Mapping the suitability of groundwater dependent vegetation in a semi-arid Mediterranean area

Inês Gomes Marques¹; João Nascimento²; Rita M. Cardoso¹; Filipe Miguéns²; Maria Teresa Condesso de Melo²; Pedro M. M. Soares¹; Cathy Kurz Besson¹

¹Instituto Dom Luiz; Faculty of Sciences, University of Lisbon, Campo Grande, Ed. C8, 1749-016, Lisbon, Portugal
²CERIS; Instituto Superior Técnico, University of Lisbon, 1049-001, Lisbon, Portugal

Correspondence to: Inês Gomes Marques (icgmarques@fc.ul.pt)

Abstract. The forecasted groundwater resource depletion under future climatic conditions will greatly influence groundwater dependent ecosystems and their associated vegetation. In the Mediterranean region this will create harsh conditions for the maintenance of agroforestry systems dependent on groundwater, such as cork oak woodlands. The threat of increasing aridity conditions will affect their productivity and eventually induce a shift in their geographical distribution. Thus, characterizing and modelling the relationship between environmental conditions and groundwater dependent vegetation (GDV) will allow to identify the main drivers controlling its distribution and predict future impacts of climate change.

In this study, we built a model that explains the distribution of deep-rooted woody species in southern Portugal from climatic, hydrological and topographic environmental variables. To achieve this, we relied on the density of *Quercus suber*, *Quercus ilex* and *Pinus pinea* as proxy species of GDV. Model fitting was performed between the proxy species Kernel density and the selected environmental predictors using 1) a simple linear model and 2) a Geographically Weighted Regression (GWR), to account for autocorrelation of the spatial data and residuals. When comparing the results of both models, the GWR modelling results showed improved goodness of fitting, as opposed to the simple linear model. Soil type was the main driver of GDV density closely followed by the aridity index. Groundwater depth did not appear to be as pertinent in the model as initially expected.

Model predictor coefficients were used as weighting factors for multicriteria analysis, to create a suitability map to the GDV in southern Portugal. A validation of the resulting map was performed using independent data of integrated potential distribution of each proxy tree species in the region and overall, there was a good accordance between areas of good suitability to GDV. The model was considered reliable to predict the distribution of the studied vegetation, however, lack of data quality and information was shown to be the main cause for suitability discrepancies between maps.

Our new methodology on mapping of GDV’s will allow to predict the evolution of the distribution of GDV according to climate change scenarios and aid stakeholder decision-making concerning priority areas of water resources management.
Keywords: Groundwater dependent ecosystems, aridity, agroforestry, suitability map
1 Introduction

Groundwater is the largest subsurface water reservoir and supports valuable ecosystems (Eamus et al., 2006). Mediterranean forests, woodlands and shrublands, mostly growing under restricted water availability, are one of the terrestrial biomes with higher volume of groundwater used by vegetation (Evaristo and McDonnell, 2017). Future predictions of decrease precipitation (Giorgi and Lionello, 2008; Nadezhdina et al., 2015), decreased runoff (Mourato et al., 2015) and aquifer recharge (Ertürk et al., 2014) in the Mediterranean region threaten the sustainability of groundwater reservoirs and the corresponding dependent ecosystems. Therefore, a sustainable management of groundwater resources and the Groundwater Dependent Ecosystems (GDE) is of crucial importance.

Mapping GDE constitutes a first and fundamental step to their active management. Several approaches have been proposed, including remote sensing techniques (e.g. Normalized Difference Vegetation Index – NDVI) (Barron et al., 2014; Eamus et al., 2015; Howard and Merrifield, 2010), remote-sensing combined with ground-based observations (Lv et al., 2013), based on geographic information system (GIS) (Pérez Hoyos et al., 2016a) or statistical approaches (Pérez Hoyos et al., 2016b). Integrated multidisciplinary methodology (Condeeso de Melo et al., 2015) has also been used. A widely used classification of GDE was proposed by Eamus et al. (2006). This classification distinguishes between three types: 1) Aquifer and cave ecosystems, which includes all subterranean waters; 2) Ecosystems reliant on surface groundwater (e.g. estuarine systems, wetlands; riverine systems) and 3) Ecosystems reliant on subsurface groundwater (e.g. systems where plants remain physiologically active during extended drought periods, without visible water source).

Despite of a wide-ranging body of literature regarding GDE, most of the studies did not included Mediterranean regions (Doody et al., 2017; Dresel et al., 2010; Münch and Conrad, 2007). Moreover, studies on ecosystems relying on subsurface groundwater frequently only focused on riparian environments (Lowry and Loheide, 2010; O’Grady et al., 2006), with few examples in Mediterranean areas (del Castillo et al., 2016; Fernandes, 2013; Hernández-Santana et al., 2008; Mendes et al., 2016). There is a clear knowledge gap concerning the identification of such ecosystems, their phreatophyte associated vegetation (Robinson, 1958) in the Mediterranean region and the management actions that should be taken to decrease the adverse effects of climate change.

In the driest regions of the Mediterranean basin, the persistent lack of water during the entire summer periods selected plants with drought-avoiding strategies, like those that reach deeper stored water up to the point of relying on groundwater (Canadell et al., 1996; Miller et al., 2010). Groundwater access by deep rooting species is often associated to hydraulic lift and/or hydraulic redistribution mechanisms (Orellana et al., 2012). Those mechanisms provide the ability to move water from deep soil layers, where water content is higher, to more shallow layers where water content is lower (Horton and Hart, 1998; Neumann and Cardon, 2012). Hydraulic lift and redistribution have been reported for several woody species of the Mediterranean basin (David et al., 2007; Filella and Peñuelas, 2004) and noticeably for Cork oak (*Quercus suber* L.) (David et al., 2013; Kurz-Besson et al., 2006; Mendes et al., 2016).
Cork oak woodlands are agro-silvo-pastoral systems of the southwest Mediterranean basin (Joffre et al., 1999) who have already been referenced has a groundwater dependent terrestrial ecosystem (Mendes et al., 2016). In the ecosystems of this geographical area, the dominant tree species are the cork oak (Quercus suber L.) and the Portuguese holm oak (Quercus ilex subsp. rotundifolia Lam.) (Pinto-Correia et al., 2011). Additionally, stone pine (Pinus pinea L.) has become a commonly co-occurrent species in the last decades (Coelho and Campos, 2009). The use of groundwater has been frequently reported for both Pinus (Filella and Peñuelas, 2004; Grossiord et al., 2016; Peñuelas and Filella, 2003) and Quercus (Barbeta and Peñuelas, 2017; David et al., 2007, 2013, Kurz-Besson et al., 2006, 2014; Otieno et al., 2006) genders. Furthermore, the contribution of groundwater to tree physiology has been shown to be of a greater magnitude for Quercus sp. as compared with Pinus sp. (del Castillo et al., 2016; Evaristo and McDonnell, 2017).

