A KNOWLEDGE BASE APPROACH TOO SITE AND VARIETY SELECTION IN VITICULTURE

A Dissertation

by

ELVIS ARREY TAKOW

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Robert N Coulson
Co-Chair of Committee,	Edward W Hellman
Committee Members,	Vatche Tchakerian
	Koushen Loh
	Simon Sheather
Head of Department,	Kathleen Kavanagh

May 2016

Major Subject: Ecosystem Science and Management

Copyright 2016 Elvis Takow

ABSTRACT

Regional site suitability for wine grape varieties is generally considered to be closely related to environmental conditions. However, the global spatial distribution of grape varieties is also strongly influenced by socio-economic factors. These factors have shaped and given prominence to the classic wine growing regions of the Old World. New World and other prospective wine growing areas have not benefitted from centuries of trial and error in the selection of appropriate varieties. The goal of this study was to develop a knowledge base for understanding the role of environmental factors in regional selection of wine grape varieties for optimal production. Decision support tools are developed to guide potential and existing growers in selecting appropriate grape varieties for their region. Voluminous environmental data from numerous sources and at varying spatial and temporal resolutions are incorporated in a broad scale spatial analysis of environmental conditions associated with wine grape varieties.

Many of the environmental indices that are widely used throughout the viticulture industry in evaluating regional suitability for grape varieties came into use before the advent of geographic information system (GIS) analysis and are relied upon due to historical precedence. We statistically analyzed the relationship of the most commonly used index of growing degree days (GDD) with regional price as a measure of viticultural success. We also assess the relationship between other commonly used environmental indices and price with several years of comprehensive data collected from

the grape crush districts of California. Finally, we propose a general broad scale approach to assessing the environmental similarity of renowned growing regions for selected varieties with prospective regions.

Our results suggest that systematic GIS analysis combined with continued collection of regional performance data of varieties is critical to the continued scientific advance of viticultural site selection. A clear and consistent measure of viticultural success is necessary. Indices such as GDD are useful guides in viticultural site selection, but should be used with caution. Viticultural site and variety selection should focus on the similarity of a broad selection of environmental variables in known Old World regions of success with those of prospective regions.

DEDICATION

This dissertation is dedicated to Dad, Mom, Gus, and James. We have all spent so much time apart in pursuit of our education thus finally achieving this goal means so much. Thank you for the love, values, and morals which you instilled in us about working hard and realizing ones dreams. I truly wish you were all here with me to share in this accomplishment. Dad, this is especially for you and I hope you are proud of the accomplishment. Mom, thank you as well and I hope we have made you proud. Gus, I missed watching you grow up while pursuing my education. I am sorry and hope to make up for the lost time. To my lovely wife Kathryn; you are the best thing in my life and I would not have made it this far without you. Thank you for the sacrifice you have personally made in order for me to achieve my goals. Your parents have been a wonderful inspiration throughout the process and were instrumental in representing symbols of motivation to succeed. To the Downey family, you hold a special place in my heart as I certainly would not be here today without you. Mr. Downey, I know you are watching over me; thank you for everything. Mrs. Downey, thank you so much for always being there and supporting me. You are a special Mom and we are very fortunate to have you in our lives. Jacob, you won little brother and I am proud to accept defeat to the best little brother. James, words cannot express how instrumental you have been in my life. Thanks for having Summer in our lives and my little nephew Jaylen. Thanks for all the love!

ACKNOWLEDGEMENTS

I would like to thank my committee chairs, Dr. Coulson, and Dr. Hellman. You have both been very instrumental in shaping my career moving forward and I would not have had this opportunity without you. Dr. Hellman, thank you for funding this project as long as you did and I appreciate the opportunity to work for and with you. Dr. Coulson, coming to KEL was a turning point in my life and I am truly grateful for the friendship and guidance. I know you will be working for many more years to come and hope I can be a part of it. To my committee members, Dr. Loh, Dr. Tchakerian, and Dr. Sheather, thank you for their guidance and support throughout the course of this research. Dr. Maria Tchakerian, thank you so much for everything. You helped start me down this path early on as I struggled with learning how to manage and use GIS software. You have been a great adviser and friend.

Thanks also go to all my friends and colleagues and the Ecosystem Science & Management department faculty and staff for making my time at Texas A&M University a great experience. I want to specifically thank everyone at the Knowledge Engineering Laboratory. Many of you have gone on to other endeavors but I am truly grateful for the contributions you have made in helping me succeed. Thank you to Vickey Chen, Chris Garza, Luis Calixto, James Tracy, Reza Teshnizi and so many others who have helped me along the way. I also want to extend my gratitude to Dr. Andrew Birt. Andrew has been the most influential person throughout this learning experience as a friend and mentor. Thank you for the countless hours of help with code, analysis, and simply just

being a friend. Finally, thanks to my Mom and Dad for their encouragement and to my wife Kathryn for her patience and love.

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vii

LIST OF FIGURES x

LIST OF TABLES xv

CHAPTER I INTRODUCTION 1

Preface 1
Background to Wine Grapes and the Wine Industry 4
Life Cycle of Wine Grapes 7
Grape Growing Regions of the World 9
Old World versus New World Viticulture 11
Current models for understanding wine grape environmental conditions 14

CHAPTER II METHODOLOGY 17

Introduction 17
Methodology 18
Data Acquisition and Collection 20
Data Integration and Aggregation 24
Data Modeling and Analysis 26
Data Interpretation 27
Results and Discussion 27
Conclusions 32
The KELSoil Web service: An application of the Soil Survey Geographic Database 34
Data Acquisition and Collection 35
Data Integration and Aggregation 35
Data Modeling and Analysis 39

CHAPTER III A CONCEPTUAL APPROACH TO BUILDING MODELS FOR SITE AND VARIETY SELECTION IN VITICULTURE 45

Introduction 45
Methodology 47
A Scientific Approach to Site and Variety Selection 47
Results and Discussion 53
Conclusions 54

CHAPTER IV THE UTILITY OF GROWING DEGREE DAYS AS AN INDEX FOR VITICULTURE 56

Introduction 56
The Importance of Bioclimatic Indices like Growing Degree Days 58
Materials and Methods 59
Estimating GDD at different temporal resolutions 61
Assessing Inter annual or year to year variation 62
Spatially interpolated versus Station data 63
Assessing the Spatial Differences in GDD 64
Assessing the relationship between GDD and temperature 65
Results 65
Temporal Resolution 65
Inter annual Variation 66
Interpolated versus Station data 68
Spatial Differences in GDD 70
GDD versus GSAT 71
Discussion 71
Conclusions 88

CHAPTER V WINE GRAPE SUITABILITY: A CASE STUDY OF THE CALIFORNIA CRUSH DISTRICTS 89

Introduction 89
California Wine Industry 90
California Crush Districts 91
Data and Methods 93
Development of Crush District Data (Dependent Data) 93
Development of Environmental Data (Independent Data) 98
Environmental Factors 100
Regression Analysis 102
Results and Discussion 106
Conclusion 130

CHAPTER VI INTEGRATION OF CLIMATIC AND EDAPHIC CHARACTERISTICS IN VITICULTURE SITE SELECTION: ENVIRONMENTAL LANDSCAPE VITICULTURE INFORMATION SYSTEM (ELVIS) 133

Introduction 133
Methods 135
System architecture 135
Environmental Database 136
Data Modeling 137
Discussion 138
Conclusion 146

CHAPTER VII CONCLUSION 149

REFERENCES 154

APPENDIX A 158

LIST OF FIGURES

Figure 1. The wine grape quality and site selection paradigm illustrates the interaction of site, variety, and quality in order to achieve appropriate site selection..........3

Figure 2. An illustration of global wine production for the top wine producing nations of the world..........4

Figure 3. A description of the annual life cycle of the grape vine showing stages of the vine through the growing season..........9

Figure 4. A depiction of the global distribution of the wine producing regions of the world.Source:www.thirtyfifty.co.uk/spotlight-climate-change.asp..........11

Figure 5. A generalization of the global distribution of Old World versus New World regions as applied to viticulture..........13

Figure 6. An overview of the data analysis process required for modern day environmental modelling..........19

Figure 7. A summary of the environmental data and sources..........22

Figure 8. A summary of the environmental data, variables and associated relevant spatial attributes of the data..........23

Figure 9. A description of the data conversion from machine readable formats to human readable interpretations..........24

Figure 10. An overview of the data processing of environmental data..........25

Figure 11. An illustration of the proprietary functions designed to query raw environmental data..........26

Figure 12. Results of the data access times and required storage space for various data sources utilized in the research..........28

Figure 13. An illustration of the file database of Daymet climate data illustrating the vast size with 45,029 files of data..........29

Figure 14. An illustration of the global distribution of weather stations representing a time frame of 1960-2012..........30

Figure 15. A depiction of the data organization and structure of global weather stations..........31

Figure 16. An illustration of the original SSURGO database showing data tables and the complex relationships 36

Figure 17. An illustration of the extracted tables and relationships in the simplified database showing the four tables 38

Figure 18. Illustrations of various methods for defining an area of interest based on single location, rectangle, and polygon 40

Figure 19. An illustration of results of soil data for a single location organized as a texture pyramid and soil profile....41

Figure 20. The results of an area of interest chosen for analysis is represented by multiple MUKEY values 42

Figure 21. Soil texture pyramid displaying additional soil data layered by soil horizon.43

Figure 22. Illustrating the significance of determining measures of suitability and environmental conditions....49

Figure 23. An illustration of the suitability plotted against environmental conditions for variety A....50

Figure 24. An illustration of the environmental conditions at a particular location are simplified into an index representative of the conditions at a particular location 51

Figure 25. Highlighting the relationship between GDD and suitability assessed at 2 different locations 52

Figure 26. A conceptual design of the process of involved in the development of a model for variety selection 53

Figure 27. A visual comparison of interpolated data and station data GDD for entire US 64

Figure 28. An illustration of the percentage difference in GDD estimates compared at different resolutions 65

Figure 29. A depiction of the average annual year to year variation in GDD across the US 66

Figure 30. A depiction of the average percentage change in annual year to year variation of GDD across the US 67

Figure 31. Results of the inter annual variation in estimates of GDD for Napa and Sonoma 67

Figure 32A scatter plot of the general relationship between interpolated GDD and station GDD 68

Figure 33. A scatter plot of the general relationship between interpolated GDD and specific station GDD 69

Figure 34. A scatter plot of the relationship between modeled GDD estimates and interpolated GDD 70

Figure 35. A graphical illustration of variation in daily and monthly accumulations of GDD over time 73

Figure 36. A graphical illustration of the variation in hourly and monthly accumulations of GDD over time 74

Figure 37. A graphical illustration of variation in daily and monthly accumulations of GDD over time 75

Figure 38. A depiction of the original Winkler map of California alongside our updated version based upon average GDD estimates for 1991-2012 77

Figure 39. California Grape Variety Suitability Based on Winkler Zone Classification for the 17 crush districts 79

Figure 40. A comparison of the relationship between inter- annual variation in GDD and changes in elevation 84

Figure 41. No predictable relationship between GDD estimates and distance of weather station to interpolated location 85

Figure 42. The relationship between GDD and solar radiation shows no direction correlation 86

Figure 43. Results indicating the lack of correlation between GDD and solar radiation across the US 87

Figure 44. The California Grape Crush Districts as defined by Clare Berryhill Grape Crush Report Act of 1976 91

Figure 45. An example of raw tonnage data of white wine grapes for the year 2012 by district and variety 96

Figure 46. Restructured generic data table for variable of interest by year, district and variety 97

Figure 47. An illustration of the methodology for data acquisition, collection, integration and aggregation of California crush data 97

Figure 48. Distribution grid of Daymet data covering the full extent of the US at spatial resolution of 2x2 decimal degrees with illustration of the California crush districts and the Texas wine growing regions....................98

Figure 49. Daymet data resampled from 2x2 degree grids to 0.125x0.125 degree grids for analysis....................99

Figure 50. An overview of the data collection and analysis of the California crush data....................100

Figure 51. A illustration of the original Winkler Regions as defined by Amerine and Winkler (1944)....................107

Figure 52. A depiction of the crush districts using Winkler classification based on Average GDD for the period of 1991-2012....................108

Figure 53. Regression plot for first level of analysis to predict Price from GDD..........109

Figure 54. A correlation matrix of coefficients illustrating the relationships of each of the environmental variables....................119

Figure 55. A depiction of grapevine climate based on average growing season temperature (source: Jones et al, 2006) 121

Figure 56. Parameter estimates and summary of model fit at second level of regression analysis....................122

Figure 57. The results of parameter estimates and a summary of model fit at the third level of regression analysis for Pinot Noir district 14 126

Figure 58. A summary of the results of the third level of analysis for all statistically significant parameters by variety and district....................127

Figure 59. An overview of the web-based system architecture of the ELVIS decision support system structured around an environmental database 135

Figure 60. An illustration of how ELVIS functions through the utilization of data for geo-spatial analysis of defined areas of interest 138

Figure 61. A depiction of the process of defining an area of interest for exploration using the rectangle tool in the ELVIS system 140

Figure 62. An illustration of the environmental summary of the area of interest characterize by factors specific to viticulture 141

Figure 63. An illustration of the outlined areas from interest for simultaneous comparison using the compare option of the ELVIS system 142

Figure 64. Graphical comparison of average annual relative humidity and annual precipitation in Burgundy and Willamette illustrating the ability to compare multiple variables instantly 143

Figure 65. A graphical comparison of average annual cumulative GDD in Burgundy and Willamette valley 144

Figure 66. An illustration of a spider chart of Burgundy and Willamette valley illustrating simultaneous comparison of multiple variables only possible through the analysis phase 144

Figure 67. A visual display of the ELVIS system functionality and the various levels of user interaction illustrating increasing levels of interpretation of data 145

LIST OF TABLES

Table 1. A comparison of GSAT to GDD using the different scenarios where GSAT is identical 71

Table 2. Winkler scale for California by Amerine & Winkler based on heat summation (GDD) 78

Table 3. GDD and GSAT by region developed for California (Amerine and Winkler 1944, Halliday 1993) 87

Table 4. Comprehensive list of environmental variables used for regression analysis modeling of viticultural suitability 103

Table 5. A description of the regression analysis modeling approach for GDD....105

Table 6. A description of the regression analysis modeling approach for environmental variables 106

Table 7. The results of the first level of a regression analysis to predict Price from GDD, while controlling for Year, District, and Varieties....109

Table 8. The results of the second level of a regression analysis to predict Price from GDD, while controlling for Year and District 110

Table 9. The results of the third level of a regression analysis to predict Price from GDD while controlling for Year 111

Table 10. A description of the variation in GDD from year to year within a given District 116

Table 11. Parameter estimates for environmental variables at first level of analysis 118

Table 12. Summary of model fit for all varieties at the first level of analysis 120

CHAPTER I

INTRODUCTION

Preface

This dissertation is divided into seven chapters. The first chapter introduces the goal and objectives of the dissertation. We also provide background information which pertains to the issue of site and variety selection in viticulture with the goal of producing quality wine grapes. We provide a brief overview of the wine grape industry along with the life cycle of the grape vine and how it relates to the problem. We also briefly discuss the concepts of Old and New World regions, Terroir and the distribution grape growing regions of the world. In the second chapter we outline an approach to the data collection process along with the analysis techniques employed to prepare data for environmental modeling. This chapter discusses the need for a scientific approach to processing voluminous amounts of data. In the third chapter we outline a conceptual but scientific approach to building models for site and variety selection. We emphasize the identification of a consistent measure of viticultural success (dependent variable). The fourth chapter examines the use of the GDD concept to describe varietal suitability in the context of viticulture. We evaluate the potential limitations in the concept of an estimate of GDD for a particular location. In the fifth chapter we undertake a case study of California by modeling the relationship between environmental conditions and a measure of viticultural suitability. We specifically assess the relationship between GDD and price as a measure of viticultural success. The sixth chapter demonstrates and describes the development of an internet based, scientifically objective tool to facilitate vineyard site assessment and grape variety selection. The core of this system is a spatially explicit environmental database relevant to wine grape production including climate, soil and topography data. The seventh chapter provides an overall summary of the conclusions. We propose an approach to assessing similarity between regions. In other words how similar is one location to another for wine growing given a set of environmental conditions. We present this approach as future work in the development of site and variety selection for viticulture.

Viticulture is perhaps the most geographically expressive of all agricultural industries (de Blij, 1983). Moreover the most important factor for producing quality wine is growing high-quality wine grapes. As such site and variety selection is the single most important decision any prospective grower will make (Gladstones, 1992). The goal of this research is to understand the role of environmental factors that drive wine grape production. Using a data driven approach and the acquired knowledge base, prospective and current wine grape growers can make scientifically objective (more informed) decisions about the selection of appropriate grape varieties for a location or appropriate locations for a specific variety. The wine grape quality and site selection paradigm focuses on the interaction between site, variety, and quality which leads to selection of the most suitable site for viticultural success. Figure 1 illustrates this interaction.

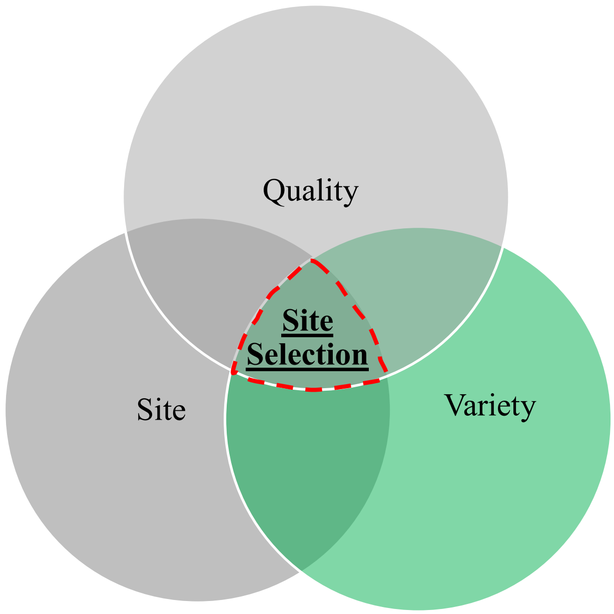
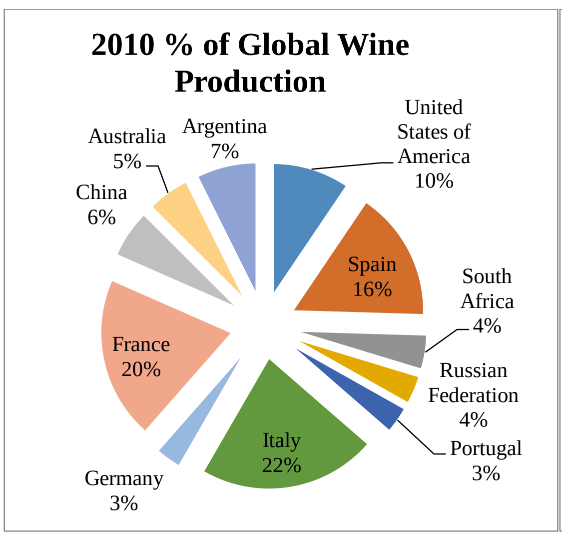


Figure 1. The wine grape quality and site selection paradigm illustrates the interaction of site, variety, and quality in order to achieve appropriate site selection

This research focuses on the development of a knowledge base that relates environmental conditions to the culture of successful grapes for wine making. At a most basic level, this will involve a review of select world-wide locations that currently and have historically supported the production of successful wine grape varieties. Presumably due to a long history of success these locations exhibit the environmental conditions necessary for successful wine grape production. This information was applied to the development of a knowledge base (models) that can be used (extrapolated) to identify suitable grape growing sites. More specifically, growers can make informed decisions about either (1) the selection of potential sites most likely to support grapes of a given variety or (2) the selection of the varieties most suitable for a particular plot of land.

Background to Wine Grapes and the Wine Industry

The global wine market is currently a billion dollar industry and has been in existence since the early 1800s. Today the top 5 wine producing nations of the world are responsible for over 50% of the world's wine production, representing the sensitive nature of land for wine grapes (OIV, 2013). Figure 2 illustrates the most recent statistics on global wine production from the International Organization of Vine and Wine.



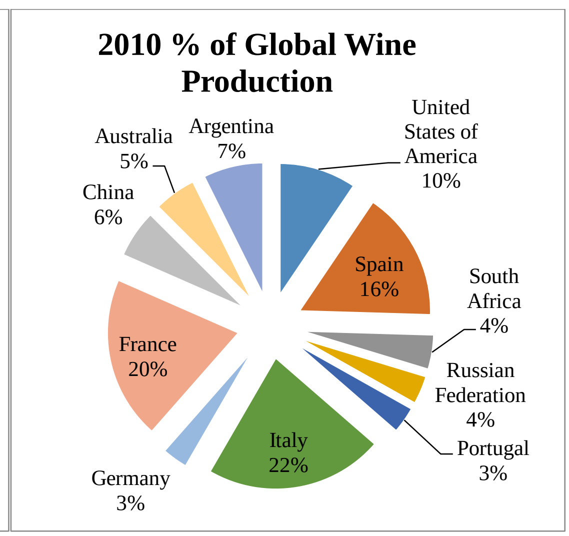


Figure 2. An illustration of global wine production for the top wine producing nations of the world

Wine is a unique commodity whose production predates recorded history (Mullins et al., 1992). Generally made from one or more varieties of the European species *Vitis vinifera,* wine grapes include well-known international varieties such as Pinot noir, Chardonnay, Cabernet Sauvignon, Sauvignon Blanc, Riesling, Semillon, Syrah/Shiraz, and Merlot. Throughout history, the best wines were reserved for the elite of society hence the image of wine as a sign of status persists even today (Mullins et al., 1992). Wine is still an integral part of the culture in many countries. When a single variety is used as the predominant wine grape (usually as defined by appellation law as minimums of 75% to 85%), the result is a "varietal" as opposed to a "blend". Wine is also made from other species of grape or from hybrids, resulting from the genetic crossing of two species. *Vitis labrusca* (the Concord grape), *Vitis aestivalis*, *Vitis ruprestris*, *Vitis rotundifolia* and *Vitis riparia* are all native North American grapes.

Grapes are divided into four broad categories: European or Vinifera (*Vitis vinifera*), French hybrids (*Vitis vinifera* crossed with *Vitis rupestris* or *Vitis Lincecumii*), American, and Muscadine (*Vitis rotundifolia*) grapes. European grapes require warm weather and a long growing season to properly mature their fruit. Characterized by tolerance to heat, drought, sandy soils, and soils with a high pH, they are also highly susceptible to winter injury, insects, and diseases (Winkler, 1962). Of particular importance is their high susceptibility to Pierce's disease. The disease is endemic in northern California, being vectored by the glassy-winged sharpshooter which is a large leafhopper insect from the family Cicadellidae.

French hybrids originated in France and are grown to a large extent in certain parts of Europe due to their resistance to fungal diseases (Winkler, 1962). American grapes however are cold hardy and resistant to a large number of diseases and insect pests. The vines are fairly vigorous, highly productive and generally mature their fruit early (Winkler, 1962). The majority of the world's vineyards though are planted with European *Vitis vinifera* vines that have been grafted onto the North American species' rootstock. Grafting became a common practice due to the resistance of North American species to phylloxera, a root louse that eventually kills the vine.

Wine is a multidimensional geographic agricultural commodity whose origin and means of dissemination is unclear. Economic and political influences as well as local cultures have historically influenced the geography of wine. Identification with specific geographic and regional environments as well as scientific study of environmental factors in viticulture literature has led to the notion of wine growing regions. In France this regional identity has been reinforced by the designation of wine growing areas known as *appellation d'origine controlee* (Johnson and Robinson, 2013). The goal of AOC system was to maintain the standards and quality for wines while retaining its regional and cultural identity (Barham, 2003). Grape production limits, pH restrictions, and percentages of specific grape varieties in wines have contributed towards establishing the identities and traditions of these regions (Bureau of Alcohol, Tobacco and Firearms, 2001, Jackson, 2008).

Life Cycle of Wine Grapes

The annual life cycle of the grape vine is the process that takes place annually in the vine yard. This process begins with bud break in the spring and culminates with leaf fall followed by dormancy in the fall. Each step of this process plays a vital role in the development of grapes suitable for wine making. The effects of environmental conditions along with vine disease are constantly monitored by viticulturist as these may either impede or facilitate the vines progress from bud break, flowering, fruit set, veraison, harvesting, leaf fall and dormancy. Variation in the amount of time spent at each stage is dependent on the climates (site) and the grape variety which together determine the resultant quality of the grapes.

The life cycle begins in the spring with bud break. This period is around March in the Northern hemisphere or September in the Southern hemisphere. Environmentally this stage is triggered by average daily temperatures surpassing 10°C (50°F) (Winkler 1962; Moncur et al. 1989). Subsequently buds on the vine start to swell and eventually shoots begin to grow. This is followed by the process of flowering which begins with small flower clusters appearing on the tips of the young shoots around May in the Northern hemisphere and November in the South. Fruit set proceeds flowering almost immediately, when the fertilized flower begins to develop a seed. In the Northern Hemisphere, this normally takes place in May and in the Southern Hemisphere in November. After fruit set, the grape berries are green and hard to the touch when they enter the stage of veraison which is beginning of the ripening process. In the Northern

Hemisphere this will be around the end of July and into August and between the end of January into February for the Southern Hemisphere. Finally based on the subjective determination of ripeness, the grapes are removed from the vine. This event is known as harvest and in the Northern Hemisphere this is generally between September and October while in the Southern Hemisphere it is generally between February and April. Figure 3 illustrates the annual life cycle of the grape vine with distinct stages critical to the quality of the grape.

It is commonly recognized that the start of bud break in deciduous fruit crops is determined by cessation of winter dormancy (Hauagge and Cummins, 1991). This stage is followed by a phase of bud growth which is also related to mean daily temperature (Lombard and Richardson, 1979). For many years agricultural scientists have sought causal links between variations in environmental conditions and the development rates of crops. This contributed to the idea that temperatures accumulated above a specific threshold or base temperature and subject to an upper limit provide a good indication of the heat requirements. Specific development stages as a result of heat requirements are central to expressive geography of wine. The life cycle of the grape vine requires that specific environmental conditions be met in order to develop quality fruit hence the very specific global range of viticulture distribution. According to Hellman (2003), the timing and duration of developmental events are subject to variations due to the grape variety, local climate, and seasonal weather. This underscores the uniqueness of wine grapes and the importance of the wine grape quality and site selection paradigm described in figure 1. The sequence of events however remains constant

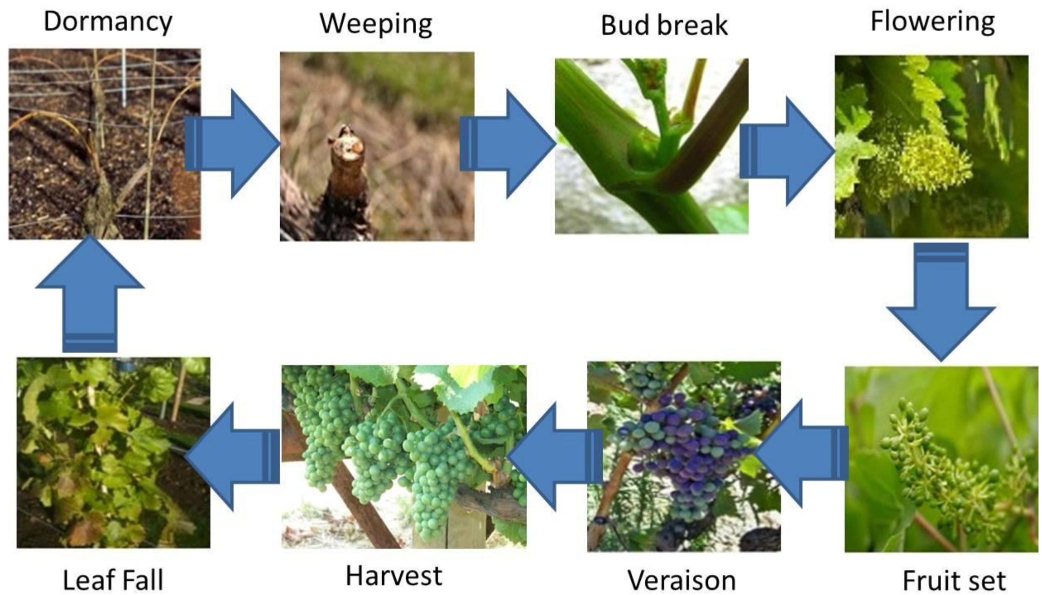


Figure 3. A description of the annual life cycle of the grape vine showing stages of the vine through the growing season

Grape Growing Regions of the World

Globally the main areas of viticulture are situated between latitudes 30°N and 50°N and between 30°S and 50°S. Presumably environmental factors have a strong influence on the spatial distribution of wine regions (Dry and Smart, 1988; Gladstones, 1992; Wolf, 1997). The delineation and consequent distribution of regions based environmental conditions and wine quality is not a new concept; it was first practiced during the Roman Empire (Bohmrich, 1996). During the early 19th century vineyards and wineries of Bordeaux were categorized on market value, and classification was generally skewed toward famous estates (Bohmrich 1996). With the creation of the now famous *Appellation d'origine controlée* (AOC) system in the 1930s, varieties and the controversial concept of 'Terroir' became globally known. Hundreds of years of experience gained by trial and error in identifying and isolating terroirs became the law in French viticulture. As such the AOC system was established as a result of the destruction of the French wine grape industry by phylloxera. This system ultimately delineated wine regions, determined the varieties grown, the amount of harvest, yield, and the alcoholic content of grapes in a region. The AOC system provided the French wine industry with a powerful marketing tool for the sale of their wines (Celine, 1998). This terroir driven system has consequently guided the current distribution of regions by serving as a guideline for production. Figure 4 provides a general overview of the current distribution of global wine regions. The general global distribution of the world's wine regions is shaped by an average temperature band of 10°-20°C. This broadly covers the overall distribution of most of the world's wine growing regions. With the advent of global climate change and technology, the world's wine map no longer consist of two neatly banded zones as regions have thrived outside of traditional growing areas.

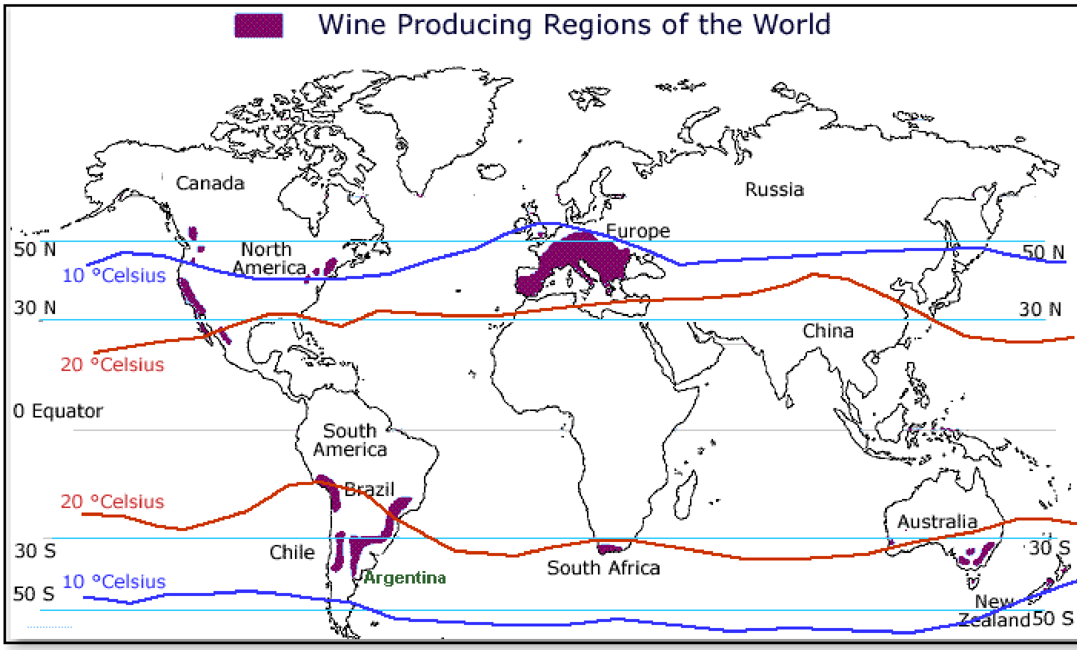


Figure 4. A depiction of the global distribution of the wine producing regions of the world.Source:www.thirtyfifty.co.uk/spotlight-climate-change.asp

Old World versus New World Viticulture

It is important to understand the fundamental differences between the terms Old and New World as used in the context of viticulture. These terms have often been used to describe differences in viticulture and winemaking philosophies. We provide a brief over view of some of the key distinctions between Old and New World viticulture.

The term Old World as applied to viticulture refers to European countries and regions like France, Spain, Italy, Germany, Portugal, Greece, North Africa and the Middle East. The Old World also includes regions around the Mediterranean Sea that exhibit a

Mediterranean climate. Wine making began in the Old World where these regions have had a long history of viticulture producing wine for thousands of years. The Old World regions are responsible for wine quality laws and were first created and implemented in France.

The New World refers to viticulture from regions such as the United States, South America, Australia, New Zealand and South Africa. New World regions are generally wine producing areas outside of the so- called traditional wine-growing regions of Europe. Characterize by science, innovation, and technology, these regions have rapidly led to the increase globalization of the wine industry. Established viticultural areas ("Old World") have benefitted from centuries of trial and error in the selection of appropriate varieties. New World regions have embraced experimentation and technology. Figure 5 illustrates a generalization of the current distribution of Old World and New World viticultural regions of the world.

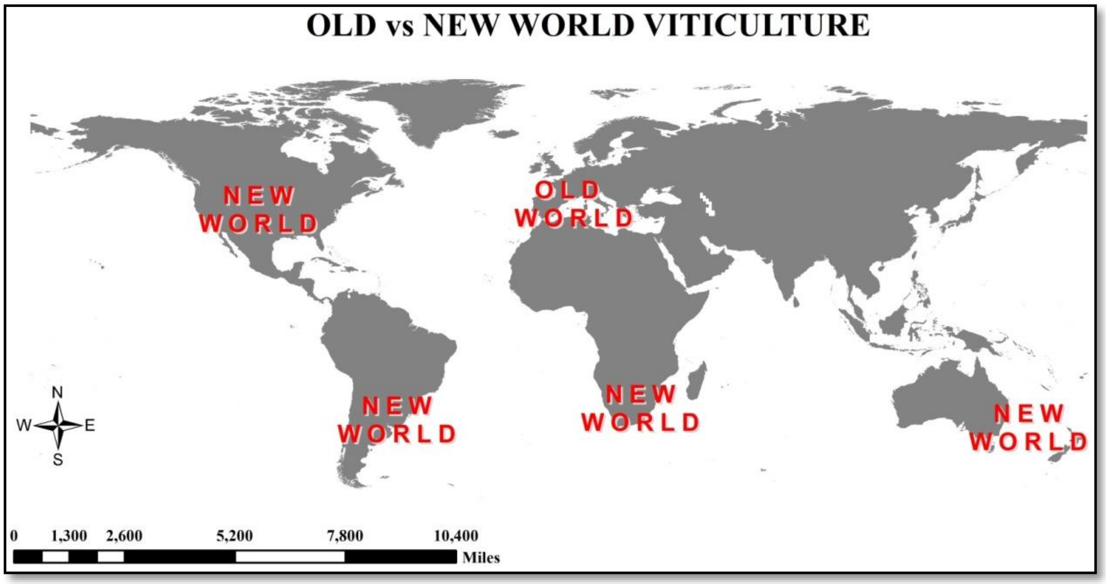


Figure 5. A generalization of the global distribution of Old World versus New World regions as applied to viticulture

In contrast to the Old World, there are fewer restrictions in New World regions. These regions don't have regulations to the same extreme as Old World regions. New World regulations provide greater freedoms for experimentation with how wines can be made as opposed to what a law says should be made. New World viticulture generally places less emphasis on historic practices or centuries of experiential knowledge, and more emphasis on viticulture practices that take advantage of modern advances in science and technology. However given the risks and timelines involved in planting vines and producing high quality grapes, the models and knowledge base we develop will enable viticulturists to use a scientific and objective approach to site selection that should benefit individual growers (either existing or potential) and thus the wine industry as a whole. This research will help in guiding the selection of appropriate locations or appropriate varieties for specific locations

Current models for understanding wine grape environmental conditions

The physical factors that influence suitability in viticulture include matching a given grape variety to its ideal climate along with optimum site characteristics of soil, elevation and slope. In order to analyze these factors a number of environmental models have been put forth. Overall, climate exerts the greatest influence on the ability of a region to produce quality grapes (van Leeuwen et al. 2004). The most frequently used models include simple to complex indices of temperature. As such our focus in this research is on the climatic factors that potentially influence suitability. Although most grapevines can be grown across a wide variety of soil types, the most important edaphic characteristics for optimum growth are good internal drainage, adequate depth, sufficient water holding capacity during dry periods, and a soil pH that is slightly less than neutral (Jones and Hellman, 2003). Our study will be limited to the aforementioned soil factors along with climatic variables that have historically been examined in previous viticulture studies. We also assume that historic Old world wine regions known for growing premium wine grapes exhibit conditions that determine which grape varieties can be grown in these regions. Furthermore, the identification of new locations which exhibit similar conditions has the potential to provide guidance in the choice of wine grape varieties for prospective growers in new world regions.

Several studies have been conducted over the years which place an emphasis on "premium" wine regions. These regions include old world locations of Europe and new world regions of Australia and western United States. The studies attempt to assess the suitability of establishing and sustaining vineyards by analyzing general climate aspects (Dry and Smart 1988; Gladstones 1992; Winkler et al 1974), and by investigating the overall terroir elements, such as climate, soils and other viticultural practices (Jones 2006; Van Leeuwen et. al, 2004; White 2003). Many other studies have focused on the regional aspects of suitability by narrowing the geographic area of study. In a 2001 study, Jackson and Schuster examined the influence of climate in order to understand how to adjust vineyard management techniques to maximize productivity (Jackson 2001; Jackson and Schuster, 2001). Jones and Davis (2000) further examined the influence of climate in Bordeaux, France on grapevine phenology, composition and wine production. Furthermore, Jackson and Cherry (1988) explored seventy-eight locations throughout Europe, Australia, New Zealand and North America for suitability based on temperature and latitude. Tonietto and Carbonneau (2004) provide one of the more comprehensive studies involved with quantifying the global wine grape-growing regions and creating a multi-criteria climatic classification (MCC) system, based on the Heliothermal Index (HI), the Cool Night Index (CI) and the Dryness Index (DI). The MCC system explores ninety-seven of the established premium wine grape growing regions and classified them according to the aforementioned indices. Several smaller scale studies have investigated site suitability throughout Oregon (Jones and Hellman, 2002; Jones et. al 2004). A similar study involved the region of the Okanagan and Similkameen Valleys in British

Columbia (Bowen et al 2005). Common to most viticulture modeling studies is the use of growing degree days (GDD) as an index of viticulture suitability.

