

Assessment of weather-related risk on chestnut productivity

M. G. Pereira^{1,2,3}, L. Caramelo^{1,3}, C. Gouveia^{2,4}, J. Gomes-Laranjo¹, and M. Magalhães⁵

¹Centre for Research and Technology of Agro-Environment and Biological Sciences (CITAB), University of Trás-os-Montes and Alto Douro, Vila Real, Portugal

²Instituto Dom Luiz – Universidade de Lisboa, Lisboa, Portugal

³Departamento de Física, Escola de Ciência e Tecnologia, Universidade de Trás-os-Montes e Alto Douro (DF-ECT UTAD) Vila Real, Portugal

⁴Escola Superior de Tecnologia, Instituto Politécnico de Setúbal, Setúbal, Portugal

⁵Departamento de Ciências Florestais e Arquitectura Paisagista, Escola de Ciência Agrárias e Veterinarias, Universidade de Trás-os-Montes e Alto Douro (DCFAP-ECAV UTAD) Vila Real, Portugal

Received: 31 January 2011 - Revised: 20 April 2011 - Accepted: 30 May 2011 - Published: 12 October 2011

Abstract. Due to its economic and nutritional value, the world production of chestnuts is increasing as new stands are being planted in various regions of the world. This work focuses on the relation between weather and annual chestnut production to model the role of weather, to assess the impacts of climate change and to identify appropriate locations for new groves. The exploratory analysis of chestnut production time series and the striking increase of production area have motivated the use for chestnut productivity. A large set of meteorological variables and remote sensing indices were computed and their role on chestnut productivity evaluated with composite and correlation analyses. These results allow for the identification of the variables cluster with a high correlation and impact on chestnut production. Then, different selection methods were used to develop multiple regression models able to explain a considerable fraction of productivity variance: (i) a simulation model (R^2 -value = 87 %) based on the winter and summer temperature and on spring and summer precipitation variables; and, (ii) a model to predict yearly chestnut productivity (R^2 -value of 63 %) with five months in advance, combining meteorological variables and NDVI. Goodness of fit statistic, cross validation and residual analysis demonstrate the model's quality, usefulness and consistency of obtained results.

1 Introduction

According to FAO statistics (FAO, 2010), Portugal was the sixth world's largest producer in 2008 with 22 000 tons; the world's largest producers are China (1000000 tons), South Korea (75000 tons) Italy and Turkey (55000 tons), and Japan (26 000 tons), but it should be noted that all these countries have a much higher land area than Portugal (Bounous, 2002b). The most noteworthy facts from world chestnut production trends in the last four decades are: (i) East Asian production continues to enlarge, mainly because of the great increase in the contribution of China (from 130000 tons in the 60s of the 20th century); (ii) a general decrease in the production in some western European countries (Portugal, 82000 tons in the 60s, Spain, France and Italy) and an increase in Turkey (40000 in the 60s); (iii) new orchards have been planted in Europe, North (USA) and South America (Brazil and Chile), Australia and New Zealand, due to the increase of the retail price for quality nuts and processed products and by the European Community funding programmes (Bounous, 2002b; Gomes-Laranjo et al., 2007).

Chestnut trees are also cultivated for their fruit and wood. With regards to the fruit, it is used in preparations of many recipes due to its dietetic value. Its wood is as strong as oak, but significantly lighter. In fact, results obtained by Jacobs et al. (2009) demonstrated that North American chestnut trees compete favourably in the aboveground allocation of biomass and carbon sequestration ability with any other species in this region. A chestnut agro-ecosystem also provides a habitat for diverse macrofungal species which support a high value of economical activity such as the mushroom harvesting (Baptista et al., 2010).



Correspondence to: M. G. Pereira (gpereira@utad.pt)

In Europe, the most important chestnut specie is Castanea sativa Mill., one of 13 species from Castanea genus. In relation to its phenology, bud break happens at the end of April, the flowering period is between June and July, being the last phase related to fruit development between August and October, time for fruit fall. This species, also called sweet chestnut, dislikes chalky soil, but appreciats sedimentary or siliceous soils. Their roots tend to decay in poorly drained soils, which help to explain why they prosper on hills and mountainsides. The European chestnut is cultivated for its nuts and wood and can be found on acidic to neutral soils, influenced by an oceanic climate which is characterised by annual mean values of sunlight spanning between 2400 and 2600 h and rainfall ranging between 600 and 1500 mm, mean annual temperature between 9 and 13 °C, 27 °C being the mean of the maximal temperature (Heiniger and Conedera, 1992; Gomes-Laranjo et al., 2008). According to Dinis et al. (2011), chestnut regions must have 1900-2200 °D between May and October. The degree-days (°D) is the sum of the temperature values in degrees Celsius with a base temperature of 6 °C (Cesaraccio et al., 2001; Dinis et al., 2011). Accordingly, in the Iberian Peninsula, this edaphoclimatic situation can be observed since sea level on seacoast regions to mountainous regions (between 600 and 1000 m a.s.l.) in the inner part of the continent. The influence of temperature and radiation on photosynthesis productivity in chestnut populations in Northeast Portugal was analysed by Gomes-Laranjo et al. (2006). Maximum photosynthetic activity occurs at 24-28 °C for adult trees, but exhibits termoinhibition when the air temperature is above 32 °C, which is frequent during summer (Gomes-Laranjo et al., 2006, 2008). All species of plants are dependent on the weather with regards to their production. However, only a few number of works have been published on weather dependence of chestnut production and none of references found on this subject intend to quantify and model portuguese chestnut production. Wilczynski and Podlaski (2007) concluded that the radial growth of horse chestnut (Aesculus hippocastanum L.) is positively related to high air temperature of August and during the previous winter and negatively related to excessive precipitation in August. The growing season is defined as the period of time when the mean 24-h temperature is greater than 5 °C. Fernández-López et al. (2005) study the geographic differentiation in adaptive traits of the wild chestnut populations in Spain resorting to climate data (e.g., temperature variation, summer precipitation/droughts and the temperature of the warmest month) and adding evidence of the role of some meteorological variables, namely frost during bud break, mean temperature of the warmest month, summer precipitation and drought, creating a xerothermic index. The influence of temperature and radiation on the photosynthesis productivity in chestnut populations in Northeast Portugal was analysed by Almeida et al. (2007).

