António Carlos Corte-Real de Sousa

ANALYZING THE INFLUENCE OF THE DOURO VALLEY WEATHER ON THE QUALITY AND YIELD OF VINTAGE PORT

Thesis submitted to the Faculty of Engineering of the University of Porto in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Industrial Engineering and Management.

Supervisor: José Luís Cabral de Moura Borges, Assistant Professor of the Faculty of Engineering of the University of Porto

Department of Industrial Engineering and Management
Faculty of Engineering of the University of Porto (FEUP)
September 2014

Na vinificação do Douro pouco ou quase nada haverá que reformar.

Os processos tradicionais e empíricos, cotejados com as mais recentes teorias científicas, dão em último resultado a perfeição. O vinicultor nem sempre saberá talvez a razão científica daquilo que faz, mas faz sempre, por hábito contraído e por costume herdado, aquilo que deve fazer.

−Ramalho Ortigão, *As Farpas I*, 1887

Acknowledgements

First, I would like to express my special appreciation and thanks to Professor **José Luís Cabral de Moura Borges** for the promptness, effective guidance and strong encouragement through the whole PhD process. He guided me with great patience, he taught me, he showed me the right path and he helped me overcome my difficulties, always pushing me to do my very best. His attention to detail showed me that I fail far more often than I would imagine. He pays great attention to aspects not necessarily related to the scope of the research, in areas that range from psychology and communication to aesthetics and art. With him, I learned that a PhD process is much broader than a deep study on one subject or area. It was a great pleasure and learning experience to work with him.

I also want to express my personal admiration and gratitude to Professor **José António Sarsfield Pereira Cabral** that with his vast knowledge on the Douro Valley region provided wise opinions whenever they have been necessary. He was a mentor for this work.

I am grateful to the co-authors of the published articles and communications related to this work: Professor José Luis Cabral de Moura Borges (FEUP), Professor José António Sarsfield Pereira Cabral (FEUP) and Professor **Gregory V. Jones** (Southern Oregon University).

To Professor **José Fernando da Costa Oliveira**, Director the Doctoral Program in Industrial Engineering and Management (PRODEIG), I express my admiration for the excellent organization of the course, for the availability, and for effectiveness in addressing students' issues.

I acknowledge with gratitude the financial support of **ADVID - Associação para o Desenvolvimento da Viticultura Duriense**.

I am especially grateful to my friend **Carlos Alberto Bragança de Oliveira** whose telephone call inviting me to join him in PRODEIG doctoral course made all this to happen.

Some friends, in crucial moments, created the environment necessary for me to keep going. I deeply acknowledge **Maria Manuela da Rocha Ferreira**, **Francisco da Rocha Ferreira**, **Raquel Costa Oliveira, João Nuno Costa Oliveira Macedo Silva** and, last but not least, **Mário Rui Noya de Andrade Maia**.

I greatly value the friendship of many anonymous friends that have helped me stay relatively sane throughout my life. I deeply appreciate their unconditional support and their presence in so many moments that made my life fulfilled and enjoyable.

I pay a special tribute to my deceased parents that raised me with love and guidance. I am aware that I have not been the easiest son to raise, but I believe that I became a person that they would be proud of.

Finally, I thank my wife **Ana Mafalda Almeida Araújo Pires Monteiro** for her love, partnership, patience, and contagious joy of living.

Abstract

Corte-Real de Sousa, António Carlos. **Analyzing the Influence of the Douro Valley Weather on the Quality and Yield of Vintage Port**. Universidade do Porto, September 2014. Advisor: José Luís Cabral de Moura Borges.

The Douro Valley is a well-known wine region. Port wine has been produced in this region since the end of 18th century and it is the region's most important economic factor. Climate characteristics determine the type of grapes that can be grown in a region while the weather specificities influence the ripening of grapes and the quality of the wine of each vintage. For several wine regions, previous research studied the influence of weather variability in vintage outcome, identifying weather factors related to quality and yield[1]. Moreover, the relation of climate long-term trends with the evolution of quality and yield has been studied. In the Douro Valley, this type of research was to be done. The aim of this research was the analysis of the influence of weather variability and climate trends on the variability of Vintage Port quality and yield, identifying how each weather factor, at each moment, improves or degrades the probabilities of a vintage becoming a high quality or high yield vintage at the end of the season. Vintage Port was used to assess vintage quality as it is exclusively produced in the Douro Valley and most vintages are rated in several renowned vintage-charts, for the period beginning at the early 80s until 2009.

This research introduced differentiating approaches to the analysis of the influence of the weather variability on vintage quality and yield. A consensus ranking was proposed as an impartial and unbiased measure of vintage quality rather than the ratings of a single wine expert / tasting panel. We proposed the definition of heat related variables using a partition of the growing season based on the heat accumulation that triggers each phenological event, instead of commonly used calendar dates. We used meteorological daily data of temperatures and precipitation amount, collected at several weather stations located within the Douro Valley instead of gridded interpolated global data. Links of weather variability and climate change to the region's economy were analyzed through the impact of quality and yield on the retail and release prices of Port wine.

Logistic regression was used to model the probability of a vintage being a high quality or a high yield vintage. To validate that the weather variables included in logistic models had influence on the vintage outcome, variable values from top ranked vintages were compared to corresponding values from bottom ranked vintages. Distinctive weather patterns were found for high quality and for high yield vintages. Trends in the region's temperatures were detected showing moderate relation to quality and no relation to yield. A strong association between retail prices of Vintage Port and experts opinions was found.

[1] There are no available data regarding Port wine yield. In this research yield refers to all types of wines in the Douro Valley.

Resumo

Corte-Real de Sousa, António Carlos. **Analyzing the Influence of the Douro Valley Weather on the Quality and Yield of Vintage Port**. Universidade do Porto, Setembro de 2014. Orientador: José Luís Cabral de Moura Borges.

O Douro é uma conhecida região de produção de vinho. O vinho do Porto é produzido na região desde o final do século 18, sendo o fator económico mais importante da região. As características climáticas determinam o tipo de uvas que podem ser cultivadas numa região enquanto as especificidades meteorológicas influenciam a maturação das uvas e a qualidade do vinho de cada colheita. Em várias regiões vinícolas, pesquisas anteriores estudaram a influência da variabilidade do tempo no resultado das colheitas, identificando fatores climáticos relacionados com a qualidade e com a produtividade. Além disso, a relação das tendências climáticas de longo prazo com a evolução da qualidade e produtividade tem sido estudada. No Vale do Douro, este tipo de pesquisa estava por fazer. O objetivo desta investigação foi a análise da influência da variabilidade do tempo na variabilidade da qualidade e produtividade[2] das colheitas, identificando como cada fator meteorológico, em cada momento, aumenta ou diminui a probabilidade de vir a acontecer uma boa colheita, no final da temporada. O Porto Vintage foi usado neste trabalho para a avaliação da qualidade das colheitas pois que é exclusivamente produzido na região do Douro e pelo facto de ter a maioria das colheitas classificadas em vintage-charts de renome, desde o início dos anos 80 até 2009.

Nesta investigação, introduzimos algumas abordagens inovadoras. Propusemos um ranking de consenso como uma medida imparcial da qualidade da colheita, em vez das classificações de um único especialista em vinho / painel de provadores. Na definição de variáveis meteorológicas relacionadas com o calor, propusemos uma divisão da estação de crescimento em intervalos baseada na acumulação de calor que determina o desencadeamento de cada acontecimento fenológico, em vez de datas de calendário. Foram usados dados meteorológicos diários de temperaturas e precipitação, obtidos em várias estações meteorológicas da região, em vez de dados meteorológicos globais interpolados para uma grelha geográfica. Ligações entre a variabilidade do tempo e a alteração do clima com a economia da região foram analisadas através do impacto da qualidade e produtividade das colheitas do vinho do Porto nos preços de retalho e nos preços de entrada no Mercado.

Utilizou-se regressão logística para modelar a probabilidade de uma colheita ser de elevada qualidade ou rendimento. Para validar que as variáveis meteorológicas incluídas nos modelos logísticos têm, de facto, influência na qualidade ou no rendimento das colheitas compararam-se os seus valores nas melhores e piores colheitas. Distintos padrões meteorológicos foram encontrados para os anos com colheitas de elevada qualidade e para os anos com colheitas de elevado rendimento. Foram verificadas tendências de aquecimento nas temperaturas da região que revelaram influência moderada na qualidade e nenhuma influência na produtividade das colheitas. Foi identificada uma forte associação entre os preços de retalho dos Porto Vintage e as classificações das colheitas emitidas por provadores e especialistas de renome.

[2] Não há dados correspondentes rendimento exclusivo do vinho do Porto. Neste trabalho referem-se os rendimentos correspondentes a todos os tipos de vinhos.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADVID	Associação para o Desenvolvimento da Viticultura Duriense
BBR	Berry Bros & Rudd Website
DC	Decanter Magazine
GDD	Growing Degree Days
IM	Instituto de Meteorologia de Portugal
INE	Instituto Nacional de Estatística
IPMA	Instituto Português do Mar e da Atmosfera
IVDP	Instituto dos Vinhos do Douro e do Porto
MAD	Median of the Absolute Deviations
MB	Michael Broadbent
NAO	North Atlantic Oscillation Index
OD	Ordinal Date
OLS	Ordinary Least Squares
PA	Precipitation Amount [mm]
QM	Quantile-Matching Adjustment
r	Correlation Coefficient
RMSE	Root Mean Square Error
SNIRH	Sistema Nacional de Informação de Recursos Hídricos
SOI	Southern Oscillation Index
SWE	Sotheby's Wine Encyclopedia
T_{agv}	Estimate of Mean Temperature [ºC]
T_{max}	Maximum Temperature [ºC]
T_{min}	Minimum Temperature [ºC]
VT	Vintages.com Website
WA	Wine Advocate Magazine

WE	Wine Enthusiast Magazine
WS	Wine Spectator Magazine

Glossary

CHANGE POINT: a moment, in a time series, where a sudden change of the mean or slope happens.

CLIMATE CHANGE: according to the Intergovernmental Panel on Climate Change (IPCC), climate change is a change in the state of the climate that can be identified (e.g., by using statistical tests) by changes in the mean and/or the variability of its properties and that persists for an extended period, typically decades or longer.

COMPLETE RANKING: a ranking where all the alternatives are ranked in order (without ties) by all judges.

DIGITIZE: to transcribe data into a digital form so that a computer can directly process it.

FULL RANKING: the same as complete ranking.

GDD (Growing Degree Days): sum of the average day temperature above a baseline temperature from start date to a given date.

HETEROSCEDASTICITY: a situation where the error terms of an explanatory variable in a regression model do not have constant variance across observations.

HOMOSCEDASTICITY: a situation where the error terms of an explanatory variable in a regression model have constant variance across observations.

LAPSE RATE: rate of change in temperature observed while moving upward through the Earth's atmosphere.

MAD: **M**edian of the **A**bsolute **D**eviations from the data's median is a robust measure of the variability.

METADATA: according to World Meteorological Organization (WMO), metadata are descriptive data necessary to allow us to find, process and use data, information and products. Metadata are data about data and should provide detailed information necessary for users to gain adequate background knowledge about the data.

ORDINAL DATE: a number that identifies the day-of-year, ranging from 1 to 365 starting January 1.

ORGANOLEPTIC: relating to perception by a sensory organ.

PARTIAL RANKING: a ranking where one or more judges do not specify completely their preferences, allowing some alternatives to be equally preferred (tied). Partial rankings are rankings where ties are allowed.

PHENOLOGY: a segment of ecology focusing on the study of periodic plant and animal life-cycle events that are influenced by climate and seasonal change in the environment.

QUALITY CONTROL (of meteorological data)**:** set of procedures used to detect erroneous observations.

R: a language and environment for statistical computing and graphics available as free software. R provides a wide variety of statistical and graphical techniques.

RELEASE PRICE OF A WINE: price of a wine when it is first released to the market.

RMSE: **R**oot **M**ean **S**quare **E**rror is a measure of the differences between values predicted by a model or an estimator and the values actually observed. RMSE is a measure of the "average" error, weighted according to the square of the error.

VINTAGE-CHARTS: tables where overall vintage quality ratings are presented by region and type of wine.

Chapter 1

Introduction

This thesis describes the subjects and procedures involved in the research of the weather factors main responsible for a good vintage of Vintage Port or for a high yield vintage. Vintage Port is produced in the Douro Valley and is widely considered the top quality style of Port wine. This work made use of data on Vintage Port quality and yield to search for links with the region's meteorological data on temperature and precipitation. The research focused on the Douro Valley, in the period from 1980 to 2009.

A better knowledge of the weather factors that influence the probability of a vintage being of high quality or high yield is vital for the definition of procedures that may mitigate in the present and in the future, some of the negative effects of weather variability and of climate change. This knowledge will have impact on the economy of wine producers and will be potentially invaluable for the economic sustainability of the Douro Valley.

In this chapter, we give some context information, we present the motivation for this research, we present the research objectives, and we pose the research questions. Additionally, we present a brief description of the methodologies used in the research, a summary of the conclusions and list the publications that arose from the work.

1.1 Context

From the second century AD until now, France dominated the world wine market. In the period from 1150 to 1700, with some interruptions, Britain was the most important client for French wines. From 1703 to 1860, tariffs restricted the French wines import by Britain. This restriction favored the establishment of wine trading from Portugal to Britain and gave the opportunity for the English people to discover Douro Valley wines.

Difficulties in maintaining Douro wine quality during the ship journeys from Porto to England led to the addition of brandy to the red wines in order to stabilize the wine, fact that originated the early days' Port wine.

The Douro Valley is a wine-producing region, situated inland in the northern portion of Portugal (Figure 1), distant 100 km from the Atlantic Ocean. The region is well known for the production of Port wine, long considered one of the best wines in the world. Presently, the region is also producing high quality table wines, some of which have been top rated in renowned wine magazines.

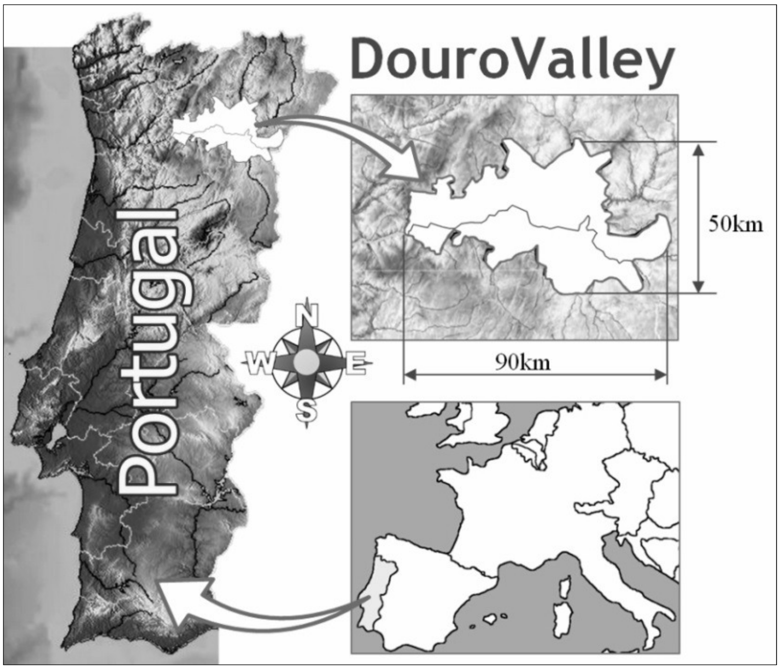


Figure 1 - Portugal and Douro Valley location (relief maps: www.maps-for-free.com).

Agriculture and tourism are the backbone of the economy of the Douro Valley. Grapes, olives, almonds, figs, oranges and cherries are the agricultural leading products of the Douro Valley (Aguiar et al. 2001). Grapes and olives are the two most important cultures (Rebelo et al. 2001). The Douro Valley is a UNESCO World Heritage Site and is becoming an important wine tourism destination.

Wine is the major economic commodity of the Douro Valley, determining the income of locals that work in agriculture producing wine or producing grapes for making wine. Additionally, wine has an indirect influence on the income of many local people who work in the industry of making and bottling wine, in the wine tourism sector, and in many tourism dependent activities. In 2011, the global production of all types of wine in the Douro Valley was 1.32 Mhl of which, 45 % was Port wine. The sales of Port wine in 2011 were EUR 353M representing 78% of the EUR 450M sales from all types of wines in the region. Vintage Port is the top, most exclusive, quality Port accounting for just 0.8% of the total Port wine production, but representing 3.7% of Port wine sales in value. In 2011, the average price of Port wine, excluding Vintage Port, was 434 €/hl while Vintage Port average price was 1854 €/hl (source IVDP: www.ivdp.pt).

The production of olives, the second most important agricultural product of the Douro Valley, was 90 Mkg in 2011, generating revenue of EUR 30M (source: *Associação de Olivicultores de Trás-os-Montes e Alto Douro*) representing 7% of the revenue derived from wine trade.

This research was not able to find institutional information on values that might give a precise indication of the importance to the Douro Valley of the wine tourism industry. However, as the revenue generated by olives, the second agricultural product of the region, represents only 7% of the revenue generated by wine, we believe it is safe to estimate that wine production and wine tourism represent together more than 80% of the global revenue of the region.

Climate characteristics determine the type of grapes that can be grown in a region and the types of wines that can be produced, while the weather specificities influence the ripening of grapes and the quality of the wine of each vintage. Climate change[3] influences both weather variability and weather averages having influence on the evolution of vintage-to-vintage quality / production and on the suitability of a region to produce certain styles of wine. Weather variability affects the vintage-to-vintage quality and production.

There are many signs that climate is changing globally: sea level rise, global temperature rise, warming oceans, shrinking ice sheets, declining Artic sea, glacial retreat and extreme events (NASA 2013). In the Douro Valley, signs of climate change are also evident with higher minimum and maximum temperatures, increase in extreme temperatures, fewer cold events, more stress events, and lower temperature range. Climate projections for the region predict changes in precipitation and temperature (Jones & Alves 2012). Climate trends may have an important impact on Douro region's

[3] A change in the state of the climate that can be identified (e.g., by using statistical tests) by changes in the mean and/or the variability of its properties and that persists for an extended period, typically decades or longer (Intergovernmental Panel on Climate Change, http://www.ipcc.ch/)

economy demanding that all efforts should be made to keep viticulture and winemaking profitable in this region.

1.2 Motivation

Despite the effort and commitment of winegrowers to yearly produce wines of the best quality, quality varies from one year to another depending on a number of controllable and uncontrollable factors. The economic relevance of the winemaking sector in many regions around the world has made the relation between the variability of the yearly weather factors and the variability of wine quality / production an important scientific research subject. Winkler et al. (1974), Corsi & Ashenfelter (2001), Nemani et al. (2001), Jones et al. (2005), Grifoni et al. (2006), Ashenfelter (2008), Makra et al. (2009), Gladstones (2011), Parker et al. (2011), Santos & Malheiro (2011), and numerous other researchers have conducted research on the relations between the meteorological variability and the annual quality and yield of the wines of a region. Their results show that temperature and precipitation are the most important factors on a vintage quality and yield. Santos and Malheiro (2011) has shown that the overall production of the vintages in the Douro Valley varies due to weather variability.

Research on the influence of the weather on wine quality and yield was conducted for several wine regions around the world. However, as far as we are aware, no research on this subject focused on the Douro Valley regarding the quality of Port wine, the most important economic factor in the region. It is important for the Douro Valley that a detailed research on the relation between the variability of the Douro Valley weather and the variability of vintage quality and yield of Port wine is conducted. A better knowledge on how weather factors historically influenced Port wine quality and yield, will highlight which factors are more influential on quality and on yield. While uncertainty exists on how climate will change in the region, such knowledge will bring insight on the definition of strategies that may help to mitigate negative influence of these factors in the future. Our research aims to fill this knowledge gap.

The process of finding an adequate measure of vintage quality is a challenging task due to inherent subjectivity in assessing quality. One option is to use the yearly vintage-charts[4] published by internationally recognized critics, magazines, or organizations, which compare and contrast wines from different properties, different regions, and/or different vintages. Vintage ratings have been used in numerous studies examining a wide range of economic, consumer, and scientific topics. The analysis of vintage charts reveals that there is not a widespread consensus on the vintage quality of a given region over the years. Each publisher has its own tasting panel, with its own criteria and perception of quality, which tastes a different set of wines, at different times and conditions. In addition, a variety of rating scales are used. The difficulty of combining the judgment of several vintage charts is even bigger when it is observed that some publishers use the same rating scale, but with different criteria. We proposed the use of a rank aggregation method to combine a collection of vintage chart ratings into a ranking of the vintages that represents the consensus of the input vintage charts (Borges et al. 2012). The method takes advantage of the information available from a set of independent sources and combines it into an impartial ranking of a region's vintages over the years. The resulting ranking provides a relative measure of a given region's vintage quality.

Most past and ongoing research on the influence of weather variability in wine quality and yield uses gridded interpolated global meteorological data. This research used meteorological daily data of temperatures and precipitation amount, collected at several weather stations located within the Douro Valley. In order to obtain a high quality meteorological dataset that would allow an accurate characterization of the region's climate and the analysis of climate trends, the raw dataset was first cleaned of erroneous values using the methodology proposed by Feng, Hu, & Qian (2004) and homogenized using the methodology proposed by Wang (2011).

In order to study the weather yearly profiles and their relation with the vintage quality and yield it is necessary to split each year growing season into smaller growth intervals in which the weather variables may be evaluated and compared to the corresponding values from other years. Some studies partition the grapevine growing season into smaller intervals using calendar defined weeks or months (e.g., from week *x* to week *y* or from March to June). Other studies partition the season into intervals whose bounds are defined using a calendar simplification of plant phenology, making

[4] Vintage-Charts are tables where overall vintage quality ratings are presented by region and type of wine, for a time period of several years.

use of accepted dates when, on average, the main phenological events happen in a region. C. Real et al. (2014) analyzed the differences in temperature and precipitation when using growth intervals with boundaries defined by historical dates of the main phenological events (used as reference) and growth intervals with boundaries defined by two methods: method 1 - by mean values of the heat requirements of the main phenological events and method 2 - by generalized calendar average dates associated with the occurrence of the main phenological events. The results showed that when there are no records on the actual dates of the phenological events, the best option is to use accumulated heat (growing degree-days) to determine the growth events and intervals between events. In this research, weather variables and indexes were defined using intervals whose boundaries are related to winegrape phenology[5].

These are the differentiating aspects of this research.

1.3 Research Objectives

This research intends to give new insight into the current knowledge on the influence of climate trends on the evolution of the quality and yield of wine vintages and on the influence of weather variability in vintage-to-vintage quality and yield. New approaches are necessary to analyze the climate / weather relation to vintage quality and yield. We seek to bring novelty by using an unbiased and impartial measure of vintage quality, by using high quality meteorological data collected at local stations, and by characterizing the weather using variables defined in time intervals bounded using heat accumulation.

This research was designed with the purpose of guiding us through the process of answering to the following research questions:

Research question 1

How to assess vintage quality? Using one expert's opinion published in wine magazines as vintage-charts ratings or consensualizing several experts' opinions? How to consensualize different ratings expressed in different rating scales?

[5] Phenology is a segment of ecology focusing on the study of periodic plant and animal life-cycle events that are influenced by climate and seasonal change in the environment (Leopold 2013)

Most past and ongoing research on the relation between weather and vintage quality uses the ratings of a single expert or expert's panel. Wine tasting is a sensory experience. It is based on personal tasting skills, culture, memory, and fashion. Several tasters will have different opinions about the same wine. A new method to assess vintages through an impartial measure that aggregates the opinions of several tasters is needed.

Research question 2

What weather dependent variables are responsible for most of the variability in quality of Vintage Port vintages? Is there a weather profile that enhances the likelihood of a vintage becoming a high quality vintage?

The outcome to this research question will be the identification of the most significant variables that explain vintage quality. Weather variables and indexes showing influence on the quality of a vintage of Vintage Port vary accordingly to methodology used in each model. Results from different methodologies will be used for validation of most influential explanatory variables. Weather variables will be defined based on grapevine phenology in order that their meaning is not only statistically and mathematically significant but also that makes sense in terms of grapevine physiology.

Research question 3

What weather dependent variables are responsible for most of the variability in yield of Vintage Port vintages? Is there a weather profile that enhances the likelihood of a vintage becoming a high yield vintage?

The outcome to this research question shares a very common path with research question two but with the focus on vintage yield, instead of vintage quality.

Research question 4

What is the best way to partition the growing season into smaller growth intervals? Using calendar dates (e.g., mean temperature in the May 15 to May 31 period) or using grapevine phenology (e.g., mean temperature from budburst to flowering)?

The dates of phenological events vary from one year to the next. In 1980-2009, in the Douro Valley, the budburst date ranged from March 1 to April 7. When defining a weather variable using calendar (e.g., mean temperature in March) that variable may in one year correspond to a certain development stage of the plant and in the following year correspond to a different development stage. Is this fact important? The answer to this question is of great importance for this and for future research on this area.

Research question 5

Is there evidence that the Douro Valley temperature, precipitation, and extreme temperatures, showed climate trends in 1980-2009? To what extent was the evolution of the quality and yield of the vintages influenced by climate trends?

As a scientific outcome of this research question, it will be possible to assess the type and the intensity of the trends in the Douro Valley climate, using meteorological data collected at weather stations located inside the region. This information will allow us to evaluate if the evolution of vintage quality and yield in 1980-2009 is related to climate trends.

Research question 6

Are retail prices of Vintage Port and release prices of Port wine related to the quality and yield of the vintages?

The outcome to this research question will provide information on the impact that the yearly weather profiles have on the Douro Valley economy by influencing the retail prices of Vintage Port and the release prices of Port wine. Association between retail prices and vintage quality / yield as well as between release prices and vintage quality / yield will be analyzed in order to answer this question.

1.4 Data

Reliable data on weather, grapevine phenology as well as wine production, sales, revenue, prices, vineyards area and yields are generally sparse for the Douro Valley region, except for the years after year 2000. This research was able to gather data that characterizes the Douro Valley weather and climate and the quality of the vintages of the vintages of Vintage Port. Additionally, data on the annual average release prices of Port wine were collected. Data on the annual average yield for the Douro Valley wines and on Vintage Port retail prices were estimated using the available data.

The quality of the vintages of Vintage Port was characterized using the ratings of several vintage-charts issued by renowned tasters and experts in wine magazines. Vintage-charts cover the main wine regions in the world. Ratings from Vintage-charts are widely used as references to assess wine quality. For Vintage Port, several renowned vintage-charts have most vintages rated for the period beginning at the early 80s until the present moment. Ratings on table wines may also be found for vintages after 1995 but from wines produced at different regions in Portugal and consequently do not allow the assessment of the overall wine quality for a single region. In 2009, at the beginning of this research, we were able to collect ratings for Vintage Port, from several vintage-charts, for a 30-year period (1980-2009) that we considered the minimum time span adequate to be used as the quality sample in this research. Even if vintage ratings would exist for most vintages before 1980, the use of a larger period, beginning before 1980, would incorporate in the ratings for the yearly quality of the vintages the variability caused by the mechanization in the vineyards and cellars that was introduced in the Douro Valley during the 1970s.

The World Meteorological Organization (www.wmo.int) also requires the calculation of averages for consecutive periods of 30 years to describe the climate of a region. For this reason, the 30-year period (1980-2009) used to characterize vintage quality, was also convenient to characterize the region’s climate. With the objective of characterizing the region’s yearly weather variability, meteorological daily datasets were collected from eight meteorological stations located within the Douro Valley. Daily datasets of maximum temperature, minimum temperature, and precipitation amount were collected from five weather stations belonging to the *Instituto de Meteorologia de Portugal* (IM) and from three weather stations belonging to *Sistema Nacional de Informação de Recursos Hídricos* (SNIRH).

Data on release prices for all Port wine types were available from two different sources, Cunha (2001), for all years in 1980 - 2001 and IVDP, for years after 2005. Estimates of current average retail prices for Vintage Port vintages in 1980-2009 were obtained averaging individual prices of Vintage Ports from 290 merchants in the UK, 265 in the USA and 624 in non-UK Europe, collected from Wine-Searcher (www.wine-searcher.com)

To assess the wine production's variability we used wine yield. We believe that wine yield is more adequate than wine production, which depends on the area of planted vineyards, which is not dependent of weather variability. Yield is a measure of the amount of grapes or wine that is produced per unit surface of vineyard. Data characterizing the Douro Valley wine production is available from the *Instituto Nacional de Estatística*, INE. In order to estimate the yearly average yield for the region it was necessary to have data on the planted area of vineyards in the Douro Valley, between 1980 and 2009. The only data on vineyard area refers to the years 1982 and 2010. As from 1982 to 2006 several financial programs were put in place for planting and restructuring vineyards in the Douro Valley, we added the vineyard area in 1982 to new vineyard areas. Making use of the data on wine production, together with the estimates of the annual area of vineyards, it was possible to estimate the annual wine yield for the region.

1.5 Methods

As most factors that influence the yearly wine quality and yield are stable over the years (grapes varieties, sites location, soils, the cultural practices in the vineyards, and the wine making process), the variability of the vintages of Port must be related to the only factor that changes in a year-over-year basis, the weather, and to its long-term evolution, climate-trends. Together with the weather variability factors, other factors have influence on vintage quality and yield such as weather extremes and vine diseases. These other factors usually have an influence that is limited in geographic and temporal scope, as they only produce effects for short periods and on small sites of the whole region. Data on most of these factors are not available for the Douro Valley.

We focused this research on weather factors that are related to temperature and to precipitation as these variables vary in a similar manner all over the region, in a year-over-year basis.

The yearly variability of the weather factors that are related to temperatures affects the whole region in a similar manner (the Douro Valley is a small region of 90 km × 50 km). Although temperatures may be different in sites located on different sub-regions inside the Douro Valley, their variability in a year-over-year basis is similar all over the region as a very hot summer in a particular site also will be a very hot summer in any other site within the region, independently of the temperature values in each site. The precipitation amount, although with a higher geographic and temporal variability than temperature, also has a similar basic pattern all over the region if we are focused on average values. The variability of average precipitation amounts, in a year-over-year basis, is similar all over the region as a very rainy winter in a particular site also will be a very rainy winter in any other site within the region, independently of the amount of rain in each site.

This research focused on the Vintage Port style of Port wine, although in some analysis data from Port wine (includes all styles of Port) are used. As the soils where grapes grow, the grape varieties, the yearly weather, and the wine making process are similar for all styles of Port, we believe that it is reasonable to assume that the quality of Port wine is highly correlated with the quality of Vintage Port.

Data Preparation Methods

In this research, in order to identify and clean the meteorological dataset from erroneous values we used of the well-established methodology proposed by (Feng et al. 2004). In order to homogenize the cleaned dataset the RHtestsV3 software package (Wang 2011) was used. After cleaning and homogenizing the meteorological dataset, we obtained a high-quality meteorological dataset that allowed an accurate characterization of the Douro Valley climate and the estimation of regional's temperature and precipitation change rates.

In order to obtain an unbiased and impartial measure of vintage quality we devised a new method that converts the ratings of each individual source into rankings and uses a rank aggregation algorithm to combine the input ranking into a consensus ranking (Borges et al. 2012). The ranking represents an impartial consensus of the collection of input vintage charts as it effectively incorporates input from numerous publishers. Different types of scoring formats and of different scales are allowed to the scoring of the vintages.

Based on the conclusions of C. Real et al. (2014), for the definition of the weather variables we partitioned the grapevine growing season into smaller growth intervals. The boundaries of each growth interval were defined by calculating for each year the dates when the historical accumulated heat necessary to trigger each phenological event was reached. The defined variables were used as predictors in regression models for vintage quality and for vintage yield as well as in the analysis of top ranked vintages comparisons to bottom ranked vintages.

Data Analysis Methods

Logistic regression was used as a technique to explore association between weather explanatory variables and vintage quality and yield, capturing the most influent variables and allowing the assessment of their relative importance. Binary logistic regression models for top n vintages (in quality) *vs* remaining vintages, bottom n vintages (in quality) *vs* remaining vintages, top n vintages (in yield) *vs* remaining vintages, and bottom n vintages (in yield) *vs* remaining vintages, are presented. Variables captured in each model depend on the number of observations included in each class (n for class Y = 1 and $30 - n$ for class Y = 2). To overcome this issue we repeated the technique for n values in the range $6 \leq n \leq 10$ that correspond to a top vintage class containing 20.0% to 33.3% of the 30 vintages.

Analysis of top *vs* bottom ranked vintages was conducted in order to identify differences between weather variables' medians of top vintages and bottom vintages. As some variables were not normally distributed and other exhibit heteroscedasticity, Mann-Whitney test was used. The same type of analysis was conducted in order to identify differences between phenology variables' averages of top vintages and bottom vintages. As phenology variables were normally distributed and homoscedastic, t-test was used.

The Douro Valley climate between 1980 and 2009 was analyzed to identify trends in annual mean temperature, growing season mean temperature, annual precipitation amount, growing season precipitation amount, annual number of days with maximum temperature above 36 ºC, and annual number of days with minimum temperature below -2 ºC. The evolution of vintage quality and yield in 1980-2009 was compared to climate evolution, searching for links.

Association between vintage quality / yield and retail prices was analyzed using Spearman's rank test.

Association between vintage quality / yield and market release prices was subjectively analyzed, comparing the evolution of average release prices in 1980-2009 with the evolution of vintage quality / yield in the same period.

1.6 Thesis Structure

Although this is not a cumulative thesis[6], some research questions will be answered based on published or submitted papers. These papers will not be included as published / submitted but adapted to the thesis structure and information flow, including supplementary information.

In the remainder of this section, we briefly point out the subjects and contributions dealt with in each chapter.

In **Chapter 2**, we review the literature on several different subjects addressed in this research:

- modelling vintage quality / yield;
- grapevine phenology;
- wine quality;
- aggregation of opinions;
- quality control of meteorological datasets;
- temperature lapse rates.

In **Chapter 3**, we present background information about the Douro Valley characteristics, about Port wine, and about the main factors that are known to affect the quality and yield of Vintage Port.

In **Chapter 4**, we present information on the methodologies for quality control of meteorological datasets, we present the meteorological dataset for the Douro Valley and the application of the cleaning and homogenization procedures to this dataset.

[6] A thesis comprising a collection of individual peer reviewed papers integrated in a bound copy.

Chapter 5 is a chapter based on the research publication "**A New Method to Obtain a Consensus Ranking of a Region's Vintages**" (Borges et al. 2012)**.** The method combines a set of input vintage chart ratings for a wine region into a ranking of the vintages over the years. The combined ranking gives an ordering of the vintage quality that represents the consensus of the ratings given by the set of publishers' vintage chart. The method is an optimization method for aggregating opinions based on the minimization of a metric that represents the overall distance from the rankings that correspond to the vintage chart ratings to a ranking candidate to be consensual.

Chapter 6 is a chapter based on the research publication "**Partitioning the Grapevine Growing Season in the Douro Valley of Portugal: Accumulated Heat better than Calendar Dates**." (C. Real et al. 2014). Different alternatives for the partitioning of grapevine growing season are compared: i) historical dates of the main phenological events for the region, ii) calendar dates, and iii) mean values of the heat requirements of the main phenological events. Background information on the grapevine annual life cycle, the main phenological events, and the growth stages of grapevine is provided. An analysis is presented showing that phenology is the better method for defining variables and that, when phenology dates are not available, the definition of variables should be based on the accumulated heat values that, on average, trigger the phenological events rather than on calendar dates.

In **Chapter 7**, we present the analysis methodologies that were used to i) relate climate trends to the evolution of quality and yield of the vintages and ii) to analyze the influence of the yearly weather on vintage quality and yield.

Before presenting these methodologies, we define a set of weather variables and a set of phenology variables, and indicate the preliminary statistical tests used to analyze variables. The knowledge of the statistical characteristics of the variables was important to the definition of the analysis methodologies and to the selection of variables that would be appropriate for each methodology.

Before conducting climate trends analysis, the meteorological data were used to characterize the weather of the Douro Valley in 1980-2009. Climate trend analysis was conducted and trends were observed in temperatures and precipitation. Correlation analysis was conducted to assess association between climate trends and the evolution of the quality and yield of the Port wine vintages in 1980-2009.

Analysis of the relation between Vintage Port retail prices and vintage quality / yield is conducted using correlation analysis and analysis of the relation between Port wine release prices and vintage quality / yield is conducted using a subjective analysis of the evolution of the release prices, vintage quality and vintage yield, in 1980-2009.

Analysis of the influence of yearly weather on vintage quality and yield is conducted using several techniques:

- Logistic Regression is used to model the probability of a vintage being a high quality or a high yield vintage;
- Top *vs* bottom Ranked Vintages (analysis of weather variables): the Mann-Whitney test is used in the analysis of differences of the medians of the weather related variables, between top *n* vintages and bottom *n* vintages;
- Top *vs* bottom Ranked Vintages (analysis of phenology variables): the t-test is used in the analysis of differences in mean values of the phenology variables, between top *n* vintages and bottom *n* vintages.

The results of the analysis of the influence of weather variability on vintage quality and on vintage yield are presented, showing that top and bottom vintages of Vintage Port, in terms of both quality and yield, have different yearly temperature and precipitation profiles. Weather profiles that the analysis results show to enhance the likelihood of a vintage being a high quality or a high yield vintage are characterized.

Analysis results and conclusions are checked using the weather, quality and yield data. The level of agreement between each yearly weather profile and the weather profiles that enhance the likelihood of a vintage being a high quality or a high yield vintage data, is calculated as a score on a 100 point scale. Scores for yearly quality and yield are compared with the corresponding measures of quality and yield of the vintages of Vintage Port, showing a significant association.

In **Chapter 8**, the main findings with regard to the research questions are summarized. Results are discussed and compared with results obtained from other research.

In **Chapter 9**, we present the main conclusions of the research summarizing the main contributions and finally, we present some suggestions for future work as a guideline on how the research could be extended and as recognition that this research is not the last word on the subject.

1.7 Summary of Results

The results of the analysis of the influence of weather variability on vintage quality and on vintage yield showed that the top and bottom vintages of Vintage Port, in terms of both quality and yield, have different yearly temperature and precipitation profiles. Furthermore, results made evident that the average retail prices of Vintage Port are highly correlated to vintage quality, showing no correlation to vintage yield. Moreover, results show that the overall quality of Vintage Port has sustainably increased in the last 35 years but that Port wine release prices, at constant prices, decreased between 1980 and 2009. Finally, results show that the climate of the Douro Valley showed trends in between 1980 and 2009, as the increase in the annual mean temperature and in the growing season (April-September) mean temperature, the decrease in the number of days with minimum temperature below -2 ºC, and the decrease the precipitation amount during the growing season. Climate trends showed no relation with the evolution of wine yield and only a moderate relation to the evolution of vintage quality.

1.8 Publications Arising from this Thesis

Three publications arose from this research. The first publication in the list was published in an international journal and the last two were submitted and are in a reviewing process:

- Borges, J., Real, A. C., Cabral, J. S., & Jones, G. V. (2012). A New Method to Obtain a Consensus Ranking of a Region's Vintages. Journal of Wine Economics, 7(01), 88–107.

  This article originated a comments article (Hulkower 2012). A response to the comments article was provided through the following article:

  Borges, J., C. Real, A., Cabral, J. S., & Jones, G. V. (2012). Condorcet versus Borda, a response to: Comment on "A New Method to Obtain a

Consensus Ranking of a Region's Vintages' Quality"**.** Journal of Wine Economics, 7(02), 245–248.

- C. Real, A. et al., 2014. Partitioning the Grapevine Growing Season in the Douro Valley of Portugal: Accumulated Heat better than Calendar Dates. International Journal of Biometeorology, (forthcoming).
- C. Real, A., Borges, J., Cabral, J. S., & Jones, G. V. (2014). Weather Influence on Quality, Yield and International Market Prices of Vintage Port. (in revision process).

A research paper was presented at an international meeting:

- Corte-Real, A., Borges, J. & Cabral, J.S. (2013). Influence of the Characteristics of Weather of Douro Region on Port Wine Vintages' Quality. In 7th Conference of the American Association of Wine Economists. Stellenbosch, South Africa.

Chapter 2

Literature Review

A review of the literature was conducted in order to establish the significance of the area of research and to identify where and how a new contribution could be made. This chapter provides a summary of those topics considered most relevant to the research problem. It begins with the review of the methodologies used in several models of wine quality. The review of this subject, as working basis to this research, was conducted in a more detailed manner than the review of the literature of the remaining subjects. Next, a review on grapevine phenology is presented. Then a review on the definition of wine quality is followed by a review on the methodologies to achieve consensus opinion. As weather datasets were used, we reviewed the main methodologies used in quality control of meteorological datasets as well as the main homogenization procedures. Finally, we briefly reviewed methodologies that relate the rates of change in temperature with elevation as the Douro Valley is a mountain region and the elevation of the weather stations range from 65 m (*Régua* station) to 715 m (*Carrazeda de Ansiães* station).

In sections based on published / submitted papers, specific literature review is presented locally. All the literature references in the following sub-sections are chronologically ordered.

2.1 <u>Modeling Vintage Quality and Yield</u>

Corsi & Ashenfelter (2001) studied how the weather determines the quality of wines produced in the Barolo and Barbaresco regions of Italy. This study used data of the 1970-1997 period.

<u>Vintage / Wine Quality Assessment</u>: wines quality was assessed using expert ratings. Vintage ratings from three sources were used: Gambero Rosso, Robert Parker

and Pauline Wasserman's. Meteorological Dataset: monthly temperature and precipitation data in 1980-1997 were collected from the weather station in Castiglione Falletto, belonging to the Regional Service. Castiglione Falletto was chosen as the most representative of the Barolo region weather. Monthly temperature data between 1970 and 1980 for the region of Barolo were estimated through linear regression, using data from the weather station in Cuneo. Monthly precipitation data between 1970 and 1980 for the region of Barolo were estimated through linear regression, using data from the weather station in La Morra. Methodology: as the assessment of vintage quality was achieved using an ordinal scale, this research used a ordinal PROBIT regression to model the quality rankings based on four explanatory variables. The used explanatory variables were the total rainfall between October and March of the season preceding the vintage, the average monthly temperature in the period March to July, the total rainfall in August and September and the average temperature in August and September. Conclusions: results showed no significant association between expert's quality indexes and winter rain or average temperatures.

Jones, White, Cooper, & Storchmann (2005) studied the impact of climate change on viticulture and on quality wine production. This study used data of the period between 1950 and 1999.

Vintage / Wine Quality Assessment: wine quality was assessed using the Sotheby's vintage ratings for 27 wine producing regions in the world, covering 28 categories of wine made from the dominant *vitis vinifera* varieties grown in each region. When lacking a vintage rating for a region, the Sotheby's data were supplemented with ratings from the Wine Enthusiast Magazine. Meteorological Dataset: monthly mean air temperature for each wine region, for 1950-1999, was obtained using $0.5^\circ \times 0.5^\circ$ gridded climatology data produced from the Global Historical Climatology Network. To examine the potential future temperature changes in the wine regions, this research used a 100-yr run (1950–2049) of the HadCM3 coupled atmosphere-ocean general circulation model developed at the Hadley Centre. Methodology: The structure, variability, and trends of growing season average temperatures and vintage ratings were examined using descriptive statistics and regression analysis. Conclusions: from 1950–1999, growing season average temperatures have increased in the world's high-quality wine producing regions by 1.26 ºC. In the majority of regions, climate variations and trends were found to influence year-over-year variations and trends in vintage quality ratings: 10 to 60% of vintage ratings were explained by growing season temperature variations. Based on a quadratic econometric modeling approach, 12 of the wine regions were found to have an optimum growing season temperature above which vintage ratings tended to decline. While the observed warming of the late 20th century appears to have been mostly beneficial for high-quality wine production worldwide, this research suggests that the impacts of future climate change will be highly heterogeneous across varieties and regions. Critically, in some regions, warming may exceed the varietal specific optimum temperature threshold such that the ability to ripen balanced fruit from the existing varieties grown and the production of current wine styles will be challenging if not impracticable

Grifoni, Mancini, & Maracchi (2006) studied the relation between meteorological information freely available on Internet and the average quality of Italian wine. Weather variables temperature and precipitation were used. The presence of teleconnections and their effect on quality was investigated by considering other variables: 500 hPa geopotential height[7], sea surface temperature, and meteorological indices such as North Atlantic Oscillation and Southern Oscillation. This study focus on the period from 1970-2002.

Vintage / Wine Quality Assessment: vintage ratings were used to define wine-quality ranking, which was based on the collection of estimates from 1 to 5 classes. The rating of a given vintage was conducted in a single blind tasting by a panel of experts. Six wines, produced in central and northern Italy, were used: three wines were produced in the Tuscany region in central Italy, two wines were produced in the Piedmont region in northwestern Italy and one wine was produced in the Veneto region in northeastern Italy. The analysis used the average of the six quality rankings. Meteorological Dataset: air temperature, cumulated precipitation, 500 hPa geopotential height, and sea surface temperature (SST) were used. The North Atlantic Oscillation Index[8] (NAO) and Southern Oscillation Index[9] (SOI) were also used. All meteorological information was provided by NOAA-CIRES Climate Diagnostics Center (http://www.cdc.noaa.gov/).

[7] The geopotencial at a point is the work that must be done against the Earth's gravitational field to rais a mass of 1 kg from sea level to the point.Geopotencial high is used as a vertical coordinate of a point (altitude) referenced to mean sea level and representing the altitude necessary to reach the given pressure source: INMET – Brasil).

[8] NAO index is the difference of sea-level pressure between two stations situated close to the centres of Iclandic Low and Azores High in the North Atlantic region (http://ossfoundation.us/projects/environment/global-warming/north-atlantic-oscillation-nao). NÃO is a measure of the strength of the westerlies across the North Atlantic.

[9] SOI is a standardized index based on the sea level pressure differences between Tahiti and Darwin. SOI gives an indication of the development and intensity of El Niño or La Niña events in the Pacific Ocean (http://www.bom.gov.au/climate/glossary/soi.shtml).

Methodology: correlations between wine quality and air temperature, precipitation and several bioclimatic indices data were calculated over the wine production area. The correlations were calculated on a monthly to a multimonthly basis. Correlation maps between wine quality and 500 hPa geopotential height or SST were used to verify the possible impact of large-scale phenomena on wine quality. The presence of teleconnections between wine quality and monthly large-scale meteorological circulation was also investigated by means of the NAO and SOI indices. Conclusions: positive effect was observed for air temperature, confirming that wines of high-quality ranking were produced during warm years. The temperatures of May to October period are highly correlated to wine quality. Rainfall was inversely correlated with wine quality. The highest correlations were obtained using the last months of the season for September to October because of the importance of weather conditions during harvest period when fungal infections and sugar accumulation can be affected by intense precipitation. High quality wines were obtained during the years characterized by 500 hPa geopotential height above the average values, in particular for the April to July period. Wine quality rankings were inversely correlated with the NAO index and no significant correlation was found between wine quality and the SOI.

Ashenfelter (2008) studied the relation between the variability in the quality and prices of Bordeaux vintages and the weather that created the grapes. It was studied how the price of wines may be predicted from data available when the grapes are picked. This study used data of the 1952-2003 period.

Vintage / Wine Quality Assessment: wine quality was assessed using an index based on price. It is stated that " knowing the reputations of the 6 chateaux and the 10 vintages gives sufficient data to determine the quality of all 60". The price index was the average price for each vintage, calculated using the market prices of *Bordeaux* wines from several *chateaux* (the *chateaux* are deliberately selected to represent the most expensive wines: *Lafite, Latour, Margaux* and *Cheval Blanc*, as well as a selection of wines that are less expensive: *Ducru, Beaucaillou, Leoville Las Cases, Palmer, Pichon Lalande, Beychevelle, Cos d'Estournel, Giscours, Gruaud-Larose,* and *Lynch-Bages*). Meteorological Dataset: average temperature in April to September, precipitation in August, precipitation in September and precipitation in October to March prior to the vintage were the used variables. Data were collected from a single station in Merignac, a part of the Bordeaux region. Methodology: a multivariate regression of the prices of the wines on the weather variables was conducted using the age of the vintage and the weather variables as explanatory variables and the price index as independent variable. Conclusions: about 80% of the variation in the average price of Bordeaux wine vintages is explained by the four variables: the age of the vintage, the average temperature over the growing season (April-September), the amount of rain in September and August, and the amount of rain in the months preceding the vintage. Analysis of the effects of age alone produces a model that explains only slightly more than 20%, suggesting that the weather is an extremely important determinant of the quality of a wine vintage and its price at maturation.

Makra et al (2009) studied the effects of climatic elements on wine quantity and quality for the winegrowing region of Tokaj-Hegyalja, Hungary. It is unclear throughout this work if the study is about all types of wines from the region Tokaj, or if it is focused on the iconic Aszú wine. The study used data of the period between 1901 and 2004.

Vintage / Wine Quality Assessment: wine quality was assessed using the ratings of WGRIT – Wine Growing Research Institute of Tarcal, Hungary. The quality scores consist partly of wine quality characteristics and partly of the quantity measure of the so-called aszú berry production in the given year. Hence, the scores are comprised of both subjective (such as sensory quality ratings—aroma, flavor) and objective components (such as alcohol, sugar free extract, titrated acid, and citric acid content). Production data of the region were also collected (source not identified). Meteorological Dataset: monthly data of three climatic variables from April to September for the years 1901 to 2004: mean monthly temperature, monthly precipitation and monthly hours of sunshine. Data were collected from the meteorological station of Tarcal operating at the Ministry of Agriculture and Rural Development. The region of Tokaj is a small region (the longest distance between two points is 52 km) and Tarcal is located in the southern part of the region Tokaj. Methodology: three statistical models were used to assess the influence of 18 independent weather variables and two dependent variables on vintage quality and quantity: factor analysis, cluster analysis and variance analysis. Conclusions: the most important factors of wine quantity in the Tokaj region are hours of sunshine in May, June, July, and August and precipitation in September. Additionally, mean temperature, precipitation, and hours of sunshine in May and September play a basic role in wine quality, as does precipitation in July and hours of sunshine in August. The weather in September is very important for aszú wine production, whereby increased rainfall during this month leads to higher occurrence of *Botrytis Cinerea* and to higher sugar and flavor levels.

Santos & Malheiro (2011) studied the impact of projected climate change for the Demarcated Region of Douro, Portugal, on wine production. Statistically significant correlations were identified between annual yield and monthly mean temperatures and monthly precipitation totals during the growing cycle. Additionally, using ensemble simulations under the A1B emission scenario, projections for GYM derived yield in the Douro Region, and for the whole of the twenty-first century, were analyzed. This study used data of the period between 1986 and 2008.

Production Assessment: production was assessed using the total yearly production (all types of wine) for the Douro Valley. Meteorological Dataset: daily time series of precipitation and temperature, recorded at the meteorological station of Vila Real in the Douro Valley. For future climate (2001–2100) two integrations following the Intergovernmental Panel on Climate Change[10] A1B scenario were selected. Methodology: a multivariate linear regression model was adjusted to grapevine yield time series using the full set of selected potential predictors (monthly mean temperatures and monthly precipitation totals) in 1986–2008. A stepwise methodology was applied to select the most significant predictors. Conclusions: high March rainfall, high temperatures and low precipitation amounts in May and June favors yield. A slight upward trend in yield is projected to occur until about 2050, followed by a steep and continuous increase until the end of the twenty first century.

Mattis (2011) tried to find correlations between weather and wine quality in such a way that would allow the prediction of wine scores based solely on weather data. He analyzed what climate conditions were involved in the correlations and contributed to the wine scores, looking for specific weather patterns that could be linked to positive or negative effects on the quality of the wine. The study focused on the region of Sonoma in California, USA, using data of 1980-2007 period.

Vintage / Wine Quality Assessment: wine quality was assessed using the ratings from Wine Spectator magazine. Meteorological Dataset: daily maximum temperature,

[10] The Intergovernmental Panel on Climate Change (IPCC) was jointly established by the World Meteorological Organization (WMO) and the United Nations Environment Programme (UNEP) to assess the scientific, technical and socio-economic information relevant for the understanding of the risk of human-induced climate change.

minimum temperature and precipitation collected from 14 different stations covering Sonoma County and surrounding areas. Only weather data from April 1 through October 31 was used. In addition to the weather station data, two different derived data were calculated: daily temperature range and accumulated heat. Methodology: Scores and weather data were combined in data points for each region and year that had both. Each data point consisted of a region name, year, score and daily weather data. A custom software program was developed based on a genetic algorithm. The analysis conducted on the data was based on the concept of a weather period and combinations of different weather periods. A weather period consists of a period during the growing season, the type of weather data to be used for this period, and the type of calculation to be performed on this data. Daily meteorological data that can be chosen to apply the calculation are T_{max}, T_{min}, precipitation, temperature range and accumulated heat. Calculations that can be chosen to apply to the meteorological data are Maximum, Minimum, Average and Sum. Each weather period generates a single value for each of the regions and for each year. The algorithm makes combinations with one several weather periods and calculates, using linear regressions, the respective correlations with the quality the vintage for that region. Conclusions: positive factors for wine quality: high temperatures in April-May period and the absence of excessive temperatures during the summer. Too much heat during ripening is bad for wine quality as it would start to raisin and burn the grapes. A period was identified, lasting about one week around the last week of August - first week September, in which any adverse meteorological factors have an extremely negative influence on wine quality.

2.2 Grapevine Phenology

Coombe (1995) proposed a system called the Modified E-L system to differentiate the main, as well as several intermediate, phenological stages of the grapevines. This is a system of measurement and description of stages of the grapevine, which copes with the dual needs for a simple listing of major stages and, at the same time, provides intermediate detailed stages.

