Testing and modelling the effects of climate on the incidence of the emergent nut rot agent of chestnut Gnomoniopsis castanea

Testing and modelling the effects of climate on the incidence of the emergent nut rot agent of chestnut *Gnomoniopsis castanea*

G. Lione, L. Giordano, F. Sillo and P. Gonthier*

Department of Agricultural, Forest and Food Sciences, University of Turin, Largo Paolo Braccini 2, I-10095 Grugliasco, Italy

*E-mail: paolo.gonthier@unito.it

Running head: Modelling incidence of *G. castanea*

Keywords: *Gnomoniopsis castanea*, *Castanea sativa*, nut rot, incidence, climate, modelling

Abstract

Gnomoniopsis castanea is an emerging fungal pathogen causing nut rot of *Castanea sativa*. This study was aimed at testing and modelling the effects of climate on disease incidence. Up to 120 ripe nuts were collected in 2011 from trees in each of 12 sites located in the north-west of Italy. The incidence of *G. castanea* in each site was expressed as the number of infected nuts on the total number of nuts sampled (%), as determined by combining the results of morphological identification of isolates obtained from nuts and their typing through a newly developed taxon-specific molecular assay. Disease incidence ranged from 20% to 93%, depending on site. Geostatistical analyses revealed that, despite the clustering of sites ($P<0.05$), disease incidence was not spatially autocorrelated ($P>0.05$). This finding suggests that the disease is influenced by site-dependent factors whose scale (~7.5-15.6 km) is consistent with the climate variability throughout the sampling region. Multivariate analyses on maximum, mean and minimum temperatures and on rainfall showed that warmer temperatures were associated with higher levels of the disease incidence. The temperatures of months before nut harvesting were selected as predictors for Partial Least Squares Regression (PLSR) models (*Gno*Mods) of *G. castanea* incidence. External validation on data collected either on sites or in years not used for models fitting showed the good predictive abilities of the *Gno*Mods (Spearman $\rho_{obs/\,pred} > 0.72$, $P<0.05$). The above findings support a relation between climate and incidence of *G. castanea*, providing statistical tools to forecast disease incidence at site level.

Introduction

Sweet chestnut (*Castanea sativa* Miller) is a widespread broadleaf species in southern and western Europe, in Maghreb, Turkey, Caucasus as well as in Australia and New Zealand. This species has been spread and cultivated for thousands of years for both fruit and wood production and plays an important economic role in many countries, being a source of food highly appreciated for both fresh consumption and processing because of appreciable organoleptic and nutritional properties.

Several fungi can cause nut rot of chestnut in pre-harvest and/or post-harvest conditions resulting in yield and economic losses, including *Botrytis cinerea* Pers., *Ciboria batschiana* (Zopf) N.F. Buchw., *Cytodiplospora castanea* Oudem., *Diplodina castaneae* Prill. & Delacr., *Dothiorella* spp., *Fusarium* spp., *Penicillium* spp., *Pestalotia* spp., *Phoma castanea* Peck, *Phomopsis endogena* (Speg.) Cif., *Phomopsis viterbensis* Camici and *Rhizopus* spp. (Washington *et al*., 1997). Since 2005, in Italy, France and Switzerland, chestnut growers have noticed an abnormal increase in the amount of rotten nuts locally affecting more than 80% of nuts (Visentin *et al*., 2012; Maresi *et al*., 2013). The huge majority of these nut rots were associated with an emerging fungal pathogen recently described as *Gnomoniopsis castanea* Tamietti, an ascomycete belonging to the family of Gnomoniaceae (Visentin *et al*., 2012). The symptoms of the nut rot caused by this fungus include a chalky aspect of the nut kernel at ripening, turning to brown as soon as the mummification progresses and the mycelium invades the kernel tissues. Besides being a parasite in the kernel of the nuts, *G. castanea* can also be found as an endophyte in the thin bark of chestnut branches and in other green tissues of the tree (Visentin *et al*., 2012). The teleomorphic stage of the fungus produces its perithecia on the burrs (Visentin *et al*., 2012). The acervuli of the anamorphic stage can be observed on necrotic galls whose formation on chestnut buds and leaves is triggered by the Asian chestnut gall wasp (*Dryocosmus kuriphilus* Yasumatsu) accidentally introduced to Europe in the early 2000s (Quacchia *et al*., 2008). A disease very similar to the one here described was observed

in New Zealand starting from 2008 (Shuttleworth *et al*., 2013). While the pathogen was described in New Zealand as *Gnomoniopsis smithogilvyi* L.A. Shuttlew. (Shuttleworth *et al*., 2012), it is still unknown whether the two congeners *G. castanea* and *G. smithogilvyi* may be the same species or not.

To date, little is known about the ecology, epidemiology, biogeography and infection biology of *Gnomoniopsis* spp. on chestnut. Despite some hypotheses on the reasons determining the spread and the severity of these pathogens in chestnut orchards (Gentile *et al*., 2009; Maresi *et al*., 2013; Shuttleworth *et al*., 2013), many aspects still need to be elucidated.

The climate has been reported to be related to pathosystems dynamics at global, regional and local scale both in agriculture and in forestry (Garrett *et al.*, 2006). Climate may affect the pathosystems influencing not only the pathogens and their hosts, but also ecosystems composition, structure and functions (Garrett *et al.*, 2006). During the last decades researchers have shown a growing interest in elucidating the role played by climate on plant diseases under a quantitative perspective. Many regression and simulation models have been proposed to explain and/or predict disease parameters as a function of the climate. Despite no general rules can be used to forecast the impact of climate on plant diseases, a vast body of literature support that temperature and rainfall figure among the most important climatic variables to model and to predict incidence, severity and spread of plant pathogens (Coakley *et al*., 1999; Kendrick, 2000; Magarey *et al.*, 2005; Garrett *et al.*, 2006). Epidemiological models including temperature and/or rainfall as predictors have been proposed for a large variety of plant pathogens as, for instance, *Alternaria alternata* (Fr.) Keissl. (Moschini *et al.*, 2006.), *Fusarium oxysporum* f. sp. *ciceris* Matuo & K. Satô (Navas-Cortés *et al.*, 2007), *Heterobasidion* spp. (Gonthier *et al*., 2005), *Phytophthora ramorum* Werres, De Cock & Man in 't Veld (Kelly *et al.*, 2007) and *Plasmopara viticola* (Berk. & M.A. Curtis) Berl. & De Toni (Lalancette *et al.*, 1988). To date such models have found many practical applications in different fields including crop production estimation, food security policy, forest management, plant disease control, risk maps development, decision making support and economic losses estimation (Gregory

116 *et al.*, 2009; Edmonds, 2013; Gonthier & Thor, 2013). In particular warming temperatures, often
117 related to the global climate change, have been identified in many cases as risk factors increasing
118 the detrimental effects of plant pathogens (Harvell *et al.*, 2002; Doohan *et al.*, 2003).
119 Some observations carried out in Italy, Australia and New Zealand suggest that climate could
120 play a role in promoting high incidence levels of *G. castanea* and *G. smithogilvyi* (Maresi *et al.*,
121 2013; Shuttleworth *et al.*, 2013). Even if dry and warm periods (Maresi *et al.*, 2013), as well as
122 rainy and warm ones (Smith & Agri, 2008; Smith & Ogilvy, 2008; Gentile *et al.*, 2009;
123 Shuttleworth *et al.*, 2013) occurring during the vegetative season have been suggested to affect the
124 incidence of nut rots, many of these hypotheses still need to be confirmed by statistical evidence.
125 Several difficulties and constraints arise when modelling the incidence of plant diseases as a
126 function of environmental variables because of sampling adequacy, spatial autocorrelation, spatial
127 pseudoreplication, high collinearity among predictors, noise, lack of model parameters
128 distributional theory and presence of restrictive assumptions regarding the statistical tests (Roy &
129 Roy, 2008; Kéry, 2010; Crawley, 2013). However, recent improvements in statistics have led to the
130 availability of methods allowing plant pathologists to carry out computational analyses that can deal
131 with many of the above cited constraints. For instance, tools once unavailable or mainly confined to
132 the borders of specific fields (e.g. chemometrics, criminology, urban planning) have recently been
133 used in plant pathology (Gonthier *et al.*, 2012a,b; Garbelotto *et al.*, 2013). These methods and tools
134 include Geographic Information Systems (GIS), spatial clustering and spatial autocorrelation
135 analyses, Partial Least Squares Regression (PLSR), cross-validation, bootstrap and Principal
136 Coordinates Analysis (PCoA).
137 Taking advantage from the above cited methods and tools, the goals of this research were: I) to
138 verify if the spatial pattern of the incidence of *G. castanea* at regional level is consistent with the
139 hypothesis of a climate influence on the disease, II) to test whether climatic parameters and
140 incidence of the disease are correlated, III) to model the incidence of the disease at site level as a
141 function of climatic parameters, and IV) to validate the models.

Materials and methods

Study sites, samplings and fungal isolations

Up to 120 ripe nuts per site (Table 1) were randomly collected at the beginning of November 2011 from the crown of 6-8 trees per site randomly chosen in 12 sweet chestnut orchards located in the north-west of Italy. The sites were selected so as to include a wide latitudinal and longitudinal extension according to the chestnut distribution in the area. Sites were located within a rectangular region of 9080 km^2 (63 km from E to W, 144 km from S to N) at a mean distance of 12 km. The precise location and the main characteristics of the study sites are reported in Table 1. Samples were transported to the laboratory and stored at 4°C before subsequent analyses.

