CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

FACULTY OF ENVIRONMENTAL SCIENCE

DEPARTMENT OF ECOLOGY

Predictive distribution modeling of European Tree species using Pliocene and future climatic data

Diploma Thesis 2023

Author: Leandro Eusebio

Supervisor: Ing. Jan Douda, Ph.D.

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Environmental Sciences

DIPLOMA THESIS ASSIGNMENT

Leandro Eusebio

Nature Conservation

Thesis title

Predictive distribution modelling of European tree species using Pliocene and future climatic data

---

Objectives of thesis

The aim of the master thesis is to perform predictive modelling of the future distribution of temperate and boreal trees based on Pliocene climatic data and the distribution of generas. The main idea of the thesis is to use Mid-Pliocene climatic and paleoecological data for predictions of the effects of ongoing climatic changes on the distribution of tree taxa. Mid-Pliocene climate was analogous to predictions of future climate and CO2 atmospheric concentration. Therefore the distribution of temperate and boreal trees in this period could be used for cross-comparison with the predictive modelling on current species distribution data. The principal question is: Will species distribution models of the Mid-Pliocene Warm Period give better insight into the future distribution of tree species?

Methodology

Species distribution models will be constructed projecting species distribution according to Mid-Pliocene and future climates. The inferred distribution by SDMs in the Mid-Pliocene will be compared with paleoe-cological fossil data. The species will be divided into biogeographical-ecological groups and changes in their distribution among the current distribution, the Mid-Pliocene distribution and the future distribution will be compared and discussed.

The proposed extent of the thesis

30-60

Keywords

climate change, Europe, fossil data, species distribution models, tree species

Recommended information sources

Svenning, J. C., Fløjgaard, C., Marske, K. A., Nógues-Bravo, D., & Normand, S. (2011). Applications of species distribution modeling to paleobiology. Quaternary Science Reviews, 30(21-22), 2930-2947.

Svenning, J. C., Normand, S., & Kageyama, M. (2008). Glacial refugia of temperate trees in Europe: insights from species distribution modelling. Journal of Ecology, 96, 1117-1127.

Svenning, J. C., & Skov, F. (2007). Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? Ecology Letters, 10, 453-460.

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD–a platform for ensemble forecasting of species distributions. Ecography, 32, 369-373.

Expected date of thesis defence

2022/23 SS – FES

The Diploma Thesis Supervisor

Ing. Jan Douda, Ph.D.

Supervising department

Department of Ecology

Electronic approval: 22. 2. 2023

prof. Mgr. Bohumil Mandák, Ph.D.

Head of department

Electronic approval: 23. 2. 2023

prof. RNDr. Vladimír Bejček, CSc.

Dean

Prague on 30. 03. 2023

I hereby declare that I wrote this thesis independently, under the direction of prof. Ing. Jan Douda, Ph.D.. I have listed all literature and publications used to acquire the information included in this thesis.

In Prague, 30.03.2023 Leandro Eusebio

Acknowledgements

I would like to thank my wife for supporting me throughout the writing of this thesis. I would also like to thank my family for their moral support. Lastly, I would like to acknowledge my supervisor Jan Douda and the rest of the Plant Ecology Lab team for their insight and guidance regarding European forest ecosystems

Abstract

Forest ecosystems, especially the species that define them, often reflect the environmental conditions surrounding them. Thus logically, prominent forest species will be found in areas in which their ecological niche is suitable for the external conditions being experienced. This may be a cause for concern in European forests due to climate change. It can be inferred that the distribution of suitable areas for prominent tree species will shift in the coming decades due to changes in the temperature and water regimes. Warm periods have occurred in the Earth's past, the most recent is the mid-Pliocene Warm Period, mPWP, between 3 to 3.25 million years ago. This period represents an interval of warm climate which is very similar to future warming scenarios RCP 4.5 and RCP 8.5(Burke et al 2018). Thus, species distribution modeling of prominent European tree species for both mPWP and future scenarios may hold insight on the distribution of forest ecosystems within the respective time periods. Visualizing the similarities between mPWP and future tree distributions may aid in the conservation and protection of these ecosystems.

Keywords: climate change, Europe, fossil data, species distribution models, tree species

Abstrakt

Lesní ekosystémy, a především druhy které je definují, často odrážejí ekologické předpoklady, které je obklopují. Logicky, prominentní lesní druhy se tak budou nacházet v oblastech, kde jsou vnější podmínky vhodné pro jejich specifické ekologické znaky. Kvůli změně klimatu toto může být důvod k obavám ohledné evropských lesů. Lze usuzovat, že vlivem změn teplot a vodních režimů se rozložení vhodných ploch pro tyto dřeviny v příštích desetiletích posune. V historii Země se vyskytovala teplá období. Naposledy to bylo období středního pliocénu, mPWP, před 3 až 3,25 miliony lety. Toto období představuje interval teplého klimatu, který je velmi podobný scénářům budoucího oteplování, RCP 4,5 a RCP 8,5 (Burke et al 2018). Modely distribuce druhů, budoucí i z pliocénu, by tak mohly nabízet pohled na rozložení lesních ekosystémů v příslušných obdobích. Vizualizace podobností mezi výskytem stromů v pliocénu a budoucnosti může pomoci při zachovávání a ochraně těchto ekosystémů.

Table of Contents

1. Introduction

It is widely known that groups of tree species are often associated with general climatic conditions of the region. For example, the boreal forests of Northern Europe are almost solely dominated by coniferous tree species such as *Picea abies* & *Pinus sylvestris*. Further associations can be made that mid-latitude temperate regions are dominated by oaks, beech, hornbeam, and maples; warm Mediterranean regions are dominated by evergreen oaks and members of the family *Cupressaceae*. It can be assumed that warming temperatures and changes to the water regimes of these tree bioregions of Europe will drastically change the suitable areas of these species in the future.

Global climate is reaching temperatures that have not been experienced since the Late-Tertiary in the Mid-Pliocene Warm Period or mPWP. This time period, 3.0 to 3.25 million years ago, represents an interval of global temperatures that exceed those we experience today. The warming experienced in the mPWP has been shown to be analogous to future climate scenarios RCP 4.5 and RCP 8.5(Burke et al 2018). Global average temperatures of the Pliocene were found to be between 1.8 ℃ to 3.6℃ warmer than modern conditions with a generally more humid climate. Future climate scenarios RCP 4.5 and RCP 8.5 represent a similar change in temperature and humidity. Thus the mPWP can be used as a representation of a stable climate phase with vegetation that has adapted and shifted their distributions during this time period’s climate.

It is likely that the distribution of Mediterranean, Temperate, and Boreal flora would have a different pattern due to different climatic conditions. There have been attempts to find the general distribution of different biomes according to pollen and macrofossil data, however these projects are largely in low resolution(Haywood et al 2002). Thus precise approximations of the biomes of Europe during the mPWP are lacking in the literature.

Dominant tree genera of Europe of this time are largely similar to the modern composition of species, with certain genera of East Asian and North American affinities becoming extinct during the Late-Pliocene and Early-Pleistocene glaciation events. The dominant tree genera however has largely been stable with principal components of European forests surviving throughout the Tertiary and into the Quaternary(Haywood et al 2002 & Salzmann et al 2011). Thus the use of modern species distribution methods and macrofossil evidence to create a high resolution visualization of Mid-Pliocene Europe that has not been seen before. Comparison of the species distribution maps from the Mid-Pliocene to future scenarios has the potential to show insight on areas of high conservation interest and high risk areas in high detail. This can aid in the future management and proactive action in high risk areas.

2. Aims of this Study

1) Review the current body of literature regarding Pliocene climate and vegetation and its comparison to future climate scenarios in the European continent.
   a) Estimations have been made for Pliocene climate and vegetation through analysis of sea sediment cores, macrofossils, and pollen. Specific species assemblages have been found in Pliocene sediments throughout Europe, but no comprehensive study has shown the general distribution of modern prominent tree species. Global estimations of Pliocene climate and vegetation have been created, although with low resolution. Thus there is lack of literature regarding the fine distribution of prominent tree species of the Pliocene
2) Creating accurate and precise estimations on habitat suitability of prominent European tree species of the Mediterranean, Temperate, and Boreal zones.
   a) Biomod2 will be used to create habitat suitability maps for prominent tree species in the three general bioclimatic regions of Europe. This will aid visualizing the potential distribution for modern, Pliocene, and future conditions from 2040 to 2100.
3) Investigate the similarities in the distribution of suitable habitat between Pliocene and future scenarios for all bioclimatic zones.
   a) The mid-Pliocene warm period ,3-3.25 million years ago, has been shown to be analogous to future climate scenarios RCP 4.5 and RCP 8.5. Thus, investigating the similarities and difference these two time periods can aid in identifying potential refugia of modern European forests in the past and future.
4) Give insight to high-risk areas and zones of high conservation interest.
   a) Shifts in suitable habitat highly affects the fitness of the organism. Investigating the shifts in vegetation in the past may aid in the identifying previously unknown suitable areas in the future. This can also visualize areas where suitable habitat has shifted out of reach for the tree species. Thus, can aid forest management to prevent future die-off and local extinctions of prominent tree species.

3. Literature Review

3.1 Ecological Niche Modeling

Niche modeling, also known as species distribution modeling, is a tool used to illustrate the relationship between a certain organism with its environment. Thus, it is commonly used as a measure of the suitable habitat in which an organism can occupy. Due to the coupling of environmental factors and species occurrence niche modeling has the ability to show the effect of shifting bioclimatic variables on the spatial distribution of suitable habitat for certain species. Niche modeling has the potential to predict suitable habitat among varying bioclimatic values in different geographic locations and geological time periods.

3.1.1 Maximum Entropy Methods

The maximum entropy method or MaxEnt is a well-established machine-learning algorithm that results in the mean probability of presence in the context of all background points. This coupled with Geographic Information System programs results in raster data where each pixel is representing the suitability of the species in focus (Svenning et al 2008).

A rising issue among niche-based models is that species occurrence data must have presence & absence. Presence data is much more accessible and in some cases absence data is not available. To overcome this issue MaxEnt was devised to perform niche-based models on presence-only data (Phillips et al 2006). Other benefits for using MaxEnt is that it can utilize both continuous & categorical variables. Furthermore, the resulting probability distribution has a clearly defined mathematical definition (Phillips et al 2006). A major drawback however is that MaxEnt, unlike other niche models ex. general linear & general boosted models, does not have a clear method of variable selection. Variable selection has a high importance due to its ability to cause statistical bias. Without the knowledge of statistical error MaxEnt models can suffer from overfitting (Phillips et al 2004). Another drawback, with high importance for this thesis, is the lowered predictive capability when applying this method in another biogeographical area or another time period from the source data (Higgins et al 2020). Nevertheless MaxEnt is by in large the most used species distribution model algorithm is maximum entropy method (Qazi et al 2022).

3.1.2 Generalized Linear Models & Generalized Boosted Models

Generalized linear models(GLM) are flexible general linear regressions that allow for the response variable, in this case species occurrence, to come in various distributions. As mentioned before MaxEnt has the ability to use presence-only data while GLMs are capable of handling presence-absence and abundance counts (Gusian et al 2000). Another advantage GLMs have over MaxEnt is the well-defined model-fitting process. The use of Analysis of Variance or Pearson's correlation coefficient has the ability to show significant and non-significant environmental factors (Antúnez 2017). This allows for the filtering of models so that only significant environmental factors can be used in the final model.

GBMs they all describe the same family of models that use both regression/classification tree algorithms and boosting (Friedman et al 2000 & Elith et al 2008). GBMs build upon GLMs potential for model fitting. GBMs use decision trees with exceedingly simple methods that employ the use of regression or categorical trees in order to predict data. Unlike general linear models GBMs do not use linear relationships that relate the predictor variables to the response; rather it uses a decision tree to make true or false statements that predict the occurrence of the species. (Friedman et al 2000 & Elith et al 2008). The final tree however is extremely difficult to interpret especially with a high number of predictors and large number of trees. This diminishes the predictor power of regression trees when used alone.

Boosting, however, can be used in conjunction with regression trees in order to overcome this issue. Boosting is a sequential evaluation of the relationship between the predictor variables and observations. Observations that are explained poorly are noted, the following trees in the sequence will then attempt to emphasize these observations in order to have a homogenous relationship between the predictor variable and the response(Elih et al 2008, De'ath 2007). The Gini impurity index is usually used to evaluate the accuracy of each tree(Daniya et al 2020). Each model is evaluated sequentially, thus the resulting tree is the best possible fit for the model. It is common to use a considerably high number of trees in order to verify the best combination of predictor variables.

