from M0 for all days increased by only 2% with ex-ante forecasting. The influence of temperature forecast error on peak time forecast error appears to be less for ISONE. This is because, unlike MeSU, ISONE is a colder region with temperatures rarely rising above 90°F, as seen in Figure 7.10. Also, it is predominantly a non-electric heating region; hence winter peaks are also not much affected by temperature. Therefore, the primary determinants of the occurrence of daily peak hours are related to human activities for ISONE.

From the test results of both case studies, we gather that M2 is a simple yet

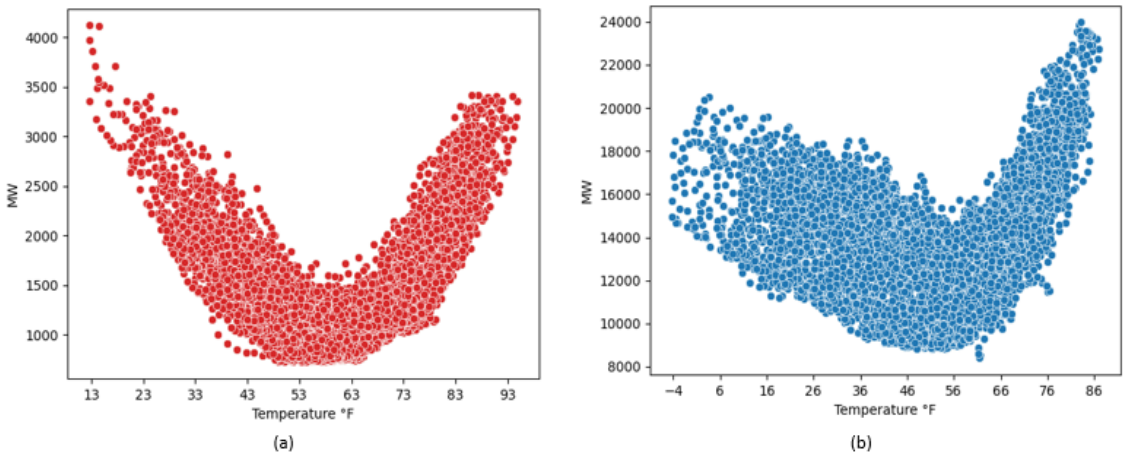


Figure 7.10: Scatterplot of load vs. temperature for a) MeSU, b) ISONE (year 2016).

robust model for ex-ante forecasting of peak hour for PLDs. It outperformed the M0 as well as the M3 for PLDs and is least sensitive to temperature forecast errors. Notably, M2 does not use any lagged load variable. Hence, M2 could be a valuable approach for short to medium-term ex-ante forecasting of peak hours. Model M3 also outperforms M0 in ex-ante forecasting for PLDs, and additionally, it improves the peak timing forecast for all days of the year. Empirical evidence demonstrates the strong performance of the M3 for day-ahead peak hour forecasting.

CHAPTER 8: CONCLUSION

Peak load forecasting is crucial in the power industry. With the imperative of consistently meeting demand, generators dispatched for peak demand typically have a very high marginal price. Most of the prior studies on peak forecasting are focused on the magnitude of peak load. As the industry gears up for more efficient and cleaner ways, forecasting the peak timing is increasingly becoming critical. Various peak shaving strategies necessitate the prior knowledge of 'when'. With the integration of DER and ESS, optimization of the load scheduling is highly reliant on the timing and shape of peak demand. And yet, there is a huge gap in the academic literature on the subject of forecasting and measuring the timing and shape of peak demand.

This dissertation aims to bridge this gap by focusing on a) defining the taxonomy of peak timing problems, b) proposing error measures to evaluate peak timing forecasts effectively, c) evaluating the peak hour forecast from traditional hourly load forecasting models, d) proposing novel methods of short-term peak hour forecasting and e) ex-ante forecast analysis.

