UHH
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DISSERTATION

Improved Ecological Niche Modelling of *Nothofagus pumilio* in the Southern Andes

vorgelegt von

Melanie Werner

2025

Dissertation zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Erdsystemwissenschaften
der Universität Hamburg

Fachbereich Erdsystemwissenschaften

Datum der Disputation:	06.02.2026
Gutachter der Dissertation:	Prof. Dr. Jürgen Böhner Prof. Dr. Udo Schickhoff
Zusammensetzung der Prüfungskommission:	Prof. Dr. Jürgen Böhner (Vorsitz) Prof. Dr. Udo Schickhoff Prof. Dr. Olena Dubovyk Prof. Dr. Céline Hadziioannou Prof. Dr. Kai Jensen
Vorsitzender des Fach-Promotionsausschusses Erdsystemwissenschaften:	Prof. Dr. Dirk Notz
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Danksagung

Zuerst möchte ich mich bei Prof. Dr. Jürgen Böhner bedanken, der mich bereits vor meinem Studium bei einer Informationsveranstaltung für Schüler:innen für das Fach Geographie begeistert hat. Während meines Studiums hat er meine Leidenschaft für die Physische Geographie nachhaltig geprägt. Ich bin sehr dankbar, meine Arbeit unter seiner Betreuung durchgeführt haben zu dürfen und für das Vertrauen, sie nach meinen eigenen Interessen und Schwerpunkten gestalten zu können. Prof. Dr. Udo Schickhoff hat meine Interessen für Biogeographie im besonderen Maße gefördert, für seine Unterstützung und die Betreuung meiner Arbeit bin ich sehr dankbar.

Dr. Maria Bobrowski und Johannes Weidinger haben wesentlich zum Gelingen meiner Arbeit beigetragen. Ich durfte nicht nur mein „methodisches Handwerk“ in ihren R-Kursen erlernen, sondern habe stets durch ihre Erfahrungen im wissenschaftlichen Arbeiten lernen dürfen. Ein besonderer Dank gilt Dr. Maria Bobrowski, die mir als Mentorin stets bei allen Fragen und Sorgen zur Seite stand. Ihre Arbeit zu *Betula utilis* war mir ein großes Vorbild für meine eigene Arbeit und nicht zuletzt hat sie einen wichtigen Anstoß zur Idee für den Instagram-Ground-Truthing-Approach geliefert. Weiterhin bedanke ich mich bei Dr. Jens Oldeland, der mich besonders im *Forests*-Review bei der Integration zusätzlicher Methoden unterstützt hat.

Den Alltag am Institut haben mir vor allem meine Kolleg:innen Nadine Kaul, Wakana Tateishi, Michèle von Kocemba, Malin Klein und Helge Jentsch versüßt. Auch in der Freizeit hatten meine Freund:innen ein offenes Ohr für mich. Vielen Dank, dass ihr mich auf diesem Weg begleitet habt!

Ich bin weiterhin sehr dankbar dafür, dass die Landesgraduiertenförderung der Freien und Hansestadt Hamburg den Abschluss meiner Arbeit vom 01. Oktober 2025 bis zum 31. März 2026 mit einem Stipendium unterstützt hat. Auf diese Weise konnte ich meine Arbeit fokussiert und sorgenfrei abschließen.

Zuletzt möchte ich meiner Familie und meinem Partner danken, ohne deren Unterstützung eine Promotion nicht möglich gewesen wäre. Mama und Papa, ich bin Euch unendlich dankbar für die Möglichkeiten, die Ihr mir im Leben ermöglicht habt und für den Halt, den Ihr mir gebt! Jede Sorge, jeden schönen und schwierigen Moment konnte ich direkt mit Dir teilen, vielen Dank Johannes, dass Du immer für mich da bist!

Zusammenfassung

Hochgebirgsregionen zählen zu den besonders vulnerablen Ökosystemen, die in besonderem Maße vom globalen Klimawandel betroffen sind. Eine Veränderungen ist der potenzielle Anstieg der alpinen Baumgrenze. Auch in den südlichen Anden ist die Baumgrenze von klimabedingten Veränderungen betroffen. Bisher wurden das radiale Wachstum sowie das Aufkommen und Überleben von Keimlingen der baumgrenzbildenden Art *Nothofagus pumilio* vor allem in kleinräumigen, dendrochronologischen und experimentellen Studien untersucht, die lokale Temperatur- und Niederschlagsmuster berücksichtigen. Ecological Niche Modelling (ENM) erlaubt dagegen eine flächenhafte Analyse aktueller Verbreitungsmuster und potenzieller zukünftiger Veränderungen über das gesamte Verbreitungsgebiet der Art. Entsprechende Modellstudien zu *N. pumilio* liegen jedoch bislang nicht vor. Ein wesentlicher Grund dafür liegt in der begrenzten Verfügbarkeit valider und möglichst unverzerrter Vorkommensdaten. Studien, die Daten durch Feldforschung *in situ* erheben, sind oftmals kosten- und zeitintensiv, insbesondere in schwer zugänglichen Hochgebirgsregionen. Daher wird häufig auf globale Biodiversitätsdatenbanken wie die Global Biodiversity Information Facility (GBIF) zurückgegriffen, die jedoch Daten mit räumlichem und zeitlichem Bias enthalten können. Ein Großteil der in solchen Datenbanken enthaltenen Daten stammt aus Citizen-Science-Projekten, bei denen Nicht-Expert:innen z.B. mit Hilfe von Smartphones Artvorkommen erfassen. Dieses Vorgehen erlaubt eine schnelle und kostengünstige Datensammlung, weist jedoch räumliche Verzerrungen auf, da Beobachtungen vor allem in urbanen oder touristisch erschlossenen Regionen erfolgen und somit kein vollständiges Bild der tatsächlichen Verbreitung von Arten liefern. Fernerkundungsmethoden stellen eine vielversprechende Alternative zur herkömmlichen Felddatenerhebung dar, da sie die Erfassung von Artvorkommen auch in abgelegenen und schwer zugänglichen Regionen ermöglichen. Allerdings ist eine Validierung der fernerkundungsbasierten Ergebnisse erforderlich, um sicherzustellen, dass die jeweilige Art tatsächlich in den identifizierten Gebieten vorkommt (Ground Truthing). Hierfür sind wiederum *in-situ*-Daten notwendig. In jüngerer Zeit rücken Soziale Medien zunehmend in den Fokus, da sie eine hohe globale Reichweite aufweisen und die Zahl georeferenzierter Beiträge stetig zunimmt. Diese Inhalte bieten ein bislang wenig genutztes Potenzial zur Generierung von Vorkommensdaten.

Ziel der vorliegenden Arbeit mit dem Titel „Improved Ecological Niche Modelling of *Nothofagus pumilio* in the Southern Andes“ ist es, die derzeitige und zukünftige potenzielle Verbreitung von *N. pumilio* in den südlichen Anden mit Hilfe eines ENM-Ansatzes zu modellieren und somit die bestehende Forschungslücke zu schließen. Dabei wurde durch die Optimierung der Eingangsdaten und den Einsatz von Machine-Learning-Algorithmen ein verbesserter ENM-Ansatz entwickelt. Die Studie erfolgte in zwei Schritten: Zunächst wurde das Potential von Sozialen Medien für die Erhebung von Artvorkommen getestet und ein neuartiger „Instagram ground truthing approach“ (IGTA)

entwickelt, der auf der Auswertung von Instagram-Beiträgen basiert (Publikation I und II). Dabei wurden Vorkommensdaten generiert, die einen geringeren räumlichen Bias aufweisen als klassische GBIF-Daten. Die IGTA-Punkte wurden anschließend zum Ground Truthing von Fernerkundungsdaten über das gesamte Verbreitungsgebiet der Art genutzt. Die zwei daraus resultierenden Datensätze gingen abschließend in die Modellierung ein (Publikation III und IV).

Aufgrund der hohen globalen Reichweite mit zwei Milliarden Nutzern weltweit und der Fokussierung auf Bildinhalte eignet sich Instagram besonders für die Erhebung georeferenzierter Beobachtungen. Auf Basis eines klar definierten Kriterienkatalogs wurden Bilder ausgewählt, auf denen *N. pumilio* eindeutig identifizierbar war und deren Standort anhand von Landschaftsmerkmalen lokalisiert werden konnte. Insgesamt wurden 1.238 Vorkommenspunkte manuell georeferenziert. Aufgrund der hohen Anzahl verfügbarer Instagram-Beiträge konnten Vorkommen über das gesamte Verbreitungsgebiet hinweg erfasst werden. Eine Bias-Analyse mit dem R-Package „sampbias“ zeigte, dass die IGTA-Daten weniger stark in Richtung urbaner Zentren verzerrt sind als GBIF-Daten. Anschließend dienten die IGTA-Daten als Referenz zum Ground Truthing von Fernerkundungsdaten, die über eine Klassifikation (supervised classification) von Sentinel-2-Level-2A-Daten generiert wurden. Dadurch entstanden zwei Vorkommensdatensätze, der IGTA-Punktdatensatz und die Sentinel-2-Rasterdaten, die für die Modellierung genutzt werden können.

Zur Modellierung wurde der Machine-Learning-Algorithmus Random Forest (RF) eingesetzt, der sich durch eine hohe Vorhersagekraft und gute Interpretierbarkeit auszeichnet. Zudem eignet er sich sowohl für die punktbasierten IGTA-Daten als auch für die kontinuierlichen Rasterdaten, wodurch ein direkter Vergleich beider Ansätze möglich war. Für den IGTA-Punktdatensatz wurde ein RF-Klassifikationsmodell, für die Rasterdaten ein RF-Regressionsmodell berechnet. Zur Erfassung der ökologischen Nische dienten die bioklimatischen Variablen von CHELSA als Prädiktoren. Beide Modelle lieferten valide Projektionen der potenziellen gegenwärtigen und zukünftigen Verbreitung der Art. Es zeigte sich eine potenzielle Verschiebung des Verbreitungsgebiets in höhere Lagen sowie ein Rückgang in trockeneren Regionen, insbesondere in Nordpatagonien. Diese Ergebnisse stimmen mit denen aus der Literatur überein, wonach die Art von höheren Temperaturen in Gebieten mit ausreichend Niederschlag profitiert, während es in trockenen Gebieten zu Dürren und zum Rückgang der Art kommt. Der Vergleich beider Modellansätze zeigte, dass die Raster-Vorkommensdaten zu einer aussagekräftigeren Variableninterpretation und vor allem zu einer besseren räumlichen Vorhersage führten.

Der Einsatz Sozialer Medien zur Erfassung von Vorkommensdaten stellt eine innovative Methode mit großem Potenzial für großräumige ökologische Analysen dar. Der IGTA verbindet die Vorteile von Citizen Science, wie schnelle und kostengünstige Datenerhebung, mit einer deutlichen Reduktion der damit häufig verbundenen Verzerrungen. Obwohl die manuelle Bildauswertung noch aufwändig ist, bieten

automatisierte Schnittstellen (APIs) und KI-basierte Erkennungsverfahren künftig großes Potenzial zur Effizienzsteigerung. Während die Nutzung von Fernerkundungsdaten zur Ableitung von Umweltvariablen im ENM bereits etabliert ist, sollte auch die Erfassung von Vorkommensdaten aus Fernerkundung künftig weiterentwickelt werden, durch die sich modellgestützte Aussagen für geeignete Arten deutlich verbessern lassen. Die vorliegende Arbeit präsentiert einen innovativen ENM-Ansatz für *N. pumilio* und schließt eine bislang bestehende Forschungslücke in der Modellierung klimabedingter Veränderungen an der Baumgrenze in den südlichen Anden.

Abstract

High mountain regions are among the most vulnerable ecosystems and are particularly affected by global climate change. One consequence is the potential upward shift of the alpine treeline. In the Southern Andes, the treeline is likewise influenced by climate-related factors. To date, radial growth, seedling emergence and survival of the treeline species *Nothofagus pumilio* have mainly been studied in small-scale dendrochronological or experimental studies that consider local temperature and precipitation patterns. In contrast, Ecological Niche Modelling (ENM) allows for comprehensive analyses of current distribution patterns and potential future changes across the entire study area. However, corresponding ENM studies for *N. pumilio* are currently lacking. One major reason is the limited availability of valid and non-biased species occurrence data. The sampling of *in situ* observations through fieldwork is often time-consuming and costly, especially in remote high mountain regions. Consequently, global biodiversity databases such as the Global Biodiversity Information Facility (GBIF) are frequently used, despite the fact that they may contain data with spatial and temporal biases. A significant proportion of the data in such databases originates from Citizen Science (CS) projects, in which non-experts collect species occurrence data for example by using smartphones. While this approach allows for fast and cost-effective data sampling, it introduces spatial bias, as observations are predominantly recorded in urban or tourist areas, resulting in incomplete representations of species' actual distributions. Remote sensing provides a promising alternative to traditional fieldwork, as it enables species detection even in inaccessible and remote regions. However, validation of remote sensing results is required to confirm that the target species actually occurs in the identified areas (ground truthing). This, in turn, requires *in situ* observations. Recently, social media has gained increasing attention in biodiversity research due to its global reach and the increasing number of georeferenced posts. These sources offer untapped potential for generating occurrence data.

The aim of this thesis, entitled "Improved Ecological Niche Modelling of *Nothofagus pumilio* in the Southern Andes," is to model the potential current and future distribution of *N. pumilio* in the Southern Andes using an ENM approach. By optimising input data and incorporating machine learning algorithms, an improved modelling framework was developed. The study was conducted in two main steps: first, the potential of social media for species occurrence data sampling was explored through the development of a novel “Instagram ground truthing approach” (IGTA) (Publication I and II). This approach generated occurrence points with reduced spatial bias compared to conventional GBIF data. The IGTA points were then used to validate remote sensing data across the entire distribution range of the species (ground truthing). Second, the two resulting datasets were used for ENM (Publications III and IV).

Due to its global reach with two billion users and its image-based format, Instagram is particularly suitable for collecting georeferenced biodiversity data. Based on a catalogue of specific criteria, images were selected in which *N. pumilio* was clearly identifiable and the location could be determined using landscape elements. A total of 1,238 occurrence points were manually georeferenced. Due to the large volume of available posts, occurrences could be documented across the entire distribution range of the species. A bias analysis using the R package “sampbias” demonstrated that the IGTA dataset exhibited less spatial bias than GBIF data. The IGTA data were then used to validate remote sensing data derived from supervised classification of Sentinel-2, Level 2A imagery. This process yielded two occurrence datasets, the IGTA point dataset and the Sentinel-2 raster dataset, suitable for use in ENM.

For the modelling process, the Random Forest (RF) machine learning algorithm was employed due to its high predictive performance and interpretability. Moreover, RF can be applied to both point-based binary and continuous raster occurrence data, enabling a direct comparison of the two approaches. An RF classification model was developed for the IGTA point dataset, while an RF regression model was created for the continuous raster dataset. CHELSA bioclimatic variables were used as environmental predictors. Both models resulted in valid predictions of the species’ ecological niche and its potential current and future distribution. The results indicate a potential shift of the distribution range to higher elevations and a decline in drier regions, particularly in northern Patagonia. These findings are consistent with existing literature indicating that the species benefits from higher temperatures in areas with sufficient precipitation, whereas population declines due to drought are expected in drier areas. The comparison of both modelling approaches showed that the continuous raster data enabled a more meaningful interpretation of predictor variables and improved spatial predictions.

The use of social media for sampling species occurrence data represents an innovative and promising method. The IGTA combines the strengths of CS, such as rapid and cost-effective data collection, with a reduction in spatial bias. Although the manual analysis of posts is time-consuming, the process could be improved through the use of automated processes with Application Programming Interfaces (APIs) or artificial intelligence (AI) in the future. While the use of remote sensing to derive environmental predictors is already well established in ENM, the sampling of species occurrence data through remote sensing also holds considerable potential and should be further developed. For suitable species, this approach can significantly improve model accuracy and ecological interpretability. Therefore, this study presents an improved ENM workflow for *N. pumilio* and addresses a previously existing research gap.

Outline

List of Figures, Tables and Abbreviations

Figures

Tables

Abbreviations

AAO	Antarctic Oscillation
AI	Artificial Intelligence
API	Application Programming Interface
AUC	Area Under the Curve
BAM	Biotic-Abiotic-Movement framework
BP	Before Present
CHELSA	Climatologies at high resolution for the earth’s land surface areas
CMIP	Coupled Model Intercomparison Project
DSM	Digital Surface Model
ENM	Ecological Niche Model(ling)
ENSO	El Niño Southern Oscillation
G	Geographic space
G_A	Abiotic suitable conditions in geographic space, fundamental niche
GBIF	Global Biodiversity Information Facility
G_I	Invadable distributional area
G_O	Occupied distributional area
IGTA	Instagram ground truthing approach
NDVI	Normalized Difference Vegetation Index
PDO	Pacific Decadal Oscillation
PDP	Partial Dependence Plots
RCP	Representative Concentration Pathways
RF	Random Forest
SAM	Southern Annular Mode
SDM	Species Distribution Model(ling)
SHAP	Shapley Additive Explanations
SSP	Shared Socioeconomic Pathways
TSS	True Skill Statistics

1. Introduction

Anthropogenic climate change is impacting ecosystems worldwide. However, certain regions of the Earth are disproportionately affected by the associated warming and its consequences (IPCC, 2023). High mountain regions are experiencing above-average temperature increases, primarily due to albedo feedback mechanisms (Pepin et al., 2015). Consequently, ecosystems near the alpine treeline are particularly vulnerable (Schickhoff et al., 2022). Emerging new abiotic and biotic conditions are leading to habitat shifts and losses (Schickhoff et al., 2022; IPCC, 2023). One consequence, which is regarded as a key indicator of climate change, is the potential upward shift of the alpine treeline (Qiu et al., 2023).

The alpine treeline marks the upper elevational limit at which tree growth remains possible (Paulsen and Körner, 2014). The principal limiting factor for tree growth is heat deficiency (Körner, 2020), therefore, the location of treelines globally closely aligns with the 6.4°C isotherm during growing season (Paulsen and Körner, 2014). Nevertheless, treelines exhibit substantial variability in both composition and structure. A basic distinction can be made between diffuse, abrupt and island-shaped treelines (Harsch and Bader, 2011). In diffuse treelines, tree growth gradually declines as a result of increasing heat limitation, eventually leading to a transition from upright trees to krummholz growth forms. In contrast, dieback processes play a significant role at abrupt or island treelines (Bader et al., 2021). Factors such as strong winds, frost events, or unfavourable soil conditions outside the protecting forest stand lead to increased seedling mortality (Harsch and Bader, 2011). With increasing temperatures, it is highly probable that the treeline position will shift to higher elevations at diffuse sites, however, this upward shift may be constrained at abrupt treelines (Harsch and Bader, 2011; Treml and Veblen, 2017). Furthermore, various additional factors can influence the position and upward shift of the treeline. Besides climatic treelines, distinctions are also made between edaphic, topographic and anthropogenic treelines (Schickhoff et al., 2020). The treeline elevation can be suppressed due to snow accumulations, cold air drainage, or edaphic and topographic factors like steep slopes, rock beds or avalanches, shallow soil depth, waterlogged, dry or nutrient-poor soils (Körner, 2007; Hadley et al., 2013; Schickhoff et al., 2020). Human activities likewise exert a significant influence on the treeline ecotone. Many treelines, particularly in the Northern Hemisphere, are anthropogenic treelines (Treml et al., 2016; Vitali et al., 2019; Miehe et al., 2023). These are shaped by disturbances such as deforestation, livestock grazing, or use for local recreation and tourism, and thus often lie below the altitudinal limit that would be climatically possible (Schickhoff et al., 2020). Globally, 66 % of treelines have been reported to be advancing (Hansson et al., 2021). In the Northern Hemisphere, 90 % of treelines are advancing (Lu et al., 2021), whereas those in the Southern Hemisphere are advancing less markedly. This difference is largely attributed to the abrupt nature of many Southern Hemisphere treelines (Hansson et al., 2023).

Treelines and their dynamics have been extensively studied in recent decades (Körner, 2012; Holtmeier and Broll, 2020). Many investigations have focused on treelines in the Northern Hemisphere, including those in the Rocky Mountains, the Alps, the Scandes, and the Himalayas. In contrast, treelines in the Southern Hemisphere remain significantly underrepresented in the scientific literature (Hansson et al., 2021, 2023). Although small-scale studies, primarily dendrochronological and experimental studies, have been conducted in the Southern Andes, there is a lack of modelling studies that capture the treeline in its entirety and simulate both current and future conditions. The treeline of the Southern Andes represents a particularly interesting research subject. It is abrupt, largely natural, and minimally influenced by human activity (Holtmeier, 2009). The species forming the treeline is the deciduous species *Nothofagus pumilio* (Poepp et Endl.) Krasser. Recent studies have already shown that this species is sensitive to climatic variations, often associated with phase shifts in Antarctic Oscillation (AAO) also known as the Southern Annular Mode (SAM), the El Niño Southern Oscillation (ENSO), and Pacific Decadal Oscillation (PDO) (Álvarez et al., 2015; Srur et al., 2016, 2018). Higher temperatures, in combination with sufficient precipitation, promote increased radial growth (Álvarez et al., 2015; Brand et al., 2022; Reiter et al., 2024) and the establishment of seedlings above the treeline (Srur et al., 2016, 2018). However, drought conditions, particularly increasing in northern Patagonia, can lead to growth limitation and mortality among both seedlings and mature individuals (Rodríguez-Catón et al., 2016; Fajardo et al., 2019). Understanding the climatic conditions at the treeline, along with the potential current and future distribution of this species, is therefore of considerable interest. These questions can be addressed using Ecological Niche Modelling (ENM). Nevertheless, such modelling studies are currently lacking.

The lack of modelling studies can largely be attributed to the limited availability of species occurrence data. ENM approaches typically rely on georeferenced point data indicating the presence or absence of species (Sillero et al., 2021). Such data can either be collected through fieldwork or obtained from online databases (Feng et al., 2019). The advantage of field-based data lies in the fact that it is usually gathered directly by experts, thereby minimising taxonomic misidentifications (e.g., Fitzpatrick et al., 2009). However, fieldwork is both time-consuming and costly and tends to be restricted to relatively small study areas. Open-access databases such as the Global Biodiversity Information Facility (GBIF) compile data from various sources (GBIF, 2025a). These include expert-verified field observations, museum collections, and increasingly contributions from Citizen Science (CS) initiatives. CS refers to projects in which “non-experts”, referred to as Citizen Scientists, participate in scientific research, for example by documenting species occurrences (Bonney, 1996; Bonney et al., 2009). This is commonly done using mobile applications that allow users to take photographs, automatically record coordinates, and generate species identifications. Despite their growing popularity, GBIF datasets are subject to several well-documented sources of bias (Beck et al., 2014; Meyer et al., 2016; Daru et al., 2018). These include spatial, taxonomic, and temporal biases. Sampling methods and observer behaviour can lead to species misidentification and to data collection being concentrated in urban areas, resulting in spatial bias, while inaccurate GPS readings can lead to coordinate errors (Di Cecco et al., 2021). Moreover, CS projects typically do not cover areas that are difficult for humans to access, such as high mountain regions with complex topography.

A promising approach that circumvents many of the aforementioned biases is the collection of species occurrence data through remote sensing techniques. Remote sensing enables comprehensive surveys of large areas, including those that are inaccessible to humans (He et al., 2015). Very high-resolution imagery (e.g., IKONOS, WorldView, RapidEye, or airborne data) is commonly used in small-scale studies to identify individual species or forest types (Fassnacht et al., 2016). For larger spatial extents, medium- to high-resolution datasets such as Landsat or Sentinel are typically employed (Immitzer et al., 2016; Immitzer et al., 2019). However, ground truthing, the verification whether the species of interest is indeed present in the area detected by remote sensing, remains necessary (Nagai et al., 2020). This, in turn, requires occurrence data that has been collected on-site. Recently, social media has gained attention as a potential tool for both collecting occurrence data and supporting ground truthing (Jarić et al., 2020). Platforms such as Facebook, Flickr, Instagram, Twitter (now X) and YouTube have been used to gather occurrences of animal and plant taxa (e.g., Hentati-Sundberg and Olsson, 2016; ElQadi et al., 2017; Pace et al., 2019; Gibson et al., 2020; Martino et al., 2021; Virić Gašparić et al., 2022; O'Neill et al., 2023). The quantity of geo-tagged images on social media is steadily increasing, offering significant potential for the analysis of suitable image and video contributions useful for species occurrence sampling. For example, on Instagram, the third most widely used social media platform worldwide (after Facebook and YouTube), with 2 billion active users (Statista, 2025), over 60 million contributions are posted daily (WirtschaftsWoche, 2014). Some platforms enable a manual search based on location tags or hashtags (metadata keywords marked with a hash symbol, e.g., #nothofaguspumilio), while others offer access via Application Programming Interfaces (APIs), allowing for automated searches (e.g., Flickr, Fox et al., 2022). Given the wide reach of social media platforms plus the fact that social media content is analysed by experts, social media occurrence data sampling has the potential to generate large volumes of less-biased occurrence data and is still largely untapped.

This thesis introduces a novel Instagram ground truthing approach (IGTA), which used the social media platform Instagram for species occurrence data sampling aiming at creating less-biased occurrence data of *N. pumilio* for ENM. The IGTA comprised two main steps. First, Instagram posts were systematically searched using a strict set of criteria to identify both the species *N. pumilio* and the location where the corresponding photo was taken. The resulting data were then georeferenced and transferred to a map as point occurrences. Due to the high volume of potentially suitable posts and the manual georeferencing of occurrence points, allowing for multiple points per post e.g., occurrences in the background of posted photos, spatial sampling bias can be reduced. Bias was detected and compared with that found in GBIF data. In the next step, to further mitigate spatial bias, the generated IGTA point dataset was subsequently used for ground truthing large-scale remote sensing data. Raster data of *N. pumilio* occurrences was derived from supervised classification of Sentinel-2 Level 2A imagery. Therefore, the IGTA approach resulted in two occurrence datasets of *N. pumilio*: a point dataset and a raster dataset covering the entire range of the species in the Southern Andes. Both datasets can be used to model the ecological niche, and the potential current and future distribution of the species. Two models were developed and compared: one using a common binary (point-based) ENM approach, and another employing a more innovative method based on continuous raster data.

The subsequent chapters provide a comprehensive overview of the methodological background and conceptual framework of Ecological Niche Modelling (Chapter 2), including a detailed presentation of the data underlying the models (Chapters 2.1 and 2.2), and the algorithms applied (Chapter 2.3). Chapter 3 introduces the study area and study species in detail. Chapter 4 presents the key methods and content of the peer-reviewed publications that form the core of this cumulative dissertation (full texts available in the appendix). Finally, Chapter 5 synthesises the findings, and Chapter 6 offers a conclusion and outlines future research priorities.

2. Ecological Niche Modelling: Concepts, Data, Limitations

One of the most important questions for biogeographers is understanding the spatial and temporal distribution patterns of species (Thuiller, 2024). In the last decades, this question has been investigated using Ecological Niche Modelling (ENM). ENM, also known as Species Distribution Modelling (SDM) and Habitat Suitability Modelling, is an approach to investigating the current, past and future distribution of species (Thuiller, 2024). This is achieved by a mathematical approach using species observations and environmental predictor variables as well as correlative or machine learning algorithms (Zurell and Engler, 2019). Typical areas of application are forecasting the effects of climate change, land use, or anthropogenic factors on biodiversity, planning of protected areas, and monitoring habitat shifts or the invasive potential of species (Araújo et al., 2019; Thuiller, 2024). ENM has therefore developed into an indispensable application for biodiversity conservation and management (Franklin, 2010; Thuiller, 2024). The terms Ecological Niche and Species Distribution Modelling are often used synonymously. However, ENM and SDM differ in terms of the subject of investigation and the underlying hypothesis (Peterson and Soberón, 2012). Peterson and Soberón (2012) highlight the differences between the concepts. SDM is primarily concerned with modelling the current distribution of a species based on observed occurrences without making any statements about the niche of a species. ENM, on the other hand, focuses primarily on modelling the underlying abiotic and biotic (ecological) environmental conditions under which a species can occur (Peterson and Soberón, 2012). According to the authors, SDM can only refer to modelling the actual current distribution of a species. In contrast, modelling under changing environmental conditions relates to the niche concept, as it involves estimating the fundamental niche and potential changes to it (Peterson and Soberón, 2012). Although many studies apply models to study changes over time, the term SDM is used more frequently than ENM. The term SDM can be understood as an umbrella concept under which numerous applications and methods have been published (Elith and Leathwick, 2009; Franklin, 2010; Bobrowski, 2018). In the following, however, the term ENM is used and the underlying concepts are described accordingly.

ENM is a valuable tool for predicting suitable conditions in space and time for a study species (Anderson, 2012). ENMs can be divided into three categories according to the extent to which biological processes are included: mechanistic, correlative, and hybrid models (Sillero et al., 2021). Mechanistic ENMs incorporate biotic variables about physiological, morphological, and behavioural processes, while correlative and hybrid ENMs use geographical data of species occurrences and environmental predictors. In addition to geographical data, hybrid ENMs also incorporate process-based influences such as biotic interaction and dispersal abilities (Sillero et al., 2021). The following work refers to correlative models. The classic correlative approach consists of data on the occurrence of the study species and environmental predictors like climate, soil, topography, and biotic data. The species data can be presence–only, presence–

background and presence–absence (Peterson et al., 2011). Some authors describe that the type of species occurrence data used determines which kind of niche can be modelled, for example the fundamental, realised, or potential niche (Franklin, 2010). The precise definition of the niche being modelled is a critical aspect of ENM, and this is where the niche concept becomes important.

Defining the term “ecological niche” has a long and contested history in ecology (Sales et al., 2021). Following its initial introduction by Grinnell in 1917, the concept was further developed, leading to numerous interpretations and applications (Peterson, 2011). Grinnell originally defined the “niche” as the environmental (abiotic) conditions that allow a species to survive and reproduce (Grinnell, 1917; Sillero et al., 2021). The Grinnellian niche is thus based on abiotic conditions, whereas a later definition by Elton describes the niche as the functional role of a species within an ecosystem and has therefore a focus on biotic conditions. The Eltonian niche refers for example to biotic interactions such as what a species consumes and what predators it faces (Elton, 1927; Sillero et al., 2021). Hutchinson later combined abiotic and biotic dimensions and was the first to introduce the concepts of the “realised” and “fundamental” niche. A fundamental niche refers to the geographic space in which a species can occur and reproduce successfully without concurrence, whereas the realised niche describes the space that is actually occupied by a species despite competition (Hutchinson, 1957; Sillero et al., 2021). Both concepts are important for ENM, since the scope of a modelling approach and the underlying hypotheses depend on them. In the 21st century the Biotic-Abiotic-Movement framework (BAM) was developed for ENM applications (Soberón and Peterson, 2005). It supplements concepts of Grinnell and Hutchinson and is the basic framework for ENM studies (Zurell and Engler, 2019).

The BAM framework is illustrated in Figure 1a. In this framework, **G** represents the geographic space, **A** denotes the abiotic conditions suitable for a species’ survival and reproduction, **B** includes biotic interactions, and **M** defines the area that is accessible to a species without dispersal limitations. The intersection of **G** and **A**, $\mathbf{G_A}$, represents the species’ fundamental (Grinnellian) niche, while the intersection of **A**, **B**, and **M**, referred to as $\mathbf{G_O}$, represents the actual occupied area (Soberón and Peterson, 2005; Peterson, 2011; Peterson and Soberón, 2012), which some authors synonymously define as the realised niche (Sillero et al., 2021). However, some authors define the realised niche strictly as the intersection of **A** and **B**, in line with Hutchinson’s classical concept (Soberón and Peterson, 2005; Soberón, 2007; Zurell and Engler, 2019). ENM aims at predicting suitable environmental conditions for a species, yet there is ongoing debate about which niche is modelled (Zurell and Engler, 2019). This distinction depends largely on the type of species occurrence data used. Correlative ENMs based on presence–only data tend to approximate the potential or realised niche, while presence–absence models can capture the occupied niche (Franklin, 2010; Zurell and Engler, 2019). In this study, a more neutral terminology is used. The presence–absence approach used here contains information on **A**, **B**, and **M**: environmental predictors represent **A** and presence data originate from $\mathbf{G_O}$, indirectly reflecting **B** and **M**. However, to estimate $\mathbf{G_O}$ precisely, additional information on **M** would be required in order to delineate **G_I**, the invadable area constrained by dispersal limitations (Peterson, 2011; Peterson and Soberón, 2012). Therefore, the ENM in this study is understood to predict the potential geographic distribution (**G_P**), defined as the area where abiotic and biotic conditions (intersection of **A** and **B**) are suitable for the species' survival and reproduction. Following the authors' recommendation, the term "potential current and future distribution" is used throughout this study to describe the modelled distribution of *Nothofagus pumilio* (Peterson and Soberón, 2012).

Lately, several guidelines for ENMs have been published to help developing an ENM workflow (e.g., Araújo et al., 2019; Feng et al., 2019; Sillero et al., 2021). Figure 1b shows the basic structure of an ENM approach. The ENM workflow consists of the preparation and integration of two input datasets, the occurrence data and the environmental data, followed by the model setup and algorithm choice, the evaluation and calibration of the model, and finally, the transfer in space or time (Anderson, 2012). Each step involves specific strengths and limitations. While the choice of methods is critical for producing accurate model outputs, potential sources of error in the input data must also be addressed. These include bias in the occurrence data, inaccuracies in the (climatic) predictors or their preprocessing, spatial autocorrelation, and multicollinearity among variables. Opportunities and limitations related to these aspects are discussed in detail in the following chapters.

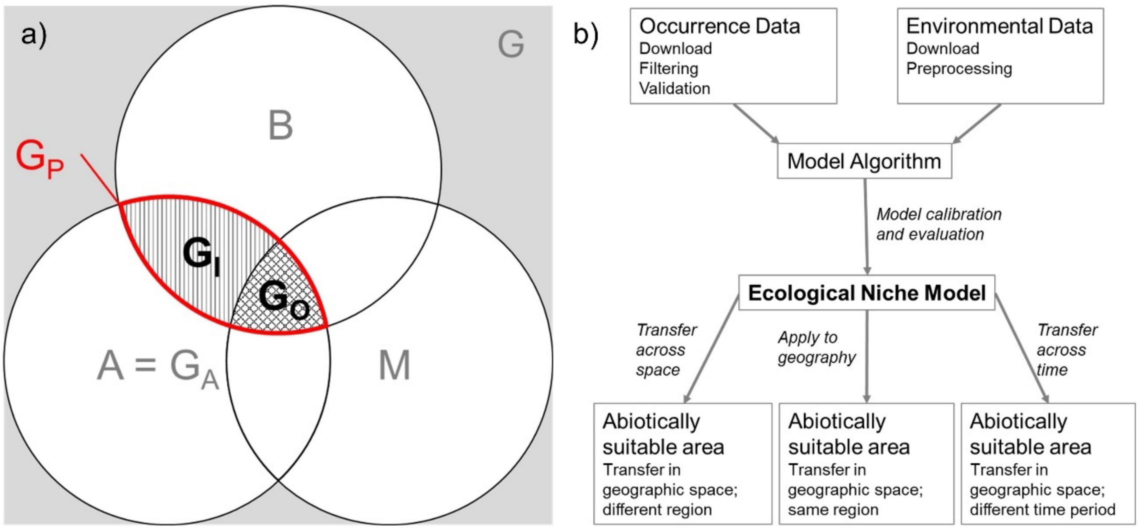


Figure 1: a) Visualisation of the Biotic-Abiotic-Movement (BAM) scheme (after Peterson 2011). In ***G****, the geographic space,* ***A*** *abiotic and* ***B*** *biotic conditions for a species are given.* ***M*** *(movement) is the space accessible for the species, where it can move without limitations.* ***G_A*** *is the fundamental niche of the species, while the intersection between* ***A****,* ***B*** *and* ***M*** *is* ***G_O*** *the occupied distributional area.* ***G_P*** *is the potential geographic distribution that includes* ***G_I****, the invadable distributional area (Soberón and Peterson 2005; Peterson, 2011; Peterson and Soberón, 2012). b) shows the typical workflow of an Ecological Niche Modelling (ENM) approach (after Anderson, 2012).*

2.1 Species Occurrence Data

Correlative ENMs require georeferenced species occurrence data. These can either be *in situ* data from fieldwork sampled by experts, digitalised data from museums, or data from online databases that come from diverse sources (Thuiller, 2024). In the last decades, online databases make biodiversity data largely available and are frequently used by ecological modellers (Michener et al., 2012; Feng et al., 2019). Expert and museum data as well as data from Citizen Science (CS) projects are available in global databases such as the Global Biodiversity Information Facility (GBIF) (Anderson et al., 2016; GBIF, 2025a). Especially CS data contribute significantly to such databases (Goldberg, 2023). CS is the active involvement of “non-experts” in scientific research (Bonney, 1996; Bonney et al., 2009). In the case of ecological modelling this is primarily the sampling of species occurrences with mobile devices (Kullenberg and Kasperowski, 2016). One example of a CS project is iNaturalist. In the iNaturalist app, photos of biodiversity observations with meta data of coordinates, date, and time can be uploaded with a smartphone or PC. Users can identify the species based on suggestions in the app, which are reviewed and updated by other users or experts (curators) (Heberling and Isaac, 2018; Goldberg, 2023; iNaturalist, 2025). After reaching a consensus by at least two users, occurrences are considered correctly identified, complete and certain, and only then are they passed on to biodiversity data repositories like GBIF for publication as valid data (Heberling and Isaac, 2018). iNaturalist serves as one of the main contributors to GBIF with over 131 million hosted occurrence records (GBIF, 2025b). Therefore, CS is a very important contribution to species occurrence data sampling (Feldman et al., 2021), offering a cost- and time-efficient method for large-scale monitoring (Sumner et al., 2019). Some authors even emphasise that CS is the only practical way to investigate range shifts of species in large scale (Dickinson et al., 2010). Others found that the data quantity increases exponentially (Pocock et al., 2017) and the data quality is improving rapidly, nearing expert level (Aceves-Bueno et al., 2017; Mesaglio and Callaghan, 2021).

Recently, social media gained attention towards its contribution to species occurrence data sampling (Deng et al., 2012; Jarić et al., 2020). With its wide reach and vast amount of posted content with geotags, social media offers a still largely untapped potential for surveying species occurrences (Deng et al., 2012). Just like CS, the use of social media content has the potential to increase species occurrence data in a cost- and time-efficient way (Edwards et al., 2021). However, CS usually requires the active involvement and information transfer to Citizen Scientists (Wiggins and Crowston, 2011). When social media content is re-used and analysed by scientists, content creators are rarely aware that they are participating in a study. Therefore, social media species occurrence data sampling can be assigned to “passive Citizen Science” (Nascimento et al., 2024). Social media platforms such as Facebook, Flickr, Instagram, X (formerly Twitter), and YouTube are used to sample occurrence data and range shifts of animals, for example data of whales, dolphins (e.g., Pace et al., 2019; Gibson et al., 2020; Martino et al., 2021), birds (e.g., Hentati-Sundberg and Olsson, 2016), snakes (e.g., Marshall and Strine, 2019), insects (e.g., Virić Gašparić et al., 2022; O’Neill et al., 2023), and plants (e.g., ElQadi et al., 2017).

However, even though data availability is increasing, data from global databases are subject to several biases that must be addressed in ENM (Boakes et al., 2010; Beck et al., 2014; Meyer et al., 2016; La Sorte et al., 2024). Species data can inhibit taxonomic, sampling, and spatial bias (Meyer et al., 2016). Field trips by experts are, on the one hand, costly and time-consuming, often limited to small study areas that cannot cover the whole range of a species, and limited to areas that are accessible for humans. On the other hand, expert data are considered to be mostly free of species misidentifications or sampling biases. When species occurrences are sampled by (unstructured) CS projects, misidentifications are common errors (Bird et al., 2014), especially for rare species (Cox et al., 2012). Furthermore, sampling biases such as wrong coordinates can occur mainly due to weak satellite signals while sampling with mobile devices (Uyeda et al., 2020). One major bias is spatial bias. Citizen Scientists record the occurrence of species primarily in areas of their everyday lives. For example, on their way to work, during walks, or on holiday (Dimson and Gillespie, 2023). This observer behaviour leads to the accumulation of occurrence points in areas that do not necessarily correspond to the actual distribution range of species (Di Cecco et al., 2021). Areas that are difficult or impossible for humans to access are usually not recorded at all. Even on a global scale, the data is spatially biased, as sampling of occurrences takes place primarily in the Northern Hemisphere (North America and Europe) (Di Cecco et al., 2021; Feldmann et al., 2021). The potential of social media species occurrence data sampling to reduce these biases effectively has not yet been tapped. As globally posted content is widely available and accessible, and the data are collected and evaluated by experts, both sampling and spatial biases may be mitigated. However, user behaviour on social media often mirrors that of Citizen Scientists, meaning that certain biases still persist.

Therefore, occurrence data must be examined for these biases in order to avoid distorting the modelling results (Meyer et al., 2016). There are several methodological solutions to address the described sources of bias in occurrence data (Bird et al., 2014, Sillero et al., 2021). First steps while and after downloading the data are intense filtering and bias detection methods (Feng et al., 2019). Within the download process, data can be checked for missing coordinates and excluded. Some databases such as GBIF indicate the coordinate uncertainty, so that data with high variations in coordinates can also be removed. Nevertheless, it is crucial to plot occurrence points and check for false deviations. For example, spatial and environmental outliers, such as terrestrial plant species occurring far from their natural distribution area or on water or glacier surfaces, indicate an erroneous recording (Feng et al., 2019). The degree of spatial bias in the data can be determined using bias detection methods such as the R package “sampbias” (Zizka et al., 2021). “Sampbias” compares the position of species records with geodata on urban areas and calculates the sampling rate. A high sampling rate in urban areas and very low values in rural areas indicate significant spatial bias (Zizka et al., 2021). Additional data can be necessary to reduce spatial bias or the proportion of “clustered” occurrences in urban area can be reduced (filtering) (Sillero and Barbosa, 2021). Detecting taxonomic or sampling bias is more difficult. Misidentified species or incorrect coordinates can rarely be identified retrospectively and are difficult to avoid entirely. While such biases can lead to substantial errors in some algorithms, others (e.g., Boosted Regression Trees, Random Forest, Maximum Entropy) are considered more robust to these issues (Sillero et al., 2021).

After downloading and filtering the occurrence data, they must be pre-processed and formatted appropriately for modelling. Data types for correlative ENMs can be presence–only, presence–background, or presence–absence/pseudo-absence data (Sillero et al., 2021). The choice of data type depends on the modelling objective and the algorithm applied. Presence–only models rely solely on species presence as the response variable and typically apply envelope algorithms to predict habitat suitability (Booth et al., 2014; Sillero et al., 2021). Presence–background data represents a special case, most commonly used in Maximum Entropy (MaxEnt) model approaches (Phillips et al., 2006). Background data refers to areas within the study area where no species occurrences have been recorded. The abiotic and biotic conditions at presence locations are compared to those at locations in the background (Phillips et al., 2009; Sillero et al., 2021). In contrast to true absence data, background data covers the whole study area, that may include sites where the species is actually present (Sillero and Barbosa, 2021). Model validity can be increased by incorporating true absences, representing locations where a species is known to be absent (Sillero et al., 2021). However, true absence information is rarely available. When working with geospatial data, it is common practice to generate so-called pseudo-absences (VanDerWal et al., 2009; Barbet-Massin et al., 2012; Broussin et al., 2024). These are absence points created in grid cells e.g., in the resolution of the species or climate data, that do not contain any recorded presence points and are therefore treated as absences for modelling purposes (Sillero et al., 2021). The choice of the number of pseudo-absence points created is important and can greatly influence the model result (Barbet-Massin et al., 2012). “Prevalence” describes the proportion of presence points in all data points (i.e. presence and absence). To gain a suitable prevalence, different presence-to-absence ratios are recommended depending on model algorithms. For example, a 1:10 ratio is recommended for a Generalised Linear Model (GLM) and 1:1 for a Random Forest (RF) algorithm (Barbet-Massin et al., 2012). If there are too many or too few absence points, the predictive power of the models can vary greatly. For presence–absence modelling, another filtering method is necessary to ensure good model quality. “Spatial thinning” ensures that only one occurrence point (as well as pseudo-absence point) is present per raster cell. This reduces spatial bias, prevents model overfitting caused by overrepresentation of duplicate occurrences, and mitigates spatial autocorrelation (discussed in detail below) (Steen et al., 2021).

2.2 Environmental Data

Environmental data of e.g., climate, soil, land cover, and land use can also be sampled *in situ,* but here again datasets available online are most commonly used (Thuiller et al., 2024). Most ENMs concentrate on climatic data. Common climate databases are WorldClim (Hijmans et al., 2005; Fick and Hijmans, 2017) and CHELSA (Karger et al., 2017). Both datasets contain information on current and future climate, as well as paleoclimate data (Last Glacial Maximum) with a resolution of up to 30 arc sec ~ 1 km. While WorldClim is based on interpolated station data and a digital elevation model (Fick and Hijmans, 2017), CHELSA data are generated through dynamic downscaling and reanalysis of ERA5 data (Karger et al., 2017). CHELSA data are more detailed and account for orographic effects, resulting in improved performance in areas with complex topography, such as high mountain regions (Bobrowski et al., 2021a). More recently, remote sensing data have contributed significantly to the development of environmental predictors for ENM. Data from sources such as MODIS, SRTM, or LiDAR provide valuable information on climate, topography, land cover, and spectral metrics (e.g., NDVI), which can be extracted and used in modelling (Wang et al., 2025).

The choice and the number of predictor variables are again critical factors influencing ENM outcomes (Čengić et al., 2020; Sillero et al., 2021). While it may be tempting to include as many abiotic variables as possible, an excessive number of predictors can lead to model overfitting, reducing the model's ability to generalise beyond the training area. A commonly cited rule of thumb suggests that the number of predictor variables should not exceed the number of species occurrence records (Sillero et al., 2021). Other authors provide estimation formulas for example $k = (n - 50)/8$ or $k = n - 104$ (Field et al., 2012) for regression-based methods or general rules like $k = n/10$ (Franklin, 2010), where *k* is the number of predictors and *n* the number of species occurrence records (Sillero et al., 2021). Additionally, multicollinearity must be taken into account. Multicollinearity describes the effect that environmental variables are highly correlated with one another. This effect can result in weak model performance and misleading interpretations (Dormann et al., 2013). It increases the risk of overfitting and may cause variables with no direct ecological relevance to exert a strong influence on model outcomes simply because they are correlated with truly influential predictors (Sillero et al., 2021). Some modelling algorithms are highly sensitive to multicollinearity, while others can handle it more effectively. Nevertheless, collinearity should always be assessed, and variables with correlation coefficients exceeding ± 0.7 to ± 0.8 should generally be excluded (Dormann et al., 2013). Another often-used solution is the transformation of the predictors into principal components and their use in the models (Principal Component Analysis, PCA). However, interpreting principal components is more challenging than interpreting the original variables, as they represent combinations of multiple factors rather than direct environmental gradients (Sillero et al., 2021).

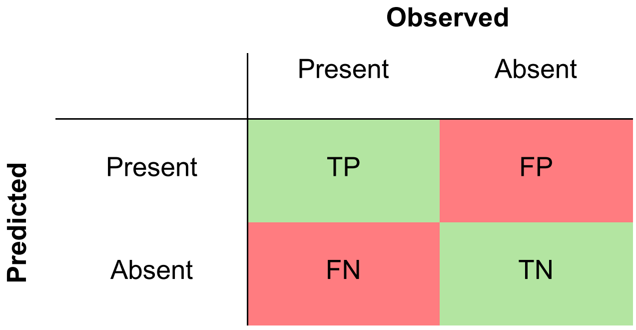
2.3 Modelling Algorithms, Evaluation and Calibration

For ENM, regression approaches like Generalised Linear Models (GLM), Generalised Additive Models (GAM), and Multivariate Adaptive Regression Splines (MARS), or machine learning approaches like Random Forest (RF), Maximum Entropy (MaxEnt), and Artificial Neural Networks (ANN) can be used. The choice of algorithm should be adapted to the data, the size of the study area, and the scope of the study. For example, less complex models tend to perform better in small study areas (Li and Wang, 2013). In linear regression approaches, all predictor-species relationships can be interpreted and plotted in response curves. Therefore, these approaches are considered particularly easy to interpret and suitable for ENM (Li and Wang, 2013). Conversely, machine learning algorithms are often said to be black boxes and difficult to interpret (Breiman, 2001b; Ryo et al., 2021). In comparison, the performance of machine learning approaches is often significantly better than that of linear regression approaches (Li and Wang, 2013). Nevertheless, interpreting the ecological niche, or at least the underlying abiotic conditions, should remain a central focus of any ENM study. However, some authors have criticised the growing tendency to prioritise distribution prediction over ecological interpretation, arguing that the original emphasis on understanding niche dynamics is increasingly being watered down (Bobrowski et al., 2021b).

Prior to model computation, the final dataset, consisting of species occurrence and environmental data, is split into training and testing subsets for model validation, commonly using a cross-validation approach. Best practice in model validation would be to test the model on independent, external data. But as these are rarely available, standard cross-validation based on data splitting is commonly applied. The model is trained on the training dataset and evaluated using the testing dataset (Yates et al., 2023). There are several methods of cross-validation, with *k*-fold cross-validation being the most commonly applied. In this approach, the dataset is divided into *k* equal subsets (folds). In each iteration, one subset is used for testing, while the remaining $k - 1$ subsets are used for training. This process is repeated *k* times, and model performance is evaluated across all *k* iterations (Yates et al., 2023). Another possibility is the leave-one-out method, where each individual data point is used once as test data while all remaining points serve as training data ($k = n$) (Yates et al., 2023). Training and testing datasets should be as independent as possible to allow for meaningful model validation (Sillero et al., 2021). However, spatial data is always spatially autocorrelated to a certain degree, which poses additional challenges for achieving full independence between training and testing sets (Legendre, 1993; Dormann et al., 2007; Roberts et al., 2016). Tobler's law applies here, that says: “[...] everything is related to everything, but near things are more related than distant things.” (Tobler, 1970). In this context, this means that abiotic conditions at closer species locations tend to be more similar than more distant ones, so that autocorrelation is more pronounced (Sillero et al., 2021). In random cross-validation splits, spatial autocorrelation is not accounted for, which can lead to overly optimistic model evaluations. Therefore, a third cross-validation method must be mentioned, the spatial or block cross-validation (Roberts et al., 2016). This approach divides the data into spatially distinct blocks, creating folds that are geographically separated, thereby reducing spatial autocorrelation between training and testing sets and providing a more realistic assessment of model performance (Roberts et al., 2016; Valavi et al., 2019). To summarize, cross-validation is a method used to test the predictive performance of the model with data that is as independent as possible. In addition, cross-validation can be used to optimise model hyperparameters by testing models with different settings, for example, testing different numbers of trees (ntree) or tested variables at each node (mtry) in a RF model (Schratz et al., 2019; Yates et al., 2023). After model validation, either the best-performing model (i.e. the one with the highest predictive performance and optimal hyperparameters) or an average model (based on the mean validation metrics across all runs) is selected and applied for model transfer in space or time (Yates et al., 2023).

Evaluation metrics depend on the data type. For regression approaches with binary data (presence data has two expressions, 1 = presence and 0 = absence) commonly applied metrics include the threshold-dependent Cohen's Kappa and True Skill Statistic (TSS), as well as the threshold-independent Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve (Mouton et al., 2010; Sillero et al., 2021). In binary modelling, a threshold must be set to determine when a predicted value is considered a presence and when it is considered an absence (e.g., presence ≥ 0.5). The model results can then be displayed in confusion matrices, as illustrated in Table 1 (Miller, 2010; Sillero et al., 2021). From this confusion matrix, sensitivity, the proportion of correctly predicted presences, and specificity, the proportion of correctly predicted absences can be derived, which in turn can be used to calculate various evaluation measures (Miller, 2010). Cohen's Kappa is a commonly used metric that measures the agreement between predicted and observed values. However, because it is highly sensitive to prevalence, the TSS is now more widely used as a more robust alternative (Mourton et al., 2010). The TSS is calculated as $\text{sensitivity} + \text{specificity} - 1$ and takes values from −1 to +1. Values close to +1 indicate perfect agreement between predicted and observed values and thus reflect high model performance, whereas values near or below 0 indicate performance no better than random guessing (Allouche et al., 2006). The choice of threshold plays an important role, as it strongly influences the result and model performance (Liu et al., 2013). The AUC is independent of thresholds and is therefore often used to evaluate prediction models (Peterson et al., 2008). The ROC curve, on which the AUC is based, plots sensitivity (y-axis) against the false positive rate (x-axis) for all possible thresholds. The AUC value describes the "area under this curve" and can take values between 0 and 1. An AUC of ≤ 0.5 means that the model is no better than random guessing (Peterson et al., 2008). Other authors recommend likelihood-based criteria such as Akaike information criterion (AIC) and the coefficient of determination (R^2) (Lawson et al., 2014). For continuous data (e.g., abundance data, variable is numerical), simple correlations between observed and modelled values can be computed, or evaluation metrics such as R^2 and the Root Mean Square Error (RMSE) can be used to assess model performance (Potts and Elith, 2006).

