UNIVERSITY OF CALIFORNIA
SANTA CRUZ

ARTIFICIAL NEURAL NETWORK FOR OPTIMIZED POWER SYSTEM MANAGEMENT

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

by

Daniel A. O'Leary

June 2015

The Dissertation of Daniel A. O'Leary
is approved:

Professor Joel Kubby, Chair

Professor Michael Isaacson

Professor Patrick Mantey

Professor Ken Pedrotti

Dean Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Artificial Neural Network for Optimized Power System Management

by

Daniel A. O'Leary

This paper proposes a model for making power purchasing and routing decisions at the prosumer level and presents forecasting techniques that can be used in the optimization of that model.

To implement this model, the author has built and installed tracking solar panels, solar radiation sensors, power measurement devices, and automated data collections systems at the University of California Santa Cruz (UCSC) and the National Aeronautics and Space Administration (NASA) Renewable Energy Lab (RE Lab) in Mountain View, CA. Further, the author has programmed data collection and database tools, Artifical Neural Networks, and data analysis tools to capitalize on the data collected at the RE Lab. This paper reviews the use of Artificial Neural Networks (ANNs), a powerful forecasting method that has proven to be among the most accurate of the machine learning algorithms, for the prediction of power production and consumption. This paper includes analysis of ANNs for the prediction of solar and wind power production and individual power consumption. Finally, a new method of feature selection is proposed and found to improve short term solar power predictions.

To my parents,

Tim and Linda O'Leary,

Their love brought me into this world and their example inspires me to improve it.

Acknowledgments

I want to thank my cohorts in Prof. Kubby's MEMS group, in particular my friend, confidant, and advisor Oscar Azucena. I would also like to thank professors Kubby, Mantey, and Isaacson - who have mentored me as a student, advisee, and TA. I have greatly benefitted from their patience and insight. Finally, I would like to thank my friend and intern, Samuel Kahn, whose work has been tireless and his contributions invaluable.

Part I

First Part

Chapter 1

Introduction

The addition of renewable energies into the aging US power grid requires new technologies and management techniques. This improved power grid, the smart grid, will forecast future power availability and needs and make decisions about routing, production, and consumption of power. These decisions were once the sole domain of the power utility. However, recent political, social, and economic incentives have exploded the prevalence of consumer side renewable energy. These new players in the energy market - both producers and consumers of power (prosumers) - need to make decisions about power distribution, how much power to buy or sell and where to route the power that has been bought or produced. These decisions are similar to the decisions made by the power utility for the overall power grid, and the prosumer needs similar tools to those used by the power utility to automate decisions - but tailored to the prosumer's unique needs. To optimize power use, the prosumer requires information about present and future power production and consumption. This paper outlines a model for "bottom

up" power management and presents new methods for power forecasting in key elements of that model.

Improvements to the existing power grid are motivated by the changing landscape of US power production. This paper examines that changing landscape, identifies the need for an improved power grid, and reviews existing literature for improving the power grid. As power consumers become power producers, and energy needs and management at the individual level increase in complexity, the forecasting and management tools of the power industry can benefit the consumer as well. A proposed forecasting and power management system is offered to implement those improvements at the individual level.

Artificial Neural Network (ANN) forecasting methods are investigated using two renewable sources, wind and solar, and one consumption model, the power usage in a home. Further analysis is provided on the most common renewable energy for individual production, solar power. An improvement to existing solar power ANNs is developed and tested.

Chapter 2

Motivation and Literature Review

2.1 Motivation

The power grid in the United States (US) is supplied from many sources, the three most pervasive sources are fossil fuels (sixty-seven percent), nuclear power (nineteen percent), and hydroelectric power (seven percent) [1]. Each source has costs and benefits:

- Fossil fuels are inexpensive, easy and safe to store and transport, and benefit from a large, mature infrastructure. However, fossil fuels have significant environmental problems inherent in their use or endemic to their production. Fossil fuels are a limited resource that is more expensive and more dangerous to produce as resources dwindle. Finally, there are significant negative political implications to our energy dependence on foreign fossil-fuel-rich countries.

- Hydroelectric power is ideal in many ways, it is renewable, inexpensive to produce,

and safe. However the construction of hydroelectric dams have several limitations. First, construction of dams require a large initial investment. Second, the environmental impact of building a dam can be significant to local plants, animals, and even people. Finally, few hydroelectric dams are planned in the future because the most advantageous locations are already in use - it is unlikely that new hydroelectric sources will be identified to meet our increasing energy needs.

- Nuclear power is, on the whole, inexpensive, safe, and environmentally friendly. However, like hydroelectric dams, the construction of nuclear power plants require a large initial investment. Although nuclear power plant accidents are rare, when a nuclear accident does occur, the results can be catastrophic. Finally, compared with the environmental waste associated with burning fossil fuels, nuclear power has less waste that is easier to contain; nevertheless the nuclear waste that is produced is radioactive, dangerous to transport and store, and has a half life of twenty-four millennia.

No one power source is an ideal solution to our power needs; however renewable energy sources, such as solar and wind, are an increasingly important power source - especially for individuals or small companies. Political, economic, and environmental factors make these options better than ever before and the indication is that our use of renewable energies will continue to grow.

- Political - The California Renewables Portfolio Standard requires that one-third of California's energy use come from renewable resources by 2020[5]. In 2013 the

President of the United States issued a memorandum as part of the Federal Leadership in Energy Management requiring that twenty percent of all federal buildings power usage come from renewable resources by 2020[17]. Further, growing concern over the safety of nuclear power and fossil fuels due to recent oil spills and nuclear accidents are increasing pressure to find alternatives to these traditional sources. One political response to the demand for more renewable energy has been to subsidize the purchase and installation of renewables, lowering the economic barrier-to-entry for solar, wind, and other renewable power production.

- Economic - The costs of harvesting renewable energy continues to diminish. Outside of subsidies, which artificially and temporarily decrease the cost of renewables, other market driven forces are permanently reducing the cost of producing renewable energy. Manufacturing changes due to technological improvements and lower costs associated with increased production have reduced the cost of solar panels and wind turbines. Increases in demand for renewable resources have resulted in the construction of new manufacturing facilities.

  These new facilities take advantage of new technologies, producing solar panels and wind turbines that are less expensive and more efficient than before, increasing production (which lowers prices) and decreasing production costs (which also lowers price). The price of renewable energy is also becoming less expensive relative to other power options. As existing fuel sources become scarcer, prices for those resources increase, making the decreasing costs of renewables seem even less

expensive by comparison.

- Environmental - Renewables are not without environmental impacts for example, wind turbines can negatively impact bird populations and migration patterns and solar panels can contain toxic chemicals and can be difficult to recycle. Generally, these problems pale in comparison to the environmental disasters found in oil and nuclear industries and the hazardous environments and greenhouse gasses of coal and other fossil fuels.

An increased reliance on renewable energy brings with it new challenges to the existing power infrastructure that need to be addressed. The power company must manage and regulate the power on the grid while handling, not only the variability of their renewable energy suppliers (solar and wind farms), but also the variability of prosumers' power production (homes and businesses with solar panels and wind turbines).

Most homes in the US are connected to a public utility power company. The power utility often offers tiered pricing structures to try to initiate Demand Side Management (DSM) to promote power conservation during peak power consumption hours. Some power companies go so far as to offer real time pricing of power, giving customers access to ever fluctuating power prices, like a stock market for home power. Many people in variable pricing power models are already performing a simple instance of power management by curtailing power usage when power prices are high. In the tiered pricing example, this management model is often very simple. Users identify appliances

whose use can be delayed or eliminated during high peak pricing times. An example would be a customer during peak pricing times that waits to do laundry until power prices go down. Such a simple model does not require software or hardware assistance. However, more complex decisions could be made that would require specialized software and hardware; for example, smart water heaters could delay water heating during high pricing times. The decision on when to heat the water in the hot water heater could be as simple as maintaining a minimum level of hot water temperature until off peak pricing comes into effect, or it could be a more complex decision based on previous water usage to determine the acceptable minimum water temperature on an ad hoc basis. In this way, even the simple tiered pricing model could warrant a software agent to forecast power needs and make decisions about power use. In the more complex instance of real time power pricing, forecasts on power pricing would also assist in making optimal power purchasing decisions.

Many homes now have solar panels that produce power, and prosumers have the choice to use that power or sell it back to the power company. In the simplest case, under flat rate power pricing, solar power management decisions are very straight forward - always use the solar power generated, and buy as little as possible. The addition of tiered or real time pricing makes the decision far more complex. In a previous example a consumer chose to delay doing laundry until power prices decreased. The addition of solar power makes that decision less clear. Is there enough solar power to do laundry without purchasing power from the power company? Will solar power be available during non-peak pricing times? Would it make more sense to delay doing

laundry, sell the solar power at the peak price, and do the laundry later under solar power? Does the previous decision make sense if there is no solar power expected during non-peak pricing hours and power must be purchased from the power utility to do laundry, albeit at non-peak prices? The answers to these questions require an understanding of the patterns of local renewable power production and utility power pricing to optimize power management decisions.

The addition of power storage adds further complexity to power management. First, there are the factors associated with the cost of storage - there are different types of storage, devices that are primarily for storage; for example, lithium ion batteries, lead acid batteries, and hydrogen fuel cells have all been used in individual homes for power storage (or load balancing DSM). There are devices that can perform power storage as a secondary role, such as vehicle to grid electric cars and thermal storage in smart appliances such as refrigerators, freezers, air conditioners, and hot water heaters. Each of these systems has different costs associated with them - the cost of maintaining a charge in a lithium ion battery is much different than the cost of maintaining a thermal charge in a hot water heater. It is also necessary to consider the impacts the charging and discharging of the storage medium has on the life of the storage medium - charge/discharge cycles on a water heater have less of an impact on the lifetime of the water heater than charge/discharge cycles on a lead acid battery have on the lifetime of that battery.

A decision must be made about the use of power, given considerations about storage availability and costs, current and forecast power prices, forecasted power usage,

and forecasted power production from renewables. Even the simple model of a prosumer with tiered power pricing could benefit from power management software. A realtime pricing model with various storage and renewable energy sources would require several complex decisions daily, if not hourly. Most consumers will need an automated decision making tool to manage home power purchases and storage. Such a model is offered in Fig. 2.1.

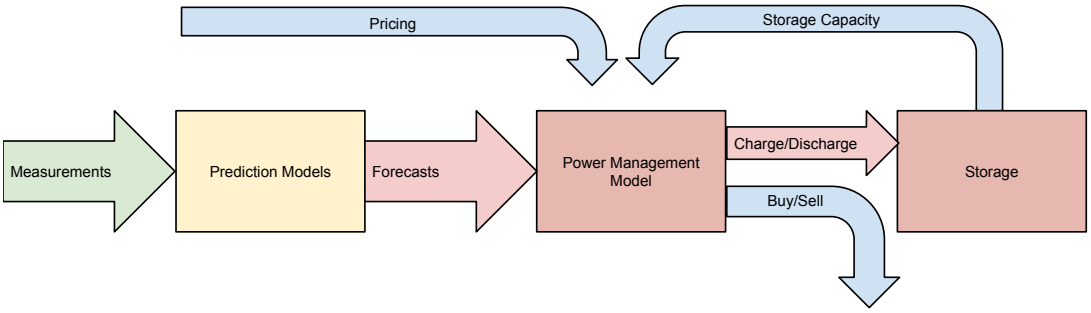


Figure 2.1: A model of the power management model decision making information flow.

In this model, measurements are inputs used to make predictions about future performance. These measurements include solar power production, solar irradiance, wind speed, and power usage. The details of the measurement techniques are outlined in the hardware chapter of this document.

From these measurements, machine learning techniques identify patterns in power use and power production. From these patterns, the software forecasts future power consumption and power generation. These forecasts, along with realtime pricing data and available storage constraints, are needed by the power management model to identify an optimum store, buy, or sell quantity. Finally, given the amount of energy to

be put into or taken out of storage, a decision is made about the best storage method to use.

This paper details the work done in the first three area of this model (measurement, prediction models, and forecasts). Future works involve the investigation of the other aspects of the proposed model.

2.2 Literature

Power companies have long looked to machine learning techniques to make power consumption and production predictions. The high variability of renewables are of particular interest, because power companies are required to implement renewable sources, but running a consistently available and dependable power grid on variable power sources is difficult. Greater foreknowledge on renewable power generation can reduce the need for expensive peaking power plants and optimize the use of renewable power farms and base power plants.

