MASTER
DATA ANALYTICS FOR BUSINESS

MASTER'S FINAL WORK
DISSERTATION

ANALYSIS OF CLIMATE DATA IN PORTUGAL:
TENDENCIES AND ASSOCIATIONS WITH
AGRICULTURAL INSURANCE LOSSES

LUCIANA DA SILVA FLORA

MARCH -2022

MASTER

DATA ANALYTICS FOR BUSINESS

MASTER'S FINAL WORK

DISSERTATION

ANALYSIS OF CLIMATE DATA IN PORTUGAL:
TENDENCIES AND ASSOCIATIONS WITH
AGRICULTURAL INSURANCE LOSSES

LUCIANA DA SILVA FLORA

ORIENTATION:

ALEXANDRA BUGALHO DE MOURA
CARLOS MIGUEL DOS SANTOS OLIVEIRA

MARCH -2022

Acknowledgment

I would like to thank my supervisors Alexandra Bugalho Moura and Carlos Miguel dos Santos Oliveira for their guidance. They were always available to discuss my questions and to give me their suggestions that have helped in developing this dissertation. To professor Anna Couto for her advices and support.

I would like to thank IPMA, for making it possible to do this research, especially to meteorologist Maria de Lourdes Bugalho for making the data available and for giving her expert input into this work. I would also like to thank the IFAP, especially Cristina Malta, head of Market Support Department, for being available to answer all my requests.

To Raquel Pepe, my mentor at Mercedes-Benz Financial Services, for all the support and understanding that made it possible to develop this work.

The realization of this work would not be possible without the support of my family and friends for which I am thankful.

Abstract

Climate change and weather-related catastrophes have been putting pressure on nature-dependent sectors such as Agriculture. The insurance business is a support mechanism for these vulnerable activities. Thus, in this work we intend to study the relationship between losses in the agricultural sector, particularly the ones partially supported by insurance companies, by analysing climate and insurance data. Because of this relationship we believe that insurance companies contract definitions should be based on scientific evidence.

To correctly understand the climate data, provided by IPMA, it is necessary to treat the collected data. That was done in this work using the CLIMATOL software and by analysing standard quality checks that guarantee the goodness of our data. We used the treated data for trend analysis. Agriculture-Insurance related data was collected from the website IFAP, which contains a publicly available dataset that concerns information about the Crop Insurance variables and Governmental aid to farmers.

We will analyse to which extent the insurance companies and Government base their budgeting and policy definition on the scientific analysis of weather data. This was done by means of regression models and analysing the impact of each created variable for different groups of crops and regions.

For the treatment and manipulation of the data, it was used inhouse R code and PowerBI as the data visualization tool.

Keywords: Climate Data, Climatol, Crop Insurance, Agricultural losses, Trend Analysis, Regression Modeling

Resumo

As alterações climáticas e as catástrofes naturais têm vindo a pôr pressão sobre os setores dependentes da natureza, nomeadamente a agricultura. As seguradoras surgem como mecanismos de suporte para estas atividades mais vulneráveis. Consequentemente, neste trabalho, através da análise dos dados do clima e dos dados de seguros, pretendemos perceber a relação que existe entre as perdas no setor agrícola, em particular aquelas que são suportadas em parte pelas companhias de seguros. Esta relação que parece existir entre os setores leva-nos a crer que a definição de contratos de seguros deve ter uma base científica.

De forma a analisar corretamente os dados do clima, disponibilizados pelo IPMA, é necessário tratar os mesmos para que possam ser utilizados. Esse tratamento de dados foi feito neste trabalho através da utilização do software CLIMATOL e da análise de critérios de qualidade de forma a garantir a qualidade dos dados a utilizar. Após tratados, os dados foram utilizados para análise de tendências. Os dados relacionados com os seguros agrícolas foram obtidos através do website do IFAP, estando disponíveis publicamente. A base de dados utilizada contém informação sobre variáveis de seguros agrícolas e apoios estatais aos agricultores.

Foram analisados até que ponto é que a definição de orçamentos e de políticas bem como dos prémios de seguro são baseados na análise científica da evolução do clima. Para tal, utilizámos modelos de regressão que estudassem estas relações para diferentes regiões e conjuntos de culturas.

Para o tratamento e manipulação dos dados foram utilizados códigos de R e o PowerBI como ferramenta de visualização.

Palavras-chaves: Dados do Clima, Climatol, Seguros Agrícolas, Perdas Agrícolas, Análise de Tendências, Modelos de Regressão

Glossary

IFAP - Finance Institute of Agriculture and Fishery

IPMA - Portuguese Institute of Sea and Atmosphere

SC - Crop Insurance

SIPAC- Integrated Weather Protection System

SVC- Crop Viticulture Insurance

Content

1. Introduction

Climate Change is a reality intensely studied over the last years. Many changes have occurred over the world and the impacts differ for different regions, see [1]. Portugal is no exception to these more extreme phenomena, and it has been registering changes in temperature and in the frequency of occurrences of drought periods, as seen in [1]. This increase in the events and their intensity has been more predominant in the last thirty years, [2].

In this thesis, we aim to understand how the Portuguese latest climate evolution impacts the agriculture-insurance line of business, and its recent development. We also intend to study to which extent both the insurance sector and Government's support mechanisms are climate-driven. On one hand, the conclusions are that for the agriculture crop insurances, the tariffs are highly related to the weather variables as well as the frequency of hazards. On the other hand, the support given by the State to the farmers for crop insurance has not been aligned with the weather evolution. Indemnities paid to the farmers cannot be explained by our models.

As mentioned in [3], the effects of climate change can be direct or indirect. In Portugal, the common tendencies significantly affect the water resources available for industrial activities and day-to-day life. A substantial seasonal and year variability makes the country vulnerable to the extreme phenomenon associated with droughts, as exposed in [4]. At the same time several works, such as [2], [1], and [5] highlight that the last decades are the ones with higher average temperatures. Between 2004 and 2006 the droughts had the most prolonged duration, affecting 100% of the Portuguese mainland. In 2012, the severe drought situation led to the rise of government and public concerns related to climate change.

More recently, in November of 2021, a meteorological drought period began and it has been getting worse in 2022. The Portuguese mainland is considered to be in a drought situation, which is related to surface water and groundwater unavailability in accordance to what is explained in [5].

A study case on the Guadiana's River in the south of Portugal, [5], highlights the vulnerability of regions that depend on Agriculture because those s are more susceptible to severe changes and higher risks in the future. This phenomenon aggravates for

Portugal because of its western Mediterranean position, which is believed to be one of the regions to feel climate change impacts firstly and more intensely, see [5].

The role of insurances as a support service is of great importance in mitigating the impact of climate change in Agriculture, as mentioned in [6] and [7]. Scientific knowledge and methods based on evidence may help in predicting climate phenomena and making the insurance sector an important figure in future adaptations. Understanding how climate change has been affecting losses and prices of insurance companies, namely regarding agricultural lines of business, is of great importance in defining new strategies for the sector and in better supporting the more vulnerable industries.

In [6], it is mentioned that many factors contribute to the higher risks of this type of insurance, namely information asymmetries. The unpredictability of weather phenomena requires more skilled and expert underwriting. Weather and climate studies may be helpful to define future budgets for disaster payment, according to [8].

In order to relate the information from both agriculture-insurance data and climate data, there are some recommended steps. The phases go from data collection, to climate analysis with several in-between steps, [9]. Climate data still has many quality problems that are even more predominant when considering daily data. Several works, such as [9] and [10], highlight as the most important steps the homogenization, data quality checks, missing data infilling and metadata study. They all impact the goodness of results and were taken into consideration for this thesis. There are many software or packages that already include features that approach each of the listed topics. The choice of the method depends on several factors that characterize the dataset of each researcher.

In this work, we use climate data from IPMA, the Portuguese Institute of Sea and Atmosphere. This institute is responsible for collecting the data from the Portuguese meteorological stations. Different methods to treat and analyse the data are studied in order to get good quality from the Portuguese Climate stations' observations. The homogenization and quality control steps were performed using the CLIMATOL package from R. This package applies, under parametrization, a homogenization algorithm and quality control checks. Using the treated data, the latest climate developments and trends are analysed. The results of different types of meteorologic observation stations, namely manual and automatic, are extrapolated from one period to the other by analysing the parallel measurement periods.

Afterwards, the analysed Portuguese climate data is related with crop insurance aid data, which is publicly available. The data can be found in IFAP, Finance Institute of Agriculture and Fishery, website. The information refers to Government support mechanisms related with Insurance Underwriting Aid. The goal is to study how the frequency of hazards and its financial amounts impact the Insurance companies, the State, and the farmers, as well as how are those variables developing. Finally, the impact of climate variables in specific insurance variables is studied with the purpose of extrapolating their influence on the insurance sector. The final step is done by means of regression models.

This thesis is organized as follows. In Chapter 2, we start with the analysis of the raw data, as well as the application of the CLIMATOL package. Still in this chapter, the tendencies for climate data are studied along with the parallel measurements' comparison. In Chapter 3, the agriculture -insurance data is analysed and related with the climate data, using regressions models. The final chapter is dedicated to conclusions and future perspectives.

2. Portuguese Climate Data: ETL & Analysis

2.1 Portuguese climate raw data analysis

Good quality data is not possible without some preliminary steps. Homogenization, missing data infilling and metadata study is essential to guarantee robust results as mentioned in [8],[6], and [9]. When performing homogenization there are several methods and techniques we can consider. Several software have been tested in [11], and our choice was on CLIMATOL, an R package, that allows for climate data treatment with flexibility to deal with different weather variables, and a user-friendly interface as exposed in [12] and [13]. More details on the process of climate data treatment and the CLIMATOL algorithm can be found in Appendix A section.

The data used in this work was collected and provided by IPMA. The data comprises observations since 1941, although not all stations have such long series of observations. Since 1941, meteorological stations have suffered modifications. The main changes are related to location and equipment. The data collected for this work is denominated as manual or automatic due to the differences in equipment. Most of manual stations have information until 2011, and automatic stations series go from 1995 to 2018.

For the data to be analysed through visuals, it had to be transformed from its original format using R as the main tool. The selected variables to study were the minimum (Tmin) and maximum (Tmax) temperature, rainfall values for twenty-four hours (Rtotal), and the maximum wind speed achieved in one day (Wspeed). Throughout the thesis, those same variables are referred to as Tmin, Tmax, Rtotal, Wspeed.

For the exploratory and visualization analysis, PowerBi tool was used due to its capabilities of building dynamic graphs that allows for instant filtering and complex graphical visualization.

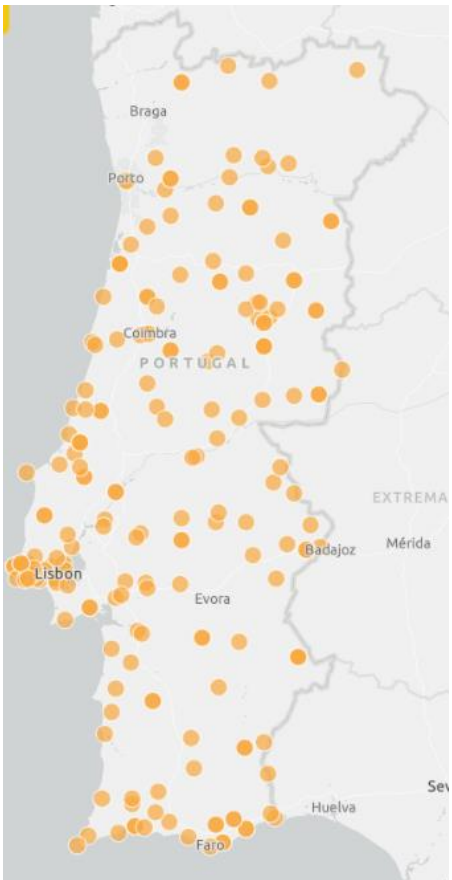

Figure 1. Distribution of the meteorological stations

We start by analysing the specificities of the raw data. The way the stations are dispersed throughout the country, the seasonal behaviour of the weather variables, and their limits are taken into account in this quality checks assessment. The data comprises observations of 90 automatic stations that cover a period of 29 years and 150 manual stations accounting for 70 years, from 1941 to 2011. In Figure 1, we can see their distribution over mainland Portugal. There exists a high density of stations in Lisbon. The Algarve region has stations along the coast. Interior and northern region of Portugal have a lower density of stations.

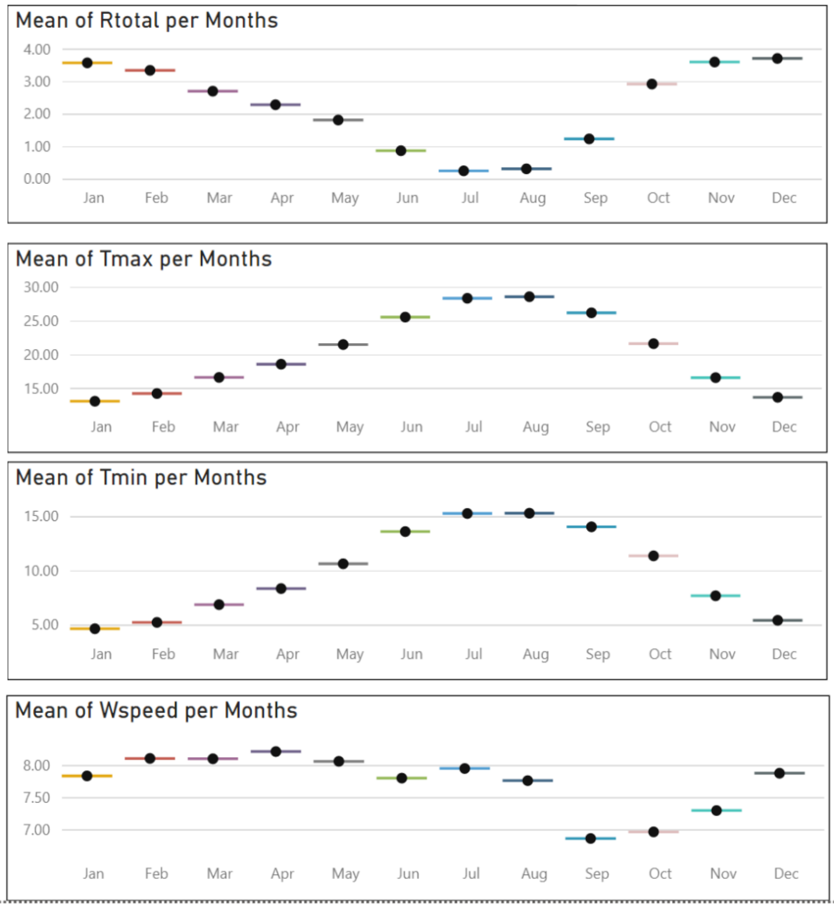

Figure 2. *Means of considered variables throughout the year-Tmax and Tmin (ºC), Rtotal (mm) and Wspeed (km/h)*

While looking at the data distribution throughout the year, in Figure 2, we can see that the data mirrors what is the expected behaviour of climate in Portugal. The winter and autumn months have lower temperatures (both maximum and minimum) and higher rainfall values. Also, the wind variable reaches higher velocities in the winter and autumn months, decreasing in the spring and summer seasons.

Another essential step performed on the raw data is the cleaning of incorrect data that cannot be used. Some data records had values that could not be accepted because they were outside logic and valid limits for the data. Both incorrect and missing data were found and may be due to misfunction of the station´s equipment. The logic limits are defined by IPMA, [14], that keeps track of each variable's extremes.

All the variables with values outside the acceptable intervals are considered unacceptable and deleted. Figure 3, which includes data of both manual and automatic stations, displays the percentages of acceptable data, and it is possible to see that the datasets are good overall.

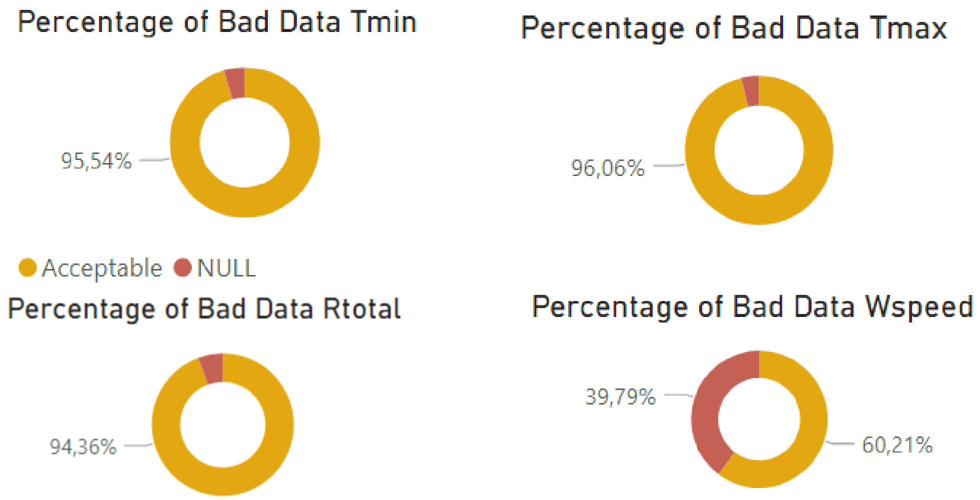

Figure 3. Percentage of correct and incorrect data for all variables

The only exception is the variable of wind speed that has the biggest proportion of incorrect data, with almost 40% of its values unsuitable to be used, while this percentage is no more than 6% for the other variables. Several studies, including [15], highlight that although temperatures and rainfall are highly studied in the scientific community, the windspeed-related variables are more difficult to analyse and understand and fewer studies exist for wind data. Indeed, reliable windspeed observations are difficult to obtain as they are affected by several factors, such as anemometer height changes or different sampling intervals, that lead to their inhomogeneities. Nonetheless, the study of changes in wind speed is also of great importance.

The tendencies of climate data should not be calculated using raw data, [9]. Nevertheless, to perform Temporal Validation, which analysis if the data in hands follows the previously described evolution or events, we analysed briefly the trend lines of the variables means, supplied by PowerBI. We observe that the maximum and minimum temperatures are increasing over the years, from 1941 to 2019, and the rainfall values are decreasing for the same period. The wind speed variable shows a significant decreasing slope, while the tendency slope of the other variables is more moderate. The variables align with what has been reported in other studies such as [2] and [1].

In the next section, we describe the homogenization process using CLIMATOL and after that we use the homogenized data to perform a tendency analysis in PowerBI.

2.2 Data treatment and homogenization using CLIMATOL

In order to use the homogenization tool of CLIMATOL, two input files are prepared, one with information about the data itself and another with information about the station codes and locations. The raw data is treated through an inhouse R code to automatically generate the input files for each variable and for the desired number of years.

Afterwards, following [12], an exploratory analysis on the data is performed to better parametrize the functions in CLIMATOL and adjust them to the variables. One of the output files, that results from the exploratory analysis function, is a report about the data that allows us to take conclusions and parametrize the necessary variables so that the homogenization process best fits our data (see [12] and [16]). This type of analysis was already done for other contexts, such as in [16]. Some parametrization tips can also be found in other papers and on the user guide for CLIMATOL [9]. The exploratory analysis is based mainly on three phases: (i) evaluation of the general range and distribution; (ii) analysis of data clusters; and (iii) analysis of the anomalies that are important for future parametrization.

Here we describe the process in detail, and more conclusions, for manual stations data that cover the period of 1941 up to 2011. For the remaining data, the exploratory analysis follows a similar procedure.

2.2.1 Range and distribution of the data

The first analysis made, with only 30 stations, showed inconsistencies, leading to the conclusions that there existed some quality issues due to the small number of stations available. As the number of stations used increased, the quality improved. Figure 4 displays the availability of the data. The white spots stand for the days with no available data; the stations are represented by specific numbers given by the software on the y-axis, and on the x-axis, time is represented in years.

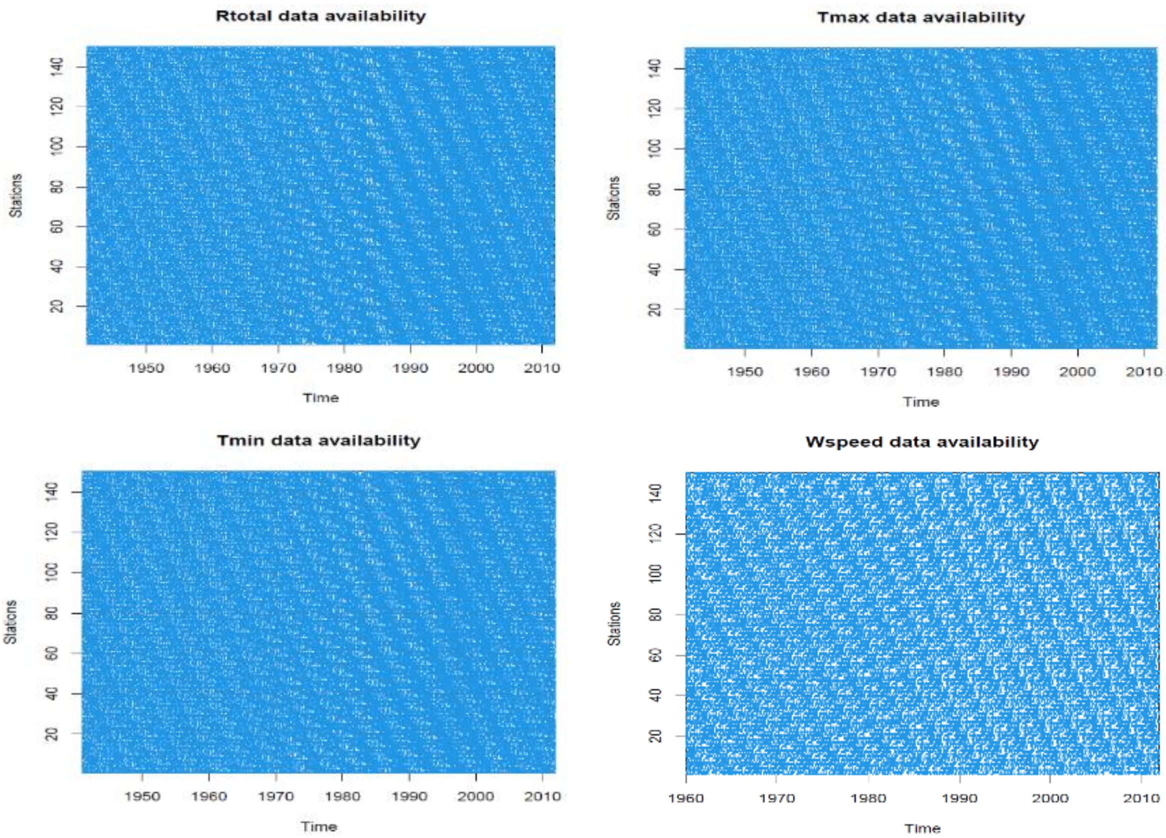

Figure 4. Data availability for each station and over the years for all variables

In terms of data availability, we verify that there are no available data for some steps, even considering the 150 manual stations. However, that availability should not be a problem for the quality of the process. As expected, the available wind speed data is lower than that of all the other variables.

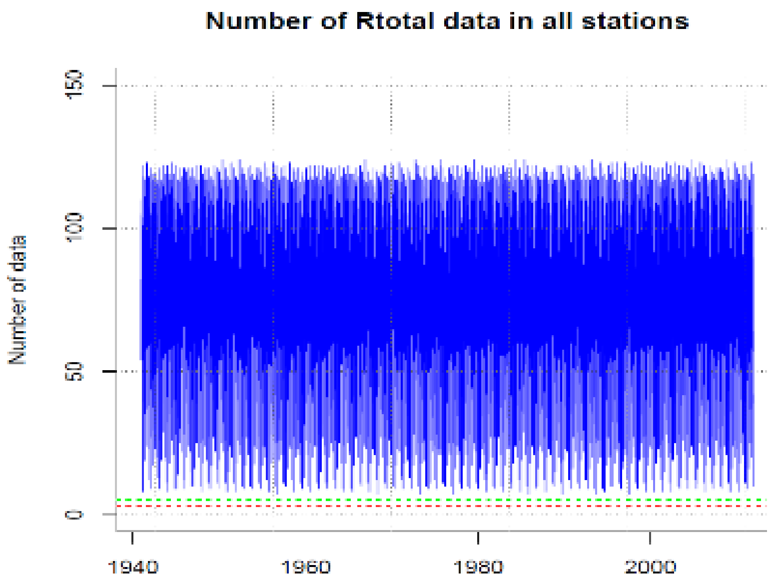

Figure 5. Data availability for each step over the years

Figure 5 shows the available data for each step. The dotted green line indicates the minimal number to have reliable homogenization processes. Only the Wind speed variable showed some steps with only two or one data available, which is explained by the difficulty in having wind data. As mentioned before, poor data quality and availability is common when considering wind variables.

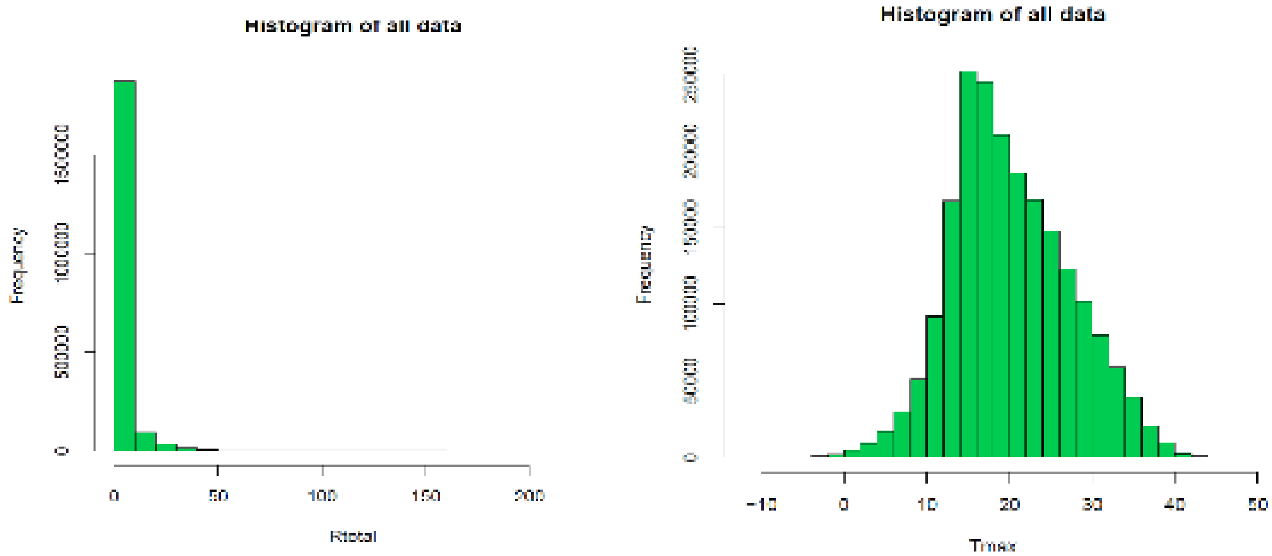

Figure 6. Data Histograms for the variables Rtotal (mm) and Tmax (ºC)

Looking at the data distribution histograms of the raw data, in Figure 6, it is possible to see that for maximum temperature the distribution is not centred and some values lean to the extremes. Similar graphs were generated for other studies, such as [16], and these variables were considered to follow a Gaussian distribution. This is also true for minimum temperatures. The situation is different for the wind speed and rainfall variables, which distributions have an “L” shape, best fit by Gamma distributions, which is characteristic of zero limited variables. These differences lead to different standard deviation parameters, as advised in [17].

2.2.2 Data correlation and clustering

Next, we analyse data similarity and clustering, starting with the correlograms of the data. The daily correlogram series, in Figure 7, show the correlation coefficient in terms of distance. It is expected that the further the stations, the lower their correlation is. The higher the correlation between stations, the more reliable the filling of the missing data is. In Figure 7, most of the correlations are around zero, and some are even negative.

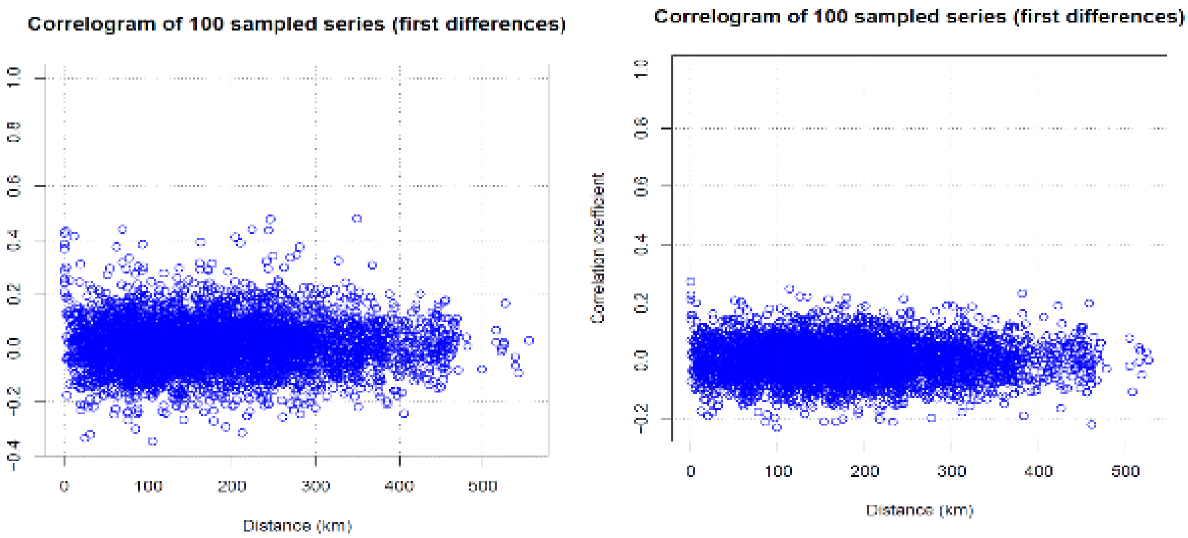

Figure 7. Data correlograms for the variables Tmin and Wspeed

This result is not the most encouraging, but it can be understandable if we consider the lack of station, namely in some regions, and how far they are from each other. Although 150 manual stations are being studied, the density of this network is not high, which may impact the final results, [18].

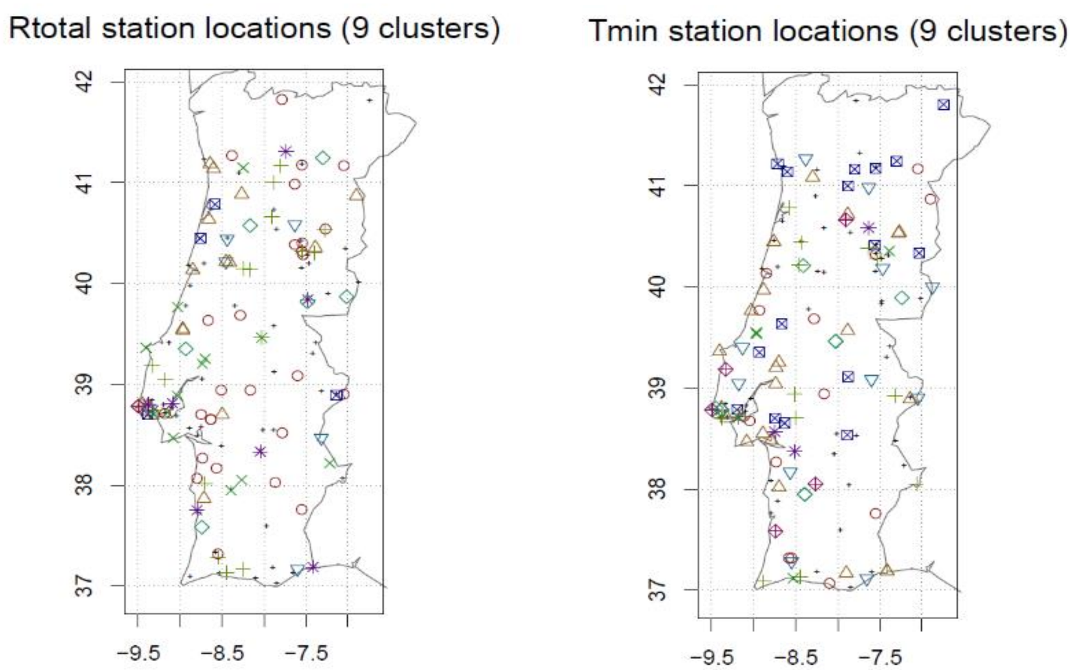

Figure 8. Map of Portugal with weather variable clusters, for the variables Rtotal and Tmin

Figure 8, displays the main variable clusters of rainfall values and minimum temperatures. We can see how the stations cluster together and how heterogeneous are the data profiles. The results are very similar for the other 2 variables. We can see, that the nine main clusters do not seem to follow any geographical rule. There are also stations of the same cluster in different regions, very dispersed. These results may explain the poor correlation between stations.