Table 1: Typical structure of a confusion matrix comparing observed and predicted values, thereby distinguishing True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). This forms the basis for calculating several model evaluation metrics (after Miller et al., 2010).



		Observed	
		Present	Absent
Predicted	Present	TP	FP
	Absent	FN	TN

After model setup, calibration, and validation, an ENM can be transferred both spatially and temporally, unlike a traditional SDM (Anderson, 2012, Peterson and Soberón, 2012). Typically, the model is first projected across the entire study area to estimate the full potential distribution of the species (Anderson, 2012). However, it can also be transferred to remote geographic regions, for example, to assess the potential spread of invasive species in novel environments (Elith and Leathwick, 2009; Anderson, 2012). Moreover, ENMs can be projected under future or past climatic conditions. This requires climate data with temporal projections (Sillero et al., 2021). Future projections are usually based on Representative Concentration Pathways (RCPs), which are categorised into mild, medium, and severe emission scenarios (van Vuuren et al., 2011). The most commonly used sources for these datasets are again WorldClim and CHELSA, both of which provide datasets in which RCP scenarios have been modelled using a variety of global climate models.

3. *Nothofagus pumilio*: Climate Change and Impacts on the Alpine Treeline

The Andes are the largest mountain range with a latitudinal extent covering over 8,000 km. After the Himalayas, they are the second-highest mountain range in the world with peaks up to 6,962 m (Cerro Aconcagua, Argentina) (Graham, 2009). Due to the high complex topography and the north-south exposition, the Andes comprise of steep climatic gradients leading to high heterogeneity of habitats. This results in a very high floral diversity, with 10 % of the world's vascular plant species in the Andes, at only 0.6 % of the global land surface (Mittermeier et al., 2011). Biodiversity is highest in northern mid-elevation cloud forests, while biomes of the Southern Andes are less diverse (Pérez-Escobar et al., 2022). Nevertheless, the Andean biodiversity hotspot is threatened by climate change in its northern and southern parts (Tovar et al., 2022). High mountain regions like the Andes are particularly affected by climate change, as warming is above the global average due to snow albedo feedback mechanisms (Pepin et al., 2015). Glaciers and snow fields are retreating and bare ground with low albedo favours warming of surrounding areas (Pepin et al., 2015; Pepin et al., 2022). Globally, an average increase in air temperatures of 0.2°C is observed per decade since 1880 with an ongoing positive trend (Allen et al., 2018). In high mountain regions, temperature increases ranging from of 0.3°C to 0.4°C per decade (Schickhoff et al., 2022). This above-average warming forces ecosystems to alter in terms of their structure and species composition, leading to habitat shifts or even their loss due to new environmental conditions (Dirnböck et al., 2011; Dullinger et al., 2012). An especially visible and documented change is the shift of alpine treeline positions (Holtmeier and Broll, 2020).

This study focuses on the Southern Andes, specifically the Andean Cordillera between 35°S and 56°S, extending to the southern tip of South America (Tierra del Fuego), a region often referred to as Patagonia in the literature. The Southern Andes reach a maximum altitude of up to 4058 m at 46°S (Borsdorf and Stadel, 2013). The altitude itself creates a steep temperature gradient and the north-south exposition in combination with the west-wind regime creates a precipitation gradient, which is considered to be the most extreme precipitation gradient of the world (Garreaud et al., 2013). These climatic gradients directly influence the distribution range of the treeline species *N. pumilio*, as well as changing conditions at the treeline in response to climate change. This chapter gives an overview of the study area, its climate and climate change implications as well as the study species *N. pumilio* and the changes of the species' distribution at treeline locations.

3.1 Climate and Climate Change in Southern South America

The climate in the Southern Andes is temperate, adjacent to mediterranean climate in the north and more cold-temperate climate at the southern tip of the South American continent (Borsdorf and Stadel, 2013). One special feature, however, is that the climate of the western and eastern side of the Andes is extremely different, which is due to the extreme precipitation gradient. Climate is maritime and hyper-humid at the westside of the Andes and more continental and arid at the east of the mountain range (Garreaud et al., 2013). Figure 2 shows a Köppen-Geiger climate classification map and the steep west-east precipitation gradient in the study area.

The Southern Andes are located between two pressure bands: the semi-permanent anticyclones of the Pacific in the north-west and the subpolar low-pressure belt in the south (Antarctic Oscillation) (Paruelo et al., 1998). These create strong and constant west winds that transport humid air masses from the Pacific. Through orogenic uplift, precipitation sums are very high at the western side of the Andes, reaching extremes from 5,000 up to 10,000 mm/a (Garreaud et al., 2013). Since most of the precipitation falls on the Chilean side of the Andes and dry air descents on the Argentinean side of the Andes, this causes high evaporation and a precipitation deficit of below 300 mm/a at the lee side of the mountain range (Garreaud et al., 2013). Seasonal changes in precipitation are due to shifts in the pressure belts. A northward shift of the Antarctic Oscillation (AAO) is associated with increased storm activity and colder, wetter conditions in the mid-latitudes, especially in winter, whereas its southward shift results in slightly lower precipitation sums in spring and summer (Holz et al., 2017). Temperatures are relatively cold due to the Antarctic-Circumpolar- and Humboldt-Current that transport cold water masses from the Antarctic (Schneider et al., 2003). Mean annual temperatures range from 12°C in the north-east to 3°C in the south of the study area due to the latitudinal gradient. Furthermore, the altitude gradient and cold winds influence local temperature patterns (Paruelo et al., 1998).

Interannual variability of temperature and precipitation in the study area are significantly influenced by expressions of large-scale phenomena like the El Niño Southern Oscillation (ENSO), La Niña events and the Antarctic Oscillation (AAO) also knows as Southern Annular Mode (SAM). Furthermore, (multi-) decadal variability is due to expressions of the Pacific Decadal Oscillation (PDO) (Garreaud, 2009). The AAO is the dominant climatic pattern influencing the extratropical regions of the Southern Hemisphere (Garreaud, 2009; Veblen et al., 2011). As outlined above, its northward and southward shifts significantly affect temperature and precipitation patterns in southern South America (Holz et al., 2017). Anomalies involving strong poleward shifts of the AAO have a significant impact on regional precipitation, leading to weakened westerlies and to drought conditions in northern Patagonia (Holz et al., 2017). ENSO events are non-cyclical reversals of the Walker circulation, and are normally associated with very high precipitation amounts and higher (water) temperatures in tropical and subtropical South America (Cai et al., 2020). But in the Southern Andes ENSO events create warm-dry conditions also leading to droughts as the westerlies are weakened especially in spring and summer month. La Niña events on the opposite create cool and rainy summers (Garreaud, 2018). The PDO refers to long-term water surface temperature changes in the Pacific that last over approximately 20 years and lead to (global) temperature rises. A negative-to-positive shift in PDO leads to increased spring and summer temperatures in the Southern Andes, while a positive-to-negative shift in PDO creates dry and cold conditions (Srur et al., 2016).

Temperature increase in southern South America is mostly associated with the positive phase (poleward) shift of the AAO. Since the latter half of the 20th century, an increased positive trend in the AAO has been observed, attributed to increased greenhouse gas concentrations and reduced stratospheric ozone in the atmosphere (Miller et al., 2006, Veblen et al., 2011). Therefore, it is most likely that the positive trend, associated with rising temperatures and regional drought conditions, persists in the 21st century (Fyfe and Saenko, 2006; Veblen et al., 2011). Decreases in precipitation can also be attributed to the AAO and ENSO, as described above. However, in addition to variations caused by phase shifts of AAO, ENSO, and PDO, trends attributable to anthropogenic climate change are evident for both, temperatures and precipitation. Since the 1950s, temperatures in northern Patagonia have risen by +1°C and precipitation has decreased by −5 % (Pessacg et al., 2020). Climate modelling using CMIP5 models in central Chile (30°S to 40°S) indicates a temperature increase of +1.2°C and a decrease in precipitation of −3 % by the end of the century in a mild scenario (RCP2.6). For the RCP8.5 scenario, temperatures increase by +3.5°C and precipitation decreases up to −30 % (Bozkurt et al., 2018). A regional CORDEX model calculates an increase of +1.5°C for annual mean temperature and a −10 to −30 % decrease in precipitation for northern Patagonia for the same period (Pessacg et al., 2020). Chile is currently experiencing its worst droughts since records (Garreaud et al., 2020), with the most severe drought occurring in 2016 (Garreaud et al., 2018). For southern Chile, Bambach et al. (2022) find warming trends in the Andes, as well as drying trends in the austral summer and winter, and even emphasise that the Andean tundra and alpine climates (classification according to Köppen and Geiger) will decline dramatically by the end of the century (Bambach et al., 2022). Climate change in the Southern Andes has a strong influence on snowfall and snow persistence, glacier thickness (Masiokas et al., 2020), water availability (Masiokas et al., 2020; Pessacg et al., 2020), droughts (Garreaud et al., 2017; Garreaud et al., 2020), and fire frequency (Holz et al., 2017; Mundo et al., 2017). Ultimately, ecosystems will be forced to adapt to new climatic conditions, resulting in habitat shifts and species extinction, particularly in vulnerable high mountain regions. The effects of climate change also have a strong impact on the study species *N. pumilio* at the alpine treeline.

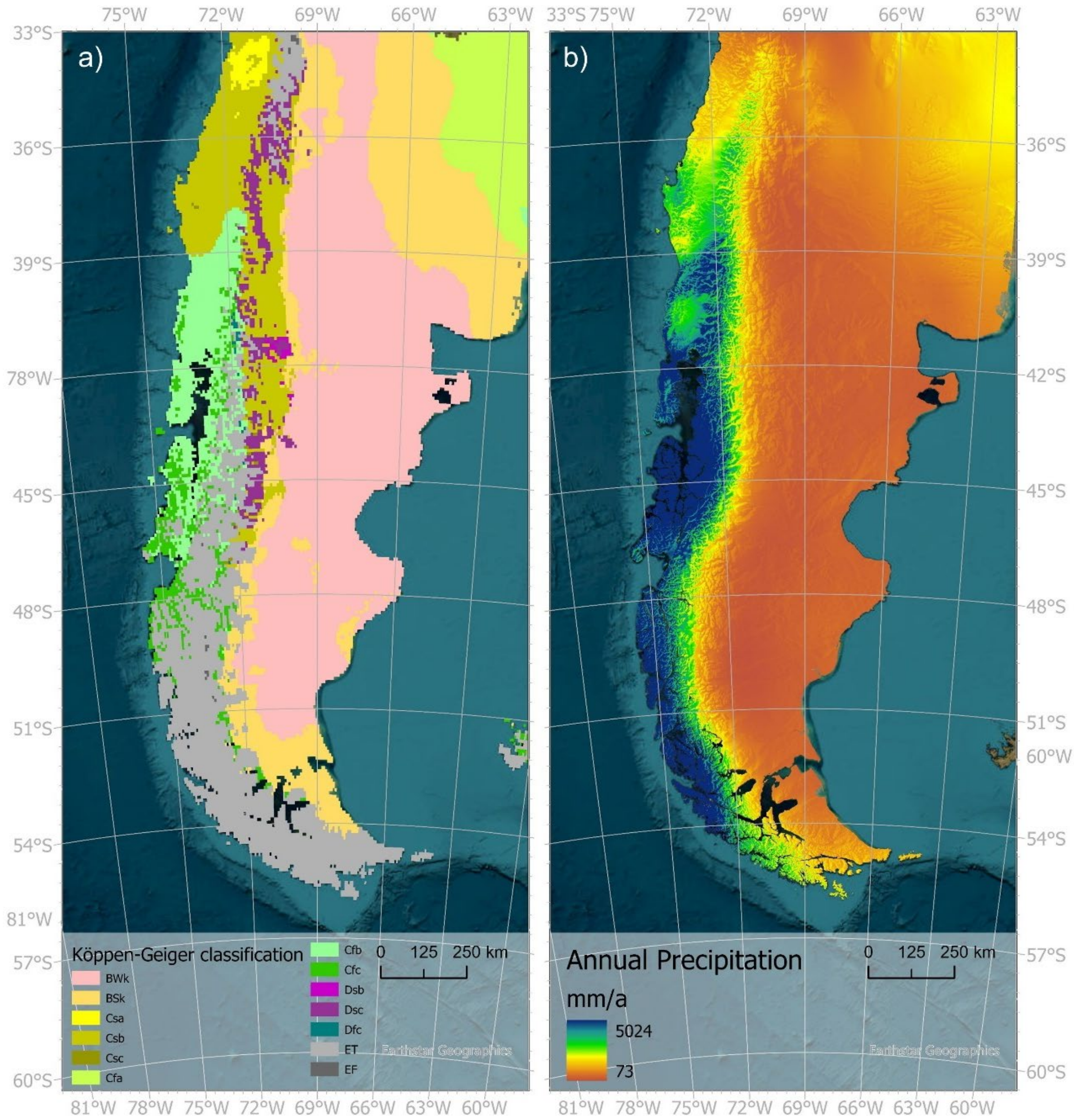


Figure 2: Study area in southern South America showing a) a climate classification map according to Köppen and Geiger, with climate types ranging from ***BWk*** *(Arid, desert, cold),* ***BSk*** *(Arid, steppe, cold),* ***Csa*** *(Temperate, dry summer, hot summer),* ***Csb*** *(Temperate, dry summer, warm summer),* ***Csc*** *(Temperate, dry summer, cold summer),* ***Cfa*** *(Temperate, no dry season, hot summer),* ***Cfb*** *(Temperate, no dry season, warm summer),* ***Cfc*** *(Temperate, no dry season, cold summer),* ***Dsb*** *(Cold, dry summer, warm summer),* ***Dsc*** *(Cold, dry summer, cold summer),* ***Dfc*** *(Cold, no dry season, cold summer),* ***ET*** *(Polar, tundra) to* ***EF*** *(Polar, frost) (data from Beck et al., 2023) and b) the steep precipitation gradient from the humid west to the arid east (Bioclim variable 12, annual precipitation sums, Karger et al., 2017).*

3.2 Treeline Species *Nothofagus pumilio*

Nothofagus pumilio (Poepp et Endl.) Krasser (southern or lenga beech) belongs to the *Nothofagaceae* family and is one of ten *Nothofagus* species on the South American continent (Amigo and Rodríguez-Guitián, 2011). The genus *Nothofagus* developed on Gondwana, with the first pollen evidence dating back to the late Cretaceous period (83 to 70 million years BP) (Manos, 1997). There are 36 *Nothofagus* species worldwide, which have diversified with the continental drift of Gondwana to the present-day continents in disjunct areas with a high degree of endemism in South America, New Zealand, Australia, and Tasmania (Manos, 1997; Premoli et al., 2012). The South American species survived the ice ages by migrating to more northern areas or to alpine refuges in the Andes (Soliani et al., 2015; Premoli et al., 2024). The ten recent South American *Nothofagus* species inhabit climates ranging from mediterranean to temperate to cold-temperate conditions (Amigo and Rodríguez-Guitián, 2011) and represent 80 % of the Patagonian-Andean forests (Veblen et al., 1996; Varela et al., 2010). Most species are deciduous (seven species), and three species are evergreen. The orophilic species *N. pumilio* and *N. antartica*, which are particularly adapted to cold conditions, are deciduous (Amigo and Rodríguez-Guitián, 2011).

N. pumilio is the most widespread endemic species in the southern Chilean and Argentinean Andes and forms pure forest stands at the alpine treeline from 35°S to 56°S. The species is the dominant subalpine tree species and bioindicator for the oro-temperate belt following the southern Andean Cordillera over a latitudinal range of 2000 km (Amigo and Rodríguez-Guitián, 2011). Along the latitudinal temperature gradient, the distribution area ranges from very high elevation stands in the north at 2000 m to forest stands at sea level at Tierra del Fuego (Lara et al., 2005). The eastern and western distribution area is defined by precipitation levels. The species does not occur in low-lying areas in the hyper-humid west, where the species *N. betuloides* becomes dominant (Young and León, 2007; Amigo and Rodríguez-Guitián, 2011). In the east, low precipitation levels define the distribution limit towards the Patagonian steppe. The species follows the forest-steppe ecotone and often forms two treelines, a common upper and a lower, xeric treeline towards the arid region (Hertel et al. 2008). *N. pumilio* mostly occurs in pure forest stands especially south of 39°S, but it also forms mixed stands in the north with *Australocedrus chilensis* and *Araucaria araucana* (Hildebrand-Vogel et al., 1990). In the north-west and further south, adjacent or overlapping habitats with other deciduous and evergreen *Nothofagus* species are possible. Between 35°S and 40°S, the deciduous species *N. alpina* occurs from the coast up to a maximum altitude of 1200 m (Pollmann, 2001), as well as deciduous *N. obliqua* in most oceanic areas (Amigo and Rodríguez-Guitián, 2011). In the (hyper-) humid west, evergreen *N. dombeyi* and *N. nitida* accompanies *N. pumilio* up to 47°S and *N. betuloides* all the way south to Tierra del Fuego (Amigo and Rodríguez-Guitián, 2011). The deciduous species *N. antarctica* accompanies *N. pumilio* across almost its entire distribution range and occupies niches at high elevations in cold depressions, in locations with poor soil conditions, and towards the arid east (Veblen et al., 1996; Amigo and Rodríguez-Guitián, 2011). While the other deciduous species are morphologically and ecologically very different from *N. pumilio*, *N. antarctica* is very similar, and even hybrids are possible (Soliani et al., 2015).

N. pumilio is characterised by its dark green, elliptical, and notched broad leaves which turn orange-red in autumn. Morphologic characteristics can be seen in Figure 3. The tree species has high phenotypic plasticity. While erect trees can reach a height of up to 35 m under optimal growing conditions, shrub-like krummholz growth occurs at the treeline (Stecconi et al., 2010). The species is monoecious, anemophilic and not self-pollinating, pollen and fruit are spread by wind. Every six to eight years, there are so-called mast fruiting years, during which a particularly large number of fruits are produced (Cuevas, 2000). Seedling stage is the most critical life stage. Seedlings germinate in biological substrate mostly in the protecting forest stand climate, which is an important limiting factor for the advance of the treeline (Batllori et al., 2009; Bürzle et al., 2018). As orophilic species, *N. pumilio* is particularly adapted to the harsh environmental conditions in high mountain regions. At treeline sites under extreme conditions, mean growing season soil temperatures correspond to the 6.6°C isotherm (Fajardo and Piper, 2014). *N. pumilio* can tolerate cold air and soil temperatures and is well adapted to a short growing season (Premoli et al., 2007). The species protects itself from the cold by concentrating carbohydrates in the xylem and thus lowering the freezing point (Fajardo and Piper, 2014). An effective adaptation to the risk of frost drought is deciduousness, as this minimises transpiration during the critical spring months when water uptake is limited (Körner, 2021). Due to advective precipitation, lee effects, and high radiation in combination with high evapotranspiration, water availability is limited (Bach and Price, 2013). The species is adapted to medium to low annual precipitation levels of 400 to 500 mm, most of which falls as snow (Veblen et al., 1996). This also explains the species distribution to the arid east (Hertel et al., 2008). Leaf size and shoot growth is significantly reduced in high elevation individuals in comparison to low elevation trees. This might be an adaptation to high wind speed and high radiation (Premoli et al., 2007). *N. pumilio* occurs on Andisols formed from volcanic ash (Hildebrand-Vogel et al., 1990). These soils are nutrient-rich, but the high content of phosphorus and potassium is largely unavailable to plants, leading to the conservation of organic matter, further slowing down the already delayed soil formation in high mountain regions (Hildebrand-Vogel et al., 1990, Premoli et al., 2007). Therefore, *N. pumilio* grows and germinates preferentially in areas with an herb layer of for example *Empetrum rubrum* stands, that provide nutrients, especially nitrogen, through organic matter (Pissolito, 2016).

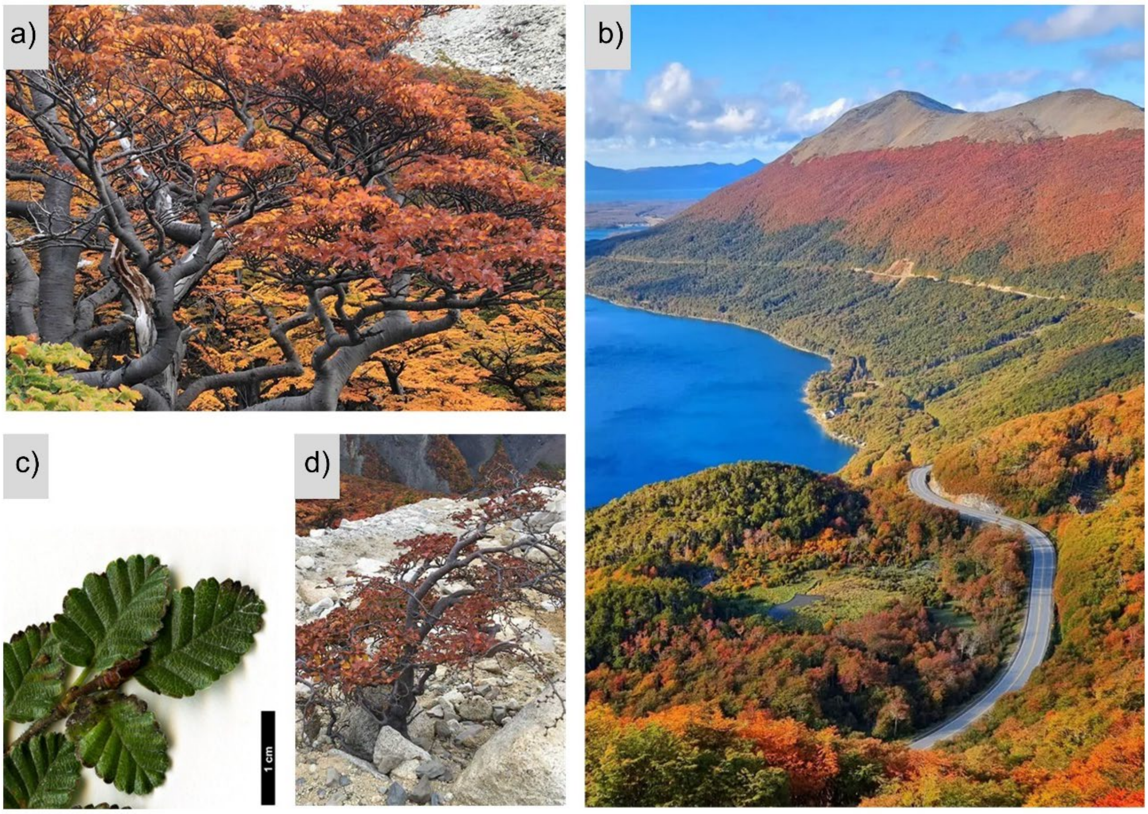


Figure 3: Morphological and ecological features of Nothofagus pumilio. a) Autumn-coloured tree crown of a mature individual (used with permission by Instagram user amb6767, 2018), b) abrupt treeline seen from the Paso Garibaldi scenic viewpoint at Lago Escondido (used with permission by Instagram user sebaingles, 2022), c) green leaves in detail (De Langhe, 2015), d) wind-crooked individual above the treeline in krummholz growth form (used with permission by Instagram user amb6767, 2018).

3.3 Changes at the Treeline of *Nothofagus pumilio*

The treeline formed by *N. pumilio* is abrupt. Abrupt treelines are the result of unsuccessful seedling establishment and dieback processes outside the protecting forest stand (Harsch and Bader, 2011; Bader et al., 2021), which can be attributed to very cold temperatures (Fajado and Pieper, 2014), a prolonged snow cover (Holtmeier and Broll, 2005), strong, cold winds, and wind shearing (Rebertus et al., 1997). Furthermore, steep topography, talus slope/blockfields, and shallow soil depth prevent the establishment of seedlings above the treeline (Hadley et al., 2013). Seedlings preferably establish in developed soils covered with an herb layer (Pissolito, 2016; Srur et al., 2018). Figure 3d visualises the harsh conditions above the abrupt treeline. A wind-crooked *N. pumilio* individual has established in the lee of larger rocks, where wind speed is reduced and substrate and soil can accumulate. At the same time, it shows the barren ground and inhospitable conditions at the treeline at Torres del Paine National Park. The treeline elevation ranges from 2000 m to 1600 m in the northernmost distribution area of *N. pumilio* to 400 m at Tiera Del Fuego (Cuevas, 2000; Lara et al., 2005). Therefore, the treeline elevation decreases almost constantly by 60 m per 1° latitudinal dispersion (Lara et al., 2005). Treeline elevation also varies longitudinal due to the precipitation gradient. At the humid western site of the Southern Andes, the treeline position is lower than at dry eastern slopes (Daniels and Veblen, 2003). But at wetter sites, tree density of newly established trees is higher than at mesic and dry sites (Srur et al., 2016).

Generally, treeline elevation of abrupt treelines is lower than expected (Harsch and Bader, 2011) and unlike diffuse treelines they are less likely to shift to higher elevations (Harsch et al., 2009). Nevertheless, last decades research found, that *N. pumilio* is highly sensitive to climate variations, which are often associated with phase shifts in AAO, ENSO, and PDO, as reflected in its radial growth patterns and seedling establishment above the treeline (Lara et al., 2001; Aravena et al., 2002; Daniels and Veblen, 2004; Masiokas and Villalba, 2004; Álvarez et al., 2015; Srur et al., 2016, 2018; Brand et al., 2022; Reiter et al., 2024). Generally, higher temperatures can lead to increased radial growth of trees and the establishment of seedlings above the current treeline elevation (e.g., Srur et al., 2018; Reiter et al., 2024). However, these effects are highly dependent on the precipitation regime (Brand et al., 2022; Srur et al., 2020). There are differences between northern and southern Patagonia, as northern Patagonia is more frequently affected by drought events (Rodríguez-Catón et al., 2016; Garreaud et al., 2020). However, the most pronounced contrasts are found between regions with differing precipitation regimes along the west-east gradient (Brand et al., 2022). Radial growth in humid and mesic locations is directly influenced by temperatures (Brand et al., 2022). In these regions, higher temperatures lead to increased radial growth. Higher spring and summer temperatures, in particular, have a positive effect on growth, as snow melt begins earlier thus extending the growing season. In dry locations, however, water availability plays a major role. Here, growth decreases due to increased evapotranspiration and associated drought stress (Brand et al., 2022). Prolonged dry conditions not only reduce growth but can also lead to increased mortality (Rodríguez-Catón et al., 2016). Reiter et al. (2024) even emphasise that treeline limitation changes from being primarily cold-limited to drought-limited.

The situation is similar with regard to the establishment of seedlings above the treeline. Higher temperatures lead to an increased occurrence of seedlings in humid, mesic, and dry locations. However, the density of seedlings is lower in mesic and dry locations (Srur et al., 2016) because sufficient water availability promotes the survival of seedlings (Srur et al., 2020). Studies show that the establishment of seedlings is primarily related to warmer spring and early summer temperatures, and the associated earlier snowmelt (Srur et al., 2016). Abrupt changes in establishment can be attributed to phases of the PDO. A negative-to-positive shift in the 1970s led to warmer spring and summer temperatures and thus to an abrupt increase in establishment, while the positive-to-negative shift from the 1990s onwards led to lower temperatures and a decrease in establishment above the treeline (Srur et al., 2016). A positive phase of the AAO leads to higher temperatures in the entire study area and is therefore also positively correlated with seedling establishment (Srur et al., 2018). Although drought poses a challenge for seedlings, it does not affect seeds or seed production (Aschero et al., 2022). Seed viability and fruit dispersal are essential prerequisites for seedling emergence and establishment above the treeline. While seed production, fruit dispersal, seed viability, and seedling establishment all decline with increasing elevation (Cuevas, 2000), both the quality and quantity of seeds tend to improve under higher temperatures, an effect attributed to ENSO and the AAO (Rodríguez-Souilla et al., 2024a; Rodríguez-Souilla et al., 2024b).

Even though changes in AAO, ENSO, and PDO cause variations in radial growth, seed quantity and quality, as well as seedling establishment, long-term trends over 100 years can be observed (Brand et al., 2022). With an average temperature increase of 0.54 °C over the last 100 years, tree growth has increased in mesic and humid locations, while it has decreased in dry locations. This trend will further continue in the 21st century (Brand et al., 2022). The general trend for seedling establishment indicates that warm temperatures favour the establishment of new individuals above the treeline, while abundant precipitation support their survival (Srur et al., 2020). Despite the abrupt nature of the Southern Andes' treeline, climate change could generally cause it to shift to higher elevations. However, it is important to note that various abiotic and biotic factors might slow this process. Observations suggest that treeline advance is limited to a range of 5 to 10 m from closed stands, as the seeds of *N. pumilio* are relatively heavy and dispersed by wind only over short distances (Srur et al., 2018). In addition, low soil depth and a lack of herbaceous vegetation can further slow down the process (Pissolito, 2016). Studying the conditions at the treeline is therefore of particular relevance for understanding and predicting potential changes under climate change.

3.4 Study Objectives and Research Questions

N. pumilio is sensitive to climate variations, making it an especially interesting research object to investigate changes at the alpine treeline due to climate change. Local studies investigating changes to *N. pumilio* are mostly restricted to local experimental or dendrochronological studies. Large-scale modelling studies investigating the entire current and future distribution of the species are currently lacking, which can be attributed to a lack of species occurrence data covering the full range of the species.

As described in the previous chapters, species data are generally scarce in areas that are difficult to access or even inaccessible, such as high mountain regions, and may be spatially biased towards urban or tourist areas. Remote sensing, along with innovative approaches such as species occurrence data sampling from social media, holds great potential for addressing these issues. Accordingly, the first main objective of this thesis was to introduce a novel Instagram ground truthing approach (IGTA), which combined the strengths of both, social media and remote sensing data. Using Instagram, one of the most widely used social media platforms worldwide, offered a new possibility of creating less-biased point occurrences of *N. pumilio.* These were subsequently used to validate the large-scale raster dataset of the species (ground truthing), which is generally free from the typical spatial biases found in Citizen Science and social media-based datasets.

This led to two main research questions:

1. Can Instagram be used to obtain occurrence data for *Nothofagus pumilio*, and is the spatial bias in the resulting IGTA dataset reduced compared to that in commonly used GBIF data?
2. Can supervised classification be used to generate raster occurrence data for *Nothofagus pumilio* across its entire range?

The IGTA resulted in two species datasets: a binary point dataset and a continuous raster dataset, both of which can be used for modelling. The point dataset enabled a typical presence–absence approach, while the continuous raster dataset allowed for a more complex modelling strategy. To enable direct comparison between both approaches, the Random Forest algorithm, suitable for both data types, was applied to model the potential current and future distribution of the treeline species. This not only addressed the existing research gap concerning treeline modelling studies in the Southern Andes but also allowed for a comparison of two modelling approaches and an evaluation of the potential benefits of using continuous raster data over binary point data. This raised three further research questions to be answered:

3. How can the ecological niche of *Nothofagus pumilio* be characterised under current climatic conditions, and what is the species' potential geographic distribution?
4. What potential range shifts are projected under future climate change scenarios?
5. What are the differences between the modelling approaches used in this study, and what are their respective advantages and disadvantages?

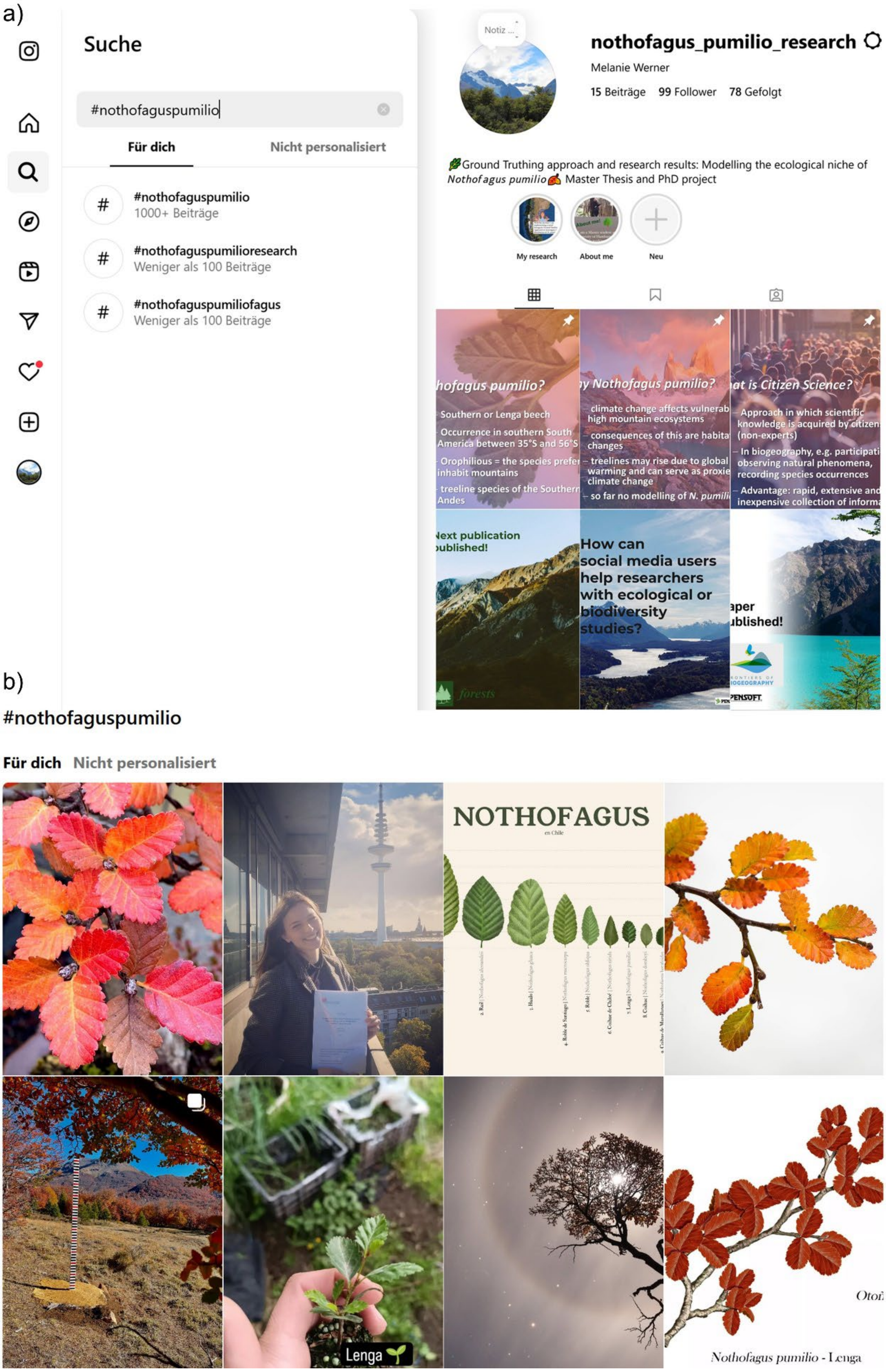
4. Improved Ecological Niche Modelling of *Nothofagus pumilio* – Methodological Overview

ENM can provide valuable insights of climatic conditions shaping treeline patterns in the Southern Andes. The PhD project "Improved Ecological Niche Modelling of *Nothofagus pumilio* in the Southern Andes" aims at modelling the current and future potential distribution of the study species *N. pumilio*. Limitations in modelling were discussed and addressed with innovative methods with a focus on creating two species occurrence datasets developed by an Instagram ground truthing approach and applying them in an ENM workflow. In Publication I, the IGTA was developed, and the resulting data was made publicly available on the open-access data provider of University of Hamburg (Publication II). Two species datasets resulted from the IGTA, a point dataset and a raster dataset, both of which can be used for modelling the ecological niche of the species. The point dataset enabled a conventional binary presence–absence model approach with randomly created pseudo-absences, but the more complex and continuous raster dataset enabled a new modelling opportunity. The bioclimatic variables from CHELSA (Bioclim, version 2.1) were used as predictors. To download and preprocess these variables, the R package "ClimDatDownloadR" was employed, which detailed description is Publication III. Both models were subsequently tested, compared, and used to analyse the ecological niche, and to predict the potential current and future distribution of the species, as presented in Publication IV.

4.1 The Instagram Ground Truthing Approach

The availability of unbiased species occurrence data is a major challenge in the field of ENM (Chauvier et al., 2021). Comprehensive field studies that could create unbiased data are costly, time-consuming and areas like high mountain regions are difficult or even impossible to access. Therefore, data is mostly used from Open-Source databases, which may contain errors of unknown magnitude, as described before (Beck et al., 2014; Meyer et al., 2016). A promising approach to mitigate spatial bias is the use of remote sensing data. Spatial data from remote sensing methods can be used to create large-scale and areal data and record occurrences in inaccessible areas. But there is still a need of a ground truthing process, the validation that a species actually occurs in the areas investigated with remote sensing (Nagai et al., 2020). In Publication I, the innovative IGTA was developed, in which sampling and spatial bias were reduced in two ways: first, by creating occurrence points from a worldwide used social media platform in a stratified approach, and second, by using this dataset for ground truthing of remote sensing data created using supervised classification.

Instagram is one of the largest social media platforms worldwide, with more than 2 billion users and over 60 million daily uploads (WirtschaftsWoche, 2014; Statista, 2025). On Instagram, only pictures and short videos (Reels) can be posted. The user's content is listed chronologically on their profiles. A search bar allows users to search for profiles, locations and hashtags. Hashtags and location tags can link posts to specific subjects or places (e.g., #nothofaguspumilio or #torresdelpaine) and can thus be found and analysed in a large number. Figure 4a shows the Instagram interface with the profile "nothofagus_pumilio_research" (www.instagram.com/nothofagus_pumilio_research/), which was created specifically for this study. The typical chronologically sorted post view is visible on the right, along with the search bar on the left side. Figure 4b illustrates the results of a search process for the hashtag #nothofaguspumilio within the Instagram user interface. By searching for different hashtags and location tags, it is not only possible to find photos taken by experts linking the species name, but also photos taken on hikes, while skiing, on a tourist trip, or on everyday photos, where the occurrence of the species was documented unintentionally. This increases the number of suitable posts even more. With a strict set of criteria (Table 2), posts were selected on which the species and the location where the photo was taken were clearly identifiable. For manual georeferencing of the occurrences with SAGA GIS (Conrad et al., 2015), landscape elements visible in the photo had to be recognisable in satellite images. Such landscape elements were, for example lakes, rivers, typical mountain ranges and peaks, glaciers, buildings, and infrastructure. Due to the ecology of the species, occurrence points were not only set at the sampling location but also in the visible background of the photo. If the treeline was clearly autumn coloured and abrupt, occurrence points were set as far as the neighbouring valley. The distance between background points was set to 1 km, corresponding to the target resolution used in the modelling. Figure 5 illustrates the post selection and point transfer into a map. The high quantity of posts and the background points have the potential to effectively reduce spatial bias. To quantify this, a bias analysis was conducted using the R package "sampbias" (Zizka et al., 2021). By incorporating geodata representing urban features, such as shapefiles of cities, roads, rivers, and lakes, the relationship between species occurrences and their distance to urban areas was calculated. This analysis produced both quantitative results about the estimated sampling rate and a visual bias map.


a)

Suche

#nothofaguspumilio

Für dich | Nicht personalisiert

- #nothofaguspumilio – 1000+ Beiträge
- #nothofaguspumilioresearch – Weniger als 100 Beiträge
- #nothofaguspumiliofagus – Weniger als 100 Beiträge

nothofagus_pumilio_research

Melanie Werner

15 Beiträge 99 Follower 78 Gefolgt

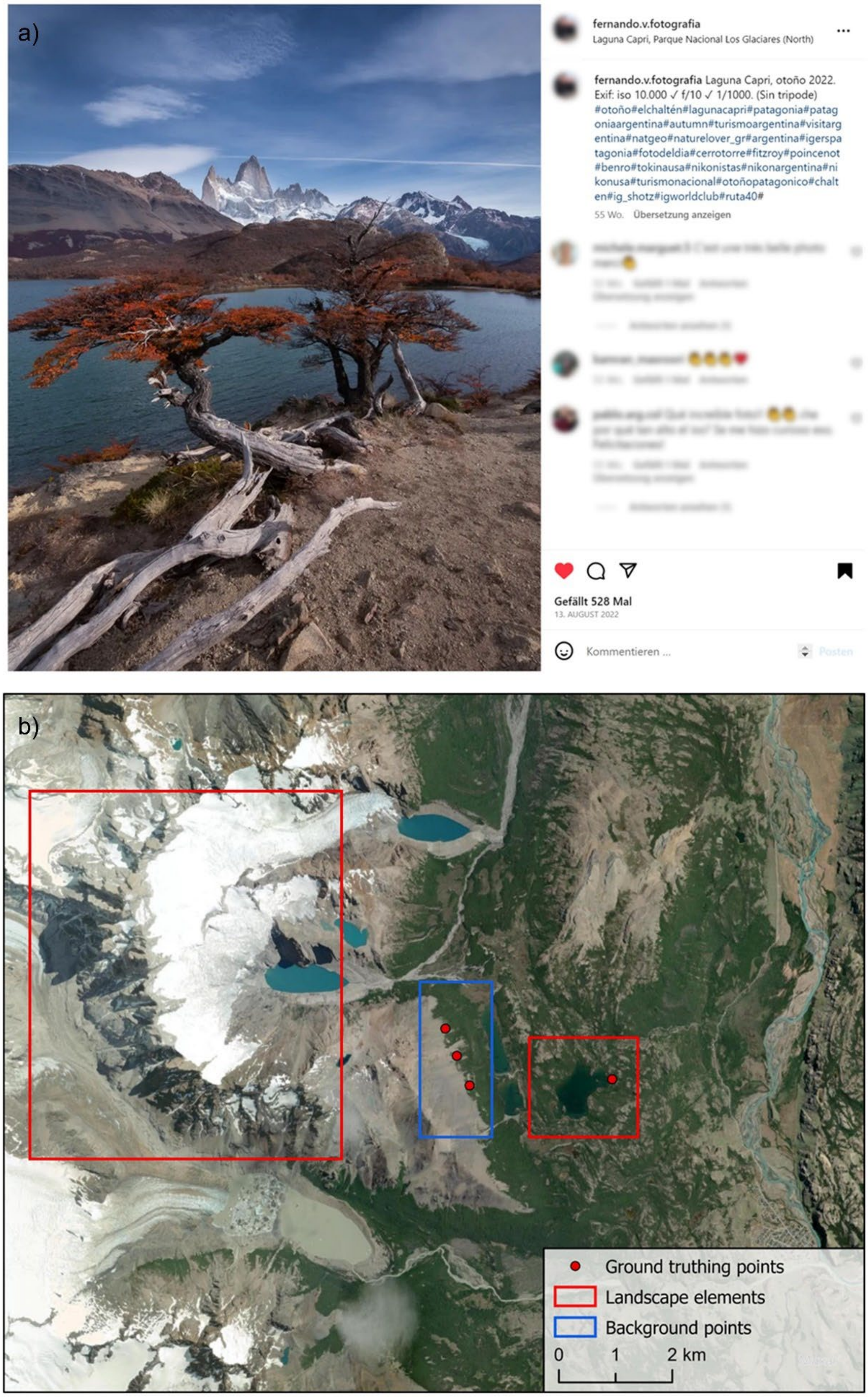
Ground Truthing approach and research results: Modelling the ecological niche of *Nothofagus pumilio* Master Thesis and PhD project

b)

#nothofaguspumilio

Für dich Nicht personalisiert

Figure 4: Screenshot of a) the Instagram profile “nothofagus_pumilio_research”, which was generated for the Instagram ground truthing approach. The profile’s photo contributions are listed chronologically. The search bar on the left is used for specific searches. The screenshot in b) shows the search results for posts tagged with #nothofaguspumilio. Here the first of a total of 1,017 posts are displayed (last accessed 28.10.2025).



fernando.v.fotografia
Laguna Capri, Parque Nacional Los Glaciares (North)

fernando.v.fotografia Laguna Capri, otoño 2022.
Exif: iso 10.000 ✓ f/10 ✓ 1/1000. (Sin tripode)
#otoño#elchaltén#lagunacapri#patagonia#patagoniaargentina#autumn#turismoargentina#visitargentina#natgeo#naturelover_gr#argentina#igerspatagonia#fotodeldia#cerrotorre#fitzroy#poincenot#benro#tokinausa#nikonistas#nikonargentina#nikonusa#turismonacional#otoñopatagonico#chalten#ig_shotz#igworldclub#ruta40#
55 Wo. Übersetzung anzeigen

Gefällt 528 Mal
13. AUGUST 2022

Kommentieren ...

Figure 5: a) Selected post at Laguna Capri in Argentina with the species visible in front (used with permission by Instagram user fernando.v.fotografia, 2022). The abrupt treeline in the background, the lake and Mount Fitz Roy are traceable landscape elements marked in red in b). The post resulted in 4 Instagram ground truthing points. One at the photo sampling position and three background points (marked in blue) at the abrupt and autumn-coloured treeline.

Table 2: Criteria for selecting Instagram posts to generate occurrence data for Nothofagus pumilio. All criteria must be met for an image to be included in the Instagram ground truthing analysis.

Criterion	**Element or Example**
Typical characteristics of *Nothofagus pumilio*	morphological characteristics (leaves, branches, habitus) autumn colouring abrupt treeline mono-species forest
Concrete location information	geographical tag location hashtag location description in the caption
Recognisable landscape elements	glaciers mountain peaks or ranges rivers, lakes roads tourist points, cities, villages coastlines
Fitting hashtags	hashtags describing the location or the species
Picture criteria	Avoid persons in focus no photo montages no emojis no extreme (colour) falsifications

The IGTA points were subsequently used for ground truthing of large-scale remote sensing data. Raster occurrence data of *N. pumilio* were created using supervised classification of Sentinel-2 Level 2A imagery at a spatial resolution of 20 m. Summer (month December and January), autumn (month April), and winter (month August) Sentinel-2 data from 2019 to 2022 were downloaded using the R package “sen2R” (version 1.5.1, Ranghetti et al., 2020). Sentinel-2 scenes covered the entire distribution range of *N. pumilio*, with most scenes having a cloud cover of 15 % or less. A maximum cloud cover threshold of 50 % was accepted. To prevent misclassification during the classification process, the Scene Classification Layer (SCL) (Figure 6b), a standard product of the Sentinel-2 Level 2A data, was used to remove all raster cells, that were not classified as vegetation. The supervised classification was then trained on the Sentinel-2 data with training areas and three classes (1 = deciduous vegetation, 2 = evergreen vegetation and 3 = low vegetation/grassland). Autumn data made autumn-coloured forest at the treeline visible, while winter scenes highlighted evergreen vegetation. Two separate classifications were conducted using commonly applied algorithms and evaluated based on Kappa values (Richards, 2022). The supervised classification was performed using SAGA GIS via the “RSAGA” package (version 1.4.0, Brenning et al., 2022) in R (version 4.4.1, R Core Team, 2024). Summer and autumn scene classifications using the RF algorithm achieved the highest performance and outperformed those obtained with the Maximum Likelihood, Minimum Distance, and Spectral Angle algorithms. The summer and autumn results for deciduous vegetation were merged into one layer at the end (Figure 6d). With this approach also other deciduous species except *N. pumilio* were recorded. Therefore, an altitude correction was applied to ensure that only high-elevation deciduous forest was classified as *N. pumilio*. The thresholds for altitude correction in the northern and central study areas (800 m from 35°S to 40°S; 500 m from 40°S to 45°S; 250 m from 45° to 50°S) were determined on the basis of literature. The raster data spans over 2000 km latitudinal extent from 33.49°S to 56.27°S and captures the whole distribution range of *N. pumilio,* including inaccessible areas. Data gaps resulted from missing Sentinel-2 data or due to cloud or shadow coverage. Finally, ground truthing was conducted by verifying whether the IGTA points coincided with the raster occurrence data. As a result, two valid species occurrence datasets were generated through the IGTA approach: a point dataset and a raster dataset with reduced spatial and sampling bias.

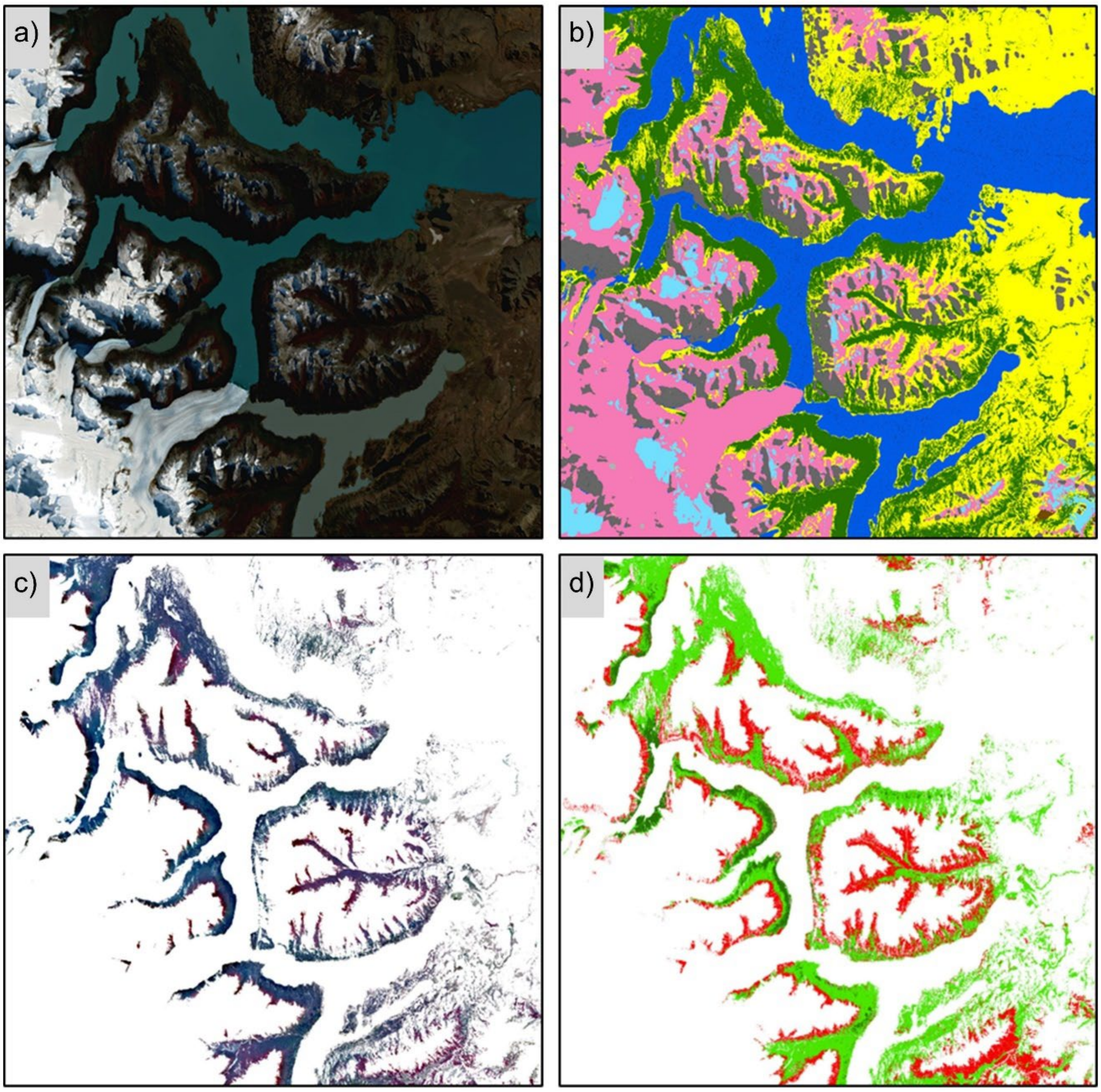


Figure 6: a) Sentinel-2 autumn scene at the Perito Moreno Glacier, Argentina, b) Scene Classification Layer of the Sentinel-2 scene: green area represents vegetation, c) the masked Sentinel-2 scene and, d) classification result with three classes (red = Nothofagus pumilio, dark green = evergreen vegetation, light green = low vegetation/grassland).