In 2005, Mellit [13] used a Radial Basis Funciton (RBF) based Artificial Neural Network (ANN) to predict daily global solar radiation with 98.9% accuracy. Using 20 years of global solar radiation data from a meteorological station in Algiers, researchers created a composite reference year. Using this reference year for training, validation and test, the scientists were able to create an ANN with one hidden layer of nine nodes that took inputs of air temperature and sunshine duration and output global daily solar radiation. On first reading, this level of accuracy seems incredulous, but a

close look at what is being predicted suggests the accomplishment is pretty straight forward. This is simply the use of how long the sun shone and how warm a day it was to determine the aggregate solar radiation for the day. According to the article, these inputs' relationship to the output is non-linear and poorly suited to other non-linear signal predictors - however the accuracy of these other methods is not discussed, so it is unknown to what degree the RBF is an improvement.

Similar solar power production forecasts have been done with no exogenous inputs [21]. Forecasting models used as input two years of power output data from a 1 MW solar facility in Merced, CA, detailing the average power output in one hour increments. The data was run through different forecasting methods with different time horizons (1 and 2 hour predictions) and found that ANNs outperformed the other techniques. In a similar paper, the use of support vector machines for time series prediction was explored [23]. Although the paper was larger in scope than just smart grids or renewable energy, a section on electric utility forecasting details several papers in the field of electrical load forecasting.

Researchers have also used back propagation of ANNs for short term photovoltaic (PV) power generation prediction [7]. Using an ANN with four inputs and two hidden layers, researchers made predictions on the power produced from a 750 W PV panel in Instanbul, Turkey. The four inputs were ambient temperature, cell temperature, diffuse solar irradiation, and power produced. A comparison was made of the increase in root mean squared error (RMSE) and correlation coefficients for varying time horizons of the prediction. The author found that a five minute time horizon held

strong prediction accuracy, and anything between five and thirty-five minutes had an acceptable RMSE and correlation coefficient. The author does not identify what is considered an acceptable correlation coefficient or RMSE, but based on the charts in the paper, it would appear to be a correlation coefficient of 0.85 or greater and a RMSE of 50 W (from a 750 W panel, roughly 15%) or less.

In 2012, a Particle Swarm Optimization (PSO) algorithm was used with a complex set of mathematical equations that modeled multiple renewable energy sources, storage, consumer demands, and market pricing to find an optimum micro grid set-up [16]. Although this is not a machine learning paper, it does represent the modeling and optimization of multiple renewable power sources, the integration of market pricing into a model, and optimization algorithm (searching for minima) is a significant portion of machine learning.

Researchers in Chile created a model of a microgrid in Huatacondo, a small isolated village in the Atacama Desert, Chile, and proposed an energy management system (EMS) based on that model [19]. This microgrid included many of the same systems used at the NASA RE lab: PV panels, a wind turbine, a battery bank, and consumer electric load demands. The microgrid also included a diesel generator and a water pump (a flexible load), which are not part of the system at NASA. The energy management system made predictions using a backpropagation artificial neural network algorithm on electricity consumption with a time horizon of two days. The observed power consumption for the village tended to be around ten to fifteen kW from midnight to six p.m., then a ramp up from six p.m. to nine p.m., hitting a peak of forty kW,

with a similar ramp down from nine p.m. to midnight. Predictions on this consumption had a mean absolute error (MAE) of 1.6 kW and a standard deviation of 1.4 kW. The percentage of error associated with these values is not given in the paper, but based on the graphs, the power consumption described, and the given MAE, a ten percent error is a reasonable estimate. Researchers were able to create a microgrid that fulfilled the power balance in the microgrid under varying consumption and generation profiles. Using the renewable energy sources, the flexible water pumping load, and the diesel battery storage, they minimized the use of the diesel generator for a savings of 2.33%.

ANNs with back propagation are also used in determining the optimal power flow for stability of a power grid with high renewable energy penetration [11]. The ANN in this work took inputs of total energy cost, bus voltage deviations, frequency deviation, tie-line flow deviations, line loadings, line losses, generator stability margins, and other indices related to the system economy, stability, and security. The output is the active and reactive power to be generated. Using this Dynamic Stochastic Optimal Power Flow (DSOPF) control algorithm in a modeled twelve bus test power system, researchers simulated an unexpected contingency of a 100 MW sixty Mvar load to trip. The DSPOF system resulted in lower fuel cost and line loss when compared to the traditional real-time active power balancing tool, the proportional-integral-controller-based automatic generation control. Further, system frequency and tie-line flow are regulated much closer to their nominal values and under-voltage load shedding is avoided.

Supporting an increasing electric vehicle (EV) fleet on the existing power system offers problems and potential solutions for power grids with increasingly variable

renewable energy production. Using mathematical models to simulate a smart grid with flexible EV charging, scientists used PSO algorithms to minimize cost and emissions from traditional power plants [22].

Just as the previous paper found an optimum load time for EV charging to balance existing power demand fluctuations, other researchers have used distributed generation to balance existing power production fluctuations [10]. Scientists explored the use of distributed generation as back-up generation for islanded microgrids (power outages) and voltage support and reactive power control when connected to the grid. Using a modeled microgrid under different islanding and reactive power fluctuation scenarios, the use of a smart grid to influence distributed generation power systems reduced down time and costs to the system.

Researchers in Sweden have modeled the impact of an increased load of an additional 2200 EVs [24] on the existing Swedish power grid. The electric vehicle fleet is modeled after a mix of taxis, commuter cars, and family vehicles - with different power consumption and time of use profiles. When the vehicles charged in an uncoordinated, direct manner, the peak power demands on the power grid for EVs coincided with the peak power demands of the base consumption on the network. The researchers proposed using a smart power system to charge EVs on a flexible schedule. In doing so, they were able to move the majority of the EV charging load to the midnight to six a.m. time frame, when base power demands were at a minimum while still meeting the driving demands of the simulated EV fleet.

The analysis of the preprocessing methods used to improve ANNs has shown

much success in other fields, including the use of visual data in the pharmaceutical field [6] and fault identification in rotating machinery [20] where wavelet analysis has been used for preprocessing ANN data.

For convenience, common ANN error assessment figures and the values found in the literature above is offered in table 2.1. The equations for these metrics can be found in the appendix.

Table 2.1: Error values for ANNs

Assessment	Value
MRE	1.5% [21]
R^2	95% [21], 85% [7]
MAE	53.49 kW [21] 10% [19]
nRMSE	15.82% [21], 7% [7]

Chapter 3

Measurements

3.1 Solar Measurements

Solar energy measurements were taken at the Renewable Energy lab at NASA Ames, in Mountain View, CA. The hardware of the RE lab consists of a solar tracker, six solar panels, a computer, and several insolation measurement instruments. These tools are used to monitor solar production and to offer students an online interactive opportunity to study renewable energies with production level equipment [18].

The solar tracker is a Wattsun AZ-225 solar tracker, capable of rotating 225 sq. ft. of solar panels. The tracker is mounted on a thirteen foot pole buried half in the ground, meaning the tracker motors are mounted on the top of the pole at roughly 6.5 feet above ground. The two motors move slowly (the top speed is about one degree every five seconds).

The solar tracker holds six Sharp NE-170U1 solar panels, capable of producing

a total of 1.02 kW of power. The Sharp NE-170U1 panel is made of seventy-two polycrystalline solar cells in series with a total efficiency of 13.1%. The power instruments of the RE lab measure only one of the six panels, for a maximum power of 170 W of power.

The hardware at the center of the RE Online Lab is a computer, the solar server. The solar server provides three functions: reporting data, collecting data, and controlling the solar tracker. The solar server reports data from a database on the solar server to the RE Online Lab website. The solar server collects data from the laboratory's measurement tools and stores that data in a database on the solar server. Finally, the solar server controls the solar tracker's orientation via a connection to custom circuitry that manipulates the solar tracker's motors.

There are three measurement devices connected to the solar server that are used as inputs to the ANN: an IV curve tracer, to measure the max power point of the solar panel, a Precision Spectral Pyronometer (PSP) and a Normal Incidence Pyrheliometer (NIP) to measure integrated and directional irradiance, respectively.

The IV curve tracer, a DS-100C IV curve tracer from Daystar, Inc., is a power measurement instrument. The power it measures is reported as an IV curve, showing the voltage and current of a solar panel under different load conditions. The IV curve tracer is connected to power lines that run to one solar panel on the solar tracker. The IV curve tracer is also connected to a Universal Serial Bus (USB) line that runs to the solar server. It receives instructions from the solar server and sends IV curve measurements to the solar server using the USB line.

The PSP and the NIP are from Eppley Laboratory, Inc. Irradiance, which is the density of sunlight power incident on a surface, is measured in $watts/meter^2$. The PSP and NIP are connected to the IV curve tracer, which allows the solar server to read data from the PSP and NIP via the USB connection to the IV curve tracer. The PSP and NIP are mounted on the solar tracker such that they always faces the same direction as the solar panels.

The PSP measures the intensity of light with a subtended solid angle of 2π (a hemisphere), depicted in Fig. 3.1. The PSP's large solid angle allows it to measure light intensity from everything in front of it, including light directly from the sun, light that bounces off the ground, and diffuse light that reflects from the clouds. Consequently, the PSP measures the intensity of light hitting the solar panel from all directions.

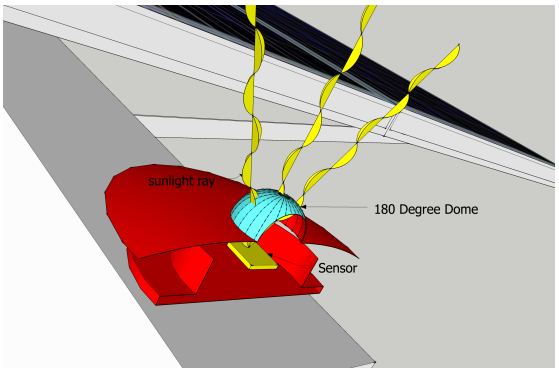


Figure 3.1: A depiction of the PSP with 2π steradians subtended solid angle.

The NIP measures the intensity of light with a subtended solid angle of 7.84 millisteradians, depicted in Fig. 3.2. The NIP's small solid angle allows it to measure only the light that comes directly at it. When the tracker is pointed at the sun, the

NIP measures only the intensity of light from the sun, blocking out any reflective light that bounces off the ground or clouds. Thus, it is reasonable that the signal from the NIP would dominate the input to the ANN on a clear, sunny day, and the signal from the PSP on a cloudy or foggy day. Using this equipment in Mountain View, CA, measurements were taken over the course of two years, from August 2012 to May 2014.

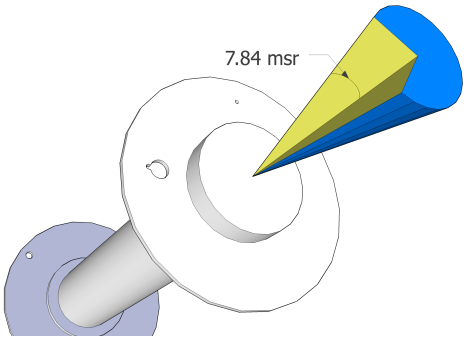


Figure 3.2: A depiction of the NIP with 7.84 millisteradians subtended solid angle.

3.2 Wind and Power Consumption Measurements

Wind data for wind prediction comes from wind speed measurements made by the weather station at Moffett Field, in Mountain View, CA. Power consumption predictions are based off the Pacific Gas and Electric (PG&E) killowatthours used at a home in Los Gatos, CA for the year 2014. This home's power included solar panel production, resulting in negative kilowatt hours consumed - during times when solar power produced exceeded the home's power consumption. Future models would be based solely on power consumed and not power produced - which should yield more

accurate predictions, as it reduces the variables in the input (solar power production).

Chapter 4

Software

The prediction models that use the measurements to generate forecasts are artificial neural networks (ANNs). Artificial neural networks are mathematical constructs based on the physical structures of a biological system, the brain. In Fig. 4.1 a generalized artificial neuron is depicted, the building block of ANNs. The neuron consists of inputs (x_i), weights (w_i), a summation (Σ), an analysis of the summation (the activation function, S), and finally an output (y).

