

A 30-YEAR AGROCLIMATIC ANALYSIS OF THE SNAKE RIVER
VALLEY AMERICAN VITICULTURAL AREA - DESCRIPTIVE
AND PREDICTIVE METHODS.

by
Charles L. Becker



A thesis
submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Hydrologic Science
Boise State University

August 2020

© 2020

Charles L. Becker

ALL RIGHTS RESERVED

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Charles L. Becker

Thesis Title: A 30-YEAR AGROCLIMATIC ANALYSIS OF THE SNAKE
RIVER VALLEY AMERICAN VITICULTURAL AREA - DE-
SCRIPTIVE AND PREDICTIVE METHODS.

Date of Final Oral Examination: 29 April 2020

The following individuals read and discussed the thesis submitted by student Charles L. Becker, and they evaluated the student's presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Alejandro Flores Ph.D. Chair, Supervisory Committee

David Wilkins Ph.D. Member, Supervisory Committee

Nancy Glenn Ph.D. Member, Supervisory Committee

The final reading approval of the thesis was granted by Alejandro Flores Ph.D., Chair of the Supervisory Committee. The thesis was approved by the Graduate College.

DEDICATION

Dedicated to my late father, Terry J. Becker.

ACKNOWLEDGMENT

I'd like to thank each and every member of the BSU Geoscience department who collectively enhanced my research and provided both academic and non-academic support in a truly enjoyable environment, as well as all my family and friends who encouraged me throughout this journey. Lastly, a special thanks to my personal mentor Jim F. and girlfriend Jenna P. for their frequent and continued support.

ABSTRACT

Climate change poses serious threats to global agriculture, however some localities and crops may benefit from increasing temperatures. Grape production in southern Idaho may be a beneficial example as vineyard acreage has increased over 300% since the designation of the Snake River American Viticultural Area (SRVAVA) in 2007. We perform a statistical characterization of agroclimate within the SRVAVA that centers around four primary objectives: utilization of a novel, 30-year high resolution climate dataset to provide insight and agrometrics unavailable at coarser resolutions, climatic implications of the unique topography within the SRVAVA, identification of statistical trends, and correlation of SRVAVA climate to large-scale climate indicators such as the El Nino Southern Oscillation (ENSO). In Chapter 3 we build on the identified correlations to large scale climate and utilize a long short-term memory (LSTM) model in conjunction with empirical mode decomposition (EMD) to create a novel, data driven method to forecast regional temperature trends with lead times up to one year. Favorable results for local viticulture include an increase in growing degree days and season length, as well as reduced frequency of freezing events. Possible disadvantages include increased risk to shoulder season freezing events with warmer winters, increased magnitude of strong freezing events, mid-season heat stress, and higher susceptibility to powdery mildew outbreaks. Additionally, with strong correlations identified with large-scale climate indicators, we find EMD an effective method

to increase modeling power by using multiple frequencies of the signals as input into a LSTM machine learning algorithm that can accurately predict temperature trends up to one year in advance. This climatic characterization and modeling framework could potentially inform many agricultural management decisions such as cultivar choice, vineyard site selection, fungicide spray timing, irrigation strategy, and canopy management.

TABLE OF CONTENTS

DEDICATION	iv
ACKNOWLEDGMENT	v
ABSTRACT	vi
LIST OF FIGURES	xi
LIST OF TABLES	xv
LIST OF ABBREVIATIONS	xvi
1 AGRICULTURE AND VITICULTURE IN THE SNAKE RIVER VALLEY OF IDAHO	1
1.1 Motivation	1
1.2 Idaho Agriculture	2
1.2.1 Snake River Valley American Viticultural Area	4
1.2.2 General Climate and Geology	5
1.3 Grape Suitability	6
1.3.1 Climatic Controls	7
1.3.2 Topographic Controls	11
1.3.3 Soil Controls	13

1.4	Summary	15
2	AGROCLIMATE ANALYSIS OF THE SNAKE RIVER VALLEY AVA	16
2.1	Introduction	16
2.2	Utility of High Resolution Data	17
2.3	Weather Research and Forecasting Model	18
2.4	Data	19
2.4.1	Data Subsets	21
2.5	General Characterization of Climate in the Snake River Valley AVA	23
2.6	Agroclimate Analysis	30
2.6.1	Growing Degree Days (24-hour method)	30
2.6.2	Frost Hours	33
2.6.3	Frost Degree Days	35
2.6.4	Powdery Mildew Index	35
2.7	Climate Trends	39
2.7.1	Linear Trend Estimations	39
2.7.2	Interannual Variability and Correlations to Large Scale Climate	46
2.8	Summary	50
3	FORECASTING REGIONAL CLIMATE FROM LARGE SCALE CLIMATE: A DATA DRIVEN APPROACH	53
3.1	Introduction	53
3.2	Large-scale Climate	55
3.2.1	El Nino Southern Oscillation	55
3.2.2	Pacific Decadal Oscillation	57

3.2.3	Climate of the Snake River Valley	59
3.3	Properties and Relationships of Climate Signals	59
3.4	Methodology	64
3.4.1	Long Short Term Memory Model Overview	64
3.4.2	Empirical Mode Decomposition	68
3.4.3	Data Preparation	70
3.4.4	Model Framework and Results	75
3.5	Discussion and Future Work	77
4	CONCLUDING SUMMARY	80
	REFERENCES	83
	APPENDICES	90
A	DESIGN OF AN INTERACTIVE WEB TOOL	91
A.1	Introduction	92
A.2	Platform	92
A.3	Data	93
A.4	Functionality	93
A.5	User Experience	96
A.6	Summary	98

LIST OF FIGURES

1.1	Geographic boundary of the Snake River Valley American Viticultural Area and Sunnyslope Wine trail (red). Established in 2007.	4
1.2	Approximate mean temperature requirements for <i>vitis vinifera</i> varieties (Jones, 2003).	10
2.1	Geographic boundary of WRF domain setup with 3km (outer) and 1km (inner) extents.	20
2.2	Mean minimum and maximum daily temperatures against elevation in the Snake River Valley AVA.	24
2.3	Number of extreme cold temperature events in the SRVAVA (1988 - 2017).	26
2.4	Probability density functions of frost timing and variance in the SRVAVA (1988 - 2017).	27
2.5	Mean temperature distribution in the SRVAVA (1988 - 2017).	29
2.6	Diurnal temperature variation in the SRVAVA (1988 - 2017).	29
2.7	Spaghetti plot of cumulative aggregated precipitation in the SRVAVA (1988-2017).	31
2.8	Differences in seasonal growing degree days estimating mean daily temperature in two various ways (TMIN/TMAX only, vs 24 hourly observations) in the SRVAVA (1988-2017).	32

2.9	Mean and standard deviation of growing degree days in the SRVAVA (1988-2017).	33
2.10	Cumulative mean aggregated freezing hours in SRVAVA (1988-2017).	34
2.11	15-year split of frost degree days against frost hours. Magnitude of freeze events is stronger per hour in most recent time period.	36
2.12	15-year differences of FROSTH and FDD. An increase of FDD, indicating an increase in the magnitude of freezing events, is noted in the lower valley where many vineyards are located.	36
2.13	Mean powdery mildew risk index. Majority of current vineyards located within the sunny slope region denoted by the black bounding box.	38
2.14	Mean temperature changes from October 1987 - September 2017. . .	40
2.15	Monthly mean temperature changes from October 1987 - September 2017. Error bars represent one standard deviation comprised of 1000 Monte Carlo Samples.	40
2.16	Cumulative distribution functions of frost timing and season length. .	43
2.17	Frost timing by year.	43
2.18	Evidence of similar low-frequency patterns in timing of both first and last frosts suggesting strong relationship to large scale climate.	47
2.19	Temperature anomalies through space and time.	48
2.20	Distribution of > 99% anomalous temperatures. Relative high anomalies are much more likely in the springtime. Clear low frequency signals appear in the annual distribution.	49

2.21	Distribution of $< 1\%$ anomalous temperatures. Relative TMIN low anomalies are much more likely in the wintertime.	49
3.1	Generalized Walker Circulation pattern in different ENSO states. (Di Liberto, 2014)	56
3.2	Historical ENSO and PDO events. Mild/moderate events shown in light grey with extreme events in dark grey. Raw signal represented by points.	58
3.3	Unfiltered autocorrelations. PDO has significantly longer memory than ENSO and regional temperature.	60
3.4	Unfiltered cross-correlations. The Southern Oscillation has a strong influence on PDO.	61
3.5	Unlagged relationship between teleconnection states and regional temperature.	63
3.6	Unlagged relationship between PDO and regional minimum temperatures. Lagged correlations are inconsistent.	63
3.7	Basic ANN Structure.	65
3.8	Basic RNN Structure.	66
3.9	LSTM Cell State.	67
3.10	Basic LSTM Structure.	67
3.11	Snowmelt dynamics represented by the cell state value of an LSTM network (Kratzert <i>et al.</i> , 2018)	68
3.12	Process to generate IMFs through EMD.	70
3.13	Intrinsic Mode Functions of SOI and PDO.	72
3.14	Example of additive decomposition regional temperature.	73

3.15	Conceptual data preprocessing flow chart.	74
3.16	SRVAVA Temperature trend forecasts with varying lead times in two separate scenarios.	76
A.1	Spatial Explorer UI.	94
A.2	Time Series UI.	95
A.3	Data can be downloaded directly from any location.	96

LIST OF TABLES

2.1	Variable subset of raw WRF output.	21
2.2	Derived variables from raw WRF output.	22
2.3	Spatially aggregated Statistical summary of each metric within the SRVAVA from October 1987 through September 2017. Linear trends calculated with annually aggregated data according to method.	42
2.4	2-sample Kolmogorov-Smirnov and Mann-Whitmann statistical test results for mean spatial distributions over the first and last 15-year periods.	45

LIST OF ABBREVIATIONS

AVA American Viticultural Area

EMD empirical mode decomposition

ENSO El Nino Southern Oscillation

GDD growing degree days

KS Kolmogorov-Smirnov

LSTM long short-term memory

MW Mann-Whitney

PDO Pacific Decadal Oscillation

PMI powdery mildew index

SLP sea-level pressure

SOI Southern Oscillation Index

SRVAVA Snake River Valley American Viticultural Area

SST sea-surface temperature

WRF Weather Research and Forecasting model

CHAPTER 1: AGRICULTURE AND VITICULTURE IN THE SNAKE RIVER VALLEY OF IDAHO

1.1 Motivation

The vulnerability of agricultural commodities is driven primarily by both climate and weather. Climate, at the regional scale, informs the general suitability of an agricultural zone. Weather events are often associated with agricultural risk. Though relationships between climate and agriculture are obvious, they are complex, and the intrinsically chaotic nature of the total climate system poses great challenges to the entire agricultural industry (Liang *et al.*, 2017). With increasing evidence of the impacts of climate change on the agricultural sector and availability of novel high-resolution datasets, there is an increasing need for examining historical climate and forecasting future climate in ways that agriculture can benefit at local scales.

Climate change has both direct (hazardous climate and weather phenomena) and indirect (increased pressure from pests and pathogens) impacts on agricultural production (Walthall CL, 2013). Current literature suggests that without adaptation strategies, these impacts will negatively impact future agricultural production on large scales. Examples include a reduction of yield in maize, rice, and wheat in both tropical and temperate regions (Challinor *et al.*, 2014); reduction of agricultural land

due to sea-level rise (Gornall *et al.*, 2010); and a potential reduction of U.S. total factor production up to 4% (Liang *et al.*, 2017). However, benefits to agriculture stemming from a changing climate also exist in such forms as increased growing season length (Jones *et al.*, 2010), increase of photosynthetic rate of C3 crops as a result of higher CO2 concentrations (Gornall *et al.*, 2010), reduced number of freezing events (Mirade Orduña, 2010), and potential expansion of suitability of certain crops (Sacchelli *et al.*, 2017). Many of these impacts are crop specific and it is beneficial to take a detailed view of climate with respect to a particular agricultural sector and region. For this study, we primarily examine climate through the lens of viticulture - the growing of grapes for wine production - and constrain our analysis to the Snake River Valley American Viticultural Area (SRVAVA) region in Southern Idaho. Although this study provides analysis through the perspective of grape growing, much of the analysis will be useful to other specialty crops or Idaho agriculture as a whole.

The remainder of this chapter provides an overview of viticulture and the conditions that control its successful production. Chapter 2 statistically characterizes climate within SRVAVA over 30 years using a high resolution dataset generated locally at Boise State University. Chapter 3 leverages the trends found within this data to produce a data driven seasonal forecasting model that may be useful as a tool for viticultural decision making.

1.2 Idaho Agriculture

Idaho has a long history of productive agriculture. Idaho agribusiness currently ranks in the top 10 for 26 separate crops in annual production (United States Department of Agriculture National Agricultural Statistics Service, 2017), accounts for 26.4 billion dollars in sales, and provides over 120,000 local jobs, which is approximately 1

in every 8 jobs within the state (Watson, 2019). When combined with food processing, this accounts for approximately 20% of Idaho's total economic output (Koong *et al.*, 2018). Potatoes, barley, sugarbeets, and Austrian winter peas are some of the top Idaho exports but other specialty crops such as grapes and hops have expanded substantially in recent years. There's a total of 11.7 million acres of land within Idaho utilized for agriculture and this number is expected to grow in coming decades (United States Department of Agriculture National Agricultural Statistics Service, 2017).

Idaho Viticulture

Viticulture, the production of grapes (*vitis vinifera*) for the purpose of making wine is considered to be one of the fastest growing sub-segments of the agricultural industry in Idaho, though it only accounts for a small percentage of current planted acreage (Idaho Wine Commission, 2018). The first known grapes planted in Idaho were in the Northwestern city of Lewiston in 1864 (Idaho Wine Commission, 2020). However, the passing of prohibition in the 1920s put a halt to productive vineyards that did not see any new grapes planted until the year 1970. In 2007, Idaho designated its first American Viticultural Area (AVA) and was quickly followed by two smaller Idaho AVA designations. An AVA can be described as a legally designated grape growing region for the mutual benefit of growers and consumers (Alcohol and Tobacco Tax and Trade Bureau, 2020). It was the designation of the Snake River Valley AVA (SRVAVA) in 2007 that sparked a boom in new Idaho vineyards and wineries which began to give Idaho a national name for the production of quality wines. As of 2017, Idaho has 60 wineries, over 1300 hundred acres of grapes planted, and produced 2942

tons of grapes with an estimated economic impact of 209.6 million dollars (Idaho Wine Commission, 2018; Watson, 2019). This is nearly a 300% increase from an economic impact of 73 million just 10 years prior.

1.2.1 Snake River Valley American Viticultural Area

The SRVAVA region encompasses 21,400 km² and is one of the largest AVAs in the U.S. It is unique in that it is the only AVA that extent is geologically bounded - by an elevation of 1050 meters above sea level - the approximate extent of Pliocene Lake Idaho (Gillerman *et al.*, 2006). Its range covers much of southern and southwestern Idaho and partially extends into eastern Oregon (Figure 1.1). As the vast majority of grapes, and agriculture in general, are planted within the bounds of the SRVAVA, our analysis will be constrained to the spatial extent of the SRVAVA.



Figure 1.1: Geographic boundary of the Snake River Valley American Viticultural Area and Sunnyslope Wine trail (red). Established in 2007.

1.2.2 General Climate and Geology

There are many geologic and climate factors that comprise a desirable agricultural zone, some general, and some specific to any given crop. Broadly, Idaho experiences a semi-arid climate with hot days and cool nights, low precipitation (yet adequate surface and groundwater), and well drained soils. Geologic variability that influences agricultural suitability may include topography (slope and aspect), as well as soil type, depth, and fertility.

Using the Koppen climate classification, the SRVAVA is primarily classified as a Bsk - a cold semi-arid (steppe) climate. This climate type is generally characterized by low amounts of precipitation, warm-to-hot summers, and cold winters. They are generally isolated from large bodies of water and exhibit large diurnal temperature variations. Nearly all of these characteristics are ideal for grapes as cold winters and low precipitation discourage fungal and bacterial pathogens, hot summers significantly increase photosynthetic potential, and large daily temperature swings provide balance between acidity and sugar levels important for the quality of wine produced (Jackson, 2014). Although many of the general characteristics of the Snake River Valley are favorable for viticulture, there exists much more variance in regional weather than other notable wine producing regions such as Napa and Willamette Valleys. This larger variance in regional weather poses a greater risk for unhealthy conditions in vineyards. Examples include major freeze events in the shoulder and dormant seasons, extreme heat in the summer, hail damage, and flash flooding. These risks are of major concern as short term hazardous weather events can have effects on production that can span several years. The frequency and magnitude of such events, as well as long term trends are thoroughly examined in the following chapter.

The climate was not always arid in the Snake River valley, as much of the region was previously underwater. The large body of water, Lake Idaho, was drained through the western edge of the valley through Hells Canyon, as it's known today, approximately two to three million years ago. But perhaps the most notorious event to shape agriculture in SRVAVA was the Bonneville flood. Just after the last glacial maximum 14,500 years ago, Lake Bonneville - a massive body of water in central/northern Utah covering the present-day Salt Lake basin - breached Red Rock Pass in Southern Idaho. The breach sent a massive amount of water towards the Snake River which ultimately led to the breaking of multiple natural dams. At its peak, a 125 meter wall of water, with an estimated volumetric flow of 930,000 m³/s at over 110 km/hour, scoured the valley and deposited the well drained, sandy-silt topsoils we see today (U.S. Geological Survey, 2002).

1.3 Grape Suitability

Broadly, the three primary factors of crop suitability can be attributed to climate, topography, and soil. Furthermore, a French concept known as *terroir*, is used to further evaluate these controls. Terroir is broadly defined as the combination of all environmental factors that affect a crop's phenotype (Gillerman *et al.*, 2006). The primary factors considered in terroir are similarly climate, soil, and cultivar. However, any factor that contributes to a physiological change in a crop can also be a factor, including anthropogenic influences such as irrigation methodology, cultivar selection, or any other vineyard management decision. The concept of terroir is more commonly implemented for crops used to make secondary products that are more scrutinized for quality, and more specifically, flavor. Examples include grapes for wine-making, coffee beans, and tobacco. American Viticultural Areas work in parallel with the terroir

concept, that often have similar geologic and climate terroir profiles at medium-to-large scales. An understanding of each of these factors are the main decision driving metrics for determining ideal site locations.

1.3.1 Climatic Controls

Although Chapter 2 provides a historical analysis of climate within the Snake River AVA, it is useful to first provide some initial context of the importance of each phenomena.

Solar Radiation

Solar insolation provides the primary energy needed for grape growth and maturation through photosynthesis. The most notable variability will again be exaggerated based on distance from the equator, with the photosynthetic period (hours of sunlight) becoming increasingly non-uniform. Despite being the primary energy input, the large variance of insolation throughout the growing season at high latitudes seems to have a relatively minor effect compared to other metrics such as mean temperature. It is quite possible that these effects would become much more meaningful at latitudes above 50 degrees, although temperatures at these latitudes do not currently support viticulture.

Temperature

Temperature is widely considered the primary metric for evaluating suitability for *vitis vinifera*. There are numerous metrics derived from temperature important for viticulture that extend beyond a simple mean monthly temperature, for example. And once again, we generally see higher variance of temperature expressions in higher

latitudes. This is compounded with the effects of regional weather patterns that are prone to develop in the mid-latitudes where the majority of vineyards are located. The most notable temperature derived metric used for suitability analysis is the heat accumulation unit known as the degree day. Other temperature-based metrics of interest are day of first and last frost (or inversely frost-free days) , diurnal variation, and the frequencies and magnitudes of minimum and maximum temperature. The combination of these temperature expressions have a wide range of impacts.

Degree days, more commonly known as growing degree days (GDD), is a heat summation unit developed for phenology that is used as a heuristic for determining basic crop suitability (Jackson, 2014). It is widely used by farmers to predict phenological development rates and timing. For grapes, this could include: bud burst, bloom, fruit set, cluster closure, véraison and fruit maturity (harvest). As different crops have different phenological demands, a specific reference temperature is utilized for specific crops. The reference temperature, also known as the base temperature, is the minimum temperature required for growth. For grapes, a warm weather crop, the base temperature is set to 10 °C. In addition to the minimum temperature, some crops also make use of a maximum reference temperature which represents the temperature at which increased photosynthesis ceases (30 °C for grapes).

An example of the calculation of GDD uses the mean daily temperature and subtracts the reference temperature of with the following constraints: you cannot have a negative GDD value and any daily mean above 30 degrees Celsius is set to 30 degrees. This is broadly represented as

$$GDD = \int (T_{(t)} - T_{(base)}) dt \quad (1.1)$$

However, this is usually approximated as

$$GDD = \frac{T_{max} + T_{min}}{2} - T_{base} \quad (1.2)$$

with T_{max} and T_{min} representing the daily maximum and minimum temperatures, and the reference (base) temperature being set to 10 °C. GDD is generally only summed over the growing season (typically April 1st through October 31st for viticulture). This is not only used as a general suitability index for a crop, but can also be useful in differentiating suitability for varietal differences, many of which have substantial differences. Figure 1.2 shows the approximate temperature requirements for a variety of *vitis vinifera* species.

The magnitude and duration of minimum or maximum temperatures will have adverse effects on *vitis vinifera*. Minimum temperatures that cause frost or plant tissue freezing are generally of more concern due to their higher impact and larger frequency. However, extreme high temperatures are known to burn both leaves and fruit (Hayman *et al.*, 2012). Any temperature below freezing is considered harmful, however, the timing of such events is of extreme importance. Depending on the annual growth stage, events of the same magnitude can have drastically different effects. A minor, harmless frost event in the midst of the dormant season may have no effect at all, whereas the same event just after bud break could significantly alter the seasonal yield. One primary concern with climate change is an early deacclimation of cold weather followed by early season frost events (Meier *et al.*, 2018). Additionally, there can be dormant season cold events of such magnitude that can completely overcome plant acclimation and do significant damage. January of 2017 was an example of this, when a large scale arctic cold front moved into the valley and brought

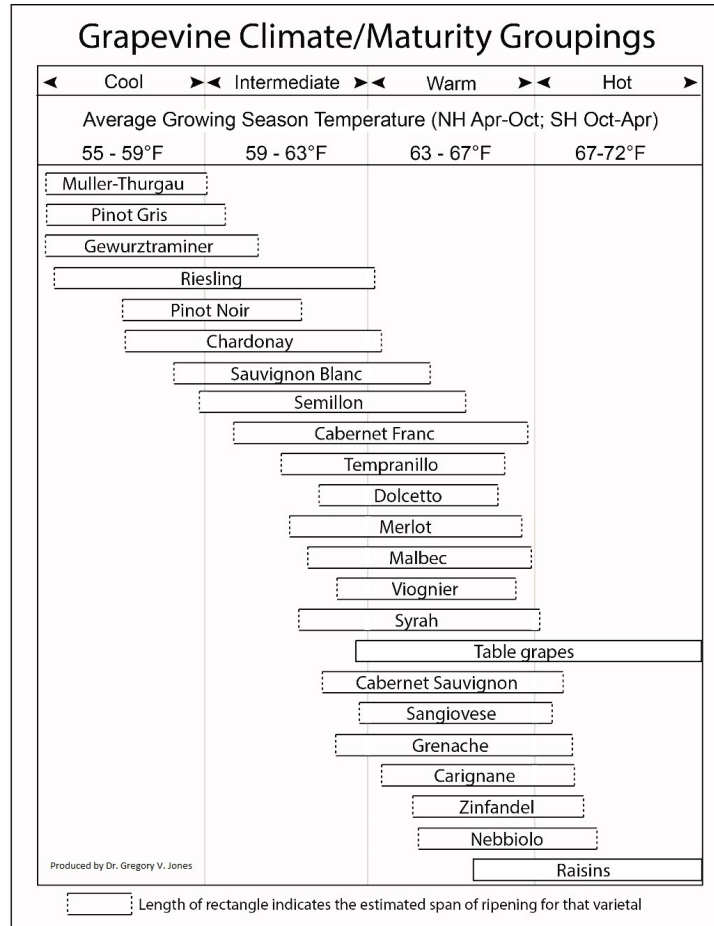


Figure 1.2: Approximate mean temperature requirements for *vitis vinifera* varietals (Jones, 2003).

temperatures in some vineyards down to nearly -30 °F. Many vines died and required replanting. However, some vineyards with slightly higher elevations and those with snow cover (insulating the roots from the extreme temperature) suffered much less damage. Events such as these can have disastrous economic impacts as cutting (trimming only the damaged part of the vine) or replanting have production effects that span years.

