MACHINE LEARNING TECHNIQUES FOR WEATHER FORECASTING

by

WILLIAM SAMUEL SANDERS

(Under the Direction of Frederick Maier)

ABSTRACT

Machine learning models were developed in order to forecast weather variables such as solar radiation, temperature, and wind speed for one to 24 hours in advance. Weather predictions and ground truth weather observations were sourced from the National Oceanic and Atmospheric Administration (NOAA) and the Georgia Automated Environmental Monitoring Network (GAEMN) for five cities in Georgia. Results indicate that incorporating weather forecasts becomes increasingly more important for accurate solar radiation prediction at longer prediction windows, and also that postprocessing of NOAA's weather forecasts can drastically improve accuracy beyond usage of the raw forecasts alone.

INDEX WORDS: weather forecasting, solar radiation, grib, postprocessing

MACHINE LEARNING TECHNIQUES FOR WEATHER FORECASTING

by

WILLIAM SAMUEL SANDERS

B.S., Florida State University, 2005

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2017

MACHINE LEARNING TECHNIQUES FOR WEATHER FORECASTING

by

WILLIAM SAMUEL SANDERS

Major Professor: Frederick Maier

Committee: Khaled Rasheed
Walter D. Potter

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2017

TABLE OF CONTENTS

Page

LIST OF TABLES vi

LIST OF FIGURES viii

CHAPTER

1 WEATHER VARIABLE FORECASTING 1

1.1 INTRODUCTION 1

1.2 REFERENCES 4

2 SOLAR RADIATION PREDICTION IMPROVEMENT USING WEATHER FORECASTS 5

2.1 ABSTRACT 6

2.2 INTRODUCTION 6

2.3 LITERATURE REVIEW 7

2.4 DATA ACQUISITION 10

2.5 DATA VISUALIZATION 11

2.6 RESULTS: ONE-HOUR PREDICTIONS 12

2.7 RESULTS: 24-HOUR PREDICTIONS 21

2.8 CONCLUSION 27

2.9 REFERENCES 29

3 A COMPARISON OF MACHINE LEARNING TECHNIQUES FOR POSTPROCESSING WEATHER FORECASTS 32

3.1 ABSTRACT...33

3.2 INTRODUCTION 33

3.3 POSTPROCESSING 35

3.4 LITERATURE REVIEW 37

3.5 METHODOLOGY 41

3.6 RESULTS: AIR TEMPERATURE 44

3.7 RESULTS: ADDITIONAL WEATHER VARIABLES 50

3.8 CONCLUSION..51

3.9 REFERENCES 53

4 CONCLUSION & FUTURE DIRECTIONS 58

4.1 CONCLUSION..58

4.2 FUTURE DIRECTIONS 59

4.3 REFERENCES 61

BIBLIOGRAPHY..62

LIST OF TABLES

Page

Table 2.1: A comparison of error rates for different model inputs to Random Forest when predicting one-hour solar radiation for five Georgia cities 15

Table 2.2: Properties of five selected Georgia cities based on GAEMN data (2003-2013) 18

Table 2.3: Average recorded solar radiation at midnight for five Georgia cities (2003-2013) 20

Table 2.4: A comparison of error rates for different model inputs to Random Forest when predicting one-hour solar radiation for five Georgia cities 20

Table 2.5: A comparison of machine learning methods for predicting 24-hour solar radiation using historical GAEMN weather observations as model inputs 24

Table 2.6: A comparison of error rates for different model inputs to Random Forest when predicting 24 hour solar radiation for five Georgia cities 26

Table 3.1: Mean Absolute Error of NOAA 24-hour forecasts for eight forecasted variable for five Georgia cities, June 10, 2016 to June 10, 2017 43

Table 3.2: Average 24-hour air temperature forecast bias by time of day 44

Table 3.3: Average 24-hour air temperature forecast bias by month 45

Table 3.4: Average 24-hour air temperature forecast bias by time of day for Griffin, Georgia 46

Table 3.5: A comparison of machine learning methods for postprocessing 24-hour NAM temperature predictions over five cities in Georgia 47

Table 3.6: Comparison of a model incorporating forecasts from surrounding GRIB cells 48

Table 3.7: Analysis of cross-predictive accuracy across cities 49

Table 3.8: 24 hour forecast error rates for eight weather variables 51

LIST OF FIGURES

Page

Figure 2.1: Solar radiation data by hour of the day collected from the Griffin, Georgia weather station (2003-2013)...11

Figure 2.2: Solar radiation data by month of the year collected from the Griffin, Georgia weather station (2003-2013)...12

Figure 2.3: A map of the five Georgia cities chosen for solar radiation prediction 17

Figure 2.4: Average solar radiation by year for five selected cities in Georgia (2003-2013) 19

CHAPTER 1

WEATHER VARIABLE FORECASTING

1.1 INTRODUCTION

The effects of weather permeate nearly every aspect of our everyday lives, from travel to commerce to government. The average U.S. adult consults weather forecasts 115 times per month, for a total of more than 300 billion forecasts used per year (Lazo, Morss, & Demuth, 2009). Providing accurate weather forecasts can lead to more effective planning and resource allocation by persons and businesses alike; it is estimated that U.S. electric utilities save $150 million annually by using 24-hour air temperature forecasts to estimate load demand (American Meteorological Society, 2015).

The purpose of this thesis was to analyze current research on weather forecasting and compare machine learning techniques in the field. Various inputs to these machine learning models were also tested to determine the usefulness of each, as measured by their contribution to lowering the difference between predicted values and the ultimate observed ground truth. Data for these analyses was sourced from the Georgia Automated Environmental Monitoring Network (GAEMN) and the National Oceanic and Atmospheric Administration (NOAA) for five cities in Georgia.

The GAEMN data consists of 16 observed weather variables: air temperature (ºC), humidity (%), dew point (ºC), vapor pressure (kPa), vapor pressure deficit (kPa), barometric pressure (kPa), wind speed (m/s), wind direction (º), SD, maximum wind speed (m/s), pan, solar radiation, total solar radiation (KJ/m^2), photosynthetically active radiation ($umole/m^2s$), and two rainfall measurements (mm). In addition, time-of-day and day-of-year inputs were added to all models. The NOAA historical weather forecasts were downloaded from their servers in the form of GRIB files, a standardized format commonly used for weather forecasts and observations (World Meteorological Organization, 2003). The machine-learning software Weka (Hall et al., 2009) was chosen for training and testing all models generated in this thesis.

This thesis attempted to analyze the effect of using NOAA predictions for air temperature, precipitation rate, visibility, wind speed, wind direction, dew point temperature, air pressure, and relative humidity as a means of forecasting solar radiation over one-hour and 24-hour time frames. While using only persistence models and historical collected data is one component of accurate solar radiation predictions, incorporating weather forecasts is undoubtedly important as well. At longer prediction windows, it makes intuitive sense that the impact of weather forecasts will begin to overtake the impact of using historical data alone; Chapter 2's goal was to quantify that impact. In addition, weather variable forecasts from areas immediately surrounding the target area were incorporated in order to smooth out forecasting error. Larson et al. (2016) reported that global horizontal irradiance (GHI) prediction models which use numerical weather prediction (NWP) data can usually be improved by averaging the GHI forecasts from NWP grid points surrounding the target site, and Chapter 2 attempts to leverage those findings.

It was found that including weather forecast data in the prediction models resulted in a 7.6% reduction in mean absolute error (MAE) for one-hour predictions when compared to using historical observations alone, and a 40.2% reduction in MAE for 24-hour predictions. Results from several machine learning techniques were compared, with Random Forests achieving the lowest error rate. In addition, inclusion of weather forecasts from nearby areas resulted in a 4.6% lower mean absolute error (MAE) in one-hour predictions and a 20.9% reduction in 24-hour predictions when averaged across the five cities studied.

Chapter 3 analyzes the effect of incorporating the historical biases of weather predictions in order to improve them, a widely-used method known as postprocessing. Due to the existence of persistent, statistically-significant time- and location-based biases, nearly all current numerical weather forecasting systems apply some form of postprocessing to their raw input data (Marzban, 2003). While linear regression Model Output Statistics (MOS) is commonly used in practice, there is a minimal number of current studies comparing its results to those achievable by more advanced models. This thesis compared several different machine learning techniques and found that an ensemble model stacking a Random Forest with an artificial neural network (ANN) was found to reduce prediction error over MOS on seven of the eight weather variables studied (air temperature, cloud cover, visibility, wind speed, wind direction, dew point temperature, air pressure, and humidity). The inclusion of additional forecasted weather variables from areas immediately surrounding the target location was not found to have an impact on prediction error, a contrast to the solar radiation prediction results found in Chapter 2.

1.2 REFERENCES

American Meteorological Society. (2015). Weather Analysis and Forecasting.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reuteman, P., & Witten, I. (2009). The WEKA Data Mining Software: An Update. SIGKDD Explorations, vol. 11 (1).

Larson, D. P., Nonnenmacher, L., & Coimbra, C. F. (2016). Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest. Renewable Energy, 91, 11-20.

Lazo, J. K., Morss, R. E., & Demuth, J. L. (2009). 300 billion served: Sources, perceptions, uses, and values of weather forecasts. Bulletin of the American Meteorological Society, 90(6), 785-798.

Marzban, C. (2003). Neural networks for postprocessing model output: ARPS. Monthly Weather Review, 131(6), 1103-1111.

World Meteorological Organization. (2003). Introduction to GRIB Edition 1 and GRIB Edition 2. Retrieved from https://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/Introduction_GRIB1-GRIB2.pdf.

CHAPTER 2

SOLAR RADIATION PREDICTION IMPROVEMENT USING WEATHER FORECASTS[1]

[1] Sanders, W., Barrick, C., Rasheed, K., and Maier, F. To be submitted to *Solar Energy*.

2.1 ABSTRACT

Prediction models were developed to generate forecasts of solar radiation, and, by proxy, expected solar plant power output, for one hour and 24 hours in the future. Data was sourced from the Georgia Automated Environmental Monitoring Network (GAEMN) and the National Oceanic and Atmospheric Administration (NOAA) for five cities in Georgia for the years 2003-2013. Early predictive models only made use of historical recorded solar radiation and other weather phenomena as inputs, while later models incorporated weather forecasts for the target area and surrounding areas. Including weather forecast data in the prediction models resulted in a 7.6% reduction in mean absolute error (MAE) for one-hour predictions when compared to using historical observations alone, and a 40.2% reduction in MAE for 24-hour predictions. Results from several machine learning techniques were compared, with Random Forests achieving the lowest error rate. These results indicate that weather forecasts are an important component of accurate solar radiation prediction even over short- and medium-term prediction timeframes, and the inclusion of the surrounding geographical area in addition to the target city is an important component of these predictions.

2.2 INTRODUCTION

Solar power - the conversion of sunlight into electricity - is forecasted to become the world's largest source of electricity by 2050 (International Energy Agency, 2014). Solar power is unlike other forms of energy production such as nuclear power or fossil fuels in that it relies heavily on an external agent as the key determiner of power output. Output is discernibly higher on sunny days during the summer, and lesser on cloudy days or during the winter when the sunlight hits the earth at a less direct angle. Making accurate predictions about expected solar power output is paramount to efficiently harvesting power from a solar plant.

There are two categories of data commonly used in predicting solar radiation. One consists of historical collected data on solar radiation, temperature, and other meteorological phenomena. The other is comprised of forecasted weather variables and atmospheric movements. Beyond very short-term prediction windows, these forecasts are crucial to making accurate solar radiation predictions, and this paper will attempt to quantify their contribution.

First, predictions for one hour and 24 hours into the future using only historical collected data will be presented and analyzed. Then, additional forecasted data will be used to make new one hour and 24 hours predictions for the same historical data range, and it will be shown that significant reductions in error rates are possible by virtue of their inclusion. Analysis will first be performed on a case study city of Griffin, Georgia, and then extended to four other cities throughout Georgia. The results will show that incorporating weather forecasts slightly increases prediction accuracy in one-hour results and greatly increases it in 24-hour results, and using forecasts from surrounding areas in addition to the target city increases accuracy further than using forecasts for the target city alone.

2.3 LITERATURE REVIEW

Solar radiation is commonly recorded in watts per meter squared (watts/m^2), and error values in this paper are reported in these units. As with most machine learning models, there are many ways to report the accuracy of the methods applied in this paper: correlation coefficient, root mean squared error, mean absolute error, relative absolute error, and others. When reviewing previous papers, no one method stood out as more popular than any others for this particular area. This results in a lack of directly comparable performance metrics throughout current solar radiation forecasting literature (Hamilton, 2016). Mean absolute error (MAE) does not overweight outliers, as opposed to other metrics such as root mean squared error (RMSE), and therefore MAE was chosen for this paper as the benchmark for all tests. All error rates reported here can be assumed to denote the mean absolute error in watts/m^2 unless stated otherwise.

Many attempts at predicting solar radiation have been done in recent years. Pedro and Coimbra (2016) attempted one-hour predictions using only one-hour averaged radiation data collected from a one-megawatt plant in Southern California, and were able to achieve an MAE of 42.96 watts/m^2 by using a genetic algorithm to mutate and mate a series of artificial neural networks (ANNs). Their time period spans from November 3, 2009 to August 15, 2011, and they note that additional weather variables, such as global horizontal irradiance, cloud cover, and wind speed and direction were not used in their study, as it was focused on the use of endogenous variables for forecasting power output.