2.2.3 Data anomaly analysis

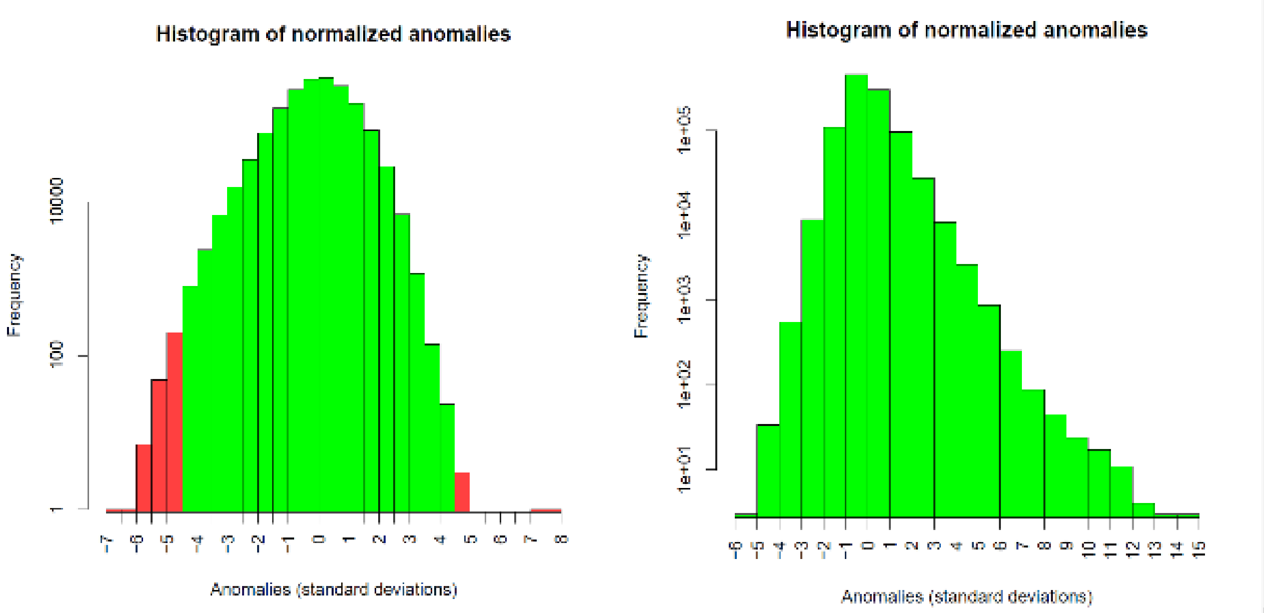

Figure 9. Histograms of normalized anomalies: (i) Tmax; (ii) Wspeed

The graphs about the distribution of the normalized anomalies, Figure 9, allow for a better parameterization of the homogenization function in CLIMATOL. These graphs allow us to decide what the threshold of anomaly acceptance should be. In this work, we gave as much range to the accepted anomalies as possible. We want to focus more on the extremes and outlier values than on the averages of the data.

The frequency distribution of the SNHT, Standard Normalised Homogeneity Test, for overlapping windows and for the overall series are more important than the frequency of the standardized anomalies.

The SNHT is a likelihood test performed on the ratios or differences between the data that will be calculated for and the reference series, as explained by [10]. These values can be parametrized and for this work a wide window was given so that more outlier values are accepted, including the extremes of the variables without compromising the quality of the homogenization and infilling process.

After the first an analysis it was possible to implement the CLIMATOL tool and the next section presents the quality evaluation of the results.

2.2.4 Data quality checks

The exploratory analysis is an interactive process that should be repeated until satisfactory results are obtained. Also, it should not be applied directly on daily data due to its high variances, as mentioned in [17]. The daily data was aggregated into monthly series that should be parametrized if necessary.

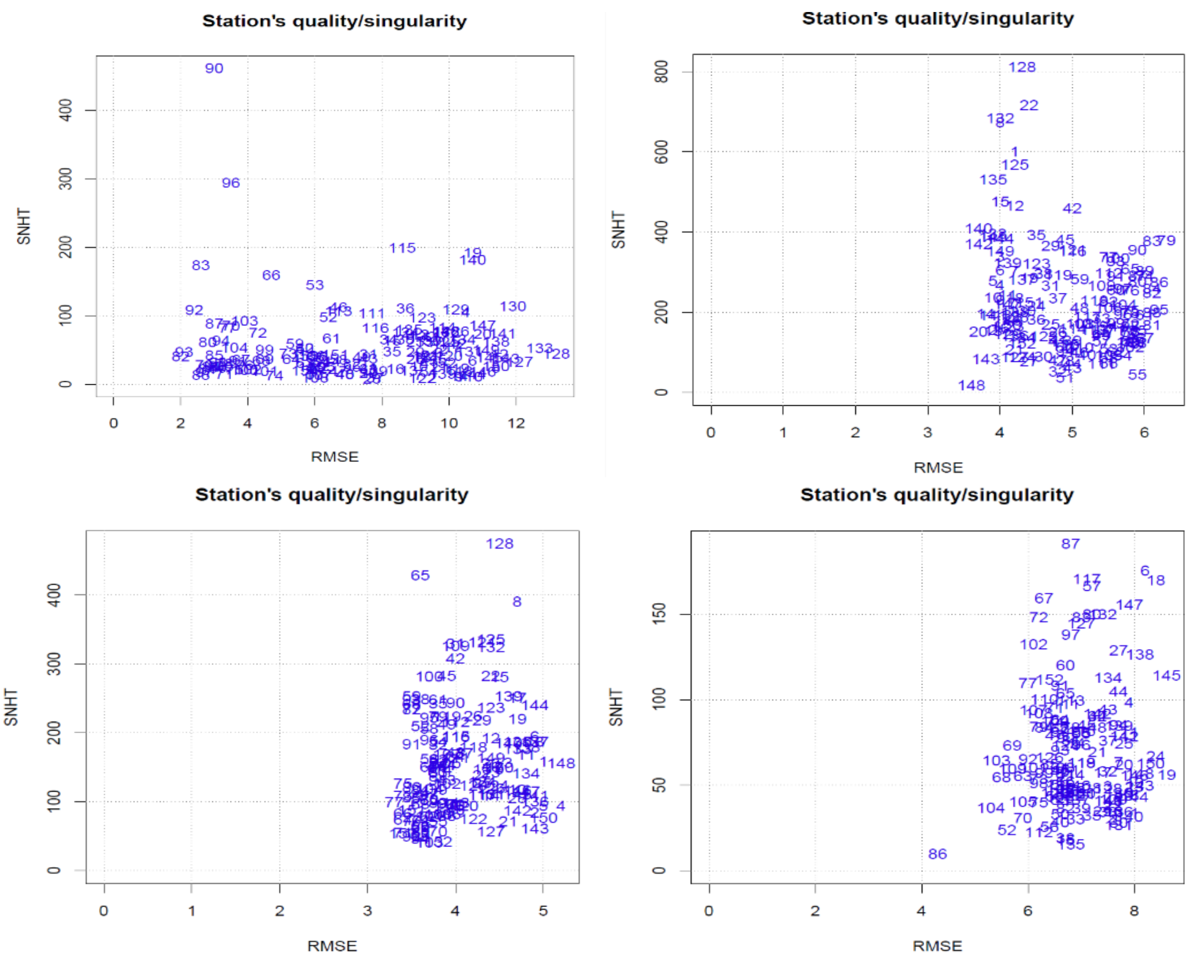

Figure 10. Quality and singularity plots-SNHT vs. RMSE values

The stations' quality and singularity graphs, in Figure 10, provide an idea of how well the stations' values were calculated and homogenized, i.e., bringing insight on the quality of the final results. The ideal situation is to have all stations in the left bottom corner, meaning a low SNHT and Root Mean Square Error (RMSE). The lack of correlation between our series can explain the higher values of RMSE. Nonetheless, the lower values for SNHT are acceptable as an indicator of the quality of the homogenization process.

Indeed, the SNHT relates with the difference of the calculated homogenized series to the actual homogenized values.

Although the final results are satisfactory other analyses on the treated data are made to assess if the data is good to use. Evaluating if the variables are within the valid limits and their seasonal the distribution allow us to access some basic validation rules, [19].

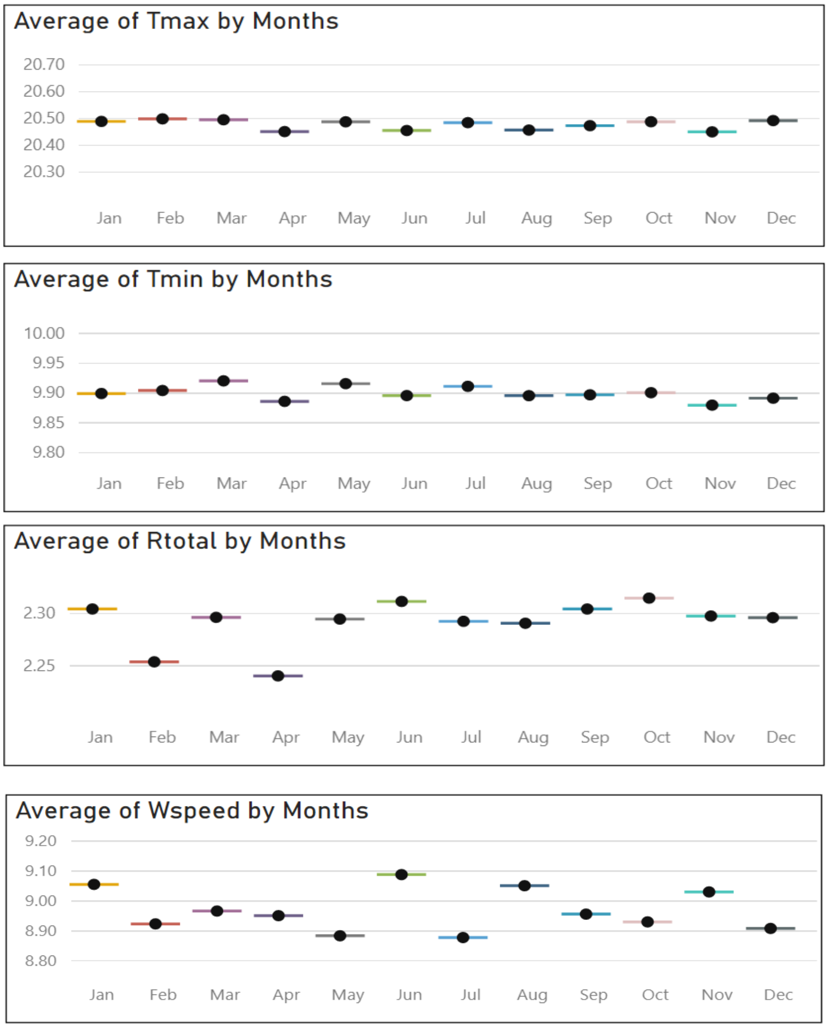

Figure 11. Evolution of the treated data for all variables by month- Tmin and Tmax (ºC), Rtotal (mm) and Wspeed (km/h)

From Figure 11 it can be seen that the distribution of climate data is no longer following what is the typical weather seasonality for the Portuguese climate, [4]. The homogenized data lost its seasonal characteristic. It is mentioned in [11] that CLIMATOL underestimates the seasonal Cycle amplitude in the adjusted data, and clearly this aspect must be taken into account for the research. The use of these results for trend analysis is still recommended and reliable. However, these distributions do not reflect the Portuguese climate characteristics on temperature and rainfall, making the data unsuitable for any seasonal analysis, and only suitable for the trend study (see [11]) as the variables are within acceptable limits defined by IPMA, [14].

It was analysed both manual and automatic station tendencies, and the main conclusions were that the treated data presents what is reported before concerning the Portuguese weather phenomena and their most recent years' evolution, mentioned by [2] and [5].

2.3 Comparing climate data of manual and automatic stations

In order to have high-quality observation data, these have to be made over a sufficiently large period, so to differentiate the patterns that relate to non-climatic factors from the ones that genuinely exploit the climatic evolution. Over the last years, for most of the Portuguese meteorological stations, there was a shift between older stations with manual instruments and newer stations with automatic instruments. The transition process comprised an overlap period for some stations, meaning that the two types, manual and automatic, performed in parallel. These parallel measurements allow for inference in the patterns of the differences between the two types of equipment (see [21]).

In [21] and [22], such comparisons are made for the German meteorologic stations network. It is studied the distribution of the differences in terms of frequency, the evolution of the differences and behaviour in terms of seasonality, as well as their mean and standard deviations. With this information, they are able to choose the best way to extrapolate the data from manual to automatic observation data, allowing to have long data series.

Hence, the parallel data for manual and automatic includes the period from 1941 to 2018. We have manual observations between 1941 to 2011 and automatic observations between 1995 to 2018. For the 240, including manual and automatic stations, around 60 correspond to the same locations or with minor differences in their location. Those stations are used for comparison. The differences are computed as the automatic stations' values minus manual stations' values.

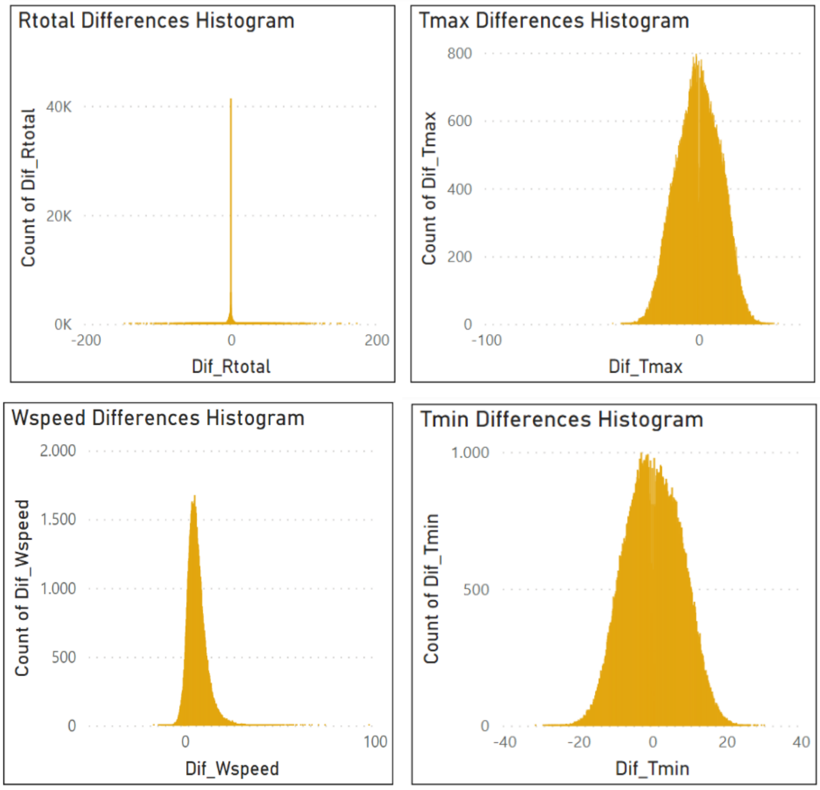

Figure 12. Variables differences between automatic and manual observations- Frequency distribution-

Tmin and Tmax (ºC), Rtotal (mm) and Wspeed (km/h)

The frequency distribution on the differences, in Figure 12, between stations is analysed first. Rainfall differences have the lowest values, with the histogram peek clearly at zero. For the minimum and maximum temperatures, the differences verify a normal distribution around zero. Nonetheless, there is a broader range of differences in the maximum and minimum temperature histograms. The differences concerning the windspeed follow a normal distribution with a slight skewness to the right. Here the differences are not centred in zero but in six units km/h.

Differences throughout the year between automatic and manual measurements are also studied. Figure 13 shows that rainfall values have lower values of these differences in summer months due to the lower rain intensity in those same periods for both stations. For the temperature the differences are higher in summer for both minimum and maximum temperatures because between the equipment used, one is more sensitive to higher temperatures than the other, reacting faster. Manual stations used mercury thermometers that take more time to react to the temperatures rising, contrary to the instant thermometer used in automatic stations. For the wind speed differences, there is no clear pattern across the year.

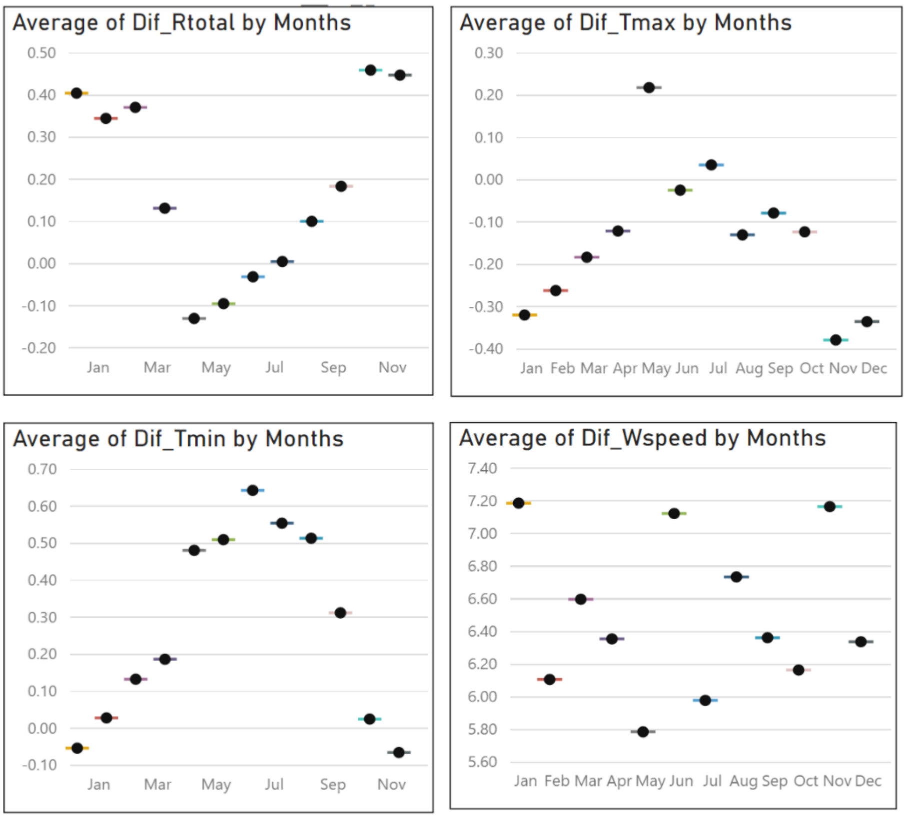

Figure 13. *Mean of differences between automatic and manual station observations per months*- Tmin and Tmax (ºC), Rtotal (mm) and Wspeed (km/h)

The average difference between automatic and manual station observations is zero for the rainfall variable, minimum and maximum temperature. For the windspeed variable, the average difference registered is around six kilometres per hour which should be taken into account when analysing the trends.

The studies in [21] and [22] showed very low standard deviation values for the differences between automatic and manual stations observation, and it was easy to extrapolate that the difference between equipment was 0 for most cases. For the Portuguese data at hand, the standard deviation of these differences is 7.84 ml/24 hours, 10.14ºC, 7.48ºC, and 5.14km/h for rainfall, maximum, minimum temperature, and wind speed, respectively. The differences of geographic and meteorologic characteristics between the two networks, such as higher thermal amplitudes in Germany in contrast with Portugal, for example, could be in the origin of these diverse outcomes.

The behaviour of the differences between automatic and manual time series indicate that we can study the time series from automatic stations as a continuation of the time series from the manual station's observations. The higher differences in summer, for temperature values, are due to the different thermometers used, leading to more step trend lines. Regarding standard deviations of the differences, the values are significant and emphasize the extensive range of differences, especially concerning the maximum temperature.

2.4 Trend analysis of the homogenized climate data

PowerBI provides a visual trend line that consists of a linear regression using time as the independent variable, and the user cannot intervene. Thus, a dynamic trend line is calculated so that we could have access to the slope values. A dynamic trend line is a linear equation that dynamically uses the time variable (adjusting the data to the periods such as year, quarter, month, etc.). More information about the approach of dynamic trend lines can be found in [19] and [20].

The average mean values over the years for wind and temperature are considered. For the rainfall values, the yearly sum is considered instead. We started by looking at the tendencies of minimal and maximum temperatures, rainfall values in 24 hours, and maximum wind speed values. As the research evolved other variables such as thermal amplitude, maximum of maximum temperature and the minimum of minimum temperature were considered. To calculate the trends of the thermal amplitude we looked at the mean values per year. For the maximum and minimum for both maximum and minimum temperatures we calculated each extreme for each station and looked at the mean per year considering all the stations. A linear equation is used to determine the trendline, an approach used in other papers such as [23]:

$$(1) \qquad Y = mX + b.$$

where the m represents the speed at which the variable in study is growing, [23].

In Figure 14, we can see trends' behaviour of the mean variables or sum (for the rainfall variable). On a closer look at the trends of the manual stations, it can be seen that the slopes are minimal, around 0 for most variables. Nevertheless, the slight tendencies seem to be negative for the minimum, maximum temperature, and the rainfall values. Looking at the whole period from 1941 to 2011, the growth rate for both minimum and maximum temperatures are - 0.001ºC per year. For the rainfall values, the trend is also very small since we are looking at the sum's values, corresponding to -91.33 millilitres per year. Regarding the windspeed, the tendency is almost -0.002 km/h, per year.

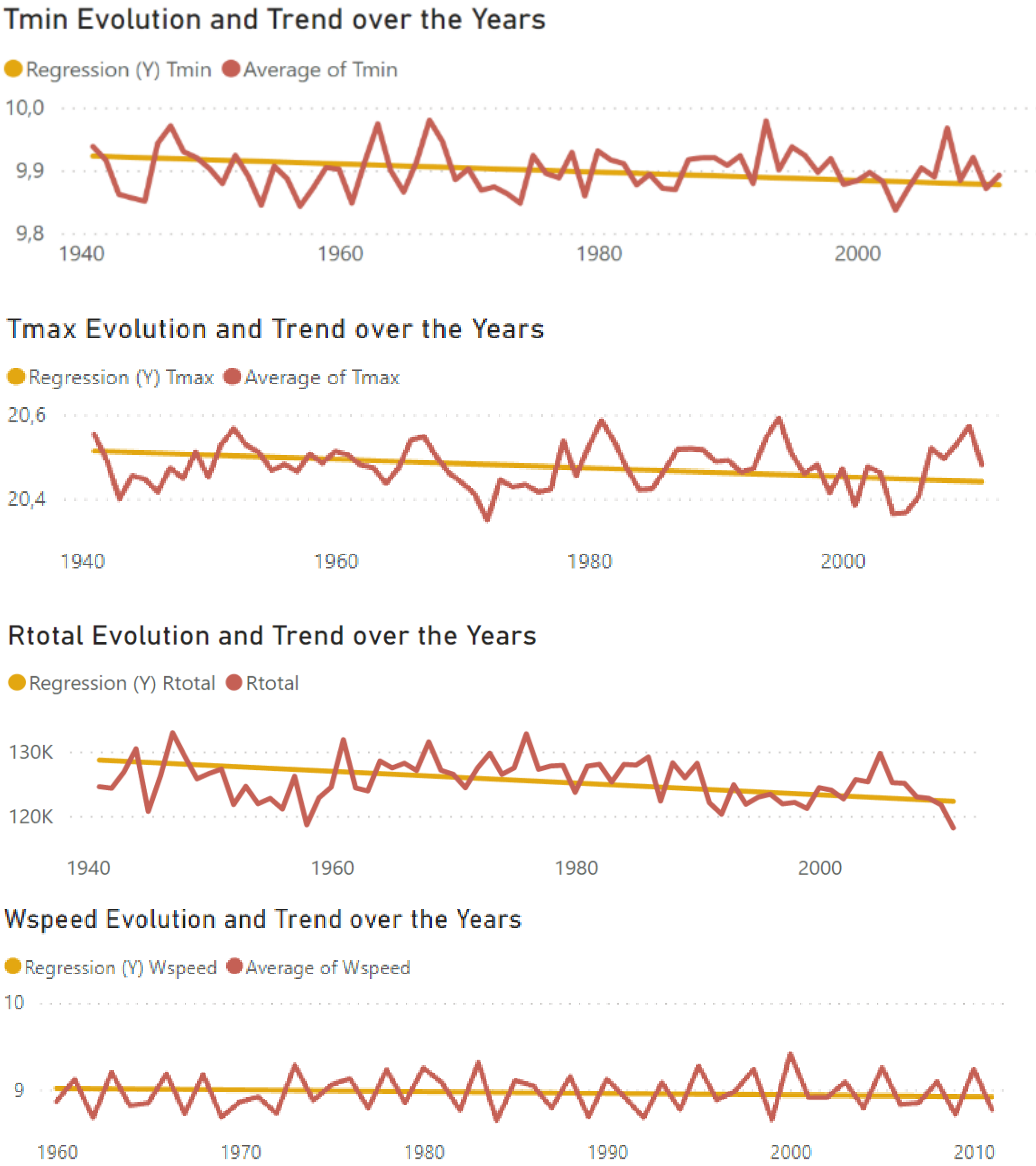

Figure 14. *Trend and Variables Evolution-manual stations, 1941 to 2011*- Mean of Tmin and Tmax (ºC), Sum of Rtotal (mm) and Mean Wspeed (km/h)

As we decrease the range of years (from 1985 to 2011), and start analysing more recent intervals, the tendencies shift from slightly negative to slightly positive for both minimum and maximum temperatures. For rainfall values, the negative slope maintains, but its value changes to -375.08 millilitres per year.

Regarding automatic stations in Figure 15, we can see that in the whole period between 1995 and 2018, the extreme temperatures, both maximum and minimum, present a negative slope, indicating that their growth rate is negative. The maximum temperature values show a decrease of 0.003ºC on average, and the minimum temperatures show a decrease of 0.001ºC. On the other hand, the rainfall tendencies increase 83 millilitres per year for the entire country, again representing minimal differences. For the wind speed variable, the trend of evolution is negative.

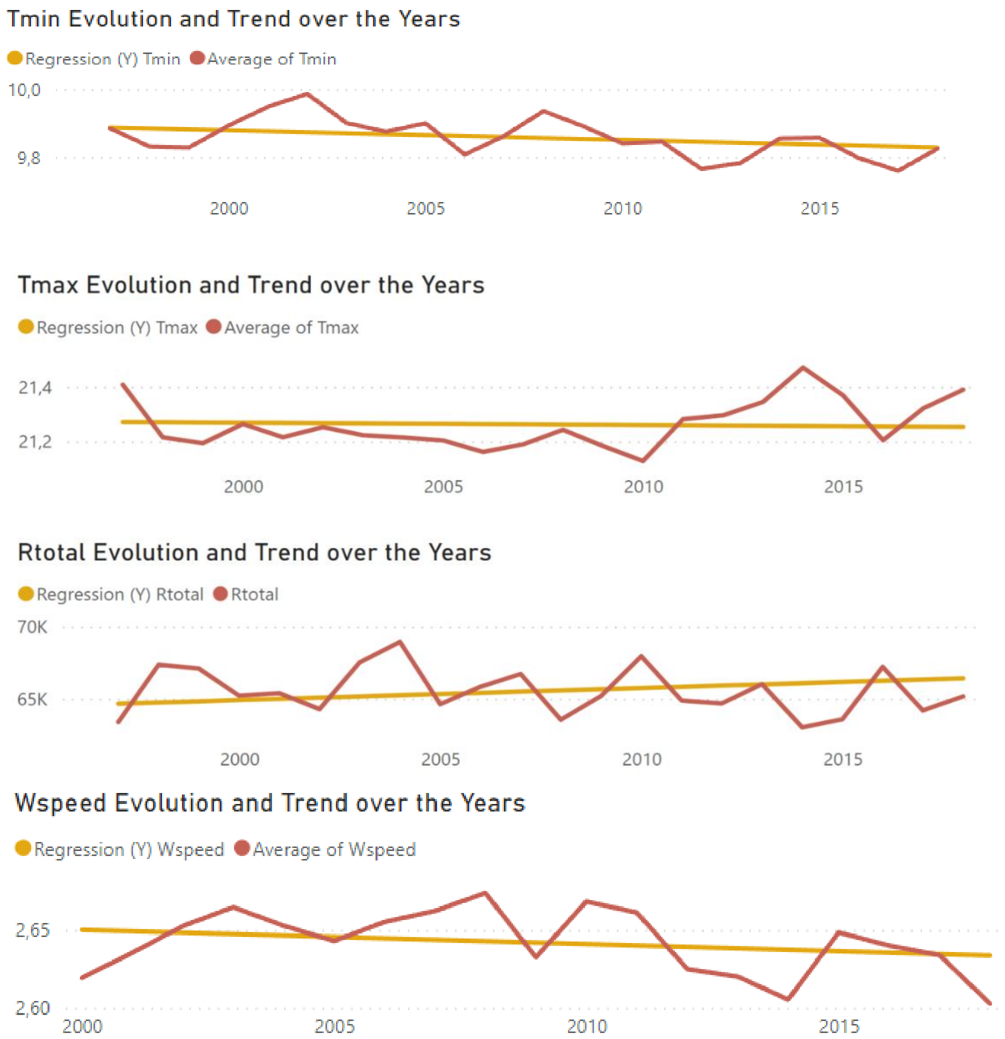

Figure 15. Trend and Variables Evolution-automatic stations, 1995 to 2018-- Mean of Tmin and Tmax (ºC), Sum of Rtotal (mm) and Mean Wspeed (km/h)

When shortening the analysis period to the last ten years of observations,2010 to 2018, we see that the variables have different behaviours. For the maximum temperature, for example, since 2010, the homogenized data is above the trend line, and for the rainfall, the sum line is almost always under the trend line. The trends have changed over the last ten years, with the slopes reaching 0.033ºC for the maximum temperature, meaning, it is expected that the average maximum temperature increases around 0.033ºC per year. For the rainfall the tendency was of -349.788ml per year. As we decrease the range of the observations, it is possible to see an intensification of those signals. For the period between 2010 and 2018, the behaviour of the minimum temperatures continues to show a decreasing trend that is now steeper with a tendency of decrease of -0.02ºC. The wind speed trend line shows a slight decrease for this period. These results are within what have been the latest developments in Portugal in terms of maximum temperature and precipitation values, which are described in [1], [2], and [3].

Next, we look at other variables such as thermal amplitude, yearly maximum of daily maximum temperature, and yearly minimum of daily minimum temperatures. Looking at the thermal amplitudes in a day, or the extremes that happen for each station in a year could give more insight into the latest climate evolution trends.

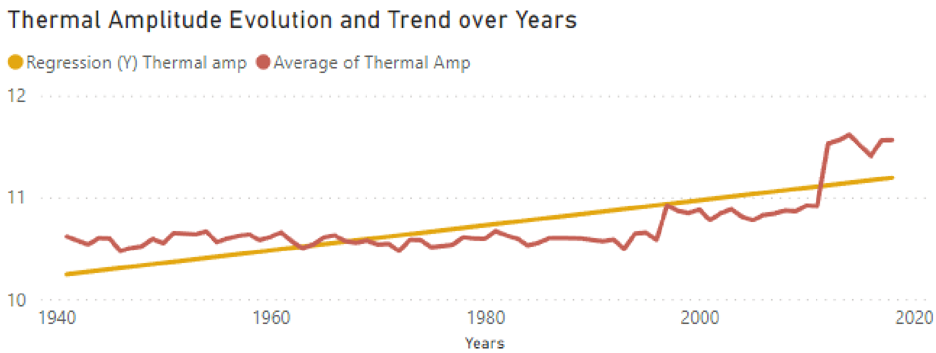

Figure 16. Trend and Variables Evolution for Thermal Amplitude-manual and automatic stations – values in ºC

In Figure 16, it is shown the evolution of the thermal amplitude yearly mean values for the last 60 years. It can be seen that the trend is positive, with a growth rate of 0.01ºC per year. As before, we see significative changes in the behaviour of the observed variables (red line) in the last ten years of observations, with a sudden increase, in 2010, which corresponds to the same year where the maximum temperatures started having higher positive trends. From 2000 to 2018, the trend line slope is about 0.04ºC, and between 2010 and 2018, it is around 0.08ºC. These trends represent an intensification of maximum temperatures, with a contrary in the minimum temperatures.