4.2 Downloading and Preprocessing Bioclimatic Predictors

A correlative ENM requires spatial data about the species occurrence and environmental predictor variables (Sillero et al., 2021). Most commonly used climate datasets for ecological applications are the WorldClim (Hijmans et al., 2005; Fick and Hijmans, 2017) and CHELSA datasets (Karger et al., 2017; Karger et al., 2021). Both contain climate data on a daily, monthly and annual basis. Furthermore, they offer bioclimatic variables, which are temperature and precipitation variables calculated for quarters and annual means, that allow for the interpretation of seasonality. CHELSA Bioclim data were chosen, because they tend to perform better in high mountain ecosystems than the WorldClim dataset (Soria-Auza et al., 2010; Bobrowski et al., 2021a). The CHELSA Bioclim data version 2.1 is a global dataset consisting of 19 variables in a resolution of 30 arc sec ~ 1 km (see Table 3). The temperature and precipitation values are averaged from climate records for the period 1981 to 2010. To download and preprocess the data, the R package “ClimDatDownloadR” (version 0.1.7.6, Jentsch, 2025) was used. The package, described in detail in Publication III, enables the download of data, clips it to a predefined spatial extent, and optionally converts the files into ESRI ASCII format. The function below illustrates how the data were downloaded and pre-processed. All available functions of the package are published on GitHub (GitHub Inc., 2025).

```r
# CHELSA Bioclim - Download ------------------------------------------

library(ClimDatDownloadR)
options(timeout=3600) # cache time to 1 hour

Chelsa.Clim.download(
  save.location = paste0(WD, "data/"),
  parameter = c("bio"),
  bio.var =  c(1:19), # all 19 Bioclim variables will be downloaded
  version.var = "2.1", # here the newer version is chosen
  clipping = TRUE, # TRUE for clipping with shapefile
  clip.shapefile = paste0(WD, "BioClimAreaLonLat.shp"),
  convert.files.to.asc = FALSE, # tiff data will be downloaded
  stacking.data = FALSE,
  combine.raw.zip = FALSE,
  delete.raw.data = TRUE,
  save.bib.file = FALSE
)
```

The 19 Bioclim variables are highly multicollinear and even though RF can handle multicollinearity well, it is recommended to use a subset of low or uncorrelated predictors for modelling. Only the variables calculated for quarters and seasonality variables were selected, since these allow for a better ecological understanding of the bioclimatic conditions at the treeline than variables for individual months or annual averages, as conclusions can be drawn about seasonality (Bobrowski et al., 2017). Furthermore, the R package "VSURF" (version 1.2.0, Genuer et al., 2015) a tool to identify which variables are important for interpretation and prediction, removed variables that were redundant from the variable subset. For modelling, eight Bioclim variables were used (see variables marked with an X in Table 3). The CHELSA version 2.1 dataset also contains data on future CMIP6 scenarios based on Representative Concentration Pathways (RCP) (Karger et al., 2021). The SSP126 (RCP2.6), SSP370 (RCP7), and SSP585 (RCP8.5) scenarios for the years 2041 to 2070 and 2071 to 2100 from the MPIESM1-2HR model were used for modelling the potential future distribution of *N. pumilio*.

Table 3: CHELSA Bioclim variables calculated for quarters and seasonality variables used in the analysis (X). The variables bio 9 and bio 16 have been excluded by the VSURF analysis.

Name	**Description**	**Used (X)**
bio 1	mean annual air temperature [°C]	
bio 2	mean diurnal air temperature range [°C]	
bio 3	isothermality [°C]	
bio 4	temperature seasonality [°C/100] [1]	X
bio 5	mean daily maximum air temperature of the warmest month [°C]	
bio 6	mean daily minimum air temperature of the coldest month [°C]	
bio 7	annual range of air temperature [°C]	
bio 8	mean daily mean air temperatures of the wettest quarter [°C]	X
bio 9	mean daily mean air temperatures of the driest quarter [°C]	excluded
bio 10	mean daily mean air temperatures of the warmest quarter [°C]	X
bio 11	mean daily mean air temperatures of the coldest quarter [°C]	X
bio 12	annual precipitation amount [kg m^{-2}]	
bio 13	precipitation amount of the wettest month[kg m-2]	
bio 14	precipitation amount of the driest month [kg m-2]	
bio 15	precipitation seasonality [kg m^{-2}] [2]	X
bio 16	mean monthly precipitation amount of the wettest quarter [kg m-2]	excluded
bio 17	mean monthly precipitation amount of the driest quarter [kg m-2]	X
bio 18	mean monthly precipitation amount of the warmest quarter [kg m-2]	X
bio 19	mean monthly precipitation amount of the coldest quarter [kg m-2]	X

[1] standard deviation of the monthly mean temperatures; [2] the coefficient of variation is the standard deviation of the monthly precipitation estimates expressed as a percentage of the mean of those estimates.

4.3 Model Algorithm and Model Procedure

The modelling of the potential current and future distribution is facilitated by the integration of IGTA occurrence data and bioclimatic predictors in Publication IV. The nature of the two species input datasets enables two model approaches: (1) a typical binary presence–absence model and (2) a continuous (abundance) model. As described before, numerous regression and machine learning approaches have been applied to ENM (Li and Wang, 2013). Here, the Random Forest algorithm was used (Breiman, 2001a), as it is capable of handling both data types, effectively managing multicollinearity, offering high predictive performance, and remaining largely interpretable (Breiman, 2001a, 2001b). Despite widespread criticism that machine learning algorithms are black boxes with limited interpretability (Li and Wang, 2013, Breiman, 2001b), RF results can be interpreted with variable importance analysis, partial dependence plots (PDP), and Shapley Additive Explanations (SHAP) analysis in order to draw conclusions about ecological site requirements of the study species.

The first steps in modelling involve the preparation of input data. For the binary presence–absence approach using RF classification, it was first ensured that only one presence point remained per 1 x 1 km Bioclim raster cell (spatial thinning). This filtering step removed 239 duplicate points, resulting in a final set of 999 presence points. Subsequently, pseudo-absence points were generated. This process involved creating an alpha hull with a 1 km buffer around the presence data to serve as a mask. To avoid placing absence points within presence raster cells, an additional 5 km buffer was applied around all presence points. Within this masked area, 2,000 pseudo-absence points were randomly generated. Main literature recommends a 1:1 ratio, or a prevalence of 0.5, for RF models, but discuss at the same time that a larger number of pseudo-absence points may be necessary when modelling across large spatial extents (Barbet-Massin et al., 2012). Since a 1:1 ratio resulted in an overprediction of presence, a 1:2 ratio was applied. For the continuous approach (RF regression), the 20 m resolution raster data were aggregated to the target resolution of 30 arc sec ~ 1 km by calculating the percentage of area covered by *N. pumilio*. This created a species variable with cover values for *N. pumilio* occurrence, ranging from close to 0 % to a maximum of 99.96 %.

Spatial data are often spatially autocorrelated (Legendre, 1993, Dormann et al., 2007), meaning that environmental conditions at nearby occurrence locations are more similar to each other than to those at more distant locations. This spatial dependence can significantly influence model results and lead to overestimation of model performance if not properly accounted for. To minimise autocorrelation effects, a spatial-cross validation approach was used (Roberts et al., 2016). Specifically, 5-fold spatial cross-validation was conducted using the R package “blockCV” (version 3.1.5, Valavi et al., 2019). For 5-fold cross-validation, the data is initially partitioned into spatial blocks of a predefined size, and each block is randomly assigned to one of five folds ($k = 5$). The model is then trained on four folds ($k - 1$) and evaluated on the remaining fold. This procedure is repeated five times, ensuring that each fold serves as both a training and a testing set.

With the function “cv_spatial_autocor” the distance of autocorrelation was first estimated. As a result, 50 km hexagonal spatial blocks were created (see Figure A1 in the appendix section). During the spatial cross-validation process, not only was model performance evaluated across the splits, but also the optimal hyperparameters for modelling were identified. Different numbers of variables considered at each split (mtry: 2, 3, 4) and varying numbers of decision trees (ntree: 100, 300, 500) were tested. This approach resulted in the calculation of 45 models (9 hyperparameter combinations and 5 folds). The best model was chosen by highest average AUC for the RF classification model and the highest R^2 for the RF regression model. For both models, the combination of hyperparameters $mtry = 2$ and $\mathrm{ntree} = 500$ yielded best performance. The optimal hyperparameter setting was then applied to the entire dataset to model the current and future distribution of *N. pumilio*.

For RF models, variable importance, PDP, and SHAP analysis can be used to interpret the model results. There are two types of variable importance: accuracy importance, which indicates most influential predictors for model accuracy, and Gini importance, which indicates the most frequently used variables for decision at nodes (Wei et al., 2015). Both were used for interpretation. PDP allow for the interpretation of the influence of individual predictors on the model outcome while all other predictors stay constant (Friedman, 2001). They visualise this relationship by plotting the values of the predictor variable on the x-axis against the corresponding values of the target variable on the y-axis. For PDP plotting, the function “partialPlot” from the “randomForest” R package (version 4.7-1.2, Breiman et al., 2002) was used. While PDP only allow for the interpretation of single predictors, SHAP summary or bee swarm plots, an analysis that originates from cooperative game theory, provide a comprehensive overview of the contribution of each predictor to the model outcome (Li et al., 2024). Here the “fastshap” R package (version 0.1.1, Greenwell, 2024) was used. SHAP summary plots display SHAP values on the x-axis, indicating a positive or negative influence of each predictor on the model output, while the predictor variables are listed on the y-axis. These analyses enable an interpretation of the ecological niche of *N. pumilio* on the basis of bioclimatic variables.

Spatial predictions under current and future climate conditions were conducted using R and visualised with ArcGIS Pro (version 2.7.0, ESRI, 2020). To quantify and evaluate the modelled treeline elevation, a Digital Surface Model (DSM) was used to extract the predicted elevation of the treeline and compare it with on-site treeline measurements provided by Lara et al. (2005) across the entire study area. To identify the highest raster cell representing treeline elevation, a threshold needed to be defined. Based on a commonly accepted definition of the treeline, where canopy cover declines to approximately 30 %, marking the upper boundary of closed forest (Holtmeier, 2009), a threshold of 30 % cover (or 0.3 predicted probability) was applied. Using the ALOS Global DSM (30 m resolution, JAXA EORC, 2025), the elevation of the highest raster cell based on the defined threshold was recorded at the 13 treeline locations documented by Lara et al. (2005).

4.4 Overview of Original Publications

Publication I

Werner, M., Weidinger, J., Böhner, J., Schickhoff, U., and Bobrowski, M. (2024). Instagram data for validating *Nothofagus pumilio* distribution mapping in the Southern Andes: A novel ground truthing approach. *Frontiers of Biogeography*, *17*, Article e140606. https://doi.org/10.21425/fob.17.140606

Abstract:

The availability of valid, non-biased species occurrence data has always been a major challenge for biodiversity research and modelling studies. Data from open-source databases or remote sensing are promising approaches to increase the availability of species occurrence data. However, these data may contain spatial, temporal, and taxonomic biases or require ground truthing. In recent years, social media has received attention for its contribution to species occurrence data sampling and ground truthing approaches. The wide reach of social media platforms allows for rapid and large-scale analyses. Here we introduce a novel Instagram ground truthing approach to validate the occurrence mapping of *Nothofagus pumilio* across its entire distribution range. The treeline species of the southern Andes has been extensively studied in small-scale studies, but large-scale modelling approaches are largely missing due to limited accessibility to treeline sites resulting in a lack of occurrence data. The content posted on the social media platform Instagram consists of images and videos in which the species *N. pumilio* and its location can be identified. By searching for suitable posts using hashtags and location tags, we created 1238 Instagram ground truthing points. We compared the performance of our dataset with open-source data from the Global Biodiversity Information Facility (GBIF) through visual, quantitative, and bias analyses, acknowledging that both social media-based and Citizen Science data are subject to sampling and spatial biases due to collection in human-accessible areas. The Instagram ground truthing points were subsequently used to validate remote sensing occurrence data, generated using Sentinel-2 level 2A data and supervised classification. The combined approach – Instagram ground truthing and remote sensing – allows for the collection of occurrence data over the entire latitudinal range of *N. pumilio*, covering approximately 2000 km.

Author Contributions:

Conceptualization, **M.W.**, J.W. J.B., U.S. and M.B.; methodology, **M.W.**, J.W. and M.B.; validation, **M.W.**, J.W. and M.B.; formal analysis, **M.W.**; investigation, **M.W.**; data curation, **M.W.**; writing—original draft preparation, **M.W.**; writing—review and editing, **M.W.**, J.W., J.B., U.S. and M.B.; visualisation, **M.W.**; supervision, J.B. and U.S. All authors have read and agreed to the published version of the manuscript.

Publication II

Werner, M., Weidinger, J., Böhner, J., Schickhoff, U., and Bobrowski, M. (2024). Instagram ground truthing approach – Spatial occurrence data of *Nothofagus pumilio* in the Southern Andes [Data set]. https://doi.org/10.25592/UHHFDM.16239

Abstract:

Valid and unbiased species occurrence data are scarce, making their availability a challenge for ecological modelling approaches. Remote sensing is a promising approach for increasing the availability of species occurrence data, but it requires ground truthing for validation. Recently, Citizen Science and social media have contributed to improving occurrence data sampling and ground truthing methods. This dataset contains remote sensing occurrence data of the tree species *Nothofagus pumilio*, which was validated by ground truthing points created using a novel Instagram ground truthing approach.

Author Contributions:

This publication is a dataset publication corresponding to Publication I (see above).

Publication III

Jentsch, H., Weidinger, J., **Werner, M**., and Bobrowski, M. (2025). ClimDatDownloadR: Accessing Climate Data Repositories for Modelling [Manuscript submitted for publication].

Systematical accessing, downloading, and pre-processing climatological data from CHELSA and WorldClim remains a challenge in different environmental disciplines like Species Distribution Modelling (SDM) and climate studies. This package provides a set of functions that allow easy access and customized selection of climate datasets. Besides downloading the raw data, also functionalities to complete pre-processing steps like clipping, rescaling, and file management are available. The applications of the package range from one-time-use to implementing the functions in automatic processing of scientific workflows.

Author Contributions:

Conceptualization, H.J., J.W., **M.W.** and M.B.; methodology, H.J., J.W.; validation, H.J., J.W., **M.W.** and M.B.; formal analysis, J.H.; writing—original draft preparation, H.J. and **M.W.**; writing—review and editing H.J., J.W., **M.W.** and M.B.; visualisation, H.J.; supervision, M.B., All authors have read and agreed to the submitted version of the manuscript.

Publication IV

Werner, M., Böhner, J., Oldeland, J., Schickhoff, U., Weidinger, J., and Bobrowski, M. (2025). Treeline Species Distribution Under Climate Change: Modelling the Current and Future Range of *Nothofagus pumilio* in the Southern Andes. *Forests*, *16*(8), 1211. https://doi.org/10.3390/f16081211

Abstract:

Although treeline ecotones are significant components of vulnerable mountain ecosystems and key indicators of climate change, treelines of the Southern Hemisphere remain largely outside of research focus. In this study, we investigate, for the first time, the current and future distribution of the treeline species *Nothofagus pumilio* in the Southern Andes using a Species Distribution Modelling approach. The lack of modelling studies in this region can be contributed to missing occurrence data for the species. In a preliminary study, both point and raster data were generated using a novel Instagram ground truthing approach and remote sensing. Here we tested the performance of the two datasets: a typical binary species dataset consisting of occurrence points and pseudo-absence points and a continuous dataset where species occurrence was determined by supervised classification. We used a Random Forest (RF) classification and a RF regression approach. RF is applicable for both datasets, has a very good performance, handles multicollinearity and remains largely interpretable. We used bioclimatic variables from CHELSA as predictors. The two models differ in terms of variable importance and spatial prediction. While a temperature variable is the most important variable in the RF classification, the RF regression model was mainly modelled by precipitation variables. Heat deficiency is the most important limiting factor for tree growth at treelines. It is evident, however, that water availability and drought stress will play an increasingly important role for the future competitiveness of treeline species and their distribution. Modelling with binary presence–absence point data in the RF classification model led to an overprediction of the potential distribution of the species in summit regions and in glacier areas, while the RF regression model, trained with continuous raster data, led to a spatial prediction with small-scale details. The time-consuming and costly acquisition of complex species information should be accepted in order to provide better predictions and insights into the potential current and future distribution of a species.

Author Contributions:

Conceptualization, **M.W.**, J.B., J.O., U.S., J.W. and M.B.; methodology, **M.W.**, J.O., J.W. and M.B.; validation, **M.W.**, J.O. and M.B.; formal analysis, **M.W.**; investigation, **M.W.**; data curation, **M.W.**; writing—original draft preparation, **M.W.**; writing—review and editing, **M.W.**, J.B., J.O., U.S. and M.B.; visualisation, **M.W.**; supervision, J.B. and U.S. All authors have read and agreed to the published version of the manuscript.

5. Synthesis and Discussion

This thesis aims at modelling the potential current and future distribution of the treeline species *Nothofagus pumilio* using an Ecological Niche Modelling (ENM) approach, thereby addressing the existing research gap in treeline modelling studies in the Southern Andes. A novel Instagram Ground Truthing Approach (IGTA) was developed to generate bias-minimised species occurrence data, resulting in both a point dataset and a raster dataset, the latter of which was validated using the IGTA points. Both datasets served as inputs for modelling with a machine learning algorithm chosen for its high predictive performance and interpretability. The resulting models were compared and the strengths and limitations of each input dataset were critically assessed. In doing so, this study applied an approach of "Improved Ecological Niche Modelling", combining less-biased occurrence data with advanced modelling techniques. This chapter addresses the research questions by summarising and discussing the main findings.

5.1 Ground Truthing with Instagram: A Novel Approach

Can Instagram be used to obtain occurrence data for *Nothofagus pumilio*, and is the spatial bias in the resulting IGTA dataset reduced compared to that in commonly used GBIF data?

N. pumilio is particularly suitable for analysis with Instagram due to its morphology, phenology, and ecology. The species occurs in pure forest stands at the treeline within a scenic topographic and highly touristic region with numerous national parks. The orange-red autumn colouring also makes the species an attractive photo motif. As a result, a large number of posts featuring *N. pumilio* were identified on Instagram and assessed for suitability as occurrence points using a strict set of criteria. Occurrence points were created when the species and the location where the photo was taken could be clearly identified by landscape elements. The IGTA resulted in 1,238 valid and transparently traceable occurrence points of *N. pumilio*. A total of 297 posts published between 2017 and 2022 were used for the analysis. These were posts linked to specific hashtags relating to the species or location, as well as posts with location tags in the distribution area. In this way, the dataset included not only posts in which experts had deliberately photographed and posted the species, but also everyday images taken by non-experts, in which the occurrence was captured unintentionally. When georeferencing the posts, points were assigned not only to the exact location where the image was taken (460 points), but also to identifiable occurrences in the background of the image, including adjacent stands at the autumn-coloured treeline (778 points). Figure 7 shows all 1,238 sampled IGTA points.

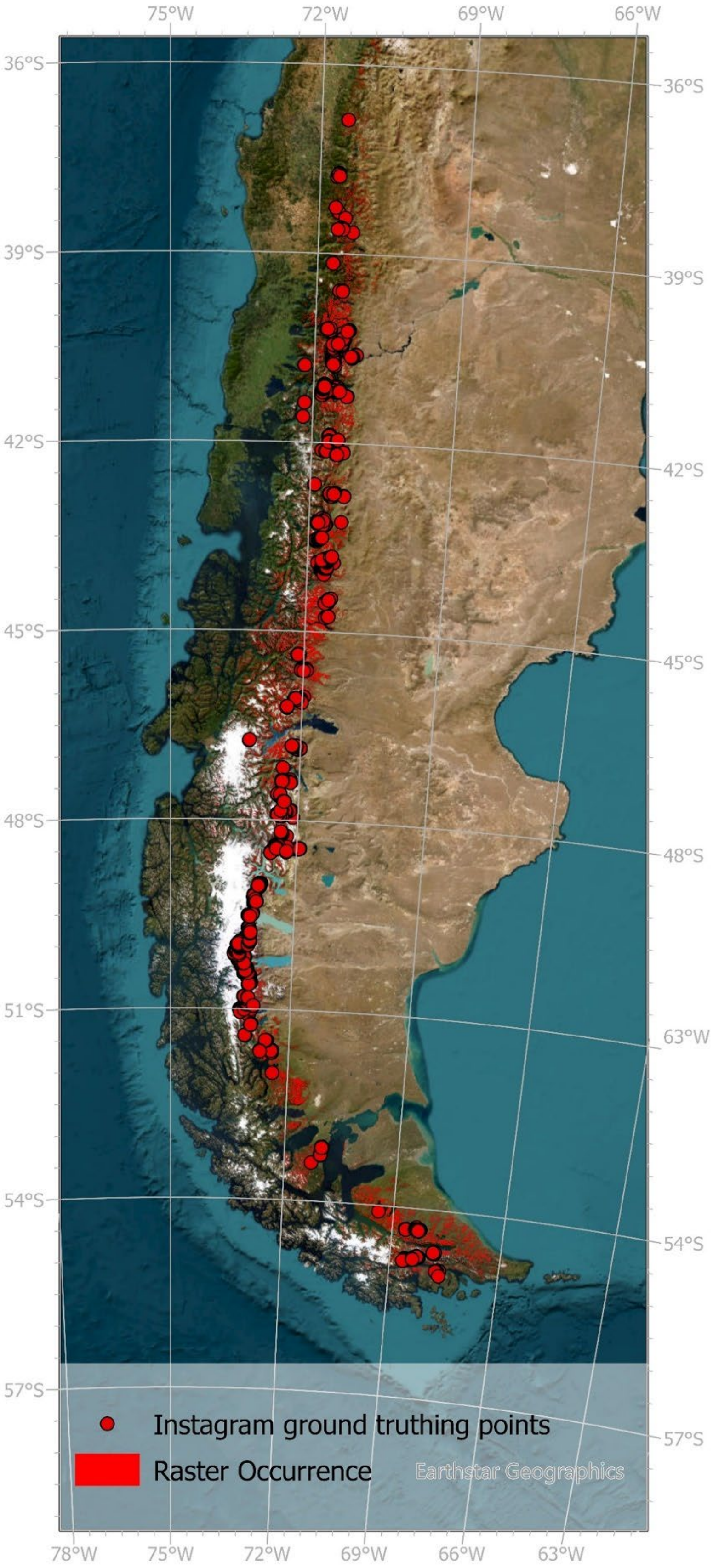


Figure 7: The two datasets resulting from the Instagram ground truthing approach (IGTA): IGTA points and Sentinel-2 raster data in a resolution of 20 m covering the entire distribution range in the Southern Andes

According to a vignette study by Haklay et al. (2021) 50 % of respondents describe the re-use of social media content as a form of CS. However, CS typically requires active involvement of participants, information transfer, and data transparency (Haklay et al., 2021). The IGTA can be framed as "passive CS" method (Nascimento et al., 2024). Social media platforms can not only be used to sample occurrence data but also to share information, communicate research findings, and recruit Citizen Scientists (Nascimento et al., 2024). Furthermore, Instagram users can be informed about the importance of using geotags and providing precise location information to support initiatives like the IGTA. For this purpose and to take CS principles into account, an Instagram profile (nothofagus_pumilio_research) was created and used to post information about the project. When a post from another user was used for occurrence data sampling, it was "liked" (by clicking the heart icon, see Figure 5a as an example), which notifies the user and draws their attention to the research account. In addition, the chat function was used to ask users about the exact location of a post or to discuss methods and research findings. To promote data transparency, the final dataset was published via an open-access data provider (Publication II), bringing the project more in line with the standards expected of CS initiatives. An advantage of passive CS is that it does not require active recruitment of participants, and data collection via social media can be conducted remotely, making it even more cost- and time-effective than traditional CS approaches (Edwards et al., 2021). Moreover, the stratified IGTA aimed to further reduce common sources of bias typically present in CS-derived datasets.

Sampling bias in global datasets, which often contain data from unstructured CS projects, arises from sampling behaviour in predominantly urban areas. As a result, the occurrence data does not always correspond to the actual distribution of a species, which can have a negative impact on the model results (Di Cecco et al., 2021). In the IGTA, the high number of available posts across the entire distribution area and manual georeferencing led to a reduction in this bias. The "sampbias" analysis (see Figure 8) showed that, in comparison to 558 filtered GBIF points with a maximum coordinate uncertainty of 1 km (GBIF, 2024), the IGTA dataset provided better coverage of rural areas. While both datasets showed increased sampling rates around urban centres, the sampling rate of IGTA points was significantly higher in non-urban regions, resulting in a more homogeneous spatial distribution and reduced bias towards urban areas. Sampling bias, such as incorrect coordinates due to errors in georeferencing of museum data (Boakes et al., 2010; Marcer et al., 2022) or weak satellite signals during on-site sampling (Uyeda et al., 2020), was avoided in the IGTA by the manual transfer of the occurrence points. However, a source of uncertainty in the IGTA data must be considered. The publication date of the posts does not necessarily correspond to the day or even the year the photo was taken. This means that temporal bias may be introduced (Nascimento et al., 2024), as the IGTA does not reliably provide information on the exact time the photograph was taken. Some information can be found in the caption or comments or must be requested individually from Instagram users using the chat function. Furthermore, unlike databases such as GBIF, which include historical records from museums and other sources, “historical” data is not available through Instagram.

In summary, the global usage and high volume of posts on Instagram, combined with the manual georeferencing of occurrences, enable a cost-effective sampling of less-biased occurrence data for *N. pumilio*. Compared to GBIF data, spatial bias is significantly reduced. However, the IGTA was still relatively time-consuming, as manually searching for posts via Instagram’s search function requires considerable effort. Some social media platforms, such as Flickr, allow access and automated searches via an Application Programming Interface (API), enabling faster identification of suitable posts and facilitating automation (e.g., Fox et al., 2022). In contrast, Meta (www.meta.com/), the company behind Instagram, Facebook, Messenger, and WhatsApp, does not provide such access. Nevertheless, the time-consuming process of manual searches is offset by the significantly broader reach of the Instagram platform and therefore coverage of potential occurrence points.

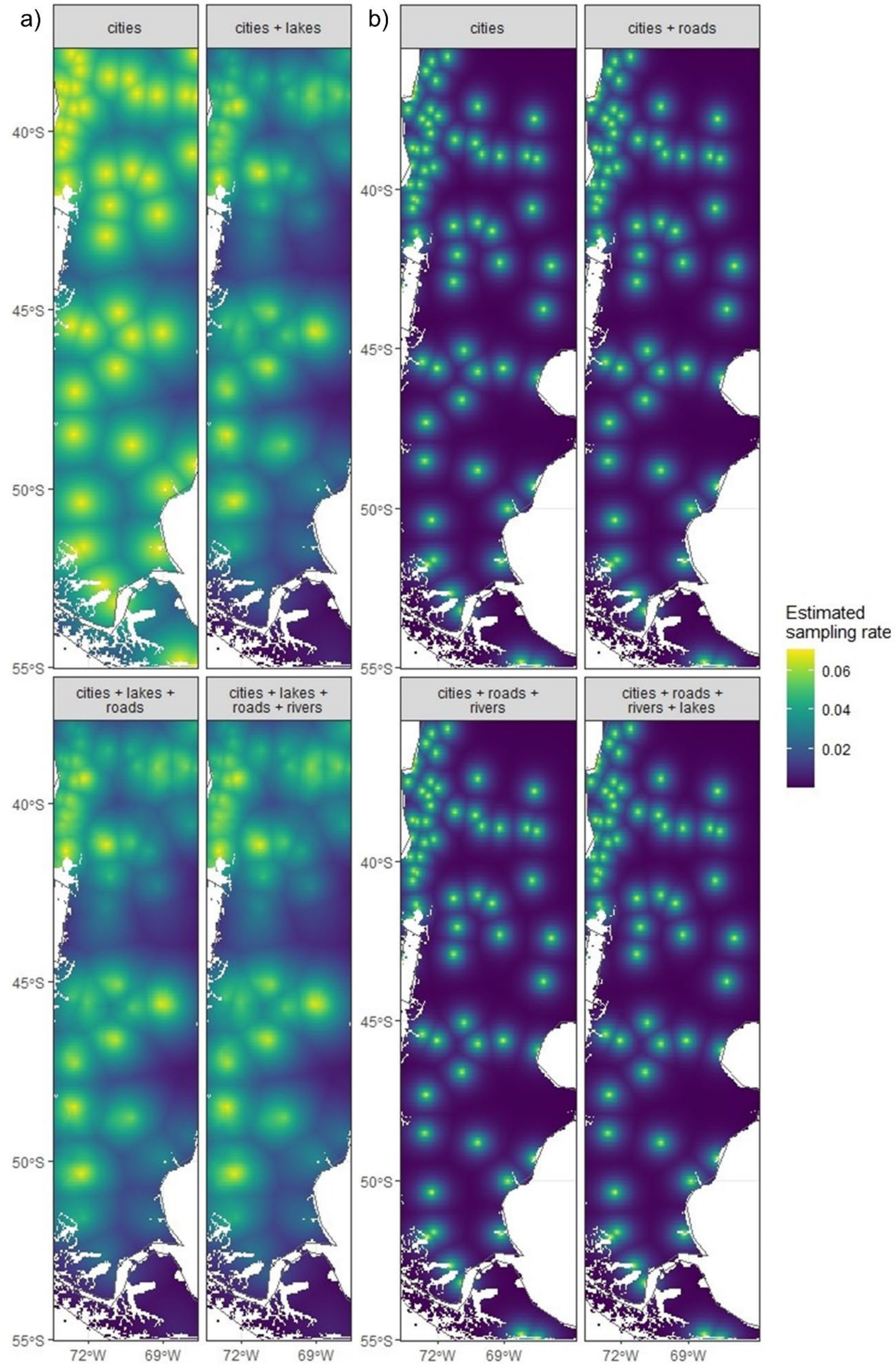


Figure 8: Results of the bias analysis using the R package "sampbias", indicating the sampling rate near urban areas for a) the Instagram ground truthing approach (IGTA) data and b) GBIF data. While high sampling rates in the GBIF dataset are concentrated near urban centres, the IGTA data shows a more homogeneous sampling across the study area.

Can supervised classification be used to generate raster occurrence data for *Nothofagus pumilio* across its entire range?

The occurrence of *N. pumilio* in pure forest stands at the treeline and its phenology enabled the creation of continuous raster data via supervised classification, resulting in over 2000 km range from 35°S to 55°S at a resolution of 20 m (Figure 7). The classification of summer and autumn scenes achieved high performance with Kappa values of 0.89 and 0.97, respectively. Potential sources of error from shadow or cloud cover were mitigated by masking the Sentinel-2 scenes using the Scene Classification Layer (SCL) before classification. However, this also introduced gaps in the dataset, as did missing Sentinel-2 data in certain areas, which could not be replaced with alternative scenes during the sensing period from 2019 to 2022. Ground truthing of the classification results with the IGTA points resulted in a 92.25 % match, meaning 1,142 IGTA points cover the determined raster data, indicating a high accuracy. To ensure that other deciduous species were not included in the raster data, an altitude correction was performed, retaining only high-elevation deciduous forest. However, one source of error that could not be avoided is inclusion of some *N. antarctica* stands. The species are morphologically and ecologically very similar. Even hybrids between the two species are possible (Soliani et al., 2015). In the future, to strictly separate the two species, external datasets could be used, e.g., from local forestry authorities (Corporación Nacional Forestal (CONAF), Chile and Ordenamiento Territorial de Bosque Nativo (OTBN), Argentina). However, at this point, the forest type “high mountain deciduous forest” was reliably determined, even if *N. antarctica* was partially included.

The IGTA facilitated the creation of large-scale raster data for *N. pumilio* and its validation through Instagram ground truthing points, resulting in two robust datasets. The approach is transferable to other ecological studies, both in terms of Instagram-based sampling and the generation of remote sensing data, provided that the study species exhibit similar characteristics to *N. pumilio.* Posted content must include visible landscape elements, which excludes close-up images of smaller plant or animal species. Similarly, remote sensing can only reliably detect and distinguish contiguous species populations or individual mature trees (Immitzer et al., 2016, 2019). However, for plant or animal species that contain images with recognisable landscapes, the IGTA offers a promising approach for generating valid species occurrence datasets.

5.2 Modelling the Ecological Niche of *Nothofagus pumilio*: A Comparison of Two Approaches

The ecological niche and potential current and future distribution of *N. pumilio* were analysed using two models: a Random Forest (RF) classification model with the IGTA point dataset and an RF regression model with the continuous raster data. A 5-fold spatial cross-validation was used to validate the models. The model quality was evaluated using the mean validation metrics and the best hyperparameters were selected. The best-performing RF classification model achieved the highest mean AUC of 0.9279 (±0.0257, 95 % confidence interval (CI): 0.8960–0.9599), an overall accuracy of 0.8466 (±0.0537, 95 % CI: 0.7799–0.9132), and a TSS of 0.6148 (±0.1582, 95 % CI: 0.4183–0.8112). The best RF regression model had an R^2 of 0.3933 (±0.0409, 95 % CI: 0.3425–0.4441). For both models, the best hyperparameters were $\mathrm{mtry} = 2$ and $\mathrm{mtree} = 500$, which were applied to the entire datasets for spatial prediction and interpretation. Variable importance, partial dependence plots (PDP) and Shapley Additive Explanations (SHAP) summary plots were used for characterising the ecological niche of *N. pumilio*. By applying the same algorithm, a direct comparison of the models and their outputs was possible, although it must be considered that the scales of the output variables differ. The RF classification model yielded predicted probability values for *N. pumilio* occurrence, and the RF regression model yielded predicted cover values for the species. These values were compared directly on the basis that low values in both datasets indicate largely unsuitable bioclimatic conditions, while high probability and cover values suggest suitable habitat. However, it is important to note that a probability of 0.01 represents only a very low likelihood of species occurrence, whereas 1 % cover indicates that the species is present, albeit in low abundance.

How can the ecological niche of *Nothofagus pumilio* be characterised under current climatic conditions, and what is the species' potential geographic distribution?

The temperature variable bio 8 (mean daily mean air temperatures of the wettest quarter) and the precipitation variable bio 19 (mean monthly precipitation amount of the coldest quarter) emerged in the RF classification model as the most important variables for model accuracy (accuracy importance ranking: bio 8, bio 19, bio 15, bio 17, bio 4, bio 10, bio 18, bio 11). In the RF regression model, bio 15 (precipitation seasonality) and bio 19 (mean monthly precipitation amount of the coldest quarter) emerged as the most important variables (accuracy importance ranking: bio 15, bio 19, bio 4, bio 11, bio 10, bio 8, bio 18, bio 17). PDP can be used to interpret the characteristics of single variables (Figure 9). In the PDP the temperature variable bio 8, which had the highest variable importance in the RF classification model, shows a high predicted probability at very low temperatures of −5 to −2°C in the wettest quarter, with a decline as temperatures increase. *N. pumilio* therefore occurs in cold areas in the wettest quarter. The influence of bio 15 on the predicted cover values contained more variation, but it can be concluded that *N. pumilio* prefers regions with more stable precipitation patterns, and the species is less likely to occur in areas with highly variable precipitation (> 70 %), e.g., with periods of drought.

SHAP plots allowed for a more comprehensive analysis of the ecological niche based on bioclimatic variables. Even though the characteristics of the SHAP plots of the models differed slightly, both showed the suitable climatic conditions for the species (Figure 9). The species *N. pumilio* is adapted to the harsh conditions in high mountain regions, as evidenced by the fact that it occurs in regions with cold winters (low values of bio 11, mean daily mean air temperatures of the coldest quarter) and mild summers (moderate values of bio 10, mean daily mean air temperatures of the warmest quarter). Heat deficiency is considered the most important limiting factor for tree growth at the alpine treeline (Körner, 2020). Globally, the alpine treeline follows the 6.4°C isotherm during growing season (Paulsen and Körner, 2014), which corresponds closely to the distribution limit of *N. pumilio* at the 6.6°C isotherm (Fajardo and Pieper, 2014). As described in Chapter 3, the species is adapted to the cold by concentrating carbohydrates in the xylem to prevent damage due to frost (Fajardo and Piper, 2014), mitigating the risk of frost drought by being a deciduous species (Körner, 2021), and furthermore due to its high phenotypic plasticity (Stecconi et al., 2010). Leaf size is reduced in high-elevation individuals (Premoli et al., 2007) and growth height ranges from 35 m at low elevations to krummholz growth at the treeline (Stecconi et al., 2010). The influence of temperature was evident not only in the SHAP plots but also in the high variable importance of bio 8 (mean daily mean air temperatures of the wettest quarter) in the RF classification model and in high GINI importance of bio 11 (mean daily mean air temperatures of the coldest quarter) in both models. However, precipitation has an important influence on the distribution of the species. The species occurs in areas with moderate to low precipitation sums, mainly in the form of snow (moderate values of bio 19, mean monthly precipitation amount of the coldest quarter). However, the species requires sufficient precipitation throughout the year (low to medium values for bio 17, mean monthly precipitation amount of the driest quarter and bio 18, mean monthly precipitation amount of the warmest quarter) and does not occur in regions with high precipitation variability (bio 15, precipitation seasonality). Very low precipitation (low values of bio 17, 18, 19) defines the boundaries of the distribution range of *N. pumilio*. Although the species is present in low-precipitation areas, such as the arid eastern regions (Hertel et al., 2008), drought conditions pose a significant challenge to its survival and regeneration (Rodríguez-Catón et al., 2016). Studies investigating the relationship between radial growth of the species and climate variations have found growth declines, especially in the northern distribution area, due to low precipitation in spring and summer (Daniels and Veblen, 2004; Álvarez et al., 2015). Droughts lead to seedling dieback and even increased mortality of mature trees (Rodríguez-Catón et al., 2016, Fajardo et al., 2019). However, a distinction must be made between northern and southern Patagonia. In more southern regions, high spring precipitation is primarily associated with delayed snowmelt and thus a shortened growing season, which in turn again limits tree growth (Villalba et al., 1997; Lara et al., 2001; Álvarez et al., 2015). The study by

Brand et al. (2022) summarises that growth in mesic locations increases with higher temperatures, followed by humid locations, while growth in dry locations decreases. Other authors describe that growth limitation at the treeline changes from cold-limited to drought-limited (Reiter et al., 2024). The model results can reflect these regional differences in precipitation patterns. Particularly, the RF regression model focused on the influence of precipitation variables, with bio 15 and bio 19 as the most important variables. Bio 15 suggests that *N. pumilio* is unlikely to occur in regions with high precipitation seasonality, which may be associated with periods of drought. Bio 19, on the other hand, may be linked to precipitation in the form of snow, the resulting snow cover, and water availability following thaw.

The models were used to predict the current potential distribution of the species. The prediction resulted in a distribution area from 36.35°S to 55.45°S for the RF classification model and a slightly smaller area from 35.24°S to 55.24°S for the RF regression model (see Figure 10). The projected distribution area of the RF classification model is somewhat more compact, while the RF regression model extends slightly further into the humid west. On Tierra del Fuego, the RF regression model shows a more homogeneous distribution than the RF classification model. The spatial predictions also differed in terms of their details and the treeline elevation. The comparison of the spatial predictions of the models is answered in detail by the last research question.

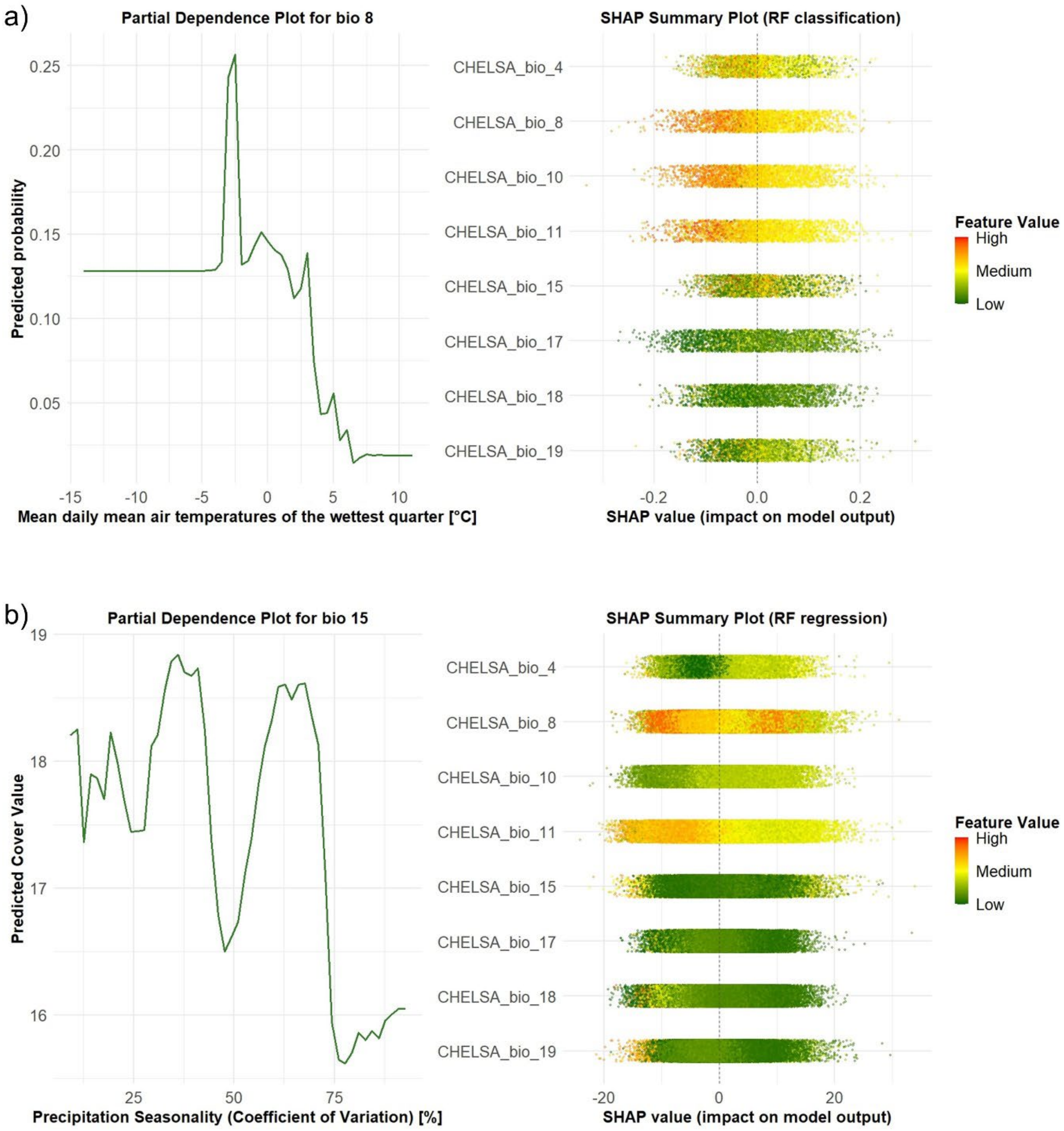


Figure 9: Partial Dependence Plots (PDP) and Shapley Additive Explanations (SHAP) summary plots for a) the Random Forest (RF) classification model and b) RF regression model. For the SHAP analysis, feature values were normalised (breakpoints: 0, 0.5, 1) to account for the differing units of temperature and precipitation variables.

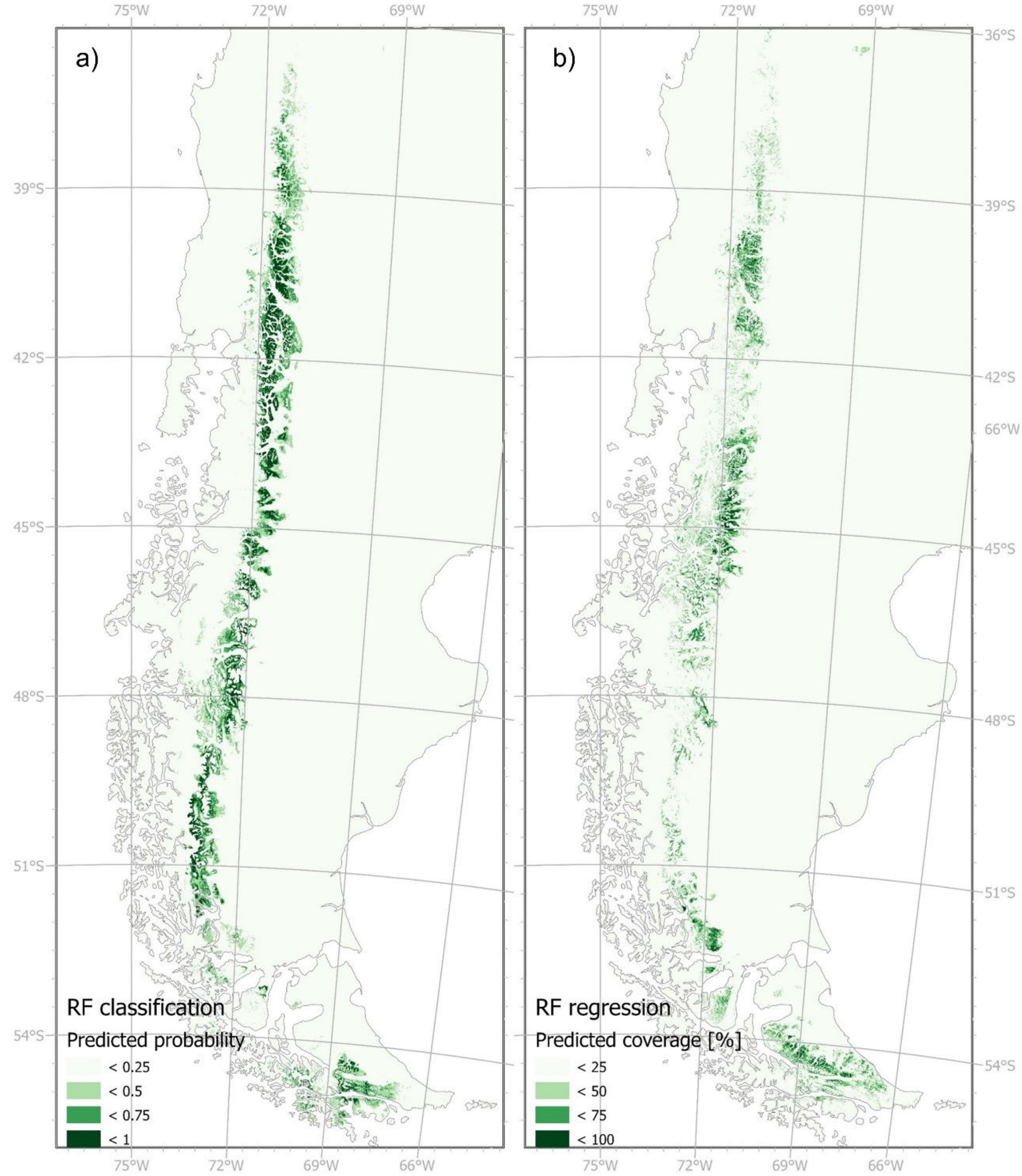


Figure 10: Potential current distribution of Nothofagus pumilio modelled by a) the Random Forest (RF) classification model and b) the RF regression model.

What potential range shifts are projected under future climate change scenarios?

CMIP6 data from CHELSA was used to predict the future distribution of *N. pumilio* under different SSP scenarios. Predictions were made for the periods 2041 to 2070 and 2071 to 2100 using the SSP126, SSP370 and SSP585 scenarios (see Figures A2 to A5 in the appendix section). Both models showed a shift of the treeline to higher elevations, and this trend intensifies with modelled time period and the SSP scenario. However, there are differences between the predictions. On the one hand, the RF classification model showed sharp declines in the total distribution area in the north and in the southern part on Tierra del Fuego, while these declines were only slight in the north and not pronounced in the south in the RF regression model. On the other hand, the RF classification model predicted an increase in currently snow- and ice-covered regions at the Southern Patagonian Icefield, while the RF regression model predicted a decrease in populations. The differences in the predicted upward shift of the treeline were clearly shown by the analysis of high-elevation occurrences (estimated treeline elevation) using a digital surface model at 13 selected locations across the study area (see Table 4 for time period 2041 to 2070 below and Table A6 for 2071 to 2100 in the appendix section). In general, the estimated treeline elevation in the RF classification model was higher than those in the RF regression model. The table also clearly shows the population decline in the north. At the second location, stands are missing in both models from the SSP370 scenario for the period 2041 to 2070.

The models are based on climatic factors only, so the analysis showed the potential climatic treeline shift in the Southern Andes. Whether this advance can actually take place also depends on other factors. Abrupt treelines generally advance less pronounced than diffuse treelines, as they are caused by seedling dieback outside the protective tree stands (Bader et al., 2021). However, higher temperatures and a prolonged growing season can improve conditions for seed production, fruit dispersal, seed viability, seedling establishment and survival, and thus the conditions for an upward shift in the treeline (Daniels and Veblen, 2004; Fajardo and Pieper, 2014; Srur et al., 2016, 2018; Aschero et al., 2022). A treeline advance of 5 to 10 m above the current treeline has already been recorded for *N. pumilio* (Srur et al., 2018). However, seedling establishment also depends heavily on water availability and soil moisture (Lett and Dorrepaal, 2018; Qiu et al., 2021). Seedling survival decreases with increasing aridity, and a deterioration in growth conditions for both seedlings and mature individuals was observed on low- and high-elevation slopes (Fajardo and Pieper, 2014; Rodríguez-Catón et al., 2016; Aschero et al., 2022; Reiter et al., 2024). Furthermore, even if temperatures and water availability are favourable for treeline advance, this can still be prevented by local topographic or edaphic factors. The inclusion of additional variables related to topography, soil, vegetation cover, and wind could further elucidate the potential upward shift.

Table 4: Estimated treeline elevation [m] at 13 locations according to Lara et al. (2005) based on the results of the Random Forest (RF) classification (Class.) and RF regression (Reg.) models under current climate conditions and CMIP6 SSP scenarios for the future period 2041 to 2070 (applied thresholds 0.3 and 30 %). NA = not available (no data recorded).

Coordinates		**Current Climate**		**SSP126 (2041–2070)**		**SSP370 (2041–2070)**		**SSP585 (2041–2070)**	
X	**Y**	**Class.**	**Reg.**	**Class.**	**Reg.**	**Class.**	**Reg.**	**Class.**	**Reg.**
−71.00	−35.36	NA	NA	NA	NA	NA	NA	NA	NA
−71.11	−37.27	1988	1949	2214	NA	NA	NA	NA	NA
−71.33	−38.42	1854	1789	2201	1709	2186	1920	2227	2045
−72.15	−40.42	1591	1437	1699	1636	1768	1674	2026	1674
−72.19	−41.48	1500	1201	1560	1464	1730	1638	1730	1720
−71.45	−43.07	1839	1440	1918	1545	2059	1725	1918	1725
−71.42	−44.39	1320	1216	1704	1324	1852	1509	1852	1591
−72.24	−47.12	1361	1197	1500	1423	1651	1439	1538	1500
−72.30	−48.30	1522	1074	1340	1098	1473	1098	1586	1209
−72.54	−50.57	1176	956	1296	961	1313	1103	1349	1124
−71.00	−53.00	543	560	592	721	NA	783	592	783
−68.45	−54.17	544	520	648	615	667	547	607	607
−67.30	−54.57	610	610	614	492	557	614	NA	614

What are the differences between the modelling approaches used in this study, and what are their respective advantages and disadvantages?

Differences between the model results existed in spatial prediction, variable importance, and thus PDP and SHAP plots, respectively. Even though a synthesis of the model results was possible to identify and interpret the suitable bioclimatic conditions for *N. pumilio*, the strengths and weaknesses of the models were particularly evident in the spatial predictions. The RF classification model tended to predict occurrences in higher, not vegetated areas, resulting in slight overprediction at high-elevation sites. This became particularly evident in a comparison of the elevations of the highest grid cells in the models and with external data. Table 5 compares the reported treeline elevation from an on-site study by Lara et al. (2005) at 13 locations across the species' distribution range with the estimated treeline elevation from the models. The treeline elevation of the RF classification model was in some cases significantly higher than the on-site data, but the data corresponded well in the southern distribution area. The treeline elevation of the RF regression model, in contrast, was only slightly higher in the north and otherwise corresponded very well with the actual measured values. Furthermore, the spatial prediction of the RF regression model showed small-scale details (see Figure 11). It distinguished better between forested valleys and bare peaks and did not overperform at high altitudes. This was also verified with external data from the Argentine forestry authority, Secretaria de Ambiente y Desarrollo Sustentable de la Nacion, which provides shapefiles on forest types (Mohr-Bell et al., 2019). A comparison of the polygons with the grid cells of the model results showed a very good match with the RF regression model results, while the grid cells of the RF classification model extended beyond the boundaries of the polygons (Figure 12). The overprediction at high altitudes can be attributed, among other things, to high variable importance of temperature-related variables, which often reflect the effects of altitude. This is also evident in the SHAP summary plots, where high temperatures exert only a negative influence on the model outcome in the RF classification model, thus predicting suitable climatic conditions predominantly in colder areas. In contrast, the RF regression model delineates extremely cold areas, such as summit regions, more precisely, as demonstrated by the predictor bio 10. In this context, low temperatures during the warmest quarter have a negative impact on the model outcome (see Figure 9).

The continuous species occurrence variable “cover values” of the RF regression model, calculated from the coverage of 20 m grid cells within a 1 km grid cell (the target resolution of the climatic predictors of the model), contains information on the species relative abundance and therefore additional unknown information on land cover. In contrast, the binary variable used in the RF classification model consists only of the values 0 or 1, indicating species absence or presence. Cover values can provide a more realistic insight into the actual conditions on site. For this reason, the RF regression model with this numerical variable provided a more accurate spatial prediction and a result that was better interpretable with more meaningful predictors. The inclusion of remote sensing data as environmental variables is already common practice in ENM. But the creation of occurrence data based on remote sensing data can improve model validity, holding great potential for future ENM studies.

Table 5: Treeline elevation sampled on site by Lara et al. (2005) compared with the modelled highest occurrence raster cells of the Random Forest (RF) classification (applied threshold 0.3) and RF regression model (threshold 30 %).