Spokas and Forcella (2006) obtained one-hour MAEs averaging 57 watts/m^2 over 18 different stations from 1997-2004 when adding climate data such as daily precipitation and daily air temperature extremes, as well as direct normal irradiance (DNI). Their site list encompassed a wide variety of climates: New York, Iowa, Australia, Florida, and more, and therefore may taken as a more accurate approximation of reasonable solar radiation prediction error rates than many other papers which just focused on one particular weather station. Sfetsos and Coonick (2000) evaluated several methods for predicting hourly solar radiation and achieved best results (RMSE 39.32 watts/m^2) with the Levenberg-Marquardt algorithm, a backpropogation technique often used to solve non-linear least squares problems. Their analysis was focused on the French island of Corsica, a Mediterranean climate, during the late spring and early summer of 1996. Although their approach began with simply using lagged solar radiation observations as well as some calculated autocorrelation and crosscorrelation statistics, they later added temperature and wind direction as additional useful exogenous parameters.

Zervas et al. (2008) investigated the correlation of weather conditions and daylight duration on the distribution of global solar irradiance (GSI) in Athens, Greece. They broke up daily weather conditions into six classifications (clear, few clouds, partly cloudy, cloudy, heavy clouds, or rainfall) and used standard equations to calculate the number of daylight hours, then fed these inputs into a radial basis function (RBF) neural network to predict daily GSI. They found a coefficient of determination R^2 of 0.985 on a validation set, highlighting the impact of cloud cover in predictions.

Feudo et al. (2015) found similar correlations between cloud cover and accuracy of solar radiation predictions when making hourly predictions from July 2013 to December 2013 in South Italy. They were able to achieve a mean absolute error of 26.7 watts/m^2 in clear conditions versus 43.1 watts/m^2 in cloudy conditions . They noted that a variability in the optical depth of clouds, suggesting cloud cover is not a uniform variable that can be expressed as a simple percentage reading, but rather could affect surface readings differently depending on the different types of clouds present.

Satellite images can also be used as model inputs; when combined with estimates for clear-sky solar radiation, these forecasts can be quite effective (Miller et al., 2013; Linares-Rodriguez et al., 2013). An alternative to these satellite-based approaches are methods using Numerical Weather Prediction (NWP) for regional solar radiation prediction (Lara-Fanego et al., 2012; Perez et al., 2013). Ruiz-Arias, Quesada-Ruiz, Fernández, and Gueymard (2015) suggest several advantages for NWP methodologies over satellite-based imagery, including the fact that NWP models comprehensively simulate the entire atmospheric system, including wind, temperature, humidity, and other variables. For forecast horizons extending beyond 6 hours, NWP models tend to outperform those without forecasts (Cheung et al., 2015).

The results found in these previous works indicate that present- and forward-looking approaches such as satellite images and forecasts of multiple weather variables are important components to predicting solar radiation, as opposed to using historical solar radiation observations alone. This paper will attempt to further quantity their contribution.

2.4 DATA ACQUISITION

The Georgia Automated Environmental Monitoring Network (GAEMN) was developed in 1990 to develop a climatic data base used for various agricultural, ecological, water management, and other environmental-based research (Hoogenboom, Verma, & Threadgill, 1990). A variety of weather- and environment-related data including air and soil temperature, barometric pressure, wind speed and direction, rainfall, and solar radiation, is collected in 15-minute intervals across the state of Georgia. This GAEMN data was analyzed for the time period of 2003 to 2013 for five cities in Georgia. The data included 43 observed fields, although only a handful were proven relevant for solar radiation prediction.

In addition to these historical data observations, this paper makes use of weather forecasts provided by The National Oceanic and Atmospheric Administration's (NOAA). This data is disseminated in the GRIB file format, a compact binary format commonly used to store historical and forecasted weather data due to its self-description and flexibility (World Meteorological Organization, 2003). Each GRIB file describes a particular geographical region for a single date, and internally splits this region into a grid of cells of a consistent size. For each cell, attribute values are listed describing weather attributes in the cell at that time, or, in the case of weather forecasts, at a specified time in the future.

As a final note, for the purposes of this paper a "forecast" will always reference a NOAA weather attribute forecasted variable as obtained through GRIB data or an interpolation, whereas a "prediction" will reference a prediction generated for a future solar radiation value.

2.5 DATA VISUALIZATION

As solar radiation measures the amount of solar power received from the sun, its value is highest during the early afternoon hours and drops to near-zero values between sunset and sunrise.

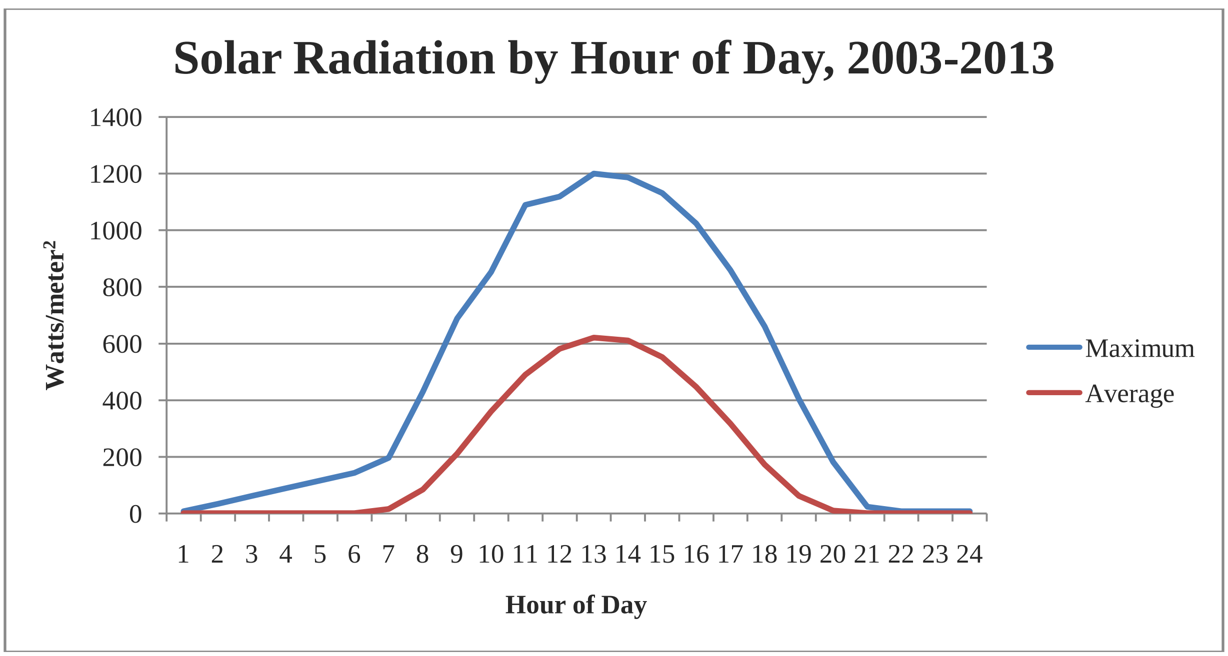

Figure 2.1. Solar radiation data by hour of the day collected from the Griffin, Georgia weather station (2003-2013).

The second significant source of seasonality in solar irradiance corresponds to the time of year. Due to the variation of tilt of the earth's axis during summer and winter months, solar radiation varies significantly throughout the year, peaking in June.

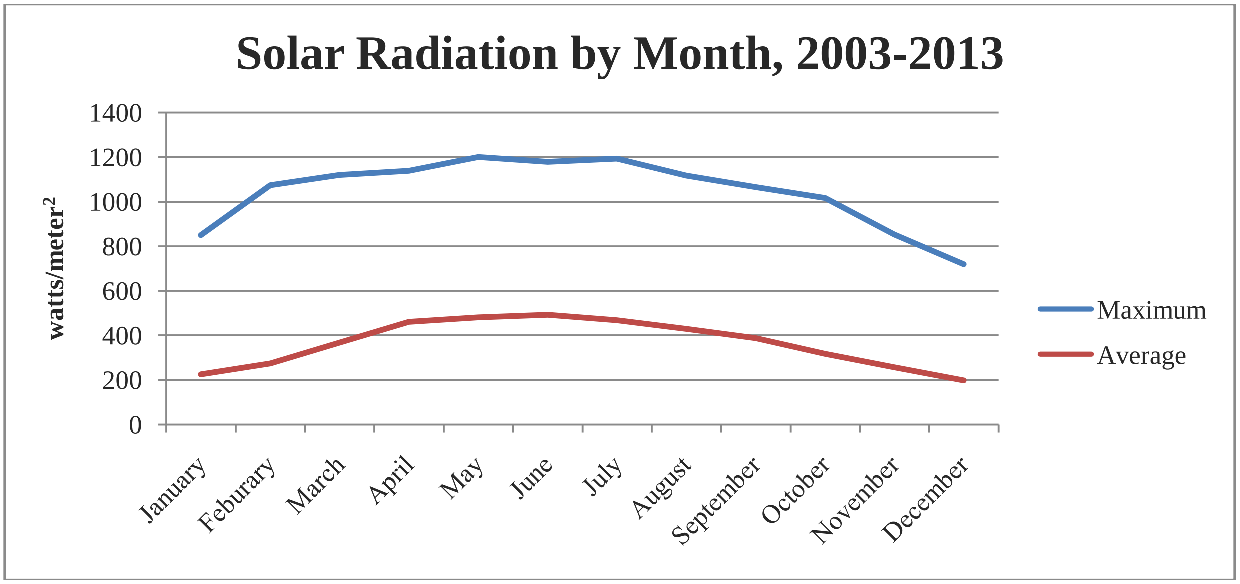

Figure 2.2. Solar radiation data by month of the year collected from the Griffin, Georgia weather station (2003-2013).

2.6 RESULTS: ONE-HOUR PREDICTIONS

In order to compare one-hour solar radiation predictions made with weather forecast data versus those using only historical data, it was necessary to find one-hour weather forecasts which overlapped with the historical GAEMN solar radiation data from 2003-2013. NOAA's Rapid Refresh (RAP) dataset[2] generates data on a 13-km resolution horizontal grid across North America and every hour makes available hourly forecasts going out 18 hours and was chosen for one-hour predictive purposes. While these hourly forecasts have been generated since 2002, only a subset is available online due to storage requirements. RAP data was downloaded for the period of June 22, 2011 to April 30, 2012, and this period was chosen for one-hour predictions. Approximately 2% of the RAP forecast files for this time frame were missing, and these periods

[2] https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/rapid-refresh-rap

were not included in any predictions. Finally, since the GAEMN recorded solar radiation data was recorded at 15-minute intervals but RAP one-hour data forecasts are generated at a one-hour granularity, synthetic RAP weather forecasts were created by linearly interpolating between the hours as appropriate. For example, to create a synthetic one-hour weather forecast at 1:15 pm for 2:15 pm, 75% of the one-hour forecast generated at 1:00 pm (for 2:00 pm) and 25% of the two-hour forecast generated at 1:00 pm (for 3:00 pm) were added together. In this way a dataset of one-hour forecasts every 15 minutes was built, while taking care to only make use of data that would have been available the time of prediction.

Performing these interpolations resulted in 28,896 instances corresponding to 15-minute intervals over 313 different days in the time frame of June 22, 2011 to April 30, 2012. It is important to note that this time frame is missing the high-variability months of May and most of June, therefore, these predictions should not be indicative of raw prediction accuracy but rather are meant to show the relative reduction in error rate when incorporating weather forecasts. Due to the relatively small size of the datasets - 28,896 instances for one-hour predictions and 67,720 instances for 24-hour predictions - 10-fold cross validation in Weka was chosen to compute expected error rates, and it is these results which are reported.

There are two naive approaches to predicting future solar radiation which can be implemented with minimal effort: one uses the current solar radiation as the prediction value and the other uses the average solar radiation for the desired day of year and time of day (as computed from 2003-2013). These persistence models are easy to calculate and serve as useful benchmarks on which to compare future, more advanced methods. Using the current solar radiation as the prediction value one hour into the future yields an MAE of 67.28, and using the average solar radiation for the desired day and time yields an MAE of 60.70. Using the average of these two numbers as the prediction value yields a lower MAE of 52.54, which became the baseline for future testing for the city of Griffin, Georgia.

To improve these predictions, 16 GAEMN weather variables recorded at the time of prediction from the Griffin weather station were added to the model's inputs: air temperature (ºC), humidity (%), dew point (ºC), vapor pressure (kPa), vapor pressure deficit (kPa), barometric pressure (kPa), wind speed (m/s), wind direction (º), SD, maximum wind speed (m/s), pan, solar radiation, total solar radiation (KJ/m^2), photosynthetically active radiation (umole/m^2s), and two rainfall measurements (mm). In addition, the time of day was converted to a decimal ranging from 0.00 to 0.99 that represents the time as a proportion of the day passed (e.g., 6:00 a.m. is encoded as 0.25) and included in the model's inputs, along with the day of the year. These 18 inputs were fed into multiple machine learning techniques in Weka (Hall et al., 2009) in order to predict solar radiation one hour in advance. (All methods used default Weka settings unless stated otherwise.)

As evidenced in *Table 2.1* below, neural nets underperformed expectations, while decision-tree models (in particular, random forest and M5P model trees) emerged as the strongest methods. This finding would prove to be a pattern throughout analysis of this dataset; accordingly, all future one-hour results reported will be from the application of Random Forest to the dataset in question unless stated otherwise. For one-hour predictions, using a Random Forest model resulted in a 10-fold cross-validation MAE of 29.96 watts/m^2 with a correlation coefficient of 0.969.

Table 2.1. A comparison of machine learning methods for predicting one-hour solar radiation for Griffin, Georgia using historical GAEMN weather observations as model inputs.

