5-2022

Forecasting Electricity Load in New Jersey with Artificial Neural Networks

Erik W. Raab

Abstract

Load forecasting is an important tool for both the energy and environmental sectors. It has progressed hand-in-hand with machine learning innovation, where recurrent neural networks, a type of artificial neural network, is primarily used. This thesis compares progressively complex, feed-forward artificial neural networks using a mix of weather and temporal data. We demonstrate that electrical load in New Jersey can be reliably predicted using memory-less algorithms with minimal predictors drawn from preexisting public data sources. The methods used in this thesis could be used to build competitive load forecasting models in other states, and if included in diverse model ensembles, may generate significant improvements.

MONTCLAIR STATE UNIVERSITY

Forecasting Electricity Load in New Jersey with Artificial Neural Networks

by

Erik W. Raab

A Master's Thesis Submitted to the Faculty of

Montclair State University

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

May 2022

College of Science and Mathematics

Department of Applied Mathematics and Statistics

Thesis Committee:

[redacted]

Dr. Eric Forgoston
Thesis Sponsor

[redacted]

Dr. Pankaj Lal
Committee Member

[redacted]

Dr. Aparna Varde
Committee Member

FORECASTING ELECTRICITY LOAD IN NEW JERSEY WITH ARTIFICIAL NEURAL NETWORKS

A THESIS

Submitted in partial fulfillment of the requirements

For the degree of Master of Science

by

ERIK W. RAAB

Montclair State University

Montclair, NJ

May 2022

Acknowledgements

I would like to thank my advisor, Dr. Eric Forgoston, and my committee members, Dr. Pankaj Lal, and Dr. Aparna Varde for their input and their support of my work.

I would also like to thank my family and friends as the source of my motivation and diligence.

Contents

List of Figures

List of Tables

1 Introduction

To keep civilization running, especially considering the impact of waste on climate change, one requires the efficient management of energy. The generation of electricity is not instantaneous and thus cannot effectively react to changes in electricity demand without advanced planning. Moving and storing electricity is expensive and prone to losses, and therefore is a poor substitute for a lack of foresight. Hence, the forecasting of electricity demand is a powerful solution for reducing costs, waste, and environmental impacts without sacrificing service quality [6].

Unfortunately, the behavior of electricity demand across a state is complex. It is the sum of millions of individual behaviors, where each individual may be represented as a unique function characterizing the impact of different factors on their behavior [7]. To approach this problem statistically, without considerable expenditure of resources and time, is to give way to an incredible amount of assumption and simplification due to its human limitations. It is for this reason that machine learning methods in electricity demand forecasting have recently become prevalent [6].

Machine learning methods are, in some ways, a black box technology whose conclusions experts are often unable to explain [8–10]. For this reason, machine learning problems are approached empirically, with the performance comparison of a variety of methods to determine the most effective solutions.

This thesis explores the effectiveness of limited public data using artificial neural networks (ANN) in forecasting hourly electricity load within New Jersey. Its purpose is to create a tenable data set from existing public data to train and evaluate ANN models for an optimal yet accessible solution. In performing this study, we encountered the following five research questions:

1. Is individual level data collection required to make accurate predictions at a state level?

2. Regarding electricity demand, does past behavior influence future behavior. In other words, should these forecasting models have memory?

3. Can load behavior related to time be better captured by translating time into multiple indicator variables?

4. To what extent is electricity demand variance explained by temperature and time?

5. Should different seasons be defined by load behavior, temperature mean, or temperature variance?

Section 2 provides background information related to this work, including electricity, predictors, machine learning, and related works. Section 3 contains an overview of the methods used in this work including the programming tools, analysis of inputs, data selection, preprocessing, and the model building process. The results, at both the regional and state levels, are presented in section 4, and section 5 closes with a discussion on the reliability of the methods used, interpretation of the results, research questions, and future research.

2 Background

2.1 New Jersey Electricity Infrastructure

The New Jersey electricity infrastructure is defined by the organizations who generate, transport, and distribute electricity. This thesis is primarily focused on the latter two branches given their access to and management of relevant data. New Jersey is split into four regions called load zones as seen in Figure 1. These load zones are defined and managed by their respective electricity distribution companies: Rockland Electric Company (RE), Jersey City Power and Lights (JCPL), Public Service Enterprise Group (PSEG), and Atlantic Electric Company (AEC). These electricity distribution companies operate within the Pennsylvania-New Jersey-Maryland (PJM) Interconnection grid, now known simply as PJM. PJM is the electricity transmission company and competitive wholesale electricity market for New Jersey. We may think of PJM as the electrical, physical, and financial connections between electricity generation and distribution companies. As such, it publishes electricity generation, demand, and pricing data, including hourly metered load values measured in megawatts (MW) for each load zone [1].

2.2 Electricity Demand and Load

Depending on who is using the word, demand and load may mean subtly different things. The U.S. Energy Information Administration (EIA) defines electricity demand as "The rate at which energy is delivered to loads and scheduling points by generation, transmission, and distribution facilities," which aligns with their definition for electric load as "An end-use device or customer that receives power from the electric system." [7] However, PJM's glossary defines load as nearly synonymous to demand except in scale; "Demand is the usage or consumption of electricity on a power system. Demand is generally expressed in kilowatt-hours or megawatt-hours. Load is the overall usage or consumption of electricity on a power supply. Load is generally expressed in kilowatt-hours or megawatt-hours." [1]

Typically electricity demand is represented as the amount of electricity requested by a community over a time period while electricity load is the measured amount of electricity being consumed by that community over a time period. Obtaining demand is ideal since it gives a complete picture of a community's electricity needs including those without access or the ability to consume. However, it is difficult to measure, which is why the industry standard is to use electricity load as a proxy for electricity demand. For this thesis we will use PJM's definitions and hereon will use load as a proxy for demand.

2.3 Machine Learning

Artificial neural networks (ANN) are computing systems named for their resemblance to biological neural networks found in the brain. The field of machine learning encompasses ANNs because these computing systems can build models from training

Figure 1: A map of New Jersey, published by PJM [1], illustrating the four load zones within the state. Each load zone is defined and managed by its respective electricity distribution company: Rockland Electric Company (RE), Jersey City Power and Lights (JCPL), Public Service Enterprise Group (PSEG), and Atlantic Electric Company (AEC).

data which are able to make decisions not explicitly defined by their programming [2]. ANNs with multiple layers belong to the field of deep learning, a sub-field of machine learning, which refers to the multiple layers used in ANNs to identify and interpret features from the inputs [8, 9]. We can visualize these two concepts with Figure 2. A diagram illustrating a simple ANN can be seen in Figure 3(a), noting that a hidden layer is so named because its values are not directly obtained from the models' inputs or outputs. Since the first ANN was developed in 1958 by Frank Rosenblatt [11], ANNs have found wide-spread use and continuous evolution. ArXiv reports an increased growth rate for submissions in computer science from 2015 to the present primarily driven by machine learning, computer vision, and computational linguistics. Machine learning itself contributed over 1,000 ArXiv submissions in 2015 increasing

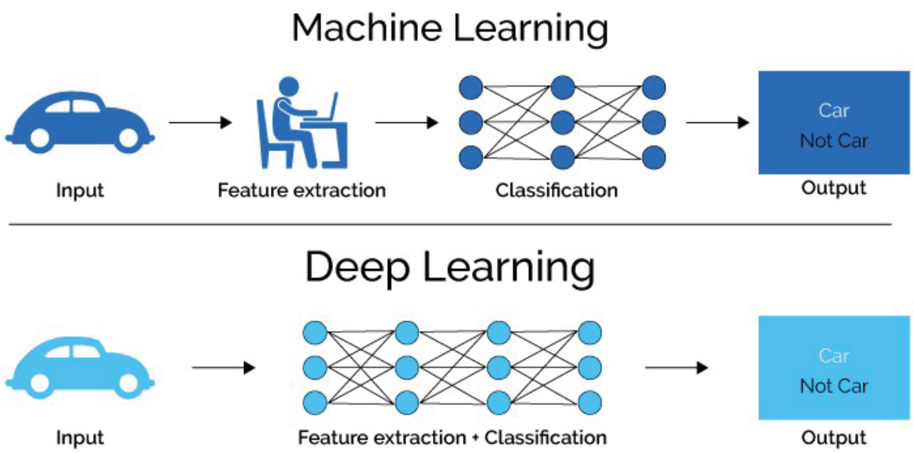


Figure 2: Example representation of machine learning and deep learning workflows [2].

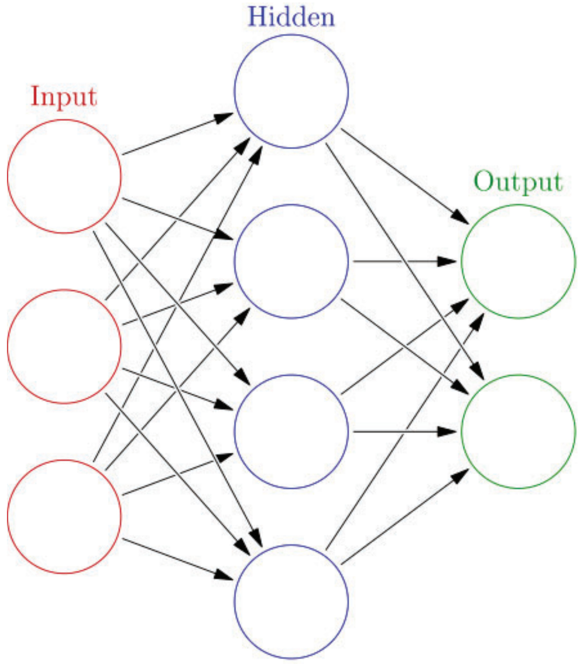
to over 11,000 in both 2020 and 2021 [12].