Our approach has been to focus on a data driven model that researchers and growers can use to study climatic and soil relationships that determine grape and wine quality. Hence we assess the utility of the most commonly used index in viticulture. While other studies assume a relationship between environmental factors and wine grape success, we sought out to examine if a significant relationship indeed exists based on actual wine production data. Conducting this type of suitability characterization requires a broad scale comprehensive data base of environmental variables. So long as environmental data is available, along with data representing viticultural success, this approach can be implemented for any prospective location. It is crucial to identify whether there exist a relationship between the success of viticulture at a particular location and the environmental conditions which characterize the location for viticulture. By so doing we can then proceed to compare how similar a new location is to a known region of viticultural success based on a set of environmental conditions.

CHAPTER II

METHODOLOGY

In this chapter we shall present an overview of the data collection process along with analysis techniques employed to prepare data for environmental modeling. We begin with a brief discussion on the need for a scientific approach to processing voluminous amounts of data. The modern day challenges of acquiring data for analysis and research are not limited to identifying and collecting data from multiple sources. Modern scientific research is characterized by 'Big Data' which describes the exponential growth and availability of data for research. We describe a data driven approach to environmental modelling that initially involves mining, collection, analysis, and interpretation of Big Data.

Introduction

Environmental modeling is important to the Texas wine industry. A growing wine industry necessitates the need to match varieties to appropriate environmental conditions. Whereas established viticultural areas ("Old World") have benefitted from many years of trial and error in the selection of appropriate varieties, Texas has not. For individual viticulturists, there are significant temporal and financial risks associated with planting grape vines. In order to minimize the inherent risk associated with viticulture, a scientifically objective approach to matching varieties to suitable locations is necessary. Environmental modelling therefore provides an efficient means of understanding the functional relationships between variety suitability and environmental conditions. The need for such research facilitated by technology is essential to addressing the issue of matching grape varieties to environmental conditions. Modern day scientific research has evolved due to innovations in technology that drive the methods by which research is now being carried out. Historically, scientific research was driven by hypotheses followed by the collection of data in an effort to design scientific experiments which address a problem. Nowadays, big data drives almost every aspect of environmental research. Data driven models have become an integral part of scientific research. As such we must develop scientifically efficient methods for analyzing voluminous sources of heterogeneous data.

The goal of this research is to describe a process for organizing large, complex, and heterogeneous data sources to enable environmental modelling given the current technological advances. Our approach towards efficient analysis of data rest on the following premise; data is a valuable currency for research hence we must derive ways to develop value from data that is incompletely or imperfectly captured for the research at hand. While most research only focuses on the analysis and modeling of data, we have taken an approach that involves maximizing the value of data by adopting a multi-step process for managing the data. Our methods described in the following section further elaborate on this approach.

Methodology

Environmental modelling is not limited by computer technology but by the availability of data and human understanding encapsulated in data driven models

(Goodchild et al. 1996). In order to effectively carry out environmental modelling for scientific research, we propose an iterative process with 4 distinct phases. Each phase of this process has specific challenges associated with preparing data for the subsequent stage of analysis. We proceed by describing each phase of the analysis process as illustrated in figure 6. Each phase of the process is stand-alone, yielding results which can be used for the subsequent stage. The end result of each phase of this process is data that is more organized and more complete than the previous stage of the process. The choice to carry on through the entire process is dependent upon the desired goals of the research and modelling.

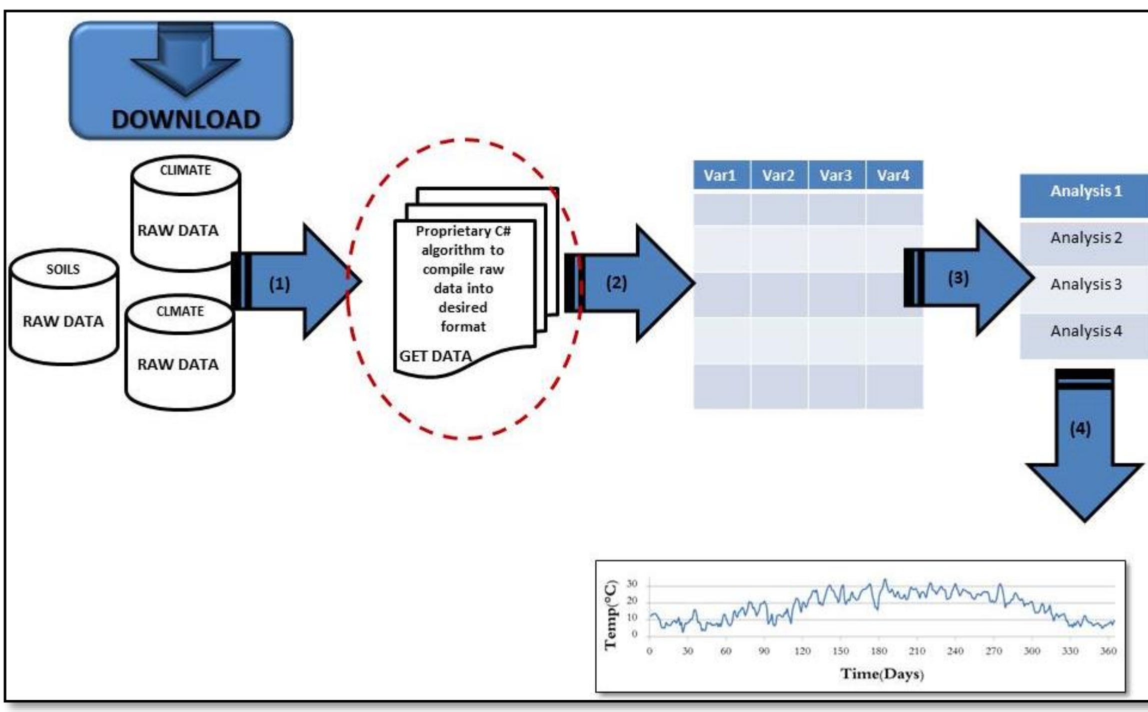


Figure 6. An overview of the data analysis process required for modern day environmental modelling

Data Acquisition and Collection

Data acquisition and collection in the context of this research involved the process of identifying and collecting raw environmental data for the purpose of environmental modeling for viticulture. This process has evolved over time from identifying a scientific problem and subsequently collecting data in order to address the problem, to mining data that has already been collected for scientific analysis. Modern technology now facilitates data collection through the existence of government research agencies like National Aeronautics and Space Administration (NASA), Environmental Protection Agency (EPA), Natural Resources Conservation Service (NRCS), United States Department of Agriculture (USDA) and United States Geological Survey (USGS). These agencies collect environmental data on a broad scale. Data acquisition no longer only involves systematically gathering and measuring information on variables of interest to address stated research questions, test hypotheses, or evaluate outcomes. It also includes systematically sorting through massive databases of information collected on a broad scale for a general geographic region.

This research used environmental data such as climate and soil from a number of different sources at varying scales. Our data sources included climate data from the Oakridge National Laboratory (ORNL), National Climatic Data Center (NCDC), and the European Climate Assessment and Dataset (ECA&D). Sources of soil data included the Harmonized World Soil Database (HWSD) and the Soil Survey Geographic Database (SSURGO). The goal of this stage of the analysis was to put together a comprehensive collection of raw environmental data in its native format. This involved identifying and downloading relevant data from the aforementioned sources. Figure 7 provides a summary of the environmental data used in this research as well as the data sources. More specific details about each data source including variables, scale, resolution and other spatial attributes of each data source have been outlined in figure 8. We identified and collected raw data from sources deemed to provide the most comprehensive coverage of the study areas. Since raw data is often voluminous, it was necessary to design algorithms which did not discard useful information as the data was collected or downloaded. The data was carefully uploaded to a structured query language (SQL) database using Microsoft SQL Server Management Studio. We also uploaded data to a simple file database structure based on a nomenclature format which included latitude and longitude. Every climate data file from ORNL was stored using a format similar to the following: "lat_25125lon_-81000.csv". For example, if a user identified a location as their area of interest and wanted daily gridded data for that location, the appropriate query would be run based on a latitude and longitude. This required searching a file database using an algorithm which takes latitude and longitude as inputs and returns daily weather variables for the location of interest. Similarly, the same procedure is used to query every data source in our database environment. Our raw data was categorized as Climate, Soils, or Topography, regardless of the data format. Figure 7 illustrates how data was categorized.

Source	Url	Category	Format	Type	Data Extent		Data Resolution	
					Spatial	Temporal	Spatial	Temporal
DAYMET	http://daymet.ornl.gov	Climate	Text file	Interpolated	US	1980-2012	1km	Daily
ECA&D	http://eca.knmi.nl	Climate	Gridded	Interpolated	Europe	1980-2013	27km	Daily
NCDC	http://www.ncdc.noaa.gov	Climate	Text file	Station	Global	1960-2012	Variable	Daily
SSURGO	http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=nrcs142p2_053550	Soils	Polygon	Vector	US	NA	na	NA
HWSD	http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/en	Soils	Gridded	Raster	Global	NA	1km	NA

Figure 7. A summary of the environmental data and sources

Data Type	Source	Variable	Description	Resolution		Extent		Format	Units	Processed	Acronymn	Size
				Spatial	Temporal	Spatial	Temporal					
Weather	National Climatic Data Center	MIN	Minimum Temp	Variable	Daily	Global	1960-2012	csv	°F	yes	NCDC	15.1GB
	National Climatic Data Center	MAX	Maximum Temp	Variable	Daily	Global	1960-2012	csv	°F		NCDC	
	National Climatic Data Center	TEMP	Average Temp	Variable	Daily	Global	1960-2012	csv	°F		NCDC	
	National Climatic Data Center	PRCP	Total Precip	Variable	Daily	Global	1960-2012	csv	in		NCDC	
	National Climatic Data Center	DEWP	Average dew point	Variable	Daily	Global	1960-2012	csv	°F		NCDC	
Weather	European Climate Assessment	TN	Minimum Temp	0.25 DD°	Daily	Europe	1980-2013	netCDF Grid	°C	no	ECA&D	
	European Climate Assessment	TX	Maximum Temp	0.25 DD°	Daily	Europe	1980-2013	netCDF Grid	°C		ECA&D	
	European Climate Assessment	TG	Average Temp	0.25 DD°	Daily	Europe	1980-2013	netCDF Grid	°C		ECA&D	
	European Climate Assessment	RR	Total Precip	0.25 DD°	Daily	Europe	1980-2013	netCDF Grid	mm		ECA&D	
	European Climate Assessment	PP	sea level pressure	0.25 DD°	Daily	Europe	1980-2013	netCDF Grid	hPa		ECA&D	10.7GB
Weather	Oakridge National Laboratory	TMIN	Minimum Temp	0.125DD°	Daily	US	1980-2012	csv	°C	no	Daymet	
	Oakridge National Laboratory	TMAX	Maximum Temp	0.125DD°	Daily	US	1980-2012	csv	°C		Daymet	
	Oakridge National Laboratory	TDAY	Average Temp	0.125DD°	Daily	US	1980-2012	csv	°C		Daymet	
	Oakridge National Laboratory	PRCP	Total Precip	0.125DD°	Daily	US	1980-2012	csv	mm		Daymet	
	Oakridge National Laboratory	VP	Vapour Pressure	0.125DD°	Daily	US	1980-2012	csv	Pa		Daymet	
	Oakridge National Laboratory	SRAD	Short Wave Radiation	0.125DD°	Daily	US	1980-2012	csv	Wm-2		Daymet	
	Oakridge National Laboratory	DAYLEN	Length of Day	0.125DD°	Daily	US	1980-2012	csv	s		Daymet	25.7GB
Soils	Harmonized	MU_GLOBAL	global mapunit	1km	N/A	Global	N/A	grid	code	yes	HWSD	1.74 GB
	Harmonized	T_Texture	texture of topsoil (0-30cm)	1km	N/A	Global	N/A	grid			HWSD	
	Harmonized	Drainage	drainage class desrciptions	1km	N/A	Global	N/A	grid			HWSD	
	Harmonized	AWC_Class	available water capacity	1km	N/A	Global	N/A	grid			HWSD	
	Harmonized	T_Sand	% by Wt of sand	1km	N/A	Global	N/A	grid	%		HWSD	
	Harmonized	T_Silt	% by Wt of silt	1km	N/A	Global	N/A	grid	%		HWSD	
	Harmonized	T_Clay	% by Wt of clay	1km	N/A	Global	N/A	grid	%		HWSD	
	Harmonized	Texture	USDA texture class	1km	N/A	Global	N/A	grid			HWSD	
	Harmonized	T_Bulk_Density	bulk density of soil	1km	N/A	Global	N/A	grid	kg/dm3		HWSD	
	Harmonized	T_pH	soil reaction	1km	N/A	Global	N/A	grid	–log(H+)		HWSD	
Soils	Natural Resources Conservation Service	OM	Organic Matter		N/A	Global	N/A	polygon	%	yes	SSURGO	5.63GB
	Natural Resources Conservation Service	Depth	drainage class desrciptions		N/A	Global	N/A	polygon	cm		SSURGO	
	Natural Resources Conservation Service	AWC	available water capacity		N/A	Global	N/A	polygon	cm		SSURGO	
	Natural Resources Conservation Service	Sand	% by Wt of sand		N/A	Global	N/A	polygon	%		SSURGO	
	Natural Resources Conservation Service	Silt	% by Wt of silt		N/A	Global	N/A	polygon	%		SSURGO	
	Natural Resources Conservation Service	Clay	% by Wt of clay		N/A	Global	N/A	polygon	%		SSURGO	
	Natural Resources Conservation Service	pH	USDA texture class		N/A	Global	N/A	polygon	–log(H+)		SSURGO	
	Natural Resources Conservation Service	Bulk_Density	bulk density of soil		N/A	Global	N/A	polygon	kg/dm3		SSURGO	
Topography	European Climate Assessment	Elev	Elevation	0.25 DD°	N/A	Europe	N/A	netCDF	km	no	ECA&D	367KB
Topography	National Oceanic and Atmospheric Administration	Elev	Elevation	1km	N/A	Global	N/A	grid	km	no	GLOBE	7.71GB

Figure 8. A summary of the environmental data, variables and associated relevant spatial attributes of the data

Data Integration and Aggregation

As described in the previous section, scientific analysis and environmental modelling often requires the collection of heterogeneous data from multiple sources. As such, the collected data must be organized in a structured manner. Data integration and aggregation was achieved by developing specific algorithms for sorting through data for an area of interest. Often data must be converted from machine or computer read formats to tabular and subsequently graphical formats which are more readily interpretable. Figure 9 illustrates a simple scenario where binary data is converted to graphical data using C# code. Such is the case with climate data from ECA&D which was downloaded as a netCDF file and converted to tabular format for interpretation.

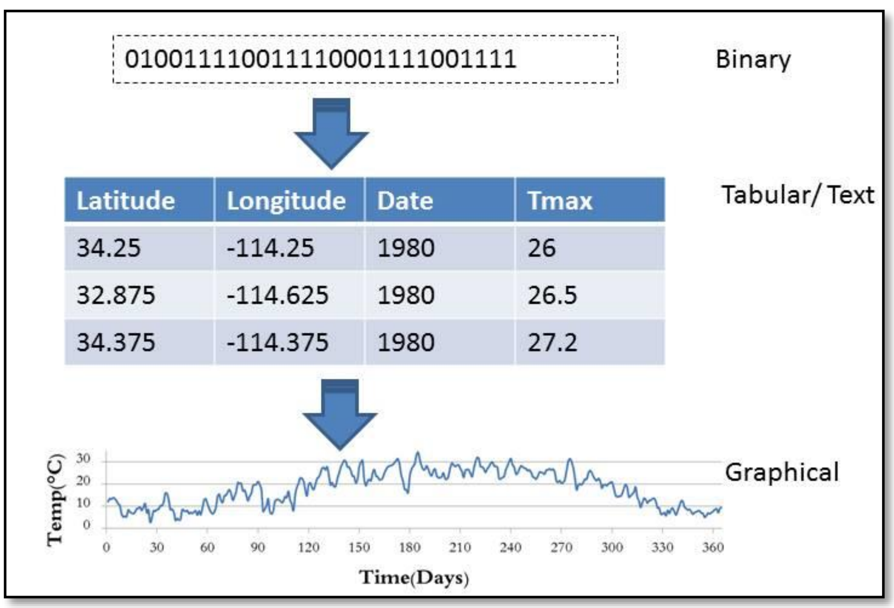


Figure 9. A description of the data conversion from machine readable formats to human readable interpretations

We designed simple algorithms that convert polygons to rasters, based on an integer assigned to a map unit or area of common soil attributes. Each pixel of the resultant raster is associated to a particular latitude and longitude. Figure 10 is an overview of the data processing performed on each data set.

Source	Category	Format	Type	Data Extent		Data Resolution		Data Processing
				Spatial	Temporal	Spatial	Temporal	
DAYMET	Climate	Text file	Interpolated	US	1980-2012	1km	Daily	yes; resampled at regualr intervals of 0.125°(13km)
ECA&D	Climate	Gridded	Interpolated	Europe	1980-2013	0.25°(27km)	Daily	no
NCDC	Climate	Text file	Station	Global	1960-2012	Variable	Daily	no
SSURGO	Soils	Polygon	Vector	US	NA	na	NA	yes rasterized from polygon to grid
HWSD	Soils	Gridded	Raster	Global	NA	1km	NA	no

Figure 10. An overview of the data processing of environmental data

Every location on the earth's surface can be described by a latitude and longitude. Our algorithms were designed to receive a single latitude and longitude as input or an array of latitudes and longitudes. This input is then used to query or search through all the available environmental data within the designated data category for the appropriate data file at the desired location. The data is then transformed and integrated from its original format using proprietary algorithms to resolve heterogeneity in data structure. For the purposes of this research and due to the voluminous nature of the data, we sought out ways to quickly query data for any given location. The result was tabular data that is uniformly interpretable and standardized to fit the needs of the analysis. This is illustrated in figure 11 where we demonstrate the use of specific C# algorithms to organize the raw data into a format amenable for our analysis. The details of the algorithms are defined by the user and are dependent on the desired structure of the data for future analysis. The user defines variables along with the temporal resolution of the data returned. For example, data can be returned as annual averages based on every day of the year for a specified time frame.

We designed two primary functions which received locational information as inputs. Depending on the location(s) of interest and category of data required, the appropriate data file was opened, sorted, and organized into a tabular form with the relevant variables for analysis.

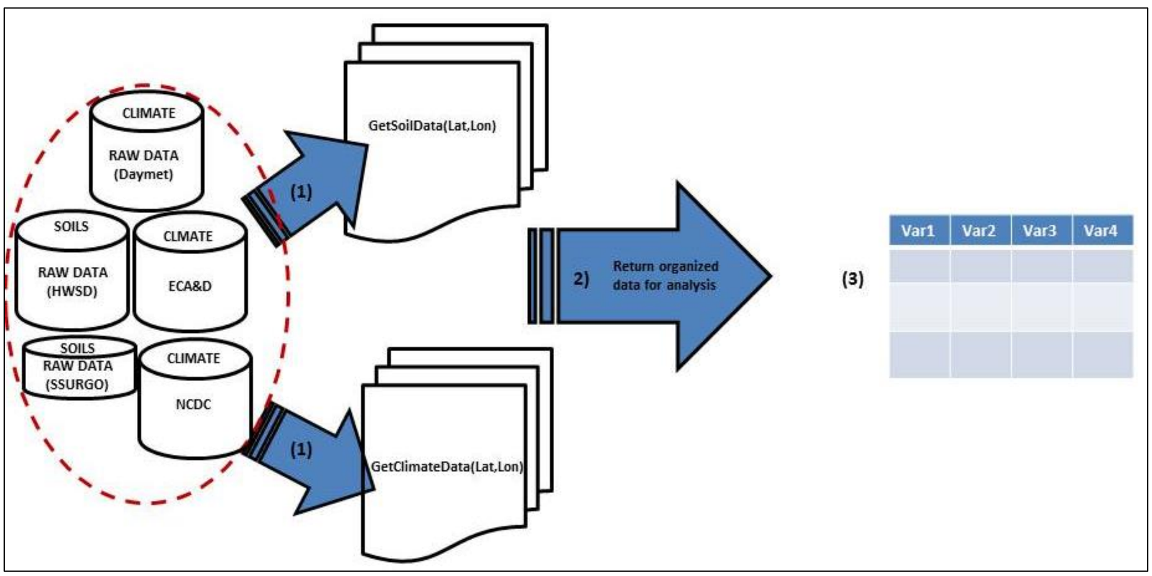


Figure 11. An illustration of the proprietary functions designed to query raw environmental data

Data Modeling and Analysis

Analyzing and querying 'big data' is fundamentally different from traditional statistical analysis on small data samples. Data analysis is therefore the process of scientifically applying statistical techniques to evaluate and describe data. Different analytical procedures allow us to draw inductive inferences from data and distinguishing the phenomenon of interest from the statistical variations present in the data. The data is often heterogeneous, inter-related, and inherent with source error hence the need for careful analysis. In this research we used statistical methods like multiple linear regression techniques to model the relationship between two or more environmental variables (independent variables) and a dependent variable. This is achieved by fitting a linear equation to observed environmental data.

Data Interpretation

Ultimately, data analysis must yield some interpretation of the data. This generally involves examining all the assumptions and retracing all the inherent sources of error associated with the data. We defer to the expertise of viticulturist and viticulture literature in order to draw conclusions about the analysis of our data and the relevance to viticulture. At this stage of the research, our interpretations of the data are limited to understanding whether the results make sense in the context of viticulture. For example, does an average annual GDD of 2500 in °C or a RPMT of 19°C make sense for a particular location?

Results and Discussion

The goal of this research was to describe a process for organizing large, complex, and heterogeneous data sources to enable rapid retrieval of data for environmental modelling of any location. We assessed the utility of our approach by quantifying our methods. Table 12 is a summary of data processing as they relate to measures taken to process data. Examples include data access times and data storage that result from our analysis of climate and soil data used in this research.

Source	Category	Data Processing	Access	Storage
DAYMET	Climate	Resampled at every 0.125° lat and lon for US	700ms	25.7GB
ECA&D	Climate	No processing; Data is queried using lat and lon	14237ms	10.7GB
NCDC	Climate	Queried on nearest lat and lon	7120ms	15.1GB
SSURGO	Soils	Rasterized then queried on lat and lon associated to attribute table	424ms	5.63GB
HWSD	Soils	Queried on lat and lon associated to attribute table	2178ms	1.74 GB

Figure 12. Results of the data access times and required storage space for various data sources utilized in the research

Daymet data in its native format was sampled at a 1 km x 1 km spatial resolution for the extent of North American. At this daily temporal resolution for the period of 1980-2012, the volume of data is not manageable for the untrained user. We therefore downloaded the full extent of the data at every 0.125° latitude (Lat) and longitude (Lon). The result was a more manageable number of data files at a consistent Lat and Lon interval. Data access time prior to the adjustment was immeasurable but after adjustments we were able to minimize access time to 700ms using a customized algorithm as illustrated in the appendix. By virtue of our approach we were able to sort through 32 years of daily climate data and 45,000 files in a matter of seconds. Figure 13 illustrates the file database of climate data which is quickly queried for any Lat and Lon over the extent of the data which represents the entire U.S. We can now query any single location or number of locations in the U.S. by simply knowing the latitude and longitude. Consequently a user defined algorithm returns an organized structure of weather data for the period of 1980-2012 in matter of seconds. This organization eliminates the need for sorting through 32 years and thousands of files of daily weather data.

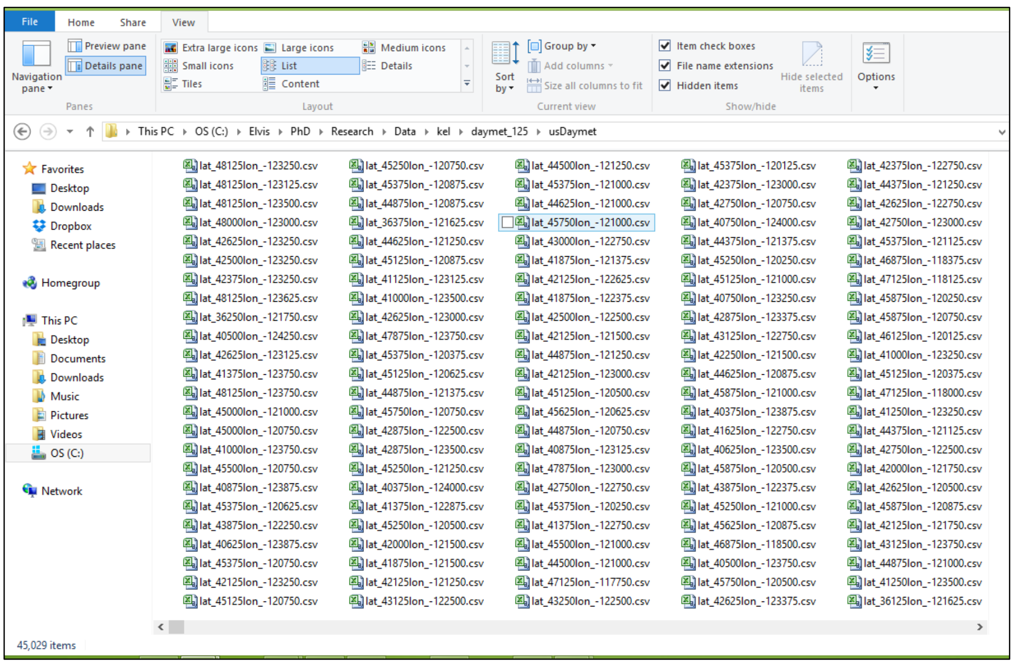


Figure 13. An illustration of the file database of Daymet climate data illustrating the vast size with 45,029 files of data

Similarly our daily global station data was queried in 7120ms, sorting through 45 years of weather data for the nearest weather station relative to the input latitude and longitude. Our data queries were based on the same principles as the Daymet data source.With this data source, every weather station has an associated latitude and longitude stored in an SQL database. When a location is requested by a user, the appropriate weather station is retured by virtue of the latitude and longitude. If the exact location is not available, the nearest weather station is returned to the user. The returned weather location is used to query the climate data base which is organized by year. Each year is queried for the appropriate weather station and all the relevant climate data is returned to the user in an organized and sturctured format. This data can now be used for further analysis. By vurtue of a user defined algorithm, 45 years and thousands of files are sorted in matter of seconds for any location as shown by the global distribution of weather stations in figure14.

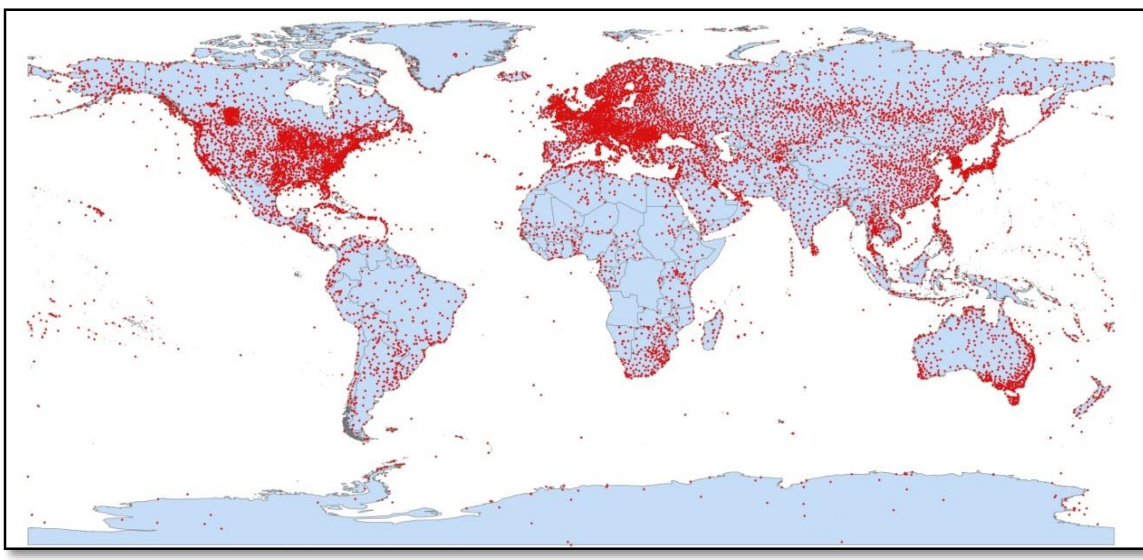


Figure 14. An illustration of the global distribution of weather stations representing a time frame of 1960-2012

Every year with available climate data is organized in a separate file by year. Figure 15 illustrates the organization of data with every year from 1960-2012 along with a list of every station for a given year. The number of stations with available data vary from year to year and from one location to the next. More developed parts of the world with advanced technolgy and research capabilities represent the greatest coverage as illustrated in Figure 14. The United States and Europe tend to have the greatest coverage of weather stations hence more comprehensive data coverage.

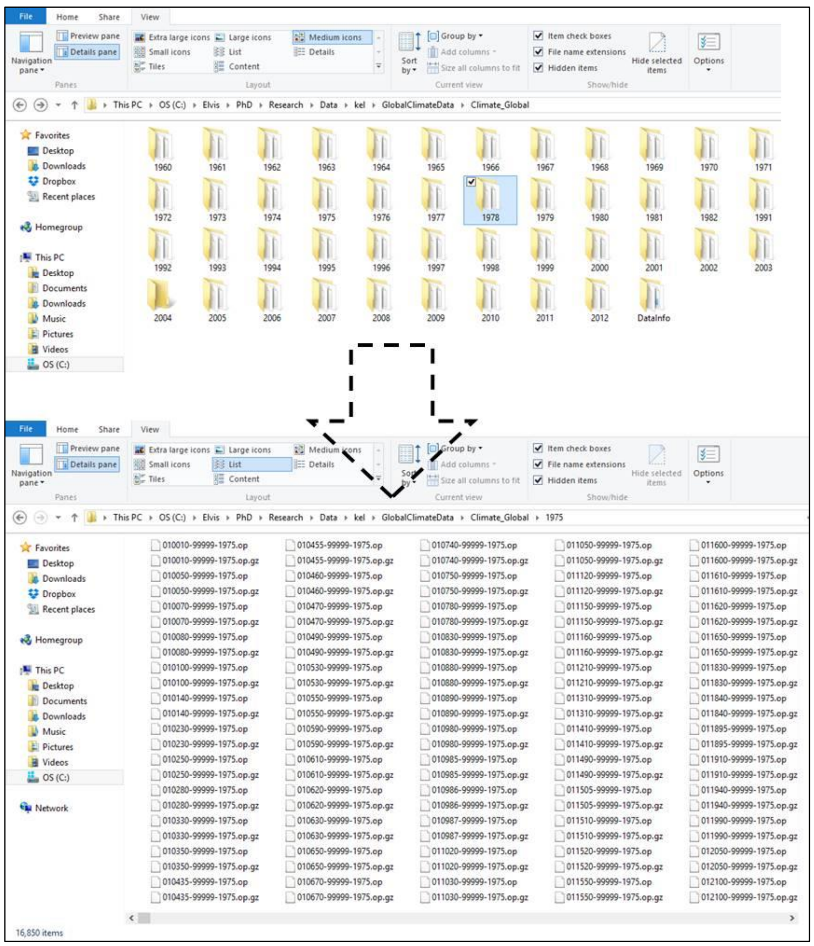


Figure 15. A depiction of the data organization and structure of global weather stations

Though most raw data used for this research was maintained in its native format, SSURGO soils data for the entire US was converted from vector to raster format. This conversion ultimately increased the required storage space but reduced data access time for a particular Lat and Lon. The original data set included thousands of soil polygons linked by an attribute table in a Microsoft access database. In its native format, SSURGO required a GIS environment, in depth knowledge of the data structure, and an ability to relate multiples tables in order to query the appropriate soil map unit for a particular location of interest. As such we can now query a gridded SSURGO soil data set of the entire U.S. which has been linked to a summarized attribute table for any particular location in a matter of seconds. Each cell or pixel at a particular location is linked to a map unit identifier called the map unit key by the latitude and longitude at that location. A unique map unit key is used to link the raster cells to an attribute table which describes the soils at the location of interest. All this information is queried and returned to the user in a matter of seconds. Refer to figure 12 for precise access times however our approach eliminates the need for relating multiple attribute tables in order to understand the detailed structure of the original soils database. The result is an organized structure of soil data which can now be used for more detailed analysis.

Conclusions

The goal of this study was to develop an approach for managing large amounts of data for environmental modelling and scientific research. We sought out to develop specific steps and general principles describing how we scientifically and efficientlly manage voluminous data in the modern era of technology driven scientific research. We began by emphasing the significance and value of data but equally important was an acknowledgement of a fundamental shift in how research and consequently how data is now collected. In today's modern era of technology, "Big Data" drives research thus our abiltiy to carry out envrionmental modelling for scientific research hinges on the ability to manage large datasets efficiently and effectively. The technical challenges of using big data are very real however the managerial challenges are even greater. Thus establishing scientifically objective principles and distinct steps for managing data at large scales is crucial to succeeding We described how we managed data for our research using a number of different data sources all adhering to the same priciples, namely (1) Data acquisition and collection, (2) Data integration and aggregation, (3) Data analysis and modelling and (4) Data interpretation. These principles ensure the perpetuity of the data for future research needs as well as underscore the value of appropriate data management

In the following section we illustrate how the methods described for managing 'Big Data' have been implemented for processing one of our many environmental data sets. The increased need to process 'Big Data' and turn it into useful information for decision making is at the forefront of this case study. The complexity of most environmental data sets typically limits use to scientific laboratories and academic institutions where the data can only be used by researchers who understand the complexities of the data. These complexities also include technical/technological challenges of managing different formats. The goal of this case study is to demonstrate the implementation of the procedures outlined in the previous section for managing 'Big Data'.

The KELSoil Web service: An application of the Soil Survey Geographic Database

Soil is an integral part of many ecosystem processes and functions. It plays an important role in decision making both by scientist and ecosystem practitioners. The Soil Survey Geographic Database (SSURGO) is a nationwide soil survey effort administered by the Natural Resources Conservation Service (NRCS). The project has generated massive amounts of spatial data describing the physical and chemical characteristics of soil. This data however presents a number of significant technical challenges to the average scientist or user. The SSURGO soils database is too large and complex for practical use in its current form. Much of the information collected is difficult to interpret without a significant understanding and background of soils. Moreover the technical challenges associated with managing SSURGO are beyond the ability of most users interested in soils data.

The goal of this study is to provide the average user with quick and easy access to SSURGO data by simplifying the data analysis process. This was achieved by implementing the principles and methods for managing big data described in this chapter thus enabling environmental modelling using SSURGO data. Our approach to this problem has been to download the SSURGO data, integrate it with other related data sources and develop sophisticated but easily interpretable outputs. This solution provides for simplified and interpreted versions of this information delivered across the web by simply clicking on a map. Furthermore using web services we have ensured that this technology can be integrated into any existing or future decision support websites or tools. In its existing form, the service provides a visual and interactive summary of soil properties at any longitude and latitude. We believe that in this form the availability of soil information will be directly and immediately useful to agriculturalists, engineers, and scientists. The type of information available as the result of our analysis includes but is not limited to soil depth of each horizon, soil pH, soil organic matter, soil available water capacity and soil texture. The iterative process of data analysis described in this chapter was used in our analysis of SSURGO data. We describe the specific steps in the following section.

Data Acquisition and Collection

SSURGO soils data is downloaded from the Natural Resources Conservation Service as soils polygons and associated attribute files in tabular format. In order to retrieve the data, a Microsoft Access soils database along with the spatial and tabular files specific to each county in the US must be downloaded to a local computer and decompressed. This resulted into a large collection of soil data with multiple tables of soil attributes

Data Integration and Aggregation

Spatial and tabular data files are then associated to each other using a database. The result is a US wide database that is several hundred megabytes in size with over 130 tables. The tables are interlaced together in a complex web of relationships. Figure 16 illustrates this complex system of tables and relationships. The database in its original format provides all soil survey documentation. This information is for a wide variety of users, consequently making it difficult to extract specific information. Ultimately there is too much information as it was designed to accommodate all users.

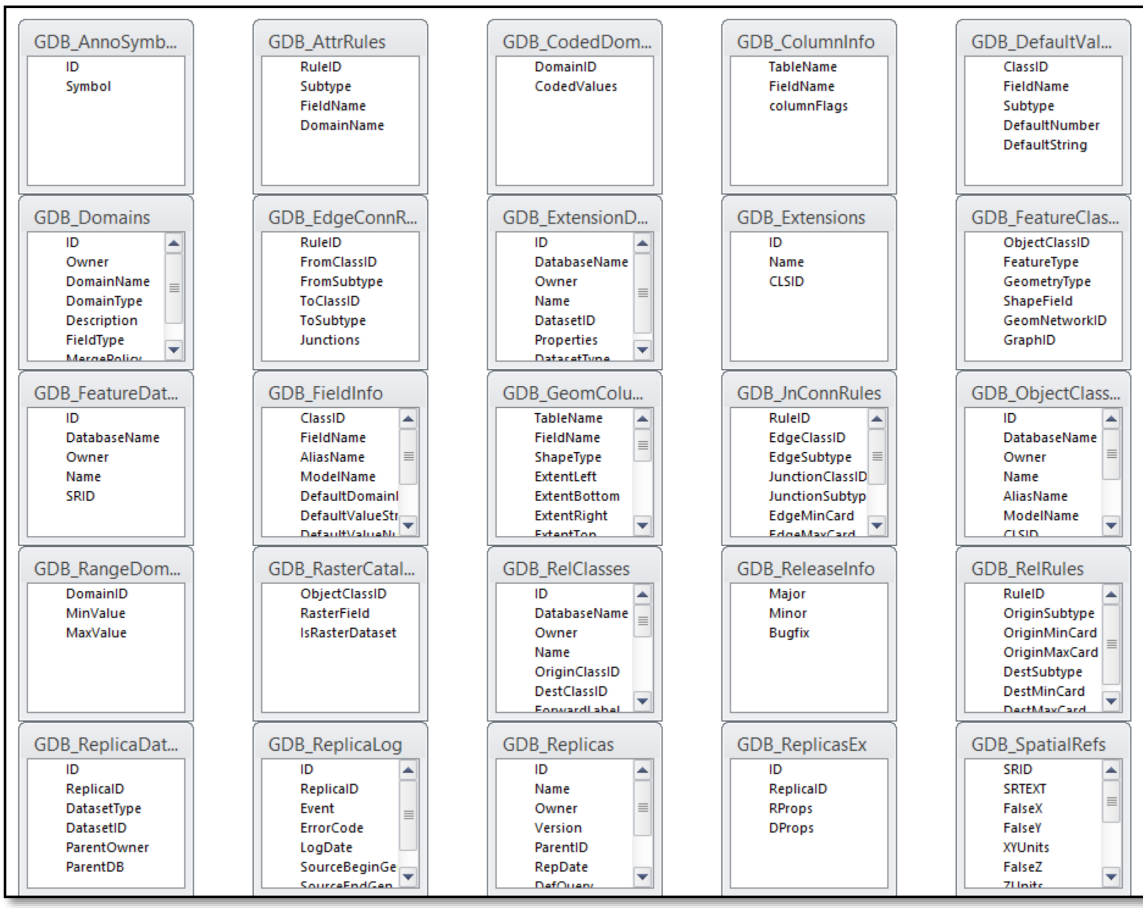


Figure 16. An illustration of the original SSURGO database showing data tables and the complex relationships

Our approach to this problem was to simplify the database using only the pertinent information for our research. This resulted in a reduction of the size of the database to a more manageable size compared to the original database at ten times the size. We prioritized the information that we wanted to retrieve and analyzed the database schema in order to understand the relationships between the tables. This resulted to 5 tables chosen to include in our new abbreviated database. These tables contained the majority of the information required. Figure 17 displays these five tables and the simple relationships between these tables. Spatial data in the form of polygons associated to the mapunit table by a unique mapunit key (MUKEY) is converted to a grid or raster. This is accomplished by using a proprietary C# algorithm which converts each soil polygon to a raster based on the MUKEY. We now have a single raster of the entire US with each pixel associated to a particular MUKEY which can be linked back to its relevant soil attributes.