Machine Learning Method	MAE (watts/m^2)	Time to run (seconds)[3]
Linear Regression	70.79	4
Multilayer Perceptron (Neural Net)[4]	66.18	750
M5P model tree, unpruned	32.00	36
M5P model tree, pruned	34.18	66
Random Tree	43.02	11
REP Tree, unpruned	37.42	7
REP Tree, pruned	37.26	11
Random Forest	29.96[5]	545
Alternating Model Trees[6]	36.92	25
Additive Regression using REP Tree, pruned	36.91	11
Additive Regression using Alternating Model Trees	37.06	341
Bagging M5P, unpruned	31.21	264
Bagging REP Tree, pruned	32.51	33

To further improve these results, sliding windows were incorporated into the model inputs in order to make use of historically recorded solar radiation data instead of simply using the current observation. In addition to the 18 inputs above, 24 inputs were added to the model, corresponding to the recorded solar radiation value for each of the past 24 hours, on the hour. Application of these new inputs to the Random Forest model yielded an MAE of 29.16, a 2.7% improvement over using GAEMN recorded data alone. This value becomes the baseline for comparing models generated with historical data only to those which incorporate weather forecasts.

As yet, only collected historical data has been used to make predictions. To test the hypothesis that using weather forecasts will increase prediction accuracy, eight one-hour RAP

[3] Time to run 10-fold cross-validation in Weka on Windows 7 with AMD-FX 8350 eight-core processor and 32GB memory.

[4] 20% validation set size, maximum 50,000 epochs, 57 hidden nodes

[5] Compare to 33.55 watts/m2 when using Random Forest to predict one-hour solar radiation for all instances in the total dataset from 2003-2013. This discrepancy could be explained by the absence of May and part of June in this smaller training set, and it is not unreasonable to assume that inflating all one-hour error rates reported in this paper by approximately 12% would be a more accurate assessment of raw prediction accuracy over a broader timeframe.

[6] With "build decision tree" set to True

weather forecast variables were added to the above weather observation and sliding window inputs, resulting in a total of 50 inputs to the one-hour model. The forecasted variables used are air temperature, precipitation rate, visibility, wind speed, wind direction, dew point temperature, air pressure, and relative humidity. This data was extracted from the relevant GRIB RAP forecast file, which consists of splitting the continental United States into cells measuring 13 kilometers square. The cell closest to the latitude and longitude of Griffin was extracted from each GRIB file and the pertinent forecast values examined. As noted above, linear interpolations between the one- and two-hour RAP forecasts were used for intra-hour forecasts. The addition of these eight forecasted weather variables to the model resulted in a 2.6% reduction in MAE from 29.16 to 28.40.

Finally, to increase prediction accuracy beyond using NOAA forecasted weather data from the Griffin cell alone, the eight forecasted values were included from each of nine cells in a three-by-three square of cells, with Griffin in the center. This was intended to take into account the not-unlikely possibility of weather in surrounding cells moving into the Griffin cell, a scenario which may not have been fully accounted for in the forecasts. The addition of these additional eight weather forecast values from each of eight surrounding cells brought the number of inputs to 114, and dropped the error rate to 27.43, a 3.4% decrease from using the single cell alone. This could be evidence of information existing in surrounding weather forecasts which is not embedded in the Griffin weather forecasts directly, or it could be caused by reducing the effect of a single inaccurate weather forecast or interpolation. Indeed, global horizontal irradiance (GHI) prediction models which use numerical weather prediction (NWP) data can usually be improved by averaging the GHI forecasts from NWP grid points surrounding the target site (Larson et al., 2016).

In total, adding 72 forecasted weather values (eight values from each of nine cells including and surrounding Griffin) reduced the 24-hour solar radiation prediction mean absolute error rate from 29.16 watts/m^2 to 27.43 as compared to using historical collected data alone, a 5.9% improvement. After these results were found applying Random Forest to extended weather forecasts, additional machine learning techniques were then applied to determine if Random Forests still produced the best results. After analysis of all techniques detailed in *Table 2.1*, there were no significant changes in ranking.

After these prediction improvements were noted for the city of Griffin, four additional cities throughout Georgia were chosen for further analysis: Jonesboro, Attapulgus, Blairsville, and Brunswick. Jonesboro sits just 20 miles north of Griffin and was used as a validation city for its results, and the other three cities were chosen to represent a broad range of weather conditions possible throughout the state of Georgia. *Figure 2.3* shows their locations.

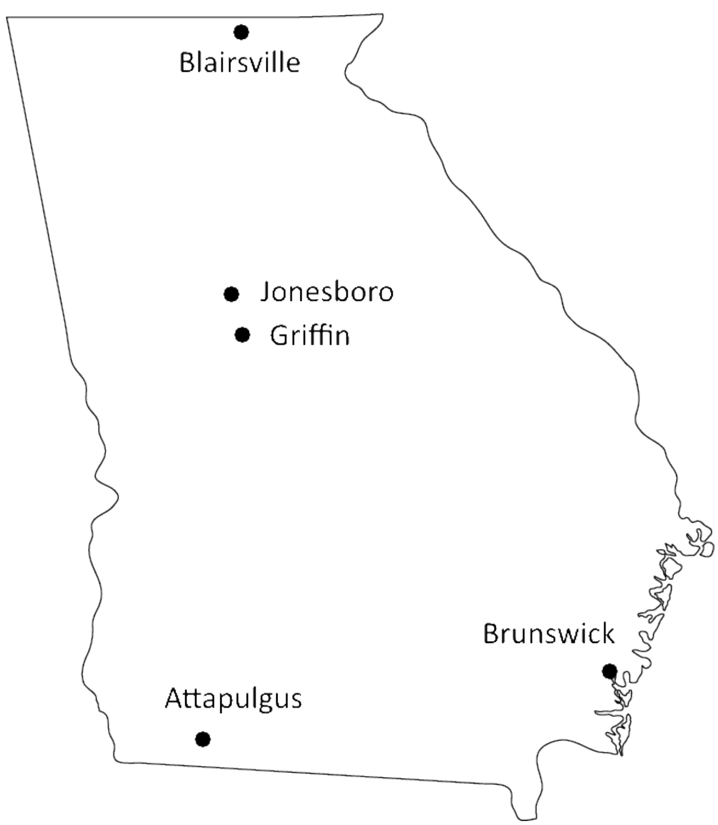

Figure 2.3. A map of the five Georgia cities chosen for solar radiation prediction.

Table 2.2. Properties of five selected Georgia cities based on GAEMN data (2003-2013).

	Griffin	Jonesboro	Attapulgus	Blairsville	Brunswick
Average solar radiation (watts/m^2)	190.4	172.4	185.1	180.7	199.1
Standard deviation of solar radiation (watts/m^2)	277.1	259.9	267.0	273.9	281.6
Average solar radiation, 8:00 a.m. - 8:00 p.m. (watts/m^2)	363.1	329.1	353.1	345.6	373.8
Average air temperature (°F)	62.1	61.6	65.6	56.1	67.3
Average air temperature, 8:00 a.m. - 8:00 p.m. (°F)	66.5	66.7	71.0	61.4	71.3
Average humidity (%)	68.8	72.8	75.3	75.3	74.9
Average wind speed (mph)	4.4	3.1	3.7	3.4	3.6

Average statistics for these five different cities based on recorded GAEMN data were calculated. Their results are shown in *Table 2.2*, and a few observations may be made here. Jonesboro was specifically chosen as the closest and purportedly most identical site to the case study city of Griffin; it lays just 21 miles northwest of Griffin and sits at a nearly identical elevation (919 feet to Griffin's 978 feet). However, its average solar radiation was the actually the furthest from Griffin's average out of these four additional cities. This could be caused by several different factors, including the type of sensors and arrays used in the weather stations, their cleaning and maintenance schedule, or suboptimal positioning relative to the sun's path.

Brunswick, sitting on the southeastern edge of the state on the coast of the Atlantic ocean, recorded the highest average solar radiation and air temperature. Blairsville, in the northeastern Georgian foothills at an elevation of 1,883 feet, recorded the lowest values outside of Jonesboro, and Attapulgus fell somewhere in the middle.

Figure 2.4 charts the average solar radiation readings by year for each city, and lends credence to the theory that cleaning and maintenance differences can result in discrepancies between readings. The readings for each of the five cities were very close to each other in 2003, the earliest year for which data was available. As time goes on, the readings diverge from one another, and Jonesboro in particular appears to have a decidedly negative slope.

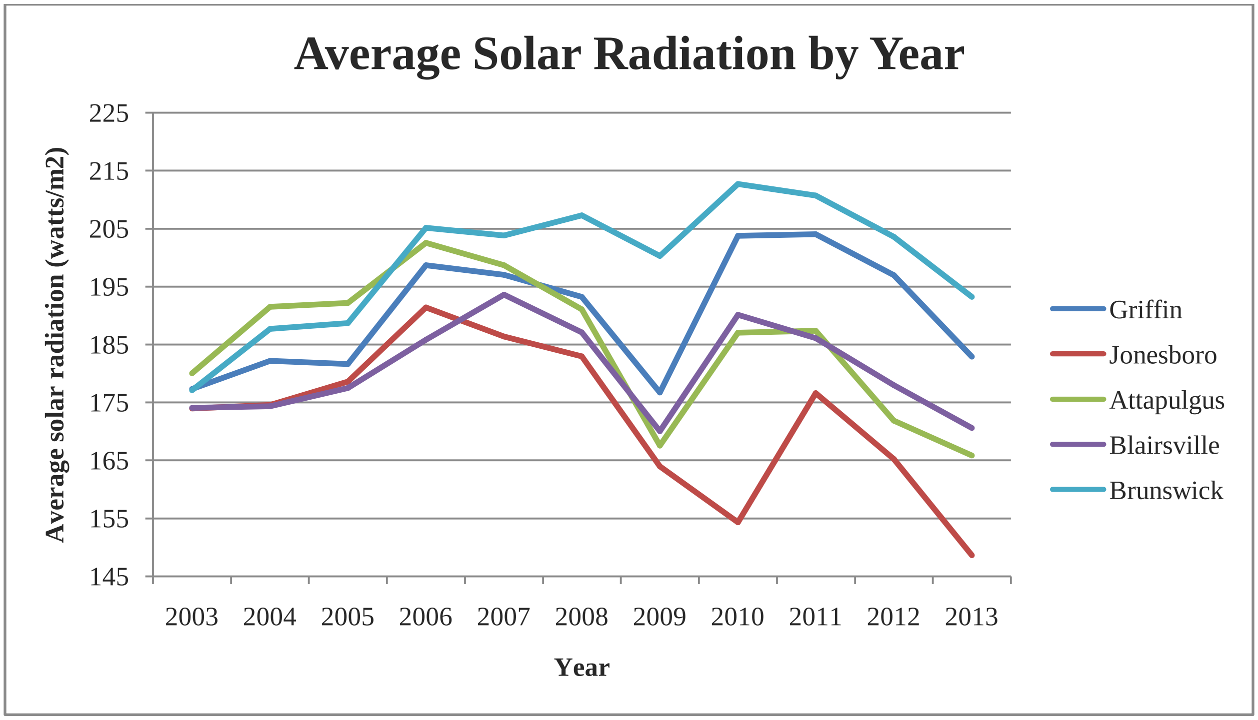

Figure 2.4. Average solar radiation by year for five selected cities in Georgia (2003-2013).

A final note on sensor differences can be made by examining average readings at nighttime across sites. Between sunrise and sunset, observed solar radiation should be zero or very near zero. *Table 2.3* displays the average recorded solar radiation at midnight for each of the five cities; note that Brunswick did not record a value less than 3.1 watts/m^2 in the entire dataset. While this is not an issue for this paper which trains and applies models on a city-by-city basis, care must be taken if models were to be trained on one city and then applied to another. The outputs would need to be calibrated for the "nighttime bias" of each site's solar radiation sensor to avoid errors that could account for a nontrivial proportion of the overall MAE for the model during nighttime hours.

Table 2.3. Average recorded solar radiation at midnight for five Georgia cities (2003-2013).

City	Average solar radiation at midnight (watts/m^2)
Griffin	2.01
Jonesboro	2.61
Attapulgus	1.93
Blairsville	0.66
Brunswick	6.28

One-hour solar radiation predictions for each of these cities were then generated using the same methods as for Griffin. GAEMN historical data and GRIB weather forecast data were collected, and GRIB interpolations were created to create synthetic intra-hour forecasts. Analysis was then run using several different sets of model inputs to discern whether future weather forecasts from the city's GRIB cell and surrounding cells reduced prediction error in these cities as it did in Griffin; these results are presented in *Table 2.4*.

Table 2.4. A comparison of error rates for different model inputs to Random Forest when predicting one-hour solar radiation for five Georgia cities.

	Griffin	Jonesboro	Attapulgus	Blairsville	Brunswick	Average
Persistence model baseline (MAE, watts/m^2)	52.54	45.53	48.45	50.94	51.97	49.89
GAEMN weather observations + 24 hour sliding window (MAE, watts/m^2)	29.16	27.63	30.17	31.44	32.00	30.08
GAEMN weather + window + single GRIB cell forecasts (MAE, watts/m^2)	28.40	26.73	29.26	30.39	31.04	29.16
Improvement when incorporating GRIB cell forecasts (%)[7]	2.6%	3.3%	3.0%	3.4%	3.0%	3.1%
GAEMN weather + window + GRIB cell and eight	27.43	25.68	28.07	28.45	29.32	27.79

[7] Over using GAEMN weather observations + 24 hour sliding window alone

surrounding cell forecasts (MAE, watts/m^2)						
Additional improvement when incorporating surrounding cell forecasts (%)[8]	3.4%	3.9%	4.1%	6.4%	5.5%	4.6%
Total improvement when incorporating GRIB cell and surrounding cell forecasts (%)[9]	5.9%	7.1%	7.0%	9.5%	8.4%	7.6%

The results from these additional four cities underscore the importance of including weather forecast variables even for prediction windows as short as one hour. Over the five cities studied, incorporating single-cell weather forecasted variables reduced MAE by an average of 3.1%, and including forecasts from surrounding cells brought the improvement up to 7.6% over using historical observations alone. Additionally, although this paper is focused on quantifying relative improvements and not on raw accuracies, the MAEs reported here appear to be relatively stable across cities.