Under a biological hood, 5 fragments per nut (approximately 1 × 1× 2 mm in size) were excised and plated in 9 cm diameter Petri dishes filled with Malt Extract Agar (MEA) as previously described (Visentin *et al*., 2012). Putative colonies of *G. castanea* were identified by examining macro and micro-morphological features including both the aspect of mycelium and acervuli and the shape and size of conidia. The incidence of *G. castanea* at site level was calculated as the ratio (%) between the number of infected nuts and the total number of nuts sampled.

Development and application of a taxon-specific molecular diagnostic assay

To confirm the morphological identification, a subset of 36 randomly selected putative colonies of *G. castanea* and all colonies showing anomalous morphological characters were typed by using a taxon-specific molecular diagnostic assay. Taxon-specific primers for *G. castanea* were designed based on alignment of ITS (Internal Transcribed Spacer) sequences of 15 species belonging to Gnomoniaceae family. In order to check their specificity, primers were also tested in an optimized

PCR assay using as template the DNA extracted from three ascomycetes fungi frequently associated with chestnut. Details of primers design, DNA extractions, PCR reactions and gel electrophoresis visualization are reported as Supplementary Material (S1).

Geostatistical analyses

The coordinates of each site were recorded with a GPS device (Magellan Mobile Mapper 6) in UTM WGS84 (zone 32N). Geostatistical analyses were implemented in CrimeStat 3.3 (Levine, 2010) to detect the spatial pattern of sites and to test the spatial autocorrelation of incidence levels of *G. castanea*.

The spatial pattern of sites was investigated with the *L(d)* transformed Ripley's *K(d)* function (Mitchell, 2005) where *d* is the geographic distance among sites. Lower ($L_{csr\text{-}lower}$) and upper ($L_{csr\text{-}upper}$) bounds were calculated for L_{csr}, that is the *L(d)* function under the assumption of complete spatial randomness (csr) with 2000 simulations and 95% confidence level. The default correction for a rectangular study area was selected. The *L(d)* function, $L_{csr\text{-}lower}$ and $L_{csr\text{-}upper}$ were plotted against *d* in order to find distance ranges of significant spatial clustering ($L(d)>L_{csr\text{-}upper}$), of significant spatial dispersion ($L(d)<L_{csr\text{-}lower}$) and the remaining ranges of random spatial pattern (Mitchell, 2005). A Nearest Neighbor Hierarchical Clustering (NNHC) analysis was performed to identify significant spatial clusters of sites and their order (Mitchell, 2005). The consistency between the results of the *L(d)* function and the NNHC analysis was assessed by measuring the distance between clustering sites (see Results).

The spatial autocorrelation of incidence levels of *G. castanea* was assessed with the General Moran's Index (I) and with the Getis-Ord General G-statistic (G) (Mitchell, 2005). The latter was calculated in a range of distances from 1 to 100 km (with 100 iterations) to detect the presence of cold and hot spots. The threshold to reject the null hypothesis of tests was set at P=0.05.

194 **Climatic analyses**

195

196 For each site, climatic data were downloaded from the nearest thermo-pluviometric station
197 (ARPA Piemonte, 2011). Those data included daily maximum, mean and minimum temperatures
198 (°C) and the total daily rainfall (mm) from January 1st 2011 to October 31st 2011. To estimate the
199 consistency between the climatic data derived from the thermo-pluviometric stations and the
200 climate of the study sites, the mean distance between the sites and their nearest thermo-pluviometric
201 stations was calculated (Table 1). Moreover, to assess the consistency between the spatial
202 distribution of the sites and the spatial distribution of their nearest thermo-pluviometric stations the
203 correlation between the geographical distance matrices among sites and among their nearest
204 thermo-pluviometric stations was tested with the simple Mantel test.

205 The correlation between the incidence of the pathogen at site level and the monthly average
206 maximum, mean, minimum temperatures and the monthly average rainfall was assessed with the
207 Spearman's ρ correlation coefficient analysis (Crawley, 2013).

208 Each of the 1200 daily values for both temperatures and rainfall was used as variable to perform
209 a Principal Coordinates Analysis (PCoA) on sites. The PCoA was performed on the Euclidean
210 distance matrix calculated from the coordinates of the sites in the space defined by the above cited
211 variables. The minimum number of principal axes accounting for more than 70% of the total
212 variance was retained and the principal coordinates of the sites were calculated. On those principal
213 coordinates a Hierarchical Cluster Analysis (HCA) based on the Euclidean distance matrix and on
214 the Ward agglomerative method (Garbelotto *et al*., 2013) was run to define groups of sites
215 characterized by similar climatic conditions. The maximum silhouette width and the minimum C-
216 index criteria were used to identify the optimal number of clusters.

217 The climate conditions between the two clusters of sites detected by the HCA (see Results) were
218 compared with the Mann-Whitney test performed with exact significance (Crawley, 2013) on the
219 average maximum, mean, minimum temperatures and average rainfall of each month. Bootstrap

bias-corrected accelerated percentile confidence intervals were calculated for each monthly average value based on 10000 iterations (Crawley, 2013).

The incidence of *G. castanea* was calculated for the two clusters. The incidence levels were then compared with a χ^2 test.

The above mentioned analyses were carried out in R programming language (R Core Team, 2013) by running the *labdsv* library for PCoA, the *NbClust* and *clValid* libraries for HCA, the *boot* library for the calculation of the bootstrap confidence intervals and the *ecodist* library for the simple Mantel test.

Model fitting and validation

A PLSR (Wold *et al*., 2001) was performed to model the incidence of *G. castanea* at site level in relation to the climatic conditions. The incidence value of the pathogen in each site was transformed through the application of the logit function before PLSR models fitting (Crawley, 2013). The logit transformed incidence at site level was used as dependent variable. The monthly average maximum, mean, minimum temperatures and the monthly average rainfall recorded at each site from January to October 2011 were considered as potential predictors. A pre-selection of predictors was performed before models fitting: only the climatic variables being significantly different between the 2 clusters of sites identified with the PCoA-HCA analyses were retained as predictors.

A first set of PLSR models was fitted to sites data including all the pre-selected predictors (hereafter in the text simply defined as predictors) and from 2 to 11 latent variables (LV) (Abdi, 2010). In addition, the null model was also fitted. Every model was identified by the acronym *Gno*Mod (*Gnomoniopsis* Model) followed by two indexes indicating the number of LV and the number of predictors. Each *Gno*Mod was expressed in terms of a vector-matrix form equation $\hat{\mathbf{Y}} = \mathbf{XB}$ (Equation 1) where $\hat{\mathbf{Y}}$ is the column vector of the predicted values of the incidence of *G. castanea* at site level (i.e. logit of the percentage of nuts infected by *G. castanea*), **X** is design

246 matrix of the predictors for a model parameterization including the intercept (i.e. a matrix of the
247 predictors values whose first column is filled by 1s) and **B** is the column vector of the β
248 coefficients (i.e. the multiplicative coefficients obtained through PLSR fitting and assigned to each
249 predictor) (Wold *et al*., 2001; Kéry, 2010).

250 For every *Gno*Mod the Akaike Information Criterion (AIC) was calculated as described by Li *et*
251 *al*. (2002) by adding a constant set to 100. LV selection was performed according to the minimum
252 AIC criterion (Li *et al*., 2002). For the resulting *Gno*Mod (i.e. the one with lowest AIC), the ΔAIC
253 between the null model and the actual model was calculated. A semiparametric bootstrap based on
254 10000 iterations was performed on the ΔAIC, deriving its 95% confidence interval (Carpenter &
255 Bithell, 2000). On the same model, the internal validation parameters Q^2 (Wold *et al*., 2001) and
256 Q^2_{cum} (Lazraq *et al*., 2003) were determined by cross-validation. The Q^2 is similar to R^2 in classical
257 Ordinary Least Squares (OLS) regression, but originates from iterative calculus and refers to the
258 estimation of predictive ability rather than to goodness of fit. Instead Q^2_{cum} provides and estimate of
259 the internal adequacy of the predictors.

260 In order to test and validate the effective predictive ability, the *Gno*Mod was run on data of a
261 validation set (i.e. data not used to fit the model) (Abdi, 2010) gathered from 8 chestnut orchards,
262 some of which were sampled more than once but in different years during a period lasting from
263 2007 to 2013 (Table 2). Samplings in these orchards were carried out at the beginning of
264 November. The incidence of *G. castanea* (i.e. observed incidence) was assessed through
265 morphological identification of isolates as previously described, while the input predictors were
266 collected for the validation set and then inserted in the *Gno*Mod equation to estimate the incidence
267 of *G. castanea* (i.e. predicted incidence) in logit scale. The predicted and the observed values
268 recorded for the validation set were used to calculate some external validation indexes including
269 their squared correlation coefficient and associated P-value ($R^2_{obs/\,pred}$) (Roy & Roy, 2008), their
270 Spearman's correlation coefficient and related P-value ($\rho_{obs/\,pred}$) (Gonthier *et al*., 2012a), the

semiparametric bootstrap 95% confidence interval for the dependent variable based on 10000 iterations (95% CI_{dv}) (Carpenter & Bithell, 2000; Abdi, 2010) and the Mean Square Error of Prediction (MSEP) (Aptula *et al*., 2005). For the 95% CI_{dv}, its mean width value (95% CI_{dvw}) was calculated as a summary measure.

