SURFACE REPRESENTATIONS OF RAINFALL AT SMALL EXTENTS: A STUDY OF RAINFALL MAPPING BASED ON VOLUNTEERED GEOGRAPHIC INFORMATION IN KONA, HAWAII

by

Jeffrey Fletcher Schroeder

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

December 2013

Acknowledgements

I would like to thank my Committee Chair, Dr. Karen Kemp, whose guidance, feedback, and patience have been invaluable to me. In addition, I would also like to thank my committee members, Dr. Darren Ruddell and Dr. Katsuhiko Oda for their contributions.

I would like to thank my wife Zoey, and my children Eva and Simon. Their patience, support, and belief in my journey have been unwavering.

Table of Contents

List of Tables

List of Figures

Abstract

Rainfall maps produced with data from widely dispersed official government weather stations are generalized maps covering broad geographical areas that provide little detail at larger scales. Little research has been completed in producing surfaces at smaller extents due to the lack of available data. A non-traditional method of obtaining additional data is through Volunteered geographic information (VGI), which presents data from non-authoritative sources that often supplement traditional data sources, and make analyses not previously considered, now possible.

This thesis used citizen collected rainfall measurements, VGI, to create rainfall surface representations of a small geographic area located within the Kona Districts on the Big Island of Hawaii. The geostatistical methods of ordinary kriging, co-kriging, and Empirical Bayesian Kriging (EBK) were used to interpolate these rainfall point location averages and create rainfall surface maps. Prediction error statistics were generated that corresponded to each surface representation and were used to determine the most accurate method.

The resulting maps that were created for the study area were at least as good as those produced by traditional, authoritative sources. An examination of a cluster of citizen rainfall gauges within a smaller sub-region of the study area was used to create rainfall maps with greater spatial variation compared to maps created from government stations. EBK provided the most accurate results nine out of twelve times, while using the least amount of input.

Chapter 1 – Introduction

1.1 Micro-scale Rainfall Mapping

As climate change has become a scientific concern worldwide, maps showing monthly or annual rainfall are used frequently as one illustration of general climatic trends. These maps are usually drawn at broad geographical scales and can provide a general indication of local variations. As this study explains, only limited research has been completed examining rainfall totals over smaller geographical areas to determine if local variations exist consistently through time. Such local scale studies need spatially-dense data station records previously not available through the sparse network of government weather stations. There are few examples of rainfall studies performed using rainfall records collected by private individuals, a kind of data often referred to as volunteered geographic information (VGI).

Collecting rainfall at a micro-scale level is not typically conducted in most parts of the world. In the United States, the National Weather Service (NWS) has a sparse network of gauges scattered across the country. As a result, rainfall map creation has used a small network of gauges, supplemented by radar data, model simulations, and vegetation patterns. The islands of Hawaii have a unique history of non-NWS rainfall monitoring due to a long history of intensive industrial sugarcane and pineapple plantations. In addition to NWS rain gauge sites, rainfall collection is estimated to be performed by over 2,000 citizens throughout the Hawaiian Islands (Giambelluca et al. 2013). Although these citizen rainfall station records are non-authoritative, they can be considered valuable sources of data because of their increased coverage over NWS sites.

VGI is a result of what is often referred to as citizen science and is not necessarily a new practice. Due to the widespread adoption of information technologies it has become the topic of intense research in the last decade. The massive increase in new sources of data is actively maturing from a social level to a scientific level. It has become common to see VGI in biodiversity monitoring, conservation biology, and early disaster response. One of the most recent emerging fields to utilize VGI is environmental monitoring. It is hoped that this study, using citizen collected data as a supplemental source of data to determine rainfall distributions will help demonstrate the value of VGI based climate research.

1.2 Purpose of This Thesis

The purpose of this study was to create monthly rainfall surface maps with a greater spatial variation than others previously produced using NWS rainfall data. To demonstrate a method of climatological monitoring using VGI, monthly mean rainfall totals were spatially interpolated in a geographic information system (GIS) to create rainfall maps. In addition, the prediction error results of three geostatistical methods of kriging were evaluated to determine the most suitable approach of spatial interpolation.

1.3 Thesis Organization

The organization of this thesis begins with a background review of relevant published information: rainfall collection, VGI, and geostatistical methods for rainfall mapping.

The next chapter discusses the methodology utilized in this study, including an overview of the study area and its rainfall characteristics, the rainfall data sources utilized, and how the GIS data was prepared. In addition, the procedure that was used is explained: exploration of the

rainfall data, a comparison of the relevant kriging methods, and the resulting rainfall surfaces generated with corresponding prediction error results.

A results chapter follows with a month-by-month analysis of the rainfall surfaces created, including how they compare with results from other months. Prediction Standard Error Maps were produced to evaluate the results. Finally, a sub-area of the study area with a denser pattern of rainfall gauges is explored.

The final chapter discusses the conclusions determined in this thesis, validity of the results, impact of the Modifiable Temporal Unit Problem in relation to the arbitrariness of the time increments used, and future work to be considered.

Chapter 2 – Background

The following background section provides a review of relevant literary information: a brief history and corresponding methods of rainfall collection; evolution of VGI and citizen science; and choosing the most suitable approach for rainfall mapping.

2.1 A Brief History of Rainfall Collection

Collecting and measuring rainfall as chronicled by Strangeways (2004; 2010) dates back to the fourth century BC. The first historical accounts were made in India, with additional isolated events occurring in Palestine in the first century AD, China during the 13th century, and Korea in the 15th century. The 17th century saw the Age of Enlightenment spread across Europe, with a new attitude toward science, and a curiosity about measuring the earth and its phenomena. This led to an increase in rainfall collection efforts across Europe.

A detailed history of weather monitoring by Fiebrich (2009) provides an account of the development of rainfall measurements in the United States. The first weather observations were made by Reverend John Campanius Holm near the Delaware River in 1644. Dr. John Lining made daily recordings of rainfall from 1737-1753 in Charleston, South Carolina, looking for a connection between weather, rainfall, and yellow fever. Notable Americans George Washington, Thomas Jefferson, and Benjamin Franklin kept detailed rainfall diaries, all at a time when there were only a handful of weather observers across the Colonies.

According to Fiebrich, weather observations and rainfall measurements spread through the U.S. in the 1800s. Several government, military, and academic measures required records to be kept. In 1870, Congress passed a joint resolution which formed the U.S. Weather Bureau and also required military posts to make weather observations. When the Weather Bureau was

transferred to the Department of Agriculture in 1890, it was recognized that climate data was needed for the entire country, and a larger network of observers would be required. The country's first official voluntary network of observers was formed in 1890, the Cooperative Observer Network (COOP).

Various observer networks continue to serve an active role in rainfall collection measurement. State networks, Mesonets, as well as private and volunteer observation networks are actively growing throughout the country (Fiebrich 2009). One notable network is the Community Collaborative Rain, Hail, and Snow Network (CoCoRaHS). As Cifelli et al. (2005) explain, CoCoRaHS was started by the Colorado Climate Center at Colorado State University (CSU) in Fort Collins in 1998 following a series of devastating storms in northeastern Colorado the previous year. Volunteers have gradually spread throughout Colorado to the Midwestern states, and many parts of the country.

The early centuries' rainfall collectors were motivated by agricultural needs (Strangeways 2010). Today's collectors still measure rainfall for agriculture, but also require the data for hydrology. And as current evidence of global climate change mounts, often indicated by changes in rainfall totals, measuring rainfall has reached a new level importance (Rodda and Dixon 2012).

2.2 Measuring Rainfall

While collecting rainfall data seems like a simple procedure, it is actually somewhat difficult to do with accuracy. As Rodda and Dixon (2012) point out, even though most of today's rain gauges resemble a simple bucket design from centuries ago, they all suffer from various forms of error. While wind is the biggest source of error, rainfall measurements are also compromised by out-splash, evaporation, and the aerodynamic forces surrounding the rain gauge

(Strangeways 2004). Unfortunately, with no standard error-proof gauge implemented, there are estimated to be 80 different types of gauges used at the national level, all prone to the errors mentioned above, including those found at most automatic recording stations (Rodda and Dixon 2012).

Development of a better gauge is much more complex than most assume. Contemporary gauge designs include manual gauges, mechanical float and weight-operated recorders, electrical tipping buckets, electronic weighing rain gauges, capacitance gauges, drop-counting gauges, aerodynamic gauges, and optical rain gauges. In addition, various wind barriers have been utilized including a turf wall and a sunken pit (Strangeways 2010). Placing the gauge in a pit with its opening near ground level provides the most accurate measurement of rainfall, regardless of the rain gauge (Strangeways 2004; Rodda and Dixon 2012).

Questions have been raised about the accuracy of rainfall measurements from the citizen science cooperative networks. CoCoRaHS encourages its members to attend a training session, or view a video or slide show on its website, to avoid making common errors (Cifelli et al. 2005). A study by Wu et al. (2005) on COOP rainfall data found significant discrepancies between COOP and Automatic Weather Data Network (AWDN) measurements. However, when a moving 7-day average was applied to the records, there was a dramatic decrease in discrepancies, suggesting that differing observation times were the cause.

2.3 Volunteered Geographic Information

Humans have been collecting and recording data for hundreds, if not thousands, of years. Scientists, as well as non-scientists, have used this collected data for record keeping, map making, and a wide variety of other pursuits. The first decade of the 21st century saw a change in the technology landscape, with internet users now able to share their collected data with a wider

audience, and the ability for private citizens to create their own geographic information. Goodchild (2007) termed this voluntary, citizen created data, volunteered geographic information (VGI). While many geographers have varying definitions of what VGI encompasses (Elwood 2012), most agree on the core principal that it is data from untrained contributors.

In its more technologically advanced form, VGI is aided in its development by a new wave of Global Positioning System (GPS) enabled devices: phones, handheld receivers, and digital cameras. As Elwood (2008) points out, GPS technology allows the layman to geotag digital data with latitude/longitude coordinates, providing the means to gather and produce VGI.

This new process of making it possible for everyday internet users to contribute content is commonly referred to as Web 2.0. As Goodchild (2007) explains, the change began when Web pages evolved from no longer consisting of content solely provided by a website's owner. Instead of a one-way (client-server) relationship, sites such as eBay, Expedia, and Craig's List allowed users to supply their own content. This user generated content, when melded with geography, can be found on sites such as Google Maps, Wikimapia, OpenStreetMap, and Flickr (Elwood et al. 2012).

VGI finds its roots in citizen science: the process of community or citizen observation of science (Goodchild 2007). A common example is the Audubon Society's Christmas Bird Count, where amateurs perform an inventory of bird populations across the United States (Flanagin and Metzger 2008; Connors et al. 2012). Started in 1900, it is considered one of the longest running citizen science projects.

As rich geographic information becomes readily available to users, it also reveals a lack of information in many locations. Often caused by a lack of public funds or staffing, but also as a result of desired temporal or spatial scales being beyond institutional capabilities, there exist

holes in existing geographic databases (Elwood 2008). As Goodchild (2007) points out, these gaps can be filled by volunteer contributions. When "official" sources are unavailable, citizens can and do act independently, creating patchwork coverage.

So, who are these people? While there is no apparent common denominator that differentiates citizen scientists, most researchers agree that the number and diversity of data creators will continue to expand (Elwood 2008). Some answers can be found in what motivates individuals to collect data, which varies upon situations. With internet related content, self-promotion is often the root (Goodchild 2007). Often the explanation is broader: professional or personal interest, intellectual stimulation, or interest in a personal investment (Connors et al. 2012). A group that has been collecting geographic data longer than most is farmers. The success or failure of their livelihoods has led to their interest in temperature and rainfall data. Many farmers have invested in technology for monitoring and data collection that far exceeds what can be found at a governmental level (Goodchild 2007). However, Gould (1999, 314) argues a simpler explanation, that "there is a geographer in most people".

Those benefitting from the resulting VGI are numerous. Geographic information scientists have a natural interest in the data, but a wide range of physical geographers potentially will as well (Elwood et al. 2012). Opportunities arise for a multitude of research topics at a variety of scales. But as Connors et al. (2012) point out, little effort has been put forth to find data from VGI networks that are independent of more formalized citizen science projects; projects that are rapidly growing in number and size. And while studies can be found of using VGI and citizen science in biodiversity monitoring and conservation biology, they are not as apparent in environmental monitoring. Involving citizens in environmental data collection gets them involved in policy creation and decision making (Connors et al. 2012).

One of the biggest issues facing VGI is credibility of the data. Non-citizen contributed data typically comes with credibility by association, as it is produced by a governmental institution or a credentialed professional (Flanagin and Metzger 2008). In addition, the requisite metadata provides essential information on evaluation and use. But efforts to carry this practice of documentation over to the citizen created data world have not caught on (Goodchild 2007). Without any citation or claims of ownership, it becomes asserted geographic information. And with a diverse spectrum of contributors, there exists the possibility of incorrect, or even contradictory, data (Elwood 2008). This reality is motivating some researchers to create automated techniques to filter data based on its likely quality or credibility (Elwood 2008). But as some consider credentialing volunteers for credibility (Flanagin and Metzger 2008), others remind geographers that citizens with functioning sense perception and the ability to compile and interpret are credible human sensors (Goodchild 2007).

2.4 Choosing a Method of Spatial Interpolation to Map Rainfall

The accurate mapping of rainfall variation over a region requires a dense network of rain gauges and measurements. Often, though, gauges are not located in all areas: mountainous regions, inaccessible topography, or private property. To create a rainfall map from sparse rainfall recording locations, it becomes necessary to estimate rainfall amounts at unrecorded sites using values from neighboring sites (Goovaerts 2000).

Many methods exist for interpolating rainfall data, including one of the simplest involving Thiessen polygons (Thiessen 1911). The U.S. National Weather Service began using an inverse square distance technique in 1972 (Bedient and Huber 1992), but neither of these methods take into account other considerations that can affect rainfall totals, such as elevation (Goovaerts 2000). McCuen (1998) describes a method of using isohyets to overcome the

previous methods' shortcomings. Local knowledge of rainfall catch environments allows for drawing lines of equal rainfall amounts; areas with no rain data can be interpolated from within that area's isohyet. The biggest drawback to this technique is the requirement for an extensive rain gauge network to create accurate isohyets (Goovaerts 2000).

Several other methods can be grouped into deterministic methods: nearest-neighbor, inverse distance-weighting, bicubic splines, and multiquadrics. As O'Sullivan and Unwin (2010) point out, these approaches assume the data at control points are exact and use a specified deterministic mathematical function as their interpolator. But because environmental measurements are subject to errors, and vary spatially, many geostatisticians argue that deterministic methods are not the correct interpolators to use for creating rainfall maps.

Geostatistics provides various methods of interpolation that take advantage of spatial autocorrelation between neighboring sites, allowing unknown site values to be predicted (Goovaerts 2000). These procedures are known as kriging, which refers to the group of generalized least-squares algorithms (Diodato 2005), including ordinary, simple, universal, probability, indicator, and disjunctive kriging. Weights as sample values that have been calculated based on parameters of a variogram model, a plot of the variance of the difference between two variables at two locations, are used to estimate unknown attribute values (Moral 2010). As Goovaerts (2000) explains, the technique's weighting recognizes that measurements close to each other tend to be more similar than those at further distances, therefore receiving a larger weight. The variogram displays the degrading of this spatial correlation among points as their separating distance increases (Moral 2009). Kriging is solved by the formula:

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \varepsilon(\mathbf{s}), \quad (1)$$

where Z(**s**) is the variable of interest, decomposed into a deterministic trend explained by the variogram μ(**s**) and autocorrelated errors form ε(**s**). The symbol **s** represents the location (Esri 2012). All of the different forms of kriging use variations of this formula. Using kriging to make predictions has proven to be more accurate than using deterministic techniques as several authors have shown (Tabios and Salas 1985; Phillips et al. 1992; Goovaerts 2000; Diodato 2005; and Moral 2010). Kriging is often considered the optimal interpolator because it produces estimated values with minimum errors. However, it is worth noting that kriging is an inexact interpolator: the predicted values that are produced may vary from the actual measured values (O'Sullivan and Unwin 2010).

If secondary information such as elevation is available, predictions can be improved with co-kriging (Goovaerts 2000; Moral 2010). Using this multivariate extension of kriging provides the ability to add more densely sampled attributes (Diodato 2005). It has been shown that rainfall amounts increase with increases in elevation: the orographic effect of mountainous areas lifts air vertically, producing condensation due to adiabatic cooling. So while the univariate algorithm of kriging only considers rainfall data, co-kriging can generate more accurate results by using rainfall and elevation data (Moral 2010). In addition, as Goovaerts (2000) suggests, this added auxiliary information can be obtained inexpensively through a digital elevation model (DEM).

Empirical Bayesian Kriging (EBK) is an additional method of kriging. As Pilz and Spöck (2007) explain, while other methods of kriging require manually adjusting parameters, EBK automatically determines these parameters, improving the accuracy of results. Additionally, EBK accounts for the error introduced when the underlying semivariogram is estimated. It is argued that EBK is more accurate for small datasets, requires minimal user manipulation, and generates more accurate standard errors of prediction.

2.5 Summary and Relevance to This Study

The motivation for rainfall collection and record keeping has remained unchanged for agricultural needs throughout history. What has changed is the recent broader use of this data to monitor climate change. Accuracy is a primary concern during the collection process. While many of the current types of rain gauges resemble those used a century ago, there is no single agreed upon "best" rain gauge design. In addition, local environmental factors and gauge location can affect the accuracy of the measurements.

Citizen collected rainfall measurements are one of the earliest forms of VGI. While the more modern forms of this citizen science involve GPS, the internet, or other electronic monitoring and communication devices, in its simplest form VGI is data that has been collected by untrained contributors. Data credibility is a significant concern, but often there is no official certified-source for a location or event, and the resulting VGI data can fill in gaps of more traditional sources.

To achieve the objective of this thesis of creating rainfall maps at greater spatial variation than those created from NWS sources, a denser network of rainfall stations was necessary, and could only be found in citizen contributed sources. And while dense networks of rainfall collectors are somewhat rare, a suitable group was found in Hawaii.

The geostatistical method of kriging is considered optimal for creating rainfall surface maps. Co-kriging can improve the accuracy of kriging results by including additional contributing factors, or covariates in its calculations. EBK also provides accurate results, without additional data, but instead by limiting user input.

As described above, condensation increases due to adiabatic cooling in mountainous areas as air is lifted vertically. For this reason, using elevation data as covariate information in

co-kriging can produce more accurate results than kriging alone. Other less impactful physical factors may have the potential to increase rainfall as well. Since the increase in elevation is at differing rates within a geographic area, that measure of steepness, slope, could be a contributing factor in differing rates of rainfall. In addition, the direction that a surface faces, its aspect, could also be a factor in the rate of rainfall. For these reasons, elevation, slope, and aspect were included as covariates in the co-kriging operations.

A comparison of the results using co-kriging to those using EBK was sought, as the literature reviewed did not favor one method over the other, but instead touted both methods as superior to all others. Ordinary kriging was also included in the interpolation comparison to serve as a check that the two preferred methods did in fact produce better results.

Chapter 3 – Methodology

3.1 Study Area

In Hawaii, rainfall measurement records have been kept by citizens for agricultural needs for at least as long as commercial plantations have existed. Access to citizen collected rainfall totals in an area within the North and South Kona Districts on the Big Island of Hawaii provided an opportunity to analyze rainfall patterns of a smaller area. With rainfall records and elevation data available across a smaller region of the Big Island of Hawaii, it was possible to map more detailed spatial variations than those previously produced.

3.1.1 The Kona Coffee Region

The region is home to an estimated 800 coffee farms, located on the western slopes of Hualālai and Mauna Loa, on the western side of the island. A tradition of family-owned farms is maintained, with the average farm size measuring less than five acres. While rainfall data is not collected or maintained at all farms, the Kona Coffee Farmers Association has a significant number of its 240 members who do.

The study area utilized was created by using the western 1/3 of the Kona Districts, as it represents an approximate equidistant border from the rainfall station record sites. As seen in Figure 1, the western border is represented by the island's western coastline, the northern and southern borders are the northern and southern borders of the Kona Districts, and the eastern border has been created as the equidistant border. The size of the study area measures 347 square miles.

The elevation of the study area increases from west to east (Figure 2). The western and

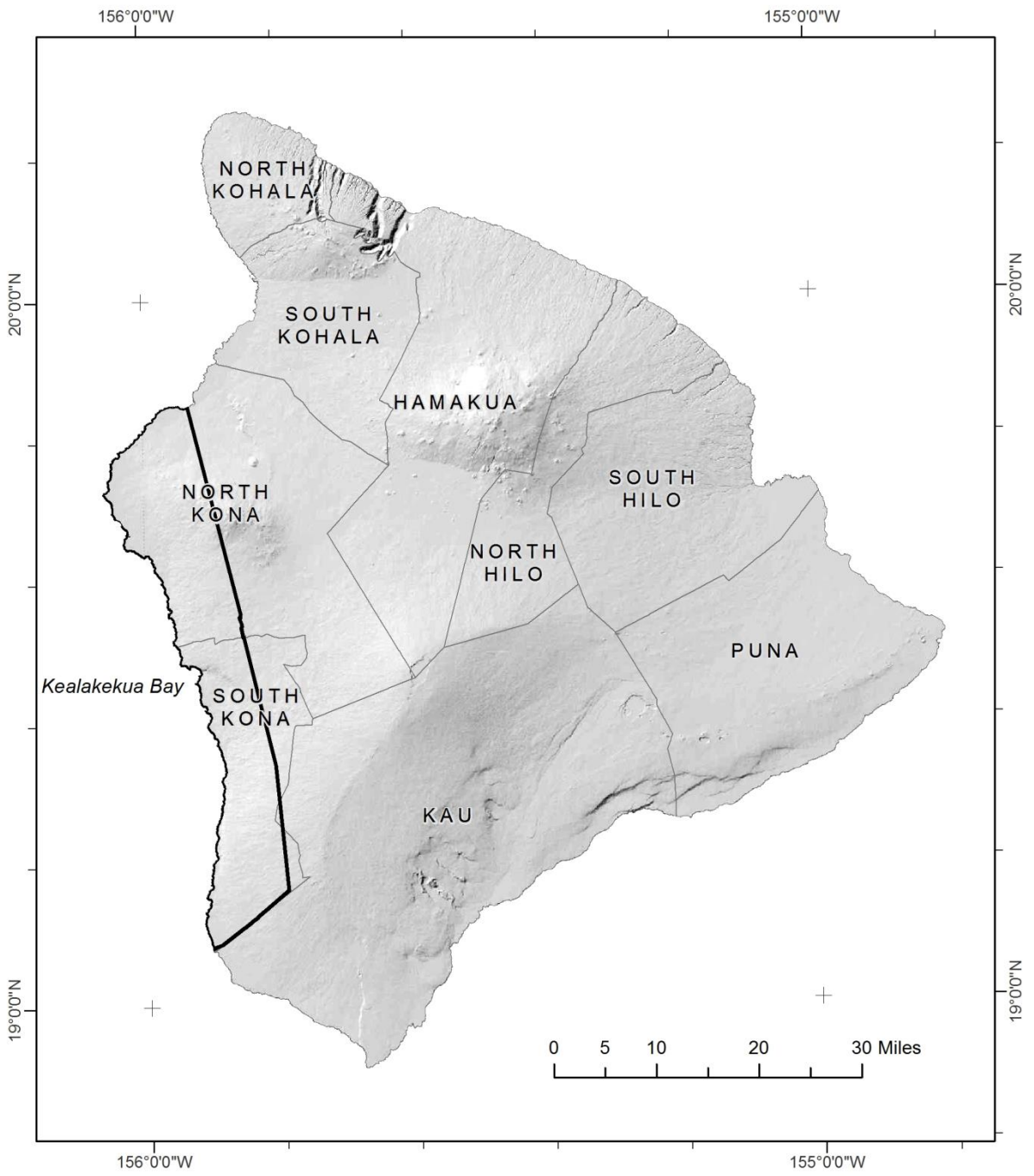


Figure 1: Study area outlined on the western coast of the Big Island of Hawaii, covering portions of the North and South Kona Districts.

northern borders are coastline, and are generally at sea level. The western slopes of the volcano Hualālai is located in northern part of the study area, and provides the highest elevations and

greatest slope angles. In the south are the most western slopes of the volcano Mauna Loa. In this region, the increase in slope angles is much less dramatic.

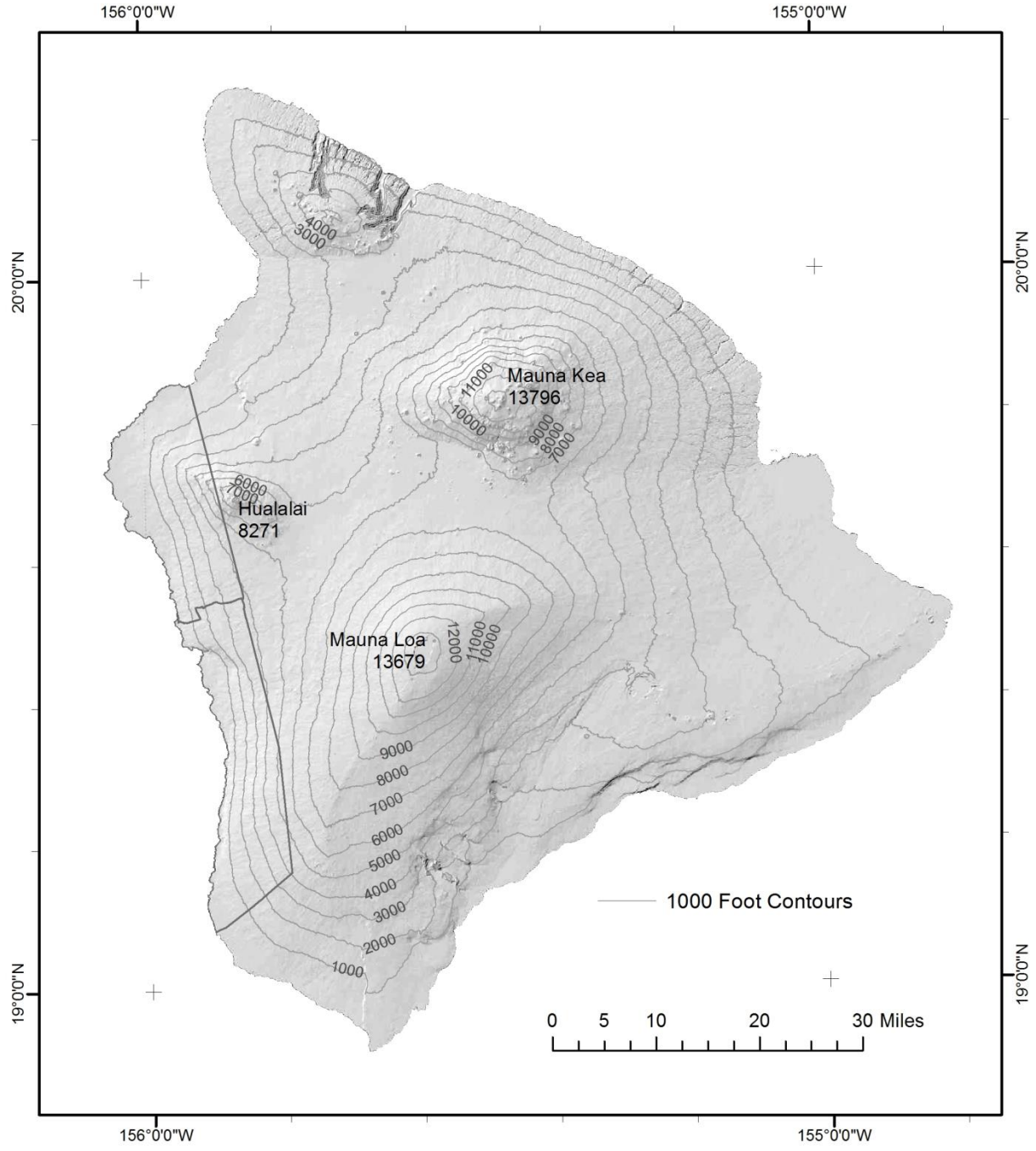


Figure 2: The Big Island of Hawaii with elevation contours spaced at 1,000 feet apart.

3.1.2 Rainfall Characteristics

The Hawaiian Islands are under the influence of northeasterly trade-winds, which produce frequent trade-wind showers, especially during the summer months. On the Big Island of Hawaii, the volcanic mountains of Mauna Kea in the north and Mauna Loa in the south act to block these winds, resulting in higher rainfall on their windward sides. On their lee sides, the resulting split airflow weakens the trade-winds, decelerating their flow, and minimizes their presence on the eastern side of the island (Chen and Nash 1994).

Along the Kona coast study area, more rainfall is typically experienced in the summer months than the winter. Daytime onshore sea-breezes ascend the mountain slopes and mix with descending trade-winds transitioning through the saddle of the two volcanic mountains, resulting in late afternoon showers (Giambelluca et al. 1986). In the evening, the local circulation is characterized by downslope land breezes, and fewer rainfall events (Chen and Nash 1994).

The Kona region experiences a less persistent sea-breeze circulation during the winter, leading to less rainfall. Cutoff lows in the upper level subtropical westerlies, also called Kona storms, bring widespread and abundant rains throughout the island (Chu and Chen 2005).

A study of rainfall patterns throughout the Hawaiian Islands was undertaken by Giambelluca et al. (2013) and published online as the Online Rainfall Atlas of Hawai'i. Using 30-year rain gauge averages from 1978-2007, the results are viewable via an easily navigated interactive website map. In addition, 8.1 arcsecond resolution raster files of individual months and an annual summary are downloadable. The study cites a decline in stations used over the 30-year period, with a high statewide of 1,067, finishing with a low of 340. To gap fill the missing data, rain gauge data were supplemented with other coarser, non-gauge data. The online Atlas is recognized as one of the best sources of regional rainfall maps of Hawaii.

3.2 Rainfall Data Sources

Traditionally, rainfall mapping studies attempt to acquire thirty years of rainfall data, as the extended duration averages out months with higher, or lower, anomalous measurements. Thirty years of data were desired for this study, but a cursory review of the study area's rainfall records and sources revealed that obtaining records for greater than five years would not be possible. With the goal of achieving the best coverage of the region with the densest network of gauges possible, a five-year average was used.

Monthly rainfall measurements for the period January 2007-December 2011 were sought for this study. Citizen collected rainfall records were primarily used, while other sources were searched for and added to create a dense network of rainfall station gauges. Rainfall records were requested from the following sources:

- College of Tropical Agriculture and Human Resources (CTAHR), University of Hawaii
- Craigslist (ad placed)
- Hawaii Farmers Union United
- Hawaii Organic Farmers
- Hawaii Tropical Fruit Growers
- Kona Coffee Council
- Kona Coffee Farmers Association
- Kona County Farm Bureau
- Kona Outdoor Circle - Master Gardeners
- National Oceanic Atmospheric Administration – National Climatic Data Center
- National Weather Service - Hydronet
- Weather Underground

The following sections describe each of the different data sources that rainfall measurements were acquired from for use in this study.

3.2.1 Citizen Collected Rainfall Data

To obtain rainfall measurement data from the various farmer and agricultural associations, a questionnaire (Appendix A) was sent out via email on 18 October 2012, and contributions were sought until 22 January 2013. Requested information included rainfall amounts, unit of measurement, type of rain gauge used, frequency of collection, and latitude/longitude of gauge. Additional comments from collectors were also accepted and reviewed.

Responses to inquiries were very enthusiastic as many people in these groups wanted to participate. Data was received during the period of 18 October 2012 through 7 February 2013. Unfortunately, many rainfall records were unusable as they only contained data for a couple of years or they were too incomplete. On the other hand, a few farmers presented meticulously maintained records that included complete daily data that dated back several decades. In addition, the records were received in a variety of formats including Excel, Word, Adobe PDF, and hand written text.

The specific rainfall gauge location on a property varied between contributors in the farmer group, but an overall strategy was prevalent: gauges were placed in open spaces, away from trees and roof edges, but not too far from the convenience of a residence or other frequently visited structure. Geographic coordinates were included from the collectors with datasets, as instructions for determining latitude and longitude via Google Maps was included in the questionnaire. It is estimated that these supplied coordinates are accurate to within 20 feet.

A few rainfall records had occasional monthly totals missing. If totals were missing for more than one month during a five year record for a specific month, the entire station record was not used. For example, if a station record was missing January 2007 and January 2008 from its

January measurements, that station's records were not used. Station records from fifteen contributors were not used for this reason. If one month was missing, the station's record was used, and the value added. To calculate a single missing month's total, a total was estimated by comparing to neighboring stations with consistently similar measurements, determining the average percentage the neighboring station varied with existing measurements, and applying an average of that difference to known measurements of neighboring stations. A formula to solve for "missing" is illustrated as follows:

	Station A	Neighbor 1 (X)		Neighbor (Y)	
Year	Actual	Actual	A's diff.	Actual	A's diff.
2007	B	F	=(B-F)/B	K	=(B-K)/B
2008	missing	G	--	L	--
2009	C	H	=(C-H)/C	M	=(C-M)/C
2010	D	I	=(D-I)/D	N	=(D-N)/D
2011	E	J	=(E-J)/E	O	=(E-O)/E
		Avg. diff.	=(sum)/4	Avg. diff.	=(sum)/4
		Applied to known (G) = X	=(G*Avg diff)+G	Applied to known (L) = Y	=(L*Avg diff)+L

missing = (X+Y)/2

(2)

Using sample rainfall measurements, "missing" is shown solved as follows:

	Station A	Neighbor 1		Neighbor 2	
Year	Actual	Actual	A's diff.	Actual	A's diff.
2007	1.1	1.2	-9%	1.1	0%
2008	missing	0.9	--	0.8	--
2009	1.3	1.4	-8%	1.3	0%
2010	1.2	1.2	0%	1.1	8%
2011	1.0	1.1	-10%	0.9	10%
		Avg. diff. =	-7%	Avg. diff. =	5%
		Applied to known (G) = X	= 0.840	Applied to known (L) = Y	= 0.837

missing = 0.838

(3)

Thirty five data station records, more than half of the data set, were supplied by farmers in the various farmer associations (Table 1); the majority of those records were received from the Kona Coffee Farmers Association. The location of these collection stations are shown, with all other data sources, in Figure 3. Importantly, this substantial portion of data used in this study is characterized by not being collected by scientists or trained rainfall collection specialists. It is volunteered geographic information that was collected and measured by citizens using their own methods and devices, at schedules and frequencies determined individually and without protocols. There is no metadata assessing its fitness for use. Errors that could not be precisely identified are likely to exist. The accuracy of the results of this study is only as accurate as those non-certified station records.

Table 1: Summary of station records provided by farmers.