2.7 RESULTS: 24-HOUR PREDICTIONS

The difficulty with attempting 24-hour predictions as opposed to one-hour predictions is that relatively slight weather changes are probable over a 60-minute period from the current environment, but much could happen over the next 24 hours, such as a storm front moving in. In this case, no historical data collected could predict the effect of a weather condition that has yet to manifest itself. Therefore, using forecasted weather data is imperative to see what is in store over the next 24 hours.

[8] Over using GAEMN weather observations + 24 hour sliding window + single-cell weather forecasts
[9] Over using GAEMN weather observations + 24 hour sliding window alone

In order to make 24-hour predictions using weather forecasts, the North American Mesoscale (NAM) Forecast System[10] dataset was chosen. This system splits the continental United States into 40-km by 40-km cells and provides weather forecasts for hundreds of variables every 12 hours for each three hours into the future, out to 60 hours. This NAM data has been generated on a consistent basis since June 2003 but only a subset of the data is available to the public. An online data request form is available but due to size constraints the training/testing period was restricted from June 2003 to May 2005. Approximately 4% of the NAM forecast files for this time frame were missing, and these periods were not included in any predictions.

The NAM forecasts dataset includes similar weather variables as the RAP dataset, but there are some differences. One important change is the inclusion of a cloud cover percentage variable; this is an important addition because it is estimated that 90% of the short-term variance in solar radiation is due to changes in cloud cover (Zack, 2014). In total, eight forecasted variables were extracted from the NAM forecasts: air temperature, cloud cover, precipitation probability, wind speed, wind direction, maximum temperature, minimum temperature, and dew point temperature.

Because NAM data forecasts are generated in three-hour intervals, in order to make forecasts every 15 minutes linear interpolation was used to generate synthetic forecasts for intra-frequency instances. While this technique was also used with RAP data for the one-hour weather forecasts, it must be noted that 15-minute linear interpolations between three-hour intervals are less accurate than those interpolating between one-hour intervals, and this may be a significant source of noise. Incorporating seasonality and non-linear models to create more accurate synthetic forecasts could be the focus of further research. Performing these interpolations

[10] More information available at https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-mesoscale-forecast-system-nam

resulted in 67,720 instances corresponding to 15-minute intervals over 720 different days in the time frame June 2, 2003 to May 25, 2005.

Naive persistence models were first applied in order to develop a baseline of performance for 24-hour predictions. Using the current solar radiation to predict the solar radiation 24 hours into the future achieved an MAE of 68.51 watts/m^2, and using the average solar radiation for the desired day of year and time of day achieved an MAE of 62.53. Using the average of these two numbers as the prediction value yields an MAE of 59.80, which became the baseline for future testing for the city of Griffin.

As with one-hour predictions, the 18 weather observations collected from GAEMN, along with time of day and day of year values, were fed into multiple machine learning techniques in Weka in order to predict the solar radiation 24 hours in advance. (All methods used default Weka settings unless stated otherwise.) *Table 2.5* illustrates that decision-tree based models were the strongest methods in predicting 24-hour solar radiation, a similar finding to the one-hour results shown in *Table 2.1*. Use of Random Forests resulted in the lowest overall MAE, and all further 24-hour results reported will be from the application of this method unless noted otherwise. For 24-hour predictions, analysis with a Random Forest model resulted in a 10-fold cross-validation MAE of 44.11 watts/m^2 with a correlation coefficient of 0.944.

Table 2.5. A comparison of machine learning methods for predicting 24-hour solar radiation using historical GAEMN weather observations as model inputs.

Machine Learning Method	MAE (watts/m^2)	Time to run (seconds)[11]
Linear Regression	82.59	19
Multilayer Perceptron (Neural Net)[12]	90.37	11,616
M5P model tree, unpruned	48.12	120
M5P model tree, pruned	54.07	121
Random Tree	58.61	17
REP Tree, unpruned	52.15	28
REP Tree, pruned	54.14	22
Random Forest	44.11[13]	646
Alternating Model Trees[14]	62.46	55
Additive Regression using REP Tree, pruned	51.92	418
Additive Regression using Alternating Model Trees	61.33	1,232
Bagging M5P, unpruned	47.09	988
Bagging REP Tree, pruned	48.50	165

Sliding windows were then added to include the past 24 hourly observations of solar radiation in the model. While this tactic improved the one-hour predictions slightly, it resulted in worse 24-hour predictions; MAE rose 4.3% from 44.11 to 45.99. This result suggests that past solar radiation readings bear very little predictive power to solar radiation 24 hours in advance when added to current climate conditions.

The next iteration of analysis incorporated NAM weather forecasted variables into the model inputs. The addition of these eight variables lowered the MAE from 45.99 when using GAEMN data and 24 historical windows to 35.40, a 23.0% improvement. Finally, eight forecasted values were included from each of nine cells in a three-by-three square of cells, with Griffin in the center, bringing the total number of model inputs to 114. These additional inputs

[11] Time to run 10-fold cross-validation in Weka on Windows 7 with AMD-FX 8350 eight-core processor and 32GB memory.

[12] 20% validation set size, maximum 50,000 epochs, 57 hidden nodes

[13] Compare to 50.61 watts/m2 when using Random Forest to predict 24-hour solar radiation for all instances in the total dataset from 2003-2013. This appears to be evidence of a particularly nonvolatile training set, and it is not unreasonable to assume that inflating all 24-hour error rates reported in this paper by approximately 15% would be a more accurate assessment of raw prediction accuracy over a broader timeframe.

[14] With "build decision tree" set to True

decreased the MAE from 35.40 to 28.70, a further 18.9% decrease from using data from the single Griffin cell alone. In total, adding 72 forecasted weather values (eight values from each of nine cells including and surrounding Griffin) reduced the 24-hour solar radiation prediction mean absolute error rate from 45.99 watts/m^2 when using historical collected data alone to 28.70, a 37.6% improvement.

After these results were found applying Random Forest to extended weather forecasts, additional machine learning techniques were then applied to determine if Random Forests still produced the best results. After analysis of all techniques detailed in Table IV, there were no significant changes in ranking.

As with one-hour predictions, data from Jonesboro, Attapulgus, Blairsville, and Brunswick were next analyzed to quantify the ramifications of incorporating weather forecasts into their 24 hour solar radiation predictions. *Table 2.6* shows the results of using various sets of model inputs into a Random Forest model. As compared to one-hour prediction results, these 24-hour results appear to provide evidence for the variability of the weather in the five different cities. Blairsville sits in the northeastern Georgia mountains and generates one of the lowest average solar radiation readings, but has the highest 24-hour prediction error rates here when examining the persistence models and historical data observations alone. Brunswick, a more stable city on the coast with the highest average solar radiation, recorded some of the lowest error rates in these models. However, Blairsville was able to notch the highest gain in accuracy when incorporating single-cell and multi-cell weather forecasts, suggesting that higher-variability cities benefit disproportionally more from weather forecasts when attempting longer-term predictions than lower-variability cities (this phenomenon did not appear to manifest itself in the one-hour prediction results).

Table 2.6. A comparison of error rates for different model inputs to Random Forest when predicting 24 hour solar radiation for five Georgia cities.

	Griffin	Jonesboro	Attapulgus	Blairsville	Brunswick	Average
Persistence model baseline (MAE, watts/m^2)	59.80	61.06	62.08	68.06	56.98	61.60
GAEMN weather observations + 24 hour sliding window (MAE, watts/m^2)	45.99	49.85	51.59	52.72	48.14	49.66
GAEMN weather + window + single GRIB cell forecasts (MAE, watts/m^2)	35.40	36.93	40.50	38.11	36.59	37.51
Improvement when incorporating GRIB cell forecasts (%)[15]	23.0%	25.9%	21.5%	27.7%	24.0%	24.4%
GAEMN weather + window + GRIB cell and eight surrounding cell forecasts (MAE, watts/m^2)	28.70	29.05	32.89	29.53	28.14	29.66
Additional improvement when incorporating surrounding cell forecasts (%)[16]	18.9%	21.3%	18.8%	22.5%	23.1%	20.9%
Total improvement when incorporating GRIB cell and surrounding cell forecasts (%)[17]	37.6%	41.7%	36.2%	44.0%	41.5%	40.2%

This analysis indicates that incorporating weather forecasts greatly enhances accuracy for 24-hour predictions, even more so than for one-hour predictions. Adding single-cell weather forecasted variables to the model inputs reduced MAE by an average of 24.4%, and including

[15] Over using GAEMN weather observations + 24 hour sliding window alone
[16] Over using GAEMN weather observations + 24 hour sliding window + single-cell weather forecasts
[17] Over using GAEMN weather observations + 24 hour sliding window alone

forecasts from surrounding cells brought the improvement up to 40.2% over using historical observations alone. In addition, as with one-hour predictions, the MAEs achieved when using weather forecasts are appear to be tightly clustered; geographical disparities notwithstanding, results from four of the cities lay within 1.52 watt/m^2 of each other.

2.8 CONCLUSIONS

The improvement in accuracy in both one-hour and 24-hour predictions shown here indicates that predictive data such as future weather forecasts or real-time satellite imagery is an important component of solar radiation prediction over short- and medium-term prediction timeframes. Large-scale collection and dissemination of this data is likely to enhance current prediction models and improve the operations and resource allocation of current and future solar power plants.

There are several clear paths forward to extending the analysis detailed in this paper. The first is made possible by improvements in NOAA's weather forecasting methodology. While the 24-hour forecasts for 2003-2005 made use of NAM data generated every 12 hours at a 40 kilometer grid size, current NAM data generates forecasts every six hours and reduces the grid size to 12 kilometers, providing both more frequent and more granular forecasts. This increased data reporting will no doubt reduce prediction error rates by virtue of more rapid updating of forecasts in addition to increased accuracy as a result of the smaller geographical cell size.

A second approach would be to use NOAA weather forecasts to directly predict NOAA weather variables, such as air temperature and precipitation probability, in effect using their published data to beat their own forecasts. The rationale for expected success here can be seen by comparing solar radiation forecasts using the single-cell weather forecasts to those using the three-by-three square of cells surrounding the weather station. The improvement in accuracy when using multi-cell forecasts suggests that there is important weather data contained in these surrounding cells which isn't reflected in the single cell forecasts alone. It is a reasonable assumption to make that because these surrounding weather forecasts improve predictions of solar radiation at a particular location, they may also be used to improve forecasts for more direct weather variables such as air temperature at that location.

A final obvious avenue of extension would analyze a broader range of climates and macroclimates. The variations in weather throughout the state of Georgia are modest in comparison to the spectrum of temperature, wind, and rainfall present throughout the world at any given time. It would be interesting to see if the improvements in accuracy noted here with the inclusion of weather forecasts continue to hold when moving into more variable - and less predictable - weather systems.

2.9 REFERENCES

Cheung, W., Zhang, J., Florita, A., Hodge, B. M., Lu, S., Hamann, H. F., ... & Lehman, B. (2015). Ensemble Solar Forecasting Statistical Quantification and Sensitivity Analysis (No. NREL/CP-5D00-67191). NREL (National Renewable Energy Laboratory (NREL), Golden, CO (United States).

Feudo, T. L., Avolio, E., Gullì, D., Federico, S., Calidonna, C. R., & Sempreviva, A. M. (2015). Comparison of Hourly Solar Radiation from a Ground–Based Station, Remote Sensing and Weather Forecast Models at a Coastal Site of South Italy (Lamezia Terme). Energy Procedia, 76, 148-155. DOI: 10.1016/j.egypro.2015.07.884.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reuteman, P., & Witten, I. (2009). The WEKA Data Mining Software: An Update. SIGKDD Explorations, vol. 11 (1).

Hamilton, C. (2016). Hourly solar radiation forecasting through neural networks and model trees. (Unpublished master's thesis). University of Georgia, Athens, Georgia.

Hoogenboom, G., Verma, B.P. & Threadgill, E.D. (1990). The development of the Georgia Automated Environmental Monitoring Network. *Proceedings of the 1991 Georgia Water Resources Conference*, held March 19 and 20, 1991, at The University of Georgia.

International Energy Agency. (2014). *Technology roadmap: Solar photovoltaic energy*. Retrieved from http://www.iea.org/publications/freepublications/publication/TechnologyRoadmapSolarPhotovoltaicEnergy_2014edition.pdf.

Ipsos. (2011). *Global citizen reaction to the Fukushima nuclear plant disaster*. Retrieved from https://www.ipsos-mori.com/Assets/Docs/Polls/ipsos-global-advisor-nuclear-power-june-2011.pdf.

Lara-Fanego, V., Ruiz-Arias, J.A., Pozo-Vazquez, D., Santos-Alamillos, F.J., & Tovar-Pescador, J. (2012). Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain). Sol. Energy 86 (8), 2200–2217.

