Improving Short-Term Local Solar Energy Forecasts for Optimizing Power Generation Using Machine Learning

by Jeffrey D. Cwagenberg

B.S.E. in Earth Systems Science and Engineering, May 2014, University of Michigan
M.Eng. in Systems Engineering, December 2017, Pennsylvania State University

A Praxis submitted to

The Faculty of
The School of Engineering and Applied Science
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Engineering

August 31, 2021

Praxis directed by

Aristos Dimitriou
Professional Lecturer of Engineering Management and Systems Engineering

Timothy Eveleigh
Professional Lecturer of Engineering Management and Systems Engineering

The School of Engineering and Applied Science of The George Washington University certifies that Jeffrey Daniel Cwagenberg has passed the Final Examination for the degree of Doctor of Engineering as of August 5, 2021. This is the final and approved form of the Praxis.

Improving Short-Term Local Solar Energy Forecasts for Optimizing Power Generation using Machine Learning

Jeffrey D. Cwagenberg

Praxis Research Committee:

Aristos Dimitriou, Professional Lecturer of Engineering Management and Systems Engineering, Praxis Co-Director

Timothy Eveleigh, Professional Lecturer of Engineering Management and Systems Engineering, Praxis Co-Director

Amir Etemadi, Associate Professor of Engineering and Applied Science, Committee Member

Dedication

This project is dedicated to everyone who has had the pleasure of driving through the Australian Outback.

Acknowledgements

The author wishes to acknowledge Leidos, Inc. for financial support in the completion of this program in addition to the faculty and staff of the George Washington University for providing guidance, support, and resources that have been invaluable to the completion of this project. Without the above, in addition to the support and encouragement of family, friends, and associates- this endeavor would not have been possible.

Abstract of Praxis

Improving Short-Term Local Solar Energy Forecasts for Optimizing Power Generation using Machine Learning

This Praxis introduces the SAFE-T (Solar Advanced Forecast Expectation Tool) as an approach to improve local area solar energy forecasts using machine learning based techniques in MATLAB. With the increasing cost competitiveness of solar versus diesel electricity production, there has been a push by utilities to deploy solar power generating capacity across areas that have not historically been popular for solar power. (Georgitsioti, et al., 2015) While typical solar forecast models have been improving for standard locations, model accuracy has not kept up across all areas where forecasts are now desired. (White, et al., 1999) Recent advancements in large area solar forecasting have come from advanced physics and statistics methods. (Du, 2019) (Verbois, Huva, Rusydi, & Walsh, 2018) These improvements, however, have left a gap for medium term forecasting (48 to 96 hours) for single point areas in parts of the world that lack significant collections of historical solar energy measurements. This research focuses on developing a model to enable electric utilities in these areas to decrease forecast error by using a GlobalSearch based machine learning optimization combined with a multiple linear regression model to improve solar forecasts. With a less error prone forecast, electric utilities will be able to better optimize the amount of backup diesel generator capacity (and fuel) to deploy for a project. In doing so, utilities will be able to decrease the deployment of wasted resources, which can reduce the total cost of electricity generated by up to 35%. (Georgitsioti, et al., 2015)

Table of Contents

List of Figures

List of Tables

List of Symbols

a_n	Model weighting
e	Vector of ones in the same size as g
f	Optimization function of interest
g(x)	Inequality constraints
h(x)	Equality constraints
H	$H = \nabla^2 f(x) \sum_i \lambda_i \nabla^2 g_i(x) + \sum_j y_i \nabla^2 h_j(x)$
J_g	Jacobian of the constraint function g
J_h	Jacobian of the constraint function h
km	Kilometers
m	Meters
M_n	WRF model forecast value at a specific location/time (watts/square meter)
n	Maximum Number of Models
p	Population proportion
s	Slack variable
S	S=diag(s), where diag is the matrix operator returning values on diagonal
ν	Gradient
W_n	Weighting coefficient of the n^{th} WRF model (fractional)
WRF_n	WRF model output for the n^{th} WRF
w/m^2	Watts per square meter
X_k	Model factor
y	Denotes the Lagrange multiplier vector associated with h.
y	Solar radiation modeled value

$\hat{y}$	Solar radiation measured value
z	Test statistic
Z	Zulu Time (UTC)
ε	Acceptable margin of error
λ	The Lagrange multiplier vector associated with constraints g
Λ	$\Lambda = \text{diag}(\lambda)$, diag is the matrix operator returning values on diagonal

List of Equations

List of Acronyms

AIRS	Archive Information Request System
AMS	American Meteorological Society
ASCII	American Standard Code for Information Interchange
BMS	NREL - Solar Baseline Measurement System
CSV	Comma Separated Value
DEG	Degree
ECMWF	European Center for Medium-Range Weather Forecasts
fmincon	Interior-Point Search Optimization Function in MATLAB
FNNs	Fuzzy Neural Networks
GA	Genetic Algorithm Optimization Function in MATLAB
GB	Gigabyte
GFS	Global Forecast Service
GIS	Geographic Information Service
GHI	Global Horizontal Irradiance
GlobalSearch	Scatter-Search Optimization Function in MATLAB
GPUs	Graphical Processing Units
GS	GlobalSearch Algorithm in MATLAB
GUI	Graphical User Interface
HTML	Hypertext Markup Language
IEEE	Institute of Electrical and Electronics Engineers
MAE	Mean Average Error

MAPE	Mean Absolute Percentage Error
MOS	Model Output Statistics
NAM	North American Mesoscale Model
NCAR	National Center for Atmospheric Research
NREL	National Renewable Energy Laboratory
NWP	Numerical Weather Prediction
particleswarm	Particle-Velocity Evolution Function in MATLAB
patternsearch	Adaptive-Mesh Function in MATLAB
RAM	Random Access Memory
RMSE	Root Mean Square Error
RNNs	Recurrent Neural Networks
SAFE-T	Solar Advanced Forecasting Tool
SSD	Solid State Drive
surrogateopt	Interpolate Objective Function in MATLAB
TB	Terabyte
TSI	Total Sky Imager
USA	United States of America
UTC	Coordinated Universal Time
WRF	Weather Research Forecast Model

Chapter 1—Introduction

1.1 Background

To prevent supply-side electrical shortages, utilities need to match electrical generating capacity to demand. Prior to the use of renewable sources for energy production, utilities primarily focused on developing tools to improve the precision and accuracy of forecasting demand side electrical needs. (Kyriakides & Polycarpou) With the decreasing cost of renewable energy technologies, there has been an increase in the deployed capacity and number of unique solar power generating facilities. As a result, utilities have faced a challenge in matching solar energy forecasting capability as solar forecast model development had not kept up with the development of energy demand models. (Zhang, Beaudin, Taheri, Zareipour, & Wood, 2015)

This praxis was completed in support of a University of Michigan project to deploy sensors across remote sections of the Australian Outback. The University of Michigan project plans to deploy a series of monitoring sensors across the Outback to improve the accuracy of an operational energy model built by IBM. The remoteness of these planned deployment locations will require each site to be electrically self-sufficient. To support this, the University of Michigan group plans to deploy solar based electrical generating capacity to minimize the total cost of deployment. The costs of deploying traditional energy generating approaches can be high as the Australian Outback poses a logistical challenge with anticipated 96-hour lead times required to deploy additional fuel or backup capacity. This praxis focused on identifying an approach to reduce error in

localized solar energy forecasts to appropriately deploy only the needed solar capacity as the minimization of costs would support deploying additional sensors.

1.2 Research Motivation

This praxis was motivated by a planned collection of sensor data from remote portions of the Australian Outback. The 2015 partnership between IBM and the University of Michigan to test a statistical based machine learning operational model identified the need for additional data collection to support improvements to the IBM developed model. (Shao, et al., 2016) This praxis research was formed to identify an approach to temporarily improve the performance of point-location forecasts for solar power generating deployments. This temporary improvement was designed in support of minimizing the cost of collecting data needed to improve the overall IBM model.

1.3 Problem Statement

Utilities cannot accurately estimate solar energy production for small projects (Nagendra, Narayanmurthy, Moser, & Singh, 2020)*, based on a doubling solar model prediction errors for small areas,* (White, et al., 1999) *resulting in increased energy production cost of up to 35%* (Georgitsioti, et al., 2015).

As the average price for solar energy generating photovoltaic panels declined from $4 per watt of energy in 2006 to under $1 per watt in 2012, there has been an increase in deployment of these technologies with new yearly deployed capacity doubling approximately every four years. (Zhang, Beaudin, Taheri, Zareipour, & Wood, 2015) (Solar Energy Industries Association, 2020). In order to deploy appropriate amounts of both solar generating capability and backup diesel capability, utilities rely on the accuracy of solar weather forecasts. While significant effort has been expended on

advances in statistical and physics-based solar prediction modeling, a lack of focus on local physical and environmental considerations increases the error when a regional or global model is used to predict for a single location. (Soares de Araujo, 2020) (White, et al., 1999). The uncertainties in forecast accuracy cause utilities to deploy excess backup generating capacity, which causes an increase in overall energy production costs by up to 35%. (Georgitsioti, et al., 2015)

1.4 Thesis Statement

Combining multiple solar prediction models using machine learning can generate solar forecasts that decrease watt-per-square-meter error for a specific location.

The Weather Research and Forecasting (WRF) model is a generic modeling framework developed and maintained by the National Center for Atmospheric Research (NCAR). The WRF model provides multiple selectable options for different underlying physics-based approaches to performing numerical weather prediction (NWP) based on both research-based and community provided inputs. (Powers, et al., 2017) These different variants of WRF model the atmosphere and propagate weather patterns using slightly different mathematical approaches. This praxis was designed to demonstrate that a machine learning based WRF ensemble model can reduce solar prediction error caused by output variations across the different WRF versions for single point-locations that are unknown to the ultimate end user of the forecast products.

1.5 Research Objectives

The purpose behind the research undertaken in this praxis was to develop a MATLAB based tool that generates improved solar radiation forecasts for use in deploying localized (single point) solar energy generation equipment. This tool was

designed to utilize machine learning techniques to select the best combination of numerical weather prediction model(s) for the deployment site of interest. The base underlying method was to use an optimization approach that leverage past performance of individual forecast models to predict the future performance of each model for the location of interest. (Li, Ren, Chen, Wang, & Qi, 2020). This base methodology enabled the investigation of using measured temperature as an alternative training variable for solar energy forecasts. This alternative measurement was investigated because large portions of the Australian Outback lack historical collected solar radiation data (Bureau of Meterology, 2021) but have measured temperature records. (Bureau of Meterology, 2021) Based on industry information, the end user expected up to a 35 percent decrease in the cost of generating electricity with the improved localized solar radiation predictions for 0–96-hour forecasts. (Georgitsioti, et al., 2015) This decreased cost would enable a larger deployment of monitoring sensors without increased budget.

1.6 Research Questions and Hypotheses

To implement these objectives, a series of questions and associated research hypothesis were formulated to provide guidance in the research process. These were based on an understanding of the existing state of technology with respect to solar forecasts and availability of wide area solar radiation collected truth data.

RQ1: Can multiple solar weather forecast models be combined in a way that reduces average solar energy prediction error over using a single model?

RQ2: Would combinations of multiple solar forecast models improve energy predictions for locations where individual models alone perform poorly?

RQ3: Can surface temperature be used to improve solar energy predictions for combined weather models in locations without collected solar measurements?

H1: Weighted combinations of Weather Research Forecast (WRF) models will reduce solar energy prediction error for at least 75% of individual locations.

H2: Weighted combinations of Weather Research Forecast (WRF) models will reduce prediction error by at least 5% on average for 25% of the worst performing locations using a single model.

H3: Weighted combinations of Temperature Based Weather Research Forecast (WRF) models will reduce solar energy prediction error on average by at least 2%.

1.7 Scope of Research

The scope of the research performed, and the capabilities of the tool developed for this praxis were designed to meet the needs of the University of Michigan end-users. This praxis focused on using the WRF model framework to enable the developed tool to support improved solar forecasts for any location on earth with either recorded temperature or solar radiation data. As the WRF framework supports global use of this model, for the purposes of praxis development, a region with high availability of truth data was selected for analysis. In this case, the United States was chosen for assessment of the praxis developed model because of ease of data availability through the National Climatic Data Center to enable performance assessment.

This praxis investigated multiple approaches for model selection, combination, and optimization. The praxis first investigated possible weather models of interest to integrate. Based on the time and resources available, the investigation selected the widely accessible Weather Research Forecast model. Six versions of the WRF forecast model

were chosen to support the research of combining multiple different weather into a weighted ensemble model. These six different models comprised the total combination of two cloud models (Ferrier and WSM 6-Class), two longwave solar radiation (RRTM and GFDL), and two shortwave radiation (Dudhia and GFDL) mathematical models taken from the models used operationally by the National Weather Service. (NOAA, 2021)

The scope of the developed research tool was also set based on end user requirements. The University of Michigan group required the tool to be developed in the MATLAB environment on engineering workstations, which limited the scope of possible combination and optimization approaches. This praxis focused on a multiple linear combination model based on literature discussed in Chapter 2 for the first portion of the foundational research. For the optimization method portion of research, the scope was limited to all seven optimization algorithms native to MATLAB. The functional/physical requirements of the developed tool were set by the end user organization based on engineering team computing workstation capabilities. The notional engineering workstation contained an Ubuntu 18.04 OS running on a quad-core Intel i-7 with 32 GB of RAM and a 1TB SSD, which was the environment for development under this praxis.

1.8 Research Limitations

This praxis attempted to encompass all weather models, combination approaches, and optimization algorithms. The objectives of the end-user team and the time/resources restrictions imposed by faculty at the George Washington University, however, limit the scope of this research. This praxis ultimately focused on a subset of the state-of-the-art available options to support the development of this tool, discussions on prioritization are found in Chapter 2. The primary limitation to the scope of this research came from the

requirement of a MATLAB based solution to interface with existing software tools. This limitation restricted the available approaches for optimization and subsequently combination methods to those contained within the MATLAB Global Optimization toolbox. Additional limitations were placed on the number and type of weather models integrated into the analysis based on time/resources and accessibility. The combination tool developed under this praxis was designed to ingest any available weather model to integrate. Due to the large data-files (>1.2GB per forecast) and run-times (>1.5 Hrs per forecast) a limitation trade-off was performed between the number of models integrated and the number of locations assessed. The optimal trade-off based on climates of interest, model resolution, and data points was set at six Models, 2401 Locations, and 219 data points. This praxis initially planned to integrate multiple weather models from different development organizations. The European Center for Medium Range Weather Forecasting (ECMWF) – a preeminent forecasting model, however, does not have a license that permits use for this application. (ECMWF, 2021) Additionally, limited ground truth availability of solar energy measurements restrict the ability to rely on only using historical solar data, so historical temperature data was also pulled forward. (Bureau of Meterology, 2021) (Bureau of Meterology, 2021) Ultimately, multiple variants of WRF, with different underlying approaches to forecasting weather were selected because of the open source availability of the code. (Skamarock, et al., 2019)

1.9 Organization of Praxis

The research under this praxis was structured to provide a detailed assessment of the process involved in developing the research products for improved localized solar

weather forecasting. Chapter one provides the introduction and background information needed to understand the purpose behind the research effort undertaken.

Chapter two provides an assessment of literature topics and approaches to improve forecasting and the prediction of solar energy. This chapter details key factors and technologies needed for understanding the development of the solar energy prediction tool. Topics covered in the Chapter two literature review include operational solar forecasting approaches (historical, current, and future). The literature survey first covered assessments of Numerical Weather Prediction (NWP) and Statistical Forecasting capabilities and drawbacks. Then the survey changed focus to identify approaches to improve solar forecasts which included ensemble combination approaches such as the multiple linear regression, optimization approaches such as GlobalSearch, and error calculation approaches such as Root-Mean-Square-Error (RMSE). The final piece of the literature survey investigated approaches to assess the performance of developed forecast tools to provide a means to validate improvements of the praxis modeling tool.

Chapter three then presents an overview of the methodologies used within the praxis for generating the research product. This chapter details the specifics of the WRF models used, the weighted linear combination approach, and the GlobalSearch optimization method. This chapter details not only the underlying approaches and error equations, but also provide a justification for the selection of the approaches.

Chapter four then dives into the results of the praxis developed tool which is based on the algorithms described in chapter three and the analysis methods cited in chapter two. Attempting to validate the hypotheses proposed in chapter one, chapter four details the systematic approach this praxis takes to collect, process, and analyze data to

confirm/reject the null hypotheses. Once the raw and statistical calculation approaches have been completed, this chapter presents the measured performance.

Chapter five then concludes the praxis with a discussion of the impacts of the results from chapter four in the scope of the praxis research objectives. This chapter will focus on providing a context to understand what impact these results will have for projects that use or extend this work. Additionally, chapter five will detail the drawbacks encountered and suggest opportunities for further investigation based on the praxis' results.

Chapter 2—Literature Review

2.1 Introduction

The literature review detailed in this chapter is designed to provide an assessment of need, analysis of previous approach applicability, and relevance of potential methodologies. The goal of this survey is to understand the existing landscape for use in the development of an improved solar energy forecasting tool. The literature review is structured in a combined thematic/methodological manner where each section provides a single theme or methodology of importance to the development of the research product. The scope of this literature review is focused on first addressing the need for this praxis research and assessing the applicability of previous efforts from adjacent problem areas for use. Based on the identification of multiple approaches, this review will focus on detailing the justification behind the underlying methodologies and validation approaches selected to be utilized. A comprehensive review of related literature was completed, however not all available sources has been specifically cited in this overview. The sources selected focused on peer-reviewed engineering and management technical journals such as from the Institute of Electrical and Electronics Engineers (IEEE) and American Meteorological Society (AMS). In addition, articles from similar professional conference proceedings and governmental documents and websites are included in the breadth of the research assessed. Sources that have been selected for this assessment are primarily less than 10 years old. Certain topics – especially foundation topic research – date back to the 1970s, but these are kept to a minimum. The outcome of this review will

demonstrate that the need exists for a tool designed to improve solar forecasts for localized areas using a machine learning optimized weighted linear combination model.

2.2 Photovoltaic Energy Generation

The share of US energy generation by solar photovoltaics has grown by approximately 1% between 2011 and 2017 resulting in more than 47 Gigawatts of total electrical generating capacity. (US Department of Energy, 2019) With this growth, solar generating capacity as of 2017 has remained a small share of total energy production. Experts forecast that the cost advantage per produced watt of electricity for natural gas over solar generation should disappear by 2050. (U.S. Energy Information Administration, 2020) The recent trends would suggest that solar photovoltaics may account for 30% of total electrical generation by 2050. (Aminifar, Shahidehpour, Alabdulwahab, Abusorrah, & Al-Turki, 2020). Major urban areas have utilized these deployments to supplement the electrical grid, but for remote locations- solar is being assessed as a primary generating source. (Taylor, Ralon, Anuta, & Al-Zoghoul, 2020) For these remote locations, diesel electrical generation has been the most common source of power generation. As the cost-benefit has swung towards solar photovoltaic approaches, electrical utilities have looked for major cost savings not only by reducing dependence on diesel fuel but also eliminating costly fuel transportation. (Lamas, 2016) One drawback for solar energy production, however, remain large fluctuations in solar energy output and associated uncertainty in forecasts which force utilities to supply excess backup capacity and fuel. (Peters, Kilper, Calais, Jamal, & von Maydell, 2018) This drawback has proven to negate much of the savings of switching to photovoltaic. (Lamas, 2016) (Yengejeh, Shahnia, & Islam, 2014) The University of Michigan end user for this praxis

research is pursuing solar energy generation for deployments in remote locations like those addressed above. This survey of previous research, therefore, concludes that a gap in capability exists for accuracy local solar forecasts in small remote solar energy deployments. The impact of this gap is of importance to the perspective end users of this praxis developed model, as minimizing the gap will provide improved outcomes.

2.3 Impacts and Drivers of Inaccurate Solar Forecasts

To understand where the gaps in local area forecast uncertainty exist, this literature survey turned to identifying drivers of overall forecast approach accuracy. Through the identification of these drivers of inaccuracy, this praxis intends to confirm the need for a method to improve forecasts. In literature, it is acknowledged that large area solar energy deployments have seen improving forecast uncertainty which have not been seen in local area forecasts. (Shakya, et al., 2018) With large-scale solar forecast model accuracy having improved by 33-66% over the last decade, the approaches taken have not been to the benefit of individual location forecast accuracy. (Blaga, et al., 2019)

A study on localized variations in forecast error was done for multiple locations across the Republic of Korea. This study found average model error to vary between 6.1% and 13.9% using a single model for different locations. (Kim, Kim, Kang, & Yun, 2018) These errors were reduced when forecast error was assessed across the entire model area of operation. This better overall performance was attributed to underlying model assumptions which caused the model to underperform for specific locations. (Kim, Kim, Kang, & Yun, 2018) Examples of these assumptions include differences in energy transport equation assumptions or local geography features that occur at a finer resolution than the data/equations utilized. (Kotamarthi, Mearns, Hayhoe, Castro, & Wuebbles,

2016) (Rodrigo, et al., 2019) Global forecast models like the Weather Research and Forecasting (WRF) model rely on these approximations to calculate atmospheric dynamics at the mesoscale and synoptic level (200m to 2000km). (NCAR, 2020) These atmospheric levels, associated with large scale weather patterns do not require the feature resolution of a local model to effectively calculate the system dynamics. (Stull, 2003) For local area forecasts, the impact of microscale features (2 to 200m) on model errors are heavily influenced by the local terrain. (Frnda, et al., 2049) Other features that have been seen to drive forecast error at the local level are attributed to insufficient resolution data including; Placement (Leeward/Windward) of a mountain, measurements of surface covering (grass, water, etc) (Kotamarthi, Mearns, Hayhoe, Castro, & Wuebbles, 2016) and/or urban surface coverings. (Meyer, et al., 2020)

Widespread availability of surface estimation parameters would not overcome all physics-based model errors because of scientific unknowns in interactions and computational limitations. (Stull, 2003) Figure 2-1, below shows raw WRF model outputs which exhibit this difference in output. The highlighted boxes in each showcase different output models (for the same time) and demonstrated this literature phenomena.

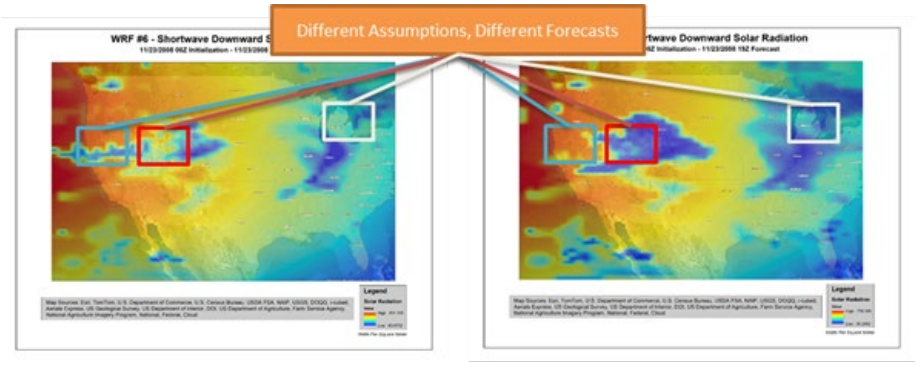


Figure 2-1. Two raw WRF Model outputs with different time-aligned forecasts. Data Source: (NCAR, 2020)

One proposed reason behind this difference in output is the different underlying approaches used to calculate initial conditions and assumptions across lateral boundary conditions. (Kumar, Kishtawal, & Pal, 2015) One area of recent forecast model development has been the use of these different performance profiles to select the best model for a location. (Frnda, et al., 2049) This research, however, is limited based on the availability of data to assess errors and use of time and location averaging to across large areas. This methodology does not directly support the praxis development because of these requirements as this praxis looks to investigate an approach for small area forecasts.