Treeline Position and Elevation [m] after Lara et al. (2005)				Treeline Elevation [m] Current Climate	
ID	**X**	**Y**	**Elevation Range**	**RF Class.**	**RF Reg.**
1	−71.00	−35.36	1530	NA	NA
2	−71.11	−37.27	1500–1720	1988	1949
3	−71.33	−38.42	1490–1650	1854	1789
4	−72.15	−40.42	1000–1300	1591	1437
5	−72.19	−41.48	1300	1500	1201
6	−71.45	−43.07	1230–1350	1839	1440
7	−71.42	−44.39	1000–1200	1320	1216
8	−72.24	−47.12	800–1180	1361	1197
9	−72.30	−48.30	1200	1522	1074
10	−72.54	−50.57	650–980	1176	956
11	−71.00	−53.00	350–600	543	560
12	−68.45	−54.17	200–600	544	520
13	−67.30	−54.57	300–600	610	610

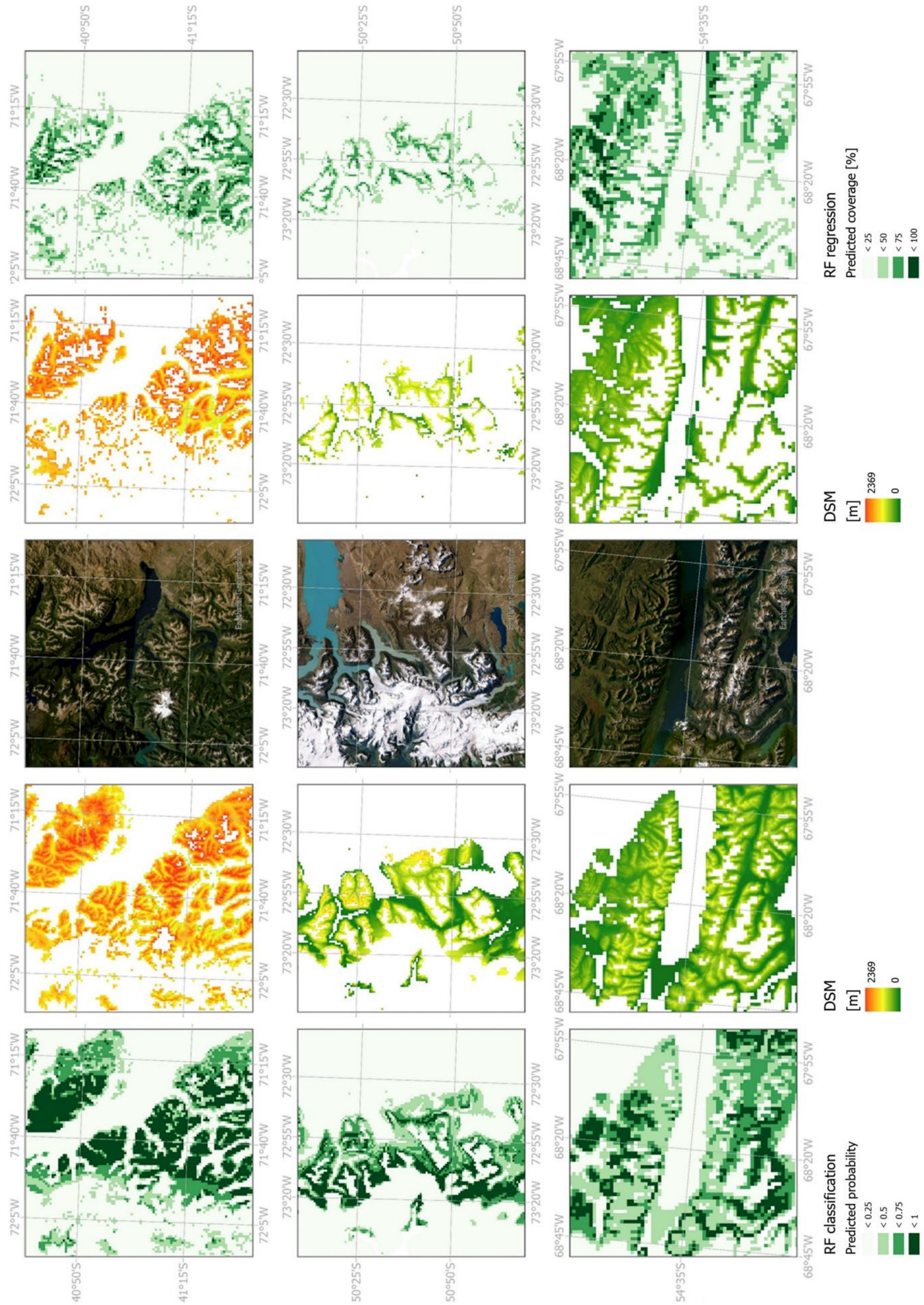


Figure 11: Details of the spatial predictions of the potential current distribution of Nothofagus pumilio from the Random Forest (RF) classification model and RF regression model, shown in comparison with a satellite basemap (centre) and a Digital Surface Model (DSM) indicating the elevation of the raster cells covered by the model outputs.

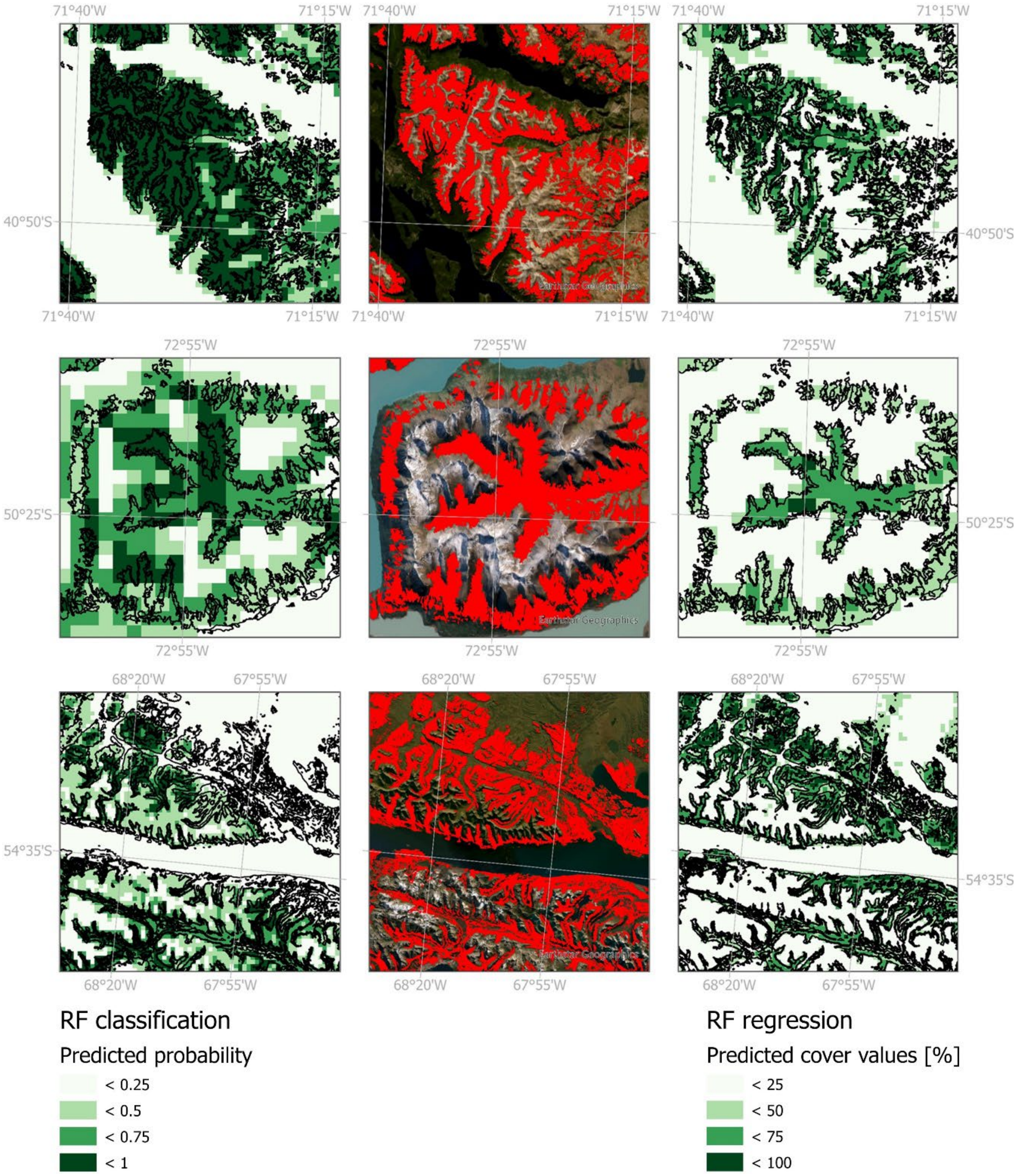


Figure 12: Comparison of spatial predictions from the Random Forest (RF) models under current climate with forest polygons of Nothofagus pumilio (red, centre) provided by the Argentinian forestry authority (Mohr-Bell et al., 2019).

6. Conclusion and Outlook

This study aimed to investigate the underlying bioclimatic requirements, and potential current and future distribution of the treeline species *Nothofagus pumilio* in the Southern Andes, using less-biased species occurrence data derived from a novel method combining social media and remote sensing. The primary objective was to improve a common modelling approach in two key steps: first, by developing the Instagram ground truthing approach (IGTA), and second, by comparing binary and continuous species datasets in Ecological Niche Modelling using a machine learning algorithm.

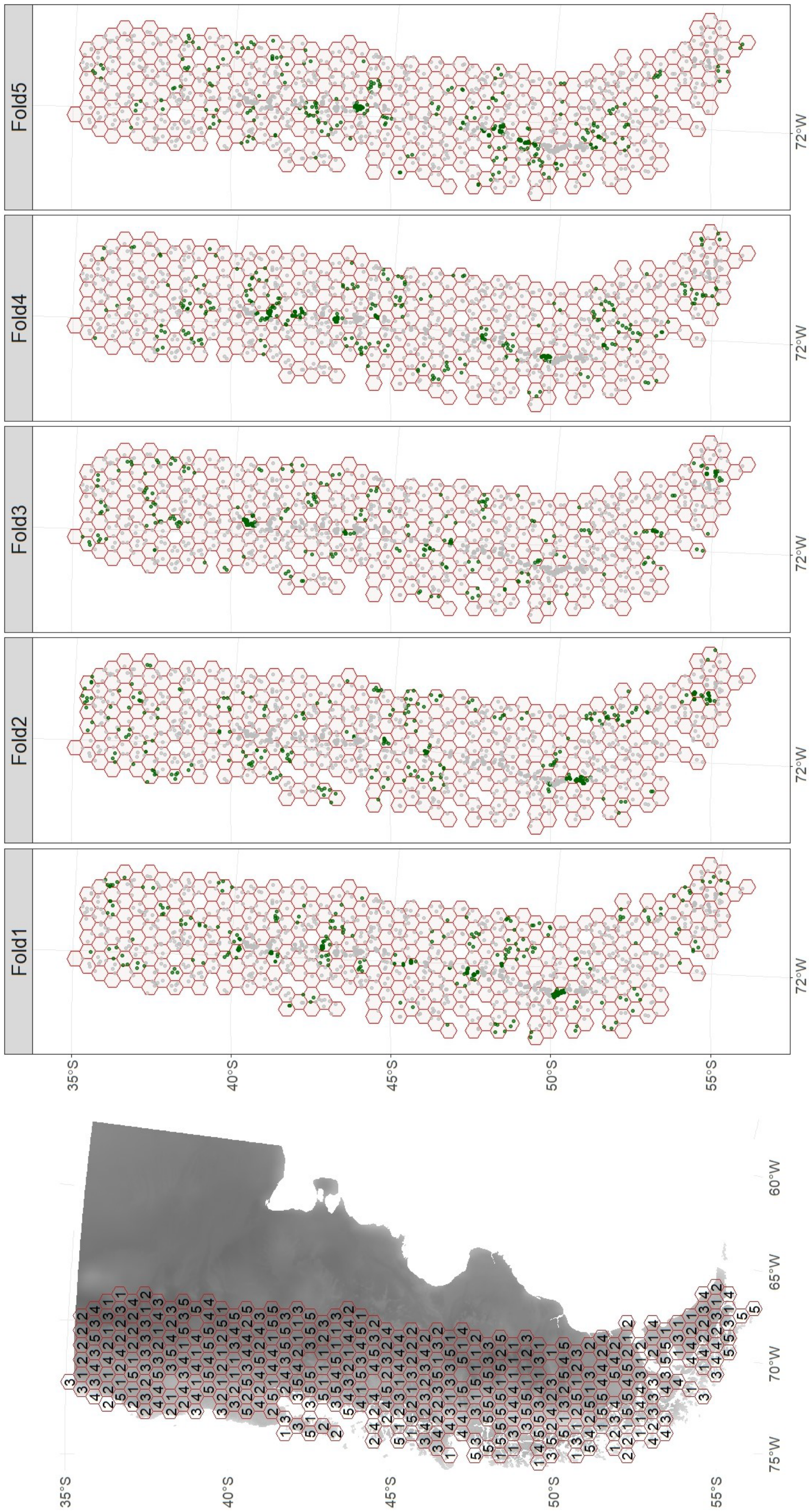
The results demonstrated that the IGTA produced less-biased species occurrence data compared to commonly used data from the Global Biodiversity Information Facility (GBIF). Moreover, the resulting IGTA point dataset proved valuable for validating large-scale remote sensing data of *N. pumilio* across its entire distribution range. Remote sensing data, in particular, can reduce spatial bias, as raster cells are sampled consistently across the landscape, rather than being concentrated in urban or tourist areas where most occurrences by Citizen Scientists or images by Instagram users are typically taken. While the advantages of the approach were clearly demonstrated above, it must be acknowledged that the manual search for suitable posts on Instagram and the georeferencing of occurrences are still time-consuming processes. Future research could explore the automation of this approach through the use of the Instagram API or the integration of AI-based image recognition techniques. Both options would require permission from the parent company, Meta, which has recently updated Instagram's terms of use to allow for AI usage (Meta, 2025). In conclusion, social media holds significant potential for species occurrence data sampling and can promote research on species in remote and high-elevation regions. The IGTA represents a novel and transferable method that can be applied to other species and regions.

In the next analysis step, both datasets were used for modelling, allowing for a direct comparison between models based on different types of species input data. This comparison highlighted the advantages of using continuous species data derived from remote sensing across the entire distribution range, as opposed to binary species data. Continuous data, here the proportion of 20 m raster cells classified as *N. pumilio* within each 1 km target resolution cell, can capture additional information related to land cover, topography, and species composition, thereby providing insights into the relative abundance of the species. In contrast, binary variables are limited to presence or absence (1 or 0), offering no detail on local density or coverage. The Random Forest regression model produced more meaningful predictor-variable relationships and yielded a more detailed spatial prediction.

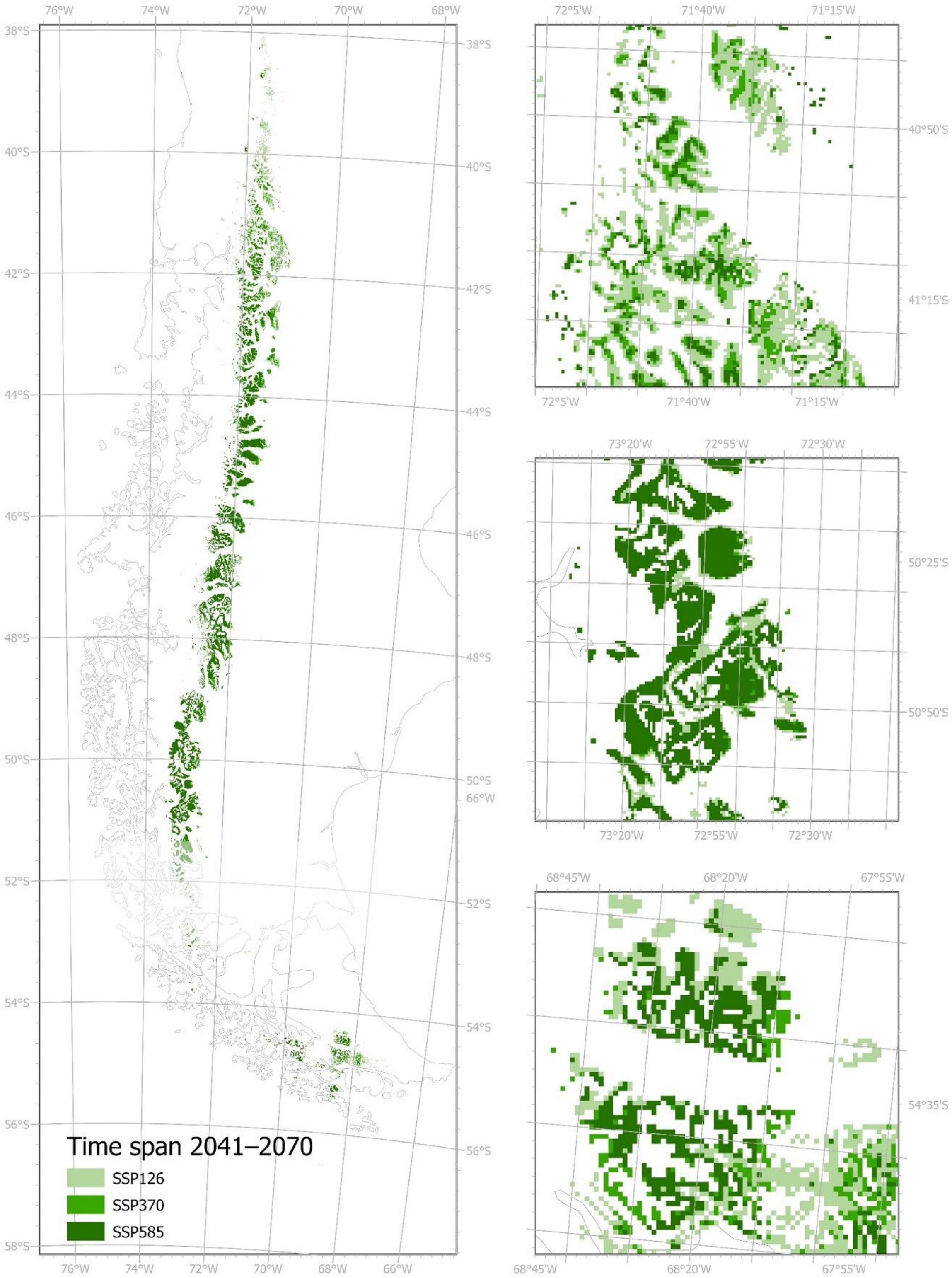
In this study, only bioclimatic variables were used to model the potential current and future distribution of *N. pumilio*. While this approach provides a comprehensive overview of temperature, precipitation, and seasonality, future research could expand the modelling framework to include additional abiotic variables, such as topography, soil characteristics, and wind exposure, as well as biotic and anthropogenic factors, including vegetation composition, fire regimes, grazing intensity, and forestry practices.

Up to date, this is the first study investigating the current state and future development of the potential distribution range of *N. pumilio* in the Southern Andes, thereby addressing an existing research gap. Common practices in ENM were successfully improved by introducing the IGTA, incorporating continuous remote sensing-derived data and using machine learning algorithms.

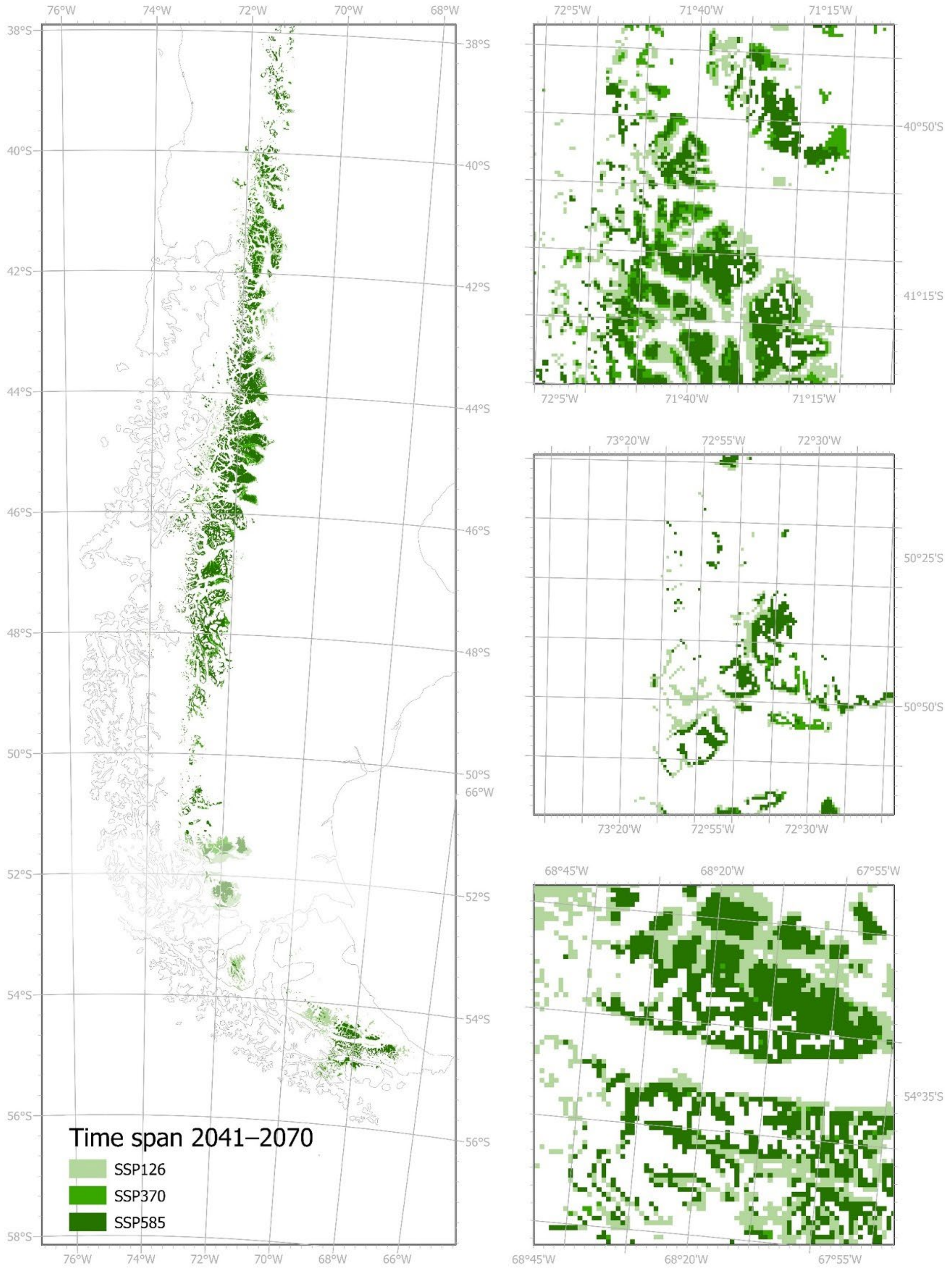
Appendix



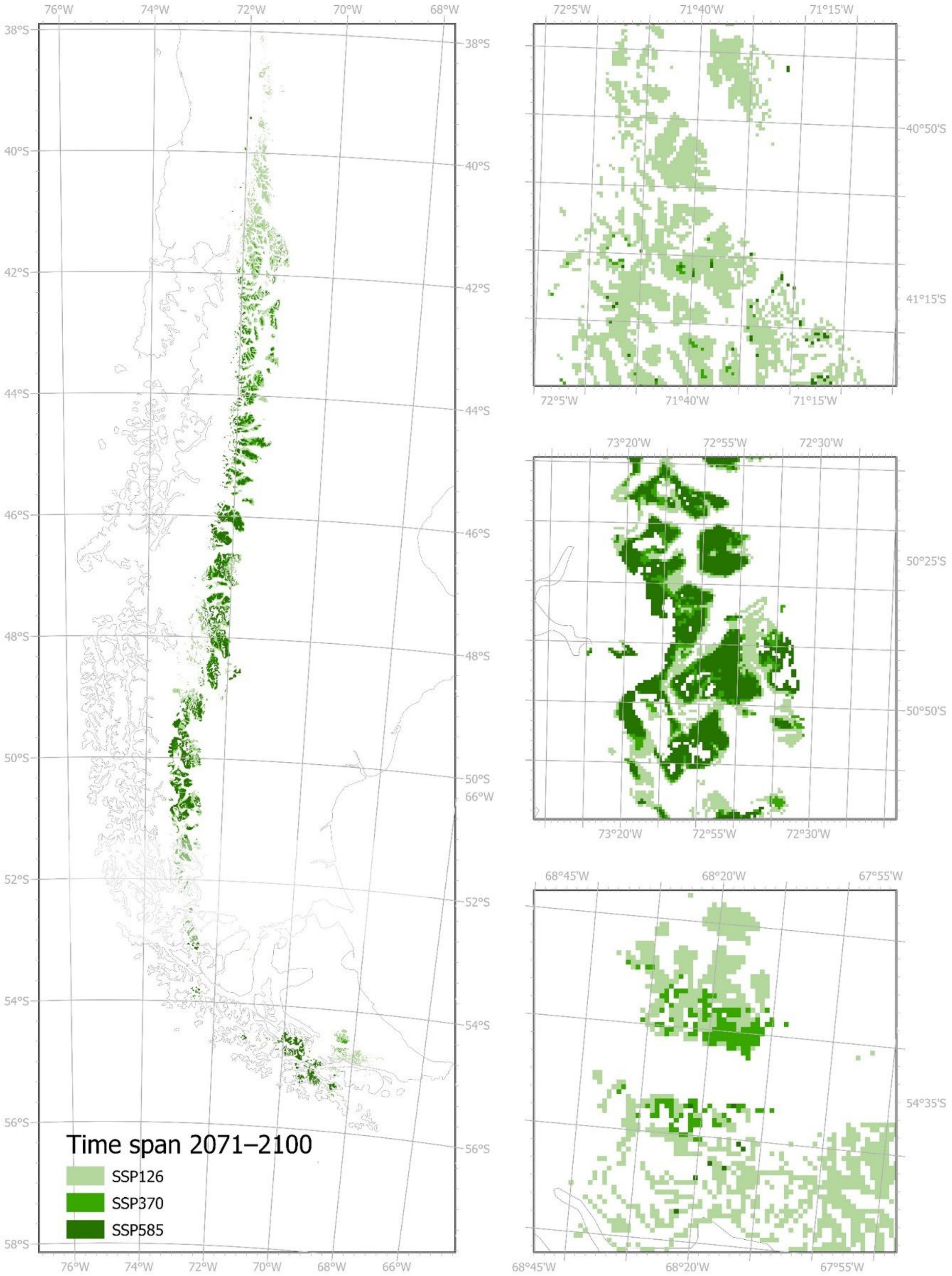
A1: Hexagonal spatial blocks used for 5-fold spatial cross-validation and the five folds with training data points (grey) and testing data points (dark green).



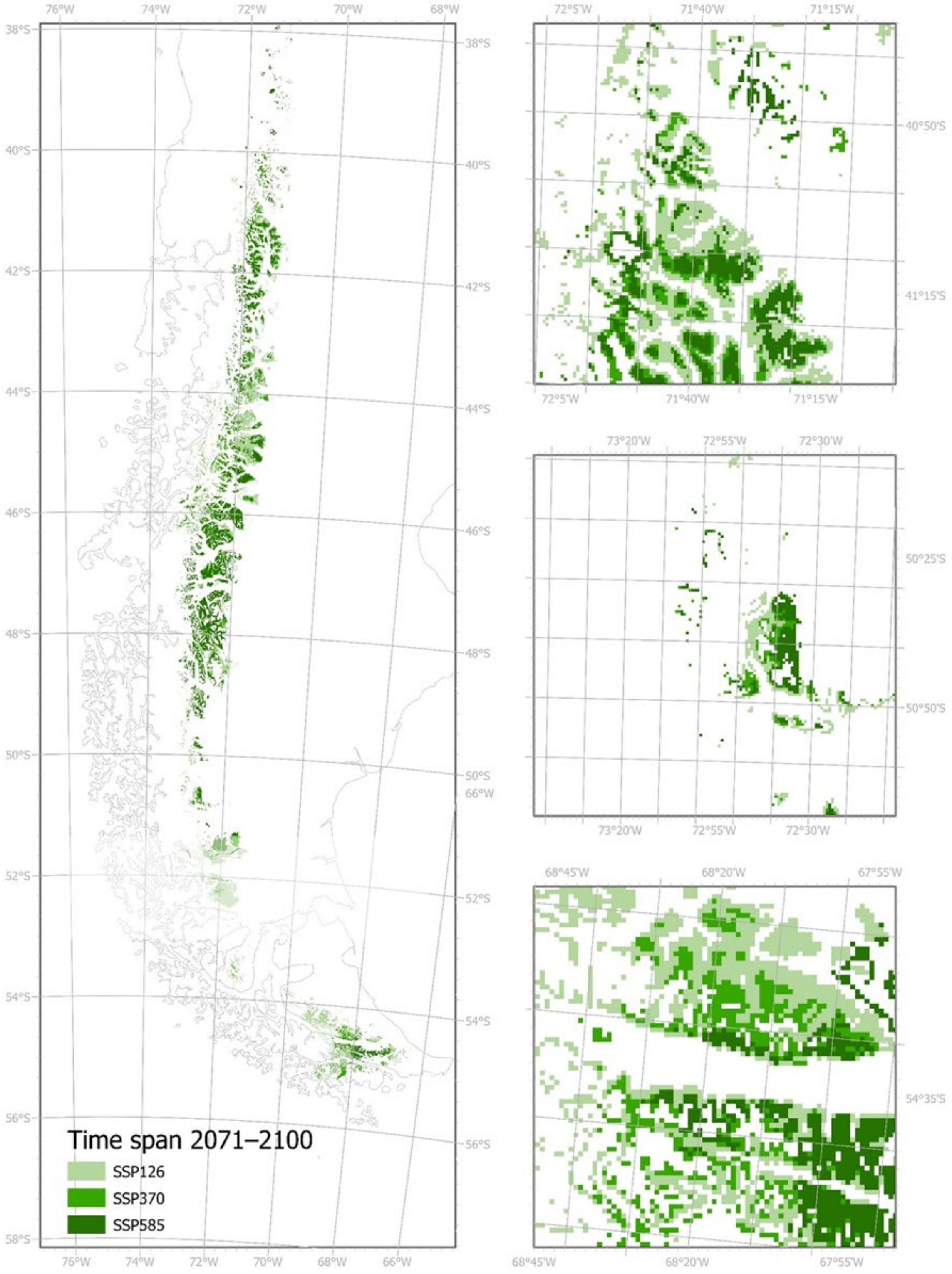
A2: Potential future distribution of Nothofagus pumilio modelled using the Random Forest classification model using CMIP6 SSP Scenarios for CHELSA Bioclim variables for the time period of 2041 to 2070.



A3: Potential future distribution of Nothofagus pumilio modelled using the Random Forest regression model using CMIP6 SSP Scenarios for CHELSA Bioclim variables for the time period of 2041 to 2070



A4: Potential future distribution of Nothofagus pumilio modelled using the Random Forest classification model using CMIP6 SSP Scenarios for CHELSA Bioclim variables for the time period of 2071 to 2100.



A5: Potential future distribution of Nothofagus pumilio modelled using the Random Forest regression model using CMIP6 SSP Scenarios for CHELSA Bioclim variables for the time period of 2071 to 2100.

A6: Estimated treeline elevation [m] at 13 locations based on the results of the Random Forest (RF) classification and RF regression models under current climate conditions and CMIP6 SSP scenarios for the future period 2071 to 2100 (applied thresholds 0.3 and 30 %). NA = not available (no data recorded).

Coordinates		Current Climate		SSP126 (2071–2100)		SSP370 (2071–2100)		SSP585 (2071–2100)	
X	**Y**	**Class.**	**Reg.**	**Class.**	**Reg.**	**Class.**	**Reg.**	**Class.**	**Reg.**
−71.00	−35.36	NA	NA	NA	NA	NA	NA	NA	NA
−71.11	−37.27	1988	1949	2328	NA	NA	NA	NA	2530
−71.33	−38.42	1854	1789	2071	1700	1780	1871	2460	2035
−72.15	−40.42	1591	1437	1679	1636	2026	1971	2026	2026
−72.19	−41.48	1500	1201	1555	1510	1917	1743	1917	1730
−71.45	−43.07	1839	1440	1955	1545	2059	1955	2059	2059
−71.42	−44.39	1320	1216	1703	1595	1952	1593	1952	1427
−72.24	−47.12	1361	1197	1439	1346	1742	1651	1901	1840
−72.30	−48.30	1522	1074	1399	1098	1578	1171	1698	1340
−72.54	−50.57	1176	956	1317	1000	1457	1229	1537	1287
−71.00	−53.00	543	560	592	728	NA	NA	NA	NA
−68.45	−54.17	544	520	648	616	615	757	NA	NA
−67.30	−54.57	610	610	609	614	NA	614	NA	614

References

Aceves-Bueno, E., Adeleye, A. S., Feraud, M., Huang, Y., Tao, M., Yang, Y., and Anderson, S. E. (2017). The Accuracy of Citizen Science Data: A Quantitative Review. *Bulletin of the Ecological Society of America*, *98*(4), 278–290. https://doi.org/10.1002/bes2.1336

Allen, M. R., Dube, O. P., Solecki, W., Aragón-Durand, F., Cramer, W., Humphreys, S., Kainuma, M., Kala, J., Mahowald, N., Mulugetta, Y., Perez, R., Wairiu, M., and Zickfeld, K. (2018). Framing and Context. In Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, V., Matthews, J., Chen, Y., Zhou, X., Gomis, M. I., Lonnoy, E., Maycock, T., Tignor, M. and Waterfield, T. (Eds.), *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* (pp. 49–92). Cambridge University Press. https://doi.org/10.1017/9781009157-940.003

Allouche, O., Tsoar, A., and Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*(6), 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Álvarez, C., Veblen, T. T., Christie, D. A., and González-Reyes, Á. (2015). Relationships between climate variability and radial growth of *Nothofagus pumilio* near altitudinal treeline in the Andes of northern Patagonia, Chile. *Forest Ecology and Management*, *342*, 112–121. https://doi.org/10.1016/j.foreco.2015.01.018

amb6767 (2018). *Two photos showing Nothofagus pumilio at Torres del Paine national park send per direct message*. https://www.instagram.com/amb6767/. Last accessed 02/11/2025.

Amigo, J., and Rodríguez-Guitián, M. A. (2011). Bioclimatic and phytosociological diagnosis of the species of the *Nothofagus* genus (Nothofagaceae) in South America. *International Journal of Geobotanical Research*, *1*(1), 1–20. https://doi.org/10.5616/ijgr110001

Anderson, R. P. (2012). Harnessing the world's biodiversity data: Promise and peril in ecological niche modeling of species distributions. *Annals of the New York Academy of Sciences*, *1260*, 66–80. https://doi.org/10.1111/j.1749-6632.2011.06440.x

Anderson, R. P., Araújo, M. B., Guisan, A., Lobo, J. M., Martinez-Meyer, E., Peterson, A. T., and Soberón, J. (2016). *Are species occurrence data in global online repositories fit for modeling species distributions? The case of the Global Biodiversity Information Facility (GBIF). Final Report of the Task Group on GBIF Data Fitness for Use in Distribution Modelling.* Global Biodiversity Information Facility (GBIF).

Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., and Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, *5*(1), eaat4858. https://doi.org/10.1126/sciadv.aat4858

Aravena, J. C., Lara, A., Wolodarsky-Franke, A., Villalba, R., and Cuq, E. (2002). Tree-ring growth patterns and temperature reconstruction from *Nothofagus pumilio* (Fagaceae) forests at the upper tree line of southern Chilean Patagonia. *Revista Chilena De Historia Natural*, *75*(2). https://doi.org/10.4067/S0716-078X2002000200008

Aschero, V., Srur, A. M., Guerrido, C., and Villalba, R. (2022). Contrasting climate influences on *Nothofagus pumilio* establishment along elevational gradients. *Plant Ecology*, *223*(4), 369–380. https://doi.org/10.1007/s11258-021-01211-8

Bach, A. J., and Price, L. W. (2013). Mountain Climate. In Price, M. F., Byers, A. C., Friend, D. A., Kohler, T. and Price, L. W. (Eds.), *Mountain geography: Physical and human dimensions.* University of California Press.

Bader, M. Y., Llambí, L. D., Case, B. S., Buckley, H. L., Toivonen, J. M., Camarero, J. J., Cairns, D. M., Brown, C. D., Wiegand, T., and Resler, L. M. (2021). A global framework for linking alpine-treeline ecotone patterns to underlying processes. *Ecography*, *44*(2), 265–292. https://doi.org/-10.1111/ecog.05285

Bambach, N. E., Rhoades, A. M., Hatchett, B. J., Jones, A. D., Ullrich, P. A., and Zarzycki, C. M. (2022). Projecting climate change in South America using variable-resolution Community Earth System Model: An application to Chile. *International Journal of Climatology*, *42*(4), 2514–2542. https://doi.org/10.1002/joc.7379

Barbet-Massin, M., Jiguet, F., Albert, C. H., and Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, *3*(2), 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Batllori, E., Camarero, J. J., Ninot, J. M., and Gutiérrez, E. (2009). Seedling recruitment, survival and facilitation in alpine *Pinus uncinata* tree line ecotones. Implications and potential responses to climate warming. *Global Ecology and Biogeography*, *18*(4), 460–472. https://doi.org/10.1111/j.1466-8238.2009.00464.x

Beck, H. E., McVicar, T. R., Vergopolan, N., Berg, A., Lutsko, N. J., Dufour, A., Zeng, Z., Jiang, X., van Dijk, A. I. J. M., and Miralles, D. G. (2023). High-resolution (1 km) Köppen-Geiger maps for 1901-2099 based on constrained CMIP6 projections. *Scientific Data*, *10*(1), 724. https://doi.org/10.1038/s41597-023-02549-6

Beck, J., Böller, M., Erhardt, A., and Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, *19*, 10–15. https://doi.org/10.1016/j.ecoinf.2013.11.002

Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., Stuart-Smith, R. D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J. F., Pecl, G. T., Barrett, N., and Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, *173*, 144–154. https://doi.org/10.1016/j.biocon.2013.07.037

Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K., and Mace, G. M. (2010). Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biology*, *8*(6), e1000385. https://doi.org/10.1371/journal.pbio.1000385

Bobrowski, M., Gerlitz, L., and Schickhoff, U. (2017). Modelling the potential distribution of *Betula utilis* in the Himalaya. *Global Ecology and Conservation*, *11*, 69–83. https://doi.org/10.1016/j.gecco.2017.04.003

Bobrowski, M. (2018). *Modelling the ecological niche of a treeline species (Betula utilis) in the Himalayan region (Dissertation).* Universität Hamburg, Fachbereich Geowissenschaften.

Bobrowski, M., Weidinger, J., and Schickhoff, U. (2021a). Is New Always Better? Frontiers in Global Climate Datasets for Modeling Treeline Species in the Himalayas. *Atmosphere*, *12*(5), 543. https://doi.org/10.3390/atmos12050543

Bobrowski, M., Weidinger, J., Schwab, N., and Schickhoff, U. (2021b). Searching for ecology in species distribution models in the Himalayas. *Ecological Modelling*, *458*, 109693. https://doi.org/10.1016/j.ecolmodel.2021.109693

Bonney, R. (1996). Citizen science: A lab tradition. *Living Bird*, *15*(4), 7–15.

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., and Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience*, *59*(11), 977–984. https://doi.org/10.1525/bio.2009.59.11.9

Booth, T. H., Nix, H. A., Busby, J. R., and Hutchinson, M. F. (2014). bioclim : the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and Distributions*, *20*(1), 1–9. https://doi.org/10.1111/ddi.12144

Borsdorf, A., and Stadel, C. (2013). *Die Anden*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-8274-2458-7

Bozkurt, D., Rojas, M., Boisier, J. P., and Valdivieso, J. (2018). Projected hydroclimate changes over Andean basins in central Chile from downscaled CMIP5 models under the low and high emission scenarios. *Climatic Change*, *150*(3-4), 131–147. https://doi.org/10.1007/s10584-018-2246-7

Brand, R., Srur, A. M., and Villalba, R. (2022). Contrasting growth trends in *Nothofagus pumilio* upper-elevation forests induced by climate warming in the Southern Andes. *Agricultural and Forest Meteorology*, *323*, 109083. https://doi.org/10.1016/j.agrformet.2022.109083

Breiman, L. (2001a). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L. (2001b). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3). https://doi.org/10.1214/-ss/1009213726

Breiman, L., Cutler, A., Liaw, A., and Wiener, M. (2002). *CRAN: Contributed Packages. randomForest: Breiman and Cutlers Random Forests for Classification and Regression.* https://doi.org/10.32614/CRAN.package.randomForest

Brenning, A., Bangs, D., and Becker, M. (2022). *RSAGA (version 1.4.0)*. CRAN: Contributed Packages. https://doi.org/10.32614/CRAN.package.RSAGA

Broussin, J., Mouchet, M., and Goberville, E. (2024). Generating pseudo-absences in the ecological space improves the biological relevance of response curves in species distribution models. *Ecological Modelling*, *498*, 110865. https://doi.org/10.1016/j.ecolmodel.2024.110865

Bürzle, B., Schickhoff, U., Schwab, N., Wernicke, L. M., Müller, Y. K., Böhner, J., Chaudhary, R. P., Scholten, T., and Oldeland, J. (2018). Seedling recruitment and facilitation dependence on safe site characteristics in a Himalayan treeline ecotone. *Plant Ecology*, *219*(2), 115–132. https://doi.org/10.1007/s11258-017-0782-2

Cai, W., McPhaden, M. J., Grimm, A. M., Rodrigues, R. R., Taschetto, A. S., Garreaud, R. D., Dewitte, B., Poveda, G., Ham, Y.-G., Santoso, A., Ng, B., Anderson, W., Wang, G., Geng, T., Jo, H.-S., Marengo, J. A., Alves, L. M., Osman, M., Li, S., Wu, L., Karamperidou, C., Takahashi, K., and Vera, C. (2020). Climate impacts of the El Niño–Southern Oscillation on South America. *Nature Reviews Earth & Environment*, *1*(4), 215–231. https://doi.org/10.1038/s43017-020-0040-3

Čengić, M., Rost, J., Remenska, D., Janse, J. H., Huijbregts, M. A. J., and Schipper, A. M. (2020). On the importance of predictor choice, modelling technique, and number of pseudo-absences for bioclimatic envelope model performance. *Ecology and Evolution*, *10*(21), 12307–12317. https://doi.org/10.1002/ece3.6859

Chauvier, Y., Zimmermann, N. E., Poggiato, G., Bystrova, D., Brun, P., and Thuiller, W. (2021). Novel methods to correct for observer and sampling bias in presence-only species distribution models. *Global Ecology and Biogeography*, *30*(11), 2312–2325. https://doi.org/10.1111/geb.13383

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, *8*(7), 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015

Cox, T. E., Philippoff, J., Baumgartner, E., and Smith, C. M. (2012). Expert variability provides perspective on the strengths and weaknesses of citizen-driven intertidal monitoring program. *Ecological Applications : A Publication of the Ecological Society of America*, *22*(4), 1201–1212. https://doi.org/10.1890/11-1614.1

Cuevas, J. G. (2000). Tree recruitment at the *Nothofagus pumilio* alpine timberline in Tierra del Fuego, Chile. *Journal of Ecology*, *88*(5), 840–855. https://doi.org/10.1046/j.1365-2745.2000.00497.x

Daniels, L. D., and Veblen, T. T. (2003). Regional and local effects of disturbance and climate on altitudinal treelines in northern Patagonia. *Journal of Vegetation Science*, *14*(5), 733–742. https://doi.org/10.1111/j.1654-1103.2003.tb02205.x

Daniels, L. D., and Veblen, T. T. (2004). Spatiotemporal influences of climate on altitudinal treeline in northern Patagonia. *Ecology*, *85*(5), 1284–1296. https://doi.org/10.1890/03-0092

Daru, B. H., Park, D. S., Primack, R. B., Willis, C. G., Barrington, D. S., Whitfeld, T. J. S., Seidler, T. G., Sweeney, P. W., Foster, D. R., Ellison, A. M., and Davis, C. C. (2018). Widespread sampling biases in herbaria revealed from large-scale digitization. *The New Phytologist*, *217*(2), 939–955. https://doi.org/10.1111/nph.14855

De Langhe, J. (2015). *Neat leaves, with the secondary veins ending in a sinus on Nothofagus pumilio (Arboretum Keracoual, France – 19 August 2015).* Arboretum Wespelaar. https://www.treesandshrubsonline.org/articles/nothofagus/nothofagus-pumilio/. Last accessed 02/11/2025.

Deng, D.-P., Chuang, T.-R., Shao, K.-T., Mai, G.-S., Lin, T.-E., Lemmens, R., Hsu, C.-H., Lin, H.-H., and Kraak, M.-J. (2012). Using social media for collaborative species identification and occurrence. In Goodchild, M. (Ed.), *ACM Conferences, Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information* (pp. 22–29). ACM. https://doi.org/10.1145/2442952.2442957

Di Cecco, G. J., Barve, V., Belitz, M. W., Stucky, B. J., Guralnick, R. P., and Hurlbert, A. H. (2021). Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science. *BioScience*, *71*(11), 1179–1188. https://doi.org/10.1093/biosci/biab093

Dickinson, J. L., Zuckerberg, B., and Bonter, D. N. (2010). Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics*, *41*(1), 149–172. https://doi.org/10.1146/annurev-ecolsys-102209-144636

Dimson, M., and Gillespie, T. W. (2023). Who, where, when: Observer behavior influences spatial and temporal patterns of iNaturalist participation. *Applied Geography*, *153*, 102916. https://doi.org/10.1016/j.apgeog.2023.102916

Dirnböck, T., Essl, F., and Rabitsch, W. (2011). Disproportional risk for habitat loss of high-altitude endemic species under climate change. *Global Change Biology*, *17*(2), 990–996. https://doi.org/10.1111/j.1365-2486.2010.02266.x

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dormann, C. F., M. McPherson, J., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W., Kühn, I., Ohlemüller, R., R. Peres-Neto, P., Reineking, B., Schröder, B., M. Schurr, F., and Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, *30*(5), 609–628. https://doi.org/10.1111/j.2007.0906-7590.05171.x

Dullinger, S., Gattringer, A., Thuiller, W., Moser, D., Zimmermann, N. E., Guisan, A., Willner, W., Plutzar, C., Leitner, M., Mang, T., Caccianiga, M., Dirnböck, T., Ertl, S., Fischer, A., Lenoir, J., Svenning, J.-C., Psomas, A., Schmatz, D. R., Silc, U., Vittoz, P., and Hülber, K. (2012). Extinction debt of high-mountain plants under twenty-first-century climate change. *Nature Climate Change*, *2*(8), 619–622. https://doi.org/10.1038/nclimate1514

Edwards, T., Jones, C. B., Perkins, S. E., and Corcoran, P. (2021). Passive citizen science: The role of social media in wildlife observations. *PloS One*, *16*(8), e0255416. https://doi.org/10.1371/journal.pone.0255416

Elith, J., and Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 677–697. https://doi.org/10.1146/-annurev.ecolsys.110308.120159

ElQadi, M. M., Dorin, A., Dyer, A., Burd, M., Bukovac, Z., and Shrestha, M. (2017). Mapping species distributions with social media geo-tagged images: Case studies of bees and flowering plants in Australia. *Ecological Informatics*, *39*, 23–31. https://doi.org/10.1016/j.ecoinf.2017.02.006

Elton, C. S. (1927). *Animal Ecology*. The Macmillan Company.

ESRI (2020). *Arc GIS Pro: Version 2.7.0*. https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview. Last accessed 27/10/2025.

Fajardo, A., Gazol, A., Mayr, C., and Camarero, J. J. (2019). Recent decadal drought reverts warming-triggered growth enhancement in contrasting climates in the southern Andes tree line. *Journal of Biogeography*, *46*(7), 1367–1379. https://doi.org/10.1111/jbi.13580

Fajardo, A., and Piper, F. I. (2014). An experimental approach to explain the southern Andes elevational treeline. *American Journal of Botany*, *101*(5), 788–795. https://doi.org/10.3732/ajb.1400166

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., and Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, *186*, 64–87. https://doi.org/10.1016/j.rse.2016.08.013

Feldman, M. J., Imbeau, L., Marchand, P., Mazerolle, M. J., Darveau, M., and Fenton, N. J. (2021). Trends and gaps in the use of citizen science derived data as input for species distribution models: A quantitative review. *PloS One*, *16*(3), e0234587. https://doi.org/10.1371/journal.pone.0234587

Feng, X., Park, D. S., Walker, C., Peterson, A. T., Merow, C., and Papeş, M. (2019). A checklist for maximizing reproducibility of ecological niche models. *Nature Ecology & Evolution*, *3*(10), 1382–1395. https://doi.org/10.1038/s41559-019-0972-5

fernando.v.fotografia (2022). Selected post at Laguna Capri in Argentina with the species visible in front. https://www.instagram.com/p/ChLmWxQuVjh/. Last accessed 12/02/2026.

Fick, S. E., and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Field, A., Miles, J., and Field, Z. (2012). *Discovering statistics using R / Andy Field, Jeremy Miles, Zoë Field*. *Core textbook*. Sage.

Fitzpatrick, M. C., Preisser, E. L., Ellison, A. M., and Elkinton, J. S. (2009). Observer bias and the detection of low-density populations. *Ecological Applications : A Publication of the Ecological Society of America*, *19*(7), 1673–1679. https://doi.org/10.1890/09-0265.1

Fox, N., Graham, L. J., Eigenbrod, F., Bullock, J. M., and Parks, K. E. (2022). Geodiversity Supports Cultural Ecosystem Services: an Assessment Using Social Media. *Geoheritage*, *14*(1). https://doi.org/10.1007/s12371-022-00665-0

Franklin, J. (2010). *Mapping species distributions: Spatial inference and prediction* (7. printing). *Ecology, biodiversity and conservation*. Cambridge Univ. Press. https://doi.org/10.1017/CBO9780511810602

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5). https://doi.org/10.1214/aos/1013203451

Fyfe, J. C., and Saenko, O. A. (2006). Simulated changes in the extratropical Southern Hemisphere winds and currents. *Geophysical Research Letters*, *33*(6), Article 2005GL025332. https://doi.org/10.1029/2005GL025332

Garreaud, R. D. (2009). The Andes climate and weather. *Advances in Geosciences*, *22*, 3–11. https://doi.org/10.5194/adgeo-22-3-2009

Garreaud, R. D. (2018). Record-breaking climate anomalies lead to severe drought and environmental disruption in western Patagonia in 2016. *Climate Research*, *74*(3), 217–229. https://doi.org/10.3354/cr01505

Garreaud, R. D., Alvarez-Garreton, C., Barichivich, J., Boisier, J. P., Christie, D., Galleguillos, M., LeQuesne, C., McPhee, J., and Zambrano-Bigiarini, M. (2017). The 2010–2015 megadrought in central Chile: impacts on regional hydroclimate and vegetation. *Hydrology and Earth System Sciences*, *21*(12), 6307–6327. https://doi.org/10.5194/hess-21-6307-2017

Garreaud, R. D., Boisier, J. P., Rondanelli, R., Montecinos, A., Sepúlveda, H. H., and Veloso-Aguila, D. (2020). The Central Chile Mega Drought (2010–2018): A climate dynamics perspective. *International Journal of Climatology*, *40*(1), 421–439. https://doi.org/10.1002/joc.6219

Garreaud, R. D., Lopez, P., Minvielle, M., and Rojas, M. (2013). Large-Scale Control on the Patagonian Climate. *Journal of Climate*, *26*(1), 215–230. https://doi.org/10.1175/JCLI-D-12-00001.1

GBIF (2024). *Occurrence Download (6 June 2024).* https://doi.org/10.15468/dl.3kv29u

GBIF (2025a). *GBIF | Global Biodiversity Information Facility: Free and open access to biodiversity data*. https://www.gbif.org/. Last accessed 27/10/2025.

GBIF (2025b). *iNaturalist.org*. https://www.gbif.org/publisher/28eb1a3f-1c15-4a95-931a-4af90ecb574d/metrics. Last accessed 27/10/2025.

Genuer, R., Poggi, J.-M., and Tuleau-Malot, C. (2015). VSURF: An R Package for Variable Selection Using Random Forests. *The R Journal*, *7*(2), 19. https://doi.org/10.32614/RJ-2015-018

Gibson, C. E., Williams, D., Dunlop, R., and Beck, S. (2020). Using social media as a cost-effective resource in the photo-identification of a coastal bottlenose dolphin community. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *30*(8), 1702–1710. https://doi.org/10.1002/aqc.3356

GitHub Inc. (2025). *ClimDatDownloadR*. https://github.com/HelgeJentsch/ClimDatDownloadR. Last accessed 27/10/2025.

Goldberg, J. K. (2023). Inaturalist is an open science resource for ecological genomics by enabling rapid and tractable records of initial observations of sequenced biological samples. *Biology Letters*, *19*(10), 20230251. https://doi.org/10.1098/rsbl.2023.0251

Graham, A. (2009). THE ANDES: A GEOLOGICAL OVERVIEW FROM A BIOLOGICAL PERSPECTIVE. *Annals of the Missouri Botanical Garden*, *96*(3), 371–385. https://doi.org/10.3417/2007146

Greenwell, B. (2024). *CRAN: Contributed Packages.* https://doi.org/10.32614/CRAN.package.fastshap

Grinnell, J. (1917). The Niche-Relationships of the California Thrasher. *The Auk*, *34*(4), 427–433. https://doi.org/10.2307/4072271

Hadley, K. S., Price, L. W., and Grabherr, G. (2013). Mountain Vegetation. In Price, M. F., Byers, A. C., Friend, D. A., Kohler, T. and Price, L. W. (Eds.), *Mountain geography: Physical and human dimensions.* University of California Press.