The PLS-bootstrap method was applied on the *Gno*Mod to perform predictors selection according to the algorithms of Amato & Esposito Vinzi (2003) and Lazraq *et al*. (2003) run in their semiparametric variant (Carpenter & Bithell, 2000). This procedure was iterated until no 95% confidence intervals of the predictors coefficients included 0. All the above described indexes were calculated for each nested *Gno*Mod obtained at every step of the PLS-bootstrap method. The collinearity of the predictors was assessed with the Steiger test.

To further assess the consistency among the climatic analyses and the *Gno*Mods equations a simulation was carried out. The simulation consisted in running the equations on the validation set after increasing the predictors values by a multiplicative constant set to 1.01, then to 1.02 and finally to 1.05 and in recording at each step the extent of variation of the mean predicted dependent variable. The effect was estimated by calculating the mean percentage of increase in the predicted dependent variable for a 1% increment of the predictors values.

The PLSR models were fitted and cross-validated with the *plsdepot* library, while the other algorithms described were compiled in R programming language (R Core Team, 2013).

Results

Incidence of *G. castanea* and taxon-specific molecular diagnostic assay

A total of 441 colonies were identified as *G. castanea* based on macro and micro-morphological features. *G. castanea* was present in nuts from all the study sites. The incidence of the disease ranged from 20.0% to 93.5%, depending on site (Table 1), while the total incidence was 64.6%.

297 A forward primer (Gc1f, 5'-AGCGGGCATGCCTGTTCGAG-3') and a reverse primer (Gc1r,
298 5'-ACGGCAAGAGCAACCGCCAG-3') were designed to amplify a 168 bp PCR product when
299 used with the following thermocycler parameters: an initial cycle with a 95°C denaturation step of 5
300 min, followed by 35 cycles, each one consisting of a 95°C denaturation step of 30 s, a 62°C
301 annealing step of 45 s and a 72°C extension step of 1 min and a final cycle with a 72°C extension
302 step of 10 min. No cross-reactivity of primers with DNA of ascomycotes frequently associated with
303 chestnut was observed (Fig. 1). The morphological identification of *G. castanea* was confirmed by
304 the results of the taxon-specific molecular diagnostic analysis for the whole subset of isolates.
305
306 **Geostatistical analyses**
307
308 The *L(d)* function analysis indicated that the sites were significantly clustered ($P<0.05$) within a
309 distance range comprised between 7.47 and 15.55 km ($L(d)>L_{csr\text{-}upper}$), while for all the other
310 distance ranges the spatial pattern of sites did not differ significantly from a random one
311 ($L_{csr-lower} \leq L(d) \leq L_{csr-upper}$) ($P>0.05$) (Fig. 2).
312 The NNHC analysis identified three significant first order spatial clusters of sites. The largest
313 cluster (A) included four sites, while the other two (B and C) were composed by only two sites each
314 (Table 1). The mean distance among sites within the clusters was 7.49 km, a value in agreement
315 with the clustering range indicated by the *L(d)* function.
316 The General Moran's Index (I) excluded the presence of spatial autocorrelation of the incidence
317 of *G. castanea* ($I=0.18$; $P>0.05$). This result was confirmed by the Getis-Ord General G-statistic,
318 that attained not significant values ranging from 0.00 to 0.82 ($P>0.05$), showing that neither hot
319 spots nor cold spots could be identified.
320
321 **Climatic analyses**
322

323 The simple Mantel test revealed a strong and significant correlation between the distance
324 matrices of sites and of their nearest thermo-pluviometric stations (R=0.99; P<0.05). The mean
325 distance between sites and their nearest thermo-pluviometric stations was 4.79 km.
326 The Spearman's ρ correlation coefficients analysis (Fig. 3) showed significant positive
327 correlations (P<0.05) between the incidence of *G. castanea* and the monthly average maximum
328 temperatures of July (ρ =0.60), August (ρ =0.62), September (ρ =0.61) and October (ρ =0.62) and
329 the monthly average mean temperatures of June (ρ =0.70) and July (ρ =0.63). Instead, no
330 significant correlations were detected between the incidence of *G. castanea* and the monthly
331 average rainfall, with the only exception of August which showed a negative correlation coefficient
332 (ρ =- 0.60; P<0.05).
333 In the PCoA only two principal axes were retained, the first one accounting for 56.2% and the
334 second for the 14.3% of the total variance. The HCA performed on the principal coordinates of sites
335 revealed that two clusters of sites sharing similar climatic conditions could be identified (Fig. 4). In
336 fact the maximum silhouette width (0.51) and the minimum C-index (0.33) were obtained when
337 sites were partitioned in two groups. The first cluster (cluster 1) included eight sites (1, 2, 5, 6, 7, 9,
338 11, 12; see Table 1 for sites codes) while the second one (cluster 2) comprised the remaining four
339 sites (3, 4, 8, 10).
340 Despite a slightly lower amount of precipitation in late spring, early and late summer in cluster 1
341 compared to cluster 2, differences between the two clusters in terms of monthly average rainfall
342 along the period from January to October were not significant (Mann-Whitney test; P>0.05).
343 Instead, many significant differences were detected for the monthly average maximum, mean and
344 minimum temperatures between the two clusters, indicating warmer climatic conditions in cluster 1.
345 The monthly average maximum temperatures were significantly higher in cluster 1 than in cluster 2
346 in every month from January to October (P<0.05), and the same was true for the monthly average
347 mean temperatures from February to October (P<0.05). Significant differences between the two

clusters were also observed in terms of monthly average minimum temperatures in the period ranging from April to July ($P<0.05$) (Fig. 5).

The incidence of *G. castanea* was 68.2% in cluster 1 and 57.8% in cluster 2. The χ^2 test revealed that the difference of incidence levels between the two clusters was significant ($P<0.05$).

Model fitting and validation

The pre-selection performed with the climatic analyses allowed for the identification of 23 predictors (i.e. the monthly average temperatures listed in the previous section that were significantly different between the two clusters). The null model attained an AIC value of 128.77, while among the models from *Gno*Mod-2-23 to *Gno*Mod-11-23 the minimum AIC (78.27) was observed in *Gno*Mod-8-23. Thus, from *Gno*Mod-8-23, the nested models *Gno*Mod-8-19, *Gno*Mod-8-16 and *Gno*Mod-8-15 were derived with the PLS-bootstrap method (Table 3). The four *Gno*Mods differed because of the number of included predictors (i.e. the monthly average maximum, mean, minimum temperatures listed for each model in Table 3). Only in *Gno*Mod-8-15 (the last step of the PLS-bootstrap method) the β coefficients were all significantly different from 0 ($P<0.05$). In all models the ΔAIC was significantly different from 0 and the Steiger test confirmed the collinearity among predictors ($P<0.05$). The four *Gno*Mods showed a constant Q^2 (0.99), while the other internal validation parameter Q^2_{cum} ranged from 0.53 to 0.88. In the 8 orchards included in the validation set the incidence of *G. castanea* was comprised between 5.0% and 83.3% depending on site and sampling year (Table 2). The external validation parameters $R^2_{obs/\,pred}$ (attaining values ranging from 0.52 to 0.65) and $\rho_{obs/\,pred}$ (comprised between 0.72 and 0.79) were significant ($P<0.05$) in each *Gno*Mod. The 95% CI_{dvw} varied from 2.95 to 3.21 and the MSEP ranged from 5.81 to 7.68 depending on the model.

For all *Gno*Mods the simulations recorded an increasing value of the predicted dependent variable at each step. On average a 1% increase of the predictors values produced a mean percentage of increase in the predicted dependent variable of 6.07% in *Gno*Mod-8-23, 5.10% in *Gno*Mod-8-19, 6.99% in *Gno*Mod-8-16 and 6.90% in *Gno*Mod-8-15.

Discussion

The nut rot caused by *G. castanea* represents a serious threat for sweet chestnut orchards, as shown in this study by the widespread occurrence and the high incidence of the pathogen in the north-west of Italy. In agreement with the results of previous surveys carried out in Italy (Visentin *et al*., 2012; Maresi *et al*., 2013), Australia (Shuttleworth *et al*., 2013) and New Zealand (Smith & Agri, 2008), *Gnomoniopsis* spp. may be considered as emerging pathogens whose detrimental effects on nut production impose a better understanding of their ecology, epidemiology, biogeography and infection biology. This research was mainly focused at elucidating and modelling the relation between climate and the incidence of *G. castanea* at site level.

The primers Gc1f and Gc1r designed and tested in this study were shown to be taxon-specific for the amplification of the DNA of *G. castanea*, resulting in the successful discrimination between *G. castanea* and other common agents of nut rot of chestnut, such as *Ciboria* sp. and *Phomopsis* sp. Since this molecular assay was designed *ad hoc* as a tool to validate previous morphological identifications of fungal colonies isolated from nut kernels, further research is needed to assess its diagnostic efficacy on DNA extracted directly from chestnut tissues.

