# Czech University of Life Sciences

Faculty of Forestry and Wood Sciences

Department of Genetics and Physiology of Forest Trees

# Ph.D. Thesis

**Promoting climate adaptation of European larch through genetic evaluation, intrapopulation adaptive response, and GxE dynamics across Austrian landscapes**

Author: Ing. Valérie Poupon

Supervisor: prof. Ing. Milan Lstibůrek, PhD.

2023

# CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Forestry and Wood Sciences

# Ph.D. THESIS ASSIGNMENT

Ing. Valérie Poupon

Forestry Engineering
Forest Biology

Thesis title

**Promoting climate adaptation of European larch through genetic evaluation, intrapopulation adaptive response, and GxE dynamics across Austrian landscapes**

---

**Objectives of thesis**

In the thesis, we will be analyzing the data of a European larch population that was planted across a wide range of environments. The main aim of the thesis is to study European larch's genotypic adaptation to its environment to support gene resource management activities in the context of environmental changes such as Climate change.
The specific objectives can be summarized as follow:
(1) to analyze the population's genetic diversity at the genotype level,
(2) to evaluate the population's genotype by environment interaction and,
(3) to estimate the impact of elevation and climate on the population and the studied genotypes.

**Methodology**

The thesis will be based on a previously collected European larch dataset and its associated reconstructed pedigree. The research will follow two main approaches to analyzing the data:

(1) the development of a new methodology that combines traditional tree breeding methods and the response function methodology,

(2) the use of factor analytic statistical modeling.

Afterward, the successes and limitations of these approaches will be discussed, as well as the potential methodological improvements. The utility of analyzing tree populations at the genotypic level will also be argued, especially in the context of climate change.

**The proposed extent of the thesis**

100 NS

**Keywords**

assisted migration, climate change, factor analytic, forest tree breeding, genetic diversity, *Larix decidua*, mixed models, multi-environment trials, response function

---

**Recommended information sources**

Calleja-Rodriguez, A., Andersson Gull, B., Wu, H.X., Mullin, T.J., Persson, T., 2019. Genotype-by-environment interactions and the dynamic relationship between tree vitality and height in northern Pinus sylvestris. Tree Genet. Genomes, 15, 1–15.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B., 2021. Climate change 2021: the physical science basis. Cambridge University Press, Cambridge, United Kingdom, and New York, 2391 pp.

McLachlan, J.S., Hellmann, J.J. and Schwartz, M.W., 2007. A framework for debate of assisted migration in an era of climate change. Conservation biology, 21(2), 297-302.

Nardin, M., Musch, B., Rousselle, Y., Guérin, V., Sanchez, L., Rossi, J.P., Gerber, S., Marin, S., Pâques, L.E. and Rozenberg, P., 2015. Genetic differentiation of European larch along an altitudinal gradient in the French Alps. Annals of Forest Science, 72, 517-527.

Oliveira, I. C. M., Guilhen, J. H. S., Ribeiro, P. C. de O., Gezan, S. A., Schaffert, R. E., Simeone, M. L. F., Damasceno, C. M. B., Carneiro, J. E. de S., Carneiro, P. C. S., Parrella, R. A. da C., & Pastina, M. M., 2020. Genotype-by-environment interaction and yield stability analysis of biomass sorghum hybrids using factor analytic models and environmental covariates. Field Crops Research, 257, 1–10.

Rehfeldt, G. E., Tchebakova, N. M., Parfenova, Y. I., Wykoff, W. R., Kuzmina, N. A., & Milyutin, L. I., 2002. Intraspecific responses to climate in Pinus sylvestris. Global Change Biology, 8, 912–929.

Smith, A., Cullis, B., & Thompson, R., 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics, 57, 1138–1147.

Wang, T., Hamann, A., Yanchuk, A., O'Neill, G. A., & Aitken, S. N., 2006. Use of response functions in selecting lodgepole pine populations for future climates. Global Change Biology, 12, 2404–2416.

White, T.L., Adams, W.T., Neale, D.B., 2007. Forest genetics. Cabi, Wallingford, Oxfordshire, UK, 591 pp.

Zobel, B., Talbert, J., 1984. Applied forest tree improvement. John Wiley & Sons, New York, 505 pp.

**Expected date**

2022/23 SS – FFWS – State Doctoral Examinations

**The Dissertation Thesis Supervisor**

prof. Ing. Milan Lstibůrek, MSc, Ph.D.

**Supervising department**

Department of Genetics and Physiology of Forest Trees

Electronic approval: 1. 9. 2023

**prof. Ing. Milan Lstibůrek, MSc, Ph.D.**

Head of department

Electronic approval: 1. 9. 2023

**prof. Ing. Milan Lstibůrek, MSc, Ph.D.**

Chairperson of Field of Study Board

Electronic approval: 1. 9. 2023

**prof. Ing. Róbert Marušák, PhD.**

Dean

Prague on 07. 09. 2023

# Declaration

I declare that the Ph.D. thesis "Promoting climate adaptation of European larch through genetic evaluation, intrapopulation adaptive response, and GxE dynamics across Austrian landscapes" was developed independently under the guidance of my supervisor. Additionally, I declare that all the sources of information used in this work were cited accordingly.

# Acknowledgment

# Dedication

*"All things are difficult before they are easy."*

– Dr. Thomas Fuller

# Abstract

Climate change's negative impact on forest communities around the globe is already visible and is expected to increase further in the near future. This is in part due to the rapid environmental changes against the lower rates of adaptation and migration capacity of tree populations. Understanding the complex relationship between climate and tree species is essential to establishing efficient management strategies, such as assisted migration, to cope with the inevitable changes. In this thesis, we focused on three commercially important traits, height, diameter at breast height, and wood density, of an Austrian European larch population that was planted across 21 sites. Based on the concept that trees within the same population of origin will react differently to the environment, the aim was to analyze its intra-population genetic variability and genotype by environment interaction (GxE) and to infer which environmental variables are likely influencing it. Two different types of methodologies were used in this thesis, the first one, which we developed, combines the response function methodology and tree breeding approach, while the second one is based on a complex factor analytic modeling that is mostly used in crop science but that is also newly used in tree species. According to our results, both methods were successful in detecting the studied genotypes' response variability and GxE. Furthermore, we were able to determine which environmental variables are likely to influence each trait. Within our studied range, height seems to be affected by temperature, altitude, and precipitation seasonality. Diameter at breast height seems to be affected by the mean diurnal range and wood density, though mildly, seems to be affected by altitude. In our opinion, both methodologies can be applied to describe the intra-population genetic variability and were able to identify key environmental variables affecting specific commercially important traits. We believe that this knowledge can be used to select well-performing genotypes that are adapted to specific conditions, and hence be used in tree breeding and conservation programs (e.g., for assisted migration) focused on helping trees to cope and adapt to changing environmental conditions.

# Table of Contents

# List of tables and figures

# List of used symbols and abbreviations

| | |
|---|---|
| Alt | Altitude |
| Age | Age |
| AvPrec | Mean annual precipitation |
| AvTmean | Mean annual temperature |
| Biol2 | Mean diurnal range |
| Biol3 | Isothermality |
| Biol7 | Annual range of the temperatures |
| Biol15 | Precipitation seasonality |
| GxE | Genotype by environment |
| MAI-DBH | Diameter at breast height mean annual increment |
| MAI-H | Height mean annual increment |
| MET | Multi-environmental trial |
| MTCM | Mean temperature of the coldest month |
| pMAI | Predicted height mean annual increment |
| PP | Wood density |
| $R^2$ | Coefficient of determination |
| Tmean2 | Mean temperature of February |
| Tmin2 | Minimum temperature of February |

# 1 Introduction

This chapter presents a brief introduction to the dissertation thesis. First, the background and context of the study are discussed. Then, the research problem is described, followed by the research aims and objectives. Afterward, the significance of the research and its limitations are argued. Finally, the general organization of the dissertation will be indicated.

Anthropogenic climate change is a well-recognized phenomenon within the scientific community and the larger public. Notably, it encompasses shifts in temperatures, globally toward higher ends with a probable increase of 2°C before the 21$^{st}$ century, and an increasing incidence and intensity of extreme events such as extreme temperatures, drought spells, and forest fires (Angélil et al., 2017; Finkel and Katz, 2018; Klein Tank and Können, 2003; Masson-Delmotte et al., 2021). In this context of rapid shifts in environmental conditions, forest ecosystems are faced with unprecedented risks as the tree natural migration rates are expected to be insufficient to match the predicted speed of environmental change(Sally N. Aitken et al., 2008; Davis and Shaw, 2001; Gugger and Sugita, 2010), which would lead to adaptational lags and weakened forest states.

Within this context, the main aim of this thesis was to analyze the climatic adaptation of European larch, one of the major European commercial and ecological forest tree species. The core of the analyses is based on growth and wood density data from a set of half-sib individuals. These individuals originate from the same population and are planted across multiple stands. Using two types of multi-environmental trial (MET) analyses, the main objectives of this study were to understand the intra-population genetic variability and the GxE interaction across the different plantation sites of the studied population and to identify potential key environmental variables influencing the studied traits.

In forestry, traditional breeding programs aim at favorable productive traits with the primary objective to maximize economic gain. However, due to the forecasted changes in climate, it is now crucial to further our understanding of tree species' adaptation to climate, so that new forest management methods (such as assisted migration), aiming at mitigating the

detrimental consequences of climate change, can be implemented. Here, the two methodologies used were able to capture the intra-population genetic variability across the studied sites and infer key variables influencing growth and wood density. Hence, we believe that it is a necessary step for promoting the mitigation process. Additionally, it seems that these methodologies are both valid methods to be utilized for forest management purposes.

Regarding the structure of the dissertation, the next chapter (Chapter Two) details the aims and objectives of the study. This is followed by the literature review (Chapter Three), which dives into the details of the study's background and context. Afterward, there are the Material and methods, Results, and Discussion chapters (Chapters Four, Five, and Six) that are axed on the two main approaches used in this work. Finally, the last chapter (Chapter Seven) will be the Conclusion of the thesis.

# 2 Aims and objectives

This research aims to investigate European larch's adaptation to its environment at the genotypic level as a means to facilitate gene resource management in the context of climate change. More specifically, the main objectives can be described as follows:

(1) to analyze the intra-population genetic diversity of a European larch population at the individual genotypic level;

(2) to evaluate the dynamics of the GxE interaction across several planting sites,

(3) to estimate and contrast the influence of altitude and climatic variables on the overall population and the specific genotypes,

# 3 Literature review

## 3.1 Climate change

More than a century ago, the Swedish physicist Svante Arrhenius wondered about the influence of the atmospheric concentration of carbon dioxide, originating from the burning of fossil fuel associated with the rapid industrialization process, on the temperatures. He then built a model that predicted that doubling the concentration of atmospheric carbon dioxide ($CO_2$) would result in an average augmentation of 5℃ of the Earth's surface temperatures (Arrhenius, 1896). However, he predicted that it would take the human race about 1000 years. Since the mid-50th, global warming induced by an anthropogenic increase of greenhouse gases has been under further consideration (Keeling et al., 1976; Manabe and Wetherald, 1967; Petit et al., 1999). The first sound recording of the constant increase in ($CO_2$) is the Keeling curve (Keeling et al., 1976). Additionally, records such as the Vostok ice core drilling indicate a clear correlation between temperatures over the last 420 k years and the atmospheric concentration of carbon dioxide (Petit et al., 1999). Moreover, the ice cores also revealed that the present concentration of ($CO_2$) is at its highest value in the last 420 k years. In their pioneer paper, Manabe and Wetherald (1967) presented a relatively sound modeling of the potential future climate. The model forecasted a 2℃ increase in the temperature for the doubling of atmospheric $CO_2$ content.

Since then, many studies have been performed regarding the present and future predicted changes in the climate. In addition to the increase in global temperature, several other changes have been observed and forecasted regarding the increase in extreme events such as floods or droughts. Nowadays, climate change, and its anthropogenic origin, is a widely accepted phenomenon among the scientific community and the general public (Angélil et al., 2017; Finkel and Katz, 2018; Masson-Delmotte et al., 2021).

From 1946 to 1999, cold extremes dropped, warm extremes rose, and variations in extremes widened after 1975 (Alexander et al., 2006; Finkel and Katz, 2018; Frich et al., 2002; Klein Tank and Können, 2003). Additionally, precipitation levels have increased at the global scale (Angélil et al., 2017). However, no real spatial coherence was identified (Alexander et al., 2006; Allan and Soden, 2008; Klein Tank and Können, 2003). An amplification of precipitation extremes has been recorded in many places with very wet days (Alexander et al., 2006; Frich et al., 2002; Giorgi et al., 2019; Klein Tank and Können, 2003). These temperature and rainfall trends are expected to continue in the future (Allan and Soden, 2008; Beniston et al., 2007; Easterling et al., 2000; Frei et al., 2006; Giorgi et al., 2019). Concerning Europe, an increase in extreme precipitations is predicted north of 45° while no changes or even a decrease in rainfall is predicted in the southern parts (Beniston et al., 2007; Frei et al., 2006). Since the 1970s, an increase in aridity in Southern Europe and Asia, East Asia, and East Africa has been observed (Dai et al., 2018). The rising of temperatures results in the increment of the atmospheric moisture demands. This process, alongside events like El Nino and anomalous tropical sea surface temperature, is likely to be at least partly responsible for the increased aridity. Aridity and droughts are expected to further increase during the 21st century (Dai et al., 2018). Another study highlights that while global warming itself may not be responsible for drought events, the rising temperature would cause quicker and more intense droughts (Trenberth et al., 2014). In Europe, according to multi-model experiments, the probability of mega-heatwaves during the summer will be multiplied by 5 to 10 in the next 40 years (Barriopedro et al., 2011). Wildfires have also increased in some parts and there is evidence that it is connected to an increase in temperature (Mansoor et al., 2022; Westerling et al., 2006).

## 3.2 Influence of climate change on forest ecosystems

Climate change, with the increase in mean temperature and occurrence of extreme events, is already affecting the Earth's ecosystems and is expected to continue further (Walther et al., 2002). For 2050, climate warming scenarios are predicting that 15 to 37% of the species

sampled in many taxa will be committed to extinction (Román-Palacios and Wiens, 2020; Thomas et al., 2004). Climate change is considered to be one of the most important biodiversity drivers of change at present and will most likely become the most important one in the future. This driver of change affects the ecosystems alongside other anthropogenic factors such as land use, nitrogen deposition, pollution, invasive species, and over-exploitation of the lands and megafauna (Jandl et al., 2019; Sala et al., 2000). Many complicated interactions can occur among those drivers, amplifying or diminishing their influences. Moreover, the changes occurring at the ecosystem level can also generate positive or negative feedback. For example, an increase in temperature alongside carbon and nitrogen fertilization is expected to boost forest productivity (Carrasco et al., 2022; Kapeller et al., 2012), increasing carbon storage and leading to a reduction of carbon dioxide in the atmosphere (negative feedback) (Reichstein et al., 2013). On the contrary, an increase in forest surface would reduce the global albedo leading to an increase in temperature (positive feedback) (Bonan, 2008; Hyvonen et al., 2007; Lashof et al., 1997; Melillo et al., 2002).

Considering water restriction, many studies have documented its negative effect on forests (Allen et al., 2015; Anderegg et al., 2015; Mansoor et al., 2022; Martínez-Vilalta and Piñol, 2002; Rebetez and Dobbertin, 2004); Allen et al. (2010) reviewed existing literature on climate-induced mortality and found worldwide records on the negative effect of drought on tree survival. The effect of water restriction can be direct, but also indirect by for example boosting invasive insect and pathogens outbreaks (Anderegg et al., 2015; Rouault et al., 2006) and increasing wildfires occurrence (Mansoor et al., 2022; Westerling et al., 2006). In the future, climate change is expected to increase the severity and frequency of these tree mortality events (Allen et al., 2015; Halofsky et al., 2020). Additionally, trees weakened by drought events are experiencing a "legacy effect" with a reduction in growth and a slow recovery for up to four years after the event (Anderegg et al., 2015). A longer recovery time will affect carbon storage and, hence, carbon-climate feedback (Frank et al., 2015).

## 3.3 Response of forest ecosystems

The geographic distribution of species across the Earth's surface depends on several parameters. One key parameter is the species' evolutionary history, which includes its place of origin and subsequent expansion patterns. Another critical parameter is biotic interaction, encompassing competition, predation, and symbiosis among species. Moreover, the range of a species is significantly influenced by abiotic factors. These include edaphic factors, such as soil condition and geography, societal factors like land use and water availability, as well as climatic factors. As previously discussed, climate change is altering temperature and precipitation patterns globally. Given these changes, it's reasonable to anticipate shifts in the geographic ranges of various species.

Confronted with changes in climatic conditions, tree species can respond in different ways. They can either persist, adapt, or migrate to maintain and promote their fitness and competitive abilities (Sally N Aitken et al., 2008; Bussotti et al., 2015a; Sáenz-Romero et al., 2021).

### 3.3.1 Persistence

Tree population can persist the shifts in climatic conditions via phenotypic plasticity and acclimatization (Bussotti et al., 2015b; Fréjaville et al., 2019). That includes, for example, changes in the root/shoot ratio with changes in temperature or watering. Higher temperatures can induce a lower allocation of biomass to the roots while reduced watering can induce the opposite response (Matías et al., 2016). Another example would be the reduction of leaf area and growth to accommodate higher temperatures and lower precipitation levels (Sáenz-Romero et al., 2021).

### 3.3.2 Adaptation

Tree populations can adapt to the new conditions through natural selection that promotes the frequency of favorable alleles associated with adaptive traits. Many traits, such as height

and bud burst, are likely controlled by a large number of genes, each of them having a small effect (Sally N. Aitken et al., 2008). Moreover, it appears that many genes are not significantly affected by gametic phase disequilibrium in natural populations, indicating a high degree of independent inheritance (Brown et al., 2004; Heuertz et al., 2006; Neale and Savolainen, 2004). Studies have shown that traits controlled by many highly polymorphic polygenes are more likely to present phenotypic differences under selection pressure than individual loci (Aitken and Bemmels, 2016). Genetic polymorphism is initially forged by the stochastic process of mutation, which is typically slow and steady. However, population allelic diversity is largely governed by more immediate processes, such as gene flow—which can increase diversity—and genetic drift—which can reduce it. Large, interconnected populations tend to boast high genetic diversity. Yet, if these populations experience different selection pressures, high gene flow can hamper the local selection process (Garcia-Ramos and Kirkpatrick, 1997).

Lynch and Lande (1993) propose that species can adapt consistently in an evolving environment if the rate of change remains below a certain threshold. Determining this threshold depends on various factors, such as effective population size, genetic diversity, strength of selection, environmental stochasticity, and fecundity rate. When the rate of environmental change surpasses this threshold, populations may experience an adaptational lag, which can ultimately lead to extinction (Lynch and Lande, 1993).

### 3.3.3 Migration

In addition to persistence and adaptation, which involve phenotypical and genetic changes within the population, trees can also migrate to another location where the conditions will be more suitable for their current level of adaptation. As sessile organisms, trees can only migrate during their seed stage. Various species-specific vectors, such as wind, water, or animals, aid the dispersal of these seeds from the parent tree.

What facilitates observable directional migration in tree species is the seeds' limited viability in unsuitable locations. As climatic conditions shift directionally, tree populations tend to move correspondingly, leaving behind areas where conditions become inhospitable.

This pattern, often referred to as 'range shifting,' is typically marked by population growth at the leading edge and decline or extirpation at the trailing edge (Hampe and Petit, 2005; Matyas, 2010). This migration process has been well-documented in studies of post-glacial recolonization.

One of the major interests of studying tree migration under previous climatic changes is to analyze the tree migration rates. These rates facilitate the forecast of tree survival abilities under shifting environments. Pioneers palynology and genetic studies regarding the last transition between the glacial and interglacial periods have estimated that tree species survived the glacial period in Europe within three southern refugia, namely in the Iberian, Italian, and Balkan peninsulas (Bennett et al., 1991; Demesure et al., 1996; Dumolin-Lapègue et al., 1997). Based on these results, postglacial tree migration rates in Europe were evaluated up to 1000m/y. In contradiction, the empirical migration rates are found to be much smaller, under 100m/y. The divergence between these two rates is also known as the Reid's paradox of rapid plant migration (Clark et al., 1998). This mismatch was explained by the occurrence of long-distance dispersal events. Such events can lead to the establishment of small founder populations genetically supported via pollen gene flow. Later on, additional studies have highlighted the possibility that some populations, called cryptic refugia, survived the cold period further north, in Central Europe (Sally N. Aitken et al., 2008; Svenning et al., 2008; Tzedakis et al., 2013; Willis and Van Andel, 2004). In this case, migration rate estimates are considerably decreased. Nowadays, studies worldwide indicate that these rates range somewhere between 10 and 200 m/y (Sally N. Aitken et al., 2008; Gugger and Sugita, 2010; McLachlan and Clark, 2004).

Unfortunately, considering these revised rates coupled with few long-distance dispersal events, it is unlikely that tree populations will cope with anthropogenic climate change. The forecasted changes in temperature only would require migration rates somewhere between 1000m/y and 5000m/y (Davis and Shaw, 2001; Malcolm et al., 2002). Moreover, these results do not include the additional negative impact of extreme events or anthropogenic forest fragmentation (Resco De Dios et al., 2007).

## 3.4 Assisted migration

In the coming years, tree populations may face migration delay and adaptational lag in response to climate change. Assisted migration is being considered as a possible line of management to mitigate the negative impacts on forests (Palik et al., 2022). Assisted migration is the artificial transfer of forest reproductive material (i.e. seeds or seedlings) from one location to another.

On the one hand, from the historical perspective, this process was not recommended because it could lead to maladaptation (Campbell, 1979). This means that seeds performing very well in one environment might present a low fitness in another because they would be adapted to the first environment but not the second one. Additionally, the insertion of forest reproductive material at one location will alter local populations via gene flow (McKay et al., 2005).

On the other hand, maladaptation is a problem that the forest population will almost certainly face because of climate change (Bower and Aitken, 2008; Sáenz-Romero et al., 2021; St Clair and Howe, 2007). Furthermore, local sources of forest reproductive material might not always be the best. For example, populations that faced events such as founder effects or bottlenecks could be a poor seed source. Finally, our understanding of the relationship between biotic/abiotic factors and tree organisms is continuously expanding. For several decades, studies have been conducted to analyze the intricate relationship between tree populations and climatic variables. These experiments focus primarily on the identification of the most productive sources of forest reproductive material per climatic region (Hamann et al., 2011). Nowadays, the data are also used for a deeper understanding of these relationships; as will be discussed in the next sections.

Considering the factors outlined above, it becomes compelling to explore strategies beyond conventional approaches. Assisted migration could serve as a valuable method for safeguarding species that are particularly vulnerable to the anticipated changes (Bucharova, 2017; Sáenz-Romero et al., 2021).

## 3.5 Forest genetics and forest tree breeding

### 3.5.1 Forest genetics

Forest genetics is a multifaceted field that encompasses various branches of genetic study, including Mendelian genetics, quantitative genetics, and population genetics ((Burley et al., 2004)). At its core, Mendelian genetics in forestry revolves around Mendel's laws of heredity: the law of segregation and the law of independent assortment. These principles deal with the relationships between progeny and parents, the segregation of descendant generations, individual gene action, and the linkage of genes. Interestingly, Mendelism in forestry was rediscovered at the dawn of the 20th century (Bernardo, 2020; Hallauer, 2007). However, a significant revelation was that Mendel's laws do not extend to quantitative traits, marking a complex divergence in the understanding of genetic behavior within forest tree species. This blend of genetic concepts forms the foundation of forest genetics, a field that continues to unveil intricate details of hereditary patterns in forest trees.

In the paper titled "The correlation between relatives on the supposition of Mendelian inheritance" (Fisher, 1919), the author emits the hypothesis that quantitative traits are controlled by genes following the Mendelian law, but that continuous trait variation was due to many genes. From there the concept of quantitative genetics emerged. A significant feature was then established by Fisher regarding the variance of quantitative traits. This variance can be in its simplistic form separated into two components:

- the additive variance, that corresponds to the average effect of the respective alleles of a gene;

- and the environmental variance.

Population genetics deals with the changes in the genetic composition of populations across successive generations. These changes are due to the four following evolutionary processes: gene flow (migration), mutation, natural selection, and genetic drift. In the context of forestry, the insights provided by population genetics are invaluable. Population genetics

provides essential information used in tree breeding, offers perspectives on both historical and anticipated genetic shifts, and guides comprehensive forest management strategies (Burley et al., 2004).

Forest genetics is a specialized field of research concentrating on the genetic composition of trees, the factors that influence genetic variation, the heredity of particular traits, genetic conservation, and tree improvement strategies. Typically, tree populations exhibit high degrees of genetic diversity, originating from large population sizes, local adaptation, and neutral processes such as genetic drift and mutations (Isabel et al., 2020). Within the field of forest genetics, traits can be divided into two primary categories:

- Qualitative Traits: These traits show discrete variations and are usually controlled by one or very few genes. Most qualitative traits are minimally affected by the environment, making them simply-inherited and relatively straightforward to comprehend and manipulate. An example of a qualitative trait is the color of a flower's petals.

- Quantitative Traits: Unlike qualitative traits, quantitative traits exhibit continuous variation and are typically governed by a substantial number of genes. These traits tend to be highly influenced by environmental factors, rendering them more complex to analyze and manipulate. In the context of forestry, it is significant to note that the most economically valuable traits, such as tree height, diameter at breast height, survival, etc., are quantitative in nature.

Through its detailed examination of these genetic attributes and underlying processes, forest genetics plays a critical role in understanding and managing the vast and intricate genetic landscape of forest ecosystems.

Historically, tree improvement was the central focus of forest genetics and tree breeding, largely driven by the prospect of economic gain (Wheeler et al., 2015). However, a paradigm shift began to unfold in the 1980s, giving rise to an equally vital concept: the conservation and restoration of genetic diversity. This new focus emerged in response to the dual threats posed by a changing climate and the escalating intensity and occurrence of pests and disease

outbreaks (Sáenz-Romero et al., 2021; Wheeler et al., 2015). While economic considerations remain important, the broader understanding of forest genetics now recognizes the essential balance between improving economically valuable traits and preserving the genetic diversity that underpins forest resilience and sustainability.

## 3.5.2 Forestry breeding

Boosting productivity in forestry has long been a primary goal within the forest industry. Tracing back to the 13th century, silviculture emerged as the specialized branch of forestry dedicated to managing the growth, composition, and quality of forests to fulfill human needs. Within the realm of silviculture, tree breeding plays a crucial role, focusing on the manipulation of the genetic composition of tree populations.

Since the mid-20th century, large-scale tree breeding activities have been developed to identify superior individuals within tree populations (Alan, 2020; Ruotsalainen, 2014; White et al., 2007; Zobel and Talbert, 1984). This process involves meticulous testing of the selected individuals through the evaluation of their offspring in breeding trials. Individuals are considered superior not only based on their productivity levels but also on favorable characteristics such as height, straightness, pest resistance, and more.

This multidimensional approach to tree breeding underscores the complex interplay between genetics and environmental factors. It reflects the forest industry's evolving understanding of how to harness genetic potential to meet both economic goals and broader ecological considerations.

Tree breeding is not without challenges, they can be summarized as follows (El-Kassaby and Klapste, 2016):

- the need for large experimental plots because of the mature trees' sizes;
- requires testing over a wide range of environments (climate, soil composition, slope steepness, etc.) because trees are highly influenced by their environments;

- experiments have to be conducted over long periods because trees need time to reach sexual maturity and also because some economically important traits can only be measured after years of growth (e.g., wood quality);

- some experiments may be damaged by pests, diseases, and extreme events (e.g., floods, droughts, strong wings, etc.).

### 3.5.3 Heritability

As previously discussed, an organism's phenotype (observable traits) is distinct from its genotype (genetic makeup). Both the genotype and environmental factors contribute to the phenotype, a relationship mathematically expressed by the equation:

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2 \qquad (1.)$$

where $\sigma_P^2$ represents the phenotypic variance, $\sigma_G^2$ is the genetic variance, and $\sigma_E^2$ is the environmental variance. The total genetic variance can be further broken down into:

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2 \qquad (2.)$$

where $\sigma_A^2$ corresponds to the additive genetic variance (where both alleles in a locus contribute equally to a trait), $\sigma_D^2$ correspond to the dominance genetic effect (interactions between alleles at a given locus), and $\sigma_I^2$ corresponds to the epistatic genetic effects (interactions among alleles from different loci).

In the context of forest tree breeding, heritability is the essential concept. It is defined as the ratio of genetic variance to the phenotypic variance It's a valuable measure to predict the potential gain from breeding. There are two types of heritability (Visscher et al., 2008):

- the broad-sense heritability $H^2$, which refers to the ratio of the total genetic variance (additive, dominance, and epistatic) to the phenotypic variance:

$$H^2 = \frac{\sigma_G^2}{\sigma_P^2}$$
(3.)

- and the narrow-sense heritability $h^2$ (Hill and Mulder, 2010), which refers to the ratio of the additive genetic variance to the phenotypic variance:

$$h^2 = \frac{\sigma_a^2}{\sigma_P^2}$$
(4.)

## 3.5.4 Genotype by Environment interaction

In reality, the phenotypic variation is more intricate than simply the sum of the genetic and environmental variances (Albecker et al., 2022; Ørsted et al., 2018). Another vital concept in forestry breeding genetics is the GxE interaction. This interaction highlights two main challenges. First, it emphasizes that phenotypic expression varies among genotypes across different environments, leading to a change in rank that makes it problematic to infer the ranking of a set of genotypes in multiple environments based on a single site. Second, GxE also involves the problematic issue of changes in variance across multiple environments, adding another layer of complexity. Such variations underscore why GxE interaction must be carefully assessed, often employing MET mixed models (Malosetti et al., 2013). Consequently, to accurately evaluate and select forest reproductive material, it is essential to conduct multi-site experiments that cover the range of planned planting environments, fully accounting for the multifaceted nature of GxE interaction.

## 3.5.5 Environmental variance and experimental layouts in forest tree breeding

As previously discussed, phenotypic variance is not solely a product of genetic factors but is also heavily influenced by environmental variance. Within a plantation trial, environmental conditions can vary significantly across different locations, affecting factors such as sun exposure, soil composition, terrain steepness, drainage level, and more. When evaluating populations from different origins (known as provenances) or specific genotypes (such as full-sib or half-sib families), their physical placement across the trial site can greatly influence the assessment of their respective performances. Given this complexity, it becomes vital to consider and account for spatial site variation in the analysis (Leites et al., 2012a; Rehfeldt et al., 2018). One approach is to standardize the response variable values within the site, essentially leveling the playing field for comparative analysis. However, this method has faced criticism within the scientific community, particularly for its potential to erase essential distinctions among response functions (O'Neill et al., 2007). An alternative and often complementary strategy involves meticulously organizing the test site according to a specific design. This design-oriented approach aims to minimize the confounding effects of spatial variability, thereby allowing more precise and unbiased estimation of the genotypes' performance. Together, these considerations underscore the intricate interplay between genetic and environmental factors in forestry genetics. They highlight the need for carefully crafted methodologies that acknowledge and address the spatial variations within a trial site, ensuring that the conclusions drawn are robust, accurate, and reflective of the true genetic potential of the populations under study.

Various design methodologies have been employed to structure test sites for forest genetic experiments, each with its unique characteristics and applications. A commonly observed approach is the randomized complete block design with $n$ replicates. In this model, the test site is segmented into blocks, with each block containing a representative of every provenance or genotype. These are then planted in random positions within the block.

However, when the study involves a large number of groups, it may become impractical to include every group within each block. In such instances, an incomplete block design might

be implemented. Another variation includes incomplete Latin square designs. These incomplete designs, while useful, can sometimes lead to imbalances, with certain groups being underrepresented (Wang et al., 2010).

When the experiment involves the study of more than one factor, a split-plot design might be employed to capture the interactions and complexities of multiple variables. It's worth noting that, in the majority of these experimental designs, the stands usually require no intervention after being set up. However, there might be situations where tree density reduction is performed after several years to manage specific conditions (Chakraborty et al., 2018).

These multifaceted design approaches reflect the nuanced requirements of forest genetic studies. They ensure a thorough and balanced representation of various genetic groups and accommodate different experimental needs, all while recognizing the potential challenges and mitigating them through thoughtful planning and execution.

At the block level, the number of trees per sample can vary, and so can the spacing between the adjacent trees. In some studies, we can find only one tree representing each group while others can follow a $n$ x $n$ pattern (McLane et al., 2011; Wu and Ying, 2004). Provenances can be represented with up to 50 trees per plot (Foff et al., 2014). Some studies have a homogenized type of design across all the test sites(McLane et al., 2011; Wu and Ying, 2004) and some do not (O'Neill et al., 2014; Persson, 1998). Different experimental designs can lead to additional variation among test sites.

Several researchers have delved into the analysis of variation at the test site level within the field of forest genetics. For instance, Kapeller et al. (2017) employed a mixture-model analysis to visualize the density distribution of tree heights, distinguishing between small and tall trees. Through this analysis, they discovered a greater degree of variation within warm and dry sites compared to the cold and moist sites. However, an intriguing inconsistency arose when the authors calculated the coefficient of variation for each site. These results contradicted the findings from the mixture-model analysis, highlighting a complex interplay between different statistical methods and real-world observations. Additionally, the study found that trees tend to be taller in warmer sites compared to colder ones, a conclusion that aligns with

the findings of other research in the field (Leites et al., 2012b). This corroborative evidence adds to the body of knowledge supporting the role of temperature as a key factor in influencing tree growth and height.

## 3.6 Linear regression

Linear regressions are a popular type of model in all fields of study. One of the reasons is that it is easy to fit and interpret. Additionally, linear regressions have been widely used (Alberto et al., 2013; Carrasco et al., 2022; Gunderson et al., 2010; Román-Palacios and Wiens, 2020; Vitasse et al., 2009), hence it is a well-studied and well-validated method.

In linear regression models, distinctions are made between various types of variables, each playing a specific role. The response variable, which must be continuous, is the outcome or dependent variable we are trying to predict or explain. Meanwhile, explanatory/predictor variables, which may be either continuous or categorical, are used to interpret or anticipate variations in the response variable. These explanatory variables can be further categorized based on their influence on the response variable:

- Variables with Fixed Effects: These are variables assumed to exert a consistent influence on the response variable. The term "fixed" refers to the fact that the effect of these variables is constant for all individuals or groups within the population under study. Essentially, variables with fixed effects are those where the specific levels or categories are of primary interest, and their effect on the response variable is unchanging.

- Variables with Random Effects: In contrast, variables with random effects are those assumed to have a varying influence on the response variable across different individuals or groups. These variables are used when the interest lies in understanding the underlying population rather than the specific levels of the variable themselves. The "random" nature of these variables allows for the accommodation of variations that might not be captured by the variables with fixed effects.

A model that incorporates both variables with fixed effects and variables with random effects is known as a mixed linear model. By employing this comprehensive approach, researchers can analyze complex relationships with greater flexibility, particularly when dealing with hierarchical or multi-level data (Borenstein et al., 2010; Hunter and Schmidt, 2000; White et al., 2007). Linear regression is based on a linear relationship between the explanatory and response variables. The basic linear equation is as follows:

$$Y = \beta_0 + \beta_1 X \qquad\qquad (5.)$$

where:

- $Y$ is the response variable;
- $\beta_0$ is the intercept;
- $\beta_1$ is the slope;
- and $X$ is the explanatory variable with a fixed effect.

Based on this model, one can estimate both $\beta_1$ and $\beta_0$ that are the least-square estimates and they are typically referred to as coefficients. Then, using these two components and the values of the explanatory variable, it is possible to model (predict) the response variable.

The equation for a basic mixed linear model, that also includes a random term can be written as follows:

$$Y = \beta_0 + \beta_1 X + \gamma \qquad\qquad (6.)$$

where $\gamma$ is the random effect variance.

In theory, the variability of the response variable can be entirely explained by the explanatory variables, in that case, the above models are complete. However, in practice, when fitting a linear or a mixed linear model, the selected explanatory variables cannot explain the whole variability. Indeed, the effect of the environment includes many different factors that

cannot all be included in the model. Additionally, when sampling phenotypes, a certain level of measurement error is to be expected. Furthermore, the predictors' values can also include errors in measurement or estimation. For these reasons, the proper way to describe the relationship between response and explanatory variables is to add a random error term to the equation such as:

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{7.}$$

where $\varepsilon$ is the residual error term.

With linear regression models, there are two ways to calculate the parameters. The first one is the Ordinary Least Square method (OLS) which is considered to be a deterministic method. This method minimizes the sum of squares of the differences between the observed and estimated values. It does not take into account the residual structure. When the error structure can be reliably characterized, it is possible to use the Maximum Likelihood Estimation method (MLE) which can be used for both linear model and mixed linear model. It aims at maximizing the probability (likelihood, thus referred to as the "maximum likelihood") of obtaining the observed values with the model parameters. In other words, this method estimates the parameters that make the observed data the most likely (i.e., provide the best fit). MLE provides estimates of the fixed and random effects simultaneously assuming an identical likelihood function, however, this method often provides biased variance estimates of random effects. In linear regression, fixed effect coefficients are called Best Linear Unbiased Estimates (BLUEs) (Luo, 2017).

In the context of mixed linear models, the estimation of model parameters often employs the Restricted Maximum Likelihood (REML) method. Unlike the Maximum Likelihood Estimation (MLE), which simultaneously estimates both fixed and random effects, REML takes a sequential approach. First, it estimates the variances associated with variables with random effects. Following that, it uses these variance estimates to derive the coefficients for variables with fixed effects.

The REML method offers several advantages. Primarily, it provides unbiased estimates for random variances, making it a preferred choice when the main focus of the analysis is on variance components. In comparison, MLE is commonly used when the objective is to compare models with different fixed effects, as REML's specific calculation process makes it unsuitable for comparing models with varying site effects. Furthermore, REML tends to be more robust against violations of model assumptions than MLE, adding to its appeal in certain situations. Within the framework of mixed linear regression, the estimates for random variance components derived from the REML method are known as Best Linear Unbiased Predictions (BLUPs) (Quaas et al., 1984).