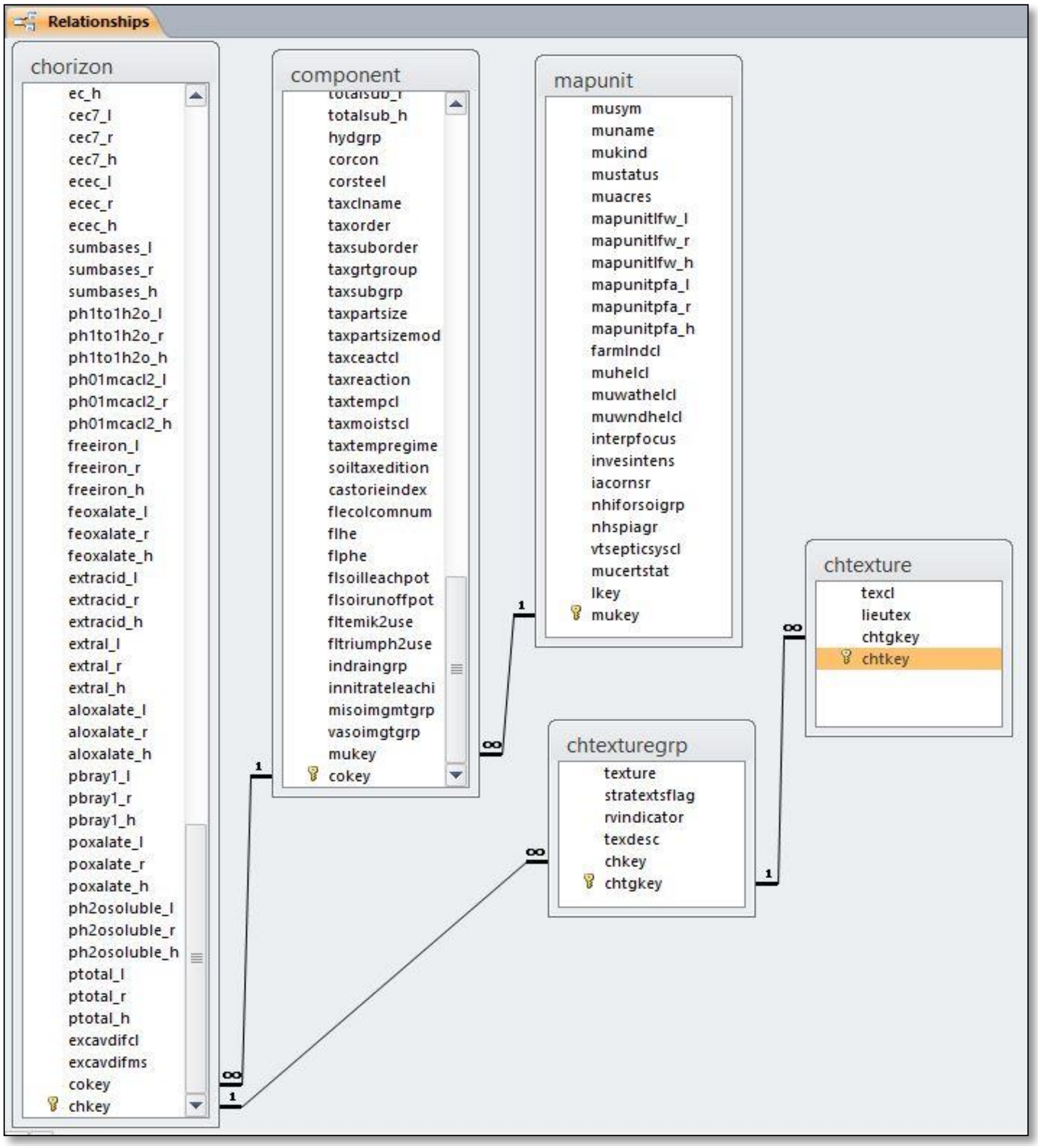


Figure 17. An illustration of the extracted tables and relationships in the simplified database showing the four tables

Data Modeling and Analysis

Five attribute tables now contain soil series information and have a 1to 1 and 1 to many relationship. There are one to many relationships between mapunit to component tables and component to chorizon tables. These relationships simply mean that there can be many soil components to a soil series and that there can be many soil horizons to a single soil component. As a result of these relationships, a GIS or map related application can drill down into the data to retrieve information about a specific location. Spatial data in raster format can now be associated to these attribute tables based on a MUKEY value associated to each pixel. We created a web-based system using the Microsoft® Bing Maps application program interface (API). This system allows a user to select and define an area of interest, consequently analyzing the location based on soil characteristics.

There are 3 options for choosing an area for analysis. A user can choose a single location, select a rectangular area or outline a very specific location, all based on latitude and longitude for the area chosen. By choosing a single location the user can select any area in the continental US. Alternatively by using the rectangle option as user can select any general area on the map. The most precise option is the polygon method with allows a user to define a very specific boundary and area of interest for analysis Figure 18 top, middle and bottom illustrates all three ways in which a user can define the boundaries of a specific area chosen.

Figure 18. Illustrations of various methods for defining an area of interest based on single location, rectangle, and polygon

The data returned is based on the latitude and longitude of the area selected. When a user clicks on the map, a latitude and longitude for that location is used to query our soil raster for the appropriate MUKEY at the clicked location. This MUKEY value is used to search through the mapunit table which is associated to the other 4 tables by MUKEY. The relevant soil attributes are returned and displayed on the screen in structured format that can be interpreted by the user. The alternative is sorting through 130 tables of soil data in an attempt to determine which relationships are valid ahead of organizing data for the location of interest. Figure 19 illustrates how the data for a single location is returned to a user for interpretation.

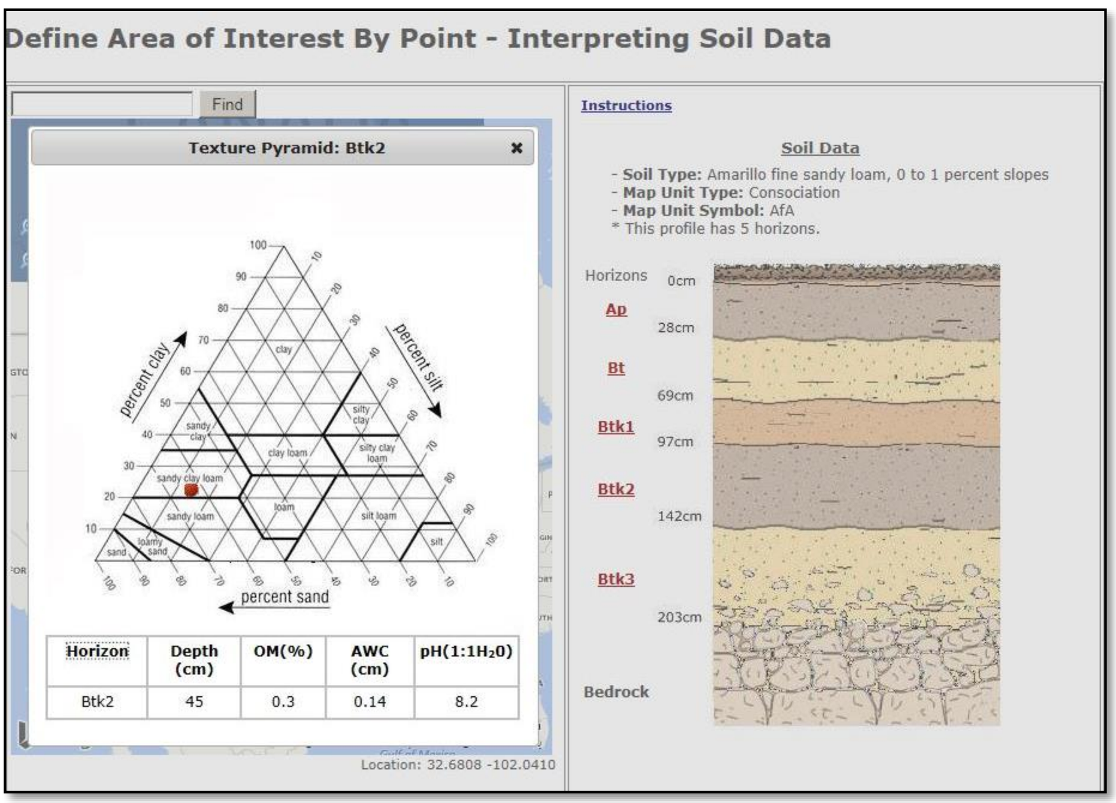


Figure 19. An illustration of results of soil data for a single location organized as a texture pyramid and soil profile

Alternatively, if a general area is selected or if an area is outline with specific boundaries, a summary of the MUKEY values are used to return data back to the user. Figure 20 illustrates a pie chart of MUKEY values for a selected location with data for the MUKEY with the greatest area represented in the profile.

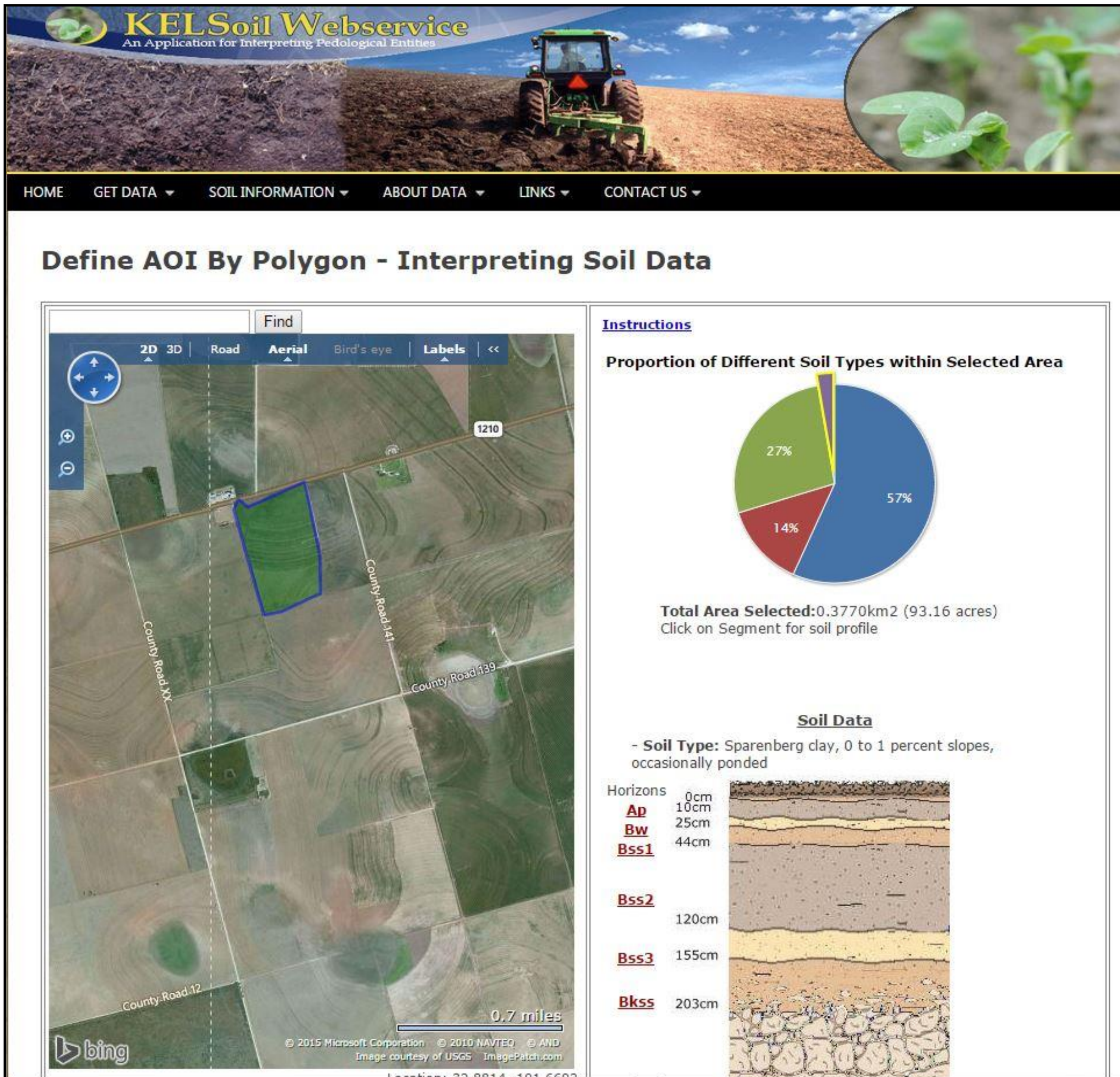


Figure 20. The results of an area of interest chosen for analysis is represented by multiple MUKEY values

Additional soil attributes for the other MUKEY values can also be viewed by clicking on a slice of the pie chart. Other attributes are layered by soil horizon and can be viewed by selecting one of multiple soil horizons. Figure 21 highlights the choice of a particular soil horizon revealing additional attributes in the soil pyramid and attribute table.

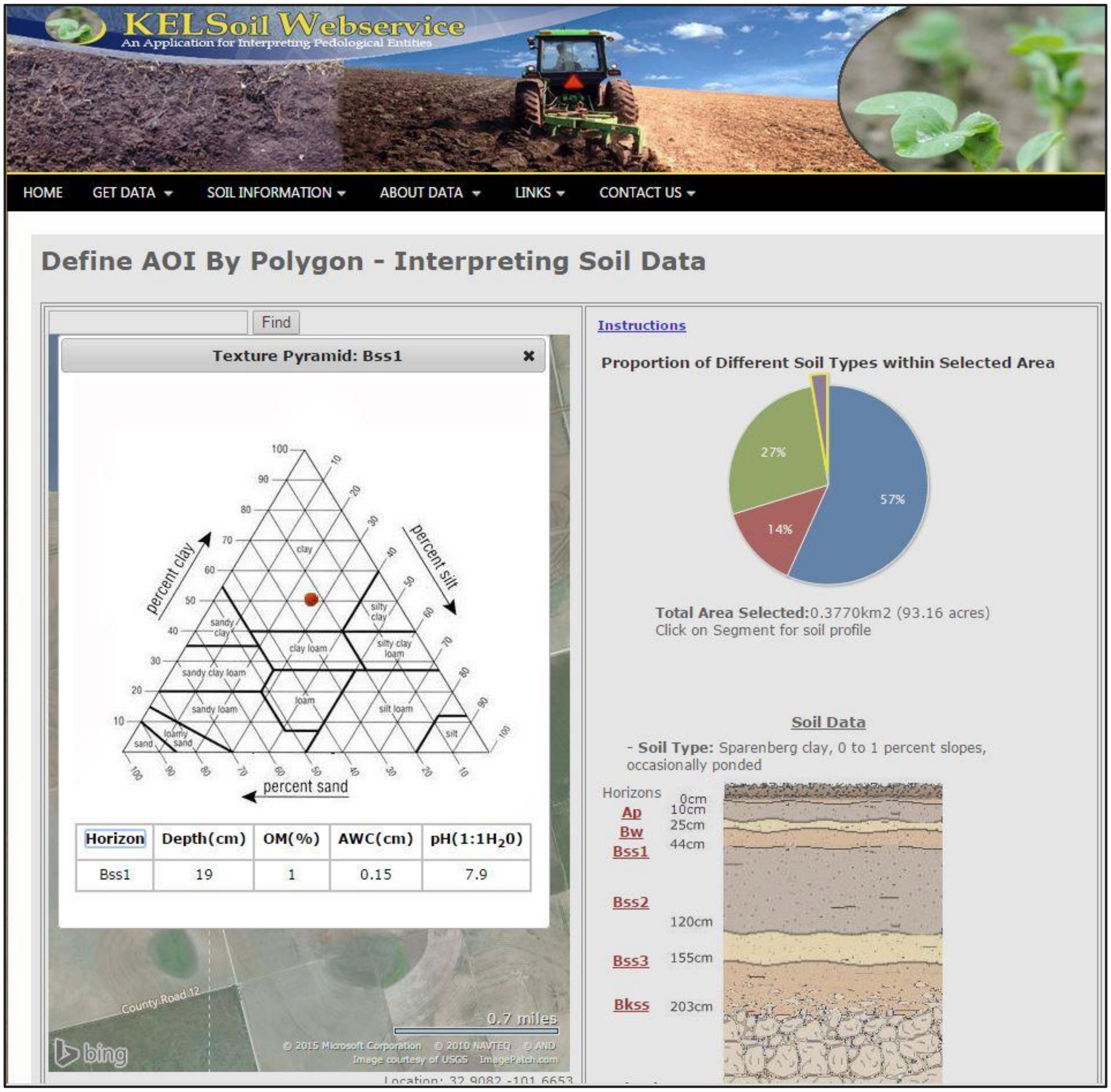


Figure 21. Soil texture pyramid displaying additional soil data layered by soil horizon

In order to provide access and use of SSURGO data to a more general audience, who may not be soil scientist or have access to GIS software we have developed a Web-based system. This system only requires a web browser and an Internet connection. The KELSoil Webservice® incorporates a large amount of complex soil data and simplifies it for use by non GIS users and scientists who have an interest in soil data. The goal was to simplify the information using the principles defined in this chapter for managing large heterogeneous data sets.

CHAPTER III

A CONCEPTUAL APPROACH TO BUILDING MODELS FOR SITE AND VARIETY SELECTION IN VITICULTURE

This chapter focuses on identifying the relevant aspects of viticulture that must be taken into consideration in order to successfully undergo the process of scientific modelling for site and variety selection. Viticulture particularly in new world regions is a relatively new industry but one that is growing and evolving. For individual viticulturists, there are significant temporal and financial risks associated with planting grape vines. These include the acquisition and preparation of land, the time taken for the vine to bear fruit, and the cost of replanting if the quality of grapes is not sufficient. In Old World regions years of trial and error in the selection of appropriate varieties has provided guidance which newer regions have not had. Given the risks and timelines involved in planting vines and producing high quality grapes, it is important to develop means by which viticulturists can use a scientific and objective approach to site and variety selection.

Introduction

A scientific model is a representation of an idea, process, or system used to describe and explain phenomena that may not be experienced directly. Modeling is central to scientific research as it guides the presentation of hypothesis and explanation of complex data. According to Schwartz et al (2009), a scientific model is a representation that abstracts and simplifies a system by focusing on key features to explain and predict scientific phenomena.

A growing wine industry in the new world necessitates the need to match varieties to appropriate environmental conditions. However in order to minimize the inherent risk a scientifically objective approach is necessary. The development of scientific models for variety selection begins with the question of ‘where can I grow a given commodity most successfully?’ This concept of success is often complex, including the economic, political and social motives of the industry. For different commodities there is often some element of productivity inherent to the success of the industry. Examples may include total production, the yield of the commodity or quite simply the price. In viticulture emphasis is generally placed on the quality of the grapes often measured in terms of yield per acre or the total price paid per ton. The goal of this chapter is to describe a conceptual approach to building scientifically objective models for site and variety selection in viticulture. We shall place emphasis on the most fundamental aspect of building models for site and variety selection, which is the choice of a consistent dependent variable.

Our approach was driven by the need to first establish a dependent variable representative of viticultural success. As such the primary objective associated with building models for site and variety selection is an understanding of the functional relationships between variety suitability, environmental conditions, and measures of success. The goal is to be able to extrapolate these models to other locations of interest.

Methodology

The first step in developing models for variety selection is to identify some measure of suitability or success of a variety. Site and variety selection usually involves multiple interacting environmental variables that influence suitability. Due to the subjective nature of suitability, creative measures must be taken in finding data sources most relevant to the problem. Biases must be dealt with often where dependent variables or measures of success are selected as surrogates for the ideal problem. For example, the price paid for a bottle of wine may serve as a substitute measure of the quality of grapes. Secondly, developing models for variety selection also involves collecting objective environmental data to define the relationship to the dependent variable. Limitations in the availability of environmental data often render this task challenging. We reviewed literature and consulted with viticultural experts in order to identify environmental data most relevant to the phenomena at hand. Takow (2008) outlines a number of factors relevant to successful wine grape production. The general approach to building models for variety selection is summarized in the following section.

A Scientific Approach to Site and Variety Selection

Site selection in viticulture often involves a number of compromises as few if any sites are ideally suited to wine grape production. In many cases site selection involves two general scenarios namely (1) Selection of potential sites most likely to support grapes of a given variety and (2) Selection of a variety that is most suitable for a particular plot of land.

Our approach applies to both scenarios. In this section we outline the steps to building a scientifically objective model for site and variety selection in viticulture. Our audience includes modern growers who may not necessarily come from an agricultural background hence decision support is the primary objective.

Establishment of Dependent variables

Data for site selection models must include a dependent variable that relates to the objective and at least 1 independent variable that represents an environmental characteristic of a site. Here the primarily goal is quite simply to establish a measure of suitability and a measure of the environmental conditions. For example, we may choose to assess the relationship between the yield of a particular variety and temperature conditions under which the variety was grown. Figure 22 illustrates a measure of suitability on the y-axis and some measure of environmental conditions on the x-axis.

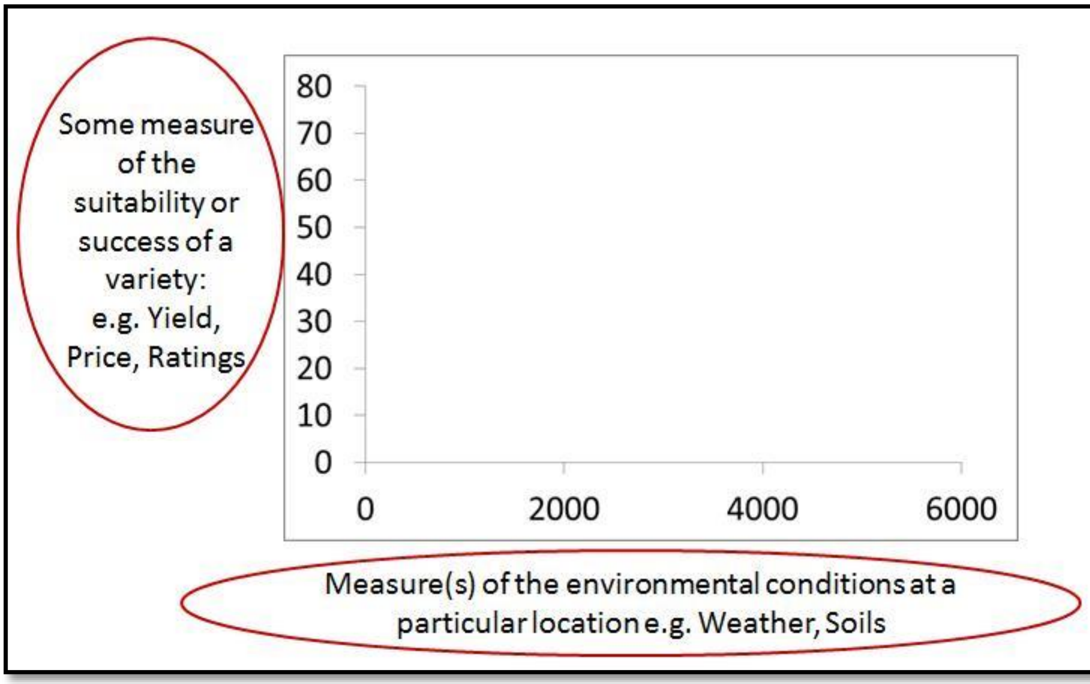


Figure 22. Illustrating the significance of determining measures of suitability and environmental conditions

A model is then fitted to the data in order to assess any relationships between the dependent variables and the independent variables (predictor) for a particular variety. A simple regression model can now be established relating measures of grape variety suitability to measures of environmental conditions at a particular location. Figures 23 shows the relationship between some measures of suitability for variety A plotted against measures of environmental conditions at a particular location.

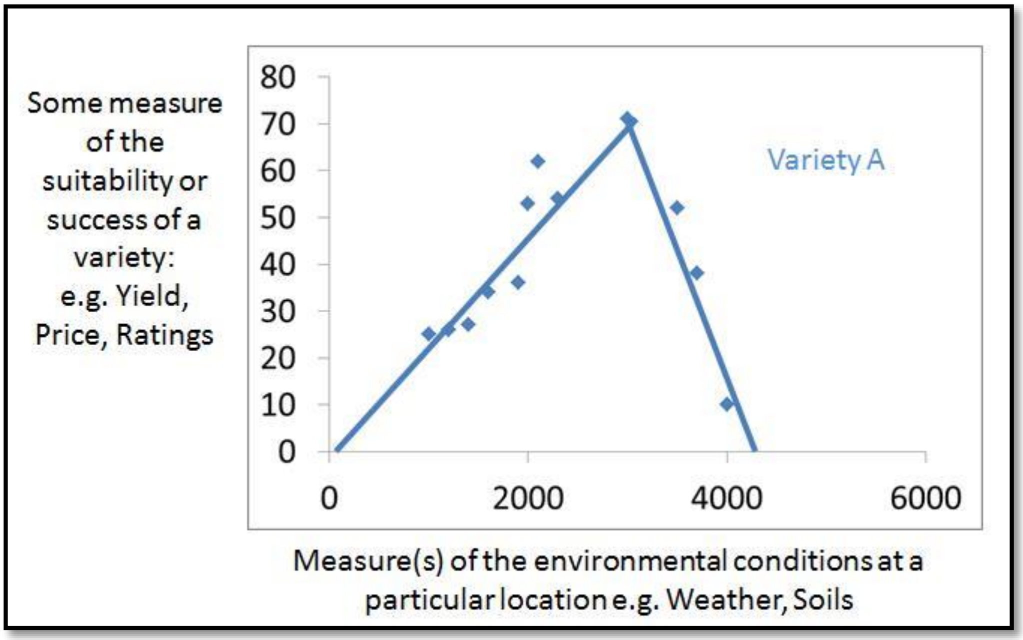


Figure 23. An illustration of the suitability plotted against environmental conditions for variety A

Derivation of Environmental Indices

Simplifying the environmental data into indices which relate back to the dependent variable is a crucial step in building scientifically objective models. The complexity of environmental conditions and associated variables like climate necessitates the derivation of indices which best represent conditions of the environment. Figure 24 illustrates how measures of the environment over time can be simplified to create an index of suitability.

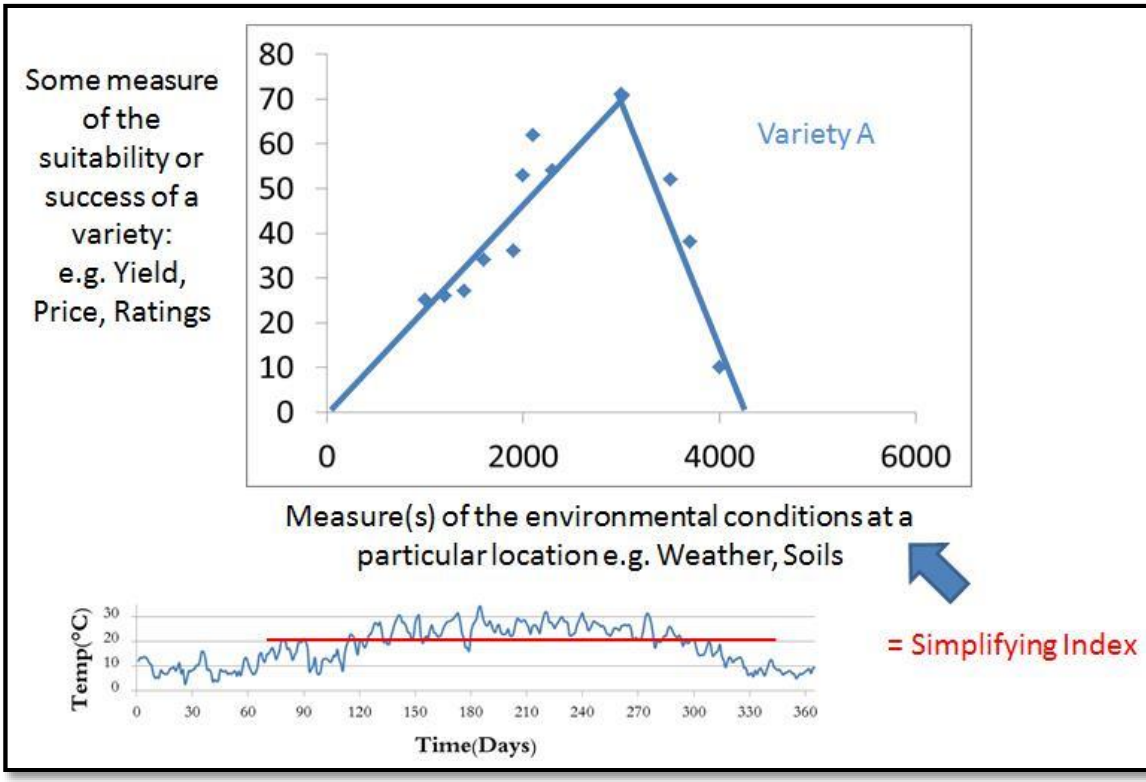


Figure 24. An illustration of the environmental conditions at a particular location are simplified into an index representative of the conditions at a particular location

Environmental data should also represent immutable mobiles, variables that do not need to be changed from one location to another. The immutable mobiles are a characteristic of scientific knowledge based systems. In the case of viticulture, environmental indices should not change depending on the location or over time. Figure 25 illustrates this concept of immutable mobiles as the same index of GDD is assessed for two different locations.

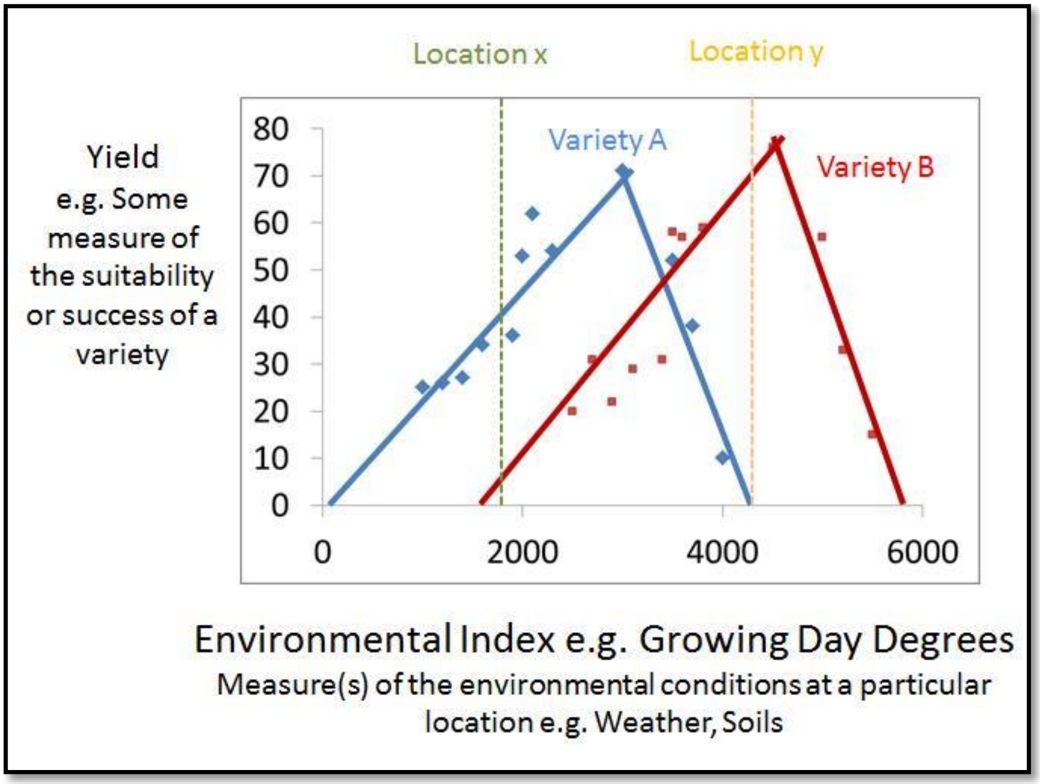


Figure 25. Highlighting the relationship between GDD and suitability assessed at 2 different locations

Model Interpretation

The final step in developing a site selection tool involves interpretation of the model results. Coupled with heuristic knowledge, the association between the dependent and independent variable should be assessed and applied to a new location. In our research this also involves an understanding of the significance of the variables in the model which is measured by the value of the individual model parameters.

Results and Discussion

A model for variety selection should be adaptable and valid in addressing the 'Problem'. It is therefore necessary to understand the importance of each variable to the model. In the case of viticulture, our model should be able to help us understand the contribution of each environmental factor thus understanding how important each variable is to the model. Figure 26 illustrates a conceptual design of the process of building scientifically objective models for variety selection in viticulture. The basic steps illustrated in the figure are applicable to any agricultural system and not restricted to viticulture.

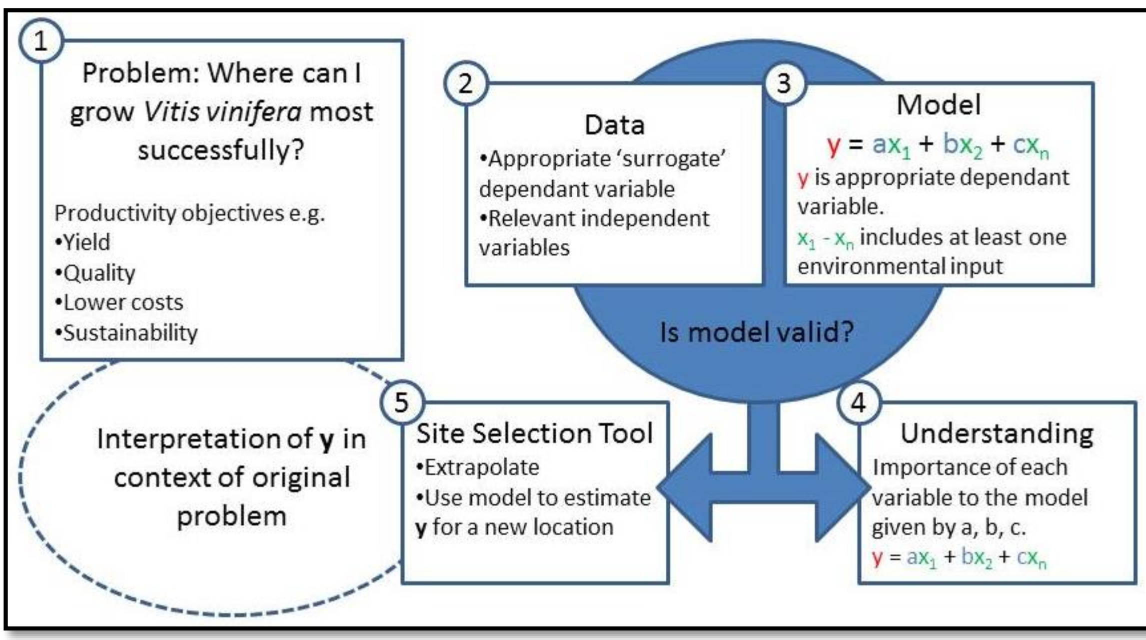


Figure 26. A conceptual design of the process of involved in the development of a model for variety selection

The significance of the variables x_1, x_2, and x_n to the model is measured by parameters a, b, and c. A valid model is one which, given $Y = ax_1 + bx_2 + \cdots cx_n + E$, 'Y' can be predicted upon establishing values for a, b, and c. These values will now allow the model to be used at any location (immutable mobiles) given known values of x_1, x_2, and x_n for that location. The ability to extrapolate the model for use at a new location is a measure of the utility of the model.

Conclusions

Overall, a scientific approach to building models for variety selection requires two types of data. Predictors or environmental data (Climate, Weather, and Soils etc.) are relatively easy to obtain, but have to be converted to environmental indices which relate back to the measure or phenomena in question. Modern scientific research faces the problem of 'Big Data' as outlined in the previous chapter hence the challenges of managing numerous large data sets from varying sources and varying scales. Additionally, which of these predictor's best represent conditions which assess the suitability of a variety of grape to a particular location? Grape quality or suitability is more difficult to quantify. Identifying the best measure of quality or success may demand the use of surrogates. In viticulture this can often include yield, bottles of wine for different locations, and the presence or absence of varieties grown at a given wine regions. In the absence of yield data the viticulture industry lacks a truly objective measure of viticultural success. Finally, the interpretation of the parameters in order to define which are most important in the context of viticulture is a necessary step in the process of building models for site and variety selection.

CHAPTER IV

THE UTILITY OF GROWING DEGREE DAYS AS AN INDEX FOR VITICULTURE

Site and variety selection in viticulture has long been recognized as important for successful grape and wine production (Jones and Hellman, 2003). This chapter presents a discussion on the use of the GDD concept to describe varietal suitability in the context of viticulture. We present a case for evaluating the limitations in the concept of an estimate of GDD for a particular location by considering of a number of factors which may influence its application for grape variety and site suitability in viticulture.

Introduction

Heat units or GDD are frequently used to describe the timing of biological processes. Temperature and time was first used by Réaumur in 1735 to describe development in plants and animals. Plants and invertebrate animals, including insects and nematodes, require a certain amount of heat to develop from one stage in their life cycles to another. The measure of accumulated heat is referred to as physiological time and provides a common reference for the development of organisms. The amount of heat required to complete a given organism's development does not vary. The combination of temperature and time will always be the same. This physiological time is often expressed and approximated in units called degree-days (°DD).Consequently, °DD models have become an integral tool in understanding insect and plant phenology. The essential assumption implicit to a degree day model is that plant and animal (poikilothermic) development is directly related to time and ambient temperature. In other words biological development is dependent on chemical reactions and temperature controls the developmental rate of many organisms. As these reactions occur over time, development proceeds. The goal of this chapter is to assess the utility of growing degree days as an index of grape variety suitability.

Numerous bioclimatic indices have been used to measure grape variety suitability and are mostly developed on the basis of climatic variables. GDD is historically and currently the most commonly used measure of climatic suitability for viticulture. It is important however to understand the limitations inherent with the use of GDD. The utility of GDD is driven by two factors, namely simplicity and applicability. The former refers to the availability, reliability and the ease with which complex environmental data can be reduced to a single index. The latter refers to the scientific and objective underpinning of GDD and the relationships to viticulture. This chapter will focus on the usefulness of GDD in viticulture for grape variety suitability. As such we shall address the following objectives:

1. The methods and relative simplicity of calculating GDD
2. The variation in GDD as a result of using data at different temporal resolutions
3. The applicability of GDD in the context of viticulture

There exist numerous methods for calculating GDD. In its most basic form, it is computed by subtracting a base temperature from the average temperature (usually rounded to the nearest degree) for the day.

