FORECASTING THE IMPACT OF CLIMATE CHANGE FOR ELECTRIC POWER MANAGEMENT IN THE SOUTHWEST

by

Paulo Minoru Tanimoto

A Thesis Submitted to the Faculty of the

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2 0 0 8

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: ______________________

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

______________________________ ______________

Dr. Bonnie Colby
Professor of Economics

Date

TABLE OF CONTENTS

TABLE OF CONTENTS – *Continued*

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Multiple challenges are gathering in the southwestern region of the U.S., resembling the formation of a storm. The explosive growth in population and economy over the past years have only aggravated the existing scarcity of natural resources in the region. Climate change is expected to amplify these issues even further. Electricity providers are particularly challenged in this situation, because demand for energy is known to be connected with many factors, including population, weather, regulations, and water. One way of improving this scenario is to promote more efficient use of inputs. Thus, we develop statistical models for forecasting short-term and medium-term usage of electricity, using data from an electric utility located in southern Arizona. Our results for the short-term show that predictions can be improved significantly when weather variables are included, especially temperature. For the medium-term, we find that climatic forecasts that are publicly available can perform well in a seasonal planning model.

CHAPTER 1

Introduction

The issue of climate and electricity in the Southwest is a matter for social and economic concern. The southwestern region of the U.S. has been characterized by high population and economic growth, and households and businesses now rely more on electricity-demanding technology than in the past. In addition, climatic change can further aggravate this problem, because it is unclear how consumers will respond to new conditions. We propose to create statistical models to forecast demand for electricity, taking into account weather and climatic information.

That climatic factors have an influence on demand for electricity there is little disagreement. Localities facing extreme summer or winter temperatures tend to employ more energy for cooling or heating. Use of electricity can also be influenced by the occurrence of severe droughts or storms. In a broader sense, human beings have developed electricity-powered devices to endure adverse conditions.

Nevertheless, most researchers in the area assume climate to be static, despite growing evidence suggesting that climate is now changing at a fast pace. Such changes can introduce greater uncertainty in the relationship between climate and energy. On the one hand, if climate change is partly attributed to human actions, the electricity sector is known to be one of the biggest polluters of the environment. On the other hand, it is not certain how individuals will react to changes in climate. For example, in the Southwest warmer temperature can result in more energy used for cooling during the summer, but less for heating in the winter.

Water and energy are also closely interrelated. Directly or indirectly, vast quantity of water is consumed in the process of producing electricity, including in the extraction of fossil fuels, the vaporization from hydroelectric power plants, or cooling for nuclear plants. In fact, the electric sector is the second largest user of water, only surpassed by agriculture (Department of Energy, 2006). Conversely, much electricity is also involved in the treatment of water for consumption and the delivery of water for agricultural, municipal, and industrial uses.

With climate change, the water-energy nexus is intensified. In the case of the Southwest, warmer temperature can increase demand for both water and energy for cooling. Because of the nexus, more use of one resource is likely to be associated with more consumption of the other.

Another aggravating element in the climate and energy issue is the economic and population growth in the Southwest region. The region is among the main destinations of migration in the country, and projections from the Census Bureau 2005 expect this trend to continue for at least the next decade. Between 1990 and 2000, the population in the state of Arizona increased by 40% and New Mexico by 20%, whereas the national rate was 8% (Merideth, 2001). By 2025, the region is projected to grow by 51%. This demographic trend is coupled, and partly associated, with intense economic growth experienced in the Southwest in recent years. Together, these factors contribute to greater demand for water and energy. All else equal, more households, commerce, and industry are likely to be accompanied by greater use of resources.

Finally, the issue is also sensitive in the supply side. By its very nature, electricity cannot be efficiently stored and transmission is limited by existing infrastructure, so the amount to be generated and delivered must be decided in advance. However, the quantity consumed may differ from what was anticipated: if the firm overproduces, the surplus is wasted unless it is sold in the spot market; if it underproduces, the deficit has to be compensated by purchasing from other companies, otherwise power outages can occur. Thus, electric power providers can benefit from knowing how much energy would be demanded. With uncertainty, the best they can do is attempt to predict future demand.

This is generally how most electricity generators proceed in practice, making use of forecasting techniques that range from expert judgment to sophisticated statistical and artificial intelligence models. However, because climate is usually assumed to be constant, electric power providers may be missing an important element in planning. Ignoring climatic effects can have severe consequences to social and economic sustainability in the future.

We propose to contribute to this area by developing statistical models that take into account weather and climatic information in forecasting future demand for electricity. We test our models against purely univariate specifications, using load data provided by an electric utility company in the Southwest region. We then evaluate our forecasts with standard measures from the economics and meteorological literature. Our expectation is that our informed models will outperform the predictions that are currently done by the company and models that do not employ weather or climatic data.

Ultimately we hope our work will bring benefits to all groups involved in the study. We have been provided exclusive and confidential access to load data that are essential to test our hypotheses. The utility company will be able to assess whether alternative techniques can improve its efficiency. Society as a whole is also expected to gain from the investigation, because better use of resources can reduce pollution and result in more reliable and cost effective provision of electricity.

1.1 Impact of Climate Change

Although the effects of climate change worldwide can be diverse, depending on the locality and the sector in question, Lenart (2007, p. 2) remarks that the outlook for the Southwest region is clearly pessimistic: global warming is likely to aggravate several problems, including scarcity of water, loss of plant and animal species, and even homes and forests due to wildfires. Besides higher temperature averages and more heat waves, the author notes the region would paradoxically suffer from both more droughts and more floods, due to the occurrence of monsoon flooding.

To support her claims, Lenart cites evidence from the latest report of the Intergovernmental Panel on Climate Change (IPCC, 2007). IPCC, a respected scientific group in the area of climate, develops projections for temperature and precipitation from multiple simulation models. In the case of North America, the results come from 21 climatic models, running under different initial conditions and specifications. The output values can then be averaged or studied for their distributional properties.

According to the report, during this century the Southwest region is expected to suffer from an increase in the annual average temperature ranging from 4.5°F to more than 7.5°F (IPCC, Figure S11.8). The average of all simulations indicates a rise of about 5°F. For precipitation, the projections have more variability, but they range from a decrease of 10% to a severe 30% drop, compared to current levels (IPCC, Figure S11.16).

In addition to the change in average levels, the occurrence of extreme conditions is also expected to increase. Lenart observes that "the number of extremely hot days is also projected to rise over the decades, leaving parts of the region with heat waves lasting an extra two weeks by the end of the century" (p. 2). Combined with the explosive urban expansion in the region, heat island effects can happen in higher frequency, accentuating the impacts of climatic change.

1.1.1 Climate Change and the Electric Sector

Demand for electricity is known to be strongly influenced by climate conditions. With climate change occurring, it remains uncertain how the electric sector will be affected. The concern is certainly not new, but only a few studies have attempted to rigorously assess the impacts of such changes. The Electric Power Research Institute (EPRI) has sponsored at least two studies and a workshop of this kind.

The first was mostly a qualitative analysis titled "Potential Impact of Climate Change on Electric Utilities" (EPRI, 1989). Using a New York and a southeastern electric utility as case studies, the investigators considered the possible consequences that climate change might create for electric utilities in 15 to 30 years, which is the typical time horizon for planning. In contrast with the approach of taking climate as constant, different scenarios were developed from results of a number of climate simulation models, and then analyzed in terms of change in fuel costs and demand.

At the time, the investigators concluded that climate change could have significant effects for the supply and demand for electricity in the following 15 to 30 years. More specifically, they speculated that by 2015 increased average temperatures alone could lead to: a) greater requirements in terms of generating capacity; b) more demand for electricity in general; c) scarcity of hydroelectric resources; and d) increased prices and production costs. For these reasons, continuing to ignore the importance of climate change in planning could be costly to electric utilities.

EPRI's own perspective on that study was that until the 1980s climate could be regarded as constant with minimal impact for utilities, which are more affected by fuel costs and demographic factors. Because the study lacked rigorous modeling, EPRI considered the results to be of preliminary nature and decided to sponsor a follow-up work (EPRI, 2008).

The second study (EPRI, 1995) involved case studies of six electric utilities, and a wider range of techniques. It also extended the time horizon to 2050, and had a stronger focus on the impacts of carbon dioxide emissions and regulations. Like the first study, instead of attempting to model the relationship between climate and electricity, the investigators considered the impacts under four climate change and four carbon dioxide scenarios.

As expected, the results varied from utility to utility, but overall the authors found that carbon dioxide emissions are likely to have a major impact for energy providers, even higher than physical climate change, largely due to regulations and difficulty in stabilizing output. In fact, the results suggest that common strategies, such as demand-side management and switching to natural gas, will not be sufficient to reduce the long-term emissions of carbon dioxide. Instead, nuclear and renewable resources would have to be contemplated. On the climate side, it was estimated that a 1°C increase in average temperature by 2050 would result in a growth in electricity demand by roughly 5%, but for periods of peak demand, the increase could range from 5% to 14%.

More recently, EPRI (2005) organized a workshop under the title "Identifying Research to Help Electric Companies Adapt to Climate Change". Again, specialists who attended the event agreed that utilities already face uncertainties from different sources in their planning. Although climate appears to play a secondary role, severe conditions, including heat waves and extreme droughts, that coincide with high-demand seasons can have large effects for power generation and even cause disruptions of service. There was a widespread consensus that more applied research on the effects of climate change is necessary, since climate forecasts have improved significantly over the past years. Besides load forecasting, other areas considered relevant by specialists were generation, transmission and distribution (T&D), and supply planning.

1.1.2 Water-Energy Nexus

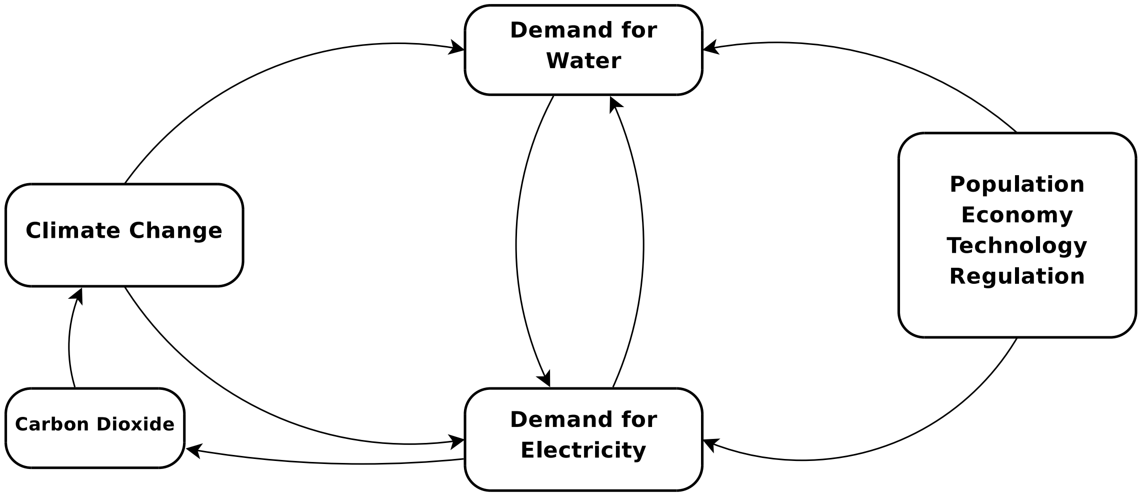
Naturally, electricity loads are linked to many factors. But among these, the water-energy nexus is particularly relevant for this discussion. Not only are the two elements intimately connected, but their usage is also sensitive to climatic and weather conditions. Environmental change can, therefore, pose serious challenges to sustainability. Acknowledging this problem, the Department of Energy published in 2006 a report to Congress titled "Energy Demands on Water Resources". Citing a large number of independent studies, the document details the interdependencies between water and electricity at different stages of processing for consumption.

Water is used, directly or indirectly, in most of the process of electricity generation. It is employed in the extraction, refining, and transportation of fuel resources. In hydroelectric plants a large volume of water is lost due to evaporation and similarly, thermoelectric generation demands water for cooling. In fact, agriculture and the electric sector have for several decades been the top destinations of water withdrawals (Thompson, 1999).

Conversely, electricity is also used in the pumping, treatment, and distribution of water for consumption. According to the report to Congress, about 4% of the total power generated in the U.S. is destined for water supply, a level that is comparable to other large industrial sectors. A study by Powicki (2002) suggests that about 75% of the cost of water treatment in municipalities comes from electricity. In investigating the nexus, the California Energy Commission (2005) has found that electricity usage associated with water consumption is even larger than energy used for water supply and treatment. The report indicates that water heating, clothes washing, and clothes drying can account for up to 14% of California's electricity consumption, mostly due to the residential sector.

With this interdependency, the impact of climate change can be amplified because it affects demand for both water and energy. Thus by increasing consumption of water it can increase usage of electricity and vice-versa, in a perverse cycle. Undoubtedly, the actual effect would vary from situation to situation. For cold regions, for instance, higher temperatures can lead to a decrease

Figure 1.1: The Water-Energy-Climate Nexus



in the use of energy for heating, but can lead to more use of water for leisure. However, for the Southwest region the picture is more pessimistic: not only are warmer days associated with more use of electricity and water at peak periods, but also with greater incidence of droughts. In other words, climate change can aggravate the likelihood of severe shortages in the region.

These relationships are depicted in Figure 1.1, which illustrates that demand for water and electricity are influenced by each other and by external factors, such as climate change, demographics, economy, and regulations. Although the complexity is daunting, recall that the focus of this study is on one specific part of the overall complexity: to investigate how knowledge of climate change can improve planning for electric utilities.

1.2 Forecast as Strategic Planning

One of the most effective ways of dealing with changing conditions is to prepare in advance. A recent example in the electric sector is the "Guide to Tools and Principles for a Dry Year Strategy", developed by the Bonneville Power Administration (BPA) in collaboration with the public. The document presents lessons that have been learned from energy crises in 2000 and 2001, and proposes concrete strategies that can be put into action in case of future electricity supply crises.

According to the guide, the most severe problem for electric utilities is not shortage of electricity in general, but the shortage to meet demand at peak levels. In the case of BPA, this issue is aggravated by the fact that the generation process is highly dependent on hydroelectric plants. When faced with a year with low storage levels in reservoirs, BPA must trade-off between supplying electricity at peak levels and having capacity to generate electricity for the rest of the year.

When explaining the strategies, a few points are mentioned in passing which are valuable lessons for the electricity planning issue. The first is that public awareness campaigns typically achieve small results and, most importantly, are short-lived. The second is that a tension can arise between reserving water for agriculture and water for generating electricity. Third is the observation that it may be advantageous to negotiate "buy-downs", agreements with major electricity-using industry to decrease their economic activity and demand for electricity under specific conditions. Finally, the guide proposes that exchange agreements between power companies can be mutually advantageous, especially when locations have peak demand in different times of the year.

These guidelines support our claim that forecasts can play a central role in the electric sector: the underlying assumption behind the strategies is that demand for energy can be estimated sufficiently in advance, so that preemptive actions can be taken. This argument is well summarized by (Feinberg and Genethliou, 2005):

> "Load forecasting has always been important for planning and operational decisions conducted by utility companies. However, with the deregulation of the energy industries, load forecasting is even more important. With supply and demand fluctuating and the changes of weather conditions and energy prices increasing by a factor of ten or more during peak situations, load forecasting is vitally important for utilities." (p. 270)

There is little doubt that knowledge about the future can be extremely valuable for governmental and corporate planning. However, accurate foresight is often an impossibility, so decision-makers use available information to attempt to predict what will happen, either through intuition based on years of observation or sophisticated quantitative methods. In the end, the goal of these forecasts remains the same: improving the set of actions, that is, the plan which the organization will implement to achieve or prevent certain outcome. Similarly, the time scale of forecasts and planning can range from short-term to medium-term. In this study, we develop models for both types of interval.

CHAPTER 2

Review of Literature

The problem of maximizing profit (or minimizing costs) for a power generator is complicated by the fact that electricity is quickly perishable, so what is produced may differ from what is demanded with consequences for costs and profits. We connect this peculiarity to the economic literature on production under demand uncertainty, which in turn suggests an appropriate statistical approach. We then review similar studies on electricity forecasting that have been done, both with qualitative and quantitative approaches. Finally, we briefly discuss the existing literature on the economic value of load forecasts for not only electric utilities, but also society as a whole.

2.1 Production Under Demand Uncertainty

The problem that electricity firms face can be framed as one of producing an output in order to maximize profits under demand uncertainty. To illustrate, we know that in a scenario of perfect information, a producer wants to maximize a profit function as given by:

$$\pi = p\,Q - C(Q) \tag{2.1}$$

where π is the profit; p is the electricity sales price, per unit; Q is the quantity demanded; and $C(\cdot)$ is the cost function.

However, the physical nature of electricity prevents any efficient type of storage. It must be consumed almost immediately, resembling a commodity that is quickly perishable. Since the producer has to decide in advance the amount of electricity that will be generated for the next period, he or she must attempt to predict the quantity that will be demanded. This quantity is the total amount of electricity actually used by consumers and, because of the element of unpredictability, it can be considered a random variable. Therefore, we must distinguish between quantity produced (Q) and quantity demanded (D).

In Equation 2.2, the profit function has been rewritten to reflect the fact that revenue will depend on the minimum value between quantity produced and demanded. The equation also includes a term, M, to represent the costs involved in modeling, such as obtaining information and generating forecasts. Thus, when the firm overproduces, it incurs production costs and it only receives payment from the amount that was actually demanded. When the firm underproduces, it will still obtain revenue only from the minimum value, but there is the additional problem that it will have to purchase electricity from other companies to meet the demand, possibly at a premium price. If it fails to do so, the customers will suffer black-outs, which are highly undesirable, and costly to the firm in ways which are difficult to quantify.

$$\pi = p \min(Q, D) - C(Q, M) \tag{2.2}$$

This basic problem of a monopolist choosing its output level under uncertainty has already been addressed in the economic literature (Sandmo, 1971; Ishii, 1977, 1991). For the electricity case, Chao (1983) discusses a general framework to cope with uncertainty in both capacity and demand simultaneously. Gardner and Rogers (1999) take a different approach and consider specific details in expanding power plants. For the purposes of this study, we are interested in investigating how simple methods can reduce uncertainty and help improve a firm's profit.

A risk-neutral firm will maximize expected profit:

$$\begin{aligned} E(\pi) &= p\ E[\min(Q,\tilde{D})] - E[C(Q,M)] \\ &= p\ \min[Q, E(\tilde{D})] - C(Q,M) \end{aligned} \tag{2.3}$$

where the new symbol, $\tilde{D}$ is a random variable corresponding to the unknown quantity to be demanded.