Typically, a preprocessing step is performed on data before entering into the ANN. It is standard to perform a min-max normalization, resulting in all inputs, outputs, and targets to fall between 0 and 1. In the simplest form, a neuron can act as an AND gate. x_1 and x_2 are the inputs to the AND gate. w_1 and w_2 are each 1/2. If either x_1 or x_2 is zero, the summation will be less than or equal to 1/2. The activation function (S) returns a 1 if the summation is greater than 1/2, 0 otherwise.

The Artificial Neural Network (ANN) is made up of layers of neurons. Fig. 4.2

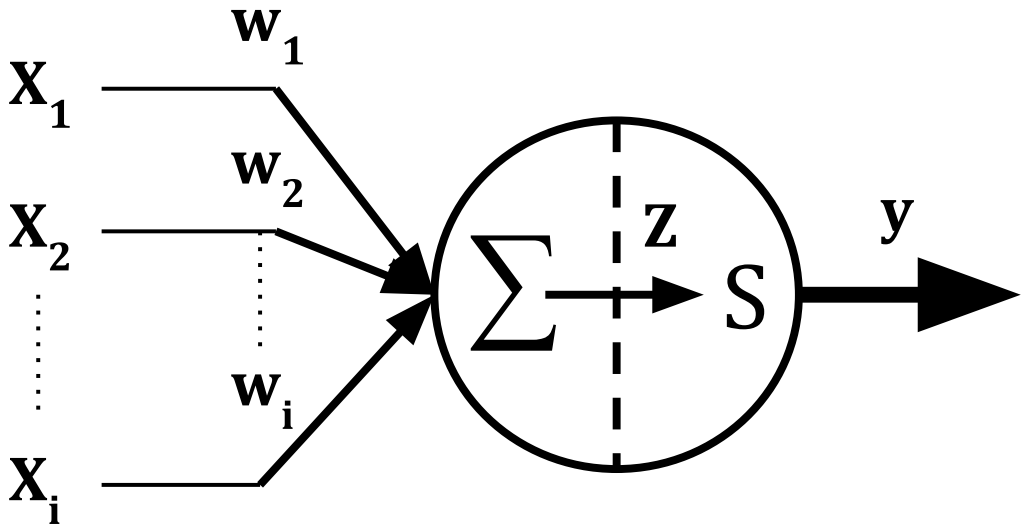


Figure 4.1: A model of an artificial neuron. The inputs x_i are multiplied by their respective weights w_i, then summed. The resulting sum is run through an activation function S producing an output y [15].

depicts a generalized version of the most common ANN structure. This ANN consists of three layers; the input layer, the hidden layer, and the output layer. Each neuron in the input layer takes in one data source. The output of each input layer neuron is input for each of the hidden layer neurons. Thus, if you have five input neurons, each neuron in the hidden layer will have five inputs. This relationship occurs a second time between the output of a neuron in the hidden layer and the inputs of the output layer. Thus, if you have twenty neurons in the hidden layer, you will have twenty inputs for each neuron in the output layer.

4.1 Analysis

4.1.1 ANNs with standard pre-processing

Standard ANN pre-processing involves the normalization of input data. The Pearson correlation is a standard measure of the dependence (linear correlation) between two quantities. Fig. 4.3 shows the correlation between the predictions of the ANN for the max power point (MPP) of the solar panel twenty minutes in the future (the prediction value, found along the vertical axis), and the actual MPP that was measured twenty minutes later (the target value, found along the horizontal axis). For example, if at 12:15 p.m., the ANN predicts that the solar panel will have a MPP of 100 Watts, and at 12:35 p.m. it is found that the solar panel produces fifty Watts, a point would appear at (x,y)=(50,100) on the graph. The graph indicates that there is a large clustering of accurate predictions in the low (zero to twenty-five) and high (125-170) Watt range.

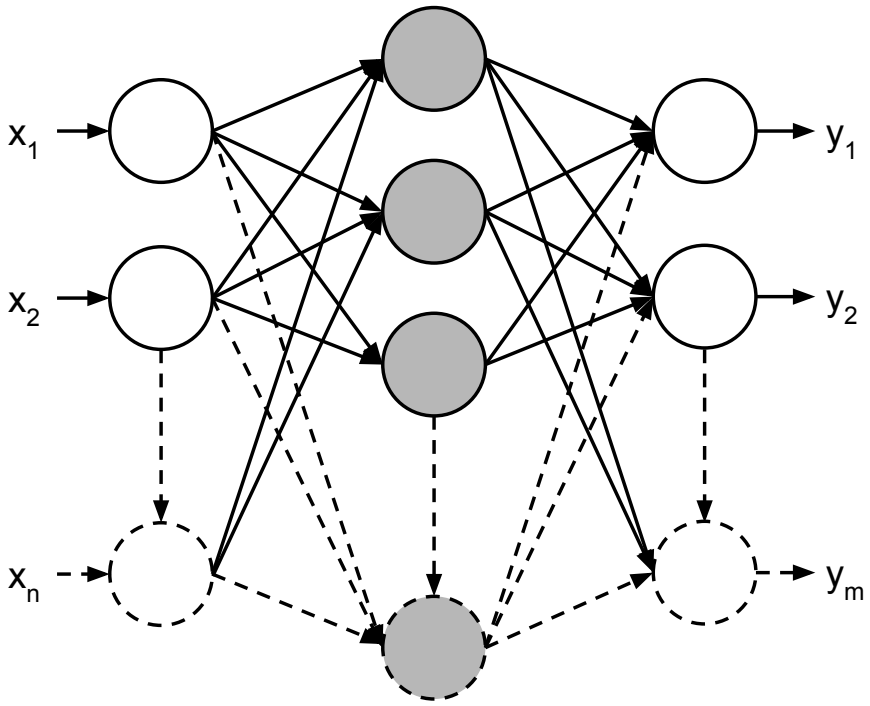


Figure 4.2: An Artificial Neural Network (ANN). Data is input to the input layer (x_i), all outputs of each layer are the inputs for each neuron of the following layer. The output layer's output (y_i) are the results of the ANN analysis [15].

This corresponds to night time, when MPP values are consistently low, and daytime during full sun, when (on sunny days, which are common in California) the MPP values are consistently high. High variability MPPs, such as during cloudy days or at dawn or dusk, are difficult to predict, which explains the lack of accuracy at the middle MPP values (50-100 W).

Fig. 4.3 shows the correlation between the predictions of the ANN for the max power point (MPP) of the solar panel twenty minutes in the future (the prediction value, found along the vertical axis), and the actual MPP that was measured twenty minutes later (the target value, found along the horizontal axis). The accuracy measurements for the ANN, with an R^2 value of 90.88%, a $RMSE$ of 16.98 W and a MAE of 6.33 W (for a 170 W panel) are in line with other ANN forecasting accuracy correlation coefficients under similar circumstances [21] and outlined in table 2.1.

The development of the ANN requires a period of experimentation to optimize the structure of the ANN. One concern in the creation of machine learning tools is the phenomenon of "overfitting". During the training phase, an ANN is trained on a training set of data, a subset of the total data to be used. For each sample in the training data set, the ANN makes a prediction, compares that prediction to the target value, back propagates through the ANN layers, and adjusts the input weights of the ANN. This process is performed until all the samples in the training data have been predicted, assessed, and adjusted and the root mean squared error (RMSE) for the training set is calculated. Each time the ANN trains on a full set of the training data, that is an epoch. After one epoch, a new data set, the validation data set, is run through the

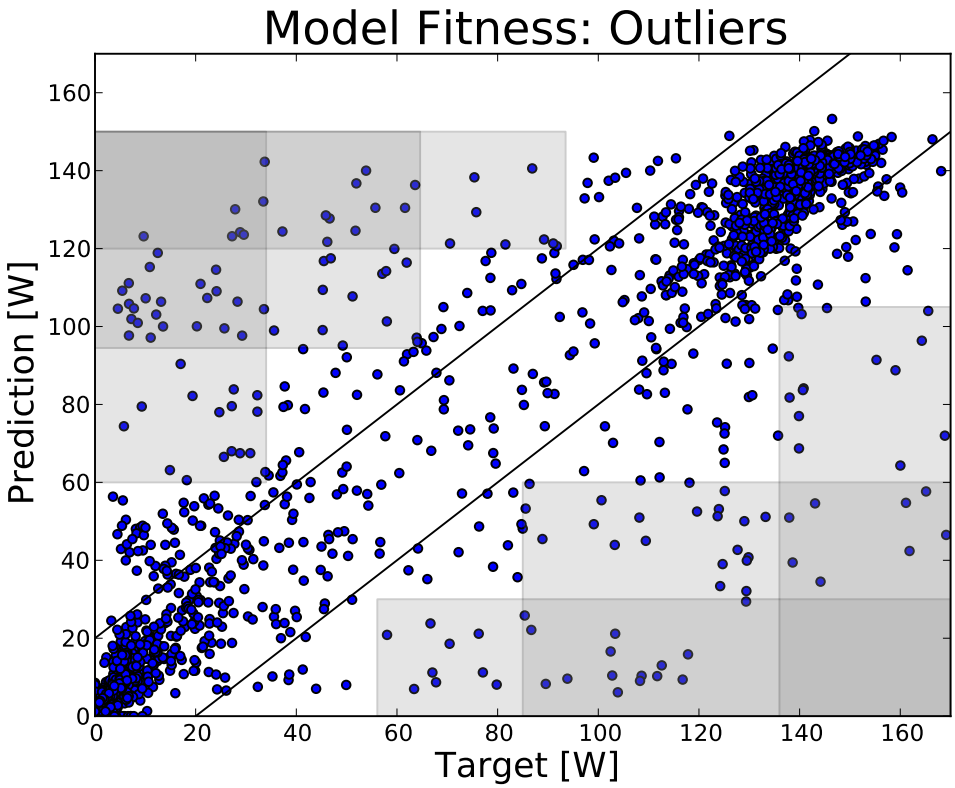


Figure 4.3: The Max Power Point (MPP) fitness analysis of a standard, normalized input ANN. Note: Overlapping target/prediction markers in this graph makes it difficult to see the high frequency of points in the zero to forty W range and the 120-160 W range that are highly accurate (eighty-five percent of the data points have less than ten percent error). This topic is discussed further in section 4.2.1. which uses the black lines and grey boxes in error categorization.

ANN, and the error is calculated. Then the process begins again. As the ANN trains through multiple epochs, the error for the training and validation sets decreases as the ANNs adjustments use the training data to identify patterns in the training data. This can be seen in early epochs of figure 4.4. Early epochs show a sharp decline in RMSE. As training continues, improvements in error continue, but at a diminished pace. Eventually, the validation set no longer improves as a result of more training, this can be seen at the minima of the second order polynomial line of best fit to the validation data. At epoch 789, the training data RMSE continues to decrease, but the validation error begins to increase. This is an example of overfitting. The ANN training after epoch 789 is causing the ANN to learn, and predict, the noise in the training set. As such, an ANN trained more than 789 epochs is more accurate at predicting the training set, but less accurate at predicting the validation set.

Another ANN feature to be customized is the number of hidden nodes in the ANN. Using the same dataset, but with different structured hidden note layers, the RMSE for validation and test data sets can indicate the impact of increased hidden nodes have on ANN accuracy and learning rate. Fig. 4.5 shows the impact of node increases on ANN RMSE for validation and test. As the number of nodes increases, the accuracy decreases, suggesting that the increase in nodes results in an overfitting of the data.