Summary: the identification of grapevine growth stages is necessary for the communication of cultural information, for decisions on establishment and cultural operations, and for use by research workers in the conduct of grapevine experiments. To be a successful system of growth stage identification, such a system should: (a) contain a succession of developmental events that always follow each other, (b) have stages that are easily described, and be clearly recognized and identified, and (c) have stages selected for consistency in assessment. Users of growth stage schemes may want descriptions of a limited number of major stages or, alternatively, a detailed set of precisely defined stages. This system combines both needs in the one scheme. The Modified E-L system is a 47 stage graphical scheme that allows the differentiation of each main phenolgical event of the grapevine (budburst, flowering, *véraison* and maturity) as well as many other minor intermediate stages.

Jones (2003b) gives information regarding plant phenology, the factors affecting phenology and on phenology, yield and quality interactions. Additionally climate change and phenology are analyzed.

Summary: while the onset and duration of each of the main phenological stages of grapevines varies spatially and for individual varieties, they are very consistent for the physiology of the main varietals in a given region and can be approximated by: Stage 1 - Shoot and inflorescence development, commencing around the end of March or first week of April, Stage 2 - Flowering, generally occurring in the first few weeks of June, Stage 3 - Berry Development, from the end of flowering in mid-June to the ripening stage, Stage 4 - Ripening, starts with *véraison*, near the end of July or the first week of August and Stage 5 – Senescence, from harvest at late September through early November and leaf fall, over the winter months leading back to bud break. Each of the major phenological stages of grapevines are governed by critical climatic influences. Temperature effects are evident in the spring where vegetative growth is initiated by prolonged average daytime temperatures above 10°C. During *floraison* and throughout the growth of the berries, extremes of heat can be detrimental to the vines. During the maturation stage, a pronounced diurnal temperature range effectively synthesizes the tannins and sugars in the grapes. Atmospheric moisture, in the form of humidity and rainfall, hastens the occurrence of fungal diseases (i.e., powdery mildew, downy mildew, botrytis bunch rot). In extreme cases, water stress resulting from high evaporative demand can manifest itself in leaf loss, severe reductions in vine metabolism, and fruit damage or loss. The occurrence of rain during critical growth stages can lead to devastating effects. In general, bud break, flowering, *véraison*, and harvest dates are the most observed events with very few growers noting any of the more detailed micro-stages in the Modified E-L system. Numerous studies based on climate parameters (mostly temperature) have been used to try to predict the dates of the individual phenological events. The rate of development between the growing season phenological stages varies with variety, climate, and topography.

van Leeuwen et al. (2008) studied the precise heat requirement data for each grapevine variety to reach each phenological event. Due to global warming, the choice of later ripening grapevine varieties might be necessary in many regions to maintain late ripening conditions favorable for *terroir* expression. Hence, it is essential information heat requirement data for most grape varieties. In this work, the Phenology (budburst, flowering, *véraison* and ripeness) and temperature data were collected for many varieties, in a wide range of locations, over a great number of vintages, and heat summations base of 10°C were calculated for each variety to reach the key phenological stages.

Summary: the timing of grape ripening is crucial in wine production. Grapes that ripen too late in the season are harvested before a desired maturity. Data covered many winegrowing regions over many vintages. Climatic data were gathered from the nearest weather station for each cultivar. 10°C is generally considered as the thermal baseline for grapevine development. Vine phenology was modelled for a wide range of cultivars by means of an agro-climatic model based on a temperature sum with a base of 10°C. Consistent classifications were produced for budbreak, flowering, *véraison* and maturity. Phenology is not only temperature related. Hence, an agro-climatic model cannot perfectly predict vine phenology. Bud break is related to pruning date. It is also related to soil type: dry soils, or soils with shallow rooting, warm up more quickly in the spring and thus speed up budding. Ripening speed, which is defined by the interval between véraison and maturity, is influenced by vine water uptake conditions. This research showed that water deficit increases berry ripening speed.

Salazar-Gutierrez, Johnson, & Chaves-Cordoba (2013) studied the use of a phenological model, in replacement of a calendar based model, for describing wheat growth. They determined the base temperature for key phenological stages of different winter wheat cultivars and developed a phenological model using the base temperature for predicting the duration in terms of thermal time for the different phenological stages.

The research concludes that the heat accumulated over time provides a more accurate physiological estimate than counting calendar days and that knowing the base temperature for each individual developmental stage for a cultivar can be useful for the development of commonly used wheat simulation models.

Summary: phenology is considered an important component in crop adaptation to local environmental conditions, and both season duration and the length of the phenological stages are important determinants of grain yield. The beginning and the end of these stages are good indicators of potential crop growth. Each plant has a specific temperature requirement before certain phenological stages are attained. Several models have been proposed that describe the effect of temperature on phenological development as improvements on the use of calendar time for predicting development. One of the most extensively used method is the accumulation of daily mean temperature above a base temperature. A basic requirement for this approach is the determination of the critical temperature below which phenological development ceases, referred to as the base temperature T_b. For wheat the use of a base temperature of 0 ºC, independent of the phenological stage, has been most common. However, when a base temperature was determined for a particular period (e.g. from floral initiation to anthesis) different values other than 0 ºC, ranging from 1.2 to 1.6 ºC, have been reported. The results showed that T_b varied depending on the development stage of the crop and cultivars and that the use of a single value of 0 °C is not recommended. There are large variations in temperature from day to day and growing season to growing season. The use of thermal time rather than calendar time considers this variability and provides an explanation for differences in crop maturity when observations from different years are compared.

2.3 Wine Quality

Charters (2003) studied qualitatively what drinkers consider the nature of wine quality and what they believe its features to be. The findings of the study suggest that different types of drinkers have different conceptualizations of quality. They concluded that drinkers tend to view quality multidimensionally. The dimensions may be intrinsic or (occasionally) extrinsic to the wine.

Summary: in this work, a detailed analysis of the concept of quality applied to wine and of the assessment of wine quality is conducted. The study used a sample with three groups of individuals (involved in wine making, involved in wine marketing and general wine consumers). Participants were asked to taste several wines and were interviewed about their perceptions and about wine quality concepts. A key conclusion of this study is that drinkers at all levels of involvement share widely differing perspectives on quality. It is stated in this research, "language informs evaluation, and shapes how it takes place". The interpretation of the differing terms used in the assessment of quality (especially the gustatory sub dimensions) seems to vary between individuals. The dimensions of wine quality may be intrinsic or extrinsic to the wine. Intrinsic quality relates to the wine-in-the-glass; thus, to what is tasted. Extrinsic classification makes use of factors extraneous to the wine as its origin and price to establish its quality. There is an apparent indecision about how the assessment of wine quality should be carried: as a scientific process, performed by the chemical analysis of the product or as an organoleptic tasting procedure. Wine consumption is an aesthetic or quasi-aesthetic process and thus less susceptible to precise, quantifiable analysis.

Keuris (2008) investigated whether fine and mass wine consumers differ in their use of signals to assess quality for both mass and fine wines. The use and importance of inherent (intrinsic) signals and non-inherent (extrinsic) signals were studied. The main intrinsic signals were sensory characteristics, appearance, age, pleasure and paradigmatic aspects. The main extrinsic signals were reputation of paradigmatic signals, certification, recommendations, promotion and price. The results show that intrinsic quality signals are mainly used by fine wine consumers and that easy-to-determine extrinsic signals are particularly used by mass wine consumers.

Summary: in this work, only signals related to still wines are analyzed. Wine is an experience product and the evaluation of quality can only occur after consumption. Consumers need intrinsic signals which they cannot assess prior to consumption. In absence of any intrinsic quality signals consumers have to rely on extrinsic quality signals. Extrinsic signals are related to the product but are not part of the physical product itself. Extrinsic signals are therefore promotional tools since they can be manipulated without changing physical product. Country and region-of-origin, grape variety / wine type, brand reputation, store reputation, certification of quality, harvest year or aging potential, controlled appellation, certified sustainability, recommendations from critics, awards, advertisement and price are extrinsic signals. The market is influenced greatly by reviews of critics as Robert Parker, Jancis Robinson and others. People like to be influenced by the opinion of critics but it is questionable if they influence the total wine market. Some critics accompany their review with a ranking to make them comparable. The influence of those ratings cannot be understated since many fine Bordeaux producers wait for Robert Parker's ratings before setting their release prices. A good rating will automatically mean a higher demand and consequently a higher price for a wine. Mass wine consumers have less knowledge about wine and are conservative in trying out new wines compared to fine wine consumers. This behavior leads to limited quality signals use for the evaluation of the quality of wine. The signals used are easy-to-determine abstract signals such as extrinsic signals; price, brand and packaging.

Duarte, Madeira, & Barreira (2010) studied how motives / attitudes, purchase and consumption behavior, as well as extrinsic attributes for wine choice, of Portuguese young adults differ from other age segments. They concluded that wine consumers can be grouped in four different clusters, according to the consumption patterns, and that these patterns are related to the consumers' age.

Summary: in this work, they studied how Portuguese consumers choose wine, what the relevant attributes for expected quality perception are, and what kind of consumer segments can be identified. They try to identify the importance of extrinsic attributes and the main sources of information for the wine purchase decision, the motives / attitudes, the frequency and occasions of consumption and how these issues relate to consumers' age. They implemented a survey using the internet in July-August, 2008, and the answers of a sample of 1160 respondents were analyzed. The statistical analysis was implemented using ANOVA tests and factor and cluster analysis. According to different consumption patterns, they could differentiate four clusters: Cluster 1 - Occasional enthusiast wine drinkers, Cluster 2 - Regular wine drinkers, Cluster 3 - Infrequent wine drinkers and Cluster 4 - Occasional convivial wine drinkers. Young adults are mainly represented in segment 3, for those aged less than 25 years and in segments 3 and 4, for those aged 25 to 34 years. Segment 2 is associated with older consumers, mainly men that drink wine every day and appreciate wine taste. For all segments, the three main extrinsic attributes for choice decision are the region of origin, having a cork stopper, and the price. Youngest consumers (less than 25) do not seem to be attracted by the taste nor even recognize the conviviality role of wine.

2.4 Aggregation of Opinions

Diaconis & Graham (1977) showed that the Kendall tau distance and the Spearman footrule distance are "equivalent," in the sense that they are within a factor.

Summary: four metrics were defined in which the Kendall tau distance and the Spearman footrule distance were included. Several properties are presented for the four metrics as well as some inequalities representing the relations between the metrics. It was shown that the Kendall tau distance and the Spearman footrule distance are "equivalent," in the sense that they are within a factor. The use of the metrics was suggested to compute the distance between permutations, as a measure of association. The research points out that although Kendall tau distance and the Spearman footrule distance are roughly similar, the Spearman footrule distance is easier to interpret and the Kendall tau distance has the advantage of having its distribution tabulated for small samples.

Cook & Kress (1986) studied the problem of combining individual preferences into a group choice or consensus. One of the fundamental models for consensus formation is based on a measure of distance between ranked preferences. In this study the fundamental model of Kemeny & Snell (1962) taking into account a strength of preference of object A over object B was used. A median consensus ranking is suggested as the ranking that minimizes the sum of the proposed distance to all the voters' rankings.

Summary: the incorporation of the strength of preference component into the model permits a more general expression of preference on the part of the ranker. If a ranker feels that object 1 is preferred to object 3 which is preferred to object 2 but also feels that, given a 10 point integer scale, object 1 should sit in the first position, object 3 in fourth position and object 2 in fifth position then the ranker strength of preference for object 1 over object 3 is much stronger than for object 3 over object 2. In the original model, the strength of preference is always equal to one or to zero when the preference is the same (tie). This new model preserves the basic structure and properties of the original model that is based on a measure of the distance on the ranking space but accommodates a strength of preference component in the ordinal structure.

H. P. Young (1988) analyzed the contribution of Condorcet (1785) to the theory of group decision making. Condorcet believed that the purpose of voting is to make a choice that is best for society. According to his view, there is one choice that is best, another that is second best and so forth. When designing a voting rule, the objective should be to choose the ranking of alternatives that is most likely to be the best. Condorcet solved this problem using a form of likelihood estimation. The study concludes that Condorcet's method is a rational way of aggregating individual choices into a collective preference ordering and that when the objective is to rank a set of alternatives the Condorcet's rule is undoubtedly better than Borda's.

Summary: first, this study summarizes Condorcet's proposal. Then it shows that Condorcet's method can be interpreted as a statistical procedure for estimating the ranking of the candidates that is most likely to be correct. Comparisons with method proposed by Borda (1781) were conducted and resulted in finding that if only one candidate (the first) is to be selected then Borda's method often gives better results but, when the problem is to rank a set of alternatives, the Condorcet's method is "undoubtedly" better than Borda's. The study concluded that Condorcet´s method is the unique ranking procedure that satisfies a variant of independence of irrelevant alternatives (Local Independence of Irrelevant Alternatives, LIIA) together with several other standard conditions in social choice theory.

Risse (2005) analyzed the dispute between the majority rule used by the Marquis de Condorcet and many others and its competitor, proposed by Borda, the Borda count. These rules are collective decision procedures used in social choice theory. This work refutes the objections of the mathematician Donald Saari to the Condorcet's majority rule arguing that the objections to majority rule fail and holds the view that defenders of Condorcet cannot muster arguments to convince supporters of Borda, and vice versa.

Summary: this work first explains that under certain conditions the majority rule is not able to produce a transitive ranking: the Condorcet's paradox. The Condorcet's proposal selects rankings supported by a maximal number of votes in pairwise votes and the Borda count assigns 0 to the last-ranked candidate, 1 to the second-last-ranked candidate and so forth until n-1 is assigned to the top-ranked candidate, and then, for each of the candidates, sums over those numbers to determine the group ranking. Several arguments are presented to support Condorcet's proposal. An example the Generalized Jury Theorem is presented in favor of Condorcet's proposal as the same rankings emerge when Condorcet's proposal or the generalized Jury Theorem are applied.

2.5 Quality Control of Meteorological Datasets

Alexandersson (1986) developed a test for the detection of changes of the mean value in a candidate series compared with a homogeneous reference series. The test tests the relative homogeneity of a undocumented (no metadata) candidate time series with respect to a reference time series developed from a group of surrounding neighbor weather stations.

Summary: a reference series is a time series of weighted averages of the same data, from several well-correlated neighbor stations. For each variable, correspondent values from neighbor stations with a significant large positive r, are used in the weighted average using a weight proportional to the r value. The number or the identity of neighboring stations is not fixed in time. For each variable, a set of neighbor stations may be optimal for a particular period and not be optimal for another period. To detect the relative inhomogeneities ratios (used with precipitation) or differences (used with temperatures) a time series of ratios / differences between the candidate series and the corresponding reference series is computed. The ratios / differences series, with data for n moments (days, weeks or months), is divided in two segments (from moment $t = 1$ to moment $t = a$ and from moment $t = a + 1$ to moment $t = n$) and each segment mean is calculated. The differences between segments' means are evaluated for every potential change-point "a" ($1 \leq a \leq n$) and a statistical test is used to detect if a significant difference exists between the means of the two segments. If one intends to correct data for the period 1, 2,…, a then the values within this period should be multiplied to (in the ratio case) or summed (in the difference case) with a constant value that brings the first segment mean to the value of the last segment mean. If the data contain only one shift, then a homogenized series where all data refer to the present measuring situation is obtained.

Reek, Doty, & Owen (1992) tested a computer program (ValHiDD -Validation of Historical Daily Data) developed by the National Climatic Data Center, in the USA, to identify, categorize, and eliminate gross digitization and observer errors. This quality control software uses a series of tests as a means of modeling the human process of data review. 138 stations of a 1300 station subset of the climatic dataset DSI-3000 were manually reviewed and closely matched the automated correction process through the ValHiDD software. The automation correction process has proven very effective and is expected to be used in the production of nearly error-free weather datasets.

Summary: the climatic dataset DSI-3000 from the National Climatic Data Center in the USA was analyzed and tens of thousands of erroneous daily values resulting from data-entry, data-recording and data-reformatting errors. ValHiDD uses several checks in the quality control of data. Temperature checks include extremes, daily maximum temperatures less than minimum temperatures, spikes and steps in a time series of daily values, continuous runs of the same temperature, and excessive diurnal ranges. Precipitation checks include extremes of precipitation and snowfall, and inconsistencies among total precipitation, snowfall and snow depth. The check for extremes compares appropriate data values to statewide period of record extremes in a given month of observed lowest and highest maximum and minimum temperatures, total precipitation and snowfall. The software outputs error codes identifying data that failed a check as well as the offending value. ValHiDD also outputs, when possible, a replacement value for the offending datum. These replacement values are given only for three conditions: the original data had a misplaced decimal point, the sign of the original data was reversed and the original data value was wrong by 100 units. For all other data that failed a check, no replacement value is given. This work concluded that ValHiDD has proven an effective tool for automatically removing errors from meteorological datasets.

Feng, Hu, & Qian (2004) examined the daily meteorological data from 726 stations in China from 1951 to 2000, and developed a climatic dataset that contains 10 daily variables: maximum and minimum surface air temperatures, mean surface air temperature, skin surface temperature, surface air relative humidity, wind speed, wind gust, sunshine duration hours, precipitation, and pan evaporation. The quality-control methods designed and used in developing the dataset were detailed. The resulting data series, as an alternative to the original series, showed both spatially and temporally consistent trends in the occurrence frequency of extreme climate events compared with the unadjusted data series.

Summary: this work used a meteorological dataset with data from 1 January 1951 to 31 December 2000, collected with data from the Chinese National Meteorological Centre. Only 60 out of the 726 stations had support metadata. The quality control methods applied to identify erroneous data resulting from sensors and observation sources were High–low extreme check for daily values; Internal consistency check; Temporal outliers check and Spatial outliers check. Missing data and suspicious data screened by the four previous checks were estimated using fitted values, by linear regressions, of neighboring stations with high data's correlation to the station where data is necessary. For each station, data from the nearest five neighbor stations were used to create reference data series used in the homogenization adjustments to the original dataset. Magnitudes of the daily adjustments vary from month to month and from variable to variable.

Wang, Wen, & Wu (2007) proposed a penalized maximal t test (PMT) for detecting undocumented mean shifts in climate data series. PMT takes the relative position of each candidate change-point into account, to diminish the effect of unequal sample sizes on the power of detection. Monte Carlo simulation studies were conducted to evaluate the performance of PMT, in comparison with the most popularly used method, the standard normal homogeneity test (SNHT). It was shown that the false-alarm rate of PMT is very close to the specified level of significance and is evenly distributed across all candidate change-points, whereas that of SNHT can be up to 10 times the specified level for points near the ends of series and much lower for the middle points. In comparison with SNHT, PMT has higher power for detecting all change-points that are not too close to the ends of series and lower power for detecting change-points that are near the ends of series but, on average, PMT has significantly higher power of detection.

Summary: this study attempted to improve a test for detecting undocumented shifts, proposing a new test statistic that treats more equally each candidate change-point in the time series being tested. This study only considered the detection of an undocumented shift in the mean and was focused on the case in which the time series being tested contains at most one change-point noting that one can implement statistical tests that are developed for the "at most one change-point" case with an appropriate recursive testing algorithm to detect multiple change-points. The task of undocumented change-point detection is to find out the most probable moment in the time series where the means before and after that moment are statistically different from each other. It is necessary search over all candidate change-points for the most probable position of an undocumented mean shift. The traditional test for this kind of problem is the likelihood ratio test that can be transformed into an equivalent test that involves the two means, before and after the candidate change-point, with a test statistic that follows Student's t distribution with (N-2) degrees of freedom. The Standard Normal Homogeneity Test (SNHT) is one case of the use of the t test. The power of the t test decreases when the two samples are of unequal size (relative to the equal-size case) and as a consequence, the maximal t test and SNHT suffer from the disadvantage that points in a homogeneous time series have different probabilities of being mistakenly identified as change-points. That is, for a change-point of certain magnitude, the test would detect it more easily when it occurs near the ends of the series than when it occurs around the middle, and the test would mistakenly declare many more change-points near the ends of a homogeneous series than around the middle. In this work, a penalized maximal t test (PMT) is proposed: PMT uses the same test statistic that follows Student's t distribution multiplied by a penalty function that even out, to a great extent, the U shape of the false-alarm rate curves of the unpenalized maximal t test (and SNHT). The new test statistic takes the relative position of each candidate change-point into account to reduce the distortion of the test statistic that is due to unequal sample sizes. Observations are treated more equally during the process of searching for the most probable change-point position/time.

2.6 Temperature Lapse Rates

Stone & Carlson (1979) studied the vertical temperature structure of the atmosphere. Early observations indicated that the lapse rate[11] is close to 6.5 ºC/km with

[11] Rate of change in temperature observed while moving upward through the Earth's atmosphere (source: Enciclopedia Britannica, www.britannica.com)

little seasonal or latitudinal variation. Hemispheric mean lapse rates were analyzed and suggest that a better choice for a constant lapse rate would be 5.1 ºC/km but point out of small seasonal changes in hemispheric mean lapse rates and the existence of strong seasonal changes in tropospheric mean lapse rates at latitudes about 50 ºN.

Summary: tropospheric mean lapse rates were calculated and compared with the lapse rates associated with moist convection and large-scale baroclinic eddies. They show that the effect on temperature structure of moist convection and baroclinic eddies varies with latitude: in low latitudes moist convection dominates; in high latitudes the baroclinic eddies dominate. The dividing point in the two regimes is 35 ºN because centered on this latitude there is a transition region, 10 degrees wide, where both mechanisms have an important effect on the temperature structure. As seasonal changes in the two mechanisms tend to counteract each other, the region with latitude between 30 and 50 ºN has very little seasonal change in mean lapse rates. The study suggests that the vertical temperature structure at all latitudes may be modelled including with two lapse rates: the moist adiabatic lapse rate and a large-scale baroclinic adjustment. This model would allow the study of the interaction between vertical and meridional temperature structure in climate problems.

Rolland (2002) as previous works revealed absence of a lapse rate seasonal pattern this research used 640 stations to reexamine monthly variations in air temperature lapse rate in Alpine regions and to quantify the improvement in the reliability of lapse rate variations by adding topographic information. Additionally this research analyzed the inconsistencies in formerly published results and assessed the accuracy of temperature interpolations based on lapse rates. Results show that lapse rates are lower in winter when compared to summer values and are lower for minimum temperatures than for maximum temperatures.

Summary: a large network of temperature stations (269 stations in northern Italy, 205 in the Tyrol area, and 166 in the Trentin–Uppe) was analyzed throughout the Italian and Austrian Alps, with 30 years data (1926-1955). Monthly average values of mean, maximum, and minimum temperatures were calculated for all the stations. The data were divided in two groups according to the location of the station: valley bottoms and slopes. Linear regression models were calculated to obtain series of linear equations relating air temperature to elevation, using lapse rates. The existence of a lapse rate seasonal trend seems to be a general phenomenon and common to minimum, mean, and maximum temperatures.

Chapter 3

The Douro Valley

3.1 Introduction

The Douro Valley is a region with a total area of 250 000 ha of which 40 000 ha are planted with vineyards. The Douro river flows westward through the Douro Valley coming from its source in Picos de Urbión, in Spain. The westernmost area of the region is located 70 km from the Atlantic Ocean. The Douro Valley extends along 90 km in the West-East direction and along 50 km in the North-South direction. Along the river, vineyards are planted on steep hillsides from the riverbanks up to 600 m elevation. As temperature decreases, on average 5 ºC / 1000m elevation increase (Rolland 2002), the hottest vineyards are located at low elevation sites, near the Douro River or its tributaries.

In this section, we present background information on the Douro Valley: history, topography, and geology. Additionally, we give information on Port wine history, styles, common grape varieties, and blending. Finally, we present the main factors that influence vintage quality and vintage yield.

3.2 History

In 1756, during the reign of José I, on the initiative of Sebastião José de Carvalho e Melo, Marquis of Pombal, the Companhia Geral de Agricultura das Vinhas do Alto Douro was created (Spence 1997). This Company implemented the demarcation of the Douro region, originating one of the first demarcated wine regions in the world. The navigability of the Douro River was initially limited to the areas where vineyards were planted because the wine transport was made through the Douro river using especially designed boats, the *rabelo* boats. At that time the river had a narrow gorge obstructed by a waterfall formed by gigantic outcrops of rock (*Cachão da Valeira,* famous for the site of Baron Forrester's death) making it impossible to navigate upstream into the remote eastern reaches of the Douro. Initially, the demarcated region was smaller than today, occupying only 40 km of the westernmost part of the present region. A new demarcation of the region, in 1907, included the area of the Douro Superior in the Demarcated Region of Douro. The completion of the railway along the Douro in 1887 meant that the *rabelo* boats were no longer the only means of transporting wine and other goods from the Douro Valley to the coast.

Throughout the 20th century, the Demarcated Douro Region has been subject to several regulatory models. The Interprofessional Committee for the Demarcated Douro Region (CIRDD) was instituted in 1995. The principal regulatory mechanism for production continues to be the system for distributing the *benefício*, according to which the amount of must that is authorized for making port wine is allocated according to the characteristics and quality of the respective vines. Mechanization was introduced in the 1970s to help with some of the more arduous tasks in the vineyard such as the scarifying of the land and bringing with it new wide, earth-banked vineyards and "vertical planting" along steeper hillsides that no longer require building walls to shore up the terraces. The aesthetic impact of these new vineyards on the landscape varies, yet the mountain viticulture of the Douro continues to be carried out almost totally by hand. The rocky nature of the soil, the steep hillsides, and the existing terraces themselves are extremely difficult to adapt to the use of machines, though the product, port wine, is today mostly made in modern, totally mechanized wineries (source: UNESCO at http://whc.unesco.org/en/list/1046).

3.3 Topography

The Douro Valley is a mountain wine region. It is a UNESCO World Heritage protected site. The valley is hilly throughout. Most of the region landscape is composed of undulated mountains. The irregular topography of the Douro Valley is generated by the geology that alternates between schist and granite. The region is characterized topographically by sloping vineyards arranged in various terraced configurations. These terraces have been created and perfected throughout the centuries enabling man to cultivate vines on the steepest slopes. Two thirds of the region's planted area is located on rocky hillsides with a gradient of over 30%. The Douro Valley has a median elevation of 470 meters.

3.4 Geology

The geology of the Douro Valley is dominated by schistous-layered rock, oriented nearly vertically, with some outcrops of granite. Vertically oriented, schistous rock behaves like stacked tiles with cracks in between each, which allows grapevine roots to penetrate deep to find nutrients and moisture. In the summer schist rock retains heat during the day keeping the vine roots zone warm during the night. In the winter schist rock allows the water from rain to penetrate and be retained deep.

3.5 Port Wine

Red Port wine is a fortified wine as brandy is added before fermentation is completed leaving some residual sugar that makes Port wine sweet and raises alcohol to a final value around 20º. The choice of the aging vessel and aging period will determine the Port wine taste and its style. Two broad styles of Port may be defined: wood aged Ports, which age in cask; and bottle aged Ports. Wood aged Ports have several styles: Ruby, Reserve, Late Bottled Vintage, Tawny and White Ports. Bottle aged Port may be found in two different styles: Vintage and Crusted Ports.

All styles of red Ports are made from several grape varieties, some always present – *Touriga Nacional, Touriga Francesa, Tinta Roriz (Tempranillo), Tinta Barroca* and *Tinto Cão*. Many other varieties, from about fifty allowed by law, may be present.

Vintage Port is the top, most exclusive, quality Port made from perfect ripen top quality grapes, grown in the best sites and only in very good vintages. Though accounting for just 1% of total Port production, Vintage Port commands the most attention and speculation from world wine markets and is, usually, the category of Port wine rated in renowned vintage-charts.

All Port wines are blended wines. Most Vintage Ports are produced from a number of different vineyards. Recently, several Port producers have begun making single-*quinta* Vintage Ports. These Ports are generally produced from their best vineyard, but not in the best years. Single-*quinta* Vintage Ports, while often excellent, are rarely as good as the "mainstream" Vintage Port from that producer.

Historically the vineyards in the Douro were planted, and some vineyards still are, with a mix of indigenous grape varieties to the point where the winemakers are not always sure of which or how many varieties are growing in each vineyard. This disorderly planting is referred to as a *field-blend* and is a notable element of Port wine. The downside of a field-blend is that not all varieties grow equal, as some varieties ripen faster than others. To overcome this, most new vineyards are planted with a single variety to ensure the ideal level of ripeness on harvest day. Even when grown in separate vineyards, several grape varieties will be blended, as each one adds some specificity to Port wine. Even top Vintage Ports are made from wines from the same vintage but made from grapes grown in different sites, at different elevation, with different orientation, with different types of soil, and maybe from different sub-regions inside Douro valley, having different climates. Grape varieties used in each Port wine blend, as well as the sites where they are grown are, usually, unknown. Port wine labels do not mention, usually, the grape varieties used in the blend.

White Port and Rosé Port are less known Ports with small productions that, together, represent 14% in volume of all Port wine production and 11.5% of all Port wine sales in 2012 (source: IVDP). These two types of Port wine will be out of the scope of this research.

3.6 Main Factors for Vintage Quality

Quality is a key concept in this research. It is an abstract concept with different definitions. The word quality, when used as an adjective, is defined as “of high standard” in “Cambridge Dictionaries Online”. Still, the definition of “high standard” is not an easy task as it may have the meaning of “excellence” which is an absolute concept or it may have the meaning of “superiority” which is a relative concept. In this research when referring to wine quality, we will not be looking at the quantifiable and measurable characteristics or attributes of a wine (e.g. chemical and biological characteristics). We will use a user-based view of wine quality, adapted from food quality (Cardello 2010) to wine, as "*the adequacy of wine organoleptic characteristics to sensory expectations of consumers with some level of knowledge and education on wine so that its consumption or tasting gives them pleasure*".

Some factors are widely accepted as influential on wine quality: i) weather and weather extremes, ii) grapevine diseases, iii) vintners' skills, and iv) soil characteristics. All, except the soil characteristics, that remain unchanged from one year to another in a region, influence the variability of vintage quality in a region. A "great vintage" is backed by an exceptional growing season entwined with the talents of region vintners.

3.6.1 Weather

The weather is the most important factor in vintage quality variability. Already in the early 1970s the seasonal conditions were considered of paramount importance to grape ripening (Winkler et al. 1974). Daily variation of maximum, minimum and average temperatures, of precipitation amount and distribution, of air humidity, of wind speed / direction, and of soil water availability determine grape ripening evolution and grape final quality that will make possible for the vintners to produce great wines. All these weather variables interact with each other and some of them influence other types of factors like weather driven diseases (Downy Mildew) which may also have important impact on wine quality and yield.

Heat

Previous research has shown that measures of accumulated heat (e.g., Lopes et al. 2008; van Leeuwen et al. 2008; Parker et al. 2011; Gladstones 2011 and others) help describe grapevine growth in numerous settings and across many varieties. These studies use a thermal time concept based on the observation that each phenological event occurs when a critical amount of accumulated heat above a critical base temperature is reached (Bonhomme 2000). While it is generally accepted that 10 ºC is the base temperature (Huglin 1978; Winkler et al. 1974; Carbonneau et al. 1992), others have found that this threshold varies by variety, location, the period of vine growth, and the water status of the plants in the season of interest (Jones 2013).

Water Deficit

Ojeda et al. (2002) studied the influence of pre and post-*Véraison* water deficit on synthesis and the concentration of skin phenolic compounds during berry growth and its relation to wine quality. They showed that there are two types of berry responses to water deficit: an indirect and always positive effect on the concentration of phenolic compounds due to berry size reduction and a direct action on biosynthesis that can be positive or negative depending on type of phenolic compound, period of application, and severity of water deficit.

Weather extremes

Weather extremes such as hail, frost, cyclones, heavy rain and extreme temperatures, usually have an influence that is limited in geographic and temporal scope, as they only produce effects during short periods and on small sites of the whole region. However, depending on their intensity and duration, may have impact on the annual growth pattern of the grapevine influencing the grapevine annual yield and/or fruit/wine quality.

Grapevine Diseases and Insect Plagues

Several diseases and insect plagues affect Douro region grapevines: i) Botrytis, ii) Downy mildew, and iii) Powdery mildew are most common fungal diseases in Douro region. Insect plagues: iv) Phylloxera, v) Leafhoppers (*Empoasca Vitis* and *Jacobiasca Lybica*), and vi) European grapevine moth (*Lobesia Botrana*) are most recurring insect plagues in Douro region.

3.6.2 Vintners' Skills

During the course of a year a vintner is faced with several jobs in the vineyards such as planting vines, training vines, pruning, ending, tying, working the soil, fertilization, irrigation, foliage management, pest control and prevention of fungal diseases, scheduling the harvest in the optimal moment for each particular vineyard. After the harvest, processing the grapes and making the wine, a vintner is faced with several decisions in the cellar with respect to maceration, extraction, additions to juice / must, fermentation conditions, filtration, clarification, blending and stabilization treatments. Douro region's viticulture is hillside viticulture. Over 70% of the Douro vineyards are planted on hillsides with slopes greater than 30%, at elevations ranging between 100m and 700m (Queiroz et al. 2008) which limits the level of mechanization of vintners tasks in the vineyards. In the Douro Valley, in the last two decades, the cellars had great evolution in terms of equipment, using now state-of-the-art vinification technology such as the stainless steel vats with temperature control and automated treading machines.

The quality of the Douro Valley region DOC wines and Port wine is improving as shown by the number of Douro wines with top ratings in international renowned Vintage-Charts. This quality improvement is possibly a consequence of both climate trends and better skills and knowledge of Douro's vintners. An adequate choice of the best varieties to grow and the improvement of the conditions for grapevines to grow perfectly and to protect them for pests and diseases are actions that the Douro's vintners implemented. In the cellars, vintners' better knowledge of wine processing, together with the adoption of sophisticated vinification technology have also a role on the quality improvement.

All factors referred in this section influence the variability of vintage quality. These factors do not act isolated but interact as, for example, Downy Mildew that is a weather driven disease. Moreover, the grapevine is a complex living system that permanently tries to adapt to surrounding environment managing resources in the most profitable way: similar to human beings, every plant is individual expressing its own genes in its surviving strategies (John Moore et al. 2008).

Chapter 4

Quality Control of Meteorological Datasets

4.1 Introduction

Quality control of a meteorological dataset consists in a set of procedures to detect erroneous observations in meteorological data series (Aguilar et al. 2003).

The objective of quality control is to verify whether a reported data value is representative of what was intended to be measured and has not been contaminated by unrelated factors. The observer, or automated observing system, should apply quality control procedures to ensure that the time and station's identification are correct, that the recorded values reliably reflect current conditions, and that the observed elements are consistent. These steps should preferably be taken prior to the recording or transmission of an observation (World Meteorological Organization 2011) but, in some regions and for some time periods, meteorological datasets provided by meteorological agencies, have poor or limited quality control.

Metadata are data about data with the purpose of providing detailed information that is necessary for users to gain adequate background knowledge about the data. In essence, metadata states who, what, when, where, why, and how about every data that are being documented. Complete metadata are necessary in order that the final data user may have no doubts about the conditions in which the data were recorded, gathered and transmitted. Metadata should always be available together with the observational data, although it is not always the case: metadata often lack accuracy and completeness and in some cases are not available at all.

As reliable time series are necessary in order to analyze weather variability or climate trends, the meteorological datasets used in this research had to undergo a quality check to detect and correct inconsistent or missing data as well as to detect and correct existing inhomogeneities.

Quality control of data series consists on: i) data cleaning procedures and ii) data homogenization procedures. In this section, we will present methodologies for quality control of meteorological data series and will present the results of the use these methodologies in quality control of the meteorological datasets from the Douro Valley.

4.2 Data Cleaning Procedures

4.2.1 Introduction

Human errors and errors resulting from measuring instruments can happen at all stages of meteorological series production process, from data acquisition to its storage, resulting in data that are not registered or is registered with errors. In the latter type of errors are, for example, maximum temperatures with values lower than minimum temperatures, same maximum or minimum temperature over several consecutive days, negative amounts of precipitation, wrong position of the decimal separator of a temperature / precipitation amount, and temperatures / precipitation amounts much higher or much lower than the corresponding values recorded at well-correlated proximity weather stations.

In order to identify and clean meteorological data from erroneous values, several methods were proposed in the last two decades by several researchers (Reek, Doty, and Owen 1992; Stooksbury, Idso, and Hubbard 1999; Feng, Hu, and Qian 2004; Zahumenský 2005; You, Hubbard, and Goddard 2008, and others).

Most methodologies developed to identify erroneous data resulting from sensors and observation sources have similarities as they propose several quality checks, based on flagging schemes, to identify data that are suspect of being erroneous. Moreover, they indicate procedures to estimate missing and deleted erroneous data.

In this research we decided to use the well-established methodology proposed by (Feng et al. 2004). This methodology was proposed to develop a gridded climatic dataset (1.0º x 1.0º) covering China, using 10 daily weather variables from 726 weather stations and consists of a five-step check:

- hi-low extreme check for daily values;
- internal consistency check;
- temporal outliers check;
- spatial outliers check;
- missing data check.

The five-step check methodology proposed by (Feng et al. 2004) will now be briefly explained.

4.2.2 The Five Steps Check

Step1: Hi-Low Extreme Check for Daily Values

Daily values of weather variables collected at several weather stations are compared with established meteorological temperature and precipitation extreme values, recorded in global or local weather and climate extremes databases.

Records with values greater than / less than the highest / lowest extreme values for the region are flagged. Flagged values are excluded from future quality control calculations.

Step2: Internal Consistency Check

This check applies to the detection of erroneous data due to digitizing[12], typos, and unit differences. Three rules are used to check the daily temperature and precipitation data of individual weather stations:

i) Internal inconsistency: identifies errors such as daily $T_{max} < T_{min}$, $T_{avg} > T_{max}$, $T_{avg} < T_{min}$, and Precipitation < 0.

[12] Digitize: to transcribe data into a digital form so that it can be directly processed by a computer (http://www.thefreedictionary.com/)

ii) Excess diurnal temperature range: identifies extraordinarily large daily temperature ranges T_{max}-T_{min} while T_{max} and T_{min} are within their reasonable ranges. Reek et al. (1992) suggests that when diurnal temperature range exceeds 24 ºC both maximum and minimum temperatures should be checked.

iii) Flat line check: identifies data of the same value for at least seven consecutive days (not applied to precipitation data). For the flat line check all the consecutive data are flagged, except the first value.

Flagged values are excluded from future quality control calculations.

Step3: Temporal Outliers Check

Temporal outliers check is only applicable on temperature data. Hi-low extreme check and internal consistency check cannot detect outliers where a value is much larger / smaller than its neighboring values (values at adjacent dates during the whole period of data) but are not much larger / smaller than the threshold for being detected by the consistency check. To identify such outliers Lanzante (1996) proposed a method based on the biweight mean and the biweight standard deviation. This method was used in the methodology proposed by the Global Daily Climatology Network, V1.0 (Gleason 2002), published by National Climatic Data Center, USA. According to the method, for each day i ($1 \leq i \leq 365$) in each year k, three daily temperatures (for example T_{max}) are used: Y_{i-1}^{k}=$T_{max}^{day\ i-1, year\ k}$, Y_{i}^{k}=$T_{max}^{day\ i, year\ k}$, and Y_{i+1}^{k}=$T_{max}^{day\ i+1, year\ k}$. For each weather station, a series with the corresponding three-day values, using the daily temperatures from all the years (at least 10 years of data) is created, Y_i = $Y_{i-1,}^{k}$ Y_{i}^{k}, $Y_{i+1}^{k}$for k = 1 to the number of years in the data series. For example, for a daily series of temperatures in a 30 years dataset, Y_i would consist on 3 × 30 = 90 daily temperature values, three values from each year. This series is used to calculate the biweight mean estimate of the mean and the biweight standard deviation estimate of the standard deviation. These two estimators are more heavily weighted towards the center of their distributions than the tails. Biweight mean and biweight standard deviation are more robust estimators of the mean and standard deviation than the obtained using the sample average and the sample standard deviation.

The Median (MED) and Median of the Absolute Deviations of the observed values from the median (MAD[13]) are estimated. A weight u_i corresponding to each of the n observations (90 in the above example) of Y_i is computed as follows:

$$u_i = \frac{Y_i - MED}{c \times MAD} \quad (1)$$

where c value is a censor value. All observations Y_i beyond a certain critical distance from MED are given zero weight. A c value of 7.5 was used, in accordance to Lanzante (1996). For any $|u_i| > 1.0$, u_i was set to 0.0 to accomplish the censoring. The biweight estimate of the mean is:

$$\bar{Y}_{bi} = MED + \frac{\sum_{i=1}^{n}(Y_i - MED)(1 - u_i^2)^2}{\sum_{i=1}^{n}(1 - u_i^2)^2} \quad (2)$$

and the biweight estimate of the standard deviation is:

$$s_{bi} = \frac{\sqrt{n\sum_{i=1}^{n}(Y_i - MED)^2(1 - u_i^2)^4}}{\left|\sum_{i=1}^{n}(1 - u_i^2)\,(1 - 5u_i^2)\right|} \quad (3)$$

Both $\bar{X}_{bi}$ and $s_{bi}$are used to calculate the z-score of a particular day i temperature observation X_i:

$$z = \frac{|Y_i - \bar{Y}_{bi}|}{s_{bi}} \quad (4)$$

Observed values with z-score values greater than 3.0 were flagged as outliers. Flagged values were excluded from future quality control calculations.

Step 4: Spatial Outliers Check

This step consists on comparing the data from one station (the candidate station[14]) to the neighboring stations data. For each month, the correlation coefficients, r, are computed between the daily data series of each candidate station and the other neighbor stations. Stations with a large positive r, that is significant at 95% confidence level, are used to create a linear regression for the same variable with the candidate station. If more than five neighboring stations show a significant correlation with the candidate

[13] MAD: robust measure of the variability defined as the **M**edian of the **A**bsolute **D**eviations from the data's median
[14] Candidate station: a weather station from where data is being tested.

station at a specific month then the five neighboring stations with the lowest Root Mean Square Error (RMSE) are chosen. Thus, with the resulting k ≤ 5 regression equations, a daily value V_i of a variable (T_{max} or T_{min}) is flagged if it falls out of the interval:

$$VF_{ij} - F \times \; RSME_j < V_i < VF_{ij} + F \times \; RSME_j \qquad (5)$$

for all selected stations. Index j = 1, 2, ... , k where k represents the number of selected regression equations, V_i is the variable observation for day i from the candidate station, and VF_{ij} is the fitted value by regression equation j for day i. F = 5 for precipitation variables and F = 3 for temperature variables.

Step5: Estimation of Missing and Flagged Data

Missing data and data that are flagged as suspicious by any of the previous checks are estimated using the following expression:

$$v_{ei} = \frac{\sum_{j=1}^{N} \frac{VF_{ij}}{RMSE_j^2}}{\sum_{j=1}^{N} \frac{1}{RMSE_j^2}} \qquad (6)$$

where v_{ei} is the estimated value for day i and the other symbols are as in equation (5). The number or the identity of neighboring stations is not fixed in time. For each variable, a set of stations may be optimal for a particular period of time and not be optimal for another period. Thus, for each weather variable and for each candidate weather station, the spatial outlier check and the estimation of missing values are applied for individual calendar months.

4.3 Data Homogenization Procedures

4.3.1 Introduction

A homogeneous weather time series is one where data variations are caused only by weather and climate variations (Conrad & Pollak 1962).

Data discontinuities may occur for several reasons in a meteorological time series. They may result from changes in measuring instruments, in the location of the weather stations or their position, changes in the surrounding environment of the weather stations or changes in observational practices. These changes in the data patterns will be referred as artificial shifts.

There are two types of artificial shifts in meteorological time series: i) documented and ii) undocumented. Documented shifts are those with known position of the shift (i.e., the time and cause of the shift are recorded in the corresponding metadata). Documented shifts are easy to assess / test, as their position is previously known. With this type of shifts, the use of regular hypothesis testing for means or variances are applicable to statistically validate their existence. Whenever metadata does not exist or incomplete, appropriate statistical tests must be used in order to detect undocumented shifts. Undocumented shifts are shifts that are statistically significant even without metadata support.

Artificial shifts should be detected and eliminated from time series. With this purpose, several methods have been developed during the last two decades (Alexandersson 1986; Solow 1987; Easterling and Peterson 1995; Lund and Reeves 2002; Wang 2003; Wang, Wen, and Wu 2007; Wang 2008a; Wang, Chen, and Wu 2010 and others).

In this research we decided to use the well-established methodologies (Wang 2003; Wang et al. 2007; Wang 2008a; Wang 2008b) integrated on the software package RHtestsV3 (Wang 2011). TPR3 maximal F test (Wang 2003) is one of the methods integrated on the software package RHtestsV3 (Wang 2011) and, according to Reeves et al. (2007), is the best method for dealing with most climate series.

As the methods integrated on RHtestsV3 package are improvements and refinements of the Standard Normal Homogeneity Test (SNHT) (Alexandersson 1986) some background on these methods will be given in section 4.3.3.

4.3.2 Reference series

For a data series containing data of a weather variable (e.g., temperature, precipitation) from a candidate station, a reference series is a time series of weighted averages of the corresponding data, collected from several well-correlated neighbor stations. For each weather variable, correspondent values from neighbor stations are used in the weighted average, using a weight proportional to its *r* value.

Let $\boldsymbol{Y} = \{y_1, y_2, \ldots, y_n\}$ be the candidate meteorological time series and y_i a specific value in moment *i*. $\boldsymbol{X_j} = \{x_{j1}, x_{j2}, \ldots, x_{jn}\}$ will denote one of the surrounding neighbor *k* reference sites and x_{ji} a specific value from this site at moment *i*. A reference time series $\boldsymbol{W} = \{w_1, w_2, \ldots, w_n\}$ is defined according to:

$$w_i = \frac{\sum_{j=1}^{k} r_j^2 x_{ji}}{\sum_{j=1}^{k} r_j^2} \tag{7}$$

The number or the identity of neighboring stations is not fixed in time. For each weather variable, a set of neighbor stations may be optimal for a particular period and not be optimal for another period.

As an example, Figure 2 shows plots of a T_{max} series collected at a candidate station and of the corresponding reference series (dashed line). It is evident from the analysis of the plot lines that during the periods form 1980 to 1996 and from 2001 to 2009 the two lines are almost coincident. Between 1997 and 2001, there is a sudden jump in the values of candidate station temperatures when compared to the reference series temperatures.

The use of a reference series for a candidate station data series, especially when no metadata is available, enables an easier detection of inhomogeneities. If metadata were available, the reasons for the inconsistency during 1997-2001 in Figure 2 should be possible to explain.

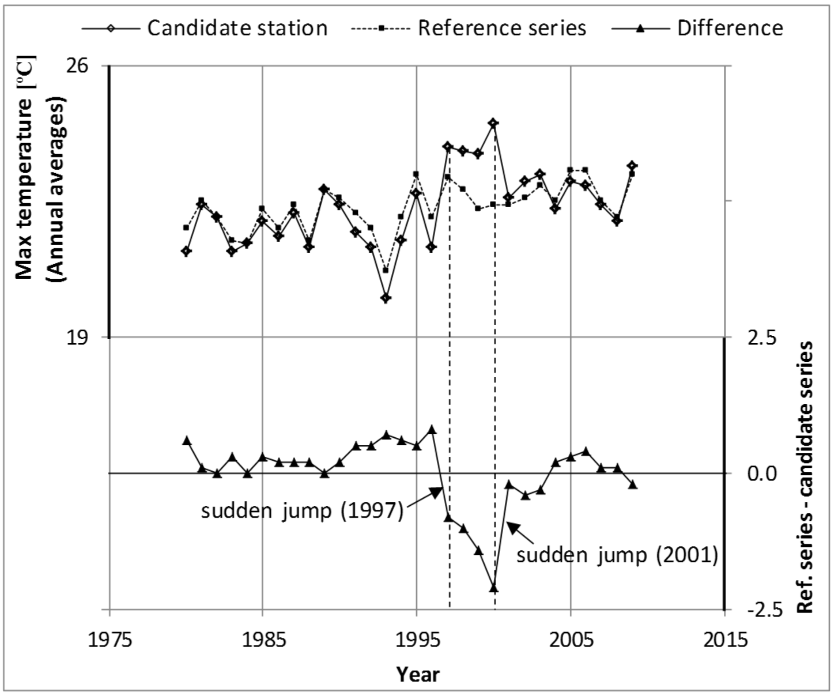


Figure 2 - Example of an inhomogeneous series.

4.3.3 Homogeneity Tests

4.3.3.1 SNHT - Standard Normal Homogeneity Test for Detecting Undocumented Mean Changes

The Standard Normal Homogeneity Test (SNHT) (Alexandersson 1986) is used to test the relative homogeneity of a undocumented (no metadata) candidate time series with respect to a reference time series developed from a group of surrounding neighbor weather stations. Let $\boldsymbol{Y} = \{y_1, y_2,\ldots, y_n\}$ be the candidate meteorological time series and y_i a specific value in moment *i*. $\boldsymbol{X_j} = \{x_{j1}, x_{j2},\ldots, x_{jn}\}$ will denote one of the surrounding neighbor reference *k* sites and x_{ji} a specific value from this site at moment *i*. SNHT may be used to detect relative inhomogeneities ratios, used with precipitation, or relative inhomogeneities differences, used with temperatures.

To detect the relative inhomogeneities ratios (used with precipitation) a time series $\boldsymbol{Q} = \{q_1, q_2,\ldots, q_n\}$ is defined according to:

$$q_i = \frac{y_i}{\frac{\sum_{j=1}^{k} \frac{r_j^2 x_{ji} \bar{Y}}{\bar{X}_j}}{\sum_{j=1}^{k} r_j^2}} \tag{8}$$

To detect the relative inhomogeneities differences (used with temperatures) a time series $\boldsymbol{Q} = \{q_1, q_2, \ldots, q_n\}$ is defined according to:

$$q_i = y_i - \frac{\sum_{j=1}^{k} r_j^2 (x_{ji} - \bar{X}_j + \bar{Y})}{\sum_{j=1}^{k} r_j^2} \tag{9}$$

The denominator in equation (8) and the second term in equation (9) are the reference series. r_j is a positive correlation coefficient between the candidate station and the neighbor station *j*. Bars denote mean values, which are incorporated for normalizing reasons. The normalization causes the q-values to fluctuate around one for equation (8) and around zero for equation (9). It is necessary that the mean values of **Y** and $\boldsymbol{X_j}$ are calculated for one common time period for all $j=1,2,\ldots,k$. Additionally, r_j needs to be estimated from the same common time period for all stations.

The standard normal homogeneity tests are applied to the standardized series:

$$z_i = \frac{q_i - \bar{Q}}{\sigma_Q} \tag{10}$$

Sudden mean shifts or gradual linear mean shifts may be tested using different versions of SNHT test.

The SNHT sudden mean shifts version (SNHTS)

A single shift of the mean level at the candidate site series $\boldsymbol{Y} = \{y_1, y_2, \ldots, y_n\}$ can be expressed as an hypothesis test with the following hypotheses:

H_0: series $\boldsymbol{Y}$ has a constant mean level (no shifts in the mean value)

$$z_i \sim N(0,1), i = 1,2, \ldots, n \tag{11}$$

H_1: series $\boldsymbol{Y}$ has one shift in the mean value (at some unknown time *a*, the mean value changes abruptly, where $\mu_1 \neq \mu_2$)

$$\begin{cases} z_i \sim N(\mu_1, 1), i = 1,2, \ldots, a \\ z_i \sim N(\mu_2, 1), i = a+1, \ldots, n \end{cases} \tag{12}$$

N denotes the normal distribution. The standard deviation is assumed not to change at change-point a. This is a simplification and in fact, it should be slightly different for the series before and after change-point a. Based upon the two hypotheses a test quantity, i.e. a quantity that is the most effective one to separate H_0 from H_1, can be derived. This is usually done by forming a likelihood ratio, i.e. the ratio of the probability that H_1 is correct, given the observed series $\{z_i\}$, to the probability that H_0 is correct. After calculations Alexandersson (1986) obtained the test statistic as:

$$T_{max} = max_{1 \leq a \leq n-1}\, T(a), T(a) = [a\bar{z}_1^2 + (n-1)\bar{z}_2^2] \qquad (13)$$

where $\bar{z}_1^2$ and $\bar{z}_2^2$ are the arithmetic averages of the $\{z_i\}$ series, before and after the shift. The value of *a* corresponding to this maximum is then the moment most probable for the break. If *T* is above a certain critical level one may say that the null hypothesis (of homogeneity) can be rejected at the corresponding significance level. According to Hawkins (1977) there is an increased probability for high T values near the ends of time series where a few low or high values of z_i make *T* large.

The two levels of the ratios or differences before and after the possible change-point *a* are then:

$$\begin{cases} \bar{q}_1 = \sigma_Q \bar{z}_1 + \bar{Q} \\ \bar{q}_2 = \sigma_Q \bar{z}_2 + \bar{Q} \end{cases} \qquad (14)$$

which are reverse uses of equation (10). If one intends to correct data for the period 1, 2,…, *a* then the values within this period should be corrected by $\frac{\bar{q}_2}{\bar{q}_1}$ in the ratio case, equation (8), and by $(\bar{q}_2 - \bar{q}_1)$ in the difference case, equation (9). If the data contains only one shift, then we obtain a homogenized series where all data refer to the present measuring situation[15].

SNHTS for single shifts is often used with a constraint: if a significant break occurs within the five first or last time periods, no corrections should be made because there are too few time periods to be able to obtain a stable correction factor $\frac{\bar{q}_2}{\bar{q}_1}$ or difference $\overline{(q_2 - \bar{q}_1)}$.

[15] A series adjusted to the present measuring situation is a series where the segment before the change-point was adjusted to the pattern of the series after the change-point.

The test for a single shift cannot properly handle series with many breaks. It is easy to generalize the test to two or more breaks (Alexandersson 1995) but an alternative is to use the single shift test on two or more consecutive parts of a complicated series.