When observing similar graphics and information for the yearly maximum and minimum for daily maximum temperature and minimum temperatures, respectively, we verify zero tendencies when considering the period of 1941 up to 2011. For the period between 2000 to 2018, the only change is seen on the growth rate of maximum temperature that is -0.02ºC. As before, it was possible to see that after 2010 the behaviour of the variables changed, but those changes do not reflect significant variations in the tendencies.

We may assume that the soft changes in the weather variables in the first observation years balance the last few years. It is noticeable that the weather variables are changing more and faster than before if considering the latest decade. Such intensification may have several impacts on several fields of life. When comparing the mean yearly evolutions with the yearly maximum and minimum evolution of the temperatures, we see that the change is more significant for mean values than the extremes. The thermal amplitude behaviour reflects the yearly mean of the extreme temperature behaviours, and its results come from faster growth in the yearly mean maximum temperature than in the yearly mean of the minimum temperatures.

3. Insurance data: agriculture lines of business

3.1 Insurance data processing and metadata

In this chapter, we explore how the yearly agriculture insurance information related with indemnities, public aid, tariffs paid, etc., evolved over the years and in which extent the weather variables can explain the last years evolution. In this study, publicly available data from the Government insurance aid to farmers is used. The data is available on the IFAP (Financial Institute of Agriculture and Fishery) website, [24]. The data is not as detailed as desired, nevertheless it allows for some analysis and conclusions.

The Government provides public aid so that the farmers can have insurances that cover their crops. The Government aid is given in the form of financial support by paying part of the insurance premium. This public help can be divided into three segments: Crop Insurance (SC), Crop Viticulture Insurance (SVC) and Integrated Weather Protection System (SIPAC). The first public aid to farmers was defined by SIPAC, which covered all types of crops and it was applied since 1996 to 2013. From 2012 on, the agreements where split into Viticulture Insurance aid and Crop Insurance aid which includes all the crops except viticulture.

The public aid covers insurances related with: (i) adverse climatic phenomena similar to natural catastrophes that destroys 30% or more of the production; (ii) adverse climatic phenomena that are not natural catastrophes; (iii) plagues and diseases that are caused by natural factors and that cannot be controlled by agricultural techniques. All this information can be found in the contracts for each of the agreements and on the several updates they suffered throughout the years, the information is in IFAP website, [25].

This data has information about the tariff's geographical region. In Table 1 of the Appendix B, is possible to see how the counties were aggregated by region. The division presented is not the original one because some counties were accounted for in two regions. In order to keep track of the real evolution in the counties and to not duplicate results, every time a county is associated with two regions it is taken out of the region with the bigger number of counties. For an easier understanding and analysis, the main segmentation used for this work is the type of crops, divided into vineyards and others, and segmentation by region.

The public information is organized into two different categories: the insurance contracts and the losses. Out of the variables is the Insured *Capital,* which is the value of the product of the production, the production in quantities, times the market price:

$$(2) \qquad Insured\ Capital = Production\ \times Market\ Price.$$

It represents how much of the good is being insured and how much it is worth. Another variable is the *Commercial Prize,* which is the value of *Insured Capital* times the Tariff.

$$(3) \qquad Insured\ Capital\ \times Tariff = (Production\ \times Market\ Price) \times Tariff.$$

The Tariff is the insurance premium applied by the insurance company, per euro of Insured Capital. The Commercial Prize value does not include taxes and it is solely based on the goods of the farmer and the price of the insurance service. Another variable is the *Bonus*. From the total amount that has to be paid to the Insurance Company, the State contributes directly paying to insurance companies, taking some responsibility from the farmers. The calculation of this value depends on the aid percentage that is applied to each case. Every program has specific application conditions. The tariff used is either the national reference tariff or the insurance company tariff in case this last one is smaller.

$$(4) \qquad Bonus = Bonus\ \% \times Commercial\ Prize$$

$$= Bonus\ \% \ \times (Tariff \times Insured\ Capital).$$

With the bonus, the amount that a farmer has to pay can be summed up in the following equation:

$$(5) \qquad Farmers\ Payment = Commercial\ Prize - (Bonus).$$

On the second set of data, there is information on the Losses. For these files the variables are *Indemnities*, which are paid values by the insurance company to the insured farmer; the *Refund of Indemnities*, which are the returned values by the insured farmer to the insurance company. The *Expenses*, that correspond to operational expenses of the insurance company; The *Refund of Expenses*, corresponding to amounts paid back to the insurance company from operational costs. All the variables' units are euros. The reasons for the refunds of indemnities and expenses were not explained in the metadata. Each record has information on the cause of the hazard and all this metadata was collected with the help and support of professionals that work at IFAP.

Besides the variables described before, the datasets are also divided into *Crop* that is covered by the insurance, *Charging Region*, consisting of five regions classified from A to E, see Table 1 of Appendix B, and the *County,* that takes into consideration the administrative division of the Portuguese national territory. The regions were also segmented by district and NUTS (Territorial Units for Statistical Purposes).

For this data to be used and analysed in Microsoft PowerBI, every file is treated, translated and merged to build the entire dataset with losses and underwriting Information.

3.2 Analysis of the insurance data

3.2.1 Data on losses

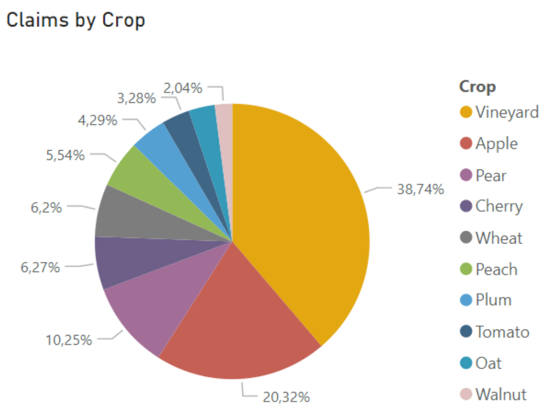

Figure 17. Percentage of claims by crop type, top 10 crops- All agreements

In this section, we evaluate the composition of the claims in terms of causes and crops in order to understand the latest years' evolution.

As shown in Figure 17, the top ten crops are displayed and account for more than 75% of the number of incidents registered between 1996 and 2019. Vineyards is the most affected culture with 38.74%, followed by apple, pear, cherry, wheat, peach, and plum culture. Looking at the causes of the incidents, Figure 18 shows that more than 70% is due to hail and frost, followed by fires (9%) and very heavy rainfall (8.2%).

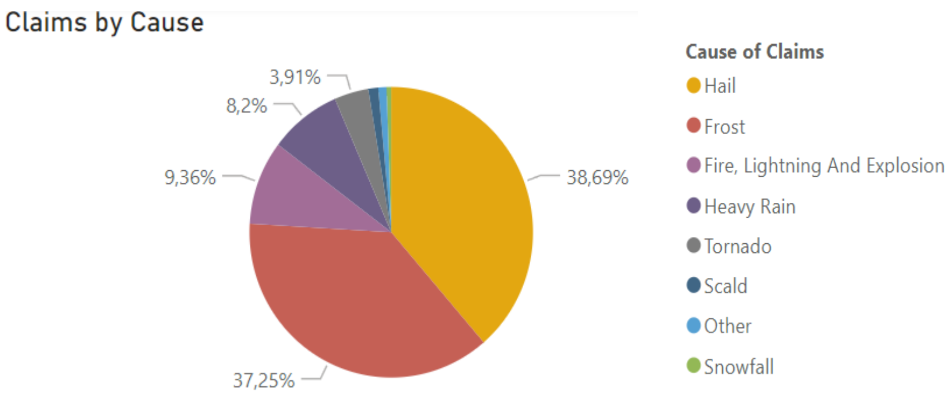

Figure 18. Percentage of claims by cause-All agreements

Most of the hazards were registered at the end of 1990. These high values, shown in Figure 19, are not the result of more incidents but rather the result of higher participation of farmers applying to the insurance public aid. The data available includes only information on insurance acquired by the farmers that applied to the public help. So, it does not always represent the actual number of hazards that affected agriculture in the country. The fact that the Government aid was very high in 1996-1999, around 85% including fiscal expenses, explains the abnormally high claim values for these years. With the public bonus decrease in the following years, as well as other alterations to the Calamity Fund, which controls the public funds that go to crop insurance aid, the farmers' participation levels decreased, and, with that, the number of registered claims. In 2000, the second-highest number of incidents was registered, not because of the contract conditions, but due to weather conditions which generated many frost claims. Since 2000 the number of registered hazards decreased until 2005, remaining roughly stable until 2019. There is no available data in 2018.

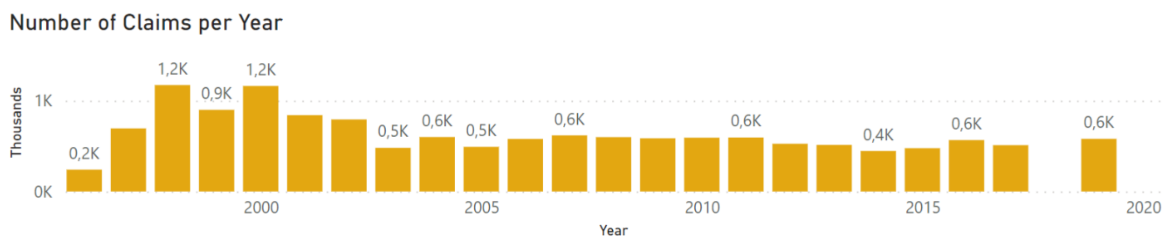

Figure 19. Number of Claims per Year- All agreements

A similar analysis can be done for the expenses and indemnities. Figure 20 illustrates the evolution of indemnities and expenses, that the insurance companies have. The bonus variable gives us the State's amount paid to the insurance company supporting the farmers with these costs. As explained before, the more significant amounts at the end of the 1990's decade are due to the greater participation of farmers applying for the aid. Interestingly, the bonus is higher than indemnities at the end of the decade and the beginning of the 2000s. However, between 2010 and 2019, this relationship changes. We can see periods where the indemnities are higher than the bonuses, which indicates that, for some periods, the insurance companies are spending more money than the State.

To better understand the impact that the weather claims have for both private and government entities, the ratio between total expenses of insurance companies, including indemnities and expenses, and bonus is calculated:

$$(6) \quad \frac{Total\ Costs}{Bonus} = \frac{Indmnities+Expenses}{Bonus}.$$

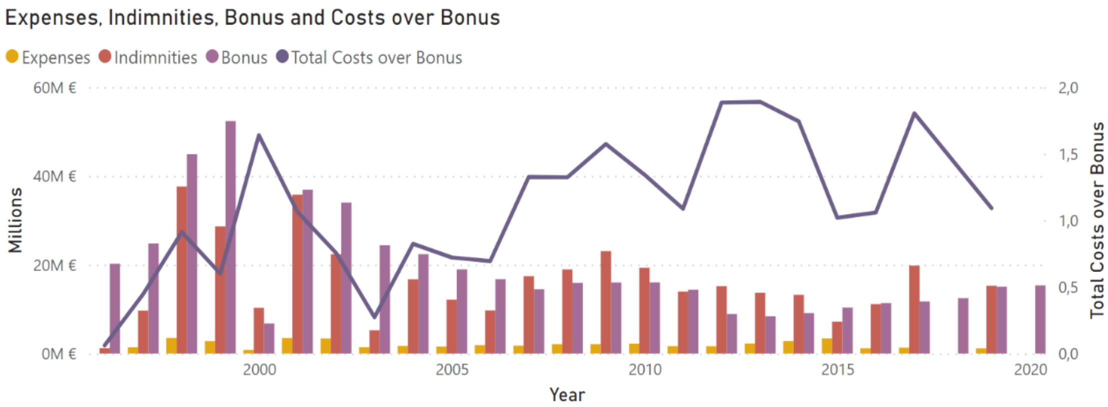

Figure 20. Evolution of expenses (€) , indemnities (€) , bonus (€) and total costs over bonus - All agreements

In Figure 20, we can see that at the same time the value of indemnities, red columns, decreases, the public aid, purple columns, presents a similar behaviour on different scales. This behaviour makes the amount spent by the insurance companies to be, sometimes, the double of public aid. In 2013 and 2017, the insurance companies paid almost twice of what was paid by the State.

The higher amounts paid by the insurance companies, compared with the State, although not linear, seem to display an increasing trend. The analysis of the expenses is similar to the indemnities, and there is no increasing trend in the last few years.

To better understand the evolution of the variables, an analysis by crop, cause of incident, and region is made. We split the time evolution in two, from 1996 to 2012, where SIPAC was the Insurance agreement in force, and from 2012 up to 2019, where SVC and SC were the insurance agreements applied.

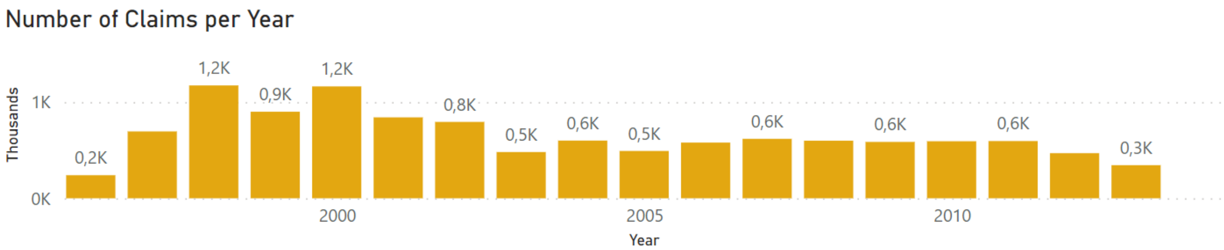

Figure 21. Number of Claims per Year- SIPAC Agreement

Looking only at the SIPAC agreement, the number of claims are decreasing from 1996 to 2013, Figure 21. In 2000, 1163 incidents were registered, accounting to 10 million euros of indemnities. In 2012, 469 incidents were registered, with the correspondent value of 14 million euros of indemnities. Although the number of claims decreased, the value paid increased about 4 million euros, which can indicate less incidents but worse weather phenomena that leads to more losses. The ratio of costs over the bonus, Figure 22, started to increase in 2003, showing that the State's expenses were less than the expenses from the insurance companies.

For SIPAC, vineyards, apple, pear, cherry, wheat, peach, and plum culture, make up for more than 70% of the total number of incidents, and hail and frost being the number one cause of all incidents.

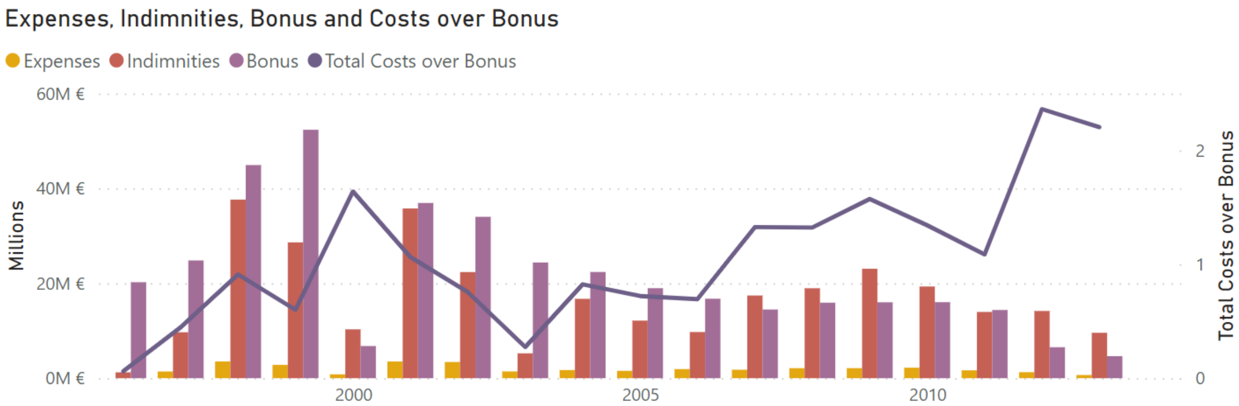

Figure 22. Evolution of expenses (€) , indemnities (€), bonus (€) and total costs over bonus – SIPAC

As displayed in Figure 23**,** for the SVC, vineyards insurance agreement, we can see an increase from 2012 up to 2019. The occurrences went from 57 in 2012 to 179. The worst year in terms of the number of claims and indemnities was in 2017, with 218 occurrences and 7 million euros spent by the insurance companies.

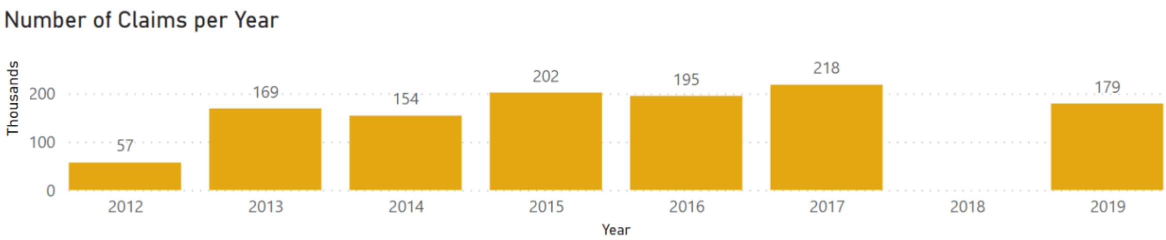

Figure 23. Number of claims per year- SVC Agreement

For this specific crop, hail and frost are the main cause of claim, but also scald and tornados are significant among the causes of losses. The sensitivity of Vineyard crop to hail and frost is mentioned in [26]. The most affected parts of the country are the north, mainly the inner countryside, mostly regions D and E. We have a small increase for the SVC insurance indemnities with 57 claims corresponded to 1 million euros of indemnities in 2012, and, in 2019, the incidents account to 2 million euros of indemnities.

Looking at the indemnities and bonus, in Figure 24, it can be seen that the bonus is always higher than indemnities, and so, for Vineyards, the public aid tends to be greater than the expenses the insurance companies have. The value of the ratio total-bonus reached its peak in 2017 with 2.4, meaning that for each euro spent by the State, the insurance companies spent 2.4 euros.

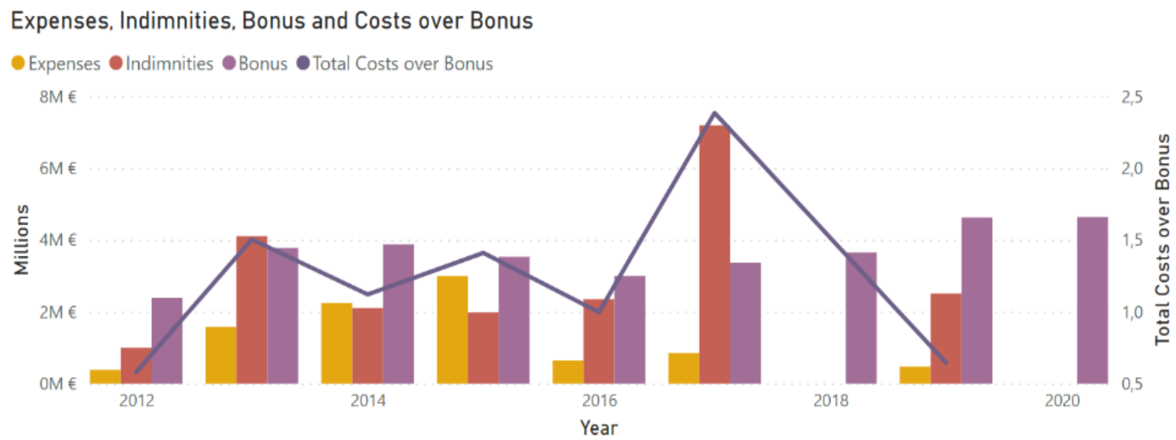

Figure 24. Evolution of expenses (€), indemnities (€), bonus (€) and total costs over bonus – SVC

Conducting a similar analysis for the SC, insurance that includes the remaining crops, we can see a slightly positive trend in the number of claims. Here, the relationship between the amount spent by insurance companies and the State is more similar, with two years, 2014 and 2017, registering more money paid by insurance companies than from the State. From 2014 to 2019, there was a slight increase in occurrences, from 292 to 403, respectively. The indemnities also have increased slightly, going from 11 million euros to 12 million euros in 2019.

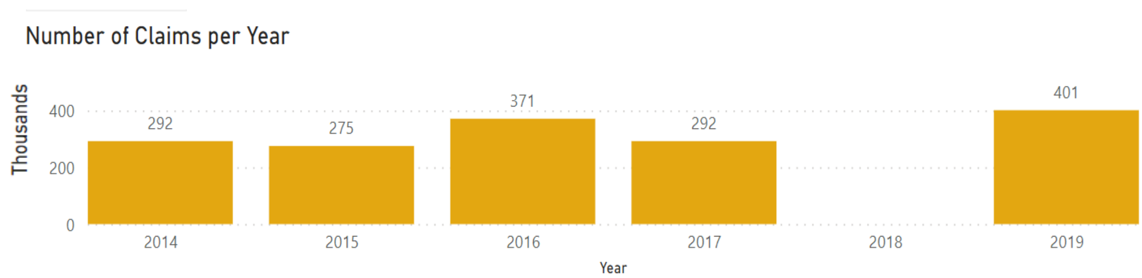

Figure 25. Number of Claims per Year- SC Agreement

For SC insurance, the observations only start from 2014 and go until 2019. For this insurance, which does not include vineyards, the main affected crops are apple, pear, cherry, and peach, making more than 50% of the whole affected crops. In terms of claim causes frost and hail are, again, the primary cause. The region where most of the claims happen is predominantly the northern countryside.

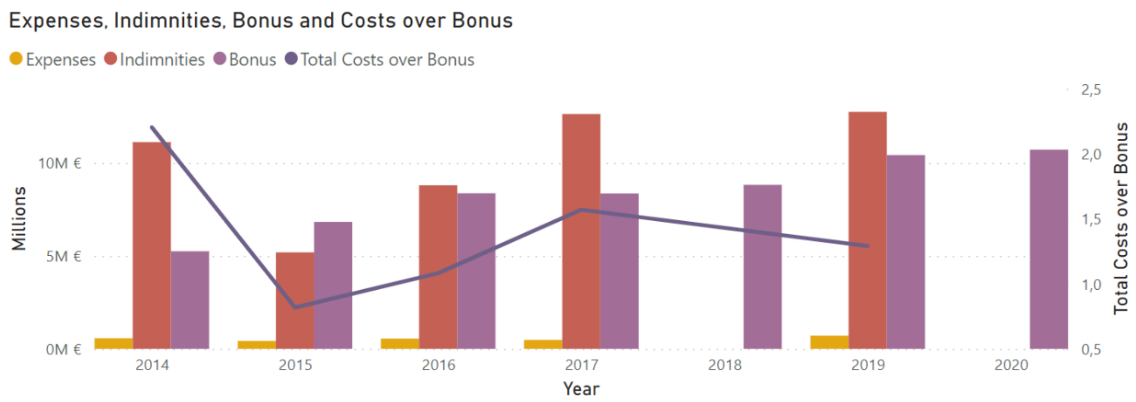

Figure 26. Evolution of Expenses (€), Indemnities (€), Bonus (€) and Total Costs over Bonus – SC

A similar analysis is done per region and per crop. We could see that in region A, which includes some counties of Lisbon and Algarve, there was a slight decrease of claims registered in the last 25 years. The ratio of insurance total costs over bonus tends to be around 1, meaning that the costs for the insurance companies and the State were very similar.

For region C, which corresponds to Alentejo, mostly countryside, the main crops and causes of claims tend to vary between insurance agreements, and fire and heavy rain are some of the principal causes of incidents. Alentejo is the driest region of Portugal, and the heavy rain occurrences are a surprising fact, but that can be explained by some south winds that bring higher levels of rainfall to the region. If these phenomena happen in a short period, that could explain damages caused by heavy rain.

For region D, the inner north of Portugal, the number of claims has been regular in the last few years. The ratio between insurance companies' and State expenses is the highest, especially in the last years. It tends to be around 1.09 euros for the insurance companies per euro spent by the State. Region E includes the counties of Vila Real, Bragança, Viseu and Guarda. It registers the most significant values for claims with a tendency to remain the same over the last ten years, around 200 claims. The ratio total costs- bonus is 1.37, which leads to this being one of the most significant regions in terms of the amount that the insurance companies pay when compared with the State. For Vineyards, in the last decade, there has been an increase in the number of incidents, however never exceeding the values registered at the end of the twentieth century. The other crops presented an irregular behaviour.

3.2.2 Contract insurance data

In this section, we analyse the insurance data defined a priori to the claims. As mentioned, for each contract we have information on the Insured Capital, the Commercial Prize, which is the total amount of insurance premium net of fiscal expenses and the bonus.

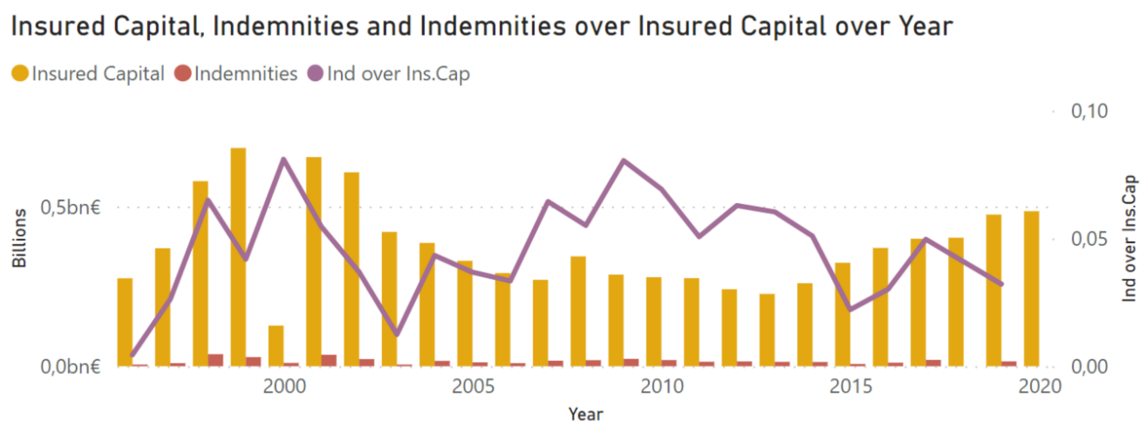

Figure 27. Evolution of Insured Capital (€), Indemnities (€) and Indemnities over Insured Capital– All

As presented in Figure 27**,** the Insured Capital is very high at the end of the 90s decade and beginning of 2000s, due to the public incentives that led to a strong participation of farmers, with more than 100 000 farmers applying for the aid. These values decreased between 2000 and 2015, but it increased again in the last five years, reaching almost half a billion euros. In order to see how the value of the goods insured relates with the actual value of the indemnities we consider the following ratio, which is represented in the graphs by the purple line:

$$\frac{Indemnities}{Insured\ Capital}. \tag{7}$$

Per euro insured by the farmers, the insurance companies only pay around 5% of that value in indemnities. The years where the Insured Capital decreases verify an increase in the ratio of indemnities over Insured Capital. In 2010 the insurance companies paid 8% of indemnities compared to the value of all insured goods.

Although there is a decrease in the Capital Insured, the incidents occurring did not decrease at the same scale. Thus, the value of the damages was more significant, which could explain the bigger participation of the farmers in the last years.

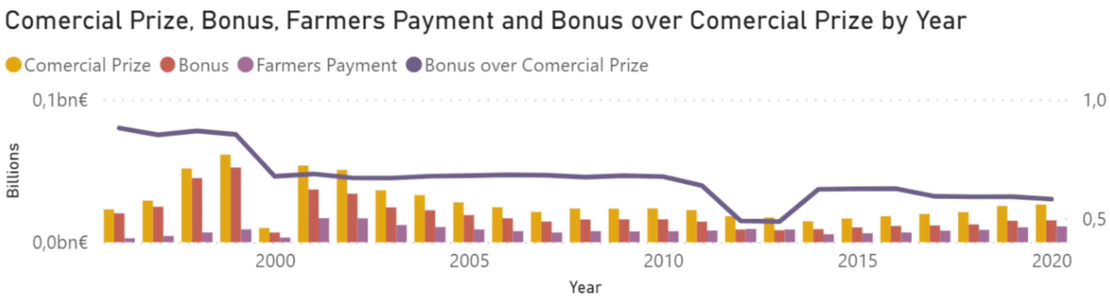

Figure 28. Evolution of Commercial Prize (€), Bonus (€) , Farmers Payment (€) and Bonus-Commercial Prize ratio over the years- All Agreements

As mentioned before, until 2000, the State covers around 85% of the farmers insurance expenses, but in the beginning of the new century, there was a cut from 85% to 65%. The worst moment for farmers was in 2012 and 2013, where the amount of help covered less than 50% of all the costs. This period paralleled with the change of insurance contracts, when they went from the SIPAC agreement to SVC and SC agreements. Afterwards, the State help remained constant, around 60%, which means that farmers support 40% of the costs with insurance. In Figure 28, we can see how the responsibilities shifted from the State to the farmers, throughout the years, contributing with 40% of the costs as opposed to the first years where they contributed with 15%.

In Figure 29, we see that the insurance companies-bonus ratio increased in the last years.

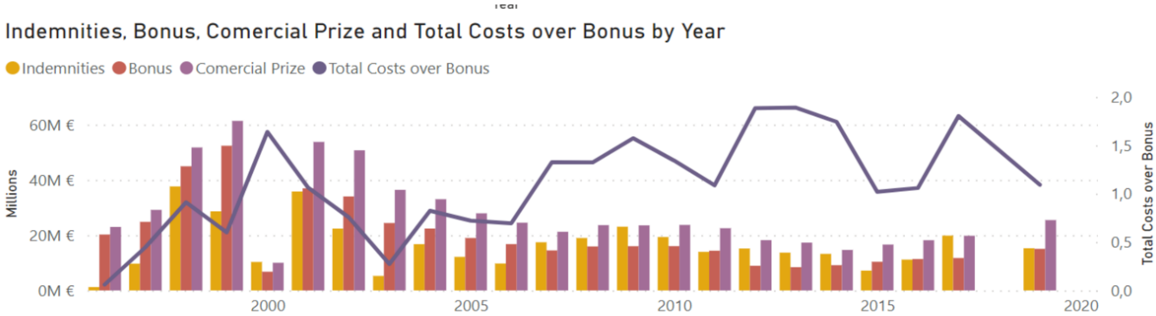

Figure 29. Evolution of Indemnities (€), Bonus (€), Commercial Prize (€) and Costs over Bonus by years- All Agreements

Doing the same analysis for SIPAC agreements, the sum of indemnities over the insured capital increased after 2004. At the same time, the indemnities over the State expenses were also increasing after that same year. On the other side, for the SVC agreement, which includes only Vineyards, the decrease of the ratio of indemnity over-insured capital, since the Capital Insured increased. The amount of support of the State remained constant, around 60%, and the value for total cost-bonus ratio slightly decreases, remaining higher than the State's payments.

For the SC insurance agreement, what is worth mentioning is that the State aid, between 2016 and 2017, suffered a clear cut, with the help decreasing from 64% to 59%, continuing to decrease in the latest years. Here, we continue to have higher expenses on the insurance companies' side than on the State's side. Compared with the other crops, the Vineyard shows the same level of State support, but not as much irregularities in the evolution.

Knowing that the Commercial Prize is the Insured Capital times a tariff, it is possible to calculate the tariff associated with each crop. By observing Figure 30, the main conclusions are that the tariff values have decreased since 1996. The average tariff started with 0.083, meaning that the insurance companies receive 0.083 euros per Insured Capital euro. Since 2000, this value has been decreasing, and it reached 0.054, in 2020.