2.4 Numerical Solar Weather Prediction

This Numerical Weather Prediction (NWP) section is organized to detail this approach to forecasting and discuss literature assessed improvements to NWP methods. Numerical Weather Prediction is the process of using physics-based equations to forecast future weather that was initially proposed in 1904 by Norwegian physicist Vilhelm Bjerknes. This approach has grown to be the predominant method for weather and solar energy forecasting. (National Ocean Service - U.S. Government, 2021) At a high level, NWP models rely on an understanding of atmospheric physics to integrate radiative transfer and fluid-dynamics equations for estimating the propagation of weather systems. (National Centers for Environmental Information, 2021) NWP models rely on inputs of assessed current conditions and surface feature estimates to feed the equations. (National Centers for Environmental Information, 2021) Small deviations in these inputs from actual conditions have led to major deviations in the resulting forecast. (Jankov, Gallus Jr., Segal, & Koch, 2007) Even with multiple approaches available to feed these data

types, for particularly remote locations the availability and resolution of data has lagged more populated centers. (Bureau of Meterology, 2021)

To initialize an NWP model, a variety of dataset types can be used. (National Centers for Environmental Information, 2021) For remote corners of the world, such as the Australian Outback, often times satellite derived observations are the only source of data. (Ohring, Lord, Deber, Mitchell, & Ji, 2002) Satellite derived data, while sufficient for many lower resolution applications, has less accuracy attributed to the lower resolution and limited type of data that is available to seed models. (Yin, Alves, & Oke, 2011). The benefit of space borne sensors that enables collecting large swaths of data over limited time periods also restricts the ability to measure parameters like soil temperature and moisture. (Xu, Qu, Hao, Zhu, & Gutenberg, 2020) Additionally, these sensors also have lower spatial resolutions typically no greater than 0.5km. (P-Tree System, 2021) For NWP models which have not been designed specifically to exclude the inputs for a particular region, outputs experience increased prediction error. (Mathiesen, Collier, & Kleissl, 2013) This data mismatch issue has been particularly prevalent during the COVID-19 pandemic as the frequency of airplane flights decreased. (Chen, 2020) Sensors mounted to aircraft have been used as a major source of weather model data since the mid-2000s. (Chen, 2020) The reduction of collected data due to the pandemic resulted in a reduction of forecast accuracy for localities where errors had been improved by the prior introduction of this data. (Chen, 2020)

Numerical Weather Prediction models, in addition to seed data errors, have seen similar errors identified due to the lack of high resolution geo-physical data. (Jee & Kim, 2017) The Weather Research and Forecasting (WRF) model is foundationally based on

using static geographical data collected from NASA's MODIS satellites at a 30 arc-second (1km) spatial resolution. (NCAR, 2020) The base WRF model is matched to this resolution, but any features of smaller resolution cell size fail to be modeled – as WRF is designed. (Moeng, Dudhia, Klemp, & Sullivan, 2007) This can cause the underlying equations to underperform in the areas where higher-resolution Geographic Information Service (GIS) data exists. (Siewert & Kroszczynski, 2020) (NASA, 2021) For the WRF model, this impact has been seen for the hydrostatic equation approximations for fluid-dynamic processes calculated near mountains. (Tolstykh, 2005) These inaccurate approximations impact the model performance for features like cloud formation because of ignored smaller-scale and/or "random" atmospheric processes and interactions. (Tolstykh, 2005) This lack of data has been assessed to degrade forecast model performance by up to seventy percent for extremely biased (poor performing) individual locations. (Jimenez, et al., 2016)

Performance patterns for different forecast models were initially noticed by weather forecasters as early as the 1970s. (Glahn & Lowry, 1972) Certain weather conditions and patterns were seen to be handled with different levels of accuracy by different forecast models. (Glahn & Lowry, 1972) Forecasters at the time developed an approach to take advantage of this phenomena called, Model Output Statistics (MOS). (Glahn & Lowry, 1972) The MOS approach utilizes a statistical assessment of which forecast model handled the "Current Conditions" best in previous instances. (Antolink, 2021) Starting with a series of multiple forecasts from different models, these outputs are statistically compared to the past performance of each weather model under similar weather conditions. (Antolink, 2021) The multiple forecasts are then combined to

generate an ensemble forecast. (Antolink, 2021) The MOS approach focuses on identifying the best overall forecast statistically for each time a new forecast run completes. (Antolink, 2021) While the MOS approach has been successful for improving prediction of large area weather systems, the MOS approach has never been directed towards single point forecasts. (Antolink, 2021) Also of note for this praxis, MOS has not been applied for solar forecasting approaches as, "historically, these models have been optimized for predicting variables such as temperature, humidity, probability of precipitation, and wind." (Tuohy, et al., 2015)

2.5 Statistical Solar Weather Prediction

The Statistical forecasting section is organized in a manner to identify literature based statistical approaches to solar and weather forecasting. Statistical based forecasting approaches, while a newer development than NWP, have been under assessment for almost a half century. (Glahn & Lowry, 1972) Using statistical assessments of past and/or present conditions, these model approaches do not rely on an understanding of atmospheric physics. (Glahn & Lowry, 1972) The basis that enabled these statistical forecast approaches comes from improvements in computer processing power, advanced software packages, and sufficiently large sets of collected meteorological data. (Weyn & Caruna, 2020) Statistical models, once operational, utilize classification and/or regression techniques to generate forecasts. (Weyn & Caruna, 2020) These techniques enable future forecasting based on data from what previously occurred. (Ryu, Ito, Ishii, & Hayashi, 2019) There is not a specific statistical based approach for generating these types of forecasts, but instead a variety of approaches that have been developed to utilize a multitude of mathematical based techniques.

The Total Sky Imager (TSI) approach has been the most utilized approach for statistical photovoltaic forecasting. (Zhang, et al., 2014) TSI was developed to improve short term (5 to 20 minute) forecasts at a single location. (Zhang, et al., 2014) This approach takes real-time data from a camera and feeds it through a predictive classification tool (such as a cognitive neural network). (Zhang, et al., 2014) Once the current conditions are matched to a previously identified meteorological pattern (e.g., cloud shapes/direction, wind, or humidity) the future projection is made based on a best fit to the previous results. (Ryu, Ito, Ishii, & Hayashi, 2019) The TSI approach is reliant on live-sky images for data, so the benefit of this forecast approach has been limited to the time-range of the local optical sensor. This has typically been under 50 miles. (Rana, Koprinska, & Agelidis, 2016) This approach, however limited in range, has outperformed traditional numerical weather prediction approaches by three to four percent when underlying requirements have been met. (Rana, Koprinska, & Agelidis, 2016)

TSI-like statistical approaches have also been attempted by researchers for longer time-span multi-day forecasts. (Wang, Koprinska, & Rana, 2017) These multi-day statistical forecast model approaches have been designed utilizing partial or full artificial neural networks to leverage previous trends in historical collected data. (Zhang, Beaudin, Taheri, Zareipour, & Wood, 2015) These approaches are similar to the TSI in the statistical approaches underlying, but differ through the use of multiple sources of off-site raw collected weather data for training/operation. (Zhang, Beaudin, Taheri, Zareipour, & Wood, 2015) The use of statistical forecasting approaches have been successfully validated for multi-day time-scales in large-area solar forecasts with four years of historical surface collected data. (Shao, et al., 2016) These techniques, however, have

only appeared for 24-hour forecasts of small-scale forecasting. (Zhang, Beaudin, Taheri, Zareipour, & Wood, 2015) A major drawback to these approaches has been that for much of the world, data availability to support this TSI- approach is lacking (Zhang, Beaudin, Taheri, Zareipour, & Wood, 2015) or non-existent. (Bureau of Meterology, 2021)

Statistical approaches have been well-positioned to provide the next generation of improved short-to-medium term forecasts due to the use of advances in data collection and processing. The lack of widespread data collection of solar radiation, though, has posed a risk to the continued development of this type of model in remote locations. Based on the literature, this statistical method is seen to generate "the best intra-hour forecasts are provided in all climates by machine learning and classical statistics models". (Blaga, et al., 2019) Issues with remote data collection prevent solely statistical methods from being used in this praxis. Therefore, without relying solely on physics-based improvements, both Numerical Weather Prediction and Statistical approaches needed to form the basis for the advanced forecasting model developed under this praxis. Section 2.6 will discuss how this praxis leveraged these existing forecasting approaches to generate an improved combination one.

2.6 Approaches to Advance Solar Weather Prediction

The success of Model Output Statistics approaches in forecasting has pushed researchers to develop additional models that leverage these learning-based approaches using both Numerical Weather Prediction and Statistical models. (Makridakis, Spiliotis, & Assimakopoulos, 2018) These hybrid models have been designed for a variety of use cases. Three of the most apparent ones include real-time informed (El-Kenawy, Ibrahim,

Mirjalili, Eid, & Hussein, 2020), classification selection (Seo, 2020), and trained ensemble (Holmstorm, Liu, & Vo, 2016).

Real-time informed models are like the previously discussed Model Output Statistics approaches. (Zhang, et al.) (Glahn & Lowry, 1972) These real-time informed methods rely on feeding current condition weather data into a trained classifier model. (Zhang, et al.) Like the MOS approach, once run, the real-time informed models generate a likelihood probability weighting for each of the standard weather models based on the prior likelihood of it forecasting correctly. (Weyn & Caruna, 2020) (Zhang, et al.) These probabilities then directed either the single best model to be used or provided probabilities for use in an ensemble combination model (Zhang, et al.). The El-Kenawy based real-time informed approach requires both live data to feed the classifier, and at least four years of collected weather data based on the needs of analogous models. (El-Kenawy, Ibrahim, Mirjalili, Eid, & Hussein, 2020) For remote locations, data availability, both live and historical, is a major obstacle, and it is unlikely to support the real-time approaches. (Bureau of Meterology, 2021)

Another approach, the classification selection method represents machine learning based approaches which classify model outputs based on historical data. (Shao, et al., 2016) This classification approach first collects NWP forecast model outputs from multiple independent models. (Shao, et al., 2016) Then, this approach uses a classification scheme to identify the model that best predicts forward events from similar conditions. (Shao, et al., 2016) A primary requirement of this methodology is the need for access to multiple independent NWP forecast models and historical ground truth. (Shao, et al., 2016) In the development of these types of schemes, researchers have

assessed several different classifier approaches for use including: Perceptrons, Nearest Neighbor, SVMs, Decision Trees, and Neural Networks. (Seo, 2020) (Wolf, O'Donncha, & Chen, 2020). The Bagged Tree approach has been seen to perform well for forecast improvement models because it was designed to reduce variance, which has often been high in weather class data. (Tiwari, Sabzehgar, & Rasouli, 2018) This classification model approach, while accurate at times, has been prone to introduce classification error for locations with limited training data. (Shao, et al., 2016) Even with available data, this model approach has been seen as resource intensive because of the assessment and classification for each new forecast. (Tiwari, Sabzehgar, & Rasouli, 2018)

Trained ensemble forecast model developments represent one-time applications of regressive training of models. (Wilson VanVoorhis & Morgan, 2007) These approaches involve the combination of multiple independent forecast models based on the historical accuracy of each forecasting tool. (Holmstorm, Liu, & Vo, 2016) This type of combination was found to utilize both numerical weather prediction and statistical forecasting models – where available – to combine into a single output. (Holmstorm, Liu, & Vo, 2016) (Glahn & Lowry, 1972) These models typically utilize a regression approach and often combined using an optimization method designed to reduce the likelihood of forecast error. Applications of this type of model were routinely seen for non-solar weather forecast improvement. (Holmstorm, Liu, & Vo, 2016). This approach benefits from the ability to leverage real time data when available but supports falling back on only NWP models when not available. (Lange & Focken, 2008) This class of model, capable of utilizing 10 times less training data than a purely statistical approach, trades-off two to three percent performance in exchange for requiring less access to data.

(Wilson VanVoorhis & Morgan, 2007) (Makridakis, Spiliotis, & Assimakopoulos, 2018) (Rana, Koprinska, & Agelidis, 2016) This combination approach has been seen successfully deployed for both small scale and large-scale areas using both statistical and numerical underlying weather models but has not been used to improve solar forecasts. (Holmstorm, Liu, & Vo, 2016) (Wolf, O'Donncha, & Chen, 2020)

The solar forecasting tools presented in this section represent the leading class of approaches based on the combination of the best available approaches. Many of the research approaches rely on large quantities of live or historical solar measurements (Shakya, et al., 2018) or have been developed for improving forecasts over large areas. (Shao, et al., 2016) Local area offline forecast approaches using ensemble combinations have seen "the best performance" as compared to other novel forecasting approaches for applications with similar requirements to the problem of this praxis. (Blaga, et al., 2019) One limiting factor is that these have been for non-solar based forecast improvement (Wolf, O'Donncha, & Chen, 2020) or have been only for training data specific to a single location in Singapore. (Verbois, Huva, Rusydi, & Walsh, 2018) These research examples, while like what this praxis set out to find, all miss at least one factor relevant to the end user's operation.

2.7 Machine Learning – Ensemble Methods

Based on the research identified in Section 2.6, this literature review continued with an investigation into ensemble combination methods as the most likely route of success for this praxis. In this domain, literature has demonstrated the use of multiple machine learning approaches for generating regressions for the combination of multiple linear parameters. This has been done across a variety problem domains which have

utilized techniques including Recurrent Neural Networks (RNNs) (Hewage, Trovati, Pereira, & Behera, 2020), Fuzzy Neural Networks (FNNs) (Zhao, Guo, Su, & Zhao, 2015), and XGBoost (Ma, Yu, Qu, Xu, & Cao, 2020). These projects have utilized these ensemble methods to identify and forecast relationships in datasets. (Ouma, Okuku, & Njau, 2020) (Pan, 2018) These regression approaches have been used to demonstrate model combination research problems for both linear and non-linear relationships typically with datasets containing at least three years of historical data. (Kraemer, Palma, Braten, & Ammar, 2020) Alternatively, research has also shown weather forecast improvements that, "In combining forecasts, a simple average of the forecasts performs well, often better than more sophisticated models" (Gaba, Testlin, & Winkler, 2017)

A variant of the simple average combination that has been utilized in forecasting dating back to the beginning of the Model Output Statistics days was the multiple weighted linear combination approach. (Glahn & Lowry, 1972) This approach used applied weights to impact the linear relationship between independent model combinations. (Glahn & Lowry, 1972) These models were developed using an optimization approach that weighted (a_k) the factors for each model (X_k). (Glahn & Lowry, 1972) These models were then combined linearly to generate the final solution, in Equation 1. (Glahn & Lowry, 1972)

$$Result = a_1 * X_1 + a_2 * X_2 + \dots a_k * X_k \qquad (1)$$

A drawback of this approach was identified for problems without enough data – those with 50 or fewer independent samples. (Wilson VanVoorhis & Morgan, 2007) In these cases, the most optimal weights were not always solvable. (Verbois, Huva, Rusydi, & Walsh, 2018) For a linear regression approach, the threshold for sufficient data of

these problems has been assessed to be almost an order of magnitude smaller than factor-analysis approaches involving training raw data. (Wilson VanVoorhis & Morgan, 2007)

The weighted optimization approach has been utilized in multiple different forms. These different approaches could be classified as either static weighted value or online learning weighted value. A static weighted value approach provides a combination approach irrespective of the underlying data. (Gaba, Testlin, & Winkler, 2017) Like the "simple average" approach, these methods use different mathematical weightings on each model combined using a mathematical approach such as summed logarithmic. (Liang, Niu, Zhou, & Fan, 2018) The most basic weighted linear combination approach impacts the relationship between models but provides equal amplitude response. (Liang, Niu, Zhou, & Fan, 2018) A review of literature did not find any support to claim non-linear combination approaches were better or worse than simple average linear combination ones. (Gaba, Testlin, & Winkler, 2017) The key to static weighted approaches is that this type of approach has been sufficient as long as the models were, "averaged just across a few short and a few long calibration window[s]". (Hong, et al., 2020) The adaptive weighted value approach uses a different approach to calculate the weightings by adding each new data-point to the classification set over time. This non-static regressive model was seen to perform better for localized weather applications, but it required real-time training. (Eschenbach, et al., 2020) (Ulmer, Adam, & Eineder, 2012)

2.8 Forecast Error for Optimization

When using a combination model, a method to identify an optimal point in a quantitative space requires an error metric to mathematically train against. (School of Engineering and Applied Science, 2021) This review investigates approaches to

optimization of error to support the end user objective of reduced watt-per-square-meter error in solar energy forecasts. An assessment of error functions used in literature over a 25-year period found Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) as the most commonly utilized techniques for forecast models. (Botchkarev, 2018) MAPE, based on Equation 2 below, has often been used because of "its very intuitive interpretation in terms of relative error". (Myttenaere, Golden, Le Grabd, & Rossi, 2016)

$$MAPE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\sum_{i=1}^{n}|y_i|} \quad (2)$$

In an assessment of error in solar radiation cases, this approach was identified as problematic due to the frequent near-zero values. (Segarra, Du, Ruiz, & Bandera, 2019) This proportional based error calculation caused unintended identification and importance in weighting of small errors near sunrise/sunset instead of large daytime errors. (Segarra, Du, Ruiz, & Bandera, 2019)

In the case of MAE, based on Equation 3 below, this approach to error calculation assesses the sum of absolute differences between the measured values and the true value. (Myttenaere, Golden, Le Grabd, & Rossi, 2016) The MAE approach was the standard for assessing performance of a model in meteorological applications because each watt-per-square-meter off the true value was consistently weighted. (Segarra, Du, Ruiz, & Bandera, 2019) For applications that intended to penalize a specific response, this error approach would not be supported as all errors are equally weighted. (Botchkarev, 2018)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (3)$$

In the case of RMSE, based on Equation 4, this approach to error calculation is viewed as a method that under most situations should be avoided. (Botchkarev, 2018)

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(|y_i - \hat{y}_i|)^2\right]^{\frac{1}{2}} \quad (4)$$

This approach to error calculation penalizes large errors more than smaller errors. (Botchkarev, 2018) For meteorological applications, RMSE is known to calculate an error metric which is "suitable for evaluating the overall accuracy of the forecasts while penalizing large forecast errors in a square order." (Zhang, et al., 2014) This approach had been avoided in meteorological literature because this metric was, "not a good indicator of average model performance". (Chai & Draxler, Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature, 2014) For approaches with an individual large difference between one forecasted value and the measured value, the RMSE approach provides a stronger negative response to that error than a slight error spread across long periods of time. (Zhang, et al., 2014) Despite historical rejection of the RMSE approach, recent applications of machine learning optimization of meteorological problems have successfully relied on this error calculation approach to provide improved model results. (Shakya, et al., 2018) (Lorenz, Hurka, Heinemann, & Beyer, 2009) While the general literature consensus supports MAE and MAPE as preferred error approaches, the RMSE features which penalize large instantaneous errors are beneficial for applications which look to validate the model to improve short-term forecasts.

2.9 Machine Learning – Optimization Algorithms

With a defined objective function based on Sections 2.7 and 2.8, to solve the optimization problem, this literature review assessed potential algorithmic approaches. In

solving optimization problems, two types of approaches are evident: Local Optimization and Global Optimization approaches. (MathWorks, 2020) Local optimization techniques solved for the nearest apparent minimum while global approaches solved for the absolute minima of a problem-space. (Horst & Tuy, 1990) Based on a requirement from the end user of this praxis, this literature review was limited to methods native to the Mathworks MATLAB software suite. This included seven approaches to optimization. (MathWorks, 2020) These approaches include; Exhaustive Search, Interior-Point Search (fmincon), Adaptive-Mesh (patternsearch), Genetic Algorithm (ga), Particle-Velocity Evolution (particleswarm) Interpolate Objective Function (surrogateopt), and Scatter-Search (GlobalSearch). (MathWorks, 2020)

The Exhaustive Search optimization method is an approach that 'brute forces' the solution based on assessing the objective at every feasible point. (Nievergelt, 2000). Implementation of exhaustive search algorithms have been problematic due to the digital approximation manner which computers store continuous numbers as rounded values which results in a pseudo-continuous set. (Kornerup, Lauter, Lefèvre, & Louver, 2010) In MATLAB, numbers are approximated using 64-Bit double precision floating points which limits precision. (Floating-Point Numbers, 2021) This limitation has caused the exhaustive search to not be exhaustive, but an interpolated one. (Sawtooth Software, 2021) This approximation, while still solving the objective, lowers the accuracy of the optima found. (Klein & Zachmann, 2005) In addition to this accuracy issue, Exhaustive Search approaches have been reviewed to be time and resource intensive. (Wilson, Harding, Hoeber, Devillers, & Banzhaf, 2012) Exhaustive searches have been sped up using expensive parallelized Graphics Processing Units (GPUs), but these speed-ups have

only managed to match the efficiency of Genetic Algorithm approaches discussed later. (Wilson, Harding, Hoeber, Devillers, & Banzhaf, 2012) When utilized on GPUs these algorithms have not seen any improved optimization performance over other algorithmic approaches. (Wilson, Harding, Hoeber, Devillers, & Banzhaf, 2012)

The remaining six optimization approaches utilize self-contained functions from within the MATLAB Global Optimization Toolbox. (Mathworks, 2021) These approaches covered a variety of machine learning techniques ranging from evolutional approaches to pattern-search ones. (MathWorks, 2020) Three of these approaches (Interior-Point, Adaptive-Mesh, and Particle-Velocity) were identified as only solving for the nearest minima in a local optimization problem. (MathWorks, 2020). These local minimization approaches do not find a lower optimal answer on the other side of a closer but less optimal solution. (MathWorks, 2021) The other three MATLAB functions (Genetic Algorithm, Interpolate Objective 'surrogateopt', and Scatter-Search 'GlobalSearch') were all designed to identify the global optimal solution of a problem irrespective of any local minima. (MathWorks, 2020) In comparing the performance of these three approaches, all three methods were reviewed to successfully identify problem optima accurately. (MathWorks, 2020) From a speed perspective, the surrogateopt and GlobalSearch approaches were identified as faster than the Genetic Algorithm, however the surrogateopt approach was known to vary widely on a problem dependent basis. (MathWorks, 2020) In assessing the scope of optimization problems supported, the Genetic Algorithm and GlobalSearch approaches supported both linear equality and inequality constraints. (MathWorks, 2020) The surrogateopt function, which was

designed for optimizing non-linear objective functions, however only supported linear inequality constraints. (MathWorks, 2020)

Of the two algorithms that have supported similar problems, the Genetic Algorithm approach was designed as a "probabilistic search […] to work on large spaces". (Goldberg & Holland, 1988) For large area forecast optimization, this design feature has made the GA a commonly used approach in Weather Forecast optimization. (Araki, Shimadera, Yamamoto, & Kondo, 2014) (Verbois, Huva, Rusydi, & Walsh, 2018) In the area of forecast optimization, the Genetic Algorithm approach has worked successfully as an optimization method for weather forecast model combination approaches in the air quality monitoring domain. (Araki, Shimadera, Yamamoto, & Kondo, 2014) In this domain, the Genetic Algorithm approach optimized the problem by generating random starting points, running the Genetic Algorithm, and then iterating until the end conditions (threshold) were crossed. (Araki, Shimadera, Yamamoto, & Kondo, 2014) (MathWorks, 2021) Despite being designed to work for large problem-spaces, the Genetic Algorithm has required additional function calls to optimize when compared to alternative approaches. (MathWorks, 2020) Each additional function call increases the run-time required to complete the optimization, and therefore the Genetic Algorithm approach has been known to be slower than other similarly performing optimizations – like GlobalSearch. (MathWorks, 2020)

The GlobalSearch approach uses an iterative approach to optimization using multiple starting points. (MathWorks, 2020) For each starting point, this approach used the interior-point method, one of the most common optimization algorithms. (Mandi & Guns, 2020) The interior-point based approach was designed to use computationally

inexpensive gradient based descent equations which iterates over new values towards the local minima until the descent thresholds were met. (Plaxco, Valdes, & Stojko, 2021) Individual interior-point algorithm optimizations were identified as practical for optimization in weather forecasting applications due to the "problem size and frequency of solution". (Janjic, Ruckstuhl, & Toint, 2019) Each individual run of the interior-point algorithm only supported the identification of a local minima, which is why the GlobalSearch algorithm utilized a "multi-start" approach with a series of starting points. (MathWorks, 2020) The GlobalSearch algorithm, when compared to the other seven solvers, tied for best optimization, but at a faster speed. (MathWorks, 2020) (Martí, Laguna, & Campos, 2005) Table A-1 in the Appendix displays the results of a direct comparison of the two different algorithms, which provided this conclusion. (MathWorks, 2020)

2.10 Model Validation

The value of a forecast model is tied to the perceived accuracy from its users. (Fader & Hardie, 2005) In this literature assessment, a survey is performed to identify approaches for validating forecast model improvements. In meteorology, statistical assessments have been a well-accepted method for validating the performance of a forecast. (Hamill, 1999) The two-sample t-Test "has been the most popular" statistical validation technique used in climate model testing. (Decremer, Chung, Ekman, & Brandefelt, 2014) But for assessments containing more factors, one way ANOVA has been used in atmospheric science assessments to identify which factors are the statistically the most important. (Siuta, West, & Stull, 2017)

To perform an ANOVA assessment, three rules need to be met: 1) Population Normality 2) Equal Variance Distributions, and 3) Independence of Data. (Eberly College of Science, 2021) For forecasting applications, the ANOVA test has been identified as a valid test for this assessment because these requirements are met: 1) Meteorological data is generally normal (Physical Sciences Labratory, 2021) 2) Variances have not been found to violate the variance requirement (Beersma & Buishand, 1999), and 3) Meteorological data samples can be considered independent. (Allan, Pereira, Raes, & Smith, 1998) If a data set satisfies the ANOVA requirements and the data is continuous, like meteorological data, (National Weather Service, 2021) the set also satisfies the requirements to be able to perform a two sample t-Test. (NCSS Statistical Software, 2021)

In addition to statistical approaches to calculate error, the use of raw calculation approaches has been acknowledged as important in identifying underlying meteorological patterns. (Versta Research, 2021) Uncertainty in the measurement or model can cause systematic error which will cause negative impacts on the statistical error calculations. Assessment of the distribution of individual performing runs can identify trends or biases that the statistical averages might not. (Meacham, et al., 2011) Using matched samples, the assessment should compare specific classes of inputs. (Gerstman, 2021)

2.11 Summary and Conclusion

This literature review has been completed to understand the research landscape in support of the development of an improved localized solar forecasting tool for independent remote locations without access to real-time data. The background literature survey covered historical solar energy production, operational forecasting approaches,

and novel forecasting proposals. Based on the gaps identified from these areas, the literature survey continued to assess methodologies which could pose a potential technical solution. These areas included machine learning based ensemble generation, error calculation, and optimization techniques.

In assessing the background of solar energy, a recent trend in the increased deployment of solar energy generating capacity was identified. (US Department of Energy, 2019) This trend was identified as leading to solar energy being a primary electric generation source in remote locations. (Taylor, Ralon, Anuta, & Al-Zoghoul, 2020) The risks to this deployment as identified in literature were that for remote independent locations, existing model designs experience an error rate double that of larger areas. (White, et al., 1999) These forecast errors have increased electricity production costs by up to 35%, (Georgitsioti, et al., 2015) which result in solar investments not being made. (Nagendra, Narayanmurthy, Moser, & Singh, 2020)

Based on this literature assessment of small area forecast error, the literature survey shifted to identify historical trends in solar forecasting and methods which improved forecasts. The literature reviewed identified two primary trends in the improvement of operational forecast models. The first - Numerical Weather Prediction (NWP) used physics-based calculations for predicting future conditions. (National Ocean Service - U.S. Government, 2021) The second - Statistical Models are more recent developments which rely on matching data trends. (Glahn & Lowry, 1972) In the literature, NWP models were identified the predominant approach to solar energy forecasting. (National Ocean Service - U.S. Government, 2021) These physics-based models have been assessed to have issues with minor deviations in inputs, such as un-

modeled nearby terrain, having major deviations in forecast accuracy. (Jankov, Gallus Jr., Segal, & Koch, 2007). The other identified literature class of models, statistical models, have been found in literature to outperform NWP approaches. (Rana, Koprinska, & Agelidis, 2016) These models, however, rely on short time-periods (Zhang, et al., 2014) and can require significant quantities of recorded data and access to live data. (Zhang, Beaudin, Taheri, Zareipour, & Wood, 2015) For locations with limited access to historical data or live data, these approaches have been unreliable. In response, researchers have developed a variety of ensemble and offline statistical based approaches to improve solar forecasting. (Tuohy, et al., 2015)

The recent identified trends in model development have focused primarily on ensemble combinations of models. These approaches either select a single model to be used or provide probabilities to be used in an ensemble combination. (Zhang, et al.) These approaches, while focused on improving local forecasts, often had drawbacks for the applicability in remote locations. One relevant example utilized principal component analysis, but it required a large dataset in a single non-movable location. (Verbois, Huva, Rusydi, & Walsh, 2018) Another research project used regression for training multiple model ensembles, but it was for non-solar forecast improvements over large areas. (Wolf, O'Donncha, & Chen, 2020) Based on the breadth of the survey, for a remote single location without access to real-time data, an ensemble combination of numerical weather prediction models was identified as the best approach, but one which had not been implemented. (Wilson VanVoorhis & Morgan, 2007) (Makridakis, Spiliotis, & Assimakopoulos, 2018) (Rana, Koprinska, & Agelidis, 2016)

Based on this identification, the literature survey turned to assess different ensemble approaches, error calculation methods, and optimization methods needed to generate this type of ensemble forecast. In reviewing a variety of combination approaches in literature, many linear and non-linear combination models were identified as applicable for similar forecast model improvements. (Ouma, Okuku, & Njau, 2020) (Pan, 2018) Based on the complexity of weather phenomena, the weighted linear combination approach was assessed in literature, however, as having the best localized weather and solar forecasting performance. (Eschenbach, et al., 2020) (Ulmer, Adam, & Eineder, 2012)

Based on this assessment, the review shifted focus to determine the best approach to identify weights for a linear combination. In the literature, three types of error calculations were identified as relevant to meteorological model assessments: Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). (Botchkarev, 2018) While not the most common, the Root-Mean-Square-Error was identified as the most relevant for reducing forecast outliers. (Botchkarev, 2018) (Segarra, Du, Ruiz, & Bandera, 2019) (Zhang, et al., 2014) This error function was the foundational function needed to optimize the weights against.