Q. suber and Q. ilex have been associated with high resilience and adaptability to hydric and thermic stress, and to recurrent droughts in the southern Mediterranean basin (Barbero et al., 1992). In Italy and Portugal, during summer droughts Q. ilex used a mixture of rain-water and groundwater and was able to take water from very dry soils (David et al., 2007; Valenti et al., 1992). An increasing contribution of groundwater in the summer has also been shown for this species (Barbeta et al., 2015). Similarly, Q. suber showed a seasonal shift in water sources, from shallow soil water in the spring to the beginning of the dry period followed by a progressive higher use of deeper water sources throughout the drought period (Otieno et al., 2006). In addition, the species roots are known to reach depths as deep as 13m in southern Portugal (David et al., 2004). Although co-occurrent to cork and holm oaks species, there is still no evidence yet that P. pinea relies on groundwater resources during the dry season. However it shows a very similar root system (Montero et al., 2004) as compared to cork oak (David et al., 2013), with large sinker roots reaching 5m depth (Canadell et al., 1996). Given the information available on water use strategies by the phreatophyte arboreous species of the cork oak woodlands, we considered Q. ilex, Q. suber and P. pinea as proxies for vegetation that belongs to GDE relying on subsurface groundwater (from here onwards designed as Groundwater Dependent Vegetation – GDV).

GDV of the Mediterranean basin is often neglected in research. Indeed, still little is known about the GDV distribution, but research has already been done on the effects of climate change in specific species distribution, such as Q. suber, in the Mediterranean basin (Duque-Lazo et al., 2018; Paulo et al., 2015). While the increase in atmospheric CO₂ and the raising temperature can boost tree growth (Barbeta and Peñuelas, 2017; Bussotti et al., 2013; Sardans and Peñuelas, 2004), water stress can have a counteracting effect on growth of both Quercus ilex (López et al., 1997; Sabaté et al., 2002) and P. pinaster (Kurz-Besson et al., 2016). Therefore, it is of crucial importance to identify geographical areas where subsurface GDV is present and characterize the environmental conditions this vegetation type is thriving in. This would contribute to the understanding of how to manage these species under unfavorable future climatic conditions.

The aim of this study was to create a suitability map of the current distribution of GDV in southern Portugal, based on the occurrence of known and foreseen subsurface phreatophyte species and well-known environmental conditions affecting groundwater storage. Several environmental predictors were
selected according to their impact on groundwater use and storage and then used in a Geographically Weighted Regression (GWR) to model the density of *Q. suber*, *Q. ilex* and *P. pinea* occurrence in the Alentejo region (NUTSII) of southern Portugal. So far, very few applications of this method have been used to model species distribution and only recently its use has spread in ecological research (Hu et al., 2017; Li et al., 2016; Mazziotta et al., 2016). The coefficients obtained from the model equation for each predictor were used as weights to build the suitability map with GIS multi-factor analysis, after reclassifying each environmental predictor.

Based on the environmental conditions of the study area and the species needs, we hypothesized that 1) groundwater depth together with climatic conditions play one of the most important environmental roles in GDV’s distribution and 2) a more superficial access to groundwater and less arid conditions should allow a higher density of GDV. Therefore, a higher suitability should be expected under such conditions.

We start by presenting the methodology used to create the environmental variables for the study area of Alentejo, followed by an explanation of how the model was constructed and lead to the GDV suitability map and subsequent validation. In the result section in chapter 3, we display the maps for the environmental variables and parameters from the model fitting, the final suitability map and respective validation. The results are discussed in the fourth chapter and the conclusions are presented in the fifth chapter.
2 Material and Methods

2.1 Study area

The administrative region of Alentejo (NUTSII) (fig01) covers an area of 31 604.9 km², between the latitude 37.22° to 39.39° N and longitude 9.00° to 6.55° W. This study area is characterized by a Mediterranean temperate mesothermic climate with hot and dry summers, defined as Csa in the Köppen classification (APA, n.d.; ARH Alentejo, 2012a, 2012b). It is characterized by a sub-humid climate, which has recently quickly drifted to semi-arid conditions (Ministério da Agricultura do Mar do Ambiente e do Ordenamento do Território, 2013). A large proportion of the area (above 40%) is covered by forestry systems (Autoridade Florestal Nacional and Ministério da Agricultura do Desenvolvimento Rural e das Pescas, 2010) providing a high economical value to the region and the country (Sarmento and Dores, 2013).

2.2 Kernel Density estimation of GDV

Presence datasets of Quercus suber, Quercus ilex and Pinus pinea of the last Portuguese forest inventory achieved in 2010 (ICNF, 2013) were used to calculate Kernel density (commonly called heat map) as a proxy to GDV suitability. Only data points with one of the three proxy species selected as primary and secondary occupation were used. The resulting Kernel density was weighted according to tree cover percentage and was calculated using a quartic biweight distribution shape, a search radius of 10km, and an output resolution of 0.018 degrees, corresponding to a cell size of 1 km. This variable was computed using QGIS version 2.14.12 (QGIS Development Team, 2017).

2.3 Environmental variables

Species distribution is mostly affected by limiting factors (controlling ecophysiological responses), disturbances and resources (Guisan and Thuiller, 2005). To characterize the study area in terms of GDV suitability, environmental variables expected to affect GDV’s density were selected according to their constraint on groundwater uptake and soil water storage. Within possible abiotic variables, landscape topography, geology, groundwater availability and regional climate were considered to map GDV density in the study area. The twelve selected variables for modeling purposes, retrieved from different data sources are listed on Table 1.

2.3.1 Topography and Geology

NASA and METI ASTER GDEM product (https://lpdaac.usgs.gov) was retrieved from the online Data Pool. Spatial Analyst Toolbox from ArcGIS® software version 10.4.1 by Esri was used to calculate the
slope from the digital elevation model. Slope was used as proxy for the identification of superficial water interaction with vegetation.

The map of soil type was obtained from the Portuguese National Information System for the Environment - SNIAmb (© Agência Portuguesa do Ambiente, I.P., 2017) and uniformized to the World Reference Base with the Harmonized World Soil Database v 1.2 (FAO et al., 2009). The vector map was converted to raster using the Conversion Toolbox from ArcGIS® software version 10.4.1 by Esri. To reduce the analysis complexity involving the several soil types present in the map, soil types were regrouped in three classes, according to their capacity to store or drain water. The classification was based on the characteristics of each soil unit (available water storage capacity, drainage and topsoil texture) from the Harmonized World Soil Database (FAO et al., 2009).

Effective soil thickness (Table 1) represents the maximum soil depth explored by the vegetation roots. It constrains the expansion and growth of the root system, as well as the available amount of water that can be absorbed by roots.