3.1.3 Random Forest

Random Forests are a similar classification method to generalized boosted models. The method, like GBMs, uses both regression and classification trees to predict a true or false statement. It differs however by its random generation of trees. GBMs employ boosting which sequentially evaluate the previous tree into the subsequent tree, however Random Forests randomly generate a large number of trees with all possible combinations of predictors (Brieman 2001). It does this by repeatedly and randomly sampling the source data in order to build a full tree with all predictors, which are much more complex when compared to the trees built by GBMs. It then continues this process with a high amount of repetition. After the desired number of trees are built each tree then gets methodologically evaluated by the Gini index. The trees are then allowed to vote based upon the Gini index score to see which complete tree best predicts the model.

Random forest also excels in variable selection as it is the most used method to methodologically select variables as suitable predictors (Senay & Worner 2019 & Spieser et al 2019) Variable selection methods are highly desirable due to the high risk of overfitting and inaccuracy when using many variables and classification trees. Random forests additionally over performs other models when predicting outcomes outside the spatial and temporal bounds of the original data(Lawler et al 2006)

3.1.4 Ensemble Modeling: biomod2

The amount and complexity of species distribution modeling methods have been increasing ever since the advent of machine learning. Thus, a valid question is proposed; which method should be used & which provide the most accurate predictions? Several studies have tried to settle the debate on which methods are better than others, however the results are usually not universally interpretable. Several methods excel in solving overfitting, variable selection, and accuracy while none provide a universal solution(Elith & Burgmann 2002, Thuiller et al 2003, & Beaumont et al 2016).

Ensemble modeling provides a method to both compare and calculate the individual and shared performance of multiple species distribution algorithms at once(Thuiller et al 2009, Maiorano et al 2019). The application of biomod2 essentially compensates for any weaknesses an individual model can suffer from.

3.2 Climatic Models

3.2.1 WorldClim

Worldclim is a highly sophisticated climate model that seeks to provide an accurate visualization of several bioclimatic factors across Earth's surface(Hijmans et al 2005). The data collected seeks to represent the bioclim variables bio1-19 from 1950- 2000. This was done by collecting weather station data from across the globe and interpolating predictions of spatial areas that are not covered by weather stations. Interpolation however comes with a high amount of uncertainty thus cross-validations, similar to those in GBMs, of close by weather stations were used to ensure predictions are as accurate as possible. The resulting data is expressed in a series of raster grids with each pixel representing the bioclimatic value of the layer.

Worldclim pairs greatly and has been used in various species distribution models. Due to its global scale and easy accessibility it has the potential of showing the present distributions of many taxa with high degree of certainty(Akyol et al 2020, Coban et al 2020, & Bazzato et al 2021)

3.2.2 Occillayers

Occillayer is a recent development in paleoclimatology. This method seeks to provide an accurate, high resolution database for paleoclimatic variables from 5.3 million years ago to present(Gamisch 2019). This was done by using a global dataset of benthic sediments which house ancient oxygen isotope ratios and global climatic models of the Last Glacial Maximum. Oxygen isotope ratios have been used as a measure of global temperature proxy in the past(Marchitto et al 2014 & O'Brien et al 2018). Thus this has the potential to show an accurate representation of the climatic surface across a large time period.

There has been a backlash on the results of the Occillayer method, especially concerning its compared of other Pliocene General Circulation Models or GCMs. The most important critique is that Occillayers results differ greatly from the past GCMs like HADCM3(Brown et al 2020). According to this GCM there were large inconsistencies in its North Atlantic and Arctic climate estimates. HADCM3 and other models that participated in PlioMIP largely underestimate North Atlantic and Arctic temperature(Haywood et al 2020, Dowsett et al 2013). Occillayers however consistently provide warmer values for North Atlantic and Arctic than PlioMIP models. Thus we can assume that the Occillayer method is more suitable for paleoclimates of the European continent in which North Atlantic and Arctic warming will have a strong effect on vegetation.

3.2.3 Future Climate Models: HadCM3 & EC-Earth3-Veg-LR

Two GCMs were chosen to represent the future climate scenarios RCP 4.5 and RCP 8.0. The Hadley Center Couple Model or HadCM3 and EC-Earth3 or EC-Earth3-Veg-LR . Both models evaluate the effect of atmospheric, oceanic, and terrestrial variables on different climatic and geophysical forces. Interpolation and extrapolation of these results can aid in the projection for future and past time periods. Both models have contributed to the CMIP or Coupled Model Intercomparison Project, which seeks to compare, evaluate, and integrate several climatic projects(Eyring et al 2016). They both have also contributed and aided IPCC climate predictions for future climate scenarios(Johns et al 2003).

HadCM3 is especially popular with its use in species distribution models as it provides an easily accessible global dataset of bioclimatic factors biologically important for plant and animal species. When paired with previously mentioned SDM algorithms, HadCM3 provides high quality background data which is used to accurately predict habitat suitability(Atorre et al 2007, Adolmaali et al 2018).

EC-Earth3 is the only European Earth System Model thus it was deemed important and novel to use climate simulations from this dataset for a European centered study (Doscher et al 2021). Much like HadCM3 EC-Earth3 has also contributed to CMIP5 thus can provide reliable predictions on future climate conditions.

3.3 Pliocene Vegetation

The composition of vegetation of past time periods on Earth can be done by investigating the pollen and fossil makeup of layers dating back to the target time period. By investigating fossil layers we can estimate the potential species composition and equally important the distribution of certain time periods. The time period in focus will be the Pliocene as it is the closest time period that is comparable to modern day warming. Global investigations of Pliocene sediments have provided a coarse low resolution visualization of the vegetation zones during that time(Salzmann et al 2011). The coarse projection found that the Warm Temperate/Mediterranean zone is now the major phytoclimatic zone in Europe with it replacing the cool mixed forests of Central and Eastern

Europe. The cool mixed forests that currently dominate most of Western, Central, and Eastern Europe are projected to be in small patches in Central Europe and a small enclave in Southern Scandinavia. Lastly boreal forests that cover Scandinavia and Northeastern Europe are now only isolated on the northern coast of Norway. This visualization, however global, is very coarse, the local distribution of phytoclimatic regions is unknown. Future study over smaller regions may lead to more detailed and higher resolution visualizations of phytoclimatic regions of the Pliocene.

3.3.1 Temperate Tree Distribution in the Pliocene

Localized evaluations of Pliocene sediment can identify with high quality the species composition of the region. As mentioned before the temperate regions were found in isolated patches in Central Europe and Southern Scandinavia. After searching the literature this hypothesis seems to be largely true. This seems to be the case across the majority of mountain ranges and uplands in Central and Western Europe with additional occurrence in Eastern Europe

Macrofossils of ancestral populations of the genera *Fagus*, *Quercus*, *Acer*, *Tilia*, and *Abies* among many others were found in the southern slopes of the Alps, Sudeten Mountains, Rhine Massif, French Massif, and the Apennines(Teodoridis et al 2017, Kvacek & Teodoridis 2008, Ferguson & Knobloch 1998, Kvacek et al 2020). An important caveat is that the area of temperate species dominance is quite small and sub-tropical species are found on the same sites as temperate species. It can be assumed that temperate mesophytic plant species were suppressed across the majority of Europe. Dominance of temperate species can only be found on the slopes of mountains and uplands as more sclerophyllous plants are distributed on the lowlands(Bertini & Martinello 2011).

Occurrence of temperate flora in Southern Scandinavia has been largely unconfirmed mainly due to lack of microfossils, this may be explained by Pleistocene glaciation misplacing the sediment. Although pollen evidence of the the Northeast Atlantic has shown the presence of the genera *Quercus*, *Alnus*, and *Abies*(Willard 1994).

Eastern European occurrence was an unexpected find due to the coarse projection lacking any presence of temperate flora. The localities housing macrofossils were found in the Middle Dnieper Basin, Upper Don Basin, and Middle Volga basin. The species composition here largely represents a temperate woody grassland. The expected genera *Quercus*,*Alnus*, *Carpinus*, *Acer*, and

Tilia were present in moderate amount with the remaining community being composed of hydrophilous plants and grasses(Velichkevich & Zastawniak 2003)

3.3.2 Mediterranean Tree Distribution in the Pliocene

Mediterranean, sclerophyllous, or thermophilic plants are well represented in Pliocene sediment with both macrofossils and pollen. This may be due to their dominance throughout most of Europe, excluding Scandinavia. While temperate plants dominated the mid-altitudes thermophilous-type plants dominated the lowlands of mainland Europe. In all of the Western and Central European plots previously mentioned thermophilous-type plants were the dominant lifeform. Mediterranean genera such as *Juniperus-type*, *Cupressus-type*, *Buxus*, *Ilex*, *Quercus ilex-type* , and *Liquidambar* were present alongside temperate taxa. Now extinct thermophilous taxa were also present such as *Lauraceae, Carya*, *Tsuga*, *Thuja*, and *Gingko* (Teodoridis et al 2017, Kvacek & Teodoridis 2008, Ferguson & Knobloch 1998, Kvacek et al 2020). Mediterranean and thermophilous taxa formed a mosaic of subtropical /warm temperate forests across the majority of mainland Europe.

Southern Europe largely followed the same trend as Western and Central Europe. Expected Mediterranean genera are present such as *Juniperus*, *Cupressus*, *Ilex*, *Liquidambar*, and *Quercus Ilex-type* with Mediterranean endemics *Olea*, *Pistacia*, and *Platanus*(Vieira et al 2018, Jost et al 2009, Vieira et al 2011, Fauqette et al 1998, and Bertini & Martinetto 2011). It should also be noted that the genus *Cupressus* has maintained a stable presence within Southern Europe from the Cretacous to present(Rundel 2019). Thus we can consider the genus *Cupressus* as an important indicator of Mediterranean-type climate.

3.3.3 Boreal Tree Distribution in the Pliocene

Coniferous forests are by in large the most displaced phytoclimatic region during the Pliocene. The modern distribution of coniferous forests is found at high altitude boreal regions of Northern Europe above 60°N. During the Pliocene coniferous genera remains as the dominant lifeform on altitudinal forests with the representative genera being *Pinus*, *Picea*, *Betula,* and *Larix*. Temperate flora are subdominant across the altitudinal range being represented by *Fagus, Alnus* and *Quercus*(Teodoridis et al 2017 & Bertini 2001). This shows a similar distribution found today, although with an extended distribution of temperate genera.

Boreal coniferous forest macrofossils are certainly lacking in the literature, this may be due to Pleistocene glaciation displacing potential macrofossils. There is however evidence of occurrence from pollen records of sea sediment. Pollen records of the North Atlantic share a similar species composition of altitudinal forests. The major genera being *Pinus* and *Picea* with a sub-dominance of *Abies* and *Betula* and rare occurrence of *Alnus* and *Quercus*(Panitz 2016 & Willard 1994). Although not showing detailed local distributions of boreal forests it does show warmer than modern conditions. Temperate genera currently do not reach this far north, however with a lack of macrofossils certainly of occurrence is unknown. The detailed distribution of boreal coniferous forests of the Pliocene is currently not known, however this study has the potential of bridging that gap.

3.4 Pliocene as a Future Analog

The Mid-Pliocene Warm Period or mPWP has been shown to be a close analog for future warming scenarios RCP 4.5 and RCP 8.0(Burke et al 2018). Until 2100 the majority of Europe will have a Pliocene-like climate. Therefore this has the potential to show how the current species composition of Europe may react to a warming climate. Species distribution models using extrapolated future climate data have been made to predict the future distribution of plants. Shifts in suitable habitat follow a trend which is highly similar to that of the Pliocene. The highest correlation is the spreading of Mediterranean-type plants northward and the shrinking of suitable areas for Temperate-type and Boreal-type plants(Ohlemuller et al 2006). Mediterranean plants are simulated to be suitable across the Western Europe while boreal coniferous trees are finding above the tree line in the Norwegian mountains(Keenan et al 2011 & Han et al 2021). The same can be said for altitudinal forests as well, as lowland plants dominate, temperate plants are pushed upward onto slopes of mountains and uplands. A study observing the species composition change found that boreal and temperate mountains have increased species richness due to upward migration of lowland plants(Pauli et al 2012). The same study also observed the local extinction of summit plants in Southern Europe which would further threaten altitudinal forests.