We have defined the taxonomy of peak timing problems into four types specific to the business application: Peak Load Day (PLD), Peak hour, on-peak period, and Peak Shape. We have critically examined the error measures used in existing literature on peak timing forecasting. We also explored traditional classification error measures employed in practice and common metrics used for event detection problems in other domains. Based on the study, we have proposed five easily implementable error measures for the four classes of peak timing forecasting problems.

In the aspect of the model development for peak timing forecasting, this research has focused on the short-term peak hour forecasting problem. In practice, daily peak

hour forecasts are obtained from the models developed for hourly load forecasting. However, there is no study to evaluate how good these forecasts are. This research assessed the state-of-art hourly load forecasting framework based on the Recency effect model [79]. It was also set as the benchmark (M0) in this study. Peak hour forecasts from M0 were evaluated using the proposed error measure to understand the model's performance to forecast the peak hour. For one of the case studies, the ex-post forecast analysis revealed that the peak hour forecasts from the M0 lie within one hour of the actual peak hour for 88% of days in a year; however, with ex-ante forecast analysis, the accuracy comes down to 70%.

We also explored the model selection for peak hour forecasting using the Recency effect forecasting framework. Instead of the traditional method of selecting a model on MAPE, several discrete and continuous error measures related to peak load, peak timing, and peak shape were used. With extensive experiments, the research concluded that with hourly load forecasting models, choosing a model by measuring the errors of the entire profile is a more robust method than picking certain hours for evaluation. However, the study also revealed that selecting a model based on ex-ante forecast analysis of the validation period can be beneficial for peak hour forecasting.

This dissertation also explored developing direct forecasting models for the problem of peak hour forecasting. Two models are proposed. Model M2, based on the classification approach, explicitly forecasts the daily peak hour. The model is very simple in complexity and computation. The model is interpretable because it was developed by understanding the factors contributing to peak occurrences. Compared to M0, the model demonstrated exceptional performance on PLDs, giving an improvement of 25-55% on peak hour forecasting accuracy. It also proved resilient to temperature forecast error, making it ideal for ex-ante forecasting on short to medium-term forecasting horizons.

Model M3 is a two-stage forecasting framework that utilizes the hourly load fore-

casts from M0 in the first stage. For the second stage, the shape forecasts are derived from the first-stage forecasts and passed to the second-stage classification model to get the final peak hour forecast. The model empirically demonstrated to improve the M0 results by 12-16% for all days in a year and 21-46% for only PLDs. This approach can be highly instrumental in optimizing load scheduling and market hedging techniques by enhancing the day-ahead forecast of peak hours.

Post-analysis of results revealed that each data set is unique due to its geographic characteristics; hence, peak timing patterns and their determinants differ. Understanding these drivers is critical for good modeling practice. It is worth highlighting that a model, like M2, exhibiting lower overall accuracy in peak hour forecasting, may paradoxically excel in the forecast of just PLDs. In conclusion, peak timing problems require specialized models and evaluation metrics. By understanding the strengths and applications of M2 and M3, researchers and practitioners can make informed choices to address diverse challenges of peak timing forecasting.

This detailed research on the topic of peak timing forecasting opens up exciting avenues for future research. As part of model development, we have only taken on the problem of peak-hour forecasting. The study can be expanded to include other peak timing problems, such as month-ahead forecasting of PLDs, day-ahead on-peak period, and peak shape forecasting. In this research, we have evaluated the peak hour forecasting models on the actual PLDs. It is intriguing to see how the model selected on forecasted PLDs performs. The research can also be extended to probabilistic forecasting for quantifying uncertainty for long-term forecasting purposes. Hence, there are numerous open questions on this subject. Hopefully, the proposed framework will serve as a foundation for developing more sophisticated models and addressing emerging real-world grid operation challenges.

REFERENCES

[1] N. O'Connell, P. Pinson, H. Madsen, and M. O'Malley, "Benefits and challenges of electrical demand response: A critical review," *Renewable & Sustainable Energy Reviews*, vol. 39, pp. 686–699, Nov 2014.