Swiss botanist A.P. de Candolle observed that vine growth started when the mean daily temperature reached 10ºC. This led to the idea of a heat summation above a base temperature defining vine growth and grape maturation. Amerine and Winkler (1944) elaborated on this concept by developing an index of heat summation for California that is now widely used as a guide for selecting appropriate grape varieties and for determining a given area's suitability to produce quality wine grapes. The heat summation index is calculated for the period of April 1 through October 31 in the Northern Hemisphere by summing each day's average temperature above the base of 50 ºC (10ºC). This time frame represents the growing season of the grape vine in the northern hemisphere as described by its life cycle. This base temperature is assumed to be the minimum temperature observed at which vine growth occurs. Amerine and Winkler consequently defined five climatic regions or Winkler zones for California in °F with recommended varieties best suited to these regions. Our goal remains to understand the applicability of GDD as an indicator of wine grape suitability. The idea is to examine whether a single number can usefully summarize the environmental conditions at a location. Hence how representative is an estimate of GDD and what is the significance to wine grape growth?

The Importance of Bioclimatic Indices like Growing Degree Days

Plant growth is clearly driven by sunlight (photosynthesis), temperature (rate of photosynthesis and respiration), moisture availability, and nutrient availability (Leopold, 1964). These environmental factors are manifested through complex weather patterns. In order to draw general conclusions about the effects of these weather patterns on a plant, we need to simplify them into indices. Simplification should provide explanatory power for a phenomenon as well as be practically possible given the availability of data.

In viticulture GDD is an index representative of the degree day accumulation for a particular location over a specified time frame (growing season). If the degree day accumulation is greater than the required accumulation for a particular variety, then the variety can grow to maturity at that location. Conversely if the degree day accumulation is less than the required accumulation for a particular variety, then the variety cannot grow to maturity at that location. GDD, as applied to viticulture is used to develop limits. These limits define which varieties can reach maturity and which cannot at a specific location. Put simply, GDD is a number that tells you whether the growing season contains enough days when the temperature is within a range that is conducive to the production of "quality" grapes.

Materials and Methods

To assess the utility of GDD as an index for viticulture, an analysis of an estimate of GDD was carried out. The research began with the construction of a climate database of weather data sourced from Daymet (Thornton *et al*., 1997). We discuss the details of this process in chapter 2 (Data Acquisition and Data Development) of this dissertation. Our weather data consisted of daily climate variables for the conterminous United States at a 1-km resolution (http://www.daymet.org). This daily gridded weather data is downloaded with variables of daily maximum and minimum temperature, day length, precipitation, solar radiation, and humidity for the period of 1980-2012. The first step for a user is to extract data for the area of interest. In our case we sampled the entire U.S. at an interval 0.125degrees latitude and 0.125 degrees longitude. The results were text files of daily weather data for the entire U.S. named according to the latitude and longitude of each location (ex. lat_25125lon_-81000.csv). Each text file was then queried using proprietary C# code to calculate the cumulative GDD of each year for the period of April 01st to October 31st. This code is detailed in appendix A. The results of each year for every location were stored in an excel spreadsheet based on the corresponding latitude and longitude.

The next stage of the analysis involved importing the results of our calculations into Esri's ArcGIS software, ArcMap. The results were imported as points corresponding to each latitude and longitude over the entire US. These points were then converted to grids at a resolution of 0.250 decimal degrees in order to visualize the spatial distribution of GDD estimates across the U.S. The result at this stage of the analysis was a grid of the average GDD at a given location in the US for the period 1980-2012. We also used the point data to assess the correlation of GDD estimates calculated at different temporal resolutions of the data. We assessed inter annual variability of GDD by calculating the coefficient of variation to determine how much an estimate of GDD varies from year to year compared to the average. We deduced a model for predicting GDD by using a multiple regression analysis with elevation, latitude and longitude as the key predictors. Finally, we assess which, if any environmental variables does GDD actually represent.

The following sections will detail the specific methods and analysis undertaken in order to assess the utility of an estimate of GDD.

Estimating GDD at different temporal resolutions

It is important to understand how an estimate of GDD depends upon the temporal resolution of data. In the modern era of scientific research, data acquisition is driven by advances in technology. As such data collection is automated with raw data often stored in databases at varying resolutions and formats. The simplicity and practicality of calculating GDD was assessed by examining estimates of GDD at different temporal resolutions of the data. Many methods of calculating GDD have been successfully used in agricultural sciences (Allen, 1976; McMaster et al, 1997, Cesaraccio, 2001). Particularly in the areas of crop phenology and development, the most commonly used equation for calculating GDD is as below:

$GDD = \Sigma[(T_{MAX}+T_{MIN}) \div 2]-T_{BASE}$…. (1)

T_{MAX} is the daily maximum air temperature, T_{MIN} is the daily minimum air temperature and T_{BASE} is the temperature below which the process of interest does not progress. In the context of viticulture, T_{BASE} has been established at 50 ºF (10ºC), the temperature below which grape vine growth does not occur.

Alternatively equation 1 can also be simplified by using the daily average air temperature. By setting the quantity $[(T_{MAX}+T_{MIN}) \div 2]$ from equation 1 equal to T_{AVG}, the result is the following adjusted equation for GDD.

$GDD = \Sigma T_{AVG}-T_{BASE}$…. (2)

In our study, estimates of GDD were calculated using daily and monthly averages of climate data from the Daymet database. GDD calculated using monthly averages was done by summing the daily average maximum and minimum temperature for each day of the month from April 01st to October 31st (growing season in the Northern Hemisphere). An estimate of GDD was then calculated for each month using equation 2 and multiplied by the number of days in the month. The cumulative GDD was determined for the year by summing up the individual GDD values over the period of April 01st to October 31st. These calculations were all carried out using proprietary C# code outlined in appendix A and discussed in chapter 2 of this dissertation.

In order to calculate hourly GDD we interpolated the daily temperature minimums and maximums. This was done by fitting a sine curve to daily maximum and minimum temperatures in order to estimate how temperature changes throughout the day (Baskerville and Emin, 1969; Cesaraccio et al., 2001). The assumption is that the temperature cycle is approximated by a sine wave. As such the temperature cycle can be (1) completely above the base temperature, (2) completely below the base temperature or (3) intercepted by the base temperature.

Assessing Inter annual or year to year variation

Climate variability impacts agriculture, particularly grape and wine production which has demonstrated a narrow niche of climatic suitability (Jones 2006). Recent research on the impacts of climate in viticulture has focused on the influences of short-term climate variability on grape vine phenology, production, and quality (Jones and Davis, 2000). Daymet data was used to derive annual GDD estimates for the entire U.S.

for every year from1980-2012. These estimates were then summed to calculate an average GDD value. We then calculated a coefficient of variation (CV) by estimating the ratio of the standard deviation to the average of GDD.

Spatially interpolated versus Station data

For any given location in the US, GDD was calculated using data from the nearest weather station. Weather station data was sourced from the National Climatic Data Center (http://www.ncdc.noaa.gov/data-access). We used a latitude and longitude intervals of 0.125 decimal degrees as a search radius in order to examine weather station locations. Using proprietary C# code in appendix A, we were able to determine the nearest station in our global climate database. The input latitude and longitude was compared to latitude and longitude of stations in our global climate station database in order to establish the nearest weather station. The weather data from the nearest station was then used to calculate an estimate of GDD at the given location.

GDD was also calculated from the interpolated Daymet data set as describe in the previous section on estimating GDD at different temporal resolutions. Differences in estimates for the same location were assessed by creating a scatter plot of interpolated versus station data as well as an ESRI grid of differences between interpolated and station data estimates of GDD. This was done by importing data points of station and interpolated weather data into ArcMap. These points and the associated estimates of GDD were then used to create a grid of station GDD and one of interpolated GDD. Figure 27 shows a visual comparison of interpolated versus station data GDD across the entire US.

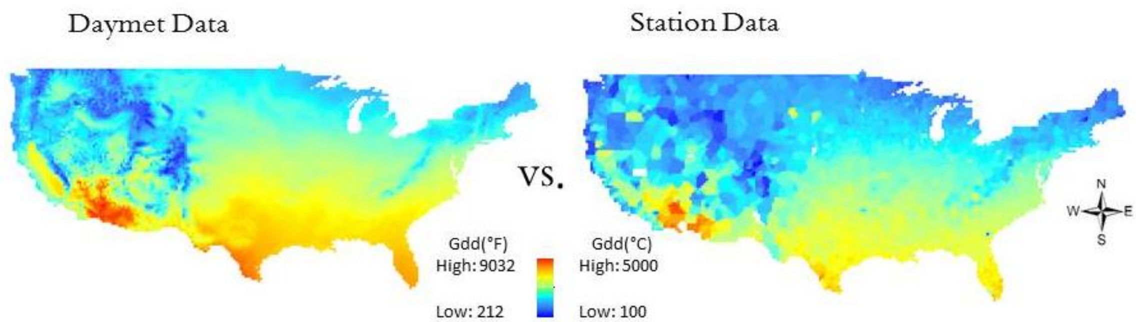


Figure 27. A visual comparison of interpolated data and station data GDD for entire US

Assessing the Spatial Differences in GDD

We assessed the spatial differences in GDD estimates by deducing a model that predicts GDD using only inputs of elevation, latitude, and longitude. This was done by using a multiple linear regression analysis to model the relationship between elevation, latitude, and longitude to an estimate of GDD. We used a statistical software package called JMP® to fit a linear equation to estimates of GDD. This equation describes how estimates of GDD change with changes in elevation, latitude and longitude. The general multiple regression model is of the form:

$y = b_0 + b_1 x_1 + b_2 x_1 + \cdots + b_n x_n + E \ldots.$ (3)

where b_0 to b_n are partial regression coefficients. x_1 to x_n are the measured variables (elevation, latitude and longitude) and E is the error term.

Assessing the relationship between GDD and temperature

Intuitively, one would think that GDD relates directly with average temperature. However the difference between GDD and GSAT can be explained by weather patterns. We assessed the relationship between GDD and growing season average temperature (GSAT) for a particular location by addressing 3 different scenarios where GSAT and GDD are measured. In each scenario the GSAT is the same however the differences in daily average temperature result in varying DD accumulations.

Results

Temporal Resolution

Across locations in the US, there is up to a 10% difference in the GDD calculated hourly versus daily versus monthly. Figure 28 shows the percentage difference in GDD estimates across the US, calculated at different temporal resolutions of the climate data.

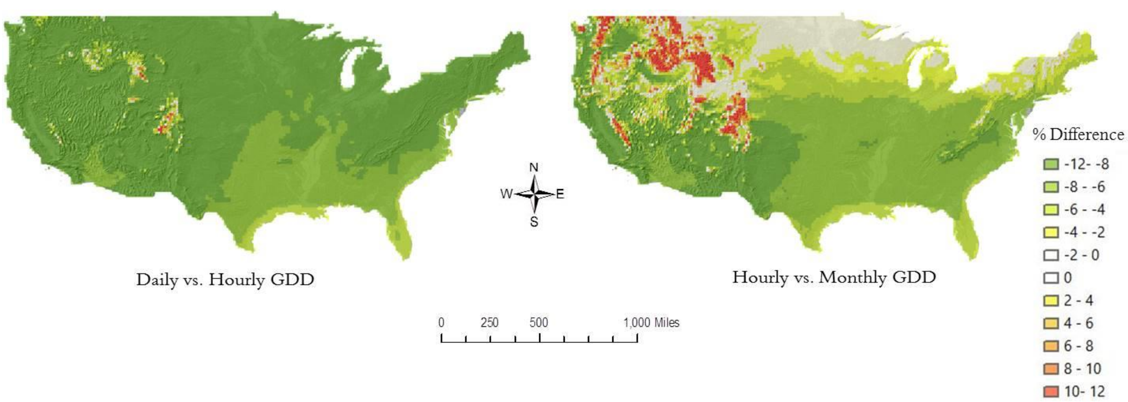


Figure 28. An illustration of the percentage difference in GDD estimates compared at different resolutions

Inter annual Variation

The GDD at a particular location varies from year to year. On an average year the GDD across the US varies as illustrated in figure 29.

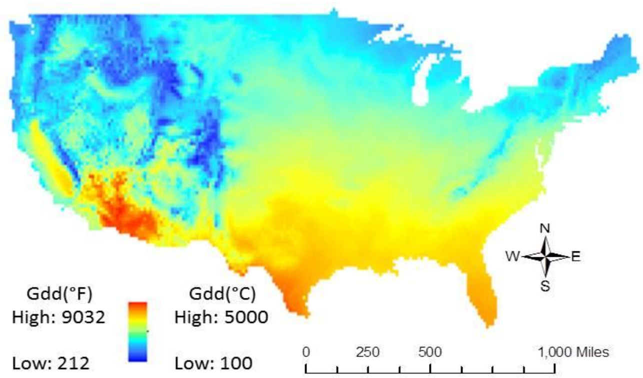


Figure 29. A depiction of the average annual year to year variation in GDD across the US

We estimated that there is as much as a 79% variation from year to year in an estimate of GDD at a particular location. Figure 30 shows the variation from the average in an estimate of GDD calculated from year to year at any given location throughout the US, as a percentage.

A closer look at the variation in GDD for two of the more renowned growing areas in the US reveals clear year to year variation in an average estimate of GDD. Napa and Sonoma valley of California are regions renown for quality wine production, both of which display annual variation in estimates of GDD. Figure 31 shows the inter-annual variation in GDD for Napa and Sonoma valley from 1991-2012. There is clear variation

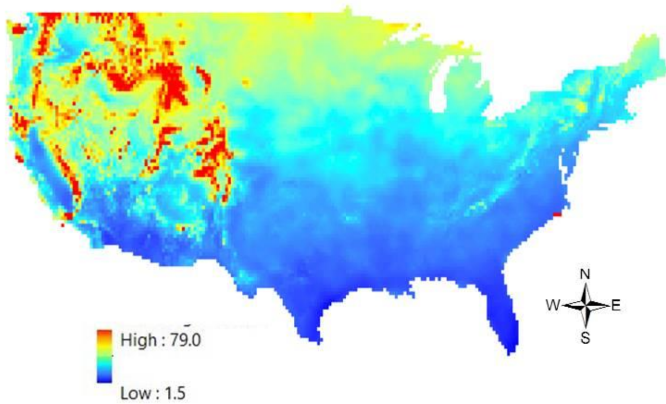


Figure 30. A depiction of the average percentage change in annual year to year variation of GDD across the US

from year to year in estimates of GDD thus it is important to understand the implications of this variation in the context of GDD.

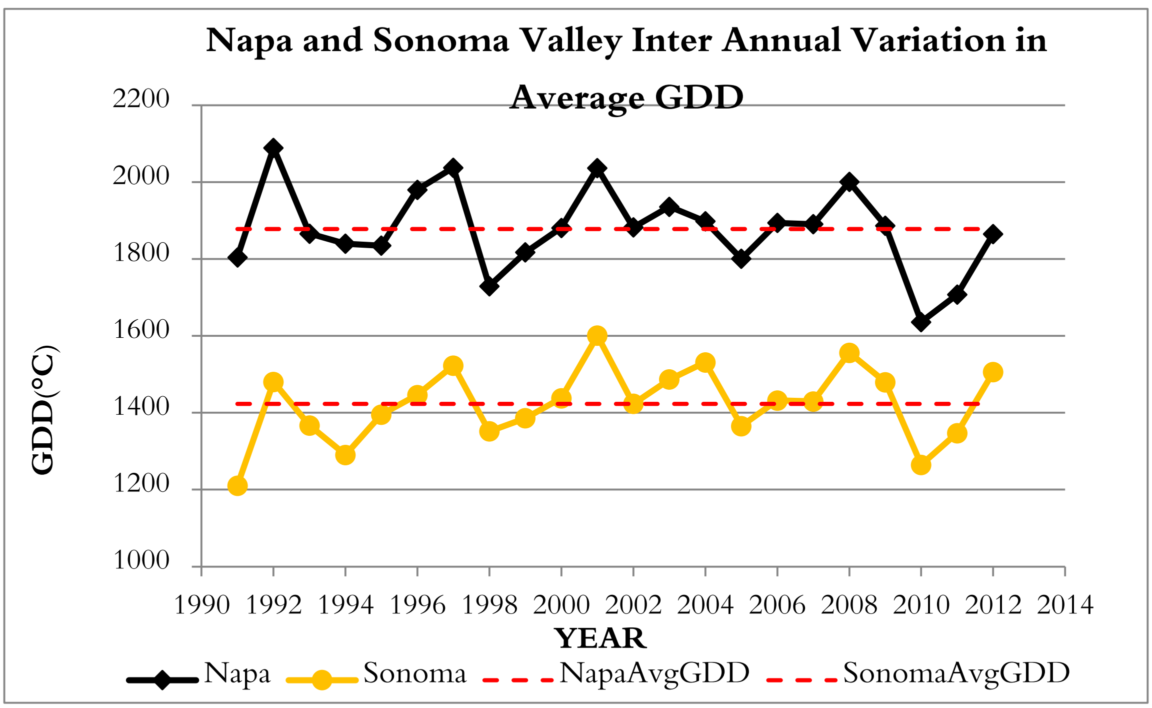


Figure 31. Results of the inter annual variation in estimates of GDD for Napa and Sonoma

Interpolated versus Station data

We detected differences in an estimate of GDD at a particular location depending on the source of the climate data. We estimated GDD with interpolated climate data (Daymet) and with station data from the nearest weather station for the same location. Figure 32 shows the general relationship between GDD estimates calculated from interpolated versus station data. Approximately 82% (R2 = 0.8212) of the variation in the data can be explained by the model.

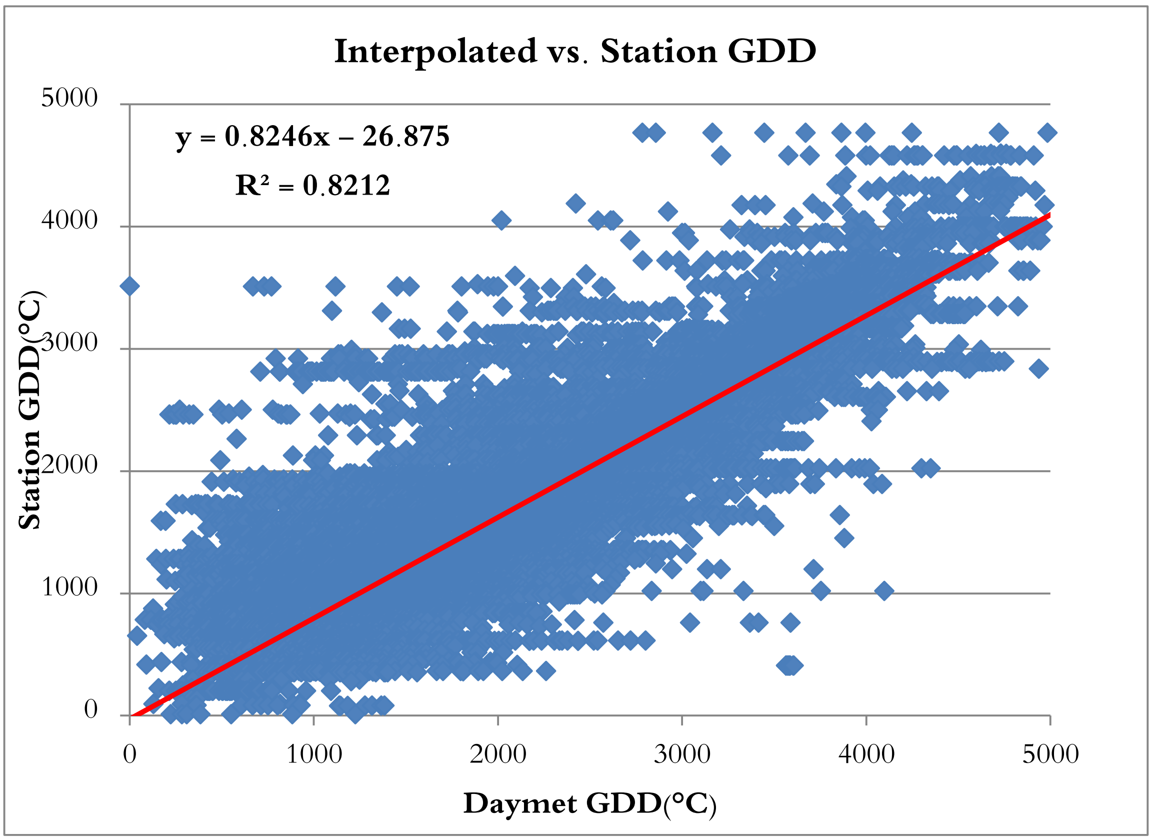


Figure 32A scatter plot of the general relationship between interpolated GDD and station GDD

Alternatively, comparing estimates at specific weather stations to interpolated GDD estimates at the same location (latitude and longitude) yields a similar linear relationship. Figure 33 illustrates the relationship between estimates at specific weather stations and the interpolated GDD estimates with only 78.69% of the variation in estimates explained by the data.

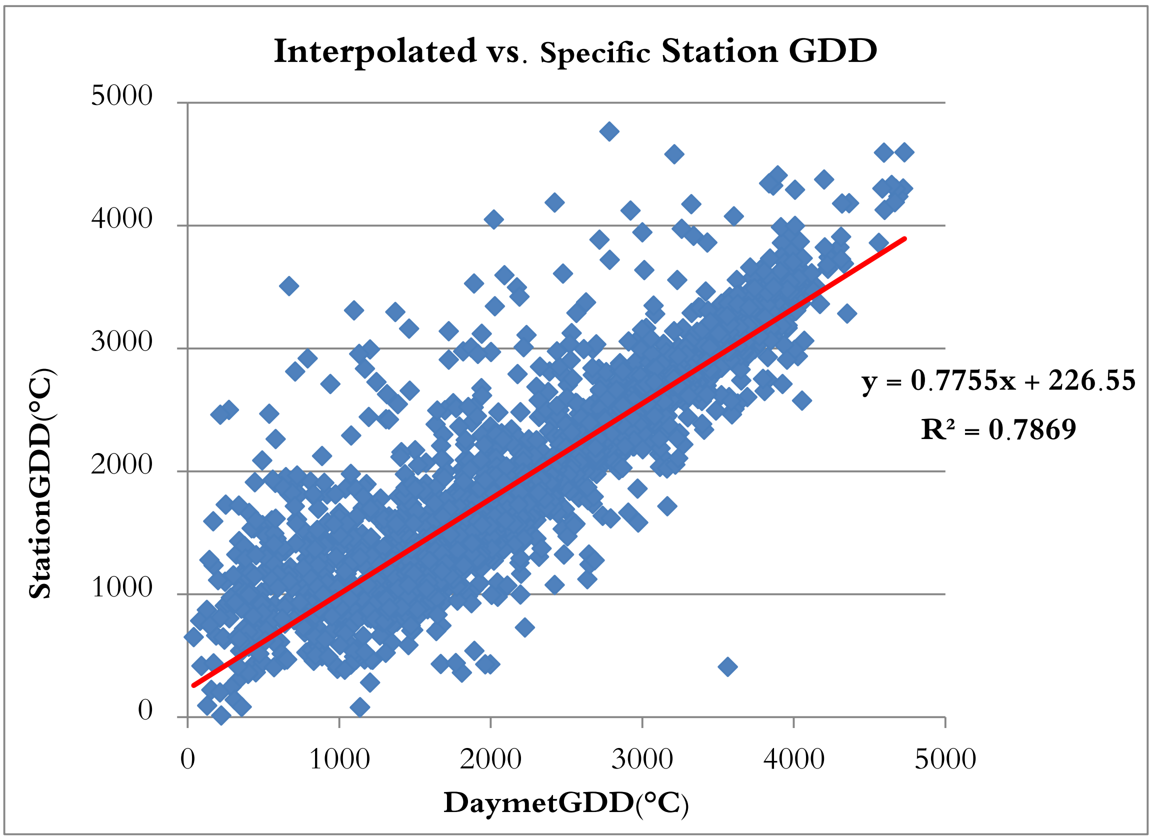


Figure 33. A scatter plot of the general relationship between interpolated GDD and specific station GDD

Spatial Differences in GDD

We deduced a model using a multiple regression analysis that explains approximately 88% (R^2 = 0.8887) of the variation in GDD across the US. Per our model results elevation, latitude, and longitude have the greatest influence on GDD estimates. The equation for estimating GDD is as below:

$$\text{GDD} = 6201.837 + (-0.663 * \text{Elevation}) + (-126.338 * \text{Latitude}) + (-16.596 * \text{Longitude}) \quad \text{.... (4)}$$

Figure 34 shows the relationship between estimates calculated from the derived model of GDD and actual estimates of GDD calculated using interpolated daily weather data.

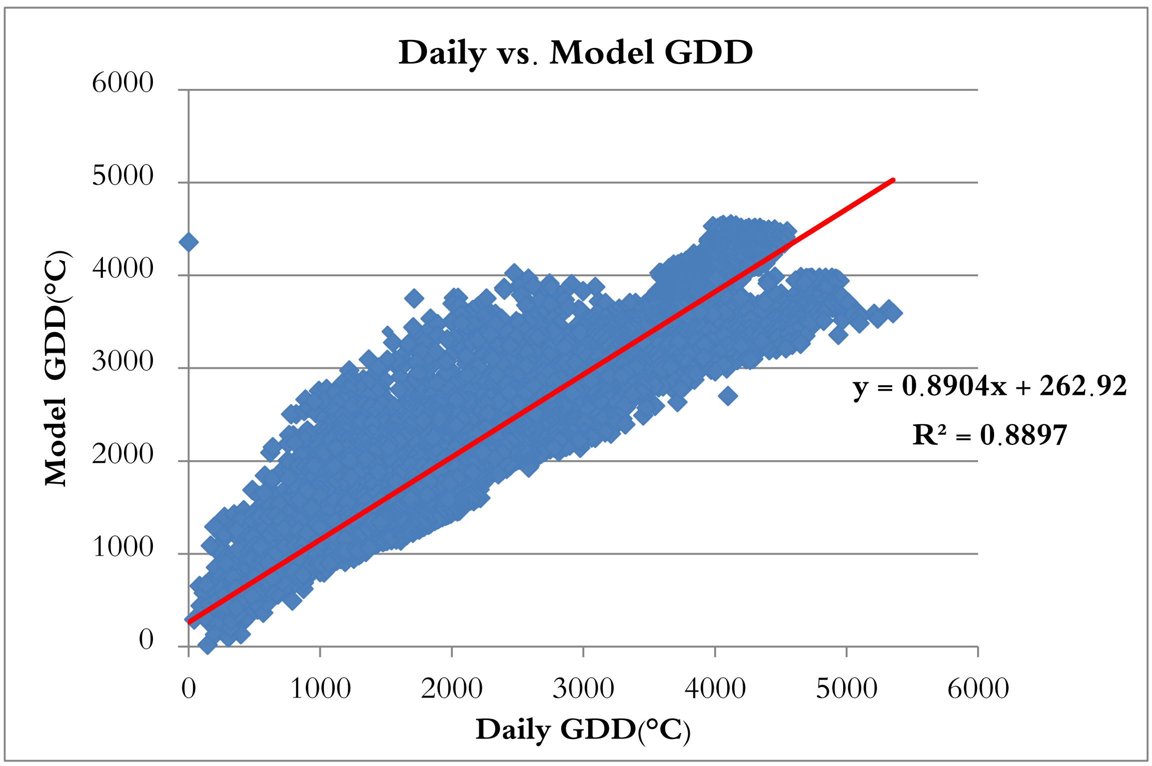


Figure 34. A scatter plot of the relationship between modeled GDD estimates and interpolated GDD

GDD versus GSAT

Growing season average temperatures (GSAT) typically define the climate-maturity ripening potential for premium quality wine varieties grown in cool, intermediate, warm, and hot climates (Jones, 2006). GSAT however does not account for fluctuations in weather patterns while GDD on the other hand maintains the patterns in weather. In other words areas with the same GSAT may have completely different estimates of GDD as illustrated in table 1.

Table 1. A comparison of GSAT to GDD using the different scenarios where GSAT is identical

Scenario	Day1(°C)	Day2(°C)	GSAT(°C)	CummDD(°C)
1	0	10	10	0
2	5	15	10	5
3	20	0	10	10

Discussion

We have illustrated differences in estimates of GDD due to the temporal resolution of the available weather data. The resolution of the data being used is not often reported along with an estimate of GDD. A graphical representation of GDD estimated at different temporal resolutions across the U.S. shows differences in accumulated GDD over time. Figure 35, 36, and 37 illustrate these differences in estimates of GDD over time at different temporal resolutions. The threshold or base temperature is indicated by a red line while temperature is indicated in blue. The area above the threshold and below the temperature curve indicates the accumulated degree day accumulation. Depending on the resolution of the data used in the calculations, these estimates of accumulated GDD vary. These differences or variations can be visualized spatially across the U.S. when we compare the estimates of GDD for the same location at different resolutions by creating a percentage difference map of the U.S. as shown in figure 28 of the results section. By examining the GDD for a given location at different temporal resolutions and calculating the percent change in the estimate across locations in the US, we found that there is up to 10% difference in the GDD calculated hourly versus daily versus monthly. In more practical terms, a GDD estimate of 2000 ºC may vary by as much as 200 ºC depending on the temporal resolution of the data used in the calculation.

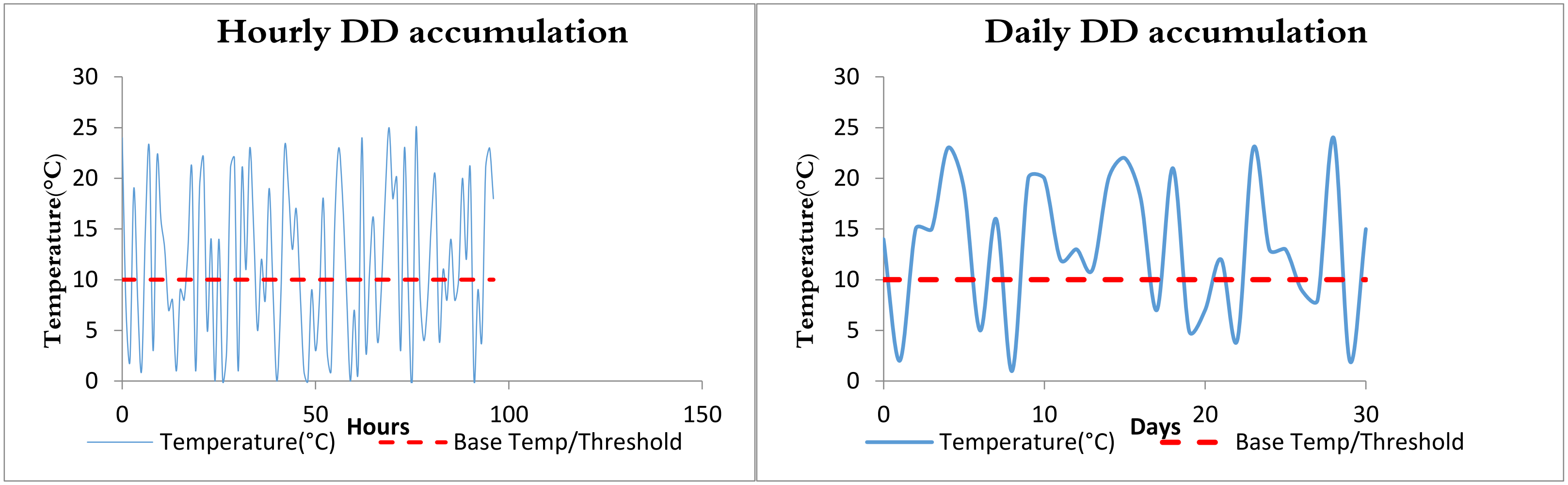


Figure 35. A graphical illustration of variation in daily and monthly accumulations of GDD over time

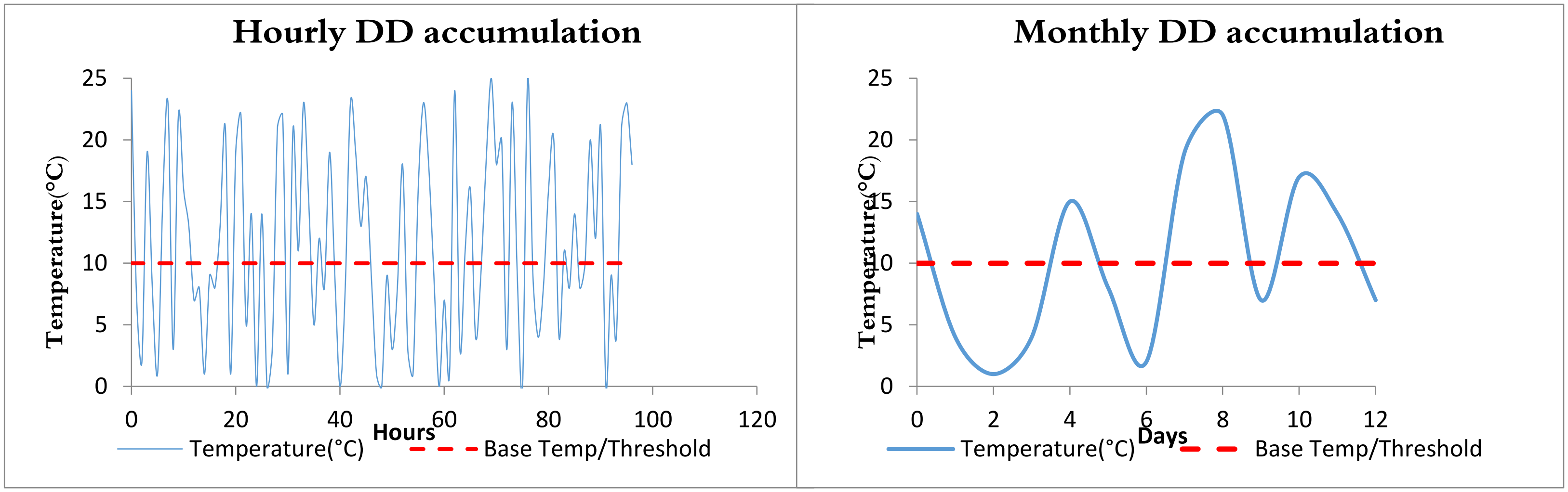


Figure 36. A graphical illustration of the variation in hourly and monthly accumulations of GDD over time

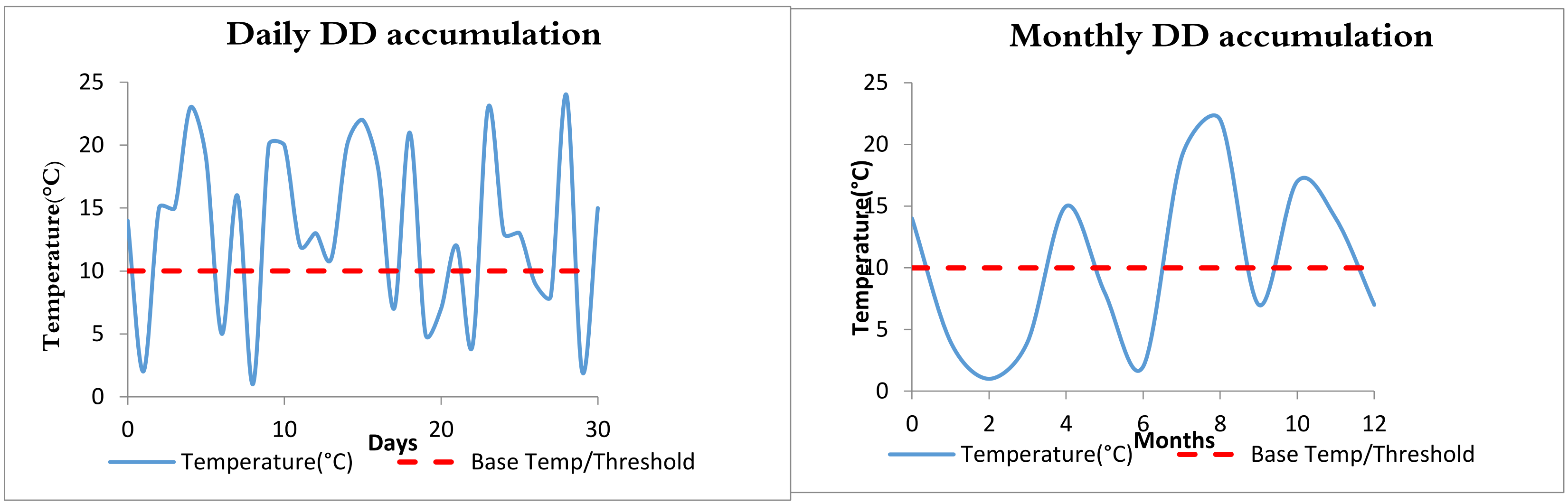


Figure 37. A graphical illustration of variation in daily and monthly accumulations of GDD over time

The implications of this variation are most evident when using GDD groupings such as the widely used Winkler scale developed for California. A 10% difference in an estimate of GDD due to resolution can result in the difference in the choice of a variety for a particular location. Amerine and Winkler developed this widely used scale based on the conditions for California however it is frequently applied as a measure of grape variety suitability throughout the U.S. and beyond. The original Winkler scale divided California into zones or climatic regions, based on a heat summation above 50°F. As such these zones were associated with specific varieties judged to be most suitable to a particular zone based upon heat summation. Figure 38 illustrates the original Winkler zones as initially described by Amerine and Winkler along with our updated depiction of the Winkler zones based upon average GDD estimates for the period of 1991-2012. California was subsequently divided in to 17 grape pricing districts. Based on Winkler's classification, varieties grown within these districts were classified into Winkler zones. As such varieties determined to be more suitable for one location versus another were classified accordingly. Table 2 illustrates the original Winkler classification based upon heat summation in both degrees Celsius and Fahrenheit. Climatic zones were designated for California based upon heat summation ranges established Amerine and Winkler.

