Albert-Ludwigs-Universität Freiburg
Faculty of Environment and Natural Resources

Douglas-Fir Growth under Future Climate

Applications of a Century of Provenance Trial Data

Raphael Habel
Matriculation No. 3536501

Supervisor: Dr. David R. Roberts
Second Supervisor: Prof. Dr. Jürgen Bauhus

28.04.2016

Acknowledgements

First of all, I want to thank Prof. Dr. Carsten Dormann and Prof. Dr. Jürgen Bauhus for supervising my project and for making this thesis possible.

A huge “Thank you!” goes out to David Roberts. Thanks for your advice and support, for having an open door and all the time in the world whenever I needed it. You were an exceptional supervisor! I have learned so much in the course of this thesis, because of you.

Moreover, there are some people, without whom this would have never been possible and who I want to thank here as well:

First of all, THANKS to Jascha for taking care of my whole life apart from the thesis; Klaus, for motivational and focusing assistance (sorry for being not focused enough to finish in time); Jonas, for being someone to count on when it is most urgent; to Luke and Sam for linguistic advice, and to all the great people, who make life apart from Uni so special.

Finally I want to thank my parents for not just enabling me to study, but also for backing every decision on the way, for the most amazing and unconditional support and all the other things too numerous to list here.

Table of Content

Table of Figures

Abbreviation Index

Provenances

C_BC	Coastal British Columbia
C_OR	Coastal Oregon
C_WA	Coastal Washington
CC_OR	Cascades Oregon
CC_WA	Cascades Washington
DC_OR	Dry-Coast Oregon
DC_WA	Dry-Coast Washington
HE_CA	High-Elevation California
LE_CA	Low-Elevation California
I	Interior
IC	Coastal Interior
IN	Interior North
IS	Interior South

Bioclimatic Variables

MAT	Mean Annual Temperature (°C)
MAP	Mean Annual Precipitation (mm)
AHM	Annual Heat-Moisture Index (MAT+10)/(MAP/1000) (°C/mm)
MWMT	Mean Warmest Month Temperature (°C)
MCMT	Mean Coldest Month Temperature (°C)
DD_0	Degree Days below 0°C; chilling Degree-Days
DD5	Degree Days above 5°C; growing Degree-Days
FFP	Frost Free Period
EMT	Extreme Minimal Temperature (°C)
SHM	Summer Heat-Moisture Index (MWMT/ (MSP/1000))
TD	Continentality Index; Difference between MWMT and MCMT (°C)
MDMP	Mean Driest Month Precipitation (mm)

Models

GLM	Generalized Linear Model
RF	Random Forests (Model)
PCA	Principal Component Analysis

PC ..... Principal Component

CV ..... Cross-Validation

VAR ..... Value Above Random

VAB ..... Value Against Best

VAC ..... Value Above Consistency

Abstract

Seed transfer and the suitability of tree provenances to the target climate are important factors for reforestation and forest management issues. These aspects are expected to gain significance considering the imminent era of anthropogenic climate change. One of the most common practices in this field of research involves response functions that calculate growth as a function of climate variables. In this bachelor thesis, I applied response functions to European forests by using data from 112 common garden experiments to investigate the future growth of Douglas-Fir and its multiple provenances under different climate change scenarios. The main findings of this thesis are that Douglas-Fir is likely to stay a promising tree species for forestry in Europe, particularly at high altitudes and latitudes. According to my models, the optimal selection of planting provenances is not going to change significantly, even though we can perceive a trend towards seed material from dryer origins. Furthermore, in terms of methodology I conclude that GLMs are a more suitable tool for modelling future tree height than predictions made with Random Forests models as a comparison.

Zusammenfassung

Der Transfer von Pflanzmaterial und die Tauglichkeit von Baumprovenienzen in Bezug auf das Zielklima sind wichtige Faktoren für Wiederaufforstung und forstwirtschaftliche Entscheidungen. Im Angesicht des anthropogenen Klimawandels ist zu erwarten, dass diese Aspekte noch an Bedeutung gewinnen. *Response Functions* sind eine weit verbreitete Methode auf diesem Forschungsgebiet. Diese berechnen Wachstum als eine Funktion klimatischer Variablen. In dieser Bachelorarbeit wende ich *response functions* auf europäische Wälder an, um das zukünftige Wachstum der Gewöhnlichen Douglasie und ihrer zahlreichen Provenienzen unter verschiedenen Klimawandel-Szenarien zu untersuchen. Dabei nutzte ich Daten aus 112 Pflanzversuchen in Europa. Die wesentlichen Ergebnisse dieser Arbeit zeigen, dass die Gewöhnliche Douglasie vorrausichtlich auch weiterhin eine vielversprechende Spezies für das Forstwesen in Europa darstellt, insbesondere in hohen Höhenlagen und Breitengraden. Nach meinem Modell wird sich die optimale Pflanzprovenienz nicht signifikant ändern, auch wenn ein Trend hin zu Samenmaterial aus trockeneren Regionen zu beobachten ist. In Bezug auf die Methodik lässt sich zudem feststellen, dass sich GLMs als ein geeigneteres Instrument für eine Modellierung des zukünftigen Wachstums erwiesen haben als Random Forest-Modelle.

1 Introduction

1.1 Forests under the Influence of Climate Change

A growing number of scientific publications have identified the severe challenges, climate change is likely to impose on health and productivity of forest ecosystems (LINDNER ET AL. 2010). As climate zones start to shift northwards, so do optimal habitat conditions for a variety of species (HAMANN & WANG 2006). The frequency of extreme weather events such as droughts, fires, or floods is likely to increase significantly in Europe and especially in the Mediterranean regions (IPCC 2013B). Together with these abiotic factors, biotic stress factors such as harmful fungi or pests will shift their distribution range and thereby affect ecosystems that will often have difficulties to react effectively to these new exposures (AITKEN ET AL. 2008). Tree species are especially prone to changing climate conditions of their environment, since due to slow reproduction rates and rotation periods, their adaptive capacities are limited (LINDNER ET AL. 2010). For example, REHFELDT ET AL. (2002) predicted that some Pine species need 12 generations or, considering the life span of individual trees, approximately 1500 years to fully adapt to a new climatic environment. Given the current rate of climate change, it is obvious that the local adaptation of tree species is too slow to keep up with the rapidity of anthropogenic global warming (DAVIS & SHAW 2001).

The same applies to migration, which is the second possible reaction of tree species to a changing environment. Even though palaeobiotic pollen analyses show that tree species have successfully reacted to changing climate conditions by shifting their habitat towards areas with optimal growing conditions, the dynamic of the current, anthropogenic climate change is already outpacing migratory rates of many tree species by an order of magnitude (DAVIS & SHAW 2001). The fragmentation of natural landscapes by human activities is an exacerbating factor, which can additionally compromise successful range shifts on a local level (MALCOLM ET AL. 2002; SCHWARTZ 1992).

Thus, many of the trees that are currently being planted in European forests will still be standing there, when the global temperature will have risen by up to 4.8°C (RCP8.5, IPCC 2013A) and the frequency of extreme weather events will have changed significantly as well (IPCC 2013B). As a result of Europe's climatic and

topographic variability, it is difficult to make general statements about the effects of climate change on European forest ecosystems. While in some areas, such as the northern and boreal zones, the outcomes might even be beneficial, most European forests are expected to increasingly show signs of maladaptation to their current and rapidly changing environment, including decreasing health and growth performance, increasing susceptibility to pests (LINDNER ET AL. 2010), and eventually dropping survival rates of individual trees and the extinction of species or populations (THOMAS ET AL. 2004; AITKEN ET AL. 2008). Against this background, especially drought and heat resistance are valuable traits for future forest trees (EILMANN ET AL. 2013). In an assessment of heat and dryness stress on forests, ALLEN ET AL. (2010) documented that the number of scientific reports on warming/drought-induced forest mortality has increased considerably.

This degradation can have dramatic consequences, as forests provide a variety of purposes for society: from an economic perspective, fading productivity rates are all the more concerning in the face of a constantly rising global demand for timber products, which is caused by the world's population and economic growth and an increasing interest in forests as renewable resources for biomass (FAO 2009). Furthermore, forests are highly important carbon sinks. Their well-being mitigates climate change, and decreasing growing and survival rates might trigger climate feedback mechanisms (PAN ET AL. 2011; ALLEN ET AL. 2010). Finally, there are concerns about biodiversity loss and the recreational benefits for society that forests provide. Against this background, the development of climate resilient forest ecosystems should be a central concern from both research and management perspectives.

1.2 Implications for Forest Management and Seed Transfer

When adaption and migration responses fail, many species would face extinction under natural circumstances. Six to eleven percent of species in natural reserves are predicted to go extinct under current emission scenarios (ARAÚJO ET AL. 2004). Another study by THOMAS ET AL. (2004) investigating the survival of endemic species in the face of climate change found extinction rates between fifteen and thirty-seven percent. However, due to intensive forestry most European forests can actually not be considered natural ecosystems. For centuries, European forests have been actively managed and altered by decision-makers, such as private land owners, municipalities or state representatives (JOHANN 2004). The fact that forest owners decide

about the composition of tree species on their land might be a solution to the impending scenario described above. When future climate conditions are taken into consideration in forest management decisions, and when seeds and saplings are chosen wisely, anthropogenic influence could counteract natural limitations of species adaptation and migration (WANG ET AL. 2010; LEDIG & KITZMILLER 1992). Natural migration rates of tree populations do not have to keep up with the shift of climate ranges. Instead, seeds can be transferred and planted at locations, where projected climate conditions better match the tree's biological requirements. Moreover, the previously mentioned expansive time spans for in-situ adaptation could be bypassed by planting seeds from tree populations that have been adapting to similar climate conditions at different sites for generations.

Tree species show an especially high variety of phenotypes and quantitative traits as a result of local adaptation to external factors such as climate, soil conditions, or inter- and intraspecific competition (SAVOLAINEN ET AL. 2007). With climatic variables as a very influential factor in this regard, tree populations have developed a great genetic variation along climatic clines (HOWE ET AL. 2003). Therefore, growth and productivity of seed material at a certain site rely on the compatibility of climatic conditions of origin and transfer location. Most research on seed and gene transfer in forestry is based on data from provenance trials, which are also known as common garden experiments. Provenance trials are long-term field studies, in which tree seeds from a variety of provenances are planted under standardized conditions at a certain site. Eventually, differences in height, health and survival rates reveal the suitability of certain populations for the environmental conditions of this specific area.

Targeted seed transfer under consideration of future climate conditions at respective sites not only offers a great opportunity, but also entails a mandate for environmental research to identify and optimize transfer models. Using transfer models as a forest management tool, means making large economic decisions based on models and predictions, which should be as accurate as possible. One of the most promising approaches in seed transfer research is the investigation of future growth performance through response functions, which will also be applied in this bachelor thesis.

1.3 Response Functions

Response functions are a well-recognized tool for modelling growth performance of tree populations as a function of predicting parameters, which usually characterize the prevailing climate. They are based on data from provenance trials and have been frequently used in current research on seed transfer (E.G. CHAKRABORTY ET AL. 2015; O'NEILL ET AL. 2008)

Response functions return the response of a certain population to the range of a climate variable. The response is usually measured as height, breast height diameter or presence-absence data, but the number and combination of explanatory climate variables can vary and depend on the data available. From a conceptual perspective, response-functions are usually inversed quadratic functions with growth performance rising from both sides towards a biological optimum. When trees of similar genetic origin are planted in several common gardens, whose locations cover different realizations of a certain climate variable, it is possible to determine the optimal growing conditions or growing locations for a given genetic group. An alternative to response functions are transfer functions, which describe the relationship between a response variable and a transfer distance (with equal climate conditions at transfer distance zero). This option was also considered but later rejected, because as WANG ET AL. (2010) stated, “a transfer from mean annual temperature (MAT) 10° to 8°C (i.e., 2°C transfer) may have a dramatically different effect on phenotypes than a transfer from 0° to 2°C (also a 2°C transfer)” (p. 154).

In this research project, provenance trial data has been used to create response functions of genetically similar Douglas-Fir populations, in order to predict the future growth of Douglas-Fir in Europe.

1.4 History of the Douglas-Fir and its Importance for Seed Transfer Research

The Douglas-Fir (*Pseudotsuga menziesii*) is an important commercial timber species, which produces wood of high-quality at high growing rates (KLEINSCHMIT & BASTIEN 1992). Its natural range covers a huge area in the west of North-America, where it is the predominant conifer species. It abundantly occurs along the coastal mountain ranges from southern British Columbia down to the coastal areas of California. Further inland it also inhabits the Rocky Mountain range from widely spread habitats in

British Columbia and the northwest of the United States and discontinuously southwards through Utah, Colorado, New Mexico and Arizona down to small fragmented populations in Mexico (Figure 1). While there are some *pseudotsuga* varieties of smaller prominence in Asia and Mexico (HOWE ET AL. 2006), two main subspecies are primarily mentioned in common literature: *Pseudotsuga menziesii* var. *menziesii*, which inhabits the coastal strip of land and is therefore also known as coastal Douglas-Fir, and *P. menziesii* var. *glauca* or the interior Douglas-Fir, which grows in the continental habitats east of the Cascades (HERMANN & LAVENDER 1999).

Throughout its widespread natural range Douglas-Fir populations have adopted to a large number of regional climates. Coastal trees from provenances west of the Rocky Mountain Range, for instance, grow under extremely wet and mild conditions, with mean annual precipitation rates of up to 4617mm/yr. Precipitation, humidity, temperatures, and length of growing season decrease from east to west, as the Douglas-Fir's habitat crosses the Rocky Mountain, Cascade, and Sierra Nevada Range (MORGENSTERN 1996). On the other side of the climate spectrum, interior Douglas-Fir can also be found in continental areas, where the annual precipitation does not exceed 114 mm/yr and temperatures can vary by more than 40K in the course of a year (ISAAC-RENTON ET AL. 2014).

In the early 19th century the Douglas-Fir was introduced to Europe. Due to its favourable traits and growth performance it became a quite important timber wood in several European countries. According to the German federal tree inventory 2015, Douglas-Fir now covers 217604 ha of forest land in Germany, which makes it the most important introduced timber species from North America (BMEL 2014). Fortunately for today's research on Douglas-Fir, its introduction to European forests was accompanied by comprehensive research efforts aiming to find the best performing seed sources of this promising new timber species (KLEINSCHMIT & BASTIEN 1992). This is one of the reasons for the sound spatial coverage of provenance trials in our dataset.

1.4.1 Differences within the Douglas-Fir species

According to MORGENSTERN (1996), provenance trials revealed "a parallel pattern of decreasing growth rate from west to east". Seeds from the coastal areas of Oregon, Washington and the southern part of British Columbia generally performed best in European climates, followed by coastal populations from montane provenances, and interior populations showing the worst growth-performance. As a result, seeds from

coastal populations were predominantly used in Europe, whereas planting the interior variety has not been pursued on a large scale. Since the mid-19th century silvicultural yield tables from forestry research in Europe exclusively mention coastal varieties (HERMANN & LAVENDER 1999).

Differences in the growth-performance of coastal and interior taxa can be explained biologically with the trade-off between growth rates and stress tolerance (ST CLAIR ET AL. 2005). The more inhospitable the climate in which a population is situated, the more energy individual trees have to invest in protective traits. As tree damage is most dangerous when the tree is actively growing, one example of an adaptive trait is the adjustment of bud-burst and growing cessation (WHITE 1987), which results in a generally shorter growing season and thus smaller yields from this particular population. Populations from mild climates can therefore invest more of their energy supplies into growth performance. In Europe, where the climate is generally less extreme than in North-America, this biological trade-off mechanism favors coastal provenances. Even among provenances of the coastal variety, which are known to be superior seed sources for European plantations EILMANN ET AL. (2013) found substantial differences in terms of seedling survival, yield, wood quality and drought tolerance.

1.5 Research Gap and Hypotheses

Douglas-Fir is a very suitable species for the investigation of seed transfer models in the European context for a number of reasons. First, the steady economic interest in Douglas-Fir timber during the past century set the ground for a large number of common garden experiments and a comprehensive amount of data. Second, Douglas-Fir is expected to further gain importance in the European forestry sector because of its favourable traits such as high productivity, wood quality, and drought resilience with the latter becoming increasingly important considering prospects of future climate conditions in Europe (EILMANN ET AL. 2013). Third, because of its huge and diverse natural range, there is a great variety of provenances to draw potential seed material from. As a consequence, there are a number of recent publications providing the scientific context in which this thesis is located.

In 2007, (ST. CLAIR & HOWE) AND HOWE investigated adaptive traits in provenance trials in order to assess the potential maladaptation of Douglas-Fir to future climates in Oregon and Washington, USA. They drew the conclusion that human intervention

would be necessary to ensure a successful adaptation and to maintain the productivity of Douglas-Fir forests. Furthermore, they recommended planting seed sources from more southern and montane provenances in order to address the challenges of a warming climate. MONTWÉ ET AL. (2015) found comparable results when conducting a similar experiment in British Columbia. Their research focus on drought tolerance did not only confirm the trade-off between drought resistance and productivity. They also call for a profound consideration of water availability for future forest management decisions, because the desired yield results of vigorously growing provenances only arise under optimal conditions, while planting more resilient populations under moist conditions might be counterproductive and cause significant productivity losses. EILMANN ET AL. (2013) stated that currently followed planting recommendations for Douglas-Fir in Europe might be outdated, because they did not consider today's rapidly changing climate conditions. Their objective to identify the most suitable provenances for future climates coincides with my research goal. However, only one common garden site in the Netherlands was used as a proxy for European climate. In addition, in their research they conducted a mixture of dendrochronological research with linear regression analysis, whereas the methodology of my approach is more similar to the research by WANG ET AL. (2010). In this study, universal response functions were used to predict the distribution of lodgepole pine under different climate change scenarios in British Columbia. I intended to create similar optimal distribution and growth-response maps, but for Douglas-Fir in Europe. Such predictions have been conducted before by ISAAC-RENTON ET AL. (2014), but on the basis of climate envelope models instead of response functions and without a quantitative analysis of the implications for forests management.

Against this background, I created response functions from a dataset of Douglas-Fir provenance trial data, in order to further research on Douglas-Fir seed transfer in European forests. For this purpose, I set the following research objectives:

1. Detect and quantify the general provenance effect in growth-response-functions between genetically similar Douglas-Fir subpopulations on the basis of the provenance trail data.
2. Use these functions to develop recommendations for forest management about the performance of different provenances under future climates.

2 Methods

2.1 Retrieving Provenance Trial Data

The dataset was compiled and adapted by ISAAC-RENTON ET AL. in 2013. In the course of a master thesis they compared growth expectations from bioclimatic envelope models with measured Douglas-Fir heights from European common garden experiments. For this purpose, they collected data from 39 publications and technical reports. The resulting data set consisted of 2795 Douglas-Fir trees, which originate from a variety of 375 provenance locations and were planted at 120 different sites in Europe. The quality of the data was validated and adjusted with the geographic information system software ArcGIS from ESRI. In order to create applicable results for forest management, the wide range of provenance locations was grouped into 14 provenance groups of similar genetic origin.