Recurrent neural networks (RNN) are a common variant of ANN found in load forecasting literature. This is likely because RNNs were adapted to work with sequential and time-series data in which data points are not independent. RNNs are specialized with their ability to store and consider information from past states. Their cyclic structure was developed to mimic memory. In contrast, other ANNs are feed-forward networks, meaning their connections do not form a cycle, and thus can only consider information from a current state [8].

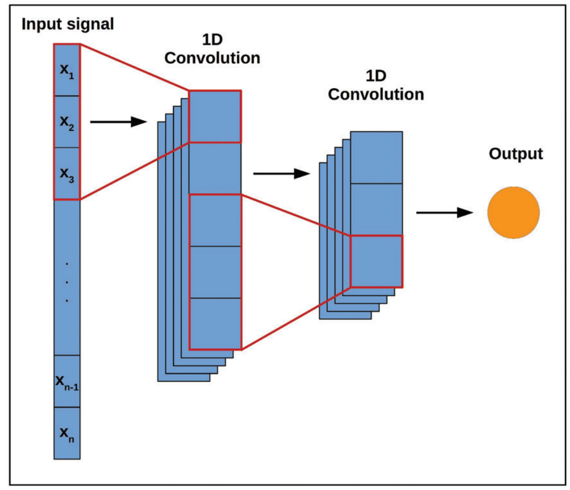
Convolutional neural networks (CNN) have risen in popularity in recent years, especially with their success in computer vision, due to their improved ability to detect and generalize features or patterns in data. CNNs are named for the convolution operations they use to transform multiple input values into a new value to better capture features. These operations can be adjusted to transform inputs in various ways so the model can consider the data from different "perspectives". A one-dimensional CNN is an architecture so named because it was designed to receive one-dimensional inputs like time series and signal data but it is still a feed-forward network [8, 10]. Figure 3(b) shows an example two-layer, one-dimensional CNN with a filter size of 3. For this thesis we focus on a CNN's lack of memory as a feed-forward network while having the ability to extract high level features from multiple points in time using convolutions.

2.4 Predictors

When an ANN is trained it is given data in the form of predictors that the user believes the ANN will find helpful in making a decision along with target values to teach the ANN how to make that decision. In this work, we construct regression ANNs so that each model builds a complex formula from the predictors it is given with the goal of estimating the target value for those predictors. Since an ANN may find a predictor unhelpful and thus reduce the weight on it to ignore it, the addition of a predictor is often associated with access to it and available computational power.



(a)



(b)

Figure 3: Example structures of (a) an ANN [3] and (b) a one-dimensional CNN [4]. The CNN uses a filter size of three which transforms three input signals into a single signal.

2.4.1 Demographics

In the United States, detailed demographic data can be difficult to obtain due to privacy laws. Although the US Census Bureau does publish demographic data, New Jersey load zones and political boundaries are not a close fit except at the state level. Demographic factors are usually considered only in long-term forecasting methods, since individual and regional demographics are unlikely to change quickly. As a result, demographics are unlikely to explain short-term load variance [6]. A common demographic predictor is the class of load consumer: residential, commercial, and industrial [1,6,13]. Consumer class is regarded as a strong predictor due to the clear differences in electricity load and consumption patterns over time from an industrial to a residential to a commercial consumer. While consumer class could be useful in short-term load forecasting, at a state level it loses this ability since the proportion of any consumption class isn't likely to change significantly over the short-term.

2.4.2 Weather

Weather is a well-established predictor, especially in short-term load forecast models [6, 13–15]. It is regarded that variance, unrelated to human behavior, in hourly load is primarily driven by heating, ventilation, and air conditioning system (HVAC) fluctuations as these systems react to the weather [7]. Common weather predictors are temperature, humidity, wind speed, wind directions, cloudiness, precipitation, and severity of the weather event.

Weather events such as a quick freeze, snow, hurricanes, and even a beautiful day alter human behavior and thus load patterns. This can result in a sudden excess of electricity generation or demand for a season and has even crippled electricity infrastructure. While clearly relevant, it is difficult to quantify the wide variety of weather events aside from providing temperature, humidity, wind speed, wind directions, precipitation, and cloudiness. This may explain why so few models include an indicator of weather events as a predictor.

In many cases, humidity, the amount of water vapor present in air at a given temperature, doesn't have a significant impact on HVAC systems. However, humidity can alter the required energy to heat or cool the air and its impact on perceived temperature, the heat index, can influence load indirectly through human behavior [1]. PJM explains absolute humidity as a potential reason for a difference in relationship between winter and summer loads with temperature. Since cold air can hold less water, on average it requires less energy to influence compared to hot air which has the potential to hold more water.

Many homes and businesses have supplemented their electricity consumption with solar panels. This means cloudiness, a measure of how clear a sky is, may be relevant to predicting local electricity generation and therefore a lower demand from PJM. However, ownership of solar panels can be regarded as a demographic factor and as such is unlikely to influence short-term forecasting at the state level.

Wind speed is another weather variable that does not change temperature directly but is thought to influence load through human behavior, increased heat loss during the winter, and potential offset from local wind generation.

2.4.3 Time

If HVAC systems account for much of the hourly load variance unrelated to human behavior, then time is a strong predictor for human activity. As expected, time is frequently used in load forecasting although often only as an index for observations rather than as a predictor. When time is a predictor, it is usually translated to day of the week and hour of the day [6, 13, 14, 16, 17]. PJM reports separate errors for winter and summer when comparing model results. Although it is unclear if they use the same model for both seasons, it is clear that season plays an important role in their model performance [1].

2.5 Related Works

ANNs are used for classification and regression in many fields, and network architectures from one field can be adapted for use in another field. Karthikeyan et al. for example, use several network designs including a variation of the ResNet used in this work to diagnose X-ray images [18]. Even limiting consideration to ANNs applied to the load forecasting problem, there is no shortage of approaches. We previously mentioned differences in terms of scope, network designs, and inputs. One also sees different types of forecasting outputs such as point and probabilistic forecasting. While many works have been focused on point forecasting, recent years have seen an increase in probabilistic forecasting [13, 14, 19, 20]. Ding et al. perform probabilistic load forecasting using a deep residual network with Inception modules [21]. Further, they test their model robustness using predicted temperature values to demonstrate its resistance to temperature or input uncertainty. Similarly, the probabilistic load forecasting model developed by Chen et al. consisted of a deep residual network with Monte Carlo dropout [22]. However, Chen et al. applied an ensemble strategy to reduce the variance of model performances.

For point forecasting, many different model architectures have been explored including combinations of architectures known as model ensembles. Recent model ensembles often include a CNN for feature identification matched with a form of RNN [13, 23–25]. Lang et al. forecasted short-term loads for dozens to hundreds of consumers using a single one-dimensional convolutional layer with dense layers which append predicted points into the inputs for the next prediction [26]. Sadaei et al. proposed a fuzzy time series methodology to create images of inputs and reduce over-fitting within the convolutional layers [27]. Tong et al. introduced the Inception structure to temporal CNNs and compared its ultra-short-term performance against 10 other models using data from Jiangsu, China and PJM [28]. Kim et al. applied a variety of temporal features including sine and cosine hour of day within an interesting network architecture: two RNN layers, a long short-term memory in this case, accept inputs before handing off to a single inception module and finally two dense layers [29]. Their goal was to mitigate the weakness of RNNs using an Inception module.

3 Methods

3.1 Programming Tools

All analysis, preprocessing, model building, and figure generation was performed using the R programming language [30]. Keras and Tensorflow are powerful and widely used machine learning libraries [31]. They are the primary packages that we used to construct, train, and validate the ANNs present in this work [32, 33]. It is worth noting that one advantage of using R, Keras, and Tensorflow is due to the fact that each has a significant presence within open-source support communities and academic journals [31, 34].