ID	GAUGE	LATDD	LONDD	ELEV (FT)	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
11	Davis Vantage Pro II	19.677172	-155.967480	1444	2.508	3.204	2.724	3.808	4.768	4.980	4.396	2.518	5.114	2.930	2.346	4.224
12	plastic Tru-Chek 6"	19.601639	-155.965506	208	1.752	1.396	0.872	1.544	2.382	2.852	3.366	2.060	2.210	2.112	1.690	3.328
13	plastic Tru-Chek 6"	19.559082	-155.931090	1538	2.114	2.708	3.168	2.630	5.364	4.370	5.314	4.612	4.880	3.382	2.860	3.542
14	10" wedge	19.157153	-155.859531	1603	1.270	1.436	1.378	0.816	1.288	1.266	1.850	1.080	1.722	1.898	1.280	3.766
15	10" wedge	19.160386	-155.854406	1560	1.180	1.394	1.454	0.948	1.022	1.146	1.764	0.962	1.338	2.122	1.346	3.886
16	10" wedge	19.140831	-155.851022	1601	1.340	1.518	1.410	0.842	1.024	1.046	1.666	0.850	1.126	1.930	1.184	3.976
17	10" wedge	19.133747	-155.845567	1635	1.282	1.452	1.392	0.932	0.838	1.018	1.598	0.774	0.972	1.788	1.238	3.782
18	10" wedge	19.143047	-155.868586	1398	1.228	1.332	1.198	0.712	0.914	1.004	1.520	0.904	2.494	2.094	1.290	3.602
19	10" wedge	19.149050	-155.861053	1490	1.054	1.300	1.304	0.974	1.148	1.134	1.854	0.946	1.708	2.236	1.322	3.706
20	10" wedge	19.132508	-155.854833	1329	1.260	1.344	1.344	0.968	0.900	1.032	1.568	0.692	1.092	2.016	1.268	3.558
21	10" wedge	19.154819	-155.869172	1301	1.208	1.464	1.362	0.816	0.948	1.034	1.688	0.778	2.314	2.310	1.286	3.708
22	10" wedge	19.147283	-155.865856	1350	1.086	1.314	1.268	0.902	1.032	1.032	1.652	0.752	1.648	2.082	1.298	3.548
23	10" wedge	19.140650	-155.865572	1240	1.152	1.484	1.240	0.906	0.938	1.086	1.762	0.758	1.558	2.242	1.240	3.872
24	10" wedge	19.133944	-155.860972	1212	1.232	1.426	1.246	0.896	0.938	1.006	1.626	0.598	1.130	2.146	1.072	3.480
25	10" wedge	19.126333	-155.853919	1242	1.226	1.522	1.274	0.828	0.800	0.878	1.382	0.732	1.008	1.994	1.138	3.518
26	10" wedge	19.144150	-155.840544	2007	1.416	1.716	1.548	0.956	1.142	1.038	1.800	1.180	1.016	1.390	1.238	4.256
27	10" wedge	19.144900	-155.830731	2287	1.484	1.696	1.798	1.084	1.144	1.050	1.954	1.122	1.004	1.480	1.338	4.828
28	10" wedge	19.151275	-155.830533	2467	1.358	1.758	1.736	1.116	1.166	0.862	1.856	0.930	1.192	1.508	1.300	4.908
29	10" wedge	19.154233	-155.832056	2441	1.360	1.838	1.906	1.302	1.372	0.884	2.112	0.950	1.306	1.636	1.394	5.018
30	10" wedge	19.163936	-155.836400	2439	1.352	1.672	1.768	1.308	1.204	0.932	1.938	1.014	1.278	1.726	1.486	4.514
31	10" wedge	19.154469	-155.841917	2112	1.184	1.506	1.544	0.984	1.192	1.056	1.748	1.100	1.176	1.482	1.268	4.168
32	10" wedge	19.159056	-155.850233	1935	1.144	1.496	1.360	0.922	1.432	1.156	1.892	1.096	1.618	1.788	1.262	3.762
33	Cylinder tube with 1" cap	19.207935	-155.867028	2223	1.918	1.296	1.908	1.282	1.874	1.532	2.128	1.362	3.022	2.614	2.062	3.676
34	2.5"x2.5"x13" Wedge	19.572500	-155.938056	1386	2.130	2.640	2.822	2.488	5.048	4.456	5.694	4.594	3.900	4.690	2.790	4.024
35	4" wedge	19.450000	-155.890000	778	1.774	2.334	4.422	3.840	4.768	5.914	7.458	5.182	5.428	5.568	4.940	6.704
36	10" wedge	19.630037	-155.949737	1442	2.050	2.472	2.998	2.578	4.750	4.650	6.112	4.254	4.736	3.324	2.056	4.086
37	7" plastic Taylor	19.450553	-155.882041	1067	1.840	2.360	3.160	1.940	3.300	2.780	4.540	3.340	3.260	3.640	4.100	4.700
38	Tru Check brand plastic	19.636232	-155.956920	1163	1.912	2.440	2.214	2.692	3.940	4.494	4.774	3.142	4.718	3.208	1.398	3.920
39	2 3/4" x 2 3/4" x 13.5" w	19.293646	-155.867050	1814	1.830	4.090	2.994	2.396	3.718	3.128	4.392	3.184	4.796	5.868	2.560	6.410
40	4" wedge	19.533789	-155.929827	1339	2.000	1.986	3.108	2.702	4.674	3.708	5.742	4.180	4.558	3.250	2.626	4.494
41	Davis Vantage Pro II	19.339789	-155.874108	1035	1.718	2.526	3.100	1.260	1.740	1.970	3.252	2.548	2.634	3.748	2.372	2.998
42	4" wedge	19.723277	-155.971616	2062	2.394	2.558	1.450	2.994	3.468	3.296	3.796	1.464	2.474	2.106	1.580	5.480
43	4" wedge	19.535436	-155.926653	1494	1.875	2.378	3.577	3.265	5.229	4.370	5.879	4.387	4.955	3.505	2.441	4.769
44	4" wedge	19.698737	-155.972530	1638	2.088	2.771	2.479	3.699	4.375	4.488	4.600	1.944	3.504	3.042	2.136	5.154
45	4" wedge	19.525041	-155.911795	1972	2.310	3.050	4.790	4.400	5.420	4.950	7.400	5.410	5.660	5.910	5.110	6.660

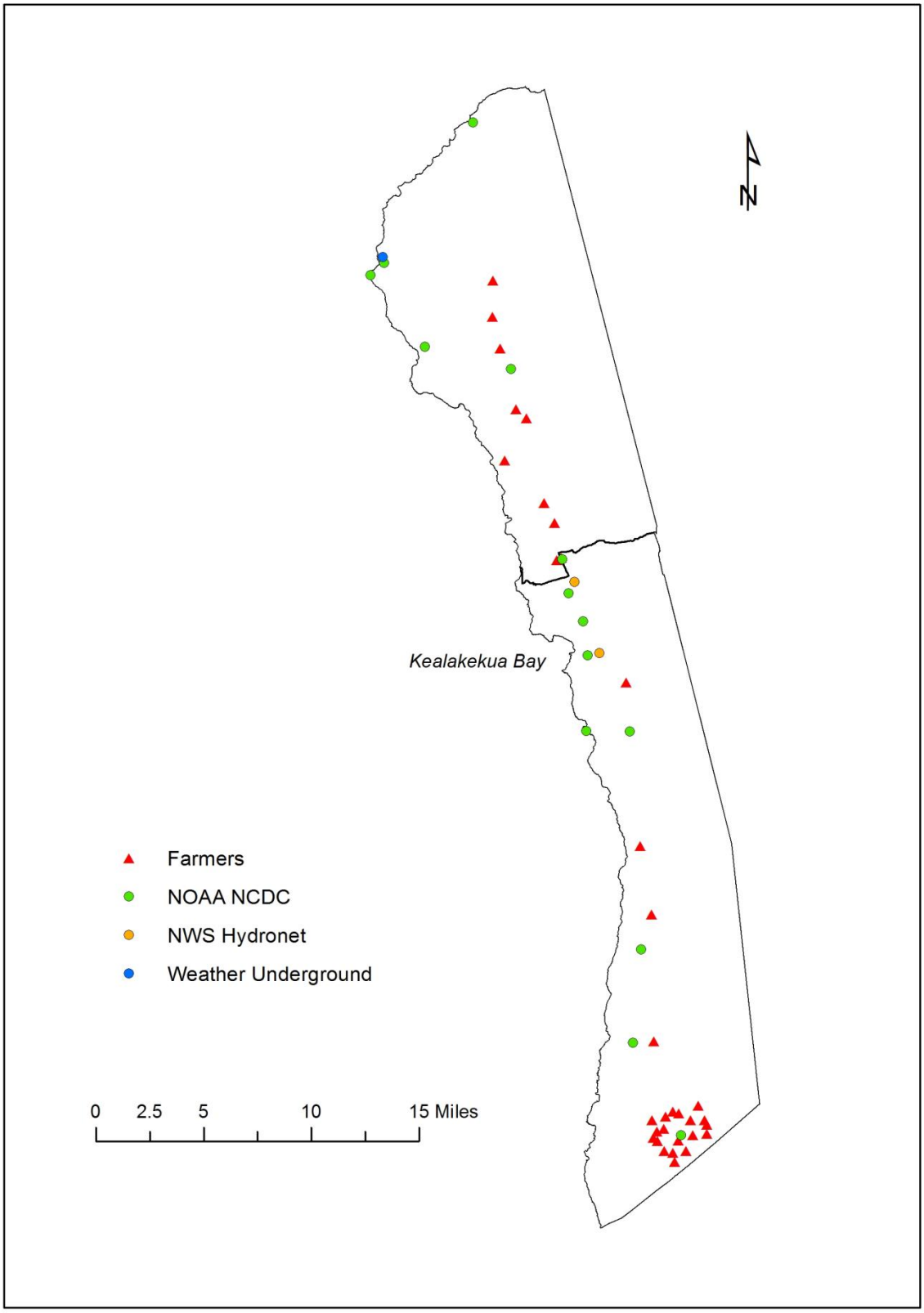


Figure 3: Rainfall station locations categorized by source.

3.2.2 National Oceanic Atmospheric Administration – National Climatic Data Center

Archived climate data was obtained from National Oceanic Atmospheric Administration's (NOAA) National Climatic Data Center (NCDC) website. The website provides daily, monthly, and annual archived global climate data from a searchable map. Individual weather station locations are displayed on the site and the user selects the specific stations and climate data needed.

The interactive site shows 243 stations on the Big Island of Hawaii with 52 stations in the Kona Districts. The site allows the user to review the time span that records are available for each station. It was discovered that not all stations contained current records, as several appeared to have been decommissioned. The records for 23 stations were complete for the study's time period of 2007-2011 and were subsequently downloaded.

Reviewing the downloaded data of the NOAA stations revealed that some stations had records with missing months. Those values were filled using the same method as described in the citizen station record data section above. Stations with more than one month missing in its five year record for a specific month were not included (e.g. if January was missing for 2008 and 2009, the station was not used). Some stations, although listed on the NOAA site as having a full record of data, were absent of rainfall data. A total of 14 stations had usable monthly rainfall data. The locations of these stations are shown in Figure 3, and their measurements are summarized in Table 2.

Table 2: Summary of NOAA NCDC rainfall station measurements.

NAME	ID	LATDD	LONDD	ELEV (FT)	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
MILOLII 2.34	46	19.207000	-155.882000	1171	1.638	2.060	1.696	1.000	0.922	1.658	2.786	1.320	2.370	2.224	1.080	2.428
OPIHIHALE 2 24.1	47	19.270000	-155.875000	1358	1.466	2.602	2.100	1.596	1.750	1.694	3.970	2.150	4.304	3.902	2.210	4.294
KE AHOLE POINT 68.13	48	19.728000	-156.059000	20	0.292	0.592	0.092	0.422	0.678	0.374	0.336	0.230	0.636	0.104	0.092	1.278
HONAUNAU 27	49	19.417000	-155.880000	938	1.216	1.586	2.194	1.636	2.544	2.800	2.732	3.014	2.332	2.660	2.438	4.176
KEALAKEKUA 26.2	50	19.492000	-155.912000	1480	0.618	1.360	2.848	1.406	2.322	1.782	3.098	2.468	1.996	2.262	3.750	2.582
LANIHAU 68.2	51	19.663000	-155.960000	1529	2.694	2.354	3.076	3.688	4.686	5.210	4.292	3.392	5.362	3.636	2.422	4.650
KAINALIU 73.2	52	19.534000	-155.926000	1499	1.548	2.248	2.930	2.532	4.222	2.292	4.650	3.626	4.106	2.372	1.864	3.494
NAPOOPOO 28	53	19.469000	-155.909000	400	0.470	0.804	1.046	1.120	1.644	1.056	1.628	1.794	1.264	1.482	2.404	2.638
KAILUA KONA KE AHOLE ARPT	54	19.736000	-156.049000	43	0.966	0.930	0.392	1.014	0.540	0.746	0.644	0.140	0.238	0.322	0.280	1.922
KONA VILLAGE 93.8	55	19.830000	-155.984000	20	0.868	1.028	0.340	0.158	0.584	0.224	0.220	0.142	0.034	0.204	0.368	3.036
HONOKOHAU HARB 68.14	56	19.679000	-156.021000	30	1.150	1.940	1.704	1.712	1.852	2.012	1.964	1.070	1.482	0.840	0.968	4.020
KEALAKEKUA 4 74.8	57	19.511000	-155.922000	1421	0.818	1.356	2.190	1.604	2.396	1.732	2.672	2.488	2.440	1.736	2.518	2.224
PUUHONUA O HONA 27.4	58	19.418000	-155.911000	16	0.322	0.536	0.976	0.734	0.840	1.078	0.840	0.744	1.034	0.488	0.464	1.810
KAPUA 2.36	59	19.144000	-155.849000	1749	1.442	1.490	1.544	1.020	1.028	1.188	1.734	0.934	1.414	1.758	1.122	4.160

3.2.3 National Weather Service - Hydronet

Hydronet data is National Weather Service (NWS) rainfall data from automated weather stations that has been archived. The system collects 15-minute rainfall totals that are provided by month in a comma delimited format. There are 70 gauges located in the Hawaiian Islands, with three in the Kona region:

- HI-82 Waiaha
- HI-84 Kealakekua
- HI-95 Honaunaau 2

The comma delimited files were downloaded from the NWS Hydronet website, converted to Excel, and then each monthly total was aggregated with the relevant month's starting and ending times. The NWS maintains there is no quality control over the data and therefore cannot certify its accuracy. In addition, it is stated that there may be some time periods where data is missing, in which case a value of -99.999 is entered.

After downloading compressed files for the Big Island of Hawaii, unzipping and tabulating them, it was discovered that station HI-82 Waiaha reported data for only the first four months of 2007. All subsequent months contained values of -99.999. It was surmised that this

station was decommissioned, and it was not used in this study. For the Kealakekua and Honaunaau 2 stations, all months for 2007-2011 were in the archived records. Both stations were used and their locations shown in Figure 3; their rainfall measurements are summarized by month in Table 3.

Table 3: Summary of NWS Hydronet rainfall station measurements.

NAME	ID	LATDD	LONDD	ELEV (FT)	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
HI-84 Kealakekua	61	19.518656	-155.917655	1644	1.816	1.996	2.602	2.588	4.616	3.150	5.040	3.908	5.060	3.874	3.050	3.984
HI-95 Honaunau 2	62	19.470488	-155.900772	724	1.716	2.100	3.104	2.252	4.042	2.670	4.490	3.564	3.730	4.722	4.406	4.924

3.2.4 Weather Underground

Weather Underground functions as a commercial, non-government weather forecast provider. They provide free weather information via the web to users worldwide. In addition, they have developed a network of citizen weather collectors, operating what they have termed "personal weather stations". They claim to have almost 23,000 stations in the United States and over 13,000 in other countries across the world. Collected data is uploaded to their website where visitors can monitor current weather conditions at participating sites, as well as view historical data.

In the Kona Districts, there are nine personal weather station sites provided in the Weather Underground network. Of the nine potential stations, only three list having rainfall records for the time period 2007-2011 used in this study. These three records were retrieved and each month's rainfall data, in addition to latitude and longitude coordinates, were input into an Excel document.

Reviewing the retrieved the data revealed that the Kahuku station was missing 19 months of data. In addition, two of the months appeared to have potentially erroneous records: totals of

14.71 and 17.03, when most months were typically close to one inch, or less. While these measurements are possible, the 19 months of missing data exceeded the gap filling requirements of a maximum of 12 months (one month of the five, per year) of missing data. For this reason, this station's data was considered incomplete and not used in this study.

The Kaloko-Honokohau station was also missing 19 months of data. In addition, one of the station's recorded months had a measurement of 65.23 inches of rain. Most readings for this station ranged from 0.5-5.0 inches. This station's data was considered incomplete and not used in this study.

The Keahole Point station appeared to be usable, as all months contained data, and all rainfall measurements were within a plausible range. The Keahole Point station was the only Weather Underground station used; its location shown in Figure 3 and its rainfall measurements are summarized in Table 4.

Table 4: Summary of Weather Underground rainfall station measurements.

NAME	ID	LATDD	LONDD	ELEV (FT)	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Keahole Point	60	19.740000	-156.050000	46	0.966	0.930	0.390	1.206	0.538	0.746	0.632	0.134	0.236	0.320	0.280	1.910

3.2.5 Summary of Rainfall Measurement Stations

A summary of the sources of rainfall gauge locations is shown in Table 5. Geographic location of the gauge locations, categorized by source, is shown in Figure 3. All monthly 5-year average data for individual stations were tabulated in a master summary rainfall file (Table 6).

Table 5: Summary of rainfall station sources.

DATA SOURCE	# OF RAINFALL STATIONS
Farmers	35
NOAA NCDC	14
NOAA Hydronet	2
Weather Underground	1
Total	**52**

Table 6: Master summary rainfall measurement file.

ID	LATDD	LONDD	ELEV (FT)	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
11	19.677172	-155.967480	1444	2.508	3.204	2.724	3.808	4.768	4.980	4.396	2.518	5.114	2.930	2.346	4.224
12	19.601639	-155.965506	208	1.752	1.396	0.872	1.544	2.382	2.852	3.366	2.060	2.210	2.112	1.690	3.328
13	19.559082	-155.931090	1538	2.114	2.708	3.168	2.630	5.364	4.370	5.314	4.612	4.880	3.382	2.860	3.542
14	19.157153	-155.859531	1603	1.270	1.436	1.378	0.816	1.288	1.266	1.850	1.080	1.722	1.898	1.280	3.766
15	19.160386	-155.854406	1560	1.180	1.394	1.454	0.948	1.022	1.146	1.764	0.962	1.338	2.122	1.346	3.886
16	19.140831	-155.851022	1601	1.340	1.518	1.410	0.842	1.024	1.046	1.666	0.850	1.126	1.930	1.184	3.976
17	19.133747	-155.845567	1635	1.282	1.452	1.392	0.932	0.838	1.018	1.598	0.774	0.972	1.788	1.238	3.782
18	19.143047	-155.868586	1398	1.228	1.332	1.198	0.712	0.914	1.004	1.520	0.904	2.494	2.094	1.290	3.602
19	19.149050	-155.861053	1490	1.054	1.300	1.304	0.974	1.148	1.134	1.854	0.946	1.708	2.236	1.322	3.706
20	19.132508	-155.854833	1329	1.260	1.344	1.344	0.968	0.900	1.032	1.568	0.692	1.092	2.016	1.268	3.558
21	19.154819	-155.869172	1301	1.208	1.464	1.362	0.816	0.948	1.034	1.688	0.778	2.314	2.310	1.286	3.708
22	19.147283	-155.865856	1350	1.086	1.314	1.268	0.902	1.032	1.032	1.652	0.752	1.648	2.082	1.298	3.548
23	19.140650	-155.865572	1240	1.152	1.484	1.240	0.906	0.938	1.086	1.762	0.758	1.558	2.242	1.240	3.872
24	19.133944	-155.860972	1212	1.232	1.426	1.246	0.896	0.938	1.006	1.626	0.598	1.130	2.146	1.072	3.480
25	19.126333	-155.853919	1242	1.226	1.522	1.274	0.828	0.800	0.878	1.382	0.732	1.008	1.994	1.138	3.518
26	19.144150	-155.840544	2007	1.416	1.716	1.548	0.956	1.142	1.038	1.800	1.180	1.016	1.390	1.238	4.256
27	19.144900	-155.830731	2287	1.484	1.696	1.798	1.084	1.144	1.050	1.954	1.122	1.004	1.480	1.338	4.828
28	19.151275	-155.830533	2467	1.358	1.758	1.736	1.116	1.166	0.862	1.856	0.930	1.192	1.508	1.300	4.908
29	19.154233	-155.832056	2441	1.360	1.838	1.906	1.302	1.372	0.884	2.112	0.950	1.306	1.636	1.394	5.018
30	19.163936	-155.836400	2439	1.352	1.672	1.768	1.308	1.204	0.932	1.938	1.014	1.278	1.726	1.486	4.514
31	19.154469	-155.841917	2112	1.184	1.506	1.544	0.984	1.192	1.056	1.748	1.100	1.176	1.482	1.268	4.168
32	19.159056	-155.850233	1935	1.144	1.496	1.360	0.922	1.432	1.156	1.892	1.096	1.618	1.788	1.262	3.762
33	19.207935	-155.867028	2223	1.918	1.296	1.908	1.282	1.874	1.532	2.128	1.362	3.022	2.614	2.062	3.676
34	19.572500	-155.938056	1386	2.130	2.640	2.822	2.488	5.048	4.456	5.694	4.594	3.900	4.690	2.790	4.024
35	19.450000	-155.890000	778	1.774	2.334	4.422	3.840	4.768	5.914	7.458	5.182	5.428	5.568	4.940	6.704
36	19.630037	-155.949737	1442	2.050	2.472	2.998	2.578	4.750	4.650	6.112	4.254	4.736	3.324	2.056	4.086
37	19.450553	-155.882041	1067	1.840	2.360	3.160	1.940	3.300	2.780	4.540	3.340	3.260	3.640	4.100	4.700
38	19.636232	-155.956920	1163	1.912	2.440	2.214	2.692	3.940	4.494	4.774	3.142	4.718	3.208	1.398	3.920
39	19.293646	-155.867050	1814	1.830	4.090	2.994	2.396	3.718	3.128	4.392	3.184	4.796	5.868	2.560	6.410
40	19.533789	-155.929827	1339	2.000	1.986	3.108	2.702	4.674	3.708	5.742	4.180	4.558	3.250	2.626	4.494
41	19.339789	-155.874108	1035	1.718	2.526	3.100	1.260	1.740	1.970	3.252	2.548	2.634	3.748	2.372	2.998
42	19.723277	-155.971616	2062	2.394	2.558	1.450	2.994	3.468	3.296	3.796	1.464	2.474	2.106	1.580	5.480
43	19.535436	-155.926653	1494	1.875	2.378	3.577	3.265	5.229	4.370	5.879	4.387	4.955	3.505	2.441	4.769
44	19.698737	-155.972530	1638	2.088	2.771	2.479	3.699	4.375	4.488	4.600	1.944	3.504	3.042	2.136	5.154
45	19.525041	-155.911795	1972	2.310	3.050	4.790	4.400	5.420	4.950	7.400	5.410	5.660	5.910	5.110	6.660
46	19.207000	-155.882000	1171	1.638	2.060	1.696	1.000	0.922	1.658	2.786	1.320	2.370	2.224	1.080	2.428
47	19.270000	-155.875000	1358	1.466	2.602	2.100	1.596	1.750	1.694	3.970	2.150	4.304	3.902	2.210	4.294
48	19.728000	-156.059000	20	0.292	0.592	0.092	0.422	0.678	0.374	0.336	0.230	0.636	0.104	0.092	1.278
49	19.417000	-155.880000	938	1.216	1.586	2.194	1.636	2.544	2.800	2.732	3.014	2.332	2.660	2.438	4.176
50	19.492000	-155.912000	1480	0.618	1.360	2.848	1.406	2.322	1.782	3.098	2.468	1.996	2.262	3.750	2.582
51	19.663000	-155.960000	1529	2.694	2.354	3.076	3.688	4.686	5.210	4.292	3.392	5.362	3.636	2.422	4.650
52	19.534000	-155.926000	1499	1.548	2.248	2.930	2.532	4.222	2.292	4.650	3.626	4.106	2.372	1.864	3.494
53	19.469000	-155.909000	400	0.470	0.804	1.046	1.120	1.644	1.056	1.628	1.794	1.264	1.482	2.404	2.638
54	19.736000	-156.049000	43	0.966	0.930	0.392	1.014	0.540	0.746	0.644	0.140	0.238	0.322	0.280	1.922
55	19.830000	-155.984000	20	0.868	1.028	0.340	0.158	0.584	0.224	0.220	0.142	0.034	0.204	0.368	3.036
56	19.679000	-156.021000	30	1.150	1.940	1.704	1.712	1.852	2.012	1.964	1.070	1.482	0.840	0.968	4.020
57	19.511000	-155.922000	1421	0.818	1.356	2.190	1.604	2.396	1.732	2.672	2.488	2.440	1.736	2.518	2.224
58	19.418000	-155.911000	16	0.322	0.536	0.976	0.734	0.840	1.078	0.840	0.744	1.034	0.488	0.464	1.810
59	19.144000	-155.849000	1749	1.442	1.490	1.544	1.020	1.028	1.188	1.734	0.934	1.414	1.758	1.122	4.160
60	19.740000	-156.050000	46	0.966	0.930	0.390	1.206	0.538	0.746	0.632	0.134	0.236	0.320	0.280	1.910
61	19.518656	-155.917655	1644	1.816	1.996	2.602	2.588	4.616	3.150	5.040	3.908	5.060	3.874	3.050	3.984
62	19.470488	-155.900772	724	1.716	2.100	3.104	2.252	4.042	2.670	4.490	3.564	3.730	4.722	4.406	4.924

3.3 Non-GIS Data Exploration

The summary of rainfall measurements data was explored prior to performing most GIS operations. A mix of visual assessments and chart creation provided a general understanding of the dataset's spatial arrangement, elevation range, and rainfall characteristics.

3.3.1 Spatial Arrangement

As shown in Figure 3, the spatial distribution of the rainfall stations generally follows a north-south line. There is an area in the northern region of the study area with fewer stations, with one station near the northernmost point. As shown in Figure 4, the north-south line of stations roughly tracks the north-south line of the state highway. While most of the stations are not located directly on the highway, they are situated on the limited network of roads within a few miles of it. To the east of the highway there are no rainfall stations, as there are no roads for access.

In the south, there is a cluster of 20 stations: 19 farmer stations and one NOAA NCDC station. The 19 farmer stations in this southern cluster came from a single source and share the same collection characteristics: same rainfall gauge, same person measuring rainfall collections, same timing of measurements and recording. In this dataset of various citizen collectors and methodologies, it seems unique to have a cluster of this size, all possibly with the same collection protocol.

3.3.2 Elevation Range

Elevations for the rainfall stations range from lowest, station 58 at 16 feet, to the highest, station 28 at 2,467 feet. There are 21 stations, almost half of the dataset, between the elevations 1,001 feet and 1,500 feet (Table 7). This presence of stations within the same elevation range

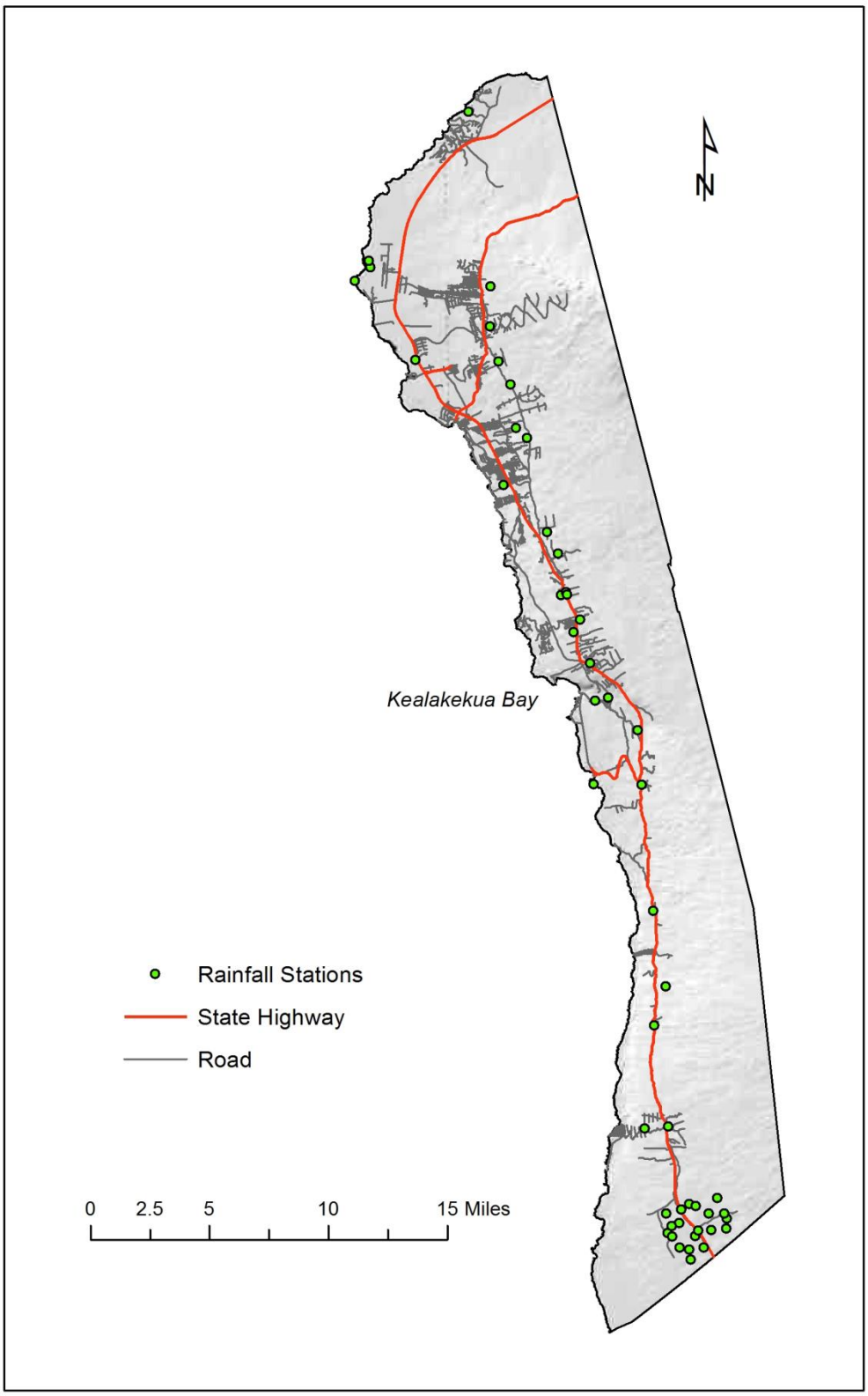


Figure 4: Distribution of rainfall stations relative to roads.

coincides with elevation contour data shown in Figure 2, where locations along a north-south line are within the same elevation range.

Table 7: Summary of rainfall stations elevation ranges.

Elevation (ft)	# of stations
0 - 500	8
501 - 1,000	3
1,001 - 1,500	21
1,501 - 2,000	12
2,001 - 2,500	8

3.3.3 Rainfall Characteristics

A review of the rainfall station averages found in Table 6 revealed that the lowest average rainfall at a station was 0.034 in., and the highest 7.458 in. A look at the seasonality lows and highs of rainfall can be seen in Figure 5, which shows an average of all 52 station averages from January through December. While the range for January – November is 1.42 in. – 2.84 in. with the highest averages occurring Spring through Summer, December had an unusually higher average than all other months at 3.85 in. A closer look at these averages broken out by year revealed the answer, as shown in Figure 6: December 2007 had a higher than average amount of rainfall, which has increased the annual average for December, as shown in Figure 5.

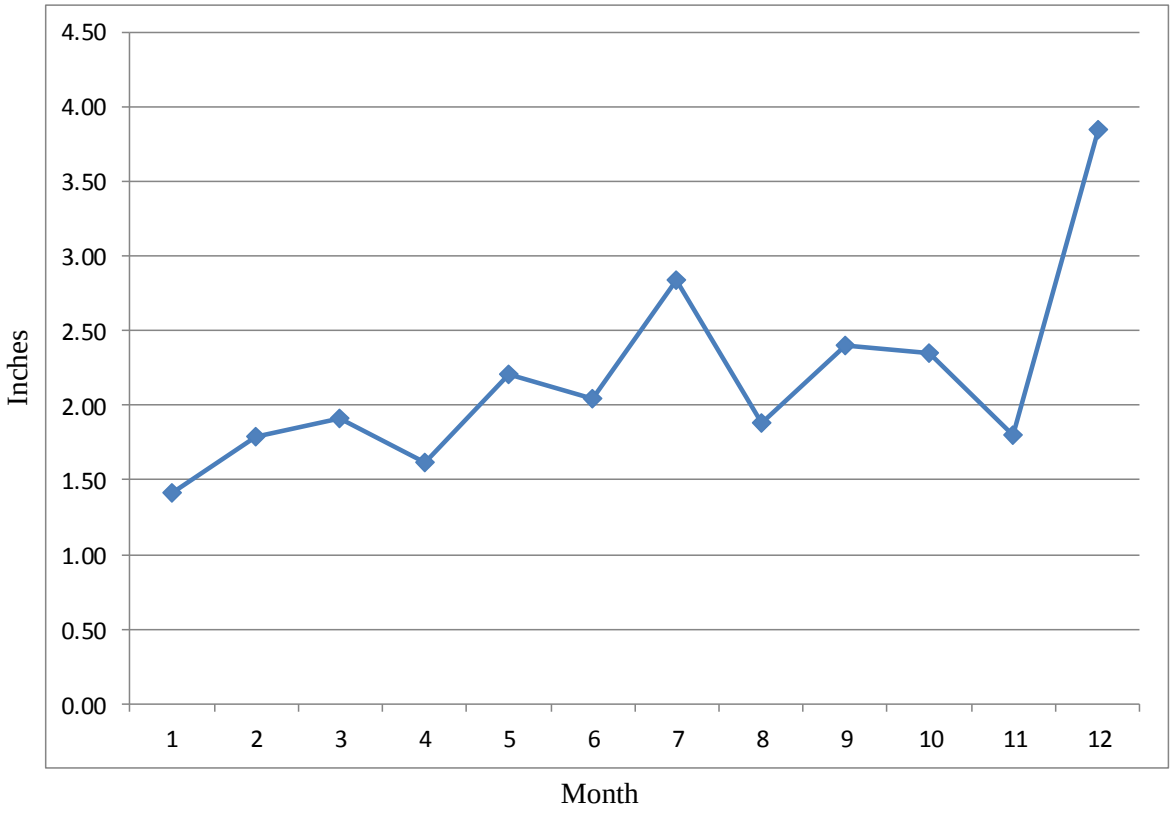


Figure 5: Monthly average of all stations 5-year averages. December shows a significantly higher rainfall average.

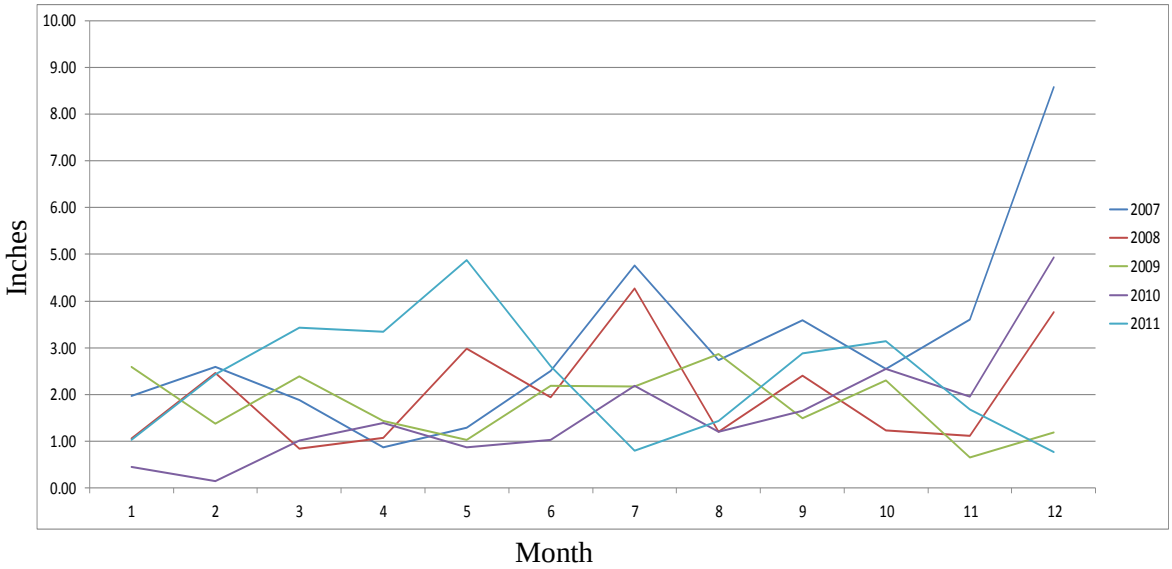


Figure 6: Monthly average of all stations individual annual averages. December 2007 had an anomalous amount of rainfall.

Figure 7 shows each station's 5-year monthly average for each month, January through December. The overlapping of data and difficulty in discerning individual months roughly demonstrates most stations falling within the same general range of values. It is worth pointing out that two stations frequently had higher totals than all others: stations 35 and 45.