This implies that a profit-maximizing producer is interested in minimizing the disparity between the quantity produced and the expected quantity demanded, assuming all else constant. This difference can be thought of as the forecasting error. Minimizing the error can be accomplished through the collection of more information or with the improvement of prediction power. It is usually less costly and easier for firms to change their forecast techniques than to invest in measurement equipment to gather previously unavailable data relevant to electricity demand.

Naturally, there are also costs associated to modeling and making forecasts, again, captured by M in the above equation. The initial costs of producing models can include hiring or training of staff, licensing statistical software, acquiring weather forecasts or data for other variables, as well as purchasing adequate computer equipment. Once these requirements are in place, the cost of making an additional forecast for any given model is very small, typically only involving updating the model with new observations and running the estimation again. The advantage of the method proposed in the next chapter is that only one model needs to be estimated, as opposed to creating multiple models for different parts of the year.

Improving efficiency with better forecasts is indeed a point where our study can prove valuable. Developing models that can better utilize available information to predict the future is a strategy that has been used for many years in the electric sector. As we will discuss below, even small improvements in the percentage error can amount to savings of millions of dollars every year for companies.

2.2 Previous Forecasting Studies

Load forecasting has a long history of cooperation between academia and the industry. While in the early days of power generation the decision of how much to produce, would be taken based solely on the experience and judgment of experts, this method alone proved of limited success as the scale and complexity of operation started to grow. An ever increasing demand for energy coupled with constrained ability to produce meant that wrong decisions could lead to system-wide black-outs or large inefficiencies. Even though until the 1980s the electric sector was considered to be a natural monopoly, in which companies enjoyed little to no competition, the industry still worked on improving forecasting methods.

With the initial growth of the electric sector, much of the work on prediction was based on simple estimates that would take into account hard facts about consumers: number of households, number of appliances, proportion of commercial and industrial sectors, and so on. In other words, these were direct assessments of how many consumers of various types would demand electricity. Although this method was very realistic, it suffered from the drawback of ignoring the varying behavior of households and businesses. Consumption, it could not be ignored, is not merely determined by the number of electrical devices in the grid.

Researchers and electric utilities shifted to a more stochastic approach by employing simple regression models, in which system load would be a function of not only consumer characteristics, but also of the time of the day and weather conditions (Papalexopoulos and Hesterberg, 1990). This method has an advantage over the previous technique in that it is capable of capturing finer fluctuation yet still in a simple manner. Gross and Galiana (1987) also note that this simplicity enables the operator to easily update the model parameters by linear regressions or linear exponential smoothing. However, the authors point out that on the negative side, "time-of-day models do not accurately represent the stochastically correlated nature of the load process, or its relation to weather variables. As a result, when weather patterns are changing rapidly, the coefficients are not appropriate, except for a short time interval into the future" (p. 1563).

During the 1980s, more purely dynamic models were proposed. Instead of attempting to uncover the precise set of explanatory variables behind demand, a task that can reach high complexity, time-series models take advantage of recurrent patterns in the data. By specifying load as a function of past load, other factors become implicitly accounted for in the model. Typically this approach utilizes techniques developed by Box and Jenkins (1970), in which autoregressive (AR) and moving-average (MA) terms are specified, as well as differencing for removing trends in the data (Hagan and Behr, 1987). One strong advantage, besides minimal requirements on external variables, is the ability of dynamic models to capture nonlinearity. However, one of the major issues of traditional time series models is the fact that the same model can often be represented in multiple alternative forms. The lack of a rigorous set of rules to guide the modeling exercise has led some authors to consider it an art, rather than science.

The 1990s witnessed the emergence of artificial intelligence techniques applied to the problem of load forecast. In particular, Artificial Neural Networks (ANN) enjoyed much acceptance by researchers for their immense flexibility in learning a variety of patterns with seemingly minimal interference by the modeler. Early studies include publications by Peng et al. (1992) and Papalexopoulos et al. (1994). Despite much initial enthusiasm, neural networks did not prove to be the panacea of modeling. Their flexibility came at the risk of converging at local minima, thus undermining its generalization power. Because of its sensitivity to structure and initial conditions, modeler expertise was found to be extremely important. Finally, different from econometric methods, the coefficients from neural networks suffered from not having an easy interpretation. Precisely because of its multiplicative structure, the effects from different parameters would be highly nonlinear.

Our study takes the middle route of creating a dynamic model extended with explanatory variables. From the time series toolbox, we take the idea that current demand for electricity can be partly predicted from the evolution of load itself. Due to inertia, it is likely that current behavior will be similar to past ones. In addition, this approach has the appealing feature that other variables affecting demand, possibly unobservable ones, are implicitly captured in the previous observations. Because usage of energy is highly sensitive to changing conditions, we also follow the regression tradition of including additional explanatory variables to make more informed decisions. As will be seen in the next chapter, this study also proposes an improved method of modeling seasonality for load data.

We refrain from using artificial intelligence techniques due to our interest in learning the relationship between climatic variables and electricity use. Even though machine learning algorithms have the advantage of being largely automatic and sensitive to nonlinear patterns, we believe that better performance can be obtained if we can specify the structural form of the model. Part of our goal is to achieve better understanding of the influence of weather and climatic factors on energy consumption, but methods such as neural networks and support vector machines work as black boxes that simply give a point prediction.

2.3 Economic Value of Forecasts

Although consensus exists that forecasts are economically important to electricity providers, not much attention has been put into rigorously quantifying their value. A study by Hobbs (1997) addressed this issue by using an earlier survey for EPRI, in 1997, in which 28 utilities were interviewed and reported estimated savings from switching to better forecasting techniques. The average value reported was \$800,000 per year. One utility, with a peak load of 35GW, reported that its 1.5% reduction in the MAPE of forecasts would correspond to an annual savings of \$7.6 million. The author notes, however, that those estimates were calculated in an informal manner, making use of different sets of assumptions.

As a follow-up, Hobbs et al. (1999) used data provided by two of those utilities to obtain a more rigorous assessment of those savings. The authors combined forecast simulation, optimization for unit commitment at each hour, and modeling of economic dispatch. They identified 12 decisions with economic implications that managers face in electricity generation (p. 1342):

- “Commitment of generating units
- Short run hydropower scheduling
- Economic dispatch of committed units
- Predictive automatic generation control
- Spinning reserve
- Fuel allocation
- Short-term energy purchases and sales
- Real-time prices
- Load interruption
- Load control
- Generator and transmission line maintenance
- Available transmission capability”

Forecasts errors, also called by the authors “regret”, naturally incur costs for the generator, since the difference of what is produced from the quantity that is actually demanded may be wasted or may be sold in the spot market for a low price. When an overforecast occurs, that is, more electricity is produced than demanded, Hobbs et al. enumerate the following sources of costs for the utility (p. 1342):

- “Units may have been unnecessarily committed, raising fuel costs

- Expensive power may have been purchased which was not needed, or a profitable opportunity to sell bulk power might have been spurned
- Hydropower may have been produced which would have been more valuable if generated at a later time
- Overly high real-time prices might have been quoted, depressing sales
- Unnecessary interruptions or load controls might be invoked, annoying consumers and lowering revenue"

On the other hand, when utilities underforecast and generate less electricity than what turns out to be demanded, the following consequences can follow (p. 1342):

- "Insufficient resources may be available for meeting security constraints, such as spinning reserve margins
- To meet the unanticipated load increase, uneconomic generation or purchases of spot power might be necessary
- Commitments to sell power may have been made at a price less than the value of that power to the utility
- Too low real-time prices might have been quoted, resulting in revenue falling short of the utility's cost"

Although the potential savings depend on the characteristics of the utilities and the behavior of consumers, the authors estimate that a 1% decrease in the MAPE can reduce variable costs of generation by .1% to .3%. In concrete terms, for a utility with 10GW peak load, the value can reach $1.6 million, annually.

Based on the previous study, Teisberg et al. (2005) extended the models and calculations to estimate the economic value of temperature forecasts for electricity generation. The government, through different agencies, provides weather forecasts free of charge, under the assumption that they are a form of public goods. As such, if left alone private organizations would have no incentive to develop and offer public forecasts and they would tend to be produced in a suboptimal quantity. Because little research exists on the value of these forecasts for the electric sector, it is interesting to compare the marginal benefits of improved weather forecasts to the marginal costs of generating them.

The authors choose to employ load data from six electric utilities, located in Vermont, Ohio, Florida, Texas, southern California, and Washington. The reason for choosing these diverse locations is to capture different geographies and climatic conditions in the United States. The weather information used consisted of daily temperature forecasts from the National Weather Service (NWS), as well as their actual values. These are compared to the naive approach of persistence forecasts, that is, assuming that tomorrow's temperature will be the same as today's. From the calculations, the authors estimate that the nationwide value of the NWS forecasts is $166 million annually, relative to persistence forecasts. A large proportion of this value (about 89%) comes from utilities located in the south, because weather variables play a significant role in the region, especially through energy used for cooling in the summer.

No information on specific costs and benefits of improved forecasts was available for this study. However, because the region of analysis is located in the Southwest, load is also highly sensitive to weather variation. Although we quantify estimate the net benefits from improved forecasts, we can compare the performance of purely univariate models with models extended with weather variables. We can, then, observe the amount of improvement in the load forecasts that is gained by including the additional variables.

2.3.1 Climate and Load Forecasts

Weather forecasting is known to be a difficult enterprise and even with the best of current models and technology, prediction power falls considerably after one or two weeks. That is because weather, in its short-term manifestation, is chaotic and highly volatile. Climate, on the other hand, is believed by researchers to be more predictable, because "the physical basis (...) lies in components that vary slowly compared with individual weather events", such as ocean and land surface events (Palmer et al., 2005, p.1996). El Niño and La Niña are examples of recurrent climatic phenomena that make seasonal outlooks more predictable. Despite these advances, little work has been done in the literature to connect demand for electricity to climate forecasts. Here we review some of the key pieces in the area.

In one of the earliest studies, Warren and LeDuc (1981) proposed to evaluate the economic value of climate for electric utilities. The authors show that even crude climatic measures, such as cooling and heating degree days can prove to be valuable for long-term planning. In fact, the authors suggest that a cursory glance at the existing literature is likely to underestimate the actual usage of climatic indices in practice:

> "Several types of models estimate energy demand by incorporating climate and economic indices. The sources of these models are primarily the individual electric utilities, followed by state public service commissions, other government agencies, and academia. Consequently, the published information on the models understates their widespread development and use" (p. 1431).

To show the adequacy of such measures, the authors develop a medium-term model based on an index of quarterly accumulated cooling degree-days. The researchers use data on natural gas sales, but note that the “climate response is similar to that of electricity consumption during the winter months”. Their model performs well, especially for residential customers. They conclude that climate information can be successfully used in statistical models, bringing support to the practice of many utilities.

However, it was not until a multidisciplinary research conducted by Weiss (1982), that researchers started to consider the value of climate forecasts, as opposed to historical indices, for electricity generation. This study involved experts in “meteorology, applied physics, statistics, system analysis, water resources engineering, political science, political economy and law” (p. 510). The list of collected data is long and included the reliability of climate forecasts back then, the effect of weather on natural gas and electricity for multiple seasons, the effect of weather on water use by power plants, the decision process in the private energy sector, and the extent of how these decisions are weather-sensitive.

The researchers concluded, contrary to their initial expectations, that climate information had “limited value to the private sector in managing supply”. The main reason for this conclusion was that, even though many corporate decisions are weather-sensitive, the climate forecasts available at the time were severely limited, often unreliable, and with no lead-time (only for the current season). Also, the format of the predictions, averages for the entire period, did not indicate concentration or duration of abnormality, which would be more important for decision-makers.

The author also cites the political costs of errors as a key drawback in adopting climate forecasts. In particular, officials do not want to take responsibility for what she calls the “two-class error”: a winter that is forecast to have above average temperature, but turns out to be below average, and a summer with above average temperature which was wrongly forecast to stay below average. In both cases, adopting those forecasts would lead to taking costly measures to prevent crises that would not occur in the first place. Thus, for a number of reasons climate predictions were not popular among utility planners.

More than a decade later, Changnon et al. (1995) conducted a survey to assess the use of climate forecasts by utility managers. The survey involved interviews with 56 decision makers in six utilities to investigate the use and potential use of climate forecasts. At the time, only 3 out of 56 actually employed the forecasts at work, but the vast majority (80%) admitted that they would be very useful for long-term planning if some changes were made and more information on them was available.

In the same year, the Climate Prediction Center (CPC) at NOAA had completely revised its climate forecasts. Instead of monthly predictions only for the current season, the agency would publicize 3-month forecasts, out to 13 overlapping periods ahead, thus attending to one of the major complaints by utility planners as reported by Weiss. The new format also featured forecasts in probability terciles, instead of point predictions. The terciles were constructed to inform the user the expected probability that the season would be below, near, or above the historical average, for both temperature or precipitation.

Although the new system was only recently available, Changnon et al. found strong resistance to the new predictions among the participants in the survey. Most planners considered the new format hard to understand and a major hindrance to adoption. Many were also skeptical of the reliability of longer-term predictions and responded that a short scientific explanation following the forecasts would increase the likelihood of acceptance. Interviewed decision-makers were also more interested in different types of information, such as "climatic profiles" announcing other conditions that could be associated with each season, and "analogs", a summary of other years with similar conditions. The authors concluded that:

> "This study revealed a gap between existing nonusage of forecasts and the potentially high value of usage. It appears that an outreach effort is needed to develop usage of climate forecasts in the utility industry and to realize the considerable potential value of forecasts for several applications within the utilities" (p. 719).

The observations by Changnon et al. remain largely valid today, despite a growing acknowledgment of the importance of climate for utilities. In the aforementioned workshop organized by EPRI (2005), load specialists have agreed that demographic and economic variables, such as fuel prices, population and economic growth, play a larger role in medium and long-term forecasts, but they also observed that climate forecasts can become increasingly important, especially with climate change taking place:

> "As regional climate forecasts improve, companies may benefit by considering climate change in their load forecasting activities. Residential energy consumption, for example, is likely to be more sensitive to climate change than commercial or industrial demand. Consumers likely would respond to higher temperatures by

> installing additional air conditioning capacity or by extending operation of existing units. This could considerably increase summer demand for electricity, especially in temperate areas where the current penetration level of air conditioning is relatively low. Some companies may experience a shift from winter to summer peak demand as a result. In addition, climate change may result in shifting demographic and demand patterns if people begin to relocate in response." (p. 2)

In the case study reported here, climate forecasts are not currently used by the electric utility in statistical models, although forecasts are available for the region that is served by its power grid. Climate can have a particularly strong impact on energy use in the Southwest, because a large part of electricity consumption is dedicated to cooling and, to a lesser degree, heating. It is, therefore, in the best interest of an organization to have reliable predictions of the climatic conditions of a future season. We develop a simple medium-term model to assess the feasibility of incorporating that type of information in planning and compare its performance with an alternative model employing actual values for the climate variables used in forecasting. Details of this procedure are given in the following chapter.

CHAPTER 3

Methodology and Estimation

3.1 Data Description

Electricity providers are still protective of their data on system load. Not only is such information considered to be of strategic value, but it also poses privacy issues for consumers. The empirical part of this research was made possible by cooperation with the Arizona Electric Power Cooperative (AEPCO), which provided data for one of its member electricity providers. For purposes of confidentiality, the name of the area served is not revealed.

AEPCO is a not-for-profit cooperative originally founded in 1961 to provide electricity to rural southeastern Arizona. It is now composed of seven member cooperatives in Arizona and California. A restructuring of its organization took place in 1999, as part of the deregulation of the energy sector. Prior to this event, AEPCO worked monolithically with both generation and transmission of electricity. With deregulation, the cooperative split its functions into three distinct headquarters: one for services, one for transmission, and the last one for power generation (AEPCO, 2008). Currently most of the energy is generated by combustion of coal and, to a lesser extent, natural gas.

The load data consist of hourly measurements of total electricity provided in a certain member-specific area. The period covered starts in January 1993 and ends in December 2006, with no gaps.

Short-term electricity load is known to be highly sensitive to weather condi-

Table 3.1: Summary Statistics

Variable	Unit	Mean	Std Dev	Minimum	Maximum
load	MW	40.10	19.11	10.40	144.28
temp	°C	20.25	10.00	-7.80	45.80
humi	%	42.62	24.93	2.50	100.00
vapor	KPas	1.81	1.55	0.00	9.50
solar	MJ/m^2	0.86	1.15	0.00	4.29
prcp	mm	0.03	0.41	0.00	42.00
wind	m/s	2.42	1.46	0.20	12.90
holiday	dummy	0.001	0.033	0	1

T = 122,639

tions (Willis, 2002). Individuals tend to respond quickly to changes in temperature, humidity, and luminosity conditions, especially when those changes are atypical. As an illustration, an unusually hot day may be expected to observe more use of electricity for cooling, but the occurrence of a mid-afternoon rain storm can reduce usage drastically. Similarly, households will turn on their lights and consume more energy in a cloudy day, or turn off air conditioners in a windy evening.

We retrieved weather data for our specific region from AZMET, a University of Arizona meteorological network that measures a large number of weather-related variables at hourly intervals. Among the 28 areas covered by AZMET is the one used in this study. The variables used in this study, shown with summary statistics in Table 3.1, include air temperature (temp), relative humidity (humi), vapor pressure deficit (vapor), solar radiation (solar), precipitation (prcp), and wind speed (wind). Dew point temperature might have been useful, but unfortunately it is only available from 2003 on, and so is not part of the data we used.

For visual inspection, we plot electricity load in different aggregation levels. First, as a broad picture, we present demand aggregated at the monthly interval. In Figure 3.1, we can observe a clear seasonal pattern: the annual peak of electricity usage occurs in the summer, whereas a second, much smaller peak is seen in the winter period. Demand of energy for cooling, and heating to a lesser extent, are the driving forces behind this recurrent phenomenon. In addition to the annual seasonality, it is evident that the time series is non-stationary across all periods. In other words, there is an upward trend in electricity usage for this area that does not revert to the long-term mean.