3.2 Data selection

Obtaining electricity demand data in the United States is a challenge as detailed data is typically not publicly available due to a number of concerns including privacy. Given the scope of our study, PJM's hourly metered load data from New Jersey 2020 was used [35]. PJM collects detailed load data from the four electricity distribution companies in New Jersey and summarizes this data into "[megawatt]-hour net energy

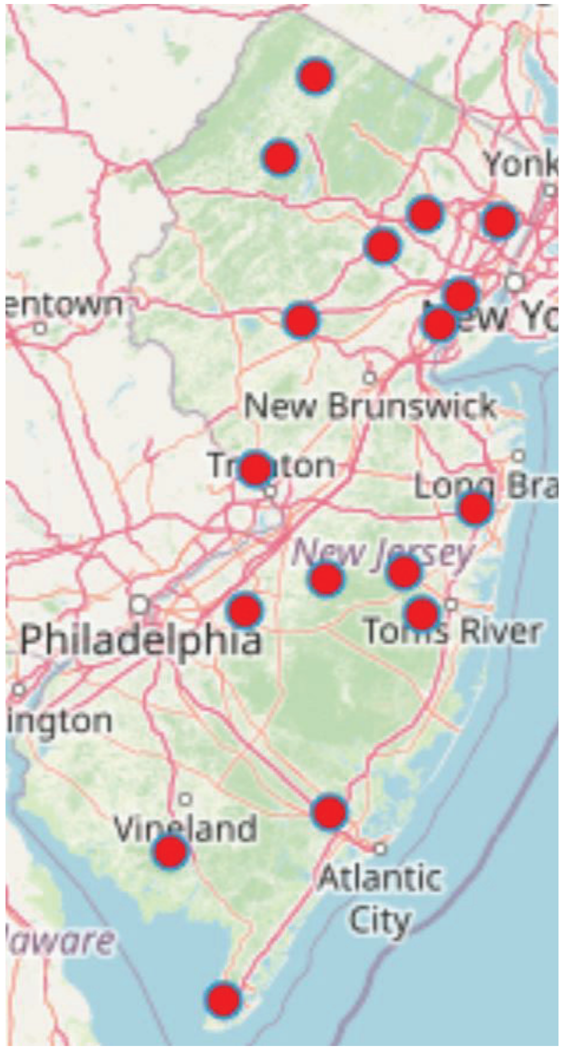


Figure 4: Map of New Jersey which illustrates the locations of ASOS monitoring stations in New Jersey as red points.

for load as consumed by the [load zones].” All load data included in training has already surpassed PJM’s 90-day cutoff for data adjustments. This load data was downloaded as a comma-separated values (csv) file using PJM’s Data Miner 2 in Coordinated Universal Time (UTC).

Hourly, historic weather records are publicly available from the National Oceanic and Atmospheric Administration’s (NOAA) Automated Surface Observing System (ASOS) which can be retrieved from the Iowa Environmental Mesonet(IEM) [36, 37]. The IEM’s ASOS archive is comprised of weather observations from monitoring stations across the United States and the world [37]. There are 17 ASOS stations located at airports in population centers of New Jersey and two stations on the southwest border of New Jersey where there are New Jersey population centers that lack airports. The monitoring station locations in New Jersey can be seen in Figure 4. These weather monitoring stations were matched geographically to load zones they bordered or intersected with to find the average weather values for that region at any given hour. This also means some weather monitoring stations were associated with multiple load zone models. Four weather predictors are extracted from the ASOS data along with the time that each measurement was collected. These weather predictors are temperature (in degrees Fahrenheit), relative humidity (in percentage), precipitation (in inches/hour), and wind speed (in knots). The hourly (in UTC) ASOS weather station data was downloaded from the IEM by station and year as a csv file.

3.3 Analysis of Inputs

As described in Section 2.3, an ANN is a black box and does not inform its users how predictors are used. Thus, an exploration of inputs can aid the discussion of model results. With the goal of the model being the prediction of load, the analysis of predictors is from the same perspective. Two types of predictors are used in this work: temporal and weather.

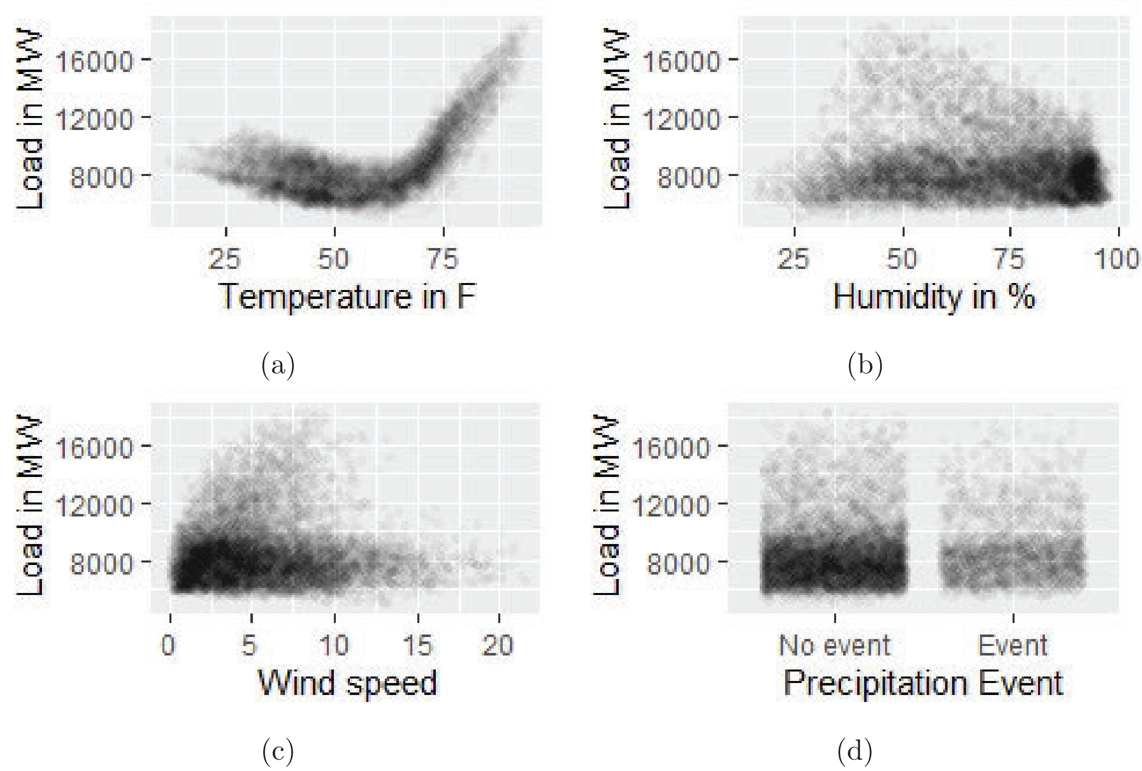


Figure 5: 2020 hourly load in New Jersey versus (a) temperature (degrees Fahrenheit), (b) relative humidity (percentage), (c) wind speed (knots), and (d) precipitation event indicator. In (a), points have a transparency of 0.0125 so that a solid black point is comprised of 80 overlapping points. In (b), (c), and (d) points have a transparency of 0.02 so that a solid black point is comprised of 50 overlapping points. As a categorical factor, (d) has been plotted with jitter to reduce the density of points and improve visualization of the data.

3.3.1 Weather Inputs

Plotting temperature against load shows interesting behavior as seen in Figure 5(a). The lowest average loads occur near 60 degrees Fahrenheit, and the increase in load as temperatures drops below 60 degrees is less than the increase in load as temperatures increase above 60 degrees. Also note that the load variance appears relatively constant throughout. While 60 degrees is too low to be a comfortable room temperature, it is a common building temperature for when no one is present or awake. According to the 2020 ASOS data, New Jersey's average relative humidity is 68.89% and the distribution changes little over the year with some periodicity across the 24 hour day [36]. One can see in Figure 5(b) that relative humidity does not appear to significantly inform load variance. The same can be observed from wind speed and precipitation events respectively in Figures 5(c) and 5(d). However it is possible that the interaction of these weather variables with each other or time could inform load fluctuations. The importance of these variables will be determined in the model building phase based on the difference in performance for models with and without these variables.

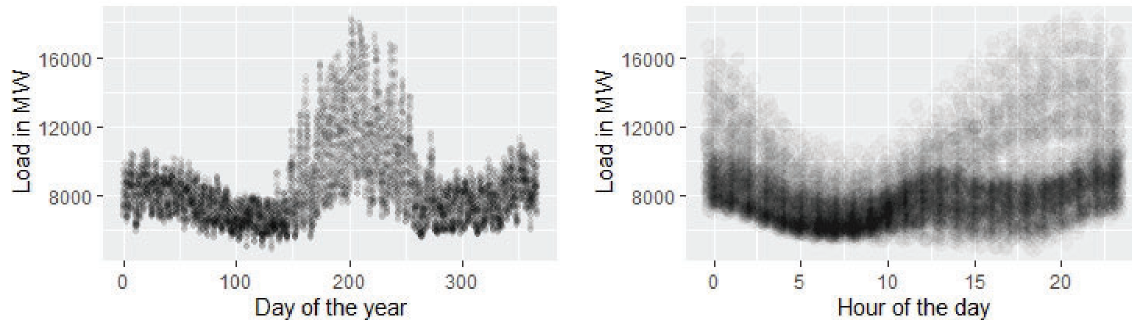
3.3.2 Time Inputs

A periodogram of PJM's hourly load data from New Jersey in 2020 reveals three large peaks at the frequencies 0.000111, 0.000222, and 0.041666. These values translate to expected periodicities of 9,000, 4,500, and 24 hours respectively. As a leap year, there are 8784 hours in 2020, so it is possible that the 9,000 hour periodicity reflects an annual load cycle as observed in Figure 6(a). The 4,500 hour periodicity likely represents a rounded biannual periodicity seen in Figure 6(a) as the two peaks and valleys. Evidence for annual and biannual periodicity supports the use of daily, monthly, yearly, and seasonal variables as predictors. The 24-hour load periodicity can be observed in Figure 6(b) where load measurements are grouped by the hour of day in which they occurred.

Plotting load over the days of the week shows a slight decrease on weekends compared to weekdays. Comparing loads for business days against weekends and holidays one can also see there is an increased load on average for business days compared to non-business days. Figures 7(a) and 7(b) demonstrate it is plausible that either day of the week or a business day indicator might aid load forecasting.