Geostatistical analyses performed on the geographic coordinates of sites i.e. *L(d)* function and NNHC clearly showed that there was a significant clustered spatial pattern of sites at a scale of a few kilometres (~7.5-15.6). This pattern is usually unfavourable in the context of inferential statistics, since it can frustrate the attempt to draw correct conclusions from data because of spatial pseudoreplication (Crawley, 2013). However, it should be noted that the risk of spatial

pseudoreplication is substantial only if the Tobler's principle holds true at the scale the study is performed. This principle states that the values of a variable (e.g. disease incidence) sampled from neighbouring locations are expected to be more similar than the ones coming from locations set far apart (Mitchell, 2005). The results of the General Moran's Index and the Getis-Ord General G-statistic revealed that the incidence of *G. castanea* violated the Tobler's principle and hence that the sampling was not affected by spatial pseudoreplication. Furthermore, the discrepancy between the geographic pattern of sites and the spatial autocorrelation pattern of the incidence of *G. castanea* indicates the scale at which factors potentially related to the disease are operating. In fact, considering that sites geographically clustered do not show similar values of disease incidence, the above mentioned factors are likely to be site-specific, hence variable from site to site at the sampling scale of few kilometres as indicated by the *L(d)* function.

As reported by previous papers focused on *Gnomoniopsis* spp. associated with chestnut, the climate might stand among the most important factors related to the incidence of nut rot in chestnut orchards (Maresi *et al*., 2013; Shuttleworth *et al*., 2013). This study tested the consistency between the spatial pattern of the incidence of *G. castanea* and the hypothesis of a climate influence on the disease. Based on the results of HCA and NNHC, the lack of spatial autocorrelation of the incidence of *G. castanea* implies also that nearer sites were not more likely to share similar values of the disease incidence. Thus the spatial pattern of incidence of *G. castanea* is consistent with the hypothesis of climate as a site-specific factor influencing the disease. It is worth noting that the average spatial variability of climate in the north-west of Italy, that is often sizeable even at a local scale, is in agreement with these findings. Even though those data came from thermo-pluviometric stations not located within the sampled chestnut orchards, the spatial distribution of these stations was highly correlated with the spatial distribution of the study sites as demonstrated by the results of the simple Mantel test. The mean distance between the study sites and their nearest thermo-pluviometric stations was also consistent with the scale of the study. Both these observations

423 demonstrate that the selected thermo-pluviometric stations were representative enough to correctly
424 describe the sites climate conditions.
425 The agreement between the spatial scale of both climate and disease incidence may suggest they
426 are associated, however it does not allow interpretation of the role and the relative importance of
427 different climatic parameters on the disease. For this reason further climatic analyses were carried
428 out. Monthly average temperatures were always positively correlated with the incidence of nut rot
429 caused by *G. castanea* and such correlation was significant for at least the maximum temperatures
430 or the mean temperatures in the period lasting from June to October. This finding suggests that
431 warmer temperatures in the second half of the vegetative season are associated with increasing
432 percentages of rotten nuts. Further evidence confirming this interpretation derives from results of
433 the PCoA and HCA. Cluster 1 was clearly characterized by warmer temperatures than cluster 2,
434 with the most notable differences detectable in the monthly average maximum and mean
435 temperatures. The incidence of *G. castanea* was significantly higher in cluster 1 than in cluster 2,
436 despite the mild magnitude of the difference (+10.4%). This significant but not substantial increase
437 of disease incidence may suggest that other factors in addition to climatic ones are likely to be
438 involved in driving infections and/or disease expression. Although the mechanisms of infection and
439 the pathways of host colonization are mostly unknown for this pathogen, some hypotheses on the
440 role played by warm temperatures on the disease may be formulated. Temperature affects fungal
441 growth and may trigger metabolic and functional changes in fungi improving their trophic balance
442 and sporulating ability (Kendrick, 2000). Such traits are pivotal for phytopathogenic fungi since
443 they are involved in host colonization and disease transmission. Interestingly, *in vitro* growth of *G.*
444 *castanea* was reported to be optimal at 25°C (Visentin *et al*., 2012), and such a temperature in this
445 study was attained in the field only in sites of cluster 1, whose disease incidence was higher.
446 However the effects of the temperature on the host side could be involved too. In fact the hypothesis
447 that warmer temperatures could be associated with stress on chestnut and consequently with an
448 increase of incidence of *G. castanea* was recently formulated (Maresi *et al*., 2013).

In a previous study the severity of the nut rot was mainly interpreted as a potential consequence of drought (Maresi *et al*., 2013), suggesting that the decrease of the water input provided by the rainfall could have played an important role. Instead, in the opposite hemisphere, abundant rainfalls during the flowering period were shown to be mildly correlated to the incidence of *G. smithogilvyi* (Shuttleworth *et al*., 2013). A comparison between the ecology of *G. castanea* and *G. smithogilvyi* may be hazardous since they occur in different biogeographical and environmental contexts, yet, at a first glance, the role of rainfall in the epidemiology of these pathogens seems to be still an argument to debate. The results of the climatic analyses performed in our study suggested that the rainfall was not significantly associated with the incidence of the nut rot. In fact no significant correlations were detected between the monthly average rainfall and the incidence of *G. castanea*, with the exception of August where the correlation was significant, but negative. Moreover the above mentioned cluster 1 and cluster 2 were never significantly different when compared in terms of monthly average rainfall. These findings cannot exclude a possible role of drought, but it is worth noting that drought does not depend only on a reduced water input, but also on the water loss which is often increased by warmer temperatures. Furthermore, since no correlation between the rainfall during the flowering period of the chestnut (June-July in the study sites) and the incidence of nut rot was detected, other factors in addition to possible floral infections should be considered to elucidate the infection biology and the epidemiology of *G. castanea*. A better understanding could be achieved with investigations performed on the abundance of the airborne inoculum of this fungus during the year in relation to the phenology of chestnut, on the potential interactions between the pathogen and other organisms affecting chestnut (e.g. the Asian gall wasp) and on the ways the pathogen penetrates into the host tissues. All these factors are, at least in theory, potentially influenced by climatic conditions, yet investigations of these aspects were beyond the aim of this study.

Four PLSR models (i.e. *Gno*Mods) were proposed in order to model the incidence of *G. castanea* at site level as the logit percent amount of infected nuts in function of monthly average

maximum, mean, minimum temperatures. Because of the high number of predictors and their collinearity, a simple OLS regression would not have been recommended. It is worth noting that a significant correlation between all the predictors and the dependent variable is not a prerequisite for PLSR fitting (Wold *et al*., 2001). However, a first pre-selection of predictors may be useful to improve the reliability of the β coefficients. The further selection of the predictors was considered advantageous since it improved the predictive performances, provided that all the four PLSR models obtained were significantly different from the null model. On one hand, the cross-validation suggested that the *Gno*Mods were interchangeable for predictive purposes (since they showed the same Q^2), but *Gno*Mod-8-15 was characterized by a better internal adequacy of the selected predictors (i.e. highest Q^2_{cum} value). On the other hand, the external validation indexes, often considered more reliable for models selection than the internal ones (Aptula *et al*., 2005), did not provide univocal response. Considering that the ideal model should maximise $R^2_{obs/\,pred}$ and $\rho_{obs/\,pred}$ while minimizing 95% CI_{dvw} and MSEP (Aptula *et al*., 2005; Roy & Roy, 2008) there is not an outstanding *Gno*Mod. However, combining the internal and external validation indexes, *Gno*Mod-8-16 and *Gno*Mod-8-15 may be the most reliable ones, especially considering the difference in MSEP with the other two models. It should be noted that all *Gno*Mods showed significant and high external validation indexes $R^2_{obs/\,pred}$ and $\rho_{obs/\,pred}$. This suggests that no substantial overfitting occurred and that *Gno*Mods are robust tools for predicting the incidence of *G. castanea* at site level even with data gathered from different sites and/or years. This finding implies that the *Gno*Mods predictions are reliable both under a spatial and under a temporal perspective. It is worth noting that a successful external validation is pivotal for all predictive models, but it is even more important in the case of models fitted on data gathered from a single-sampling session to ensure that no biased coefficients have been obtained. Moreover the simulations carried out with all *Gno*Mods demonstrated the consistency between the association of warmer temperatures with increasing

disease incidence (as identified by the climatic analyses) and the effects of increasing temperatures on the models response.

Modelling the incidence of the nut rot caused by *G. castanea* as a function of the climate may be interesting under many perspectives. Since *G. castanea* is an emerging pathogen whose ecology is still partially unknown, the fact that significant and robust models endowed with satisfactory predictive performances can be obtained is *per se* a relevant result enlightening there is a quantitative relation between the climate and the incidence of *G. castanea* at site level. Moreover, the *Gno*Mods could be practical tools to predict the incidence before nut harvesting. Such an estimate of the amount of rotten nuts could allow nut growers to evaluate the related economic losses and thus the convenience of nut harvesting. A similar approach has already been proposed, for instance in the estimation of the direct financial losses related to the incidence of hearth rot caused by *Heterobasidion annosum s.l.* in Alpine conifer stands (Gonthier *et al.*, 2012a.). It should be noted that despite the computational complexity for fitting the *Gno*Mods to experimental data, their application to new datasets is fairly trivial since to obtain the prediction of the logit percent amount of nuts infected by *G. castanea* at site level only the matrix **X** needs to be compiled with the required monthly average maximum, mean and minimum temperatures, whose values are easy to download from widely available meteorological databases.