Based on other forecast improvement research, an optimization approach would be needed to identify the regression parameters for ensemble. (Holmstorm, Liu, & Vo, 2016) This review assessed a variety of methods contained within the Mathworks MATLAB Global Optimization Toolbox to calculate weights in an ensemble equation. Of these algorithms, literature identified that half of the algorithms in the toolbox support only local optimization methods. (MathWorks, 2020) These methods, which can get

stuck in a local minimum, have not been useful when the global minimization is needed. (MathWorks, 2020) The remaining three algorithms, Genetic Algorithm, Interpolate Objective, and Scatter-Search all have been assessed as approaches to find global optima. (MathWorks, 2020) In assessing these three, literature reviews identify all three as having equal optimization success. The Genetic Algorithm was seen to perform slower than the rest, while the Interpolate Objective approach does not support linear programming problems with equality constraints, which were needed. (MathWorks, 2020)

This praxis is based on the predicate that an improved forecasting tool is needed. The survey of literature has identified significant developments in solar forecasting and weather forecasting which can be applicable to the development of this tool. The review of literature has identified a gap in research for an approach to generate localized solar energy forecasts for remote locations that lack real-time solar radiation data. The bulk of research in solar forecasting has been seen to support wide area forecasts, statistical approaches requiring real-time data, and forecast improvements for non-solar forecasting needs. The prevalence of forecasting errors for these remote deployments, though has the potential to increase costs for deployed systems to exceed baseline estimates by up to 35 percent (Georgitsioti, et al., 2015)

Chapter 3—Methodology

3.1 Introduction

"All models are wrong, but some are useful". (Box, n.d.) In forecasting realms, one of the most commonly held beliefs is that forecast models provide useful information only if you understand the model underlying phenomena. With experience, a forecaster can leverage this prior intuition to identify which models will likely perform best for a particular location. The solar energy forecast tool developed in this praxis is designed to automate this approach using analytical methods to improve accuracy of forecasts for remote single point-locations. This tool takes historical third party forecast model outputs and measured ground truth weather data as inputs. The tool then uses multiple algorithms to combine these independent models in an optimal way that improves forecast accuracy. From a foundation perspective, if this model improves the accuracy of local single point solar forecasts, the praxis methodology should validate the proposed hypotheses.

This chapter contains a description of the research methodology for this praxis. This methodology forms the architecture of the praxis research product and contains justifications for why this model is sufficient for the end user problem. A notional diagram of the underlying model is detailed in Figure 3-1. The chapter begins with a description of the functional architecture and data required to run the model. The focus then shifts to detail and justify the methodology. This includes the data subset, analytical combination, and analytical optimization methods selected. The chapter then turns to a description of the operation of the tool before concluding with an example of what the model looks like to the end user.

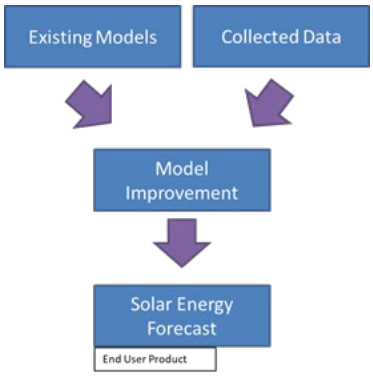


Figure 3-1. Fundamental Research Methodology

3.2 Research Tool Functional Architecture

Based on existing research methods identified in Chapter 2, this praxis model required a niche set of requirements. These requirements were not supported by existing developments in the forecasting trade-space. The methodologies and architecture presented in this chapter address the combination of analytical methods selected for use to fill this gap based on combinations of approaches from the literature. This developed research product leverages weighted linear combinations and machine learning based linear programming optimization techniques from prior reviewed research. The developed tool requires two process flows for operation, as depicted in Figure 3-2. The first flow follows the purple arrows. This "Operational Use – End User" flow is the day-to-day use of the research tool. For this, the research tool collects input forecasts from existing Weather Research and Forecasting (WRF) models. From there, these models are fed through the research combination model and generate enhanced forecasts. Before "Operational Use", "The Setup – Training Regime" flow that follows the orange arrows is required. This process is run only once for each new location of interest. To run, this process takes historical data and generates optimal weights that minimize forecast error for the combination of WRF model forecasts at the location of interest.

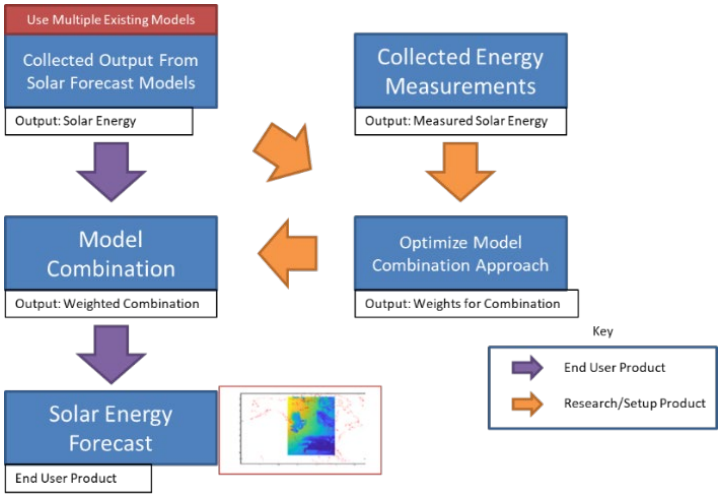


Figure 3-2. Detailed Research Methodology

For the Setup – Training regime, a set of historical model predictions and resulting ground truth are fed through a weighted combination optimization method. This method calculates how accurate the forecast has been and adjusts weights to each forecast model. The weights are adjusted in a manner to identify the ones that minimize forecast error for the resulting combination of models. The "Operational Use" method then takes those weights and combines current model outputs using a weighted sum approach to generate enhanced solar prediction forecasts. These two processes, when combined under this praxis's MATLAB tool development, were designed to improve solar energy forecast accuracy. The following sections detail the design and operation of the forecasting tool.

3.3 Model Environment & Data Ingestion

To support the end user organization's operational requirements, several computer environment restrictions were placed on the development of this praxis's enhanced forecast generation tool. This section presents a description of both the computing environment and the enablers and limiters of the process to ingest data for processing. At a high level, the end user required that all tools which interface with existing products

must run on the Ubuntu 18.04 Linux operating system and be fully integrated into the Mathworks MATLAB R2019a suite. The operating system and software development requirements both enable and restrict options for the research tool foundational blocks in Data Collection & Generation, Weighted Linear Combination, and GlobalSearch Optimization. Appendix B provides details and links to all pre-requisite software installations and data sources used in this praxis's enhanced forecasting tool.

The first block; Data Collection & Generation is designed to provide the training method with historical forecasts and ground truth measurements. WRF model initialization and ground truth data are readily available through both The National Renewable Energy Laboratory (NREL) and National Center for Atmospheric Research (NCAR). (NCAR, 2020) (National Renewable Energy Labratory, 2021) Specific sources utilized in this praxis are found in Appendix B. To support the download of these files for use in the praxis, Ubuntu's WGET function is leveraged. WGET supports downloading comma separated value (CSV) and American Standard Code for Information Interchange (ASCII) formatted files which contain ground truth and limited historical models. At the same time, WGET supports Hypertext Markup Language (HTML) files which can include historical model initialization files to generate expired forecasts.

The design for the praxis research tool can support the use of both the Weather Research and Forecasting (WRF) model framework, and it can support both pre-run and newly generated forecasts from other models. Based on availability, the primary model integrated is the WRF framework model. The WRF model is selected for use due to the open-source code access, support of locations in the USA and Australia, resolution, and ability to be run on a standard laptop computer. (NCAR, 2020) Other models were

considered, including the European Center for Medium-Range Weather Forecasts (ECMWF), but do not support open access for data or processing model runs. In operation, WRF is natively supported in Ubuntu 18.04. Running it, however, requires additional software pre-requisites to be installed. These include MPICH for use in the parallelization of runs to speed up the number of data points that could be collected, (High-Performance Portable MPI, 2021) NETCDF for use in Encoding/Decoding the scientific input/output data, (UCAR, n.d.) and other low level programming packages needed to compile including Jasper, linpng, and zlib. (UCAR, n.d.) Once pre-requisites are installed, the National Center for Atmospheric Research's WRF User Guide's process is followed to build and run the model. (UCAR, n.d.) With WRF setup complete, the model can generate historical forecasts to use in training and operation of the praxis developed model. Section 3.5 will detail the specific requirements for model input data.

The final portion of the data input block is the data alignment method. Once forecasts and truth data are generated/collected, it needs to be loaded into the MATLAB processing environment. Using MATLAB's native file read functionality for both text-based (fscanf) and NETCDF-based (ncread) files, the historical data and forecasts are read into time-location-model 3-dimmensional data matrix structures. These data structures will be utilized by the linear combination weighting and GlobalSearch optimization methods which will be discussed in Section 3.6 and 3.7, respectively.

3.4 Data Inputs

For the purpose of calculating the set of linear combination weights as part of the praxis developed model, both an input and validation set of data are needed. For these two sets, both input historical model forecasts and ground truth measurements are

required. The historical model inputs take the form of WRF model initialization files. These files are time-definite and provide initial condition estimates that are structured to support the needs of WRF to generate a forecast. The ground truth measurements are measured resulting conditions time and location matched to the generated model forecasts. This praxis set out to improve solar irradiance forecasting using historical truth measurements from solar sensors, therefore solar radiation measurements are required as inputs. However, for locations where solar collections are lacking – as detailed in Section 1.8 Research Limitations, this praxis turns to temperature measurements.

On the historical model front, the praxis research tool is developed to utilize the latest WRF 3.8 model. Leveraging the WRF model open framework which contains at least $2x10^{19}$ different independent model combinations, a sub-set of models which each select a combination of the relevant underlying mathematical assumptions and/or equations is pulled forward. Each parameter choice represents a different approach to calculating the propagation of energy or moisture through the environment. (NCAR, 2020) Based on a review of the associated parameters, three were defined as processes associated with the prediction of solar energy: mp_physics [Cloud Formation], ra_lw_physics [Longwave Solar Energy], and ra_sw_physics [Shortwave Solar Energy]. Further detail on these underlying WRF models can be found in Appendix B. Within each of these, multiple calculation approaches exist, and details about the selection of which specific parameters have been used in particular models is found in Section 3.5.

To generate each of the input forecasts, each of the selected WRF models are processed with an initialization dataset. These datasets are generated by meteorological organizations around the world at least 4 times per day. Using a common GRIB2 file

standard, these initialization files are gridded binary files containing assessed conditions throughout the surface and atmosphere. An example of this machine readable format is; "50:33939039:d=2017072501:CNWAT:surface:an1". Multiple types of GRIB 2 initialization files exist which differ in underlying meteorological data sources selected to be integrated. (NCAR, 2020). This praxis uses the GFS based GRIB 2 files which are known as the "most commonly utilized parameters". (NOAA, 2021) The GFS based files support the two needs for praxis assessment – significant stores of historical initialization files to generate 'archive' forecasts and operational real-time ones to test against.

The training approach which this praxis utilizes requires a set of both the model and truth data for solar radiation and temperature to generate forecast model weights. This truth data must be temporally and spatially matched to the models generated in order to optimize combined forecast weights. Truth data for this praxis is drawn from the Surface Observations dataset at the National Renewable Energy Laboratory's Solar Baseline Measurement System (BMS). This database contains 200+ locations around the continental USA which have recorded surface observation data like the data in Table 3-1. In addition to this recorded data, the BMS contains an API which returns interpolated surface conditions based on the data collected at those locations along with satellite derived observations. This database provides both solar radiation and temperature measurements which are accessible across a span of at least 2004-2021 for most locations across the continental USA. (National Renewable Energy Labratory, 2021)

Table 3-1. BMS Surface Observation Example

Year	Month	Day	HH	MM	SS	Temp (C)	Radiation (w/m^2)
1998	12	28	13	00	65	N/A	874

(National Renewable Energy Labratory, 2021)

3.5 Data Selection

With large quantities of available data, the praxis research tool needs a sub-set of data to be able to run and assess performance. As historical and near real-time meteorological data is widely available across large areas, processing resource limitations are often encountered in meteorology. (NOAA, 2021) This section details the process and justification behind which specific datasets and quantities are selected for use in development, testing, and operation of the praxis research tool. These data selection assessments are performed to identify the optimal number and type of independent WRF forecast models, number of model initialization files, and number of locations/times for truth/validation. Specific standards and assessment or rules of thumb are addressed.

As addressed in Section 3.4, the WRF model is a framework that is available in multiple pre-defined variants built upon a series of underlying equations. (NCAR, 2020) Each combination of different underlying models represented a potential run-case of a different model. In analyzing the potential subsets of the three WRF parameters identified in Section 3.4, two options per parameter are selected for use because each one originates from either the current or future operational forecast models from the National Weather Service. (NCAR, 2020) Due to the limitations denoted in Section 1.8, additional WRF underlying physics models are not considered because these models are not operationally used ones. Based on this, the full replicate of the six selected parameter-model combinations provides the independent input models for this praxis. A blocked partial run-set is not considered for this tool as the available run-time for the full replicate is supported. The description of parameter options selected is found in Table 3-2 with the mapping of underlying model to research WRF version in Table 3-3.

Table 3-2. WRF Solar Parameters

Parameter	NWS Status	ID	Description
mp_physics	NMM Tested - NEW	6	WSM 6-class graupel scheme
	NMM Operational - OLD	5	Ferrier scheme
ra_lw_physics	NMM Tested - NEW	1	RRTM scheme
	NMM Operational - OLD	99	GFDL scheme
ra_sw_physics	NMM Tested - NEW	1	Dudhia scheme
	NMM Operational - OLD	99	GFDL scheme

Table 3-3. WRF Run Table

Parameter	WRF-1	WRF-2	WRF-3	WRF-4	WRF-5	WRF-6
mp_physics	6	6	6	5	5	5
ra_lw_physics	1	1	99	99	1	1
ra_sw_physics	1	99	1	99	1	99

To generate an input model set for these WRF variants, a subset of the NOAA National Centers for Environmental Information "NOMADS" data identified in Section 3.4 is required. To accurately assess the praxis hypothesis, the amount of initialization data for the WRF model needs to be pre-planned. The questions of 1) how frequent and 2) how many need to be answered for the dataset containing all testing and training initialization files. For this praxis, to meet the rule of thumb requirements, 219 different initialization files are required with a four-day spacing. The rule of thumb here is based on the minimum sample size for a machine learning problem of 50. (Voorhis & Morgan, 2007) A total of 216 initialization files using a 70/30 split between training and testing results in 50 Testing and 166 Training files. Using the average error of solar radiation measurements at 0.59% (Vignola, et al., 2016), an additional 1% of data is added to account for potential filtered/rejected data. Data for this tool is only rejected if the dates do not have files available. Because the forecasts are operational ones, any outliers or "bad forecasts" are a figment of the process this tool trains to avoid. This results in 219 total initializations and associated model-start times.

In assessing the timespan between input files, an approach based on sampling frequency of continuous signals was selected. In literature, it has been shown that weather patterns can be reconstructed from sparse data with less than 10% of the rate. (van Bussel, et al., 2016) However, using sampling theory sets a stronger data requirement which requires data at 200% of the frequency of the pattern. This 200% frequency is applied to the weather cycles of interest for the end user; diurnal (day/night), mesoscale (week-long), and annual (single year). While the sun also has an 11-year solar cycle, the difference in energy of up to 2-watts-per-square-meter does not have a meaningful impact when the delta of other weather impacts is up to 1367 watts. (NASA, 2021) Therefore, these are the weather cycles of interest. To identify the inputs that would meet this requires looking to the model runs themselves. As each WRF model run is for 96 one-hour forecast resolution, the diurnal needs are natively met with more than one forecast every 12 hours (1200% frequency). On the annual side, to meet the 200% ratio the forecasts should cover at least two years. Lastly, for mesoscale features which occur in seven-day patterns, these require at least 2x samples per week. With 96 hours of forecasts per week, a minimum of two samples exists per week with at least one forecast. Based on this, a dataset required would contain 2 years of once a week 96-hour long forecasts. For this praxis, a data set was picked at random using 0Z (Midnight in the UTC Zone) initializations every seven days for 1 degree rated GFS Files from October 1, 2005 through December 27, 2009. These initialization files would be downloaded from the National Center for Environmental Information and is sufficient for the model.

The last data set this praxis research requires is the ground truth data. In selecting ground truth locations, the minimum number of locations needed to satisfy basic

statistical tests is 30. (Sedlmeier & Gigerenzer, 1997) To assess the hypothesis, data must be time and location matched to the model forecasts. Using the National Renewable Energy Laboratory (NREL) BMS API, the data download is limited to approximately 2,401 locations based on the API limit of 5,000 (location/time) requests per day. Based on the desire for continental USA coverage for hypothesis testing, the 49 by 49 grid was set in 100km spacing with a center point of NREL Colorado, as shown in Figure 3-3. This dataset provides at least 30 samples from each of the climatic schemas desired. (Desert, Coastal, Plains, and Mountain) Section 3.6 details the development and features of this underlying model and uses of this data.

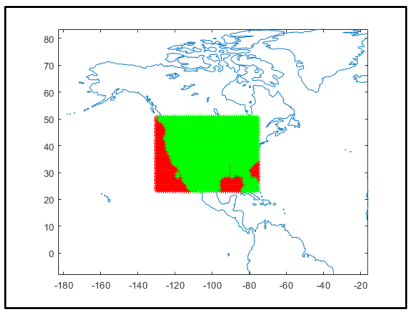


Figure 3-3. WRF Modeled Area (Red), WRF + Ground Truth (Green)

3.6 Combination Model

The combination model forms the foundation of both the training and operation modes of this praxis research. Based on a similar approach to the Model Output Statistics (Glahn & Lowry, 1972), this methodology combines outputs from multiple independent forecast models. The multi-criteria decision analysis approach uses a weighted linear combination that leverages the outputs from multiple WRF solar forecasting models in a weighted manner. This combination model, based on Equation 5 takes the output from each of the six WRF models (WRF_n) and multiplies that by an importance value (W_n)

before summing the resulting values. The weights of the independent models are designed to promote or demote the importance of each independent model based on past accuracy. This equation is used as an objective function to identify weights in the training mode and then used to generate forecasts in the operational mode.

$$F_{Combined} = W_1 * WRF_1 + W_2 * WRF_2 + W_3 * WRF_3 + W_4 * WRF_4 + W_5 * WRF_5 + W_6 * WRF_6 \quad (5)$$

The training mode, as depicted in Figure 3-4 is the setup operation of the weighted linear combination. As each underlying WRF model has different accuracy implications across geographically dispersed locations, (Kim, Kim, Kang, & Yun, 2018) the weights (W_n) are calculated for each location of interest. The weights, representing a fractional importance of the model are set to values between 0 and 1. Error is assessed for the resulting historical forecasts and measured ground truth. This model is the basis for the optimization approach selected to calculate these weights in Section 3.7.

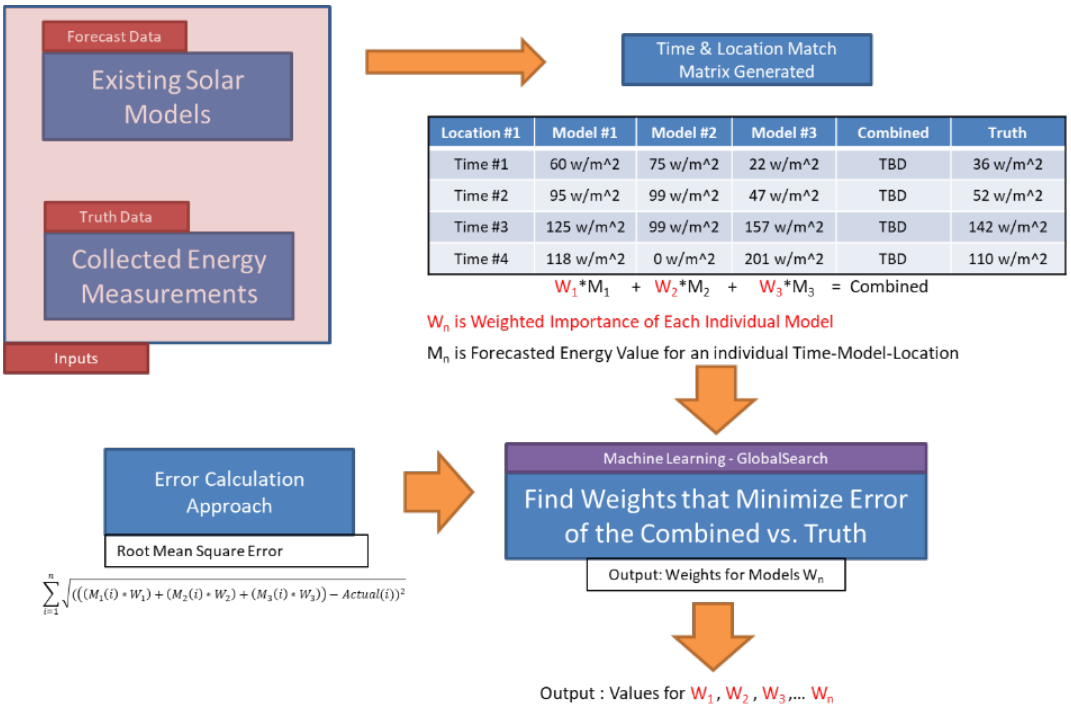


Figure 3-4. Combination Model - Training

The operational mode, as depicted in Figure 3-5 is the day-to-day use mode for the weighted linear combination model. Using the weights calculated in the training method, this mode generates new forecasts using current initialization data and calculated weights. The output from this model is a single forecast in both tabular graphical formats.

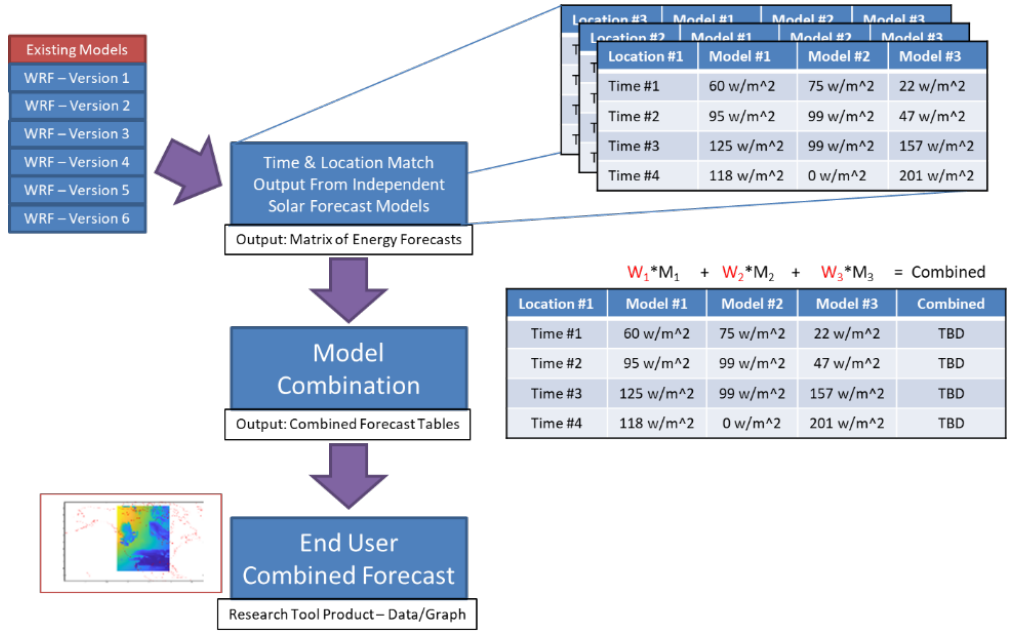


Figure 3-5. Combination Model – Operation

3.7 Optimization Model – Training & Testing

The identification of optimal weights for use in the above defined combination model requires a numerical optimization approach. This optimization method would utilize the set of historical forecast models and measured truth data to calculate weighted value which minimize combined forecast error for each specific location of interest. Based on the literature assessment in Chapter 2, the optimization algorithm utilized in this praxis model is MATLAB's GlobalSearch algorithm. The GlobalSearch algorithm is a machine learning based optimization algorithm designed to solve Linear Programming problems. This method uses a scattered search approach where multiple starting points

are identified, and a local Interior-Point based optimization is performed to identify multiple local minima. (MathWorks, 2020) Using a scoring function, this algorithm evaluates points on a gradient descent until the algorithm goal is met, and an optimal solution is identified. (Glover, et al., 2011)

The MATLAB version of the GlobalSearch interior-point based optimization method starts with a random selection of multiple starting trial points within the scope of the problem defined equality/inequality constraints. Once the initial points are identified, this algorithm uses two methods to solve - factorization and gradient descent. (Agnarsson, Sunde, & Ermilova, 2013) To use these methods, the optimization problem must have objective and constraint functions that are both continuous with continuous first derivatives. (MathWorks, 2021) Additionally, all values must be real value in both the constraint function and objective function. (UC Berkeley EECS, 2021) With these standards met, the factorization and gradient descent is performed using the fmincon (interior-point) solver. (Agnarsson, Sunde, & Ermilova, 2013) This solver is designed to take the function of interest (f) and the constraints/slack variables (s) and solved for the minimization based on the equation below; (Agnarsson, Sunde, & Ermilova, 2013)

$$\min_{x,s} f_{\mu}(x,s) = \min_{x,s} f_{\mu}(x) - \mu \sum_i \ln(s_i) \qquad (6)$$

The solver attempts to directly solve using Cholesky factorization (Agnarsson, Sunde, & Ermilova, 2013) which leverages symmetry and definiteness through matrix transformation (Higham, 2009), but switches to a conjugate gradient descent method when the data locally around the point is not convex. (Agnarsson, Sunde, & Ermilova, 2013) The gradient direct step method solves for the matrix of equations in order to

identify the change in step size and change in objective function using a linearized Lagrangian of the equation below where $A * B = -C$. (7) (MathWorks, 2021)

Table 3-4. Lagrangian Matrix Operations

Matrix A				Matrix B	Matrix C
H	0	J_h^T	J_g^T	Δx	$\nabla f + J_h^T y + J_g^T \lambda$
0	$S\Lambda$	0	S	$S^{-1}\Delta s$	$S\lambda - \mu e$
J_h	0	0	0	Δy	h
J_g	S	0	0	$\Delta\lambda$	g+s

Note: See symbols section for description of the above symbols

For each local minimization calculation across each of the random starting points a score is generated consistent with the value of the objective function at that point. (Agnarsson, Sunde, & Ermilova, 2013) This score is based on equation 8 where g(x) are the inequality constraints and h(x)=0. (MathWorks, 2021) For the output of each updated starting point, another round of optimization is run if the point is not in/near a minima, the point has a lower score than the threshold, and the point satisfies function constraints. (Agnarsson, Sunde, & Ermilova, 2013) Once the threshold values are passed and the rate of change in score is below minimum threshold or the maximum number of iterations is passed, the best point is declared optimal. (Agnarsson, Sunde, & Ermilova, 2013)

$$f_{\mu}(x, s) + v(h(x), g(x) + s) \quad (8)$$

The objective function used in this GlobalSearch algorithm is an error function based on the deviation between solar forecast and measured value. The equation used in this praxis tool is based on the Root-Mean-Square-Error (RMSE) of the weighted combination. This equation, number nine below, takes the square root of the average of the sum of the squares of differences between actual and forecasted values. This

approach, while not the standard for meteorological error calculations benefits this particular use case draws from. Description of RMSE is presented along with Equation 4.

$$\left[\frac{1}{n}\sum_{i=1}^{n}(|y_i - \hat{y}_i|)^2\right]^{\frac{1}{2}} \quad (9)$$

The RMSE approach utilized here penalizes single large errors more than small average errors. For this use case, because the end user is deploying electrical generating capacity having a large instantaneous error in prediction risks insufficient supply of electricity, causing a loss of data recording continuity, a major issue. The objective function values in (y_i) and ($\hat{y}_i$) are the weighted linear combinations.