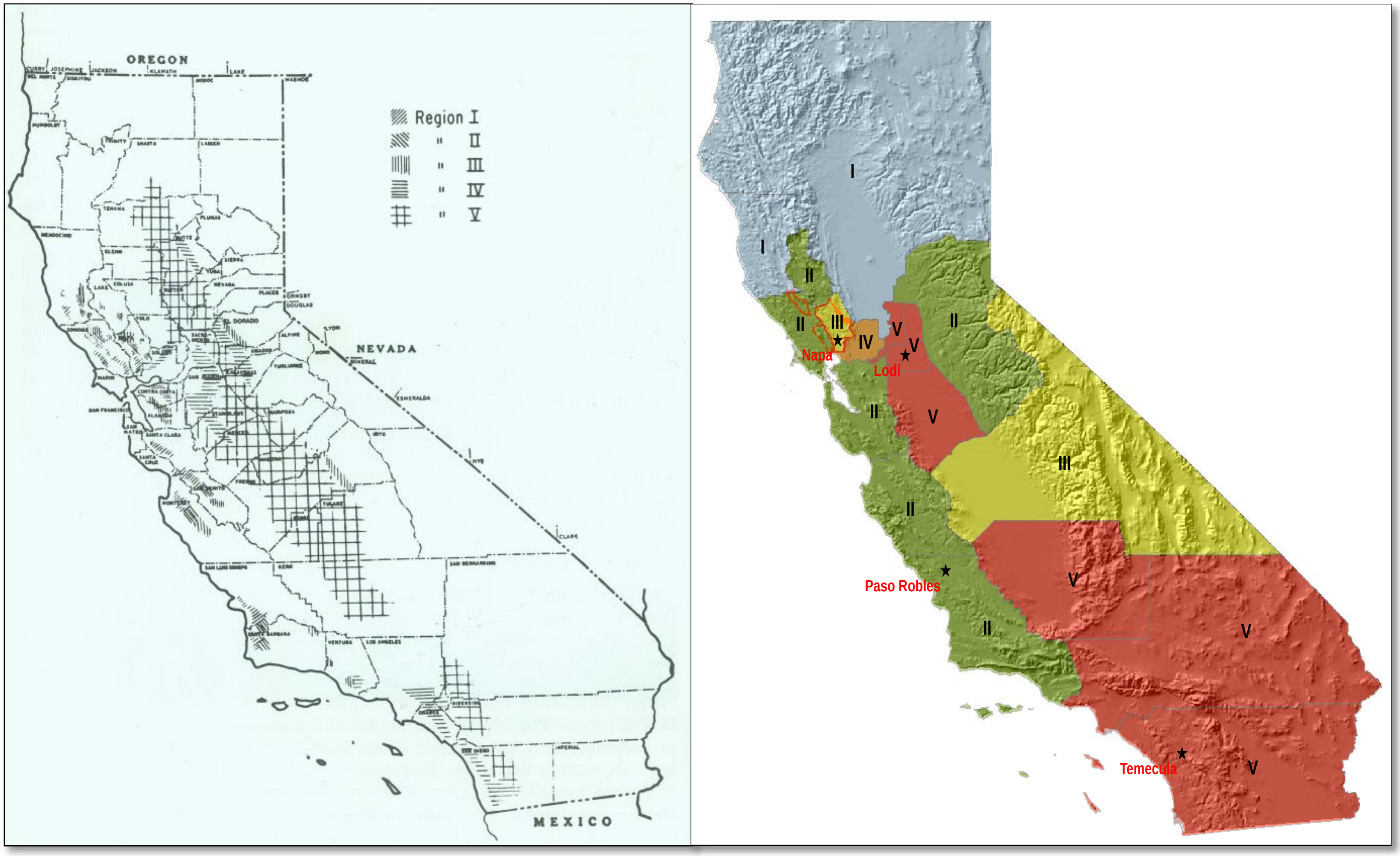


Figure 38. A depiction of the original Winkler map of California alongside our updated version based upon average GDD estimates for 1991-2012

Table 2. Winkler scale for California by Amerine & Winkler based on heat summation (GDD)

Climatic Zone	GDD Range
Region I	<1390°C (2500°F)
Region II	<1390-1670°C (2501-3000°F)
Region III	<1671-1940°C (3001-3500°F)
Region IV	<1941-2220°C (3501-4000°F)
Region V	<2200°C (>4000°F)

Based upon the heat summation ranges illustrated in table 2, wine grape varieties from the 17 grape pricing districts of California were classified into Winkler zones. Differences in estimates due to resolution of the data being used will affect the decision of a prospective grower interested in growing a particular variety of grape. It is therefore important to understand how these differences in an estimate may affect the choice of variety or the location of a potential vineyard. Figure 39 shows the different grape varieties grown in the California crush districts according to the 2013 grape crush report. Using Winkler's scale we have classified each variety in the appropriate Winkler zone by district. These results will vary significantly based upon the resolution of the data used ultimately affecting the choice of a location (district) or the choice of the variety grown.

1		2	3			4		
I	III	IV	I	II	III	II	III	IV
Cabernet Sauvignon	Aligote	Aleatico*	Cabernet Sauvignon	Aligote	Aligote	Aligote	Aligote	Aleatico*
Chardonnay	Barbera	Barbera	Chardonnay	Cabernet Sauvignon	Barbera	Cabernet Sauvignon	Barbera	Barbera
Gamay	Cabernet Sauvignon*	Cinsault*	Gamay	Chardonnay*	Cabernet Sauvignon*	Chardonnay*	Cabernet Sauvignon*	Cinsault*
Pinot noir	Carignane*	Carignane	Pinot noir	Folle blanche*	Carignane*	Folle blanche*	Carignane*	Carignane
Gewurztraminer*	French Colombard*	French Colombard	Gewurztraminer*	Mondeuse*	French Colombard*	Mondeuse*	French Colombard*	French Colombard
Refosco*	Gros Manzenc*	Grignolino*	Refosco*	Petite Sirah	Gros Manzenc*	Petite Sirah	Gros Manzenc*	Grignolino*
Sauvignon blanc	Muscat Canelli	Gros Manseng*	Sauvignon blanc	Pinot blanc	Muscat Canelli	Pinot blanc	Muscat Canelli	Gros Manseng*
White Riesling	Peverella*	Inzolia*	White Riesling	Red Veltliner*	Peverella*	Red Veltliner*	Peverella*	Inzolia*
	Refosco	Malvasia bianca		Refosco	Refosco	Refosco	Refosco	Malvasia bianca
	Sangiovese*	Mission*		Sauvignon blanc	Sangiovese*	Sauvignon blanc	Sangiovese*	Mission*
	Sauvignon blanc	Muscat Blanc		Semillion	Sauvignon blanc	Semillion	Sauvignon blanc	Muscat Blanc
	Semillion	Orange Muscat		Sylvaner	Semillion	Sylvaner	Semillion	Orange Muscat
	Trousseau	Palomino		Tannat*	Trousseau	Tannat*	Trousseau	Palomino
		Peverella*		White Riesling*		White Riesling*		Peverella*
		Refosco						Refosco
		Tinta Cao*						Tinta Cao*
		Tinta Madeira						Tinta Madeira
		Trousseau						Trousseau
		Tempranillo*						Tempranillo*

Figure 39. California Grape Variety Suitability Based on Winkler Zone Classification for the 17 crush districts

5	6			7		8		
IV	I	II	III	I	II	I	II	IV
Aleatico*	Cabernet Sauvignon	Aligote	Aligote	Cabernet Sauvignon	Aligote	Cabernet Sauvignon	Aligote	Aleatico*
Barbera	Chardonnay	Cabernet Sauvignon	Barbera	Chardonnay	Cabernet Sauvignon	Chardonnay	Cabernet Sauvignon	Barbera
Cinsault*	Gamay	Chardonnay*	Cabernet Sauvignon*	Gamay	Chardonnay*	Gamay	Chardonnay*	Cinsault*
Carignane	Pinot noir	Folle blanche*	Carignane*	Pinot noir	Folle blanche*	Pinot noir	Folle blanche*	Carignane
French Colombard	Gewurztraminer*	Mondeuse*	French Colombard*	Gewurztraminer*	Mondeuse*	Gewurztraminer*	Mondeuse*	French Colombard
Grignolino*	Refosco*	Petite Sirah	Gros Manzenc*	Refosco*	Petite Sirah	Refosco*	Petite Sirah	Grignolino*
Gros Manseng*	Sauvignon blanc	Pinot blanc	Muscat Canelli	Sauvignon blanc	Pinot blanc	Sauvignon blanc	Pinot blanc	Gros Manseng*
Inzolia*	White Riesling	Red Veltliner*	Peverella*	White Riesling	Red Veltliner*	White Riesling	Red Veltliner*	Inzolia*
Malvasia bianca		Refosco	Refosco		Refosco		Refosco	Malvasia bianca
Mission*		Sauvignon blanc	Sangiovese*		Sauvignon blanc		Sauvignon blanc	Mission*
Muscat Blanc		Semillion	Sauvignon blanc		Semillion		Semillion	Muscat Blanc
Orange Muscat		Sylvaner	Semillion		Sylvaner		Sylvaner	Orange Muscat
Palomino		Tannat*	Trousseau		Tannat*		Tannat*	Palomino
Peverella*		White Riesling*			White Riesling*		White Riesling*	Peverella*
Refosco								Refosco
Tinta Cao*								Tinta Cao*
Tinta Madeira								Tinta Madeira
Trousseau								Trousseau
Tempranillo*								Tempranillo*

Figure 39. Continued

9		10		11	12	
IV	V	III	IV	IV	IV	V
Aleatico*	Aleatico*	Aligote	Aleatico*	Aleatico*	Aleatico*	Aleatico*
Barbera	Cinsault*	Barbera	Barbera	Barbera	Barbera	Cinsault*
Cinsault*	Grenache	Cabernet Sauvignon	Cinsault*	Cinsault*	Cinsault*	Grenache
Carignane	Inzolia*	Carignane*	Carignane	Carignane	Carignane	Inzolia*
French Colombard	Malvasia bianca	French Colombard*	French Colombard	French Colombard	French Colombard	Malvasia bianca
Grignolino*	Mission	Gros Manzenc*	Grignolino*	Grignolino*	Grignolino*	Mission
Gros Manseng*	Orange Muscat	Muscat Canelli	Gros Manseng*	Gros Manseng*	Gros Manseng*	Orange Muscat
Inzolia*	Palomino	Peverella*	Inzolia*	Inzolia*	Inzolia*	Palomino
Malvasia bianca	Saint Macaire*	Refosco	Malvasia bianca	Malvasia bianca	Malvasia bianca	Saint Macaire*
Mission*	Salvador	Sangiovese*	Mission*	Mission*	Mission*	Salvador
Muscat Blanc	Tinta Cao*	Sauvignon blanc	Muscat Blanc	Muscat Blanc	Muscat Blanc	Tinta Cao*
Orange Muscat	Tinta Madeira	Semillion	Orange Muscat	Orange Muscat	Orange Muscat	Tinta Madeira
Palomino	Trousseau*	Trousseau	Palomino	Palomino	Palomino	Trousseau*
Peverella*			Peverella*	Peverella*	Peverella*	
Refosco			Refosco	Refosco	Refosco	
Tinta Cao*			Tinta Cao*	Tinta Cao*	Tinta Cao*	
Tinta Madeira			Tinta Madeira	Tinta Madeira	Tinta Madeira	
Trousseau			Trousseau	Trousseau	Trousseau	
Tempranillo*			Tempranillo*	Tempranillo*	Tempranillo*	

Figure 39. Continued

13		14	15		16		17	
IV	V	V	IV	V	IV	V	IV	V
Aleatico*	Aleatico*	Aleatico*	Aleatico*	Aleatico*	Aleatico*	Aleatico*	Aleatico*	Aleatico*
Barbera	Cinsault*	Cinsault*	Barbera	Cinsault*	Barbera	Cinsault*	Barbera	Cinsault*
Cinsault*	Grenache	Grenache	Cinsault*	Grenache	Cinsault*	Grenache	Cinsault*	Grenache
Carignane	Inzolia*	Inzolia*	Carignane	Inzolia*	Carignane	Inzolia*	Carignane	Inzolia*
French Colombard	Malvasia bianca	Malvasia bianca	French Colombard	Malvasia bianca	French Colombard	Malvasia bianca	French Colombard	Malvasia bianca
Grignolino*	Mission	Mission	Grignolino*	Mission	Grignolino*	Mission	Grignolino*	Mission
Gros Manseng*	Orange Muscat	Orange Muscat	Gros Manseng*	Orange Muscat	Gros Manseng*	Orange Muscat	Gros Manseng*	Orange Muscat
Inzolia*	Palomino	Palomino	Inzolia*	Palomino	Inzolia*	Palomino	Inzolia*	Palomino
Malvasia bianca	Saint Macaire*	Saint Macaire*	Malvasia bianca	Saint Macaire*	Malvasia bianca	Saint Macaire*	Malvasia bianca	Saint Macaire*
Mission*	Salvador	Salvador	Mission*	Salvador	Mission*	Salvador	Mission*	Salvador
Muscat Blanc	Tinta Cao*	Tinta Cao*	Muscat Blanc	Tinta Cao*	Muscat Blanc	Tinta Cao*	Muscat Blanc	Tinta Cao*
Orange Muscat	Tinta Madeira	Tinta Madeira	Orange Muscat	Tinta Madeira	Orange Muscat	Tinta Madeira	Orange Muscat	Tinta Madeira
Palomino	Trousseau*	Trousseau*	Palomino	Trousseau*	Palomino	Trousseau*	Palomino	Trousseau*
Peverella*			Peverella*		Peverella*		Peverella*	
Refosco			Refosco		Refosco		Refosco	
Tinta Cao*			Tinta Cao*		Tinta Cao*		Tinta Cao*	
Tinta Madeira			Tinta Madeira		Tinta Madeira		Tinta Madeira	
Trousseau			Trousseau		Trousseau		Trousseau	
Tempranillo*			Tempranillo*		Tempranillo*		Tempranillo*	

Figure 39. Continued

The GDD at a particular location varies from year to year. Areas of greatest variation tend to be areas most influenced by the local topography such as changes in elevation. Figure 40 illustrates changes in elevation across the U.S. and the corresponding percentage change in average GDD as ratio of the standard deviation to the mean of GDD. This ratio, termed the coefficient of variation (CV) estimates how much GDD varies from year to year compared to the average GDD. Higher CV implies greater dispersion or variation in GDD from the average. From a viticulture perspective, inter annual variation in GDD can influence the choice of a particular variety for a given location as some years can be ±79% different in GDD from the previous or following year. Hence simply based on the Winkler scale, variety choice or location choice may vary considerably from year to year.

GDD was calculated using data from the nearest weather station and data from the interpolated Daymet data set. Differences in estimates for the same location were observed using interpolated versus station data. At weather station locations, Daymet tends to give higher estimates of GDD compared to estimates from weather stations. Thus an estimate of GDD is also dependent on whether it was derived from station data or from interpolated data. In most cases vineyards are not located at the same geographic location of a weather station hence the need for interpolated weather data. We assessed the relationship between distance and differences in estimates of GDD from interpolated versus station data. This assessment led to the conclusion that distance has no predictable influence on estimates of GDD between an interpolated location and the

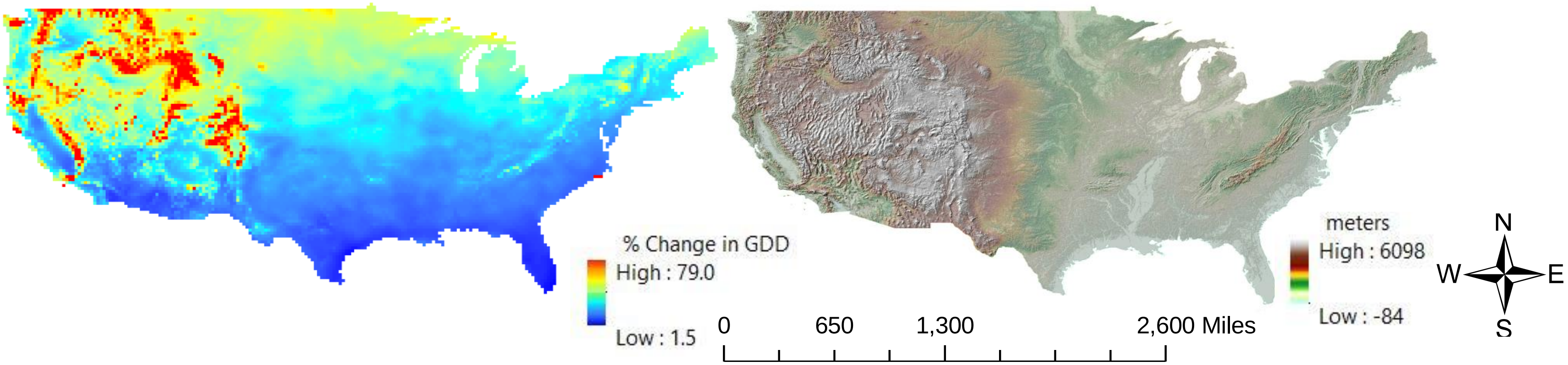


Figure 40. A comparison of the relationship between inter- annual variation in GDD and changes in elevation

location of the weather station. This is illustrated in figure 41 where there is no clear relationship between differences in GDD and differences in distance. Spatially interpolated data (Daymet) is therefore very valuable for anyone interested in calculating the GDD at a particular location.

We determined the spatial differences in GDD to be driven by elevation, latitude and longitude. At higher elevations, there is cooling with increasing height as the ground is the earth's heat source. With less atmospheric pressure at higher elevations, the result is lower temperatures. Latitudinal change on the other hand correlates with changes in the angle of the sun. As we approach the poles, there's less solar radiation which correlates with temperature. As such we would expect GDD to correlate directly with changes in solar radiation.

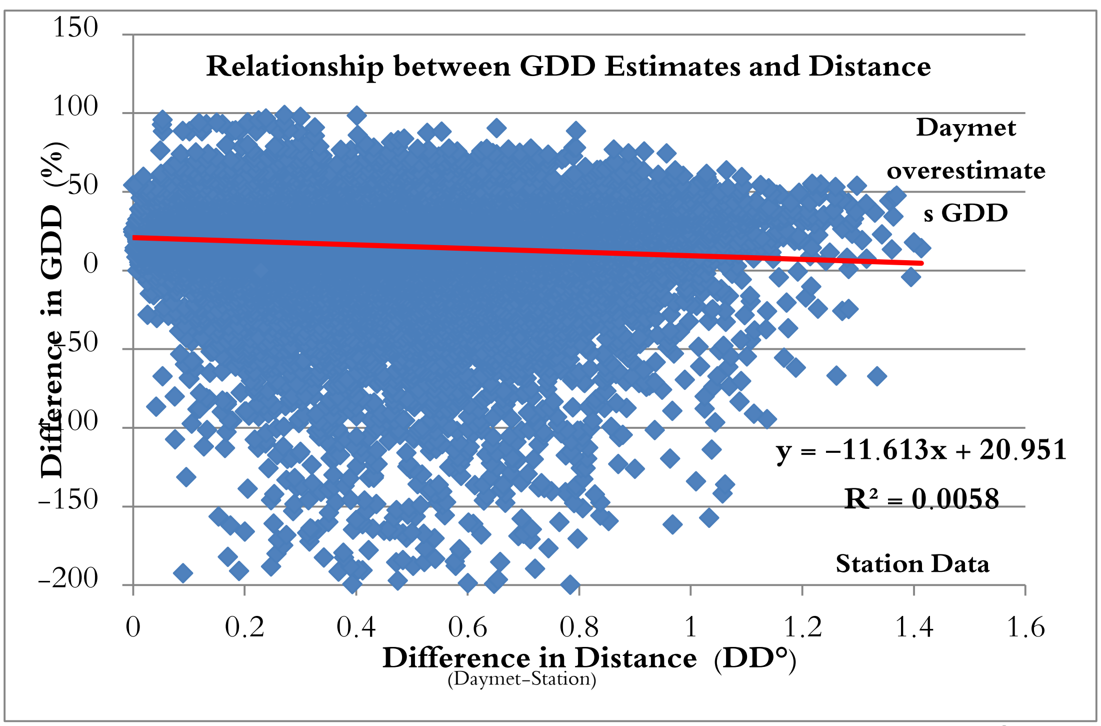


Figure 41. No predictable relationship between GDD estimates and distance of weather station to interpolated location

In assessing the relationship between GDD and solar radiation, we find no direct correlation as shown in Figure 42.

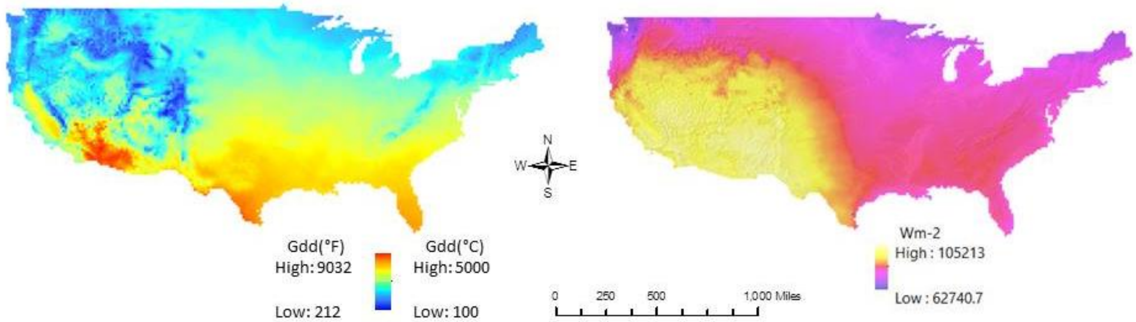


Figure 42. The relationship between GDD and solar radiation shows no direction correlation

A graphical illustration of the relationship between estimates of GDD and solar radiation confirms our conclusion from figure 42 as there is no clear relationship between GDD and solar radiation. This is illustrated in figure 43 with an R^2 value of 12%.

The goal of this research remains to ultimately understand how much accuracy should be expected when applying an estimate of GDD in viticulture. In other words, how accurate is a single value of GDD as an indicator of wine grape production success. The answer may depend upon the use of scales such as the often used Winkler scale. Table 3 shows the Winkler scale developed for California describing the GDD groupings and the average temperature range for each region. Depending upon calculation method, GDD estimates for a particular location can vary by as much as +-10%.

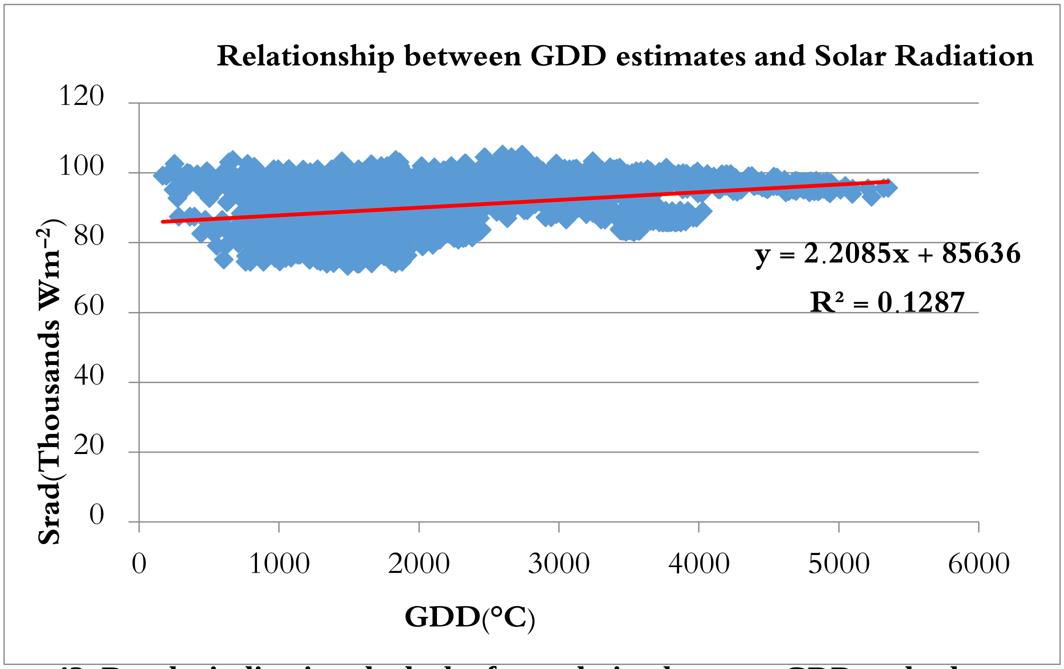


Figure 43. Results indicating the lack of correlation between GDD and solar radiation across the US

Table 3. GDD and GSAT by region developed for California (Amerine and Winkler 1944, Halliday 1993)

Climatic Zone	GDD Range	Average Temperature Range
Region I	<1390°C (2500°F)	16.5°C (61.7°F)
Region II	<1390-1670°C (2501-3000°F)	16.5-17.8°C(61.7-64.04°F)
Region III	<1671-1940°C (3001-3500°F)	17.8-19.07°C(64.04-66.33°F)
Region IV	<1941-2220°C (3501-4000°F)	19.07-20.28°C(66.33-68.50°F)
Region V	<2200°C (>4000°F)	>20.28°C(68.50°F)

Conclusions

Data availability and technology drive the use of GDD as a concept in agriculture. Growing degree day calculations are sensitive to the temporal resolution of weather data. As such, the interpretation of GDD as a concept in viticulture is reliant on both an understanding of the method of calculation and the resolution of the data used for calculation. GDD is not just about temperature however it is simply an estimate that requires a start date, end date and a threshold or base temperature. An estimate of GDD is subject to errors in the temporal and spatial resolution of the data as well as the year to year variation. These errors do not imply that estimates of GDD are incorrect. However it is simply an expression of how literally estimates should be applied to variety selection and the suitability of a region. Elevation, latitude, and longitude account for approximately 88% of the variation in an estimate of GDD calculated at one location to the next. In other words we can conclusively estimate GDD to within 88% accuracy with simply knowledge of the elevation, latitude, and longitude of a location. An estimate of GDD for a particular location is therefore dependent upon an understanding both the practical (simplicity) and theoretical (applicability) implications. Though generally useful in outlining limits for general suitability, GDD estimates must be used with caution in the selection of appropriate varieties.

CHAPTER V

WINE GRAPE SUITABILITY: A CASE STUDY OF THE CALIFORNIA CRUSH DISTRICTS

In this chapter we utilize grape production data obtained from the California office of the USDA's National Agricultural Statistic Service (NASS) to model the relationship between environmental conditions and a measure of viticultural suitability. The NASS conducts hundreds of surveys every year and prepares reports covering virtually every aspect of U.S. agriculture. Charged with the responsibility of providing objective information, the NASS along with the USDA provides timely, accurate, and useful statistics about wine grapes as reported in the annual grape crush report. This report constitutes the most comprehensive and objective grape production data set describing wine grape production in the US. As such it provides broad scale temporal and spatial coverage of the most significant wine growing area in the US.

Introduction

Environmental conditions determine to a large extent which grape cultivars (variety) can be grown where. Grapevines are grown in distinct climate regimes with ideal conditions to produce quality grapes. Furthermore, GDD is historically and currently the most commonly used measure of climatic suitability (variety adaptation) for viticulture. The goal of this research was to investigate the relationship between variability in environmental conditions and a measure of suitability in viticulture. Moreover using price as a measure of the suitability, we specifically assess the relationship between GDD and price. For the purposes of this research, price refers to the amount of money paid for a ton of a particular variety of grapes as described in the grape crush report. Suitability or “success” in viticulture is a subjective concept but should be measurable and quantifiable. Measures of viticultural success usually refer to average yield or wine production per hectare (Jackson 2008). In the absence of accurate acreage planted yield cannot serve as a measure; thus surrogates such as production and price per ton have been used as suitable substitutes (Jones and Davis 2000). In the absence of yield we used a data driven approach to model environmental conditions of a location that may affect successful wine grape production. As such we investigate the relationship between GDD and measures of viticultural success using data from the California grape crush reports of 1991- 2012. This is a comprehensive database of grape production data that has been objectively collected since 1976. The information contained in these reports was supplied by grape processors to fulfill the reporting requirements of Section 55601.5 of the Food and Agricultural Code. The preliminary grape crush reports include all grape tonnage crushed during the 1991 -2012 seasons. It also includes purchased tonnage and pricing information for grapes with final prices for California.

California Wine Industry

California is the leading producer of wine grapes in the US, responsible for 90% of all US wine. As of 2013 there were over 4100 bonded wineries, 214.6 million cases sold, and an estimated retail value $23.1 billion in wine sales in the US. California is consequently the 4th leading global wine producer after France, Italy, and Spain with wine grapes grown in 134 federally approved AVAs. The annual impact of the wine industry in California has resulted in over 330,000 jobs and $61.5 billion in revenue. The wine grape industry in California accounts for a remarkable 57% share of the US market, implying that nearly 3 out of every 5 bottles of wine sold in the US comes from California (MFK Research, 2009). The rationales for utilizing wine grape production data and prices from California are obvious given the economic impact of the California wine industry.

California Crush Districts

Every processor who crushes grapes in California is required to report certain information to the California Department of Food and Agriculture (CDFA). As such 17 grape crush districts of California were mapped as a result of the Clare Berryhill Grape Crush Report Act of 1976. Figure 44 illustrates the 17 'Grape pricing districts' which are in turn aggregates of AVAs for California.

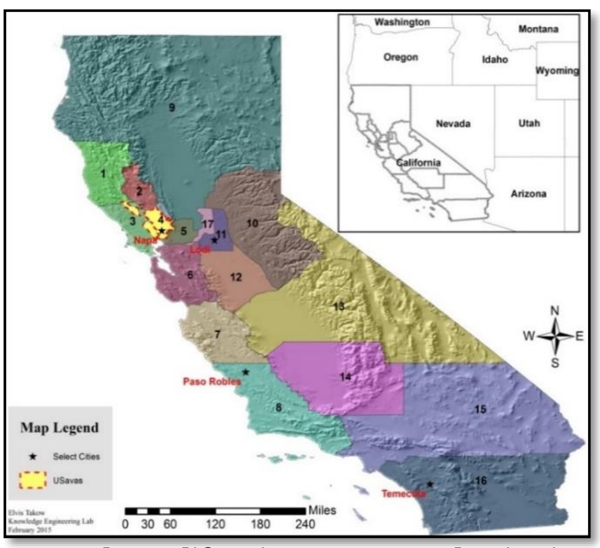


Figure 44. The California Grape Crush Districts as defined by Clare Berryhill Grape Crush Report Act of 1976

It is important to establish that the crush districts represent the general area within which grapes are grown and not specific wine growing regions. Some districts represent an aggregation of several wine growing regions.

In most agricultural systems, technological advances account for increases in yield and crop production over time. Viticulture is no exception and according to Gladstones (1992), after accounting for technology, climate is the main control on the yield of the grapevine. Climate is a prevalent factor in the success of all agricultural systems, as it influences the suitability of a crop within a given region. In viticulture, the influence of climate is largely responsible for controlling production and quality (Jones et al., 2005; White et al., 2006). Consequently, consumers have often used a product's price as a measure of the product's quality (Monroe 1973; Olson 1976). Wine grape quality is rewarded by the price point that is achieved by the finished product (Webb et al., 2008). Consumers and wineries would otherwise not pay high prices for wine grapes if they could not achieve a corresponding high price for the resulting wine. As such the price differential has been linked to wine and wine grape quality in some studies (Golan and Shalit, 1993; Oczkowski, 2001; Haeger and Storchmann, 2006; Ashenfelter, 2008). No one has scientifically or objectively tested whether GDD is actually a good index of variety suitability. The California crush data is comprehensive with broad scale temporal and spatial coverage and allows for such an analysis. We used price as a measure of suitability to scientifically assess climatic variation. The objectives were to 1) outline the climatic variables from the viticulture literature that may influence the production of quality wine grapes 2) evaluate the spatial temporal relationship between GDD and price as a measure of success, and 3) examine the spatial temporal relationship between environmental factors and price as a measure of viticultural success. We assume that wine grape price is a valid indicator of grape variety suitability to a geographic region. Higher prices are therefore paid for a given grape variety produced in a region where it is well suited. We also assume that environmental variability and specifically GDD is a valid indicator of grape variety suitability.

Data and Methods

Development of Crush District Data (Dependent Data)

This analysis used grape crush data from the California office of the USDA's NASS. Grape crush reports for the period of 1991-2012 were downloaded in excel format and aggregated into a single excel worksheet. The data contain details of the crushed tonnage, degrees brix, and weighted average prices reported by grape type, variety, and grape pricing districts. The 17 grape crush (pricing) districts refer to the area in which the grapes were grown as defined in the Administrative Code. State totals and averages for the preceding crop year dating back to 1976 are aggregated in tables for comparison. For the purposes of this analysis we only used data from 1991-2012 as data before 1991 was not available in excel format. We were specifically interested in 3 main tables of data namely tonnage, weighted average degrees brix, and weighted average price per ton for all grapes crushed and purchased for wine. The aforementioned tables contain data which represent different categories or kinds of our dependent data.

The reports contain 10 tables of data on all grape tonnage crushed for every season for each of the 17 districts for all varieties grown. The data are broken down by red and white grape varieties for each of the 17 districts. The table illustrated in figure 45 represents a sample for 2012 crop year for all white wine grape varieties grown in the 17 districts. A similar sample is provided for red wine grape varieties and this represents the raw data as reported by the grape processors. Upon downloading the aforementioned data as an excel spreadsheet, the data was restructured to represent the variables of interest for our research. This was achieved by reorganizing the raw data illustrated in figure 45 thereby only representing each year, district, variety, and the appropriate dependent variable. This reorganization was based on year, variety and district which form the basis of the subsequent steps in our analysis.

Figure 46 illustrates the adjusted format of the restructured data table with a dependent variable of interest for all years, varieties, and districts. Based on the desired data structured outlined in figure 46, we proceeded by importing each restructured data table into SQL Server 2014 Management Studio® database where individual data tables are joined based on the common fields of ‘year’, ‘district’ and ‘variety’. Finally we sorted our data by variety based on the top 8 varieties grown in California by production. The overall process of downloading raw data from NASS and organizing it for analysis based upon the aforementioned fields is illustrated in figure 47. This process follows the principles and methods for managing big data described in chapter 2 of this dissertation. This was a four step process which involved the following: (1) Data acquisition and collection, (2) Data integration and aggregation, (3) Data analysis and modelling and (4)

Data interpretation. The later stages of analysis, modelling, and interpretation will be described in greater detail at the conclusion of the methods section of this analysis.

TABLE 2: TONS OF GRAPES CRUSHED BY CALIFORNIA PROCESSORS
FROM THE 2012 CROP BY TYPE, VARIETY, AND REPORTING DISTRICT WHERE GROWN, WITH COMPARISONS

Type and Variety	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	Tons										Tons						
WINE GRAPES (WHITE):																	
Albarino	2.6	0.8	24.1	64.0	8.1	0	439.7	355.0	11.1	19.2	40.3	0	38.1	0	0	11.3	364.5
Arneis	37.1	0.0	107.2	0.1	0.0	0.0	23.5	26.8	0.5	0.0	0	3.4	0.0	0.0	0.0	17.8	0.0
Burger *	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2,884.2	2,381.8	26,017.4	1,821.1	0.0	0.0	2,159.0
Chardonnay *	24,447.9	1,725.9	81,603.2	31,944.5	6,109.4	7,943.8	103,330.5	46,341.2	22,963.7	767.3	155,503.7	66,621.0	96,164.3	40,972.4	51.8	234.7	49,089.0
Chenin Blanc	96.6	0.0	19.2	95.9	1,449.7	0.0	1,240.9	978.4	112.1	50.8	3,579.5	1,892.1	30,084.8	5,154.0	0.0	17.6	13,308.5
Cortese	2.9	0.0	0.0	0.0	0.0	0.0	0.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.1	0.0
Emerald Riesling	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	665.0	0.0	0.0	0.0
Fiano	0.0	0.0	4.3	0.4	2.2	8.9	0.0	1.3	3.0	2.8	2.0	0.0	0.0	0.0	0.0	0.0	5.5
Flora	1.5	0.0	0.0	28.6	0.0	0.0	0.0	0.0	0.0	2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Folle Blanche	0.0	0.0	18.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
French Colombard	67.6	0.0	116.6	0.8	0.0	2.2	0.0	0.0	152.8	0.0	2,773.6	10,758.4	262,012.7	35,661.6	0.0	0.0	65.2
Gewurztraminer	1,030.8	437.5	640.0	33.8	123.0	5.7	8,486.4	918.0	18.7	6.8	140.1	2,296.4	76.3	0.0	0.0	37.4	2,590.0
Gray Riesling *	0.0	0.0	41.1	5.2	0.6	0.0	0.0	0.0	0.0	0.0	32.6	0.0	0.0	0.0	0.0	0.0	0.0
Grenache Blanc	0.0	0.0	27.9	29.5	2.1	3.6	45.6	888.3	23.8	51.2	52.1	0.0	0.0	0.0	0.0	38.7	0.0
Gruner Veltliner	0.0	0.0	1.7	10.2	0.2	1.2	359.6	89.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	93.3
Kerner	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.3	0.0	0.0	0.0	0.0	0.0	0.0
Malvasia Bianca	0.0	0.0	33.4	9.2	0.0	34.8	739.2	113.0	1.1	3.3	1,189.3	3,932.3	1,063.5	0.0	0.0	8.5	0.0
Marsanne	20.8	75.5	57.0	54.7	0.0	15.5	81.7	272.4	2.9	54.0	151.5	0.0	2.8	0.0	0.0	1.7	0.0
Melon	0.0	0.0	19.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Moscato Gaillo *	0.6	0.0	5.4	0.0	0.0	0.0	0.0	0.0	1.1	4.0	65.5	0.0	0.0	0.0	0.0	6.9	0.0
Muscat Blanc *	269.9	340.6	101.6	404.7	60.1	9.8	829.1	1,081.4	9.5	60.1	2,349.8	1,997.5	10,339.3	18,229.5	4.8	190.0	371.7
Muscat Orange	32.1	0.0	0.0	13.5	16.8	27.7	98.1	162.0	0.0	66.5	608.1	752.8	635.2	0.0	0.4	5.5	42.1
Muscat of Alexandria	0.0	0.0	1.8	0.0	0.0	0.0	0.0	0.9	4,959.4	1.2	299.5	965.7	42,769.4	29,416.8	0.0	2.9	0.0
Palomino *	0.0	0.0	4.1	0.0	0.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0	1,202.6	0.0	13.0	20.4	0.0
Picpoul Blanc	0.0	0.3	4.7	1.0	0.0	0.0	9.9	38.7	1.2	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Pinot Blanc	113.5	11.6	368.8	59.5	0.0	6.8	881.4	299.3	0.0	2.5	0.0	0.0	0.0	54.6	0.0	0.0	0.0
Pinot Gris *	546.1	475.0	2,701.0	997.2	5,066.5	1,331.1	8,961.5	3,536.4	2,781.4	132.5	51,933.7	41,599.0	33,529.8	25,553.0	1.2	95.8	16,211.5
Ribolla Gialla *	0.0	0.0	2.7	23.1	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Roussanne	87.4	44.3	95.4	68.5	0.0	14.2	57.4	637.1	11.6	121.1	64.7	0.0	4.5	0.0	0.0	70.8	0.0
Sauvignon Blanc	3,251.5	8,592.1	17,162.4	14,805.5	842.4	1,481.3	8,542.4	6,252.4	1,028.3	669.5	22,516.2	5,353.5	3,611.8	3,555.1	32.0	169.7	15,407.0
Sauvignon Musque	0.5	149.5	363.9	348.2	20.9	15.6	49.0	0.1	0.2	0.0	4.2	0.0	0.5	0.0	0.0	0.0	0.0
Sauvignon Vert *	0.0	0.0	3.9	8.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Scheurebe	0.0	0.0	1.4	13.1	0.0	0.0	0.0	2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Semillon	55.4	189.6	470.9	885.9	0.9	104.9	50.1	38.4	122.7	60.1	1,924.5	0.0	1,329.2	0.0	2.0	3.6	1,276.0
St. Emilion *	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.0	0.4	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Sylvaner	0.0	0.0	13.6	0.0	0.0	0.0	0.0	13.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Symphony	1.7	0.0	0.0	0.0	143.5	0.0	0.0	0.0	44.3	58.0	2,547.7	14,134.0	5,499.0	2,474.9	0.0	0.0	1,298.6
Tocai Friulano	30.0	0.0	28.1	7.2	0.0	0.0	38.1	32.1	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Torrontes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3	2.2	0.0	31.7	0.0	0.0	0.0	0.0	0.0	0.0
Triplett Blanc	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1,046.0	21,935.0	0.0	0.0	0.0	0.0
Verdelho	0.0	0.0	0.0	24.9	58.7	15.2	5.3	46.3	4.3	60.1	1,246.9	0.0	3.7	0.0	0.0	11.7	185.2
Vermentino *	2.5	2.8	35.7	3.5	0.0	0.0	2.6	69.8	9.1	33.4	348.1	0.0	0.0	0.0	0.0	49.5	0.0
Vernaccia	1.0	0.0	48.5	0.0	0.0	0.0	0.0	4.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Viognier	476.1	204.5	843.5	349.2	248.1	281.4	442.4	2,408.4	826.8	422.0	11,456.7	3,837.1	753.0	0.0	11.7	173.4	1,425.5
White Riesling *	510.3	379.3	363.8	460.9	162.6	15.0	14,813.2	2,232.3	21.4	18.4	3,188.5	9,391.3	612.0	0.4	0.0	149.1	4,609.3
Other White 1/	0.5	4.6	42.2	10.9	0.8	10.5	9.7	15.8	237.6	9.5	2,256.1	0.0	776.8	1,654.6	12.0	4.0	3.1
Total White	31,086.9	12,633.9	105,376.3	50,762.0	14,316.6	11,340.2	149,537.3	66,869.6	33,351.7	2,683.2	267,198.1	166,962.3	538,461.7	165,213.0	128.9	1,327.1	108,505.0

Figure 45. An example of raw tonnage data of white wine grapes for the year 2012 by district and variety

YEAR	DISTRICT	VARIETY	Price
YEAR	DISTRICT	VARIETY	Tons
YEAR	DISTRICT	VARIETY	˚Brix

Figure 46. Restructured generic data table for variable of interest by year, district and variety

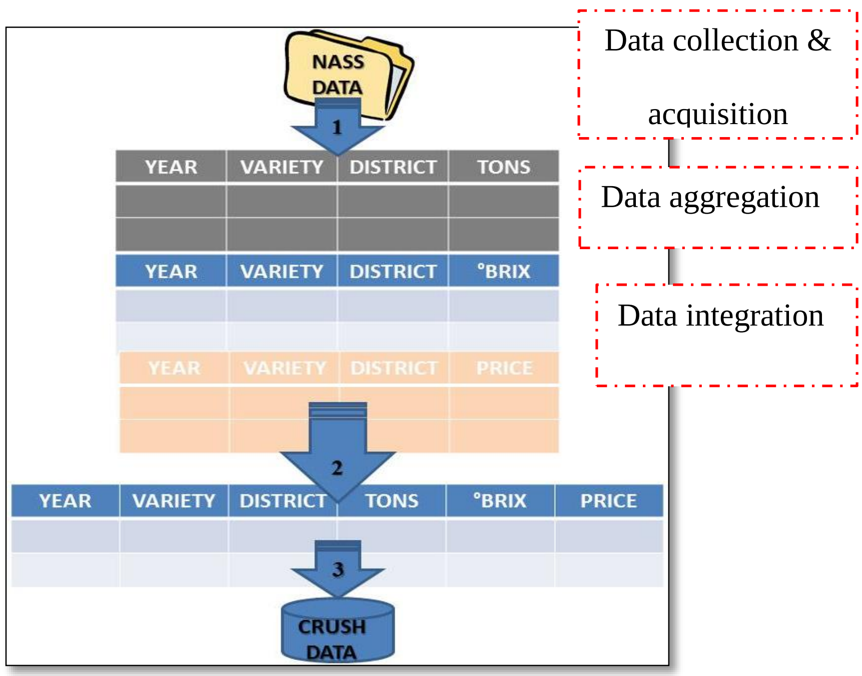


Figure 47. An illustration of the methodology for data acquisition, collection, integration and aggregation of California crush data

Development of Environmental Data (Independent Data)

Environmental data was obtained from Daymet as daily surface weather data for the entire US for the period of 1980-2012. This data were obtained as csv files of daily weather parameters over large regions at a spatial resolution of 0.125 decimal degrees (~8.5miles). Figure 48 illustrates a distribution grid of raw Daymet data which covers the entire US extent. The raw data are distributed from Daymet in 2 degree by 2 degree tile subsets as shown in figure 49.