More insight is gained when we visualize the evolution of hourly load during a week. In Figure 3.2, average load and temperature are plotted for each hour of the day, for each day of the week, during the summer of 2005. Put in a different way, each point in the graph represents the mean of several values of the same period in time. This is a simple way of displaying the recurring fluctuations that occur during a week. In the case of the summer, a strong relationship between load and temperature can be seen. In fact, the two variables appear to move closely together throughout the period. Additionally, note that the shape of the load curve is unimodal, with a peak occurring in the afternoon (around 7:00). This spike in demand can be linked to individuals returning to their houses and turning on electric devices, including air conditioners. Although it cannot be easily seen, consumption of electricity tends to decrease during the weekend and on Fridays, but even between those days the reduction is different.

Figure 3.1: Monthly Load for 1993 to 2006

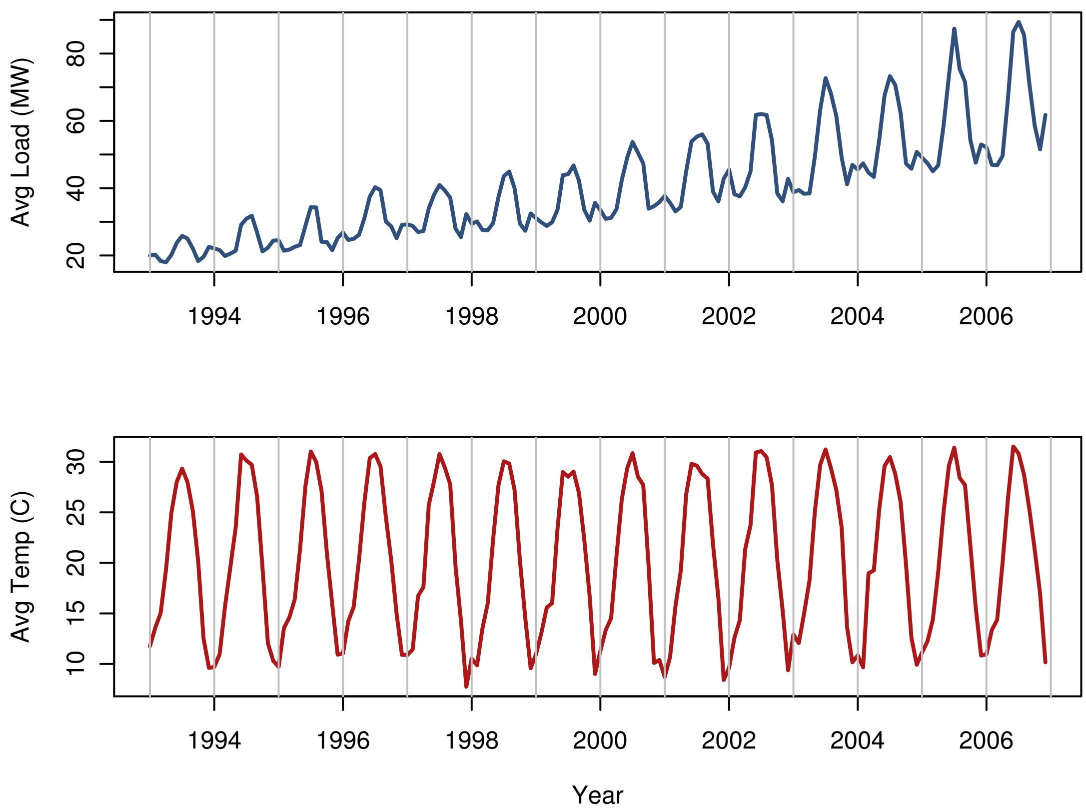
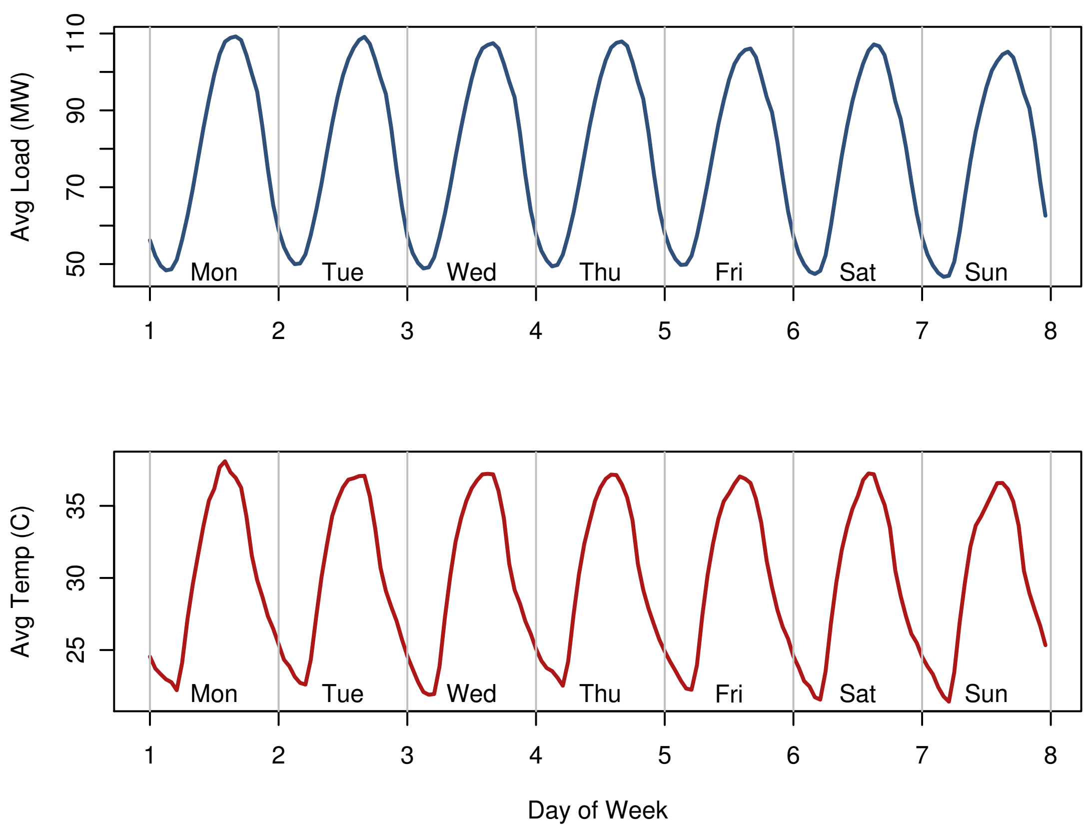


Figure 3.2: Average of Weekly Load for Summer 2005



A similar plot for the winter of 2005 is shown in Figure 3.3. Again, we observe a daily pattern, but this time the load curve is bimodal, with one peak in the morning and another one, slightly more accentuated, in the evening. Those correspond to hours when the temperature is low, but individuals are home and active, demanding energy for heating. It is worth reminding the reader that these plots are averages of several days, while in reality the relationship between load and temperature in the winter is not as straightforward as it may seem. The distinction between weekdays and weekends is now even more evident, and again each day appears to have a profile of its own. A last feature of the two previous figures is the fact that at any given point in time, there is a base level of electricity that continues to be used independent of the daily fluctuation. This base level itself is increasing across the years for the region of study.

Finally, in Figure 3.4 we present a plot of average daily temperature against average daily load, for the entire year of 2005. Consistent with the literature (Willis, 2002), temperature appears to be related to load, but not in a linear fashion. Weron (2006) suggests that for California this relationship can be thought as a “hockey stick”, in which load is insensitive to temperature up to a certain level, after which the relationship can be captured in a linear fashion. This is, in other words, a piecewise representation. In our dataset, the interaction of temperature and load is well depicted by a quadratic function. As can be observed in the figure, below and above a certain level, a change in temperature is associated with increasingly more use of electricity. In our case, the neutral point for temperature (the minimum of the quadratic curve) is approximately 15°C.

Figure 3.3: Average of Weekly Load for Winter 2005

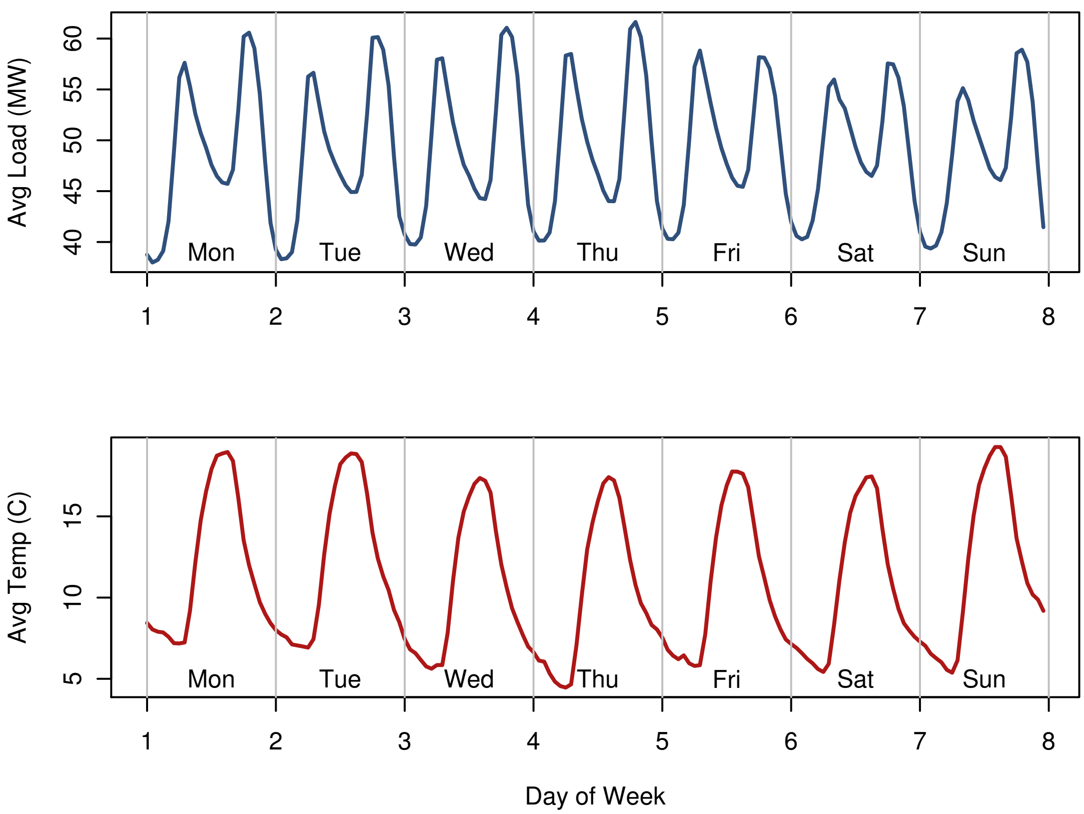
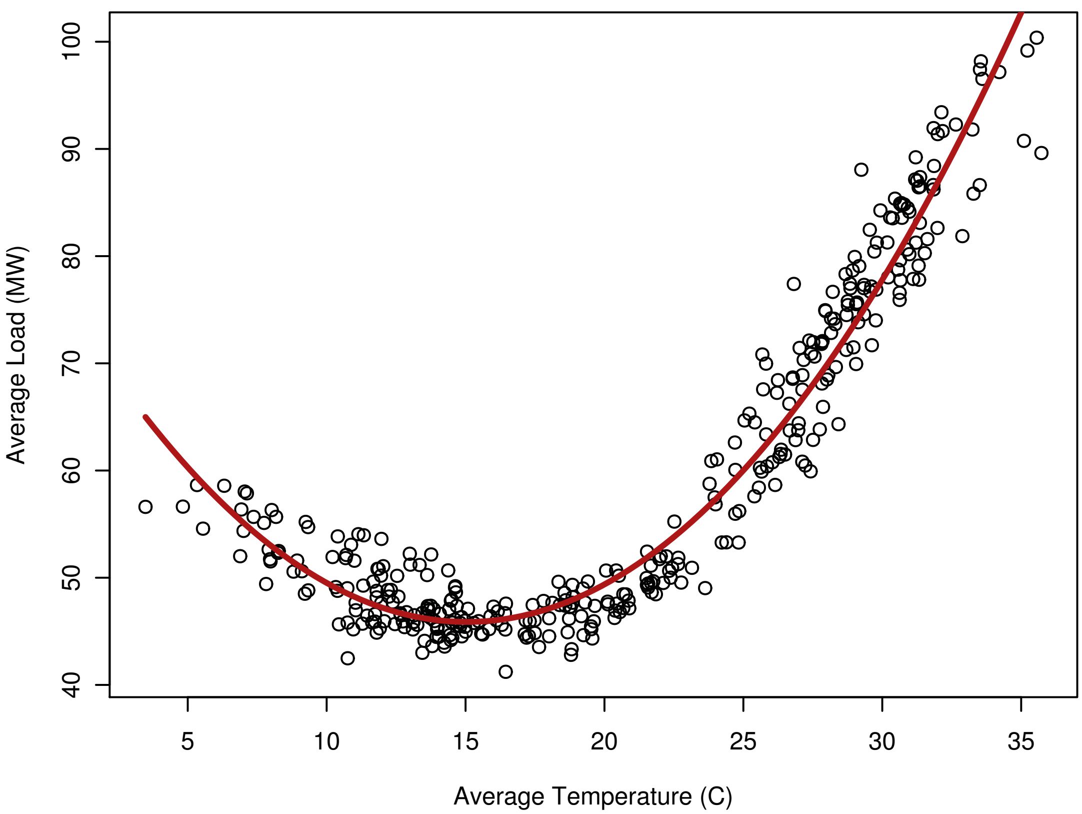


Figure 3.4: Daily Average Temperature and Daily Average Load for 2005



General inspection of the data appears to be in line with findings of other studies. Demand for electricity appears to be sensitive to weather conditions, especially temperature. Also, seasonality is present in the series at at least three levels: a) an annual pattern, including a peak in the summer and another one in the winter; b) a daily periodicity that changes from unimodal in the summer to bimodal in the winter; c) and a weekly "seasonality", in which weekdays are different from weekends. Finally, even visual inspection reveals that the process is nonstationary, with an upward trend. The literature points to population migration, economic growth, and technology, as leading factors, but climate change is likely to intensify the nonstationary process even further.

3.2 Forecast Evaluation

One of the main goals of this research is to develop models that can make good forecasts. These are desirable because they can help planning decisions, reduce costs, and increase economic efficiency. In considering how forecasts can be evaluated, we identify two points. First, we present measures of how close forecasts are to actual values. Second, we discuss the costs of making and improving the forecasts.

The issue of how to measure the goodness of forecasts is a perennial question in the statistical literature, and a vast number of methods have been proposed to tackle the problem. Despite this diversity of opinions, most agree that forecasts must be evaluated by their statistical properties. Defining the error as the difference between the predicted value and the actual value, we are interested in a measure that summarizes the error of various predictions, not a single prediction.

More formally, the expected value of the error, $E(\varepsilon)$, measures how wrong an estimator is on average. This term is also known as bias, and its reciprocal is the accuracy of an estimator. In the second moment, the variance of the error, $V(\varepsilon)$, can be thought of as a measure of the dispersion of the error, with its reciprocal giving the precision. Aside from these summary statistics, an important consideration is the weights associated with positive and negative discrepancies. While it is customary to assume that over- and underpredictions have equally severe consequences, this need not be the case. With electricity generation, the decision of underproducing can cause black-outs, while overproducing would be inefficient.

A commonly used measure of forecast error is the Root Mean Squared Error (RMSE). It is given by the square root of the mean squared error (MSE), a closely related measure of error. The formula is presented below:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{Y}_t)^2} \tag{3.1}$$

where Y_t is the actual observation for period t, $\hat{Y}_t$ is the corresponding predicted value, and T is the total number of periods. There are a few features that are worth discussing about the RMSE. One attractive point over MSE, for example, is that the values are given in the same units as the variable of interest, thus making is easier to visualize the dispersion of the errors. Another characteristic is that, because of squaring, larger values are given more weight than smaller values. Finally and also because of squaring, negative and positive errors are treated identically. That is, an underprediction and overprediction of the same magnitude are considered to be the same error in this formula.

In the electricity industry, the most widely employed measure of error is the Mean Absolute Percentage Error (MAPE), which can be seen in the formula below:

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{|Y_t - \hat{Y}_t|}{Y_t} \tag{3.2}$$

Like the previous measure, MAPE also exhibits a number of properties. First, its values are given in percentage (unitless), therefore making it comparable across different situations. As in RMSE, there is no distinction between positive and negative errors, all are considered to have equal weights in absolute terms. Unlike the previous error rate, the impact of small or large deviations is always relative to the actual value, capturing the idea that a forecast error that is severe for a small utility may be less important for a large company. One point to notice is that the error is divided by the actual value, so this would be problematic in cases where the denominator is close to zero. From Table 3.1, we know that this is not a problem for our data, since all load values are well above zero. In evaluating our models we will report both RMSE and MAPE, but we should note that both measures are largely equivalent: in most cases an increase or decrease in the forecast error will lead to a change in both scores in the same direction, albeit at a different rate.

An additional measure of the performance of predictions comes from the meteorological literature. Forecast skill, as it is called, is "a statistical evaluation of the accuracy of forecasts or effectiveness of detection techniques" (American Meteorological Society, 2008). The skill score (SS) is given by the formula:

$$SS = 1 - (RMSE_f / RMSE_r) \tag{3.3}$$

where $RMSE_f$ is the root mean squared error of the forecasts being evaluated, while $RMSE_r$ is the corresponding measure of error for values that are taken as a reference. A positive value for SS indicates that the model offers some skill over the reference technique. A common practice is to use "persistence forecasts" as the reference, known as "naive forecasts" in the economic literature, in which observations of similar periods are used as the predictions for the period of interest. For example, one can assume that tomorrow's temperature will be the same that was recorded today, or that next summer will have the same climate characteristics as the current one.

In addition to errors in forecasts, there are also costs due to time and resources allocated to estimating statistical models. In economic terms, the marginal gains from spending one extra hour in modeling should be greater or equal to the marginal costs of doing so. Even at a more fundamental level, ultimately there is a time constraint within which the decision of how much electricity to produce must be taken, otherwise black-outs can occur. To address this issue, we will include an extra benchmark measure: time of estimation, in minutes. This is merely the time that an estimation takes to converge to a solution. Although this does not take into account the time spent in setting up the parametric form of the model, it can be important in determining the feasibility of generating a large number of alternative specifications of models.

3.3 Modeling Considerations

In this section, we derive the models that will be used for forecasting system load. The first step in time series modeling is to ensure the data conform to the assumptions required by the methodology. One of the most important of them is the requirement of stationarity. Seasonality, which is technically a violation of that assumption, deserves special attention in our model. Each of the topics are discussed in turn.