Beyond the practical applications, these models could also provide the researcher with equations able to quantify the disease incidence under different climate change scenarios, possibly helping in the interpretation of the epidemiology of *G. castanea*. Assuming that the global climate change implies for the future a long-term warming of the temperatures, according to our results we might expect on average an increase of the incidence of *G. castanea* in analogy with documented case studies involving other plant pathogens (Harvell *et al.*, 2002; Doohan *et al.*, 2003). Despite our results showed that temperatures are associated with the incidence of *G. castanea*, we cannot exclude that other climatic variables not investigated in our study could play a role. Relative humidity, wind and solar radiation have been reported to be related to fungal spores dispersion and

525 survival (Munk, 1981; Rotem *et al.*, 1985; Kendrick, 2000), yet those climatic variables are often
526 not available. In fact only a few thermo-pluviometric stations belonging to the official networks
527 managed by regional or national agencies are equipped with the devices needed to measure those
528 variables, and this is particularly true in the mountain areas where chestnut orchards are located.
529 In conclusion, this study showed that climate is a site-specific factor that, at a scale of a few
530 kilometres, can affect the incidence of nut rot caused by *G. castanea*. It was shown that warm
531 temperatures during the months before nut harvesting are associated with increasing amount of
532 rotten nuts and that the incidence of the disease can be modelled based on temperature values.
533

534 ## Acknowledgements

535

536 This study was supported by a grant of Regione Piemonte through the activity of the "Chestnut
537 Growing Center". The authors wish to thank Giovanni Bosio (Servizio Fitosanitario Regione
538 Piemonte), Giacomo Tamietti and Silvia Gentile (University of Torino) for technical assistance as
539 well as the anonymous Referees that, with their suggestions, contributed to improve the paper.
540

541 ## References

542

543 Abdi H, 2010. Partial least squares regression and projection on latent structures regression (PLS
544 Regression). *Wiley Interdisciplinary Reviews - Computational Statistics* **2**, 97–106.
545 Amato S, Esposito Vinzi V, 2003. Bootstrap-based $\hat{Q}^2_{kh}$ for the selection of components and
546 variables in PLS Regression. *Chemometrics and Intelligent Laboratory Systems* **68**, 5–16.
547 Aptula AO, Jeliazkova NG, Schultz TW, Cronin MTD, 2005. The better predictive model: high q^2
548 for the training set or low root mean square error of prediction for the test set? *QSAR &*
549 *Combinatorial Science* **24**, 385–396.

ARPA Piemonte, 2011. Banca Dati Meteorologica. [http://www.arpa.piemonte.it]. Accessed 15 May 2014.

Carpenter J, Bithell J, 2000. Bootstrap confidence intervals, when, which, what? *Statistics in Medicine* **19**, 1141–1164.

Coakley SM, Scherm H, Chakraborty S, 1999. Climate change and plant disease management. *Annual Review of Phytopathology* **37**, 399–426.

Crawley MJ, 2013. The R book – 2nd edition. Chichester, UK: John Wiley and Sons Ltd.

Doohan FM, Brennan JB, Cooke M, 2003. Influence of climatic factors on *Fusarium* species pathogenic to cereals. *European Journal of Plant Pathology* **109**, 755–768.

Edmonds RL, 2013. General strategies of forest disease management. In: Gonthier P, Nicolotti G, eds. *Infectious Forest Diseases*, Wallingford, UK: CAB International, 29–49.

Garbelotto M, Guglielmo F, Mascheretti S, Croucher P, Gonthier P, 2013. Population genetic analyses provide insights on the introduction pathway and spread patterns of the North American forest pathogen *Heterobasidion irregulare* in Italy. *Molecular Ecology* **22**, 4855–4869.

Garrett KA, Dendy SP, Frank EE, Rouse MN, Travers SE, 2006. Climate change effects on plant disease: genomes to ecosystems. *Annual Review of Phytopathology* **44**, 489–509.

Gentile S, Valentino D, Visentin I, Tamietti G, 2009. *Discula pascoe* infections of sweet chestnut fruits in north-west Italy. *Australian Nutgrower* **23**, 23–25.

Gonthier P, Brun F, Lione G, Nicolotti G, 2012a. Modelling the incidence of *Heterobasidion annosum* butt rots and related economic losses in alpine mixed naturally regenerated forests of northern Italy. *Forest Pathology* **42**, 57–68.

Gonthier P, Garbelotto MM, Nicolotti G, 2005. Seasonal patterns of spore deposition of *Heterobasidion* species in four forests of the western Alps. *Phytopathology* **95**, 759–767.

Gonthier P, Lione G, Giordano L, Garbelotto M, 2012b. The American forest pathogen *Heterobasidion irregulare* colonizes unexpected habitats after its introduction in Italy. *Ecological Applications* **22**, 2135–2143.

576 Gonthier P, Thor M, 2013. Annosus root and butt rots. In: Gonthier P, Nicolotti G, eds. *Infectious*
577 *Forest Diseases*, Wallingford, UK: CAB International, 128–158.
578 Gregory PJ, Johnson SN, Newton AC, Ingram JSI, 2009. Integrating pests and pathogens into the
579 climate change/food security debate. *Journal of Experimental Botany* **60**, 2827–2838.
580 Harvell CD, Mitchell CE, Ward JR, Altizer S, Dobson AP, Ostfeld RS, Samuel MD, 2002. Climate
581 warming and disease risks for terrestrial and marine biota. *Science* **296**, 2158–2162.
582 Kelly M, Guo Q, Liu D, Shaari D, 2007. Modeling the risk for a new invasive forest disease in the
583 United States: An evaluation of five environmental niche models. *Computers, Environment and*
584 *Urban Systems* **31**, 689–710.
585 Kendrick B, 2000. The Fifth Kingdom – Third edition. Newburyport, MA: Focus Publishing.
586 Kéry M, 2010. Introduction to WinBUGS for Ecologists – A Bayesian Approach to Regression,
587 ANOVA, Mixed Models and Related Analysis. London, UK: Academic Press – Elsevier.
588 Lalancette N, Ellis MA, Madden LV, 1988. Development of an infection efficiency model
589 for *Plasmopara viticola* on American grape based on temperature and duration of leaf wetness.
590 *Phytopathology* **78**, 794–800.
591 Lazraq A, Cléroux R, Gauchi JP, 2003. Selecting both latent and explanatory variables in the PLS1
592 regression model. *Chemometrics and Intelligent Laboratory Systems* **66**, 117–126.
593 Levine N, 2010. CrimeStat: A spatial statistics program for the analysis of crime incident locations
594 (v 3.3), Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington,
595 DC. [http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/2824]. Accessed 11 March 2014.
596 Li B, Morris J, Martin EB, 2002. Model selection for partial least squares regression. *Chemometrics*
597 *and Intelligent Laboratory Systems* **64**, 79–89.
598 Magarey RD, Sutton TB, Thayer CL, 2005. A simple generic infection model for foliar fungal plant
599 pathogens. *Phytopathology* **95**, 92–100.
600 Maresi G, Oliveira Longa CM, Turchetti T, 2013. Brown rot on nuts of *Castanea sativa* Mill: an
601 emerging disease and its causal agent. *iForest* **6**, 294–301.

Mitchell A, 2005. The ESRI® guide to GIS analysis – volume 2 – Spatial measurements and statistics. Redlands, CA, USA: ESRI Press.

Moschini RC, Sisterna MN, Carmona MA, 2006. Modelling of wheat black point incidence based on meteorological variables in the southern Argentinean Pampas region. *Crop and Pasture Science* **57**, 1151–1156.

Munk L, 1981. Dispersal of erysiphe graminis conidia fromwinter barley. *Grana* **20**, 215–217.

Navas-Cortés JA, Landa BB, Méndez-Rodríguez MA, Jiménez-Díaz RM, 2007. Quantitative modeling of the effects of temperature and inoculum density of *Fusarium oxysporum* f. sp. *ciceris* races 0 and 5 on development of Fusarium wilt in chickpea cultivars. *Phytopathology* **97**, 564–573.

Quacchia A, Moriya S, Bosio G, Scapin I, Alma A, 2008. Rearing, release and settlement prospect in Italy of *Torymus sinensis*, the biological control agent of the chestnut gall wasp *Dryocosmus kuriphilus*. *BioControl* **53**, 829–839.

R Core Team, 2013. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. [http://www.r-project.org/]. Accessed 12 June 2013.

Rotem J, Wooding B, Aylor DE, 1985. The role of solar radiation, especially ultraviolet, in the mortality of fungal spores. *Phytopathology* **75**, 510–514.

Roy PP, Roy K, 2008. On some aspects of variable selection for partial least squares regression models. *QSAR & Combinatorial Science* **27**, 302–313.

Shuttleworth LA, Liew ECY, Guest DI, 2012. *Gnomoniopsis smithogilvyi* Fungal Planet 108. In: Crous PW, Summerell BA, Shivas RG, Burgess TI, Decock CA, Dreyer LL, Granke LL, Guest DI, Hardy GEStJ, Hausbeck MK, Hüberli D, Jung T, Koukol O, Lennox CL, Liew ECY, Lombard L, McTaggart AR, Pryke JS, Roets F, Saude C, Shuttleworth LA, Stukely MJC, Vánky K, Webster BJ, Windstam ST, Groenewald JZ, eds. *Fungal Planet Description Sheets: 107-127*. *Persoonia* **28,** 107–127.