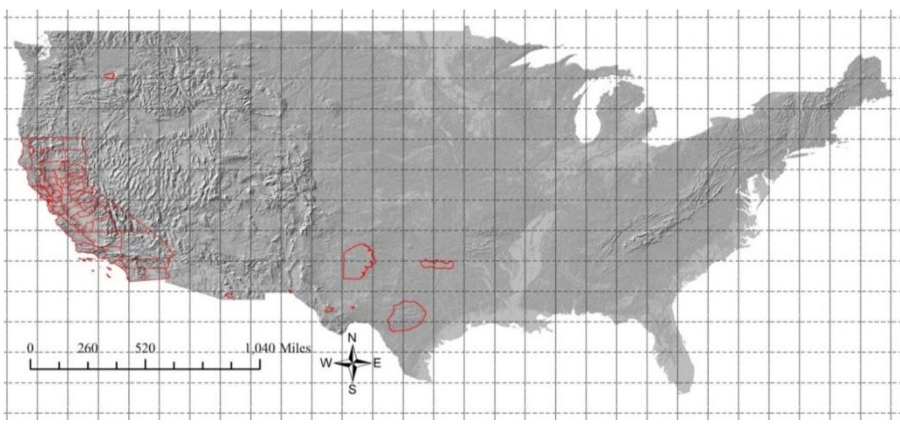


Figure 48. Distribution grid of Daymet data covering the full extent of the US at spatial resolution of 2x2 decimal degrees with illustration of the California crush districts and the Texas wine growing regions

The distribution depicted in figure 49 is at an interval of 0.125 by 0.125 degrees. We then utilized a script developed by Daymet to extract data for multiple latitude and longitudes at an interval of 0.125 degrees latitude and longitude.

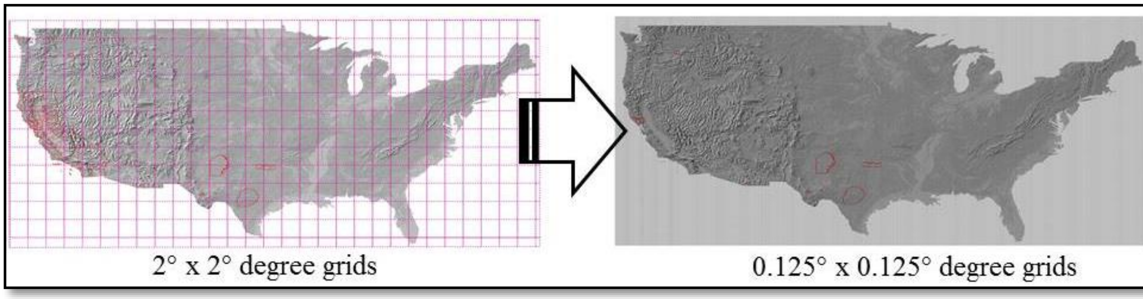


Figure 49. Daymet data resampled from 2x2 degree grids to 0.125x0.125 degree grids for analysis

The data was then stored into a file geodatabase as described in chapter 2 and was accessed using C# code outline in appendix A. The entry point of the C# code is a single latitude and longitude or multiple latitudes and longitudes for the area of interest. At this stage, the result is several thousand csv files of the entire US extent at 0.125 degree interval of latitude and longitude stored in a file geo-database. Daymet was developed to fulfill the need for continuous surfaces of daily weather data necessary for plant growth model inputs. As such, the Daymet output variables included minimum and maximum temperature, precipitation, water vapor pressure, shortwave radiation, and snow water equivalent. These output variables were used to derive additional variables which were used in our subsequent regression analysis to examine the relationship between measures of viticultural success and environmental variability over time and space. Figure 50 illustrates a complete overview of our methodology from data collection, aggregation, to integration of the California crush data and environmental data.

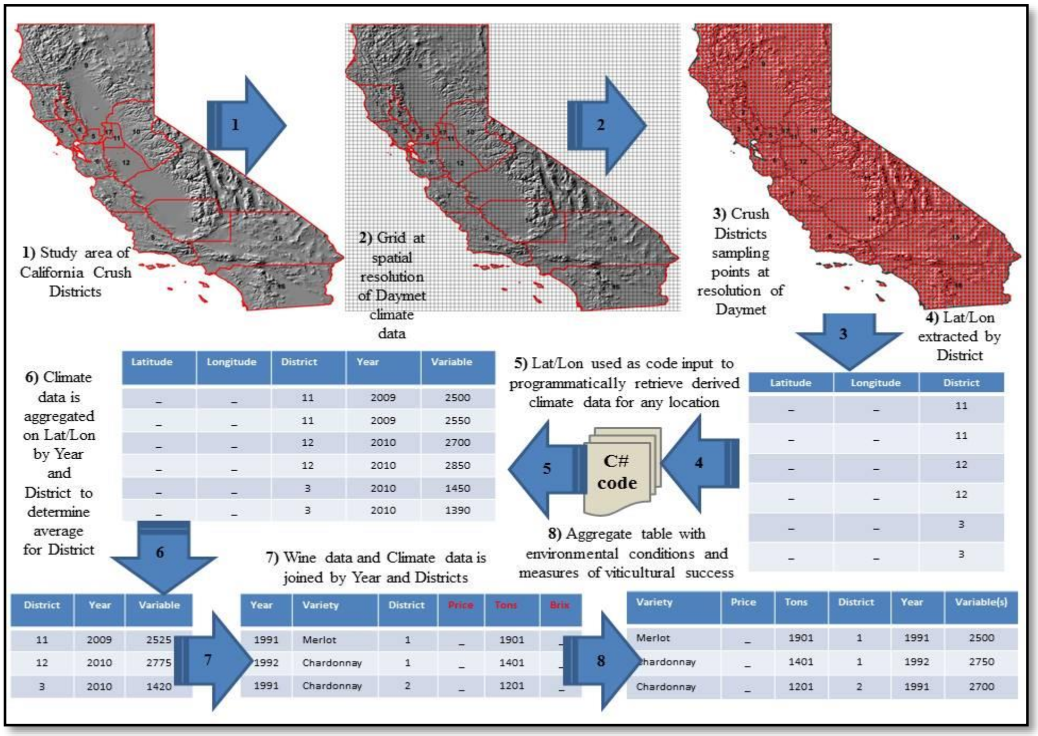


Figure 50. An overview of the data collection and analysis of the California crush data

Environmental Factors

A number of studies have been conducted over the years which focus on the premium global wine regions the world (Fanet 2004; Hancock 1999; Jones 2006; Van Leeuwen et. al, 2004; White 2003). In order to assess the relationship between measures of viticultural success and environmental variability over time and space, it is important to evaluate the influence of environmental factors on wine grape growing regions. Wine grapes are a climatically sensitive crop whereby quality production is achieved across a fairly narrow geographic range. While there is a general understanding that climate drives successful wine grape production, it is still unclear as to which aspects of climate contribute most to matching a region to suitable varieties. Globally, the average climatic conditions of a region determine which grape cultivars can be grown there, while wine production and quality are mostly influenced by site specific factors, husbandry decisions and short-term climate variability (Jones and Hellman, 2003). Environmental factors affecting grape growth, production, and wine quality include average temperatures, temperature extremes, solar radiation, heat accumulation, precipitation, humidity, and soil water balance characteristics (Jones et al 2012). We assembled a comprehensive list of climatic and edaphic factors that most completely capture the environmental conditions of a region. These factors are based on the prevailing viticultural literature (Jackson and Spurling 1988; Jackson and Schuster 1987; Jones 2006; Jones and Davis, 2000(a) and (b); Jones et al, 2012; Smart and Dry 1980; Tonietto and Carbonneau 2004; Van Leeuwen et. al, 2004; Winkler et al 1974) and consultation with experts in the field of viticulture. Plant growth and development is the result of both photosynthesis and respiration. Hence it is important to account for the night time accumulation of heat units as well. Temperature is the most influential environmental factor affecting the rate of respiration with increasing temperature causing a progressive increase in respiration rate up to a point where tissue damage occurs (Hellman, 2003; Mullins et al., 1992). As such we propose an index which accounts for both the day and night time accumulation. Net GDD is therefore the difference in heat units accumulated in the day during photosynthesis and at night while the grape vine undergoes respiration. The result is what we term Net GDD and is calculated as follows:

$$netGDD = \text{Daytime GDD} - \text{Nighttime GDD}$$

We propose this index as alternative to the traditional GDD estimate which does not account for the vines activities during respiration at night.

As such, our list of environmental factors is provided in table 4 but is limited by the availability of complete and comprehensive data coverage. This list does not exhaust the scope of environmental variables that may influence successful wine grape growth. Our list of environmental factors provides a preliminary selection of environmental variables that may influence the choice of a particular variety of grape for a specific location.

Regression Analysis

Scientific analysis can be driven by hypotheses about relationships or casual mechanisms between the phenomena in question. Analysis can also be exploratory and seek to derive predictive relationships concerning the general variation of spatially distributed phenomena. Based on an exploratory theme we chose the following analysis. Correlation and multiple regression analysis were conducted to examine the relationship between price and various potential predictors or environmental factors. Our primary goal was to assess the relationship between GDD and Price. We also attempt to derive an equation from that relationship that can be used to predict Price from known values of several independent variables (environmental factors). Based on the current viticulture literature and previous studies, we compiled a subset of independent variables by including all factors identified to be remotely related to successful grape production.

Table 4. Comprehensive list of environmental variables used for regression analysis modeling of viticultural suitability

Variable	Units	Description
[2]Tmin	deg C	Average annual minimum temperature
[2]Tmax	deg C	Average annual maximum temperature
[2]GSAT	deg C	Growing Season Average Temperature (April-Oct)
[2]GDD	deg C	Average annual cumulative degree day accumulation for the growing season (April-Oct)
YrPrcp	deg C	Total annual precipitation in mm
[2]RPMT	deg C	Average ripening period temperature (Jul-Sept)
[2]GTmin	deg C	The average daily minT during the growing season of April- Oct
[2]GTmax	deg C	The average daily maxT during the growing season of April- Oct
[2]DUTR	deg C	The average difference between the maxT & minT on a given day during the growing season of April- Oct
NFrostDays	integer	The No of frost days in the growing season …number of days during the growing season with temp below freezing
netGDD	deg C	Net GDD accumulated during the day and night (Daytime GDD-Nighttime GDD)
Elevation	meters	Height or Distance above sea level in meters
RH	%	average daily relative humidity during the ripening period July-Sept
RainDays	integer	number days of rainfall during the growing season
Srad	Wm-2	total solar radiation for growing season April - Oct
[3]T_pH	–log(H+)	soil reaction of the top soil measured as –log(H+)
[3]Bulk Density	kg/dm3	bulk density of soil measured in kg/dm3
[13]Texture	Categorical	USDA texture class name
[3]T_Sand	% by Wt.	% by Wt. of sand in the topsoil
[3]T_Silt	% by Wt.	% by Wt. of silt in the topsoil
[3]T_Clay	% by Wt.	% by Wt. of clay in the topsoil
[3]AWC	mm/m	available water capacity in mm/m of the soil unit

[1]Categorical
[2]Temperature based indices in °C
[3]Edaphic variables

Conversely we also included as few variables as possible because each irrelevant independent variable decreases the precision of the estimated coefficients and predicted values. The goal of variable selection becomes one of parsimony as we strived to achieve a balance between simplicity (as few independent variables as possible) and fit (as many independent as needed). We thus provide a frame work for evaluating if a variable(s) makes a significant contribution to the prediction of Price. The problem comes down to finding the best function of the form represented by equation 1, the general equation for a multiple regression where b_0 to b_n are partial regression coefficients, X_1 to X_n are measured environmental variables and E is the error term.

$$y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \cdots + b_nX_n + E \ldots\ldots (1)$$

Our analysis was initially carried out simply using GDD as an independent variable and Price as the dependent variable. We set out to test the following hypothesis: GDD is not a reliable predictor of wine grape quality as it does not control for location and inter-annual variation in weather. This analysis was carried out at 3 different modeling levels by controlling for different factors at each level of the analysis. Table 5 illustrates the different levels of the modeling analysis and the factors we controlled for at each level of the analysis.

Table 5. A description of the regression analysis modeling approach for GDD

Model Level	Model Parameters	Dependent	Controls
Level 1	Price vs. GDD	Price	Years, Districts, Varieties
Level 2	Price vs. GDD By Variety	Price	Years, Districts
Level 3	Price vs. GDD By Variety &District	Price	Years

We continued our analysis by also assessing the list of environmental factors we complied using the same frame work applied in the analysis of GDD. Table 6 outlines the frame work used in our analysis of environmental variables. We initially establish a correlation matrix of all the variables and assess the variance inflation factor (VIF) in order to determine multicollinearity. A VIF > 5 generally implies there is an indication for multicollinearity which occurs when two or more predictors are correlated and provide redundant information about the response (Sheather, 2009). Our approach involved evaluation of R^2 and the p-value (p value <0.05) to determine which set of environmental variables best influence price.

Table 6. A description of the regression analysis modeling approach for environmental variables

Model Level	Model Parameters	Dependent	Controls
Level 1	Price vs. All Environmental Vars	Price	Years, Districts, Varieties
Level 2	Price vs. All Environmental Vars By Variety	Price	Years, Districts
Level 3	Price vs. All Environmental Vars By Variety &District	Price	Years

Results and Discussion

Regression analysis was conducted to examine the relationship between price and GDD at multiple levels. Our data consisted of an average annual GDD value for each of the 17 crush districts for the period of 1991 to 2012. These data represented the predictors in our study while Price paid per ton of a variety of grape was our dependent variable. We had previously determined in chapter 4 that there is as much as a 79% variation in an estimate of GDD from year to year. We also reported that estimates of GDD vary spatially. As such, using the Winkler zone classification we illustrated the characterization of California based Winkler's original GDD summations. Figure 51 illustrates the spatial variation of GDD per the Winkler classification based on the original Winkler scale.

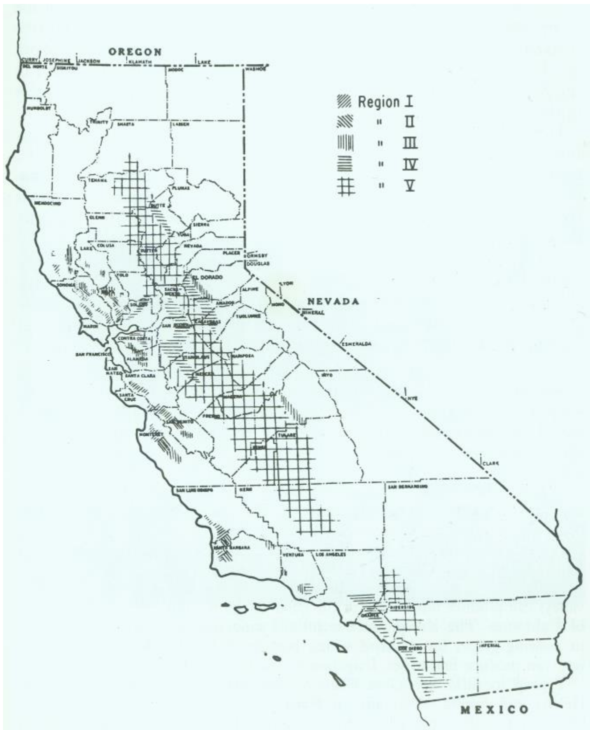


Figure 51. A illustration of the original Winkler Regions as defined by Amerine and Winkler (1944)

Using data for California from the period of 1991-2012 we re-characterized California using the same Winkler scale in order to visualize how GDD varies over time due to the temporal component of the data, but more importantly how GDD varies spatially. Figure 52 illustrates the recreation of Winkler zones based on the crush district delineations and average annual GDD values in California from 1991-2012.

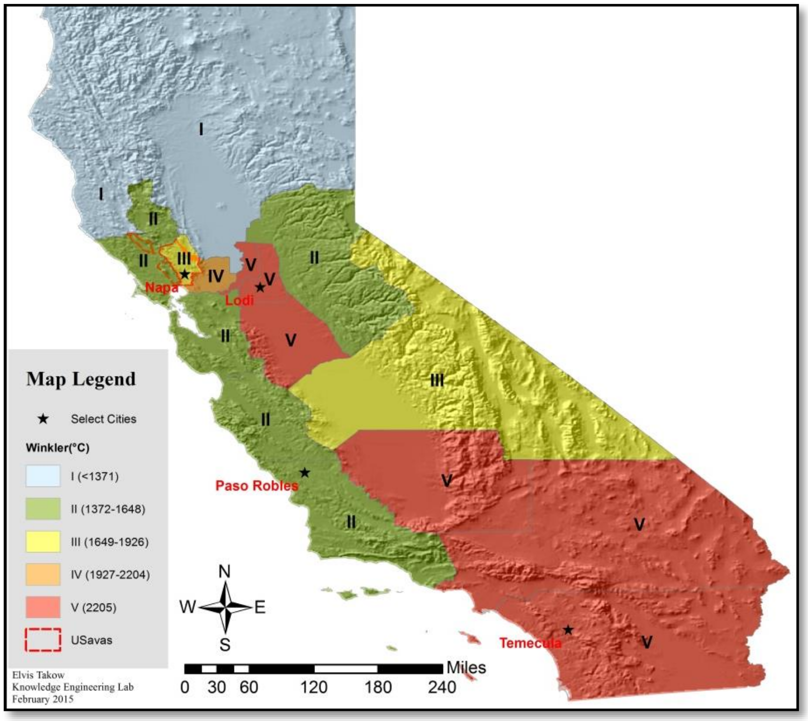


Figure 52. A depiction of the crush districts using Winkler classification based on Average GDD for the period of 1991-2012

Given this change over time and space, we used a linear regression to predict Price based on GDD. Our goal was to investigate the degree to which GDD predicts Price (success). We achieved this at three different levels of analysis by 1) controlling for every year, district, and variety of grape 2), controlling for every year and district, and 3) controlling for every year. At the first level of analysis, a statistically significant (p-value<0.05) regression equation was found with an R^2 of 0.05. Table 7 shows the results of the first level of the analysis. Though GDD significantly predicts Price at this level of the analysis, it only explains 5% of the variability in the data.

Table 7. The results of the first level of a regression analysis to predict Price from GDD, while controlling for Year, District, and Varieties

Level 1

GDD vs Price			
N	**Slope**	**P value**	**R2**
2705	-0.23423	0.0001	0.05

Figure 53 illustrates the regression plot and the relationship between Price and GDD at the first level of analysis and the predictive equation for Price

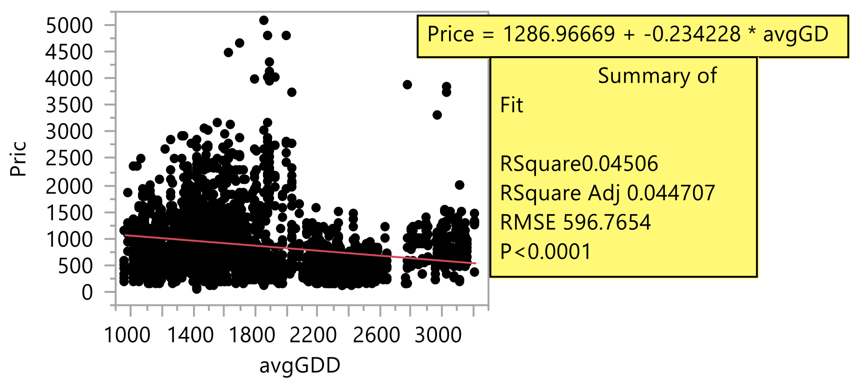


Figure 53. Regression plot for first level of analysis to predict Price from GDD

At the second level of the analysis we found significant regression equations for each variety by controlling for both year and district. We thus discovered GDD to be statistically significant (p-value <0.05) for all varieties. R^2 values however range from 0.05 to 0.12 ($0.05 < R^2 < 0.12$). This implies that GDD only accounts for 5 to 12% of the variation in Price. Table 8 illustrates the results of this second level of analysis implying that GDD alone would not be a reliable predictor of Price given prior knowledge of the variety.

Table 8. The results of the second level of a regression analysis to predict Price from GDD, while controlling for Year and District

Level 2

GDD + Variety vs PRICE				
Variety	**N**	**Slope**	**P p-value**	**R2**
Cabernet Sauvignon	370	-0.31431	0.0001	0.05
Chardonnay	368	-0.21687	0.0001	0.07
Chenin Blanc	327	-0.13627	0.0001	0.07
French Columbard	228	-0.14312	0.0001	0.12
Merlot	368	-0.19938	0.0001	0.05
Pinot Noir	311	-0.48368	0.0001	0.1
Sauvignon Blanc	358	-0.14466	0.0001	0.05
Zinfandel	375	-0.23463	0.0001	0.06

At the third and final level of the analysis we controlled for year by assessing the relationship between GDD and Price for each variety and each of the 17 districts. Overall, there are few statistically significant regression equations for predicting Price from GDD at any stage of this level of analysis. More specifically we have significant results for Cabernet Sauvignon District 3, Chardonnay District 6 and 16, Merlot District 3 and 7, Pinot Noir District 17, Sauvignon Blanc District 3 and Zinfandel District 3. Table 9 illustrates the results of our third level of analysis by each District and each variety of grape.

Table 9. The results of the third level of a regression analysis to predict Price from GDD while controlling for Year

GDD + Variety + District vs PRICE					
District		**N**	**Slope**	**P value**	**R2**
1	Cabernet Sauvignon	22	0.366548	0.5857	0.02
2		22	0.22095	0.6738	0.01
3		22	2.623925	0.0228	0.23
4		22	-1.85734	0.4808	0.03
5		22	-0.53716	0.3312	0.05
6		22	0.073118	0.8397	0
7		22	0.38648	0.2295	0.07
8		22	0.100037	0.7661	0
9		22	-0.50617	0.3169	0.05
10		22	0.404648	0.2782	0.06
11		22	-0.14261	0.6093	0.01
12		22	-0.14674	0.6065	0.01
13		22	-0.35071	0.1978	0.08
14		22	-0.53067	0.1006	0.13
15		18	1.401838	0.0635	0.2
16		22	0.321605	0.5487	0.02
17		22	-0.17638	0.4457	0.03
1	Chardonnay	22	0.347299	0.3544	0.04
2		22	0.313693	0.3139	0.05
3		22	1.21588	0.0682	0.16
4		22	-0.28255	0.7243	0.01
5		22	0.459204	0.2729	0.06
6		22	0.951202	0.0496	0.18
7		22	0.537628	0.161	0.1
8		22	0.231801	0.4606	0.03
9		22	-0.43898	0.3362	0.05
10		22	0.139315	0.6901	0.01
11		22	-0.09645	0.7002	0.01
12		22	0.039528	0.8876	0
13		22	-0.46925	0.0877	0.14
14		22	-0.41439	0.1658	0.09
15		16	1.720727	0.0267	0.3
16		22	0.210102	0.6657	0.01
17		22	0.090559	0.6239	0.01

Table 9. Continued

GDD + Variety + District vs PRICE					
District		**N**	**Slope**	**P value**	**R2**
1	Chenin Blanc	22	0.428517	0.2514	0.07
2		12	-0.5743	0.0719	0.29
3		22	-0.49757	0.3578	0.04
4		22	-1.05368	0.148	0.1
5		22	0.177486	0.3238	0.05
6		11	-0.59161	0.7304	0.01
7		22	0.019822	0.9735	0
8		22	0.08114	0.7627	0
9		22	-0.04897	0.8686	0
10		19	0.51674	0.463	0.03
11		22	-0.06924	0.5524	0.02
12		22	-0.0563	0.609	0.01
13		22	0.061523	0.6044	0.01
14		22	-0.0785	0.4587	0.03
15		3	0.080073	0.0908	0.98
16		18	0.253498	0.5645	0.02
17		22	-0.01439	0.8867	0
1	French Columbard	22	0.069159	0.8392	0
2		-	-	-	-
3		22	0.413452	0.0347	0.2
4		13	-0.01728	0.9827	0
5		20	-0.04397	0.895	0
6		7	-0.70571	0.2297	0.27
7		15	0.894191	0.1371	0.16
8		-	-	-	-
9		21	-0.07373	0.55	0.02
10		7	0.218021	0.0656	0.52
11		22	-0.01019	0.9153	0
12		22	-0.08123	0.4257	0.03
13		22	0.06262	0.581	0.02
14		22	-0.05524	0.5584	0.02
15		-	-	-	-
16		-	-	-	-
17		11	0.015978	0.91	0

Table 9. Continued

GDD + Variety + District vs PRICE					
District		**N**	**Slope**	**P value**	**R2**
1	Merlot	22	0.212529	0.7237	0.01
2		23	0.24196	0.5563	0.02
3		22	1.18618	0.0372	0.2
4		22	-0.21299	0.8283	0
5		22	0.675463	0.1739	0.09
6		22	0.63791	0.1318	0.11
7		22	0.747085	0.0141	0.27
8		22	0.503485	0.1536	0.1
9		22	-0.43182	0.4221	0.03
10		22	0.161517	0.6136	0.01
11		22	-0.04532	0.8828	0
12		22	0.141858	0.6709	0.01
13		22	-0.58983	0.1255	0.11
14		22	-0.40836	0.4636	0.03
15		15	0.691017	0.2205	0.11
16		22	0.135706	0.641	0.01
17		22	0.086232	0.6765	0.01
1	Pinot Noir	22	0.403875	0.7514	0.01
2		15	-1.96208	0.1974	0.12
3		22	2.759538	0.1333	0.11
4		22	-1.02291	0.3899	0.04
5		22	-0.07935	0.8717	0
6		22	-0.43701	0.7	0.01
7		22	-0.37933	0.667	0.01
8		22	-0.51989	0.675	0.01
9		22	0.669033	0.5295	0.02
10		22	0.816977	0.1873	0.09
11		16	-0.23522	0.4349	0.04
12		19	0.102999	0.6271	0.01
13		20	0.108857	0.7922	0
14		12	0.06689	0.8657	0
15		6	-1.51424	0.8061	0.02
16		10	0.05566	0.9632	0
17		15	-0.75263	0.007	0.44

Table 9. Continued

GDD + Variety + District vs PRICE					
District		**N**	**Slope**	**P value**	**R2**
1	Sauvignon Blanc	22	0.19251	0.6292	0.01
2		22	0.188207	0.5769	0.02
3		22	1.427113	0.0312	0.21
4		22	-0.68875	0.4358	0.03
5		22	0.060829	0.7857	0
6		22	0.284183	0.2078	0.08
7		22	0.050875	0.8715	0
8		22	0.005808	0.9835	0
9		22	0.13052	0.6913	0.01
10		22	0.291891	0.3271	0.05
11		22	0.060377	0.6328	0.01
12		22	0.089189	0.4836	0.02
13		22	0.052468	0.4241	0.03
14		22	-0.01155	0.813	0
15		6	1.575363	0.2441	0.32
16		22	0.457246	0.4917	0.02
17		22	0.075908	0.5492	0.02
1	Zinfandel	22	0.266985	0.6595	0.01
2		23	-0.12581	0.7743	0
3		22	3.107541	0.0466	0.18
4		21	-1.87933	0.2792	0.06
5		23	0.229125	0.3922	0.04
6		22	-0.38621	0.5131	0.02
7		22	0.288592	0.5431	0.02
8		22	-0.26597	0.5534	0.02
9		22	-0.06788	0.5015	0.02
10		22	0.623671	0.1481	0.1
11		22	0.028922	0.8796	0
12		22	-0.00924	0.9279	0
13		22	-0.07991	0.5477	0.02
14		22	-0.03212	0.7249	0.01
15		22	0.382539	0.3965	0.04
16		22	1.056968	0.1374	0.11
17		22	0.094185	0.2227	0.07

After 3 levels of analysis we have concluded that GDD on its own is not a good predictor of price at any level of the analysis. We offer the following explanations to support the results of our analysis:

The choice of a measure of success in viticulture or better yet a measure of the suitability of a variety of grape to a particular location is subjective by its very nature. Consequently our choice of Price as a measure of success is undoubtedly subjective. In the absence of a definitive measure of viticultural success, surrogates like Price and Yield have often been used. Ultimately Price may not be a suitable measure of viticultural success.

Price as reported in the California crush district data represents the aggregate price of a particular variety for the entire district. The price reported does not necessarily represent the actual price paid at specific locations where wine grapes are grown. Price represents the average Price paid throughout the entire District and not necessarily representative of where wine grapes are grown within a District. We assume that the aggregate Price paid throughout the District also represents specific locations where wine grapes are actually grown.

Estimates of GDD also represent the average for the entire District and do not represent the environmental conditions of a specific location where wine grapes may actually be grown. Site specific locations may have GDD estimates that vary substantially from the calculated aggregate GDD estimate of the District as reported in our data. The data does not account for the inter-annual variation in an estimate of GDD which we reported in chapter 4 to account for as much as a 79% variation across a location in the US. Table 10 illustrates a representation of the variation in GDD from year to year within a district. The coefficient of variation (CV) shows how GDD varies from year to year. Higher CV values imply greater dispersion in the estimate of GDD.

Table 10. A description of the variation in GDD from year to year within a given District

GDD(°C)						
District	Min	Max	Mean	StDev	Range	CV (%)
1	989.25	1483.817	1166.867	121.5383	494.5667	10.41578
2	1189.458	1757.063	1442.324	142.7614	567.6042	9.898013
3	1210.703	1600.875	1423.322	96.73688	390.1719	6.796554
4	1636.344	2089.219	1878.006	109.801	452.875	5.846683
5	1971.542	2447.375	2217.854	115.6887	475.8333	5.216243
6	1405.305	1870.242	1623.337	110.9626	464.9375	6.835462
7	1285.776	1695.058	1423.432	103.9783	409.2821	7.30476
8	1327.218	1797.141	1580.514	127.7999	469.9234	8.08597
9	958.3301	1294.418	1136.518	88.0376	336.0878	7.746257
10	1336.274	1740.098	1557.652	112.2115	403.8232	7.203891
11	2199.611	2653.5	2424.319	118.7833	453.8889	4.899656
12	2130.379	2584.735	2372.181	117.424	454.3561	4.950045
13	1574.725	2022.834	1826.936	109.622	448.1085	6.000317
14	1948.09	2477.769	2265.448	129.4535	529.6793	5.714256
15	2636.486	3128.459	2949.542	135.5279	491.9729	4.594879
16	2802.02	3220.586	3048.691	107.4897	418.5662	3.525766
17	2157.6	2623.15	2390.639	115.3522	465.55	4.825161

In the second part of our analysis we utilized our compiled list of environmental variables. Based on communication with viticulture experts and previous research, we identified these variables to be relevant to the success of a variety at a particular location or necessary for a location to grow a particular variety of grape. We evaluated multicollinearity by putting together a correlation matrix and also assessing VIF. Highly correlated variables ($\pm 0.80 \geq r \leq \pm 1.0$) were then removed from our initial list in order to develop a final list of environmental variables. As such we removed GSAT, GTmin, GTmax, RPMT, Warmest, and DUTR as these variables fundamentally capture the same climate information. T_pH, T_Bulk, Texture, T_Clay, T_Silt, and T_Sand were also removed from our list as statistical testing found them to be biased. Visual inspection of the correlation matrix illustrated in figure 54 showed these variables to be correlated with one or more of the included variables.

At the first level of analysis, we again controlled for every year, district and variety of grape. The results of this level of analysis are illustrated in table 11. All model parameters are statistically significant (P<.0001) at the first level of the analysis with the exception of NFrostDays, DUTR, T_Silt, and AWC which are not statistically significant.