2.3.2 Groundwater availability

Root access to groundwater is one of the most limiting factors for GDV’s growth and survival, especially during the dry season. The map of depth to water table was interpolated from piezometric observations from the Portuguese National Information System on Water Resources (SNIRH) public data base (http://snirh.apambiente.pt, last accessed on March 31st 2017) and the Study of Groundwater Resources of Alentejo (ERHSA) (Chambel et al., 2007). Data points of large-diameter wells and piezometers were retrieved for the Alentejo region (fig02) and sorted into undifferentiated, karst or porous geological types to model groundwater depth. Due to the large heterogeneity of geological media, groundwater depth was calculated separately for each sub-basin. A total of 3158 data points corresponding to large wells and piezometers were used, with uneven measurements between 1979 and 2017. For each piezometer an average depth was calculated from the available observations and used as a single value. In areas with undifferentiated geological type, piezometric level and elevation were highly correlated (>0.9), thus a linear regression was applied to interpolate data. Ordinary kriging was preferred for the interpolation of karst and porous aquifers, combining large wells and piezometric data points. To build a surface layer of the depth to water table, the interpolated surface of the groundwater level was subtracted from the digital elevation model. Geostatistical Analyst ToolBox from ArcGIS® software version 10.4.1 by Esri was used for this task.

Drainage density is a measure of how well the basin is drained by stream channels. It is defined as the total length of channels per unit area. Drainage density was calculated for each of the six hydrographic basins of the Alentejo region, by the division between the total stream length (L) in km and the basin area (A) in km², as in Eq. (1).

\[ Dd = \frac{L}{A} \]  \hspace{1cm} (1)
2.3.3 Regional Climate

Temperature and precipitation datasets were obtained from the E-OBS (http://eca.knmi.nl/download/ensembles/ensembles.php, last accessed on March 31\textsuperscript{st} 2017) public database (Haylock et al., 2008). Standardized Precipitation Evapotranspiration Index (SPEI), Aridity Index (AI) and Ombrothermic Indexes were computed from long-term (1951-2010) monthly temperature and precipitation observations. The computation of potential evapotranspiration (PET) was performed according to Thornthwaite (1948) and was assessed using the SPEI package (Beguería and Vicente-Serrano, 2013) in R program software version 3.4.2 (R Development Core Team, 2016).

SPEI multi-scalar drought index (Vicente-Serrano et al., 2010) was calculated over a 6 month interval to characterize drought severity in the area of study using SPEI package (Beguería and Vicente-Serrano, 2013) for R program. SPEI is based on the normalization of the water balance calculated as the difference between cumulative precipitation and PET for a given period at monthly intervals. Normalized values of SPEI typically range between -3 and 3. Drought events were considered as severe when SPEI values were between -1.5 and -1.99, and as extreme with values below -2 (Mckee et al., 1993). Severe and extreme SPEI predictors were computed as the number of months with severe or extreme drought, counted along the 60 years of the climate time-series.

Aridity index gives information related to evapotranspiration processes and rainfall deficit for potential vegetative growth. It was calculated following Eq. (2) according to Middleton et al. (1992), where PET is the average annual potential evapotranspiration and P is the average annual precipitation, both in mm for the 60 years period of the climate time-series. Dry lands are defined by their degree of aridity in 4 classes: Hyperarid (AI<0.05); Arid (0.05<AI<0.2); Semi-arid (0.2<AI<0.5) and Dry Subhumid (0.5<AI<0.65) (Middleton et al., 1992).

\[
AI = \frac{P}{PET} \quad (2)
\]

Ombrothermic Indexes were used to better characterize the bioclimatology of the study region (Rivas-Martínez et al., 2011), by evaluating soil water availability for plants during the driest months of the year. Four ombrothermic indexes were calculated following Eq. (3), where Pp is the positive annual precipitation (accumulated monthly precipitation when the average monthly mean temperature is higher than 0°C) and Tp is the positive annual temperature (total in tenths of degrees centigrade of the average monthly temperatures higher than 0°). If the ombrothermic index presents values above 2 for the analyzed months, the area cannot be considered as Mediterranean bioclimatically. For non-Mediterranean areas, there is no dry period in which, for at least two consecutive months, the precipitation is less than or equal to twice the temperature. Each ombrothermic index differed in the examined period of the year (Table1).

\[
i_o = \frac{P_p}{T_p} \quad (3)
\]

2.4 Predictors selection
The full set of environmental variables were evaluated as potential predictors for the suitability of GDV (based on the Kernel density of the proxy species). A preliminary selection was carried out, first by computing Pearson’s correlation coefficients between environmental variables and second by performing a Principal Components Analysis (PCA) to detect multicollinearity. Covariates were discarded for modeling according to a sequential procedure. Whenever pairs of variables presented a correlation value above 0.4, the variable with the highest explained variance on the first axis of the PCA was selected. Variables showing low correlations and explaining a higher cumulative proportion of variability with the lowest number of PCA axis were later selected as predictors for modeling. PCA was performed using the GeoDa Software (Anselin et al., 2006) and Pearson’s correlation coefficients were computed with Spatial Analyst Tool from ArcGIS software version 10.4.1 by Esri.

2.5 Model development

When fitting a linear regression model to the selected variables, we must assure a normal distribution and stationarity of the model residuals. However, spatial autocorrelation and non-stationarity are common when using linear regression on spatial data. To overcome these issues, Geographically Weighted Regression (GWR) was used to allow model coefficients to adjust to each location of the dataset, based on the proximity of sampling locations (Stewart Fotheringham et al., 1996). In this study, simple linear regression and GWR were both applied to the dataset and their performances compared. Models were fitted on a 5% random subsample of the entire dataset (6242 data points), due to computational restrictions.

Adaptive Kernel bandwidths for the GWR model fitting were used due to the spatial irregularity of the random subsample. Bandwidths were obtained by minimizing the CrossValidation score (Bivand et al., 2008). To analyze the performance of the GWR model alone, the local and global adjusted R-squared were considered. To compare between the GWR model and the simple linear model, we considered the distribution of the model residuals, e.g. whether there were visible clustered values. The second-order Akaike Information Criterion (AICc) was also contemplated. The spatial autocorrelation of the models residuals was evaluated with the Moran’s I test (Moran, 1950) using the Spatial Statistics Tool from ArcGIS software version 10.4.1 by Esri, and also graphically. GWR model was fitted using the spgwr package from R program version 3.4.2 (Bivand and Yu, 2017).