Haklay, M., Fraisl, D., Greshake Tzovaras, B., Hecker, S., Gold, M., Hager, G., Ceccaroni, L., Kieslinger, B., Wehn, U., Woods, S., Nold, C., Balázs, B., Mazzonetto, M., Ruefenacht, S., Shanley, L. A., Wagenknecht, K., Motion, A., Sforzi, A., Riemenschneider, D., Dorler, D., Heigl, F., Schaefer, T., Lindner, A., Weißpflug, M., Mačiulienė, M., and Vohland, K. (2021). Contours of citizen science: A vignette study. *Royal Society Open Science*, *8*(8), 202108. https://doi.org/10.1098/rsos.202108

Hansson, A., Dargusch, P., and Shulmeister, J. (2021). A review of modern treeline migration, the factors controlling it and the implications for carbon storage. *Journal of Mountain Science*, *18*(2), 291–306. https://doi.org/10.1007/s11629-020-6221-1

Hansson, A., Yang, W.-H., Dargusch, P., and Shulmeister, J. (2023). Investigation of the Relationship Between Treeline Migration and Changes in Temperature and Precipitation for the Northern Hemisphere and Sub-regions. *Current Forestry Reports*, *9*(2), 72–100. https://doi.org/10.1007/s40725-023-00180-7

Harsch, M. A., and Bader, M. Y. (2011). Treeline form - a potential key to understanding treeline dynamics. *Global Ecology and Biogeography*, *20*(4), 582–596. https://doi.org/10.1111/j.1466-8238.2010.00622.x

Harsch, M. A., Hulme, P. E., McGlone, M. S., and Duncan, R. P. (2009). Are treelines advancing? A global meta-analysis of treeline response to climate warming. *Ecology Letters*, *12*(10), 1040–1049. https://doi.org/10.1111/j.1461-0248.2009.01355.x

He, K. S., Bradley, B. A., Cord, A. F., Rocchini, D., Tuanmu, M.-N., Schmidtlein, S., Turner, W., Wegmann, M., and Pettorelli, N. (2015). Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation*, *1*(1), 4–18. https://doi.org/10.1002/rse2.7

Heberling, J. M., and Isaac, B. L. (2018). Inaturalist as a tool to expand the research value of museum specimens. *Applications in Plant Sciences*, *6*(11), e01193. https://doi.org/10.1002/aps3.1193

Hentati-Sundberg, J., and Olsson, O. (2016). Amateur photographs reveal population history of a colonial seabird. *Current Biology : CB*, *26*(6), R226-8. https://doi.org/10.1016/j.cub.2016.02.007

Hertel, D., Therburg, A., and Villalba, R. (2008). Above- and below-ground response by *Nothofagus pumilio* to climatic conditions at the transition from the steppe–forest boundary to the alpine treeline in southern Patagonia, Argentina. *Plant Ecology & Diversity*, *1*(1), 21–33. https://doi.org/10.1080/17550870802257026

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Hildebrand-Vogel, R., Godoy, R., and Vogel, A. (1990). Subantarctic-Andean *Nothofagus pumilio* forests. *Plant Ecology*, *89*(1), 55–68. https://doi.org/10.1007/BF00134434

Holtmeier, F.-K. (2009). *Mountain timberlines: Ecology, patchiness, and dynamics*. *Advances in global change research: Vol. 36*. Springer.

Holtmeier, F.-K., and Broll, G. (2005). Sensitivity and response of northern hemisphere altitudinal and polar treelines to environmental change at landscape and local scales. *Global Ecology and Biogeography*, *14*(5), 395–410. https://doi.org/10.1111/j.1466-822X.2005.00168.x

Holtmeier, F.-K., and Broll, G. (2020). Treeline Research—From the Roots of the Past to Present Time. A Review. *Forests*, *11*(1), 38. https://doi.org/10.3390/f11010038

Holz, A., Paritsis, J., Mundo, I. A., Veblen, T. T., Kitzberger, T., Williamson, G. J., Aráoz, E., Bustos-Schindler, C., González, M. E., Grau, H. R., and Quezada, J. M. (2017). Southern Annular Mode drives multicentury wildfire activity in southern South America. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(36), 9552–9557. https://doi.org/10.1073/pnas.1705168114

Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, *22*(0), 415–427. https://doi.org/10.1101/SQB.1957.022.01.039

Immitzer, M., Neuwirth, M., Böck, S., Brenner, H., Vuolo, F., and Atzberger, C. (2019). Optimal Input Features for Tree Species Classification in Central Europe Based on Multi-Temporal Sentinel-2 Data. *Remote Sensing*, *11*(22), 2599. https://doi.org/10.3390/rs11222599

Immitzer, M., Vuolo, F., and Atzberger, C. (2016). First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing*, *8*(3), 166. https://doi.org/10.3390/rs8030166

iNaturalist (2025). *Wie es funktioniert*. https://www.inaturalist.org/. Last accessed 30/10/2025.

IPCC (2023). *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.* https://doi.org/10.59327/IPCC/AR6-9789291691647

Jarić, I., Correia, R. A., Brook, B. W., Buettel, J. C., Courchamp, F., Di Minin, E., Firth, J. A., Gaston, K. J., Jepson, P., Kalinkat, G., Ladle, R., Soriano-Redondo, A., Souza, A. T., and Roll, U. (2020). Iecology: Harnessing Large Online Resources to Generate Ecological Insights. *Trends in Ecology & Evolution*, *35*(7), 630–639. https://doi.org/10.1016/j.tree.2020.03.003

JAXA, EORC (Japan Aerospace Exploration Agency, Earth Observation Research Center) (2025). AW3D30 DSM data map. https://www.eorc.jaxa.jp/ALOS/en/aw3d30/data/index.htm. Last accessed 04/11/2025.

Jentsch, H. (2025). *HelgeJentsch/ClimDatDownloadR: 1.0.0* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.15663182

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., and Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*, 170122. https://doi.org/10.1038/sdata.2017.122

Karger, D. N., Wilson, A. M., Mahony, C., Zimmermann, N. E., and Jetz, W. (2021). Global daily 1 km land surface precipitation based on cloud cover-informed downscaling. *Scientific Data*, *8*(1), 307. https://doi.org/10.1038/s41597-021-01084-6

Körner, C. (2007). Climatic treelines: conventions, global patterns, causes. *Erdkunde*, *61*(4), 316–324. https://doi.org/10.3112/erdkunde.2007.04.02

Körner, C. (2012). *Alpine treelines: Functional ecology of the global high elevation tree limits ; with 22 tables*. Springer. https://doi.org/10.1007/978-3-0348-0396-0

Körner, C. (2020). Climatic Controls of the Global High Elevation Treelines. In *Encyclopedia of the World's Biomes* (pp. 275–281). Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.11998-0

Körner, C. (2021). *Alpine Plant Life*. Springer International Publishing. https://doi.org/10.1007/978-3-030-59538-8

Kullenberg, C., and Kasperowski, D. (2016). What Is Citizen Science?--A Scientometric Meta-Analysis. *PloS One*, *11*(1), e0147152. https://doi.org/10.1371/journal.pone.0147152

La Sorte, F. A., Cohen, J. M., and Jetz, W. (2024). Data coverage, biases, and trends in a global citizen-science resource for monitoring avian diversity. *Diversity and Distributions*, *30*(8), Article e13863. https://doi.org/10.1111/ddi.13863

Lara, A., Aravena, J. C., Villalba, R., Wolodarsky-Franke, A., Luckman, B., and Wilson, R. (2001). Dendroclimatology of high-elevation *Nothofagus pumilio* forests at their northern distribution limit in the central Andes of Chile. *Canadian Journal of Forest Research*, *31*(6), 925–936. https://doi.org/10.1139/cjfr-31-6-925

Lara, A., Villalba, R., Wolodarsky-Franke, A., Aravena, J. C., Luckman, B. H., and Cuq, E. (2005). Spatial and temporal variation in *Nothofagus pumilio* growth at tree line along its latitudinal range (35°40′–55° S) in the Chilean Andes. *Journal of Biogeography*, *32*(5), 879–893. https://doi.org/10.1111/j.1365-2699.2005.01191.x

Lawson, C. R., Hodgson, J. A., Wilson, R. J., and Richards, S. A. (2014). Prevalence, thresholds and the performance of presence–absence models. *Methods in Ecology and Evolution*, *5*(1), 54–64. https://doi.org/10.1111/2041-210X.12123

Legendre, P. (1993). Spatial Autocorrelation: Trouble or New Paradigm? *Ecology*, *74*(6), 1659–1673. https://doi.org/10.2307/1939924

Lett, S., and Dorrepaal, E. (2018). Global drivers of tree seedling establishment at alpine treelines in a changing climate. *Functional Ecology*, *32*(7), 1666–1680. https://doi.org/10.1111/1365-2435.13137

Li, M., Sun, H., Huang, Y., and Chen, H. (2024). Shapley value: from cooperative game to explainable artificial intelligence. *Autonomous Intelligent Systems*, *4*(1). https://doi.org/10.1007/s43684-023-00060-8

Li, X., and Wang, Y. (2013). Applying various algorithms for species distribution modelling. *Integrative Zoology*, *8*(2), 124–135. https://doi.org/10.1111/1749-4877.12000

Liu, C., White, M., and Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, *40*(4), 778–789. https://doi.org/10.1111/jbi.12058

Lu, X., Liang, E., Wang, Y., Babst, F., and Camarero, J. J. (2021). Mountain treelines climb slowly despite rapid climate warming. *Global Ecology and Biogeography*, *30*(1), 305–315. https://doi.org/10.1111/geb.13214

Manos, P. S. (1997). Systematics of *Nothofagus* (*Nothofagaceae*) based on rDNA spacer sequences (ITS): taxonomic congruence with morphology and plastid sequences. *American Journal of Botany*, *84*(8), 1137–1155. https://doi.org/10.2307/2446156

Marcer, A., Chapman, A. D., Wieczorek, J. R., Xavier Picó, F., Uribe, F., Waller, J., and Ariño, A. H. (2022). Uncertainty matters: ascertaining where specimens in natural history collections come from and its implications for predicting species distributions. *Ecography*, *2022*(9), Article e06025. https://doi.org/10.1111/-ecog.06025

Marshall, B. M., and Strine, C. T. (2019). Exploring snake occurrence records: Spatial biases and marginal gains from accessible social media. *PeerJ*, *7*, e8059. https://doi.org/10.7717/peerj.8059

Martino, S., Pace, D. S., Moro, S., Casoli, E., Ventura, D., Frachea, A., Silvestri, M., Arcangeli, A., Giacomini, G., Ardizzone, G., and Jona Lasinio, G. (2021). Integration of presence-only data from several sources: a case study on dolphins' spatial distribution. *Ecography*, *44*(10), 1533–1543. https://doi.org/10.1111/-ecog.05843

Masiokas, M., Rabatel, A., Rivera, A., Ruiz, L., Pitte, P., Ceballos, J. L., Barcaza, G., Soruco, A., Bown, F., Berthier, E., Dussaillant, I., and MacDonell, S. (2020). A Review of the Current State and Recent Changes of the Andean Cryosphere. *Frontiers in Earth Science*, *8*, Article 99. https://doi.org/10.3389/feart.2020.00099

Masiokas, M., and Villalba, R. (2004). Climatic significance of intra-annual bands in the wood of *Nothofagus pumilio* in southern Patagonia. *Trees*, *18*(6), 696–704. https://doi.org/10.1007/s00468-004-0355-6

Mesaglio, T., and Callaghan, C. T. (2021). An overview of the history, current contributions and future outlook of iNaturalist in Australia. *Wildlife Research*, *48*(4), 289–303. https://doi.org/10.1071/WR20154

Meta (2025). *Terms of Use (November 2025)*. https://www.facebook.com/help/-instagram/581066165581870/?helpref=hc_fnav. Last accessed 02/11/2025.

Meyer, C., Weigelt, P., and Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, *19*(8), 992–1006. https://doi.org/10.1111/ele.12624

Michener, W. K., Allard, S., Budden, A., Cook, R. B., Douglass, K., Frame, M., Kelling, S., Koskela, R., Tenopir, C., and Vieglais, D. A. (2012). Participatory design of DataONE—Enabling cyberinfrastructure for the biological and environmental sciences. *Ecological Informatics*, *11*, 5–15. https://doi.org/10.1016/j.ecoinf.2011.08.007

Miehe, G., Mao, K., Hasson, S. u., Böhner, J., and Schickhoff, U. (2023). What do we know about treelines of the Anthropocene in High Asia? *Plant Diversity.* Advance online publication. https://doi.org/10.1016/j.pld.2023.08.005

Miller, J. (2010). Species Distribution Modeling. *Geography Compass*, *4*(6), 490–509. https://doi.org/10.1111/j.1749-8198.2010.00351.x

Miller, R. L., Schmidt, G. A., and Shindell, D. T. (2006). Forced annular variations in the 20th century Intergovernmental Panel on Climate Change Fourth Assessment Report models. *Journal of Geophysical Research: Atmospheres*, *111*(D18), Article 2005JD006323. https://doi.org/10.1029/2005JD006323

Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., and Gascon, C. (2011). Global Biodiversity Conservation: The Critical Role of Hotspots. In F. E. Zachos (Ed.), *Biodiversity hotspots: Distribution and protection of conservation priority areas* (1. ed., pp. 3–22). Springer. https://doi.org/10.1007/978-3-642-20992-5_1

Mohr-Bell, D., Díaz, G., Príncipe, R., Gonzalez, C., Bono, J., Ciuffoli, L., Strada, M., Parmuchi, G., Chomnalez, F., Montenegro, C., Loguercio, G., and Bava, J. (2019). *Monitoreo de la Superficie de Bosque Nativo de la República Argentina, Región Forestal Bosque Andino Patagónico; Tomo I Informe*. Secretaría de Ambiente y Desarrollo Sustentable de la Nación.

Mouton, A. M., Baets, B. de, and Goethals, P. L. (2010). Ecological relevance of performance criteria for species distribution models. *Ecological Modelling*, *221*(16), 1995–2002. https://doi.org/10.1016/j.ecolmodel.2010.04.017

Mundo, I. A., Villalba, R., Veblen, T. T., Kitzberger, T., Holz, A., Paritsis, J., and Ripalta, A. (2017). Fire history in southern Patagonia: human and climate influences on fire activity in *Nothofagus pumilio* forests. *Ecosphere*, *8*(9), Article e01932. https://doi.org/10.1002/ecs2.1932

Nagai, S., Nasahara, K. N., Akitsu, T. K., Saitoh, T. M., and Muraoka, H. (2020). Importance of the Collection of Abundant Ground-Truth Data for Accurate Detection of Spatial and Temporal Variability of Vegetation by Satellite Remote Sensing. In K. Dontsova, Z. Balogh-Brunstad, & G. Le Roux (Eds.), *Geophysical Monograph Series. Biogeochemical Cycles* (pp. 223–244). Wiley. https://doi.org/10.1002/9781119413332.ch11

Nascimento, L. S., Nogueira Júnior, M., Hara, C. S., and Almeida Noernberg, M. (2024). Passive citizen science: social media as a tool for marine wildlife observation. *Marine Ecology Progress Series*, *740*, 219–233. https://doi.org/10.3354/meps-14629

O'Neill, D., Häkkinen, H., Neumann, J., Shaffrey, L., Cheffings, C., Norris, K., and Pettorelli, N. (2023). Investigating the potential of social media and citizen science data to track changes in species' distributions. *Ecology and Evolution*, *13*(5), e10063. https://doi.org/10.1002/ece3.10063

Pace, D. S., Giacomini, G., Campana, I., Paraboschi, M., Pellegrino, G., Silvestri, M., Alessi, J., Angeletti, D., Cafaro, V., Pavan, G., Ardizzone, G., and Arcangeli, A. (2019). An integrated approach for cetacean knowledge and conservation in the central Mediterranean Sea using research and social media data sources. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *29*(8), 1302–1323. https://doi.org/10.1002/aqc.3117

Paruelo, J. M., Beltrán, A., Jobbágy, E., Sala, O. E., and Golluscio, R. A. (1998). The climate of Patagonia : general patterns and controls on biotic processes. *Ecología Austral*, *008*(02).

Paulsen, J., and Körner, C. (2014). A climate-based model to predict potential treeline position around the globe. *Alpine Botany*, *124*(1), 1–12. https://doi.org/10.1007/s00035-014-0124-0

Pepin, N., Arnone, E., Gobiet, A., Haslinger, K., Kotlarski, S., Notarnicola, C., Palazzi, E., Seibert, P., Serafin, S., Schöner, W., Terzago, S., Thornton, J. M., Vuille, M., and Adler, C. (2022). Climate Changes and Their Elevational Patterns in the Mountains of the World. *Reviews of Geophysics*, *60*(1), Article e2020RG000730. https://doi.org/10.1029/2020RG000730

Pepin, N., Bradley, R. S., Diaz, H. F., Baraër, M., Caceres, E. B., Forsythe, N., Fowler, H., Greenwood, G., Hashmi, M. Z., Liu, X. D., and Miller, JR (2015). Elevation-dependent warming in mountain regions of the world. *Nature Climate Change*, *5*(5), 424–430. https://doi.org/10.1038/nclimate2563

Pérez-Escobar, O. A., Zizka, A., Bermúdez, M. A., Meseguer, A. S., Condamine, F. L., Hoorn, C., Hooghiemstra, H., Pu, Y., Bogarín, D., Boschman, L. M., Pennington, R. T., Antonelli, A., and Chomicki, G. (2022). The Andes through time: Evolution and distribution of Andean floras. *Trends in Plant Science*, *27*(4), 364–378. https://doi.org/10.1016/j.tplants.2021.09.010

Pessacg, N., Flaherty, S., Solman, S., and Pascual, M. (2020). Climate change in northern Patagonia: critical decrease in water resources. *Theoretical and Applied Climatology*, *140*(3-4), 807–822. https://doi.org/10.1007/s00704-020-03104-8

Peterson, A. T. (Ed.). (2011). *Monographs in population biology: no. 49*. *Ecological niches and geographic distributions*. Princeton University Press.

Peterson, A. T., Papeş, M., and Soberón, J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, *213*(1), 63–72. https://doi.org/10.1016/j.ecolmodel.2007.11.008

Peterson, A. T., and Soberón, J. (2012). Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right. *Natureza & Conservação*, *10*(2), 102–107. https://doi.org/10.4322/natcon.2012.019

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., and Araújo, M. B. (2011). Concepts of Niches. In A. T. Peterson (Ed.), *Monographs in population biology: no. 49. Ecological niches and geographic distributions.* Princeton University Press. https://doi.org/10.23943/-princeton/9780691136868.003.0002

Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3-4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., and Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications : A Publication of the Ecological Society of America*, *19*(1), 181–197. https://doi.org/10.1890/07-2153.1

Pissolito, C. (2016). *Sucesión Primaria de la Vegetación en Bosques de Nothofagus spp. en los Andes Patagónicos: Patrones Espacio Temporales de Establecimiento de Plantas e Implicancias Para la Restauración*. Ph.D. Thesis, Universidad Nacional de Cuyo, Mendoza, Argentina.

Pocock, M. J. O., Tweddle, J. C., Savage, J., Robinson, L. D., and Roy, H. E. (2017). The diversity and evolution of ecological and environmental citizen science. *PloS One*, *12*(4), e0172579. https://doi.org/10.1371/journal.pone.0172579

Pollmann, W. (2001). *Vegetationsökologie und Dynamik temperierter Nothofagus alpina - Wälder im südlichen Südamerika (Chile, Argentinien): Mit 29 Tabellen im Text, als Anhang sowie auf 2 Beilagen*. Zugl.: Münster (Westfalen), Univ., Diss., 2001. *Dissertationes botanicae: Vol. 348*. Cramer.

Potts, J. M., and Elith, J. (2006). Comparing species abundance models. *Ecological Modelling*, *199*(2), 153–163. https://doi.org/10.1016/j.ecolmodel.2006.05.025

Premoli, A. C., Mathiasen, P., Acosta, M. C., and McCulloch, R. D. (2024). Two sides of the same coin? Transient hybridization in refugia and rapid postglacial ecological divergence ensure the evolutionary persistence of sister *Nothofagus*. *Journal of Evolutionary Biology*, *37*(10), 1181–1193. https://doi.org/10.1093/jeb/voae100

Premoli, A. C., Quiroga, P., Mathiasen, P., and Kitzberger, T. (2012). Ecological niche modeling meets phylogeography to unravel hidden past history of key forest genera in plant geography: *Podocarpus* and *Nothofagus*. *Natureza & Conservação*, *10*(2), 160–168. https://doi.org/10.4322/natcon.2012.036

Premoli, A. C., Raffaele, E., and Mathiasen, P. (2007). Morphological and phenological differences in *Nothofagus pumilio* from contrasting elevations: Evidence from a common garden. *Austral Ecology*, *32*(5), 515–523. https://doi.org/10.1111/-j.1442-9993.2007.01720.x

Qiu, J., Feng, S., and Yuan, W. (2023). Upward-moving mountain treelines: An indicator of changing climate. *Global Change Biology*, *29*(24), 6832–6833. https://doi.org/10.1111/gcb.16954

Qiu, T., Sharma, S., Woodall, C. W., and Clark, J. S. (2021). Niche Shifts From Trees to Fecundity to Recruitment That Determine Species Response to Climate Change. *Frontiers in Ecology and Evolution*, *9*, Article 719141. https://doi.org/10.3389/-fevo.2021.719141

R Core Team (2024). *A Language and Environment for Statistical Computing: Version 4.4.1*. https://www.r-project.org/. Last accessed 27/10/2025.

Ranghetti, L., Boschetti, M., Nutini, F., and Busetto, L. (2020). “sen2r”: An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data. *Computers & Geosciences*, *139*, 104473. https://doi.org/10.1016/j.cageo.-2020.104473

Rebertus, A. J., Kitzberger, T., Veblen, T. T., and Roovers, L. M. (1997). Blowdown History And Landscape Patterns In The Andes Of Tierra Del Fuego, Argentina. *Ecology*, *78*(3), 678–692. https://doi.org/10.1890/0012-9658(1997)078[0678:-BHALPI]2.0.CO;2

Reiter, E. J., Weigel, R., Walentowski, H., Loguercio, G. A., Fierke, J., Winter, A. F. N., Simon, A., Kotowska, M. M., Joelson, N. Z., Caselli, M., and Leuschner, C. (2024). Climate vulnerability of *Nothofagus pumilio, Nothofagus dombeyi* and *Austrocedrus chilensis* in northern Patagonia’s temperate forests. *Forest Ecology and Management*, *572*, 122261. https://doi.org/10.1016/j.foreco.2024.122261

Richards, J. A. (2022). *Remote Sensing Digital Image Analysis*. Springer International Publishing. https://doi.org/10.1007/978-3-030-82327-6

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F. (2016). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Rodríguez-Catón, M., Villalba, R., Morales, M., and Srur, A. M. (2016). Influence of droughts on *Nothofagus pumilio* forest decline across northern Patagonia, Argentina. *Ecosphere*, *7*(7), Article e01390. https://doi.org/10.1002/ecs2.1390

Rodríguez-Souilla, J., Cellini, J. M., Roig, F. A., Lencinas, M. V., Chaves, J. E., Peri, P. L., and Martínez Pastur, G. J. (2024a). Climate influence seed production in managed and unmanaged *Nothofagus pumilio* forests of Southern Patagonia. *New Zealand Journal of Forestry Science*, *54.* https://doi.org/10.33494/nzjfs542024x295x

Rodríguez-Souilla, J., Chaves, J. E., Lencinas, M. V., Cellini, J. M., Roig, F. A., Peri, P. L., and Martinez Pastur, G. (2024b). Quality evaluation of *Nothofagus pumilio* seeds linked to forest management and climatic events. *Ecological Processes*, *13*(1). https://doi.org/10.1186/s13717-024-00485-w

Ryo, M., Angelov, B., Mammola, S., Kass, J. M., Benito, B. M., and Hartig, F. (2021). Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models. *Ecography*, *44*(2), 199–205. https://doi.org/10.1111/ecog.05360

Sales, L. P., Hayward, M. W., and Loyola, R. (2021). What do you mean by “niche”? Modern ecological theories are not coherent on rhetoric about the niche concept. *Acta Oecologica*, *110*, 103701. https://doi.org/10.1016/j.actao.2020.103701

Schickhoff, U., Bobrowski, M., Mal, S., Schwab, N., and Singh, R. B. (2022). The World’s Mountains in the Anthropocene. In Schickhoff, U. (Ed.), *Sustainable Development Goals Ser. Mountain Landscapes in Transition: Effects of Land Use and Climate Change* (pp. 1–144). Springer International Publishing AG. https://doi.org/10.1007/978-3-030-70238-0_1

Schickhoff, U., Bobrowski, M., and Schwab, N. (2020). Alpine Waldgrenzen im Klimawandel – Wie sind die heterogenen Reaktionsmuster zu erklären? In Lozán, J. L., Breckle, S. W., Escher-Vetter, H., Grassl, H., Kasang, D., Paul, F., and Schickhoff, U. (Eds.), *Warnsignal Klima: Hochgebirge im Wandel* (pp. 232–238). doi:10.25592/warnsignal.klima.hochgebirge-im-wandel.35. Last accessed.

Schneider, C., Glaser, M., Kilian, R., Santana, A., Butorovic, N., and Casassa, G. (2003). Weather Observations Across the Southern Andes at 53°S. *Physical Geography*, *24*(2), 97–119. https://doi.org/10.2747/0272-3646.24.2.97

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., and Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecological Modelling*, *406*, 109–120. https://doi.org/10.1016/j.ecolmodel.2019.06.002

sebaingles (2022). *Abrupt treeline seen from the Paso Garibaldi scenic viewpoint at Lago Escondido*. https://www.instagram.com/p/Cb_M3rHszwA/. Last accessed 02/11/2025.

Sillero, N., Arenas-Castro, S., Enriquez-Urzelai, U., Vale, C. G., Sousa-Guedes, D., Martínez-Freiría, F., Real, R., and Barbosa, A. (2021). Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. *Ecological Modelling*, *456*, 109671. https://doi.org/10.1016/j.ecolmodel.2021.-109671

Sillero, N., and Barbosa, A. M. (2021). Common mistakes in ecological niche models. *International Journal of Geographical Information Science*, *35*(2), 213–226. https://doi.org/10.1080/13658816.2020.1798968

Soberón, J., and Peterson, A. T. (2005). Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodiversity Informatics*, *2*(0). https://doi.org/10.17161/bi.v2i0.4

Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, *10*(12), 1115–1123. https://doi.org/10.1111/j.1461-0248.2007.01107.x

Soliani, C., Tsuda, Y., Bagnoli, F., Gallo, L. A., Vendramin, G. G., and Marchelli, P. (2015). Halfway encounters: Meeting points of colonization routes among the southern beeches *Nothofagus pumilio* and *N. Antarctica*. *Molecular Phylogenetics and Evolution*, *85*, 197–207. https://doi.org/10.1016/j.ympev.-2015.01.006

Soria-Auza, R. W., Kessler, M., Bach, K., Barajas-Barbosa, P. M., Lehnert, M., Herzog, S. K., and Böhner, J. (2010). Impact of the quality of climate models for modelling species occurrences in countries with poor climatic documentation: a case study from Bolivia. *Ecological Modelling*, *221*(8), 1221–1229. https://doi.org/10.1016/j.ecolmodel.2010.01.004

Srur, A. M., Amoroso, M. M., Mundo, I., Morales, M. S., Rodríguez-Catón, M., Aschero, V., and Villalba, R. (2020). Forest Dynamics in the Argentinean Patagonian Andes: Lessons Learned from Dendroecology. In Pompa-García, M. and Camarero, J.J. (Eds.), *Latin American Dendroecology* (pp. 171–201). Springer International Publishing. https://doi.org/10.1007/978-3-030-36930-9_8

Srur, A. M., Villalba, R., Rodríguez-Catón, M., Amoroso, M. M., and Marcotti, E. (2016). Establishment of *Nothofagus pumilio* at Upper Treelines Across a Precipitation Gradient in the Northern Patagonian Andes. *Arctic, Antarctic, and Alpine Research*, *48*(4), 755–766. https://doi.org/10.1657/AAAR0016-015

Srur, A. M., Villalba, R., Rodríguez-Catón, M., Amoroso, M. M., and Marcotti, E. (2018). Climate and *Nothofagus pumilio* Establishment at Upper Treelines in the Patagonian Andes. *Frontiers in Earth Science*, *6*, Article 57. https://doi.org/10.3389/feart.2018.00057

Statista (2025). *Social Media: nach der Anzahl der Nutzer, in Millionen*. https://de.statista.com/themen/1842/soziale-netzwerke/. Last accessed 27/10/2025.

Stecconi, M., Puntieri, J. G., and Barthélémy, D. (2010). An architectural approach to the growth forms of *Nothofagus pumilio* (*Nothofagaceae*) along an altitudinal gradient. *Botany*, *88*(8), 699–709. https://doi.org/10.1139/B10-040

Steen, V. A., Tingley, M. W., Paton, P. W. C., and Elphick, C. S. (2021). Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution*, *12*(2), 216–226. https://doi.org/10.1111/2041-210X.13525

Sumner, S., Bevan, P., Hart, A. G., and Isaac, N. J. (2019). Mapping species distributions in 2 weeks using citizen science. *Insect Conservation and Diversity*, *12*(5), 382–388. https://doi.org/10.1111/icad.12345

Thuiller, W. (2024). Ecological niche modelling. *Current Biology : CB*, *34*(6), R225-R229. https://doi.org/10.1016/j.cub.2024.02.018

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, *46*, 234. https://doi.org/10.2307/143141

Tovar, C., Carril, A. F., Gutiérrez, A. G., Ahrends, A., Fita, L., Zaninelli, P., Flombaum, P., Abarzúa, A. M., Alarcón, D., Aschero, V., Báez, S., Barros, A., Carilla, J., Ferrero, M. E., Flantua, S. G. A., Gonzáles, P., Menéndez, C. G., Pérez-Escobar, O. A., Pauchard, A., Ruscica, R. C., Särkinen, T., Sörensson, A. A., Srur, A. M., Villalba, R., and Hollingsworth, P. M. (2022). Understanding climate change impacts on biome and plant distributions in the Andes: Challenges and opportunities. *Journal of Biogeography*, *49*(8), 1420–1442. https://doi.org/10.1111/jbi.14389

Treml, V., Šenfeldr, M., Chuman, T., Ponocná, T., and Demková, K. (2016). Twentieth century treeline ecotone advance in the Sudetes Mountains (Central Europe) was induced by agricultural land abandonment rather than climate change. *Journal of Vegetation Science*, *27*(6), 1209–1221. https://doi.org/10.1111/jvs.12448

Treml, V., and Veblen, T. T. (2017). Does tree growth sensitivity to warming trends vary according to treeline form? *Journal of Biogeography*, *44*(7), 1469–1480. https://doi.org/10.1111/jbi.12996

Uyeda, K. A., Stow, D. A., and Richart, C. H. (2020). Assessment of volunteered geographic information for vegetation mapping. *Environmental Monitoring and Assessment*, *192*(8), 554. https://doi.org/10.1007/s10661-020-08522-9

Valavi, R., Elith, J., Lahoz-Monfort, J. J., and Guillera-Arroita, G. (2019). blockCV : An r package for generating spatially or environmentally separated folds for k -fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, *10*(2), 225–232. https://doi.org/10.1111/2041-210X.13107

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., and Rose, S. K. (2011). The representative concentration pathways: an overview. *Climatic Change*, *109*(1-2), 5–31. https://doi.org/10.1007/s10584-011-0148-z

VanDerWal, J., Shoo, L. P., Graham, C., and Williams, S. E. (2009). Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling*, *220*(4), 589–594. https://doi.org/10.1016/j.ecolmodel.2008.11.010

Varela, S. A., Gyenge, J. E., Fernández, M. E., and Schlichter, T. (2010). Seedling drought stress susceptibility in two deciduous *Nothofagus* species of NW Patagonia. *Trees*, *24*(3), 443–453. https://doi.org/10.1007/s00468-010-0412-2

Veblen, T. T., Hill, R. S., and Read, J. (1996). *The ecology and biogeography of Nothofagus forests*. Yale University Press.

Veblen, T. T., Holz, A., Paritsis, J., Raffaele, E., Kitzberger, T., and Blackhall, M. (2011). Adapting to global environmental change in Patagonia: What role for disturbance ecology? *Austral Ecology*, *36*(8), 891–903. https://doi.org/10.1111/j.1442-9993.2010.02236.x

Villalba, R., Boninsegna, J. A., Veblen, T. T., Schmelter, A., and Rubulis, S. (1997). RECENT TRENDS IN TREE-RING RECORDS FROM HIGH ELEVATION SITES IN THE ANDES OF NORTHERN PATAGONIA. *Climatic Change*, *36*(3-4), 425–454. https://doi.org/10.1023/A:1005366317996

Virić Gašparić, H., M. Mikac, K., Pajač Živković, I., Krehula, B., Orešković, M., Galešić, M. A., Ninčević, P., Varga, F., and Lemić, D. (2022). Firefly Occurrences in Croatia – One Step Closer from Citizen Science to Open Data *Interdisciplinary Description of Complex Systems*, *20*(2), 112–124. https://doi.org/10.7906/-indecs.20.2.4

Vitali, A., Garbarino, M., Camarero, J. J., Malandra, F., Toromani, E., Spalevic, V., Čurović, M., and Urbinati, C. (2019). Pine recolonization dynamics in Mediterranean human-disturbed treeline ecotones. *Forest Ecology and Management*, *435*, 28–37. https://doi.org/10.1016/j.foreco.2018.12.039

Wang, L., Diao, C., and Lu, Y. (2025). The role of remote sensing in species distribution models: a review. *International Journal of Remote Sensing*, *46*(2), 661–685. https://doi.org/10.1080/01431161.2024.2421949

Wei, P., Lu, Z., and Song, J. (2015). Variable importance analysis: A comprehensive review. *Reliability Engineering & System Safety*, *142*, 399–432. https://doi.org/10.1016/j.ress.2015.05.018

Wiggins, A., and Crowston, K. (2011). From Conservation to Crowdsourcing: A Typology of Citizen Science. In R. H. Sprague (Ed.), *2011 44th Hawaii International Conference on System Sciences: (HICSS 2011) ; Kauai, Hawaii, USA, 4 - 7 January 2011* (pp. 1–10). IEEE. https://doi.org/10.1109/HICSS.2011.207

WirtschaftsWoche (2014). *Anzahl der täglich bei Instagram hochgeladenen Fotos in den Jahren 2011 bis 2014 (in Millionen)*. https://de.statista.com/statistik/daten/-studie/312273/umfrage/taeglich-auf-instagram-hochgeladene-und-geteilte-fotos/. Last accessed 30/10/2025.

Yates, L. A., Aandahl, Z., Richards, S. A., and Brook, B. W. (2023). Cross validation for model selection: A review with examples from ecology. *Ecological Monographs*, *93*(1), Article e1557. https://doi.org/10.1002/ecm.1557

Young, K. R., and León, B. (2007). Tree-line changes along the Andes: Implications of spatial patterns and dynamics. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1478), 263–272. https://doi.org/10.1098/rstb.2006.1986

Zizka, A., Antonelli, A., and Silvestro, D. (2021). sampbias a method for quantifying geographic sampling biases in species distribution data. *Ecography*, *44*(1), 25–32. https://doi.org/10.1111/ecog.05102

Zurell, D., and Engler, J. O. (2019). Ecological niche modelling. In P. O. Dunn & A. P. Møller (Eds.), *Effects of Climate Change on Birds* (pp. 60–73). Oxford University PressOxford. https://doi.org/10.1093/oso/9780198824268.003.0006

List of Publications

Main publications of the thesis

Jentsch, H., Weidinger, J., **Werner, M**., and Bobrowski, M. (2025). ClimDatDownloadR: Accessing Climate Data Repositories for Modelling [Manuscript submitted for publication].

Werner, M., Böhner, J., Oldeland, J., Schickhoff, U., Weidinger, J., and Bobrowski, M. (2025). Treeline Species Distribution Under Climate Change: Modelling the Current and Future Range of *Nothofagus pumilio* in the Southern Andes. *Forests*, *16*(8), 1211. https://doi.org/10.3390/f16081211

Werner, M., Weidinger, J., Böhner, J., Schickhoff, U., and Bobrowski, M. (2024). Instagram data for validating *Nothofagus pumilio* distribution mapping in the Southern Andes: A novel ground truthing approach. *Frontiers of Biogeography*, *17*, Article e140606. https://doi.org/10.21425/fob.17.140606

Werner, M., Weidinger, J., Böhner, J., Schickhoff, U., and Bobrowski, M. (2024). Instagram ground truthing approach – Spatial occurrence data of *Nothofagus pumilio* in the Southern Andes [Data set]. https://doi.org/10.25592/UHHFDM.16239

Additional publications

Werner, M., Weidinger, J., Böhner, J., Schickhoff, U., and Bobrowski, M. (2025). Instagram Ground Truthing Approach – Ein innovativer Social-Media-Ansatz zur Generierung von Vorkommensdaten für Species Distribution Modelling. Berichte Der Reinhold-Tüxen-Gesellschaft(34) [accepted].

Conference contributions

Institute Colloquium (5. May 2025), presentation: Instagram Ground Truthing Approach – Ein innovativer Social-Media-Ansatz zur Generierung von Vorkommensdaten für die Modellierung der ökologischen Nische von *Nothofagus pumilio* in den südlichen Anden

Reinhold-Tüxen-Gesellschaft (2. May 2025), presentation: Instagram Ground Truthing Approach – Ein innovativer Social-Media-Ansatz zur Generierung von Vorkommensdaten für Species Distribution Modelling

Annual Meeting of working group Biogeography and Macroecology (4. May 2023), presentation: Instagram Occurrence Data Sampling for Species Distribution Modelling

Arbeitskreis Klima (29. October 2022), poster presentation: Instagram Occurrence Data Sampling for Ecological Niche Modelling

Original Publications

Publication I

Werner, M., Weidinger, J., Böhner, J., Schickhoff, U., and Bobrowski, M. (2024). Instagram data for validating *Nothofagus pumilio* distribution mapping in the Southern Andes: A novel ground truthing approach. *Frontiers of Biogeography*, *17*, Article e140606. https://doi.org/10.21425/fob.17.140606

FRONTIERS OF BIOGEOGRAPHY
The scientific journal of
The International Biogeography Society

Frontiers of Biogeography 17, 2024,
e140606 | DOI 10.21425/fob.17.140606

RESEARCH ARTICLE

Instagram data for validating *Nothofagus pumilio* distribution mapping in the Southern Andes: A novel ground truthing approach

Melanie Werner[1], Johannes Weidinger[1], Jürgen Böhner[2], Udo Schickhoff[1], Maria Bobrowski[1]

1 *University of Hamburg, Center for Earth System Research and Sustainability (CEN), Institute of Geography, Bundesstraße 55, 20146 Hamburg, Germany*
2 *University of Hamburg, Center for Earth System Research and Sustainability (CEN), Cluster of Excellence "Climate, Climatic Change, and Society" (CLICCS), Institute of Geography, Bundesstraße 55, 20146 Hamburg, Germany*

Corresponding author: Melanie Werner (melanie.werner@uni-hamburg.de)

Editor Janet Franklin
Received 2 March 2024 ♦ **Accepted** 22 October 2024 ♦ **Published** 12 November 2024

Abstract

The availability of valid, non-biased species occurrence data has always been a major challenge for biodiversity research and modelling studies. Data from open-source databases or remote sensing are promising approaches to increase the availability of species occurrence data. However, these data may contain spatial, temporal, and taxonomic biases or require ground truthing. In recent years, social media has received attention for its contribution to species occurrence data sampling and ground truthing approaches. The wide reach of social media platforms allows for rapid and large-scale analyses.

Here we introduce a novel Instagram ground truthing approach to validate the occurrence mapping of *Nothofagus pumilio* across its entire distribution range. The treeline species of the southern Andes has been extensively studied in small-scale studies, but large-scale modelling approaches are largely missing due to limited accessibility to treeline sites resulting in a lack of occurrence data. The content posted on the social media platform Instagram consists of images and videos in which the species *N. pumilio* and its location can be identified. By searching for suitable posts using hashtags and location tags, we created 1238 Instagram ground truthing points. We compared the performance of our dataset with open-source data from the Global Biodiversity Information Facility (GBIF) through visual, quantitative, and bias analyses, acknowledging that both social media-based and Citizen Science data are subject to sampling and spatial biases due to collection in human-accessible areas. The Instagram ground truthing points were subsequently used to validate remote sensing occurrence data, generated using Sentinel-2 level 2A data and Supervised Classification. The combined approach – Instagram ground truthing and remote sensing – allows for the collection of occurrence data over the entire latitudinal range of *N. pumilio*, covering approximately 2000 km.

Highlights

- The use of social media content provides potentially important contributions to species occurrence data sampling and ground truthing
- In our study we introduce a novel ground truthing approach for species occurrence data sampling based on Instagram data
- Instagram ground truthing points, combined with Supervised Classification generate species occurrence data of *Nothofagus pumilio* over its entire distribution range in the southern Andes
- The performance of the Instagram ground truthing points is evaluated by comparison with existing data from the GBIF database.
- Our Instagram ground truthing approach demonstrates a new way of sampling species occurrence data and can be applied to other suitable species and study areas.

Keywords

Citizen Science, Ecological Modelling, Ground Truthing, Instagram, Occurrence Data Sampling, Remote Sensing, Social Media, Southern Andes, Supervised Classification

PENSOFT®

Introduction

Quantifying spatial and temporal distribution of species and analysing underlying ecological requirements has become increasingly important in high elevation regions due to climate and environmental change (Schickhoff et al. 2022). Worldwide, ecological modelling studies are applied to model treeline species under present, past and future climate conditions (e.g., Dullinger et al. 2004; Bobrowski et al. 2017; Akobia et al. 2022). Whereas high mountains of the Northern Hemisphere are well represented in treeline related research, those of the Southern Hemisphere, such as the Andes, have rather been neglected (Hansson et al. 2021; Hansson et al. 2023). Recently published studies on the southern Andes have focused on local treeline sites (e.g., Daniels and Veblen 2003; Fajardo and Piper 2014; Srur et al. 2016, 2018), while large-scale modelling studies investigating treeline species or vegetation at higher elevations are very rare (Nagy et al. 2023). The limited accessibility of treeline sites due to remoteness and complex topography might have impeded such studies.

Species occurrence data are mainly collected through field research and made available in publications and databases (Feng et al. 2019). However, large-scale vegetation sampling is often costly and time-consuming. Therefore, species occurrence data are frequently downloaded via open-source databases such as the Global Biodiversity Information Facility (GBIF, gbif.org), which hosts datasets compiled from various sources (Edwards 2004; Boakes et al. 2010). However, these data may contain unknown taxonomic, spatial or sampling biases and are seldom evaluated or revisited (Beck et al. 2014; Meyer et al. 2016; Daru et al. 2018). Ensuring the quality of species data is essential for accurate modelling outcomes (Chauvier et al. 2021). Therefore, prior to utilisation, thorough examination, filtration, and potential supplementation of the data are imperative steps.

More recently, Citizen Science projects and social media are becoming crucial for surveying species occurrences (Jarić et al. 2020; Goldberg 2023). Citizen science involves the participation of citizens in scientific processes, such as collecting species data (Bonney 1996; Bonney et al. 2009). The number of Citizen Science projects, particularly those that are computer- or app-based, along with the resulting data, is increasing exponentially (Pocock et al. 2017). Such projects can generate large amounts of occurrence data in a comparatively short time (Sumner et al. 2019). Dickinson et al. (2010) even take the view that Citizen Science is the only practical way to study distribution patterns and range shifts of species over large areas. Despite the improving quality of the data, which is increasingly nearing that of expert-recorded data (Aceves-Bueno et al. 2017), bias persists (Di Cecco et al. 2021). Data collection is concentrated in urban and tourist areas, with inaccessible or remote locations rarely being recorded. Additionally, the data is collected solely by individuals engaged in Citizen Science projects. Leveraging social media contributions can unlock further potential in data collection. On social media, an increasing number of geotagged image files are used. Images posted by both experts and non-experts can be analysed in large quantities. Recently, social media platforms such as Facebook, Flickr, Instagram, Twitter and YouTube are used to sample occurrence data and range shifts of animals, for example data of whales, dolphins (e.g., Pace et al. 2019; Gibson et al. 2020; Martino et al. 2021), birds (e.g., Hentati-Sundberg and Olsson 2016), insects (e.g., Virić Gašparić et al. 2022; O'Neill et al. 2023) and plants (e.g., ElQadi et al. 2017). Through the usage of geotagged social media content, as well as identifiable landscape elements and descriptions of the posted image, the actual location of the posts can be traced. In combination with clearly recognisable plant characteristics, it is possible to identify the location of species, representing a promising new ground truthing possibility. With the increasing use of social media and the spread of high-quality digital cameras, a huge potential can be tapped to make, in addition, important contributions to biodiversity monitoring and to the evaluation of potential protected areas (Chowdhury et al. 2023).

While both Citizen Science and social media enhance the sampling of occurrence data, the data remain spatially biased, as records are predominantly collected from locations accessible to humans (Meyer et al. 2016; La Sorte et al. 2024). Within this context, utilising remotely sensed species occurrences emerges as a promising method for reducing such bias and examining large study areas, particularly in regions with limited accessibility, such as high-elevation areas. One example is the use of remote sensing to classify tree species. Very high-resolution data like IKONOS, WorldView, RapidEye and airborne images are mostly used for small-scale studies (Fassnacht et al. 2016) and medium high-resolution data like Landsat and Sentinel data can successfully be used for larger areas (Immitzer et al. 2016; Immitzer et al. 2019). Despite the wide availability of high-resolution data and highly developed remote sensing methods, ground truthing remains indispensable for validating species data accuracy through field validation (on-site sampling) or validation processes with existing datasets (Nagai et al. 2020).

The availability of data for treeline species is limited; however, such species are likely to be suitable candidates for analysis using remote sensing and social media-based species occurrence data sampling. *N. pumilio* forms an abrupt treeline in the orotemperate belt of the southern Andes in mono-species forest stands (Amigo and Rodríguez-Guitián 2011) and is distributed in a region with many touristic areas where photos are taken, for example, during hiking. Therefore, the species can be recognised in satellite images as well as in Instagram posts. As *N. pumilio* responds to climate variations by changing radial growth patterns (Lara et al. 2001; Aravena et al. 2002; Daniels and Veblen 2004; Masiokas and Villalba 2004; Álvarez et al. 2015) and seedling establishment patterns above the treeline (Fajardo and Piper 2014), it is a suitable target species also for modelling approaches aiming at analysing treeline sensitivity and treeline shift due to climate change.

In this study, we demonstrate the large-scale sampling of *N. pumilio* occurrence data using Sentinel-2 imagery and Supervised Classification. To validate the spatial occurrence data, we introduce a novel Instagram ground truthing approach, leveraging occurrence points derived from the social media platform Instagram (www.instagram.com). We hypothesise that this Instagram-based method, due to a high volume of potentially suitable posts and our sampling approach, will generate more occurrence points with reduced spatial bias compared to datasets from the open-source GBIF database. Spatial bias in the resulting species occurrence data is further mitigated by incorporating remote sensing data, which enhances both the quantity and spatial coverage of occurrence information. Unlike presence-only point datasets, remote sensing data provide presence-absence datasets, offering more comprehensive opportunities for ecological modelling.

Material and methods

Study area and target species

Nothofagus pumilio (Poepp et Endl.) Krasser (southern or lenga beech) is the dominant subalpine tree species in the southern Chilean and Argentinean Andes between 35°S and 56°S, encompassing a longitudinal distribution range of more than 500 km (Masiokas and Villalba 2004; Lara et al. 2005; Rodríguez-Catón et al. 2016). Out of all *Nothofagus* species, *N. pumilio* is the most orophilic (Amigo and Rodríguez-Guitián 2011). The dark green, elliptical, and notched broad leaves of the deciduous species turn into an orange-reddish colour in autumn, which is a reliable distinguishing feature of the species in comparison to other evergreen *Nothofagus* species in this region (Hildebrand-Vogel et al. 1990; Amigo and Rodríguez-Guitián 2011).

The distribution area of the species along the Andean cordillera follows an elevational gradient from north to south, while the west-east expansion is also dependent on precipitation. *N. pumilio* forms mono-species forests located between 1600 m up to 2000 m in the northern parts, whereas the elevational limit decreases to 400 m in the southernmost range at Tierra Del Fuego (Cuevas 2000; Lara et al. 2005). The treeline elevation decreases constantly by 60 m per 1° latitude (Lara et al. 2005). The west-east distribution area is defined by the extreme precipitation gradient from the windward to the leeward side of the Andes (Hertel et al. 2008). The eastern distribution boundary is characterised by low precipitation, following the forest-steppe ecotone (Rodríguez-Catón et al. 2016). Often, two treelines are formed in the eastern regions – a common upper and a lower, xeric treeline towards the arid region (Hertel et al. 2008). On the western side of the Andes, the distribution is restricted to high elevations. In these hyperhumid areas, the species does not occur at lower elevations, where *Nothofagus betuloides* is the dominant tree species (Young et al. 2007; Amigo and Rodríguez-Guitián 2011). Especially the autumn colouring of the deciduous species and its occurrence in mono-species forests at the treeline are important characteristics for the Instagram ground truthing approach, as these features aid in identifying the species in Instagram posts and make them recognisable in satellite images.

Instagram ground truthing approach

As a first step, we developed the Instagram ground truthing approach to ensure proper validation of large-scale remotely sensed occurrence data of *N. pumilio*. Additionally, the Instagram ground truthing points are quantitatively compared with existing occurrence data from the GBIF database. We used the social media platform Instagram (www.instagram.com) to sample the Instagram ground truthing points. Although other social media platforms have been utilised in studies sampling species occurrences, Instagram has largely been overlooked. Nonetheless, we identify a significant advantage in using Instagram. Instagram users have the possibility to post both photos and short videos. The platform's lack of text-only posts makes it especially suitable for our approach. At the same time, Instagram is one of the largest social media platforms with 2 billion users worldwide (We Are Social et al. 2024), allowing for the analysis of a significantly larger volume of content compared to lesser-known platforms that exclusively host visual content (e.g., Flickr). Moreover, users can localise their posts with a geographical tag and add descriptions where the content and, e.g., the location can be specified with text or hashtags. The aim of the ground truthing analysis was to identify locations in which the species *N. pumilio* is clearly identifiable, the location can be reliably traced and transferred to a map.

We started the Instagram ground truthing approach in 2021 and repeated it in 2022. The analysis consisted of 3 steps: 1) Potential contributions from publicly accessible profiles were searched for using the search bar embedded in the Instagram user interface and two specific search options: hashtags (#nothofaguspumilio and #lenga for exact species information) and places or landscape features (by location tags, locations in hashtags or usernames for exact locations). 2) Posts were checked using a strict catalogue of criteria (Table 1), ensuring that the species can be clearly identified, and that the location is traceable. Autumn pictures were preferentially included in the analysis, as the identification of the species is particularly reliable during this season. In addition, the typical abrupt treeline and mono-species forests were main criteria for selecting the posts. Furthermore, the Instagram posts include a publication date. Due to the chronological structure of the Instagram interface, our analysis has primarily focused on the most recent posts. However, the publication date does not guarantee that the photo was taken in that year. 3) Occurrences verified by the approach were transferred to a map. We first analysed all posts with specific species information (#nothofaguspumilio and #lenga). Next, we searched for specific locations within the species' distribution range, filling gaps and adding points to the already set occurrence points. This structured approach ensures a homogeneous distribution of the sampled occurrences.

When we manually transferred the locations to a map in step 3), simple descriptions of the locations were not sufficient. The locations still had to be clearly traceable by landscape features (see Table 1). Such landscape elements in Patagonia could be glaciers, characteristic mountain peaks, roads, urban areas, touristic sites, waterbodies, and coastlines. These features should be so characteristic, that they can also be recognised in a satellite image. To avoid mistakes, the hashtags and descriptions in the post should match the given location. Images centring people, as well as those altered through the usage of filters, colour modifications, or emojis, were excluded.

Table 1. Criteria for selecting Instagram posts to generate *Nothofagus pumilio* occurrence data. All these criteria must be fulfilled for the image to be included in the analysis.