Shuttleworth LA, Liew ECY, Guest DI, 2013. Survey of the incidence of chestnut rot in south-eastern Australia. *Australasian Plant Pathology* **46**, 63–72.

628 Smith HC, Agri M, 2008. The life cycle, pathology and taxonomy of two different nut rot fungi in
629 chestnut. *Australian Nutgrower* **22**, 11–15.
630 Smith HC, Ogilvy D, 2008. Nut rot in chestnuts – Scientific research identifies major cause of rot in
631 Australia/New Zealand chestnuts as *Gnomonia pascoe* not *Phomopsis*. *Australian Nutgrower* **22**,
632 10.
633 Visentin I, Gentile S, Valentino D, Gonthier P, Tamietti G, Cardinale F, 2012. *Gnomoniopsis*
634 *castanea* sp. nov (Gnomoniaceae, Diaporthales) as the causal agent of nut rot in sweet chestnut.
635 *Journal of Plant Pathology* **94**, 411–419.
636 Washington WS, Allen AD, Dooley LB, 1997. Preliminary studies on *Phomopsis castanea* and
637 other organisms associated with healthy and rotted chestnut fruit in storage. *Australasian Plant*
638 *Pathology* **26**, 37–43.
639 Wold S, Sjöström M, Eriksson L, 2001. PLS-regression: a basic tool of chemometrics.
640 *Chemometrics and Intelligent Laboratory Systems* **58**, 109–130.
641
642

Figures legends

Figure 1. Cross reactivity test for taxon-specific primers Gc1f/Gc1r. *Gnomoniopsis castanea* MUT 455 (lanes 1 and 5), *Cryphonectria parasitica* (lanes 2 and 6), *Ciboria* sp. (lanes 3 and 7) and *Phomopsis* sp. (lanes 4 and 8) were amplified with primers combination ITS1f and ITS4 and with primers combination Gc1f and Gc1r. No bands were observed with primers combination Gc1f and Gc1r for *C. parasitica*, *Ciboria* sp. and *Phomopsis* sp. Negative Controls (NC) were also included. M is the molecular weight marker 100-bp DNA Ladder.

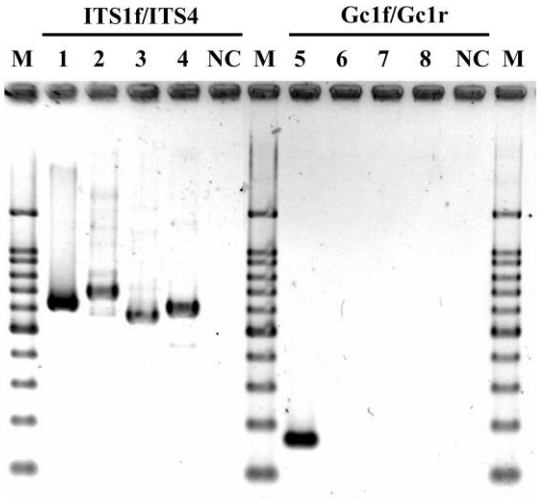
Figure 2. Spatial pattern of sites investigated with the *L(d)* transformed Ripley's *K(d)* function. The *L(d)* function is plotted against the geographic distance (*d*) among sites as well as the upper and lower bounds ($L_{csr\text{-}upper}$ and $L_{csr\text{-}lower}$) of the 95% confidence interval simulated under the assumption of complete spatial randomness. A significant spatial clustering of sites occurs in the interval between 7.47 and 15.55 km, where $L(d)>L_{csr\text{-}upper.}$

Figure 3. Spearman's ρ correlation analysis between the incidence of *G. castanea*, temperatures and rainfall. The Spearman's ρ correlation coefficient is indicated for each climatic parameter (monthly average maximum, mean, minimum temperatures and rainfall) from January to October. Asterisks show significant ρ values ($P<0.05$).

Figure 4. Multivariate analyses of sites with similar climatic conditions. a) Each site (see codes in Table 1) is projected as a point in a bi-dimensional space defined by the Principal Coordinates Analysis (PCoA). Nearer points share more similar climatic conditions than farther ones. b) The Hierarchical Cluster Analysis (HCA) performed on the principal coordinates of the sites shows that two clusters of sites sharing similar climatic conditions can be identified (cluster 1 and cluster 2). Sites belonging to the same cluster are also circled in the principal coordinates space (a).

Figure 5. Comparisons of temperatures and rainfalls between the clusters identified with PCoA and HCA. The monthly average maximum, mean and minimum temperatures and the monthly average rainfall (a, b, c and d, respectively) were compared between cluster 1 and cluster 2. The 95% bootstrap confidence intervals are reported for each value. For each month, different letters next to the plotted points indicate a significant difference detected by the Mann-Whitney exact test ($P<0.05$).