Precipitation and Humidity

Excessive moisture and precipitation is generally a more serious issue for grapes than conditions that are too arid, especially when irrigation is available. Excess water in the growing season can lead to abundant canopy growth which can lead to oversized berries lacking quality flavor. Additionally, excess precipitation can lead to higher humidity which greatly enhances the risk of fungal pathogens such as powdery mildew. This is one reason there are very few vineyards near the east coast of the U.S. despite adequate seasonal temperatures. Although, during the off-season in areas that are at risk for extreme cold events, large snow events can help insulate the trunk and roots during major freeze events. Quick draining soils also help mitigate humidity and excess available water. Where irrigation is utilized (all vineyards within the SRVAVA are irrigated), vines are often “stressed” by limiting the water availability. This will restrict shoot growth and berry maturation which can lead to smaller more desirable fruit (Jackson, 2014). However, too little water can affect the vine negatively by leaving the berries vulnerable to sun damage or worse not having enough water to deliver nutritional content. Lastly, strong thunderstorms that produce hail can cause significant damage to vineyards and farms alike. However, the spatial distribution of hail is sparse and most large wine regions don’t have a climate structure that is conducive to regularly generating large hail producing storms.

1.3.2 Topographic Controls

Topography influences vine and grape production primarily through controlled radiative input, frost severity, and water drainage. Potential disadvantages may include lack of uniformity across a site, snow cover reduction or a premature winter

deacclimation, and soil erosion.

The two most important attributes of topography are slope (degree vertical incline) and aspect (compass direction of the slope). The most significant metric derived directly from topographic orientation is radiative input. This is especially true for higher (or lower in the southern hemisphere) latitudes as the seasonal variance of solar radiation is directly proportional to latitude. Equatorial facing slopes are exposed to much more solar input than their poleward counterparts. In topographically complex regions, or even individual sites, slope and aspect can have an extreme impact on varietal suitability of the area.

Reduced frost severity from topography is primarily driven by two factors: increased radiation from equatorial facing slopes and pathways for cold air drainage. In the former, the additional hours, or even minutes, of direct solar input can shorten the duration of freezing conditions harmful to the fruit or vine. Though an exception to this may be the additional melting of a layer of snow on the ground that was formerly insulating the trunk and roots, and could thus expose them to harsher winter conditions. Cold air drainage is mostly driven by slope and by planting row orientation. The lack of solar heating after sunset aids in allowing dense, cool air to sink and settle in ground level valleys and troughs. If this air mass has no mechanism to be transported out of the zone, it can have devastating effects depending on the magnitude of temperature and exposure time. Thus, an unobstructed, gravitational outlet is needed for effective transport of lethal cold air pools. The higher the magnitude of the slope (5-10 degrees is common), the more velocity (and thus momentum) the sinking air has to be transported. Thus, it is preferable to have an outer boundary of a vineyard as the lowest elevation which would operate as the exit point for cold

pools. It is also important to have the exit path be free of obstacles. To achieve this, rows are often planted more parallel to the aspect, as perpendicular (to the aspect) orientations could have the vine canopy itself blocking the transport of cool air.

Proper water drainage is critical for keeping a proper amount of water in the soil that is available to the roots. If there is too much water available to the plant, an excess of energy is sent to the canopy rather than the fruit which can yield undesirable results. In viticulture, it's common practice to "stress" the vines by limiting their water availability. A sloped site provides a gravity fed drainage of excess soil water similar to cold air drainage. This is increasingly important in regions that have shallow hardpans, prolonged precipitation events, or high intensity precipitation events such as thunderstorms.

Lastly, it is worth mentioning altitude when discussing topography. Altitude plays a significant role, however, it is primarily expressed through climate factors such as temperature that are discussed independently. There are potentially some biological effects of the lower air pressure at high altitude, but those impacts are minor and outside the scope of this thesis.

1.3.3 Soil Controls

Of geologic and climatic influences, soil characteristics are one of the least important factors influencing the quality of grapes (Maltman, 2008; Wahl, 1988). Soil attributes affecting vine and fruit growth include texture composition, depth, pH, drainage (permeability), and nutrient availability. These attribute influences can be generally expressed as heat retention, water availability, and nutritional status. For example, different texture compositions and color can have a different planetary albedo which influences the soil heat retention. The amount of retained heat in the soil

then has effects on the local microclimate with regard to frost protection. Although soil characteristics may be considered less important than atmospheric conditions, they still have direct impacts. Consistent conditions, or soil uniformity, across a vineyard may be more important than the overall characteristics themselves (Jackson, 2014).

Soil textures, in agricultural soil, are broadly classified by the relative volume fractions of sand, silt and clay. The texture component directly influences all three primary characteristics: heat, water availability and nutritional status. Heavier soils (clay dominated) are known to have a lower albedo and retain more heat, hold much more water due to a lower permeability, and have a higher capacity to hold nutrients. The opposite is generally true for lighter (sand dominated) soils. Although clay retains significantly more water than sand, it is often unavailable to plants as the surface area to volume ratio is large and the water bonding is too strong for the plant root to overcome. For this reason, silt is often considered the best texture for plants to thrive in as it balances water retention and availability. However, grapes are known to thrive in lighter soils with light water retention as the improved water drainage promotes heat radiation from the soil protecting from frost. Additionally, high permeability leads to less berry fissures which has an increased risk from fungal and bacterial pathogens. Lastly, limited soil fertility is known to be highly beneficial for quality grape production. Soils with restricted (yet adequate) nutrients restrict vegetative growth and direct more sugars (photosynthates) to the fruit (Jackson, 2014).

In addition to soil texture, soil depth has a direct effect on water availability of roots. Hardpans close to the surface have an inherent risk of water logging, despite

soil texture, if large amounts of water are delivered to the ground. Natural soil compaction over time also creates problems by increasing soil density to levels that water and roots have a difficult time moving through. A common practice to deal with this problem is soil “ripping” - mechanically breaking up compacted soil and/or shallow hardpans - prior to planting new vines.

1.4 Summary

With the establishment of the SRVAVA in 2007, Idaho has experienced enormous growth in the viticultural sector. Its success is primarily controlled through factors expressed through soil, topography, and climate. Current literature points to many negative implications of climate change on future agriculture and food production. However, viticulture in the Snake River Valley could potentially benefit from current climate trends due to reduced risk of cold-damage and longer growing seasons.

CHAPTER 2:

AGROCLIMATE ANALYSIS OF THE SNAKE RIVER VALLEY AVA

2.1 Introduction

A climatological normal is a concept in applied climatology defined as a 30-year period of weather that serves as a comparison to other time periods that is both long enough to smooth out internal variability (when taking a mean) and short enough to see trends within that variability (WMO Climatological Normals, 2020). However, there is a general trade-off between resolution (spatially and temporally) and the length of time periods of publicly available climate products. As a result, most gridded climate products that span 30 years or more have coarse resolutions that do not capture important phenomena impactful to agriculture, and specifically, viticulture. This is especially important in topographically complex regions such as the Snake River Valley AVA, where large variance in temperature and other relevant metrics exist across very small extents. We address this issue by utilizing a novel 30-year, high-resolution regional climate dataset generated by the Weather Research and Forecasting model (WRF) that has the advantages of resolving phenomena at smaller scales and is long enough to statistically characterize trends through time.

This chapter focuses on seeking agriculturally relevant insight from the internal

variability of climate in the Snake River Valley AVA from October 1987 through September 2017. The analysis primarily highlights four distinct areas: creating novel agroclimate metrics and insights as a direct result of higher resolution data, the unique geologic layout of the Snake River AVA and the resulting climatic influences, identifying agroclimate trends and impacts, and correlating interannual variability as a result of large scale climate phenomena.

2.2 Utility of High Resolution Data

The chaotic nature of the atmosphere and its interactions and feedbacks with the land surface can change quickly. These rapid changes in the atmosphere can result in many important events that can resolve over the course of 24 hours and be unseen at coarse temporal resolutions. Examples include: convective thunderstorms producing hail and flash floods, sea and land breezes, propagation of squall lines, and cold air transport and/or pooling. Temperatures are particularly relevant for viticulture, and hourly data provides a much better picture of the *duration* of these events, which is often as important as the extremes. There are of course flows and small scale air masses that resolve within sub-hourly time scales, but the magnitude and duration of temperature swings within an hour (yet not seen with hourly resolution) are unlikely to significantly affect health and production of a grapevine. Thus, hourly data appears to be an appropriate resolution to characterize temperature driven events important to viticulture.

Furthermore, hourly data allows the derivation of agroclimate metrics otherwise not possible. Examples include metrics to better characterize the magnitude, frequency and duration of frost and freezing temperatures - the primary threat to vineyards in the SRVAVA - as well as other hazard indicators such as the Powdery Mildew

Risk Index that specifically requires hourly input. Additionally, other standard metrics such as mean temperature and growing degree days can be improved on with hourly data. For example, growing degree days are most frequently calculated with the minimum and maximum daily temperatures as opposed to a mean calculated with all 24 daily observations.

Spatially, many climate products are often statistically downscaled to a finer resolution with a variety of interpolation methods. However, these methods can be highly inaccurate and inappropriate if there are land surface complexities (such as topography) between interpolation points or if there are a wide variety of terrain types across a domain. A spatial resolution of 1 km generated directly from a physically based model such as WRF, allows for a much finer representation of the complexities of a land surface and their resulting influence on the climate. It can accurately capture large variances in data over small spatial scales important to crops, such as cold pools and differences in radiative input due to terrain slope. As such, high spatially resolved data can more accurately denote suitable agricultural locations and the associated risks within a topographically complex region such as the SRVAVA.

2.3 Weather Research and Forecasting Model

The Weather Research and Forecasting (WRF) model is an open-source numerical weather prediction model used for both operational forecasting, historical hindcasting, and research (Skamarock *et al.*, 2008). Most global models operate on large grids ($> 0.25^\circ$) which don't resolve complex interactions between the atmosphere and the land surface, specifically complex topography, that are important to agriculture. A primary advantage to using WRF is its ability to dynamically downscale data from coarse to fine resolutions in both space and time using a higher resolution grid of the

land surface. Coarser climate data is used as lateral boundary conditions in a limited area domain which is downscaled through a series of nested domains that generally resolve equations of motion at smaller and smaller scales. The availability and usage of long-term reanalysis data for input to the WRF model allows for a limited area, long-term, climate reconstruction with resolutions that provide insight not available at typical climate resolutions.

Other advantages of WRF specific data often not available in other climate datasets include precipitation partitioning (differentiating between liquid and frozen precipitation), vertical atmospheric profiles (up to 50 layers), and additional variables such as humidity, wind speed/direction, and snow parameters (depth, cover, SWE), all contained within a single dataset. Furthermore, other difficult to measure agroclimate metrics such as evapotranspiration (ET) can be estimated and derived from WRF output (using radiation, soil moisture, specific humidity, wind, and temperature). The primary disadvantage of running large climate simulations with WRF is the high computational cost and the storage requirements. However, rapidly evolving cloud-computing frameworks such as Amazon Web Services are helping mitigate these challenges.

2.4 Data

A full regional climatological dataset was generated with WRF by the Lab for Ecohydrology and Applied Forecasting (LEAF) laboratory at Boise State University (Flores *et al.*, 2016), and spans water years 1988 - 2017. The model was forced with the North American Regional Reanalysis (NARR), a gridded data assimilation product with a 32 km spatial and 3-hour temporal resolution produced by the National Centers for Environmental Prediction (NCEP) (Kalnay et al., 1996). The WRF setup

featured a nested domain with a 3 km outer domain and an inner domain of 1 km. Temporal resolution was downscaled from 3-hours to 1-hour throughout the entire domain. Each water year (October 1st - September 31st) was run as a separate simulation with a model spin-up period of two weeks. Extents for each domain are illustrated in figure 2.1.

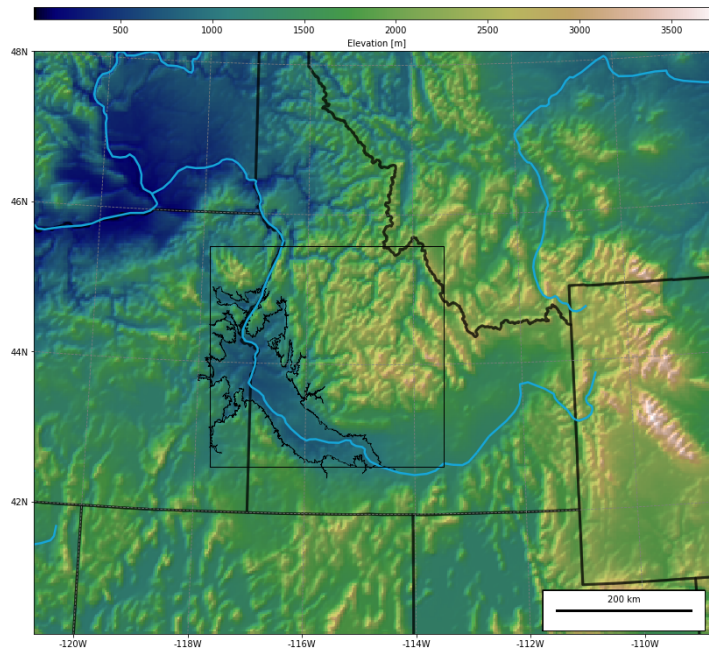


Figure 2.1: Geographic boundary of WRF domain setup with 3km (outer) and 1km (inner) extents.

Our primary area of study, the Snake River Valley AVA, is nearly entirely contained within the inner, highest resolution (1km) domain. Only the very western tip of the AVA overreaches the inner domain boundary. With over 98% of the AVA area contained within the inner domain, we choose to only use data from the inner domain and exclude the small portion not lying within the domain. To our knowledge, this is the highest resolution (both temporally and spatially) climate dataset in the region and provides the ability to conduct novel data analyses that are useful in an

agricultural context.

2.4.1 Data Subsets

WRF output files contain over 260 variables, some of which are 4-dimensional (x, y, z, time), resulting in output of daily files, each 13 GB in size. However, this application requires analysis of only a few of these variables. After specific variables were subset, the data was resampled from hourly to daily data to further the size reduction and make the data more accessible. Table 2.1 shows the subset of raw WRF output variables of interest.

Table 2.1: Variable subset of raw WRF output.

Raw WRF Variable	Description
T2	2-Meter Temperature
Q2	2-Meter absolute humidity
SWDOWN	Downward shortwave radiation
SWNORM	Slope-dependent downward shortwave radiation
RAINNC	Accumulated total precipitation (since beginning of simulation)
I.RAINNC	Bucket for accumulated precipitation
SNOWNC	Accumulated total snowfall

For each 24-hour period, a sum, mean, or other functional transformation was applied to the raw output to derive other useful metrics. In addition to some basic transformations of raw WRF output, there are four unique variables that we derived directly as a result of having hourly resolution. (1) 24-hour method of calculating

Table 2.2: Derived variables from raw WRF output.

Derived Variable	Description	Calculation
TMIN	Minimum 2-meter temperature	Minimum T2 of 24 daily values
TMAX	Maximum 2-Meter temperature	Maximum T2 of 24 daily values
TMEAN	Mean 2-meter temperature	Mean T2 of 24 daily values
TSPREAD	Mean Diurnal temperature variation	$TMAX - TMIN$
PRCP	Hourly total precipitation accumulation	$(RAINNC + I_RAINNC)_t - (RAINNC + I_RAINNC)_{t-1}$
SNOW	Hourly snowfall accumulation	$SNOWNC_t - SNOWNC_{t-1}$
GDD	Growing degree days	$\frac{TMIN+TMAX}{2} - 10^\circ$
GDD24	Growing degree days (24-hour method)	$TMEAN - 10^\circ$
FROSTH	Frost hours	Number of T2 output below 0° per day
FDD	Frost degree days	Sum of $(0^\circ - T2)$ for all T2 values $< 0^\circ$ per day
PMI	Powdery mildew risk index	See section 2.6.4

growing degree days (GDD24), growing degree days calculated from a mean daily temperature of all hourly data points (as opposed to a mean calculated from the minimum and maximum temperatures). (2) Frost hours (FROSTH), number of hours below 0°C . (3) Frost degree days (FDD), an index we created to help combine the

magnitude and duration of freezing temperatures. (4) Powdery mildew index (PMI), an index created by the University of California Davis to help control the outbreak of powdery mildew in vineyards. Each new variable and the relationship between the derived quantities and the raw output is shown in table 2.2. Analysis of each is discussed in the section 2.5.

Lastly, the entire new dataset was masked with a shape file of the Snake River Valley AVA. Thus, any spatial means taken in the subsequent analysis will always represent the AVA as a whole.

2.5 General Characterization of Climate in the Snake River Valley AVA

In Chapter 1 we discussed the importance of topography and slope as a mechanism for airflow pathways at the vineyard scale. However, this phenomenon also exists at the regional scale. With the large valley structure of the SRVAVA, we can expect the cooler, denser air at the boundaries of the AVA to be gravitationally fed into the valley in the evenings as the sun drops below the horizon and solar radiation ceases. This is demonstrated in figure 2.2 comparing daily maximum and minimum temperatures with elevation. Daily maximum temperatures show a clear correlation, with the low elevations (and latitudes, as denoted by color) having higher values, as expected due to the typical atmospheric profile of decreasing temperatures with increases in elevation. Minimum temperatures, however, have a much different relationship due to the common temperature inversions known to take place in valleys when the denser cool air flows towards the bottom of the valley reversing the atmospheric profile (for our purposes, the term *inversion* means a reversal of the atmospheric temperature

profile at *any* point in time, not necessarily a long lasting event that many Boise locals think of when they speak of inversions). The correlation is not quite an inverse of the TMAX relationship, however, many of the lower elevation locations have a notable shift to cooler temperatures. The coldest mean temperatures are still clustered around 1000 m but the temperature gradient between high and low temperatures has been reduced. A couple of possible explanations for this is that there is 1) not enough average momentum for the cool day time air to flow to the bottom of the valley, and/or 2) not enough time (overnight) for the dense air to make it to the bottom of the valley.

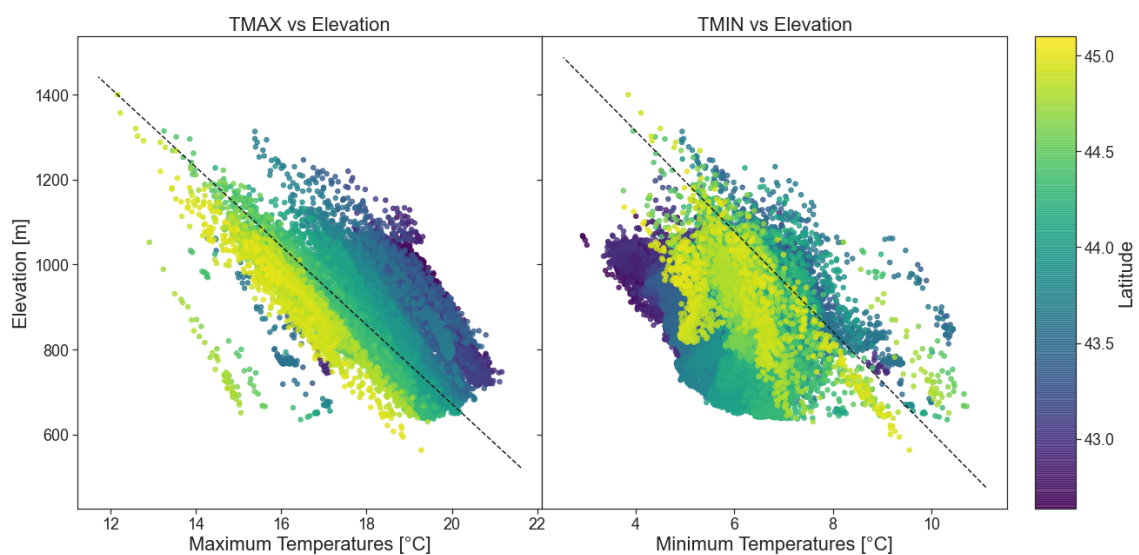


Figure 2.2: Mean minimum and maximum daily temperatures against elevation in the Snake River Valley AVA.

The primary risks associated with these daily temperature inversions are frost and freezing damage due to the freezing air flowing into the base of the valley where the majority of vineyards are located. These daily inversions are common but are usually reversed to a typical atmospheric profile the next day when the solar radiation

heats the ground in the morning and afternoon. Occasionally, these conditions come with intense fog and low cloud cover which can prevent daytime solar insolation from fully penetrating to the ground (especially in the winter when solar incidence angles are low), thus keeping the temperature inversion in place (Yao & Zhong, 2009). This mechanism can provide more time for the coolest air to be fed to the lowest elevation and lead to an increased chance of freezing the plant tissue depending on the magnitude and duration of the temperatures. Figure 2.3 show the locations of extreme freezing events highlighting the inversion effect and the magnitude of these extremes. We see that the most extreme of the events (such as the one in January of 2017), recording temperatures below -35°C , are contained near the base of the valley around the Snake River. Looking at the number of events in the less extreme plots, we see a shift to many more events in the eastern segment of the AVA. This is likely a product of the steeper slope of the canyon, along with cooler base temperatures due to the higher slightly higher elevation, providing much more momentum for the air to travel down and may not require the extended inversion periods to get the coldest air to the base of the valley. It is also interesting to note that the cooler air in the eastern AVA tends to settle well north of the Snake River, and that area south of the river, specifically from longitudes of approximately 115.5°W to 116.5°W , has a similar record of extreme frost as the well known sunny slope region.

The growing season length is often characterized by the timespan between the last frost of the spring and the first frost of the fall season, also known as consecutive frost-free days. Figure 2.4 provides a look at the total distribution of seasonal frost timing and it's standard deviation through time. We see a bi-modal distribution for both the first and last frosts, with the "short" portions of the bi-modal distribution

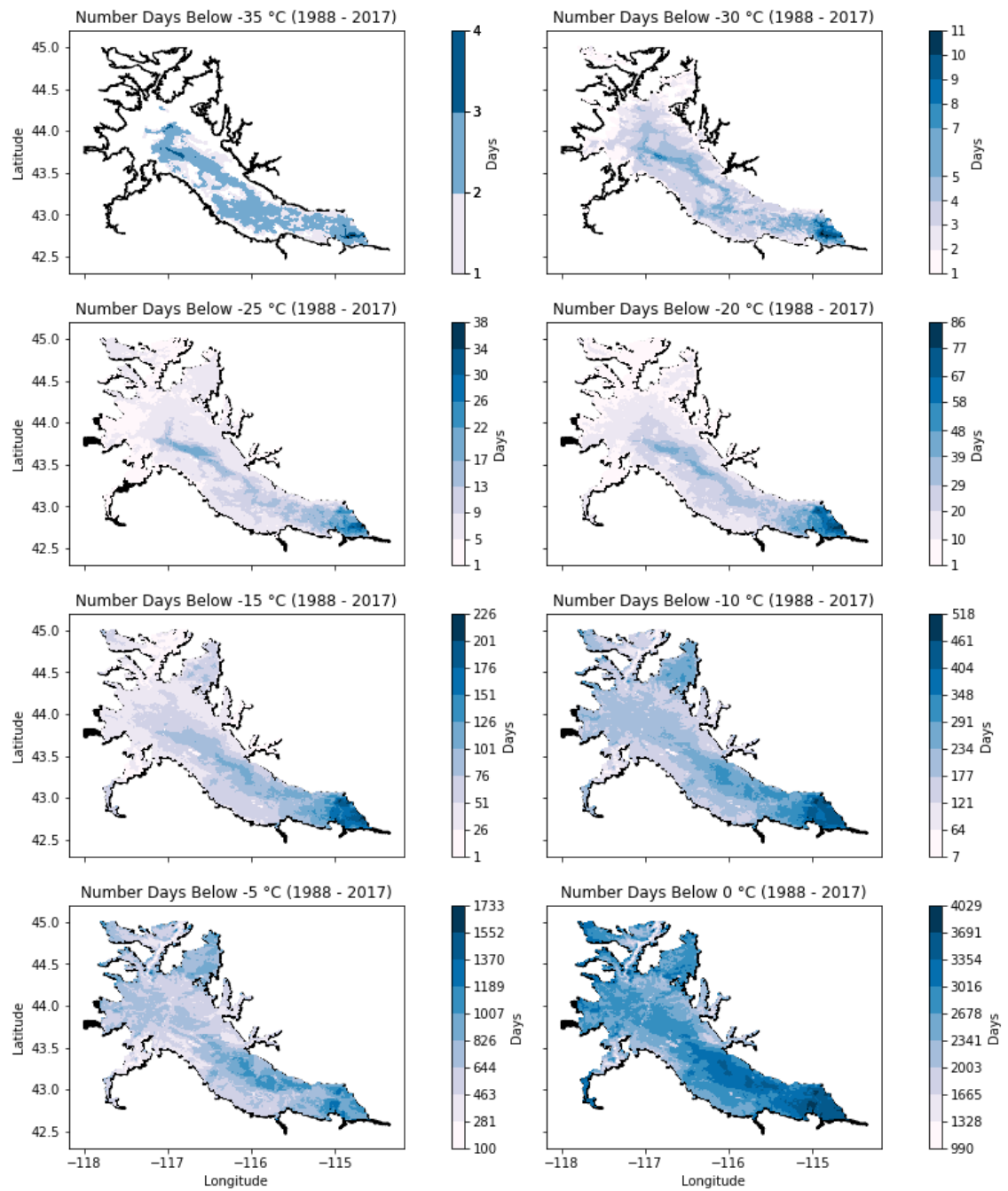


Figure 2.3: Number of extreme cold temperature events in the SRVAVA (1988 - 2017).