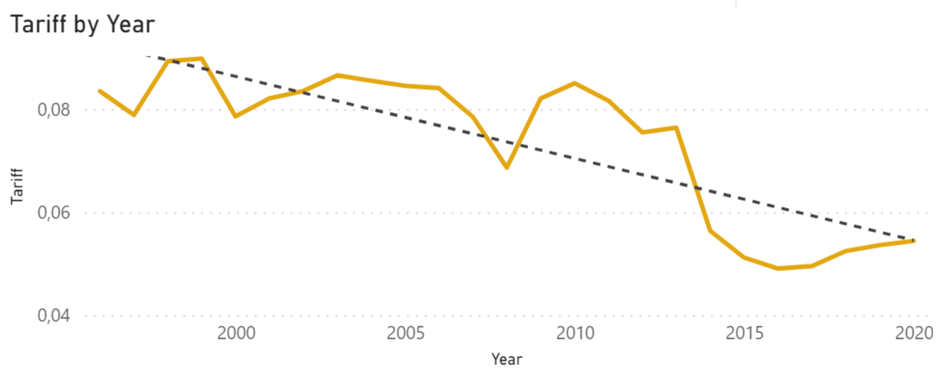

Figure 30. Evolution of tariffs (€)- All Agreements

In Figure 31, we can see the regions in the north of Portugal with higher tariffs, by the size of the circles, which can indicate a bigger probability of claims. Indeed, it is the region with the higher number of claims. Region A corresponds to the orange dots, region B to the green dots and regions C, D and E to the blue, red and purple dots, respectively.

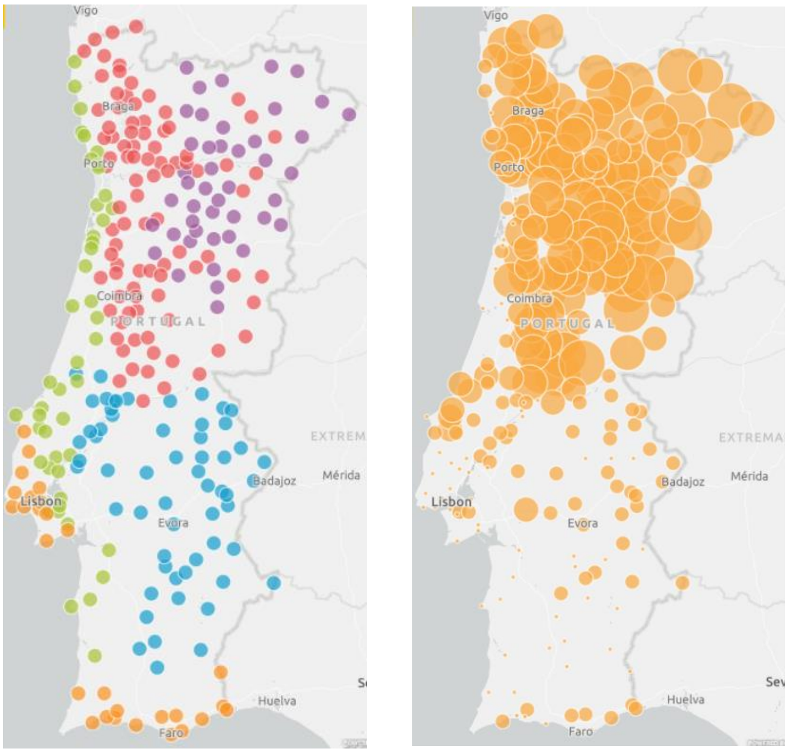

Figure 31. Tariffs (€) in Portugal

Looking at the differences in the evolution per region, in Figure 32, region E has an increasing tendency of tariffs' value. Although following more or less the behaviour of other regions, after 2014 it starts to increase while the others decrease or continue relatively stable. Region E relates to Portugal's innermost north part.

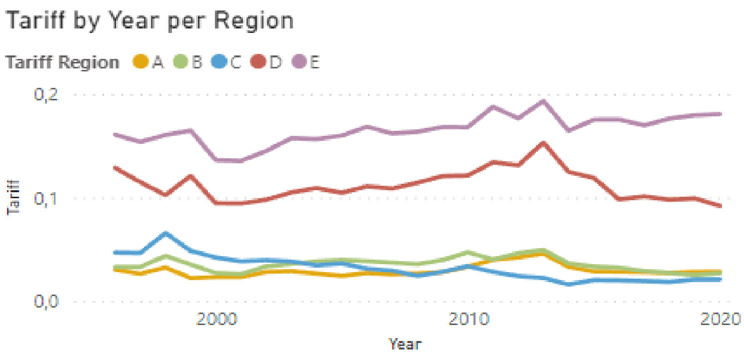

Figure 32. Tariffs (€) per Region over the Years

There is a relatively stable tendency for region D with the average tariff being 0.109. Region C has a decreasing tendency on tariffs and relatively low values for it. This region corresponds mainly to Alentejo. In Region B we verify very irregular fluctuations of the tariff throughout the years, that end up creating a relatively stable tendency. Finally, region A, which includes Lisbon and Algarve areas, shows an increasing tendency for the value of the tariffs, which are on average 0.029.

In the SIPAC agreement, the average tariff for all products is around 0.08. The tariffs have a decreasing tendency with very irregular fluctuations. The highest tariffs applied were in 1999 with 0.09, and the lowest in 2008 with 0.07. In 2006 there is a drastic drop followed by a drastic increase in 2009.

The most significant tariffs are associated with Walnut, Hazelnut, and Almond with tariffs values of more than 0.2 euros. The lowest tariffs were paid in crops such as Safflower, Canary seed, and Horticulture, with 0.02 euros. The regions' evolution, for SIPAC Agreement, shows a similar behaviour, with regions E and D standing out with the most significant values and the biggest increase in 2013.

For the SVC insurance agreement, the average tariff was 0.051. Starting with a tariff of 0.087 in 2012 and ending with a tariff of 0.042 euros in 2020, we have an evident decreasing tendency. Region E is again outstanding, with the highest values and a very steady evolution compared with region D. The only culture in this agreement is the Vineyard which is mainly located in the north of Portugal.

Finally, for the SC insurance Agreement, there is an evident tendency of increase of tariffs values, on average, for all regions, with regions E and D being those with higher tariffs. The crops that stand out the most are Walnut, Cherry, Quince, and Peach, with tariff values of 0.246,0.238,0.22 and 0.213, respectively. On the other hand, we have rice, Sorghum, and Barley crops with very low tariffs. The average tariff here is around 0.055 euros.

Looking into the crops with higher tariffs, Figure 33, Walnut production has an average tariff of 0.243, and it is produced mainly in regions C, D, and E. The tariff tendency has been steadily increasing over the years. Quince culture is very restricted in the areas where it is produced, mainly in Viseu and Guarda, and it only started to appear after 2016. Nonetheless, it has 0.22 of average tariff. The Cherry is common in the inner North of Portugal. Its average tariff is around 0.218, and its tendency is increasing over the years.

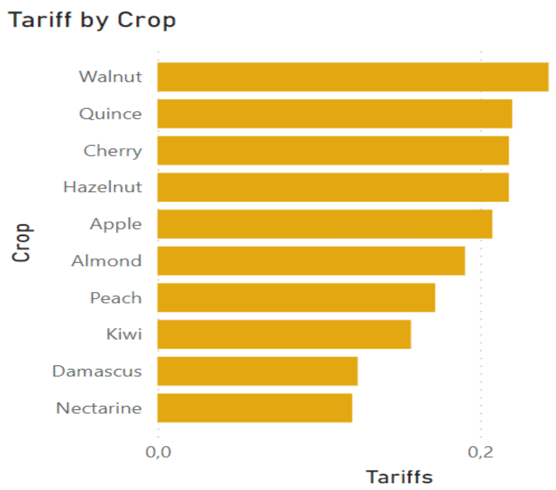

Figure 33. Tariffs (€) per Crop over the Years

3.3 Impact of climate data in insurance data

In this section, we analyse the impact of weather variables on agriculture-related insurance.

3.3.1 Methodology

There are several pieces of evidence in literature that link weather phenomena and their climate variables with agriculture, leading to “disaster payments (..) affected by weather and long-term climate variables”, as mentioned in [8]. Many studies, such as [27], define their variables keeping in mind the thresholds defined in insurance contracts, that set limits for weather disasters. In the data used for this study, the applicability of the insurance indemnities does not depend on weather variable limits, but on production losses thresholds instead.

One of the approaches indicated in [8] and [28] is to use models that directly study crop yields against weather variables. Another possible methodology is to use other variables, of socioeconomic nature, to do studies at an aggregate level. In this work, we have access to payment per culture and region, as well as weather data. Thus, we choose to plot climate data against insurance-related data.

We consider a regression model approach with equations of the form of:

$$Y_i = f(X_{wi}) + e_i. \quad (8)$$

Where the Y_i is the dependent variable under study and X_{wi} are all the weather variables for each period i. f represents the function used on the explaining variables and e_i represents the data not consider in the function f that still explains Y_i.

Type	Variable Name	Description	Variable Acronym
Dependent Variables	Tariffs	The price per unit insured, paid to the insurance companies	Tariffs
	Bonus over Commercial Prize	The proportion of what is paid by the State as part of the total amount of Commercial Prize that the Insurance companies receive	BoCP
	Indemnities	Total amount paid by insurance companies in case of incidents	Indemnities
	Claims over Number of Contracts	The proportion of claims for the total number of contracts registered	CoNC

Table 1. Dependent Variables

We aim at explaining variables such as (i) the Tariffs; (ii) Indemnities; (iii) Bonus over Commercial Prize, and (iv) Claims over the number of contracts through climate variables. The definition of the dependent variables is presented in Table 1.

In order to combine the yearly information from the insurance data with the daily information from the weather climate variables, the later are transformed into yearly variables. This transformation is done by finding measures that translate the evolution, tendencies, and outliers found in *Chapter 2*. The definition of the climate variables can be found in Table 2.

Type	Variable Name	Description	Variable Acronym
Independent Variables	Days_above_treshold_x	Number of days for each year that are above the third quantile (limit of the 25% highest observations) of the whole variable distribution, between 1996 and 2020	DATx
	Days_under_treshold_x	Number of days for each year that are under the first quantile (limit of the 25% lowest observations) of the whole variable distribution, between 1996 and 2020	DUTx
	Days_above_mean_x	Number of days for each year that are above the mean of the whole variable distribution, between 1996 and 2020	DAMx

	Min(x)	Minimum values per year for each variable. The values are calculated per group of region, stations and year.	Minx
	Max(x)	Maximum values per year for each variable. The values are calculated per group of region, stations and year.	Maxx
	Mean(x)	Mean values per year for each variable. The values are calculated per group of region, stations and year.	Meanx
	Fst_qtl_x_year	First quantile value for each variable distribution per year. The values are calculated per group of region, stations and year.	Fstx
	Trd_qtl_x_year	Third quantile value for each variable distribution per year. The values are calculated per group of region, stations and year.	Trdx
	Ratio_x	Number of available data (correct data) over the number of days in a year (365)	Rx

Table 2. Independent Variables.

The *x* represents the climate variable being referred to. The variables can be represented by *Rt* or *Rtotal* for rainfall; min or *Tmin* for minimum temperature; max or *Tmax* for maximum temperature; *Wspeed* or *W* for windspeed related variables and *dif* or *ThermA* referring to thermal amplitude.

In the study [8], the 99th and 1st percentile were defined for maximum and minimum temperature, respectively, so to capture the impacts of the highest and lowest temperatures. With variables such as *FstRtotal*, *FstTmin*, *TrdTmax*, and *TrdThermA*, among others, we tried to capture the evolution of the highest and lowest observations for all-weather phenomena registered. In [28], abnormal values of temperature and precipitation were used to study the farmers' decision to contract insurance services. These variables are used to quantify extreme weather situations. In our study, the outliers were defined as days above or under the thresholds representing the number of days in each year, below the 25% lowest observations or above the 25% highest observations. The findings throughout the process led to the creation of new variables, such as the mean and thermal amplitude-related variables. As new variables are created and new approaches experimented, several examples were tested. The primary analysis was done on initial variables such as DATrt, DATmin, DATmax, DATw, DUTrt, DUTmin, DUTmax, DUTw, MinRt, MinTmin, MinTmax, MinWspeed, MaxRt, MaxTmin, MaxTmax, MaxWspeed, FstRt, FstTmin, FstTmax, FstWspeed, TrdRt, TrdTmin, TrdTmax, TrdWspeed, RTmin, RTmax, and RWspeed.

The weather data analysed in *Chapter 2* was considered unfit for the regressions, because the outlier observations became inexistent. The initial study on the raw weather datasets allowed us to check that the collected information was in accordance to what is the Portuguese latest evolutions throughout the years and its seasonal distribution, see [4], [5], and [3]. The incorrect data was removed, taking into account the acceptable limits registered by IPMA, [14].

Since the insurance data comprised the years from 1996 to 2020, only the weather data from those years was considered, mainly the variables from automatic stations only, which cover a period between the 1995 and 2018. From the analysis in the previous section, it is clear that there exist differences in the insurance variables' evolution when segmented by tariff regions and crops. Thus, the regressions' analysis is first performed at an aggregate level for all regions and all cultures combined, and after, by region and type of crop. The regions go from A to E, and the stations included in each can be consulted in Table 2 of the Appendix B. The crops are divided into two big groups, the Vineyard crops and all other except Vineyard. Since we want the data to be segmented for region and crop, all the yearly weather variables are calculated for each combination of region and crop. The climate and insurance data were combined by the Stations' Location and the insurance tariff regions, as shown in Table 2 in the Appendix B.

The correlation between independent variables in Figure 34, shows a strong correlation between all Ratio_x variables and the other variables, as expected. Because of such correlations, the ratios were not included in the regressions. After which, all correlations decreased significantly.

The group of DAT and DUT variables also presented correlations between them. However, they are not so significant and were still included in the models, and the results confirmed those variables to be significant and important.

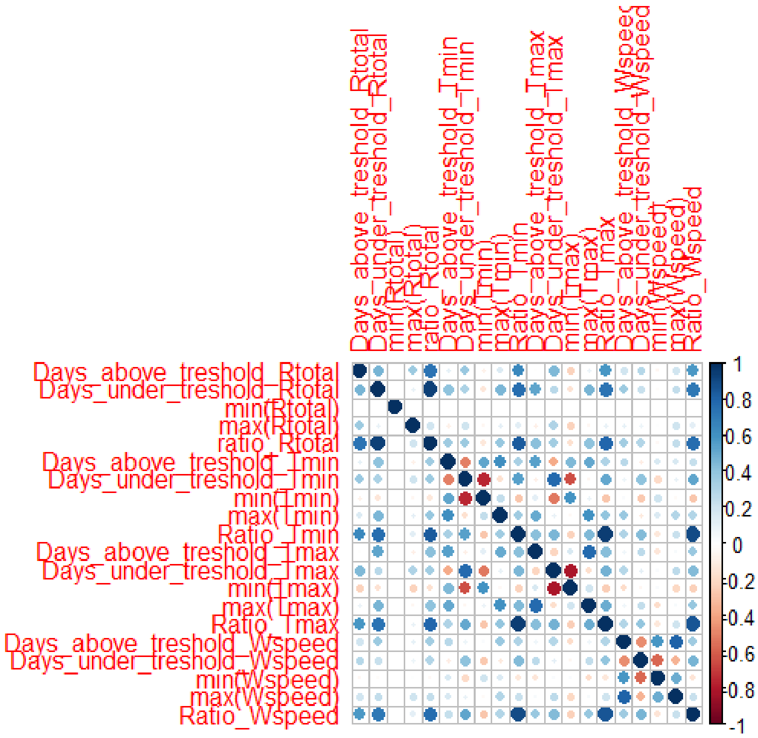

Figure 34. Correlogram of initial variables

For all the dependent variables being studied, scatterplots against the independent variables are generated. The main conclusion is that there is no clear relationship, as we can see in the example of Figure 35. The relationships had similar behaviour for all the dependent variables except for the indemnities.

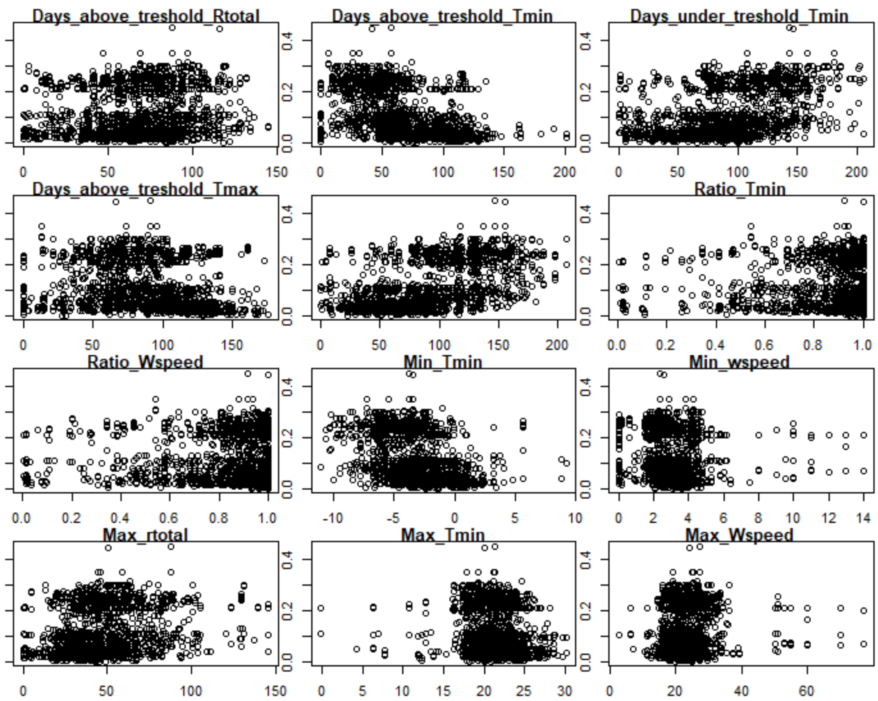

Figure 35. Scatter Plots of Independent variables against the Tariffs

First, to find the final models, all-weather variables are considered. Looking at the statistical significance of each variable as well as the significance of the model as a whole, some variables are dropped. This is done recursively until reaching the final model. R^2, adjusted R^2 and F-tests are analysed to understand the total capacity of the model to explain the dependent variable. A similar approach is made in [28], to access the variables' goodness for regression models with the variables' significance level and the F-tests levels for the significance of the models as a whole.

With the entire set of variables, those showing no significance are dropped, one at the time. The improvements of the model significance are analysed. When R^2 improvement stagnates, the model is considered good. The remaining variables are then tested for quadratic relationships, to understand if those explain better the behaviour of the dependent variable.

3.3.2 Models and Results

3.3.2.1 Regressions for variables a priori to hazards

The results from the regression models obtained to explain the tariffs values, as expected, show differences between regions and crops. As explained before, we first study models that include all tariff regions and all crops, and then study smaller datasets with specific regions and groups of crops, Vineyards or all the others except Vineyards. The first regression considered is as follows:

(9) $Tariffs = \beta_0 + \beta_1 DATRt + \beta_2 DATRt^2 + \beta_3 MaxRt + \beta_4 MaxRt^2 + \beta_5 DATmin + \beta_6 DATmin^2 + \beta_7 MinTmin + \beta_8 MinTmin^2 + \beta_9 DUTmax + \beta_{10} DUTmax^2 + \beta_{11} MinTmax + \beta_{12} MinTmax^2 + \beta_{13} MaxTmax + \beta_{14} MaxTmax^2 + \beta_{15} DATw + \beta_{16} DATw^2 + \beta_{17} DUTw + \beta_{18} DUTw^2 + \beta_{19} MinWspeed + \beta_{20} MinWspeed^2 + \beta_{21} MaxWspeed + e.$

```
Residuals:
      Min        1Q    Median        3Q       Max 
-0.229796 -0.043641 -0.005935  0.044512  0.252953 

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)                          0.0864323  0.0071358  12.113  < 2e-16 ***
poly(a$Days_above_treshold_Rtotal, 2)1 -0.3621926  0.0969742  -3.735 0.000191 ***
poly(a$Days_above_treshold_Rtotal, 2)2  0.1456215  0.0807232   1.804 0.071321 .  
poly(a$`max(Rtotal)`, 2)1            0.7308829  0.0778350   9.390  < 2e-16 ***
poly(a$`max(Rtotal)`, 2)2           -0.2529922  0.0784787  -3.224 0.001277 ** 
poly(a$Days_above_treshold_Tmin, 2)1   -1.6161841  0.1157555 -13.962  < 2e-16 ***
poly(a$Days_above_treshold_Tmin, 2)2    0.1867056  0.0817951   2.283 0.022512 *  
poly(a$`min(Tmin)`, 2)1              0.0956035  0.1009414   0.947 0.343641    
poly(a$`min(Tmin)`, 2)2              0.3316598  0.0784414   4.228 2.41e-05 ***
poly(a$Days_under_treshold_Tmax, 2)1    1.2667522  0.1518467   8.342  < 2e-16 ***
poly(a$Days_under_treshold_Tmax, 2)2    0.3906128  0.0914580   4.271 2.00e-05 ***
poly(a$`min(Tmax)`, 2)1             -1.7471487  0.1295763 -13.484  < 2e-16 ***
poly(a$`min(Tmax)`, 2)2             -0.5710889  0.0891769  -6.404 1.71e-10 ***
poly(a$`max(Tmax)`, 2)1             -0.1092764  0.1064184  -1.027 0.304557    
poly(a$`max(Tmax)`, 2)2             -0.3618947  0.0814171  -4.445 9.06e-06 ***
poly(a$Days_above_treshold_Wspeed, 2)1 -0.8261913  0.1592506  -5.188 2.24e-07 ***
poly(a$Days_above_treshold_Wspeed, 2)2  0.4618632  0.1074484   4.298 1.77e-05 ***
poly(a$Days_under_treshold_Wspeed, 2)1  0.2744296  0.1206718   2.274 0.023013 *  
poly(a$Days_under_treshold_Wspeed, 2)2  0.3188418  0.0785323   4.060 5.01e-05 ***
poly(a$`min(Wspeed)`, 2)1           -0.2189516  0.1174398  -1.864 0.062350 .  
poly(a$`min(Wspeed)`, 2)2            0.4510745  0.0936583   4.816 1.52e-06 ***
a$`max(Wspeed)`                      0.0018252  0.0002988   6.107 1.12e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06634 on 3613 degrees of freedom
  (39 observations deleted due to missingness)
Multiple R-squared:  0.5509,	Adjusted R-squared:  0.5483 
F-statistic: 211.1 on 21 and 3613 DF,  p-value: < 2.2e-16
```

Figure 36. Results of Regressions on Tariffs for all regions and all crops

This regression takes into account all regions and crops resulting in a model explanatory capacity of 55%. The results can be seen in Figure 36. The most significant variables are DATRt, DATmin, MinWspeed, DUTmax, DUTwspeed, and the maximum and minimum for most of the weather variables. The relationships are mainly quadratic, and almost all variables are significant at a five percent level. The coefficients of the variables show substantial impacts on the Tariffs, bearing in mind that the units are in euros. For all crops and regions, MinTmax negatively impacts tariffs, and DUTmax has the most significant positive impact on tariffs. We can observe that tariffs are vulnerable to the changes in the weather variables.

The regression model used to explain the data for all regions and all other crops except Vineyard, represented in Equation 10, reached very high levels of explanatory capacity, with an R^2 of 70% approximately, see Figure 47. The F-test values indicate the model to be significant at a 5% level. The main highlights are that for all-weather variables, the DAT and DUT variables seem to be present and significant.

(10) $Tariffs = \beta_0 + \beta_1 DATRt + \beta_2 DATRt^2 + \beta_3 MaxRt + \beta_4 MaxRt^2 + \beta_5 DATmin + \beta_6 DATmin^2 + \beta_7 DUTmin + \beta_8 DUTmin^2 + \beta_9 MinTmin + \beta_{10} MinTmin^2 + \beta_{11} DATmax + \beta_{12} DATmax^2 + \beta_{13} DUTmax + \beta_{14} DUTmax^2 + \beta_{15} MinTmax + \beta_{16} MinTmax^2 +$

$\beta_{17}DATw + \beta_{18}DATw^2 + \beta_{19}DUTw + \beta_{20}DUTw^2 + \beta_{21}MinWspeed + \beta_{22}MinWspeed^2 + \beta_{23}MaxWspeed + e.$

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.253666 -0.038000 -0.004884  0.038296  0.218341

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                              0.0798144  0.0072587  10.996  < 2e-16 ***
poly(a$Days_above_treshold_Rtotal, 2)1 -0.3225211  0.0877486  -3.676 0.000242 ***
poly(a$Days_above_treshold_Rtotal, 2)2  0.1844419  0.0721465   2.556 0.010628 *
poly(a$`max(Rtotal)`, 2)1                0.7840402  0.0699070  11.215  < 2e-16 ***
poly(a$`max(Rtotal)`, 2)2               -0.1430000  0.0709328  -2.016 0.043900 *
poly(a$Days_above_treshold_Tmin, 2)1   -1.5996390  0.1199126 -13.340  < 2e-16 ***
poly(a$Days_above_treshold_Tmin, 2)2    0.3225103  0.0856803   3.764 0.000171 ***
poly(a$Days_under_treshold_Tmin, 2)1   -0.0475766  0.1734169  -0.274 0.783838
poly(a$Days_under_treshold_Tmin, 2)2   -0.4892481  0.1028289  -4.758 2.06e-06 ***
poly(a$`min(Tmin)`, 2)1                  0.3949917  0.1088268   3.630 0.000289 ***
poly(a$`min(Tmin)`, 2)2                  0.4695302  0.0802021   5.854 5.37e-09 ***
poly(a$Days_above_treshold_Tmax, 2)1   -0.2274943  0.0984035  -2.312 0.020861 *
poly(a$Days_above_treshold_Tmax, 2)2   -0.2870929  0.0752073  -3.817 0.000138 ***
poly(a$Days_under_treshold_Tmax, 2)1    1.4877082  0.1623416   9.164  < 2e-16 ***
poly(a$Days_under_treshold_Tmax, 2)2    0.6241937  0.0963715   6.477 1.11e-10 ***
poly(a$`min(Tmax)`, 2)1                 -2.0187944  0.1261472 -16.003  < 2e-16 ***
poly(a$`min(Tmax)`, 2)2                 -0.7189450  0.0820160  -8.766  < 2e-16 ***
poly(a$Days_above_treshold_Wspeed, 2)1 -0.8406396  0.1413184  -5.949 3.06e-09 ***
poly(a$Days_above_treshold_Wspeed, 2)2  0.6696901  0.0955196   7.011 2.98e-12 ***
poly(a$Days_under_treshold_Wspeed, 2)1  0.5717670  0.1251826   4.567 5.16e-06 ***
poly(a$Days_under_treshold_Wspeed, 2)2  0.2263490  0.0731188   3.096 0.001984 **
poly(a$`min(Wspeed)`, 2)1               -0.3232016  0.1047208  -3.086 0.002047 **
poly(a$`min(Wspeed)`, 2)2                0.3965389  0.0845832   4.688 2.89e-06 ***
a$`max(Wspeed)`                          0.0025823  0.0003044   8.483  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05899 on 2689 degrees of freedom
  (39 observations deleted due to missingness)
Multiple R-squared:  0.7031,	Adjusted R-squared:  0.7006
F-statistic: 276.9 on 23 and 2689 DF,  p-value: < 2.2e-16
```

Figure 37. Results of Regressions on Tariffs for all regions and all crops except Vineyards

The higher capacity of the model to explain the level of the tariffs may indicate that the crops besides Vineyards are more sensitive to the changes that occur from year to year. This conclusion is under what is mentioned in [5], where simulations on future climate scenarios and agriculture approaches show how different crops depend on climate variability. In that study, it is interesting to see that the crops that are more dependent on water resources, such as horticulture, maize, and fruit trees, are more sensitive to climate changes. On the other hand, crops such as olive and grapevine are less influenced by the climate scenarios because they are better adapted to Mediterranean conditions.

For the models exposed in Figures 36 and 37, most variables explain the tariffs better when having a quadratic behaviour.

A higher explanatory capacity for other crops except Vineyards is observed for region A. The R^2 reaches 70% with a p-value for the F-test that indicates the model is significant in explaining the Tariffs. Equation 11, for region A, all crops but Vineyards, most variables presented quadratic relationships with positive coefficients, as we can see in Figure 38. A continuous increase of DATRt, DATmin, DATmax, and MinTmax leads to a significant increase in the Tariffs. The growth of maximum temperature and wind speed also positively impacts the Dowsed Tariffs.

$$(11)\quad Tariffs = \beta_0 + \beta_1 DATRt + \beta_2 DATRt^2 + \beta_3 DATmin + \beta_4 DATmin^2 + \beta_5 DATmax + \beta_6 DATmax^2 + \beta_7 DUTmax + \beta_8 DUTmax^2 + \beta_9 MinTmax + \beta_{10} MinTmax^2 + \beta_{11} MaxTmax + \beta_{12} MaxWspeed + e.$$

```
Residuals:
       Min         1Q     Median         3Q        Max
-0.0089861 -0.0027501  0.0000209  0.0025827  0.0164679