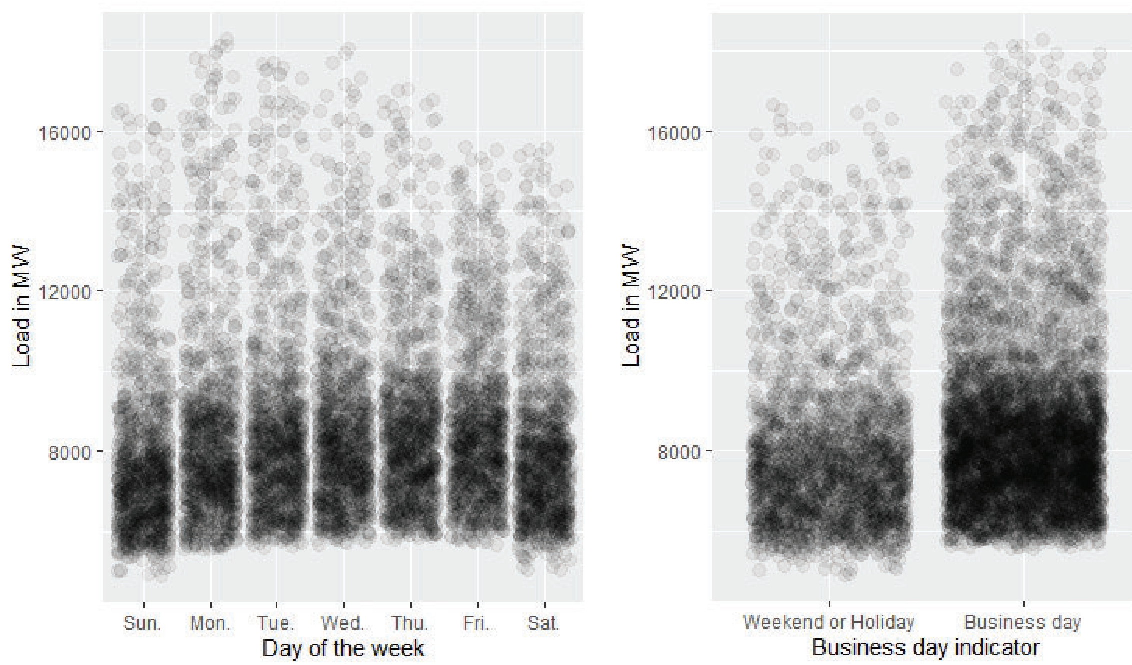
Plotting load by month, in Figure 8, we expected to see something similar to days of the year and found perhaps a clearer example of seasonal load behavior. We expected spring and fall months, classified as transitory months, to have the lowest load and variance since they represent the mildest time of year for New Jersey. May, initially classified as a summer month, appears to share load characteristics with transitory and summer months. This realization led us to develop a more informative definition of season, which we now describe.

Improving model performance over the summer time appears to be critical for improving load forecasting models. Despite accounting for weather variables and the time of year, load forecasting models consistently report larger errors during the



(a) (b)

Figure 6: 2020 hourly load in New Jersey grouped by (a) day of the year, and (b) hour of the day. In (a) points have a transparency of 0.05 so that a solid black point is comprised of 20 overlapping points, while in (b) points have a transparency of 0.02 so that a solid black point is comprised of 50 overlapping points. Additionally, (b) has been plotted with jitter to reduce the density of points and improve visualization of the data.



(a) (b)

Figure 7: 2020 hourly load in New Jersey grouped by (a) day of the week and (b) binary business day indicator. In both (a) and (b), points have a transparency of 0.05 so that a solid black point is comprised of 20 overlapping points. Additionally, both (a) and (b) have been plotted with jitter to reduce the density of points and improve visualization of the data.

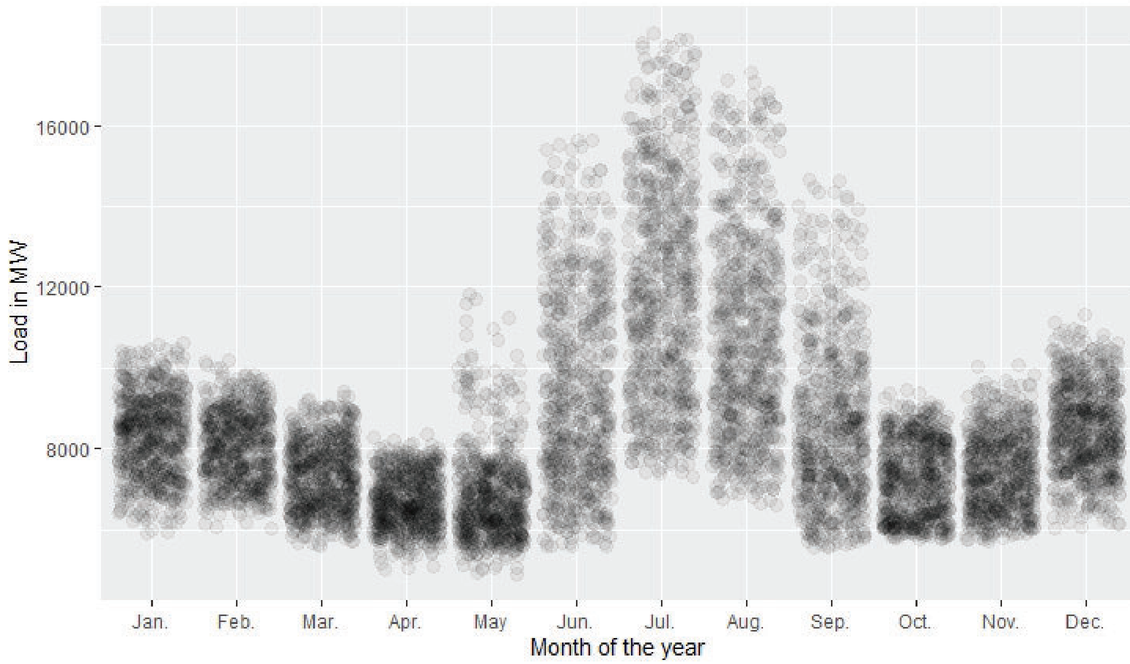
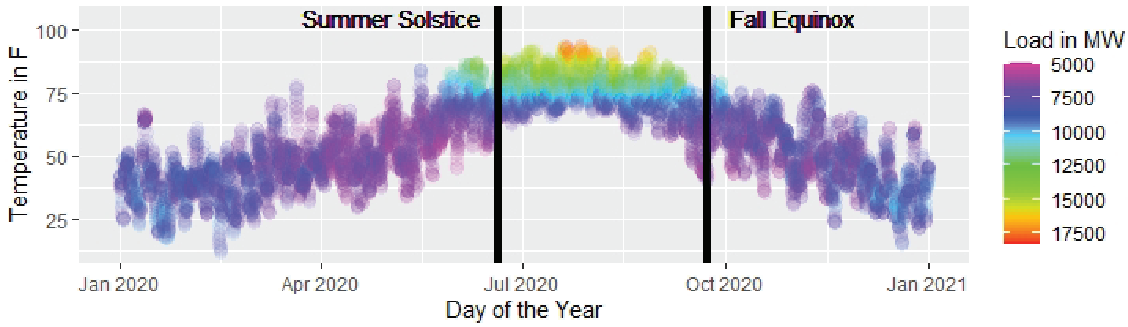


Figure 8: 2020 hourly load observations grouped by month in New Jersey. Points have a transparency of 0.05 so that a solid black point is comprised of 20 overlapping points. Additionally, points have been plotted with jitter to reduce the density of points and improve visualization of the data.

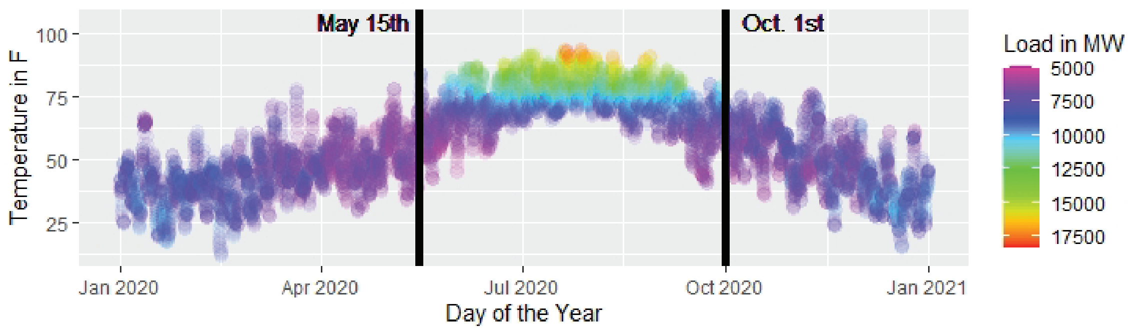
summer than during the winter. Comparing hours with the highest absolute errors across models, we find very little in common other than season, high temperatures, and high load variance. National Geographic and the UK's National Physical Laboratory (NPL) agree that seasons can be defined as periods of time during a year which are separated by climate or light patterns related to the movement of the sun around the Earth [38, 39]. A standard definition of seasonal changes is the four solstices/equinoxes. Figure 9(a) highlights the location of the summer solstice and fall equinox in relation to 2020 load and temperature readings. Note that the window defined by the summer solstice and fall equinox doesn't appear to capture the entirety of the summer's characteristic temperature and load behaviors. Defining summer in terms of days with high load variance produces Figure 9(b), in which summer's high daily load variance is centered between May 15th and October 1st. In fact, if season is overlaid with temperature and load behavior as in Figure 10, we find that load and temperature observations are strongly clustered by season, further encouraging the use of an encoded season predictor. This load oriented definition for summer was used to train models in this thesis.