4. Methodology

4.1 Species selection

Tree species were selected according to their prevalence, representativeness, and presence of an ancestral species in the Pliocene. This is done under the assumption that ancestral forms of modern species have largely conserved their niche. This can be supported by modern species during the Pleistocene having similar ecological breath to present time(Meyer & Peterson 2006). Species that are present in the Pliocene are also useful to validate the models that will be created for the focus time period.

Mediterranean species were selected according to their geographic distribution and their prevalence in locations with mild humid winters and hot dry summers. These species were chosen due to their circum-Mediterranean distribution and their similar ecological characteristics of being thermophilous, light demanding, and drought tolerant. The species that were chosen were *Quercus ilex, Quercus pubescens, Pinus halepensis, Pinus brutia, Pinus pinea, Cupressus sempervirens,* and *Juniperus oxycedrus.*

Temperate species were similarly selected due to their geographic distribution in mainland Europe up to between 45-60°N and their mesophilic characteristics. A key characteristic that was accounted for during species selection is the species prevalence in an area with low precipitation seasonality and the presence of cold winters. The species chosen were *Fagus sylvatica, Quercus robur, Quercus petraea, Acer pseudoplatanus, Carpinus betulus,* and *Tilia cordata.* These species, especially *F. sylvatica Q. robur* are principal components of broadleaf and mixed forests throughout all of Europe.

Boreal-continental trees are characterized by their northward distribution, Northern Scandinavia, and long cold winters with short warm growing periods. Boreal winters are long and have frequent frosts in spring, thus species are selected for frost tolerance and hardiness. The species chosen were *Picea abies* and *Pinus sylvestris.* It must be noted that the aforementioned species are azonal and often occur in the lowlands and altitudinal forests of the Temperate and Mediterranean zone. Dominance of conifers and frost hardy trees however, do become more apparent in Northern and Eastern Europe. In Northern Europe especially coniferous species create pure stands with large geographic distribution.

4.2 Tree occurrence data

Presence data for the selected tree species were collected from two existing databases EU-Forest and EUFORGEN. The first being EU-Forest which is an ensemble of three previous European tree datasets: Forest Focus Regulation, Biosoil Project, and National Forest Inventories(NFI). EU-Forest seeks to provide accurate and precise, 1x1 km, tree occurrence for 242 woody species within the European Union(Mauri et al 2017). This dataset however lacks high density records in Poland, Belarus, Slovenia, Bulgaria, and Greece. EU-forest also completely lacks records for the Balkan countries, Turkey, and Russia. This lack of occurrence data is supplemented by the second dataset, EUFORGEN. EUFORGEN provides the general contiguous distributions of 190 trees and woody plants throughout Europe, including Russia to the Ural mountains and North Africa(Caudullo et al 2017). This fills the gaps of information left by EU-Forest and presents a general overview of species occurrence. This dataset however includes native and naturalized populations, thus only native distributions were used in the final models. These two datasets were then combined and used to estimate tree occurrence of the selected tree species.

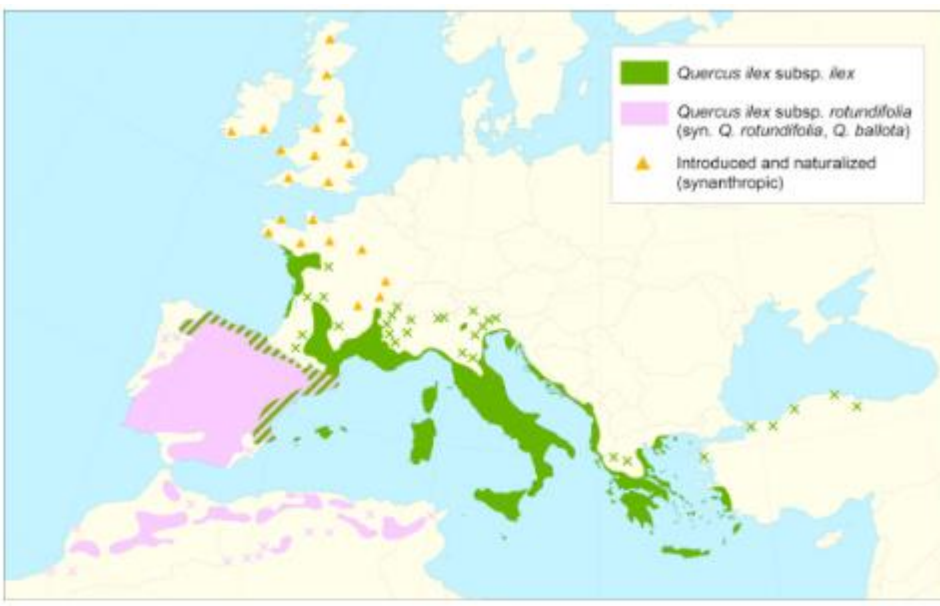

Figure 1. Example of *Quercus ilex* with the native and naturalized distribution(Caudullo et al 2017)

4.3 Bioclim variable selection

Bioclimatic variables were extracted from WorldClim 2, an interpolated dataset which assesses 19 different bioclimatic variables concerning global temperature and precipitation from 1970-2000(Fick & Hijmans 2017). This however does not account for additional warming from the past two decades. WorldClim 2 has proved to be a reliable dataset that has been used in species distribution modeling for plants and animals.

Bioclimatic variable selection is commonly species specific. Machine learning systems, like randomForest, are used in assessing variable importance; although randomForest and other machine learning systems are limited to finding variable importance in single species. This is reflected in species specific ecological needs, although cannot be applied to multi-species models. A standard selection of variables that take into account a broad range of ecological characteristics across the Mediterranean, Temperate, and Boreal zones must be used. Thus the selected variables are bio1, bio4, bio5, bio6, bio15, and bio16. This group of variables have proven to be reliable predictors for species in the Holarctic(Petitpierre et al 2017)).

Bio1 is the annual average temperature from the historical time period 1970-2000. Bio4 represents the difference in temperature throughout the year or temperature seasonality. Bio5 and Bio6 are the mean maximum temperature of the warmest month and mean minimum temperature of the coldest month respectively. Bio15 is the change in precipitation throughout the year, or precipitation seasonality. Lastly, bio16 is the precipitation of the wettest quarter.

4.4 Climatic Variables for Pliocene and Future Climate

Pliocene climate will be taken from Oscillayers, a dataset representing oscillations of climate from the last 5.4 million years in 10,000 year intervals(Gamish 2019). Thus climatic data was averaged from 3.25-3.0 million years ago to estimate the climate of the mid-Pliocene Warm Period. For future climate scenarios two datasets were used: HadGEM3-GC31-LL and EC-Earth3-Veg(Roberts 2017, Döscher et al 2021). Both datasets were used in CMIP6 which is a reliable predictor of future climate that is used by the IPCC. Each dataset is in 2.5 arc minute resolution.

4.5 Model building

An ensemble model using biomod2 will be utilized for the creation of mid-Pliocene Warm Period, Present, RCP4.5, and RCP8.5 habitat suitability maps for each species. Biomod2 provides a framework in which multiple species distribution models can work individually and collectively to predict species suitability(Thuiller et al 2003). Biomod2 has various model options although the chosen methods are randomForest, Generalized Boosted Models, and MaxEnt. MaxEnt was chosen for its strength with its default settings and its flexibility with presence-only data((Phillips et al 2006). This is useful since our occurrence data lacks absence data. Generalized Boosted Models uses sequential evaluation of variables and results in models with regression trees that have high predictive power(Friedman et al 2000,Elith et al 2008). Finally randomForest is used for a similar purpose of GBMs, although regression tree creation is random, thus reducing any form of bias in the model creation(Senay & Worner 2019, Speiser et al 2019).

Occurrence data will be coming in the form of point coordinates from the aforementioned datasets and bioclimatic variables will be coming from the WorldClim 2, Plio-Occilayers, HadCM3 and EC-Earth3 datasets. These will represent the temperature and precipitation regimes for present, Pliocene, and future scenarios respectively. Pliocene bioclimatic variables will be averaged from 3.25 million years ago to 3.0 million years ago. After both future climate datasets, HadCM3 and EC-Earth3, will be averaged and will be created into an ensemble dataset. Then the new ensemble dataset will then be separated into two time periods 2040-2060 and 2080-2100. This is done to show the effects of mid-century to late-century warming of both RCP4.5 and RCP8.5 scenarios.

Each modeling technique will be run ten times thus 30 model runs for each species grouping. 100 trees were set as the maximum with 10 cross-validations for GBM and RandomForest regression trees. 70% of the data is used for training while the remaining 30% will be used for model building. Final models will be made for each of the three species groups for each climate scenario.

4.6 Data model comparison & Model Evaluation

To validate the models of the mid-Pliocene Warm Period, macrofossil and pollen data will be reviewed from across Europe. As mentioned before, in species selection, tree taxa were chosen on the basis of an ancestral form being present during the Pliocene. A body of scientific literature of macrofossils and pollen was reviewed to identify similar tree species present today. This was then used to show definitive presence within a geographic area for a certain species. For example, ancestral *Fagus, Picea* or *Quercus* populations would likely be distributed differently in the Pliocene, thus macrofossil and pollen data would aid in visualizing the difference in distribution when compared to modern forms of these taxa. Presence of species in the Pliocene will then be compared to the suitability maps created for the Pliocene.

True Skill Statistic(TSS) and Receiver Operating Characteristic(ROC) were used to evaluate the predictive performance of the resulting models. TSS ranges from -1 to 1 with scores above 0.5 indicating suitable performance(Allouche et al 2006). Furthermore, ROC ranges from 0 to 1 with values above 0.7 representing acceptable model performance(Komac et al 2016, Bradley 1997). Thus only model runs with scores equal to or above the aforementioned values were used in the final estimation of suitable habitat.

5. Results

5.1 Model Performance Evaluation

Evaluation scores differed between species groups and modeling methods although all final models resulted in evaluation scores above our recommended thresholds. The Mediterranean species group had the highest evaluation metrics with an average ROC of 0.9194 and an average TSS of 0.7709. This is followed by the Boreal species grouping with an average ROC of 0.8469 and average TSS of 0.5724. The Temperate species grouping had the lowest evaluation metrics with an average ROC of 0.8302 and an average TSS of 0.5441. All model outputs exceeded our recommended thresholds thus all species grouping models were used in final model building.

Evaluation metrics were also compared between the modeling methods RandomForest, MaxEnt, and Generalized Boosted Models. RandomForest and MaxEnt both consistently resulted in evaluation metrics above our thresholds while GBM averaged metrics close to our thresholds. RandomForest had an average ROC value of 0.9085 and an average TSS of 0.6832. This is followed by MaxEnt with an average ROC of 0.8930 and an average TSS of 0.6359. GBM resulted in the lowest average in both ROC and TSS with the values being 0.7950 and 0.5683 respectfully.