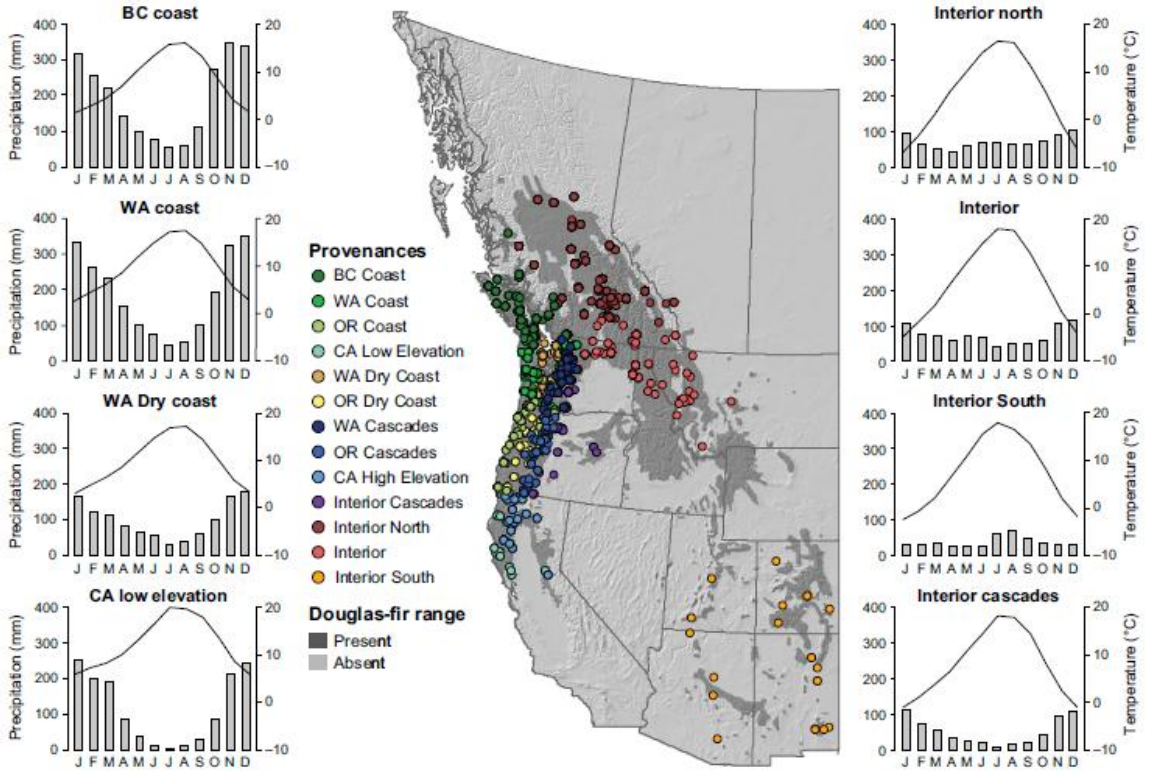

Figure 1 (from ISAAC-RENTON ET AL. 2014, p. 2610): Distribution of Douglas-Fir in grey with classification of provenances

As described previously, the climatic conditions of a tree population's environment have a strong influence on the tree's genetic traits and their growth performance under different climates. Since the desired outcome of this grouping procedure was a

set of genetically similar subpopulations that would perform equally if planted at a given site, climatic resemblance worked as a key differentiator in this classification. Principal component analysis and multivariate regression tree analysis was performed with the provenance's climate data to determine climatic proximity. In some cases, political boundaries were also used to separate provenance populations. Yet, this criterion was mainly used in the coastal habitat, where not only the habitat's huge North-South expansion requires certain latitudinal partitioning, but where an administrative division procedure also facilitated the research project's application of seed acquisition. Another criterion besides climatic analogy was geographic proximity. The actual possibility to exchange genes is a mandatory condition for sharing a common genepool. Similar to the grouping of tree origin by provenance climate, a site group had been assigned to all planting sites as well. Sample trees younger than 5 years were excluded from the data, because climate transfer distances need a certain time to reveal the investigated impacts through differentiated tree heights (ISAAC-RENTON 2013).

2.2 Retrieving Climate Data

Climate data was retrieved as open source data from the website of Andreas Hamman, Professor at the Faculty of Agricultural, Life, and Environmental Sciences at the University of Alberta. This climate data has been created with two software packages called ClimateWNA and ClimateEU, which are based on methodology described by HAMANN ET AL. (2013) and allow free access to a database of high-resolution climate data for western North America and Europe. Geographic climate surfaces for the European continent are available for 22 bioclimatic variables as well as 48 monthly meteorological variables. The climate normal period of 1961 to 1990 was used as the reference basis for climate data before anthropogenic global warming and also as the training data for growth-expectation models. Future European climate data, coordinate grids containing climate variable values in 1km resolution, was downloaded for three decades (2020s, 2050s, 2080s) and two emission scenarios (RCP4.5 and RCP8.5). The climate models on the Hamman's website are based on the average of 15 Atmosphere-Ocean Global Circulation Models of the CMIP5 multimodel dataset corresponding to the IPCC AR5.

2.3 Adjustment of the Initial Data

The first steps of data analysis required some preparatory work on the data set. The provenance group “mexican” was merged with “interior south”. With a sample size of only five trees it was impossible to generate a reliable model, but adding them to “interior south” also improved this relatively weak sample to 27 entries. The loss of accuracy and applicability due to this conjunction should be acceptable, because Mexican provenances will be genetically more similar to the southern interior populations than to any coastal or northern variety, while the southern interior provenance group is already stretching over a wide geographic distance. Obviously flawed data was identified mostly in residual plots and if a site showed a consistent unreliability, it was removed from the dataset. For instance, at one site close to Saint-Julien-le-Petit in France (site ID #80) height had been documented as standardized values, which showed up as negative growth values. Another site near Kirchzarten, Germany (site ID #71) had been adjusted the wrong coordinates. Finally, measurements from a site near Bande in Spain (site ID # 97) were excluded because they included obviously wrong data, such as trees with height 10m at the age of 5 years. Two other singular trees were removed after an analysis of their impact on the models. Even though removing outliers is a questionable procedure, I decided to remove them for a number of reasons; mainly, because they severely impacted the models of the provenances "I" (tree ID #180) and "IC" (tree ID #102). I identified them in residual plots, where for different GLMs they had repeatedly differed by 4 (GLM) to 8 (RF) standard deviations in otherwise evenly distributed data values. Moreover, their cook’s distance was more than 28 (#180) and 135 (#102) times higher than the following values. They massively violated an even distribution of “age” as the most important predictor and finally, these extremely deviating data points belonged to provenances with a relatively small sample size (see Appendix). In the end, my dataset consisted of 2731 trees from 362 provenances planted at 112 European test sites.

Contrary to the first analysis of this dataset by ISAAC-RENTON ET AL. (2014), I refrained from using standardized heights, but decided to use actual heights, and include “age” as a predictor in my models. I believe that it produces better results, because standardized heights reflect the height of a given tree relative to the average tree height at a site. Whereas this method takes site-specific factors like soil conditions or tree care

out of the equation, standardized values depend of the choice of provenances present at a site, which is not at all consistent.

Whereas trees of less than 5 years of age were already excluded from the dataset, I also set the maximum limit of my data to 50 years. The growth rate of trees is a function of age. As a tree grows towards its natural maximum height, the growth rate decreases resulting in a saturation curve, as shown in Figure 2. By modelling only within the linear growth period I was able to treat tree age as a linear model predictor, which simplified the models without losing accuracy.

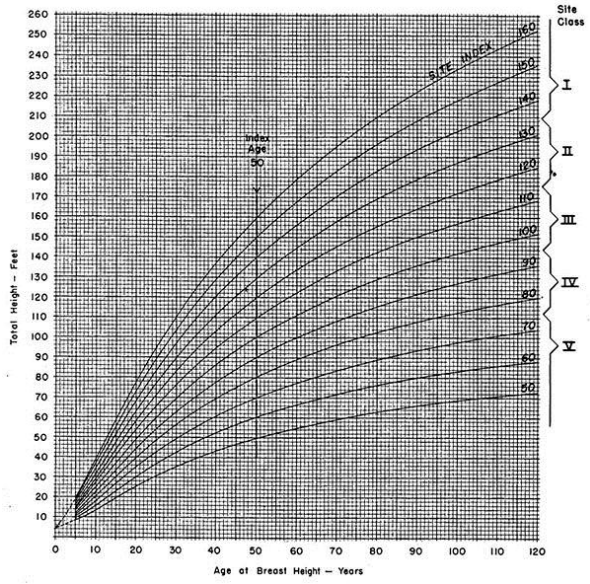

Figure 2 (from KING, J.E. 1966): Douglas-Fir growth-rate against age

2.4 Transforming Climate Data

The ASCII-file-packages containing climate grids had to be adjusted as well. I made a preselection of important bioclimatic variables based on previous research and publications. Apart from the variables in Table 1, I included “frost free period” (FFP) and “extreme minimum temperature” (EMT) into my model to account for traits like bud-set and frost resistance. All bioclimatic variables were highly correlated, because they are combinations of monthly and yearly measurements of temperature and precipitation. As a result, they were unsuitable as predictors, as correlated predictors cause problems with collinearity and variance inflation (DORMANN 2013). To avoid collinearity I performed a principal component analysis with R.

Table 1: Bioclimatic Indicators for Tree Growth and where they have been used in recent studies

Literature		Selected Variables			
CHAKRABORTY	(2013)	MAT	TD	SHM	
EILMANN	(2013)	MDMP			
GRIESBAUER	(2013)	MAT	MAP	AHM	
LEITES	(2012)	MCMT			
MONTWE	(2015)	MWMT	MDMP		
O'NEILL	(2008)	MCMT			
REHFELDT	(2003)	DD5 / DD_0	MAT	AHM	TD
REHFELDT	(2002)	DD5	DD_0	MAT	AHM
WANG	(2010)	MAT	AHM		

I downscaled the data to a 10 km resolution and I implemented a PCA of the selected variables using prcomp in the {stats}-package. Then I transformed all climate data, the 1961–1990 as well as the six future scenarios with my PCA. Every climate surface was downscaled from a 1km to a 2km grid size. Given the size of the climate surface files I had to sacrifice some accuracy in order to reduce the otherwise excessive computing time. I continued to work with the first 5 PCs (see chapter 3.1)

2.5 Model Development and Validation

Even though from a biological reasoning growth response functions display an inverted quadratic function, I decided to apply not only a GLM with quadratic predictors (as the mathematical equivalent), but for comparative reasons also a Random Forest (RF) model to my data.

RF is a method for classification or regression (in my case the latter), which is based on a repeated creation of decision trees. For each tree, data is repeatedly split by the predictor which explains the most variance until splitting the data any further does not account for a significant reduction of variance. The response value of the data points building a terminal end node are averaged and represent the output value of new data applied to the model. RF models produce robust results by averaging over a large number of trees with randomly selected subsets of data and predictor sets. GLMs are a bit easier to interpret, because the predictor's coefficients depict their relationship to the response variable. The interpretive output of RFs is an importance table, which shows the significance of different predictors for the splitting process. The more often and the earlier in the classification process a predictor served as division criterion, the higher the predictor's importance for the response value.

For the GLMs, I used age and the PCs as linear, and the PC's squared terms as quadratic predictors for height. As my research objectives require my models to extrapolate over 90 years and for a large geographic range, the fragile balance between accuracy and overfitting was a major concern. This is why I tested GLMs and RF-models with 3 to 5 PCs in order to find the best predicting models. The function stepAIC in {MASS} is often used in R to determine the most important predictors, but its procedure of excluding one predictor at a time repeatedly changes the significance of all other predictors. Therefore, the final combination depends on the order of exclusion. For the GLMs, I used the dredge-command in the R-package {MuMIn}, which tests all predictor combinations and returns the most accurate model. RF models were built with 2000 classification trees. Afterwards, each model was validated through a 16-fold cross-validation.

Cross-validation (CV) is a method to test for a model's prediction capacity. Methods like AIC or likelihood can quantify a model's fit to a dataset, but a model with a great fit is not necessarily a model that predicts equally well. In a CV, the dataset is split into several folds. Points from all but one fold build the training data for calibrating the model, which is then used to predict values for the left-out fold. This step is repeated until values for the whole dataset have been predicted. This way, the model is repeatedly confronted with unknown data, which had not been used to create the model itself. In the end, prediction quality can be estimated through the comparison of observed and predicted values. At best, the folds represent independent datasets, which in my case means different climatic conditions. A model that predicts well for unknown climates should also perform well when confronted with future climates. Furthermore, we can also see how well the models extrapolate to areas where there is no data available. Even though our dataset has a fair spatial distribution of planting sites (see Figure 3), especially in the south and east of Europe the model has to extrapolate from the data available - an ability which is also tested by CV.

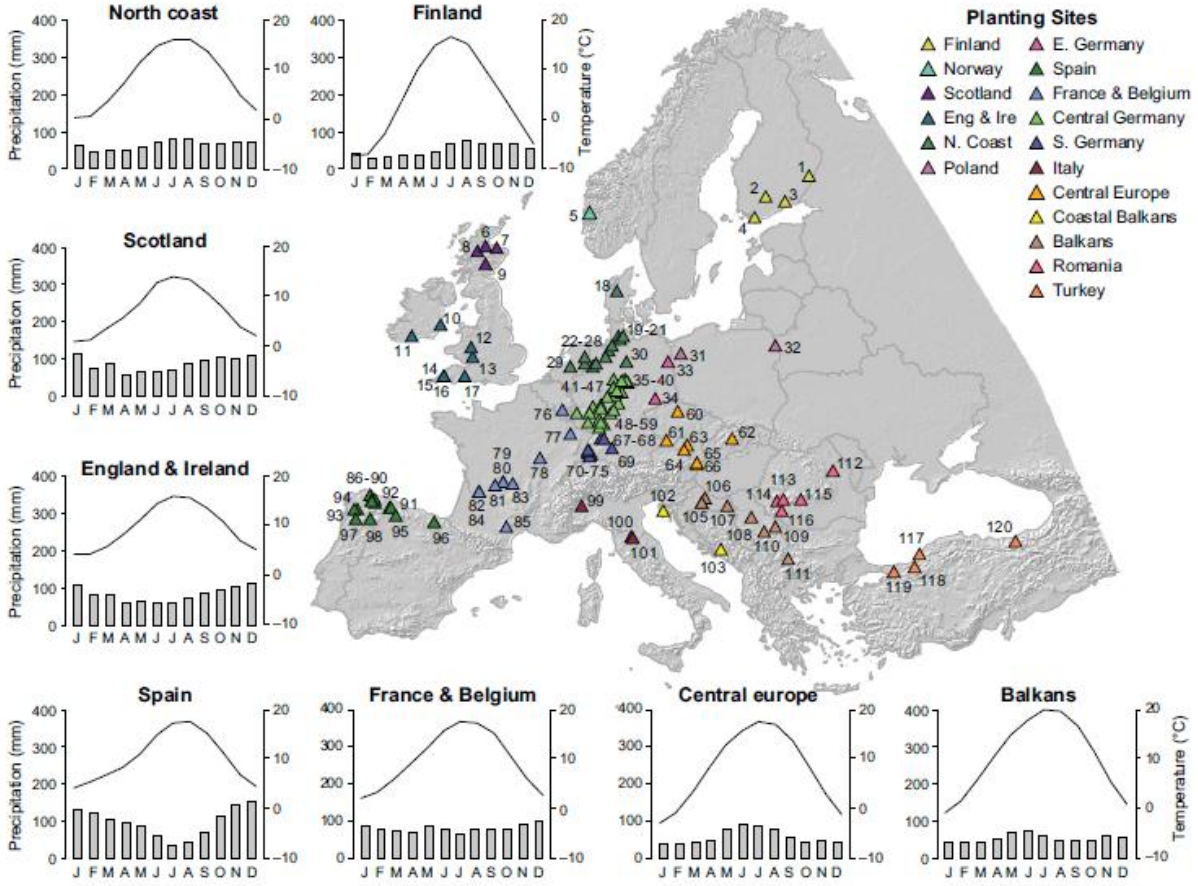

Figure 3 (from ISAAC-RENTON ET AL. 2014, p.2609): Spatial distribution and climatic classification of provenance trial locations in Europe

The planting sites had already been grouped by ISAAC-RENTON ET AL. (2014) into 16 climatically similar areas. Since some groups only contained very few sites, I also tried to merge all groups with less than 5 sites to get folds of similar size and repeated the procedure. However, clustering different climates resulted in very bad predictions for the merged groups, so I abandoned this approach. Then, the predicted results of the 16-fold cross-validation were evaluated via RMSE for the model's quantitative accuracy, and correlation indices (Pearson's r and Spearman's rank coefficient) for the qualitative accuracy, respectively (Table 3). Finally, both a GLM, and a Random Forests model was created for every provenance group with the provenance trial data.

TELFORD & BIRKS (2005) stated that many publications on response functions are unreliable, as they neglect the effect of spatial correlation between their observations. A general assumption in statistics is the independence of samples. Since my models run on climate data, absolute independency of samples would require spatially inde-

pendent climate data, which is an unrealizable assumption. A certain degree of spatial autocorrelation is yet not problematic, as long as there is no far-reaching spatial structure in the model's error. By testing for a random distribution of the model's residuals across the sites we can see if the model correctly accounts for all those coherences that it can avoid. Thus, I performed a "Moran's I"- test on special autocorrelation of the models' residuals to prevent biased results due to autocorrelation of my samples.

2.6 Height Prediction and Validation

With the individual growth functions I predicted the height of an average tree from each provenance for the 61-90 normal period and the six future climate scenarios, three decades (2020s, 2050s, and 2080s) and two emission scenarios (RCP4.5 and RCP8.5) with the factor "age" set to 30 years. I ran the models with the previously created future climate surfaces and saved the results as ASCII-files containing height predictions.

In order to test my results I implemented a number of tests. The first one is called "value above average" (VAR) and determines if the choice of provenances made by my models results in a better growth performance than a random sample of planting material. I split the provenance trial data by planting location, recorded the combination of provenance groups planted at each site, and modelled height with my models, the observed tree age, and the 61-90 climate data. Then, I ranked the provenances according to my predicted heights and took the mean of the first three genetic groups. If a site only contained three provenances, the choice size was reduced to two and if there were only two provenance groups at a site, I still tried to find the superior one. Then, I calculated the mean observed height from a random sample of three (two, or one depending on the number of provenances at the respective site) provenances, and calculated the difference between modelled-best and random-observed height values. To account for the random selection of comparison-provenances in the VAR estimation, I repeated the step 10 times and averaged the results. VAR investigates if following the models' recommendations results in a better wood harvest than choosing planting material randomly. The higher the overall VAR, the better the model works.

However, by taking VAR as the only validation criteria, massively overestimating models would appear to provide good guidance for management decisions, whereas the bias towards high values would be an inherent mistake and would not reflect a realistic scenario. This is why I performed a second test, a “value against best”- test (VAB), which is similar to VAR, but instead of taking a random sample of the observed values, I built the mean of the three best performing provenances. A good prediction model should produce a VAB close to zero. The predictions should not be higher than the actual values, because a strong positive value would be an indicator for an overly optimistic model. A negative value either shows that the model is underestimating tree heights, or that it chooses wrong and badly performing provenances.

In order to preclude the latter possibility, it is crucial not only to compare tree heights, but the compliance of observed and modelled provenance types. Therefore, I created error-of-confusion matrices for the accuracy with which both models correctly identified the best performing provenance.

2.7 Application

My research objective was to investigate future Douglas-Fir growth in Europe. The main assumption in this regard is that those provenances that have been performing best in the past might not necessarily be the superior genotypes for future European climates. In order to put that assumption to the test, I implemented a final test: “value above consistency” (VAC). I computed average future growth-expectations scenarios using the response functions of the three currently best performing provenances as observed in provenance trials. Then, I repeated the procedure with the three best provenances recommended by my models. Subtracting the entries of these two ASCII-files produced a locally defined prediction on where and how much forest owners would benefit from changing habits according to my model predictions.

3 Results

3.1 Transforming Climate Data

Figure 4 shows a bar plot of the first 7 principal components of my data. Y-axis and bar length display explained variance, while the bar’s labels show the cumulative explained variance. The loadings shown in Table 2 indicate the importance of each variable for each PC, thereby allowing for an interpretation of the predictive models.