It should be noted that for testing absolute homogeneity of undocumented meteorological time series, instead of relative homogeneity, the $\boldsymbol{Q} = \{q_1, q_2, \ldots, q_n\}$ series does not have to be a ratio or difference series defined using a reference series but can be the candidate time series itself, $\boldsymbol{Y} = \{y_1, y_2, \ldots, y_n\}$.

The SNHT gradual linear mean shifts version (SNHTT)

To test if the mean level of the Q-series changes linearly from time *a* to *b* is testing for a trend of arbitrary length of Q-series.

The test hypotheses may be stated as:

H_0: series $\boldsymbol{Y}$ has a constant mean level (no shifts in the mean value)

$$z_i \sim N(0,1), i = 1,2,\ldots,n \qquad (15)$$

H_1: series $\boldsymbol{Y}$ has one shift in the mean value (at some unknown arbitrary time *a*, a change in the mean value takes place gradually, as a linear trend, ending at arbitrary time *b*)

$$\begin{cases} z_i \sim N(\mu_1, 1) & i = 1, 2, \ldots, a \\ z_i \sim N(\mu_1 + (i-a)(\mu_2 - \mu_1)/(b-a), 1) & i = a+1, \ldots, b \\ z_i \sim N(\mu_2, 1) & i = b+1, \ldots, n \end{cases} \qquad (16)$$

Deriving a likelihood ratio, i.e. the ratio of the probability that H_1 is correct, given the observed series $\{z_i\}$, to the probability that H_0 is correct and maximizing it with respect to μ_1, μ_2, a, b gives the test statistic for the trend test (Alexandersson 1986):

$$T_{max} = max_{a,b:\,1 \le a < b \le n}\, T(a,b),$$
$$T(a,b) = [-a\mu_1^2 + 2a\mu_1\bar{z}_1 - \mu_1^2 SB - \mu_2^2 SA + 2\mu_1 SZB + 2\mu_2 SZA - 2\mu_1\mu_2 SAB - (n-b)\mu_2^2 + 2(n-b)\mu_2\bar{z}_2] \qquad (17)$$

Where

$$SA = \sum_{i=a+1}^{b} (i-a)^2/(b-a)^2 \qquad (18)$$

$$SB = \sum_{i=a+1}^{b} (b-i)^2/(b-a)^2 \tag{19}$$

$$SZA = \sum_{i=a+1}^{b} z_i(i-a)\,/(b-a) \tag{20}$$

$$SZB = \sum_{i=a+1}^{b} z_i(b-i)\,/(b-a) \tag{21}$$

$$SAB = \sum_{i=a+1}^{b} (b-i)(i-a)\,/(b-a)^2 \tag{22}$$

$$SL = \frac{(n-b)\bar{z}_2 + SZA}{SA + n - b} \tag{23}$$

$$SK = \frac{-SAB}{SA + n - b} \tag{24}$$

$$\mu_1 = \frac{a\bar{z}_1 + SZB - SL \times SAB}{a + SB + SK \times SAB} \tag{25}$$

$$\mu_2 = \mu_1 SK + SL \tag{26}$$

where $\bar{z}_1^2$ and $\bar{z}_2^2$ are the arithmetic averages of the $\{z_i\}$ series, before and after the trend section. μ_1 and μ_2 must be used in equations (14) to obtain the two fixed levels of $\bar{q}_1$ and $\bar{q}_2$ before and after the trend period. It is wise to require a trend period of more than 5 time periods, to accept it as a real gradual change.

Complicated series with multiple shifts or mixed shifts and trends, are difficult to handle. Such series have to be tested in subsections (Alexandersson 1986).

4.3.3.2 PMT - Penalized Maximal t Test for Detecting Undocumented Mean Changes (PMT)

Maximal two-sample t test, tests the homogeneity of an undocumented (no metadata) shift in the mean of an time series with zero trend and identically and independently distributed (IID) Gaussian errors. Let ***Y*** denote an IID Gaussian time series. A single shift of the mean level at the candidate site series ***Y*** = $\{y_1, y_2, \ldots, y_n\}$ can be expressed as an hypothesis test with the following hypotheses:

H_0: series ***Y*** has a constant mean level (no shifts in the mean value)

$$Y \sim IIDN(\mu, \sigma^2), i = 1,2,\ldots,n \tag{27}$$

H_1: series $\boldsymbol{Y}$ has one shift in the mean value (at some unknown time a, the mean value changes abruptly, where $\mu_1 \neq \mu_2$)

$$\begin{cases} Y \sim IIDN(\mu_1, \sigma^2), i = 1,2, \dots, a \\ Y \sim IIDN(\mu_2, \sigma^2), i = a+1, \dots, n \end{cases} \tag{28}$$

N denotes the normal distribution. The standard deviation is assumed not to change at change-point *a.* Based upon the two hypotheses a test quantity, i.e. a quantity that is the most effective one to separate H_0 from H_1, can be derived. This is usually done by forming a likelihood ratio, i.e. the ratio of the probability that H_1 is correct, given the observed series $\{z_i\}$, to the probability that H_0 is correct. After calculations the test statistic is obtained (Wang et al. 2007, p.917 cites Csörgö & Horváth 1997):

$$T_{max} = max_{1 \leq a \leq n-1} T(a), T(a) = \frac{1}{\hat{\sigma}_a} \sqrt{\frac{a(n-a)}{n}} |\bar{Y_1} - \bar{Y_2}| \tag{29}$$

where

$$\bar{Y_1} = \frac{1}{a} \sum_{i=1}^{a} y_i, \quad \bar{Y_2} = \frac{1}{n-a} \sum_{i=a+1}^{n} y_i \tag{30}$$

and

$$\hat{\sigma}_a^2 = \frac{1}{n-2} \left[\sum_{i=1}^{a} (y_i - \bar{Y_1})^2 + \sum_{i=a+1}^{n} (y_i - \bar{Y_2})^2 \right] \tag{31}$$

This test is called the maximal two-sample t test. T_{max} values can be generated using Monte Carlo simulations. Maximal two-sample t test and SNHTS are equivalent (Wang et al. 2007).

The power of t test decreases considerably when samples are of unequal size (Gardner 1975) and, as a consequence, maximal two-sample t test and SNHTS suffer from the this disadvantage. In a homogeneous time series, points near the end of the series where the sizes of samples to the left and to the right of those points are considerably different have higher probabilities to be mistakenly identified as change-points. Wang et al. (2007) used simulation to estimate change-point false-alarm rate (FAR) as a function of time series length and change-point position *a*. $M_\alpha(a)$ denotes the number of cases where $T(a) > T_{max}(\alpha)$, that is, for which point *a* is mistakenly identified as change-point at α significance level. The false-alarm rate for point a is estimated as:

$$FAR_{\alpha}(a) = \frac{M_{\alpha}(a)}{M} \quad (32)$$

where M is the number of homogeneous IID Gaussian time series generated in the simulation. Simulations showed that FAR graphs are U shaped (Figure 3).

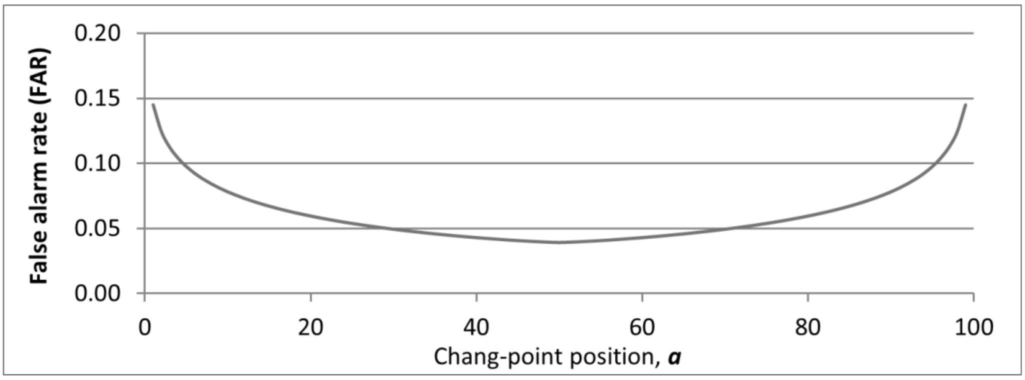


Figure 3 - FAR at significance α = 0.05, as a function on change-point position, a, for time series length of 100.

Identical curves are obtained for $FAR_{\alpha}(a)$ when the SNHTS test statistic is used instead of the maximal two-sample t test statistic. The larger the series length is, the flatter the curves are. The U shape curves indicate that for points near the ends of a homogeneous series, the chance of being mistakenly identified as change-points is much larger than for those in the center of the series. $FAR_{\alpha}(a)$ is the effective level of significance and, ideally, it should be constant and equal to the value of the significance level, α. The U shape of $FAR_{\alpha}(a)$ curves indicates that near the ends of the time series the significance level is larger than the select level for the hypothesis test and that in the center of the series it may be smaller than the select level for the hypothesis test. It is highly desirable to have the same level of significance on the detected change-points regardless of their position in the time series. To overcome the inequality of the significance level with the position of a detected change-point Wang et al. (2007) used a penalty factor *P(a)* to even out the U shape of $FAR_{\alpha}(a)$ curves and proposed the following Penalized Maximal t Test (PMT):

$$PT_{max} = max_{1 \le a \le n-1}\, PT(a)\, , PT(a) = \frac{P(a)}{\sigma_a}\sqrt{\frac{a(n-a)}{n}}\,|\overline{Y}_1 - \overline{Y}_2| \quad (33)$$

The mathematical definition of the penalty factor function *P(a)* is described in Wang et al. (2007).

Comparison of PMT with SNHT

To assess the ability of each method to detect a known shift, Wang et al. (2007) applied PMT and SNHTS methods to two time series, with documented mean shifts, of monthly and annual mean pressure recorded at Burgeo, Canada. Homogeneous highly correlated series, recorded at Yarmouth Airport in Nova Scotia, Canada, were used as the reference series. The hit rates showed that the improvement of PMT over SNHTS ranges from 14% to 25% for detecting small shifts ($\Delta < \sigma$) and up to 5% to detect medium shifts ($\sigma < \Delta < 1.5\sigma$) in time series of length $n < 100$.

Both PTM and SNHTS assume that errors on the time series are IID Gaussian errors. Meteorological data series typically present autocorrelation, periodicity, and trend. Periodicity and trend can be greatly diminished by using a homogeneous reference series that have the same periodicity and trend of the base series. Autocorrelation is not diminished by using a homogeneous reference series and must be accounted using dedicated means (Wang 2008a).

4.3.3.3 TPR3 - Maximal F Test for Detecting Mean Changes without Trend Change

Wang (2003) suggested that instead of the situation where the model is prepared to detect a mean shift that may be accompanied with a trend change, a more common and simpler situation is the detection of a mean shift that is not accompanied with a trend change. This is a two-phase linear regression (Solow 1987; Easterling and Peterson 1995; Vincent 1998 and others) where both trends $\alpha_1 = \alpha_2 = \alpha$. This is a linear regression scheme (TPR3) for a time series $\boldsymbol{Y} = \{y_1, y_2, \ldots, y_n\}$ that can be expressed in the form:

$$y_i = \begin{cases} \mu_1 + \alpha i + \varepsilon_i, i = 1,2,\ldots,a \\ \mu_2 + \alpha i + \varepsilon_i, i = a+1,\ldots,n \end{cases} \quad (34)$$

where ε_i are zero mean independent random Gaussian errors with constant variance.

The existence of a single change-point at the candidate site series $\boldsymbol{Y} = \{y_1, y_2, \ldots, y_n\}$ can be expressed as an hypothesis test with the following hypotheses:

H_0: series $\boldsymbol{Y}$ has no change-point (has constant mean)

$$y_i = \mu + \alpha i + \varepsilon_i, i = 1,2, \ldots, n \tag{35}$$

H_1: series $\boldsymbol{Y}$ has a step change (mean change) at time a

$$y_i = \begin{cases} \mu_1 + \alpha i + \varepsilon_i, i = 1,2, \ldots, a \\ \mu_2 + \alpha i + \varepsilon_i, i = a+1, \ldots, n \end{cases} \tag{36}$$

where μ_1 and μ_2 are the location parameters of the $\boldsymbol{Y}$ series, before and after the shift at time a.

The following test statistic can be derived:

$$F_{max} = max_{1 \le a \le n-1} F(a), F(a) = \frac{(n-3)(SSE_{Red} - SSE_{Full})}{SSE_{Full}} \tag{37}$$

where

$$SSE_{Full} = \sum_{i=1}^{a} (y_i - \hat{\mu}_1 - \hat{\alpha}_1 i)^2 + \sum_{i=a+1}^{n} (y_i - \hat{\mu}_2 - \hat{\alpha}_2 i)^2 \tag{38}$$

$$SSE_{Red} = \sum_{i=1}^{n} (y_i - \hat{\mu}_{Red} - \hat{\alpha}_{Red} i)^2 \tag{39}$$

$$\hat{\alpha}_{Red} = \frac{12 \sum_{i=1}^{n} [(y_i - \bar{Y}) i]}{n(n+1)(n-1)} \tag{40}$$

$$\hat{\mu}_{Red} = \frac{\sum_{i=1}^{n} (y_i - \hat{\alpha}_{Red} i)}{n} \tag{41}$$

Wang (2003) states that results from comparing TPR3 with two-phase linear regression show that TPR3 has a higher power of detection, especially in short length time series.

PMFT - Penalized Maximal F Test for Detecting Mean Changes without Trend Change

The power of maximal two-sample t test, see equation (29), decreases considerably when samples are of unequal size (Gardner 1975) and, as a consequence, maximal two-sample t test and the equivalent SNHTS suffer from the this disadvantage.

TPR3 test has W shape $FAR_{\alpha}(a)$ curves. The W shape of $FAR_{\alpha}(a)$ curves indicates that near the ends of the time series (within the first or last n/10 points) the significance level is larger than the select level for the hypothesis test with lower significance near points 0.22n from either of the ends of the series and increasing moderately in the center of the series.

Similarly to the motivation for the Penalized Maximal t Test (PMT) development, to improve the two-phase linear regression test for detecting a mean change without trend change (TPR3), Wang (2008b) constructed empirically a penalty function and imposed it to the test statistic of TPR3. With this, she was able to overcome the inequality of the significance level with the position of a detected change-point. A penalty factor *P(a)* was used to even out the W shape of $FAR_{\alpha}(a)$ curves (U shape for the $FAR_{\alpha}(a)$ curves of T_{max}) and the following Penalized Maximal F Test (PMFT) test statistic was proposed:

$$PF_{max} = max_{1 \leq a \leq n-1} PF(a) \, , PF(a) = P(a) \frac{(n-3)(SSE_{Red} - SSE_{Full})}{SSE_{Full}} \quad (42)$$

where SSE_{Red} and SSE_{Full} are the same as in equations (38) to (41).

The mathematical definition of the penalty factor function *P(a)* is described in Wang (2008b).

Comparison of TPR3 with PMFT

To assess the ability of each method to detect a known shift, Wang et al. (2007) applied TPR3 and PMFT methods to two time series of monthly mean pressure series, one obtained at Greenwood airport, Canada and the other obtained at Daniels Harbor, Canada. The pressure series from Greenwood airport had a documented change-point near the middle of the series. The pressure series from Daniels Harbor had a documented change-point near the end of the series. The change-point in the former series was detected by PMFT but was not detected by TPR3 showing that PMFT outperforms TPR3 when the change-point is located by the center of the series. The change-point in the latter series was detected by PMFT and by TPR3 indicating that both tests perform similarly when the change-point is located near the ends of the time series. The hit rates showed that the improvement of PMFT over TPR3 can be larger than 10% for detecting small shifts ($\Delta < \sigma$).

Wang (2008a) alerts that autocorrelation is not diminished by using a homogeneous reference series and must be accounted using dedicated means.

4.3.3.4 Techniques for Detecting Changes in Daily Precipitation Series

All the presented methods along with other commonly used methods for the homogenization of meteorological time series are based on the following assumptions: i) identically and independently distributed (IID) Gaussian errors, ii) constant variance across all time period, iii) piecewise linearity: the time series is linear in time both before and after a change-point. Violation of these assumptions may have great impact on the efficiency and even on the validity of the detection procedure. These assumptions are often not met in climate applications.

The presented methods are adequate to homogenize temperature series of annual, monthly, weekly or daily periodicity and precipitation series of annual and monthly precipitation (however, in this case, using a logarithmic transformation of data).

As daily precipitation data are not normally distributed and is highly skewed, a transformation that is able to bring data closer to the above assumptions is necessary. Wang et al. (2010) proposed a Box-Cox power transformation (Box and Cox 1964) which is defined as follows:

$$x_i = h(y_i, \lambda) = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda}, \lambda \neq 0 \\ \log(y_i), \lambda = 0 \end{cases} \tag{43}$$

The estimation of parameter λ is made using an exhaustive search algorithm in range [-1.0, +1.0].

After the Box-Cox transformation of the original daily precipitation series (Wang et al. 2010) uses a stepwise testing algorithm that uses the previously presented Penalized Maximal F Test (PMFT) to test the series for single or multiple change-points and accounts for autocorrelation (Wang 2008a).

4.3.4 The RHtestsV3 package

Several statistical tests exist to detect inhomogeneities in weather and climatic time series but there is no universally agreed best homogenization test. Taking into account test complexity and general performance, Reeves et al. (2007) concluded that the TPR3 maximal F test (Wang 2003) is best for dealing with most climate series.

TPR3 method is one of the methods integrated on the software package RHtestsV3 (Wang 2011) that is freely downloadable and is updated and maintained by the Climate Research Division, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada Toronto, Ontario, Canada. In RHtestsV3 software the TPR3 maximal F test (Wang 2003), the penalized maximal t test (Wang et al. 2007) and the penalized maximal F test (Wang 2008b) are embedded in a recursive testing algorithm (Wang 2008a), with the lag-1 autocorrelation of the time series being empirically accounted for. The RHtests_dlyPrcp software package is similar to the RHtestsV3 package, except that it is specifically designed for homogenization of daily precipitation data series. It is based on the transPMFred algorithm (Wang et al. 2010) which integrates a data adaptive Box-Cox transformation procedure to the original time series. RHtestsV3 and RHtests_dlyPrcp are based on R, a language and environment for statistical computing and graphics that runs on Microsoft Windows (R Core Team 2005).

RHtestsV3 is available from the CCI/CLIVAR/JCOMM Expert Team (ET) at http://etccdi.pacificclimate.org/software.shtml

4.4 Cleaning and Homogenization of the Douro Valley Dataset

4.4.1 Weather Dataset for the Douro Valley

Data characterizing the weather variables in the Douro region for the period 1980-2009 were collected from eight local meteorological stations (Figure 4). Five of those eight stations were used as main stations and the other three were used as auxiliary stations since they contained large segments of missing values. The auxiliary stations were used to supplement the datasets from the five main stations. Daily datasets of Maximum temperature (T_{max}), Minimum temperature (T_{min}), and Precipitation Amount (PREC) were collected. Mean temperatures datasets, T_{avg}, were estimated by averaging T_{max} and T_{min}.

The five main meteorological stations are: *Carrazeda de Ansiães, Mirandela, Pinhão, Régua,* and *Vila Real*. Data from these five stations were provided by *Instituto de Meteorologia de Portugal* (IM), now the *Instituto Português do Mar e da Atmosfera* (IPMA).

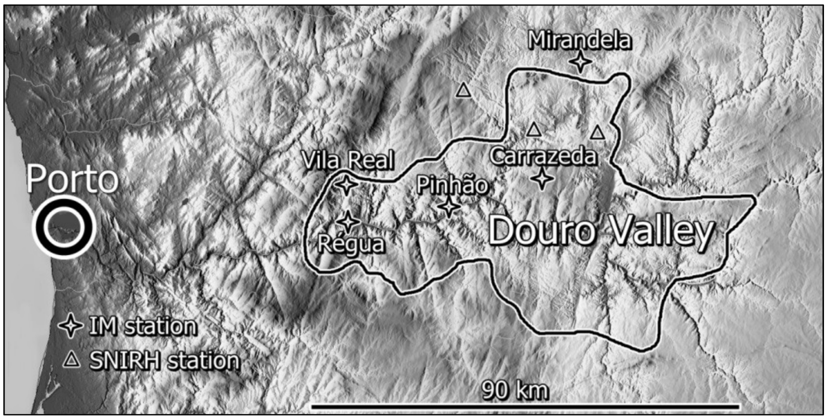


Figure 4 – Location of the meteorological stations (relief map: www.maps-for-free.com).

The three auxiliary meteorological stations are: *Folgares, Junqueira,* and *Minas de Jales*. These stations belong to the National Information System for Water Resources – *Sistema Nacional de Informação de Recursos Hídricos* (SNIRH). Meteorological data from these stations were available until 2011 from SNIRH webpage (http://snirh.pt/index.php?idMain=2&idItem=1). The valid segments of data from these three stations were used in the quality control procedures of the IM data series.

Data provided by IM consists of 10 950 records per variable, per station, summing up 164 250 records. IM data had 4.97% of missing values. SNIRH data had 53.2% of missing values. The proportion of missing values for each weather variable and for each weather station is shown in Table 1.

Table 1 - Proportion of missing values for the meteorological data series.

		Siatema Nacional de Informação de Serviços Hídricos (SNIRH)			Instituto de Meteorologia de Portugal (IM)				
		Folgares	Junqueira	M. Jales	Carrazeda	Mirandela	Pinhão	Régua	Vila Real
Data Series	T_{max}	47.1%	100%	40.0%	10.6%	8.4%	2.4%	2.3%	0.0%
	T_{min}	47.1%	100%	48.0%	10.6%	5.5%	5.0%	2.5%	0.2%
	PREC	11.8%	72.8%	12.3%	17.4%	5.1%	2.0%	2.5%	0.0%

The datasets provided by IM and by SNIRH had no corresponding metadata where it could be possible to check what, if any, quality control tests were previously applied. Moreover, IM did not give any information on the dataset quality. Preliminary analysis of data revealed a huge discrepancy between the 2.72% missing values assumed by IM and the 4.97% observed missing values. As reliable time series are necessary in order to analyze weather variability or climate trends, the datasets provided by IM and by SNIRH had to undergo quality checks to detect and correct inconsistent or missing data as well as to detect and correct existing inhomogeneities. Quality control of meteorological datasets was performed through: i) data cleaning procedures and ii) data homogenization procedures, as previously presented previously in this Chapter.

4.4.2 Cleaning the Douro Valley Meteorological Dataset

The Douro Valley meteorological dataset was cleaned from erroneous values and the cleaned and missing values were replaced with estimated values using the five-step check methodology proposed by (Feng et al. 2004), presented in section 4.2.2. The application of the methodology was conducted using a Microsoft Excel spreadsheet. Three datasets of meteorological data series were cleaned: maximum temperature (T_{max}), minimum temperature (T_{min}), and precipitation amount (PREC).

Step 1: Hi-low extreme check for daily values

Temperature extremes for the Douro Valley region, Portugal, were collected at Wallén (1970) and precipitation extremes for the Douro Valley region, Portugal, were collected at (Brandão et al. 2001): extreme for T_{max} = 42 ºC, extreme for T_{min} = -11 ºC, and extreme for $PREC_{max}$ = 112 $mm/_{24h}$.

T_{max} dataset: 0.0183% of all observations were flagged.

T_{min} dataset: 0.0018% of all observations were flagged.

PREC dataset: no observation was flagged.

Step 2: Internal consistency check

Internal inconsistency:

Tmax dataset: no observation was flagged.

Tmin dataset: 0.5443% of all observations were flagged.

PREC dataset: no observation was flagged.

Excess diurnal temperature range:

Tmax dataset: no observation was flagged.

Tmin dataset: no observation was flagged.

PREC dataset: no observation was flagged.

Flat line check:

Tmax dataset: 0.0128% of all observations were flagged.

Tmin dataset: no observation was flagged.

PREC dataset: no observation was flagged.

Step 3: Temporal outliers check

T_{max} dataset: 0.1114% of all observations were flagged.

T_{min} dataset: 0.1315% of all observations were flagged.

PREC dataset: no observation was flagged.

As an example, Figure 5 shows the boundary lines for T_{max} for the weather station of *Pinhão*. The boundary lines were defined as the daily biweight estimate of mean temperature ± 3 biweight estimates of the standard deviation, see eq. (2) and eq. (3) on page 51. The markers represent the observed daily maximum temperatures for the 30 years period, from 1980 to 2009. The two arrows show some of the observations flagged as temporal outliers.

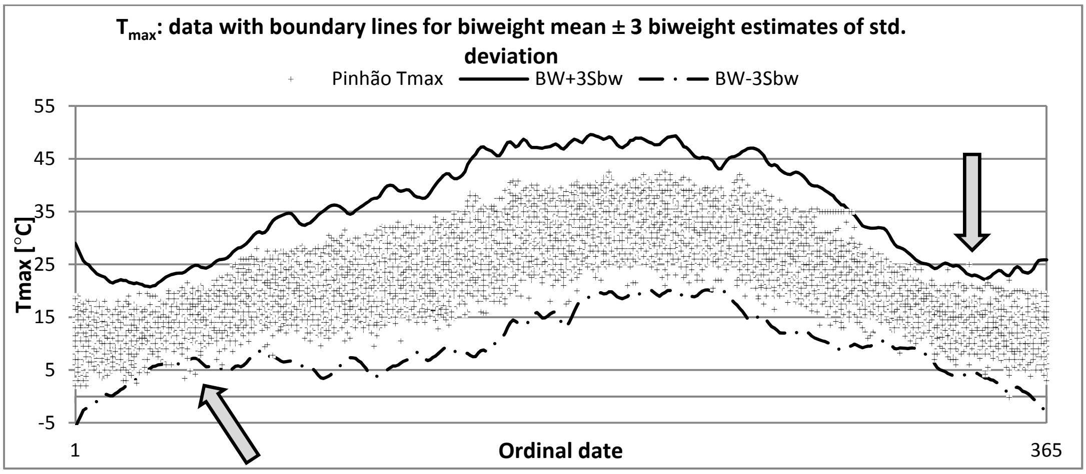


Figure 5 - Tmax observations at *Pinhão* and boundary lines for flagging outliers.

Step 4: Spatial outliers check

T_{max} dataset: 0.3233% of all observations were flagged

T_{min} dataset: 0.06758% of all observations were flagged

PREC dataset: 0.0311% of all observations were flagged

Step 5: Estimation of Missing and Flagged Data

Flagged observations (0.41% of IM data) and missing values (4.97% of IM data) were estimated using a weighted average of the corresponding values from the neighboring meteorological stations with a significant correlation for the estimated variable. Average weights are proportional to the inverse of the Root Mean Square Error (RMSE) for the estimated variable, using eq. (6) on page 52.

4.4.3 Homogenizing the Douro Valley Meteorological Dataset

RHtestsV3 software package (Wang 2011) was used to homogenize the Douro Valley meteorological dataset provided by IM. The process of homogenization of the time series using RHtestsV3 and RHtests_dlyPrcp is mostly automatic.

Temperature Series Homogenization

The functions of RHtestsV3 package can handle annual / monthly / daily series with Gaussian errors with or without metadata support. All the Douro Valley meteorological datasets were provided without the corresponding metadata. RHtestsV3 detects two types of change-points: type 0 and type 1. Type 0 change-points are significant only if they are supported by reliable metadata and type 1 change-points are significant at a pre-set level of confidence (0.95 was used) even without metadata support. Data may have zero or linear trend throughout the whole period of record. The graphic interface of RHtestsV3 incorporates buttons for several functions as shown in Figure 6.

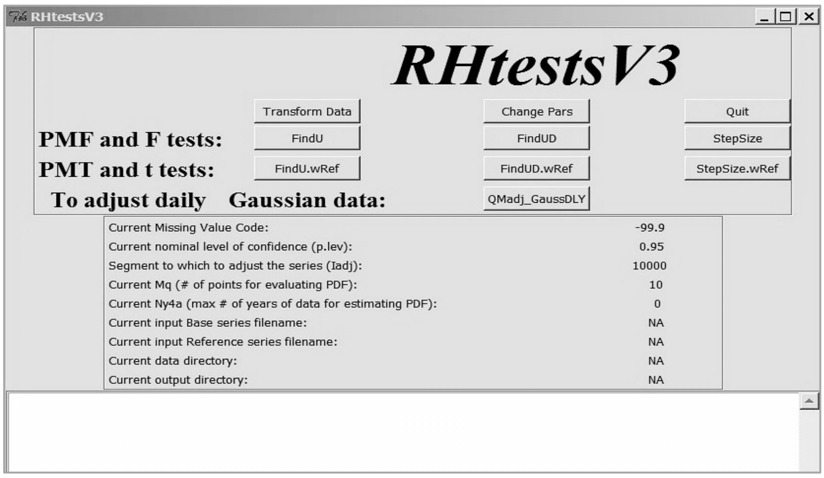


Figure 6 - Graphic Interface of RHtestsV3.

The homogenization procedure for the temperatures series T_{max} and $T_{min,}$ was conducted in six steps:

i) Daily reference series of T_{max} and T_{min} were created for each meteorological station.

ii) Monthly reference series of T_{max} and T_{min} were created for each meteorological station by averaging the reference daily series.

iii) Monthly reference series of T_{max} and T_{min} were tested for homogeneity. Adjustments for data segments with identified change-points were automatically applied.

iv) Daily reference series of T_{max} and T_{min} were tested for homogeneity (using the results of the monthly data homogeneity tests). Adjustments for data segments with identified change-points were automatically applied.

v) Monthly Series of T_{max} and T_{min} were tested for homogeneity using the corresponding homogenized monthly reference series. Adjustments for data segments with identified change-points were automatically applied.

vi) Daily series of T_{max} and T_{min} were tested for homogeneity using the corresponding homogenized daily reference series and using the results of the monthly data homogeneity tests. Adjustments for data segments with identified change-points were automatically applied.

Temperature datasets collected at the weather station located in *Vila Real* showed no signs of inhomogeneities. Temperature datasets collected at the weather station located in *Carrazeda de Ansiães* showed eight change-points, the largest amount observed in all the stations (Figure 7).

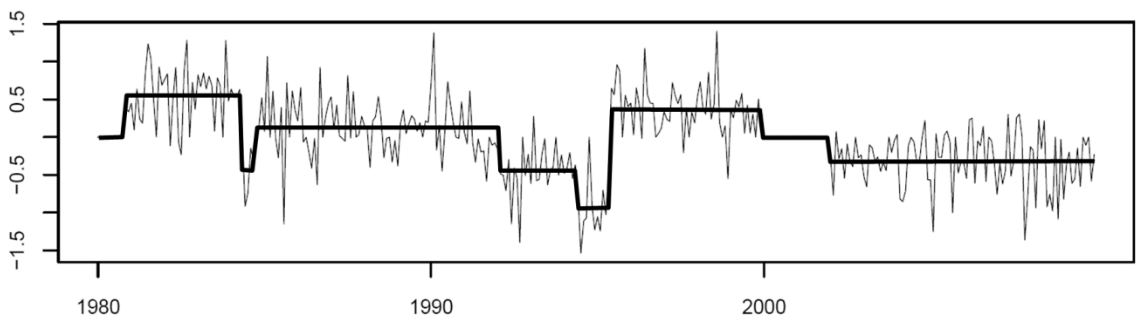


Figure 7 - Tmin monthly series for *Carrazeda de Ansiães* (difference between the *Carrazeda de Ansiães* series and the reference series).

Daily Precipitation Series Homogenization

The RHtests_dlyPrcp software package is similar to the RHtestsV3 package, except that it is specifically designed for the homogenization of daily precipitation data series.

In this case, no reference series was used. The homogenization process is mostly automatic.

Figure 8 is an example of a graph from an output file of RHtests_dlyPrcp revealing two change-points that were detected in the daily precipitation series of *Mirandela*. Change-points occurred at 1986 and 2006. The process of homogenization is similar to the process for temperatures. Figure 9 shows the homogenized series.

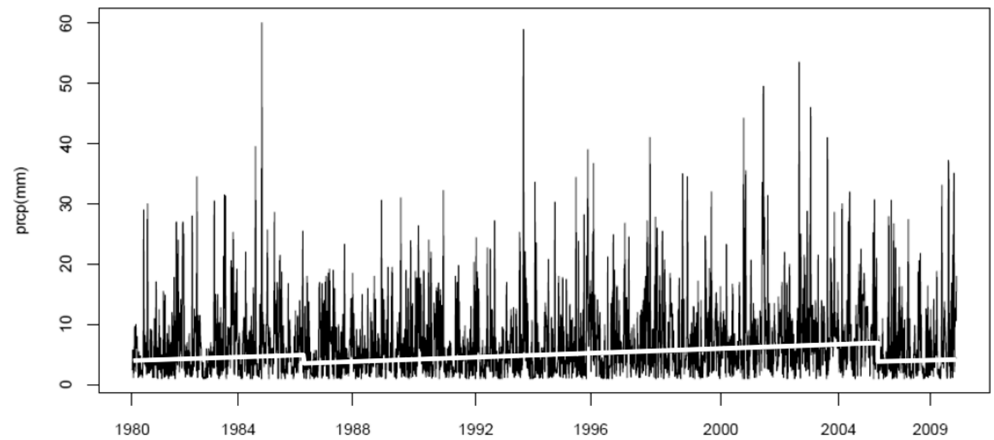


Figure 8 - Mirandela daily precipitation with two change-points.

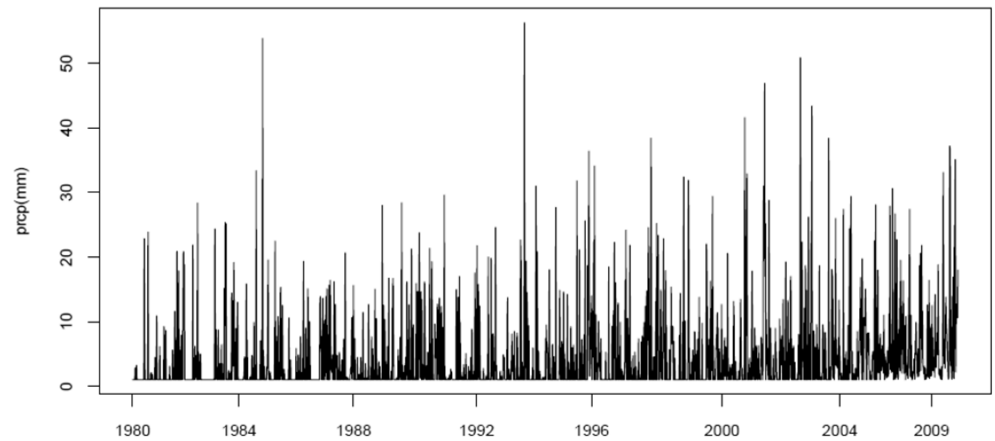


Figure 9 - Mirandela daily precipitation (homogenized).

4.4.4 Comparing Meteorological Raw Data with Quality Controlled Data

The number of missing values in the data for the three weather variables was 4.97% and the overall flagged observations for these variables was 0.52% resulting in 5.49% of observations to be estimated. Flagged observations and missing values were estimated resulting in five cleaned datasets of continuous daily data without missing values for each variable.

The obtained datasets for *Vila Real, Régua*, *Pinhão, Carrazeda de Ansiães,* and *Mirandela* underwent a homogenization procedure through the use of the software package RHtestsV3 (Wang 2011). The homogenization process was automatic, requiring little interaction with the analyst, thus the precise number of corrected segments (data segments between change-points) is not available. However, the number of corrected segments was smaller for the data collected at *Régua*, *Pinhão* and *Mirandela* when comparing to the dataset from *Carrazeda de Ansiães*. The dataset from *Vila Real* showed no signs of inhomogeneities. All change-points were corrected.

Lund & Reeves (2002) state that "*Change-points can substantially alter conclusion made from climatic series. Change-point information is the single most important factor for obtaining an accurate estimate of the linear temperature change rate. Linear change rate is typically on the order of 1 ºC to 2 ºC century*$^{-1}$ *and a single change point can induce a mean temperature shift of a few degrees Celsius*".

Chapter 5

A New Method to Obtain a Consensus Ranking for Vintage Quality

This section is based on: Borges, J., Real, A. C., Cabral, J. S., & Jones, G. V. (2012). **A New Method to Obtain a Consensus Ranking of a Region's Vintages**. Journal of Wine Economics, 7(01), 88–107.

The paper was adapted to the thesis structure and information flow and was extended with supplementary information. The "introduction" section was adapted to the thesis flow and organization. In section 5.2.2 supplementary information on voting systems and on the treatment of the ranking aggregation problem as a problem of optimization is given. In the original paper, the "results" section illustrated the proposed method with its application to the data from three wine production regions: *Piedmont*, *White Burgundy* and *Champagne*. This section title is now "Aggregation of Vintage Port Ratings in 1980-2009" and its content was adapted replacing the data from *Piedmont*, *White Burgundy* and *Champagne* with the data from Vintage Port in 1980-2009.

5.1 Introduction

Wine production is a highly variable agricultural endeavor with yield and quality variations being largely driven by climate (Jones & Davis 2000). Understanding vintage quality variability and its influences are important for the economic sustainability of producers, consumer purchasing decisions, investor portfolio holdings, and researchers examining the myriad of drivers of quality. However, the process of finding an adequate measure of the vintage quality is a challenging task due to the inherent subjectivity in assessing quality. One option is to use the yearly vintage charts published by internationally recognized critics, magazines, or organizations which compare and contrast wines from different properties, different regions, and/or different vintages. Examples of very influential vintage charts are the *Wine Spectator* Vintage Chart (Spectator 2011) or the *Michael Broadbent*'s Vintage Wine Companion (Broadbent 2007). A vintage chart assigns a score to each year representing an overall rating for the quality of the vintage typically for an entire region or a category of wine (i.e., red or white) and, in general, no information is provided on the scores given by the tasting panel to the individual wines tasted.

Vintage ratings have been used in numerous studies examining a wide range of economic, consumer, and scientific topics. For example, Wanhill (1995) found that vintage ratings were significant predictors of the hammer prices for a collection of vintages from a single chateau in Bordeaux. Other research by Landon & Smith (1998) examining Bordeaux wines found that reputation far exceeds current quality (ratings) in terms of the price premium achieved. Also in Bordeaux, Jones & Storchmann (2001) found differences in the sensitivity of ratings between cultivars with Cabernet Sauvignon fruit quality being more influential than Merlot on ratings. Schamel & Anderson (2003) found that regional reputations in Australia and New Zealand have become increasingly differentiated over time and that vintage ratings by James Halliday and Winestate magazine have a highly significant effect on the price premium paid by consumers. Exploiting a delay in the publishing of the ratings by Robert Parker in 2003, Ali et al. (2008) estimated the ‘Parker effect’ to be 2.80 euros per bottle for Bordeaux wines. Gergaud & Ginsburgh (2008) studied the role of technology and *terroir* in wine quality, finding that technological choices in wine production affect quality much more than natural endowments (e.g., aspects of *terroir*). In addition, Gokcekus & Nottebaum (2011) found that consumer scores on wine quality tend to correlate higher with certain experts, but that the correlation between consumer scores and expert ratings are less than those observed between the expert ratings alone.

A large area of study of vintage ratings includes weather and climate relationships. Research using vintage ratings has found that they accurately reflect the weather factors long known to determine wine quality, and ultimately influences market or futures prices (Ashenfelter et al. 1995; Ashenfelter & Jones 2000; Corsi & Ashenfelter 2001). Jones & Davis (2000), examining numerous regions and chateaux in Bordeaux found strong connections between climate, grapevine phenology, fruit composition, and vintage ratings during 1952-1997. Examining vintage ratings for Napa Valley, Nemani, White, Cayan, & Jones (2001) also showed how Wine Spectator ratings impact price with an average rating increase of 10 points (on a 0 to 100 scale), resulting in a 220% increase in price per bottle for the 1995 vintage. Furthermore, Jones, White, Cooper, & Storchmann (2005) found that vintage ratings for 27 wine regions worldwide have shown trends of increasing overall quality with less vintage-to-vintage variation and that vintage ratings are strongly correlated with growing season temperatures. In 24 wine regions in Australia, Sadras, Soar, & Petrie (2007) also found that higher vintage ratings and a reduction in vintage-to-vintage quality variability were related to temperatures during the growing season. While varying some by the region and wine type, the average marginal effect of growing season temperatures shows that a 1.0°C warmer vintage results in an average 13 rating point increase (Jones, White, Cooper, & Storchmann, 2005). Examining climate variability mechanisms Jones & Goodrich, (2008) found significant variability in Wine Spectator vintage-to-vintage ratings and that much of it could be explained by conditions in the El Niño-Southern Oscillation and the Pacific Decadal Oscillation.

The analysis of vintage charts reveals that there is not a widespread consensus on the vintage quality of a given region over the years. Each publisher has its own tasting panel, with its own criteria and perception of quality, which tastes a different set of wines, at different times and conditions. In addition, a variety of rating scales are used. While some publishers use a 5-star rating scale, others use a 10-point or 20-point scale, and still others a 100-point rating scale. The difficulty of combining the judgment of several vintage charts is even bigger when it is observed that some publishers use the same rating scale, but with different criteria. For example, both the Wine Spectator Vintage Chart (Spectator 2011) and the Robert Parker Vintage Guide (Parker 2011) use a 100-point scale in which ratings below 50 are not considered. However, while the former splits the top half of the scale into 7 intervals, the latter splits the same top half of the scale into 6 intervals. As a result, for the Wine Spectator, 95 points correspond to a rating in the top tier while for the Wine Advocate the same rating is in the second tier. Therefore, combining the ratings provided by a set of vintage charts into a single absolute score that represents the production quality of a vintage is a process that has to be based on a set of questionable and arbitrary assumptions. Such assumptions are necessary to define the process of converting every rating scale into a common range of values. Also, it would be difficult to generalize such a process to an arbitrary collection of vintage charts.

In order to assess the degree of consensus among the ratings provided by a set of vintage charts, Table 2 gives the correlation coefficient for the scores given by several vintage charts for the three wine regions that we will use to illustrate our method; DC: Decanter (Decanter 2011); WS: Wine Spectator (Spectator 2011); WA: Wine Advocate (Parker 2011); VC: Vintages (Spirits 2011); AB: Addy Bassin's (Addy Bassin 2011); MB: Michael Broadbent's (Broadbent 2007). For the sake of this example, the correlations were calculated with the original scores, that is, without normalizing the scores. The results show that for the Piedmont region the correlations vary between 0.76 and 0.95, for white Burgundy between 0.47 and 0.80, and for the Champagne between 0.17 and 0.79. The higher the correlation coefficient the higher is the consensus among publishers; the results show that in some cases the consensus is low.

Table 2 - Correlation coefficients for the scores given by several publishers to three wine regions.

Piedmont 1985-2006					
	DC	**WS**	**WA**	**VC**	**AB**
DC	1.00	0.77	0.76	0.77	0.84
WS		1.00	0.95	0.84	0.93
WA			1.00	0.90	0.89
VC				1.00	0.88
AB					1.00

White Burgundy 1982-2005						
	DC	**WS**	**WA**	**VC**	**AB**	**MB**
DC	1.00	0.80	0.61	0.78	0.73	0.53
WS		1.00	0.68	0.75	0.77	0.59
WA			1.00	0.47	0.50	0.52
VC				1.00	0.80	0.53
AB					1.00	0.62
MB						1.00

Champagne 1982-2003				
	DC	**WS**	**VC**	**MB**
DC	1.00	0.17	0.59	0.54
WS		1.00	0.48	0.52
VC			1.00	0.79
MB				1.00

(DC: Decanter; WS: Wine Spectator; WA: Wine Advocate; VC: Vintages; AB: Addy Bassin; MB: Michael Broadbent).

Therefore, we propose the use of a rank aggregation method to combine a collection of vintage chart ratings into a ranking of the vintages that represents the consensus of the input vintage charts. The method takes advantage of the information available from a set of independent sources and combines it into an impartial ranking of a region's vintages over the years. The resulting ranking provides a relative measure of a given region’s vintage quality. The method is general in the sense that it can be used with an arbitrary set of distinct input vintage charts, each having its own ordinal rating scale. We illustrate the method with the scores given by up to six different vintage charts to three different wine regions.

Several papers have been published using rank aggregation methods to study wine classifications. For example, Balinski & Laraki (2011) proposed the Majority Grade method that can be used to induce a ranking of the wines tasted by a given panel of judges using the same classification language. In our context this requirement is not always met since several vintage chart publishers use different rating scales.

The method we propose has the advantage of making use of the information available in the form of vintage charts for a given wine region, each potentially using a different rating scale. Thus, we believe that the proposed method has the potential to be a useful tool for researchers who need an impartial measure of the wine production quality for a given region over the years.

5.2 Materials and Methods

We proposed a method that combines a set of input vintage chart ratings for a wine region into a ranking of the vintages over the years. The combined ranking gives an ordering of the vintage quality that represents the consensus of the ratings given by the set of publishers' vintage chart. The method will be described by means of an example. In Table 3a we give the scores for white wines from the Burgundy region between 1983 and 1988 according to three publishers, Decanter (DC) (Decanter 2011), the Wine Spectator (WS) (Spectator 2011) and the Wine Advocate (WA) (Parker 2011). The analysis of the scores reveals that the three publishers give the top score among the six years to the 1985 vintage. Also, the DC gives an identical score to the 1986 vintage, the WS gives the second best score to 1986 and WA gives only the third best score to that vintage. Thus, we can say that there is a consensus regarding the best year but not regarding the second best year.

Table 3 - An illustration of the conversion of the vintage chart scores (a) into ranks with the scores for white wines from the Burgundy wine region (b). Acronyms are as given in the footnote in Table 2.

(a) The vintage chart scores

	DC	WS	WA
1983	3	85	85
1984	2	78	
1985	4	94	89
1986	4	92	82
1987	2	84	79
1988	3	86	82

(b) The rankings corresponding to the scores

	DC	WS	WA
1983	3	4	2
1984	5	6	6
1985	1	1	1
1986	1	2	3
1987	5	5	5
1988	3	3	3

5.2.1 Converting the Vintage Charts' scores into rankings

The goal of the method is to induce a relative measure of the vintage quality that takes into account the information given by the publishers in an impartial way. One possibility could be to convert the scores given by the publishers into a common scale and to compute the average score. Such a process would require some undesirable and arbitrary assumptions on which scores in one scale correspond to which scores in another scale. We propose to convert the scores given by each publisher into a ranking of the vintages. Therefore, for each publisher we construct a ranking of the years in which the ranks represent their preferences with respect to the vintage quality, originally expressed as a score. The year with the highest score is assigned the top rank and the year with the lowest score the bottom rank. If the same score is assigned to two different years, such years are assigned the same rank (for example 1985 and 1986 from DC in Table 2). We note that the rank of a given year gives the number of years which are better than it plus one. Thus, since in the ranking for DC there are two years tied in the first place, the number two is not assigned to any year, while years 1983 and 1988 are tied in third place. The option to give years the same ranking instead of the common practice of adopting the average rank is due to the more natural interpretation in our particular context of having two years tied in the first place than having two years tied in the 1.5th place. However, the proposed method can be modified to adopt the average rank without any loss of generality.

As a result of this first step, we obtain a set of input rankings such that each ranking represents an ordering of the vintage quality over the years as perceived by the corresponding publisher's tasting panel. Table 3b gives the rankings of the years corresponding to the scores given by each of the publishers.

The only assumption of this first step is that each vintage chart's publisher uses, at least, an ordinal scale and has evaluating criteria that remains stable over the years, in such a way that it is possible compare the perceived quality of the vintages by the comparison of their scores. We believe that this is a reasonable assumption.

We also note that vintage charts occasionally present missing values. Missing values occur if there are years for which a score is not provided. There are several methods available to deal with missing values. One option is to assume that if a provider did not rate a given year it was because it was decided that the perceived quality of the corresponding harvest was not sufficiently good to justify the tasting. In this case such years are assigned the bottom rank. In the context of vintage charts, such an assumption is often reasonable since when some vintages are perceived as uninteresting in the early stages, the publishers decide that its quality does not justify the effort associated with its evaluation. For example, the description of the *Wine Advocate* vintage chart, (Parker 2011) refers to the fact that wines with a score below 50 are not reviewed and *Vintages* (Spirits 2011) states that wines with a score below 4 are not rated. In the example of Table 3, the WA did not provide a score for the year 1984. A closer look shows that 1984 gets low scores by the other two publishers and therefore it is reasonable to assign the bottom rank to 1984.

In cases where the assumption of a missing year being the lowest quality is not defensible, classic methods to deal with missing values can be used. More precisely, if most of the publishers do not provide a score for a particular year that year can be removed from the analysis, since there is not sufficient information to rank that year. In the case where there is a publisher that does not provide a score for most of the years, such publisher could be removed since it does not provide sufficient information for the method. Finally, there is the possibility of filling one particular missing value with the average ranking of the other available publishers for that particular year. In this case it may be necessary to re-rank the years below the year that was filled with the average rank. We note, however, that the method we propose is able to handle any of the above options without any loss of generality.

5.2.2 Aggregating the input rankings into a consensus ranking

The rank aggregation problem is defined as the task of combining many different rank orderings into the ranking that is closest to the set of input rankings (Lin 2010). This is a classic problem from voting theory that has gained interest recently due to its application to the problem of combining the search results of a collection of web search engines. The aggregation of several rankings into one single ranking is, from the point of view of an optimization process, a process that finds a ranking, δ, that minimizes $\sum d(\delta, R_i)$, where d is measure of agreement between rankings. For this purpose, according to Monjardet (1997), there are three ways to quantify the agreement between two rankings: using appropriate correlation coefficients; using a normalized distance between rankings (a "least moves" distance is often appropriate); associating two rankings with a third that represents their agreement and then use the parameters of this agreement structure. At times, these three approaches may coincide as in the case of Kendall tau.

In the research, the second alternative was used, treating the process ranking aggregation as a problem of optimization of a function defined using distances between the rankings. In this sense, we have looked at the ranks of several rankings as a set of points in a metric space. The function to be optimized is the global distance between a set of rankings $R_1, R_2, R_3, \ldots, R_n$ issued by n judges and a candidate ranking, δ. The ranking δ that best represents all the rankings will be the consensus ranking. This is equivalent to the problem of finding the rank $\delta = f(R_1, R_2, R_3, \ldots, R_n)$ that minimizes the global distance $D = \sum d(\delta, R_i)$, to all the n individual rankings.

5.2.2.1 Distances between Full Rankings[16]

A metric space is any set provided with a sensible notion of the "distance" between points.

A nonempty set X with a map d: X × X → R, is called a metric space if the map d has the properties of i) *non-negativity*: $d(x, y) \geq 0$, ii) *symmetry*: $d(x, y) = d(y, x)$, iii) *identity of indiscernibles*: $d(x, y) = 0$ if and only if $x = y$ and iv) the *triangle inequality* $d(x, z) \leq d(x, y) + d(y, z)$ (Shirali & Vasudeva 2006).

Spearman's footrule and Kendall's tau are two well-established distances between full rankings.

[16] Full ranking is a ranking where all the alternatives are ranked in order (no ties)

Spearman's Footrule

Consider a set of K rankings R_1, R_2,..., R_k each with an ordering of the same m alternatives. Let $r^{R_1}(t)$ be the rank of alternative t in ranking R_1. Spearman's footrule is a metric that sums the absolute values of the ranks of all m alternatives in two rankings:

$$F(R_i, R_j) = \sum_{t=1}^{m} \left| r^{R_i}(t) - r^{R_j}(t) \right| \quad (44)$$

As an example, let ranking R_1 = (a, b, c, d) and ranking R_2 = (a, d, c, b). Calculation of the distance between rankings R_1 and R_2, measured by Spearman's footrule is shown in Table 4.

Table 4 - Example of Spearman's footrule calculation.

Element t	$r^{R_1}(t)$	$r^{R_2}(t)$	$\lvert r^{R_1}(t) - r^{R_2}(t) \rvert$
a	1	1	0
b	2	4	2
c	3	3	0
d	4	2	2
$r^{R_1}(t), r^{R_2}(t)$ – rank of t in rankings R_1 and R_2			Σ = $F(R_1, R_2)$ = 4

The distance between rankings R_1 and R_2, measured by Spearman's footrule equals 4. Spearman's footrule can be normalized dividing $F(R_i, R_j)$ by its largest possible value, $\frac{m^2}{2}$, where m is the number of alternatives in both rankings. The normalized Spearman's footrule is given by:

$$\frac{2F(R_i, R_j)}{m^2} \quad (45)$$

Normalized Spearman's footrule metric can be generalized to a distance between a ranking, δ, and a set of k rankings R_1, R_2, R_3,…, R_k, as:

$$F[\delta, (R_1, R_2, R_3, \ldots, R_k)] = \frac{2}{km^2} \sum_{i=1}^{k} F(\delta, R_i) \quad (46)$$

Kendall's tau (τ)

Kendall τ, $K(R_i, R_j)$, is a metric that counts the number of disagreements between the ordering of every pairwise combination of t and u alternatives in two rankings R_i and R_j. If the two elements t and u have the same ordering in both lists, then no penalty is incurred (a good scenario). If the element *t* precedes *u* in the first list and *u* precedes *t* in the second list (or *vice-versa*), then a penalty of one is imposed (a bad scenario).

$$K_{t,u} = \begin{cases} 1, t, u \; with \; = ordering \\ 0, t, u \; with \; \neq \; ordering \end{cases} \quad (47)$$

Kendall τ for two rankings R_i and R_j is given by the following expression:

$$K(R_i, R_j) = \sum_{t,u \in R_i \cup R_j} K_{t,u} \quad (48)$$

As an example, let ranking R_1 = (a, b, c, d) and ranking R_2 = (a, d, c, b). Calculation of $K(R_1, R_2)$ is shown in Table 5.