2.6 Suitability map building

To create the suitability map we proceeded with the classification of all predictor layers included in the GWR model, similarly to Condesso de Melo et al. (2015) and Aksoy et al. (2017). The likelihood of an interaction between the vegetation and groundwater resources was scored from 0 to 3 for each predictor. Scores were assigned after bibliographic review and expert opinion. The higher the score, the higher the likelihood, 1 corresponding to a weak likelihood and 3 indicating very high likelihood. Groundwater depth was divided in two classes, according to the accessibility to superficial water above 1.5m and the maximum rooting depth for Mediterranean woody species reaching 13 m, reported by Canadell et al.
The minimum score was given to areas where groundwater depth was too shallow (below 1.5m). This allowed to identify species dependent on more superficial groundwater which were considered to belong to other types of groundwater dependent vegetation. Areas with steep slope were considered to have superficial water flow (e.g. areas with permanent water table close to the surface due to proximity to permanent streams) and influence negatively tree density (Costa et al., 2008). Those areas were treated as less suitable to GDV. Aridity Index and Ombrothermic Index of the summer quarter and the immediately previous month (Ios4) values were split in 3 classes according to Jenks natural breaks, with higher suitability scores corresponding to higher aridity.

A direct compilation of the predictor layers could have been performed, however not all predictors influence in the same measure the distribution of this type of vegetation. Therefore, there was a need to define weighting factors for each layer of the final GIS multicriteria analysis. Yet, due to the intricate relations between all environmental predictors and their effects on the GDV, experts and stakeholders provided very different scoring for a same layer. Subsequently, we instead chose to use the coefficients of the GWR model (Eq. 4) as weighting factors. The final GIS multicriteria analysis was performed using the Spatial Analyst Tool ArcGIS® software version 10.4.1 by Esri, resulting in the final suitability map.

In the latter, lower values indicate a lower probability of GDV occurrence while higher values correspond to a higher suitability. To allow for an easier interpretation, the data on suitability to GDV was subsequently classified based on their distribution value, according to Jenks natural breaks. This resulted in 5 suitability classes: “Very poor”, “Poor”, “Moderate”, “Good” and “Very Good”.

2.7 Map validation

To assess the quality of the suitability map obtained in the present study, independent maps of integrated suitability to Q. suber, Q. ilex and P. Pinea were retrieved from the EPIC WebGIS Portugal (http://epic-webgis-portugal.isa.ulisboa.pt/) public data base (Magalhães et al., 2015a, 2015b, 2015c). Those distribution maps represent the suitability to a tree species according to bioclimatic, soil morphological conditions and best silvicultural practices (Magalhães et al., 2015a). By overlapping the maps of the three species in ArcGIS, we obtained a synthetic independent map where it was possible to identify suitable areas to none, one, two or three of the tree species, considered good proxies of GDV (fig. C1). Artificialized areas, rocky outcrops, rivers and humid areas were eliminated from the evaluation and validated maps before performing an analytical comparison using the Analysis Tool ArcGIS® software version 10.4.1 by Esri.
3 Results

3.1 Kernel Density

Within the studied region of Portugal, the phreatophyte species Quercus suber, Quercus ilex and the suspected phreatophyte species Pinus pinea were not distributed uniformly throughout the territory. Areas with higher Kernel density (or higher distribution likelihood) were mostly spread between the northern part of Alentejo region and the western part close to the coast, with values ranging between 900 and 1200 (fig03). Two clusters of high density also appeared below the Tagus river. The remaining study area presented mean density values, with a very low density in the area of the river Tagus.

3.2 Environmental conditions

The exploratory analysis of the variables, performed through the PCA and Pearson correlation matrix confirmed the presence of multicollinearity. From the initial variables (Table 1), Thickness, Drainage Density, Spei_severe, Spei_extreme, Annual Ombrothermic Index (Io), Ombrothermic Index of the hottest month of the summer quarter (Ios1) and Ombrothermic Index of the summer quarter (Ios3) were discarded, while the variables slope, soil type, depth, AI and Ios4 were maintained for analysis (fig. A1 and Table A1 in appendix). Therefore, five environmental variables out of the initial 12 considered (fig04) were endorsed to explain the variation of the Kernel density of GDV in Alentejo: soil type, ombrothermic index of the summer quarter and the immediately previous month (Ios4), slope, aridity index and groundwater depth.

The Alentejo region showed high heterogeneity of soil types, with 27 different categories (fig04a). In most part of the region, slope was below 10% (fig04b), coastal areas presenting the lowest values and variability. Highest values of groundwater depth (fig04c), reaching a maximum of 255m, were found in the Atlantic margin of the study area, mainly in Tagus and Sado river basins. Several other small and confined areas in Alentejo also showed high values, corresponding to aquifers of porous or karst geological types. Most of the remaining study area showed groundwater depths ranging between 1.5m and 15m. Figures 4d and 4e indicate the southeast of Alentejo as the driest area, given by minimum values of the aridity index (0.618), and potential evapotranspiration much higher that precipitation. Besides, Ios4 presented a maximum value (0.714) for this region (meaning that water availability in the soil was not compensated by the precipitation of the previous months).

Combining all variables, it was possible to distinguish two sub-regions with distinct conditions: the southeast of Alentejo and the Atlantic margin. The latter is mainly composed of podzols and regosols, low slope areas and more humid climatic conditions than the southeast of Alentejo.

3.3 Regression models
The Kernel density of the proxy GDV species, *Q. suber*, *Q. ilex* and *P. pinea*, showed a skewed normal distribution. Therefore, a square-root normalization of the data was applied on this response variable, before model fitting. To be able to compare the resulting model coefficients and use them as weighting factors, the multi-criteria analysis to build the suitability map, the predictor variables were normalized using the z-score function. The equation resulting from the GWR model fitting, featuring the predictor coefficients (Table 2) used later for the computation of the GDV suitability map corresponds to Eq. (4).

Local adjusted R-squared was highly variable throughout the study area, ranging from 0.25 to 0.95 (fig05). Lower R-squared values were clustered, near the Tagus river basin and in central and northern Alentejo. The overall fit of the GWR model was high (Table 3). The adjusted regression coefficient indicated that 92% of the variation in the data was explained by the GWR model, while only 11% was explained by the simple linear model (Table 3). Accordingly, GWR had a substantially lower AICc when compared with the simple linear model, indicating a much better fit.