Larson, D. P., Nonnenmacher, L., & Coimbra, C. F. (2016). Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest. Renewable Energy, 91, 11-20.

Linares-Rodriguez, A., Ruiz-Arias, J.A., Pozo-Vazquez, D., & Tovar-Pescador, J. (2013). An artificial neural network ensemble model for estimating global solar radiation from Meteosat satellite images. Energy 61, 636–645.

Miller, S.D., Heidinger, A.K., & Sengupta, M., 2013. Physically based satellite methods. In: Kleissl, J. (Ed.), Solar Energy and Resource Assessment. Elsevier, Oxford, pp. 49–79.

Paz, J. O., & Hoogenboom, G. (2008). Climate and weather information for Georgia farmers. Cooperative Extension, the University of Georgia.

Pedro, H. T., & Coimbra, C. F. (2012). Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, *86*(7), 2017-2028.

Perez, R., Lorenz, E., Pelland, S., Beauharnois, M., Van Knowe, G., et al. (2013). Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. Sol. Energy 94, 305–326.

Ruiz-Arias, J. A., Quesada-Ruiz, S., Fernández, E. F., & Gueymard, C. A. (2015). Optimal combination of gridded and ground-observed solar radiation data for regional solar resource assessment. Solar Energy, 112, 411-424.

Sfetsos, A., & Coonick, A. H. (2000). Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. Solar Energy,68(2), 169-178

Spokas, K. & F. Forcella. (2006.) Estimating hourly incoming solar radiation from limited meteorological data. Weed Science 54:182–189.

University of Georgia Office of Sustainability. (2017). *Renewable Energy*. Retrieved from http://sustainability.uga.edu/what-were-doing/renewable-energy/.

World Meteorological Organization. (2003). *Introduction to GRIB Edition 1 and GRIB Edition 2*. Retrieved from https://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/Introduction_GRIB1-GRIB2.pdf.

Zack, J. (2014). Solar power production forecasting: overview of methods and input data needs. *ERCOT ETWG Meeting*, April 30, 2014, Austin, Texas.

Zervas, P.L., Sarimveis, H., Palyvos, J. A. & Markatos, N. C. G. (2008). Prediction of daily global solar irradiance on horizontal surfaces based on neural-network techniques. Renewable Energy, vol. 33, pp. 1796-1803.

CHAPTER 3

A COMPARISON OF MACHINE LEARNING TECHNIQUES FOR POSTPROCESSING WEATHER FORECASTS[18]

[18] Sanders, W., Barrick, C., Rasheed, K., and Maier, F. To be submitted to *International Journal of Energy Research*.

3.1 ABSTRACT

Machine learning models were developed and compared to increase the accuracy of 24-hour forecasts of eight weather variables (air temperature, cloud cover, visibility, wind speed, wind direction, dew point temperature, air pressure, and humidity) for five cities in Georgia for the time period June 10, 2016 to June 10, 2017. Sinusoidal representations of the hour-of-day and day-of-year being forecasted were found to significantly decrease prediction errors when included as model inputs in order to extract cyclical patterns. The industry-standard technique of linear regression Model Output Statistics (MOS) was analyzed as a baseline performance measure, and an ensemble model stacking a Random Forest with an artificial neural network (ANN) was found to reduce prediction error over MOS on seven of the eight weather variables studied. The inclusion of additional forecasted weather variables from areas immediately surrounding the target location was not found to have an impact on prediction error.

3.2 INTRODUCTION

Short-range weather forecasting is of vital importance to a wide range of consumers, including pilots, drivers, local and federal emergency services, recreation facilities, the construction industry, and the military (National Center for Atmospheric Research, 2017). With such a large impact surface, even small improvements in prediction accuracy can translate into noticeable economic impacts on these sectors. One such example where accurate forecasts have led to direct financial benefits is in predicting demand for electricity; it is estimated that U.S. electric utilities save $150 million annually by using 24-hour air temperature forecasts to estimate load demand (American Meteorological Society, 2015).

This paper's goal was to compare machine learning techniques to detect and correct for systematic biases in 24-hour forecasts generated and disseminated by The National Oceanic and Atmospheric Administration (NOAA), an approach referred to as postprocessing. In addition, an attempt was made to incorporate forecasted variables from areas surrounding the target location in order to improve forecast accuracy.

Weather forecasting is inherently difficult. The chaotic nature of weather leads to a considerable sensitivity on initial conditions, and the path to accurate forecasts is limited as much by the measurement error of initial conditions as it is by model error (Stensrud et al., 1999). However, the influence that weather has on our lives cannot be overstated. Up to 90% of the emergencies declared by the Federal Emergency Management Agency (FEMA) and 7,000 road fatalities annually can be attributed to weather, and weather-related air traffic delays cost Americans about $6 billion per year (American Meteorological Society, 2015). For these reasons, enormous amounts of manpower and computing power have been allocated to weather prediction over the past few decades, and this focus has paid off. The British Meteorological Service (2017) claims that their current four day forecasts are as accurate now as their one day forecasts were 30 years ago. This statement mirrors the estimate that forecast horizons have been steadily increasing at approximately one day per decade (Nurmi et al., 2012). These improvements in accuracy, coupled with the large number of individuals, businesses, and governments who rely on weather forecasts every day, Bauer, Thorpe, and Brunet (2015) claim that the impact of weather predictions at high as that of any other physical science.

Numerical Weather Prediction (NWP) data are a common form of computer-generated weather forecasts created by feeding current weather observations into mathematical models to develop estimates for the future state of the weather. These models generally represent a threedimensional projection of various weather elements for a specified time in the future. NOAA alone generates eight different NWP models on a continuous basis which are widely used both on a standalone basis and as part of ensemble models by both the National Weather Service and third-party weather forecast vendors such as Weather Underground and AccuWeather.

3.3 POSTPROCESSING

While vast amounts of research and processing power go into generating these NWP forecasts (Lin, Atlas, and Yeh, 2004), there have been a number of attempts at improving their accuracy further by finding and detecting various time- or location-based biases which may exist by postprocessing the data. Postprocessing is commonly used to refine near-surface forecasts, as forecasts of NWP models are generated at a relatively coarse horizontal resolution and may only crudely approximate the specific physical processes that occur at or near ground level. In addition, NWP models represent the world simply as an array of gridpoints, which simplifies and homogenizes surface conditions while possibly ignoring the small-scale effects that topological features or small bodies of water may have on hyper-local forecasts (Wilks, 2011).

One of the most elementary forms of correction is simply to add (subtract) a forecast's mean historical error to the current forecast value in order perform a bias correction if a forecast is consistently too low (high). Generally, however, more advanced techniques are applied. Model Output Statistics (MOS) is a multiple linear regression postprocessing technique which involves correlating local and regional weather observations with numerical model outputs in an attempt to reduce bias and increase accuracy by factoring in local climatology (Klein & Glahn, 1974). At its essence, MOS assumes consistent error patterns and attempts to correct current weather forecasts by analyzing errors from past forecasts (Mathiesen & Kleissl, 2011), and has historically been the preferred method for NWP postprocessing (Wilks, 2011).

MOS models for widely-recorded weather variables such as temperature and wind speed are commonly generated for each individual weather station in a region (Rudack & Ghirardelli, 2010). The National Weather Service (NWS) has long applied MOS statistical techniques to forecasts generated by the National Centers for Environmental Prediction, resulting in more accurate guidance by using an ensemble technique of melding multiple models followed by statistical adjustments to correct for model biases and to properly weight each model's contribution (Dallavalle & Erickson, 1999). Such methods can even be used to infer forecasts for additional weather variables which weren't explicitly predicted in the initial dataset.

One drawback for these postprocessing models is the necessity of including sufficient data in the training set. Wilks and Hamill (2007) found "significant" improvement in forecast skill when making use of 15- and 25-year training sets for MOS corrections as compared to using shorter periods of 1 to 5 years. If the underlying NWP models are updated on a more frequent basis in an attempt to remove bias or otherwise provide more accurate forecasts, the MOS model would no longer be applicable and would need to be retrained using the forecasts from the new NWP model. An additional drawback specific to MOS and other linear regression models is the necessity of creating multiple models to account for different time-of-day or seasonal biases. For example, a bulletin from the National Weather Service Office of Meteorology (1995) on an MOS approach to predicting cloud ceiling heights refers to a "cool season" model and a "warm season" model, in effect creating a model tree.

In addition to numerous academic studies, postprocessing is widely used in practice. A NOAA whitepaper addressing and encouraging postprocessing reports that many National

Weather Service (NWS) organizations have substantially improved their services via judicious use of postprocessing to minimize noise and correct for biases (Hamill et al., 2015). The authors note that it is inevitable that raw guidance will necessarily include systematic errors, and that leveraging historical biases along with incorporating additional modeling systems (if available) have substantially improved the quality of forecasts.

Dennstaed (2006) found that small airports commonly add their own regional expertise to numerical weather forecasts by accounting for historical observations to correct for possible model biases. In addition, many local weather forecasters do not merely reiterate forecasts from the National Weather Service directly but instead apply their own experience and regional knowledge to manually postprocess the data they receive from the NWS. Samenow and Fritz (2015) reference a TV meteorologist in Washington who, each morning, tracks the models' forecasts for different variables and layers of the atmosphere to find and correct for potential biases. They note that this is not an isolated case but that many local meteorologists have impressive academic credentials and take pride in applying their regional knowledge to craft their own forecasts.

3.4 LITERATURE REVIEW

While the use of standard multiple linear regression Model Output Statistics postprocessing is still widespread, in recent years more advanced machine learning techniques have been applied to weather datasets in order to evaluate their efficacy. Guarnieri, Pereira, and Chou (2006) analyzed day-ahead solar radiation forecasts collected from the Brazilian Center of Weather Forecast and Climate Studies (CPTEC) for several weather stations in southern Brazil. They fed 36 weather-related predictor variables into an two artificial neural networks (ANNs) with different structures in an attempt to improve the Root Mean Squared Error (RMSE) of the raw solar radiation forecasts, and found that each ANN eliminated the systematic bias observed from the raw values and dropped the RMSE by over 30%. As they did not perform a comparison to results obtained by more traditional techniques such as linear regression, it is not known if neural networks outperformed these methods; however, this result is evidence that advanced machine learning techniques do show promise in the domain of weather post-processing.

Similarly, Lauret, Diagne, and David (2014) performed analysis on historical NWP global horizontal solar irradiance (GHI) day-ahead predictions collected over one full year from La Reunion Island, a French territory in the Indian Ocean. They found mean bias errors related to cloud cover forecasts and the angle of the sun for the forecast period. Specifically, for forecasted clear-sky conditions, GHI was overpredicted, while for forecasted cloudy predictions it was underpredicted, suggesting the model over-exaggerates the effect of cloud cover or the absence thereof. They found that the use of an ANN as a bias-correction step reduced the RMSE of GHI predictions by 17.7% as compared to the NWP models alone, and reduced average bias to nearly zero. Additional attempts to improve upon this result by using spatial averaging techniques as the starting point led to comparable results as those given by the bias removal with no spatial averaging. This paper will take a similar approach and attempt to improve accuracy results by including information from the GRIB cells immediately surrounding the target site.

Mathiesen and Kleiss (2011) also found that NAM (North American Mesoscale) models systematically overforecast global horizontal irradiance during clear sky conditions by up to 40%. In addition, clear sky conditions were overpredicted, resulting in many more false clear sky days than false cloudy sky days. These two phenomena resulted in a general hour-ahead overforecasting of GHI by 41.0 - 77.5 watts/m^2 for seven sites across the United States . By applying a more advanced fourth-order multivariate MOS correction based on solar zenith angle and forecast clear sky index, the researchers were able to nearly eliminate mean bias while reducing the RMSE of the forecasts by an average of 10.3%. The authors state that their work may serve as a useful reference point to contrast to other, more advanced, products, but do not provide this comparison.

Galanis et al. (2006) applied Kalman filter postprocessing to NWP model forecasts of air temperature and wind speed in two unnamed locations in southern Europe and noted that the low resolution of the model renders it susceptible to systematic interpolation error. They report finding that Kalman filters with low-order polynomials appear to perform well in predicting air temperature and wind speed from 24 hours to 120 hours ahead. Their results indicate dropping absolute errors approximately 64% in temperature prediction and 6% in wind speed prediction on Case 1, and 46% and 39%, respectively, in Case 2. In both cases the mean bias was dropped to nearly zero, fulfilling the one of the primary goals of postprocessing. It must be noted that Case 1 consisted of just one month of forecasts (December 2003) and Case 2 of just three months (March-May 2003), which are comparatively smaller data sets than most other experiments analyzed.

Marzban (2003) analyzed forecasts from the Advanced Regional Prediction System, a model which was found to exhibit significant temperature biases depending on the forecast hour. To correct these systematic errors, they used customized neural networks on each of 31 weather stations and included other weather variables to achieve a 40% reduction in mean-squared temperature forecast error across all stations as well as reductions in bias and variance for time frames from one hours to 24 hours. For a small number of sites, postprocessing did not appear to materially affect results, indicating that some geographic regions may be more susceptible to systematic biases than others. Additionally, the optimal number of nodes in the hidden layer of the neural network was calculated for each weather station; it was found that most stations required between two and eight hidden nodes, while one didn't require any. Marzban makes the observation that this represents an implicit comparison between a neural network and traditional MOS, as MOS is based on linear regression and can be represented by a neural network with no hidden nodes. The point is made that when the optimal number of hidden nodes is greater than zero - which was the case for all but one weather station analyzed - it is indicative that a properly-tuned neural network will outperform a linear model.