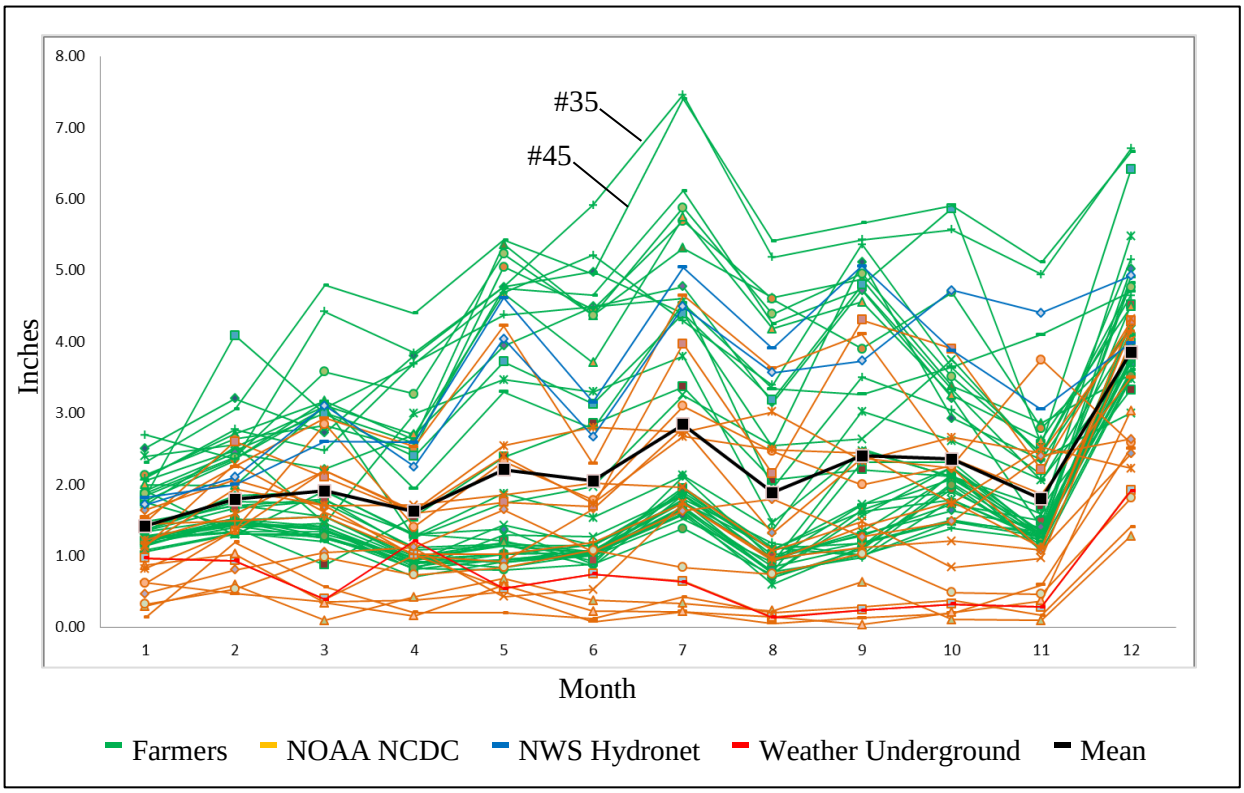


Figure 7: Individual station's 5-year monthly average for each month, January through December. Stations 35 and 45 have averages higher than all others.

3.4 GIS Data Preparation

To conduct the spatial interpolation necessary for this study, a GIS database was developed in Esri's ArcGIS 10.1. The rainfall data was aggregated into a usable tabular format with its corresponding latitude and longitude coordinates, and subsequently converted into point data.

The elevation data for the study area was obtained from a National Elevation Dataset (NED) digital raster file derived from a United States Geological Survey (USGS) Digital Elevation Model (DEM). The 10-meter resolution data was downloaded from the USGS National Map Viewer website. DEM accuracy is explained by Gesch (2007): NED datasets have been "tested by comparing it with an independent reference source of very high accuracy." Two types of accuracy are accessed: absolute vertical accuracy and relative vertical accuracy. The absolute vertical accuracy of the NED has a root mean square error (RMSE) of 2.44 meters. The relative vertical accuracy is a point-to-point assessment and is generally considered a more important measure of accuracy when calculating slope and aspect. The relative vertical accuracy for the NED is 1.64 meters. Some USGS DEM files will include accuracy statistics in the header records of the source files, but no such files were included in the DEM used in this study. Therefore, the assessment of the vertical accuracy of the DEM used in this study is the general assessment cited above.

Slope was calculated in ArcMap using the Slope tool. The tool determines the maximum change in elevation between neighboring cells by using a 3 x 3 cell neighborhood surrounding the center cell, thereby using the eight surrounding neighbors. An average maximum technique, using the rate of change in the x and y directions, calculates a slope value in degrees or percent (percent rise). Terrain that is flat with no elevation changes from cell to cell will have a slope of

zero. Increasing steepness of terrain will have correspondingly higher values of slope. The Slope tool produces a raster with similar dimensions from a DEM with slope as the cell value. The resulting slope raster (Figure 8) was output in degrees.

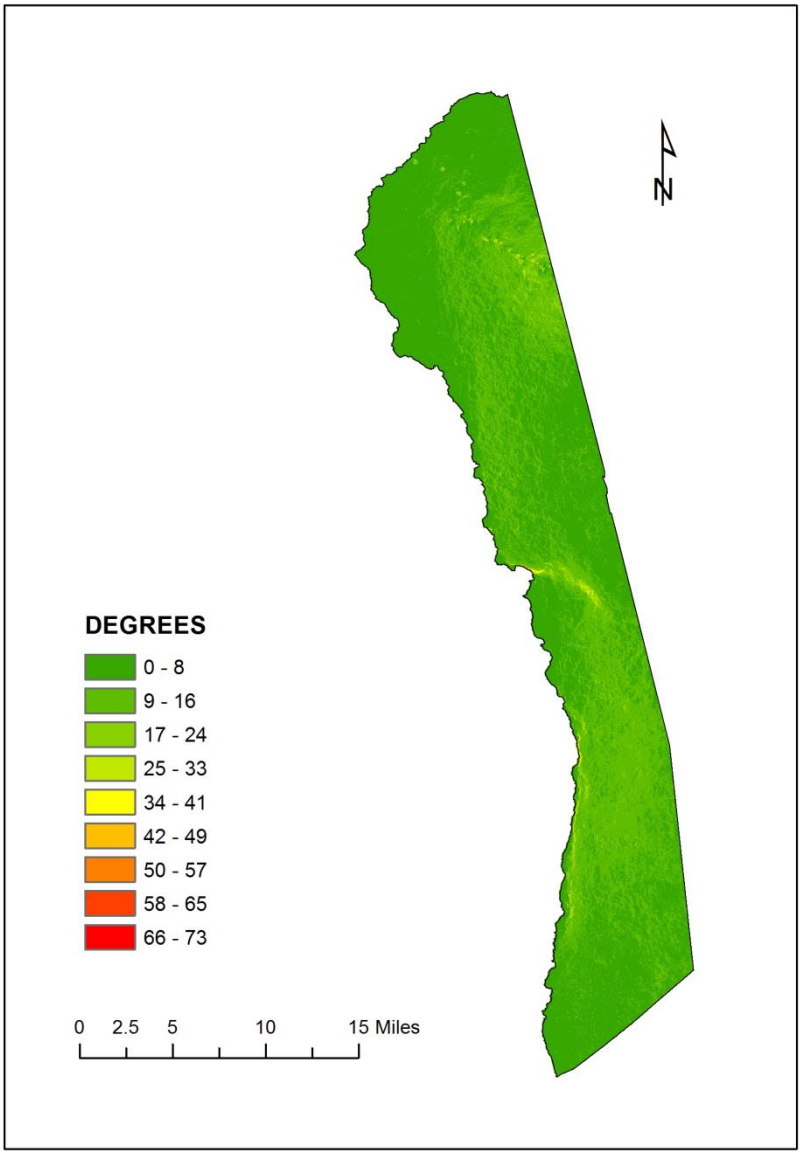


Figure 8: Slope raster generated for the study area using the DEM.

The aspect of a surface is defined by Esri (2012) as the "downslope direction of the maximum rate of change in value from each cell to its neighbors", or simply put, the slope

direction. The aspect tool in ArcMap uses a 3 x 3 moving window, incorporating the values of the eight surrounding neighbors of a cell. An algorithm calculates aspect, and creates a raster with values reflecting the directions on a compass, from 0 to 360. Flat areas with no slope are given a value of -1. An aspect raster was created for the study area using the DEM (Figure 9).

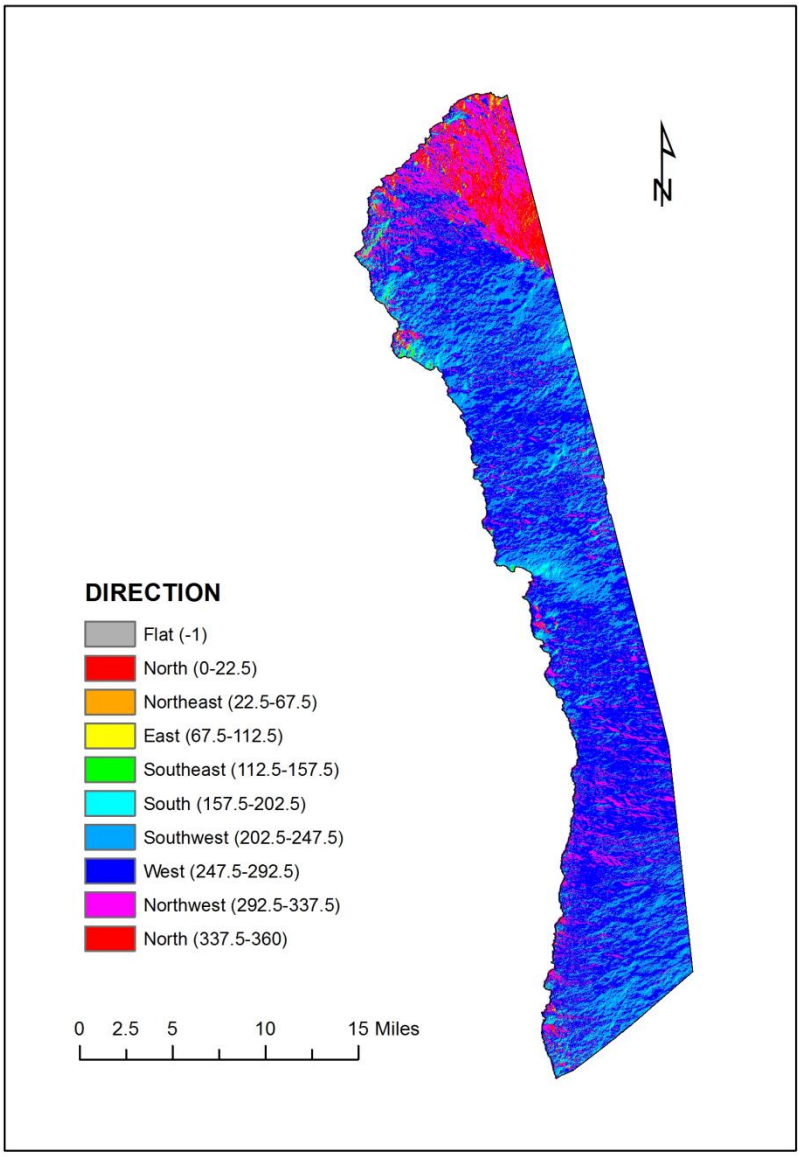


Figure 9: Aspect raster generated for study area using the DEM.

All files in this study conformed to the North American Datum 1983 (NAD 83) and were projected to the coordinate system UTM 5N.

3.5 Procedure

3.5.1 Exploring the Data: Slope and Aspect

Initial inspection of the slope output raster (Figure 8) revealed the lowest slope (flattest terrain) in the north and northwest coastal region of the study area, in addition to intermittent sections along the west coast and southwest. The highest slope values were seen extending inland from Kealakekua Bay, approximately on the mid-point of north and south on the west coast. A very small sliver of steep coastline with a slope in the range of 67-73 degrees is seen bordering the north of Kealakekua Bay. Other areas of high slope values were scattered throughout the north and south of the study area, as the terrain sloped upward in an easterly direction toward the summits of Hualālai in the north and Mauna Loa in the south.

To aid in the visual assessment of the aspect raster (Figure 9), color symbology is added in increments of 45, so that the traditional points on a compass are easily discernible (N, NE, E, SE, S, SW, W, and NW). In addition, the flat value (-1) is given a separate color to easily distinguish it. As would be expected from slopes on the west side of a volcanic island, the aspect raster shows a surface predominantly facing west, with much variation to the south and southwest. Several northwest facing areas were found in the northern and south central region. The northern tip of the study area faced predominantly north and northwest.

3.5.2 Exploring the Data: Semivariogram Clouds, Histograms, and Outliers

Using the Geostatistical Analyst functions in ArcMap, a Semivariogram Cloud and Histogram were applied to the individual months of rainfall point data. The Semivariogram Cloud helps evaluate the spatial dependence of a dataset. The Histogram is used to examine the distribution of a dataset and provides summary statistics. Together these tools can aid in uncovering any potential dataset outliers, which have extreme or erroneous values which do not belong in the dataset or which are correct but will influence the results in unexpected ways.

An example of the Semivariogram Cloud is shown in Figure 10, the corresponding result for July rainfall averages. The x-axis represents the distance between rainfall stations, and the y-axis the squared difference of corresponding rainfall averages. Each possible pair of the 52 July average rainfall points was plotted, resulting in 1,326 pairs of points. Local outliers appear as points that have low values on the x-axis, but are high on the y-axis; they are points that are close together in proximity, but not in their rainfall average value. As the distance increases between stations (moving right on the x-axis), it is expected that the difference in rainfall averages can also increase (up on the y-axis).

The result for July shows several points with low values on the x-axis and corresponding high values on the y-axis. In Figure 11, these points have been shaded as light blue dots and a circle drawn around them. A map of the study area to the right shows the location of these stations, again in light blue, as well as the neighboring stations with significantly different values. Blue lines are drawn to connect these stations, displaying the paired relationship. Station numbers are included to identify each station.

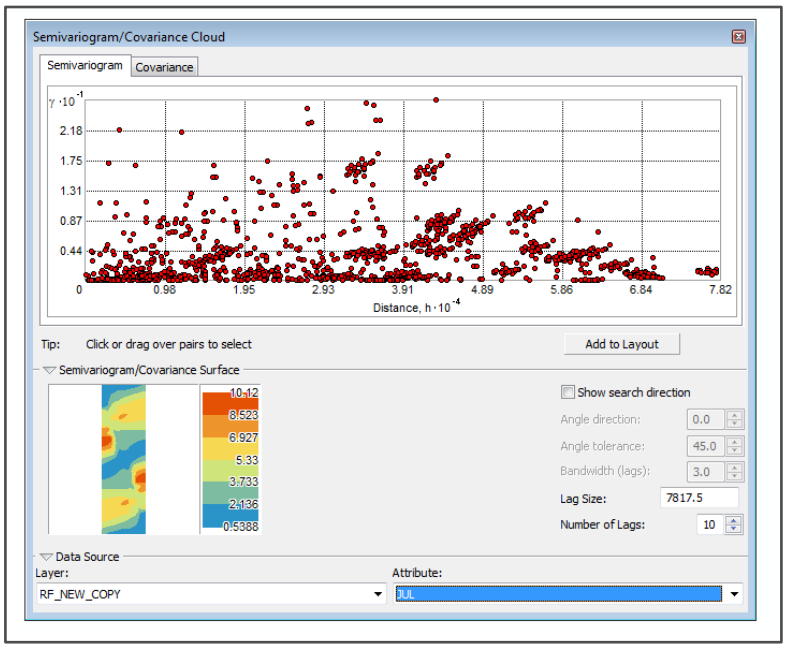


Figure 10: Semivariogram Cloud for July. The x-axis represents the distance between stations, and the y-axis the square of the difference of rainfall averages.

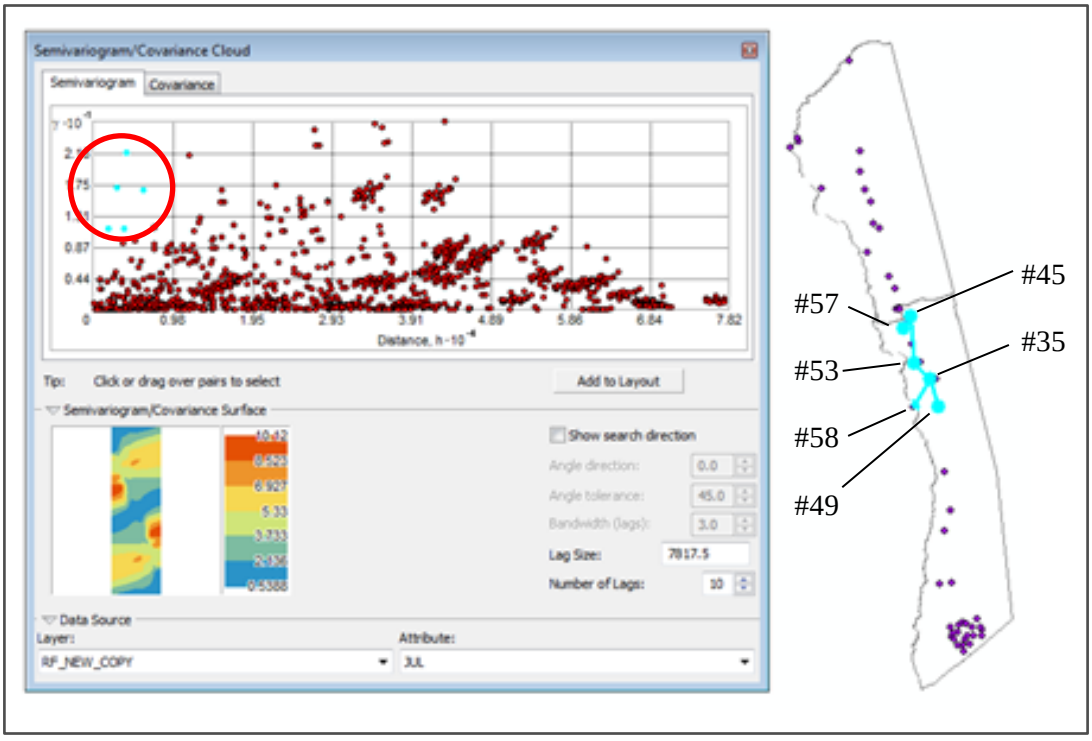


Figure 11: Semivariogram Cloud for July with potential outliers circled and corresponding locations identified on map at right.

The corresponding Histogram is shown in Figure 12, where the corresponding station values from the Semivariogram are again highlighted in light blue. The bars represent the range of rainfall averages in ten intervals along the x-axis, with the relative proportion (frequency) of those averages on the y-axis. The specific station each highlighted bar corresponds to is indicated. A station that is highlighted within a taller bar indicates there are a greater number of stations within that bar's value range. Stations 49, 53, 57, and 58 have rainfall averages for July that are in the same range as several other stations. Highlighted bars that are to the far right have the greatest value, and should be the lowest value. For July, stations 35 and 45 have higher values that don't correspond to the decreasing number of stations and values.

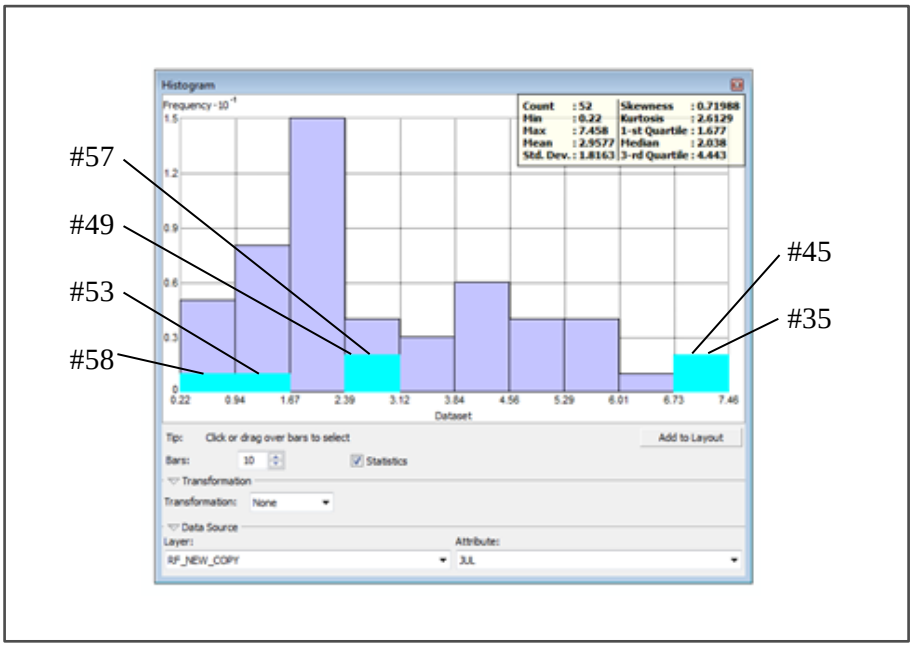


Figure 12: Histogram of July with corresponding stations highlighted and numbered.

In a review of the Semivariograms for all twelve months of monthly average rainfall data, it was noted that stations #35 and 45 tended to have higher monthly rainfall totals than all other stations in the period April-December. This is consistent with the results discussed in Section

3.3.3 and illustrated in Figure 7. As shown in Table 6, station #45 is one of the higher elevation stations at 1,972', which is a possible explanation of the higher rainfall values. Station #35 is at 778', which is a lower than average station elevation, and should have lower than average rainfall. January-March also revealed inconsistencies, but those could be explained by the higher elevation of the corresponding stations. Stations #35 and 45 were noted as potential outliers and held under consideration to be removed from the rainfall dataset. Month-by-month Semivariogram Clouds and Histograms are found in Appendix B.

3.5.3 Exploring the Data: Exploratory Regression

Exploratory Regression is a GIS tool located in the Spatial Statistics toolbox of ArcMap. Using Ordinary Least Squares (OLS) regression analysis, the tool evaluates all possible combinations of a given set of independent variables to determine the best combination to model a specified dependent variable. In addition, the output states the percentage of the dependent variable that can be attributed to that combination. The tool is not required for the use of, nor has any relationship with, the geostatistical method of co-kriging. It was used in this study as a means of determining the most relevant covariates for interpolating rainfall averages (Lynch 2003). As discussed earlier, elevation, slope, and aspect may have an effect on rainfall amounts; Exploratory Regression was used to help determine if any or any combination of those factors has any correlation with the rainfall data used in this study.

In order to run the Exploratory Regression tool, all variables must be in a single data file. The master rainfall data file included latitude and longitude point locations of each station record station, as well as rainfall averages for each month of the year. Elevation information was previously extracted from the DEM for each station location, and was included in the main rainfall data file.

Slope and aspect data were in the format of separate raster files that were created from a DEM (as previously discussed). The Extract Multi Values to Points tool in the ArcMap Spatial Analyst toolkit extracts the cell values from the raster files at the point location from the input file. For example, station #1corresponds with cells of slope=3.0 and aspect=270. The tool adds fields to the original input file for a slope value of 3.0 and aspect value of 270. Using the Extract Multi Values to Points tool, a single file was created that contained all variables.

Exploratory Regression was run twelve times, using monthly rainfall averages as the dependent variable. The independent variables used were elevation, slope, aspect, latitude, and longitude. A report was generated for each month, each detailing the best results as indicated by the highest adjust r-squared result for all possible combinations of the independent variables. In addition, the overall goal of a properly specified OLS model was examined.

The results in Table 8 display the best combination achieved for each individual month. Only two months were able to achieve properly specified OLS models: March and April. This was accomplished with the combination of aspect, latitude, and longitude. The corresponding adjusted r-squared result was 0.57 for March and 0.58 for April, indicating that 57% of March's rainfall and 58% of April's could be attributed to aspect, latitude, and longitude.

An additional series of Exploratory Regression calculations were conducted for all twelve months with the two potential outlier rainfall stations, previously discussed in the Semivariogram and Histogram exploration section, removed from the dataset. No other changes were made to the dataset. As shown in Table 9, removal of the stations slightly improved the overall results, with an additional month achieving a properly specified OLS model. December's passing model accounted for 53% of rainfall with the combination of elevation and slope. April's result improved to 66% using the combination of latitude, elevation, and slope. March's result slipped

Table 8: Exploratory Regression results using elevation, slope, aspect, latitude, and longitude as independent variables. March and April were the only two passing models.

Month	Passing model? (properly specified)	Highest % (adjusted R^2)	Best Combination (adjusted R^2)
Jan	No	50%	Elevation + Slope + Aspect + Latitude
Feb	No	48%	Elevation + Slope + Aspect + Latitude
Mar	Yes	57%	Aspect + Latitude + Longitude
Apr	Yes	58%	Aspect + Latitude + Longitude
May	No	70%	Elevation + Aspect + Latitude + Longitude
Jun	No	61%	Elevation + Slope + Aspect + Latitude + Longitude
Jul	No	56%	Elevation + Slope + Aspect + Latitude + Longitude
Aug	No	59%	Aspect + Latitude + Longitude
Sep	No	51%	Elevation + Aspect + Latitude + Longitude
Oct	No	41%	Slope + Aspect + Latitude + Longitude
Nov	No	51%	Elevation + Aspect + Latitude + Longitude
Dec	No	44%	Elevation + Slope + Latitude + Longitude

a single point with a 56% result using the same previous combination of latitude, longitude, and aspect. The remaining nine months had resulting combinations with adequate adjusted R^2 results, but were unable to achieve properly specified models. As a consequence of the improved results by removing the potential outlier stations 35 and 45, these stations were removed from the dataset for the remainder of the study.

Table 9: Exploratory Regression with two potential outlier rainfall stations removed. March, April, and December were passing models.

Month	Passing model? (properly specified)	Highest % (adjusted R^2)	Best Combination (adjusted R^2)
Jan	No	49%	Elevation + Slope + Aspect + Latitude
Feb	No	47%	Elevation + Slope + Aspect + Latitude
Mar	Yes	56%	Aspect + Latitude + Longitude
Apr	Yes	66%	Elevation + Slope + Latitude
May	No	69%	Elevation + Slope + Aspect + Latitude + Longitude
Jun	No	64%	Elevation + Slope + Aspect + Latitude + Longitude
Jul	No	59%	Elevation + Slope + Aspect + Latitude + Longitude
Aug	No	57%	Aspect + Latitude + Longitude
Sep	No	48%	Elevation + Aspect + Latitude + Longitude
Oct	No	43%	Slope + Aspect + Latitude + Longitude
Nov	No	51%	Slope + Aspect + Latitude + Longitude
Dec	Yes	53%	Elevation + Slope

3.5.4 Kriging Comparison: Methods Used

With the goal of achieving the most accurate rainfall surfaces for the data used, three methods of kriging were performed side-by-side: ordinary kriging, ordinary co-kriging, and Empirical Bayesian Kriging (EBK). As discussed earlier, ordinary co-kriging and EBK are considered the superior methods of kriging when interpolating rainfall point data to create a rainfall surface map. Ordinary kriging was included in this study as a comparison to validate this claim.

Ordinary kriging and EBK solely use the rainfall point data to interpolate rainfall surfaces; no other data or variables are used. Ordinary co-kriging also uses the rainfall point data, plus independent variables (covariates) in its interpolation. The covariates applied were elevation, slope, and aspect. Latitude and longitude were not added in the co-kriging as these are implicit in the kriging process. To ensure the kriging operations account for similarity among data values, the rate values change with distance, and whether or not those rates change with changes in direction, a model is fit to the data points forming the semivariogram. This process can remove broad regional trends and improve value estimates by focusing on local variation (Scott and Getis 2008). During the kriging and co-kriging operations, the semivariogram was fit to the Gaussian model, as the continuous curve that was produced started relatively flat before increasing and subsequently leveling off flat.

Ordinary co-kriging was run three additional times for the months of March, April, and December using the covariates determined through properly specified OLS models in Exploratory Regression. Table 10 summarizes all kriging procedures that were conducted.

Table 10: Summary of kriging methods used and corresponding ordinary co-kriging covariates.

Month	Method	Co-kriging Covariates
January	Ordinary kriging	
	Ordinary co-kriging	- w/ elevation, slope, aspect
	Empirical Bayesian Kriging	
February	Ordinary kriging	
	Ordinary co-kriging	- w/ elevation, slope, aspect
	Empirical Bayesian Kriging	
March	Ordinary kriging	
	Ordinary co-kriging	- w/ aspect
	Ordinary co-kriging	- w/ elevation, slope, aspect
	Empirical Bayesian Kriging	
April	Ordinary kriging	
	Ordinary co-kriging	- w/ elevation, slope
	Ordinary co-kriging	- w/ elevation, slope, aspect
	Empirical Bayesian Kriging	
May	Ordinary kriging	
	Ordinary co-kriging	- w/ elevation, slope, aspect
	Empirical Bayesian Kriging	
June	Ordinary kriging	
	Ordinary co-kriging	- w/ elevation, slope, aspect
	Empirical Bayesian Kriging	
July	Ordinary kriging	
	Ordinary co-kriging	- w/ elevation, slope, aspect
	Empirical Bayesian Kriging	
August	Ordinary kriging	
	Ordinary co-kriging	- w/ elevation, slope, aspect
	Empirical Bayesian Kriging	
September	Ordinary kriging	
	Ordinary co-kriging	- w/ elevation, slope, aspect
	Empirical Bayesian Kriging	
October	Ordinary kriging	
	Ordinary co-kriging	- w/ elevation, slope, aspect
	Empirical Bayesian Kriging	
November	Ordinary kriging	
	Ordinary co-kriging	- w/ elevation, slope, aspect
	Empirical Bayesian Kriging	
December	Ordinary kriging	
	Ordinary co-kriging	- w/ elevation, slope
	Ordinary co-kriging	- w/ elevation, slope, aspect
	Empirical Bayesian Kriging	

3.5.5 *Surfaces Generated and Prediction Error Results*

Thirty nine rainfall surfaces were produced and then symbolized in an identical manner: on an equal interval scale from 0.0 in to 8.0 in, with increments of 0.5 in. Symbolizing all months with the same scale allows for side-by-side visual comparison of the generated surfaces. Note that, in ArcGIS, kriging results are retained and symbolized for visual analysis in the native format of the interpolation procedure, not output directly as rasters.

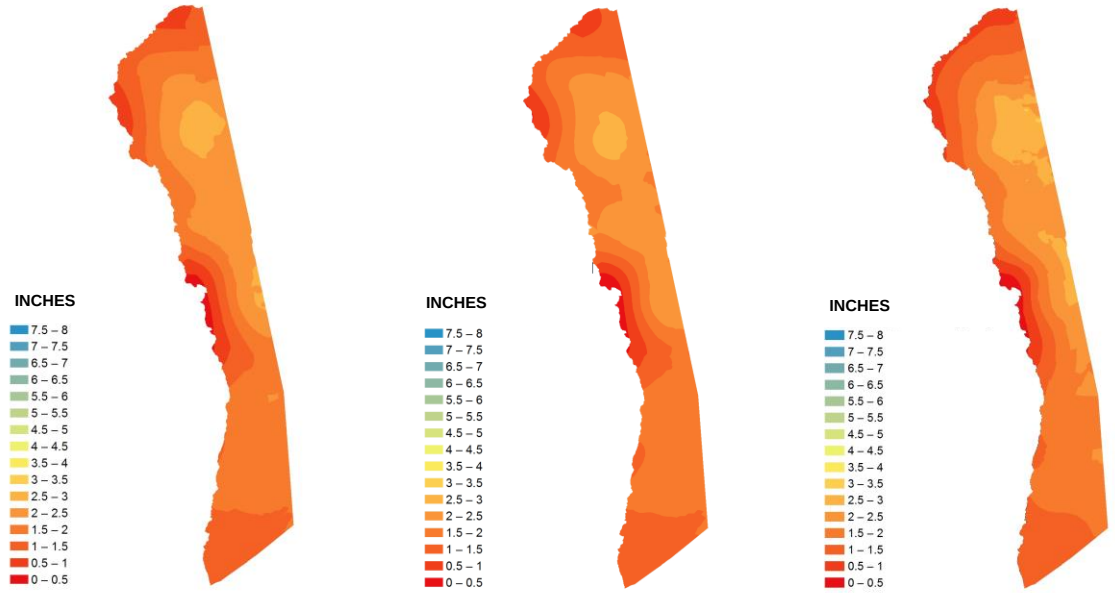
While visual comparison of the generated surfaces may prove satisfying, this procedure is subjective, and will not accurately determine the best method utilized. Conversely, generating prediction error statistics provides an objective method for comparing and evaluating surfaces, and determining the best model. In the final step of all kriging methods, a cross-validation dialogue box appears. This process sequentially omits a point in the dataset, uses the remaining data to predict a value for that point, and then compares the two values. The difference between the measured value and the predicted value is the prediction error, and statistics based on the aggregation of these errors provide diagnostics on the generated model (Esri 2012).

For all surfaces generated, prediction error statistics were generated and recorded for the following:

Prediction Error Statistic	Best model
Standardized mean	nearest to 0
Root-mean-squared (RMS) prediction error	smallest
Average standard error	nearest to RMS prediction error
Standardized root-mean-squared prediction error	nearest to 1

Figure 13 shows the surfaces generated for January for each kriging method and their corresponding prediction error statistics. The criteria for determining the best model is in the left column. Highlighting has been applied to indicate the model that adheres to the criteria closest, and an arrow below pointing to the model which adheres to the most categories of those criteria. As shown for January, the EBK method has the best prediction error statistics for three out of four of the categories. Therefore, EBK is the best method used in this study for mapping rainfall measurements for the January dataset. Figures similar to Figure 13 showing generated surfaces for all months and their corresponding prediction error statistics used for evaluation are shown in Appendix C.

January



	Objective	Ordinary Kriging	Ordinary Co-kriging (w/ Elevation, Slope, Aspect)	Empirical Bayesian Kriging
Standardized Mean	nearest 0	0.0790	0.0233	0.0136
Root-Mean-Squared Prediction Error	lowest	0.5062	0.3176	0.3228
Average Standard Error	nearest R-M-S Pred Error	0.2536	0.2497	0.3432
Standardized Root-Mean-Squared Prediction Error	nearest 1	1.4929	1.3896	0.9976

Figure 13: Rainfall surfaces generated for January and the corresponding prediction error statistics. The EBK model most closely adheres to the prediction error criteria, indicating it is the best model of these methods

Chapter 4 - Results

The best method of kriging for each month of this dataset is summarized in Table 11. EBK was the best method for nine months of the data, and tied (two results with best results) for two months. Ordinary kriging was the method for one month, and tied for one month. Ordinary co-kriging, with all covariates (elevation, slope, aspect) was tied for best method only one month. Ordinary co-kriging using covariates selected through Exploratory Regression never provided the best results.

Table 11: Summary of the best method of kriging for each month, as determined by prediction error results.

Month	Best Method of Kriging
January	Empirical Bayesian Kriging
February	Empirical Bayesian Kriging
March	Empirical Bayesian Kriging
April	Empirical Bayesian Kriging
May	Empirical Bayesian Kriging
June	Empirical Bayesian Kriging; Ordinary Co-kriging (tie)
July	Empirical Bayesian Kriging
August	Empirical Bayesian Kriging; Ordinary Kriging (tie)
September	Ordinary Kriging
October	Empirical Bayesian Kriging
November	Empirical Bayesian Kriging
December	Empirical Bayesian Kriging

The resulting rainfall map for each month is reviewed in the next section, discussing the rainfall surface pattern created using the most applicable method of kriging, as summarized above. Prediction standard error maps, a method of assessing the level of uncertainty in the results, are examined in the following section titled Evaluating Results.

4.1 Overall Rainfall Patterns

A review of the surfaces generated for all months revealed several standard patterns that are fairly consistent from month to month. Generally, the lowest amounts of rainfall were seen along the central-western, northern, and southwest coasts. These driest regions seldom received a monthly average greater than 1 inch of rain. Quite often, less than 0.5 in. was measured.

The wettest region was found extending through the center of the study area in a north-south pattern. As the distance from the west coast increased, so did the average amount of monthly rainfall totals. Elevation also increases as the distance from the coast increases, as previously shown in Figure 2. This increase in rainfall can be explained as a result of orographic lifting. The presence of the volcano Hualālai in the northern region of the study area reflects a similar pattern. As the elevation of its western slope increases, an increase in measured rainfall is generally found.