Coefficients:
                                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                              -0.0347744  0.0108899  -3.193 0.001988 **
poly(a$Days_above_treshold_Rtotal, 2)1  0.0684866  0.0086113   7.953 8.06e-12 ***
poly(a$Days_above_treshold_Rtotal, 2)2  0.0138658  0.0084679   1.637 0.105322
poly(a$Days_above_treshold_Tmin, 2)1   -0.0419515  0.0102803  -4.081 0.000103 ***
poly(a$Days_above_treshold_Tmin, 2)2    0.0206510  0.0118053   1.749 0.083936 .
poly(a$Days_above_treshold_Tmax, 2)1   -0.0540613  0.0089876  -6.015 4.66e-08 ***
poly(a$Days_above_treshold_Tmax, 2)2    0.0162743  0.0092383   1.762 0.081815 .
poly(a$Days_under_treshold_Tmax, 2)1    0.0263858  0.0100430   2.627 0.010249 *
poly(a$Days_under_treshold_Tmax, 2)2    0.0162943  0.0085388   1.908 0.059814 .
poly(a$`min(Tmax)`, 2)1                 0.0280597  0.0076290   3.678 0.000416 ***
poly(a$`min(Tmax)`, 2)2                 0.0143044  0.0067788   2.110 0.037855 *
a$`max(Tmax)`                           0.0015081  0.0002746   5.493 4.24e-07 ***
a$`max(Wspeed)`                         0.0006128  0.0001307   4.687 1.07e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.005065 on 83 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.7477,    Adjusted R-squared:  0.7112
F-statistic:  20.5 on 12 and 83 DF,  p-value: < 2.2e-16
```

Figure 38. Results of Regressions on Tariffs for Region A and all crops except Vineyards

For regions C, D, and E, the models with higher explanatory capacity are those of Vineyard crops. This is expected because these regions, especially regions D and E, have a predominancy of Vineyards. Nonetheless, the capacity of the models to explain the variable tariff is not as high as it is for the whole country. Despite lower values for the R^2, the models are significant at a five percent level. The details of the models for regions C, D and E can be found in the Appendix C in Figures 1, 2 and 3.

For the different crop groups comparison, the main insights are that DATmin has mainly positive effects on the evolution of the tariff's values, for all crops except Vineyards, and mainly negative effects for Vineyards. DUTmin does not seem to have much importance on the models that explain the tariffs, concerning Vineyard crops. Only for region E there are variables such as, MinTmin, resulting in increases of the tariffs, as suggested in [29] where the sensitivity of Vineyard crops to low negative temperatures is highlighted.

DATRt has mainly positive effects for both groups of crops, meaning it increases tariff values as more days verify rainfall values above the limit for the 25% highest observations. MaxTmin has mainly positive effects for Vineyard crops while for other crops except Vineyards this is not the case. In [26], is referred that fruit trees differ in their sensitivity to chilling conditions and although some grow better for cooler temperatures, others are well adapted to warmer conditions. This heterogeneity seems to be represented in our results where there is no pattern that variables like DATmax or MaxTmax follow.

Regarding the coefficients, it was noticed that DAT variables are present in almost all of the final models, with DATRt mostly linear and with positive signs. DATmin, DATmax, and DATw have quadratic relationships, with the signs differing between models. DUTmin had almost always a quadratic effect on the dependent variable of tariffs. MinTmin, MinTmax, and MinWspeed have mostly positive linear effects on the evolution of Tariffs. The impact of MinTmin and DUTmin are different from what is seen in [8]. In their study, the decrease of minimum temperatures is associated with higher disaster payments. At the same time, an increase in maximum temperatures also leads to increases in disaster payments. In our work DATmax, for half of the models has a positive relationship with the tariff's variable. For region C, this does not verify. Another important conclusion is the DATRt, is mainly related to increases in the tariffs, which is confirmed in [8], where higher precipitation levels lead to an increase in disaster payments. Again, for region C, this does not verify, which is explained by the fact that the regions included are mostly in Alentejo, known for lower values of rainfall. Hence, more days of rainfall above the threshold is considered beneficial for the crops taking into account the very low values it normally verifies.

In order to better explain the variable of the tariffs, it was assumed that, because the tariffs are values defined a priori to the hazards, it could be true that the previous year's weather variables have a higher impact on the tariff's levels for a specific year. To analyse that, four regions and culture groups are chosen to see what the impacts are at a first try. The results can be found on the following Table:

Experience	Description	Original R^2	Final R^2
1	All tariff Regions and all crops except Vineyards	70%	65.5%
2	Region D and all crops	20%	28%
3	Region E and all crops	6%	7.2%
4	Region B and all crops except Vineyards	42%	71.9%

Table 3. Comparison of Results of Regressions for Tariffs with and without lag years

The first experience is done for all tariff regions and all crops but Vineyards, which previously had the highest R^2. The second experience is performed on region D for all crops, where before it registered an explanatory capacity of around 20%. The third experience is for region E and all crops, where the R^2 was no more than 0.06. For the last experience, region B is chosen for all crops, but Vineyards with a previous explanatory capacity of 42%. More detailed information about the model results for each scenario, can be found in the Appendix C.

When applying the new regressions of tariffs with the previous years' weather variables, we see that the final results remained very similar for all the examples. In some cases, the explanatory capacity of the models decreased. In others, it increased, and only for the model of region B the improvements were significant. In terms of significance of the variables there are no substantial changes that indicate that the variables that are not significant before became significant for the models with a one-year lag.

Nonetheless, it is seen that DAT and DUT of the weather variables are still the ones that are always present after the selection process. Once more, these variables perform better when applying a quadratic relationship. Details on the calculated models, can be found in the Appendix C, in Figures 4,5, 6 and 7. These results are expected if we believe that the weather does not have drastic changes from one year to the other. The similarities over consecutive years can explain the similarities in the model's significance levels. Future studies may be necessary on relating not only the previous year but also some past consecutive years on the study of the tariff level.

As the research process evolves, there are improvements made to the dataset. In an initial phase, the variables chosen are related mainly to more extreme weather observations. The next step is to verify if more central variables better explain the tariffs. The mean of the variables for each year is introduced, and the number of observations in a year above that mean. The latest is calculated for all variables, and a new weather phenomenon is introduced, the thermal amplitude. The choice of the thermal amplitude results from the fact that hail is a very present cause of claim all over the country. In the absence of information related to humidity, the thermal amplitude could explain it to some extent.

To analyse the significance of new variables, they are introduced in the models that cover all regions. The first try is for all regions and all crops, in the second try the model covers data for all regions and vineyard's crop, and for the last try, the dataset used is for all regions and all other crops except vineyards.

Experience	Description	Original R^2	Final R^2
1	All regions and all crops	54%	60%
2	All regions and Vineyard Crop	41%	45%
3	All regions and all crops except Vineyard Crops	70%	78%

Table 4. Comparison of Results of Regressions for Tariffs with and without new variables

In Table 4, we may see the final results after introducing new variables for these three scenarios. The new variables showed significance and improved the models. For the first try, the model achieved an R^2 of 0.6, showing a higher explanatory power over the tariff's values. The Adjusted R^2 was also higher. For the second try, the improvements were minor, but the new variables showed significance at a 5% level. For the last test, the R^2 reaches 0.78, which gives an excellent explanatory capacity to this model. More details on the models generated can be found in the Appendix C, in Figures 8,9 and 10.

Introducing the new variables could be considered an improvement to all regressions and so considered essential to look at when defining tariff values for insurance contracts.

One of the other variables studied is the ratio of BoCP. Here, we intend to understand how the weather variables can explain how the State's aid varies. The scatterplots for BoCP do not show any specific behaviour of the independent variables. The model that includes all regions and all crops is as follows:

(12) $Bonus\ over\ Commercial\ Prize = \beta_0 + \beta_1 DATRt + \beta_2 DATRt^2 + \beta_3 TrdRt + \beta_4 TrdRt^2 + \beta_5 DATmin + \beta_6 DATmin^2 + \beta_7 DUTmin + \beta_8 DUTmin^2 + \beta_9 MaxTmin + \beta_{10} MaxTmin^2 + \beta_{11} DATmax + \beta_{12} DATmax^2 + \beta_{13} DUTmax + \beta_{14} DUTmax^2 + \beta_{15} TrdTmax + \beta_{16} TrdTmax^2 + \beta_{17} DATw + \beta_{18} DATw^2 + \beta_{19} DUTw + \beta_{20} DUTw^2 + \beta_{21} MaxWspeed + \beta_{22} MaxWspeed^2 + \beta_{23} TrdWspeed + \beta_{24} TrdWpseed^2$*+e.*

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.44909 -0.06151  0.00401  0.06671  0.39926

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                               0.629257   0.001685 373.376  < 2e-16 ***
poly(a$Days_above_treshold_Rtotal, 2)1 -1.556950   0.166705  -9.340  < 2e-16 ***
poly(a$Days_above_treshold_Rtotal, 2)2 -0.956557   0.131524  -7.273 4.30e-13 ***
poly(a$trd_qtl_Rtotal_year, 2)1          1.225919   0.134407   9.121  < 2e-16 ***
poly(a$trd_qtl_Rtotal_year, 2)2         -1.024024   0.135527  -7.556 5.24e-14 ***
poly(a$Days_above_treshold_Tmin, 2)1    -0.677597   0.237400  -2.854 0.004339 **
poly(a$Days_above_treshold_Tmin, 2)2    -0.254182   0.152502  -1.667 0.095652 .
poly(a$Days_under_treshold_Tmin, 2)1    -0.506627   0.269318  -1.881 0.060032 .
poly(a$Days_under_treshold_Tmin, 2)2    -0.471472   0.161837  -2.913 0.003599 **
poly(a$`max(Tmin)`, 2)1                  0.672235   0.227972   2.949 0.003211 **
poly(a$`max(Tmin)`, 2)2                  0.668866   0.192573   3.473 0.000520 ***
poly(a$Days_above_treshold_Tmax, 2)1    -1.560553   0.342434  -4.557 5.36e-06 ***
poly(a$Days_above_treshold_Tmax, 2)2    -1.295573   0.181278  -7.147 1.07e-12 ***
poly(a$Days_under_treshold_Tmax, 2)1     2.181268   0.442492   4.930 8.62e-07 ***
poly(a$Days_under_treshold_Tmax, 2)2     1.251835   0.207446   6.035 1.76e-09 ***
poly(a$`min(Tmax)`, 2)1                 -1.885769   0.220023  -8.571  < 2e-16 ***
poly(a$`min(Tmax)`, 2)2                  0.999853   0.147141   6.795 1.26e-11 ***
poly(a$`max(Tmax)`, 2)1                 -2.626863   0.349309  -7.520 6.87e-14 ***
poly(a$`max(Tmax)`, 2)2                  0.810560   0.304103   2.665 0.007724 **
poly(a$fst_qtl_Tmax_year, 2)1            3.404341   0.449873   7.567 4.81e-14 ***
poly(a$fst_qtl_Tmax_year, 2)2           -0.460810   0.196358  -2.347 0.018990 *
poly(a$trd_qtl_Tmax_year, 2)1            1.234822   0.431191   2.864 0.004211 **
poly(a$trd_qtl_Tmax_year, 2)2           -0.347947   0.307865  -1.130 0.258470
poly(a$Days_above_treshold_Wspeed, 2)1 -1.426403   0.355523  -4.012 6.14e-05 ***
poly(a$Days_above_treshold_Wspeed, 2)2  0.522834   0.316087   1.654 0.098198 .
poly(a$Days_under_treshold_Wspeed, 2)1  1.548236   0.242378   6.388 1.90e-10 ***
poly(a$Days_under_treshold_Wspeed, 2)2  0.427031   0.126960   3.364 0.000778 ***
poly(a$`max(Wspeed)`, 2)1                0.764083   0.214492   3.562 0.000372 ***
poly(a$`max(Wspeed)`, 2)2               -0.580208   0.159901  -3.629 0.000289 ***
poly(a$trd_qtl_Wspeed_year, 2)1          2.214904   0.391746   5.654 1.69e-08 ***
poly(a$trd_qtl_Wspeed_year, 2)2         -0.657406   0.256143  -2.567 0.010311 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1016 on 3604 degrees of freedom
  (39 observations deleted due to missingness)
Multiple R-squared:  0.2927,    Adjusted R-squared:  0.2869
F-statistic: 49.72 on 30 and 3604 DF,  p-value: < 2.2e-16
```

Figure 38. Results of Regressions on BoCP for all Regions and all crop

We conclude that the models showed less capacity to explain the bonus over Commercial Prize when compared with tariffs' models. As shown in Figure 38, the variables that are more present are the DAT and DUT ones. Almost all variables are significant at a level of 5%, and all the variables were better represented with a quadratic relationship to explain the dependent variable. This model presented a capacity to explain around 30% of the values of the ratios being studied.

For all regions and all crops except Vineyards the model got more robust results. This improvement is also noticed in the F-statistics where for Vineyard crops is 17.17 and for all crops except Vineyards is 48.18. Stronger models for all crops but Vineyards are also true for region A where we get an R^2 of 0.85. For regions C, D, and E, we were able to get greater values of explanatory capacity when considering only the Vineyards. The models enumerated can be found with more details in the Appendix C, in Figures 11 to 15.

In terms of coefficients and their signals, the most significant findings are that most models had DATRt and DATmin with quadratic behaviours and negative signals. This indicates that as the number of days above the third quantile for rainfall and minimum temperature increases, the proportion of what the State pays decreases. This relationship although not linear and not true for all the calculated models, is the most predominant. DAT and DUT variables are present in most of the models. The DUTw variable has positive coefficients that indicate that while the DUTw increases, the BoCP increases. The first and third quantile of the variables for each year were also significant for most models, with the TrdTmin having mostly negative linear impacts and TrdTmax with positive linear impacts.

In terms of the signals of the coefficients, there is no pattern that the models follow. For the BoCP models the new variables that concern the means, days above mean, and measures for the thermal amplitude are not used due to time constraints. Nonetheless, we believe that the same improvements previously seen for the tariffs would be verified here.

Although the bonus variable was not studied alone, it is believed that the high values of explanation of the models result from the fact that Commercial Prize is related to tariffs, as illustrated by Equation 4. Consequently, the variables that explain the tariffs end up having significance for the BoCP values. The models did not reach higher R^2 because the State's money depends on many conditions, mainly the funds made available for such programs and not solely on the impact of the weather variables.

There are no specific patterns of variables signals for the two groups of crops being analysed. For both groups, we verify that MaxRt leads to increases in the BoCP and MaxTmax leads to decreases for both Vineyard crops and all crops but Vineyard.

3.4.2.2 Regressions for variables a posteriori to hazards

In order to study variables that result from the hazards, an initial analysis on the correlations and variables scatter plots was done for indemnities.

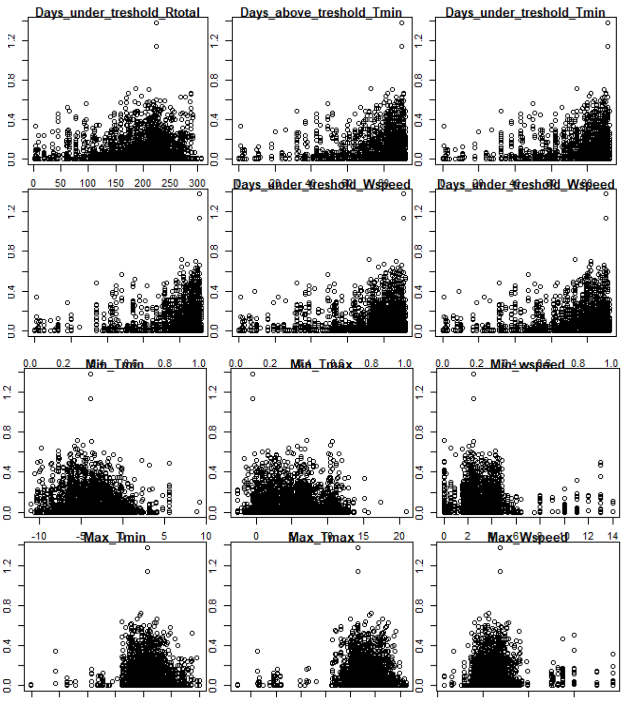

Figure 39. Example of Scatter Plots of Independent variables against the Indemnities

In Figure 39, the scatter plots suggest better and clearer relationships between the variables and the indemnities. Initially, it was considered that the indemnities variable would have the most robust models considering the R^2 and F-statistics. To reinforce this belief, we know that tariffs and bonus are defined a priori to the weather phenomena. For indemnities, its value is defined a posteriori and the existing relationship could be more direct.

Contrary to our assumptions, the regressions for indemnities or ratios using the indemnities have the poorest results. Neither the entire dataset nor the segmentation by region and crop improved the models. We obtain no significance for almost any variable. The models have an R^2 no bigger than 0.07, which gives almost no importance to the weather phenomena in explaining the values of Indemnities.

Looking at the plots of the variables, it could be considered that some follow an exponential behaviour. Thus, it was created the logarithm of indemnities to use in the regressions. The results continue to be unsatisfying, with poor significance for the variables individually and the entire models. Neither the corrections on the datasets nor the introduction of new variables lead to improvements in the significance of the models.

The indemnities are the combination of several factors and are aggregated values from several insurance companies that may define the indemnities payment differently. Such differences and constraints may be affecting the results of the regressions, indicating that this dependent variable is not directly correlated with weather evolution.

The last variable defined is the ratio of claims over the number of contracts. Using the number of claims alone leads to biased conclusions, because they depend on the number of farmers that applied to the State aid and not on the total number of hazards in the country. We chose to use a ratio that relates the number of claims in a year with the number of contracts made. In this way, we can see the proportion of hazards compared to the number of contracts, taking out some of the bias of the number of claims.

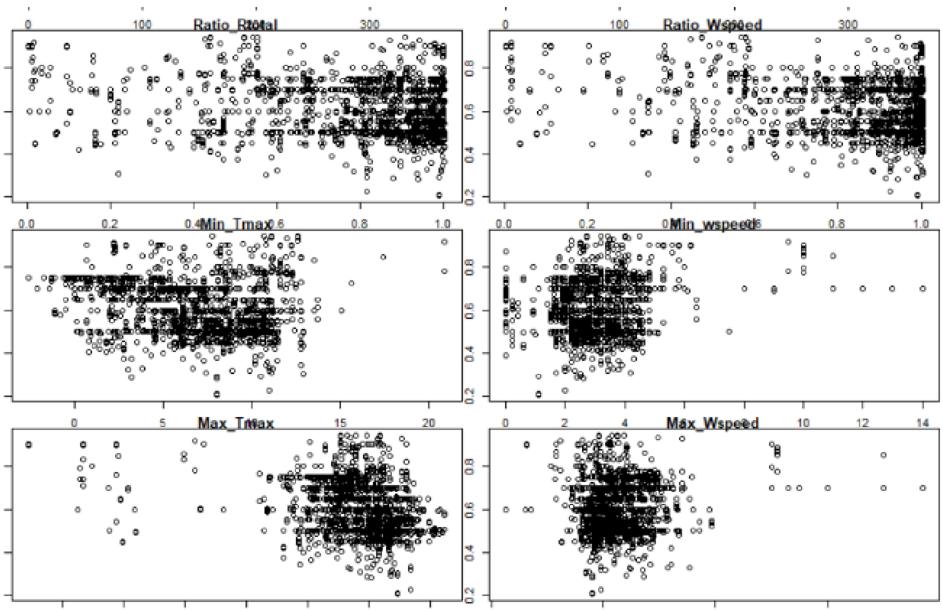

Figure 40. Scatter Plots of Independent variables against the CoNC

In figure 40, we see how the independent variables relate with effect variable. There is no clear relation that can be assumed. When accessing the regressions, most of the variables seem significant, as well as the models that present a high R^2.

For region A, all crops, the model is as follows:

(13) $Claims\ over\ Number\ of\ Contracts = \beta_0 + \beta_1 DATRt + \beta_2 MaxRt + \beta_3 MeanRt + \beta_3 TrdRt + \beta_4 TrdRt^2 + \beta_5 DATmin + \beta_6 DATmin^2 + \beta_7 DUTmin + \beta_8 DUTmin^2 + \beta_9 MaxTmin + \beta_{10} MaxTmin^2 + \beta_{11} DATmax + \beta_{12} DATmax^2 + \beta_{13} DUTmax + \beta_{14} DUTmax^2 + \beta_{15} TrdTmax + \beta_{16} TrdTmax^2 + \beta_{17} DATw + \beta_{18} DATw^2 + \beta_{19} DUTw + \beta_{20} DUTw^2 + \beta_{21} MaxWspeed + \beta_{22} MaxWspeed^2 + \beta_{23} TrdWspeed + \beta_{24} TrdWpseed^2 + e.$

```
Residuals:
     Min       1Q   Median       3Q      Max 
-0.66825 -0.00284  0.00075  0.06496  0.40655 

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                 -30.25064    4.18067  -7.236 1.90e-10 ***
a$Days_above_treshold_Rtotal  0.40894    0.06221   6.574 3.77e-09 ***
a$`max(Rtotal)`              -0.16499    0.02772  -5.952 5.77e-08 ***
a$`mean Rtotal`               7.38590    1.24406   5.937 6.14e-08 ***
a$Days_above_mean_Rtotal     -0.59560    0.08833  -6.743 1.77e-09 ***
a$Days_above_treshold_Tmin   -0.68573    0.13147  -5.216 1.27e-06 ***
a$`min(Tmin)`                -2.12659    0.39474  -5.387 6.27e-07 ***
a$`max(Tmin)`                -0.34442    0.08021  -4.294 4.63e-05 ***
a$`mean Tmin`                -1.44082    0.68137  -2.115 0.037393 *  
a$Days_above_mean_Tmin        0.22748    0.04416   5.151 1.66e-06 ***
a$fst_qtl_Tmin_year           3.08425    0.58588   5.264 1.04e-06 ***
a$trd_qtl_Tmin_year           3.85826    0.97214   3.969 0.000150 ***
a$Days_above_treshold_Tmax    0.52087    0.09475   5.497 3.97e-07 ***
a$`min(Tmax)`                 1.54840    0.28185   5.494 4.03e-07 ***
a$`max(Tmax)`                 2.28316    0.36892   6.189 2.06e-08 ***
a$`mean Tmax`               -18.14443    3.39846  -5.339 7.66e-07 ***
a$Days_above_mean_Tmax        0.04219    0.01071   3.939 0.000167 ***
a$fst_qtl_Tmax_year          12.92759    2.33593   5.534 3.40e-07 ***
a$trd_qtl_Tmax_year           2.28818    0.55121   4.151 7.82e-05 ***
a$Days_above_treshold_Wspeed -0.29127    0.05490  -5.305 8.81e-07 ***
a$`min(Wspeed)`               4.55042    0.89856   5.064 2.35e-06 ***
a$`max(Wspeed)`               0.87141    0.16984   5.131 1.80e-06 ***
a$`mean Wspeed`             -18.21341    3.56230  -5.113 1.93e-06 ***
a$Days_above_mean_Wspeed      0.10543    0.02103   5.013 2.89e-06 ***
a$fst_qtl_Wspeed_year        11.86468    2.30947   5.137 1.75e-06 ***
a$trd_qtl_Wspeed_year         7.15089    1.36560   5.236 1.17e-06 ***
a$Days_above_treshold_diff    0.05604    0.01234   4.540 1.84e-05 ***
a$`min thermal_amp`           6.95674    1.23808   5.619 2.38e-07 ***
a$`max thermal_amp`          -0.93242    0.14201  -6.566 3.90e-09 ***
a$`mean thermal_amp`        -20.32264    3.91910  -5.186 1.44e-06 ***
a$Days_above_mean_diff       -0.08380    0.01482  -5.653 2.06e-07 ***
a$fst_qtl_diff_year           4.41002    0.89170   4.946 3.78e-06 ***
a$trd_qtl_diff_year          10.52774    1.91638   5.494 4.03e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1685 on 85 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.868,	Adjusted R-squared:  0.8183 
F-statistic: 17.46 on 32 and 85 DF,  p-value: < 2.2e-16
```

Figure 41. Results of Regressions on CoNC for Region A and all crops

As we see in Figure 41, the model for region A reaches an R^2 of 0.86, which indicates that the weather variables explain in almost 87% the number of claims over contracts that occur. All the variables are more significant for region A when represented with a linear relationship. The means of the variables are the ones with the lowest coefficients leading to the conclusion that the higher the means, the lower the ratio of claims over the number of contracts.

Region B also reaches a value of 0.83, an excellent indicator that we can use the weather variables to explain the number of claims. For regions C, D, and E, we observe high significance values for the variables and the models as a whole.

In terms of the coefficient's values and signals, the main observations are that DATRt presents a negative coefficient on its quadratic variable, which represents that a continuous increase of the DATRt variable leads to a decrease in the CoNC. On the other hand, DATmax and DATw have a positive linear impact that leads to the increase of CoNC. The DUT variables have no significance for any of the models. For Minimum of Minimum and Maximum temperature, the impacts are mostly positive and linear, meaning that these variables increase claims over contracts.

The thermal amplitude has a negative linear impact, which indicates that the higher the difference between maximum and minimum, the more claims over contracts should be expected. It is believed that thermal amplitude has impact on the occurrence or frost, which is one of the main causes of hazards.

For CoNC, the first and third quantile of the variables were significant in most models, but their signals vary. The only common point is for the third quantile of maximum temperature that has a positive coefficient for all models, leading to more significant increases of claims. DAMRt and DAMmin have mostly negative coefficients which lead to lower claims as they increase, although for DAMRt, these conclusions are not always valid.

For the model that represents all regions and all crops, that can be accessed in the Appendix C, Figure 16, the best model had all rainfall-related variables linear. For almost all the variables with quadratic behaviours, we can observe negative values that indicate that CoNC decreases as the independent variables increase. For variables such as the first and third quantile of the maximum temperature, we identify the contrary behaviour. For the minimum thermal amplitude, an increase represents a positive growth in the number of claims. Almost all variables showed significance to explain, to a great extent, the number of claims over the number of contacts per year.

Considering an analysis per region we can see that, for region A, tariffs and claims over contracts seem to be aligned and share the same coefficient signals. This is true for

DATRt, DATmin and MinTmin. For the first two the BocP signals are inverse meaning that while the tariffs and CoNC increase with DATRt and DATmin, the BoCP decreases. For MaxTmin all the dependent variables are negatively impacted leading to decreases in tariffs, CoNC and bonus ratio. Here, it may happen that lower risk for the framers, that translates in lower tariffs, lowers the help the State may give.

An analysis for all other regions, show that, for all variables that are common for the models of tariffs and CoNC it is possible to verify, for most of the cases, that if a variable influence positively the tariffs it also influences positively the CoNC, for example. For BoCP there seems to exist a contrary behaviour where the ratio decreases if the other variables increase, for example. This could be explained by the fact that the Commercial Prize increases with tariffs leading to lower ratios and sometimes, for the sustainability of the aid system the State may have to reduce its participation, reducing the bonus while the Commercial Prize increases. We see this happening in the year of 1999, as exposed in section 3.1, where the State's aid went from 85% to almost 65%. This relationship is not in all cases and it may result from the weaker relationship between the bonus and the weather variables that weakens the evolution of the ratio when compared with the impact on the tariffs or number of claims.

4. Conclusions

The main goals of this dissertation were to relate weather data with agriculture-insurance data, in order to understand to which extend the frequency of hazards, the indemnities, the bonus of the Government and the insurance premiums, develop alongside with the climate evolution, reflecting the latest years of climate change and extreme phenomena intensification.

Through quality control, homogenization, and missing data infilling on of the data collected from IPMA, it was possible to create a dataset that led to reliable results in the analysis of the climate in Portugal in the last decades.

Parallel measurements analysis for manual and automatic stations, which have different instruments, was essential to understand how results can be extended from one period to the other. With this analysis we could observe that the instruments represent differences that are zero on average. Contrary to what was found in other studies, the standard deviation of these differences was high, which may be explained by the high thermal amplitudes and geographic position of Portugal.

A trend analysis on climate variables, such as temperature, rainfall and, wind speed were, performed and we verified an increase of the maximum temperature and a decrease of minimum temperatures alongside with the decrease of rainfall values. When looking at the whole period, from 1941 to 2018, the changes in the data did not look significant. However, in the last ten years there was a very clear intensification of the tendencies.

Regarding the analysis of the insurance data, the main outcome was that the different tariff regions and cultures have different associated evolutions and specificities, and such segmentation is important for the analysis of the results. The regressions led to the conclusion that the weather evolution is important to explain the definition of prices by the insurance companies and the frequency of claims. When it comes to the State aid, the weather is not as important. Also, the indemnities were the variable least explained by the climate variables through our models.

The variables related with Vineyard crop were less explained by weather phenomena than other crops due to its suitability to Mediterranean climate. The influence of each climate variable on each group of region and crop varied immensely, due to the big heterogeneity that each crop has in terms of growth ideal conditions.

Future studies should include more segmented crop groups, that have similar ideal growth conditions. A seasonal analysis should also provide a more realistic relationship between climate variables and agriculture-hazards.

References

[1] V. Pires, J. Marques, L. F. Nunes, T. Cota and L. Mendes, “CLIMA DE PORTUGAL CONTINENTAL - TENDÊNCIAS,” Instituto de Meteorologia, Lisbon.

[2] V. Pires, A. Silva and L. Mendes, "Risco de Secas em Portugal Continental," *Territorium,* vol. 17, pp. 27-34, 2010.

[3] J. Nascimento, L. Ribeiro, L. V. Cunha and O. R. P., "Impacto das Alterações Climáticas nos Recursos Hídricos Subterrâneos de Portugal Continental: Alguns Resultados Preliminares," Lisbon, Portugal, 2004.

[4] I. Lima, F. Santo, S. Cunha and Á. Silva, "Tendências nos extremos de precipitação em Portugal Continental em 1941-2012. Riscos Naturais Antrópicos e Mistos-Homenagem ao Professor Doutor Fernando Rebelo," Departamento de Geografia, Faculdade de Letras, Universidade de Coimbra, Coimbra, Portugal, 325-337, 2013.

[5] P. Valverde, R. Serralheiro, M. de Carvalho, R. Maia, B. Oliveira and V. Ramos, "Climate change impacts on irrigated agriculture in the Guadiana river basin (Portugal)," *Agricultural Water Management,* vol. 152, pp. 17-30, 2015.

[6] J. Chikaire, J. Echetama, R. Ihenacho, F. Nnadi, P. Umunnakwe and C. Utazi, "Agricultural insurance: A strategic tool for climate change adaptation in the agricultural sector," *Net Journal of Agricultural Science,* vol. 1, pp. 1-9, 2013.

[7] S. L. Jørgensen, M. Termansen and U. Pascual, "Natural insurance as condition for market insurance: Climate change adaptation in agriculture," *Ecological Economics,* vol. 169, no. 106489, 2020.

[8] D. Nadolnyak and V. Hartarska, “Agricultural disaster payments in the southeastern US: do weather and climate variablility matter?,” *Applied Economics,* vol. 44:33, pp. 4331-4342, 2012.

[9] V. Venema, B. Trewin, X. Wang, T. Szentimrey, M. Lakatos, E. Aguilar, I. Auer, J. A. Guijarro, M. Menne, C. Oria, W. S. R. L. Louamba and G. Rasul, "Guidance on the homogenisation of climate station data," 2018.

[10] F. Acquaotta and S. Fratianni, "THE IMPORTANCE OF THE QUALITY AND RELIABILITY OF THE HISTORICAL TIME SERIES FOR THE STUDY OF CLIMATE CHANGE," Revista Brasileira de Climatologia, 2014.

[11] O. Skrynyk, E. Aguilar, J. Guijarto, L. Y. A. Randriamarolaza and S. Bubin, "Uncertainty evaluation of Climatol’s adjustment algorithm applied to daily air temperature time series," *Internation Journal of Climatology,* vol. 41, pp. E2395-E2419, 2021.

[12] J. A. Guijarro, "User guide to climatol," State Metereological Agency,Ballearic Islands Office, Spain, 2011.

[13] J. A. Guijarro, "Homogenization of climatic series with Climatol," State Metereological Agency (AEMET), Balearic Islands Office,Spain, 2018.

[14] IPMA, "Extremos Climatológicos - Continente," [Online]. Available: https://www.ipma.pt/pt/oclima/extremos.clima/index.jsp?page=extreme_co.xml.

[15] M. Laapas and A. Venallainen, "Homogenization and trend analysis of monthly mean and maximum wind speed time series in Finland, 1959–2015," *International Journal of Climatology 37(14), 4803-4813,* 2017.

[16] S. Trilles, P. Juan, C. Diaz-Avalos, S. Ribeiro and M. Painho, "Reliability Evaluation of the Data Acquisition Potential of a Low-Cost Climatic Network for Applications in Agriculture," vol. 20(22), 2020.

[17] J. Guijarro, "Daily Series Homogenization and Gridding with Climatol v.3," 2017.

[18] S. Gluber, S. Hunziker, M. Begert, M. Croci-Maspoli, T.Konzelmann, S.Broennimann, C.Schwierz, C.Oria and G. Rosas, "The influence of station density on climate data homogenization," *International Journal of Climatology,* p. 37: 4670–4683, 2017.

[19] D. Maslyuk, "Simple linear regression in DAX-XXL BI," [Online]. Available: https://xxlbi.com/blog/simple-linear-regression-in-dax/. [Accessed january 2022].

[20] G. Beaumont, "Use Custom DAX to create a Linear Regression Trendline with a Dynamic Date Hierarchy in Power BI," Microsoft, [Online]. Available: https://techcommunity.microsoft.com/t5/healthcare-and-life-sciences/use-custom-dax-to-create-a-linear-regression-trendline-with-a/ba-p/2405245. [Accessed january 2022].

[21] L. Hannak, K. Friedrich, F. Imbery and F. Kaspar, "Comparison of manual and automatic daily sunshine duration measurements at German climate reference stations," *Advances in Science and Research,* vol. 16, pp. 175-183, 2019.

[22] F. Kaspar, L. Hannak and K.-J. Schreiber, "Climate reference stations in Germany:Status, parallel measurements and homogeneity of temperature time series," *Advances in Science and Research,* vol. 13, pp. 163-171, 2016.

[23] Á. J. Back, "Aplicação de análise estatística para identificação de tendências climáticas," Brasília, 2000.

[24] IFAP, "Ajudas Apoios," 2021. [Online]. Available: https://www.ifap.pt/web/guest/seguro1. [Accessed october 2021].

[25] IFAP , "Crop Insurances," 2021. [Online]. Available: https://www.ifap.pt/web/guest/sipac-contratacao. [Accessed october 2021].

[26] R. Costa, H. Fraga and J. A. Santos, "Climate change impacts on thermal growing conditions of main fruit species in Portugal," *Climatic Change,* vol. 140(2), pp. 273-286, 2016.

[27] B. Barnett, K. Coble and W. Steven, "Insurance, Developing and Pricing Precipitation," *Journal of Agricultural and Resource Economics,* vol. 26, pp. 261-274, 2001.

[28] J. R. S. a. M. A. M. Haiping Luo, "Insurance, Weather Information and the Potential for Intertemporal Adverse Selection in Crop," *Oxford Journals,* vol. 16, pp. 441-451, 1994.

[29] H. Fraga, I. G. d. C. Atauri, A. C. Malheiro, J. Moutinho-Pereira and J. A. Santos, "Viticulture in Portugal: A review of recent trends and climate change projections," *OENO One, Institut des Sciences de la Vigne et du Vin (Université de Bordeaux),* vol. 51 (2), pp. 61-69, 2017.

[30] R. Junqueira, J. d. S. Amorim and A. S. d. Oliveira, "COMPARAÇÃO ENTRE DIFERENTES METODOLOGIAS PARA PREENCHIMENTO DE FALHAS EM DADOS PLUVIOMÉTRICOS," *Sustentare,* vol. 2(1), pp. 198-210, 2018.

[31] R. K. Baba, M. S. M. G. Vaz and J. d. Costa, "CORREÇÃO DE DADOS AGROMETEOROLÓGICOS UTILIZANDO MÉTODOS ESTATÍSTICOS," *Revista Brasileira de Metereologia,* vol. 29, pp. 515-526, 2014.

[32] A. Mamara and A. Argiriou A. A., "Detection and correction of inhomogeneities in Greek climate," *International Journal of Climatology 34: 3024–3043,* 2014.