The analysis of spatial autocorrelation, given by the Moran Index, showed a Z-score of 107.79 for the GWR model, with a considerable reduction of the Moran Index between models, from 0.94 in the simple linear regression model to 0.67 in the GWR model. From figures 06a and 06b there is an evident decrease in clustered residual values from the simple linear model to GWR. In the linear model (fig06a), positive residuals were condensed in the right side of Tagus and Sado river basins, while negative values were mainly present on the left side of the Tagus river and in the center-south of Alentejo. In GWR model (fig06b) the condensed positive and negative residual values were much more scattered throughout the study region, highlighting a much better performance of the GWR, which minimized residual autocorrelation.

\[
\text{Density} = 23.88 + 0.22 \text{Las4} - 1.61 \text{Al} + 0.06 \text{Depth} + 1.33 \text{Soil type 2} + 2.46 \text{Soil type 3} + 0.14 \text{Slope} , \quad (4)
\]

### 3.4 GDV Suitability map

The classification of the 5 endorsed environmental predictors is presented in Table 4 and their respective maps in figure B1 in appendix. Rivers Tagus and Sado had a positive impact on GDV’s suitability to each predictor. This is due to a higher water availability reflected by the values of omborhermic and aridity indexes (figs. B1a and B1b in appendix) and a higher groundwater depth in the surroundings of the rivers (fig. B1c in appendix). Optimal conditions for groundwater access were mainly gathered in the interior of the study region (fig07), with the exception for some confined aquifers. Favorable slopes for GDV were mostly highlighted in the Tagus river basin area, where a good likelihood of interaction between GDV and groundwater could be identified (fig. B1d in appendix). However, this high likelihood was hindered by the type of soil present in that area (fig. B1e in appendix).

The final map illustrating the suitability to GDV is shown in Fig. 7. The largest part of the study area (17 538 km²), representing more than half of the total area (55.8%) showed a very good suitability to the occurrence of GDV. The rest of the territory showed a “Moderate” to “Poor” suitability, representing...
5,037 km² and 4,313 km², respectively. Altogether, 1/3 of the total area showed “Very poor” to “Moderate” suitability to GDV, corresponding to the most southern and eastern part of the study region. The suitability to GDV in the Alentejo region was mainly driven by soil type, given by the similar distribution pattern between the suitability map and the soil type predictor (fig04a and fig07). This was also confirmed by the high coefficient obtained for the soil type predictor in the GWR model equation. The aridity index also showed a strong influence on GDV’s suitability, mostly for the intermediate and good classes. Areas with high suitability classes corresponded to the most northern and coastal areas of Alentejo region. Areas with intermediate class in the north of the study region mostly matched with soil type polygons, with score 1 and 2 (figB1e in appendix), while high aridity values restricted GDV’s suitability in the south. Areas with a good suitability mostly coincided with polygons of soil type 3 and with lower values of aridity index in the northern region of Alentejo.

### 3.5 Map validation

To assess the quality of the suitability map developed in the present study, we compared our results with integrated suitability maps, from different data sources than those used in this study, for each of the previously considered proxy species. The integrated suitability maps of each proxy species were aggregated into one validation map. Both the result and validation maps were highly coincident, especially with respect to areas with lower and moderate suitability to GDV (Table 5). Areas with very poor GDV suitability corresponded to almost 76% of the non-suitable areas for proxy species in the validation map. Accordingly, poor suitability areas for GDV matched 36.65% of the non-suitable areas for proxy species and 45.27% of areas suitable to only one of the proxy species. Besides, areas with moderate GDV suitability matched almost half of the suitable areas for two of the proxy species in the validation map. Classes with higher GDV suitability did not show a good agreement with the validation map.

When juxtaposing both maps (fig07 and fig08), there was an overall correspondence between areas with higher suitability to the proxy species. In both maps the northern and coastal area of the Alentejo region, south of the Tagus river basin, showed a matching higher suitability to the proxy species and the GDV. The Sado region was a common area of high suitability in both maps as well. The largest mismatches between maps were found in the center and southeast of the study region. Temporary irrigated areas matched non-suitable areas for proxy species in the validation map (fig C1 in appendix). This could explain some of the mismatches highlighted before, particularly where a large percentage of good and very good GDV suitability (28 and 41% respectively) corresponds to a non-suitable area for each of the proxy species in the validation map (Table 5).
4 Discussion

4.1 Modeling approach

Mapping the suitability of regional Groundwater Dependent Vegetation in southern Portugal proved to be a challenge because of the intricate relations between topographical, hydrological and biotic conditions in this specific area of the Mediterranean basin. Only 50% of the initial predictors were assigned for model fitting, due to a high collinearity between variables of the same type (e.g. aridity index and SPEI variables). Nevertheless, the small number of elected predictors for modeling will provide a higher reliability of the forecast of GDV suitability under future predicted environmental conditions.

Despite the exclusion of redundant predictors, the spatial distribution of residues after fitting the simple linear model still showed a significant clustered pattern, which violated the basic assumption of independence between samples. Therefore, a Geographically Weighted Regression model was used according to Stewart Fotheringham et al. (1996). This spatial variation of the linear model has been used before in ecological studies (Li et al., 2016; Mazzotta et al., 2016), but never for the mapping of GDV, to our knowledge. This approach considerably improved the goodness of fit, with a coefficient of regression $R^2$ increasing from 0.11 to 0.92 at the global level, and an obvious reduction of residual clustering. Despite those improvements, it has not been possible to completely eliminate the residual autocorrelation after fitting the GWR model.

Kernel density for the study area provided a strong indication of presence and abundance of the tree species considered as GDV proxy for modeling. Mediterranean cork woodlands are very particular agroforestry systems present in a confined area of the Mediterranean basin, where sparse tree distributions dominate, because of silvicultural management to increase cork and acorn production, while providing a large grassland area for cattle (Bugalho et al., 2009). However, anthropologic management of agroforestry systems in this region has not been considered in the model. This could, at least partially, explain the non-randomness of the residual distribution after GWR model fitting.

Another explanation of the reminiscent autocorrelation after GWR fitting could be the lack of groundwater dependent species in the model. For example, we decided to exclude Pinus pinaster Aiton due to its more humid distribution in Portugal, and due to conflicting conclusions driven from previous studies to pin point the species as a potential groundwater use (Bourke, 2004; Kurz-Besson et al., 2016).

In addition, only recently the use of groundwater by an olive orchard has been proved (Ferreira et al., 2018), however with little groundwater flow amount of the daily root flow, and so with no significant impact for the species physiological conditions.

Methods previously used by Doody et al., (2017) and Condesso de Melo et al. (2015) to map specific vegetation relied solely on expert opinion, e.g. Delphi panel, to define weighting factors of environmental information for GIS multicriteria analysis. In our study, we used GWR to assess weighting factors for each environmental predictor in the study area, to build a suitability map for the GDV in southern
Portugal. This allowed an empirical determination of the relevance of each environmental predictor in GDV distribution, thus avoiding the inevitable subjectivity of Delphi panels.

### 4.2 Suitability to Groundwater Dependent Vegetation

As shown by the simulations of future climate conditions based on RCP4.5 and RCP8.5 emission scenarios (Soares et al., 2015, 2017), a significant decrease of precipitation for the Guadiana basin and overall decrease for the southern region of Portugal are expected. Agroforestry systems relying on groundwater resources, such as cork oak woodlands, may show a decrease in productivity and ecosystem services or even face sustainability failure. Therefore, linking GDV to key environmental drivers and especially climate, will allow to forecast ecological changes under future climatic conditions and spot priority areas for adaptation and stakeholder decision.