Table 2: PCA Loadings of bioclimate variables on each PC

	PC1	PC2	PC3	PC4	PC5
MAT	-0.3676	0.0272	0.0862	-0.1541	-0.0014
AHM	-0.2688	-0.3162	-0.2272	-0.1662	0.5715
MWMT	-0.3086	-0.2325	0.4061	-0.2659	-0.0225
MCMT	-0.3426	0.2025	-0.1715	0.0635	-0.0214
MAP	0.0156	0.5014	0.5059	0.2064	-0.0217
DD5	-0.3524	-0.1175	0.2588	-0.0592	-0.1531
DD_0	0.3222	-0.2035	0.1295	0.2552	-0.2627
FFP	-0.3541	0.0552	0.0784	-0.0680	-0.4073
EMT	-0.3467	0.1811	-0.1156	0.0728	-0.1634
SHM	-0.2421	-0.2293	0.2414	0.7868	0.3415
TD	0.1958	-0.4134	0.5004	-0.2682	0.0098
MDMP	0.0958	0.4882	0.2805	-0.2487	0.5197

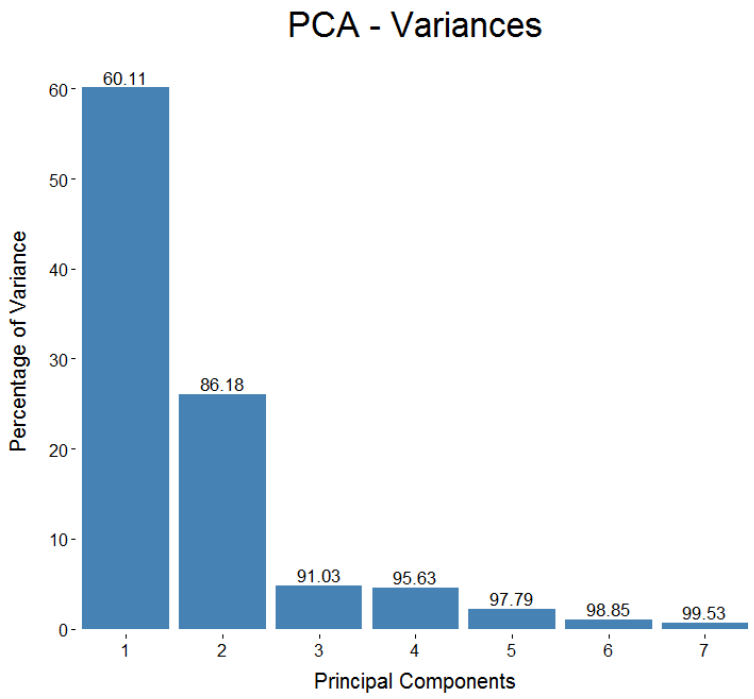

Figure 4: Bar graph of the first 7 principal components

3.2 Model Development and Validation

The results of the spatial autocorrelation test are presented in Figure 5, where I plotted the model's residuals against distance between test sites. The graphs show that I have not missed an essential predictor in my selection of climate variables. We can see at low x-values that climate is always spatially dependent to a certain degree (0.5 for GLM and 0.02 for RF), but the correlation drops quickly to values close to zero at

around 50 km lag-distance. This means that apart from close local proximity I have not missed any overarching trends, which would spatially influence the models' residuals. This test revealed another beneficial feature of my data, besides the low I-value at close distance: As most of the samples are located further apart than 50 km, they lie entirely out of the correlated range.

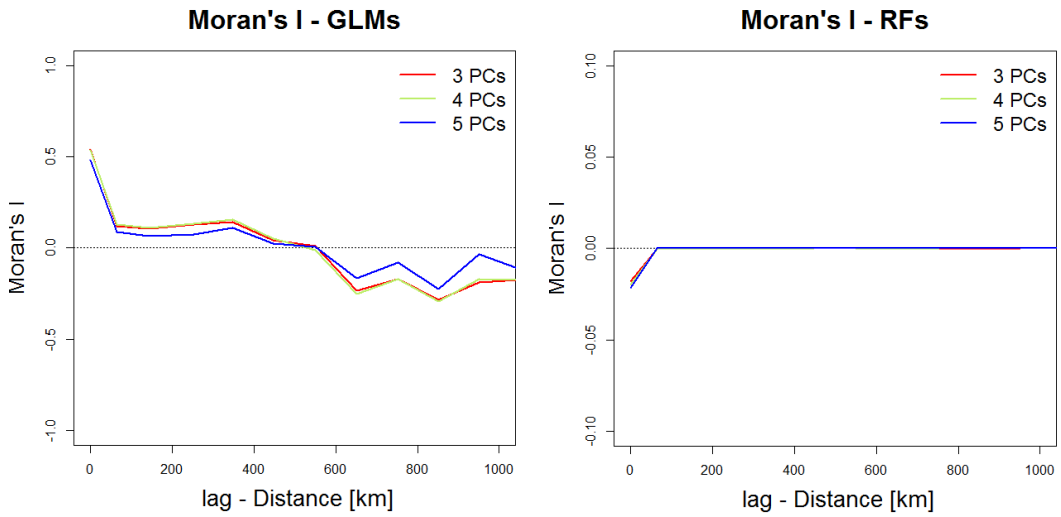

Figure 5: Moran's I (Test on Spatial Autocorrelation) for GLM and RF-model

The results from the cross-validation shown in Table 3 identify the GLM with three PCs and the Random Forest with five PCs as the superior models within their group. RMSE values quantify the average difference between projected and observed values. The higher the RMSE, the higher will be the absolute mistakes regarding tree heights in our projections. The correlation coefficients indicate if the model gets the trend in our data correctly, which is equally important, as its final application should be a classification of species according to their rank against others. Figure 6 is a chart of the cross-validation's residuals, which should ideally be equally distributed around and as close to zero. Whereas the GLMs look almost identical, the superiority of RF5 can be seen in the visual analysis as well.

Table 3: Results from 16-fold cross-validation of different models

	PCs	RMSE	Pearson's p	Spearman's rho
GLM	3	0.04017	0.94854	0.85600
GLM	4	0.04178	0.94412	0.85423
GLM	5	0.04383	0.93911	0.86782
RF	3	0.06315	0.89586	0.66857
RF	4	0.06714	0.88477	0.70915
RF	5	0.04651	0.94579	0.85460

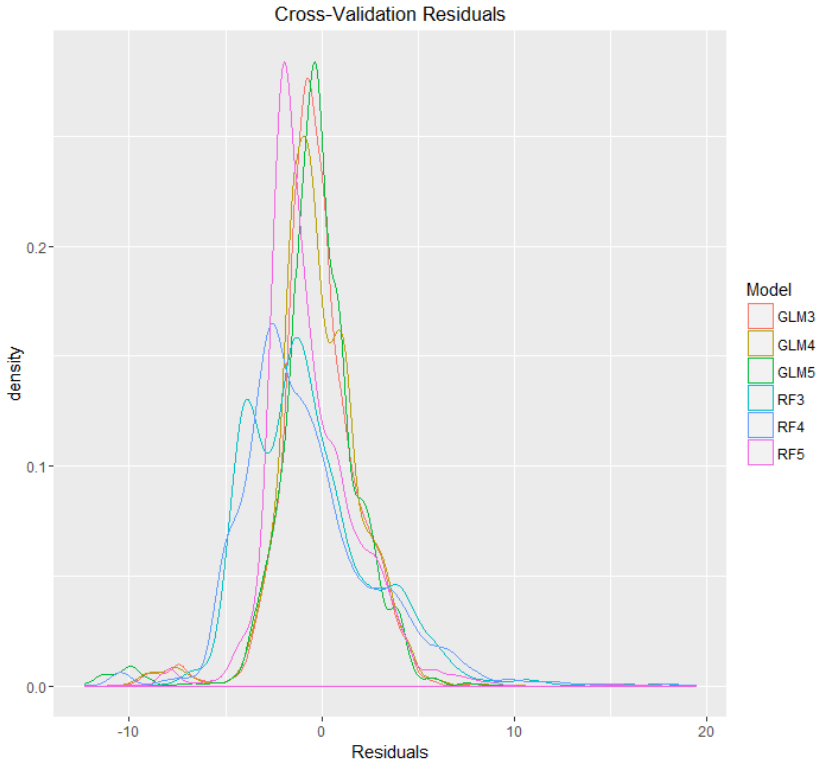

Figure 6: Residual Analysis from Cross-Validation Values

The predictor's importance values for RF5 in Table 4 show that "age" is clearly the most important distinguishing factor, but it is followed by PC4 with more than a seventh of the predictive capacity of "age". Mapping the distribution of PC4 values across Europe (Figure 7) revealed a strong spatial concentration of high values in the Mediterranean climate. According to the PCA rotation matrix (Table 2), the factor which dominates PC4 is "summer heat moisture" (0.78), which makes sense, as dryness is a major concern for tree growth in the dry Mediterranean climate.

Table 4: Importance measures of RF5

Predictor	Importance Measures
AGE	81791.209
SITE_PC1	3819.490
SITE_PC2	5753.564
SITE_PC3	6151.757
SITE_PC4	11739.593
SITE_PC5	7320.985

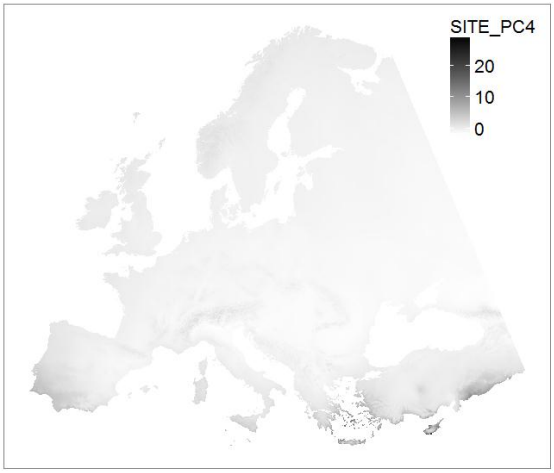

Figure 7: Spatial distribution of PC4

Even though it reduced the overall prediction accuracy of the model a little bit (Table 3), I still included PC4 into my GLM, because firstly, the reduction of precision did not occur before the third decimal place. Secondly, excluding this factor would have compromised the model's validity of all of south Europe. PC5, however, was waived, because I did not want to sacrifice more accuracy for a factor that also accounted for dryness values (AHM, SHM, MDMP; Table 2) and only accounted for 2% of explained variance (Figure 4).

The results of the multiple regression analysis for GLM4 are presented in Table 5. Of course, the trees' age explains the main share of height variance with more than 92%. However, as I predicted tree height for a set age of 30 years, this factor is taken out of the equation when the performances of different provenances are compared.

Table 5: Multiple regression analysis of GLM4

Predictors	Estimate	Std. Error	t-Value	p-Value	Expl. Dev. in %
Intercept	-4.808177	0.180029	-26.708	< 2e-16	
Age	0.713159	0.003696	196.949	< 2e-16	92.6577
PC1	-1.009865	0.033166	-15.285	< 2e-16	1.0287
$PC1^2$	-0.198455	0.033166	-5.984	2.47e-09	0.0195
PC2	0.202117	0.091656	2.205	0.0275	0.2451
$PC2^2$	0.079813	0.017999	4.434	9.60e-06	0.0698
PC3	-0.241932	0.106980	-2.261	0.0238	0.0253
$PC3^2$	-0.711040	0.081921	-8.680	<2e-16	0.1605
PC4	-0.301703	0.166841	-1.808	0.0707	0.0227
$PC4^2$	-0.830603	0.294980	-2.816	0.0049	0.0168
Whole Model					94.2067

3.3 Height Prediction and Validation

After identifying GLM4 and RF5 as the superior models, they were applied to the actual purpose of this study – predicting the best performing provenances and their expected height compared to each other, and to the current Douglas-Fir growth habits. Both models were fed with PCA transformed climate data from different emission scenarios and timeframes including the reference basis of climate data from 1961 to 1990 and subsequently, the validation tests described in “2.6. Height Prediction and Validation” were implemented.

As described in more detail in chapter 2.6., VAR is supposed to test if following the models’ recommendations for planting Douglas-Fir seeds would result in an increase in forest productivity at a given site, whereas VAB investigates if the model overestimates its predictions. VAR histograms of GLM and RF are presented in Figure 8 and VAB histograms in Figure 9.

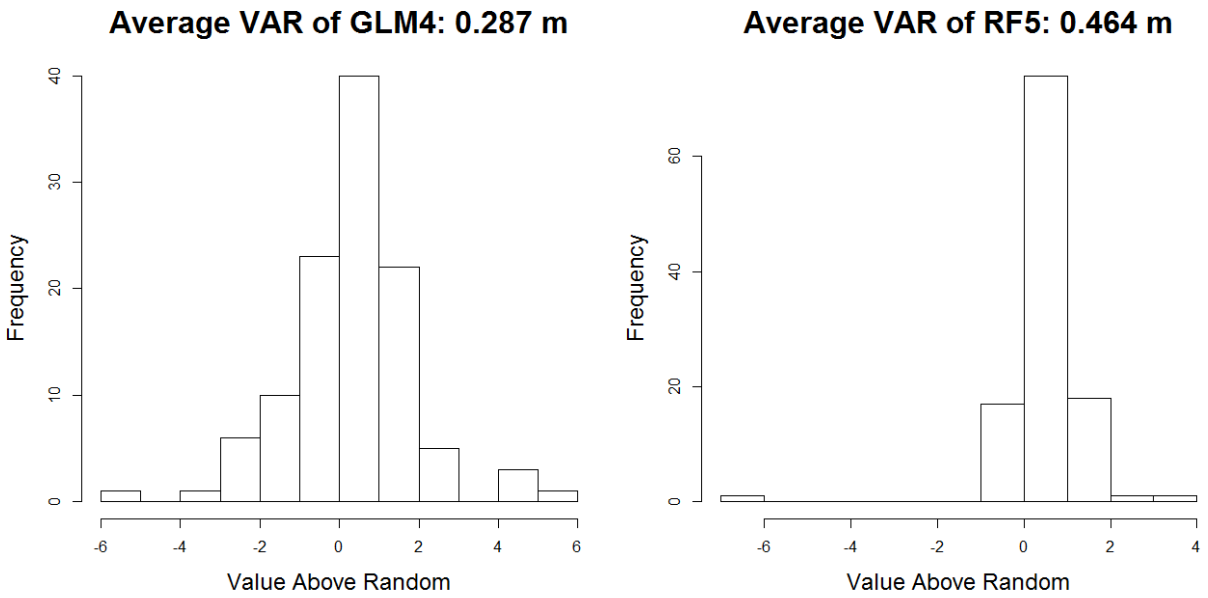

Figure 8: VAR-values in comparison between GLM and RF

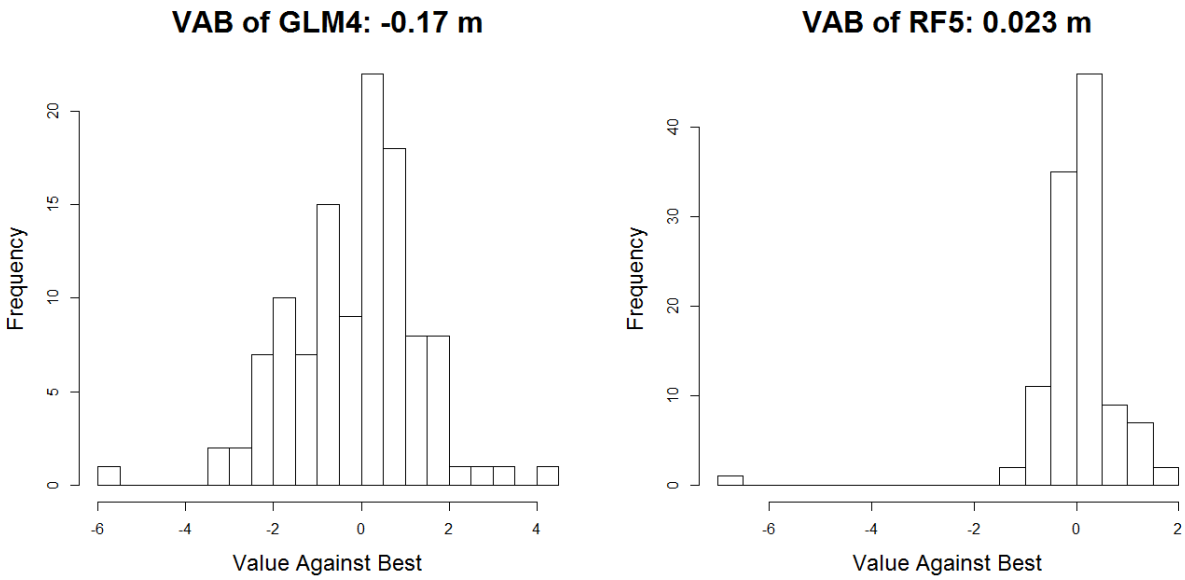

Figure 9: VAB-values in comparison between GLM and RF

According to the VAR test, choosing the models' provenance selection over a random pick of seed material increases the average height by a bit more than 28 cm for the GLM and almost half a meter for the RF. Moreover, the RF's larger value does not seem to be due to overestimation, because its VAB is with 2,3 cm very close to zero. The VAB of the GLM model, on the other hand, is with minus 17 cm significantly below zero, even though values seem to be evenly distributed around zero.

Table 6 and Table 7 show how well both models identified the best performing provenance. Correct classifications are marked bold diagonally across the table. At first glance, each models' match rate seems to be quite unreliable. Some general observations are that RF provides better results than GLM, and provenances with larger sample sizes are identified with a higher accuracy than small samples. In order to include the classification procedure into the validation, I grouped provenances into five subgroups (second row). In doing so, I revealed provenance distinctions, which were made according to administrative boundaries rather than climatic clines. A mismatch such as for C_WA in Table 6 (only 12 in C_WA, but 30 overall in the coastal type) is less severe than the misclassification in row 9 of INs as C_WA, because the only concrete difference between the former three regions are the state borders of Washington state. When assessing the models' reliability for a certain provenance, it is worth looking also at the values in the overarching category.

Table 6: Correct classification of optimal provenance made by GLM

	Predicted												
	C			CC		DC		I		CA			
Observed	C_BC	C_OR	C_WA	CC_OR	CC_WA	DC_OR	DC_WA	IC	IN	LE_CA	HE_CA	n	Match rate %
Coastal													
C_BC	**1**	1	4	0	0	0	0	1	0	0	1	8	13
C_OR	1	**10**	2	1	0	0	1	0	0	0	1	16	62
C_WA	1	17	**12**	2	0	2	5	0	1	2	0	41	30
Cascades													
CC_OR	0	1	0	**1**	0	0	1	0	0	1	0	3	33
CC_WA	o	2	5	0	**0**	0	0	1	0	0	0	8	0
Dry Coast													
DC_OR	1	1	1	1	1	**1**	0	0	2	1	1	8	13
DC_WA	1	6	1	1	0	1	**3**	1	0	1	0	19	16
Interiors													
IC	0	1	0	0	0	0	0	**1**	0	0	0	2	50
IN	0	0	3	0	0	0	0	0	**1**	0	0	5	20
California													
LE_CA	0	0	0	0	0	0	1	0	0	**1**	0	**2**	50

Table 7: Correct classification of optimal provenance made by RF

	Predicted											
	C			CC		DC		I		CA		
Observed	C_BC	C_OR	C_WA	CC_OR	CC_WA	DC_OR	DC_WA	IN	IS	LE_CA	n	Match rate %
Coastal												
C_BC	**2**	2	0	2	1	1	0	0	0	0	8	25
C_OR	0	**9**	3	1	1	0	2	0	0	0	16	56
C_WA	0	5	**28**	2	2	0	2	0	1	1	41	58
Cascades												
CC_OR	0	0	0	**2**	0	0	1	0	0	0	3	66
CC_WA	1	1	3	0	**1**	1	0	0	0	1	8	12
Dry Coast												
DC_OR	0	0	2	0	0	**2**	2	0	2	0	8	25
DC_WA	0	3	2	2	0	0	**12**	0	0	0	19	63
Interiors												
IC	0	0	0	0	0	2	0	0	0	0	2	0
IN	0	0	1	0	0	0	0	**4**	0	0	5	80
California												
LE_CA	0	0	0	0	0	0	0	0	0	**2**	**2**	100

3.4 Application

The following test is already a trial of its potential application as a forest management tool. Furthermore, it is an approach to verify my initial research objective. At the outset there was the question of whether the optimal composition of Douglas-Fir provenances in Europe might diverge from its current state due to climate change. The VAC-test quantifies this change in species composition and locates it. Only the re-

sults from the 2020's and the 2080's are going to be shown. These two scenarios combined make underlying trends clearly visible, whereas the 2050's values usually only represent a transition scenario.