In addition to the error function, the optimization problem requires constraints. These constraints bound the scope of the weights being calculated. In meteorological applications, ratios are often used to describe relationships between factors. (Mixing Ratio, Cloud Fraction, Relative Humidity) Based on these ratios, and the linearity of solar energy, the weights are set to be between 0 and 1 through inequality constraints. These constraints restrict the available settings of the weights to be greater or equal to zero but less than or equal to 1. The last constraint is set with respect to the total sum of the weights. Because each independent model is already defined based on what should be 100% of the forecasted solar energy, the sum of these models (and therefore sum of the weights) should equal 100%. To accomplish this, an equality constraint is defined as the sum of the weights equal to 1. Based on the objective function and constraints, the following is defined as the optimization problem. This equation, Equation 10, provides the application of the minimization of the error function, by combining the Equation 1 and Equation 4. Below that, the Equality and Inequality constraints are defined.

$$\min f(W_1,\ W_2, W_3, W_4, W_5, W_6)$$

$$= \sum_{i=1}^{n} \sqrt{\left(\left((M_1(i)W_1) + (M_2(i)W_2) + (M_3(i)W_3) + (M_4(i)W_4) + (M_5(i)W_5) + (M_6(i)W_6)\right) - Act(i)\right)^2} \tag{10}$$

$$\text{subject to} \quad \begin{gathered} W_1 + W_2 + W_3 + W_4 + W_5 + W_6 = 1 \\ W_1, W_2, W_3, W_4, W_5, W_6 \geq 0 \\ W_1, W_2, W_3, W_4, W_5, W_6 \leq 1 \\ M_1, M_2, M_3, M_4, M_5, M_6; Matrix\ of\ Static\ Values \\ Act(i); Matrix\ of\ Actual\ Static\ Values \end{gathered}$$

The solution to this problem for the dataset involved are the optimal weights for each of the defined WRF model at the location of interest. In this state, the optimization problem is put together in the format of a MATLAB function in a script file. This script operates the functions which follow the data flow-process in Figure 3-6. This training approach ingests multiple time-based historical forecast and ground truth files which are set for a single location. This data is then formulated into a matrix and passed to the GlobalSearch optimization function. This optimization function then uses the scatter-shot approach described above to identify the optimal weights to minimize the objective function. As detailed above, the data which is used in this system must be from real, continuous measurements, and the objective functions must be continuous and differentiable. In this research model, these conditions are met using forecast model outputs that are real continuous values between 0 and 1367 w/m^2 (Missouri Department of Natural Resources, 2021) with the objective function and constraint functions being differentiable. Once the GlobalSearch optimization sends a message that the exit point is reached, this set of values represents the forecast model weighting values which result in the minimum amount of error given the training set of data. The weights then get passed back to the Operational Mode to be used in future forecasts.

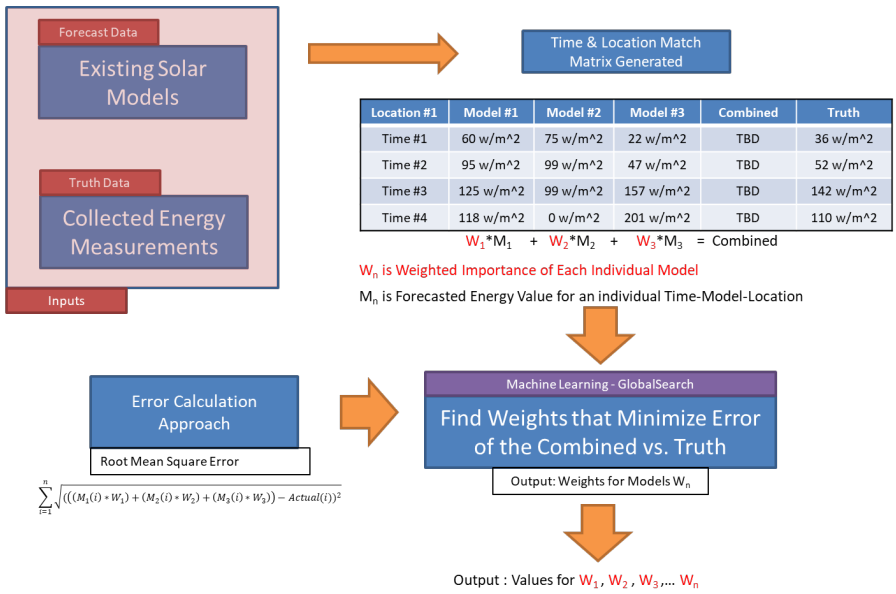


Figure 3-6. Combined WRF Training Algorithm

3.8 Forecast Process & Output

The research tool developed based on the above underlying methodologies makes both the process of generating and collecting data, as well as the processing of that data to calculate weights, opaque to the end user. This approach enables the end user to focus on the output from the model, not the intermediate steps. An end user sees two steps - weight generation and forecast generation. To generate the weights, a user enters a latitude and longitude of interest into the tool and starts the training regime. Behind the scenes, the training regime automatically reaches out to the respective data sources and downloads needed model initialization and ground truth data. Once downloaded, at least 219 model initialization files are processed through the six different WRF models. The forecast results and ground truth data are then ingested into MATLAB where the GlobalSearch optimization function identifies the linear combination weights for each of the six WRFs. These weights are then saved for later use. This initialization process takes 2-3 days for each new site of interest, but the weights remain valid going forward.

Once the weights are set, the tool is enabled for operational use. An end user only needs to select the pre-trained location which a forecast is desired and then starts the forecast request. The tool then downloads the most recently available forecast initialization file and process it through all six WRF models. This six-to-eight-hour process, once complete, loads the output files into MATLAB where weights are applied to each of the individual WRF forecasts before the results are summed. Once summed, the final portion of the model generates the graphical and tabular output before the end user is notified. The forecast is output to a GIFF movie file and saved as CSV and MAT tabular files. A single frame of the GIFF is presented in Figure 3-7, below.

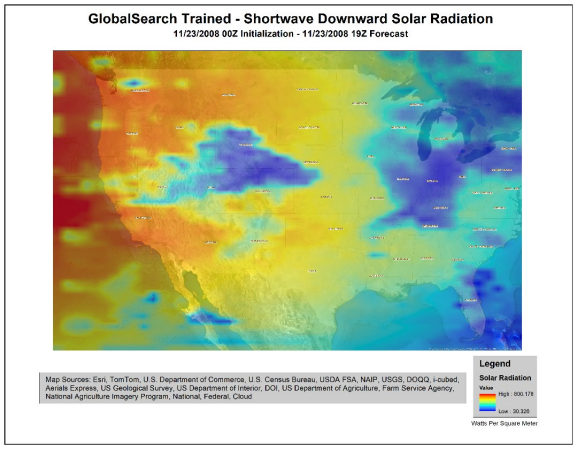


Figure 3-7. Combined WRF Graphical Output

3.9 Product Validation

After completing the initial model runs and collecting a dataset, the model needs to be validated to confirm reliability and performance prior to being utilized for testing hypotheses. This praxis takes a split validation approach using a combination of raw calculations and statistical analysis for this. This section will detail the process involved in validation with results detailed in Chapter 4. The first validation method looks to

assess the performance of the model from a functional level – to meet the end user buy-off. This method utilizes the raw assessment of trained weights of WRF forecast models for each location of interest. In counting all models with a weight of greater than 1% (0.01) this method looks to confirm that the weighted sum method utilizes multiple models in combination as opposed to selecting a single model. Locations where the optimal weighting combination result in only one model selected represent locations where this methodology did not work. A result where most of the locations had >1% impact from multiple models indicate a validation of this basic approach.

The next method for validation focuses on proving that the combined model performs better than independent WRF models. This validation approach will also support assessment of the hypotheses later. For validating if the combined model reduces forecast error compared to independent WRF approaches, both statistical and raw calculation methods are utilized. First, a one-way ANOVA (herein refereed to simply as an ANOVA) analysis is performed using the error values of each individual model and the combined WRF model. This approach looks to identify if there is a statistically significant difference in the means of errors from the different modeling approaches. Three different end-user use cases exist for comparing the performance of the combined model compared to the independent ones. If the combined model has statistically lower error than all independent models, it will always be the best choice. If the combined model has statistically lower error than half of the independent models, it will perform better than an end user randomly selecting which model to rely on. Lastly, if the combined model has statistically lower error than at least one independent model, it will perform better only when the end user picks the worst model to use. These three use

cases are important because they infer the level of training an operator might need in order to out-perform the model. This same ANOVA approach will also be used to assess if the combined model performs statistically worse, as this is also important to remain low in order to ensure end user buy-off.

The ANOVA tests will be performed in MATLAB using the ANOVA1 function. For running the ANOVA, a statistical confidence level is needed. This alpha value is used to define the error bars for each distribution. Typical meteorological applications rely on an Alpha value of 0.05 (Halquist, 2006), but for this application an Alpha value of 0.1 is used because it is also a statistically acceptable value (Greenland, et al., 2016), and an analysis of historical meteorological data suggests that less than 2% of results would be impacted by this change as shown in Figure A-3. [39 of 2100] For each individual location, an ANOVA test will be run on the resultant model errors. The outputs of each ANOVA are in the format of Figure 3-8, where blue are error bars on the combined model, red on the WRFs that perform statistically better/worse and gray on the ones that perform equally. These results will be tabulated to identify location trends in errors.

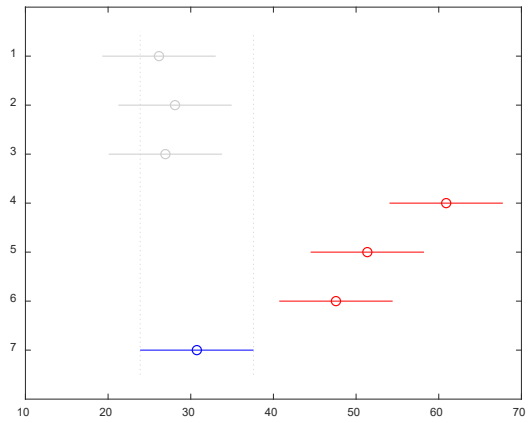


Figure 3-8. Example ANOVA Graphic [X Axis – Error Means in w/m^2]

The next validation assessment to be performed is a raw calculation based one. Where the ANOVA test assesses error distributions, this raw calculation approach assesses individual time-location errors across forecasting methods. This assessment takes the individual model predictions for a location and time and compares them to the truth value. Based on the law of large numbers, as the number of testing samples exceeds 30 (Sedlmeier & Gigerenzer, 1997), the raw number of instances where the combined model results in a lower error can be reasonably compared to those where it does not. This identifies the percent of occurrences in testing where the combined model trained on other data realized a better forecast than independent WRF models. While not a statistical based test, the end user's interest is in the accuracy of each time-point forecast.

The last test looks to identify for the third hypothesis if the performance of the combined forecast model is still improved over individual models if trained using temperature measurements, a more widely available dataset. (Bureau of Meterology, 2021) This analysis looks to compare two combinations. The first combination is trained model performance for locations where WRFs perform poorly. The second is a combination model trained on temperature data instead of solar radiation, but assessed for solar performance. For these two analyses the two-sample t-Test approach is utilized. Like the ANOVA tests above, this statistical test looks to identify trends in error distributions between the different approaches to determine if one use case is meaningfully better or worse than another one. This approach will be used to compare performance of the combined model across bottom 25^{th} percentile versus top 75^{th} and solar trained model versus temperature trained ones. A case where the distributions are

not statistically different indicates both approaches are equivalent. Validation of the research project model will be presented in Chapter 4.

3.10 Example Use-Case

The Solar Advanced Forecasting Tool (SAFE-T), known as the praxis research tool is opened by the end user initializing MATLAB as a "root" or administrative user. Once MATLAB has finished initializing, the user then opens the application by running the "SAFEt_Forecast.mlapp" application. This brings the operator to the home screen depicted in Figure 3-9. This display provides a GUI based control for the user to train (top left) and operate (bottom) the combination model and optimization algorithm.

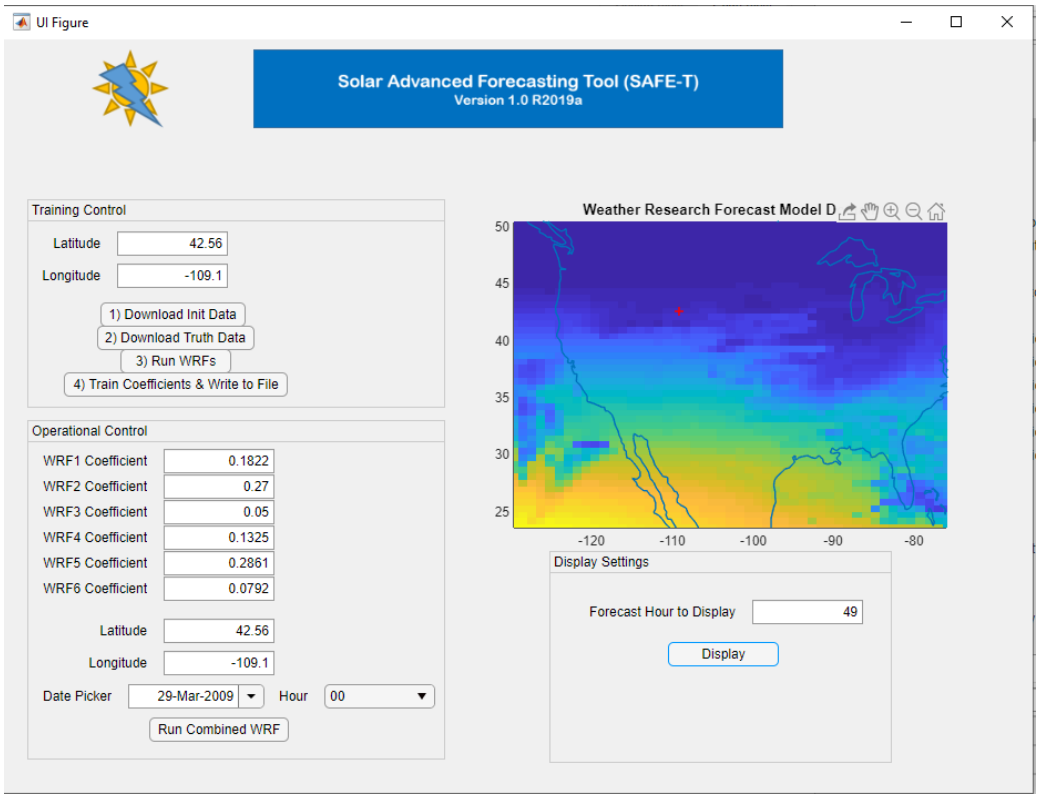


Figure 3-9. SAFE-T Main Display

To operate the training portion, the operator will input a location of interest and then follow the ordered steps within the "Training Control" section of the interface.

These steps command the Linux based programs to: 1) download historical initialization files, 2) download ground truth data for the location of interest, 3) run six versions of WRF at the location of interest for all downloaded initialization files, and 4) train the weighted sum model using these data points. The end user will then be able to move on to the "Operational Control" box where they are able to pick a date and time which they would like a combined forecast to be initialized at (present or past). The application will automatically download the appropriate initialization files, run the six versions of WRF, and post-process combine the files. The resulting data is then saved in a .mat file for use in other MATLAB programs or viewable for quick-look access in the figure to the right. The end user will not be expected to utilize the visual output, but the numerical forecast results (w/m^2) will be fed into another tool which validates if a solar energy generation facility deployment will generate sufficient electricity to meet energy demand on-site.

Chapter 4—Results

4.1 Introduction

This chapter provides a detailed analysis for the results and efficacy of the enhanced solar energy forecast tool developed in this praxis. Section 4.2 begins with a description of the data sources utilized and the process followed to generate the specific set of praxis model inputs for use in the analysis and hypothesis testing. Section 4.3 then provides an assessment of the validation for both the praxis model and the proposed statistical tests for analyzing the datasets. Once the validity of statistical methods and the model are confirmed this chapter will then detail the processes and results for proving or rejecting the hypotheses identified in Chapter 1. Each section will detail the data, data validation method, and statistical assessment processes used. Once the results are generated and an assessment is completed, each section will conclude with a statement detailing if the statistical tests were proven or failed to be proven. A proven statement will indicate that the praxis research model satisfies the research hypothesis, while a failed to be proven one will indicate inconclusive results to state that the hypothesis was proven. To recap, the three hypotheses from Chapter 1 are as follows;

1. Weighted combinations of Weather Research Forecast (WRF) models will reduce solar energy prediction error for at least 75% of individual locations.
2. Weighted combinations of Weather Research Forecast (WRF) models will reduce prediction error by at least 5% on average for 25% of the worst performing locations using a single model.

3. Weighted combinations of Temperature Based Weather Research Forecast (WRF) models will reduce solar energy prediction error on average by at least 2%.

4.2 Output Generation & Process

Valid hypothesis testing requires sufficient output forecast data to be generated using the praxis research model tool in order to accept or reject the hypothesis. As introduced in Section 3.9, the dataset used must contain sufficient time-samples and location samples in order to remain valid. For time-based samples, the literature (Hanke & Wichern, 2014) provides an assessment of required sample sizes for time series data. Hanke & Wichern's research identified the need for only 50 time-based samples to confirm a hypothesis with a 95% confidence interval. In assessing the number of geographic locations needed the literature (Sullivan, 2021) presents an approach based on binomial outcomes (true / false) to identify minimum sample size with Equation 7.

$$n = p(1-p) * \left[\frac{Z}{\varepsilon}\right]^2 \qquad (11)$$

Three parameters are required to calculate the estimated minimum sample size; p (population proportion), Z (the test statistic) and ε (acceptable margin of error). To calculate the Z value, a 90% confidence interval (Z=1.645) is chosen based on the (Halquist, 2006) statement that both 90% and 95% confidence intervals are widely accepted in meteorological applications. For the ε value, a margin of error equal to 2% is selected as the smallest measure of improvement contained within the three hypothesis which are being tested against. As this research has an unknown population proportion (p), a calculation is needed to identify the worst-case value to use instead. To do this, P*(1-P) is plotted in Figure 4-1. Based on Equation 7 and Figure 4-1, the most stringent p for setting the sample size requirement is a p-value of 0.5. Using the above calculated

values, the total minimum sample size for a geographically diverse assessment while still sustaining a less than 2% margin of error is 1692 total assessed geographic locations.

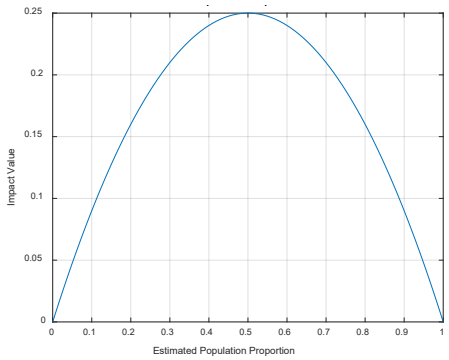


Figure 4-1. Population Parameter Estimation

To meet these time and location data requirements, an analysis region of 49 by 49 locations (2400 data-points) with 219 time-based files (65 Testing Time-Initializations) is setup. This analysis domain is geographically centered at the National Renewable Energy Laboratory's Colorado facility (39.7407° N, 105.1686° W). The grid has spacing from assessment point to point that is spaced at 100km in both latitude and longitude directions. This assessment spacing supports the desire of this praxis to identify any trends that might exist due to geographic or environmental differences across the continent. Based on these parameters, the analysis grid covers most of the continental United States as seen in Figure 4-2. Data to seed the time-variant model runs is sourced through the National Oceanic and Atmospheric Administration's – National Centers for Environmental Information's - Archive Information Request System (AIRS). Details on acquiring the data used in this praxis are found in Appendix B. The data source drawn from contains Global Forecast Service (GFS) 1 degree class model initialization files for the Weather Research and Forecasting model (WRF). The GFS provides initialization coverage using multiple integrated sensors for whole earth coverage. (Global Forecast

System (GFS), 2021) The availability of historical files and requirements of the statistical assessments result in the use of initialization files being used from October 1, 2005 (0Z) through December 27, 2009 (0Z) with a new file collected every seven days. The praxis model also requires corresponding day-time-location ground truth data to be collected in order to train the model. To meet the requirements of validation and testing, ground truth solar radiation and temperature is collected from the National Renewable Energy Laboratory (NREL)'s - Solar Radiation Research Laboratory's (SRRL) - Baseline Measurement System (BMS) (Stoffel & Andreas, 2015). The data collected matched the locations and times detailed above, along with 96 hours following each model initialization time to represent up to 4 days of forecasts. Detailed descriptions of where this data is sourced from can be found in Appendix B. While 2401 locations have been modeled for use under this validation, as depicted in Figure 4-2, details in red the locations where only modeled data is available and in green where both modeled and ground truth data exist. To meet the 1692 data-point threshold, 70% of the modeled area needs to also contain ground truth data. Based on the figure below, the dataset collected indicates 76.22% - which passes the test for calculation margin of error.

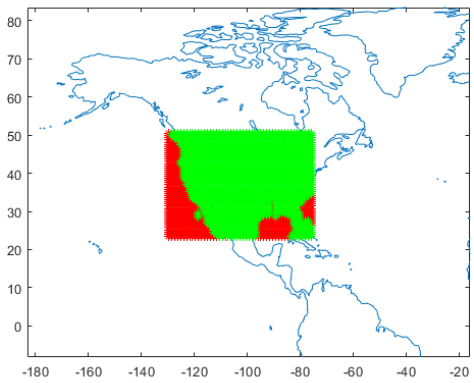


Figure 4-2. Research Assessment Area [Red – Model Only, Green – Truth Data]

The above data, based on the Chapter 3 methodology, provides sufficient quantities to train, validate, test, and assess the research praxis model and hypotheses. To begin the praxis research tool operation, each of the time-based GFS 1 Degree initialization files is processed through six different versions of the WRF model. The inputs to the WRF model are the GFS files and the output from each is a location and time correlated solar energy and temperature forecast. Each output for a single GFS file input results in a 49 by 49 by 96 matrix representing the location and time-based forecast from that particular model and initialization file. Figure 4-3 displays two outputs from different WRF models for the same initialization and forecast time. This process is completed for each of the 219 GFS files for each of the six variants of WRF.

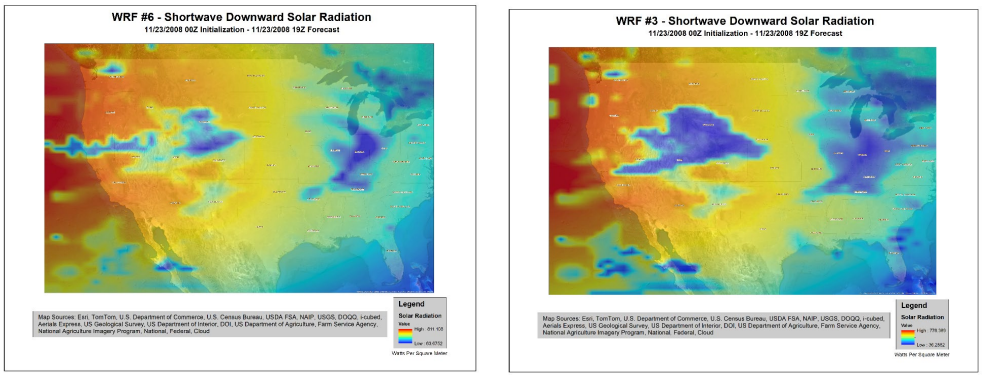


Figure 4-3. Two Independent Weather Research Forecast Model Forecasts.