[33] K. P. Fante and J. L. S. Neto, "TÉCNICAS ESTATÍSTICAS PARA A HOMOGENEIZAÇÃO DE DADOS DE TEMPERATURA EM SÉRIES TEMPORAIS CLIMATOLÓGICAS.," vol. 18, 2016.

[34] D. Allcroft, C. Glasbey and M. Durban, "Modelling Weather Data," SCRI annual report, 2001.

[35] J. Coll, P. Domonkos, J. Guijarro, M. Curley, E. Rustemeier, E. Aguilar, S. Walsh and J. Sweeney, "Application of homogenization methods for Ireland's monthly precipitation records:Comparison of break detection results," vol. 40(14), pp. 6169-6188, 2020.

Appendix A

Data treatment, homogenization and quality control

Good quality data is not possible without some preliminary steps. First, we should analyse the raw data to understand which series and periods have well enough data to perform studies on them, [30]. There is the need to homogenize the climate time-series data by removing systematic biases. In [15], is identified that the changes that lead to breakpoints may occur due to non-climatic factors and it is the researchers' job to distinguish the climate phenomena resulting from human or equipment errors from true phenomena. In the end, we want to achieve comparability between data, in order to make correct conclusions, as mentioned in [9].

Since the changes in data may result from external factors, direct analysis on the raw data could be dangerous if not appropriately studied. The homogenization process is always recommended to study climate variables. Previous studies, such as [9], refer that this can be done in three steps: detection, adjustments, and validation. By applying all these we should be able to detect inhomogeneities, compare the stations with neighbour stations, giving more certainty to the results, and critically assess the work that has been done.

In [31], is suggested several validation rules for quality control on the data. The validation checks can be (i) basic, including limit, logic and per period validation; (ii) temporal validation; or (iii) spatial.

Although, long-term time series for climate data is the most accurate data recorded in the past, older datasets always bring several challenges and so metadata is extremely important to understand what happened over time, see [32]. Parallel measurements between old and new setups are advisable when starting the homogenization process (see [9]).

The weights of each reference series should reduce the white noise, the inhomogeneities, and respect the regional climate signal, see [9]. The theoretical minimum number of stations for statistical homogenization is considered three, but in practice, five is the value to achieve good results.

Some works, such as [9] and [18] suggest that we should have in mind that the distances from reference stations to the stations under analysis have impacts on the results. The density of the station's network depends on many factors, as for instance the size and development of a country.

In [33] and [10] is mentioned the necessity to overcome other challenges such as the missing data for stations' time series. Some software, such as CLIMATOL, include missing data infilling in their homogenization algorithms.

In the work [32], it is described that choosing the best method for data treatment is a subjective choice of the researchers. There are many methods and inter-comparisons of the techniques that we can study to decide which process best suits their work. Most techniques are recommended generally for annual and monthly data because daily data presents more statistical bias, as mentioned in [33]. In the research in [33] and [30], the suggested approaches were the *arithmetic average*; *Regional Weighting;* the *Simple Linear Regression* and the *Multiple Linear Regression*; the *Inverse Distance Weighting (IDW)*. An important conclusion from those comparisons was that the bigger the datasets, the better the results for any technique.

The methods presented can be applied manually or automatically and in [9] is highlighted those manual methods as being more labour-intensive and demanded of a more experienced user. The latest developments of automatic methods increase efficiency while decrease the chances for human errors.

CLIMATOL: Software for climate data treatment

Specificities about the weather variables have to be taken into account when choosing the best tool for the data homogenization and missing data infilling, see [34]. The study of [32] describes the topic of statistical packages and software that do homogenization and filling of missing data. These methods were studied by a coordinated European initiative, which assessed their validity. Software such as HOMER, MASH, ACMANT, PRODIGE, and CLIMATOL are mentioned and compared to see which one fits better the homogenization task.

The homogenization processes can be performed based on statistical testing or using numerical studies, as mentioned in [9]. When assessing statistical testing, we can choose several types of tests: (i) the t-test; (ii) the Standard Normalised Homogeneity Test, SNHT, which is used by CLIMATOL which is the software used in this work; (iii) and Penalized Maximal T-test, PMT.

In [11], it is done a direct comparison between methods, and the main conclusion was that the tools differed on the ratios of homogeneous series. Different breaking points are detected according to the different algorithms and strategies the tools intrinsically use. CLIMATOL showed to be the software that approximate better the actual scenario. In addition, the CLIMATOL tool is the most suitable for several types of weather variables, being the tool with most support material available as well as more user-friendly interface.

In the work [13], the creator of the CLIMATOL package described the tool as being able to “provide functions to facilitate the homogenization of climatological variables at any temporal scale”. The R package of CLIMATOL contains quality control functions, homogenization, and infilling of missing data.

CLIMATOL allows for resolution in daily data, which is the type of data we use in this work, by using composite reference data. The primary operations are automatic. Authors such as [9] established the tool’s good results and accuracy. One of its best specificities is that it can be used automatically, [11], while handling mid-size networks that go up to 100 stations network, while removing several types of errors.

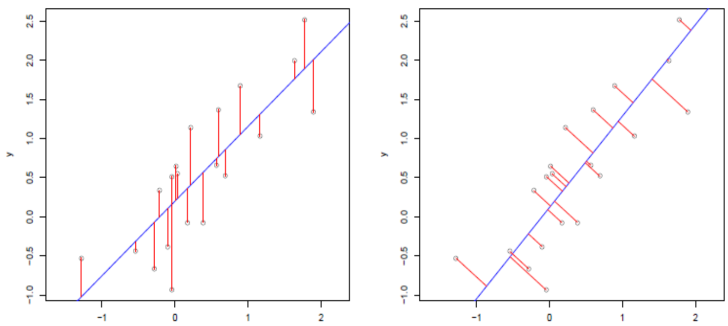

Figure 1. Type I and Type II Regression. Source: Guijarro [13]

CLIMATOL uses a type II regression (RMA) instead of type I regression for the homogenization. The orthogonal regression (RMA) minimizes the perpendicular distance of the scatter points to the linear regression line instead of the vertical distance, as illustrated in Figure 1. Also, in CLIMATOL tool, both the dependent and independent variables have been standardized. It is possible to use several reference data for the same point and the weights of each reference data are defined according to the distance to the candidate series, [12].

The method used by CLIMATOL for missing data, allows for flexibility by using nearby data while adapting to the different availability of information in stations nearby.

After estimating all the data, the following step is outlier and shift detection and correction. The outliers correspond to points for which the anomalies are greater than five, by default, and the value used in this work and above which outliers are deleted. The SNHT, Standard Normalised Homogeneity Test, is mostly used for series with one breakpoint but with unknown dates, it is a likelihood test performed on the ratios or differences between the data that is calculated for and the reference series, [10]. The maximum values of SNHT are stored, as well as their locations, and when the statistic series of SNHT is higher than a certain threshold, the series is split at that point, creating a new series with the same coordinates, [12].

The same happens with the threshold that rejects anomalous data, set as five. This value should be set up to at least twenty when dealing with daily data, especially for precipitation, because of its significant variability. The last step of the process is devoted to recalculating the missing data, including the data that was deleted in the process. The process that unrolls in CLIMATOL is summarised in the flow-chart in Figure 2.

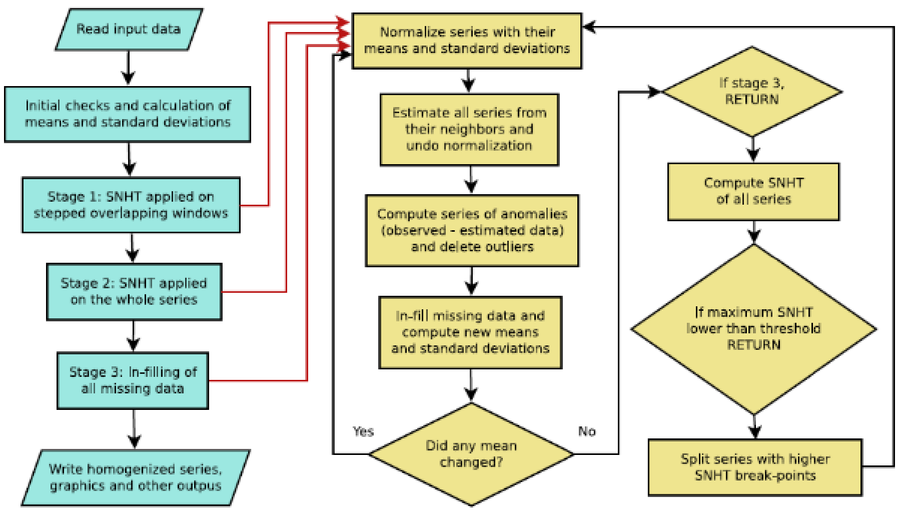

Figure 2. Flowchart of Climatol software. Source: Guijarro [13]

Although the process is similar, we have to consider the specificities of dealing with daily data, which is known to have a higher variability, [17]. Thus, it is better to homogenize monthly data and use the metadata for daily homogenization In CLIMATOL this is done by using the breakpoints from the monthly data into the daily series.

In [11] is referred that the assessment of the certainty on the results of a homogenization process, although very important, is many times overlooked. Understanding the proper signal of a climate variable is not a straightforward procedure, however CLIMATOL has shown to be good in removing most of the trend errors, which leads to more solid results on the trend analysis.

Appendix B

Tariff Region	Region	Counties
A	Faro; Lisbon; Setubal	ALBUFEIRA; ALCOUTIM; ALJEZUR; CASTRO MARIM; FARO; LAGOA; LAGOS; MONCHIQUE; LOULE; OLHAO; PORTIMAO; SAO BRAS DE ALPORTEL; SILVES; VILA DO BISPO; TAVIRA; VILA DO BISPO; VILA REAL DE SANTO ANTONIO; AMADORA; CASCAIS; LISBOA; LOURES; LOURINHA; MAFRA; ODIVELAS; OEIRAS; SINTRA; TORRES VEDRAS; ALMADA; SEIXAL; SESIMBRA; SETUBAL
B	Sesimbra; Santarém; Setúbal; Viana do Castelo	AVEIRO; ESPINHO; ESTARREJA;SANTA MARIA DA FEIRA;ILHAVO;MURTOSA;OLIVEIRA DE AZEMEIS;OVAR; SAO JOAO DA MADEIRA;VAGOS; ODEMIRA; ESPOSENDE; FIGUEIRA DA FOZ;MIRA;MONTEMOR O VELHO;SOURE; ALCOBACA;BOMBARRAL;CALDAS DA RAINHA;LEIRIA;MARINHA GRANDE;NAZARE;OBIDOS;PENICHE; POMBAL;PORTO DE MOS; ALENQUER;ARRUDA DOS VINHOS;CADAVAL;SOBRAL DE MONTE AGRACO;VILA FRANCA DE XIRA; MAIA;MATOSINHOS;PORTO;POVOA DE VARZIM; VILA DO CONDE; VILA NOVA DE GAIA ; RIO MAIOR;AZAMBUJA; ALCACER DO SAL; ALCOCHETE; BARREIRO; GRANDOLA; MOITA; MONTIJO; PALMELA;S ANTIAGO DO CACEM; SINES; CAMINHA;VIANA DO CASTELO
C	Setubal; Santarem;	ALJUSTREL; ALMODOVAR; ALVITO; BARRANCOS; BEJA; CASTRO VERDE; CUBA; FERREIRA DO ALENTEJO; MERTOLA; MOURA; OURIQUE; SERPA;VIDIGUEIRA; ALANDROAL; ARRAIOLOS; BORBA; ESTREMOZ; EVORA; MONTEMOR O NOVO; MORA; MOURAO; PORTEL; REDONDO; REGUENGOS DE MONSARAZ; VENDAS NOVAS; VIANA DO ALENTEJO; VILA VICOSA; BATALHA; ALTER DO CHAO; ARRONCHES; AVIZ; CAMPO MAIOR; CASTELO DE VIDE; CRATO; ELVAS ;FRONTEIRA; GAVIAO; MARVAO; MONFORTE; NISA; PONTE DE SOR; PORTALEGRE; SOUSEL; ALCANENA; ALMEIRIM; ALPIARCA; BENAVENTE; CARTAXO; CHAMUSCA; CONSTANCIA; CORUCHE; ENTRONCAMENTO; GOLEGA; OUREM; SALVATERRA DE MAGOS; SANTAREM; TORRES NOVAS; VILA NOVA DA BARQUINHA
D	Aveiro; Braga; Bragança; Castelo Branco; Coimbra; Setubal do Castelo; Vila Real; Viseu	ALBERGARIA A VELHA; ANADIA; AROUCA; AGUEDA; CASTELO DE PAIVA; MEALHADA; OLIVEIRA DO BAIRRO; SEVER DO VOUGA; VALE DE CAMBRA; AMARES; BARCELOS; BRAGA; CABECEIRAS DE BASTO; CELORICO DE BASTO; FAFE; GUIMARAES; POVOA DE LANHOSO; TERRAS DE BOURO; VIEIRA DO MINHO; VILA NOVA DE FAMALICAO; VILA VERDE; VIZELA; ALFANDEGA DA FE; MIRANDELA ;BELMONTE; CASTELO BRANCO; IDANHA A NOVA; OLEIROS; PENAMACOR; PROENCA A NOVA; SERTA; VILA DE REI; VILA VELHA DE RODAO; ARGANIL; CANTANHEDE; COIMBRA; CONDEIXA A NOVA; GOIS; LOUSA; MIRANDA DO CORVO; PAMPILHOSA DA SERRA; PENACOVA; PENELA; TABUA; VILA NOVA DE POIARES; GOUVEIA; MEDA; SABUGAL; SEIA; VILA NOVA DE FOZ COA; ALVAIAZERE; ANSIAO; CASTANHEIRA DE PERA; FIGUEIRO DOS VINHOS; PEDROGAO GRANDE; AMARANTE; BAIAO;FELGUEIRAS; GONDOMAR;LOUSADA;MARCO DE CANAVESES;PACOS DE FERREIRA;PAREDES;PENAFIEL;SANTO TIRSO; TROFA;VALONGO; ABRANTES; FERREIRA DO ZEZERE; MACAO;SARDOAL; TOMAR; ARCOS DE VALDEVEZ; MELGACO; MONCAO; PAREDES DE COURA; PONTE DA BARCA; PONTE DE LIMA; VALENCA; VILA NOVA DE CERVEIRA; MESAO FRIO; MONDIM DE BASTO; PESO DA REGUA;SANTA MARTA DE PENAGUIAO;VALPACOS; ARMAMAR; CARREGAL DO SAL; CINFAES;MORTAGUA; NELAS;LIVEIRA DE FRADES; RESENDE; SANTA COMBA DAO; SAO JOAO DA PESQUEIRA; SAO PEDRO DO SUL
E	Bragança; Guarda; Vila Real; Viseu; Castelo Branco; Coimbra;	BRAGANCA; CARRAZEDA DE ANSIAES; FREIXO DE ESPADA A CINTA; MACEDO DE CAVALEIROS; MIRANDA DO DOURO; MOGADOURO; TORRE DE MONCORVO; VIMIOSO; VINHAIS; AGUIAR DA BEIRA; ALMEIDA;CELORICO DA BEIRA;FIGUEIRA CASTELO RODRIGO; FORNOS DE ALGODRES; GUARDA; MANTEIGAS; PINHEL; TRANCOSO; ALIJO; BOTICAS; CHAVES; MONTALEGRE; MURCA; RIBEIRA DE PENA; SABROSA; VILA POUCA DE AGUIAR; VILA REAL; CASTRO DAIRE; MOIMENTA DA BEIRA; PENALVA DO CASTELO; PENEDONO; SATAO; SERNANCELHE; TAROUCA;VILA NOVA DE PAIVA; CARRAZEDA DE ANSIAES; VILA FLOR; CARRAZEDA DE ANSIAES; VILA FLOR; COVILHA; FUNDAO; OLIVEIRA DO HOSPITAL; AGUIAR DA BEIRA; ALMEIDA; CELORICO DA BEIRA; FORNOS DE ALGODRES; GUARDA; PINHEL; TRANCOSO; ALIJO; CHAVES; MURCA; SABROSA; VILA REAL; AROUCA; TONDELA; VILA NOVA DE PAIVA; VISEU; VOUZELA; CASTRO DAIRE; LAMEGO; MANGUALDE; MOIMENTA DA BEIRA ; PENALVA DO CASTELO; PENEDONO; SATAO; SERNANCELHE; TABUACO

Table 1. Counties and Tariff Regions

Regions	Meteorological Stations Number
A	535; 739; 740; 746; 770; 865; 867; 869
B	531; 702; 713; 718; 720; 726; 729; 742; 766; 767; 776; 783
C	558; 562; 571; 734; 744; 824; 826; 35;837; 840; 847; 848;850; 863; 864
D	549; 570; 605; 622; 630; 632; 655; 657; 668; 685; 697; 705; 707; 716; 724; 800;803; 06;812
E	566; 616; 619; 644; 651; 663; 666; 671; 680; 683; 687; 690; 698

Table 2. Tariff Regions and Meteorological Stations

Appendix C

```
lm(formula = a$Tariffs ~ a$Days_above_treshold_Rtotal + a$`max(Rtotal)` +
    a$Days_above_treshold_Tmin + poly(a$Days_under_treshold_Tmin,
    2) + a$`min(Tmin)` + a$`max(Tmin)` + a$Days_above_treshold_Tmax +
    poly(a$Days_under_treshold_Tmax, 2) + a$`min(Tmax)` + poly(a$Days_above_treshold_Wspeed,
    2) + a$`min(Wspeed)`)

Residuals:
      Min        1Q    Median        3Q       Max
-0.062628 -0.013790 -0.001539  0.011573  0.048460

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                              5.864e-02  3.061e-02   1.916 0.057393 .
a$Days_above_treshold_Rtotal            -1.828e-04  1.185e-04  -1.543 0.125030
a$`max(Rtotal)`                         -2.179e-04  1.236e-04  -1.763 0.080066 .
a$Days_above_treshold_Tmin               6.383e-05  1.335e-04   0.478 0.633163
poly(a$Days_under_treshold_Tmin, 2)1     1.353e-01  3.989e-02   3.392 0.000896 ***
poly(a$Days_under_treshold_Tmin, 2)2    -4.695e-02  2.910e-02  -1.613 0.108811
a$`min(Tmin)`                           -1.659e-03  1.127e-03  -1.472 0.143130
a$`max(Tmin)`                            2.640e-03  1.242e-03   2.125 0.035279 *
a$Days_above_treshold_Tmax              -9.602e-04  1.198e-04  -8.014 3.35e-13 ***
poly(a$Days_under_treshold_Tmax, 2)1    -2.259e-01  4.714e-02  -4.792 4.04e-06 ***
poly(a$Days_under_treshold_Tmax, 2)2     4.877e-02  2.933e-02   1.663 0.098528 .
a$`min(Tmax)`                            2.015e-03  1.082e-03   1.863 0.064526 .
poly(a$Days_above_treshold_Wspeed, 2)1   3.689e-02  3.408e-02   1.082 0.280841
poly(a$Days_above_treshold_Wspeed, 2)2  -5.663e-02  2.655e-02  -2.133 0.034642 *
a$`min(Wspeed)`                          8.588e-03  3.805e-03   2.257 0.025496 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0197 on 145 degrees of freedom
Multiple R-squared:  0.5297,    Adjusted R-squared:  0.4843
F-statistic: 11.67 on 14 and 145 DF,  p-value: < 2.2e-16
```

Figure 1. Results of Regressions on Tariffs for Region C and Vineyards

```
Call:
lm(formula = a$Tariffs ~ a$Days_above_treshold_Rtotal + poly(a$Days_above_treshold_Tmin,
    2) + poly(a$`max(Tmin)`, 2) + a$Days_above_treshold_Tmax +
    a$`min(Tmax)` + a$Days_under_treshold_Wspeed + poly(a$`min(Wspeed)`,
    2) + poly(a$`max(Wspeed)`, 2))

Residuals:
      Min        1Q    Median        3Q       Max
-0.073055 -0.010878  0.001726  0.013510  0.063354

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                              9.827e-02  1.076e-02   9.137  < 2e-16 ***
a$Days_above_treshold_Rtotal             1.122e-04  6.347e-05   1.767 0.078198 .
poly(a$Days_above_treshold_Tmin, 2)1    -9.016e-02  4.704e-02  -1.917 0.056204 .
poly(a$Days_above_treshold_Tmin, 2)2    -1.587e-01  3.292e-02  -4.822 2.24e-06 ***
poly(a$`max(Tmin)`, 2)1                  2.150e-02  3.752e-02   0.573 0.567039
poly(a$`max(Tmin)`, 2)2                  1.022e-01  2.983e-02   3.427 0.000693 ***
a$Days_above_treshold_Tmax              -1.483e-04  7.555e-05  -1.963 0.050560 .
a$`min(Tmax)`                            1.275e-03  5.959e-04   2.140 0.033166 *
a$Days_under_treshold_Wspeed            -1.560e-04  3.894e-05  -4.006 7.75e-05 ***
poly(a$`min(Wspeed)`, 2)1               -1.090e-02  3.926e-02  -0.278 0.781481
poly(a$`min(Wspeed)`, 2)2                3.942e-02  2.760e-02   1.428 0.154226
poly(a$`max(Wspeed)`, 2)1               -6.799e-02  4.219e-02  -1.611 0.108144
poly(a$`max(Wspeed)`, 2)2               -6.771e-02  2.925e-02  -2.315 0.021254 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02339 on 307 degrees of freedom
Multiple R-squared:  0.2911,    Adjusted R-squared:  0.2634
F-statistic: 10.51 on 12 and 307 DF,  p-value: < 2.2e-16
```

Figure 2. Results of Regressions on Tariffs for Region D and Vineyards

```
Call:
lm(formula = a$Tariffs ~ a$Days_above_treshold_Rtotal + poly(a$`max(Rtotal)`,
    2) + poly(a$Days_above_treshold_Tmin, 2) + poly(a$`min(Tmin)`,
    2) + poly(a$`max(Tmin)`, 2) + a$Days_above_treshold_Tmax +
    poly(a$Days_under_treshold_Tmax, 2) + a$Days_above_treshold_Wspeed +
    poly(a$Days_under_treshold_Wspeed, 2))

Residuals:
      Min        1Q    Median        3Q       Max
-0.091953 -0.032379  0.001281  0.031618  0.109015

Coefficients:
                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                            0.0547338  0.0137213   3.989 7.94e-05 ***
a$Days_above_treshold_Rtotal           0.0010513  0.0001166   9.017  < 2e-16 ***
poly(a$`max(Rtotal)`, 2)1             -0.1482020  0.0478229  -3.099 0.002084 **
poly(a$`max(Rtotal)`, 2)2             -0.1335012  0.0489186  -2.729 0.006642 **
poly(a$Days_above_treshold_Tmin, 2)1  -0.1609820  0.0721495  -2.231 0.026238 *
poly(a$Days_above_treshold_Tmin, 2)2  -0.2168560  0.0505768  -4.288 2.28e-05 ***
poly(a$`min(Tmin)`, 2)1               -0.1547670  0.0487268  -3.176 0.001612 **
poly(a$`min(Tmin)`, 2)2                0.1578225  0.0476926   3.309 0.001023 **
poly(a$`max(Tmin)`, 2)1                0.0008840  0.0625044   0.014 0.988723
poly(a$`max(Tmin)`, 2)2                0.2310931  0.0563698   4.100 5.05e-05 ***
a$Days_above_treshold_Tmax             0.0003597  0.0001294   2.779 0.005712 **
poly(a$Days_under_treshold_Tmax, 2)1  -0.1486520  0.0740792  -2.007 0.045481 *
poly(a$Days_under_treshold_Tmax, 2)2  -0.1798830  0.0530197  -3.393 0.000763 ***
a$Days_above_treshold_Wspeed          -0.0003910  0.0001066  -3.668 0.000279 ***
poly(a$Days_under_treshold_Wspeed, 2)1 -0.7826065  0.0836398  -9.357  < 2e-16 ***
poly(a$Days_under_treshold_Wspeed, 2)2  0.2286184  0.0435509   5.249 2.52e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0395 on 387 degrees of freedom
Multiple R-squared:  0.3887,    Adjusted R-squared:  0.365
F-statistic:  16.4 on 15 and 387 DF,  p-value: < 2.2e-16
```

Figure 3. Results of Regressions on Tariffs for Region E and Vineyards

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.178474 -0.043061 -0.004166  0.041174  0.207594

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                             4.148e-01  4.534e-02   9.150  < 2e-16 ***
poly(a$Days_above_treshold_Rtotal, 2)1 -2.782e-01  8.468e-02  -3.286 0.001046 **
poly(a$Days_above_treshold_Rtotal, 2)2  2.327e-01  6.955e-02   3.345 0.000846 ***
poly(a$`max(Rtotal)`, 2)1               3.803e-01  6.694e-02   5.681 1.66e-08 ***
poly(a$`max(Rtotal)`, 2)2              -2.308e-01  6.806e-02  -3.391 0.000717 ***
poly(a$Days_above_treshold_Tmin, 2)1   -2.260e+00  1.356e-01 -16.659  < 2e-16 ***
poly(a$Days_above_treshold_Tmin, 2)2    4.887e-01  8.376e-02   5.834 6.85e-09 ***
poly(a$Days_under_treshold_Tmin, 2)1   -1.527e+00  1.691e-01  -9.035  < 2e-16 ***
poly(a$Days_under_treshold_Tmin, 2)2   -2.580e-01  9.818e-02  -2.628 0.008687 **
poly(a$`max(Tmin)`, 2)1                 3.519e-01  1.140e-01   3.086 0.002069 **
poly(a$`max(Tmin)`, 2)2                -3.371e-01  7.282e-02  -4.628 4.07e-06 ***
poly(a$Days_above_treshold_Tmax, 2)1    9.887e-01  1.177e-01   8.398  < 2e-16 ***
poly(a$Days_above_treshold_Tmax, 2)2   -4.056e-01  8.091e-02  -5.013 6.13e-07 ***
poly(a$Days_under_treshold_Tmax, 2)1    2.304e+00  1.561e-01  14.765  < 2e-16 ***
poly(a$Days_under_treshold_Tmax, 2)2    1.023e-01  9.541e-02   1.072 0.283936
a$`max(Tmax)`                          -7.933e-03  1.162e-03  -6.829 1.32e-11 ***
a$Days_above_treshold_Wspeed            1.895e-04  5.021e-05   3.773 0.000168 ***
a$Days_under_treshold_Wspeed            7.022e-04  6.935e-05  10.127  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06054 on 1270 degrees of freedom
  (8 observations deleted due to missingness)
Multiple R-squared:  0.6594,    Adjusted R-squared:  0.6549
F-statistic: 144.6 on 17 and 1270 DF,  p-value: < 2.2e-16
```

Figure 4. Results of Regressions on Tariffs for all Regions and all crops except Vineyards. Y-1 weather variables

```
Call:
lm(formula = a$Tariffs ~ poly(a$Days_above_treshold_Rtotal, 2) +
    poly(a$`max(Rtotal)`, 2) + poly(a$Days_above_treshold_Tmin,
    2) + poly(a$Days_under_treshold_Tmin, 2) + poly(a$`max(Tmin)`,
    2) + poly(a$Days_above_treshold_Tmax, 2) + poly(a$Days_under_treshold_Tmax,
    2) + a$`max(Tmax)` + a$Days_above_treshold_Wspeed + a$Days_under_treshold_Wspeed)

Residuals:
      Min        1Q    Median        3Q       Max
-0.167921 -0.043174  0.000174  0.042214  0.141764

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                             0.2517976  0.1013282   2.485  0.01341 *
poly(a$Days_above_treshold_Rtotal, 2)1  0.1019874  0.1187463   0.859  0.39098
poly(a$Days_above_treshold_Rtotal, 2)2 -0.0856456  0.0990191  -0.865  0.38764
poly(a$`max(Rtotal)`, 2)1               0.2203303  0.0812814   2.711  0.00703 **
poly(a$`max(Rtotal)`, 2)2              -0.1152932  0.0889428  -1.296  0.19571
poly(a$Days_above_treshold_Tmin, 2)1   -0.4300387  0.1844562  -2.331  0.02028 *
poly(a$Days_above_treshold_Tmin, 2)2    0.6674833  0.0969827   6.883 2.59e-11 ***
poly(a$Days_under_treshold_Tmin, 2)1    0.1003166  0.1836174   0.546  0.58517
poly(a$Days_under_treshold_Tmin, 2)2    0.0954451  0.1186432   0.804  0.42165
poly(a$`max(Tmin)`, 2)1                 0.3500684  0.1666605   2.100  0.03638 *
poly(a$`max(Tmin)`, 2)2                 0.0158122  0.0991252   0.160  0.87335
poly(a$Days_above_treshold_Tmax, 2)1    0.4203225  0.1520595   2.764  0.00600 **
poly(a$Days_above_treshold_Tmax, 2)2   -0.2500885  0.1178966  -2.121  0.03458 *
poly(a$Days_under_treshold_Tmax, 2)1    0.1272636  0.1822242   0.698  0.48538
poly(a$Days_under_treshold_Tmax, 2)2    0.2177360  0.1183348   1.840  0.06658 .
a$`max(Tmax)`                          -0.0015298  0.0025781  -0.593  0.55330
a$Days_above_treshold_Wspeed           -0.0001957  0.0000661  -2.961  0.00326 **
a$Days_under_treshold_Wspeed           -0.0002055  0.0001350  -1.522  0.12883
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06394 on 363 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.3127,    Adjusted R-squared:  0.2805
F-statistic: 9.716 on 17 and 363 DF,  p-value: < 2.2e-16
```

Figure 5. Results of Regressions on Tariffs for Region D and all crops. Y-1 weather variables

```
Call:
lm(formula = a$Tariffs ~ a$Days_above_treshold_Rtotal + a$`max(Rtotal)` +
    a$Days_above_treshold_Tmin + a$Days_under_treshold_Tmin +
    a$`min(Tmin)` + a$`max(Tmin)` + a$Days_above_treshold_Tmax +
    a$Days_under_treshold_Tmax + a$`min(Tmax)` + a$`max(Tmax)` +
    a$Days_above_treshold_Wspeed + a$Days_under_treshold_Wspeed +
    a$`min(Wspeed)` + a$`max(Wspeed)`)

Residuals:
     Min       1Q   Median       3Q      Max
-0.20243 -0.03675  0.02187  0.04451  0.23963

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                   1.012e-01  5.115e-02   1.978  0.04816 *
a$Days_above_treshold_Rtotal  6.332e-04  1.204e-04   5.257 1.80e-07 ***
a$`max(Rtotal)`               1.232e-04  1.254e-04   0.983  0.32602
a$Days_above_treshold_Tmin   -7.335e-04  2.473e-04  -2.966  0.00309 **
a$Days_under_treshold_Tmin    3.484e-04  1.754e-04   1.987  0.04720 *
a$`min(Tmin)`                -4.939e-04  1.481e-03  -0.334  0.73883
a$`max(Tmin)`                 1.740e-03  1.767e-03   0.985  0.32508
a$Days_above_treshold_Tmax    2.248e-04  1.883e-04   1.194  0.23283
a$Days_under_treshold_Tmax   -4.391e-04  1.714e-04  -2.562  0.01057 *
a$`min(Tmax)`                 8.691e-04  1.464e-03   0.594  0.55278
a$`max(Tmax)`                 1.720e-03  1.510e-03   1.139  0.25480
a$Days_above_treshold_Wspeed -6.408e-05  1.401e-04  -0.457  0.64743
a$Days_under_treshold_Wspeed -2.647e-04  1.204e-04  -2.199  0.02812 *
a$`min(Wspeed)`               1.397e-02  2.669e-03   5.234 2.04e-07 ***
a$`max(Wspeed)`              -1.039e-03  7.080e-04  -1.468  0.14251
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06572 on 958 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.08582,   Adjusted R-squared:  0.07246
F-statistic: 6.424 on 14 and 958 DF,  p-value: 1.739e-12
```

Figure 6. Results of Regressions on Tariffs for Region E and all crops. Y-1 weather variables

```
Call:
lm(formula = a$Tariffs ~ poly(a$Days_above_treshold_Rtotal, 2) +
    poly(a$`max(Rtotal)`, 2) + a$Days_above_treshold_Tmin + poly(a$Days_under_treshold_Tmin,
    2) + a$`max(Tmin)` + poly(a$Days_above_treshold_Tmax, 2) +
    poly(a$Days_under_treshold_Tmax, 2) + poly(a$Days_under_treshold_Wspeed,
    2) + a$`max(Wspeed)`)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0231355 -0.0042315  0.0008879  0.0050620  0.0240901