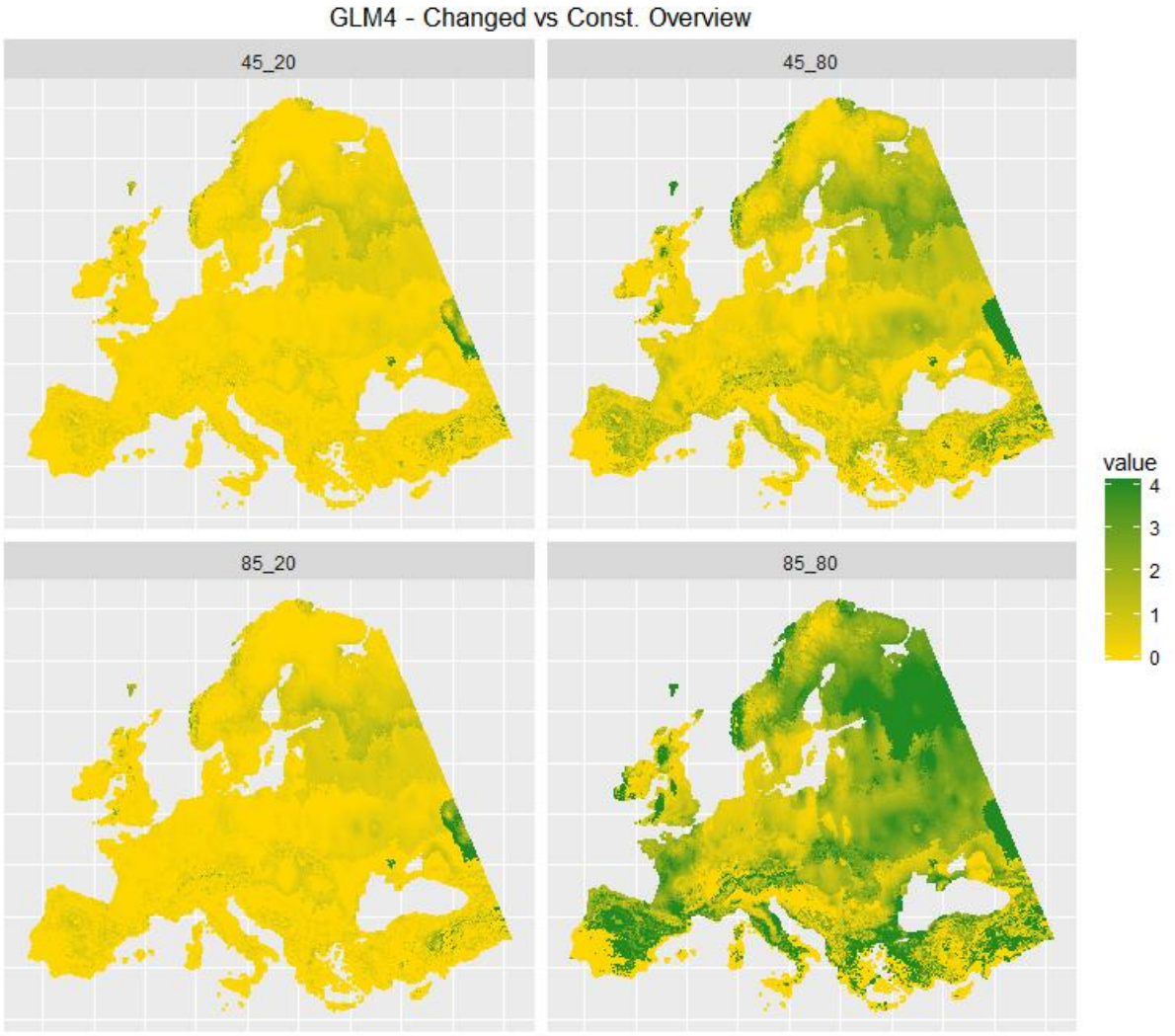

Figure 10: Visualization of VACs from GLM predictions

The graphs in Figure 10 show GLM-predicted VAC values across Europe. The development from left to right shows the expected long term trend, whereas the comparison between upper and lower graphs display the influence of the intensity of climate change on the results, which is remarkable. Whereas the difference might not yet be visible in the 2020, in the course of just 60 years a change in planting patterns would result in a significant height increase. Especially in high latitudes, Spain, Italy and the southern Balkans, precautionary guided planting would make a significant difference. Figure 11 show a comparison between current and future optimal tree heights, if management policies were altered towards the optimal provenances. The

largest gains compared to the current state could therefore be achieved in the North-East of Europe, the Scandinavian Coast, southern Spain and Greece, as well as in the high altitudes of central Europe.

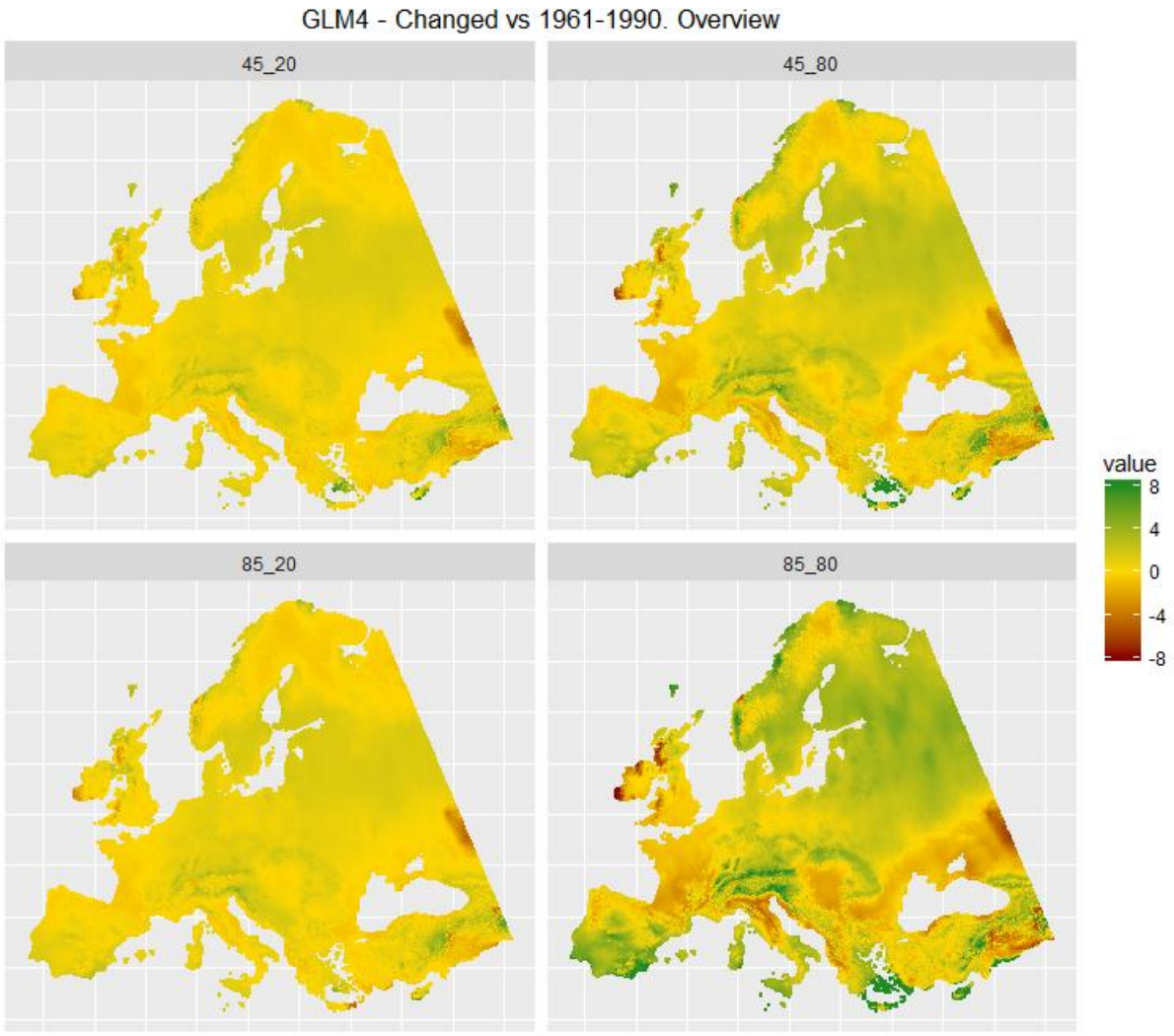

Figure 11: Visualization of GLM-recommended provenances' growth vs. current optimal growth heights

Red areas indicate regions in which even with optimal management implementations Douglas-Fir is going to lose habitat as a result of climate change. An interesting observation in this regard are the conditions in coastal west Europe and especially in some parts of the UK, where according to Figure 10 choosing the "right" mixture of provenances would make a large difference, but where productivity is going to decrease either way.

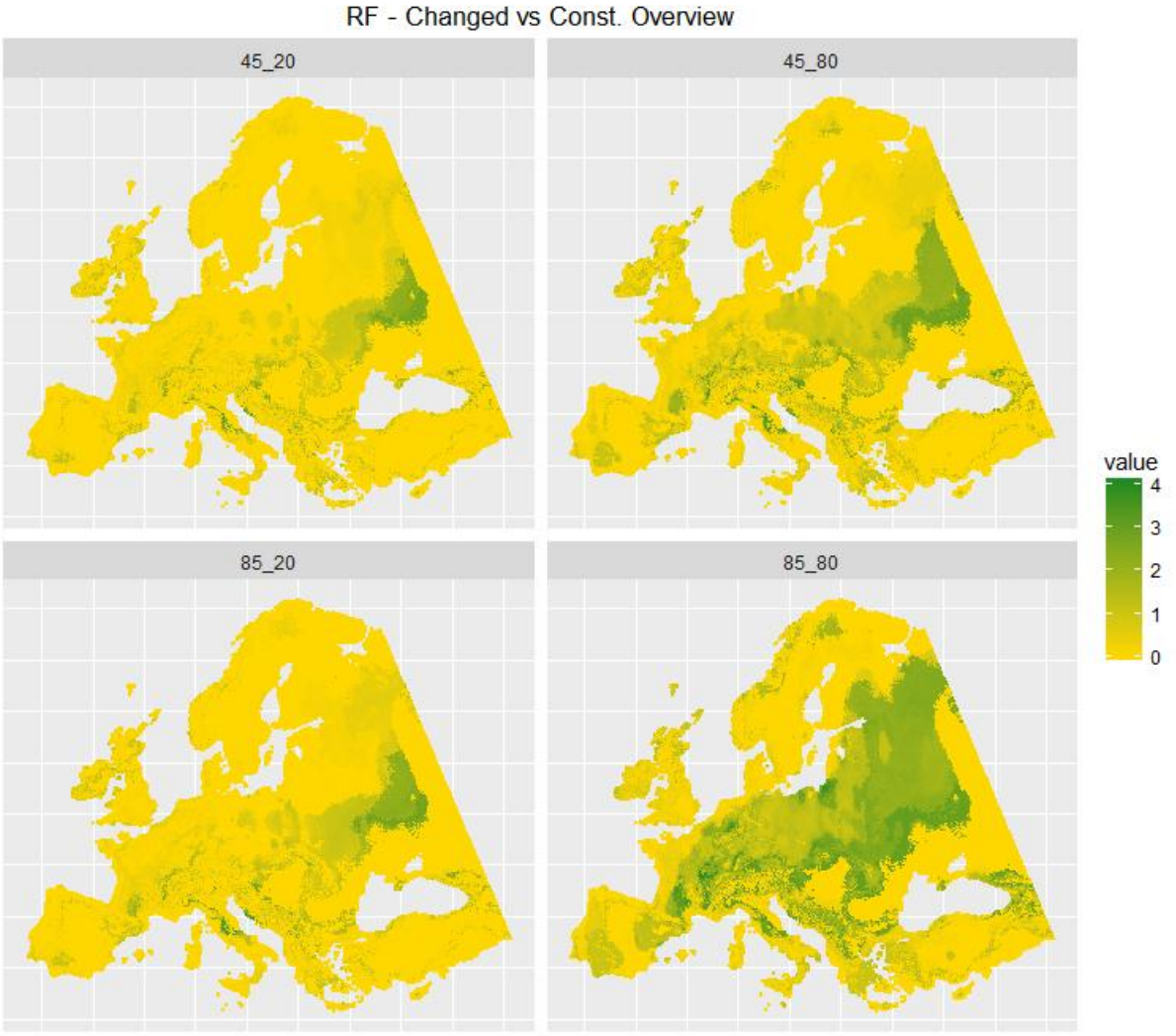

Figure 12: Visualization of VACs from RF predictions

Also according to the RF model, a change in planting policies would have exclusively positive outcomes. The distribution of benefits, however, is different from the projections made with GLMs. Firstly, the absolute values are less optimistic. Even though all predictions of the RF are positive, the main share lies within a height gain of 10 cm. (Figure 14). Therefore, with between 20 cm (4.5/2020) and 90 cm (8.5/2080), the average height gain predictions are also much smaller than those from the GLM, which lie between 55 cm and 154 cm (even though except for comparing models these average height gains are of limited significance, because the average is taken over the whole and very diverse European climate). A look at VAC maps in Figure 12 shows that the RF's optimistic predictions are strong punctual improvements in central and southern Europe, besides the extensive area in the North-East of Europe, which the GLM also identifies as a region, which yields a high potential for Douglas-Fir.

Concerning the overall suitability of Douglas-Fir the two models coincides in the two major observations (Figure 11, Figure 13). Both identify the Atlantic coast, England, and Ireland as areas, which are likely to become increasingly unsuitable for growing Douglas-Fir. On the other hand, large areas in the North-West have a tremendous potential for enhanced Douglas-Fir growth in near future. Nevertheless, there are also some regions where the RF's overall estimation of suitability of Douglas-Fir differs from the findings of the first model. Apart from a generally more positive height expectation in east Europe, the projected height gain for North-East Europe is much stronger than in the GLM data. Another interesting observation is how fast climate change will drive species northwards. For 2080, there is a massive difference between RCP4.5 and RCP8.5 regarding the northbound expansion of optimal height gain towards the Baltic, North Russia and Finland, which is not that striking just 60 years earlier.

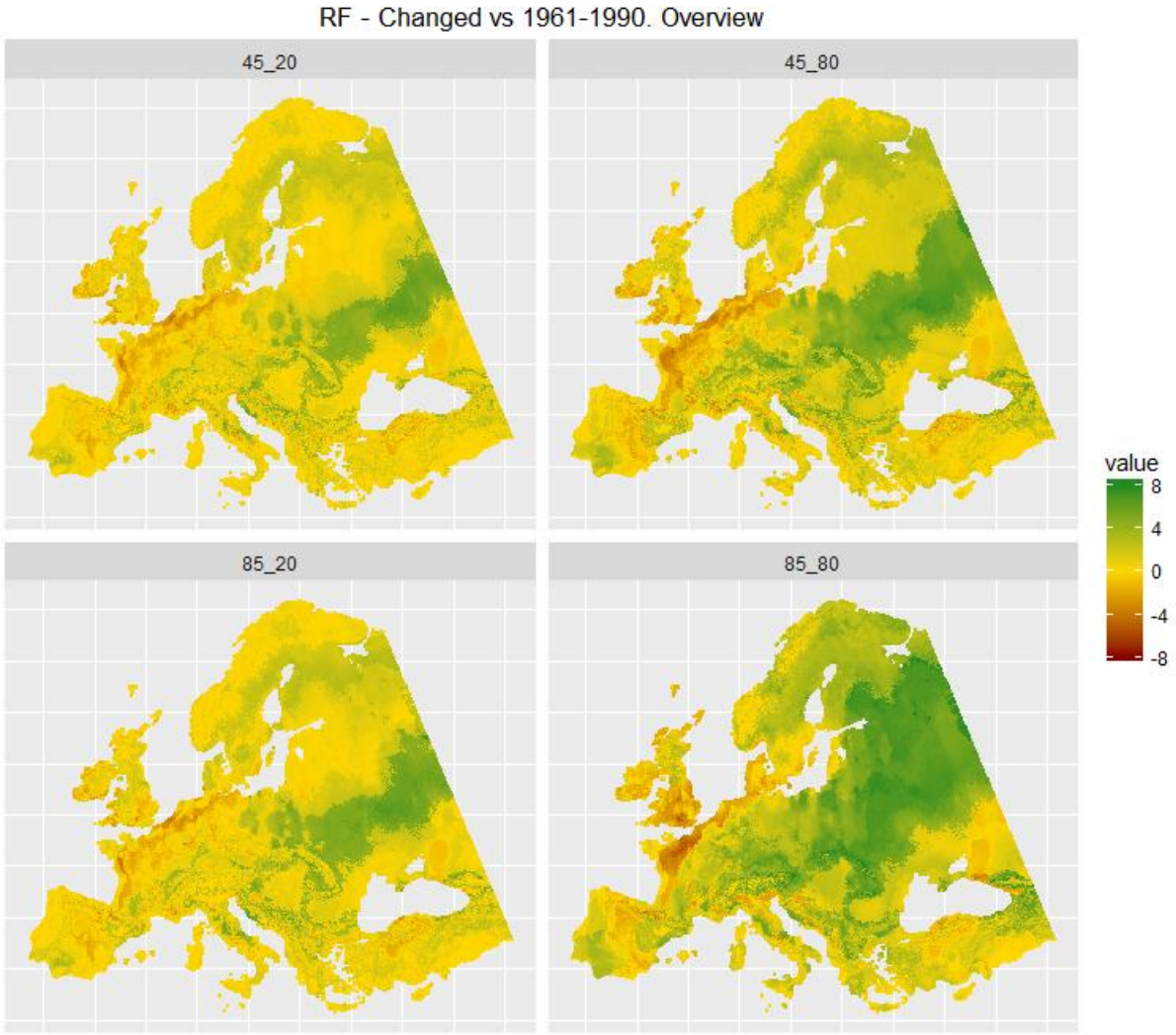

Figure 13: Visualization of RF-recommended provenances' growth vs. current optimal growth heights

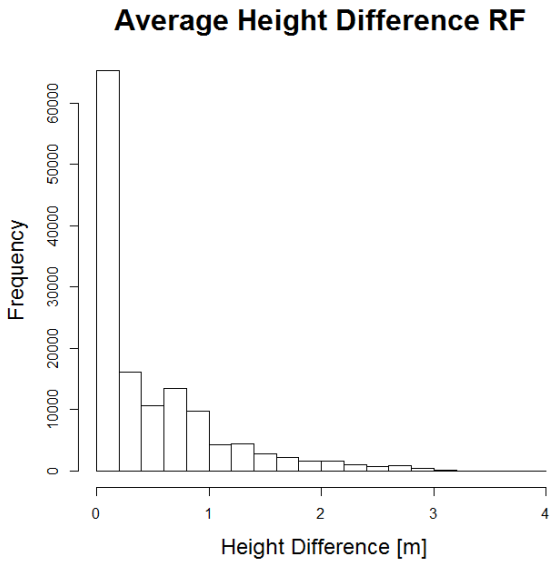

Figure 14: Histogram of VAC-Values form RF predictions

While the previous graphs have visualized the prosperity of the Douglas-Fir as a species represented by its best subpopulations, the figures on the following pages are supposed to unravel the prospects for individual provenances and illustrate their ranking. Figure 15 we can see histograms of models extrapolations for the current performance of each provenance. The black line indicates the average tree height, so that not only comparisons between provenances, but their height against the mean can be estimated. In compliance with past forestry research and Douglas-Fir management, both models correctly locate the interior types substantially below average and identify the coastal and dry-coastal populations as the well performing types. The graphs not only reveal insights about provenances, but about the underlying models as well. The RF projection is a much more precise in its ranking, whereas for some provenances such as high (HE_CA) and low elevation California (LE_CA) the GLM projection expects the whole range from almost zero to 20 m to appear on the European continent. As these overview plots only depict the frequency of expected data realization, Figure 16 and Figure 17 provide the corresponding spatial analysis.