3.4 Preprocessing

To ease preprocessing, the date-time information for the load data and each weather station data set was formatted to the POSIXlt data class in the R programming



(a)



(b)

Figure 9: Temperature as a function of time, and colored by load value. One can see that the summer has low temperature variance but high load variance. Both (a) and (b) have vertical lines indicating the boundaries of summer as defined by the summer solstice and fall equinox in (a) and load variance in (b). Points have a transparency of 0.05 so that a solid point is comprised of 20 overlapping points. Additionally, points have been plotted with jitter to reduce the density of points and improve visualization of the data.

language. Then every weather station data set was stacked by rows into a single set of weather data. At times, a weather station may have as many as 15 consecutive hours of missing, trace, or bad data. The ASOS encoding of missing and trace values is the letter M and T respectively. Thus, by forcing the weather variables to be treated as numeric, all missing and trace values are set to NA. This was confirmed by checking the count of M and T observations against the count of introduced NA values. Bad values, such as negative humidity, were also replaced with NA by identifying numeric values outside the possible range of each variable.

The load data consisted of a single observation per hour (MW-hour). Since the weather observations were associated with the actual second that a given reading was recorded, the weather observation times were rounded to the nearest hour. At the weather monitoring station level this resulted in some hours with multiple observations and other hours without observations. To account for multiplicity, the weather measurements were grouped by station and by hour, and then were averaged to obtain a single weather observation per station. PJM's hourly metered load data contains

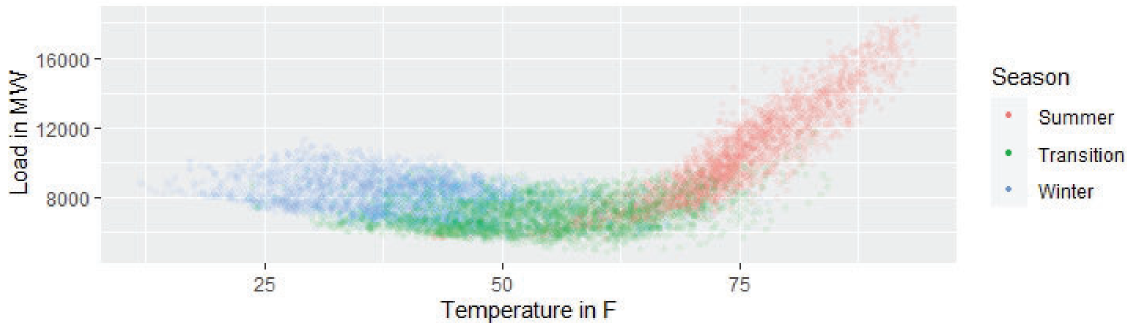


Figure 10: 2020 hourly load observations in New Jersey plotted against temperature in degrees Fahrenheit, and colored by season to demonstrate the seasonal clustering of load and temperature observations. Points have a transparency of 0.05 so that a solid point is comprised of 20 overlapping points. Additionally, points have been plotted with jitter to reduce the density of points and improve visualization of the data.

no missing values. The PJM metadata explains the lack of missing values with the following statement: "This data contains [electricity distribution] company submitted values or PJM generated estimated values substituted where company data is unavailable."

At this point in the preprocessing, each hour will have the same number of weather observations as there are weather stations, although some observations may still report missing values. Depending on the region of the model, either New Jersey or a particular load zone, the appropriate stations were selected first, and the weather data was grouped by hour. The weather station observations were averaged so that each hour had a single weather observation relevant to its region, and the weather data set can be joined with the load data set.

A left outer join is performed on the load data, appending the hourly weather observations with matching hours. An exponentially weighted and centered moving average was applied using four nearest neighbors and a maximum gap of three consecutive NA values. Any remaining hours still reporting NA were dropped from the merged data set.

Considering that the presence of rainfall at any station in a region means the average rainfall for that region is non-zero this numerical variable was transformed into a binary indicator for the presence of any precipitation. The POSIXlt data class time variable was sub-divided into thirteen different time variables to account for the different ways time can influence load. This includes hour of the day as an integer, a sine wave, a cosine wave, common work shifts which are one-hot encoded into three variables, day of the week as an integer, a business day indicator, month of the year as an integer, and season which is one-hot encoded into three variables.

The hour of the day is presented as an integer from 0 through 23 as well as a sine and cosine fluctuation to ensure there is continuity from hour 23 to hour 0. The formulae used to generate these alternate variables are $\sin(\text{hour} \cdot (2\pi/24))$ and $\cos(\text{hour} \cdot (2\pi/24))$. Model training was performed using the load data, the four weather variables described previously, and the thirteen time variables described previously. The rows, which contain the relevant data at each hour, were shuffled to avoid overfitting the model to the features present in the first half of the year. Finally, the merged data set was divided so that 82% of the data was used for training, 18% for validation, and 10% of the data was used for testing. Feature means and standard deviations are captured from the training input set and used to normalize all features in both the training and test inputs data sets. The Keras functional API requires a three-dimensional tensor despite only using two dimensions in the training. Thus, the last preprocessing step was to add a third dimension to the training input and target data sets such that their dimension is increased in the following way: (number of observations or hours, number of training variables, 1).

3.5 Model Building

Over forty unique networks and more than 400 models were compared throughout this study. Model building began with data from the smallest load zone and a network designed with the least number of input variables and simplest architecture. Then network and input complexity was increased with subsequent iterations. If an added network feature consistently improved performance it was prioritized in following network iterations. The mean absolute error (MAE) was used as the primary metric for model performance and network comparisons within a load zone. Since the range of load varies significantly from region to region the most reliable measure of quality between load zones is the relative error. The learning rate used in this work is 0.001. Where RMSProp was applied, the network was trained without momentum and where Adam was applied an exponential decay rate for the first moment is 0.9 and for the second moment is 0.999.

The first model design used data from the Rockland Electric Company (RE) load zone using only temperature as the predictor variable, and a Keras sequential API with 2 dense layers consisting of 8 and 4 nodes respectively. Table 1 shows the development of network designs. Initial dense layers contained 16 nodes and subsequent dense layers had half as many nodes. Except as noted in Table 1, convolutional layers had a kernal size of three, and named networks reflected their namesake designs. Varying activation and optimization functions showed no significant changes so RMSProp and ReLu persisted to the final networks. Beyond ANN iteration 15 in Table 1, increasing the depth of a given network yielded limited results except in the case of ResNet and InceptionTime [40]. At the end of the study the 1D ResNet, Inception, and InceptionTime inspired architectures performed the best using temperature, wind speed, relative humidity, a precipitation indicator, hour of the day, sine hour, cosine hour, work shift, business day, day of the week, day of the year, month of the year, and season.

Due to the randomization in both shuffling observations and the selection of training, validation, and testing data sets, each resulting model's training and thus performance will be different. Therefore, it is critical to train a number of models with a shared design to estimate the skill or typical model performance derived from that design. Further, training multiple models can also help identify a network's tendency

Iteration	Avg. MAE	Summary of Major Changes in ANN Design
1	87.55	Two layer sequential ANN, rmsprop, and temperature.
2	44.90	Added relative humidity, wind speed, and precipitation in inches.
3	34.84	Added day of the year, hour of the day, work shifts, day of the week, business day, month of the year, and season.
4	25.67	Redefined season in terms of load variance.
5	26.96	One-hot encoded hour of day.
6	23.13	Added sine and cosine hour of the day and returned hour of day to integer.
7	23.56	Removed integer hour of the day.
8	23.18	Returned integer hour of the day and added Adam optimizer.
9	20.21	Introduced input normalization without Adam optimizer.
10	19.07	Normalize with three layers and batch size of 12.
11	18.60	Normalize with three layers and batch size of 2.
12	19.55	Re-added Adam optimizer.
13	8.38	Single layer CNN with kernel size of 3.
14	8.93	Single layer CNN with kernel size of 5.
15	8.10	Two convolutions with a max pooling layer between.
16	8.99	Two convolutions with an average pooling layer between.
17	7.09	Two convolutions with a max pooling layer between and 64 filters.
18	7.98	Two convolutions with a max pooling layer between and 32 filters.
19	7.72	Two convolutions with a max pooling layer between, ReLu activation, and 64 filters; one dense layer with linear activation.
20	8.48	Two convolutions with 32 filters, kernel size of 5, Adam optimizer
21	8.45	Two convolutions with 32 filters, kernel size of 5, and RMSProp.
22	7.97	Two layer ResNet with 64 filters and kernel size of 2.
23	5.45	Five layer ResNet with 64 filters and kernel size of 2.
24	5.95	Single Inception v1 module.
25	6.41	Single Inception v1 module without season indicator.
26	5.52	One module InceptionTime.
27	4.59	Three module InceptionTime.