### 3.6.1 Assumptions of the linear models

The linear regression models are based on five key assumptions that must be met (Poole and O'Farrell, 1971):

- random sample, where the observations have to be independent of one another;
- independence between the explanatory variables;
- homoscedasticity, where the model residuals have to be constant;
- normal distribution of residuals;
- a linear relationship between the explanatory and response variables;

Additionally, several other factors are necessary to make valid inferences:

- the sample has to be representative of the population;
- when making predictions, it is better to stay within the studied range, extrapolating is not recommended (Masana and Baranyi, 2000).

### 3.6.2 Model evaluation and model comparison

When developing a model, it is possible to use any type of predictor for a response variable. However, many of them would not be relevant, so we have to choose which ones are the most suitable. First, one has to use logic and select variables that are likely to be influencing

the response (e.g., temperature to explain tree growth). Second, it is possible to evaluate how well the model fits the data. Finally, when considering several models, there are several statistical ways to compare them.

## 3.6.2.1  Model evaluation

There are two common ways to evaluate the goodness of fit of a model. The first one involves the analysis of residual diagnostic plots (Butler et al., 2017). It is possible to obtain the residuals' histogram to verify that they follow, to a certain extent, a normal distribution. It is also possible to plot a normal Q-Q plot that can also be used to assess the normality of the residuals. When the residuals are aligned on the diagonal line, it means that they follow a normal distribution. Then there is the plot fitting the residuals against the fitted values, which is used to assess the homogeneity of variances between observed and predicted values. There is homogeneity when the residuals form a uniform horizontal band. Finally, there is the plot fitting the residuals against the unit number (each measurement) that is similar to the residuals against fitted values plot, but the residuals are ordered by unit (their ID number) instead of by the predictor's values. The residual plots can also be used to determine the potential presence of outliers that could be removed from the analysis.

The second one is to use both the coefficient of determination $R^2$ and the $p$ values of the model. The $R^2$ determines the proportion of variance of the response variable that is explained by the explanatory variable (Kapeller et al., 2017; Ørsted et al., 2018). The higher the $R^2$, the better the model seems to fit the data. The model's $p$ values help to determine if the relationships observed in the tested sample also exist at the larger population level. For each predictor, the null hypothesis, which states that there is no relationship between the predictor and the response variables, is tested. If the obtained $p$ value is inferior to the assumed significance level (e.g., 0.05 or 0.001), it indicates that there is sufficient evidence to reject the null hypothesis and vice-versa.

3.6.2.2 Model comparison

One common way to select a model is to choose the one that minimizes the squared difference between the observed values and the estimated values (Malosetti et al., 2013). This is called the least square method. Another way is to use the $R^2$. Generally, a higher $R^2$ indicates a better model. Other ways, when comparing nested models (when one model is an extension of the other; full versus reduced models), are the F-test, likelihood ratio test, Akaike information criterion (AIC), and lastly the Bayesian information criterion (BIC) (Chakraborty et al., 2018; Joyce and Rehfeldt, 2017; Lu and Zhang, 2010; Oliveira et al., 2020; Sáenz-Romero et al., 2017). The F-test is a hypothesis test with a null hypothesis stating that the full model does not explain the variance in response variable better than the reduced model. The log-likelihood ratio test is based on the ratio of probabilities to obtain the observed data between the full and the reduced model. AIC and BIC are based on the log-likelihood of the model and the number of explanatory parameters. As the log-likelihood is always higher when incorporating new parameters, both AIC and BIC are attributing penalties for each additional explanatory variable. Higher values are indicative of better models.

## 3.7 Multi-environmental trials analyses

In a regular historical context, tree populations are typically adapted to their environment. This means that within the same tree species, populations are displaying different levels of local adaptation (e.g., to frost, drought, soil type, etc.) across its geographical range. Hence, for optimal top phenotype selection in tree breeding, it is necessary to perform MET analyses to test provenances and genotypes in a set of trials covering the environmental range for which the seeds are intended. Additionally, the notion of "locally adapted" infers that one population performing best in one area can be outperformed in other parts of the species' climatic range which is a form of GxE interaction (Matyas, 1996).

However, the "local is the best" statement is not always true (Rehfeldt et al., 2018; Schmidtling, 1994; Wang et al., 2006a) for reasons such as the population history, adaptational

lag, and interaction between interspecific competition and local adaptation to climate (Leites et al., 2012b; Matyas, 1996). In several occurrences, a population originating from another area can outperform the local population (Matyas, 1994). This population can originate from a similar climate and is better adapted, or it is a population that generally performs better, than the local one.

These provenance trials can also be used to analyze the effect of climate change on tree populations (Matyas, 1994; Rehfeldt et al., 2001; Wang et al., 2006b). Indeed, a provenance that is tested within different climatic conditions, if its local climatic conditions were to change towards similar ones, can hint at what would be its performance. Unfortunately, since provenance trials were not specifically designed for this type of analysis, scientists are facing a major issue. It lies in the fact that breeding trials are usually established to identify top populations or genotypes for afforestation projects within the species range. Hence, a limited amount of tests are conducted at the boundaries of the species range, and close to none are established further. This leads to model inaccuracies when studying the effect of conditions situated outside of the species' ranges (Wang et al., 2006b). There are two other issues regularly faced by foresters and scientists when studying provenance trials. The first one is the occurrence of poor testing designs, and the second one is the lack of systematic and regular sampling. Both of these issues lead to the need for more complex analyses and larger results uncertainties.

Nowadays, several studies have been done using mature tree provenance data trials; and those are the experiments I will focus on within this literature. However, the same analyses can be performed on seedlings' provenance trials which were specifically designed within the climate change framework. In these trials, the designs allow simpler and more efficient analyses. However, trees have long life spans and these trials are still at a seedling/young trees stage which may not reveal their long-term performance (Frank et al., 2017a; Matías et al., 2016; Robakowski et al., 2005).

Besides the MET analyses, it is also possible to use empirical observations of natural stands over the whole species range to predict the impact of climate change. In this type of

study, we identify the differences in performances between populations and correlate them to the climatic conditions (Vitasse et al., 2009). However, with this method, we assume that "local is the best" because each population is tested only once, in their natural environment. Hence, it is not possible to know if they would perform better somewhere else or if they would be outperformed in their natural environment by other populations.

Furthermore, in some areas, there are already tangible records of the effect of climate change such as an increase in drought events (Dai, 2011; Dai et al., 2018; Walther et al., 2002). Examining the direct consequences of climate change enables us to forecast, within certain boundaries, how individual populations and entire species may be affected.

When studying the impact of climate change on tree species and identifying suitable seed sources for new afforestation projects, it is possible to use this information to redefine seed delineation patterns and breeding zones based on future climate scenarios and not on current zone or country borders. Adjusting seed delineation patterns and breeding zones may help mitigate the predicted harmful impact of climate change on forests (Matyas, 1994; Wang et al., 2006a). Furthermore, reviewing seed delineation patterns could also result in a productivity boost in some areas (Rweyongeza et al., 2007).

### 3.7.1 Studied traits

The focus of this literature review, as previously mentioned, is mostly on European mature tree data from tree breeding provenance trials. While not the most suitable material, for reasons mentioned in the previous section, studying the impact of climate change is a recent field of study and many years are necessary to establish and evaluate suitable experiments specially designed for MET analyses. During the 20th century, many provenance trials have been implemented in Eurasia and North America for several economically valuable tree species (e.g., *Picea abies* – Norway spruce, *Fagus sylvatica* – European beech, *Larix decidua* – European larch, etc.).

The purpose of provenance testing is to select populations and genotypes that are performing well across a specific range of environmental conditions. Accordingly, we want to study traits that are genetically regulated. Height is the most common trait studied (Gömöry et al., 2012; Kapeller et al., 2012; O'Neill et al., 2007; Rehfeldt et al., 2001; Schueler et al., 2013; Thomson and Parker, 2008; Wu and Ying, 2004). While it is only heritable around 20 to 30%, it is the most heritable growth-related trait for many tree species (Wang et al., 2010). Height is representative of the overall tree productivity as well as competitive ability, and it has been established that it has a reliable response to selection (Kapeller et al., 2012; Wu and Ying, 2004). Other commonly studied traits are survival (Rehfeldt et al., 2001; Sáenz-Romero et al., 2017; Schueler et al., 2013), diameter at breast height (Foff et al., 2014; Rehfeldt et al., 2001; Thomson and Parker, 2008), tree ring data (McLane et al., 2011; Suvanto et al., 2016), basal area per hectare (Chakraborty et al., 2018), photosynthetic capacity, time of bud set, and more.

## 3.7.2  Climatic variables

Climatic variables are found to be relatively well correlated with the variation in growth and phenological traits, indicating a certain adaptive pressure. Interestingly, traits can be independently affected by climate and thus show different patterns of adaptation within the same provenance (Matyas, 1996).

When selecting climatic variables, it is usually best to obtain yearly averages over periods that are representative of the length of the studied trials. The climatic data can be obtained from different sources. One of these sources is to use weather station records. In this instance, the data can be extracted from the closest station (Rehfeldt et al., 2002) or from many stations situated around the test and provenance sites (Hasenauer et al., 1999). Using only one or relatively distant stations can be problematic in regions with steep climatic gradients. For example, in mountainous areas, the climate is changing quickly along the altitudinal gradient and several kilometers can easily make a considerable difference.

Climatic data can also be extracted from climatic models. One such model is the WorldClim, which has a 30 arc-second (app. 1km) spatial resolution all over the global land

areas (Hijmans et al., 2005). Another model is the ClimateBC which covers most of North America with a 150 arcsec resolution (approximately 5km) covering a period from 1961 to 1990 (Wang et al., 2006a). Additionally, when estimating the potential impact of climate change, researchers are using predictive climate models, that are often developed for different periods (e.g., 2020s, 2050s, and 2080s) and different potential scenarios. EURO-CODEX is a European (approximately 12.5km) climate change model (Jacob et al., 2014). Due to the climatic uncertainties of the future, predictive models tend to have a lower resolution. As for the weather stations, coarse resolutions of climatic models also lead to issues linked with steep environmental gradients. Interestingly, some models are considering this issue. For example, the ClimateBC was developed with a point value in the middle of each tile of the grid with a scale-free gradient between points to avoid large steps between tiles (Wang et al., 2006a).

Several climatic data are used in the literature, with temperature variables often having the greatest correlations with the response variables (Matyas, 1996). Commonly found temperature-related variables are included, such as the mean annual temperature, the mean temperature of the coldest or warmest month, the sum of degree- days superior to 5°C, and the sum of degree days inferior to 0°C (Leites et al., 2012b; Rweyongeza et al., 2007; Schueler et al., 2013). Even though they often show milder correlations (Kapeller et al., 2017), precipitation variables are also found to be significantly correlated with growth traits. A common precipitation variable used is the mean annual precipitation. In the literature, we also find studies using variables combining temperature and precipitation components, such as the annual heat moisture index (Chakraborty et al., 2015; Kapeller et al., 2012; Wang et al., 2006b). Furthermore, geographic factors associated with the climate, such as latitude, longitude, or altitude, are also often found in the literature (Kapeller et al., 2017; Matyas, 1996). There is a large number of potential variables that can be used in the MET analyses; in some cases, only one variable is selected, while in other cases, several can be used (O'Neill et al., 2014). When selecting more than one variable, it is important to test the correlation among them. Indeed, collinear variables used in the same model cannot independently predict the value of the response variable. This signifies that part of the variance in the response variable can be

explained by both variables, consequently, it may result in a reduction of their statistical significance. When two variables are collinear, it is preferable to remove one of them. For example, temperature or precipitation variables are often highly collinear among themselves and will often yield similar results and explain the same variation (Matyas, 1996). There are several ways to reduce the number of variables to obtain a set of relatively independent ones. One way is to use a stepwise multiple regression procedure, which fits a set of regression models that will automatically eliminate variables based on prespecified criteria (Schmidtling, 1994; Wang et al., 2010). One of the criteria is the significance level required for the variable to be entered into the model, and the second one is the required significance level for the variable to remain in the model. Wang et al. (2010), for example, have set these values to 0.01 and 0.005 respectively. Another way to avoid confounded variables is to use principal component analyses (PCA) that allow the integration of several variable effects (Kapeller et al., 2012; Matyas, 1994; Rweyongeza et al., 2007). When using a PCA with temperature and precipitation variables, the first component, which explains the most variability in the data, is usually composed of temperature variables while the second one is often composed of precipitation/drought variables. An example from Kapeller et al. (2017) is presented in Fig. 3.1. Interestingly, we can see that altitude is negatively correlated with the other temperature-related variables. This makes sense as an increase in altitude levels usually translates into lower temperatures.

In some studies, authors have used the principal components' values as ecological distance (Matyas, 1994). In this case, they calculated the differences between sites' values and used them as ecological values. In this situation, the greater the difference between the values of two sites, the greater the ecological value, and the greater the difference between environments. The ecological distance represents a measure of the difference between two locations (i.e., the origin of the provenance or the test site), and it can also be calculated differently. Gömöry et al. (2012), for example, have used altitude levels as ecological value.
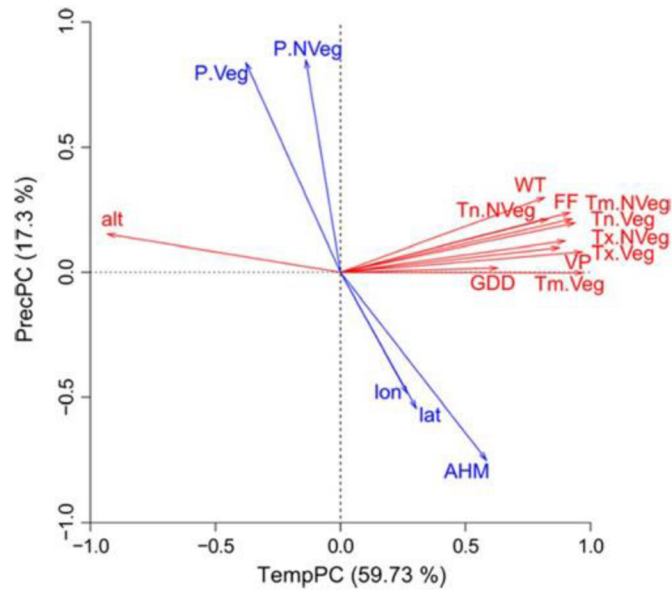
**Fig. 3. 1** Results of a PCA using a set of 14 climatic parameters. In this plot, the temperature-related factors are colored in red, and the precipitation-related ones are colored in blue (Kapeller et al., 2017).

### 3.7.3 Other environmental variables

To study the impact of climate change, it is logical to focus on climatic variables. However, as was mentioned previously, tree population adaptation to the environment also includes adaptation to other abiotic factors (e.g., soil composition, steepness, sun exposure, etc.) and biotic factors (e.g., specific symbiosis with microorganisms or insects)(McLane et al., 2011; O'Neill et al., 2014; Sáenz-Romero et al., 2021). Taking into account these factors when modeling tree adaptation to climate is advantageous for at least two reasons. First, adding other types of environmental variables than temperature and precipitation-related ones in the model will likely increase its overall accuracy as well as the estimation and significance accuracy of all the predictors. Second, when considering moving populations outside of their current climatic adaptational range, we are considering moving them to several other new conditions (e.g., soil type). Furthermore, other factors such as windfall or pest damage will likely influence the analyses' results (Sáenz-Romero et al., 2021). Hence, considering all the available environmental factors in MET analysis is advisable.

45

### 3.7.4  Fixed and random effects in MET

In tree breeding, the linear MET models usually include the site effect (environmental effect) as a fixed effect. The site effect can be included in the model either as a categorical variable, with each site having its own effect, or as a continuous explanatory variable (e.g., mean annual temperature, ecological distance, etc.) (Foff et al., 2014; Poupon et al., 2021; Rehfeldt et al., 2018; Wang et al., 2010). Assuming the simple linear equation mentioned earlier (Equation (5.)), the model will provide BLUEs of the fixed site effects. When we use each site as a different category, the model output will give a site mean for each site (Cullis et al., 2014; Poupon et al., 2023). Then, it is possible to estimate the response variable in site $s$ as the sum of the intercept, also called the overall mean, and the mean of site $s$. When we use a continuous variable as a site effect, the model will provide the estimate as a single component. In this case, it will estimate the value of the response variable for a specific value of the explanatory variable; thus it is possible to sum the overall mean with the component value multiplied by the explanatory variable value. When using explanatory variables, it is not possible to encompass the whole site variability, however, the results are more easily interpretable and it is possible to predict performances on different sites.

In tree breeding, MET mixed linear regression models can have several components included as random. They notably include the population or genotype effect. When studying a set of genotypes (e.g., full-sib or half-sib families), it is possible to incorporate the pedigree information into the model. Additionally, I mentioned that one way to take into account the spatial variation within a breeding trial is to organize it according to a specific design (e.g., blocks). In this case, the information is entered into the model as a random effect as well. Finally, the GxE, which refers to the variation in performance ranking across trials, is also included as a random effect.

## 3.8 Response function

The development and analysis of response functions is a recent field of study connected with the increasing interest in understanding and mitigating the impact of climate change on forests (Rehfeldt et al., 2018; Sáenz-Romero et al., 2021; Suvanto et al., 2016). This methodology uses provenance trials population data. The Response Function (RF) approach technically includes response functions and transfer functions analyses. In both analyses, there is a heritable trait (e.g., height, diameter at breast height, or photosynthesis level) plotted and statistically analyzed as a function of an independent climatic variable (e.g., mean annual temperature or annual heat-moisture index) using linear regression fitting. However, the first one analyses the performance of several provenances at one planting site while the second one analyses the performance of one provenance at different planting sites. More recently, both functions have been combined into Universal Response Functions (URF) that take into account both the environmental effects, with the planting site climate, and the genetic effects, with the seed source climate (Chakraborty et al., 2018; O'neill et al., 2008; Wang et al., 2010). The URF model enables the visualization of climatic limits of seed transfer for populations since it allows the estimation of a trait response for any seed source at any location under specific climates.

There are several variants used in the elaboration of RF. First, several kinds of data, traits, and climatic variables are available and can be used. Second, several types of mathematical functions can be used for the modeling, depending on which best fits the data. Two kinds of functions are often used in RF studies. In certain instances, authors have chosen to employ a linear function, as detailed in works by Schmidling et al. (1994), Matyas (1984), Kapeller (2016), and Frank (2018). Alternatively, there are cases where either quadratic or Gaussian functions, both of which exhibit a parabolic shape, may be selected (Gömöry et al., 2012; Kapeller et al., 2012; Sáenz-Romero et al., 2017). Furthermore, in some situations, it can be beneficial to manipulate the data, such as through age correction or data transformation, before engaging in RF analysis. Lastly, there exists a variety of analyses that can be conducted in conjunction with RFs to enhance the overall comprehension of the data. The linear function's shape is quite different from the two other functions listed above, but this can be explained as

follows: if we were to imagine an example where we have a provenance's growth data available from sites covering the whole studied species range, then when looking the trait response, we will observe a parabolic response (Schmidtling, 1994). Indeed, one population usually has a climate for which it grows best, and deviations from this climate, in both directions, will translate into a decrease in productivity. However, if we only plot test sites with a smaller environmental range, the observable response might only represent one side of the parabola, and this will be best represented by a linear fitting. It is possible to test the different functions to determine which one fits the data best (Kapeller et al., 2012). When evaluating the goodness of fit, the authors usually refer to the $R^2$ and to the $p$ value of the curve's fitting.

### 3.8.1 Quadratic function

A quadratic function, which refers to a polynomial function of degree two, is a type of linear regression. The model equation can be written as:

$$y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon \qquad (8.)$$

where:

- $y$ is the response variable;
- $\beta_0$ is the intercept;
- $\beta_1$ and $\beta_1$ are the model's coefficients;
- $X$ is the predictor variable;
- and $\varepsilon$ is the residual error.

In a study by Gömöry et al., (2012), the authors used a Norway spruce provenance test, that included five test sites and fifteen provenances and that was situated along a strong altitudinal gradient. In this provenance trial, several traits (e.g., height, survival, diameter at breast height, and tree volume) were measured, and height was measured at five different ages between 10 and 45 years. The authors used quadratic modeling to fit each trait (and each age)

48

with several explanatory variables (e.g., temperature, precipitations, growing degree days, etc.) used as the eco-distance.

In Table 3.1, we can visualize the optimums' average transfer rates (based on the polynomial curves' peaks) for the spruce provenances; it indicates that, generally, the transfer of populations to lower and warmer altitudes and with less precipitation levels would increase productivity. The use of several measurement ages for height led to interesting results. Indeed, it seems as if the growth was influenced differently by the tested environmental variables depending on the measurement age. Generally, the differences between the optimum conditions and the provenances conditions seem to decrease from age 10 to 38 to then increase again. Another interesting result is that, for both survival and diameter at breast height, none of the quadratic regression models were significant.

**Table 3. 1** This table reports the RF's average results of all provenances where the optimal transfer rate values of the explanatory variables are indicated for each trait (and each age). The values in bold represent the significant quadratic fittings (Gömöry et al., 2012).

| Trait | ALT (m) | $P_{JUL}$ (mm) | $P_{VEG}$ (mm) | $P_Y$ (mm) | $T_{JUL}$ (°C) | $T_{VEG}$ (°C) | $T_Y$ (°C) | $GDD_0$ | $GDD_5$ | $GDD_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $H_{10}$ | **−317.6** | **−36.6** | −190.9 | −244.4 | **2.476** | **2.356** | **1.736** | 21.3 | 25.8 | 38.1 |
| $H_{15}$ | **−214.1** | −23.6 | −73.6 | 168.4 | 2.192 | **2.021** | **1.505** | 17.2 | 19.7 | 27.0 |
| $H_{20}$ | **−161.9** | 10.7 | −236.5 | −234.7 | 1.605 | 1.502 | **1.138** | 12.7 | 14.3 | 20.2 |
| $H_{38}$ | **−147.3** | **−27.1** | **−137.9** | **−139.5** | 1.506 | 1.352 | 0.945 | 9.9 | 11.8 | **15.2** |
| $H_{45}$ | **−242.9** | **−27.6** | **−150.9** | **−168.9** | 1.877 | **1.751** | **1.264** | **15.0** | 17.7 | 23.4 |
| $D_{45}$ | 55.8 | −6.0 | −15.1 | 72.3 | −0.017 | −0.093 | −0.156 | −3.4 | −4.0 | −7.3 |
| $V_{45}$ | −109.0 | −16.3 | −85.1 | −43.3 | 1.087 | 0.986 | **0.668** | 7.7 | 9.5 | 10.8 |
| $S_{45}$ | 20.9 | −222.5 | −147.9 | −95.7 | 3.251 | 2.453 | 1.792 | 14.0 | 28.5 | 66.2 |

Additionally, in their study, the authors found that, at the provenance level, the results indicate that provenances from the driest sites would benefit from more precipitation while the opposite trend is observed for populations originating from the wettest sites.

### 3.8.2 Gaussian function

The Gaussian regression fitting is using the following equation:

$$y = ae^{\frac{-(X-b)^2}{2c^2}} \qquad (9.)$$

where:

- $y$ is the response variable;

- $X$ is the predictor variable;

- and $a$, $b$, and $c$ are the model parameters.

When utilizing the Gaussian function for the development of RFs, the parameters can be interpreted in an ecologically meaningful way (Kapeller et al., 2012). Indeed, the parameter $a$ corresponds to the maximum value of the response variable (i.e., response variable value at the top of the parabola), the parameter $b$ corresponds to the predictor's value for which the response variable is at its maximum (i.e., predictor's value at the top of the parabola), and parameter $c$ corresponds to the standard deviation of the Gaussian distribution and can be interpreted as an index of the climate sensitivity (the higher the value, the wider the curve, the lower the climate sensitivity).

In the study by Kapeller et al. (2012), the authors used measurements at the age of 15 of 379 provenances planted across 29 test sites. They developed Gaussian response functions for each test site using the annual heat moisture index as the explanatory variable. Their results demonstrated that growth performance is affected by both the region of origin (genetic adaptation) and/or the planting site's climatic conditions (environmental effect). This was also highlighted by Schueler and Kapeller (2010). The results indicate a rise in productivity in conjunction with an increase in the heat moisture index, as demonstrated by two standard provenances in Fig. 3.2. However, it appears that beyond a heat moisture index of approximately 30, the response variable reaches a plateau. At this point, further increases in the heat moisture index do not correspond to any additional growth in tree height.
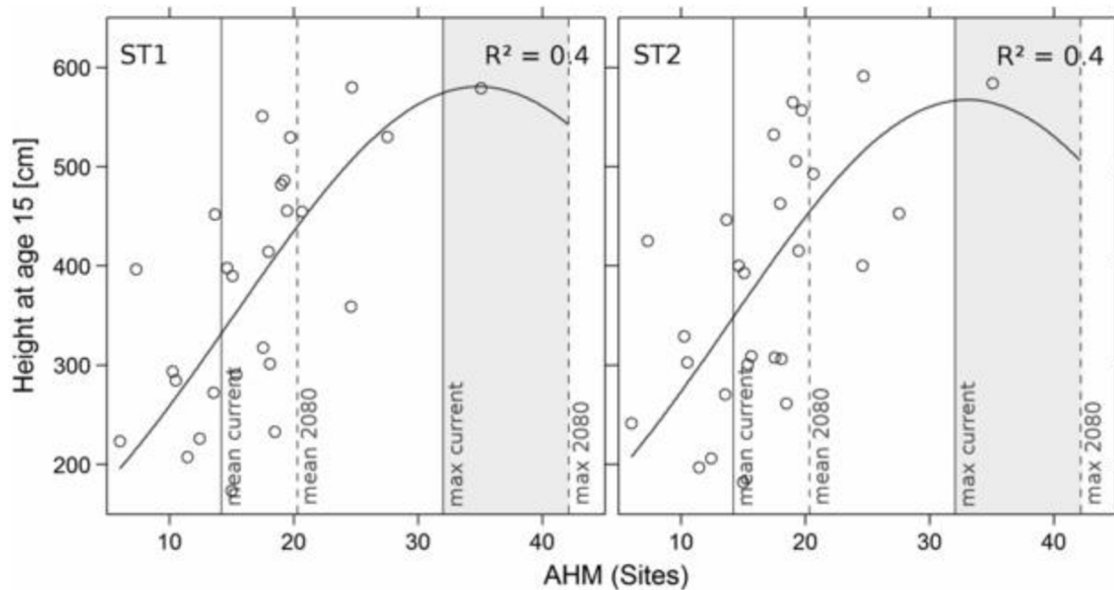
**Fig. 3. 2** Annual heat moisture index response functions of two standard provenances (ST1 = Schneegattern; ST2 = Murau) of Norway spruce using the 90 % percentile. The vertical lines are indicating mean and maximal annual heat moisture index values for Norway spruce in Austria, for both the present and the forecasted future climate (Kapeller et al., 2012).

The authors also used the results from the response functions to analyze how the predicted climate change is expected to affect Norway spruce in Austria. They predicted that, up to 2080, an increase of productivity is expected by up to 45% and could increase by 11% more by optimizing seed source management. Another interesting result from their study is that, while survival rates varied significantly among planting sites, they found no significant differences based on the populations' origins.

Another study, performed based on the same provenance trials, investigated the instar-specific genetic variation of Norway spruce in response to climate (Schueler et al., 2013). Gaussian response functions were also fitted with heights as a function of the annual heat moisture index. The authors found that the climate at the planting site and the region of origin strongly affected the height at the age of 15 with correlation coefficients ranging between 0.61 and 0.69. In Fig. 3.3, we can see that the variability among provenance clusters is increasing with an increase in the predictor variable. The author expects that this trend will continue to increase further in the future.

51
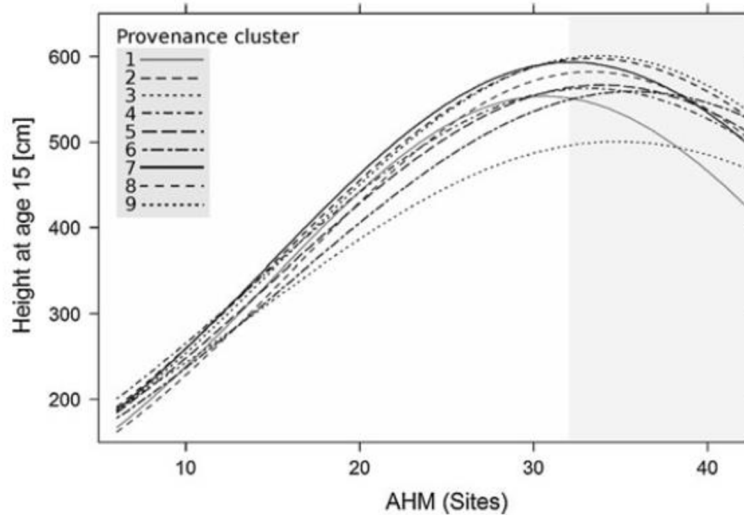
**Fig. 3. 3** Annual heat moisture index response functions of provenance clusters for Norway spruce using the 90 % percentile. The area in grey denotes values of the annual heat moisture index that are beyond the current maximal limits (Schueler et al., 2013).

Based on these results, they determined the two clusters that are expected to perform the best in future conditions based on the forecasted changes in the annual heat moisture index. In the study, they also focused on analyzing the location of the genetic conservation units based on the forecasted best-performing provenance clusters. They found that they are relatively rare in the regions from which these clusters originate. Additionally, they found that the most genetically diverse populations were not always represented by a high number of conservation units. This indicates that more genetic conservation units could be established with regard to climate change.

### 3.8.3  Linear function

Several studies utilized the linear RF of the form:

$$y = \beta_0 + \beta_1 X \tag{10.}$$

where:

- $y$ is the response variable;

- $\beta_0$ is the intercept;

- $\beta_1$ is the slope;

- and $X$ is the fixed effect variable.

In an early RF study (Mátyás, 1994), the authors plotted both types of RFs: response functions and transfer functions. They used data from three Jack pine provenances planted in seven different sites and measured at the age of 15. They were able to fit both the RFs of the three provenances in each site and the transfer functions of the three provenances across the seven sites. They used polynomial fitting for the former, with three reference points, and linear fitting for the latter with seven reference points. As we can see in Fig. 3.4, they used both types of function to develop a three-dimensional model that can be considered as an early attempt at modeling a URF that was later on sophisticated (Chakraborty et al., 2016). The results are interesting, however, when considering the polynomial fitting, the use of only three provenances is most likely associated with high model uncertainties. In the literature, researchers rarely use less than six reference points (sites or provenances) when developing RF to increase the modeling accuracy (Frank et al., 2017b; Gömöry et al., 2012; Kapeller et al., 2017; Kramer et al., 2017). Additionally, in the presented study, the authors did not indicate any goodness of fit information (e.g., $R^2$ or curve fitting $p$ value), hence the results need to be interpreted cautiously.
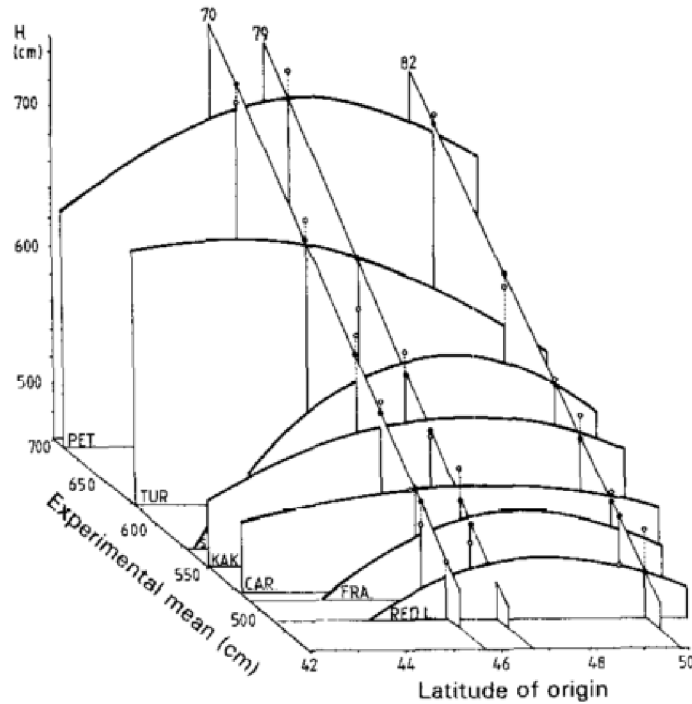
**Fig. 3. 4** Three-dimensional model of three Jack pine provenances planted across seven sites (Matyas, 1994).

Another study by Schmidtling et al. (1994) focused on Loblolly pine and used data from 14 provenances tested over 10 sites spread across Florida and Georgia (USA). The provenance trial ranges across a large part of the species distributional range and was measured at the age of 20. Transfer functions, using both linear and quadratic regression, were fitted in three selected sites: in the southern, middle, and northern parts of the study's range. The $R^2$ of the regressions, ranging from 0.30 to 0.75, are indicative of moderately to good model fittings. In the southern site, they used linear regression, as it was best fitting the data, and the results indicated a continuous increase in growth performance with an increase in the minimum temperature of the provenance (Fig. 3.5). The central and Northern planting sites were best represented by quadratic regression models. Interestingly, when testing provenances in cooler sites, the trend is inversing and the provenances originating from colder regions start to perform better than provenances originating from warmer regions. This result supports the concept that tree populations are locally adapted and that temperature is a major factor influencing tree adaptation.

54
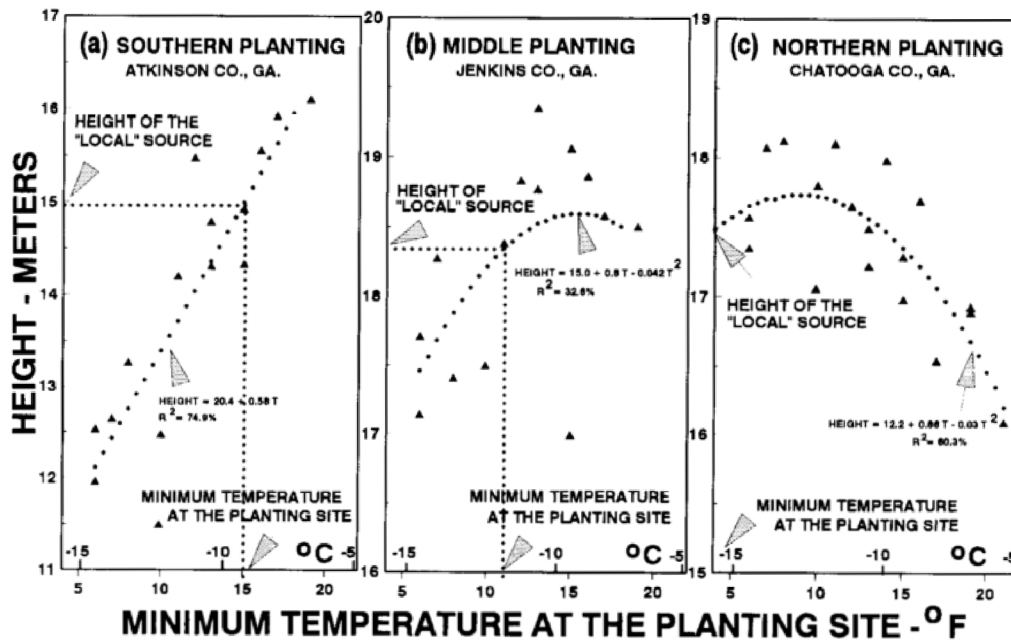
**Fig. 3. 5** Response functions for three Loblolly pine testing sites. The average provenances' height is plotted against the provenances' yearly average minimum temperature (full triangles) (Schmidtling, 1994).

### 3.8.4 Anchor points

Since the provenance trials were not established to study the effect of climate change, we often lack provenances and test sites originating and situated at the species' distributional range. Therefore, in those cases, the fitting of a response function cannot provide reliable results for these regions. One way to cope with this problem is described by Rehfeldt et al. (2001). Their study was based on 125 *Pinus cordata* provenances tested disparately across 60 sites. In their study, the authors first plotted the provenances that were tested across the broadest range of climatic conditions. Then, they calculated the breadth of the parabolas using the two values of *x* for which *y* equaled zero. From these results, they used the widest breadth as a reference to force the other response functions, of the less widely represented provenances, to not exceed this breadth length (Rehfeldt et al., 2001). In Fig. 3.6, we can observe an example where the original response function is represented by a dashed line, and the constrained response function is represented by a solid line.
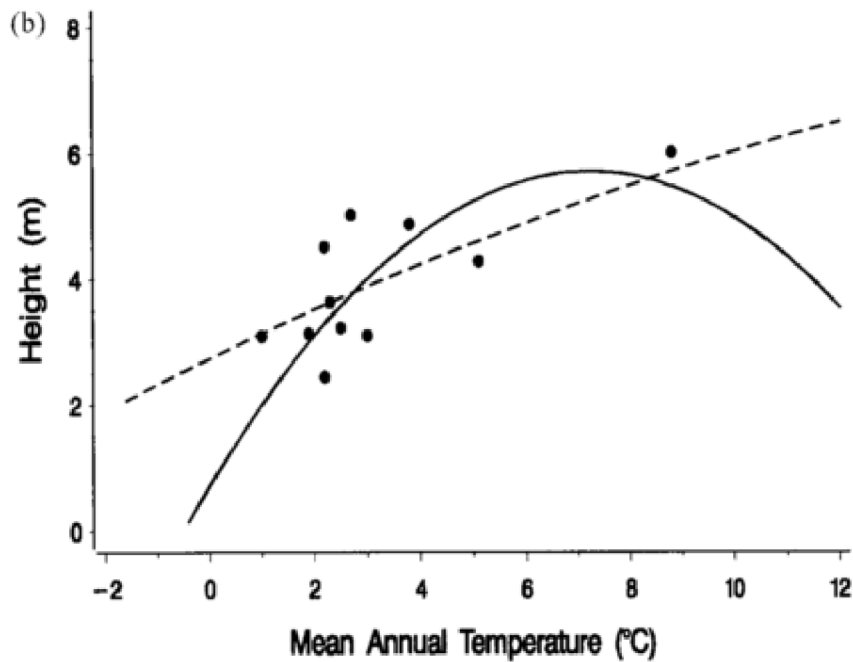
**Fig. 3. 6** *Pinus cordata* response functions fitting height as a function of the mean annual temperature of the testing sites. The dashed line represents the unconstrained response function and the solid line represents the adjusted one (Rehfeldt et al., 2001).