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                              7.351e-02  1.989e-02   3.695 0.000419 ***
poly(a$Days_above_treshold_Rtotal, 2)1  3.093e-02  1.715e-02   1.803 0.075390 .
poly(a$Days_above_treshold_Rtotal, 2)2 -4.846e-02  1.731e-02  -2.800 0.006508 **
poly(a$`max(Rtotal)`, 2)1              -3.242e-02  1.338e-02  -2.423 0.017833 *
poly(a$`max(Rtotal)`, 2)2              -5.691e-02  1.670e-02  -3.408 0.001061 **
a$Days_above_treshold_Tmin              1.262e-04  8.376e-05   1.507 0.136087
poly(a$Days_under_treshold_Tmin, 2)1   -3.390e-02  1.799e-02  -1.884 0.063488 .
poly(a$Days_under_treshold_Tmin, 2)2    1.150e-02  1.695e-02   0.679 0.499533
a$`max(Tmin)`                          -5.829e-05  1.131e-03  -0.052 0.959042
poly(a$Days_above_treshold_Tmax, 2)1   -3.081e-02  1.346e-02  -2.289 0.024949 *
poly(a$Days_above_treshold_Tmax, 2)2   -1.955e-02  1.498e-02  -1.305 0.196010
poly(a$Days_under_treshold_Tmax, 2)1   -1.029e-01  1.625e-02  -6.333 1.67e-08 ***
poly(a$Days_under_treshold_Tmax, 2)2    8.716e-02  1.557e-02   5.599 3.47e-07 ***
poly(a$Days_under_treshold_Wspeed, 2)1  6.619e-02  1.836e-02   3.606 0.000562 ***
poly(a$Days_under_treshold_Wspeed, 2)2 -5.743e-02  1.427e-02  -4.025 0.000136 ***
a$`max(Wspeed)`                        -1.433e-03  5.282e-04  -2.713 0.008292 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.009755 on 74 degrees of freedom
Multiple R-squared:  0.7667,    Adjusted R-squared:  0.7194
F-statistic: 16.21 on 15 and 74 DF,  p-value: < 2.2e-16
```

Figure 7. Results of Regressions on Tariffs for Region B and all crops except Vineyards. Y-1 weather variables

```
Call:
lm(formula = a$Tariffs ~ poly(a$Days_above_treshold_Rtotal, 2) +
    a$`max(Rtotal)` + poly(a$Days_above_mean_Rtotal, 2) + a$fst_qtl_Rtotal_year +
    poly(a$trd_qtl_Rtotal_year, 2) + poly(a$Days_above_treshold_Tmin,
    2) + poly(a$`min(Tmin)`, 2) + poly(a$`max(Tmin)`, 2) + poly(a$`mean Tmin`,
    2) + poly(a$fst_qtl_Tmin_year, 2) + poly(a$trd_qtl_Tmin_year,
    2) + poly(a$Days_above_treshold_Tmax, 2) + poly(a$`min(Tmax)`,
    2) + poly(a$`max(Tmax)`, 2) + poly(a$fst_qtl_Tmax_year, 2) +
    a$trd_qtl_Tmax_year + poly(a$Days_above_treshold_Wspeed,
    2) + poly(a$`min(Wspeed)`, 2) + a$`max(Wspeed)` + a$`mean Wspeed` +
    a$Days_above_mean_Wspeed + a$fst_qtl_Wspeed_year + a$trd_qtl_Wspeed_year +
    poly(a$Days_above_treshold_diff, 2) + poly(a$`min thermal_amp`,
    2) + a$`max thermal_amp` + poly(a$`mean thermal_amp`, 2) +
    poly(a$fst_qtl_diff_year, 2) + poly(a$trd_qtl_diff_year,
    2))

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                              1.510e-01  5.897e-02   2.560 0.010499 *
poly(a$Days_above_treshold_Rtotal, 2)1  3.129e-01  3.028e-01   1.033 0.301585
poly(a$Days_above_treshold_Rtotal, 2)2  7.330e-01  2.008e-01   3.650 0.000266 ***
a$`max(Rtotal)`                          3.637e-04  5.778e-05   6.295 3.44e-10 ***
poly(a$Days_above_mean_Rtotal, 2)1     -3.243e-01  3.028e-01  -1.071 0.284154
poly(a$Days_above_mean_Rtotal, 2)2     -4.037e-01  2.045e-01  -1.974 0.048489 *
a$fst_qtl_Rtotal_year                    5.484e-01  1.564e+00   0.351 0.725948
poly(a$trd_qtl_Rtotal_year, 2)1        -4.158e-01  9.896e-02  -4.202 2.71e-05 ***
poly(a$trd_qtl_Rtotal_year, 2)2         2.782e-01  1.250e-01   2.225 0.026159 *
poly(a$Days_above_treshold_Tmin, 2)1   -1.465e+00  2.915e-01  -5.025 5.29e-07 ***
poly(a$Days_above_treshold_Tmin, 2)2    5.022e-01  1.131e-01   4.440 9.27e-06 ***
poly(a$`min(Tmin)`, 2)1                 2.695e-01  1.353e-01   1.993 0.046372 *
poly(a$`min(Tmin)`, 2)2                 1.954e-01  1.039e-01   1.880 0.060173 .
poly(a$`max(Tmin)`, 2)1                 6.867e-01  1.793e-01   3.830 0.000131 ***
poly(a$`max(Tmin)`, 2)2                -3.669e-01  1.831e-01  -2.004 0.045139 *
poly(a$`mean Tmin`, 2)1                -2.637e+00  9.328e-01  -2.827 0.004723 **
poly(a$`mean Tmin`, 2)2                -2.198e+00  5.326e-01  -4.126 3.77e-05 ***
poly(a$fst_qtl_Tmin_year, 2)1           2.321e+00  4.743e-01   4.893 1.04e-06 ***
poly(a$fst_qtl_Tmin_year, 2)2           7.539e-01  2.632e-01   2.865 0.004196 **
poly(a$trd_qtl_Tmin_year, 2)1           5.357e-01  5.330e-01   1.005 0.314971
poly(a$trd_qtl_Tmin_year, 2)2           1.173e+00  4.229e-01   2.774 0.005572 **
poly(a$Days_above_treshold_Tmax, 2)1   -9.495e-01  2.716e-01  -3.496 0.000478 ***
poly(a$Days_above_treshold_Tmax, 2)2   -4.819e-01  1.108e-01  -4.349 1.40e-05 ***
poly(a$`min(Tmax)`, 2)1                -1.400e+00  1.458e-01  -9.605  < 2e-16 ***
poly(a$`min(Tmax)`, 2)2                -8.207e-01  9.325e-02  -8.802  < 2e-16 ***
poly(a$`max(Tmax)`, 2)1                 1.240e-01  2.097e-01   0.591 0.554508
poly(a$`max(Tmax)`, 2)2                -8.175e-01  1.550e-01  -5.273 1.42e-07 ***
poly(a$fst_qtl_Tmax_year, 2)1          -1.674e+00  2.882e-01  -5.808 6.86e-09 ***
poly(a$fst_qtl_Tmax_year, 2)2           1.014e+00  1.113e-01   9.117  < 2e-16 ***
a$trd_qtl_Tmax_year                      2.343e-03  1.772e-03   1.322 0.186202
poly(a$Days_above_treshold_Wspeed, 2)1  6.138e-01  4.097e-01   1.498 0.134184
poly(a$Days_above_treshold_Wspeed, 2)2  4.047e-01  1.727e-01   2.343 0.019187 *
poly(a$`min(Wspeed)`, 2)1              -1.629e-01  1.368e-01  -1.191 0.233686
poly(a$`min(Wspeed)`, 2)2               3.480e-01  9.399e-02   3.703 0.000216 ***
a$`max(Wspeed)`                          1.279e-03  3.728e-04   3.430 0.000611 ***
a$`mean Wspeed`                         -1.987e-02  9.818e-03  -2.023 0.043116 *
a$Days_above_mean_Wspeed                -3.115e-04  1.003e-04  -3.104 0.001923 **
a$fst_qtl_Wspeed_year                    9.963e-03  5.002e-03   1.992 0.046482 *
a$trd_qtl_Wspeed_year                    1.001e-02  4.696e-03   2.131 0.033158 *
poly(a$Days_above_treshold_diff, 2)1    1.879e+00  2.908e-01   6.463 1.17e-10 ***
poly(a$Days_above_treshold_diff, 2)2    5.223e-01  1.120e-01   4.662 3.24e-06 ***
poly(a$`min thermal_amp`, 2)1           1.774e-01  1.448e-01   1.225 0.220695
poly(a$`min thermal_amp`, 2)2           1.011e+00  1.363e-01   7.419 1.47e-13 ***
a$`max thermal_amp`                     -3.872e-03  8.877e-04  -4.362 1.32e-05 ***
poly(a$`mean thermal_amp`, 2)1         -2.986e-01  1.044e+00  -0.286 0.774827
poly(a$`mean thermal_amp`, 2)2         -3.418e+00  5.517e-01  -6.194 6.51e-10 ***
poly(a$fst_qtl_diff_year, 2)1           2.399e-01  4.405e-01   0.545 0.585982
poly(a$fst_qtl_diff_year, 2)2           1.551e+00  2.663e-01   5.823 6.29e-09 ***
poly(a$trd_qtl_diff_year, 2)1           1.480e-01  6.353e-01   0.233 0.815814
poly(a$trd_qtl_diff_year, 2)2           1.711e+00  3.746e-01   4.568 5.10e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06253 on 3585 degrees of freedom
  (39 observations deleted due to missingness)
Multiple R-squared:  0.6042,    Adjusted R-squared:  0.5988
F-statistic: 111.7 on 49 and 3585 DF,  p-value: < 2.2e-16
```

Figure 8. Results of Regressions on Tariffs for All Regions and all crops. Introduction of new weather Variables

```
Call:
lm(formula = a$Tariffs ~ a$Days_above_treshold_Rtotal + a$Days_above_mean_Rtotal +
    a$Days_above_treshold_Tmin + a$`min(Tmin)` + a$`max(Tmin)` +
    poly(a$Days_above_mean_Tmin, 2) + a$trd_qtl_Tmin_year + poly(a$Days_above_treshold_Tmax,
    2) + a$`min(Tmax)` + poly(a$`max(Tmax)`, 2) + poly(a$fst_qtl_Tmax_year,
    2) + poly(a$Days_above_treshold_Wspeed, 2) + poly(a$Days_above_treshold_diff,
    2) + a$`min thermal_amp` + a$`max thermal_amp` + a$Days_above_mean_diff +
    a$trd_qtl_diff_year)

Residuals:
      Min        1Q    Median        3Q       Max
-0.092342 -0.026023 -0.003703  0.019710  0.117261

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                         -0.0296106  0.0671665  -0.441 0.659425
a$Days_above_treshold_Rtotal         0.0009856  0.0001974   4.994 7.12e-07 ***
a$Days_above_mean_Rtotal            -0.0009565  0.0002464  -3.882 0.000111 ***
a$Days_above_treshold_Tmin          -0.0009507  0.0001634  -5.818 8.26e-09 ***
a$`min(Tmin)`                       -0.0011312  0.0009316  -1.214 0.224959
a$`max(Tmin)`                        0.0019221  0.0009467   2.030 0.042616 *
poly(a$Days_above_mean_Tmin, 2)1     0.4498158  0.1311790   3.429 0.000633 ***
poly(a$Days_above_mean_Tmin, 2)2    -0.0498287  0.0792413  -0.629 0.529626
a$trd_qtl_Tmin_year                  0.0067050  0.0025087   2.673 0.007661 **
poly(a$Days_above_treshold_Tmax, 2)1   -0.3765745  0.1307875  -2.879 0.004080 **
poly(a$Days_above_treshold_Tmax, 2)2   -0.2682008  0.0665945  -4.027 6.12e-05 ***
a$`min(Tmax)`                       -0.0014704  0.0007170  -2.051 0.040573 *
poly(a$`max(Tmax)`, 2)1             -0.1203678  0.1112133  -1.082 0.279403
poly(a$`max(Tmax)`, 2)2              0.1189422  0.0702598   1.693 0.090823 .
poly(a$fst_qtl_Tmax_year, 2)1       -0.4238383  0.1178483  -3.596 0.000340 ***
poly(a$fst_qtl_Tmax_year, 2)2        0.1964702  0.0456528   4.304 1.87e-05 ***
poly(a$Days_above_treshold_Wspeed, 2)1  0.0470570  0.0498141   0.945 0.345090
poly(a$Days_above_treshold_Wspeed, 2)2  0.1271980  0.0456751   2.785 0.005468 **
poly(a$Days_above_treshold_diff, 2)1    0.3417686  0.2108970   1.621 0.105466
poly(a$Days_above_treshold_diff, 2)2    0.1332342  0.0592042   2.250 0.024664 *
a$`min thermal_amp`                  0.0037766  0.0017807   2.121 0.034204 *
a$`max thermal_amp`                 -0.0044984  0.0009433  -4.769 2.16e-06 ***
a$Days_above_mean_diff              -0.0003670  0.0001044  -3.516 0.000459 ***
a$trd_qtl_diff_year                  0.0121071  0.0024060   5.032 5.86e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03828 on 898 degrees of freedom
Multiple R-squared:  0.4688,    Adjusted R-squared:  0.4552
F-statistic: 34.46 on 23 and 898 DF,  p-value: < 2.2e-16
```

Figure 9. Results of Regressions on Tariffs for All Regions and Vineyard crops. Introduction of new weather Variables

```
Call:
lm(formula = a$Tariffs ~ poly(a$Days_above_treshold_Rtotal, 2) +
    poly(a$`max(Rtotal)`, 2) + poly(a$`mean Rtotal`, 2) + poly(a$Days_above_mean_Rtotal,
    2) + a$fst_qtl_Rtotal_year + a$trd_qtl_Rtotal_year + poly(a$Days_above_treshold_Tmin,
    2) + poly(a$`min(Tmin)`, 2) + poly(a$`max(Tmin)`, 2) + poly(a$`mean Tmin`,
    2) + poly(a$Days_above_mean_Tmin, 2) + poly(a$fst_qtl_Tmin_year,
    2) + poly(a$trd_qtl_Tmin_year, 2) + poly(a$Days_above_treshold_Tmax,
    2) + poly(a$`min(Tmax)`, 2) + poly(a$`max(Tmax)`, 2) + poly(a$`mean Tmax`,
    2) + a$trd_qtl_Tmax_year + a$Days_above_treshold_Wspeed +
    poly(a$`min(Wspeed)`, 2) + a$`max(Wspeed)` + a$`mean Wspeed` +
    poly(a$Days_above_mean_Wspeed, 2) + poly(a$fst_qtl_Wspeed_year,
    2) + a$trd_qtl_Wspeed_year + poly(a$Days_above_treshold_diff,
    2) + poly(a$`min thermal_amp`, 2) + poly(a$`max thermal_amp`,
    2) + poly(a$`mean thermal_amp`, 2) + poly(a$Days_above_mean_diff,
    2) + poly(a$fst_qtl_diff_year, 2) + poly(a$trd_qtl_diff_year,
    2))

    Coefficients:
                                            Estimate Std. Error t value Pr(>|t|)
    (Intercept)                            0.1070826  0.0759301   1.410 0.158574
    poly(a$Days_above_treshold_Rtotal, 2)1 1.3043572  0.2668193   4.889 1.08e-06 ***
    poly(a$Days_above_treshold_Rtotal, 2)2 0.5716934  0.1715265   3.333 0.000871 ***
    poly(a$`max(Rtotal)`, 2)1              0.2684482  0.0853350   3.146 0.001675 **
    poly(a$`max(Rtotal)`, 2)2              0.1973230  0.0699815   2.820 0.004843 **
    poly(a$`mean Rtotal`, 2)1              0.5051996  0.1435654   3.519 0.000441 ***
    poly(a$`mean Rtotal`, 2)2             -0.4216014  0.0898325  -4.693 2.83e-06 ***
    poly(a$Days_above_mean_Rtotal, 2)1    -1.7372221  0.2862655  -6.069 1.47e-09 ***
    poly(a$Days_above_mean_Rtotal, 2)2    -0.1316888  0.1736112  -0.759 0.448203
    a$fst_qtl_Rtotal_year                  7.2569432  1.4640420   4.957 7.62e-07 ***
    a$trd_qtl_Rtotal_year                 -0.0120788  0.0027747  -4.353 1.39e-05 ***
    poly(a$Days_above_treshold_Tmin, 2)1  -1.5081612  0.2931249  -5.145 2.87e-07 ***
    poly(a$Days_above_treshold_Tmin, 2)2   0.9603856  0.1218298   7.883 4.62e-15 ***
    poly(a$`min(Tmin)`, 2)1                0.2861624  0.1150714   2.487 0.012950 *
    poly(a$`min(Tmin)`, 2)2                0.3451747  0.0849997   4.061 5.03e-05 ***
    poly(a$`max(Tmin)`, 2)1                0.6565780  0.1544764   4.250 2.21e-05 ***
    poly(a$`max(Tmin)`, 2)2               -0.8322121  0.1528086  -5.446 5.62e-08 ***
    poly(a$`mean Tmin`, 2)1               -3.0077793  0.8318780  -3.616 0.000305 ***
    poly(a$`mean Tmin`, 2)2               -2.3836265  0.4616384  -5.163 2.60e-07 ***
    poly(a$Days_above_mean_Tmin, 2)1      -0.3502239  0.2650611  -1.321 0.186517
    poly(a$Days_above_mean_Tmin, 2)2      -0.9379576  0.1744254  -5.377 8.21e-08 ***
    poly(a$fst_qtl_Tmin_year, 2)1          2.2089811  0.4077416   5.418 6.58e-08 ***
    poly(a$fst_qtl_Tmin_year, 2)2          1.1490487  0.2352399   4.885 1.10e-06 ***
    poly(a$trd_qtl_Tmin_year, 2)1          0.2545462  0.4760148   0.535 0.592871
    poly(a$trd_qtl_Tmin_year, 2)2          1.4156797  0.3626827   3.903 9.72e-05 ***
    poly(a$Days_above_treshold_Tmax, 2)1   0.1714669  0.2462071   0.696 0.486218
    poly(a$Days_above_treshold_Tmax, 2)2  -0.3941708  0.0965891  -4.081 4.62e-05 ***
    poly(a$`min(Tmax)`, 2)1               -1.4299813  0.1292280 -11.066  < 2e-16 ***
    poly(a$`min(Tmax)`, 2)2               -0.8868501  0.0758023 -11.700  < 2e-16 ***
    poly(a$`max(Tmax)`, 2)1                0.3345317  0.1897582   1.763 0.078026 .
    poly(a$`max(Tmax)`, 2)2               -0.6899620  0.1598699  -4.316 1.65e-05 ***
    poly(a$`mean Tmax`, 2)1               -0.7447283  0.5133160  -1.451 0.146949
    poly(a$`mean Tmax`, 2)2                0.3099934  0.1290103   2.403 0.016336 *
    a$trd_qtl_Tmax_year                    0.0028414  0.0021846   1.301 0.193477
    a$Days_above_treshold_Wspeed           0.0001540  0.0001387   1.110 0.267081
    poly(a$`min(Wspeed)`, 2)1             -0.2364463  0.1145190  -2.065 0.039049 *
    poly(a$`min(Wspeed)`, 2)2              0.1866499  0.0860311   2.170 0.030128 *
    a$`max(Wspeed)`                        0.0013745  0.0003599   3.820 0.000137 ***
    a$`mean Wspeed`                       -0.0290353  0.0094422  -3.075 0.002126 **
    poly(a$Days_above_mean_Wspeed, 2)1    -1.0441679  0.3268767  -3.194 0.001418 **
    poly(a$Days_above_mean_Wspeed, 2)2     0.5950959  0.1380559   4.311 1.69e-05 ***
    poly(a$fst_qtl_Wspeed_year, 2)1        1.0636318  0.6032262   1.763 0.077975 .
    poly(a$fst_qtl_Wspeed_year, 2)2        0.3184118  0.1057129   3.012 0.002619 **
    a$trd_qtl_Wspeed_year                  0.0173728  0.0044379   3.915 9.28e-05 ***
    poly(a$Days_above_treshold_diff, 2)1   1.9146913  0.3665519   5.224 1.89e-07 ***
    poly(a$Days_above_treshold_diff, 2)2   1.2499810  0.1406598   8.887  < 2e-16 ***
    poly(a$`min thermal_amp`, 2)1          0.1815561  0.1270393   1.429 0.153084
    poly(a$`min thermal_amp`, 2)2          1.1532272  0.1265322   9.114  < 2e-16 ***
    poly(a$`max thermal_amp`, 2)1         -1.2186182  0.1521938  -8.007 1.74e-15 ***
    poly(a$`max thermal_amp`, 2)2          0.5116127  0.1011779   5.057 4.56e-07 ***
    poly(a$`mean thermal_amp`, 2)1        -2.3197125  0.9176608  -2.528 0.011534 *
    poly(a$`mean thermal_amp`, 2)2        -3.5758067  0.5057415  -7.070 1.97e-12 ***
    poly(a$Days_above_mean_diff, 2)1       0.3921978  0.3510106   1.117 0.263950
    poly(a$Days_above_mean_diff, 2)2      -1.3217574  0.1835403  -7.201 7.72e-13 ***
    poly(a$fst_qtl_diff_year, 2)1          0.7433315  0.3725243   1.995 0.046102 *
    poly(a$fst_qtl_diff_year, 2)2          2.0791353  0.2356682   8.822  < 2e-16 ***
    poly(a$trd_qtl_diff_year, 2)1          0.2469606  0.5463199   0.452 0.651274
    poly(a$trd_qtl_diff_year, 2)2          1.4242023  0.3444107   4.135 3.66e-05 ***

    Residual standard error: 0.05093 on 2655 degrees of freedom
      (39 observations deleted due to missingness)
    Multiple R-squared:  0.7815,    Adjusted R-squared:  0.7769
    F-statistic: 166.6 on 57 and 2655 DF,  p-value: < 2.2e-16
```

Figure 10. Results of Regressions on Tariffs for All Regions and All crops except Vineyards. Introduction of new weather Variables

```
Call:
lm(formula = a$`Bonus over Comercial Prize` ~ a$Days_above_treshold_Rtotal +
    poly(a$`max(Rtotal)`, 2) + a$trd_qtl_Rtotal_year + a$Days_above_treshold_Tmin +
    a$`min(Tmin)` + poly(a$`max(Tmin)`, 2) + poly(a$trd_qtl_Tmin_year,
    2) + a$Days_above_treshold_Tmax + a$Days_under_treshold_Tmax +
    a$`min(Tmax)` + poly(a$fst_qtl_Tmax_year, 2) + a$Days_above_treshold_Wspeed +
    a$fst_qtl_Wspeed_year)

Residuals:
     Min        1Q    Median        3Q       Max
-0.066878 -0.020027  0.001757  0.014458  0.087087

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                   0.1438001  0.0578752   2.485 0.015109 *
a$Days_above_treshold_Rtotal -0.0018383  0.0001985  -9.260 3.38e-14 ***
poly(a$`max(Rtotal)`, 2)1     0.2822127  0.0593145   4.758 8.82e-06 ***
poly(a$`max(Rtotal)`, 2)2    -0.3445371  0.0532836  -6.466 8.01e-09 ***
a$trd_qtl_Rtotal_year         0.1807467  0.0145672  12.408  < 2e-16 ***
a$Days_above_treshold_Tmin   -0.0016266  0.0003326  -4.891 5.27e-06 ***
a$`min(Tmin)`                -0.0416712  0.0052542  -7.931 1.28e-11 ***
poly(a$`max(Tmin)`, 2)1       0.4493866  0.1412041   3.183 0.002097 **
poly(a$`max(Tmin)`, 2)2      -0.6313649  0.1271615  -4.965 3.95e-06 ***
poly(a$trd_qtl_Tmin_year, 2)1 -0.6449067  0.1897097  -3.399 0.001067 **
poly(a$trd_qtl_Tmin_year, 2)2  1.3904750  0.1823618   7.625 5.00e-11 ***
a$Days_above_treshold_Tmax    0.0003310  0.0003116   1.062 0.291423
a$Days_under_treshold_Tmax   -0.0020285  0.0008346  -2.430 0.017381 *
a$`min(Tmax)`                 0.0104011  0.0030093   3.456 0.000889 ***
poly(a$fst_qtl_Tmax_year, 2)1  0.3068077  0.1353561   2.267 0.026181 *
poly(a$fst_qtl_Tmax_year, 2)2 -0.5265532  0.0835272  -6.304 1.61e-08 ***
a$Days_above_treshold_Wspeed  0.0002722  0.0001634   1.666 0.099656 .
a$fst_qtl_Wspeed_year         0.0676133  0.0074087   9.126 6.14e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03059 on 78 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.8579,    Adjusted R-squared:  0.8269
F-statistic:  27.7 on 17 and 78 DF,  p-value: < 2.2e-16
```

Figure 11. Results of Regressions on BoCP for Region A and all crops except Vineyards

```
Call:
lm(formula = a$`Bonus over Comercial Prize` ~ a$Days_above_treshold_Rtotal +
    poly(a$Days_above_treshold_Tmin, 2) + poly(a$`max(Tmin)`,
    2) + poly(a$fst_qtl_Tmin_year, 2) + a$trd_qtl_Tmin_year +
    poly(a$Days_under_treshold_Tmax, 2) + a$`max(Tmax)` + a$fst_qtl_Tmax_year +
    a$trd_qtl_Tmax_year + poly(a$Days_under_treshold_Wspeed,
    2) + a$`min(Wspeed)` + a$`max(Wspeed)` + a$fst_qtl_Wspeed_year)

Residuals:
     Min        1Q    Median        3Q       Max
-0.229213 -0.033589 -0.001007  0.032794  0.179664

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                         -4.427e-01  2.854e-01  -1.551 0.122223
a$Days_above_treshold_Rtotal         1.336e-05  2.395e-04   0.056 0.955552
poly(a$Days_above_treshold_Tmin, 2)1 -1.224e-01  2.222e-01  -0.551 0.582097
poly(a$Days_above_treshold_Tmin, 2)2  2.619e-01  1.671e-01   1.568 0.118259
poly(a$`max(Tmin)`, 2)1             -1.815e-01  2.175e-01  -0.835 0.404720
poly(a$`max(Tmin)`, 2)2             -1.526e-01  1.326e-01  -1.151 0.251046
poly(a$fst_qtl_Tmin_year, 2)1       -1.010e+00  1.478e-01  -6.832 6.93e-11 ***
poly(a$fst_qtl_Tmin_year, 2)2       -5.957e-01  1.855e-01  -3.211 0.001506 **
a$trd_qtl_Tmin_year                  2.884e-02  1.152e-02   2.504 0.012953 *
poly(a$Days_under_treshold_Tmax, 2)1  7.783e-01  2.127e-01   3.659 0.000312 ***
poly(a$Days_under_treshold_Tmax, 2)2  3.662e-01  1.183e-01   3.095 0.002206 **
a$`max(Tmax)`                       -6.552e-03  2.265e-03  -2.892 0.004180 **
a$fst_qtl_Tmax_year                  7.949e-02  1.550e-02   5.127 6.09e-07 ***
a$trd_qtl_Tmax_year                 -8.417e-03  4.388e-03  -1.918 0.056280 .
poly(a$Days_under_treshold_Wspeed, 2)1 -2.587e-01  1.423e-01  -1.817 0.070399 .
poly(a$Days_under_treshold_Wspeed, 2)2  1.915e-01  1.157e-01   1.656 0.099139 .
a$`min(Wspeed)`                      1.322e-02  6.559e-03   2.016 0.044913 *
a$`max(Wspeed)`                      7.677e-03  1.532e-03   5.012 1.05e-06 ***
a$fst_qtl_Wspeed_year               -6.463e-02  1.745e-02  -3.705 0.000263 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06847 on 238 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.4705,    Adjusted R-squared:  0.4304
F-statistic: 11.75 on 18 and 238 DF,  p-value: < 2.2e-16
```

Figure 12. Results of Regressions on BoCP for Region B and all crops except Vineyards

```
Call:
lm(formula = a$`Bonus over Comercial Prize` ~ a$Days_above_treshold_Rtotal +
    poly(a$Days_under_treshold_Tmin, 2) + poly(a$`max(Tmin)`,
    2) + a$trd_qtl_Tmin_year + poly(a$Days_above_treshold_Tmax,
    2) + poly(a$Days_under_treshold_Tmax, 2) + poly(a$`max(Tmax)`,
    2) + poly(a$fst_qtl_Tmax_year, 2) + a$Days_above_treshold_Wspeed +
    a$Days_under_treshold_Wspeed + poly(a$`min(Wspeed)`, 2) +
    a$trd_qtl_Wspeed_year)

Residuals:
      Min        1Q    Median        3Q       Max
-0.204540 -0.051764 -0.004108  0.033674  0.231797

Coefficients:
                                       Estimate Std. Error t value Pr(>|t|)
(Intercept)                          -0.7542248  0.3842897  -1.963  0.05167 .
a$Days_above_treshold_Rtotal         -0.0003258  0.0005125  -0.636  0.52598
poly(a$Days_under_treshold_Tmin, 2)1  0.5625193  0.2362004   2.382  0.01859 *
poly(a$Days_under_treshold_Tmin, 2)2 -0.1738560  0.1400990  -1.241  0.21670
poly(a$`max(Tmin)`, 2)1              -0.0059605  0.1890011  -0.032  0.97489
poly(a$`max(Tmin)`, 2)2              -0.2199438  0.1156524  -1.902  0.05926 .
a$trd_qtl_Tmin_year                   0.0110557  0.0170148   0.650  0.51691
poly(a$Days_above_treshold_Tmax, 2)1 -1.3909217  0.1752960  -7.935 6.11e-13 ***
poly(a$Days_above_treshold_Tmax, 2)2  0.4242685  0.1433239   2.960  0.00361 **
poly(a$Days_under_treshold_Tmax, 2)1 -0.8303885  0.4555256  -1.823  0.07045 .
poly(a$Days_under_treshold_Tmax, 2)2 -0.4214842  0.2229791  -1.890  0.06079 .
poly(a$`max(Tmax)`, 2)1               0.0556491  0.1373596   0.405  0.68600
poly(a$`max(Tmax)`, 2)2              -0.2787573  0.1351762  -2.062  0.04104 *
poly(a$fst_qtl_Tmax_year, 2)1         0.2700772  0.4503093   0.600  0.54964
poly(a$fst_qtl_Tmax_year, 2)2         0.6585589  0.2251353   2.925  0.00402 **
a$Days_above_treshold_Wspeed         -0.0017738  0.0008247  -2.151  0.03320 *
a$Days_under_treshold_Wspeed          0.0016110  0.0006837   2.356  0.01984 *
poly(a$`min(Wspeed)`, 2)1             0.1971826  0.1257157   1.568  0.11903
poly(a$`min(Wspeed)`, 2)2            -0.1660579  0.0933366  -1.779  0.07739 .
a$trd_qtl_Wspeed_year                 0.1086454  0.0343711   3.161  0.00193 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08898 on 140 degrees of freedom
Multiple R-squared:  0.6593,	Adjusted R-squared:  0.6131
F-statistic: 14.26 on 19 and 140 DF,  p-value: < 2.2e-16
```

Figure 13. Results of Regressions on BoCP for Region C and Vineyards Crops

```
Call:
lm(formula = a$`Bonus over Comercial Prize` ~ a$Days_above_treshold_Rtotal +
    poly(a$`max(Rtotal)`, 2) + a$Days_under_treshold_Tmin + poly(a$`max(Tmin)`,
    2) + poly(a$fst_qtl_Tmin_year, 2) + a$trd_qtl_Tmin_year +
    a$Days_above_treshold_Tmax + a$Days_under_treshold_Tmax +
    poly(a$trd_qtl_Tmax_year, 2) + a$Days_above_treshold_Wspeed +
    a$Days_under_treshold_Wspeed + a$`max(Wspeed)`)

Residuals:
     Min       1Q   Median       3Q      Max
-0.25281 -0.03760  0.00262  0.04913  0.22128