3.3.1 Stationarity

Traditional time series techniques require that the input series be stationary. Intuitively, stationarity implies that in the long-term the series does not wander away indefinitely from its mean. When the series is nonstationary, the variance is said to be explosive, and predictions are not reliable.

Granger and Newbold (1974) and Phillips (1986) showed that tests based on conventional critical t values were overly optimistic (Greene, 2003). Dickey and Fuller (1979) studied the distribution of a series under different unit processes. More specifically, they simulated critical values for random walk, random walk with a drift, and a single trend. We use the augmented version of the Dickey-Fuller test, which removes the structural effect prior to testing. Its formula is given by:

$$DF_{\tau} = \frac{\hat{\gamma} - 1}{StdErr(\hat{\gamma})} \tag{3.4}$$

The load series was detrended by taking the first difference of the logged hourly load. In the ADF tests, presented in Table 3.2, the null hypothesis is that a unit root is present in the data, that is to say, the series is nonstationary. Because of the complexity of our dataset, we performed tests out to 168 hours, that is, one week. In all cases, the null hypothesis can be rejected at any conventional statistical levels of confidence.

Table 3.2: Augmented Dickey-Fuller Tests

Type	Lag	τ	$P < \tau$
Zero Mean	1	-177.36	0.0001
	2	-156.42	0.0001
	3	-169.53	0.0001
	24	-95.77	< .0001
	168	-34.34	< .0001
Single Mean	1	-177.36	0.0001
	2	-156.42	0.0001
	3	-169.53	0.0001
	24	-95.77	< .0001
	168	-34.34	< .0001
Trend	1	-177.36	0.0001
	2	-156.42	0.0001
	3	-169.53	0.0001
	24	-95.77	< .0001
	168	-34.34	< .0001

T = 122,639

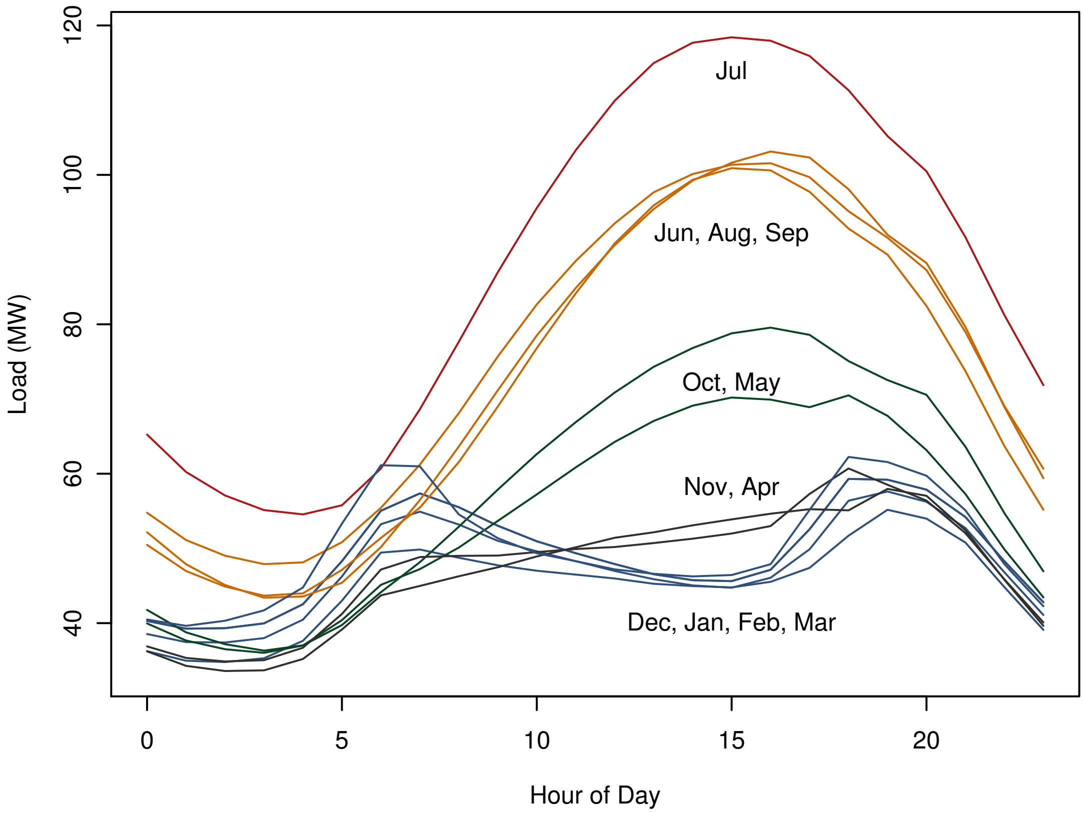
3.3.2 Seasonality

As mentioned before, even a visual inspection of the data is sufficient to detect systematic patterns that appear to be associated with daily, weekly, and annual cycles. In the statistical literature, these phenomena are commonly referred to as seasonality, or periodicity. A broad definition has been proposed by Hylleberg (1992):

> "Seasonality is the systematic, although not necessarily regular, intra-year movement caused by the changes of the weather, the calendar, and timing of decisions, directly or indirectly through the production and consumption decisions made by the agents of the economy. These decisions are influenced by endowments, the expectations and preferences of the agents, and the production techniques available in the economy." (p. 4)

One interesting characteristic of our data, and demand for electricity in general, is that the daily periodicity changes over time in a systematic manner. Recall from Figures 3.2 and 3.3, the presence of a unimodal curve during the summer, and a bimodal shape in the winter. Figure 3.5 reveals more details about the different shapes of the load curve, month by month. We can loosely cluster the curves into five groups: a) a clearly bimodal interval, coinciding with the winter months (December to March); b) a transition phase, in which the curve is flat, but the two peaks can still be observed (November and April); c) a first, low unimodal period (October and May); d) a second, intermediate unimodal period (June, August, and September); e) a third unimodal period, which is the highest across the entire year (July). Note that not only does the shape

Figure 3.5: Monthly-Averaged Hourly Load for 2005



change over the months, but the range of the curves is significantly different. The minimum point of July, for example, is larger than the maximum of most winter months.

This peculiarity poses a challenge to modeling for a number of reasons. First, it is not completely captured by weather variables, because there are other fluctuations in the behavior of individuals, such as business cycles and migration, that are exogenous to our model. Thus, the issue cannot be ignored and some form of seasonal modeling is necessary if our goal is forecasting. Second, to the best of our knowledge a model with changing periodicity has not received attention from the literature and there is no established method of dealing with it.

A crude way of capturing this changing seasonality would be to create dummy variables for specific periods of the data, for instance, one for each month. However, this approach can be problematic, because the transition from one season to another is not known in advance and varies from year to year. Additionally, even if that problem is solved, using dummy variables would result in abrupt changes in arbitrary periods, say, from March 31 to April 1 in Figure 3.5, while we expect the transition to be continuous and perhaps smooth.

Another option would be to employ some method of seasonal adjustment to filter the data, then proceed to modeling. Many economic studies rely on seasonally-adjusted series, in the hope of removing the irrelevant noise prior to modeling. Examples of this procedure include data from the Census Bureau (using the X-11 or X-12 method) and other macroeconomic datasets. However, in the last decade several researchers have found strong reasons against this strategy. First, it is usually the case that the econometrician producing the final model will simply use the transformed series without knowing exactly how it was generated, implicitly accepting assumptions that the original modeler made. Second, it is not clear how seasonal adjustments will alter the properties of the series and even introduce additional noise.

Sims (1992), among others, endorses this position and cites an early article by Roberts (1978), where the author observes that "surely the route to better scientific understanding is to incorporate the seasonality directly into multivariate models that are formulated in terms of unadjusted data so the source, transmission, and effects of seasonal variations can be better understood". Sims also notes that in other fields, such as meteorology, scientists directly model phenomena with high degree of seasonality without undertaking adjustment as a separate step.

One innovation of this study is to propose a way of accounting for a multi-dimensional seasonality in a time series dataset, that can be estimated in one single pass. We build upon the work of Aradhyula and Ergun (2004) and Aradhyula and Tronstad (2006), who developed a general approach for dealing with seasonality. As in those studies, we estimate a time series model together with high order polynomials. Different from prior studies, we generalize the framework to handle periodicity in more than one dimension.

Modeling seasonality with a high-order polynomial is not entirely a novel technique and authors have warned about its pitfalls in the literature. Sims (1992) notes, for example that: "there are some drawbacks to using polynomial-seasonal interactions rather than a corresponding number of trigonometric terms for removing seasonal power. If [the number of coefficients] is large, the polynomial terms may be highly collinear". The issue of multicollinearity, as will be shown, can be easily solved. The greatest advantage of favoring polynomials over trigonometric functions, however, is the minimal user intervention necessary to modeling. In the case of the former approach, the forecasts would have to employ spectral analysis which requires not only training and experience, but usually a few trials and errors.

In a general form, given a variable s that cycles with certain periodicity and takes values between 0 and 1, we can create a polynomial S of order p by:

$$S = \alpha_1 s_t + \alpha_2 s_t^2 + \ldots + \alpha_p s_t^p = \sum_{i=1}^{p} \alpha_i s_t^i \tag{3.5}$$

where α is the variable coefficient.

We then make the function continuous by requiring that the starting and ending points coincide:

$$S|_{s=0} = S|_{s=1}$$
$$\sum_{i=1}^{q} \alpha_i = 0$$

Smoothness can be enforced as well, by ensuring that:

$$\sum_{i=2}^{p} i\alpha_i = 0$$

The polynomial proposed so far is extremely flexible and is able to adjust itself to highly nonlinear patterns, given sufficient order terms. Contrasted with spectral analysis, it requires knowledge of only the period of the variable, as opposed to finding a combination of sine and cosine functions, with different periods, amplitude, and shift. A problem that can surface when high order polynomials are used is multicollinearity, since each additional term is a (nonlinear) combination of the previous one. However, this can be easily solved by creating orthogonal polynomials instead, in which the inner product of any two vector terms is zero: $\langle s_i, s_j \rangle = 0, \forall i \neq j$.

In other words, all terms are created so that they are linearly independent from each other. Orthogonal polynomials have a rich literature of their own, and several specifications have been proposed, including the Legendre, Chebyshev, Jacobi, and Leguerre forms, to cite a few (Hochstrasser, 1965). Most statistical programs have canned procedures to produce such polynomials automatically. The Chebyshev polynomials of the first kind, in particular, exhibit the appealing property of being bound between a fixed interval, making them the most appropriate for our model of periodicity. They are given by the following recursive definition:

$$T_0(x) = 1$$

$$T_1(x) = x$$

$$T_i(x) = 2xT_{i-1}(x) - T_{i-2}(x)$$

where i is the order of the polynomial. For convenience, the full expansion of the definition up to the ninth order is presented in Appendix B.

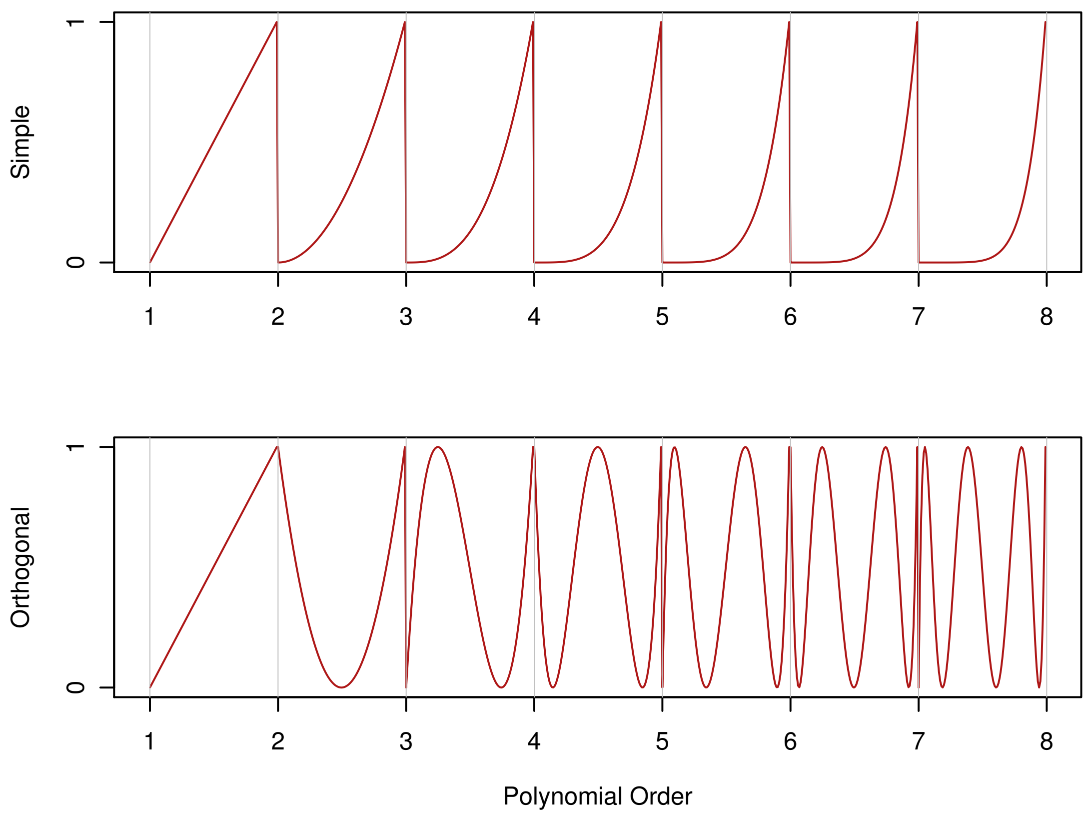
For the sake of illustration, a graphical comparison of simple and Chebyshev polynomials are given in Figure 3.6. The orthogonal polynomial, besides the guarantee of not possessing multicollinearity, appears to capture a greater wealth of variability than the simple form for the same interval. In fact, the Chebyshev polynomials of the first kind satisfy the following trigonometric identity (Hochstrasser, 1965), thereby putting them close to traditional spectral analysis[1]:

$$T_i(cos(i\theta)) = cos(i\theta)$$

In a practical sense, the described polynomial model should be able to compete directly with its spectral counterpart. However, in their traditional form both suffer from a severe limitation: the seasonality is only allowed to affect the dependent variable in an additive, linear fashion. To see why we would want to waive this constraint, consider again the case of hourly electricity load plotted for all months of the year in Figure 3.5. Suppose we specify one season-

[1]In fact, the recursive definition of the polynomial can also rewritten in trigonometric form.

Figure 3.6: Comparison of Simple and Orthogonal Polynomials



ality form for the day (or month) of the year, and a separate one for the hour of the day. Then, because the model is additive, moving across one seasonality curve at a time will simply shift the other one up or down. However, we have seen that seasonality of the load shape itself changes over the year. There is, therefore, a periodic change in the shape of the seasonality that requires a more flexible model.

Such a model can be achieved in different ways. Perhaps the most strictly pure form is to estimate a model with time varying parameters, for instance:

$$y_t = \omega_{0,t} + \omega_{1,t} x_t + \varepsilon_t \tag{3.6}$$

$$\omega_{i,t} = \omega_{i,t-1} + \epsilon_{i,t-1} \tag{3.7}$$

where $\omega_{i,t}$ is a coefficient that is allowed to change across time (in this case, as a random walk process), and ε_t and $\epsilon_{i,t-1}$ are white noise. It is evident that estimating a model such as this would be a challenging problem, since most econometric software are designed for linear models. State space models can handle this type of estimation, at the price of added complexity (Durbin and Koopman, 2001).

An alternative way of generating the same effect is to include in a linear model all the interactions between the seasonality terms. For instance, given two vectors of orthogonal polynomials, one for the day of the year, D of order m, and another for the hour of the day, H with polynomial order n, we can produce the interactions by simple matrix algebra:

$$\begin{aligned}\mathbf{1}'D'H\mathbf{1} &= \begin{bmatrix}1 & 1 & 1 & \cdots & 1\end{bmatrix}\begin{bmatrix}1 & h & h^2 & \cdots & h^n \\ d & dh & dh^2 & \cdots & dh^n \\ d^2 & d^2h & d^2h^2 & \cdots & d^2h^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d^m & d^mh & d^mh^2 & \cdots & d^mh^n\end{bmatrix}\begin{bmatrix}1 \\ 1 \\ 1 \\ \vdots \\ 1\end{bmatrix} \\ &= \sum_{i=0}^{m}\sum_{j=0}^{n} d^i h^j\end{aligned}$$

where $\mathbf{1}$ is a column vector of 1s, the superscripts for d and h denote the order of the polynomial, possibly in an orthogonal form, and the matrix multiplication is set up to be conformable.

Naturally, when the order of the seasonality for the day of year, m, is zero, the model reduces to the hourly seasonality; likewise, when n is zero, only the annual periodicity is being modeled.

This specification is much less parsimonious than the model with varying parameters, costing more degrees of freedom. However, in our case this sacrifice should not pose a problem, since the available dataset is sufficiently large. The advantage of being able to keep the familiar linear model is valuable, since our goal is to propose a model that can be easily implemented by an electric utility.

The multiple seasonality model can be more easily understood by means of a graph. To that end, we present two graphs. The first, for comparison, is a three dimensional plot of actual values across hour of day and day of year (Figure 3.7). The second is a plot of predicted values estimated from a seasonality for daily load multiplied by another for the annual cycle (Figure 3.8). The graph generated from the polynomial model possesses the pattern described above: the shape of the daily load curve changes across the year. During the winter (beginning and end of the year), demand for electricity follows a bimodal form, with peaks in the morning and evening. In the summer (middle of the year), the load changes into a unimodal form, with much higher minimum and maximum values. Thus, the innovation of this modeling technique is to capture the transition between these two seasons in a continuous and smooth manner. Note that in the intermediate region, the shape of the load curve is much flatter.

Although, the model appears to capture the two seasonalities as expected, it remains to be seen how it will perform in practice. This is left for the next chapter, where load forecasts are made using the specification discussed above.

3.4 Model Variables

Having discussed the transformations and adjustments to the data, we now proceed to specifying the functional forms of the variables in the model. We also include a brief discussion on the expectation for the sign of each variable's coefficient in the final estimation. The reasoning behind this is to avoid mere experimentation with the data and potential overfitting for our sample.