These independent model outputs are the input to the enhanced ensemble combination method which forms the basis of the praxis research methodology to improve solar energy forecasts. To support validation and testing, 30% of the output WRF forecasts are randomly selected and reserved for testing later. The remaining 70% are used to train the praxis research tool. This tool, as detailed in Chapter 3, generates a linear weighted combination model for each of the locations of interest using the training

data and ground truth data. Once the training data is fed through the praxis tool in training mode, weights are generated for each WRF model at each geographic location. These weights are then saved for later use in generating combined forecast outputs using new input data for validating the model and hypothesis. These weights are used to linearly weight and combine each of the six WRF models to then calculate error for each forecast time-location period. An example of one time-based model combination across all locations is presented in Figure 4-4. This output is forecast data, which the end user of the research product will see. The hypothesis testing performed in later sections of this chapter relies on this data of combined forecast as compared to independent forecasts by using the ground truth data to assess error in confirmation or rejection of the hypotheses.

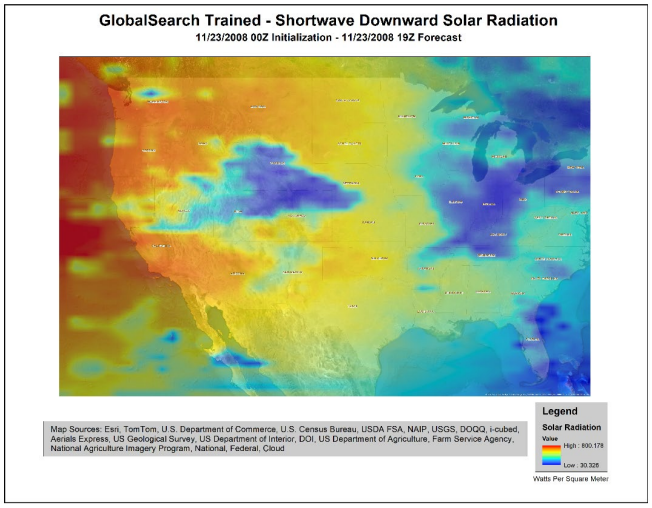


Figure 4-4. Trained & Combined Forecast.

4.3 Model Validation & Statistical Testing

The above combined forecasts, while valid outputs from the model (i.e., demonstrating the model does in fact run), do not represent a confirmed approach to solar energy forecasting. This section first addresses the statistical tests planned to provide

confirmation of validity and then dives into detail confirms that the underlying data of this praxis meets the rules for using these identified test methods. Then, the section will switch to detailing the validation of the underlying enhanced solar forecasting approach. The planned analysis of the validity of the three hypotheses, detailed in sections 4.4-4.6, is designed to utilize primarily two statistical tests. (Two sample t-Tests and One way ANOVA – Analysis of Variance) Chapter 2.10 outlined the validity of tests based on the literature survey (Hamill, 1999) (Decremer, Chung, Ekman, & Brandefelt, 2014) (Siuta, West, & Stull, 2017) (Eberly College of Science, 2021) (Physical Sciences Labratory, 2021) (Beersma & Buishand, 1999) (Allan, Pereira, Raes, & Smith, 1998) (National Weather Service, 2021) (NCSS Statistical Software, 2021) (Versta Research, 2021) (Meacham, et al., 2011) (Gerstman, 2021). This section performs analysis on the data which was collected and modeled to certify these statistical tests are valid when ultimately used to validate or reject hypotheses.

The following rules represent data attributes which are needed for the planned statistical tests to remain valid. The rules for these two tests are sourced from various pieces of literature and are outlined below. The ANOVA method typically contains three rules that need to be met: 1) Population Normality, 2) Equal Variance Distributions, and 3) Independence of Data. (Eberly College of Science, 2021) Based on the literature of Chapter 2, for forecasting applications, the ANOVA test has been identified as a valid test because these requirements are met through the following justifications: 1) meteorological data is generally normal (Physical Sciences Labratory, 2021), 2) variances have not been found to violate the variance requirement (Beersma & Buishand, 1999), and 3) meteorological data samples can be considered independent. (Allan, Pereira, Raes,

& Smith, 1998) To confirm these rules are met in the data used, the data for the praxis research tool will be assessed for equal variance and normality. Independence will not be assessed as the forecast measurements are all time-location-measurement variant.

The two-sample t-Test has a similar set of rules to the ANOVA test. If a data set satisfies the ANOVA requirements and the data is continuous [like meteorological data] (National Weather Service, 2021), then the set also satisfies the requirements to be able to perform a two sample t-Test. (NCSS Statistical Software, 2021) Based on these rules, if equal variance and normality are not rejected for the dataset, then the validity of both the One Way ANOVA and two-sample t-Test statistical assessments are confirmed.

The first data test performed in looking to validate these statistical approaches is conformity of equal variance. This test looks to identify for each individual test-location if the daytime solar energy forecast error data satisfies conditions of equal variance. This assessment uses the Bartlett's test for equal variance using MATLAB's "vartestn" function. For each location, a matrix of forecast error vector values (the first six columns with one for data from each independent model and the seventh for the combined model) are fed into the function, and the resulting test p-value is output. For low p-values (less than the specified alpha), the assessment of equal variance is rejected. The plots below in Figure 4-5 depict the p-values calculated based on thresholds of alpha = 0.1 (left) and alpha = 0.05 (right). For 81.48% of locations, equal variance is not rejected with an alpha of 0.1. For 88.31% of locations, equal variance is not rejected with alpha of 0.05.

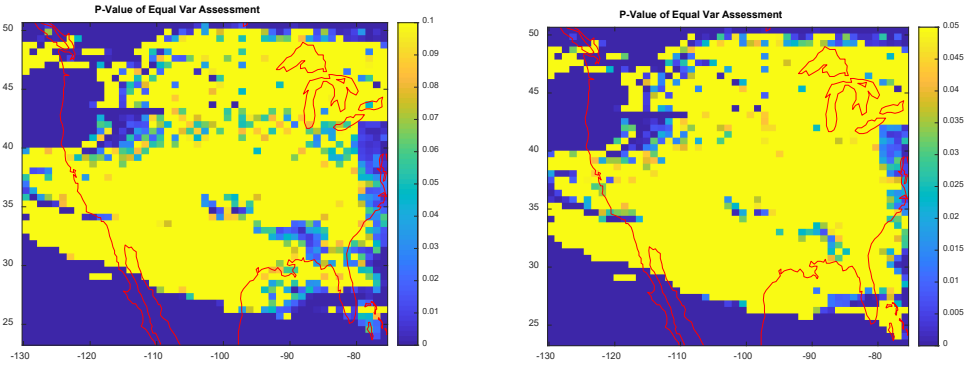


Figure 4-5. P-Value of Equal Variance. [0 to 0.1 (Left) | 0 to 0.05 (Right)]

This assessment, while not finding universal confirmation of equal variance, does not reject equal-variance data for a majority of locations. Based on the non-zero number of locations where an equal variance data assumption is rejected, before throwing out those locations - further investigation into the impact of this test is needed. Looking into Bartlett's test for equal variance, "The test's reliability is sensitive (not robust) to non-normality" (Borkowski, 2021). Based on this, the rejection of normality for these locations could be caused by small deviations from normality for the data. Before assessing the normality of data and impact of potentially non-normal data, this praxis completed an investigation into what the implication might mean. When assessing the ANOVA's test sensitivity from literature it is identified that, "Moderate deviations from the assumption of equal variances do not seriously affect the results in the ANOVA. Therefore, the ANOVA is robust to small deviations from the [sic] assumption. We only need to be concerned about large deviations from the [sic] assumption." (Borkowski, 2021) Based on an NIH study of the qualitative - quantitative link of the term "moderate," a deviation under 30% would qualify. (Selby, et al., 2009) Because of this,

with under 20% of locations which failed to identify as equal-variance for a sensitive test, this data is determined to be sufficient for continued use of the ANOVA method.

Even with the above, this praxis research continues to assess the data set for the ANOVA requirements including normality of data. To check for normality, the Anderson-Darling test is used on a sample of the forecasting data. Using the MATLAB command "adtest" multiple length vectors of randomly selected forecast errors from the combined model are assessed. The table below showcases the results from the test. For hypothesis=1, the null hypothesis of normality is rejected. For hypothesis=0, the hypothesis of normality of data fails to be rejected. This test is run against sample sizes ranging from 10 through 16,000 of daytime solar forecast errors to generate the following hypothesis results in Table 4-1.

Table 4-1. Normality of Data Hypothesis Testing – Palmarejo, MX

Sample Size:	10	100	250	500	1000	2500	5000	7500	10000	12000	13000	16000
hypothesis	0	0	0	0	0	1	1	1	1	1	1	1

When the Anderson-Darling test is run for the entire dataset (16,000 error samples) a rejection of normality occurs. This test, when run against smaller subsets of up to 44 days of data, does not reach the same rejection of normality. Based on a survey of literature, this pattern may be caused by a natural phenomenon such as seasonal time-variant changes in the meteorological distributions. (Jewson & Caballero, 2003) Other research questions the validity of tests like Anderson-Darling and others for large sample sizes, stating that: "The formal normality tests including Shapiro-Wilk test and Kolmogorov-Smirnov test may be used from small to medium sized samples (e.g., $n < 300$)" (Kim H.-Y. , 2013). From this, alternative assessments to identify or reject the normality of the data are performed. The first approach takes the raw error data for

forecasts and plots using Q-Q plots of normality. The two figures below are Q-Q plots for solar forecast (Figure 4-6) and combined model solar forecast (Figure 4-7) error for near a single randomly selected location in Chínipas de Almada, Centro Chihuahua, Mexico. In a generic Q-Q plot, a perfectly normal distribution of data is depicted when the blue marks line up with the red line. (Wang, Steele, & Zhang, 2016)

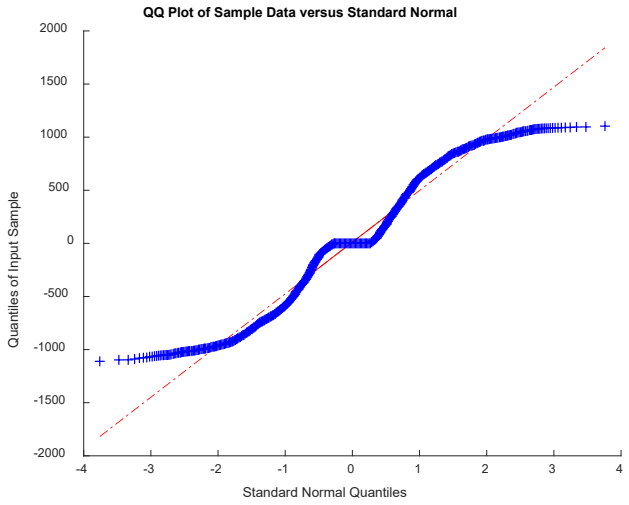


Figure 4-6. Q-Q Plot of Solar Forecast Data.

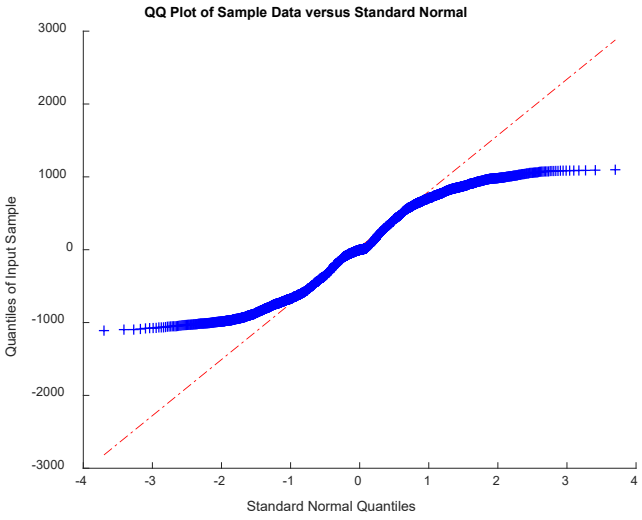


Figure 4-7. Q-Q Plot of Combined Solar Forecast Error.

Based on the plots above, the pattern identified is consistent with a minimally skewed distribution with moderate kurtosis, along the lines of a uniform distribution. (Wang, Steele, & Zhang, 2016) To further assess the distributions and identify if these distributions truly represent uniform distributions or are simply non-perfect normal distributions with uniform tendencies, histograms of the data are plotted. Figure 4-8 plots the forecast-error occurrences for each of the six Weather Research and Forecast models used in the combination model. These distributions are then assessed against depictions of standard normal distributions to determine a qualitative level of normality.

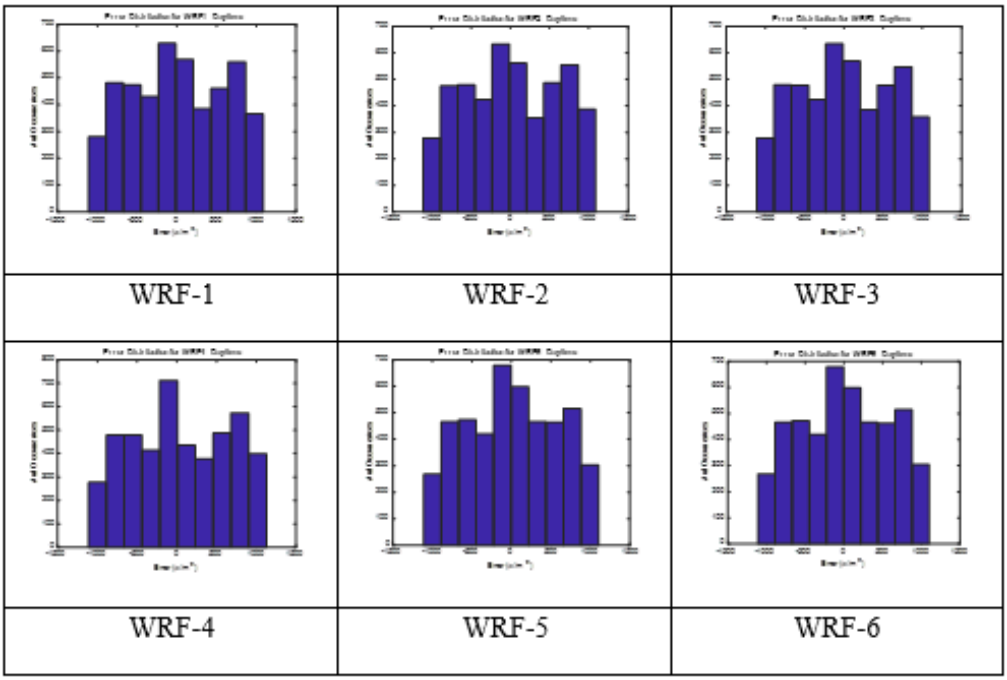


Figure 4-8. Histogram of Independent WRF Solar Forecast Error.

Based on comparing these plots to uniform distributions in (Wang, Steele, & Zhang, 2016), no conclusive pattern of normality is identified. As identified in the Q-Q plot assessment, the data appears to be between a normal distribution and a uniform one. From here, an additional approach based on (George & Mallery, 2010) is brought

forward to assess normality. In this method, an approach was identified to assess normality based on bounding acceptable kurtosis measurements which still exhibit normality, a Kurtosis value of plus or minus two. The first measurement of kurtosis is done for the Palmarejo, MX dataset. In this case, the Kurtosis = 1.9291, which implies normality of data. To complete the assessment, the Kurtosis is measured for all datasets generated, and the average Kurtosis for the multiple models is plotted in Figure 4-9.

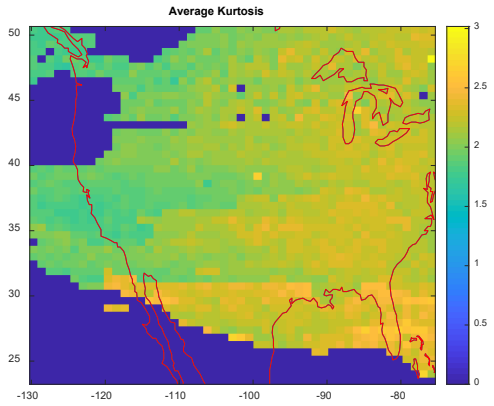


Figure 4-9. Forecast Error Average Kurtosis.

This calculation identifies that the dataset meets the George & Mallery threshold for normality for only 75% of locations. Upon further review, based on (Uchida, et al., 2017) this 25% rejection rate when using a less than two Kurtosos threshold could be due to time-varying phenomena in atmospheric science models identified earlier. An alternative Kurtosis based approach identified by (Kim H.-Y. , 2013) found that if the absolute Kurtosis does not exceed seven (software Kurtosis of four), “we cannot substantiate a conclusion of non-normality”. Based on this, when the output forecast errors from the praxis research model assessed against a Kurtosis threshold of four, the analysis in this praxis fails to reject normality.

From the above analysis and literature, this dataset has been validated to meet the identified rules for the statistical assessment methods planned. These methods will be used in both validating the underlying model and confirming/rejecting null hypothesis. Turning to validation, the first assessment looks to determine if the model combination approach benefited from the use of multiple models (i.e., if the training method found it beneficial to integrate outputs from multiple models). To complete this assessment, the weights calculated at each of the 2,401 locations when the training regime is completed are gathered for all locations of interest. At each location, weights are stored in a table like the one below. For each WRF model with a non-zero weight, the model was identified as having an impact on the combined output solar forecast.

Table 4-2. Calculated WRF Weights – Palmarejo, MX

Location	WRF-1	WRF-2	WRF-3	WRF-4	WRF-5	WRF-6
27.423000, -108.378000	0.2424	0.0189	0.0072	0.0055	0.3518	0.3742

To identify if multiple models are commonly combined to generate a lower error model, the number of non-zero weighted locations are counted. This calculation is performed for all the locations where the research model had been run for, and the map in Figure 4-10 is generated. For each of the pixels on the map, a non-blue color represents a location where at least one WRF model was successfully run, and ground truth data exists. The other colors [Tan through Red] indicate locations where at least one WRF model is used in generating the lowest error combined solar energy forecast. The closer to tan the color, the fewer the number of models combined in the optimal solution. In assessing the complete praxis training dataset, the average number of WRFs combined is 4.165 per location (out of 6 available WRF models). The percent of locations where only one WRF (i.e., selecting one model to use for that location) is the optimal solution measured at only 1.2%. This implies that for 98.8% of locations, combining multiple

WRF models results in an error which is lower than the error of any individual WRF model by itself. Not statistically significant, this implies that for most locations, combined forecasts perform more optimally than using a single forecast output.

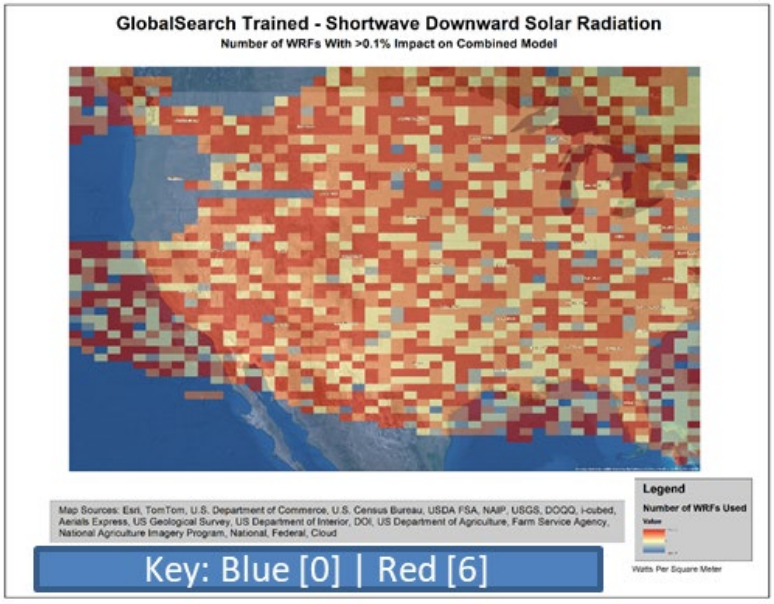


Figure 4-10. Number of WRF Models Combined per Training Location

While this indicates potential efficacy of the combined ensemble approach this research praxis tool uses, the next validation approach begins the statistical assessment. In Chapter 2 of this praxis, multiple acceptable significance levels (alpha) were identified as commonly used in the meteorology field (Halquist, 2006). Before running with one value or the other, an assessment is performed using the data output from the praxis combined forecast model to identify if either Alpha = 0.05 or Alpha = 0.1 presents meaningfully different results. To run this assessment, an ANOVA is performed comparing the error distributions of the six independent WRF models to the error of the trained combination model at each location. The left column of plots in Figure 4-11 showcase locations in red that perform statistically better when using the praxis

developed combined forecast model than all independent WRF models. The center column depicts in red locations that the praxis model performs statistically better than half of independent WRF models. The right column depicts in red locations where the praxis model performs statistically better than at least one independent WRF model. The assessments in the top row of Figure 4-11 utilize an ANOVA with Alpha = 0.1 and bottom row an ANOVA with Alpha = 0.05.

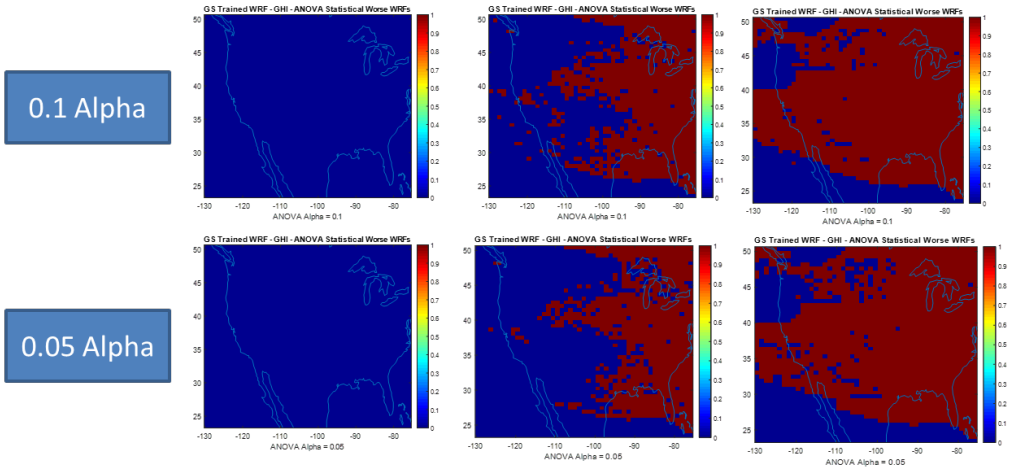


Figure 4-11. Alpha Assessment

Based on these calculations, the trend is identified where only 39/2100 locations which experience a difference in result based on the Alpha = 0.05 vs. Alpha = 0.1. Locations which experienced this difference can predominantly be found in the desert southwest and mountain northwest but represent a small fraction of overall locations. Generally, while 0.05 is considered the meteorological standard (Halquist, 2006), based on these results and the earlier literature in Chapter 2 rejecting a material difference – the difference is small enough that given the margin of error identified earlier in this chapter, the hypothesis tests would all be satisfied using either value.

With the parameters for ANOVA determined, the statistical validation of the praxis research tool begins with the purpose of validating that the performance is at least as good as independent WRF models. The primary validation approach here is the use of the ANOVA test to determine if the means of forecast errors have statistically significant differences. Prior to completing the assessment over the entire forecast error, an initial test is done at one specific location, the NREL Bolder Colorado office. This trial is to demonstrate the intermediate analysis steps which will underlie the complete hypothesis testing later in this chapter. In this case, the forecast errors for each of the independent WRF models and the combined research tool model are run through an ANOVA. The output result is presented in Figure 4-12. The color-scheme has blue representing the error bounds of the praxis combined WRF approach, gray representing individual WRF error bounds who do not have statistically different means, and red for individual WRF error bounds which have error distributions that are statistically better or worse than the praxis developed combination approach.

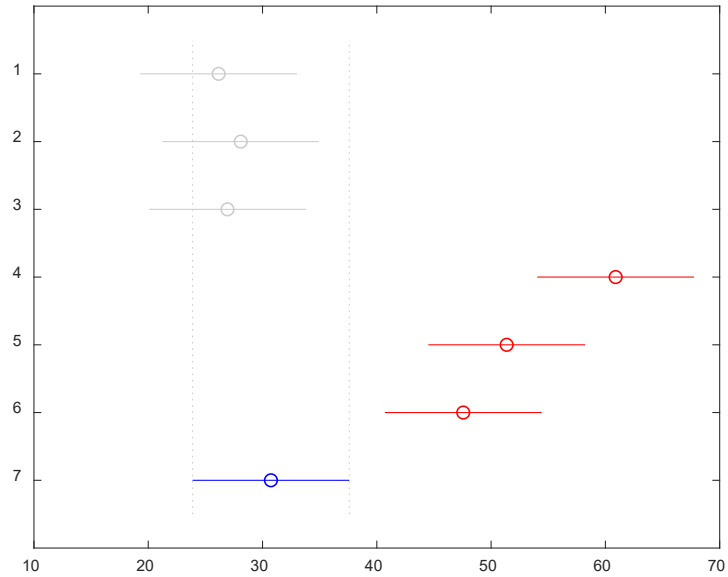


Figure 4-12. ANOVA Assessment of Error Variances. [X-Mean Error w/m^2]

For each of the locations where both WRF models have been run and ground truth data exists, one of these plots are generated to determine the statistical significance. To support later calculations, the number of independent WRFs identified as statistically better, statistically worse, and statistically similar are counted and logged. Additionally, the ANOVA p-values and statistics are generated for each location, like the data found in the table below. Once generated, these are logged for use in assessing the strength of the differences in mean errors. For a low p-value like the one found in the table below, the null hypothesis is rejected, and the statement can be made that the error distributions are not equal across all the individual and combined models for this location. These p-values are computed for all modeled locations in Table 4-3 and plotted in Figure 4-13.

Table 4-3. ANOVA Table for NREL Colorado

Source	SS	*df*	*MS*	*F*	*Prob>F*
Forecasts	7.12E06	6	1184735.8	5.77	5.13E-06
Error	8.49E09	41349	205312.4		
Total	8.50E09	41355			

Note. Small P-Value indicated Null-Hypothesis Rejected, Distributions not equal

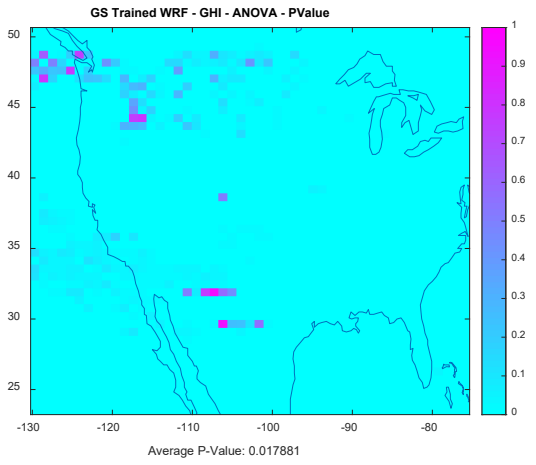


Figure 4-13. ANOVA Model Operation Area P-Values.