At the same time, remote-sensing technology has been developing steadily and its products can provide many

applications in drought monitoring (Vicente-Serrano et al., 2006; Gouveia et al., 2009), agriculture, crop growth monitoring, yield modelling (Gouveia and Trigo, 2008) and crop identification. In particular, the time-series of satellite imagery efficiently provide a synoptic view of vegetation dynamics, namely the chestnut vegetative cycle that may be used for chestnut management. Phenological information is, in fact, essential for decision making during many of the phases of growing, namely on management planning, pest and disease control (Gouveia et al., 2011). In this context, several vegetation indices have been used in order to describe the phenology, namely the Normalized Difference Vegetation Index (NDVI) as derived from remote-sensed information. The NDVI was designed to capture the contrast between red and near-infrared reflection of solar radiation by vegetation, and is an indicator of the amount of green leaf area (Asrar et al., 1984; Myneni et al., 1995). Despite its simplicity, NDVI has been widely used in studies of vegetation phenology and interannual variability of vegetation greenness (Gouveia et al., 2008).

This work aims to identify the favourable/adverse weather conditions to chestnut production as well as helping to assess risk and to identify appropriate measures for adaptation to climate change. In this sense, the three specific objectives of this work are: (i) to characterise the chestnuts production in Portugal; (ii) to quantify the role of weather and climate on chestnut production; and, (iii) to develop simulation and prediction models of chestnut production based on meteorological variables and vegetation indices.

2 Study area and datasets

This work uses chestnut production, vegetation index and meteorological datasets which cover the 1982–2006 period and includes:

- annual values of the total chestnut production and the production area in Portugal, provided by the Portuguese National Institute of Statistics (INE, 2010) (http://www. ine.pt);
- daily values of several meteorological variables registered in the Bragança weather station, located at 41°47′28″ N and 6°45′43″ W, 740 m a.s.l. (Fig. 1), namely maximum, minimum and mean temperature, wind speed, total (rain and/or melted snow) precipitation, as well as a set of weather parameters indicative of the occurrence of hail, snow, fog and storm. This data was obtained from the Meteored web site (METEORED, 2009);
- Monthly Normalized Difference Vegetation Index (NDVI) dataset, at 8-km resolution, from the Advanced Very High Resolution Radiometers (AVHRR), provided by the Global Inventory Monitoring and Modelling System (GIMMS) group (GIMMS, 2009).



Fig. 1. Digital elevation model of Continental Portugal with spatial resolution of $1 \text{ km} \times 1 \text{ km}$. The colourbar represents the altitude (m). Location of Bragança weather station is identified with a black circle.

In Portugal, chestnut orchards may be found in the NE quarter of the country which is also characterised by irregular topography (Fig. 1) with Mediterranean Csb type of climate, according to the Koppen-Geiger climate classification (Peel et al., 2007). A description of climate, vegetation, soil specific characteristics and of this specific region can be found in Gomes-Laranjo et al. (2007, 2009).

The 5th National Forestry Inventory (NFI5), provided by the Portuguese National Forestry Authority (Autoridade Florestal Nacional, AFN) was used to determine the exact location of chestnut orchards in Portugal (AFN, 2010). The NFI5 was based on a digital aerial-photo coverage obtained during the 2004–2006 period and on a ground survey performed from December 2005 to June 2006 allowing the definition of homogeneous land parcels from the soil occupation standpoint.

Based on this information, the spatial distribution of chestnut orchards (as main occupation land use) located above 500 m were produced. Then, the location and density of chestnut trees was computed based on the number of Portuguese Forestry Inventory (NFI5) chestnut land parcels located inside each NDVI pixels, with a size of $8 \text{ km} \times 8 \text{ km}$ (Fig. 2).

The high density of chestnut trees is located in the administrative districts of Bragança and Vila Real, which constitutes the Trás-os-Montes and Alto Douro region, in an area of low mountains (Marão, Montezinho) and narrow valleys, in particular, in a sub-region characterised by low air temperature during the winter named Terra-Fria (Cold Land) (Fig. 1, right panel). Approximately 85 % (25 644 ha) of the 30 097 ha of the Portuguese chestnut tree area in 2006 was located in this region, which motivates the use of



Fig. 2. Location of the pixels $(8 \text{ km} \times 8 \text{ km})$ with chestnut trees as the main occupation with an altitude above 500 m. The colour bar represents the number of land parcels with chestnut trees as the main occupation, identified by the 5th National Forestry Inventory, NFI5 (AFN, 2010), inside a GIMMS pixel (GIMMS, 2009); only the pixels with more than 10 chestnut trees were selected.

meteorological data registered in Terra Fria. In addition, previous studies about chestnut weather-dependence were performed on chestnut orchards in the Trás-os-Montes region (Pires et al., 1995; Martins et al., 2005; Raimundo, 2003; Raimundo et al., 2001, 2004, 2009; Fonseca et al., 2004; Gomes-Laranjo et al., 2005, 2006, 2008). For these reasons, we decide to use the data from the Bragança weather station which is included in the meteorological observation national network, is well-situated in this region and can characterise weather local features.

Preliminary analysis on the meteorological dataset for Bragança weather station reveals only a small fraction of suspicious (extremely low) or missing values for temperature (0.45%). Missing values for wind speed, precipitation also accounts for a minute fraction of the total number of records (0.16, 0.21 and 0.47%, respectively). Monthly means and medians were evaluated taking into account the missing values. However, weather parameters, indicative of the occurrence of hail, snow, fog and storm, should be used carefully since they have a much larger amount of missing values and present some inconsistencies with meteorological variables.

Based on the evidence found in the literature about the meteorological influence on the chestnut production/productivity (Bounous, 2002a; Fernandez-López et al., 2005; Wilczynski and Podlaski, 2007; Gomes-Laranjo et al., 2006, 2008), we compute a large set of meteorological variables (e.g., maximum, minimum and mean temperature and precipitation averages for specific group of months) and meteorological parameters, such as monthly number of days with maximum, minimum and mean temperature above/below/between defined thresholds (e.g., number of days with minimum air temperature below 0 °C in January, N_{Davs} ($T_{\text{Min}}(1) < 0^{\circ}$ C), number of days in August with maximum air temperature above $32 \degree C$, $N_{\text{Days}}(T_{\text{Max}}(8) > 32 \degree C)$ or the number of days with maximum temperature between 24 °C and 28 °C in June, N_{Days} (24 °C < T_{Max} (6) < 32 °C) and monthly number of days with hail, snow, fog and storm. The numbers in parenthesis (separated by comas) represents the month when the data used to compute the variable was obtained.