3.4.1 Dependent Variable

Hourly Load The dependent variable of our model is electricity load measured in an hourly interval for a specific region in Arizona. Because of the observed exponential growth in demand, we transform the variable into its natural log form. To ensure stationary, we can either take a first difference or

Figure 3.7: 3D-Plot of Load by Hour of Day and Day of Year

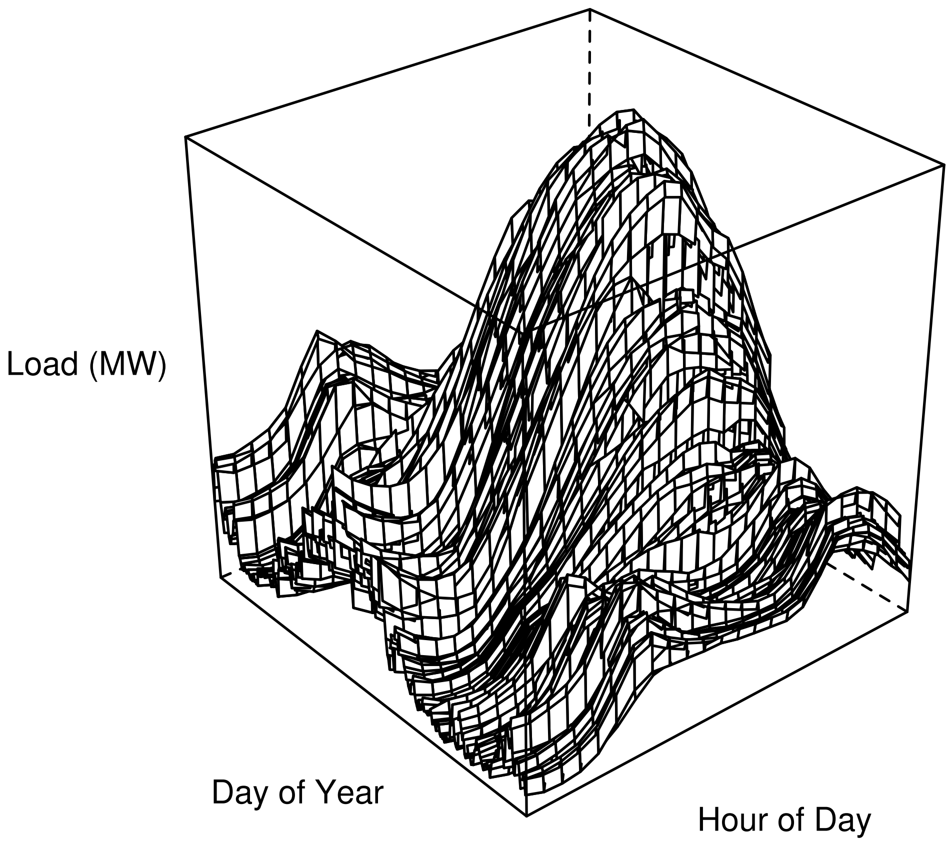
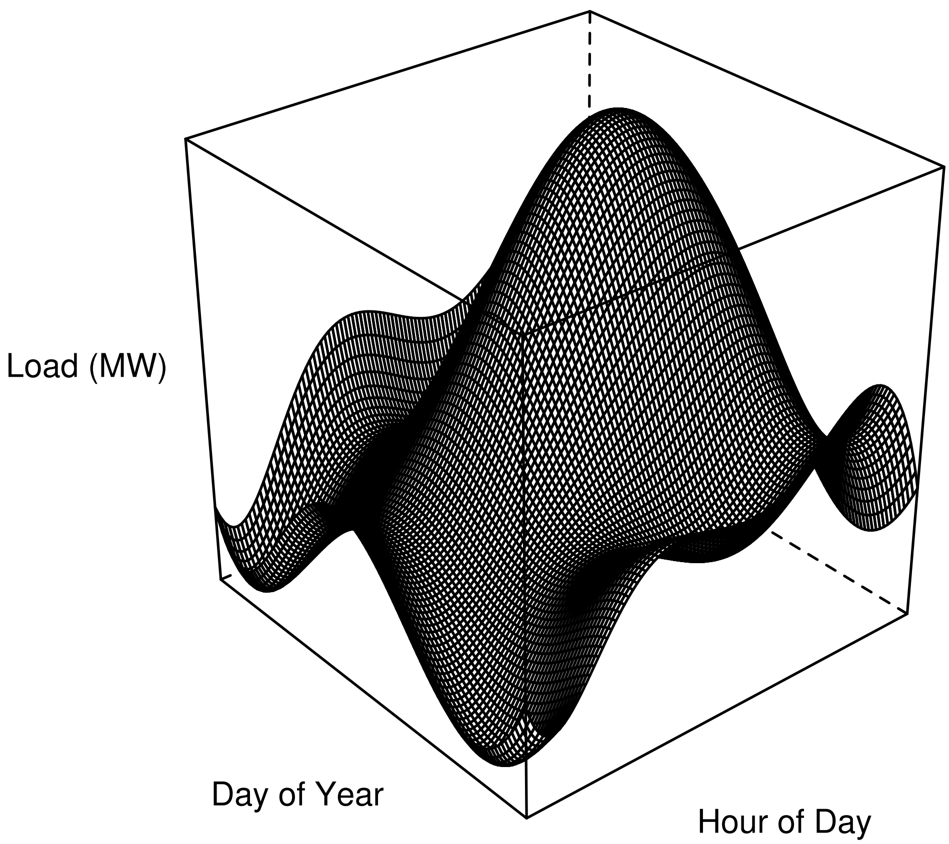


Figure 3.8: 3D-Graph of Polynomial Model by Hour of Day and Day of Year



include a linear trend variable in our model. After simple experimentation, we opted for the second, as it would give better prediction results. Once the model is estimated and forecasts are made, we transform the values back via anti-log. Note, however, that by Jensen's Inequality[2], this transformation is not without its problems. Nevertheless, the introduced bias is negligible and is not enough to justify a correction. This decision is supported by numerous other studies that have taken a similar approach (Brocklebank and Dickey, 2003).

3.4.2 Explanatory Variables

Trend As explained above, we included a term for trend due to the increasing usage of electricity for the region of study. Because this growth is steady, we expect to find a robust positive coefficient.

Day of the Week From visual inspection we have learned that Friday, Saturday, and Sunday have consistently lower demand for electricity compared to weekdays. The difference is not as notable within the other weekdays, but because weekly variability appears to follow a pattern, we decided to create a dummy for each day of the week, leaving Monday as the baseline. Our expectation is that the coefficients for the weekend are negative, while the others can move in either direction.

Holiday Demand for electricity tends to be affected during special events, such as holidays. We therefore collected all official holidays for the state of

[2] That is, the expected value of a function is not equal in general to the function of the expected value. For a concave function $f(\cdot)$, such as log, $E(f(x)) \leq f(E(x))$ and there is a potential for downward bias.

Arizona from the Human Resources Division. Unfortunately, data was not available prior to 2004, so we reconstructed the holidays from the Arizona Revised Statutes (ARS), Sections 1-301[3]. For each year there are ten paid holidays, that are moved back or forward so that they do not coincide with weekends. Thus, holidays imply non-business days and with industry and part of the commerce shut down we would expect a decrease in the usage of electricity. However, certain holidays may also be associated with more demand by residential customers, as can be the case of New Year's day and sports events. Because our data include all classes, we cannot tell which effect will be larger. We still expect the coefficient for holiday to be negative.

Temperature For this particular region, the relationship between load and temperature appears to follow a quadratic pattern. Other studies have reported that cubic and piecewise linear modeling offered a more appropriate fit (Weron, 2006). We favor the quadratic form due to its simplicity. Temperature and its squared value are both included in the model, with the expectation that the quadratic coefficient will be negative. In other words, we believe that temperature levels that are far from a certain neutral point lead to more consumption of electricity for cooling or heating. As previously discussed, the insensitive level found in exploratory analysis was around 15°C.

Humidity Humidity plays a secondary yet important role in the influence of weather on electricity load. More specifically, the sensation caused by humidity can compensate for high or low temperature, therefore affecting use of cooling or heating equipment. Because this variable is not so well studied in the liter-

[3]The source code of the program for generating the holidays is available in Appendix C.

ature, we try different specifications in the model: a simple linear coefficient, a quadratic relationship as useful for temperature, and the interaction term between temperature and humidity.

Precipitation Because the region under study naturally receives small amount of rain, the distribution of this variable is highly skewed towards zero. The fact that the unit of analysis is measured at hourly intervals also contributes to this sparseness. Hence, we decided to recode the variable as a dummy: precipitation takes the value of 1 if there was any rain during that period, and 0 otherwise. Similar to humidity, we expect less demand for electricity during rainy intervals.

Wind Speed Our final variable, wind speed, is assumed to work analogously to the previous two: rapid air movement generates a chilling sensation that can decrease the need for electric cooling, or conversely, can augment the use of heating during the winter. To support these claims, the coefficient of this variable should appear as negative in our model.

3.5 Estimation Process

Once we have decided on the appropriate model, we use data from 1993 to 2005 to forecast next-day electricity load for the entire year of 2006, in a rolling sample scheme. More specifically, on each new day we exclude the oldest 24 hours of our sample and include the actual values of the 24 hours that have passed since the last estimation. We then re-estimate the model and produce new forecasts for the next 24 hours. This process is repeated until the reserved sample is over.

The rationale behind this organization is two-fold. First, it attempts to mimic how the forecasts would be generated in a utility, in a day-to-day routine. The difference is that we already have the entire year of 2006 available to our model, so the procedure can be fully automated, whereas in practice an individual would have to feed in the new values. Second is the idea of updating the model as time passes. This point is of key importance, because, as we have seen, usage of electricity is very sensitive to a number of factors, including current weather conditions and even recent consumption, due to inertia.

Although old load values are discarded in the rolling sample setup, technically that does not necessarily have to be the case. It is interesting to consider why this tradition started. One of the fundamental properties of time series is that recent observations have greater influence on current values, while distant points should have an ever decreasing impact. With a large sample, it is likely that the initial 24 hours would have a negligible effect on the model and on forecasts. From a more practical perspective, the oldest observations are taken out so that the sample size remains constant across all estimations, thus facilitating comparisons of statistics that are sensitive to the number of observations, such as the Akaike Information Criterion (AIC) or the Schwartz Bayesian Criterion (SBC), and even the standard errors of coefficients.

The predictions, still in natural log form, are transformed back and stored, one-by-one, in a separate dataset. Once the entire process is finished, the forecasts are merged with their corresponding actual values and residuals are calculated. From the residuals, the measures of prediction error are created, in this case, MAPE and RMSE. Note that despite the fact that these statistics are the common way of evaluating a model, it is also valuable to generate a scatterplot of the residuals, which can be inspected for systematic patterns.

CHAPTER 4

Analysis of Results

4.1 Short-term Forecasts

To be able to compare performance, we ran three different specifications of the model: a full version, with all variables, one with only the temperature variables, and the last one with no weather variables. In all cases we modeled seasonality as discussed in the previous chapter, that is, by including interaction terms for orthogonal polynomials. The polynomial order for the day of the year was 10, while the term for hour of the day was 9. Estimation results are presented in Table 4.1. For the sake of clarity and due to space constraints, the seasonality coefficients have been omitted from that table, but are included in Appendix A. We also remind the reader that the dependent variable was the log of hourly load, therefore the coefficient and standard errors cannot be easily interpreted in their magnitude.

As expected, the model was able to capture the positive trend in the load for our region. Modeling the trend in this explicit form gave better results than the usual approach of taking the first difference.

Besides the trend and the polynomials for seasonality, we also included dummy variables for each day of the week, taking Monday as the baseline. As expected, Friday, Saturday, and Sundays tend to have lower load, while the effect for the other days is not so clear.

The variable for holiday also had a negative coefficient throughout. But, as

Table 4.1: Estimation Results

Variable	Full Model	Temp Only	No Weather
constant	2.8686***	2.8739***	2.7756***
	(0.0194)	(0.0194)	(0.0486)
lag1	-0.7822***	-0.7878***	-0.8803***
	(0.0018)	(0.0018)	(0.0013)
lag24	-0.1810***	-0.1767***	-0.1070***
	(0.0018)	(0.0018)	(0.0013)
trend	0.00001***	0.00001***	0.00001***
	(0.000001)	(0.000001)	(0.000001)
tuesday	0.0023*	0.0022*	0.0030**
	(0.0009)	(0.0009)	(0.0011)
wednesday	0.0031**	0.0031*	0.0036*
	(0.0012)	(0.0012)	(0.0014)
thursday	0.0022	0.0020	0.0028
	(0.0013)	(0.0013)	(0.0016)
friday	-0.0030*	-0.0031*	-0.0019
	(0.0013)	(0.0013)	(0.0016)
saturday	-0.0123***	-0.0124***	-0.0100***
	(0.0012)	(0.0012)	(0.0014)
sunday	-0.0123***	-0.0123***	-0.0104***
	(0.0009)	(0.0009)	(0.0011)
holiday	-0.0038*	-0.0052**	-0.0030
	(0.0016)	(0.0017)	(0.0020)
temp	-0.0230***	-0.0229***	
	(0.0002)	(0.0002)	
temp2	0.0006***	0.0006***	
	(0.00001)	(0.00001)	
humi	0.0012***		
	(0.0001)		
humi2	0.00001***		
	(0.00001)		
prcp	0.0019		
	(0.0011)		
wind	-0.0009***		
	(0.0001)		
Forecasts for 2006			
MAPE	4.177	4.276	5.666
RMSE	3.897	4.100	5.989
SS	0.05	0.32	0.22
Time	8′23′′	8′04′′	8′01′′

Significance Levels: *.05 **.01 ***.001

discussed above, in practice forecasts for special days may require more information and possibly judgment from an expert. Although it is widely known that special events, such as holidays, tend to have different behavior in the load curve, and therefore should be treated with care, the relationship is not obvious. As previously discussed, some holidays have less use in electricity, while others will have higher demand.

The quadratic relationship between temperature and load was adequately captured by the model. As expected, the second coefficient was positive, indicating higher demand being associated with large deviations from a certain temperature value. Both temperature coefficients were statistically significant at the .001 level, suggesting a robust relationship throughout the sample.

Much to our surprise, humidity presented a similar pattern to that of temperature. To our knowledge, this relationship has not been reported in other studies of load forecasting. In this case, the pattern is the inverse of temperature: extremely high or low levels of humidity are associated with less demand for electricity. As temperature, this relationship seems to be robust and stable in all runs of the models.

It is also worth mentioning that we also tested an interaction term between temperature and humidity, in the expectation that the combination of both might have a specific effect. However, that relationship was not found and we opted for not reporting the estimation results in the table.

Precipitation, as a dummy variable, was expected to have a negative coefficient sign, but in the table the variable had the opposite sign and is not significant. Although the results appear to contradict this expectation, there are a few points to be made. First, as previously noted, the variable is highly skewed,

Table 4.2: Correlation Matrix for Explanatory Variables

	temp	temp2	humi	humi2	prcp	wind
temp	1.00					
temp2	-0.87	1.00				
humi	-0.05	0.24	1.00			
humi2	0.09	-0.22	-0.96	1.00		
prcp	-0.03	0.07	0.09	-0.15	1.00	
wind	-0.15	0.11	-0.01	0.04	-0.09	1.00

T = 122,639

with more than 98% of cases equal to zero. Second, it is not evident how consumers will respond to rain: although it may serve as a means of cooling, it might also drive individuals indoors and lead to more usage of electricity. We, unfortunately, cannot identify each case separately, but because the variable does seem to reduce the forecast errors, we decided to keep it in the model. Third, it can be seen in Table 4.2 that humidity and precipitation do not appear to suffer from a problem of multicollinearity: the correlation with precipitation is only about .09 for humidity and $-.15$ for humidity squared. In fact, apart from the squared terms, no coefficients are correlated at any alarming level.

Finally, wind speed met our prior expectations of having a negative relationship with load. Even with a high temperature, a windy day can bring about the sensation of chillness that can reduce the demand for electricity for cooling.

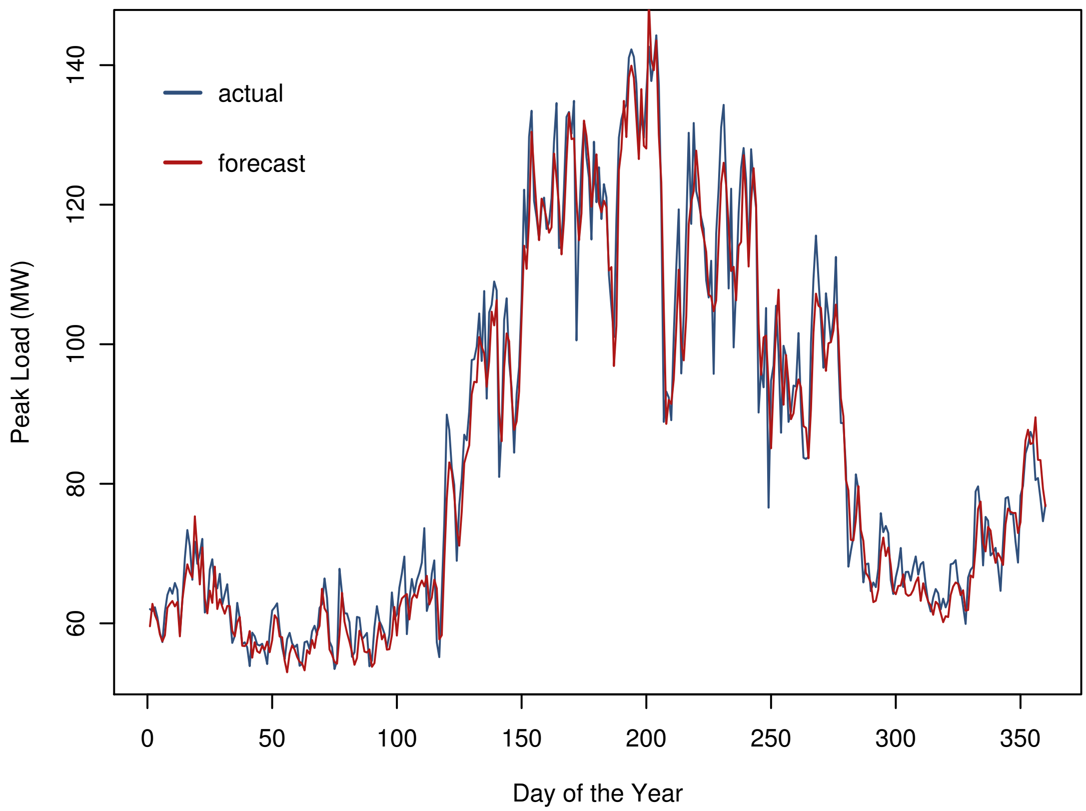
It is clear in comparing the alternative specifications, that weather holds information that is important for modeling and forecasting demand for electricity. Although our baseline model did not perform so well (MAPE of 5.67%), it is not far from what is reported in the literature and by AEPCO. Merely including temperature in the model is sufficient to reduce the error rate by almost 1.4%. The gains from additional variables, that is, humidity, precipitation, and wind speed, are not as accentuated (only .1%), but for a large utility may be enough to bring significant savings.