The cluster of twenty stations in the most southern region discussed in Section 3.3.1 presents an interesting sub-study within their small region. While some months had consistent rainfall amounts across all stations, other months had more variety. These resulting differences are explained in more detail in the individual monthly summaries found below. Additionally, the cluster is analyzed independently later in this chapter.

The station with the highest elevation discussed previously, station 28 (Figure 14), produced results that standout compared to others within the region. In nearly every month, this station recorded higher rainfall amounts than its direct neighbors to the north and south. As a result, a higher rainfall pattern is produced for most months in this region.

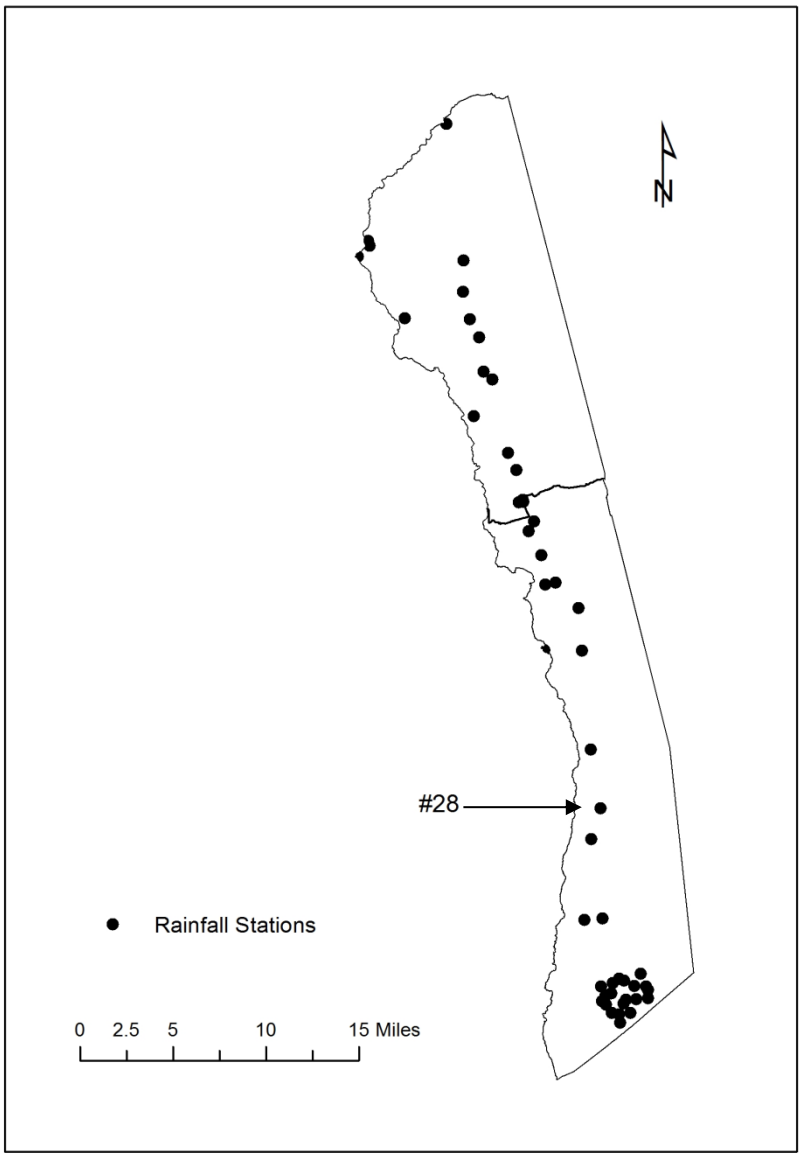


Figure 14: Location of station 28, which generated higher rainfall patterns compared to its neighbors.

4.1.1 January

The rainfall pattern generated for January (Figure 15) with EBK followed the general pattern outlined above. While this month had the least amount of rainfall, increases could still be found through the center north-south, as well as the western slope of Hualālai. The southern cluster of stations appeared to have the same values, and therefore there was no color variation within them.

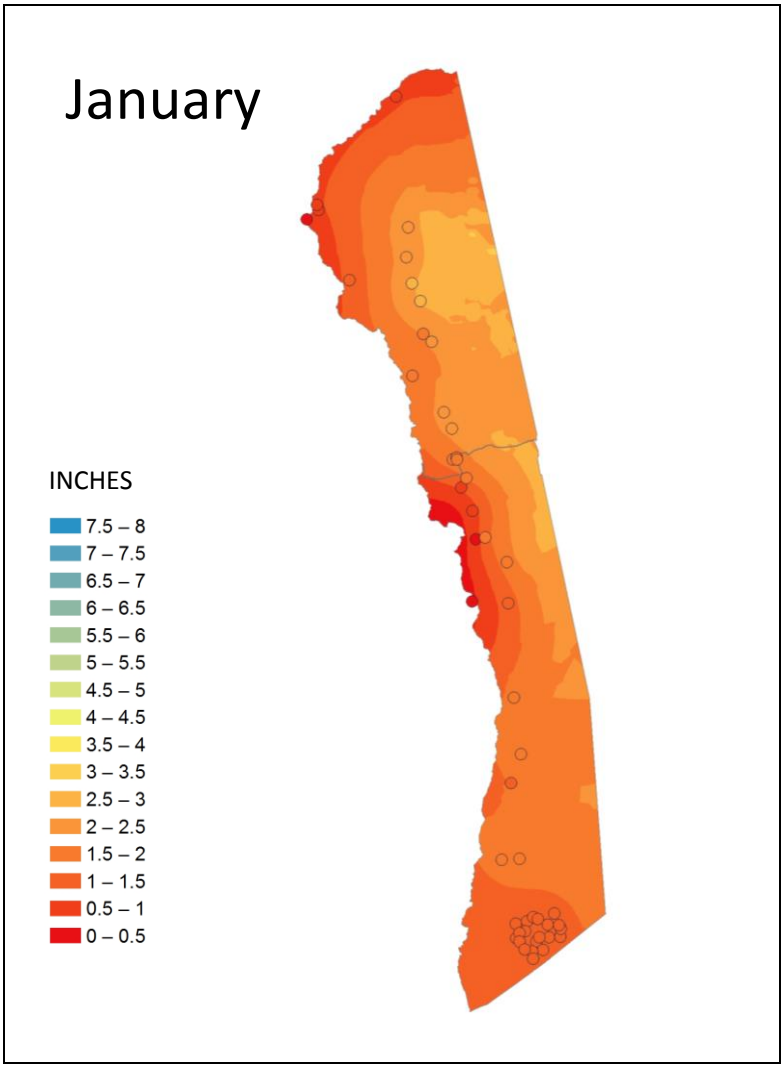


Figure 15: January rainfall surface generated with Empirical Bayesian Kriging.

4.1.2 February

The February result (Figure 16) using EBK showed a slight increase in rainfall over January, but was still drier than most months. The north-south pattern of higher rainfall was present, but showed more differentiation due to the variation in corresponding rainfall amounts found in the stations directly to the west. The higher amount of rainfall measured at station 28 resulted in a color pattern of its own, with the highest amount of rainfall throughout the study area. The southern cluster had slightly mixed totals, resulting in variation of the spatial patterns.

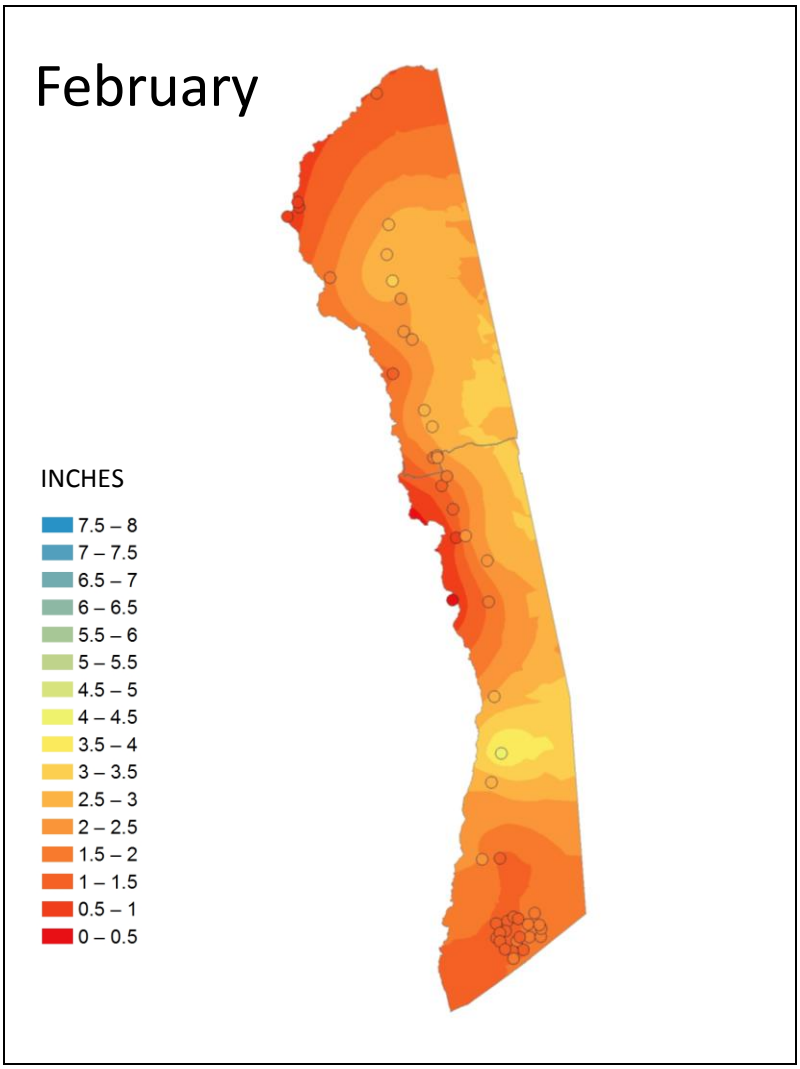


Figure 16: February rainfall surface generated with Empirical Bayesian Kriging.

4.1.3 March

The March result (Figure 17) generated using EBK showed an increase in the center of the study area, while the north, west, and south regions indicated no increase whatsoever. The southern cluster had values split into two measurement increments, and results in two separate ranges splitting the cluster.

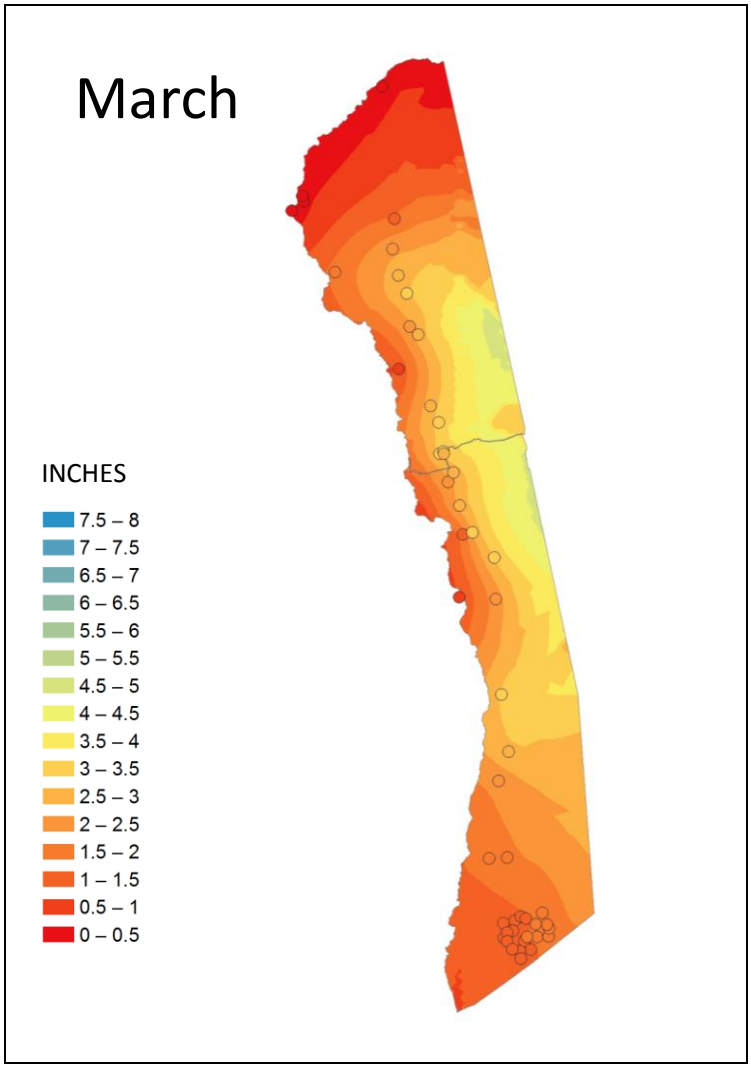


Figure 17: March rainfall surface generated with Empirical Bayesian Kriging.

4.1.4 April

The April result (Figure 18) using EBK revealed the highest amount of rainfall on the rising slopes of Hualālai. In contrast, the southern tip of the study area was at its driest, with more than half of the southern cluster stations measuring less than 0.5 inches of rainfall. Station 28 again had a higher amount of rainfall compared to its direct neighbors, causing a higher color pattern to bend through the region.

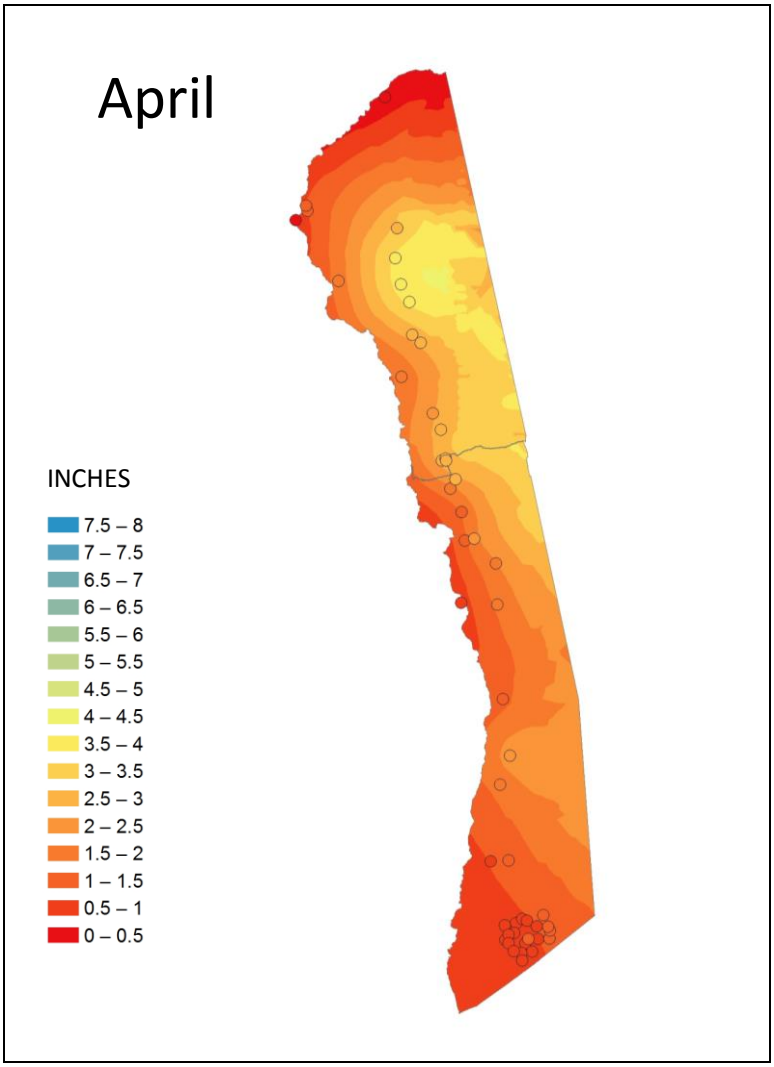


Figure 18: April rainfall surface generated with Empirical Bayesian Kriging.

4.1.5 May

The May surface (Figure 19) generated with EBK showed some of the highest rainfall found throughout the year in the central section, while still maintaining the lowest levels in the north, west, and south. As a result, the rainfall color ranges changed from level to level in a short distance. Station 28 again showed a higher level than its neighbors, influencing the patterns found in the region. The southern cluster was again divided in values, and resulted in divided ranges.

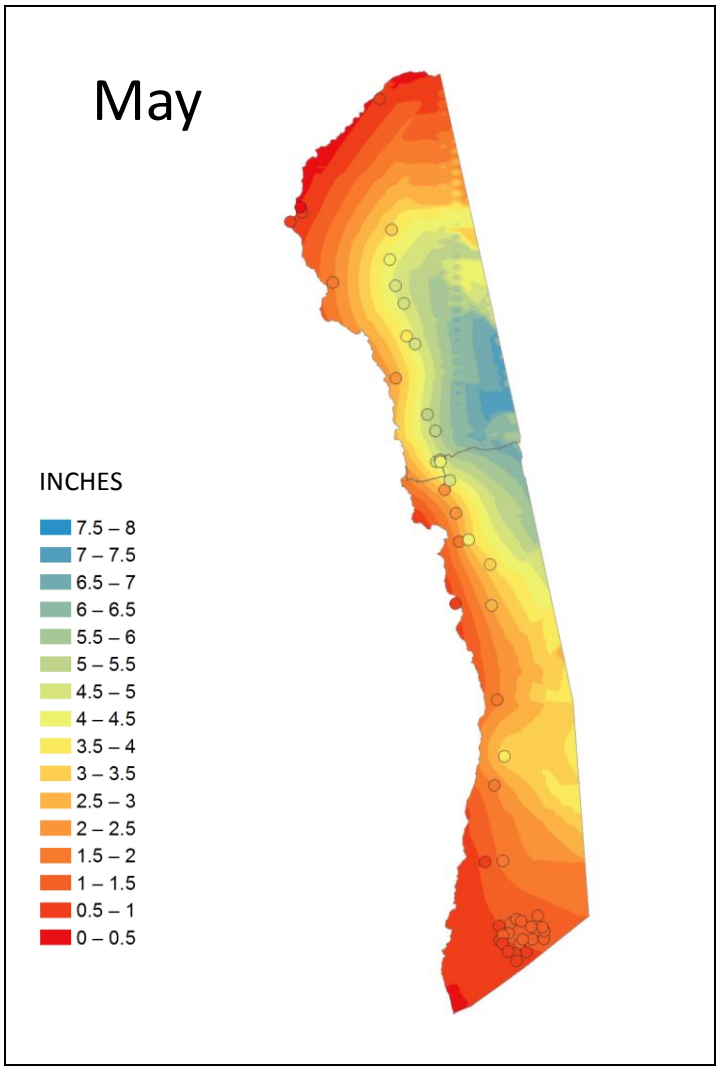


Figure 19: May rainfall surface generated with Empirical Bayesian Kriging.

4.1.6 June

Evaluation of the prediction error statistics for the June surfaces did not present a clear best option, but instead there are two surfaces to consider. While both surfaces exhibited the same general patterns as outlined in the beginning of this section, the two surfaces did not appear identical. The result for ordinary co-kriging, as seen in the left of Figure 20, was a more generalized result with less variation in its rainfall color ranges. While a higher amount of rainfall was seen in the north-central region, little variation of color ranges were present from the center moving south to the southern cluster. Station 28 had a value higher than its surrounding neighbors, but there was no resulting change in the underlying interpolated surface.

By contrast, the result for EBK (right side of Figure 20) displayed a larger area of the highest rainfall in the north, with higher values extending from north to south. In addition, the regions with the least amount of rainfall in the north, west, and south were larger than those found in the OCK method. Station 28 again created its own pattern, differing from its neighbors. The variation of the southern cluster created a pattern of color that wrapped around the southeastern stations, representing the variation in values of those stations.

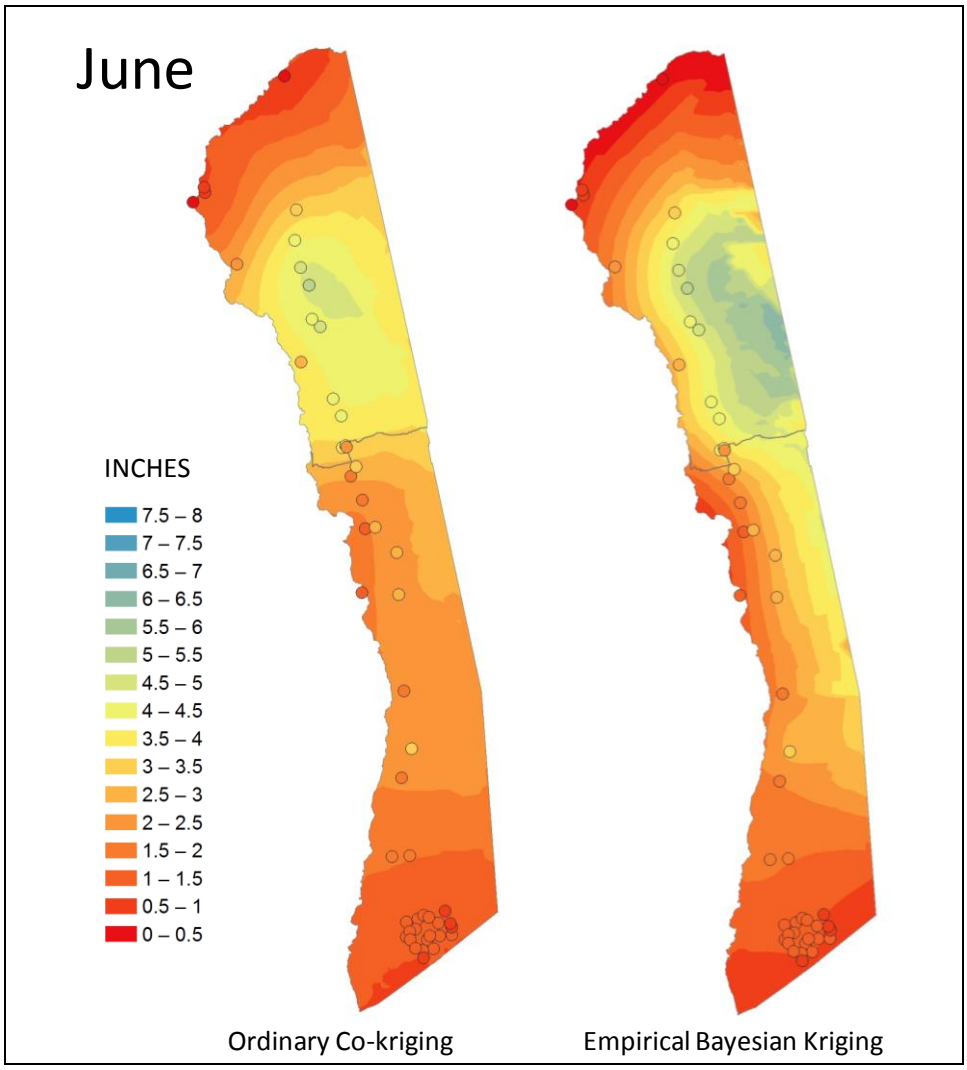


Figure 20: June rainfall surfaces generated with ordinary co-kriging on the left and Empirical Bayesian Kriging on the right.

4.1.7 July

The July surface (Figure 21) generated with EBK looked very similar to May, with values spanning the scale from lowest to highest in a relatively short space. The north-south pattern of high rainfall showed some of the highest rainfall amounts of the year, while still maintaining very little rain in the north and west. The southern cluster predominantly fell into one color, with eighteen of its values within the same measurement range.

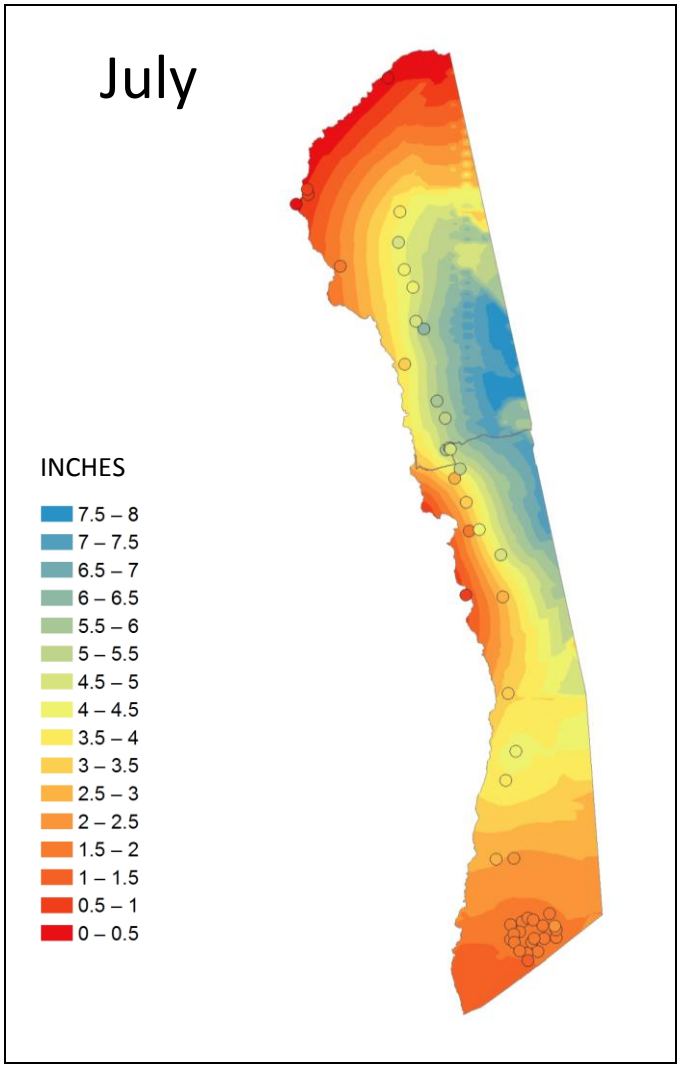


Figure 21: July rainfall surface generated with Empirical Bayesian Kriging.

4.1.8 August

August was another month with no clear best option deduced from the prediction error statistics evaluation. The two results, ordinary kriging and EBK (Figure 22), had very similar resulting surfaces. Both followed the general patterns described above. Most of the differences between the two surfaces were found on the eastern side of the region, where no measurement stations were located.

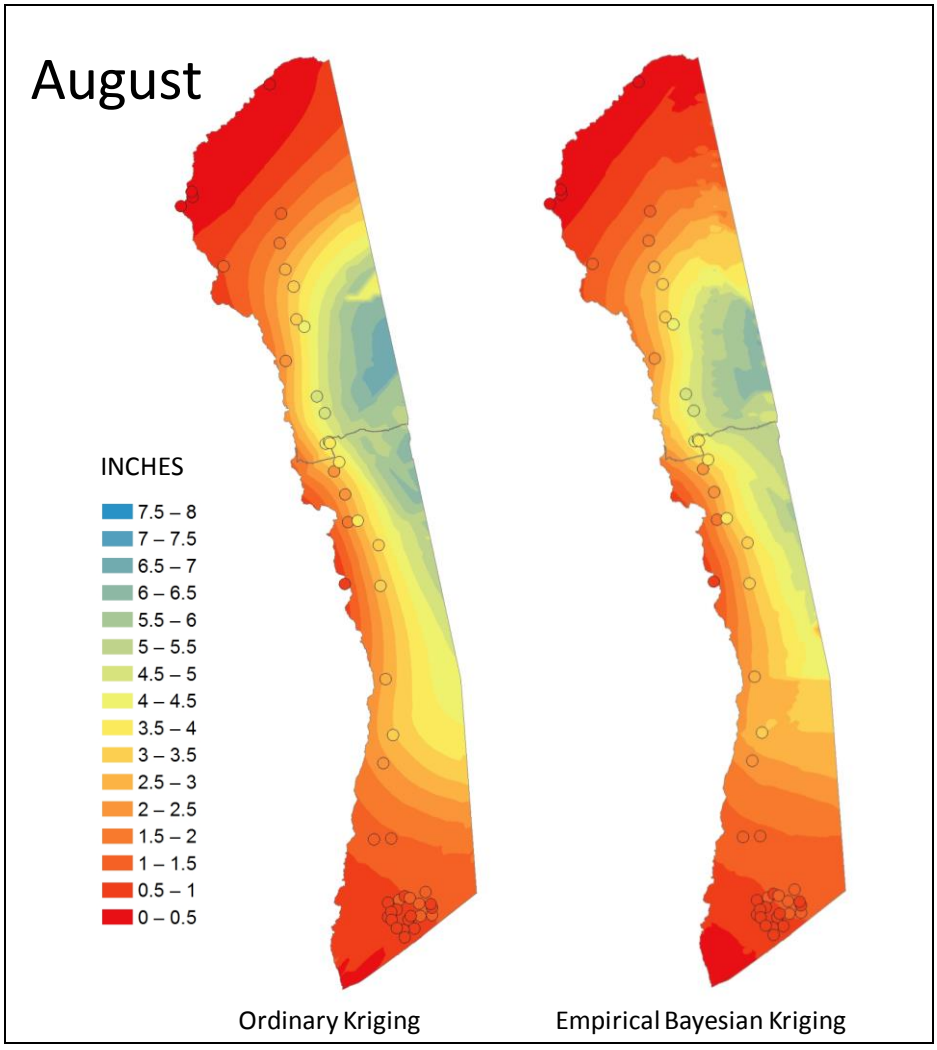


Figure 22: August rainfall surfaces generated with ordinary kriging on the left and Empirical Bayesian Kriging on the right.

4.1.9 September

The September surface (Figure 23) generated with ordinary kriging displayed the overall pattern seen in all preceding months, but with a few significant differences. Station 28 had a value that was one of the highest in the study area, which produced a pattern of the same measurement level in the southern region. This grazed station 28, but the station sat outside of it in a lower rainfall range. The southern cluster had values in four different classifications, causing significant variations in the colors found in the southern tip of the study area.

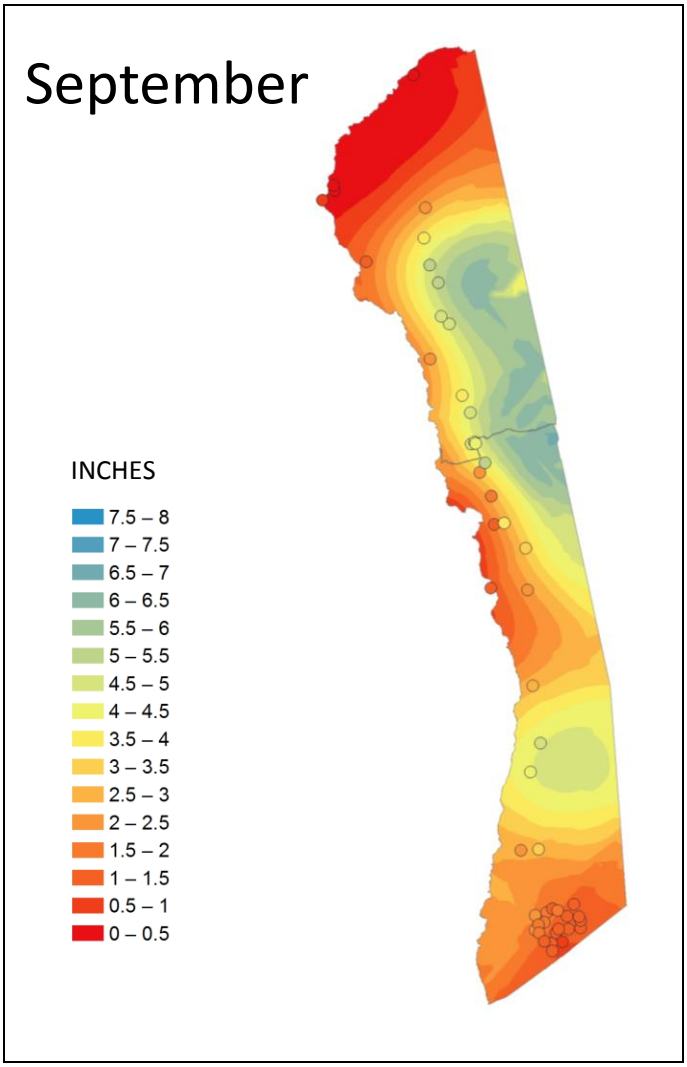


Figure 23: September rainfall surface generated with ordinary kriging.

4.1.10 October

The October surface (Figure 24) generated using EBK displayed a moderate amount of rainfall through the center of the region, but very dry in the north. The southern cluster was somewhat reversed from most other months, with the lowest values found in the eastern stations and higher in the western. The resulting spatial patterns appeared directionally reversed from most other months, although remaining on the lower end of the rainfall measurement range.

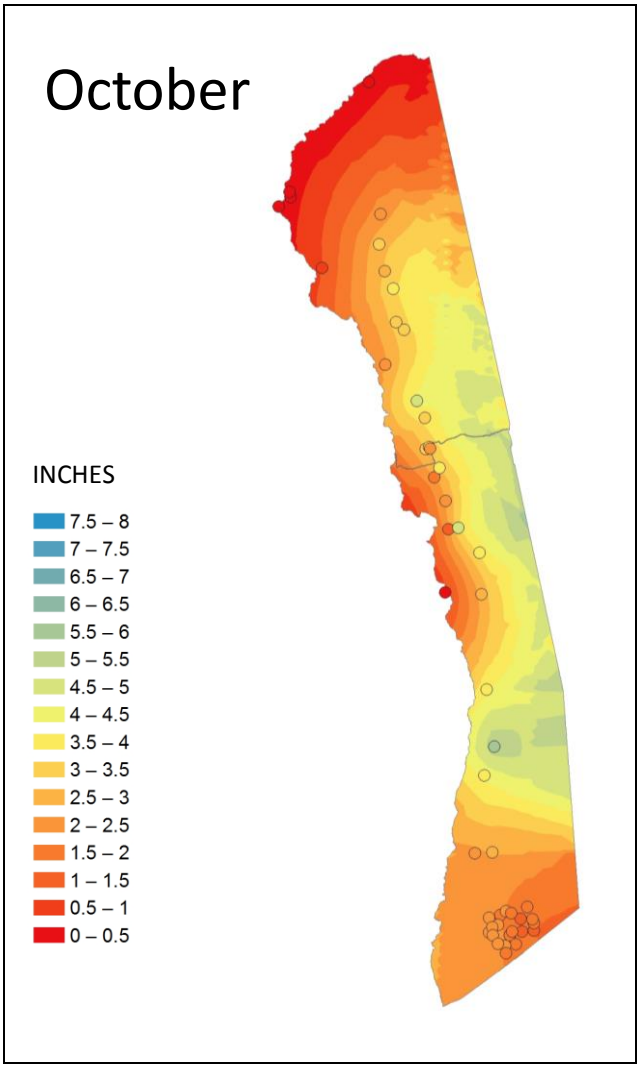


Figure 24: October rainfall surface generated with Empirical Bayesian Kriging.

4.1.11 November

November was a relatively drier month, with the surface generated with EBK (Figure 25) having the appearance of the drier months earlier in the year, especially March. A slight increase was found on the slopes of Hualālai, and a larger increase through the center of the study area. All stations in the southern cluster had the same value, and were found in a single value range.