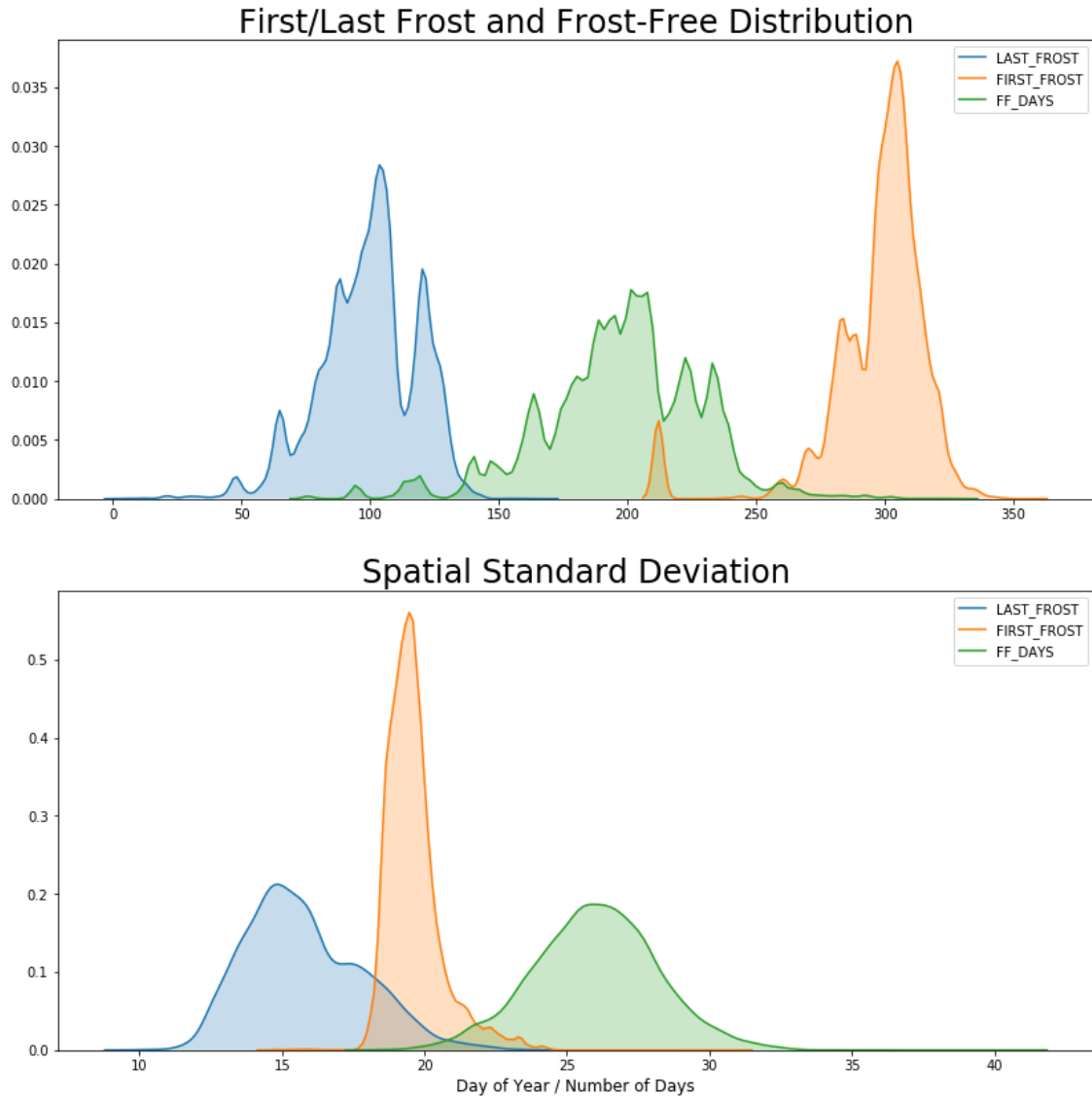


Figure 2.4: Probability density functions of frost timing and variance in the SRVAVA (1988 - 2017).

being areas of much higher risk for viticulture. Those distributions have frost events extending into May and starting as early as September which would be unsuitable for most vines, although, some species can ripen sooner than others and not require as long of growing season. Traditionally, the growing season for grapes is classified as

April 1st through October 31st, a season length of 214 days. The mean season length of points within the AVA is 201 days (not shown), and the bulk of the distribution lies within a 150 and 250 day season length with half of the area within the AVA having a mean season length of over 200 days. Fall frosts have much narrower distribution in both space and time which indicates that the weather of the AVA changes much more as a whole in the fall and is prone to more intraregional variance in the spring. Despite a significantly narrower distribution in the fall, the overall magnitude of the variance is significantly higher in the fall with a mean deviation of about 19 days. The wide distribution of timing in the spring has a mean of approximately 15 days. The combination of both shows there's a mean standard deviation of 25 days for the total growing season. This could help inform some vineyard planting and management decisions that still allow for successful production when growing seasons deviate to a shorter length by more than a month.

General mean temperatures in the SRVAVA are shown in figure 2.5 and shows a small seasonal lag in the summertime with August being the warmest month despite the largest amount of solar radiation occurring in July. December is the coldest month, but not lagged as it also provides the least amount of solar heating. Outliers in low mean temperatures are much more common in the winter months and are most likely more evidence of the positive feedback of long lasting temperature inversion patterns known to be more frequent in the wintertime.

As discussed in chapter 1, large diurnal temperature swings in the growing season has positive benefits to viticulture as it improves grape quality by balancing sugar and acid levels in the fruit. The SRVAVA has some of the largest diurnal swings out of any area that is otherwise suitable for viticulture in the Pacific Northwest with

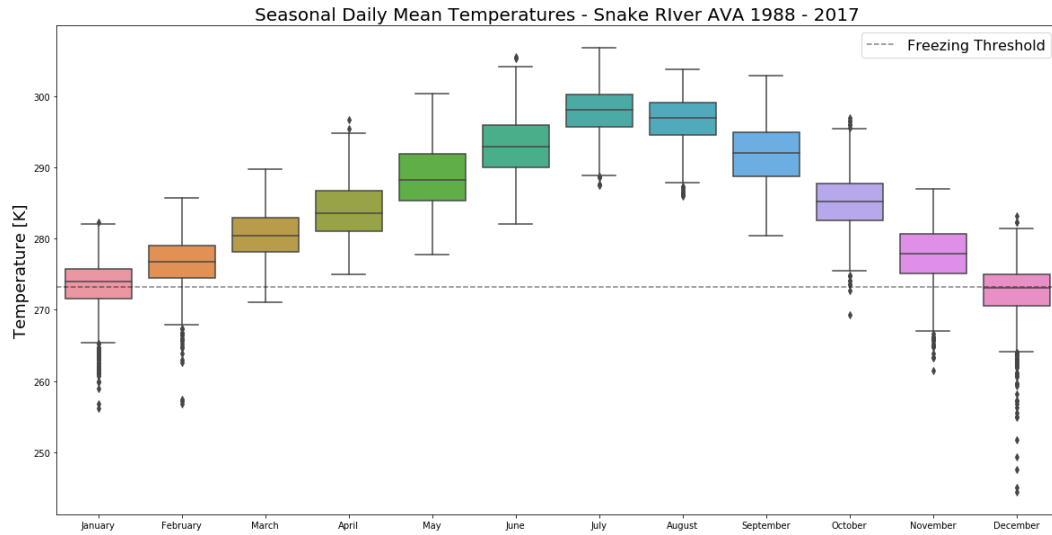


Figure 2.5: Mean temperature distribution in the SRVAVA (1988 - 2017).

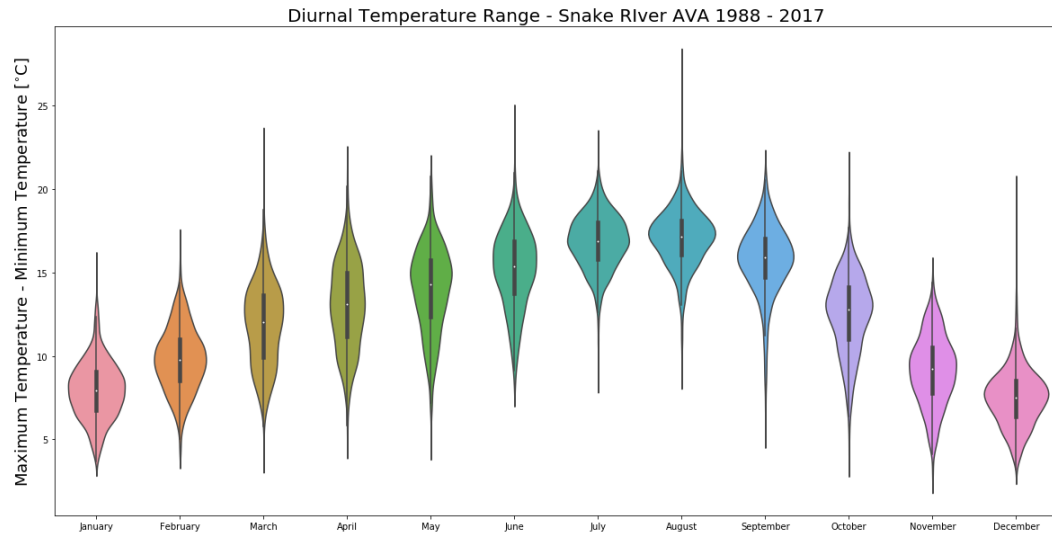


Figure 2.6: Diurnal temperature variation in the SRVAVA (1988 - 2017).

its seasonal distributions shown in figure 2.6. It follows a similar seasonal pattern as mean temperature with the highest spreads in late summer and early fall when it's most important for the ripening fruit. Additionally, you can see narrower distributions in the spring that is evident of the more localized convective weather present in the

spring.

All vineyards within the SRVAVA are fully irrigated and do not rely on direct precipitation, but rather on mountainous snowpack. However, there have been some drought years that did not provide enough snowmelt to the irrigation canals for some vineyards without priority water rights. Drought is a normal and complex component of Idaho climate, however the Idaho State Hazard Mitigation Plan report (State of Idaho Hazard Mitigation Plan, 2018) has noted that future increases in the magnitude and frequency of drought events are likely.

Annual AVA precipitation distributions are shown in figure 2.7. Although rare, when significant droughts do take place and water is not available towards the end of the season, it can impact production of current and future grape production. Current year production can be forced to harvest ahead of schedule with lower yields and lower quality fruit. Future years can be affected by not having the ability to “flood” the vineyards after harvest, a common practice to reduce the impact of early frost events due to the high heat capacity of water.

2.6 Agroclimate Analysis

2.6.1 Growing Degree Days (24-hour method)

Growing degree days are classified as a heat summation unit that serves as a heuristic tool to represent the total relevant heat used for plant growth and development. Growing degree days are most frequently calculated by estimating a daily mean temperature by taking an average of the daily minimum and maximum temperatures. However, the distribution of temperatures throughout a day is often skewed in one direction that averaging the minimum and maximum temperature would not

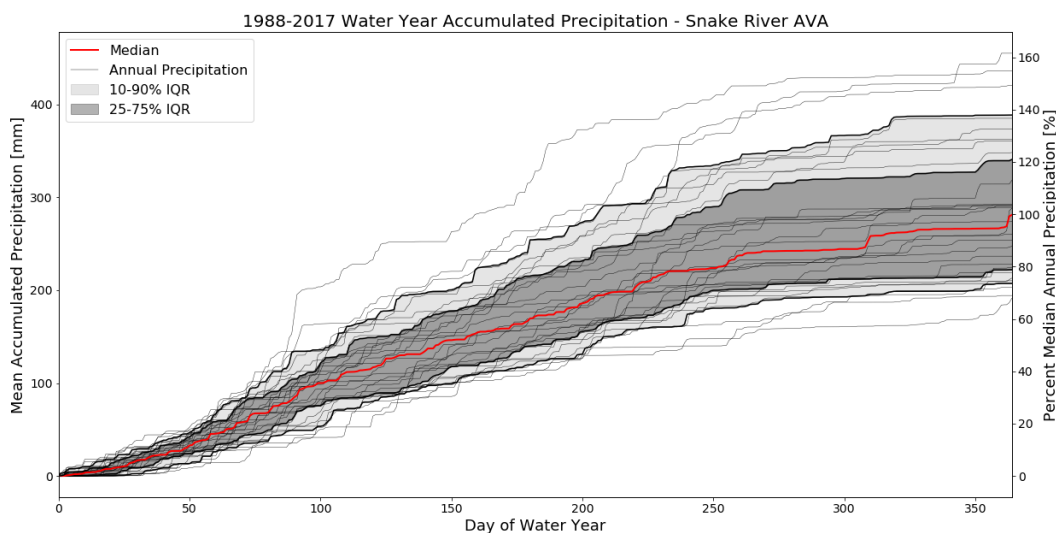


Figure 2.7: Spaghetti plot of cumulative aggregated precipitation in the SRVAVA (1988-2017).

accurately reflect a mean daily temperature. Furthermore, the difference in skewness of daily temperature profiles across different locations and/or regions can lead to the metric being better or worse suited in certain locations.

Literature has shown that utilizing hourly data to calculate growing degree days is a more accurate representation of plant phenology timing (Gu *et al.*, 2014). Thus, we employ a method to utilize all 24 hourly temperatures as shown earlier in table 2.2. Figure 2.8 shows the seasonal sums of growing degrees using the traditional method vs the 24-hour method which shows a significant overestimation of GDD using the traditional method. Differences of over 100 GDD °C are noted in multiple years which could potentially alter planting and management decisions by farmers if historical data was given using a 24-hour method. Additionally, this difference between the methods is getting smaller through time which is due to climate change and the relative changes to minimum and maximum temperatures used with the traditional method.

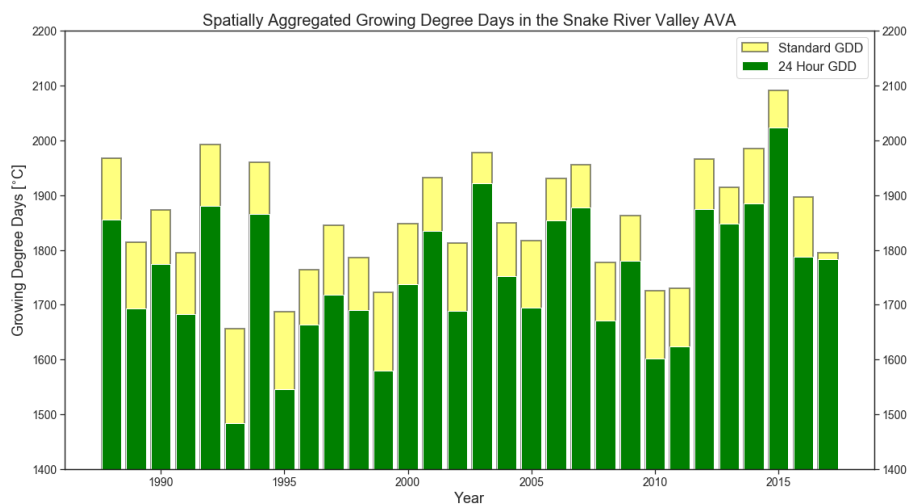


Figure 2.8: Differences in seasonal growing degree days estimating mean daily temperature in two various ways (TMIN/TMAX only, vs 24 hourly observations) in the SRVAVA (1988-2017).

The traditional method for calculating GDD is still considered the industry standard, presumably due to the simple requirement of only needing a minimum and maximum daily temperatures. As data at higher temporal resolution becomes increasingly common, other methods may become more popular. However, for the remainder of this analysis, all references to GDD will be shown using the 24-hour method and summed during the growing season (April 1st - October 31st).

The SRVAVA provides a similar amount of growing degree days to other famed wine growing regions such as Napa Valley (figure 2.9). The approximate area of the sunny slope region is shown by the black bounding box and on average has about 1900 GDD °C (3420 °F) which is suitable for most grapes outside those with the highest heat demands such as Zinfandel or Grenache. However, there is a standard deviation of about 130 GDD °C (234 °F) which can pose risk of certain varieties not fully ripening (even without the risk of frost). Mean deviation is smaller in the southeast portion of the AVA, but as seen earlier, this area experiences a higher risk

for severe frost. Also of note is the large area of GDD over 2100 °C (colored in yellow) which represents the urban heat island effect of the Boise metropolitan area and is captured by the land surface model used by WRF.

2.6.2 Frost Hours

Freezing temperatures are the primary threat to vineyards in the SRVAVA. However, the frequency, magnitude and duration of freezing temperatures are all impor-

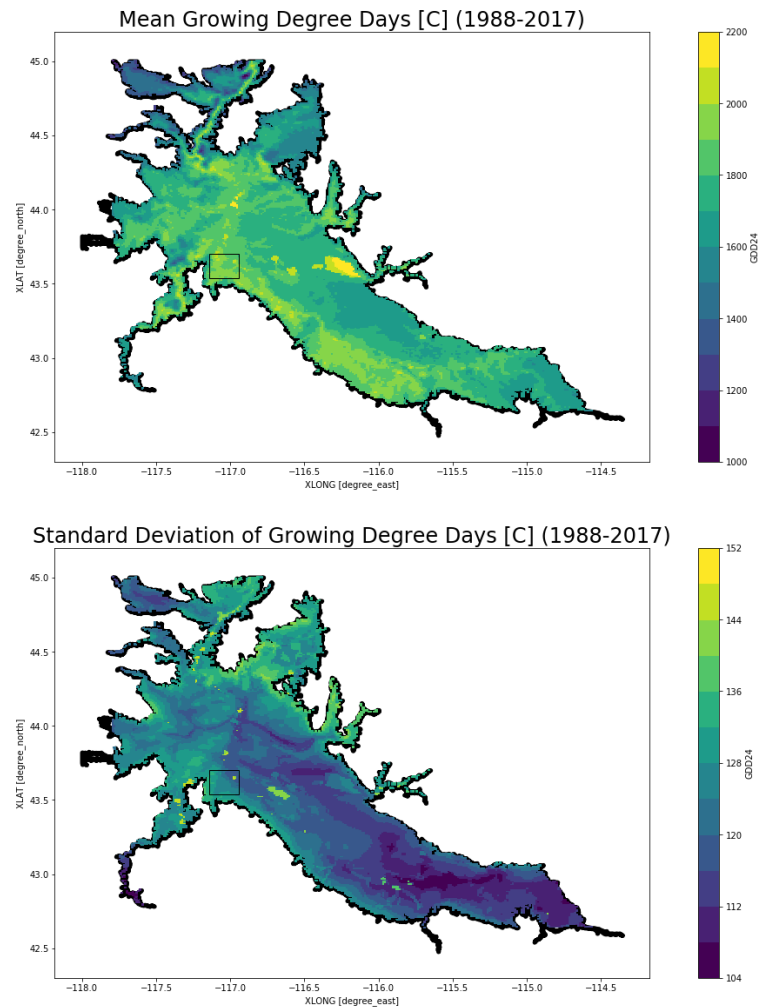


Figure 2.9: Mean and standard deviation of growing degree days in the SRVAVA (1988-2017).

tant factors and it can be difficult to tease apart relative contributions of each. Frost hours (FROSTH) aim to characterize the frequency and duration components by simply summing binary values where any hour that is below 0 °C is ‘1’, and all non-freezing temperatures are ‘0’. This is particularly helpful compared to daily data, as the effects of only a single hour of freezing temperatures or the entire day being below freezing (undetermined with daily data) could have significant effects on the health of a crop. This is especially important for the shoulder seasons when vines are not fully acclimated to cold temperatures and are much more vulnerable. Figure 2.10 highlights the large variance and timing of total cumulative frost hours. Other use cases for frost hours is determining the length of the growing season which is commonly referred to the period between the last (spring) and first (fall) frosts of the season. Although frost hours are useful in determining periods that may pose a risk to viticulture, it does not take into account the magnitude of the freezing temperatures, which is the subject of the next metric.

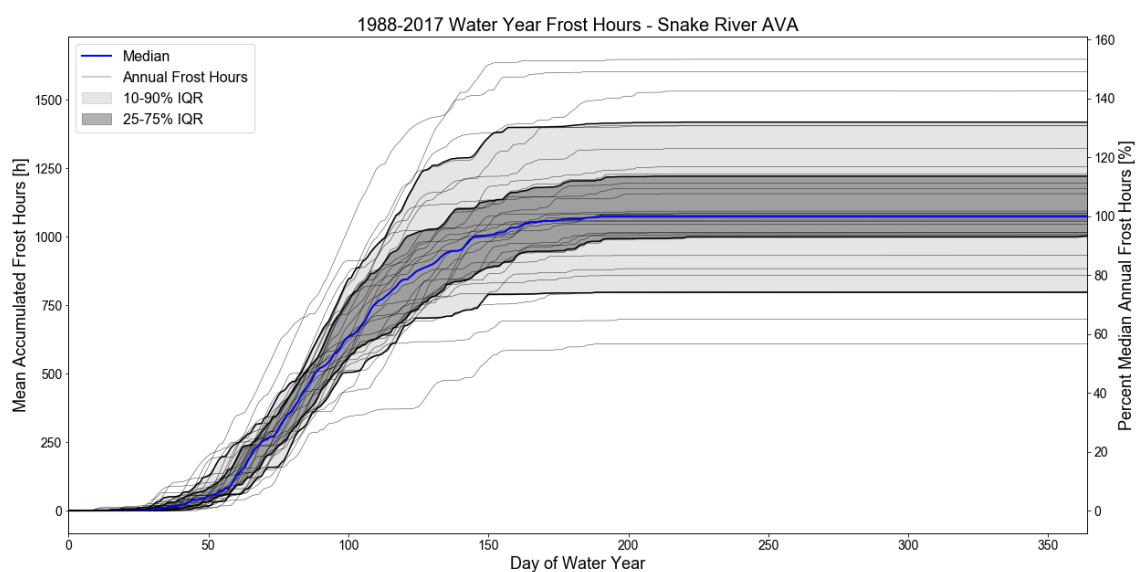


Figure 2.10: Cumulative mean aggregated freezing hours in SRVAVA (1988-2017).

2.6.3 Frost Degree Days

The frost degree day (FDD) is a metric computed from hourly data to better characterize the magnitude of freezing temperatures without the need of creating multiple thresholds in addition to the freezing threshold. It is calculated by summing the difference between 0 °C and any negative temperature for each hour of the day and is represented by the formula in table 2.2. A simplified hypothetical example would be if the temperature was -2 °C for the entire day (24 hours) the FDD would be $(0 - (-2)) * 24 = 48$. If the temperature was -10 for the entire day, the daily FDD would be $(0 - (-10)) * 24 = 240$, and represent the magnitude of the temperatures. The previously introduced frost hour metric would be 24 for both cases. Combining both duration and magnitude into a single metric provides a more complete picture of how vines have been affected by certain hazardous conditions. A look at the relative difference between frost hours and frost degree days (figure 2.11) demonstrates a shift to more high impact events despite the same amount of time spent below freezing. This is emphasized in figure 2.12 showing a universal drop in the amount of hours below freezing, yet many low-elevation areas (particularly in the lower valley) experienced an increase in total FDD.

