ANALYZING THE EFFICIENCY OF HORIZONTAL PHOTOVOLTAIC CELLS AT AIR FORCE INSTALLATIONS IN VARIOUS CLIMATE REGIONS

THESIS

Parker A. Hines, 2d Lt, USAF

AFIT-ENV-MS-19-M-179

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-MS-19-M-179

ANALYZING THE EFFICIENCY OF HORIZONTAL PHOTOVOLTAIC CELLS AT AIR FORCE INSTALLATIONS IN VARIOUS CLIMATE REGIONS

THESIS

Presented to the Faculty

Department of Engineering Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Engineering Management

Parker A. Hines
2d Lt, USAF

March 2019

Abstract

The U.S. Air Force relies heavily on electrically powered facilities to ensure mission-critical capabilities can be carried out. Without a resilient electrical system in place, the U.S. Air Force could be without power for extended periods, resulting in severe implications. Installing a photovoltaic array on an installation is one strategy to improve energy resiliency. The amount of power a photovoltaic array can produce is reliant on its geographical location, position and weather characteristics. This research presents the development of novel linear regression models based upon 14 case studies from global Department of Defense (DoD) installations to predict horizontal photovoltaic power output. The model incorporates Köppen-Geiger climate classifications with location-specific weather and geographical variables to predict horizontal photovoltaic power production. Both Köppen-Geiger climate and weather variables were determined to provide added value to the model. From the analysis, the ideal climate classification was determined to be the Cfb or a fully humid, warm temperate area with warm summers. Additionally, a goodness-of-fit of the full and reduced models was conducted on a validation dataset. This analysis determined that weather variables were able to account for 22% more variation within the validation set compared to climate variables.

Acknowledgments

I would like to thank my parents who have always encouraged me and instilled in me a hard work ethic at a young age. Without them and other family members I would not be where I am today.

Parker A. Hines

Table of Contents

Page

List of Figures

Page

List of Tables

Page

List of Equations

Page

ANALYZING THE EFFICIENCY OF PHOTOVOLTAIC CELLS AT AIR FORCE INSTALLATIONS IN VARIOUS CLIMATE REGIONS

I. INTRODUCTION

Background

Daily operations for the United States Air Force relies heavily on energy consumption including electrical power. If the power becomes severed, it could severely impact or halt daily operations leading to grave consequences to the United States and its allied nations. Due to this essential energy demand, the Air Force has increased its goal to promote and increase its energy resiliency [1]. Resilience, as defined in Executive Order 13693, means "the ability to anticipate, prepare for, and adapt to changing conditions and withstand, respond to, and recover rapidly from energy disruptions" [2]. Resilience incorporates providing a constant supply of power to mission-critical areas, including airfield operations, medical facilities, cybersecurity, nuclear installations, and remotely piloted aircraft operations. A disruption of power to any of these areas could be the difference between life or death, making it crucial to ensure operations can continue during a power disruption.

In the event of a disruption that severs the main power supply to an installation, one of the main backup methods is to utilize gas and diesel-powered generators. When this situation occurs, the amount of power becomes limited to the amount of fuel stored, efficiency of generators, and the amount of fuel that can be resupplied. The logistics of transferring fuel to a base requires extensive planning and can become costly over an extended period. For example, the power to Incirlik Air Force Base in Turkey was cut off

in July 2016 [3]. This led to the base supplying power internally in an attempt to sustain daily operations. Due to the Turkish government closing the air space to military aircraft, daily operations against the Islamic State of Iraq and the Levant were shut down [3]. Even with limited daily operations, base officials did not believe operations could be sustained through internal power generation [3]. This situation demonstrates how on-base energy production can be crucial for Air Force installations. Alternatives to fossil fuels are renewable energy sources, such as wind, hydroelectric, or solar power. These sources of power could reduce the need of costly logistic planning and could provide energy for a sustained period if the main power supply to base were severed. When paired with a storage device, such as a large battery, these sources of energy could have the potential to provide power indefinitely.

If the Air Force chooses to install additional sources of renewable energy, the Air Force will be able to better comply with Executive Order 13693. This order requires 25% of electrical energy on government installations to be produced by renewable means in fiscal year 2025 [2]. This order was revoked but may be reinstated in the future. Still the Air Force adopted the same goal into its 2017-2036 Flight Plan requiring 25% of electrical energy use to be generated from renewable energy sources by fiscal year 2025 [1]. With increased demands for renewable energy, the Air Force will be investing a considerable amount of money into renewable energy technologies. Solar energy is one renewable energy technology already commercially available across multiple continents. Solar energy uses photovoltaic cells to capture the energy from the sun's radiation and transforms it into electricity [4].

Problem Statement

With a potential large investment linked to investing in solar infrastructure, there is motivation for the Air Force to identify locations that would provide the best environment to obtain the highest return on their investment. This ensures that tax payer dollars are being effectively used and not wasted on an infeasible project. To obtain this high efficiency, the Air Force must first identify how solar technology can be implemented onto an installation without disrupting daily operations. Once a feasible location on an installation is identified, the next step in the implementation is to determine optimal solar panel orientation. This research only focuses on solar panels utilizing a horizontal orientation, based off recent research proposing that this orientation may be more efficient at capturing diffuse solar radiation [5]. Horizontal orientation will also ensure a standard test system set-up throughout the data collection in this research, allowing the focus to primarily be on how site characteristics affect power production. Finally, the Air Force must identify what climate regions will provide the best environmental conditions to obtain the highest efficiency for horizontal panels.

Research Questions

To identify locations which promote the highest efficiency for horizontal photovoltaic cells, a consistent classification system must be used. One of the most used systems is the Köppen-Geiger climate classification system [6]. This system is based upon many factors that affect the efficiency of solar panels including temperature and precipitation [7]. There are 38 different climates on Earth within this climate classification system. [8]. This system separates the world into easily identifiable areas based upon vegetation, precipitation and temperature [8]. In turn, this allows for an easy prediction of

yearly weather patterns and the environment that a photovoltaic cell would experience. This leads to the following questions:

1. How effective are Köppen-Geiger climate classifications on predicting the power output of horizontal photovoltaic cells?
2. Which Köppen-Geiger climate classification regions and sub-regions are best suited for horizontal photovoltaic cells?

Methodology Overview

A test apparatus was used to collect data to determine which climate is best suited for horizontal photovoltaic cells. This apparatus was assembled and distributed by previous researchers [9], [10] The system consisted of a Raspberry Pi computer, two types of photovoltaic solar panels and an exterior power source. Computer measurements were recorded at 15-minute intervals and the data was recorded on a micro secured digital card. These test systems were distributed to 38 different locations worldwide to collect data for a time frame of at least one year. The locations were selected using a Pareto analysis to organize and determine which locations have the highest concentration of Air Force installations [11]. Data was collected from these systems approximately every two months. The final data set included up to 15 months of power and weather measurements. Finally, a linear model was developed using site temperature and humidity and compared with a model using Köppen-Geiger climate classifications to compare model accuracy at predicting photovoltaic power output.

Assumptions

Several assumptions were made throughout this research. The first assumption is climate zones did not change since the last Köppen-Geiger climate classification had been

updated in 2006 [8]. Climate zones are constantly changing around the world and should be treated as a fluid system [8]. As a result, climate zone boundaries can shift, which could place a test system in a different zone. This assumption was mitigated by placing the test systems as close to the center of the classification zone as possible [11]. Another assumption made was that all test systems worked according to specifications and correctly recorded data. This assumes that variables like snow, dust and or shadowing were limited. These variables were limited by instructing test site point of contacts (POC) to inspect the system and clear off snow or large amounts of dust from the panels.

Limitations

The extent of the data used in this research ranged from May 2017 to September 2018; however, this is not consistent with every location as there were many data gaps. These gaps were due to various reasons, including but not limited to equipment malfunctions, loss of power and point of contact turn over. Collecting data for 15 months helped mitigate variations between the different seasons. However, there is a chance that weather for a region could have been atypical during data collection. The last limitation of this research is that not every climate zone was taken into consideration within the analysis. This was due to the inability to establish contact at certain test locations along with losing contact with several test sites over the data collection period.

II. LITERATURE REVIEW

Due to federal policy change in the last several years, federal installations have been pressured to reduce their global footprint. Renewable energy systems are one of the potential sources many agencies are exploring to reduce their energy demands. The Department of Defense (DoD) and more specifically the United States Air Force have taken these policies and established their own goals to increase its use of renewable energy. One potential source of renewable energy the Air Force could implement on installations is photovoltaic arrays, which capture solar energy. With new technology being developed within the photovoltaic realm, such as solar pavements, the Air Force is researching the best application of photovoltaic arrays to maximize the power output from their investment. Power production from photovoltaic cells highly depends on their orientation and the specific site characteristics such as humidity and temperature of the location where they are utilized. Humidity and temperature fall under a broader classification of climate that in turn help identify areas that have the ideal conditions to utilize photovoltaics. This review will identify current federal and Air Force renewable energy policies, the science behind photovoltaic cells, factors that affect photovoltaic power generation, climate classifications systems and basic regression analysis.

Federal Policy

The Air Force was directed to reevaluate their energy procurement and consumption after the Energy Policy Act of 2005. This act required federal agencies to purchase power from renewable energy sources that were “economically feasible and

technically practical" [12]. It also established a standard for the amount renewable energy to be used and a reduction in energy consumption. These standards stated that 3% of federal electricity use had to be from renewable sources in Fiscal Year 2007 and progressively increased to 7.5% by 2013 along with a 20% decrease in energy consumption by 2015 compared to 2003 [12]. However, many agencies could easily make the argument that renewable energy systems were not economically feasible due to the high cost of these systems compared to the overall lower cost of fossil fuels. Even with the monetary barrier, Air Force personnel were determined to meet these requirements. As a result, one of the first solar array systems in the Air Force was constructed. The created system consisted of a small solar array with a capacity of only 414 kW. It was completed in June 2007 at March Air Reserve Base, California [13].

Figure 1: Photovoltaic array on two carports at March Air Reserve Base [13]

Standards became more restrictive in 2007 with the Independence and National Security Act. This policy called for a 55% fossil fuel energy reduction in newly constructed buildings by 2010 and increases to a net zero (100% reduction) fossil fuel consumption in new buildings by 2030 [14]. With increasing federal requirements, the Air Force decided to further invest in solar array technology, which led to the construction of the largest solar array in North America of its time [15]. The solar array was successfully constructed in 2007 at Nellis Air Force Base outside of Las Vegas, Nevada. The 14 Megawatt (MW) array provides 25% of the base's energy demand resulting in a savings of approximately $1 million dollars per year [15]. A second project expanded the array in 2015 by 15 MW. With this second array, the base can now completely operate on solar energy during daytime operations when there is adequate sun [16]. This highly successful large-scale solar array paved the way for future Air Force and Department of Defense solar array research and implementation.

Figure 2: Second solar array constructed at Nellis, AFB [17]

Current Air Force Policy

The Air Force eventually developed and created its own energy plan in 2010 [18]. Within this guidance the Air Force identified multiple objectives to meet by 2030. These goals include mitigating greenhouse gas emissions, promoting energy security with on-base production, increasing forward operating base renewable energy capability, and providing cost effective technologies to reduce energy demand and increase supply [18]. Four of these goals can be met using solar energy alone. However, these goals have progressed, and new goals have emerged. The most recent energy goals established by the Air Force can be seen in its 2017-2036 Energy Flight Plan [1]. The three main objectives of the energy flight plan are to improve resiliency, optimize demand, and assure supply [1]. Again, solar energy can be utilized to assist in meeting these objectives. Solar energy improves resiliency by supplementing the main energy supply to a base, which can be utilized if the main power supply is severed. The resiliency is further increased if the panels are used in conjunction with a battery in order to continuously supply power during periods without sunlight. With the potential large investment in solar energy, the Air Force needs to understand where to best utilize solar energy and how to implement it on an installation.

Today, several Department of Defense (DoD) installations are utilizing Power Purchase Agreements (PPA) to supplement their energy demand. A PPA is an agreement between a private energy company and the Air Force that allows the company to lease under-utilized land from the Air Force to develop a renewable energy system and sell the energy back to the Air Force and local community for a discounted fixed rate [19]. The

benefits gained from utilizing PPAs include promoting the use of renewable energy in local communities, helping meet federal and DoD energy goals, obtaining a reasonably priced source of energy, and improving energy resiliency on bases. A recent example of a successful implementation of this process can be seen with the Gulf Coast Solar Center in Northwest Florida that was completed in August 2017. This solar array encompasses three different Naval and Air Force installations, with a capacity of 120 MW, the largest in the DoD [20]. The Gulf Coast Solar Center cost the contractor, Coronal Energy, $235 million [21]. By utilizing a PPA, the Air Force and Navy were able to avoid the high front-end cost of the array and the annual maintenance costs and. They were also able to benefit from a lower renewable energy price, compared to fossil fuel sources. Before future extensive funding is invested in solar pavements at installations, research should be conducted to identify areas where horizontal arrays are most efficient due to most solar panel power production models being calculated from angled arrays. First, the science behind photovoltaic cells, along with the different types and factors affecting their power output, will be discussed.

Photovoltaic Cell Technology

The basic concept behind the conversion of sunlight to electricity comes from the photovoltaic effect. To induce this effect, a light absorbing material must be present within the structure of the solar cell. The photovoltaic effect can then take place when energy from sunlight, in the form of a photon, strikes a photovoltaic cell and transfers energy to the negatively charged atomic particles, valence electrons, on the outer edge of the cell [4]. With enough energy transferred, these valence electrons can be freed from their lower

energy level. Finally, with a semi-conducting, potential barrier in place within the structure, these electrons can flow to produce a voltage and as a result drive a current through a circuit [4]. This potential barrier is created by having two different forms of silicon in place, n-type and p-type. N-type silicon is doped with another metal in order to provide a higher number of valence electrons that the photons will easily be able to knock free; p-type silicon is doped with a metal that has fewer valence electrons [22]. The lack of valence electrons in the p-type silicon creates holes and a positive charge that will attract other electrons from the n-type silicon to fill its holes, resulting in a potential barrier and a driven current. This effect can be carried out by using several different semi-conducting materials, including silicon, gallium arsenide (GaAs), copper indium diselenide (CuInSe2), cadmium telluride (CdTe) and indium phosphide [23]. Today the most efficient and commercially available material used for photovoltaic cells are crystalline silicon cells [23]. The two main types of crystalline silicon cells include mono-crystalline cells, which range in efficiency from 16-22%, and poly-crystalline cells, which range in efficiency from 14-18% [24]. Other silicon base panels include thin-film and third-generation photovoltaics; however, these types are not as readily available compared to silicon crystalline solar arrays. Due to availability, the two types of panels used in this research are mono-crystalline and poly-crystalline photovoltaic cells. Silicone panels were used for this research because they accounted for 70 – 90% of the market for photovoltaic panels during acquisition, meaning they would most likely be used in large scale solar array construction [25].

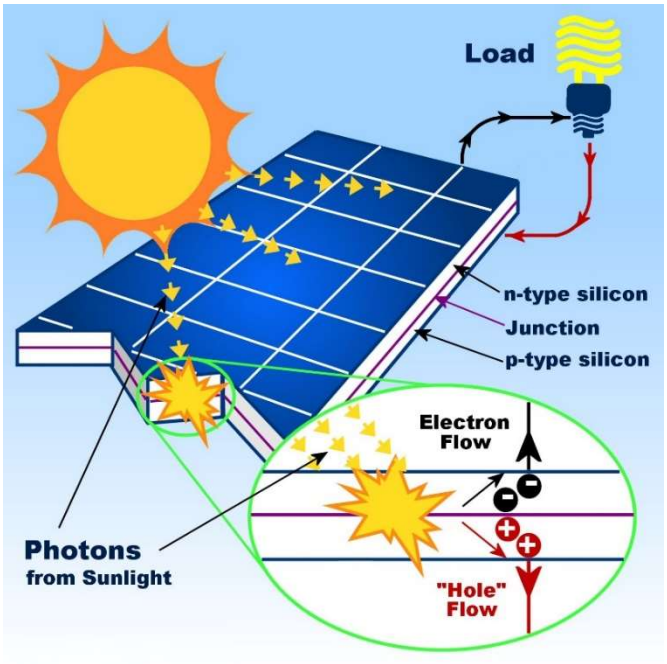


Figure 3: Photovoltaic effect [26]

Figure 4: Mono-crystalline (Left) and poly-crystalline (Right) [9]

Not only is solar cell material important for efficient power production but the orientation at which the cell is positioned can have a significant impact on the power production as well. The most efficient position that allows solar panels to utilize the highest amount of solar irradiance is when its surface is perpendicular to the sun [27]. The

optimum angle of fixed panels can be achieved by setting the tilt angle, the angle between the panel and the ground, equal to the latitude of the panel. This allows the panel to produce the largest amount of power over a year provided that the weather is reasonable. However, tilt angle mainly affects the amount of direct or reflected solar irradiance that the panel is receiving, but, diffuse irradiance is independent of the panel's position [28], [29]. Diffusion of solar irradiance occurs when molecules in the air scatter the sun's rays. These molecules occur due to humidity, cloud coverage, aerosols, etc. One study conducted discovered that daily effective solar irradiance captured increased up to 16% by orientating photovoltaic panels horizontally during overcast and broken cloud situations [5]. The results in Figure 5 show that the highest gains were during times where $0.15 < KT < 0.30$ which correspond with overcast conditions [5]. KT refers to the clearness index of a location, which is a ratio of the solar irradiance received compared to the maximum value that is available on a perfectly clear day. Values range from 0 to 1, with 1 representing a perfectly clear day and 0 representing no solar irradiance being transmitted through the atmosphere. From these results it can be concluded that horizontal photovoltaic arrays could be utilized to better collect diffuse solar irradiance in areas where the weather is consistently cloudy.