Another way to deal with the issue of the limited available studied environmental range is described by Wang et al. (2006b). In this study, for the provenances that were not tested in the coldest environments, they estimated what would be their responses using transfer functions. Five of these transfer functions were elaborated for the five coldest testing sites, with the provenances heights plotted against their regions of origin' mean annual temperature. Then, for provenances that were not tested in the coldest sites, they used the five transfer functions' predicted heights based on their origin mean annual temperature as anchor points.

## 3.9 Factor analytic

As already mentioned, mixed-linear models are often used in forest tree breeding to analyze MET and select superior populations or genotypes. Generally, mixed models tend to increase in complexity when adding more components such as the number of test sites, number of predictor variables, number of traits (bivariate models), and site designs (e.g., blocks). Indeed, the more components, the more means and variances to calculate. Additionally, model

56

complexity will vary depending on the objectives of the model. For example, do we want to obtain the GxE information? Do we want to obtain GxE correlations between each pair of sites? How do we want to model the residual structure? Do we want to include the pedigree information? Complex models with a large number of variance components to estimates will often have convergency and computational issues. Additionally, models dealing with unbalanced datasets will be more difficult to fit.

One way to overcome the computational and convergency issues when studying the GxE interaction is to use the factor analytic (FA) variance structure. This structure has been widely used in plant breeding, and more recently in tree breeding (Oliveira et al., 2020; Smith et al., 2001b). It has been found to provide a flexible way to describe the GxE in a parsimonious manner that calculates less than a full rank variance structure (smaller number of parameters than an unstructured structure) (Kelly et al., 2007; Smith et al., 2001b; Smith and Cullis, 2018; Zhang et al., 2020). Additionally, it allows the incorporation of pedigree information and the specification of the test site designs (Smith and Cullis, 2018). Also, this model structure can properly accommodate unbalanced, missing, or incomplete data (Burgueño et al., 2011, 2008). Furthermore, compared to other model types (e.g., unstructured), the FA model has been found to provide estimates with higher accuracy in the presence of a substantial GxE, and hence it improves the selection of top genotypes under these circumstances (Kelly et al., 2007).

Simply explained, FA modeling is a type of statistical method that describes the variability among one or several observed variables (e.g., height, wood density, etc.) using a small set of factors. The number of factors has to be specified in the model, and with a higher number, more variance components will be estimated. In a MET analysis, we will obtain a variance-covariance matrix that includes a specific variance component for each site, and a factor loading for each site associated with each factor. The factor loadings correspond to the correlations between the sites and the factors and the factors are representative of unknown underlying variables. Using this information, several graphical tools can be used to visualize the GxE (Cullis et al., 2010; Oliveira et al., 2020; Smith and Cullis, 2018).

For example, it is possible to compute the correlation matrix between sites that can be visualized using heatmaps. Cullis et al. (2014) used an FA of order 3 to analyze the diameter at breast height of 77 *Pinus radiata* breeding trials. Their model accounted for 86% of the variability in their data, indicating a good model fit. The average overall additive genetic correlation was estimated at 0.54, which indicates a relatively high GxE between sites. In addition, the authors assembled a heatmap, that is presented in Fig. 3.7, displaying all the additive genetic pairwise correlations between sites. The pairwise correlations are greatly variable, i.e., the values range from 0 (no correlation between these sites) to 1 (no differences between these sites).
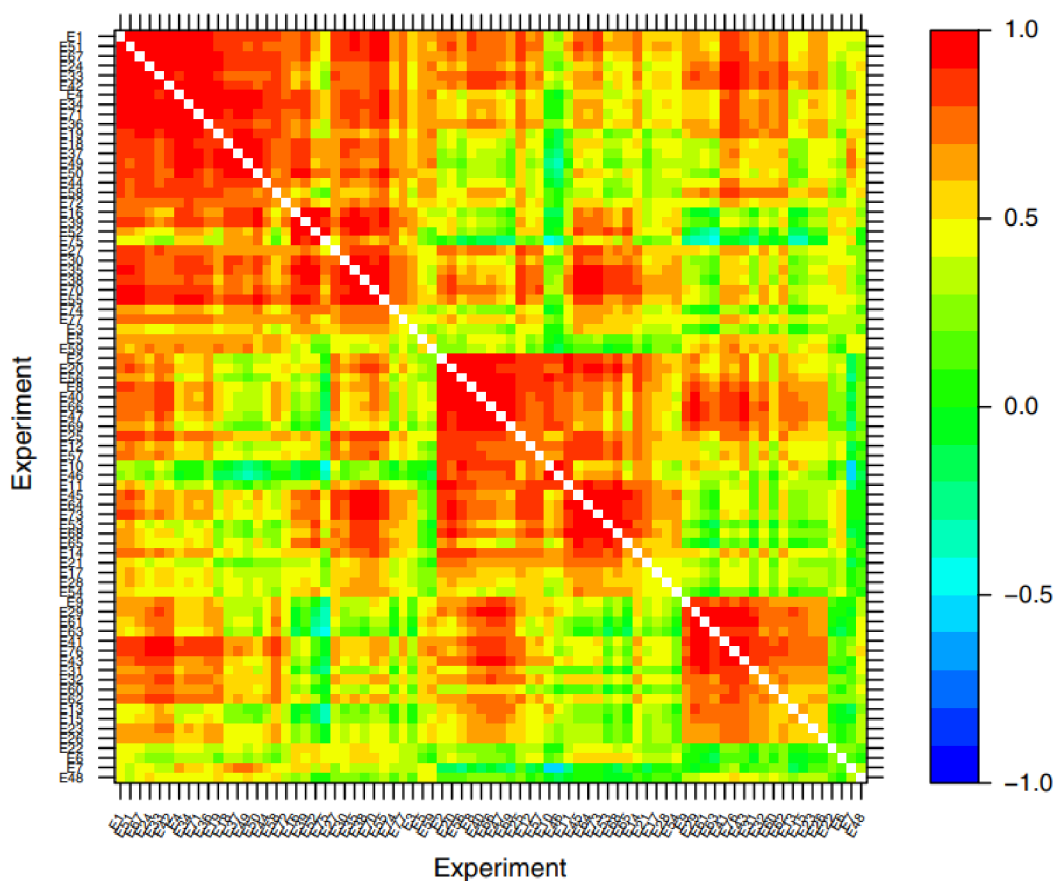


**Fig. 3. 7** Heatmap representing the pairwise additive genetic correlations for 77 *Pinus radiata* breeding trials in New Zealand. The tiles in red are indicative of a low GxE and the tiles in green are indicative of a high GxE (Cullis et al., 2014).

Another graphical tool that can be used to visualize the GxE in the data is the latent regression plot (Thompson et al., 2003) where a factor loading is plotted against a predicted response variable (e.g., site means, genotypic breeding values per site). However, before plotting factor loadings, it is appropriate to first rotate their loadings to maximize the percentage of genetic variation that is explained by the first-factor loading, then by the second-factor loading, and so on (Cullis et al., 2010). One popular way to do it is to use the Varimax function (Kaiser, 1958). When analyzing the latent regression plots, a fitted horizontal line indicates that the response variable remains constant along the environments clustered in the plotted factor. However, if the fitted line is inclined, it indicates that the response variable is influenced by the environmental effects clustered in this factor. Furthermore, when analyzing a set of genotypes, it is possible to compare their latent regression plots. Indeed, when the regression lines all follow a similar inclination, it indicates a low GxE, and when their inclinations are variable, it indicates a higher GxE (Cullis et al., 2014; Oliveira et al., 2020). In their study of the *Pinus radiata* breeding trails mentioned in the previous paragraph, Cullis et al. (2014) used latent regression plots to analyze the response of several genotypes along the three factors.

Fig. 3.8 presents their results along the first-factor loading that explains the largest part of the variability in the data (59.1%). There, we can see that the diameter at breast height of these genotypes is mostly positively influenced by higher values of the first-factor loading.
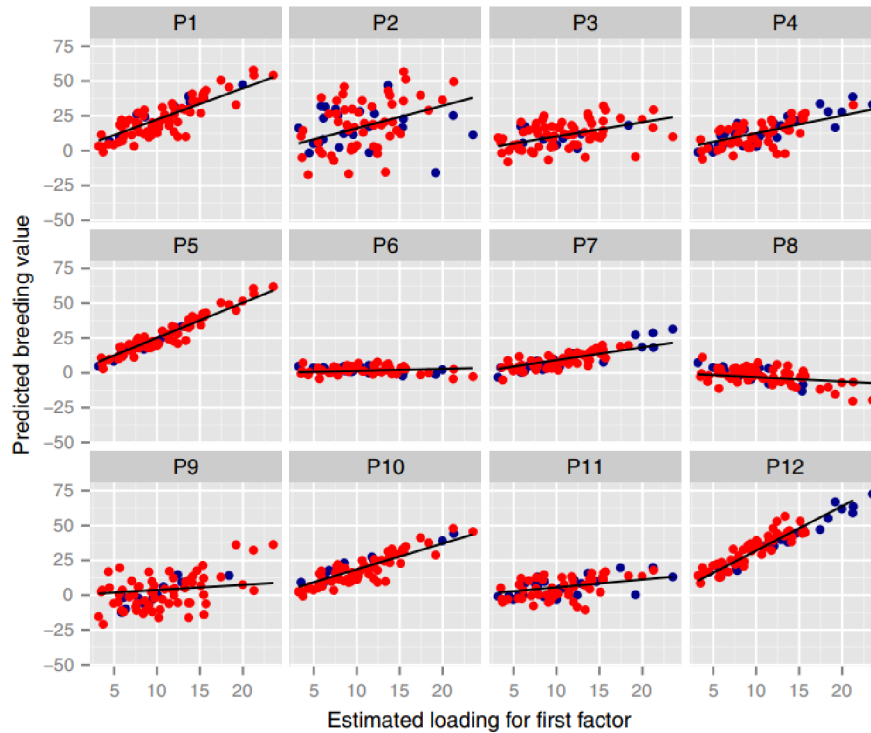
**Fig. 3. 8** Predicted breeding values for 12 *Pinus radiata* genotypes plotted against the first-factor loading of the FA model of order 3 fitted by Cullis et al. (2014).

Another method has been developed to investigate which environmental variable may be clustered in each factor loading (Oliveira et al., 2020), and hence which environmental variables are likely to be influencing genotypic performance across environments and GxE. In their study on sorghum biomass, they used an FA structure of the third order to analyze the GxE for 55 genotypes across 29 trials. Additionally, the authors have selected a set of environmental variables from the National Aeronautics and Space Administration Prediction of Worldwide Energy Resource (NASA POWER) database. To analyze which variable might be clustered in each factor, they used these variables to calculate Pearson correlation coefficients with each factor loading. They obtained interesting results with several climatic variables being moderately but significantly correlated with each factor loading (Table 3.2). Then, they were able to infer, based on the correlations between the genotypes' breeding values and the factor loadings (latent regression plots) and the correlations between the factor loadings and environmental variables, how the studied environmental variables may be influencing sorghum biomass in 55 genotypes.

**Table 3. 2** Pearson correlation coefficients between 9 environmental variables (*covariate*) and 3 factor loadings (*factors 1,2, and 3*) based on a sorghum provenances trials FA model (Oliveira et al., 2020). The values in bold correspond to the significant correlations (significance level not given).

| Covariate | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| **Precipitation** | **0.362** | 0.094 | -0.009 |
| **Maximum temperature** | 0.083 | 0.044 | -0.012 |
| **Minimum temperature** | **0.353** | **-0.391** | 0.199 |
| **Average temperature** | 0.252 | -0.193 | 0.102 |
| **Solar radiation** | 0.024 | **0.305** | -0.107 |
| **Crop's growth cycle** | 0.130 | 0.069 | **-0.221** |
| **Humidity** | 0.175 | -0.279 | 0.076 |
| **Wind speed** | **-0.390** | 0.250 | 0.051 |
| **Altitude** | -0.237 | **0.454** | -0.254 |

# 3.10 European larch

In this dissertation thesis, we are focusing on European larch (*Larix decidua*) and its adaptation to climate. This species is endemic to Central Europe and is characterized by a fragmented distribution (Foff et al., 2014; Saulnier et al., 2019). Native to the Alps and the Carpathian mountains, the species naturally occurs in the Alps, Carpathian, Sudetes, as well as Polish lowlands (Danek and Danek, 2022; Lstiburek et al., 2020). European larch is one of the most common tree species in the subalpine regions, along with Norway spruce and Swiss stone pine, at altitudes situated from 1500 to 2000 m asl. However, the species can be found from foothills to mountainous areas with altitudes ranging from 450 to 2300 m asl (George et al., 2017; Obojes et al., 2022). Compared to its historical range, European larch's distributional range has considerably expanded due to forest management practices (Zeidler et al., 2022). This species largely grows in pure forest stands, however, it can also occur in mixed forests with Norway spruce, Swiss stone pine, and silver fir (Saulnier et al., 2019).

European larch is a fast-growing and shade-intolerant species, that needs light during all its developmental stages, with a pioneering strategy that predisposes it to colonize disturbed soils (e.g., due to landslides, avalanches, grazing, etc.) and form open forests (Lstiburek et al.,

2020; Obojes et al., 2022). European larch is an economically important species due to its fast-growing ability combined with its superior wood quality such as wood density, strength, and water resistance. Its wood has been used to build houses, furniture, water-resistant objects such as outdoor objects or shipbuilding, and high-quality paper. Under the right environmental conditions, its average annual increment rate can reach up to 10 m3 ha−1 (Zeidler et al., 2022), and it is considered to be the seventh-largest forest production volume, in comparison with the other European tree species (Danek and Danek, 2022). For these reasons, European larch has been widely used for afforestation projects outside of its natural range.

Regarding its adaptation to climate, this species has been described as having a relatively robust response to climatic stressors such as frost and drought (George et al., 2017; Obojes et al., 2022; Vacek et al., 2022). However, several studies have found that European larch is substantially negatively affected by low water availability, high temperatures, and droughts (Danek and Danek, 2022; Obojes et al., 2022; Swidrak et al., 2013). Nevertheless, larch has been identified as positively associated with increasing temperatures, particularly spring months temperatures (Danek and Chuchro, 2019; Danek and Danek, 2022; Foff et al., 2014; Izworska et al., 2022). Additionally, larch was found to be generally negatively affected by increasing precipitation levels (Foff et al., 2014), even though several studies indicated different responses depending on the studied months, with for example the positive influence of increased precipitation levels in November and July, and the negative influence in September (Izworska et al., 2022; Swidrak et al., 2013). This is supported by studies indicating that growth seems to be negatively influenced by higher altitudes (Li et al., 2003). Finally, another interesting finding is that larch has displayed an increasing climatic sensitivity with aging (Carrer and Urbinati, 2004).

From a genetic perspective, European larch has shown high geographic genetic variation (Belletti et al., 1997; Matras and Paques, 2010; Nardin et al., 2015), which is one of the factors leading to the fact that provenance research on the influence of climate has begun from the firsts larch provenances experiments (the elaboration of the first international experiment took place in 1944) (Foff et al., 2014). This variation can be explained by the fragmented nature of

the species distribution and its survival of the last glacial period in at least 6 glacial refugia that led to strong genetic differentiation via local adaptation and genetic drift (Wagner et al., 2015). This high genetic variation among provenances suggests high prospects in terms of mitigating the impact of climate change via selection and assisted migration. However, currently, most of the species' gene resource management effort remains focused on gene conservation and seed provision (Pâques, 2013).

# 4  Material and Methods

## 4.1  Material

The phenotypic and pedigree data used in both studies originate from Lstibůrek et al. (2020). All the statistical analyses were performed in R (R Core Team, 2020).

### 4.1.1  Seed orchard

The European larch trees used in both studies originated from the Hamet clonal seed orchard, listed as "Lä P3 (4.2/sm-tm)" in the Austrian national list of seed orchards. The orchard is situated in the North-East of Austria, close to Klausen-Leopoldsdorf in the Vienna woods (48° 3' 36" N, 15° 34' 12" E), at an elevation of 520 m.a.s.l. It spans an area of 3.15 ha. It was established in 1954 using 1,666 vegetative propagules (grafts) originating from 42 phenotypically superior trees (genotypes, denoted as clones) selected based on their level of straightness (Schueler, 2011). Those trees were selected from a local European larch provenance in the Austrian provenance region "4.2, North border of the Alps – East part" (Fig. 4.1).

Open-pollination among the clones in the orchard provides valuable seeds that have been heavily harvested since the end of the 1960s (Geburek, 2021); the seeds are mainly used in lower altitude levels, and it is considered one of the most valuable European larch seed provenances for the northern Alpine foreland's mountainous area. However, due to its partially advanced age, its productivity is currently decreasing. From its establishment, the Austrian seed orchard program for larch aims to secure seed supply with little genetic testing. Hence, progeny tests and controlled pollinations were not performed.
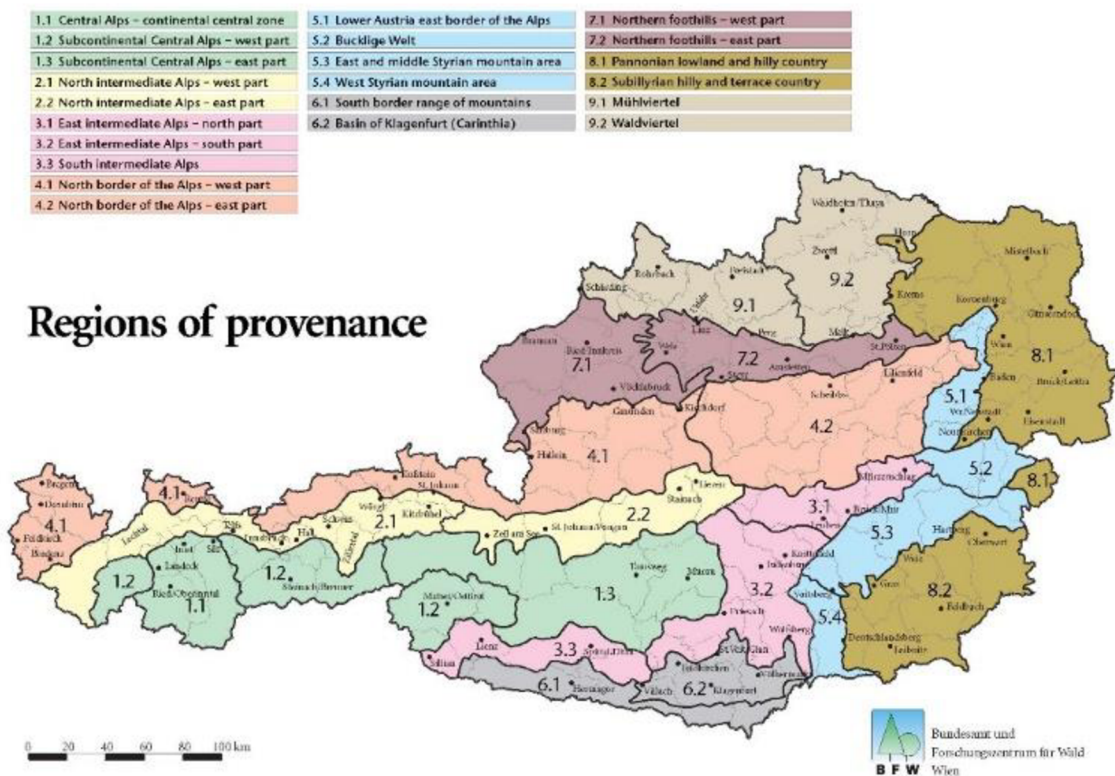
**Fig. 4. 1** Map of the Austrian regions of provenance (Salzmann et al., 2023)

## 4.1.2  Test sites

For their study, Lstibůrek et al. (2020) identified 21 typical forest stands, established using seeds originating from the above-mentioned seed orchard (Fig. 4.2). The stands were selected based on several criteria:

- a minimum of 200 trees per stand,

- stands composed mostly or exclusively of European larch,

- a low level of environmental variation within each stand,

- and similar site structure across all stands.

The selected Austrian forest's stands are geographically spread over an area of approximately 110 km by 170 km. The sites' altitudes ranged from 280 m to 760 m a.s.l., and the ages, at the time of the sampling, varied from 25 to 37 years (Table 4.1).

65
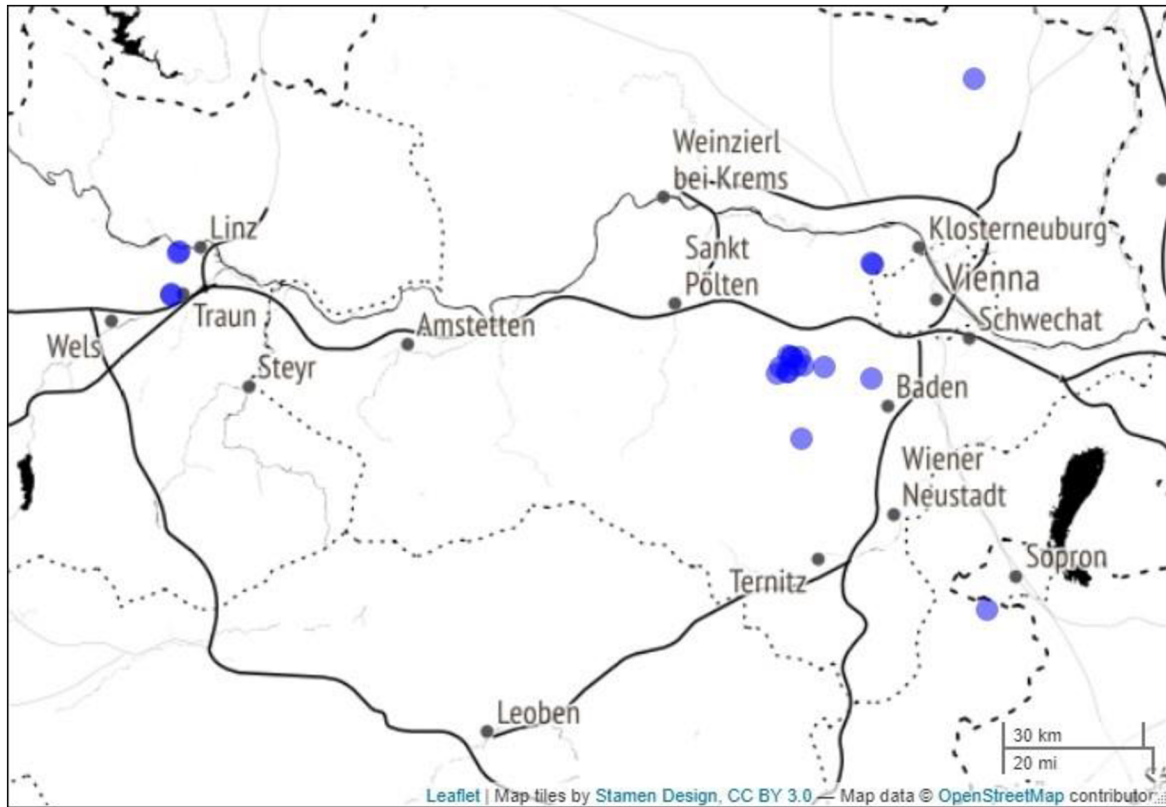
**Fig. 4. 2** Map of the 21 reforestation sites (blue circles) retrieved from Poupon et al. (2021).

**Table 4. 1** Basic descriptive parameters of the 21 forest stands. *Site* is listing the names of the sites, *Nb* is the number of sampled trees per site, *Lat* and *Lon* are the latitudes and longitudes of the test sites, *Alt* is the altitude recorded at the time of the sampling, and *Age* is the age of the stands at the time of the sampling.

| Site | Nb | Lat | Lon | Alt | Age |
|------|-----|-------|-------|-----|-----|
| A | 209 | 48.62 | 16.48 | 324 | 27 |
| B1 | 208 | 48.28 | 16.19 | 345 | 35 |
| B2 | 200 | 48.28 | 16.20 | 360 | 25 |
| B3 | 204 | 48.10 | 15.99 | 474 | 35 |
| B4 | 202 | 48.08 | 16.00 | 473 | 30 |
| B5 | 201 | 48.10 | 15.97 | 473 | 30 |
| B6 | 201 | 48.10 | 15.96 | 439 | 25 |
| B7 | 201 | 48.09 | 15.98 | 470 | 37 |
| B9 | 198 | 48.08 | 16.06 | 474 | 25 |
| B11 | 201 | 48.07 | 15.92 | 617 | 32 |
| B12 | 201 | 48.07 | 15.95 | 580 | 27 |
| B13 | 210 | 48.10 | 15.96 | 532 | 30 |
| B16 | 205 | 47.95 | 15.99 | 724 | 30 |
| B18 | 200 | 48.08 | 15.93 | 760 | 35 |
| B20 | 200 | 48.07 | 15.95 | 614 | 25 |
| T1 | 218 | 48.22 | 14.20 | 279 | 25 |
| T2 | 195 | 48.22 | 14.20 | 284 | 25 |
| W3 | 170 | 48.30 | 14.22 | 450 | 27 |
| W4 | 204 | 48.30 | 14.23 | 420 | 33 |
| H2 | 208 | 48.06 | 16.19 | 350 | 30 |
| N | 231 | 47.62 | 16.52 | 330 | 28 |

### 4.1.3  Phenotypic data collection

The phenotypic data collection on the 21 sites was performed by the BFW institute (Bundesforschungszentrum für Wald - Vienna). They sampled a total of 4 267 trees for the three following traits:

- height, in cm, using a hypsometer,

- diameter at breast height, in cm, using a caliber,

- and wood density (PP), in mm, using a 6J Forest Pilodyn (Cown, 1978), via the pilodyn penetration method. A higher PP corresponds to a lower density and vice-versa. They used a standard pin, measuring 2.5 mm in diameter at a height of 1.3 m. They did not remove the bark to avoid damage to the trees. They took two samples per tree and then averaged the value (more details are given by Škorpík et al., 2018).


### 4.1.4  Genetic data collection

In addition to the phenotypic data, the authors also genotyped a subset of the above-mentioned measured trees. First, both a random and a top-phenotypic subset of offspring individuals across all stands were selected based on the methodology described by El-kassaby and Lstibůrek (2009); Lstibůrek et al. (2012); and Lstibůrek et al. (2015). The number of individuals selected was based on meeting three crucial criteria:

- obtaining a genetic response to selection comparable to traditional recurrent selection that uses structured control crosses (White et al., 2007).

- to be able to reconstruct a pedigree with sufficient accuracy (Kalinowski et al., 2007; Marshall et al., 1998).

- achieve the necessary effective population size for genetic diversity in the authors' target for a seed production population.

Additionally, the authors considered the pollen contamination from external sources when deciding on the sample size (Lstibůrek et al., 2012). The authors genotyped a total of 1 252 individuals across the 21 sites.

Then, they proceeded to the DNA extraction. They first collected tissue samples of the trees using a 15 mm hole punch. In doing so, they obtained cambium cells (unspecialized meristem cells situated between the bark and the wood) that they dried and then stored in silica gel. Afterward, they extracted the DNA following a modified version of the CTAB protocol, as described by Lefort and Douglas (1999). They used approximately 100 mg of frozen tissue that was previously ground using a Mixer Mill MM20. Finally, using three microsatellite multiplexes, they fingerprinted the extracted DNA. The microsatellites used are denominated as follow: Ld30, bcLK189, bcLK228, bcLK263, Ld56, Ld31, Ld50, bcLK211, bcLK253, Ld58, Ld42, Ld101, and 4 Ld45. These microsatellite multicomplexes were designed by Wagner et al. (2012).

## 4.1.5  Pedigree reconstruction

The authors used the Cervus likelihood-based method to reconstruct the pedigree (Marshall et al., 1998). The parameters that were used for the pedigree reconstruction were as follows:

- unknown sexes of the offspring,

- no assumption regarding the maternal contributions,

- LOD score, which corresponds to the overall likelihood ratio's natural logarithm,

- and Delta, which corresponds to the difference in LOD scores between the two candidate parents that were the most likely.

These parameters reflected the possibility of genotyping errors and the fact that, due to pollen contamination, the parental population was incomplete.

Then, to evaluate the robustness of the parentage assignment, they performed an initial set of simulations of the parentage analysis for 10 000 offspring based on the 53 candidate parents' genotypes. They used six parameter scenarios that included:

- six different possibilities for proportions of the parental population, ranging from 0.5 to 1,
- and two different maximal error rates for the genotyping, either 0.01 or 0.1.

Additionally, the following parameters remained constant:

- a minimum of six typed loci,
- a monoecious species (individuals having both male and female reproductive organs) with a polygamous mating (female mating with several pollen donors and pollen donor mating with several females),
- a possibility of selfing (self-pollination),
- a 99% confidence regarding the parental assignment.

The authors kept the family assignments consistent across all six scenarios as the final pedigree.

## 4.1.6 Climatic data

The two studies described below were based on the same climatic dataset. It was extracted from the WorldClim database (Hijmans et al., 2005). This dataset has a spatial resolution of 1 km$^2$ and covers a 50 years' period, from 1950 to 2000. It consists of:

- altitudes
- an average of the minimum temperatures for each month,
- an average of the mean temperatures for each month,
- an average of the maximum temperatures for each month,
- a set of 19 biological indicators described in detail in Table 4.2.

**Table 4. 2** Description of the 19 biological relevant indicators from the WorldClim dataset (Hijmans et al., 2005).

|  | **Description** |
|---|---|
| **Biol1** | Annual mean temperature |
| **Biol2** | Mean diurnal range (mean of monthly (max temp - min temp)) |
| **Biol3** | Isothermality (bio2/bio7) (* 100) |
| **Biol4** | Temperature seasonality (standard deviation *100) |
| **Biol5** | Max temperature of the warmest month |
| **Biol6** | Min temperature of the coldest month |
| **Biol7** | Temperature annual range (bio5 - bio6) |
| **Biol8** | The mean temperature of the wettest quarter |
| **Biol9** | The mean temperature of the driest quarter |
| **Biol10** | The mean temperature of the warmest quarter |
| **Biol11** | The mean temperature of the coldest quarter |
| **Biol12** | Annual precipitation |
| **Biol13** | Precipitation of the wettest month |
| **Biol14** | Precipitation of the driest month |
| **Biol15** | Precipitation seasonality (coefficient of variation) |
| **Biol16** | Precipitation of the wettest quarter |
| **Biol17** | Precipitation of the driest quarter |
| **Biol18** | Precipitation of the warmest quarter |
| **Biol19** | Precipitation of the coldest quarter |

## 4.2 Method - Response function

In this study, we aimed to combine the traditional response function methodology (Chakraborty et al., 2015; Mátyás, 1994; McLane et al., 2011; Wang et al., 2006b) with genetic information (i.e., pedigree) into a new methodology. This methodology aims to support the selection of optimum seeds for re-forestation or afforestation projects, especially under climate change. This new methodology enables the selection of specific parental trees for defined climatic conditions. This is different from the conventional implementation of the RF methodology, where the selection is based on the population level (Poupon et al., 2021).

## 4.2.1 Environmental data selection

From the climatic variables presented in section 4.1.6 and the site's altitude, we needed to select one or more variables to fit our subsequent model. To that effect, we decided to use the Random Forest algorithm, as it can give a highly accurate determination of the importance of many explanatory variables (on a studied dataset). This algorithm is based on building many decision trees, created based on random subsets of explanatory variables and data, which results are then averaged. Simply put, a decision tree for variable selection will, at each node, separate the data in two based on the high or low importance of an explanatory variable. Eventually, we can obtain the importance of each variable on the data. However, a decision tree can have poor accuracy, for example, in the case of complex datasets (e.g., complex variable interactions). Hence, the use of the Random Forests algorithm is more appropriate. There are two types of measures of the importance of each variable:

- the "mean decrease accuracy", which corresponds to the accuracy lost when excluding one variable. A higher value indicates a better explanatory variable.
- The "mean decrease impurity" (or gini importance), which corresponds to the average gain in purity (a pure node split data 100% to 0%) by the splits of a given variable.

As the importance/ranking of the variables had some variation when re-running the model, we ran it ten times and selected the most recurring variables.

After this preliminary selection, we calculated pair-wise correlations between the selected variables. For the highly collinear variables (> 0.85), preliminary response function models (see description in section 4.2.3) were used to choose the most important one.

## 4.2.2 Genetic evaluation

Using the pedigree information (see section 4.1.5) combined with the height and PP measurements (see section 4.1.3), we used a bivariate mixed linear model, to obtain estimations of the traits' narrow sense heritability estimates as well as the individual additive genetic breeding values.

First, our data included forest stands with different ages (between 25 and 37 years old). Provided that age highly influences height, we transformed the height data into the mean annual increment data (MAI-H) to obtain comparable results across sites. No correlation between age and PP was found, so no transformation was performed for the wood density data. Then, we performed a bivariate analysis using the mixed linear genetic animal model within the ASReml-R package (Butler et al., 2017). We used the protocol described by Henderson (1984). The model equation is as follows: (Henderson, 1984)

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e} \tag{11.}$$

where:

- $\boldsymbol{y}$ is the vector of the bivariate phenotypic observation;

- $\boldsymbol{b}$ is the fixed effect (site and trait means) with respective incidence matrix $\boldsymbol{X}$;

- $\boldsymbol{Z}$ is the genetic additive relationship matrix;

- $\boldsymbol{a}$ is the random vector of additive genetic values (BV) that is described as $\boldsymbol{a} \sim N(0, \sigma_a^2)$;

- $\boldsymbol{e}$ is the random vector of the residual effect that is described as $\boldsymbol{e} \sim N(0, \sigma_e^2)$.

The covariance matrix of the random additive genetic effects assumed the heterogeneous covariance structure described as:

$$\sigma_a^2 = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_1 a_2} \\ \sigma_{a_1 a_2} & \sigma_{a_2}^2 \end{bmatrix} \otimes \mathbf{A} \tag{12.}$$

where:

- $\boldsymbol{A}$ is the average numerator relationship matrix;

- $\sigma_{a_1 a_2}$ is the additive genetic covariance between the two traits (1 and 2);

- and $\otimes$ is the Kronecker product operator.

To model the effect of the random residual error, we assumed the unstructured covariance matrix described as:

$$\sigma_e^2 = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_1 e_2} \\ \sigma_{e_1 e_2} & \sigma_{e_2}^2 \end{bmatrix} \otimes \mathbf{I} \tag{13.}$$

where the residual covariance between traits 1 and 2 is defined as $\sigma_{e_1 e_2}$. Here, we assumed independence of the random effects (zero off-diagonal elements).

A comparison was performed between a reduced bivariate model (without the genetic information) and the bivariate model described above using the Akaike Information Criterion, Bayesian Information Criterion (BIC), and restricted maximum likelihood ratio test. Additionally, the narrow-sense heritability estimates for both traits were estimated using the following equation:

$$h^2 = \sigma_a^2 \, / \, (\sigma_a^2 + \sigma_e^2) \tag{14.}$$

where:

- $h^2$ is the heritability;
- $\sigma_a^2$ is the additive genetic variance;
- and $\sigma_e^2$ is the error variance.

Furthermore, the pairwise differences between sites were calculated using the coefficients of the fixed effect of sites from the model output.

## 4.2.3  Response function

To model the subsequent response functions, we first removed the half-sib families that were not sufficiently represented (present at less than six sites). This was done to increase family representation and to provide sufficient data for the modeling. A similar number of test

sites were selected by several other studies (Foff et al., 2014; Kapeller et al., 2017; Suvanto et al., 2016). Afterward, we determined the predicted phenotypic performance (pMAI-H) at each site for each individual and the trait MAI-H. This was calculated by summing the overall mean, the site effect, and the individual additive genetic breeding values obtained from the bivariate model.

Using the pMAI-H values, we then developed both population- and individual-level response functions (RF). The objectives were to observe how MAI-H varied for the population across a climatic gradient and to observe the genetic variation within this population. First, to select the model, we tested the three main models used in previous studies (Leites et al., 2012b; O'Neill et al., 2014, 2007; Sáenz-Romero et al., 2017; Wang et al., 2006b): linear, quadratic, and Gaussian. The results indicate that the linear model wasn't a suitable fit, while the other two were more appropriate. The quadratic and Gaussian models showed almost identical curve shapes and Akaike Information Criteria. The quadratic model was selected for the subsequent analyses. The equation for this model is:

$$v_{jk} = \beta_0 + \beta_1 c_j + \beta_2 c_j^2 \tag{15.}$$

where:

- $v$ is the estimated response variable at the site $j$ for parent $k$;
- and $\beta_0$, $\beta_1$, and $\beta_2$ are the regression coefficients at site $j$;
- and $c$ is the associated climatic variable.

## 4.3 Method - Factor analytic

In this study, we investigated the response of European larch to its environment to facilitate gene resource management, especially in the context of environmental changes (e.g., different planting sites, climate change, etc.). More specifically, we wanted to (1) study the GxE for the three sampled traits (see section 4.1.3) by fitting a multi-environmental FA model;

(2) evaluate the impact of the environment by correlating the models' estimated loadings against pre-selected environmental variables; and finally (3) analyze the GxE dynamics of specific genotypes.

## 4.3.1  Environmental data selection

To study the GxE within the European larch dataset described in section 4.1.3 and make inferences on the influence of environmental variables; we first needed to select a set of environmental variables. As mentioned in section 4.1.6, the climatic data were extracted from The WorldClim dataset, and together with the altitudes and ages of the sampled forest stands, they formed the basis of the selection process. This process was based on a series of Principle Component Analyses (PCA) aiming at reducing the number of variables in each round. As the variables were not on the same scale, we used PCAs on correlation matrices (rather than on covariance matrices) for their standardization. For each analysis, we plotted the associated biplot. When two or more variables demonstrated redundancy, appearing overlapping on the biplot, we selected only one for the subsequent PCA analysis. This process was repeated until we were left with a set of non-redundant variables. The approach to choosing between the redundant variables is ultimately subjective and we decided to keep the ones with the broader meaning. For example, if both specific monthly temperatures and the mean annual temperature overlap, we would choose the yearly mean temperature.