[2] R. J. Hyndman and S. Fan, "Density forecasting for long-term peak electricity demand," *IEEE Transactions On Power Systems*, vol. 25, pp. 1142–1153, May 2010.

[3] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 376–388, 2020.

[4] T. Hong, "Crystal ball lessons in predictive analytics," 2015. EnergyBiz Mag, vol. 12, no. 2, pp. 35-37.

[5] T. Hong, J. Wilson, and J. Xie, "Long term probabilistic load forecasting and normalization with hourly information," *IEEE Transactions on Smart Grid*, vol. 5, pp. 456–462, Jan 2014.

[6] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016.

[7] J. Carrasco, L. Franquelo, J. Bialasiewicz, E. Galvan, R. PortilloGuisado, M. Prats, J. Leon, and N. Moreno-Alfonso, "Power-electronic systems for the grid integration of renewable energy sources: A survey," *IEEE Transactions on Industrial Electronics*, vol. 53, no. 4, pp. 1002–1016, 2006.

[8] P. R. White, J. D. Rhodes, E. J. H. Wilson, and M. E. Webber, "Quantifying the impact of residential space heating electrification on the texas electric grid," *Applied Energy*, vol. 298, 2021.

[9] M. Muratori, "Impact of uncoordinated plug-in electric vehicle charging on residential power demand," *Nature Energy*, vol. 3, no. 3, pp. 193–201, 2018.

[10] J. Twitchell, "A review of state-level policies on electrical energy storage," *Current Sustainable/Renewable Energy Reports*, vol. 6, no. 2, pp. 35–41, 2019.

[11] W. Kempton and J. Tomic, "Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy," *Journal of Power Sources*, vol. 144, no. 1, pp. 280–294, 2005.

[12] C. W. Gellings, "The concept of demand-side management for electric utilities," *Proceedings of the IEEE*, vol. 73, no. 10, pp. 1468–1470, 1985.

[13] Department of Energy, U.S., "Benefits of demand response in electricity markets and recommendations for achieving them," 2005. A Report submitted to the United States Congress Pursuant to Section 1252 of the Energy Policy Act of 2005.

[14] J. Granderson, M. Sharma, E. Crowe, D. Jump, S. Fernandes, S. Touzani, and D. Johnson, "Assessment of model-based peak electric consumption prediction for commercial buildings," *Energy and Buildings*, vol. 245, 2021.

[15] K. Mahmud, J. Ravishankar, M. J. Hossain, and Z. Y. Dong, "The impact of prediction errors in the domestic peak power demand management," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4567–4579, 2020.

[16] R. F. Engle, C. Mustafa, and J. Rice, "Modeling peak electricity demand," *Journal of Forecasting*, vol. 11, no. 3, pp. 241–251, 1992.

[17] J. Moral-Carcedo and J. Perez-Garcia, "Integrating long-term economic scenarios into peak load forecasting: An application to spain," *Energy*, vol. 140, pp. 682–695, Dec 1 2017.

[18] J. Hoffer and M. Prill, "On the models of peak load forecast uncertainty in probabilistic production costing algorithms," *International Journal of Electrical Power & Energy Systems*, vol. 18, no. 3, pp. 153–160, 1996.

[19] P. McSharry, S. Bouwman, and G. Bloemhof, "Probabilistic forecasts of the magnitude and timing of peak electricity demand," *IEEE Transactions On Power Systems*, vol. 20, pp. 1166–1172, May 2005.

[20] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 798–805, 2001.

[21] A. Papalexopoulos and T. Hesterberg, "A regression-based approach to short-term system load forecasting," *IEEE Transactions On Power Systems*, vol. 5, pp. 1535–1547, Nov 1990.

[22] Y. T. Chae, R. Horesh, Y. Hwang, and Y. M. Lee, "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings," *Energy and Buildings*, vol. 111, pp. 184–194, Jan 1 2016.