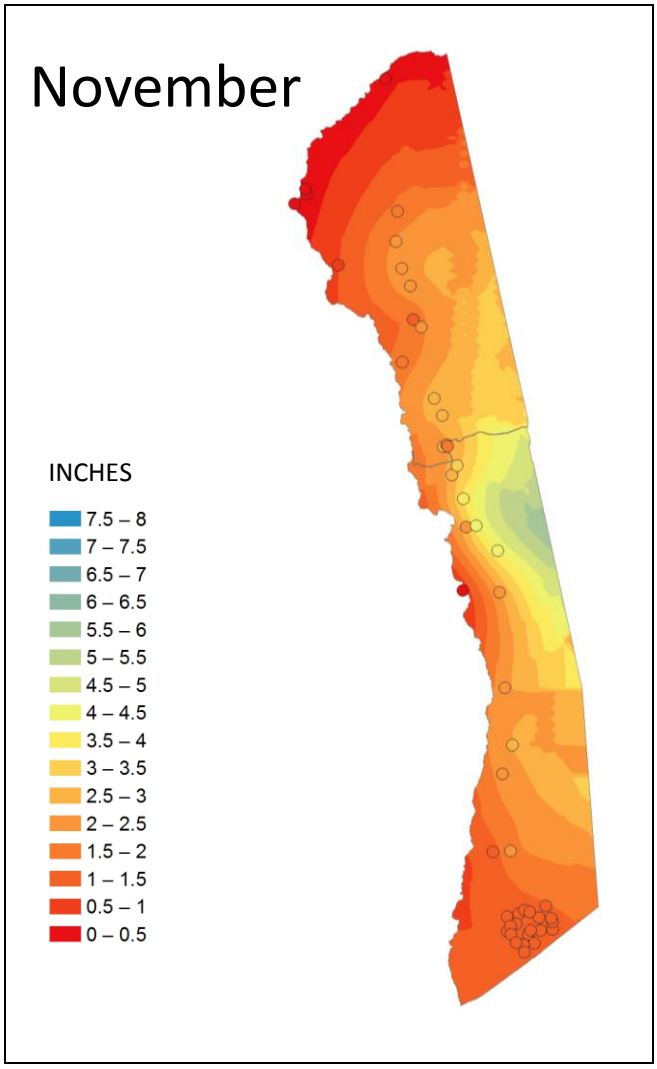


Figure 25: November rainfall surface generated with Empirical Bayesian Kriging.

4.1.12 December

The December surface (Figure 26) generated with EBK was the most dramatic of all months, as it had the highest amount of rainfall throughout the study area compared to the other eleven months. Peak rainfall values were found stretching from Hualālai in the north, down to the southern border. Dry areas were relatively sparse, and were still found in the same areas, but had higher values than previously seen. Station 28 had one of the highest values on the map, and resulted in high value color ranges correspondingly created. The southern cluster's values

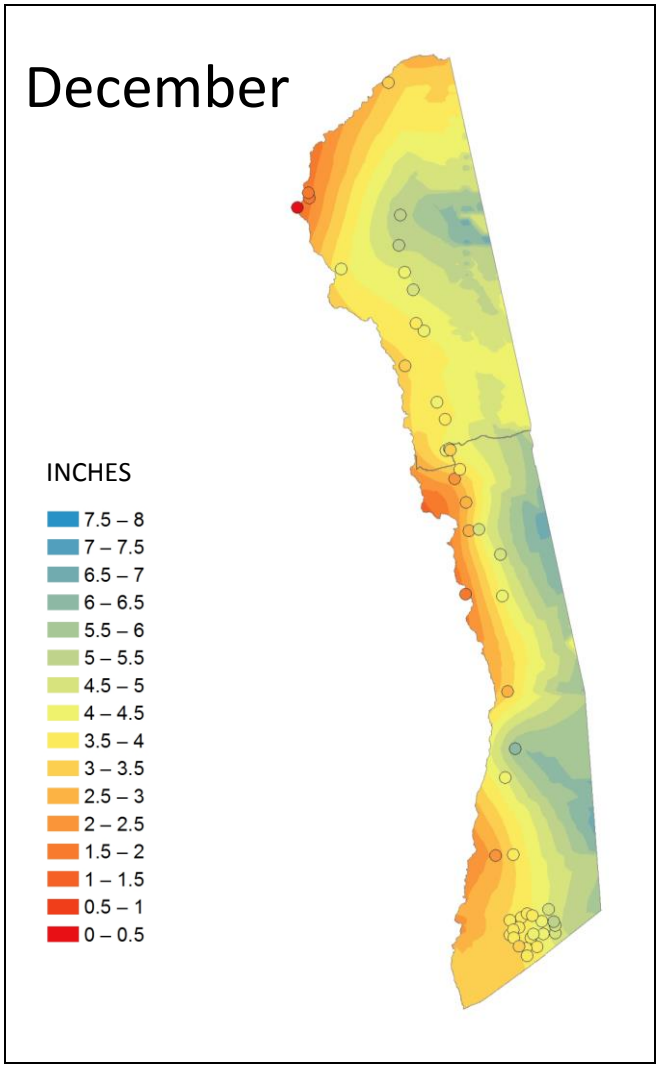


Figure 26: December rainfall surface generated with Empirical Bayesian Kriging.

increased from west to east, and rainfall ranges representing those increasing values were appropriately found in that region.

The surface created for December would seem to indicate that December is the wettest month of the year, with rainfall amounts far surpassing all other months of the year. As discussed in Section 3.3.3 and illustrated in Figures 5 and 6, December 2007 was an exceptionally wet month. Review of the individual, non-averaged December rainfall totals for each year revealed that 2007 had values that were 50% to 500% higher than the four year average of December 2008-2011. Figure 27 shows December 2008-2011, with 2007 removed. The result looks more similar to the other 11 months of the year. As discussed in Section 3.2, a 30-year average is preferred for rainfall-average studies to average out anomalous measurements in months like these. Since, for this study area, only a handful of records were available with 30 or more year's records, only five years were used to increase the amount of stations, with the intention of achieving better coverage of the region.

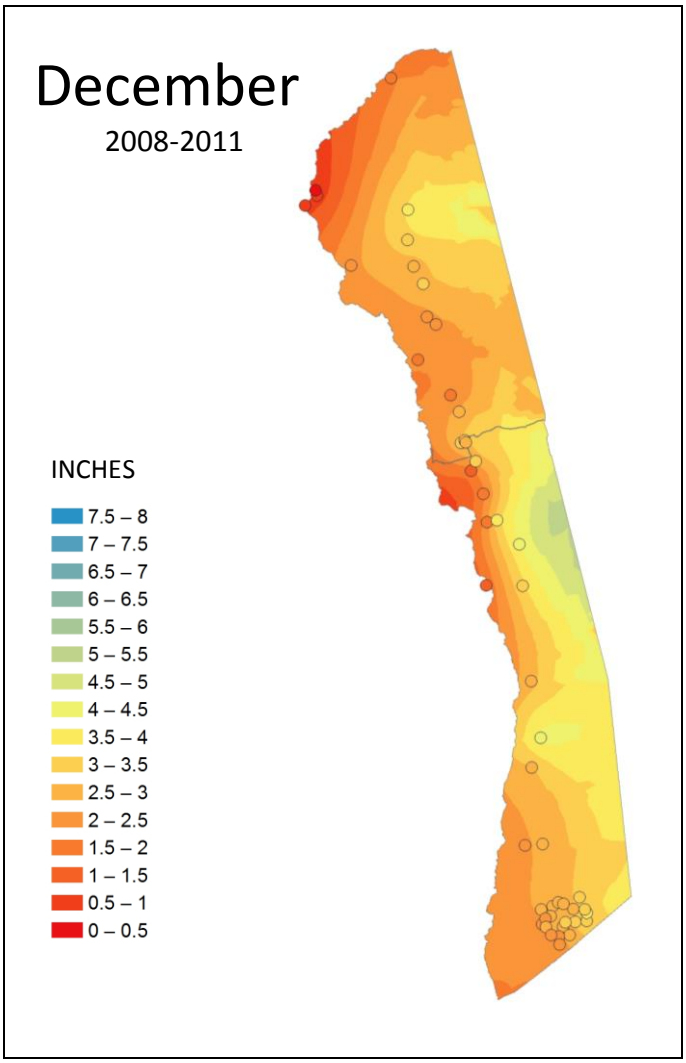


Figure 27: December rainfall surface generated with 4-year average, 2008-2011.

4.1.13 Side-by-side comparison of all months

Positioning the rainfall maps side-by-side, reveals the extremes of month-to-month variations in rainfall, and is shown in Figure 28.

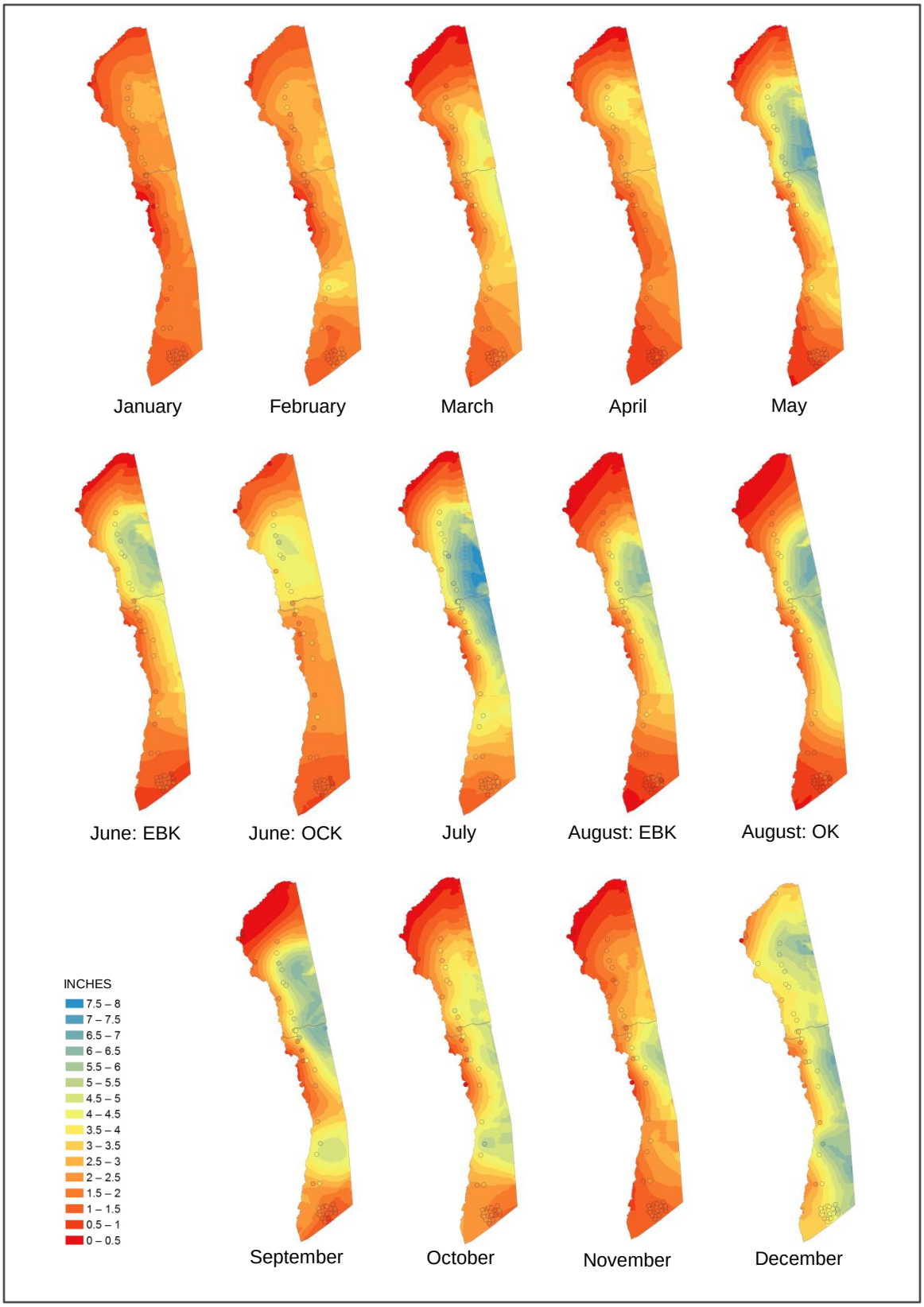


Figure 28: Side-by-side comparison of rainfall maps.

4.2 Evaluating the Results

Two primary methods were considered to evaluate the results obtained: visually reviewing whether or not station rainfall totals match underlying interpolated surfaces, and production of Prediction Standard Error maps.

4.2.1 Do Stations Match Interpolated Surfaces?

Since the rainfall stations supplied the values that the interpolated surfaces are modeled from, it would seem to follow that the color of the corresponding rainfall ranges should be the same as the rainfall stations. For instance, if a rainfall station received a monthly average of 1.5 in., then it logically follows that the underlying surface should also reflect 1.5 in. As mentioned in the month-by-month results section above, most monthly rainfall maps contained several stations and surfaces that did not match.

Contrary to this common thinking, Krivoruchko (2011) explains when working with imprecise and error prone data like rainfall or elevation, "A new prediction at the data location is more accurate than the original noisy measurement." As pointed out earlier in this thesis, kriging is not an exact interpolator, and predicted values may be different than measured values. What may seem counterintuitive is explained by kriging using information from neighboring measurements that are spatially correlated to improve what is known about actual measured values (Kresic and Mikszewski 2013). So although stations and surfaces didn't always match, it is possible that the created interpolated surface is more accurate than the station measurement.

4.2.2 Prediction Standard Error Maps

As an additional assessment of the accuracy of the produced surfaces, kriging has the ability to produce maps of the prediction standard error. As explained by Esri (2013), "A simple

rule of thumb is that 95 percent of the time, the [value] will be inside the interval formed by the predicted value ± 2 times the prediction standard error." A result of zero is ideal, and indicates the highest level of interpolator accuracy. Increased point density and shortest distances to points will have the lowest prediction standard error values, while areas with sparser points will produce higher values.

The prediction standard error map for the July result using EBK is shown as an example in Figure 29. The lowest values, indicating more accurate interpolators, are displayed as the lightest colors. Where the density of stations is higher, the values are at their lowest. These values are found in the north-south line of rainfall measurement stations and the cluster of stations in the south. Darker colors express a greater level of uncertainty and generally increase as the distance from the rainfall stations increase.

A side-by-side comparison of prediction standard error maps for each month is shown in Figures 30-34. The range of values used for all maps was 0.0-4.5 in., as 4.5 was the highest value produced. Two maps were produced for June (Figure 32) and August (Figure 33), as each of these months had a tie for best method based on prediction error statistics. In both June and August, the prediction standard error map for EBK shows a higher level of uncertainty.

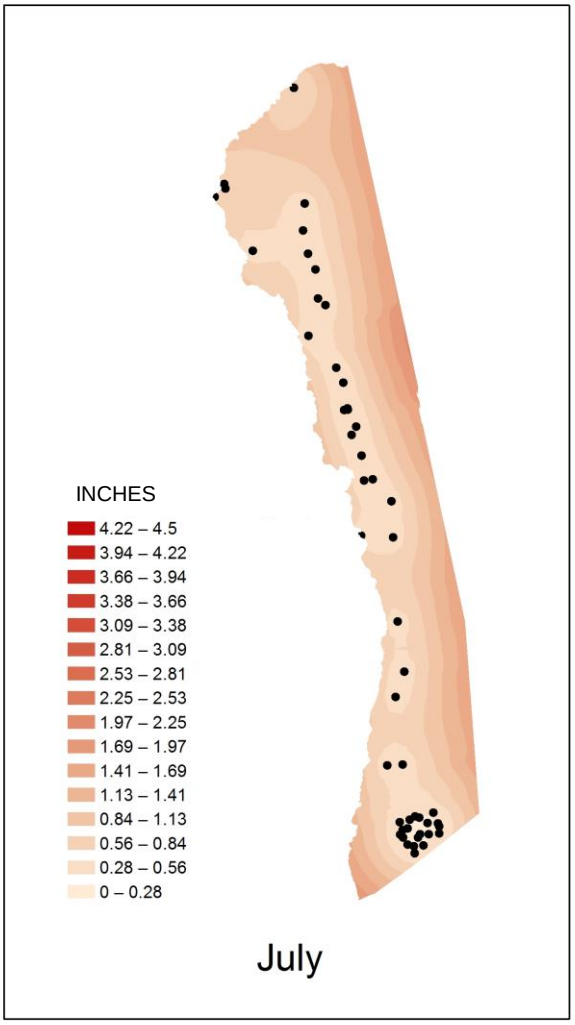


Figure 29: Prediction standard error map for July result using EBK.

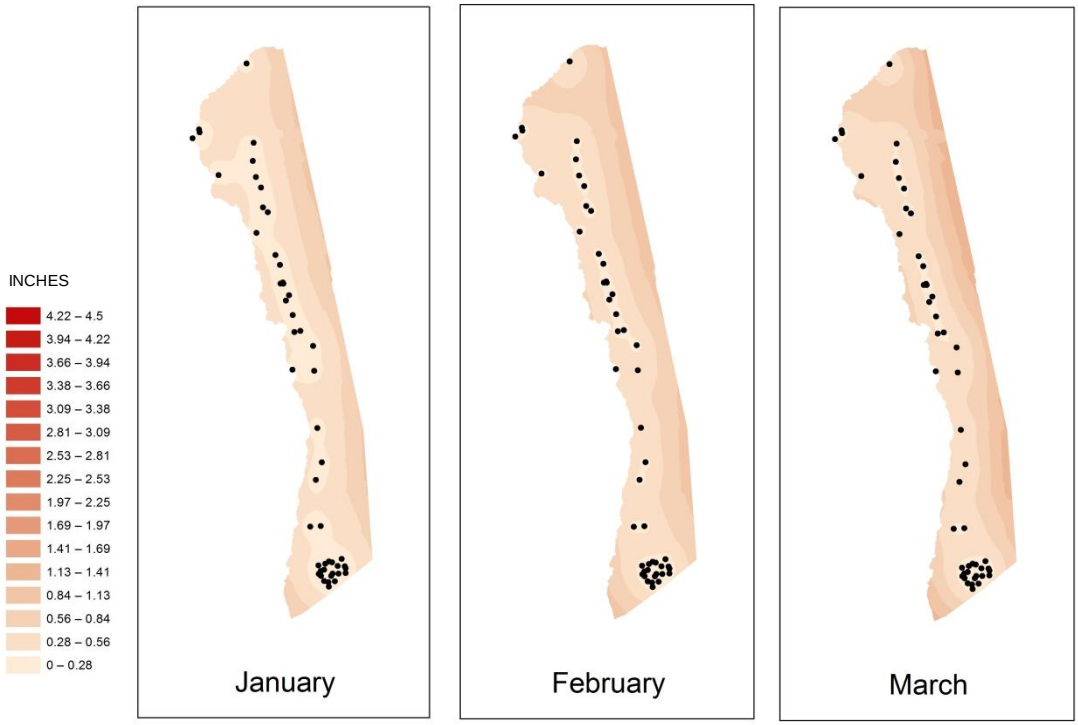


Figure 30: Prediction standard error maps generated for January, February, and March.

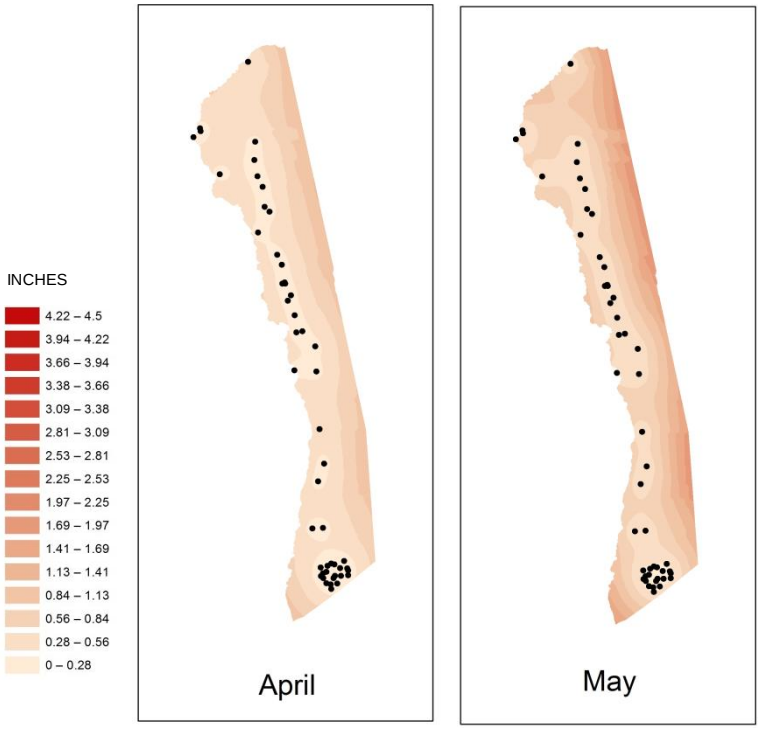


Figure 31: Prediction standard error maps generated for April and May.

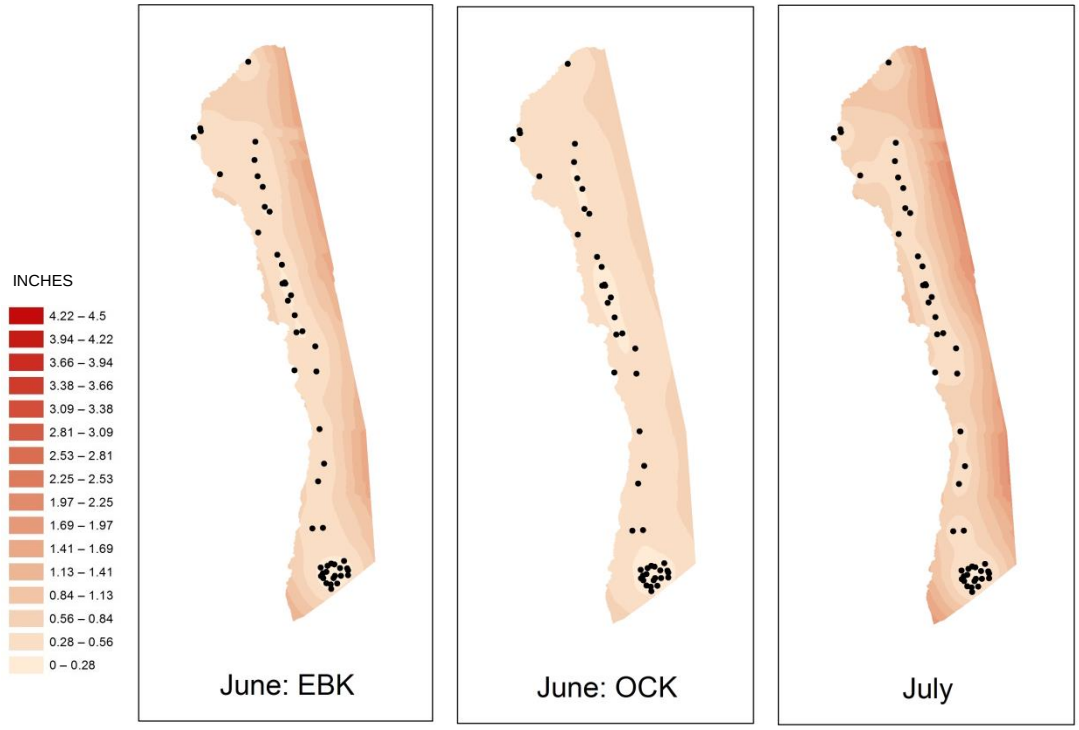


Figure 32: Prediction standard error maps generated for June and July.

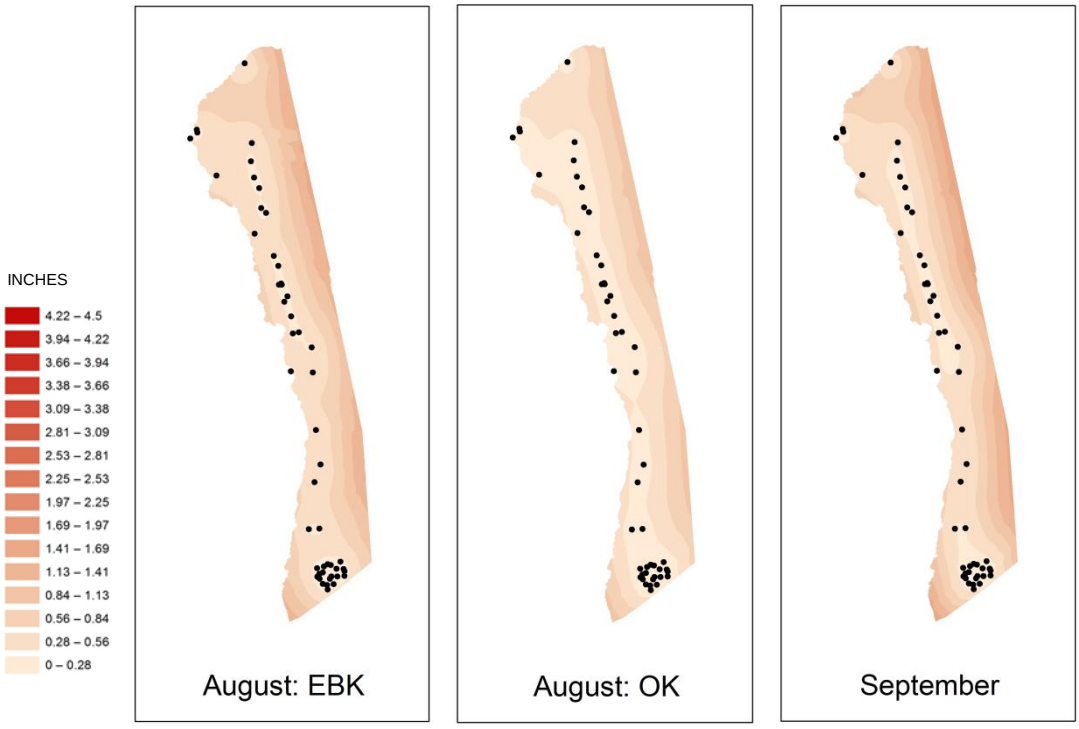


Figure 33: Prediction standard error maps generated for August and September.

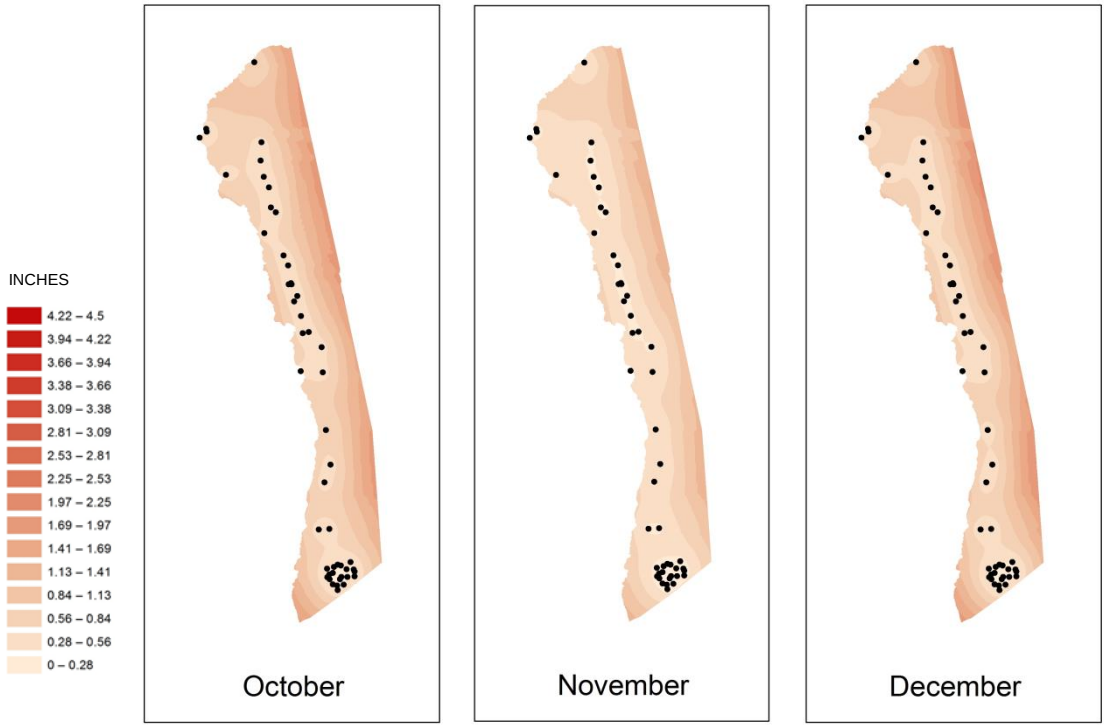


Figure 34: Prediction standard error maps generated for October, November, and December.

4.3 Analyzing the Southern Cluster

As discussed previously, the cluster of stations in the southern region presented an opportunity to create rainfall maps for a denser spatial distribution of rainfall measurements. With the possibility of identical collection protocol for 19 of the 20 stations, aspects of citizen induced error may not be removed, but they should be consistent throughout the dataset.

The surface area of the southern cluster is substantially smaller than the original study area, and is dwarfed by the entire island of Hawaii (Figure 35). All stations are within a five mile diameter. When viewed at a smaller scale throughout the previous analysis, many stations appeared to overlap. However, with the closer view possible in this analysis, it is evident they are

spread apart (Figure 36). The lowest elevation of the area is just under 1000 ft. in the southwest and slopes upwards in a northeasterly direction to over 2600 ft.

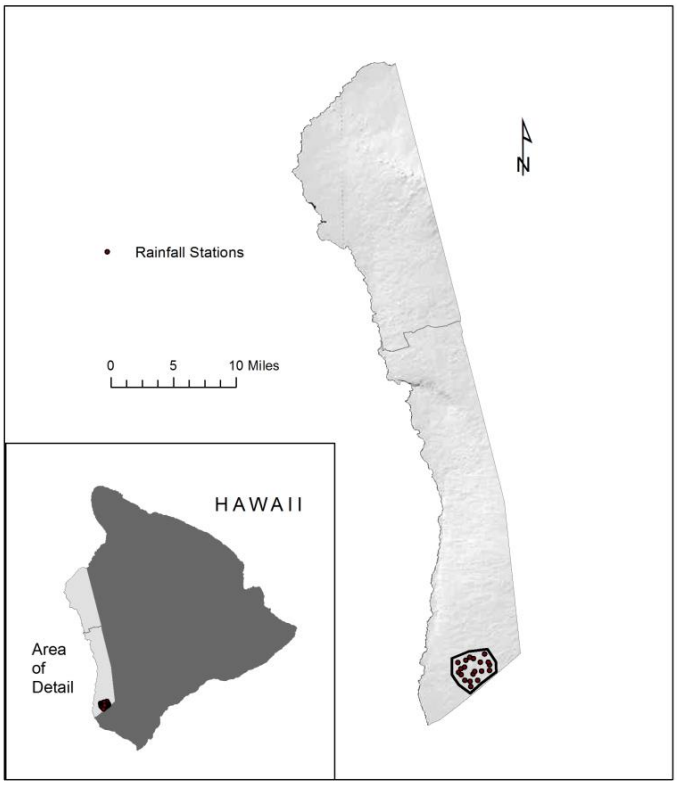


Figure 35: Southern cluster of rainfall stations' location and size in relation to the study area and island of Hawaii.

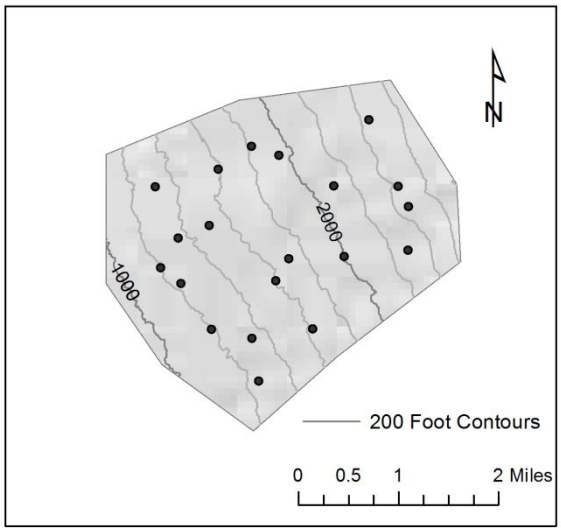


Figure 36: Closer view of southern cluster of rainfall stations.

The rainfall averages for all twelve months were interpolated using EBK. EBK was used for this exercise as it produced the majority of best models in the preceding results, as discussed in Chapter 4. The results for the southern cluster are shown in Figure 37. To provide an equitable month-to-month of comparison, each rainfall surface was symbolized by using the same rainfall value range: the low and high values were used from the lowest and highest monthly station averages within the area. The resulting range, 0.598-5.018 in., was subsequently divided equally into 16 rainfall ranges.

In seven of the maps (Feb.-May, Jul., Aug. and Dec.), a pattern of increased rainfall followed a similar southwest to northeast pattern, the same direction as the rising elevation depicted in Figure 36. The January, June, and November maps showed little variation across the surface. The October and September maps exhibited a pattern opposite of the first seven: monthly rainfall averages increased, as elevation decreased. December was the wettest month, as has been discussed in previous sections.

In addition, prediction standard error maps were produced for each month to provide an assessment of each map's accuracy (Figure 38). The lowest and highest values across all months were determined, resulting in a range of 0.031-0.361 in., and used as the value range for all maps. September, one of the two months with a rainfall map with increasing rainfall over decreasing elevation, had the prediction error map with the most uncertainty. The December result also showed a high level of uncertainty overall, but with better results observed closer to the stations. The remaining months showed a similar overall pattern of the highest amount of uncertainty around the study area's perimeter and increased accuracy closer to stations.

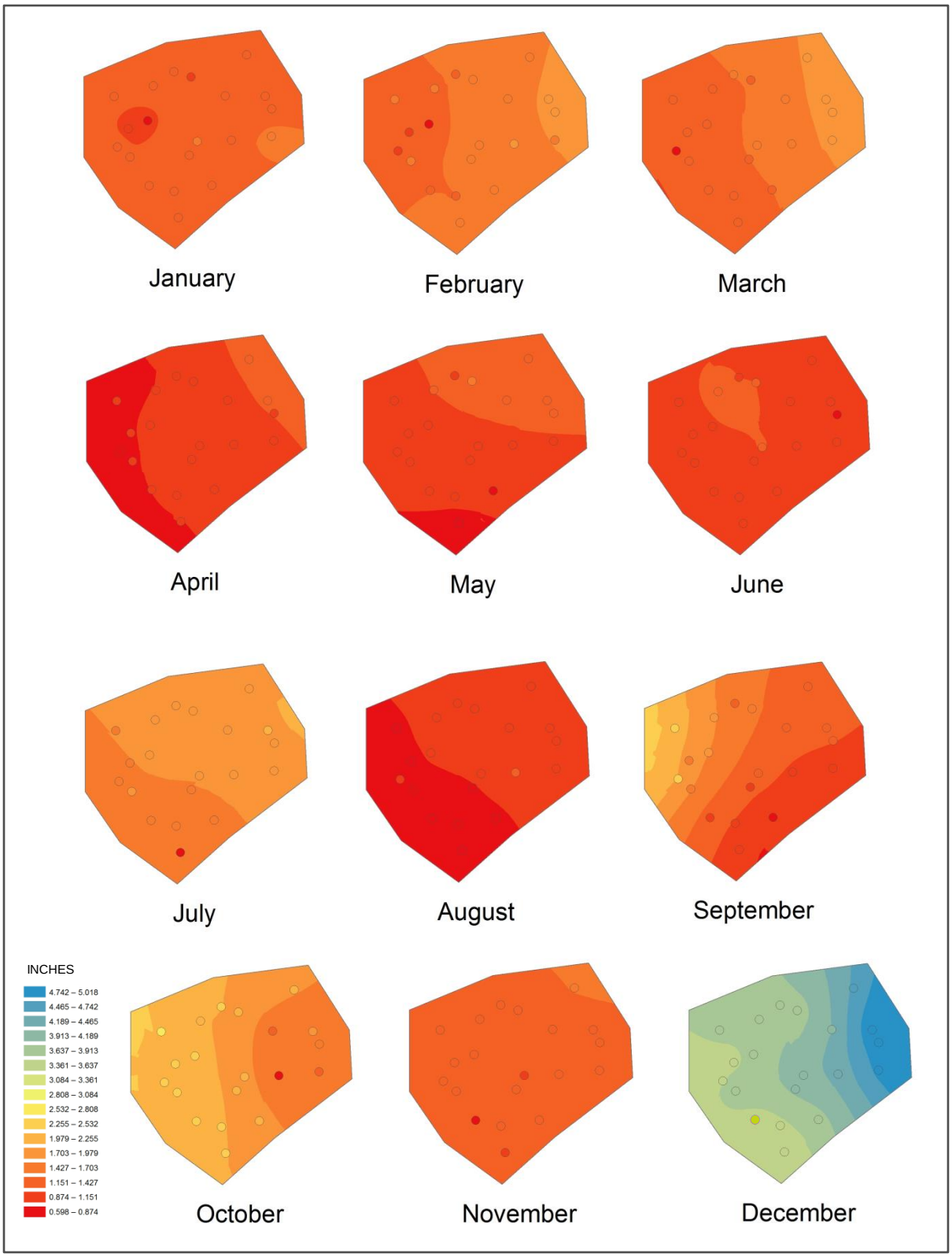


Figure 37: Rainfall surfaces for southern cluster sub-study area.