2.6.4 Powdery Mildew Index

Powdery mildew is the most widely known fungal pathogen to affect grapes, although many other harmful pathogens exist such as: black rot, bunch rot, and downy mildew. Powdery mildew can have a number of negative effects including reduced vine growth, yield, fruit quality, and cold-hardiness. It thrives in warm, humid conditions with an ideal temperature range of 21 - 30 °C (70 - 85 °F). It's common practice for vineyard managers to widely use the fungicide copper sulfate ($CuSO_4$) as a measure

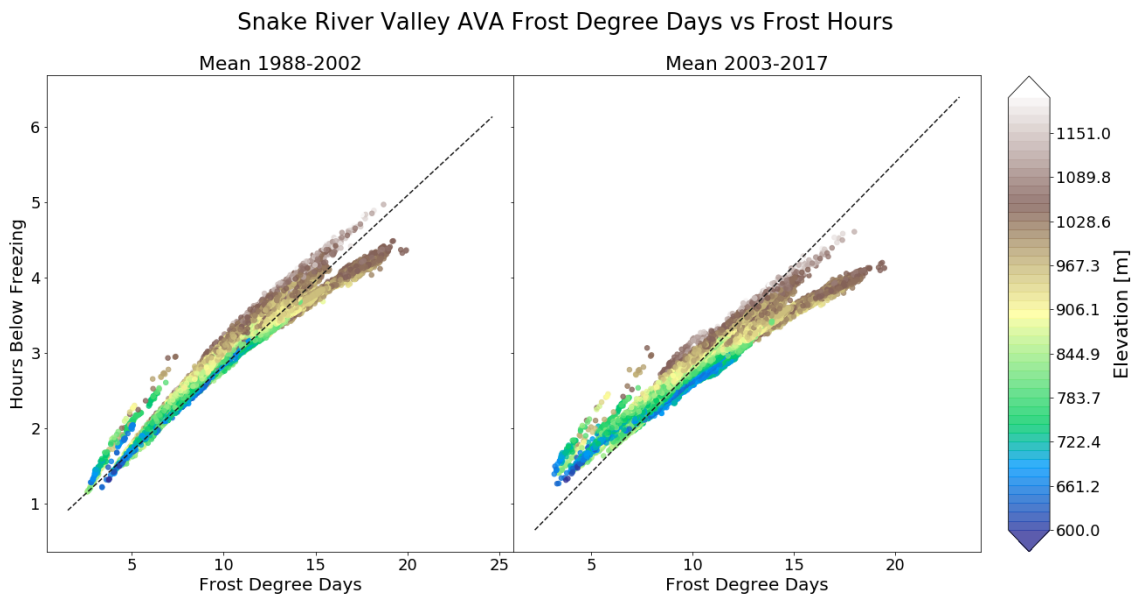


Figure 2.11: 15-year split of frost degree days against frost hours. Magnitude of freeze events is stronger per hour in most recent time period.

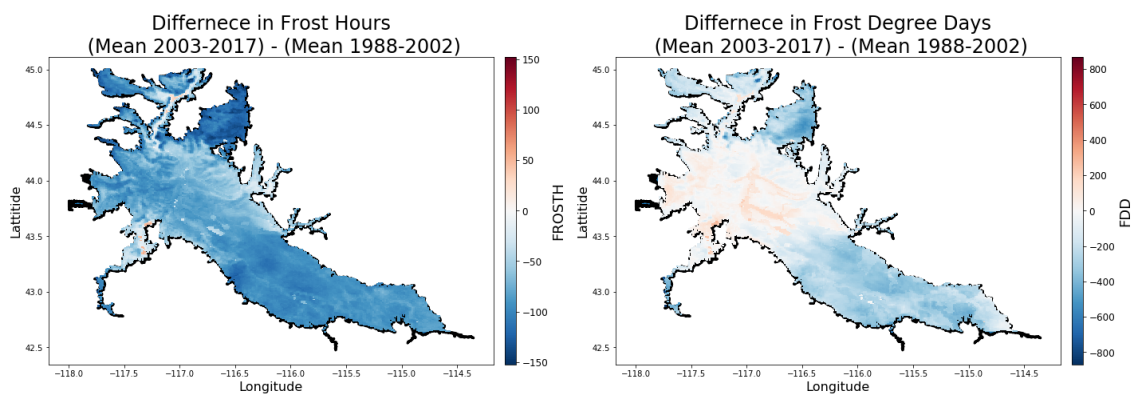


Figure 2.12: 15-year differences of FROSTH and FDD. An increase of FDD, indicating an increase in the magnitude of freezing events, is noted in the lower valley where many vineyards are located.

to mitigate powdery mildew outbreaks, which is both expensive and often damaging to local aquatic ecosystems. Despite the much drier conditions of the SRVAVA compared to other famed wine regions, powdery mildew is still a considerable threat.

UC Davis created the powdery mildew index (PMI) as a tool to help determine

disease pressure and adjust fungicide spray timings accordingly (DR & Bettiga LJ, Revised continuously). The index ranges from 0 - 100, with values lower than 30 to be considered mild pressure, 30 - 60 moderate pressure, and greater than 60 high pressure. The risk index calculation requires hourly data and is calculated with a two step process, initialization and adjustment. The initialization process is started with the following procedure:

1. There must be three consecutive days with 6 or more continuous hours of canopy temperatures between 70 - 85 °F
2. Once, number 1) has been satisfied, starting with the index at 0 on the first day, add 20 points for each day with 6 or more continuous hours of temperatures between 70 and 85 °F
3. Until the index reaches 60, if a day has fewer than 6 continuous hours of temperatures between 70 and 85 °F, reset the index to 0 and continue
4. Once the index reaches 60, the index has been initialized and adjustment procedures should be followed

Adjustment (after initialization) constrained with the following guidelines:

- If the index is already at 100, you can't add points
- If the index is already at 0, you can't subtract points
- You can't add more than 20 points a day
- You can't subtract more than 10 points a day

- If fewer than 6 continuous hours of temperatures occurred between 70 and 85 °F, subtract 10 points
- 6 or more continuous hours of temperatures occurred between 70 and 85 °F, add 20 points
- If temperatures reached 95 °F for more than 15 minutes, subtract 10 points
- If there are 6 or more continuous hours with temperatures between 70 and 85 °F AND the temperature rises to or above 95 °F for at least 15 minutes, add 10 points (this is the equivalent of combining points 2 and 3 above)

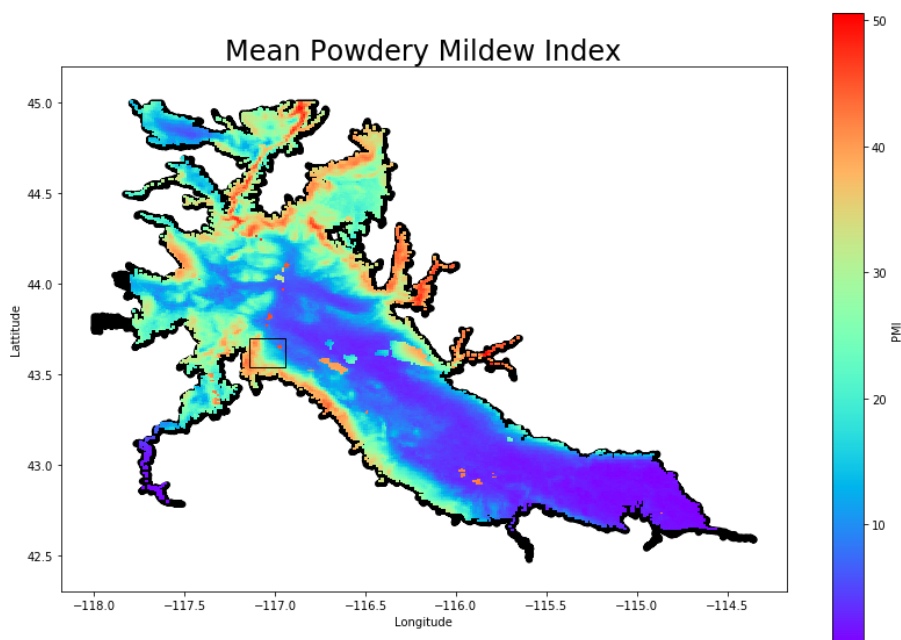


Figure 2.13: Mean powdery mildew risk index. Majority of current vineyards located within the sunny slope region denoted by the black bounding box.

To our knowledge, this is the only regional climate dataset that has been used to produce a spatio-temporal map of PMI. With the complex topography in the

SRVAVA, PMI can aid in site selection by identifying areas with higher or lower risk, on average, and areas more prone to extreme events. There is high variance in PMI due to the complex topography and flow paths in the SRVAVA as seen in figure 2.13. Of particular interest is the large variance of PMI within the “sunny slope” region, where the majority of vineyards are located, as denoted by the black bounding box.

2.7 Climate Trends

A large amount of natural and anthropogenic variance exists within the earth’s climate and occurs at a variety of timescales. Our trend analysis seeks to describe the overall variance within the 30 year climate normal. Our aim is two-fold: 1) to perform linear trend estimation and other statistical tests to understand the long term changes that can help better understand future climate and its agricultural implications, and 2) to investigate the intraperiod variability and potential linkages to large scale climate with a cyclical nature.

2.7.1 Linear Trend Estimations

According to the National Oceanic and Atmospheric Administration (NOAA), global mean temperatures have risen by approximately 0.07 °C per decade since 1880, and by 0.18 °C per decade since 1981 (Lindsey, 2020). These estimates include warming to both the land and ocean, which would have significant differences if looked at separately, as water requires more heat than the land or atmosphere to rise to the same temperature. Thus, we may expect temperatures in the SRVAVA to see increases higher than 0.18 °C per decade as our analysis is nearly entirely constrained to a land surface. This is confirmed with SRVAVA linear trend estimates of +0.72 (TMEAN), +0.80 (TMAX), and +0.62 °C (TMIN) which break down to +0.24, +0.27, and +0.21

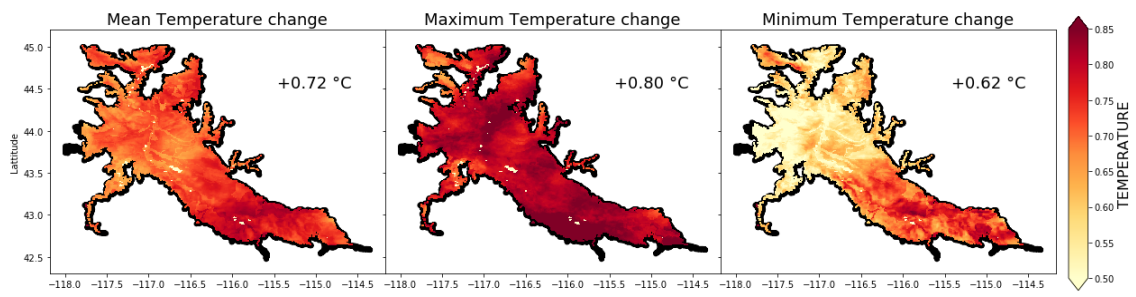


Figure 2.14: Mean temperature changes from October 1987 - September 2017.

°C per decade increases (figure 2.14). Daily maximums are increasing over 29% faster than daily minimums, which is contrary to the majority of the globe, where minimums are increasing faster than maximums (Lindsey, 2020). This results in a larger diurnal temperature spread that can potentially benefit vineyards by helping balance sugar and acidity in the fruit. However, the distribution of temperature changes are not evenly distributed by month (figure 2.15).

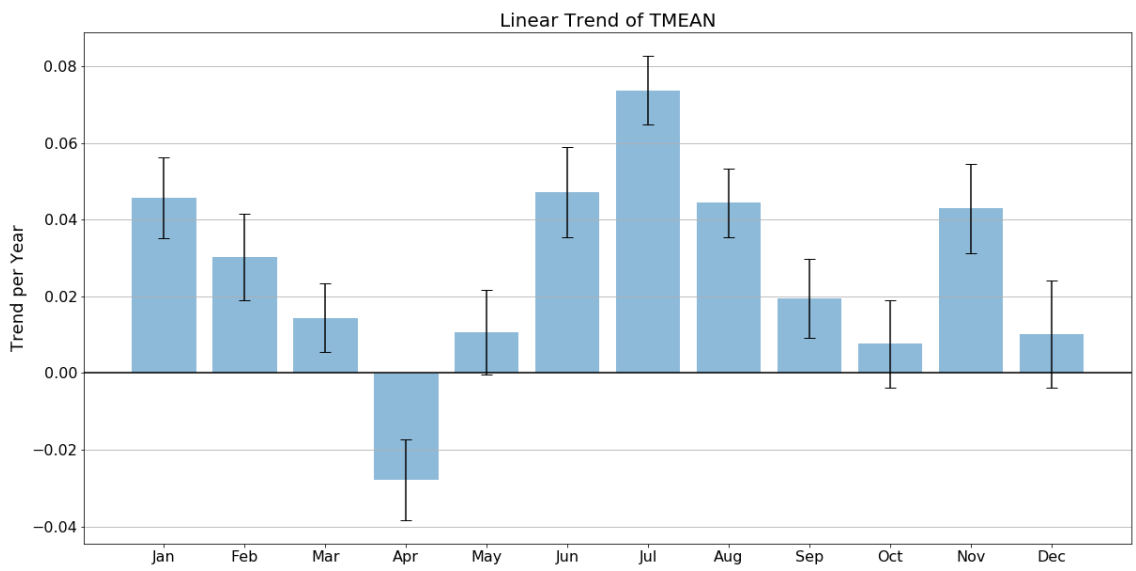


Figure 2.15: Monthly mean temperature changes from October 1987 - September 2017. Error bars represent one standard deviation comprised of 1000 Monte Carlo Samples.

Error bars are represented as one standard deviation pulled from a Monte Carlo

simulation using random samples of 80% of the data to help constrain the effects of high magnitude events. Summers within the SRVAVA are warming faster than winters, with the agricultural shoulder seasons warming the slowest. April is the only month that is cooling, and could have negative impacts for viticulture by earlier deacclimation to cold weather in late winter making them slightly more vulnerable to freezing events in spring. Additionally, the larger increases in maximum temperatures in the summer poses additional threat of burning the leaves or fruit. A statistical summary of each metric including its associated linear trend (per decade) and accompanying p-value is presented in table 2.3.

GDD24 was the only metric found to have a statistically significant linear trend (p-value < 0.10), though raw temperature and frost metrics have relatively low p-values compared to other metrics. One explanation for the lack of significance in linear trends is the relatively low sample size (30, as everything was annually aggregated) coupled with a noisy signal which is typical in climate dynamics. A small presence of high magnitude anomalous years could highly impact significance within a relatively short climate period. It is entirely possible that a longer time series would result in many metrics having a significant linear trend. Despite the lack of significance of this particular statistical t-test, the general trends may have some qualitative usefulness.

Total GDD are increasing at a rate of 43 heat units [$^{\circ}\text{C}$] per decade (128 over 30 years) which is approximately equal to the annual standard deviation over the same period. Benefits of this include new suitability for more heat demanding varieties such as malbecs and zinfandels, as well as reduced risk for more moderate demanding varieties to reach optimal GDD on years that deviate below the mean. Additionally, as climate warming is widely expected to continue increasing at similar or higher

Table 2.3: Spatially aggregated Statistical summary of each metric within the SRVAVA from October 1987 through September 2017. Linear trends calculated with annually aggregated data according to method.

Metric	Method	Mean	Min	25%	50%	75%	Max	STD	Trend (units per decade)	P-val	% Change w/ respect to STD
TMIN	Annual mean	6.04	4.74	5.62	6.29	6.41	7.59	0.66	0.21	0.14	95%
TMEAN	Annual mean	11.69	10.08	11.03	11.88	12.12	13.31	0.73	0.24	0.12	100%
TMAX	Annual mean	18.59	16.61	17.85	18.80	19.13	20.33	0.85	0.27	0.14	96%
TSPREAD	Annual mean	12.6	11.6	12.3	12.5	12.9	13.3	0.44	0.06	0.52	42%
GDD	Seasonal sum	1857	1656	1788	1849	1949	2090	105	25.1	0.26	73%
GDD24	Seasonal sum	1756	1484	1684	1763	1855	2024	123	42.6	0.10	106%
Season Length	Consecutive days	131	85	117	130	146	183	22	2.54	0.16	35%
First Frost (Fall)	Date threshold	Sep 26	Aug 25	Sep 15	Sep 25	Oct 6	Oct 26	17	4.54	0.22	80%
Last Frost (Spring)	Date threshold	May 17	Apr 18	May 7	May 15	May 30	Jun 16	16	2.01	0.56	38%
FROSTH	Annual sum	1104	609	1003	1075	1222	1649	249	-48.6	0.36	-60%
FDD	Annual sum	4006	1319	2713	3860	4637	8193	1872	-203.0	0.62	-33%
PMI	Annual mean	9.62	5.34	7.70	9.16	11.11	15.35	2.66	0.47	0.41	54%
PRCP	Annual sum	518	115	344	438	497	598	760	9.87	0.69	26%
Q2	Annual mean	4.51e ⁻³	4.05e ⁻³	4.36e ⁻³	4.48e ⁻³	4.67e ⁻³	5.25e ⁻³	2.57e ⁻⁴	4.2e ⁻⁵	0.45	50%
SWNORM	Annual mean	217.7	211.4	215.5	218.3	219.9	222.9	3.25	-0.178	0.80	-17%

rates, these benefits will be further enhanced.

Both frost related metrics, FROSTH and FDD are shown to be decreasing through time. However, in addition to the general reduction of freezing temperatures, the timing of such events is important. Figures 2.16 and 2.17 show the spatially aggregated annual timings of the first and last frost and the growing season length (span between last and first frost) and the associated trends. Interestingly, there is a seasonal shift towards the end of the calendar year for both the first and last frost. However, the trend for spring has moved by approximately 6 days forward and the fall frost shifting

by about 14 days, thus having the total growing season increased by almost 8 days. Two potential implications of this would be, 1) traditional planting and harvesting dates may not be as valuable and require adjustment, and 2) a longer total growing season offers more time for fruit to develop.

Cummulative Distribution Functions of Season Length and Frost Timing

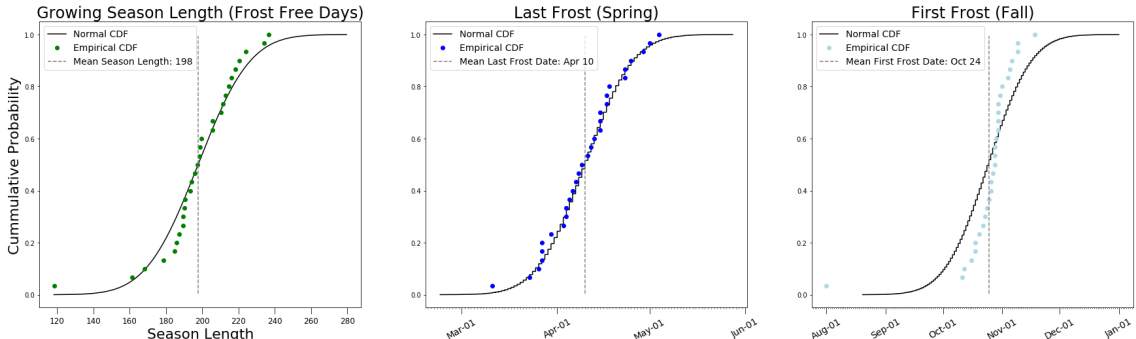


Figure 2.16: Cumulative distribution functions of frost timing and season length.

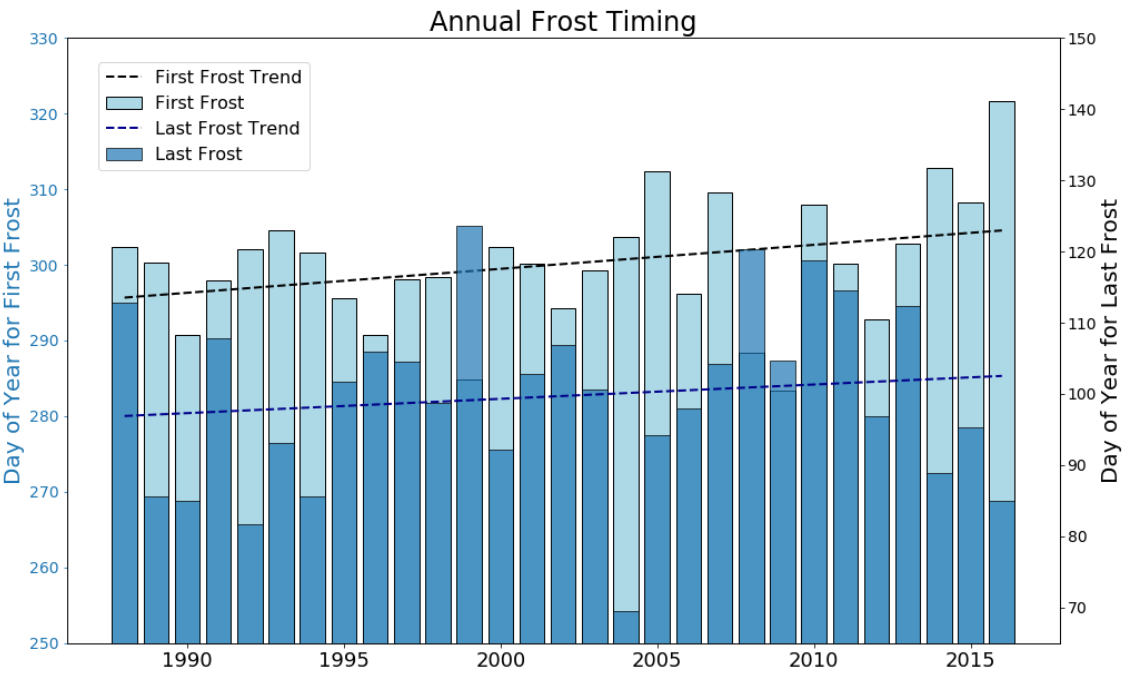


Figure 2.17: Frost timing by year.

Overall, we see important linear trends across all temperature driven metrics that by and large, are potentially beneficial to viticulture in the SRVAVA. This includes an increase of growing degree days, reduction of freezing events, and a longer growing season. Additional statistical tests can also be applied to further characterize changes through time.

Spatial Distribution Tests

Although SRVAVA annual climate does not seem to change in a strict linear fashion as denoted by insignificant student t-tests with linear regression, other statistical tests are applied to better understand changes through time. In table 2.4 we apply two statistical tests, the 2-sample Kolmogorov-Smirnov (KS) test and Mann-Whitney (MW) rank test, over spatial SRVAVA distributions aggregated over 15-year time periods. We compare most of the same metrics as a single tailed test over the mean distribution of water years 1988-2002 vs 2003-2017. The hypothesis for each single tailed test is determined as either 'Greater' or 'Less' for the relative means of each 15-year period. For example, if the first 15-year mean is less than the last 15-year mean, the hypothesis will be tested as 'Greater' than, and vice versa.

Both the KS and the MW tests are nonparametric tests to check if independent samples are drawn from the same continuous distribution. However, their methods are sensitive to different properties of the distributions. The KS test is sensitive to any change in the distribution, such as standard deviation, skewness, or median, and is calculated by looking at the maximum difference of the cumulative distribution functions of each set of data. The MW test is largely only sensitive to changes in median, and is calculated based on the numeric rank (low to high) of all data points,

Table 2.4: 2-sample Kolmogorov-Smirnov and Mann-Whitmann statistical test results for mean spatial distributions over the first and last 15-year periods.

Metric	30YR Mean	First 15YR Mean	Last 15YR Mean	Hypothesis	KS Statistic	KS P-value	MW Statistic	MW P-value
TMIN	6.04	5.88	6.20	Greater	0.13	0.00	2.74e ⁸	0.00
TMEAN	11.69	11.48	11.89	Greater	0.24	0.00	3.03e ⁸	0.00
TMAX	18.59	18.35	18.84	Greater	0.21	0.00	2.91e ⁸	0.00
GDD	1858	1830	1885	Greater	0.19	0.00	2.95e ⁸	0.00
GDD24	1756	1713	1798	Greater	0.30	0.00	3.20e ⁸	0.00
FROSTH	1105	1147	1063	Less	0.21	0.00	1.73e ⁸	0.00
FDD	4006	4076	3936	Less	0.07	0.00	2.15e ⁸	0.00
PRCP	286.25	291.91	280.58	Less	0.07	0.00	2.10e ⁸	0.00
SWNORM	217.72	217.61	217.83	Greater	0.02	1.25e ⁻²	2.37e ⁸	0.00
Q2	4.51e ⁻³	4.49e ⁻³	4.52e ⁻³	Greater	0.22	0.00	2.89e ⁸	0.00

and then calculates the mean rank of each distribution. The test statistics for both tests represent a rejection of the null hypothesis with high test statistic values and low p-values. The range for the KS test statistic is 0-1 and the max of the MW statistic is the sum of the ranks of the distribution.

P-values across all metrics for each test indicated significance at the 1% level (p-value < 0.01), however SWNORM has a low KS statistic (0.02) indicating a lack of significance. KS test statistics were highest for GDD24, TMEAN, and TMAX which is in agreement with the largest total linear change in table 2.3. TMIN on the hand, has relatively low KS statistic, but a relatively high MW statistic, indicating the total median has likely increased through time as well as a change in the shape of distribution. One cause could be a significant difference in long-lasting temperature inversion patterns that can be common in the valley. Other metrics such as PRCP

and FDD show reductions in total median as well, however the strength of the test statistics are lower. These tests do not quantify by how much the metrics have changed over time, but rather if they have *any* statistically different changes present, and provides further support of the validity of the linear trends noted earlier.

Although these statistical tests have been useful for seeing big picture trends, it does not characterize the large variance of climate well, which is evident by the large standard deviations seen earlier in table 2.3. The following section takes a closer look at the annual variance of climate in the SRVAVA.