Table 5 - Example of Kendall's tau calculation.

t, u	$r^{R_1}(t)$	$r^{R_1}(u)$	$r^{R_2}(t)$	$r^{R_2}(u)$	R_1	R_2	$K_{t,u}$
a, b	1	2	1	4	a < b	a < b	0
a, c	1	3	1	3	a < c	a < c	0
a, d	1	4	1	2	a < d	a < d	0
b, c	2	3	4	3	b < c	b > c	1
b, d	2	4	4	2	b < d	b > d	1
c, d	3	4	3	2	c < d	c > d	1
$r^{R_1}(t), r^{R_1}(u)$ – rank of *t* and *u* in ranking R_1 $r^{R_2}(t), r^{R_2}(u)$ – rank of *t* and *u* in ranking R_2							Σ = $K(R_1, R_2) = 3$

For full rankings the Kendall distance is equivalent to the *bubble sort* distance, i.e., the number of pairwise adjacent transpositions needed to transform from one list to the other. The bubble sort makes multiple passes through a list, compares adjacent items and exchanges those that are out of order. Using the same example, R_1 = (a, b, c, d) and ranking R_2 = (a, d, c, b). To transform R_2 into R_1 three passes would be needed: R_2 = (a, d, c, b), pass one $R_{2,1}$ = (a, d, **b**, **c**), pass two $R_{2,2}$ = (a, **b**, **d**, c) and pass three $R_{2,3}$ = (a, b, **c**, **d**) = R_1.

Kendall's τ, can be normalized dividing $K(R_i, R_j)$ by its largest possible value, $\frac{m(m-1)}{2}$, where *m* is the number of alternatives in both rankings. The normalized Kendall's τ is given by:

$$\frac{2K(R_i, R_j)}{m(m-1)} \quad (49)$$

Normalized Kendall's τ metric can be generalized to a distance between a ranking, δ, and a set of *k* rankings $R_1, R_2, R_3, \ldots, R_k$, as:

$$K[\delta,(R_1,R_2,\ldots,R_k)] = \frac{2}{km(m-1)}\sum_{i=1}^{k} K^p(\delta,R_i) \tag{50}$$

The aggregation obtained by optimizing Kendall distance is also called Kemeny optimal aggregation (Kemeny 1959). Kemeny optimal aggregation is, in terms of computational complexity, NP-Hard (Bartholdi et al. 1989).

H. P. Young and Levenglick (1978) demonstrated that Kemeny's rule is the unique preference function that is neutral, consistent and Condorcet.

5.2.2.2 Kendall's Distance between Partial Rankings[17]

A distance measure on partial rankings is a near-metric if there is a constant c, such that for all $n > 1$ and x, z, $x_1,\ldots, x_{n-1}$ in the domain, satisfies the relaxed polygonal inequality: $d(x, z) \leq c[d(x, x_1) + d(x_1, x_2)+\ldots+ d(x_{n-1}, z)]$ (Fagin et al. 2004).

Variations of Kendall's τ metric (eq. (47) may be defined to partial rankings as a generalization of the full ranking metric. Kendall τ distance between partial rankings counts the number of disagreements between the ordering of every pairwise combination of t and u alternatives in two rankings R_i and R_j: if the two elements t and u have the same ordering in both lists, then no penalty is incurred (a good scenario). If the element t precedes u in the first list and u precedes t in the second list (or *vice-versa*), then a penalty of one is imposed (a bad scenario). If in one list u and t are tied and in the other list are not tied, then a soft penalty p is incurred (a not good / not bad scenario).

$$K_{t,u}^{p} = \begin{cases} 0 \; if \; r^{R_i}(t) < r^{R_i}(u), r^{R_j}(t) < r^{R_j}(u) \\ 0 \; if \; r^{R_i}(t) > r^{R_i}(u), r^{R_j}(t) > r^{R_j}(u) \\ 1 \; if \; r^{R_i}(t) < r^{R_i}(u), r^{R_j}(t) > r^{R_j}(u) \\ 1 \; if \; r^{R_i}(t) > r^{R_i}(u), r^{R_j}(t) < r^{R_j}(u) \\ p \; if \; r^{R_i}(t) \neq r^{R_i}(u), r^{R_j}(t) = r^{R_j}(u) \\ p \; if \; r^{R_i}(t) = r^{R_i}(u), r^{R_j}(t) \neq r^{R_j}(u) \end{cases} \tag{51}$$

In equation (51) $r^{R_1}(t), r^{R_1}(u)$ are the ranks of t and u in ranking R_1 and $r^{R_2}(t), r^{R_2}(u)$ are the ranks of t and u in ranking R_2.

Kendall τ for two partial rankings R_i and R_j is given by the following expression:

[17] Partial ranking is a ranking where ties are allowed.

$$K^p(R_i, R_j) = \sum_{t,u \in Ri \cup Rj} K_{t,u}^p \quad (52)$$

Kendall's τ distance can be generalized to a distance between a ranking, δ, and a set of k partial rankings R_1, R_2, R_3,…, R_k, as:

$$K^p[\delta, (R_1, R_2, \dots, R_k)] = \sum_{i=1}^{k} K^p(\delta, R_i) \quad (53)$$

K^p is a metric when $0.5 \le p \le 1$, is a near-metric when $0 < p < 0.5$, and is not a distance measure when $p = 0$ (Fagin et al. 2004).

Herein, we propose the use of a method that meets the Condorcet criterion (Condorcet 1785). Methods based on the Condorcet criterion, rank each candidate by measuring the number of competitors that would be beaten by it in a two candidate election.

In order to respect the Condorcet property, the rank aggregation problem has been defined as the task of minimizing the number of pairwise disagreements between the input rankings and the resulting ranking (Kemeny 1959). This formulation is known as the Kemeny rank aggregation (Young 1988) and it has been shown to verify the Condorcet property (see sections 5.2.2.1 and 5.2.2.2).

We will now give a definition of the rank aggregation problem for our context.

Kendall's τ distance measure $K[\delta, (R_1, R_2, \dots, R_k)]$ defined as in equations (51) to (53) was used. The rank aggregation problem seeks to minimize Kendall's τ distance measure. Since the problem is NP-hard there is no algorithm to find the optimal solution in polynomial time, (Schalekamp & van Zuylen 2009). However, several algorithms to find close to optimum solutions are available. To illustrate the process we adopt the Quicksort with local search optimization approach described in Schalekamp & van Zuylen (2009). In the first step of the algorithm a matrix of weights w, is defined such that $w_{i,j}$, with $i, j: 1, \dots, n$, gives the number of input rankings that prefer i to j. In the second step of the process, a classic Quicksort algorithm (see, Hoare 1962) sorts elements in the ranking in such a way that i is preferred to j when $w_{ij} \ge w_{ji}$ and j is preferred to i when $w_{ij} < w_{ji}$. The final step of the algorithm is a local search that consists of swapping pairs of elements in the ranking given by the Quicksort step and verifies if it improves the Kendall-tau metric. Since the Quicksort algorithm has a random nature, the algorithm should be run several times and the best solution according to the Kendall-tau metric chosen.

The method is applied in a predefined time window. If the analyst needs to extend the time window, for example by including more recent years, the overall ranking has to be recomputed. In such a case, the ranks of the previously existing years may be altered. However, the Condorcet property ensures that the relative positioning of such years is maintained. Finally, we note that the algorithm described above was chosen due to its simplicity but any of the rank aggregation algorithms described in (Schalekamp & van Zuylen 2009) can be used for the second step of the proposed method.

The method can be adapted to incorporate weighting schemes if there is the need to differentiate the relative importance attributed to the publishers.

5.3 Aggregation of Vintage Port Ratings in 1980-2009

Data characterizing the quality of the vintages were collected from two different types of sources listed in Table 6: (i) eight publicly available Vintage-Charts published by renowned institutions or well known individual tasters, (ii) Instituto dos Vinhos do Douro e do Porto (IVDP).

Table 6 - Sources of Vintage ratings.

Source		Rating scale
Berry Bros & Rudd (**BBR**)	(Rudd 2013)	1 - 10
Decanter (**DC**)	(Decanter.com 2013)	1 - 5
Michael Broadbent (**MB**)	(Broadbent 2007)	0 - 5
Sotheby's Wine Encyclopedia (**SWE**)	(Stevenson 2011)	0 - 100
Vintages.com (**VT**)	(Vintages.com 2013)	0 - 10
Wine Enthusiast (**WE**)	(Enthusiast 2013)	50 - 100
Wine Advocate (**WA**)	(Parker 2013)	50 - 100
Wine Spectator (**WS**)	(Spectator 2013)	50 - 100
Instituto dos Vinhos do Douro e do Porto (**IVDP**)	(IVDP 2013)	0 - 1

Vintage-charts assign a score to each vintage representing the overall rating for the quality of several tasted wines of that particular vintage, for a certain type of wine, for a limited wine region. Original vintage-chart ratings for Vintage Port are shown in Table 7.

Table 7 - Original vintage-chart ratings for Vintage Port.

	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09
DC	3		3	4		3		3				4	4		5	1	2	4	3		5	4	2	4	4	5	3	5	3	
WS	90		84	92		93		88				93	94		99	92		96			97			98				98		
WE					81	90	84	85	83	86	85	92	93		96	91	85	93	87	86	90	84	84	96	90	91	89	95	89	95
SWE	85		80	95		95						95	85		95	88		90	80	75	95	86	70	94	88	80	70	95		
VT	6		7	8		9		8				9	9		10	9		10			10	8		10	9	8				
MB	3		4	4		4		3		3	3	4	4		4	3	3	4	3	3	5	3	3	5	4	5				
WA	84		86	92		92						90	95		92			89			92			90				93	90	94
BBR	7		6	8		8						7	8		9			8	6		9			8				9	1	8
IVDP	1	0	1	1	0	1	0	1	0	1	0	1	1	0	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0

The analysis of vintage-charts reveals that there is not a widespread consensus about vintage quality among publishers. Additionally, publishers use a variety of different rating scales and some vintages are not rated by all sources, resulting in missing values.

In order to have all scores in a common [0, 1] scale, the normalized scores corresponding to Table 7 were calculated for each source using: $x_i^{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}}$.

To assess the degree of consensus among the used vintage charts, the pairwise correlation coefficients for the normalized scores of Vintage Port vintages were calculated (Table 8).

Table 8 - Correlation coefficients for the normalized scores of used sources.

	BBR	DC	MB	SWE	VT	WA	WE	WS
BBR	**1.00**	0.66	0.68	0.79	0.74	0.34	0.55	0.89
DC		**1.00**	0.71	0.48	0.33	0.58	0.54	0.63
MB			**1.00**	0.55	0.53	0.41	0.73	0.69
SWE				**1.00**	0.63	0.54	0.67	0.74
VT					**1.00**	0.68	0.76	0.84
WA						**1.00**	0.21	0.60
WE							**1.00**	0.86
WS								**1.00**

Correlation values as low as 0.21were obtained between the scores published by Wine Advocate (WA) and the scores published by Wine Enthusiast (WE). Correlation values as high as 0.89 were obtained between the scores published by Berry Bros & Rudd (BBR) and the scores published by Wine Spectator (WS).

In this section we illustrate the application of the method with the data for the Douro Valley region, for 1980-2009 time period, for Vintage Port. The procedure proposed in sections 5.2.1 and 5.2.2 was applied to the vintage-charts ratings for Vintage Port. The method is generic in the sense that an arbitrary number of publishers using any rating scale could be included. For this experiment we considered that all vintage charts are equally important and, therefore, are assigned the same weight (input scores could be weighted according to the publishers' importance, as perceived by the analyst).

Original vintage-chart ratings for Vintage Port (Table 7) were converted into rankings (Table 9). As most sources do not rate every year, several missing values occur in the rankings.

Table 9 - Rankings corresponding to original ratings.

	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09
DC	12		12	5		12		12				5	5		1	21	19	5	12		1	5	19	5	5	1	12	1	12	
WS	11		13	9		7		12				7	6		1	9		5			4			2				2		
WE					25	10	21	18	24	16	18	7	5		1	8	18	5	15	16	10	21	21	1	10	8	13	3	13	3
SWE	12		14	1		1						1	12		1	9		8	14	17	1	11	18	7	9	14	18	1		
VT	15		14	10		5		10				5	5		1	5		1			1	10		1	5	10				
MB	12		10	4		4		20		12	20	4	10		4	12	12	4	12	12	1	12	12	1	4	1				
WA	13		12	4		4						8	1		4			11			4			8				3	8	2
BBR	10		12	4		4						10	4		1			4	12		1			4				1	14	4
IVDP	1		1	1		1		1		1		1	1		1			1			1			1				1		

Missing values were estimated using the following rule: i) estimated last rank if no source rates the vintage; ii) if at least one source rates the vintage the missing value was estimated as the average of all ranks for that vintage. Resulting values are shown in Table 10.

Table 10 - Rankings with estimated missing values.

	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09
DC	12	30	12	5	25	12	21	12	24	9	19	5	5	30	1	21	19	5	12	15	1	5	19	5	5	1	12	1	12	3
WS	11	30	13	9	25	7	21	12	24	9	19	7	6	30	1	9	16	5	13	15	4	11	17	2	6	6	14	2	11	3
WE	10	30	11	4	25	10	21	18	24	16	18	7	5	30	1	8	18	5	15	16	10	21	21	1	10	8	13	3	13	3
SWE	12	30	14	1	25	1	21	12	24	9	19	1	12	30	1	9	16	8	14	17	1	11	18	7	9	14	18	1	11	3
VT	15	30	14	10	25	5	21	10	24	9	19	5	5	30	1	5	16	1	13	15	1	10	17	1	5	10	14	1	11	3
MB	12	30	10	4	25	4	21	20	24	12	20	4	10	30	4	12	12	4	12	12	1	12	12	1	4	1	14	1	11	3
WA	13	30	12	4	25	4	21	12	24	9	19	8	1	30	4	10	16	11	13	15	4	11	17	8	6	6	14	3	8	2
BBR	10	30	12	4	25	4	21	12	24	9	19	10	4	30	1	10	16	4	12	15	1	11	17	4	6	6	14	1	14	4
IVDP	1	30	1	1	25	1	21	1	24	1	19	1	1	30	1	10	16	1	13	15	1	11	17	1	6	6	14	1	11	3

Each source ranking was re-ranked to accommodate the estimated values for the missing values. Resulting values are shown in Table 11.

Table 11 - Re-ranked rankings.

	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09
DC	14	29	14	6	28	14	25	14	27	13	22	6	6	29	1	25	22	6	14	21	1	6	22	6	6	1	14	1	14	5
WS	15	29	19	12	28	10	26	18	27	12	25	10	7	29	1	12	23	6	19	22	5	15	24	2	7	7	21	2	15	4
WE	11	29	15	5	28	11	24	21	27	19	21	8	6	29	1	9	21	6	18	19	11	24	24	1	11	9	16	3	16	3
SWE	15	29	18	1	28	1	26	15	27	10	25	1	15	29	1	10	21	9	18	22	1	13	23	8	10	18	23	1	13	7
VT	21	29	19	13	28	7	26	13	27	12	25	7	7	29	1	7	23	1	18	21	1	13	24	1	7	13	19	1	17	6
MB	15	29	12	6	28	6	26	24	27	15	24	6	12	29	6	15	15	6	15	15	1	15	15	1	6	1	23	1	14	5
WA	19	29	17	4	28	4	26	17	27	13	25	10	1	29	4	14	23	15	19	22	4	15	24	10	8	8	21	3	10	2
BBR	13	29	17	4	28	4	26	17	27	12	25	13	4	29	1	13	23	4	17	22	1	16	24	4	10	10	20	1	20	4
IVDP	1	29	1	1	28	1	26	1	27	1	25	1	1	29	1	17	23	1	20	22	1	18	24	1	15	15	21	1	18	14

The overall distance, $K^p[\delta,(R_1,R_2,\ldots,R_k)] = \sum_{i=1}^{k} K^p(\delta,R_i)$, where $K^p(\delta,R_i)$ is defined using equation (52). $K^p[\delta,(R_1,R_2,\ldots,R_k)]$ is calculated summing, for each ranking δ candidate to be the consensus ranking, the Kendall's penalties in equation (51). The number of possible candidate rankings, δ, is m! if no ties are allowed (in this research 30! ≈ 2.65e+32) and is $\frac{\mathrm{m}!}{2[\ln(2)]^{(\mathrm{m}+1)}}$ if ties are allowed (in this research ≈ 1.14e+37), as shown by Cameron, Kang, and Stark (2010). In this research, no ties were allowed in the consensus ranking. In the aggregation of the quality of the vintages of Vintage Port, a penalty $p = 0.5$ was used when two vintages had the same ordering (the same rank) in one of the lists, see equation (51).

As the number of possible candidate rankings δ is very large, the problem of finding the optimal solution is NP-hard. To obtain a good solution for the optimization problem the Quicksort algorithm with local search optimization approach described in Schalekamp & van Zuylen (2009) was used.

The best approximation for the Vintage Port consensus ranking δ is shown in Table 12 and Table 13:

Table 12 - Consensus ranking for Vintage Port vintages in 1980-2009, sorted by Vintage.

Vintage	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09
Rank	15	30	17	7	27	8	26	18	28	14	25	9	10	29	1	13	24	6	20	23	3	19	22	4	12	11	21	2	16	5

Table 13 - Consensus ranking for Vintage Port vintages in 1980-2009, sorted by Rank.

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Vintage	94	07	00	03	09	97	83	85	91	92	05	04	95	89	80	08	82	87	01	98	06	02	99	96	90	86	84	88	93	81

This consensus ranking will be used throughout this research as the relative measure of Vintage Port quality in 1980-2009.

5.4 Discussion and Conclusions

While vintage-chart ratings such as those used in this research may not be perfect, consumers have come to use these numbers as a general rule of thumb for purchasing wine. Producers have also incorporated vintage ratings into numerous aspects of the economics of their businesses (e.g., winemakers are given bonuses based upon achieving higher scores) and the marketing of their wines or regions. In addition, much important economic and scientific research is also based upon quality metrics such as ratings. Therefore, the problem of assessing the vintage quality over the years for a given wine region is an important research topic.

In this work, we proposed a method that takes advantage of the numerous vintage charts that are published yearly by renowned wine rating critics, magazines, and organizations. The method converts the ratings of each individual source into rankings and uses a rank aggregation algorithm to combine the input ranking into a consensus ranking. As a result, we are able to produce a ranking of the vintage quality that can be seen as a measure of their relative quality. The ranking represents an impartial consensus of the collection of input vintage charts, in the sense that no assumption is made on how each vintage chart was formulated. The method effectively incorporates input from numerous publishers, using different types of scoring formats (ordinal, interval or ratio), with different scales (e.g., 0-5, 0-10, or 0-100), and with different assumptions with respect to the bottom of the scale and the interval that constitutes an extraordinary wine (see Cicchetti & Cicchetti, 2009). It should be noted that a limit of the method is that it only provides a relative measure instead of an absolute measure of the vintage quality.

To sum up, we believe that the proposed method has the potential to be a useful tool for wine research that requires an impartial assessment of the vintage quality for a given wine region.

The use of this method to obtain a consensus ranking for the vintages of Vintage Port for 1980-2009 was an innovative approach to assess vintage quality in wine studies. In the analyses that will be presented in Chapter 7 the consensus quality ranking for the 1980-2009 vintages is used to assess vintage quality as it represents an impartial consensus of the input vintage-charts.

Chapter 6

Partitioning the Growing Season using Accumulated Heat

This section is based on: C. Real, A., Borges, J., Cabral, J. S., & Jones, G. V. (2014). **Partitioning the Grapevine Growing Season in the Douro Valley of Portugal: Accumulated Heat better than Calendar Dates**. International Journal of Biometeorology, (forthcoming).

The paper was modified and extended with supplementary information. The "introduction" section was adapted to the thesis flow and organization. In section 6.2 supplementary information on grapevine phenology is given. The analysis of the ability of the length of growth intervals with boundaries based on grapevine heat requirements to differentiate the best from the worst vintage years was removed from the original paper text as it will be conducted in Chapter 7. The conclusions section was adapted.

6.1 Introduction

In order to study the yearly weather profiles and their relation with the vintage quality, yield and price it is necessary to split each year's growing season into smaller growth intervals. In these intervals, weather variables may be evaluated and compared to the corresponding values from other years. Three alternatives may be used to define the boundaries of the growth intervals: i) historical dates of the main phenological events for the region, ii) commonly accepted calendar dates, and iii) mean values of the heat requirements of the main phenological events.

The development cycle of the grapevine is usually divided into three major phases: inflorescence development, berry development, and ripening. These three phases are bounded by the phenological events that determine the beginning/ending of each phase (Jones 2013): budburst, flowering, *véraison*, and maturity. The most common means to determine the dates of each phenological event are based on observations of grapevines using the guidance of a growth descriptive system. Several descriptive systems have been used to identify grapevine growth stages: (Baggiolini 1952), (Eichhorn & Lorenz 1977), and BBCH scale for grape (Lorenz et al. 1994).

While there are some long-term observations of grapevine phenology in various places worldwide (Jones 2013), in many regions data are often only collected for one event (e.g., maturity) or likely only for a few years. This is the case for the Douro Valley in that there are no readily available consistent records of the dates of any of the main phenological events for an extended period. In cases such as this, where grapevine phenological event data are not available or are limited, researchers often use calendar defined periods to partition the growing season and to examine weather and/or climate influences (e.g., Corsi & Ashenfelter 2001; Grifoni et al. 2006; Makra et al. 2009; Mattis 2011). Some studies partition the grapevine growing season into smaller intervals using calendar defined weeks or months (e.g., from week *x* to week *y* or from March to June). Other studies partition the season into intervals whose bounds are defined using a calendar simplification of plant phenology, making use of accepted dates when, on average, the main phenological events happen in a region (e.g., budburst by the end of March, flowering by the beginning of June, *véraison* by the end of July and maturity by mid-September). While this method may provide some insight into the relationships between climate, vine growth, production, and quality, it would be arguably better to base the division on plant responses to the weather in a given vintage.

Previous research has shown that measures of accumulated heat help describe grapevine growth in numerous settings and across many varieties (e.g., Lopes et al. 2008; van Leeuwen et al. 2008; Parker et al. 2011; Gladstones 2011 and others). These studies use a thermal time model, based on the observation that each phenological event occurs when a critical amount of accumulated heat above a critical threshold temperature is reached (Bonhomme 2000). While it is generally accepted that 10 ºC is the threshold temperature (Winkler et al. 1974; Huglin 1978; Carbonneau et al. 1992), others have found that this threshold varies by variety, location, the period of vine growth, and the water status of the plants in the season of interest (Jones 2013). To consider dormant period influences, some models incorporate the effect of chilling temperatures during the winter on the breaking of buds in the spring (e.g., Chuine 2000; Cesaraccio et al. 2004; Fila et al. 2012 and others). However, good agreement between phenology dates estimated using the thermal time model and the historical average phenology dates in the Douro Valley, as well as the simplicity of this model justify its choice for this research.

Jones & Davis (2000) suggested the use of grapevine phenological events to define growth periods as they give more insight into the crop/climate relationship than calendar date divisions. Growth intervals boundaries defined using fixed calendar dates are expected to have weak agreement with growth interval boundaries defined using the observed dates. Salazar-Gutierrez et al. (2013) consider that heat accumulated over time provides a more accurate physiological estimate than counting calendar days. We note that, using the region heat summation to define the phenological intervals has been criticized for not taking into account site to site variability that may depend not only on temperature but also on the grape variety, soils, site orientation and water uptake conditions (van Leeuwen et al. 2008) and that, in some cases, heat summations may lack significance in the relation to vine physiology (Jones & Davis, 2000). However, since detailed records on grapevine physiology are not available in general, the use of heat summation to define when each phenological event occurs should be considered a good approximation to define the growth intervals.

The purpose of this research was to analyze differences in temperature and precipitation when using growth intervals with boundaries defined by the estimates of historical dates of the main phenological events (used as reference) and growth intervals with boundaries defined by two methods: method 1 - by mean values of the heat requirements of the main phenological events and method 2 - by generalized calendar average dates associated with the occurrence of the main phenological events.

6.2 Grapevine Phenology

6.2.1 Grapevine Annual Growing Cycle

A grapevine's annual cycle is the natural process of fruit production, dormancy, and regeneration. The process happens in four main growth stages (Jones 2003a): stage 1 – Shoot and Inflorescence Development; stage 2 - Berry development; stage 3 - Ripening; and stage 4 - Senescence. These growth stages are bounded by the four main phenological events: Budburst, Flowering, *Véraison* and Maturity, see Figure 10.

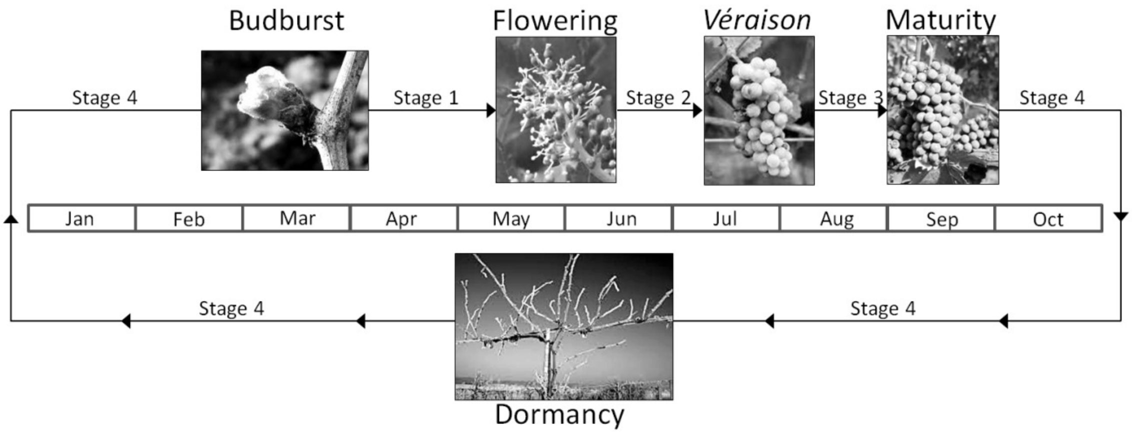


Figure 10 - Stages of grapevine annual growth cycle.

In the Northern Hemisphere, by the end of March, previously dormant buds[18] begin to grow. This event is called budburst. Growth of dormant buds is the result of the commencement of expansion of internodes and leaves pre-formed in the previous season. After budburst, the initial shoot[19] growth is slow (up to the time at which the first 12 leaves have separated). The relatively slow early growth period is followed by a massive growth of shoots and leaves later in spring. It is during this time that lateral shoots may form, adding to the general leafiness of the vine. Shoot growth is greatest just before flowering, after which it declines as the vine begins to direct its energies towards fruit production. Development of individual flower parts starts just before budburst and continues until flowering.

[18] A bud is a growing point that develops in the leaf axil
[19] The shoot consists of stems, leaves, tendrils, and fruit and is the primary unit of vine growth

By the end of May / first weeks of June once the flower parts are mature, flowering occurs (also known as bloom or anthesis). Flowering usually occurs six to ten weeks after the beginning of shoot growth. The process of flowering begins with small flower clusters appearing on the tips of the young shoots. *Vitis vinifera* varieties generally have perfect hermaphrodite flowers. Once the grapevines have flowered, pollination and fertilization can take place. During this stage, vines begin the process of pollination. Grapevines are self-pollinating, so bees and other insects do not appear to play an active role in pollination, but prolonged cool weather or rain can prevent the flowers from pollinating completely or cause them to be fertilized unevenly which can mean the fruit clusters will be sparse, uneven, or in the worst case, non-existent. Rain during bloom can physically inhibit pollination and fertilization by dilution of the stigmatic surface, which is to receive pollen from the flower's anthers. An individual grape inflorescence (flower cluster) contains hundreds of flowers. However, not all of those flowers will set fruit and develop into berries. On average, 40-50 percent of flowers within an inflorescence set fruit and become berries. After fruit set, fleshy grape berries grow and ripen throughout summer until the harvest moment in September or October.

By mid to late July, approximately five to seven weeks after fruit set, *véraison* begins. The color on red cultivars is readily apparent, while the visual indicators of maturity on white cultivars are more subtle. During the next four to six weeks during the berry ripening period, sugar, pigments, and other flavor compounds increase in the maturing fruit, while organic acids decrease.

A berries growth rate is biphasic (Figure 11); one rapid growing phase of about 4 weeks beginning after fruit set, a no growth phase of about two weeks that ends at the beginning of *véraison* and another rapid growing phase of about six weeks (Dokoozlian 2000).

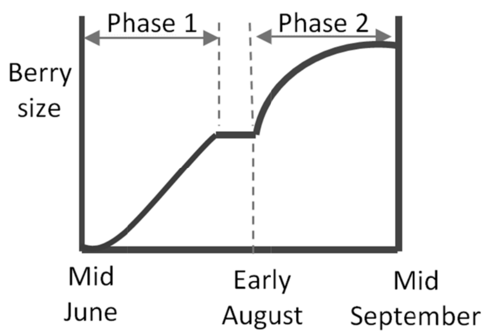


Figure 11 - Change in berries size.

During phase I berries grow through both cell division and cell enlargement, the sugar content of the berries remains low and organic acids[20] accumulate. During lag phase berries growth nearly stops and the organic acids concentration reaches its maximum level. In phase II (ripening phase) berries growth resumes only by cell enlargement. In colored varieties red pigments accumulate in skin, sugar accumulates, organic acid concentrations decline and aroma and flavor components accumulate. In the phase II the formation of flavor compounds is limited by low night temperatures and extreme day temperatures evaporate and degrade these components (Gladstones 2011).

Maturity is the phenological event with a more subjective definition as it depends on the ripeness target that winegrowers would like grapes to achieve, for the type of wine they want to produce. Timing of harvest is a matter of determining the ripeness point that best fits the winemaker’s objective for the wine to be produced. Harvest typically happens by mid-September (in Douro, in 1980-2009, between day 231 and day 285, after January 1) to late October in cooler years or regions.

After grapes are removed during harvest, vines continue the process of photosynthesis, creating carbohydrate reserves to store in the vine's roots and trunks. It will continue doing this until an appropriate level of reserves have been stored. Leaves change color from green to yellow. As temperatures further decrease, they undergo a number of changes in preparation for ‘shutting down’ for the colder months: levels of water in various tissues decrease, soluble proteins in bark increase, enzymes adjust their make-up to withstand temperature changes, and cell membranes alter their function. Basically vines set themselves up with the biological equivalent of ‘anti-freeze’ to ensure live tissue remains for the renewal of growth in the following spring. If vine tissues freeze, the cells can explode or damage cell contents or membranes, and enzymes and other proteins that control metabolic functions can be destroyed. Freezing damage to buds can affect the vine’s growth and fruitfulness over the coming season. After this modification process the vine will go, as a whole, into dormancy; the state of bud dormancy that is of most interest for vine management. Vines are not dead or completely inactive when dormant. They do not photosynthesize as they have no leaves, but they are breathing to maintain basic metabolic functions.

[20] Organic acids in grape berries: tartaric and malic acids (69 to 92% of all organic acids) and minor amounts of citric, succinic, lactic and acetic acids (Conde et al. 2007)

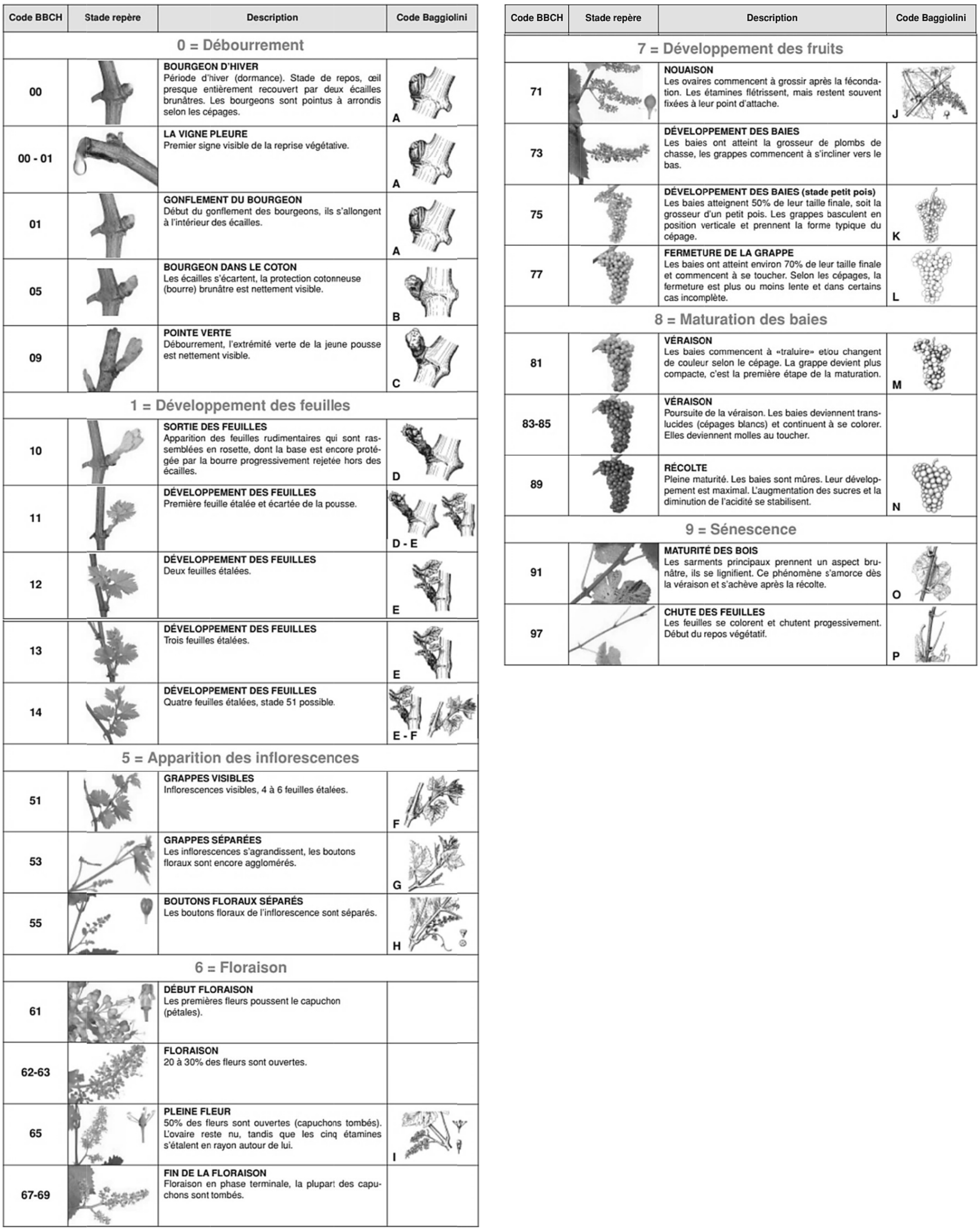
As temperatures warm in late winter, stored starch is converted to sugar, sap begins to move in the vine and pruning wounds begin to “bleed". As temperatures warm, buds begin to swell, then burst (break). It is the end of dormancy and the beginning of a new cycle.

6.2.2 Phenology Descriptive Systems

The most common means to determine the dates of each phenological event are based on the observation of the grapevines using the guidance of a growth descriptive system. Several descriptive systems may be used on the identification of grapevine growth stages: Baggiolini (1952), Eichhorn & Lorenz (1977), and the BBCH scale for grape vines (Lorenz et al., 1994). Descriptive systems provide a sequence of distinctive grapevine development elements, clearly recognized, described in an unambiguous and widely understood language that allow the identification of each stage (Coombe 1995) as shown in Figure 12.

Users of descriptive systems may want descriptions of a limited number of major stages or a detailed set of precisely defined stages. The Baggiolini system was the first to be proposed and was widely adopted because of the clear sketches and clear description of ten stages between budburst and setting. The Eichhorn & Lorenz system was a more comprehensive system with 47 numbers corresponding to an equal number of stages characterized by silhouette drawings and accompanying word descriptions.

The BBCH system was derived from a proposal within the European Union to adopt a uniform code. In the adaptation of the BBCH system to the grapevine, nine macro stages are used, from germination / budburst to senescence / dormancy. Within each macro stage, up to 10 secondary stages are defined with numbers from 0 to 99. The BBCH system is being advocated for use in European countries.



Code BBCH	Stade repère	Description	Code Baggiolini
0 = Débourrement			
00		**BOURGEON D'HIVER** Période d'hiver (dormance). Stade de repos, œil presque entièrement recouvert par deux écailles brunâtres. Les bourgeons sont pointus à arrondis selon les cépages.	A
00 - 01		**LA VIGNE PLEURE** Premier signe visible de la reprise végétative.	A
01		**GONFLEMENT DU BOURGEON** Début du gonflement des bourgeons, ils s'allongent à l'intérieur des écailles.	A
05		**BOURGEON DANS LE COTON** Les écailles s'écartent, la protection cotonneuse (bourre) brunâtre est nettement visible.	B
09		**POINTE VERTE** Débourrement, l'extrémité verte de la jeune pousse est nettement visible.	C
1 = Développement des feuilles			
10		**SORTIE DES FEUILLES** Apparition des feuilles rudimentaires qui sont rassemblées en rosette, dont la base est encore protégée par la bourre progressivement rejetée hors des écailles.	D
11		**DÉVELOPPEMENT DES FEUILLES** Première feuille étalée et écartée de la pousse.	D - E
12		**DÉVELOPPEMENT DES FEUILLES** Deux feuilles étalées.	E
13		**DÉVELOPPEMENT DES FEUILLES** Trois feuilles étalées.	E
14		**DÉVELOPPEMENT DES FEUILLES** Quatre feuilles étalées, stade 51 possible.	E - F
5 = Apparition des inflorescences			
51		**GRAPPES VISIBLES** Inflorescences visibles, 4 à 6 feuilles étalées.	F
53		**GRAPPES SÉPARÉES** Les inflorescences s'agrandissent, les boutons floraux sont encore agglomérés.	G
55		**BOUTONS FLORAUX SÉPARÉS** Les boutons floraux de l'inflorescence sont séparés.	H
6 = Floraison			
61		**DÉBUT FLORAISON** Les premières fleurs poussent le capuchon (pétales).	
62-63		**FLORAISON** 20 à 30% des fleurs sont ouvertes.	
65		**PLEINE FLEUR** 50% des fleurs sont ouvertes (capuchons tombés). L'ovaire reste nu, tandis que les cinq étamines s'étalent en rayon autour de lui.	I
67-69		**FIN DE LA FLORAISON** Floraison en phase terminale, la plupart des capuchons sont tombés.	
7 = Développement des fruits			
71		**NOUAISON** Les ovaires commencent à grossir après la fécondation. Les étamines flétrissent, mais restent souvent fixées à leur point d'attache.	J
73		**DÉVELOPPEMENT DES BAIES** Les baies ont atteint la grosseur de plombs de chasse, les grappes commencent à s'incliner vers le bas.	
75		**DÉVELOPPEMENT DES BAIES (stade petit pois)** Les baies atteignent 50% de leur taille finale, soit la grosseur d'un petit pois. Les grappes basculent en position verticale et prennent la forme typique du cépage.	K
77		**FERMETURE DE LA GRAPPE** Les baies ont atteint environ 70% de leur taille finale et commencent à se toucher. Selon les cépages, la fermeture est plus ou moins lente et dans certains cas incomplète.	L
8 = Maturation des baies			
81		**VÉRAISON** Les baies commencent à «traluire» et/ou changent de couleur selon le cépage. La grappe devient plus compacte, c'est la première étape de la maturation.	M
83-85		**VÉRAISON** Poursuite de la véraison. Les baies deviennent translucides (cépages blancs) et continuent à se colorer. Elles deviennent molles au toucher.	
89		**RÉCOLTE** Pleine maturité. Les baies sont mûres. Leur développement est maximal. L'augmentation des sucres et la diminution de l'acidité se stabilisent.	N
9 = Sénescence			
91		**MATURITÉ DES BOIS** Les sarments principaux prennent un aspect brunâtre, ils se lignifient. Ce phénomène s'amorce dès la véraison et s'achève après la récolte.	O
97		**CHUTE DES FEUILLES** Les feuilles se colorent et chutent progessivement. Début du repos végétatif.	P

Figure 12 - Main stages of BBCH scale for grape descriptive system and correspondence to all 16 (A to P) Baggiolini stages (Source: station de recherche Agroscope Changins-Wädenswil, www.acw.admin.ch).

6.3 Data and Methods

This section describes the data and the methods used to divide the growing season into four smaller growth intervals bounded by January 1 and by the main phenological events of budburst, flowering, *véraison* and maturity.

6.3.1 Weather Data

All meteorological data series used in this research were examined and cleaned of erroneous values using the methodology proposed by Feng et al. (2004) presented in section 4.2.2 and homogenized by using a software package for data homogenization - RHtestsV3 (Wang 2011) presented in section 4.3.4. Mean temperatures data series were obtained by averaging maximum and minimum temperatures. Grapevine heat requirements to reach each key phenological stage were defined in growing degree-days (GDD) – the sum of the average daily temperature above a threshold temperature from January 1 to a given date: $\mathrm{GDD}\,[°\mathrm{C}] = \sum_{i=1\,(\mathrm{January}\ 1)}^{\mathrm{O.Date}} \left(\mathrm{T_{avg}} - \mathrm{T_{base}}\right)$, T_{avg} is the average temperature in °C and T_{base} is a temperature used as threshold. While a threshold temperature of 3.5 ºC has been used for budburst (Lopes et al. 2008), 10 ºC is the most commonly used base temperature in viticulture

6.3.2 Consensus Ranking for Vintage Port

In order to select the best and the worst vintages for Port wine in the 1980-2009 period the consensus ranking method proposed in Borges et al. (2012) was used (see Chapter 5). The consensus ranking for Port wine vintages during 1980-2009 was obtained using this method and is shown in Table 14 was used as a relative measure of vintage quality.

Table 14 - Vintage quality consensus ranking for Port wine in 1980-2009.

Vintage	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09
Rank	15	30	17	7	27	8	26	18	28	14	25	9	10	29	1	13	24	6	20	23	3	19	22	4	12	11	21	2	16	5

6.3.3 Grapevine Phenological Data

In this section, the approach used to estimate the yearly dates of the main phenological events from the historical average dates and the two methods used to divide the growing season into smaller intervals are described.

6.3.3.1 Observations: Characterizing the Main Phenological Events from Estimates of Historical Dates

Data covering the phenology of grapevine for the entire region over a long time period are not available for the Douro Valley. However, the research was able to obtain the average observed dates of the main events for the city of *Régua* (Fig. 2) from ADVID – *Associação para o Desenvolvimento da Viticultura Duriense* (ADVID 2012), which are shown in Table 15 (no information is given on the number of years used in the averages nor on the phenological scale used). In addition, this research also used the observed dates of the main events for the *Touriga Franca* variety, in 2001-2012, at *Quinta de Santa Bárbara* (QSB), located in *Pinhão* (Fig. 2) and the values are also presented in Table 15.

Table 15 - The average dates of the main phenological events obtained from ADVID and from QSB.

	ADVID		QSB	
Event	**Average Ordinal Date**	**Calendar Date**	**Average Ordinal Date**	**Calendar Date**
Budburst	75	March 16	83	March 24
Flowering	140	May 20	143	May 24
Véraison	191	July 10	205	July 25
Maturity	247	September 4	246	September 3

The values show that the average dates are quite similar for budburst, bloom, and maturity, yet there is a difference of 14 days for *véraison*. However, since the ADVID records are referenced to a larger area and represent an average value that incorporates several grape varieties, as opposed to the QSB data that represent a single grape variety, the ADVID data were used as the reference values for the main phenological events in the Douro Valley in this research. To produce yearly dates of the main phenological events for the region, the average heat accumulation, needed to reach each event was used. Heat accumulation values are expressed in GDD – cumulative Growing Degree Days.

ADVID's average phenology dates correspond to the following average heat accumulations: budburst - 60 GDD, flowering – 400 GDD, *véraison* – 1100 GDD, and maturity (harvest) – 1750 GDD.

6.3.3.2 Method 1 - Characterizing the Main Phenological Events from Experimental Heat Requirements of Representative Grape Varieties

Lopes et al. (2008) and van Leeuwen et al. (2008) studied the heat requirements for several grape varieties. While the van Leeuwen et al. (2008) study covered the most widely planted varieties in the world collected throughout Europe, the Lopes et al. (2008) study focused on 34 varieties of the Portuguese Ampelographic Collection that includes the main winegrape varieties grown in the Douro Valley. The van Leeuwen et al. (2008) study collected data from a wide range of cultivars in many winegrowing regions, mostly in Europe, over many vintages. For bud break, GDD were calculated when 50% of the buds reached Baggiolini's B stage (Baggiolini 1952). For flowering, GDD were calculated when 50% of the flowers were open. For *véraison*, GDD were calculated when 50% of the berries changed color (red varieties) or softened (white varieties). Harvest dates in regions where each specific variety is widely planted were treated as the date of maturity for analysis. The Lopes et al. (2008) study collected data from Quinta da Almoinha, Estação Vitivinícola Nacional just north of Lisbon (39º 02' N and 9º 11' W) during 1990-2006 observing phenology and climate protocols of the OIV (1983). The heat requirements in both studies are presented in Table 16.

The only comparable grape variety in the two studies is *Tinta Roriz* (*Tempranillo*). The results show that the heat requirements to the budburst and flowering events are equivalent according to the two sources (budburst GDD = 50 and flowering GDD = 355). In addition, the heat requirements for the *véraison* event are very similar (1030 GDD vs 1027 GDD, $\Delta \approx 0.3\%$). The value for maturity is not given in van Leeuwen et al. (2008).

Table 16 - Heat requirements for common grape varieties grown in the Douro Valley.

			Heat requirement to reach each key phenological stage [GDD]											
Source	Phenological Event	T_{base} : threshold temperature [°C]	Tinta Roriz (Tempranillo)	Σ	Tinta Barroca	Σ	Tinto Cão	Σ	Tinta Amarela (Trincadeira)	Σ	Touriga Franca	Σ	Touriga Nacional	Σ
Lopes et al. (2008)	Budburst	3.5	582	582	544	544	545	545	614	614	528	528	552	552
	Budburst*	10.0	50	50	47	47	48	48	54	54	43	43	45	45
	Flowering	10.0	305	355	300	347	294	342	311	365	292	335	293	338
	Véraison	10.0	675	1030	620	967	771	1113	715	1080	618	953	778	1116
	Maturity	10.0	486	1516	617	1584	470	1583	559	1639	673	1626	535	1651
van Leeuwen et al. (2008)	Budburst	3.5												
	Budburst	10.0	50	50										
	Flowering	10.0	305	355										
	Véraison	10.0	722	1027										
	Maturity	10.0	-	-										

* T_{base}=10.0 ºC estimated values from T_{base}=3.5 ºC values

The agreement of the two sources concerning the heat requirements for these three events suggests that, in general, the two sources are comparable. We note that the *Tinta Roriz* variety represents 12% of the Douro Valley vineyard area while the most planted variety, *Touriga Franca*, represents 22% of the Douro Valley vineyard area (Magalhães 2003; Copello 2010). Figure 13 shows the relative heat requirements for the six grape varieties most commonly grown in the region according to Lopes et al. (2008). The two main varieties are each indicated separately while the other four are not individually identified.

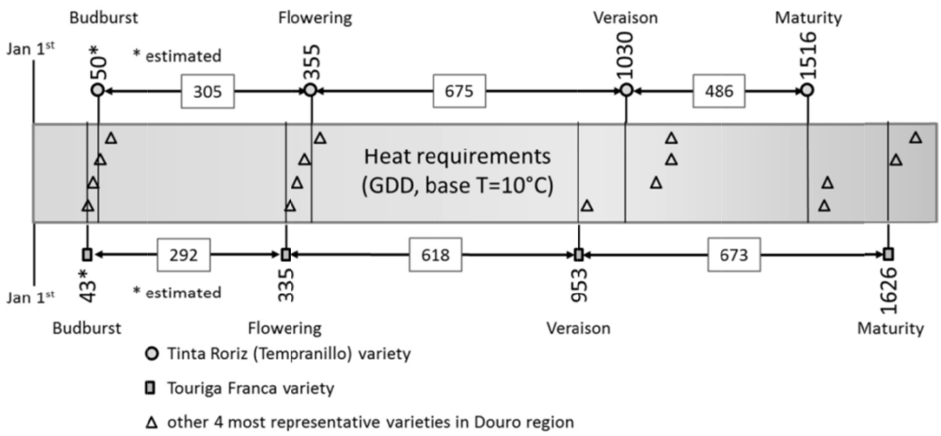


Figure 13 - Relative heat requirements for the six recommended varieties for Port wine.

For defining the heat boundaries for growth intervals, analysis of Figure 13 suggests that *Tinta Roriz* is the variety that best represents the Douro Valley's most planted varieties for the first three phenological events, but not so for maturity. *Tinta Roriz* is an early maturity variety that ripens much earlier than the average of the other varieties. In fact, Figure 13 shows that the heat requirements to maturity of the *Touriga Franca* variety better represents the region since it is the most planted variety and its heat requirement value is close to the average heat requirement of all six varieties for maturity (1608 GDD). Thus, we chose to use GDD=1626, the *Touriga Franca* value, as the heat requirement for maturity in the region. In summary, the heat requirements used as representative for all varieties in the Douro Valley are: budburst - 50 GDD, flowering – 355 GDD, *véraison* – 1030 GDD, and maturity – 1626 GDD.

6.3.3.3 Method 2 - Characterizing the Main Phenological Events from Average Calendar Dates

When there are no historical phenological data available or reference heat requirement values for a region, a common approach is to consider the generally accepted average dates for the main phenological events. For example, it is generally accepted that in the Northern Hemisphere the maturity event happens, on average, between the middle of September to the middle of October depending on region and variety. For simplicity Northern Hemisphere average dates were used as a reference of the Douro Valley's boundary dates for the growth intervals (Jones 2013). Ordinal dates and the corresponding calendar dates for the main phenological events are presented in Table 17.

Table 17 - Generally accepted average dates for main phenological events in the Northern Hemisphere.

Event	Average Ordinal Date	Calendar Date
Budburst	91	April 1
Flowering	161	June 10
Véraison	206	July 25
Maturity	258	September 15

6.3.3.4 Summary of the Two Methods

This section presents a summary of the methods presented in the previous sections. Each vintage growing season was partitioned into four smaller growth intervals. The last three growth intervals are coincident with the three major phases of grapevine development. These intervals are based on the main phenological events and include: 1) End of Dormancy Interval – the time from the beginning of January to budburst, 2) Inflorescence Development Interval – the time from budburst to flowering, 3) Berry

Development Interval – the time from flowering to *véraison*, and 4) Ripening Interval – the period from *véraison* to maturity. A summary of growth interval boundaries according to each of the methods is given in Table 18.

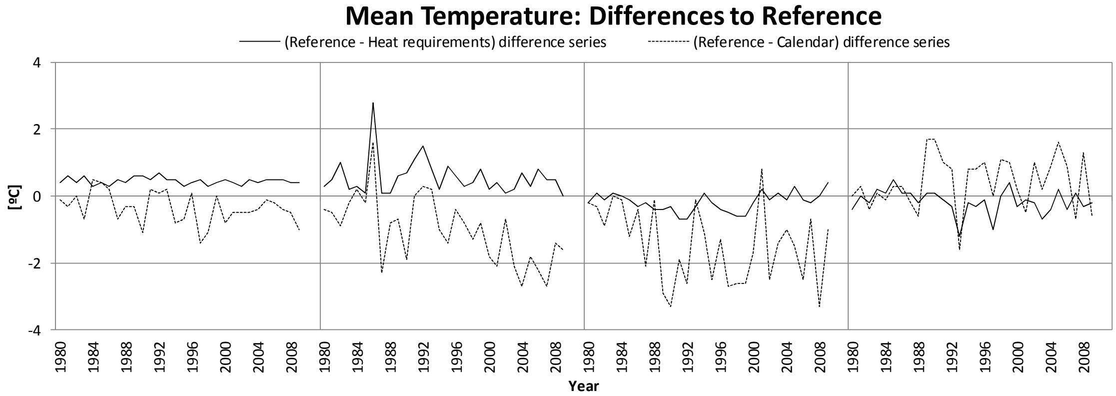
Table 18 - Summary for growth intervals boundaries based on GDD (Method 1) and calendar dates (Method 2).

			Growth Interval				
			End of Dormancy	Inflorescence Development	Berry Development	Ripening	
Method	**Boundaries**	**Units**	**Jan 1**	**Budburst**	**Flowering**	***Véraison***	**Maturity**
Reference	Estimated yearly dates	GDD	0	60	400	1100	1750
1	Cultivar heat requirements	GDD	0	50	355	1030	1626
2	Calendar dates	Date	1	91	161	206	258

6.4 Comparing the Methods for the Definition of Growth Intervals Boundaries

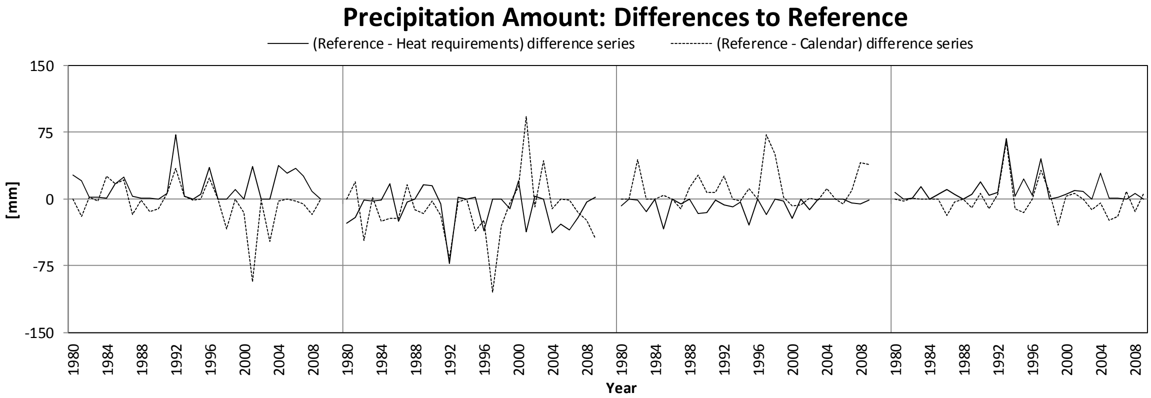
For each year during 1980-2009, the growing season was partitioned into the four growth intervals (End-of-Dormancy, Inflorescence Development, Berry Development and Ripening) using the previously described methods. For each growth interval, two weather variables commonly used in characterizing weather profiles - mean temperature and precipitation amount - were assessed. The values obtained from yearly historical data were used as reference values. The values obtained from the common varieties experimental heat requirements (Method 1) and the values obtained from average calendar dates (Method 2) were compared to the corresponding reference values.