Table 11. Parameter estimates for environmental variables at first level of analysis

Term	Estimate	Std Error	t Ratio	P-value	VIF
Intercept	-1679.221	1271.324	-1.32	0.1867	.
Coldest	70.384925	7.73869	9.10	<.0001*	4.3498191
YrPrcp	0.3412429	0.048158	7.09	<.0001*	4.1988003
Srad	-0.04437	0.0093	-4.77	<.0001*	11.661044
RH	47.686445	11.10771	4.29	<.0001*	17.753011
NFrostDays	-2.576282	4.902476	-0.53	0.5993	39.628904
RainDays	-15.87434	3.079911	-5.15	<.0001*	6.6951787
netGDD	2.3880287	0.439993	5.43	<.0001*	87.654185
Elev	0.4705256	0.097487	4.83	<.0001*	17.33148
DUTR	-76.06068	50.77838	-1.50	0.1343	54.49258
T_Silt	-9.670446	6.941254	-1.39	0.1637	11.526764
T_Bulk	-550.8404	263.5873	-2.09	0.0367*	6.9505539
Texture	334.80111	40.81839	8.20	<.0001*	6.9176892
AWC	-1.133525	0.713262	-1.59	0.1121	3.0017368
GDD	-0.973036	0.084789	-11.48	<.0001*	22.896017

Correlation Matrix	AWC	T_pH	T_Bulk	Texture	T_Clay	T_Silt	T_Sand	DUTR	Elev	GDD	netGDD	GSAT	GTmax	GTmin	NFrostDays	RainDays	RPMT	YrPrcp	Srad	RH	Warmest	Coldest
AWC	1	-0.0247	-0.2444	0.26	-0.4165	0.147	0.0636	0.2032	-0.4216	0.1139	0.3127	0.1521	0.2048	0.0937	-0.211	0.07	0.0899	0.1629	-0.3928	-0.2434	0.073	0.1822
T_pH	-0.0247	1	0.4858	0.008	-0.502	-0.4654	0.523	-0.0908	-0.2134	0.8467	0.5157	0.8436	0.8219	0.8138	-0.5002	-0.6388	0.8343	-0.7654	0.374	0.4224	0.7771	0.5578
T_Bulk	-0.2444	0.4858	1	0.5546	-0.4682	-0.7914	0.7361	-0.2532	0.3734	0.2969	-0.13	0.2624	0.1981	0.3072	-0.0342	-0.3623	0.3272	-0.4018	0.5184	0.4789	0.3313	0.076
Texture	0.26	0.008	0.5546	1	-0.7962	-0.6261	0.7517	-0.1241	0.4286	0.0784	-0.3211	0.0276	-0.0039	0.0556	0.2824	0.0902	0.1311	0.091	0.1182	0.1712	0.176	-0.2175
T_Clay	-0.4165	-0.502	-0.4682	-0.7962	1	0.6499	-0.8492	0.1022	-0.0928	-0.5765	-0.1078	-0.5406	-0.5156	-0.5318	0.0982	0.2126	-0.6053	0.2685	-0.1415	-0.2842	-0.6055	-0.169
T_Silt	0.147	-0.4654	-0.7914	-0.6261	0.6499	1	-0.9532	0.2273	-0.5517	-0.3497	0.2228	-0.2927	-0.2351	-0.3293	-0.1066	0.3086	-0.4101	0.362	-0.5618	-0.4677	-0.4404	0.0299
T_Sand	0.0636	0.523	0.7361	0.7517	-0.8492	-0.9532	1	-0.1985	0.4201	0.4723	-0.1119	0.4184	0.3684	0.4403	0.035	-0.299	0.5257	-0.3583	0.4466	0.438	0.5468	0.0465
DUTR	0.2032	-0.0908	-0.2532	-0.1241	0.1022	0.2273	-0.1985	1	0.0448	-0.1367	0.5217	-0.1359	0.1201	-0.3676	0.3815	-0.1221	-0.1284	0.0753	0.2754	-0.8615	-0.0867	-0.2679
Elev	-0.4216	-0.2134	0.3734	0.4286	-0.0928	-0.5517	0.4201	0.0448	1	-0.3435	-0.6122	-0.4386	-0.428	-0.4224	0.7919	0.1308	-0.2509	0.125	0.6168	-0.0271	-0.1429	-0.6931
GDD	0.1139	**0.8467	0.2969	0.0784	-0.5765	-0.3497	0.4723	-0.1367	-0.3435	1	0.612	0.9927	0.9597	0.9646	-0.6067	-0.5783	0.9767	-0.6954	0.1739	0.4139	0.9143	0.5437
netGDD	0.3127	0.5157	-0.13	-0.3211	-0.1078	0.2228	-0.1119	0.5217	-0.6122	0.612	1	0.6592	0.7943	0.4936	-0.5332	-0.5437	0.5433	-0.4205	0.0426	-0.2801	0.4756	0.4499
*GSAT	0.1521	0.8436	0.2624	0.0276	-0.5406	-0.2927	0.4184	-0.1359	-0.4386	0.9927	0.6592	1	0.9672	0.9713	-0.6865	-0.5897	0.9579	-0.6844	0.1062	0.4105	0.8838	0.6107
*GTmax	0.2048	**0.8219	0.1981	-0.0039	-0.5156	-0.2351	0.3684	0.1201	-0.428	0.9597	0.7943	0.9672	1	0.8791	-0.5901	-0.622	0.927	-0.6664	0.1769	0.1905	0.8635	0.5432
*GTmin	0.0937	0.8138	0.3072	0.0556	-0.5318	-0.3293	0.4403	-0.3676	-0.4224	0.9646	0.4936	0.9713	0.8791	1	-0.736	-0.5244	0.93	-0.6606	0.0336	0.5922	0.8504	0.6377
NFrostDays	-0.211	-0.5002	-0.0342	0.2824	0.0982	-0.1066	0.035	0.3815	0.7919	-0.6067	-0.5332	-0.6865	-0.5901	-0.736	1	0.3905	-0.5158	0.3652	0.3609	-0.4724	-0.3939	-0.7993
RainDays	0.07	-0.6388	-0.3623	0.0902	0.2126	0.3086	-0.299	-0.1221	0.1308	-0.5783	-0.5437	-0.5897	-0.622	-0.5244	0.3905	1	-0.5045	0.7219	-0.5715	-0.2882	-0.4383	-0.2812
*RPMT	0.0899	**0.8343	0.3272	0.1311	-0.6053	-0.4101	0.5257	-0.1284	-0.2509	0.9767	0.5433	0.9579	0.927	0.93	-0.5158	-0.5045	1	-0.6554	0.2094	0.3818	0.9595	0.4947
YrPrcp	0.1629	-0.7654	-0.4018	0.091	0.2685	0.362	-0.3583	0.0753	0.125	-0.6954	-0.4205	-0.6844	-0.6664	-0.6606	0.3652	0.7219	-0.6554	1	-0.4671	-0.4153	-0.6007	-0.3099
Srad	-0.3928	0.374	0.5184	0.1182	-0.1415	-0.5618	0.4466	0.2754	0.6168	0.1739	0.0426	0.1062	0.1769	0.0336	0.3609	-0.5715	0.2094	-0.4671	1	0.0491	0.243	-0.2915
RH	-0.2434	0.4224	0.4789	0.1712	-0.2842	-0.4677	0.438	-0.8615	-0.0271	0.4139	-0.2801	0.4105	0.1905	0.5922	-0.4724	-0.2882	0.3818	-0.4153	0.0491	1	0.3124	0.4113
Warmest	0.073	0.7771	0.3313	0.176	-0.6055	-0.4404	0.5468	-0.0867	-0.1429	0.9143	0.4756	0.8838	0.8635	0.8504	-0.3939	-0.4383	0.9595	-0.6007	0.243	0.3124	1	0.3879
Coldest	0.1822	0.5578	0.076	-0.2175	-0.169	0.0299	0.0465	-0.2679	-0.6931	0.5437	0.4499	0.6107	0.5432	0.6377	-0.7993	-0.2812	0.4947	-0.3099	-0.2915	0.4113	0.3879	1
	*multicollinearity																					
	**chance correlation																					

Figure 54. A correlation matrix of coefficients illustrating the relationships of each of the environmental variables

Conversely our resultant model only explains approximately 28% of the variation in Price ($R^2 = 0.2859$). A summary of the model at this first level of analysis is presented in table 12. Given the low coefficient of determination (R^2), the majority of the variation in Price cannot be explained by the selected environmental variables. This implies that our model is not a 'good fit' to the data hence variables are not good predictors of price.

Table 12. Summary of model fit for all varieties at the first level of analysis

Summary of Fit All Varieties	
RSquare	0.285901
RSquare Adj	0.282451
Root Mean Square Error	517.203
Mean of Response	845.0342
Observations	2705

At the second level of the analysis we re-examined the environmental variables in table 11 by controlling for every year and district in order to assess the influence on Price. A summary of the results indicate R^2 ranged from 33% to 59% ($0.33 < R^2 < 0.59$). Across all varieties, we found most of the variation in Price remains unexplained by the environmental variables. We found varieties such as Chardonnay, Pinot Noir, Cabernet Sauvignon, and Merlot to exhibit more broad environmental suitability as these varieties showed a wide range of statistical significance in environmental parameters. Most of the parameters were statistically significant for Chardonnay, Pinot Noir, Cabernet Sauvignon and Merlot implying a broader range of environmental suitability. Jones et al (2006) have suggested that grape varieties can be characterized by maturity grouping based average growing season temperature. Figure 55 shows a depiction of the broad climate range of varieties like Chardonnay, Pinot Noir, Cabernet Sauvignon and Merlot.

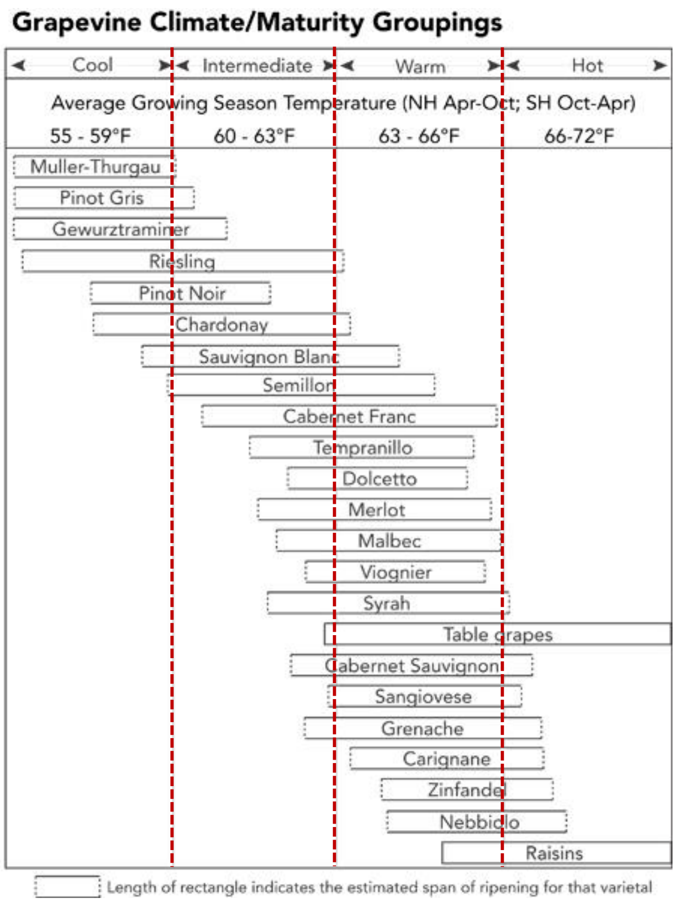
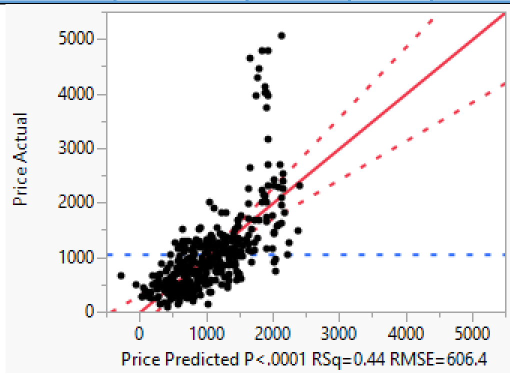


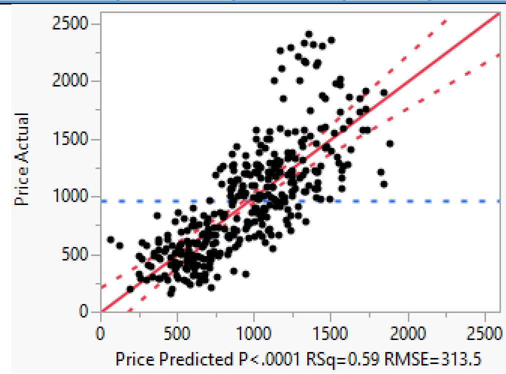
Figure 55. A depiction of grapevine climate based on average growing season temperature (source: Jones et al, 2006)

We present the second level of analysis results in figure 56 with parameter estimates, regression plot and a summary of the proposed model fit.

Cabernet Sauvignon				
Term	**Estimate**	**StdError**	**tRatio**	**Prob>|t|**
Intercept	-4486.035	4074.939	-1.10	0.2717
Coldest	91.632979	24.91518	3.68	0.0003*
YrPrcp	0.5167661	0.156693	3.30	0.0011*
Srad	-0.06589	0.03016	-2.18	0.0296*
RH	43.36624	35.78188	1.21	0.2263
NFrostDays	-1.47226	15.87586	-0.09	0.9262
RainDays	-16.79343	9.930181	-1.69	0.0917
netGDD	5.3169536	1.397006	3.81	0.0002*
Elev	1.1730216	0.312153	3.76	0.0002*
DUTR	-333.0191	161.323	-2.06	0.0397*
T_Silt	21.450937	20.75038	1.03	0.3020
T_Bulk	309.83317	832.044	0.37	0.7098
Texture	502.51195	130.3256	3.86	0.0001*
AWC	2.3778812	2.292531	1.04	0.3003
GDD	-1.556175	0.265057	-5.87	<.0001*

Chardonnay				
Term	**Estimate**	**StdError**	**tRatio**	**Prob>|t|**
Intercept	-5169.706	2114.998	-2.44	0.0150*
Coldest	76.216766	12.91596	5.90	<.0001*
YrPrcp	0.4872643	0.081044	6.01	<.0001*
Srad	-0.021344	0.01564	-1.36	0.1732
RH	68.113671	18.55842	3.67	0.0003*
NFrostDays	-2.797426	8.271852	-0.34	0.7354
RainDays	-12.50171	5.144018	-2.43	0.0156*
netGDD	2.9823646	0.736733	4.05	<.0001*
Elev	0.3772683	0.162624	2.32	0.0209*
DUTR	-98.265	85.33394	-1.15	0.2503
T_Silt	-13.88047	11.16006	-1.24	0.2144
T_Bulk	-857.6443	438.4251	-1.96	0.0512
Texture	380.33541	68.1043	5.58	<.0001*
AWC	-2.413385	1.192586	-2.02	0.0438*
GDD	-1.13331	0.1406	-8.06	<.0001*





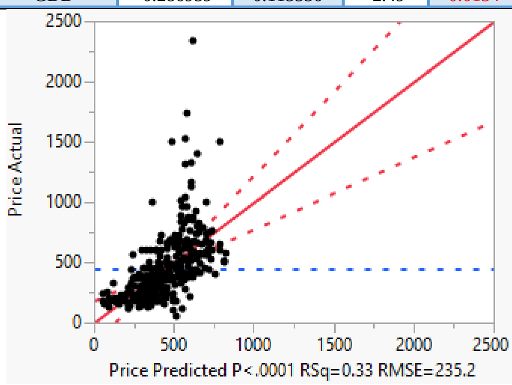
Summary of Fit Cabernet Sauvignon	
RSquare	0.443911
RSquare Adj	0.421981
Root Mean Square Error	606.4046
Mean of Response	1056.356
Observations	370

Summary of Fit Chardonnay	
RSquare	0.585973
RSquare Adj	0.569553
Root Mean Square Error	313.4973
Mean of Response	962.0787
Observations	368

Figure 56. Parameter estimates and summary of model fit at second level of regression analysis

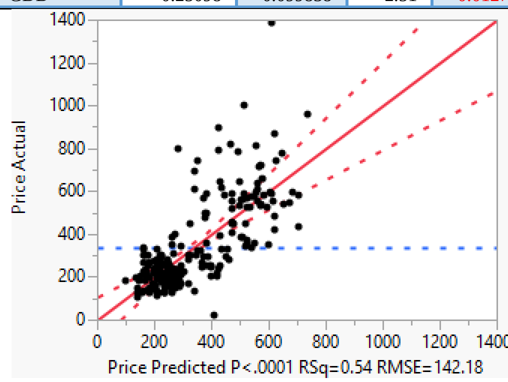
Chenin Blanc

Term	Estimate	StdError	tRatio	Prob>|t|
Intercept	-628.6743	1704.699	-0.37	0.7125
Coldest	37.615224	10.33186	3.64	0.0003*
YrPrcp	0.1353071	0.064275	2.11	0.0361*
Srad	-0.024568	0.012247	-2.01	0.0457*
RH	28.164157	14.74239	1.91	0.0570
NFrostDays	-1.854722	6.519657	-0.28	0.7762
RainDays	-5.787868	4.175269	-1.39	0.1667
netGDD	0.1639715	0.588916	0.28	0.7809
Elev	0.080251	0.13102	0.61	0.5406
DUTR	66.535019	66.91723	0.99	0.3209
T_Silt	1.9320439	10.63671	0.18	0.8560
T_Bulk	165.72916	357.1288	0.46	0.6429
Texture	48.278768	53.45767	0.90	0.3672
AWC	0.0184443	0.928398	0.02	0.9842
GDD	-0.286959	0.115336	-2.49	0.0134*



French Colombard

Term	Estimate	StdError	tRatio	Prob>|t|
Intercept	-848.420	1445.806	-0.59	0.5579
Coldest	6.9713673	8.648013	0.81	0.4211
YrPrcp	0.1961358	0.045439	4.32	<.0001*
Srad	0.001793	0.009073	0.20	0.8435
RH	-12.1845	11.29186	-1.08	0.2818
NFrostDays	2.030687	5.066799	0.40	0.6890
RainDays	-2.73534	3.147191	-0.87	0.3857
netGDD	0.8418629	0.484545	1.74	0.0838
Elev	0.0998609	0.113231	0.88	0.3788
DUTR	-87.2753	55.44118	-1.57	0.1169
T_Silt	10.24957	11.22708	0.91	0.3623
T_Bulk	485.36243	344.6447	1.41	0.1605
Texture	16.164362	57.43802	0.28	0.7787
AWC	3.9491837	1.026017	3.85	0.0002*
GDD	-0.25098	0.099838	-2.51	0.0127*



Summary of Fit Chenin Blanc	
RSquare	0.325195
RSquare Adj	0.294916
Root Mean Square Error	235.203
Mean of Response	446.2173
Observations	327

Summary of Fit French Colombard	
RSquare	0.54489
RSquare Adj	0.514976
Root Mean Square Error	142.1818
Mean of Response	337.2799
Observations	228

Figure 56. Continued

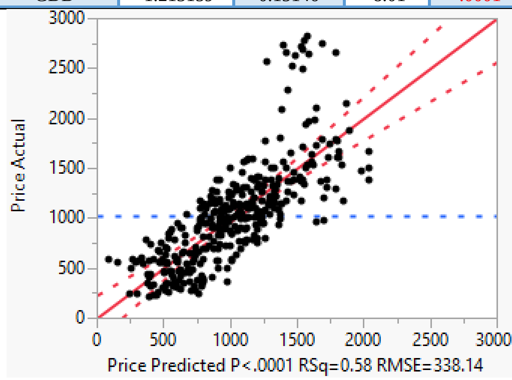
Merlot	Pinot Noir

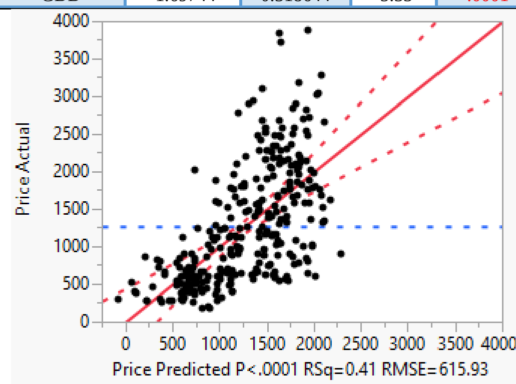
Merlot

Term	Estimate	StdError	tRatio	Prob>|t|
Intercept	-7055.313	2267.807	-3.11	0.0020*
Coldest	65.81358	13.74723	4.79	<.0001*
YrPrcp	0.6695665	0.08731	7.67	<.0001*
Srad	-0.016506	0.016855	-0.98	0.3281
RH	64.421119	19.96747	3.23	0.0014*
NFrostDays	5.5288658	8.853788	0.62	0.5327
RainDays	-9.942929	5.5236	-1.80	0.0727
netGDD	4.3562906	0.790383	5.51	<.0001*
Elev	0.4669726	0.175424	2.66	0.0081*
DUTR	-246.0338	91.96782	-2.68	0.0078*
T_Silt	-9.046237	12.1646	-0.74	0.4576
T_Bulk	-301.659	474.5168	-0.64	0.5254
Texture	398.17412	73.54723	5.41	<.0001*
AWC	-0.702698	1.28738	-0.55	0.5855
GDD	-1.213159	0.15146	-8.01	<.0001*

Pinot Noir

Term	Estimate	StdError	tRatio	Prob>|t|
Intercept	12241.125	4548.039	2.69	0.0075*
Coldest	110.17723	28.02273	3.93	0.0001*
YrPrcp	0.2007393	0.17006	1.18	0.2388
Srad	-0.124677	0.033459	-3.73	0.0002*
RH	50.523614	38.62356	1.31	0.1919
NFrostDays	22.874151	17.09207	1.34	0.1818
RainDays	-36.26807	10.82428	-3.35	0.0009*
netGDD	3.6524063	1.60939	2.27	0.0240*
Elev	0.0377838	0.350118	0.11	0.9141
DUTR	-143.9258	187.2047	-0.77	0.4426
T_Silt	-117.7821	27.58969	-4.27	<.0001*
T_Bulk	-2186.051	1036.522	-2.11	0.0358*
Texture	368.85783	162.6624	2.27	0.0241*
AWC	-6.791768	2.662666	-2.55	0.0113*
GDD	-1.69744	0.318644	-5.33	<.0001*





Summary of Fit Merlot

RSquare	0.584593
RSquare Adj	0.568118
Root Mean Square Error	338.1372
Mean of Response	1021.533
Observations	368

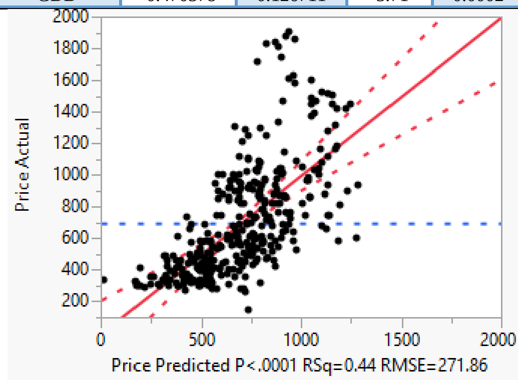
Summary of Fit Pinot Noir

RSquare	0.406809
RSquare Adj	0.378753
Root Mean Square Error	615.928
Mean of Response	1268.483
Observations	311

Figure 56. Continued

Sauvignon Blanc

Term	Estimate	StdError	tRatio	Prob>|t|
Intercept	-4467.934	1878.837	-2.38	0.0180*
Coldest	56.069632	11.31071	4.96	<.0001*
YrPrcp	0.2434609	0.070567	3.45	0.0006*
Srad	-0.020648	0.01365	-1.51	0.1313
RH	48.961767	16.34429	3.00	0.0029*
NFrostDays	-15.17013	7.275413	-2.09	0.0378*
RainDays	-6.884703	4.511128	-1.53	0.1279
netGDD	0.5715789	0.660607	0.87	0.3875
Elev	0.4567303	0.142082	3.21	0.0014*
DUTR	127.43638	76.05407	1.68	0.0947
T_Silt	9.546557	10.99834	0.87	0.3860
T_Bulk	-616.6278	396.8389	-1.55	0.1211
Texture	245.4929	60.19672	4.08	<.0001*
AWC	-0.086795	1.042767	-0.08	0.9337
GDD	-0.470378	0.126711	-3.71	0.0002*

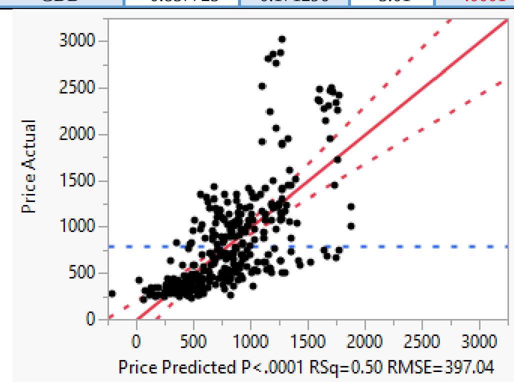


Summary of Fit Sauvignon Blanc

RSquare	0.442325
RSquare Adj	0.419563
Root Mean Square Error	271.8618
Mean of Response	698.4658
Observations	358

Zinfandel

Term	Estimate	StdError	tRatio	Prob>|t|
Intercept	-2502.331	2645.091	-0.95	0.3448
Coldest	66.159245	16.09079	4.11	<.0001*
YrPrcp	0.3024628	0.102361	2.95	0.0033*
Srad	-0.036929	0.019675	-1.88	0.0613
RH	40.030528	23.27029	1.72	0.0862
NFrostDays	-23.86574	10.26458	-2.33	0.0206*
RainDays	-19.85332	6.468656	-3.07	0.0023*
netGDD	1.1099937	0.897515	1.24	0.2170
Elev	0.7639217	0.203796	3.75	0.0002*
DUTR	-3.7125	104.4103	-0.04	0.9717
T_Silt	22.274709	13.14333	1.69	0.0910
T_Bulk	-612.1281	541.2082	-1.13	0.2588
Texture	380.08781	85.04594	4.47	<.0001*
AWC	2.924016	1.497608	1.95	0.0517
GDD	-0.857725	0.171296	-5.01	<.0001*



Summary of Fit Zinfandel

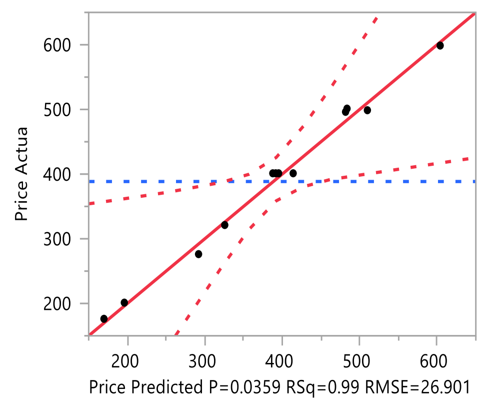
RSquare	0.503364
RSquare Adj	0.48405
Root Mean Square Error	397.037
Mean of Response	793.6926
Observations	375

Figure 56. Continued

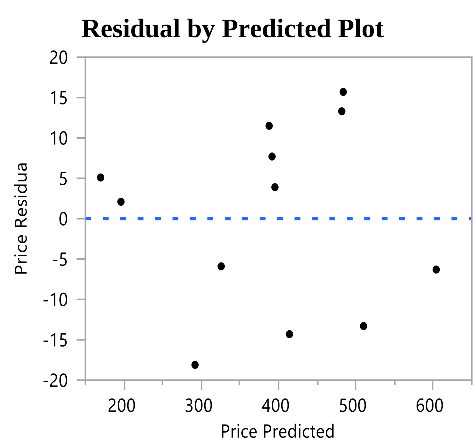
At the third level of analysis, there is little to no statistical significance in any of the environmental variables by district or by variety. This means that regardless of the variety or district chosen to plant wine grapes in California, our selected environmental variables were not reliable predictors of price. Upon examination of the R2 values a statistically significant ($p<0.05$) regression equation was found for Pinot Noir in district 14. The derived model explains approximately 99% of the variation in Price. We displayed the results of our analysis in figure 57 illustrating parameter estimates, regression plot and a summary of the proposed model fit.

Pinot Noir District 14

Term		Estimate	StdError	tRatio	Prob>|t|
Intercept	Biased	6088.9226	3174.763	1.92	0.1951
Coldest		79.605366	11.28837	7.05	0.0195*
YrPrcp		-1.604471	0.161154	-9.96	0.0099*
Srad		-0.115792	0.025333	-4.57	0.0447*
RH		105.81401	22.04579	4.80	0.0408*
NFrostDays		-595.3883	45.96261	-12.95	0.0059*
RainDays		-39.65002	6.157909	-6.44	0.0233*
netGDD		-41.2531	3.104645	-13.29	0.0056*
DUTR		5213.5697	401.3494	12.99	0.0059*
GDD		1.6313836	0.314584	5.19	0.0352*



Summary of Fit	
RSquare	0.991911
RSquare Adj	0.955512
Root Mean Square Error	26.90104
Mean of Response	388.5292
Observations	12



Prediction Equation

$$\boldsymbol{Price} = 6088.922 + 79.605 * Coldest + -1.604 * YrPrcp + -0.11579 * Srad + 105.814 * RH + -595.388 * NFrostDays + -39.650 * RainDays + -41.253 * netGDD + 5213.5697 * DUTR + 1.631 * GDD$$

Figure 57. The results of parameter estimates and a summary of model fit at the third level of regression analysis for Pinot Noir district 14

We have summarized the results of the third level of the regression analysis on the relationship between environmental variables and price in figure 58.

District	Variety	Parameter(s)	P-value	Estimate	R^2	adjR^2	N
1	Cabernet Sauvignon						
	Chardonnay						
	Chenin Blanc						
	French Columbard						
	Merlot	YrPrcp, RH	0.0491, 0.0225	-0.044217, 143.23972	0.545453	0.204542	22
	Pinot Noir						
	Sauvignon Blanc						
	Zinfandel						
4	Cabernet Sauvignon						
	Chardonnay	NFrostDays	0.064	443.57998	0.586035	0.275561	22
	Chenin Blanc	Srad, NFrostDays	0.0400, 0.0166	-0.286334, 538.81546	0.5524449	0.216786	22
	French Columbard						
	Merlot						
	Pinot Noir	NFrostDays	0.027	713.02731	0.62886	0.350505	22
	Sauvignon Blanc	NFrostDays	0.0465	482.07284	0.602972	0.305201	22
	Zinfandel						
5	Cabernet Sauvignon						
	Chardonnay	YrPrcp, RainDays	0.0121, 0.0418	1.0441395, -39.15684	0.681636	0.485719	22
	Chenin Blanc						
	French Columbard						
	Merlot	YrPrcp	0.002	1.5510472	0.720513	0.548521	22
	Pinot Noir						
	Sauvignon Blanc	RainDays	0.0328	-28.71788	0.447167	0.106962	22
	Zinfandel						

Figure 58. A summary of the results of the third level of analysis for all statistically significant parameters by variety and district

District	Variety	Parameter(s)	P-value	Estimate	R^2	$adjR^2$	N
6	Cabernet Sauvignon	YrPrcp, NFrostDays	0.0205, 0.0348	0.8994167, 156.83553	0.545644	0.204877	22
	Chardonnay	YrPrcp, Srad	0.0028, 0.0176	1.3400728, 0.2014354	0.740244	0.545426	22
	Chenin Blanc						
	French Columbard						
	Merlot	Coldest, YrPrcp	0.0436, 0.0015	117.67558, 1.3687292	0.690083	0.457646	22
	Pinot Noir						
	Sauvignon Blanc						
	Zinfandel	Srad	0.0391	-0.297755	0.412631	-0.0279	22
7	Cabernet Sauvignon						
	Chardonnay						22
	Chenin Blanc	NFrostDays	0.0231	167.41812	0.559814	0.229674	22
	French Columbard						
	Merlot						
	Pinot Noir	Coldest, YrPrcp, NFrostDays	0.0017, 0.0166, 0.0277	-276.846, '-1.125316, 164.29066	0.791311	0.634794	22
	Sauvignon Blanc	NFrostDays	0.032	72.954948	0.393394	112.4957	22
	Zinfandel						
8	Cabernet Sauvignon						
	Chardonnay	NFrostDays	0.0459	82.063272	0.646477	0.381334	22
	Chenin Blanc	YrPrcp	0.0436	-0.394346	0.565174	0.239054	22
	French Columbard						
	Merlot						22
	Pinot Noir	Coldest, YrPrcp	0.0117, 0.0037	-334.664, '-2.23802	0.740964	0.546687	22
	Sauvignon Blanc	Coldest, YrPrcp, Srad	0.0293, 0.0135, 0.0224	-71.05273, '-0.459344, '-0.128476	0.670884	0.424046	22
	Zinfandel	Coldest, YrPrcp, Srad	0.0162, 0.0071, 0.0140	-117.4604, '-0.752902, '-0.206369	0.726506	0.521386	22

Figure 58. Continued

District	Variety	Parameter(s)	P-value	Estimate	R^2	$adjR^2$	N
9	Cabernet Sauvignon						
	Chardonnay	NFrostDays, netGDD, DUTR, GDD	0.0129, 0.0145, 0.0077, 0.0288	-118.6877, '-21.59726, 2464.0919, 7.2125263	0.608931	0.315628	22
	Chenin Blanc	RainDays	0.0235	-17.11062	0.677444	0.435527	22
	French Columbard						
	Merlot						
	Pinot Noir						
	Sauvignon Blanc						
	Zinfandel						
10	Cabernet Sauvignon	YrPrcp, RH, RainDays	0.0296, 0.0282, 0.0392	0.4428273, 223.58883, '-26.58188	0.669804	0.422157	22
	Chardonnay	YrPrcp, RH	0.0229, 0.0395	0.5358973, 237.24705	0.492398	0.111696	22
	Chenin Blanc						
	French Columbard						
	Merlot	YrPrcp, RH, GDD	0.0010, 0.0352, 0.0267	0.5554733, 152.87488, '-1.764502	0.762433	0.584258	22
	Pinot Noir	RainDays	0.0381	-41.58302	0.714839	0.500969	22
	Sauvignon Blanc	RH	0.0222	201.82584	0.616089	0.328156	22
	Zinfandel	YrPrcp, RainDays	0.0461, 0.0225	0.3705602, '-27.92394	0.79169	0.635457	22
11	Cabernet Sauvignon						
	Chardonnay	YrPrcp	0.0369	0.8016892	0.352495	-0.04597	22
	Chenin Blanc	RH	0.0468	-39.85108	0.505966	0.201945	22
	French Columbard	RH	0.0129	-42.2931	0.517221	0.220126	22
	Merlot	YrPrcp	0.0133	1.0353682	0.526097	0.234464	22
	Pinot Noir						
	Sauvignon Blanc						
	Zinfandel						

Figure 58. Continued

District	Variety	Parameter(s)	P-value	Estimate	R^2	adjR^2	N
14	Cabernet Sauvignon						
	Chardonnay	Coldest	0.0404	88.8109	0.57705	0.259838	22
	Chenin Blanc						
	French Columbard						
	Merlot	YrPrcp	0.0263	1.9174763	0.630509	0.35339	22
	Pinot Noir						
	Sauvignon Blanc						
	Zinfandel						
16	Cabernet Sauvignon						
	Chardonnay						
	Chenin Blanc						
	French Columbard						
	Merlot						
	Pinot Noir						
	Sauvignon Blanc	YrPrcp	0.0462	-1.890171	0.627999	0.348998	22
	Zinfandel						
17	Cabernet Sauvignon	YrPrcp	0.0456	0.6901631	0.382984	0.003282	22
	Chardonnay	YrPrcp	0.0384	0.5469368	0.435733	0.088491	22
	Chenin Blanc						
	French Columbard						
	Merlot	YrPrcp	0.0069	0.723625	0.591159	0.339564	22
	Pinot Noir						
	Sauvignon Blanc						
	Zinfandel						

Figure 58. Continued

Conclusion

Defining the relationship between environmental variability and measures of viticultural success provides wine producers, wine grape growers and researchers information to assess an areas potential for growing wine grapes. GDD is characteristically the most widely used index of variety suitability however no published study has scientifically or objectively tested whether GDD is actually a good index of variety suitability. Furthermore measures of success in viticulture represent a subjective notion often best described by surrogate measures such as yield, price of a bottle of wine, or the price paid for a ton of a particular variety of grapes. This research focused on primarily addressing the hypothesis that GDD is not a reliable index of variety suitability. We further assessed other commonly used environmental variables in order to define the relationship between environmental variability and the success or suitability of a variety. We used price as a measure of the success or suitability of a variety. Price refers to the average price paid by producers for a particular variety of grape.

Our approach focused on three separate levels of analysis by controlling for 1) variety, district and year 2) district and year 3) year. Though GDD was statistically significant at the first and second level of the analysis, much of the variation in Price could not be explained by GDD. Consequently at the third level of the analysis, GDD was not statistically significant regardless of prior knowledge of the variety and district. While GDD is useful in determining the annual phenological development of the grape vine, we conclude that GDD alone is not a good predictor of price as a measure of viticultural success at any level of the analysis.

On a broader environmental scale, we assessed the relationship between success (price) and a number of predetermined environmental variables. Though some varieties showed a greater range of environmental tolerance, we conclude that environmental variation is not a reliable predictor of wine grape suitability. We could not conclude whether different varietal selection processes have taken place in different locations due to environmental variation. It is important to clarify though that appropriate measures of wine grape suitability or success must be established. Viticulture has often used indirect measures to assess success such as price which can be influenced by a number of market forces. The effects of supply and demand on the economy or the influence of consumer preference often has an influence on price. More appropriate and direct measures of suitability in viticulture must be determined in order to properly assess relationships between environmental variability and viticultural success. We defined viticultural success as either determining locations suitable for a particular variety of grape or determining varieties best suited for a particular location. We can conclusively say that viticultural success is also driven by other processes that are more disconnected from the environment such as socio-economic factors. There is no doubt the variation in environmental conditions drives the global distribution of wine grapes (Gladstones 1992; Wilson 1998; Winkler et al 1974; van Leeuwen et al. 2004; van Leeuwen and Seguin 2006). However it is difficult to assess whether this variation is significant enough is to be modelled and more importantly whether we can determine measures to appropriately capture this variation. We assumed that the influence of environmental variability was strong enough that we could model these processes. It’s worth noting that wine growing regions as described in this chapter do not refer to wine regions referred to in the Code of Federal Regulations better known as American Viticultural Areas (AVA).

CHAPTER VI

INTEGRATION OF CLIMATIC AND EDAPHIC CHARACTERISTICS IN VITICULTURE SITE SELECTION: ENVIRONMENTAL LANDSCAPE VITICULTURE INFORMATION SYSTEM (ELVIS)

Introduction

In this chapter we discuss the development of a decision support system for viticulture. This system is grounded upon the principles of data management outlined in chapter 2. The goal of the system is to provide growers with instant access to data and consequently information and knowledge to make more informed decisions about what varieties to grow and where to grow them.

Agriculture is based on the principles of maximizing yield, reducing cost, and being a steward of the land. Historically viticulture has always involved trial and error in making choices regarding where to plant as well as varietal selections. Given the financial burden associated with growing grapes, site selection is consequently considered the most important decision when establishing a vineyard. The process of matching grape varieties to environmental conditions will affect yields and profitability for the life of the vineyard and is a determining factor for economic success in wine grape production. It is therefore necessary to ensure the vineyard site is evaluated for adequate climatic conditions, soil quality, and varietal appropriateness. These factors directly impact the long-term sustainability of any potential viticulture endeavor. The goal of this chapter is to describe how the development of a decision support system for viticulture can be used to make more informed decisions about where or what to plant.

The Environmental Landscape Viticulture Information System (ELVIS) was conceived to allow current and prospective growers the ability to rapidly explore, compare, and analyze environmental conditions relevant to grapevine growth and varietal selection. At the most basic level, this will allow users to reach more informed decisions about the selection of potential sites most likely to support grapes of a given variety or selection of the varieties that are most suitable for a particular location. The foundation of this technology is grounded upon 3 functional principles that will allow viticulturists to effectively explore, compare, and analyze a scientific approach to vineyard selection. The process of site selection or matching varieties to a location is becoming an increasingly exact science that involves careful objective analysis of climate, soil, and topographic data. For any location around the world, a user can explore raw environmental data, compare between one or several locations and analyze locations using pre-existing models and or novel environmental indices.

Decision Support Systems (DSS) like ELVIS will consequently allow users to better understand grape varietal appropriateness for a particular location by facilitating the evaluation of locations most appropriate for growing a particular variety of grape. ELVIS facilitates the decision making process by providing instant access to environmental data. By utilizing a series of analysis ranging from linear regression to simple comparisons of similarity between locations, ELVIS enables a better understanding of variety appropriateness.

Methods

System architecture

The system was designed to support research through Web-based GIS applications structured around a spatially explicit environmental database. The client/server architecture proposed in Figure 59 was implemented using open source tools in order to guarantee the web application's sustainability. Implementation of customized geospatial and analytical functions were easily developed given a design which allows user specified data queries driven by a comprehensive collection of environmental factors relevant to grape vine growth.

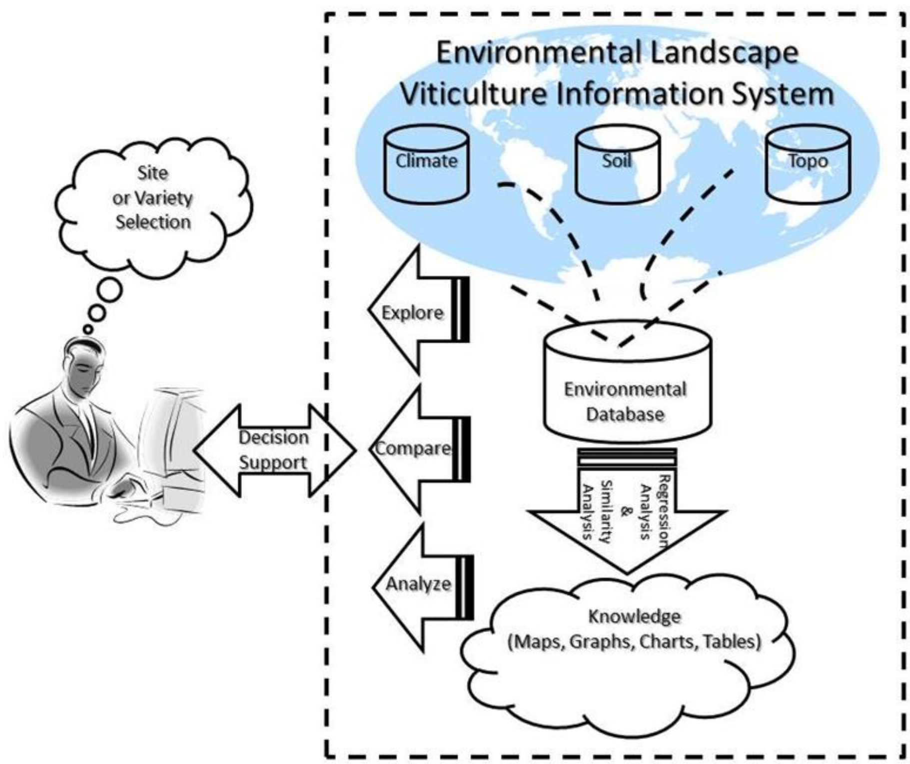


Figure 59. An overview of the web-based system architecture of the ELVIS decision support system structured around an environmental database

The client server structure provides growers with the information they need at a scale they can use for managing spatially based information. This architecture played a key role in integrating spatially based information from a variety of sources. A central database of relevant climate, soil, and topographic variables as well as grape variety information lies at the core of this system architecture. The ability to query all environmental factors relevant to viticulture within one central location is a novelty of this system architecture. We discuss the development of the data and consequent management system in chapter 2 of this dissertation.