The monthly NDVI dataset covers the area between 10° W to 0° E and 35° to 45° N, and respect to the 25-yr long period from 1982 to 2006. Details on the quality of GIMMS dataset may be found in Kaufmann et al. (2000) and Zhou et al. (2001). We have used the information about vegetation density from NFI5 in order to select the NDVI pixels in this region and corresponding to chestnut orchards (as main occupation land use) located above 500 m. (Fig. 2). The colour bar in Fig. 2 represents the number of land parcels with chestnut trees as the main occupant, identified by NFI5, inside a GIMMS-NDVI pixel; only the pixels with more than 10 chestnut trees were selected. Monthly composites of NDVI and corresponding anomalies for the considered period were accordingly computed for the period and pixels considered.

3 Method

After a preliminary quality and exploratory analysis of the raw data, composite and correlation analysis were used to identify the meteorological variables and parameters with the potential to influence chestnut production/productivity in Portugal. Composite analysis is used to recognize the meteorological variables that present significant differences between years of extreme positive and extreme negative chestnut production/productivity. Composite analysis is based on the arithmetic mean of the meteorological variables/parameters for selected yearly values of chestnut production/productivity and, eventually, it is followed by an anomaly analysis (defined as the difference between the composite and the arithmetic mean evaluated using all records) or by the assessment of the relative difference (RD) between composites obtained for extreme positive (C^+) and extreme negative years (C^{-}), defined as RD = ($C^{+} - C^{-}$)/ C^{-} . Composite analysis is widely and lengthily used in atmospheric sciences and climatic research (Jury and Pathack, 1991; Bauer and Del Genio, 2006; Pereira et al., 2005; Trigo et al., 2006).

The extreme positive/negative years were defined as the years for which production or productivity time series are greater/smaller or equal to the time series arithmetic mean value plus/minus the standard deviation. Using this criterion, the years of 1989, 1990, 2000 and 2003 were classified as extreme positive while the years of 1991, 1992 1993 and 2005 were classified as those that present extreme negative productivity. The criterion used to identify the extreme years is shown to be very suitable as it allows the classification of the same number of positive and negative extreme years that are relatively well spanned within the study period. Other criteria (e.g., production or productivity above/below previous time series value plus/minus 10%) were tested, but no significant changes were found when the same number of extreme years had to be considered.

The correlation analysis objective is to measure the strength of the linear relationship between chestnut production/productivity and meteorological variables/parameters through the evaluation of the Spearman correlation coefficient. This complementary technique was applied for different subperiods (e.g., 1982–1990, 1991–1999, 1999–2006), for moving subperiods with a constant length (e.g., 5, 10 and 15 yr) within the period of analysis, for periods with increasing and decreasing length, onward and backward, starting from the first and last years.

After the identification of the potential meteorological predictors, we use SAS software (Statistical Analysis System, v9.1.3) to develop a multiple linear regression model to simulate and to predict chestnut production/productivity with different selection methods (e.g., stepwise, forward and backward). A comprehensive description of these predictor selection methods can be found in Austin and Tu (2004), Miller (1984, 2002) and Hocking (1976). In this context,

Fernandez-López et al. (2005) have used a linear regression analysis between Spanish chestnut population performance, climatic and geographic data.

Among other assumptions, linear regression requires that predictors are linearly independent (collinearity), samples are representative of the population and that the error terms have zero mean, constant variance (homoscedasticity), normality (for hypothesis testing purposes) and be uncorrelated. Shapiro-Wilk test, Kolmogorov-Smirnov test as well as Lillefors test (which is an adaptation of the Kolmogorov-Smirnov test) will be used to test if the null hypothesis that data come from a normally distributed population. On the other hand, predictors and predictant frequently present a natural sequence (e.g., weather parameters), which means that the errors in time series data exhibit serial correlation, i.e., are autocorrelated. The Durbin-Watson statistic (d) is used to test the presence of autocorrelation and can assume values between 0 and 4. When d = 2 is indicative of no autocorrelation while, when d < 2 there is evidence of positive serial correlation, which means that succeeding error terms are, on average, similar to one another. On the other hand, when d > 2 the following error terms are, typically, much different to one another, i.e., negatively correlated which can imply an underestimation of the level of statistical significance.

To assess the goodness-of-fit, residual analysis was performed and several statistics were evaluated. Since the objective of this work is to study the influence of weather on chestnut productivity, it is important to consider the coefficient of multiple determination (R^2) which accounts how successful the fit is in explaining the variation of the data, that is, how much chestnut productivity time series variance is explained by the developed regression model. Adding predictors to the model, in general, increases the R^2 value, but not necessarily the usefulness of the model, in the sense of the prediction of future outcomes. To take into account eventual over fitting, R^2 should be adjusted (R^2_{Adj}) taking into account the residual degrees of freedom, which can be used for proper comparison between models with different numbers of independent variables.

In addition, other usual statistics were also determined, namely: the sum square error (SSE), the mean square error (MSE) or root mean square error, and Mallows' CP. The RMSE is just the square root of MSE which is defined as the quotient between SSE and residual degrees of freedom. Values of RMSE, MSE and SSE closer to 0 are indicative of better models (useful for prediction). Mallows' CP, (Mallows, 1973) can be used as a predictor selection criterion and to assess the model fit, in particular, with respect to over fitting since the estimates of the mean squared prediction error does not necessarily decrease as more variables are included in the model like other error measures (e.g., SSE).

Statistical models, developed with relatively short time series, are particularly prone to overfitting problems (Wilks, 1995), to solve this caveat it is advisable to apply cross-validation techniques, i.e., to split the available time series into calibration and validation periods. The evaluation of model performance and the prevention of overfitting was done by means of leave-one-out cross-validation technique, i.e., by using a single observation from the original sample as the validation data and the remaining observations as the training data.