[23] I. Drezga and S. Rahman, "Short-term load forecasting with local ann predictors," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 844–850, 1999.

[24] R. Cottet and M. Smith, "Bayesian modeling and forecasting of intraday electricity load," *Journal of the American Statistical Association*, vol. 98, no. 464, pp. 839–849, 2003.

[25] S. Chen, D. YU, A. Moghaddamjo, C. LU, and S. Vemuri, "Weather sensitive short-term load forecasting using nonfully connected artificial neural network," *IEEE Transactions On Power Systems*, vol. 7, pp. 1098–1105, Aug 1992.

[26] M. Bessec and J. Fouquau, "Short-run electricity load forecasting with combinations of stationary wavelet transforms," *European Journal of Operational Research*, vol. 264, pp. 149–164, Jan 1 2018.

[27] J. R. Xie and T. Hong, "Wind speed for load forecasting models," *Sustainability*, vol. 9, no. 5, 2017.

[28] J. Xie and T. Hong, "Load forecasting using 24 solar terms," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 2, pp. 208–214, 2018.

[29] J. Dorrington, I. Finney, T. Palmer, and A. Weisheimer, "Beyond skill scores: exploring sub-seasonal forecast value through a case-study of french month-ahead energy prediction," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 733, pp. 3623–3637, 2020.

[30] T. Haida, S. Muto, Y. Takahashi, and Y. Ishi, "Peak load forecasting using multiple-year data with trend data processing techniques," *Electrical Engineering in Japan*, vol. 124, no. 1, pp. 7–16, 1998.

[31] Y. Li, Y.-Q. Bao, B. Yang, C. Chen, and W. Ruan, "Modification method to deal with the accumulation effects for summer daily electric load forecasting," *International Journal Of Electrical Power & Energy Systems*, vol. 73, pp. 913–918, Dec 2015.

[32] C. A. Maia and M. M. Goncalves, "Application of switched adaptive system to load forecasting," *Electric Power Systems Research*, vol. 78, no. 4, pp. 721–727, 2008.

[33] Y. Amara-Ouali, M. Fasiolo, Y. Goude, and H. Yan, "Daily peak electrical load forecasting with a multi-resolution approach," *International Journal of Forecasting*, vol. 39, no. 3, pp. 1272–1286, 2023.

[34] B. J. Chen, M. W. Chang, and C. J. Lin, "Load forecasting using support vector machines: A study on eunite competition 2001," *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1821–1830, 2004.

[35] N. Amjady and F. Keynia, "Mid-term load forecasting of power systems by a new prediction method," *Energy Conversion and Management*, vol. 49, no. 10, pp. 2678–2687, 2008.

[36] C. C. Hsu and C. Y. Chen, "Regional load forecasting in taiwan - applications of artificial neural networks," *Energy Conversion and Management*, vol. 44, no. 12, pp. 1941–1949, 2003.

[37] S. Pezzulli, P. Frederic, S. Majithia, S. Sabbagh, E. Black, R. Sutton, and D. Stephenson, "The seasonal forecast of electricity demand: A hierarchical bayesian model with climatological weather generator," *Applied Stochastic Models in Business and Industry*, vol. 22, no. 2, pp. 113–125, 2006.

[38] Y. C. Jang, E. Byon, E. Jahani, and K. Cetin, "On the long-term density prediction of peak electricity load with demand side management in buildings," *Energy and Buildings*, vol. 228, 2020.

[39] 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.

[40] L. M. Saini and M. K. Soni, "Artificial neural network-based peak load forecasting using conjugate gradient methods," *IEEE Transactions on Power Systems*, vol. 17, no. 3, pp. 907–912, 2002.

[41] P. K. Dash, A. C. Liew, and S. Rahman, "Peak load forecasting using a fuzzy neural-network," *Electric Power Systems Research*, vol. 32, no. 1, pp. 19–23, 1995.