Table 1: Summary of the major changes in ANN design and respective performance metrics for the Rockland Electric Company (RE) load zone. Since each iteration listed here represents the same load zone the performance metric shown is iteration's average model mean absolute error (MAE). For context, the range of loads in 2020 for RE was from 84.8 to 397.5 MW and an average home load is 2 to 4 kW per the New Jersey Energy Master Plan [5].

to overfit the data. A k-fold cross-validation was the primary vehicle to meet this goal. Since 2020 is comprised of 8,784 hours there are plenty of observations to cross-validate ten models at a time with a 90% training-testing split. Note in Figure 1 that the PSEG load zone splits the JCPL load zone into two pieces. Since the JCPL load data from PJM is not independently georeferenced for each part, models were developed for JCPL and PSEG combined instead of JCPL independently.

4 Results

4.1 Model Development

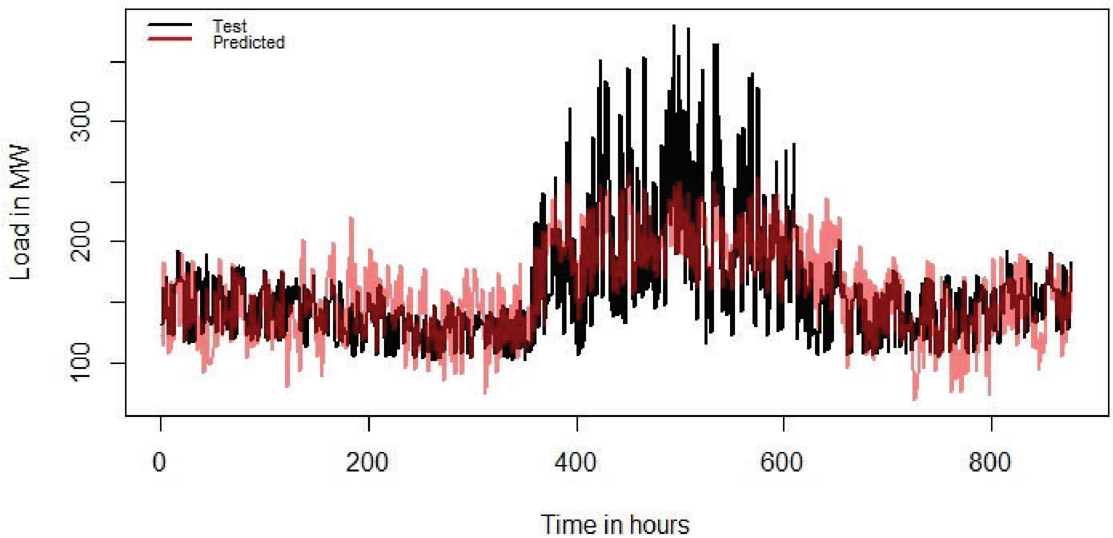
The early model designs had two things in common: (1) they were sequential models, which (2) struggled during the summer season. The progression in performance from the fourth design to the eleventh design (see Table 1), seen in Figures 11a and 11b respectively, shows an overall improvement in load forecasting as well as improved ability to capture the variance of the summer season. The introduction of convolutions using convolutional neural network (CNN) architectures resulted in a significant performance increase especially in regards to seasonal adaptability. Figures 11b and 12 illustrate the change in performance from ANN to CNN from network iteration 11 to 13. Finally a five layer ResNet, single module Inception network, and a three module InceptionTime network were found to be the top performers.

4.2 Final designs

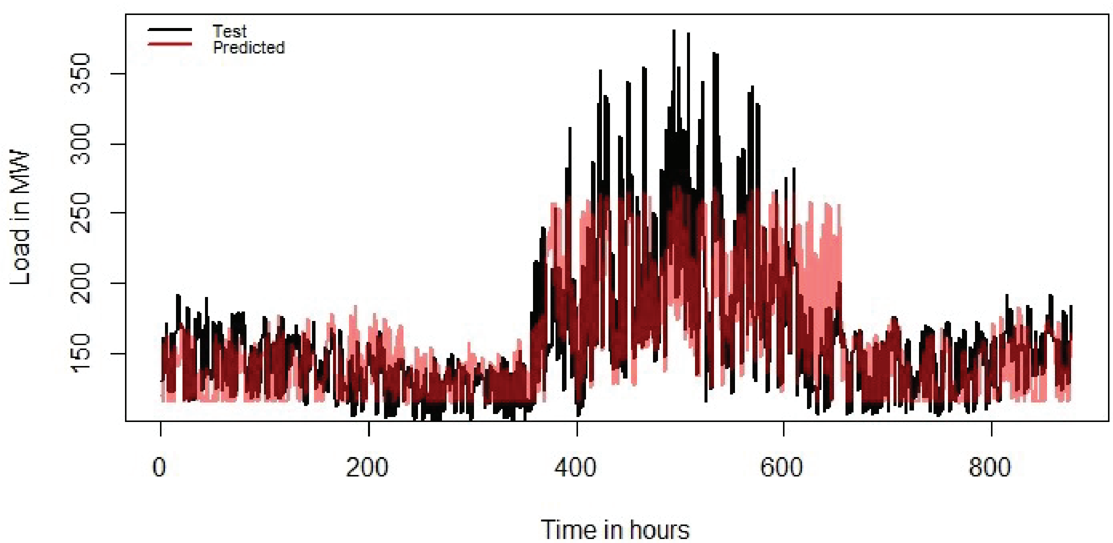
The top performing networks for New Jersey as a statewide region were inspired by Inception, ResNet, and InceptionTime. Their performance metrics are shown in Table 2. The New Jersey models designed after InceptionTime achieved an average relative error of 3.16%. Figures 13(a) and 13(b) demonstrate its performance. In particular, note the distribution of error spikes and lessened error concentration during the summer season. Table 3 lists the smaller load zones and their respective performance metrics using the InceptionTime architecture. Since PJM reports relative errors between 1% and 6% for various load forecasts, the model performances produced in this work appear to be competitive with industry.

Architecture	Average Model Mean Absolute Error	Average Model Relative Error
ResNet	281.063 MW	3.24%
Inception	394.82	4.57%
InceptionTime	281.11	3.16%

Table 2: Average architecture performance for New Jersey statewide models. The average model mean absolute error is the absolute difference between predicted and actual load values, while relative error is the absolute error at a given hour divided by the actual load value.



(a)



(b)

Figure 11: Load versus time of the test data (black) compared with the model prediction (red). The predicted load was found using three different Rockland Electric Company models trained and tested using the 4th, 11th, and 13th network iterations (see Table 1). The 4th iteration (Figure 11a) is a two dense layer sequential network, the 11th iteration (Figure 11b) feeds normalized inputs into three dense layers, and the 13th iteration (see Figure 12) is a single convolutional layer in a sequential network.

5 Summary

5.1 Method Reliability

The load data used in this thesis is complete and well-documented in terms of collection and validation by PJM. The weather data collected by the ASOS monitoring

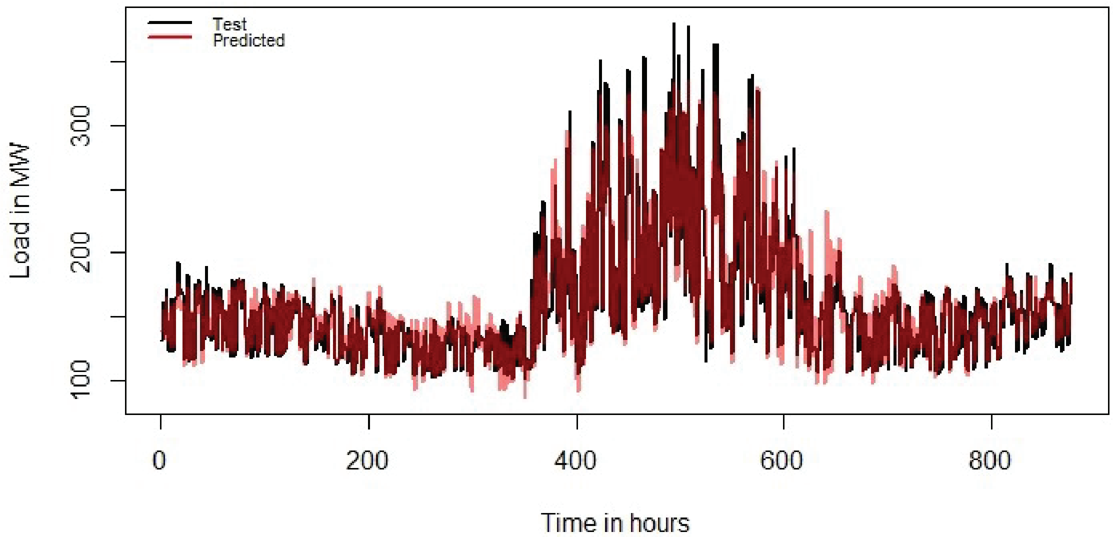
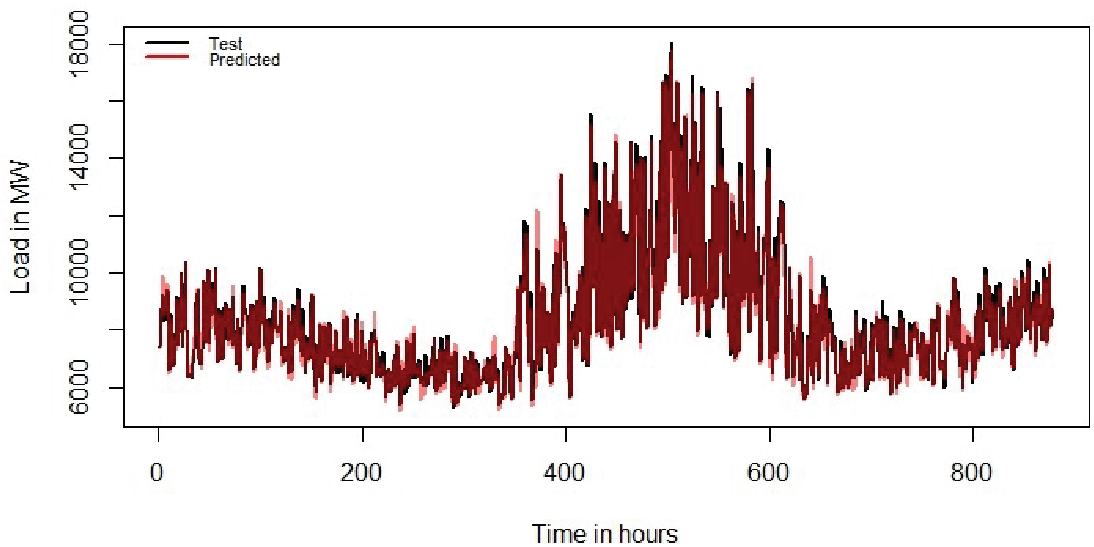