4 Results

Apparently, the location of orchards are determined by soil and climate conditions. In continental Portugal, 86% of the total number of pixels, with chestnut trees as the main soil occupant, is located between 500 and 1000 m in altitude, where the chestnut trees may have found suitable conditions for their development, which helps to explain the strong resemblance between the location of chestnut orchards (Fig. 2) and the Portuguese elevation map (Fig. 1). The great majority of the chestnut production comes from areas of higher altitude namely the Terra Fria included in the Trás-os-Montes and Alto Douro region, where the landscape is dominated by the low slopes of the plateau Transmontano. High density of chestnut trees in Portugal and in Spain are found in the same geographical region, extending from NE Portugal and Galicia to Navarra, through coastal NW Spain (Fernandez-López et al., 2005). The exceptional increase in the production area time series between 1991 and 1999 (Fig. 3) has a profound impact on the chestnut production time series variability reflected in the increase of the average chestnut production from 18000 ton during the first third of the period to 30 000 ton in the last third, which has been related to European Community funding programmes (EEC Regulation N° 2080/92). This fact does not allow a proper comparison between chestnut production values in different periods which induce the use of chestnut productivity instead of chestnut production in our analysis.

In general, it is assumed that the trend in the chestnut productivity time series can be due to factors that do not remain constant during the study period such as the introduction of new agriculture techniques, pesticides, laws and government policies, crop diseases and plagues (Portela et al., 1999; Gentile et al., 2009; Ghezi et al., 2010). In fact, chestnut diseases, such as ink and cancer, which reached Portugal in the 19th century and ended in the 20th century, respectively, could be among the possible reasons for the long-term linear decreasing tendency (Kiple and Ornelas, 2000). Removing these trends, we are expecting that variability of the detrended chestnut productivity is only due to climate variability, since changes in climate or soil are expected to have impacts on much longer time scales. A similar procedure was followed to correct the chestnut tree-ring chronology which shows a constant decreasing trend as trees became older (Wilczynski and Podlaski, 2007).



Fig. 3. Observed values of: (a) chestnut production, (in $\times 10^3$ ton) and production area (in $\times 10^3$ ha); and, (b) chestnut productivity and detrended chestnut productivity (in ton ha⁻¹), in Portugal.

The obtained productivity time series (Fig. 3) present an outlier in the year of 1993 for two main reasons: unfavourable climate conditions and inertial delay in the effect of production area increasing. In fact, in this year, the mean air temperature during summer (June, July, August and September) was $2.7 \,^{\circ}$ C below the average. In addition, chestnut production in 1993 seems to follow the decreasing trend registered in the previous period since the effect of the increase of the production area is not yet felt because it is unlikely that the recently planted chestnut trees are at their maximum production capacity.

4.1 Chestnut production characteristics and potential predictors

The chestnut production area in Portugal is not constant during the study period. In the 1982–1990 and 1999–2006 subperiods, the production area time series presents a small increasing trend of 195 ha yr⁻¹ ($R^2 = 97$ %) and 215 ha yr⁻¹ ($R^2 = 94$ %), respectively, without statistical significance. However, it is evident an abrupt positive trend between 1991 and 1999 of 1.5×10^3 ha yr⁻¹, ($R^2 = 99$ %, statistical



Fig. 4. Annual cycle of monthly NDVI values for the standard hydrological year spanning from September of year n - 1 to August of year n (with n from 1982 to 2006). The bottom/top indicates the lower/upper quartiles, and the band near the middle of the box is the median. The lower/upper end of the whiskers represents the lowest/highest observed value still within 1.5 of the interquartile range of the lower/upper quartile.

significant at 97.73 % level), when the production area almost doubled its value, from 15 000 ha to 29 000 ha (Fig. 3a). To circumvent this difficulty, we decide to analyse the chestnut productivity instead of the chestnut production. Chestnut productivity is defined as the ratio between yearly chestnut production and correspondent yearly production area (Fig. 3b).

The productivity time series presents a linear negative trend of -13.5×10^{-3} ton ha⁻¹ yr⁻¹ (statistically significant at 99.50 % level) which corresponds to an approximate decrease of 300 ton yr⁻¹, if the arithmetic mean value of the production area of 22 000 ha is considered. On the other hand, it is expected that the detrended time series variability is only due to factors that presents constant variability during the study period, such as meteorological variables/parameters and soil characteristics (Cantelaube et al., 2004; Gouveia and Trigo, 2008). For these reasons, we decided on the detrended chestnut productivity and, hereafter, all the following results are referred to the Corrected Productivity (CP) time series (Fig. 3b).

The annual cycle of NDVI monthly mean values is represented in Fig. 4 by means of boxplots and can be used to illustrate the intra-annual variability of chestnut productivity. In this case, we have adopted the standard hydrological year, spanning from September of year n-1 to August of year n (with n from 1982 to 2006). The bottom/top indicates the lower/upper quartiles, and the band near the middle of the box is the median. The lower/upper end of the whiskers represents the lowest/highest observed value still within 1.5 of the interquartile range of the lower/upper quartile. It may be noted that winter shows a higher variability than in summer, presenting February and May with the most concentrated distribution. The NDVI cycle presents the maximum in early summer (June) and the minimum during winter (December), a feature that is related to the vegetative cycle of chestnut.

Based on preliminary results obtained in the composite and correlation analysis, additional meteorological parameters were computed to obtain additional parameters that present higher correlation coefficient with the CP time series for the reason that the SAS software tends to select the variables/parameters which present high correlation with CP. For example, accumulated precipitation in April, P(4) and July, P(7) were merged into the accumulated precipitation in those two months, P(4,7) aiming to have a variable with higher correlation coefficient with CP than P(4) or P(7). Failing to present all results from the composite and correlation analysis, we decided to present only those for a set of selected variables (Table 1).

Composite analysis reveals that, in general, higher values of the relative difference between composites obtained for extreme positive and extreme negative years were obtained for precipitation, minimum temperature and parameters based on the number of days respecting some criteria, than for NDVI, mean or maximum air temperature. This finding is consistent with the nature of these variables in the sense that variables of the former set have a more irregular behaviour and, even the same absolute changes produce lower relative changes in the latter set. Some of the most significant results obtained from composite analysis were found for N_{Days} ($T_{\text{Min}}(2) < 10^{\circ}$ C), $T_{\text{Min}}(1,2)$, $N_{\text{Days}}(24^{\circ}$ C $< T_{\text{Max}}(3) < 28^{\circ}$ C), P(12), P(4) and P(7) with RD of -179%, 123 %, 100 %, 69 %, 67 % and 60 %, respectively.