Figure 14 shows, for each growing interval and for a given year, the difference between the average temperature when the interval is defined from historical phenology data (reference) and when the interval is defined by each of the two alternative methods (Method 1 and Method 2). The results show that the differences in the mean temperatures are, on average, much smaller and have smaller variability when interval boundaries are defined from the heat requirements (Method 1) than when using calendar dates (Method 2). Calendar dates produce the greatest deviation for the inflorescence and berry development intervals, significantly overpredicting the temperatures during these intervals. Moreover, for the ripening interval, calendar dates tend to underpredict the temperatures. Similarly, Figure 15 shows that differences in precipitation are, on average, smaller and have smaller variability when interval boundaries are defined by using Method 1 than when using Method 2.



a) End-of-Dormancy b) Inflorescence Development c) Berry Development d) Ripening

Figure 14 - Mean temperature differences for Methods 1 and 2 growth intervals from reference growth intervals.



a) End-of-Dormancy b) Inflorescence Development c) Berry Development d) Ripening

Figure 15 - Precipitation amount differences for Methods 1 and 2 growth intervals from reference growth intervals.

Figure 16 shows the yearly mean temperature in each growth interval when computed using intervals with boundaries defined according to the method based on historical phenology data *vs* intervals with boundaries defined according to each of the two alternative methods. For each growth interval, regression lines were plotted for the series representing historical data (reference) *vs* heat requirements (Method 1); and for the series representing historical data (reference) *vs* calendar average dates (Method 2). Results show a very high level of association between the temperature series obtained from growth intervals with boundaries defined using historical phenology dates and corresponding temperature series obtained from growth intervals with boundaries defined using experimental heat requirements (R^2 of 0.99, 0.56, 0.96, and 0.91).

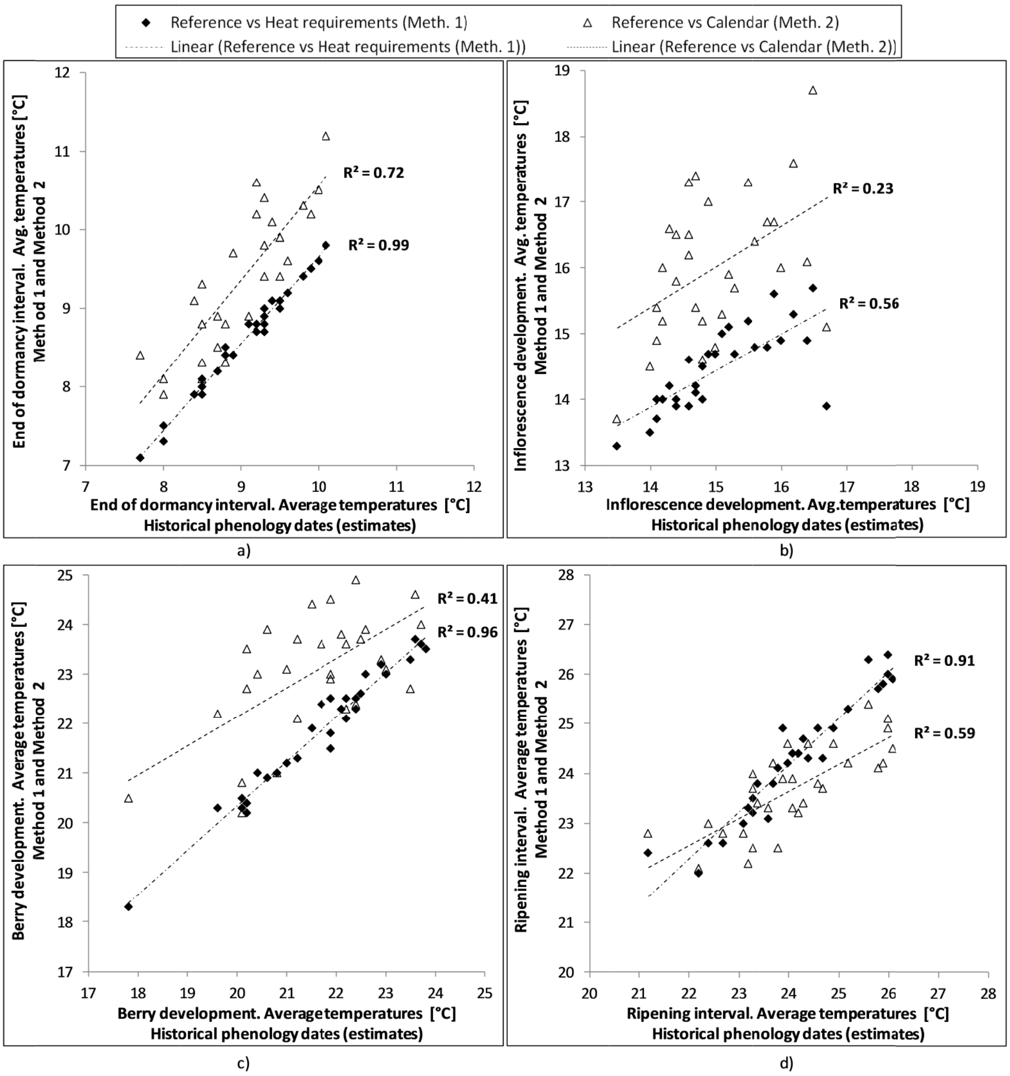


Figure 16 - Plots of mean temperature for growth intervals with boundaries defined using historical phenology dates *vs* corresponding values for growth intervals with boundaries defined by method 1 and method 2: a) End of dormancy interval, b) Inflorescence development interval, c) Berry development interval, and d) Ripening interval.

Figure 17 shows the yearly values of accumulated precipitation in each growth interval when computed using intervals with boundaries defined according to the method based on historical phenology data (reference) *vs* intervals with boundaries defined according to each of the two alternative methods. Regression analysis was conducted and the results show that growth intervals with boundaries based on experimental heat requirements are able to better estimate the accumulated precipitation in growth intervals with boundaries defined using phenology historical dates.

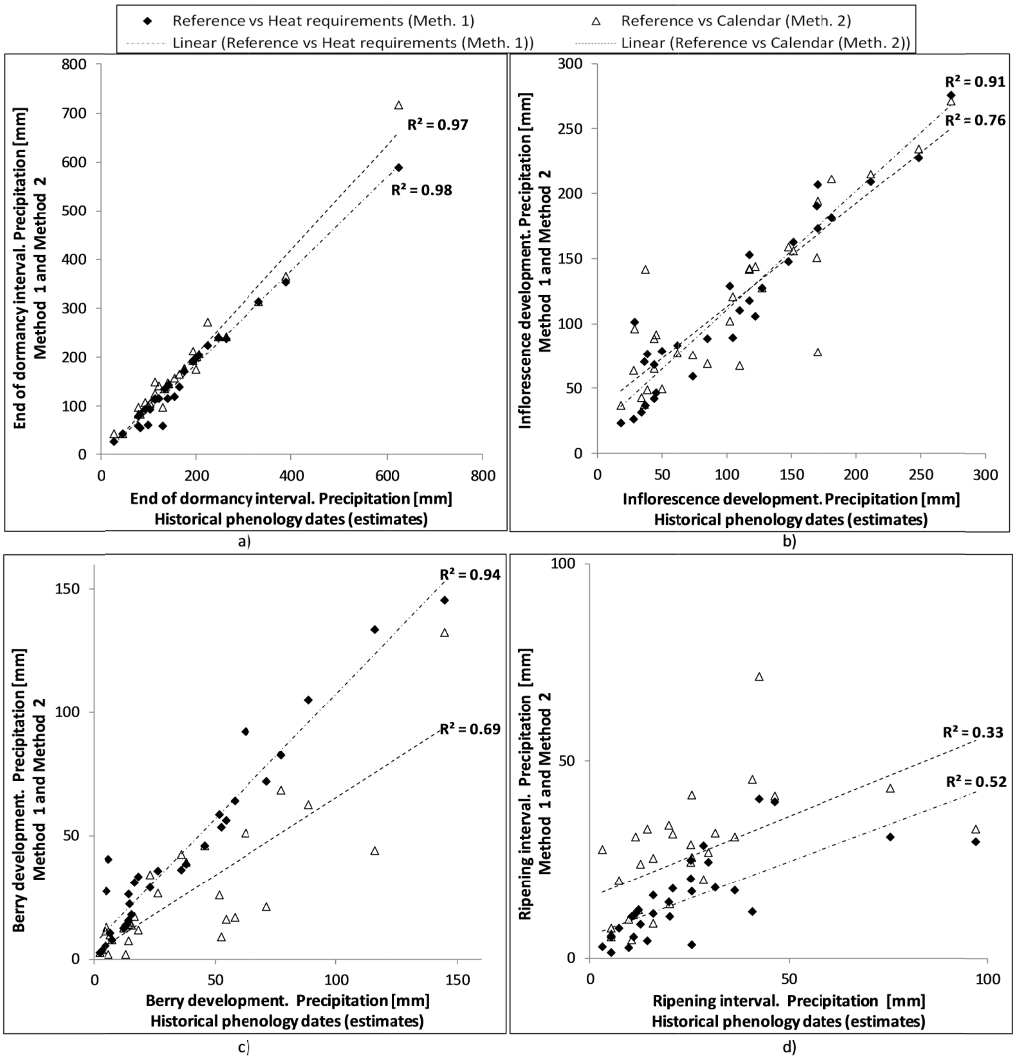


Figure 17 - Plots of accumulated precipitation for growth intervals with boundaries defined using historical phenology dates *vs* corresponding values for growth intervals with boundaries defined by method 1 and method 2: a) End of dormancy interval, b) Inflorescence development interval, c) Berry development interval, and d) Ripening interval.

The plots shown in Figure 16 and Figure 17 exhibit that there is a much higher agreement between temperature and precipitation series between growth intervals with boundaries defined using grapevine experimental heat requirements and growth intervals with boundaries defined using historical phenology observed dates. A method that is able to obtain weather profiles similar to those obtained by using historical yearly data should be preferred in the absence of observations.

6.5 Conclusions

Partitioning the grapevine growing season into smaller growth intervals is necessary for studying the relationships of wine quality to weather and climate variability. When no data on historical phenological dates are available, the partitioning of a growing season may be achieved by defining interval boundaries using different methods: i) by mean values of heat requirements to reach each main phenological stage and ii) by generalized calendar average dates.

In general, it is difficult to have access to consistent data with the dates of the four main developmental stages for grapevines that covers a whole region for an extended period. However, this research was able to obtain the average dates of the main phenological events for the city of *Régua* and yearly data for the *Touriga Franca* variety from 2001 to 2012 at *Quinta de Santa Bárbara*, near the village of *Pinhão*, in the Douro Valley of Portugal. These data were used to estimate the observed yearly dates of the four main developmental phenological events for the region.

Using the available data, the research assessed the accuracy of determining the main growth intervals by means of the heat requirements of the main phenological events and by means of generalized calendar average dates. The results show, that when there are no records on the actual dates of the phenological events, the best option is to use accumulated heat (growing degree-days) to determine the growth events and intervals between events. Partitioning based on calendar dates should be used only when there is no knowledge of grapevine heat requirements or when there are no daily records of temperatures.

Previous research has shown that both climate variability and change play strong roles in wine production and quality in many regions (e.g., Jones et al. 2005 and others). A better understanding of the way weather and climate factors affect the variability of vintages will potentially be invaluable for decreasing the vulnerability of producers in wine regions, and ultimately providing insights in appropriate adaptive measures that will aid in the economic sustainability of wine regions. The results obtained in this work highlight the need for regional monitoring of grapevine growth stages and maintaining consistent historical phenological data for a significant period. Better phenological data and better understanding of the roles weather and climate play in phenological timing, and therefore vintage quality and production variability, will be useful to the winemakers of the Douro Valley and other wine regions.

In the analyses that will be presented in Chapter 7, the boundaries of each growth interval were defined by calculating, for each year, the dates when the heat values that represent the estimates of the historical accumulated heat that is necessary to trigger each phenological event, were reached. This is a partition of the growing season based on the grapevine phenology. This was an innovative approach as most research in this area usually uses calendar dates to partition the growing season.

Chapter 7

Climate and Weather Effects on Vintage Quality, Yield and Price

7.1 Introduction

Understanding the linkages between weather or climate variability and change on vintage quality has become an important scientific research subject. Primault (1969), Winkler et al. (1974), Bindi et al. (1996), Jones & Davis (2000), Corsi & Ashenfelter (2001), Grifoni et al. (2006), Lopes et al. (2008), Ashenfelter (2008), Makra et al. (2009), Mattis (2011), Gladstones (2011), Parker et al. (2011) and numerous other researchers have conducted research on modeling the relationships between meteorological variability and grapevine annual yield and fruit / wine quality.

Research results show that grapevine phenological timing and length of time between events is strongly tied to temperature-based measures such as degree-days and other bioclimatic indices. Still others have used the strong relations between climate and quality to examine climate change impacts (Jones et al. 2005; Duchêne & Schneider 2005; Webb et al. 2007; Schultz & Jones 2010; Tomasi et al. 2011). Their results show that grapevine phenology has generally trended earlier (approximately 5 - 10 days per 1°C of warming), with shorter interphases between events (shortening of 10 - 20 days) which has been related to higher sugar content, lower acidity, and changes in vintage ratings.

The relation between weather or climate variability and change on vintage yield has been studied by Nemani et al. (2001), Adams et al. (2003), Lobell et al. (2006), Santos & Malheiro (2011) and others. Nemani et al. (2001) found that reduced spring frost and lower heat stress during the summer helped yields in Napa Valley grow 34% during 1963-1996. Adams et al. (2003) project yield increases of 90% by 2100 for the coastal regions of California, including the Napa and Sonoma valleys. Results from Lobell et al. (2006) indicate that climate change in California is very likely to put downward pressure on the yields of table grapes by 2050.

The relations between vintage quality based on expert ratings and market prices has also been studied by Costa & Brito Cunha (2011) indicating that an increase of 1 point in Wine Spectator rating results in 29% increase in the release price of bottle-matured Port in the US market. Nemani et al. (2001) also found that a rating increase of 10 points by the Wine Spectator (on a scale from 50 to 100) translated to a 220% increase in price per bottle for Napa Valley wines. This subject was also studied by Ashenfelter (2008), Gibbs et al. (2009), Ashenfelter & Jones (2013) and others showing that wine prices are sensitive to expert ratings and to the historical reputation of the producer and the information of the weather characteristics of the vintage.

The purpose of this research was to bring a better understanding on the relation between the variability of the yearly weather and the annual quality and yield of Vintage Port and, on a larger scale, on the relation between climate trends and the evolution of quality and yield of Vintage Port, over time.

In this section, we first present the definition of the variables and their preliminary analysis and then explore two types of analysis:

- analysis of the Douro Valley climate trends and of links between climate trends and the evolution of Vintage Port quality and yield in the same period;
- analysis of the relation between the yearly variability of the Douro Valley weather to the variability, in terms of quality and yield, of the vintages of Vintage Port.

The Douro Valley climate was characterized in 1980-2009 and trend analysis was conducted to identify trends in temperatures and precipitation. The evolution of vintage quality and yield was compared to climate trends, searching for links. We conducted correlation analysis in order to identify association between the evolution of quality / yield and the trends in climate during 1980-2009. As quality was assessed using a ranking, all climate factors were ranked and their association to vintage quality was tested using the Spearman's rank test. Implications of quality and yield on prices were evaluated using correlation analysis.

Different approaches have been used to analyze the relationship between weather variability and vintage quality as well as between weather variability and vintage yield:

- Logistic Regression: we used logistic regression to model the probability of a vintage to be a high quality or a high yield vintage;
- Top *vs* bottom ranked vintages (analysis of weather variables): Mann-Whitney test was used in the analysis of differences in the median values of the weather related variables, between top *n* vintages and bottom *n* vintages;
- Top *vs* bottom ranked vintages (analysis of phenology variables): t-test was used in the analysis of differences in mean values of the phenology variables between top *n* vintages and bottom *n* vintages

The quality of the vintages of Vintage Port during 1980-2009 was assessed using a consensus ranking that aggregated the different ratings collected from eight vintage-charts (see Chapter 5). The use of an independent response variable (quality) expressed in an ordinal scale imposed limitations on the type of analysis that could be conducted. In terms of regression analysis, Generalized Linear Models (GLM) can handle proportions, binary and ordinal response variables. The most widely used GLM model, the logistic regression model, was used as regression model. Logistic regression models can handle as regressor variables that are not normally distributed and present some heteroscedasticity, but are sensitive to multicollinearity and sample size.

The weather specificities of the best vintages and of the worst vintages were assessed. Comparisons between the weather variables of top *n* and bottom *n* ranked vintages and between the phenology variables of top *n* and bottom *n* ranked vintages was conducted. Differences between the central tendency of the variables was tested. As some weather variables were not normally distributed, the non-parametric Mann-Whitney test was used. As all phenology variables were normally distributed and homoscedastic, the t-test for two independent samples was used.

The association between vintage quality and yield to retail prices was analyzed using Spearman's rank test and the association between vintage quality and yield to market release prices was subjectively analyzed, comparing the evolution of average release prices in 1980-2009 with the evolution of vintage quality and yield in the same period.

7.2 Definition of Variables

7.2.1 Weather Variables

We reviewed the most significant literature on the relationships between meteorological variability and grapevine annual fruit / wine quality. Most research models in this area are based on a measure for vintage quality and on a set of weather related variables defined using a meteorological dataset. Most research uses calendar dates in the definition of temperature and precipitation variables (e.g, Corsi & Ashenfelter 2001; Grifoni et al. 2006; Ashenfelter 2008; Makra et al. 2009; Santos & Malheiro 2011 and others). In this research, we defined a set of temperature related variables, a set of precipitation related variables and one index based on both temperature and precipitation. The defined variables were used in logistic regression models for vintage quality and vintage yield and in the analysis of differences between top ranked vintages and bottom ranked vintages.

Based on the conclusions of section 6.5, the partitioning of the growing season that was used to define the boundaries of the heat related variables was based on plant phenology. This was a differentiating aspect of this research. Instead of using average values of mean temperatures during time intervals whose boundaries are limited by calendar dates we used an indirect assessment of the evolution of mean temperatures along the year, using the lengths of time intervals bounded by heat accumulation amounts related to the plant phenology. For example, instead of defining variables such as the mean temperature in March we defined variables such as the length of the interval from budburst to flowering. In the former example, the time window for the variable maintains over the years and the mean value for temperature varies over the years. In the latter example, the time window for the variable varies over the years and the heat accumulation (related to mean temperature) maintains. In the latter example, the budburst and the flowering events are defined using the average heat accumulation that triggers, in the Douro Valley, these events - 60 GDD for budburst and 400 GDD for flowering. The heat accumulation values that represent, for the Douro Valley, an estimate of the average historical accumulated heat amount necessary to trigger each main phenological event are: Budburst at 60 GDD, Flowering at 400 GDD, *Véraison* at 1100 GDD and Maturity (harvest) at 1750 GDD (see section 6.3.3.1, page 102).

The definition of the intervals to be associated to the heat related variables was based on the heat accumulation for the four main events. However, in order to have better discriminant capacity for the evolution of heat accumulation during the year, the intervals with boundaries defined by each couple of events were split in two, resulting in a partition of the period from January 1 to September 30 into the seven smaller intervals presented in Table 19a. These eight intervals were used in the definition of the heat related variables.

Table 19 - Partition of the period from January 1 to September for a) heat variables and b) precipitation variables.

	Heat	
	Interval boundaries	
Interval	Left	Right
1	January 1	60 GDD
2	60 GDD	230 GDD
3	230 GDD	400 GDD
4	400 GDD	750 GDD
5	750 GDD	1100 GDD
6	1100 GDD	1425 GDD
7	1425 GDD	1750 GDD

a)

	Precipitation	
	Interval boundaries	
Interval	Left	Right
1	January 1	March 31
2	April 1	June 30
3	July 1	September 30

b)

As the boundaries of the intervals used in the definition of heat related variables are based on fixed amounts of heat, the lengths of the intervals, measured in number of days, vary from one year to another. This length variability would make difficult to interpret differences in precipitation between years. To overcome this difficulty the partitioning used to evaluate precipitation was calendar-based.

As rainwater penetrates deep into the soil where it is retained, being usable by plant roots long after the moment when precipitation happens, the definition of precipitation variables does not need to be based on the growth phenological stages of the grapevine. Additionally, as the rainfall that happens in one day is retained and still useable by the plant one or two months later, there is no need to have a partition in very small intervals. For the definition of precipitation variables we used three intervals, corresponding to the three first trimesters of the year. These intervals are presented in Table 19b.

Heat Related Variables

A set of 13 heat related variables was considered adequate to characterize the yearly weather variability related to temperature. Next, we define these variables, noting that the ones bounded in terms of GDD use a phenology partition of the grapevine growing season (see Table 19).

HI Huglin Index (Huglin 1978) is a viticultural climate index that sums accumulated heath over the period from April 1 to September 30. k is the length of day coefficient (1.02 for the Douro Valley).

$$HI = \sum_{i=1 \text{ April}}^{30 \text{ Sept}} \frac{\left(T_{avg(day\ i)} - 10\right) + \left(T_{max(day\ i)} - 10\right)}{2} k \tag{54}$$

CI Cool Night Index (the average of the daily minimum temperatures in September): is a climate index considered to be related to the intensity of flavor, aroma and color of wine (Tonietto & Fialho 2012).

NGS growing season's length [days] .... (60 to1750 GDD)

JB0 1st January to budburst length [days].. (0 to 60 GDD)

BF1 length of first ½ of budburst to flowering [days] .... (60 to 230 GDD)

BF2 length of second ½ of budburst to flowering [days]........... (230 to 400 GDD)

FV1 length of first ½ flowering to *véraison* [days].................... (400 to 750 GDD)

FV2 length of second ½ of flowering to *véraison* [days]......... (750 to 1100 GDD)

VM1 length of first ½ ripening period [days].......................... (1100 to 1425 GDD)

VM2 length of second ½ of ripening period [days]................. (1425 to 1750GDD)

NT1 number of days with $T_{max} > 25°C$ from budburst to flowering

NT2 number of days with $T_{max} > 33°C$ from flowering to *véraison*

NT3 number of days with $T_{max} < 36°C$ from *véraison* to maturity

Precipitation Variables

Precipitation variables were defined using a calendar partition. In order to characterize the yearly weather variability a set of four variables was defined.

PJS accumulated precipitation from Jan 1 to September 30 [mm]

PT1 accumulated precipitation from Jan 1 to March 31 [mm]

PT2 accumulated precipitation from April 1 to June 30 [mm]

PT3 accumulated precipitation from July 1 to September 30 [mm]

Variables Related to Both Heat and Precipitation

SPEI - Standardized Precipitation Evapotranspiration Index is a water balance index between precipitation and potential evapotranspiration, developed by Vicente-Serrano, Beguería, and López-Moreno (2010). SPEI is the number of standard deviations that the observed value would deviate from the long-term median. Negative SPEI values correspond soil drought. Three month SPEI, for periods Jan 1 to March 31, April 1 to June 30, and July 1 to September 30 were computed using a free SPEI software tool (available at http://digital.csic.es/handle/10261/10002):

SPEI3 three month SPEI for Jan 1 to March 31

SPEI6 three month SPEI for April 1 to June 30

SPEI9 three month SPEI for July 1 to September 30

Figure 18 shows the SPEI values in 1980-2009. Moderate to severe drought during the growing season are visible at 1981, 1982, 1986, 1990, 1991, 1992, and 2004. 2005 was the most severe drought in 1950-2010, in the Douro Valley.

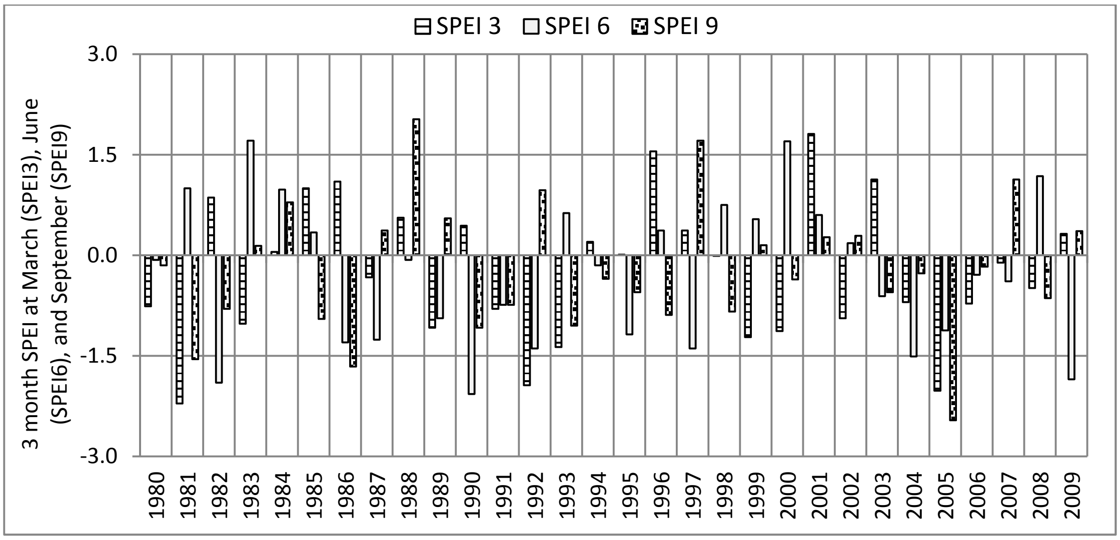


Figure 18 - SPEI for grapevine growing season in the Douro Valley during 1980-2009.

In Table 20 we present the values for all weather dependent variables and indexes.

Table 20 - Values of variables used as candidate explanatory variables and their averages, maximum and minimum values.

	Variables marked with a X present high level of multicollinearity and, by this reason, have not been used as potential explanatory variables in any model																			
	X											X	X	X	X	X	X			
Year	**NGS**	**JB0**	**BF1**	**BF2**	**FV1**	**FV2**	**VM1**	**VM2**	**NT1**	**NT2**	**NT3**	**HI**	**CI**	**SPEI3**	**SPEI6**	**SPEI9**	**PJS**	**PT1**	**PT2**	**PT3**
1980	181	87	45	24	37	29	21	25	13	12	40	2728	13.2	-0.76	-0.07	-0.15	305	165	115	25
1981	172	81	52	27	27	25	20	21	8	27	49	2853	13.8	-2.21	1.00	-1.55	320	98	156	66
1982	174	81	41	24	36	28	22	23	21	12	41	2874	13.9	0.86	-1.90	-0.80	268	83	96	89
1983	200	72	71	19	31	29	26	24	14	10	45	2586	14.0	-1.02	1.71	0.14	405	82	276	47
1984	174	97	26	40	29	28	23	28	10	21	48	2627	11.9	0.05	0.98	0.79	391	192	170	29
1985	166	93	39	26	31	26	24	20	12	19	34	2827	14.9	1.00	0.34	-0.95	470	314	144	12
1986	171	91	50	16	31	25	24	25	24	26	47	2706	14.7	1.10	-1.30	-1.66	378	241	73	64
1987	179	67	51	23	38	26	19	22	16	20	32	2999	15.3	-0.33	-1.26	0.37	388	212	88	88
1988	197	78	52	25	41	28	25	26	5	5	45	2614	12.2	0.56	-0.07	2.03	441	144	268	29
1989	165	81	47	20	34	24	19	21	18	20	29	2958	12.7	-1.08	-0.94	0.55	315	107	179	29
1990	180	60	57	22	36	26	18	21	18	15	24	3098	14.7	0.44	-2.07	-1.08	280	125	80	75
1991	161	90	47	15	33	24	20	22	22	22	28	2908	14.3	-0.80	-0.74	-0.74	331	242	36	53
1992	175	80	47	16	44	24	19	25	23	16	31	2859	12.3	-1.94	-1.39	0.97	274	107	103	64
1993	195	90	46	27	31	26	22	43	9	22	61	2553	11.7	-1.37	0.63	-1.05	381	44	236	101
1994	187	73	48	30	33	26	23	27	14	20	45	2747	11.6	0.20	-0.15	-0.35	376	206	132	38
1995	167	74	30	27	40	30	20	20	26	6	31	2990	12.0	0.01	-1.18	-0.55	339	176	90	73
1996	171	89	40	26	29	26	21	29	8	23	49	2737	12.2	1.55	0.37	-0.89	560	386	122	52
1997	185	66	31	25	52	33	20	24	21	4	34	2970	14.2	0.37	-1.39	1.71	385	135	147	103
1998	179	64	52	25	39	26	19	18	9	13	22	2993	14.6	-0.01	0.75	-0.84	464	150	215	99
1999	163	84	40	22	34	24	22	21	13	20	40	2905	14.1	-1.22	0.54	0.15	413	109	151	153
2000	179	69	57	20	32	26	21	23	12	18	35	2963	13.4	-1.13	1.70	-0.36	336	43	236	57
2001	174	74	44	30	29	29	20	22	9	24	36	2935	13.6	1.81	0.60	0.27	837	718	78	41
2002	168	82	36	28	33	27	21	23	17	20	37	2893	14.0	-0.94	0.18	0.29	326	193	47	86
2003	163	73	44	25	30	28	19	17	13	17	18	3144	15.8	1.13	-0.61	-0.55	402	272	101	29
2004	171	69	49	26	27	27	18	24	17	20	33	3042	14.9	-0.70	-1.51	-0.27	197	111	39	47
2005	152	84	37	25	27	24	20	19	9	30	28	3132	14.3	-2.02	-1.12	-2.46	161	83	51	27
2006	147	84	33	19	31	26	18	20	16	13	26	3203	14.9	-0.72	-0.29	-0.17	331	156	70	105
2007	185	66	47	24	40	31	20	23	16	5	37	2866	13.9	-0.11	-0.39	1.13	329	146	134	49
2008	192	62	54	26	38	27	23	24	12	14	43	2808	13.2	-0.49	1.18	-0.64	403	140	210	53
2009	173	72	47	23	32	30	22	19	22	16	29	3058	14.5	0.32	-1.85	0.36	310	200	98	12
Avg	**175**	**78**	**45**	**24**	**34**	**27**	**21**	**23**	**15**	**17**	**36**	**2886**	**13.7**	**-0.25**	**-0.28**	**-0.21**	**371**	**180**	**131**	**60**
Max	**200**	**97**	**71**	**40**	**52**	**33**	**26**	**43**	**26**	**30**	**61**	**3203**	**15.8**	**1.81**	**1.71**	**2.03**	**837**	**718**	**276**	**153**
Min	**147**	**60**	**26**	**15**	**27**	**24**	**18**	**17**	**5**	**4**	**18**	**2553**	**11.6**	**-2.21**	**-2.07**	**-2.46**	**161**	**43**	**36**	**12**

7.2.2 Phenology Variables

The yearly dates of the main phenological events for the region were calculated (see section 6.3.3, page 102. A set of variables adequate for representing the main phenological events and lengths of the growth stages between events was defined (Table 21). The dates of the main phenological events in 1980-2009 and the corresponding interval lengths were estimated (Table 22)

Table 21 - Definition of phenology variables.

Dates of main phenological events	Growth intervals length
• **BB** – Budburst date	• **Jan-BB** – January 1 to budburst length [days]
• **Fl** – Flowering date	• **BB-Fl** – Budburst to flowering length [days]
• **Vr** – *Véraison* date	• **Fl-Vr** – Flowering to *véraison* length [days]
• **Mt** – Maturity date	• **Vr-Mt** – *Véraison* to maturity length [days]

Table 22 - Dates of main phenological events and length of corresponding growth intervals.

	Phenology ordinal dates				Growth interval length [days]			
Year	**BB**	**Fl**	**Vr**	**Mt**	**Jan-BB**	**BB-Fl**	**Fl-Vr**	**Vr-Mt**
1980	87	156	222	268	87	69	66	46
1981	81	160	212	253	81	79	52	41
1982	81	146	210	255	81	65	64	45
1983	72	162	222	272	72	90	60	50
1984	97	163	220	271	97	66	57	51
1985	93	158	215	259	93	65	57	44
1986	91	157	213	262	91	66	56	49
1987	67	141	205	246	67	74	64	41
1988	78	155	224	275	78	77	69	51
1989	81	148	206	246	81	67	58	40
1990	60	139	201	240	60	79	62	39
1991	90	152	209	251	90	62	57	42
1992	80	143	211	255	80	63	68	44
1993	90	163	220	285	90	73	57	65
1994	73	151	210	260	73	78	59	50
1995	74	131	201	241	74	57	70	40
1996	89	155	210	260	89	66	55	50
1997	66	122	207	251	66	56	85	44
1998	64	141	206	243	64	77	65	37
1999	84	146	204	247	84	62	58	43
2000	69	146	204	248	69	77	58	44
2001	74	148	206	248	74	74	58	42
2002	82	146	206	250	82	64	60	44
2003	73	142	200	236	73	69	58	36
2004	69	144	198	240	69	75	54	42
2005	84	146	197	236	84	62	51	39
2006	84	136	193	231	84	52	57	38
2007	66	137	208	251	66	71	71	43
2008	62	142	207	254	62	80	65	47
2009	72	142	204	245	72	70	62	41
Avg	78	147	208	253	78	70	61	44
Max	97	163	224	285	97	90	85	65
Min	60	122	193	231	60	52	51	36

7.2.3 Preliminary Analysis of Variables

In order to be able to select the statistical analysis techniques that may be adequate to the task, a pre-analysis of the weather variables and of the phenology variables was conducted.

Statistical models (e.g., linear regression) and tests (e.g., t-test for means) rely upon a set of assumptions about the variables used in the analysis: absence of outliers, moderate level of multicollinearity, normality, and homoscedasticity. Logistic regression models do not require normality or homoscedasticity but require a moderate level of multicollinearity.

The weather variables were tested for multicollinearity, normality, and homoscedasticity. The phenology variables were tested for normality and homoscedasticity but not for homoscedasticity since these variables were not used in regression analysis. The existence of outliers was not checked as all meteorological datasets had already undergone an outliers cleaning process (see section 4.4).

7.2.3.1 Multicollinearity

A major consideration in regression models is multicollinearity. Multicollinearity may result in a regression model that has an overall significant F test without significant t tests for individual regressors.

Multicollinearity between potential predictors was analyzed using the Variance Inflation Factor (VIF), a widely used measure of the degree of multicollinearity of the i^{th} independent variable with the other independent variables. The VIF associated with the i^{th} predictor, X_i, is given by:

$$VIF_i = \frac{1}{1 - r_i^2} \tag{55}$$

where r_i^2 is the coeficiente of determination of the i^{th} predictor, X_i, on the remaining n-1 explanatory variables.

One rule of thumb indicates a VIF value of 5 to be used as a threshold, indicating that multicollinearity may exist (Menard 2001). Another commonly used rule of thumb indicates a VIF value of 10, stating that VIF values greater than 10 indicate that multicollinearity may be unduly influencing the least squares estimates (Neter et al. 1996). Although widely used, O'Brien (2007) suggests that VIF value should not be regarded as the unique measure to decide on whether or not, to exclude a potential predictor. We used the mean value of the two most commonly used rules, a threshold VIF value of 7.5.

VIF was calculated for the 20 weather related variables. Iteratively, the variable with the highest VIF > 7.5 was removed and VIF for the remaining variables was recalculated. The process was repeated until all the remaining variables had VIF ≤ 7.5.

From the initial 20 potential explanatory variables 13 variables show a small to moderate level of multicollinearity: JB0, BF1, BF2, FV1, FV2, VM1, VM2, NT1, NT2, NT3, PT1, PT2, and PT3 (see section 7.2). The VIF values for the selected variables are shown in Figure 19. This set of 13 variables will be tested for normality and homoscedasticity.

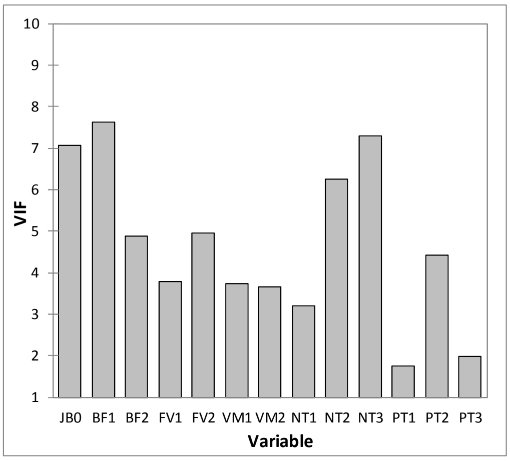


Figure 19 - VIF values for potential variables to be included in models.

7.2.3.2 Normality and Homoscedasticity

Normality was tested using the Shapiro-Wilk test which is considered superior to the K-S test when samples are small, having less than 50 elements (Razali & Wah 2011). Homoscedasticity was tested using the White test (White 1980). A significance level of 5% was used in both tests.

From the 13 selected weather variables, five were considered non-normally distributed (see Table 23): BF2, FV1, VM2, PT1, and PT2. All the variables were considered homoscedastic. Variables that fail the tests of normality or homoscedasticity do not meet the assumptions needed to be used in linear regression. Variables that fail the test of normality do not meet the assumptions needed to be used in the t-test for means.

Table 23 - Results of normality and heteroscedasticity tests for weather variables (Y – yes, N – no).

	JB0	BF1	BF2	FV1	FV2	VM1	VM2	NT1	NT2	NT3	PT1	PT2	PT3
Non-normal	N	N	**Y**	**Y**	N	N	**Y**	N	N	N	**Y**	**Y**	N
Heteroscedastic	N	N	N	N	N	N	N	N	N	N	N	N	N

The tests for all phenology dates and growth interval lengths did not reject the hypothesis of normality or homoscedasticity (see Table 24).

Table 24 - Results of normality and heteroscedasticity tests for phenology variables (Y – yes, N – no).

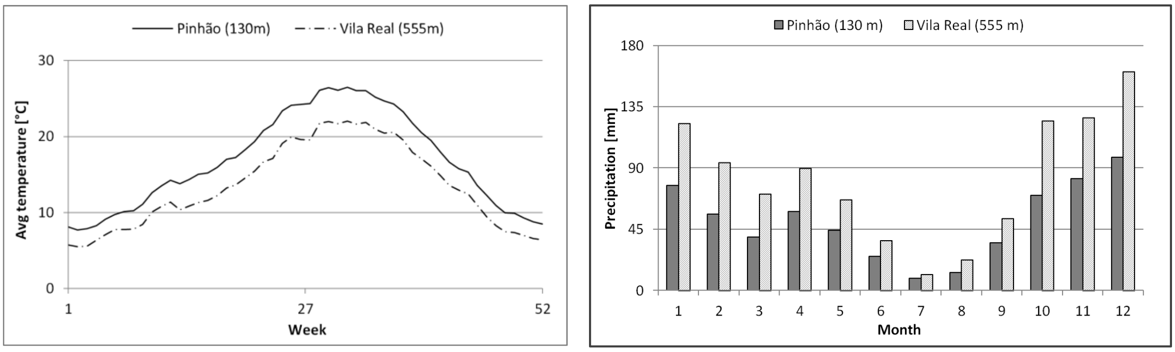
	Phenology dates				Growth interval length			
	BB	**Fl**	**Vr**	**Mt**	**Jan-BB**	**BB-Fl**	**Fl-Vr**	**Vr-Mt**
Non-normal	N	N	N	N	N	N	N	N
Heteroscedastic	N	N	N	N	N	N	N	N

7.3 The Influence of Climate Trends on Vintage Quality and Yield

7.3.1 The Douro Valley Climate in 1980-2009

The Douro Valley is sheltered from Atlantic wet and cold winds by two mountain ranges, *Marão* and *Montemuro*, located at its western border, enhancing a Mediterranean like climate. Temperature increases from West to East and precipitation decreases from West to East. The westernmost regions inside the Douro Valley are nearer to the Atlantic Ocean and are more affected by the moist maritime winds. The easternmost regions inside the Douro Valley, near Spain, are farther from the Atlantic

Ocean and have a more continental climate. As an example, we present the weekly average temperature (Figure 20a) and monthly average precipitation (Figure 20b), in 1980-2009, in two different locations within the Douro Valley: *Vila Real* and *Pinhão*.



a) b)

Figure 20 – a) weekly average temperatures in 1980-2009, in *Vila Real* and *Pinhão* and b) monthly average precipitation in *Vila Real* and *Pinhão*.

Vila Real is located at westernmost edge of the Douro Valley (see Figure 4, page 67) and the weather station from where data were collected is situated at an elevation of 555 meters. *Pinhão* is located at center of the Douro Valley, at the Douro River right bank, and the weather station from where data were collected is situated at an elevation of 130 meters.

Elevation plays a very important role in the temperature values collected at one site as temperature tends to decrease on average 5 ºC / 1000m elevation increase (Rolland 2002). It is of paramount importance to know the elevation of the meteorological stations when comparing the temperatures of two sites, as a part of the temperature difference may be explained by the difference in elevation of the stations. Figure 21 shows the weekly average temperatures in 1980-2009, in *Vila Real* and *Pinhão* when referred to a common elevation of 250 meters by means of temperature lapse rates (Rolland 2002).

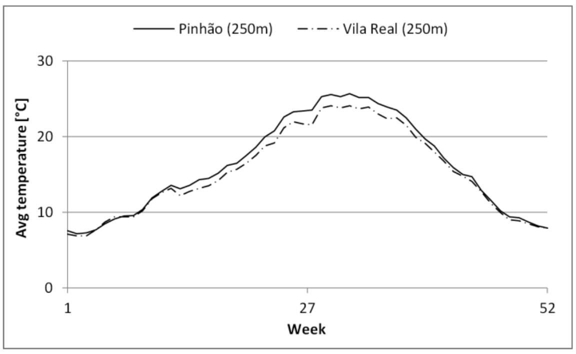


Figure 21 - Weekly average temperatures in 1980-2009, in *Vila Real* and *Pinhão* referred to a common elevation of 250 meters by means of temperature lapse rates.

Figure 20a shows the actual average temperature profiles of both locations and Figure 21 shows the average temperature profiles of both locations when referred to the same elevation of 250m using temperature lapse rates (Rolland 2002). As Figure 20a shows, the difference in the weekly average temperatures of these two locations is significant, reaching 5 ºC in the summer months. The difference in elevation between the weather stations from which the data were collected is 425 meters, which is responsible for a part of the temperature differences between the two locations. When referring both temperatures to a same elevation of 250 meters, the temperatures in *Vila Real* increases, as a consequence of the decline in elevation, whereas the temperature of *Pinhão* decreases, as a consequence of the elevation increase, making the temperature differences between both locations much smaller.

Throughout the remaining of this thesis, we will need to characterize the Douro Valley weather and climate. Meteorological data were collected at five weather stations inside the Douro Valley: *Carrazeda de Ansiães, Mirandela, Pinhão, Régua,* and *Vila Real*. As both precipitation and temperature vary inside the Douro Valley region, we chose three from the five weather stations to be representative of the three generally accepted climatic sub-regions of the Douro Valley: *Baixo Corgo*, *Cima Corgo* e *Douro Superior*. The weather stations of *Régua*, *Pinhão* and *Mirandela* were selected. The weather station of *Vila Real* is located together with the *Régua* station in the sub-region of *Baixo Corgo* but is located at westernmost edge of the Douro Valley, next to the mountain range of *Marão*, having a climate that is not representative of the average Douro Valley climate. The weather station of *Carrazeda de Ansiães* is located in the Douro Superior sub-region but the corresponding meteorological dataset had the largest proportion of missing values (above 12%) and by this reason the station of *Mirandela* (6% of missing values), located at the edge of the Douro Superior sub-region was selected to represent this sub-region. Weather data from these three stations were averaged and the averaged values used as representative of the whole region.

All meteorological data for the Douro Valley used in this research are the average of the data from *Régua*, *Pinhão* and *Mirandela* stations. All temperatures are referred to a common virtual elevation of 250m to uncouple measured temperatures from elevation (Rolland 2002).

Figure 22 shows the weekly average temperatures in 1980-2009, in *Régua*, *Pinhão*, *Mirandela,* and their average referred as Douro.

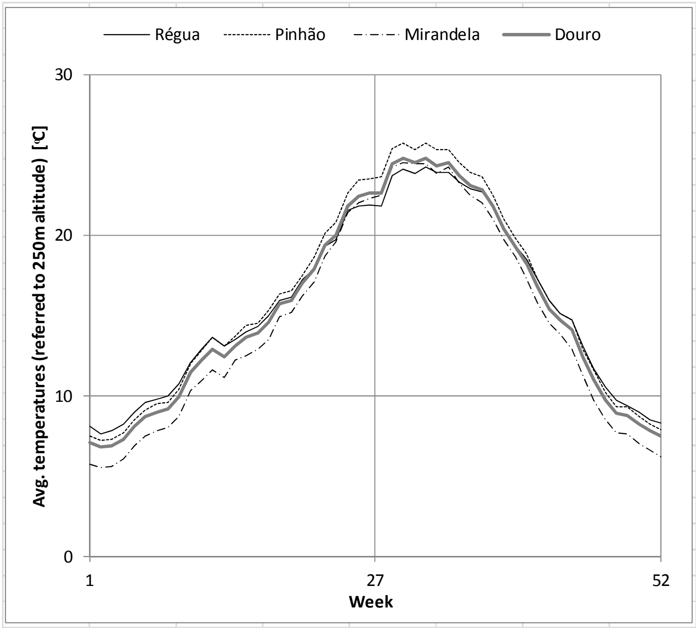


Figure 22 - Weekly average temperatures in 1980-2009, in *Régua*, *Pinhão* and *Mirandela*, referred to a common elevation of 250 meters by means of temperature lapse rates.

Figure 23 shows the weekly average precipitations in 1980-2009, in *Régua*, *Pinhão*, *Mirandela,* and their average referred as Douro.

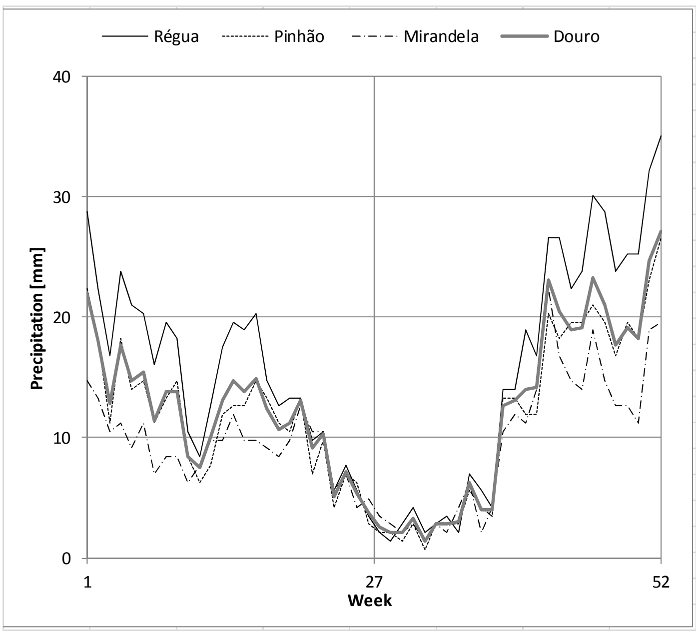


Figure 23 - Weekly average precipitations in 1980-2009, in *Régua*, *Pinhão* and *Mirandela.*

In 1980-2009 the Douro Valley climate was a warm temperature climate with dry, warm summers, a Csb type in *Köppen* Climate Classification System[21] and a HI+2/DI+2/CI+1 type in MCC System[22], with climate characteristics presented in Table 25.

Table 25 - Climate of Douro region (average values in 1980-2009).

Temperatures referred to a reference elevation = 250m		
Variable / Index		Average value
T_{mean} annual	[ºC]	15.4
T_8 average T_{max} warmest month (August)	[ºC]	32.1
T_1 average T_{min} coldest month (January)	[ºC]	2.7
Average annual precipitation	[mm]	624.0
Average precipitation of driest month (July)	[mm]	11.2
Annual Thermal Amplitude	[ºC]	18.2
HI – Huglin Index	[ºC day]	2740.0
DI - Dryness Index	[mm]	-126.0
CI – Cool night index	[ºC]	13.6
T_{mean} (April-September period)	[ºC]	20.6
Precipitation (April-September period)	[mm]	193.0
Annual average sunshine	[hours]	2500.0

7.3.2 Climate Trends in the Douro Valley

For the Douro Valley, the climate projections indicate a range of growing season warming of 0.8 to 1.8 °C by 2020, of 1.8 to 4.3 °C by 2050 and of 2.5 to 6.6 °C by 2080. With respect of precipitation the projections predict a decrease in the precipitation in the driest and warmest areas of the Douro Superior (Jones & Alves 2012).

The evolution of temperature and precipitation was investigated, from 1980 to 2009, within the Douro Valley in order to identify patterns of change. In this analysis, we considered the calendar definition of the grapevine growing season, that includes the six month time period from April 1 to September 30.

[21] World Map of the Köppen-Geiger climate classification (Kottek et al. 2006)

[22] MCC – Geoviticulture Multicriteria Climatic Classification System is a group of 3 indexes that characterizes the viticultural climate of a region (Tonietto & Carbonneau 2004)

7.3.2.1 Temperature

Annual Temperature

In order to capture the trend in the evolution of the mean annual temperature, the following procedure was implemented. First, daily data series was smoothed using a MA365 - moving average of 365 days (1 year) of daily averages. Trend analysis was conducted using MA365. The slope of the regression line was 0.0009 (p-value < 0.0001) that corresponds to an increase of 0.33 ºC/decade. Using this slope, adapted to the 30-year study, the annual mean temperature in the Douro Valley increased 1.0 ºC in 1980-2009, from an average value of 14.9 ºC in the early 80s to 15.9 ºC at present time.

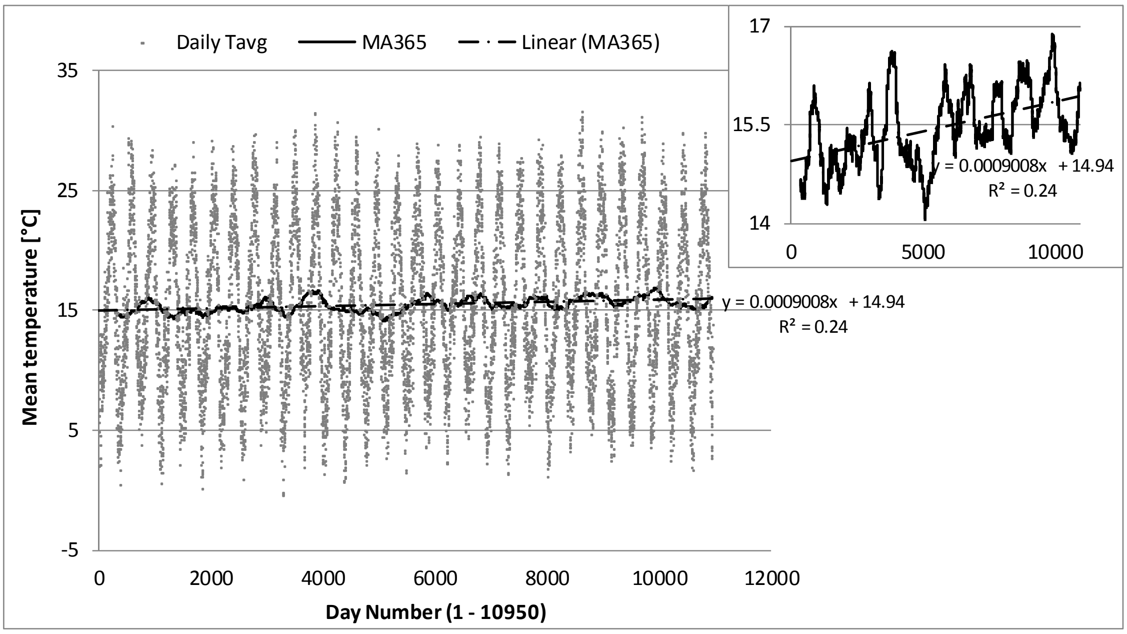


Figure 24 - Daily average temperature in the Douro Valley, in 1980-2009. Zoomed graph shows the 365 days moving average and the corresponding regression line.

Temperature during the Growing Season

Monthly time series containing the daily temperature values, using the 30 years data, were created according to eq. (56).

$$T^{January} = T_{1980}^{Jan\,1}, \dots, T_{1980}^{Jan\,31}, T_{1981}^{Jan\,1}, \dots, T_{1981}^{Jan\,31}, \dots, T_{2009}^{Jan\,1}, \dots, T_{2009}^{Jan\,31} \qquad (56)$$

Series were created for both maximum and minimum temperatures. Regression analysis of these data series, revealed how temperatures evolved monthly, from 1980 to

2009. For each month, the slope of its regression lines was calculated and the hypothesis of being equal to zero tested (t-test, p-value < 0.05). Only slopes significantly different from zero were considered as rates of change of T_{max} or T_{min}. Slopes with p-value ≥ 0.05 were considered zero. Values of average temperature trends for the Douro Valley, in 1980-2009, are shown in Table 26. From 1980 to 2009, the mean temperatures in the Douro Valley increased in every month except November and December. November had a decrease of 0.1 ºC/decade in mean temperature and December had no variation. Every month from February to June, had an increase in monthly mean temperatures greater or equal to 0.2 ºC/decade, having a maximum value of 0.9 ºC/decade in May. In this period, the increase in mean temperatures was caused by increases in both maximum and minimum temperatures, being the former larger than the latter. Every month from July to October, had an increase in monthly mean temperatures smaller or equal to 0.31 ºC/decade. In this period, the increase in mean temperatures was caused exclusively by an increase in minimum temperatures.