677 Fig. 1



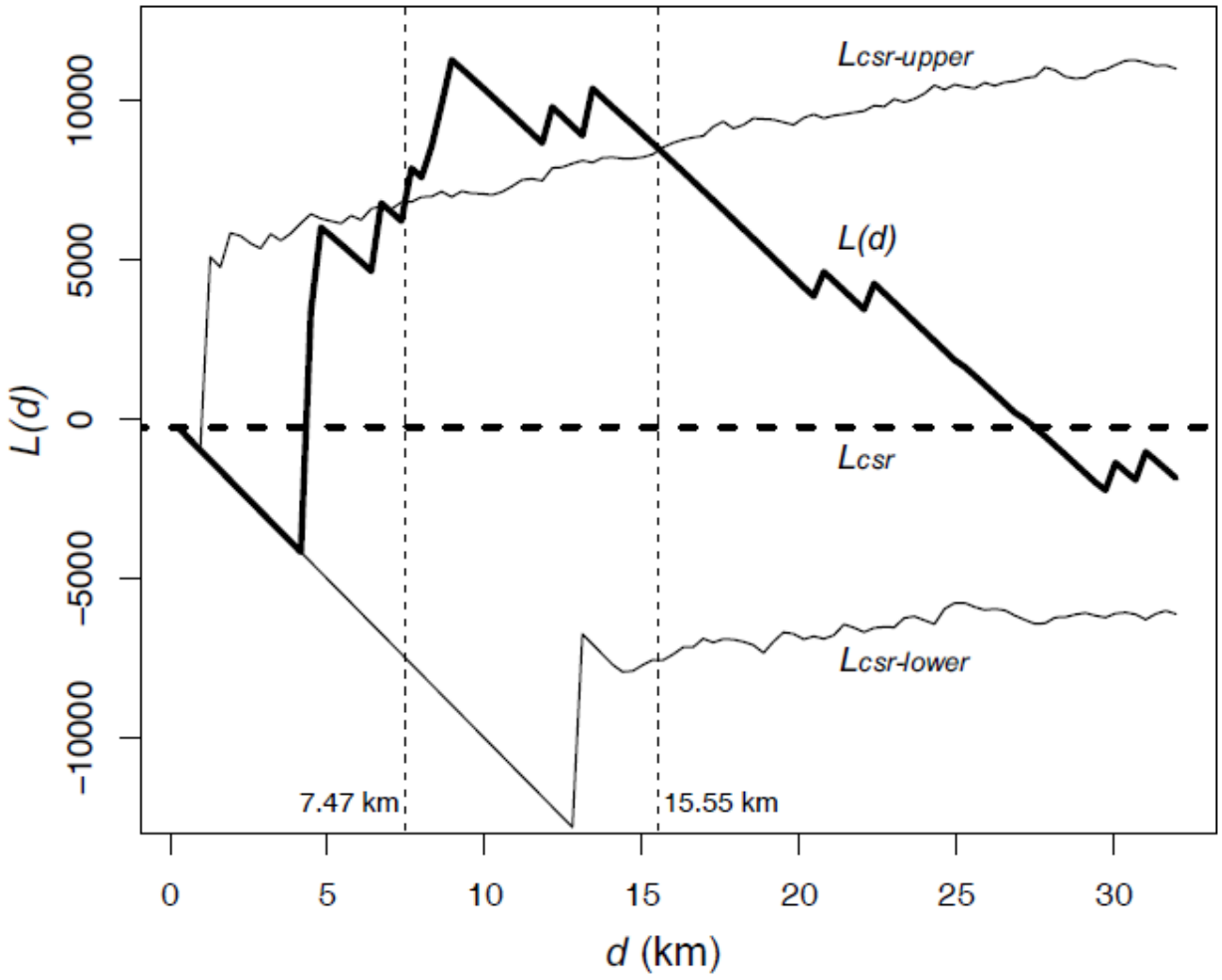
678

679

680

681 Fig. 2

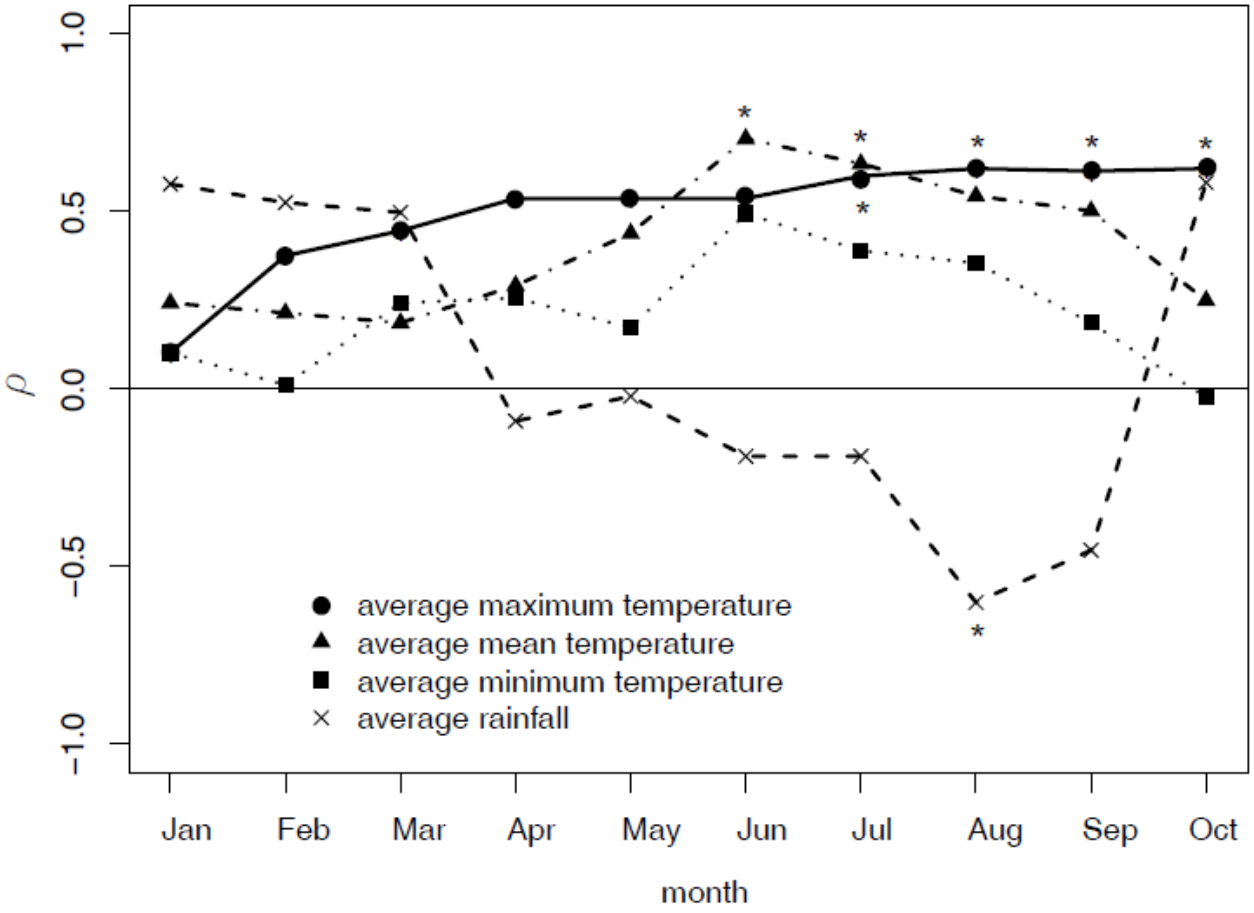
682



683

684

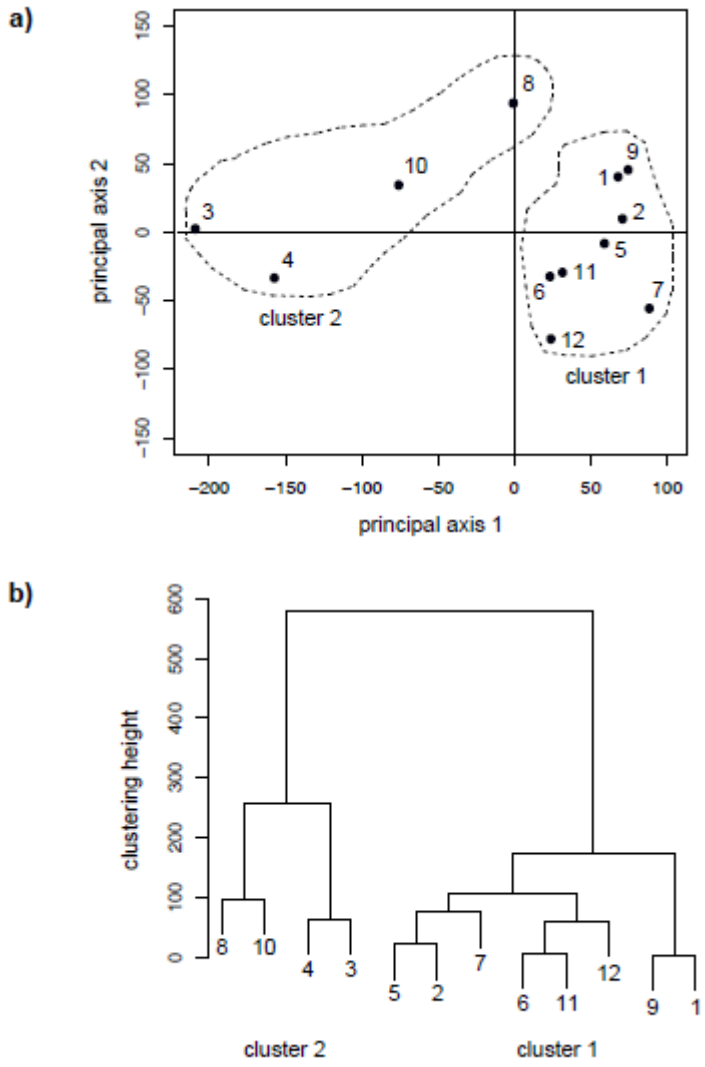
685 Fig. 3



686

687

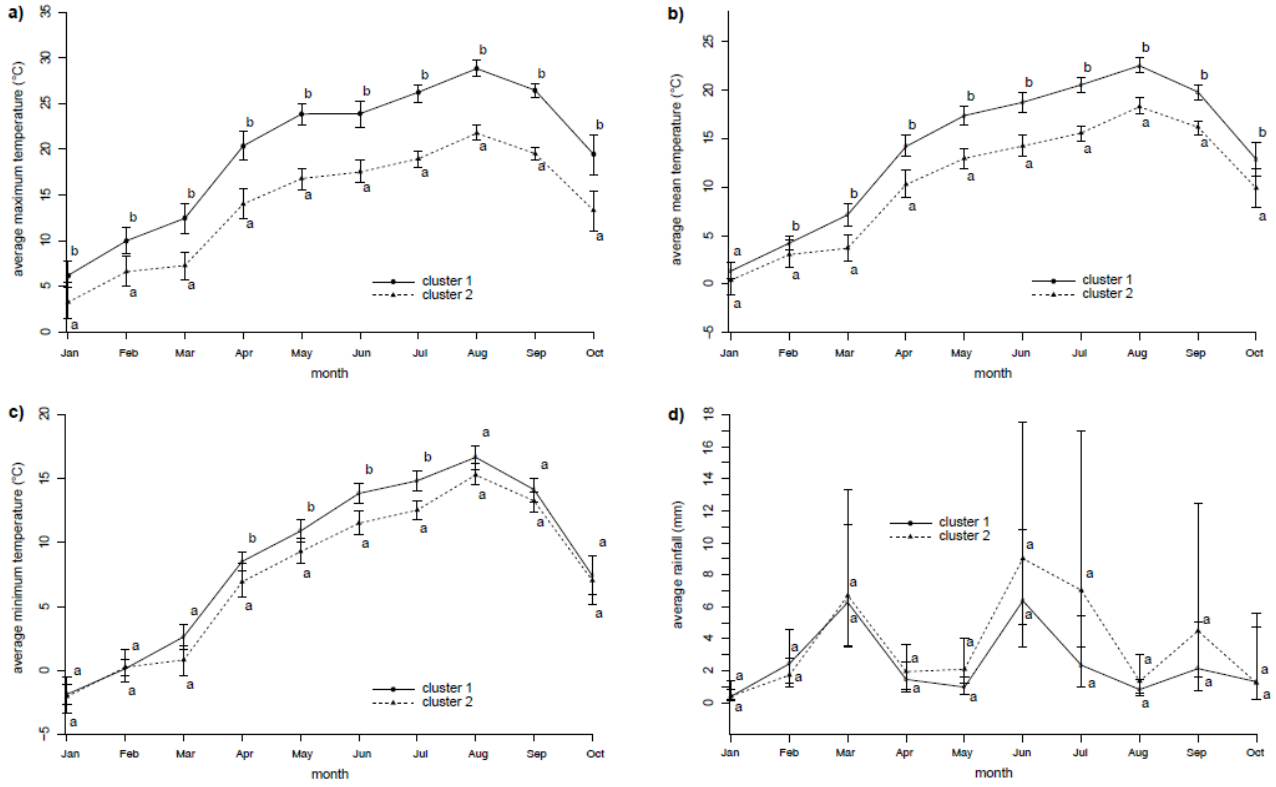
688 Fig. 4



689

690

691 Fig. 5



692

693

Table 1. Main characteristics of study sites sampled in 2011 for the assessment of the incidence of *Gnomoniopsis castanea*. For each site, the incidence of *G. castanea* and the results of the Nearest Neighbor Hierarchical Clustering (NNHC) are reported. Sites included in the same geographical cluster are marked with the same capital letter, while sites not included in any cluster are labelled with –.