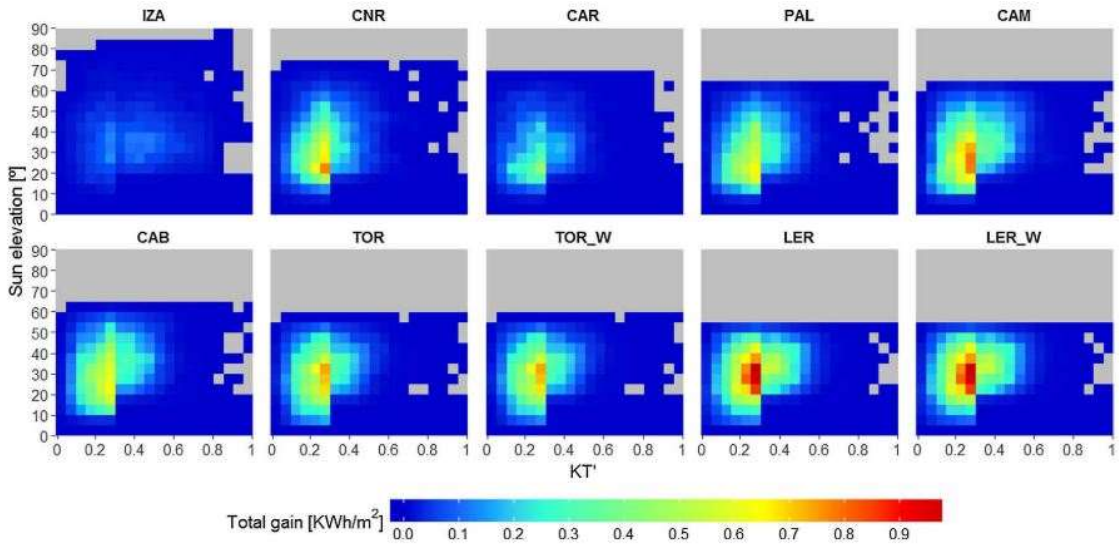


Figure 5: Annual irradiation gains distribution with respect to solar elevation and clearness index [5]

Further research indicates that using latitude as an optimal angle is not nearly as efficient in latitudes above 45 degrees [30]. In another study, Matrix Laboratory (MATLAB) software was used to calculate the optimum tilt angle, and it was discovered that 99.5% of solar irradiance can be collected by adjusting a panels latitude tilt degree six times in a year [31]. The results from the studies above demonstrate that fixed latitude-tilt arrays may not be the best way to utilize solar irradiance. A solution to obtain a higher amount of the sun's solar radiance is to install a tracking system that slowly tilts the array to maintain an optimal angle to follow the sun's rays throughout the day. However, studies have also shown that this may not be efficient in every condition. One of these studies produced results that horizontal photovoltaic cells generated 50% more power than a dual axis solar tracking system during cloudy conditions [23]. This shows that horizontal photovoltaic cells could be used in certain areas where there tends to be a higher percentage of cloud coverage during the day to obtain a higher power production compared to tracking systems. Additionally, axis tracking systems require a large initial

investment and recurring maintenance cost when compared to horizontal mounting systems. Overall, horizontal photovoltaic arrays could be a cost-effective alternative to increase DoD energy resilience.

Site Characteristics

The material and mounting system of photovoltaic cells are not the only factors that play an important role in an array's power production. As mentioned earlier site-specific characteristics can significantly impact the power capabilities of photovoltaic cells. Several site factors that have a large effect on power are temperature, cloud coverage and humidity. If the temperature deviates from the nominal operating cell temperature (NOCT), it can have adverse effects on the efficiency of the panel. The standard conditions for the cell to reach its NOCT include: irradiance of 800 W/m^2, ambient temperature of 20 ºC, wind velocity of 1 m/s, and an open back side mount [32]. An increase in temperature can cause the panel to overheat, resulting in a significant loss of efficiency. For example, every 1 ºC increase in temperature can decrease the efficiency between 0.38% and 0.45% of a mono-crystalline panel [23]. This has led to both operating temperature and ambient temperature being taken into consideration with photovoltaic power and efficiency correlation equations [7].

Other studies have found that cloud coverage also has a significant effect on the power output of photovoltaic cells due to a limitation on the amount of solar irradiance that reaches the panels from clouds [33]. Clouds can scatter the solar irradiance, disrupting its direct path towards a photovoltaic cell restricting the amount of direct irradiance it receives. Cloud coverage can be measured in various ways, but the most known format is

based on the Okta scale from 0-9, where the number represents the number of eighths of the sky that is obscured by clouds [34]. For example, 0 is a perfectly cloudless sky whereas 8 describes a completely overcast day. 9 is used to represent another weather phenomenon that is blocking the sun such as a dense fog. However, this scale does not take into consideration the opaqueness of the clouds which has a severe impact on how much irradiance can pass through [34]. For example, an Okta value of 8 with transparent clouds could potentially allow more irradiance to pass through compared to opaque clouds with an Okta value of 4. Cloud coverage has been previously measured by human observers; however, over the last several years many locations now have integrated a fully automated system [34].

For the reasons listed above, cloud coverage can be an unreliable measurement depending on how the value was measured and may not be available for locations with limited resources. A similar cloud measurement that could be influential in photovoltaic power prediction is cloud ceiling, which measures the height of a cloud cluster at a location [35]. Cloud ceiling has data more readily available from the National Oceanic and Atmospheric Administration (NOAA) and a consistent measurement technique when compared to cloud coverage [36]. For NOAA, cloud ceiling is measured using a ceilometer with values ranging from 0-722 in hundreds of feet. This device only reads opaque clouds with 5/8's or more coverage of the sky, where a reading of 722 indicates a clear sky due to a clouds inability to form higher than 72,200 ft [36]. Cloud ceiling could provide a more direct correlation to photovoltaic power output than cloud coverage due to the impact cloud ceiling has on the distance from the photovoltaic cell solar irradiance is scattered. Overall, clouds play an important role in the amount of irradiance that reaches

the surface and should be taken into consideration when deciding where to place a photovoltaic array.

Similarly, humidity should also be taken into consideration when deciding where to place photovoltaic cells. Humidity is one source that generates diffuse irradiance, which causes direct solar irradiance to scatter and can significantly impact the amount of irradiance that reaches a photovoltaic panel [37]. This scattering of the irradiance happens more specifically from the water molecules in the air refracting, reflecting, or diffracting the sun's rays. Overall, it can be concluded that site specific characteristics, such as temperature, cloud ceiling and humidity can make a significant impact on the efficiency and, as a result, the power production of photovoltaic cells.

Soiling Effects

The efficiency of photovoltaic panels can also be affected by characteristics other than weather, such as dust accumulation and other soiling factors. A soiling factor is something that partially covers or blocks a panel, decreasing the amount of solar irradiance the panel can receive and lowering its overall power capability [38]. Different particles that can cover up an array include leaves, bird droppings, snow, dust and even smog. Soiling factors can play a large role in the power output of a panel and their effects should be taken into consideration before a large array is constructed.

Research has shown that solar panels with a lower inclination collect more dust on their surface which leads to a lower efficiency than panels with a higher tilt angle [39]–[41]. There are benefits and faults with the orientation and tilt angle of a solar panel and how dust accumulates; however, the amount of dust, rainfall, wind, and type of dust is

dependent on the geographic location that the panel is located in [23], [42]. These effects could be taken into consideration with climate classification due to the dust in an area being related to the amount of precipitation an area receives. Using climatic classification as a qualitative variable in a model to predict power output could be beneficial and help account for variance in model predictions due to soiling from dust accumulation.

Soiling also can be attributed to the amount of smog in the air for a local region. These particles affect photovoltaic efficiency by limiting the intensity of solar irradiance on the panel [38]. Dust and other fine particles in the air can produce this same effect, which is classified as soft shading compared to hard shading, which is when larger particles block the irradiance from reaching the panel completely [38]. In conclusion, inefficiencies from soiling should be taken into consideration when deciding the location of a solar array. This can be accomplished by incorporating climate classifications into the model for the power production of a solar array.

Köppen-Geiger Climate Classification

Temperature, cloud coverage, humidity and other weather features are part of a broader classification system that can be used to describe certain areas of the globe. The most widely used and accepted climate classification system is known as the Köppen-Geiger climate classification system [6]. This system incorporates temperature and precipitation to categorize the world into 31 distinct climate zones in Figure 6 below. The Köppen-Geiger climate classification system uses three different criteria to classify its regions, including general vegetation type, precipitation and temperature. First, they are divided using general climate type, which was originally based upon general vegetation

for the area. These types include arid, warm temperate, snow, polar, and equatorial [8]. These regions are then subdivided further using six precipitation groups, including desert, steppe, fully humid, summer dry, winter dry, and monsoonal [8]. Finally, the zones are further delineated using eight temperature classes, including hot arid, cold arid, hot summer, warm summer, cool summer, extremely continental, polar frost, and polar tundra [8]. Using this climate classification system, areas that have ideal photovoltaic cell conditions can be identified. In conclusion, little research has been conducted using weather and power production data collected by placing horizontal photovoltaic panels around the globe.

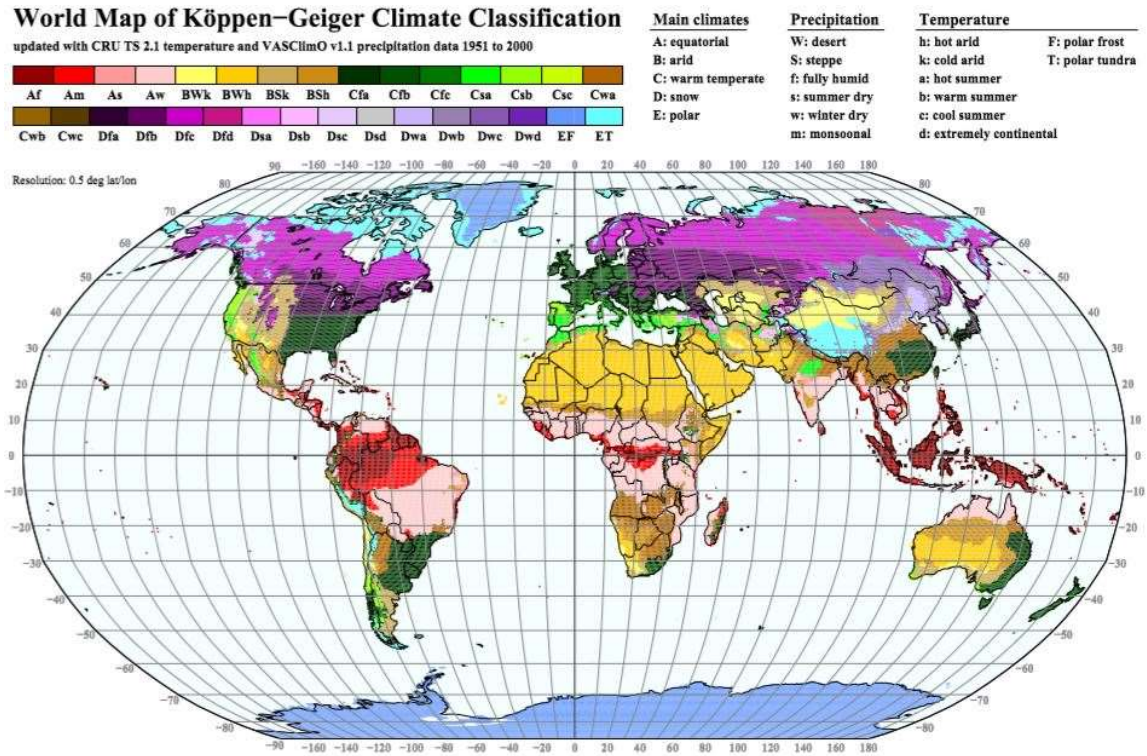


Figure 6: Koppen-Geiger climate classification world map [8]

Regression Analysis

Regression analysis is "a statistical methodology that utilizes the relation between two or more quantitative variables so that a response or outcome variable can be predicted from the other, or others" [43]. This methodology was first developed by Sir Francis Galton towards the end of the 19th century [43]. He discovered this while studying the relation between heights of parents and children, noting that the heights of children regressed towards the mean of the group, regardless of whether they had tall or short parents. From Galton's research, he developed a mathematical relationship to describe this occurrence, a precursor of current regression models. Today regression analysis is used in many businesses and other organizations to help predict certain outcomes utilizing data collected on the subject. Regression analysis has become far more complex by incorporating quantitative, qualitative and quadratic variables and all their interactions. Fortunately, multiple software programs have been created to help perform these analyses to develop linear models to predict outcomes utilizing millions of data points collected. This type of analysis can be applied to help predict solar panel power production using data such as climate zone, latitude, longitude, month, temperature and many other inputs. Overall, this prediction method can help identify ideal site characteristics that will increase the efficiency of photovoltaic production.

Multiple modeling attempts have been conducted to predict the power output of photovoltaic cells, including linear regression. From a recent literature review of modeling equations, none of the 26 models incorporated Köppen-Geiger or an equivalent climate classification system into predicting the power output of photovoltaic panels [44]. These

modeling techniques may be accurate; however, they do not identify an ideal location for photovoltaic technology. Furthermore, these models and equations were not created for horizontal photovoltaic technology. One recent study incorporated Köppen-Geiger climate classifications; however, they generalized the climate classifications to six general climates compared to the 31 distinct climates classifications provided in Figure 6. This research did not incorporate the climate system directly into a photovoltaic performance model, it rather utilized the Köppen-Geiger system with irradiance classification levels to correlate power performance to specific regions [45]. The power performance for the research was also conducted utilizing different prediction models rather than actual power measurements from test equipment. Although this research was able to identify potential high efficiency areas, these regions were based off a fixed-angle panel orientation. This research will address the limitations of current photovoltaic performance modeling and expand current knowledge of horizontal photovoltaic performance modeling.

In conclusion, the analysis of the data acquired from this study would be beneficial to identify locations that are most practical to utilize horizontal photovoltaic arrays or solar pavements. The U.S. Air Force, DoD, and other international businesses could use this research to invest in solar energy to supplement their energy supply to meet stricter policies requiring the use of renewable energy. Particularly, the Air Force and DoD would benefit from this research to increase its goals of energy resiliency and assurance in case an installation's main power supply is severed. In summary, this research aims at creating a model to predict photovoltaic power output and using it to identify the climate zones that promote an environment that maximizes the efficiency of horizontal photovoltaic arrays.

III. METHODOLOGY

The purpose of this chapter is to communicate the procedures and equipment used in this research. The design, assembly and implementation of the test system are discussed along with how the data was collected from these systems. Finally, the data compilation and analysis techniques are discussed to identify how the results were determined.

Test System Design

The test system hardware and software were designed in prior research by Captain John Nussbaum and the Electrical Engineering Department at the Air Force Institute of Technology [11]. The systems themselves were manufactured at TecEdge Works located in Dayton, OH. The main components of the system consist of an ALEKO 25 Watt, 12 Volt mono-crystalline solar panel, a Renogy 50 Watt, 12 Volt poly-crystalline solar panel, a Raspberry Pi 3, model B, version 1.2 computer inside a weatherized case, a weather probe, and an external power source.

A Raspberry Pi is a fully functioning computer all located on a small chip that houses its own central processing unit (CPU), memory and graphics card, as shown in Figure 7 [10], [46], [47]. As one can see, this computer's small size, along with its price and capabilities make it ideal for reading and recording data for this research. The Raspberry Pi computer took measurements at 15-minute intervals. These measurements included the current and voltage of the panels, ambient air temperature, and humidity of the site. The current was read using a noninvasive current Hall sensor. A Hall sensor detects the magnetic field that is produced by the current [48]. This magnetic force is proportionate to the measured current based on the size and type of Hall sensor used [48]. Next, voltage was measured by measuring the voltage drop across a known resistance.