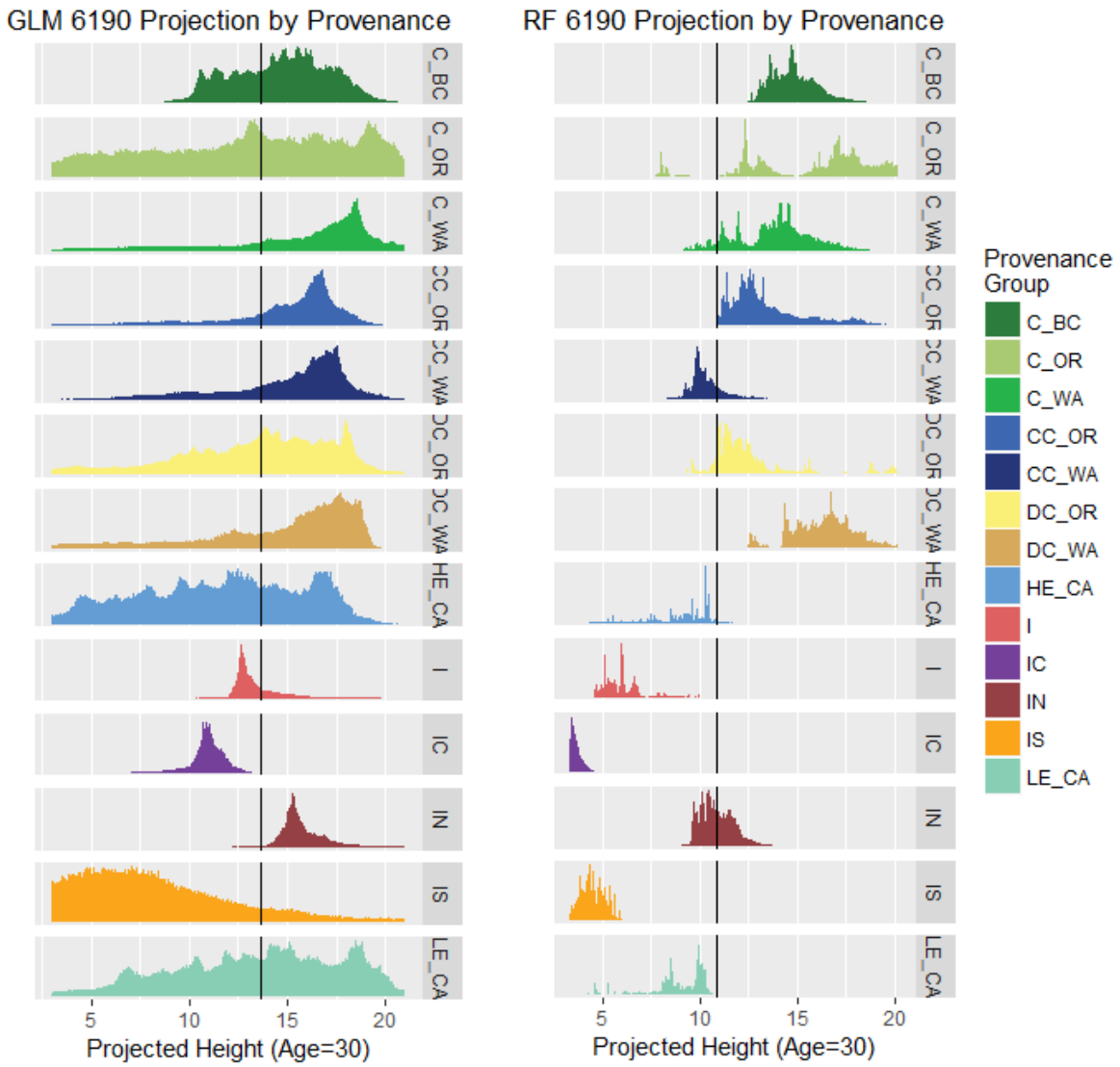

Figure 15: Distribution of response values by provenance for reference climate condition (1961 - 1990)

Here (Figure 16) we can see the reason for the widely stretched data sets in the GLMs. Expectations for HE_CA and LE_CA, for example, cover the whole range of height realizations in almost equal shares across the continent. The long tails towards low values in all histograms of coastal types except C_BC (Figure 15) originate from exclusively bad growth expectations in Scandinavia and the Mediterranean. IN, I and C_BC seem to be a good choice for plantations in southern Europe, where all other provenances have severe problems. Apart from that, differences between all provenances from C_OR to DC_WA appear to be minor. A questionable property of the extrapolations is the occurrence of extreme gradients in transition areas, where height expectations are expected to drop or increase by 15 m within few kilometers. The RF model results show a more outbalanced extrapolation within each group and

stronger distinctions between them. C_OR and DC_WA are the clear front runner with adaptation problems in extreme northern and southern locations. C_BC and C_WA follow with a more evenly distributed above-average growth height, while C_BC performs better in cold regions. DC_OR and CC_OR appear to grow especially well in the plains of Italy, the Balkans and East Europe. All other provenances seem to be uninteresting for successful forestry purposes with IC, IS and I lagging particularly far behind.

Looking at predictions for 2020 and 2080 (Figure 18, Figure 19) confirms observations from VAC and comparison against current height analyses. Optimal growing habitats of all provenances shift northwards, which occurs a lot stronger under RCP8.5 than RCP4.5. In order to fit future projections on two pages, I left out the “Interior South” map, since these trees were far from catching up with any other provenance at any time or location. I further chose the RCP8.5 emission scenario, because trends were similar, but stronger and therefore more discernible than in RCP4.5.

According to GLM projections, large parts of previously unsuitable forest land in Scandinavia are going to become habitable for several Douglas-Fir provenances. One could say C_BC keeps it status as "universal" population that grows well everywhere in west Europe. There are local differences between the otherwise similarly top performing types C_WA, CC_OR, and the dry-coast provenances. C_OR grows especially well in central Europe, DC_OR a bit further south around the Alps and down the coast of the Balkan peninsula, where also LE_CA seems to be a good alternative. Trees from Washington are predicted to grow pretty well everywhere except for the south. The bad growth characteristics of the otherwise superior provenances in the dry climate of the Mediterranean increases and expands. Here, IC and IN would be the best choice.

The RF-projections are, once again, generally more consistent. Even though all populations experience the same northbound shift in their habitat range, this trend is most obvious for CC_OR; DC_OR, and DC_WA. According to Figure 19, the latter develops an almost universal applicability combined with exceptional height reaching more than 20 m at age 30. DC_OR shows a similar development, except for worse values in oceanic West Europe and Scandinavia, where especially on the lee side of the Skagen mountain range C_OR is predicted to grow well. All other provenances lag considerably behind these three dominant subspecies.

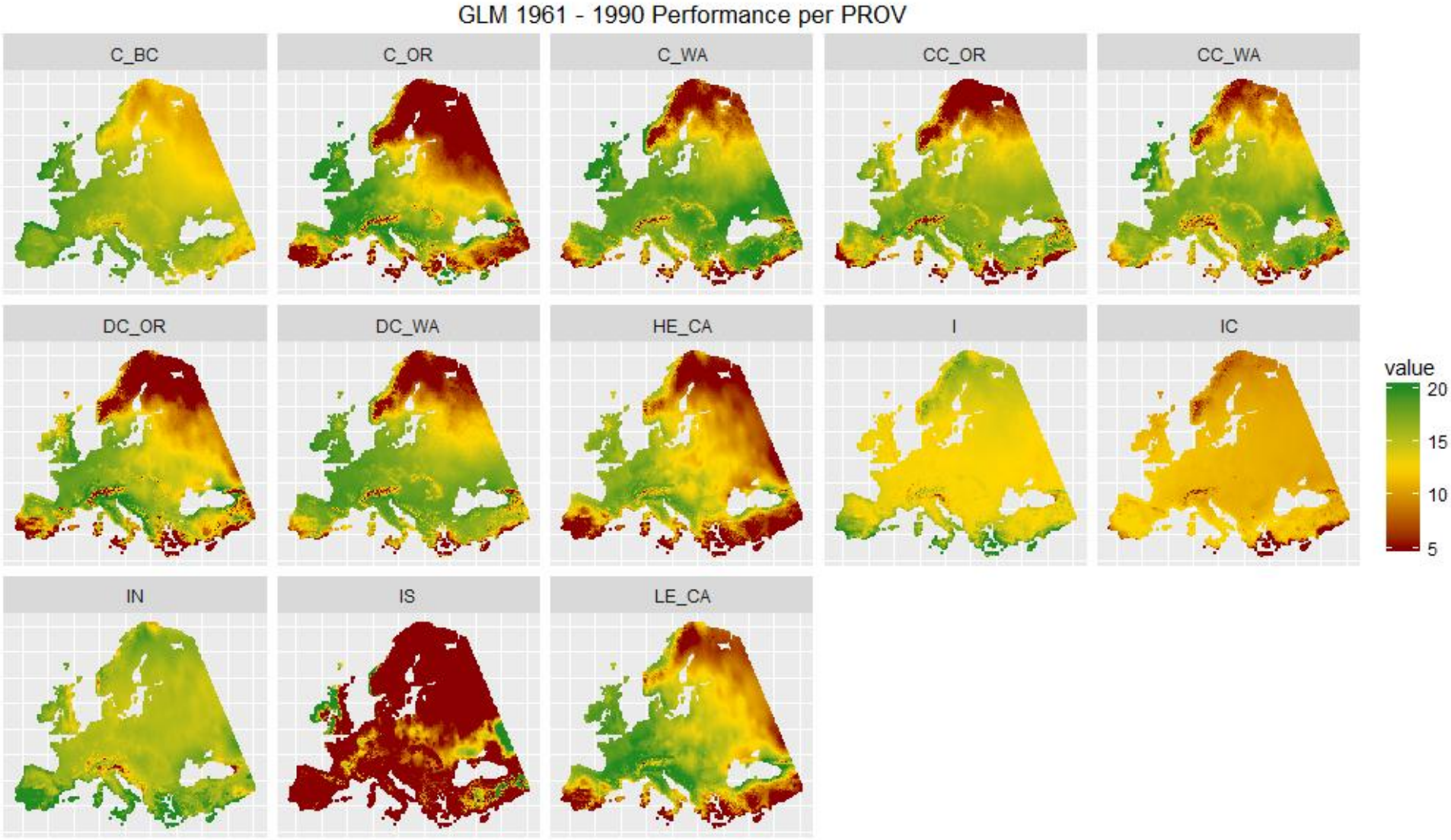

Figure 16: 1961-1990 extrapolation by provenance made with GLMs

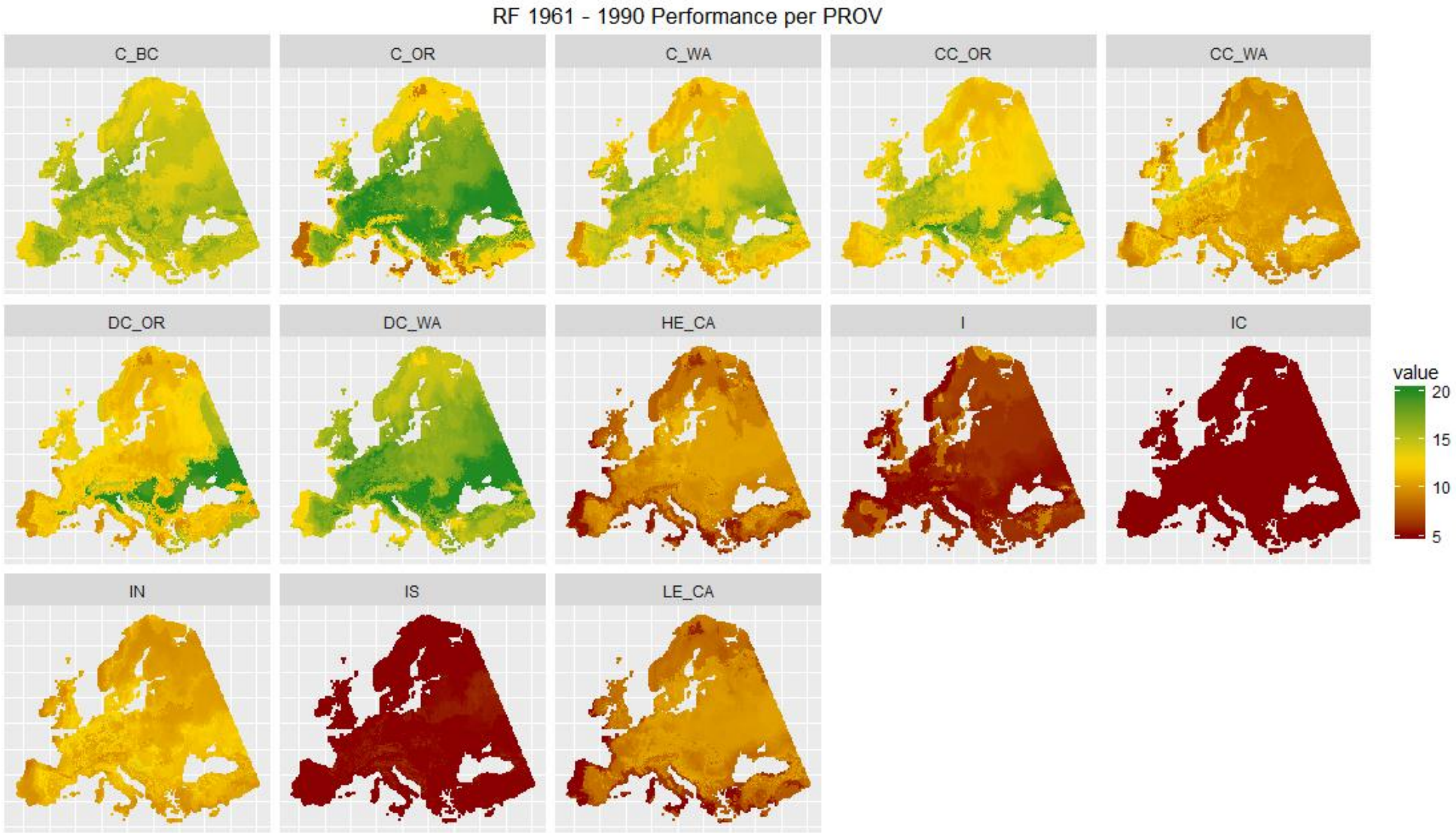

Figure 17: 1961-1990 extrapolation by provenance made with Random Forests

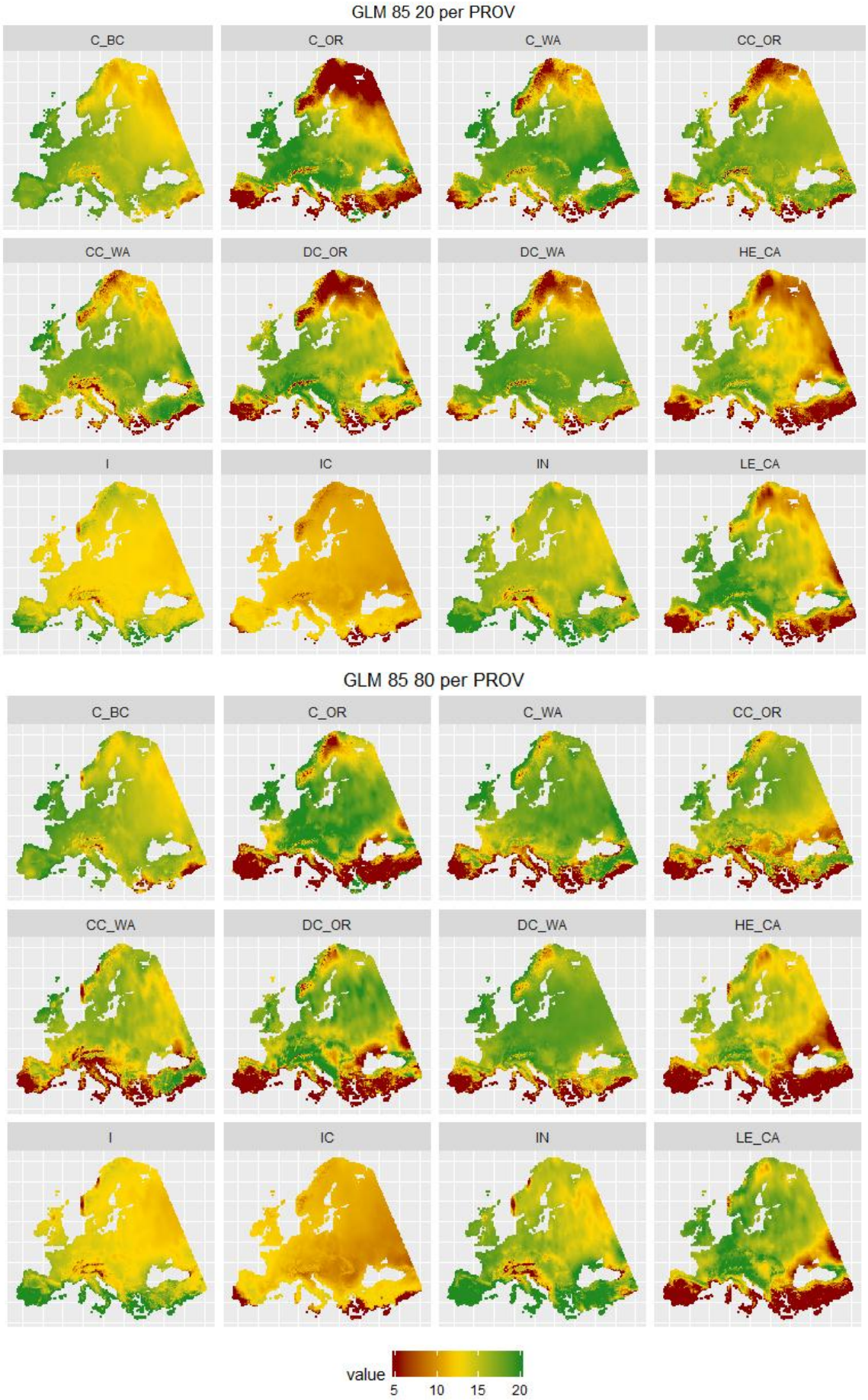

Figure 18: Future GLM predictions for RCP 8.5 by provenance

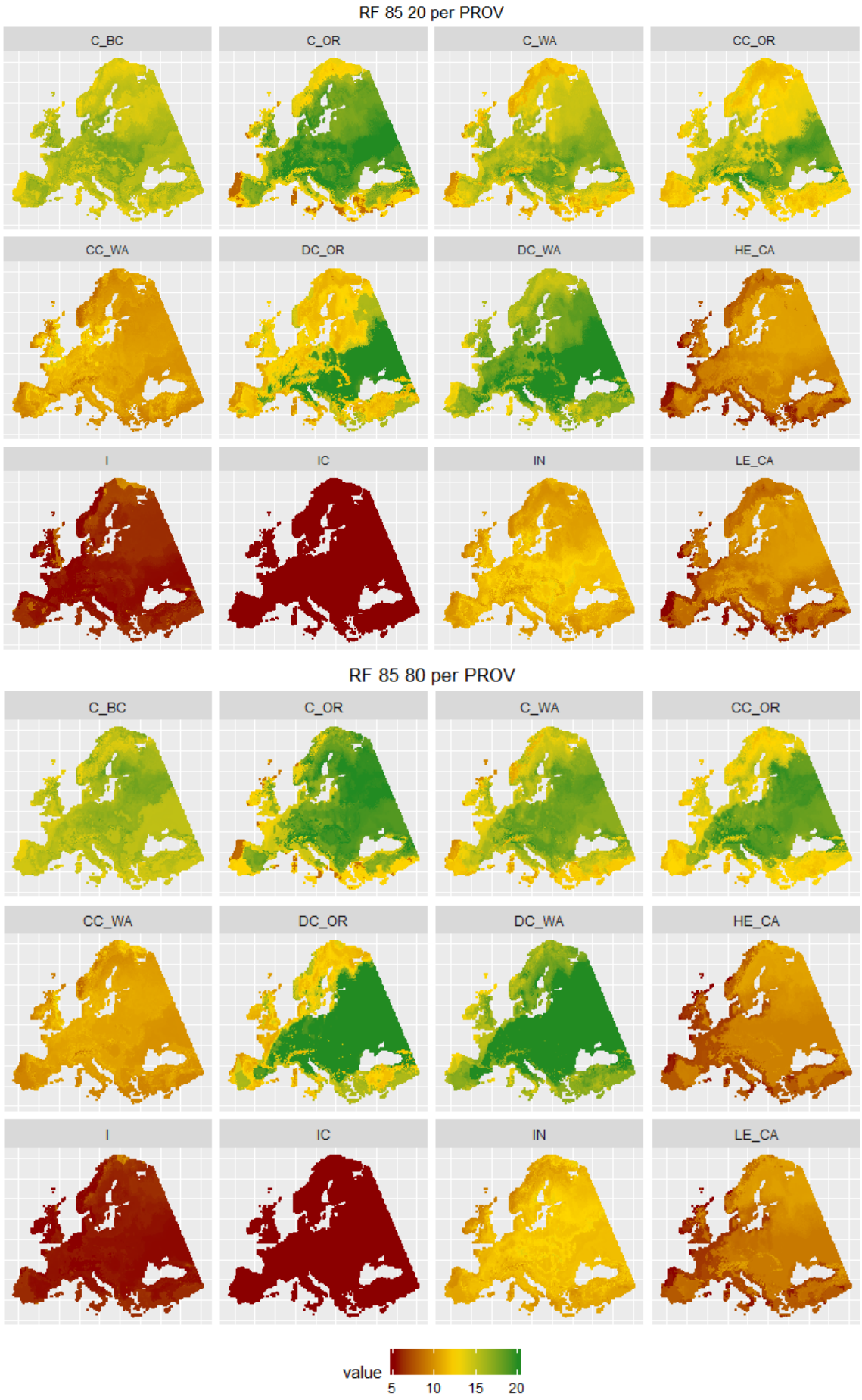

Figure 19: Future RF predictions for RCP 8.5 by provenance

4 Discussion

4.1 Findings of my Height Predictions

In this study, I used GLMs and a Random Forests model to predict future height growth of different Douglas-Fir provenances in Europe. The sometimes glaring differences between populations, which are projected to intensify in the future, emphasize the importance of seed transfer origin for forestry management decisions in the face of climate change. Even though results differed depending on the model applied, there were some observations, which were consistent in both approaches.