Criterion	Element or Example
Typical characteristics of *Nothofagus pumilio*	morphological characteristics (leaves, branches, habitus) autumn colouring abrupt treeline mono-species forest
Concrete location information	geographical tag location hashtag location description in the caption
Recognisable landscape elements	glaciers mountain peaks or ranges rivers, lakes roads tourist points, cities, villages coastlines
Fitting hashtags	hashtags describing the location or the plant
Picture criteria	Avoid persons in focus no photo montages no emojis no extreme (colour) falsifications

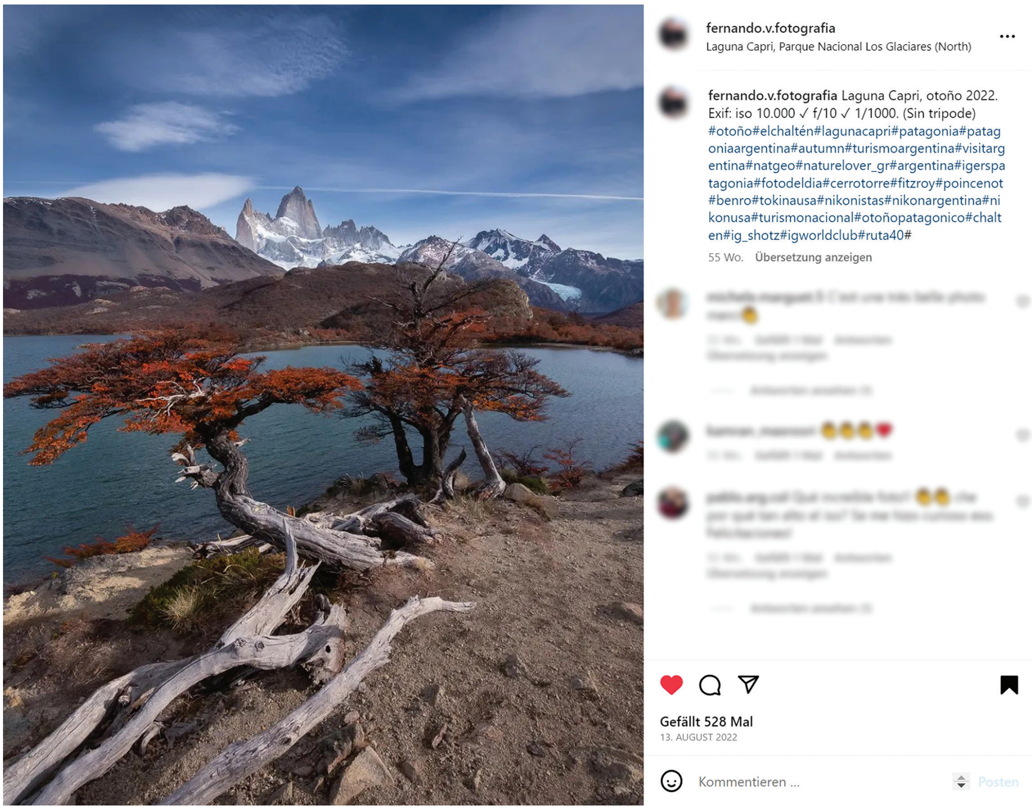
If all conditions were met, we manually transferred the determined occurrence to a map with SAGA GIS (Conrad et al. 2015, https://saga-gis.sourceforge.io). When creating Instagram ground truthing points, at least one point was created at the actual location where the photo was taken and *N. pumilio* was identified. However, additional points were created if other occurrences of *N. pumilio* were visible at the posted image. This is particularly the case when forest or even the treeline of *N. pumilio* can be seen in the background of landscape photographs. The species' ecology enables the identification of these occurrences, as its presence at the treeline, combined with autumn colouring, makes the species recognisable, particularly when morphological features are clearly visible in the foreground of the images. In some cases, points were set as far as the neighbouring valley, when the treeline was clearly autumn-coloured in the used satellite image. Background points were assigned to a 1 km grid, as this is commonly the target resolution for model analyses, thereby maintaining the number of background points at a reasonable level. The selection of posts is exemplified in Fig. 1. The landscape elements used to locate the posts, along with the transferred points, are shown in Fig. 2.

Remote sensing occurrence data sampling with supervised classification

Analysing multispectral, medium spatial resolution satellite data like Sentinel-2 leads to cost-efficient and robust results in tree species classification over large spatial extents (Fassnacht et al. 2016; Immitzer et al. 2016; Immitzer et al. 2019). Therefore, we compiled large-scale occurrence data over the entire distribution range of *N. pumilio* using Sentinel-2 level 2A data (BOA) in a resolution of 20 m. As *N. pumilio* occurs in mono-species forests at the treeline, a medium spatial resolution was sufficient to classify the forest type while allowing analysis over almost 2000 km latitude. Furthermore, the high temporal resolution data allowed phenological differentiation with autumn and summer images. Using the R Package "sen2r" (Ranghetti et al. 2020) we downloaded summer (months December and January) and autumn (month April) Sentinel-2 data. The sensing period was 2019 to 2022. Sentinel-2 scenes with up to 50 % cloud cover were included. Most scenes had a cloud cover of 15 % or less. We used the Sentinel-2 level 2A data "Scene Classification Layer" product (SCL, Fig. 3B) to mask all raster cells not classified as vegetation. This ensured that only data relevant to the analysis was included in the classification. Downloading numerous Sentinel-2 data with different acquisition times ensured that as many vegetation pixels as possible from a Sentinel-2 scene were included in the classification, despite high cloud cover.

We trained our Supervised Classification with training areas including three classes (1 = *N. pumilio*, 2 = Evergreen vegetation, 3 = Low vegetation). Training areas were created using summer, autumn and winter Sentinel-2 scenes at selected sites across the range. The winter data made evergreen vegetation clearly recognisable. Autumn colouring at the treeline indicated *N. pumilio*. We tested various classification algorithms for Supervised Classification, including well-performing standard algorithms like Maximum Likelihood, Minimum Distance, and Spectral Angle, as well as the decision tree-based Random Forest algorithm. The performance of these algorithms was measured using overall accuracy and the Kappa value (Richards 2022). To cover a wide spectral range, we used all the spectral bands from 2 to 7, 8a, 11 and, 12. All preprocessing steps and the Supervised Classification, which are visualised in Fig. 3, were carried out in R (R Core Team 2023) and SAGA GIS (Conrad et al. 2015). Map visualisation was performed in ArcGIS Pro software by Esri, version 2.7.0.

We classified summer and autumn data separately and subsequently extracted and merged the result of



fernando.v.fotografia
Laguna Capri, Parque Nacional Los Glaciares (North)

fernando.v.fotografia Laguna Capri, otoño 2022.
Exif: iso 10.000 ✓ f/10 ✓ 1/1000. (Sin tripode)
#otoño#elchaltén#lagunacapri#patagonia#patagoniaargentina#autumn#turismoargentina#visitargentina#natgeo#naturelover_gr#argentina#igerspatagonia#fotodeldia#cerrotorre#fitzroy#poincenot#benro#tokinausa#nikonistas#nikonargentina#nikonusa#turismonacional#otoñopatagonico#chalten#ig_shotz#igworldclub#ruta40#

55 Wo. Übersetzung anzeigen

Gefällt 528 Mal
13. AUGUST 2022

Figure 1. Example of the Instagram ground truthing approach at Laguna Capri, Argentina. *Nothofagus pumilio* can be identified by its habitus and leaves in the foreground, its autumn-colouring and the abrupt treeline in the background. The lake itself and Mount Fitz Roy are reliable landscape elements. A location tag, the post description and hashtags also refer to the location (used with permission by Instagram user fernando.v.fotografia 2022).

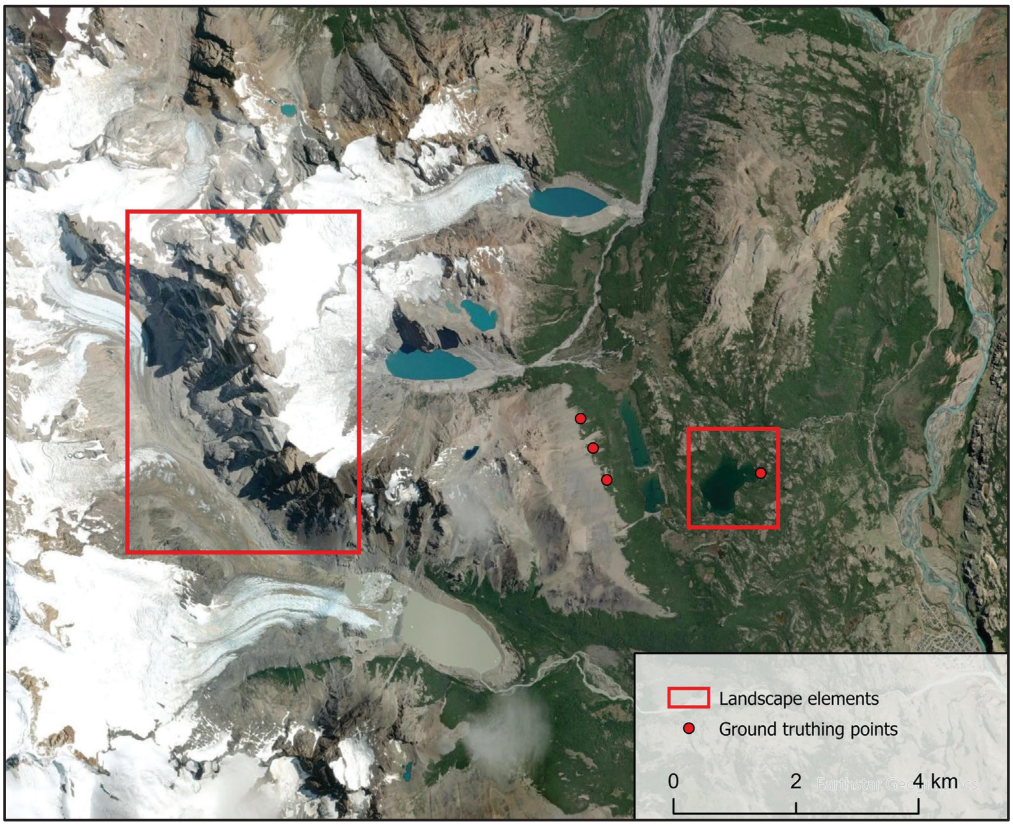


Figure 2. The transmitted points (red) visible at Laguna Capri and at the treeline. Red boxes indicate the landscape elements, Laguna Capri and Mount Fitz Roy, that allowed to identify the location of the Instagram post shown in Fig. 1.

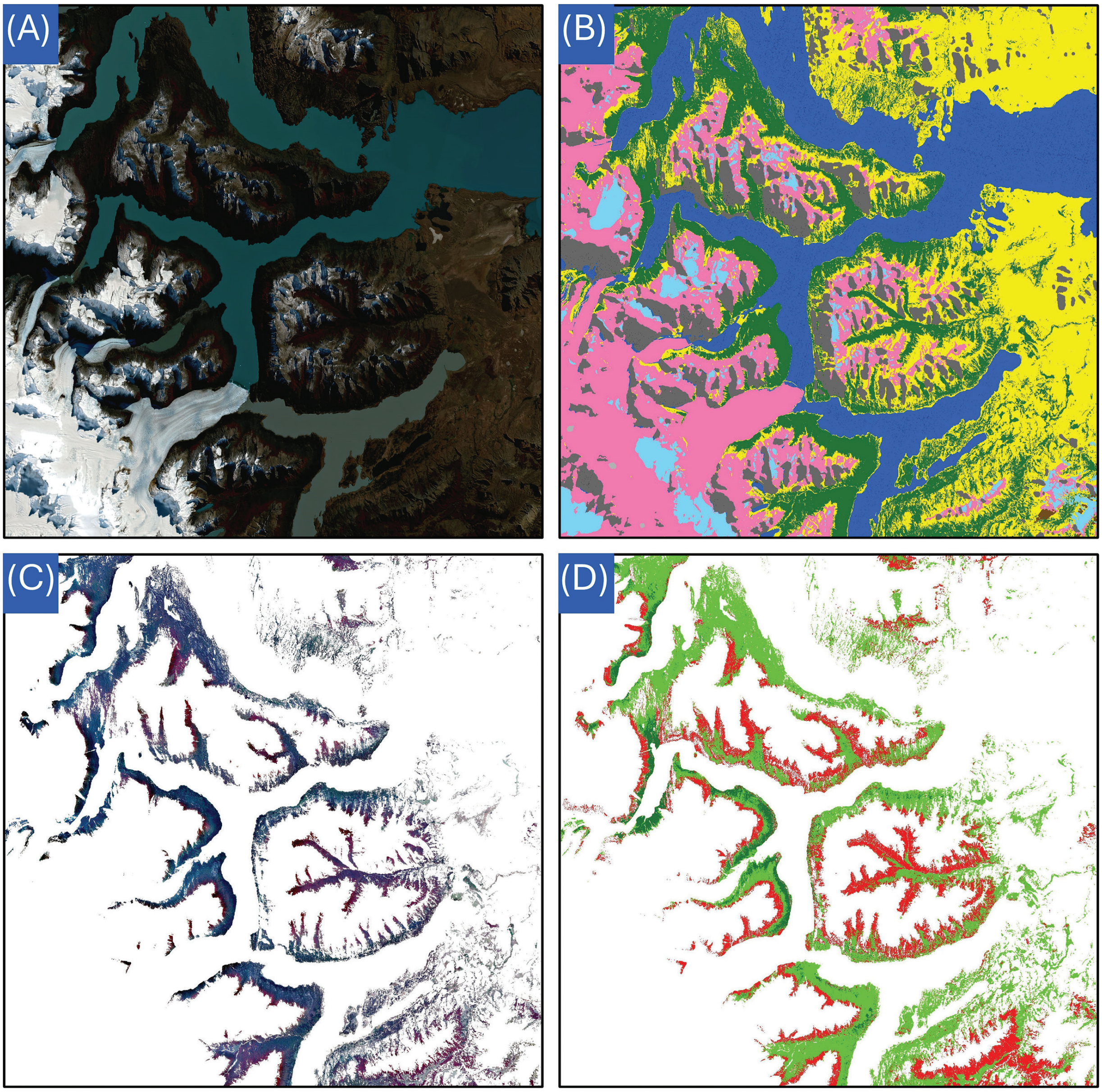


Figure 3. (**A**) Sentinel-2 autumn image at the Perito Moreno Glacier, Argentina, (**B**) Scene Classification Layer of the Sentinel-2 scene; the green area, class 4, shows vegetation, (**C**) the masked Sentinel-2 scene and, (**D**) classification result with three classes (red = *Nothofagus pumilio*, dark green = evergreen vegetation, light green = low vegetation/grassland).

the *N. pumilio* occurrence into one layer. The result was further refined using three different masks. As the classification of the class "low vegetation/grassland" was particularly reliable in the summer classification and the classification of the class "evergreen" in the autumn classification, the result was masked by the result of these classes. Therefore, any pixels that may have been misclassified have been removed. In the north of the study area, *N. pumilio* occurs only at higher elevations, so that other deciduous species at lower elevations were misclassified as *N. pumilio*. To remove this occurrence, a Digital Surface Model (DSM, ALOS Global Digital Surface Model "ALOS World 3D – 30m (AW3D30)", Jaxa EORC 2023) was used to remove occurrences below high-elevation mono-species forests (thresholds: 800 m from 35°S to 40°S; 500 m from 40°S to 45°S; 250 m from 45° to 50°S).

GBIF data and validation process

The large-scale remote sensing data on the occurrence of *N. pumilio* was validated using the Instagram ground truthing points. This process involved verifying whether the Instagram ground truthing points align with the spatial distribution derived by Supervised Classification. Additionally, we used occurrence data from the GBIF database to also validate the spatial distribution and to compare it with the Instagram ground truthing points visually, quantitatively and with a sampling bias analysis. The GBIF database provides data on species of all taxa according to the open-source principle. The Secretariat in Copenhagen coordinates data from various sources, such as museums, research publications, and Citizen Science projects, and makes them available (GBIF 2024a). A search for the species *Nothofagus pumilio* (Poepp.

& Endl.) Krasser resulted in 1,616 entries (as of 6 June 2024, GBIF 2024b). These entries come primarily from the Citizen Science platform iNaturalist, but also from other Citizen Science projects, universities, botanical gardens/arboreta, state institutions, and archives. All available data were downloaded and filtered according to two criteria. First, coordinates had to be provided, which led to the removal of 573 entries. Second, the coordinate uncertainty had to be less than or equal to 1,000 m, resulting in the removal of additional 485 entries. After filtering, 558 occurrences remained in the GBIF dataset. The data sources included two Citizen Science projects (iNaturalist with 546 occurrences and naturgucker with 5 occurrences) and two museums (Museo Argentino de Ciencias Naturales "Bernardino Rivadavia" with 6 occurrences and NTNU University Museum with 1 occurrence). The sampling years range between 1981–2024. Instagram ground truthing points and GBIF points were compared visually, quantitatively and by elevation using a DSM. The R package "sampbias" (Zizka et al. 2021) was used to assess whether the spatial sampling bias, influenced by site accessibility to humans, is reduced in the Instagram ground truthing point dataset. The package quantifies geographic bias and estimates the sampling rate across the study area using a Bayesian approach. Cities, roads, rivers, and lakes from Natural Earth Data (https://www.naturalearthdata.com/) were utilised as bias factors (gazetteers).

Usage of AI

ChatGPT (GPT-4 and GPT-3.5; available at https://chat.openai.com/) was used to enhance sentence structure and grammar in individual sentences.

Results

Instagram ground truthing approach

Numerous posts found by hashtags and location tags were reviewed in 2021 and 2022, resulting in 1238 traceable occurrence points. In total we found 297 suitable posts published between 2017 and 2022. Most posts were published in 2021. A total of 460 points were placed at the actual location of the posts, and 778 points were placed in the visible background area (mainly autumn coloured treeline locations). Posts with specific species information using the hashtags #nothofaguspumilio or #lenga provided 61 occurrence points. Posts with detailed location information, where *N. pumilio* is recognisable, contributed significantly to the occurrence data.

Comparison with GBIF data

Fig. 4 illustrates that the Instagram ground truthing points from the structured approach are less scattered. This is because the approach targeted the species' distribution range and gaps between already set points, when searching for specific locations, such as cities, national parks, mountain peaks, and lakes within that range. In contrast, the GBIF points are somewhat more scattered, with a few separate "outliers" visible in the west. The average elevation of the GBIF points is 559.82 m. Nevertheless, the GBIF dataset also includes very high-elevation locations, with the highest recorded location at 2123 m, compared to 1952 m for the Instagram ground truthing points. Instagram ground truthing points are higher on average at 1049.25 m, as images with a visible treeline were preferably selected. However, the differences in mean elevation are mainly due to the fact that the GBIF points were predominantly recorded in the southern part of the study area (45°S to 55°S). Of the 558 points, 419 are located in this area, with only 139 located to the north. In contrast, the Instagram ground truthing points are more homogeneously distributed: 661 of the 1238 points are located between 45°S and 55°S, while 577 points are in the northern part of the study area. However, the GBIF data also supplement the Instagram ground truthing data, particularly in the northernmost parts of the distribution area. 16 points are located further north of the Instagram ground truthing dataset. In some cases, GBIF data augment locations with Instagram ground truthing points by adding several GBIF points. With 252 points nearly half of the GBIF points are concentrated in a few tourist/urban areas. For example, 159 points are located in and around Ushuaia, 24 points in and around Bariloche, 53 points in Torres del Paine National Park, and 16 points solely at a parking area with viewing platform for the Perito Moreno Glacier. To quantify this observed bias, we used the "sampbias" package, which assesses sampling bias based on factors (gazetteers) indicating the influence of human-accessible locations (Zizka et al. 2021). The analysis revealed a significant impact of cities on both datasets, with the effect being higher for the GBIF dataset (bias weight, IGTA = 0.3455, GBIF = 0.4611). Additionally, roads and rivers exert a greater influence on the GBIF dataset. The Instagram ground truthing dataset, however, is more biased only in relation to lakes, which can be attributed to numerous points in national parks such as Torres del Paine and Los Glaciares. Fig. 5 presents the estimated sampling rates for both datasets. In Fig. 5 (B) the undersampled area between cities can be clearly seen. The bias weights and further results from the "sampbias" analysis are included as supplemental material (see Suppl. material 1).

Supervised classification

We found that the OpenCV Supervised Classification and Random Forest algorithms demonstrated the best performance. For the summer classification, the overall accuracy was 0.93 with a Kappa value of 0.89, and for the autumn classification, the overall accuracy was 0.97 with a Kappa value of 0.96 (see Suppl. material 1 for details

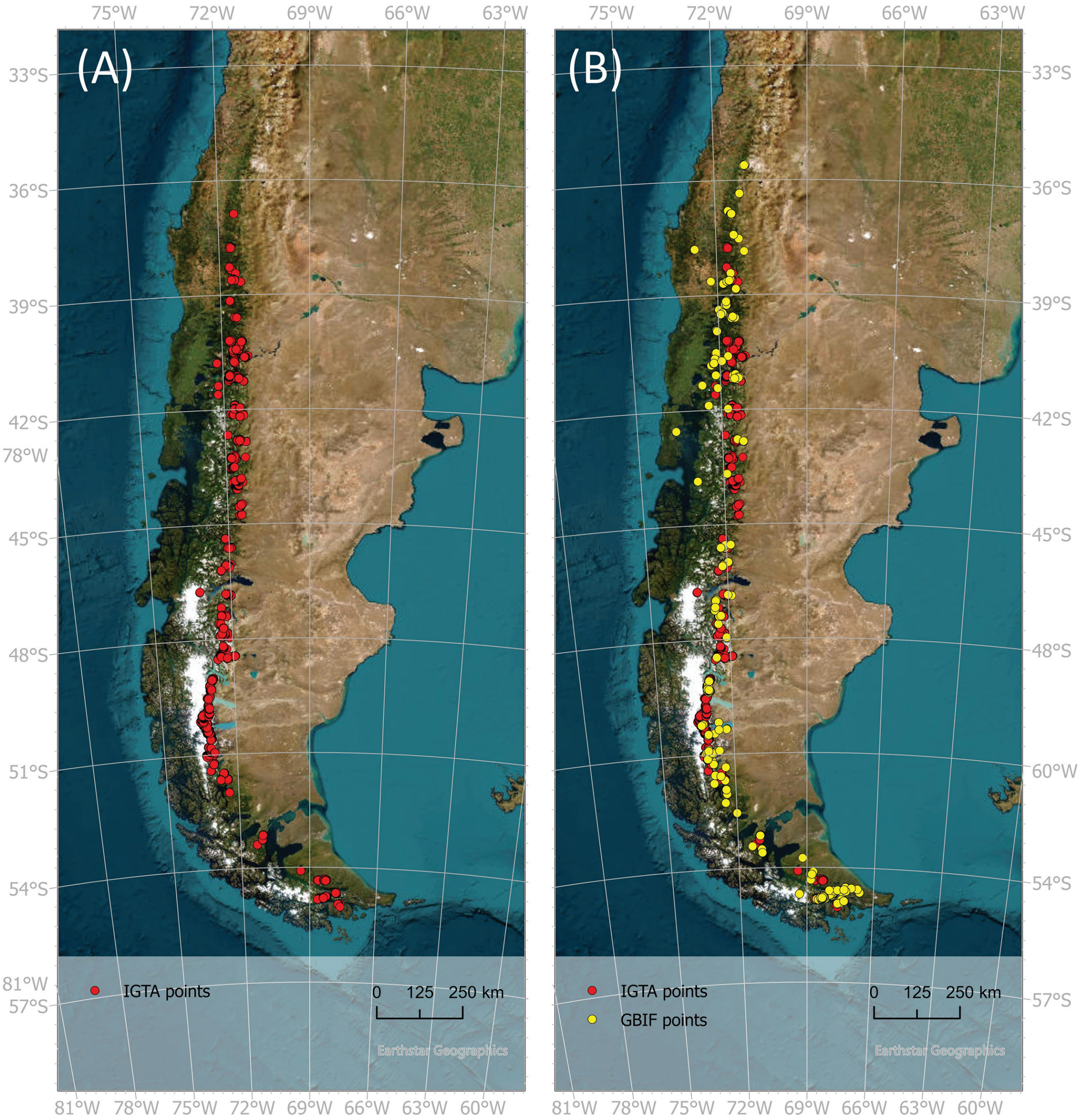


Figure 4. (**A**) 1238 points created by the Instagram ground truthing approach (IGTA) and (**B**) IGTA points (red) and 558 occurrence points from GBIF database (yellow, GBIF 2024b).

on the Supervised Classification validation). The Supervised Classification of summer and autumn Sentinel-2 data resulted in a spatial distribution map of the species *N. pumilio* from 35°10'S to 55°59'S (Fig. 6A). Detailed strengths (Fig. 6B) and weaknesses (Fig. 7) can be reviewed in the following figures.

In particular, the area extending east of the Northern and Southern Patagonian Ice Fields shows an accurate classification result: Three tree species dominate these areas, with *N. pumilio* and *N. antarctica* being deciduous species and *N. betuloides* being an evergreen species (Veblen et al. 1996). The occurrence of *N. pumilio* at the treeline was clearly distinguished from the occurrence of the evergreen species *N. betuloides* (class 2) below the treeline and areas of low vegetation (class 3). A clear distinction is also made in more arid regions in the east of the study area. Here, *N. pumilio* occurrences are clearly separated from scrub- and grassland. In the northern part of the study area, *N. pumilio* was reliably recorded at the treeline. However, in the valleys of this area there are also deciduous species that are misclassified as *N. pumilio*. We have removed these occurrences by an elevation correction, so that only occurrences of *N. pumilio* that could be unequivocally identified as such remained. However, at higher elevation, it was not possible to distinguish between the two deciduous and morphologically and ecologically similar species, *N. pumilio* and *N. antarctica*.

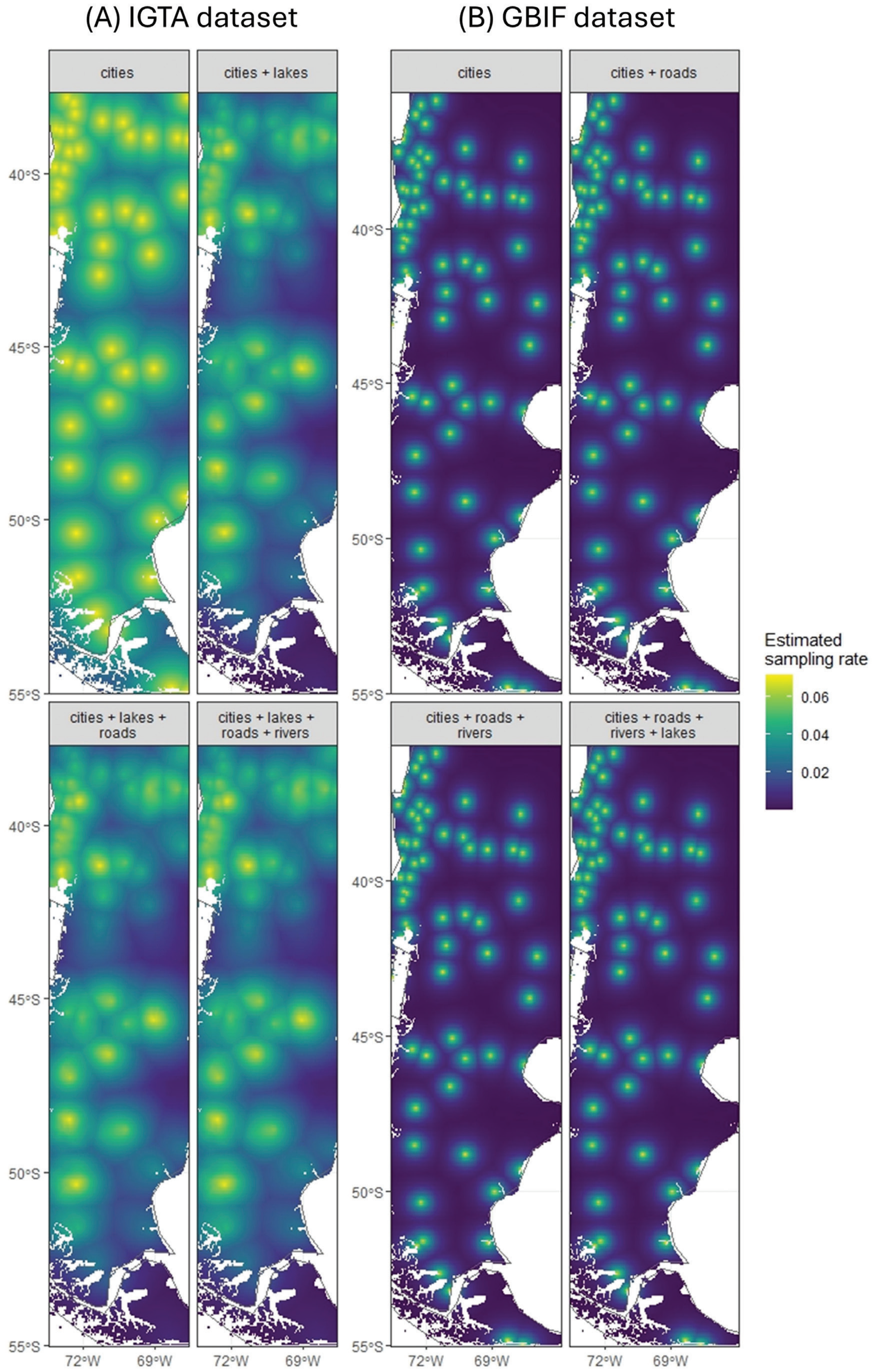


Figure 5. Visualised results of the "sampbias" analysis, indicating the sampling rate based on the influence of bias factors (gazetteers: cities, roads, rivers and lakes). In comparison, the IGTA dataset (**A**) displays more homogeneous sampling, whereas the GBIF dataset (**B**) shows undersampled areas, represented by dark blue regions.

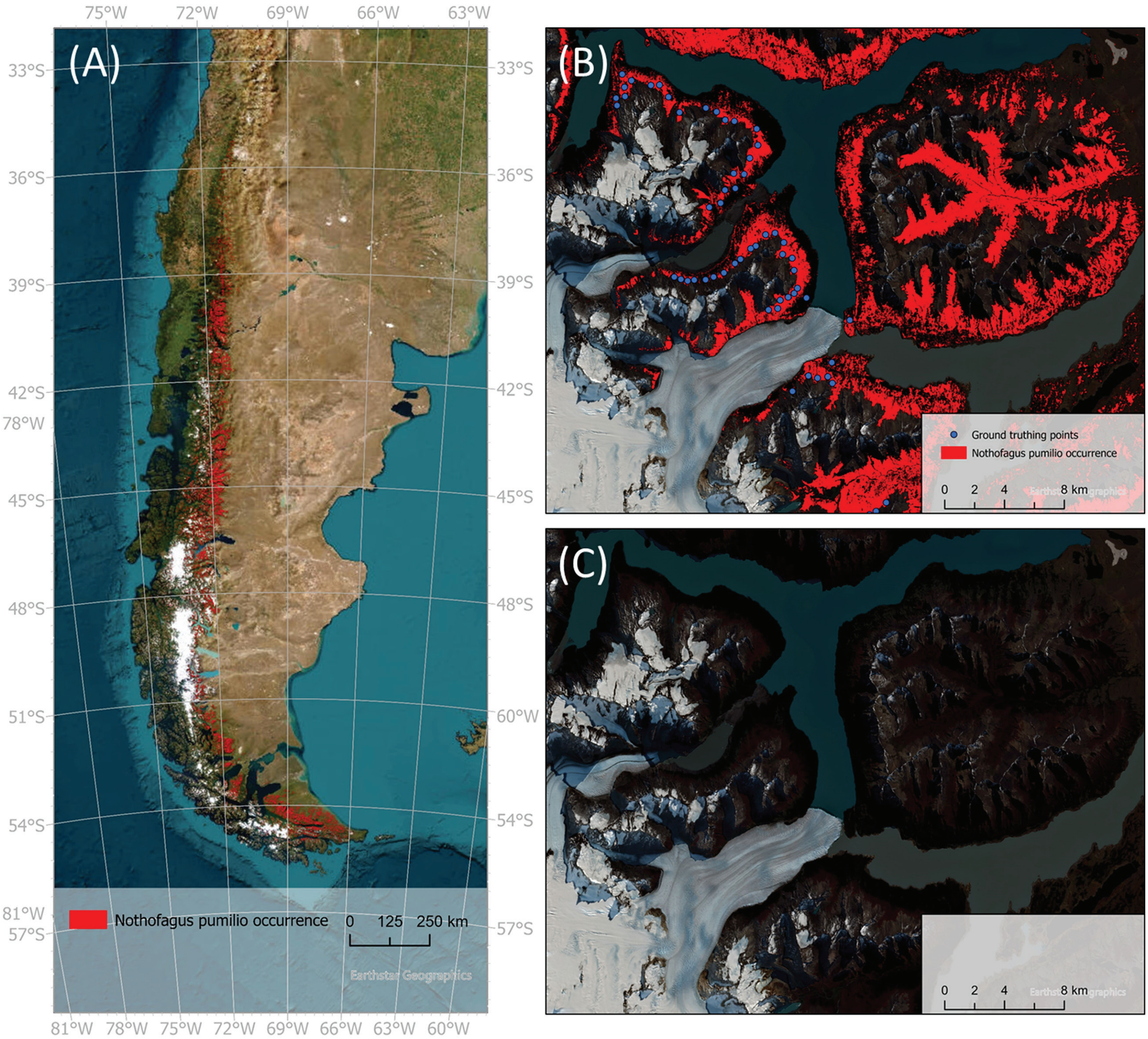


Figure 6. (**A**) *Nothofagus pumilio* occurrence determined by Supervised Classification and (**B**) occurrence of *Nothofagus pumilio* at Perito Moreno Glacier, with a classification result corresponding to the natural conditions, shown in comparison with the autumn Sentinel-2 scene in (**C**). All Instagram ground truthing points (blue, **B**) cover the identified occurrence.

Errors occur mainly due to high cloud cover and shadow effects. At the southern tip of Chile and Argentina, high cloud cover leads to data coverage problems. An area where the Sentinel-2 scene is not fully available in the sensing period of 2019 to 2022 and other scenes were almost completely covered by clouds is shown in Fig. 7A). A case where data are missing due to mountain shadows is visualised in Fig. 7B). Using the SCL resulted in shadow or shaded valleys being excluded from the analysis. Errors in validation with Instagram ground truthing points occur precisely in these areas.

Validation

We validated the classification result with the Instagram ground truthing and GBIF points by checking whether the points match the spatial occurrence. Out of 1238 Instagram ground truthing points, 1142 points are congruent with the remote sensing data, which is 92.25 %. 96 points (7.75 %) lie outside the areas classified as *N. pumilio*. These errors are probably due to mountain shadows and missing data, as we show in the results. Of the GBIF points, 157 (28.14 %) align with the spatial occurrence, while 401 (71.86 %) do not. However, many of these points lie just outside the determined spatial occurrence. Errors can also occur due to shadows and missing data. Other reasons may include the uncertainty of the coordinates, the image being recorded on roads or paths next to the occurrence rather than directly in the plant stand, or individual trees or stands being recorded in urban areas, evergreen forest stands, or open areas with low vegetation, which the classification does not categorise as *N. pumilio* areas. Fig. 8 provides two examples that support these hypotheses.

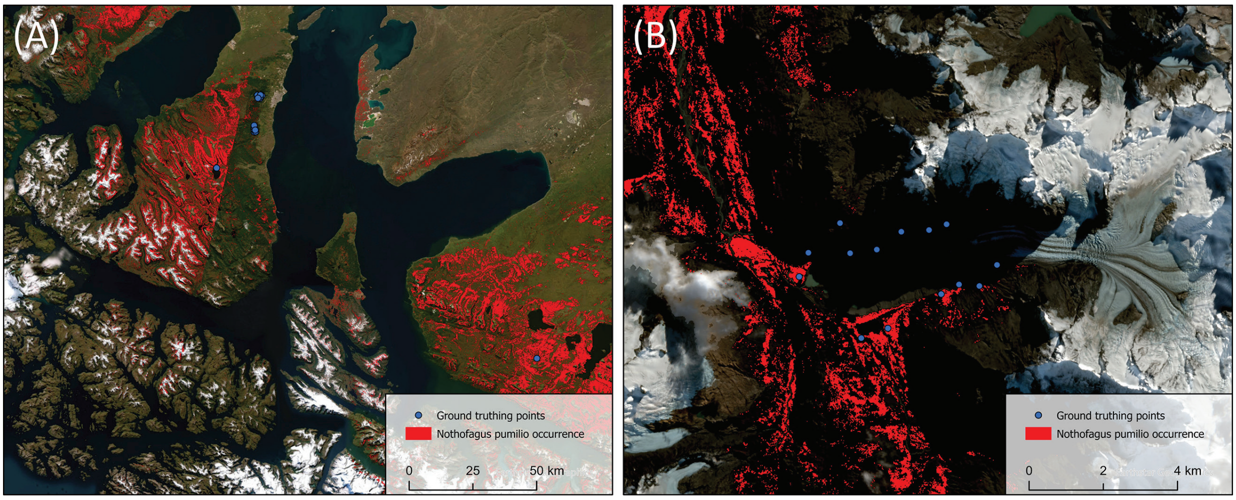


Figure 7. (**A**) A data gap at the southern tip of Chile. This is caused by missing Sentinel-2 data in the sensing period between 2019 and 2022 and very high cloud cover. (**B**) A valley with mountain shadows. To avoid errors in the spectral signals, these areas were removed during analysis using the Sentinel-2 Scene Classification Layer. However, this leads to a gap in the classification result and errors in the validation with the Instagram ground truthing points (blue).

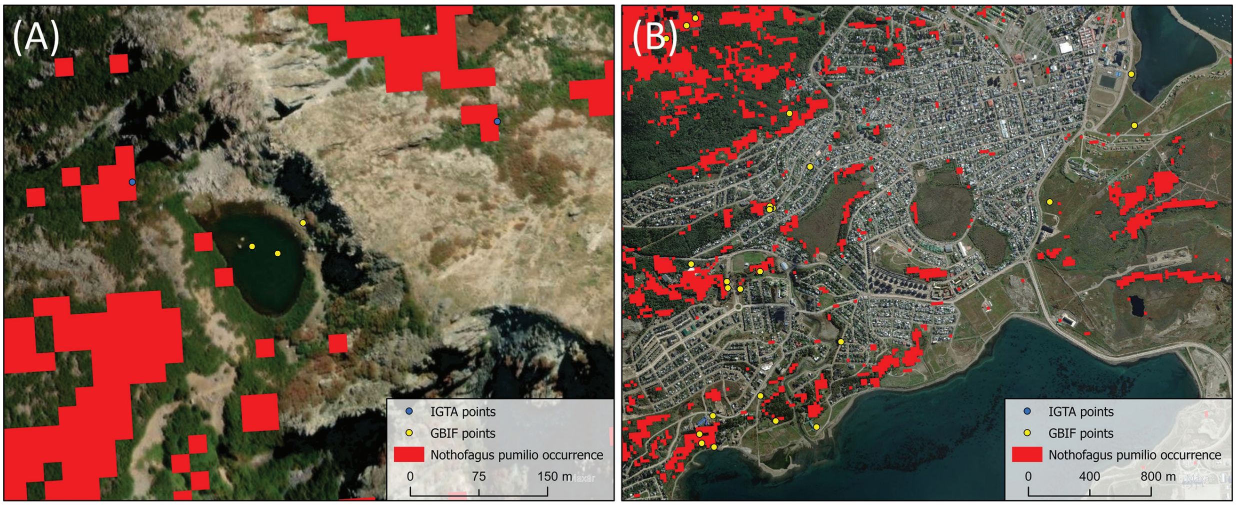


Figure 8. (**A**) Example of *Nothofagus pumilio* occurrence points from the GBIF database (GBIF points, yellow) with an uncertainty in coordinates, so that the points are located in a lake and not at the actual sampling location and (**B**) GBIF points in Ushuaia, where some raster cells, at locations of GBIF points, were not classified as vegetation.

Discussion

Instagram ground truthing approach as novel method for ground truthing

The availability of sufficient, non-biased species occurrence data has always been a major problem for ecological modelling studies (Bobrowski et al. 2021; Chauvier et al. 2021). An increasing wealth of information on the occurrences of species is becoming available through global databases (Michener et al. 2012; Feng et al. 2019). As field research is costly and time-consuming, open-source occurrence databases like the GBIF, with data compiled by Citizen Science are increasingly used in such studies (Feldman et al. 2021). However, even these advances do not replace the need for ground truthing procedures, as these data may contain spatial, temporal, and taxonomic biases (Beck et al. 2014; Meyer et al. 2016). In this study, we tested a novel Instagram ground truthing approach to generate ground truthing points for the species *N. pumilio* to validate species occurrence data derived from remote sensing. Furthermore, the Instagram ground truthing points were compared with existing GBIF data which led us to the conclusion that our approach offers the possibility to generate occurrence data with the potential to increase and improve existing data.

We include the sampling of ground truthing points on Instagram, and thus the re-use of social media posts in the realm of Citizen Science, although this is controversially discussed. According to the vignette study by Haklay et al. (2021), approximately 50 % of respondents describe the re-use of social media as Citizen Science. While Citizen

Science requires the active involvement of participants (Wiggins and Crowston 2011), in social media, content creators are rarely aware that they are participating in a study. Rather, permission to use imagery repurposed for occurrence information was given passively or unintentionally (Jarić et al. 2020). In addition, information about one's own study and its results (data transparency) is an important point for Citizen Science (Haklay et al. 2021). In our current work, we considered these criticisms as much as possible. By creating an own public Instagram account (www.instagram.com/nothofagus_pumilio_research/), where our study and results are presented, we aim at making a knowledge exchange possible. In addition, all images sourced from publicly accessible profiles used for our analysis were "liked" to draw the attention of content creators to the account. Users whose images were used in publications were asked for permission in advance and informed of the purpose. Furthermore, Instagram's chat function was used to exchange information with Instagram users about a photo's location or the species itself. Despite the discourse on terminology, we primarily focus on discussing the advantages and disadvantages of Citizen Science occurrence data sampling and exploring potential improvements to our method.

N. pumilio is suitable for the Instagram approach due to its distinctive characteristics and visibility in satellite images at the treeline in mono-species forest stands. Additionally, its occurrence in national parks, where tourists often take photos, increases the number of Instagram posts. The species' autumn colouring further enhances its aesthetic appeal, leading to even more posts. These advantages, also benefit occurrence data sampling in unstructured Citizen Science projects. In unstructured projects, user behaviour of Citizen Science participants leads to observations with spatial bias (Di Cecco et al. 2021), mainly concentrated in tourist locations or urban areas, such as near cities or along roads (Reddy and Dávalos 2003; Graham et al. 2004; Fithian et al. 2015; Chauvier et al. 2021). Since the direct sampling location of the Citizen Science participants is usually recorded, remote locations that are difficult to access are not included. Other studies have already addressed this spatial bias in GBIF data and highlighted potential pitfalls for ecological conclusions (Boakes et al. 2010; Beck et al. 2014; Meyer et al. 2016), which can be further emphasized with the analysed dataset where nearly half of the points are located in urban or touristic locations. Points derived from Instagram are also not free of this bias, but as the Instagram ground truthing approach allowed to identify occurrence points in the background of suitable posts and additional points at autumn-coloured treelines, this spatial sampling bias is efficiently reduced. Additionally, the use of Instagram, with 2 billion users worldwide (We Are Social et al. 2024) and millions of photos uploaded daily (60 million daily uploads by 2014, WirtschaftsWoche 2014), allows us to analyse a large number of posts that are potentially suitable for analysis. The large number of possibly suitable posts on Instagram and a structured sampling approach further reduced spatial bias. After posts were found using specific hashtags (#nothofaguspumilio and #lenga), gaps between these occurrences were then closed by searching for specific locations. Contributions were found where the species was not specified by the Instagram user but was clearly recognisable. This ensured that the study area was covered as homogeneously as possible. The improvement of the Instagram ground truthing approach compared to other Citizen Science projects lies particularly in the inclusion of images taken by non-experts, where the species was recorded even though the social media user might not be aware of it. Instagram exclusively shares photos and videos and is one of the largest social media platforms, offering the possibility to analyse a large quantity of suitable posts. This is also reflected in the quantitative comparison. While the GBIF database contains 558 data points from 1981 to 2024 after filtering, the Instagram ground truthing approach was able to create 1238 points from posts dated 2017 to 2022.

Data from Citizen Science projects are improving and even approaching the quality of expert data (Mesaglio and Callaghan 2021). For example, species identifications in iNaturalist, the main source of GBIF data used in this study, must first be confirmed by a 2/3 consensus. iNaturalist also offers suggestions for species identification based on examples, which are reviewed and updated by experts (curators). Only after occurrences are considered complete and certain are they passed on to the GBIF database for publication as valid data (Heberling and Isaac 2018). Although the sampling bias in species identification is decreasing, the sampling bias in terms of coordinate accuracy still needs improvement. Coordinate uncertainties result from georeferencing methods of museum data (Marcer et al. 2022) or weak satellite signals while sampling with mobile devises (Uyeda et al. 2020). We included GBIF data with an uncertainty of up to 1 km, although such a deviation is substantial, especially in a highly complex ecosystem like high mountains. The validation process showed that deviations in the coordinates lead to errors. Precise filtering and selection of data from databases are necessary. Although the manual creation of occurrences during the Instagram ground truthing approach was time-consuming, this reduced such bias.

The limitation of the Instagram ground truthing approach is the time-consuming and non-automated process. While sampling by Citizen Science projects is a way of data collection in which data can be collected particularly quickly and cost-effectively with a large reach (Kullenberg and Kasperowski 2016; Sumner et al. 2019), the manual search for suitable posts and map transfer is time-consuming. Other studies using social media to investigate species occurrences carried out an automated search of suitable posts via the Application Programming Interface (API) of the chosen social media platform (e.g., Flickr: Alampi Sottini et al. 2019; August et al. 2020). The Instagram API is not as easy to access. At this stage, manual analysis remains unavoidable when using Instagram, and it is similarly required for platforms such as Facebook and Twitter, as demonstrated by O'Neill et al. (2023). In the future, AI-based image recognition processes could be an efficient way to find suitable posts on Instagram more quickly. Recently, Instagram's terms of usage were changed to allow for this purpose. By using Instagram, users agree that their content may be analysed by AI (Meta 2024).

Another limitation is the accuracy of the recording date. Caution is needed regarding the exact timing of when a photo was taken and posted on Instagram, as the publication date may not correspond to the actual date the photo was captured. In our analysis, we primarily focused on recent posts, covering the period from 2017 to 2022, while the GBIF data includes records dating as far back as 1981. Verifying the accuracy of the Instagram date is essential (if necessary, by contacting the post creators), particularly for temporal distribution analyses of species. "Historical" data may not be available on Instagram at all. Additionally, discrepancies in acquisition dates between Sentinel-2 and Instagram data may introduce potential sources of error in the validation process. We used Sentinel-2 data from 2019 to 2022, while Instagram posts date back to 2017. Changes in forest stands, such as deforestation or forest fires, could result in discrepancies.

Supervised classification

The occurrence in mono-species forest stands at the treeline and the phenology of *N. pumilio* allows a precise creation of training areas and a reliable result of the Supervised Classification. With the Sentinel-2 level 2A data in a resolution of 20 m we achieved an accurate classification over a very large study area of about 2000 km latitudinal extent in the southern Andes. Such a resolution is sufficient for subsequent modelling, which often uses climate data at a resolution of 30 arcseconds (~1 km), for example, WorldClim and Bioclim data (e.g., Bobrowski et al. 2018). Furthermore, the high temporal resolution of the Sentinel-2 data provided many scenes, that were analysed in the Supervised Classification. By masking the Sentinel-2 scenes with the SCL, sources of error due to too many non-vegetation classes were avoided. It also removed vegetation grid cells covered by clouds, aerosols, and shadows, which can lead to classification errors. Consequently, the three vegetation classes (1 = *N. pumilio*, 2 = Evergreen vegetation, 3 = Low vegetation) were trained using training areas that contained only the spectral information of representative vegetation raster cells. However, this also created gaps in the classification result that could not be filled even with the large number of Sentinel-2 scenes.

A different source of error is that in some cases stands of *N. antarctica* were classified as stands of *N. pumilio* in high and low elevation areas. The two deciduous species are very similar, both phenologically and ecologically, and often share the same range. Hybrids of the two species are also possible (Soliani et al. 2015). Inclusion of *N. antarctica* occurrence data, e.g., data from local forestry services (e.g data from Corporación Nacional Forestal (CONAF), Chile and "Ordenamiento Territorial de Bosque Nativo" (OTBN), Argentina), or repeating the Instagram ground truthing approach for this species could further specify the result. Nevertheless, the forest type was determined with certainty at this stage. High mountain deciduous forest was correctly identified even if *N. antarctica* was partially included.

In more southerly areas, *N. pumilio* and *N. antarctica* dominate as deciduous species. In the north, however, many other deciduous species occur at lower elevations, while *N. pumilio* occurs only at the treeline. For this reason, an elevational correction of the result was necessary. Occurrences below subalpine forest stands of *N. pumilio* have been removed. Individual stands below these are also less relevant for treeline modelling studies. The thresholds (800 m, 500 m, 250 m) were estimated from literature data and from the classification result in order to obtain the most accurate classification result with the least data loss.

Conclusion

Citizen Science and social media-based occurrence sampling is developing and improving rapidly, becoming an important source of species occurrence data, especially for large-scale modelling approaches where alternatives are limited. However, resulting data are not free from bias and need to be filtered and verified before being used in applications such as ecological models. We conclude that using social media posts on Instagram in a structured Instagram ground truthing approach leads to less-biased occurrence data for *N. pumilio* in comparison with GBIF data. Sampling biases are further minimised by combining the Instagram ground truthing method with Supervised Classification, as large-scale occurrence data are generated across the entire distribution range of the species, rather than just in urban or tourist locations where most pictures are taken building the basis for Citizen Scientist observations or Instagram posts. We further conclude that the Instagram ground truthing approach is a novel method that can complement occurrence data sampling methods and be applied to other suitable species. However, it is essential that landscape elements are visible in the posts, which is more likely for landscape images and less so for detailed images of smaller herbaceous plant or animal species. Future work could focus on creating an automated search for Instagram posts using Instagram API and AI technology to replace the time-consuming manual search and further increase the availability of suitable posts. We believe that using social media can unlock significant potential for species occurrence data sampling and thus promote research on species in remote and high-elevation regions. Furthermore, the spatial occurrence data of *N. pumilio* enables presence-absence modelling approaches, that can provide detailed insights into the current and future distribution of *N. pumilio*.

Acknowledgments

We would like to thank all the Instagram users who have supported us with their posts, with information about the location where their pictures were taken and sharing information about the occurrence of *N. pumilio* and *N. antarctica*. Special thanks to user fernando.v.fotografia for allowing us to publish his post.

Author contributions

All authors participated during project design. MW conceived and developed the idea, collected, and analysed the data and drafted the first manuscript, JW and MB were significantly involved in the conception of the idea and the development and implementation of the analyses, MB also managed the whole project, JB and US were the reviewers and editors of the manuscript, providing their expertise on the topic and assistance with the literature. All authors have read and agreed to the published version of the manuscript.

Data accessibility statement

The Instagram ground truthing points and the spatial species occurrence data generated in our study are currently in the process of being published with the open access data provider of the University of Hamburg.

Supplementary material: The following materials are available as part of the online article at https://escholarship.org/uc/fb.

References

Aceves-Bueno E, Adeleye AS, Feraud M, Huang Y, Tao M, Yang Y, Anderson SE (2017) The Accuracy of Citizen Science Data: A Quantitative Review. Bulletin Ecological Society of America 98: 278–290. https://doi.org/10.1002/bes2.1336

Akobia I, Janiashvili Z, Metreveli V, Zazanashvili N, Batsatsashvili K, Ugrekhelidze K (2022) Modelling the potential distribution of subalpine birches (*Betula* spp.) in the Caucasus. Community Ecology 23: 209–218. https://doi.org/10.1007/s42974-022-00097-4

Alampi Sottini V, Barbierato E, Bernetti I, Capecchi I, Fabbrizzi S, Menghini S (2019) Rural environment and landscape quality: an evaluation model integrating social media analysis and geostatistics techniques. Aestimum 74: 43–62. https://doi.org/10.13128/aestim-7379

Álvarez C, Veblen TT, Christie DA, González-Reyes Á (2015) Relationships between climate variability and radial growth of *Nothofagus pumilio* near altitudinal treeline in the Andes of northern Patagonia, Chile. Forest Ecology and Management 342: 112–121. https://doi.org/10.1016/j.foreco.2015.01.018

Amigo J, Rodríguez-Guitián MA (2011) Bioclimatic and phytosociological diagnosis of the species of the *Nothofagus* genus (Nothofagaceae) in South America. International Journal of Geobotanical Research 1: 1–20. https://doi.org/10.5616/ijgr110001

Aravena JC, Lara A, Wolodarsky-Franke A, Villalba R, Cuq E (2002) Tree-ring growth patterns and temperature reconstruction from *Nothofagus pumilio* (Fagaceae) forests at the upper tree line of southern Chilean Patagonia. Revista chilena de historia natural 75: 361–376. https://doi.org/10.4067/S0716-078X2002000200008

August TA, Pescott OL, Joly A, Bonnet P (2020) AI Naturalists Might Hold the Key to Unlocking Biodiversity Data in Social Media Imagery. Patterns 1: 100116. https://doi.org/10.1016/j.patter.2020.100116

Beck J, Böller M, Erhardt A, Schwanghart W (2014) Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. Ecological Informatics 19: 10–15. https://doi.org/10.1016/j.ecoinf.2013.11.002

Boakes EH, McGowan PJK, Fuller RA, Chang-qing D, Clark NE, O'Connor K, Mace GM (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biology 8: e1000385. https://doi.org/10.1371/journal.pbio.1000385

Bobrowski M, Bechtel B, Böhner J, Oldeland J, Weidinger J, Schickhoff U (2018) Application of Thermal and Phenological Land Surface Parameters for Improving Ecological Niche Models of *Betula utilis* in the Himalayan Region. Remote Sensing 10: 814. https://doi.org/10.3390/rs10060814

Bobrowski M, Gerlitz L, Schickhoff U (2017) Modelling the potential distribution of *Betula utilis* in the Himalaya. Global Ecology and Conservation 11: 69–83. https://doi.org/10.1016/j.gecco.2017.04.003

Bobrowski M, Weidinger J, Schwab N, Schickhoff U (2021) Searching for ecology in species distribution models in the Himalayas. Ecological Modelling 458: 109693. https://doi.org/10.1016/j.ecolmodel.2021.109693

Bonney R (1996) Citizen science: a lab tradition. Living Bird 15: 7–15.