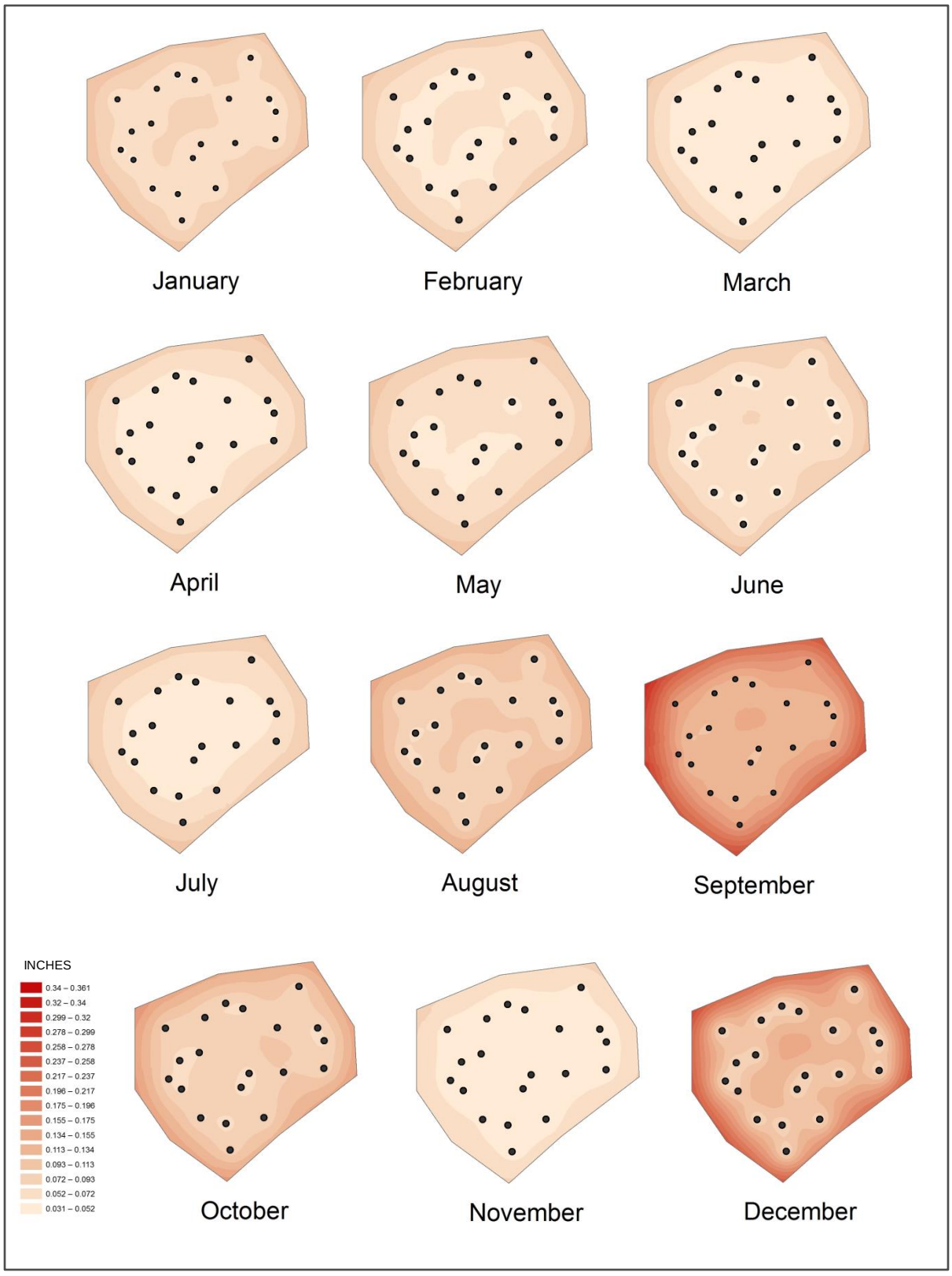


Figure 38: Prediction standard error maps for southern cluster sub-study area.

4.4 Comparison to Online Rainfall Atlas of Hawai'i

As discussed previously, the Online Rainfall Atlas of Hawai'i (Giambelluca et al 2013) presents one of the most detailed rainfall mapping studies to date of the Hawaiian Islands. While the Atlas used 30 years of rainfall data (1978-2007), versus five years of data used in this study (2007-2001), it is useful to compare the results of this study with the Atlas. One of the goals of this study was to produce rainfall maps with higher spatial variation for the study area than those found in the online Atlas. Rainfall maps from this study were compared to those of the online Atlas for the months of February, July, and September.

The downloaded Atlas raster data is shown symbolized using the same color scheme used in their online Atlas website (Figure 39). Filled contours were used instead of a continuous color range, as the Atlas maps needed to be in the same format as those in this study for visual comparison. As noted above in Section 3.5.5, in ArcGIS, kriging results are not output directly as rasters, but instead in the native format of the interpolation procedure. Therefore, it is not possible to compare the spatial resolution of these two datasets.

The study area of this study is shown outlined along the Kona coast. The range of rainfall values necessary to cover the island's wide range of July rainfall, 0.233-28.97 in., causes the study area to be symbolized with little variation. The network of rainfall gauges used by the online Atlas (Figure 40) consists of 53 stations. Unlike this study in which missing records caused stations to be removed from the analysis, the Atlas used a sophisticated process to calculate missing data over the much longer period of record used. This allowed them to retain a dense set of stations, even though not all of them persisted for the entire period of analysis. Indeed, as of the 2011 when the Atlas was published, 41 of the stations had been discontinued, with 12 remaining operational. Many of the NOAA NCDC stations used in this study are found

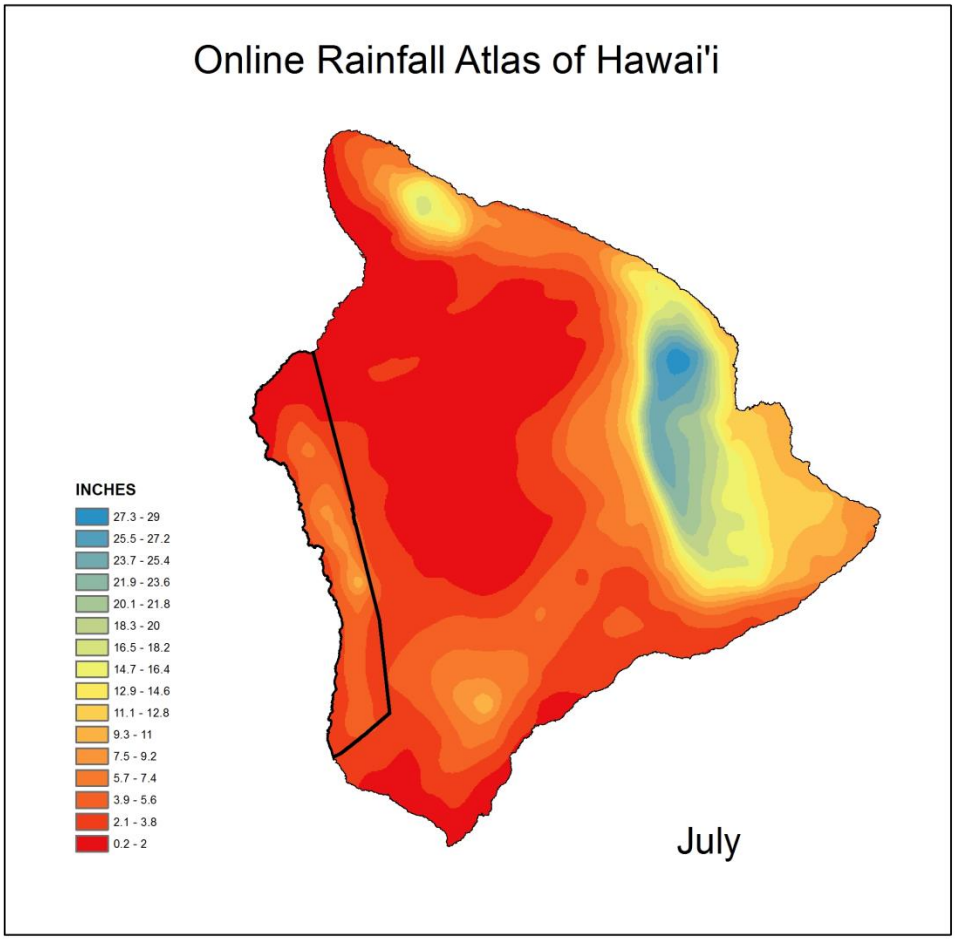


Figure 39: July rainfall map from the Online Rainfall Atlas of Hawai'i with study area outlined.

in the Atlas's network, and can be seen in a side-by-side comparison (Figure 40).

While the number of stations used in each map is relatively close, the spatial distribution differs significantly. The online Atlas's gauges are located predominantly in the center of the study area, providing a denser network than that of this study. There are several more gauges that were used in this study found along the southwest coast, and a substantially larger group of stations in the south.

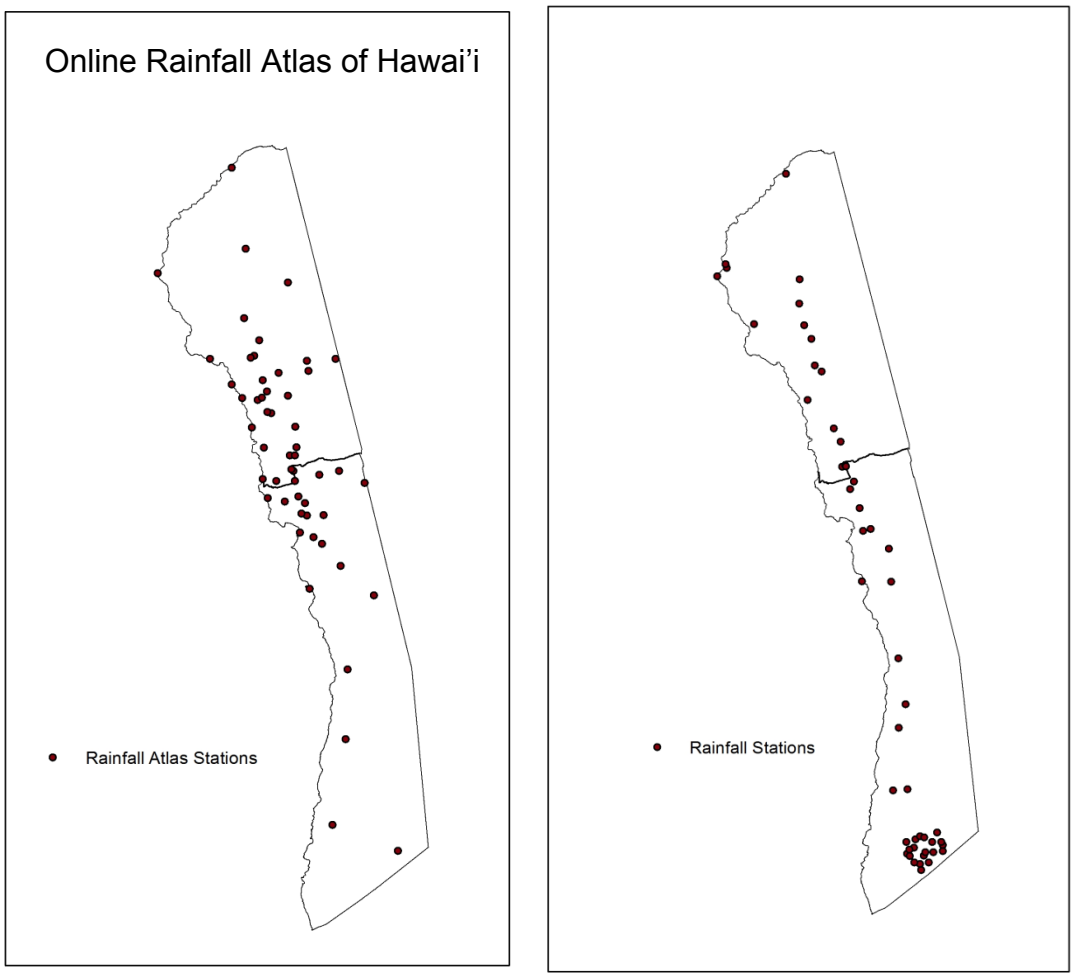


Figure 40: Comparison of rainfall station locations from the Online Rainfall Atlas of Hawai'i on the left, and this study's on the right.

Since the Atlas's data covers a different period, different rainfall values exist in the same locations for each set of maps. Rainfall value ranges using the minimum and maximum of each displayed map were used to illustrate its spatial variation for comparison. A side-by-side visual review of the February maps (Figure 41) shows a similar level of spatial variation throughout the top two-thirds of the study area. The maps for July (Figure 42) and September (Figure 43) show the Atlas's maps have a greater level of spatial variation in this region. This study's maps have an area of higher rainfall represented in the central south region as a result of the previously discussed (Section 4.1; Figure 14) station 28. While this citizen station was noted for its higher

elevation location as the cause for higher rainfall, the comparisons to the Atlas maps provide an indication that there may have been a data recording problem; a circumstance of using VGI data. In the far south, the maps for all three months produced for this study show a different pattern of rainfall than the Atlas's, and significantly more spatial variation.

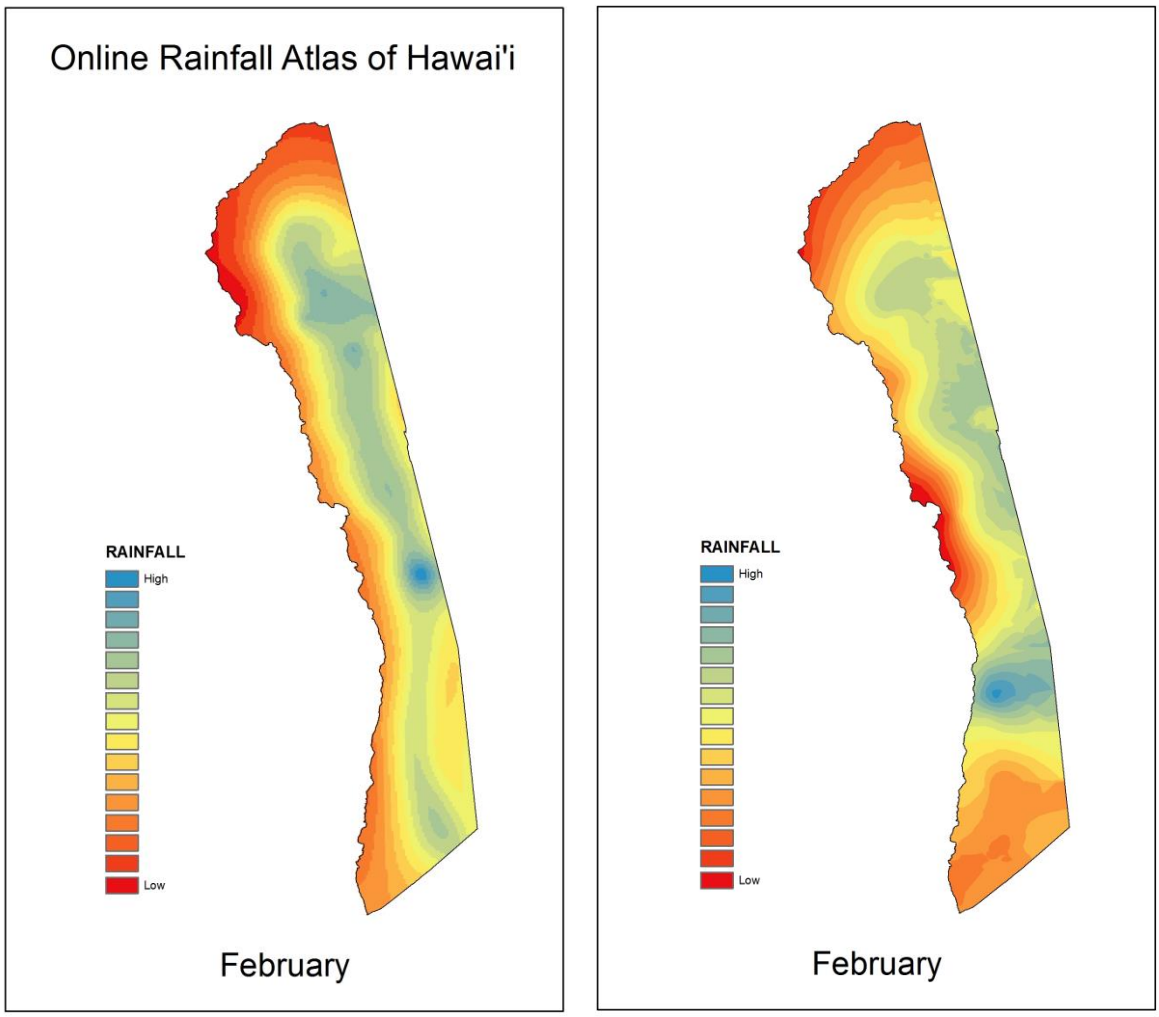


Figure 41: Comparison of February rainfall maps for the Online Rainfall Atlas of Hawai'i on the left, and this study's map on the right.

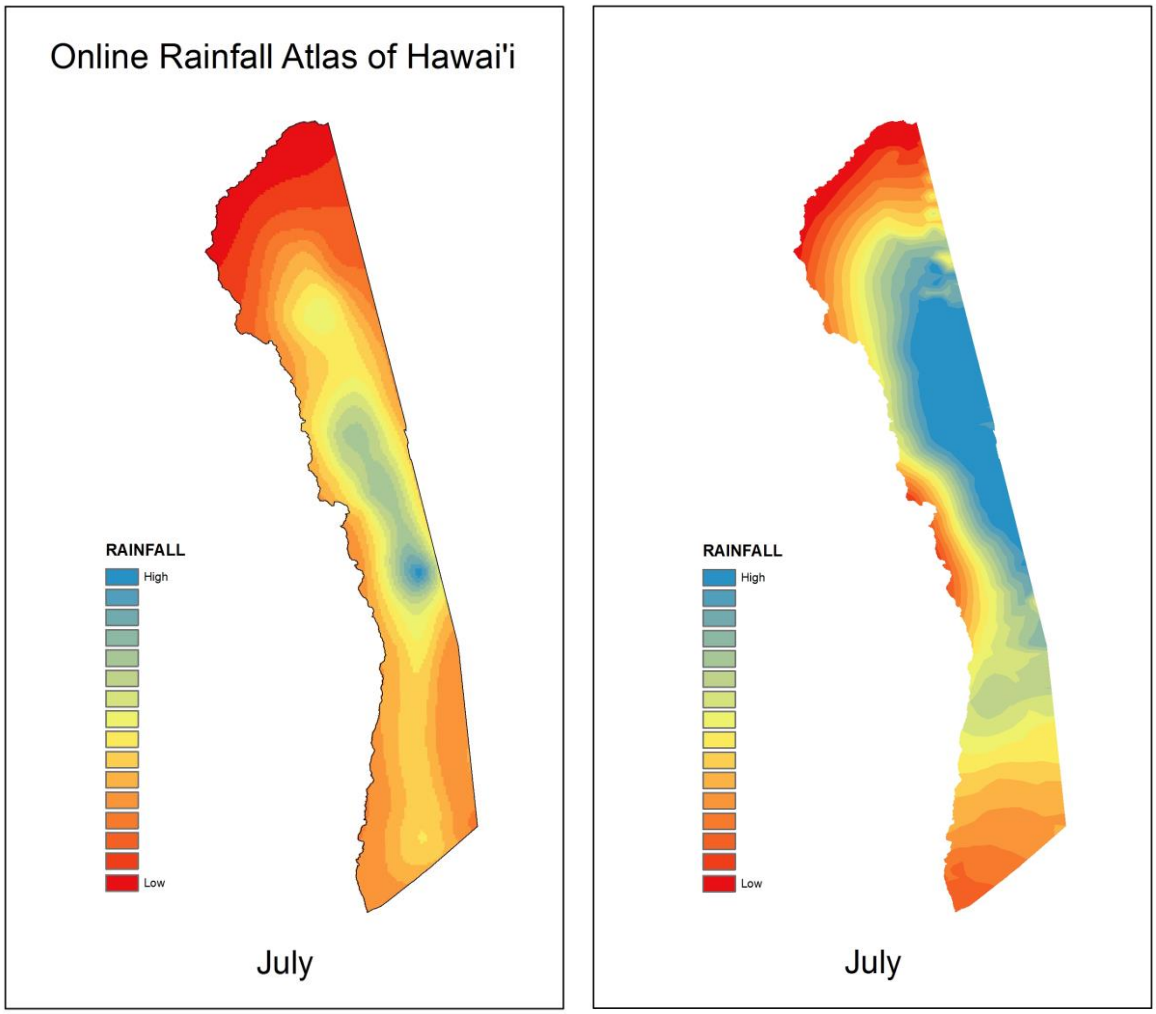


Figure 42: Comparison of July rainfall maps for the Online Rainfall Atlas of Hawai'i on the left, and this study's map on the right.

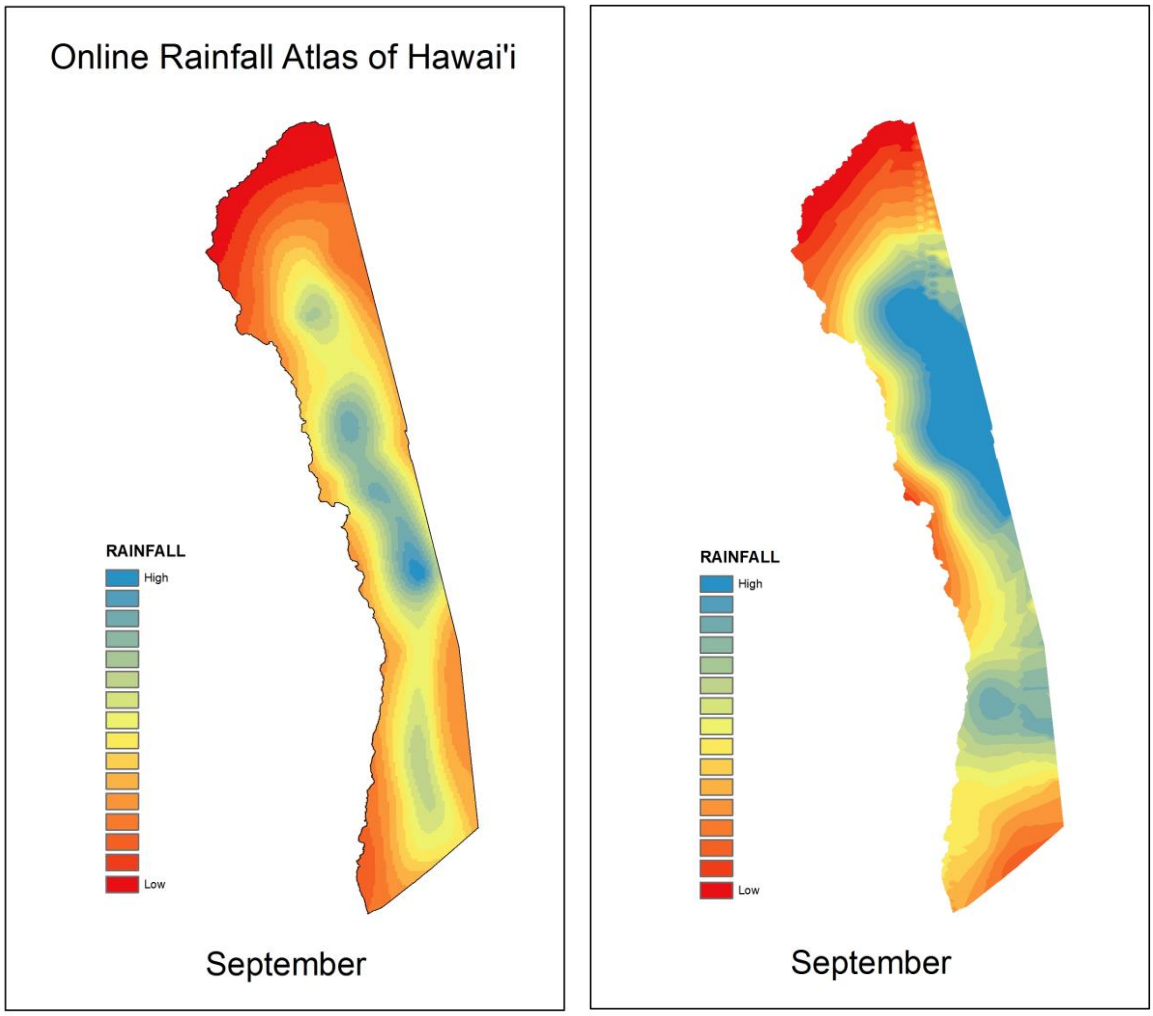


Figure 43: Comparison of September rainfall maps for the Online Rainfall Atlas of Hawai'i on the left, and this study's map on the right.

A comparison of the southern cluster sub-study area was also conducted. However, no Atlas rainfall stations were located within the sub-study area (Figure 44); one station was on the eastern perimeter and a second to the north. This contrasts with the dense network of 20 stations used in this study, as shown in Figure 36. All maps were symbolized using the same minimum and maximum technique used above, which allowed for a larger range of rainfall values to be colorized (Figures 45-47).

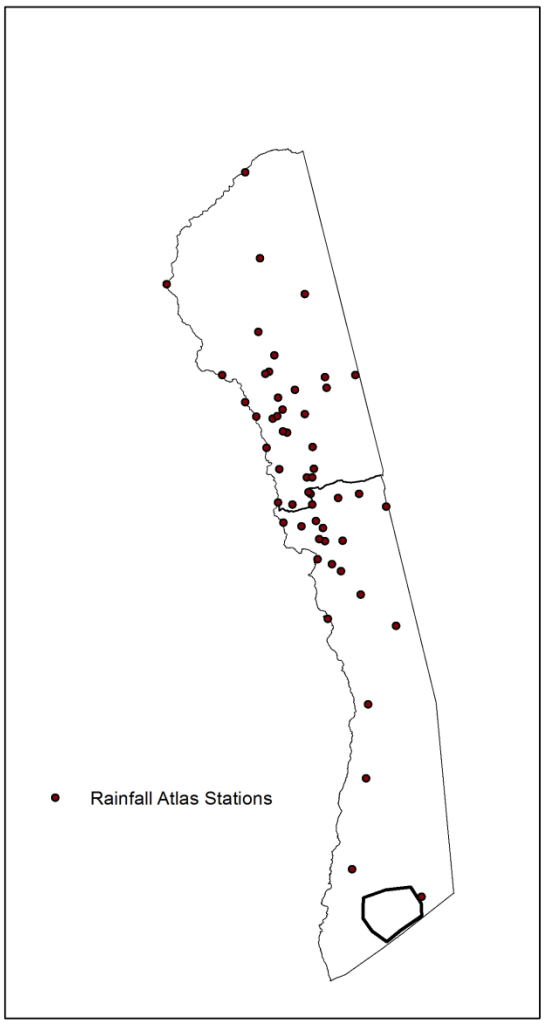


Figure 44: Southern sub-study area outline in relation to Online Rainfall Atlas of Hawai'i rainfall station gauge locations.

The maps produced for this study show a higher level of spatial variation, as the colored ranges twist and turn throughout the sub-study area. By contrast, the Atlas's patterns appeared to have been affected by the value of a single station, with values gradually decreasing as distances increased from them. Even though the values changed, there was little variation in the resulting patterns.

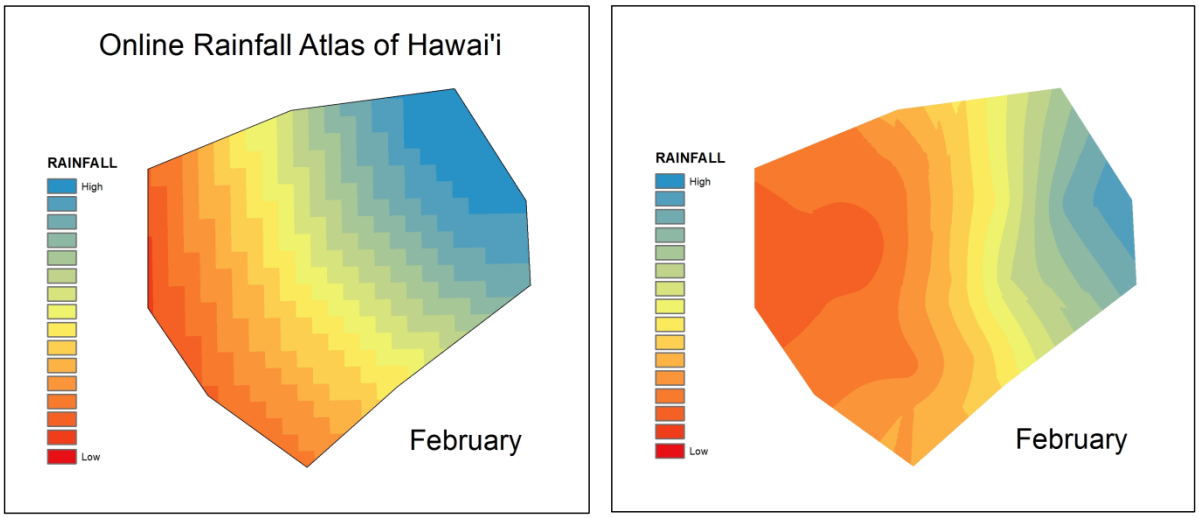


Figure 45: February rainfall maps for the southern cluster. The Online Rainfall Atlas of Hawai'i is on the left, and this study's map on the right.

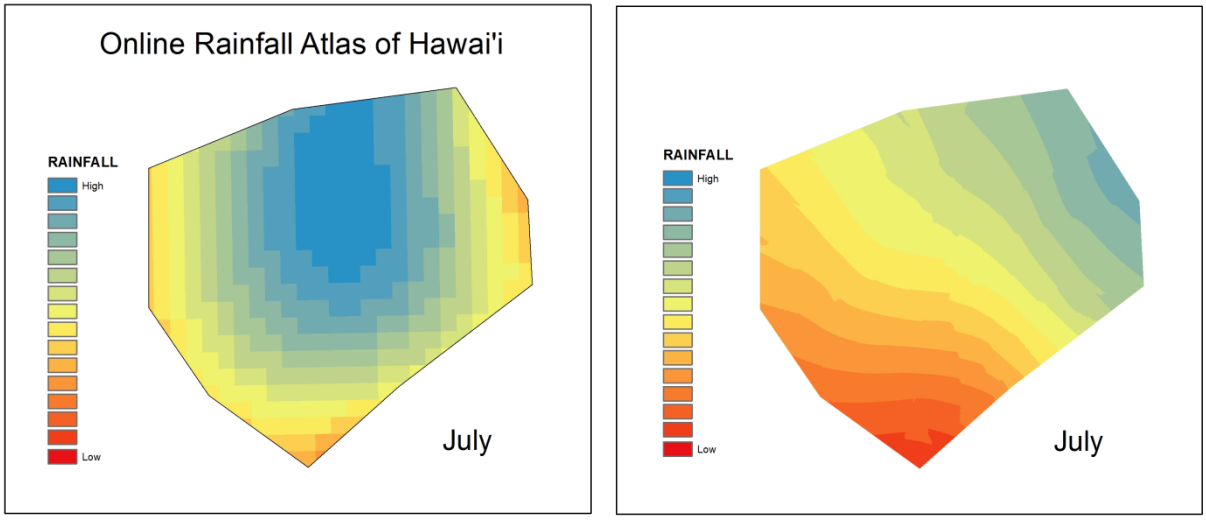


Figure 46: July rainfall maps for the southern cluster. The Online Rainfall Atlas of Hawai'i is on the left, and this study's map on the right.

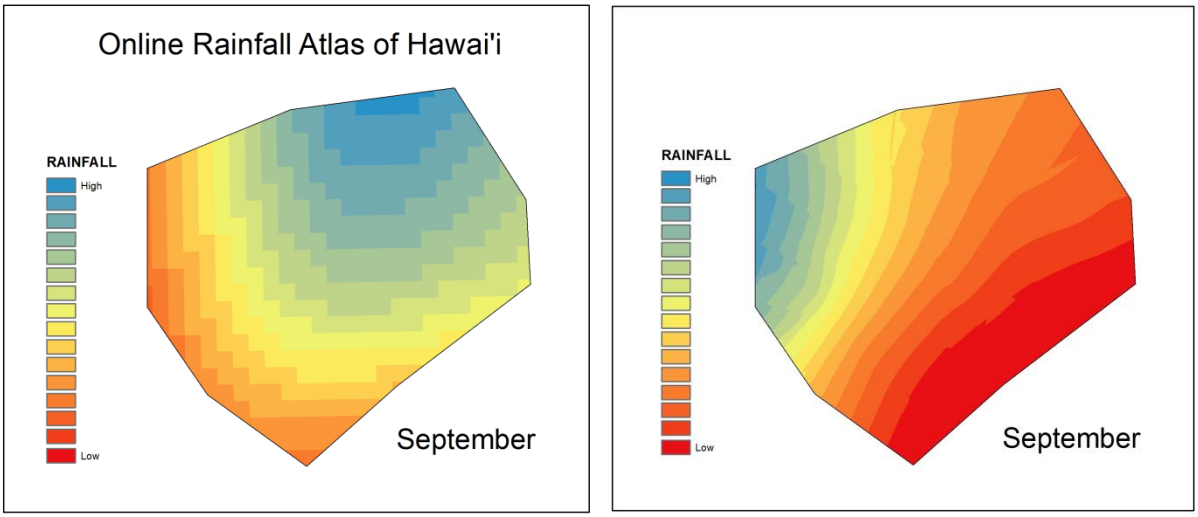


Figure 47: September rainfall maps for the southern cluster. The Online Rainfall Atlas of Hawai'i is on the left, and this study's map on the right.

Chapter 5 – Conclusion and Future Work

This study has created monthly rainfall surface maps exhibiting areas with greater spatial variation than others previously produced using NWS rainfall data. A comparison of the results of this study created for the Kona study area with the same area of the Online Rainfall Atlas of Hawai'i, showed higher levels of spatial variation were achieved in both datasets when a denser network of rainfall gauge stations was used. A closer examination of a sub-study area with a dense cluster of rain gauges in the southern region of the Kona study area made it possible to create rainfall surface maps with a higher level of spatial variation than the online Atlas.

In addition, this study demonstrated that VGI can be used to create rainfall surfaces, a tool used in climatological monitoring. Citizen collected rainfall measurements were the primary data source used in this study, and ultimately provided the dense network of gauges necessary for examination of the rainfall of a small geographic area.

Prediction error results were used to evaluate three geostatistical methods of kriging, providing a side-by-side comparison of quantifiable results to determine the most suitable approach of spatial interpolation for this dataset. EBK proved to be the best method for nine of the months while the best result for the remaining three months was accomplished by the use of ordinary co-kriging or EBK, ordinary kriging or EBK, and ordinary kriging. While ordinary kriging was included as a benchmark to compare the results of ordinary co-kriging and EBK to, it provided the best results for one month and tied for best in another. The use of Exploratory Regression to determine a better combination of variables did not provide results that improved any of the outcomes.

5.1 Validity of Results

The validity of the results presented can be challenged based on the lack of station coverage of the study area. While the overall number of rainfall stations used in this study were similar to those used in the respected Online Rainfall Atlas of Hawai'i, there were some areas that were sparse in the context of this micro-scale study. Even though the southern region had a cluster of stations nearly overlapping, most other areas had greater distances between stations. Better results for the study area could have been obtained by using a denser set of stations.

The spatial arrangement of the stations also compromised the final surfaces. The pattern of the stations followed a north-south line, with a few bends to the east and west. As indicated by the prediction standard error maps, the results to the east of the line of points were less reliable. A more evenly distributed, broken up pattern of stations would have produced more accurate results. The general spread of population, as indicated by existing roads within the study area, is within a few miles of the north-south main highway, limiting the existence of rainfall collectors in areas far from the highway. Land directly to the east of this area is uninhabited and at much greater elevations.