According to our results, more than half of the study area appears suitable for GDV. However, one third of the studied area showed the lowest suitability to GDV. The lower suitability to this vegetation in the eastern part of the studied area can be explained by less favorable climatic and geological conditions, resulting from the combination of a high aridity index and low water retention at deep soil layers. Soils with lower water capacity further lowered GDV suitability in the most southeastern region of the studied area.

An increase in aridity and drought frequency for the Mediterranean (Spinoni et al., 2017) will most probably induce plant physiological adaptations (Peñuelas and Filella, 2001) as well as species substitution (Lloret et al., 2004). In the Mediterranean environment, Peñuelas et al. (2011) distinguished two plant communities regarding water consumption, one with deep roots, able to constantly access water and nutrients and a second community with shallow roots, depending on superficial water from rainfall. Due to climate change conditions, in the Alentejo region of Portugal, we should expect GDV of the less suitable areas to be replaced by the community with shallow roots using rainfall water exclusively. This has already been reported for a Mediterranean woodland of the Iberian Peninsula, where extreme drought conditions led to a shift in vegetation cover from deep-rooting species to water spending species (Caldeira et al., 2015). Groundwater reservoirs would thus no longer be a constraint for plant survival during summer droughts, because the supplanting vegetation community, namely annuals that stop their growing cycle or die before the onset of the dry season, would no longer need constant access to water. Such species substitution would be associated with ecological and biodiversity costs, by shifting from woodland to shrubland ecosystems.

In environments with scarce water sources such as the Mediterranean basin, many tree species have adapted to the precipitation’s seasonality and its large variability by developing dimorphic root systems. When comparing different water limited ecosystems from a global dataset, Schenk and Jackson (2002) showed that rooting depth increased with aridity. Our results agree with these findings since the aridity index was the second most important predictor of GDV density, according to our equation. Nevertheless, the soil type turned out to be the most important predictor of GDV density. This is comprehensible because the soil type defines the capacity for groundwater storage and the accessibility for deep root
system (Centenaro et al., 2017; Grimaldi et al., 2015). However, the soil type component is not expected
to change as dramatically as the aridity index in response to climate change, leaving the aridity index as
the main driver for the GDV density under climatic changes in southern Portugal.

As stated by our model equation, groundwater depth appeared to have little influence on GDV density.
This disagrees with our initial hypothesis. However, this disagreement should be regarded cautiously due
to the poor quality of the data used. On one hand, data points in the study region were highly
heterogeneous, and certain areas showed a better statistical representation than others. Moreover, the high
variability in geological media, topography and vegetation cover at the regional scale did not allow to
account for small changes in groundwater depth (<15m deep), which has a huge impact on GDV
suitability (Canadell et al., 1996; Stone and Kalisz, 1991). Indeed, a high spatial resolution of
hydrological database is essential to characterize the spatial dynamics of groundwater depth between
hydrographic basins (Lorenzo-Lacruz et al., 2017). However, such resolution was not available for our
study area. In addition, the lack of temporal data did not allow the calculation of seasonal trends in
groundwater depth, which are essential under Mediterranean conditions to build a reliable interpolation of
observed data. Temporal data would also further help discriminate areas of optimal suitability to GDV,
either during the wet and the dry seasons.

Overall, the map of suitability to GDV showed good results, compared with the validation maps showing
the proxy species integrated suitability (Magalhães et al., 2015a, 2015b, 2015c). However, areas of high
suitability to GDV matched areas of the validation that were non-suitable areas for the proxy species. This
can be explained by the lack of information in the model concerning main land occupation and land
management in the studied region. We found that areas where the main land occupation is non-
silvicultural (e.g. temporary irrigation fields), corresponded to non-suitable areas for proxy species in the
validation map. Several other discrepancies can be explained by additional information considered in the
validation map by their authors, such as the current occupation type (e.g. olive orchards, vineyards or
urban).

The main areas showing good suitability are mostly matching in both maps. Furthermore, our results
highly agree with Paulo et al. (2015) who predicted site productivity index and soil variables for cork oak
(Q. suber) stands in Portugal with a stochastic modelling approach. This allows us to apply the
methodology to extend our findings for larger geographical areas such as the Iberian Peninsula. Also, the
model equation can be considered reliable to simulate the impact of future climatic conditions on the
distribution of GDV in southern Portugal.

### 4.3 Key limitations

With the methodology applied in this study, weighting factors can be easily evaluated solely from local
and regional observations of the studied area. Nonetheless, either the computation of model coefficients
or expert opinion to assess weighting factors, require update, and/or environmental data, species
distribution and revised expert knowledge (Doody et al., 2017).
Changes in climate conditions only represent part of the water resources shortage issue in the future. Global-scale changes in human populations and economic progresses also rule water demand and supply, especially in arid and semi-arid regions (Vörösmarty et al., 2000). A decrease in useful water resources for human supply can induce an even higher pressure on groundwater resources (Döll, 2009), aggravating the water table drawdown caused by climate change (Ertürk et al., 2014). Therefore, additional updates of the model should include human consumption of groundwater resources, identifying areas of higher population density or intensive farming. Future model updates should also account for the interaction of deep rooting species with the surrounding understory species. In particular, shrubs surviving the drought period, which can benefit from the redistribution of groundwater by deep rooted species (Dawson, 1993; Zou et al., 2005).
5 Conclusions

The current pressure applied by human consumption of water sources has reinforced the concern on the future of economic activities dependent on groundwater resources. To address this issue, several countries have developed national strategies for the adaptation of water sources for Agriculture and Forests against Climate Change, including Portugal (FAO, 2007). In addition, local drought management as long-term adaptation strategy has been one of the proposals of Iglesias et al. (2007) to reduce the climate change impact on groundwater resources in the Mediterranean. The preservation of Mediterranean agroforestry systems, such as cork oak woodlands and the recently associated P. pinea species, is of great importance due to their high socioeconomic value and their supply of valuable ecosystem services (Bugalho et al., 2011). Management policies on the long-term should account for groundwater resources monitoring, accompanied by defensive measures to ensure agroforestry systems sustainability and economical income from these Mediterranean ecosystems are not greatly and irreversibly threatened.

Our present study, and novel methodology, provides an important tool to help delineating priority areas of action for species and groundwater management, at regional level, to avoid the decline of productivity and cover density of the agroforestry systems of southern Portugal. This is important to guarantee the sustainability of the economical income for stakeholders linked to the agroforestry sector in that area. Furthermore, mapping vulnerable areas at a small scale (e.g., by hydrological basin), where reliable groundwater depth information is available, should provide further insights for stakeholder to promote local actions to mitigate climate change impact on GDV.