Site name	**Site code**	**UTM WGS84 coordinates (m)**	**Altitude (m a.s.l.)**	**Exposure**	**Soil type (Soil Taxonomy)**	**Number of sampled nuts**	***G. castanea* incidence (%)**	**NNHC cluster**	**Distance from the nearest thermo-pluviometric station (km)**
Borgo San Dalmazzo	1	E 378203.3 N 4909837.6	655	ENE	Typic Hapludalf	40	85.0	A	6.17
Boves	2	E 385186.1 N 4907245.0	783	E	Typic Hapludalf	120	69.2	A	2.98
Donato	3	E 414851.2 N 5043995.9	1011	SSW	Typic Dystrudept	120	55.0	–	2.32
Donnas	4	E 402474.5 N 5048801.2	848	SE	n.a.*	37	59.5	–	8.86
Envie	5	E 371375.2 N 4950168.6	285	flat	Typic Hapludalf	80	77.5	B	10.51
Mattie	6	E 351141.2 N 4995572.5	1170	ENE	Typic Dystrudept	40	20.0	C	6.50
Monteu Roero	7	E 414064.5 N 4960599.5	350	NE	Psammentic Haplustalf	46	93.5	–	4.53
Peveragno	8	E 389871.2 N 4907514.9	680	NNW	Typic Hapludalf	40	80.0	A	2.18
Robilante	9	E 381773.9 N 4904511.4	695	NNE	Typic Hapludalf	40	75.0	A	2.26
Sanfront	10	E 365472.8 N 4944613.2	607	SW	Typic Dystrudept	40	42.5	B	3.87
Torre Pellice	11	E 357449.6 N 4965227.1	725	SSW	Typic Hapludalf	40	65.0	–	3.85
Villar Focchiardo	12	E 359474.5 N 4995073.5	1150	WNW	Typic Dystrudept	40	45.0	C	3.48

* not available

Table 2. Main characteristics of validation set sites sampled from 2007 to 2013 for the assessment of the incidence of *G. castanea* and for the external validation of the *Gno*Mods.

Site name	**UTM WGS84 coordinates (m)**	**Altitude (m a.s.l)**	**Exposure**	**Soil type (Soil Taxonomy)**	**Number of sampled nuts**	**Sampling year**	***G. castanea* incidence (%)**
Bastianetti (Italy)	E 420752.2 N 4896438.9	608	SSE	Typic Hapludalf	40	2012	35.0
					51	2013	31.5
Boves (Italy)	E 385186.1 N 4907245.0	783	E	Typic Hapludalf	40	2007	80.0
					40	2012	27.5
Gaiola (Italy)	E 371742.5 N 4910445.3	815	ESE	Typic Hapludalf	40	2012	32.5
					40	2013	5.0
Peveragno (Italy)	E 389871.2 N 4907514.9	680	NNW	Typic Hapludalf	102	2007	69.6
					40	2013	42.5
Robilante (Italy)	E 381773.9 N 4904511.4	695	NNE	Typic Hapludalf	37	2008	59.5
					40	2012	32.5
					40	2013	5.0
San Giorio di Susa (Italy)	E 357285.4 N 4997786.6	544	NNE	Typic Dystrudept	40	2013	10.0
Saint Auban (France)	E 315409.7 N 4855943.8	1240	N	n.a.*	40	2011	52.5
Valdieri (Italy)	E 371447.2 N 4904194.7	886	E	Typic Dystrudept	60	2007	83.3
					44	2008	18.2

* not available

Table 3. Coefficients and indexes of the Partial Least Squares Regression (PLSR) models *Gno*Mods. The β coefficients are associated with the predictors indicated in subscripts where tmax, tmed and tmin stand for monthly average maximum, mean and minimum temperatures followed by the abbreviation of the month they refer to. Next to the β coefficients their 95% confidence intervals are shown. The ΔAIC with its 95% confidence interval, the internal validation indexes Q^2, Q^2_{cum} as well as the external ones $R^2_{obs/\,pred}$, $\rho_{obs/\,pred}$, 95% CI_{dvw} and MSEP are reported for all models. After a coefficient or a parameter the symbol * indicates significance ($P<0.05$), no symbol indicates no significance ($P\geq0.05$), while (~) indicates that no test is associated with the value. The symbol – replacing coefficients values indicates that their associated predictors were removed from the model based on the outcomes of the PLS-bootstrap analysis.

	***Gno*Mod-8-23**	***Gno*Mod-8-19**	***Gno*Mod-8-16**	***Gno*Mod-8-15**
β_0	-7.97* (-8.98; -6.96)	-6.92* (-8.53; -5.32)	-6.62* (-8.40; -4.84)	-6.92* (-8.71; -5.13)
$\beta_{\text{tmax-jan}}$	0.01 (-0.11; 0.13)	–	–	–
$\beta_{\text{tmax-feb}}$	0.10 (-0.08; 0.28)	–	–	–
$\beta_{\text{tmax-mar}}$	-0.11* (-0.17; -0.04)	-0.05 (-0.13; 0.03)	–	–
$\beta_{\text{tmax-apr}}$	-0.15* (-0.24; -0.06)	-0.08 (-0.23; 0.08)	–	–
$\beta_{\text{tmax-may}}$	0.03 (-0.02; 0.09)	–	–	–
$\beta_{\text{tmax-jun}}$	0.52* (0.46; 0.58)	0.52* (0.47; 0.58)	0.50* (0.43; 0.57)	0.50* (0.43; 0.57)
$\beta_{\text{tmax-jul}}$	0.40* (0.37; 0.43)	0.45* (0.39; 0.50)	0.46* (0.37; 0.55)	0.46* (0.36; 0.56)
$\beta_{\text{tmax-aug}}$	0.05 (-0.01; 0.11)	–	–	–
$\beta_{\text{tmax-sep}}$	0.04* (0.01; 0.08)	0.05* (0.01; 0.09)	0.04 (-0.03; 0.12)	–
$\beta_{\text{tmax-oct}}$	0.37* (0.27; 0.48)	0.39* (0.25; 0.54)	0.33* (0.17; 0.49)	0.36* (0.19; 0.52)
$\beta_{\text{tmed-feb}}$	0.32* (0.06; 0.57)	0.42* (0.14; 0.70)	0.58* (0.33; 0.84)	0.52* (0.27; 0.78)
$\beta_{\text{tmed-mar}}$	-1.20* (-1.28; -1.13)	-1.18* (-1.3; -1.07)	-1.21* (-1.35; -1.07)	-1.20* (-1.34; -1.06)
$\beta_{\text{tmed-apr}}$	-0.42* (-0.52; -0.33)	-0.44* (-0.54; -0.34)	-0.46* (-0.59; -0.33)	-0.46* (-0.59; -0.33)
$\beta_{\text{tmed-may}}$	-0.15* (-0.19; -0.11)	-0.14* (-0.20; -0.07)	-0.14* (-0.22; -0.05)	-0.14* (-0.23; -0.05)
$\beta_{\text{tmed-jun}}$	0.34* (0.23; 0.45)	0.30* (0.17; 0.43)	0.21* (0.03; 0.38)	0.20* (0.01; 0.40)
$\beta_{\text{tmed-jul}}$	-0.32* (-0.38; -0.25)	-0.29* (-0.37; -0.21)	-0.30* (-0.41; -0.19)	-0.31* (-0.43; -0.20)
$\beta_{\text{tmed-aug}}$	-0.11* (-0.19; -0.02)	-0.12 (-0.26; 0.03)	–	–
$\beta_{\text{tmed-sep}}$	-0.39* (-0.47; -0.31)	-0.42* (-0.54; -0.30)	-0.39* (-0.64; -0.15)	-0.35* (-0.65; -0.05)
$\beta_{\text{tmed-oct}}$	-1.12* (-1.21; -1.02)	-1.16* (-1.27; -1.06)	-1.23* (-1.41; -1.06)	-1.22* (-1.39; -1.04)
$\beta_{\text{tmin-apr}}$	0.54* (0.41; 0.66)	0.47* (0.30; 0.65)	0.49* (0.32; 0.66)	0.46* (0.28; 0.64)
$\beta_{\text{tmin-may}}$	0.55* (0.39; 0.71)	0.61* (0.42; 0.79)	0.57* (0.38; 0.76)	0.58* (0.39; 0.77)
$\beta_{\text{tmin-jun}}$	1.93* (1.72; 2.15)	1.88* (1.59; 2.16)	1.74* (1.45; 2.03)	1.79* (1.48; 2.10)
$\beta_{\text{tmin-jul}}$	-1.24* (-1.36; -1.11)	-1.22* (-1.36; -1.08)	-1.19* (-1.35; -1.04)	-1.19* (-1.35; -1.04)
ΔAIC	50.51* (24.26; 77.06)	50.66* (28.93; 72.39)	48.34* (31.75; 64.93)	48.82* (28.80; 68.84)
Q^2	0.99(~)	0.99(~)	0.99(~)	0.99(~)
Q^2_{cum}	0.53(~)	0.78(~)	0.79(~)	0.88(~)
$R^2_{obs/\,pred}$	0.52*	0.59*	0.65*	0.63*
$\rho_{obs/\,pred}$	0.78*	0.79*	0.72*	0.79*
95% CI_{dvw}	3.21(~)	2.98(~)	2.95(~)	3.05(~)
MSEP	7.68(~)	8.08(~)	5.81(~)	6.00(~)