[42] S. M. Maksimovich and V. M. Shiljkut, "The peak load forecasting afterwards its intensive reduction," *IEEE Transactions On Power Delivery*, vol. 24, pp. 1552–1559, Jul 2009.

[43] J. Nagi, K. S. Yap, F. Nagi, S. K. Tiong, and S. K. Ahmed, "A computational intelligence scheme for the prediction of the daily peak load," *Applied Soft Computing*, vol. 11, pp. 4773–4788, Dec 2011.

[44] T. Chen, J. M. Lehr, O. Lavrova, and M. Martinez-Ramon, "Distribution feeder-level day-ahead peak load forecasting methods and comparative study," *IET Generation Transmission & Distribution*, vol. 12, pp. 3270–3278, Jul 31 2018.

[45] M. R. Haq and Z. Ni, "A new hybrid model for short-term electricity load forecasting," *IEEE Access*, vol. 7, pp. 125413–125423, 2019.

[46] C. Fan, F. Xiao, and S. W. Wang, "Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques," *Applied Energy*, vol. 127, pp. 1–10, 2014.

[47] O. Avatefipour and A. Nafisian, "A novel electric load consumption prediction and feature selection model based on modified clonal selection algorithm," *Journal Of Intelligent & Fuzzy Systems*, vol. 34, no. 4, pp. 2261–2272, 2018.

[48] T. Hong, "Short term electric load forecasting," 2010. North Carolina State University.

[49] R. Ramanathan, R. Engle, C. Granger, F. VahidAraghi, and C. Brace, "Short-run forecasts of electricity loads and peaks," *International Journal Of Forecasting*, vol. 13, pp. 161–174, Jun 1997.

[50] T. Haida and S. Muto, "Regression-based peak load forecasting using a transformation technique," *IEEE Transactions on Power Systems*, vol. 9, no. 4, pp. 1788–1794, 1994.

[51] A. Goia, C. May, and G. Fusai, "Functional clustering and linear regression for peak load forecasting," *International Journal Of Forecasting*, vol. 26, pp. 700–711, Oct-Dec 2010.

[52] Y. Zong-chang, "Modelling and forecasting of electric daily peak load movement based on the elliptic-orbit model with weekly periodic extension: a case study," *IET IET Generation Transmission & Distribution*, vol. 8, pp. 2046–2054, Dec 2014.

[53] J. Hoffer and P. Dorfner, "Reliability and production cost calculation with peak load forecast uncertainty," *International Journal of Electrical Power & Energy Systems*, vol. 13, no. 4, pp. 223–229, 1991.

[54] L. Xu, S. Wang, and R. Tang, "Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load," *Applied Energy*, vol. 237, pp. 180–195, Mar 1 2019.

[55] V. A. Evangelopoulos and P. S. Georgilakis, "Probabilistic spatial load forecasting for assessing the impact of electric load growth in power distribution networks," *Electric Power Systems Research*, vol. 207, 2022.

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

[57] A. Laouafi, M. Mordjaoui, F. Laouafi, and T. E. Boukelia, "Daily peak electricity demand forecasting based on an adaptive hybrid two-stage methodology," *International Journal Of Electrical Power & Energy Systems*, vol. 77, pp. 136–144, May 2016.

[58] L. M. Saini, "Peak load forecasting using bayesian regularization, resilient and adaptive backpropagation learning based artificial neural networks," *Electric Power Systems Research*, vol. 78, no. 7, pp. 1302–1310, 2008.

[59] S. M. Baek, "Mid-term load pattern forecasting with recurrent artificial neural network," *IEEE Access*, vol. 7, pp. 172830–172838, 2019.

[60] M. H. Choueiki, C. A. MountCampbell, and S. C. Ahalt, "Implementing a weighted least squares procedure in training a neural network to solve the short-term load forecasting problem," *IEEE Transactions on Power Systems*, vol. 12, no. 4, pp. 1689–1694, 1997.