Power consumption analysis gives a classically shaped model fitness chart, similar to that found in the solar data. Fig. 4.6 shows a tight grouping along the forty-five degree line for typical energy use. As should be expected, peaking (times

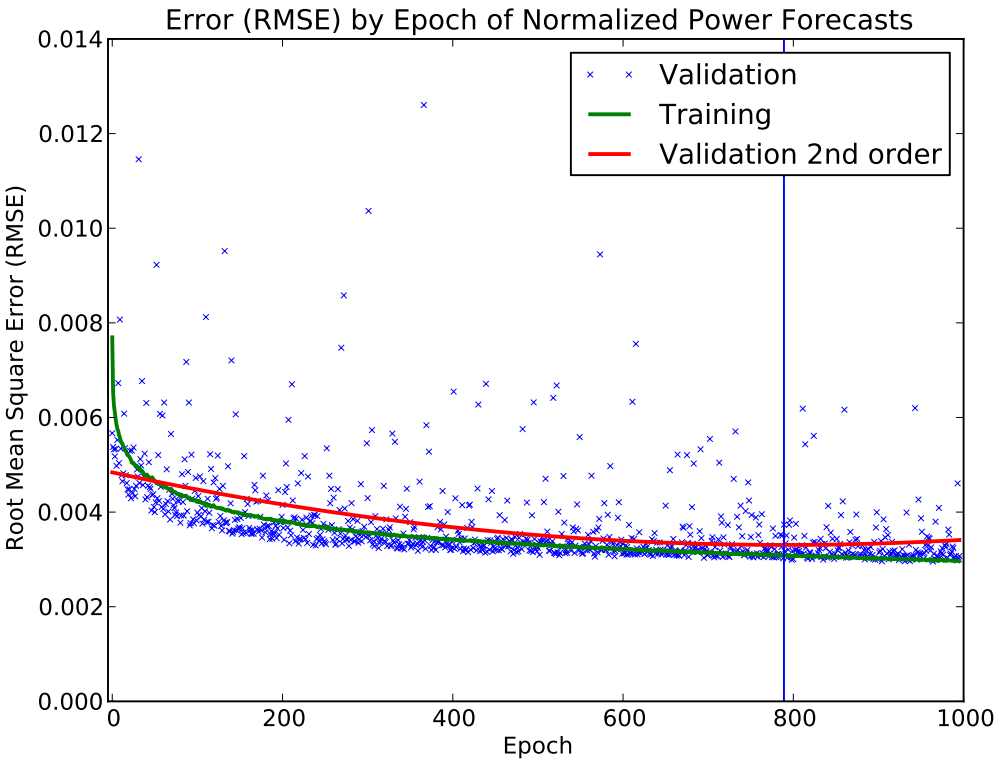


Figure 4.4: The training and validation root mean square error chart shows a decline in validation error until epoch 789, when validation error begins to increase due to overtraining

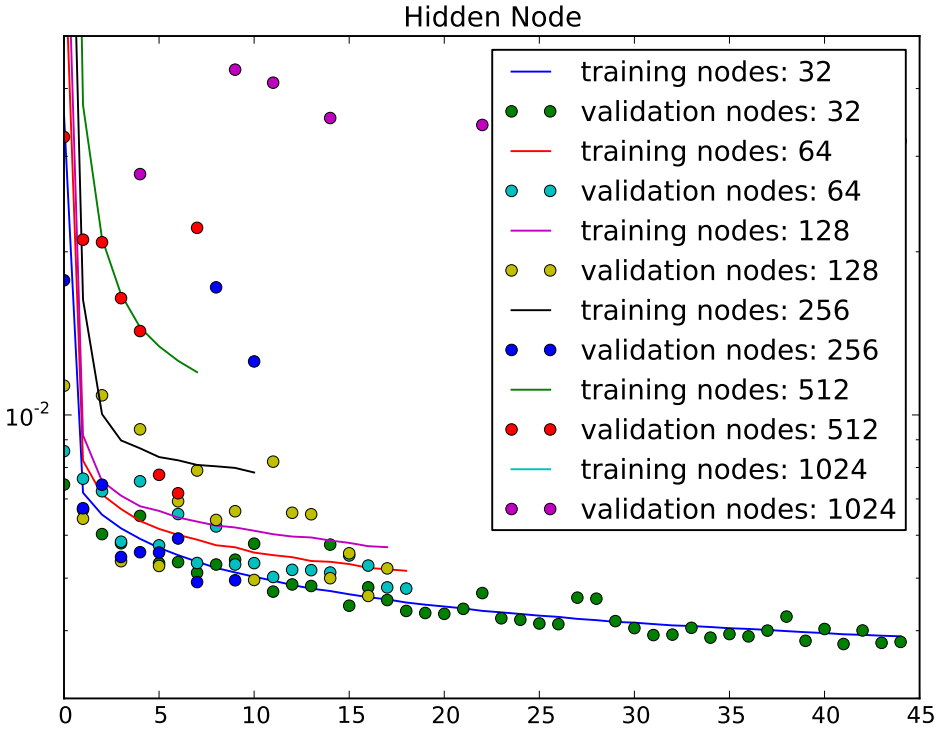


Figure 4.5: A comparison of several ANNs, showing the impact of hidden node quantity on accuracy

when power consumption goes uncharacteristically high), is difficult to predict. The challenge of predicting peaking presents itself in the lopsidedness of the predictions in the high power target instances. When the home uses the most power (further to the right of the x-axis) the ANN more often forecasts power usage below the target value (the target prediction point is below the forty-five degree line). Essentially, the ANN takes a conservative estimate of power consumption, developing a model that under predicts rare instances of high power consumption in exchange for less error during instances of typical power production.

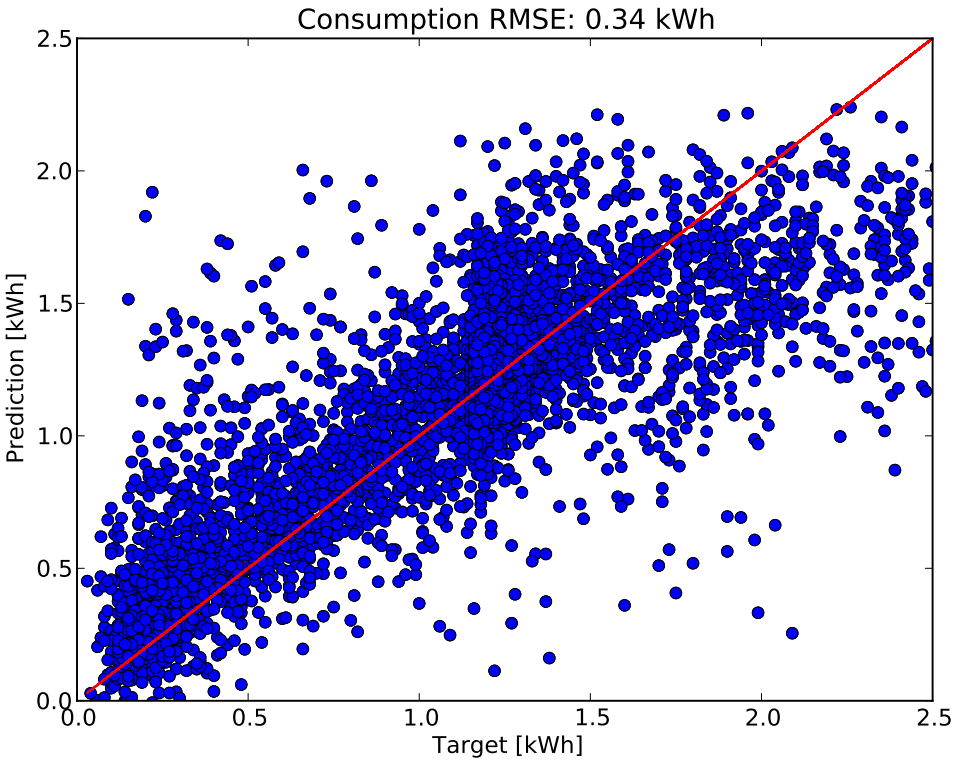


Figure 4.6: The power consumption fitness analysis of a standard, normalized input ANN.

Wind prediction data displays some notable differences from the solar data. The wind speed data from the METAR station is discrete in nature - such that measurements fall into one of twenty-four wind speed values. Consequently, target values for the fitness graph Fig. 4.7 show targets limited to these values.

The wind data differs from the other ANN data in this paper, because the wind data is discrete. Wind values are recorded from the Moffett Field Airfare Base METAR weather station in one of twenty-one speed ranges, from zero to 31.9 mph. The target vs prediction graph used in other analysis looks very odd with these discrete values Fig. 4.7.

Given that classifications of the discrete target values are known, the prediction values are filtered to the nearest target classification. Thus, any prediction between zero and 1.7 mph would be classified as zero mph, 1.7 to four mph would be classified as 3.4 mph, and so on. The result is a set of discrete points, demonstrated in Fig 4.8. However, now the chart indicates the classifications that are used, it does not give a sense of the frequency at which such classifications occur. The table 4.1 offers an insight to the frequency of the various prediction, target pairs and offers a new RMSE for the classified prediction values, down to 2.61 [mph].

The table reveals the characteristic grouping of high instances at low wind speeds (common for the area) and along the forty-five degree axis (indicating accurate predictions). There were 322 instances where the target and prediction agreed on a wind speed between 1.7 and four mph. There were 278 instances where the target value between four and 5.2 mph was under predicted by the ANN as being in the 1.7 to four

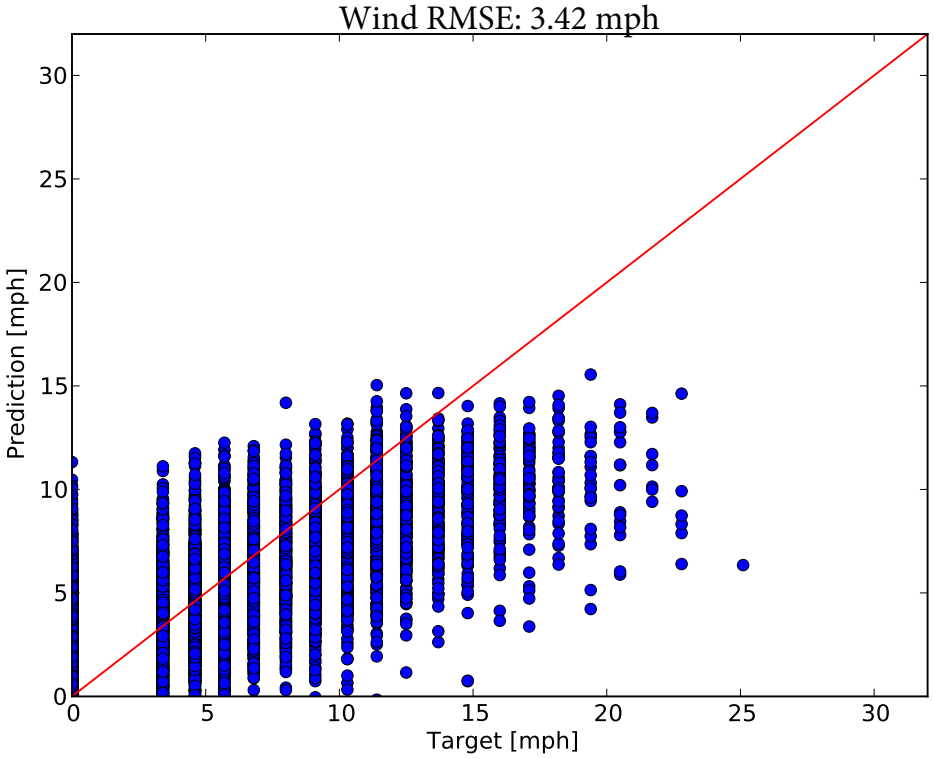


Figure 4.7: The fitness plot for wind speed looks "streaky" because the ANN is making a regression analysis of discrete targets.

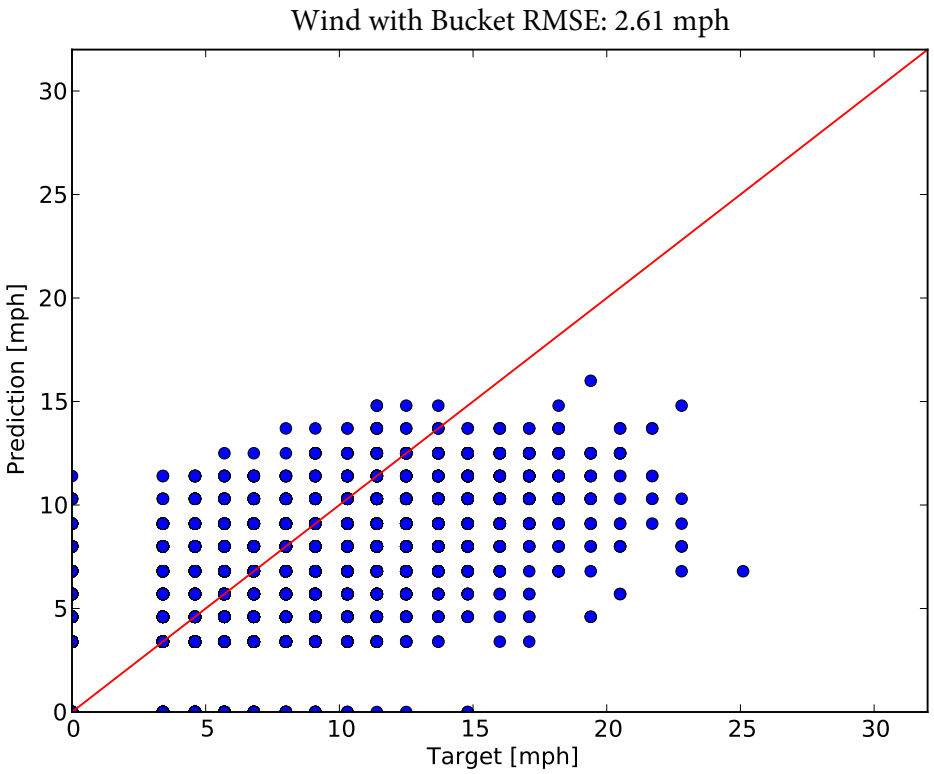


Figure 4.8: The fitness plot for wind speed with discrete targets and predictions.