## 4.3.2  Statistical analyses

This study used the three sampled traits from the European larch dataset described in section 4.1.3: height, diameter at breast height, and wood density. Because of the multi-ages sites described in section 4.2.2, we transformed height and diameter at breast height into mean annual increments (MAI-H and MAI-DBH, respectively). We also used the pedigree described in section 4.1.5.

We analyzed each site separately before proceeding to the multiple-site mixed linear models. A single-site model was considered for each trait:

$$\boldsymbol{y} = \mathbf{1}_n \mu + \boldsymbol{Z}\mathbf{a} + \boldsymbol{e} \tag{16.}$$

where:

- $\mathbf{y}$ is the response variable ($n \times 1$ vector);
- $\boldsymbol{n}$ is the number of observations;
- $\mu$ is the overall mean;
- $\boldsymbol{a}$ is the vector ($m \times 1$) of the random additive genetic effects of $m$ genotypes, where $\boldsymbol{a} \sim N(0, \sigma_a^2 \boldsymbol{A})$, and $\sigma_a^2$ is the additive genetic variance;
- $\boldsymbol{e}$ is the vector ($n \times 1$) of the random residual effects, where $\boldsymbol{e} \sim N(0, \sigma_e^2 \boldsymbol{I_n})$, and $\sigma_e^2$ is the error variance;
- $\boldsymbol{Z}$ is the genotype effects' incidence matrix;
- $\mathbf{1}_n$ is the vector of ones;
- $\boldsymbol{A}$ is the ($m \times m$) pedigree-based relationship matrix;
- and $\boldsymbol{I_n}$ is the number of observations matrix.

For each site, diagnostic plots were analyzed to assess the data quality and remove several outliers from the subsequent analyses. Narrow-sense heritability estimates were also calculated and used to evaluate each site's data. The heritability estimates were calculated as :

$$h^2 = \sigma_a^2 \, / \, (\sigma_a^2 + \sigma_e^2) \tag{17.}$$

where:

- $h^2$ is the heritability;
- $\sigma_a^2$ is the additive genetic variance;
- and $\sigma_e^2$ is the error variance.

The same two equations were also used on the whole dataset.

To model the MET, we used the FA structure for each trait. Details on this type of modeling can be found in section 3.9. Simply put, it is a statistical method that describes the variability among one or several observed variables using a small set of factors. These factors are proxies for many explanatory variables that cannot be included directly in the analysis (e.g., due to computational issues). The model is as follows:

$$\boldsymbol{y} = \mathbf{1}_{n^*}\mu + \boldsymbol{X}\boldsymbol{s} + \boldsymbol{Z}\text{a.s} + \boldsymbol{e} \tag{18.}$$

where:

- $\boldsymbol{y}$ is the vector ($n^* \times 1$) of the response variable across $s$ environments;
- $\boldsymbol{n^*}$ is the total number of individuals;
- $\mu$ is the overall mean;
- $\boldsymbol{s}$ is the vector ($s \times 1$) of the fixed effects of sites;
- $\boldsymbol{a.s}$ is the vector ($ms \times 1$) of the random additive genetic effects (breeding values) of the genotypes $m$ nested within the environments $s$, and is described as $\boldsymbol{a.s} \sim N(0, G \otimes \boldsymbol{A})$;
- $\boldsymbol{e}$ is the vector ($n^* \times 1$) of the residual effects, and is described as $\boldsymbol{e} \sim N(0, \sigma_e^2 \boldsymbol{I}_{n^*})$, with the error variance $\sigma_e^2$;
- $\boldsymbol{G}$ is the ($s \times s$) variance-covariance matrix for the genotypic effect nested within the environments (see details below);
- $\boldsymbol{A}$ is the matrix ($m \times m$) of the relationship matrix that is pedigree-based;
- $\otimes$ is the Kronecker product.
- $\boldsymbol{X}$ is the incidence matrix for the environmental effects;
- $\boldsymbol{I}_n$ is the matrix of the number of observations;
- and $\mathbf{1}_n$ is the vector of ones.

The percentage of the genetic variance that is explained by each model, from factors 1 to 4, was calculated as:

$$\bar{v} = 100 \times t(\Lambda\Lambda^{T})/t(\Lambda\Lambda^{T} + \psi) \tag{19.}$$

where:

- $\bar{v}$ is the overall percentage of the genetic variance;
- $\Lambda$ is the matrix ($s \times k$) of factor loadings $\{\lambda_{sk}\}$;
- $\lambda_{sk}$ correspond to the factor loading $k$ ($k = 1, 2, 3,$ or $4$);
- and $\psi$ is the ($s \times s$) diagonal matrix, with a specific variance for each environment $s$.

The variance-covariance matrix $G$ was modeled as a factor analytic of order $k$ (FA$_k$) following the structure:

$$G = (\Lambda\Lambda^{T} + \psi) \tag{20.}$$

where:

- $\Lambda$ is the matrix ($s \times k$) of factor loadings $\{\lambda_{sk}\}$;
- $\lambda_{sk}$ corresponds to the factor loading $k$ ($k = 1, 2, 3,$ or $4$);
- and $\psi$ is the ($s \times s$) diagonal matrix, with a specific variance for each environment $s$.

For each trait, we tested the significance of the models using likelihood ratio tests (each $k+1$ model was tested against the model of order $k$). Additionally, we considered the percentage of the variance explained by each model, the Akaike information criteria, and the log-likelihood for the final model selection.

For the selected models, the site-to-site additive genetic correlations for the selected models were calculated to assess the possible presence of GxE (Oliveira et al., 2020). If the

sites are highly correlated, it suggests a low GxE interaction. Conversely, lower correlations between sites are indicative of a higher GxE interaction. These correlations between pairs of sites were calculated using the **G** matrices obtained from the FA models as:

$$\rho_{ss'} = COV_{ss'} \sqrt{\sigma_s^2 \sigma_{s'}^2} \qquad (21.)$$

where:

- $\rho_{ss'}$ is the genetic correlation for each pair of sites;

- $COV_{ss'}$ is the genetic covariance between pairs of sites $s$ and $s'$;

- and $\sigma_s^2$ and $\sigma_{s'}^2$ are the genetic variances for the trials $s$ and $s'$, respectively.

Additionally, the average narrow-sense heritability estimate of each trait was calculated as:

$$\overline{h^2} = \overline{\sigma_a^2} / (\overline{\sigma_a^2} + \sigma_e^2) \qquad (22.)$$

where:

- $\overline{h^2}$ is the average heritability;

- $\overline{\sigma_a^2}$ is the average additive genetic variance;

- and $\sigma_e^2$ is the error variance.

Then, to correct for the G×E, the overall unbiased $h^2$ was calculated using the following expression:

$$h_u^2 = \overline{h^2} \times \overline{\rho_a}$$
(23.)

where:

- $h_u^2$ is the overall unbiased narrow-sense heritability;

- $\overline{h^2}$ is the average heritability;

- and $\overline{\rho_a}$ is the mean site-to-site additive genetic correlation.

Afterward, the relationship between the genotypic expression and the environmental conditions represented by the factor loadings was investigated for each trait. First, the loadings were rotated, as described by Cullis et al. (2010), using the Varimax function (Kaiser, 1958). The purpose was to maximize the percentage of genetic variation accounted for by the first loading and then the second loading. Then, latent regression plots were built (Cullis et al., 2014; Thompson et al., 2003), both at the provenance and the genotypic levels. At the provenance level, the rotated loadings were plotted against the predicted site means and at the genotypic level against the individual breeding values.

Finally, as presented by Oliveira et al. (2020), we calculated the Pearson correlation coefficients (and associated significance levels) between the selected environmental variables and the rotated loadings. These coefficients were used to identify which variable may be responsible for the variation of the genotypic performance.

# 5 Results

In this section, a detailed examination of the results will be presented, adhering to the same organizational framework employed in the methodology section. The initial segment involves a review of the response function study's results. Here, a novel methodology was both developed and tested, aiming at quantifying the intra-population additive genetic variance and aligning specific genotypes with certain climatic variables (Poupon et al., 2021). This method resulted from a fusion of the response function approach with traditional tree breeding methodologies.

The subsequent segment offers an overview of the findings from the application of FA modeling. The purpose was to comprehend the GxE interaction within the examined population, pinpoint the major environmental variables influencing the growth and wood density of European larch, and visualize the effects of these environmental variables (Poupon et al., 2023).

## 5.1 Results - Response function

In this section, we will first go through the results concerning the environmental variables selection. Then, we will review the bivariate model output and the random additive variance within the studied population. Afterward, we will focus on our findings from the response functions at the population and individual levels.

### 5.1.1 Environmental data selection

The occurrence of the most important variables over the 10 random forest models can be seen in Table 5.1. Based on these results, the seven most frequent variables were selected for further analyses:

- altitude,

- minimum temperature of January,

- minimum temperature of December,

- mean temperature of the coldest month (MTMC),

- mean temperature of December,

- maximum temperature of January and December,

- and maximum temperature of December.

These variables are explaining 93.9% of the variability in our dataset. However, the pairwise correlations between the variables indicated high levels of collinearity (> 0.85). Hence only one variable was selected based on the occurrence score and preliminary response functions: MTCM. This variable, using the random forest algorithm, explains 69.4% of the variability in the dataset.

**Table 5. 1** Frequency occurrence *Freq* of the 68 tested environmental variables *EV* after running 10 random forest models. The occurrence of both the mean decreases in accuracy *MDA* and the mean decrease in impurity *MDI* of all models are recorded. The seven most recurrent variables are indicated in bold.

| EV | MDA | MDI | Freq |
|---|---|---|---|
| **Alt** | **10** | **8** | **18** |
| **Biol4** | 3 | 2 | 5 |
| **Biol9** | 1 | 0 | 1 |
| **Biol12** | 1 | 2 | 3 |
| **Prec5** | 1 | 0 | 1 |
| **Tmax1** | **5** | **1** | **6** |
| **Tmax2** | 4 | 0 | 4 |
| **Tmax6** | 1 | 2 | 3 |
| **Tmax11** | 1 | 0 | 1 |
| **Tmax12** | **6** | **9** | **15** |
| **Tmean1** | **9** | **10** | **19** |
| **Tmean5** | 0 | 2 | 2 |
| **Tmean9** | 2 | 3 | 5 |
| **Tmean12** | **6** | **8** | **14** |
| **Tmin1** | **8** | **9** | **17** |
| **Tmin5** | 0 | 1 | 1 |
| **Tmin9** | 2 | 3 | 5 |
| **Tmin12** | **4** | **4** | **8** |

## 5.1.2 Genetic evaluation

In Table 5.2, we compare the full and the reduced bivariate models. They indicate that the former, which includes the pedigree information, is significantly improved over the latter. More specifically, the Akaike Information Criterion and the Bayesian Information Criterion are both smaller for the full model, and the restricted maximum likelihood ratio test indicates a significant improvement ($P < 0.001$).

**Table 5. 2** Statistics of the bivariate genetic animal model $B_1$ versus the reduced model $B_2$, with the Akaike Information Criterion *AIC*, the Bayesian Information Criterion *BIC*, and the *p* value *P* of the restricted maximum likelihood ratio test.

| Model | AIC | BIC | P |
|:---:|:---:|:---:|:---:|
| $B_0$ | 8111 | 8128 | |
| $B_1$ | 7946 | 7981 | **< 0.001** |

In Table 5.3, we present the summary of variable components from the full bivariate model. A negligible (0.09) and most likely not significant (z ratio < 2) additive genetic correlation was found between the traits. The heritability estimates were 0.27 (SE = 0.07) for MAI-H and 0.30 (SE = 0.07) for PP. The pairwise differences between sites are reported in Table S1, 87% of them were significant ($P < 0.05$).

**Table 5. 3** Bivariate genetic animal model variance components *VC* values, standard errors *SE*, and z ratios. *COVa* is the additive genetic correlation between the traits, *Va* is the additive genetic variance for both MAI-H and PP, *COVe* is the residual correlation between the traits, and *Ve* is the residual variance of both MAI-H and PP.

| VC | Value | SE | z ratio |
|:---:|:---:|:---:|:---:|
| **COVa** | 0.090 | 0.199 | 0.453 |
| **Va MAI-H** | 11.880 | 3.425 | 3.469 |
| **Va PP** | 0.698 | 0.195 | 3.576 |
| **COVe** | 0.312 | 0.054 | 5.773 |
| **Ve MAI-H** | 32.370 | 2.515 | 12.870 |
| **Ve PP** | 1.660 | 0.138 | 12.011 |

The parental breeding values for MAI-H are visible in Table 5.4 and Fig. 5.1. They illustrate the extent of the additive genetic variation that is present within the set of parental trees (31 parents). We can see that the breeding values range from -5.0 (parent L11) to 6.4 cm (parent S7;). This means that a given offspring of parent L11 has the potential to grow an average of 2.5cm less than the offspring of parent S21, and 5.7 cm less than the offspring of parent S7 (expressed per year). Additionally, each breeding value is associated with its respective confidence interval. Here, the confidence intervals are quite homogenous among parents and they can be interpreted as follows: Parent S21 has a confidence interval of 1.2cm, which indicates a possible range of the breeding value from -1.2 to +1.2 cm under assumed significance level alpha = 0.05.
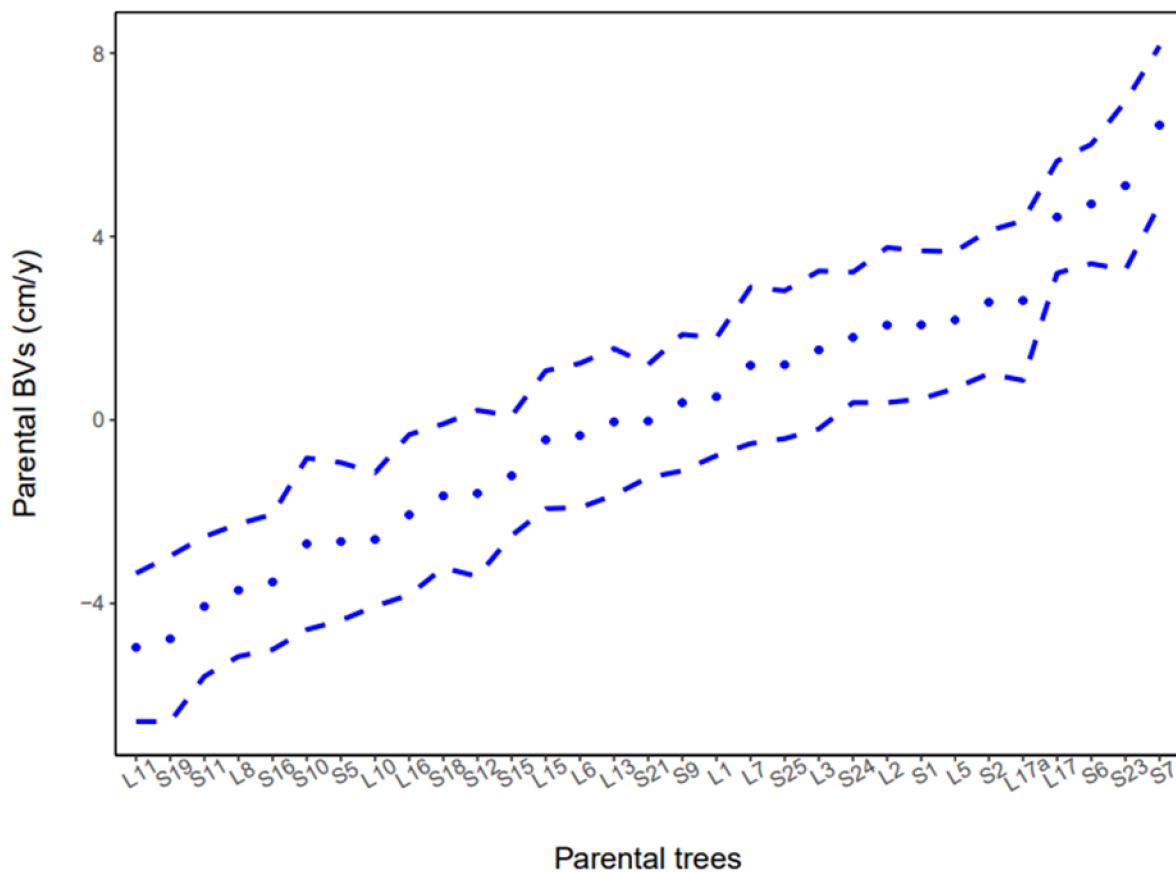


**Fig. 5. 1** Plot of the estimated parental breeding values (blue dots), ranked from the smallest to the highest, with their associated confidence interval (blue dashes).

**Table 5. 4** List of the estimated parental breeding values *BV* and their associated standard errors *SE*.

| Parent | BV | SE |
|--------|------|-----|
| L11 | -5.0 | 1.6 |
| S19 | -4.8 | 1.8 |
| S11 | -4.1 | 1.5 |
| L8 | -3.7 | 1.4 |
| S16 | -3.5 | 1.5 |
| S10 | -2.7 | 1.9 |
| S5 | -2.7 | 1.7 |
| L10 | -2.6 | 1.5 |
| L16 | -2.1 | 1.7 |
| S18 | -1.7 | 1.6 |
| S12 | -1.6 | 1.8 |
| S15 | -1.2 | 1.3 |
| L15 | -0.4 | 1.5 |
| L6 | -0.3 | 1.6 |
| L13 | 0.0 | 1.6 |
| S21 | 0.0 | 1.2 |
| S9 | 0.4 | 1.5 |
| L1 | 0.5 | 1.3 |
| L7 | 1.2 | 1.7 |
| S25 | 1.2 | 1.6 |
| L3 | 1.5 | 1.7 |
| S24 | 1.8 | 1.4 |
| L2 | 2.1 | 1.7 |
| S1 | 2.1 | 1.6 |
| L5 | 2.2 | 1.5 |
| S2 | 2.6 | 1.6 |
| L17a | 2.6 | 1.7 |
| L17 | 4.4 | 1.2 |
| S6 | 4.7 | 1.3 |
| S23 | 5.1 | 1.8 |
| S7 | 6.4 | 1.7 |

### 5.1.3  Response function

In Fig. 5.2, we present the quadratic response function at the population level, with an adjusted $R^2$ of 0.32 and a significant *p* value ($p < 0.001$). The curve's 95% confidence interval is represented by the gray band, we can see that on the extremities of the curve, the interval is

wider. The pMAI-H culminates at around 65 cm per year for an associated MTMC of -2.2°C. In the figure, each boxplot describes the pMAI-H distribution of the studied offspring (the black dots represent extreme values). Site B16, in the lower-left corner, is associated with the lowest pMAI-H values and is represented by the lowest MTCM of the studied sites. The site W3, on the other hand, displays the highest pMAI-H values. While some variation is visible, all the sites have similar distributions. However, we can notice that several sites with the same MTCM are showing contrasting ranges of pMAI-H.
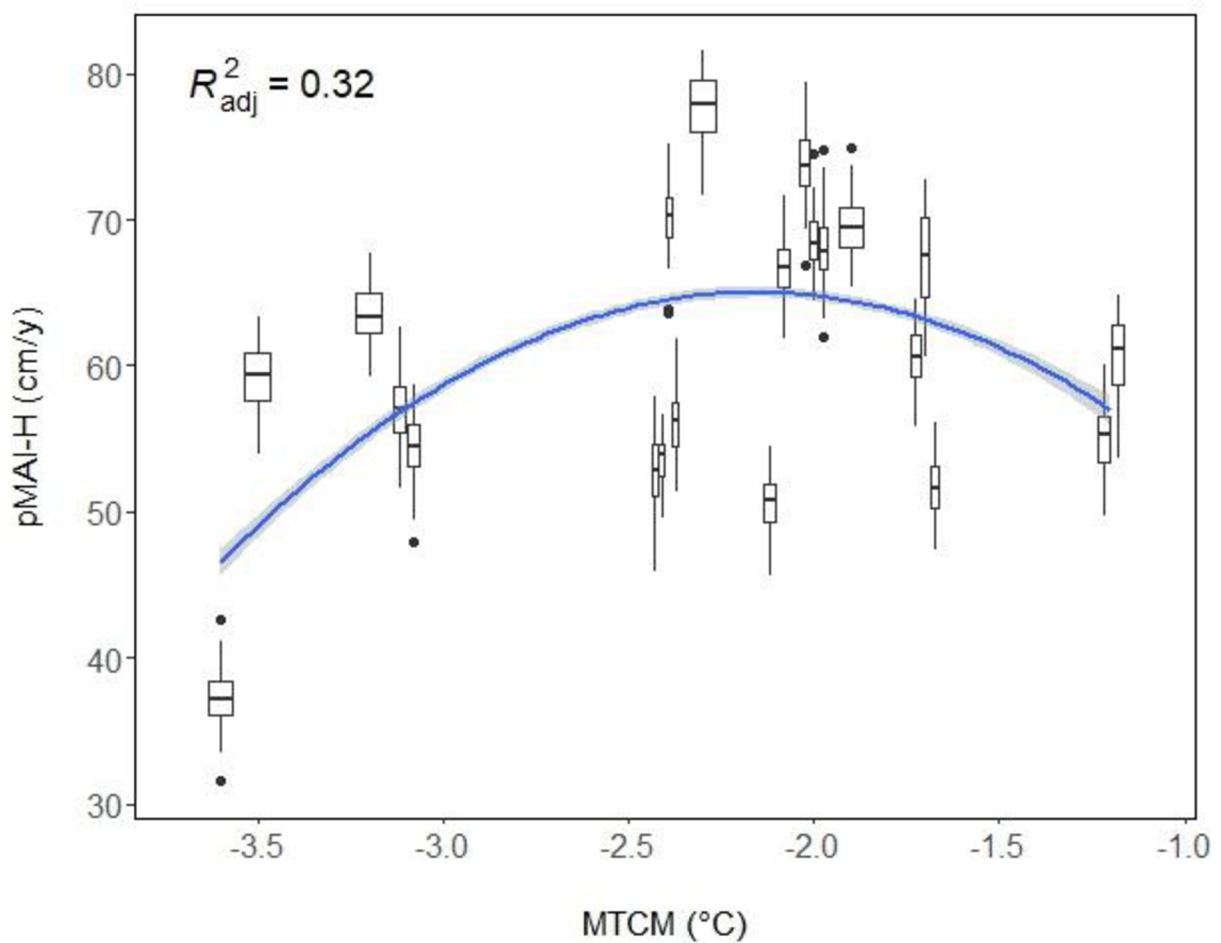


**Fig. 5. 2** Response function at the population level. The boxplots are illustrating the distribution of the pMAI-H in each site. The black dots are outliers. The $R^2_{adj}$ is the adjusted coefficient of determination. The blue line is the fitted quadratic function and the gray area represents its 95% confidence interval.

In Table 5.5 we have details regarding the quadratic fitting at the individual level. In the second column, we can see the number of individuals representing each parent (per half-sib family). This number is varying from 35 to 120 individuals. While there is an uneven representation of the families, we can see in Fig. 5.3 that the mean of the adjusted correlation coefficients does not depend on the half-sib families' sizes. Nonetheless, the variability of these coefficients is higher for smaller half-sib families.



$$y = -0.0006x + 0.322$$
$$R^2 = 0.0099$$
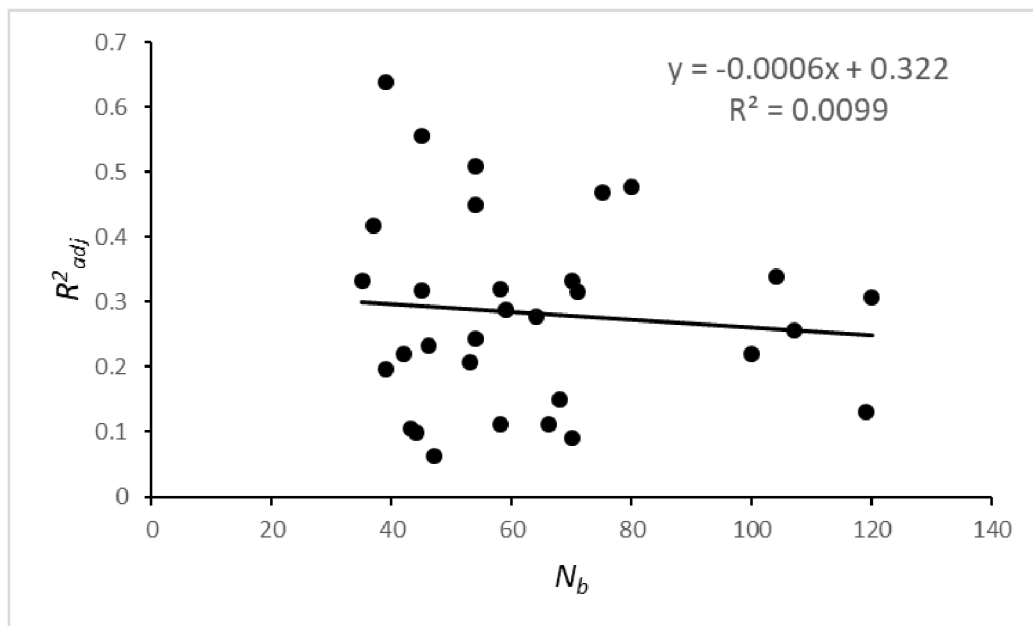
**Fig. 5. 3** Number of individuals *Nb* in each half-sib family as a function of the adjusted correlation coefficients $R^2_{adj}$ of each response function.

In the third column, we provide the adjusted correlation coefficients for each quadratic individual response function. The values are ranging from 0.06 to 0.64 with a median value of 0.27. Finally, in the last column, we can see that all the curves' *p* values are significant, with *p* < 0.05, except for the genotype L2 with the *p* value of 0.09.

**Table 5. 5** Summary statistics of the individual response functions. *Par* is representing the studied genotypes, *Nb* is the number of individuals per half-sib family, the $R^2_{adj}$ is the adjusted coefficient of determinations of each curve, and *P* is the associated *p* value, with significance levels as follows: * = <0.05, ** = < 0.01, ***= < 0.001

| Par | Nb | $R^2_{adj}$ | P |
|---|---|---|---|
| L1 | 107 | 0.26 | *** |
| L10 | 71 | 0.31 | *** |
| L11 | 54 | 0.51 | *** |
| L13 | 53 | 0.21 | *** |
| L15 | 66 | 0.11 | ** |
| L16 | 42 | 0.22 | ** |
| L17 | 119 | 0.13 | *** |
| L17a | 43 | 0.10 | * |
| L2 | 47 | 0.06 | 0.09 |
| L3 | 45 | 0.56 | *** |
| L5 | 70 | 0.33 | *** |
| L6 | 58 | 0.11 | * |
| L7 | 46 | 0.23 | *** |
| L8 | 75 | 0.47 | *** |
| S1 | 54 | 0.24 | *** |
| S10 | 35 | 0.33 | *** |
| S11 | 64 | 0.28 | *** |
| S12 | 39 | 0.20 | ** |
| S15 | 104 | 0.34 | *** |
| S16 | 70 | 0.09 | * |
| S18 | 58 | 0.32 | *** |
| S19 | 39 | 0.64 | *** |
| S2 | 59 | 0.29 | *** |
| S21 | 120 | 0.31 | *** |
| S23 | 37 | 0.42 | *** |
| S24 | 80 | 0.48 | *** |
| S25 | 54 | 0.45 | *** |
| S5 | 45 | 0.32 | *** |
| S6 | 100 | 0.22 | *** |
| S7 | 44 | 0.10 | * |
| S9 | 68 | 0.15 | *** |

A set of nine individual response functions is shown in Fig. 5.4 We notice that the different half-sib families have different types of curves across the studied range of MTCM. Additionally, we can see that their confidence intervals are fanning out at the extremities as for
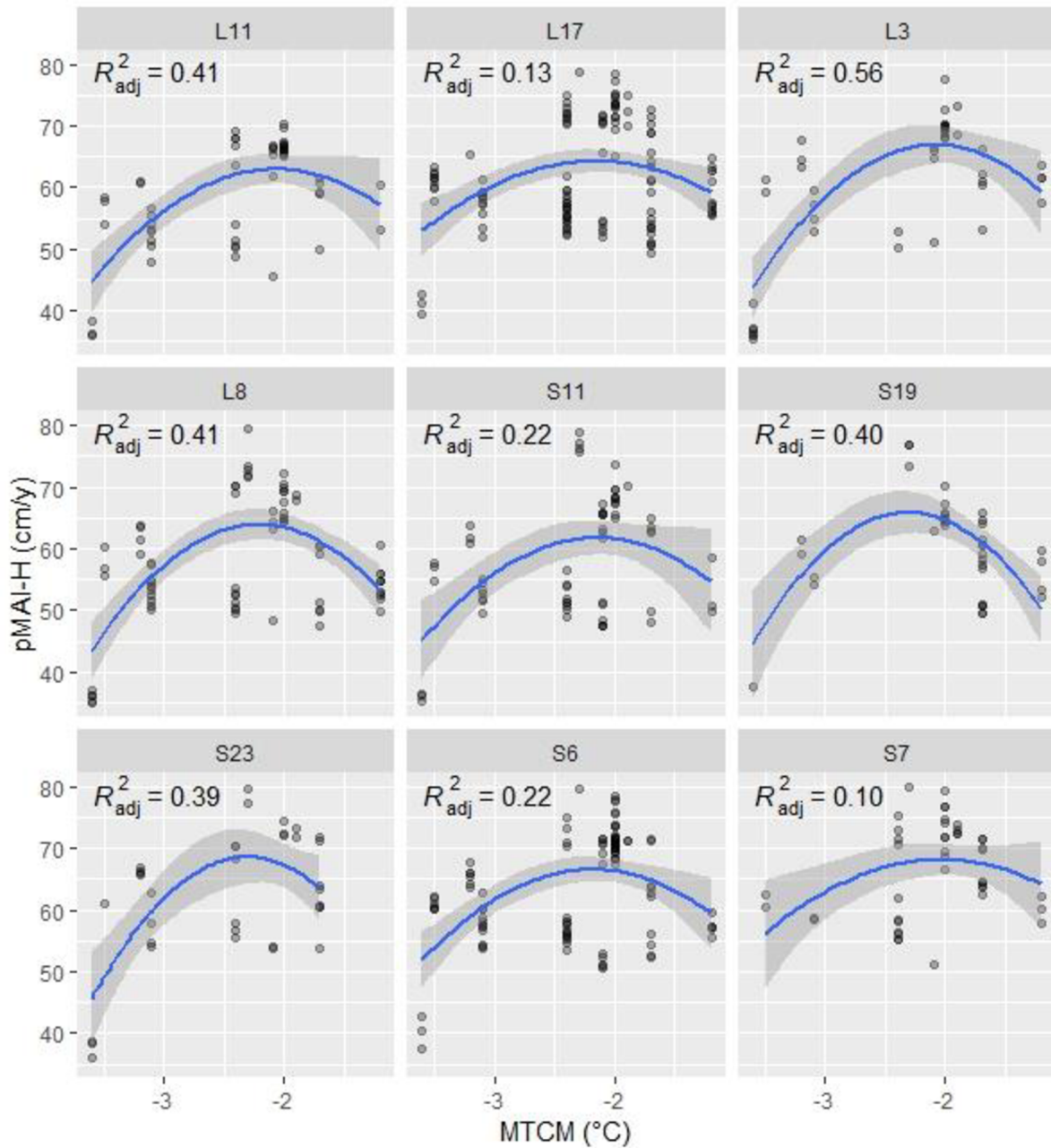
**Fig. 5. 4** Set of nine individual response functions. The blue lines are representing the fitted quadratic functions and their respective gray band are representing their 95% confidence intervals. $R^2_{adj}$ is the adjusted coefficient of determination of each curve.

The entire collection of quadratic response functions at the individual level is illustrated in Fig. 5.5, providing a comprehensive overview of the pMAI-H variability both within and among each half-sib family. As previously noted, various types of curves can be observed. For

example, by examining S23, we notice that this genotype achieves the highest peak among all others. However, its dominance is apparent only within a limited range of MTCM; outside this range, its performance lags behind other families. In contrast, genotype S7 may not reach the same heights as S23, but it surpasses the majority of the studied genotypes across the observed range of MTCM, exhibiting a rounder curve.

When examining all the curves collectively (Fig. 5.5), it becomes evident that the apexes are distributed between 62 (genotypes L2/S11) and 70 cm (genotype S23) per year. This pattern is mirrored in the differences in maximum pMAI-H, where we find a spectrum of optimums across the MTCM gradient. The coldest optimum is identified at a temperature of -2.4°C for genotype S23, while the warmest aligns with a temperature of -1.6°C for genotype S12.
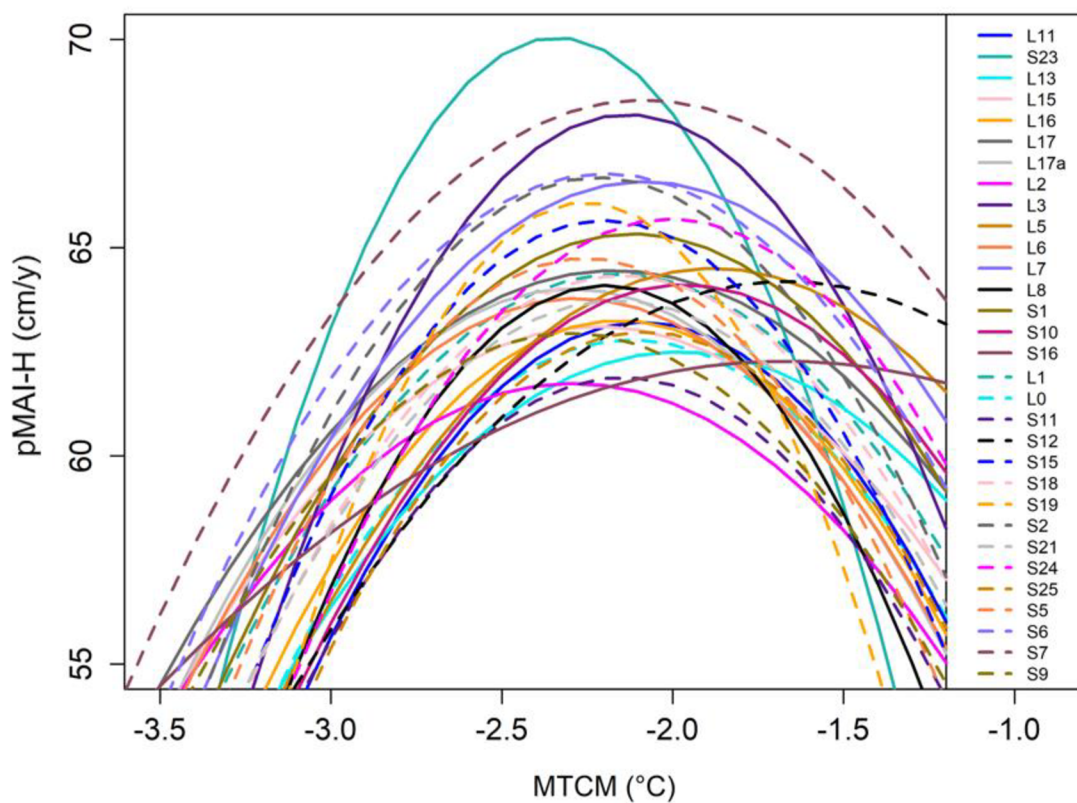


**Fig. 5. 5** Response functions at the individual level are plotted together. On the right side, there is the legend for each half-sib family.

## 5.2 Results - Factor analytic

### 5.2.1 Environmental data selection

Based on the final PCA, we pre-selected 10 environmental variables out of the 68 environmental variables and the stands' ages (Age):

- the annual average of the monthly precipitations (AvPrec);

- the annual average of the mean monthly temperatures (AvTmean);

- the mean temperature of February (Tmean2);

- the minimum temperature of February (Tmin2);

- the mean diurnal range (Biol2);

- the isothermality (Biol3);

- The annual range of the temperatures (Biol7);

- the precipitation seasonality (Biol15);

- the altitudes (Alt);

- and Age.

In Table 5.6, we highlight the importance of the 10 first components of the final PCA. The first four are explaining more than 96% of the variance, with the first component explaining 46.4%, the second 34.0%, the third 9.5%, and the fourth 7.1%.