The temporal aspect of the rainfall data may also have limited the accuracy of the results. While thirty years of data is preferred for rainfall average studies, only five years of data was used in this study. While accurate averages can still be produced using five years, the average is more susceptible to non-average results from anomalous periods of heavier or lighter rainfall. The use of thirty years of data averages-out these high and low anomalies. The decision was made in this study to use five years of data, as that greatly increased the number of rainfall stations, and led to better coverage of the study area. Only a limited number of stations were able to provide rainfall records for a longer period of time.

5.2 The Use of VGI

The use of VGI as a primary source of data cannot be overlooked in this study. While each citizen data collector is likely to have had the best intentions to collect the most accurate data possible, the accuracy of that data cannot be determined. Variations in rainfall gauge types, time of measurements, methods of measurements, tabulation of records, monitoring of personal gauges, and proper gauge placement, could have compromised the accuracy of the collected data.

However, using semivariogram clouds paired with histograms was successful in identifying VGI data that did not fall within the value ranges of the authoritative data. Out of 35 citizen measurements, only two were deemed questionable, and were removed from the dataset. After removal of this potential outlier data, all of the VGI data values were consistent with the values from the authoritative sources.

If VGI were not used in this study, there would have been 17 rainfall stations within the study area, instead of 52. Within the southern sub-study area, there would have been one station, instead of 20. With this decrease in stations, the spatial distribution and density of existing stations would have been significantly lower than previous efforts to map rainfall, such as the Online Rainfall Atlas of Hawai'i. The use of citizen collected data helped make this study possible and worthwhile.

5.3 Modifiable Temporal Unit Problem

In studies of spatially aggregated data, it has been demonstrated that the size and structure of spatial units may affect the results of analyses, an issue which is referred to as the Modifiable Areal Unit Problem (MAUP) (Openshaw and Taylor 1979). Similarly, temporal units are subject to aggregation problems. As de Jong and de Bruin (2012) point out in a study measuring various levels of temporal aggregation of vegetative cycles, the Modifiable Temporal

Unit Problem (MTUP) may also affect the result of analyses. Start and stop times, the length of temporal units, and level of aggregation, if used carelessly can lead to spurious results.

The temporal unit used for this study was the calendar month. But because a month can be 28, 29, 30, or 31 days, it is not a constant value. For this reason, the months with 31 days will have the potential to receive more rainfall than those months with fewer days. For this reason alone, it is not an equitable comparison to compare rainfall totals from month-to-month.

In addition, the start and end dates of a month are arbitrary when comparing to weather cycles. If the first day of a month was to occur just a few days later than its actual start date, there is the potential that a heavy rainfall event may have occurred in those days, and therefore be attributed to a different month. This is particularly important in this area since high rainfall month's totals are significantly impacted by one single day's heavy rainfall. For example, one farmer recorded a single day's rainfall of 3.75 in. that pushed his May 2013 total to the highest recorded in his period of recording, 8.03 in., and contrasts sharply with the unusually dry months of around 1 in. on either side of it (personal communication: Keola Childs, 1 June 2013, via email).

These somewhat arbitrary dates serve as the framework of this study. A researcher could divide a year into equal parts; for example 12 months of 30 days, or 13 months of 28 days. But these numbers are also arbitrary, as one could use any number of days they chose. A method to lessen the effects of this would be to use a rolling multiday average and then aggregate the averages. However, in this study, the majority of data collectors tabulated their data in standard calendar months, and it was impossible to manipulate them temporally. Furthermore, since each month is calculated and evaluated separately, conclusions within monthly bases are not affected by this issue.

5.4 Future Work to Be Considered

Any future work of this nature would benefit from a denser, more evenly distributed network of stations. Unfortunately, additional existing stations in this study area are likely to be along the same north-south line, as farms, residents and rainfall stations do not exist much above 3,000 ft. elevation. An increase in the number of stations in this same area would increase the density of stations, and improve the results in this study area, but limitations would still be encountered due to the lack of wide spatial distribution.

The existence of other regions with a dense network of citizen rainfall gauges are likely, and worth investigating. Sparse locations of government manned weather stations provide generalized rainfall maps to regions that could benefit from better climatological monitoring.

References

Bedient, P.B. & Huber, W.C. (1992). *Hydrology and Floodplain Analysis*, 2nd ed. Reading, MA: Addison-Wesley.

Cifelli, R., Doesken, N., Kennedy, P., Carey, L.D., & et al. (2005). The Community Collaborative Rain, Hail, and Snow Network. *Bulletin of the American Meteorological Society, 86*(8), 1069.

Chen, Y., & Nash, A. J. (1994). Diurnal variation of surface airflow and rainfall frequencies on the island of Hawaii. *Monthly Weather Review, 122*(1), 34-56. doi: 10.1175/1520-0493(1994)122<0034:DVOSAA>2.0.CO;2

Chu, P., & Chen, H. (2005). Interannual and interdecadal rainfall variations in the Hawaiian islands. *Journal of Climate, 18*(22), 4796-4813. doi: 10.1175/JCLI3578.1

Connors, J.P., Lei, S., & Kelly, M. (2012). Citizen science in the age of neogeography: Utilizing volunteered geographic information for environmental monitoring. *Association of American Geographers. Annals of the Association of American Geographers, 102*(6), 1267.

De Jong, R. & de Bruin, S. (2012). Linear trends in seasonal vegetation time series and the modifiable temporal unit problem. *Biogeosciences*, (9) 71-77. doi: 10.5194/bg-9-71-2012

Diodato, N. (2005). The influence of topographic co-variables on the spatial variability of precipitation over small regions of complex terrain. *International Journal of Climatology*, 25: 351-363. doi: 10.1002/joc.1131

Elwood, S. (2008). Volunteered geographic information: Future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal, 72*(3), 173-183. doi: 10.1007/s10708-008-9186-0

Elwood, S., Goodchild, M.F., & Sui, D.Z. (2012). Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Association of American Geographers. Annals of the Association of American Geographers, 102*(3), 571. doi: 10.1080/00045608.2011.595657

Esri (2012). What are the different kriging models? ArcGIS Resources. Last accessed on 22 November 2012 at <http://resources.arcgis.com/en/help/main/10.1/index.html#//00310000003q000000>

Esri (2012). Extensions, Geostatistical Analyst, Exercise 3: Mapping Ozone Concentration. ArcGIS Resources. Last accessed on 20 June 2013 at <http://resources.arcgis.com/en/help/main/10.1/index.html#//0031000000pz000000>

Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal, 72*(3), 137-148. doi: 10.1007/s10708-008-9188-y

Fiebrich, C. A. (2009). History of surface weather observations in the United States. *Earth Science Reviews, 93*(3), 77-84. doi: 10.1016/j.earscirev.2009.01.001

Gesch, D.B. (2007). Chapter 4 – The National Elevation Dataset, in *Digital Elevation Model Technologies and Applications: The DEM Users Manual,* 2nd ed., ed. by D. Maune, 99-118. Bethesda, MD: American Society for Photogrammetry and Remote Sensing.

Giambelluca, T., Chen, Q., Frazier, A., Price, J., Chen, Y., Chu, P., Eischeid, J., and Delparte, D. (2013). Online Rainfall Atlas of Hawai'i. *Bull. Amer. Meteor. Soc.* 94, 313-316, doi: 10.1175/BAMS-D-11-00228.1.

Giambelluca, T., Nullet, M., & Schroeder, T. (1986). *Rainfall Atlas of Hawai'i, Report R76.* Honolulu, HI: State of Hawaii, Department of Land and Natural Resources.

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal, 69*(4), 211-221. doi: 10.1007/s10708-007-9111-y

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology, 228*(1), 113-129. doi: 10.1016/S0022-1694(00)00144-X

Gould, P. (1999). *Becoming a Geographer*. Syracuse, NY: Syracuse University Press.

Krivoruchko, K. (2011). *Spatial Statistical Data Analysis for GIS Users.* 928 pp. Redlands, CA: Esri Press.

Kresic, N., & Mikszewski, A. (2013). *Hydrogeological Conceptual Site Models: Data Analysis and Visualization.* Boca Raton, FL: CRC Press.

Lynch, S.D. (2003). Development of a Raster Database of Annual, Monthly and Daily Rainfall for Southern Africa. Water Research Commission, Pretoria, South Africa, WRC Report 1156/1/03. pp78.

Moral, F.J. (2010). Comparison of different geostatistical approaches to map climate variables: application to precipitation. *International Journal of Climatology*, 30: 620-631. doi: 10.1002/joc.1913

McCuen, R.H. (1998). *Hydrologic Analysis and Design*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

Openshaw, S. & Taylor, P.J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: *Statistical applications in the spatial sciences*, ed. By N. Wrigley, 127-144. London: Pion.

Phillips, D.L., Dolph, J., & Marks, D. (1992). A comparison of geostatistical procedures for spatial analysis of precipitation in mountain terrain. *Agricultural Forestry and Meteorology*, 58: 119-141.

Pilz, J., and G. Spöck (2007). Why do we need and how should we implement Bayesian kriging methods. *Stochastic Environmental Research and Risk Assessment* 22 (5): 621–632.

Rodda, J. C., & Dixon, H. (2012). Rainfall measurement revisited. *Weather, 67*(5), 131-136. doi: 10.1002/wea.875

O'Sullivan, D., & Unwin, D. (2010). *Geographic Information Analysis*, 2nd ed., 239-313. Hoboken, New Jersey: John Wiley & Sons, Inc.

Scott, L., & Getis, A. (2008). Spatial Statistics. In *Encyclopedia of Geographic Information Science*, ed. K. Kemp, 439. Thousand Oaks, CA: Sage Publications, Inc.

Strangeways, I. (2004). Improving precipitation measurement. *International Journal of Climatology, 24*(11), 1443-1460. doi: 10.1002/joc.1075

Strangeways, I. (2010). A history of rain gauges. *Weather, 65*(5), 133-138. doi: 10.1002/wea.548

Tabios, G.Q. & Salas, J.D. (1985). A comparative analysis of techniques for spatial interpolation of precipitation. *Water Resources Bulletin*, 21(3), 365-380.

Thiessen, A.H. (1911). Precipitation averages for large areas. *Monthly Weather Review,* 39(7), 1082-1084.

Wu, H., Hubbard, K. G., & You, J. (2005). Some concerns when using data from the cooperative weather station networks: A Nebraska case study. *Journal of Atmospheric and Oceanic Technology, 22*(5), 592-602. doi: 10.1175/JTECH1733.1

Appendices

Appendix A

Citizen Rainfall Measurement Questionnaire

Thank you for offering your help with this exciting local-scale rainfall mapping project. Your contribution is what makes it possible. Again, please be assured that no names or addresses will be publicly associated with the data used.

Below are a few simple questions about your collection process and your data. This should only take a few minutes to complete and there are no wrong answers. You can reply to this email with just your answers with 1, 2, etc. Or if it's easier, just reply and insert them in the original text below.

For the actual rainfall data, ideally we'd like to get monthly totals for the past 5+ years (i.e. back to January 2007). If you have more than this, or even daily data in a simply transmitted format, we'll take it (we love data!). If you have less, that's fine, too. It will all be part of the stew we'll be cooking. People have told us their data is in a variety of formats: calendars, Excel spreadsheet, notepad, digital station record, etc. Again, the goal is to keep this easy for you, so please provide your monthly data in whatever format is most convenient. We could even transcribe from scanned documents if necessary (though not preferred, of course). If you want to start from scratch, you can use a simple table such as the one shown at the bottom of this form. Attach a separate document, or just fill that table in.

Thank you!

Jeffrey Schroeder, MSGIS student, jfschroe@usc.edu

Karen Kemp, Professor, kakemp@usc.edu or karenkemp@geokemp.net

Spatial Sciences Institute, University of Southern California

----------Questionnaire----------

Please provide the following background information on your collection:

1. Your TMK#:

2. Rainfall gauge description (e.g. 4 inch wedge, 8 inch copper cylinder, something electronic from Brookstone, etc.)

3. Units of measurement (inches, millimeters, etc.):

4. Frequency of collection:

5. Optional, provide the location of your rain gauge in latitude/longitude:

(see below for instructions on how to figure this out from Google Map)

6. Any comments and ideas you think we might like to know about your data or neighborhood rainfall.

7. Please provide any details about your data collection process that you think would be important for us to know.

8. Please add your data to this response by attaching a separate document, pasting your data here, or filling in the table below.

----------Finding the latitude and longitude of your gauge----------

To determine the lat/long numbers for the exact location of where you have your rain gauge on your property:

1. Go to http://maps.google.com.

2. You should be looking at a satellite image. --- If it opens on the road map, click on the box on the right that says Satellite.

3. Find your farm or house. --- If you're zoomed out, type your address or the name of a nearby community in the search box right above the map, click the blue magnifying glass, and then click and drag the picture around until you see your house.

4. Zoom in on your place as far as it will let you.

5. Now look around and point with your mouse to the exact place where you have your rain gauge outside.

6. Right click at the point where your gauge is located.

7. In the menu that pops up, point at and click on "What's here?"

8. This will put a green arrow right at your gauge's location!

9. Now, click on that green arrow and the numbers shown are your gauge's lat/long. Copy the top, bold pair of numbers from this box and paste into the questionnaire.

----------Sample Data Table----------

Year	Month	Amount

Appendix B

Semivariogram Clouds, Corresponding Stations, and Histograms

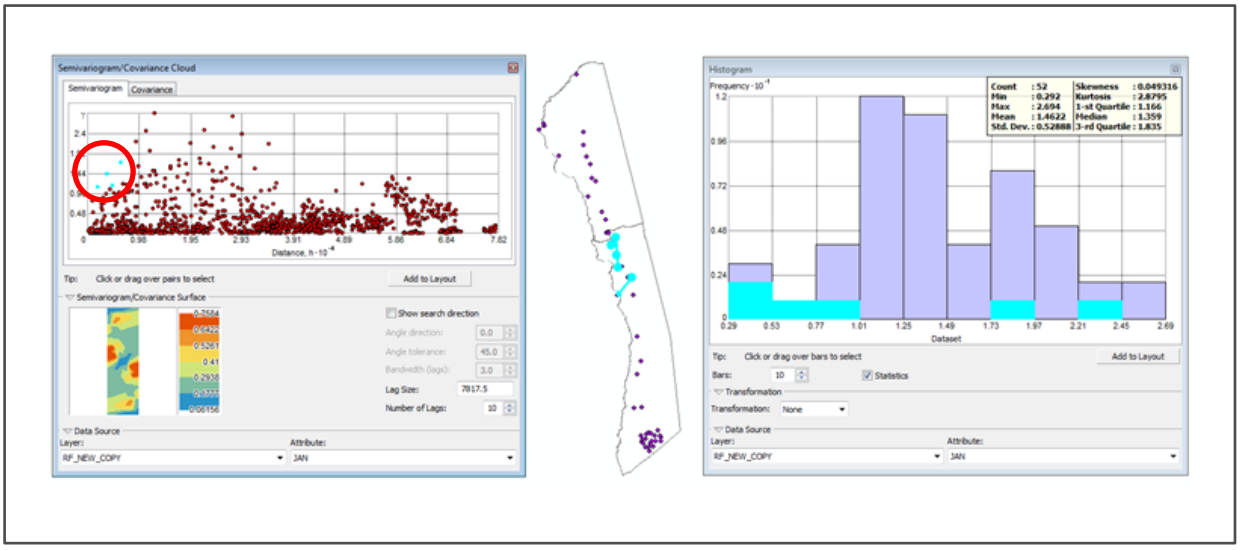


Figure B.1: Semivariogram Cloud, stations, and Histogram for January.

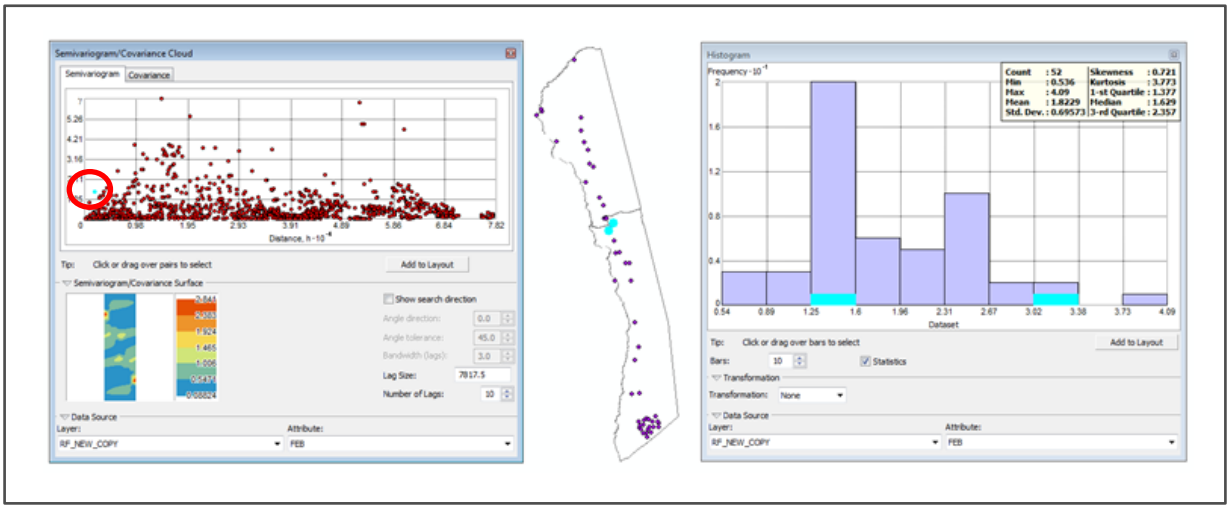


Figure B.2: Semivariogram Cloud, stations, and Histogram for February.

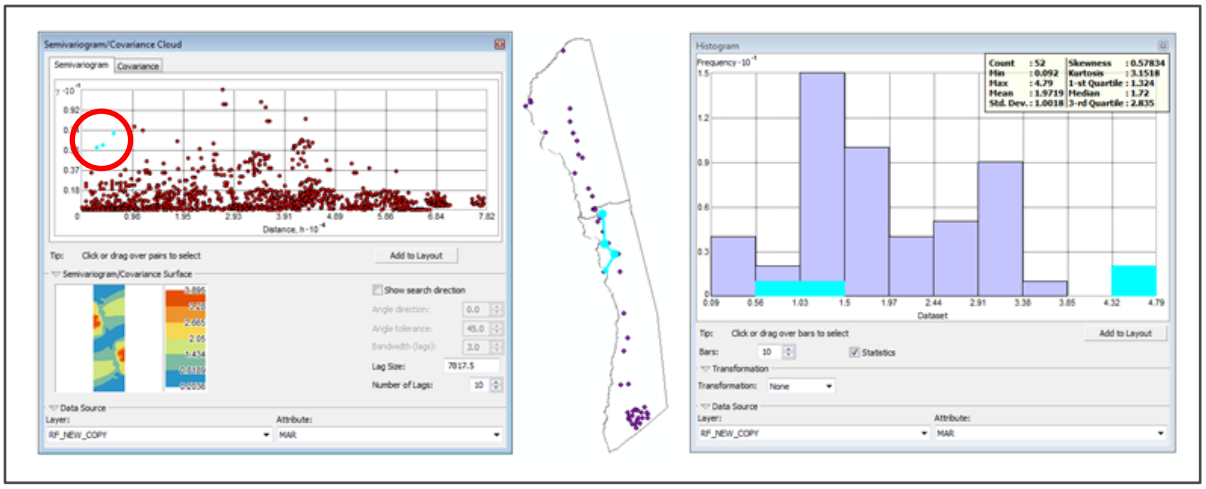


Figure B.3: Semivariogram Cloud, stations, and Histogram for March.

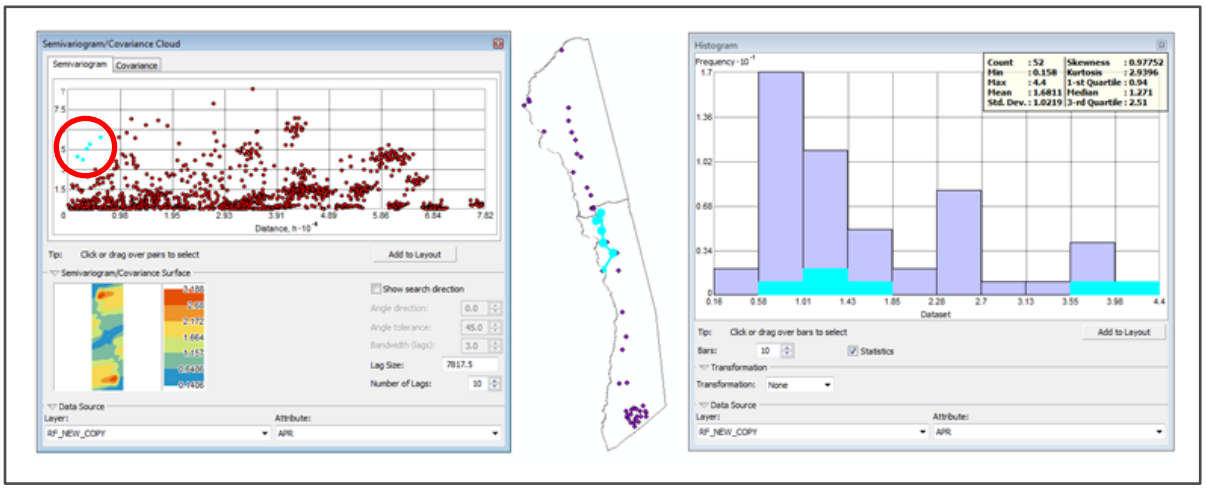


Figure B.4: Semivariogram Cloud, stations, and Histogram for April.

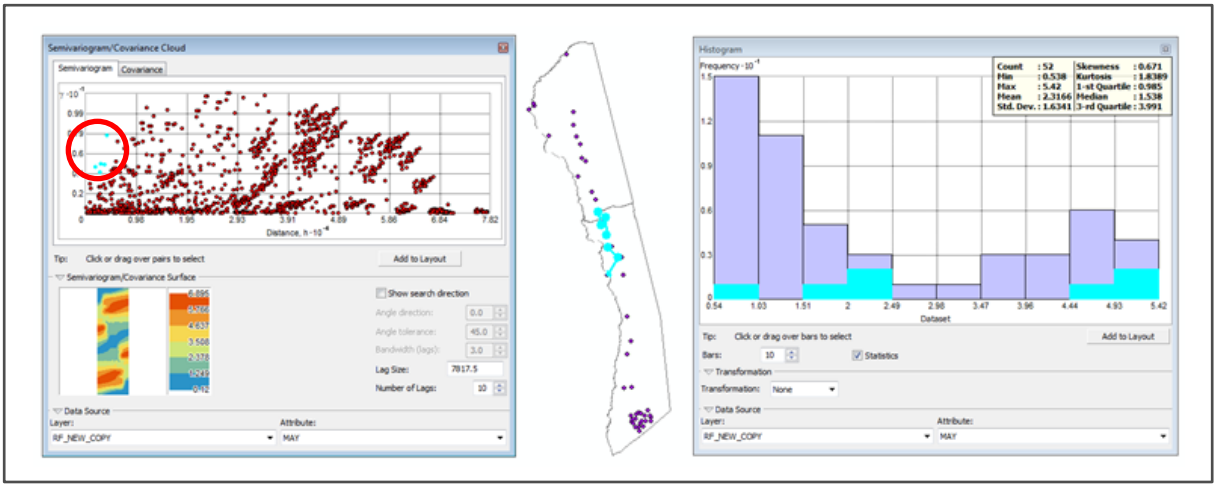


Figure B.5: Semivariogram Cloud, stations, and Histogram for May.

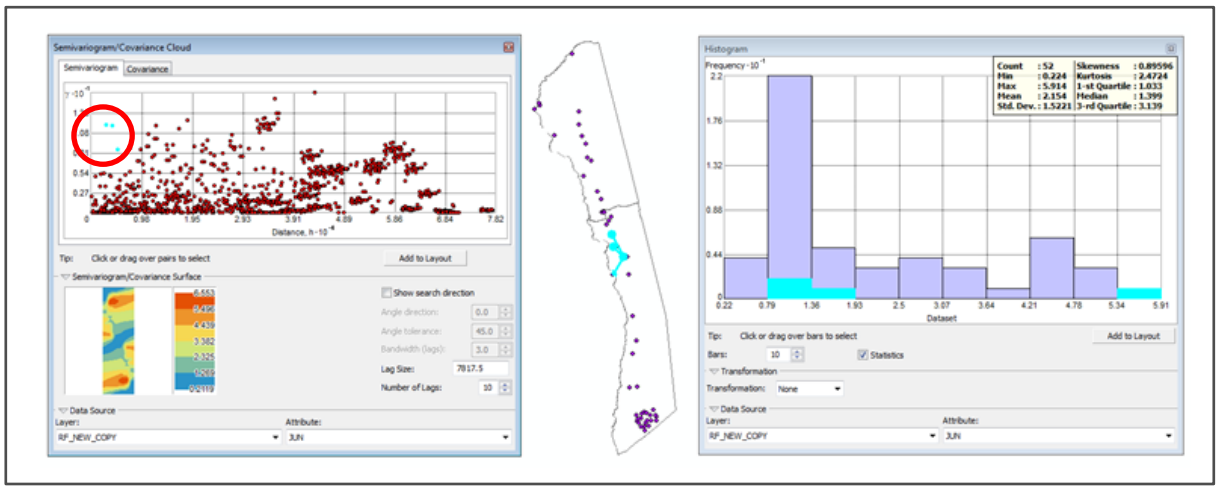


Figure B.6: Semivariogram Cloud, stations, and Histogram for June.

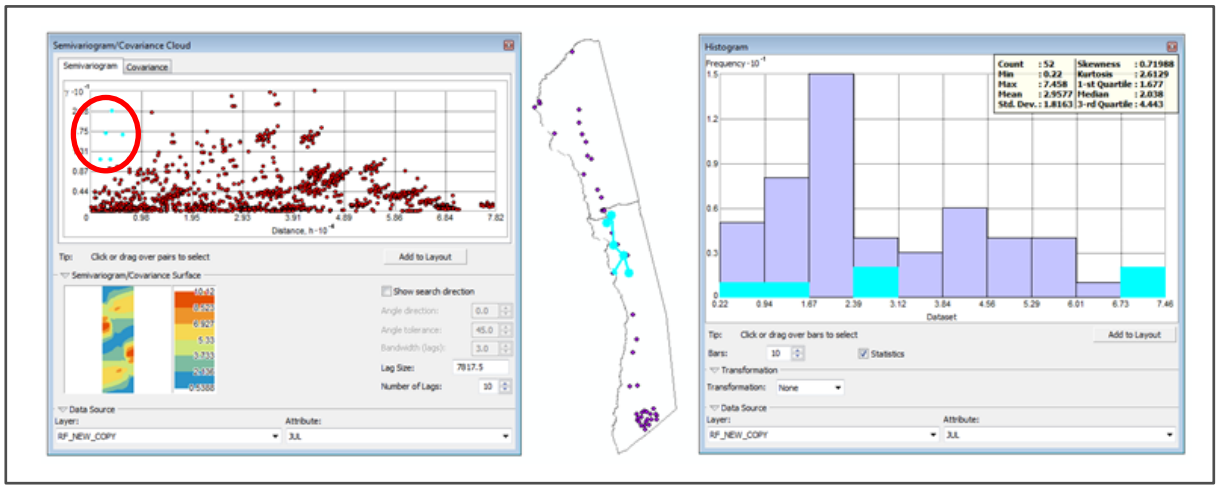


Figure B.7: Semivariogram Cloud, stations, and Histogram for July.

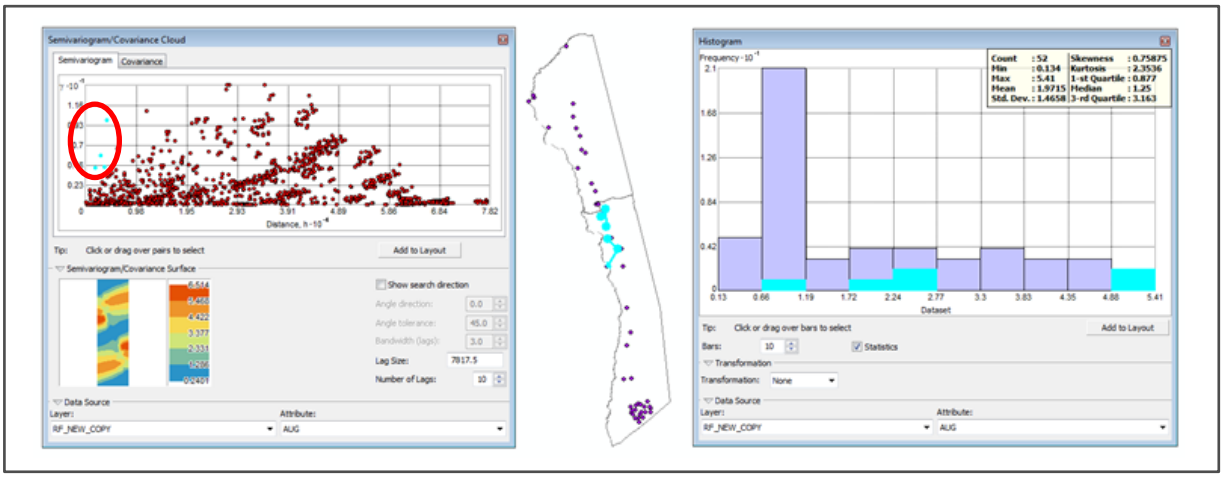


Figure B.8: Semivariogram Cloud, stations, and Histogram for August.

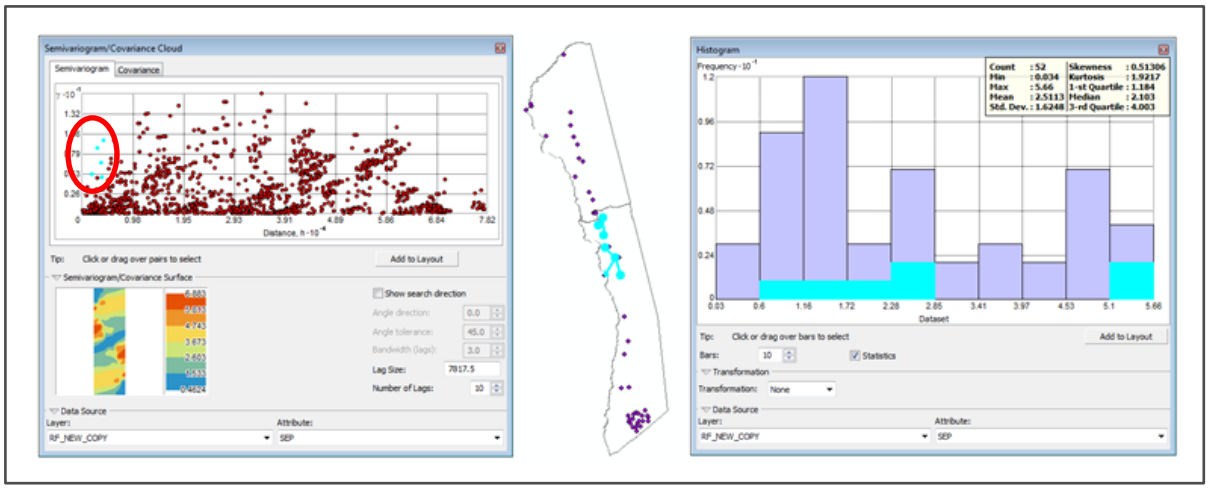


Figure B.9: Semivariogram Cloud, stations, and Histogram for September.

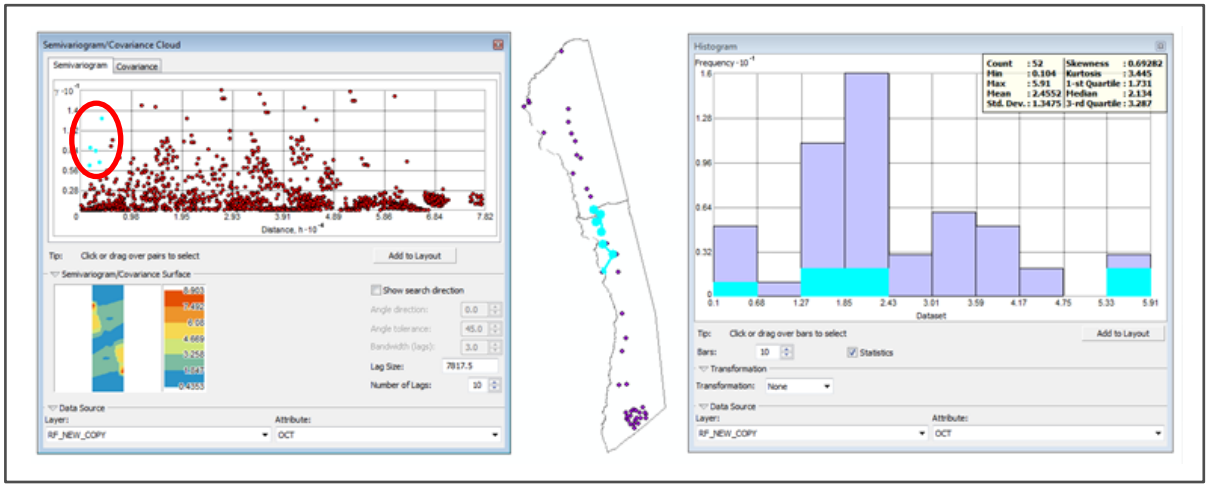


Figure B.10: Semivariogram Cloud, stations, and Histogram for October.

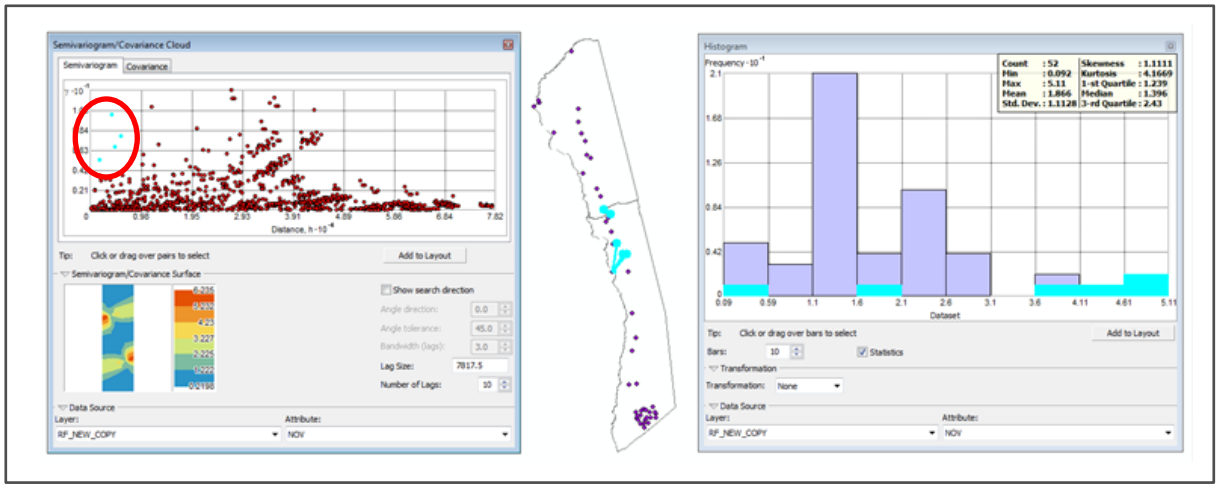


Figure B.11: Semivariogram Cloud, stations, and Histogram November.