Finally, ambient air temperature and humidity was measured using a probe located on the outside of the weatherized case. Data collection was conducted using a micro secured digital (SD) card equipped with the software designed by the electrical engineers to read and record data.

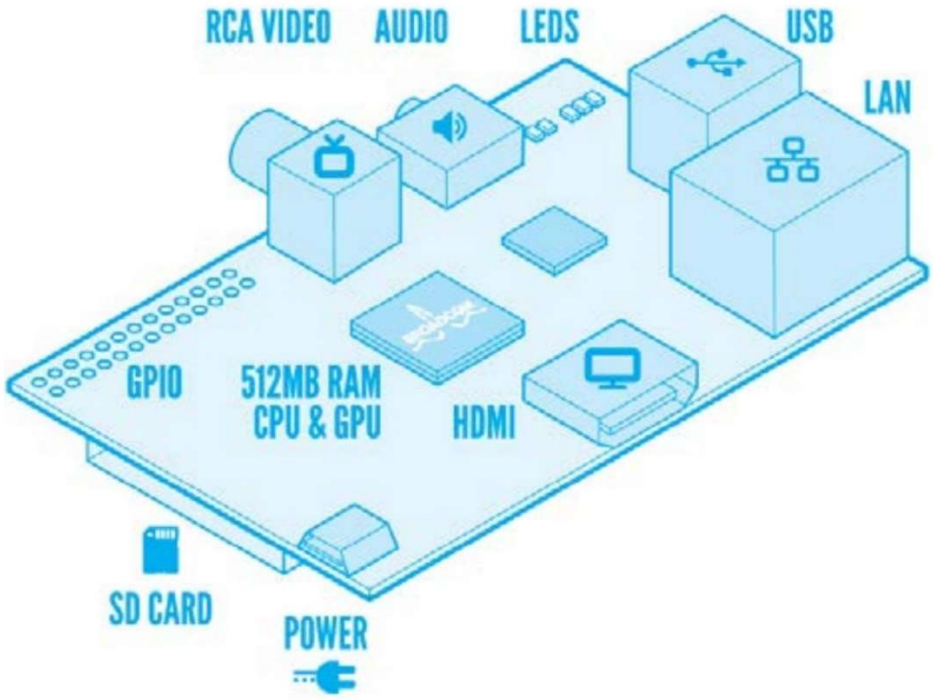


Figure 7: Raspberry Pi components [47]

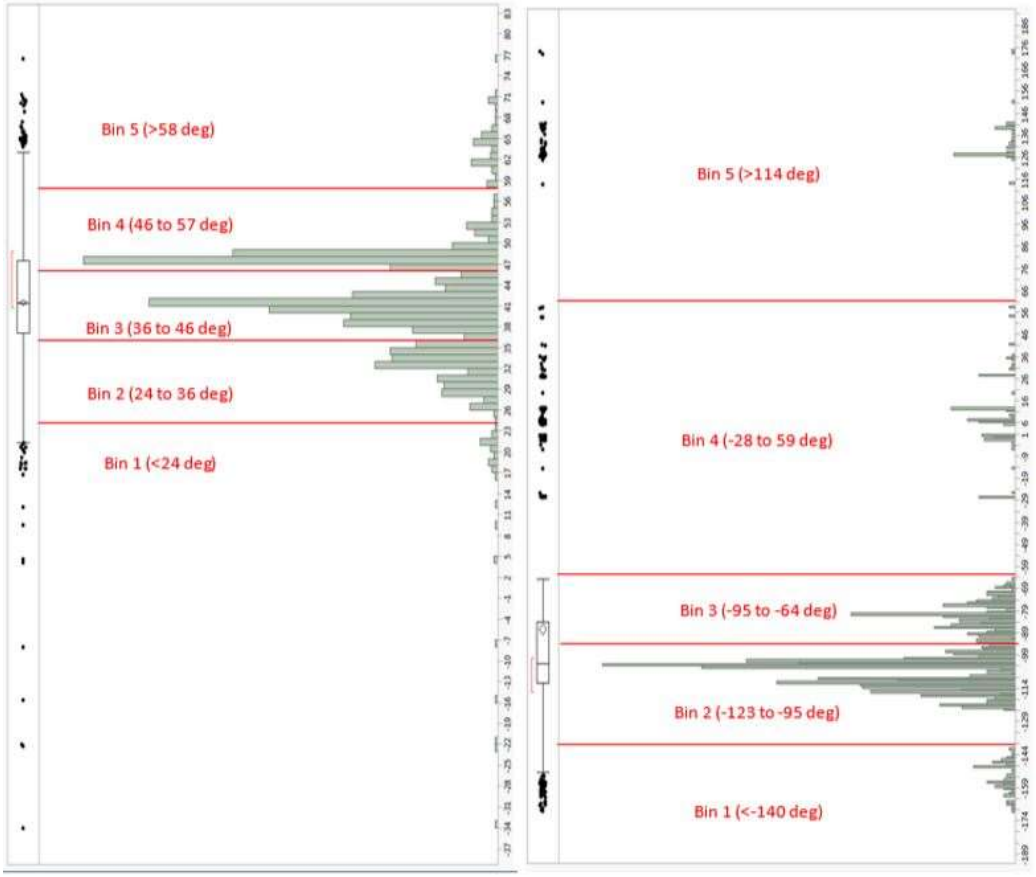
The system was powered by two possible sources depending on the location of the system. The first source of power was from an extension cord that was connected to an outlet of a nearby building. The second source of power, if the system was not located next to a building, consisted of a third photovoltaic panel and battery unit. From this power source the panel can supply power during the day and charge the battery to supply continuous power during periods without sunlight. Due to limited availability during purchasing, the mono-crystalline and poly-crystalline panels used for this research were

not the same size. As a result, the poly-crystalline panel had a rating of 50 watts while the mono-crystalline was only rated at 25 watts. The panel size difference will be addressed later in this chapter.

Finally, three informative light emitting diodes (LEDs) that can be seen outside the case were also equipped on the test system. The lights consist of three colors: red, yellow, and green. When the red LED is lit, it indicates that an error has occurred within the system or during one of the readings. Yellow indicates that the system is operational. Lastly, green indicates that a reading is taking place. Overall, this design allowed for reliable data to be gathered throughout the year.

Site Selection

The locations for placing these systems were selected from current Air Force installations by past researchers using a statistical analysis software called JMP [11]. This analysis began by creating a histogram for latitude and longitude using the 1,763 installations. After the histograms were created an analysis of variance was conducted to create five different bins within each latitude, as seen in Figure 8. From this analysis, 25 distinct regions labeled A through Y were produced. Every region was a combination of the bins developed prior, five for latitude and five for longitude. These 25 regions can be seen in Figure 9 below. With the use of these regions, previous researchers identified which zones to place the test systems in. Using these larger areas limited the number of climates analyzed; however, this process was sufficient due to the goal of analyzing climates that contain a majority of Air Force installations.



(a) Latitudes **(b) Longitudes**

Figure 8: Histogram breakdown of Air Force installation coordinates [10], [11]

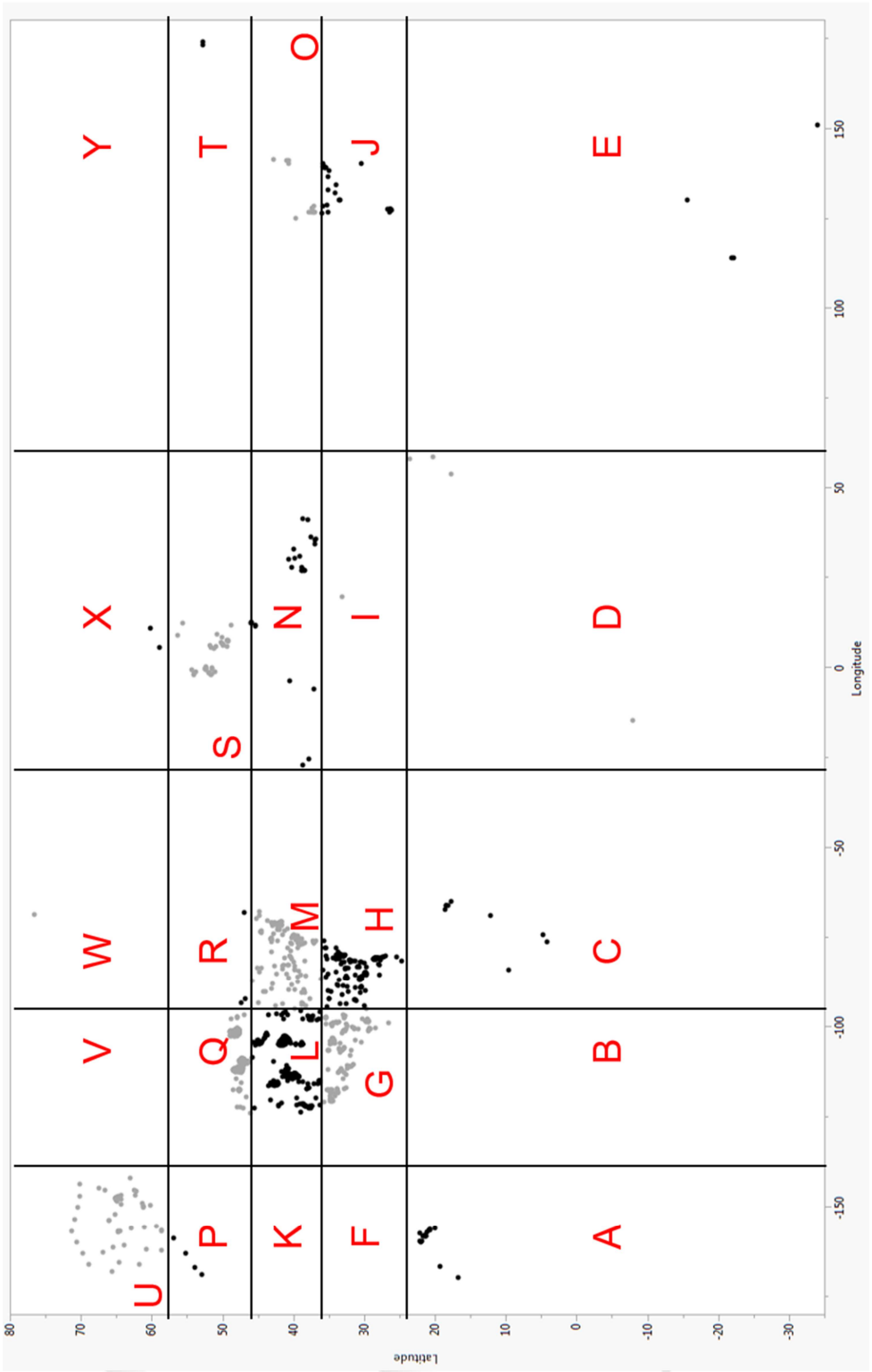


Figure 9: Installation breakdown into 25 regions [10], [11]

Only twenty locations were selected from the regions identified above because five regions did not have an Air Force installation within their boundaries. This limited the ranges of latitude and longitude that the results of this analysis can be generalized to. Seventeen further sites were selected using a Pareto analysis and the Köppen-Geiger climate classification system. A Pareto analysis uses multiple criteria with different weights to develop a solution that optimizes the overall objective of the problem [49]. The objective of the Pareto analysis was to select air force installations from Köppen-Geiger climates that had the highest population of Air Force installations within their boundaries [11]. Adding additional sites allowed for the analysis to maximize the number of climate types incorporated into the research. From the Pareto analysis, it was discovered that the Air Force had installations in 14 of 31 climate classifications. Due to budget constraints limiting the number of systems constructed and shipped, the final climates were selected carefully. With spare parts, one additional test system was set up near Wright-Patterson Air Force Base and maintained by the researchers. The final test location sites are shown as red dots on the map in Figure 10. Some of the locations in Figure 10 are close to one another, making it appear as one dot, such as the U.S. Air Force Academy and Peterson, AFB.

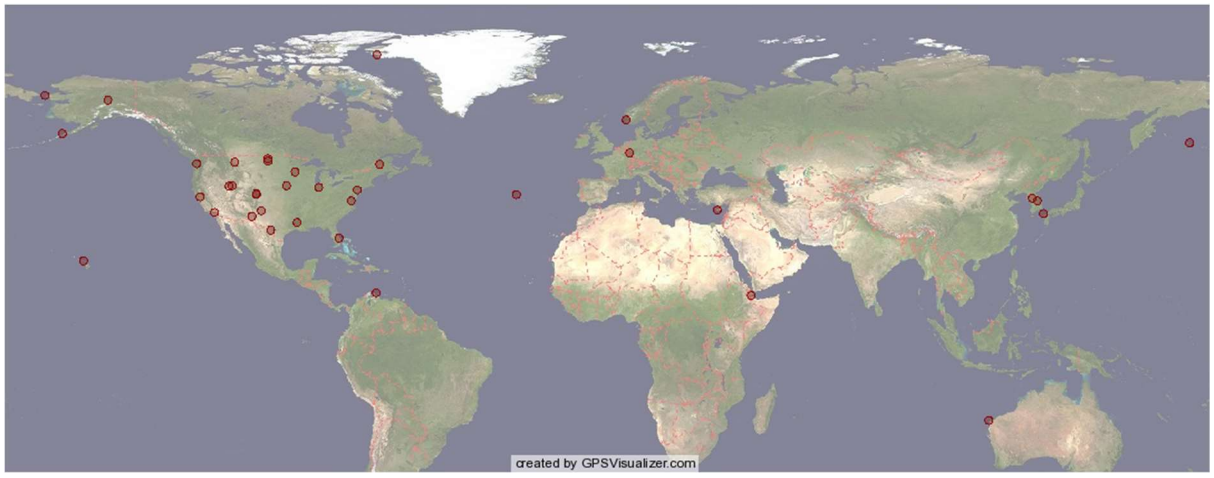


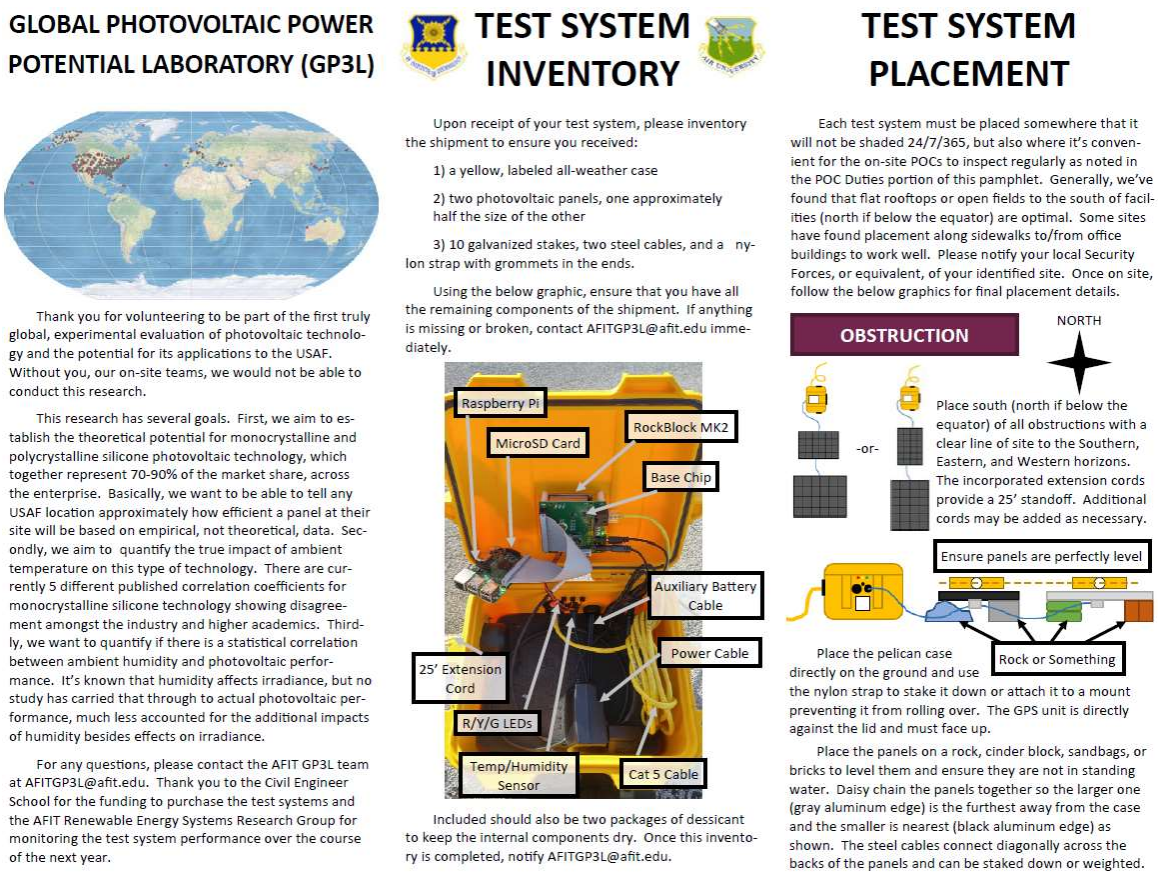
Figure 10: Test site locations [10]

Test System Assembly and Setup

Construction, assembly, and shipping for the systems occurred in late 2016 and early 2017. Most of the assembly occurred using volunteers during a Science, Technology, Engineering and Math (STEM) event for high school students at the Air Force Institute of Technology (AFIT). After the assembly was completed, each system was thoroughly checked and sealed with epoxy or a silicone adhesive.