In terms of skill score (SS), we find that each of the specifications bring improvements over the previous one. In the case of the model with no weather, the reference is the RMSE of simple persistence forecasts. In other words, our simplest model brings an improvement in skill of .22 over merely assuming that the load for the next 24 hours will be the same as the previous 24 observations. The model with temperature alone possesses a skill score of .32 over the model with no weather. Finally, the model with all weather variables offers more skill over only using temperature of .05. Note that performance measured by skill level is in line with the results in terms of MAPE, as discussed in the previous paragraph.

The time of estimation did not vary significantly for each of the models. All of them finished generating the next-day hourly forecasts for 2006 in about 8 minutes. As expected, the models including more variables were slightly slower to generate forecasts, but the difference from best case to worst case is only 22 seconds. These measures include the time that is required to create the polynomial terms, the variable transformations, the model estimation and forecasts, and the calculation of MAPE and RMSE. Note that the main advantage advocated here is that a single model can be used for the entire year, as opposed to separate models for different seasons.

To further illustrate the performance of the forecasts, a plot of actual peak

Figure 4.1: Actual versus Predicted – Daily Peak

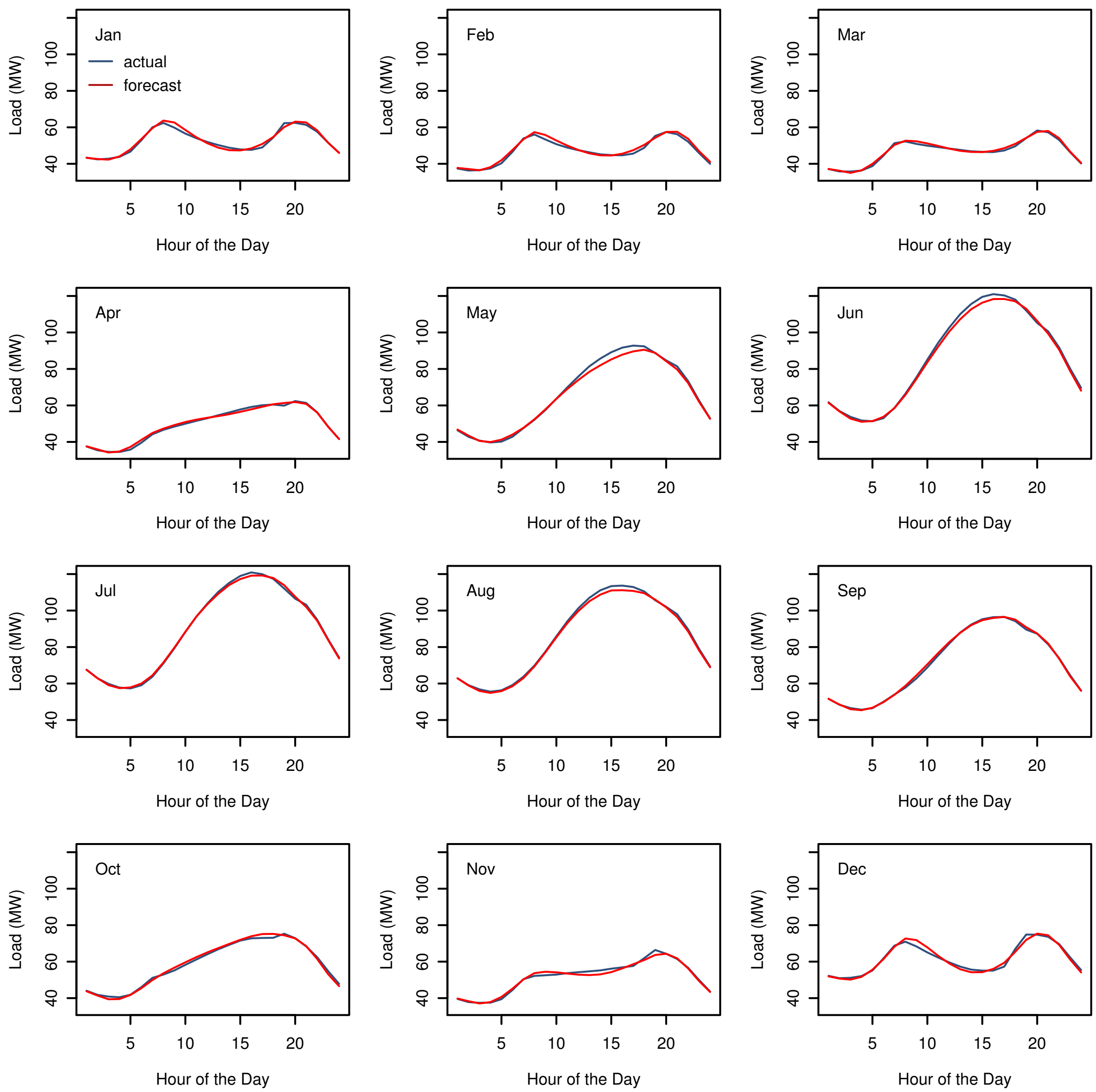


versus predicted values is presented in Figure 4.1. For all days of the out-of-sample period, we selected the daily maximum load and the equivalent value from the forecasts. This aggregation was done partly because the fine scale of one-hour interval would make the graph overly crowded, to the detriment of understanding. The peak was chosen because it constitutes the most interesting value for AEPCO (Thompson and Cathers, 2005). The figure indicates that the predicted values appear to be consistently close to the actual load throughout the year. The models performs relatively well even during the periods of transition between seasons, although initial underforecasts can be seen.

Finally, we plot actual and predicted values averaged by month in Figure

4.2. Each box in that figure presents hourly load averaged over every month. This series of plots should be comparable with what was observed in Figure 3.5, except that: a) these graphs cover the forecasting period; b) values for each month are now shown separately. These plots should provide the reader with evidence for the success of polynomials in capturing the change in seasonality that occurs across the year. The technique for dealing with seasonality advocated in this study is able to adjust itself to unimodal, bimodal, and intermediate shapes of the load curve in a smooth manner.

Figure 4.2: Actual versus Predicted – Monthly-Averaged Hourly Load



4.2 Seasonal Forecasts

To evaluate the value of seasonal forecasts, we model and estimate five specifications of monthly average load, using the same dataset as before. In addition, one year-ahead seasonal forecasts developed by the Climate Prediction Center (CPC) at NOAA are included, for both temperature and precipitation[1]. The variables employed in the models have their descriptive statistics shown in Table 4.3. Load, the dependent variable, is averaged at the monthly interval, yielding 133 periods. Actual temperature and precipitation are aggregated in a similar fashion, and at this level exploratory analysis did not find a nonlinear relationship, in contrast to the short-term models. The CPC forecasts are given in probability terciles, that is, the likelihood that a particular season will observe condition that is below, near, or above the historical average. Because our interest lies on measuring how much a unit increase in temperature or precipitation will affect electricity consumption, we transform these predictions by multiplying the terciles by the historical averages.

Again, to achieve stationarity the dependent variable is logged, and an explicit term for trend is included in the model, to control for the growth in demand that the region has been facing. Because we only have 13 years of data and lost a few when matching with the CPC forecasts, we refrain from reserving part of the dataset for out-of-sample prediction. Instead, we will evaluate the models by their in-sample properties.

The estimation results for all models are presented in Table 4.4. Model 1 only includes autoregressive terms and the trend variable. Each additional model expands the previous with more variables. Models 2 and 4 are estimated

[1]The CPC forecasts are available at `http://www.cpc.ncep.noaa.gov/`

Table 4.3: Summary Statistics for Seasonal Variables

Variable	Mean	Std Dev	Minimum	Maximum
load-avg	44.39	14.69	24.61	89.39
actual-temp	20.24	7.59	7.76	31.51
actual-prcp	0.02	0.04	0.00	0.25
pred-temp-low	16.80	4.48	7.40	29.55
pred-temp-up	29.34	7.68	16.01	51.92
pred-prcp-low	0.26	0.28	0.00	1.27
pred-prcp-up	0.26	0.28	0.00	1.27

T = 133

with actual values for average temperature and precipitation, while Models 3 and 5 use the CPC forecasts.

By comparing Model 1 to Models 2 and 3, we can observe a large improvement in the explanatory power due to temperature alone. The improvement is supported by both R^2 and the Schwartz Bayesian Criterion (SBC). For both models, the temperature variables are statistically significant and have the expected sign. Months with higher average temperature tend to be followed by larger demand for electricity, again in accordance with the literature and with our own previous findings. Interestingly, it appears that the CPC forecasts, despite the long lead of one year, are able to predict the seasonal average with good accuracy. Not only are the coefficients fairly similar, but also the R^2 and SBC values are not distant from each other.

Actual and predicted values for precipitation are introduced in Models 4 and 5. This time, however, the variable did not appear to bring any improvement to the fit. Even in Model 4, in which the real monthly average for precipitation was used, the coefficient was not statistically significant at any conventional level. In Model 5, the forecast values for lower and upper terciles had diverging signs. Accordingly, by judging from the SBC scores, we would favor the previous models for their better fit and better parsimony.

However, as before we make the caveat that the precipitation variables, as available for this study, may not be adequate for capturing the relationship with demand for electricity. To be sure, we have hypothesized that dry periods would pose higher demand for electricity for different reasons, including evaporative, energy-intensive treatment of water and artificial irrigation (the water-energy nexus). Conversely, periods with more precipitation would lead to less use of electricity, by naturally satisfying those needs.

Nevertheless, precipitation as a measure of condensation, may not necessarily capture the effect of water scarcity in the region. If this is the case, a variable for drought intensity might prove to be more adequate. The relationship can also be countered by preemptive actions taken by the public. If the seasonal forecasts serve to prepare the government and individuals in general, we would not be able to observe the effect in the model. Due to these limitations, we cannot conclude from our estimations whether or not precipitation would play an important role in long-term planning.

In general we have found support for the idea that climate forecasts are reliable enough to be used in medium-term modeling and planning. Predictions made by CPC can accurately capture the movement in seasonal average of temperature and, to a lesser extent, precipitation. With good assessments of how the next year season will be, decision-makers can have better chances of

Table 4.4: Estimation Results for Seasonal Model

Variable	Model 1 Only Trend	Model 2 (1) + Temp	Model 3 (1) + P Temp	Model 4 (2) + Prcp	Model 5 (3) + P Prcp
constant	3.3069***	2.9269***	2.6218***	2.9167**	2.5983***
	(0.0425)	(0.0292)	(0.0546)	(0.0285)	(0.0568)
trend	0.0065***	0.0063***	0.0060***	0.0062**	0.0060***
	(0.0006)	(0.0002)	(0.0003)	(0.0002)	(0.0003)
actual-temp		0.0196***		0.0199**	
		(0.0012)		(0.0012)	
actual-prcp				0.2406	
				(0.1883)	
pred-temp-low			0.0124***		0.0131***
			(0.0024)		(0.0024)
pred-temp-up			0.0174***		0.0174***
			(0.0014)		(0.0014)
pred-prcp-low					-0.9368
					(1.9529)
pred-prcp-up					0.9721
					(1.9520)
DF	129	128	127	127	125
R^2	0.8829	0.9237	0.9130	0.9229	0.9144
SBC	-196.90	-249.83	-228.02	-243.69	-220.41

Significance Levels: *.05 **.01 ***.001

allocating necessary resources and take preemptive measures to handle energy crises. Without these techniques, managers have higher risk of being caught off-guard by unexpected variations in climate. With tighter regulations controlling the expansion of powerplants, the room for error becomes narrow, calling for more sophisticated methodology.

The simple models presented here can be put to use with little effort, as it simply requires the utility to retrieve information that is publicly available on the CPC website. Because the service also provides historical information, back to many decades, more thorough validation can be done if more data on electricity load is available.

CHAPTER 5

Conclusion

Sustainability in the Southwest is increasingly becoming a matter of concern for the government, business, advocacy organizations, and the public in general. The region faces rapid growth in its population and economy, while access to many natural resources remains limited by physical and, more recently, regulatory constraints. In particular, electricity embodies the essence of this problem, not only because of the dependency of technology on this form of energy, but also due to the intimate connection that power generation has with other elements, such as coal, gas, and water.

An additional source of pressure that has been traditionally taken for granted is climate, and the potential impacts of climate change to this already complicated scenario. To be sure, electric utilities have long known that usage of energy is strongly influenced by the type of climate of the locality that is served. Places with extremely cold winters or hot summers tend to have the peak of demand in different seasons, due to use of heating or cooling, and over the years companies have adjusted to those cycles. However, with, say, warmer temperatures and less precipitation, the seasonal patterns can change drastically. Of key significance is the possibility of increase in the occurrence of extreme conditions, such as heat waves and droughts, that coincide with periods when utilities are operating at their maximum capacity. Service disruptions due to energy crises, as the ones that occurred in California in 2001, can follow from lack of preparation to adverse situations.

To help electricity suppliers adapt to climate change, this study investigated the feasibility of developing statistical models that use weather and climatic information to provide planners with better input for decisions. Different models for forecasting short and medium-term electricity load have been created, leading to promising results. For both time horizons, we found that temperature alone is one of the most influential factors in explaining patterns of demand for electricity. Compared to a naive univariate specification, next-day forecasts were improved by about 1.4%, measured by MAPE, when temperature was included in the model. Augmenting the model with humidity, precipitation, and wind speed reduced the rate of error by .1%. Despite being small percentage changes, researchers have found that similar improvements can help utilities save millions of dollars annually. These results also add to research examining the economic value of weather and climate forecasts being freely provided by governmental agencies.

For the medium-term model, availability of only a few years of load data limited the range of possible models that could be developed. However, even a simple autoregressive model was able to generate interesting findings for the study. Perhaps the most outstanding finding is the observation that a model using seasonal forecasts provided by the Climate Prediction Center (CPC) at NOAA performed similarly to an alternative model using actual values for the average variables. This result should bring comfort to utility planners who, according to a previously mentioned survey (Changnon et al., 1995), were uncertain about the reliability of seasonal forecasts. The medium-term model was not without its shortcomings. The most severe of them is the omission of demographic and economic variables, which are not collected at monthly intervals, despite the fact that they have been found to play major roles in demand for energy in any region. This issue was handled by explicitly modeling trend in electricity consumption in the data.

Indeed, inspection of the data reveals that usage of electricity has been increasing steadily over the past years. Not only is the base load much higher today, but even more alarmingly so are peak loads at their highest in the summer. Although this study was not able to scrutinize these relationships, it is evident that population migration to the Southwest, especially during the 1990s, and accompanying advances in personal income can explain large part of this trend. Our argument, supported by several other studies, is that such issues can be further aggravated by the connection between water and energy, in which demand for one is closely associated with demand for the other. If water crises occur, decision-makers are likely to face the trade-off between reserving the resource for consumption, irrigation, and power generation, to cite a few. Thus, by indirect means, a crisis in the energy sector could follow.

The energy sector is also responsible for a large share of carbon dioxide emissions in the U.S., mostly due to combustion of fossil fuels still used for electricity generation. The sector is potentially doubly affected by emissions, either via climate change increasing demand for cooling, or through environmental regulations created for the purpose of limiting the release of pollutants to the atmosphere. Usually the most effective way of solving this problem is to replace existing generators with alternative ones, such as photovoltaic panels, wind turbines, or nuclear power plants. Evidently, this solution is often prohibitively expensive. In addition, industry may be conservative regarding adopting a new technology. But again, preemptive planning can be obtained by improving existing methods of assessing future scenarios.

As an initial effort in a largely unexplored field, it is our hope that other studies will continue to investigate the issues that were convered in this research. Experience with the short-term model suggests at least three areas of interest. First, we have used actual values for the weather variables, while in practice one would have to substitute them with forecasts, such as the ones provided by the National Weather Service (NWS). Fortunately, NWS offers hourly point predictions of up to 5 days ahead and the entire process of retrieving and feeding the model for estimation can be automated via the Internet. In addition, the new research could compare the performance of the NWS forecasts with a model using the actual values.

Second, more research could be used in evaluating other modeling techniques, in particular artificial intelligence algorithms. These methods have traditionally won forecasting competitions sponsored by energy organizations, and they have the double appeal of capturing nonlinearity and being largely automatic. Their major drawback, however, is the difficulty in interpreting the results, as they are designed to be used as black-box tools. For an electric utility, for whom the main interest is to minimize the rate of error, this may not be an impeding issue, but it was an obstacle for our goal of understanding the relationships between load, climate and other variables.

Third, this study proposed a method of handling multiple dimensions of seasonality that has not been investigated in the literature. Naturally, more research is necessary to assess the advantages and disadvantages of this method and how generalizable the results are. To be sure, consumption of energy follows a daily pattern, but that pattern changes across the year. It is likely that we would not observe similar cycles in financial markets, but agriculture may be subject to them. Furthermore, this technique can prove useful in modeling variability as well, through ARCH or GARCH models, and again summer and winter provide examples where this would be applicable.

Finally, if a researcher can have access to load data for multiple locations, she can take advantage of that for cointegration analysis. This technique exploits the fact that quite often when two or more time series are nonstationary, stationarity can still be reached by some linear combination of them. If an electric utility is willing to provide disaggregated data for different stations, an econometrician can even engage in spatial cointegration, in which details are investigated via GIS mapping.

In the case of the medium-term, seasonal models, more could be done if a dataset with a longer period were available. With the possibility of aggregating at annual intervals, one would be able to match the time interval with the wealth of information collected by the U.S. Census Bureau, such as population and income projections. Again, a more detailed spatial analysis would be possible, because the Census data includes mapping in GIS format. This can prove useful, since we cannot expect that all localities in the Southwest would change homogeneously.

A pertinent set of issues that has not been investigated empirically in this study is the water-energy nexus. The link can, however, be studied in various levels depending on data availability. For instance, if one can have measurement of the amount of water that is wasted due to overproduction of electricity, it would be possible to estimate the quantity saved by improving load forecasts. Similarly, it would be interesting to evaluate the amount of water that is saved when a utility switches from one type of generator to another. So far, most studies of this kind have addressed the issue of carbon dioxide emissions, not water.