Target [MPH]	Prediction [MPH]													
	0	1.7	4	5.2	6.3	7.4	8.6	9.7	10.9	12	13.1	14.3	15.4	
	1.7	4	5.2	6.3	7.4	8.6	9.7	10.9	12	13.1	14.3	15.4	16.6	Totals
0 - 1.7	978	912	288	156	89	44	14	5	1	0	0	0	0	2487
1.7 - 4	207	322	178	123	89	35	25	3	2	0	0	0	0	984
4 - 5.2	125	278	126	140	114	44	25	10	8	0	0	0	0	870
5.2 - 6.3	48	159	137	122	94	70	26	12	8	1	0	0	0	677
6.3 - 7.4	21	111	84	91	84	58	40	28	11	1	0	0	0	529
7.4 - 8.6	9	57	65	71	94	72	57	23	7	1	1	0	0	457
8.6 - 9.7	8	35	39	67	65	74	62	32	11	9	1	0	0	403
9.7 - 10.9	3	15	29	45	43	62	67	40	18	7	2	0	0	331
10.9 - 12	1	13	24	32	40	62	47	53	31	18	8	2	0	331
12 - 13.1	1	5	9	16	37	39	50	33	22	21	2	1	0	236
13.1 - 14.3	0	2	5	8	21	25	29	38	20	12	3	1	0	164
14.3 - 15.4	2	0	4	6	14	18	17	28	33	20	2	0	0	144
15.4 - 16.6	0	1	1	2	5	8	18	21	13	13	4	0	0	86
16.6 - 17.7	0	1	2	2	1	3	12	16	14	15	2	0	0	68
17.7 - 18.8	0	0	0	0	4	3	5	13	7	5	5	1	0	43
18.8 - 20	0	0	2	0	1	2	3	4	4	4	0	0	1	21
20 - 21.1	0	0	0	2	0	3	2	1	2	4	2	0	0	16
21.1 - 22.3	0	0	0	0	0	0	1	2	2	0	2	0	0	7
22.3 - 26.8	0	0	0	0	1	2	1	1	0	0	0	1	0	6
26.8 - 31.4	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Totals	1403	1911	993	883	797	624	501	363	214	131	34	6	1	7861

Table 4.1: The frequency of wind predictions for each prediction (columns) and target (rows) classification.

mph range. This table begins to show the impact and success of the ANN forecasting. A graph of the frequency of each classification in fig 4.9 shows that the ANN predicts the 0 to 1.7 mph range with less frequency than it occurs in the target values about 1000 times, and then predicts the 1.7 to four mph range about 900 times more than that classification is measured in targets.

Another way to look at this data would be a three dimensional histogram of the prediction/target pairs. Fig 4.10(a) gives the traditional fitness plot, with the frequency indicated by the height of the graph. Not unexpectedly, there is a high density of accurate predictions in the low wind speeds, the sheer number of which out shadows predictions in the mid to high wind speed range. The mid to high wind speed predictions

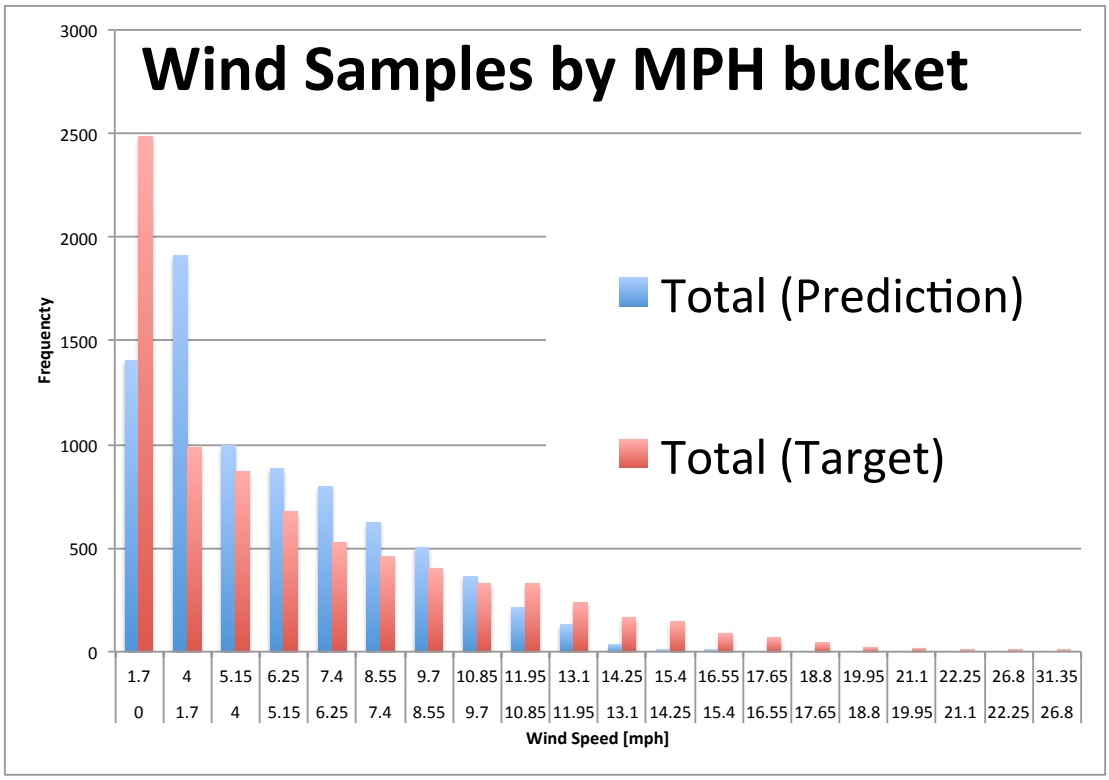


Figure 4.9: The target and prediction frequency for each wind speed classification.

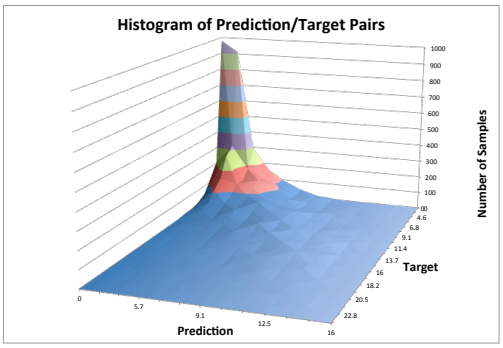
can be revealed through a log scale height axis, found in Fig 4.10(b), which diminishes the impact of the overpoweringly high density of accurate low wind speed predictions.

The logarithmic scale on the height axis allows the graph to have greater gradients for smaller frequency values and creates a more nuanced view of the mid and high speed wind predictions. With this new color gradient, a top down view of the logarithmic scaled histogram (Fig 4.10(c))will create a similar graph to the traditional fitness plot found in consumption and solar graphs.

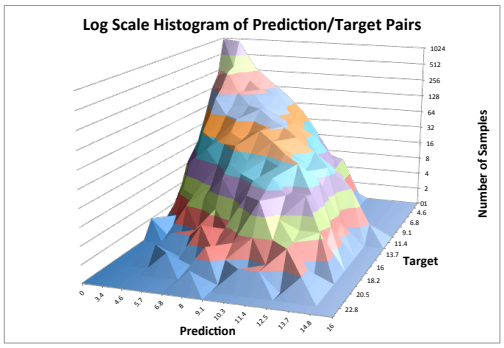
Fig 4.10(c) shows an ANN that is predicting aggressively at low speeds (for wind speeds less than four mph, when the ANN prediction is incorrect, it is more often predicting a higher value than the target value) and conservatively in the mid to high ranges (at wind target speeds greater than four mph, when the ANN is incorrect, the incorrect prediction is more often below the target value). The few instances where the target value exceeds 16 mph are essentially ignored by the ANN in favor of lower error values for more frequent predictions. Essentially, the ANN has learned to ignore gusts in favor of lower wind speed values that are more common.

4.1.2 Categorizing Inputs to Improve ANN Accuracy

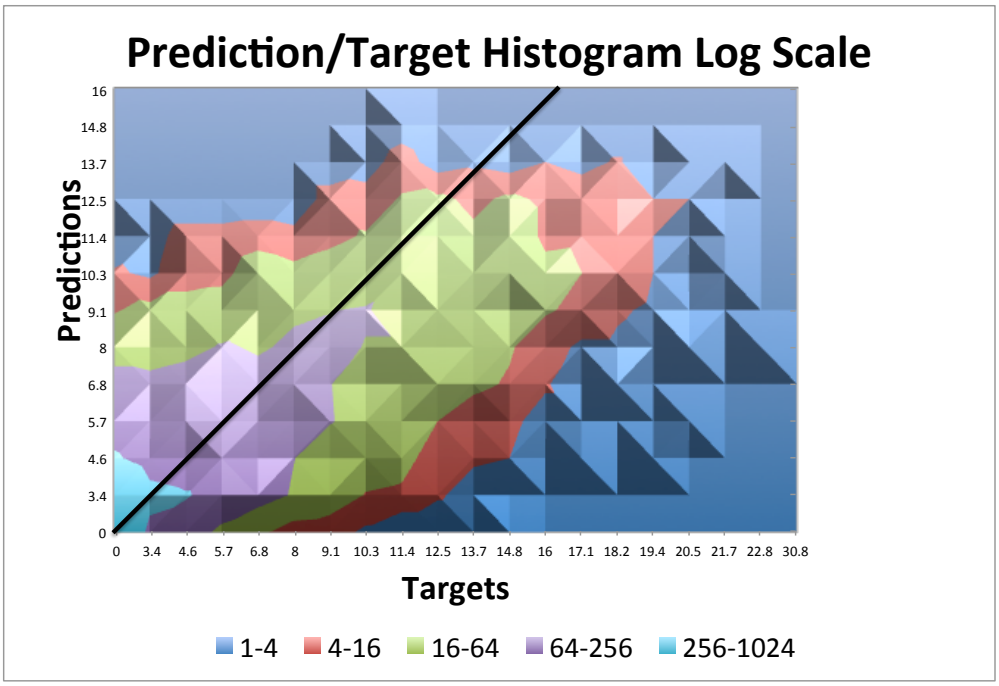
Neurons in the hidden layer of the ANN will find different local minima in the input space of the ANN. These local minima may correspond to different environmental scenarios - one neuron may find a local minima that is associated with cloudy days, another may find a minima associated with night time, etc. The output of the ANN is the sum of the neurons of prior layers. Consequently, neurons in the hidden layer



(a) A histogram of wind samples.



(b) A histogram with z axis in log scale.



(c) The fitness plot along the x-y plane, with frequency color coding in log scale.

Figure 4.10: A 3-D representations of the wind data histograms, showing the progression to logarithmic z axis scaling and color coding, resulting in a color coded fitness graph in line with expectations for accurate wind predictions.

that are well suited to predicting MPP of the solar panel on sunny days will still add to the output of the ANN on cloudy days. This results in an increase in error in an otherwise predictable environment. In order to allow the ANN to minimize the impact of hidden layer neurons that are not predictive of the current environment of the solar panel, masked inputs were designed that identify which neurons to emphasize and which neurons to minimize.