Upon receiving the test systems, participants were provided with a set of instructions for setup shown in Figure 11 below. These instructions also provided the rationale behind the study to help participants understand the extent of the research and the importance of the data being collected. The instructions recommended the test systems be installed on a rooftop or in an open field; however, participants were able to make the final decision where the panels were placed. This freedom was limited with the requirement that the panels had to have a clear view of the eastern, western, and southern horizons and be placed in a location that ensured direct sunlight the entire day. Locations were also provided with stakes and tie downs to lower the chance of the panels being

blown over. Participants were also directed to inspect the system daily to insure it was still functioning and to clear off any snow, heavy dust, or other debris on the panels. Finally, some participants created their own mounting systems for the panels, which can be seen in Figure 12 and Figure 13. Overall, all locations followed the setup guidance.



GLOBAL PHOTOVOLTAIC POWER POTENTIAL LABORATORY (GP3L)

Thank you for volunteering to be part of the first truly global, experimental evaluation of photovoltaic technology and the potential for its applications to the USAF. Without you, our on-site teams, we would not be able to conduct this research.

This research has several goals. First, we aim to establish the theoretical potential for monocrystalline and polycrystalline silicone photovoltaic technology, which together represent 70-90% of the market share, across the enterprise. Basically, we want to be able to tell any USAF location approximately how efficient a panel at their site will be based on empirical, not theoretical, data. Secondly, we aim to quantify the true impact of ambient temperature on this type of technology. There are currently 5 different published correlation coefficients for monocrystalline silicone technology showing disagreement amongst the industry and higher academics. Thirdly, we want to quantify if there is a statistical correlation between ambient humidity and photovoltaic performance. It's known that humidity affects irradiance, but no study has carried that through to actual photovoltaic performance, much less accounted for the additional impacts of humidity besides effects on irradiance.

For any questions, please contact the AFIT GP3L team at AFITGP3L@afit.edu. Thank you to the Civil Engineer School for the funding to purchase the test systems and the AFIT Renewable Energy Systems Research Group for monitoring the test system performance over the course of the next year.

TEST SYSTEM INVENTORY

Upon receipt of your test system, please inventory the shipment to ensure you received:

1) a yellow, labeled all-weather case

2) two photovoltaic panels, one approximately half the size of the other

3) 10 galvanized stakes, two steel cables, and a nylon strap with grommets in the ends.

Using the below graphic, ensure that you have all the remaining components of the shipment. If anything is missing or broken, contact AFITGP3L@afit.edu immediately.

Raspberry Pi
MicroSD Card
RockBlock MK2
Base Chip
Auxiliary Battery Cable
Power Cable
25' Extension Cord
R/Y/G LEDs
Temp/Humidity Sensor
Cat 5 Cable

Included should also be two packages of dessicant to keep the internal components dry. Once this inventory is completed, notify AFITGP3L@afit.edu.

TEST SYSTEM PLACEMENT

Each test system must be placed somewhere that it will not be shaded 24/7/365, but also where it's convenient for the on-site POCs to inspect regularly as noted in the POC Duties portion of this pamphlet. Generally, we've found that flat rooftops or open fields to the south of facilities (north if below the equator) are optimal. Some sites have found placement along sidewalks to/from office buildings to work well. Please notify your local Security Forces, or equivalent, of your identified site. Once on site, follow the below graphics for final placement details.

OBSTRUCTION

NORTH

-or-

Place south (north if below the equator) of all obstructions with a clear line of site to the Southern, Eastern, and Western horizons. The incorporated extension cords provide a 25' standoff. Additional cords may be added as necessary.

Ensure panels are perfectly level

Rock or Something

Place the pelican case directly on the ground and use the nylon strap to stake it down or attach it to a mount preventing it from rolling over. The GPS unit is directly against the lid and must face up.

Place the panels on a rock, cinder block, sandbags, or bricks to level them and ensure they are not in standing water. Daisy chain the panels together so the larger one (gray aluminum edge) is the furthest away from the case and the smaller is nearest (black aluminum edge) as shown. The steel cables connect diagonally across the backs of the panels and can be staked down or weighted.

Figure 11: GP3L test system instructions [10]

Figure 12: Grid connected test system, Spangdahlem Air Base, Germany

Figure 13: Grid connected test systems, St. Paul Air National Guard Base, MN (Left) and Grissom Air Reserve Base, IN (Right)

Data Collection

Data was planned to be collected monthly from participants at the 38 test locations. Each participant was reminded at the beginning of the month to send the previous month's

data if they had not already submitted it. The micro SD cards were capable of storing more than a year of data, but collection was set to a monthly basis due to problems sending large files over email and allowing researchers ample time to clean and compile the data. Most data files were sent via email, but if a file was too large, participants were able to send it through the Aviation and Missile Research, Development and Engineering Center Safe Access File Exchange site.

Data Compilation

Upon reception of data from participants, it was cleaned, processed, and compiled before analysis. Over each 15-minute interval 64 readings for voltage and current were recorded for each panel respectively. After multiplying these values together, 64 readings of power were available. The maximum value of the 64 readings was obtained and used as the power reading for the panel. Finally, the file was simplified to just 9 columns of data: date, time, humidity, ambient air temperature, internal temperature, temperature of the mono-crystalline and poly-crystalline panel and power of the mono-crystalline and poly-crystalline panels.

Upon completion of data compilation, other data was incorporated into the dataset to provide another variable for analysis. The variable added into the data was cloud ceiling data gathered from national oceanic and atmospheric administration (NOAA). Even though this data was not recorded at the same time interval, each recording was matched to the closest time stamp of the data line received for that specific location using a statistical software called RStudio. Other variables that were incorporated into the dataset were the Köppen-Geiger main and sub climate classifications, latitude, longitude, altitude, and season. Month and hour were also given individual columns to aid in accounting for

further variation when modeling. Next, the time for every location was adjusted from military or "Zulu" time to its respective time zone. Zulu time was the default time setting on each Raspberry Pi computer. Daylight savings was also not accounted for when time was recorded and will not be considered for this research. This requires users to adjust for daylight savings during the application process of the final model. Finally, season was kept the same for all locations except Learmonth solar observatory, located in western Australia, due to its seasons being opposite of the Northern Hemisphere. For all installations in the Northern Hemisphere, fall was identified as September 1 to November 30, winter December 1 to February 28, spring March 1 to May 31, and summer June 1 to August 31. Learmonth's seasons were the opposite with the fall and spring timeframes reversed along with winter and summer dates. Ultimately, every location's respective data was compiled into one single excel document.

Initial Data Examination

After the final dataset was acquired, the data was examined in order to determine what locations could be utilized in creating a linear model. From this initial analysis of the data it became apparent that the monocrystalline solar panel had inconsistent recordings and was not a reliable source of data to draw conclusions from. As a result, it was removed from the dataset and no inferences were made. Next, many of the locations did not have a continuous supply of data and several locations did not even have six months of recorded data. These locations with low amounts of discontinuous data would not be able to create a linear model that accurately predicted photovoltaic performance throughout an entire year. As a result, the locations with low and/or discontinuous months of data were removed.

The next location that was removed was Learmonth Solar Observatory in Northwestern Australia. This location was removed because it was the only site in the southern hemisphere. The problem with having this location in the dataset is that the seasons are opposite of the northern hemisphere. The southern hemisphere's seasons would affect the model and counteract the seasonal effects of the northern hemisphere. As a result, the true effect of season and month on photovoltaic power output would be obscured. After removal, the model's final application and interpretation would be limited to the northern hemisphere.

Next, the data was examined and configured in order to determine what ranges were acceptable for each variable and if there were any outliers that needed to be removed. When first looking at the data, it appeared that some of the Raspberry Pi computers consistently recorded high readings. These readings appeared suspicious, since they were higher than the panels rated wattage of 50 W. Although higher readings were possible, they were highly unlikely and upon further investigation the researchers discovered that the Raspberry Pi's calibration was off for several sites within the first several months. This was discovered by populating histograms of the power output in chronological order. The Raspberry Pi corrected its calibration later, which could have been due to a break-in period of the system or a replacement of the category 5 (CAT) cable. The specifics of the discovery will be addressed further within the results section; however, this problem was addressed by removing the identified dates from the data compilation.

The last discrepancy identified in the data was associated with low temperature readings of the system. The lowest temperature that was recorded was -40°C. This was an extremely low temperature and is a possible recording at some of the sites. However, this

reading appeared to be associated with an error because it was recorded at John Dickins Missile Tracking Station in Florida. Reading this temperature in Florida is impossible and suggests an error in the Raspberry Pi computer when these extremely low temperature values were recorded. When looking closer at this potential error, one potential explanation for northern latitudes is that this error could have occurred at low temperatures causing the Raspberry Pi to malfunction. Based upon these reasons, data with temperature readings lower than -39.3 °C were removed. Along with these low temperature readings, there were several high temperature and power readings that were removed due to these points being extreme outliers and were easily identified as error readings.

Finally, the last data removed was the entire dataset associated with Curacao. Curacao was removed because the NOAA did not have any data available on Curacao when cloud ceiling data was being retrieved. With the removal of Curacao there were also a magnitude of other data points removed that did not have cloud ceiling data available for the specific time stamps of the test system. As a result, the final sites were limited to fourteen locations.

Model Variables

After narrowing the test sites to useful continuous power data and removing outliers, individual variables were addressed to create a simplistic model that allowed for ease in interruption and could be applied to all sites. First, the final variables for the model were selected. These variables included poly-crystalline panel power output, test site latitude, Köppen-Geiger climate classification, altitude, month, hour, temperature, cloud

ceiling, and humidity. Next the rational for the inclusion of each variable into the model will be discussed.

The first variable is what the model will be trying to predict. This is the power output of the poly-crystalline panel. The panel is rated at 50 W, but after removing the discrepancies from the break-in period this value was never obtained at any of the sites. On the lower range of the power output, any values lower than 0.25 W were removed from the dataset. This data was removed in order to account for potential error in the readings and limit the prediction range of the model to instances when the sun was present. Values less than 0.25 W during daylight hours can occur from very dense cloud coverage, snow or other debris accumulating on the panel. These readings were also removed to account for the potential error of the panel recording power from an artificial light source. Potential artificial light sources near the panels could be from street and sidewalk lamps. From Table 1 and 2 below, the highest recommended lux, SI unit of illuminance, for a roadway is 34 lux. The solar cell usable irradiance value from a halogen lamp is 2.03 W/m^2 as seen in Table 3. However, this is from a 500 lux lamp source. Using the irradiance from a halogen lamp with an area of 0.37 m^2, area of photovoltaic cells in the poly-crystalline panel, and a maximum efficiency of 18%, the maximum power the artificial light could generate is 0.14 W [24]. This source combined with other potential calibration errors within the test system itself is why 0.25 W or a 0.5% error was eliminated from the data readings. Overall, power is what the models will be predicting in this research and will be treated as a continuous variable measured in Watts.

Road and Pedestrian Conflict Area		Pavement Classification (Minimum Maintained Average Values)			Uniformity Ratio E_{av}/E_{min}	Veiling Luminance Ratio L_{vma}/L_{av}
Road	Pedestrian Conflict	R1 lux/fc	R2 & R3 lux/fc	R4 lux/fc		
Freeway Class A		6.0/0.6	9.0/0.9	8.0/0.8	3.0	0.3
Freeway Class B		4.0/0.4	6.0/0.6	5.0/0.5	3.0	0.3
Expressway	High	10.0/1.0	14.0/1.4	13.0/1.3	3.0	0.3
	Medium	8.0/0.8	12.0/1.2	10.0/1.0	3.0	0.3
	Low	6.0/0.6	9.0/0.9	8.0/0.8	3.0	0.3
Major	High	12.0/1.2	17.0/1.7	15.0/1.5	3.0	0.3
	Medium	9.0/0.9	13.0/1.3	11.0/1.1	3.0	0.3
	Low	6.0/0.6	9.0/0.9	8.0/0.8	3.0	0.3
Collector	High	8.0/0.8	12.0/1.2	10.0/1.0	4.0	0.4
	Medium	6.0/0.6	9.0/0.9	8.0/0.8	4.0	0.4
	Low	4.0/0.4	6.0/0.6	5.0/0.5	4.0	0.4
Local	High	6.0/0.6	9.0/0.9	8.0/0.8	6.0	0.4
	Medium	5.0/0.5	7.0/0.7	6.0/0.6	6.0	0.4
	Low	3.0/0.3	4.0/0.4	4.0/0.4	6.0	0.4

Table 1: Table of illuminance recommended values adapted from National Optical Astronomy Observatory [50]

Illuminance for Intersections				
Functional Classification	Average Maintained Illumination at Pavement by Pedestrian Area Classification lux/fc			E_{avg}/E_{min}
	High	Medium	Low	
Major/Major	34.0/3.4	26.0/2.6	18.0/1.8	3.0
Major/Collector	29.0/2.9	22.0/2.2	15.0/1.5	3.0
Major/Local	26.0/2.6	20.0/2.0	13.0/1.3	3.0
Collector/Collector	24.0/2.4	18.0/1.8	12.0/1.2	4.0
Collector/Local	21.0/2.1	16.0/1.6	10.0/1.0	4.0
Local/Local	18.0/1.8	14.0/1.4	8.0/0.8	6.0

Table 2: Table of recommended illuminance for intersections adapted from National Optical Astronomy Observatory [50]

Spectra at 500 lux	Visible irradiance (380-780 nm) (W/m^2)	Solar-cell-usable irradiance (300-1200 nm) (W/m^2)	Total irradiance (W/m^2)
Fluorescent tube	1.66 (70%)	1.85 (47%)	1.85
Halogen lamp (with cold reflector)	1.88 (80%)	2.03 (47%)	2.03
AM 1.5	2.36 (100%)	3.95 (100%)	5
Planck's black-body 3000 K (halogen/incandescent lamp)	3.21 (140%)	9.77 (250%)	22.35

Table 3: Table of irradiance values of the spectral distributions at 500 lux adapted from Virtuani (2006) [51]

The next variable incorporated into the model is latitude. Latitude was treated as a continuous variable measured in degrees. Latitude was selected to account for the angle of the sun's irradiance. This angle accounts for the surface area the irradiance is striking. The ideal angle for the irradiance to strike the panel in order to maximize the area exposed is 90°. This ideal angle is why many fixed solar panels are tilted at an angle equivalent to their latitude because irradiance strikes a horizontal panel directly at 90° at 0° latitude or on the equator. However, longitude was not incorporated into this model because longitude does not affect the angle at which irradiance strikes the panel. Longitude affects the location of the solar panel, which can be accounted in other ways such as the climate region or the altitude of the solar panel.

Köppen-Geiger climate classifications were included in the model to identify how effective they predict solar panel power production. Climate classification can account for location-specific characteristics that may not be included in other variables, such as wind speed, precipitation, vegetation and geographical landmarks such as mountains. Other weather variables that climate classifications would incorporate, such as temperature,

cloud ceiling and humidity were added to the model due to their significance in solar power production. Temperature affects how efficient the panel is at generating power while cloud ceiling affects how much irradiance the panel receives. Humidity affects both the efficiency of the panel and the amount of irradiance the panel receives. As mentioned earlier, this is because water vapor in the air affects the amount of diffuse irradiance that reaches the panel and humidity can also have a soiling effect on the panel if water vapor seeps into the glass casing of the panel. The Köppen-Geiger climate classifications will be treated as categorical variables while temperature, cloud ceiling and humidity will be treated as continuous variables. Temperature will be measured in degrees, cloud ceiling in hundreds of feet, and humidity will be expressed as a percentage.