Table 26 - Monthly average rates of change per decade for maximum, minimum, and mean temperatures in the Douro Valley, in 1980-2009.

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Rate of change per decade [°C]	T_{max}	0.30	0.44	0.50	0.76	0.98	0.81	0.03	0.00	0.00	0.00	0.00	0.00
	T_{min}	0.51	0.00	0.44	0.46	0.76	0.85	0.43	0.60	0.25	0.32	-0.24	0.00
	T_{mean}	0.41	0.22	0.47	0.61	0.90	0.83	0.24	0.31	0.12	0.27	-0.12	0.00

A visual representation of values in Table 26 is shown in Figure 25.

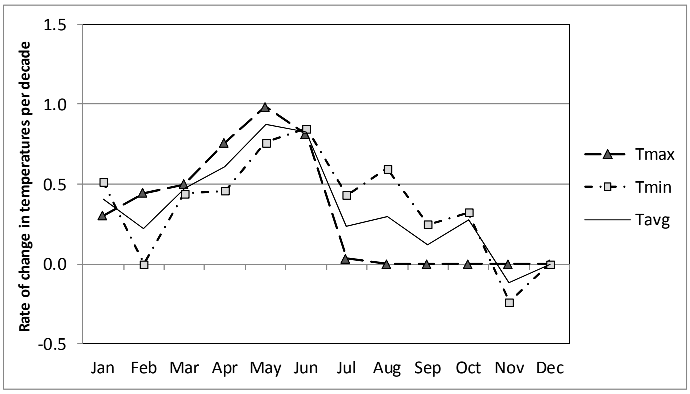


Figure 25 - Monthly average rates of change per decade for maximum, minimum, and average temperatures in the Douro Valley, in 1980-2009.

From Table 26 we can compute the rate of change of the mean temperature during the growing season (April to September) as 0.50 ºC/decade; 0.78 ºC/decade during the first three months of the season and 0.22 ºC/decade during the last three months of the season. The equivalent values for the 30 year period in study (1980-2009) are 1.5 ºC average increase for the growing season mean temperature, from an average value of 19.6 ºC in the early 80s to 21.1 ºC at present time. This increase was not homogeneous over the whole season, with 2.34 ºC average increase in the mean temperature of the first trimester of the growing season (April to June), and 0.66 ºC average increase in the mean temperature of the last trimester of the growing season (July to September).

As the increase in mean temperatures in April to June was the consequence of an increase in maximum temperature that was larger than the increase of minimum temperature, an increase of 0.48 $ºC/_{decade}$ was induced on the average thermal amplitude of the first three months of the growing season. Similarly, as the increase in mean temperatures in July to September was the consequence of an increase in minimum temperature with no change in maximum temperature, a decrease of 0.63 $ºC/_{decade}$ was induced in the average thermal amplitude of the last three months of the growing season.

Extreme Temperatures

We analyzed the evolution of the number of days where maximum temperature was over 36 ºC and of the number of days where minimum temperature was under -2 ºC, in 1980-2009 (see Table 27). A significant trend was detected on the evolution of the annual number of days where minimum temperature was under -2 ºC, showing a decreasing rate of change of 2.4 days per decade, from an average value of 9 days per year in the early 80s to 2 days per year, at 2009. We note that we are using temperatures referred to an elevation (virtual) of 250m to uncouple temperature from elevation effect.

Table 27 - Number of days with temperature above 36 °C and under -2 °C in 1980-2009.

	Year																													
	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09
Tmax > 36 °C	10	21	7	3	3	12	12	23	5	18	22	24	22	10	4	9	10	8	15	11	13	15	9	27	13	20	23	6	9	16
Tmin < -2 °C	16	13	0	10	3	12	3	10	3	5	7	8	12	5	6	4	2	3	4	0	1	11	0	2	0	7	4	6	0	5

If a different elevation was used as reference the number of days with temperature above 36 ºC and under -2 ºC would be different but the trend would be the same.

7.3.2.2 Precipitation

Annual Precipitation Amount

Over the 30-year period, the annual accumulated precipitation (Figure 26), did not change significantly either in the amount or in the distribution pattern. In 1980-2009 the annual precipitation amount in the Douro Valley had a mean value of 624 mm, a median value of 550 mm, ranging from 380 mm in 2005 to 1400 mm in 2001.

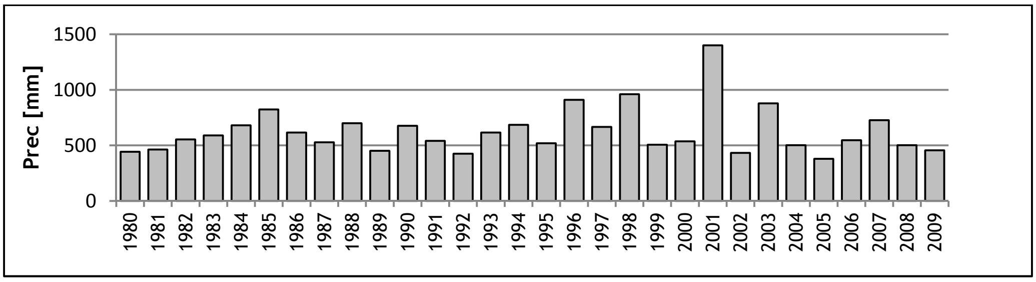


Figure 26 - Annual precipitation in the Douro Valley.

Precipitation Amount during Winter

Over the 30-year period the accumulated precipitation during the January to March three-month period (Figure 27) did not change significantly, either in the amount or in the distribution pattern. In 1980-2009 the annual precipitation amount in the Douro Valley had a mean value of 177 mm, a median value of 147 mm, ranging from 42 mm in 2000 to 718 mm in 2001.

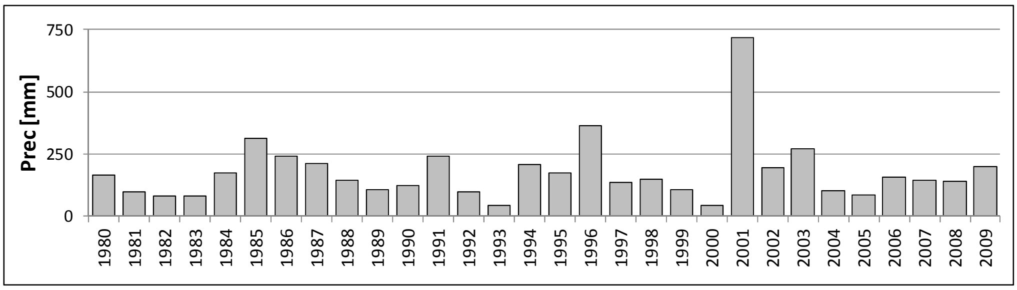


Figure 27 - Precipitation during Winter in the Douro Valley.

Precipitation Amount during the Growing Season

Over the 30-year period the accumulated precipitation during the April to September six-month period (Figure 28) did not change significantly in the distribution pattern. In 1980-2009 the annual precipitation amount in the Douro Valley had a mean value of 193 mm, a median value of 177 mm, ranging from 78 mm in 2005 to 339 mm in 1993.

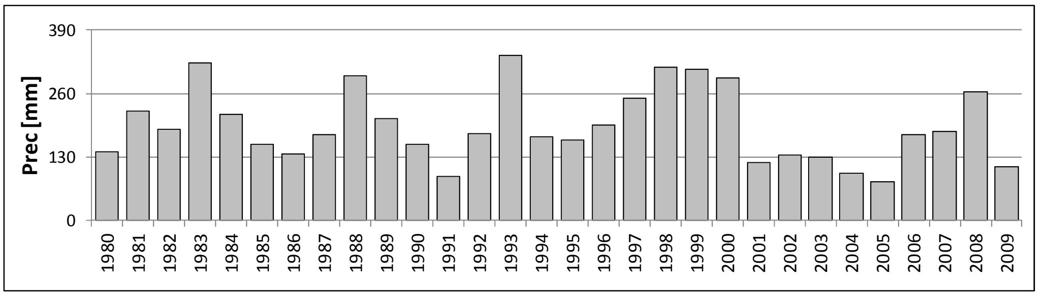


Figure 28 - Precipitation during the growing season in the Douro Valley.

Although there is not a clear pattern of change in Figure 28 apparently, from 2000 to 2009, there was a decrease in precipitation during the growing season.

7.3.2.3 Phenology

In this section, we analyze the evolution of the main phenological event dates and of the corresponding growth interval lengths, throughout the 30 years during 1980-2009, in the Douro Valley. The yearly dates of the main phenological events were obtained using the average heat accumulation (GDD) for each event (see section 7.2.2). A graphical representation of the amplitude of the boundaries of the main grapevine phenological events in Douro Valley during 1980-2009 is shown in Figure 29. The moments when the main events occur, as well as the length of the intervals between these events, vary yearly and may be used as indicators of the overall temperature profile.

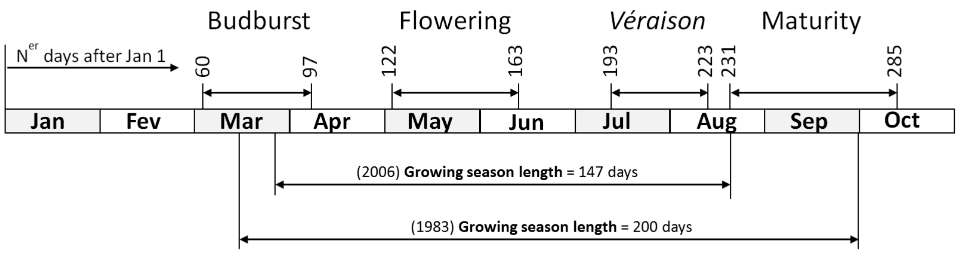


Figure 29 - Major phenological event dates and intervals for grapevines grown in the Douro Valley during 1980-2009.

The distribution of the dates in which each of the four main events occurred is as follows: the budburst event median date is March 12 (72 OD) with an interquartile range of 14.3 days; the flowering event median date is May 20 (140 OD) with an interquartile range of 14.0 days; the *véraison* median date is July 17 (199 OD) and exhibits the smallest variability of the four events with an interquartile range of 8.8 days; the maturity event median date is August 27 (240 OD) with an interquartile range of 10.8 days.

The distribution of the lengths of the four growth intervals and of the growing season is as follows: the end of dormancy interval median length is 72 days with an interquartile range of 14.3 days; the inflorescence development interval median length is 68 days with an interquartile range of 13.8 days; the berry development interval median length is 56 days with an interquartile range of 7.8 days; the ripening interval has a median length of 42 days and the smallest variability in interval lengths with an interquartile range of 5.0 days; the growing season median length is 168 days with an interquartile range of 13.8 days. During this period the range between the shortest and longest growing season was 53 days with the shortest occurring in 2006 (147 days) and the longest occurring in 1983 (200 days).

It is interesting to note that during 1980-2009 the dates of the four major phenological events show a tendency for occurring earlier (see Figure 30). The estimated dates of the phenological events reveal statistically significant decreasing trends (t test, $p < 0.05$), confirming that there is a tendency for an earlier grapevine physiology driven by changes in heat accumulation in the region (Jones & Alves 2012). During the time period, the events have trended 4.2 to 7.5 days earlier per decade, with the maturity dates changing the most.

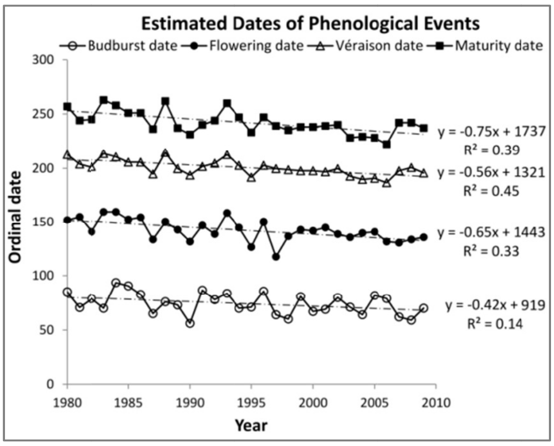


Figure 30 - Trends in the phenological events dates (estimated) during 1980-2009.

For the length of the growth intervals, the Inflorescence Development, the Berry Development and the Ripening intervals showed no significant trend during 1980-2009. However, the analysis of the length of the end-of-dormancy interval (from January 1 to budburst) that is coincident with the date of budburst, in 1980-2009 revealed a statistically significant decreasing trend (t test, p=0.04). Overall budburst today is occurring, on average, 13 days earlier than in the early 1980's.

7.3.2.4 Summary

Jones & Alves (2012) stated that in the Douro Valley “signs of climate change are already evident with higher minimum temperatures, increase of extreme temperatures, fewer cold events, more stress events and smaller temperature range”. The analysis of temperatures and precipitation series confirmed several of these conclusions showing increasing temperatures and an apparent increase in drought. Increase of extreme temperatures was not confirmed. Moreover, the four major phenological events show a tendency for occurring earlier.

From 1980 to 2009 the annual mean temperature increased 1.0 ºC, the growing season (April-September) mean temperature increased 1.5 ºC, the number of days with minimum temperature below -2 ºC decreased 7 days, the annual precipitation maintained and the precipitation during the growing season decreased in the last decade of the study period.

7.3.3 Analyzing the Influence of Climate Trends on Quality and Yield

In section 7.3.2 the Douro Valley climate was characterized. Trend analysis on temperature and precipitation was conducted for assessment of climatic trends in 1980-2009. The analysis revealed trends in the annual mean temperature, with the growing season mean temperature, and with the precipitation amount during the growing season.

Next, we will analyze the evolution of the quality of the vintages and the evolution of wine yield during the same period, looking for trends. Additionally, we compare trends in climate to trends in quality and yield, looking for association.

7.3.3.1 Data on Vintage Quality and Yield

Quality

In the assessment of vintage quality, the consensus ranking obtained in section 5.3 was used.

Yield

To assess wine productionvariability we used wine yield. We believe that wine yield is more adequate than wine production since the latter depends on the area of planted vineyards, which is not dependent of weather variability. Moreover, yield is not affected by commercial policies. Yield is a measure of the amount of grapes or wine that is produced per unit surface of vineyard. In order to estimate region's yearly yield we collected data on yearly wine production and planted area of vineyards.

The annual Port wine production is determined by the IVDP, which takes into consideration the sales and remaining stock from previous years as well as the yield forecast and commercial expectations for the year. In 1980-2009 the overall yield (for all types of wines) in the Douro Valley varied from an average of 21.4 hl/ha, in the least productive vintage, to an average of 56.6 hl/ha in the most productive vintage. Grape-growers tend to allocate as much as is possible of their crop to Port wine production since the price paid for grapes for Port wine is much higher than the price paid for grapes used in still wines. Thus, Port wine annual production variability is relatively smaller than the variability of the global Douro Valley annual production and, consequently, less influenced by weather-variability. For this reason, in order to analyze the influence of the weather on the variability of the yield of Port wine, we believe that is more adequate to consider the average yield of all types of wines produced in the region, as opposed to consider the average yield of Vintage Port.

Data characterizing the Douro Valley wine production are available from the *Instituto Nacional de Estatística*, INE. In order to calculate the yearly average yield for the region it was necessary to have data on the planted area of vineyards in the Douro Valley, in 1980-2009. According to Mayson (2013), in 1982 there were 30 000 ha of vineyards. In addition, according to IVDP, in 2010 there were 43 000 ha. In order to have a clear picture of the evolution in the vineyards area between 1982 and 2010 we collected records from financial programs that were put in place for planting and restructuring of new vineyards in the region. According to (Santos & Azevedo 2004), the PDRITM program added 2 850 ha of new vineyards from 1984 to 1989, the *Programa Operacional* added 1 900 ha of new vineyards from 1990 to 1993, the PAMAF project added 2 800 ha of new vineyards from 1994 to 1999, and the VITIS program added 4 000 ha of new vineyards from 2000 to 2006. Values of planted vineyards were added to the 1982 area to obtain the values in Table 28.

Table 28 - Calculated values for vineyards area in Douro Valley.

Year	Vineyards Area [ha]	Reconverted area of vineyards	Source
1982	30 000		Mayson (2013)
1988	32 850	PDRITM Program 1984-1989, 2850 ha	
1993	34 750	Programa Operacional program 1990-1993, 1900 ha	
1999	37 550	PAMAF project 1994-1999, 2800 ha	
2006	41 550	VITIS program 2000-2006, 4000 ha	
2010	43 000		IVDP

A linear regression estimation on the vineyard area values was used (R^2=0.99) in order to have an estimate of the vineyard area in the Douro Valley for each year in 1980-2009. Estimated yield was computed dividing the yearly production by the estimated vineyard area values. The yields and corresponding ranks sorted by vintage are presented in Table 29 and sorted by rank in Table 30.

Table 29 - Yield [hl/ha] for all types of Douro Valley wines in 1980-2009, sorted by Vintage.
(rank: 1- highest yield, 30 – lowest yield)

Vintage	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09
Yield	45	34	37	36	36	42	35	50	25	36	57	44	35	23	25	33	49	27	21	40	34	46	30	38	35	37	36	29	28	28
Rank	5	19	10	12	15	7	16	2	28	14	1	6	17	29	27	21	3	26	30	8	20	4	22	9	18	11	13	23	24	25

Table 30 - Yield [hl/ha] for all types of Douro Valley wines in 1980-2009, sorted by Rank.
(rank: 1- highest yield, 30 – lowest yield)

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Yield	57	50	49	46	45	44	42	40	38	37	37	36	36	36	36	35	35	35	34	34	33	30	29	28	28	27	25	25	23	21
Vintage	90	87	96	01	80	91	85	99	03	82	05	83	06	89	84	86	92	04	81	00	95	02	07	08	09	97	94	88	93	98

7.3.3.2 The Influence of Climate Trends on Vintage Quality

We analyzed the evolution of vintage quality in 1980-2009 using both the obtained quality consensus ranking (see Table 12, page 90) and the Wine Enthusiast scores (Table 31). Wine Enthusiast was selected as we needed a classification of the vintages in a non-ordinal scale (an interval scale, in the case) and its vintage-chart has the largest number of rated vintages in 1980-2009 (26 rated in 30 vintages).

Table 31 - Vintage Port ratings from Wine Enthusiast. na – not available.

Vintage	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10
Score	81	90	84	85	83	86	85	92	93	**na**	96	91	85	93	87	86	90	84	84	96	90	91	89	95	89	95	92

In Figure 31, the grey square markers represent the ratings from Wine Enthusiast magazine for Vintage Port vintages in 1984-2009 and the black diamond markers represent a consensus ranking of Vintage Port vintages in 1980-2009. For both series, trend lines were plotted in order to facilitate the perception of an increasing or decreasing pattern of the associated data evolution along time.

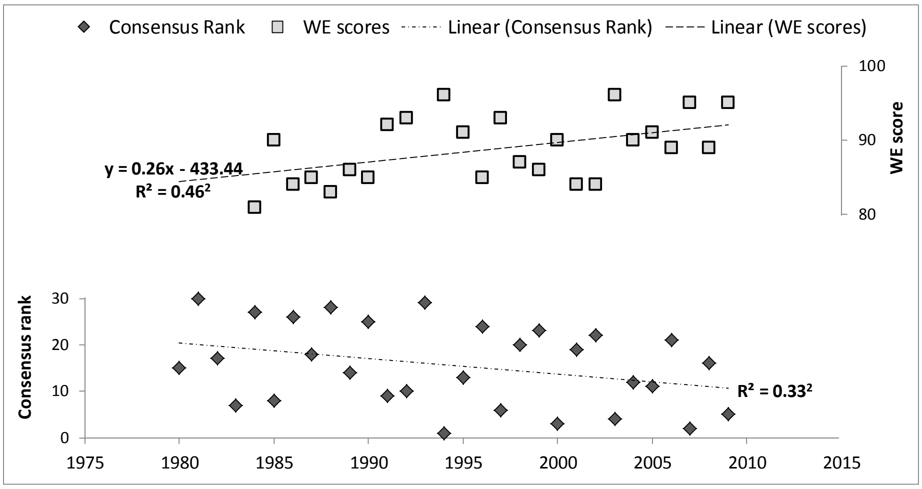


Figure 31 - Vintage Port consensus ranks and Wine Enthusiast scores during 1980-2009.

Wine Enthusiast (WE) ratings show a steady increasing pattern with a trend line having a slope of 0.26, indicating that, on average, the ratings for Vintage Port have increased 2.6 points/decade in 1984-2009. The consensus ranks in 1980-2009 show a steady decreasing pattern (lower ranks mean better quality perception), enforcing the indication given by the Wine Enthusiast ratings.

The association between quality indicators and time evolution in 1980-2009, was assessed using Pearson's correlation test for the WE scores and using Spearman's rank correlation test for the consensus ranking. Pearson's correlation coefficient of 0.46 (p-value = 0.018) was obtained for WE, and Spearman's rank correlation coefficient of -0.33 (p-value = 0.073) was obtained for the consensus ranking in 1980-2009. The results show that in 1980-2009, the overall quality of Vintage Port vintages has steadily increased.

We conducted correlation analysis to identify association between quality and the climate factors that showed trends in 1980-2009: i) annual mean temperature, ii) growing season mean temperature, and iii) precipitation during the growing season. As quality was assessed using a ranking, all climate factors were ranked and association was tested using Spearman's rank test.

A significant correlation (p-value = 0.035) was found between quality and the growing season mean temperature (Figure 32). Vintage quality, in 1980-2009, in the Douro Valley, showed a positive association to the growing season mean temperature, meaning that vintage quality and growing season mean temperature move in the same direction (higher growing season mean temperature is associated with better vintage quality). Moreover, a significant correlation (p-value = 0.032) was found between quality and the amount of precipitation during the growing season (Figure 32). Vintage quality, in 1980-2009, in the Douro Valley, showed a negative relation with the amount of precipitation during the growing season, meaning that vintage quality and growing season's precipitation move in opposite directions (smaller amount of precipitation during the growing season is associated with better vintage quality).

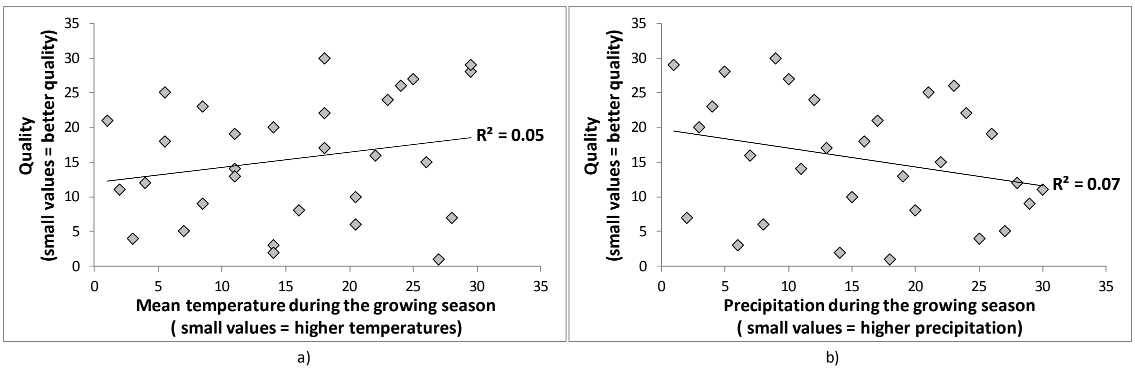


Figure 32 - Association between vintage quality evolution and a) growing season mean temperature [ºC] and b) precipitation during the growing season [mm].

No significant correlation was found between quality and the annual mean temperature. Similarly, no significant correlation was found between quality and the annual precipitation amount.

The analysis showed that vintage quality is associated to the growing season mean temperature and to the amount of precipitation during the growing season. The determination coefficients for both pairs of variables (quality *vs* growing season mean temperature and quality *vs* growing season precipitation amount) are small, indicating that only a small part of the increase in vintage quality may be explained by trends on these variables.

As all the remaining factors that may have influence in wine and vintage quality are generally constant (soil, location, and grape varieties), the only factors that may explain the consistent increase in vintage quality over time is the human factor: vintners skills and better technology.

7.3.3.3 The Influence of Climate Trends on Vintage Yield

We analyzed the evolution of the yearly yield (for all types of wines) in the Douro Valley, in 1980-2009 (see Table 29, page 137). This yield represents the average global wine yield for all types of wines in the region of the Douro Valley and its evolution in 1980-2009 is shown in Figure 33. The regression line shows an apparent decrease in the average yield values from 39.7 hl/ha in the early 80s to 31.7 hl/ha at present time. Yield values show high year-to-year variability making the slope of the regression line, -0.2743, to be considered different from zero with a low significance, p-value = 0.12.

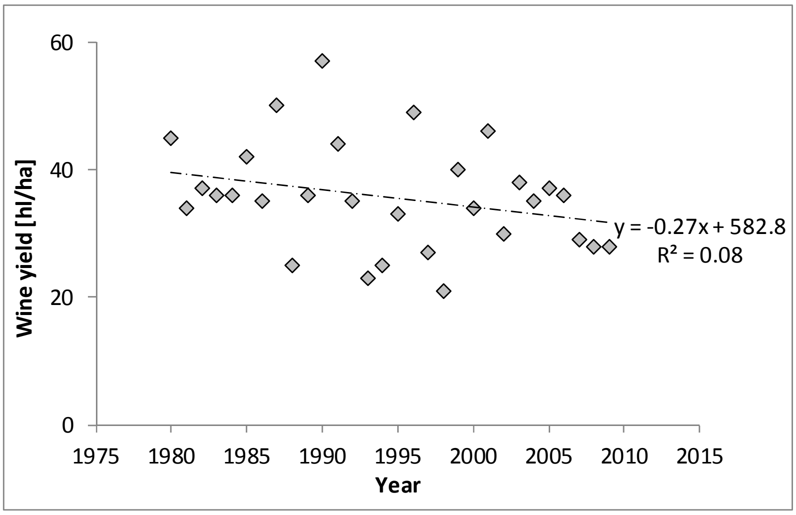


Figure 33 - Wine yield (all types of wine) in the Douro Valley, in 1980-2009.

This research was not able to explain the considerable decrease in yield, over the 30-year period, from 1980 to 2009.

We conducted correlation analysis to identify association between yield and the climate factors that showed significant trends during in 1980-2009: i) annual mean temperature, ii) growing season mean temperature, and iii) precipitation during the growing season.

Significant correlation (p-value < 0.01) was found between yield and the amount of precipitation during the growing season (Figure 34). Wine yield, in 1980-2009, showed a positive association with the amount of precipitation during the growing season, indicating that yield and the amount of precipitation during the growing season move in the same direction.

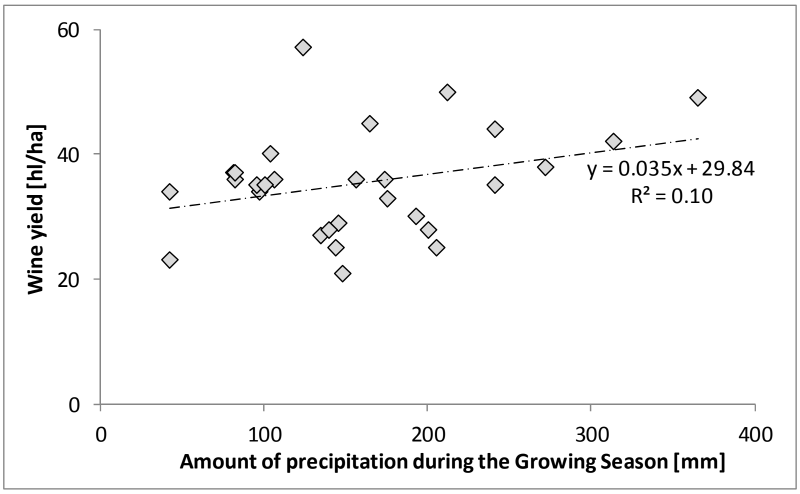


Figure 34 - Association between yield evolution and precipitation during the growing season.

No significant correlation was found between yield and the annual mean temperature nor between yield and the growing season mean temperature.

As a decrease in the precipitation amount during the growing season happened during 1999-2009 (Figure 28), it would be expected, based on the positive association between yield and the amount of precipitation during the growing season, that wine yield would have decreased in the same period and maintained in 1980-1999. However, the decrease pattern in yield maintained throughout the 30-year period of 1980-2009 (see Figure 33), decreasing 20.2% from an average value of 39.7 hl/ha in the early 80s to 31.7 hl/ha at present time. This fact may indicate that yield is decreasing not as a consequence of climate trends in the region but as a consequence of human related factors. Some possible causes of wine yield decrease may be the aging of the vineyards, or a policy of region's vintners to make their vines produce fewer grapes in order to improve quality.

7.3.3.4 Summary

In a region, wine and vintage quality are not exclusively dictated by weather and climate, but by vintners knowledge and technology. During the course of a year, vintners are faced with several jobs in the vineyards that culminate with the harvest. After the harvest, in the cellar, grapes are processed to make wine. In the last two decades, the cellars had great evolution in terms of equipment, using now the state-of-the-art vinification technology.

While it is not possible to quantify the influence of climate trends in the evolution of vintage quality, but certainly they have-influence on it. From the early 80s until present time Vintage Port vintages have been increasing in quality. In the same period, the mean temperature during the growing season increased 1.5 ºC and the annual mean temperature increased 1.0 ºC. Moreover, annual precipitation maintained and the precipitation amount during the growing season decreased in the last decade of 1980-2009.

It is likely that the observed climate trends in the Douro Valley during 1980-2009 were responsible for a moderate portion of the increase in quality of the vintages of Vintage Port and other types of Port wine. Better skills of the region's vintners together with the use of modern technology could have been an important part of the improvement in vintage quality. However, climate trends do not appear to be related to the evolution of the yield in the region.

7.3.4 Analyzing the Influence of Quality and Yield on Price

In this section, we inspect if the high quality of a vintage or the abundance of a harvest yield are related to the average retail prices of a Vintage Port or to the average release prices of Port wine.

7.3.4.1 Data on Retail and Release Prices

Data on production, sales, revenue and prices are in general not available for the Douro Valley region, except for the years after 2000 at IVDP (www.ivdp.pt). Data on release prices for all Port wine types were available from two different sources: (Cunha 2001), for all years in 1980 - 2009 and IVDP, for years after 2005. As the data from both sources are similar for the years after 2005, the data from Cunha (2001) were used. Release prices for Port wine are presented in column 3 of Table 32. The same prices, reduced to 2012 constant prices are displayed in column 4. For the calculation of the constant 2012 prices, the inflation rates for Portugal during 1980-2009 were collected from INE.

Estimates for the current average retail prices for Vintage Port vintages were obtained by averaging the prices of numerous bottles of Vintage Port, for each vintage in 1980-2009, from 290 merchants in the UK, 265 in the USA and 624 in non-UK Europe. Prices were collected from Wine-Searcher (www.wine-searcher.com). Average retail prices for Vintage Port are presented in column 5 of Table 32.

Table 32 - Port wine release prices (cols 3-4) and of current international retail prices for Vintage Port (cols 5-6).

	Quality	Release prices of Port wine		Current retail prices of Vintage Port	
col 1	col 2	col 3	col 4	col 5	col 6
Year	Ranking (1-best, 30-worst)	When released [€/hl]	at constant 2012 prices [€/hl]	March 2014 [€/bottle]	March 2014 (detrended) [€/bottle]
1980	15	58	620	93	66
1981	30	67	596	40	15
1982	17	83	609	70	46
1983	07	109	639	74	52
1984	27	130	599	56	35
1985	08	155	591	70	51
1986	26	177	611	50	32
1987	18	197	613	62	45
1988	28	217	619	38	23
1989	14	241	605	60	46
1990	25	304	675	30	17
1991	09	307	602	59	47
1992	10	302	556	116	105
1993	29	321	536	22	13
1994	01	326	518	98	89
1995	13	330	508	51	43
1996	24	353	529	39	32
1997	06	388	571	59	53
1998	20	393	566	35	30
1999	23	415	590	27	23
2000	03	433	594	58	54
2001	19	429	553	35	32
2002	22	456	566	30	28
2003	04	439	531	65	63
2004	12	431	513	35	34
2005	11	432	506	38	37
2006	21	432	484	34	33
2007	02	429	472	61	61
2008	16	421	451	33	33
2009	05	421	455	56	56

The analysis of the retail prices reveals two main components: a trend component and an oscillating component (Figure 35). The decreasing trend is not explained by vintage quality evolution, which steadily increased from 1980 to 2009 (see section 7.3.3.2, page 137). The reason for the decreasing trend of prices should be related to the fact that Vintage Port increases in quality while in the bottle and to the fact that older vintages are usually more difficult to find in the market.

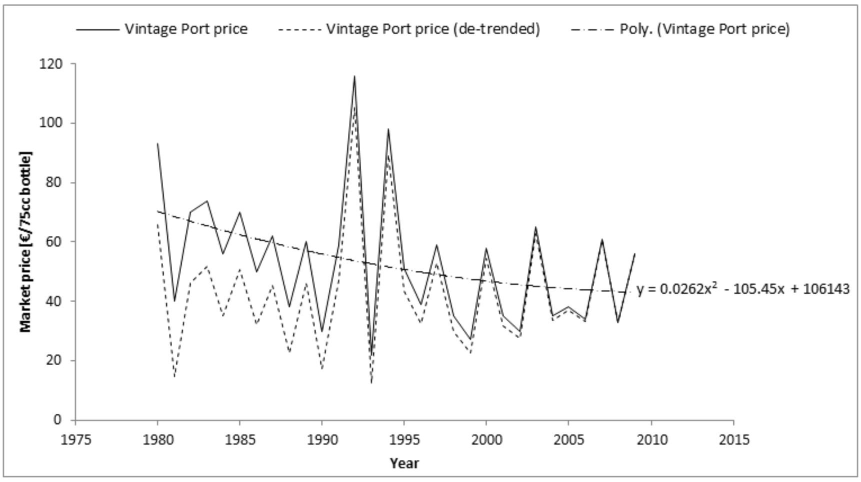


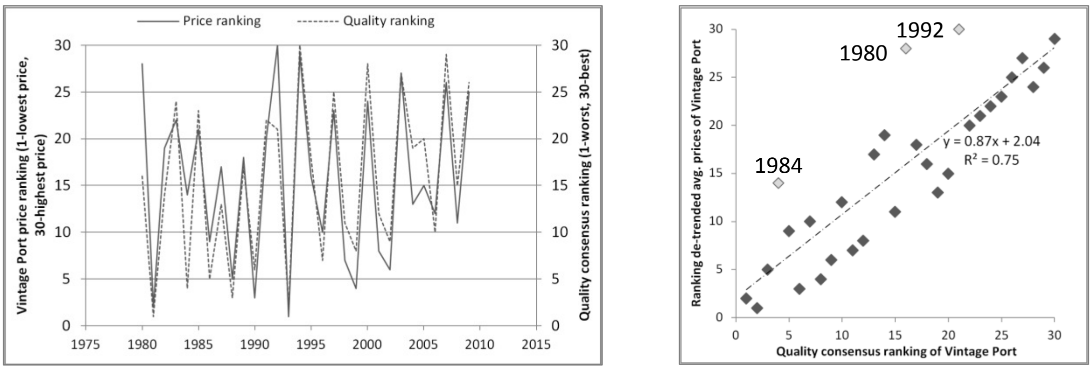
Figure 35 - Market retail prices of Vintage Port.

Detrending the original time series we obtained the oscillating component of the prices. Although the trend removal may also have removed some effects of climate trends in vintage quality that also influence the prices, we believe that in order to study the effect of weather on prices, the removal of the trend component was required. Values of the detrended prices are presented in column 6 of Table 32.

7.3.4.2 The Influence of Quality and Yield on Retail Prices

For the analysis of the influence of the quality and yield of the vintages on Vintage Port retail prices, we used the estimates of the current retail prices of Vintage Port vintages in 1980-2009 (Table 32). As previously explained, in the analysis we used detrended average retail prices as estimates of the retail prices of Vintage Port. Since vintage quality was assessed through a consensus ranking, we ranked the Vintage Port retail prices in order to establish comparisons.

Figure 36a shows a stacked line chart where the dashed line represents the evolution of the ranks of Vintage Port detrended retail prices and the solid line represents the evolution of the ranks of Vintage Port quality consensus ranking, during 1980-2009. Figure 36b presents a scatter plot of the ranks of Vintage Port detrended retail prices *vs* the ranks of Vintage Port quality consensus ranking, during 1980-2009. The concordance between lines in Figure 36a is almost perfect ($r^2 = 0.75$), indicating a very strong association between retail prices and vintage quality, confirmed by the scatter plot in Figure 36b that shows the scatters positioned in the vicinity of the identity line y = x. The slope of regression line in Figure 36b is 0.868 with the 95% confidence interval [0.676, 1.060] including the value 1.0 and the intercept value of the regression line is 2.041 with the 95% confidence interval [-1.368, 5.450] including the value 0.0. The hypothesis of the regression line to be the identity line y = x should not be rejected at a 95% confidence level, not giving statistical evidence that the evolution of the ranks of Vintage Port retail prices is different from the evolution of the ranks of Vintage Port quality ranking. The association between the underlying populations of vintage quality and retail prices was analyzed using Spearman's rank test. A very significant Spearman correlation coefficient of 0.87 ($p < 0.0001$) was found between the two rankings, indicating a very strong relationship between vintage quality and retail price



a) b)

Figure 36 - Charts of a) price vs quality consensus ranking and b) price ranking vs quality consensus ranking.

There are three outliers in the scatterplot in Figure 36b: the vintages of 1980, 1984 and 1992. These outliers represent vintages that have a retail price that is much higher than expected from their quality level.

Similar to the analysis of association between vintage quality and retail prices, the association between the underlying populations of vintage yield and retail prices was also analyzed using Spearman's rank test but no significant association was found.

7.3.4.3 The Influence of Quality and Yield on Release Prices

As the vintage quality ranking regards only the Vintage Port type, for the analysis of the existence of relationships between vintage quality / yield and Port wine release prices we assumed that a good vintage for Vintage Port is usually a good quality vintage for most Port wine styles. We analyzed subjectively how the market release prices of Port wine evolved in 1980-2009, looking for links to both quality and yield.

The evolution of the average market release prices of Port wine at constant 2012 prices is shown in Figure 37. The average release prices decreased from the end of the 1970s throughout 2009.

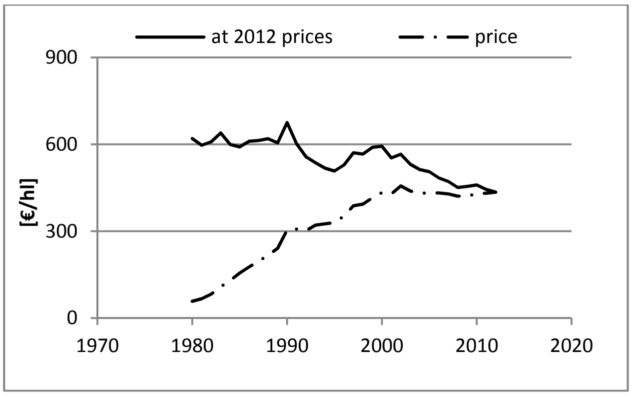


Figure 37 - Evolution of Port wine average release price per hectoliter.

We concluded, in section 7.3.3.1, that the overall quality of Vintage Port vintages increased steadily throughout 1980-2009 and, in section 7.3.3.3, we concluded that wine yield decreased in the same period. Smaller quantities of higher quality Port wine should have influenced prices to increase and cannot explain the sustained decrease in the release prices of Port wine supporting that there is not a direct relation between wine quality and Port wine release prices or between wine yield and Port wine release prices.

7.4 The Influence of the Yearly Weather on Vintage Quality and Yield

In order to analyze the relation between the yearly weather characteristics and the quality and yield of the vintages of Vintage Port we conducted a regression analysis and, additionally, investigated differences between the weather variables values in top ranked vintages and the corresponding values in bottom ranked vintages. The use of an independent response variable (quality) expressed in an ordinal scale imposed limitations on the type of regression analysis that could be implemented. The logistic regression model, the most widely used of Generalized Linear Models (GLM) was considered adequate to model the response variable (quality) using as predictors the weather variables that showed a low to moderate level of multicollinearity.

7.4.1 Modelling Quality and Yield using Logistic Regression

7.4.1.1 Method Presentation

Ordinary least squares (OLS) linear regression predicts the expected value of a given continuous response variable with *k* observations, $\boldsymbol{Y}$, see equation (57), as a linear combination of a set of *n* observed values (predictors). Bold type letters will be used for vectors.

$$\boldsymbol{Y^T} = [y_1, y_2, y_3, \dots y_k] \tag{57}$$

A standard linear regression model has the following form:

$$Y_j = \beta_0 + \beta_1 X_{1,j} + \beta_2 X_{2,j} + \cdots + \beta_n X_{n,j} + e_j \tag{58}$$

or, using vector notation:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e} \tag{59}$$

The ordinary least squares (OLS) procedure is adequate to compute the vector $\hat{\boldsymbol{\beta}}$ as an estimate of the $\boldsymbol{\beta}$ vector:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y},\ \hat{\boldsymbol{\beta}} \approx \boldsymbol{\beta} \qquad (60)$$

Under certain assumptions, the estimates of the regression coefficients obtained by OLS procedure, $\hat{\boldsymbol{\beta}}$, are the "best" estimates in the sense that of all estimates that are unbiased, they have minimum variance. The most important assumptions are the following:

- errors, e, have the same variance, σ^2;
- errors, e, are independent;
- errors, e, are independent of the explanatory variables, X_i;
- errors, e, are Normally distributed with E(e)=0;
- the values of the explanatory variables, X_i, are known without error;
- explanatory variables interrelation is weak or inexistent;
- the response variable is continuous, unbounded, and expressed on an interval or ratio scale.

If all these assumptions are verified then the response variable $\boldsymbol{Y}$ is modeled as a normal random variable. A further assumption that ought to be satisfied is that the values of the explanatory variable are non-stochastic (their values should be known in advance). Most of the times this is not the case and inferences are assumed, in practice, as to be conditional of the values of the explanatory variables (Everitt & Dunn 2001):

$$\hat{\boldsymbol{Y}} = \boldsymbol{E}(\boldsymbol{Y}|\boldsymbol{X}) \qquad (61)$$

Such a model is inadequate in situations when the response variable $\boldsymbol{Y}$ is not a normal random variable (e.g., a response variable that is ordinal or categorical with two or several possible outcomes). To overcome this limitation, models that allows for response variables that have error distribution other than normal were developed. Generalized linear models are a class of models for relating responses for linear combinations of explanatory variables that, in addition to ordinary regression models for continuous response variables, can handle proportions, binary, and ordinal response variables. The linear model is a linear combination with the form of equation (62), where η is called the linear predictor for the model and represents the response variable or some transformation of it.

$$\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{62}$$

The function that makes the transformation of the response variable is called the *link function*. Consider the following examples: $\eta = Y$ is the identity link and leads to the general linear model for continuous outcomes and $\eta = \ln(y)$ is the logarithmic link. $\eta = \ln\left(\frac{p}{1-p}\right)$ is the logistic link where p is the probability of $Y = 1$:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{63}$$

Equation (63) is the model for the logistic regression and $\eta = \ln\left(\frac{p}{1-p}\right)$ is the logistic or LOGIT transformation. Solving equation (63) for p gives:

$$p = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_j X_j)}} \tag{64}$$

If the probability $p \geq 0.5$ then the observation $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is assigned to category $Y = 1$ and otherwise to category $Y = 2$. Consequently, the logistic regression is a linear classifier. β_j estimates are derived assuming a binomial distribution for the observations and maximizing the Likelihood Function $L(\boldsymbol{\beta})$:

$$L(\boldsymbol{\beta}) = \prod \theta(X_j)^{Y_j} [1 - \theta(X_j)]^{(1-Y_j)} \tag{65}$$

The quality of the 30 vintages in 1980-2009 has been assessed using a consensus ranking expressed in a ordinal scale and we are looking for a methodology that is able to model the probability that a vintage becomes a high quality / high yield vintage (expressed as a ranking) using as predictor the weather related variables expressed in numeric continuous scales. Logistic regression is an adequate methodology.

7.4.1.2 Selecting the Type of LOGIT Regression

A consensus quality ranking (see Chapter 5) was used as a relative measure of Port wine vintage quality in 1980-2009. This ranking was used as the response variable for the regression model. This response variable was expressed in an ordinal scale and the 13 potential predictor variables were expressed in numeric continuous scales. Therefore, different formulations of multiple logistic regression could be used, depending on the number of defined classes for the response variable and on the type of scale of those classes:

- two classes (r = 2) expressed in a categorical scale - dichotomous LOGIT,
- several classes (r > 2) expressed in a nominal scale – polytomous LOGIT,
- several classes (r > 2) expressed in a ordinal scale – polytomous ordinal LOGIT.

In order to use a dichotomous LOGIT model we must define two classes of vintages, for example, the top *n* vintages (class 1) and the remaining 30 - *n* vintages (class 2). A similar model should be used to model the bottom vintages, bottom *n* vintages (class 1) and remaining 30 - *n* vintages (class 2). LOGIT regression is especially appropriate for the analysis of dichotomous dependent variables (Menard 2001).

In order to use a polytomous LOGIT model we must define at least three classes of vintages, for example, the top *n* vintages (class 1), the bottom *n* vintages (class 2) and the remaining 30 - 2*n* vintages (class 3). In this case, we have to consider that the defined classes are expressed in a nominal scale, accepting that there was no hierarchy between the classes. This is reasonable when assuming that top quality, bottom quality and the remaining vintages quality represent only three different types of perceived quality, which are only subjective classifications having no hierarchy between categories. Polytomous LOGIT regression models are a mathematical extension of dichotomous models. One of the multiple nominal classes is designated as the reference class and the probability of membership for other classes is compared to the probability of membership in the reference class. The model consists in all the comparisons relative to the reference class. Models for nominal outcomes are sometimes avoided because of the number of parameters and perceived difficulty in their interpretation (Long 2012). Begg & Gray (1984) performed a series of separate dichotomous logistic regressions to replace a polytomous logistic regression and concluded that, in general, the individualized method is highly efficient and facilitates variable selection.

In order to use a polytomous ordinal LOGIT model at least three classes of vintages have to be defined having a hierarchy between the classes or ordering. For example, three classes of vintages could be the top *n* vintages (class 1- high quality vintages), the bottom *n* vintages (class 3 – low quality vintages) and the remaining 30 - 2*n* vintages (class 2 – common vintages). Most common polytomous ordinal LOGIT models use the proportional odds or cumulative odds assumption. Proportional odds restrict the nature of the relations between regressors and outcomes imposing the existence of parallel regressions assumption. This assumption states that regardless of the J - 1 splits of the data to obtain J ordinal classes, if several dichotomous logistic regressions are fit to the data they will show different intercepts but identical slopes (Long 2012). In other words, an ordered logistic regression assumes that the coefficients that describe the relation between the lowest versus all higher classes of the response variable are the same as those that describe the relation between the next lowest class and all higher classes. If the main assumption of proportional odds is not fulfilled, the application of the proportional odds ordinal LOGIT model is invalid and yields misleading results (Bender & Grouven 1998). Most software packages for data analysis provides a score test for the proportional odds assumption but, when using continuous predictor variables, it nearly always indicates rejection of the assumption (O'Connell 2006).

Herein, dichotomous logistic regression was selected as the technique to model the top n vintages as well as the bottom n vintages, in terms of both quality and yield. The objective was to identify significant predictors (weather variables) of quality / yield and to assess the relative influence that each weather variable has on the probability of a vintage to become a high quality or high yield vintage.

In the dichotomous logistic regression model, vintages were coded in two classes: one coded as $Y = 1$ (top n vintages or bottom n vintages) and the other coded as $Y = 2$ (the remaining 30 - n bottom vintages or 30 - n top vintages). Weather variables X_i were used as predictors. The response variable of the model was the conditional probability P($Y = 1 | X_1, X_2, \ldots, X_n$). A cutoff value of 0.5 was used for classification purposes: if the probability $P(Y = 1|X_1, X_2, \cdots, X_n)$ was smaller than 0.5 the observation was assigned to category $Y = 2$ (non top n vintage) otherwise the observation was assigned to category $Y = 1$ (top n vintage).

The Interpretation of the Logistic Regression Coefficients

A logistic regression model allows us to establish a relationship between a binary outcome variable and a group of predictor variables. It models the logit-transformed probability as a linear relationship with the predictor variables. Let $X_1, \ldots, X_n$ be a set of predictor variables. The logistic regression estimates, via the maximum likelihood method, the coefficient values for β_1, β_2, … , β_n, see equation (64). When the regressor X_j with a regression coefficient β_j, increases by one unit, controlling for the other variables, then the odds, $p/(1-p)$, increases by a multiplicative amount of e^{β_j}, where p is the probability of $Y = 1$ (success).

To compare effects of quantitative predictors having different units, it is helpful to report standardized coefficients obtained by fitting the model replacing each X_j by its standardized value. Then, each standardized coefficient represents the effect on the odds, $p/(1-p)$, of a standard deviation change in a predictor controlling for the other variables.

Selecting the Predictors

As referred in section 7.2.1, a set of temperature related variables was defined using grapevine phenology: the dates of the main phenological events estimated based on heat accumulation and the corresponding growth intervals lengths. Additionally, a set of precipitation related variables and one index based on both temperature and precipitation. From the 20 potential explanatory variables initially defined, the 13 variables having shown a small to moderate level of multicollinearity were selected to be used as potential predictors (see section 7.2.3.1) in the LOGIT models: JB0, BF1, BF2, FV1, FV2, VM1, VM2, NT1, NT2, NT3, PT1, PT2, and PT3 (see section 7.2.1, page 116, for a description of the variables).

7.4.1.3 Analysis Procedure

Both for quality and for yield, two logistic dichotomous models were fitted, resulting in four models:

- model 1: class $Y = 1$, top quality n vintages; class $Y = 2$, remaining $30 - n$ vintages;
- model 2: class $Y = 1$, bottom quality n vintages; class $Y = 2$, remaining $30 - n$ vintages;
- model 3: class $Y = 1$, top yield n vintages; class $Y = 2$, remaining $30 - n$ vintages;

- model 4: class $Y = 1$, bottom yield n vintages; class $Y = 2$, remaining 30 - n vintages

As the sample size, m = 30 observations, was smaller than the recommended size for LOGIT models of 100 observations (Long 1997), we considered predictors as significant when having a p-value < 0.15, in models with a goodness of fit p-value < 0.15, as suggested in Long (1997).

It was necessary to select the number of vintages, n, to include in class $Y = 1$. This number should be small enough to include only extreme quality or yield vintages. However, a very small number of n would make possible an undesirable high influence of the characteristics of one or two non-typical vintages on the global characteristics of class $Y = 1$. To comply with these limitations regarding the number of vintages to be included in class $Y = 1$ we considered it adequate to include a number of vintages in the range $6 \leq n \leq 10$ that corresponds to 20.0% - 33.3% of the vintages.

In order to capture the regressors related to the best vintages, as well as those related to the worst vintages, we conducted five dichotomous logistic regressions for each model. Each regression had the top (or bottom) n vintages, $6 \leq n \leq 10$, in class $Y = 1$ and the remaining $30 - n$ vintages in class $Y = 2$. For the five regressions we kept the significant predictors from the regression with $n = 6$ vintages. We tried to keep as few regressors as possible, regarding they were able to produce a model with a significant goodness of fit.

In order to be able to compare the regressors' influence on the response variable, we used standardized coefficients (see page 151).

7.4.1.4 Results

In this section, we present a summary of the five regressions for $6 \leq n \leq 10$ for each model (top quality, bottom quality, top yield, and bottom yield).

Variables with negative coefficients increase the probability of a vintage to become a class $Y = 1$ vintage when they have small values. Variables with positive coefficients increase the probability of a vintage to become a class $Y = 1$ vintage when they have large values. The table cells with grey background highlight insignificant values.

For each regression in Table 33, Table 34, Table 35, and Table 36 the following information is presented:

- the number of vintages in each class;
- a classification table showing the number of well-classified and miss-classified observations for both classes;
- the sensitivity, specificity and the overall percentage of well-classified observations;
- the coefficients for the significant predictors. The five predictors were selected from the regression having six vintages in class $Y = 1$ and 24 vintages in class $Y = 2$, both for quality models and yield models;
- the goodness of fit of the model (-2LL).

In all the five regressions for each model, the number of non-significant regression predictors increases as the number of vintages in class $Y = 1$, n, becomes closer to the number of vintages in class $Y = 2$, $30 - n$.

Model 1: top quality *n* vintages *vs* remaining vintages

Table 33 - Regressors, standardized coefficients, predicting accuracy, and goodness of fit for five dichotomous logistic regressions.