The northbound shift of optimal habitats has been observed by many other researchers and although using a different approach, the study by ISAAC-RENTON ET AL. (2014) projects the same results for Douglas-Fir. The rough estimation of a 300 to 500 km shift per century made by DAVIS & SHAW (2001) seems approximately right, when the middle of the green band between the unsuitable habitats in southern and northern Europe is interpreted as optimal growing range. Consequently, new areas for enhanced Douglas-Fir plantation are likely to emerge in the north-east of Europe. Regardless of which method is applied, coastal Douglas-Fir types are unlikely to be widely outperformed by interior provenances and continue to be the best choices for the improving planting sites in the North of Europe. The success of dry-coast Douglas-Fir, which has also been observed by ISAAC-RENTON ET AL. (2014) could be interpreted as a genetic response to a generally dryer climate in the future. It is further in line with the recommendation by MONTWÉ ET AL. (2015) to use drought-resilient species, as well as the findings by ST. CLAIR & HOWE (2007) to mix local populations with more southern provenances. If these changes are being considered, there is reason to believe that given proactive and well-guided management decisions, Douglas-Fir will stay and might become an even more important species for European forestry, especially in north Europe.

4.2 Model Assessment

Since the outcomes of the two models are quite different, it is necessary to ask which model is more reliable. When put to the test of correctly identifying the ranking of provenances in general (Table 6, Table 7) RF is a lot more accurate than its GLM equivalent. Regarding the VAB (Figure 9) as quantitative evaluation of the first three

choices, RF does better as well. Therefore, it seems obvious that in terms of reproduction of test results and classification of provenances, the RF clearly outperforms GLM. However, VAB and Confusion tables actually just test model fit. Even though the advantage in the two finally chosen models (GLM4 and RF5) does not become noticeable before the third decimal place, especially Figure 6 gives reason to believe that GLMs are generally the preferable tool for predicting to future and unknown climate. In the cross-validation, GLMs proved to be significantly more accurate than RFs. This ambiguity between classification accordance and prediction capacity might be explained by inherent characteristics of the two model approaches.

RF creates thousands of decision trees with the data at hand. The model increases accuracy by re-evaluating random data samples in many trees and produces estimates of extremely good fit. When confronted with independent data, the RF is constrained by the range of the bottom level of its trees. A new and very extreme point being estimated by the model can only go consistently right or left in the decision tree, but the limit is the tree's edges. Therefore, RFs are very robust against outliers, which might be good for model fit (see VAB and Confusion-Tables), but could result in overly conservative predictions. What we can see in Figure 17 and Figure 19 is that all provenances stick to a certain range and there is not much variety in terms of time and space within each provenance. While this might as well be the case for future climate, an alternative explanation could be that because I built a RF model for each provenance and RFs tend to be conservative predictors, each future climate realization will result in a height value similar to those used in the creation of the model (the end value of the multiple trees' nodes). On the contrary, GLMs are second-order polynomials, which can never reach the same level of accuracy as a 1000-fold classification tool, when trying to fit to a large number of realizations during model creation, therefore worse values in data identification. They are also susceptible to outliers, because their coefficients are just multiplied with predictor values, which can drive the response to extreme values. However, this feature might also make it a better predictor for climate change projections, where predictor (and response) variable might indeed attain extreme values. I was first sceptical about differences in height projections of 15 m within few kilometers, but the scale of local change coincides with the projections made by WANG ET AL. (2010), who also found height differences in short distances of about 8 m at age for 20 year old trees.

Whereas, it seems like RF is the better choice if we want to get an accurate estimate of the provenances' performance compared to each other (or the ranking and best choice of provenances), I conclude that the GLMs future projections are more reliable. One should consider that the model does not need to always identify the first choice correctly. First and second rank might just be distinguished by a small and, given the difference in wood quantity, insignificant height difference. One has to account for the option that the model correctly identifies the best *n* provenances, but in a different order. By testing only if the best predicted provenance matches the best observed provenances, we would over-penalize models that capture trends correctly, but fail to get the exact order, which might even be not that important. Forest owners rarely plant one homogeneous genotype, but rather a mixture of populations. Furthermore, as we cannot know the future, our proxy-method for unknown climate is cross-validation, where GLMs perform better. In this context, it is worth asking what the main purpose of this project's modelling attempts is and which model feature is more important in this regards. Unfortunately, the main application is a combination of both: to *find the best* (RF) selection of provenances for *future* (GLM) climates. I believe that the better accuracy of future predictions as more important. The lack of accuracy in rank estimations is a trade-off of GLMs that could be addressed, for example, by planting a larger number of provenances.

4.3 Model Boundaries

Apart from the inherent advantages and disadvantages of model types, there are some general limitations to statements made in this thesis. The modelling approach in this bachelor thesis is based solemnly on the factors age and climate. Other crucial factors for tree growth, such as soil conditions, exposition, pests and fungi, or competition with other trees, are not included. Worth mentioning in this regard is *Rhabdocline pseudotsungae*, a pathogenic fungus causing needle cast in Douglas-Fir. According to KLEINSCHMIT & BASTIEN (1992), coastal provenances are less vulnerable to this pest, northern interiors show great variety and southern interior are the least resistant type, which has been another reason for the planting preference of coastal Dougla-Fir in Europe (EILMANN ET AL. 2013). Such aspects must be assessed for each individual case by people with knowledge about the respective area. In this context, climate models such as this one can work as a complementary guidance, yet not as the only decisive factor.

4.4 Data Accuracy

First and foremost, a potential source of fallacies is the dataset itself. The data stems from a large variety of provenance trials, conducted by various institutions and also at different points in time, which compromises consistency in terms of standardized procedures of measurement or treatment during growth. While some obviously differing data sets could be identified and excluded, double checking and adjusting all data sources would have extended the scope of this bachelor thesis. On the other hand, the positive side of this variety of data sources is a decent coverage of provenance and planting climates. The set derives its power from the number of sites and observations. Additionally, my promising validation results (VAB; Confusion Matrices, and CV) give reason to believe that the overarching trends are being correctly identified.

Furthermore, in terms expectations towards actual height predictions, it is important to note that the largest factor of uncertainty is probably neither dataset nor model choice, but climate data. For some parts of Europe there were considerable differences between RCP4.5 and RCP8.5 projections. Keeping track of the development of climate change is hence the main requirement for making valid statements. Not to mention the level of uncertainty within the climate projections as well. Quantitative statements on the basis of my projection should therefore be made in due consideration of these confidence levels.

4.5 Height as Response Variable

There are several ways to measure ecological suitability of tree species for their environment. The most common indicators are height, diameter at breast height, survival rates or presence-absence data. According to ISAAC-RENTON ET AL. (2014) height had been used as response variable, because it was consistently reported. A further advantage of height is that it presages mortality and allows for a faster detection of maladaptation (GRAY ET AL. 2011). However, using tree height as indicator for ecological suitability can also be misleading. As described in more detail in chapter 1.4.1., some tree species invest some of their energy in stress resilience rather than height growth. Interior Douglas-Fir is known to be more drought (PHARIS & FERRELL 1966) and cold (REHFELDT 1977) resilient. Since we only include trees in our measurement which have survived the particularly vulnerable sapling period, our results might underestimate the environmental fitness of interior provenances. In other terms, because sur-

vival is not accounted for in our height data, the models might extrapolate optimistic growing expectations of growth-intensive types to regions, where they actually would not survive. Considering the expected increase in drought events in climate change scenarios, we might miss some crucial factors in this regard.

4.6 Further Research Prospects and Applications

Both methods, GLM and RF have revealed shortcomings in this study. Whereas GLMs appear to produce better future predictions, RFs are a lot more accurate in ranking provenances (Table 7). My research objective actually requires a model that combines both properties. Further research projects could try to merge the benefits of both approaches into one model. When provenance trial data from the 2020s will soon be available, it will also be possible to evaluate the prediction quality of different models directly instead of by using CV as a proxy.

Another obstacle of seed transfer under climate change stems from the attempt to make predictions for a certain point in time under gradually changing conditions. Making long-term planting decisions is tantamount to shooting at a moving target. GRAY ET AL. (2011) advise not to plan further ahead than 20 years. They argue that the uncertainty of climate projection increases with growing time scales, besides the fact that saplings are most vulnerable to get damaged because of adaptation lag. Hence, near climate projections should guide decisions rather than the conditions at harvest age. Apart from the time target issues, they call for a mixture of research approaches in order to compensate shortcomings of individual models with multiple information sources. Explicitly mentioned are remote sensing, bioclimatic modelling, and dendrochronology. I agree that a tree ring analysis of the sample trees might indeed clarify some open questions concerning the influence of drought hardiness (for example the provenances' response to the drought events in 2003 and 2011).

Of course, it would also be interesting to apply the same population specific spatial analysis to other tree species, on the one hand for commercial purposes, but also from a conservational perspective. A concept that often comes up in this context is *assisted migration*. The term refers to actively relocating species towards more suitable habitats, if they are unable to adapt quickly enough to their changing environment. The topic is highly controversial and judging the practice itself shall not be part of this thesis (for a comprehensive review see MCLACHLAN ET AL. (2007)). Still, it is worth

noticing that response functions can contribute to finding the best target location for endemic species that are driven out of their limited range, or abundant species experiencing a growing adaptation lag (GRAY ET AL. 2011).

5 Conclusion

In conclusion, my study shows that proactive management decisions, which consider future climate conditions as well as the genetic source of the planting material, might improve yields from Douglas-Fir in European forests substantially. Regarding the choice of provenances, differences between current and future recommendations are minor, but present.

Coastal provenances such as C_WA, C_BC, and especially C_OR will continue to be top performers for European forestry, even though their optimal planting range will considerably shift northwards, where increasingly productive areas might emerge in Scandinavia, the Baltics and north Russia. A trend towards dryness-resistant coastal types (such as DC_OR and DC_WA) in central and southern-eastern Europe could be read as the symbiosis between superior growth performance of coastal types and the need of future seed material to be more adapted to a generally dryer and warmer European climate. GLMs projections confirm the mentioned selection for central (DC_OR and C_OR) and northern (DC_WA and C_WA) Europe, but include LE_CA as a provenance worth considering for central Europe as well. The GLMs further predict IN and I to be able to cope best with the intensifying Mediterranean climate.

Given the revealing and fairly accurate results, Tests in this study have shown, that independent of the area of application, an application of targeted seed transfer based on response functions had exclusively positive outcomes compared to the status quo of provenance selection.

I believe that apart from Douglas-Fir in Europe this methodology could also be applied to analyse other species and locations. Even though it further yields potential to guide conservational efforts of transferring endangered species to climatically more suitable habitats, the availability of comprehensive provenance trial data might restrict its application to commercial forestry. Nevertheless, when set within a context of site-specific growing conditions, such as soils, microclimate, or biotic side factors, response functions can be a valuable additional assessment tool for planting policies in Europe.

6 Publication bibliography

Aitken, Sally N.; Yeaman, Sam; Holliday, Jason A.; Wang, Tongli; Curtis‐McLane, Sierra (2008): Adaptation, migration or extirpation: climate change outcomes for tree populations. In *Evolutionary Applications* 1 (1), pp. 95–111. DOI: 10.1111/j.1752-4571.2007.00013.x.

Allen, Craig D.; Macalady, Alison K.; Chenchouni, Haroun; Bachelet, Dominique; McDowell, Nate; Vennetier, Michel et al. (2010): A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. In *Adaptation of Forests and Forest Management to Changing ClimateSelected papers from the conference on "Adaptation of Forests and Forest Management to Changing Climate with Emphasis on Forest Health: A Review of Science, Policies and Practices", Umeå, Sweden, August 25-28, 2008* 259 (4), pp. 660–684. DOI: 10.1016/j.foreco.2009.09.001.

Araújo, Miguel B.; Cabeza, Mar; Thuiller, Wilfried; Hannah, Lee; Williams, Paul H. (2004): Would climate change drive species out of reserves? An assessment of existing reserve‐selection methods. In *Global change biology* 10 (9), pp. 1618–1626. DOI: 10.1111/j.1365-2486.2004.00828.x.

BMEL (2014): Der Wald in Deutschland. Ausgewählte Ergebnisse der dritten Bundeswaldinventur. Edited by Bundesministerium für Ernährung und Landwirtschaft. Berlin, Germany. Available online at https://www.bundeswaldinventur.de/fileadmin/SITE_MASTER/content/Dokumente/Downloads/BMEL_Wald_Broschuere.pdf, checked on 27.04.16.

Chakraborty, Debojyoti; Wang, Tongli; Andre, Konrad; Konnert, Monika; Lexer, Manfred J.; Matulla, Christoph; Schueler, Silvio (2015): Selecting Populations for Non-Analogous Climate Conditions Using Universal Response Functions: The Case of Douglas-Fir in Central Europe. In *PloS one* 10 (8), pp. e0136357. DOI: 10.1371/journal.pone.0136357.

Davis & Shaw (2001): Range shifts and adaptive responses to Quaternary climate change. In *Science (New York, N.Y.)* 292 (5517), pp. 673–679. DOI: 10.1126/science.292.5517.673.

Dormann, Carsten F. (2013): Parametrische Statistik: Verteilungen, maximum likelihood und GLM in R. Berlin, Heidelberg: Springer-Verlag.

Eilmann, Britta; Vries, Sven M.G. de; den Ouden, Jan; Mohren, Godefridus M.J.; Sauren, Pascal; Sass-Klaassen, Ute (2013): Origin matters! Difference in drought tolerance and productivity of coastal Douglas-fir (Pseudotsuga menziesii (Mirb.)) provenances. In *Forest Ecology and Management* 302, pp. 133–143. DOI: 10.1016/j.foreco.2013.03.031.

FAO (2009): State of the World's Forests 2009. Rome, Italy. Available online at ftp://ftp.fao.org/docrep/fao/011/i0350e/i0350e.pdf, checked on 4/27/2016.

Gray, Laura K.; Gylander, Tim; Mbogga, Michael S.; Chen, Pei-yu; Hamann, Andreas (2011): Assisted migration to address climate change. Recommendations for aspen reforestation in western Canada. In *Ecological Applications* 21 (5), pp. 1591–1603. DOI: 10.1890/10-1054.1.

Hamann, Andreas; Wang, Tongli; Spittlehouse, David L.; Murdock, Trevor Q. (2013): A Comprehensive, High-Resolution Database of Historical and Projected Climate Surfaces for Western North America. In *Bull. Amer. Meteor. Soc.* 94 (9), pp. 1307–1309. DOI: 10.1175/BAMS-D-12-00145.1.

Hamann & Wang (2006): Potential Effects of Climate Change on Ecosystem and Tree Species Distribution in British Columbia. In *Ecology* 87 (11), pp. 2773–2786.

Hermann & Lavender (1999): Douglas-fir planted forests. In *New Forests* 17 (1-3), pp. 53–70. DOI: 10.1023/A:1006581028080.

Howe, Glenn; Aitken, Sally N.; Neale, David B.; Jermstad, Kathleen D.; Wheeler, Nicholas C.; Chen, Tony H. H. (2003): From genotype to phenotype. Unraveling the complexities of cold adaptation in forest trees. In *Can. J. Bot.* 81 (12), pp. 1247–1266. DOI: 10.1139/b03-141.

Howe, Glenn; Jayawickrama, Keith; Cherry, Marilyn; Johnson, G. R.; Wheeler, Nicholas C. (2006): Breeding Douglas‐Fir: John Wiley & Sons, Inc (27). In *Plant Breeding Reviews*, 2006.

IPCC (2013b): Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. 2013: Long-term Climate Change: Projections, Commitments and Irreversibility. AR5. Cambridge, United Kingdom and New York, NY, USA.: Cambridge University Press.

IPCC (2013a): Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Observations: Atmosphere and Surface. AR5. Cambridge, United Kingdom and New York, NY, USA.: Cambridge University Press.

Isaac-Renton, Miriam (2013): Guiding Douglas-fir seed selection in Europe under changing climates:bioclimatic envelope model predictions versus growth observed inprovenance trials. Master Thesis. University of Alberta, Edmonton, Alberta. Department of Renewable Resources.

Isaac-Renton, Miriam G.; Roberts, David R.; Hamann, Andreas; Spiecker, Heinrich (2014): Douglas-fir plantations in Europe: a retrospective test of assisted migration to address climate change. In *Global change biology* 20 (8), pp. 2607–2617. DOI: 10.1111/gcb.12604.

Johann, Elisabeth (2004): Forest History in Europe. In Dietrich Werner (Ed.): Biological Resources and Migration. Berlin, Heidelberg: Springer, pp. 73–82.

Kleinschmit & Bastien (1992): IUFRO's role in Douglas-fir [PseudotsugaMenziesii (Mirb.) Franco] tree improvement. In *Silvae Genetica* (41), pp. 161–173. Available online at http://www.allgemeineforstundjagdzeitung.com/fileadmin/content/dokument/archiv/silvaegenetica/41_1992/41-3-161.pdf.

Ledig & Kitzmiller (1992): Genetic strategies for reforestation in the face of global climate change. In *Forest Ecology and Management* 50 (1-2), pp. 153–169. DOI: 10.1016/0378-1127(92)90321-Y.

Lindner, Marcus; Maroschek, Michael; Netherer, Sigrid; Kremer, Antoine; Barbati, Anna; Garcia-Gonzalo, Jordi et al. (2010): Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. Adaptation of Forests and Forest Management to Changing Climate. Selected papers from the conference on “Adaptation of Forests and Forest Management to Changing Climate with Emphasis on Forest Health: A Review of Science, Policies and Practices”, Umeå, Sweden, August 25-28, 2008. In *Forest Ecology and Management* 259 (4), pp. 698–709. DOI: 10.1016/j.foreco.2009.09.023.

Malcolm, Jay R.; Markhan, Adam; Neilson, Ronald P.; Garaci, Michael (2002): Estimated Migration Rates under Scenarios of Global Climate Change. In *Journal of Biogeography* 29 (7), pp. 835–849.

McLachlan, Jason S.; Hellmann, Jessica J.; Schwartz, Mark W. (2007): A framework for debate of assisted migration in an era of climate change. In *Conservation biology : the journal of the Society for Conservation Biology* 21. DOI: 10.1111/j.1523-1739.2007.00676.x.