Next, altitude was incorporated into the model to help account for the intensity of the irradiance on the panel. As irradiance travels to Earth, it can be deflected and diffused by water vapor and other particles in the air. With a higher altitude there is less chance for irradiance to be deflected and diffused resulting with a higher amount of direct irradiance hitting the solar panel compared to panels at lower altitudes. Altitude will be measured in meters measured from sea level and treated as a continuous variable.

Finally, time was incorporated into the model to account for the position of the sun throughout the day and its seasonal affects. Time was accounted for by using the variables hour and month. Hour accounted for the position of the sun as it traverses the sky from east to west across the panel. Minute was not included because the position of the sun does not change significantly between the 15-minute measurements compared to its position after 60 minutes. As a result, hour was treated as a categorical variable with values between 0-23. The time frame for this model will be further limited between 10:00AM

and 3:45PM (10-15) or daylight hours. Creating a standard time frame helped eliminate bias in variable coefficients when the sun was not present due to northern locations having a shorter daylight period during the winter solstice. This decision was validated by other research limiting its timeframe to daylight hours as well [25]. Month helped account for seasonal changes throughout the year as well as the sun's elevation in the sky with reference to the southern horizon. The elevation change affected the angle of the irradiance hitting the panel, which as stated earlier affects the amount of area the irradiance is hitting. Month was also treated as a categorical variable with values between 1-12. In conclusion, these variables aided in analyzing the effect of climate classification on horizontal solar panel power output while holding influential variables constant. After developing the model, each climate classification's effect should be easily identifiable and its ability to predict power will be compared to the prediction capability of weather data.

Linear Models

Once the model variables were finalized 1,000 points were randomly removed to provide a validation set to confirm the model's predictive ability. Next, a conceptual, additive model was created that was later replaced with the actual input variables to create the final statistical model used during analysis. After the statistical model was completed, a full and reduced model were developed from the statistical model to conduct the final Wald test. The Wald test was used in place of the analysis of variance (ANOVA) test because not all the assumption for an ANOVA were met during initial data examination. The Wald test compares F-values similarly to ANOVA, however the homoscedasticity and autocorrelation assumptions did not need to be met. After this transition in tests, the full and reduced models were compared using RStudio and conclusions were made on the

effectiveness of the Köppen-Geiger climate classification system to predict photovoltaic power output compared to weather data.

The first conceptual model can be seen in Equation 1. This model simplifies the model into three specific factors that impact photovoltaic power production as expressed earlier in the paper. These factors can be broken into specific variables to better understand their influence on power, as shown in Equation 2. This equation contains 28 different variables and associated coefficients. The variables clumped together represent categorical variables that have multiple dummy variables. Dummy variables are either equivalent to 0 or 1 in an equation such that one variable is represented at a time. There is also one less dummy variable than there are categories for each group due to one variable being the baseline, which is taken into account by the intercept or β_0. Finally, each numbered coefficient is defined in Table 4 below.

$$power = f(sun\ angle, weather, climate)$$

Equation 1: Conceptual photovoltaic power prediction model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{3-13} X_{3-13} + \beta_{14-1}\ X_{14-18} + \beta_{19} X_{19} + \beta_{20} X_{20} + \beta_{21} X_{21} + \beta_{22-}\quad X_{22-27}$$

Equation 2: Statistical photovoltaic power prediction model

Coefficient	Variable	Coefficient	Variable
β_0	Intercept	β_{14}	Hour: 11 (11:00 AM)
β_1	Latitude	β_{15}	Hour: 12 (12:00 PM)
β_2	Altitude	β_{16}	Hour: 13 (1:00 PM)
β_3	Month: 2 (Feb)	β_{17}	Hour: 14 (2:00 PM)
β_4	Month: 3 (Mar)	β_{18}	Hour: 15 (3:00 PM)
β_5	Month: 4 (Apr)	β_{19}	Temperature
β_6	Month: 5 (May)	β_{20}	Humidity
β_7	Month: 6 (Jun)	β_{21}	Cloud Ceiling
β_8	Month: 7 (Jul)	β_{22}	Climate: BSk
β_9	Month: 8 (Aug)	β_{23}	Climate: Cfb
β_{10}	Month: 9 (Sep)	β_{24}	Climate: Csa
β_{11}	Month: 10 (Oct)	β_{25}	Climate: Csb
β_{12}	Month: 11 (Nov)	β_{26}	Climate: Dfa
β_{13}	Month: 12 (Dec)	β_{27}	Climate: Dfb

Table 4: Statistical model coefficient definitions

Next, assumptions were tested to determine if the proposed model could be viable. Assumptions that were tested include multicollinearity, serial correlation, normality, homoscedasticity, and coefficient significance. These were all tested in RStudio to determine if further analysis could be carried out. If these assumptions were not met, appropriate measures needed to be taken to draw valid conclusions from the model. This would most likely be the result of an important variable being withheld from the model, a specific transform or variable interaction not included, or that the data behaves in a non-linear manner.

After all assumptions were tested, a reduced model was created for both weather and climatic variables in order to conduct Wald test. The Wald test compared these reduced models to the full model to determine if the variables provided any value in the prediction of power. The reduced models can be seen below in Equation 3 and 4.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{3-13} X_{3-13} + \beta_{14-1}\ X_{14-18} + \beta_{19} X_{19} + \beta_{20} X_{20} + \beta_{21} X_{21}$$

Equation 3: Reduced prediction model without Köppen-Geiger climate classifications

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{3-13} X_{3-13} + \beta_{14-1}\ X_{14-18} + \beta_{22-27} X_{22-}$$

Equation 4: Reduced prediction model without weather variables

Within the Wald test a hypothesis test, dependent on the model's F-value, was conducted for each reduced model. Below the null (H_0) and alternative (H_a) hypotheses can be seen for each Wald test performed. Depending on the F-value and the associated critical F-value, the null hypothesis can be rejected or fail to be rejected. If the null hypothesis is rejected, the test would be concluded with the alternative hypothesis, meaning that the coefficients are not equal to zero and add value to the power prediction model. If the null hypothesis cannot be rejected, the conclusion would be that the coefficients are equivalent to zero and add no value to the power prediction model. In other words, the variables tested do not have any effect on the power output of horizontal photovoltaic cells.

$$H_0: \beta_{22} = \beta_{23} = \beta_{24} = \beta_{25} = \beta_{26} = \beta_{27} = 0$$

$$H_a: \beta_{22} = \beta_{23} = \beta_{24} = \beta_{25} = \beta_{26} = \beta_{27} \neq 0$$

Equation 5: Köppen-Geiger climate classification variables hypotheses

$$H_0: \beta_{19} = \beta_{20} = \beta_{21} = 0$$

$$H_a: \beta_{19} = \beta_{20} = \beta_{21} \neq 0$$

Equation 6: Weather variables hypotheses

Finally, the two reduced models were compared and analyzed to determine their effectiveness. First, each model's goodness-of-fit was tested by calculating each model's R-squared values after inputting the 1,000 random validation points. The R-squared values

explain how well each model fits the validation set utilizing the estimated coefficient values for each model. Next the models' predictive abilities were tested. Each model's root mean squared error and mean absolute error were calculated to explain how well the models were able to predict the power given the input variables and actual power recorded. In conclusion, the results obtained from these tests determined the effectiveness of incorporating Köppen-Geiger climate classifications.

IV. ANALYSIS AND RESULTS

Initial Data Quality and Analysis

The final data set was secured in October of 2018. The longest timeframe of data collected from a test location was 14 months. The dates of this data range between June 2017 and September 2018. Table 5 below displays the data collected from all locations. A green filled box represents one month of data, an orange filled box represents approximately a half month of data, and an empty or white box represents three days or less data or no data recorded at all. Data was never received from 10 of the sites and 5 more sites dropped off shortly after data collection began. This was believed to be the result of transition and turnover of personnel at these locations. Multiple sites had sporadic data as well, due to malfunctions of the test equipment where no data was recorded for up to a month. These malfunctions were mitigated by sending new CAT5 cables to all locations in hopes to fix the errors and prevent future errors. Another malfunction occurred at Minot and Langley AFB where the computer recorded not only the wrong time but also the wrong month and year. This misreading occurred at random instances making it extremely difficult to correct and adjust the time accurately, resulting in the data being removed. Based on this inconsistent data, the preliminary final location selection was narrowed down to 16 sites to use in the analysis seen in Table 6 below. From these sites 528,569 data points were obtained.

LOCATION	Jun-17	Jul-17	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17	Jan-18	Feb-18	Mar-18	Apr-18	May-18	Jun-18	Jul-18	Aug-18	Sep-18
Camp Murray, Washington, USA																
Camp Red Cloud, South Korea																
Cannon Air Force Base (AFB), New Mexico, USA																
Chabelley Airfield, Djibout																
Claiborne Air Force Range, Louisiana, USA																
Clear Air Force Station, Alaska, USA																
Cold Bay Long Range Radar Site, Alaska, USA																
Eareckson Air Station, Alaska, USA																
former Limestone AFB, Maine, USA																
Forward Operating Location Curacao, Curacao																
Grissom Air Reserve Base, Indiana, USA																
Hill AFB, Utah, USA																
Holloman AFB, New Mexico, USA																
Itazuke Auxiliary Airfield, Japan																
Johnathan Dickinson Missile Tracking Annex, Florida, USA																
Joint Warfare Centre, Norway																
Lajes Field, Portugal																
Langley AFB, Virginia, USA																
Laughlin AFB, Texas, USA																
Learmonth Solar Observatory, Australia																
Malmstron AFB, Montana, USA																
March Air Reserve Base, California, USA																
Maui Air National Guard Communications Station, Hawaii, USA																
McGuire AFB, New Jersey, USA																
Minneapolis-Saint Paul Joint Air Reserve Station, Minnesota, USA																
Minot AFB, North Dakota, USA																
Minot Air Force Missile Site, North Dakota, USA																
Offut AFB, Nebraska, USA																
Osan Air Base, South Korea																
Peterson AFB, Colorado, USA																
Royal Air Force Akrotiri, Cyprus																
Spangdahlem Air Base, Germany																
Thule Air Base, Greenland																
Tin City Long Range Radar Site, Alaska, USA																
Travis Air Force Base, California, USA																
US Air Force Academy, Colorado, USA																
Utah Training and Test Range, Utah, USA																
Wright-Patterson AFB, Ohio, USA																

Table 5: Test site data collection

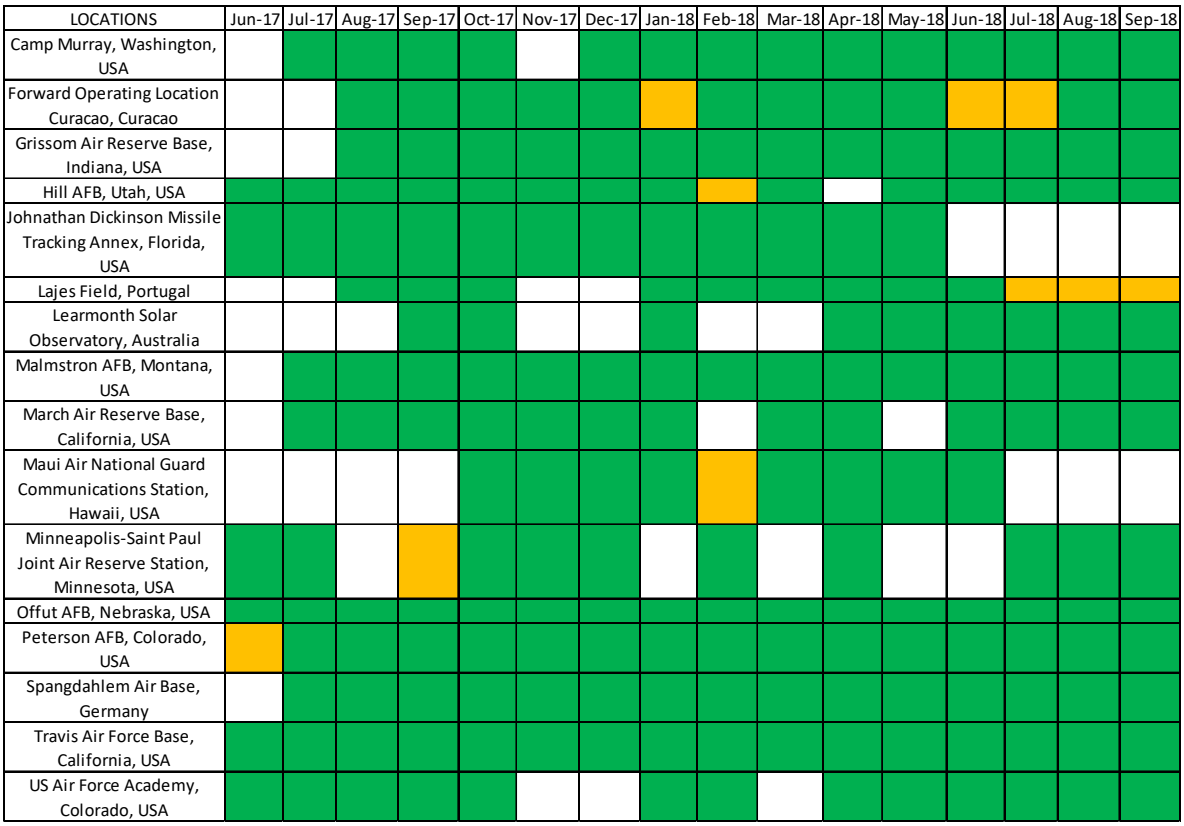


Table 6: Preliminary analysis site selection

After this preliminary data was selected it was observed that the power output recorded for the mono-crystalline panel was unreliable. The values ranged from 0 W to 500 W for a 25 W rated panel. For multiple locations the data consistently read over 25 W. In total there were over 64,000 readings of 25 W or higher, which is impossible for the panel to consistently read at such a high power output. As a result, this data was removed, and no further analysis was performed on the mono-crystalline power output.

Next, as mentioned earlier, Learmonth Solar Observatory was removed due to it being the only test site located in the southern hemisphere. 21,086 points were removed from the dataset when this location was removed. Following the removal of Learmonth, the calibration break-in period of several locations was identified, and discrepancies were

removed. Figure 14 shows that the test system located at Camp Murray in Washington was able to produce over 20W of energy during the night. This is impossible for the panel to produce almost half of its rated wattage without any irradiance from the sun. Displaying the data in a different manner showed that 5 months produced power readings above 50 W while all other months could not produce higher than 35 W, see Figure 15. Upon further investigation a clear discrepancy can be seen. Camp Murray recorded data for the months of June, July and August for both 2017 and 2018. These months were highlighted in the computer program JMP and displayed in Figure 17; the same points can be seen in Figure 16. From Figure 16 and the associated dates of the dataset, the discrepancy can be seen between the two years. In 2017, the minimum power was between 10 W and 25 W with a maximum reading above 70 W, while in 2018 the minimum power was 0 W with a maximum reading no higher than 35 W. Using the associated date and time stamp, the exact date and time could be identified for the discrepancies and removed from the dataset. Below, in Table 7, all locations' discrepancies can be identified with red, while blue represents reliable data and white represents only three or less days of reliable data or no data record at all. In the end 68,338 of data points were removed from the dataset.