A cursory look at Fig. 4.3 could easily lead the reader to the conclusion that the ANN has a high forecasting error, despite the strong correlation coefficient. However, the model fitness diagram points overlap in areas where the prediction/target points cluster, at low and high power production times. Consequently, the vast majority (ninety percent of points) fall in between the two black lines indicating a target/prediction pair with an absolute error of less than twenty watts. The pie chart in Fig. 4.11 breaks out the percentage of samples in different absolute error ranges. Less than fifteen percent of all predictions are off by more than ten watts.

A closer examination of the high error forecasts shows a clustering of high error points around specific times of day. Dividing the model fitness outliers into groups based on their location in the model fitness graph (shaded in grey in Fig. 4.3) yields a correlation between high error and time of day.

Fig. 4.12(a) and Fig. 4.12(c) identify regions where the ANN predicts more than it should have. The corresponding histograms in Fig. 4.12(b) and Fig. 4.12(d) show the frequency of these forecasts throughout the day. Peaks in the fourteen and fifteen hour range (two p.m. to four p.m.) indicate that these over predictions are

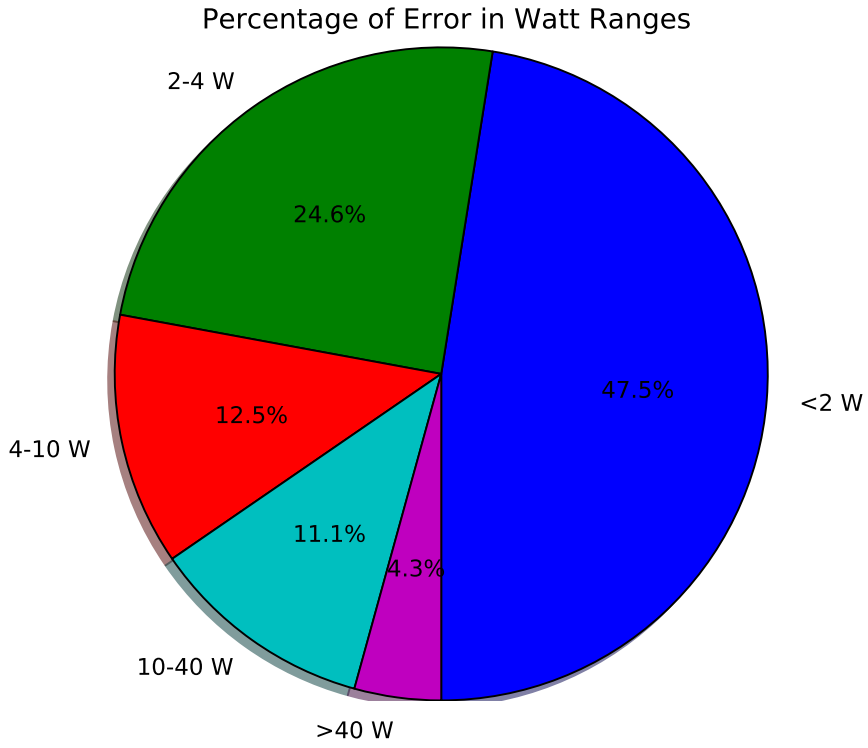
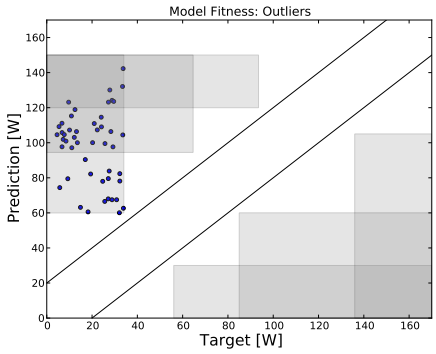


Figure 4.11: A pie chart of the proportion of predictions in different error categories. Nearly half (47.5%) of all predictions have an error of less than two [W].

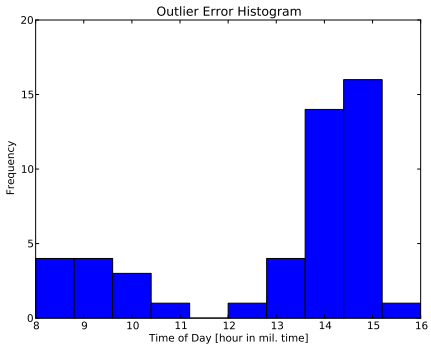
consistent with the time that the panel begins to loose light due to the setting sun and local obstructions. Similarly, Fig. 4.13(a) and Fig. 4.13(c) define under predictions that, when reviewed by time of day in Fig. 4.13(b) and Fig. 4.13(d) , show a high frequency of these errors occur in the six and seven hour range (six a.m. to eight a.m.), corresponding to the time the solar panel first begins to get light in the mornings. Based on these peaks, two other ranges are identified - night, from four p.m. to six a.m. and day, from eight a.m. to two p.m.

4.1.3 Defining time frames

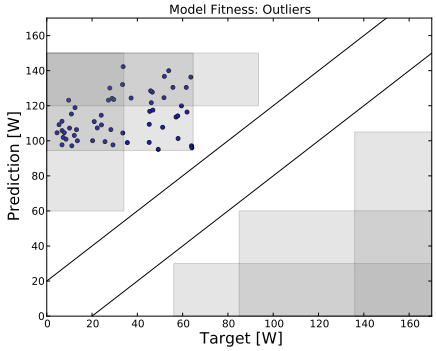
The new model proposes the addition of a pre-processing stage that masks ANN inputs to restrict an input to specific times of day associated with high or low error ranges in previous predictions. Reviewing the error values of the standard preprocessing ANN, error increases occur between six and eight a.m. and two and four p.m., which correspond to the dawn and dusk perceived by the solar panel. These time frames are the most volatile, as they are the time frame when the irradiance changes the greatest (from night to day or day to night) and they are very sensitive to season (the time of sunrise and sunset fluctuate throughout the year). By comparison, the time frames of night and day are relatively stable, their volatility being due to the cloud cover and seasonal intensity of the sun, factors that are equally applicable to the volatility of the sunrise and sunset timeframes. Outside of the peaks, two ranges appear - night, from four p.m. to six a.m. and day, from eight a.m. to two p.m. Analysis of the standard pre-processing ANN shows four distinct time frames, shown in Fig. 4.14 :



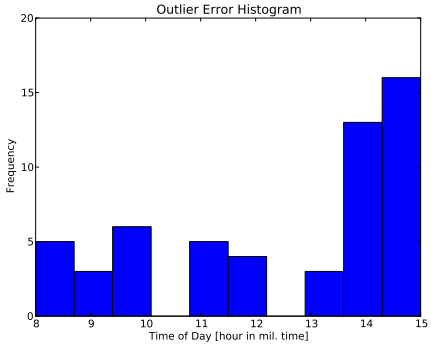
(a) Wide Prediction Range, Low Targets



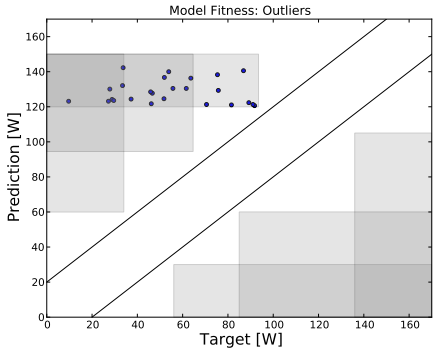
(b) Significantly more dusk errors



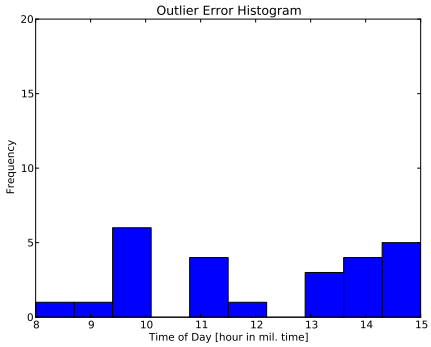
(c) Greatest Overpredictors



(d) Significantly more dusk errors

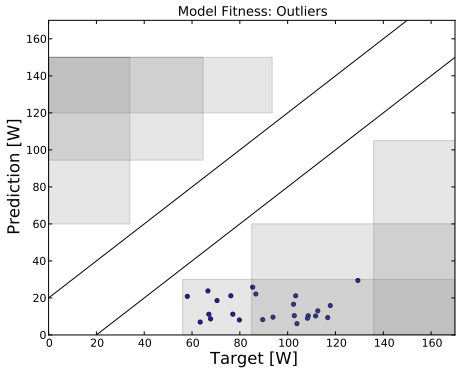


(e) Wide Target Range, High Predictions

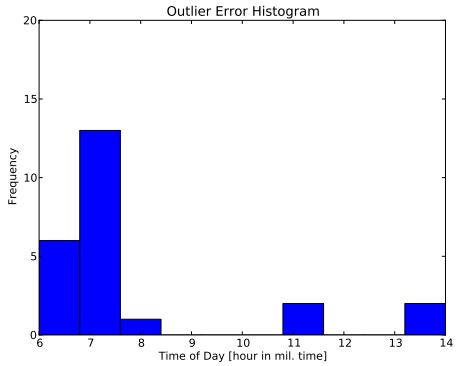


(f) Possible dawn/dusk grouping

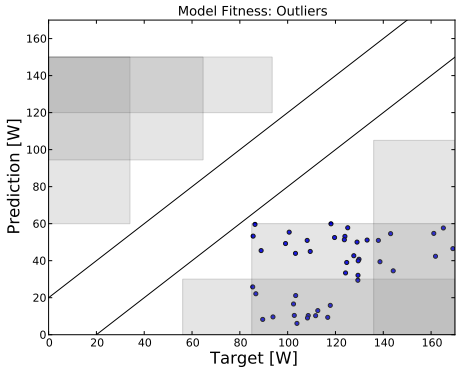
Figure 4.12: Overpredictors: Outliers are filtered by region on the model fitness graph and their frequency by time of day is assessed in corresponding histograms.



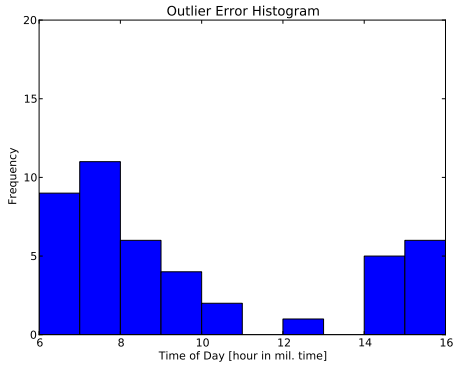
(a) Wide Target Range, Low Predictions



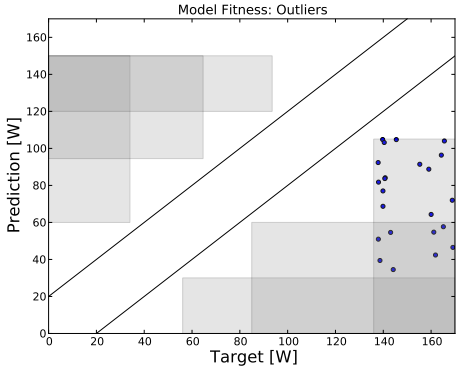
(b) Significantly more dawn errors



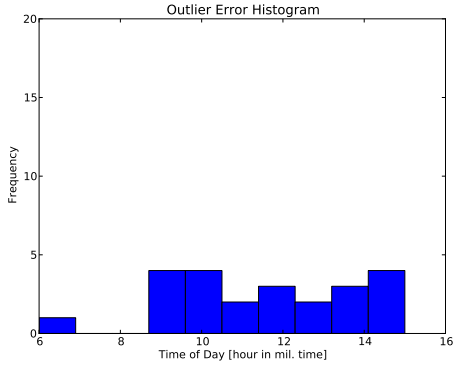
(c) Greatest Underpredictors



(d) Significantly more dawn errors



(e) Wide Prediction Range, High Targets



(f) Even error distribution

Figure 4.13: Underpredictors: Outliers are filtered by region on the model fitness graph and their frequency by time of day is assessed in corresponding histograms.