[61] W. Y. Zhang, W.-C. Hong, Y. Dong, G. Tsai, J.-T. Sung, and G.-F. Fan, "Application of svr with chaotic gasa algorithm in cyclic electric load forecasting," *Energy*, vol. 45, pp. 850–858, Sep 2012.

[62] Y. M. Huang, N. Hasan, C. R. Deng, and Y. K. Bao, "Multivariate empirical mode decomposition based hybrid model for day-ahead peak load forecasting," *Energy*, vol. 239, 2022.

[63] T. H. Ouyang, Y. S. He, H. J. Li, Z. Y. Sun, and S. Baek, "Modeling and forecasting short-term power load with copula model and deep belief network," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 2, pp. 127–136, 2019.

[64] C. S. Lai, Z. Y. Mo, T. Wang, H. L. Yuan, W. W. Y. Ng, and L. L. Lai, "Load forecasting based on deep neural network and historical data augmentation," *IET Generation Transmission & Distribution*, vol. 14, no. 24, pp. 5927–5934, 2020.

[65] K. D. Zhu, Y. P. Li, W. B. Mao, F. Li, and J. H. Yan, "Lstm enhanced by dual-attention-based encoder-decoder for daily peak load forecasting," *Electric Power Systems Research*, vol. 208, 2022.

[66] Z. Y. Yu, Z. W. Niu, W. H. Tang, and Q. H. Wu, "Deep learning for daily peak load forecasting-a novel gated recurrent neural network combining dynamic time warping," *IEEE Access*, vol. 7, pp. 17184–17194, 2019.

[67] M. R. Amin-Naseri and A. R. Soroush, "Combined use of unsupervised and supervised learning for daily peak load forecasting," *Energy Conversion And Management*, vol. 49, pp. 1302–1308, Jun 2008.

[68] P. R. S. Jota, V. R. B. Silva, and F. G. Jota, "Building load management using cluster and statistical analyses," *International Journal of Electrical Power & Energy Systems*, vol. 33, no. 8, pp. 1498–1505, 2011.

[69] M. Moazzami, A. Khodabakhshian, and R. Hooshmand, "A new hybrid day-ahead peak load forecasting method for iran's national grid," *Applied Energy*, vol. 101, pp. 489–501, Jan 2013.

[70] M. Tamimi and R. Egbert, "Short term electric load forecasting via fuzzy neural collaboration," *Electric Power Systems Research*, vol. 56, no. 3, pp. 243–248, 2000.

[71] S. E. Papadakis, J. B. Theocharis, S. J. Kiartzis, and A. G. Bakirtzis, "A novel approach to short-term load forecasting using fuzzy neural networks," *IEEE Transactions on Power Systems*, vol. 13, no. 2, pp. 480–489, 1998.

[72] J. Nowotarski, B. Liu, R. Weron, and T. Hong, "Improving short term load forecast accuracy via combining sister forecasts," *Energy*, vol. 98, pp. 40–49, Mar 1 2016.

[73] H. Saxena, O. Aponte, and K. T. McConky, "A hybrid machine learning model for forecasting a billing period's peak electric load days," *International Journal of Forecasting*, vol. 35, no. 4, pp. 1288–1303, 2019.

[74] O. Aponte and K. McConky, "Peak electric load days forecasting for energy cost reduction with and without behind the meter renewable electricity generation," *International Journal of Energy Research*, vol. 45, no. 13, pp. 18735–18753, 2021.

[75] C. Sigauke and M. M. Nemukula, "Modelling extreme peak electricity demand during a heatwave period: a case study," *Energy Systems-Optimization Modeling Simulation and Economic Aspects*, vol. 11, no. 1, pp. 139–161, 2020.

[76] D.-C. Wu, A. Amini, A. Razban, and J. Chen, "Arc algorithm: A novel approach to forecast and manage daily electrical maximum demand," *Energy*, vol. 154, pp. 383–389, Jul 1 2018.