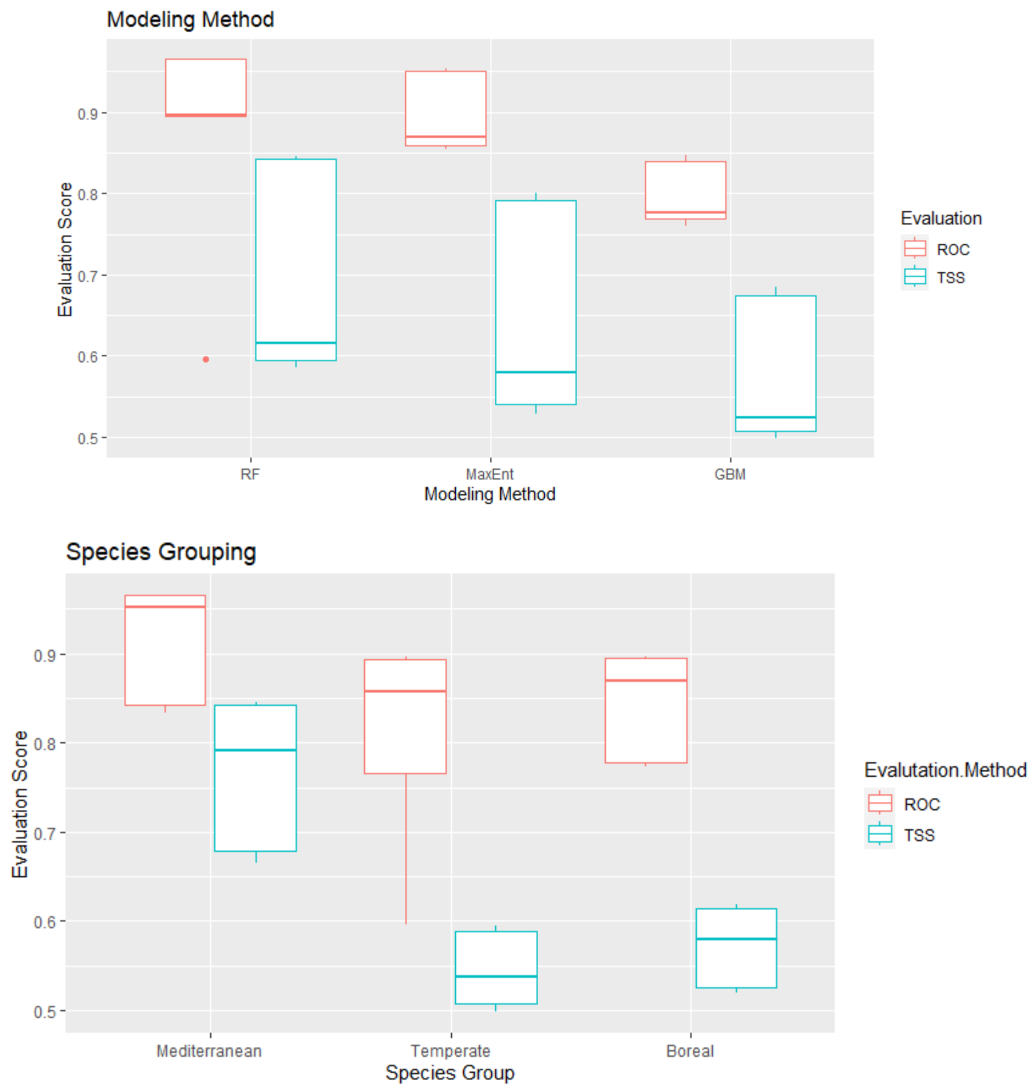

Figure 2. Evaluation metric scores between modeling method(above) and species grouping(below)

5.2 Habitat Suitability Maps

Species distribution maps were created for each species group for modern, Mid-Pliocene, 2050, and 2100 using RCP 4.5 and RCP 8.5 for the future scenarios. The resulting models visualized the probability of occurrence with a range of of 0-1000. Five suitability classifications will be created by separating suitability values into five distinct classes of equal interval. 0-200 will represent very low or no suitability, 201-400 will represent low suitability, 401-600 will represent moderate suitability, 601-800 will present high suitability, and finally 801-1000 will present very high or perfect suitability.

5.2.1 Habitat Suitability for Present Climate

The Mediterranean species group is largely contained to the Mediterranean Basin. With the core areas of high suitability in the Maghreb, Iberian, Italian, and Balkan peninsulas. Areas of southwestern France, specifically the historical area of Aquitaine also have highly suitable habitats. Suitable habitat wanes in the eastern Mediterranean with areas of the Middle East with low to no suitability. Moderately suitable habitats can be found further inland especially in western France and the inland areas of the Balkans.

The Temperate species group has a much far-ranging distribution. High suitability areas occupy the inland areas of mainland Europe, altitudinal forests of Southern Europe, and the Caucasus Mountain range. Moderately suitable areas dominate most of Central and Eastern Europe. The core suitable areas are found in mainland Central and Western Europe with enclaves of high to moderately suitable habitat in southern Sweden. Suitable habitat sharply declines in Northern and Eastern Europe with much of the area being dominated by habitats of low suitability. There is however small pockets of moderate suitability spread across Eastern Europe. The present northern limit of Temperate species according to our models is estimated at 60° N which is expected.

The Boreal species grouping also have a large geographic distribution. Core areas of high suitability are separated into two clusters. The first being in Central Europe, especially the countries of Germany, Austria, the Czech Republic, and regions of the Alps and Carpathians. The second cluster dominates the Fennoscandian peninsula below 66° N. The majority of Sweden, Finland, Norway, and Baltics are dominated by highly suitable habitats. Notable areas of high suitability not mentioned in the clusters are the mountain ranges of Southern Europe such as the Pyrenees, Apennines, and the Dinaric Alps. Habitat suitability decreases in Northeastern Europe with the Kola peninsula and Northern Russia being dominated by areas of moderate to low suitability. Suitability also declines in the Central regions of Russia with a modest boundary separating the Temperate and Boreal zones approximately at 55 °- 60 °N.

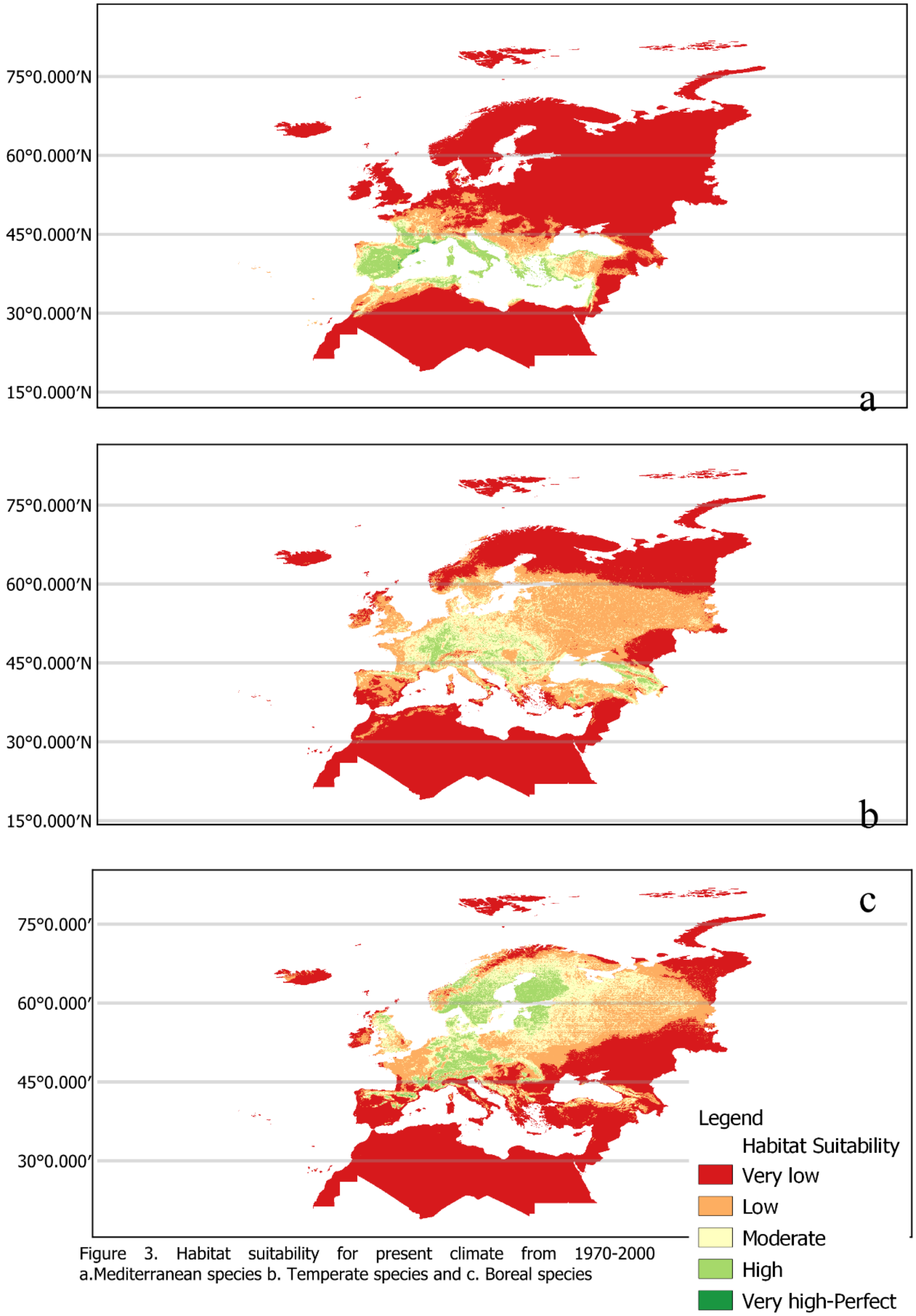

Figure 3. Habitat suitability for present climate from 1970-2000 a.Mediterranean species b. Temperate species and c. Boreal species

5.2.2 Habitat Suitability for Pliocene Climate

The distribution of suitable habitat for the Mediterranean species grouping has drastically changed. The core areas in the Mediterranean Basin are largely the same with minor decreases in suitability in the Maghreb, Iberia, Greece, and the Balkans. Clusters of high suitability are now found in Western and Central Europe. A large grouping of highly suitable habitats can be found in Germany and western Poland. Moderately suitable areas dominate Central and Western Europe. Expansion of suitable areas can now be found in the inland areas of the Balkans and Southern Sweden. Lastly there are low suitable areas stretching as a band across northern Russia and in Scandinavia.

The Temperate species grouping have also drastically shifted their distribution. The former core areas of Western and Central Europe have largely fragmented. With the majority of high suitability being replaced by moderately to low suitable areas. Moderately suitable habitat still remains in Germany and Czech Republic. Few habitats of high suitability remain with the largest being the highly elevated areas near the western and eastern Alps. Two other large suitable areas can be found in the Pyrenees and the inland highlands of the Balkans. The majority of the moderate to high suitable areas shifted towards the north and east. There is a band of high suitability across the Pontic Steppe and in the Central Russian Upland. Similar moderate to high suitable habitats can be found in highlands of Scandinavia. Low suitable areas are ubiquitous across the majority of Europe with areas of no suitability being found in Southern Sweden and the Baltic Sea coast of Poland.

The distribution of Boreal species in the Pliocene has drastically decreased. The first core cluster in Central Europe has essentially disappeared with some pockets of moderate suitability being found in higher elevations. The remaining moderately suitable areas are found in the mountains surrounding the Czech Republic, German Highlands, and the Carpathians. Large areas of high suitability can still be found in the Alps, Pyrenees, Caucasus, and Dinaric Alps although in much higher elevations when compared to the modern distribution. Suitable areas in Sweden, Finland, and the Baltics have completely disappeared. The largest area of moderate to high suitable habitat can be found in the mountainous areas of Norway. This band continued northward and ultimately ended at the Kola Peninsula.