Figure 12: Load versus time of the test data (black) compared with the model prediction (red). The predicted load was found using three different Rockland Electric Company models trained and tested using the 4th, 11th, and 13th network iterations (see Table 1). The 4th iteration (see Figure 11a) is a two dense layer sequential network, the 11th iteration (see Figure 11b) feeds normalized inputs into three dense layers, and the 13th iteration (Figure 12) is a single convolutional layer in a sequential network.

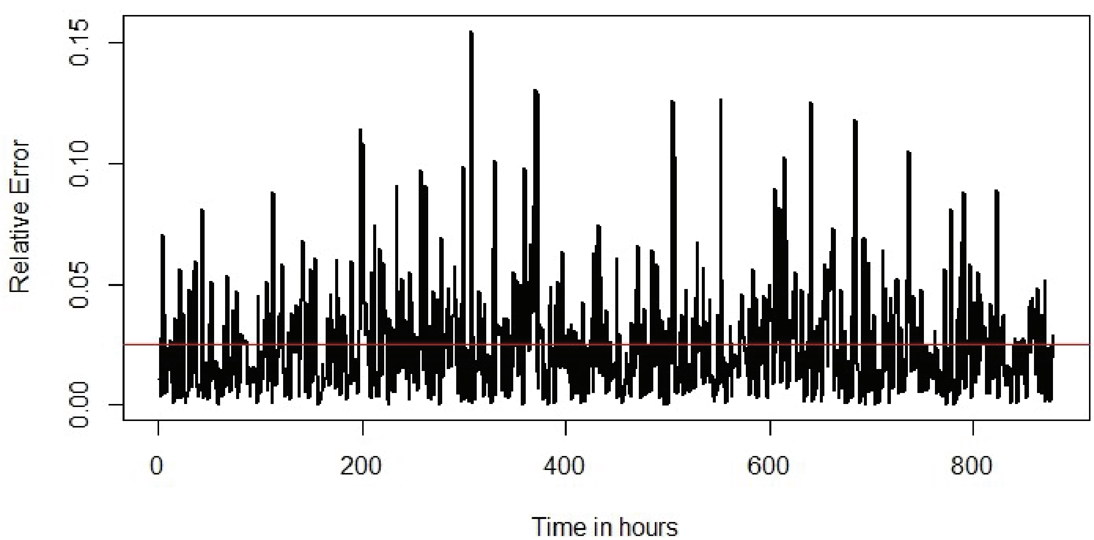
stations, while occasionally incomplete, is weighted by observations from multiple stations and is used by both NOAA and the respective airports these stations are located at. Oddly, the number of unique values for weather measurements seem rather low for their degree of precision. Temperature recordings from a single station in 2020 may have as few as 112 unique values despite over 8,000 observations over a continuous range from 12 to 95 degrees. This highlights a data redundancy concern with regards to complex relationships. An ANN learns best from diverse data sets, though it is unclear how to diagnose a model with a data redundancy concern. Fortunately we expect that aggregating weather values from twenty different stations helps to remediate this potential concern.

Load Zone	Load Range	Average Model Mean Absolute Error	Average Model Relative Error
RE	84.8 - 397.5 MW	4.59	3.50%
AEC	442.9 - 2489.5 MW	30.28	3.99%
PSEG	2950.3 - 9557.3 MW	97.74	2.57%
JCPL & PSEG	4351.5 - 15420.1 MW	152.03	3.31%

Table 3: Performance metrics of the InceptionTime ANN across load zones.



(a)



(b)

Figure 13: (a) Load versus time of the test data (black) compared with the model prediction (red). The illustrated performance involves a statewide New Jersey InceptionTime model. (b) Relative error for each of the training hours visible in (a). This particular model expressed an average relative error of 2.48%, indicated by the red horizontal line.

The preprocessing methods described in the methods section were designed with multiple quality control checks throughout the R programming script to minimize missing or false values while maximizing ANN comprehension. The quality of the preprocessed data is backed by the resulting skill of the models trained. However, seasonality error remains a concern. Even when models were assigned a season and

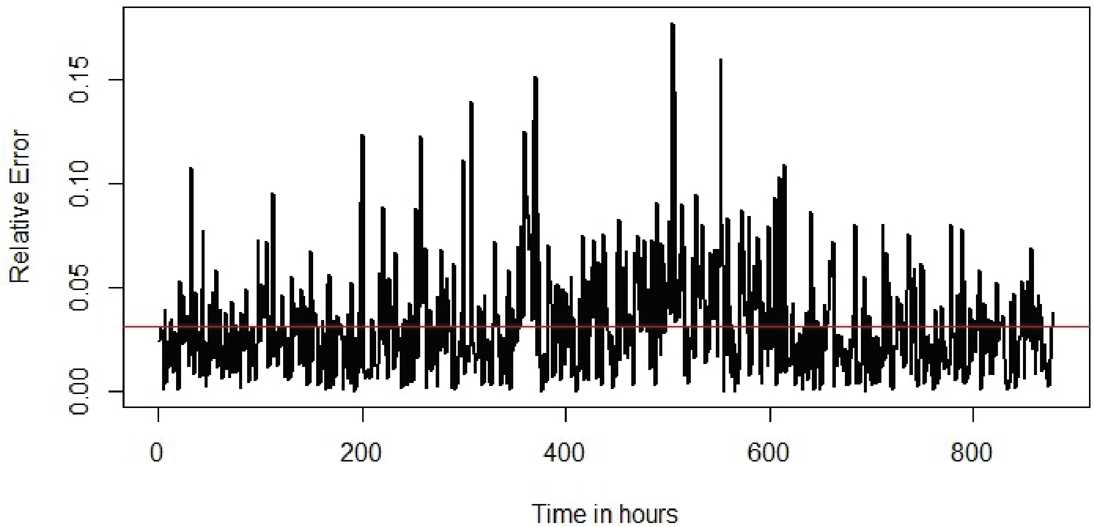


Figure 14: Relative error for each training hour for a model that uses a three module InceptionTime network. This particular model had higher than average error during the summer months and expressed an average relative error of 3.24%, indicated by the red horizontal line.

trained only on that season, the summer forecasts typically performed worse than the rest of the year. The New Jersey InceptionTime models, while occasionally suffering from seasonality error as seen in Figure 14, often overcomes this hurdle as demonstrated in Figure 13(b) with a balanced error distribution about the average relative error through the year.

5.2 Research Findings

We now revisit the overarching research questions presented in the introduction and provide responses based on the work and results presented in this thesis.

1. To make accurate predictions at a state level, is data collected at a finer scale such as individual or county required?

   The accuracy of the models produced in this thesis, even the rudimentary models, suggests that state load forecasting only requires state-scale inputs to be competitive. This matches results seen in other load forecasting papers. However, the average model relative error for PSEG is significantly lower than that of AEC as seen in Table 3. This correlates with the differences in distribution of weather monitoring stations in relation to population centers between load zones. Thus, assuming negligible differences in management, it could be theorized that small-scale weather changes play a significant role.

2. Regarding electricity demand, does past behavior influence future behavior. In other words, should these forecasting models have memory?

This thesis has shown that a memory-less load forecasting model can be successful. At a deeper level, however, our work raises the question if the information captured by a convolutional neural network is the same as that of an recurrent neural network. Moreover, would the two types of network be better off paired together, and which part should accept which inputs?

3. Can load behavior related to time be better captured by translating time into multiple indicator variables?

   In this thesis, the accuracy of models which lacked various time features was poorer than models which possessed the features. However, we did not include recurrent neural networks, nor did we cover the full breadth of possible machine learning methods. Although we are unable to fully answer the question from a general standpoint, we have shown in this work that various time features do improve model accuracy, especially in the case of CNNs.