Based on the methodology applied in this work, future predictions on GDV suitability, according to the RCP4.5 and RCP8.5 emission scenarios will be shortly computed, providing guidelines for future management of these ecosystems in the allocation of water resources.
6 Acknowledgements

The authors acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (http://ensembles-eu.metoffice.com) and the data providers in the ECA&D project (http://www.ecad.eu). The authors also wish to acknowledge the ASTER GDEM data product, a courtesy of the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, https://lpdaac.usgs.gov/data_access/data_pool. We are grateful to ICNF for sharing inventory database performed in 2010 in Portugal continental. We also thank Célia Gouveia, Cristina Catita, Ana Russo and Patrícia Páscoa for the advice and helpful comments. We are very grateful to Eric Font for the useful insights on soil properties. I Gomes Marques and research activities were supported by the Portuguese National Foundation for Science and Tecnology (FCT) through the PIEZAGRO project (PTDC/AAG-REC/7046/2014).

The authors declare that they have no conflict of interest.
References


Autoridade Florestal Nacional and Ministério da Agricultura do Desenvolvimento Rural e das Pescas: 5o Inventário Florestal Nacional, 2010.


Figure and Table Legends

Table 1: Environmental variables for the study area characterization in suitability to Groundwater Dependent Vegetation.

Table 2: Coefficients of determination resulting from the application of GWR model between GDV density and the selected predictive variables.

Table 3: Comparison of Adjusted R-squared and second-order Akaike Information Criterion (AICc) between simple regression and GWR models.

Table 4: Classification scores for each predictor. A score of 1 was given to areas less suitable and 3 to highly suitable areas.

Table 5: Interception (in %) between the classes of the GDV suitability map classes and the Overlapped Integrated suitability map. Value of “0” in overlapped integrated suitability map represent the non-suited area for all the proxy species; value of “1” represent the suitable area for 1 of the proxy species; value of “2” represent the suitable area for 2 of the proxy species and value of “3” represent the suitable area for all the proxy species.

Table A1: Squared correlations between predictor variables and principal components axis. The most important predictors for each axis (when squared correlation is above 0.3) are showed in bold.

Figure 01: Study area. On the left the location of Alentejo in the Iberian Peninsula; on the right, the elevation characterization of the study area with the main river courses from Tagus, Sado and Guadiana basins. Names of the main rivers are indicated near to their location in the map.

Figure 02: Large well and piezometer data points used for Water Table Depth calculation. Squares represent piezometers data points and triangle represent large well data points.

Figure 03: Map of Kernel Density weighted by cover percentage of Q. suber, Q. ilex and P. pinea.

Figure 04: Map of environmental layers used in model fitting. (a) – Soil type; (b) – Slope; (c) – Groundwater Depth (Depth); (d) – Ombrothermic Index of the summer quarter and the immediately previous month (Ios4); (e) – Aridity Index (AI).

Figure 05: Spatial distribution of local R^2 from the fitting of the Geographically Weighted Regression.

Figure 06: Spatial distribution of residuals from the fitting of the Simple Linear model (a) and Geographically Weighted Regression (b).

Figure 07: Suitability map for Groundwater Dependent Vegetation.

Figure 08: Validation map corresponding to the juxtaposition of the integrated suitability maps for each of the proxy species Q. suber, Q. ilex and P. pinea. Areas suitable for more than 1 or more proxy species are represented with a gradient of brown colors. Rivers and dams are indicated in blue and artificialized areas in grey.

Figure A2: Correlation plot between predictors used for fitting the simple linear model and the GWR model. AI is Aridity Index; Depth is Groundwater Depth (Depth) and Ios4 is the Ombrothermic Index of the summer quarter and the immediately previous month.
Figure B1 – Predictors maps after classification. (a) – Ombrothermic Index of the summer quarter and the immediately previous month (Ios4); (b) – Aridity Index (AI); (c) – Groundwater Depth (Depth); (d) – Slope; (e) – Soil type.
Table 1: Environmental variables for the study area characterization in suitability to Groundwater Dependent Vegetation.

<table>
<thead>
<tr>
<th>Variable code</th>
<th>Variable type</th>
<th>Source</th>
<th>Resolution and Spatial extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope</td>
<td>Slope (%)</td>
<td>This work</td>
<td>0.000256 degrees (25m) raster resolution</td>
</tr>
<tr>
<td>Soil type</td>
<td>Soil type in the first soil layer</td>
<td>SNIAmb (© Agência Portuguesa do Ambiente, I.P., 2017)</td>
<td>Converted from vectorial to 0.000256 degrees (25m) resolution raster</td>
</tr>
<tr>
<td>Thickness</td>
<td>Soil thickness (cm)</td>
<td>EPIC WebGIS Portugal (Barata et al., 2015)</td>
<td>Converted from vectorial to 0.000256 degrees (25m) resolution raster</td>
</tr>
<tr>
<td>Depth</td>
<td>Depth to groundwater (m)</td>
<td>This work</td>
<td>0.000256 degrees (25m) raster resolution</td>
</tr>
<tr>
<td>Dd</td>
<td>Drainage Density</td>
<td>This work</td>
<td>0.000256 degrees (25m) raster resolution</td>
</tr>
<tr>
<td>Spei_severe</td>
<td>Number of months with severe SPEI</td>
<td>This work</td>
<td>Time coverage 1950-2010</td>
</tr>
<tr>
<td>SPEI_extreme</td>
<td>Number of months with extreme SPEI</td>
<td>This work</td>
<td>0.000256 degrees (25m) raster resolution</td>
</tr>
<tr>
<td>A1</td>
<td>Aridity Index</td>
<td>This work</td>
<td>0.000256 degrees (25m) raster resolution</td>
</tr>
<tr>
<td>Io</td>
<td>Annual Ombrothermic Index</td>
<td>This work</td>
<td>0.000256 degrees (25m) raster resolution</td>
</tr>
<tr>
<td>Ios1</td>
<td>Ombrothermic Index of the hottest month of the summer quarter (J, J and A)</td>
<td>This work</td>
<td>0.000256 degrees (25m) raster resolution</td>
</tr>
<tr>
<td>Ios3</td>
<td>Ombrothermic Index of the summer quarter (J, J and A)</td>
<td>This work</td>
<td>0.000256 degrees (25m) raster resolution</td>
</tr>
<tr>
<td>Ios4</td>
<td>Ombrothermic Index of the summer quarter and the immediately previous month (M, J, J and A)</td>
<td>This work</td>
<td>0.000256 degrees (25m) raster resolution</td>
</tr>
</tbody>
</table>
Table 2: Coefficients of determination resulting from the application of GWR model between GDV density and the selected predictive variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Minimum</th>
<th>1st Quartile</th>
<th>Median</th>
<th>3rd Quartile</th>
<th>Maximum</th>
<th>Global</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-48.55</td>
<td>16.01</td>
<td>23.88</td>
<td>29.16</td>
<td>94.65</td>
<td>13.86</td>
</tr>
<tr>
<td>IoS4</td>
<td>-18.31</td>
<td>-2.47</td>
<td>0.22</td>
<td>3.13</td>
<td>16.29</td>
<td>-0.22</td>
</tr>
<tr>
<td>AI</td>
<td>-48.27</td>
<td>-11.22</td>
<td>-1.61</td>
<td>5.48</td>
<td>64.87</td>
<td>-0.72</td>
</tr>
<tr>
<td>Depth</td>
<td>-32.30</td>
<td>-1.08</td>
<td>0.06</td>
<td>0.95</td>
<td>33.25</td>
<td>0.43</td>
</tr>
<tr>
<td>Soil type (2)</td>
<td>-19.78</td>
<td>-3.4</td>
<td>0.33</td>
<td>3.97</td>
<td>24.32</td>
<td>3.98</td>
</tr>
<tr>
<td>Soil type (3)</td>
<td>-20.18</td>
<td>-0.48</td>
<td>2.46</td>
<td>5.13</td>
<td>23.17</td>
<td>7.62</td>
</tr>
<tr>
<td>Slope</td>
<td>-2.88</td>
<td>-0.18</td>
<td>0.14</td>
<td>0.68</td>
<td>4.75</td>
<td>-0.13</td>
</tr>
</tbody>
</table>