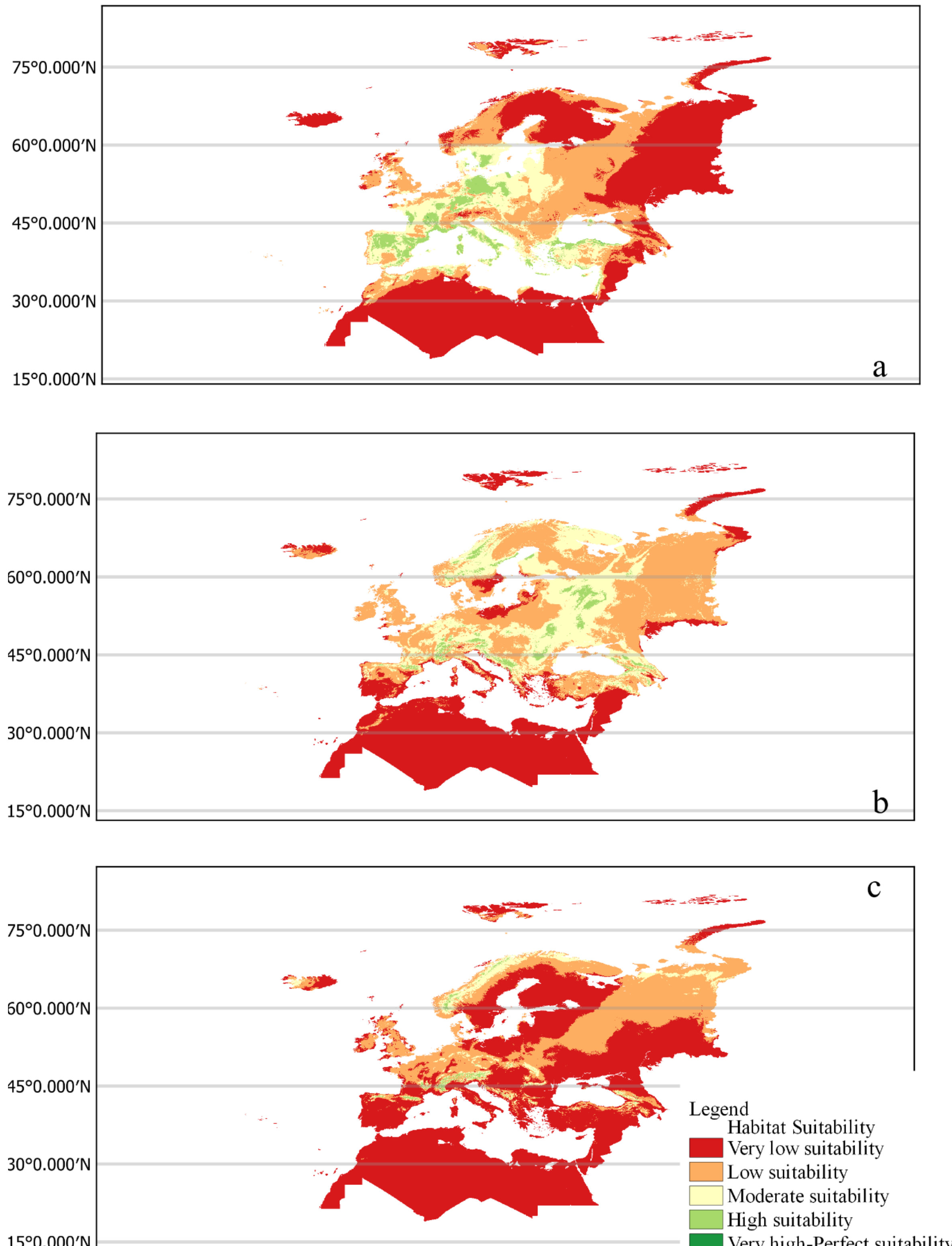

Figure 4. Habitat suitability of the mid-Pliocene Warm Period a. Mediterrranean species b. Temperate species and c. Boreal species

5.2.3 Habitat Suitability for RCP 4.5 2040-2060

Mediterranean species have shifted greatly in the moderate greenhouse emission scenario. The core suitable areas in the Mediterranean Basin are still largely present, however there are some declines in the Maghreb and Balkan Peninsula. A novel area of high suitability can be seen dominating Western Europe and Southern England. Suitability decreases eastward with low suitability areas ending in a similar distribution as the Temperate grouping. Low suitable areas can also be seen in Southern Sweden, the Baltics, and the Norwegian coast.

The Temperate grouping has a differing response to future warming than that of the Pliocene. Patterns of suitability are largely similar in mainland Europe. Central Europe is a mixture of moderate to high suitable habitat with large areas of high suitable habitat being found in Southern Czech Republic, German Highlands, Western Alps, and the Carpathians. Novel areas of moderate to high suitability can be found in Fennoscandia and the Baltics with some large pockets of moderate suitability reaching the Central Russian Highland. The northern limit of suitability can be found approximately at 70° N.

The Boreal species grouping has a similar response to future warming as Mid-Pliocene warming. The majority of Central European clusters of high suitable habitat has been lost with small refugia being found in Alps, Carpathians, Pyrenees, and the mountains surrounding the Czech Republic. The core cluster of high suitable areas in Fennoscandia has largely diminished with new areas of high suitability being found along the Norwegian Coast. A novel area of high suitability can be found in the Kola Peninsula and areas of northern Russia.

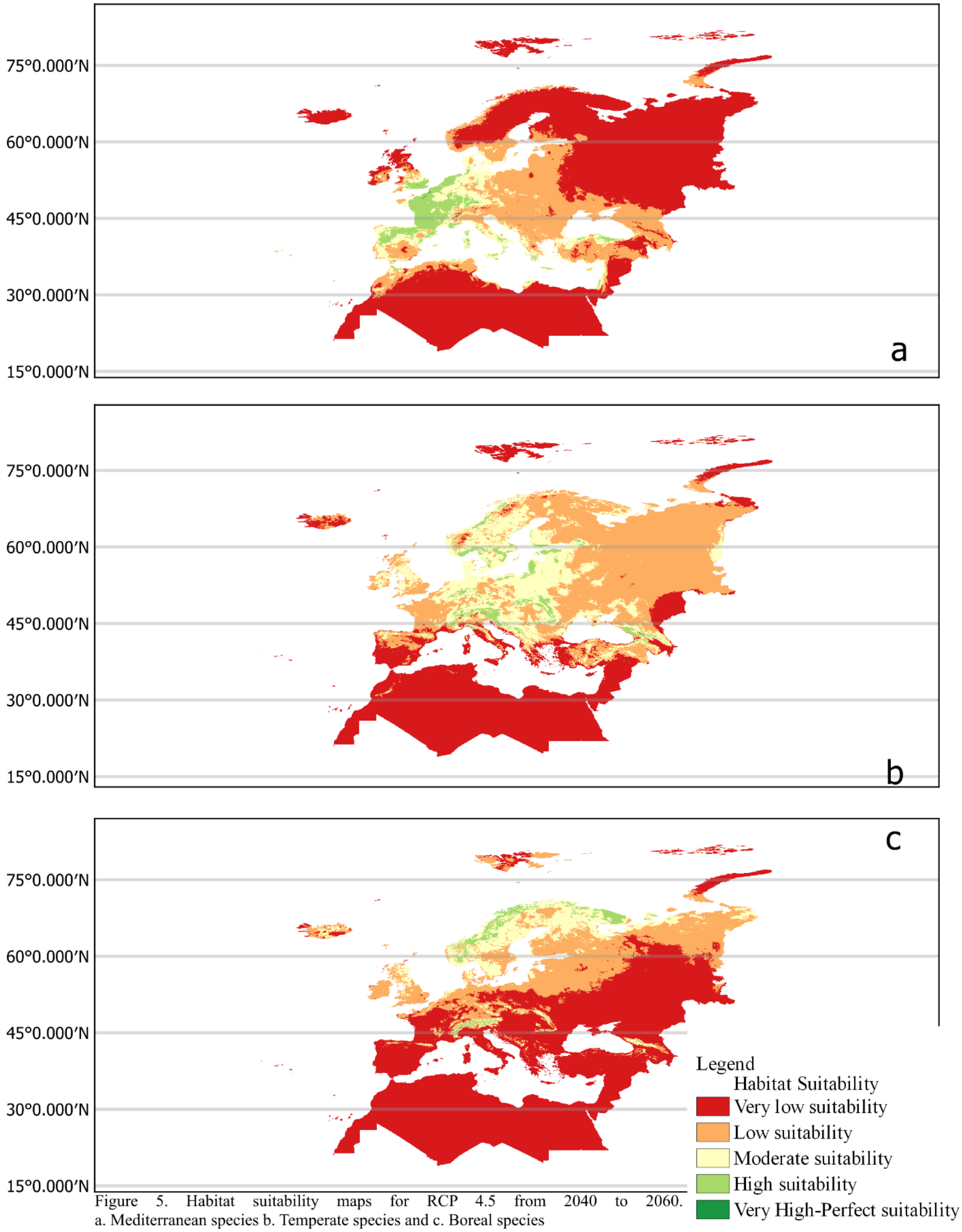

Figure 5. Habitat suitability maps for RCP 4.5 from 2040 to 2060. a. Mediterranean species b. Temperate species and c. Boreal species

5.2.4 Habitat Suitability for RCP 4.5 2080-2100

The decline of suitable areas continues to decline in the core areas in the Mediterranean Basin. Noticeably declines can be found in the Greek and Italian peninsulas. New suitable areas however have developed in the inland areas of the Balkans. The Western European cluster developed in 2050 has continued expanding north and eastward with highly suitable areas being as far north as Northern England and Southern Sweden. The eastern expansion is dominated by moderate suitable habitats with small pockets of high suitability in the Carpathian Mountains. There have also been noticeable increases in suitable areas in the Black Sea coast of Turkey.

Moderate to highly suitable habitat continues to wane for Temperate species in Central Europe. The cluster of moderate to highly suitable habitat has dramatically decreased in size and the only remaining refugia of highly suitable habitat can be found in the Ore Mountains, German Highlands, and the Western/Eastern Carpathians. The complex of suitable habitat along the Baltic Coast of Poland and the Batlics have strongly decreased with moderately suitable habitat still remaining in small inland pockets. The majority of the moderately suitable habitat has shifted further eastward from the Batlics towards the Central Russian Upland. The core habitat area in Fennoscandia has gradually increased with highly suitable habitat being found far inland in both Sweden and Finland. The northern regions of Norway and the Kola Peninsula are now dominated by moderately suitable habitat.

The complex of suitable habitat in the mountainous regions of Central Europe has all been lost for the Boreal species grouping. Only very small pockets of moderately suitable habitat remain in the Western Carpathians, and the mountains surrounding the Czech Republic. The largest area of suitable habitat is still present in the Alps although suitability has diminished in the lower elevated areas of the mountain range. The remaining moderately suitable habitat in Sweden and Finland have largely disappeared. A similar decline can be seen in the mountain ranges of Norway as well. The novel area of suitability in the Kola Peninsula has also disappeared.

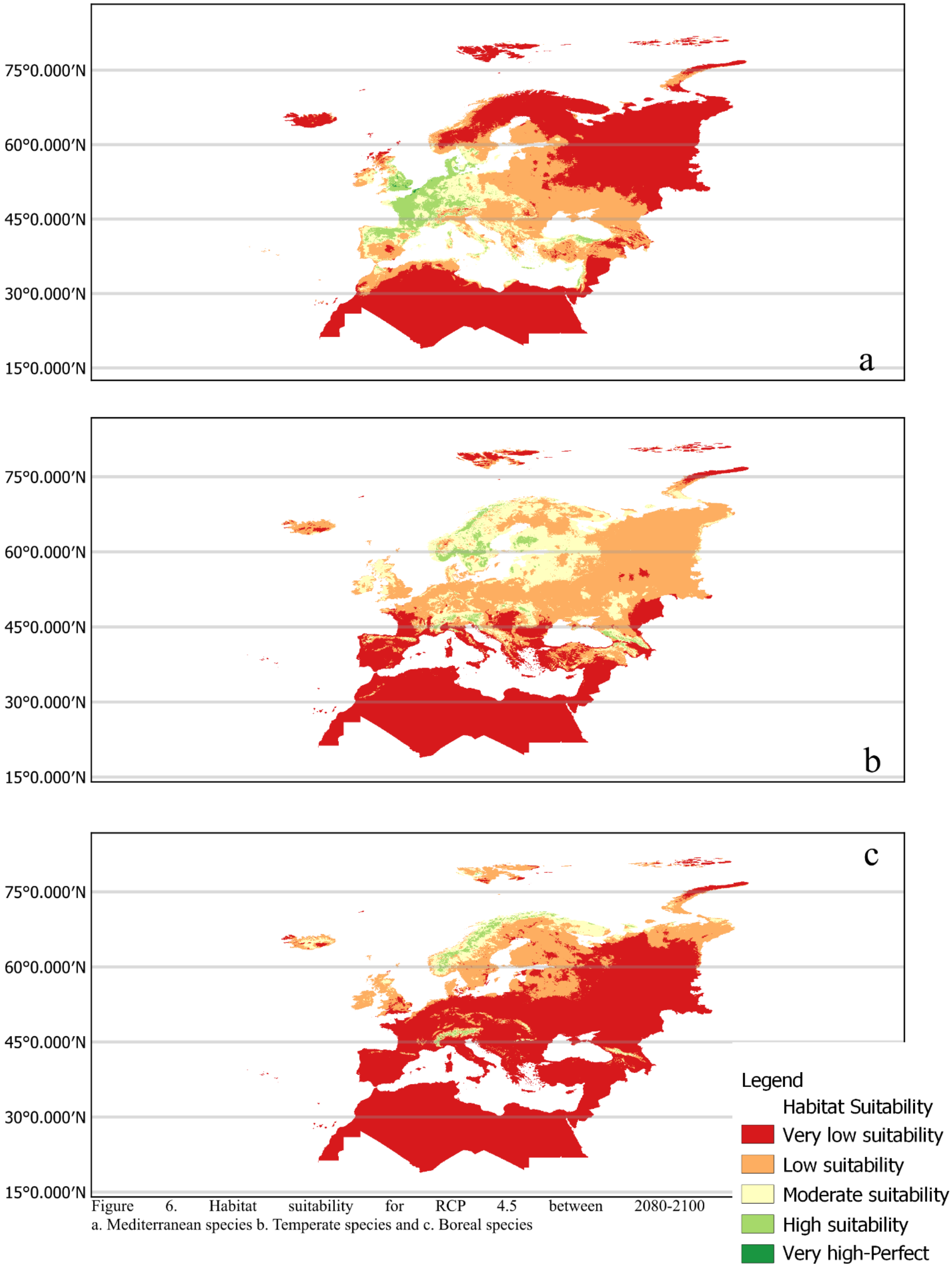

Figure 6. Habitat suitability for RCP 4.5 between 2080-2100
a. Mediterranean species b. Temperate species and c. Boreal species

5.2.5 Habitat Suitability for RCP 8.5 2040-2060

The Mediterranean species grouping has a noticeable accelerated shift in distribution. Much like the previous climate scenario the core areas in the Mediterranean Basin are mostly still intact. There is however a slight decrease in highly suitable habitat in the Italian Peninsula, Northern Portugal, and the Peloponnese. This is coupled with an increase in moderately suitable habitat in Western Europe and the Balkans, largely mimicking the distribution during 2100 of the RCP 4.5 scenario. There is a noticeable increase in moderately suitable habitat in Eastern Germany extending into Poland and lowland areas of the Carpathians. There is also marked increase of suitability in southern Sweden mostly in close proximity to the coast.