**Table 5. 6** The 10 first principal components of the final PCA. The *Standard deviation* is representing the variance in the data represented by the principal components.

|  | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 | PC9 | PC10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Standard deviation** | 2.153 | 1.843 | 0.975 | 0.840 | 0.506 | 0.207 | 0.077 | 0.065 | 0.036 | 0.013 |
| **Proportion of variance** | 0.464 | 0.340 | 0.095 | 0.071 | 0.026 | 0.004 | 0.001 | 0.000 | 0.000 | 0.000 |
| **Cumulative proportion** | 0.464 | 0.803 | 0.898 | 0.969 | 0.995 | 0.999 | 0.999 | 1.000 | 1.000 | 1.000 |

In Fig. 5.6, we present the influence of the selected variables on the principal components PC1 and PC2. AvTmean and Alt are both strongly influencing PC1, but not PC2. AvPrec, on the other hand, strongly influences PC2 while only lightly influencing PC1. The other variables

are showing different degrees of relevance for both components. In the same plot, we can see the PCA scores of the studied sites. We can cluster them into four groups. The first one, distributed on the negative side of PC2 and the positive one of PC1, includes the sites W3, W4, T1, and T2. The second one, with sites N, H2, B1, B2, and A, is situated on the opposite side of PC2. The third one, with the sites B18, B16, B12, B20, and B11, is situated along the negative side of PC1. Finally, the last cluster, which includes the remaining sites, is situated in the center of the plot.
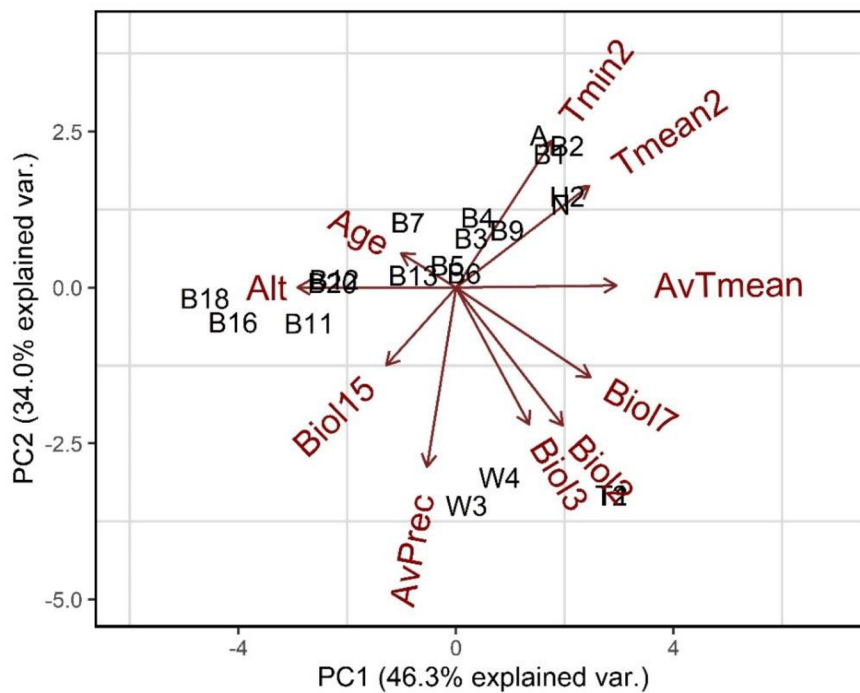


**Fig. 5. 6** Final PCA biplot with the 10 pre-selected environmental variables. On the x-axis, there is the first principal component *PC1,* and on the y-axis the second principal component *PC2*. The red arrows are representing the vector projections of the environmental variables on the PC1 and PC2 axis. The size of the arrows is representing the strength of the correlation with each component and the direction is indicating whether the correlation is positive or negative. The position of the 21 sites (A, B1, B2, etc.), marked in black, is representing the PC1 and PC2 scores for the respective sites.

In Table 5.7, we present the details of the importance of each variable in calculating the first four components. For the first two components, we report the same information as in Fig. 5.6, however, interestingly, we notice that while Age has a limited influence on the first two

components, it has a major influence on the third component. Similarly, Biol15 has its strongest influence on the fourth component.

**Table 5. 7** Importance of variables on the first 4 components of the final PCA. *EV* stands for environmental variables.

| EV | PC1 | PC2 | PC3 | PC4 |
|---|---|---|---|---|
| **Biol2** | 0.306 | -0.403 | 0.067 | -0.011 |
| **Biol3** | 0.210 | -0.400 | 0.286 | 0.105 |
| **Biol7** | 0.385 | -0.262 | -0.084 | -0.216 |
| **Biol15** | -0.200 | -0.227 | -0.392 | -0.825 |
| **AvPrec** | -0.084 | -0.524 | 0.092 | 0.138 |
| **AvTmean** | 0.460 | 0.007 | 0.029 | -0.121 |
| **Tmean2** | 0.382 | 0.297 | 0.014 | -0.131 |
| **Tmin2** | 0.276 | 0.430 | -0.028 | -0.115 |
| **Alt** | -0.456 | 0.001 | -0.012 | 0.048 |
| **Age** | -0.158 | 0.100 | 0.861 | -0.442 |

## 5.2.2 Statistical analyses

Assuming the single-site analyses, between 5 and 14 outliers were removed for each trait, based on the models' residual distributions. An illustration of the MAI-H data for site B3 is presented in Figs. 5.7 and 5.8. Fig. 5.7 showcases the residual plots containing all the data for site B3, while Fig. 5.8 depicts the same plots for site B3 after the removal of one outlier. This outlier can typically be found in the lower left corner of each plot.
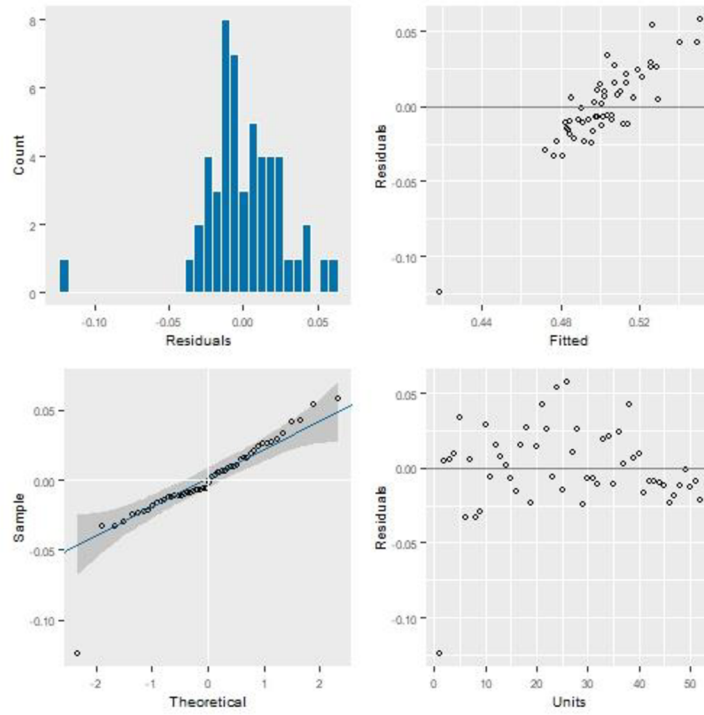
**Fig. 5. 7** Residual plots for the variable MAI-H at site B3 with all the data for this site.
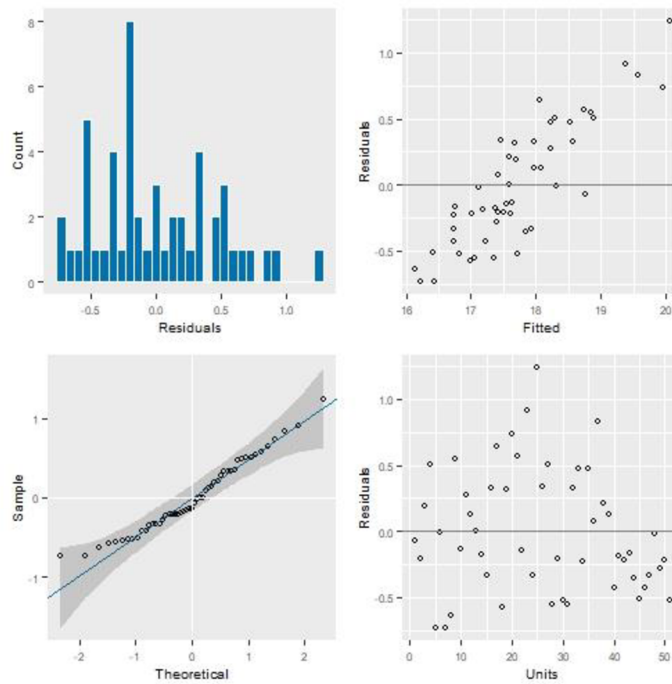


**Fig. 5. 8** Residual plots for the variable MAI-H at site B3, after removing one outlier.

The heritability estimates for each site are reported in Table 5.8; they vary from 0.01 to 0.99, with standard errors ranging from 0.01 to 0.68. For the whole dataset, the heritability estimates are 0.26 (SE= 0.07), 0.14 (SE= 0.05), and 0.27 (SE= 0.07) for MAI-H, PP, and MAI-DBH, respectively.

**Table 5. 8** Single sites narrow-sense heritability estimates $h$ with their associated standard errors $SE$.

| Site | MAI-H | | MAI-DBH | | PP | |
|---|---|---|---|---|---|---|
| | h | SE | h | SE | h | SE |
| A | 0.07 | 0.35 | 0.01 | 0.01 | 0.23 | 0.35 |
| B1 | 0.3 | 0.34 | 0.01 | 0.28 | 0.01 | 0.01 |
| B11 | 0.01 | 0.01 | 0.10 | 0.32 | 0.76 | 0.38 |
| B12 | 0.45 | 0.40 | 0.29 | 0.41 | 0.89 | 0.38 |
| B13 | 0.65 | 0.41 | 0.42 | 0.46 | 0.92 | 0.36 |
| B16 | 0.23 | 0.36 | 0.07 | 0.31 | 0.25 | 0.39 |
| B18 | 0.19 | 0.33 | 0.04 | 0.30 | 0.19 | 0.31 |
| B2 | 0.22 | 0.37 | 0.07 | 0.33 | 0.71 | 0.46 |
| B20 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| B3 | 0.45 | 0.43 | 0.15 | 0.39 | 0.67 | 0.43 |
| B4 | 0.24 | 0.28 | 0.29 | 0.28 | 0.40 | 0.31 |
| B5 | 0.34 | 0.44 | 0.48 | 0.42 | 0.56 | 0.41 |
| B6 | 0.58 | 0.39 | 0.17 | 0.33 | 0.50 | 0.41 |
| B7 | 0.36 | 0.37 | 0.09 | 0.41 | 0.01 | 0.01 |
| B9 | 0.47 | 0.52 | 0.98 | 0.44 | 0.33 | 0.47 |
| H2 | 0.01 | 0.01 | 0.02 | 0.36 | 0.61 | 0.42 |
| N | 0.99 | 0.27 | 0.34 | 0.22 | 0.62 | 0.25 |
| T1 | 0.99 | 0.01 | 0.99 | 0.01 | 0.62 | 0.34 |
| T2 | 0.99 | 0.01 | 0.95 | 0.56 | 0.16 | 0.68 |
| W3 | 0.68 | 0.52 | 0.01 | 0.01 | 0.01 | 0.01 |
| W4 | 0.6 | 0.35 | 0.28 | 0.33 | 0.27 | 0.32 |
| **Total** | **0.26** | **0.07** | **0.14** | **0.05** | **0.28** | **0.07** |

FA models from order 1 to 4 (FA1 to FA4) were fitted for each trait (see the summary in Table 5.9). The explained variance varied from 64.8% to 91.0%, 56.0% to 94.4%, and 65.9% to 92.2%, for MAI-H, MAI-DBH, and PP, respectively. For each trait, the log-likelihood consistently increased from FA1 to FA4 indicating better fits. On the contrary, the Akaike Information Criterion are indicating that the lower-order models have better fits. However, the likelihood ratio tests indicated, for both MAI-H and PP, that the FA2 models were significantly

better ($p < 0.10$) compared to the FA1 models, with respective $p$ values of 0.022 and 0.094. Since the higher-order models did not demonstrate any significant enhancement, we chose to utilize the second-order models for each trait in the subsequent analyses.

**Table 5. 9** Summary statistics of the fitted FA models (order 1 to 4), for the three studied traits: MAI-H, MAI-DBH, and PP. In the third column, we have the cumulative variance explained VE, then we have the loglikelihood of each model LogL, the Akaike's Information Criteria AIC, the p-value P of model k+1 over model k, based on a significance level of 0.10, and the number of variance components Parameters which are estimated by the respective models.

| Trait | Model | VE | LogL | AIC | P | Parameters |
|-------|-------|------|-------|----------|------|------------|
| MAI-H | FA1 | 64.8 | -2829 | **5728** | | 42 |
| | FA2 | 77.4 | -2813 | 5732 | **0.02** | 63 |
| | FA3 | 87.4 | -2804 | 5743 | 0.26 | 84 |
| | FA4 | 91.0 | **-2798** | 5762 | 0.68 | 105 |
| MAI-DBH | FA1 | 56.0 | -3817 | **7703** | | 42 |
| | FA2 | 67.9 | -3807 | 7719 | 0.35 | 63 |
| | FA3 | 89.3 | -3798 | 7731 | 0.26 | 84 |
| | FA4 | 94.4 | **-3790** | 7743 | 0.29 | 105 |
| PP | FA1 | 65.9 | -1093 | **2259** | | 42 |
| | FA2 | 77.3 | -1082 | 2268 | **0.09** | 63 |
| | FA3 | 85.6 | -1075 | 2285 | 0.59 | 84 |
| | FA4 | 92.2 | **-1066** | 2297 | 0.27 | 105 |

The averaged heritability estimates, based on the outputs of the selected FA2 models, were estimated at 0.79, 0.63, and 0.63 for MAI-H, MAI-DBH, and PP, respectively. The overall unbiased narrow-sense heritability estimates, corrected for GxE, were estimated at 0.41, 0.16, and 0.35 for MAI-H, MAI-DBH, and PP, respectively.

The genetic correlations are illustrated in Figs. 5.9, 5.10, and 5.11. In the case of MAI-H, the correlations are ranging from -0.53 to 0.99, with a mean of 0.53 (Table S2). In Fig. 5.9, we can see that sites A, B1, B11, B2, and B13 are displaying correlations approaching 1 which indicates a low GxE. The sites B5, B9, B16, B18, and T2, on the other hand, displayed close to null correlations indicating a high level of GxE. For this trait, 6.2 % of the correlations were negative, all of them being with site N.
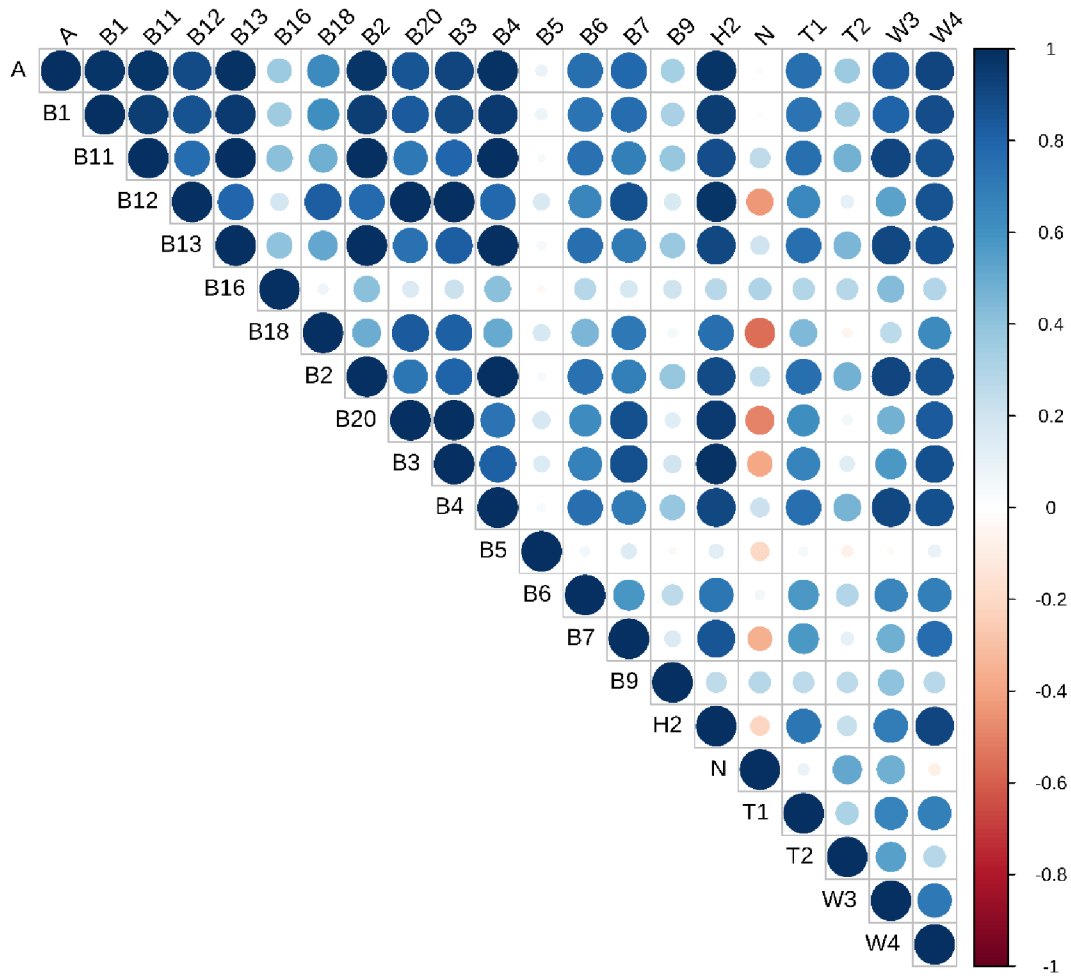
**Fig. 5. 9** Additive site-to-site genetic correlations generated by the FA2-MAI-H model. The blue color indicates positive correlations and the red color indicates negative ones. The size is also indicative of the strength of the correlation, the large circles signify correlations close to 1 or -1, and the small circles signify correlations close to 0.

For MAI-DBH (Fig. 5.10), the genetic correlations are ranging from -0.87 to 0.99, with a mean of 0.25 (Table S3). For this trait, the differences between sites are more visible, pointing to high GxE, especially for the sites H2, N, B4, B5, B6, B9, and B20. For this trait, there is a high proportion of negative values corresponding to 32.8% of the correlations.
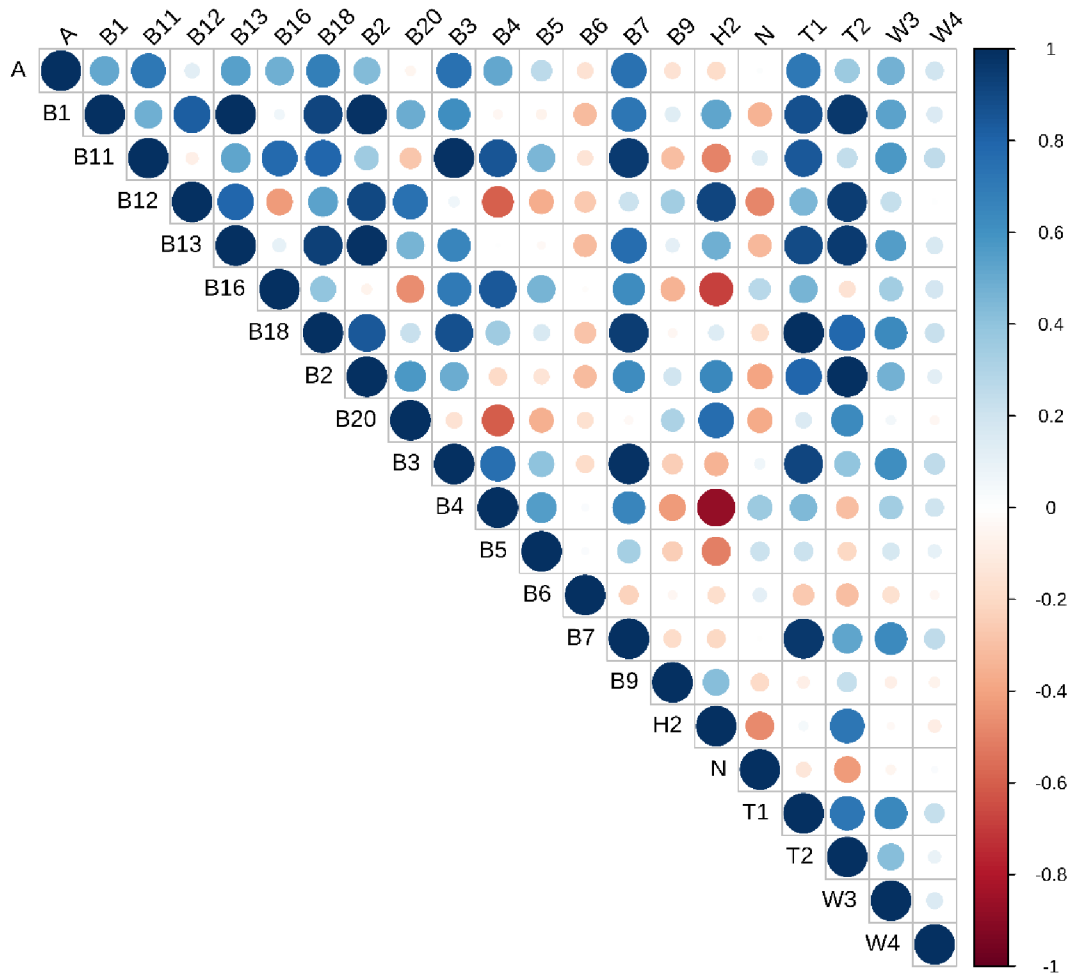
**Fig. 5. 10** Additive site-to-site genetic correlations generated by the FA2-MAI-DBH model. The blue color indicates positive correlations and the red color indicates negative ones. The size is also indicative of the strength of the correlation, the large circles signify correlations close to 1 or -1, and the small circles signify correlations close to 0.

In the case of PP (as depicted in Fig. 5.11), the genetic correlations span a range from -0.17 to 0.99, with an average value of 0.56 (Table S4). The majority of the sites exhibit moderate to strong correlations, with only 4.3% of them being negative. The sites H2 and B20, in particular, are associated with the lowest correlation values.
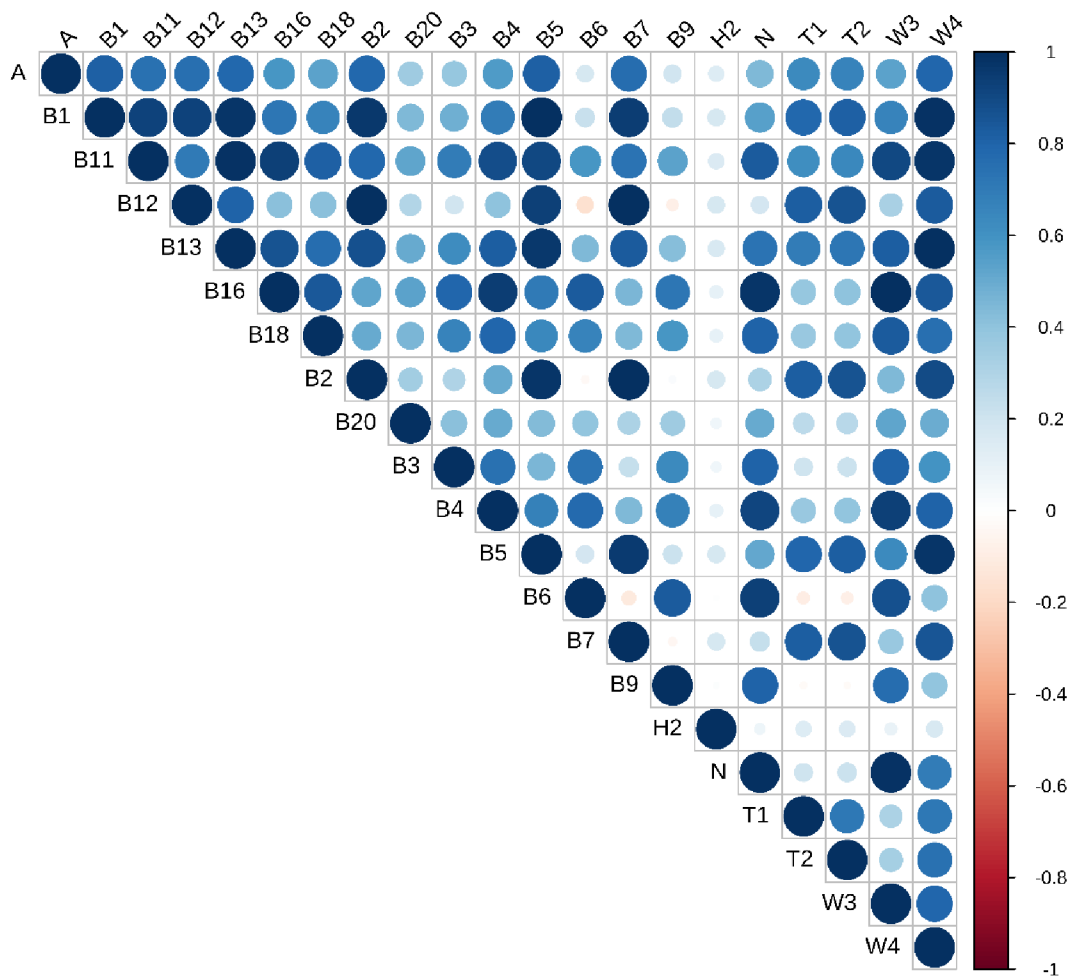
**Fig. 5. 11** Additive site-to-site genetic correlations generated by the FA2-PP model. The blue color indicates positive correlations and the red color indicates negative ones. The size is also indicative of the strength of the correlation, the large circles signify correlations close to 1 or -1, and the small circles signify correlations close to 0.

The latent regression plots at the provenance level for each trait are illustrated in Figs. 5.12, 5.13, and 5.14. In Fig. 5.12, for MAI-H, we can observe that the predicted site means are positively and significantly correlated with the first-factor loading, with a $R^2$ of 0.34 and a $p$ value of 0.003. However, the correlation with the second-factor loading is almost null, with a $R^2 < 0.01$ and a $p$ value of 0.563.
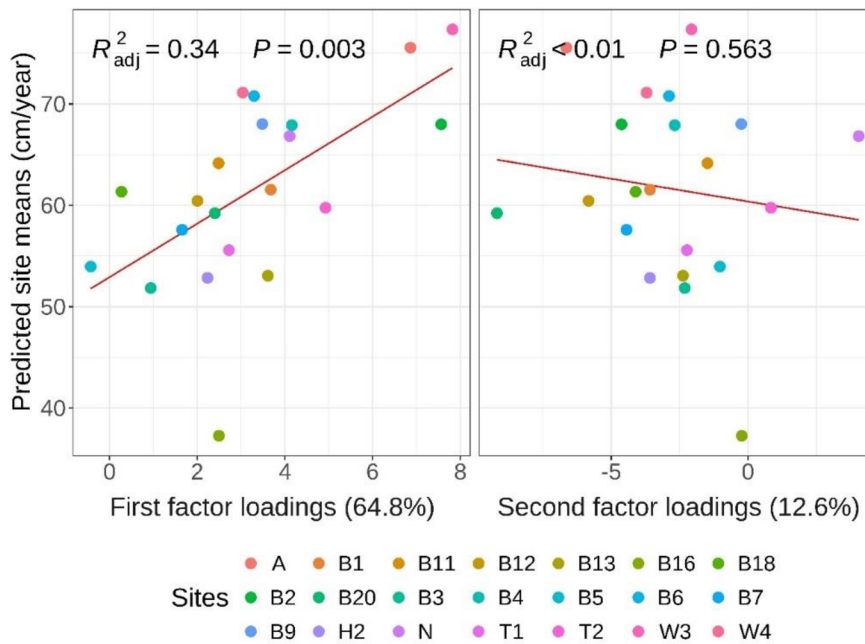
**Fig. 5. 12** Latent regression plots at the provenance level with the first and second-factor loadings for the FA2-MAI-H model. The red lines represent the linear regressions, the $R^2_{adj}$ correspond to their associated adjusted coefficient of determination, and the $P$ indicates the significance level of the correlation.

In Figs. 5.13 and 5.14, for the two other traits, low (from 0.11 to <0.01) and non-significant (from 0.08 to 0.37) correlations were found with both the first and the second-factor loadings. The observations are widely scattered around the plot with no discernable patterns. Noticeably, for the two growth traits, we can see that site 16 is showing considerably lower mean values (Figs 5.12, 5.13, and Table 5.10).
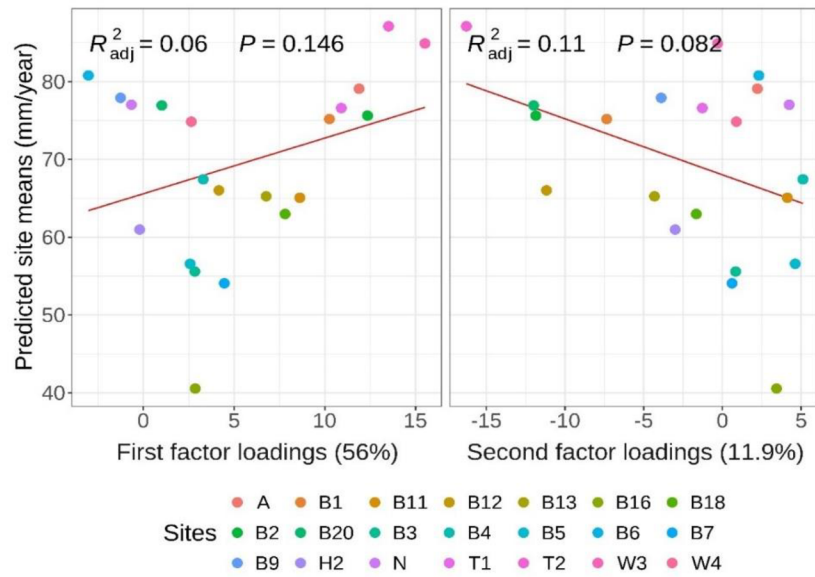
**Fig. 5. 13** Latent regression plots at the provenance level with the first and second-factor loadings for the FA2-MAI-DBH model. The red lines represent the linear regressions, the $R^2_{adj}$ correspond to their associated adjusted coefficient of determination, and the $P$ indicates the significance level of the correlation.
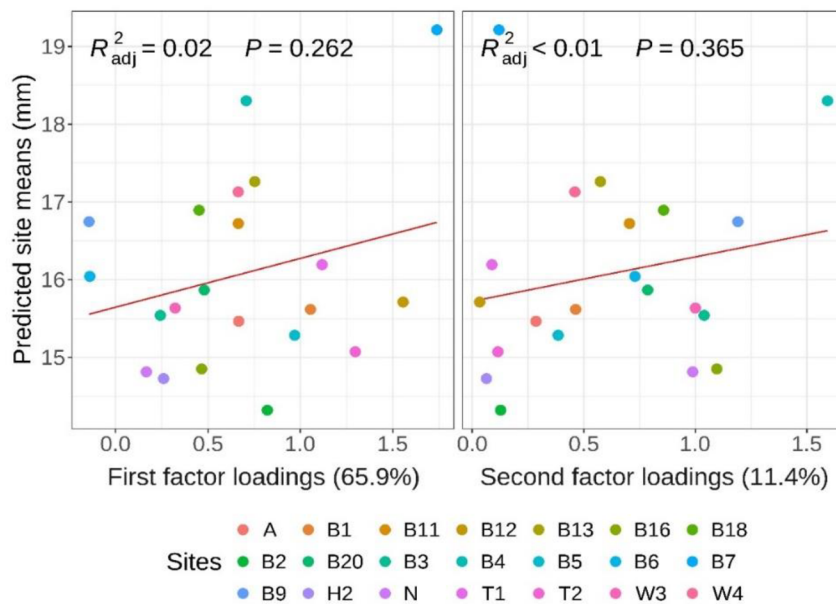


**Fig. 5. 14** Latent regression plots at the provenance level with the first and second-factor loadings for the FA2-PP model. The red lines represent the linear regressions, the $R^2_{adj}$ correspond to their associated adjusted coefficient of determination, and the $P$ indicates the significance level of the correlation.

**Table 5. 10** Phenotypic measurements summary for MAI-H (cm/year), MAI-DBH (mm/year), and PP (mm), at each site; *Nb* is the number of sampled individuals per site, and *min*, *max*, and *mean* are corresponding to the minimum, maximum and mean values for each trait. The numbers in bold indicate the sites with the lowest values of min, max, and mean.

| | **MAI-H** | | | | **MAI-DBH** | | | | **PP** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Site** | **Nb** | **min** | **max** | **mean** | **Nb** | **min** | **max** | **mean** | **Nb** | **min** | **max** | **mean** |
| **A** | 56 | 0.51 | 0.90 | 0.74 | 58 | 0.30 | 1.10 | 0.78 | 57 | 12.50 | 18.00 | 15.35 |
| **B1** | 63 | 0.44 | 0.70 | 0.61 | 63 | 0.33 | 0.95 | 0.73 | 63 | 12.50 | 19.50 | 15.61 |
| **B11** | 52 | 0.53 | 0.72 | 0.64 | 53 | 0.34 | 0.92 | 0.65 | 53 | 13.50 | 19.50 | 16.64 |
| **B12** | 54 | 0.38 | 0.68 | 0.57 | 54 | 0.29 | 0.94 | 0.63 | 54 | 12.00 | 20.00 | 15.54 |
| **B13** | 56 | 0.39 | 0.65 | 0.53 | 56 | 0.48 | 0.97 | 0.64 | 56 | 14.50 | 21.00 | 17.22 |
| **B16** | 57 | **0.22** | **0.46** | **0.37** | 57 | **0.15** | **0.64** | **0.41** | 57 | 12.00 | 18.00 | 14.96 |
| **B18** | 51 | 0.43 | 0.68 | 0.59 | 53 | 0.29 | 0.85 | 0.61 | 53 | 14.50 | 19.50 | 16.92 |
| **B2** | 47 | 0.49 | 0.86 | 0.67 | 47 | 0.46 | 1.06 | 0.73 | 47 | 12.00 | **17.00** | **14.30** |
| **B20** | 64 | 0.30 | 0.72 | 0.54 | 65 | 0.28 | 1.10 | 0.75 | 65 | 12.00 | 20.00 | 15.63 |
| **B3** | 51 | 0.44 | 0.61 | 0.51 | 51 | 0.37 | 0.79 | 0.56 | 52 | 11.50 | 19.50 | 15.60 |
| **B4** | 59 | 0.52 | 0.77 | 0.67 | 58 | 0.43 | 0.92 | 0.69 | 59 | 14.50 | 23.50 | 18.23 |
| **B5** | 40 | 0.42 | 0.66 | 0.54 | 40 | 0.31 | 0.82 | 0.59 | 38 | 12.00 | 18.50 | 15.11 |
| **B6** | 59 | 0.52 | 0.82 | 0.70 | 60 | 0.38 | 1.10 | 0.81 | 60 | 14.00 | 19.00 | 16.07 |
| **B7** | 49 | 0.42 | 0.67 | 0.56 | 48 | 0.34 | 0.75 | 0.54 | 49 | 15.00 | 24.50 | 19.19 |
| **B9** | 47 | 0.50 | 0.85 | 0.69 | 46 | 0.55 | 1.12 | 0.76 | 47 | 12.50 | 21.50 | 16.83 |
| **H2** | 53 | 0.39 | 0.59 | 0.52 | 53 | 0.39 | 0.83 | 0.61 | 51 | **10.50** | 19.00 | 14.71 |
| **N** | 195 | 0.48 | 0.87 | 0.68 | 198 | 0.47 | 1.05 | 0.79 | 199 | 11.00 | 18.50 | 14.73 |
| **T1** | 59 | 0.42 | 0.67 | 0.55 | 59 | 0.40 | 1.05 | 0.74 | 59 | 12.50 | 20.50 | 16.11 |
| **T2** | 31 | 0.37 | 0.72 | 0.61 | 31 | 0.27 | 1.14 | 0.80 | 31 | 12.00 | 18.00 | 14.90 |
| **W3** | 38 | 0.54 | 0.91 | 0.78 | 40 | 0.23 | 1.24 | 0.83 | 40 | 13.00 | 19.00 | 15.79 |
| **W4** | 58 | 0.56 | 0.78 | 0.70 | 58 | 0.45 | 1.12 | 0.75 | 58 | 14.50 | 20.00 | 17.13 |

In Fig. 5.15, for the trait MAI-H, we can see latent regression plots at the individual level with the eight genotypes that are represented by the highest number of individuals. Interestingly we can see several types of responses among individuals. If we take a look at the first-factor loading (Fig. 5.15 A), some individuals (i.e. L17, S24, and S6) are showing strong and positive correlations. Others (i.e. L8 and S21) are showing strong but negative correlations. The remaining individuals (i.e. L1, L10, S9) show relatively constant expressions along the factor loading with close to null correlations. Additionally, certain individuals possess higher breeding values than others across all sites. For example, genotype L1 mostly exhibits positive breeding values, in contrast to genotype L8, which primarily has negative values. Similar types of responses are found with the second-factor loading (Fig. 5.15 B). However, when comparing

the first and second-factor loadings, we can see that some individuals share the same expression across one of the factor loadings while having different expressions across the other one, with for example genotypes L10 and L17, or genotypes L17 and L8. Finally, it is interesting to note that while some genotypes are not correlated with one of the factor loadings, they are with the other one (e.g., L1 and S21).

Concerning MAI-DBH (Fig. 5.16), similar types of expression can be seen. However, for several genotypes, such as S21, L8, and S9, it appears that the observations are more scattered around the regression line, indicating wicker correlations.