Environmental Database

One of the initial steps in this endeavor was to assimilate a spatial database of environmental indices most relevant to viticulture. This process involved collecting climate, soil and topographic data from various sources for careful construction of a database management system. Climate data was sourced from the Oakridge National Laboratory (ORNL), National Climatic Data Center (NCDC), and the European Climate Assessment and Dataset (ECA&D). Data obtained from Daymet consisted of a collection of algorithms and computer software that interpolate and extrapolate daily meteorological data using digital elevation models (DEMs) and a data set of daily observations from ground-based meteorological stations (Thornton et al., 1997). Climate indices were derived by applying mathematical formulas written in C# to calculate values for individual observations stored as csv files for the entire spatial extent of the U.S. at a specific resolution of latitude and longitude. For the purposes of our research, the climate data ranged from 1980 to December of 2013 with daily elements of mean temperature, dew point, maximum temperature, minimum temperature, precipitation, and elevation. Soils data was derived primarily from the Soil Survey Geographic (SSURGO) database which provides the most detailed level of soil information for the continental U.S.A. Global soil coverage beyond the continental US was obtained from the Harmonized World Soil Database for global soil data. Parameters included in the soil database were soil texture, pH, soil depth, and water holding capacity. One of the goals of this project was to provide a useful, geographically coherent, multi-source and site-specific data base to support viticulture. The environmental database is a relational database designed to store, query, and manipulate geographic information and spatial data relevant to viticulture. The primary advantage of spatial databases, over file-based data storage, is that they allow implementation of geospatial functions and GIS procedures necessary to model viticultural environments best suited for specific varieties. This includes support for SQL and the ability to generate complex geospatial queries. Moreover the database's client/server architecture supports multiple users and allows them to query and visualize the database with a standard web browser

Data Modeling

The conceptual design of the database is based on the idea of georeferenced queries and data retrieval for statistical analysis of site-specific environmental variations with the potential to expand the relations given new data for new growing regions. Figure 60 depicts a conceptual visualization of how the system design and functions are centered on a data driven geo-spatial perspective.

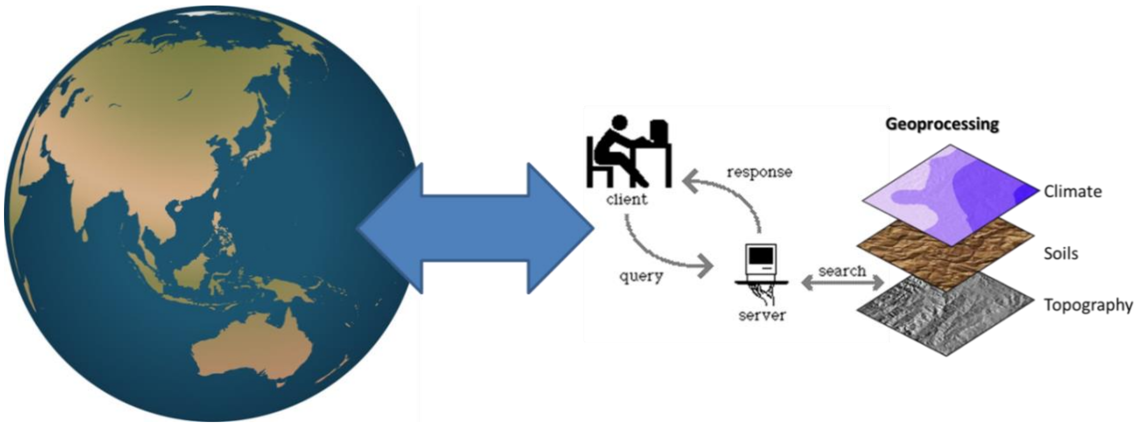


Figure 60. An illustration of how ELVIS functions through the utilization of data for geo-spatial analysis of defined areas of interest

Our efforts reflect continued research into the development of information management systems to gain a more complete understanding of regional factors and their influences on grapevine growth and 'success'. Techniques for identifying the interactions among these factors range from modeling specific effects of the environment within and among vineyards. Simple yet novel indices describing summary variations include cumulative degree days (GDD), annual precipitation (YrPrcp), growing season average temperature (GSAT), relative humidity (RH) estimations and ripening period mean temperature (RPMT) can be established for any location within the global map interface.

Discussion

The development of this system centered upon 3 main stages which were critical to the success of this technology. Simply stated, users of the system can explore, compare, and analyze data within and between locations. In the exploratory phase, raw data of environmental factors can be downloaded in both tabular and graphical formats. This is then returned to the user as an excel spreadsheet in its most native form prior to any user specified analysis. We explore a location by clicking on the map interface or defining an area of interest. Concurrently environmental indices derived from climatic and edaphic data are selected from a precompiled list or drop down menu. These indices are used to explore specific environmental conditions for the chosen location. In figure 61 we have used the rectangle tool to outline an area in the Willamette Valley of Oregon. A user can now proceed by characterizing this area based on specific environmental factors relevant to viticulture. This characterization is only made possible by the instant of access of user defined data for a particular location or area of interest.

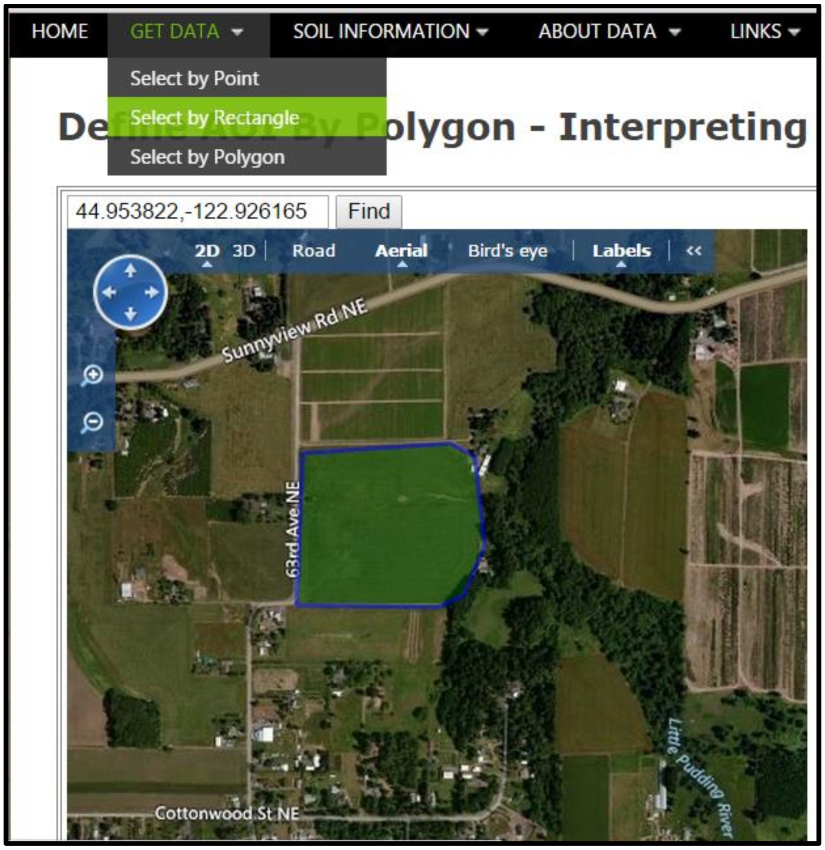


Figure 61. A depiction of the process of defining an area of interest for exploration using the rectangle tool in the ELVIS system

Figure 62 illustrates a soil and climate summary of our area of interest in the Willamette valley. User defined variables have been chosen for characterization of the region of interest. By implementing the principles of data management discussed in chapter 2 we are now able to rapidly summarize a region based on climatic and edaphic factors specific to viticulture.

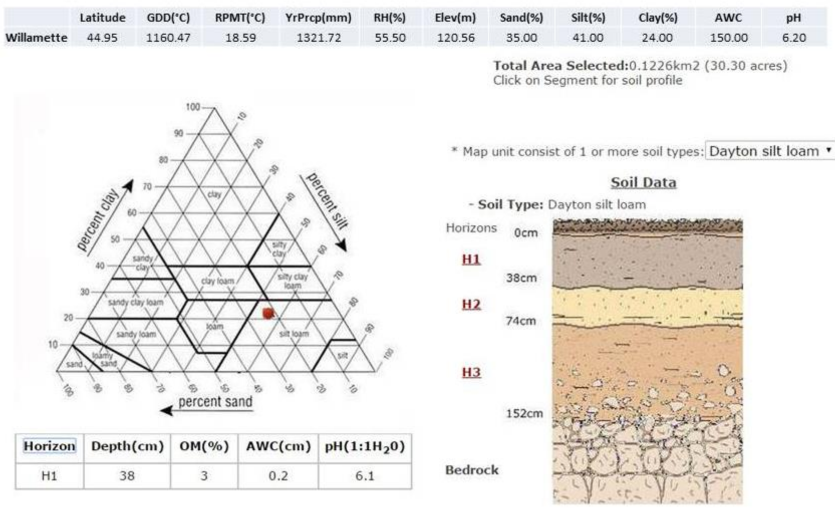


Figure 62. An illustration of the environmental summary of the area of interest characterize by factors specific to viticulture

The comparative phase allows a user to compare specific indices for two or more regions of interest. The graphic user interface allows for simultaneous comparison of locations Figure 63 illustrates two areas of interest outlined using the rectangle tool from our ELVIS interface.

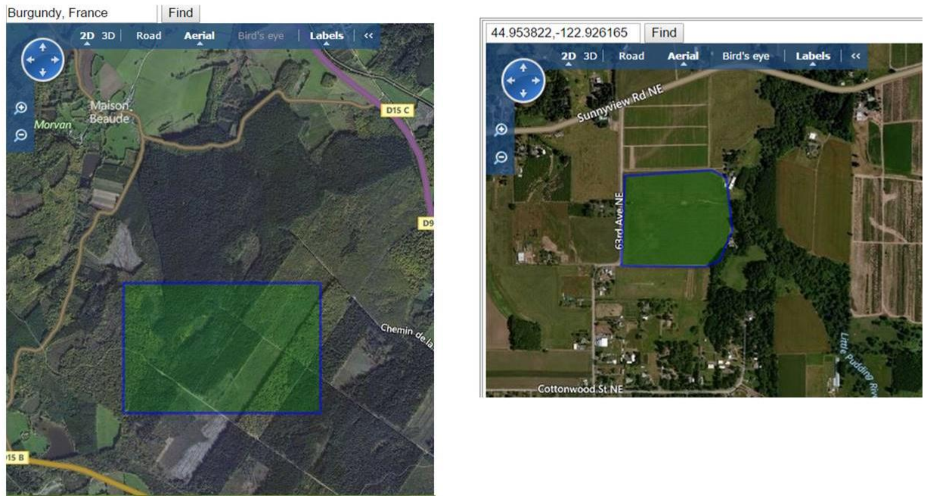
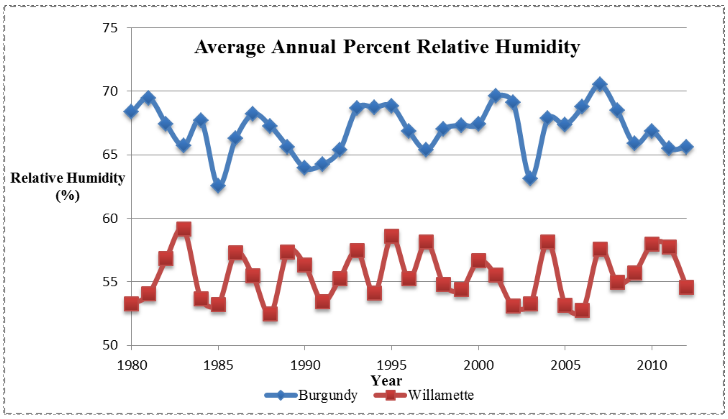


Figure 63. An illustration of the outlined areas from interest for simultaneous comparison using the compare option of the ELVIS system

Based on the selection of these two areas of interest, we can graphically compare the locations based on similarities or dissimilarities in environmental conditions relevant to viticulture. Figure 64, 65, and 66 depict a simple visual comparative analysis of multiples variables for two regions renown for growing Pinot noir, namely Burgundy, France and the Willamette Valley in Oregon. By simultaneously comparing these two locations, we can quickly assess how these two areas renowned for growing a particular variety of grape compare to each other for select environmental conditions. We are instantly able to contrast how the same environmental variable compares over two spatially variable locations. We would otherwise only be able to achieve such an assessment through the efforts of several ours of data collection. The unique without instant access to environmental data

In the analysis phase we utilize scientific approaches towards interpretation of data and model results. This includes analysis techniques such as regression which allows for extrapolation and delineation of other locations. The implication for viticulture is that we are now able to rapidly assess the potential for growing a particular variety of grape at another location presumably also suitable for viticultural ‘success.



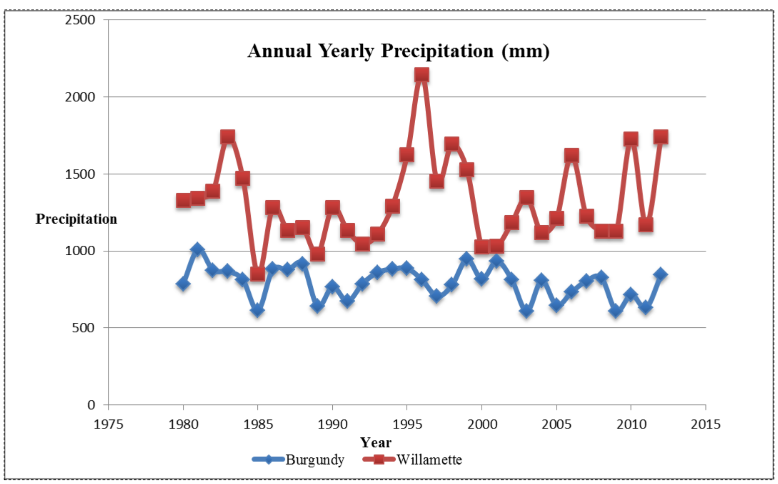


Figure 64. Graphical comparison of average annual relative humidity and annual precipitation in Burgundy and Willamette illustrating the ability to compare multiple variables instantly

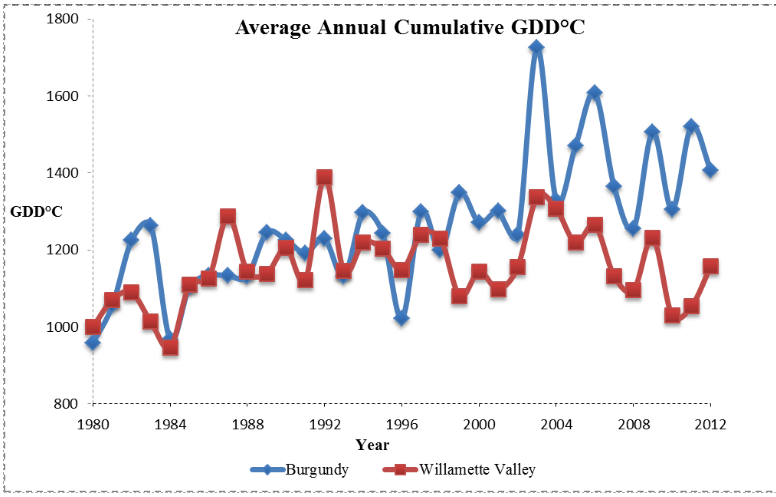


Figure 65. A graphical comparison of average annual cumulative GDD in Burgundy and Willamette valley

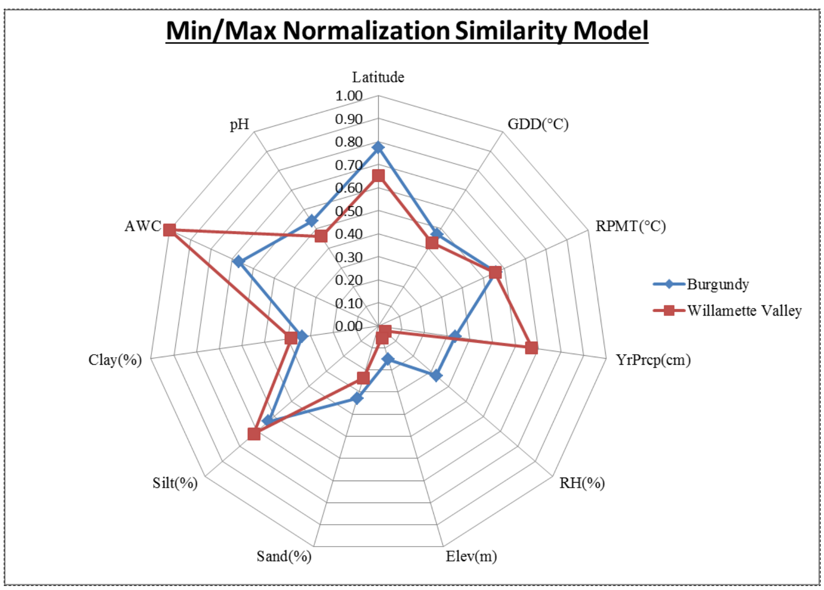


Figure 66. An illustration of a spider chart of Burgundy and Willamette valley illustrating simultaneous comparison of multiple variables only possible through the analysis phase

Users can choose any global location for analysis based on derived environmental indices. The ability to graphically compare variations at one or several locations by analysis of indices such as degree days or soil texture type, demonstrates one of the key objectives of this research. In chapter 5 we outline an approach for modeling the relationship between environmental variability and 'success' in viticulture. Emphasis was placed on the development of a multi-level modeling approach where we control for the year, location, and variety of grape. This analysis is made possible by the unique 3 tier framework of the ELVIS system.

The consequent extrapolation of results out to new locations illustrates the final stage of the design functionality.

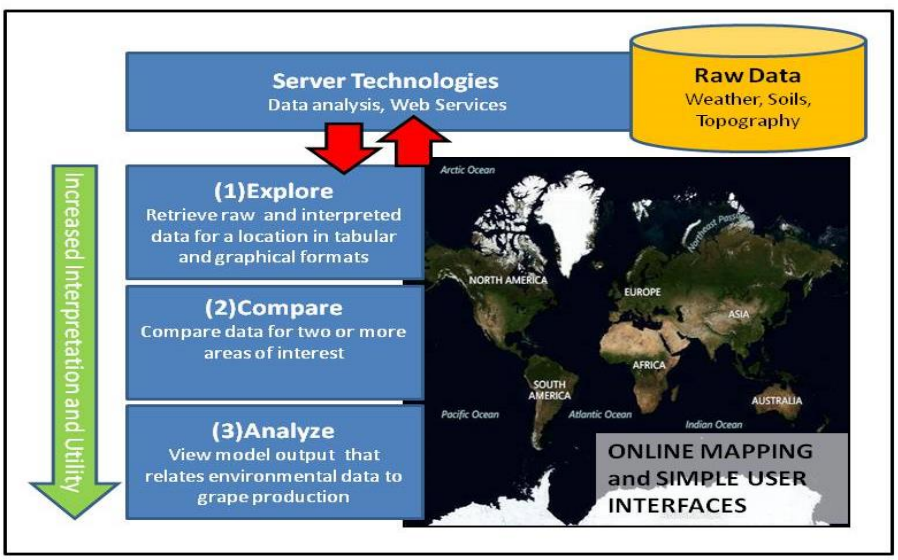


Figure 67. A visual display of the ELVIS system functionality and the various levels of user interaction illustrating increasing levels of interpretation of data

There is an increased level of interpretation as the user interacts with the technology lending to the uniqueness of the system. This increased level of user interaction and interpretation is illustrated in figure 67. With increased interpretive utility, our models provide a scientific approach towards determination of adapted varieties to a particular location. As such with the instant access to environmental data current and prospective growers are able to undertake more informed decisions about where and what varieties to grow.

Conclusion

Data acquisition, visualization and modeling of data in viticulture were utilized to help current and prospective growers make more informed decisions. As such growers are able to more conclusively evaluate which varieties are most suited to a particular site or which sites are most suited to particular varieties. By implementing dynamic web-based technology, viticulturist and researchers can access geographic information and data using a standard desktop computer without installing expensive GIS software. A centralized database of environmental variables allows instant access to data and information for any location in the world. User interpretation of data and model results allows for extrapolation in order to delineate future locations most suited for growing a particular variety. Exploratory and comparative analysis of environmental variation between locations is one of the key aspects of the system technology. Moreover, future developments of this web application will address a fully customizable GUI with statistical and analysis tools beyond the current descriptive indices.

The current research concludes that data management is an essential part of responsible research. The value of data increases as it is aggregated into collections and as data becomes more available for re-use in addressing new or challenging research questions. The value of data is greatly diminished without proper organization; as such we proposed the ELVIS system to facilitate decision making for site and variety selection in viticulture.

The goal of this research was to extend current knowledgebase into the relationships between local climate and successful viniferous grape production. Our approach involved identifying areas of the world that currently have established, successful wine producing areas, and obtain a number of simple climatic indices and relationships that most parsimoniously explain the success of grape varieties grown in such regions. Utilization these indices and relationships is prevalent throughout the viticulture literature as the basis for evaluating new areas for the production of different grape varieties.

Due to the quality and quantity dimension of viniferous grape production, short term weather is arguably the most important factor in the successful, sustained production of grapes. As such inter-annual variations in climate can have specific effects on the quality and quantity of any single years grape harvest, which in turn has a large effect on the long term success of a vineyard. We proposed a web-based system grounded upon the management and development of an environmental database. This system enables rapid access to appropriate environmental factors which facilitate decisions about where and what grape varieties to grow. This is achieved by summarizing complex, short term weather patterns into indices simple enough to develop into models of grape production for better decision making. With the use of ELVIS we are able to scientifically and objectively model these indices for any particular location.

CHAPTER VII

CONCLUSION

The goal of this research was to understand the relationship between environmental variability and viticultural success. As such we set out to provide a knowledge base for evaluating either the most suitable location for growing a particular variety of wine grape or the most appropriate variety for a particular property of land. To this end we outlined an approach to managing voluminous amounts of data towards the development of an environmental database.

The modern era of “big data” and technology however necessitates the need for a scientifically objective approach towards managing “big data”. Due to the broad scale distribution of wine grapes, a comprehensive collection of environmental data was required that covers a broad geographic area. This often required multiple data sets of varying spatial and temporal resolutions in order to develop an environmental data base that covers a broad enough spatial and temporal extent which exhibits environmental variability. Our approach for managing “big data” can be applied to any large scale modelling endeavor than spans an extensive geographic region over time. We consequently established that efficient and effective data management in the modern era of “big data” is dependent on understanding the value of data hence adhering to the following principles and steps: (1) Data acquisition and collection, (2) Data integration and aggregation, (3) Data analysis and modelling and (4) Data interpretation. These steps ensure the perpetuity of the data for future research use.

We described a conceptual approach to building models for site and variety selection in viticulture. We established that the first and most significant step involved determining a clear and consistent dependent variable or in the case of this research a measure of viticultural success. The next objective associated with building models for site and variety selection was an understanding of the functional relationships between variety suitability, environmental conditions, and measures of success. As such the second step involved collecting objective environmental data in order to determine these relationships.

We underscore the broad use of GDD in viticulture as a measure of the suitability of a region to growing a particular variety of grape or the suitability of variety to a particular location. We therefore presented a case for evaluating limitations in the concept of an estimate of GDD by considering of a number of factors which may influence its application for grape variety and site suitability in viticulture. Our findings suggest that data availability and technology drive the use of GDD as a concept in agriculture. However GDD calculations are sensitive to the temporal resolution of weather data thus interpretation of GDD as a concept in viticulture is reliant on both an understanding of the method of calculation and the resolution of the data used for calculation. We concluded that GDD is not just about temperature that requires a start date, end date and a threshold or base temperature but it is subject to errors in the year to year variation. These errors in GDD estimates do not imply that estimates are incorrect, but rather express how literally estimates should be applied to variety selection and the suitability of a region. We determined elevation, latitude, and longitude to account for approximately 88% of the variation in an estimate of GDD calculated at one location to the next. The implications of our findings mean that GDD for a particular location is therefore dependent upon an understanding both the practical (Simplicity) & theoretical (Applicability) implications. Though widely used and accepted in viticulture, an estimate of GDD must be applied with caution and a thorough understanding of the limitations mentioned above.

As a result of our findings in some of the limitations of GDD as an estimate in viticulture, we assessed the relationship between other environmental conditions and a measure of viticultural suitability in California. We determined GDD to not be a reliable index of wine grape variety suitability using price as a measure of success in California. Our analysis was conducted at three different levels. At the first level of analysis we controlled for the variety, the location (district) and the year. At the second level we controlled for the location (district) and the year. At the third level of our analysis we controlled for the year. Though GDD is generally useful in determining the annual phenological development of the grape vine, we concluded that GDD alone is not an adequate predictor of price as a measure of viticultural success at any level of the analysis. The lack of a clear and consistent measure of viticulture success and the subjective nature of current measures may contribute to our results. We also assessed the relationship between success (price) and a number of predetermined environmental variables. Some varieties showed a greater range of environmental tolerance (they were grown ‘successfully’ in a number of different Districts) but overall we concluded that environmental variation on its own is not a reliable predictor of wine grape suitability.

Human factors such as viticultural and enological techniques influence success in viticulture (Seguin 1986). As such the combined influences of environmental and human factors play the greatest role in driving success in viticulture. The influence of the environment alone does not play a significant enough role in the viticultural success in order for us to model these influences. Additionally the current viticultural indices may not adequately capture the environmental conditions which drive ‘success’ in viticulture. Future research should include the evaluation of new and novel indices that more accurately quantify variation in environmental conditions that may drive site and variety selection

Perhaps a more appropriate approach to the problem of site and variety selection in viticulture is an understanding of how similar two regions are in the context of viticulture. As opposed to only understanding the influence of environmental factors on wine grape production, we therefore present a more general approach by quantifying the similarity between two or more locations. The goal would be to develop an index of similarity grounded on the premise that Old World viticultural regions with an extended history of success are presumed suitable for growing specific varieties of wine grapes. As such New World and prospective locations which exhibit similar conditions are potentially also suitable for growing the same varieties. Similarity however is a subjective measure of how different or similar two objects are relative to some contextual and often subjective standard. As such a first step in assessing the similarity between two regions is a clear and consistent definition of the viticultural context. Given the stated premise of our approach, context in this case would refer to all or select Old

World viticultural regions. These regions, whether selected by variety or simply by location would form the foundation for a subjective but consistent understanding of which environmental factors are important to viticultural success. We therefore believe this broad scale approach of similarity analysis is more practical is assessing viticultural suitability.

Finally we described the design of a decision support system for site and variety selection in viticulture. This system is grounded upon the principles of data management outlined in chapter 2 of this dissertation. Emphasis should be placed on the idea that modeling and consequent site and variety selection in viticulture can only be achieved through proper data management. As such the ability to rapidly and instantaneous query data for any user defined location is central to the ELVIS system. Consequent analysis of data is only possible due to a central environmental database of spatially explicit factors deemed relevant to ‘successful’ viticulture.

REFERENCES

Allen, Jon C. "A modified sine wave method for calculating degree days. “Environmental Entomology 5.3 (1976): 388-396.

Amerine, Maynard Andrew, and Albert Julius Winkler Composition and quality of musts and wines of California grapes. Berkley: University of California, 1944.

Ashenfelter, Orley. "Predicting the quality and prices of Bordeaux wine*." The Economic Journal 118.529 (2008): F174-F184.

Barham, Elizabeth. "Translating terroir: the global challenge of French AOC labeling." Journal of rural studies 19.1 (2003): 127-138.

Baskerville, G. L., and P. Emin. "Rapid estimation of heat accumulation from maximum and minimum temperatures." Ecology (1969): 514-517.

Bohmrich, Roger. "Terroir: Competing perspectives on the roles of soil, climate and people." Journal of Wine research 7.1 (1996): 33-46.

Bowen, P. A., et al. "Geology and wine 10: Use of geographic information system technology to assess viticulture performance in the Okanagan and Similkameen Valleys, British Columbia." *Geoscience Canada* 32.4 (2005).

Céline AL. The Appellation d'Origine Contrôlée (AOC) and other official product identification standards. University of Kentucky, Department of Agricultural Economics; 1998.

Cesaraccio, Carla, et al. "An improved model for determining degree-day values from daily temperature data." International Journal of Biometeorology 45.4 (2001): 161-169.

Code of Federal Regulations: Alcohol, Tobacco Products, and Firearms, 2001, (Title 27,

de Blij, Harm Jan. "Geography of viticulture: rationale and resource." Journal of Geography 82.3 (1983): 112-121.

Dry, P. R., and R. E. Smart. Vineyard site selection: in. Winetitles, Adelaide, Australia, 1988.

Fanet, Jacques. Great wine terroirs. University of California Press, 2004.

Gladstones, John. Viticulture and environment. Winetitles, 1992.

Golan, Amos, and Haim Shalit. "Wine quality differentials in hedonic grape pricing." Journal of Agricultural Economics 44.2 (1993): 311-321.

Goodchild, Michael F., Louis T. Steyaert, and Bradley O. Parks. GIS and environmental modeling: progress and research issues. John Wiley & Sons, 1996.

Haeger, John W., and Karl Storchmann. "Prices of American Pinot Noir wines: climate, craftsmanship, critics." Agricultural Economics 35.1 (2006): 67-78.

Hancock, Jake. "Feature review." Journal of Wine Research 10.1 (1999): 43-49.

Hauagge, Roberto, and James N. Cummins. "Phenotypic variation of length of bud dormancy in apple cultivars and related Malus species." Journal of the American Society for Horticultural Science 116.1 (1991): 100-106.

Hellman, Edward W. "Grapevine structure and function." Oregon Viticulture. Hellman, EW (Ed.). Oregon State University Press, Corvallis (2003): 5-19.

Jackson, D. I., and M. B. Spurling. "Climate and viticulture in Australia." Viticulture. Winetitles, Adelaide (1988): 91-106.

Jackson, Ronald S. Wine science: principles and applications. Academic press, 2008.

Johnson, Hugh, and Jancis Robinson. The world atlas of wine. New York, NY.: Simon & Schuster, 1994.

Jones, Gregory V., and Robert E. Davis. "Climate influences on grapevine phenology, grape composition, and wine production and quality for Bordeaux, France." American Journal of Enology and Viticulture 51.3 (2000): 249-261.

Jones, G. V. "Climate and terroir: Impacts of climate variability and change on wine." Geoscience Canada reprint series 9 (2006): 203-217.

Jones, Gregory V., and E. Hellman. "Site assessment." Oregon Viticulture: Oregon State University Press, Corvallis, Oregon (2003): 44-50.

Jones, Gregory V., and Robert E. Davis. "Using a synoptic climatological approach to understand climate-viticulture relationships." International Journal of Climatology 20.8 (2000): 813-837.

Jones, Gregory V., Ryan Reid, and Aleksander Vilks. "Climate, grapes, and wine: structure and suitability in a variable and changing climate." The Geography of Wine. Springer Netherlands, 2012. 109-133.

Leopold, Aldo Carl. "Plant growth and development." Plant growth and development. (1964).

Lombard, P., and E. A. Richardson. "Physical principles involved in controlling phenological development." Modification of the aerial environment of plants. BJ Barfield and JF Gerber (Eds.) (1979): 429-440.

McMaster, Gregory S., and W. W. Wilhelm. "Growing degree-days: one equation, two interpretations." Agricultural and Forest Meteorology 87.4 (1997): 291-300.

MKF Research (2009), Economic impact of California wine 2009, St. Helena California

Moncur, M. W., et al. "Base temperatures for bud break and leaf appearance of grapevines." American Journal of Enology and Viticulture 40.1 (1989): 21-26.

Monroe, Kent B. "Buyers' subjective perceptions of price." Journal of marketing research (1973): 70-80.

Mullins, Michael G., Alain Bouquet, and Larry E. Williams. Biology of the grapevine. Cambridge University Press, 1992.

Oczkowski, Edward. "Hedonic wine price functions and measurement error." Economic record 77.239 (2001): 374-382.

Olson, Jerry Corrie. Price as an informational cue: Effects on product evaluations. No. 43. College of Business Administration, Pennsylvania State University, 1976.

Reaumur, RAF de. "Observation du thermometer, faites à Paris pendant l'année 1735, compares avec celles qui ont été faites sous la ligne, à l'Isle de France, à Alger et en quelques-unes de nos isles de l'Amérique." Paris: Mémoires de l'Académie des Sciences 1735 (1935).

Schwarz, Christina V., et al. "Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners." (2009).

Sheather, Simon. A modern Approach to Regression with R. Springer Science & Business Media, 2009.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database. Available online at http://sdmdataaccess.nrcs.usda.gov/. Accessed [8/4/2010].

Takow, Elvis Arrey. Utility Of The American Viticultural Areas Of Texas Information Systems (AVATXIS) As A Tool In The Characterization Of The Texas Wine Regions. Diss. Texas A&M University, 2008.

Thornton, Peter E., Steven W. Running, and Michael A. White. "Generating surfaces of daily meteorological variables over large regions of complex terrain." Journal of Hydrology 190.3 (1997): 214-251.

Thornton, Peter E., et al. "Daymet: Daily surface weather on a 1 km grid for North America, 1980-2008." Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center for Biogeochemical Dynamics (DAAC) 1 (2012).

Tonietto, Jorge, and Alain Carbonneau. "A multicriteria climatic classification system for grape-growing regions worldwide." Agricultural and Forest Meteorology 124.1 (2004): 81-97.

Van Leeuwen, Cornelis, and Gerard Seguin. "The concept of terroir in viticulture." Journal of Wine Research 17.1 (2006): 1-10.

Van Leeuwen, Cornelis, et al. "Influence of climate, soil, and cultivar on terroir." American Journal of Enology and Viticulture 55.3 (2004): 207-217.

Webb, L. B., P. H. Whetton, and E. W. R. Barlow. "Modelling the relationship between climate, wine grape price and wine grape quality in Australia." Climate Research 36.2 (2008): 89.

White, Michael A., et al. "Extreme heat reduces and shifts United States premium wine production in the 21st century." Proceedings of the National Academy of Sciences 103.30 (2006): 11217-11222.

Wilson, James E. Terroir: The role of geology, climate and culture in the making of French wines. Univ of California Press, 1998.

Winkler, Albert Julius. General viticulture. Univ of California Press, 1962.

Wolf, Tony K. "Site selection for commercial vineyards." Publication-Virginia Cooperative Extension Service (USA) (1988).

APPENDIX A

Climate data for this research included data which covers a spatial scale spanning the entire US and most of Europe. The temporal scale of the data varied as the US data covers every day for the period of 1980-2012. The European data covers every day for the period of 1980-2013. A tertiary data set of station data also provides daily data on a less consistent basis for the period of 1960-2012. Due to the voluminous amount of data, algorithms were derived to sort through each data source. The results were data in a form amenable for further analysis.

A summary of the pseudo code used to develop C# code for our analysis at every level of this research is presented below. SQL Server Management Studio© was used as a database structure for organizing all the data.

```
       //Here we are basically sorting through all the data based on latitude and
longitude extent. The function GetInterpolatedClimateData calls other functions which
define the user format in which the data is returned to the user.
        public int yearfrom = 1980;
        public int yearto = 2013;
        public double outputLat = 0;
        public double outputLon = 0;

          public void GetInterpolatedClimateData(double Lat, double Lon)
        {
            if (Lat >= 0)
            {
                RawData.Capacity = 100000;
                if ((Lat >= 24.00 && Lat <= 49.50) && (Lon <= -66.50 && Lon >= -
125.00))//U.S.(daymet) latitudinal and longitudinal extent
                {
                    OpenDayMetFile(Lat, Lon);
                }
                else if ((Lat >= 25.00 && Lat <= 75.00) && (Lon >= -40.00 && Lon <=
75.00))//Europe(European Climate Assessment) latitudinal and longitudinal extent
                {
                    OpenNCDF_Elevation(Lat, Lon);
                    OpenNCDF(Lat, Lon, 1980, 1994);
                    OpenNCDF(Lat, Lon, 1995, 2013);
                }
            }
```

```
            else if (Lat < 0)
            {
                RawData.Capacity = 100000;
                OpenAussieClimateFile(Lat, Lon);
            }
        }

       //Here we are returning the data in user defined format in order to continue
with the analysis

////////Calculate Daily GDD...(April-Oct)
        public ArrayList CalculateDayDegrees()
        {
            DateTime startdate = new DateTime(this.yearfrom, 4, 1);
            DateTime enddate = new DateTime(this.yearfrom, 10, 31);
            DateTime stopdate = new DateTime(this.yearfrom, 11, 01);
            if (outputLat < 0)// only use if you have data in the southern hemisphere
            {
                startdate = new DateTime(this.yearfrom, 10, 01);
                enddate = new DateTime(this.yearfrom + 1, 4, 30);
                stopdate = new DateTime(this.yearfrom + 1, 5, 01);
            }
            ArrayList GDD = new ArrayList();
            double cumDD = 0;
            if (this.RawData.Count > 0)
            {
                for (int z = 0; z < this.RawData.Count; z++)
                {
                    DailyWeatherVars DayDegrees = (DailyWeatherVars)this.RawData[z];

                    if (DayDegrees.date >= startdate && DayDegrees.date <= stopdate)
                    {
                        if (DayDegrees.Tavg != -999.99)
                        {
                            double DD = DayDegrees.Tavg - 10;
                            if (DD < 0)
                            {
                                DD = 0;
                            }
                            else
                            {
                                cumDD += DD;
                            }
                        }
                        //double DD = ((DayDegrees.Tmax-DayDegrees.Tmin)/2) - 10;
                    }

                    if (DayDegrees.date == stopdate)
                    {
                        ////store data
                        double[] yeardata = new double[2];
                        yeardata[0] = DayDegrees.date.Year;
                        yeardata[1] = cumDD;
                        GDD.Add(yeardata);
                        cumDD = 0; //reset
                        startdate = startdate.AddYears(1);// = new
DateTime(startdate.Year + 1);
```

```
                stopdate = stopdate.AddYears(1);

            }

        }
    }

    return GDD;
}
```