This chart of p-values identifies 48 locations where p-values are high enough that it cannot be stated that distribution means are unequal. This implies for those locations no forecast model or combined model is better or worse than others. For the locations where there is a statistical difference, the impact of the significance is found in using the earlier logged ANOVA data. Figure 4-14 below showcases three scenarios based on counting the number of statistically significant differences for each location. These plots showcase in blue locations where the praxis research forecast combination model performs statistically better than independent WRF models. The left figure represents the case where the praxis model is statistically better than all independent WRFs. This represents locations where the developed model performs better than an end user already knowing what the best model to use is. The middle figure represents locations where the combined model is better than the end user picking a model randomly. The right figure represents locations where the combined model is better than the end user picking the worst model. These indicate that depending on the use case of the end user, the developed model will range from not providing statistically improved results to providing statistically improved results for 98.8% of locations. The end users of this praxis tool desire the ability to improve forecasts without requiring a subject matter expert to identify the best solar weather forecasting model for use at each location. Based on this, the reach goal of the end user organization for the praxis developed combination forecast model tool is to improve forecast accuracy with respect to all independent WRF models. This, however, is not the only threshold which defines success. If the praxis forecast tool could statistically reduce the error as compared to randomly selecting a model or selecting the worst independent model, this would still be considered a success by the end user.

0% Improvement 68% Improvement 99.8% Improvement

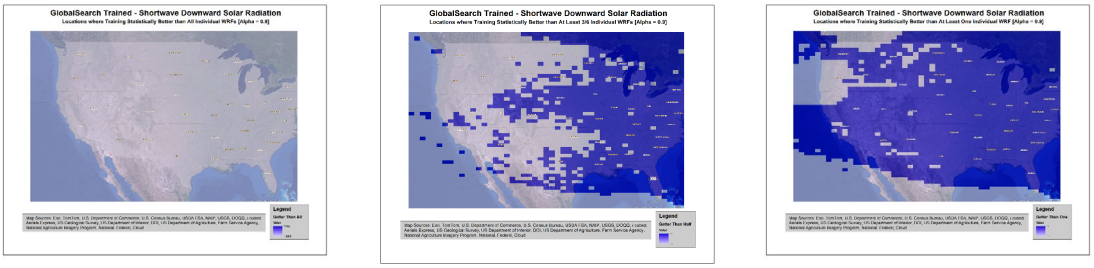


Figure 4-14. Combined Model Statistical Improvements.

In addition to understanding the improvement that the praxis developed model offers, the end user is highly sensitive to cases where the new model might be statistically worse. So, to further validation for the end user, a similar assessment of research tool is performed to then identify when it performs statistically worse than individual WRF models. This assessment counts the number of WRFs which the combined model performs statistically worse in. The plots in Figure 4-15 showcase the percent of locations where picking the combined model performed statistically worse than independent WRF models. The left plot in blue identifies locations where the combined model performs statistically worse than all individual WRF models. The middle plot showcases in blue where the combined model performs statistically worse than half of individual WRF models (i.e. random selection by the end user). The right plot showcases in blue locations where the combined model performed statistically worse than at least one individual WRF (i.e. compared to the selection of best model by the end user). Based on the comparison of the plots assessing increased performance (above) and degraded performance (below), the combined research model is more likely to statistically improve performance (up to 99.8%) while providing low incidences of statistically significant decreased forecasting performance. (up to 2.6%)

0% Degradation <1% Degradation 2.6% Degradation

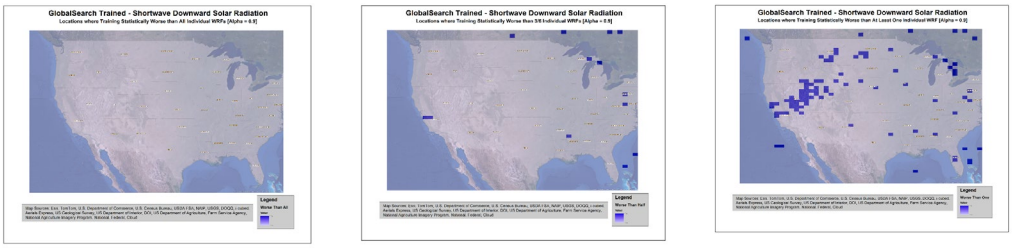


Figure 4-15. Combined Model Statistical Degraded Performance.

From the above, the performance of the model should provide statistically improved forecasting ability for the end user while having low occurrence of degraded performance. For an end user group with no a priori knowledge of independent WRF (or other forecast) model performance for a particular location, this praxis research tool satisfies the research objective of the end user. From here, the next three sections will continue statistical assessments for validating or rejecting the proposed research hypothesis. These assessments will use portions of the dataset generated for the above assessment while performing specific calculations. The validation of one or more of these hypotheses will guide the specific use-cases which the end user can direct this research product to be used in.

4.4 Weighted Combination – Overall Forecast Improvement Assessment

This section begins the validation or rejection of the research hypotheses. The first hypothesis this research tool is being analyzed against is that; “Weighted combinations of Weather Research Forecast (WRF) models will reduce solar energy prediction error for at least 75% of individual locations.” The implication of this hypothesis is that the research tool is better than existing individual WRF model forecasts for at least 75% of locations in the representative domain. The sub-set of data used in the

confirmation or rejection of this null hypothesis comprises parts of the dataset described in Section 3.5. This includes all six WRF forecast model outputs from 219 initialization files covering all 2401 locations run over 96 hours for each initialization file. The data pulled from these outputs comprises time (UTC) – location (Latitude/Longitude) – Solar Forecast (Watt Per Square Meter). Additionally, ground truth data matching the above format is collected for the assessment. This data is then split in a 70%/30% ratio and processed through the combined model for both testing and validation. The results from the validation will be used to confirm or reject the null hypothesis using both a statistical ANOVA method and raw forecast-hour by forecast-hour comparisons. As mentioned in Chapter 3, in weather and solar forecasting domains both statistical and raw performance are important metrics.

Assessment of the model performance with respect to the first hypothesis begins with the ANOVA data matrix generated in Section 4.3. In recap, this data identifies locations which based on the statistical ANOVA assessment had error distributions with statistically higher or lower forecast error means for individual WRF models than the praxis tool's combined model. For the purpose of testing this hypothesis, the collected counts can be used to identify locations which have statistically significant improvements. Once collected, the counts of these locations are then compared to the overall performance across the modeled area to determine percent of locations which are statistically improved. Figure 4-16, taken from the previous model validation section, presents the three end user scenarios of interest for understanding model performance. These are: 1) percent of locations where the research product had a forecast error distribution statistically lower than all independent WRF models, [Left] 2) percent of

locations where the research product had a forecast error distribution statistically lower than half of independent WRF models, [Center] and 3) percent of locations where the research product had a forecast error distribution statistically lower than at least one independent WRF model. [Right] In each of these images, a blue square indicates a location where the combined research tool forecast has statistically lower error while white cells are not improved or lacking sufficient data.

0% Improvement 68% Improvement 99.8% Improvement

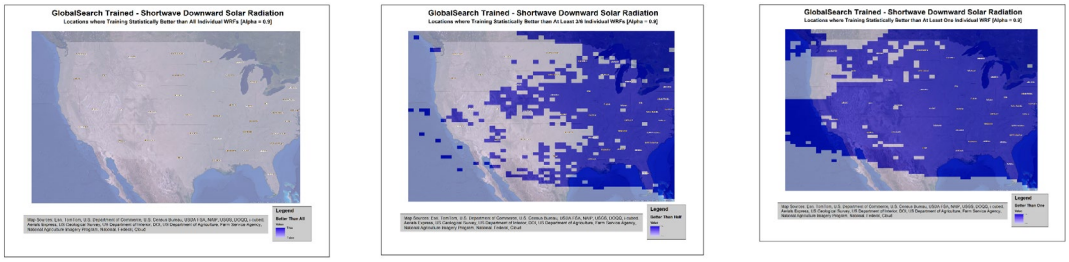


Figure 4-16. Combined Model Statistical Improvements.

Each of these three end-user cases represent a different level prior knowledge about how independent WRF models perform at a particular location. For the remote location new deployment case this praxis tool is developed for - not having prior knowledge the operational usefulness of even the “better than the worst” provides information to the end user. As seen from the charts above, this research product does not ever provide a statistically improved result when compared to the best performing individual WRF model. As the end user will not know which model will perform best, however, the other two cases of random model draw (68% improvement) and improvement over the worst model selection (99.8% improvement) are more applicable because they still reduce risk for the end user. The initial finding of this testing identifies

that the first hypothesis is only validated for the case comparing the research product against the worst performing WRF model, and not the ‘random draw’ or ‘always’ cases.

The statistical performance of the model, as identified above, only provides insight into the performance of the distribution of errors for the models. As mentioned earlier, the end user group is highly driven by the instantaneous comparisons of WRF independent models versus combined research tool forecast performance. This praxis performance assessment continues looking at how the combined model performs as compared to individual WRF models in hour-by-hour forecasts (comparing same time-location across different models). For each individual time-location forecast, the error of the research product combined model is compared to the error of each individual WRF forecast. The number of times the research tool has lower error is counted, and a 90% confidence interval is used. The computed results are presented in the Figure 4-17, below.

77.76% of Locations 99.45% of Locations 99.51% of Locations

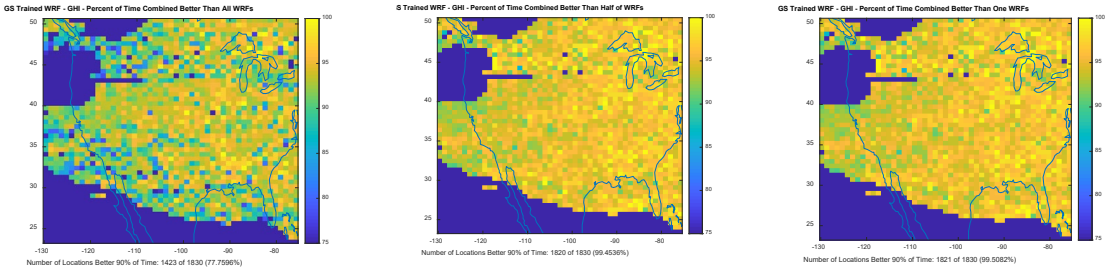


Figure 4-17. Individual Forecast Improvement vs. Individual WRFs.

These outputs represent the percent of locations which experience improvement as compared to the same three end user cases. Each, however, is based on forecast error improvements of raw forecast-by-forecast comparison. The performance of the three end user cases is as follows: 1) for 77.76% of locations, the combined research model has lower raw error forecasts than all independent WRF models [Left], 2) for 99.45% of

locations, the combined model had lower raw error forecasts than at least half of independent WRF models [Center], and 3) for 99.51% of locations, the combined model had lower raw error forecasts than at least one independent WRF model. [Right] Based on both the statistical and raw assessments, this praxis can make the following claims. For the research praxis model with a statistical assessment alpha of 0.1, the null hypothesis is rejected for 100% of location that this combined forecast model improves forecasts when compared to every independent WRF and for 32% of locations when compared to half of independent WRF models. However, for 99.8% of locations with an alpha of 0.1, this praxis fails to reject the null hypothesis that this combination model statistically improves forecasts when compared to the worst performing independent WRF model. From a raw analysis perspective, the hypothesis threshold is met for 77.76% of locations when compared to all WRF models, 99.45% of locations when compared to half of WRF models, and 99.51% of locations when compared to the worst independent WRF model. Therefore, using a raw assessment fails to reject the null hypothesis.

4.5 Weighted Combination – Worst Performing Forecast Improvement Assessment

This section assesses the praxis research model against the second hypothesis that; "Weighted combinations of Weather Research Forecast (WRF) models will reduce prediction error by at least 5% on average for 25% of the worst performing locations using a single model." The implication of this hypothesis is that locations where a single independent WRF performs poorly (in the bottom 25^{th} percentile) that the combined model approach used in this praxis would improve overall forecast error by at least 5% on average. This represents an improvement in-line with other forecast improvement projects. (Solar Forecasting 2 - Solar Energy Technologies Office, 2021) The dataset to

be used in confirming/rejecting this null hypothesis is a sub-set of the dataset used in Section 4.4. This sub-set contains the locations where the average solar energy WRF forecast error across all instances are in the top 25% for forecast errors. Surveying the raw error data, the locations with the highest mean error are isolated. These locations of interest are plotted in Figure 4-18, where white locations represent the locations of interest with the worst performing independent WRFs and gray represent the remainder modeling area not used in this assessment. This data to be used includes all six selected WRF forecast model outputs from 219 initialization files covering all 625 locations run over 96 hours for each initialization file. The data pulled from these outputs comprises time (UTC) – location (Latitude/Longitude) – Solar Forecast (Watt Per Square Meter). Additionally, ground truth data for the above format is collected for the assessment. This data will be split 70%/30% and processed through the combined model in a testing/validation split. The results from the validation split will be analyzed and then used to confirm or reject the null hypothesis.

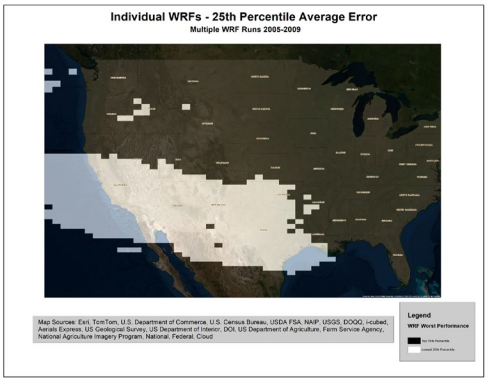


Figure 4-18. 25th Percentile of Individual WRF Performance Locations.

The analysis needed to validate or reject this hypothesis first starts with a two-sample t-Test. If the error means are not statistically different between locations in the

bottom 25th percentile and the overall dataset, this can leverage the previous ANOVA assessment and error data to validate this null hypothesis. The two-sample t-Test, like ANOVA, requires testing for both equal variance and normality of data in order to perform statistical assessments. The equal variance (Bartlett's) and normality (Kurtosis) tests used on this sub-set of the Section 4.4 dataset are done to confirm this sub-set still maintains equal variance and normality. Bartlett's test is run across all locations of interest to confirm equal variance. Based on the outputs, Figure 4-19 (below) is generated. In this figure, a black square in the area of interest represents a location where the null hypothesis is rejected for the Bartlett test that the data is of equal variance. Based on these results, less than 4% of these locations of interest exhibit distributions that can be considered of potential equal variance. This indicates that the standard two-sample t-Test is not suitable for this dataset.

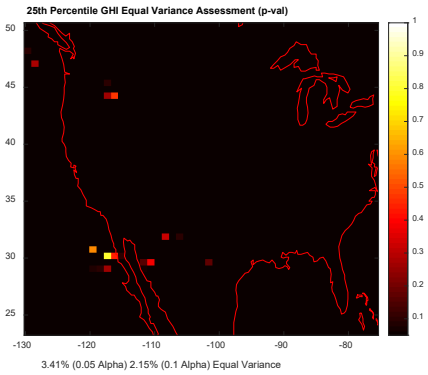


Figure 4-19. 25th Percentile Equal Variance Assessment.

Because of the unequal variance, Welch's two-sample t-Test is used. As Welch's test is designed to work in the case where populations are normal, if the data observations are independent and normally distributed, it is suitable. (Albright, 2021) As addressed earlier in this chapter, as each time-location-data-point represents an independent

measurement, independence of observations is assured. Normality is again assessed through a Kurtosis calculation. Figure 4-20 plots Kurtosis for locations of interest. With all having a Kurtosis below 4, normality is assumed. (Kim H.-Y. , 2013)

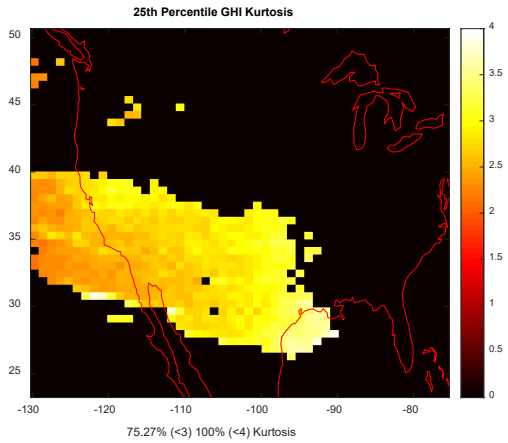


Figure 4-20. 25th Percentile Kurtosis.

Based on the above, this assessment satisfies the requirements for the use of Welch's two-sample t-Test. This test is then performed by comparing the praxis combined model forecast errors of locations where original error was in the worst 25th percentile (locations of interest) with those in the best 75th percent (all other locations). The output of Welch's test is in Table 4-4. Based on the p-value of 0; the null hypothesis is rejected that these two sets of errors come from the same population. This implies that the resulting forecast errors of the bottom 25th percent of originally performing locations continue to remain worse than errors for the top 75% of originally performing locations.

Table 4-4. Welch's 2-Sample t-Test Improvement of Entire Grid vs. 25th Percentile

T-stat	*df*	*sd*	*CI*	*PValue*
-23.5934	1828	0.9004	-1.1539	0

The next step in validating or rejecting this hypothesis is to test the distributions of errors for the combined model against the error distributions for each independent WRF. Given the failure to confirm normality of data, this assessment turns to the Kruskal-Wallis test of equal distributions instead of the ANOVA test. The Kruskal Wallis test only requires independent samples with independent observations (Mathworks, 2011) which are guaranteed given the time-location-model independence of data collected. The Kruskal-Wallis test is performed using errors from all independent WRFs at a location compared to the praxis research tool model. For each location in the lowest 25th percentile, the p-values are logged and plotted in Figure 4-21. The null hypothesis for this test states that all distribution samples in the set come from the same distribution. Therefore, a low p-value rejects that all samples come from the same original distribution. Based on the results, 86.56% of locations fail to reject the hypothesis. This implies that for most locations tested where independent models originally perform poorly, the combined model is not statistically different in error distribution than the independent models. This implies the combined model does not perform better here.

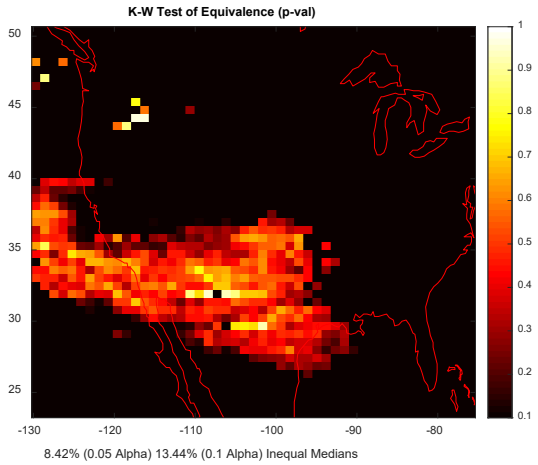
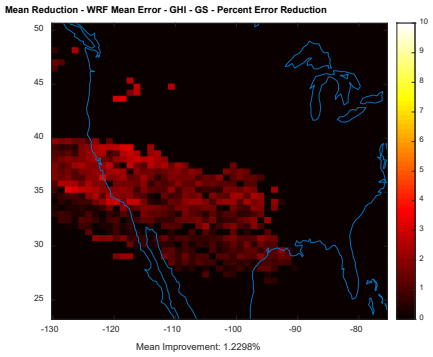


Figure 4-21. 25th Percentile Kruskal Wallis

With the statistical based assessment showing a rejection of the hypothesis, one further assessment is performed using the raw time-location by time-location analysis which was used in the first hypothesis. This approach looks to identify instances where the developed combination model generated a more accurate forecast than individual WRF models for locations where the individual model performance is in the 25th percentile. This assessment calculated the mean instantaneous percent improvement or degradation in error for the combination model as opposed to the independent WRF models. These percent improvements and degradations are then averaged to identify the average improvement or degradation rate for each location of interest. The plot in Figure 4-22 depicts the results for average combined model error improvement versus mean independent WRF error [Left] and average combined model error improvement versus max independent WRF error [Right]. The improvement in error for the worst performing 25th percentile of locations when using the research product combined model is 1.23% [vs. Mean Independent WRF Error] and 6.51% [vs. Max Independent WRF Error]. This demonstrates when comparing raw forecasts across models instead of overall error distributions, the combined WRF model does perform better than independent models.

Mean Error Reduction



Maximum Error Reduction

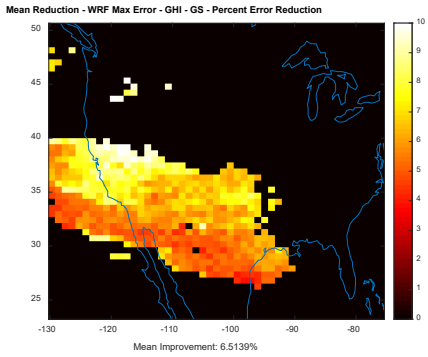
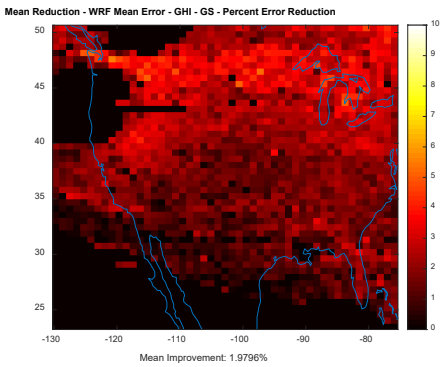


Figure 4-22. Combined Model Improvement 25th Percentile.

These results are then compared to the error improvement performance of the overall model to assess if the percentages could be meaningful. The same calculations are performed across the entire modeled area from hypothesis #1. Figure 4-23 below exhibits the average error improvements over the entire continental US modeled area. Comparing the two different areas (bottom 25th percentile and the rest of the locations), a pattern is evident. For locations in the bottom 25th percentile, there is a clearly demonstrated degradation in improvement – as identified with Welch's test. The implication of these results is that while the combination forecast model developed for this praxis improves solar forecasts for all locations, it performs statistically better in locations where independent WRF models already exhibit lower error occurrences.

Mean Error Reduction



Maximum Error Reduction

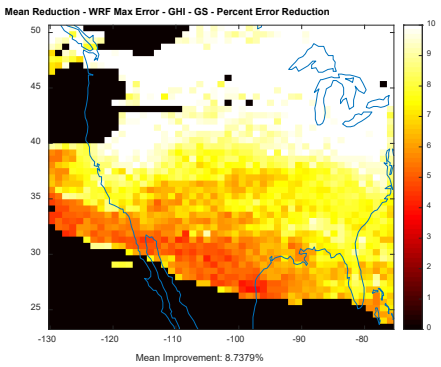


Figure 4-23. Combined Model Improvement Full Run Area.

The limited improvement in forecasting seen for locations with already bad independent model performance is a limiting factor for this model. Based on this assessment, however, these locations that perform the worst for independent WRFs do experience demonstrable improvement in forecast performance. Based on this analysis, with an alpha of 0.1 the p-value is found to round down to zero, so statistically the null

hypothesis that the combined model performs statistically better at improving forecasts for locations that have the worst independent WRF performance is rejected. Based on raw forecast hour by hour calculations, however it is confirmed numerically that the average improvement for the combined model vs. half of independent WRFs is 1.23% and when compared to the worst performing independent WRF is 6.51%. This demonstrates model improvement but rejects a notional early hypothesis that the praxis model would perform better in locations with particularly poor individual WRF model performance. Based on these results the null hypothesis failed to be proven statistically, but was confirmed through the use of raw calculations.

4.6 Weighted Combination – Temperature Data Forecast Improvement Assessment

This section assesses the praxis developed combination solar energy forecast model against the third hypothesis which is that; "Weighted combinations of Temperature Based Weather Research Forecast (WRF) models will reduce solar energy prediction error on average by at least 2%." The implication of this hypothesis is that for locations where no solar ground truth data exist to train the praxis ensemble model, temperature can be used for training that when used operationally for solar forecasting will result in solar forecast improvement of at least 2% as compared to independent WRF models. The data set used in the confirmation or rejection of the null hypothesis comes from the same data as the set used in Section 4.3. This includes all six identified WRF forecast model outputs from 219 initialization files covering all 2401 locations run over 96 hours for each initialization file. The data pulled from these outputs comprises time (UTC) – location (latitude/longitude) – Solar Forecast (watt per square meter) – Temperature Forecast (Fahrenheit). Additionally, ground truth data matching the above

format is collected for the assessment. This data will be split 70%/30% and processed through the combined model in a testing and validation split. Prior to beginning the assessment of solar energy forecast errors generated using temperature data trained models, an assessment involving confirming that the statistical methods are applicable to the dataset is performed. For this, first an assessment identify if the underlying data supports both equal variance and normality to support the use of the ANOVA method is required. Similar to the earlier hypotheses, Bartlett's method for identifying equal variance is used to assess the variances between independent WRF forecast temperatures and ground truth data. Figure 4-24 below displays the p-value of Bartlett's test for the modeled locations. This plot demonstrates that for 74.7% of locations where alpha = 0.1 and 78.8% of locations where alpha = 0.05, we fail to reject the null hypothesis that the distributions come from normal distributions with the same variance. (Mathworks, 2021) To further validate the normality, the Kurtosis based assessment method is also utilized. Based on the Kurtosis <4 normality method; the normality of the data is confirmed. The graphical depiction of the Kurtosis measurements is found in Appendix A; Figure A-4.

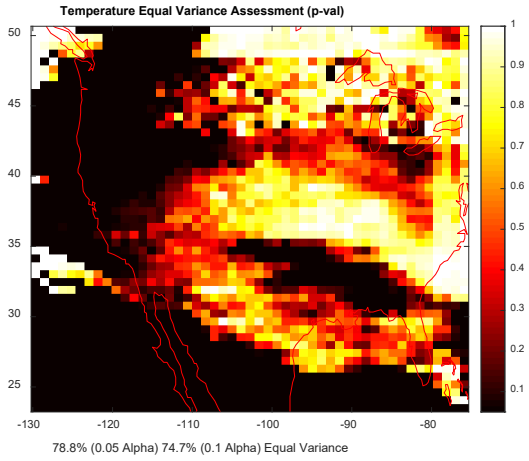


Figure 4-24. Temperature Data Equal-Variance Assessment.