The distribution of suitable habitat for Temperate species largely follows the same patterns of distribution as the previous RCP 4.5 scenario albeit with a reduced amount of highly suitable habitat in Central Europe. Areas such as Southern Bohemia, German Highlands, and Eastern Alps have reduced coverage if moderately suitable habitat. The core distribution of moderately suitable habitat has shifted further east with large areas of high suitability surrounding the Gulf of Finland. This continues into southern Finland with a similar distribution in Sweden and Norway.

Much like the Temperate species grouping, the spatial pattern of Boreal species for RCP 8.5 is very similar to the lower emissions scenario RCP 4.5. The Alps, Carpthains, and Caucasus mountains remain as refugia with high to moderate habitat suitability dominating these regions. The core area of suitable habitat remains in Norway following an Atlantic distribution until reaching the Kola peninsula. Patches of moderately suitable habitat stretch towards the northern foothills of the Urals in the east.

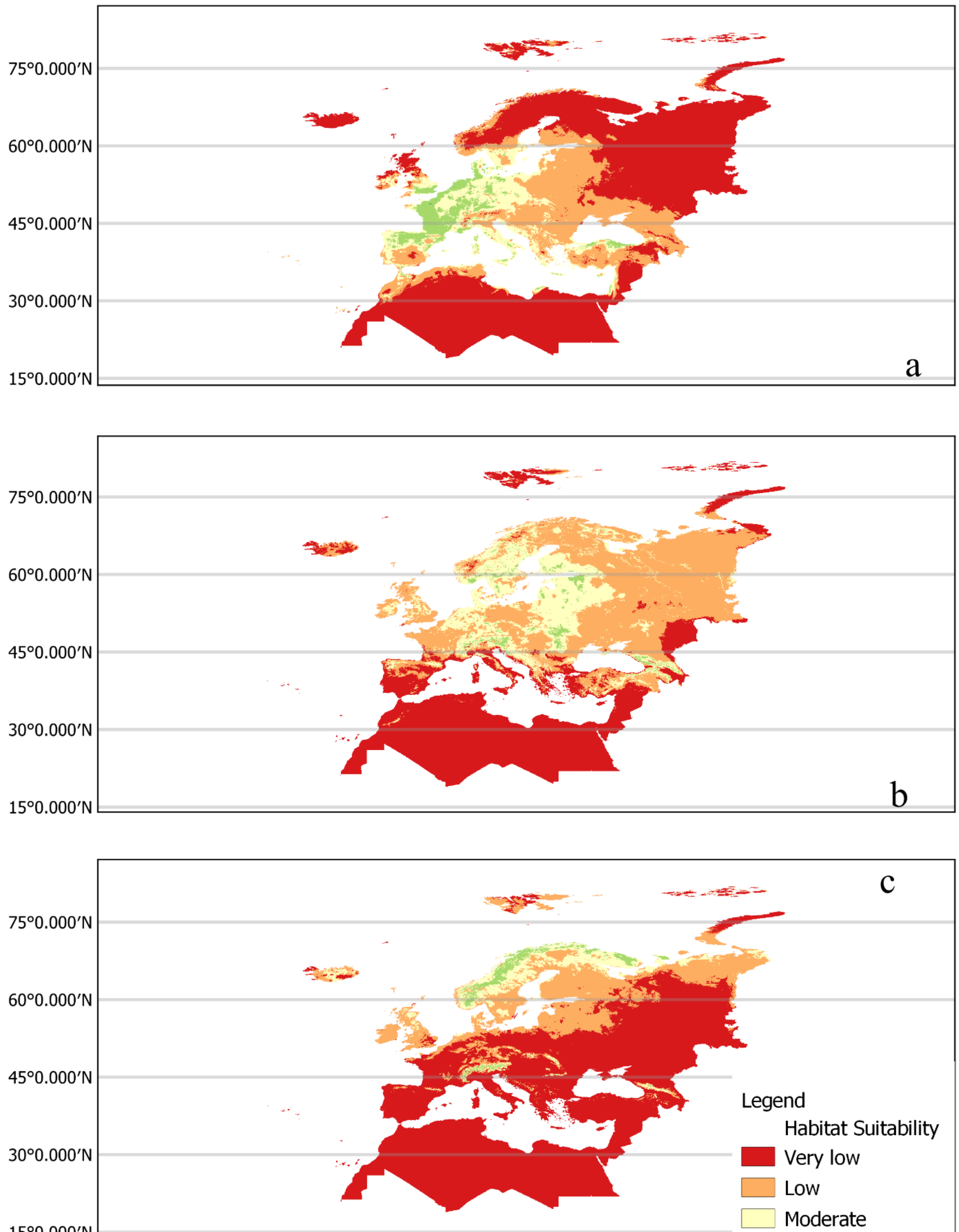

Figure 7. Habitat suitability for RCP 8.5 between 2040-2060 a. Mediterranean species b. Temperate species and c. Boreal species

5.2.6 Habitat Suitability for RCP 8.5 2080-2100

The suitability of Mediterranean species have drastically decreased in its core habitat of the Mediterranean Basin. Small pockets of high to moderately suitable habitat remain in Portugal, the Pyrenees, and a few large islands such as Corsica, Sardinia, Sicily, Crete, and Cyprus. The majority of the Iberian, Italian, and Balkan peninsulas have been reduced to low suitable habitat. Furthermore the former suitable areas of Western Europe have largely diminished. The remaining areas of high to moderate suitability are found in a strong Atlantic distribution. The core areas of suitable habitat are now found across the coasts of France, Belgium, the Netherlands, and Denmark. Furthermore there are novel suitable areas that dominate the British Isles. The Atlantic distribution continues towards the coast of Norway and Sweden then extends as far north as 70° N.

The suitability of Temperate species continues to shift northward with the core area in Central Europe essentially disappearing. Areas of moderately suitable habitat can still be found in Western Czech Republic, German Highlands, and the Carpathians although greatly diminished. The Alps, Pyrenees, and Caucasus mountains are still stable refugia of Temperate species albeit with a reduced area. The largest areas of high to moderately suitable habitat can be found throughout Fennoscandia and the Kola peninsula. This area of suitability extends eastward along the Arctic coast before terminating at Novaya Zemlya in northern Russia.

The Boreal species grouping has drastically diminished its suitable area in Central Europe. All suitable habitat, spare the Alps, have been eliminated in Central Europe. Refugia in the Pyrenees and Caucasus mountains have also been eliminated. The core area of suitable habitat in Norway and the Kola Peninsula remains with noticeable decreases of habitat.

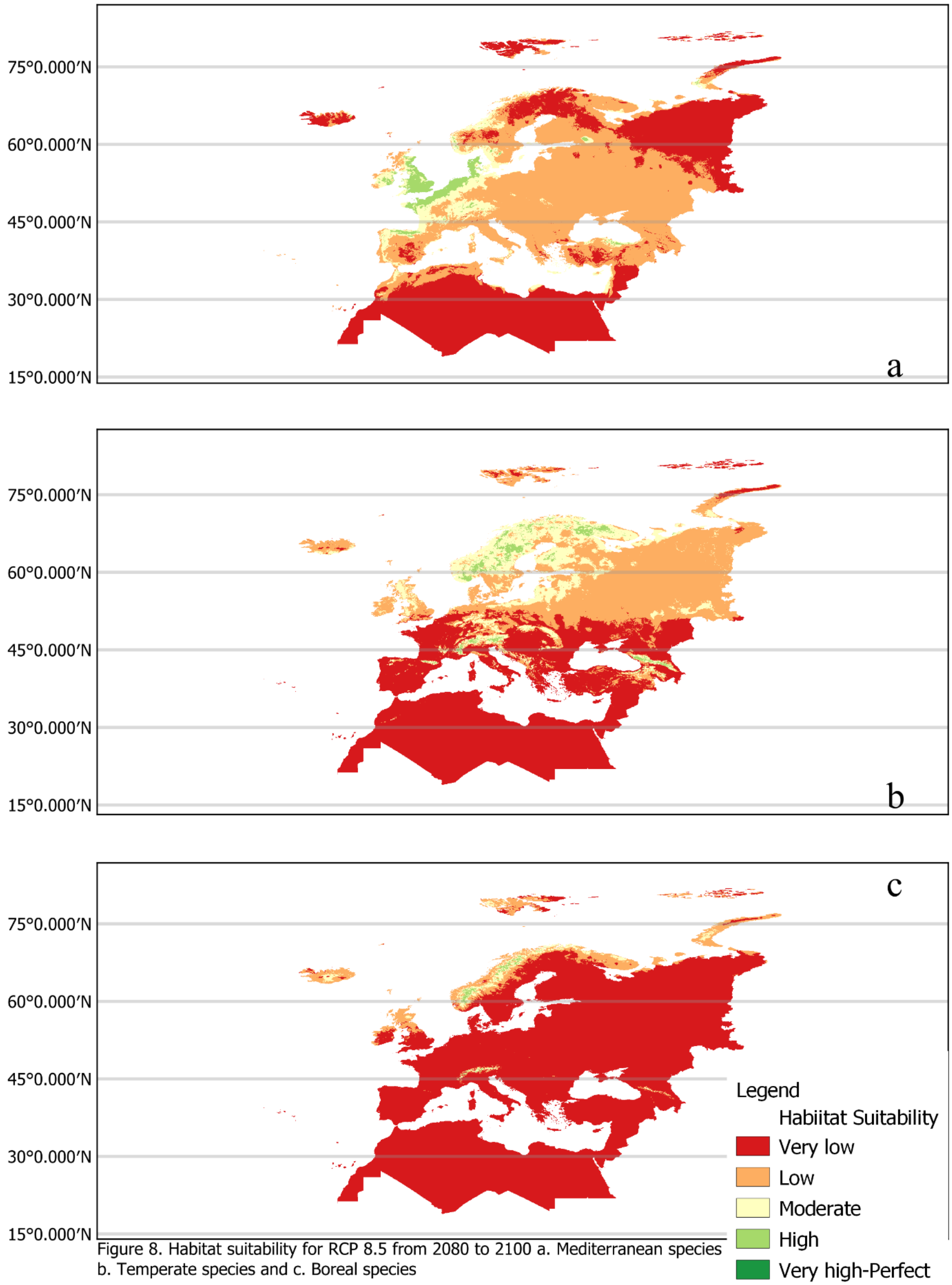

Figure 8. Habitat suitability for RCP 8.5 from 2080 to 2100 a. Mediterranean species b. Temperate species and c. Boreal species

6. Discussion

6.1 Vegetation of the mid-Pliocene

Environmental conditions in Southern Europe during the mid-Pliocene were likely similar to present conditions with the majority of Mediterranean genera showing potenial for existence throughout the Tertiary and into the Quaternary(Rundel 2019). Modern drought-tolerant conifers and shrubs would likely dominate the lowlands with evergreen broadleaf species residing on altitudinal forests. Furthermore, now extinct sub-tropical genera such as *Pseudotsuga*, *Sciadopitys*, *Carya*, *Taxodium* and many others were codominant with Mediterranean flora according to pollen and macrofossil data(Vieira et al 2018, Jost et al 2009, Vieira et al 2011, Fauqette et al 1998, and Bertini & Martinetto 2011). Our final models support this hypothesis by showing core areas of highly suitable habitat for modern Mediterranean tree species in the Mediterranean Basin under mid-Pliocene climate.