It is important to reiterate that our research intended to give continuation to the work of Thompson and Cathers (2005). In their concluding remarks, the authors suggestions for future researchers to improve their achievements. In particular, it was noted that their five ARIMA models did not use weather variables, partly due to the complexity of incorporating them via transfer functions, partly because weather forecasts were not available for all regions investigated. We proposed a method of employing those variables in a way that is familiar for econometricians, and a model that can handle seasonality in more than one dimension, thus eliminating the need for creating multiple models for arbitrarily defined seasons. It is our hope that we have been able to inspire other researchers to improve our findings even further.

APPENDIX A

Model Estimates

A.1 Estimates for Full Model

Yule-Walker Estimates

SSE	70.5935562	DFE	86298
MSE	0.0008180	Root MSE	0.02860
SBC	-367784.97	AIC	-368965.22
Regress R-Square	0.8139	Total R-Square	0.9939
Durbin-Watson	1.1083		

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	3.1843	0.008980	354.61	<.0001
trend	1	8.5987E-6	1.0629E-7	80.90	<.0001
tuesday	1	0.002683	0.000864	3.10	0.0019
wednesday	1	0.003999	0.001105	3.62	0.0003
thursday	1	0.003512	0.001202	2.92	0.0035
friday	1	-0.001533	0.001205	-1.27	0.2032
saturday	1	-0.0121	0.001108	-10.87	<.0001
sunday	1	-0.0126	0.000868	-14.47	<.0001
holiday	1	-0.005459	0.001540	-3.55	0.0004
temp	1	-0.0214	0.000198	-108.31	<.0001
temp2	1	0.000591	4.2785E-6	138.12	<.0001
humi	1	0.001231	0.0000675	18.25	<.0001
humi2	1	-9.792E-6	5.3432E-7	-18.33	<.0001
prcp	1	0.002402	0.000993	2.42	0.0156
wind	1	-0.000856	0.000120	-7.16	<.0001
d0_h1	1	11476	37.5214	305.86	<.0001
d0_h2	1	-8509	47.7831	-178.07	<.0001
d0_h3	1	-4506	32.7003	-137.79	<.0001
d0_h4	1	-308.0791	24.0319	-12.82	<.0001
d0_h5	1	-4700	18.6048	-252.61	<.0001
d0_h6	1	452.9034	14.8950	30.41	<.0001
d0_h7	1	1792	12.8665	139.28	<.0001
d0_h8	1	-84.8518	10.4968	-8.08	<.0001
d0_h9	1	340.8971	9.6908	35.18	<.0001
d1_h0	1	2125	217.2560	9.78	<.0001
d1_h1	1	732.3483	32.6329	22.44	<.0001

d1_h2	1	-411.4070	38.1639	-10.78	<.0001
d1_h3	1	-1125	26.4118	-42.58	<.0001
d1_h4	1	385.3418	20.4972	18.80	<.0001
d1_h5	1	646.9496	16.7857	38.54	<.0001
d1_h6	1	-40.6973	14.3257	-2.84	0.0045
d1_h7	1	28.9209	12.5082	2.31	0.0208
d1_h8	1	-9.6999	10.3305	-0.94	0.3478
d1_h9	1	-178.4781	9.4724	-18.84	<.0001
d2_h0	1	-8679	236.6363	-36.68	<.0001
d2_h1	1	-2535	34.7007	-73.06	<.0001
d2_h2	1	787.2066	42.2047	18.65	<.0001
d2_h3	1	3012	29.5897	101.78	<.0001
d2_h4	1	-1830	21.7099	-84.30	<.0001
d2_h5	1	-1833	17.7936	-103.00	<.0001
d2_h6	1	995.2627	14.5587	68.36	<.0001
d2_h7	1	803.8289	12.6984	63.30	<.0001
d2_h8	1	-207.3431	10.4180	-19.90	<.0001
d2_h9	1	-152.6057	9.6369	-15.84	<.0001
d3_h0	1	-3566	205.5484	-17.35	<.0001
d3_h1	1	-311.9031	32.7384	-9.53	<.0001
d3_h2	1	601.4649	37.7807	15.92	<.0001
d3_h3	1	895.8356	26.2720	34.10	<.0001
d3_h4	1	-457.2704	20.5080	-22.30	<.0001
d3_h5	1	-713.4952	16.7106	-42.70	<.0001
d3_h6	1	105.0122	14.2863	7.35	<.0001
d3_h7	1	102.1910	12.4669	8.20	<.0001
d3_h8	1	-41.0666	10.2958	-3.99	<.0001
d3_h9	1	78.9655	9.4441	8.36	<.0001
d4_h0	1	7467	221.1737	33.76	<.0001
d4_h1	1	207.8927	32.9039	6.32	<.0001
d4_h2	1	525.4385	38.8438	13.53	<.0001
d4_h3	1	-553.0160	26.7067	-20.71	<.0001
d4_h4	1	624.7717	20.5759	30.36	<.0001
d4_h5	1	541.9089	16.8463	32.17	<.0001
d4_h6	1	-500.0805	14.3445	-34.86	<.0001
d4_h7	1	-22.6771	12.5909	-1.80	0.0717
d4_h8	1	173.4494	10.3579	16.75	<.0001
d4_h9	1	-212.9633	9.5235	-22.36	<.0001
d5_h0	1	4140	196.4640	21.07	<.0001
d5_h1	1	-644.7001	32.3925	-19.90	<.0001
d5_h2	1	136.3569	37.0051	3.68	0.0002
d5_h3	1	351.1849	25.8989	13.56	<.0001
d5_h4	1	90.0781	20.2515	4.45	<.0001
d5_h5	1	80.9963	16.6202	4.87	<.0001
d5_h6	1	-133.1702	14.1796	-9.39	<.0001
d5_h7	1	-25.9977	12.4230	-2.09	0.0364
d5_h8	1	95.4483	10.2593	9.30	<.0001
d5_h9	1	-19.1636	9.4153	-2.04	0.0418
d6_h0	1	-2052	208.9044	-9.82	<.0001
d6_h1	1	606.2987	32.9812	18.38	<.0001

d6_h2	1	-514.3408	38.7368	-13.28	<.0001
d6_h3	1	-634.6086	26.5823	-23.87	<.0001
d6_h4	1	203.7755	20.5585	9.91	<.0001
d6_h5	1	227.0236	16.8620	13.46	<.0001
d6_h6	1	-80.0653	14.3435	-5.58	<.0001
d6_h7	1	-91.5847	12.5466	-7.30	<.0001
d6_h8	1	42.4947	10.3521	4.10	<.0001
d6_h9	1	48.2502	9.4881	5.09	<.0001
d7_h0	1	-1117	187.9678	-5.94	<.0001
d7_h1	1	704.9596	32.0521	21.99	<.0001
d7_h2	1	-512.1968	36.2485	-14.13	<.0001
d7_h3	1	-684.5667	25.6340	-26.71	<.0001
d7_h4	1	144.0192	20.0716	7.18	<.0001
d7_h5	1	210.0250	16.5343	12.70	<.0001
d7_h6	1	20.2680	14.1064	1.44	0.1508
d7_h7	1	-17.5886	12.3659	-1.42	0.1549
d7_h8	1	-20.2456	10.2261	-1.98	0.0477
d7_h9	1	-4.1294	9.3875	-0.44	0.6600
d8_h0	1	-728.5858	193.8831	-3.76	0.0002
d8_h1	1	-199.3724	33.1930	-6.01	<.0001
d8_h2	1	221.9641	38.8608	5.71	<.0001
d8_h3	1	175.5444	26.6321	6.59	<.0001
d8_h4	1	46.4234	20.5828	2.26	0.0241
d8_h5	1	-67.1668	16.8570	-3.98	<.0001
d8_h6	1	-43.8218	14.3434	-3.06	0.0022
d8_h7	1	62.2190	12.5344	4.96	<.0001
d8_h8	1	11.3136	10.3488	1.09	0.2743
d8_h9	1	-54.5922	9.4820	-5.76	<.0001
d9_h0	1	-630.8112	166.0980	-3.80	0.0001
d9_h1	1	-592.3610	31.4297	-18.85	<.0001
d9_h2	1	28.1601	35.3310	0.80	0.4254
d9_h3	1	499.4927	25.2031	19.82	<.0001
d9_h4	1	-82.1396	19.8614	-4.14	<.0001
d9_h5	1	-75.7827	16.4096	-4.62	<.0001
d9_h6	1	89.5930	14.0330	6.38	<.0001
d9_h7	1	-12.0391	12.3100	-0.98	0.3281
d9_h8	1	-46.9801	10.1914	-4.61	<.0001
d9_h9	1	8.4086	9.3571	0.90	0.3689
d10_h0	1	-113.3768	167.1119	-0.68	0.4975
d10_h1	1	18.2309	32.7272	0.56	0.5775
d10_h2	1	-178.7835	38.7995	-4.61	<.0001
d10_h3	1	-43.6597	26.5671	-1.64	0.1003
d10_h4	1	95.5523	20.5817	4.64	<.0001
d10_h5	1	21.3546	16.8579	1.27	0.2053
d10_h6	1	-20.8519	14.3394	-1.45	0.1459
d10_h7	1	46.6175	12.5290	3.72	0.0002
d10_h8	1	-5.1880	10.3411	-0.50	0.6159
d10_h9	1	-51.2120	9.4727	-5.41	<.0001

A.2 Estimates for Model with Temperature

Yule-Walker Estimates

SSE	70.6609937	DFE	86302
MSE	0.0008188	Root MSE	0.02861
SBC	-367747.95	AIC	-368890.73
Regress R-Square	0.8147	Total R-Square	0.9939
Durbin-Watson	1.1111		

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	3.2273	0.008689	371.40	<.0001
trend	1	8.5836E-6	1.1023E-7	77.87	<.0001
tuesday	1	0.002701	0.000875	3.09	0.0020
wednesday	1	0.004013	0.001118	3.59	0.0003
thursday	1	0.003572	0.001217	2.93	0.0033
friday	1	-0.001286	0.001220	-1.05	0.2918
saturday	1	-0.0117	0.001122	-10.39	<.0001
sunday	1	-0.0122	0.000879	-13.91	<.0001
holiday	1	-0.005180	0.001559	-3.32	0.0009
temp	1	-0.0218	0.000181	-119.85	<.0001
temp2	1	0.000575	4.1363E-6	138.94	<.0001
d0_h1	1	11518	37.3677	308.24	<.0001
d0_h2	1	-8574	47.3645	-181.03	<.0001
d0_h3	1	-4532	32.5967	-139.04	<.0001
d0_h4	1	-267.6873	23.8839	-11.21	<.0001
d0_h5	1	-4711	18.4190	-255.75	<.0001
d0_h6	1	437.2504	14.7371	29.67	<.0001
d0_h7	1	1813	12.7435	142.26	<.0001
d0_h8	1	-82.9237	10.3444	-8.02	<.0001
d0_h9	1	326.4295	9.6078	33.98	<.0001
d1_h0	1	2295	224.8513	10.21	<.0001
d1_h1	1	742.5683	32.4826	22.86	<.0001
d1_h2	1	-431.2008	37.9846	-11.35	<.0001
d1_h3	1	-1136	26.2841	-43.22	<.0001
d1_h4	1	394.1922	20.3863	19.34	<.0001
d1_h5	1	649.3123	16.6917	38.90	<.0001
d1_h6	1	-43.1402	14.2370	-3.03	0.0024
d1_h7	1	31.2893	12.4360	2.52	0.0119
d1_h8	1	-9.7090	10.2710	-0.95	0.3445
d1_h9	1	-180.2379	9.4173	-19.14	<.0001
d2_h0	1	-9143	242.5084	-37.70	<.0001
d2_h1	1	-2550	34.4124	-74.10	<.0001
d2_h2	1	845.6532	41.8647	20.20	<.0001
d2_h3	1	3032	29.2714	103.58	<.0001
d2_h4	1	-1858	21.5831	-86.10	<.0001
d2_h5	1	-1847	17.6488	-104.65	<.0001

d2_h6	1	994.8047	14.4649	68.77	<.0001
d2_h7	1	805.8358	12.6235	63.84	<.0001
d2_h8	1	-199.5273	10.3473	-19.28	<.0001
d2_h9	1	-149.9518	9.5800	-15.65	<.0001
d3_h0	1	-3881	211.5240	-18.35	<.0001
d3_h1	1	-339.1009	32.5217	-10.43	<.0001
d3_h2	1	637.5367	37.5328	16.99	<.0001
d3_h3	1	919.6610	26.0856	35.26	<.0001
d3_h4	1	-476.2164	20.3553	-23.40	<.0001
d3_h5	1	-715.0652	16.6198	-43.02	<.0001
d3_h6	1	114.0137	14.1885	8.04	<.0001
d3_h7	1	97.5219	12.3946	7.87	<.0001
d3_h8	1	-43.2354	10.2381	-4.22	<.0001
d3_h9	1	82.2440	9.3884	8.76	<.0001
d4_h0	1	7632	228.9031	33.34	<.0001
d4_h1	1	217.9767	32.7276	6.66	<.0001
d4_h2	1	501.3939	38.6517	12.97	<.0001
d4_h3	1	-565.4053	26.5524	-21.29	<.0001
d4_h4	1	629.1993	20.4678	30.74	<.0001
d4_h5	1	547.2367	16.7509	32.67	<.0001
d4_h6	1	-496.2601	14.2627	-34.79	<.0001
d4_h7	1	-21.6921	12.5190	-1.73	0.0831
d4_h8	1	171.7679	10.2997	16.68	<.0001
d4_h9	1	-213.4026	9.4705	-22.53	<.0001
d5_h0	1	4367	202.7132	21.54	<.0001
d5_h1	1	-620.9325	32.2188	-19.27	<.0001
d5_h2	1	118.9533	36.7374	3.24	0.0012
d5_h3	1	336.5028	25.7304	13.08	<.0001
d5_h4	1	104.5289	20.1087	5.20	<.0001
d5_h5	1	75.7617	16.5285	4.58	<.0001
d5_h6	1	-140.7271	14.0956	-9.98	<.0001
d5_h7	1	-18.0781	12.3483	-1.46	0.1432
d5_h8	1	96.1630	10.2052	9.42	<.0001
d5_h9	1	-23.6434	9.3595	-2.53	0.0115
d6_h0	1	-2031	216.2301	-9.39	<.0001
d6_h1	1	600.9414	32.8338	18.30	<.0001
d6_h2	1	-502.3442	38.5481	-13.03	<.0001
d6_h3	1	-628.8027	26.4504	-23.77	<.0001
d6_h4	1	204.8795	20.4499	10.02	<.0001
d6_h5	1	224.4355	16.7690	13.38	<.0001
d6_h6	1	-81.3429	14.2643	-5.70	<.0001
d6_h7	1	-91.6383	12.4757	-7.35	<.0001
d6_h8	1	41.4074	10.2937	4.02	<.0001
d6_h9	1	47.2142	9.4347	5.00	<.0001
d7_h0	1	-1261	194.3229	-6.49	<.0001
d7_h1	1	687.3488	31.8898	21.55	<.0001
d7_h2	1	-500.3002	36.0421	-13.88	<.0001
d7_h3	1	-675.6236	25.4800	-26.52	<.0001
d7_h4	1	136.1681	19.9556	6.82	<.0001
d7_h5	1	217.1436	16.4491	13.20	<.0001

d7_h6	1	24.1387	14.0312	1.72	0.0854
d7_h7	1	-25.7655	12.2954	-2.10	0.0361
d7_h8	1	-20.1604	10.1730	-1.98	0.0475
d7_h9	1	-0.2307	9.3343	-0.02	0.9803
d8_h0	1	-736.8734	200.4919	-3.68	0.0002
d8_h1	1	-198.8503	33.0550	-6.02	<.0001
d8_h2	1	216.7173	38.6772	5.60	<.0001
d8_h3	1	176.8303	26.5040	6.67	<.0001
d8_h4	1	43.5457	20.4732	2.13	0.0334
d8_h5	1	-67.0340	16.7648	-4.00	<.0001
d8_h6	1	-43.1459	14.2636	-3.02	0.0025
d8_h7	1	62.2010	12.4641	4.99	<.0001
d8_h8	1	12.2327	10.2904	1.19	0.2345
d8_h9	1	-54.2312	9.4289	-5.75	<.0001
d9_h0	1	-565.3058	171.4934	-3.30	0.0010
d9_h1	1	-591.5728	31.3601	-18.86	<.0001
d9_h2	1	32.7710	35.2086	0.93	0.3520
d9_h3	1	500.1377	25.1182	19.91	<.0001
d9_h4	1	-83.8136	19.7782	-4.24	<.0001
d9_h5	1	-79.7734	16.3352	-4.88	<.0001
d9_h6	1	89.7666	13.9657	6.43	<.0001
d9_h7	1	-9.1908	12.2475	-0.75	0.4530
d9_h8	1	-46.9365	10.1401	-4.63	<.0001
d9_h9	1	8.1900	9.3083	0.88	0.3789
d10_h0	1	-97.8145	172.2992	-0.57	0.5702
d10_h1	1	19.7945	32.5826	0.61	0.5435
d10_h2	1	-184.0532	38.6328	-4.76	<.0001
d10_h3	1	-46.9891	26.4468	-1.78	0.0756
d10_h4	1	99.0606	20.4730	4.84	<.0001
d10_h5	1	23.8935	16.7655	1.43	0.1541
d10_h6	1	-22.3718	14.2595	-1.57	0.1167
d10_h7	1	46.2831	12.4591	3.71	0.0002
d10_h8	1	-4.8583	10.2832	-0.47	0.6366
d10_h9	1	-51.7526	9.4199	-5.49	<.0001

A.3 Estimates for Naive Model

Yule-Walker Estimates

SSE	81.5059788	DFE	86304
MSE	0.0009444	Root MSE	0.03073
SBC	-355430.78	AIC	-356554.83
Regress R-Square	0.8078	Total R-Square	0.9930
Durbin-Watson	1.0872		