In the case of PP (Fig. 5.17), we observe similar types of expressions as the two growth traits. Contrarily to MAI-DBH, for several genotypes such as L10, L8, and S24, the predicted breeding values are strongly correlated with both factor loadings with observation very close to the regression line.
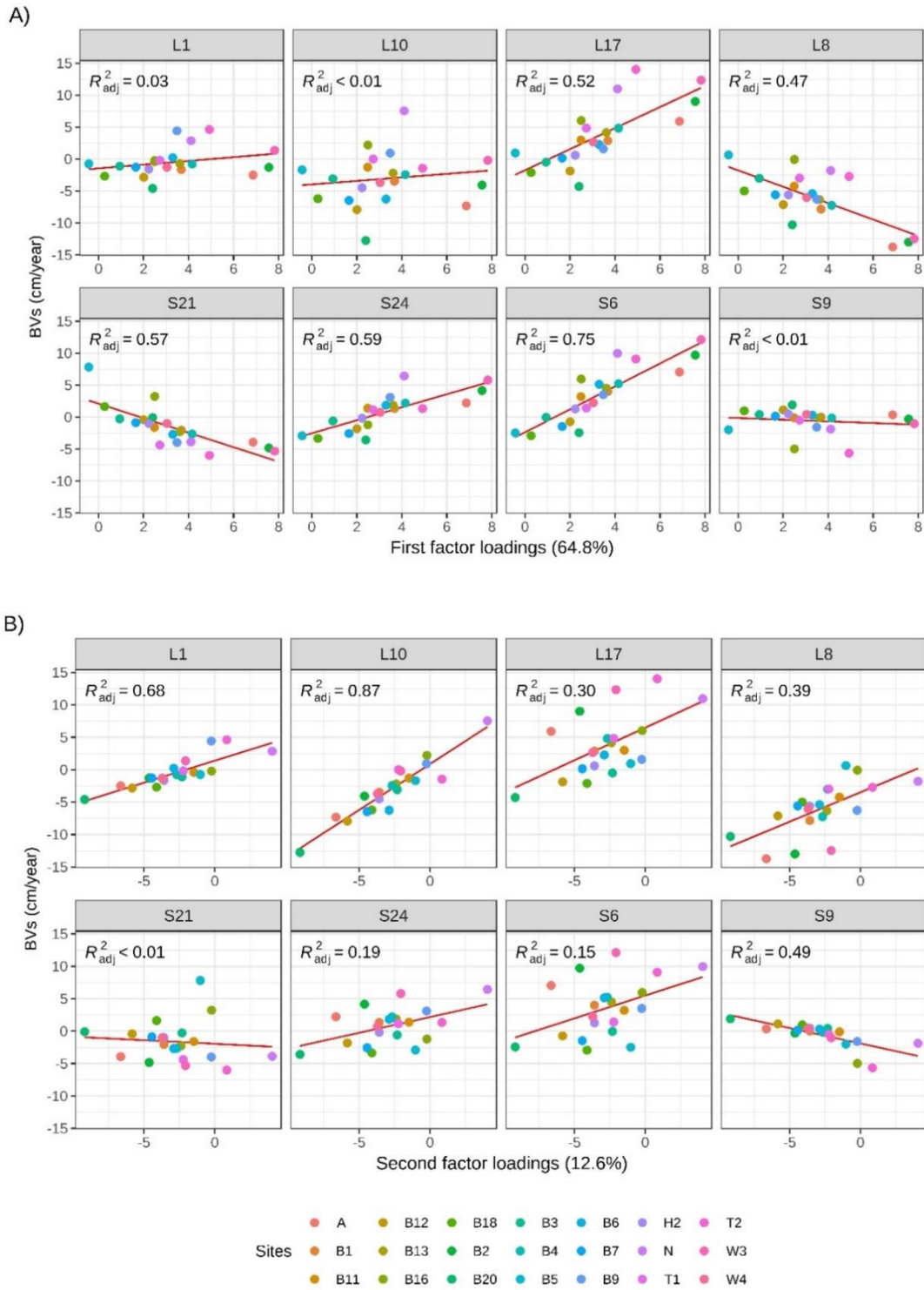
**Fig. 5. 15** Latent regression plots at the genotype level for the FA2-MAI-H model. The loadings of the first- and the second-factors are plotted in A and B, respectively. The red lines represent the linear regressions, the $R^2_{adj}$ correspond to their associated adjusted coefficient of determination.
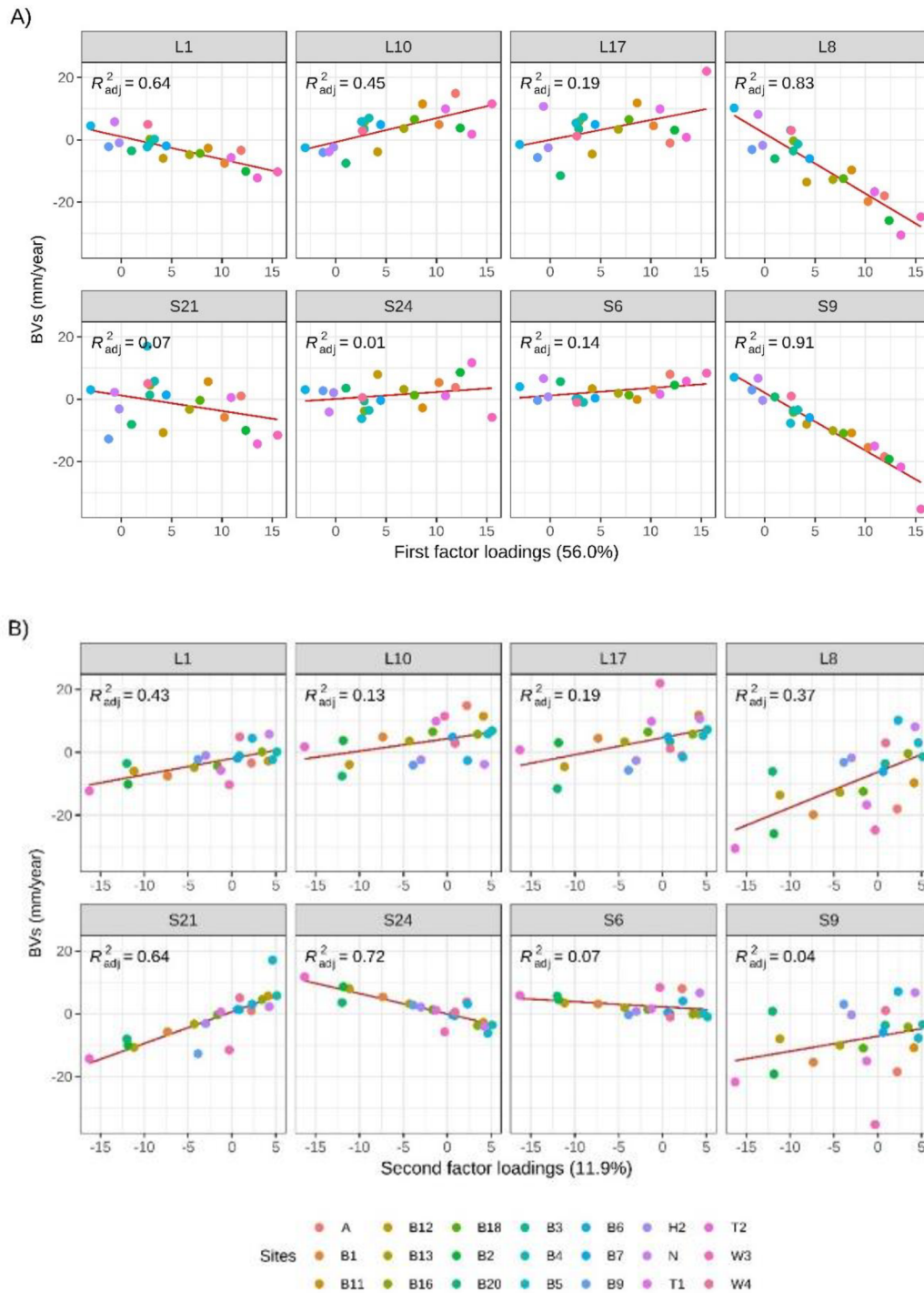
**Fig. 5. 16** Latent regression plots at the genotype level for the FA2-MAI-DBH model. The loadings of the first- and the second-factors are plotted in A and B, respectively. The red lines represent the linear regressions, the $R^2_{adj}$ correspond to their associated adjusted coefficient of determination.
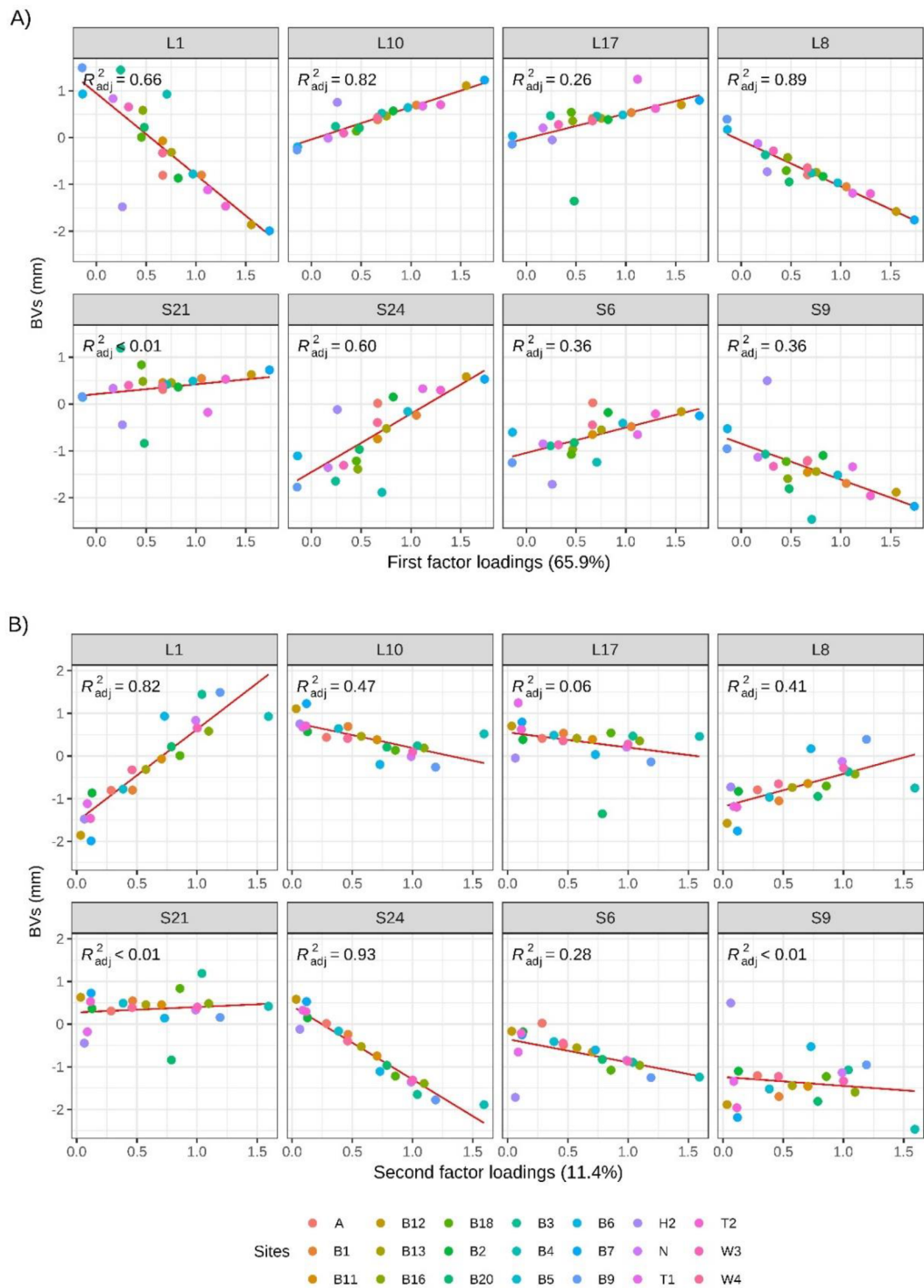
**Fig. 5. 17** Latent regression plots at the genotype level for the FA2-PP model. The loadings of the first- and the second-factors are plotted in A and B, respectively. The red lines represent the linear regressions, the $R^2_{adj}$ correspond to their associated adjusted coefficient of determination.

The results of the Pearson correlation coefficients, with their associated $p$-values, between the rotated loadings and the selected environmental variables are presented in Tables 5.11, 5.12, and 5.13, for MAI-H, MAI-DBH, and PP, respectively. For MAI-H, we established that five variables were significantly and moderately to highly correlated with the first-factor loading:

- Alt,

- Age,

- Biol7,

- Biol15,

- AvTmean.

Alt, Age, and Biol7 were negatively correlated while the other two variables were positively correlated. The absolute values of these correlations were situated between 0.40 and 0.50. AvTmean was also moderately and positively correlated with the second-factor loading (0.38).

**Table 5. 11** Pearson's correlation coefficients between the selected environmental variables *EV* and the loadings of the first- and second-factors (L1 and L2), for MAI-H. *P* represents the significance of the correlations based on a significance level of 0.10. The values in black are representing the significant correlations.

| EV | L1 | P | L2 | P |
|---|---|---|---|---|
| Alt | -0.48 | **0.03** | -0.20 | 0.39 |
| Age | -0.50 | **0.02** | -0.03 | 0.89 |
| Biol2 | 0.31 | 0.18 | 0.31 | 0.17 |
| Biol3 | 0.11 | 0.65 | 0.30 | 0.19 |
| Biol7 | 0.40 | **0.07** | 0.38 | **0.09** |
| Biol15 | -0.41 | **0.06** | 0.36 | 0.11 |
| Tmean2 | 0.30 | 0.19 | 0.16 | 0.48 |
| Tmin2 | 0.21 | 0.37 | 0.05 | 0.82 |
| AvPrec | 0.01 | 0.97 | 0.08 | 0.72 |
| AvTmean | 0.45 | **0.04** | 0.26 | 0.26 |

In the case of MAI-DBH, we identified that one variable, Biol2, was positively and significantly correlated with the first-factor loading with a correlation of 0.41 and a $p$-value of 0.06. No significant correlations were found for the second-factor loading.

**Table 5. 12** Pearson's correlation coefficients between the selected environmental variables *EV* and the loadings of the first- and second-factors (L1 and L2), for MAI-DBH. *P* represents the significance of the correlations based on a significance level of 0.10. The values in black are representing the significant correlations.

| EV | L1 | P | L2 | P |
|---|---|---|---|---|
| Alt | -0.29 | 0.21 | 0.07 | 0.76 |
| Age | -0.28 | 0.22 | 0.11 | 0.63 |
| Biol2 | 0.41 | **0.06** | -0.30 | 0.19 |
| Biol3 | 0.29 | 0.20 | -0.26 | 0.26 |
| Biol7 | 0.34 | 0.14 | -0.27 | 0.23 |
| Biol15 | -0.22 | 0.34 | -0.21 | 0.35 |
| Tmean2 | -0.04 | 0.87 | 0.14 | 0.56 |
| Tmin2 | -0.17 | 0.46 | 0.23 | 0.32 |
| AvPrec | 0.35 | 0.12 | -0.33 | 0.14 |
| AvTmean | 0.25 | 0.28 | -0.04 | 0.85 |

Finally, in the case of PP, we found that Alt was positively and significantly correlated with the second-factor loading with a correlation of 0.40 and a p-value of 0.07. No significant correlations were found with the first-factor loading.

**Table 5. 13** Pearson's correlation coefficients between the selected environmental variables *EV* and the loadings of the first- and second-factors (L1 and L2), for PP. *P* represents the significance of the correlations based on a significance level of 0.10. The values in black are representing the significant correlations.

| EV | L1 | P | L2 | P |
|---|---|---|---|---|
| Alt | -0.12 | 0.60 | **0.40** | **0.07** |
| Age | 0.22 | 0.34 | 0.07 | 0.76 |
| Biol2 | 0.15 | 0.52 | -0.25 | 0.28 |
| Biol3 | -0.04 | 0.87 | -0.16 | 0.48 |
| Biol7 | 0.12 | 0.59 | -0.22 | 0.34 |
| Biol15 | 0.05 | 0.84 | 0.23 | 0.32 |
| Tmean2 | -0.08 | 0.75 | -0.17 | 0.46 |
| Tmin2 | -0.13 | 0.59 | -0.09 | 0.69 |
| AvPrec | 0.15 | 0.52 | -0.02 | 0.95 |
| AvTmean | 0.05 | 0.82 | -0.28 | 0.22 |

# 6 Discussion

## 6.1 Discussion - Response function

In this study, we wanted to combine two known processes for studying the genetic variation within tree populations. The first one is the response function methodology, which focuses on the genetic adaptive variation across environments at the population level, and the second one is the conventional breeding, which focuses mainly on the intrapopulation genetic variation. Combining both methodologies aimed at capturing the intrapopulation adaptive response of European larch across the environmental gradient. Additionally, we aimed to propose a practical way to implement this methodology in the field.

At the population level, our results indicate a clear phenotypic variation across the studied environments. Additionally, we were able to illustrate how the genetic differences among individuals affected their performance across these environments. These results suggest that the methodology proposed here was successful in understanding the intrapopulation genetic variation of the studied European larch provenance in a specific set of environments.

In this study, we considered several environmental variables to represent our climatic gradient, and we chose to use MTCM (MTCM is the mean temperature of the coldest month; here the temperature of January). In the study by Chakraborty et al. (2018), the authors indicate how to use more than one climatic variable with the response function modeling. However, the set of variables selected using the random forest algorithm indicated high levels of collinearity, and for this reason, we only used one. Temperature is a major factor affecting tree growth, specifically European larch (Danek and Danek, 2022; Izworska et al., 2022). Additionally, the use of this specific variable is supported by Foff et al. (2014). In their study, the authors identified that cold temperature is an important factor limiting growth in the European larch species.

The use of the bivariate mixed linear animal genetic model, with the traits MAI-H and PP, indicated a significant improvement over the univariate model using only MAI-H

measurements. Derived from the model, we obtained typical heritability estimates, for both height and wood density (0.27 and 0.30 for MAI-H and PP, respectively). Similar genetic variation values are found in conifers (White et al., 2007) and specifically in European larch (Lstiburek et al., 2020; Ratcliffe et al., 2014). This indicates that more than a quarter of the phenotypic variation is attributable to direct allelic effects. The GxE was not significant in the bivariate model, this can either be because there are no such interactions in our population, or because we do not have sufficient data to detect it. If a significant GxE was found, it would be possible to calculate first-order partial derivatives taking into account the environmental variables of both the planting location and the provenance origin (Chakraborty et al., 2015; Wang et al., 2010).

The model regressed the phenotypic data onto random genetics (breeding values), and fixed site effects factors. The breeding values, as we can see in Fig. 5.1, unravel a relatively large gradient between the parents, with a difference of 11.4 cm/years between parents L11 and S7, and rather narrow and constant 95% confidence intervals (around ± 1.2 cm/year). This indicates that there is a high genetic diversity within the orchard, even though the trees were originally selected based, at least partially, on their heights. Reasonably, we can infer that the original provenance area showed similar or even greater genetic diversity, which is supported by several studies of European larch in the Alpine region (Belletti et al., 1997; Nardin et al., 2015; Pâques, 2013).

The fitting of the quadratic response function at the population level was significant, with a moderate value of the adjusted $R^2$. Similar values were published in several papers, including response function fittings for European larch and Norway spruce (Foff et al., 2014; Kapeller et al., 2017). The 95% confidence interval of the curve is very narrow across the curve but is wider on the extremities. We posit that it is due to having only a few sites on the colder and warmer ends of the MTCM gradient. This is supported by several publications where greater uncertainties were found on the edges of the response functions curves when having only a few sites, or even none, for those extreme conditions (Rehfeldt et al., 2001; Wang et al., 2006b). Rehfeldt et al. (2001), proposed to use the curve breadth size (for y = 0) of a well-represented

provenance (across the climatic gradient) as the maximum breadth size for under-represented provenances. Wang et al. (2006b) proposed another approach consisting of using the average trait values of the extreme sites for all the response functions, including the ones for provenance not represented in these extremes.

When fitting the quadratic response functions at the individual level, we obtained a large range of coefficients of variations depending on the parent under study, and all but one curve were significant. The average value of these coefficients is 0.28, similar to the value of the population analysis. The number of individuals representing each parent varied quite largely (35 to 120) because the planted trees are the products of natural crosses in the orchard, resulting that the size of each family could only be determined when reconstructing the pedigree. Interestingly, an examination of the coefficient values as a function of the number of offspring available for each parent reveals a particular pattern. While the mean value remains stable, the variability is significantly greater when there are only up to around 80 individual records, as compared to when the number is higher. This pattern suggests that, given similar study parameters (such as the environmental range), having a minimum of 80 to 100 individuals appears to be sufficient to achieve robust results. However, we did not find any study corroborating or arguing against this claim.

In the next step, we wanted to compare the different responses of the half-sib families, to determine whether the offspring behaved similarly across environments or not. When plotting all the quadratic response functions together (Fig. 5.5), we made several observations. We noticed that half-sib families performed differently in terms of the maximum pMAI-H, which, as for the simple ranking of the breeding values, provides information on better parents for higher growth. Noticeably, the ranking of breeding values in Fig. 5.1 is similar to the order of the curves. However, while most curves tend to keep their ranking throughout the environmental gradient, some exceptions are clearly visible. Some curves tend to be steeper, reaching their highest values for a narrow part of the climatic gradient while rapidly dropping in terms of performance on both sides of that optimum. We can postulate that the selection of these individuals would be interesting in localized areas. Other curves tend to be wider,

indicating more stable genotypes across environments. Selecting stable and highly performant genotypes is generally preferred in forestry and tree breeding as they can be used in larger areas. However, one may also select sets of genotypes that include both stable and environment-specific material leading to higher site variety.

In this study, the predicted phenotypes were used to plot individual response functions, which gives insight into the intra-population genetic diversity. Compared to the response functions methodology, this method provided additional and useful information for selecting valuable afforestation material for specific environments and can be used to further boost the adaptive response of forests to climate change. Additionally, compared to traditional breeding trials, this approach minimizes the resources (costs and time) needed to establish the actual experiments because all the activities, such as phenotyping and genotyping, would take place in operational afforestation sites implemented with designated seed sources. Furthermore, the unbalanced gametic contribution among and within the respective sites would be optimally taken into account within the combined genetic evaluation protocol (i.e., multi-site animal genetic model). While traditional breeding programs rely upon seed transfer among and within fixed seed zones, this approach is flexible. The delineation of seed transfer is dynamic, corresponding to the particular climate change development.

In this study, we selected only one environmental variable as the others available were found to be highly colinear. However, other types of data such as soil composition, sun exposure, slope steepness, and pest damage were not recorded for this study even though they are known to also influence tree growth (McLane et al., 2011; O'Neill et al., 2014; Sáenz-Romero et al., 2017). For instance, in this study, sites B5, B7, and B13 generally showed inferior performance than site B6, while sharing the same MTCM, indicating that other factors are likely at play. Additionally, while the studied sites were selected across an altitudinal gradient, the range of environmental conditions is relatively small. Hence, it is likely that more complete environmental records, and test sites selected on a broader scale, would give additional insights into the studied genotypes and their adaptation to the environment. Therefore, we believe that it would be an interesting and relevant research idea.

Moreover, the proposed strategy itself has several pitfalls that will be addressed here. First, while used in many response function models (Gömöry et al., 2012; Sáenz-Romero et al., 2017), the quadratic model fitting represents a simplification of traits' response to the environment. It assumes an increase in traits' values to a maximum, followed by an immediate decrease (Leites et al., 2012). In reality, the responses stem from complex genetic architectures of quantitative traits. The true underlying functions are likely non-parametric and non-linear.

Another limitation is associated with the future adaptation of new plantations established from the half-sib families of the selected parents. While these would most likely present a better adaptation to the climatic conditions in the short term, the evaluation of the long-term selection response across several generations is more complicated. Repeated selection cycles would influence environmental sensitivity according to the characteristics of the reaction norms (Kolmodin et al., 2002). The optimization of this methodology across the repeated selection cycles could be the aim of future research.

The particular finding of our investigation can be viewed as a case study that demonstrates the success of combining the response function methodology with in-situ large-scale genetic evaluation. We propose that forest tree breeders can implement this method according to the following steps. (1.) Identification of several stands afforested using a common seed source that is representing a specific population of origin, i.e., a provenance, (2.) measurement and evaluation of the phenotypes across sites followed by the reconstruction of the pedigree via DNA analysis (Lstibůrek et al., 2015), (3.) utilization of the phenotypic measurements performed across multiple sites coupled with the pedigree information in a multivariate statistical analysis (using an animal model such as ASReml) to predict the genetic worth of individual parental trees, (4.) selection of the main environmental gradients that are influencing the studied traits, (5.) development of the response functions, both at the individual- and population-level, to describe the genetic variability across the selected environmental gradient(s), (6.) selection of the best-adapted material for specific reforestation areas while also

accounting for the genetic diversity (Funda et al., 2009), and (7.) utilization and monitoring of the adapted forest reproductive material in-situ.

## 6.2 Discussion - Factor analytic

This study aimed at the investigation of a European larch provenance response to its environment. The main goal is the facilitation of gene resource management to allow the species to cope with environmental changes either in different planting sites or for climate change impact mitigation. More specifically, we wanted to fit a MET statistical model to assess the magnitude of GxE for three traits: height, diameter at breast height, and wood density. Furthermore, we sought to assess the impact of climate on this species by employing a complex factor analytic GxE model that took into account essential environmental covariates. Finally, we wanted to determine the GxE response dynamics for specific genotypes.

The factor analytic models fitting in this study were successful, and we found evidence of GxE for each trait. In addition, we found that genetic correlations between sites were more pronounced for the growth traits than for the wood density. Regarding the impact of the environment, we found several significant and moderately strong correlations between environmental variables and each of the traits. Our results suggest that height is positively influenced by temperature, balanced precipitation regimes, and decreasing altitude levels. In the case of diameter at breast height, our results indicate that the mean diurnal range is an important factor. Finally, for wood density, our analyses identified that altitude seems to be a significant influence. Interestingly, our results indicate highly variable responses among the genotypes of each trait, emphasizing the presence of GxE.

When looking at the relative positions of the environmental covariables within the final PCA plot, we can see that altitude and temperature-related variables are situated along the first component, which explains almost half of the variability. Hence, it indicates that they are the main drivers of the difference among test sites. We notice that altitude and temperature variables vectors are directed in opposite directions; this is consistent as the selected sites are situated along a relatively steep altitudinal gradient (about 500m), and that temperature

variables are usually negatively associated with altitude. Precipitation-related variables, on the other hand, are distributed along the second component that explains over a third of the variability, indicating lower importance. Since the geographical area covered by the study does not exceed 170km in length, it is expected to observe a low precipitation gradient, thus explaining these results. In the plot, we can see that age is represented by a small vector, which indicates that, even though relevant, it accounts for less variability between the sites than the other covariables. A substantial number of variables exhibited very high absolute correlations with each other. Consequently, the chosen set of seven variables seems to be a reasonable selection for the subsequent analyses.

The results indicated that the overall unbiased narrow-sense heritability estimates for the three traits were 0.41, 0.16, and 0.35 for MAI-H, MAI-DBH, and PP, respectively. These results are similar to previously reported heritability estimates for European larch (Lstiburek et al., 2020; Poupon et al., 2021; Ratcliffe et al., 2014).

Regarding the genetic correlations, we found interesting results with values ranging from -0.87 to 0.99. Some sites presented strong and positive correlations (> 0.60), some had lower values, and finally, others showed negative ones. The former type of interaction suggests a strong agreement concerning the genotypic ranking among sites. The second type suggests high GxE. The third type could be explained by genotypes exhibiting a degree of maladaptation to specific environmental conditions; alternatively, it could be explained by poor data quality on some of the sites (Oliveira et al., 2020). For both growth traits, it seems that we have mild to high GxE among many sites. However, in the case of MAI-DBH, we observed a high proportion (32.8%) of negative correlations. One possible interpretation can be that these values are connected with the inter-tree competition that may have introduced a certain level of noise in the model. Indeed, tree diameter growth is known to be highly correlated with the neighboring tree density (Ledermann, 2010). Regarding PP, we found that the majority of the genetic correlations were strong. This was expected as generally, wood density is found to be less sensitive to GxE (Fukatsu et al., 2010; Škorpík et al., 2018; Wielinga et al., 2009).

The latent regression plots, where the predicted site means were plotted as a function of the first- and second-factor loadings, were used to analyze the relationship between the larch provenance and the environments clustered in the loadings. For MAI-H, we found a significant but mild correlation between the site predictions and the first-factor loading (Fig. 5.12). This suggests that the studied genotypes have a resembling expression along the environments represented by the first-factor loading, with low to moderate GxE. Contrastingly, no clear pattern is observable along the second-factor loading. In the case of both MAI-DBH and PP, we also cannot see specific or significant patterns with either factor loading (Figs. 5.13 and 5.14). Interestingly, we can observe that site B16 consistently exhibits the lowest min, max, and mean prediction values for the growth traits (Figs. 5.12 and 5.13), which indicates poor growing conditions for this specific site.

The latent regression plots at the individual level were used to analyze patterns of genotypic expressions, such as GxE, for each trait (Figs. 5.15, 5.16, and 5.17). We can observe a wide range of responses. This can be interpreted as different adaptability levels and the presence of strong GxE between the genotypes. It is interesting to notice that some genotypes are displaying relatively constant breeding values along the studied environments, which indicates stable performances. This factor is crucial to consider when making informed breeding decisions. This is because stable genotypes, associated with high breeding values, are preferably selected by breeding programs in the forest industry to obtain superior planting seeds that are expected to grow and perform well in wide ranges of environmental conditions.

In the presented methodology, which aimed at identifying which environmental variables are influencing the genotypic performances along a gradient of environmental conditions, we calculated Pearson correlation coefficients between each of the selected variables and the two factor loadings, for each trait. To have a proper understanding of this methodology, one has to keep in mind that when a variable is positively correlated with a factor loading, and that this factor loading is also positively correlated with the predictions of the population or a specific genotype, then it is likely that there is a positive correlation between the variable and the population or this specific genotype. However, in the same case scenario, if the variable or the

factor loading is negatively correlated with the factor loading or the population/genotype, respectively, then it is likely that the environmental variable and the population or the specific genotype are negatively correlated. Analogous reasoning can be applied to all the other possible combinations of correlations.

We found that, for each trait, one or more environmental covariables were significantly as well as moderately highly correlated with the first- and/or second-factor loadings. In the case of MAI-H, five environmental variables were correlated with the first-factor loading. AvTmean and Biol7 (annual temperature range) were determined to be positively correlated. This indicates that the temperature probably influences height growth. Additionally, since the first-factor loading is also positively correlated with the population site means (Fig. 5.12), it suggests that an augmentation of the temperature has a positive effect on tree growth. This finding was not surprising as it is largely supported by various research (Carrer and Urbinati, 2006). However, dendroclimatic studies submitted that seasonal temperature variation, with for instance high summer temperatures combined with drought events, could have an adverse effect on radial growth (George et al., 2017; Koprowski, 2012). Therefore, we can infer that it might similarly affect height growth. Age was found to be negatively correlated which suggests a negative correlation with tree growth. This is quite interesting because age was accounted for in the model, by utilizing mean annual increment values instead of cumulative heights. Nevertheless, the annual growth rate is commonly known to diminish over time, and our model did not take that into account. Consequently, this could explain the importance of age in our results.

At the level of the genotypes, still regarding MAI-H, we observed that the majority of the individuals are, to some extent, positively influenced by temperature which is congruent with the results at the population level. However, for a few of them, a decreasing trend was observed (Fig. 5.15). This can be interpreted in several ways. For instance, previous studies identified significant genetic variation correlated to drought response among, as well as within, larch populations (George et al., 2017; Schueler et al., 2021). Therefore, some genotypes might indicate some resistance to higher temperatures, which might be more likely connected to

drought. Additionally, the specific genotypic differences in the phenological pattern of needle enfolding, with the early- and late-wood growth, as well as with the growth cessation may cause a certain level of interactions with temperature-related variables (George et al., 2019). Altitude was found to be negatively correlated with the first-factor loading (Table 5.11), nevertheless, as mentioned previously, an increase in altitude level is generally coupled with a decrease in temperature, hence it is evidence of consistent results. Now, regarding Biol15, which corresponds to the precipitation seasonality and is calculated as the coefficient of variation of precipitation for each month, we found that it is negatively correlated with the first-factor loading. This means that a more uneven precipitation distribution throughout the year, including potential drought spells, may have a negative influence on growth. Noticeably, no significant correlations were discovered with AvPrec.

In the case of MAI-DBH (Table 5.12), only one variable was significantly positively correlated with the first-factor loading: Biol2 (mean diurnal range). It is a temperature-related climatic variable that expresses the amplitude between the monthly maximum and minimum temperatures. The variable values are increasing with a greater difference. As mentioned previously, no clear pattern was found between the factor loadings and this trait at the population level, however, many strong correlations, either positive or negative, are observable at the genotypic level (Fig. 5.16). This indicates that, while this variable may have an influence on MAI-DBH, it does not seem to be a good predictor variable for this species. However, it might still be used when selecting specific genotypes.

Regarding the PP (Table 5.13), we found no correlations with the first-factor loading, but a positive correlation was found between Alt and the second-factor loading. In this model, only around 12% of the variability of PP is explained by the second-factor loading. Hence, it is important to consider that our results do not strongly support this correlation. However, several studies have already reported correlations between altitude and wood density for several species (Chave et al., 2006; Topaloğlu et al., 2016; Zhang et al., 2022).

In this research, we used a complex type of modeling to provide insight into this dataset. Substantial documentation recognizes the superiority of factor analytic modeling over the

traditional linear models used in MET analysis (Beeck et al., 2010; Cullis et al., 2014; Smith et al., 2001b, 2001a). Here, two factors explained on average 74% of the variation found in each trait, and coupled with the narrow-sense heritability estimates, this indicates a successful characterization of the GxE dynamics in this larch provenance. Additionally, the latent regression plots (Figs. 5.15, 5.16, and 5.17) exhibited contrasting but strong correlations between the genotypes and the factor loadings, which emphasizes the explanatory power of this type of modeling for GxE characterization and interpretation.

In tree breeding, it is fundamental to understand the dynamics of the GxE since tree selection usually aims at afforestation undertakings across a wide range of environmental conditions and landscape types. This is because a high GxE, which is represented by genetic correlations close to zero, indicates an unstable ranking of the studied genotypes across environments. Hence, it is necessary to perform rigorous GxE analyses, in MET, to enable improved accuracy and ameliorated ability for optimal genotypic selection. In the event of additional tests in the future, we suggest that sites sharing high correlations (such as the sites B1, B2, B11, B13, and A for MAI-H) could be clustered into a unique breeding zone. This would require less sampling and reduce the overall testing effort. Thus, it would allow the diversion and optimization of the available resources. For instance, more data could be obtained in each site, or more sites could be sampled, to improve the predictive power of the GxE modeling and thus, to improve the selection process.

In tree breeding, as previously discussed, genotypic stability across environments is considered to be an important factor for selection (White et al., 2007). Furthermore, when we take into account the climate change context, the trees' ability to endure changing climatic conditions is becoming more and more essential. We are hypothesizing that stable genotypes, as well as genotypes that are projected to perform increasingly well along the predicted changes in the future, might be more likely to present higher resistance (tree ability to resist external stressors) and resilience (tree ability to promptly return to their initial state). We believe that further research in that direction would be beneficial in the field of tree breeding.

The reconstructed pedigree that was used in this study resulted, as previously mentioned, in the calculation of reasonable heritability estimates for the three studied traits. Furthermore, the data analyses performed based on these additive relationships were successful in revealing clear and detailed patterns of genotypic behaviors. Thus, we are recommending the utilization of this (or equivalent) methodology of pedigree reconstruction for observational analyses that are based on tree measurements across a set of forest stands. Additionally, the reconstructed pedigree was adequate to support the selection of superior genotypes for afforestation projects. Finally, it supports the conclusion of Lstibůrek et al. (2020) that this method has the potential to be faster and financially superior to the conventional forest breeding approach.

The present study, because of the relatively steep altitudinal gradient across the Northern Alps foothills, encompasses an interesting range of environmental conditions situated around the lowest altitudinal ranges for European larch. Several studies have established that the warmest parts of tree species' distribution ranges are those that are the most likely to be negatively impacted by climate change and climate warming (Schueler and Liesebach, 2014). This highlights a strong point of this study. However, the covered geographical range remains relatively limited, and it only encompasses a small-scale portion of the European larch's range. Therefore, we postulate that identifying, sampling, and analyzing farther forest stands, afforested using the same provenance's seeds, could be an interesting way to extend the applicability of this study.

Regarding the data collection, we had the opportunity to use mature tree stands data which are excellent sources of information that, for example, new seedling test sites cannot always provide. Additionally, obtaining mature tree data, with known parentage and from specifically established study sites, is an excessively time-consuming and expensive process. The presented methodology allows the use of regular afforestation material, originating from a standard seed orchard, which greatly reduces the time and economic resources required. However, one disadvantage of this method is the unbalanced characteristic of the dataset (with a higher or lower representation of the genotypes). On one hand, due to its flexible nature, the factor analytic structure was able to provide interesting results. On the other hand, a more even

dataset would have likely provided more robust and precise results. In our opinion, we expect that sampling a larger number of stands, and/or trees per stand, would mitigate this issue.

Understanding the influence of the environmental covariable on the dataset was, from our perspective, quite successful. In this research, we had the possibility to study together the potential influence of several temperature and precipitation variables along with the elevation. For each of the studied traits, we established that one or several variables were significantly correlated (moderately to highly). Nonetheless, we believe that one other possibility to further our understanding of the data in future analyses would be to incorporate other types of covariables and indicators, such as slope aspect and steepness, soil composition and structure, homogeneity of the stands, and sun exposition levels to mention a few.

# 7 Conclusion

The two interrelated studies that constitute this dissertation have each contributed unique insights into the understanding of adaptive variation and environmental response within the European Larch (*Larix decidua*).

The first study delved into an area often neglected in forestry: the exploration of tree-to-tree variation within a population. Traditional response function methodology has been limited to capturing adaptive variation among populations, thereby missing the intra-population nuances. In breaking this paradigm, the first study presented a novel methodology that synergized response functions with in-situ breeding, employing extensive progeny trials of European larch. The findings of this research demonstrated that intra-population genetic variance can be harnessed to select trees that are genetically adapted to specific climatic conditions. This innovative approach holds broad applicability in both breeding and conservation programs. Importantly, it also offers a pathway to boost adaptation speed under the ever-evolving challenges of climate change, all while preserving high genetic diversity.

The second study in this dissertation aimed at comprehending the dynamics of European Larch's response to various environmental conditions. By analyzing data from 1253 trees planted across 21 sites using seeds from a single seed zone, the research evaluated key traits such as height, diameter at breast height, and wood density. The study revealed typical heritability estimates for European larch and discovered varying genetic correlations between sites, with growth traits generally exhibiting stronger genotype-by-environment interactions (GxE) than wood density. Notably, height was found to be positively influenced by factors like higher average monthly mean temperatures and balanced yearly precipitation but negatively affected by altitude. While the conclusions for diameter at breast height and wood density were more constrained, the research still identified key variables explaining a substantial portion of the variability, such as mean diurnal range for the former and altitude for the latter.

These findings have broader implications, not only in understanding the complex responses of tree species to their environment but also in informing tree breeding programs for the selection

of suitable genotypes. These genotypes can be tailored to specific present and future conditions, thus enhancing the resilience and adaptability of forest ecosystems. Together, these two research papers complement one another in providing a multifaceted view of the adaptive dynamics within European Larch populations. The methodologies and insights gleaned from both studies offer valuable tools for the forestry industry, paving the way for more informed and effective breeding and conservation strategies. In addressing both intra-population genetic variance and the nuanced interplay between genotype and environment, this dissertation underscores the importance of a harmonized approach to tree adaptation and conservation. As the world grapples with the challenges of climate change, such insights are pivotal in ensuring the sustainable management and resilience of forest ecosystems.