Results of the correlation analysis does not exactly match those from composite analysis, in the sense that variables/parameters with the highest RD do not present the highest correlation coefficient with CP. In most cases, it is possible to obtain higher values of the correlation coefficient for variables using data from two or more months than for monthly variables/parameters. This fact motivated the use of data from several months to get better correlated variables with CP. In fact, meteorological variables/parameters which present high correlation coefficient for the 1982-2006 analysis period $T_{\text{Max}}(1,9), T_{\text{Max}}(1,2,9), N_{\text{Days}}(T_{\text{Max}}(11) < 28 \,^{\circ}\text{C}),$ are: $N_{\text{Days}}(24 \,^{\circ}\text{C} < T_{\text{Max}}(5,6,7) < 28 \,^{\circ}\text{C}), P(4,7) \text{ and } T_{\text{Mean}}(2,9)$ with ρ equals to 63 %, 60 %, 56 %, 49 %, 48 % and 48 %, respectively. The same type of variable/parameters, e.g., accumulated precipitation, could present positive correlation coefficient for one period of the year and negative correlation for another, reflecting the role of precipitation in the different moments of the vegetative cycle.

4.2 Chestnut productivity simulation and prediction models

All the meteorological variables/parameters produced were tested with the predictor selection/elimination methods in the regression analysis and the best simulation regression model (R^2 -value = 87 %) of the detrended chestnut productivity

time series (CP_{sim}), is obtained with six meteorological predictors selected as follows,

$$CP_{sim} = 0.012 \times N_{Days} (T_{Max}(9) < 28^{\circ}C) -0.007 \times N_{Days} (T_{Min}(2) < 10^{\circ}C) + +0.06 \times T_{Max}(1,9) +0.035 \times T_{Max}(7) +0.002 \times P(4,7) + 0.001 \times P(9) - 2.268$$
(1)

where N_{Days} ($T_{\text{Max}}(9) < 28 \,^{\circ}\text{C}$) represent the number of days in September with maximum air temperature below 28 °C (# of days), N_{Days} ($T_{\text{Min}}(2) < 10^{\circ}$ C) represent the number of February days with minimum air temperature below 10°C (# of days), $T_{Max}(1,9)$ is the mean maximum air temperature in the months of January and September (°C), $T_{Max}(7)$ is the mean maximum air temperature in the month of July (°C), P(4,7) and P(9) accounts for the accumulated precipitation (mm) in the months of April and July, and September, respectively. This model is obtained with both forward and stepwise selection methods and all variables included are statistical significant at 0.05 level. Observed and simulated CP obtained with the more robust approach from the leaveone-out cross validation are shown in Fig. 5, while values of goodness-of-fit measures are presented in Table 2. The results of the composite and correlation analysis for these predictors are shown in Table 1. The good agreement between the modelled time series by the regression model and the one obtained by the cross-validation (R = 0.99) is an indication of the robustness of the developed model. This is further supported by the relatively slight decrease (from 0.93 to 0.88) of the correlation between the original and the two modelled time series (by simple regression and by cross-validation).

With the objective to develop a chestnut productivity model with prediction capability, we repeat the process but restricting the meteorological variables/parameters predictors to those that use basic information before the month of June, which means, at least, five months before the collecting period. Regression analysis, results in a model to predict the detrended chestnut productivity time series (CP_{pred}) with an R^2 -value of 63 %, based on four predictors as follows,

$$CP_{pred} = 0.019 \times T_{Mean}(1, 2, 3) +0.002 \times P(4) - 0.002 \times P(5) +0.001 \times NDVI500(5) + 0.146$$
(2)

where, $T_{\text{Mean}}(1,2,3)$ is the mean air temperature in the months of January, February and March (°C), P(4) and P(5)are the precipitation registered in the months of April and May (mm), respectively, and, finally, NDVI500(5) accounts for the NDVI for the pixels with chestnut trees as the main occupant located above 500 m in May. Observed (CP_{obs}) and simulated (CP_{pred}) chestnut productivity with the leave-oneout cross-validation model is shown in Fig. 5, while model robustness measures are presented in Table 2. Results of the composite and correlation analysis for these predictors are

Table 1. Values of: (i) the relative difference between composites obtained for positive and negative years; and, (ii) the Spearman correlation coefficients between selected meteorological parameters and corrected chestnut productivity. T_{Mean} , T_{Max} and T_{Min} are the average of the mean, maximum and minimum air temperature; *P* is the accumulated precipitation, N_{Days} accounts for the number of days that meets the condition in brackets and NDVI500 represents the NDVI for the pixels with the chestnut trees as the main occupant located above 500 m. All variables/parameters were computed for a specific month or group of months indicated in parenthesis.

Meteorological parameters	Composite analysis	Correlation analysis
	$RD = (C^+ - C^-)/C^- (\%)$	1982-2006
$T_{\text{Mean}}(1,2,3)$	21.4	30
$T_{Max}(1,9)$	8.7	63
$T_{\text{Max}}(7)$	-0.3	7
<i>P</i> (4,7)	66.3	48
P(4)	66.7	46
<i>P</i> (5)	-29.2	-30
<i>P</i> (9)	-18.2	1
$N_{\rm Days} (T_{\rm Min}(2) < 10^{\circ}{\rm C})$	-179	-26
$N_{\rm Days} (T_{\rm Max}(9) < 28 ^{\circ}{\rm C})$	0.0	-10
NDVI500 (5)	6.9	30
NDVI500 (1)	2.5	27

Table 2. Values of the goodness-of-fit statistics (coefficient of multiple determination, $-R^2$ -, adjusted R^2 , $-R^2_{Adj}$ -, Mallows' CP, Durbin Watson, *d*, sum square error, SSE, and the mean square error, MSE) for simulation and prediction chestnut productivity multiple regression models.

Goodness-of-fit statistics	<i>R</i> ²	$R_{\rm adj}^2$	СР	d	MSE	SSE
Simulation Model	0.8651	0.8201	7.0000	2.272	0.0035	0.0633
Prediction Model	0.5755	0.4907	5.0000	2.034	0.0100	0.1993

presented in Table 1. For this model, cross-validation results are slightly lower, but what is still noticeable is the great resemblance between both modelled time series (R = 0.97) and the small reduction (from 0.79 to 0.65) of the correlation between the observed and the two modelled time series.