2.7.2 Interannual Variability and Correlations to Large Scale Climate

In the previous section, we identified strong linear trends in many agriculturally relevant metrics in the SRVAVA, but largely glanced over the annual variability. Characterizing the interannual variability is important to help understand the associated agricultural risk. Furthermore, we seek to understand if there is an identifiable non-linear signal in regional climate that accounts for some of the annual variance. The objective of this section is to take a qualitative approach to look at the same metrics as the previous section, and argue that there is a strong overlying signal that guides the variance of climate within the Snake River AVA.

One method to help identify a low frequency signal, is to apply a high pass filter over the same data. We first demonstrate this by adding a filter to the same frost timing figure (figure 2.18) we used in the previous section. Here, we see a clear signal highlighting the nonlinear variance of frost timing in both the spring and fall. Of particular interest, is the clear inverse relationship between the two signals which are represented by periods over seven months apart. This negative covariance is

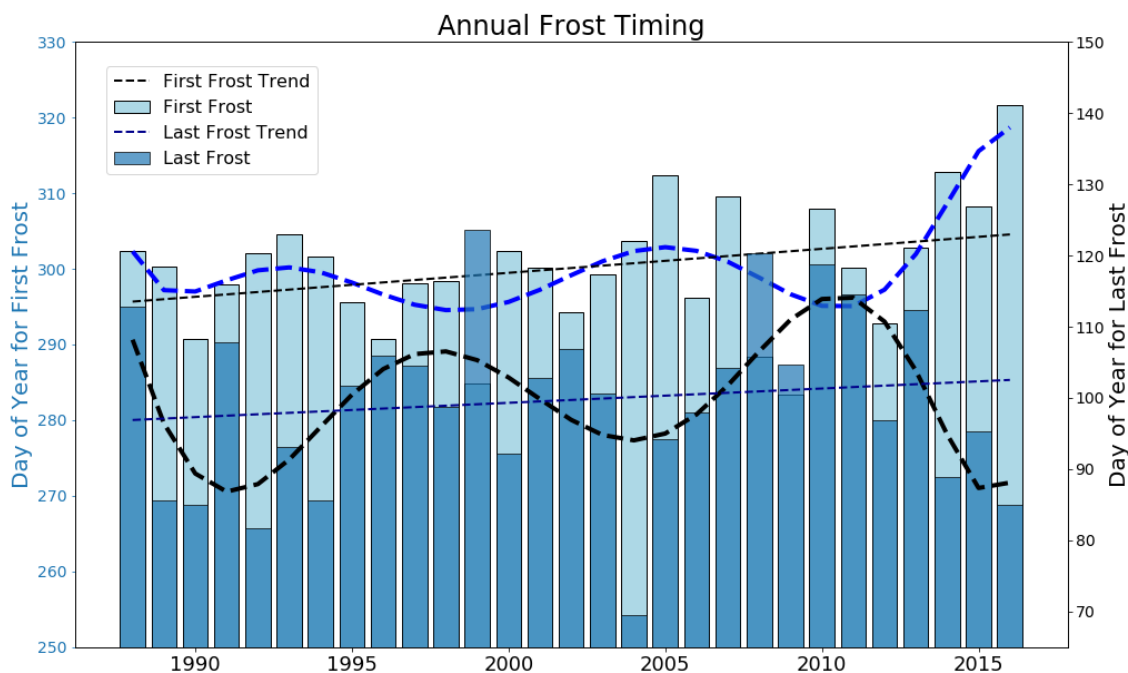


Figure 2.18: Evidence of similar low-frequency patterns in timing of both first and last frosts suggesting strong relationship to large scale climate.

interpreted as having a similar effect based on overall state of the climate within the entire year. In other words, the seasonal frost dates move together with relation to temperature. Specifically, if there is a late frost in spring, it's more likely for an early frost to occur in the fall and vice versa. This may indicate that overall variance in annual temperature may have a signal with a frequency longer than six months, which would explain the negative covariance of first and last frost dates. Further evidence of lower frequency signals are verified with a plot of temperature anomalies by year, where of all 30 years, there are 5 consecutive periods of at least four years where the anomalies are of the same sign (figure 2.19). In addition to the presence of a low frequency signal, we see more evidence that the overall variability within the AVA varies more as a whole region than within the AVA itself as noted by the small spatial

variance of each year.

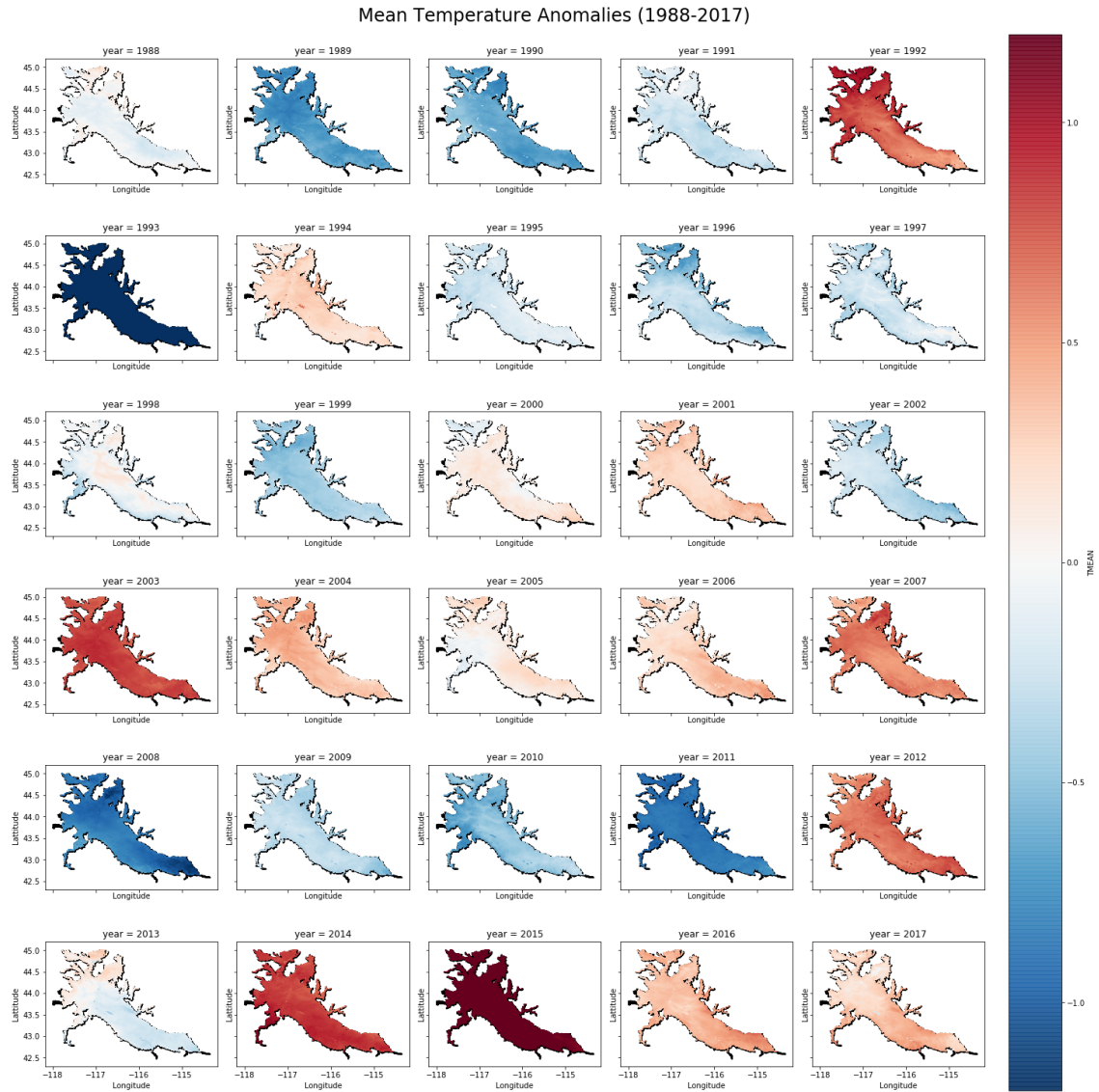


Figure 2.19: Temperature anomalies through space and time.

In addition to identifying a clear low frequency signal in annual temperature and other temperature derived metrics, it is of interest to see if the frequency of extreme anomalous events within a year also has a cyclical nature. Figures 2.21 and 2.20 show the number of daily temperature events that rise above and below the 1 and

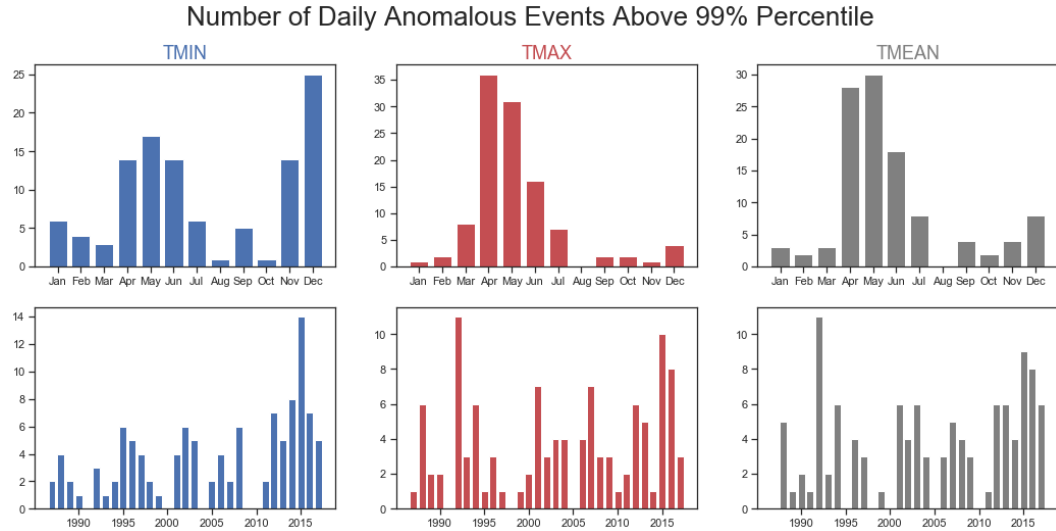


Figure 2.20: Distribution of $> 99\%$ anomalous temperatures. Relative high anomalies are much more likely in the springtime. Clear low frequency signals appear in the annual distribution.

99 percentiles relative to the mean for that same day of year (there are 10958 days in our 30 year dataset, thus we see 1% of the data, or 110 days, in each figure).

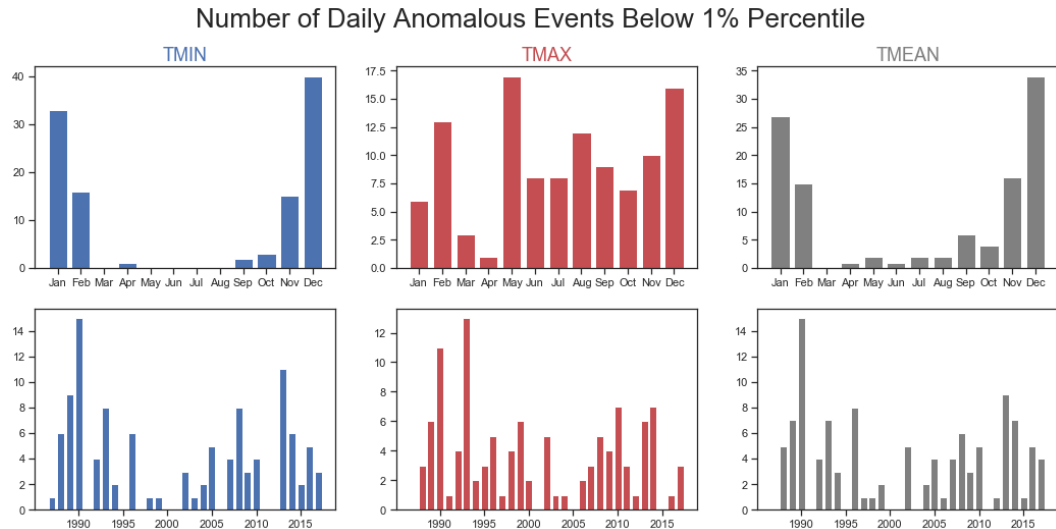


Figure 2.21: Distribution of $< 1\%$ anomalous temperatures. Relative TMIN low anomalies are much more likely in the wintertime.

They are then summed by either month or year. Starting with the yearly sums above the 99th percentile, minimum temperatures (years where the number of days extreme minimum temperatures are well above normal) show the clearest evidence of a low oscillating signal, that also has a notable positive trend. Mean and maximum temperatures show similar wave like signals, though not quite as clear as the minimum temperatures. The same can be seen in the $< 1\%$ percentile events in all categories. This has implications that regional weather fronts (cold or warm) that are responsible for extreme heat waves and cold air outbreaks are also correlated with other larger scale climate mechanisms that vary slower through time.

With evidence that both mean metrics and outlier frequency are externally driven, we can hypothesize a correlation with known large scale climate phenomena such as the Pacific Decadal Oscillation (PDO) and/or El Nino Southern Oscillation (ENSO). Correlations of these large scale climate indicators and regional climate are well documented. Examples include relationships of ENSO to drought in the southwestern United States (DeFlorio *et al.*, 2013) and PDO to total snowpack in Idaho (Kunkel & Pierce, 2006). Despite these apparent correlations between regional and large scale climate, the working relationship between them is not well understood. Chapter 3 takes a closer look at the relationships between regional climate in the Snake River AVA and large scale climate and proposes a modeling technique to better utilize the relationships between them.

2.8 Summary

Climate in the SRVAVA is well suited for viticulture with its adequate heat accumulation throughout the growing season, arid conditions that discourage fungal outbreaks, high diurnal temperature swings, and complex topography to enhance

airflow. Freezing temperatures pose the greatest risk, particularly from long lasting temperature inversion patterns common to the lower parts of the valley due to complex topography, though other risks include extreme high temperatures in the summertime and powdery mildew.

Using a 30-year, high-resolution climate dataset produced by the Weather and Research Forecasting model provides numerous advantages such as: representing both climate and weather components across a climatological normal, resolving important phenomena within a topographically complex region such as cold air transport, and allowing the use and creation of many metrics otherwise not available that can help better characterize an agricultural zone. These characterizations include using new methods to accurately calculate growing degree days (which show a 10-15% overestimation using the traditional method), a custom frost index (FDD) to better represent the magnitude and duration of freeze risk, mapping potential threat of powdery mildew, and more precise spatial mapping useful for site comparison within complex topography.

After examining the data for long term trends for a variety of relevant viticultural metrics, we found statistically significant changes in all temperature driven metrics. These include a substantial increase in growing degree days, an increase in total growing season length, a forward shift in the growing season, an increase in mean diurnal temperature swings, and a reduction of time enduring freezing temperatures (however the magnitude of these events could be increasing in the lower valley). Broadly, these changes should be beneficial to viticulture within the SRVAVA as overall risk is decreasing and more aggressive varieties are more viable than they were 30 years ago. As global climate change is expected to increase for the foreseeable future, we can

expect local temperature driven trends to continue. However, this is all predicated on there being enough snowpack to provide adequate irrigation. As precipitation variance is widely expected to increase as a result of global climate change, this could increase the number of drought years and put an increased importance on seniority of water rights.

Additionally, the SRVAVA climate experiences large interannual variance which can have large implications on fruit yield and quality, although the variance is becoming less of a risk due to the aforementioned trends. Furthermore, we identify strong low frequency signals in both mean annual climate and number of extreme weather events. These signals tend to oscillate at a multi-year (but sub-decadal) scale which can help inform stakeholder decisions. For example, a late spring frost can indicate a higher likelihood for an early fall frost of the same year which may inform some canopy management techniques to prepare for a potential early harvest. We suggest these low frequency signals may be related to large-scale climate indicators such as the El Nino Southern Oscillation (ENSO) which we investigate in Chapter 3.

CHAPTER 3:

FORECASTING REGIONAL CLIMATE FROM LARGE SCALE CLIMATE: A DATA DRIVEN APPROACH

3.1 Introduction

Large scale climate indicators, also known as atmospheric teleconnections, are used to characterize climate variability at a global scale. Teleconnections specifically refer to climate anomalies - usually sea-surface temperature (SST) and/or sea-level pressure (SLP) - that are related at long distances (usually over 1000 km) and are often simplified to a positive, negative, or neutral mode depending on the strength of the relative anomalies. The most widely known teleconnection, the El Nino Southern Oscillation (ENSO), has been significantly correlated with many regional climate phenomena across the globe. For example, effects related to the warm phase of ENSO have been shown to be associated with: drought conditions in South Africa, southern India, Australia, and the Philippines (Richard *et al.*, 2001), stronger hurricanes in the Pacific (Zheng, 1995), and flooding in the southern United States and northeastern South America (Ward *et al.*, 2014). Examples of correlations using other teleconnections include a strong correlation between the negative phase of the Arctic Oscillation

(AO) and a weakening of the polar vortex (Kennedy & Lindsey, 2014), the frequency of heavy precipitation events in the mediterrainian and the North Atlantic Oscillation (NAO) (Ouachani *et al.*, 2013), and similarly the Pacific Decadal Oscillation (PDO) and precipitation in the Pacific Northwest (DeFlorio *et al.*, 2013).

Despite the many identified correlations between large-scale and regional climate, using teleconnections as the sole input to forecast the timing of specific regional events, specifically at the monthly time scale, has been limited compared to forecasts at the seasonal+ scale. Potential reasons for this include the highly variable lead and lag periods in the correlation between teleconnections and regional events, and the widely distributed covariance between different teleconnections across different time spans. Although using large-scale climate to forecast regional events at the monthly time scale has not yet been widely successful, the usefulness of teleconnections is demonstrated by the substantial effort that has been put into forecasting ENSO and other teleconnections with lead times up to two years (see: <https://psl.noaa.gov/enso/enso.forecast.html>).

This chapter aims to first identify correlations between regional climate in the SRVAVA and specific teleconnections, and secondly, to explore a large-scale-data-driven modeling technique to forecast regional climate at the monthly scale with lead times up to one year. We address the former by identifying an underlying relationship between monthly temperature trends in the SRVAVA to both the PDO and the Southern Oscillation Index (SOI) component of ENSO. We address the latter by implementing a long short-term memory (LSTM) model, a specific type of recurrent neural network, that utilizes sequential data to help overcome the inconsistent lead and lag times present between large-scale and regional phenomena. Additionally, we utilize

an empirical mode decomposition (EMD) algorithm on the PDO and SOI signals to identify and feed our network with signals at varying frequencies. By decomposing the signals with EMD, we provide a much more robust representation of the signal to help minimize the effect of noise and maximize the effect of separate frequencies within the raw signals.

3.2 Large-scale Climate

Throughout the last few decades, ENSO has been found to be one of the most impactful climate factors when correlating large-scale and regional-scale phenomena at various locations on earth. Other teleconnections have a history of strong correlations with certain events as well, but the strength of these correlations are often localized. For example, most documented correlations with the Madden-Julian Oscillation (MJO) are constrained to the tropics, and most influences of the Pacific Decadal Oscillation (PDO) are related to countries bordering the Pacific Ocean. There are multiple studies that demonstrate the relationship of PDO with regional climate in the greater Pacific Northwest, as well as Idaho and the Snake River plain (Goodrich, 2007; Mantua *et al.*, 1997; DeFlorio *et al.*, 2013; Kunkel & Pierce, 2006). It is for these reasons - the popularity and widely distributed impact of ENSO, and the relative distance from the Pacific - that we choose to focus on these two Pacific teleconnections.

3.2.1 El Nino Southern Oscillation

The phase of ENSO is determined by a combination of differences between relative anomalies for both SLP and SST near Darwin, Australia and Tahiti in the east Pacific. The El Nino (EN) portion of the ENSO is the SST anomaly which was first noted by Peruvian fisherman seeing the strong variability in quality fishing as a result of the

upwelling (or lack thereof) of cool, nutrient rich water to the surface for fish to feed on. The Southern Oscillation aspect, known as the Southern Oscillation Index (SOI), of ENSO represents the SLP anomalies. Changes in SLP and SST in this scenario are directly related to one another through a phenomena known as the Walker circulation (figure 3.1).

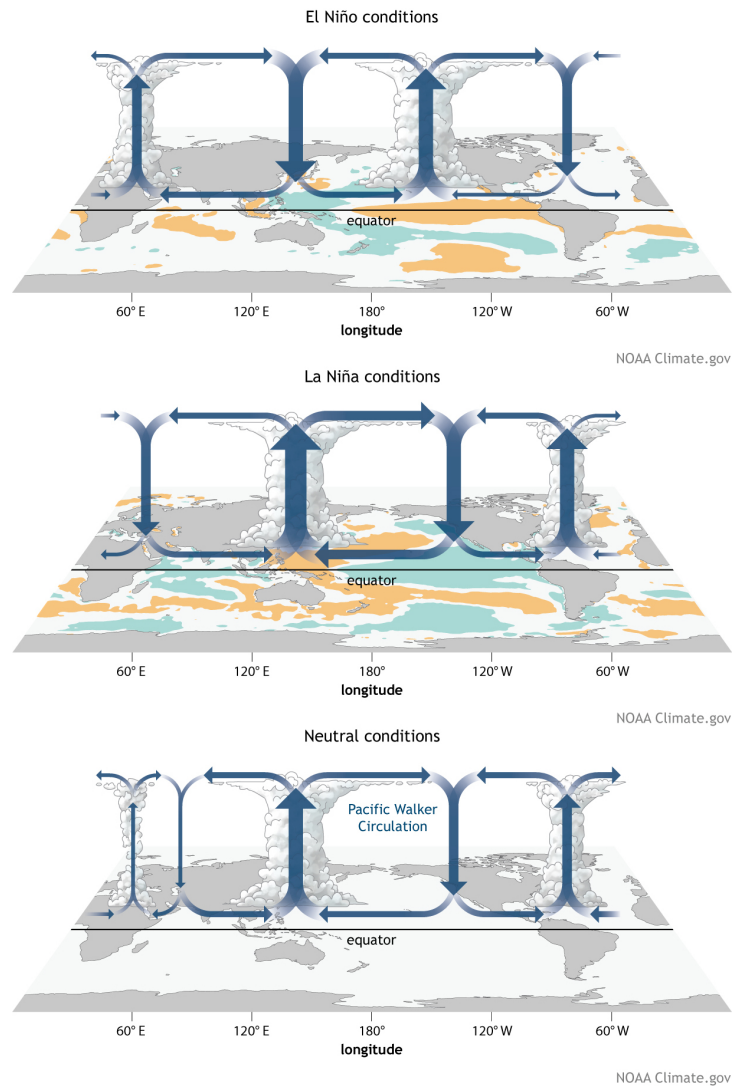


Figure 3.1: Generalized Walker Circulation pattern in different ENSO states. (Di Liberto, 2014)

The Walker circulation is broadly driven by differences in mean land and ocean temperatures. This is amplified by an ocean-atmosphere feedback: pressure gradients (largely in place from atmospheric Rossby wave flow) driving the trade winds, and the trade winds forcing warm surface waters towards the coast, thus mitigating upwelling along that coastline and further encouraging the pressure gradient with the warm waters contributing to convective, unstable air.

A variety of related sub-indices are used to help define and characterize an El Nino event, partially due to different countries having varying effects and definitions for an El Nino event (in other words, there is no globally standardized El Nino index). Despite a variety of different indices that capture ENSO behavior, including the Nino 3.4, TNI, ONI, BEST, SOI (see: <https://psl.noaa.gov/data/climateindices/list/>) the most common definition of the onset of an El Nino / La Nina event is five consecutive months of a three month moving average of SST anomaly (centered 30-year base periods updated every 5 years) being greater (El Nino) than 0.5 °C or less (La Nina) than 0.5 °C.

Event duration for El Nino events is typically 9-12 months and for La Nina events, up to two years. The difference in duration between the two events not currently understood. Frequency of each event spans approximately two to seven years, which is longer than many other identified tropical teleconnections, but smaller than other large scale modes that change phases on a decadal scale.

3.2.2 Pacific Decadal Oscillation

The Pacific Decadal Oscillation (PDO) emerged as an additional indicator of climate variability in the late 1990s when researchers noted relationships between Alaskan salmon production, ocean temperatures, and climate (Mantua *et al.*, 1997). It is

similar to ENSO in that PDO measures climate in the Pacific, although PDO is only a measure of SST anomalies. Additionally, the PDO represents changes in SST over the mid-latitudes rather than the tropics. It is formally calculated as the time series of the principal component analysis (PCA) of the standardized mean SST anomalies in the east and west Pacific Ocean above 20 degrees north. Unlike ENSO, its phase shift is on a decadal scale of approximately 20 to 30 years. One of the major mechanisms for the slower phase shift is the significantly reduced speed of oceanic Rossby waves in poleward latitudes. A time series demonstrating the frequency, magnitude and duration of historical ENSO and PDO events is shown below (3.2).