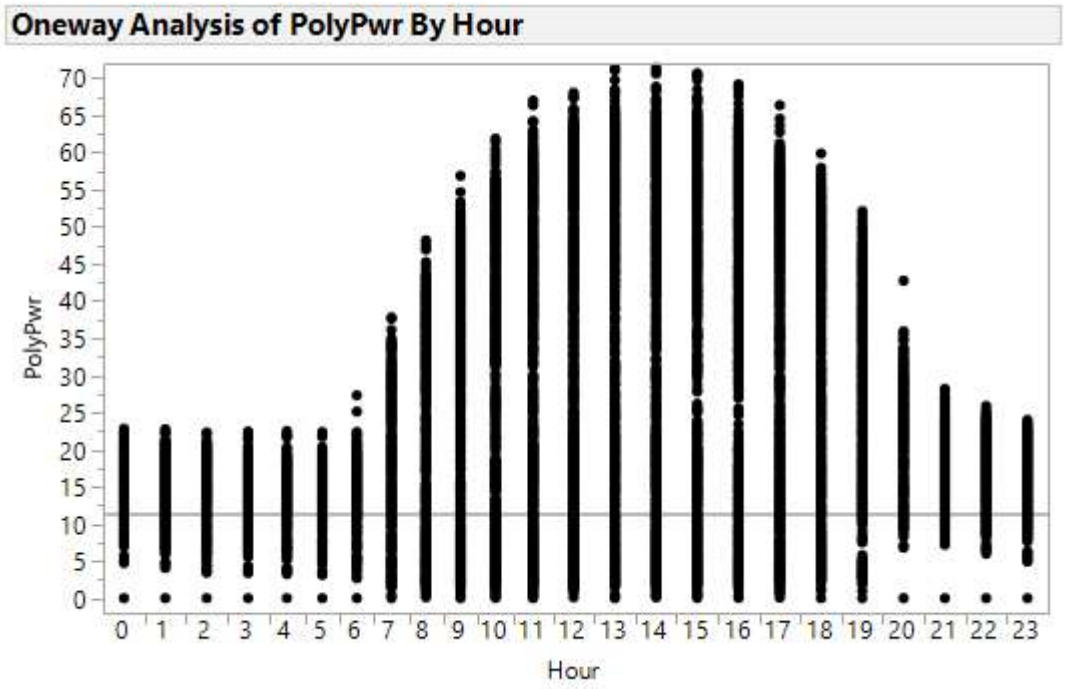


Figure 14: Camp Murray poly-crystalline power output by hour

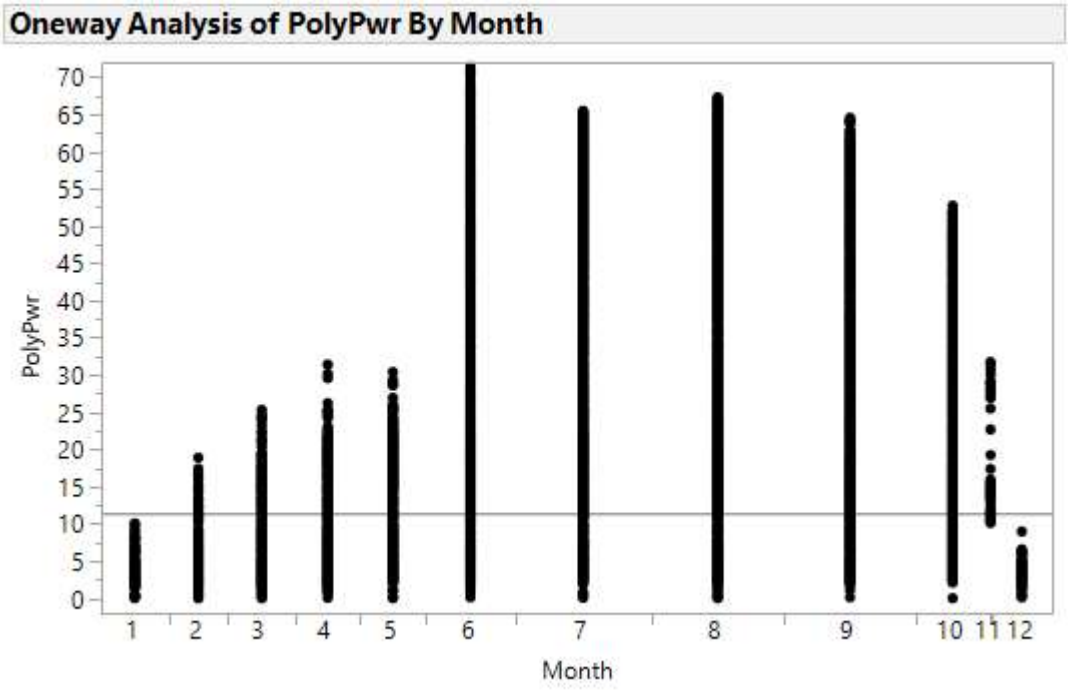


Figure 15: Camp Murray poly-crystalline power output by month

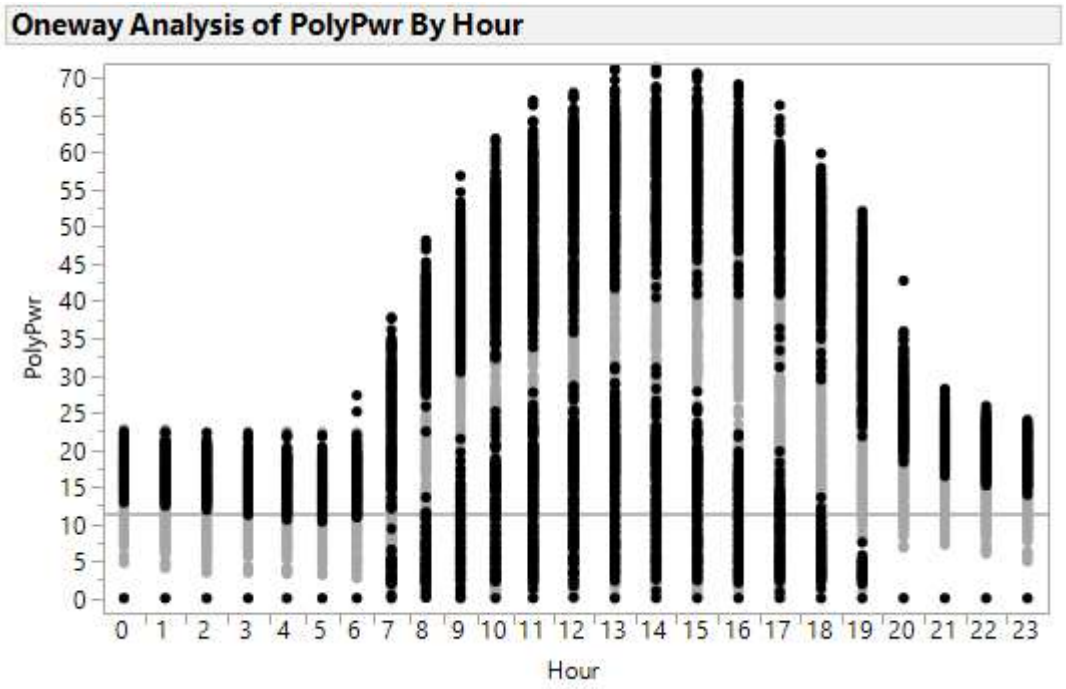


Figure 16: Camp Murray hourly poly-crystalline power output highlighted for the months of June, July and August

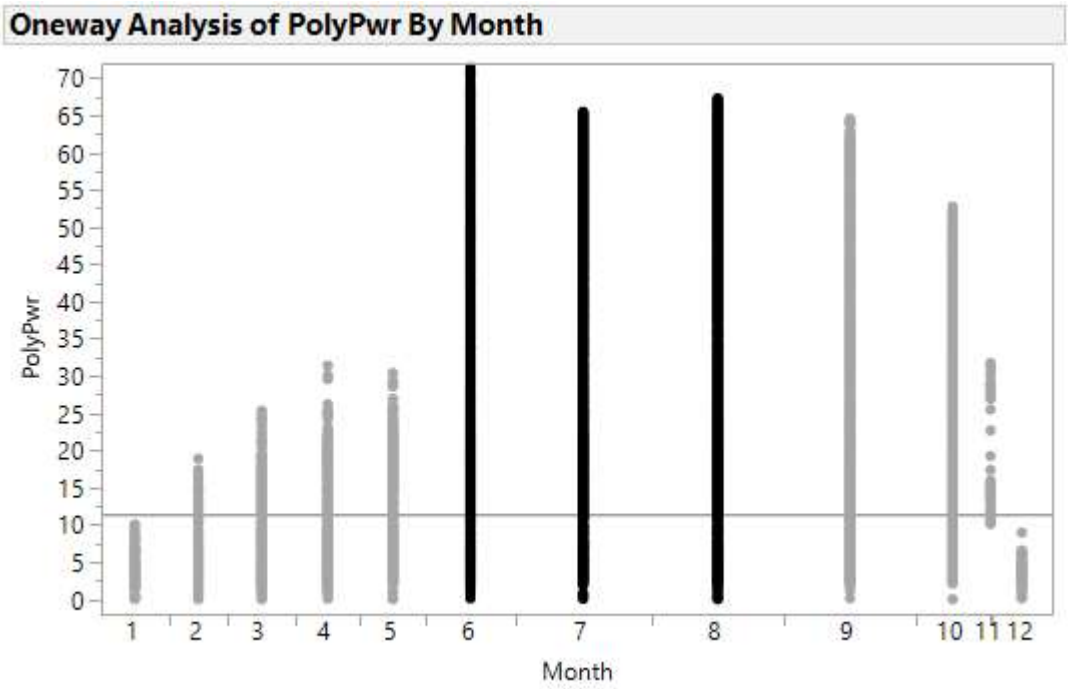


Figure 17: Camp Murray poly-crystalline power output highlighted for the months of June, July and August

LOCATIONS	Jun-17	Jul-17	Aug-17	Sep-17	Oct-17	Nov-17	Dec-17	Jan-18	Feb-18	Mar-18	Apr-18	May-18	Jun-18	Jul-18	Aug-18	Sep-18
Camp Murray, Washington, USA																
Forward Operating Location Curacao, Curacao																
Grissom Air Reserve Base, Indiana, USA																
Hill AFB, Utah, USA																
Johnathan Dickinson Missile Tracking Annex, Florida, USA																
Lajes Field, Portugal																
Malmstron AFB, Montana, USA																
March Air Reserve Base, California, USA																
Maui Air National Guard Communications Station, Hawaii, USA																
Minneapolis-Saint Paul Joint Air Reserve Station, Minnesota, USA																
Offut AFB, Nebraska, USA																
Peterson AFB, Colorado, USA																
Spangdahlem Air Base, Germany																
Travis Air Force Base, California, USA																
US Air Force Academy, Colorado, USA																

Table 7: Reliable location data after break-in period

Next temperatures below -39.3 °C were removed from the data. Along with these low temperatures, 4 other outliers were identified for either having high power readings near 300 W, temperature readings of 124 °C or a combination of high power and high or low temperature measurements. There were also 13 error readings, resulting in a ZD or ZE measurement for temperature and humidity respectively. In all, 3,806 data points were removed from the dataset, resulting in 435,339 points remaining. Following the removal of outliers, Curacao and other recordings without a cloud ceiling measurement were removed from the data. This resulted in only 14 sites being used for the final analysis. Finally, as stated previously, any power measurements below 0.25 W were removed along with any time stamps outside the times of 10:00 AM to 3:45 PM to account for errors and daylight hours respectively. In conclusion, the final dataset used in the model for the

analysis consisted of 24,179 data points. The summary statistics of the final data set can be seen in the tables below.

Months	n
Jan	1445
Feb	1195
Mar	1855
Apr	2305
May	2335
Jun	2680
Jul	3151
Aug	3184
Sep	2592
Oct	966
Nov	1292
Dec	1179

Climate Classifications	n
Af	952
BSk	6842
Cfb	3019
Csa	5131
Csb	2218
Dfa	3574
Dfb	2443

Hours	n
10	3456
11	3867
12	4117
13	4258
14	4228
15	4253

Table 8: Categorical variables summary statistics

Variable	Unit	Mean	St Dev	Min	Max
Latitude	Degree	39.13	6.62	20.89	50.02
Altitude	Meter	730.84	749.85	1.00	1947.00
Ambient Temperature	Degree Celsius	28.54	12.37	-19.98	65.74
Humidity	Percentage	38.90	24.41	0.00	99.99
Cloud Ceiling	Hundreds of Feet	497.72	310.40	0.00	722.00
Power	Watt	12.63	7.12	0.26	34.29

Table 9: Quantitative variables summary statistics

Model Analysis

First, 1,000 data points were randomly selected and removed from the data set to be used for validation after modeling was completed. Next, the full model, see Equation 1, was estimated using the remaining 23,179 data points. Then, multicollinearity was tested amongst the independent variables to determine if any the variables were dependent upon each other. The variation inflation factor (VIF) was determined for each variable. An ideal VIF for a variable would be 1; however, VIFs under 10 are acceptable. After first running the test, two variables within the model had a VIF above 10, see Table 10 below. Climate classification had a VIF of 37.64 while altitude had a VIF of 12.49. First, altitude was removed from the model, which drastically changed the VIF for climate classification. It decreased the VIF from 37.64 to 4.20, indicating that altitude and climate classification were highly correlated. The correlation could be the result of climate classifications and altitudes having similarities holding temperature, humidity and cloud ceiling constant.

Variables	Initial VIF	Final VIF
Altitude	12.49	-
Latitude	2.72	2.67
Month	2.78	2.76
Hour	1.16	1.15
Humidity	3.36	3.33
Ambient Temperature	4.49	4.41
Cloud Ceiling	1.63	1.61
Climate	37.64	4.20

Table 10: Variable inflation factors

Besides this dramatic change none of the other VIF values changed by more than 0.08. After the removal of altitude, the next highest VIF was temperature with a value of 4.41. Climate classification was also removed, and altitude was reinserted into the model to identify its effect. It also lowered all variables' VIFs under 10, with temperature having the highest VIF of 3.69. Due to climate classification being the investigated variable within this research, altitude was ultimately selected to remain out of the model. The updated model is shown in Equation 7 below.

$$Y = \beta_0 + \beta_1 X_1 + \beta_{3-13} X_{3-13} + \beta_{14-18} X_{14-} \quad + \beta_{19} X_{19} + \beta_{20} X_{20} + \beta_{21} X_{21} + \beta_{22-27} X_{22-}$$

Equation 7: Modified statistical photovoltaic power prediction model

After multicollinearity, autocorrelation also known as serial correlation was tested. Autocorrelation occurs when the residuals of the model are dependent upon each other in relation to time. To test this assumption the data was first organized alphabetically by each location and then within each location the data was organized chronologically by date and time. Upon completion of this ordering, each datapoint was chronologically assigned a number, starting at 1 and ending at 23,179. Finally, a plot of the residuals was generated in numerical order by setting the independent variable equal to the number assigned to each datapoint. This plot can be seen in Figure 18 below. From this graph autocorrelation can be clearly identified by the tendency of the data to continually stay above or below the x-axis at y = 0, labeled by the red line. Although the residuals alternate above and below the line, the general tendency of the line can be identified as a slow transition around the red line. This transition can be seen by the high grouping of points, which almost appears to be a solid line. Due to this evident trend, it was concluded that the residuals were

correlated, and no test was conducted. Like non-constant variance, autocorrelation was corrected for by conducting a robust regression to correct the standard error of each variable within the model. The robust regression was again completed within RStudio utilizing the package "sandwich" and command "coeftest," but specifying within the command to correct for both autocorrelation and heteroscedasticity. This correction will not only adjust the standard errors, but also each variables' associated p-value to determine its significance.

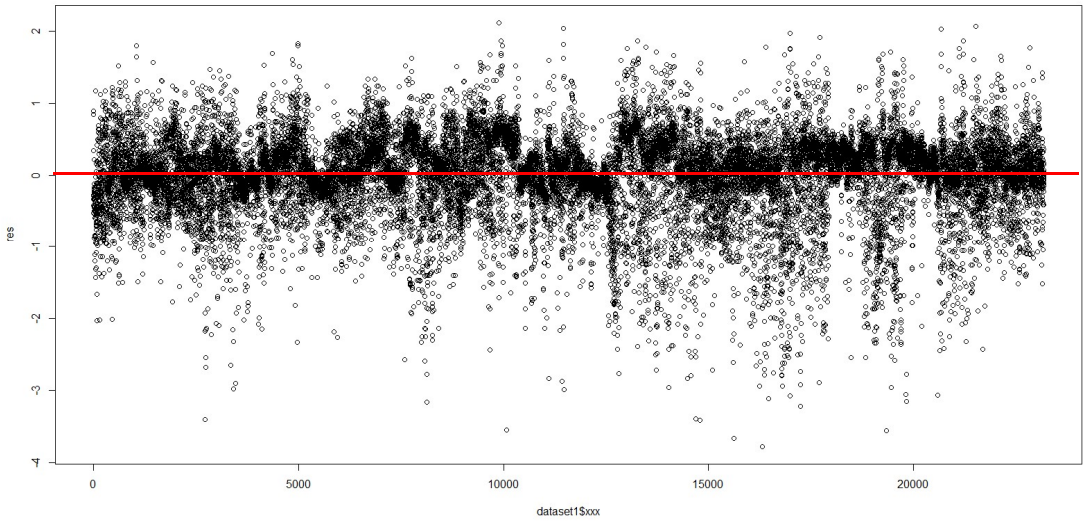


Figure 18: Residuals chronologically graphed

Next, normality was tested for the dataset. This was down by creating a quantile-quantile (Q-Q) plot, see Figure 19. A Q-Q plot is a plot of the sample quantiles, against the theoretical quantiles. In the plot below, the sample quantiles are the standardized residuals along the vertical axis. Ideally, if the dataset was normally distributed, the graphed points would follow the slanted, dotted line across the plot; yet, the tail ends of the plotted points stray from the ideal line, indicating a non-normal distribution. However, the dataset is still considered to be normally distributed due to the Central Limit Theory,

which suggests a large sample with random variables approaches normality regardless of the shape of the population distribution [52].