- Night - when solar energy production is lowest, and most easily predicted
- Sunrise - one of the two time zones with the highest error rate due to the high volatility of the solar energy production data
- Day - when solar energy is consistent (on sunny days) and therefore more predictable than sunrise or sunset
- Sunset - one of the two time zones with the highest error rate due to the high volatility of the solar energy production data

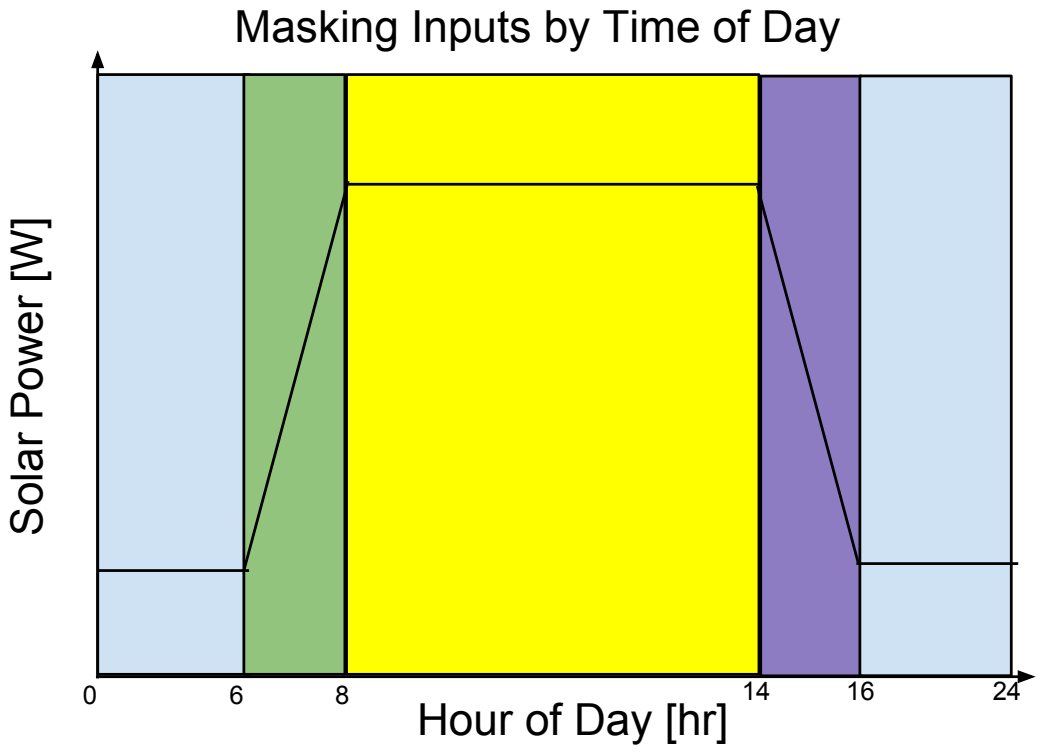


Figure 4.14: The four timeframes of the day, selected by analysis of the cumulative error reviewed by hour. Hour 0 is midnight, hour 12 is noon.

4.2 ANNs with enhanced pre-processing

It should be noted that our use of the terms "sunrise" and "sunset" are not referencing the time of day that the sun encounters the horizon - but the time of day that the sun encounters our instruments - it is sunrise and sunset *from the perspective of our instruments*. In a perfectly flat, empty landscape, these would be the same time - but obstructions, such as trees and buildings, can delay sunrise or hasten sunset from the perspective of our solar panels and irradiance measurement devices.

Using these time frames as the basis for masking the ANN inputs to time frames, the PSP input is replaced with four inputs PSP_night, PSP_dawn, PSP_day, and PSP_dusk. PSP_dawn is the same as the original PSP for timeframes between six a.m. and eight a.m., but zero for all other timeframes. Similarly, PSP_day has the same values as the original PSP for timeframes between eight a.m. and two p.m., but is zero for all other timeframes. The other masked PSP values are created in a similar manner, such that the sum of all masked PSP values would result in the original PSP data set, as shown in Fig. 4.15.

The same masking procedure is performed for MPP and NIP values. Consequently, the original ANN had 3 inputs, PSP, NIP, and MPP; the resulting masked ANN has 12 inputs, a night, dawn, day, and dusk for each of the original ANN inputs of PSP, NIP, and MPP.

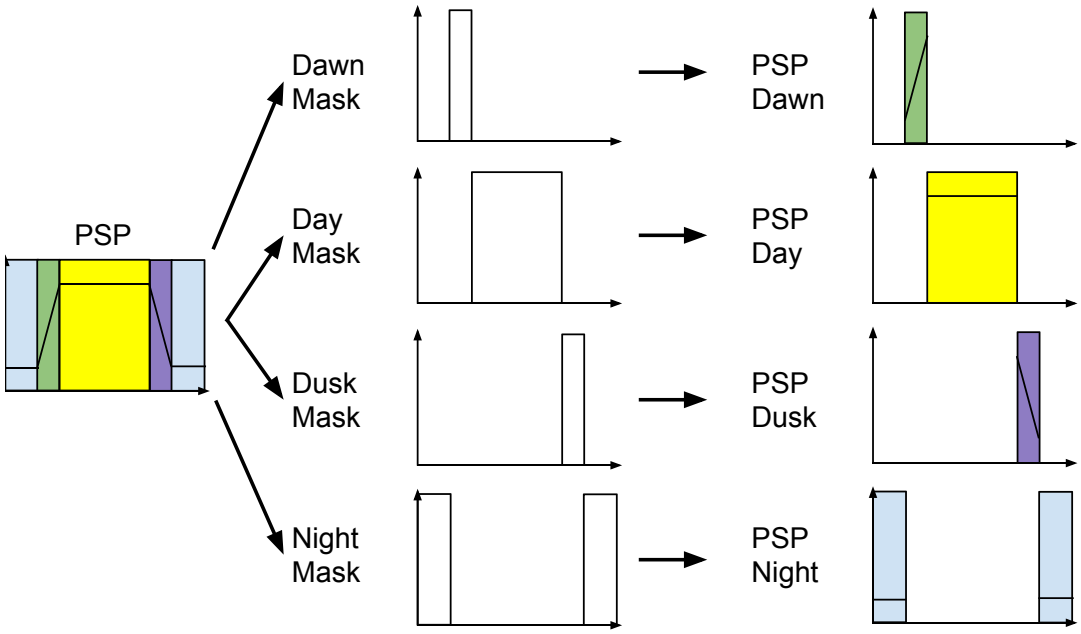


Figure 4.15: The impact of masking on the PSP inputs. The same masking was applied to the NIP and MPP inputs.

Chapter 5

Results and Conclusion

After training the ANN on the new masked inputs, a comparison of the prediction vs. target values results in a correlation for the new inputs. The accuracy measurements for the masked ANN, with an R^2 value of 92.2%, a $RMSE$ of 15.82 W and a MAE of 5.99 W show a marked improvement over the non-masked ANN. Depicted in Fig. 5.1, all error values show an improvement, with the R error value decreasing by over fourteen percent. The result is an improved correlation of roughly 1.33%, from the unmasked correlation of 90.88% to the masked correlation of 92.2%. There was a potential of improvement 9.12%, of which, a 1.33% is an improvement of 14.5% of the total possible improvement possible.

Increases in prosumer renewable energy will require software to automate the management of power produced and consumed in the home. The model presented offers a solution to expected challenges of incorporating these changes at the prosumer level. The development of the RE Lab at NASA Ames has resulted in a data warehouse

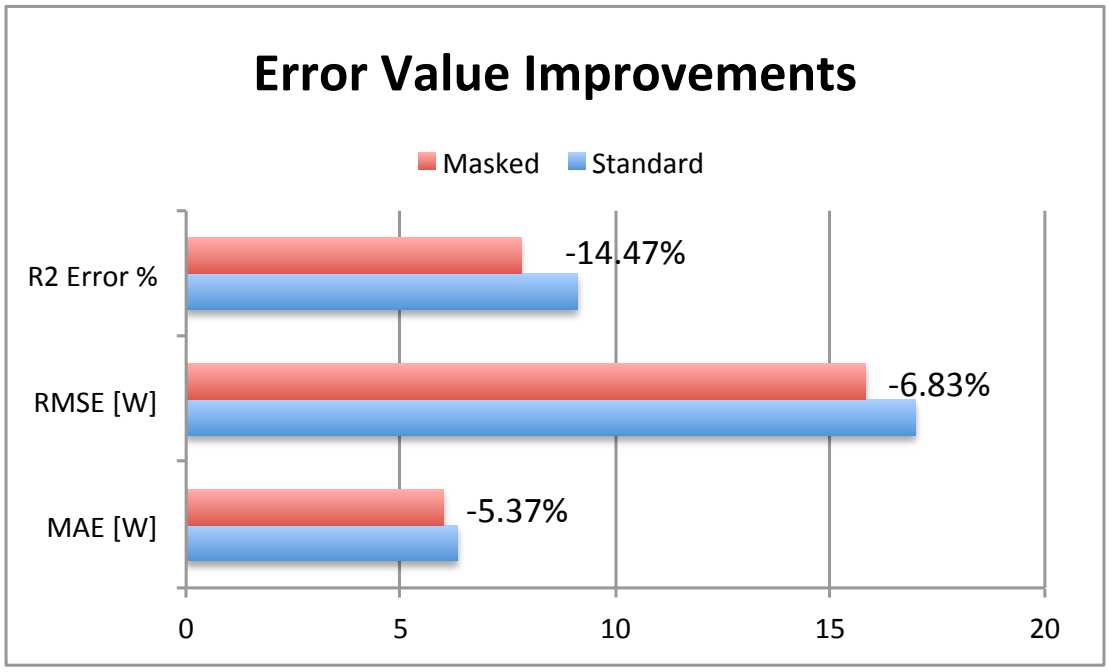


Figure 5.1: Improvements in standard ANN error values using masked pre-processing

of renewable energy and solar irradiance measurements that have been used in ANNs to identify patterns in data history that allow accurate forecasting of future events. Forecasts of power consumption in a home were developed using standard ANN models that predicted future power use within one-third of a kWh. Wind data forecasts based on wind speed classifications available at NASA Ames's Moffett Field Airfare base proved to be accurate within 2.5 mph. Finally, solar ANNs have been developed for the prediction of solar power production that use new masking techniques to improve the accuracy of forecasts by 1.3%, a decrease in correlation coefficient dissociation of nearly 14.5%.

There are many opportunities for future research on this topic. There are several other machine learning techniques for renewable energy and power consumption forecasting. Preliminary work done by Samuel Kahn on the use of Random Forest algorithms for improved solar prediction looks promising [9]. The analysis applied to solar power ANN inputs can be tailored to the consumption and wind power models presented here, with the potential of improving forecast accuracy. Increased input sources, such as weather data and weather forecasts can be used to improve solar and wind forecasts. Further, weather data can also be a valuable input to power consumption ANN forecasts for individual homes. The addition of power pricing data will allow work to begin on creating a prosumer power management algorithm and the development of smart appliance and power storage models would offer opportunities to develop a storage decision model.

Bibliography

[1] U.S. Energy Information Administration. Monthly energy review december 2014. Technical Report DOE/EIA-0035(2014/12), U.S. Department of Energy, Washington, DC, December 2014.

[2] F Almonacid, Eduardo F Fernández, P Rodrigo, PJ Pérez-Higueras, and C Rus-Casas. Estimating the maximum power of a high concentrator photovoltaic (hcpv) module using an artificial neural network. *Energy*, 2013.

[3] Ali N Celik and Tariq Muneer. Neural network based method for conversion of solar radiation data. *Energy Conversion and Management*, 67:117–124, 2013.

[4] AU Chávez-Ramírez, V Vallejo-Becerra, JC Cruz, R Ornelas, G Orozco, R Muñoz-Guerrero, and LG Arriaga. A hybrid power plant (solar–wind–hydrogen) model based in artificial intelligence for a remote-housing application in mexico. *International Journal of Hydrogen Energy*, 38(6):2641–2655, 2013.

[5] California Public Utilities Commission. Renewables portfolio standard quarterly report. Technical report, California Public Utilities Commission, San Francisco, CA, October 2014.

[6] S. Gosav and M. Praisler. The influence of input data preprocessing and of learning error on the performances of ANN systems identifying amphetamines. In *Automation Quality and Testing Robotics (AQTR), 2010 IEEE International Conference on*, volume 2, pages 1–5, May 2010.

[7] Ercan Izgi, Ahmet Öztopal, Bihter Yerli, Mustafa Kemal Kaymak, and Ahmet Duran Şahin. Short–mid-term solar power prediction by using artificial neural networks. *Solar Energy*, 86(2):725–733, 2012.