Casaioli et al. (2003) tested different statistical approaches to the post-processing of surface air temperatures in the Italian region of Puglia collected from 1985 to 1996. In their paper they outline and analyze results obtained from no postprocessing, a vertical interpolation, a Kalman filter, a linear neural network (i.e., no hidden layers), and a non-linear neural network. While they found that the non-linear neural network resulted in the lowest mean absolute deviation (or, mean absolute error) of day-ahead temperature forecasts (1.5° C), just beating out the Kalman filter (1.6° C), it was also reported that using no post-processing at all resulted in just a 1.5° C mean average error. As the majority of other analyses of post-processing efficacy find significant improvement over using raw weather forecasts alone, it must be noted that their results do not fall in line with those reported by others.

Marzban, Leyton, and Colman (2007) analyzed cloud ceiling and visibility data collected during 2001 to 2005 over a broad area of the Pacific Northwest to predict the probability of low or high cloud ceilings 6 to 12 hours ahead. They analyze 39 different weather stations and state that for stations with sufficient data, neural networks outperform both MOS and logistic regression, although the results are displayed graphically by station and no summary statistics are given. Neural networks beat MOS on 9/14 weather stations with statistically significant six-hour predictions (64%), and 34/34 on twelve-hour predictions (100%), indicating that weather forecasting biases become more non-linear at longer prediction intervals. As this paper will attempt day-ahead predictions, these results would indicate that non-linear methods may dominate linear methods at this longer time-frame.

The study of related works in this field found two shortfalls which this paper will attempt to address. The first is the prevalence of solar radiation as the predicted target variable of choice, with fewer papers dedicated to applying advanced techniques towards forecasting more well-known variables such as temperature and wind speed. The second, and the primary focus of this paper, is the dearth of research in comparing these advanced techniques to the industry standard multiple linear regression technique, which is computationally fast and easy to interpret. It is widely accepted and published that numerical weather forecasts exhibit biases that can be corrected through various statistical techniques, both old and new, and this paper will attend to compare their results on a direct basis.

3.5 METHODOLOGY

Murphy (1993) lays out three criteria for evaluating the "goodness" of a weather forecast: type 1 (consistency), type 2 (quality), and type 3 (value). It is type 2 which this paper will attempt to refine. Mean absolute error (MAE) has been chosen as the target function to optimize, although it may be noted that many consumers of weather forecasts are nearly as concerned with the range of possible outcomes as they are with the single predicted value, and so reducing the width of the probability distribution of forecasts would be an important area of inquiry for future studies.

In order to develop a baseline for current forecast accuracy, a historical dataset must be sourced which includes both historical weather forecasts as well as the actual ultimate weather observations for the forecasted time period. This paper makes use of weather forecasts provided by The National Oceanic and Atmospheric Administration (NOAA). Their data is disseminated in the GRIB file format, a compact binary format commonly used to store historical and forecasted weather data (World Meteorological Organization, 2003). Each GRIB file describes a particular geographical region for a single date, and internally splits this region into a grid of cells of a consistent size. For each cell, attribute values are listed describing weather attributes in the cell at that time, or, in the case of weather forecasts, at a specified time in the future. These GRIB files can denote either current weather variable observations or predicted weather variables for a specified time in the future.

There are several datasets published regularly by NOAA, and this paper has chosen the North American Mesoscale Forecast System (NAM) GRIB data set for analysis. This dataset was available on NOAA's public FTP servers for download and included both 24-hour forecasts and 0-hour observations for the United States for the time period June 10, 2016 to June 10, 2017. NAM forecasts are partitioned into cells of 12 kilometer by 12 kilometer resolution and are generated every six hours, at 12:00 a.m., 6:00 a.m., 12:00 p.m., and 6:00 p.m. Coordinated Universal Time (UTC). The five cities chosen for analysis in this paper all abide by Eastern Standard Time (EST) and Eastern Daylight Time (EDT), which respectively sit five and four hours behind UTC. This results in the four daily forecasts being generated at eight different possible hours, which will be addressed throughout this paper as follows: 1:00 a.m. or 2:00 a.m. (nighttime), 7:00 a.m. or 8:00 a.m. (morning), and 1:00 p.m. or 2:00 p.m. (midday), and 7:00 p.m. or 8:00 p.m. (evening). For each of the five cities, 24-hour forecasts for their geographic centers were compared to the eventual observations for the aforementioned timeframe in order to analyze the quality of their predictions.

The machine-learning software Weka (Hall et al., 2009) was chosen for training and testing all models generated in this paper. A number of machine learning techniques were run on each of the five cities independently, with 10-fold cross-validation Mean Absolute Error (MAE) used throughout as a means of assessing model accuracy.

Table 3.1. Mean Absolute Error of NOAA 24-hour forecasts for eight forecasted variable for five Georgia cities, June 10, 2016 to June 10, 2017.

	Griffin	Jonesboro	Attapulgus	Blairsville	Brunswick	Average
Air Temperature (°F)	2.50	2.46	2.24	2.26	1.96	2.28
Cloud Cover (%)	24.41	24.63	25.40	23.67	26.35	24.89
Visibility (m)	1,490	1,331	2,065	2,918	1,847	1,930
Wind Speed (m/s)	1.20	1.21	1.09	1.32	1.07	1.18
Wind Direction (°)	38.23	37.85	39.53	40.03	32.42	37.61
Dew Point Temperature (°F)	2.67	2.64	2.74	2.38	2.42	2.57
Air Pressure (Pa)	64.93	63.78	58.94	49.37	62.19	59.84
Relative Humidity (%)	5.95	6.05	5.73	6.49	5.94	6.03

Table 3.1 displays the initial 24 hour MAE for each of eight forecasted weather variables over five Georgia cities, which appear to be similar across all five cities. In order to improve on these forecasts, a two-step approach was taken. The first was an attempt to extract any existing biases - patterns in which the predicted temperature systematically overforecasted or underforecasted the actual observed variable - by creating machine learning models with time-of-day and day-of-year as inputs in addition to the NAM forecasts for air temperature and other variables. The second approach made additional use of forecasted values for the immediate areas ("cells", in GRIB parlance) surrounding the target area. Sanders (2017) found that including temperature, humidity, and other weather variables from GRIB cells surrounding the target area decreased the error in 24-hour solar radiation prediction by 20.9% as opposed to using forecasted weather variables from the target cell alone. He hypothesized that improvement in accuracy

when using multi-cell forecasts suggests that there is important weather data contained in these surrounding cells which isn't reflected in the single cell forecasts alone. It is a reasonable assumption to make that because these surrounding weather forecasts improve predictions of solar radiation at a particular location, they may also be used to improve forecasts for more direct weather variables such as air temperature at that location.

3.6 RESULTS: AIR TEMPERATURE

Air temperature, one of the most commonly referenced weather variables in everyday life, was the first variable chosen for improvement. The initial MAE of 2.28° Fahrenheit from the NOAA forecasts aligns very closely with results reported by Samenow and Fritz (2015), who found that one-day temperature forecasts are typically accurate to within 2.0° to 2.5° Fahrenheit. As mentioned before, the NAM dataset generates four 24-hour forecasts, at nighttime, morning, midday, and evening. *Table 3.2* shows the average bias of the air temperature data when broken down by time of day.

Table 3.2. Average 24-hour air temperature forecast bias by time of day.

Time of Day	Average Prediction	Average Observation	Average bias (predicted-observed)	Standard deviation of bias	Number of Observations	P-Value of bias
All	66.36	66.15	0.21	2.98	5615	<0.001
Nighttime	63.74	61.97	1.77	2.69	1375	<0.001
Morning	62.00	61.14	0.86	2.27	1395	<0.001
Midday	70.06	71.84	-1.78	2.92	1380	<0.001
Evening	69.50	69.48	0.02	2.83	1465	0.394

These results indicate that the coolest times (nighttime and morning) tend to be overforecasted, while the warmest time (midday) tends to be underforecasted. Given that the extremes led to the largest biases on an intraday level, the next analysis performed was to determine if any seasonal biases existed in warmer months versus cooler months.

Table 3.3. Average 24-hour air temperature forecast bias by month.

Month	Average Prediction	Average Observation	Average bias (predicted-observed)	Standard deviation of bias	Number of Observations	P-Value of bias
January	51.77	52.17	-0.40	2.88	520	<0.001
February	56.71	57.31	-0.60	3.04	465	<0.001
March	54.70	55.56	-0.86	2.79	435	<0.001
April	68.42	68.29	0.13	3.87	535	0.218
May	71.67	71.42	0.25	3.23	620	0.027
June	77.42	77.08	0.33	3.02	470	0.009
July	81.33	79.70	1.63	2.92	355	<0.001
August	79.98	78.93	1.06	2.38	515	<0.001
September	77.25	76.12	1.14	2.47	415	<0.001
October	67.81	67.52	0.28	2.33	495	0.004
November	59.28	58.87	0.42	2.43	270	0.002
December	50.81	51.13	-0.32	2.72	520	0.003

Table 3.3 shows that a bias is apparent in this analysis which has the effect of counteracting the previous time-based bias. The warmer months are overforecasted, while the cooler months are underforecasted. The Pearson correlation coefficient of a month's average observed temperature and its historical temperature forecast bias (predicted temperature less observed temperature) is 0.845.

One explanation for these presence of these biases is that they are random artifacts of this particular dataset which would not persist when analyzing other cities or time frames in out-of-sample data. However, a look at the p-values for each row shows that most of these results are highly statistically significant, and such results would be improbable for data without biases. As a counter-argument, these p-values may somewhat overstate statistical significance because they most likely violate the assumption of independency - if the temperature for one city in Georgia is overforecasted, it is likely that temperatures for the other cities are as well. To address this concern, intraday timeframes from just one city (Griffin, Georgia) were analyzed in order to assure independency in the forecasted instances. *Table 3.4* displays these single-city findings, which mirror the ones obtained over all cities in maintaining statistical significance.

Table 3.4. Average 24-hour air temperature forecast bias by time of day for Griffin, Georgia.

	Average Prediction	Average Observation	Average bias (predicted-observed)	Standard deviation of bias	Number of Observations	P-Value of bias
All	66.31	65.74	0.57	3.19	1123	<0.001
Nighttime	63.56	61.21	2.35	3.03	275	<0.001
Morning	61.22	60.15	1.06	2.23	279	<0.001
Midday	70.15	71.50	-1.35	3.01	276	<0.001
Evening	70.13	69.90	0.23	3.23	293	0.111

The appearance of these biases explains the popularity of postprocessing raw forecasts, as detailed earlier. One explanation of the continued existence of these biases is that regional patterns may exist which cancel each other out when analyzing nationwide data as a whole, thus making detection difficult unless each region is analyzed independently. There is some precedence for regional biases existing in weather forecasting. The National Weather Service (2017) outlines a few known regional flaws in their various models, such as the Global Forecast System (GFS) overestimating the strength of weather systems crossing the Sierra mountains in the Southwestern United States.

In order to test the effectiveness of various post-processing techniques, models were run using eight NAM predicted weather variables (air temperature, cloud cover, visibility, wind speed, wind direction, dew point temperature, air pressure, and humidity) as inputs to predicting the actual observed temperature 24 hours later. In addition, to extract cyclical patterns from the data such as those detailed in *Table 2,* 10 temporal inputs were included into the models: day-of-year, hour-of-day, and the first two harmonics of the sine and cosine of each; this approach was outlined in Klein (1974). All methods were run with default Weka settings unless indicated otherwise.

Table 3.5. A comparison of machine learning methods for postprocessing 24-hour NAM temperature predictions over five cities in Georgia. Results indicate Mean Absolute Error (MAE) in °F.

Technique	Griffin	Jonesboro	Attapulgus	Blairsville	Brunswick	Average
Baseline (no postprocessing)	2.50	2.46	2.24	2.26	1.96	2.28
Model Output Statistics (Linear Regression)	1.97	1.94	1.77	1.72	1.65	1.81
Multilayer Perceptron (Neural Net)	2.19	2.03	1.83	1.85	1.82	1.94
Multilayer Perceptron (Neural Net)*	1.94	1.94	1.65	1.72	1.69	1.79
Support Vector Machine (SVM) for Regression	1.95	1.93	1.77	1.72	1.66	1.81
M5P model tree, unpruned	1.95	2.18	7.87	2.73	2.47	3.44
M5P model tree, pruned	1.96	1.94	1.77	1.73	1.65	1.81
REP Tree, unpruned	2.44	2.46	2.24	2.16	2.15	2.29
REP Tree, pruned	2.34	2.42	2.19	2.16	2.12	2.25
Random Forest	2.00	1.97	1.66	1.82	1.64	1.82
Stacking Random Forest with MLP*†	1.83	1.80	1.55	1.63	1.52	1.67

* 20% validation size, max 50,000 epochs, learning rate 0.05
† Linear regression used as a meta-classifier

The results are shown in *Table 3.5* and proved to be relatively consistent across all five cities. The baseline linear regression model most commonly used in Model Output Statistics postprocessing significantly improved on the original NOAA forecasts, dropping the average MAE 20.6% from 2.28 to 1.81. Stacking Random Forests with a customized neural network produced the best results in all five cities and logged an average MAE of 1.67, 7.7% better than the linear regression model. It may be noted here that one potential advantage to using linear regression is that the output from the model is human-readable, as contrasted with the other black-box techniques applied in this analysis.