With the normality and variance assessments confirming the general applicability, focus turns to processing the resulting model data to validate or reject the third hypothesis. To complete this assessment, the praxis enhanced solar forecasting combination model is run with temperature forecast outputs from the independent WRF models and temperature ground truth data. This dataset is used to train the weights of the forecast linear combination model methodology. Once temperature-based weights are calculated, the testing solar forecast data is processed using the weights from the temperature trained combination. The resulting output is a set of solar forecast error values based on the temperature trained combination model. The first assessment in validating the third hypothesis looks to identify the relative performance of the solar trained praxis combination model and the temperature trained praxis combination model. To identify if there is a statistical deviation of this temperature trained method versus the solar trained method and/or independent WRF models, an ANOVA assessment is completed comparing the forecast error values between all seven approaches. The table below showcases the output of the ANOVA assessment which again identified the low-p value indicative of non-equal error distributions. This indicates that the error distributions are not considered equal, and one or more models may perform statistically better.

Table 4-5. ANOVA Table for Model Combination Comparison vs. Temp

Source	SS	*df*	*MS*	*F*	*Prob>F*
Forecasts	7.15E06	7	1022043	5	1.139E-05
Error	9.67E09	47256	204590.4		
Total	9.68E09	47263			

Note. Small P-Value indicated Null-Hypothesis Rejected, Distributions not equal

The rejection of the null hypothesis could either be due to a difference between the two methods of training or still a hold-over of differences across independent WRFs. Like Section 4.4, a calculation is done to identify the number of independent WRFs which performed statistically better or worse than each of the two training approaches (temperature vs. solar). The following plot in Figure 4-25 identifies locations (in white) where the solar trained combination model performs statistically better than the temperature trained one when compared to more independent WRF models. The 8.71% reduction in locations indicates that the temperature trained model performs worse than the solar trained combination model. This indicates further assessment is needed in order to validate or reject the null hypothesis, and that previous tests cannot be directly used.

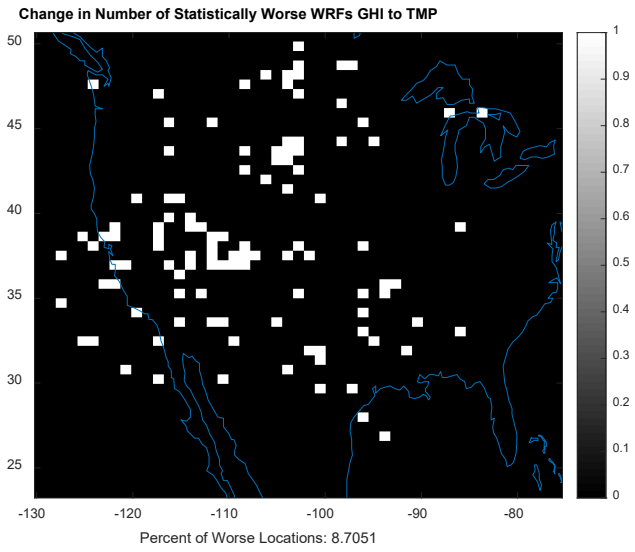
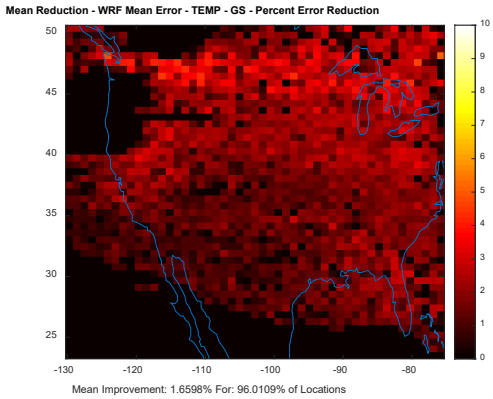


Figure 4-25. Locations with Temperature Statistically Worse.

The identification of at least one location where the temperature training method performs worse requires us to perform the error calculation for the temperature trained model. To accomplish this, the solar forecast generated using testing data with weights from the praxis combined model trained on temperature are individually compared to the

independent WRFs solar forecasts. These comparisons are performed on a raw forecast-by-forecast basis where the error is calculated at each time-location forecast. Figure 4-26 below is generated by using these recordings and calculating the average percentage forecast improvement (or decrease) of the temperature trained model against independent WRFs. The left figure plots percent improvement of the combined model forecast error with respect to the average independent WRF error. The right figure plots the combination model error with respect to the maximum independent WRF error. Based on calculations of the error improvement across the geographic domain, the temperature trained model reduces error by 1.66% on average compared to the average independent model forecast error and 8.59% when compared to average maximum independent error.

Mean Error Reduction



Maximum Error Reduction

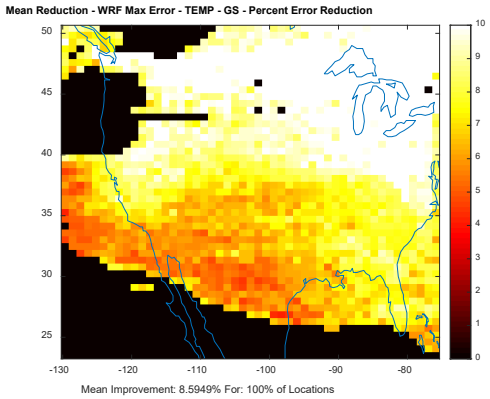


Figure 4-26. Temperature Trained Model Improvement.

Based on these results, the temperature trained combination forecast model has proven to be statistically more error prone than the solar radiation trained combination model for 8.7051% of locations when tested statistically with an alpha of 0.1. When assessing the raw forecast hour-by-hour data, this temperature trained combination model reduces error by 8.59% on average when compared to maximum individual worst WRF

error but only 1.66% when compared to the average individual WRF error. When looking at the WRF Mean Error, 96% of locations experienced an improved result where 100% saw improved forecasts as compared to the WRF Max Error. Based on these results, the null hypothesis is rejected that there is a 5% statistical error improvement when using temperature trained combinations against independent WRFs with an alpha of 0.1. However, from a raw error perspective, we fail to reject the null hypothesis for all cases except when comparing the temperature trained combination to all independent WRFs.

Chapter 5—Discussion and Conclusions

5.1 Discussion

This praxis detailed the background, development and assessment of a solar forecast improvement tool designed to reduce the cost of electrical generation for remote solar panel deployments. The praxis first identified the need and formed the basis of the developed research model through the literature review in Chapter 2. This literature survey identified a variety of existing solar energy forecasting and generic weather forecasting approaches. These approaches included advancements which have been under development across academia and industry which generally took the form of machine learning statistical approaches (Shao, et al., 2016) or numerical model approaches (Verbois, Huva, Rusydi, & Walsh, 2018). The review of these other research efforts identified useful approaches to improving weather forecasting, however none met the specific requirements of the praxis end user. Approaches that were identified either required large quantities of historical data or were for non-solar forecasting problem areas such as wind energy. Based on this literature survey, improvements to solar energy forecasting for individual locations that have less than four years of collected historical ground truth data was an identified gap. Literature associated with previous approaches and adjacent problem areas, while not providing direct solutions, were able to provide useful techniques and approaches to be pulled into the praxis research model. These approaches included those such as the ensemble model combination ones in (Wolf, O'Donncha, & Chen, 2020) which formed the basis of the praxis developed model.

In Chapter 3, a description of the research methodology was presented detailing the weighted linear combination model which was optimized using a GlobalSearch algorithm against a Root Mean Square Error function. This model was formed based on the literature recommended methods which were then validated and tested against the research hypotheses in the results section of Chapter 4.

The assessment of this praxis developed model was completed against the three praxis hypotheses and associated research questions to confirm or reject the efficacy of the model. Confirmation of model performance against these hypotheses will support the integration of this model developed into the end user's power generation deployment tool. This end user tool takes the resulting forecast outputs from the praxis model and determines the amount of solar energy generating capacity that needs to be deployed to remote locations for field experiments. The results identified in Chapter 4 found that the praxis combination forecast model improves the accuracy of solar energy forecasts when assessed under specific hypothesis-based conditions. The three-hypothesis cases and associated research questions assessed each represent a specific use case where the end user has expressed interest in deploying this enhanced model: 1) remote locations where no previous subject matter expert knowledge of best model exist, 2) locations where existing tools underperform, and 3) locations with only historical collection of temperature measurements exist. To test the model against these cases, a variety of both statistical and raw numerical calculation methodologies were selected.

Based on the results calculated in Chapter 4, with these approaches developed into the research praxis tool, the resulting approach demonstrates lower solar energy forecast errors than independent Weather Research and Forecasting (WRF) models.

These results confirm the original research questions which were set-up so that confirmation will ensure the reduction of costs associated with deploying unnecessary excess electricity generating equipment by using the praxis developed model. With confirmation of the first hypothesis, this praxis has demonstrated that a combination model can reduce solar energy prediction error over using just one model. Confirmation of the second hypothesis demonstrates that the combination model improves forecasts in those locations where an independent model performs in the bottom 25th percentile. The initial investigation set out to demonstrate that these poorly performing locations would improve the most, however – this was not found to be true. Finally, this praxis demonstrated through confirmation of the third hypothesis that temperature can be used as a training variable to improve solar energy forecast accuracy for combination forecast models. These results demonstrate the impact that the praxis developed model has at improving forecast accuracy. As utilities have faced a challenge in matching solar energy forecasting capability as solar forecast model development (Zhang, Beaudin, Taheri, Zareipour, & Wood, 2015), this praxis has demonstrated the underlying capabilities needed to close that research gap- ultimately lowering the cost of energy for the end user.

5.2 Conclusions

Weighted linear combinations of the Weather Research and Forecasting (WRF) model trained on solar radiation have demonstrated the ability to improve resulting solar energy forecast errors. This conclusion is backed up by statistical and raw error assessments performed in Chapter 4 which have shown the efficacy of using these types of advanced forecasting approaches across for reducing solar energy forecast error. This approach, while not tested against all potentialities was confirmed for the three

hypotheses using three different end user cases of interest. The one use-case which statistically lagged the others in performance was that the combination model improves forecasts better than a subject matter expert selecting the best independent model for a particular location. Based on the results in Chapter 4, this praxis developed model never met that threshold statistically, and only met that case when assessing using raw calculations for the baseline solar trained model. The other two use-cases assessed, "Better than Half of Independent WRFs and "Better than One Independent WRF", meanwhile represented cases where the praxis model did generate more success and will be detailed below. These two use cases represent cases where the praxis developed model provides an improvement in forecasting errors over an end user selecting either the single worst independent WRF model or doing better than random selection. As the end user does not expect to have a subject matter expert to select the best model for all deployments, these two cases represented the most likely deployments. The results detailed in Chapter 4 confirm through raw and/or statistical methods that for a majority of the hypotheses cases the praxis model is more accurate method than random guess or worst selection. For the end user of this tool, this result means for most deployments without a trained operator, the model will provide more accurate forecasting results.

The forecast model improvements developed and assessed in this praxis were ultimately designed to enable the reduced cost of deploying solar energy generating equipment for remote field-tests. While the total cost savings will ultimately depend on the capacity need for each solar deployment could, the unit cost savings is estimated to be upwards of $0.03 per kilowatt hour of electricity based on (Georgitsioti, et al., 2015) and results from this analysis in Chapter 4. The results here demonstrate that before including

savings from logistics due to the smaller associated deployments, this combination model approach has been validated to improve forecasts and therefore improve costs.

5.3 Contributions to Body of Knowledge

1. Contributed an approach to improve localized solar energy forecasts using only limited amounts of offline ground truth data in the MATLAB software environment.
2. Developed a portable training & operation approach designed to improve solar energy forecasts for a specific location at any point on the globe.
3. Demonstrated and validated an approach to utilize historical offline temperature ground truth measurements to improve localized solar energy forecasts for remote locations in the MATLAB software environment.
4. Confirmed that weighted linear combinations of the WRF model can be used to reduce solar energy forecast error.
5. Proved that weighted combination models of solar forecast increase accuracy even for locations where independent models perform poorly.

5.4 Recommendations for Future Research

This praxis developed research model has satisfied the initial needs of the end user. The literature survey and model development/analysis identified additional gaps in support which are recommended below as future work to continue this development.

Case: Combined Statistical & Numerical Forecast Approach

Recommendation: Develop an integrated model that could leverage statistical forecasting methods for locations where sufficient data exists to run which could use numerical trained model combinations to supplement.

Case: Additional Input Forecast Models

Recommendation: Additional training using additional forecast models, including other variants of the Weather Research and Forecasting model and the European Center for Medium Weather Forecasts (ECMWF). This research could assess importance of 2nd and lower order effects on solar energy models using other WRF parameters or could leverage other differences in model assumptions outside the WRF framework.

Case: Incorporate Additional Training Parameters

Recommendation: Training/Validation using hybrid combinations of additional meteorological parameters to train the ensemble combination model (solar and temperature and humidity, etc.). This type of future research could enable a single model to use any and/or all available data at a location to train from either limited available or best expected result data.

Case: Combination Model Methods

Recommendation: Assess additional combination model approaches – different linear, non-linear, and other machine-learning based combinations for methods to combine multiple independent forecast models.

References

Agnarsson, J., Sunde, M., & Ermilova, I. (2013). *Parallel Optimization in Matlab.* Uppasala Universitet.

Albright, E. A. (2021). *Two Independent Samples Unequal Variance (Welch's Test).* Retrieved from Duke | Nicholas School of the Environment: https://sites.nicholas.duke.edu/statsreview/means/welch/

Allan, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). *Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56.* FAO.

Aminifar, F., Shahidehpour, M., Alabdulwahab, A., Abusorrah, A., & Al-Turki, Y. (2020). The Proliferation of Solar Photovoltaics: Their Impact on Widespread Deployment of Electric Vehicles. *IEEE Electrification*, 79-91.

Antolink, M. (2021, 02 07). *Model Output Statistics - MDL - Virtual Lab*. Retrieved from NOAA - The Meterological Development Labratory: https://vlab.ncep.noaa.gov/web/mdl/mos

Araki, S., Shimadera, H., Yamamoto, K., & Kondo, A. (2014). AIR MONITORING NETWORK OPTIMIZATION METHOD. *Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes.* Varna, Bulgaria: HARMO.

Beersma, J. J., & Buishand, T. A. (1999). A Simple Test for Equality of Variances in Monthly Climate Data. *Journal of Climate*, 1770-1779.

Blaga, R., Sabadus, A., Stefu, N., Dughir, C., Paulescu, M., & Badescu, V. (2019). A current perspective on the accuracy of incoming solar energy forecasting. *Progress in Energy and Combustion Science*, 119-144.

Borkowski, J. J. (2021, 04 24). *Tests for Homogenity of Variance*. Retrieved from Montana State University - Department of Mathematical Sciences: https://math.montana.edu/jobo/st541/sec2e.pdf

Botchkarev, A. (2018). Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology. *ArXiv*.

Box, G. E. (n.d.). Retrieved from https://www.lacan.upc.edu/admoreWeb/2018/05/all-models-are-wrong-but-some-are-useful-george-e-p-box/

Bureau of Meterology. (2021, 01 17). *One Minute Solar Data - List of stations.* Retrieved from Commonwealth of Australia - Bureau of Meteorology: http://www.bom.gov.au/climate/data/oneminsolar/stations.shtml

Bureau of Meterology. (2021, 01 17). *Weather Station Directory.* Retrieved from Commonwealth of Australia - Bureau of Meteorology: http://www.bom.gov.au/climate/data/stations/

Chai, T. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 1247-1250.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 1247-1250.

Chen, Y. (2020). COVID-19 Pandemic Imperils Weather Forecast. *Geophysical Research Letters*.

Decremer, D., Chung, C. E., Ekman, A. M., & Brandefelt, J. (2014). Which significance test performs the best in climate simulations? *Tellus A: Dynamic Meteorology and Oceanography*, 23139.

Di, Z., Duan, Q., Shen, C., & Xie, Z. (2020). Improving WRF Typhoon Precipitation and Intensity Simulation Using a Surrogate-Based Automatic Parameter Optimization Method. *Atmosphere*.

Du, P. (2019). Ensemble Machine Learning-Based Wind Forecasting to Combine NWP Output With Data From Weather Station. *IEEE Transactions on Sustainable Energy*, 2133-2141.

Eberly College of Science. (2021, 03 21). *10.2.1 - ANOVA Assumptions*. Retrieved from Penn State: https://online.stat.psu.edu/stat500/lesson/10/10.2/10.2.1

ECMWF. (2021, 02 13). *Licenses available*. Retrieved from European Centre for Medium-Range Weather Forecasts: https://www.ecmwf.int/en/forecasts/accessing-forecasts/licences-available

El-Kenawy, E.-S. M., Ibrahim, A., Mirjalili, S., Eid, M., & Hussein, S. (2020). Novel Feature Selection and Voting Classifier Algorithms for COVID-19 Classification in CT Images. *IEEE Access*, 179317-179335.

Eschenbach, A., Yepes, G., Tenllado, C., Gomez-Perez, J. I., Pinuel, L., Zarazalejo, L. F., & Wilbert, S. (2020). Spatio-Temporal Resolution of Irradiance Samples in Machine Learning Approaches for Irradiance Forecasting. *IEEE Access*, 51518-51531.

Fader, P. S., & Hardie, B. G. (2005). The value of simple models in new product forecasting and customer-base analysis. *Applied Stochastic Models in Business and Industry*, 461-473.

Floating-Point Numbers. (2021, 02 20). Retrieved from MathWorks: https://www.mathworks.com/help/matlab/matlab_prog/floating-point-numbers.html

Frnda, J., Durica, M., Nedoma, J., Zabka, S., Martinek, R., & Kostelansky, M. (2049). A Weather Forecast Model Accuracy Analysis and. *Sensors*.

Gaba, A., Testlin, I., & Winkler, R. L. (2017). Combining interval forecasts. *Decicion Analysis*, 1-20.

George, D., & Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference.* Boston, MA: Pearson.

Georgitsioti, T., Pillai, G., Pearsall, N., Putrus, G., Forbes, I., & Anand, R. (2015). Short-term perfrmance variations of different photovoltaic system technologies under the humid subtropical climate of Kanput in India. *IET Renewable Power Generations*.

Gerstman, B. B. (2021, 03 21). *7: Paired Samples*. Retrieved from San Jose State University: https://www.sjsu.edu/faculty/gerstman/StatPrimer/paired.pdf

Glahn, H. R., & Lowry, D. A. (1972). The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. *Journal of Applied Meterology and Climatology*.

Global Forecast System (GFS). (2021, 05 08). Retrieved from NOAA National Centers for Environmental Information: https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs

Glover, F., Lasdon, L., Plummer, J., Duarte, A., Marti, R., Laguna, M., & Rego, C. (2011). Pseudo-Cut Strategies for Global Optimization. *International Journal of Applied Metaheuristic Computing*, 1-12.

Goldberg, D. E., & Holland, J. H. (1988). Genetic Algorithms and Machine Learning. *Machine Learning*, pp. 95-99.

Gradišar, D., Grašič, B., Božnar, M., Mlakar, P., & Kocijan, J. (2016). Improving of local ozone forecasting by integrated models. *Environmental Science and Pollution Research*, 18439-18450.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, 337-350.

Halquist, J. B. (2006). Use of HPC QPF confidence interval forecasts to produce an ensemble of river forecasts. *20th Conference on Hydrology.* Atlanta, GA: American Meterological Society.

Hamill, T. M. (1999). Hypothesis Tests for Evaluating Numerical Precipitation Forecasts. *Weather and Forecasting*, 155-167.

Hanke, J. E., & Wichern, D. (2014). *Business Forecasting.* Essex: Pearson New International.

Hewage, P., Trovati, M., Pereira, E., & Behera, A. (2020). Deep learning-based effective fine-grained weather forecasting model. *Pattern Analysis and Applications*, 343-366.

Higham, N. J. (2009). *Cholesky factorization.* John Wiley & Sons, Inc.

High-Performance Portable MPI. (2021, 03 06). Retrieved from MPICH: https://www.mpich.org/

Holmstorm, M., Liu, D., & Vo, C. (2016). *Machine Learning Applied to Weather Forecasting.* Stanford University.

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., & Zareipour, H. (2020). Energy Forecasting: A Review and Outlook. *IEEE Power and Energy*, 376-388.

Horst, R., & Tuy, H. (1990). *Global Optimization: Deterministic Approaches.* New York & Tokyo: Springer.

Janjic, T., Ruckstuhl, Y., & Toint, P. L. (2019). An algorithm for optimization with disjoint linear constraints and its application for predicting rain. *arXiv*.

Jankov, I., Gallus Jr., W. A., Segal, M., & Koch, S. E. (2007). Influence of Initial Conditions on the WRF–ARW Model QPF Response to Physical Parameterization Changes. *Weather and Forecasting*, 501-519.

Jee, J.-B., & Kim, S. (2017). Sensitivity Study on High-Resolution WRF Precipitation Forecast for a Heavy Rainfall Event. *Atmosphere*, 96.

Jewson, S., & Caballero, R. (2003). Seasonality in the statistics of surface air temperature and the pricing of weather derivatives. *Meteorolgical Applications*, 367-376.

Jimenez, P. A., Hacker, J. P., Dudhia, J., Haupt, S. E., Ruiz-Arias, J. A., Gueymard, C. A., . . . Deng, A. (2016). WRF-SOLAR Description and Clear-Sky Assessment of an Augmented NWP Model for Solar Power Prediction. *Bulletin of the American Meterological Society*, 1249-1264.

Kim, C. K., Kim, H.-G., Kang, Y. H., & Yun, C. (2018). Toward Improved Solar Irradiance Forecasts: Evaluation of Operational Numerical Weather Prediction

model for Solar Irradiance over the Korean Peninsula. *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)* (pp. 2317-2319). Waikoloa: IEEE.

Kim, H.-Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry & Endodontics*, 52-54.

Klein, J., & Zachmann, G. (2005). Interpolation Search for Point Cloud Intersection.

Kornerup, P., Lauter, C., Lefèvre, V., & Louver, N. (2010). Computing Correctly Rounded Integer Powers. *ACM Transactions on Mathematical Software*.

Kotamarthi, R., Mearns, L., Hayhoe, K., Castro, C. L., & Wuebbles, D. (2016). *Use of Climate Information for Decision-Making and Impacts Research: State of Our Understanding.* U.S. Government - Strategic Environmental Research and Development Program.

Kraemer, F. A., Palma, D., Braten, A. E., & Ammar, D. (2020). Operationalizing Solar Energy Predictions for Sustainable, Autonomous IoT Device Management. *IEEE Internet of Things*, 11803 - 11814.

Kumar, P., Kishtawal, C., & Pal, P. K. (2015). Impact of ECMWF, NCEP, and NCMRWF global model analysis on the WRF model forecast over Indian Region. *Theoretical and Applied Climatology*, 127.

Kyriakides, E., & Polycarpou, M. (n.d.). Short Term Electric Load Forecasting: A Tutorial. In *Trends in Neural Computation* (pp. 391-418). Berlin: Springer, Berlin, Heidelberg.

Lamas, G. (2016). Energy supply in isolated areas: an outline of the current situation and the potential of hydrogen technologies for distributed power generation. *Global Resource Management*, 31-53.

Lange, M., & Focken, U. (2008). New developments in wind energy forecasting. *IEEE Power and Energy*, 1-8.

Li, W., Ren, H., Chen, P., Wang, Y., & Qi, H. (2020). Key Operational Issues on the Integration of Large-Scale Solar Power Generation—A Literature Review. *Energies | Photovoltaic Systems: Modelling, Control, Design and Applications*.

Liang, Y., Niu, D., Zhou, W., & Fan, Y. (2018). Decomposition Analysis of Carbon Emissions from Energy Consumption in Beijing-Tianjin-Hebei, China: A Weighted-Combination Model Based on Logarithmic Mean Divisia Index and Shapley Value. *Sustainability - MDPI*, 2535.

Lorenz, E., Hurka, J., Heinemann, D., & Beyer, H. G. (2009). Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems. *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING*.

Ma, J., Yu, Z., Qu, Y., Xu, J., & Cao, Y. (2020). Application of the XGBoost Machine Learning Method in PM2.5 Prediction: A Case Study of Shanghai. *Aerosol and Air Quality Research*.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS One*.

Mandi, J., & Guns, T. (2020). Interior Point Solving for LP-based prediction+optimisation. *arXiv*.

Martí, R., Laguna, M., & Campos, V. (2005). Scatter Search vs. Genetic Algorithms. In R. Sharda, S. Voß, C. Rego, & B. Alidaee, *Metaheuristic Optimization via Memory and Evolution* (pp. 263-282). Norwell, Massachusetts: Springer Science & Business Media.

Mathiesen, P., Collier, C., & Kleissl, J. (2013). A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance. *Solar Energy*.

Mathworks. (2011). *Kruskal-Wallis*. Retrieved from Matlab Help Center: https://www.mathworks.com/help/stats/kruskalwallis.html#btv4oqy-10

MathWorks. (2020, 12 13). *Comparison of Six Solvers*. Retrieved from https://www.mathworks.com/help/gads/example-comparing-several-solvers.html

Mathworks. (2021, 04 25). *"vartestn" - Multiple-sample tests for equal variances*. Retrieved from MathWorks Help Center: https://www.mathworks.com/help/stats/vartestn.html

MathWorks. (2021, 05 02). *Constrained Nonlinear Optimization Algorithms*. Retrieved from Matlab Help Center: https://www.mathworks.com/help/optim/ug/constrained-nonlinear-optimization-algorithms.html#brnpd5f

MathWorks. (2021, 05 02). *fmincon*. Retrieved from Matlab Help Center: https://www.mathworks.com/help/optim/ug/fmincon.html?s_tid=doc_ta

MathWorks. (2021, 02 20). *How the Genetic Algorithm Works*. Retrieved from MATLAB - Help Center: https://www.mathworks.com/help/gads/how-the-genetic-algorithm-works.html#:~:text=The%20following%20outline%20summarizes%20how,to%20create%20the%20next%20population.