Linear regression assumptions were not violated given that: complete chestnut productivity time series was used in this work; exploratory analysis of the residuals reveals that residuals of both models have zero mean and constant variance; all normality tests performed (Shapiro-Wilk, Kolmogorov-Smirnov and Lilliefors) confirm that the null hypothesis that residuals of both models have a normal distribution cannot be rejected. In addition, values of the Durbin-Watson statistic (d) for both models are not very different from the value (d=2) of uncorrelated error terms but, while for simulation model it could be indicative of negative autocorrelation (d = 2.272), for prediction model (d = 1.855) it could be a sign of positive autocorrelation. A test for positive and negative autocorrelation (which are not frequent), at significance α , is based on the comparison of statistic (4-d)with lower and upper critical values ($d_{\rm L}$, α and $d_{\rm U}$, α), depending on the sample size and the number of regressors. Results for simulation model $[d_{L0.05} < (4-d) < d_{U0.05}]$ reveal that there is no statistical evidence that the error terms are negatively nor positively correlated while for prediction model $[(4-d) > d_{U0.05}]$ there is statistical evidence that the error terms are not negatively autocorrelated.

5 Discussion and conclusions

Composite and correlation analysis allows the identification of the meteorological variables/parameters with the highest potential to have impacts on the chestnut productivity. Actually, variables with the highest RD and correlation coefficient with CP, were selected by the selection methods during the regression analysis. However, it is clear that not only variables with high RD and, simultaneously, highly correlated with CP were selected as predictors of simulation and prediction models.

The regression analysis results obtained are dependent on the size of the pool of potential predictors used in the analysis, as well as on the selection method adopted. In fact, the use of a backward elimination method on a sufficient large pool of predictors, allows us to obtain a perfect model



Fig. 5. Values of observed (dashed with solid diamonds) and modelled with cross-validation (CV) with simulation (solid with grey circles) and prediction (solid light grey with white diamonds) model of detrended chestnut productivity in Portugal, for the 25-yr period, defined from 1982 to 2006. R^2 -values between observed and modelled time series with CV are also shown.

(R^2 -value of 100%) for both the simulation and prediction of CP. However, these models make use of a large number of predictors which is not acceptable from the statistical or physical point of view. In order to obtain more realistic models, instead of reducing the number of potential predictors with any ad-hoc criterion, we decide to restrict the solutions provided by the forward and stepwise selection methods.

It should be underlined that all the predictors retained in the models by both selection methods are statistically significant at the 0.0500 level and do not employ the same basic information (for example, variables such as $T_{\text{Mean}}(7,8,9)$, i.e., the mean air temperature in July, August and September and $T_{\text{Mean}}(7)$, i.e., the mean temperature in July, were not used simultaneously in the model. In addition, the presented models are parsimonious and useful, in the sense that it does not include a large number of predictors and present the highest R^2 -value which means that these models explains the highest percentage of chestnut productivity inter-annual variability, with the predictors tested.

Furthermore, it is possible to find a phenological interpretation for each predictor used in the models. In fact, it should be expected that high productivity is associated with a warm and relatively long growing season and a mild winter (Wilczynski and Podlaski, 2007), which is explained by the positive values of the estimated parameters for $T_{\text{Max}}(1,9)$, $T_{\text{Max}}(7)$, $T_{\text{Mean}}(1,2,3)$ and the negative value estimated for the parameter of $N_{\text{Days}}(T_{\text{Min}}(2) < 10 \text{ °C})$. This result is also associated with the fact that high chestnut production requires, at least, 6 months with mean air temperature above 10 °C (Gomes-Laranjo et al., 2008). On the other hand, $N_{\text{Days}}(T_{\text{Max}}(9) < 28 \text{ °C})$ predictor evaluation was based on the temperature range where chestnut trees have a maximum photosynthetic activity and reflects the inexistence of

abnormal high temperatures in September, which allows the growth of the nuts and avoids the thermo inhibition of the trees (Gomes-Laranjo et al., 2005, 2006). The use of several precipitation predictors (P(4), P(5), P(4,7) and P(9)) in both simulation and prediction models reflects the importance of rainfall for chestnut productivity. The relative abundance of precipitation in April and May provides the appropriate soil humidity conditions that favours budbreak. On the other hand, the occurrence of precipitation in July and September, during the chestnut development, also reflects the existence of mild temperature during the summer and, since precipitation in these months are usually of small amounts, does not compromise the flushing. In fact, summer drought/precipitation was also identified as an important factor in the relation between flushing and its coefficient of variation in Spanish chestnut orchards (Fernandez-López et al., 2005). Finally, NDVI500(5) reflects the physiological state of chestnut trees in the end of Spring, that results from the combined effect of mild temperature and the relative abundance of precipitation during the previous winter and early spring.

Near zero value of the sum of squared error and of the mean squared error reveals that simulation and prediction model have small random error component and that are useful for prediction. The small decrease of the adjusted R^2 in relation to R^2 for simulation (4.5%) and prediction model (8.5%) as well as the continuously decreasing values of Mallow's CP during predictor's selection process suggest that both models should not be especially affected by over fitting. For models not suffering from appreciable lack of fit (bias), the CP values should be similar to the number of predictors, which is the case for both models. Better results obtained for a simulation model are unsurprisingly due to the use of additional and more pertinent predictors which is reflected by higher values of R^2 .

In summary, the main conclusions from this work are: (i) time series of chestnut production, and production area present statistical characteristics that led to the use of productivity time series; (ii) the use of composite and correlation analysis allows the identification of the meteorological parameters with high impact on chestnut productivity, in good agreement with previous results; (iii) regression analysis enables the selection of the predictors to be used in simulation and prediction models; (iv) all the predictors retained are statistically significant at 0.05 level and obtained models are simple and parsimonious (linear and with few parameters); (v) productivity time series is well reproduced by simulations which means that weather (during all the vegetative cycles of the chestnut trees) explains a relatively high percentage of original time series variance; (vi) all regression verification procedures (goodness-of-fit, residual analysis and crossvalidation results confirms the quality, usefulness and robustness of the models.

The establishment of the relation between weather and chestnut productivity is not a simple task due to several

factors. Production depends on the production area, but also on many other factors that were not taken into account in the work such as chestnut variety (Gomes-Laranjo et al., 2006), tree age, altitude of the orchards and solar exposition (Almeida et al., 2007), government policies, soil degradation (Portela et al., 1999, Raimundo 2009), chestnut tree diseases – ink and cancer – which seems to be responsible for the decreasing trend in production and, consequently, in productivity (in periods when production area remains constant).