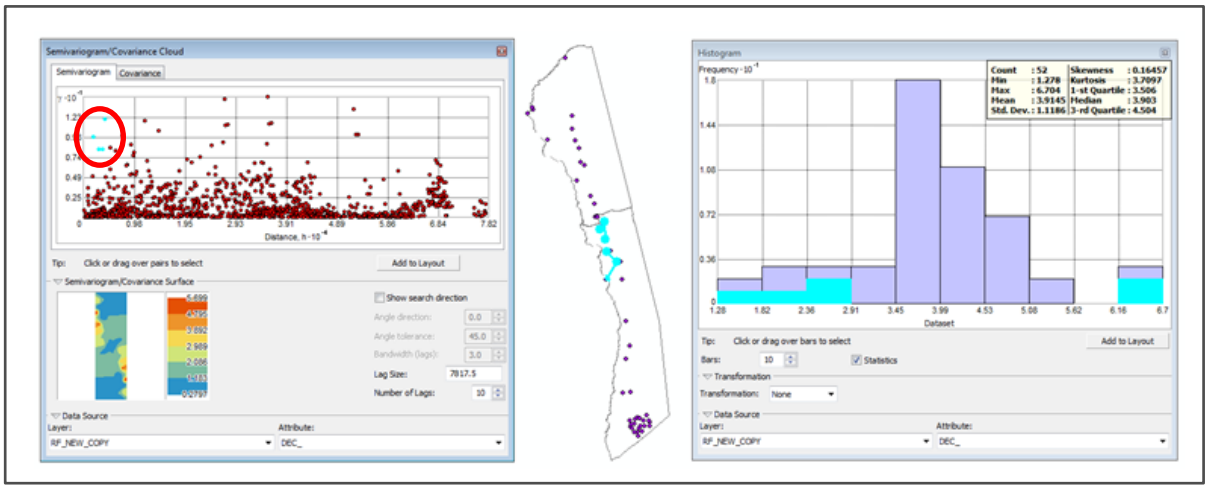
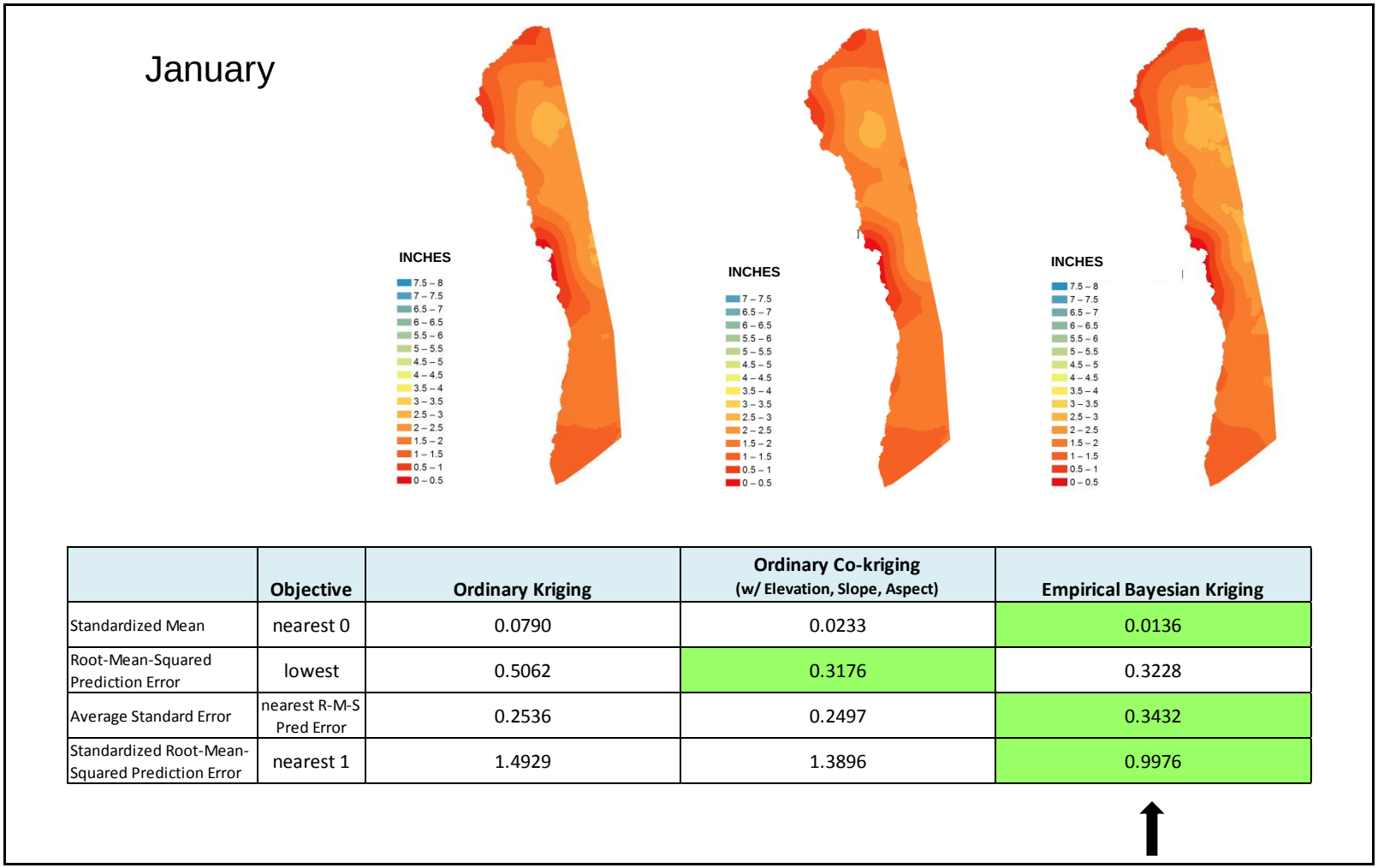


Figure B.12: Semivariogram Cloud, stations, and Histogram for December.

Appendix C

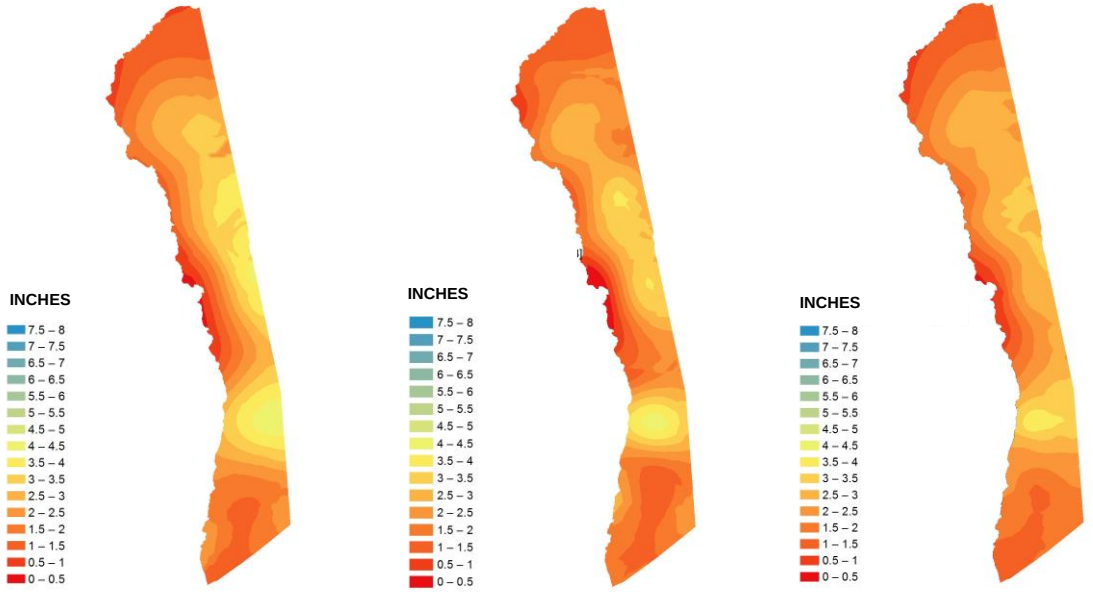
Comparison of rainfall surfaces generated for kriging methods and corresponding prediction error statistics.



	Objective	Ordinary Kriging	Ordinary Co-kriging (w/ Elevation, Slope, Aspect)	Empirical Bayesian Kriging
Standardized Mean	nearest 0	0.0790	0.0233	0.0136
Root-Mean-Squared Prediction Error	lowest	0.5062	0.3176	0.3228
Average Standard Error	nearest R-M-S Pred Error	0.2536	0.2497	0.3432
Standardized Root-Mean-Squared Prediction Error	nearest 1	1.4929	1.3896	0.9976

Figure C.1: Comparison of surfaces and prediction error statistics for January.

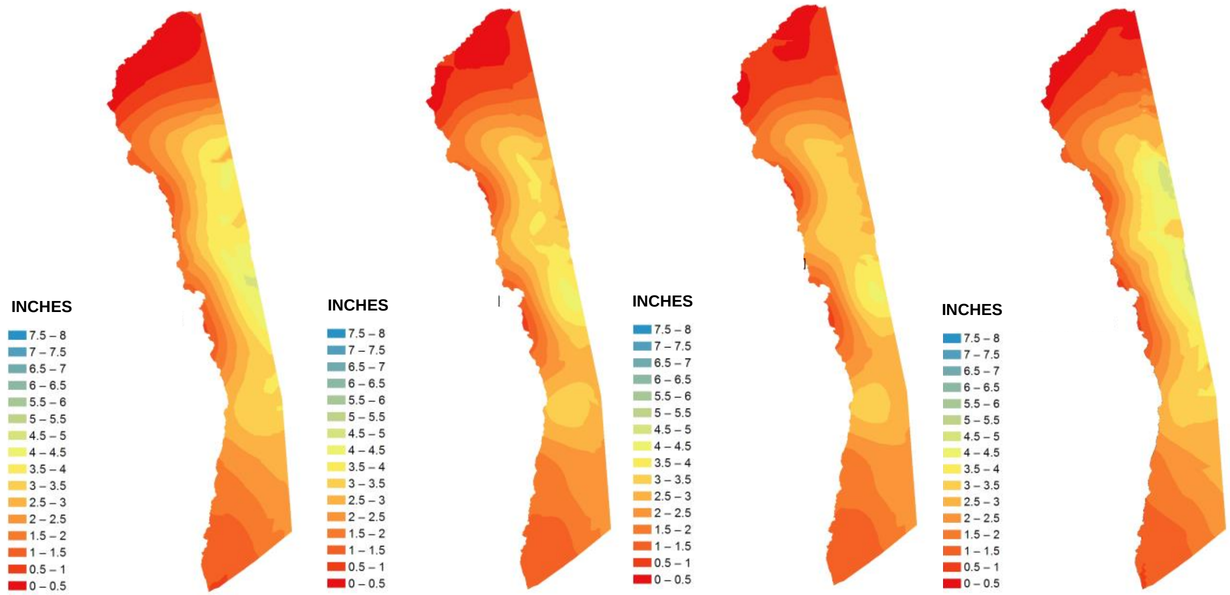
February



	Objective	Ordinary Kriging	Ordinary Co-kriging (w/ Elevation, Slope, Aspect)	Empirical Bayesian Kriging
Standardized Mean	nearest 0	0.0281	0.1205	0.0321
Root-Mean-Squared Prediction Error	lowest	0.4440	0.4726	0.4392
Average Standard Error	nearest R-M-S Pred Error	0.2544	0.2914	0.3943
Standardized Root-Mean-Squared Prediction Error	nearest 1	1.8607	1.6745	1.0809

Figure C.2: Comparison of surfaces and prediction error statistics for February.

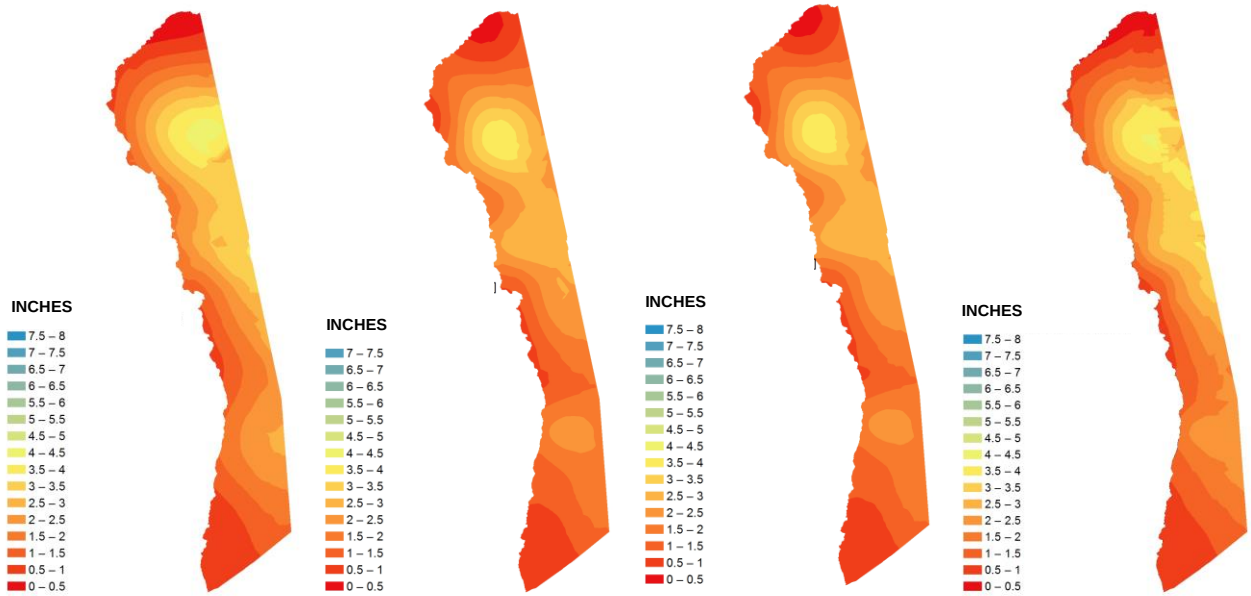
March



	Objective	Ordinary Kriging	Ordinary Co-kriging (w/ Aspect)	Ordinary Co-kriging (w/ Elevation, Slope, Aspect)	Empirical Bayesian Kriging
Standardized Mean	nearest 0	0.0259	0.0542	0.0534	0.0204
Root-Mean-Squared Prediction Error	lowest	0.4439	0.4597	0.4672	0.4751
Average Standard Error	nearest R-M-S Pred Error	0.3491	0.3522	0.3751	0.4946
Standardized Root-Mean-Squared Prediction Error	nearest 1	1.3782	1.4298	1.3534	1.0671

Figure C.3: Comparison of surfaces and prediction error statistics for March.

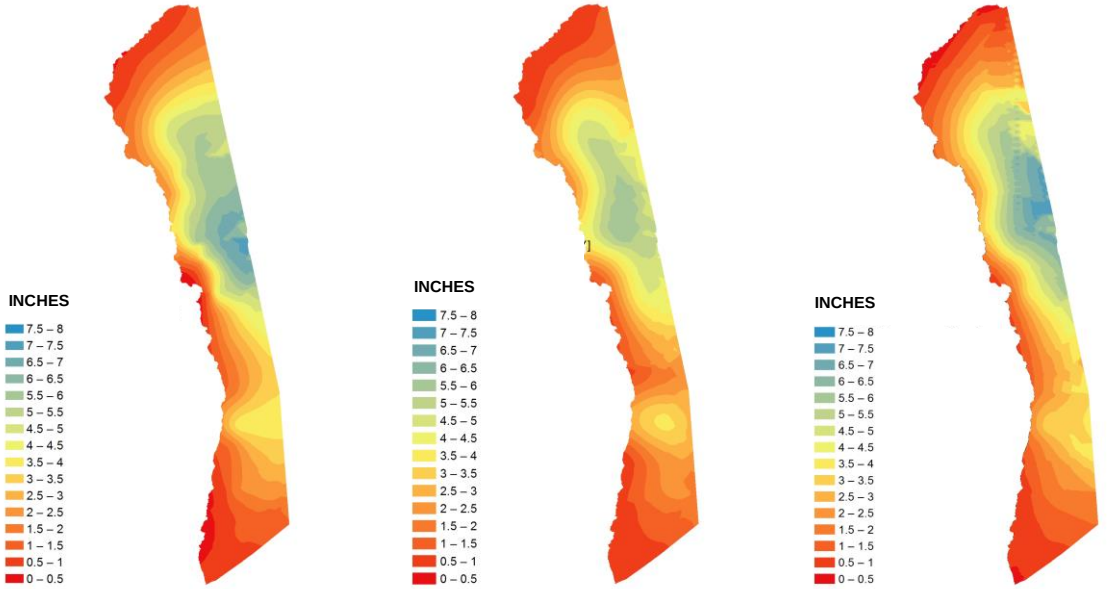
April



	Objective	Ordinary Kriging	Ordinary Co-kriging (w/ Elevation, Slope)	Ordinary Co-kriging (w/ Elevation, Slope, Aspect)	Empirical Bayesian Kriging
Standardized Mean	nearest 0	0.0362	0.0963	0.0950	0.0550
Root-Mean-Squared Prediction Error	lowest	0.4356	0.4580	0.4579	0.3925
Average Standard Error	nearest R-M-S Pred Error	0.3005	0.3409	0.3471	0.3966
Standardized Root-Mean-Squared Prediction Error	nearest 1	1.2944	1.1105	1.0910	1.0006

Figure C.4: Comparison of surfaces and prediction error statistics for April.

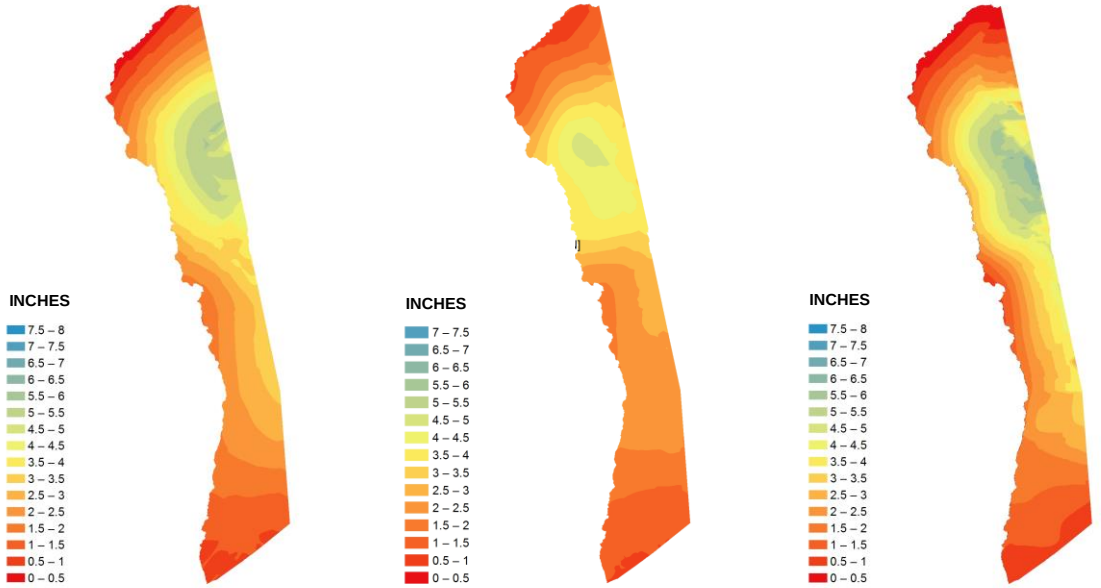
May



	Objective	Ordinary Kriging	Ordinary Co-kriging (w/ Elevation, Slope, Aspect)	Empirical Bayesian Kriging
Standardized Mean	nearest 0	0.0751	0.0846	0.0384
Root-Mean-Squared Prediction Error	lowest	0.7772	0.7889	0.6819
Average Standard Error	nearest R-M-S Pred Error	0.4661	0.5581	0.6885
Standardized Root-Mean-Squared Prediction Error	nearest 1	2.5698	1.3397	1.0800

Figure C.5: Comparison of surfaces and prediction error statistics for May.

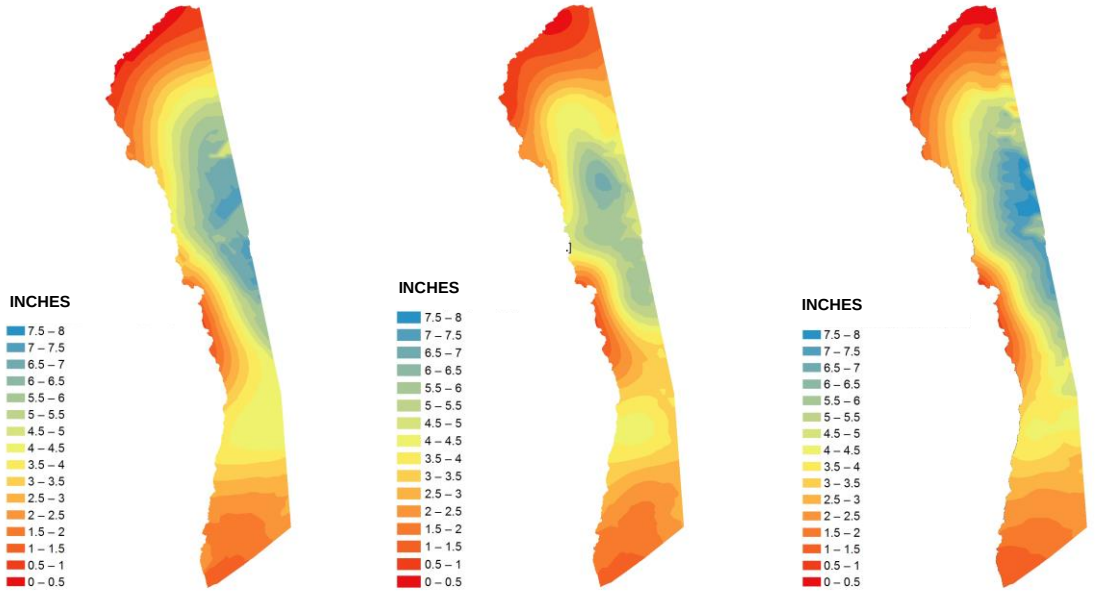
June



	Objective	Ordinary Kriging	Ordinary Co-kriging (w/ Elevation, Slope, Aspect)	Empirical Bayesian Kriging
Standardized Mean	nearest 0	0.0278	0.0496	0.0168
Root-Mean-Squared Prediction Error	lowest	0.6008	0.6766	0.5430
Average Standard Error	nearest R-M-S Pred Error	0.5199	0.6326	0.6111
Standardized Root-Mean-Squared Prediction Error	nearest 1	1.1274	1.0009	1.0062

Figure C.6: Comparison of surfaces and prediction error statistics for June.

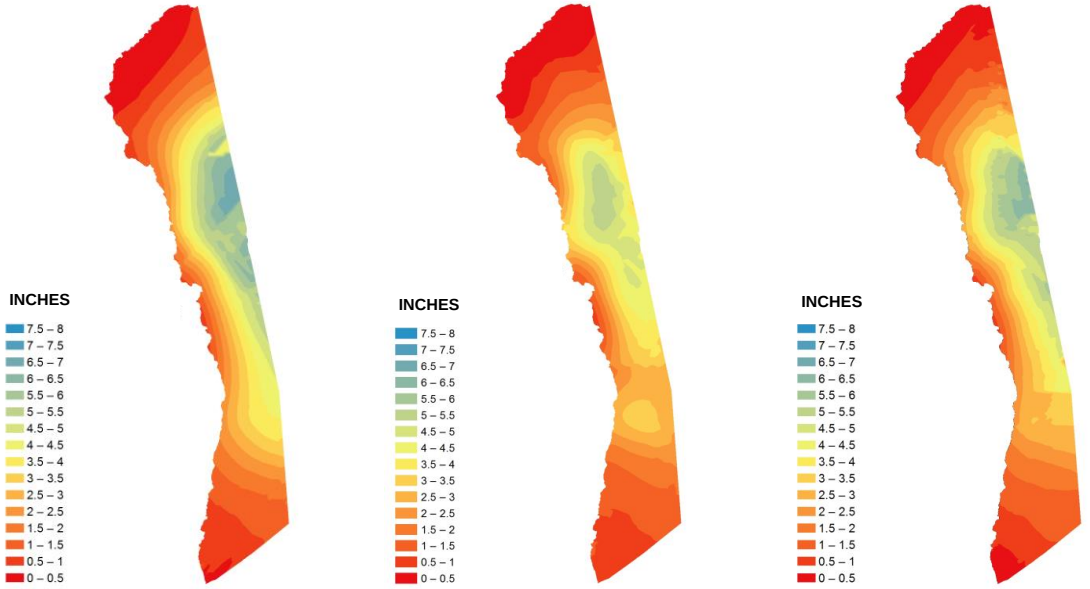
July



	Objective	Ordinary Kriging	Ordinary Co-kriging (w/ Elevation, Slope, Aspect)	Empirical Bayesian Kriging
Standardized Mean	nearest 0	0.0199	0.0626	0.0251
Root-Mean-Squared Prediction Error	lowest	0.7190	0.7657	0.6830
Average Standard Error	nearest R-M-S Pred Error	0.5625	0.6520	0.7707
Standardized Root-Mean-Squared Prediction Error	nearest 1	1.3419	1.2026	1.0185

Figure C.7: Comparison of surfaces and prediction error statistics for July.

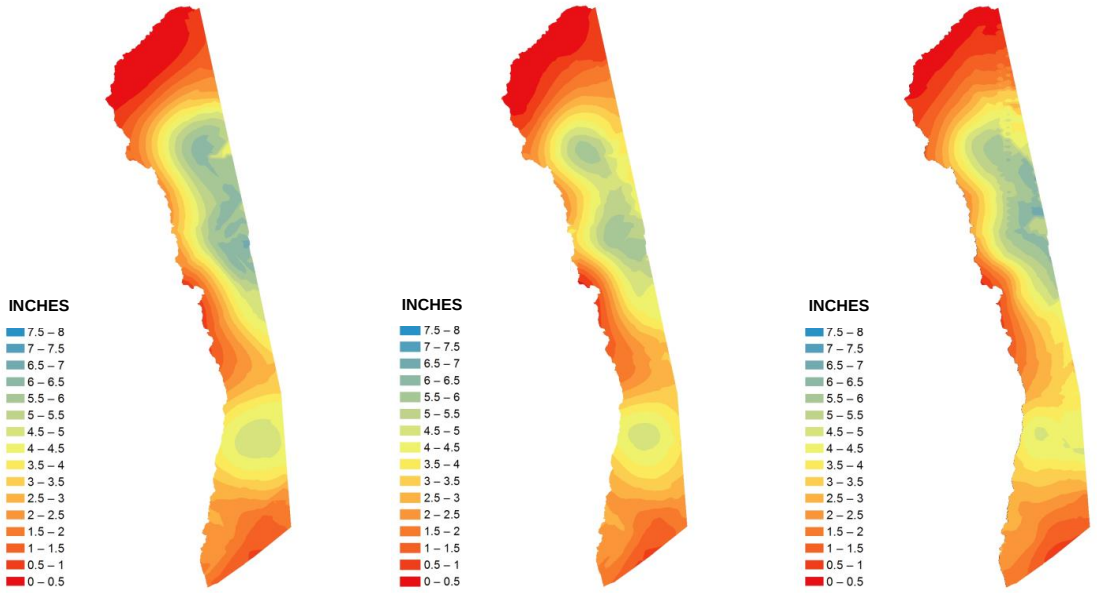
August



	Objective	Ordinary Kriging	Ordinary Co-kriging (w/ Elevation, Slope, Aspect)	Empirical Bayesian Kriging
Standardized Mean	nearest 0	0.0196	0.0526	0.0354
Root-Mean-Squared Prediction Error	lowest	0.4756	0.5412	0.4973
Average Standard Error	nearest R-M-S Pred Error	0.3654	0.4499	0.5227
Standardized Root-Mean-Squared Prediction Error	nearest 1	1.4626	1.2507	1.0449

Figure C.8: Comparison of surfaces and prediction error statistics for August.

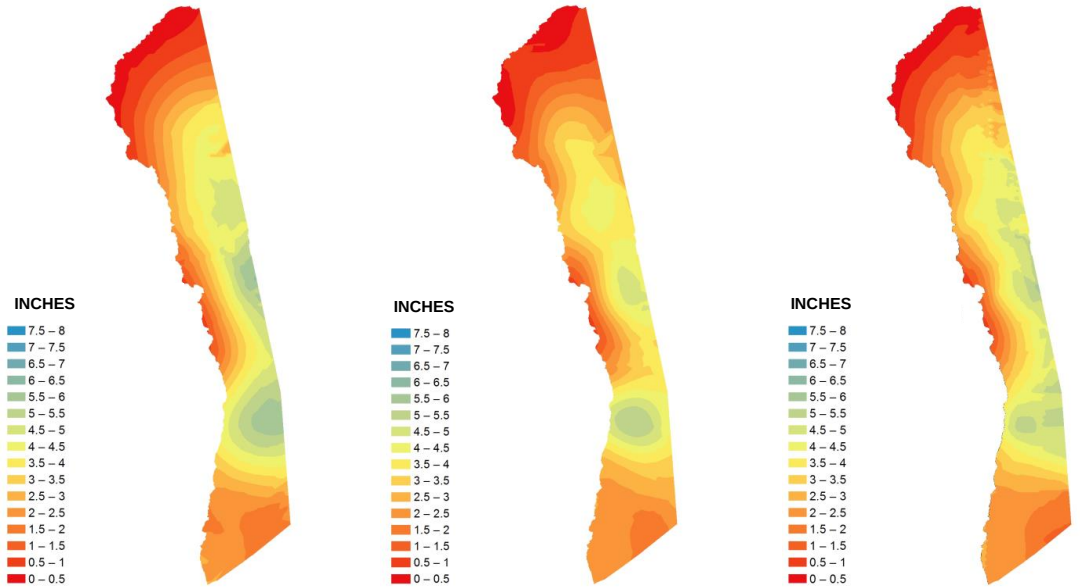
September



	Objective	Ordinary Kriging	Ordinary Co-kriging (w/ Elevation, Slope, Aspect)	Empirical Bayesian Kriging
Standardized Mean	nearest 0	0.0280	0.0678	0.0376
Root-Mean-Squared Prediction Error	lowest	0.6288	0.6702	0.6857
Average Standard Error	nearest R-M-S Pred Error	0.6000	0.6577	0.7784
Standardized Root-Mean-Squared Prediction Error	nearest 1	1.1876	1.1236	1.0134

Figure C.9: Comparison of surfaces and prediction error statistics for September.

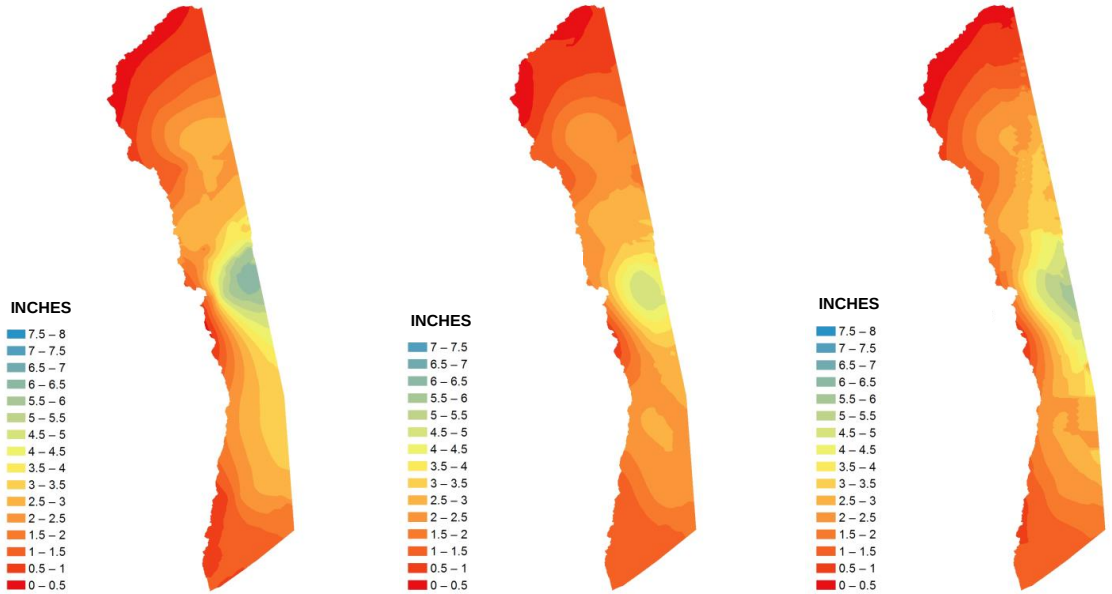
October



	Objective	Ordinary Kriging	Ordinary Co-kriging (w/ Elevation, Slope, Aspect)	Empirical Bayesian Kriging
Standardized Mean	nearest 0	0.0378	0.0664	0.0251
Root-Mean-Squared Prediction Error	lowest	0.7345	0.8338	0.7750
Average Standard Error	nearest R-M-S Pred Error	0.5290	0.6193	0.7820
Standardized Root-Mean-Squared Prediction Error	nearest 1	1.4945	1.3744	1.0629

Figure C.10: Comparison of surfaces and prediction error statistics for October.

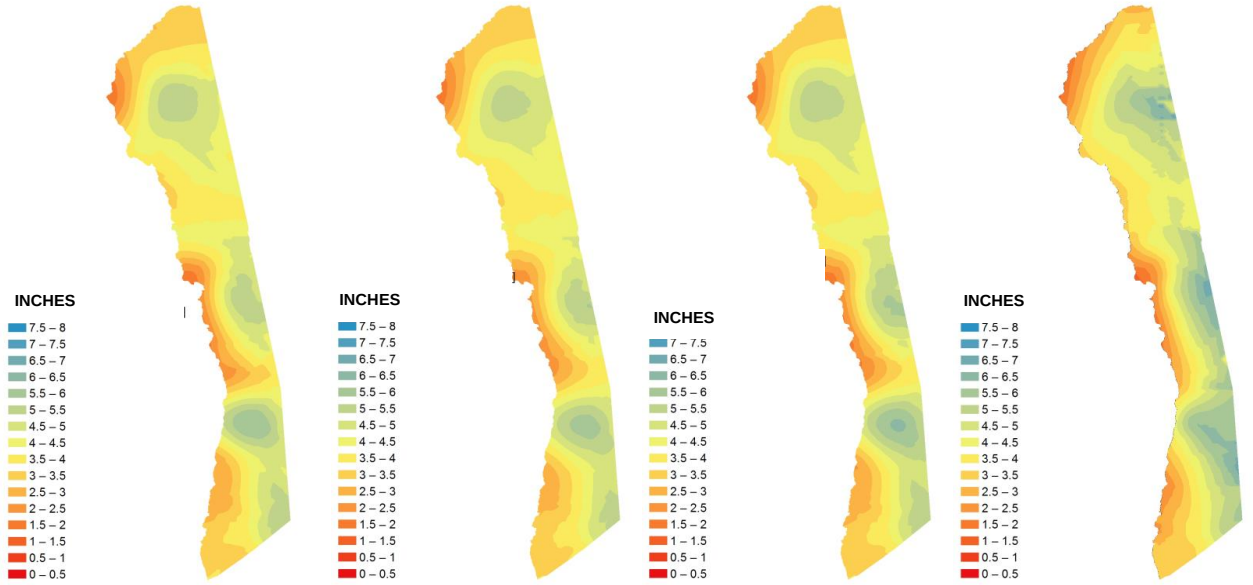
November



	Objective	Ordinary Kriging	Ordinary Co-kriging (w/ Elevation, Slope, Aspect)	Empirical Bayesian Kriging
Standardized Mean	nearest 0	-0.0292	0.0509	0.0074
Root-Mean-Squared Prediction Error	lowest	0.4503	0.4600	0.4361
Average Standard Error	nearest R-M-S Pred Error	0.3041	0.3718	0.5011
Standardized Root-Mean-Squared Prediction Error	nearest 1	2.7768	1.2862	0.9992

Figure C.11: Comparison of surfaces and prediction error statistics for November.

December



	Objective	Ordinary Kriging	Ordinary Co-kriging (w/ Elevation, Slope)	Ordinary Co-kriging (w/ Elevation, Slope, Aspect)	Empirical Bayesian Kriging
Standardized Mean	nearest 0	0.0494	0.0358	0.0418	0.0172
Root-Mean-Squared Prediction Error	lowest	0.8702	0.8615	0.8761	0.7828
Average Standard Error	nearest R-M-S Pred Error	0.5613	0.5784	0.5533	0.7408
Standardized Root-Mean-Squared Prediction Error	nearest 1	1.4442	1.3754	1.4813	1.0807

Figure C.12: Comparison of surfaces and prediction error statistics for December.