4. To what extent is electricity demand variance explained by temperature and time?

   Based on the average model relative error, we could infer that more than 95% of state-wide load variation can be explained by weather and time features. However, each model has been trained to forecast for a region using historical data from that region, and what one regional model learns about the relationship between weather, time, and load may not fully apply to another region.

5. Should different seasons be defined by load behavior, temperature mean, or temperature variance?

   This question remains unanswered but we can confidently say that season and its definition play a role in load forecasting as an indicator for changes in climate and/or human behavior. At a minimum, defining season in terms of load variance for the model instead of month or solstice did improve the average model's performance.

5.3 Conclusions

This thesis demonstrates how a competitive ANN might be developed to forecast electricity load in New Jersey using publicly available data. We have also found it is possible to achieve a consistently high level of accuracy using a feed-forward or memory-less architecture. A significant cause for poor forecasting performance is likely the inability of a model to adequately understand load behavior during the summer. From this work, we expect that the ability to accurately classify summer in terms of load behavior and extract as much information on load behavior from time will prove helpful to developing future electricity load forecasting models.

5.4 Future research

Francois Chollet stated in their book, Deep Learning with R, that "to be competitive in [deep learning] competitions, one must use an ensemble of accurate AND diverse models." They placed heavy emphasis on diverse, collaborating models because even the most accurate models are still limited in focus and scope [10]. The immediate next step for this work would be to perform a deeper exploration of model ensembles. One possible ensemble might have a model group specialized to forecast expected load from temperature and time paired with another model group specialized to forecast the noise or variations due to human behaviors with indirect weather impacts and time features like holidays. An ensemble of point and probabilistic load forecasting models would also be an interesting addition to the field. Observations related to research questions 2, 3, and 4 suggest a stochastic neural network may also yield interesting results in a load forecasting ensemble. The addition of a model like InceptionTime may also help to reduce error associated with seasonality.

While ANN architecture and algorithm innovation continues, it is useful to also innovate which inputs are delivered to an ANN. A thorough investigation of all the information contained within time related to load behavior and interactions with weather could be useful. In PJM's 2019 forecasting reports, relative error was higher for summer than winter and the models produced in this work shared a similar flaw. No discernible pattern was detected among hours with high absolute error other than season and high load variance. Perhaps the reason is related to the size of the region being forecasted or differences in input-to-output scale since the state-wide models using load zone inputs experienced less seasonal error on average than the individual load zone models using load zone inputs. None the less, we believe a targeted study on the influence of time predictors would be beneficial to future load forecasting models. The difference in performance between PSEG and AEC suggests it would also be beneficial to investigate the impact of weather monitoring station distributions on model performance.

References

[1] PJM Load Analysis Subcommittee, "PJM RTO Load Forecast Model Accuracy," Dec 2019.

[2] S. Agrawal, "Online) 2581-9429 issn (print) 2581-xxxx," International Journal of Advanced Research in Science, Communication and Technology, 02 2021.

[3] S. Dey, S. Nandi, and G. Trivedi, "Powerplanningdl: Reliability-aware framework for on-chip power grid design using deep learning," CoRR, vol. abs/2005.01386, 2020.

[4] A. Shenfield and M. Howarth, "A novel deep learning model for the detection and identification of rolling element-bearing faults," Sensors, vol. 20, no. 18, 2020.

[5] "New Jersey Energy Master Plan," 2021. nj.gov. `https://www.nj.gov/emp/` (accessed Jan. 22, 2022).

[6] E. A. Feinberg and D. Genethliou, Load Forecasting, pp. 269–285. Boston, MA: Springer US, 2005.

[7] US Energy Information Administration, "U.S. Energy Information Administration EIA." `https://www.eia.gov`, 2020. Accessed: 2020.

[8] P. Flach, Machine Learning: The Art and Science of Algorithms That Make Sense of Data. USA: Cambridge University Press, 2012.

[9] L. Deng and D. Yu, "Deep learning: Methods and applications," Tech. Rep. MSR-TR-2014-21, Microsoft, May 2014.

[10] F. Chollet, Deep learning with R. Shelter Island, New York: Manning Publications, 2018.

[11] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain.," Psychological review, vol. 65, no. 6, p. 386, 1958.

[12] Arxiv, "Arxiv.org submissions by category since 2009+." `https://arxiv.org/about/reports/submission_category_by_year`, Aug 2021.

[13] I. Ghalehkhondabi, E. Ardjmand, G. R. Weckman, and W. A. Young, "An overview of energy demand forecasting methods published in 2005–2015," Energy Systems, vol. 8, no. 2, pp. 411–447, 2017.

[14] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," International Journal of Forecasting, vol. 32, pp. 914–938, July 2016.

[15] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," International Journal of Forecasting, vol. 32, p. 896–913, Jul 2016.

[16] S. Khan, N. Javaid, A. Chand, A. B. M. Khan, F. Rashid, and I. U. Afridi, "Electricity Load Forecasting for Each Day of Week Using Deep CNN," in Web, Artificial Intelligence and Network Applications (L. Barolli, M. Takizawa, F. Xhafa, and T. Enokido, eds.), vol. 927, pp. 1107–1119, Cham: Springer International Publishing, 2019.

[17] A. J. Balaji, D. S. Harish Ram, and B. B. Nair, "A deep learning approach to electric energy consumption modeling," Journal of Intelligent & Fuzzy Systems, vol. 36, pp. 4049–4055, Jan. 2019.

[18] D. Karthikeyan, A. S. Varde, and W. Wang, "Transfer learning for decision support in covid-19 detection from a few images in big data.," 2020 IEEE International Conference on Big Data (Big Data), Big Data (Big Data), 2020 IEEE International Conference on, pp. 4873 – 4881, 2020.

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

[20] Y. Chen, Y. Kang, Y. Chen, and Z. Wang, "Probabilistic forecasting with temporal convolutional neural network," Neurocomputing, vol. 399, pp. 491–501, July 2020.

[21] D. Aijia, L. Tingzhang, and Z. Xue, "Integration of ensemble googlenet and modified deep residual networks for short-term load forecasting.," Electronics, vol. 10, no. 2455, p. 2455, 2021.

[22] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and H. Jinliang, "Short-term load forecasting with deep residual networks," IEEE Transactions on Smart Grid, vol. PP, 05 2018.

[23] Z. A. Khan, T. Hussain, A. Ullah, S. Rho, M. Lee, and S. W. Baik, "Towards Efficient Electricity Forecasting in Residential and Commercial Buildings: A Novel Hybrid CNN with a LSTM-AE based Framework," Sensors, vol. 20, p. 1399, Jan. 2020.

[24] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-Term Load Forecasting With Deep Residual Networks," IEEE Transactions on Smart Grid, vol. 10, pp. 3943–3952, July 2019.

[25] M. Sajjad, Z. A. Khan, A. Ullah, T. Hussain, W. Ullah, M. Y. Lee, and S. W. Baik, "A Novel CNN-GRU-Based Hybrid Approach for Short-Term Residential Load Forecasting," IEEE Access, vol. 8, pp. 143759–143768, 2020.

[26] C. Lang, F. Steinborn, O. Steffens, and E. W. Lang, "Electricity load forecasting – an evaluation of simple 1d-cnn network structures.," 2019.

[27] H. J. Sadaei, P. C. de Lima e Silva, F. G. Guimarães, and M. H. Lee, "Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series.," 2019.

[28] C. Tong, L. Zhang, H. Li, and Y. Ding, "Temporal inception convolutional network based on multi-head attention for ultra-short-term load forecasting.," IET Generation, Transmission Distribution (Wiley-Blackwell), vol. 16, no. 8, pp. 1680 – 1696, 2022.

[29] J. Kim, J. Moon, E. Hwang, and P. Kang, "Recurrent inception convolution neural network for multi short-term load forecasting.," Energy Buildings, vol. 194, pp. 328 – 341, 2019.

[30] R Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2017. https://www.R-project.org/.

[31] M. Gaunt, "GH archive." https://github.com/igrigorik/gharchive.org, 2020. Distributed by GitHub.

[32] F. Chollet et al., "Keras," 2015. https://keras.io.

[33] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," arXiv preprint arXiv:1603.04467, 2016.

[34] Stack Overflow, "Stack Overflow Developer Survey 2021." https://insights.stackoverflow.com/survey/2020#technology-most-loved-dreaded-and-wanted-languages-dreaded, 2021.

[35] "Data Miner 2." PJM, 2020. [Online]. Available: https://dataminer2.pjm.com/.

[36] A. Smith, N. Lott, and R. Vose, "The Integrated Surface Database: Recent Developments and Partnerships," Bulletin of the American Meteorological Society, vol. 92, pp. 704–708, June 2011.

[37] "Iowa Environmental Mesonet." Iowa State University, 2013. [Online]. Available: https://mesonet.agron.iastate.edu/.

[38] "When do the four seasons officially begin?," 2007. npl.co.uk. https://www.npl.co.uk/resources/q-a/when-do-the-four-seasons-begin/ (accessed Aug. 22, 2021).

[39] "Encyclopedic entry: Season," 2011. nationalgeographic.org. https://www.nationalgeographic.org/encyclopedia/season/ (accessed Aug. 22, 2021).

[40] H. I. Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P.-A. Muller, and F. Petitjean, "InceptionTime: Finding AlexNet for time series classification," Data Mining and Knowledge Discovery, vol. 34, pp. 1936–1962, sep 2020.