Table 3: Comparison of Adjusted R-squared and second-order Akaike Information Criterion (AICc) between simple regression and GWR models.

<table>
<thead>
<tr>
<th>Model</th>
<th>R-squared</th>
<th>AICc</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.11</td>
<td>42276</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>GWR</td>
<td>0.92 *</td>
<td>27795</td>
<td>-</td>
</tr>
</tbody>
</table>

*Quasi-global R²

Table 4: Classification scores for each predictor. A score of 1 was given to areas less suitable and 3 to highly suitable areas.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Class</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope</td>
<td>0%-5%</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>5%-10%</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>&gt;10%</td>
<td>1</td>
</tr>
<tr>
<td>Soil type</td>
<td>Eutric Cambisols; Dystric Regosol; Humic Cambisols; Haplic Luvisols; Gleyic Luvisols; Ferric Luvisols; Chromic Luvisols associated with Haplic Luvisols; Ortic Podzols</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Calcaric Cambisols; Dystric Regosol associated with Umbric Leptosols; Eutric Regosol; Vertic Luvisols; Eutric Planosols; Cambic Arenosols</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Chromic Cambisols; Eutric fluvisols; Chromic Luvisols; Gleyic Solonchak; Eutric Vertisols</td>
<td>1</td>
</tr>
<tr>
<td>Groundwater Depth</td>
<td>&gt;15 m</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.5m-15m</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>≤1.5m</td>
<td>1</td>
</tr>
<tr>
<td>Aridity Index</td>
<td>0.6-0.68</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>0.68-0.75</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>≥0.75</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>&lt;0.28</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>0.28-0.64</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>≥0.64</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 5: Interception (in %) between the classes of the GDV suitability map classes and the Overlapped Integrated suitability map. Value of “0” in overlapped integrated suitability map represent the non-suitable area for all the proxy species; value of “1” represent the suitable area for 1 of the proxy species; value of “2” represent the suitable area for 2 of the proxy species and value of “3” represent the suitable area for all the proxy species.

<table>
<thead>
<tr>
<th>GDV suitability</th>
<th>Validation map (Integrated Suitability for 0 to 3 of the proxy species)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Very Poor</td>
<td>75.67</td>
</tr>
<tr>
<td>Poor</td>
<td>36.65</td>
</tr>
<tr>
<td>Moderate</td>
<td>33.17</td>
</tr>
<tr>
<td>Good</td>
<td>38.38</td>
</tr>
<tr>
<td>Very Good</td>
<td>41.124</td>
</tr>
</tbody>
</table>
Figure 01: Study area. On the left the location of Alentejo in the Iberian Peninsula; on the right, the elevation characterization of the study area with the main river courses from Tagus, Sado and Guadiana basins. Names of the main rivers are indicated near to their location in the map.

Figure 02: Large well and piezometer data points used for Water Table Depth calculation. Squares represent piezometers data points and triangle represent large well data points.
Figure 03: Map of Kernel Density weighted by cover percentage of Q. suber, Q. ilex and P. pinea.
Figure 04: Map of environmental layers used in model fitting. (a) – Soil type; (b) – Slope; (c) – Groundwater Depth; (d) – Ombrothermic Index of the summer quarter and the immediately previous month (Ios4); (e) – Aridity Index (AI).

Figure 05: Spatial distribution of local R² from the fitting of the Geographically Weighted Regression.

Figure 06: Spatial distribution of residuals from the fitting of the Simple Linear model (a) and Geographically Weighted Regression (b).
Figure 07: Suitability map for Groundwater Dependent Vegetation.

Figure 08: Validation map corresponding to the juxtaposition of the integrated suitability maps for each of the proxy species *Q. suber*, *Q. ilex* and *P. pinea*. Areas suitable for more than 1 or more proxy species are represented with a gradient of brown colors. Rivers and dams are indicated in blue and artificialized areas in grey.
Appendix A

Table A1: Squared correlations between predictor variables and principal components axis. The most important predictors for each axis (when squared correlation is above 0.3) are showed in bold.

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope</td>
<td>&lt;0.01</td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Al</td>
<td>0.67</td>
<td>0.02</td>
<td>&lt;0.001</td>
<td>&lt;0.01</td>
<td>0.31</td>
</tr>
<tr>
<td>Ios4</td>
<td>0.18</td>
<td></td>
<td>0.24</td>
<td>0.03</td>
<td>0.10</td>
</tr>
<tr>
<td>Depth</td>
<td>0.43</td>
<td>&lt;0.01</td>
<td>0.06</td>
<td></td>
<td>0.45</td>
</tr>
<tr>
<td>Soil type</td>
<td>0.33</td>
<td>0.25</td>
<td>0.05</td>
<td>0.29</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Figure A2: Correlation plot between predictors used for fitting the simple linear model and the GWR model. AI is Aridity Index; Depth is Groundwater Depth (Depth) and Ios4 is the Ombrothermic Index of the summer quarter and the immediately previous month.
Appendix B

![Map Image (a)]

![Map Image (b)]

![Map Image (c)]

![Map Image (d)]

Classification Score:
- Red: NoData
- Light Green: 1
- Medium Green: 2
- Dark Green: 3

![Map Image (e)]
Figure B1 – Predictors maps after classification. (a) – Ombrothermic Index of the summer quarter and the immediately previous month (Ios4); (b) – Aridity Index (AI); (c) – Groundwater Depth (Depth); (d) – Slope; (e) – Soil type.