[8] Sungmoon Jung and Soon-Duck Kwon. Weighted error functions in artificial neural networks for improved wind energy potential estimation. *Applied Energy*, 111:778–790, 2013.

[9] Samuel Kahn. private communication, 2013.

[10] Alireza Kahrobaeian and YA-RI Mohamed. Interactive distributed generation interface for flexible micro-grid operation in smart distribution systems. *Sustainable Energy, IEEE Transactions on*, 3(2):295–305, 2012.

[11] Jiaqi Liang, Ganesh K Venayagamoorthy, and Ronald G Harley. Wide-area measurement based dynamic stochastic optimal power flow control for smart grids with high variability and uncertainty. *Smart Grid, IEEE Transactions on*, 3(1):59–69, 2012.

[12] Hui Liu, Hong-qi Tian, Di-fu Pan, and Yan-fei Li. Forecasting models for wind

speed using wavelet, wavelet packet, time series and artificial neural networks. *Applied Energy*, 107:191–208, 2013.

[13] A Mellit, M Menghanem, and M Bendekhis. Artificial neural network model for prediction solar radiation data: application for sizing stand-alone photovoltaic power system. In *Proceedings of IEEE power engineering society*, volume 1, pages 40–44. IEEE, June 2005.

[14] A Mellit, S Sağlam, and SA Kalogirou. Artificial neural network-based model for estimating the produced power of a photovoltaic module. *Renewable Energy*, 60:71–78, 2013.

[15] Kostas Metaxiotis. *Intelligent information systems and knowledge management for energy applications for decision support, usage, and environmental protection.* IGI Global, Hershey, PA, 2010.

[16] M Mohammadi, SH Hosseinian, and GB Gharehpetian. Optimization of hybrid solar energy sources/wind turbine systems integrated to utility grids as microgrid (MG) under pool/bilateral/hybrid electricity market using PSO. *Solar energy*, 86(1):112–125, 2012.

[17] President of the United States of America. Federal leadership on energy management. Presidential memorandum, Office of the President, United States of America, Washington, DC, December 2013.

[18] D.A. O'Leary, J. Shattuck, and J. Kubby. An online, interactive renewable energy laboratory. *Education, IEEE Transactions on*, 55(4):559–565, Nov 2012.

[19] R. Palma-Behnke, C. Benavides, E. Aranda, J. Llanos, and D. Saez. Energy management system for a renewable based microgrid with a demand side management mechanism. In *Computational Intelligence Applications In Smart Grid (CIASG), 2011 IEEE Symposium on*, pages 1–8, April 2011.

[20] BA Paya, II Esat, and MNM Badi. Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. *Mechanical systems and signal processing*, 11(5):751–765, 1997.

[21] Hugo TC Pedro and Carlos FM Coimbra. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, 86(7):2017–2028, 2012.

[22] Ahmed Yousuf Saber and Ganesh Kumar Venayagamoorthy. Plug-in vehicles and renewable energy sources for cost and emission reductions. *Industrial Electronics, IEEE Transactions on*, 58(4):1229–1238, 2011.

[23] Nicholas Sapankevych and Ravi Sankar. Time series prediction using support vector machines: a survey. *Computational Intelligence Magazine, IEEE*, 4(2):24–38, 2009.

[24] Olle Sundstrom and Carl Binding. Flexible charging optimization for electric vehicles considering distribution grid constraints. *Smart Grid, IEEE Transactions on*, 3(1):26–37, 2012.

[25] A Will, J Bustos, M Bocco, J Gotay, and C Lamelas. On the use of niching genetic algorithms for variable selection in solar radiation estimation. *Renewable Energy*, 50:168–176, 2013.

Appendix A

Explanation of Graphs

Included below are explanations of common graphing techniques used in the assessment of ANN predictions.

For an ANN making a forecast, two values are used to assess the quality of the forecast, the prediction and the target. If an ANN is designed to offer forecasts of power consumption of a home for two hours in the future, it would receive several inputs for past and present power data. From this data, it would offer a prediction about power consumption in two hours. This is the "prediction" value. Two hours later, the measurement of how much power is consumed is the "target" value. The prediction is the forecasted value, the target is the resulting true value. The difference between the prediction and the target is the error - how incorrect the ANN prediction turned out to be.

The fitness graph is a standard analysis tool that displays the target and prediction values and gives a sense of the accuracy of the model. The ANN is given

a set of data (test data) upon which it makes predictions. For each prediction, a target/prediction pair (target, prediction) is generated and graphed, with target values along the x-axis and predictions along the y-axis. If the ANN makes perfect predictions, where every prediction is equal to every target, the result is a perfect line. Fig. A.1 shows such an outcome.

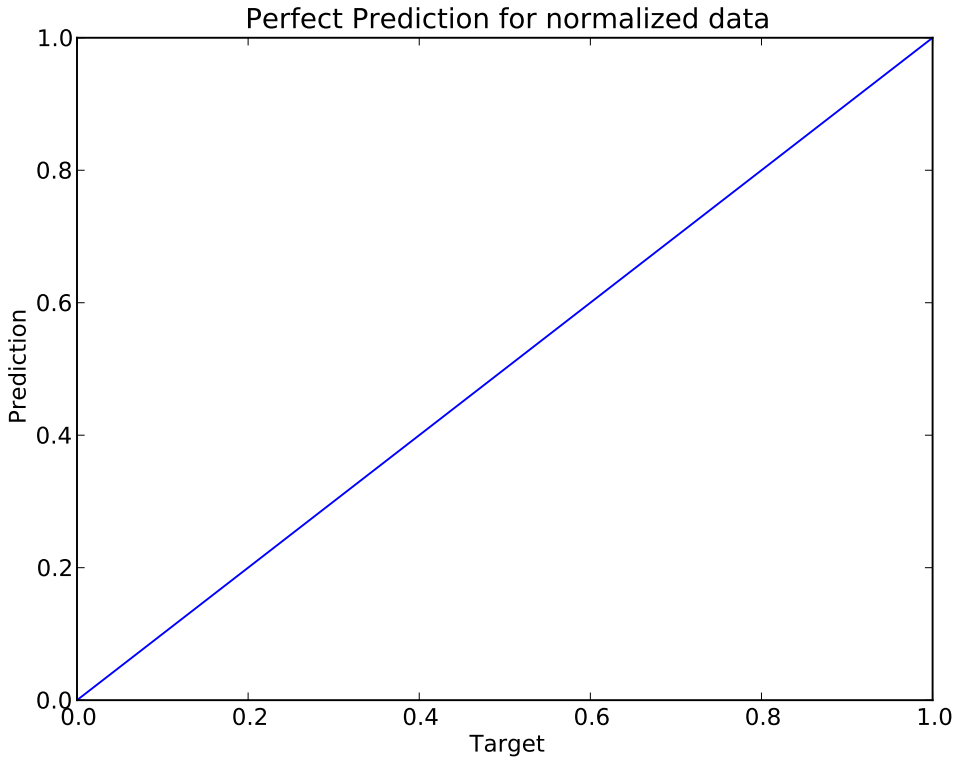


Figure A.1: An ANN that makes perfect predictions creates a forty-five degree line in the fitness plot.

Of course, such perfect predictions are rare, and not very interesting. An example of an excellent prediction, with enough variation to help familiarize yourself with the basics of this tool is offered in Fig. A.2. These values are in line with the

literature review values found in table 2.1.

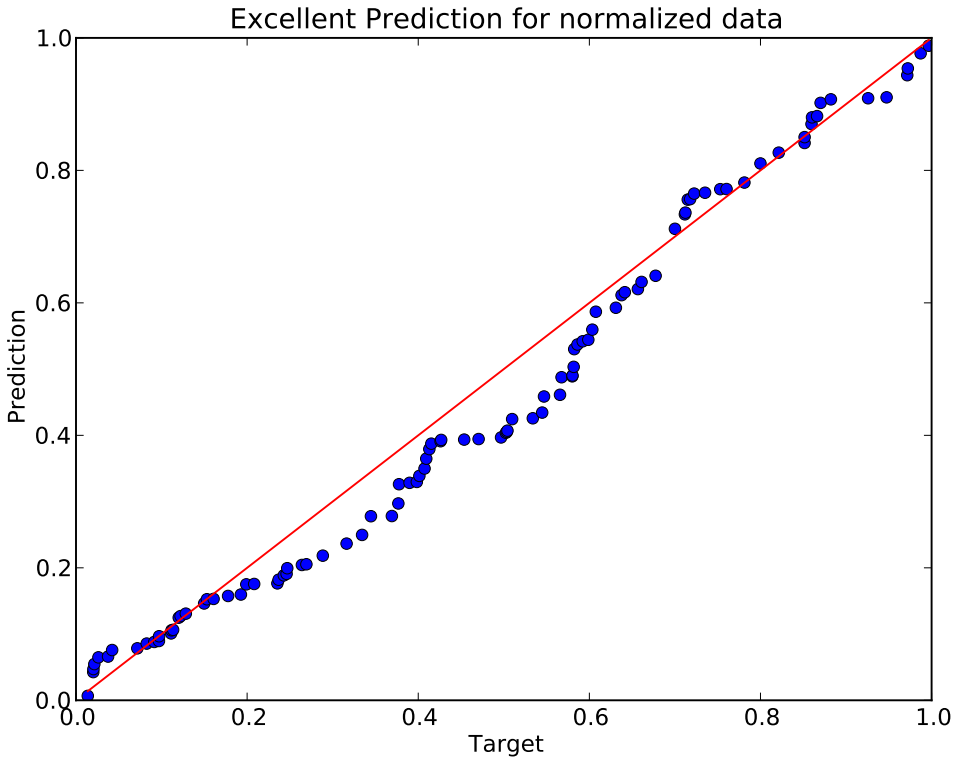


Figure A.2: An ANN that makes perfect predictions creates a forty-five degree line (red) in the fitness chart, the predictions made that were close, but not perfect, are shown as blue circles in the graph.

Appendix B

ANN Accuracy Assessment Metrics

Several benchmarks are used to compare the quality of different ANN and machine learning tools. Table B.1 defines the terms used in the ANN assessments.

Variable	Definition
m	The number samples in the evaluation
t	The sample index
P_t	Power produced at time t
$\hat{P}_t$	The forecasted power for time t
$\bar{P}_t$	Average Power $= \frac{1}{m}\sum_{t=1}^{m} P_t$

Table B.1: A glossary of variables used in ANN error equations

Common error assessment metrics are:

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{m}\sum_{t=1}^{m}\left|P_t - \hat{P}_t\right| \tag{B.1}$$

- Mean Bias Error (MBE)

$$MBE = \frac{1}{m}\sum_{t=1}^{m} P_t - \hat{P}_t \tag{B.2}$$

- Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{m}\sum_{t=1}^{m}\left(P_t - \hat{P}_t\right)^2} \tag{B.3}$$

- normalized Root Mean Square Error (nRMSE)

$$nRMSE = \sqrt{\frac{\sum_{t=1}^{m}\left(P_t - \hat{P}_t\right)^2}{\sum_{t=1}^{m} P^2}} \tag{B.4}$$

- Correlation Coefficient (R)

$$R^2 = 1 - \frac{\sum_{t=1}^{m}\left(P_t - \hat{P}_t\right)^2}{\sum_{t=1}^{m}\left(P_t - \bar{P}_t\right)^2} \tag{B.5}$$

Appendix C

Review of the Software and Hardware of the RE Lab at NASA Ames

The RE Lab online learning website is primarily served by a Raspberry Pi computer controlling custom circuits and interfacing with measuring devices. The custom circuits allow it to control the wattsun az 225 solar tracking array's orientation. The measurement connections allow it to pull data and store it in the database. Primarily, you should consider the separate functions of the device based on connection:

- bluetooth - connection to the oreintometer
- USB - connection to the solarPro IV curve measurement device - also provides irradiance measurements from the PSP and NIP
- USB - connection to custom circuits that take digital inputs from the raspberry pi and convert them to analog signals that the circuitry of the AZ 225 interprets

as incoming light data. A command to the raspberry pi to move the az 225 to the north is sent to an arduino nano via USB. The nano converts the digital request to an analog signal to the AZ 225. The AZ 225 receives that signal as an input from it's light sensor indicating that the sun is north of the tracker's current position. The tracker then moves.

- RJ 45 - Ethernet connection to the USCS network. allows us to send data to the UCSC SQL databases.