MathWorks. (2021, 03 21). *Matlab*. Retrieved from Local vs. Global Optima: https://www.mathworks.com/help/optim/ug/local-vs-global-optima.html

Mathworks. (2021, 03 21). *Optimization Toolbox*. Retrieved from MathWorks: https://www.mathworks.com/products/optimization.html

Meacham, F., Boffelli, D., Dhahbi, J., Martin, D. I., Singer, M., & Pachter, L. (2011). Identification and correction of systematic error in high-throughput sequence data. *BMC Bioinformatics*.

Meyer, D., Schoetter, R., Riechert, M., Verrelle, A., Tewari, M., Dudhia, J., . . . Grimmond, S. (2020). WRF-TEB: Implementation and Evaluation of the Coupled Weather Research and Forecasting (WRF) and Town Energy Balance (TEB) Model. *Journal of Advances in Modeling Earth Systems*.

Missouri Department of Natural Resources. (2021, 05 02). *How Much Solar Energy Is Available?* Retrieved from Division of Energy: https://energy.mo.gov/solar/how-much-solar-energy-is-available

Moeng, C. H., Dudhia, J., Klemp, J., & Sullivan, P. (2007). Examining Two-Way Grid Nesting for Large Eddy Simulation of the PBL Using the WRF Model. *Monthly Weather Review*, 2295-2311.

Mutavhatsindi, T., Sigauke, C., & Mbuvha, R. (2020). Forecasting Hourly Global Horizontal Solar Irradiance in South Africa Using Machine Learning Models. *IEEE Access*, 198872 - 198885.

Myttenaere, A. D., Golden, B., Le Grabd, B., & Rossi, F. (2016). Mean Absolute Percentage Error for regression models. *Neurocomputing*, 38-48.

Nagendra, N. P., Narayanmurthy, G., Moser, R., & Singh, A. (2020). Open Innovation Using Satellite Imagery for Initial Site Assessment of Photovoltaic Projects. *IEEE Transactions on Engineering Managment*.

NASA. (2021, 03 14). *Glory - Understanding Earth's Energy Balance*. Retrieved from NASA - Total Solar Irradiance: https://www.nasa.gov/mission_pages/Glory/solar_irradiance/total_solar_irradiance.html

NASA. (2021, 02 07). *Terra & Aqua Moderate Resolution Imaging Spectroradiometer (MODIS)*. Retrieved from NASA - Earth Data: https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/modis/

National Centers for Environmental Information. (2021, 03 20). *Numerical Weather Prediction*. Retrieved from National Oceanic and Atmospheric Admistration: https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/numerical-weather-prediction

National Ocean Service - U.S. Government. (2021, 02 07). *NOAA 200th Foundations: Weather, Ocean, and Climate Prediction*. Retrieved from National Oceanic and Atmospheric Administration: https://celebrating200years.noaa.gov/foundations/numerical_wx_pred/welcome.html#:~:text=The%20roots%20of%20numerical%20weather,the%20father%20of%

20modern%20meteorology.&text=A%20British%20mathematician%20named%20Lewis,procedures%20to%20solve%20these%20equ

National Renewable Energy Labratory. (2021, 03 06). *Baseline Measurement System*. Retrieved from Measurement and Instrumentation Data Center: http://midcdmz.nrel.gov/apps/sitehome.pl?site=BMS

National Weather Service. (2021, 03 21). *Climate Time Series - Random Variables*. Retrieved from Weather.Gov: https://training.weather.gov/pds/climate/pcu2/statistics/Stats/part1/CTS_random.htm#:~:text=For%20example%2C%20temperature%20has%20a,is%20measured%20on%20interval%20scale.

NCAR. (2020, 07 10). *How to Compile WRF*. Retrieved from National Center for Atmospheric Research: https://www2.mmm.ucar.edu/wrf/OnLineTutorial/compilation_tutorial.php

NCSS Statistical Software. (2021, 03 21). Chapter 206 - Two-Sample T-Test.

Nievergelt, J. (2000). Exhaustive Search, Combinatorial Optimization and Enumeration: Exploring the Potential of Raw Computing Power. *Lecture Notes in Computer Science*, 87-125.

NOAA. (2021, 03 07). *Data Access*. Retrieved from National Centers for Environmental Information: https://www.ncdc.noaa.gov/data-access

NOAA. (2021, 05 06). *Earth Systems Research Labratory*. Retrieved from U.S. Department of Commerce: https://esrl.noaa.gov/gsd/wrfportal/namelist_input_options.html

NOAA. (2021, 03 04). *NCEP Central Optations*. Retrieved from National Weather Service: https://www.nco.ncep.noaa.gov/pmb/products/gfs/#GFS

NOAA. (2021, 02 12). *WRF Portal*. Retrieved from NOAA - Earth Systems Research Labratory: https://esrl.noaa.gov/gsd/wrfportal/namelist_input_options.html

Ohring, G., Lord, S., Deber, J., Mitchell, K., & Ji, M. (2002). Applications of satellite remote sensing in numerical weather and climate prediction. *Advances in Space Research*, 2433-2439.

Ouma, Y. O., Okuku, C. O., & Njau, E. N. (2020). Use of Artificial Neural Networks and Multiple Linear Regression Model for the Prediction of Dissolved Oxygen in Rivers: Case Study of Hydrographic Basin of River Nyando, Kenya. *Open Access*.

Pan, B. (2018). Application of XGBoost algorithm in hourly PM2.5 concentration Prediction. *IOP Conference Series: Earth and Environmental Science.*

Peters, D., Kilper, T., Calais, M., Jamal, T., & von Maydell, K. (2018). Solar Short-Term Forecasts for Predictive Control of Battery Storage Capacities in Remote PV Diesel Networks. In *Transition Towards 100% Renewable Energy* (pp. 325-333).

Physical Sciences Labratory. (2021, 02 14). *Distributions of Daily Meteorological Variables*. Retrieved from National Oceanic and Atmospheric Administration: https://psl.noaa.gov/data/writ/distributions/background/index.html#:~:text=Often%20meteorological%20data%20is%20normally,towards%20the%20left%20(cold).&text=Distributions%20other%20than%20Gaussian%20have%20been%20used%20to%20model%20data.

Plaxco, J., Valdes, A., & Stojko, W. (2021, 02 20). *Interior-point method for LP*. Retrieved from Northwestern University Open Text Book on Process Optimization: https://optimization.mccormick.northwestern.edu/index.php/Interior-point_method_for_LP

Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., . . . Duda, M. G. (2017). The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions. *Bulletin of the American Meteorological Society*, 1717-1737.

P-Tree System. (2021, 03 20). *JAXA Himawari Monitor*. Retrieved from Japan Meteorological Agency: https://www.eorc.jaxa.jp/ptree/userguide.html

Rana, M., Koprinska, I., & Agelidis, V. G. (2016). Solar Power Forecasting Using Weather Type Clustering and Ensembles of Neural Networks. *2016 International Joint Conference on Neural Networks (IJCNN).* Vancouver, BC, Canada: IEEE.

Rodrigo, J. S., Chávez Arroyo, R. A., Moriarty, P., Churchfield, M., Kosović, B., Réthoré, P.-E., . . . Rife, D. (2019). Mesoscale to microscale wind farm flow modeling and evaluation. *WIREs Energy and Environment*.

Ryu, A., Ito, M., Ishii, H., & Hayashi, Y. (2019). Preliminary Analysis of Short-term Solar Irradiance Forecasting by using Total-sky Imager and Convolutional Neural Network. *2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia).* Bangkok, Thailand: IEEE.

Sawtooth Software. (2021, 03 21). *Search Algoritms*. Retrieved from Sawtooth Software: https://sawtoothsoftware.com/help/lighthouse-studio/manual/optimization_introduction_2.html

School of Engineering and Applied Science. (2021, 03 21). *OFRE - Optimization*. Retrieved from Princeton University: https://orfe.princeton.edu/research/optimization

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions About Sample Size: The Empirical Law of Large Numbers. *Journal of Behavioral Decision Making*.

Segarra, E. L., Du, H., Ruiz, G. R., & Bandera, C. F. (2019). Methodology for the Quantification of the Impact of Weather Forecasts in Predictive Simulation Models. *Energies*.

Selby, D., Cascella, A., Gardiner, K., Do, R., Moravan, V., Meyers, J., & Chow, E. (2009). A single set of numerical cutpoints to define moderate and severe symptoms for the Edmonton Symptom Assessment System. *J Pain Symptom Manage*, 241-249.

Seo, B. C. (2020). A Data-Driven Approach for Winter Precipitation Classification Using Weather Radar and NWP Data. *Atmosphere*.

Shakya, A., Semhar, M., Saunders, C., Armstrong, D., Pandey, P., Chalise, S., & Tonkoski, R. (2018). Solar Irradiance Forecasting in Remote Microgrids Using Markov Switching Model. *2018 IEEE Power & Energy Society General Meeting (PESGM).* Portland, Oregon: IEEE.

Shao, X., Lu, S., van Kessel, T. G., Hamann, H. F., Daehler, L., Cwagenberg, J., & Li, A. (2016). Solar irradiance forecasting by machine learning for solar car races. *2016 IEEE International Conference on Big Data (Big Data).* Washington, DC: IEEE.

Siewert, J., & Kroszczynski, K. (2020). GIS Data as a Valuable Source of Information for Increasing Resolution of the WRF Model for Warsaw. *Remote Sensing*.

Siuta, D., West, G., & Stull, R. (2017). WRF Hub-Height Wind Forecast Sensitivity to PBL Scheme, Grid Length, and Initial Condition Choice in Complex Terrain. *Weather and Forecasting*, 493-509.

Skamarock, W. C., Klemp, J. B., Dudhiac, J., Gill, D. O., Liu, Z., Berner, J., . . . Huang, X.-y. (2019). *A Description of the Advanced Research WRF Model Version 4.* NCAR Technical Notes.

Soares de Araujo, J. M. (2020). Performance comparison of solar radiation forecasting between WRF and LSTM in Gifu, Japan. *Environmental Research Communications*.

Solar Energy Industries Association. (2020). *Solart Market Insight Report.* Washington, DC.

Solar Forecasting 2 - Solar Energy Technologies Office. (2021, 05 12). Retrieved from U.S. Department of Energy - Office of Energy Efficiency & Renewable Energy: https://www.energy.gov/eere/solar/solar-forecasting-2

Stoffel, T., & Andreas, A. (2015). *NREL Solar Radiation Research Laboratory (SRRL): Baseline Measurement System (BMS)*. Retrieved from National Renewable Energy Laboratory: https://data.nrel.gov/submissions/7

Stull, R. B. (2003). *An Introduction to Boundary Layer Meterology.* Dorderecht, NL: Kluwer Academic Publishers.

Sullivan, L. (2021, 04 23). *Power and Sample Size Determination*. Retrieved from Boston University School of Public Health: https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_power/bs704_power_print.html#:~:text=The%20formula%20for%20determining%20sample%20size%20to%20ensure%20that%20the,%3D%200.975%20and%20Z%3D1.960.

Taylor, M., Ralon, P., Anuta, H., & Al-Zoghoul, S. (2020). *Renewable Power Generation Costs in 2019.* Abu Dhab: International Renewable Energy Agency.

Tiwari, S., Sabzehgar, R., & Rasouli, M. (2018). Short Term Solar Irradiance Forecast Using Numerical Weather Prediction (NWP) with Gradient Boost Regression. *2018 9th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG).* Charlotte, NC: IEEE.

Tolstykh, M. (2005). Some Current Problems in Numerical Weather Prediction. *Izvestiya Atmospheric and Oceanic Physics*.

Tuohy, A., Zack, J., Haupt, S. E., Sharp, J., Ahlstrom, M., Dise, S., . . . Collier, C. (2015). Solar Forecasting: Methods, Challenges, and Performance. *IEEE Power and Energy*, 50-59.

U.S. Energy Information Administration. (2020). *Annual Energy Outlook 2020 - Office of Energy Analysis.* Washington, DC: U.S. Department of Energy.

UC Berkeley EECS. (2021, 05 02). *EECS 141.* Retrieved from UC Berkeley: http://bwrcs.eecs.berkeley.edu/Classes/icdesign/ee141_s03/Project/Project1_solutions/fmincon.pdf

UCAR. (n.d.). *Network Common Data Form (NetCDF)*. Retrieved from unidata.

Uchida, J., Mori, M., Hara, M., Satoh, M., Goto, D., Kataoka, T., . . . Nakajima, T. (2017). Impact of Lateral Boundary Errors on the Simulation of Clouds with a Nonhydrostatic Regional Climate Model. *Monthly Weather Review*, 5059-5082.

Ulmer, F.-G., Adam, N., & Eineder, M. (2012). Ensemble based atmospheric phase screen estimation using least squares. *IEEE International Geoscience and Remote Sensing Symposium.* Munich, DE: IEEE.

US Department of Energy. (2019). *Solar Energy Technologies Office - SunShot 2030.*

US Department of Energy. (2021, 03 07). *Solar Radiation Basics*. Retrieved from Office of Energy Efficiency & Renewable Energy: https://www.energy.gov/eere/solar/solar-radiation-basics

van Bussel, L. G., Ewert, F., Zhao, G., Hoffmann, H., Enders, A., Wallach, D., . . . Tao, F. (2016). Spatial sampling of weather data for regional crop yield simulations. *Agricultural and Forest Meteorology*, 101-115.

Verbois, H., Huva, R., Rusydi, A., & Walsh, W. (2018). Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning. *Solar Energy*, 265-277.

Versta Research. (2021, 03 21). *Why You Need Raw Data*. Retrieved from Versta Research: https://verstaresearch.com/blog/why-you-need-your-raw-data/

Vignola, F., Peterson, J., Kessler, R., Mavromatakis, F., Dooraghi, M., & Sengupta, M. (2016). Use of Pyranometers to Estimate PV Module Degradation Rates in the Field. *Presented at the 43rd IEEE Photovoltaic Specialists Conference.* Portland, OR: National Renewable Energy Labratory.

Voorhis, C. R., & Morgan, B. L. (2007). Understanding Power and Rules of Thumb for Determining Sample Sizes. *Tutorials in Quantitative Methods for Psychology*, 43-50.

Wang, Y., Steele, T., & Zhang, E. (2016). *QQ Plot.* Retrieved from The University of Illinois: https://math.illinois.edu/system/files/inline-files/Proj9AY1516-report2.pdf

Wang, Z., Koprinska, I., & Rana, M. (2017). Solar Power Forecasting Using Pattern Sequences. *Lecture Notes in Computer Science.* IEEE.

Weise, T., Wu, Y., Chiong, R., & Tand, K. (2016). Global versus local search: the impact of population sizes on evolutionary algorithm performance. *Journal of Global Optimization*.

Weyn, J. A., & Caruna, R. (2020). Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere. *Journal of Advancs in Modeling Earth Systems*.

White, B. G., Paegle, J., Steenburgh, W. J., Horel, J. D., Swanson, R. T., Cook, L. T., & Miles, J. G. (1999). Short-Term Forecast Validation of Siz Models. *American Meterological Society*.

Wilson VanVoorhis, C. R., & Morgan, B. L. (2007). Understanding Power and Rules of Thumb for Determining Sample Size. *Tutorials in Quantitative Methods for Psychology*.

Wilson, G. C., Harding, S., Hoeber, O., Devillers, R., & Banzhaf, W. (2012). Parallel exhaustive search vs. evolutionary computation in a large real world network search space. *Evolutionary Computation.* IEEE.

Wolf, S., O'Donncha, F., & Chen, B. (2020). Statistical and machine learning ensemble modeling to forecast sea surface temperature. *Journal of Marine Systems*, 208.

Xu, C., Qu, J. J., Hao, X., Zhu, Z., & Gutenberg, L. (2020). Surface soil temperature seasonal variation estimation in a forested area using combined satellite observations and in-situ measurements. *International Journal of Applied Earth Observation and Geoinformation*, 102156.

Yengejeh, H. H., Shahnia, F., & Islam, S. M. (2014). Contributions of Single–Phase Rooftop PVs on Short Circuits Faults in Residential Feeders. *Australasian Universities Power Engineering Conference* (pp. 1-6). Perth: AUPEC 2014.

Yin, Y., Alves, O., & Oke, P. R. (2011). An Ensemble Ocean Data Assimilation System for Seasonal Prediction. *Monthly Weather Review*, 786-808.

Zhang, J., Florita, A., Hodge, B.-M., Lu, S., Hamann, H. F., Banunarayanan, V., & Brockway, A. M. (2014). A suite of metrics for assessing the performance of solar power forecasting. *Solar Energy*, 111.

Zhang, X., Li, Y., Lu, S., Hamann, H. F., Hodge, B.-M., & Lehman, B. (n.d.). A Solar Time Based Analog Ensemble Method for Regional Solar Power Forecasting.

Zhang, Y., Beaudin, M., Taheri, R., Zareipour, H., & Wood, D. (2015). Day-Ahead Power Output Forecasting for. *IEEE TRANSACTIONS ON SMART GRID*, 2253-2262.

Zhao, J., Guo, Z.-H., Su, Z., & Zhao, Z.-Y. (2015). An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed. *Applied Energy*, 162.

Appendix A

Figure A-1. GlobalSearch vs. Genetic Algorithm Error Distributions

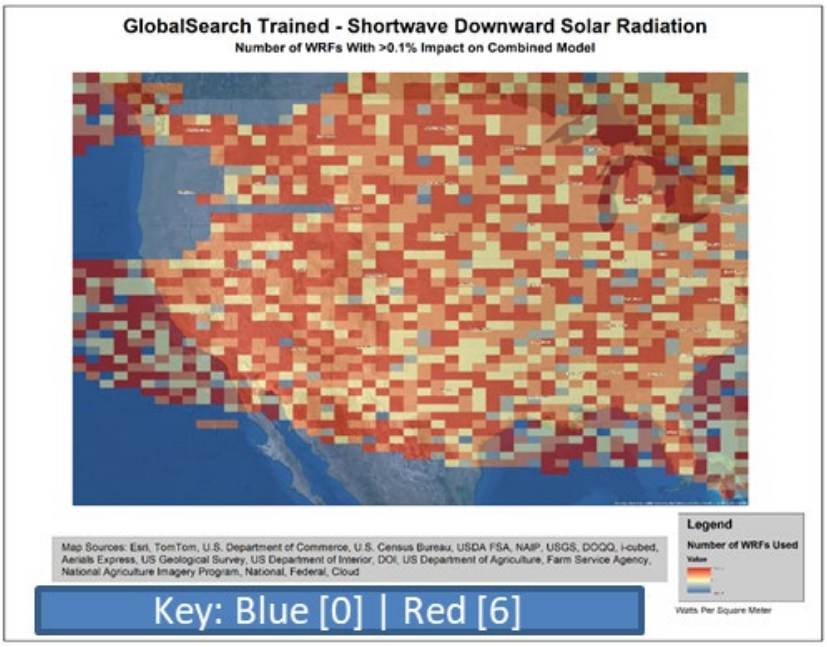


Figure A-2. Number of WRFs Integrated per Site of Interest

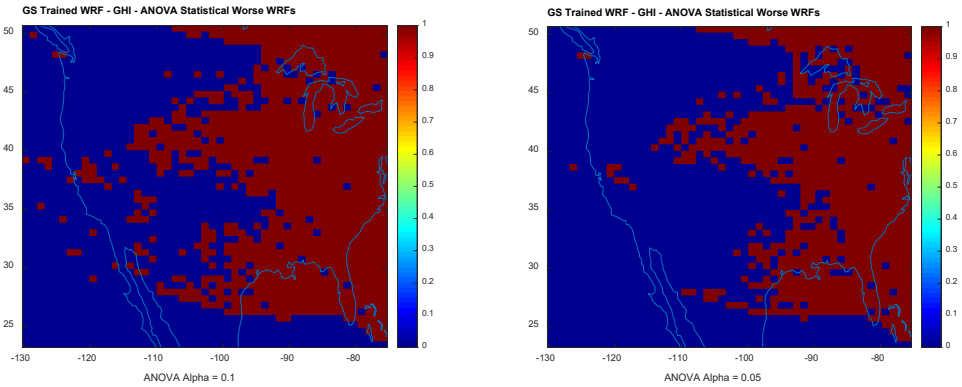


Figure A-3. Comparison of Alphas [0.05 vs. 0.1]

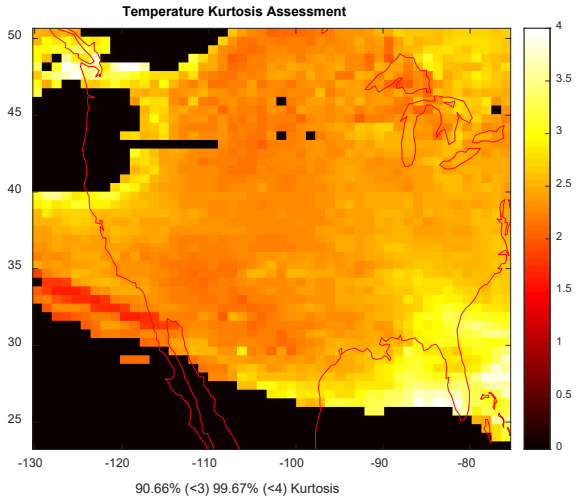


Figure A-4. Temperature Data Kurtosis.

Table A-1. Matlab R2019a Optimization Toolbox – Solver Options

Method	Function	Optimization	Error[1]	Fn Calls*	Other
Interior-Point Search	fmincon	Local Minima	12.9	15	N/A
Adaptive-Mesh	patternsearch	Local Minima	5	174	N/A
Genetic Algorithm	ga	Global Minima	0	9453	N/A
Particle-Velocity Evolution	particleswarm	Local Minima[2]	1	1140	N/A
Interpolate Objective Function	surrogateopt	Global Minima	0	200[3]	Constraint Issue[4]
Scatter-Search with Scoring Algorithm	GlobalSearch	Global Minima	0	2178	N/A

Note. Values highlighted in red represent issues with the use/implementation of the function based on the requirements and end user system capabilities.

(Weise, Wu, Chiong, & Tand, 2016) (MathWorks, 2020)

[1] Function evaluation given an optimal value of 0 (MathWorks, 2020)
[2] Subject to sticking (MathWorks, 2020)
[3] surrogateopt function calls required for optimization are based on the optimization problem (Weise, Wu, Chiong, & Tand, 2016) (MathWorks, 2020)
[4] surrogateopt designed for complex non-linear objective functions and constraints. (Di, Duan, Shen, & Xie, 2020)

Appendix B – Model Data Pre-Requisites

1) Computing Environment
   - Computer: Dell Precision 5510
   - Processor: Intel Xeon E3-1505M 2.8GHZ
   - Memory: 32GB RAM
   - Data Storage (Onboard): 512GB PCIE SSD
   - Data Storage (Offboard): Western Digital My Passport 2TB USB3.0
   - Operating System: Ubuntu 18.04 Desktop
     - OS Source: https://ubuntu.com/download/desktop
2) Weather Research and Forecasting Model Environment

- Installation Guide:
  https://www2.mmm.ucar.edu/wrf/OnLineTutorial/compilation_tutorial.php
- Ubuntu Added Packages [installed using command "apt-get install <name>" from within the Ubuntu Terminal]: gfortran, cpp, gcc, m4
- Pre-Requisite Software Packages [Installed using above "Installation Guide"]
  - MPICH 3.0.4:
    https://www2.mmm.ucar.edu/wrf/OnLineTutorial/compile_tutorial/tar_files/mpich-3.0.4.tar.gz
  - NETCDF 4.1.3:
    https://www2.mmm.ucar.edu/wrf/OnLineTutorial/compile_tutorial/tar_files/netcdf-4.1.3.tar.gz

- JASPERLIB 1.900.1: https://www2.mmm.ucar.edu/wrf/OnLineTutorial/compile_tutorial/tar_files/jasper-1.900.1.tar.gz
- LIBPNG 1.2.50: https://www2.mmm.ucar.edu/wrf/OnLineTutorial/compile_tutorial/tar_files/libpng-1.2.50.tar.gz
- ZLIB 1.2.7: https://www2.mmm.ucar.edu/wrf/OnLineTutorial/compile_tutorial/tar_files/zlib-1.2.7.tar.gz

- Weather Research and Forecasting Model version 3.8 [WPS and WRF Packages]
  - Source: http://www2.mmm.ucar.edu/wrf/users/download/get_source.html
  - Geographic Data (Complete Dataset): https://www2.mmm.ucar.edu/wrf/users/download/get_sources_wps_geog_V3.html
  - Weather Condition Initialization Data (GFS 1 Degree GRID 3): https://www.ncei.noaa.gov/has/HAS.FileAppRouter?datasetname=GFS3&subqueryby=STATION&applname=&outdest=FILE

3) Praxis Research Model Environment

Mathworks Matlab R2019a

Toolboxes: Simulink, Antenna Toolbox, Communications Toolbox, Curve Fitting Toolbox, DSP Systems Toolbox, Data Acquisition Toolbox, Deep Learning Toolbox, Fuzzy Logic Toolbox, Global Optimization Toolbox, Image Processing Toolbox, Mapping Toolbox, Optimization Toolbox, Parallel

Computing Toolbox, Phased Array System Toolbox, RF Toolbox, Sensor Fusion and Tracking Toolbox, Signal Processing Toolbox, Statistics and Machine Learning Toolbox, Symbolic Math Toolbox, System Identification Toolbox, and Wavelet Toolbox.

4) Truth Data

- Solar Radiation & Temperature Data (Ground Truth) : https://nsrdb.nrel.gov/

Appendix C – Model Parameter Selection

Parameters: https://esrl.noaa.gov/gsd/wrfportal/namelist_input_options.html

In running the Weather Research and Forecasting model, six variants are used for this praxis. These six are based on three underlying physics assumptions (below) with two selected options each. These three approaches are selected because of the impact of each on the calculation of cloud cover and transfer of solar energy through the atmosphere. These six phsics models have been selected because they are the approaches which are utilized by the National Weather Service' NAM Model. (operational) (NOAA, 2021)

Table C-1. WRF Calculation Equation Selections

Parameter	Description	#	Physics Models	Impact
mp_physics	NMM Tested - NEW	6	WSM 6-class graupel scheme	Cloud Formation
	NMM Operational - OLD	5	Ferrier scheme	
ra_lw_physics	NMM Tested - NEW	1	RRTM scheme	Longwave Solar Energy
	NMM Operational - OLD	99	GFDL scheme	
ra_sw_physics	NMM Tested - NEW	1	Dudhia scheme	Shortwave Solar Energy
	NMM Operational - OLD	99	GFDL scheme	