[77] ISONE, USA, "Maps and diagrams - load zones," 2023. https://www.iso-ne.com/about/key-stats/maps-and-diagrams/dispatch-zones, Last accessed on 2023-08-10.

[78] D. Yang, W. Wang, and T. Hong, "A historical weather forecast dataset from the european centre for medium-range weather forecasts (ecmwf) for energy forecasting," *Solar Energy*, vol. 232, pp. 263–274, 2022.

[79] P. Wang, B. Liu, and T. Hong, "Electric load forecasting with recency effect: A big data approach," *International Journal of Forecasting*, vol. 32, pp. 585–597, Jul-Sep 2016.

[80] *Introduction to the Logistic Regression Model*, ch. 1, pp. 1–33. John Wiley & Sons, Ltd, 2013.

[81] S. Haben, J. Ward, D. V. Greetham, C. Singleton, and P. Grindrod, "A new error measure for forecasts of household-level, high resolution electrical energy consumption," *International Journal of Forecasting*, vol. 30, pp. 246–256, Apr-Jun 2014.

[82] S. Aman, Y. Simmhan, and V. K. Prasanna, "Holistic measures for evaluating prediction models in smart grids," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 2, pp. 475–488, 2015.

[83] T. Yunusov, G. Giasemidis, and S. Haben, "Smart storage scheduling and forecasting for peak reduction on low-voltage feeders," in *Energy Management-Collective And Computational Intelligence With Theory And Applications* (C. Kahraman and G. Kayakutlu, eds.), vol. 149 of *Studies in Systems Decision and Control*, pp. 83–107, 2018.

[84] C. Singleton and P. Grindrod, "Forecasting for battery storage: Choosing the error metric," *Energies*, vol. 14, no. 19, 2021.

[85] B. Casati, L. J. Wilson, D. B. Stephenson, P. Nurmi, A. Ghelli, M. Pocernich, U. Damrath, E. E. Ebert, B. G. Brown, and S. Mason, "Forecast verification: current status and future directions," *Meteorological Applications*, vol. 15, no. 1, pp. 3–18, 2008.

[86] M. AH, "Probabilities, odds, and forecasts of rare events - reply," *Weather and Forecasting*, vol. 6, p. 575, Dec 1991.

[87] G. W. Brier, "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, vol. 78, no. 1, pp. 1 – 3, 1950.

[88] M. W. Liemohn, A. D. Shane, A. R. Azari, A. K. Petersen, B. M. Swiger, and A. Mukhopadhyay, "RMSE is not enough: Guidelines to robust data-model comparisons for magnetospheric physics," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 218, p. 105624, 2021.

[89] R. N. Hoffman, Z. Liu, J. F. Louis, and C. Grassotti, "Distortion representation of forecast errors," *Monthly Weather Review*, vol. 123, no. 9, pp. 2758–2770, 1995.

[90] C. Keil and G. C. Craig, "A displacement and amplitude score employing an optical flow technique," *Weather and Forecasting*, vol. 24, no. 5, pp. 1297–1308, 2009.

[91] Energy Department, USA, "Evaluating your utility rate options," 2023. https://www.energy.gov/femp/evaluating-your-utility-rate-options, Last accessed on 2023-08-10.

[92] X. Lu, Z. Y. Dong, and X. Li, "Electricity market price spike forecast with data mining techniques," *Electric Power Systems Research*, vol. 73, no. 1, pp. 19–29, 2005.

[93] "ISONE glossary and acronyms." https://www.iso-ne.com/participate/support/glossary-acronyms/OPCR. Accessed: 2023-07-25.

[94] Y. Li, "Short term ex-ante load forecasting," 2022. University of North Carolina at Charlotte.

[95] "Final 2023 heating electrification forecast." https://www.iso-ne.com/static-assets/documents/2023/04/heatFx2023$_f$*inal.pdf*. *Accessed* : $2023 - 07 - 25$.