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    0.2012624  0.1094371   1.839 0.066883 .
a$Days_above_treshold_Rtotal   0.0001724  0.0002819   0.611 0.541388
poly(a$`max(Rtotal)`, 2)1      0.0413516  0.1190881   0.347 0.728655
poly(a$`max(Rtotal)`, 2)2      0.3589646  0.1031904   3.479 0.000578 ***
a$Days_under_treshold_Tmin    -0.0006098  0.0004517  -1.350 0.177993
poly(a$`max(Tmin)`, 2)1       -0.5306833  0.2223981  -2.386 0.017638 *
poly(a$`max(Tmin)`, 2)2        0.3026874  0.1529522   1.979 0.048725 *
poly(a$fst_qtl_Tmin_year, 2)1 -1.0372301  0.2306810  -4.496 9.85e-06 ***
poly(a$fst_qtl_Tmin_year, 2)2  0.4669730  0.1202618   3.883 0.000127 ***
a$trd_qtl_Tmin_year            0.0518747  0.0075977   6.828 4.72e-11 ***
a$Days_above_treshold_Tmax    -0.0012355  0.0003451  -3.580 0.000401 ***
a$Days_under_treshold_Tmax    -0.0002557  0.0003152  -0.811 0.417952
poly(a$trd_qtl_Tmax_year, 2)1 -0.6251129  0.2196758  -2.846 0.004735 **
poly(a$trd_qtl_Tmax_year, 2)2 -0.3635743  0.1619291  -2.245 0.025472 *
a$Days_above_treshold_Wspeed   0.0001997  0.0001645   1.214 0.225709
a$Days_under_treshold_Wspeed  -0.0003060  0.0001450  -2.111 0.035619 *
a$`max(Wspeed)`               -0.0011285  0.0011516  -0.980 0.327920
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08158 on 303 degrees of freedom
Multiple R-squared:  0.4957,	Adjusted R-squared:  0.469
F-statistic: 18.61 on 16 and 303 DF,  p-value: < 2.2e-16
```

Figure 14. Results of Regressions on BoCP for Region D and Vineyards Crops

```
Call:
lm(formula = a$`Bonus over Comercial Prize` ~ poly(a$Days_above_treshold_Rtotal,
    2) + poly(a$`max(Rtotal)`, 2) + poly(a$trd_qtl_Rtotal_year,
    2) + a$Days_above_treshold_Tmin + poly(a$`max(Tmin)`, 2) +
    a$trd_qtl_Tmin_year + a$Days_above_treshold_Tmax + poly(a$Days_under_treshold_Tmax,
    2) + a$`min(Tmax)` + a$`max(Tmax)` + poly(a$Days_under_treshold_Wspeed,
    2) + poly(a$`min(Wspeed)`, 2) + a$`max(Wspeed)` + poly(a$fst_qtl_Wspeed_year,
    2))

Residuals:
     Min       1Q   Median       3Q      Max
-0.28980 -0.05572  0.01320  0.06500  0.24275

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                             0.6518214  0.1565786   4.163 3.89e-05 ***
poly(a$Days_above_treshold_Rtotal, 2)1 -1.0945484  0.2159963  -5.067 6.31e-07 ***
poly(a$Days_above_treshold_Rtotal, 2)2 -0.1853835  0.1298001  -1.428 0.154049
poly(a$`max(Rtotal)`, 2)1              -0.7883446  0.1266014  -6.227 1.26e-09 ***
poly(a$`max(Rtotal)`, 2)2               0.3164085  0.1296234   2.441 0.015103 *
poly(a$trd_qtl_Rtotal_year, 2)1         0.5985424  0.1491872   4.012 7.25e-05 ***
poly(a$trd_qtl_Rtotal_year, 2)2        -0.1874117  0.1495463  -1.253 0.210903
a$Days_above_treshold_Tmin             -0.0025361  0.0006270  -4.045 6.35e-05 ***
poly(a$`max(Tmin)`, 2)1                 0.2656043  0.1776727   1.495 0.135768
poly(a$`max(Tmin)`, 2)2                 0.3121744  0.1512208   2.064 0.039662 *
a$trd_qtl_Tmin_year                     0.0438692  0.0083741   5.239 2.69e-07 ***
a$Days_above_treshold_Tmax              0.0006135  0.0004403   1.394 0.164275
poly(a$Days_under_treshold_Tmax, 2)1    0.9851712  0.2219620   4.438 1.19e-05 ***
poly(a$Days_under_treshold_Tmax, 2)2   -0.2226060  0.1218782  -1.826 0.068565 .
a$`min(Tmax)`                          -0.0186347  0.0029422  -6.334 6.77e-10 ***
a$`max(Tmax)`                          -0.0137819  0.0035936  -3.835 0.000147 ***
poly(a$Days_under_treshold_Wspeed, 2)1 -0.2447674  0.3942463  -0.621 0.535071
poly(a$Days_under_treshold_Wspeed, 2)2  0.3480674  0.1695903   2.052 0.040815 *
poly(a$`min(Wspeed)`, 2)1              -0.4011095  0.1707618  -2.349 0.019338 *
poly(a$`min(Wspeed)`, 2)2              -0.2942169  0.1444734  -2.036 0.042395 *
a$`max(Wspeed)`                         0.0050954  0.0014614   3.487 0.000546 ***
poly(a$fst_qtl_Wspeed_year, 2)1        -1.0483560  0.3646183  -2.875 0.004265 **
poly(a$fst_qtl_Wspeed_year, 2)2        -0.4862974  0.1434628  -3.390 0.000773 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09658 on 380 degrees of freedom
Multiple R-squared:  0.4363,    Adjusted R-squared:  0.4036
F-statistic: 13.37 on 22 and 380 DF,  p-value: < 2.2e-16
```

Figure 15. Results of Regressions on BoCP for Region E and Vineyards Crops

```
Call:
lm(formula = a$`Claims over Contracts` ~ a$Days_above_treshold_Rtotal +
    a$`mean Rtotal` + a$Days_above_mean_Rtotal + a$Days_above_treshold_Tmin +
    a$`min(Tmin)` + poly(a$`mean Tmin`, 2) + a$Days_above_mean_Tmin +
    a$fst_qtl_Tmin_year + a$Days_above_treshold_Tmax + poly(a$`min(Tmax)`,
    2) + poly(a$`max(Tmax)`, 2) + poly(a$`mean Tmax`, 2) + poly(a$fst_qtl_Tmax_year,
    2) + poly(a$trd_qtl_Tmax_year, 2) + a$Days_above_treshold_Wspeed +
    poly(a$`min(Wspeed)`, 2) + a$`max(Wspeed)` + a$Days_above_mean_Wspeed +
    poly(a$fst_qtl_Wspeed_year, 2) + poly(a$trd_qtl_Wspeed_year,
    2) + poly(a$Days_above_treshold_diff, 2) + poly(a$`min thermal_amp`,
    2) + a$`max thermal_amp` + a$`mean thermal_amp` + a$fst_qtl_diff_year)

Residuals:
    Min      1Q  Median      3Q     Max
-1.6428 -0.3367 -0.0478  0.2327  3.2668

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                          2.536e+00  5.076e-01   4.996 6.14e-07 ***
a$Days_above_treshold_Rtotal         1.916e-02  1.661e-03  11.534  < 2e-16 ***
a$`mean Rtotal`                      3.417e-02  1.713e-02   1.995 0.046103 *
a$Days_above_mean_Rtotal            -2.399e-02  2.338e-03 -10.262  < 2e-16 ***
a$Days_above_treshold_Tmin          -2.919e-03  1.268e-03  -2.301 0.021434 *
a$`min(Tmin)`                        3.817e-02  7.842e-03   4.868 1.18e-06 ***
poly(a$`mean Tmin`, 2)1             -1.604e+01  5.487e+00  -2.924 0.003477 **
poly(a$`mean Tmin`, 2)2             -1.857e+00  1.067e+00  -1.741 0.081833 .
a$Days_above_mean_Tmin              -3.772e-04  9.000e-04  -0.419 0.675181
a$fst_qtl_Tmin_year                  1.058e-01  2.671e-02   3.959 7.67e-05 ***
a$Days_above_treshold_Tmax          -4.295e-03  1.184e-03  -3.626 0.000292 ***
poly(a$`min(Tmax)`, 2)1             -1.873e+01  1.373e+00 -13.635  < 2e-16 ***
poly(a$`min(Tmax)`, 2)2             -4.941e+00  8.037e-01  -6.147 8.75e-10 ***
poly(a$`max(Tmax)`, 2)1             -2.798e+00  1.955e+00  -1.431 0.152499
poly(a$`max(Tmax)`, 2)2             -4.748e+00  1.618e+00  -2.935 0.003360 **
poly(a$`mean Tmax`, 2)1              4.288e+01  8.579e+00   4.998 6.06e-07 ***
poly(a$`mean Tmax`, 2)2             -1.328e+01  3.188e+00  -4.165 3.19e-05 ***
poly(a$fst_qtl_Tmax_year, 2)1       -2.057e+01  4.803e+00  -4.283 1.89e-05 ***
poly(a$fst_qtl_Tmax_year, 2)2        1.311e+01  2.107e+00   6.222 5.46e-10 ***
poly(a$trd_qtl_Tmax_year, 2)1       -1.162e+01  4.686e+00  -2.480 0.013201 *
poly(a$trd_qtl_Tmax_year, 2)2        6.959e+00  2.631e+00   2.645 0.008210 **
a$Days_above_treshold_Wspeed         9.038e-03  1.231e-03   7.341 2.62e-13 ***
poly(a$`min(Wspeed)`, 2)1            1.857e+00  1.244e+00   1.492 0.135744
poly(a$`min(Wspeed)`, 2)2           -1.729e+00  9.820e-01  -1.761 0.078370 .
a$`max(Wspeed)`                      4.883e-03  2.979e-03   1.639 0.101331
a$Days_above_mean_Wspeed            -7.969e-03  8.640e-04  -9.224  < 2e-16 ***
poly(a$fst_qtl_Wspeed_year, 2)1     -1.660e+00  3.477e+00  -0.477 0.633125
poly(a$fst_qtl_Wspeed_year, 2)2      3.526e+00  1.699e+00   2.075 0.038078 *
poly(a$trd_qtl_Wspeed_year, 2)1     -6.058e+00  3.170e+00  -1.911 0.056041 .
poly(a$trd_qtl_Wspeed_year, 2)2     -3.880e+00  1.764e+00  -2.200 0.027895 *
poly(a$Days_above_treshold_diff, 2)1  1.472e+01  2.557e+00   5.757 9.27e-09 ***
poly(a$Days_above_treshold_diff, 2)2 -3.405e+00  7.785e-01  -4.373 1.26e-05 ***
poly(a$`min thermal_amp`, 2)1        2.656e+00  1.195e+00   2.222 0.026325 *
poly(a$`min thermal_amp`, 2)2        8.014e+00  1.094e+00   7.324 2.97e-13 ***
a$`max thermal_amp`                 -2.463e-02  8.030e-03  -3.067 0.002181 **
a$`mean thermal_amp`                 5.391e-02  4.729e-02   1.140 0.254416
a$fst_qtl_diff_year                 -1.316e-01  2.948e-02  -4.465 8.27e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6013 on 3598 degrees of freedom
  (39 observations deleted due to missingness)
Multiple R-squared:  0.3828,    Adjusted R-squared:  0.3766
F-statistic: 61.98 on 36 and 3598 DF,  p-value: < 2.2e-16
```

Figure 16. Results of Regressions on CoNC for All Regions and all cultures

```
Call:
lm(formula = a$`Claims over Contracts` ~ poly(a$Days_above_treshold_Rtotal,
    2) + a$`mean Rtotal` + poly(a$Days_above_mean_Rtotal, 2) +
    a$trd_qtl_Rtotal_year + a$Days_above_treshold_Tmin + a$`min(Tmin)` +
    a$`max(Tmin)` + poly(a$`mean Tmin`, 2) + poly(a$Days_above_mean_Tmin,
    2) + poly(a$fst_qtl_Tmin_year, 2) + a$Days_above_treshold_Tmax +
    a$`mean Tmax` + poly(a$trd_qtl_Tmax_year, 2) + a$Days_above_treshold_Wspeed +
    a$`min(Wspeed)` + poly(a$`max(Wspeed)`, 2) + poly(a$`mean Wspeed`,
    2) + poly(a$trd_qtl_Wspeed_year, 2) + poly(a$Days_above_treshold_diff,
    2) + poly(a$`max thermal_amp`, 2) + poly(a$`mean thermal_amp`,
    2) + poly(a$fst_qtl_diff_year, 2) + poly(a$trd_qtl_diff_year,
    2))
```

```
Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)                             -8.502032   2.943951  -2.888 0.004266 ** 
poly(a$Days_above_treshold_Rtotal, 2)1   4.404661   1.605921   2.743 0.006597 ** 
poly(a$Days_above_treshold_Rtotal, 2)2  -1.995393   0.904833  -2.205 0.028476 *  
a$`mean Rtotal`                         -0.209211   0.063370  -3.301 0.001123 ** 
poly(a$Days_above_mean_Rtotal, 2)1      -3.163612   1.572227  -2.012 0.045425 *  
poly(a$Days_above_mean_Rtotal, 2)2       5.047658   0.943795   5.348 2.23e-07 ***
a$trd_qtl_Rtotal_year                    0.078542   0.045029   1.744 0.082520 .  
a$Days_above_treshold_Tmin              -0.025150   0.002693  -9.340  < 2e-16 ***
a$`min(Tmin)`                           -0.103228   0.017312  -5.963 9.83e-09 ***
a$`max(Tmin)`                           -0.169511   0.022003  -7.704 4.52e-13 ***
poly(a$`mean Tmin`, 2)1                 10.444244   4.161639   2.510 0.012809 *  
poly(a$`mean Tmin`, 2)2                 12.567850   2.081529   6.038 6.62e-09 ***
poly(a$Days_above_mean_Tmin, 2)1        -2.028547   1.231314  -1.647 0.100896    
poly(a$Days_above_mean_Tmin, 2)2        -6.612165   1.151834  -5.741 3.12e-08 ***
poly(a$fst_qtl_Tmin_year, 2)1          -16.797257   1.642363 -10.227  < 2e-16 ***
poly(a$fst_qtl_Tmin_year, 2)2           -4.185841   1.113845  -3.758 0.000220 ***
a$Days_above_treshold_Tmax               0.001955   0.002867   0.682 0.495987    
a$`mean Tmax`                            0.668790   0.133547   5.008 1.13e-06 ***
poly(a$trd_qtl_Tmax_year, 2)1           -8.646156   2.115923  -4.086 6.15e-05 ***
poly(a$trd_qtl_Tmax_year, 2)2            7.164362   1.048581   6.832 8.15e-11 ***
a$Days_above_treshold_Wspeed            -0.009461   0.003012  -3.141 0.001915 ** 
a$`min(Wspeed)`                         -0.057367   0.027951  -2.052 0.041320 *  
poly(a$`max(Wspeed)`, 2)1               -1.113053   0.438802  -2.537 0.011891 *  
poly(a$`max(Wspeed)`, 2)2               -1.299413   0.657759  -1.976 0.049465 *  
poly(a$`mean Wspeed`, 2)1               -1.537739   1.737819  -0.885 0.377199    
poly(a$`mean Wspeed`, 2)2              -26.296862   4.929378  -5.335 2.38e-07 ***
poly(a$trd_qtl_Wspeed_year, 2)1          3.549582   2.479774   1.431 0.153737    
poly(a$trd_qtl_Wspeed_year, 2)2         26.385758   5.218177   5.057 9.01e-07 ***
poly(a$Days_above_treshold_diff, 2)1    12.129689   1.800986   6.735 1.42e-10 ***
poly(a$Days_above_treshold_diff, 2)2    -2.434067   0.728543  -3.341 0.000981 ***
poly(a$`max thermal_amp`, 2)1            2.069644   1.382969   1.497 0.135957    
poly(a$`max thermal_amp`, 2)2           -4.235358   0.868913  -4.874 2.10e-06 ***
poly(a$`mean thermal_amp`, 2)1         -44.050431   8.303468  -5.305 2.75e-07 ***
poly(a$`mean thermal_amp`, 2)2          48.392829   7.955526   6.083 5.20e-09 ***
poly(a$fst_qtl_diff_year, 2)1           -3.982227   1.783262  -2.233 0.026553 *  
poly(a$fst_qtl_diff_year, 2)2          -21.940686   2.641307  -8.307 1.02e-14 ***
poly(a$trd_qtl_diff_year, 2)1            8.153492   5.922636   1.377 0.170021    
poly(a$trd_qtl_diff_year, 2)2          -23.944573   6.056510  -3.954 0.000104 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2048 on 219 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.8394,	Adjusted R-squared:  0.8123 
F-statistic: 30.95 on 37 and 219 DF,  p-value: < 2.2e-16
```

Figure 17. Results of Regressions on CoNC for Region B and all cultures

```
Call:
lm(formula = a$`Claims over Contracts` ~ poly(a$Days_above_treshold_Rtotal,
    2) + poly(a$`max(Rtotal)`, 2) + poly(a$`mean Rtotal`, 2) +
    a$Days_above_mean_Rtota + poly(a$trd_qtl_Rtotal_year, 2) +
    poly(a$Days_above_treshold_Tmin, 2) + poly(a$`min(Tmin)`,
    2) + a$`max(Tmin)` + poly(a$`mean Tmin`, 2) + poly(a$Days_above_mean_Tmin,
    2) + poly(a$fst_qtl_Tmin_year, 2) + poly(a$trd_qtl_Tmin_year,
    2) + a$Days_above_treshold_Tmax + a$`min(Tmax)` + a$`max(Tmax)` +
    a$`mean Tmax` + poly(a$Days_above_mean_Tmax, 2) + poly(a$trd_qtl_Tmax_year,
    2) + a$Days_above_treshold_Wspeed + a$`min(Wspeed)` + a$`max(Wspeed)` +
    a$`mean Wspeed` + a$trd_qtl_Wspeed_year + poly(a$Days_above_treshold_diff,
    2) + a$`min thermal_amp` + a$`max thermal_amp` + a$`mean thermal_amp` +
    poly(a$Days_above_mean_diff, 2) + a$fst_qtl_diff_year)

Residuals:
     Min       1Q   Median       3Q      Max
-0.94156 -0.11827  0.03772  0.15263  0.88281

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                            -4.228e+00  9.012e-01  -4.692 3.13e-06 ***
poly(a$Days_above_treshold_Rtotal, 2)1  5.645e+00  1.184e+00   4.767 2.19e-06 ***
poly(a$Days_above_treshold_Rtotal, 2)2 -9.630e-01  4.412e-01  -2.182 0.029337 *
poly(a$`max(Rtotal)`, 2)1              -1.455e+00  4.830e-01  -3.012 0.002667 **
poly(a$`max(Rtotal)`, 2)2               1.970e+00  6.420e-01   3.068 0.002221 **
poly(a$`mean Rtotal`, 2)1               2.854e+00  1.106e+00   2.581 0.010023 *
poly(a$`mean Rtotal`, 2)2              -5.594e+00  1.688e+00  -3.314 0.000958 ***
a$Days_above_mean_Rtota                -1.354e-02  2.544e-03  -5.323 1.29e-07 ***
poly(a$trd_qtl_Rtotal_year, 2)1         1.730e+00  1.454e+00   1.189 0.234606
poly(a$trd_qtl_Rtotal_year, 2)2         3.647e+00  1.023e+00   3.564 0.000385 ***
poly(a$Days_above_treshold_Tmin, 2)1    3.736e+00  1.266e+00   2.951 0.003256 **
poly(a$Days_above_treshold_Tmin, 2)2    1.159e+00  6.217e-01   1.864 0.062623 .
poly(a$`min(Tmin)`, 2)1                 3.098e+00  4.905e-01   6.316 4.25e-10 ***
poly(a$`min(Tmin)`, 2)2                 9.788e-01  3.667e-01   2.669 0.007740 **
a$`max(Tmin)`                           6.083e-02  7.479e-03   8.133 1.40e-15 ***
poly(a$`mean Tmin`, 2)1                 1.632e+01  3.429e+00   4.760 2.26e-06 ***
poly(a$`mean Tmin`, 2)2                 1.886e+00  2.378e+00   0.793 0.427734
poly(a$Days_above_mean_Tmin, 2)1       -2.767e+00  1.947e+00  -1.421 0.155725
poly(a$Days_above_mean_Tmin, 2)2       -3.412e+00  9.858e-01  -3.461 0.000564 ***
poly(a$fst_qtl_Tmin_year, 2)1          -1.504e+01  1.879e+00  -8.003 3.78e-15 ***
poly(a$fst_qtl_Tmin_year, 2)2           1.830e+00  1.071e+00   1.709 0.087884 .
poly(a$trd_qtl_Tmin_year, 2)1          -8.448e+00  2.225e+00  -3.798 0.000156 ***
poly(a$trd_qtl_Tmin_year, 2)2          -4.768e+00  1.836e+00  -2.597 0.009549 **
a$Days_above_treshold_Tmax              2.123e-03  1.747e-03   1.216 0.224465
a$`min(Tmax)`                           1.035e-02  7.249e-03   1.428 0.153768
a$`max(Tmax)`                          -5.066e-02  8.004e-03  -6.329 3.92e-10 ***
a$`mean Tmax`                           2.416e-01  2.945e-02   8.205 8.11e-16 ***
poly(a$Days_above_mean_Tmax, 2)1        6.367e+00  1.582e+00   4.025 6.19e-05 ***
poly(a$Days_above_mean_Tmax, 2)2       -3.748e+00  1.136e+00  -3.299 0.001008 **
poly(a$trd_qtl_Tmax_year, 2)1          -1.270e+01  2.113e+00  -6.012 2.67e-09 ***
poly(a$trd_qtl_Tmax_year, 2)2           2.613e+00  1.204e+00   2.169 0.030314 *
a$Days_above_treshold_Wspeed            2.551e-03  6.012e-04   4.243 2.43e-05 ***
a$`min(Wspeed)`                         5.668e-02  2.369e-02   2.392 0.016949 *
a$`max(Wspeed)`                         1.341e-02  2.883e-03   4.653 3.78e-06 ***
a$`mean Wspeed`                         2.341e-01  5.559e-02   4.211 2.80e-05 ***
a$trd_qtl_Wspeed_year                  -2.717e-01  4.525e-02  -6.005 2.79e-09 ***
poly(a$Days_above_treshold_diff, 2)1   -2.896e+00  1.878e+00  -1.542 0.123442
poly(a$Days_above_treshold_diff, 2)2   -5.395e+00  8.353e-01  -6.458 1.75e-10 ***
a$`min thermal_amp`                     7.375e-02  1.638e-02   4.502 7.63e-06 ***
a$`max thermal_amp`                    -3.859e-02  1.056e-02  -3.654 0.000273 ***
a$`mean thermal_amp`                    3.438e-01  5.958e-02   5.769 1.10e-08 ***
poly(a$Days_above_mean_diff, 2)1       -1.943e+00  2.173e+00  -0.894 0.371550
poly(a$Days_above_mean_diff, 2)2        5.222e+00  1.161e+00   4.496 7.84e-06 ***
a$fst_qtl_diff_year                    -3.619e-01  3.857e-02  -9.384  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2452 on 886 degrees of freedom
  (28 observations deleted due to missingness)
Multiple R-squared:  0.5816,	Adjusted R-squared:  0.5613
F-statistic: 28.64 on 43 and 886 DF,  p-value: < 2.2e-16
```

Figure 18. Results of Regressions on CoNC for Region C and all cultures

```
Call:
lm(formula = a$`Claims over Contracts` ~ poly(a$Days_above_treshold_Rtotal,
    2) + a$`max(Rtotal)` + a$Days_above_mean_Rtotal + a$trd_qtl_Rtotal_year +
    poly(a$Days_above_treshold_Tmin, 2) + poly(a$`min(Tmin)`,
    2) + a$fst_qtl_Tmin_year + poly(a$trd_qtl_Tmin_year, 2) +
    a$Days_above_treshold_Tmax + a$`min(Tmax)` + poly(a$Days_above_mean_Tmax,
    2) + poly(a$Days_above_treshold_Wspeed, 2) + poly(a$`mean Wspeed`,
    2) + poly(a$Days_above_mean_Wspeed, 2) + poly(a$fst_qtl_Wspeed_year,
    2) + poly(a$trd_qtl_Wspeed_year, 2) + a$`min thermal_amp` +
    a$`max thermal_amp` + poly(a$Days_above_mean_diff, 2))

Residuals:
    Min      1Q  Median      3Q     Max
-2.3879 -0.4536 -0.0604  0.5566  1.4839

Coefficients:
                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                             0.167894   0.944498   0.178 0.859036
poly(a$Days_above_treshold_Rtotal, 2)1  7.419055   3.950444   1.878 0.061384 .
poly(a$Days_above_treshold_Rtotal, 2)2 -3.538288   1.168208  -3.029 0.002677 **
a$`max(Rtotal)`                         0.001591   0.002578   0.617 0.537755
a$Days_above_mean_Rtotal               -0.022639   0.009975  -2.269 0.023979 *
a$trd_qtl_Rtotal_year                   0.222163   0.048627   4.569 7.28e-06 ***
poly(a$Days_above_treshold_Tmin, 2)1    6.915154   3.046184   2.270 0.023938 *
poly(a$Days_above_treshold_Tmin, 2)2   -6.178222   1.553796  -3.976 8.85e-05 ***
poly(a$`min(Tmin)`, 2)1                 6.300942   1.709928   3.685 0.000273 ***
poly(a$`min(Tmin)`, 2)2                 0.003619   1.149677   0.003 0.997490
a$fst_qtl_Tmin_year                     0.072244   0.064701   1.117 0.265096
poly(a$trd_qtl_Tmin_year, 2)1         -10.105726   2.577113  -3.921 0.000110 ***
poly(a$trd_qtl_Tmin_year, 2)2           6.106874   2.168081   2.817 0.005186 **
a$Days_above_treshold_Tmax              0.002209   0.004713   0.469 0.639639
a$`min(Tmax)`                          -0.170171   0.026274  -6.477 4.02e-10 ***
poly(a$Days_above_mean_Tmax, 2)1       -9.638227   3.223330  -2.990 0.003028 **
poly(a$Days_above_mean_Tmax, 2)2        7.141621   2.114873   3.377 0.000834 ***
poly(a$Days_above_treshold_Wspeed, 2)1 41.772859  14.029652   2.977 0.003153 **
poly(a$Days_above_treshold_Wspeed, 2)2  9.007680   5.011701   1.797 0.073327 .
poly(a$`mean Wspeed`, 2)1              86.830948  30.282434   2.867 0.004444 **
poly(a$`mean Wspeed`, 2)2             -14.724565   6.560676  -2.244 0.025565 *
poly(a$Days_above_mean_Wspeed, 2)1    -40.251476  11.231844  -3.584 0.000397 ***
poly(a$Days_above_mean_Wspeed, 2)2    -14.909125   6.259630  -2.382 0.017876 *
poly(a$fst_qtl_Wspeed_year, 2)1       -28.513744  12.478384  -2.285 0.023034 *
poly(a$fst_qtl_Wspeed_year, 2)2         6.630055   3.347048   1.981 0.048554 *
poly(a$trd_qtl_Wspeed_year, 2)1       -64.125821  16.752245  -3.828 0.000158 ***
poly(a$trd_qtl_Wspeed_year, 2)2         8.847217   4.054569   2.182 0.029912 *
a$`min thermal_amp`                    -0.044998   0.065260  -0.690 0.491044
a$`max thermal_amp`                     0.132830   0.027748   4.787 2.71e-06 ***
poly(a$Days_above_mean_diff, 2)1        6.888475   3.420732   2.014 0.044963 *
poly(a$Days_above_mean_diff, 2)2       -5.411029   2.127274  -2.544 0.011491 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7366 on 289 degrees of freedom
Multiple R-squared:  0.5345,	Adjusted R-squared:  0.4862
F-statistic: 11.06 on 30 and 289 DF,  p-value: < 2.2e-16
```

Figure 19. Results of Regressions on CoNC for Region D and all cultures

```
Call:
lm(formula = a$`Claims over Contracts` ~ poly(a$Days_above_treshold_Rtotal, 
    2) + poly(a$`max(Rtotal)`, 2) + poly(a$`mean Rtotal`, 2) + 
    poly(a$`min(Tmin)`, 2) + poly(a$`max(Tmin)`, 2) + a$`mean Tmin` + 
    a$Days_above_mean_Tmin + a$trd_qtl_Tmin_year + poly(a$Days_above_treshold_Tmax, 
    2) + a$`min(Tmax)` + poly(a$`mean Tmax`, 2) + a$Days_above_mean_Tmax + 
    a$fst_qtl_Tmax_year + poly(a$trd_qtl_Tmax_year, 2) + a$Days_above_treshold_Wspeed + 
    a$`min(Wspeed)` + a$`max(Wspeed)` + a$`mean Wspeed` + a$fst_qtl_Wspeed_year + 
    a$trd_qtl_Wspeed_year + poly(a$`max thermal_amp`, 2) + a$`mean thermal_amp`)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.53882 -0.41362  0.01923  0.40063  1.83831 

Coefficients:
                                       Estimate Std. Error t value Pr(>|t|)    
(Intercept)                            9.926208   2.297158   4.321 2.00e-05 ***
poly(a$Days_above_treshold_Rtotal, 2)1   8.541007   1.443689   5.916 7.50e-09 ***
poly(a$Days_above_treshold_Rtotal, 2)2  -2.098455   0.843228  -2.489 0.013263 *  
poly(a$`max(Rtotal)`, 2)1               0.862186   1.005529   0.857 0.391753    
poly(a$`max(Rtotal)`, 2)2              -2.928796   0.970062  -3.019 0.002710 ** 
poly(a$`mean Rtotal`, 2)1              -5.120707   1.356534  -3.775 0.000186 ***
poly(a$`mean Rtotal`, 2)2               7.381129   1.426614   5.174 3.76e-07 ***
poly(a$`min(Tmin)`, 2)1                 1.750316   1.355044   1.292 0.197264    
poly(a$`min(Tmin)`, 2)2                -3.421412   1.148716  -2.978 0.003087 ** 
poly(a$`max(Tmin)`, 2)1                -2.742586   1.175662  -2.333 0.020193 *  
poly(a$`max(Tmin)`, 2)2                -1.364809   1.191863  -1.145 0.252903    
a$`mean Tmin`                          -0.320573   0.140930  -2.275 0.023494 *  
a$Days_above_mean_Tmin                 -0.011730   0.003745  -3.132 0.001872 ** 
a$trd_qtl_Tmin_year                     0.401674   0.114867   3.497 0.000528 ***
poly(a$Days_above_treshold_Tmax, 2)1    9.871746   2.898747   3.406 0.000733 ***
poly(a$Days_above_treshold_Tmax, 2)2   -2.708523   1.418256  -1.910 0.056936 .  
a$`min(Tmax)`                           0.079241   0.024051   3.295 0.001080 ** 
poly(a$`mean Tmax`, 2)1                41.323210   8.306256   4.975 1.00e-06 ***
poly(a$`mean Tmax`, 2)2                 3.553766   2.163002   1.643 0.101235    
a$Days_above_mean_Tmax                 -0.006784   0.004399  -1.542 0.123882    
a$fst_qtl_Tmax_year                    -0.438836   0.103771  -4.229 2.96e-05 ***
poly(a$trd_qtl_Tmax_year, 2)1         -27.386925   4.825062  -5.676 2.78e-08 ***
poly(a$trd_qtl_Tmax_year, 2)2           4.913000   2.677670   1.835 0.067335 .  
a$Days_above_treshold_Wspeed            0.008455   0.003742   2.260 0.024431 *  
a$`min(Wspeed)`                         0.179901   0.049579   3.629 0.000325 ***
a$`max(Wspeed)`                         0.048748   0.014125   3.451 0.000622 ***
a$`mean Wspeed`                        -1.509985   0.357221  -4.227 2.98e-05 ***
a$fst_qtl_Wspeed_year                   0.753842   0.177615   4.244 2.77e-05 ***
a$trd_qtl_Wspeed_year                   0.539940   0.197443   2.735 0.006544 ** 
poly(a$`max thermal_amp`, 2)1           2.644320   2.185851   1.210 0.227147    
poly(a$`max thermal_amp`, 2)2          -3.532500   1.034512  -3.415 0.000709 ***
a$`mean thermal_amp`                   -0.183581   0.097958  -1.874 0.061705 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.655 on 371 degrees of freedom
Multiple R-squared:  0.3931,	Adjusted R-squared:  0.3424 
F-statistic: 7.753 on 31 and 371 DF,  p-value: < 2.2e-16
```

Figure 20. Results of Regressions on CoNC for Region E and all cultures