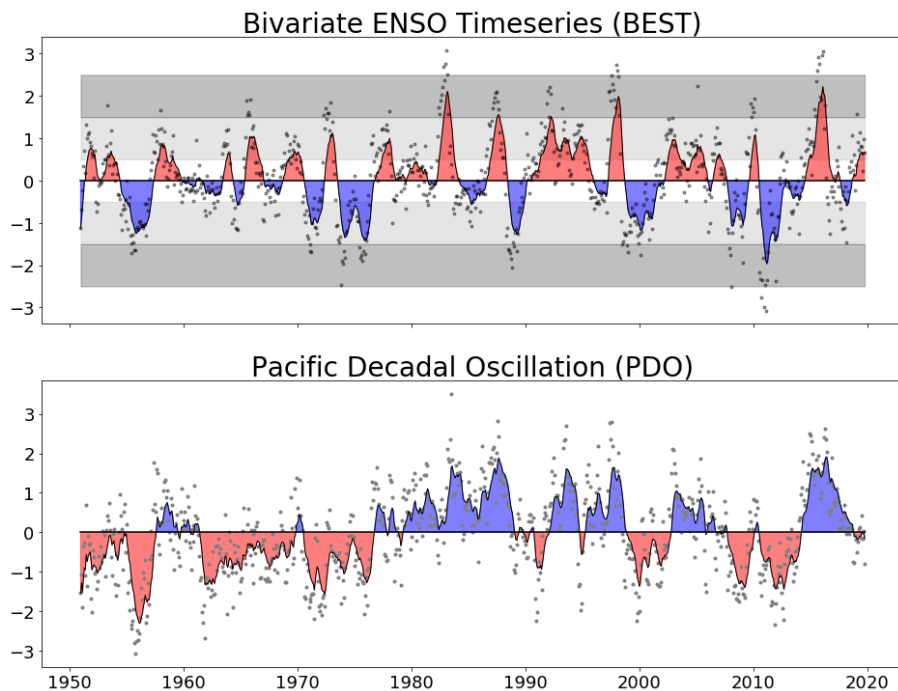


Figure 3.2: Historical ENSO and PDO events. Mild/moderate events shown in light grey with extreme events in dark grey. Raw signal represented by points.

Also unlike ENSO, the PDO index does not formally define the duration of a phase event similar to an ‘El Nino’ or ‘La Nina’ event that is brought on by meeting

sequential thresholds through time. Subsequently, its phase is determined by the aggregated monthly anomalies being either positive (warm in the western pacific) or negative (cool in the western pacific). As a result of quasi-regional weather events in the Pacific, the phase of PDO switches at intervals significantly shorter than a decadal time scale. In other words, the PDO signal is noisy and itself contains signals with varying frequencies. This idea proves useful for our purposes though, as we are interested in linking regional events that may be linked to both the low and high frequency signals.

3.2.3 Climate of the Snake River Valley

Our objective is not only to identify correlations between climate in the SRVAVA and large scale teleconnections, but to also utilize the teleconnections to forecast the regional climate at a seasonal scale. As most of the low-frequency signals we identified in chapter two were metrics derived from temperature, it is appropriate for us to start with temperature as the primary metric to be forecasted. As teleconnection indices are produced by NOAA's Earth Systems Research Laboratory at the monthly time scale (although computed from daily values), we must upsample our climate data to match the monthly time scale. Additionally, as climate indices are normalized anomalies, we remove the seasonal signal and noise of regional temperature which is discussed further in the methodology section.

3.3 Properties and Relationships of Climate

Signals

The low frequency signals present within teleconnection indices inherently imply a high degree of autocorrelation in the signals. Autocorrelation refers to the corre-

lation coefficient between a signal and the same signal at preceding/proceeding time intervals. Figure 3.3 demonstrates the high autocorrelation found in both the SOI and PDO index, which persists for much longer in the PDO than SOI, as expected for a slower oscillating signal. SOI is clearly correlated for up to approximately one year, and PDO over two years. Correlation exists for the SRVAVA regional climate anomalies as well, but is only present for two months. Additionally, when autocorrelations are extended throughout the length of the data set, clear oscillations in the total signal are also observed and function similarly to a low pass filter.

Autocorrelations of PDO, SOI, and SRVAVA Temperature from 1988 - 2017

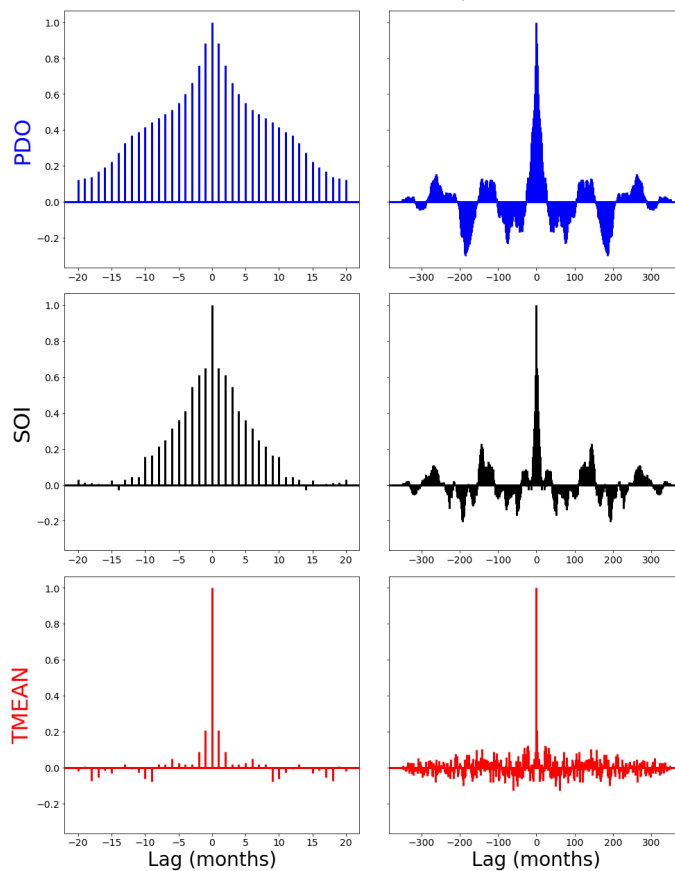


Figure 3.3: Unfiltered autocorrelations. PDO has significantly longer memory than ENSO and regional temperature.

Cross-correlation reflects the degree of correlation of one signal to a different signal as a function of time displacement (technically, autocorrelation is a cross-correlation with itself). Figure 3.4 presents the cross-correlations between PDO and SOI themselves, as well as their respective correlation to SRVAVA temperature. Between SOI and PDO, represented as $\text{PDO}(t)$ vs $\text{SOI}(t \pm n)$, we see strong negative correla-

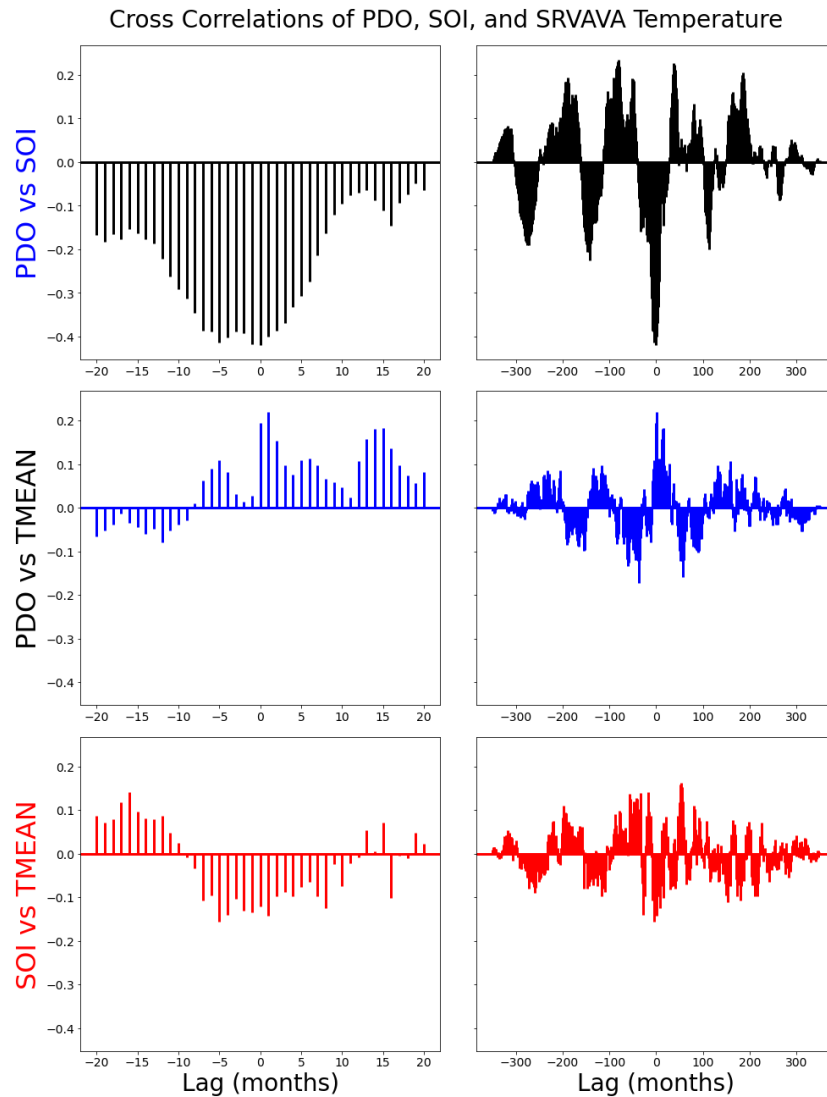


Figure 3.4: Unfiltered cross-correlations. The Southern Oscillation has a strong influence on PDO.

tions (the fact they are negative is only a product of how the teleconnection anomalies were defined). Of particular interest is the much stronger and longer lasting correlations in the negative direction, indicating that PDO is partially, dynamically driven by the ENSO/SOI signal, which has been thoroughly examined in other research as well (Newman *et al.*, 2003). However, the lingering existence of some correlation in the positive direction could be an indicator that SOI/ENSO is also influenced by PDO. Correlation between the teleconnections and regional temperature, $\text{PDO}(t \pm n)/\text{SOI}(t \pm n)$ vs $\text{TMEAN}(t)$, are of a lower magnitude, but still clearly present.

Although there are clear correlations between the signals, they are of relatively small magnitude that could potentially be driven by an excess of noise or other higher frequency signals. Figure 3.5 demonstrates the relatively weak correlation between all three components: PDO, SOI, and TMEAN. The negative linear correlation between each respective teleconnection is present which also aligns with the magnitude of regional temperature anomalies; high magnitude events with a large PDO and low SOI, and vice versa. However, there is some clear overlap and many events happening in the center or neutral state as well.

To evaluate the correlation of the signal with a reduction of noise (or high frequency signals), we applied a low-pass filter to the PDO signal and compared the time series of it to that of the SRVAVA climate which shows significantly higher correlation between the two signals of nearly 0.7 (figure 3.6).

The Comparison shows a significantly higher correlation between the two signals of nearly 0.7. This provides evidence that certain frequencies of the large-scale signals may be more useful than the raw signal itself, and a combination of them may help better forecast or historically reconstruct other climate signals. For example, there is a

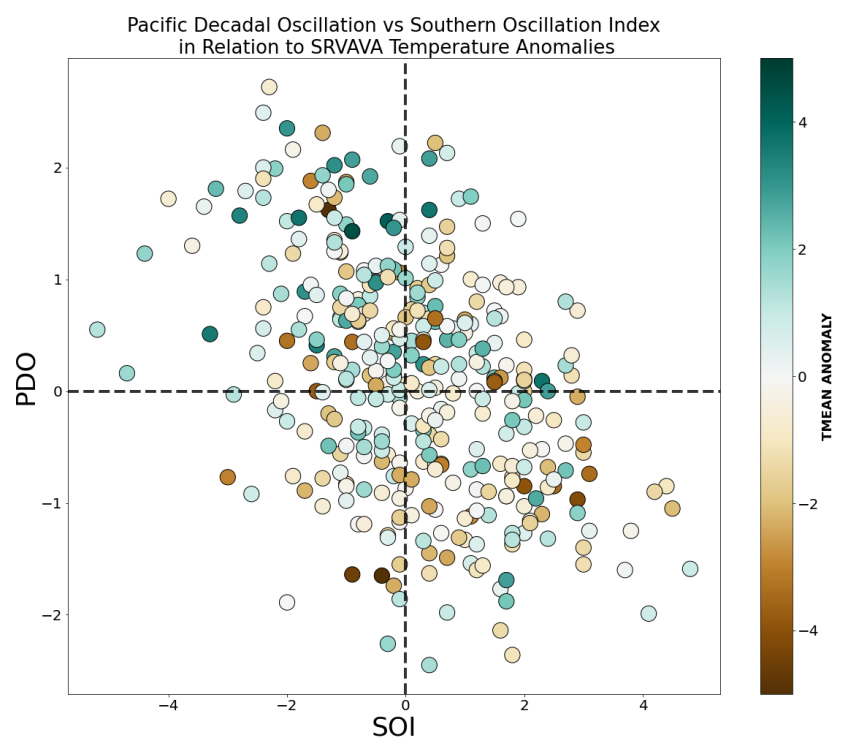


Figure 3.5: Unlagged relationship between teleconnection states and regional temperature.

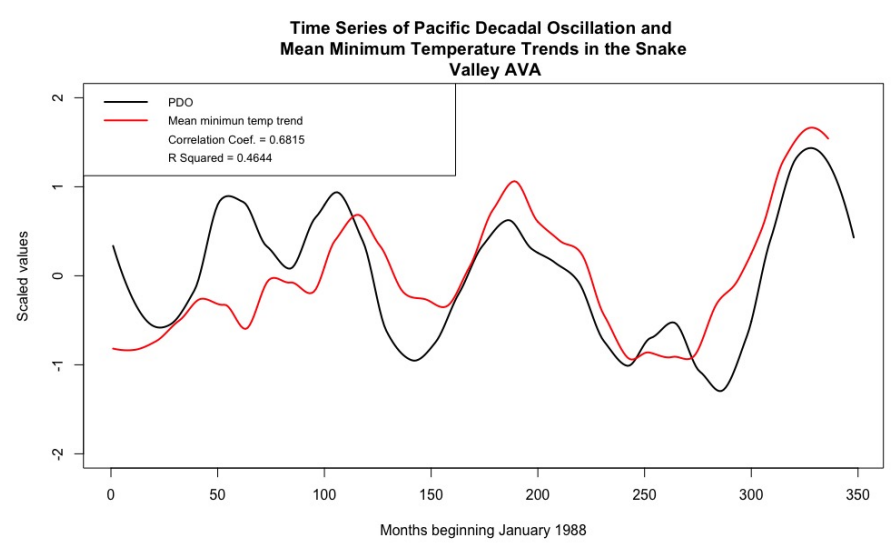


Figure 3.6: Unlagged relationship between PDO and regional minimum temperatures. Lagged correlations are inconsistent.

clear lag present in the PDO to temperature signal present throughout the time series, but the length of the lag is clearly inconsistent (as noted by longer lags earlier in time). This problem of inconsistent lags and correlations serves as the primary motivation to use an LSTM network for representing sequences of data in time, in conjunction with a deconstructed large-scale signal using Empirical Mode Decomposition.

3.4 Methodology

Although many relationships have been identified between various teleconnections and regional phenomena in the Pacific Northwest, efforts to use teleconnections to accurately forecast regional phenomena have been limited. Contributing factors that complicate this goal include inconsistent lag periods of correlations (most regression models require static lag periods) and noise or varying frequencies of the large scale signal weighting models that don't accurately capture meaningful signals. We address these issues by 1) utilizing a long short term memory (LSTM) neural network - a special type of recurrent network - that can represent long sequences of time series data and properly represent varying lag periods, and 2) decompose the teleconnections into multiple signals with an empirical mode decomposition (EMD) to be input as individual features (and thus weight the model based on multiple frequencies).

3.4.1 Long Short Term Memory Model Overview

Artificial neural networks (ANNs) are a popular tool to model highly non-linear and complex systems. However, when ANNs are applied to time series data, no sequential information is retained. The importance of the prior sequential information, or memory of the signal, can easily be demonstrated with a natural language processing (NLP) problem, the field where LSTMs first gained popularity. For example,

take the sentence, “I lived in France as a teenager, where I studied the language and became fluent in -----.” It is easy for a human to make this prediction of the word ‘French’ based on the context of living in France. In other words, our thoughts have persistence. ANNs have no persistence, thus a recurrent neural network (RNN) was developed to address the lack of persistence. In this example, the word French must be primarily informed by the word France 13 sequence steps prior.

Figure 3.7 is a common diagram to represent the architecture of a basic feed-forward neural network, where each input layer is connected to each hidden node, and each hidden node connected to each output layer, all propagating forward. If each input layer were considered a separate time step in this scenario, the important takeaway is that there is no exchange of information between the hidden layers themselves, resulting in no persistence of the sequence of information.

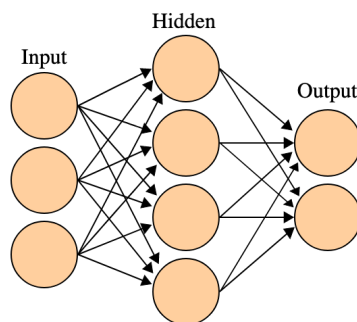


Figure 3.7: Basic ANN Structure.

Figure 3.8, on the other hand, shows a simplified unrolled RNN that connects each time step by each cell passing information to both the output cell and the adjacent sequential cell allowing the sequence to persist. Thus, each hidden cell after the first of the entire sequence is informed by two inputs rather than one.

Although RNNs have been developed for memory persistence, the length of their

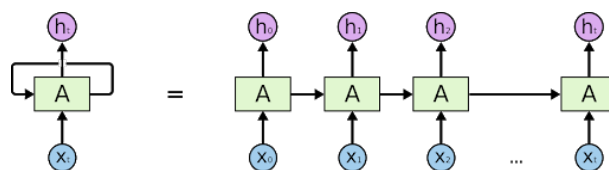


Figure 3.8: Basic RNN Structure.

meaningful memory is less than 10 steps long due to the vanishing gradient problem (Bengio & Simard, 1994). Most neural networks are trained using stochastic gradient descent which uses the current prediction error of the model to estimate a gradient to more accurately reflect weight changes with each iteration. This error gradient is then back-propagated from the output to the input. Since gradients are calculated based on their partial derivatives and propagated via the chain rule, the magnitude of the gradients diminish rather quickly and become close to zero after only a few steps away from the output, thus having a “short memory.”

LSTMs were designed specifically to increase the memory length of RNNs. Each node within the hidden layer is replaced with an LSTM cell, which has an additional output referred to as the cell state. The cell state is a vector influenced by three specific gates: a forget gate (f_t), input gate (i_t), and output gate (O_t), all with their own sigmoid activation function. Each gate can conceptually be thought to act as a binary switch as determined by its respective sigmoid function: the input gate is used to control whether the memory cell is updated or not, the forget cell is used to control if the vector is reset to zero, and the output gate is used to control whether the cell state is visible or not (figures 3.9 and 3.10). Thus, it is the cell state vector itself that contains the memory of previous time steps at any length (up to the full length of the sequence fed as input), and thus unaffected by a vanishing gradient. Furthermore, it allows the state to only use the past memory state when determined

it is useful. A detailed explanation of the architecture of an LSTM is outside the scope of this thesis, however many resources exist that cover this topic in detail.

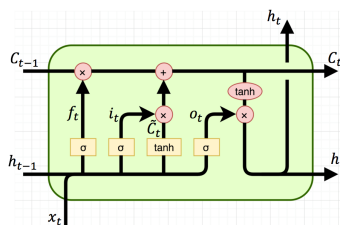


Figure 3.9: LSTM Cell State.

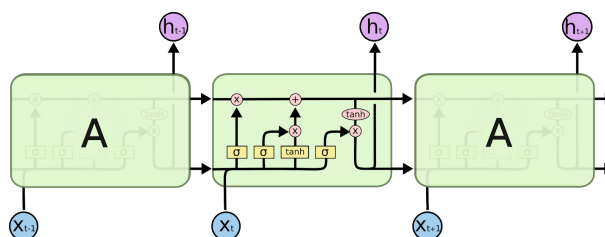


Figure 3.10: Basic LSTM Structure.

Although use of LSTMs gained popularity in the NLP domain for speech recognition, use of LSTMs within the earth sciences, and hydrology specifically, has grown substantially in recent years. Kratzert *et al.* (2018) demonstrated the effective use of an LSTM as a hydrological model to predict daily streamflow in a variety of basins. Conceptually, streamflow hydrology is an excellent use case for LSTMs as there is long persisting memory at varying temporal scales. By using sequence lengths of 365 days, the LSTM cell can begin to learn seasonal patterns such as snow melt dynamics, as well as lagged delays in snowmelt runoff. Additionally, they demonstrated snowmelt dynamics can be visualized by plotting the LSTM cell state values in conjunction with mean temperature, which helps address the large “black box” criticism of neural networks (figure 3.11).

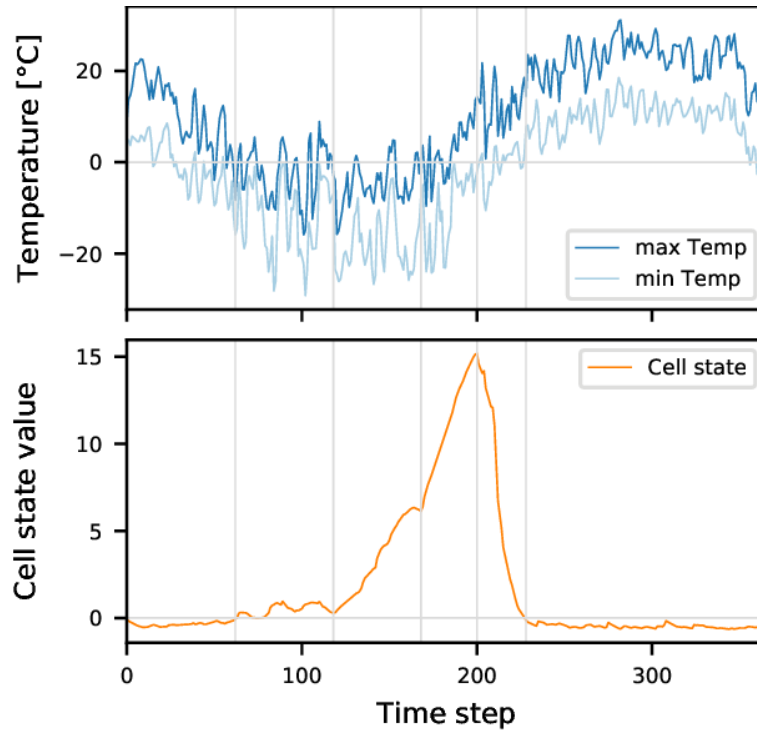


Figure 3.11: Snowmelt dynamics represented by the cell state value of an LSTM network (Kratzert *et al.*, 2018)

As lag times appear to be inconsistent for relationships between large scale climate indicators and regional climate, and further complicated by noise or other higher frequency signals, LSTMs are a good candidate for exploiting the persistent temporal relationships between them. However, trials to map the raw teleconnection data alone (as the only input) to regional climate via LSTM were unsuccessful, possibly due to an abundance of noise. Thus, we chose to first decompose the large scale climate data in an effort to inform our LSTM with varying frequency signals.

3.4.2 Empirical Mode Decomposition

Empirical mode decomposition (EMD) is an adaptive method of breaking down a time series signal into multiple signals of equal length without leaving the time

domain. The resulting signals, referred to as intrinsic mode functions (IMFs), sum to form the complete signal and do not rely on any pre-existing assumptions about the data. Other methods to break down a signal exist such as singular spectrum analysis and wavelet and Fourier transforms, but do not offer the finite oscillation time scale components that can often directly represent different processes that influence the signal as a whole (Lambert *et al.*, n.d.). Each IMF represents information on a different scale of the original data and must meet the following criteria: 1) throughout the entire raw signal, the number of zero-crossings must be not higher than one more of the number of extrema; 2) at any point the mean of the envelope determined by the extremas must be zero. The general sifting procedure used to create a complete set of IMFs is as follows (figure 3.12):

1. Identify all local minima and maxima.
2. Use a three-spline interpolation function to create the upper and lower envelopes.
3. Calculate the mean values of the upper and lower envelopes designated as $m_1(t)$.
4. The difference, $h_1(t)$, between the signal $x(t)$ and the mean $m_1(t)$, expressed as $h_1(t) = x(t) - m_1(t)$, is regarded as the primary definition of the first IMF.
5. To determine the first IMF more accurately, $h_1(t)$ is treated as a new signal, its upper and lower envelopes, and their new mean $m_2(t)$ are calculated, and a new difference $h_2(t) = h_1(t) - m_2(t)$ is determined. This $h_2(t)$ is again treated as a new signal, iterated by k until the IMF criterion satisfies. $h_k(t)$ is the first IMF, designated by $H_1(t)$.