Montwé, David; Spiecker, Heinrich; Hamann, Andreas (2015): Five decades of growth in a genetic field trial of Douglas-fir reveal trade-offs between productivity and rought tolerance. In *Tree Genetics & Genomes*. Available online at DOI 10.1007/s11295-015-0854-1.

Morgenstern, E. K. (1996): Geographic Variation in Forest Trees. Genetic Basis and Application of Knowledge in Silviculture. Vancouver, British Columbia, Canada: University of British Columbia Press.

O'Neill, Gregory A.; Hamann, Andreas; Wang, Tongli (2008): Accounting for Population Variation Improves Estimates of the Impact of Climate Change on Species' Growth and Distribution. In *Journal of Applied Ecology* 45 (4), pp. 1040–1049.

Pan, Yude; Birdsey, Richard A.; Fang, Jingyun; Houghton, Richard; Kauppi, Pekka E.; Kurz, Werner A. et al. (2011): A large and persistent carbon sink in the

world's forests. In *Science (New York, N.Y.)* 333 (6045), pp. 988–993. DOI: 10.1126/science.1201609.

Pharis & Ferrell (1966): Differences in Drought Resistance between Coastal and Inland Sources of Douglas Fir. In *Can. J. Bot.* 44 (12), pp. 1651–1659. DOI: 10.1139/b66-177.

Rehfeldt, G. E. (1977): Growth and cold hardiness of intervarietal hybrids of douglas-fir. In *Theoret. Appl. Genetics* 50 (1), pp. 3–15. DOI: 10.1007/BF00273790.

Rehfeldt, Gerald E.; Tchebakova, Nadejda M.; Parfenova, Yelena I.; Wykoff, William R.; Kuzmina, Nina A.; Milyutin, Leonid I. (2002): Intraspecific responses to climate in Pinus sylvestris. In *Global change biology* 8 (9), pp. 912–929. DOI: 10.1046/j.1365-2486.2002.00516.x.

Savolainen, Outi; Pyhäjärvi, Tanja; Knürr, Timo (2007): Gene Flow and Local Adaptation in Trees. In *Annu. Rev. Ecol. Evol. Syst.* 38 (1), pp. 595–619. DOI: 10.1146/annurev.ecolsys.38.091206.095646.

Schwartz, Mark W. (1992): Modelling effects of habitat fragmentation on the ability of trees to respond to climatic warming. In *Biodivers Conserv* 2 (1), pp. 51–61. DOI: 10.1007/BF00055102.

St Clair, J. Bradley; Mandel, Nancy L.; Vance-Borland, Kenneth W. (2005): Genecology of Douglas fir in western Oregon and Washington. In *Annals of botany* 96 (7), pp. 1199–1214. DOI: 10.1093/aob/mci278.

St. Clair & Howe (2007): Genetic maladaptation of coastal Douglas ‐ fir seedlings to future climates. In *Global change biology* 13 (7), pp. 1441–1454. DOI: 10.1111/j.1365-2486.2007.01385.x.

Telford & Birks (2005): The secret assumption of transfer functions: problems with spatial autocorrelation in evaluating model performance. In *Quaternary Science Reviews* 24 (20–21), pp. 2173–2179. DOI: 10.1016/j.quascirev.2005.05.001.

Thomas, Chris D.; Cameron, Alison; Green, Rhys E.; Bakkenes, Michel; Beaumont, Linda J.; Collingham, Yvonne C. et al. (2004): Extinction risk from climate change. In *Nature* 427 (6970), pp. 145–148. DOI: 10.1038/nature02121.

Wang, Tongli; O'Neill, Gregory A.; Aitken, Sally N. (2010): Integrating environmental and genetic effects to redict responses of tree populations to climate. In *Ecological Applications* 20 (1), pp. 153–163.

White, Timothy L. (1987): Drought Tolerance of Southwestern Oregon Douglas-Fir. In *Forest Science* 33 (2), pp. 283–293.

7 Appendix

7.1 Residual Analysis of Excluded Outliers

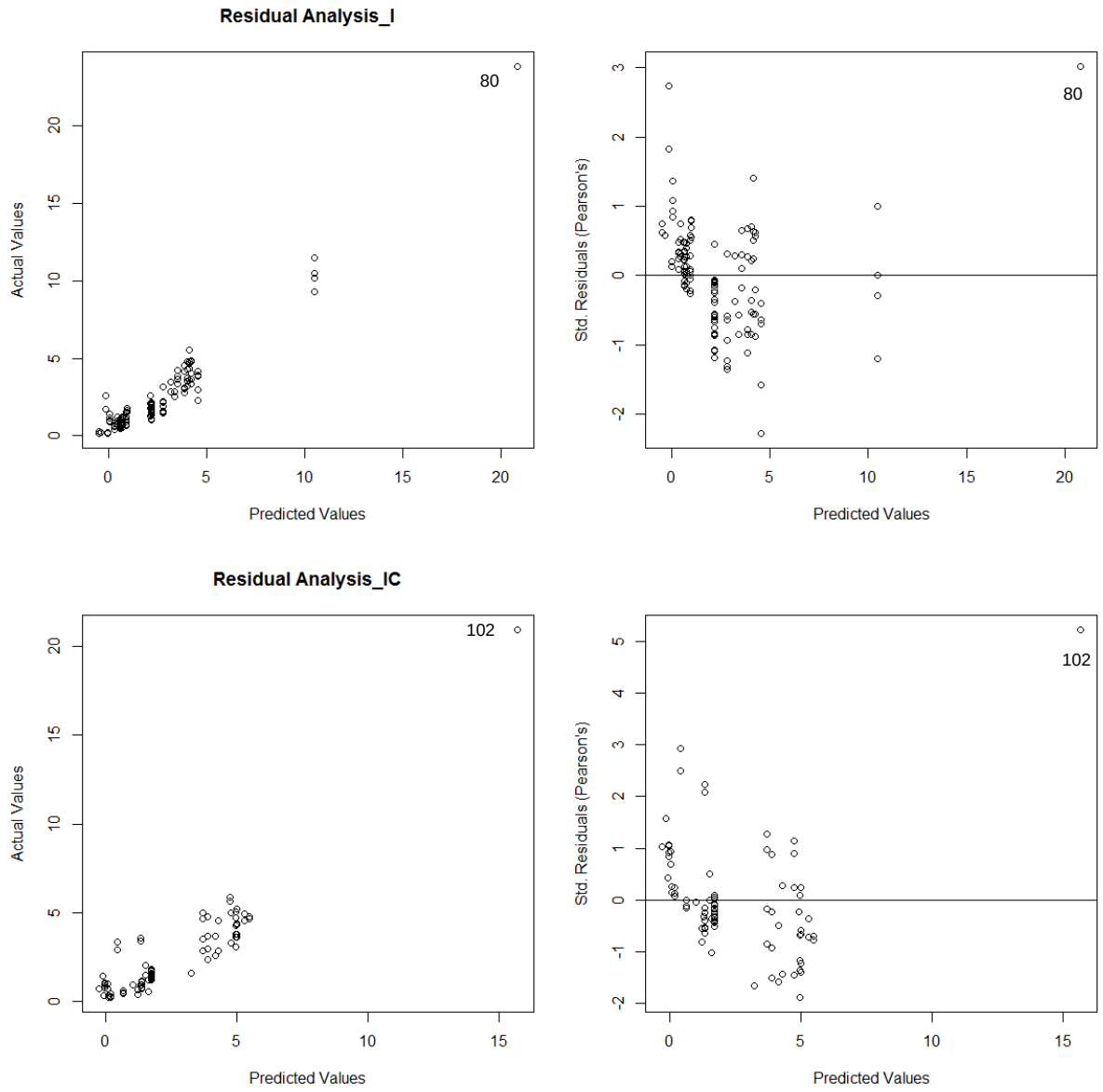

Figure 20: Residual analysis of excluded outliers

Table 8: Cook's distance of removed outliers

Tree_ID	Provenance	Cook's dist.	Cook's d of the following points
80	I	3.68e+00	1.36e-01 1.00e-01
102	IC	7.73e+00	5.70e-02 4.18e-02