LOGIT regression	Class Y	Observed	Predicted Y=1	Predicted Y=2	Correct		Standardized coefficients for variable JB0	BF1	VM2	NT3	PT3	Goodness of fit -2LL=-2Log(Likelihood) -2LL --> Chi^2
top 6 vintages	1	6	6	0	100.00%	Coefficient	-8.42	-2.64	-2.48	3.16	-3.85	32.60, p-value<0.01
remaining 24 vintages	2	24	2	22	91.67%	p-value <	0.00	0.00	0.08	0.01	0.00	
		Total	**8**	**22**	**93.33%**							
top 7 vintages	1	7	7	0	100.00%	Coefficient	-3.18	-0.37	-1.13	1.72	-1.29	26.10, p-value<0.01
remaining 23 vintages	2	23	4	19	82.61%	p-value <	0.00	0.29	0.30	0.06	0.06	
		Total	**11**	**19**	**86.67%**							
top 8 vintages	1	8	6	2	75.00%	Coefficient	-1.12	-0.26	-0.47	0.83	-0.96	17.70, p-value<0.01
remaining 22 vintages	2	22	5	17	77.27%	p-value <	0.01	0.37	0.48	0.11	0.02	
		Total	**11**	**19**	**76.67%**							
top 9 vintages	1	9	7	2	77.78%	Coefficient	-0.33	0.06	-0.34	0.33	-0.50	8.70, p-value<0.12
remaining 21 vintages	2	21	5	16	76.19%	p-value <	0.15	0.81	0.48	0.39	0.09	
		Total	**12**	**18**	**76.67%**							
top 10 vintages	1	10	7	3	70.00%	Coefficient	-0.27	0.07	0.04	-0.02	-0.47	5.20, p-value<0.39
remaining 20 vintages	2	20	8	12	60.00%	p-value <	0.37	0.79	0.92	0.95	0.09	
		Total	**15**	**15**	**63.33%**							

(shaded) p-value larger than the threshold of 0.15

Model 2: bottom quality *n* vintages *vs* remaining vintages

Table 34 - Regressors, standardized coefficients, predicting accuracy, and goodness of fit for five dichotomous logistic regressions.

LOGIT regression	Class Y	Observed	Predicted Y=1	Predicted Y=2	Correct		Standardized coefficients for variable FV1	VM1	VM2	NT3	PT1	Goodness of fit -2LL=-2Log(Likelihood) -2LL --> Chi^2
bottom 6 vintages	1	6	5	1	83.33%	Coefficient	-0.78	1.36	3.29	-2.73	-0.67	18.10, p-value<0.01
remaining 24 vintages	2	24	6	18	75.00%	p-value <	0.02	0.07	0.00	0.01	0.21	
		Total	**11**	**19**	**76.67%**							
bottom 7 vintages	1	7	6	1	85.70%	Coefficient	-1.16	0.79	4.25	-2.86	0.04	21.90, p-value<0.01
remaining 23 vintages	2	23	5	18	78.26%	p-value <	0.00	0.28	0.00	0.01	0.94	
		Total	**11**	**19**	**80.00%**							
bottom 8 vintages	1	8	7	1	87.50%	Coefficient	-1.10	-0.33	3.36	-1.39	0.03	20.90, p-value<0.01
remaining 22 vintages	2	22	5	17	77.27%	p-value <	0.00	0.55	0.00	0.14	0.95	
		Total	**12**	**18**	**80.00%**							
bottom 9 vintages	1	9	7	2	77.78%	Coefficient	-0.74	-0.59	2.00	-0.42	0.06	17.00, p-value<0.01
remaining 21 vintages	2	21	5	16	76.19%	p-value <	0.02	0.26	0.01	0.53	0.85	
		Total	**12**	**18**	**76.67%**							
bottom 10 vintages	1	10	6	4	60.00%	Coefficient	-0.67	-0.68	1.25	-0.11	0.03	14.30, p-value<0.02
remaining 20 vintages	2	20	5	15	75.00%	p-value <	0.02	0.16	0.06	0.86	0.91	
		Total	**11**	**19**	**70.00%**							

p-value larger than the threshold of 0.15

Model 3: top yield *n* vintages *vs* remaining vintages

Table 35 - Regressors, standardized coefficients, predicting accuracy, and goodness of fit for five dichotomous logistic regressions.

LOGIT regression	Class Y	Observed	Predicted Y=1	Predicted Y=2	Correct		Standardized coefficients for variable BF2	FV1	NT1	PT1	PT2	Goodness of fit -2LL=-2Log(Likelihood) -2LL --> Chi^2
top 6 vintages	1	6	1	5	83.33%	Coefficient	-1.01	1.13	-1.60	0.90	-1.87	22.40, p-value<0.01
remaining 24 vintages	2	24	5	19	79.17%	p-value <	0.12	0.03	0.01	0.03	0.00	
		Total	**6**	**24**	**80.00%**							
top 7 vintages	1	7	6	1	85.71%	Coefficient	-0.75	0.91	-1.32	1.44	-1.45	22.40, p-value<0.01
remaining 23 vintages	2	23	4	19	82.71%	p-value <	0.18	0.07	0.03	0.00	0.02	
		Total	**10**	**20**	**83.33%**							
top 8 vintages	1	8	6	2	75.00%	Coefficient	-0.89	0.46	-1.11	0.88	-0.83	16.90, p-value<0.01
remaining 22 vintages	2	22	7	15	68.18%	p-value <	0.05	0.23	0.03	0.02	0.06	
		Total	**13**	**17**	**70.00%**							
top 9 vintages	1	9	7	2	77.78%	Coefficient	-0.88	0.42	-1.19	1.25	-0.88	19.60, p-value<0.01
remaining 21 vintages	2	21	5	16	76.19%	p-value <	0.06	0.31	0.02	0.01	0.06	
		Total	**12**	**18**	**76.67%**							
top 10 vintages	1	10	7	3	70.00%	Coefficient	-0.48	0.33	-0.66	0.77	-0.72	13.40, p-value<0.02
remaining 20 vintages	2	20	5	15	75.00%	p-value <	0.19	0.34	0.12	0.04	0.07	
		Total	**12**	**18**	**73.33%**							

p-value larger than the threshold of 0.15

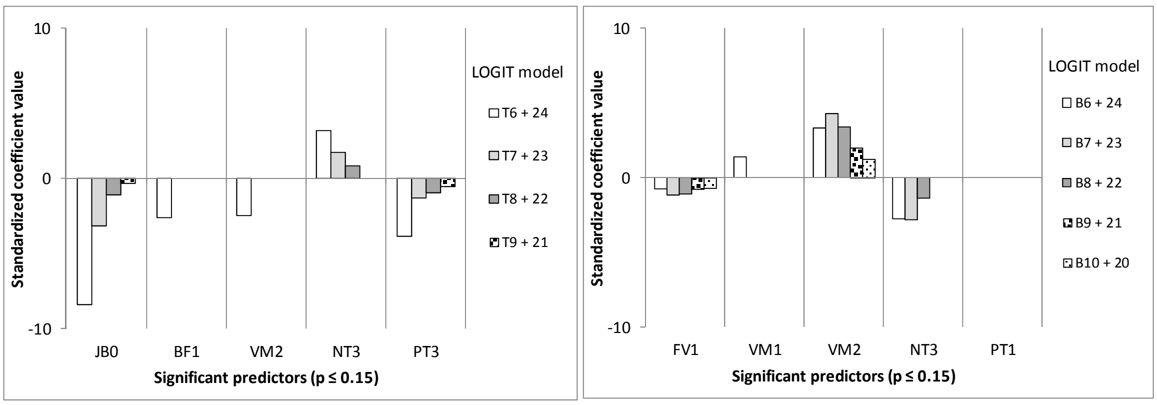
Model 4: bottom <u>yield</u> *n* vintages *vs* remaining vintages

Table 36 - Regressors, standardized coefficients, predicting accuracy, and goodness of fit for five dichotomous logistic regressions.

			Predicted				Standardized coefficients for variable					Goodness of fit -2LL=-2Log(Likelihood)
LOGIT regression	Class Y	Observed	Y=1	Y=2	Correct		JB0	BF2	VM1	VM2	NT3	-2LL --> Chi^2
bottom 6 vintages	1	6	6	0	100.00%	Coefficient	-2.89	1.32	3.08	3.82	-3.81	22.30, p-value<0.01
remaining 24 vintages	2	24	3	21	87.50%	p-value <	0.00	0.05	0.01	0.00	0.01	
		Total	**9**	**21**	**90.00%**							
bottom 7 vintages	1	7	7	0	100.00%	Coefficient	-3.44	1.56	3.18	4.24	-4.05	25.50, p-value<0.01
remaining 23 vintages	2	23	3	20	86.96%	p-value <	0.00	0.04	0.02	0.00	0.02	
		Total	**10**	**20**	**90.00%**							
bottom 8 vintages	1	8	8	0	100.00%	Coefficient	-4.51	1.94	3.57	4.72	-4.14	29.10, p-value<0.01
remaining 22 vintages	2	22	1	21	95.45%	p-value <	0.00	0.03	0.01	0.00	0.03	
		Total	**9**	**21**	**96.67%**							
bottom 9 vintages	1	9	7	2	77.78%	Coefficient	-1.52	0.56	0.37	1.22	-0.52	19.30, p-value<0.01
remaining 21 vintages	2	21	5	16	76.19%	p-value <	0.00	0.15	0.48	0.02	0.46	
		Total	**12**	**18**	**76.67%**							
bottom 10 vintages	1	10	8	2	80.00%	Coefficient	-4.66	-0.23	0.23	0.30	0.48	17.50, p-value<0.01
remaining 20 vintages	2	20	6	14	70.00%	p-value <	0.00	0.10	0.41	0.04	0.44	
		Total	**14**	**16**	**73.33%**							

(shaded) p-value larger than the threshold of 0.15

To allow an easier comparison of the similarities and differences between models we present a graphical representation of the standardized coefficients. Coefficients in Figure 38 were collected from Table 33 and Table 34 (quality analysis).



a) b)

Figure 38 - Standardized coefficients for a) top **quality** vintages (the regression T10 + 20 is not presented as its goodness of fit was not significant) and b) bottom **quality** vintages.

Coefficients in Figure 39 were collected from Table 35, and Table 36 (yield analysis).

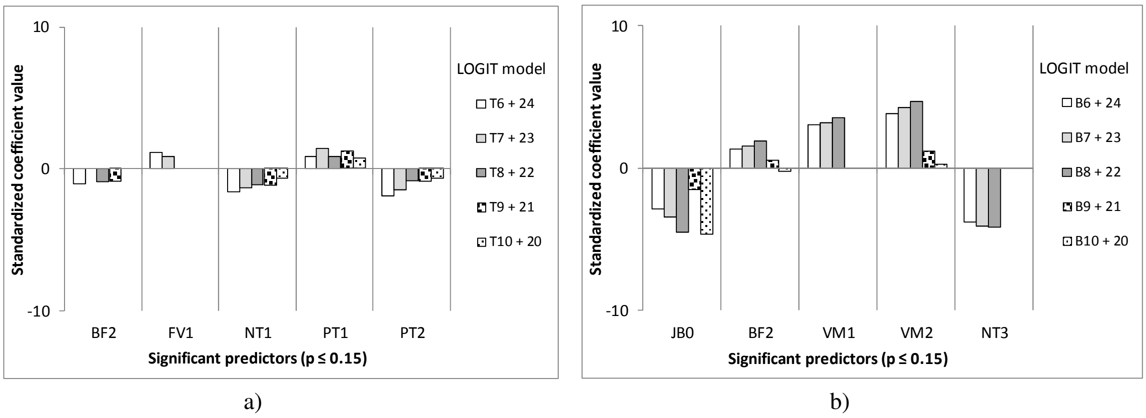


Figure 39 - Standardized coefficients for a) top **yield** vintages and b) bottom **yield** vintages.

From the analysis of Figure 38 and Figure 39 it is possible identify the variables that are most influential on vintage quality and on vintage yield, and to assess their relative influence. We will only consider coefficients that are significant on at least three of the five regressions.

High quality vintages are influenced by the following weather variables:

- JB0 - number of days from January 1 to the day with an accumulation of heat of 60 GDD (budburst). Small values of JB0 highly enhance the probability of vintage to be a quality vintage. In 1980-2009 the average value of JB0 was 78 days, ranging from 60 to 97 days;
- NT3 - number of days with Tmax < 36°C from *véraison* (1100 GDD) to Maturity (1750 GDD). Large values of NT3 enhance the probability of vintage to be a quality vintage. In 1980-2009 the average value of NT3 was 36 days, ranging from 18 to 61 days;
- PT3 - accumulated precipitation from July 1 to September 30. Small values of PT3 enhance the probability of vintage to be a quality vintage. In 1980-2009 the average value of PT3 was 60 mm, ranging from 12 to 153 mm.

Years with a warm Winter that promotes an early budburst, with a small number of days having very high temperatures during the period after the *véraison*, and with a dry trimester from the end of June to the harvest in September, enhance the likelihood that the outcome of the vintage may be of high quality.

High yield vintages are influenced by the following weather variables:

- BF2 – number of days from the day with an accumulation of heat 230 GDD (half way from Budburst to Flowering) to the day with an accumulation of heat of 400 GDD (Flowering). Small values of BF2 enhance the probability of vintage to be a high yield vintage. In 1980-2009 the average value of BF2 was 24 days, ranging from 15 to 40 days;
- NT1 – number of days with Tmax > 25°C from Budburst (60 GDD) to Flowering (400 GDD). Small values of NT1 enhance the probability of vintage to be a high yield vintage. In 1980-2009 the average value of NT1 was 15 days, ranging from 5 to 26 days;
- PT1 – accumulated precipitation from January 1 to March 31. Large values of PT1 enhance the probability of vintage to be a high yield vintage. In 1980-2009 the average value of PT1 was 180 mm, ranging from 43 to 718 mm;
- PT2 – accumulated precipitation from April 1 to June 30. Small values of PT2 enhance the probability of vintage to be a high yield vintage. In 1980-2009 the average value of PT2 was 131 mm, ranging from 36 to 276 mm.

  Years with a warm spring that simultaneously has a small number of days with high temperatures, having a rainy trimester from January to March, and a dry trimester from April to June, enhance de likelihood that the outcome of the vintage may be of high yield.

As the variables that enhance the probability of a vintage to become a high quality vintage are not the same variables that enhance the probability of a vintage to become a high yield vintage, it is expected that a high quality vintage is unlikely to be a high yield vintage. This reasoning is enforced by the fact that opposite precipitation profiles are related to quality and yield and also by the fact that an early budburst (small value of JB0) is related to both top quality vintages and bottom yield vintages.

7.4.2 Top *vs* Bottom Ranked Vintages

As the number of vintages used as sample size, $m = 30$ (vintages in 1980-2009), was smaller than the recommended size for LOGIT models we complemented the regression analysis with the analysis of central tendency of weather variables and phenology variables in top *n* and bottom *n* ranked vintages, looking for significant differences.

7.4.2.1 Comparing Weather Variables

A comparison between the weather variables for top *n* and bottom *n* ranked vintages was conducted in order to assess differences between the central tendency of the variables in better and worse vintages. As some variables were not normally distributed, the non-parametric Mann-Whitney test, with a significance level of 0.05, was used to compare the median values of top *n* and bottom *n* vintages.

In order to analyze to what extent the *n* medians of top *n* vintages were different from the medians of bottom *n* vintages, five comparisons were performed for *n* values in top *n* and bottom *n* vintages in the range $6 \leq n \leq 10$, corresponding to 20.0% - 33.3% of all vintages in 1980-2009.

Results

Results of Mann-Whitney tests, indicating tests decisions and corresponding p-values are presented for two models: i) top quality *n* vintages *vs* bottom quality *n* vintages and ii) top yield *n* vintages *vs* bottom yield *n* vintages. Only the variables that presented at least one significant test result in one of the two models are shown in Table 37 and Table 38. If the median of the top *n* vintages was smaller than the median of the bottom *n* vintages Table 37 and Table 38 indicate T < B, otherwise T > B.

Model 1: top quality *n* vintages *vs* bottom quality *n* vintages

Table 37 - Mann-Whitney test results for top **quality** *n* vintages vs bottom **quality** *n* vintages.

		JB0		BF2		FV1		FV2		VM1		NT2		NT3		PT1		PT2		PT3	
		result	p value	result	p value	result	p value	result	p value	result	p value	result	p value	result	p value	result	p value	result	p value	result	p value
Quality	median T6 vs median B6	T<B	0.03					T>B	0.05												
	median T7 vs median B7	T<B	0.02					T>B	0.02			T<B	0.03								
	median T8 vs median B8	T<B	0.03					T>B	0.00			T<B	0.02	T<B	0.05						
	median T9 vs median B9	T<B	0.05					T>B	0.02			T<B	0.02	T<B	0.03					T<B	0.04
	median T10 vs median B10	T<B	0.03									T<B	0.04							T<B	0.03

Model 2: top yield *n* vintages *vs* bottom yield *n* vintages

Table 38 - Mann-Whitney test results for top **yield** *n* vintages vs bottom **yield** *n* vintages.

		JB0		BF2		FV1		FV2		VM1		NT2		NT3		PT1		PT2		PT3	
		result	p value	result	p value	result	p value	result	p value	result	p value	result	p value	result	p value	result	p value	result	p value	result	p value
Yield	median T6 vs median B6									T<B	0.04					T>B	0.05	T<B	0.00		
	median T7 vs median B7															T>B	0.02	T<B	0.00		
	median T8 vs median B8					T<B	0.04	T<B	0.04							T>B	0.05	T<B	0.01		
	median T9 vs median B9					T<B	0.04					T>B	0.04			T>B	0.03	T<B	0.03		
	median T10 vs median B10			T<B	0.02	T<B	0.02					T>B	0.04					T<B	0.03		

In the analysis of the tables' results it is important to be attentive to the high or low positioning (row number) of the test entries for each variable. For example, in Table 37 the test entries for variable PT3 are placed in rows 4 and 5 (from top to bottom), indicating that the differences in the medians happen when comparing closer groups of vintages (top 9 *vs* bottom 9, and top 10 *vs* bottom 10). The high positioning of the entries of variable FV2 (rows 1, 2, 3, and 4) refers to comparison between groups of vintages that are further apart from each other (top 6 *vs* bottom 6 to top 9 *vs* bottom 9 vintages). Variables having significant test results located higher in the tables refer to comparisons between groups of vintages that are further apart from each other and, by this reason are more discriminating than variables having test decisions located lower in the tables.

The concentration of the entries of a variable may also vary. Variables having significant test decisions located together (or in positions close to each other) are more discriminating than variables having test decisions located sparsely.

In order to have a measure that may give an indication of the overall "relative discriminant capability" of a variable to differentiate top vintages from bottom vintages the following procedure, using two types of weights for each variable, was adopted:

- W_1 is a weight to differentiate entries located higher from entries located lower. W_1 was defined assigning an integer value from 18 (top 6 vs bottom 6 comparisons) to 10 (top 10 vs bottom 10 comparisons), using a step of two units;
- W_2 is a weight to differentiate entries that are concentrated close to each other from entries that have a sparse positioning. W_2 was defined calculating the concentration of the test entries for each variable by dividing the difference

between the higher and the lower entry by the number of entries for that variable ($W_2 \geq 1.0$. Value one indicates that all entries for a variable are adjacent);

- test entries were replaced by -1 if T < B or by +1 if T > B;
- a global value for each variable was obtained as a weighted average, summing, for each variable, the product of the entries, -1 or +1, by W_1 and dividing the result by the sum of W_1;
- the weighted average values for each variable were divided by W_2.

Using the above procedure, Table 37 and Table 38 were transformed into the equivalent Table 39 and Table 40 where the relative discriminant capability of each variable is presented using normalized values in a [-10, +10] scale. Negative values of the relative discriminant capability refers to variables that have a median significantly smaller on the top *n* vintages than on the bottom *n* vintages. Positive values of the relative discriminant capability refer to variables that have a median significantly larger on the top *n* vintages than on the bottom *n* vintages. Large absolute values of the relative discriminant capability of a variable indicate high discriminant capability to differentiate top vintages from bottom vintages.

Model 1: top quality *n* vintages *vs* bottom quality *n* vintages

Table 39 - Equivalent form of Table 37.

		i	W_1	JB0	BF2	FV1	FV2	VM1	NT2	NT3	PT1	PT2	PT3
Quality	median T6 vs median B6	1	18	-1			1						
	median T7 vs median B7	2	16	-1			1		-1				
	median T8 vs median B8	3	14	-1			1		-1	-1			
	median T9 vs median B9	4	12	-1			1		-1	-1			-1
	median T10 vs median B10	5	10	-1					-1				-1
				-1.00	0.00	0.00	0.86	0.00	-0.74	-0.37	0.00	0.00	-0.31
			W_2	1.00			1.00		1.00	1.00			1.00
	Relative discriminative capability			-10.0			8.6		-7.4	-3.7			-3.1

Model 2: top yield *n* vintages *vs* bottom yield *n* vintages

Table 40 - Equivalent form of Table 38.

		i	W_1	JB0	BF2	FV1	FV2	VM1	NT2	NT3	PT1	PT2	PT3
Yield	median T6 vs median B6	1	18					-1			1	-1	
	median T7 vs median B7	2	16								1	-1	
	median T8 vs median B8	3	14			-1	-1				1	-1	
	median T9 vs median B9	4	12			-1			1		1	-1	
	median T10 vs median B10	5	10		-1	-1			1			-1	
				0.00	-0.14	-0.51	-0.20	-0.26	0.31	0.00	0.86	-1.00	0.00
			W_2		1.00	1.00	1.00	1.00	1.00		1.00	1.00	
	Relative discriminative capability				-1.4	-5.1	-2.0	-2.6	3.1		8.6	-10.0	

From the analysis of Table 39 and Table 40 it was possible identify the variables that are most influential on vintage quality and on vintage yield as well as to assess their relative influence.

High quality vintages appear to be related to the following weather variables:

- JB0 - number of days from January 1 to the day with an accumulation of heat of 60 GDD (Budburst). Small values of JB0 highly enhance the probability of vintage to be a quality vintage. In 1980-2009 the average value of JB0 was 78 days, ranging from 60 to 97 days;
- FV2 - number of days from the day with an accumulation of heat 750 GDD (half way from Flowering to Véraison) to the day with an accumulation of heat of 1100 GDD (Véraison). Large values of FV2 highly enhance the probability of vintage to be a quality vintage. In 1980-2009 the average value of FV2 was 27 days, ranging from 24 to 33 days;
- NT2 – number of days with Tmax > 33°C from Flowering (400 GDD) to Véraison (1100 GDD). Small values of NT2 highly enhance the probability of vintage to be a high quality vintage. In 1980-2009 the average value of NT2 was 17 days, ranging from 4 to 30 days;
- NT3 - number of days with Tmax < 36°C from Véraison (1100 GDD) to Maturity (1750 GDD). Small values of NT3 enhance the probability of vintage to be a quality vintage. In 1980-2009 the average value of NT3 was 36 days, ranging from 18 to 61 days;
- PT3 - accumulated precipitation from July 1 to September 30. Small values of PT3 enhance the probability of vintage to be a quality vintage. In 1980-2009 the average value of PT3 was 60 mm, ranging from 12 to 153 mm.

High yield vintages appear to be related to the following weather variables:

- BF2 – number of days from the day with an accumulation of heat 230 GDD (half way from Budburst to Flowering) to the day with an accumulation of heat of 400 GDD (Flowering). Small values of BF2 enhance the probability of vintage to be a high yield vintage. In 1980-2009 the average value of BF2 was 24 days, ranging from 15 to 40 days;

- FV1 - number of days from the day with an accumulation of heat 400 GDD (Flowering) to the day with an accumulation of heat of 750 GDD (half way from Flowering to Véraison). Small values of FV2 enhance the probability of vintage to be a high yield vintage. In 1980-2009 the average value of FV2 was 27 days, ranging from 24 to 33 days;
- FV2 - number of days from the day with an accumulation of heat 750 GDD (half way from Flowering to Véraison) to the day with an accumulation of heat of 1100 GDD (Véraison). Small values of FV1 enhance the probability of vintage to be a high yield vintage. In 1980-2009 the average value of FV1 was 34 days, ranging from 27 to 52 days;
- VM1 - number of days from the day with an accumulation of heat 1100 GDD (Véraison) to the day with an accumulation of heat of 1425 GDD (half way from Véraison to Maturity). Small values of FV1 enhance the probability of vintage to be a high yield vintage. In 1980-2009 the average value of FV1 was 34 days, ranging from 27 to 52 days;
- NT2 – number of days with Tmax > 33°C from Flowering (400 GDD) to Véraison (1100 GDD). Large values of NT2 enhance the probability of vintage to be a high yield vintage. In 1980-2009 the average value of NT2 was 17 days, ranging from 4 to 30 days;
- PT1 – accumulated precipitation from January 1 to March 31. Large values of PT1 highly enhance the probability of vintage to be a high yield vintage. In 1980-2009 the average value of PT1 was 180 mm, ranging from 43 to 718 mm;
- PT2 – accumulated precipitation from April 1 to June 30. Small values of PT2 highly enhance the probability of vintage to be a high yield vintage. In 1980-2009 the average value of PT2 was 131 mm, ranging from 36 to 276 mm.

Summary of the Weather Variables Comparisons

Figure 40 summarizes the results presented in Table 39 and Table 40 showing relative discriminant capability of each variable. Two different patterns arise from the analysis of top *n* vintages *vs* bottom *n* vintages; one for the variables that tend to promote vintage quality and another for the variables that tend to promote vintage yield.

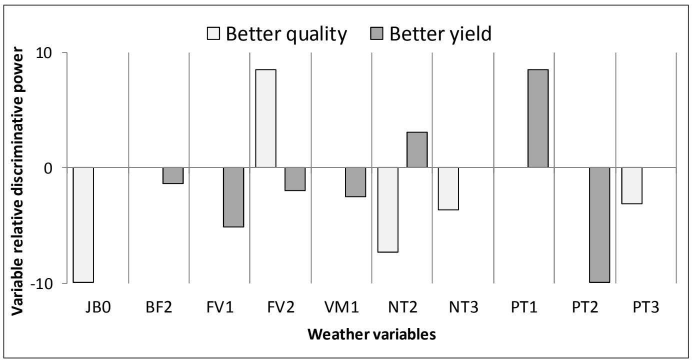


Figure 40 - Overall relative capability to discriminate top ranked vintages from bottom ranked vintages Negative values for variables meaning that the medians of bottom n vintages are smaller than the correspondent medians of top n vintages.

Analysis of Figure 40 shows that a set of five weather-related variables tend to promote high quality vintages: JB0, FV2, NT2, NT3, and PT3. Years with a warm winter that promotes an early budburst, with lower temperatures from flowering to *véraison,* with a small number of days having very high temperatures during the period after the *véraison*, and with a dry trimester from the end of June to the harvest in September, enhance the likelihood that the outcome of the vintage may be of high quality.

A different set of seven weather related variables tend to promote high yield vintages: BF2, FV1, FV2, VM1, NT2, NT3, PT1, and PT2. Years having a warm spring with the first two summer months warmer than average, having a small number of days with very high temperatures during the period after the *véraison*, with a rainy trimester from January to March, and a dry trimester from April to June, enhance the likelihood that the outcome of the vintage may be of high yield.

The weather profile that enhances the likelihood of a vintage to become a high quality vintage is quite different compared to the weather profile that enhances the likelihood of a vintage to become a high yield vintage.

7.4.2.2 Comparing Phenology Variables

Phenology variables representing the yearly dates of the main phenological events and on the lengths of the corresponding growth intervals have been used. In this research, as discussed in section 7.2.2, the grapevine growing season was partitioned into four smaller growth intervals using the main phenological events as boundaries for each growth interval:

- End-of-dormancy interval: time period bounded by January 1 and the budburst event (60 GDD);
- Inflorescence development interval: time period bounded by the budburst event (60 GDD) and the flowering event (400 GDD);
- Berry development interval: time period bounded by the flowering event (400 GDD) and the véraison event (1100 GDD);
- Ripening interval: time period bounded by the véraison event (1100 GDD), and maturity / harvest (1750 GDD).

Values of phenology variables during 1980-2009 are presented in Table 22, page 121. All variables were previously tested for normality and for homoscedasticity (see section 7.2.3.2) with tests' results not rejecting the hypotheses of normality or homoscedasticity for any of the variables.

In order to assess if grapevine phenology influences the overall outcome of a vintage in terms of both quality and yield we analyzed, in 1980-2009, in the Douro Valley, the ability of the main phenological dates and of growth intervals lengths to differentiate the best from the worst vintages. Using t-test with a significance level of 0.10, we compared differences between the average dates of the main phenological events in the top *n* vintages and the corresponding average dates in the bottom *n* vintages, varying *n* in the range $6 \le n \le 15$. A similar procedure was used to compare differences between the average lengths of the four growth intervals in the top *n* vintages and the corresponding average lengths in the bottom *n* vintages.

We repeated the tests from top 6 *vs* bottom 6 to top 15 *vs* bottom 15, in order to identify to what extent (number of vintages in the top group and in the bottom group) the differences between the averages of the two groups would be significantly different. We note that when increasing the number of elements considered in the average calculations from n to $n + 1$ vintages, only the element ranked $n + 1$ is new, being the only responsible for a change in the average value. For example, comparing the average of the top 6 vintages with the average of the top 7 vintages, only the value of the vintage ranked 7 is new in the average calculation, meaning that any difference in the average was induced only by the value of element 7.

We anticipated that increasing the number of vintages of both groups (top *n* and bottom *n*) the differences between groups' averages, if they existed, would get smaller.

Results

Results of the comparisons between top and bottom ranked vintages are presented for the four models.

Model 1: Events' average dates of the top quality *n* vintages compared to the bottom quality *n* vintages

In Figure 41 we present a graphical representation of the evolution of the main phenological events' average dates, for the top quality *n* and bottom quality *n* vintages, when $6 \leq n \leq 15$. For all events, the absolute difference between the average dates of the top *n* and bottom *n* vintages gets smaller as the number of vintages in both groups increases.

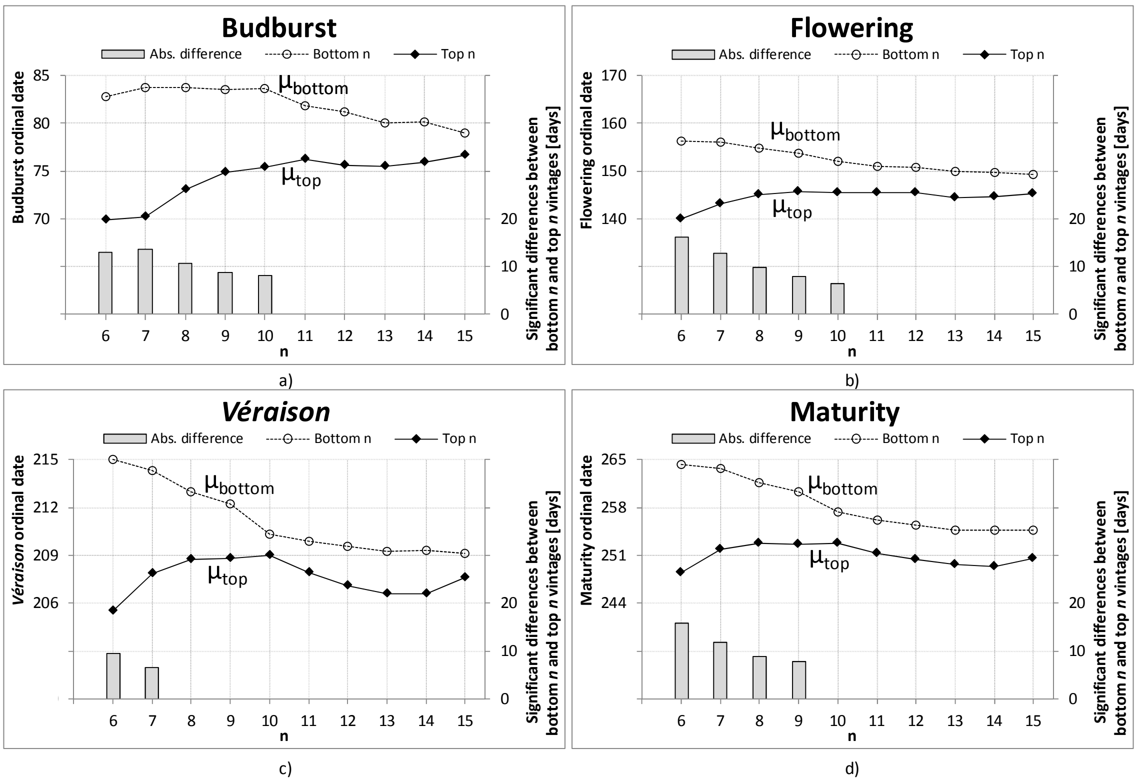


Figure 41 - Average dates of the major phenological events for top *n* vintages and bottom *n* vintages.

Figure 41a shows that regarding the budburst dates, there are significant differences between the average dates of top *n* vintages when compared to the average dates of the bottom *n* vintages, when the number vintages in both groups, *n*, is in the range $6 \leq n \leq 10$.

Figure 41b shows that regarding the flowering dates, there are significant differences between the average dates of top *n* vintages when compared to the average dates of the bottom *n* vintages, when the number vintages in both groups, *n*, is in the range $6 \le n \le 10$.

Figure 41c shows that regarding the *véraison* dates, there are significant differences between the average dates of top *n* vintages when compared to the average dates of the bottom *n* vintages, when the number vintages in both groups, *n*, is in the range $6 \le n \le 7$.

Figure 41d shows that regarding the maturity dates, there are significant differences between the average dates of top *n* vintages when compared to the average dates of the bottom *n* vintages, when the number vintages in both groups, *n*, is in the range $6 \le n \le 9$.

Model 2: Growth intervals' average lengths of the top quality *n* vintages compared to the bottom quality *n* vintages

In Figure 42 we present a graphical representation of the evolution of the growth intervals' average lengths, for the top quality *n* and bottom quality *n* vintages, when $6 \le n \le 15$.

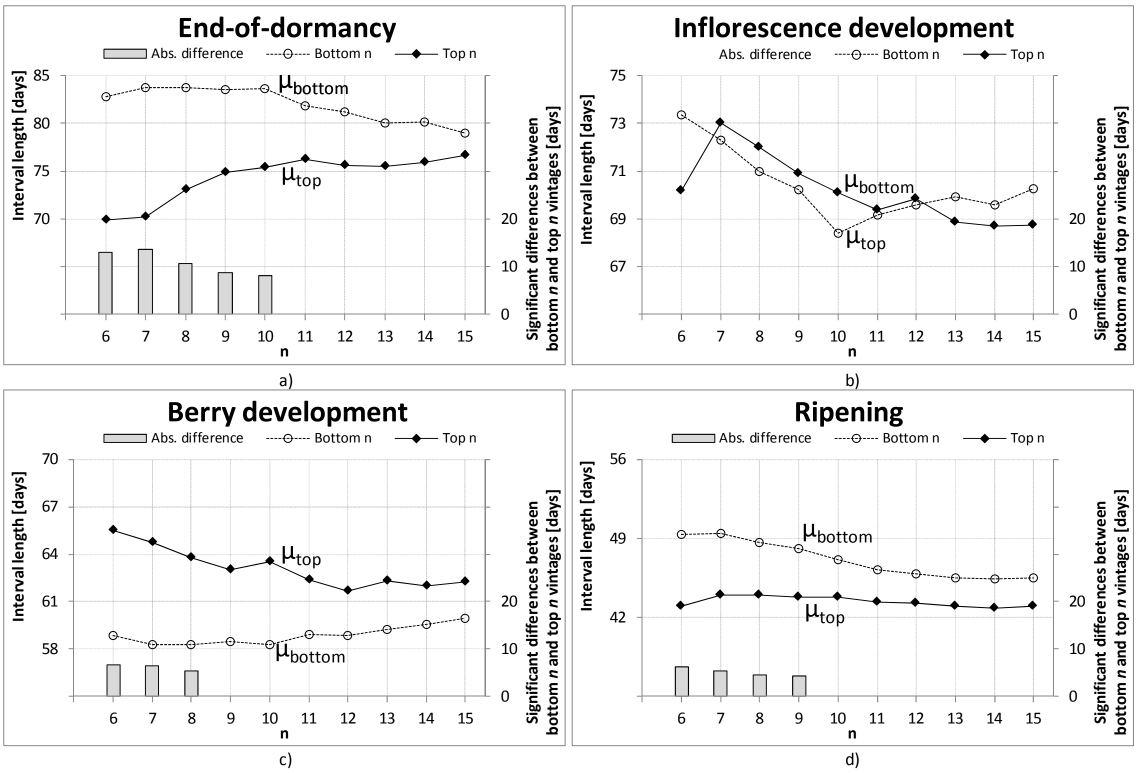


Figure 42 - Growth intervals' average lengths for top *n* vintages and bottom *n* vintages.

For all growth intervals, except the inflorescence development interval, the absolute difference between the average intervals' lengths of the top n and bottom n vintages gets smaller as the number of vintages in both groups increases.

Figure 42a shows that regarding the length of the end dormancy interval, there are significant differences between the average lengths of top n vintages when compared to the average lengths of the bottom n vintages, when the number vintages in both groups, n, is in the range $6 \leq n \leq 10$.

Figure 42b shows that regarding the length of the inflorescence development interval, there are no significant differences between the average lengths of top n vintages when compared to the average lengths of the bottom n vintages.

Figure 42c shows that regarding the length of the berry development interval, there are significant differences between the average lengths of top n vintages when compared to the average lengths of the bottom n vintages, when the number vintages in both groups, n, is in the range $6 \leq n \leq 8$.

Figure 42d shows that regarding the length of the ripening interval, there are significant differences between the average lengths of top n vintages when compared to the average lengths of the bottom n vintages, when the number vintages in both groups, n, is in the range $6 \leq n \leq 9$.

Interestingly, while the lengths of the end dormancy interval and of the ripening interval are smaller in better vintages, the length of the berry development interval is smaller in worse vintages, indicating that cool temperatures during the flowering to *véraison* period tend to promote vintage quality.

There are no significant differences between the main phenological events' average dates nor between growth interval lengths for the models corresponding to comparisons between the n top yield vintages and the n bottom yield vintages.

Summary of the Phenology Variables Comparisons

The analysis between the average values of the phenology variables from top yield vintages and bottom yield vintages did not reveal significant differences between the dates of the main phenological events nor the lengths of corresponding growth intervals.

Apparently, the yield of a vintage has little association phenology. As the phenology of a vintage is mainly determined by heat accumulation (Bonhomme 2000; van Leeuwen et al. 2008; Gladstones 2011), this lack of association may indicate that the yield of a vintage is mainly related to the precipitation amount and its distribution during the year.

In the analysis between top quality vintages and bottom quality vintages the average dates of the main phenological events showed significant differences between top quality and bottom quality vintages, for all events, indicating that high quality vintages have, on average, an earlier phenology when compared to low quality vintages. Figure 43 shows the evolution of heat accumulation during the year, for the top 6 quality vintages and for the bottom 6 quality vintages, where it is observable the tendency for earlier accumulation of heat in the best vintages (smaller ordinal dates), when compared to the worst vintages.

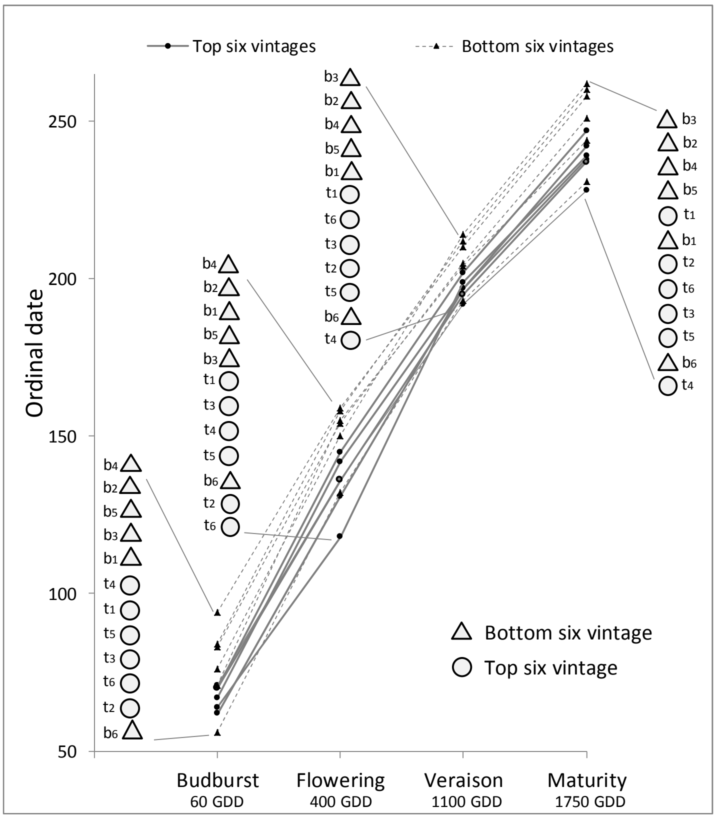


Figure 43 - Evolution of heat accumulation for top 6 vintages and bottom 6 vintages (in quality) and corresponding averages.

The fact that the average dates of budburst in 1980-2009 had a significant advance tends to advance the dates of all the following events. This fact makes the comparisons between growth interval lengths more informative than the comparisons between the phenology dates.

The analysis of the growth interval lengths, comparing the average lengths of the growth intervals in top quality and bottom quality vintages showed that the lengths of the end-of-dormancy intervals are smaller in better vintages, indicating that an early budburst (related to a warm winter) tend to promote a good outcome for the quality of the vintage. A similar situation happens with the average lengths of the ripening intervals, where interval lengths are shorter in better vintages, indicating that warm temperatures after *véraison* tend to promote a good outcome for the quality of the vintage. The average lengths of the berry development intervals also showed to have significant differences between top quality vintages and bottom quality vintages. Interestingly, the lengths of berry development intervals, differently from the other growth intervals, are larger in better vintages indicating that cool temperatures in the period from the flowering event to the *véraison* event tend to promote a good outcome for the vintage.

For all growth intervals, the differences between the average lengths of top *n* quality vintages and the average lengths of bottom *n* quality vintages tend to decrease when the number of vintages in top *n* and bottom *n* vintages increases. This fact indicates that as the two groups of vintages become closer, the differences tends to fade away, becoming the two groups not differentiable in terms of growth interval lengths.

7.5 Synthesis of the Analysis Results

The standardized coefficients form the logistic regression models in section 7.4.1, the relative influence values from top *vs* bottom ranked vintages analysis in section 7.4.2, and the information on the growth intervals' length in section 7.4.2.2, express different assessments of the relative importance that each weather variable has in influencing the likelihood of a vintage to be a high quality vintage or a high yield vintage. The analysis of their relative values reveals good concordance between the variables captured by three methodologies.

In order to congregate the information from the three methods regarding the relative importance of each weather related variable we counted the number of times that each variable was selected as significant in the three methodologies (Figure 44). Negative values indicate that small values of the corresponding variable promote quality or yield and positive values indicate that large values of the corresponding variable promote quality or yield. As the analysis in section 7.4.2.2 did not contemplate precipitation variables, the three variables related to precipitation (PT1, PT2, and PT3) were counted from the results of the analysis in sections 7.4.1 and 7.4.2 and may have a maximum absolute value of two while for the remaining variables this value is three. We note that the values for each variable have no meaning other than to express the relative "strength" for a variable to be more or less influential than the other in the outcome of a vintage.

The analysis of Figure 44 reveals two distinct weather profiles; one that promotes the likelihood of a vintage to be a high quality vintage, and other that promotes the likelihood of a vintage to be a high yield vintage.

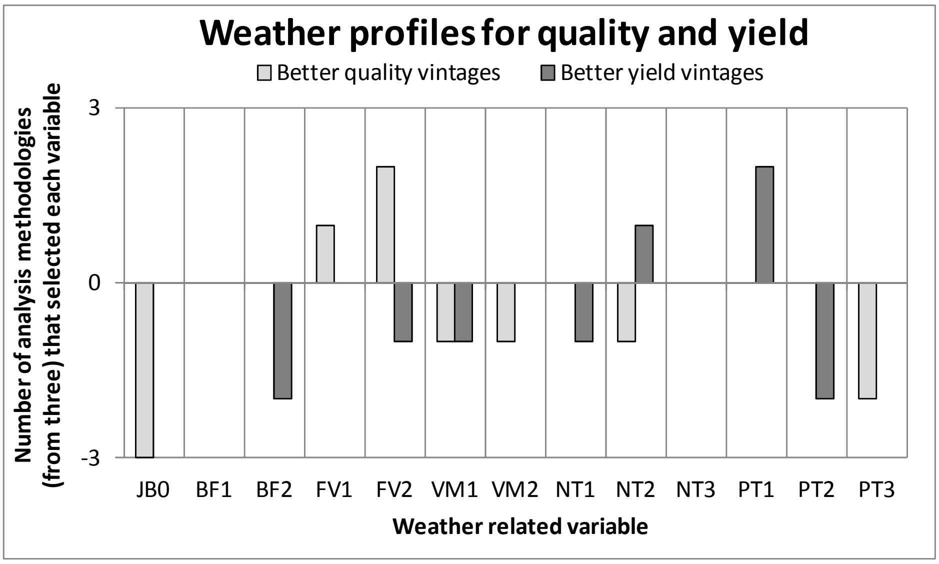


Figure 44 - Weather profiles for high quality and for high yield vintages.

Weather profiles for high quality and high yield vintages are very different, showing variables with opposite signs and variables that are present in one profile but not in the other. This fact supports the idea that a high quality vintage is most often a low yielding vintage.

In order to allow a simpler interpretation of the weather profiles related to high quality vintages and to high yield vintages shown in Figure 44 we will next present these profiles, adopting two commonly used weather variables; mean temperature and average precipitation amount. The scale of y-axis has no meaning other than showing if the temperature or the precipitation amount are below, or above, their average values.

The temperature profiles that enhance the probability of a vintage being a quality vintage or to be a high yield vintage are shown in Figure 45. The reference moments in x-axis are the main phenological events that delimit the main growth intervals. The main phenological events are not fixed in time, varying from one year to another depending on accumulated heat but, to facilitate this presentation, we will relate them to their observed average dates. The boundaries of temperature rectangles in Figure 45 may not coincide with a phenological event as some weather variables were defined dividing growth intervals in two halves. For example, variable FV2 was defined as the length of second half of the Flowering to *Véraison* interval and the corresponding temperature rectangle in Figure 45 begins half way from the Flowering and *Véraison* events and ends at the *Véraison* event.

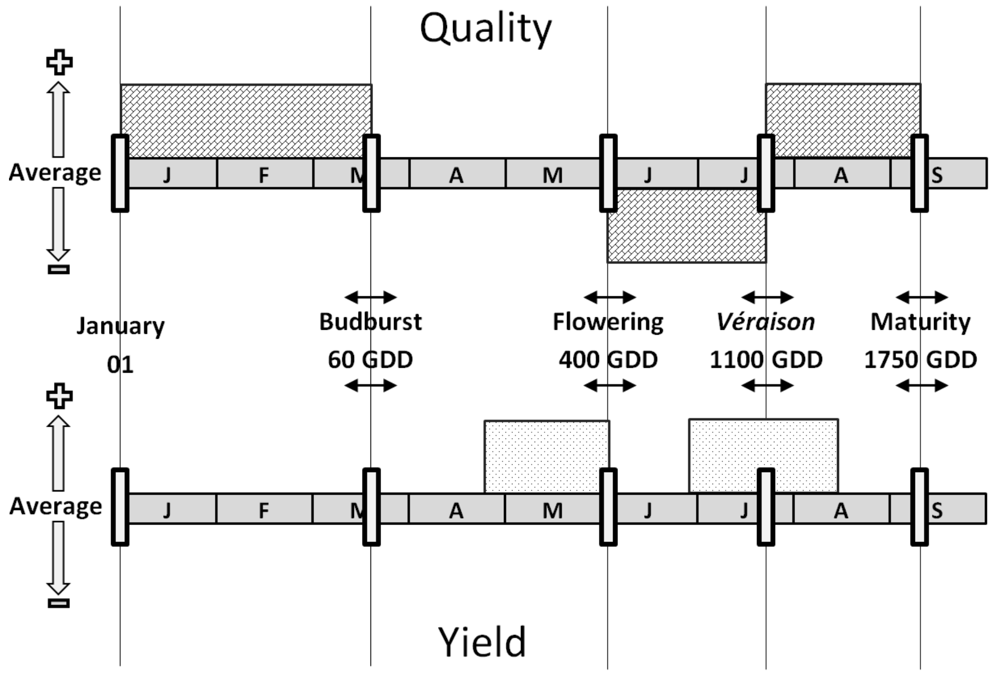


Figure 45 - Temperature profiles of high quality vintages (top) and high yield vintages (bottom).

The temperature profile that enhances the probability of a vintage to be a quality vintage has mean temperatures above average from January to March, below average from mid June to the end of July, and above average from early August to the harvest in mid-September. The temperature profile that enhances the probability of a vintage to become a high yield vintage has mean temperatures above average in May and from July to mid-August.

The precipitation profiles that enhance the probability of a vintage being a quality vintage or to be a high yield vintage are shown in Figure 46. The profile that enhances the probability of a vintage being a quality vintage has average precipitation in the first two trimesters of the year (January to June) and precipitation below average from July to the harvest. The profile that enhances the probability of a vintage being a high yield vintage has precipitation above average from January to the March, precipitation below average from April to June, and average precipitation from July to the harvest.

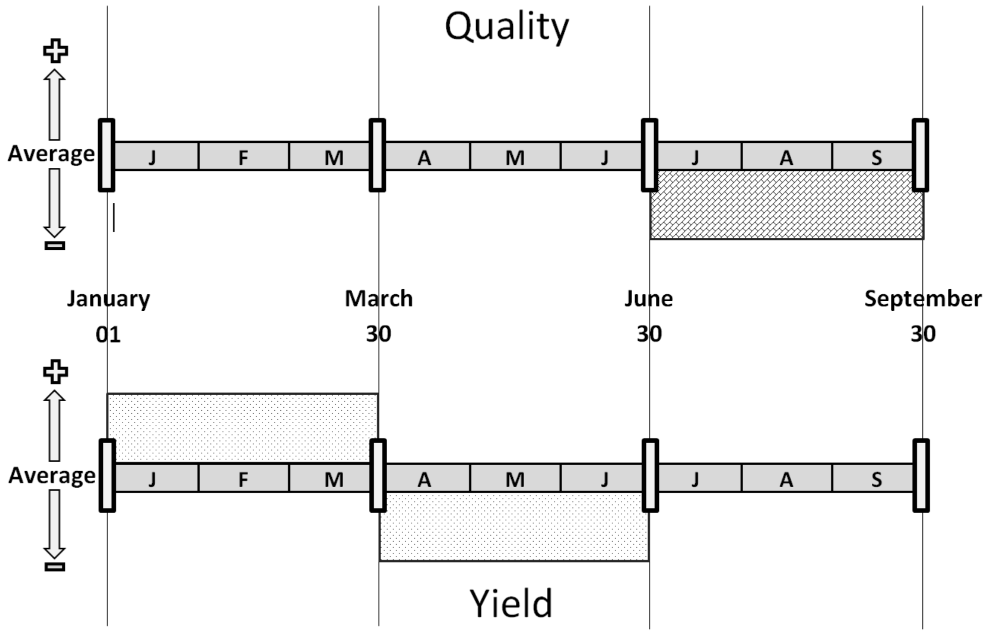


Figure 46 - Precipitation profiles of high quality vintages (top) and high yield vintages (bottom).

For the definition of precipitation related variables we decided to use calendar dates (trimesters, in the case) as explained in page 117. It would be difficult to interpret differences in the precipitation profiles if, instead of using a growing season partition based on calendar, we had adopted a partition based on phenology.

7.6 Validation of the Results using the Weather Dataset

An analysis was conducted in order to validate our results and conclusions regarding the weather profiles that promote the likelihood of a vintage to be a high quality vintage or a high yield vintage (see Figure 45 and Figure 46). We analyzed the agreement between the observed values for all weather variables in each vintage (see Table 20, page 120) and the variables' values that according the conclusions in section 7.4 should promote quality or yield. First, we calculated the z-scores for all yearly variables' values in Table 20 to obtain the number of standard deviations that each variable was yearly, above or below its average value for the 30-year research period (Table 41). We only considered z-scores with an absolute value above a threshold value (we used 0.5 which corresponded to retaining 62% of the variable values in Table 41). We considered that smaller z-scores correspond to variable values that are very close to their average values and should not be accounted as values that significantly promote quality or yield. In Table 41, cells with gray background correspond to absolute values of z-scores below the considered threshold.

Table 41 – z-scores of variables in Table 20.