The Mediterranean subtropical ecosystem also dominated the modern temperate region of Europe being fairly ubiquitous throughout all pollen and macrofossil studies in Central Europe. This is further supported by our final models showing the Mediterranean species groups largely mimicking the general distribution of modern temperate species. Western and Central Europe likely followed an altitudinal gradient of subtropical and Mediterranean genera dominating lowlands area while temperate broadleaf and boreal coniferous genera are found on upland refugia. Temperate broadleaf forests were likely pushed towards the foothills of mountains representing only a fraction of the fossils in Pliocene deposits. This likely indicates that temperate species were quite rare in mainland Europe during the Pliocene. This can likely be due to increased precipitation and temperature aiding the dominance of Mediterranean species in Central Europe.

Temperate broadleaf forests were rarely dominant in mainland Europe, however there are sediments that show core areas of habitation in Eastern Europe especially the Volga River Basin and Central Russian Upland (Velichkevich & Zastawniak 2003). Temperate genera such as *Fagus Quercus* and *Acer* show high probability of presence likely on the hills above rivers and the uplands of Central Russia. This is coupled with the presence of mesophilic wetland species such as *Alnus* and *Ulmus*. This can indicate a species composition very similar to modern assemblages. The dominant presence of Temperate mesophilic species this far east can indicate shift of mesophilic conditions into continental areas. This is supported by the increase in summer temperatures in Eastern Europe. Analyzing bio5 for the mid-Pliocene shows a strong band of warmth throughout Eastern Europe and Russia this likely led to the suitability of Temperate species so far East. It is unknown why this pattern is observed thus further research into the paleoclimatic patterns of continental Europe must be performed.

Northern Europe, especially the Norwegian mountains represent the largest suitable area for Temperate and Boreal species. Pollen analysis of sea sediment near Norway indicates a cool-mixed forest throughout Scandinavia (Panitz 2016 & Willard 1994). Our final models support the coexistence of Temperate and Boreal species in the Atlantic and mountainous regions of Norway as far north as 70°N. These ecosystems are likely very similar to modern forests of Central Europe with broadleaf trees making up the lowland and coniferous trees dominating higher elevations. This ecosystem however stretches far into the Arctic Circle indicating much warmer temperatures than today. Suitable habitat for both Temperate and Boreal species dominant the Arctic regions where tundra is found today. Arctic tundra was likely very rare or has no presence in Northern Europe.

Thus, Mediterranean species have a largely Western to Central European distribution largely replacing the modern location of Temperate species. Temperate species find refugia in high elevations of Central Europe. Novel areas of high suitability can be found further eastward in the Volga River Basin. Lastly Temperate and Boreal species likely created cool-mixed forests across Norway and the Arctic Coasts of Fennoscandia. This indicates a world with drastically shifted vegetation due to increased temperature and humidity.

6.2 Vegetation of RCP 4.5 2040-2100

RCP 4.5 represents a climate scenario that is analogous to Pliocene warming thus it is expected that European vegetation will respond in a similar way, however this is not always the case. Mediterranean species have a largely different response to future warming when compared to Pliocene warming. In the Pliocene Mediterranean species dominated in large to small pockets spread throughout Western and Central Europe. In the RCP 4.5 climate scenario Mediterranean suitability shows a clear Atlantic distribution. As the scenario progresses the distribution of habitat remains the same with slight increases to the East. This can mean a weakening of the Gulf Stream which is an important source of precipitation in winter for Europe(Patler 2015). A weakened Gulf

Stream can lead to overall drying conditions in Europe, especially regions in which precipitation is regular throughout the year. This can be advantageous for drought-tolerant species like those found in the Mediterranean.

Temperate species show a similar response to future warming when compared to Pliocene warming. Refugia are found in the same locations in Central Europe. This suggests that Temperate forests would find their new optimum environment in altitudinal forests of Central Europe, Carpathians, Caucasus, and the Alps. The similarities continue with the largest area of suitable habitat being found in the same general distribution in Scandinavia. The similarities end there however, with the large area of suitable habitat near the Volga River Basin. This is likely due to the warming being more evenly distributed throughout the continent in future climate scenarios. Nevertheless the distribution of Temperate species in their former core area has drastically shifted.

The Boreal species group has the most similar and analogous response towards future warming. The distribution of suitable habitat largely, if not entirely, follows the same pattern present in the mid-Pliocene. Thus it can be strongly assumed that cool mixed forest of broadleaves and conifers dominate Scandavia as far north as 66-70°N. This may indicate a world as warm as the mid-Pliocene as early as 2040 in Northern Europe.

6.3 Vegetation of RCP 8.5 2040-2100

Responses to warming climate start to deviate greatly under the future climate scenario RCP 8.5. The distribution of the three species groupings largely had a similar pattern for RCP 4.5 between 2040-2060, but the similarities end between 2080-2100.

Mediterranean species have largely lost suitability in its former core area in the Mediterranean Basin. This decline would likely be the sharpest decline in Mediterranean species in the Mediterranean Basin since the Neogene(Rundel 2019). Suitable areas in Western Europe have been lost as well. The remaining areas of suitability follow a strong Atlantic distribution with the areas of highest suitability being found along the English Channel. Moreover the majority of England houses highly suitable habitat for Mediterranean species. The English Channel and the North Sea are now the core for Mediterranean species in Europe. This can indicate a severe shift in precipitation patterns over its former core area. It is unknown what will replace the vegetation in the Mediterranean Basin, but it can be inferred that desertification will lead to large arid landscapes throughout southern Europe.

Temperate species still retain some refugia in Central Europe, likely still finding suitable habitat on altitudinal forests. The Alps and Caucasus mountains remain as core suitable areas. Thus, we can assume that broadleaf forests would not be prominent in mainland Europe at the start of the 22nd century. Novel areas in Fennoscandia have developed large areas of moderate to high suitability. The distribution of suitable habitat largely resembles that of Boreal species at present. Thus, broadleaved forests in Southern Sweden, Norway, and Finland may experience great expansions in the coming decades. The cool mixed forests that once dominated Central Europe may find suitable habitat here in the future.

The suitability of Boreal species in Central Europe has almost entirely disappeared. The last refugia of Boreal species would likely be found past the current treeline in the Alps and Caucasus mountains. The remaining core areas in Norway have diminished as well as the suitable habitats on the Kola peninsula. It is possible that Boreal trees may coexist in pockets all throughout Europe, but Norway may be the last place where expansive taiga will be found in Europe.

6.4 Conclusions

We are still unable to predict which climate scenario is our most likely pathway although we can somewhat assume that warming will continue and present for many decades to come. This will have profound impacts on the prominent forest vegetation that is currently present in Europe. As mentioned before, the mid-Pliocene Warm Period represents a time interval that experienced warming analogous to predicted future warming.

There seems to be some similar responses to Pliocene warming when compared to the two future climate scenarios RCP 4.5 and RCP 8.5. Boreal species have the strongest similarities between the Pliocene and future warming. Even fossil evidence shows coniferous species being present in the far north of Norway. The response of Boreal species in Central Europe is also similar with the refugia being found on higher elevations. Thus lowland populations of coniferous forest may be in danger in the future.

Temperate species had moderately strong similarities with Pliocene warming. The distribution of refugia along highlands and altitudinal forests was shared between Pliocene and future warming. In the Pliocene temperate trees were subordinate under sub-tropical trees in the lowlands being pushed onto the hill and uplands of Central Europe. This is largely similar in future warming conditions. This could however spell disaster for many areas of Temperate forest as they will lose the majority of their suitable habitat. Areas such as southern Bohemia, the western Alps, Carpathians, and other mountainous areas will be refugia for Temperate species in mainland Europe. The shift of habitable area northward is also shared between the Pliocene and the future. Temperate species find suitable areas all throughout Fennoscandia in future scenarios while habitable regions were confined to Norway and the Kola peninsula for the Pliocene. Populations of broadleaf flora in Fennoscandia will have the opportunity to expand greatly as Boreal forests contract in the future. An area of habitation present in the Pliocene, but not in our future predictions is the Volga River Basin. Fossil evidence also supports the presence of Temperate and subtropical flora in a band that correlates with our Pliocene predictions(Velichkevich & Zastawniak 2003). Further study can be done to illuminate why this pattern is present.

Mediterranean species had the most differing results when comparing the Pliocene and future warming. Both saw a spread of suitability into Western and Central Europe however the pattern of distribution is drastically different. The Pliocene saw Mediterranean species being highly suitable in large pockets throughout Western and Central Europe with no clear core area. This can likely be due to the presence of subtropical flora competing with Mediterranean species. Future warming scenarios however show a strong Atlantic distribution with the core area being the land surrounding the English Channel and the North Sea. The core area of the Mediterranean Basin was also largely suitable in both Pliocene and future warming scenarios. This changes however with drastic declines under the RCP 8.5 scenario. Mediterranean species will no longer find suitable habitat in the Mediterranean basin likely due to strengthened desertification.

It is evident that prominent tree species of Europe will change their distribution according to differing climatic regimes, however the results of my study have shown areas where certain species groups will likely disappear without reduction of greenhouse gases and human disturbance. The remaining suitable habitat for each species grouping can aid in the management of these ecosystems in the future and possibly adapt to future warming conditions.

7. References

Abolmaali, S., Mohammad, R., Tarkesh, M., & Bashari, H. (2017). *MaxEnt modeling for predicting suitable habitats and identifying the effects of climate change on a threatened species, Daphne mucronata , in central Iran*. Ecological Informatics. 43. 10.1016/j.ecoinf.2017.10.002.

Allouche, O., Tsoar, A., & Kadmon, Ronen. (2006). Assessing the accuracy of species distribution models: prevalence, Kappa and the True Skill Statistic (TSS). Journal of Applied Ecology. 43. 1223 - 1232. 10.1111/j.1365-2664.2006.01214.x.

Akyol, A., Örücü, Ö., & Arslan, E. (2020). *Habitat suitability mapping of stone pine (Pinus pinea L.) under the effects of climate change*. Biologia. 75. 10.2478/s11756-020-00594-9.

Antúnez, P., Hernandez-Díaz, J., Wehenkel, C., & Clark-Tapia, R. (2017). *Generalized Models: An Application to Identify Environmental Variables That Significantly Affect the Abundance of Three Tree Species. Forests*. 8. 10.3390/f8030059.

Attorre, F., Francesconi, F., Scarnati, L., Sanctis, M., Alfo, M. & Franco, B. (2008). P*redicting the effect of climate change on tree species abundance and distribution at a regional scale*. iForest - Biogeosciences and Forestry. 1. 132-139. 10.3832/ifor0467-0010132.

Bazzato, E., Rosati, L., Canu, S., Fiori, M., Farris, E., & Marignani, M. (2021). *High spatial resolution bioclimatic variables to support ecological modelling in a Mediterranean biodiversity hotspo*t. Ecological Modelling. 441. 10.1016/j.ecolmodel.2020.109354.

Beaumont, L., Graham, E., Duursma, D., Wilson, P., Cabrelli, A., Baumgartner, J., Hallgren, W., Esperón-Rodríguez, M., Nipperess, D., Warren, D., Laffan, S., Vanderwal, J. (2016). *Which species distribution models are more (or less) likely to project broad-scale, climate-induced shifts in species ranges?*. Ecological Modelling. 342. 135-146. 10.1016/j.ecolmodel.2016.10.004.

Bertini, Adele & Martinetto, Edoardo. (2011). *Reconstruction of vegetation transects for the Messinian–Piacenzian of Italy by means of comparative analysis of pollen, leaf and carpological records.* Palaeogeography Palaeoclimatology Palaeoecology . 304. 230-246. 10.1016/j.palaeo.2010.09.005.