Bonney R, Cooper CB, Dickinson J, Kelling S, Phillips T, Rosenberg KV, Shirk J (2009) Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. BioScience 59: 977–984. https://doi.org/10.1525/bio.2009.59.11.9

Chauvier Y, Zimmermann NE, Poggiato G, Bystrova D, Brun P, Thuiller W (2021) Novel methods to correct for observer and sampling bias in presence-only species distribution models. Global Ecology and Biogeography 30: 2312–2325. https://doi.org/10.1111/geb.13383

Chowdhury S, Fuller RA, Ahmed S, Alam S, Callaghan CT, Das P, Correia RA, Di Marco M, Di Minin E, Jarić I, Labi MM, Ladle RJ, Rokonuzzaman M, Roll U, Sbragaglia V, Siddika A, Bonn A (2023) Using social media records to inform conservation planning. Conservation Biology 38: e14161. https://doi.org/10.1111/cobi.14161

Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, Wehberg J, Wichmann V, Böhner J (2015) System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geoscientific Model Development 8: 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015

Cuevas JG (2000) Tree recruitment at the *Nothofagus pumilio* alpine timberline in Tierra del Fuego, Chile. Journal of Ecology 88: 840–855. https://doi.org/10.1046/j.1365-2745.2000.00497.x

Daniels LD, Veblen TT (2003) Regional and local effects of disturbance and climate on altitudinal treelines in northern Patagonia. Journal of Vegetation Science 14: 733–742. https://doi.org/10.1111/j.1654-1103.2003.tb02205.x

Daniels LD, Veblen TT (2004) Spatiotemporal Influences of Climate on Altitudinal Treeline in Northern Patagonia. Ecology 85: 1284–1296. https://doi.org/10.1890/03-0092

Daru BH, Park DS, Primack RB, Willis CG, Barrington DS, Whitfeld TJS, Seidler TG, Sweeney PW, Foster DR, Ellison AM, Davis CC (2018) Widespread sampling biases in herbaria revealed from large-scale digitization. The New Phytologist 217: 939–955. https://doi.org/10.1111/nph.14855

Di Cecco GJ, Barve V, Belitz MW, Stucky BJ, Guralnick RP, Hurlbert AH (2021) Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science. BioScience 71: 1179–1188. https://doi.org/10.1093/biosci/biab093

Dickinson JL, Zuckerberg B, Bonter DN (2010) Citizen Science as an Ecological Research Tool: Challenges and Benefits. Annual Review of Ecology, Evolution, and Systematics 41: 149–172. https://doi.org/10.1146/annurev-ecolsys-102209-144636

Dullinger S, Dirnböck T, Grabherr G (2004) Modelling climate change-driven treeline shifts: relative effects of temperature increase, dispersal and invasibility. Journal of Ecology 92: 241–252. https://doi.org/10.1111/j.0022-0477.2004.00872.x

Edwards JL (2004) Research and Societal Benefits of the Global Biodiversity Information Facility. BioScience 54: 485–486. https://doi.org/10.1641/0006-3568(2004)054[0486:RASBOT]2.0.CO;2

ElQadi MM, Dorin A, Dyer A, Burd M, Bukovac Z, Shrestha M (2017) Mapping species distributions with social media geo-tagged images: Case studies of bees and flowering plants in Australia. Ecological Informatics 39: 23–31. https://doi.org/10.1016/j.ecoinf.2017.02.006

Fajardo A, Piper FI (2014) An experimental approach to explain the southern Andes elevational treeline. American Journal of Botany 101: 788–795. https://doi.org/10.3732/ajb.1400166

Fassnacht FE, Latifi H, Stereńczak K, Modzelewska A, Lefsky M, Waser LT, Straub C, Ghosh A (2016) Review of studies on tree species classification from remotely sensed data. Remote Sensing of Environment 186: 64–87. https://doi.org/10.1016/j.rse.2016.08.013

Feldman MJ, Imbeau L, Marchand P, Mazerolle MJ, Darveau M, Fenton NJ (2021) Trends and gaps in the use of citizen science derived data as input for species distribution models: A quantitative review. PLOS ONE 16: e0234587. https://doi.org/10.1371/journal.pone.0234587

Feng X, Park DS, Walker C, Peterson AT, Merow C, Papeş M (2019) A checklist for maximizing reproducibility of ecological niche models. Nature Ecology & Evolution 3: 1382–1395. https://doi.org/10.1038/s41559-019-0972-5

fernando.v.fotografia (2022) Instagram post. Example of the Instagram ground truthing approach at Laguna Capri. https://www.instagram.com/p/ChLmWxQuVjh/ [Last accessed on 29/10/2024]

Fithian W, Elith J, Hastie T, Keith DA (2015) Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods in Ecology and Evolution 6: 424–438. https://doi.org/10.1111/2041-210X.12242

GBIF (2024a) What is GBIF? https://www.gbif.org/what-is-gbif [Last accessed on 29/10/2024]

GBIF (2024b) GBIF Occurrence Download (6 June 2024). https://doi.org/10.15468/dl.3kv29u

Gibson CE, Williams D, Dunlop R, Beck S (2020) Using social media as a cost-effective resource in the photo-identification of a coastal bottlenose dolphin community. Aquatic Conservation: Marine and Freshwater Ecosystems 30: 1702–1710. https://doi.org/10.1002/aqc.3356

Goldberg JK (2023) iNaturalist is an open science resource for ecological genomics by enabling rapid and tractable records of initial observations of sequenced biological samples. Biology Letters 19: 20230251. https://doi.org/10.1098/rsbl.2023.0251

Graham CH, Ferrier S, Huettman F, Moritz C, Peterson AT (2004) New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology & Evolution 19: 497–503. https://doi.org/10.1016/j.tree.2004.07.006

Haklay M, Fraisl D, Greshake Tzovaras B, Hecker S, Gold M, Hager G, Ceccaroni L, Kieslinger B, Wehn U, Woods S, Nold C, Balázs B, Mazzonetto M, Ruefenacht S, Shanley LA, Wagenknecht K, Motion A, Sforzi A, Riemenschneider D, Dorler D, Heigl F, Schaefer T, Lindner A, Weißpflug M, Mačiulienė M, Vohland K (2021) Contours of citizen science: a vignette study. Royal Society Open Science 8: 202108. https://doi.org/10.1098/rsos.202108

Hansson A, Dargusch P, Shulmeister J (2021) A review of modern treeline migration, the factors controlling it and the implications for carbon storage. Journal of Mountain Science 18: 291–306. https://doi.org/10.1007/s11629-020-6221-1

Hansson A, Shulmeister J, Dargusch P, Hill G (2023) A review of factors controlling Southern Hemisphere treelines and the implications of climate change on future treeline dynamics. Agricultural and Forest Meteorology 332: 109375. https://doi.org/10.1016/j.agrformet.2023.109375

Heberling JM, Isaac BL (2018) iNaturalist as a tool to expand the research value of museum specimens. Applications in Plant Sciences 6: e01193. https://doi.org/10.1002/aps3.1193

Hentati-Sundberg J, Olsson O (2016) Amateur photographs reveal population history of a colonial seabird. Current Biology 26: R226–R228. https://doi.org/10.1016/j.cub.2016.02.007

Hertel D, Therburg A, Villalba R (2008) Above- and below-ground response by *Nothofagus pumilio* to climatic conditions at the transition from the steppe–forest boundary to the alpine treeline in southern Patagonia, Argentina. Plant Ecology & Diversity 1: 21–33. https://doi.org/10.1080/17550870802257026

Hildebrand-Vogel R, Godoy R, Vogel A (1990) Subantarctic-Andean *Nothofagus pumilio* Forests. Distribution Area and Synsystematic Overview; Vegetation and Soils as Demonstrated by an Example of a South Chilean Stand. Vegetatio 89: 55–68. https://doi.org/10.1007/BF00134434

Immitzer M, Neuwirth M, Böck S, Brenner H, Vuolo F, Atzberger C (2019) Optimal Input Features for Tree Species Classification in Central Europe Based on Multi-Temporal Sentinel-2 Data. Remote Sensing 11: 2599. https://doi.org/10.3390/rs11222599

Immitzer M, Vuolo F, Atzberger C (2016) First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. Remote Sensing 8: 166. https://doi.org/10.3390/rs8030166

Jarić I, Correia RA, Brook BW, Buettel JC, Courchamp F, Di Minin E, Firth JA, Gaston KJ, Jepson P, Kalinkat G, Ladle R, Soriano-Redondo A, Souza AT, Roll U (2020) iEcology: Harnessing Large Online Resources to Generate Ecological Insights. Trends in Ecology & Evolution 35: 630–639. https://doi.org/10.1016/j.tree.2020.03.003

Jaxa EORC (2023) ALOS Global Digital Surface Model "ALOS World 3D - 30m (AW3D30)", Ver.4.0, Japan Aerospace Exploration Agency. https://www.eorc.jaxa.jp/ALOS/en/dataset/aw3d30/aw3d30_e.htm [Last accessed on 29/10/2024]

Kullenberg C, Kasperowski D (2016) What Is Citizen Science? - A Scientometric Meta-Analysis. PLOS ONE 11: e0147152. https://doi.org/10.1371/journal.pone.0147152

Lara A, Aravena JC, Villalba R, Wolodarsky-Franke A, Luckman B, Wilson R (2001) Dendroclimatology of high-elevation *Nothofagus pumilio* forests at their northern distribution limit in the central Andes of Chile. Canadian Journal of Forest Research 31: 925–936. https://doi.org/10.1139/cjfr-31-6-925

Lara A, Villalba R, Wolodarsky-Franke A, Aravena JC, Luckman BH, Cuq E (2005) Spatial and temporal variation in *Nothofagus pumilio* growth at tree line along its latitudinal range (35°40'-55° S) in the Chilean Andes. Journal of Biogeography 32: 879–893. https://doi.org/10.1111/j.1365-2699.2005.01191.x

La Sorte FA, Cohen JM, Jetz W (2024) Data coverage, biases, and trends in a global citizen-science resource for monitoring avian diversity. Diversity and Distributions 30: e13863. https://doi.org/10.1111/ddi.13863

Marcer A, Chapman AD, Wieczorek JR, Xavier Picó F, Uribe F, Waller J, Ariño AH (2022) Uncertainty matters: ascertaining where specimens in natural history collections come from and its implications for predicting species distributions. Ecography 2022: e06025. https://doi.org/10.1111/ecog.06025

Martino S, Pace DS, Moro S, Casoli E, Ventura D, Frachea A, Silvestri M, Arcangeli A, Giacomini G, Ardizzone G, Lasinio GJ (2021) Integration of presence-only data from several sources: a case study on dolphins' spatial distribution. Ecography 44: 1533–1543. https://doi.org/10.1111/ecog.05843

Masiokas M, Villalba R (2004) Climatic significance of intra-annual bands in the wood of *Nothofagus pumilio* in southern Patagonia. Trees 18: 696–704. https://doi.org/10.1007/s00468-004-0355-6

Mesaglio T, Callaghan CT (2021) An overview of the history, current contributions and future outlook of iNaturalist in Australia. Wildlife Research 48: 289–303. https://doi.org/10.1071/WR20154

Meta (2024) Terms of Usage (June 2024). https://www.facebook.com/help/instagram/581066165581870/?helpref=hc_fnav [Last accessed on 29/10/2024]

Meyer C, Weigelt P, Kreft H (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology Letters 19: 992–1006. https://doi.org/10.1111/ele.12624

Michener WK, Allard S, Budden A, Cook RB, Douglass K, Frame M, Kelling S, Koskela R, Tenopir C, Vieglais DA (2012) Participatory design of DataONE—Enabling cyberinfrastructure for the biological and environmental sciences. Ecological Informatics 11: 5–15. https://doi.org/10.1016/j.ecoinf.2011.08.007

Nagai S, Nasahara KN, Akitsu TK, Saitoh TM, Muraoka H (2020) Importance of the Collection of Abundant Ground-Truth Data for Accurate Detection of Spatial and Temporal Variability of Vegetation by Satellite Remote Sensing. In: Dontsova K, Balogh-Brunstadt Z, Le Roux G (Eds) Biogeochemical cycles. Ecological drivers and environmental impact. Wiley, 223–244. https://doi.org/10.1002/9781119413332.ch11

Nagy L, Eller CB, Mercado LM, Cuesta FX, Llambí LD, Buscardo E, Aragão LEOC, García-Núñez C, Oliveira RS, Barbosa M, Ceballos SJ, Calderón-Loor M, Fernandes GW, Aráoz E, Muñoz AMQ, Rozzi R, Aguirre F, Álvarez-Dávila E, Salinas N, Sitch S (2023) South American mountain ecosystems and global change – a case study for integrating theory and field observations for land surface modelling and ecosystem management. Plant Ecology & Diversity 16: 1–27. https://doi.org/10.1080/17550874.2023.2196966

O'Neill D, Häkkinen H, Neumann J, Shaffrey L, Cheffings C, Norris K, Pettorelli N (2023) Investigating the potential of social media and citizen science data to track changes in species' distributions. Ecology and Evolution 13: e10063. https://doi.org/10.1002/ece3.10063

Pace DS, Giacomini G, Campana I, Paraboschi M, Pellegrino G, Silvestri M, Alessi J, Angeletti D, Cafaro V, Pavan G, Ardizzone G, Arcangeli A (2019) An integrated approach for cetacean knowledge and conservation in the central Mediterranean Sea using research and social media data sources. Aquatic Conservation: Marine and Freshwater Ecosystems 29: 1302–1323. https://doi.org/10.1002/aqc.3117

Pocock MJO, Tweddle JC, Savage J, Robinson LD, Roy HE (2017) The diversity and evolution of ecological and environmental citizen science. PLOS ONE 12: e0172579. https://doi.org/10.1371/journal.pone.0172579

Ranghetti L, Boschetti M, Nutini F, Busetto L (2020) "sen2r": An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data. Computers & Geosciences 139: 104473. https://doi.org/10.1016/j.cageo.2020.104473

R Core Team (2023) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reddy S, Dávalos LM (2003) Geographical sampling bias and its implications for conservation priorities in Africa. Journal of Biogeography 30: 1719–1727. https://doi.org/10.1046/j.1365-2699.2003.00946.x

Richards JA (2022) Image Classification in Practice. In: Richards JA (Ed.) Remote Sensing Digital Image Analysis. Cham: Springer International Publishing, 447–502. https://doi.org/10.1007/978-3-030-82327-6_11

Rodríguez-Catón M, Villalba R, Morales M, Srur A (2016) Influence of droughts on *Nothofagus pumilio* forest decline across northern Patagonia, Argentina. Ecosphere 7: e01390. https://doi.org/10.1002/ecs2.1390

Schickhoff U, Bobrowski M, Mal S, Schwab N, Singh RB (2022) The World's Mountains in the Anthropocene. In: Schickhoff U, Singh RB, Mal S (Eds) Mountain Landscapes in Transition. Sustainable Development Goals Series. Springer, Cham., 1–144. https://doi.org/10.1007/978-3-030-70238-0_1

Soliani C, Tsuda Y, Bagnoli F, Gallo LA, Vendramin GG, Marchelli P (2015) Halfway encounters: meeting points of colonization routes among the southern beeches *Nothofagus pumilio* and *N. antarctica*. Molecular Phylogenetics and Evolution 85: 197–207. https://doi.org/10.1016/j.ympev.2015.01.006

Srur AM, Villalba R, Rodríguez-Catón M, Amoroso MM, Marcotti E (2016) Establishment of *Nothofagus pumilio* at Upper Treelines Across a Precipitation Gradient in the Northern Patagonian Andes. Arctic, Antarctic, and Alpine Research 48: 755–766. https://doi.org/10.1657/AAAR0016-015

Srur AM, Villalba R, Rodríguez-Catón M, Amoroso MM, Marcotti E (2018) Climate and *Nothofagus pumilio* Establishment at Upper Treelines in the Patagonian Andes. Frontiers in Earth Science 6: 57. https://doi.org/10.3389/feart.2018.00057

Sumner S, Bevan P, Hart AG, Isaac NJ (2019) Mapping species distributions in 2 weeks using citizen science. Insect Conservation and Diversity 12: 382–388. https://doi.org/10.1111/icad.12345

Uyeda KA, Stow DA, Richart CH (2020) Assessment of volunteered geographic information for vegetation mapping. Environmental monitoring and assessment 192: 554. https://doi.org/10.1007/s10661-020-08522-9

Veblen TT, Hill RS, Read J [Eds] (1996) The ecology and biogeography of *Nothofagus* forests. Yale University Press.

Virić Gašparić H, M. Mikac K, Pajač Živković I, Krehula B, Orešković M, Galešić MA, Ninčević P, Varga F, Lemić D (2022) Firefly Occurrences in Croatia – One Step Closer from Citizen Science to Open Data. Interdisciplinary Description of Complex Systems 20: 112–124. https://doi.org/10.7906/indecs.20.2.4

We Are Social, DataReportal, Meltwater (2024) Most popular social networks worldwide as of January 2024, ranked by number of monthly active users (in millions). Statista. Statista GmbH. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [Last accessed on 29/10/2024]

Wiggins A, Crowston K (2011) From Conservation to Crowdsourcing: A Typology of Citizen Science. In: 2011 44th Hawaii International Conference on System Sciences. IEEE, 1–10. https://doi.org/10.1109/HICSS.2011.207

WirtschaftsWoche (2014) Anzahl der täglich bei Instagram hochgeladenen Fotos in den Jahren 2011 bis 2014 (in Millionen). Statista. Statista GmbH. https://de.statista.com/statistik/daten/studie/312273/umfrage/taeglich-auf-instagram-hochgeladene-und-geteilte-fotos/ [Last accessed on 29/10/2024]

Young KR, León B (2007) Tree-line changes along the Andes: implications of spatial patterns and dynamics. Philosophical transactions of the Royal Society of London, Series B, Biological sciences 362: 263–272. https://doi.org/10.1098/rstb.2006.1986

Zizka A, Antonelli A, Silvestro D (2021) sampbias, a method for quantifying geographic sampling biases in species distribution data. Ecography 44: 25–32. https://doi.org/10.1111/ecog.05102

Supplementary materials

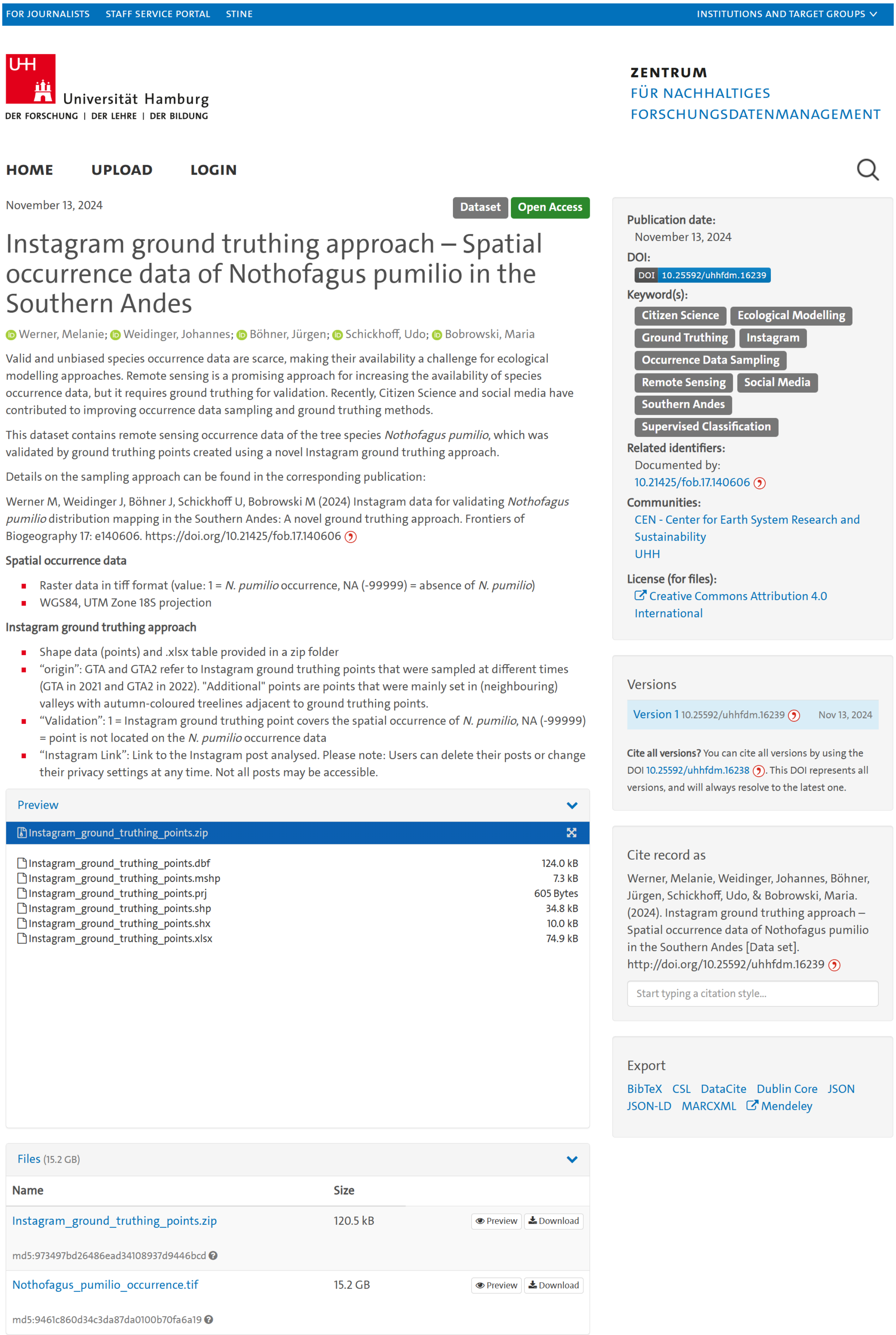
Supplementary material 1

Results of the “sampbias” analysis and information on the performance of the Supervised Classification results testing different classification algorithms (.xlsx)

Link: https://doi.org/10.21425/fob.17.140606.suppl1

Publication II

Werner, M., Weidinger, J., Böhner, J., Schickhoff, U., and Bobrowski, M. (2024). Instagram ground truthing approach – Spatial occurrence data of *Nothofagus pumilio* in the Southern Andes [Data set]. https://doi.org/10.25592/UHHFDM.16239



November 13, 2024

Instagram ground truthing approach – Spatial occurrence data of Nothofagus pumilio in the Southern Andes

Werner, Melanie; Weidinger, Johannes; Böhner, Jürgen; Schickhoff, Udo; Bobrowski, Maria

Valid and unbiased species occurrence data are scarce, making their availability a challenge for ecological modelling approaches. Remote sensing is a promising approach for increasing the availability of species occurrence data, but it requires ground truthing for validation. Recently, Citizen Science and social media have contributed to improving occurrence data sampling and ground truthing methods.

This dataset contains remote sensing occurrence data of the tree species *Nothofagus pumilio*, which was validated by ground truthing points created using a novel Instagram ground truthing approach.

Details on the sampling approach can be found in the corresponding publication:

Werner M, Weidinger J, Böhner J, Schickhoff U, Bobrowski M (2024) Instagram data for validating *Nothofagus pumilio* distribution mapping in the Southern Andes: A novel ground truthing approach. Frontiers of Biogeography 17: e140606. https://doi.org/10.21425/fob.17.140606

Spatial occurrence data

- Raster data in tiff format (value: 1 = *N. pumilio* occurrence, NA (-99999) = absence of *N. pumilio*)
- WGS84, UTM Zone 18S projection

Instagram ground truthing approach

- Shape data (points) and .xlsx table provided in a zip folder
- “origin”: GTA and GTA2 refer to Instagram ground truthing points that were sampled at different times (GTA in 2021 and GTA2 in 2022). "Additional" points are points that were mainly set in (neighbouring) valleys with autumn-coloured treelines adjacent to ground truthing points.
- “Validation”: 1 = Instagram ground truthing point covers the spatial occurrence of *N. pumilio*, NA (-99999) = point is not located on the *N. pumilio* occurrence data
- “Instagram Link”: Link to the Instagram post analysed. Please note: Users can delete their posts or change their privacy settings at any time. Not all posts may be accessible.

Preview

Instagram_ground_truthing_points.zip

Instagram_ground_truthing_points.dbf	124.0 kB
Instagram_ground_truthing_points.mshp	7.3 kB
Instagram_ground_truthing_points.prj	605 Bytes
Instagram_ground_truthing_points.shp	34.8 kB
Instagram_ground_truthing_points.shx	10.0 kB
Instagram_ground_truthing_points.xlsx	74.9 kB

Files (15.2 GB)

Name	Size	
Instagram_ground_truthing_points.zip md5:973497bd26486ead34108937d9446bcd	120.5 kB	Preview Download
Nothofagus_pumilio_occurrence.tif md5:9461c860d34c3da87da0100b70fa6a19	15.2 GB	Preview Download

Publication date:
November 13, 2024

DOI:
DOI 10.25592/uhhfdm.16239

Keyword(s):
Citizen Science, Ecological Modelling, Ground Truthing, Instagram, Occurrence Data Sampling, Remote Sensing, Social Media, Southern Andes, Supervised Classification

Related identifiers:
Documented by:
10.21425/fob.17.140606

Communities:
CEN - Center for Earth System Research and Sustainability
UHH

License (for files):
Creative Commons Attribution 4.0 International

Versions

Version 1 10.25592/uhhfdm.16239 Nov 13, 2024

Cite all versions? You can cite all versions by using the DOI 10.25592/uhhfdm.16238. This DOI represents all versions, and will always resolve to the latest one.

Cite record as

Werner, Melanie, Weidinger, Johannes, Böhner, Jürgen, Schickhoff, Udo, & Bobrowski, Maria. (2024). Instagram ground truthing approach – Spatial occurrence data of Nothofagus pumilio in the Southern Andes [Data set]. http://doi.org/10.25592/uhhfdm.16239

Start typing a citation style...

Export

BibTeX CSL DataCite Dublin Core JSON JSON-LD MARCXML Mendeley

Publication III

Jentsch, H., Weidinger, J., Werner, M., and Bobrowski, M. (2025). ClimDatDownloadR: Accessing Climate Data Repositories for Modelling [Manuscript submitted for publication].

The Journal of Open Source Software

ClimDatDownloadR: Accessing Climate Data Repositories for Modelling

Helge Jentsch[1,2*¶], **Johannes Weidinger**[2*], **Melanie Werner**[2*], and **Maria Bobrowski**[2*]

1 Climate Geography (CliG), Albert-Ludwigs-Universität Freiburg, Freiburg i. Breisgau, Germany **2** Centrum for Earth System Research and Sustainability (CEN), University of Hamburg, Hamburg, Germany ¶ Corresponding author * These authors contributed equally.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

In partnership with

AMERICAN ASTRONOMICAL SOCIETY

This article and software are linked with research article DOI 10.3847/xxxxx <- update this with the DOI from AAS once you know it., published in the Astrophysical Journal <- The name of the AAS journal..

Summary

Systematical accessing, downloading, and pre-processing climatological data from CHELSA (Karger et al., 2017, 2021; Karger et al., 2018) and WorldClim (Fick & Hijmans, 2017; Hijmans et al., 2005) remains a challenge in different environmental disciplines like Species Distribution Modelling (SDM) and climate studies. This package provides a set of functions that allow easy access and customized selection of climate data sets. Besides downloading the raw data, also functionalities to complete pre-processing steps like clipping, rescaling, and file management are available. The applications of the package range from one-time-use to implementing the functions in automatic processing of scientific workflows.

Statement of need

The climatology datasets CHELSA and WorldClim contribute as crucial data bases for studies in various scientific fields. Primarily used in studies with focus on ecology (~4,200 publications[1]), environmental sciences (>2,200 publications), and biodiversity conservation (>1,600 publications), usages extend to a wide variety of scientific disciplines. The main usage of the datasets, however, lies in Species Distribution Modelling (SDM) and Ecological Niche Modelling (ENM). Their free availability and frequent citation in widely referenced papers on SDM and ENM strategies (e.g., Randin et al., 2020; Zurell et al., 2020) have contributed to their widespread adoption, facilitating comparability between modelling studies at different spatial and temporal scales.

The high resolution global climatological datasets (30 arc-sec. ~ 1km) include downscaled and bias-corrected data from 30-year time-periods, providing always monthly mean, minimum, and maximum values of temperature and monthly precipitation sums for analysis[2]. Additionally, 19 bioclimatic parameters are accessible, which enable conclusions about seasonality.

Since their initial releases in 2018 (CHELSA V1.2), the CHELSA (Karger et al., 2017; Karger et al., 2018) datasets were cited in more than 2,800 peer reviewed papers, indexed on the Web of Science (source, Aug. 2025). The latest release of WorldClim 2 in 2017 (Fick & Hijmans, 2017) was cited more than 10,600 times (source, Aug. 2025).

---

[1] Following the Web of Science Categories, citations of Karger et al. (2018) (Data from CHELSA 2.1) had 1,155 citations in the field of Ecology. The WorldClim 2 data (Fick & Hijmans, 2017) has 3,044 citations in the same Web of Science category. Both numbers are of the date 17.05.2025. The "Web of Science Categories are assigned at the journal level", meaning the publishing journal defines the category (source).

[2] Function `Chelsa.timeseries.download` supports also the download of potential evapotranspiration (PET) from CHELSA 2.1 (Karger et al., 2018)

CHELSA and WorldClim datasets are commonly utilized in models predicting the potential past, current, and future distribution of species, particularly in studies on monitoring distribution shifts under climate change (e.g., Bobrowski et al., 2017; Twala et al., 2023; Werner et al., 2025), tracking endangered species and planning conservation strategies (e.g., Franklin, 2013; Muscatello et al., 2021), assessing the spread of invasive species (Srivastava et al., 2019), and management strategies in forestry and agriculture (e.g., Agbezuge & Balakrishnan, 2024; Pecchi et al., 2019).

Recent studies have also assessed the performance of these datasets in SDM/ENM approaches, highlighting their respective strengths and limitations (e.g., Bobrowski, Weidinger, & Schickhoff, 2021; Bobrowski, Weidinger, Schwab, et al., 2021; Bobrowski & Schickhoff, 2017; Datta et al., 2020; Rodríguez-Rey & Jiménez-Valverde, 2024). Given that dataset performance may vary depending on the research scope, it is recommended to test multiple datasets to to ensure their suitability for the research target and region.

For these applications, `ClimDatDownloadR` offers key advantages by enabling efficient retrieval from both dataset providers and pre-processing steps such as partial selection of parameters, months, and bioclimatic parameters, temporal subsets of timeseries, customized extent, and included file management as well as an output of the provider's respective citation file. In addition to time-saving aspects, the storage usage and management played a key role in the development of the `ClimDatDownloadR`.

The implemented data management creates a hierarchical, clear, and reproducible data structure for analyses during the processing. Downloaded data can be kept as is, deleted, or packed in a zip-archive file. All of raised *ease-of-use* add-ons contribute to the primary goal of `ClimDatDownloadR` to enable more scientists and other users or organisations to download and pre-process CHELSA and WorldClim data to gain more experience in geodata handling and applications.

Since the official release in 2023, the use of `ClimDatDownloadR` steadily increased (Bobrowski, Weidinger, & Schickhoff, 2021; Chen et al., 2025; Costa-Saura et al., 2025; Maitner et al., 2023; Santi et al., 2024; Twala et al., 2023; Werner et al., 2025). Further, the need of having software for downloading and pre-processing of freely available data is shown by the steady stream of interested visitors on ResearchGate (3,399 unique visits, 04.08.2025), Zenodo (>1000 views, > 150 downloads) (Jentsch et al., 2023), and citations in peer-reviewed papers.

The package implements the datasets CHELSA V1.2 , V2.1, WorldClim V1.4, and V2.1. More specifically the CHELSA Climatologies, Timeseries, CRU Timeseries (CHELSAcruts), and WorldClim Histclim datasets for present data. For past data, the CHELSA PIMP3 data from CHELSA V1.2 is also available. For future data, both CHELSA and WorldClim provide datasets incorporating various CMIP 5 and 6 global circulation models with various emission scenarios and reference periods. An overview as well as a introduction to the usage of the functions is provided in the Readme of the package on GitHub.

Acknowledgements

We acknowledge the thorough testing efforts by the co-authors and Nadine Kaul, and the open access efforts by the Eidg. Forschungsanstalt für Wald, Schnee und Landschaft WSL [Chelsa; Karger et al. (2017)] and University of California, Davis [WorldClim; Fick & Hijmans (2017)] working groups. Also we want to acknowledge the wide user group that implemented the `ClimDatDownloadR` in their unpublished scientific work or motivated us to implement new datasets or update broken links. In addition we want to thank the developers of the various R-packages for their contributions that made this package possible. In alphabetical order these were: `curl` (Ooms, 2025), `httr` (Wickham, 2023a), `ncdf4` (Pierce, 2024), `RCurl` (Temple Lang, 2025), `RefManageR` (McLean, 2014, 2017), `sf` (E. Pebesma, 2018; E. Pebesma & Bivand, 2023), `stringr` (Wickham, 2023b), `sp` (Bivand et al., 2013; E. J. Pebesma & Bivand, 2005), `terra` (Hijmans, 2025), and the development team of R (R Core Team, 2025).

References

Agbezuge, E. Y., & Balakrishnan, P. (2024). Application of species distribution modelling in agriculture: A review. In A. Swaroop, Z. Polkowski, S. D. Correia, & B. Virdee (Eds.), *Proceedings of data analytics and management* (pp. 173–188). Springer Nature Singapore. ISBN: 978-981-99-6547-2

Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R, Second edition*. Springer, NY. https://asdar-book.org/

Bobrowski, M., Gerlitz, L., & Schickhoff, U. (2017). Modelling the potential distribution of Betula utilis in the Himalaya. *Global Ecology and Conservation*, *11*, 69–83. https://doi.org/10.1016/j.gecco.2017.04.003

Bobrowski, M., & Schickhoff, U. (2017). Why input matters: Selection of climate data sets for modelling the potential distribution of a treeline species in the Himalayan region. *Ecological Modelling*, *359*, 92–102. https://doi.org/10.1016/j.ecolmodel.2017.05.021

Bobrowski, M., Weidinger, J., & Schickhoff, U. (2021). Is new always better? Frontiers in global climate datasets for modeling treeline species in the himalayas. *Atmosphere*, *12*(543). https://doi.org/10.3390/atmos12050543

Bobrowski, M., Weidinger, J., Schwab, N., & Schickhoff, U. (2021). Searching for ecology in species distribution models in the Himalayas. *Ecological Modelling*, *458*, 109693. https://doi.org/10.1016/j.ecolmodel.2021.109693

Chen, S. H., Stevens, L., Gooden, B., Rafter, M. A., Knerr, N., Thrall, P. H., Ord, L., & Schmidt-Lebuhn, A. N. (2025). PhyloControl: A phylogeny visualisation platform for risk analysis in weed biological control. *Biological Control*, 105859. https://doi.org/10.1016/j.biocontrol.2025.105859

Costa-Saura, J. M., Midolo, G., Ricotta, C., Baudena, M., Calfapietra, C., Elia, M., Fiorucci, P., Mereu, S., Sirca, C., Spano, D., Vivaldo, G., & Ottaviani, G. (2025). Are trait responses of tree species across pyroregions indicative of fire-modulated plant functional strategies? *Perspectives in Plant Ecology, Evolution and Systematics*, *67*, 125867. https://doi.org/10.1016/j.ppees.2025.125867

Datta, A., Schweiger, O., & Kühn, I. (2020). Origin of climatic data can determine the transferability of species distribution models. *NeoBiota : Advancing Research on Alien Species and Biological Invasions*, *59*, 61–76. https://doi.org/10.3897/neobiota.59.36299

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Franklin, J. (2013). Species distribution models in conservation biogeography: Developments and challenges. *Diversity and Distributions*, *19*(10), 1217–1223. https://doi.org/10.1111/ddi.12125

Hijmans, R. J. (2025). *Terra: Spatial data analysis* [Manual]. https://CRAN.R-project.org/package=terra

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*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Jentsch, H., Weidinger, J., & Bobrowski, M. (2023). *ClimDatDownloadR: Downloads Climate Data from Chelsa and WorldClim*. Zenodo. https://doi.org/10.5281/ZENODO.7924342

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*(1), 170122. https://doi.org/10.1038/sdata.2017.122

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2018). *Data from: Climatologies at high resolution for the earth's land surface areas*. Dryad. https://doi.org/10.5061/DRYAD.KD1D4

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2021). *Climatologies at high resolution for the earth's land surface areas*. EnviDat. https://doi.org/10.16904/envidat.228

Maitner, B., Gallagher, R., Svenning, J.-C., Tietje, M., Wenk, E. H., & Eiserhardt, W. L. (2023). A global assessment of the Raunkiæran shortfall in plants: Geographic biases in our knowledge of plant traits. *New Phytologist*, *240*(4), 1345–1354. https://doi.org/10.1111/nph.18999

McLean, M. W. (2014). *Straightforward bibliography management in R using the RefManager package* [Manual]. https://arxiv.org/abs/1403.2036

McLean, M. W. (2017). RefManageR: Import and manage BibTeX and BibLaTeX references in R. *The Journal of Open Source Software*. https://doi.org/10.21105/joss.00338

Muscatello, A., Elith, J., & Kujala, H. (2021). How decisions about fitting species distribution models affect conservation outcomes. *Conservation Biology : The Journal of the Society for Conservation Biology*, *35*(4), 1309–1320. https://doi.org/10.1111/cobi.13669

Ooms, J. (2025). *Curl: A modern and flexible web client for R* [Manual]. https://CRAN.R-project.org/package=curl

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *The R Journal*, *10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

Pebesma, E. J., & Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, *5*(2), 9–13. https://CRAN.R-project.org/doc/Rnews/

Pebesma, E., & Bivand, R. (2023). *Spatial data science: With applications in R*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016

Pecchi, M., Marchi, M., Burton, V., Giannetti, F., Moriondo, M., Bernetti, I., Bindi, M., & Chirici, G. (2019). Species distribution modelling to support forest management. A literature review. *Ecological Modelling*, *411*, 108817. https://doi.org/10.1016/j.ecolmodel.2019.108817

Pierce, D. (2024). *Ncdf4: Interface to unidata netCDF (version 4 or earlier) format data files* [Manual]. https://doi.org/10.32614/cran.package.ncdf4

R Core Team. (2025). *R: A language and environment for statistical computing* [Manual]. R Foundation for Statistical Computing. https://doi.org/10.32614/r.manuals

Randin, C. F., Ashcroft, M. B., Bolliger, J., Cavender-Bares, J., Coops, N. C., Dullinger, S., Dirnböck, T., Eckert, S., Ellis, E., Fernández, N., Giuliani, G., Guisan, A., Jetz, W., Joost, S., Karger, D. N., Lembrechts, J., Lenoir, J., Luoto, M., Morin, X., … Payne, D. (2020). Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. *Remote Sensing of Environment*, *239*, 111626. https://doi.org/10.1016/j.rse.2019.111626

Rodríguez-Rey, M., & Jiménez-Valverde, A. (2024). Differing sensitivity of species distribution modelling algorithms to climate data source. *Ecological Informatics*, *79*, 102387. https://doi.org/10.1016/j.ecoinf.2023.102387

Santi, F., Testolin, R., Zannini, P., Di Musciano, M., Micci, V., Ricci, L., Guarino, R., Bacchetta, G., Fernández-Palacios, J. M., Fois, M., Kougioumoutzis, K., Kunt, K. B., Lucchi, F., Médail, F., Nikolić, T., Otto, R., Pasta, S., Panitsa, M., Proios, K., … Chiarucci, A. (2024). MEDIS—A comprehensive spatial database on Mediterranean islands for biogeographical and evolutionary research. *Global Ecology and Biogeography*, *33*(8), e13855. https://doi.org/10.1111/geb.13855

Srivastava, V., Lafond, V., & Griess, V. C. (2019). Species distribution models (SDM): Applications, benefits and challenges in invasive species management. *CABI Reviews*, 1–13. https://doi.org/10.1079/PAVSNNR201914020

Temple Lang, D. (2025). *RCurl: General network (HTTP/FTP/...) client interface for R* [Manual]. https://doi.org/10.32614/CRAN.package.RCurl

Twala, T. C., Fisher, J. T., & Glennon, K. L. (2023). Projecting Podocarpaceae response to climate change: We are not out of the woods yet. *AoB PLANTS*, *15*(4), plad034. https://doi.org/10.1093/aobpla/plad034

Werner, M., Böhner, J., Oldeland, J., Schickhoff, U., Weidinger, J., & Bobrowski, M. (2025). Treeline Species Distribution Under Climate Change: Modelling the Current and Future Range of Nothofagus pumilio in the Southern Andes. *Forests*, *16*(8), 1211. https://doi.org/10.3390/f16081211

Wickham, H. (2023a). *Httr: Tools for working with urls and HTTP* [Manual]. https://doi.org/10.32614/cran.package.httr

Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations* [Manual]. https://doi.org/10.32614/cran.package.stringr

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera–Arroita, G., Guisan, A., Lahoz–Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra–Diaz, J. M., Thuiller, W., … Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, *43*(9), 1261–1277. https://doi.org/10.1111/ecog.04960

Publication IV

Werner, M., Böhner, J., Oldeland, J., Schickhoff, U., Weidinger, J., and Bobrowski, M. (2025). Treeline Species Distribution Under Climate Change: Modelling the Current and Future Range of *Nothofagus pumilio* in the Southern Andes. *Forests*, *16*(8), 1211. https://doi.org/10.3390/f16081211

forests

Article

Treeline Species Distribution Under Climate Change: Modelling the Current and Future Range of *Nothofagus pumilio* in the Southern Andes

Melanie Werner [1,*], Jürgen Böhner [2], Jens Oldeland [3], Udo Schickhoff [1], Johannes Weidinger [1] and Maria Bobrowski [1]

[1] Institute of Geography, Center for Earth System Research and Sustainability (CEN), University of Hamburg, Bundesstraße 55, 20146 Hamburg, Germany; udo.schickhoff@uni-hamburg.de (U.S.); johannes.weidinger@uni-hamburg.de (J.W.); maria.bobrowski@uni-hamburg.de (M.B.)

[2] Institute of Geography, Cluster of Excellence "Climate, Climatic Change, and Society" (CLICCS), Center for Earth System Research and Sustainability (CEN), University of Hamburg, Bundesstraße 55, 20146 Hamburg, Germany; juergen.boehner@uni-hamburg.de

[3] Institute for Globally Distributed Open Research and Education (IGDORE), Burgunderweg 9d, 22453 Hamburg, Germany; jens.oldeland@igdore.org

* Correspondence: melanie.werner@uni-hamburg.de

Abstract

Although treeline ecotones are significant components of vulnerable mountain ecosystems and key indicators of climate change, treelines of the Southern Hemisphere remain largely outside of research focus. In this study, we investigate, for the first time, the current and future distribution of the treeline species *Nothofagus pumilio* in the Southern Andes using a Species Distribution Modelling approach. The lack of modelling studies in this region can be contributed to missing occurrence data for the species. In a preliminary study, both point and raster data were generated using a novel Instagram ground truthing approach and remote sensing. Here we tested the performance of the two datasets: a typical binary species dataset consisting of occurrence points and pseudo-absence points and a continuous dataset where species occurrence was determined by supervised classification. We used a Random Forest (RF) classification and a RF regression approach. RF is applicable for both datasets, has a very good performance, handles multicollinearity and remains largely interpretable. We used bioclimatic variables from CHELSA as predictors. The two models differ in terms of variable importance and spatial prediction. While a temperature variable is the most important variable in the RF classification, the RF regression model was mainly modelled by precipitation variables. Heat deficiency is the most important limiting factor for tree growth at treelines. It is evident, however, that water availability and drought stress will play an increasingly important role for the future competitiveness of treeline species and their distribution. Modelling with binary presence–absence point data in the RF classification model led to an overprediction of the potential distribution of the species in summit regions and in glacier areas, while the RF regression model, trained with continuous raster data, led to a spatial prediction with small-scale details. The time-consuming and costly acquisition of complex species information should be accepted in order to provide better predictions and insights into the potential current and future distribution of a species.

Keywords: climate change; *Nothofagus pumilio*; Random Forest algorithm; Species Distribution Modelling; Southern Andes; treelines

check for updates

Academic Editors: Jesús Julio Camarero and Gerhard Wieser

Received: 23 May 2025
Revised: 18 July 2025
Accepted: 21 July 2025
Published: 23 July 2025

Citation: Werner, M.; Böhner, J.; Oldeland, J.; Schickhoff, U.; Weidinger, J.; Bobrowski, M. Treeline Species Distribution Under Climate Change: Modelling the Current and Future Range of *Nothofagus pumilio* in the Southern Andes. *Forests* **2025**, *16*, 1211. https://doi.org/10.3390/f16081211

1. Introduction

The application of modelling approaches, including Species Distribution Modelling (SDM), has recently rapidly increased in order to generate insights into the sensitivity and shifts in treelines in response to climate change [1]. Warming rates in high mountain regions are, on average, greater than the global mean, resulting in ecosystems being particularly stressed by changing climatic conditions [2]. The natural elevational position of the treeline is defined by heat deficiency [3] and, globally, is approximately aligned with the 6.4 °C isotherm [4]. Consequently, treeline shift in response to warming is frequently investigated and widely recognised as a key indicator of climate change [5]. However, observed responses are rather inconsistent, spanning the entire gradient from static treelines with insignificant responses to dynamic treelines with substantial treeline advance [6–8]. Globally, the proportion of advancing elevational treelines has been increasing from 52% [9] to 66% [10,11]. In the Northern Hemisphere, 90% of treelines are reported to be advancing [12], whereas treelines in the Southern Hemisphere are responding weakly to climatic changes [11]. Certain relationships between treeline form, which can vary between gradual and abrupt, and treeline dynamics have been suggested [13]. Factors such as seedling mortality and dieback processes play critical roles in shaping treeline form and influencing possible shifts [14]. While gradual treelines are more likely to advance to higher elevations, abrupt treelines, as formed by *Nothofagus* in the Southern Hemisphere, are more stable due to increased seedling mortality above closed forest stands or due to anthropogenic disturbances [14].

Treelines and the shift in treelines have extensively been studied in recent decades [1,15]; however, comparatively few studies focused on treelines of the Southern Hemisphere [10,11]. A recent review study examining the impact of climate change on Andean biomes found that those in the southern Andes remain the least studied [16]. For example, to our knowledge, there is no SDM study investigating the entire current and future distribution of important treeline species in the Southern Andes. *Nothofagus pumilio* (Poepp et Endl.) Krasser (southern or lenga beech) is forming an abrupt treeline over approximately 2000 km latitudinal extent from 35° S to 56° S in the Southern Andes. The treeline is naturally abrupt due to seedling dieback outside the protecting tree stands [17], and in some cases the treeline is influenced by anthropogenic disturbances like grazing, forestry or fire, resulting in more diffuse treeline ecotones [14]. The uppermost trees are often in a krummholz growth form [18]. The treeline follows a 6.6 °C isotherm [19], while increases from 1.2 to 4.0 °C in mean annual temperatures and decreases of up to 30% in mean annual precipitation are predicted for high-elevation catchments in the southern Andes [20,21]. It is therefore of interest to study the changes in the Southern Andean treeline as a result of global warming.

Recent research has explored changes at the treeline of *N. pumilio* through small-scale dendrochronological and experimental studies [22–26]. *N. pumilio* is highly sensitive to variations in temperature and precipitation, which are often associated with phase shifts in Antarctic Decadal Oscillation (AAO) also known as the Southern Annular Mode (SAM), the El Niño Southern Oscillation (ENSO), and Pacific Decadal Oscillation (PDO), as highlighted in recent studies [22–24]. Increases in radial growth due to rising temperatures have been observed when precipitation levels are sufficient [22,25,26]. Warm and dry springs lead to an increased tree growth at humid treeline sites but to a decrease in tree growth and an increase in tree mortality due to drought at drier treeline sides. Furthermore, high precipitation in late spring often connected with a prolonged snow cover results in a decrease in tree growth [22]. Accordingly, tree growth is highest at mesic sites, followed by wetter sites, while growth rates at drier sites continue to decline [25]. Suitable climatic conditions, particularly rising mean spring and summer temperatures, also promote the establishment of *N. pumilio* seedlings above the current treeline, on both humid and dry slopes, thereby facilitating treeline advance [23,24]. Additionally, warmer springs can improve seed quantity and quality [27] which further increases the possibility of seedlings becoming established. However, drought or soils with low water capacity are important controlling factors, causing drought stress for seedlings and adult trees [28] and ultimately preventing a treeline advance.

In this study, we model the potential current and future distribution of *N. pumilio* based on the species' suitable bioclimatic conditions, following fundamental concepts of Species Distribution Modelling (SDM). SDM models are typically constructed using binary species occurrence data and environmental variables, often climate data [29]. While global climate datasets, such as the CHELSA bioclimatic variables that we use here [30,31], are readily available, the availability of unbiased species occurrence data is a major challenge for SDM studies. Field studies, which can generate reliable occurrence data, are both time-consuming and costly. Moreover, many study sites, particularly in topographic complex regions like high mountains are inaccessible. When species data are not collected through field studies, they are primarily obtained from open-source databases such as the Global Biodiversity Information Facility (GBIF). Although the quantity [32] and quality [33] of data in databases are increasing, these sources often still contain various forms of bias, as highlighted in recent studies [34,35]. Consequently, using these point occurrences in SDM approaches, without addressing these issues, can lead to inaccurate or misleading model results [36]. A promising approach for the investigation of large study areas, especially in regions with limited accessibility, is the use of remote sensing to survey species occurrences [37–39]. Remote sensing data on a species can provide more complex, continuous data and thus further modelling opportunities. However, there is still a need for ground truthing to verify that the species of interest are actually present in the remotely sensed area. In a previous study, we developed an Instagram ground truthing approach, that created less-biased occurrence points, that were subsequently used to validate remote sensing occurrence data of *N. pumilio*, resulting in two valid occurrence datasets [40].

Here, we present an initial holistic approach to model the potential distribution of *N. pumilio* based on two input datasets: a binary point dataset and a continuous raster occurrence dataset derived from supervised classification. While we adopted a standard modelling approach, we also test an innovative technique incorporating continuous raster data. We hypothesise that this approach will yield more detailed insights into the species' potential current and future distribution due to the increased complexity of input data.

To investigate the effect of different species input datasets, our aims are (1) to model the current distribution of *N. pumilio* under prevailing climate conditions, and (2) assess potential range shifts under climate change conditions, and (3) evaluate model performance and model complexity with regard to ecological site factors.

2. Materials and Methods

2.1. Study Species and Study Area

N. pumilio is the most orophilous and widespread species of the *Nothofagus* genus on the South American continent, extending from 35° S to the southernmost tip of Tierra del Fuego (see Figure 1). As an indicator of the orotemperate belt, it forms mono-species forests at the treeline [41]. The species is sometimes accompanied by the morphologically and ecologically similar species *N. antarctica*, with which it can form mixed stands. Hybrids between the two species are also known [42]. The evergreen *N. betuloides* dominates in the lowlands and especially in the (hyper-) humid west.