Year	JB0	BF1	BF2	FV1	FV2	VM1	VM2	NT1	NT2	NT3	PT1	PT2	PT3
1980	0.9			0.5	0.9				-0.8				-1.1
1981		0.8	0.6	-1.3	-0.9	-0.5		-1.3	1.5	1.4	-1.1		
1982						0.5		1.1	-0.8	0.5	-1.2	-0.5	0.9
1983	-0.6	2.8	-1.0	-0.5	0.9	2.4			-1.1	0.9	-1.3	2.2	
1984	1.9	-2.1	3.3	-0.9		1.0	1.1	-0.9	0.6	1.3		0.6	-1.0
1985	1.5	-0.7		-0.5		1.4	-0.6	-0.6			1.7		-1.5
1986	1.3	0.5	-1.6	-0.5	-0.9	1.4		1.6	1.4	1.2	0.8	-0.9	
1987	-1.1	0.7		0.7		-1.0			0.5			-0.6	0.9
1988		0.8		1.3		1.9	0.6	-1.8	-1.8	0.9	-0.5	2.0	-1.0
1989			-0.8		-1.3	-1.0		0.6	0.5	-0.7	-0.9	0.7	-1.0
1990	-1.8	1.3				-1.4		0.6		-1.3	-0.7	-0.8	0.5
1991	1.2		-1.8		-1.3	-0.5		1.3	0.8	-0.8	0.8	-1.4	
1992			-1.6	1.8	-1.3	-1.0		1.5		-0.5	-0.9		
1993	1.2		0.6	-0.5		0.5	4.3	-1.1	0.8	2.6	-1.7	1.6	1.3
1994	-0.5		1.2			1.0	0.9		0.5	0.9			-0.7
1995		-1.6	0.6	1.1	1.3	-0.5	-0.6	2.0	-1.7	-0.5		-0.6	
1996	1.1	-0.5		-0.9			1.3	-1.3	0.9	1.4	2.6		
1997	-1.2	-1.5		3.2	2.7	-0.5		1.1	-2.0		-0.6		1.3
1998	-1.4	0.8		0.9		-1.0	-1.1	-1.1	-0.6	-1.5		1.3	1.2
1999	0.6	-0.5			-1.3	0.5			0.5		-0.9		2.9
2000	-0.9	1.3	-0.8					-0.6			-1.8	1.6	
2001			1.2	-0.9	0.9	-0.5		-1.1	1.1		6.9	-0.8	-0.6
2002		-1.0	0.8						0.5			-1.3	0.8
2003	-0.5			-0.7		-1.0	-1.3			-1.9	1.2	-0.5	-1.0
2004	-0.9			-1.3		-1.4			0.5		-0.9	-1.4	
2005	0.6	-0.9		-1.3	-1.3	-0.5	-0.9	-1.1	2.0	-0.8	-1.2	-1.2	-1.0
2006	0.6	-1.3	-1.0	-0.5		-1.4	-0.6		-0.6	-1.1		-0.9	1.4
2007	-1.2			1.1	1.8	-0.5			-1.8				
2008	-1.6	1.0		0.7		1.0		-0.6	-0.5	0.7	-0.5	1.2	
2009	-0.6				1.3	0.5	-0.9	1.3		-0.7		-0.5	-1.5

In Table 41, every z-score larger than 0.5 was replaced by a one and the remaining by a zero, meaning that, independently of the z-score value, a variable value is either considered as promoting quality or yield, or it is not. Table 42 shows the transformed table.

For each vintage in 1980-2009, two weighted averages were calculated, one for quality and one for yield using the yearly transformed values in Table 42. The weights used for each variable are the values shown in Figure 44 and represent the "strength" of each weather variable to promote quality or yield.

Table 42 – Transformed values of Table 41.

Year	JB0	BF1	BF2	FV1	FV2	VM1	VM2	NT1	NT2	NT3	PT1	PT2	PT3
1980	1	0	0	1	1	0	0	0	1	0	0	0	1
1981	0	1	1	1	1	1	0	1	1	1	1	0	0
1982	0	0	0	0	0	1	0	1	1	1	1	1	1
1983	1	1	1	1	1	1	0	0	1	1	1	1	0
1984	1	1	1	1	0	1	1	1	1	1	0	1	1
1985	1	1	0	1	0	1	1	1	0	0	1	0	1
1986	1	1	1	1	1	1	0	1	1	1	1	1	0
1987	1	1	0	1	0	1	0	0	1	0	0	1	1
1988	0	1	0	1	0	1	1	1	1	1	1	1	1
1989	0	0	1	0	1	1	0	1	1	1	1	1	1
1990	1	1	0	0	0	1	0	1	0	1	1	1	1
1991	1	0	1	0	1	1	0	1	1	1	1	1	0
1992	0	0	1	1	1	1	0	1	0	1	1	0	0
1993	1	0	1	1	0	1	1	1	1	1	1	1	1
1994	1	0	1	0	0	1	1	0	1	1	0	0	1
1995	0	1	1	1	1	1	1	1	1	1	0	1	0
1996	1	1	0	1	0	0	1	1	1	1	1	0	0
1997	1	1	0	1	1	1	0	1	1	0	1	0	1
1998	1	1	0	1	0	1	1	1	1	1	0	1	1
1999	1	1	0	0	1	1	0	0	1	0	1	0	1
2000	1	1	1	0	0	0	0	1	0	0	1	1	0
2001	0	0	1	1	1	1	0	1	1	0	1	1	1
2002	0	1	1	0	0	0	0	0	1	0	0	1	1
2003	1	0	0	1	0	1	1	0	0	1	1	1	1
2004	1	0	0	1	0	1	0	0	1	0	1	1	0
2005	1	1	0	1	1	1	1	1	1	1	1	1	1
2006	1	1	1	1	0	1	1	0	1	1	0	1	1
2007	1	0	0	1	1	1	0	0	1	0	0	0	0
2008	1	1	0	1	0	1	0	1	1	1	1	1	0
2009	1	0	0	0	1	1	1	1	0	1	0	1	1

The variables' weights corresponding to Figure 44 are presented in Table 43.

Table 43 - Weights for the weighted averages used in quality and yield prediction.

	JB0	BF1	BF2	FV1	FV2	VM1	VM2	NT1	NT2	NT3	PT1	PT2	PT3
Quality weights	-3	0	0	1	2	-1	-1	0	-1	0	0	0	-3
Yield weights	0	0	-2	0	-1	-1	0	-1	1	0	3	-3	0

The calculated weighted averages represent an assessment on a scale from 0 (no agreement) to 1 (complete agreement) of the level of agreement between each vintage variables' values and the weather profiles that promote good quality vintages or high yield vintages. These weighted averages were multiplied by 100 to obtain, in a 0-100 scale, a predicting score for vintage quality and vintage yield (Table 44).

Table 44 - Predicting scores for vintage quality and vintage yield (larger = better).

Year	Quality score	Yield score
1980	58	0
1981	8	33
1982	8	25
1983	54	33
1984	25	25
1985	33	33
1986	0	83
1987	42	33
1988	42	8
1989	33	33
1990	33	33
1991	8	92
1992	17	33
1993	4	17
1994	50	0
1995	50	33
1996	0	42
1997	54	13
1998	58	17
1999	13	8
2000	25	25
2001	50	63
2002	0	25
2003	67	33
2004	33	33
2005	42	58
2006	25	50
2007	67	8
2008	33	8
2009	75	25

Predicted scores for vintage quality and yield were ranked in order to be comparable to the consensus ranking for quality and with the ranking of the vintage yields based on the observations from the collected data (Table 45). Ties in the obtained rankings were untied using the rankings of the consensus ranking.

Association between the underlying populations of observed and predicted quality and between the underlying populations of observed and predicted yield was analyzed using the Spearman's rank test. Test results for the null-hypothesis of no-association between the underlying populations of vintage quality and predicting scores for quality and the underlying populations of vintage yield and predicting scores for yield showed that, both for quality and yield, the null-hypothesis was rejected with p-value < 0.001. The results indicate a statistically significant lack of independence between predicted quality and yield resulting from the analysis models (based on the weather and phenology variables) and the observed vintage quality and yield, supporting the existence of association.

Table 45 - Rankings for quality and yield, calculated using the observed data and the predicting scores.

	Quality rankings			Yield rankings		
	Observed	Predicted		Observed	Predicted	
Year	Q_Rank1	Q_Rank2	Abs. difference	Y_Rank1	Y_Rank2	Abs. difference
1980	15	4	11	5	29	24
1981	30	26	4	19	15	4
1982	17	25	8	10	17	7
1983	7	7	0	12	11	1
1984	27	21	6	15	18	3
1985	8	14	6	7	9	2
1986	26	30	4	16	2	14
1987	18	12	6	2	8	6
1988	28	13	15	28	28	0
1989	14	16	2	14	12	2
1990	25	18	7	1	7	6
1991	9	24	15	6	1	5
1992	10	22	12	17	13	4
1993	29	27	2	29	22	7
1994	1	8	7	27	30	3
1995	13	9	4	21	16	5
1996	24	29	5	3	6	3
1997	6	6	0	26	24	2
1998	20	5	15	30	23	7
1999	23	23	0	8	25	17
2000	3	19	16	20	19	1
2001	19	10	9	4	3	1
2002	22	28	6	22	20	2
2003	4	3	1	9	10	1
2004	12	15	3	18	14	4
2005	11	11	0	11	4	7
2006	21	20	1	13	5	8
2007	2	2	0	23	26	3
2008	16	17	1	24	27	3
2009	5	1	4	25	21	4
			Σ = 170			Σ = 156

In order to have a measure of the degree of agreement between underlying populations of the observed and predicted rankings on quality and yield we calculated Kendall's Tau (τ). This non-parametric measure is based on the probabilities of observing concordant[23] and discordant pairs of ranks among the C_2^n possible ways of selecting distinct pairs of the n ranked items ($n = 30$ vintages, in this case). It can be interpreted in terms of probability - it is the difference between the probabilities that the variables vary in the same direction and the probabilities that the variables vary in the opposite direction. For the quality rankings, Q_Rank1 and Q_Rank2 in Table 45, $\tau = 0.45$ and for the yield rankings, Y_Rank1 and Y_Rank2 in Table 45, $\tau = 0.53$, rejecting the null-hypothesis of no-association between the underlying populations with p-value < 0.001. For the quality rankings, τ indicates that in 72.2% of the C_2^{30} possible ways of selecting distinct pairs of vintages, the ordering in Q_Rank1 and Q_Rank2 is the same. For the yield rankings, τ indicates that in 76.6% of the C_2^{30} possible ways of selecting distinct pairs of vintages, the ordering in Y_Rank1 and Y_Rank2 is the same.

To complement the association analysis, we compared the top 15 lists of vintages in quality and yield using the consensus quality ranking, the quality ranking based on the predicting scores, the observed yield ranking, and the yield ranking based on the predicting scores. Table 46 presents the vintages in top 15 lists for observed and predicted values, showing that for both quality and yield, 11 out of the 15 vintages are common to the observed and predicted lists, representing a concordance of 73.3%.

According to the hypergeometric distribution function, 0.0134 is the probability that in two top 15 lists of ranked vintages, selected from a group of 30 vintages, 11 or more than 11 vintages are common. This value represents the p-value for the rejection of the hypothesis that the observed 11 common vintages just happened by chance.

Both the results of association analysis and the results of the analysis of the common vintages in top 15 lists reveal a significant association between the underlying populations of observed and predicted quality and between the underlying populations of observed and predicted yield. This high level of association between observed and predicted values shows that, in the analysis performed in section 7.4, the variables found to have significant influence on vintage quality and yield are able to explain a substantial part of the variability of the vintages of Vintage Port, in terms of both of quality and yield.

---

[23] Concordant pair of observations: a pair of observations (a pair of vintages, in the present case) with the same ordering in both rankings.

Table 46 – Top 15 lists for a) vintage quality and b) vintage yield.

15 top quality vintages

Observed Year	Predicted Year	Common
1994	2009	1980
2007	2007	1983
2000	2003	1985
2003	1980	1994
2009	1998	1995
1997	1997	1997
1983	1983	2003
1985	1994	2004
1991	1995	2005
1992	2001	2007
2005	2005	2009
2004	1987	
1995	1988	
1989	1985	
1980	2004	

a)

15 top yield vintages

Observed Year	Predicted Year	Common
1990	1991	1991
1987	1986	2001
1996	2001	2005
2001	2005	2006
1980	2006	1996
1991	1996	1990
1985	1990	1987
1999	1987	1985
2003	1985	2003
1982	2003	1983
2005	1983	1989
1983	1989	
2006	1992	
1989	2004	
1984	1981	

b)

Chapter 8

Discussion

8.1 Introduction

In this chapter, the main findings with regard to the research questions are summarized and a critical reflection on the strengths and limitations of the selected approaches is provided. We proposed a method to obtain a measure of vintage quality congregating the information available from a set of independent sources into an impartial consensus ranking. A method for the partitioning of the growing season when historical phenology data are not available, based on heat accumulation, was proposed. The results of climate trends analysis have shown trends in the Douro Valley climate in 1980-2009: a moderate increase in the annual mean temperature and a large increase in the mean temperature during the grapevine growing season. We showed that vintage quality increased steadily during 1980-2009 and was moderately linked to climate trends but that the evolution wine yield showed no links to climate trends. Moreover, a high association between the retail prices of Vintage Port and published expert ratings was highlighted. No association was found between the evolution of Port wine release prices in 1980-2009 and the evolution in Port wine quality. Finally, we were able to relate different weather profiles of temperature and precipitation to high quality and to high yield vintages.

8.2 The Assessment of Vintage Quality

The process of finding an adequate measure of the vintage quality is a challenging task due to the availability of information and the inherent subjectivity in assessing quality. Most research on the influence of the weather on vintage quality makes use of the ratings of a single expert or expert's panel. Wine tasting is a sensory experience based on personal tasting skills, culture, memory, and fashion. Several tasters may have different opinions about the same wine. A new method was proposed to overcome this issue. The proposed method uses a rank aggregation technique to combine a collection of vintage chart ratings into a ranking of the vintages that represents the consensus of the input vintage charts. The method makes use of the information available from a set of independent sources and congregates it into an impartial ranking of a region's vintages over the years. We believe that the consensus raking resulting from the proposed method is a better alternative to the use of the ratings of a single expert or expert's panel since we have shown that expert ratings are far from being consensual. The proposed method has the potential of being a useful tool for wine research that requires an impartial assessment of the vintage quality for a given wine region.

8.3 Partitioning of the Growing Season using Heat Accumulation

Partitioning the grapevine growing season into smaller growth intervals is necessary for studying the relationships of wine quality to weather and climate variability. In this research we were able to estimate the historical phenological dates based on their average heat accumulation values. However, in general it is difficult to have access to consistent data with the dates of the four main developmental stages for grapevines that covers a whole region for an extended period. When no historical data are available and it is not possible to estimate them, the partitioning of a growing season may be achieved by defining interval boundaries using different methods: i) by mean values of the heat requirements of the main phenological events and ii) by generalized calendar average dates associated with the occurrence of the main phenological events.

We analyzed differences in temperature and precipitation when using growth intervals with boundaries defined by the estimates of historical dates of the main phenological events (used as reference) and growth intervals with boundaries defined by two methods. The results showed high concordance between the temperature and precipitation profiles obtained using historical dates of the main phenological events and the corresponding values obtained from growth intervals defined using heat accumulation. A much smaller concordance was obtained when using growth intervals defined by means of generalized calendar average dates associated with the occurrence of the main phenological events.

When looking for links between wine characteristics and the weather, the partitioning of the grapevine growing season into smaller intervals where variables are evaluated and compared to the same intervals in different years is inevitable. Many researchers have used a partitioning based on calendar dates to compare weather variables. However, the same calendar period in different years may refer to different stages of the annual cycle of the grapevine, making the comparisons less accurate. The use of heat accumulation and its relation to grapevine phenology in the partitioning of the growing season makes variables' comparisons between different years refer to the same stage of the grapevine annual cycle.

8.4 Climate Trends in the Douro Valley

We analyzed the cleaned and homogenized datasets of daily maximum temperature, minimum temperature and precipitation amount. Temperatures were referred to an elevation of 250 m to uncouple them from elevation. Average values of the weather stations of *Régua, Pinhão* and *Mirandela* were used to represent the overall Douro Valley weather. We are aware that the choice of a reference elevation other than 250 m would change the results of climate analysis for temperatures. However, the results on climate trends, related to variations in time, would maintain for both temperatures and precipitation (not affected by the reference elevation).

Our results point out an increase of 1.5 ºC in the growing season mean temperature, from an average value of 19.6 ºC in the early 80s to an average value of 21.1 ºC in 2009. Jones (2012) showed results that indicate an increase in growing season mean temperature for the Douro Valley of 1.7 ºC in 1960-2005. Our results point out an increase of 1.0 ºC from an average value of 14.9 ºC in the early 80s to 15.9 ºC at present time. Jones (2012) also showed that an increase in the annual mean temperature of 0.8 ºC. In his work, Jones used meteorological datasets of daily maximum temperature, minimum temperature and precipitation amount collected at the three local weather stations, *Vila-Real*, *Régua, and Pinhão*, the last two used in this research. We believe that the use of data from the weather station of *Mirandela,* in replacement with the station of *Vila-Real* may explain the differences between our results and the results obtained by Jones. *Mirandela* is one of the hottest places inside the Douro Valley, located near the Douro Superior sub-region and *Vila-Real* is located in a cooler zone of the Douro Valley. Regarding annual precipitation, Jones (2012) did not find significant trends as our results also point out. Our results showed that the increase in the growing season mean temperature was not homogeneous, showing larger increase in March-June (2.34 ºC from 1980 to 2009) and smaller increase in July-September (0.66 ºC from 1980 to 2009). The results also showed that in March-June the increase in mean temperatures was caused by an increase in both maximum and minimum temperatures and that in July-September the increase in mean temperatures was exclusively caused by an increase in minimum temperature. The Jones (2012) results are not as detailed as our, but overall confirm our results.

We obtained evidence of trends in the growing season mean temperature, during 1980-2009. The results showed that the growing season mean temperature in the region is reaching the upper limit of 21 ºC for fortified wines for high quality wine production (Jones 2012). Further increase in growing season temperature will gradually place the region outside its theoretical optimum. The Douro Valley economy depends essentially on the wine sector, making inevitable that wine producers prepare a contingency plan for the next decades. The region's topography may positively provide some possibilities of mitigating the effects of climate change, by moving vineyards to sites with an orientation and elevation that promote a better weather profile. Additionally, adjusting agricultural practices, adequating canopy management, shading, selecting rootstocks better adapted to water limitation and warmer climate, growing varieties with different thermal requirements and higher summer stress resistance, and adopting adequate irrigation schemes (in the Douro Valley, natural rainfall is presently considered the only source of water) should be considered.

Great attention should be devoted to the monitoring of climate change, as climate is determinant of the quality / yield of vintages and indirectly to the Douro Valley economy.

8.5 Weather Relation to Vintage Quality and Yield

Three different approaches have been used to analyze the relation between weather variability and vintage quality / yield variability: i) logistic regression, ii) top *vs* bottom ranked vintages analysis for weather variables, and iii) top *vs* bottom ranked vintages analysis for phenological variables.

Preliminary analysis of the data was performed to check the assumption of moderate multicollinearity with respect to the weather variables used as predictors in the Logit models. Logit models produced a collection of mathematical expressions that represent the probability of the quality or yield of a vintage to belong to the top *n* ranking of the 30 vintages as a function of the values of a set of significant weather variables. The size of our sample (data from 30 vintages) was insufficient to perform a cross-validation of models results. For this reason, the obtained logistic regression models should be considered as descriptive models. However, the overall correct classification of the sample data by the obtained models was very high (see Table 33 to Table 36), globally above 75%.

As a strategy to validate that the weather variables selected as significant by the logistic regression models were also selected as significant by other methodologies, we analyzed the central tendency of the weather variables and of the phenology variables in top and bottom ranked vintages, looking for differences.

Both logistic models and top *vs* bottom ranked analysis identified with good agreement a set of weather variables related to high quality vintages and a different set of variables related to high yield vintages. These sets of variables define two different seasonal weather patterns related to high quality vintages and to high yield vintages. In Figure 44, page 171, we showed these two different weather patterns in terms of the relative "strength" of the correspondent weather variables.

The temperature characteristics that enhances the probability of a vintage being a quality vintage has mean temperatures above average from January to March, below average from mid-June to the end of July, and above average from early August to the harvest in mid-September. The temperature characteristics that enhance the probability of a vintage being a high yield vintage has mean temperatures above average in May and from July to mid-August.

The precipitation characteristics that enhances the probability of a vintage being a quality vintage has average precipitation in the first two trimesters of the year (January to June) and precipitation below average from July to the harvest. The precipitation characteristics that enhances the probability of a vintage being a high yield vintage has precipitation above average from January to the March, precipitation below average from April to June, and average precipitation from July to the harvest.

When comparing these profiles to the results from different research, we found that the precipitation profile that enhances the probability of a vintage becoming a high yield vintage is a perfect match with the results of Santos & Malheiro (2011). With respect to the temperature profile that enhances the probability of a vintage to be a high quality vintage, our results are an almost perfect match with the results of Mattis (2011), although his research was conducted in Sonoma County in California.

8.6 Influence of Vintage Quality and Yield in Retail Price

This research showed that the average values of Vintage Port retail prices are strongly related to vintage quality and are not related to vintage yield. A significant correlation ($r = 0.87$) exists between Vintage Port international retail detrended prices and a consensus ranking that represents the best common perception of quality from eight wine experts.

Statistical analysis of the agreement between the quality of a wine perceived by expert tasters when compared to the quality of the same wine perceived by non-expert tasters shows very small concordance (Schiefer & Fischer 2008; Goldstein & Almenberg 2008). The existence of very limited agreement between expert and non-expert tasters together with the fact that the average retail price of the vintages of Vintage Port has an almost perfect match with the vintage consensus ranks suggests that the influence of expert ratings, published on wine magazines and internet sites, is strongly determinant on Vintage Port retail price formation.

The fact that Vintage Port retail prices are on average 250% to 350% higher than general Port wine average release prices, pushes the region's vintners to declare every year (subject to confirmation of IVDP) some of their brands as Vintage Port. The "30% rule" states (*arts. 21.º e 22.º do Regulamento da Denominação de Origem vinho do Porto, Decreto-Lei n.º 166/86*) that in each year a Port house will be allowed to sell 30% of that year's production and that the remaining 70% of it will go to stock and will be sold gradually in the following years. Vintage Port prices tend to increase with the time span to the moment that a vintage first arrives to market. Cumulatively, the perceived quality of a particular vintage influences strongly the average retail price of that vintage. The top six vintages of Vintage Port in 1980-2009 have, presently, an average retail price around 63 €/bottle, while the bottom six vintages have an average retail price around 22 €/bottle.

Nature, dictating the weather characteristics of a vintage is indirectly dictating the market price that a particular vintage will have a few years later. Vintage Port prices may go well over 70 €/bottle if a particular vintage is of top quality and under 20 €/bottle if a particular vintage is of poor quality. This difference will certainly affect the revenue of the Port houses in Douro Valley as well as the economy of the region.

8.7 Influence of Vintage Quality and Yield in Port Wine Release Price

Analysis of the evolution of the quality of the vintages of Vintage Port showed that vintage quality steadily improved during 1980-2009. As the quality of the vintages of Vintage Port depends on the yearly characteristics of the weather and on vintners' skills and knowledge, it is difficult separate both influences, when analyzing the improvement of vintage quality.

The evolution of the average release prices of Port wine, at constant prices, showed a decrease in prices from the end of the 1970s throughout 2012. The reason for the decrease in the average release price of Port wine in 1980-2009 is not related with quality, as this increased steadily. In the same time period, wine yield decreased steadily in the Douro Valley. Smaller quantities of higher quality Port wine would be expected to influence prices to increase and not to decrease. These facts support that there is no relation between wine yield and Port wine release prices.

The negative evolution of the average release prices of Port wine could possibly be related to the positioning of Portuguese wines and particularly of Port wine in international wine markets. The brand Portugal and Port wine should be better promoted internationally and a consistent policy of quality production, price formation and well-oriented marketing for the wine sector should be undertaken.

Chapter 9

Conclusions

This research was set out to explore the influence of the yearly weather variability on vintage quality and yield. To the best of our knowledge, this subject has never been studied for the Douro Valley using Vintage Port quality and wine yield from all types of region's wines. The research also sought to know whether the Douro Valley climate showed trends in 1980-2009 and what will be the implications for the region’s viticulture, if past trends persist. Additionally, the relation of vintage quality and yield to retail and release prices was investigated.

We identified the common means, used in wine research, to assess vintage quality and proposed a new method that makes use of a rank aggregation method to combine a collection of vintage chart ratings into a consensus ranking of the vintages. The vintage quality consensus ranking for Vintage Port in 1980-2009 was computed.

We analyzed several methods for partitioning the growing season into smaller growth intervals where weather variables may be evaluated and compared between different years. A method based on heat accumulation was proposed for the partitioning in the absence of historical phenology data.

We collected meteorological data series of daily maximum temperature, minimum temperature and precipitation. The data series were cleaned and homogenized, prior to use. Using the cleaned and homogenized weather data, the climate of the Douro Valley was analyzed searching for trends in the annual mean temperature, in the growing season mean temperature, in the annual precipitation amount, in the precipitation during the growing season, and in extreme temperatures. We investigated the association between climate trends and the evolution of vintage quality and wine yield in 1980-2009.

Data that were sparse was integrated into a single dataset. Data on production, yield, release prices of Port wine, and on the retail prices of Vintage Port, allowed us to analyze the impact that the variability of the Douro Valley weather has, indirectly, on the prices of the wine.

Preliminary statistical analysis was conducted in order to determine the properties of the defined variables and to select appropriate analysis methods. A subset of the defined variables was considered adequate to be used in the analysis. The relation of vintage quality to the subset of weather variables was conducted using the following statistical techniques:

- logistic regression;
- top *vs* bottom ranked vintages: comparisons of weather variables;
- top *vs* bottom ranked vintages: comparisons of phenology variables (dates and interval lengths).

9.1 Main Contributions

The following, are the main research contributions of this thesis:

- **Assessment of wine quality**: most researchers use the ratings of a single taster or tasting panel as a measure of a vintage quality. This measure may not represent the overall opinions on that vintage. We proposed a consensus ranking as a better measure of vintage quality;
- **Partitioning of grapevine growing season using phenology**: we analyzed the partitioning the growing season into smaller growth intervals where variables are assessed and compared to the corresponding values in other years. Most researchers partition the growing season using calendar-defined boundaries. As plant phenology varies from one year to another influenced by the weather we showed that, when the weather variables to be assessed are related to heat, the use of growth intervals with boundaries fixed in time is not the best alternative to the partitioning. We proposed a partition of the growing season based on historical phenology dates, if available. When historical phenology dates are not available, the partitioning should be based on the grapevine heat requirements;

- **Modelling quality / yield using weather variables**: we modeled the top vintages as well as the bottom vintages, both for quality and yield, using logistic regression models. These models identified a set of weather variables that enhance the probability of vintage to be a high quality vintage and a different set of weather variables that enhance the probability of a vintage to be a high yield vintage. These models, complemented with the information that the comparison of the weather variables of top and bottom vintages produced, allowed us to identify several weather dependent variables that have great influence on wine quality and on wine production and the periods when its influence has the strongest effect;
- **Distinctive weather profiles**: we presented two distinct weather profiles, one that enhances the likelihood that a vintage becomes a high quality vintage and another that enhances the likelihood that a vintage becomes a high yield vintage. Variables that enhance the probability of a vintage becoming a high quality vintage are not the same variables that enhance the probability of a vintage becoming a high yield vintage. Precipitation characteristics that are related to high quality vintages and to high yield vintages are nearly opposite;
- **Douro Valley climate trends in 1980-2009**: we detected significant rising trends in the temperatures of the Douro Valley;
- **Vintage Port quality evolution in 1980-2009**: we showed that the quality of Vintage Port vintages steadily increased;
- **Relation between quality evolution and climate trends**: we showed that there is a moderate association between the climate trends observed in the Douro Valley in 1980-2009 and the increase in the quality of Vintage Port. This moderate association suggests that the increase in quality may also be related to the vintners skills and knowledge and to the evolution of the vinification technology;
- **Dependence of retail prices of Vintage Port to experts' assessment of vintage quality**: we showed that the average retail prices of Vintage Port are highly correlated with the consensus opinion of several expert tasters / tasting panes. This fact shows that formation of the prices for the wines from each vintage of Vintage Port is highly influenced by the information that the expert tasters pass to the buyers.

9.2 Future Work

In this work, we analyzed the relation influence of the weather on vintage quality and on vintage yield, considering the Douro Valley as whole. Although the Douro Valley is a small region, it has different climates in different sub-regions. A finer study of the relations between weather and vintage quality and yield, considering each sub-region, would be an important complement to the present research. Moreover, in this work, the assessment of the quality of the vintages was based on expert's ratings in a single moment. Assessment should be carried-out, based on periodic tastings conducted by an impartial and renowned institution such as IVDP, on the evolution of the quality of each vintage with aging. An important line of research would be the study of how the weather factors influence the evolution of the quality of the wines with aging.

The interaction between weather variables should be included to improve models. For example, the interaction effects between some weather variables and weather driven factors such as vine diseases should be better studied and included in the models.

In this work, a partitioning of the growing season using heat accumulation was suggested as the better partitioning method in the absence of historical phenology data. The quantification of vine heat requirements needs deeper study in order to differentiate all varieties, to quantify the influence of the type of soil, to differentiate different profiles of heat accumulation, to validate the 10 ºC generally used as baseline temperature and to define temperature cutoff values to be used in the heat summation process.

This work focused on the period from 1980 to 2009. Longer periods should be studied in order to achieve a better understand the relation of the weather in vintage quality and yield. Further work should be done with respect to the organization of a database with quality data from the Douro Valley. Such a database should keep records on: meteorological data - temperatures, precipitation amount, hours of sunshine, wind intensity, wind direction, hail occurrences, and frost occurrences; soils - physical and chemical characteristics, granulometry, drainage and water holding capacity; vineyards - area and location; grapevine diseases - dates and location where diseases happened; release prices - values for each type of wine produced in the region; and production - values for each type of wine, for each sub-region, for each grape variety.

References

Adams, R., Wu, J. & Houston, L., 2003. *The Effects of Climate Change on Yields and Water Use of Major California Crops*, California Energy Commission, Sacramento, USA.

Addy Bassin, 2011. Vintage Chart 1978 - 2008. *bassins.com*. Available at: http://www.bassins.com/resources/vin_chart.html [Accessed December 2, 2011].

ADVID, 2012. Report: Ano Vitícola 2012. *ADVID - Douro Region Wine Cluster*, p.6. Available at: http://www.advid.pt/imagens/boletins/13560261113826.pdf [Accessed September 12, 2013].

Aguiar, F. et al., 2001. *Plano Intermunicipal de Ordenamento do Território do Alto Douro Vinhateiro: Volume I - Diagnóstico da Situação*,

Aguilar, E. et al., 2003. *Guidelines on Climate Metadata and Homogenization. WCDMP-53*,

Alexandersson, H., 1986. A homogeneity test applied to precipitation data. *International Journal of Climatology*, 6(6), pp.661–675.

Alexandersson, H., 1995. Homogeneity testing, multiple breaks and trends. In *Proceedings of the 6th International Meeting on Statistical Climatology*. Galway, pp. 439–441.

Ali, H., Lecocq, S. & Visser, M., 2008. The Impact of Gurus: Parker Grades and En Primeur Wine Prices. *Journal of Wine Economics*, 5(1), pp.22–39.

Anon, Cambridge Dictionaries Online. Available at: http://dictionary.cambridge.org/.

Ashenfelter, O., 2008. Predicting the Quality and Prices of Bordeaux Wine. *The Economic Journal*, 118(529), pp.174–184.

Ashenfelter, O., Ashmore, D. & Lalonde, R., 1995. Bordeaux wine vintage quality and the weather. *Chance*, 8(4), pp.7–14.

Ashenfelter, O. & Jones, G. V., 2013. The demand for expert opinion: Bordeaux wine. *Journal of Wine Economics*, pp.1–9.

Ashenfelter, O. & Jones, G. V., 2000. The Demand for Expert Opinions: Bordeaux Wine. In *Cahier Scientifique N°3 de Observatoire des Conjonctures Vincoles Europeenenes*.

Baggiolini, M., 1952. Les stades repères dans le développement annuel de la vigne et leur utilisation pratique. *Revue romande d'agriculture, de viticulture et d'arboriculture*, 8, pp.4–6.

Balinski, M. & Laraki, R., 2011. *Majority Judgment: Measuring, Ranking, and Electing* 1st ed., Mit Press.

Bartholdi, J., Tovey, C.A. & Trick, M.A., 1989. Voting schemes for which it can be difficult to tell who won the election. *Social Choice and Welfare*, 6, pp.157–165.

Begg, C.B. & Gray, R., 1984. Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika*, 71(1), pp.11–18.

Bender, R. & Grouven, U., 1998. Using binary logistic regression models for ordinal data with non-proportional odds. *Journal of clinical epidemiology*, 51(10), pp.809–819.

Bindi, M. et al., 1996. Modeling the impact of future climate scenarios on yield and variability of grapevine. *Climate Research*, 7, pp.213–224.

Bonhomme, R., 2000. Bases and limits to using “degree.day” units. *European Journal of Agronomy*, 13, pp.1–10.

Borda, J. de, 1781. Memoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, 2, p.85.

Borges, J. et al., 2012. A New Method to Obtain a Consensus Ranking of a Region's Vintages. *Journal of Wine Economics*, 7(01), pp.88–107.

Box, G. & Cox, D., 1964. An analysis of transformations. *Journal of the Royal Statistical Society*, 26(B), pp.211–246.

Brandão, C., Rodrigues, R. & Costa, J., 2001. *Análise de fenómenos extremos, precipitações intensas em Portugal Continental*, Lisboa.

Broadbent, M., 2007. *Vintage Wine* Harcourt Trade Publishers, ed., London: Pavilion Books.

C. Real, A. et al., 2014. Partitioning the Grapevine Growing Season in the Douro Valley of Portugal: Accumulated Heat better than Calendar Dates. *International Journal of Biometeorology*, (forthcoming).

Cameron, P., Kang, M. & Stark, D., 2010. Random preorders and alignments. *Discrete Mathematics*, 310, pp.591–603.

Carbonneau, A. et al., 1992. *Agrométéorologie de la vigne en France*, Bruxelles, Luxembourg: Commission des Communautés Européennes.

Cardello, A., 2010. Perception of Food Quality. In I. A. T. and R. P. Singh, ed. *Food storage stability*. CRC Press, pp. 1–38.

Cesaraccio, C. et al., 2004. Chilling and forcing model to predict bud-burst of crop and forest species. *Agricultural and Forest Meteorology*, 126, pp.1–13.

Charters, S., 2003. *Perceptions of wine quality*. Edith Cowan University.

Chuine, I., 2000. A unified model for budburst of trees. *Journal of theoretical biology*, 207, pp.337–347.

Cicchetti, D. & Cicchetti, A., 2009. Wine rating scales: Assessing their utility for producers, consumers, and oenologic researchers. *International Journal of Wine Research*, 1, pp.73–83.

Conde, C. et al., 2007. *Biochemical changes throughout grape berry development and fruit and wine quality*, Global Science Books.

Condorcet, J.A.N. de C., 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la probabilité des voix*, Paris: L'imprimerie royale.

Conrad, V. & Pollak, L., 1962. *Methods in Climatology* Harvard Un., Massachusettes.

Cook, W.D. & Kress, M., 1986. Ordinal ranking and preference strength. *Mathematical Social Sciences*, (11), pp.295–306.

Coombe, B.G., 1995. Adoption of a system for identifying grapevine growth stages. *Australian Journal of Grape and Wine Research*, 1, pp.100–110.

Copello, M., 2010. As 100 castas do Douro. *Mar de Vinho*. Available at: http://www.mardevinho.com.br/colunas/castas-douro [Accessed October 5, 2013].

Corsi, A. & Ashenfelter, O., 2001. Predicting Italian Wines Quality fromWeather Data and Expert ratings. In *Cahier Scientifique N°4 de Observatoire des Conjonctures Vincoles Europeenenes*. 4.

Costa, A. & Brito Cunha, J., 2011. The influence of critics' ratings on price formation in international wine markets. In *34th World Congress of Vine and Wine*. International Organisation of Vine and Wine (OIV).

Csörgö, M. & Horváth, L., 1997. *Limit theorems in change-point analysis* Wiley, ed., New York.

Cunha, M., 2001. *Previsão de Colheitas em Viticultura. Integração de Modelos Aeropolínicos e Bioclimáticos*. Universidade do Porto.

Decanter, 2011. vintage port. *decanter.com*. Available at: http://www.decanter.com/wine-learning/vintage-guides/regions/Vintage-Port [Accessed December 2, 2011].

Decanter.com, 2013. Vintage Guides: Vintage Port. *Webpage*. Available at: http://www.decanter.com/wine-learning/vintage-guides/regions/Vintage-Port.

Diaconis, P. & Graham, R.L., 1977. Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society Series B Methodological*, 39(2), pp.262–268.

Dokoozlian, N., 2000. Grape berry growth and development. In A. and N. Resources, ed. *Raisin Production Manual*. University of California, pp. 30–37.

Duarte, F., Madeira, J. & Barreira, M.M., 2010. Wine Purchase and Consumption in Portugal-an Exploratory Analysis of Young Adults' Motives/Attitudes and Purchase Attributes. *Ciência Técnica Vitivinícola*, 25(2), pp.63–73.

Duchêne, E. & Schneider, C., 2005. Grapevine and climatic changes: a glance at the situation in Alsace. *Agronomy for Sustainable Development*, 25, pp.93–99.

Easterling, D. & Peterson, T., 1995. A new method for detecting undocumented discontinuities in climatological time series. *International Journal of Climatology*, 15(4), pp.369–377.

Eichhorn, K.W. & Lorenz, H., 1977. Phänologische Entwincklungsstadien der Rebe. *Nachrichtenblatt des Deutschen Pflanzenschutzdienstes (Braunschweig)*, 29, pp.119–120.

Enthusiast, W., 2013. 2013 Vintage Chart. *Webpage*. Available at: http://www.winemag.com/PDFs/Vintage_Chart_2013.pdf.

Everitt, B. & Dunn, G., 2001. *Applied multivariate data analysis* 2nd ed. Arnold, ed.,

Fagin, R. et al., 2004. Comparing and aggregating rankings with ties. In *Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. ACM, pp. 47–58.

Feng, S., Hu, Q. & Qian, W., 2004. Quality control of daily meteorological data in China, 1951–2000: a new dataset. *International Journal of Climatology*, 24(7), pp.853–870.

Fila, G., Lena, B. Di & Gardiman, M., 2012. Calibration and validation of grapevine budburst models using growth-room experiments as data source. *Agricultural and Forest Meteorology*, 160, pp.69–79.

Gardner, P., 1975. Scales and statistics. *Review of Educational Research*, 45(1), pp.43–57.

Gergaud, O. & Ginsburgh, V., 2008. Natural Endowments, Production Technologies and the Quality of Wines in Bordeaux. Does Terroir Matter?*. *The Economic Journal*, 118(529), pp.142–157.

Gibbs, M., Tapia, M. & Warzynski, F., 2009. Globalization, Superstars, and Reputation: Theory & Evidence from the Wine Industry. *Journal of Wine Economics*, 4(1), pp.49–64.

Gladstones, J., 2011. Temperature: The Driving Force. In *Wine, Terroir and Climate Change*. Kent Town: Wakefield Press, pp. 5–26.

Gleason, E., 2002. *Global daily climatology network, V1. 0*, Asheville.

Gokcekus, O. & Nottebaum, D., 2011. The buyer’s dilemma: To whose rating should a wine drinker pay attention? In *AAWE - American Association of Wine Economists. Working Paper No. 91*.

Goldstein, R. & Almenberg, J., 2008. Do more expensive wines taste better? Evidence from a large sample of blind tastings. *Journal of Wine Economics*, 3(1), pp.1–9.

Grifoni, D., Mancini, M. & Maracchi, G., 2006. Analysis of Italian Wine Quality Using Freely Available Meteorological Information. *American Journal of Enology and Viticulture*, 57(3), pp.339–346.

Hawkins, D., 1977. Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, 72(357), pp.180–186.

Hoare, C., 1962. Quicksort. *The Computer Journal*, 5(1), pp.10–15.

Huglin, P., 1978. Nouveau mode d’évaluation des possibilites héliothermiques d'un milieu viticole. In *Symposium International sur l’Ecologie de la Vigne. Ministere de l'Agriculture et de l'Industrie Alimentaire*. Constança, Roumanie, pp. 89–98.

Hulkower, N., 2012. Comment on “A New Method to Obtain a Consensus Ranking of a Region’s Vintages' Quality.” *Journal of Wine Economics*, 7(2), pp.241–244.

IVDP, 2013. Vintages. *Webpage*. Available at: http://www.ivdp.pt/pagina.asp?codPag=98&codSeccao=2&idioma=0 [Accessed January 18, 2013].

John Moore et al., 2008. Wine biotechnology in South Africa: Towards a systems approach to wine science. *Biothecnology*, (3), pp.1355–1367.

Jones, G. V., 2012. *A climate assessment for the Douro wine region: an examination for the past, present and future conditions for wine production*, ADVID, Peso da Régua.

Jones, G. V. et al., 2005. Climate Change and Global Wine Quality. *Climatic Change*, 73(3), pp.319–343.

Jones, G. V., 2003a. Phenology: an integrative environmental science. In M. Schwartz, ed. *Tasks for Vegetation Science*. Springer Netherlands, pp. 523–539.

Jones, G. V., 2003b. Winegrape Phenology. In M. D. Schwartz, ed. *Phenology: An Integrative Environmental Science*. Kluwer Academic Publishers, pp. 523–539.

Jones, G. V., 2013. Winegrape phenology. In M. Schwartz, ed. *Phenology: An Integrative Environmental Science*. Springer Netherlands, p. 610.

Jones, G. V. & Alves, F., 2012. Impact of climate change on wine production: a global overview and regional assessment in the Douro Valley of Portugal. *International Journal of Global Warming*, 4(3/4), pp.383–406.

Jones, G. V. & Davis, R.E., 2000. Climate influences on grapevine phenology, grape composition, and wine production and quality for Bordeaux, France. *American Journal of Enology and Viticulture*, 51(3), pp.249–261.

Jones, G. V. & Goodrich, G., 2008. Influence of climate variability on wine regions in the western USA and on wine quality in the Napa Valley. *Climate Research*, 35(3), pp.241–254.

Jones, G. V. & Storchmann, K., 2001. Wine market prices and investment under uncertainty: an econometric model for Bordeaux Crus Classés. *Agricultural Economics*, 26(2), pp.115–133.

Kemeny, J., 1959. Mathematics without numbers. *Daedalus*, 88(4), pp.577–591.

Kemeny, J. & Snell, J.L., 1962. *Mathematical models in the Social Sciences* 1st ed., New York: Blaisdell Publishing Company - Ginn and Company.

Keuris, E., 2008. *Quality Assessments by consumers in the Wine Industry*. University of Groningen.

Kottek, M. et al., 2006. World map of the Koppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3), pp.259–263.

Landon, S. & Smith, C., 1998. Quality expectations, reputation, and price. *Southern Economic Journal*, 64(3), pp.628–647.

Lanzante, J., 1996. Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station. *International Journal of Climatology*, 16, pp.1197–1226.

Van Leeuwen, C. et al., 2008. Heat requirements for grapevine varieties is essential information to adapt plant material in a changing climate. In *7 Congrès International des terroirs viticoles*. Changins, Switzerland, pp. 222–227.

Leopold, A., 2013. The Aldo Leopold Foundation. *Webpage*. Available at: http://www.aldoleopold.org/programs/phenology.shtml [Accessed February 4, 2013].

Lin, S., 2010. Rank aggregation methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5), pp.555–570.

Lobell, D. et al., 2006. Impacts of future climate change on California perennial crop yields: Model projections with climate and crop uncertainties. *Agricultural and Forest Meteorology*, 141, pp.208–218.

Long, J.S., 1997. Regression Models for Categoirical and Limited Dependent Variables. In *Paper Series on Quantitative Applications in the Social Sciences, series no. 07, 7*. Thousand Oaks, CA, USA: Sage Publications, p. 297.

Long, J.S., 2012. *Regression models for nominal and ordinal outcomes*,

Lopes, J. et al., 2008. Exigências térmicas, duração e precocidade de estados fenológicos de castas da colecção ampelográfica nacional. *Ciência e Técnica Vitivinícola*, 23(1), pp.61–71.

Lorenz, H. et al., 1994. Phaenologische Entwicklungsstadien der Weinrebe (Vitis vinifera L. ssp. vinifera). Codierung und Beschreibung nach der erweiterten BBCH-Skala. *Viticulture and Enology Science*, 49, pp.66–70.

Lund, R. & Reeves, J., 2002. Detection of undocumented changepoints: A revision of the two-phase regression model. *Journal of Climate*, 15(17), pp.2547–2554.

Magalhães, N.P., 2003. Caracterização e Condução de Castas Tintas na Região Demarcada do Douro. In UTAD, ed. *Douro - Estudos e Documentos*. pp. 163–174.

Makra, L. et al., 2009. Wine Quantity and Quality Variations in Relation to Climatic Factors in the Tokaj (Hungary) Winegrowing Region. *American Journal of Enology and Viticulture*, 60(3), pp.312–321.

Mattis, N., 2011. *Analyzing Weather Patterns to Predict Wine Quality for Sonoma County Pinot Noir*. University of California.

Mayson, R., 2013. *Port and the Douro*, Oxford, UK: Infinite Ideas Limited.

Menard, S., 2001. Applied Logistic Regression Analysis, 2nd ed. In *Paper Series on Quantitative Applications in the Social Sciences, series no. 07-106*. Thousand Oaks, CA, USA: Sage Publications.

Monjardet, B., 1997. Concordance between two linear orders: The Spearman and Kendall coefficients revisited. *Journal of Classification*, 14(2), pp.269–295.

NASA, 2013. Climate Change: Evidence. *Nasa. Climate change: How do we know?* Available at: http://climate.nasa.gov/evidence [Accessed June 10, 2013].

Nemani, R. et al., 2001. Asymmetric warming over coastal California and its impact on the premium wine industry. *Climate Research*, 19, pp.25–34.

Neter, J. et al., 1996. *Applied linear statistical models* 5th ed., McGraw-Hill/Irwin.

O'Brien, R., 2007. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41, pp.637–690.

O'Connell, A.A., 2006. Logistic Regression Models for Ordinal Response Variables. In T. Liao, ed. *Paper Series on Quantitative Applications in the Social Sciences, series no. 07-146*. Thousand Oaks, CA, USA: Sage Publications.

OIV, 1983. *Code des caractères descriptifs des variétés et espèces de Vitis*, Paris.

Ojeda, H. et al., 2002. Influence of pre-and postveraison water deficit on synthesis and concentration of skin phenolic compounds during berry growth of Vitis vinifera cv. Shiraz. *American Journal of Enology and Viticulture*, 53(4), pp.261–267.

Parker, A.K. et al., 2011. General phenological model to characterise the timing of flowering and veraison of Vitis vinifera L. *Australian Journal of Grape and Wine Research*, 17(2), pp.206–216.

Parker, R., 2011. The Wine Advocate Vintage Guide 1996 - 2010. *eRobertparker.com*. Available at: https://www.erobertparker.com/newsearch/vintageChart1.aspx [Accessed December 2, 2011].

Parker, R., 2013. Wine Advocate. *Webpage*. Available at: http://www.erobertparker.com/newsearch/vintagechart1.aspx [Accessed January 18, 2013].

Primault, B., 1969. Le climat et la viticulture. *International Journal of Biometeorology*, 13(1), pp.7–24.

Queiroz, J. et al., 2008. Steep slope viticulture: training systems in narrow terraces, Touriga Nacional, Douro Region. In CERVIM.

R Core Team, 2005. *R: A language and environment for statistical computing*, Vienna, Austria: R Foundadtion for Statistical Computing.

Razali, N. & Wah, Y., 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lillefors and Anderson-Darling test. *Journal of Statistical Modeling and Analytics*, 2(1), pp.21–33.

Rebelo, J. et al., 2001. *Plano Intermunicipal de Ordenamento do Território Alto Douro Vinhateiro: Caracterização Sócio-Económica*, Vila Real.

Reek, T., Doty, S. & Owen, T., 1992. A deterministic approach to the validation of historical daily temperature and precipitation data from the cooperative network. *Bulletin of the American Meteorological Society*, 73(6), pp.753–762.

Reeves, J. et al., 2007. A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, 46, pp.900–915.

Risse, M., 2005. Why the count de Borda cannot beat the Marquis de Condorcet. *Social Choice and Welfare*, 25, pp.95–113.

Rolland, C., 2002. Spatial and seasonal variations of air temperature lapse rates in Alpine regions. *Journal of Climate*, 16.

Rudd, B.B.&, 2013. Vintage Chart. *Webpage*. Available at: http://www.bbr.com/vintage-chart.

Sadras, V., Soar, C. & Petrie, P., 2007. Quantification of time trends in vintage scores and their variability for major wine regions of Australia. *Australian Journal of Grape and Wine Research*, 13(2), pp.117–123.

Salazar-Gutierrez, M.R., Johnson, J. & Chaves-Cordoba, B., 2013. Relationship of base temperature to development of winter wheat. *International Journal of Plant Production*, 7(4), pp.741–762.

Santos, F. & Azevedo, J., 2004. Mecanização das Vinhas na Região Demarcada do Douro: Situação Actual e Contributos para o Futuro. *Boletim Informativo Agricultura Transmontana*, pp.10–11.

Santos, J. & Malheiro, A., 2011. Statistical modelling of grapevine yield in the Port Wine region under present and future climate conditions. *International Journal of Biometeorology*, 55(2), pp.119–131.

Schalekamp, F. & van Zuylen, A., 2009. Rank aggregation: Together we' re strong. *Proceedings of 11th ALENEX*, pp.38–51.

Schamel, G. & Anderson, K., 2003. Wine quality and varietal, regional and winery reputations: hedonic prices for Australia and New Zealand. *Economic Record*, 79(246), pp.357–369.

Schiefer, J. & Fischer, C., 2008. The gap between wine expert ratings and consumer preferences. *International Journal of Wine Business Research*, 20(4), pp.335–351.

Schultz, H.R. & Jones, G. V., 2010. Climate induced historic and future changes in viticulture. *Journal of Wine Research*, 21, pp.137–145.

Shirali, S. & Vasudeva, H.L., 2006. *Metric Spaces* Springer, ed.,

Solow, A., 1987. Testing for climate change: An application of the two-phase regression model. *Journal of Climate and Applied Meteorology*, 26(10), pp.1401–1405.

Spectator, W., 2011. Vintage chart. *winespectator.com*. Available at: http://www.winespectator.com/servefile/serve/id/41618 [Accessed April 1, 2011].

Spectator, W., 2013. Wine Ratings. *Webpage*. Available at: http://www.winespectator.com/vintagecharts/search.

Spence, G., 1997. *O guia do vinho do Porto*, livros e livros.

Spirits, V.F., 2011. Vintage chart. *vintages.com*. Available at: http://www.vintages.com/circular/vintage-chart.pdf [Accessed December 2, 2011].

Stevenson, T., 2011. *The Sotheby's wine encyclopedia* 4th ed., Dorling Kindersley.

Stone, P. & Carlson, J., 1979. Atmospheric lapse rate regimes and their parameterization. *Journal of the Atmospheric Sciences*, 36, pp.415–423.

Stooksbury, D., Idso, C. & Hubbard, K., 1999. The effects of data gaps on the calculated monthly mean maximum and minimum temperatures in the continental United States: A spatial and temporal study. *Journal of Climate*, 12, pp.1524–1533.

Tomasi, D. et al., 2011. Grapevine phenology and climate change: relationships and trends in the Veneto region of Italy for 1964–2009. *American Journal of Enology and Viticulture*, 62(3), pp.329–339.

Tonietto, J. & Carbonneau, A., 2004. A multicriteria climatic classification system for grape-growing regions worldwide. *Agricultural and Forest Meteorology*, 124(1-2), pp.81–97.

Tonietto, J. & Fialho, F., 2012. The Geoviticulture MCC System and its Internet Site. In J. Tonietto, V. Sotés Ruiz, & V. Gómez-Miguel, eds. *Clima, zonificación y tipicidad del vino en regiones vitivinícolas Iberoamericanas*. Madrid: CYTED, pp. 23–38.

Vicente-Serrano, S., Beguería, S. & López-Moreno, J., 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of Climate*, 23(7), pp.1696–1718.

Vincent, L., 1998. A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate*, 11(5), pp.1094–1104.

Vintages.com, 2013. Vintage Chart. *Webpage*. Available at: http://www.vintages.com/circular/vintage-chart.html [Accessed January 18, 2013].

Wallén, C., 1970. *Climates of northern and western Europe*, Elsevier Science Ltd.

Wang, X., 2008a. Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or F test. *Journal of Applied Meteorology and Climatology*, (47), pp.2423–2444.

Wang, X., 2003. Comments on "Detection of undocumented changepoints: A revision of the two-phase regression model." *Journal of Climate*, 16(20), pp.3383–3385.

Wang, X., 2008b. Penalized maximal F test for detecting undocumented mean shift without trend change. *Journal of Atmospheric and Oceanic Technology*, 25(3), pp.368–384.

Wang, X., 2011. RHtestsV3.

Wang, X., Chen, H. & Wu, Y., 2010. New techniques for the detection and adjustment of shifts in daily precipitation data series. *Journal of Applied Meteorology and Climatology*, 49, pp.2416–2436.

Wang, X., Wen, Q. & Wu, Y., 2007. Penalized maximal t test for detecting undocumented mean change in climate data series. *Journal of Applied Meteorology and Climatology*, 46(6), pp.916–931.

Wanhill, S., 1995. A test of Claret price formulation in auction markets. *International Journal of Wine Marketing*, 7(1), pp.17–22.

Webb, L.B., Whetton, P.H. & Barlow, E.W.R., 2007. Modelled impact of future climate change on the phenology of winegrapes in Australia. *Australian Journal of Grape and Wine Research*, 13(3), pp.165–175.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 48(4), pp.817–838.

Winkler, A. et al., 1974. *General viticulture* 4 th ed., Berkley: University of California Press.

World Meteorological Organization, 2011. *Guide to Climatological Practices (WMO-n°100)* Third edit. World Meteorological Organization, ed., Geneve, Switzerland.

You, J., Hubbard, K. & Goddard, S., 2008. Comparison of methods for spatially estimating station temperatures in a quality control system. *International Journal of Climatology*, 28, pp.777–787.

Young, H.P., 1988. Condorcet's theory of voting. *The American Political Science Review*, 82(4), pp.1231–1244.

Young, H.P. & Levenglick, A., 1978. A consistent extension of Condorcet's election principle. *SIAM Journal on Applied Mathematics*, 35, pp.285–300.

Zahumenský, I., 2004. Guidelines on quality control procedures for data from automatic weather stations. In *WMO Technical Conference: The Role of Instruments in the Earth Observation Systems*. Bucharest, Romania.