For example, below is the air temperature linear regression model for the city of Griffin for this time period:

```
target = -0.714  * cosine_day +
0.2028 * sine_day +
-0.1968 * cosine_twice_day +
-0.1816 * hour +
-2.1299 * cosine_hour +
-3.4201 * sine_hour +
-0.5143 * sine_twice_hour +
0.8483 * predicted_temp +
0      * predicted_visibility +
0.0612 * predicted_dew_point +
-0.0181 * predicted_humidity +
8.7202
```

To further enhance accuracy of these forecasts, the inclusion of air temperature forecasts for the areas surrounding the target location was analyzed. Eight additional inputs were added to the existing model, which represent eight forecasted temperatures from the NAM cells lying to the northwest, north, northeast, east, southeast, south, and southwest of our target cell. As stacking Random Forests with a customized neural network resulted in the best performance in the previous section, this technique was run again with the additional surrounding inputs in order to determine if any performance increase was noted.

Table 3.6. Comparison of a model incorporating forecasts from surrounding GRIB cells. Results indicate Mean Absolute Error (MAE) in °F.

City	Initial model run	Inclusion of temperature from surrounding cells
Griffin	1.83	1.78
Jonesboro	1.80	1.78
Attapulgus	1.55	1.59
Blairsville	1.63	1.62
Brunswick	1.52	1.57
Average	1.67	1.67

As *Table 3.6* indicates, inclusion of these eight forecasted temperatures did not decrease prediction error when averaged across all five cities. It appears that there is not enough additional information available in the surrounding GRIB cells to improve on the original temperature forecasts. This finding mirrors results published by Lauret, Diagne, and David (2014), who applied MOS analysis to prediction of solar radiation for La Reunion Island, a French oversea territory located in the Indian Ocean. They found that including spatial averaging by taking the mean of irradiance over several grids centered around the station did not enhance accuracy further than applying MOS correction alone. Merely including time- and date-based inputs in addition to seven additional forecasted weather variables results in the lowest predictive error for air temperature over this dataset.

At this point, an attempt was made to determine if there was any cross-predictive accuracy among cities in our dataset. If a model was trained on City A, how would it perform when attempting to predict City B? The structure of the data required a slight shift in methodology. Simply training a model over all dates in City A and then applying it to City B would allow information leakage from the training set into the test set for the matching dates and times. To remedy this, all cities were split into two sets based on the day-of-week of the time to be predicted: Sundays through Wednesdays were used for training, and Thursdays through Saturdays for testing. The ensemble model was trained on each city's training set and then applied to every city's testing set to determine how robust the models were to changes in location.

Table 3.7. Analysis of cross-predictive accuracy across cities. Results indicate Mean Absolute Error (MAE) in °F.

		Testing City					
		Griffin	Jonesboro	Attapulgus	Blairsville	Brunswick	Average
Training City	Griffin	1.88	1.90	2.52	2.03	2.81	2.23
	Jonesboro	1.80	1.81	2.79	1.98	3.13	2.30
	Attapulgus	2.10	2.20	1.56	2.58	1.82	2.05
	Blairsville	2.10	2.15	1.99	1.90	1.97	2.02
	Brunswick	2.43	2.56	1.97	3.45	1.70	2.42
	Average	2.06	2.12	2.17	2.39	2.29	

Table 3.7 shows that the five models which were tested on the same cities on which they were trained recorded an average MAE of 1.77, whereas the 20 models tested on different cities than their training recorded an MAE of 2.31, a 30.5% increase. However, further information can be gleaned by examining the geographical disparity of the cities. Jonesboro and Griffin lie just 20 miles apart in the middle of Georgia and experience very similar weather conditions; it is most likely not coincidental that models trained on either of these cities perform very similarly when tested on the other. Further study can be done to determine if cities very close to one another with similar weather would benefit from combining their data to exploit the inherent machine learning advantage of more numerous training instances.

3.7 RESULTS: ADDITIONAL WEATHER VARIABLES

After air temperature was analyzed, the best-performing model was applied to the seven other weather variables chosen for analysis. For each variable, the inputs to the model were the seven other predicted variables in addition to the 10 calculated temporal fields as detailed in an earlier section. *Table 3.8* shows the results of forecasting all eight weather variables as compared to the original NOAA forecasts.

Table 3.8. 24 hour forecast error rates for eight weather variables.

Weather Variable	NOAA Forecasts	Linear Regression (MOC)	MOC Improvement over NOAA (%)	Stacking Model	Stacking Improvement over NOAA (%)	Stacking Improvement over MOS Linear Regression (%)
Air Temperature	2.28	1.81	20.5%	1.67	26.8%	7.7%
Cloud Cover (%)	24.89	26.79	-7.6%	25.51	-2.5%	4.8%
Visibility	1,930	2,865	-48.4%	1,901	1.5%	33.6%
Wind Speed	1.18	0.80	32.3%	0.75	36.4%	6.3%
Wind Direction	37.61	44.59	-18.6%	38.50	-2.4%	13.7%
Dew Point Temperature	2.57	2.06	19.8%	1.87	27.2%	9.2%
Air Pressure	59.84	50.19	16.1%	52.25	12.7%	-4.1%
Relative Humidity	6.03	5.87	2.7%	5.74	4.8%	2.2%

As a result of applying the stacking model, error rates were improved over NOAA raw forecasts on six out of the eight variables studied, and over the industry standard MOS on seven of them.

3.8 CONCLUSION

Postprocessing of weather forecasts is undoubtedly a common means of refining the data and increasing forecast accuracy. Marzban (2003) claims that most current numerical weather forecasting systems employ some sort of statistical postprocessing in order to improve performance. In this paper, simple time-of-day and day-of-year analysis uncovered highly statistically significant biases for air temperature forecasts over five cities in Georgia. Using these two variables as inputs into an ensemble machine learning models yielded a decrease in error of 26.8% over using NOAA's temperature forecasts alone and 7.7% over using the industry-standard Model Output Statistics (MOS) postprocessing technique.

Applying this technique to all eight weather variables chosen for analyses resulted in decreases in error for six of them over using NOAA's forecasts and for seven of them over using MOS alone. There appears to be extractable patterns in this data which NOAA's forecasts do not take into account in their forecasts, and there also appears to be the potential to improve upon the results obtained by standard Model Output Statistics.

There are a few clear steps forward to extending the research presented in this paper. While it was found that the addition of predicted weather from the GRIB cells surrounding the target area did not improve air temperature forecasts, there were also slight improvements in accuracy noted when combining data from geographically similar cities to build a single model. More research could be done to quantify this effect; if it persists, it could prove useful when working with data sets of limited size.

In addition, supplementary data from predicted times other than the target time could be integrated into the model. Klein (1978) notes that numerical weather forecasting models can be systematically slow or fast, and therefore predictors within a short period of the predictand time could potentially be useful as inputs.

Finally, Glahn (2014) noted that using historical data decay factors between 0.025 and 0.1 improved the MAE of 72-hour temperature forecasts over using MOS alone, and posited that the ever-changing nature of NWP models can render post-processing techniques stale if they do not underweight older data. More research could be done when using the more advanced techniques detailed in this paper to quantify the effect of this phenomenon.

3.9 REFERENCES

American Meteorological Society. (2015). Weather Analysis and Forecasting.

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. Nature, 525(7567), 47-55.

British Meteorological Service. (2017). How Accurate are our Public Forecasts? Retrieved from http://www.metoffice.gov.uk/about-us/who/accuracy/forecasts

Casaioli, M., Mantovani, R., Proietti Scorzoni, F., Puca, S., Speranza, A., & Tirozzi, B. (2003). Linear and nonlinear post-processing of numerically forecasted surface temperature. Nonlinear processes in geophysics, 10(4/5), 373-383.

Dallavalle, J. P., & Erickson, M.C. (1999). MOS-2000: A new era in interpretive weather guidance? Abstracts 24th Annual Meeting of the National Weather Association, Biloxi.

Dennstaed, S. 2006. Model Output Statistics Provide Essential Data for Small Airports. The Front, vol. 6, no. 2, 2006.

Casaioli, M., Mantovani, R., Proietti Scorzoni, F., Puca, S., Speranza, A., & Tirozzi, B. (2003). Linear and nonlinear post-processing of numerically forecasted surface temperature. Nonlinear processes in geophysics, 10(4/5), 373-383.

Galanis, G., Louka, P., Katsafados, P., Pytharoulis, I., & Kallos, G. (2006, October). Applications of Kalman filters based on non-linear functions to numerical weather predictions. In Annales Geophysicae (Vol. 24, No. 10, pp. 2451-2460).

Glahn, B. (2014). Determining an optimal decay factor for bias-correcting MOS temperature and dewpoint forecasts. Weather and Forecasting, 29(4), 1076-1090.

Guarnieri, R. A., Pereira, E. B., & Chou, S. C. (2006). Solar radiation forecast using artificial neural networks in South Brazil. Proceedings of the 8th ICSHMO, 24-28.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reuteman, P., & Witten, I. (2009). The WEKA Data Mining Software: An Update. SIGKDD Explorations, vol. 11 (1).

Hamill, T. M., Gilbert, K., Peroutka, M. Myrisk, D., Farrar, M. & Tolman, H. 2015. Statistical post-processing in NOAA: Key changes necessary to make the national blend of models and reforecasting successful. Retrieved from https://www.esrl.noaa.gov/psd/people/tom.hamill/WhitePaperImplementationofReforecastsandNationalBlend.pdf

Hamill, T. M., Engle, E., Myrick, D., Peroutka, M., Finan, C., & Scheuerer, M. (2017). The US National Blend of Models for Statistical Postprocessing of Probability of Precipitation and Deterministic Precipitation Amount. Monthly Weather Review, 145(9), 3441-3463.

Klein, W. H., & Glahn, H. R. (1974). Forecasting local weather by means of model output statistics. Bulletin of the American Meteorological Society, 55(10), 1217-1227.

Klein, W. H. (1978). Objective forecasts of local weather by means of model output statistics. In Proc. ECMWF Seminar on the Interpretation and Use of Large-Scale Numerical Forecast Products (pp. 186-220).

Lauret, P., Diagne, M., & David, M. (2014). A neural network post-processing approach to improving NWP solar radiation forecasts. Energy Procedia, 57, 1044-1052.

Lin, S. J., Atlas, R., & Yeh, K. S. (2004). Global weather prediction and high-end computing at NASA. Computing in Science & Engineering, 6(1), 29-34.

Mathiesen, P., & Kleissl, J. (2011). Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. Solar Energy, 85(5), 967-977.

Marzban, C. (2003). Neural networks for postprocessing model output: ARPS. Monthly Weather Review, 131(6), 1103-1111.

Marzban, C., Leyton, S., & Colman, B. (2007). Ceiling and visibility forecasts via neural networks. Weather and forecasting, 22(3), 466-479.

Model output statistics. (2017, April 5). In Wikipedia, The Free Encyclopedia. Retrieved 15:25, September 8, 2017, from https://en.wikipedia.org/w/index.php?title=Model_output_statistics&oldid=773904369

Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather and forecasting, 8(2), 281-293.

National Center for Atmospheric Research. (2017). Short Term Weather Forecasting. Retrieved from https://ral.ucar.edu/solutions/capability/short-term-weather-forecasting.

National Weather Service. (2017). Subjective List of Model Performance Characteristics. Retrieved from http://www.wpc.ncep.noaa.gov/mdlbias/biastext.shtml.

National Weather Service Office of Meteorology. (1995). NGM-Based MOS Ceiling Height Guidance for the Contiguous United States. Silver Spring, MD: Miller, D.

Nurmi, V., Perrels, A., Nurmi, P., Michaelides, S., Athanasatos, S., & Papadakis, M. (2012). Economic value of weather forecasts on transportation–Impacts of weather forecast quality developments to the economic effects of severe weather. ewent. vtt. fi/Deliverables D, 5.

Pelland, S., Galanis, G., & Kallos, G. (2013). Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale numerical weather prediction model. Progress in photovoltaics: Research and Applications, 21(3), 284-296.

Rudack, D. E., & Ghirardelli, J. E. (2010). A Comparative verification of localized aviation model output statistics program (lamp) and numerical weather prediction (nwp) model forecasts of ceiling height and visibility. Weather and Forecasting, 25(4), 1161-1178.

Sanders, S. (2017). Solar radiation prediction improvement using weather forecasts. Unpublished.

Samenow, J. & Fritz, A. (January 2, 2015). Five myths about weather forecasting. Retrieved from https://www.washingtonpost.com/opinions/five-myths-about-weather-forecasting/2015/01/02/e49e8950-8b86-11e4-a085-34e9b9f09a58_story.html.

Stensrud, D. J., Brooks, H. E., Du, J., Tracton, M. S., & Rogers, E. (1999). Using ensembles for short-range forecasting. Monthly Weather Review, 127(4), 433-446.

Wilks, D. S., & Hamill, T. M. (2007). Comparison of ensemble-MOS methods using GFS reforecasts. Monthly Weather Review, 135(6), 2379-2390.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.

World Meteorological Organization. (2003). Introduction to GRIB Edition 1 and GRIB Edition 2. Retrieved from https://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/Introduction_GRIB1-GRIB2.pdf.

CHAPTER 4

CONCLUSION AND FUTURE DIRECTIONS

4.1 CONCLUSION

The results presented here indicate that properly-chosen and tuned machine learning algorithms can lead to a reduction in predictions errors in several weather forecasting scenarios. Chapter 2 suggests that predictive data such as future weather forecasts or real-time satellite imagery is an important component of solar radiation prediction over short- and medium-term prediction timeframes. Large-scale collection and dissemination of this data is likely to enhance prediction models and improve the operations and resource allocation of current and future solar power plants. In addition, the inclusion of weather forecasts from areas immediately surrounding the target area was found to increase predictive accuracy in both one-hour and 24-hour solar radiation predictions.