6. The first residual, $d_1(t) = x(t) - H_1(t)$, is calculated with steps 1 through 4 to obtain the second IMF $H_2(t)$ and is repeated until the last residual shows little to no variation.
7. The number of IMFs for a given signal is thus variable depending on the complexity of the original signal.

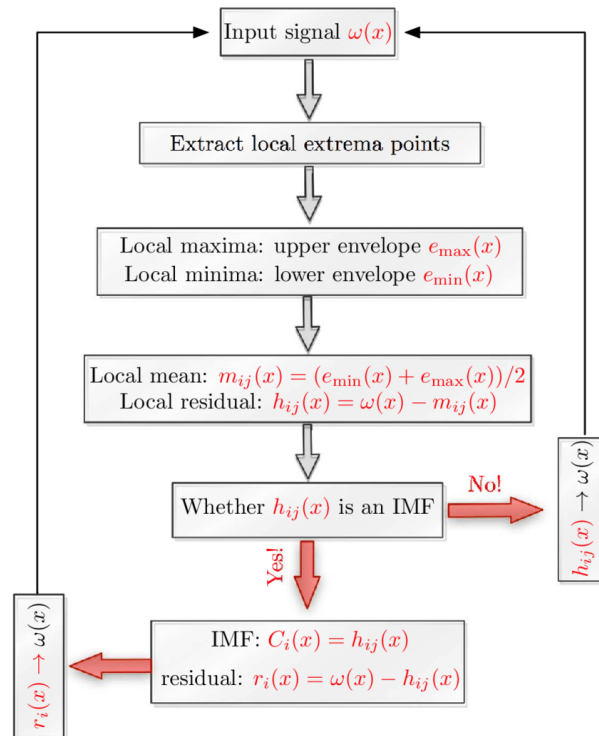


Figure 3.12: Process to generate IMFs through EMD.

3.4.3 Data Preparation

As mentioned earlier, the ENSO signal is composed of both SST and SLP components, and PDO is composed of only an SST component. We chose to only incorporate the SLP component, the Southern Oscillation Index (SOI), of ENSO into our model

for two primary reasons: 1) it is known to be highly correlated to the SST signal of ENSO via the positive ocean-atmosphere feedback, and thus still broadly represents the overall process; 2) to attempt to reduce the correlation between PDO and ENSO by using two separate components that may better represent the influence of each index. We incorporate the EMD algorithm to both the SOI and PDO monthly signals from October 1987 to September 2017 producing 9 IMFs for each signal (figure 3.13) to be used as input to our model.

As the primary goal of this modeling effort is to determine if we can successfully map large scale teleconnection data to regional climate in a way useful for forecasting, we chose to only use the trend component of our regional temperature data by removing the noise and seasonal component. We accomplish this by applying an additive time series decomposition method that decomposes the series into three components: seasonal, trend, and residual that all sum to the original signal. Additive methods, as opposed to multiplicative methods, are utilized when there is not an underlying trend that is proportional to the entire level of the time series. In other words, if there is a relatively consistent trend throughout the time series, a multiplicative method is preferred. As climate change could potentially be considered a continuous trend, we chose to address this by removing the linear trend in temperature first and then applying an additive decomposition.

Seasonality is computed by taking the mean temperature for each month of the year over the entire record, and then subtracting it from the absolute mean monthly temperature. The trend component is calculated by a simple centered moving average based on the frequency of the data within a period (12 months), thus taking an average of the first 12 values and iterating by one throughout the entire index. As a

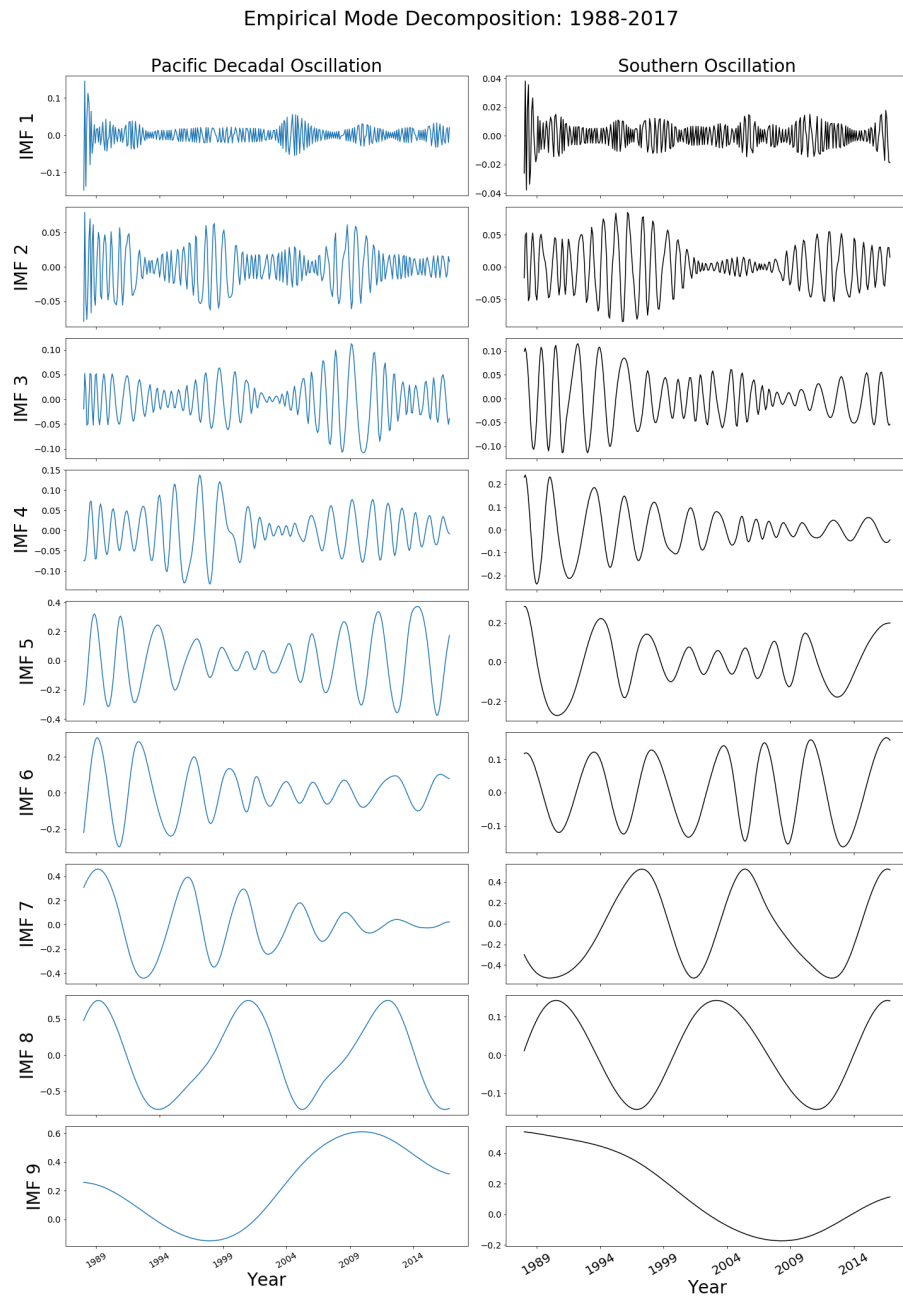


Figure 3.13: Intrinsic Mode Functions of SOI and PDO.

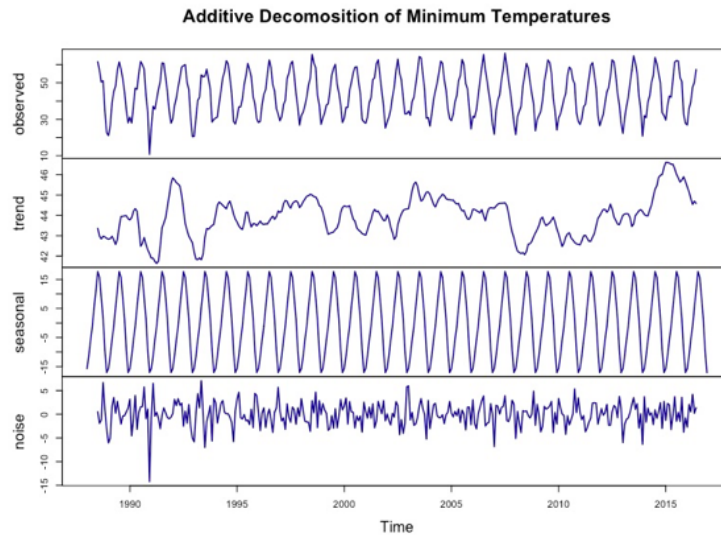


Figure 3.14: Example of additive decomposition regional temperature.

result (figure 3.14), you will end up with a time series that has a length shorter than the original signal by one frequency (6 months truncated off each end for monthly data with a frequency of 12). The resulting residual, or noise, is then represented as $residual = original\ signal - trend - seasonal\ component$. This results in 348 months (April 1988 - May 2017) of mean aggregated temperature trends in the SRVAVA to be used as the dependent variable in our model. To match the length of our input and output data, we additionally truncate the first and last six months of teleconnection data. Lastly, as most neural networks perform optimally with scaled data, we scale all 30 years of data to values between -1 and 1. Our full data preprocessing procedure is as follows (figure 3.15):

1. Upsample daily aggregated temperature data to mean monthly data for our 30 year regional data in the SRVAVA
2. Apply a simple linear model to all 360 months, and remove the linear trend

3. Apply additive decomposition and pull only the trend component
4. Truncate the first and last six months of large scale data to match the regional temperature data
5. Scale all data (independently) from -1 to 1

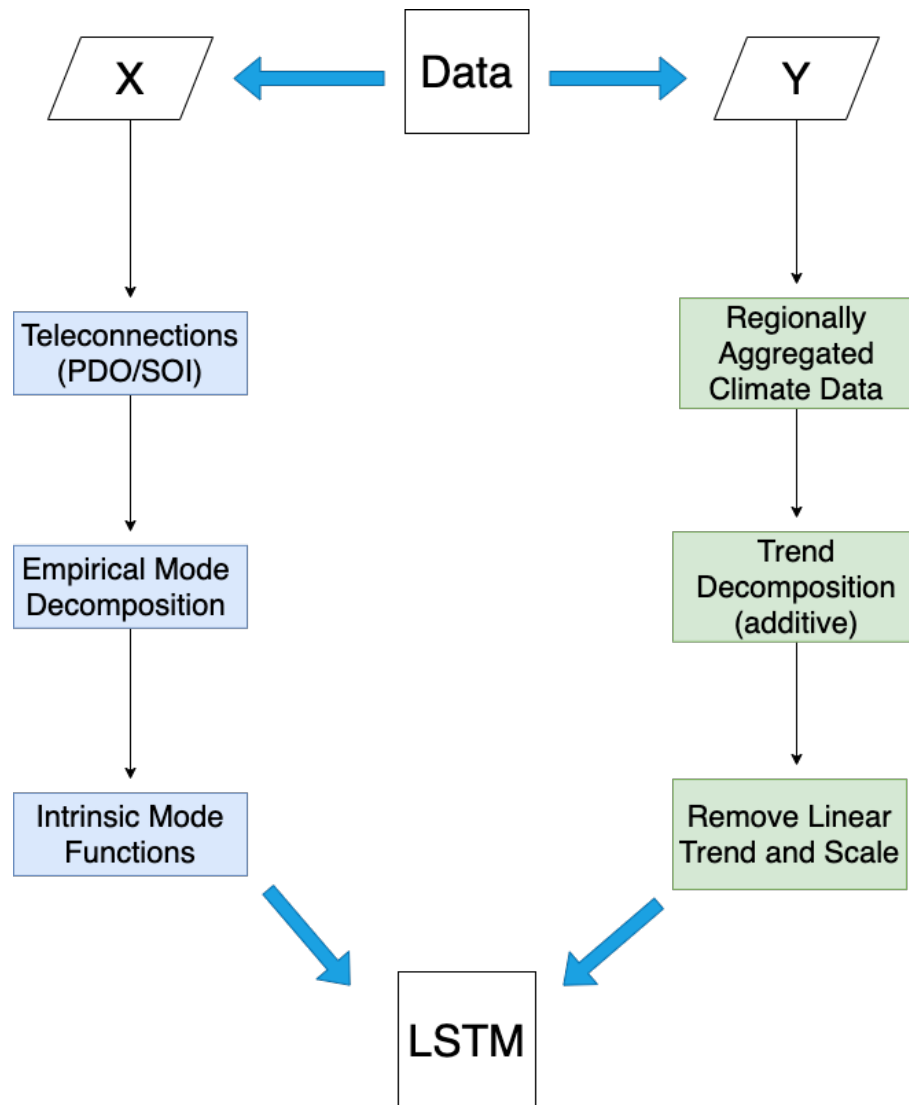


Figure 3.15: Conceptual data preprocessing flow chart.

3.4.4 Model Framework and Results

The hyperparameter tuning of neural networks is a very active area of research, as each model and the accompanying data pose unique challenges. Although our model is relatively light in respect to data volume, the focus is not necessarily on optimal performance, but more evidence that LSTMs can be useful in climatic forecasting. Thus, a rigorous hyperparameter tuning scheme was not the main priority of this research. A trial and error method of tuning hyperparameters was utilized for the most commonly impactful parameters.

The model was developed using Python with the ‘keras’ deep neural network module with a Tensorflow backend. We trained the model using the ‘Adam’ optimizer algorithm with a learning rate of 0.008 over 40 epochs and a batch size of 192. We used a single LSTM layer of size 50 (nodes), and found the optimal sequence length to be 6 (months). Other parameters not listed were used as the default ‘keras’ values and the model was evaluated with mean squared error function (MSE).

With only 348 monthly data points, we chose to separate our data into only training and testing sets - as opposed to training, validation, and testing - to supply as much training data as possible. With sequential time series data (vector inputs that require sequential time steps), traditional randomized cross-validation techniques are not possible as it is necessary to have one continuous length of data for both training and testing. As a response to this issue, we choose to perform two experiments split by 25 years of training and 5 years of testing. The first uses 1994 - 2017 as training and 1988 - 1993 as testing, and the second using 1988 - 2012 as training and 2013 - 2017 as testing. We run the model with the dependent variable lagged by periods of 1, 3, 6, 9, and 12 months. Additionally, we choose to run eight ensemble runs per

experiment to better constrain the uncertainty of each model.

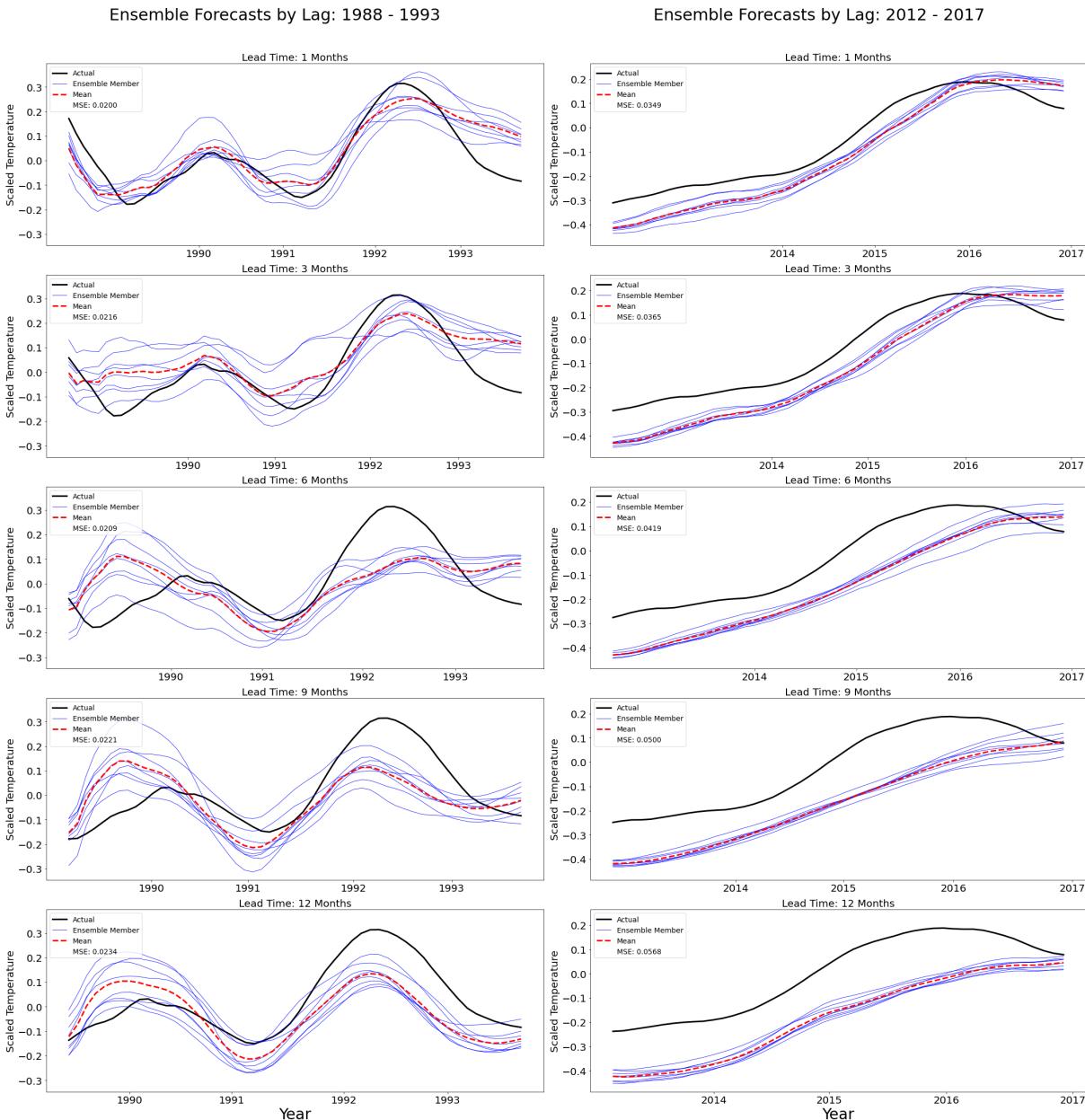


Figure 3.16: SRVAVA Temperature trend forecasts with varying lead times in two separate scenarios.

Results indicate good pattern adherence at all lead times, up to one year (we did

not test at further lead times), with general decay in performance with increasing lead time, as expected (figure 3.16). The testing period from 1988 - 1993 had multiple local maximas/minimas which made for a more interesting signal to model and caused a larger variance between ensemble runs. This appears to be a good use case for ensemble runs, as the mean of the ensemble runs has a lower RMSE (not shown) than the majority of individual runs. There is some disagreement on the 6-month lead time run in the early test set, which could relate to the chosen sequence length of six months representing some seasonal differences in SOI or PDO. We also tested using longer sequences (up to two years), however, the overall best performance was achieved with a six month sequence. The pattern of the more recent testing set was also captured at all lead times including small variance in the ensemble runs, likely as a result of the fairly simple signal.

3.5 Discussion and Future Work

Over the last couple of decades many correlations between climate teleconnections and agriculturally relevant regional climate events have been identified. However, the large amount of noise present in teleconnection signals pose challenges when attempting to use them to forecast regional climate at the monthly time scale. Our attempts to forecast temperature trends within the SRVAVA using a LSTM network with raw SOI and PDO time series as the sole input were unsuccessful, which led us to the EMD algorithm to decompose PDO and SOI into a collection of signals that may better inform our model.

Efforts to combine the use of LSTM in conjunction with EMD as a mechanism to forecast climate time-series has been previously studied (Zhang *et al.*, 2018). However, these methods have generally been used to forecast time-series data using prior

time steps of the same input (decomposed and reconstructed with EMD) to improve performance compared to other long established methods such as ARIMA models. Our model aims to accurately map large-scale climate data to regional climate data without using previously known regional data to infer a useful relationship between data at different scales.

Results indicate a strong pattern representation of regional climatic trends for two test cases with lead times up to twelve months using 30 years of decomposed monthly climate data. The EMD decomposition technique greatly improves accuracy that may indicate multiple important processes being represented by varying frequencies within large-scale teleconnections. Additionally, the use of sequential series of both PDO and SOI can help to discern the importance of the relationship between PDO and SOI from a modeling perspective. We believe this novel approach of using EMD on large-scale phenomena to inform and forecast changes to a smaller spatial scale can be a useful building block for future climate forecasting.

Practical applications of this climate model could help inform agricultural decisions such as fungicide and pesticide spray timing, irrigation strategies, and vine canopy management. Such actions could potentially increase yield and/or quality of the product, reduce overall risk of fungal and physical damage by extreme temperatures, and possibly reduce operating costs by more efficient spray timing and irrigation.

There are many potential extensions of this work including: more extensive hyperparameter tuning, analysis of the relative impact of each IMF frequency and its physical representation within climate dynamics, a longer time-series for training and testing, and testing on other important metrics other than temperature. Lastly, EMD

could be replaced with Ensemble Empirical Mode Decomposition (EEMD) - a method of adding white noise to standard EMD IMFs to better constrain the true signal.

CHAPTER 4:

CONCLUDING SUMMARY

Global climate change has increased stress on the agricultural sector. Changes in the magnitude, frequency, and spatial distribution of hazardous weather and climate events, as well as conditions that allow pests and pathogens to thrive, threaten the yield and quality of global crop production. Although mitigation of climate change is perhaps the most important challenge to overcome, some agricultural adaptation at local scales can simultaneously capitalize on the warming climate.

Viticulture within Idaho's Snake River Valley American Viticultural Area (SR-VAVA) is an example of a relatively new, local agricultural sector that has grown rapidly in recent years. Favorable climate conditions compliment well-drained soils, and sloping hills found in the area provide substantial opportunities for vineyards. To examine changes in overall agricultural suitability and risk we used a 30-year, high resolution climate dataset generated by the Weather Research and Forecasting (WRF) model to characterize spatiotemporal changes in seasonal agroclimate, as well as the magnitude and frequency of anomalous, short-term weather events impactful to viticulture.

The high-resolution of our dataset (1 km spatial and 1 hour temporal) provides a number of novel applications, specifically with the complex valley terrain of the SRVAVA. Such novelties include a more accurate way to calculate growing degree

days, multiple metrics that can only be derived with hourly data, and a better spatial representation of cold airflow across the valley. A statistical characterization of changes in agroclimate metrics across 30 years found most temperature driven metrics have changed in a way that is beneficial to viticulture. These include a substantial increase in GDD, an increase in mean diurnal temperature, a longer mean growing season, and a reduction of total hours spent in hazardous cold temperatures. Other neutral or disadvantageous changes include a later average spring frost, possible increase in severity of freeze events in the lower valley, and increased duration of conditions where powdery mildew thrives.

Different grape cultivars have a wide range of climate and weather related characteristics such as growing degree day demands, susceptibility to powdery mildew, and cold-hardiness. Thus, insight from this analysis may be useful to better determine site cultivar selection and a better evaluation of risk from powdery mildew and extreme temperatures which are common near the base of the valley.

In addition to the analysis provided in this thesis, much of the data we used is provided to the public and can be explored with use of an interactive web-viewer we developed at <http://bwc.boisestate.edu/climate-viewer> (see Appendix A). Use of this tool can facilitate exploring a variety of metrics at specific locations within the Snake River Valley suited to the specific interests of the viewer. This tool also provides the ability to download annual metrics and statistical summaries at any SRVAVA location.

In Chapter 3 we also noted clear, low-frequency signals present in both annual mean metrics and frequency of anomalous weather events that are weakly correlated to large-scale climate indicators SOI and PDO. Building upon these correlations, we

introduce a novel forecasting method using Empirical Mode Decomposition (EMD) to decompose raw SOI and PDO data into a larger collection of time-series signals as input into a Long Short Term Memory (LSTM) neural network to forecast regional temperature trends in the SRVAVA. Results show a significant improvement in modeling power when decomposing large scale climate data with EMD and provide accurate temperature trend forecasts with lead times up to one year.

In complement to Chapter 2, which aims to inform the overall agricultural suitability and risk over long periods, this data-driven model may be useful to inform agricultural management decisions associated with short-term risk due to seasonal to sub-seasonal changes in climate. Such applications may help inform changes to irrigation strategy, canopy management, and/or spray timing.

It is worth noting the interaction and roles of both physically-based models and statistically-driven forecasting methods. This thesis utilized a physically-based model (WRF) to “forecast” historical climate observations at significantly better spatiotemporal resolutions, which in turn provided us with a historical dataset to be used as the dependent variable in a statistically driven forecast. Independently combining both types of forecasting methods can provide certain benefits. One benefit is that physically-based models only need initial boundary conditions (and no historical data) to generate the missing, incomplete, inadequate, or impossible to collect data that statistical models depend on for training. A second key benefit to independently combining approaches is the ability of a statistical model to mimic relationships that we don’t have a great physical understanding of, which many exist within the food-energy-water nexus.