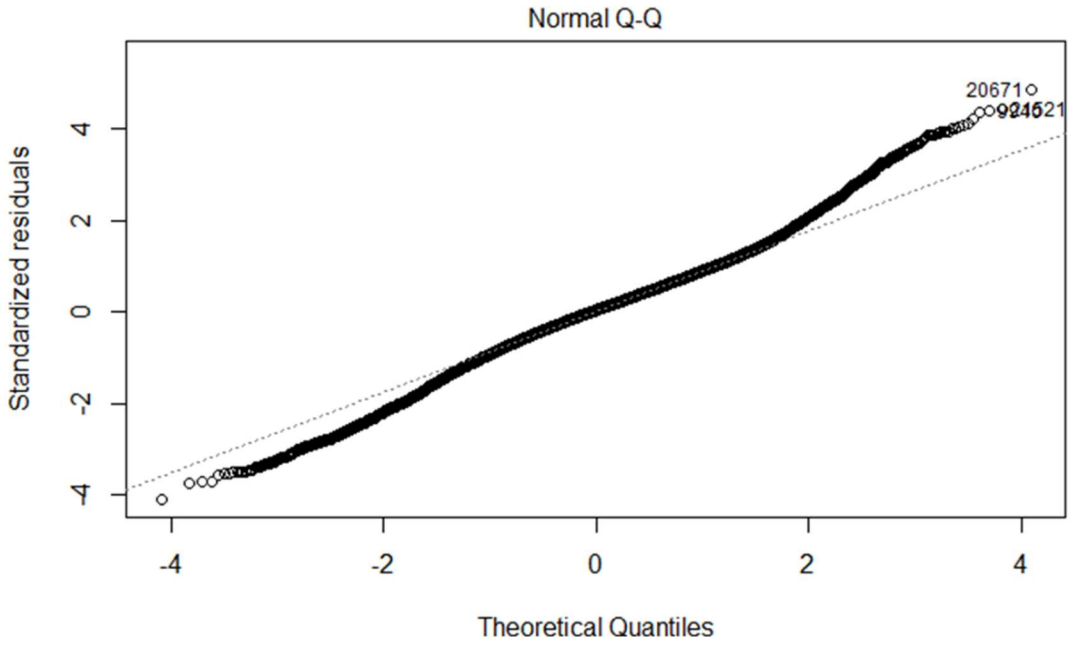


Figure 19: Normality Q-Q plot

Following normality, homoscedasticity, or constant variance of the residuals, was tested by first looking at a plot of the residuals versus the fitted values of the model, seen in Figure 20. In the plot, the values are at first closely grouped to one another, but progressively spread further apart moving left to right across the plot. This plot is depictive of heteroscedasticity, or non-constant variance. For constant variance the values would ideally follow a random pattern with no specific clustering throughout the plot. Heteroscedasticity was confirmed with a Breusch-Pagan test. Like autocorrelation, non-constant variance was dealt with by conducting a robust regression upon the model to adjust the standard error of the variables within the model to determine the correct p-value for each variable and its significance. The robust regression was again completed within RStudio utilizing the package "sandwich" and command "coeftest," but specifying within

the command to correct for both autocorrelation and heteroscedasticity. The results produced from this correction can be seen in Table 11 below.

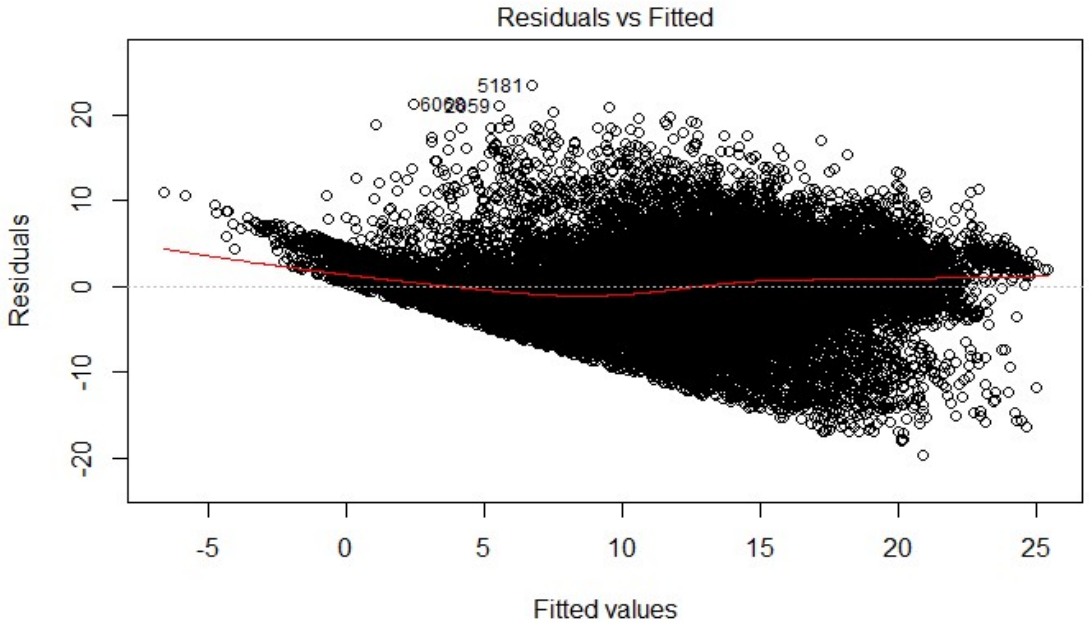


Figure 20: Residuals vs fitted values of modified model

Finally, each variable's significance was determined based upon a t-test using an alpha of 0.05 to determine associated p-value's significance. Any p-value lower than 0.05 would result in an insignificant variable. The original p-values for each variable can be seen along with their respective adjusted values below, in Table 11. After the robust regression, climate classification variables and Month 11 had the largest change in p-value. Of the climate classification variables, the p-value of Csa increased the most, making the variable become more insignificant. This insignificance leads to the conclusion that there is no difference on the effect of power production between Csa and the Af (the base climate of the model). The locations recorded with the climate classification Csa are both located in California, while the location recorded with the climate classification Af is in Hawaii. The insignificance can potentially be explained by

the two climates sharing similarities in local weather patterns and other climatic features while controlling for temperature, humidity and cloud ceiling. Similarly, Dfa and Csb also became insignificant after the robust regression. Again, there could be similarities between these climates and Af such as precipitation or wind speed while controlling for temperature, humidity and cloud ceiling. These variables were not removed from the model in order to maximize the number of DoD installations the model can be applied to. In conclusion, the climates Csa, Dfa, and Csb could not be differentiated against Af in predicting the power output of horizontal photovoltaic panels.

Variable	Coefficient	Coefficient Estimate	Original Standard Error	Corrected Standard Error	Original P-Value	Corrected P-Value
Intercept	β_0	11.96	0.410	0.741	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Latitude	β_1	-0.24	0.008	0.015	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Month (2)	β_3	2.04	0.193	0.343	$< 2.0e^{-16}$	$2.6e^{-9}$
Month (3)	β_4	4.34	0.172	0.312	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Month (4)	β_5	5.26	0.168	0.318	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Month (5)	β_6	6.05	0.176	0.324	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Month (6)	β_7	5.9	0.179	0.330	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Month (7)	β_8	5.12	0.183	0.333	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Month (8)	β_9	4.12	0.180	0.314	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Month (9)	β_{10}	2.48	0.178	0.307	$< 2.0e^{-16}$	$5.8e^{-16}$
Month (10)	β_{11}	0.93	0.210	0.347	$9.1e^{-6}$	0.007
Month (11)	β_{12}	-0.28	0.188	0.325	0.139	0.390
Month (12)	β_{13}	-1.55	0.194	0.321	$1.4e^{-15}$	$1.4e^{-6}$
Hour (11)	β_{14}	2.05	0.115	0.100	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Hour (12)	β_{15}	3.36	0.115	0.110	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Hour (13)	β_{16}	3.5	0.115	0.127	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Hour (14)	β_{17}	2.69	0.117	0.134	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Hour (15)	β_{18}	1.01	0.117	0.136	$< 2.5e^{-16}$	$1.0e^{-13}$
Humidity	β_{19}	-0.06	0.002	0.004	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Ambient Temperature	β_{20}	0.11	0.005	0.009	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Cloud Ceiling	β_{21}	0.01	0.0001	0.0002	$< 2.0e^{-16}$	$< 2.0e^{-16}$
Climate (BSk)	β_{22}	-1.11	0.239	0.512	$3.8e^{-6}$	0.031
Climate (Cfb)	β_{23}	3.63	0.221	0.500	$< 2.0e^{-16}$	$4.2e^{-13}$
Climate (Csa)	β_{24}	0.25	0.215	0.479	0.238	0.596
Climate (Csb)	β_{25}	1.06	0.272	0.564	$9.5e^{-5}$	0.059
Climate (Dfa)	β_{26}	0.85	0.237	0.525	$3.2e^{-4}$	0.104
Climate (Dfb)	β_{27}	2.34	0.253	0.534	$< 2.0e^{-16}$	$1.2e^{-5}$

Table 11: Model coefficient statistics

The only other variable within the model, besides climate classification, that was insignificant was Month 11 (November) with a p-value of .390. This higher p-value could be due to November having a similar effect on horizontal photovoltaic power prediction when compared to the base month of the model (January) while controlling for the other variables in the model. This similarity in effect can be partially explained by the solar elevation of the sun being similar for these two months. Again, the solar elevation is the height of the sun in the sky with respect to the southern horizon. The sun's solar elevation is at its lowest during the winter solstice, which occurs in December. As a result, the solar elevation of the sun is at a similar height leading up to and following the winter solstice. These time frames match up with November and January. However, if this were true, there would be a pair of months that are equal distance away from the winter and summer solstice, Jan/Nov, Feb/Oct, Mar/Sept, etc. As a result, these months should have a similar effect or coefficient estimate on the model. Most of these pairs differ by approximately 1 W as seen in Table 11, which points out that there must be another effect that month is accounting for that the other variables in the model are not. In conclusion, due to goodness-of-fit and predictive abilities of the model being the focus of the model, November was not removed.

In conclusion, the data obtained from the different test locations was developed into a simplified linear horizontal photovoltaic power model. The model was unable to meet the correlation and homoscedasticity assumptions, requiring a robust regression to accurately calculate each variables standard error and p-value. From this model each quantitative coefficient can be interpreted as a wattage increase in horizontal photovoltaic power output for a unit increase of the specific variable. For example, the coefficient for

latitude represents that with every degree increase in latitude, power output will decrease on average by 0.24 W, holding all other variables constant. Categorical coefficients can be interpreted as the average effect that the specific category has on the horizontal photovoltaic power output compared to the base variable. For example, the climate classification coefficient for BSk represents that on average this region will decrease power output by 1.11 W, while holding all other variables constant. The intercept represents the power output of all the base levels for the coefficients. Therefore, 11.96 Watts represents the average power output of a horizontal photovoltaic panel in January, at 10:00 AM on the equator in an equatorial, fully humid climate (Af), with a temperature of 0 °C, humidity of 0%, and cloud ceiling of 0. This interpretation is not fully intuitive because an equatorial climate is highly unlikely to reach a temperature of 0 °C and a cloud ceiling of 0 would imply low cloud coverage at a hundred feet or else, which would limit the irradiance hitting the panel. Overall, the Köppen-Geiger climate classification that was determined to have the highest positive effect on horizontal photovoltaic power production compared to Af is Cfb or warm temperature, fully humid, and warm summer.

Model Comparison

With the completion of an initial model, two reduced models were compared against the final model to determine if climate classification and the weather variables temperature, humidity and cloud ceiling added value to the model. The two models can be seen in Equation 8 and Equation 9 below. This analysis was conducted using a Wald test in order to properly account for the model's robust standard errors. These errors were due to the model having heteroscedastic and serial correlation issues. Both robust Wald tests

were completed with the null hypothesis stating that the removed variables in the reduced models did not add value to the model. In both instances, the test concluded with the rejection of the null hypothesis. With the null hypotheses rejected, the alternative hypotheses are concluded. In conclusion, it was determined that the Köppen-Geiger climates and weather variables, temperate, humidity and cloud ceiling add value to the full model.

$$Y = \beta_0 + \beta_1 X_1 + \beta_{3-13} X_{3-13} + \beta_{14-18} X_{14-1} + \beta_{19} X_{19} + \beta_{20} X_{20} + \beta_{21} X_{21}$$

Equation 8: Adjusted, reduced prediction model without Köppen-Geiger climate classifications

$$Y = \beta_0 + \beta_1 X_1 + \beta_{3-13} X_{3-13} + \beta_{14-1} X_{14-} + \beta_{22-27} X_{22-27}$$

Equation 9: Adjusted, reduced prediction model without weather variables

Next the models' abilities were tested based upon its ability to fit and predict the validation sets. First, the R-squared and adjusted R-squared values of the full and reduced models were compared to determine each model's goodness-of-fit or how well the model fit the validation data. A R-squared value ranges between 0 and 1, with 1 representing a model that accounts for 100% of the variation in the predicted response. The adjusted R-squared value was utilized in this analysis. This is because as the number of coefficients used in a model increases, regardless of their significance, the R-squared value will also increase. From the values in Table 12 below, the model with weather variables and no climate variables can explain, on average, 21.75% more variance of the power output of horizontal photovoltaic cells compared to the model with climate variables and no weather variables from the dataset. Although the weather variables have a better fit, incorporating climate did increase the amount of variation the model can explain by 3.05%. Each model's goodness-of-fit can be seen in Figure 21 below. In the figure, the predicted power

output is graphed against the actual power recorded. The data recorded was for March Air Reserve Base between Nov 28 and Nov 30, 2017. Due to the time frame of the model being between 1000 and 1500, times outside this period all were represented as 0 W.

Model Type	R-Squared	Adjusted R-Squared
Full Model	0.5472	0.5468
Reduced Model without Weather	0.3028	0.3021
Reduced Model without Climate	0.5201	0.5196

Table 12: Model measures of fit

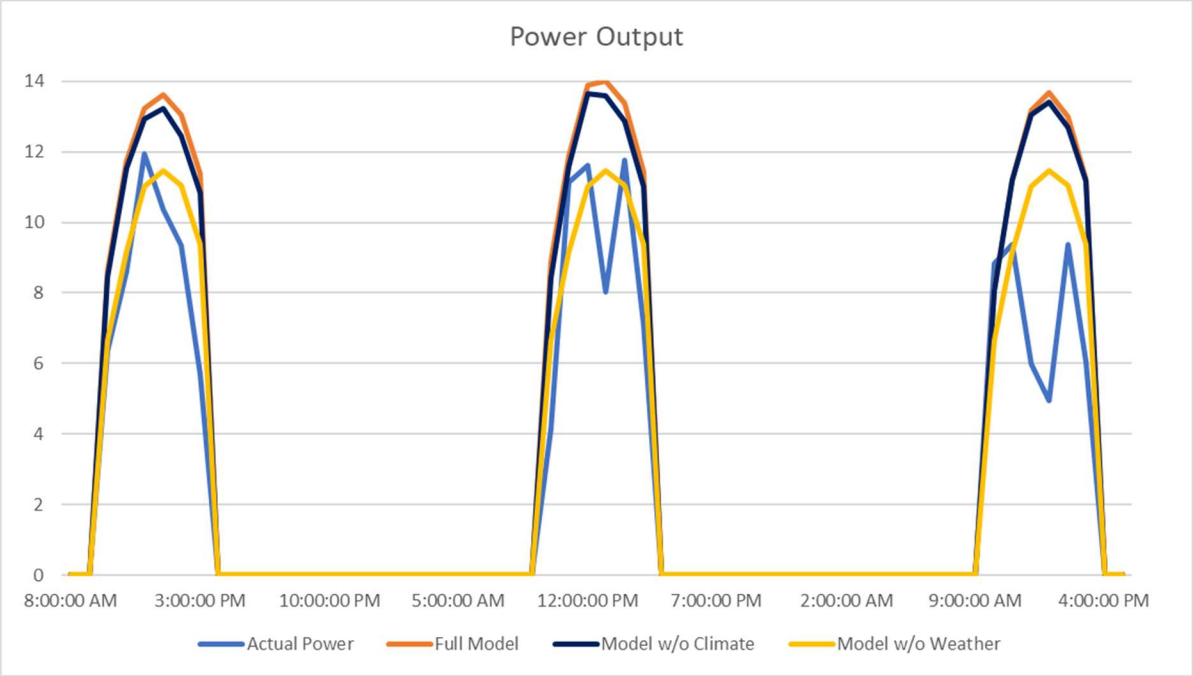


Figure 21: Model predictive power output

The last comparison conducted, was a comparison of the predictive abilities of the full and reduced models. This was completed by inputting the validation dataset into the final models and measuring the difference between the actual values recorded and the predictive values produced from the models. The values calculated to compare the models

were root mean squared error (RMSE) and mean absolute error (MAE). These values were obtained by utilizing RStudio combined with the package "forecast" and command "accuracy." Running the 1,000 validation points through the models produced the values in Table 13 below. The results are similar to the R-Squared values above, showing that the reduced model with weather variables but no climate variables produced less error while predicting horizontal photovoltaic power. However, when climate variables were added, resulting in the full model equation, the error decreased further. In conclusion, weather variables within the model were able to fit the validation data better while producing less error compared to climate variables.