In Chapter 3, simple time-of-day and day-of-year analysis uncovered highly statistically significant biases for forecasts of eight weather variables (air temperature, cloud cover, visibility, wind speed, wind direction, dew point temperature, air pressure, and humidity) for five cities in Georgia. Correction of these biases in raw forecast data is known as postprocessing, and the most common postprocessing technique currently used in practice is a standard linear regression model termed Model Output Statistics (MOS). Incorporating the sinusoidal representation of the time-of-day and day-of-year into an MOS model results in a 20.5% reduction in Mean Absolute Error (MAE) in air temperature prediction and reductions in five of the eight weather variables analyzed. Additionally, an ensemble machine learning model consisting of a Random Forest stacked with an artificial neural network (ANN) yielded a decrease in error of 7.7% over using MOS for air temperature prediction, and a reduction in error over MOS in seven of the eight weather variables analyzed.

As stated at the beginning of this thesis, the effects of weather and weather forecasting affect millions of lives daily. Improvements in the accuracy of published weather forecasts could have a positive net economic and social benefit on their many users. The results reported here indicate that advanced machine learning techniques such as Random Forests and Artificial Neural Networks can compete with or surpass the predictions given by the state-of-the-art models currently in production. As more weather-related data is collected every day, at ever-larger scales and ever-smaller granularities, the advantages held by these advanced models may grow further still.

4.2 FUTURE DIRECTIONS

One obvious avenue of research extension would be to attempt weather predictions for a broader range of climates and macroclimates. The variations in weather throughout the five cities in Georgia analyzed in this thesis are relatively modest in comparison to the wide range of atmospheric conditions present throughout the world. It would be interesting to see if the results published here are applicable to more variable weather systems.

Supplementary data from predicted times other than the target time could also be integrated into the models presented here. Klein (1978) notes that numerical weather forecasting models can be systematically slow or fast, and therefore predictors within a short period of the predictand time could potentially be useful as inputs. If this concept were explored, it must be noted that NOAA's data is produced at infrequent intervals, which could limit the effectiveness of additional data inputs. Other data sources could be researched and added as well.

Chapter 2 noted that inclusion of weather forecasts from areas immediately surrounding the target area improved solar radiation prediction accuracy for both one-hour and 24-hour predictions, but Chapter 3 showed that those improvements did not translate into an error reduction when included in postprocessing models to predict air temperature. However, it was found that there were slight improvements in accuracy when combining data from geographically similar cities into a single model. More research could be done to quantify this effect; if it persists, it could prove useful when working with data sets of limited size. Many historical weather data sets for a single target area are limited in size, especially once seasonality effects are taken into account. Any techniques which could leverage additional data to add further statistical significance to model results would be a valuable step forward.

Finally, NOAA and other providers of raw weather forecasts update their models on a frequent basis in order to improve accuracy and reduce the variation of residual errors. Glahn (2014) reported that using historical data decay factors between 0.025 and 0.1 improved the MAE of 72-hour temperature forecasts over using MOS alone, and posited that the ever-changing nature of NWP models can cause post-processing techniques to become stale if they do not underweight older data. More research could be done on this phenomenon to determine how the more advanced techniques detailed in this thesis react to changes in the underlying raw forecast models.

4.3 REFERENCES

Glahn, B. (2014). Determining an optimal decay factor for bias-correcting MOS temperature and dewpoint forecasts. Weather and Forecasting, 29(4), 1076-1090.

Klein, W. H. (1978). Objective forecasts of local weather by means of model output statistics. In Proc. ECMWF Seminar on the Interpretation and Use of Large-Scale Numerical Forecast Products (pp. 186-220).

BIBLIOGRAPHY

American Meteorological Society. (2015). Weather Analysis and Forecasting.

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. Nature, 525(7567), 47-55.

British Meteorological Service. (2017). How Accurate are our Public Forecasts? Retrieved from http://www.metoffice.gov.uk/about-us/who/accuracy/forecasts

Casaioli, M., Mantovani, R., Proietti Scorzoni, F., Puca, S., Speranza, A., & Tirozzi, B. (2003). Linear and nonlinear post-processing of numerically forecasted surface temperature. Nonlinear processes in geophysics, 10(4/5), 373-383.

Cheung, W., Zhang, J., Florita, A., Hodge, B. M., Lu, S., Hamann, H. F., ... & Lehman, B. (2015). Ensemble Solar Forecasting Statistical Quantification and Sensitivity Analysis (No. NREL/CP-5D00-67191). NREL (National Renewable Energy Laboratory (NREL), Golden, CO (United States).

Dallavalle, J. P., & Erickson, M.C. (1999). MOS-2000: A new era in interpretive weather guidance? Abstracts 24th Annual Meeting of the National Weather Association, Biloxi.

Dennstaed, S. 2006. Model Output Statistics Provide Essential Data for Small Airports. The Front, vol. 6, no. 2, 2006.

Feudo, T. L., Avolio, E., Gullì, D., Federico, S., Calidonna, C. R., & Sempreviva, A. M. (2015). Comparison of Hourly Solar Radiation from a Ground–Based Station, Remote Sensing and Weather Forecast Models at a Coastal Site of South Italy (Lamezia Terme). Energy Procedia, 76, 148-155. DOI: 10.1016/j.egypro.2015.07.884.

Galanis, G., Louka, P., Katsafados, P., Pytharoulis, I., & Kallos, G. (2006, October). Applications of Kalman filters based on non-linear functions to numerical weather predictions. In Annales Geophysicae (Vol. 24, No. 10, pp. 2451-2460).

Glahn, B. (2014). Determining an optimal decay factor for bias-correcting MOS temperature and dewpoint forecasts. Weather and Forecasting, 29(4), 1076-1090.

Guarnieri, R. A., Pereira, E. B., & Chou, S. C. (2006). Solar radiation forecast using artificial neural networks in South Brazil. Proceedings of the 8th ICSHMO, 24-28.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reuteman, P., & Witten, I. (2009). The WEKA Data Mining Software: An Update. SIGKDD Explorations, vol. 11 (1).

Hamill, T. M., Gilbert, K., Peroutka, M. Myrisk, D., Farrar, M. & Tolman, H. 2015. Statistical post-processing in NOAA: Key changes necessary to make the national blend of models and reforecasting successful. Retrieved from https://www.esrl.noaa.gov/psd/people/tom.hamill/WhitePaperImplementationofReforecastsandNationalBlend.pdf

Hamill, T. M., Engle, E., Myrick, D., Peroutka, M., Finan, C., & Scheuerer, M. (2017). The US National Blend of Models for Statistical Postprocessing of Probability of Precipitation and Deterministic Precipitation Amount. Monthly Weather Review, 145(9), 3441-3463.

Hamilton, C. (2016). Hourly solar radiation forecasting through neural networks and model trees. (Unpublished master's thesis). University of Georgia, Athens, Georgia.

Hoogenboom, G., Verma, B.P. & Threadgill, E.D. (1990). The development of the Georgia Automated Environmental Monitoring Network. *Proceedings of the 1991 Georgia Water Resources Conference*, held March 19 and 20, 1991, at The University of Georgia.

International Energy Agency. (2014). *Technology roadmap: Solar photovoltaic energy*. Retrieved from http://www.iea.org/publications/freepublications/publication/TechnologyRoadmapSolarPhotovoltaicEnergy_2014edition.pdf.

Ipsos. (2011). *Global citizen reaction to the Fukushima nuclear plant disaster*. Retrieved from https://www.ipsos-mori.com/Assets/Docs/Polls/ipsos-global-advisor-nuclear-power-june-2011.pdf.

Klein, W. H., & Glahn, H. R. (1974). Forecasting local weather by means of model output statistics. Bulletin of the American Meteorological Society, 55(10), 1217-1227.

Klein, W. H. (1978). Objective forecasts of local weather by means of model output statistics. In Proc. ECMWF Seminar on the Interpretation and Use of Large-Scale Numerical Forecast Products (pp. 186-220).

Lara-Fanego, V., Ruiz-Arias, J.A., Pozo-Vazquez, D., Santos-Alamillos, F.J., & Tovar-Pescador, J. (2012). Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain). Sol. Energy 86 (8), 2200–2217.

Larson, D. P., Nonnenmacher, L., & Coimbra, C. F. (2016). Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest. Renewable Energy, 91, 11-20.

Lauret, P., Diagne, M., & David, M. (2014). A neural network post-processing approach to improving NWP solar radiation forecasts. Energy Procedia, 57, 1044-1052.

Lazo, J. K., Morss, R. E., & Demuth, J. L. (2009). 300 billion served: Sources, perceptions, uses, and values of weather forecasts. Bulletin of the American Meteorological Society, 90(6), 785-798.

Lin, S. J., Atlas, R., & Yeh, K. S. (2004). Global weather prediction and high-end computing at NASA. Computing in Science & Engineering, 6(1), 29-34.

Linares-Rodriguez, A., Ruiz-Arias, J.A., Pozo-Vazquez, D., & Tovar-Pescador, J. (2013). An artificial neural network ensemble model for estimating global solar radiation from Meteosat satellite images. Energy 61, 636–645.

Mathiesen, P., & Kleissl, J. (2011). Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. Solar Energy, 85(5), 967-977.

Marzban, C. (2003). Neural networks for postprocessing model output: ARPS. Monthly Weather Review, 131(6), 1103-1111.

Marzban, C., Leyton, S., & Colman, B. (2007). Ceiling and visibility forecasts via neural networks. Weather and forecasting, 22(3), 466-479.

Miller, S.D., Heidinger, A.K., & Sengupta, M., 2013. Physically based satellite methods. In: Kleissl, J. (Ed.), Solar Energy and Resource Assessment. Elsevier, Oxford, pp. 49–79.

Model output statistics. (2017, April 5). In Wikipedia, The Free Encyclopedia. Retrieved 15:25, September 8, 2017, from https://en.wikipedia.org/w/index.php?title=Model_output_statistics&oldid=773904369

Murphy, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather and forecasting, 8(2), 281-293.

National Center for Atmospheric Research. (2017). Short Term Weather Forecasting. Retrieved from https://ral.ucar.edu/solutions/capability/short-term-weather-forecasting.

National Weather Service. (2017). Subjective List of Model Performance Characteristics. Retrieved from http://www.wpc.ncep.noaa.gov/mdlbias/biastext.shtml.

National Weather Service Office of Meteorology. (1995). NGM-Based MOS Ceiling Height Guidance for the Contiguous United States. Silver Spring, MD: Miller, D.

Nurmi, V., Perrels, A., Nurmi, P., Michaelides, S., Athanasatos, S., & Papadakis, M. (2012). Economic value of weather forecasts on transportation–Impacts of weather forecast quality developments to the economic effects of severe weather. ewent. vtt. fi/Deliverables D, 5.

Paz, J. O., & Hoogenboom, G. (2008). Climate and weather information for Georgia farmers. Cooperative Extension, the University of Georgia.

Pedro, H. T., & Coimbra, C. F. (2012). Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, *86*(7), 2017-2028.

Pelland, S., Galanis, G., & Kallos, G. (2013). Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale numerical weather prediction model. Progress in photovoltaics: Research and Applications, 21(3), 284-296.

Perez, R., Lorenz, E., Pelland, S., Beauharnois, M., Van Knowe, G., et al. (2013). Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. Sol. Energy 94, 305–326.

Rudack, D. E., & Ghirardelli, J. E. (2010). A Comparative verification of localized aviation model output statistics program (lamp) and numerical weather prediction (nwp) model forecasts of ceiling height and visibility. Weather and Forecasting, 25(4), 1161-1178.

Ruiz-Arias, J. A., Quesada-Ruiz, S., Fernández, E. F., & Gueymard, C. A. (2015). Optimal combination of gridded and ground-observed solar radiation data for regional solar resource assessment. Solar Energy, 112, 411-424.

Sanders, S. (2017). Solar radiation prediction improvement using weather forecasts. Unpublished.

Samenow, J. & Fritz, A. (January 2, 2015). Five myths about weather forecasting. Retrieved from https://www.washingtonpost.com/opinions/five-myths-about-weather-forecasting/2015/01/02/e49e8950-8b86-11e4-a085-34e9b9f09a58_story.html.

Sfetsos, A., & Coonick, A. H. (2000). Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. Solar Energy,68(2), 169-178

Spokas, K. & F. Forcella. (2006.) Estimating hourly incoming solar radiation from limited meteorological data. Weed Science 54:182–189.

Stensrud, D. J., Brooks, H. E., Du, J., Tracton, M. S., & Rogers, E. (1999). Using ensembles for short-range forecasting. Monthly Weather Review, 127(4), 433-446.

University of Georgia Office of Sustainability. (2017). *Renewable Energy*. Retrieved from http://sustainability.uga.edu/what-were-doing/renewable-energy/.

Wilks, D. S., & Hamill, T. M. (2007). Comparison of ensemble-MOS methods using GFS reforecasts. Monthly Weather Review, 135(6), 2379-2390.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.

World Meteorological Organization. (2003). Introduction to GRIB Edition 1 and GRIB Edition 2. Retrieved from https://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/Introduction_GRIB1-GRIB2.pdf.

Zack, J. (2014). Solar power production forecasting: overview of methods and input data needs. *ERCOT ETWG Meeting*, April 30, 2014, Austin, Texas.

Zervas, P.L., Sarimveis, H., Palyvos, J. A. & Markatos, N. C. G. (2008). Prediction of daily global solar irradiance on horizontal surfaces based on neural-network techniques. Renewable Energy, vol. 33, pp. 1796-1803.