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|

Intercept	1	3.0804	0.0253	121.79	<.0001
trend	1	8.558E-6	3.3086E-7	25.87	<.0001
tuesday	1	0.003059	0.001127	2.71	0.0066
wednesday	1	0.004333	0.001443	3.00	0.0027
thursday	1	0.004520	0.001573	2.87	0.0041
friday	1	0.002032	0.001576	1.29	0.1973
saturday	1	-0.004466	0.001448	-3.08	0.0020
sunday	1	-0.006480	0.001132	-5.72	<.0001
holiday	1	-0.004005	0.002006	-2.00	0.0459
d0_h1	1	11678	31.8371	366.79	<.0001
d0_h2	1	-9522	37.8527	-251.55	<.0001
d0_h3	1	-5065	25.9958	-194.84	<.0001
d0_h4	1	49.3336	20.0209	2.46	0.0137
d0_h5	1	-4063	16.3819	-248.01	<.0001
d0_h6	1	538.5596	13.9328	38.65	<.0001
d0_h7	1	1444	12.1755	118.60	<.0001
d0_h8	1	-215.8730	10.0501	-21.48	<.0001
d0_h9	1	423.5645	9.2135	45.97	<.0001
d1_h0	1	2415	645.2182	3.74	0.0002
d1_h1	1	1051	32.7156	32.14	<.0001
d1_h2	1	-717.2763	37.6272	-19.06	<.0001
d1_h3	1	-1388	25.9353	-53.52	<.0001
d1_h4	1	581.1391	19.9877	29.07	<.0001
d1_h5	1	721.3288	16.3590	44.09	<.0001
d1_h6	1	-155.0685	13.9152	-11.14	<.0001
d1_h7	1	0.5338	12.1611	0.04	0.9650
d1_h8	1	29.4289	10.0390	2.93	0.0034
d1_h9	1	-171.9332	9.2033	-18.68	<.0001
d2_h0	1	-9523	700.5444	-13.59	<.0001
d2_h1	1	-4192	32.0917	-130.62	<.0001
d2_h2	1	3183	37.8556	84.08	<.0001
d2_h3	1	4794	25.9957	184.40	<.0001
d2_h4	1	-2833	20.0197	-141.49	<.0001
d2_h5	1	-2505	16.3805	-152.94	<.0001
d2_h6	1	1313	13.9314	94.23	<.0001
d2_h7	1	754.6297	12.1742	61.99	<.0001
d2_h8	1	-293.7993	10.0490	-29.24	<.0001
d2_h9	1	58.5878	9.2125	6.36	<.0001
d3_h0	1	-4196	587.3074	-7.14	<.0001
d3_h1	1	-688.8829	33.3262	-20.67	<.0001
d3_h2	1	1158	37.3114	31.03	<.0001
d3_h3	1	1228	25.8459	47.52	<.0001
d3_h4	1	-805.2981	19.9388	-40.39	<.0001
d3_h5	1	-800.2451	16.3254	-49.02	<.0001
d3_h6	1	248.4415	13.8895	17.89	<.0001
d3_h7	1	114.9978	12.1400	9.47	<.0001
d3_h8	1	-67.0618	10.0229	-6.69	<.0001
d3_h9	1	83.1667	9.1884	9.05	<.0001
d4_h0	1	9104	638.9612	14.25	<.0001

d4_h1	1	728.4887	32.8544	22.17	<.0001
d4_h2	1	-76.8117	37.8935	-2.03	0.0427
d4_h3	1	-956.9268	26.0127	-36.79	<.0001
d4_h4	1	803.2105	20.0277	40.10	<.0001
d4_h5	1	510.7092	16.3854	31.17	<.0001
d4_h6	1	-499.4665	13.9348	-35.84	<.0001
d4_h7	1	143.9463	12.1768	11.82	<.0001
d4_h8	1	144.1061	10.0508	14.34	<.0001
d4_h9	1	-325.4766	9.2139	-35.32	<.0001
d5_h0	1	4742	560.4824	8.46	<.0001
d5_h1	1	-635.1931	33.5464	-18.93	<.0001
d5_h2	1	102.7347	36.9638	2.78	0.0054
d5_h3	1	368.7351	25.7470	14.32	<.0001
d5_h4	1	178.3622	19.8886	8.97	<.0001
d5_h5	1	17.1306	16.2927	1.05	0.2931
d5_h6	1	-179.6783	13.8654	-12.96	<.0001
d5_h7	1	11.6858	12.1209	0.96	0.3350
d5_h8	1	97.8345	10.0087	9.77	<.0001
d5_h9	1	-20.4491	9.1756	-2.23	0.0258
d6_h0	1	-2892	572.4026	-5.05	<.0001
d6_h1	1	499.0706	33.6200	14.84	<.0001
d6_h2	1	-533.4188	37.9456	-14.06	<.0001
d6_h3	1	-632.6515	26.0291	-24.31	<.0001
d6_h4	1	256.4038	20.0308	12.80	<.0001
d6_h5	1	336.4059	16.3844	20.53	<.0001
d6_h6	1	-151.0519	13.9324	-10.84	<.0001
d6_h7	1	-177.0377	12.1738	-14.54	<.0001
d6_h8	1	85.3465	10.0478	8.49	<.0001
d6_h9	1	80.0858	9.2104	8.70	<.0001
d7_h0	1	-1557	524.7736	-2.97	0.0030
d7_h1	1	848.9522	33.6559	25.22	<.0001
d7_h2	1	-783.6246	36.5708	-21.43	<.0001
d7_h3	1	-881.1980	25.6284	-34.38	<.0001
d7_h4	1	252.8832	19.8307	12.75	<.0001
d7_h5	1	320.0990	16.2562	19.69	<.0001
d7_h6	1	8.5366	13.8393	0.62	0.5373
d7_h7	1	-28.9814	12.1006	-2.40	0.0166
d7_h8	1	-26.6498	9.9939	-2.67	0.0077
d7_h9	1	-32.6691	9.1629	-3.57	0.0004
d8_h0	1	-682.0718	511.6750	-1.33	0.1825
d8_h1	1	-123.9037	34.0823	-3.64	0.0003
d8_h2	1	160.6717	38.0362	4.22	<.0001
d8_h3	1	118.5356	26.0626	4.55	<.0001
d8_h4	1	52.2025	20.0456	2.60	0.0092
d8_h5	1	-60.4537	16.3917	-3.69	0.0002
d8_h6	1	-14.7800	13.9363	-1.06	0.2889
d8_h7	1	69.1235	12.1758	5.68	<.0001
d8_h8	1	-9.0296	10.0486	-0.90	0.3689
d8_h9	1	-59.3626	9.2102	-6.45	<.0001
d9_h0	1	-824.5824	408.4670	-2.02	0.0435

d9_h1	1	-710.7226	33.6345	-21.13	<.0001
d9_h2	1	324.9409	36.1461	8.99	<.0001
d9_h3	1	621.2534	25.4951	24.37	<.0001
d9_h4	1	-135.2549	19.7670	-6.84	<.0001
d9_h5	1	-118.7089	16.2174	-7.32	<.0001
d9_h6	1	59.1993	13.8123	4.29	<.0001
d9_h7	1	-33.7998	12.0803	-2.80	0.0051
d9_h8	1	-23.1645	9.9795	-2.32	0.0203
d9_h9	1	41.6916	9.1509	4.56	<.0001
d10_h0	1	-275.2522	388.8559	-0.71	0.4790
d10_h1	1	-55.1061	32.1305	-1.72	0.0863
d10_h2	1	-91.4511	38.0950	-2.40	0.0164
d10_h3	1	16.0239	26.0724	0.61	0.5388
d10_h4	1	68.6361	20.0496	3.42	0.0006
d10_h5	1	17.7160	16.3904	1.08	0.2798
d10_h6	1	-24.7411	13.9328	-1.78	0.0758
d10_h7	1	31.0305	12.1712	2.55	0.0108
d10_h8	1	0.9398	10.0439	0.09	0.9254
d10_h9	1	-43.9166	9.2045	-4.77	<.0001

APPENDIX B

Chebyshev Polynomials of First Kind

For convenience, we provide the calculations of the first nine Chebyshev polynomials of first kind. Recall that the rescursive definition is:

$$T_0(x) = 1$$

$$T_1(x) = x$$

$$T_i(x) = 2xT_{i-1}(x) - T_{i-2}(x)$$

where i is the order of the polynomial.

Then, by simple expansion the polynomal terms can be shown to be:

$$
\begin{aligned}
T_0(x) &= 1 \\
T_1(x) &= x \\
T_2(x) &= 2x^2 - 1 \\
T_3(x) &= 4x^3 - 3x \\
T_4(x) &= 8x^4 - 8x^2 + 1 \\
T_5(x) &= 16x^5 - 20x^3 + 5x \\
T_6(x) &= 32x^6 - 48x^4 + 18x^2 - 1 \\
T_7(x) &= 64x^7 - 112x^5 + 56x^3 - 7x \\
T_8(x) &= 128x^8 - 256x^6 + 160x^4 - 32x^2 + 1 \\
T_9(x) &= 256x^9 - 576x^7 + 432x^5 - 120x^3 + 9x
\end{aligned}
$$

APPENDIX C

Program in Haskell for Calculating Holidays

```
module Holiday where
----------------------------------------------------------------------
-- Holiday.hs
-- A simple module for calculating holidays.
--
-- Author:  Paulo Tanimoto <ptanimoto@gmail.com>
-- Date:    2008-02-29
-- License: BSD3
----------------------------------------------------------------------

----------------------------------------------------------------------
-- Import
----------------------------------------------------------------------
import Data.Time
import System.Locale (defaultTimeLocale)

----------------------------------------------------------------------
-- Date Functions
----------------------------------------------------------------------
-- sunday=0, saturday=6
weekday :: Day -> Integer
weekday = read . formatTime defaultTimeLocale "%w"

findFirst :: Integer -> Int -> Integer -> Day
findFirst y m k | k < k'    = addDays (7 + k - k') d
                | otherwise = addDays (k - k') d
  where d  = fromGregorian y m 1
        k' = weekday d

findLast :: Integer -> Int -> Integer -> Day
findLast y m k | k > k'    = addDays (k - k' - 7) d
               | otherwise = addDays (k - k') d
  where d  = fromGregorian y m (gregorianMonthLength y m)
        k' = weekday d

findWeek :: Integer -> Int -> Integer -> Integer -> Day
findWeek y m w k = addDays x day
  where day = findFirst y m k
        x = 7 * (w-1)

findWeekR :: Integer -> Int -> Integer -> Integer -> Day
findWeekR y m w k = addDays (-x) day
```

```
  where day = findLast y m k
        x = 7 * (w-1)

toWeekday :: Integer -> Int -> Int -> Day
toWeekday y m d | wday == 0 = addDays ( 1) day
                | wday == 6 = addDays (-1) day
                | otherwise = day
  where day  = fromGregorian y m d
        wday = weekday day


-----------------------------------------------------------------------
-- Holidays for Arizona
-----------------------------------------------------------------------
holidays ys = [[ toWeekday y  1  1    -- new year
               , findWeek  y  1  3  1 -- martin luther king
               , findWeek  y  2  3  1 -- president's day
               , findWeekR y  5  1  1 -- memorial day
               , toWeekday y  7  4    -- independence day
               , findWeek  y  9  1  1 -- labor day
               , findWeek  y 10  2  1 -- columbus' birthday
               , toWeekday y 11 11    -- veterans' day
               , findWeek  y 11  4  4 -- thanksgiving
               , toWeekday y 12 25  ] -- christmas
               | y <- ys ]
```

REFERENCES

American Meteorological Society, "Glossary of Meteorology," `http://amsglossary.allenpress.com/glossary/` 2008.

Aradhyula, Satheesh and Russell Tronstad, "Estimating Seasonality in the Mean and Variance," *Essays in Honor of Stanley R. Johnson*, September 2006, (20).

Aradhyula, Satheesh V. and A. Tolga Ergun, "Trading Collar, Intraday Periodicity and Stock Market Volatility," *Applied Financial Economics*, 2004, *14*, 909–913.

Arizona Electric Power Cooperative, "About Us," `http://www.aepco.coop/` 2008.

Box, George and Gwilym Jenkins, *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1970.

Brocklebank, John C. and David A. Dickey, *SAS for Forecasting Time Series*, 2 ed., SAS Institute, 2003.

Census Bureau, "State Interim Population Projections by Age and Sex: 2004 - 2030," Technical Report 2005.

Changnon, Stanley A., Joyce M. Changnon, and David Changnon, "Uses and Applications of Climate Forecasts for Power Utilities," *Bulletin of the American Meteorological Society*, May 1995, *76* (5), 711–720.

Chao, Hung-po, "Peak Load Pricing and Capacity Planning with Demand and Supply Uncertainty," *The Bell Journal of Economics*, 1983, *14* (1).

Commission, California Energy, "Integrated Energy Policy Report: Committee Draft Report," Technical Report 2005.

Department of Energy, "Energy Demand on Water Resources: Report to Congress on the Interdependency of Energy and Water," Technical Report 2006.

Dickey, D. and W. Fuller, "Distribution of theEstimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 1979, *74*, 427–431.

Durbin, J. and S. J. Koopman, *Time Series Analysis by State Space Methods*, Oxford: Oxford University Press, 2001.

Electric Power Research Institute, "Potential Impacts of Climate Change in Electric Utilities," Technical Report 88-2, EPRI December 1989.

__ , "Potential Effects of Climate Change on Electric Utilities," Technical Report 105005, EPRI 1995.

__ , "Climate Change Impacts and the Electric Sector," Technical Report, EPRI 2005.

__ , "Electric Power Research Institute Website," March 2008.

Feinberg, Eugene A. and Dora Genethliou, "Load Forecasting," in "Applied Mathematics for Restructured Electric Power Systems," Springer, 2005.

Gardner, Douglas T. and J. Scott Rogers, "Planning Electric Power Systems Under Demand Uncertainty with Different Technoloy Lead Times," *Management Science*, October 1999, *45* (10).

Granger, Clive W. J. and P. Newbold, "Spurious Regressions in Econometrics," *Journal of Econometrics*, 1974, *2*, 111–120.

Greene, William H., *Econometric Analysis*, 5 ed., New York, NY: Pearson, 2003.

Gross, George and Francisco D. Galiana, "Short-Term Load Forecasting," *Proceedings of the IEEE*, December 1987, *75* (12).

Hagan, Martin T. and Suzanne M. Behr, "The Time Series Approach to Short Term Load Forecasting," *IEEE Transactions on Power Systems*, August 1987, *2* (3).

Hobbs, B. F., "Identification and Quantification of Benefits of Using ANNSTLF," Technical Report, Electric Power Research Institute (EPRI) October 1997.

Hobbs, Benjamin F., Suradet Jiprapaikulsarn, and Dominic J. Maratukulam, "Analysis of the Value for Unit Commitment of Improved Load Forecasts," *IEEE Transactions on Power Systems*, November 1999, *14* (4), 1342–1347.

Hochstrasser, Urs W., "Orthogonal Polynomials," in Milton Abramowitz and Irene A. Stegun, eds., *Handbook of Mathematical Functions*, New York: Dover, 1965, pp. 771–802.

Hylleberg, S., "General Introduction," in S. Hylleberg, ed., *Modelling Seasonality*, Oxford: Oxford University Press, 1992, pp. 3–13.

Intergovernmental Panel on Climate Change, "Climate Change 2007: The Physical Sciences Basis," Technical Report, IPCC 2007.

Ishii, Yasunori, "On the Theory of the Competitive Firm Under Price Uncertainty: Note," *The American Economic Review*, September 1977, *67* (4), 768–769.

__ , "On the Theory of Monopolistic Competition Under Demand Uncertainty," *Journal of Economics*, 1991, *54* (1), 21–32.

Lenart, Melanie, "Global Warming in the Southwest: An Overview," in Melanie Lenart, ed., *Global Warming in the Southwest: Projections, Observations, and Impacts*, Climate Assessment for the Southwest, 2007, pp. 2–5.

Merideth, Robert, *A Primer on Climatic Variability and Change in the Southwest*, Tucson, AZ: Udall Center for Studies in Public Policy, March 2001.

Palmer, T. N., F. J. Doblas-Reyes, R. Hagedorn, and A. Weisheimer, "Probabilistic Prediction of Climate Using Multi-Model Ensembles: From Basics to Applications," *Philosophical Transactions of The Royal Society, B*, October 2005, *360*, 1991–1998.

Papalexopoulos, Alex D. and Timothy C. Hesterberg, "A Regression-Based Approach to Short-Term System Load Forecasting," *IEEE Transactions on Power Systems*, November 1990, *5* (4).

__ , S. Hao, and T. M. Peng, "An Implementation of a Neural Network Based Load Forecasting Model for the EMS," *IEEE Transactions on Power Systems*, 1994, *9*, 1956–1962.

Peng, M., N. F. Hubele, and G. G. Karady, "Advancement in the Application of Neural Networks for Short-Term Load Forecasting," *IEEE Transactions on Power Systems*, 1992, *7*, 250–257.

Phillips, P., "Understanding Spurious Regressions," *Journal of Econometrics*, 1986, *33*, 311–340.

Powicki, C., "The Water Imperative," Technical Report, Electric Power Research Institute (EPRI) 2002.

Roberts, H. V., "Comments on "The Analysis of Single and Related Time Series into Componenets: Proposals for Improving X-11"," in A. Zellner, ed., *Seasonal Analysis of Economic Time Series*, Washington, DC: US Department of Commerce, Bureau of the Census, 1978, pp. 161–170.

Sandmo, Agnar, "On the Theory of the Competitive Firm Under Price Uncertainty," *The American Economic Review*, March 1971, *61* (1), 65–73.

Sims, Christopher A., "Seasonality in Regression," in "Modelling Seasonality," Oxford: Oxford University Press, 1992, pp. 61–81.

Teisberg, Thomas J., Rodney F. Weiher, and Alireza Khotanzad, "The Economic Value of Temperature Forecasts in Electricity Generation," *Bulletin of the American Meteorological Society*, December 2005, *86* (12), 1765–1771.

Thompson, Gary D. and Clifford A. Cathers, "Forecasting Short-Term Electricity Load Profiles," Technical Report, University of Arizona December 2005.

Thompson, Stephen Andrew, *Water Use, Management, and Planning in the United States*, Academic Press, 1999.

Warren, Henry E. and Sharon K. LeDuc, "Impact of Climate on Energy Sector in Economic Analysis," *Journal of Applied Meteorology*, December 1981, *20* (12), 1431–1439.

Weiss, Edith Brown, "The Value of Seasonal Climate Forecasts in Managing Energy Resources," *Journal of Applied Meteorology*, 1982, *21*, 510–517.

Weron, Rafal, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, West Sussex: Wiley, December 2006.

Willis, H. Lee, *Spatial Electric Load Forecast*, 2 ed., New York, NY: Marcel Dekker, 2002.