Burke, K.D., Williams, J.W., Chandler, M.A, Haywood, A.M., Lunt, D., & Otto-Bliesner, BL. (2018). *Pliocene and Eocene provide best analogs for near-future climates.* Proceedings of the National Academy of Sciences. 115. 10.1073/pnas.1809600115

Bradley, Andrew(1997). *The use of the area under the ROC curve in the evaluation of machine learning algorithms.* Pattern Recognition. 30. 1145-1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Breiman, L.(2001). *Random Forests. Machine Learning.* 45. 5-32. 10.1023/A:1010950718922.

Brown, J., Hill, D., & Haywood, A. (2020). *A critical evaluation of the Oscillayers methods and datasets*. Global Ecology and Biogeography. 29. 10.1111/geb.13103.

Coban, O., Örücü, Ö.K., & Arslan, E. (2020). *MaxEnt Modeling for Predicting the Current and Future Potential Geographical Distribution of Quercus libani Olivier*. Sustainability. 12. 1-19. 10.3390/su12072671.

Daniya, T., Geetha, M & Kumar, S. (2020). *Classification and regression trees with gini index. Advances in Mathematics Scientific Journal*. 9. 1857-8438. 10.37418/amsj.9.10.53.

De'ath, Glenn. (2007). *Boosted Trees for Ecological Modeling and Prediction.* Ecology. 88. 243-51. 10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., Bergmann, T., Bernadello, R., Bousetta, S., Caron, L., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F., Docquier, D., Echevarria, P., & Zhang, Q. (2021). The EC-Earth3 Earth System Model for the Climate Model Intercomparison Project 6. 10.5194/gmd-2020-446.

Dowsett, H., Foley, K., Stoll, D., Chandler, M., Sohl, L., Bentsen, M., Otto-Bliesner, B., Bragg, F., Chan, W., Contoux, C., Dolan, A., Haywood, A., Jonas, J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D., Nisancioglu, K., & Abe-Ouchi, A., & Zhang, Z. (2013). *Sea Surface Temperature of the mid-Piacenzian Ocean: A Data-Model Comparison*. Scientific reports. 3. 2013. 10.1038/srep02013.

Elith, Jane & Burgman, Mark. (2002). *Predictions and their validation: Rare plants in the Central Highlands, Victoria, Australia*. The University of Melbourne

Elith, J., Leathwick, J & Hastie, T. (2008). *A Working Guide to Boosted Regression Trees. The Journal of animal ecology*. 77. 802-13. 10.1111/j.1365-2656.2008.01390.x.

Eyring, V., Bony, S., Meehl, G., Senior, C., Stevens, B., Ronald, S., & Taylor, K.. (2015). *Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation*. Geoscientific Model Development Discussions. 8. 10539-10583. 10.5194/gmdd-8-10539-2015.

Fauquette, S., Joel, G., Suc, J. (1998). *A method for climatic reconstruction of the Mediterranean Pliocene using pollen data.* Palaeogeography Palaeoclimatology Palaeoecology. 144. 183-201. 10.1016/S0031-0182(98)00083-2.

Ferguson, David & Knobloch, Erwin. (1998). *A fresh look at the rich assemblage from the Pliocene sink-hole of Willershausen, Germany*. Review of Palaeobotany and Palynology. 101. 271-286. 10.1016/S0034-6667(97)00078-X.

Friedman, J., Hastie, T., Tibshirani, R. (2000). *Additive Logistic Regression: A Statistical View of Boosting. The Annals of Statistics*. 28. 337-407. 10.1214/aos/1016218223.

Gamisch, Alexander. (2019). *Oscillayers: A dataset for the study of climatic oscillations over Plio-Pleistocene time-scales at high spatial-temporal resolution*. Global Ecology and Biogeography. 28. 10.1111/geb.12979.

Guisan, Antoine & Zimmerman, Nikkaurs E.(2000). *Predictive habitat distribution models in ecology*. *Ecological Modelling* 135 (2000) 147–18

Han, Q., Keeffe, G., & Cullen, S. (2021). Climate Connectivity of European Forests for Species Range Shifts. Forests. 12. 940. 10.3390/f12070940.

Haywood, A., Valdes, P., Francis, J.E., & Sellwood, B. (2002). Global middle Pliocene biome reconstruction: A data/model synthesis. Geochemistry Geophysics Geosystems. 3. 1-18. 10.1029/2002GC000358.

Higgins, SI., Larcombe, MJ., Beeton, NJ., Conradi, T., & Nottebrock, H.(2020) *Predictive ability of a process-based versus a correlative species distribution model.* Ecol Evol. 10: 11043– 11054. https://doi.org/10.1002/ece3.6712

Hijmans, R., Cameron, S., Parra, J., Jones, P., Jarvis, A. (2005). *Very high resolution interpolated climate surfaces of global land areas. International Journal of Climatology*. 25. 1965-1978. 10.1002/joc.1276.

Johns, T., Gregory, J., Ingram, W., Johnson, C., Jones, A., Lowe, J., Mitchell, J., Roberts, D., Sexton, D., Tett, S. & Woodage, M.. (2003). *Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios*. Climate Dynamics. 20. 583-612. 10.1007/s00382-002-0296-y.

Jost, A., Fauquette, S., Kageyama, M., Krinner, G., Ramstein, G., Suc, J., & Violette, S. (2009). *High resolution climate and vegetation simulations of the Mid-Pliocene, a model-data comparison over western Europe and the Mediterranean region.* Climate of the Past. 5. 585-606. 10.5194/cp-5-585-2009

Keenan, T., Serra-Diaz, J.M., LLoret, F., Ninyerola, M., & Sabaté, S. (2011). *Predicting the future of forests in the Mediterranean under climate change, with niche- and process-based models: CO2 matters!*. Global Change Biology. 17. 565-579.

Kvacek, Z., Teodoridis, V., & Gregor, H.-J. (2008). *The Pliocene leaf flora of Auenheim, Northern Alsace (France)*. Documenta Naturae. 155. 1-108.

Kvacek, Z., Teodoridis, V., & Denk, T. (2019). *The Pliocene flora of Frankfurt am Main, Germany: taxonomy, palaeoenvironments and biogeographic affinities*. Palaeobiodiversity and Palaeoenvironments. 100. 10.1007/s12549-019-00391-6.

Komac, B., Esteban, P., Trapero, L., & Caritg, R. (2016). *Modelization of the Current and Future Habitat Suitability of Rhododendron ferrugineum Using Potential Snow Accumulation*. PloS one. 11. e0147324. 10.1371/journal.pone.0147324.

Lawler, J., White, D., Neilson, R., & Blaustein, A. (2006). *Predicting climate-induced range shifts: Model differences and model reliability*. Global Change Biology. 12. 1568 - 1584. 10.1111/j.1365-2486.2006.01191.x.

Maiorano, L., Chiaverini, L., Falco, M., & Ciucci, P. (2019). *Combining multi-state species distribution models, mortality estimates, and landscape connectivity to model potential species distribution for endangered species in human dominated landscapes*. Biological Conservation. 237. 19-27. 10.1016/j.biocon.2019.06.014.

Marchitto, T.M., Curry, W.B., Lynch-Stieglitz, J., Bryan, S., Cobb, K., & Lund, D.(2014). *Improved oxygen isotope temperature calibrations for cosmopolitan benthic foraminifera*. Geochimica et Cosmochimica Acta. 130. 1–11. 10.1016/j.gca.2013.12.034

Mauri, Achille & Strona, Giovanni & San-Miguel-Ayanz, J.. (2017). *EU-Forest, a high-resolution tree occurrence dataset for Europe*. Scientific Data. 4. 160123. 10.1038/sdata.2016.123.

O'Brien, C., Inglis, G., Robinson, S., Pancost, R., Huber, B., Farnsworth, A., Lunt, D. (2018). *Cretaceous-Paleocene-Eocene Sea surface temperature evolution at higher latitudes: constraints from tex 86 and planktonic foraminiferal oxygen isotopes*. 10.1130/abs/2018AM-323378

Ohlemüller, R., Gritti, E., Sykes, M., & Thomas, C. (2006). *Quantifying components of risk for European woody species under climate change.* Global Change Biology. 12. 1788 - 1799. 10.1111/j.1365-2486.2006.01231.x.

Palter, Jaime. (2015). *The Role of the Gulf Stream in European Climate. Annual review of marine science*. Annual Review of Marine Science. 7. 113-37. 10.1146/annurev-marine-010814-015656.

Pauli, H., Gottfried, M., Dullinger, S., Abdaladze, O., Akhalkatsi, M., Benito, A., José, L., Coldea, G., Dick, J., Erschbamer, B., Calzado, F., Ghosn, D., Holten, J., Kanka, R., Kazakis, G., Kollar, J., Larsson, P., Moiseev, P., Moiseev, D., Molau, U., & Grabherr, G. (2012). *Recent Plant Diversity Changes on Europe's Mountain Summits*. Science (New York, N.Y.). 336. 353-5. 10.1126/science.1219033.

Petitpierre, B., Broennimann, O., Kueffer, C., Daehler, C., & Guisan, A. (2016). *Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions: Which predictors increase the transferability of SDMs?*. Global Ecology and Biogeography. 26. 10.1111/geb.12530.

Phillips, S.J., Dudik, M., & Schapire, R.E.(2004). *A maximum entropy approach to species distribution modeling.* ICML '04: Proceedings of the twenty-first international conference on Machine learning

Phillips, S.J., Anderson, R.P., & Schapire, R.E.(2006). *Maximum entropy modeling of species geographic distributions.* Ecological Modelling 190 (2006) 231–259

Qazi, A.W., Saqib, Z. & Zaman-ul-Haq, M.(2022). *Trends in species distribution modelling in context of rare and endemic plants: a systematic review*. Ecol Process 11, 40 (2022). https://doi.org/10.1186/s13717-022-00384-y

Rundel, Philip. (2019). A Neogene Heritage: Conifer Distributions and Endemism in Mediterranean-Climate Ecosystems. Frontiers in Ecology and Evolution. 7. 10.3389/fevo.2019.00364.

Roberts, Malcolm (2017). *MOHC HadGEM3-GC31-LL model output prepared for CMIP6 HighResMIP hist-1950*. Version *YYYYMMDD*[1].Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.6042

Salzmann, U., Williams, M., Haywood, A., Johnson, A., Kender, S. & Zalasiewicz, J. (2011). *Climate and environment of a Pliocene warm world.* Palaeogeography, Palaeoclimatology, Palaeoecology. 309. 10.1016/j.palaeo.2011.05.044

Senay, Senait & Worner, Sue. (2019). *Multi-Scenario Species Distribution Modeling. Insects*. 10. 65. 10.3390/insects10030065.

Speiser, J., Miller, M., Tooze, J., & Ip, E. (2019). *A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling*. Expert Systems with Applications. 134. 10.1016/j.eswa.2019.05.028.

Svenning J.C., Normand, S., & Kageyama, M. (2008). *Glacial refugia of temperate trees in Europe: insights from species distribution modelling*. Journal of Ecology 2008, 96, 1117–1127. doi: 10.1111/j.1365-2745.2008.01422.x

Teodoridis, V., Martinetto, E., Bruch, A., Vassio, E., Kvacek, Z., & Stuchlik, L. (2015). *Plio-Pleistocene floras of the Vildštejn Formation in the Cheb Basin, Czech Republic — A floristic and palaeoenvironmental review.* Palaeogeography Palaeoclimatology Palaeoecology. 467. 10.1016/j.palaeo.2015.09.038.

Thuiller, W., Vayreda, J., Pino, J., Sabaté, S., Lavorel, S., & Gracia, C. (2003). *Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain)*. Global Ecology and Biogeography. 12. 10.1046/j.1466-822X.2003.00033.x.

Thuiller, W., Georges, D., & Engler, R. (2009). *biomod2: Ensemble platform for species distribution modelling*. Ecography. 32:369-373

Vieira, M., Poças, E.,& Pais, J. (2011). *Pliocene flora from S. Pedro da Torre deposits (Minho, NW Portugal)*. Geodiversitas. 33. 71-85. 10.5252/g2011n1a5.

Vieira, M., Pound, M., & Pereira, D. (2018). *The late Pliocene palaeoenvironments and palaeoclimates of the western Iberian Atlantic margin from the Rio Maior flora*. Palaeogeography, Palaeoclimatology, Palaeoecology. 10.1016/j.palaeo.2018.01.018.