# 8 References

Aitken, S.N., Bemmels, J.B., 2016. Time to get moving: Assisted gene flow of forest trees. Evol Appl 9, 271–290. https://doi.org/10.1111/eva.12293

Aitken, Sally N., Yeaman, S., Holliday, J.A., Wang, T., Curtis-McLane, S., 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol Appl 1, 95–111. https://doi.org/10.1111/j.1752-4571.2007.00013.x

Aitken, Sally N, Yeaman, S., Holliday, J.A., Wang, T., Curtis-mclane, S., 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol Appl 1, 95–111. https://doi.org/10.1111/j.1752-4571.2007.00013.x

Alan, M., 2020. Silviculture and tree breeding for planted forests. Eurasian Journal of Forest Science 8, 74–83. https://doi.org/10.3195/ejejfs.661352

Albecker, M.A., Trussell, G.C., Lotterhos, K.E., 2022. A novel analytical framework to quantify co-gradient and countergradient variation. Ecol Lett 25, 1521–1533. https://doi.org/10.1111/ele.14020

Alberto, F.J., Aitken, S.N., Alía, R., González-Martínez, S.C., Hänninen, H., Kremer, A., Lefèvre, F., Lenormand, T., Yeaman, S., Whetten, R., Savolainen, O., 2013. Potential for evolutionary responses to climate change - evidence from tree populations. Glob Chang Biol 19, 1645–1661. https://doi.org/10.1111/gcb.12181

Alexander, L. V., Zhang, X., Peterson, T.C., Caesar, J., Gleason, B., Klein Tank, A.M.G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D.B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., Vazquez-Aguirre, J.L., 2006. Global observed changes in daily climate extremes of temperature and precipitation. Journal of Geophysical Research Atmospheres 111, 1–22. https://doi.org/10.1029/2005JD006290

Allan, R.P., Soden, B.J., 2008. Atmospheric warming and the amplification of precipitation extremes. Science (1979) 321, 1481–1484. https://doi.org/10.1126/science.1160787

Allen, C.D., Breshears, D.D., McDowell, N.G., 2015. On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene. Ecosphere 6, 1–55. https://doi.org/10.1890/ES15-00203.1

Allen, C.D., Macalady, A.K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., Kitzberger, T., Rigling, A., Breshears, D.D., Hogg, E.H. (Ted), Gonzalez, P., Fensham, R., Zhang, Z., Castro, J., Demidova, N., Lim, J.H., Allard, G., Running, S.W., Semerci, A., Cobb, N., 2010. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. For Ecol Manage 259, 660–684. https://doi.org/10.1016/j.foreco.2009.09.001

Anderegg, W.R.L., Schwalm, C., Biondi, F., Camarero, J.J., Koch, G., Litvak, M., Ogle, K., Shaw, J.D., Shevliakova, E., Williams, A.P., Wolf, A., Ziaco, E., Pacala, S., 2015. Pervasive drought legacies in forest ecosystems and their implications for carbon cycle models. Science (1979) 349, 528–532. https://doi.org/10.1126/science.aab1833

Angélil, O., Stone, D., Wehner, M., Paciorek, C.J., Krishnan, H., Collins, W., 2017. An independent assessment of anthropogenic attribution statements for recent extreme temperature and rainfall events. J Clim 30, 5–16. https://doi.org/10.1175/JCLI

Arrhenius, S., 1896. XXXI. On the influence of carbonic acid in the air upon the temperature of the ground. Philosophical Magazine Series 5 41, 237–276. https://doi.org/10.1080/14786449608620846

Barriopedro, D., Fischer, E.M., Luterbacher, J., Trigo, R.M., García-Herrera, R., 2011. The hot summer of 2010: Redrawing the temperature record map of Europe. Science (1979) 332, 220–224. https://doi.org/10.1126/science.1201224

Beeck, C.P., Cowling, W.A., Smith, A.B., Cullis, B.R., 2010. Analysis of yield and oil from a series of canola breeding trials. Part I. Fitting factor analytic mixed models with pedigree information, in: Genome. pp. 992–1001. https://doi.org/10.1139/G10-051

Belletti, P., Lanteri, S., Leonardi, S., 1997. Genetic variability among European larch (larix decidua mill.) populations in piedmont, North-Western Italy.

Beniston, M., Stephenson, D.B., Christensen, O.B., Ferro, C.A.T., Frei, C., Goyette, S., Halsnaes, K., Holt, T., Jylhä, K., Koffi, B., Palutikof, J., Schöll, R., Semmler, T., Woth, K., 2007. Future extreme events in European climate: An exploration of regional climate model projections. Clim Change 81, 71–95. https://doi.org/10.1007/s10584-006-9226-z

Bennett, K., Tzedakis, P., Willis, K., 1991. Quaternary refugia of North European trees. J Biogeogr 18, 103–115. https://doi.org/10.2307/2845248

Bernardo, R., 2020. Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. Heredity (Edinb). https://doi.org/10.1038/s41437-020-0312-1

Bonan, G.B., 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. Science (1979) 320, 1444–1449. https://doi.org/10.1126/science.1155121

Borenstein, M., Hedges, L. V., Higgins, J.P.T., Rothstein, H.R., 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods 1, 97–111. https://doi.org/10.1002/jrsm.12

Bower, A.D., Aitken, S.N., 2008. Ecological genetics and seed transfer guidelines for Pinus albicaulis (Pinaceae). Am J Bot 95, 66–76. https://doi.org/10.3732/ajb.95.1.66

Brown, G.R., Gill, G.P., Kuntz, R.J., Langley, C.H., Neale, D.B., 2004. Nucleotide diversity and linkage disequilibrium in loblolly pine. Proc Natl Acad Sci U S A 101, 15255–15260. https://doi.org/10.1073/pnas.0404231101

Bucharova, A., 2017. Assisted migration within species range ignores biotic interactions and lacks evidence. Restor Ecol 25, 14–18. https://doi.org/10.1111/rec.12457

Burgueño, J., Crossa, J., Cornelius, P.L., Yang, R.-C., 2008. Using Factor Analytic Models for Joining Environments and Genotypes without Crossover Genotype × Environment Interaction. Crop Sci 48, 1291–1305. https://doi.org/10.2135/cropsci2007.11.0632

Burgueño, J., Crossa, J., Cotes, J.M., Vicente, F.S., Das, B., 2011. Prediction assessment of linear Mixed models for multi-environment trials. Crop Sci 51, 944–954. https://doi.org/10.2135/cropsci2010.07.0403

Burley, J., Evans, J., Youngquist, J.A. (Eds.), 2004. Encyclopedia of forest sciences. Academic Press.

Bussotti, F., Pollastrini, M., Holland, V., Brüggemann, W., 2015a. Functional traits and adaptive capacity of European forests to climate change. Environ Exp Bot 111, 91–113. https://doi.org/10.1016/j.envexpbot.2014.11.006

Bussotti, F., Pollastrini, M., Holland, V., Brüggemann, W., 2015b. Functional traits and adaptive capacity of European forests to climate change. Environ Exp Bot 111, 91–113. https://doi.org/10.1016/j.envexpbot.2014.11.006

Butler, D.G., Cullis, B.R., Gilmour, A.R., Gogel, B.J., Thompson, R., 2017. ASReml-R reference manual version 4. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Campbell, R.K., 1979. Genecology of Douglas-fir in a watershed in the Oregon cascades. Ecology 60, 1036–1050. https://doi.org/10.2307/1936871

Carrasco, G., Almeida, A.C., Falvey, M., Olmedo, G.F., Taylor, P., Santibañez, F., Coops, N.C., 2022. Effects of climate change on forest plantation productivity in Chile. Glob Chang Biol 28, 7391–7409. https://doi.org/10.1111/gcb.16418

Carrer, M., Urbinati, C., 2006. Long-term change in the sensitivity of tree-ring growth to climate forcing in Larix decidua. New Phylologist 170, 861–872. https://doi.org/10.1111/j.1469-8137.2006.01703.x

Carrer, M., Urbinati, C., 2004. Age-dependent tree-ring growth responses to climate in Larix decidua and Pinus cembra. Ecology 85, 730–740. https://doi.org/10.1890/02-0478

Chakraborty, D., Schueler, S., Lexer, M.J., Wang, T., 2018. Genetic trials improve the transfer of Douglas-fir distribution models across continents. Ecography. https://doi.org/10.1111/ecog.03888

Chakraborty, D., Wang, T., Andre, K., Konnert, M., Lexer, M.J., Matulla, C., Schueler, S., 2015. Selecting populations for non-analogous climate conditions using universal response functions: The case of douglas-fir in central Europe. PLoS One. https://doi.org/10.1371/journal.pone.0136357

Chakraborty, D., Wang, T., Andre, K., Konnert, M., Lexer, M.J., Matulla, C., WeiÃŸenbacher, L., Schueler, S., 2016. Adapting Douglas-fir forestry in Central Europe: evaluation, application, and uncertainty analysis of a genetically based model. Eur J For Res 135, 919–936. https://doi.org/10.1007/s10342-016-0984-5

Chave, J., Muller-Landau, H.C., Baker, T.R., Easdale, T.A., Steege, H.T., Webb, C.O., 2006. Regional and phylogenetic variation of wood density across 2456 neotropical tree species. Ecological Applications 16, 2456–2367. https://doi.org/10.1890/1051-0761(2006)016[2356:rapvow]2.0.co;2

Clark, J.S., Fastie, C., Hurtt, G., Jackson, S.T., 1998. Reid ' s paradox of rapid plant migration. Bioscience 48, 13–24.

Cown, D.J., 1978. Comparison of the pilodyn and torsiometer methods for the rapid assessment of wood density in living trees. N Z J For Sci 8, 384–391.

Cullis, B.R., Jefferson, P., Thompson, R., Smith, A.B., 2014. Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a Pinus radiata breeding programme. Theoretical and Applied Genetics 127, 2193–2210. https://doi.org/10.1007/s00122-014-2373-0

Cullis, B.R., Smith, A.B., Beeck, C.P., Cowling, W.A., 2010. Analysis of yield and oil from a series of canola breeding trials. Part II. Exploring variety by environment interaction using factor analysis, in: Genome. pp. 1002–1016. https://doi.org/10.1139/G10-080

Dai, A., 2011. Drought under global warming: A review. Wiley Interdiscip Rev Clim Change 2, 45–65. https://doi.org/10.1002/wcc.81

Dai, A., Zhao, T., Chen, J., 2018. Climate change and drought: a precipitation and evaporation perspective. Curr Clim Change Rep. https://doi.org/10.1007/s40641-018-0101-6

Danek, M., Chuchro, M., 2019. Temporal variations in tree-ring growth response to climate of European larch (Larix decidua Mill.): An example from Sudetes, Poland, in: IOP Conference Series: Earth and Environmental Science. Institute of Physics Publishing. https://doi.org/10.1088/1755-1315/221/1/012117

Danek, M., Danek, T., 2022. Recent changes in the climate-growth response of European larch (Larix decidua Mill.) in the Polish Sudetes. Trees - Structure and Function 36, 803–817. https://doi.org/10.1007/s00468-021-02251-3

Davis, M.B., Shaw, R.G., 2001. Range shifts and adaptive responses to quaternary climate change. Science (1979) 292, 673–679. https://doi.org/10.1126/science.292.5517.673

Demesure, B., Comps, B., Petit, R.J., 1996. Chloroplast DNA phylogeography of the common beech ( Fagus sylvatica L .) in Europe. Evolution (N Y) 50, 2515–2520.

Dumolin-Lapègue, S., Démesure, B., Fineschi, S., Le Corre, V., Petit, R.J., 1997. Phylogeographic structure of white oaks throughout the european continent. Genetics 146, 1475–1487.

Easterling, D.R., Meehl, G.A., Parmesan, C., Changnon, S.A., Karl, T.R., Mearns, L.O., 2000. Climate extremes: Observations, modeling, and impacts. Science (1979) 289, 2068–2074. https://doi.org/10.1126/science.289.5487.2068

El-Kassaby, Y., Klapste, J., 2016. Tree breeding – challenges and opportunities, in: Forestry for the Curious: Why Study Forestry? The curious academic publishing, pp. 1–152.

El-kassaby, Y.A., Lstibůrek, M., 2009. Breeding without breeding. Genet Res (Camb) 91, 111–120. https://doi.org/10.1017/S001667230900007X

Finkel, J.M., Katz, J.I., 2018. Changing world extreme temperature statistics. International journal of climatology 38, 2613–2617. https://doi.org/10.1002/joc.5342

Fisher, R.A., 1919. XV.—The correlation between relatives on the supposition of Mendelian Inheritance. Transactions of the Royal Society of Edinburgh 52, 399–433. https://doi.org/10.1017/S0080456800012163

Foff, V., Weiser, F., Foffová, E., Gömöry, D., 2014. Growth response of European larch (Larix decidua Mill.) populations to climatic transfer. Silvae Genet 63, 67–75. https://doi.org/10.1515/sg-2014-0010

Frank, A., Howe, G.T., Sperisen, C., Brang, P., Clair, J.B.S., Schmatz, D.R., Heiri, C., 2017a. Risk of genetic maladaptation due to climate change in three major European tree species. Glob Chang Biol 23, 5358–5371. https://doi.org/10.1111/gcb.13802

Frank, A., Sperisen, C., Howe, G.T., Brang, P., Walthert, L., Clair, J.B.S., Heiri, C., 2017b. Distinct genecological patterns in seedlings of Norway spruce and silver fir from a mountainous landscape. Ecology 98, 211–227. https://doi.org/10.1002/ecy.1632

Frank, Dorothea, Reichstein, M., Bahn, M., Thonicke, K., Frank, David, Mahecha, M.D., Smith, P., van der Velde, M., Vicca, S., Babst, F., Beer, C., Buchmann, N., Canadell, J.G., Ciais, P., Cramer, W., Ibrom, A., Miglietta, F., Poulter, B., Rammig, A., Seneviratne, S.I., Walz, A., Wattenbach, M., Zavala, M.A., Zscheischler, J., 2015. Effects of climate extremes on the terrestrial carbon cycle: Concepts, processes and potential future impacts. Glob Chang Biol 21, 2861–2880. https://doi.org/10.1111/gcb.12916

Frei, C., Schöll, R., Fukutome, S., Schmidli, J., Vidale, P.L., 2006. Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models. Journal of Geophysical Research Atmospheres 111. https://doi.org/10.1029/2005JD005965

Fréjaville, T., Fady, B., Kremer, A., Ducousso, A., Benito Garzón, M., 2019. Inferring phenotypic plasticity and population responses to climate across tree species ranges using forest inventory data. Global Ecology and Biogeography 28, 1259–1271. https://doi.org/10.1111/geb.12930

Frich, P., Alexander, L. V., Della-Marta, P., Gleason, B., Haylock, M., Tank Klein, A.M.G., Peterson, T., 2002. Observed coherent changes in climatic extremes during the second half of the twentieth century. Clim Res 19, 193–212. https://doi.org/10.3354/cr019193

Fukatsu, E., Tamura, A., Takahashi, M., Fukuda, Y., Nakada, R., Kubota, M., Kurinobu, S., 2010. Efficiency of the indirect selection and the evaluation of the genotype by environment interaction using Pilodyn for the genetic improvement of wood density in Cryptomeria japonica. Journal of forest research 16, 128–135. https://doi.org/10.1007/s10310-010-0217-6 Pages

Funda, T., Lstibůrek, M., Lachout, P., Klápště, J., El-Kassaby, Y.A., 2009. Optimization of combined genetic gain and diversity for collection and deployment of seed orchard crops. Tree Genet Genomes 5, 583–593. https://doi.org/10.1007/s11295-009-0211-3

Garcia-Ramos, G., Kirkpatrick, M., 1997. Genetic models of adaptation and gene flow in peripheral populations. Evolution (N Y) 51, 21–28. https://doi.org/10.1111/j.1558-5646.1997.tb02384.x

Geburek, T., 2021. "Der apfel fällt nicht weit vom stamm!" lärche XXL - neues forschungsprojekt zur Leistungssteigerung. [WWW Document]. URL https://bfw.ac.at/rz/bfwcms.web?dok=9859 (accessed 8.28.23).

George, J., Grabner, M., Campelo, F., Karanitsch-ackerl, S., Mayer, K., Klumpp, R.T., Schüler, S., 2019. Intra-specific variation in growth and wood density traits under water-limited conditions: Long-term-, short-term-, and sudden responses of four conifer tree species. Science of the Total Environment 660, 631–643. https://doi.org/10.1016/j.scitotenv.2018.12.478

George, J.P., Grabner, M., Karanitsch-Ackerl, S., Mayer, K., Weißenbacher, L., Schueler, S., 2017. Genetic variation, phenotypic stability, and repeatability of drought response in European larch throughout 50 years in a common garden experiment. Tree Physiol 37, 33–46. https://doi.org/10.1093/treephys/tpw085

Giorgi, F., Raffaele, F., Coppola, E., 2019. The response of precipitation characteristics to global warming from climate projections. Earth System Dynamics 10, 73–89. https://doi.org/10.5194/esd-10-73-2019

Gömöry, D., Longauer, R., Hlásny, T., Pacalaj, M., Strmeň, S., Krajmerová, D., 2012. Adaptation to common optimum in different populations of Norway spruce (Picea abies Karst.). Eur J For Res 131, 401–411. https://doi.org/10.1007/s10342-011-0512-6

Gugger, P.F., Sugita, S., 2010. Glacial populations and postglacial migration of Douglas-fir based on fossil pollen and macrofossil evidence. Quat Sci Rev 29, 2052–2070. https://doi.org/10.1016/j.quascirev.2010.04.022

Gunderson, C.A., O'hara, K.H., Campion, C.M., Walker, A. V., Edwards, N.T., 2010. Thermal plasticity of photosynthesis: The role of acclimation in forest responses to a warming climate. Glob Chang Biol 16, 2272–2286. https://doi.org/10.1111/j.1365-2486.2009.02090.x

Hallauer, A.R., 2007. History, contribution, and future of quantitative genetics in plant breeding: Lessons from maize, in: Crop Science. https://doi.org/10.2135/cropsci2007.04.0002IPBS

Halofsky, J.E., Peterson, D.L., Harvey, B.J., 2020. Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the Pacific Northwest, USA. Fire Ecology. https://doi.org/10.1186/s42408-019-0062-8

Hamann, A., Gylander, T., Chen, P. yu, 2011. Developing seed zones and transfer guidelines with multivariate regression trees. Tree Genet Genomes 7, 399–408. https://doi.org/10.1007/s11295-010-0341-7

Hampe, A., Petit, R.J., 2005. Conserving biodiversity under climate change: The rear edge matters. Ecol Lett. https://doi.org/10.1111/j.1461-0248.2005.00739.x

Hasenauer, H., Nemani, R.R., Schadauer, K., Running, S.W., 1999. Forest growth response to changing climate between 1961 and 1990 in Austria. For Ecol Manage 122, 209–219. https://doi.org/10.1016/S0378-1127(99)00010-9

Henderson, C.R., 1984. Applications of linear models in animal breeding. Guelph.

Heuertz, M., De Paoli, E., Källman, T., Larsson, H., Jurman, I., Morgante, M., Lascoux, M., Gyllenstrand, N., 2006. Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [Picea abies (L.) Karst]. Genetics 174, 2095–2105. https://doi.org/10.1534/genetics.106.065102

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965–1978. https://doi.org/10.1002/joc.1276

Hill, W.G., Mulder, H.A., 2010. Genetic analysis of environmental variation. Genet Res (Camb) 92, 381–395. https://doi.org/10.1017/S0016672310000546

Hunter, J.E., Schmidt, F.L., 2000. Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. International Journal of Selection and Assessment 8, 275–292. https://doi.org/10.1111/1468-2389.00156

Hyvonen, R., Agren, G.I., Linder, S., Persson, T., Cotrufo, M.F., Ekblad, A., Freeman, M., Grelle, A., Janssens, I.A., Jarvis, P.G., Kellomaki, S., Lindroth, A., Loustau, D., Lundmark, T., Norby, R.J., Oren, R., Pilegaard, K., Ryan, M.G., Sigurdsson, B.D., Stromgren, M., van Oijen, M., Wallin, G., 2007. The likely impact of elevated [CO2], nitrogen deposition, increased temperature and management on carbon sequestration in temperate and boreal forest ecosystems: a literature review. New Phytol 173, 463–480. https://doi.org/10.1111/j.1469-8137.2007.01967.x

Isabel, N., Holliday, J.A., Aitken, S.N., 2020. Forest genomics: advancing climate adaptation, forest health, productivity, and conservation. Evol Appl 13, 3–10. https://doi.org/10.1111/eva.12902

Izworska, K., Muter, E., Fleischer, P., Zielonka, T., 2022. Delay of growth release after a windthrow event and climate response in a light-demanding species (European larch Larix decidua Mill.). Trees - Structure and Function 36, 427–438. https://doi.org/10.1007/s00468-021-02218-4

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O.B., Bouwer, L.M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., Yiou, P., 2014. EURO-CORDEX: New high-resolution climate change projections for European impact research. Reg Environ Change 14, 563–578. https://doi.org/10.1007/s10113-013-0499-2

Jandl, R., Spathelf, P., Bolte, A., Prescott, C.E., 2019. Forest adaptation to climate change—is non-management an option? Ann For Sci 76. https://doi.org/10.1007/s13595-019-0827-x

Joyce, D.G., Rehfeldt, G.E., 2017. Management strategies for black spruce (Picea mariana (Mill.) B.S.P.) in the face of climate change: Climatic niche, clines, climatypes, and seed transfer. Forestry 90, 594–610. https://doi.org/10.1093/forestry/cpx018

Kaiser, H.F., 1958. The varimax criterion for analytic rotation in factor analysis.

Kalinowski, S.T., Taper, M.L., Marshall, T.C., 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16, 1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Kapeller, S., Dieckmann, U., Schueler, S., 2017. Varying selection differential throughout the climatic range of Norway spruce in Central Europe. Evol Appl 10, 25–38. https://doi.org/10.1111/eva.12413

Kapeller, S., Lexer, M.J., Geburek, T., Hiebl, J., Schueler, S., 2012. Intraspecific variation in climate response of Norway spruce in the eastern Alpine range: Selecting appropriate

provenances for future climate. For Ecol Manage. https://doi.org/10.1016/j.foreco.2012.01.039

Keeling, B.C.D., Bacastow, R.B., Bainbridge, A.E., Ekdahl, C.A., Guenther, P.R., Waterman, L.E.E.S., 1976. Atmospheric carbon dioxide variations at Mauna Loa Observatory , Hawaii.

Kelly, A.M., Smith, A.B., Eccleston, J.A., Cullis, B.R., 2007. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. Crop Sci 47, 1063–1070. https://doi.org/10.2135/cropsci2006.08.0540

Klein Tank, A.M.G., Können, G.P., 2003. Trends in indices of daily temperature and precipitation extremes in Europe, 1946-99. J Clim 16, 3665–3680. https://doi.org/10.1175/1520-0442(2003)016<3665:TIIODT>2.0.CO;2

Kolmodin, R., Strandberg, E., Madsen, P., Jensen, J., Jorjani, H., 2002. Genotype by environment Interaction nordic dairy cattle studied using reaction norms. Acta Agric Scand A Anim Sci 52, 11–24. https://doi.org/10.1080/09064700252806380

Koprowski, M., 2012. Long-term increase of March temperature has no negative impact on tree rings of European larch (Larix decidua) in lowland Poland. Trees 26, 1895–1903. https://doi.org/10.1007/s00468-012-0758-8

Kramer, K., Ducousso, A., Gömöry, D., Hansen, J.K., Ionita, L., Liesebach, M., Lorenţ, A., Schüler, S., Sulkowska, M., de Vries, S., von Wühlisch, G., 2017. Chilling and forcing requirements for foliage bud burst of European beech (Fagus sylvatica L.) differ between provenances and are phenotypically plastic. Agric For Meteorol 234–235, 172–181. https://doi.org/10.1016/j.agrformet.2016.12.002

Lashof, D.A., Deangelo, B.J., Resources, N., Council, D., Saleska, S.R., Harte, J., 1997. Terrestrial ecosystem feedbacks to global climate change. Defense 75–118. https://doi.org/10.1146/annurev.energy.22.1.75

Ledermann, T., 2010. Evaluating the performance of semi-distance-independent competition indices in predicting the basal area growth of individual trees. Canadian Journal of Forest Research 40, 2010. https://doi.org/10.1139/X10-026

Lefort, F., Douglas, Gerard.C., 1999. An efficient micro-method of DNA isolation from mature leaves of four hardwood tree species Acer, Fraxinus, Prunus and Quercus. Ann For Sci 56, 259–263. https://doi.org/10.1051/forest:19990308

Leites, L.P., Rehfeldt, G.E., Robinson, A.P., Crookston, N.L., Jaquish, B., 2012a. Possibilities and limitations of using historic provenance tests to infer forest species growth responses to climate change. Nat Resour Model. https://doi.org/10.1111/j.1939-7445.2012.00129.x

Leites, L.P., Robinson, A.P., Rehfeldt, G.E., Marshall, J.D., Crookston, N.L., 2012b. Height-growth response to climatic changes differs among populations of Douglas-fir: a novel analysis of historic data. Ecological Applications 22, 154–165.

Li, M.H., Yang, J., Kräuchi, N., 2003. Growth responses of Picea abies and Larix decidua to elevation in subalpine areas of Tyrol. Canadian Journal of Forest Research 33, 653–662. https://doi.org/10.1139/x02-202

Lstibůrek, M., Hodge, G.R., Lachout, P., 2015. Uncovering genetic information from commercial forest plantations — making up for lost time using "Breeding without Breeding." Tree Genet Genomes 11. https://doi.org/10.1007/s11295-015-0881-y

Lstibůrek, M., Klápště, J., Kobliha, J., El-Kassaby, Y.A., 2012. Breeding without breeding: effect of gene flow on fingerprinting effort. Tree Genet Genomes 8, 873–877. https://doi.org/10.1007/s11295-012-0472-0

Lstiburek, M., Schüler, S., El-Kassaby, Y., R. Hodge, G., Stejskal, J., Korecky, J., Skorpik, P., Konrad, H., Geburek, T., 2020. In situ genetic evaluation of European larch across climatic regions using marker-based pedigree reconstruction. Front Genet 11, 1–8. https://doi.org/10.3389/fgene.2020.00028

Lu, Y., Zhang, G., 2010. The equivalence between likelihood ratio test and F test for testing variance component in a balanced one way random effects model. Journal of Statistical Computation and Simulation 80, 443–450. https://doi.org/10.1080/00949650802695664

Luo, X., 2017. A comparison of three estimation methods in linear regression analysis, in: Advances in Computer Science Research. pp. 498–502. https://doi.org/10.2991/icmmita-16.2016.92

Lynch, M., Lande, R., 1993. Evolution and extinction in response to environmental change. Biotic Interactions and Global Change.

Malcolm, J.R., Markhan, A., Neilson, R.P., Garaci, M., 2002. Special paper: estimated migration rates under scenarios of global climate change. J Biogeogr 29, 835–849. https://doi.org/10.2307/827374

Malosetti, M., Ribaut, J.M., van Eeuwijk, F.A., 2013. The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. Front Physiol 4 MAR. https://doi.org/10.3389/fphys.2013.00044

Manabe, S., Wetherald, R.T., 1967. Thermal equilibrium of the atmosphere with a given distribution of relative humidity. J Atmos Sci 24, 241–259.

Mansoor, S., Farooq, I., Kachroo, M.M., Mahmoud, A.E.D., Fawzy, M., Popescu, S.M., Alyemeni, M.N., Sonne, C., Rinklebe, J., Ahmad, P., 2022. Elevation in wildfire frequencies with respect to the climate change. J Environ Manage. https://doi.org/10.1016/j.jenvman.2021.113769

Marshall, T.C., Slate, J., Kruuk, L.E.B., Pemberton, J.M., 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7, 639–655. https://doi.org/10.1046/j.1365-294x.1998.00374.x

Martínez-Vilalta, J., Piñol, J., 2002. Drought-induced mortality and hydraulic architecture in pine populations of the NE Iberian Peninsula. For Ecol Manage 161, 247–256. https://doi.org/10.1016/S0378-1127(01)00495-9

Masana, M.O., Baranyi, J., 2000. Adding new factors to predictive models: The effect on the risk of extrapolation. Food Microbiol 17, 367–374. https://doi.org/10.1006/fmic.1999.0326

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B., 2021. Climate change 2021: the physical science basis. Cambridge University Press, Cambridge, United Kingdom, and New York. https://doi.org/9781009157896

Matías, L., Gonzalez-díaz, P., Quero, J.L., Camarero, J.J., Lloret, F., Jump, A.S., 2016. Role of geographical provenance in the response of silver fir seedlings to experimental warming and drought. Tree Physiol 36, 1236–1246. https://doi.org/10.1093/treephys/tpw049

Matras, J., Paques, L., 2010. Larix decidua. EUFORGEN technical guidelines for genetic conservation and use 0–6.

Matyas, C., 2010. Forecasts needed for retreating forests. Nature 464, 1271–1271.

Matyas, C., 1996. Climatic adaptation of trees: rediscovering provenance tests. Euphytica 92, 45–54.

Matyas, C., 1994. Modeling climate change effects with provenance test data. Tree Physiol 14, 797–804.

Mátyás, C., 1994. Modelling climate change effects with provenance test data. Tree Physiol 14, 797–804. https://doi.org/10.1093/treephys/14.7-8-9.797

McKay, J.K., Christian, C.E., Harrison, S., Rice, K.J., 2005. "How local is local?" A review of practical and conceptual issues in the genetics of restoration. Restor Ecol 13, 432–440. https://doi.org/10.1111/j.1526-100X.2005.00058.x

McLachlan, J.S., Clark, J.S., 2004. Reconstructing historical ranges with fossil data at continental scales. For Ecol Manage 197, 139–147. https://doi.org/10.1016/j.foreco.2004.05.026

McLane, S.C., Lemay, V.M., Aitken, S.N., 2011. Modeling lodgepole pine radial growth relative to climate and genetics using universal growth-trend response functions. Ecological Applications 21, 776–788. https://doi.org/10.1890/10-0131.1

Melillo, J.M., Steudler, P.A., Aber, J.D., Newkirk, K., Lux, H., Bowles, F.P., Catricala, C., Magill, A., Ahrens, T., Morrisseau, S., 2002. Soil warming and carbon-cycle feedbacks to the climate system. Science (1979) 298, 2173–2176. https://doi.org/10.1126/science.1074153

Nardin, M., Musch, B., Rousselle, Y., Guérin, V., Sanchez, L., Rossi, J.P., Gerber, S., Marin, S., Pâques, L.E., Rozenberg, P., 2015. Genetic differentiation of European larch along an altitudinal gradient in the French Alps. Ann For Sci 72, 517–527. https://doi.org/10.1007/s13595-015-0483-8

Neale, D.B., Savolainen, O., 2004. Association genetics of complex traits in plants. Trends Plant Sci 9, 325–330. https://doi.org/10.1111/j.1469-8137.2010.03593.x

Obojes, N., Meurer, A.K., Newesely, C., Tasser, E., Oberhuber, W., Mayr, S., Tappeiner, U., 2022. Swiss stone pine growth benefits less from recent warming than European larch at a dry-inner alpine forest line as it reacts more sensitive to humidity. Agric For Meteorol 315. https://doi.org/10.1016/j.agrformet.2021.108788

Oliveira, I.C.M., Guilhen, J.H.S., Ribeiro, P.C. de O., Gezan, S.A., Schaffert, R.E., Simeone, M.L.F., Damasceno, C.M.B., Carneiro, J.E. de S., Carneiro, P.C.S., Parrella, R.A. da C., Pastina, M.M., 2020. Genotype-by-environment interaction and yield stability analysis of biomass sorghum hybrids using factor analytic models and environmental covariates. Field Crops Res 257, 1–10. https://doi.org/10.1016/j.fcr.2020.107929

O'neill, G.A., Hamann, A., Wang, T., 2008. Accounting for population variation improves estimates of the impact of climate change on species' growth and distribution Gregory. Journal of Applied Ecology 45, 1040–1049. https://doi.org/10.1111/j.1365-2664.2008.0

O'Neill, G.A., Nigh, G., Wang, T., Ott, P.K., 2007. Growth response functions improved by accounting for nonclimatic site effects. Canadian Journal of Forest Research 37, 2724–2730. https://doi.org/10.1139/X07-100

O'Neill, G.A., Stoehr, M., Jaquish, B., 2014. Quantifying safe seed transfer distance and impacts of tree breeding on adaptation. For Ecol Manage 328, 122–130. https://doi.org/10.1016/j.foreco.2014.05.039

Ørsted, M., Rohde, P.D., Hoffmann, A.A., Sørensen, P., Kristensen, T.N., 2018. Environmental variation partitioned into separate heritable components. Evolution (N Y) 72, 136–152. https://doi.org/10.1111/evo.13391

Palik, B.J., Clark, P.W., D'Amato, A.W., Swanston, C., Nagel, L., 2022. Operationalizing forest-assisted migration in the context of climate change adaptation: Examples from the eastern USA. Ecosphere 13. https://doi.org/10.1002/ecs2.4260

Pâques, L.E., 2013. Forest tree breeding in Europe. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-007-6146-9

Persson, B., 1998. Will climate change affect the optimal choice of Pinus sylvestris provenances? Silva Fennica 32, 121–128.

Petit, R.J., Jouzel, J., Raynaud, D., Barkov, N.I., Barnola, J.-M., 1999. Climate and atmospheric history of the past 420 , 000 years from the Vostok ice core , Antarctica. Nature 399, 329–340.

Poole, M.A., O'Farrell, P.N., 1971. The assumptions of the linear regression model. Transactions of the Institute of British Geographers 52, 145–158. https://doi.org/10.2307/621706

Poupon, V., Chakraborty, D., Stejskal, J., Konrad, H., Schueler, S., Lstibůrek, M., 2021. Accelerating adaptation of forest trees to climate change using individual tree response functions. Front Plant Sci 12, 1–7. https://doi.org/10.3389/fpls.2021.758221

Poupon, V., Gezan, S.A., Schueler, S., Lstibůrek, M., 2023. Genotype x environment interaction and climate sensitivity in growth and wood density of European larch. For Ecol Manage 545. https://doi.org/10.1016/j.foreco.2023.121259

Quaas, R.L., Anderson, R.D., Gilmour, A.R., 1984. BLUP School Handbook - Use of mixed models for prediction and for estimation of (co)variance components. University of New England.

R Core Team, 2020. R: A language and environment for statistical computing.

Ratcliffe, B., Hart, F.J., Kláppště, J., Jaquish, B., Mansfield, S.D., El-kassaby, Y.A., 2014. Genetics of wood quality attributes in Western Larch. Ann For Sci 71, 415–424. https://doi.org/10.1007/s13595-013-0349-x

Rebetez, M., Dobbertin, M., 2004. Climate change may already threaten Scots pine stands in the Swiss Alps. Theor Appl Climatol 79, 1–9. https://doi.org/10.1007/s00704-004-0058-3

Rehfeldt, G.E., Leites, L.P., Joyce, D.G., Weiskittel, A.R., 2018. Role of population genetics in guiding ecological responses to climate. Global change biology 24, 858–868. https://doi.org/10.1111/gcb.13883

Rehfeldt, G.E., Tchebakova, N.M., Parfenova, Y.I., Wykoff, W.R., Kuzmina, N.A., Milyutin, L.I., 2002. Intraspecific responses to climate in Pinus sylvestris. Glob Chang Biol. https://doi.org/10.1046/j.1365-2486.2002.00516.x

Rehfeldt, G.E., Wykoff, W.R., Ying, C.C., 2001. Changing climate on Pinus contorta. Clim Change 50, 355–376. https://doi.org/10.1023/A:1010614216256

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M.D., Seneviratne, S.I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D.C., Papale, D., Rammig, A., Smith, P., Thonicke,

K., Van Der Velde, M., Vicca, S., Walz, A., Wattenbach, M., 2013. Climate extremes and the carbon cycle.

Resco De Dios, V., Fischer, C., Colinas, C., 2007. Climate change effects on mediterranean forests and preventive measures. New For (Dordr) 33, 29–40. https://doi.org/10.1007/s11056-006-9011-x

Robakowski, P., Montpied, P., Dreyer, E., Robakowski, P., Montpied, P., Dreyer, E., 2005. Responses to temperature and shade in Abies alba seedlings from diverse provenances. Scand J For Res 20, 459–470. https://doi.org/10.1080/02827580500430408

Román-Palacios, C., Wiens, J.J., 2020. Recent responses to climate change reveal the drivers of species extinction and survival. PNAS 117, 4211–4217. https://doi.org/10.5061/dryad.4tmpg4f5w

Rouault, G., Candau, J.-N., Lieutier, F., Nageleisen, L.-M., Martin, J.-C., Warzée, N., 2006. Effects of drought and heat on forest insect populations in relation to the 2003 drought in Western Europe. Ann For Sci 63, 613–624. https://doi.org/10.1051/forest:2006044

Ruotsalainen, S., 2014. Increased forest production through forest tree breeding. Scand J For Res 29, 333–344. https://doi.org/10.1080/02827581.2014.926100

Rweyongeza, D.M., Yang, R.-C., Dhir, N.K., Barnhardt, L.K., Hansen, C., 2007. Genetic variation and climatic impacts on survival and growth of white spruce in Alberta, Canada. Silvae Genet 56, 117–127. https://doi.org/10.1515/sg-2007-0018

Sáenz-Romero, C., Lamy, J.B., Ducousso, A., Musch, B., Ehrenmann, F., Delzon, S., Cavers, S., Chałupka, W., Dağdaş, S., Hansen, J.K., Lee, S.J., Liesebach, M., Rau, H.M., Psomas, A., Schneck, V., Steiner, W., Zimmermann, N.E., Kremer, A., 2017. Adaptive and plastic responses of Quercus petraea populations to climate across Europe. Glob Chang Biol 23, 2831–2847. https://doi.org/10.1111/gcb.13576

Sáenz-Romero, C., O'neill, G., Aitken, S.N., Lindig-Cisneros, R., 2021. Assisted migration field tests in canada and mexico: Lessons, limitations, and challenges. Forests. https://doi.org/10.3390/f12010009

Sala, O.E., Iii, F.S.C., Armesto, J.J., Berlow, E., Dirzo, R., Huber-sanwald, E., Huenneke, L.F., Robert, B., Kinzig, A., Leemans, R., Lodge, D.M., Mooney, H.A., Oesterheld, Martín, Poff, N.L., Sykes, M.T., Walker, B.H., Walker, M., Wall, D.H., Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., Leemans, R., Lodge, D.M., Mooney, H.A., Oesterheld, Martin, Poff, N.L., Sykes, M.T., Walker, B.H., Walker, M., Wall, D.H., 2000. Global Biodiversity Scenarios for the Year 2100 Global Biodiversity Scenarios for the Year 2100. Science (1979) 287, 1770–1774. https://doi.org/10.1126/science.287.5459.1770

Salzmann, S., Wurzer, C., Golesch, G., 2023. Forest reproductive material. [WWW Document]. URL https://www.bundesamt-wald.at/en/forest-reproductive-material.html (accessed 8.28.23).