7.2 R-Code

```
########################################
##   Bachelor's Thesis Source code    ##
########################################

# Set parameters:
Data <- read.csv("Data.csv")
load("PCA_EU.RData")

bioclimlist <- c("MAT","AHM","MWMT","MCMT","MAP", "DD5","DD_0", "FFP", "EMT","SHM","TD", "MDMP")
bioclimlist2 <- c("MAT","AHM","MWMT","MCMT","MAP", "DD5","DD_0", "FFP", "EMT","SHM","TD")
yearvec <- c(20,50,80)
rcpvec <- c(45,85)
modelvec <- c("GLM","RF")
GROUP_ORDER <- as.character(sort(unique(Data$PROV_GROUP_MI)))

# For colours
prov.cols <- data.frame(prov=c("C_BC","C_WA","C_OR","LE_CA","DC_WA","DC_OR","CC_WA","CC_OR","HE_C
A","IC","IN","I","IS"), col=c("#2C7C3D","#27B34B","#A9CA72","#88CEB5","#D6A95B","#FAF078","#26367
8","#3E67B1","#639DD4","#744099","#954044","#DE6061","#FAA51A"), stringsAsFactors=F)
prov.cols <- prov.cols[c(1,3,2,8,7,6,5,9,12,10,11,13,4),]

####################################
## PCA
for (k in bioclimlist2){
  if ( k == "MAT"){
    dat <- asc2dataframe(paste0(k, ".asc"))
    dat <- subset(dat,
                  (dat$x/10000)==round((dat$x/10000)) &
                    (dat$y/10000)==round((dat$y/10000)))
    names(dat)[3] <- k
  } else {
    dat2 <- asc2dataframe(paste0(k, ".asc"))
    dat2 <- subset(dat2,
                   (dat2$x/10000)==round((dat2$x/10000)) &
                     (dat2$y/10000)==round((dat2$y/10000)))
    dat[,ncol(dat) + 1] <- dat2[,3]
    names(dat)[ncol(dat)] <- k
  }
}
# withdraw precipitation ascii-files, downscale and convert them
for (k in 1:12){
  if ( k == 1){
    dat3 <- asc2dataframe(paste0("PPT",k,".asc"))
    dat3 <- subset(dat3,
                   (dat3$x/10000)==round((dat3$x/10000)) &
                     (dat3$y/10000)==round((dat3$y/10000)))
  } else {
    dat4 <- asc2dataframe(paste0("PPT",k,".asc"))
    dat4 <- subset(dat4,
                   (dat4$x/10000)==round((dat4$x/10000)) &
                     (dat4$y/10000)==round((dat4$y/10000)))
    dat3[,ncol(dat3) + 1] <- dat4[,3]
  }
}
# room for storing the new variable
MDMP <- rep(0,86666)
dat3 <- cbind(dat3, MDMP)


# create the maximum (MDMP) and minimum(MWMT) -value column
dat3[,15] <- apply(dat3[,c(3:14)], 1, min)

# add MDMP to main frame
dat <- cbind(dat, dat3[,15])
names(dat)[ncol(dat)] <- "MDMP"

# conduct and store PCA
pca.dat <- dat[,-c(1,2)]
```

```
pca <- prcomp(pca.dat, scale=T)
save(pca, file = "PCA_EU.RData")

# graphics:
eig <- (pca$sdev)^2
variance <- eig*100/sum(eig)
cumvar <- cumsum(variance)
PCs <- c(1:12)
pcaplot <- data.frame(eig = eig, variance = variance, cumvariance = cumvar)
pcaplot <- cbind(PCs, pcaplot)

# graphics:
graphics.off(); x11(w=11,h=10)
p <- ggplot(data=pcaplot[1:10,], aes(x=PCs, y=variance)) +
  geom_bar(stat="identity", fill="steelblue")+
  geom_text(aes(label=round(cumvariance,2)), vjust=-0.3, size=3.5)+
  scale_x_continuous(breaks = c(1:10),
                     name="Principal Components",
                     expand = c(0.01,0)) +
  scale_y_continuous(name="Percentage of Variances",
                     expand=c(0.01, 0),
                     limits= c(0,64),
                     breaks = c(0,10,20,30,40,50,60))+
  ggtitle("PCA - Variances")+
  theme_classic()+
  theme(plot.margin = unit(c(1,1,1,1),"cm"),
        legend.position="none")
print(p)
savePlot("Principle Component Screeplot_PC_EU.png",type="png")


#####################################
## Moran's I

# Create and dredge GLM: (RF accordingly, but with rf <- randomforest(HEIGHT ~ …)
PROVglm <- glm(HEIGHT ~ AGE + SITE_PC1 + I(SITE_PC1^2) + SITE_PC2 + I(SITE_PC2^2)+ SITE_PC3 + I(S
ITE_PC3^2)+ SITE_PC4 + I(SITE_PC4^2), data = Daten, na.action = "na.fail")
bestglm <- get.models(dredge(PROVglm), 1)[[1]]

PROVglm2 <- glm(HEIGHT ~ AGE + SITE_PC1 + I(SITE_PC1^2) + SITE_PC2 + I(SITE_PC2^2)+ SITE_PC3 + I
(SITE_PC3^2)+ SITE_PC4 + I(SITE_PC4^2), data = Daten, na.action = "na.fail")
bestglm2 <- get.models(dredge(PROVglm2), 1)[[1]]

PROVglm3 <- glm(HEIGHT ~ AGE + SITE_PC1 + I(SITE_PC1^2) + SITE_PC2 + I(SITE_PC2^2)+ SITE_PC3 + I
(SITE_PC3^2)+ SITE_PC4 + I(SITE_PC4^2)+ SITE_PC5 + I(SITE_PC5^2), data = Daten, na.action = "na.f
ail")
bestglm3 <- get.models(dredge(PROVglm3), 1)[[1]]

# Spacial Autocorrelation: Moran's I GLM
x = SITE_X/1000 # take x coordinates
y = SITE_Y/1000 # take y coordinates
z1 = bestglm$residuals  # and the residuals of the predictions
z2 = bestglm2$residuals
z3 = bestglm3$residuals

co <- correlog(x,y,z=z1, increment = 100, resamp=0)  # implement analysis of autocorrelation
co2 <- correlog(x,y,z=z2, increment = 100, resamp=0)
co3 <- correlog(x,y,z=z3, increment = 100, resamp=0)

# visualize results
x11(10,10)
plot(0,type="n",col="black",ylab="Moran's I",xlab=" lag - Distance [km]",xlim=c(0,1000),ylim=c(-1,
1), main=("Moran's I - GLMs"), cex.lab=1.75, cex.main=2)
abline(h=0,lty="dotted")
lines(co$correlation~co$mean.of.class,col="red",lwd=2)
lines(co2$correlation~co2$mean.of.class,col="darkolivegreen2",lwd=2)
lines(co3$correlation~co3$mean.of.class,col="blue",lwd=2)
legend('topright', mods ,
       lty=1, lwd =2, col=c('red', 'darkolivegreen2', 'blue'), bty='n', cex=1.5)

savePlot("Moran's I_GLM.png",type="png")
```

```
#######################################
## Creating Climate Surfaces

for (i in yearvec){
  for (j in rcpvec){
    for (k in bioclimlist2){
      if ( k == "MAT"){
        dat <- asc2dataframe(paste0(k, ".asc"))

        #downscaling the data
        dat <- subset(dat,
                      (dat$x/2000)==round((dat$x/2000)) &
                        (dat$y/2000)==round((dat$y/2000)))
        names(dat)[3] <- k
      } else {
        dat2 <- asc2dataframe(paste0(k, ".asc"))
        dat2 <- subset(dat2,
                       (dat2$x/2000)==round((dat2$x/2000)) &
                         (dat2$y/2000)==round((dat2$y/2000)))
        dat[,ncol(dat) + 1] <- dat2[,3]
        names(dat)[ncol(dat)] <- k
      }
    }

    # add MDMP (because it was also created exclusively from PPT-ASCIIs)
    dat2 <- asc2dataframe("MDMP.asc")
    dat <- cbind(dat, dat2[3])
    names(dat)[ncol(dat)] <- "MDMP"

    # transform with pca
    dat[c(paste0("SITE_PC", 1:12))] <- predict(pca, newdata= dat[,3:14])

    # remove climate variables to reduce file size
    dat <- dat[,c(1,2,15:19)]

    # create empty column for Heigth
    PRED_HEIGHT <- rep(0, nrow(dat))
    dat <- cbind(dat,PRED_HEIGHT)

    #save the output
    write.csv(dat, file = paste("Climate surface",j,i, ".csv", sep="_"), row.names=F)
  }
}
# ... plus the same for 6190 Data



#######################################
## Cross validation

# Create columns for predictions and residual errors in “Daten”
PRED_HEIGHT <- rep(0, nrow(Daten))
RES <- rep(0, nrow(Daten))
Daten <- cbind(Daten,PRED_HEIGHT,RES)
SITEGROUP_LIST <- as.character(unique(Daten$SITE_GROUP_MI))

# Create matrix for storing results:
rows <- c(SITEGROUP_LIST, "ALL")
cols <- c("No. of Sites", "RMSE", "COR_PEAR", "COR_SPEAR")
crossval_results <- matrix(nrow = 11, ncol = 4)
rownames(crossval_results) <- rows
colnames(crossval_results) <- cols

## GLM3 ##
##########
for(i in SITEGROUP_LIST){
  testIndexes <- which(Daten$SITE_GROUP_MI==i)
  crossval_results[i, 1] <- length(testIndexes)
```

```
  PROVglm <- glm(HEIGHT ~ AGE + SITE_PC1 + I(SITE_PC1^2) + SITE_PC2 + I(SITE_PC2^2)+ SITE_PC3 + I
(SITE_PC3^2), data = Daten[-testIndexes,], na.action = "na.fail")

  # Dredge Model to get the best model:
  bestglm <- get.models(dredge(PROVglm), 1)[[1]]

  # test glm with train set
  Daten[testIndexes, "PRED_HEIGHT"] <- predict(bestglm, newdata = Daten[testIndexes, ], type="res
ponse")
  Daten[testIndexes,"RES"] <- Daten[testIndexes,"HEIGHT"]-Daten[testIndexes, "PRED_HEIGHT"]

  crossval_results[i,2] <- sqrt( sum( Daten[testIndexes,"RES"]^2  )) / length(testIndexes)
  crossval_results[i,3] <- cor(Daten[testIndexes,"HEIGHT"], Daten[testIndexes, "PRED_HEIGHT"], me
thod = "pearson")
  crossval_results[i,4] <- cor(Daten[testIndexes,"HEIGHT"], Daten[testIndexes, "PRED_HEIGHT"], me
thod = "spearman")
}

crossval_results[11,1] <- sum(crossval_results[1:10,1])
crossval_results[11,2] <- sqrt( sum( Daten[,"RES"]^2  )) / nrow(Daten)
crossval_results[11,3] <- cor(Daten["HEIGHT"], Daten["PRED_HEIGHT"], method = "pearson")
crossval_results[11,4] <- cor(Daten["HEIGHT"], Daten["PRED_HEIGHT"], method = "spearman")
write.csv(crossval_results, file = "Cross-Validation_GLM3.csv")

## ... and so on for GLM 4,5 and RF 3-5


########################################
## Developing individual provenances' GLMs:

for (i in GROUP_ORDER){
  testtrees <- which(Daten$PROV_GROUP_MI == i)

  # Create test set of Provenance Trees
  testData <- Daten[testtrees, ]

  # Run "best" (=automatically generated) glm with PROV-Data from 05a
  PROVglm <- glm(HEIGHT ~ AGE + SITE_PC1 + I(SITE_PC1^2) + SITE_PC2 + I(SITE_PC2^2)+ SITE_PC3 + I
(SITE_PC3^2)+SITE_PC4 + I(SITE_PC4^2), data = testData, na.action = "na.fail")

  # Dredged Models:
  bestglm <- get.models(dredge(PROVglm), 1)[[1]]

  # add Validation:
  predresults <- predict(bestglm, testData, type="response")
  rP <- residuals(bestglm, type = "pearson")

  # Capture output in files (txt and RData) and save coefficients and std.errs
  capture.output(summary(bestglm), file= paste0(i,"_varglm_Summary.txt"))
  save(bestglm, file = paste0(i,"_varglm.RData"))

  # Add Residual Analysis
  graphics.off(); x11(w=20,h=10)
  par(mfrow=c(1,2), mar=c(5,5,4,1))
  plot(testData[,"HEIGHT"]~predresults, ylab="Actual Values", xlab="Predicted Values", main= past
e0("Residual Analysis_",i))
  plot(rP~predresults, ylab="Std. Residuals (Pearson's)", xlab="Predicted Values")
  abline(0,0)

  savePlot(paste0("Residual Analysis_", i, ".png"),type="png")
}


########################################
## Conducting Height Predictions:

for (k in modelvec){
  # 6190-Data:
  # Get respective climate surface
  dat <- read.csv("Climate surface_6190.csv")
  dat <- subset(dat,
```

```
                (dat$x/4000)==round((dat$x/4000)) &
                  (dat$y/4000)==round((dat$y/4000)))

  for (l in GROUP_ORDER){
    # Get respective glm
    load(paste0(l,"_varglm.RData"))

    # set AGE to a fixed value
    AGE <- rep(30, nrow(dat))

    # cbind PROV data
    dat <- cbind(dat, AGE)

    # predict with it and the data the expected heights
    dat["PRED_HEIGHT"] <- predict(bestglm, newdata = dat)

    #create data frames to match the requirements of dataframe2asc
    Height_frame <- cbind(dat[,c(1,2)],dat["PRED_HEIGHT"])

    # store the frames
    dataframe2asc(Height_frame,filenames = paste0(l, "_PRED_Height.asc"))
  }

  ## Data from the Future plus Height prediction:
  for (i in yearvec){
    for (j in rcpvec){
      # Get respective climate surface
      dat <- read.csv(paste("Climate surface",j,i, ".csv", sep="_"))
      dat <- subset(dat,
                    (dat$x/4000)==round((dat$x/4000)) &
                      (dat$y/4000)==round((dat$y/4000)))
      # set AGE to a fixed value
      AGE <- rep(30, nrow(dat))
      # cbind PROV data
      dat <- cbind(dat, AGE)

      for (l in GROUP_ORDER){
        # Get respective glm
        load(paste0(l,"_varglm.RData"))

        # predict
        dat["PRED_HEIGHT"] <- predict(bestglm, newdata = dat)

        #create data frames to match the requirements of dataframe2asc
        Height_frame <- cbind(dat[,c(1,2)],dat["PRED_HEIGHT"])

        # store the frames
        dataframe2asc(Height_frame,filenames = paste(l,j,i,"_PRED_Height.asc", sep="_"))
      }
    }
  }
}


########################################
## Prediction Tables
#  so later I didn't have to load all Height_ASCIIs in R seperately

for (l in modelvec){
  for (k in GROUP_ORDER){
    if ( k == "C_BC"){
      dat <- asc2dataframe(paste0(k,"_PRED_Height.asc"))
      names(dat)[3] <- k
    } else {
      dat2 <- asc2dataframe(paste0(k,"_PRED_Height.asc"))
      dat[,ncol(dat) + 1] <- dat2[,3]
      names(dat)[ncol(dat)] <- k
    }
  }
  write.csv(dat,file="6190_All_Preds_Table_GLM.csv", row.names=F)
```

```
  ## Overview Tables for all scenarios
  for (i in yearvec){
    for (j in rcpvec){
      for (k in GROUP_ORDER){
        if ( k == "C_BC"){
          dat <- asc2dataframe(paste("C_BC", j, i,"PRED_Height.asc", sep="_"))
          names(dat)[3] <- "C_BC"
        } else {
          dat2 <- asc2dataframe(paste(k,j,i,"PRED_Height.asc", sep="_"))
          dat[,ncol(dat) + 1] <- dat2[,3]
          names(dat)[ncol(dat)] <- k
        }
      }
      write.csv(dat,file=paste("All_Preds", j, i, "Table_GLM.csv", sep="_"), row.names=F)
    }
  }
}


########################################
## Validation Tests: VAR

# Create average tree height per provenance and site
sum.PROV_GROUP_MI <- ddply(Daten[,c("ID", "SITE_ID", "PROV_GROUP_MI", "HEIGHT", "AGE", paste0("SI
TE_PC",1:5))], .(SITE_ID,PROV_GROUP_MI), numcolwise(mean))

# Ranking
sum.PROV_GROUP_MIorder <- sum.PROV_GROUP_MI[with(sum.PROV_GROUP_MI, order(SITE_ID, -HEIGHT)), ]

# Room for Model Predictions
PRED_HEIGHT <- rep(NA, nrow(sum.PROV_GROUP_MIorder))
sum.PROV_GROUP_MIorder <- cbind(sum.PROV_GROUP_MIorder, PRED_HEIGHT)

# Create List to look at each site seperately
site.rank <- split(sum.PROV_GROUP_MIorder, f=sum.PROV_GROUP_MIorder[, "SITE_ID"])

# Factor to refer to sites
sites <- unique(Daten$SITE_ID)

# Create Matrix to store Results
val.results <- matrix(NA, nrow = 113, ncol=8)
columns <- as.vector(c("SITE", "VAR", "obs1", "obs2", "obs3", "mod1", "mod2", "mod3")) # for best
 three observed and best three modelled
colnames(val.results) <- columns

# Create Matrix to store average Results from repeated VAR
VAR_REPEATS <- matrix(NA, nrow = 113, ncol=11)
columns <- as.vector(c(1:10, "SUM"))
colnames(VAR_REPEATS) <- columns

## Implement VAR
for (z in 1:10){
  for (i in 1:112){
    val.results[i,1] <- sites[i]
    # for if I have a choice of more than three
    if (nrow(site.rank[[i]])> 3){
      obs <- mean(site.rank[[i]][sample(nrow(site.rank[[i]]), 3),"HEIGHT"])
      for (j in 1: nrow(site.rank[[i]])){
        load(paste0(site.rank[[i]][j, 2],"_varglm.RData"))
        site.rank[[i]][j, 10] <- predict(bestglm, newdata=site.rank[[i]][j,])
      }
      mod <- mean(tail(sort(site.rank[[i]][, 10]),3))
      modnames <- site.rank[[i]][order(site.rank[[i]][, 10], decreasing = T),]
      val.results[i,2] <- mod-obs
      val.results[i,6] <- as.character(modnames[1,2])
      val.results[i,7] <- as.character(modnames[2,2])
      val.results[i,8] <- as.character(modnames[3,2])
    }

    # if I only have three, I want to identify the best two
    if (nrow(site.rank[[i]]) == 3){
```

```
      obs <- mean(site.rank[[i]][sample(nrow(site.rank[[i]]), 2),"HEIGHT"])
      for (j in 1: 3){
        load(paste0(site.rank[[i]][j, 2],"_varglm.RData"))
        site.rank[[i]][j, 10] <- predict(bestglm, newdata=site.rank[[i]][j,])
      }
      mod <- mean(tail(sort(site.rank[[i]][, 10]),2))
      modnames <- site.rank[[i]][order(site.rank[[i]][, 10], decreasing = T),]
      val.results[i,2] <- mod-obs
      val.results[i,6] <- as.character(modnames[1,2])
      val.results[i,7] <- as.character(modnames[2,2])
    }

    #if I only have two, I want to identify the better provenance
    if (nrow(site.rank[[i]]) == 2){
      obs <- mean(site.rank[[i]][sample(nrow(site.rank[[i]]), 1),"HEIGHT"])
      for (j in 1: 2){
        load(paste0(site.rank[[i]][j, 2],"_varglm.RData"))
        site.rank[[i]][j, 10] <- predict(bestglm, newdata=site.rank[[i]][j,])
      }
      mod <- max(site.rank[[i]][, 10])
      modnames <- site.rank[[i]][order(site.rank[[i]][, 10], decreasing = T),]
      val.results[i,2] <- mod-obs
      val.results[i,6] <- as.character(modnames[1,2])
    }

  }
  val.results[113,1] <- "ALL"
  val.results[113,2] <- sum(val.results[1:112,2])/112
  VAR_REPEATS[,z] <- val.results[,2]
}
# Histogram:
graphics.off(); x11(10,10)
hist(VAR_REPEATS[1:112,11], main= paste0("Average VAR of GLM: ", round(VAR_REPEATS[113,11],3), "
m"), xlab= "Value Above Random", ylab= "Frequency", cex.lab= 1.5, cex.main=2)
savePlot(filename = "VAR_GLM_hist.png", type="png")


#######################################
## Validation Tests: VAB

for (i in 1:112){
  val.results[i,1] <- sites[i]

  # for if I have a choice of more than three, note that obs is site.rank[[i]][1:3,] instead of r
andom sample
  if (nrow(site.rank[[i]])> 3){
    obs <- mean(site.rank[[i]][1:3,"HEIGHT"])
    for (j in 1: nrow(site.rank[[i]])){
      load(paste0(site.rank[[i]][j, 2],"_varglm.RData"))
      site.rank[[i]][j, 10] <- predict(bestglm, newdata=site.rank[[i]][j,])
    }
    mod <- mean(tail(sort(site.rank[[i]][, 10]),3))
    modnames <- site.rank[[i]][order(site.rank[[i]][, 10], decreasing = T),]
    val.results[i,2] <- mod-obs
    val.results[i,3] <- as.character(site.rank[[i]][1,2])
    val.results[i,4] <- as.character(site.rank[[i]][2,2])
    val.results[i,5] <- as.character(site.rank[[i]][3,2])

    val.results[i,6] <- as.character(modnames[1,2])
    val.results[i,7] <- as.character(modnames[2,2])
    val.results[i,8] <- as.character(modnames[3,2])
  }

  if (nrow(site.rank[[i]]) == 3){
    obs <- mean(site.rank[[i]][1:2,"HEIGHT"])
    for (j in 1: 3){
      load(paste0(site.rank[[i]][j, 2],"_varglm.RData"))
      site.rank[[i]][j, 10] <- predict(bestglm, newdata=site.rank[[i]][j,])
    }
    mod <- mean(tail(sort(site.rank[[i]][, 10]),2))
    modnames <- site.rank[[i]][order(site.rank[[i]][, 10], decreasing = T),]
```

```
    val.results[i,2] <- mod-obs
    val.results[i,3] <- as.character(site.rank[[i]][1,2])
    val.results[i,4] <- as.character(site.rank[[i]][2,2])

    val.results[i,6] <- as.character(modnames[1,2])
    val.results[i,7] <- as.character(modnames[2,2])
  }

  if (nrow(site.rank[[i]]) == 2){
    obs <- site.rank[[i]][1,"HEIGHT"]
    for (j in 1: 2){
      load(paste0(site.rank[[i]][j, 2],"_varglm.RData"))
      site.rank[[i]][j, 10] <- predict(bestglm, newdata=site.rank[[i]][j,])
    }
    mod <- max(site.rank[[i]][, 10])
    modnames <- site.rank[[i]][order(site.rank[[i]][, 10], decreasing = T),]
    val.results[i,2] <- mod-obs
    val.results[i,3] <- as.character(site.rank[[i]][1,2])

    val.results[i,6] <- as.character(modnames[1,2])
  }

}
val.results[113,1] <- "ALL"
val.results[113,2] <- sum(val.results[1:112,2])/112

# Save output
write.csv(val.results, file = "VAB_GLM.csv")

# Histogram of Results
graphics.off(); x11(10,10)
hist(val.results[,2], main= paste0("VAB of GLM4: ", round(val.results[113,2],3), " m"), xlab= "Va
lue Against Best", ylab= "Frequency", breaks=15, cex.lab= 1.5, cex.main=2)
savePlot(filename = "VAB_GLM_hist.png", type="png")

#######################################
## Validation Tests: Confusion Tables:

# Create a larger val-results matrix, so that all classes can be recorded:
val.results <- matrix(NA, nrow = 112, ncol=27)
columns <- as.vector(c("SITE", paste0("obs",1:13), paste0("mod", 1:13)))
colnames(val.results) <- columns

# Similar to VAB, just that all observed and predicted PROVs are recorded:
for (i in 1:112){
  val.results[i,1] <- sites[i]

  for (j in 1: nrow(site.rank[[i]])){
    load(paste0(site.rank[[i]][j, 2],"_varglm.RData"))
    site.rank[[i]][j, 10] <- predict(bestglm, newdata=site.rank[[i]][j,])
  }
  modnames <- site.rank[[i]][order(site.rank[[i]][, 10], decreasing = T),]
  val.results[i,2:14] <- as.character(site.rank[[i]][1:13,2])
  val.results[i,15:27] <- as.character(modnames[1:13,2])
}
write.csv(val.results, file = "confusion_matrix_GLM.csv")


#######################################
## Validation Tests: VAC

for (l in modelvec){
  dat <- read.csv(paste0("6190_All_Preds_Table_",l,".csv"))
  dat <- subset(dat,
                (dat$x/8000)==round((dat$x/8000)) &
                  (dat$y/8000)==round((dat$y/8000)))

  rank1 <- apply(dat[,-1:-2],1,function(x){names(sort(x))[13]}) #create ranking order of best ...
  rank2 <- apply(dat[,-1:-2],1,function(x){names(sort(x))[12]}) #...second best...
  rank3 <- apply(dat[,-1:-2],1,function(x){names(sort(x))[11]}) #...third best provenance
```

```
  for(i in 1:nrow(dat)){
    if(i==1){heights <- data.frame(H1=NA,H2=NA,H3=NA)}
    heights[i,c("H1","H2","H3")] <- dat[i,c(rank1[i],rank2[i],rank3[i])]
  }
  # get the average height from the best 3 provenances from 6190
  heights$AVG <- rowMeans(heights)
  heights[c("y","x")] <- dat[1:nrow(heights),c("y","x")]
  reference_height <- heights[,c("y","x","AVG")]
  dataframe2asc(reference_height, filenames= paste0(l,"_6190_Reference Height.asc"))


  # Create Rank from the predictions of each Time and Combination
  for (i in yearvec){
    for (j in rcpvec){
      dat2 <- read.csv(paste("All_Preds", j, i, "Table", l, ".csv", sep="_"))
      dat2 <- subset(dat2,
                     (dat2$x/8000)==round((dat2$x/8000)) &
                       (dat2$y/8000)==round((dat2$y/8000)))

      # ranking order of the predicted values
      rankA <- apply(dat2[,-1:-2],1,function(x){names(sort(x))[13]})
      rankB <- apply(dat2[,-1:-2],1,function(x){names(sort(x))[12]})
      rankC <- apply(dat2[,-1:-2],1,function(x){names(sort(x))[11]})

      # Height-frame with the best provenances from the predicted data (rankA-C)
      for(k in 1:nrow(dat2)){
        if(k==1){heightsmod <- data.frame(H1=NA,H2=NA,H3=NA)}
        heightsmod[k,c("H1","H2","H3")] <- dat2[k,c(rankA[k],rankB[k],rankC[k])]
      }
      heightsmod$AVG <- rowMeans(heightsmod)

      #Height-frame with the best provenances from 6190 (rank1-3)
      for(k in 1:nrow(dat2)){
        if(k==1){heightscon <- data.frame(H1=NA,H2=NA,H3=NA)}
        heightscon[k,c("H1","H2","H3")] <- dat2[k,c(rank1[k],rank2[k],rank3[k])]
      }
      heightscon$AVG <- rowMeans(heightscon)

      # Comparison Best_Modelled vs. Consistent_Modell
      VAC <- heightsmod$AVG - heightscon$AVG
      VAC_frame <- cbind(dat[,c("y","x")],VAC )
      dataframe2asc(VAC_frame, filenames= paste0(l,j,i, "HeightDiff_Best_vs_Cons.asc"))

      # Comparison Best_Modelled vs. 6190_Modell
      trend <- heightsmod$AVG - heights$AVG
      trend_frame <- cbind(dat[,c("y","x")], trend)
      dataframe2asc(trend_frame, filenames= paste0(l,j,i,"HeightDiff_Best_vs_6190.asc"))
    }
  }
}
# Visualization on the example of GLM 45_20:

#get the data
test1 <- asc2dataframe("GLM_45_20_HeightDiff_Best_vs_6190.asc")
test2 <- asc2dataframe("GLM_45_80_HeightDiff_Best_vs_6190.asc")
test3 <- asc2dataframe("GLM_85_20_HeightDiff_Best_vs_6190.asc")
test4 <- asc2dataframe("GLM_85_80_HeightDiff_Best_vs_6190.asc")

#get in shape for ggplot
test <- cbind(test1, test2[,3],test3[,3],test4[,3])
colnames(test)[3:6] <- c("45_20", "45_80", "85_20", "85_80")
plotdat <- melt(test, id.vars=c("x", "y"), measure.vars=c("45_20", "45_80", "85_20", "85_80"))

# create graph with ggplot
p <- ggplot(data=plotdat, aes(x=x, y=y, col= value)) +
  geom_point(size=0.002)+
  facet_wrap(~variable,ncol=2)+
  coord_equal() +
  scale_color_gradientn(colours=c("darkred", "gold","forestgreen"), limits=c(-8, 8), oob=squish)+
  theme(axis.title.x = element_blank(),
        axis.title.y = element_blank(),
```

```
        axis.text.x = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks = element_blank(),
        plot.margin = unit(c(0, 0, 0, 0), "cm")
  )+
  ggtitle("GLM - Changed vs 6190 Overview")
graphics.off(); x11(w=58,h=50)
print(p)
savePlot("GLM5 - Best vs 6190.png", type="png")


#######################################
## Overview Histograms

for (l in modelvec){
  test <- read.csv(paste0("6190_All_Preds_Table_",l,".csv"))
  plotdat <- melt(test, id.vars=c("x","y"), measure.vars=GROUP_ORDER)
  plotdat$variable <- factor(plotdat$variable, levels=GROUP_ORDER)
  ps <- quantile(plotdat$value, c(0.02, 0.98)) # to only show the important range

  # Overviewplot with 2-98-Quantile spectrum
  graphics.off(); x11(h=10,w=6)
  p <- ggplot(plotdat, aes(x=value,fill=variable)) +
    geom_histogram(bins=500) +
    geom_vline(xintercept=median(plotdat$value)) +
    facet_grid(variable~., scales="free_y") +
    scale_y_continuous("", breaks=NULL) +
    scale_x_continuous("Projected Height (Age=30)", limits=c(ps[1],ps[2])) +
    scale_fill_manual("Provenance\nGroup", values=prov.cols$col) +
    ggtitle(paste0(l," 6190 Projection by Provenance"))
  print(p)
  savePlot(paste0(l," 6190_Overview_Percentile.png",type="png"))
}


#######################################
## Visualization of Predictions per PROV

for (l in modelvec){
  for (i in yearvec){
    for (j in rcpvec){
      test <- read.csv(paste("All_Preds",j,i,"Table" ,l, ".csv", sep="_"))
      test <- subset(test,
                     (test$x/10000)==round((test$x/10000)) &
                       (test$y/10000)==round((test$y/10000))) #downscale to 10km, as plots are sm
all and data is abundant

      ## get shaped and ggplot:
      plotdat <- melt(test, id.vars=c("x", "y"), measure.vars=GROUP_ORDER2)

      p <- ggplot(data=plotdat, aes(x=x, y=y, col=value)) +
        geom_point(size=0.002)+
        facet_wrap(~variable,ncol=4)+
        coord_equal() +
        scale_color_gradientn(colours=c("darkred", "gold","forestgreen"), limits=c(5, 20), oob=sq
uish)+
        theme(axis.title.x = element_blank(),
              axis.title.y = element_blank(),
              axis.text.x = element_blank(),
              axis.text.y = element_blank(),
              axis.ticks = element_blank(),
              plot.margin = unit(c(0, 0, 0, 0), "cm")
        )+
        ggtitle(paste(l,j,i,"per PROV",sep=" "))+
        graphics.off(); x11(w=40,h=31)
      print(p)
      savePlot(paste(l,j,i, " Maps.png"), type="png")
    }
  }
}
```

Declaration

I declare that I have produced this thesis by myself and that I have not used other sources as those listed in the references. Passages taken verbatim or in meaning from other sources are identified as such and the sources are acknowledged and cited. All figures were either produced by myself or are attributed with a reference. Furthermore, I certify that this thesis in this or a similar form has not been previously submitted to any other graduation institution.

Raphael Habel

Freiburg, 27. April 2016