Model Type	RMSE	MAE
Full Model	4.614	3.534
Reduced Model without Weather	5.732	4.656
Reduced Model without Climate	4.749	3.620

Table 13: Model measures of prediction

V. CONCLUSION

In conclusion, horizontal panel power output and weather data was collected from 28 test locations around the globe between June 2017 and September 2018. The model started with over a half billion data points collected from 16 reliable sites. However, the data was narrowed down to 14 test sites resulting in 24,179 usable data points, leading to the development of a linear model to predict power output. The model incorporated site-specific weather and geographical characteristics, along with Köppen-Geiger climate classifications in order to determine the effect of adding climate to the model. After performing a Wald test between the full model and reduced model without Köppen-Geiger climate variables, it was determined that the variables did provide added value to the full model. Although adding Köppen-Geiger variables provided added value to the model, these variables were less effective on fitting and predicting the validation dataset.

After analyzing each models' goodness-of-fit and predictive abilities, it was concluded that the weather variables: cloud ceiling, temperature, and humidity were on average, able to account for more variation with less error compared to Köppen-Geiger climates. However, adding climate to these weather variables further increased the amount of variation explained by 3% and lowered the overall error within the model. The best Köppen-Geiger climate classification was Cfb which was able to produce, on average, 3.6 more watts compared to the climate Af. Similarly, the worst climate was BSk which was able to produce, on average, 1.1 less watts compared to climate Af. Overall, the model can predict the power output of horizontal poly-crystalline photovoltaic panels as 1,213 DoD installations between the hours of 1000-1500. Even though the final model was able to predict the power for these locations, it was only able to account for approximately 55%

of the variation within the data. Before this model is utilized in the implementation of solar infrastructure at DoD installations, its accuracy should be increased.

The model's accuracy could be increased during future research by implementing a pyranometer or utilizing a different fixed effect modeling process. A pyranometer would measure the amount of irradiance the panel is receiving. Irradiance has a direct effect on power output of photovoltaic panels, which would increase the accuracy of the model. Due to monetary constraints, a pyranometer was not included with the test equipment for this research. The model's accuracy could also be improved by utilizing a fixed effect modeling process, see equation 10. This type of model incorporates a fixed variable at the end of the model represented by alpha. Alpha accounts for potential fixed effects on power production for every location (i). Fixed effects could be the result of geographical features at a test location, such as a mountain range, which could limit when the sun rises or sets. By incorporating the fixed variable, it could show the true effect of other variables that were accounting for location specific effects. Overall, utilizing a fixed effect model could improve the accuracy of the predictive model. With improved accuracy, researchers could accurately predict the power output and calculate a benefit to cost ratio of horizontal photovoltaic cells. The results could be compared to different orientations to determine which types of panels would be most beneficial to different climates and installations. In conclusion, with improved model accuracy, DoD installations could make informed decisions on future solar infrastructure investment.

$$Y_{it} = \beta X_{it} + \varepsilon_{it} + \alpha_i$$

Equation 10: Base fixed effect model

References

[1] Department of the Air Force, "United States Air Force Energy Flight Plan," 2017.

[2] B. Obama, "Excutive Order 13693: Planning for Federal Sustainability in the Next Decade," vol. 80, no. 57, 2015.

[3] R. Sisk, "Turkey Cuts Power to Incirlik Air Base, Suspends Military Flights," *Military.com*, 2016. [Online]. Available: www.military.com/daily-news/2016/07/16/turkey-cuts-power-to-incirlik-air-base-suspends-military-flights.html. [Accessed: 30-Jul-2018].

[4] B. Parida, S. Iniyan, and R. Goic, "A review of solar photovoltaic technologies," *Renew. Sustain. Energy Rev.*, vol. 15, no. 3, pp. 1625–1636, 2011.

[5] J. Antonanzas, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas, "Optimal solar tracking strategy to increase irradiance in the plane of array under cloudy conditions: A study across Europe," *Sol. Energy*, vol. 163, pp. 122–130, 2018.

[6] M. Belda, E. Holtanová, T. Halenka, and J. Kalvová, "Climate classification revisited: From Köppen to Trewartha," *Clim. Res.*, vol. 59, no. 1, pp. 1–13, 2014.

[7] E. Skoplaki and J. A. Palyvos, "On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations," *Sol. Energy*, vol. 83, no. 5, pp. 614–624, 2009.

[8] M. Kottek, J. Grieser, C. Beck, B. Rudolf, and F. Rubel, "World map of the Köppen-Geiger climate classification updated," *Meteorol. Zeitschrift*, vol. 15, no. 3, pp. 259–263, 2006.

[9] C. J. Booker, "Analysis of Temperature and Humidity Effects on Horizontal Photovoltaic Panels," Air Force Institute of Technology, 2018.

[10] J. A. Applebee, "Determining the Viability and Efficiency of GP3L Photovoltaic System Study At Air Force Installations in Various Climate Regions Air Force Institute of Technology," Air Force Institute of Technology, 2018.

[11] J. H. Nussbaum, "Analyzing the Viability of Photovoltaic Pavement Systems: A Study In Structural Testing Methods, Measuring Potential Power, and Quantifying the Risks of Implementation," Air Force Institute of Technology, 2017.

[12] M. Holt and C. Glover, “CRS Report for Congress Energy Policy Act of 2005 : Summary and Analysis,” *Energy Policy*, 2006.

[13] A. Abbott, “Solar energy helps power base,” *Air Force Reserve Command*, 2007. [Online]. Available: http://www.afrc.af.mil/News/Article-Display/Article/158743/solar-energy-helps-power-base/. [Accessed: 15-Aug-2018].

[14] Congressional Budget Office, *Energy Independence and Security Act of 2007: A Summary of Major Provisions, Congressional Research Service Report to Congress*. 2007.

[15] Department of Energy, “Nellis Air Force Base solar array provides model for renewable projects,” *Energy.gov*, 2010. [Online]. Available: https://www.energy.gov/articles/nellis-air-force-base-solar-array-provides-model-renewable-projects. [Accessed: 15-Aug-2018].

[16] NVEnergy, “Nellis Solar Array II Generating Station,” 2017. [Online]. Available: https://www.nvenergy.com/publish/content/dam/nvenergy/brochures_arch/about-nvenergy/our-company/power-supply/Nellis-Fact-Sheet.pdf.

[17] K. Elliot, “Nellis breaks ground on DoD’s largest solar array,” *U.S. Air Force*, 2015. [Online]. Available: https://www.af.mil/News/Article-Display/Article/583512/nellis-breaks-ground-on-dods-largest-solar-array/. [Accessed: 22-Oct-2018].

[18] U.S. Air Force, “Air Force Energy Plan: 2010,” 2010.

[19] B. Millberg, “Solar Power Purchase Agreements,” in *GreenPower Partnership Seminar*, 2009, pp. 0–2.

[20] R. DelaHaya, “Gulf Power celebrates completion of largest DOD solar installations with Flip the Switch ceremony,” *Gulf Power News*, 2017. [Online]. Available: https://www.gulfpowernews.com/gulf-power-celebrates-completion-largest-dod-solar-installations-flip-switch-ceremony/. [Accessed: 15-Aug-2018].

[21] J. Bebon, “Coronal Energy Closes $235M Financing For Florida Portfolio,” *Zackin Publications Inc*, 2018. [Online]. Available: https://solarindustrymag.com/coronal-energy-closes-235m-financing-florida-portfolio/. [Accessed: 18-Nov-2018].

[22] T. Sugiura, “Barrier Potential across Semiconductor P-N Junction and Resting Membrane Potential,” *J. Arrhythmia*, vol. 27, no. 4, pp. 353–355, 2011.

[23] M. M. Fouad, L. A. Shihata, and E. S. I. Morgan, “An integrated review of factors influencing the performance of photovoltaic panels,” *Renew. Sustain. Energy Rev.*, vol. 80, pp. 1499–1511, 2017.

[24] M. A. Green, K. Emery, Y. Hishikawa, W. Warta, and E. D. Dunlop, “Solar cell efficiency tables (version 43),” *Prog. Photovoltaics Res. Appllications*, vol. 22, pp. 1–9, 2015.

[25] A. Nagengast, C. Hendrickson, and H. Scott Matthews, “Variations in photovoltaic performance due to climate and low-slope roof choice,” *Energy Build.*, vol. 64, pp. 493–502, 2013.

[26] “Solar power overview,” *Ruaha Energy*, 2016. [Online]. Available: http://ruahaenergy.com/solar/solar-overview/. [Accessed: 22-Oct-2018].

[27] O. Zogou, “Experimental and computational investigation of the thermal and electrical performance of a new building integrated photovoltaic concept,” 2011.

[28] J. D. Mondol, Y. G. Yohanis, and B. Norton, “The impact of array inclination and orientation on the performance of a grid-connected photovoltaic system,” *Renew. Energy*, vol. 32, no. 1, pp. 118–140, 2007.

[29] D. King, W. Boyson, and J. Kratochvil, “Analysis of factors influencing the annual energy production of photovoltaic systems,” in *Photovoltaic Specialists, IEEE Conference*, 2002, pp. 1356–1361.

[30] J. Viitanen, “Energy efficient lighting systems in buildings with integrated photovoltaics,” Aalto University, 2015.

[31] T. O. Kaddoura, M. A. M. Ramli, and Y. A. Al-Turki, “On the estimation of the optimum tilt angle of PV panel in Saudi Arabia,” *Renewable and Sustainable Energy Reviews*, vol. 65. pp. 626–634, 2016.

[32] E. Cuddihy, C. Coulbert, A. Gupta, and R. Liang, “Flat Plate Solar Array Project Final Report Volume VII : Module Encapsulation,” Jet Propulsion Laboratory, 1986.

[33] F. H. Gandoman, S. H. E. Abdel, N. Omar, A. Ahmadi, and F. Q. Alenezi, “Short-term solar power forecasting considering cloud coverage and ambient temperature variation effects,” *Renew. Energy*, vol. 123, pp. 793–805, 2018.

[34] C. J. Smith, J. M. Bright, and R. Crook, “Cloud cover effect of clear-sky index distributions and differences between human and automatic cloud observations,” *Sol. Energy*, vol. 144, pp. 10–21, 2017.

[35] P. Kuhn *et al.*, “Benchmarking three low-cost , low-maintenance cloud height measurement systems and ECMWF cloud heights against a ceilometer,” *Sol. Energy*, vol. 168, pp. 140–152, 2018.

[36] NOAA, “National Oceanic and Atmospheric Administration.” [Online]. Available: www.noaa.gov. [Accessed: 14-Jan-2018].

[37] A. J. Gutiérrez-trashorras, E. Villicaña-ortiz, E. Álvarez-álvarez, J. M. González-caballín, J. Xiberta-bernat, and M. J. Suarez-lópez, “Attenuation processes of solar radiation. Application to the quantification of direct and diffuse solar irradiances on horizontal surfaces in Mexico by means of an overall atmospheric transmittance,” *Renew. Sustain. Energy Rev.*, vol. 81, pp. 93–106, 2018.

[38] M. R. Maghami, H. Hizam, C. Gomes, M. A. Radzi, M. I. Rezadad, and S. Hajighorbani, “Power loss due to soiling on solar panel: A review,” *Renew. Sustain. Energy Rev.*, vol. 59, pp. 1307–1316, 2016.

[39] H. Lu and W. Zhao, “Effects of particle sizes and tilt angles on dust deposition characteristics of a ground-mounted solar photovoltaic system,” *Appl. Energy*, vol. 220, pp. 514–526, 2018.

[40] A. A. Babatunde, S. Abbasoglu, and M. Senol, “Analysis of the impact of dust, tilt angle and orientation on performance of PV Plants,” *Renew. Sustain. Energy Rev.*, vol. 90, pp. 1017–1026, 2018.

[41] E. Abdeen, M. Orabi, and E. S. Hasaneen, “Optimum tilt angle for photovoltaic system in desert environment,” *Sol. Energy*, vol. 155, pp. 267–280, 2017.

[42] J. J. John, S. Warade, G. Tamizhmani, and A. Kottantharayil, “Study of soiling loss on photovoltaic modules with artificially deposited dust of different gravimetric densities and compositions collected from different locations in India,” *IEEE J. Photovoltaics*, vol. 6, no. 1, pp. 236–243, 2016.

[43] M. Kutner, C. Hachtsheim, J. Neter, and W. Li, *Applied Linear Statistical Models*, 5th ed. McGraw-Hill Irwin, 2005.

[44] U. Kumar, K. Soon, M. Seyedmahmoudian, and S. Mekhilef, “Forecasting of photovoltaic power generation and model optimization : A review,” *Renew. Sustain. Energy Rev.*, vol. 81, no. June 2017, pp. 912–928, 2018.

[45] J. A. Vásquez, K. Brecl, and M. Topič, “Köppen -Geiger-Photovoltaic Climate Classification,” *7th World Conf. Photovolt. Energy Convers. Kona, Hawaii*, 2018.

[46] F. P. Tso, D. R. White, S. Jouet, J. Singer, and D. P. Pezaros, “The Glasgow raspberry Pi cloud: A scale model for cloud computing infrastructures,” *Proc. - Int. Conf. Distrib. Comput. Syst.*, pp. 108–112, 2013.

[47] V. Vujovic, V. Vujović, and M. Maksimović, “Raspberry Pi as a Wireless Sensor node : Performances and constraints Raspberry Pi as a Wireless Sensor Node : Performances and Constraints,” *2014 37th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2014 - Proc.*, pp. 26–30, 2014.

[48] Y. Sharon, B. Khachatryan, and D. Cheskis, “Low current Hall Effect Sensor,” 2018.

[49] Y. Ji, J. Li, and C. Dia, “An Improvement Evolutionary Algorithm Based on Grid-Based Pareto Dominance for Many-Objective Optimization,” *2018 14th Int. Conf. Comput. Intell. Secur.*, pp. 15–19, 2018.

[50] National Optical Astronomy Observatory, “Recommended Light Levels (Illuminance) for Outdoor and Indoor Venues.” [Online]. Available: https://www.noao.edu/education/QLTkit/ACTIVITY_Documents/Safety/LightLevels_outdoor+indoor.pdf.

[51] A. Ã. Virtuani, E. Lotter, and M. Powalla, “Influence of the light source on the low-irradiance performance of Cu (In , Ga) Se 2 solar cells,” vol. 90, pp. 2141–2149, 2006.

[52] J. McClave, P. Benson, and T. Sincich, *Statistics for Business and Economics*, 12th ed. Pearson, 2014.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
21-03-2019	Master's Thesis	Sept 2017 - March 2019

TITLE AND SUBTITLE
Analyzing the Efficiency of Horizontal Photovoltaic Cells at Air Force Installations in Various Climate Regions

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Hines, Parker A., Second Lieutenant, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/ENV)
2950 Hobson Way, Building 640
WPAFB OH 45433-8865

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENV-MS-19-M-179

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Intentionally left blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED..

13. SUPPLEMENTARY NOTES
This work is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT
The U.S. Air Force relies heavily on electrically powered facilities to ensure mission-critical capabilities can be carried out. Without a resilient electrical system in place, the U.S. Air Force could be without power for extended periods, resulting in severe implications. Installing a photovoltaic array on an installation is one strategy to improve energy resiliency. The amount of power a photovoltaic array can produce is reliant on its geographical location, position and weather characteristics. This research presents the development of linear regression models based upon 14 case studies from global Department of Defense (DoD) installations to predict horizontal photovoltaic power output. The model incorporates Köppen-Geiger climate classifications with location-specific weather and geographical variables to predict horizontal photovoltaic power production. Both Köppen-Geiger climate and weather variables were determined to provide added value to the model. From the analysis, it was determined that weather variables were able to account for 22% more variation within the validation set compared to climate variables.

15. SUBJECT TERMS
Köppen-Geiger, horizontal photovoltaics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Lt Col Torrey Wagner, AFIT/ENV
U	U	U	UU	82	19b. TELEPHONE NUMBER *(Include area code)* (937) 255-3636, x4611 torrey.wagner@afit.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18