Particular unfavourable weather conditions in relatively short periods (weekly scale) and on specific regions can occur without having been "detected" in a monthly temporalbase analysis and can plight the results obtained. Furthermore, combining total (country) values of chestnut production and production area with meteorological data from one specific location, even if representative of the region of the majority of the production, constitute an additional difficulty that could be circumvented if chestnut production data were available just for that region.

Finally, this study points out the need for further work to analyse the small-scale temporal and spatial effects, namely, to consider a smaller study area, to study the production of a single variety or varieties with similar weather dependence, to analyse data from orchards unaffected by diseases or to perform the analysis on a smaller time-scale (weekly). However, results from this work can be very useful to chestnut producers, related industry and agricultural/forest policy makers in the sense that the developed models provide useful information about how the weather factors control the chestnut annual production. Furthermore, in spite of the caveats and limitations of this study, the simulation model is able to reproduce 87 % of chestnut productivity in Portugal while a prediction model is able to estimate the chestnut production with more than five months in advance with a R^2 -value of 63 % which could constitute a firm basis for the assessment of climate change impacts on chestnut production.

Acknowledgements. The Satellite data was kindly provided by the Global Inventory Monitoring and Modelling System (GIMMS). We are thankful to the Portuguese National Forestry Authority (Autoridade Florestal Nacional) for providing access to the 5th National Forestry Inventory (NFI5) data.

Edited by: A. M. Tarquis Reviewed by: R. Moratiel and E. Vidal Vázquez

References

- AFN: available at: http://www.afn.min-agricultura.pt/portal/ifn, last access: December 2010.
- Almeida, P., Dinis, L. T., Coutinho, J., Pinto, T., Anjos, R., Ferreira-Cardoso, J., Pimentel-Pereira, M., Peixoto, F., and Gomes-Laranjo, J.: Effect of temperature and radiation on photosynthesis productivity in chestnut populations (Castanea sativa Mill. cv. Judia), Acta Agron. Hung., 55, 193–203, 2007.

- Asrar, G., Fuchs, M., Kanemasu, E. T., and Hatfeld, J. L.: Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat, Agron. J., 76, 300–306, 1984.
- Austin, P. C. and Tu, J. V.: Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality, J. Clin. Epidemiol., 57, 1138–1146, 2004.
- Baptista, P., Martins, A., Tavares, R. M., and Lino-Neto, T.: Diversity and fruiting pattern of macrofungi associated with chestnut (Castanea sativa) in the Trás-os-Montes region (Northeast Portugal), Fungal Ecol., 3, 9–19, 2010.
- Bauer, M. and Del Genio, A. D.: Composite Analysis of Winter Cyclones in a GCM: Influence on Climatological Humidity, J. Climate, 19, 1652–1672, 2006.
- Bounous, G.: Il castagno: coltura, ambiente ed utilizzazione in Italiae nel mondo, Ed. Agricole, Bologna, Italy, 2002a.
- Bounous, G.: Inventory of Chestnut Research, Germplassm and References, FAO, Rome, Italy, 2002b.
- Cantelaube P., Terres, J.-M., and Doblas-Reyes, F. J.: Influence of climate variability on European agriculture – analysis of winter wheat production, Clim. Res., 27, 135–144, 2004.
- Cesaraccio, C. D., Duce, P., and Snyder, R. L.: An improved model for determining degree-day values from daily temperature data, Biometeorology, 45, 161–169, 2001.
- Dinis, L. T., Peixoto, F., Pinto, T., Costa, R. Bennett, R. N., and Gomes-Laranjo, J.: Study of morphological and phenological diversity in chestnut trees ("Judia" variety) as a function of temperature sum, Environ. Exp. Bot., 70, 110–120, 2011.
- FAO: available at: http://faosat.fao.org/default.aspx, last access: January 2010.
- Fernández-lópez, J., Zas, R., Diaz, R., Villani, F., Cherubini, M., Aravanopoulos, F. A., Alizoti, P. G., Eriksson, G., Botta, R., and Mellano, M. G.: Geographic variability among extreme European wild chestnut populations, Acta Hort., 693, 181–186, 2005.
- Fonseca, T. F., Abreu, C. G., and Parresol, B. R.: Soil compaction and chestnut ink disease, Forest Pathol., 34, 273–283, doi:10.1111/j.1439-0329.2004.00371.x, 2004.
- Gentile, S., Valentino, D., and Tamietti, G.: Control of ink disease by trunk injection of potassium phosphite, J. Plant Pathol., 91, 565–571, 2009.
- Ghezi, E., Khodaparast, S. A., and Zare, R.: Distribution and severity of damage by Cryphonectria parasitica in the chestnut stands in Guilan province, Iran, Forest Pathol., 40, 450–457, 2010.
- GIMMS: available at: http://www.glcf.umd.edu/data/gimms/, last access: July 2009.
- Gomes-Laranjo, J., Salgado, P., Sang, H. W. W., Kraayenhof, R., and Torres-Pereira, J.: Isolation of chestnut chloroplasts: Membrane potentials of chestnut and spinach thylakoids, Photosynthetica, 43, 237–246, 2005.
- Gomes-Laranjo, J., Peixoto, F., and Sang, H. W. W. F.: Study of the temperature effect in three chestnut (Castanea sativa Mill.) cultivars' behaviour, J. Plant Physiol., 163, 945–955, 2006.
- Gomes-Laranjo, J., Ferreira-Cardoso, J., Portela, E., and Abreu, C. G.: Castanheiros, Universidade Trás-os-Montes e Alto Douro, Vila Real, Portugal, 2007.
- Gomes-Laranjo, J., Coutinho, J. P., Galhano, V., and Ferreira-Cardoso, J. V.: Differences in photosynthetic apparatus of leaves from different sides of the chestnut canopy, Photosynthetica, 46, 63–72, 2008.