Although our example of generating temperature data with a physically based

model to be used within a data-driven model is fairly simple, it is easy to imagine other cases of combining modeling techniques that would be useful for the agricultural industry. For example, the forecasting of evapotranspiration (ET), which is agriculturally desirable yet notoriously difficult to collect observationally. Using a regional climate model to generate ET (WRF provides all the necessary energy output to derive ET) could then be used as training data for a light-weight geostatistical forecasting model of ET.

Many additional extensions of this work are possible. One natural extension would be to continue to run additional WRF simulations (same domain setup) to the respective present time to be used as analysis or forcing data for future agroclimate projects. Subsequently, additional data could be continuously provided to the public via the climate web-application we developed, along with additional analysis features. Other opportunities include testing our modeling framework with a variety of agroclimate metrics, robust hyperparameter tuning, and investigating statistical relationships between large scale IMFs and physical processes.

REFERENCES

- Alcohol and Tobacco Tax and Trade Bureau. 2020. *American Viticultural Areas (AVAs)*. <https://www.ttb.gov/wine/american-viticultural-area-ava>. Accessed: 2020-02-15.
- Bengio, Y, & Simard, Patrice Y. 1994. Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks*, **05**(02).
- Challinor, A J, Watson, J, Lobell, D B, Howden, S M, Smith, D R, & Chhetri, N. 2014. A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, **04**(04), 287–291.
- DeFlorio, Michael J., Pierce, David W., Cayan, Daniel R., & Miller, Arthur J. 2013. Western U.S. extreme precipitation events and their relation to ENSO and PDO in CCSM4. *Journal of Climate*, **26**(12), 4231–4243.
- Di Liberto, Tom. 2014. *The Walker Circulation: ENSO's atmospheric buddy*. <https://www.climate.gov/news-features/blogs/enso/walker-circulation-ensos-atmospheric-buddy>. Accessed: 2020-03-03.
- DR, Haviland, & Bettiga LJ, Varela LG, Baldwin RA Roncoroni JA Smith RJ Westerdahl BB Bentley WJ Daane KM Ferris H Gubler WD Hembree KJ Ingels CA

- Zalom FG Zasada I. Revised continuously. UC IPM Pest Management Guidelines Grape. UC ANR Publication 3448. Oakland, CA.
- Flores, Alejandro, Masarik, Matt, & Watson, Katelyn. 2016. *30-Year, Multi-Domain High-Resolution Climate Simulation Dataset for the Interior Pacific Northwest and Southern Idaho [Data set]*. <http://dx.doi.org/10.18122/B2LEAFD001>.
- Gillerman, Virginia S., Wilkins, David, Shellie, Krista, & Bitner, Ron. 2006. Geology and wine 11. Terroir of the Western Snake River Plain, Idaho, USA. *Geoscience Canada*, **33**(1), 37–48.
- Goodrich, Gregory B. 2007. Influence of the Pacific Decadal Oscillation on Winter Precipitation and Drought during Years of Neutral ENSO in the Western United States. *Weather and Forecasting*, **22**(1), 116–124.
- Gornall, Jemma, Betts, Richard, Burke, Eleanor, Clark, Robin, Camp, Joanne, Willett, Kate, & Wiltshire, Andrew. 2010. Implications of climate change for agricultural productivity in the early twenty-first century. *Philosophical Transactions of the Royal Society B*, **365**(1554), 2973–2989.
- Gu, Sanliang, Sabuwala, Adnan, & Gohil, Hemant. 2014 (07). Comparison of Growing Degree Hours Based on Hourly Average Temperatures with Growing Degree Days Based on Daily Minimum and Maximum or Average Temperatures to Interpret Heat Summation.
- Hayman, Peter, McCarthy, Michael, Thomas, Dane, & Longbottom, Mardi. 2012. Managing Grapevines During Heatwaves Factsheet. *Australian Grape and Wine Authority*, 1–6.

Idaho Wine Commission. 2018 (Sep). *Economic Impact Study, 2018*. Tech. rept.

Idaho Wine Commission. 2020. *Idaho Wine - 150 Years in the Making*. <https://idahowines.org/about/history/>. Accessed: 2020-02-15.

Jackson, Ronald S. 2014. *Wine Science. Principals and Applications. 4th Edition*. Elsevier.

Jones, Gregory V. 2003. Paper 4 from the “Terroir, Geology and Wine: A Tribute to Simon J. Haynes” Session held at the Geological Society of American Annual Meeting, Seattle, Washington, November 2, 2003.

Jones, Gregory V, Duff, Andrew A, Bitner, Ron, Vineyards, Bitner, Wilkins, David, Shellie, Krista, & Shatz, Moya. 2010. *The Climate and Landscape Potential for Quality Wine Production in the Snake River Valley AVA*. Tech. rept. Washington State Dept. of Fish and Wildlife Mill Creek.

Kalnay et al. 1996. *The NCEP/NCAR Reanalysis 40-year Project*. Tech. rept. National Center for Atmospheric Research.

Kennedy, Caitlyn, & Lindsey, Rebecca. 2014 (January). *How is the polar vortex related to the Arctic Oscillation?* <https://www.climate.gov/news-features/event-tracker/how-polar-vortex-related-arctic-oscillation>. Accessed: 2020-02-01.

Koong, Dennis, Benz, Sue, Sigrist, Bill, Losh, Dave, Huston, Kelly, & Reichert, Jennifer. 2018. *2018 IDAHO ANNUAL STATISTICAL Estimates*. Tech. rept. United States Department of Agriculture National Agricultural Statistics Service Northwest Regional Field Office.

- Kratzert, Frederik, Klotz, Daniel, Brenner, Claire, Schulz, Karsten, & Herrnegger, Mathew. 2018. Rainfall – runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, **22**, 6005–6022.
- Kunkel, M. L., & Pierce, J. L. 2006 (Dec.). The Influence of ENSO and PDO on Idaho’s Snowpack. *Pages GC41A–1047 of: AGU Fall Meeting Abstracts*, vol. 2006.
- Lambert, Max, Engroff, Andrew, Dyer, Matt, & Byer, Ben. *Empirical Mode Decomposition*. <https://www.clear.rice.edu/elec301/Projects02/empiricalMode/index.html>. Accessed: 2019-08-01.
- Li, Shaomeng, Jaroszynski, Stanislaw, Pearse, Scott, Orf, Leigh, & Clyne, John. 2019. VAPOR: A Visualization Package Tailored to Analyze Simulation Data in Earth System Science. *Atmosphere*, **10**(9).
- Liang, Xin-zhong, Wu, You, Chambers, Robert G, Schmoldt, Daniel L, Gao, Wei, Liu, Chaoshun, & Liu, Yan-an. 2017. Determining climate effects on US total agricultural productivity. *Proceedings of the National Academy of Sciences*, **114**(12).
- Lindsey, Rebecca, Dahlman LuAnn. 2020 (January). *Climate Change: Global Temperature*. <https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature>. Accessed: 2020-02-01.
- Maltman, Alex. 2008. The Role of Vineyard Geology in Wine Typicity The Role of Vineyard Geology in Wine Typicity. *Jouranal of Wine Research*, **19**(01), 1–17.
- Mantua, Nathan J, Hare, Steven R, Zhang, Yuan, Wallace, John M, & Francis, Robert C. 1997. A Pacific Interdecadal Climate Oscillation with Impacts on Salmon

- Production. *Joint Institute for the Study of the Atmosphere and Oceans, University of Washington, Seattle, Washington*, **379**.
- Meier, Michael, Fuhrer, Jürg, & Holzkämper, Annelie. 2018. Changing risk of spring frost damage in grapevines due to climate change? A case study in the Swiss Rhone Valley. *International Journal of Biometeorology*, **62**(06), 991–1002.
- Mira de Orduña, Ramón. 2010. Climate change associated effects on grape and wine quality and production. *Food Research International*, **43**(7), 1844–1855.
- Newman, Matthew, Compo, Gilbert, & Michael, Alexander. 2003. ENSO-Forced Variability of the Pacific Decadal Oscillation. *Journal of Climate*, **16**(23), 3853–3857.
- Ouachani, Rim, Bargaoui, Zoubeida, & Ouarda, Taha. 2013. Power of teleconnection patterns on precipitation and streamflow variability of upper Medjerda Basin. *International Journal of Climatology*, **33**(1), 58–76.
- Richard, Yves, Fauchereau, Nicolas, Pocard, Isabelle, Rouault, Mathieu, & Trzaska, Sylwia. 2001. 20th Century Drought in Southern Africa: Spatial and Temporal Variability, Teleconnections with Oceanic and Atmospheric Conditions. *International Journal of Climatology*, **21**(7), 873–885.
- Sacchelli, Sandro, Fabbrizzi, Sara, Bertocci, Marco, Marone, Enrico, Menghini, Silvio, & Bernetti, Iacopo. 2017. A mix-method model for adaptation to climate change in the agricultural sector: A case study for Italian wine farms. *Journal of Cleaner Production*, **166**, 891–900.

- Skamarock, W. C., J. B., Klemp, Dudhia, J., D. O., Gill, D. M., Barker, M. G., Duda, X.-Y., Huang, Wang, W., & Powers, J. G. 2008. *A Description of the Advanced Research WRF Version 3*. Tech. rept. National Center for Atmospheric Research. NCAR/TN-475+STR, 113 pp. doi:10.5065/D68S4MVH.
- State of Idaho Hazard Mitigation Plan. 2018. *Risk Assessment: Drought*. Tech. rept. Idaho Office of Emergency Management. pp. 1–45.
- United States Department of Agriculture National Agricultural Statistics Service. 2017. *Agricultural Statistics 2017*. Tech. rept.
- U.S. Geological Survey. 2002. *Description: Lake Bonneville and the Bonneville Flood*. Tech. rept. Cascades Volcano Observatory.
- Wahl, K. 1988. Climate and soil effects on grapevine and wine : the situation on the northern border of viticulture – the example Franconia. In Proceedings of the Second International Symposium for Cool Climate Viticulture and Oenology. pp 1–5.
- Walthall CL, et al. 2013. Climate Change and Agriculture in the United States: Effects and Adaptation. USDA Technical Bulletin 1935. Washington, D.C.
- Ward, Philip J, Jongman, Brenden, Kummu, Matti, Dettinger, Michael D, & Sperna, Frederiek C. 2014. Strong influence of El Niño Southern Oscillation on flood risk around the world. *Proceedings of the National Academy of Sciences*, **111**.
- Watson, Philip. 2019. *Economic Contribution of Idaho Agribusiness - 2018*. Tech. rept. University of Idaho.

- WMO Climatological Normals. 2020. *World Climate Data and Monitoring Programme*. http://www.wmo.int/pages/prog/wcp/wcdmp/GCDS_1.php. Accessed: 2020-03-01.
- Yao, W, & Zhong, S. 2009. Nocturnal temperature inversions in a small, enclosed basin and their relationship to ambient atmospheric conditions. *Meteorology and Atmospheric Physics*, **103**, 195–210.
- Zhang, Xike, Zhang, Qiuwen, Zhang, Gui, Nie, Zhiping, Gui, Zifan, & Que, Huafei. 2018. A Novel Hybrid Data-Driven Model for Daily Land Surface Temperature Forecasting Using Long Short-Term Memory Neural Network Based on Ensemble Empirical Mode Decomposition. *International Journal of Environmental Research and Public Health*, **15**(05).
- Zheng, Xiaogu. 1995. Tropical Cyclones in the Southwest Pacific : Spatial Patterns and Relationships to Southern Oscillation and Sea Surface Temperature. *Journal of Climate*, **08**(05), 1249–1260.

APPENDIX A:
DESIGN OF AN INTERACTIVE WEB TOOL

A.1 Introduction

Interactive data visualization has become increasingly popular, primarily due to advances in software accessibility. Previously, such visualization projects often required a software and/or network engineering background. However, with the recent development of many open source software packages and libraries, the programming requirements and knowledge needed for effective interactive, web-based visualizations has been reduced making it easier for the broader scientific community to share their data and results.

This thesis was primarily funded through a grant provided by the Idaho State Department of Agriculture with a primary deliverable of an interactive, web-based tool designed for public exploration and statistical characterization of the Snake River Valley American Viticultural Area. The remainder of this appendix presents the methodology of development, current status, and future development of the tool.

A.2 Platform

R and Python are two of the most popular open source software programs used in the scientific community. Both programs offer interactive, visualization packages suitable for our purposes such as: ‘R Shiny’, ‘Bokeh’, and ‘Dash.’ We choose to use ‘R Shiny’ for three primary reasons: 1) Excellent documentation and large user base, 2) R is often considered to have a slightly smaller learning curve for newer developers, and 3) easy and well-defined web hosting options. However, one potential downside of using R is that it can be slightly slower when loading or computing on large datasets (and lacks in native parallelization compared to Python).

Although R Shiny offers its own web-hosting service, we choose to host the appli-

cation locally at Boise State University. We created our server on a virtual machine that allows for multiple R environments, thus multiple apps can be created on the same server with different web domain mappings. This can be useful for testing and development, as well as the ability to host a variety of applications without the need to set up different servers. The domain for this application is hosted at: <http://bwc.boisestate.edu/climate-viewer/>

A.3 Data

Data used within the application is largely the same as outlined in the data section of chapter two - daily aggregated data for the variables of interest from October 1987 - September 2017. One notable difference is that we do not constrain the data to within the Snake River American Viticultural Area, and include data for all of domain 01 and 02. Although the deliverable for this project is primarily concerned with the Snake River Valley AVA, future projects may be concerned with areas included in the greater Pacific Northwest. Analysis within the application has the option to subset by either the Snake River Valley AVA, domain 02, or domain 01.

A.4 Functionality

We chose to split the application into four primary categories: 1) spatial analysis, 2) temporal analysis, 3) statistical analysis, and 4) background information. The spatial explorer utilizes the 'leaflet' map package, which is an open-source javascript library for interactive maps, similar to google maps. It is mobile device friendly, zoomable, and has native layers such as land type and street names. Spatial data from various projections can be plotted as an additional layer and is highly customizable. We plot a geoJSON file of the Snake River Valley AVA and the netCDF climate data

as a separate layer using the ‘raster’ package in R.

Functionality within the spatial explorer includes the ability to plot any interval of daily data, of any metric, within a specified calendar year. Additionally, it includes the option to plot monthly or annual anomalies for any specified month and year. Anomalies are compared against the mean values for that month/year for the entire 30 year climate normal. A screenshot of the spatial explorer can be seen in figure A.1.

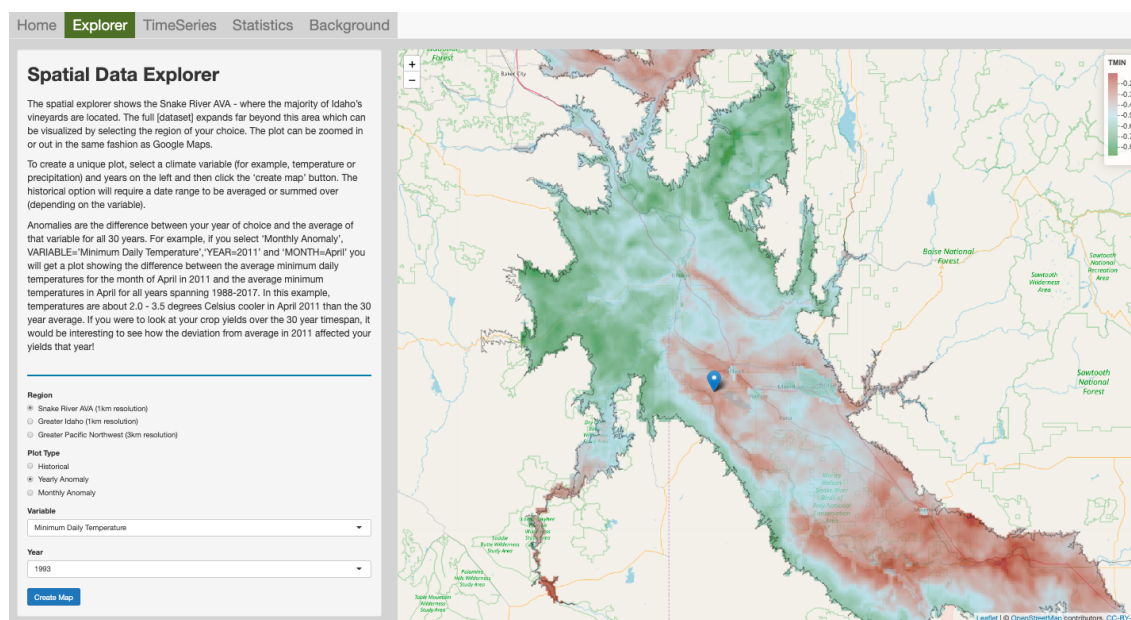


Figure A.1: Spatial Explorer UI.

The temporal component of the explorer utilizes ‘dygraphs’, an open-source interactive charting library written in javascript. It can show precise values at any point in time by mousing over the chart and can zoom to see the appropriate amount of detail. All data for use in dygraphs is spatially aggregated over any of the three previously defined regions. Any variable can be charted with up to five years charted simultaneously (figure A.2).

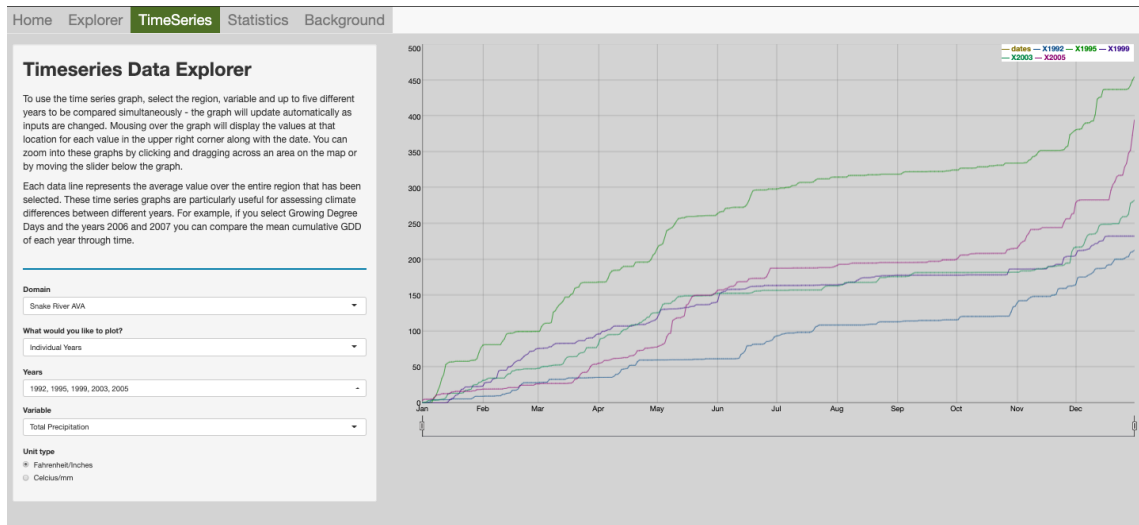


Figure A.2: Time Series UI.

The statistical analysis included two tabs, ‘Statistics’ and ‘Download Data.’ The statistics panel includes spatially aggregated statistics for the region of choice in an interactive data table. Statistics can be grouped by calendar year or by month and include means or cumulative sums for each variable. The ‘Download Data’ tab allows for more precision control of point statistics in space. It provides a leaflet map with a mouseover ability to query any point in space by latitude and longitude and bring up statistical options for that point. The two features available are an annual summary and a statistical summary. The annual summary provides means and cumulative sums for each variable for all 30 years. The statistical summary provides a basic statistical overview for each variable. It includes quantiles, minimum values, maximum values, linear trends, and the associated p-values. Each report can then be downloaded as a formatted PDF with a simple click (figure A.3).

The final ‘Background’ tab includes information on the purpose of the project, generation of the data, useful links for additional information, and personal contacts

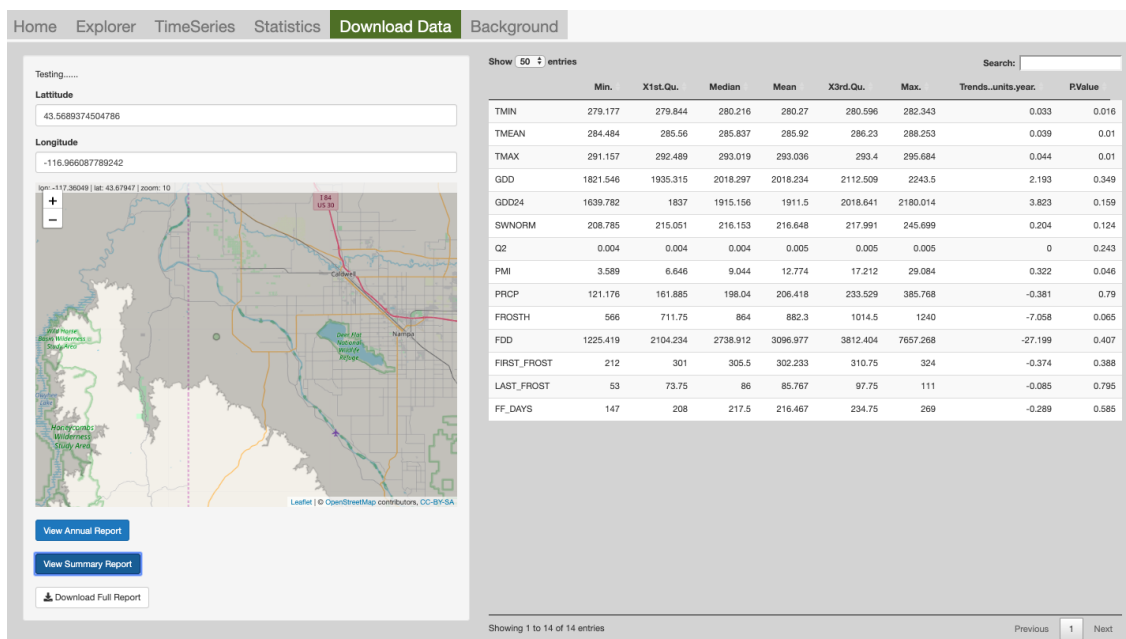


Figure A.3: Data can be downloaded directly from any location.

for the project. Additionally, it includes an animation of a significant freezing event in January of 2017 using VAPOR, a visualization tool created at the National Center for Atmospheric Research (Li *et al.*, 2019).

A.5 User Experience

As this tool is designed primarily for farmers and other interested parties that are not necessarily scientists, careful consideration for how the users will use this tool must be taken. In an effort to most closely align the functionality, ease of use, and interoperability of this tool to its primary user base, we designed a user experience (UX) survey to gain perspective from users not related to the development process.

Our survey was designed using Google Forms and included 20 questions to assess various aspects of usability. Questions were categorized into four main types: appearance and connectivity, usability, interpretability, and additional thoughts. Appearance

ance and connectivity questions were primarily designed to gather differences between different browsers and view window sizes as well as network connection errors. Usability questions were designed to assess the level of intuitiveness and amount/type of documentation needed. Interpretability questions had each participant answer a specific climate related question that required usage of the tool. Lastly, we provided open-ended questions that could be used to provide detailed feedback.

The UX survey was informally given to a variety of volunteering undergraduate and graduate students in both scientific and non-scientific disciplines. As the size of the participation was relatively small, aggregated results were used in a general qualitative way and not formally quantified. The large takeaways provided three main themes. 1) Clarity and extensibility of documentation needed improvement for both function and interpretation. Examples include the explanation of some scientific terminology such as ‘anomaly’ which may not be fully comprehended by a non-scientific population, and a wide range of answers for some simple analysis questions. 2) Many different tools (tabs - spatial explorer, time-series, etc) were used to answer the same quantitative questions, some proving more effective than others. 3) Additional bugs (some known and some new ones) and design improvements were discovered such as legend errors, inconsistent color mappings, and grammatical mistakes, and minor problems with HTML/CSS rendering.

The UX survey provided invaluable feedback for us developers to not only get a non-developer perspective, but one outside of a scientific background. Inline documentation was changed extensively as a direct result of this survey, as well as the identification and resolution of the aforementioned (and others) bugs.

A.6 Summary

The Snake River Valley Climate Viewer is an interactive web-tool designed in R-Shiny for farmers and other non-scientists to explore local historical climate data at a fine temporal and spatial resolution. It includes a multitude of visualization and analysis styles and the ability to download data summaries and statistics at precise locations that could help inform future agricultural decisions. Although it began as a viticulture centric project, other specialty crop block grants in the future could easily extend the usage of this application to include different spatial regions or other more relevant statistics and is currently underway.