Saulnier, M., Corona, C., Stoffel, M., Guibal, F., Edouard, J.L., 2019. Climate-growth relationships in a Larix decidua Mill. network in the French Alps. Science of the Total Environment 664, 554–566. https://doi.org/10.1016/j.scitotenv.2019.01.404

Schmidtling, R.C., 1994. Use of provenance tests to predict response to climatic change: Loblolly pine and Norway spruce. Tree Physiol 14, 805–817. https://doi.org/10.1093/treephys/14.7-8-9.805

Schueler, S., 2011. Herkunft lärche - eine kardinalfrage. [WWW Document]. URL https://www.waldwissen.net/de/waldwirtschaft/waldbau/waldgenetik/herkunft-laerche (accessed 8.28.23).

Schueler, S., George, J.P., Karanitsch-Ackerl, S., Mayer, K., Klumpp, R.T., Grabner, M., 2021. Evolvability of drought response in four native and non-native conifers: opportunities for forest and genetic resource management in Europe. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.648312

Schueler, S., Kapeller, S., 2010. Klima-response von fichtenherkünften im Alpenraum. Eine adaptionsmöglichkeit für die österreichische forstwirtschaft.

Schueler, S., Kapeller, S., Konrad, H., Geburek, T., Mengl, M., Bozzano, M., Koskela, J., Lefèvre, F., Hubert, J., Kraigher, H., Longauer, R., Olrik, D.C., 2013. Adaptive genetic diversity of trees for forest conservation in a future climate: A case study on Norway spruce in Austria. Biodivers Conserv 22, 1151–1166. https://doi.org/10.1007/s10531-012-0313-3

Schueler, S., Liesebach, M., 2014. Latitudinal population transfer reduces temperature sum requirements for bud burst of European beech. Plant Ecol 216, 111–122. https://doi.org/10.1007/s11258-014-0420-1

Škorpík, P., Konrad, H., Geburek, T., Schuh, M., Vasold, D., Eberhardt, M., Schueler, S., 2018. Solid wood properties assessed by non-destructive measurements of standing European larch (Larix decidua Mill.): Environmental effects on variation within and among trees and forest stands. Forests 9. https://doi.org/10.3390/f9050276

Smith, A., Cullis, B., Gilmour, A., 2001a. The analysis of crop variety evaluation data in australia. Aust N Z J Stat 43, 129–145. https://doi.org/10.1111/1467-842X.00163

Smith, A., Cullis, B., Thompson, R., 2001b. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57, 1138–1147.

Smith, A.B., Cullis, B.R., 2018. Plant breeding selection tools built on factor analytic mixed models for multi-environment trial data. Euphytica 214. https://doi.org/10.1007/s10681-018-2220-5

St Clair, J.B., Howe, G.T., 2007. Genetic maladaptation of coastal Douglas-fir seedlings to future climates. Glob Chang Biol 13, 1441–1454. https://doi.org/10.1111/j.1365-2486.2007.01385.x

Suvanto, S., Nöjd, P., Henttonen, H.M., Beuker, E., Mäkinen, H., 2016. Geographical patterns in the radial growth response of Norway spruce provenances to climatic variation. Agric For Meteorol 222, 10–20. https://doi.org/10.1016/j.agrformet.2016.03.003

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

Swidrak, I., Schuster, R., Oberhuber, W., 2013. Comparing growth phenology of co-occurring deciduous and evergreen conifers exposed to drought. Flora: Morphology, Distribution, Functional Ecology of Plants 208, 609–617. https://doi.org/10.1016/j.flora.2013.09.004

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F., De Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., Van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L., Williams, S.E., 2004. Extinction risk from climate change. Nature 427, 145–148. https://doi.org/10.1038/nature02121

Thompson, R., Cullis, B., Smith, A., Gilmour, A., 2003. A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. Aust N Z J Stat 45, 445–459. https://doi.org/10.1111/1467-842X.00297

Thomson, A.M., Parker, W.H., 2008. Boreal forest provenance tests used to predict optimal growth and response to climate change. 1. Jack pine. Canadian Journal of Forest Research 38, 157–170. https://doi.org/10.1139/X07-122

Topaloğlu, E., Ay, N., Altun, L., Serdar, B., 2016. Effect of altitude and aspect on various wood properties of Oriental beech (Fagus orientalis Lipsky) wood. Turkish journal of agriculture and forestry 40, 397–406. https://doi.org/10.3906/tar-1508-95

Trenberth, K.E., Dai, A., Van Der Schrier, G., Jones, P.D., Barichivich, J., Briffa, K.R., Sheffield, J., 2014. Global warming and changes in drought. Nat Clim Chang 4, 17–22. https://doi.org/10.1038/nclimate2067

Tzedakis, P.C., Emerson, B.C., Hewitt, G.M., 2013. Cryptic or mystic? Glacial tree refugia in northern Europe. Trends Ecol Evol 28, 696–704. https://doi.org/10.1016/j.tree.2013.09.001

Vacek, Z., Bílek, L., Remeš, J., Vacek, S., Cukor, J., Gallo, J., Šimůnek, V., Bulušek, D., Brichta, J., Vacek, O., Drábek, O., Zahradník, D., 2022. Afforestation suitability and production potential of five tree species on abandoned farmland in response to climate change, Czech Republic. Trees. https://doi.org/10.1007/s00468-022-02295-z

Visscher, P.M., Hill, W.G., Wray, N.R., 2008. Heritability in the genomics era - Concepts and misconceptions. Nat Rev Genet. https://doi.org/10.1038/nrg2322

Vitasse, Y., Delzon, S., Dufre, E., 2009. Leaf phenology sensitivity to temperature in European trees : Do within-species populations exhibit similar responses ? Agric For Meteorol 149, 735–744. https://doi.org/10.1016/j.agrformet.2008.10.019

Wagner, S., Gerber, S., Petit, R.J., 2012. Two highly informative dinucleotide SSR multiplexes for the conifer Larix decidua (European larch). Mol Ecol Resour 12, 717–725. https://doi.org/10.1111/j.1755-0998.2012.03139.x

Wagner, S., Litt, T., Sánchez-Goñi, M.-F., Petit, R.J., 2015. History of Larix decidua Mill. (European larch) since 130 ka. Quat Sci Rev 124, 224–247. https://doi.org/10.1016/j.quascirev.2015.07.002

Walther, G.R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.M., Hoegh-Guldberg, O., Bairlein, F., 2002. Ecological responses to recent climate change. Nature 416, 389–395. https://doi.org/10.1038/416389a

Wang, T., Hamann, A., Spittlehouse, D.L., Aitken, S.N., 2006a. Development of scale-free climate data for western Canada for use in resource management. International Journal of Climatology 26, 383–397. https://doi.org/10.1002/joc.1247

Wang, T., Hamann, A., Yanchuk, A., O'Neill, G.A., Aitken, S.N., 2006b. Use of response functions in selecting lodgepole pine populations for future climates. Glob Chang Biol 12, 2404–2416. https://doi.org/10.1111/j.1365-2486.2006.01271.x

Wang, T., O 'Neill, G.A., Aitken, S.N., 2010. Integrating environmental and genetic effects to predict responses of tree populations to climate. Ecological Applications 20, 153–163.

Westerling, A.L., Hidalgo, H.G., Cayan, D.R., Swetnam, T.W., 2006. Warming and earlier spring increase Western U.S. forest wildfire activity. Science (1979) 313, 940–943. https://doi.org/10.1126/science.1128834

Wheeler, N.C., Steiner, K.C., Schlarbaum, S.E., Neale, D.B., 2015. The evolution of forest genetics and tree improvement research in the United States. J For 113, 500–510. https://doi.org/10.5849/jof.14-120

White, T.L., Adams, W.T., Neale, D.B., 2007. Forest genetics. Cabi, Wallingford, Oxfordshire, UK. https://doi.org/10.1079/ 9781845932855.0000

Wielinga, B., Raymond, C.A., James, R., Matheson, A.C., 2009. Genetic parameters and genotype by environment interactions for green and basic density and stiffness of Pinus radiata D. Don estimated using acoustics. Silvae Genet 58, 112–122. https://doi.org/10.1515/sg-2009-0015

Willis, K.J., Van Andel, T.H., 2004. Trees or no trees? The environments of central and eastern Europe during the last glaciation. Quat Sci Rev 23, 2369–2387. https://doi.org/10.1016/j.quascirev.2004.06.002

Wu, H.X., Ying, C.C., 2004. Geographic pattern of local optimality in natural populations of lodgepole pine. For Ecol Manage 194, 177–198. https://doi.org/10.1016/j.foreco.2004.02.017

Zeidler, A., Vacek, Z., Cukor, J., Boruvka, V., Vacek, S., Prokupkova, A., Linda, R., Vacek, O., 2022. Is European larch (Larix decidua Mill.) a suitable substitute for Norway spruce

(Picea abies (L.) Karst.) for agricultural land afforestation? For Ecol Manage 517, 1–14. https://doi.org/10.1016/j.foreco.2022.120257

Zhang, H., Zhang, S., Chen, S., Xia, D., Yang, C., Zhao, X., 2022. Genetic variation and superior provenances selection for wood properties of Larix olgensis at four trials. J For Res (Harbin) 33, 1867–1879. https://doi.org/10.1007/s11676-021-01449-y

Zhang, R., Han, D., Hu, X., 2020. Analyzing the performance of corn in China using a factor-analytic variance-covariance structure with multiple factors. Crop Sci 60, 190–201. https://doi.org/10.1002/csc2.20090

Zobel, B., Talbert, J., 1984. Applied forest tree improvement. John Wiley & Sons, New York.

# 9 Supplementary tables

**Table S 1** Pairwise correlations between each pair of sites. *Diff* represents the difference of pMAI-H between the sites, *SD* is the standard error, and *P* is the *p* value of the correlation.

| Site 1 | Site 2 | Diff | SD | P |
|---|---|---|---|---|
| A | B1 | 13.01 | 1.16 | 0.000 |
| A | B11 | 9.64 | 1.22 | 0.000 |
| A | B12 | 16.30 | 1.21 | 0.000 |
| A | B13 | 21.18 | 1.20 | 0.000 |
| A | B16 | 35.81 | 1.20 | 0.000 |
| A | B18 | 14.33 | 1.22 | 0.000 |
| A | B2 | 6.27 | 1.26 | 0.000 |
| A | B20 | 19.13 | 1.16 | 0.000 |
| A | B3 | 22.98 | 1.22 | 0.000 |
| A | B4 | 6.44 | 1.19 | 0.000 |
| A | B5 | 19.38 | 1.33 | 0.000 |
| A | B6 | 3.82 | 1.18 | 0.001 |
| A | B7 | 17.88 | 1.24 | 0.000 |
| A | B9 | 5.05 | 1.25 | 0.000 |
| A | H2 | 22.16 | 1.23 | 0.000 |
| A | N | 5.30 | 0.98 | 0.000 |
| A | T1 | 18.67 | 1.18 | 0.000 |
| A | T2 | 12.25 | 1.41 | 0.000 |
| A | W3 | -4.30 | 1.34 | 0.001 |
| A | W4 | 3.97 | 1.19 | 0.000 |
| B1 | B11 | -3.37 | 1.18 | 0.002 |
| B1 | B12 | 3.29 | 1.18 | 0.003 |
| B1 | B13 | 8.18 | 1.18 | 0.000 |
| B1 | B16 | 22.81 | 1.16 | 0.000 |
| B1 | B18 | 1.32 | 1.19 | 0.134 |
| B1 | B2 | -6.74 | 1.22 | 0.000 |
| B1 | B20 | 6.13 | 1.12 | 0.000 |
| B1 | B3 | 9.98 | 1.19 | 0.000 |
| B1 | B4 | -6.57 | 1.16 | 0.000 |
| B1 | B5 | 6.37 | 1.30 | 0.000 |
| B1 | B6 | -9.19 | 1.15 | 0.000 |
| B1 | B7 | 4.88 | 1.21 | 0.000 |
| B1 | B9 | -7.96 | 1.22 | 0.000 |
| B1 | H2 | 9.15 | 1.20 | 0.000 |
| B1 | N | -7.71 | 0.94 | 0.000 |

| | | | | |
|-----|-----|--------|------|-------|
| B1  | T1  | 5.67   | 1.15 | 0.000 |
| B1  | T2  | -0.76  | 1.38 | 0.292 |
| B1  | W3  | -17.30 | 1.30 | 0.000 |
| B1  | W4  | -9.04  | 1.16 | 0.000 |
| B11 | B12 | 6.66   | 1.23 | 0.000 |
| B11 | B13 | 11.55  | 1.22 | 0.000 |
| B11 | B16 | 26.17  | 1.21 | 0.000 |
| B11 | B18 | 4.69   | 1.23 | 0.000 |
| B11 | B2  | -3.37  | 1.28 | 0.004 |
| B11 | B20 | 9.49   | 1.18 | 0.000 |
| B11 | B3  | 13.34  | 1.23 | 0.000 |
| B11 | B4  | -3.20  | 1.20 | 0.004 |
| B11 | B5  | 9.74   | 1.34 | 0.000 |
| B11 | B6  | -5.82  | 1.21 | 0.000 |
| B11 | B7  | 8.24   | 1.26 | 0.000 |
| B11 | B9  | -4.59  | 1.26 | 0.000 |
| B11 | H2  | 12.52  | 1.24 | 0.000 |
| B11 | N   | -4.34  | 1.01 | 0.000 |
| B11 | T1  | 9.03   | 1.20 | 0.000 |
| B11 | T2  | 2.61   | 1.44 | 0.035 |
| B11 | W3  | -13.94 | 1.35 | 0.000 |
| B11 | W4  | -5.67  | 1.21 | 0.000 |
| B12 | B13 | 4.88   | 1.20 | 0.000 |
| B12 | B16 | 19.51  | 1.21 | 0.000 |
| B12 | B18 | -1.97  | 1.23 | 0.055 |
| B12 | B2  | -10.03 | 1.27 | 0.000 |
| B12 | B20 | 2.83   | 1.16 | 0.007 |
| B12 | B3  | 6.68   | 1.24 | 0.000 |
| B12 | B4  | -9.86  | 1.20 | 0.000 |
| B12 | B5  | 3.08   | 1.34 | 0.011 |
| B12 | B6  | -12.48 | 1.19 | 0.000 |
| B12 | B7  | 1.58   | 1.25 | 0.102 |
| B12 | B9  | -11.25 | 1.26 | 0.000 |
| B12 | H2  | 5.86   | 1.24 | 0.000 |
| B12 | N   | -11.00 | 1.00 | 0.000 |
| B12 | T1  | 2.37   | 1.19 | 0.024 |
| B12 | T2  | -4.05  | 1.42 | 0.002 |
| B12 | W3  | -20.60 | 1.35 | 0.000 |
| B12 | W4  | -12.33 | 1.21 | 0.000 |
| B13 | B16 | 14.63  | 1.20 | 0.000 |
| B13 | B18 | -6.86  | 1.22 | 0.000 |
| B13 | B2  | -14.91 | 1.25 | 0.000 |
| B13 | B20 | -2.05  | 1.16 | 0.038 |
| B13 | B3  | 1.80   | 1.22 | 0.071 |

| | | | | |
|---|---|---|---|---|
| B13 | B4 | -14.74 | 1.19 | 0.000 |
| B13 | B5 | -1.81 | 1.34 | 0.088 |
| B13 | B6 | -17.36 | 1.18 | 0.000 |
| B13 | B7 | -3.30 | 1.24 | 0.004 |
| B13 | B9 | -16.14 | 1.25 | 0.000 |
| B13 | H2 | 0.97 | 1.23 | 0.214 |
| B13 | N | -15.89 | 0.99 | 0.000 |
| B13 | T1 | -2.51 | 1.18 | 0.017 |
| B13 | T2 | -8.94 | 1.42 | 0.000 |
| B13 | W3 | -25.48 | 1.35 | 0.000 |
| B13 | W4 | -17.22 | 1.20 | 0.000 |
| B16 | B18 | -21.49 | 1.22 | 0.000 |
| B16 | B2 | -29.54 | 1.25 | 0.000 |
| B16 | B20 | -16.68 | 1.15 | 0.000 |
| B16 | B3 | -12.83 | 1.22 | 0.000 |
| B16 | B4 | -29.37 | 1.18 | 0.000 |
| B16 | B5 | -16.44 | 1.32 | 0.000 |
| B16 | B6 | -31.99 | 1.19 | 0.000 |
| B16 | B7 | -17.93 | 1.24 | 0.000 |
| B16 | B9 | -30.76 | 1.25 | 0.000 |
| B16 | H2 | -13.66 | 1.22 | 0.000 |
| B16 | N | -30.52 | 0.98 | 0.000 |
| B16 | T1 | -17.14 | 1.18 | 0.000 |
| B16 | T2 | -23.57 | 1.41 | 0.000 |
| B16 | W3 | -40.11 | 1.33 | 0.000 |
| B16 | W4 | -31.85 | 1.19 | 0.000 |
| B18 | B2 | -8.06 | 1.29 | 0.000 |
| B18 | B20 | 4.81 | 1.18 | 0.000 |
| B18 | B3 | 8.66 | 1.24 | 0.000 |
| B18 | B4 | -7.89 | 1.20 | 0.000 |
| B18 | B5 | 5.05 | 1.35 | 0.000 |
| B18 | B6 | -10.51 | 1.21 | 0.000 |
| B18 | B7 | 3.56 | 1.26 | 0.002 |
| B18 | B9 | -9.28 | 1.27 | 0.000 |
| B18 | H2 | 7.83 | 1.25 | 0.000 |
| B18 | N | -9.03 | 1.02 | 0.000 |
| B18 | T1 | 4.35 | 1.21 | 0.000 |
| B18 | T2 | -2.08 | 1.44 | 0.074 |
| B18 | W3 | -18.62 | 1.36 | 0.000 |
| B18 | W4 | -10.36 | 1.21 | 0.000 |
| B2 | B20 | 12.86 | 1.22 | 0.000 |
| B2 | B3 | 16.71 | 1.29 | 0.000 |
| B2 | B4 | 0.17 | 1.25 | 0.446 |
| B2 | B5 | 13.11 | 1.39 | 0.000 |

| | | | | |
|-----|-----|--------|------|-------|
| B2 | B6 | -2.45 | 1.24 | 0.024 |
| B2 | B7 | 11.61 | 1.30 | 0.000 |
| B2 | B9 | -1.22 | 1.31 | 0.176 |
| B2 | H2 | 15.89 | 1.29 | 0.000 |
| B2 | N | -0.97 | 1.05 | 0.178 |
| B2 | T1 | 12.40 | 1.25 | 0.000 |
| B2 | T2 | 5.98 | 1.46 | 0.000 |
| B2 | W3 | -10.57 | 1.39 | 0.000 |
| B2 | W4 | -2.30 | 1.27 | 0.035 |
| B20 | B3 | 3.85 | 1.18 | 0.001 |
| B20 | B4 | -12.69 | 1.15 | 0.000 |
| B20 | B5 | 0.25 | 1.30 | 0.425 |
| B20 | B6 | -15.31 | 1.14 | 0.000 |
| B20 | B7 | -1.25 | 1.20 | 0.149 |
| B20 | B9 | -14.08 | 1.22 | 0.000 |
| B20 | H2 | 3.03 | 1.19 | 0.005 |
| B20 | N | -13.84 | 0.94 | 0.000 |
| B20 | T1 | -0.46 | 1.14 | 0.344 |
| B20 | T2 | -6.88 | 1.39 | 0.000 |
| B20 | W3 | -23.43 | 1.31 | 0.000 |
| B20 | W4 | -15.16 | 1.16 | 0.000 |
| B3 | B4 | -16.54 | 1.20 | 0.000 |
| B3 | B5 | -3.60 | 1.35 | 0.004 |
| B3 | B6 | -19.16 | 1.21 | 0.000 |
| B3 | B7 | -5.10 | 1.26 | 0.000 |
| B3 | B9 | -17.93 | 1.27 | 0.000 |
| B3 | H2 | -0.82 | 1.25 | 0.254 |
| B3 | N | -17.69 | 1.02 | 0.000 |
| B3 | T1 | -4.31 | 1.21 | 0.000 |
| B3 | T2 | -10.73 | 1.44 | 0.000 |
| B3 | W3 | -27.28 | 1.36 | 0.000 |
| B3 | W4 | -19.01 | 1.21 | 0.000 |
| B4 | B5 | 12.94 | 1.32 | 0.000 |
| B4 | B6 | -2.62 | 1.18 | 0.013 |
| B4 | B7 | 11.44 | 1.22 | 0.000 |
| B4 | B9 | -1.39 | 1.23 | 0.129 |
| B4 | H2 | 15.72 | 1.21 | 0.000 |
| B4 | N | -1.14 | 0.97 | 0.120 |
| B4 | T1 | 12.23 | 1.17 | 0.000 |
| B4 | T2 | 5.81 | 1.40 | 0.000 |
| B4 | W3 | -10.74 | 1.32 | 0.000 |
| B4 | W4 | -2.47 | 1.18 | 0.018 |
| B5 | B6 | -15.56 | 1.32 | 0.000 |
| B5 | B7 | -1.49 | 1.37 | 0.138 |

| | | | | |
|---|---|---|---|---|
| B5 | B9 | -14.33 | 1.38 | 0.000 |
| B5 | H2 | 2.78 | 1.36 | 0.020 |
| B5 | N | -14.08 | 1.15 | 0.000 |
| B5 | T1 | -0.71 | 1.32 | 0.296 |
| B5 | T2 | -7.13 | 1.53 | 0.000 |
| B5 | W3 | -23.67 | 1.45 | 0.000 |
| B5 | W4 | -15.41 | 1.32 | 0.000 |
| B6 | B7 | 14.06 | 1.22 | 0.000 |
| B6 | B9 | 1.23 | 1.24 | 0.161 |
| B6 | H2 | 18.34 | 1.21 | 0.000 |
| B6 | N | 1.48 | 0.97 | 0.063 |
| B6 | T1 | 14.85 | 1.17 | 0.000 |
| B6 | T2 | 8.43 | 1.41 | 0.000 |
| B6 | W3 | -8.12 | 1.33 | 0.000 |
| B6 | W4 | 0.15 | 1.18 | 0.450 |
| B7 | B9 | -12.83 | 1.29 | 0.000 |
| B7 | H2 | 4.27 | 1.27 | 0.000 |
| B7 | N | -12.59 | 1.04 | 0.000 |
| B7 | T1 | 0.79 | 1.23 | 0.261 |
| B7 | T2 | -5.64 | 1.45 | 0.000 |
| B7 | W3 | -22.18 | 1.38 | 0.000 |
| B7 | W4 | -13.92 | 1.23 | 0.000 |
| B9 | H2 | 17.11 | 1.28 | 0.000 |
| B9 | N | 0.25 | 1.05 | 0.406 |
| B9 | T1 | 13.62 | 1.24 | 0.000 |
| B9 | T2 | 7.20 | 1.46 | 0.000 |
| B9 | W3 | -9.35 | 1.38 | 0.000 |
| B9 | W4 | -1.08 | 1.25 | 0.194 |
| H2 | N | -16.86 | 1.02 | 0.000 |
| H2 | T1 | -3.49 | 1.21 | 0.002 |
| H2 | T2 | -9.91 | 1.44 | 0.000 |
| H2 | W3 | -26.45 | 1.36 | 0.000 |
| H2 | W4 | -18.19 | 1.23 | 0.000 |
| N | T1 | 13.38 | 0.97 | 0.000 |
| N | T2 | 6.95 | 1.24 | 0.000 |
| N | W3 | -9.59 | 1.15 | 0.000 |
| N | W4 | -1.33 | 0.98 | 0.089 |
| T1 | T2 | -6.42 | 1.40 | 0.000 |
| T1 | W3 | -22.97 | 1.33 | 0.000 |
| T1 | W4 | -14.70 | 1.18 | 0.000 |
| T2 | W3 | -16.54 | 1.53 | 0.000 |
| T2 | W4 | -8.28 | 1.41 | 0.000 |
| W3 | W4 | 8.27 | 1.33 | 0.000 |

**Table S 2** Additive site-to-site genetic correlations generated by the FA2-MAI-H model. The values in bold are the maximum and minimum correlations.

| | A | B1 | B11 | B12 | B13 | B16 | B18 | B2 | B20 | B3 | B4 | B5 | B6 | B7 | B9 | H2 | N | T1 | T2 | W3 | W4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **A** | A | | | | | | | | | | | | | | | | | | | | |
| **B1** | 0.97 | B1 | | | | | | | | | | | | | | | | | | | |
| **B11** | 0.97 | 0.95 | B11 | | | | | | | | | | | | | | | | | | |
| **B12** | 0.89 | 0.87 | 0.76 | B12 | | | | | | | | | | | | | | | | | |
| **B13** | 0.98 | 0.95 | 1.00 | 0.79 | B13 | | | | | | | | | | | | | | | | |
| **B16** | 0.36 | 0.35 | 0.42 | 0.19 | 0.41 | B16 | | | | | | | | | | | | | | | |
| **B18** | 0.63 | 0.62 | 0.48 | 0.82 | 0.52 | 0.06 | B18 | | | | | | | | | | | | | | |
| **B2** | 0.98 | 0.95 | 1.00 | 0.77 | 1.00 | 0.42 | 0.49 | B2 | | | | | | | | | | | | | |
| **B20** | 0.85 | 0.83 | 0.71 | 1.00 | 0.74 | 0.16 | 0.84 | 0.72 | B20 | | | | | | | | | | | | |
| **B3** | 0.91 | 0.89 | 0.80 | 1.00 | 0.82 | 0.21 | 0.81 | 0.80 | 0.99 | B3 | | | | | | | | | | | |
| **B4** | **0.98** | 0.95 | 1.00 | 0.79 | 1.00 | 0.41 | 0.51 | 1.00 | 0.74 | 0.82 | B4 | | | | | | | | | | |
| **B5** | 0.08 | 0.08 | 0.03 | 0.17 | 0.04 | -0.03 | 0.17 | 0.03 | 0.18 | 0.16 | 0.04 | B5 | | | | | | | | | |
| **B6** | 0.76 | 0.74 | 0.75 | 0.66 | 0.75 | 0.28 | 0.46 | 0.75 | 0.63 | 0.68 | 0.75 | 0.05 | B6 | | | | | | | | |
| **B7** | 0.79 | 0.77 | 0.68 | 0.87 | 0.71 | 0.18 | 0.72 | 0.69 | 0.87 | 0.87 | 0.70 | 0.14 | 0.58 | B7 | | | | | | | |
| **B9** | 0.33 | 0.32 | 0.39 | 0.17 | 0.38 | 0.20 | 0.05 | 0.38 | 0.14 | 0.19 | 0.38 | -0.03 | 0.26 | 0.16 | B9 | | | | | | |
| **H2** | 0.97 | 0.94 | 0.89 | 0.97 | 0.91 | 0.28 | 0.75 | 0.89 | 0.95 | 0.99 | 0.90 | 0.13 | 0.73 | 0.86 | 0.25 | H2 | | | | | |
| **N** | 0.02 | 0.02 | 0.25 | -0.43 | 0.21 | 0.30 | **-0.56** | 0.24 | -0.50 | -0.38 | 0.22 | -0.21 | 0.05 | -0.36 | 0.29 | -0.22 | N | | | | |
| **T1** | 0.76 | 0.74 | 0.75 | 0.65 | 0.76 | 0.29 | 0.44 | 0.75 | 0.62 | 0.67 | 0.76 | 0.05 | 0.58 | 0.58 | 0.27 | 0.72 | 0.08 | T1 | | | |
| **T2** | 0.37 | 0.36 | 0.48 | 0.10 | 0.46 | 0.28 | -0.05 | 0.47 | 0.05 | 0.13 | 0.46 | -0.07 | 0.30 | 0.10 | 0.26 | 0.24 | 0.51 | 0.31 | T2 | | |
| **W3** | 0.83 | 0.81 | 0.92 | 0.53 | 0.91 | 0.44 | 0.26 | 0.92 | 0.47 | 0.57 | 0.91 | -0.03 | 0.65 | 0.48 | 0.41 | 0.70 | 0.49 | 0.66 | 0.54 | W3 | |
| **W4** | 0.91 | 0.89 | 0.87 | 0.86 | 0.88 | 0.30 | 0.64 | 0.87 | 0.84 | 0.88 | 0.88 | 0.10 | 0.69 | 0.76 | 0.27 | 0.91 | -0.08 | 0.69 | 0.28 | 0.71 | W4 |

**Table S 3** Additive site-to-site genetic correlations generated by the FA2-MAI-DBH model. The values in bold are the maximum and minimum correlations.

| | A | B1 | B11 | B12 | B13 | B16 | B18 | B2 | B20 | B3 | B4 | B5 | B6 | B7 | B9 | H2 | N | T1 | T2 | W3 | W4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 1.00 | 0.51 | 0.72 | 0.13 | 0.54 | 0.49 | 0.69 | 0.43 | -0.06 | 0.74 | 0.51 | 0.27 | -0.16 | 0.74 | -0.16 | -0.19 | 0.01 | 0.71 | 0.36 | 0.47 | 0.19 |
| B1 | | 1.00 | 0.48 | 0.83 | **0.99** | 0.06 | 0.92 | **0.99** | 0.50 | 0.61 | -0.05 | -0.06 | -0.31 | 0.73 | 0.13 | 0.53 | -0.34 | 0.87 | 0.97 | 0.53 | 0.15 |
| B11 | | | 1.00 | -0.09 | 0.53 | 0.77 | 0.79 | 0.35 | -0.27 | **0.99** | 0.85 | 0.46 | -0.14 | 0.95 | -0.30 | -0.49 | 0.14 | 0.85 | 0.24 | 0.58 | 0.26 |
| B12 | | | | 1.00 | 0.80 | -0.42 | 0.54 | 0.90 | 0.74 | 0.07 | -0.60 | -0.36 | -0.26 | 0.22 | 0.34 | 0.91 | -0.48 | 0.46 | 0.94 | 0.24 | 0.01 |
| B13 | | | | | 1.00 | 0.11 | 0.94 | 0.98 | 0.47 | 0.66 | 0.00 | -0.03 | -0.31 | 0.76 | 0.11 | 0.48 | -0.32 | 0.90 | 0.95 | 0.55 | 0.16 |
| B16 | | | | | | 1.00 | 0.40 | -0.06 | -0.47 | 0.71 | 0.85 | 0.47 | -0.01 | 0.63 | -0.35 | -0.69 | 0.28 | 0.46 | -0.16 | 0.34 | 0.19 |
| B18 | | | | | | | 1.00 | 0.85 | 0.22 | 0.88 | 0.36 | 0.16 | -0.28 | 0.94 | -0.04 | 0.14 | -0.17 | **0.99** | 0.78 | 0.64 | 0.22 |
| B2 | | | | | | | | 1.00 | 0.58 | 0.49 | -0.19 | -0.14 | -0.31 | 0.62 | 0.19 | 0.64 | -0.39 | 0.80 | **0.99** | 0.48 | 0.12 |
| B20 | | | | | | | | | 1.00 | -0.16 | -0.61 | -0.36 | -0.16 | -0.04 | 0.31 | 0.76 | -0.38 | 0.15 | 0.63 | 0.05 | -0.05 |
| B3 | | | | | | | | | | 1.00 | 0.76 | 0.40 | -0.18 | **0.99** | -0.25 | -0.35 | 0.07 | 0.92 | 0.39 | 0.62 | 0.26 |
| B4 | | | | | | | | | | | 1.00 | 0.56 | 0.02 | 0.65 | -0.42 | **-0.87** | 0.37 | 0.44 | -0.30 | 0.34 | 0.20 |
| B5 | | | | | | | | | | | | 1.00 | 0.03 | 0.34 | -0.24 | -0.51 | 0.22 | 0.21 | -0.20 | 0.17 | 0.11 |
| B6 | | | | | | | | | | | | | 1.00 | -0.22 | -0.05 | -0.17 | 0.11 | -0.27 | -0.30 | -0.16 | -0.05 |
| B7 | | | | | | | | | | | | | | 1.00 | -0.19 | -0.20 | -0.01 | 0.97 | 0.53 | 0.64 | 0.25 |
| B9 | | | | | | | | | | | | | | | 1.00 | 0.42 | -0.19 | -0.08 | 0.23 | -0.08 | -0.07 |
| H2 | | | | | | | | | | | | | | | | 1.00 | -0.48 | 0.05 | 0.73 | -0.03 | -0.10 |
| N | | | | | | | | | | | | | | | | | 1.00 | -0.13 | -0.42 | -0.05 | 0.02 |
| T1 | | | | | | | | | | | | | | | | | | 1.00 | 0.72 | 0.64 | 0.23 |
| T2 | | | | | | | | | | | | | | | | | | | 1.00 | 0.42 | 0.09 |
| W3 | | | | | | | | | | | | | | | | | | | | 1.00 | 0.16 |
| W4 | | | | | | | | | | | | | | | | | | | | | 1.00 |

156

**Table S 4** Additive site-to-site genetic correlations generated by the FA2-PP model. The values in bold are the maximum and minimum correlations.

| | B1 | B11 | B12 | B13 | B16 | B18 | B2 | B20 | B3 | B4 | B5 | B6 | B7 | B9 | H2 | N | T1 | T2 | W3 | W4 |
|-----|----|-----|-----|-----|-----|-----|----|-----|----|----|----|----|----|----|----|----|----|----|----|----|
| A | 0.81 | 0.75 | 0.76 | 0.79 | 0.59 | 0.54 | 0.79 | 0.36 | 0.39 | 0.56 | 0.81 | 0.17 | 0.77 | 0.19 | 0.14 | 0.44 | 0.64 | 0.67 | 0.54 | 0.80 |
| B1 | | 0.92 | 0.92 | 0.97 | 0.73 | 0.66 | 0.97 | 0.44 | 0.48 | 0.70 | **0.99** | 0.22 | 0.94 | 0.24 | 0.18 | 0.55 | 0.78 | 0.82 | 0.66 | 0.98 |
| B11 | | | 0.70 | **0.99** | 0.94 | 0.82 | 0.79 | 0.53 | 0.70 | 0.89 | 0.91 | 0.59 | 0.73 | 0.54 | 0.15 | 0.83 | 0.61 | 0.64 | 0.90 | 0.98 |
| B12 | | | | 0.81 | 0.41 | 0.41 | **0.99** | 0.29 | 0.20 | 0.40 | 0.94 | **-0.17** | **0.99** | -0.08 | 0.18 | 0.19 | 0.83 | 0.86 | 0.33 | 0.83 |
| B13 | | | | | 0.87 | 0.77 | 0.88 | 0.50 | 0.62 | 0.83 | 0.96 | 0.45 | 0.83 | 0.43 | 0.16 | 0.73 | 0.70 | 0.73 | 0.82 | **0.99** |
| B16 | | | | | | 0.84 | 0.53 | 0.53 | 0.79 | 0.94 | 0.70 | 0.83 | 0.45 | 0.73 | 0.11 | 0.97 | 0.38 | 0.41 | **0.99** | 0.85 |
| B18 | | | | | | | 0.51 | 0.46 | 0.66 | 0.80 | 0.64 | 0.66 | 0.45 | 0.58 | 0.10 | 0.81 | 0.38 | 0.40 | 0.84 | 0.75 |
| B2 | | | | | | | | 0.35 | 0.30 | 0.51 | 0.97 | -0.03 | **0.99** | 0.03 | 0.18 | 0.32 | 0.83 | 0.86 | 0.45 | 0.90 |
| B20 | | | | | | | | | 0.41 | 0.50 | 0.43 | 0.40 | 0.31 | 0.35 | 0.07 | 0.50 | 0.26 | 0.28 | 0.52 | 0.49 |
| B3 | | | | | | | | | | 0.74 | 0.46 | 0.74 | 0.24 | 0.63 | 0.07 | 0.80 | 0.20 | 0.22 | 0.80 | 0.60 |
| B4 | | | | | | | | | | | 0.67 | 0.77 | 0.44 | 0.68 | 0.10 | 0.91 | 0.37 | 0.39 | 0.94 | 0.80 |
| B5 | | | | | | | | | | | | 0.19 | 0.95 | 0.21 | 0.18 | 0.52 | 0.79 | 0.83 | 0.64 | 0.97 |
| B6 | | | | | | | | | | | | | -0.12 | 0.84 | 0.01 | 0.94 | -0.09 | -0.09 | 0.88 | 0.41 |
| B7 | | | | | | | | | | | | | | -0.04 | 0.18 | 0.23 | 0.83 | 0.86 | 0.37 | 0.86 |
| B9 | | | | | | | | | | | | | | | 0.02 | 0.80 | -0.03 | -0.02 | 0.76 | 0.39 |
| H2 | | | | | | | | | | | | | | | | 0.07 | 0.15 | 0.15 | 0.09 | 0.17 |
| N | | | | | | | | | | | | | | | | | 0.20 | 0.22 | **0.99** | 0.70 |
| T1 | | | | | | | | | | | | | | | | | | 0.71 | 0.32 | 0.72 |
| T2 | | | | | | | | | | | | | | | | | | | 0.34 | 0.75 |
| W3 | | | | | | | | | | | | | | | | | | | | 0.79 |