- Gomes-Laranjo, J., Peixoto, F., Costa, R., and Ferreira-Cardoso, J.: Portugal, Following chestnut footprints, Cultivation and culture, folklore and history, traditions and uses, Scripta Horticulturae, 9, 106–111, 2009.
- Gouveia, C. and Trigo, R. M.: Infuence of Climate Variability on Wheat Production in Portugal, in: geoENV VI – Geostatistics for Environmental Applications, edited by: Soares, A., Pereira, M. J., and Dimitrakopoulos, R., Springer, 335–345, 2008.
- Gouveia, C., Trigo, R. M., DaCamara, C. C., Libonati, R., and Pereira, J. M. C.: The North Atlantic Oscillation and European Vegetation Dynamics, Int. J. Climatol., 28, 1835–1847, 2008.
- Gouveia, C., Trigo, R. M., and DaCamara, C. C.: Drought and vegetation stress monitoring in Portugal using satellite data, Nat. Hazards Earth Syst. Sci., 9, 185–195, doi:10.5194/nhess-9-185-2009, 2009.
- Gouveia, C., Liberato, M. L. R., DaCamara, C. C., Trigo R. M., and Ramos, A. M.: Interannual variability of wine production along the Douro valley in Portugal, Clim. Res., 48, 349–362, doi:10.3354/cr01006, 2011.
- Heiniger, U. and Conedera, M.: Chestnut forests and chestnut cultivation in Switzerland, in: Proceedings of the International Chestnut Conference, West Virginia University, Morgantown, 10–14 July 1992, 175–178, 1992.
- Hocking, R. R.: The analysis and selection of variables in linear regression, Biometrics, 32, 1–49, 1976.

INE: available at: http://www.ine.pt, last access: December 2010.

- Jacobs, D. F., Selig, M. F., and Severeid, L. R.: Aboveground carbon biomass of plantation-grown American chestnut (Castanea dentata) in absence of blight, Forest Ecol. Manage., 258, 288– 294, 2009.
- Jury, M. R. and Pathack, B.: A study of climate and weather variability over the tropical southwest Indian Ocean, Meteorol. Atmos. Phys., 47(1), 37–48, doi:10.1007/BF01025825, 1991.
- Kaufmann, R. K., Zhou, L., Knyazikhin, Y., Shabanov, N., Myneni, R. B., and Tucker, C. J.: Effect of orbital drift and sensor changes on the time series of AVHRR vegetation index data, IEEE T. Geosci. Remote, 38, 2584–2597, 2000.
- Kiple, K. F. and Ornelas K. C.: The Cambridge World History of Food, Cambridge University Press, New York, USA, 2000.
- Mallows, C. L.: Some Comments on CP, Technometrics, 15, 661– 675, doi:10.2307/1267380, 1973.
- Martins, A., Linhares, I., Raimundo, F., Borges, O., Coutinho, J. P., Gomes-Laranjo, J., and Sousa, V.: The importance of deep soil layers to supply water to agro-forestry systems: a case study of a mature chestnut orchard in northern Portugal, Acta Horticulturae, 693, 663–670, 2005.
- METEORED: available at: http://www.meteored.com/, last access: December 2009.
- Miller, A. J.: Selection of subsets of regression variables, J. R. Stat. Soc. Ser. A, 147, 389–425, 1984.
- Miller, A. J.: Subset selection in regression, Chapman & Hall/CRC, Boca Raton (FL), USA, 2002.

- Myneni, R. B., Hall, F. G., Sellers, P. J., and Marshak, A. L.: The interpretation of spectral vegetation indexes, IEEE T. Geosci. Remote, 33, 481–496, 1995.
- Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci., 11, 1633–1644, doi:10.5194/hess-11-1633-2007, 2007.
- Pereira, M. G., Trigo, R. M., DaCamara, C. C., Pereira, J. M. C., and Leite, S. M.: Synoptic patterns associated with large summer forest fires in Portugal, Agr. Forest Meteorol., 129, 11–25, 2005.
- Pires, A. L., Portela, E., and Martins, A.: Nutrient cycling in chestnut groves in the Trás-os-Montes region, Sustainability of Mediterranean Ecosystems, Case Study of the Chestnut Forest, European Commission, Ecosystem Research Report 19, 1995.
- Portela, E., Aranha, J., Martins, A., and Pires, A. L.: Soil factors, farmers practices and chestnut ink disease: Some interactions, Acta Horticulture, 494, 433–441, 1999.
- Raimundo, F.: Sistemas de Mobilização do Solo em Soutos. Influência na Produtividade de Castanha e nas Características Físicas e Químicas do solo, Ph.D. thesis, Universidade de Trásos-Montes e Alto Douro, Vila Real, Portugal, 2003.
- Raimundo, F., Branco, I., Martins, A., and Madeira, M.: Efeito da intensidade de preparação do solo na biomassa radical, regime hídrico, potencial hídrico foliar e produção de castanha de soutos do Nordeste Transmontano, Revista de Ciências Agrárias, 24, 415–423, 2001.
- Raimundo, F., Madeira, M., Coutinho, J., and Martins, A.: Efeito de sistemas de mobilização do solo na mineralização do N, na biomassa microbiana e na respiração do solo de soutos do Nordeste de Portugal, Revista de Ciências Agrárias, 27, 361– 375, 2004.
- Raimundo, F., Pires, A. L., Fonseca, S., Martins, A., and Madeira, M.: Chestnut and litterfall production and leaf nutrient concentration in chestnut groves submitted to different soil tillage systems, Revista de Ciências Agrárias, 32, 245–257, 2009.
- Trigo, R. M., Pereira, J. M. C., Pereira, M. G., Mota, B., Calado, T. J., DaCamara, C. C., and Santo, F. E.: Atmospheric conditions associated with the exceptional fire season of 2003 in Portugal, Int. J. Climatol., 26, 1741–1757, 2006.
- Vicente-Serrano, S. M., Cuadrat-Prats, J. M., and Romo, A.: Early prediction of crop production using drought indices at different time-scales and remote sensing data: application in the Ebro valley (North-East Spain), Int. J. Remote, 27, 511–518, 2006.
- Wilczynski, S. and Podlaski, R.: The effect of climate on radial growth of horse chestnut (*Aesculus hippocastanum* L.) in the Swietokrzyski National Park in central Poland, J. For. Res., 12, 24–33, 2007.
- Wilks, D. S.: Statistical methods in the atmospheric sciences: an introduction, Academic Press, San Diego, USA, 1995.
- Zhou, L., Tucker, C. J., Kaufmann, R. K., Slayback, D., Shabanov, N. V., and Myneni, R. B.: Variations in northern vegetation activity inferred from satellite data of vegetation index during 1981 to 1999, J. Geophys. Res., 106, 20069–20083, 2001.