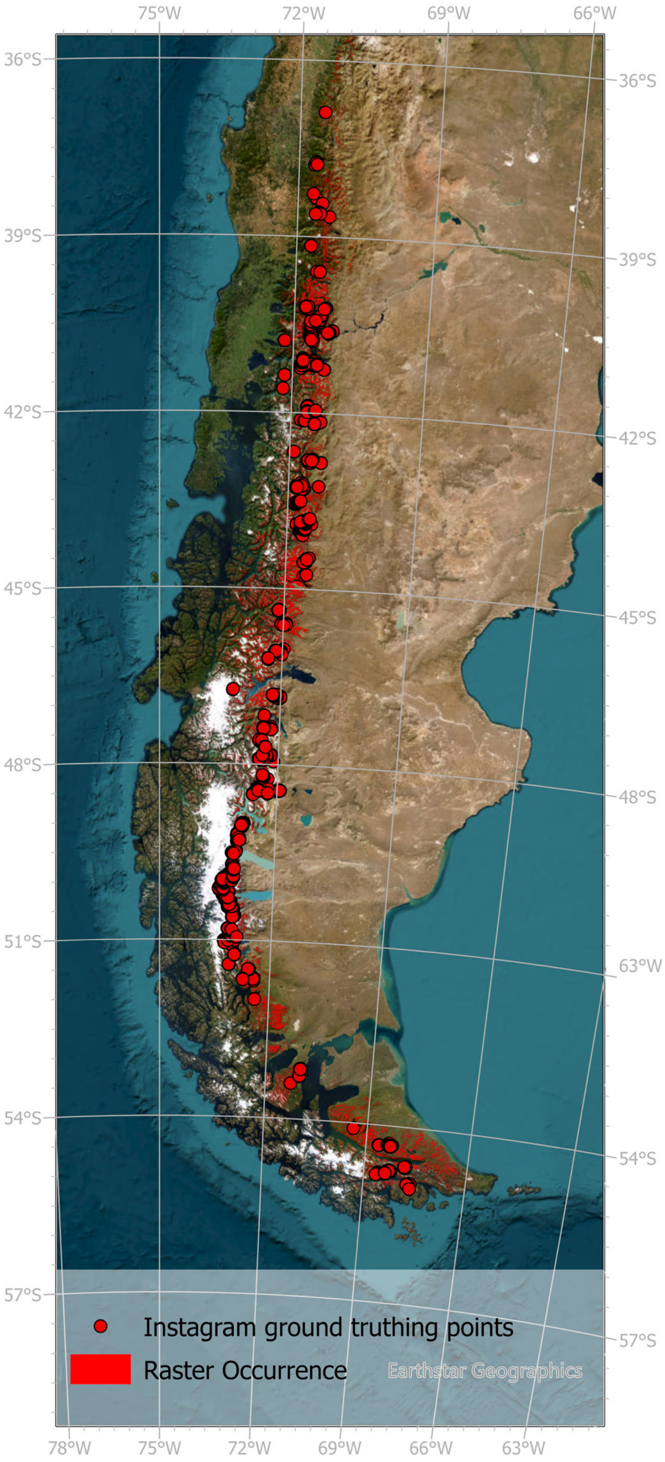


Figure 1. Point and raster occurrence data of *Nothofagus pumilio* created by a novel Instagram ground truthing approach [40].

The study area is characterised by two extreme gradients. One is the temperature gradient, which results from the elevation of the southern Andean Cordillera (up to 3000 m), and the other is the precipitation gradient, which is considered to be the most extreme precipitation gradient on Earth. While precipitation extremes of up to 10,000 mm/year occur on the windward side, west of the Andes, there is a precipitation decrease to less than 300 mm/year on the leeward side, east of the Andes [43]. Northern Patagonia has been substantially affected by the effects of climate change. Mean annual temperatures have risen by up to 1 °C since 1950. While there are no negative trends in precipitation in most areas, precipitation totals are decreasing by 5% in Northern Patagonia. Climate models predict a decrease in precipitation of up to 30% and an increase in temperature of between 1.5 and 3 °C [21].

2.2. Input Data

2.2.1. Species Data

The two species occurrence datasets used were generated in a previous study using a novel Instagram ground truthing approach (IGTA, [40]) designed to reduce sampling and spatial bias often present in existing databases [34,44]. The IGTA aimed at reducing this bias by using a very public and worldwide used social media platform (https://www.instagram.com/) and by including remote sensing data. The study species and area are particularly well-suited for an Instagram-based analysis, as deciduous *N. pumilio* forms mono-species forests at the treeline, making it an especially attractive photo motif in autumn and occurs in a highly touristic region, where tourists and hikers frequently take and share photos on social media. We compiled 1238 occurrence points extending from 36.88° S to 55.03° S by searching for suitable posts uploaded between 2017 and 2022 with the species and the photos' location clearly identifiable, as well as a strict catalogue of criteria. Spatial bias commonly present in datasets derived from citizen science or social media, typically due to sampling near urban centres or in easily accessible areas [45,46], was reduced through the IGTA. This reduction in bias is primarily due to the high number of posts, which included not only intentional but also incidental records of the species. Furthermore, owing to the ecology and phenology of *N. pumilio*, occurrence points were identifiable not only at the exact location where the photo was taken, but also in the background, where the abrupt treeline and the species' autumn colouring were visible. Although some spatial bias remains, since posts are still limited to human-accessible areas, analysis using the R package "sampbias" (version 2.0.0) [47] indicates that the bias was effectively reduced in comparison to 558 points (after filtering for missing coordinates and a coordinate uncertainty of 1 km) from often used open-source database Global Biodiversity Information Facility (GBIF). In addition, the manual georeferencing of occurrence points further reduced coordinate uncertainty. To minimise spatial bias, remote sensing data generated using Sentinel-2 level 2A data and supervised classification were created and subsequently validated using the IGTA occurrence points as ground truth. Training areas were defined and trained using all relevant spectral bands (bands 2 to 7, 8a, 11, and 12) at a spatial resolution of 20 m. Only vegetation raster cells were included in the supervised classification, which distinguished between three classes (deciduous vegetation/*N. pumilio*, evergreen vegetation, and low vegetation/grassland). An altitude correction was applied to ensure that only deciduous vegetation in high elevation was classified as *N. pumilio*. With this approach, two datasets were created: a point occurrence dataset and a spatial raster dataset of the species *N. pumilio* (Figure 1).

For the IGTA point dataset, the first step in modelling was to ensure that only one point was set in a raster cell at the target resolution of 30 arc sec, ~1 km (raster cell size of the climate data). After deleting duplicate points, 999 points remained in the model as "presence" data. For "absence" data, 2000 pseudo-absence points (PA) were generated. The number and the location of PA points have great influence on the model output [48]. We initially tested a PA ratio of 1:1 as suggested for RF models in relevant literature [49]. However, using only 1000 PA points across a large study area resulted in substantial overprediction by the RF classification model, a known issue [49]. Consequently, we adopted a ratio of 1:2 (2000 PA points). The process of PA creation involved first constructing an alpha hull around the presence points and applying a 1 km buffer, within which PA points were randomly generated. To prevent the overwriting of occurrence cells during this process, a 5 km buffer was placed around the presence points. For the final modelling dataset, values from the climate dataset were extracted for both presence and absence points. The two datasets were then merged and supplemented with a binary indicator variable specifying whether *N. pumilio* was present (presence = 1) or absent (absence = 0).

The second dataset was derived from a supervised classification (Kappa values: summer scenes classification 0.89, autumn scenes classification 0.96), in which occurrences of deciduous forest at the treeline were classified to represent the distribution of *N. pumilio* [40]. Gaps in the dataset resulted from missing Sentinel-2 data or areas affected by shadows and cloud cover. The original dataset has a spatial resolution of 20 m. For modelling purposes, *N. pumilio* cover values were aggregated to the target resolution (~1 km), resulting in percentage values ranging from close to 0% up to a maximum of 99.96%. The raster data covers a latitudinal range from 33.49° S to 56.27° S.

2.2.2. Bioclimatic Predictors

Global climatological datasets such as CHELSA [30,31] and Worldclim [50,51] are standards for large-scale SDM studies. Due to their free accessibility, the datasets are widely used and cited, allowing for some comparability of modelling studies. Since other studies suggest that the CHELSA dataset performs better in topographically highly complex areas such as high mountains [52,53], we decided to use this dataset. We utilised the 19 Bioclim variables from version 2.1 with a 30 arc sec (~1 km) resolution [30,31]. The dataset includes temperature and precipitation variables calculated on a daily, monthly, or annual basis, averaged from climate records for the period 1981 to 2010. The “ClimDatDownloadR” R package was used to download and pre-process the data (version 0.1.7.6) [54,55]. As the Bioclim variables were highly multicollinear, we decided to use a subset of the data. To enhance ecological interpretability, we included only variables derived for quarters. This allows for a better ecological understanding of the bioclimatic conditions at the treeline than variables for individual months or annual averages, as conclusions can be drawn about seasonality [56]. At high elevations, the growth and survival of treeline species are primarily determined by conditions during the short growing season [1]. Quarterly variables can isolate this critical period, whereas annual means combine summer and winter extremes, potentially obscuring the actual limiting factors. To further mitigate multicollinearity and exclude irrelevant variables, we applied the “VSURF” R package (version 1.2.0), which follows a two-step procedure. First, it identifies variables relevant for interpretation, and subsequently, it eliminates redundant variables for prediction [57]. Through this approach, two additional variables were removed from the initial subset. The variables used in the model are listed and described in Table 1.

For future prediction, CHELSA version 2.1 provides selected CMIP6 scenarios of the bioclimatic variables. Future Bioclim variables were created using representative concentration pathway (RCP) scenarios, that represent a range of potential greenhouse gas emission pathways, from a low-emission (RCP2.6) to a high-emission development (RCP8.5). We used the SSP126 (RCP2.6), SSP370 (RCP7), and SSP585 (RCP8.5) scenarios for the years 2041 to 2070 and 2071 to 2100 from the MPIESM1-2HR model [31].

Table 1. CHELSA Bioclim variables calculated for quarters and seasonality variables used in the analysis (X). The variables bio 9 and bio 16 have been excluded by the VSURF analysis.

Short Name	Long Name	Used in Analysis
bio 4	temperature seasonality [°C/100] [1]	X
bio 8	mean daily mean air temperatures of the wettest quarter [°C]	X
bio 9	mean daily mean air temperatures of the driest quarter [°C]	excluded by VSURF
bio 10	mean daily mean air temperatures of the warmest quarter [°C]	X
bio 11	mean daily mean air temperatures of the coldest quarter [°C]	X
bio 15	precipitation seasonality [kg m^{-2}] [2]	X
bio 16	mean monthly precipitation amount of the wettest quarter [kg m^{-2} $month^{-1}$]	excluded by VSURF
bio 17	mean monthly precipitation amount of the driest quarter [kg m^{-2} $month^{-1}$]	X
bio 18	mean monthly precipitation amount of the warmest quarter [kg m^{-2} $month^{-1}$]	X
bio 19	mean monthly precipitation amount of the coldest quarter [kg m^{-2} $month^{-1}$]	X

[1] standard deviation of the monthly mean temperatures; [2] the coefficient of variation is the standard deviation of the monthly precipitation estimates expressed as a percentage of the mean of those estimates.

2.3. Model Approach and Model Algorithm

We follow the basic concepts of Species Distribution Modelling (SDM) to analyse the relationship between the species' current distribution and suitable abiotic conditions, as well as its potential future distribution under climate change scenarios [58,59].

Several algorithms have been established for SDMs. In addition to linear regression approaches such as Generalised Linear Models (GLMs), Generalised Additive Models (GAMs), and Multivariate Adaptive Regression Splines (MARS), machine learning algorithms such as Random Forest (RF), Maximum Entropy (MaxEnt), and Artificial Neural Networks (ANN) are established methods [60]. We chose RF because it can be used for classifying binary data (Random Forest Classification), as well as for regression approaches with continuous data (Random Forest Regression) [61]. Although other algorithms would also be suitable for modelling the point occurrences, using the same algorithm for both datasets allows for direct comparison of the approach, subsequent analyses, and consistent interpretation of the results. Furthermore, machine learning approaches such as RF not only demonstrate strong predictive power and model performance but are also unaffected by multicollinearity, making them particularly well-suited for climatic datasets with many highly correlated variables [62]. However, some authors criticise that the interpretability of the models decreases as machine learning methods are "black boxes" compared to simpler linear approaches (e.g., GLMs) [56,60,62]. Random Forest combines both strengths: it is a robust and powerful approach that remains largely interpretable. In the following, we analysed two models: (1) RF classification with the point dataset and (2) RF Regression with continuous data from the raster dataset. We apply k-fold spatial cross-validation to identify the optimal model, using appropriate validation metrics for each modelling approach. Model outcomes are interpreted using variable importance measures, partial dependence plots, and SHAP (Shapley Additive Explanations) analysis. The results of the models are subsequently compared visually.

Data processing and modelling were conducted in R (version 4.4.1 [63]). Maps were created using SAGA GIS (version 9.3.2 [64]) and ArcGIS Pro (version 2.7.0 [65]).

2.4. Model Calibration and Evaluation

Spatial data, such as species and climate data, are often spatially autocorrelated [66,67]. Consequently, when spatial dependence is present in a dataset, spatial (or block) cross-validation is recommended [68]. We performed spatial cross-validation using the R package "blockCV" (version 3.1.5) [69]. For 5-fold cross-validation, the data is initially partitioned into spatial blocks of a predefined size, and each block is randomly assigned to one of five folds (k = 5). The model is then trained on four folds (k - 1) and evaluated on the remaining fold. This procedure is repeated five times, ensuring that each fold serves as both a training and a testing set. To determine an appropriate block size, we initially used the function "cv_spatial_autocor" to calculate the spatial autocorrelation of our species data. Spatial autocorrelation differed slightly between the two datasets (29.6 km for the point dataset and 24.4 km for the raster dataset). Therefore, we opted for a larger block size of 50 km to minimise potential autocorrelation effects and to ensure comparability between the two datasets. The spatial arrangement of the hexagonal blocks is displayed in the Appendix A (Figure A1). To validate the models, we used threshold-independent metrics such as AUC and overall accuracy, as well as threshold-dependent metrics like the True Skill Statistic (TSS), using the maximum sensitivity plus specificity threshold, for the RF classification approach. For the RF regression approach, the root mean square error (RMSE) and the coefficient of determination (R^2) were used as validation metrics. In addition to evaluating model performance on the cross-validation splits, we also assessed the models' hyperparameters. We evaluated models with different numbers of variables considered at each split (mtry: 2, 3, 4) and varying numbers of trees (ntree: 100, 300, 500). Through hyperparameter tuning in combination with spatial cross-validation, we were able to identify the optimal model while minimising the risk of spatial overfitting. The model with the highest average AUC resp. R^2 was selected and subsequently used to predict the potential current and future distribution of *N. pumilio* across the entire dataset.

3. Results

3.1. Current Distribution Range of *N. pumilio*

The optimal model for predicting the current distribution of *N. pumilio* was identified using 5-fold spatial cross-validation. The cross-validation results for both models are presented in Tables A1 and A2 in the Appendix A. Model quality was assessed based on the mean validation metrics across all five folds (for the RF classification model, the optimal model was selected based on the highest AUC; for the RF regression model, selection was based on the highest R^2). For the RF classification model, the best-performing model, with hyperparameters mtry = 2 and ntree = 500, achieved the highest mean AUC of 0.9279 (±0.0257, 95% confidence interval (CI): 0.8960–0.9599), an overall accuracy of 0.8466 (±0.0537, 95% CI: 0.7799–0.9132), and a TSS of 0.6148 (±0.1582, 95% CI: 0.4183–0.8112). The final model was subsequently trained using these hyperparameters on the entire dataset. For the RF regression model, the highest mean R^2 determined by spatial cross-validation was 0.3933 (±0.0409, 95% CI: 0.3425–0.4441), also indicating mtry = 2 and ntree = 500 as the optimal hyperparameters. The models trained with the optimal hyperparameters and on the complete datasets were then used for variable importance analysis as well as for spatial predictions of current and future distributions.

Bioclimatic variables bio 8 (mean daily mean air temperature of the wettest quarter) and bio 19 (mean monthly precipitation amount of the coldest quarter) emerged as the most influential predictors for model accuracy of the RF classification model (Accuracy Importance ranking, from most to least important bioclimatic variable: bio 8 (mean daily mean air temperatures of the wettest quarter), bio 19 (mean monthly precipitation amount of the coldest quarter), bio 15 (precipitation seasonality), bio 17 (mean monthly precipitation amount of the driest quarter), bio 4 (temperature seasonality), bio 10 (mean daily mean air temperatures of the warmest quarter), bio 18 (mean monthly precipitation amount of the warmest quarter), and bio 11 (mean daily mean air temperatures of the coldest quarter)). Additionally, Gini Importance was calculated to assess the most frequently used variables for decision at nodes, with bio 11 (mean daily mean air temperatures of the coldest quarter) and bio 8 (mean daily mean air temperatures of the wettest quarter) emerging as the primary split criteria for the RF classification model (Gini Importance ranking: bio 11, bio 17, bio 8, bio 10, bio 18, bio 19, bio 4, bio 15). While a temperature variable was the most important predictor in the RF classification model, precipitation-related variables, particularly bio 15 (precipitation seasonality) and bio 19 (mean monthly precipitation amount of the coldest quarter), played a dominant role in the RF regression model (Accuracy Importance ranking: bio 15, bio 19, bio 4, bio 11, bio 10, bio 8, bio 18, bio 17). Gini Importance analysis for the RF regression model indicated that bio 11 (mean daily mean air temperatures of the coldest quarter) was the most critical variable for splits, followed by bio 15 (precipitation seasonality) (Gini Importance ranking: bio 11, bio 15, bio 8, bio 18, bio 4, bio 10, bio 19, bio 17).

A RF model consists of multiple individual decision trees, making it challenging to interpret the specific thresholds used at each node. However, the extraction of partial dependence plots (PDP) for individual variables allows for an interpretation of the influence of specific bioclimatic predictors within the RF model. We employed the “partialPlot” function from the “randomForest” R package (version 4.7-1.2, [70]) to assess the influence of the most important variable in each model. The resulting PDPs are presented in Figure 2a,c. For the RF classification model, the x-axis displays the values of the bioclimatic variable bio 8 (mean daily mean air temperatures of the wettest quarter), while the y-axis represents the predicted probability for class 1 (i.e., presence of *N. pumilio*). At low temperatures in the wettest quarter (below −5 °C), the predicted probability of occurrence is high, clearly decreasing towards 5 °C and remaining consistently low above this threshold. This indicates that the model predicts the presence of *N. pumilio* primarily in colder environments during the wettest quarter. In the RF regression model, the x-axis shows the values of bio 15 (precipitation seasonality), and the y-axis represents the predicted cover values of *N. pumilio*. The plot reveals greater variability, but cover values are relatively high in areas with moderate precipitation seasonality (0% to 30%). However, there is a marked decline in predicted cover values in regions with high precipitation seasonality (70% to 100%). Two key assumptions emerge: cover values are greater in areas with lower precipitation seasonality, suggesting that *N. pumilio* prefers regions with more stable precipitation patterns and the species is less likely to occur in areas with highly variable precipitation, e.g., with phases of drought.

These effects are also evident in the SHAP summary (bee swarm) plots, created using the “fastshap” R package (version 0.1.1) [71]. SHAP (Shapley Additive Explanation) analysis, which originates from cooperative game theory, provides a comprehensive overview of the contribution of each predictor to the model outcome (see Figure 2b,d) [72]. In the RF classification model, the influence of temperature variables is consistent: high temperatures have a negative effect on the model (indicating absence), while low temperatures have a positive effect (indicating presence). In the RF regression model, the influence of temperature variables varies. For bio 8 (mean daily mean air temperatures of the wettest quarter) and bio 11 (mean daily mean air temperatures of the coldest quarter), low temperatures also have a positive influence on the model outcome (higher cover values), whereas this effect is reversed for bio 10. For bio 10 (mean temperature of the warmest quarter), low temperatures result in a decrease in cover values, while intermediate temperatures lead to an increase. Bio 11, particularly in the RF regression model, exhibits a wide range, indicating a strong influence on both models. This is further supported by the variable importance analysis, which identifies bio 11 as the most important splitting criterion at nodes based on Gini importance. The pattern for precipitation variables is less distinct;

however, the trend is similar in both models: low to medium precipitation totals have a positive effect on the model. A similar trend can also be observed for seasonality variables bio 4 (temperature seasonality) and bio 15 (precipitation seasonality). Low to moderate values of these variables are associated with species occurrence, whereas very low or very high values result in a reduced probability of presence and lower cover values.

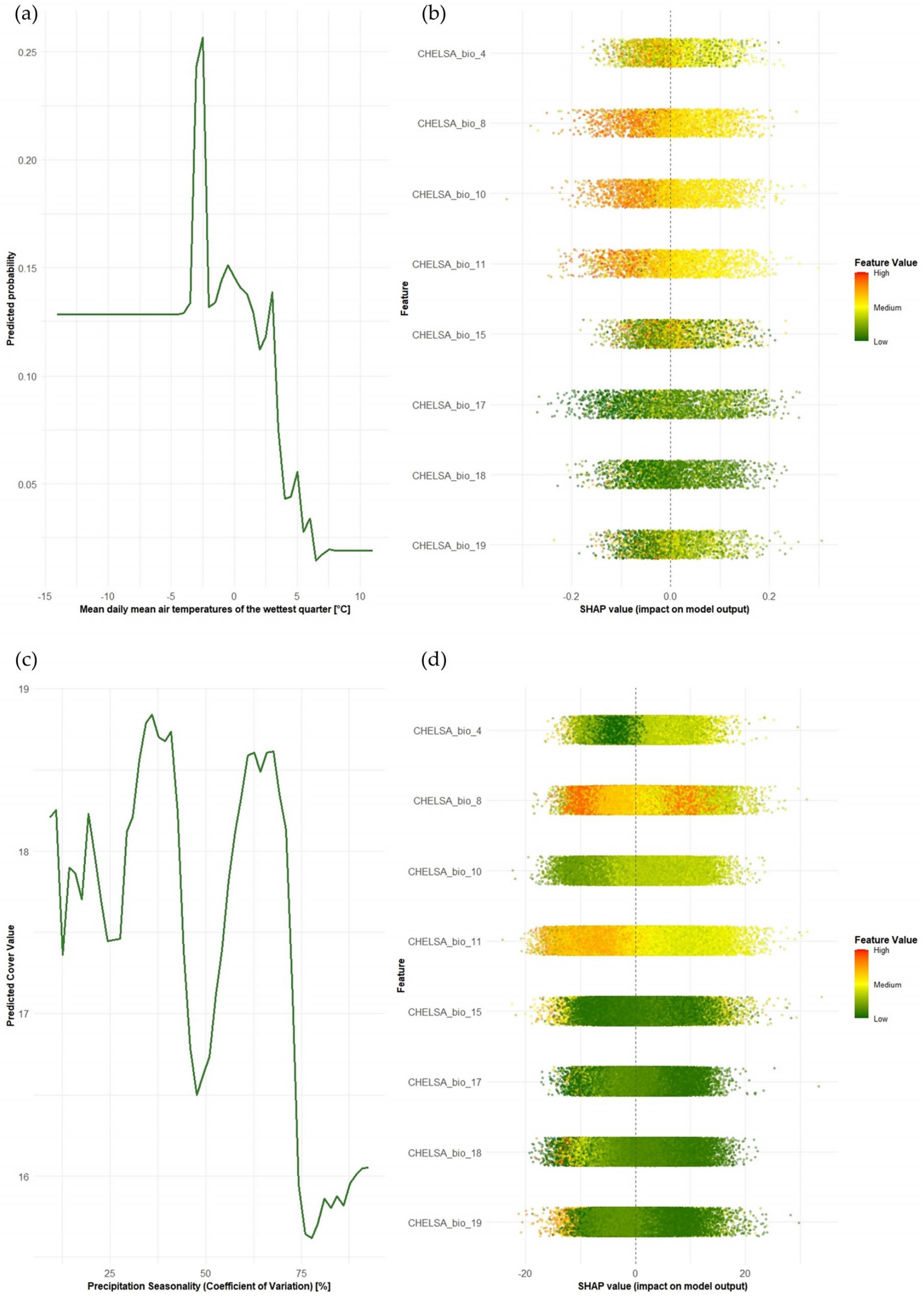


Figure 2. Partial Dependence Plot (PDP) of (**a**) bio 8 (mean daily mean air temperatures of the wettest quarter) for the Random Forest Classification model and (**c**) bio 15 (precipitation seasonality) for Random Forest regression model indicating the influence of the most important variables. SHAP summary plots (**b**,**d**) indicate the contribution of each predictor/feature to the model outcome. Feature values were normalised (breaks: 0, 0.5, 1) due to different units of temperature and precipitation variables.

The spatial predictions of the two models revealed a more compact distribution in the RF classification model, with a slightly broader extent towards the west and east for the RF regression model. The binary RF classification model predicted a northern extent of up to 36.35° S and a southern extent of 55.45° S, whereas the RF regression model extended from 35.24° S to 55.24° S (Figure 3). In the southern regions, including Tierra del Fuego, the RF regression model depicted a more homogeneous distribution. IGTA points were lacking in this area resulting in gaps in the predicted occurrence. While the RF classification model tended to overpredict in unvegetated summit regions and glaciated areas, the RF regression model provided a more fine-grained representation, capturing vegetated valleys more accurately. When compared with a digital surface model (DSM, ALOS Global Digital Surface Model, 30 m), it was evident that the RF classification model predicted suitable climatic conditions in higher (unvegetated) areas as distribution areas, while in the RF regression model they were omitted. The described small-scale differences between the models are illustrated in Figure 4.

To further validate the spatial predictions of the models, we conducted two additional comparisons. First, we performed a visual comparison of the model outputs with independent data provided by the Argentinian forestry authority (Secretaría de Ambiente y Desarrollo Sustentable de la Nación, [73]). Second, we compared the elevation of high-altitude raster cells from both models with reported treeline elevations from 48 plots across 13 locations, as published by Lara et al. (2005) [74]. Figure 5 presents the visual comparison between the raster outputs of the models and the *N. pumilio* forest polygons. The RF classification model shows slightly more deviation and tends to extend beyond the polygon boundaries, while the RF regression model closely matches the reference polygons. Table 2 compares the treeline elevations from on-site measurements [74] with the elevations of the highest raster cells predicted by the models (based on a DSM resampled to 1 km), which are intended to approximate the treeline position. To identify these treeline raster cells, a threshold is required. Based on the range of definitions compiled in the literature, the treeline can be defined as the elevation at which tree canopy cover declines to approximately 30%, representing the uppermost margin of closed forest [18]. Accordingly, we applied a threshold of 30% (resp. 0.3 for RF classification) to the model outputs. In the RF classification model, treeline elevations are, in some cases, significantly higher than the treeline elevations measured on site. However, in the southern regions, the predicted treeline aligns well with the literature. In contrast, the RF regression model predicts treeline elevations that are only slightly higher in the north and overall correspond closely to the reported values.

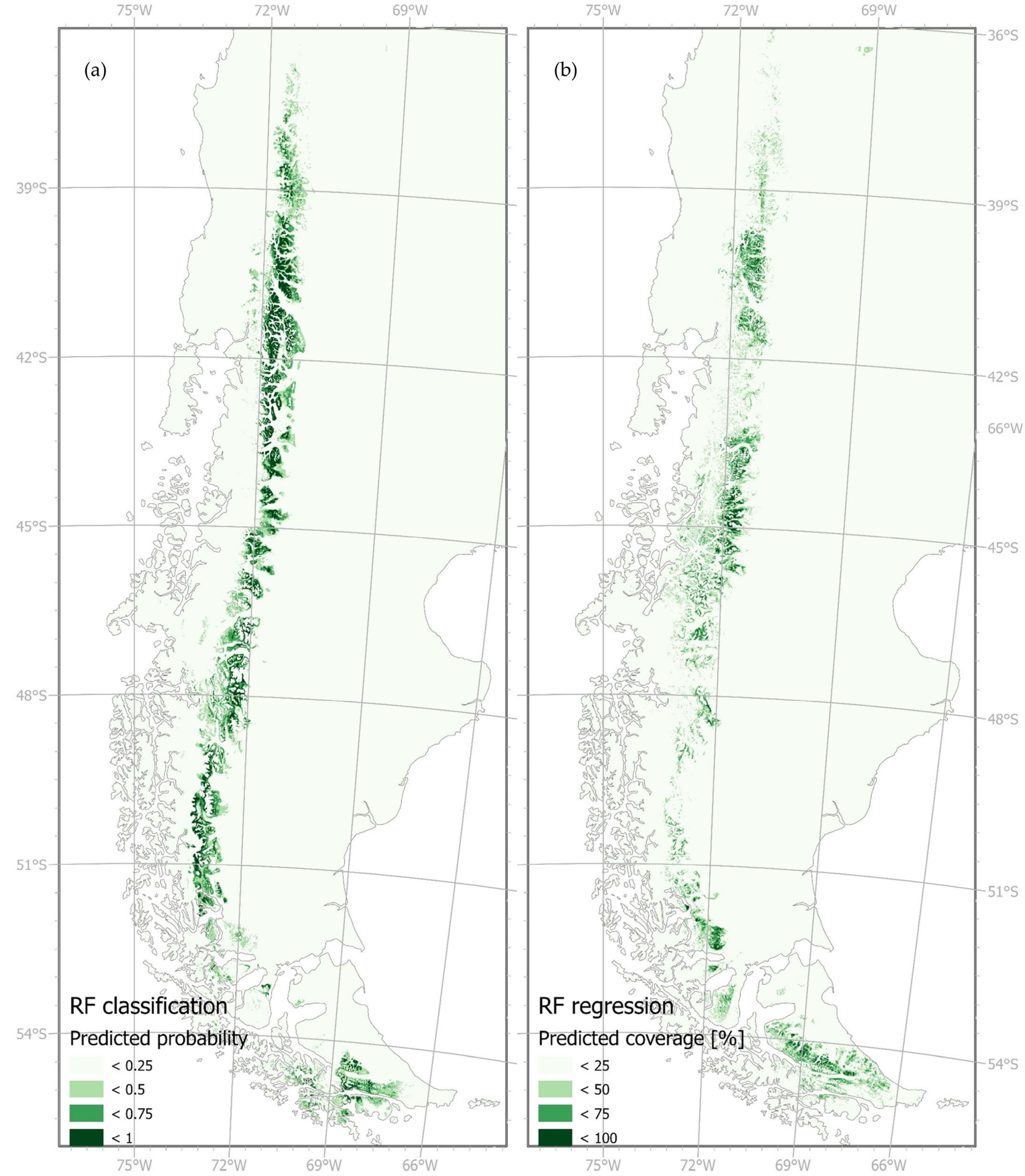


Figure 3. Predicted probability of *Nothofagus pumilio* occurrence of (**a**) the Random Forest classification model and (**b**) predicted cover values of *N. pumilio* of the RF regression model.

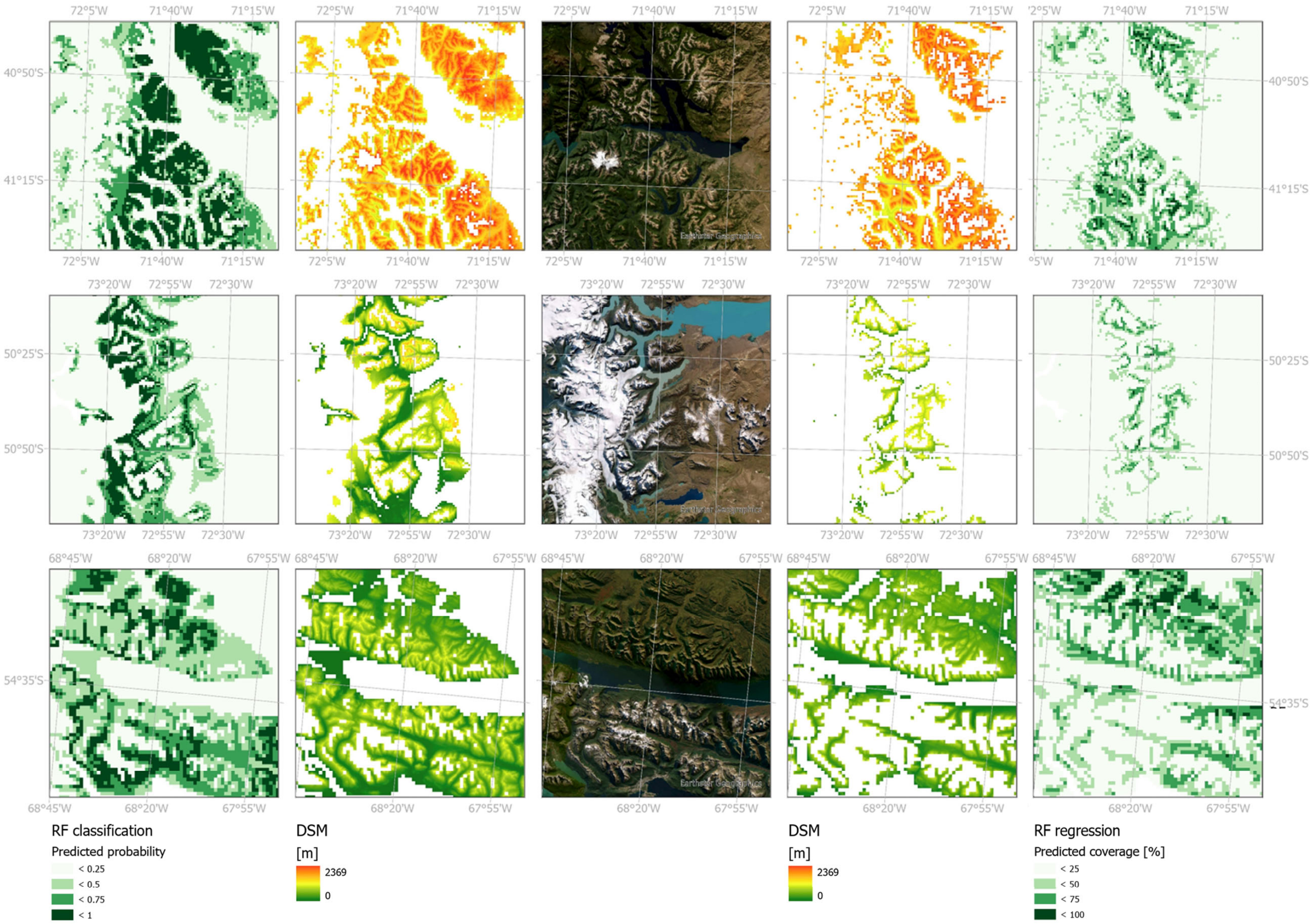


Figure 4. Small-scale details of the Random Forest classification model and Random Forest regression model compared with a satellite basemap (centre) and a Digital Surface Model (DSM), which shows the elevation of the raster cells covered by the model results.

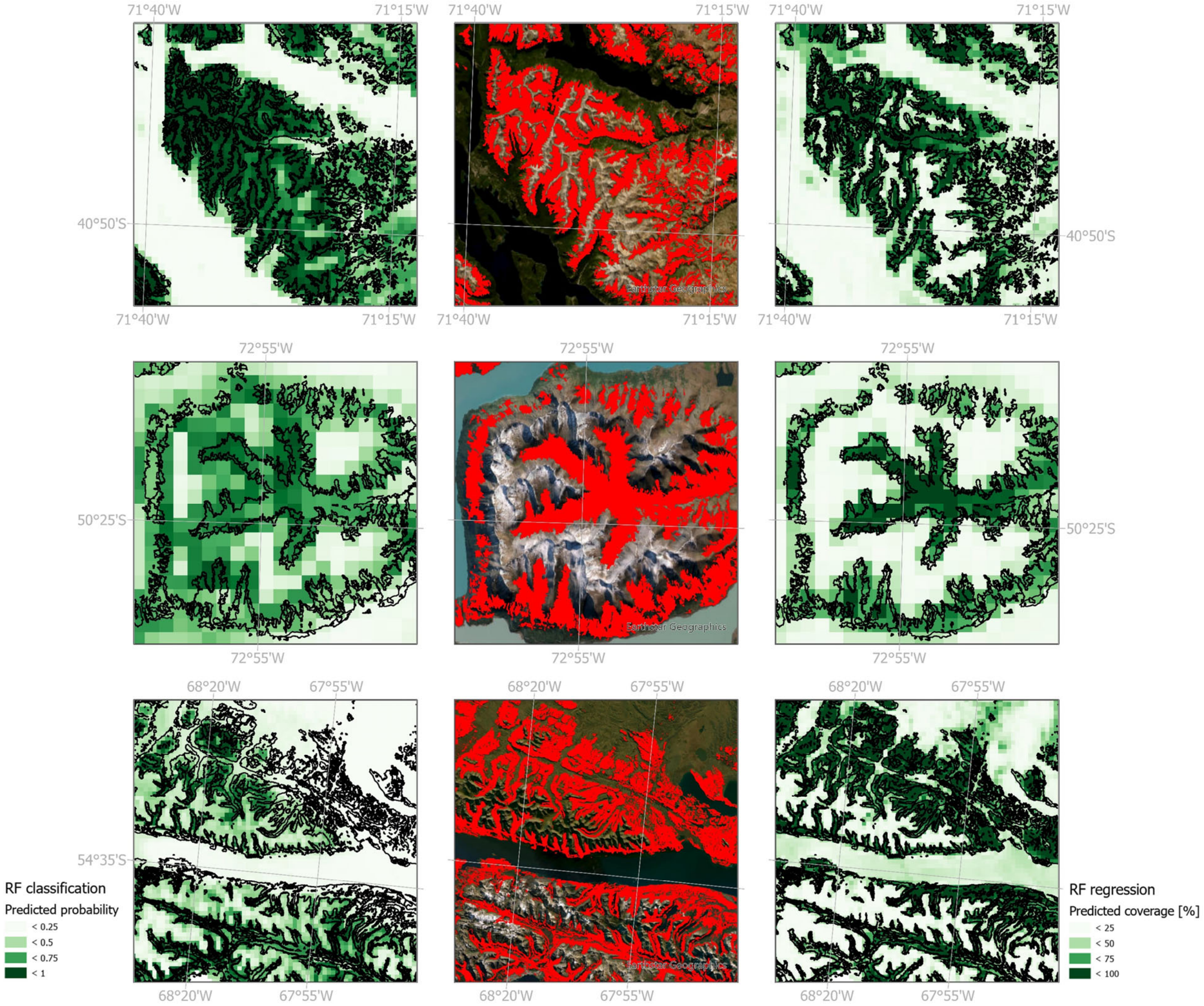


Figure 5. Comparison of the model results with *Nothofagus pumilio* forest polygons (red, centre) from Argentinian forestry authority (Secretaría de Ambiente y Desarrollo Sustentable de la Nación, [73]).

Table 2. Treeline positions based on field data from 48 plots at 13 locations [74], compared with the elevation of the highest raster cell from the model outputs (Random Forest classification and RF regression) in the adjacent mountain range corresponding to each plot location. NA = not available (no data recorded).

Treeline Position and Elevation [m] After Lara et al., 2005 [74]				Treeline Elevation [m] Current Climate	
ID	**X**	**Y**	**Elevation Range**	**RF Class.**	**RF Reg.**
1	−71.00	−35.36	1530	NA	NA
2	−71.11	−37.27	1500–1720	1988	1949
3	−71.33	−38.42	1490–1650	1854	1789
4	−72.15	−40.42	1000–1300	1591	1437
5	−72.19	−41.48	1300	1500	1201
6	71.45	−43.07	1230–1350	1839	1440
7	−71.42	−44.39	1000–1200	1320	1216
8	−72.24	−47.12	800–1180	1361	1197
9	−72.30	−48.30	1200	1522	1074
10	−72.54	−50.57	650–980	1176	956
11	−71.00	−53.00	350–600	543	560
12	−68.45	−54.17	200–600	544	520
13	−67.30	−54.57	300–600	610	610

3.2. Future Distribution Range of N. pumilio

Using CMIP6 data provided by CHELSA for the SSP scenarios, we predicted the potential future distribution of *N. pumilio* with both models. Predictions were generated for the SSP126, SSP370, and SSP585 scenarios for two future periods: 2041–2070 and 2071–2100. Figure 6a,b illustrate the potential distribution of *N. pumilio* for 2041–2070, while Figure A2a,b in the Appendix A depict projections for 2071–2100. For the RF classification model, the SSP scenarios for the period 2041–2070 already indicate a potential shift in distribution towards currently unvegetated summit areas. This upward shift to higher elevations becomes more pronounced with increasingly severe scenarios (SSP370, SSP585). Simultaneously, there is a progressive decline in occurrences throughout the northern part of the species' range, and a slight decrease in the southernmost distribution areas. Additionally, the occurrences are predicted to shift towards the more humid western regions. These trends become even more marked in the period from 2071 to 2100. During this later time span, a decrease in occurrence area is evident, particularly at lower elevations, resulting in an overall stronger decline in the north. Both models consistently indicate a reduction in occurrences in northern regions. However, in contrast to the RF classification model, the RF regression model suggests that *N. pumilio* is more likely to persist at higher elevations in the north, and that the decline in occurrences in the southernmost parts of the range is less pronounced. The total decreases in distribution area and the westward shift are less marked in the RF regression model, with eastern occurrences more likely to remain stable. Nevertheless, the trend of occurrences shifting towards higher elevations is observed in both models, and this trend intensifies with the severity of the scenario and over time. While the RF classification model predicts an expansion into higher, currently snow- and ice-covered areas of the Southern Patagonian Icefield, the RF regression model forecasts a decline in occurrences in this region. Despite this, under the scenarios for 2071–2100, the RF regression model also projects a reduction in occurrences at lower elevations, which ultimately results in a net decrease in the total distribution area of the species. Thus, both modelling approaches reveal potential changes in the distribution of *N. pumilio* under future climate scenarios, particularly in lowland and northern areas, but they differ slightly with respect to the potential persistence of the species at higher elevations in the north and in the southern part of its range.

To assess treeline shifts in numbers, we again refer to the comparison of treeline elevations at the 13 locations. In Table 3, we compare the elevation of the highest raster cells at the treeline sites under current climatic conditions and under the SSP scenarios for the period 2041–2070 (for the time span 2071–2100, please refer to Table A3 in the Appendix A). Even under current climatic conditions, no occurrence was recorded at the northernmost site (site no. 1), and this remains the case across all scenarios. The predicted decline in the northern distribution range continues, with no remaining occurrences at the second site under the SSP370 scenario. Overall, the estimated treeline elevations in the RF classification model are generally higher than those predicted by the RF regression model. As climate scenarios progress, a general upward shift in treeline elevation is observed, with a few exceptions where treeline elevation either stagnates, mainly due to already having reached the highest local topography, or shows a slight decrease. These trends are also evident in the period 2071–2100, with even higher treeline elevations in most cases, or occasional decreases due to an overall loss of suitable area in the region.

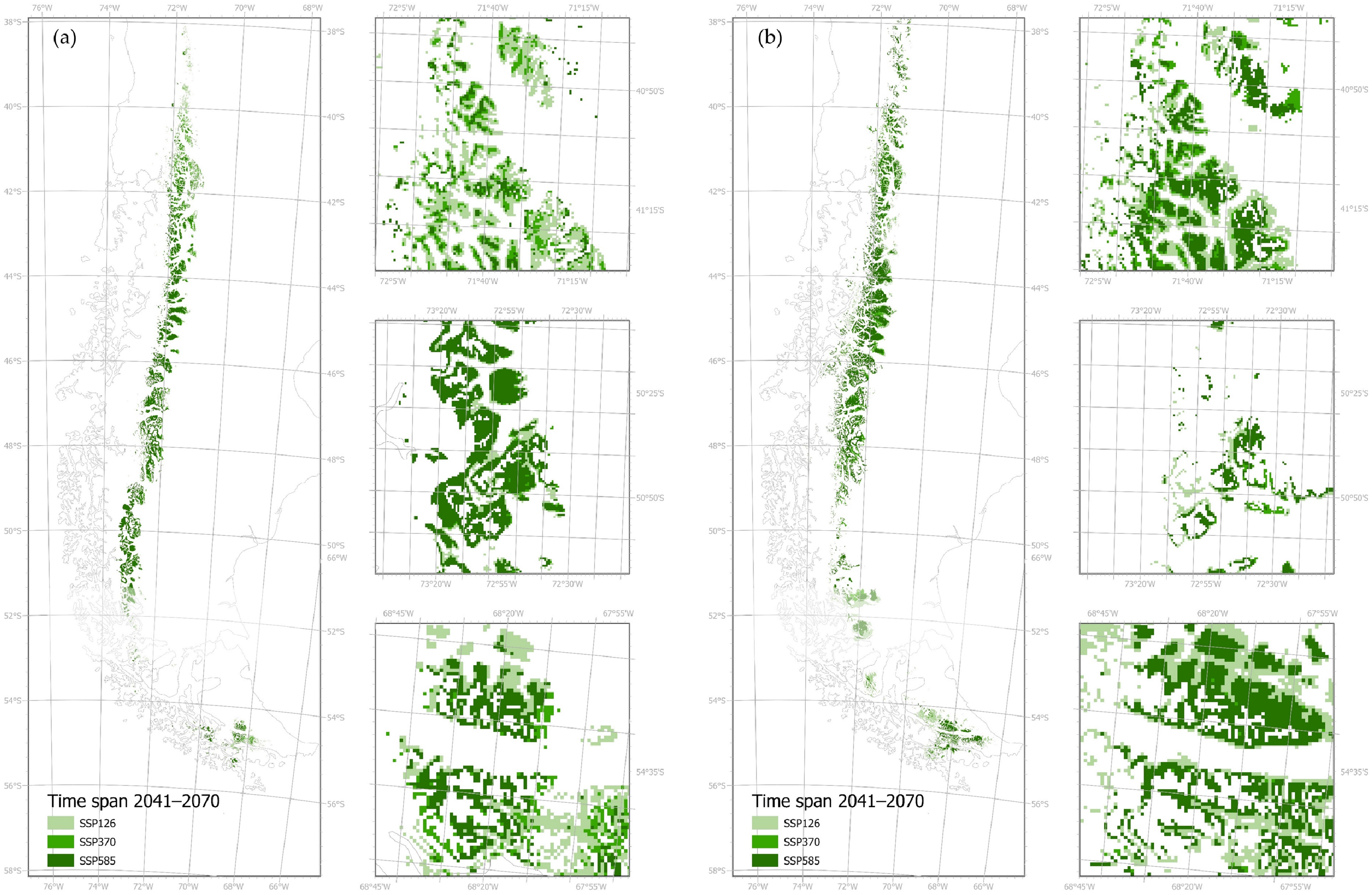


Figure 6. Potential future distribution of *Nothofagus pumilio* modelled by (**a**) the Random Forest classification and (**b**) RF regression model using CMIP6 SSP Scenarios for CHELSA Bioclim variables for the time span of 2041 to 2070.

Table 3. Treeline elevation estimates based on the highest raster cells from the model outputs (Random Forest classification and RF regression) under current climatic conditions and CMIP6 SSP scenarios for the future period 2041–2070. NA = not available (no data recorded).

Coordinates		Treeline Elevation [m] Current Climate		Treeline Elevation [m] SSP126 (2041–2070)		Treeline Elevation [m] SSP370 (2041–2070)		Treeline Elevation [m] SSP585 (2041–2070)	
X	Y	RF Class.	RF Reg.	RF Class.	RF Reg.	RF Class.	RF Reg.	RF Class.	RF Reg.
−71.00	−35.36	NA	NA	NA	NA	NA	NA	NA	NA
−71.11	−37.27	1988	1949	2214	NA	NA	NA	NA	NA
−71.33	−38.42	1854	1789	2201	1709	2186	1920	2227	2045
−72.15	−40.42	1591	1437	1699	1636	1768	1674	2026	1674
−72.19	−41.48	1500	1201	1560	1464	1730	1638	1730	1720
71.45	−43.07	1839	1440	1918	1545	2059	1725	1918	1725
−71.42	−44.39	1320	1216	1704	1324	1852	1509	1852	1591
−72.24	−47.12	1361	1197	1500	1423	1651	1439	1538	1500
−72.30	−48.30	1522	1074	1340	1098	1473	1098	1586	1209
−72.54	−50.57	1176	956	1296	961	1313	1103	1349	1124
−71.00	−53.00	543	560	592	721	NA	783	592	783
−68.45	−54.17	544	520	648	615	667	547	607	607
−67.30	−54.57	610	610	614	492	557	614	NA	614

4. Discussion

The treeline species *Nothofagus pumilio* is highly sensitive to climate variations, as reflected in its radial growth patterns and seedling establishment above the treeline. Consequently, research into the species' treeline dynamics in response to climate change has already gained some attention. Many studies have examined growth variations using dendrochronology, providing insights into the species' response to climate variations, mostly related to expressions in ENSO, PDO, and AAO/SAM over the past century [22,25,26]. Other research has focused on seedling establishment [23,24,75] or assessed the quantity and quality of *N. pumilio* seeds [17,75,76]. However, large-scale Species Distribution Modelling (SDM) approaches investigating the current and future development of the species are lacking. In this study, we calculated two Random Forest (RF) models using a binary and a continuous species dataset to model the current and future potential distribution of the species.

Both models, RF classification and RF regression, achieved reasonable results and good performance values. The RF classification model achieved an AUC of 0.93 (95% CI: 0.90–0.96), whereas the RF regression model explained R^2 = 0.39 (95% CI: 0.34–0.44) of the variance. This discrepancy is primarily due to the greater noise and structural complexity inherent in the continuous response variable, as well as the inherently stricter nature of R^2 as a performance metric. The use of spatially blocked cross-validation further amplifies this issue, because the model must extrapolate beyond clusters of spatially autocorrelated observations along the treeline. Consequently, an R^2 of around 0.4 can already be considered good performance in ecological regression tasks. The RF algorithm is well-suited for this study due to its strong predictive power and, more importantly, its applicability to both datasets, enabling direct comparability. Moreover, RF is a well-established algorithm that facilitates comparisons between different modelling approaches within the field of SDM [60,77]. While binary approaches are mainly used in SDM, the use of a continuous target variable generated from remote sensing, rather than the generation of abiotic predictors, is still a novel approach. The continuous data were derived from 20 m raster cells, with coverage aggregated to the target resolution of 1 km. While the binary variable contains only information on presence or absence, the continuous data reflect additional influences from land cover and species composition, which affect cover values and provide ecologically meaningful information. Other studies investigating topographic complex regions have already discussed the loss of knowledge due to the use of binary data at 1 km spatial resolution [56]. Here, we demonstrated an information gain using 1 km resolution data while incorporating continuous variables.

It should be noted that the scales of the modelling approaches (RF classification and RF regression) differ. However, we compare the model outputs, predicted probability and predicted cover values, on the basis that both a low predicted probability and a low cover value indicate largely unsuitable bioclimatic conditions, whereas high probability values and cover values point to suitable conditions. Nevertheless, it is important to acknowledge that a probability of 0.01 reflects only a very low likelihood of the species occurring at all, whereas 1% cover implies the species is present, albeit in low abundance. Although the models differ in their spatial predictions and variable importance, the correlation between predicted probability and cover values is relatively high, with a Pearson's r of 0.6. We therefore treat the two metrics as complementary, not interchangeable, and interpret model outputs jointly.

4.1. Current Distribution Range of N. pumilio

Differences between the two models were apparent across all analyses. In terms of spatial prediction, the RF regression model captured fine-scale details more accurately and was less prone to overprediction at high elevations. Additionally, the key predictors for the RF models varied. While a temperature variable had the highest importance (Accuracy Importance) in the RF classification model, precipitation variables had the highest importance in the RF regression model. Although the SHAP plots differ slightly in their expression between the models, the suitable bioclimatic conditions for *N. pumilio* are clearly evident in both. The species benefits from cold winters (bio 11, mean daily mean air temperatures of the coldest quarter) combined with moderate levels of precipitation, presumably in the form of protective snow cover (bio 19, mean monthly precipitation amount of the coldest quarter), and cool summers (moderate values of bio 10 (mean daily mean air temperatures of the warmest quarter)). Moreover, *N. pumilio* tends to occur in regions with sufficient overall precipitation and low precipitation seasonality, reflected by low to medium values for bio 17 (mean monthly precipitation amount of the driest quarter), bio 18 (mean monthly precipitation amount of the warmest quarter), and bio 15 (precipitation seasonality). Extreme heat or pronounced drought conditions inhibit the species' presence. Both temperature and precipitation seasonality are low, indicating that the species does not occur in areas characterised by extreme temperature fluctuations or erratic precipitation patterns, such as extended dry periods. The orophilous species *N. pumilio* is particularly adapted to the harsh climatic conditions of high mountain ecosystems. The species shows high phenotypic plasticity. While occurring at lower elevations as an erect tree up to a height of 35 m, it shows krummholz growth forms at the treeline [78]. As a deciduous species, it reduces transpiration in the months when frost-drought can be a problem [79]. Heat deficiency is considered to be the most important site factor for treeline formation worldwide. The treeline of *N. pumilio* follows an isotherm of 6.6 °C [19]. The influence of temperature was shown in the RF classification model by the high importance of bio 8 (mean daily air temperature of the wettest quarter), by the fact that the variable bio 11 (mean daily mean air temperatures of the coldest quarter) has a very important influence on the decisions at the nodes of both models (Gini Importance), and in the SHAP plots. However, when modelling a species in high mountains, the influence of elevation can also be represented by temperature variables. In the SHAP plot of the RF classification model, only very high temperature values in the warmest quarter (bio 10) result in a reduction in predicted probability. In contrast, in the RF regression model, very low temperatures during the warmest quarter lead to a decrease in predicted cover values. This difference is also reflected in the spatial predictions, with the RF classification model showing slight overprediction in summit regions. However, high temperatures during the warmest quarter can also lead to drought events, particularly in high mountain ecosystems where insolation and consequently evapotranspiration is very high. Authors analysing the sensitivity of *N. pumilio* to climate change using changes in tree rings have found a correlation with precipitation regimes in addition to temperatures. The species occurs in humid to arid regions and is well adapted to medium to low precipitation sums that occur at high elevations due to advective precipitation. This is also shown by its occurrence as far as the arid east, where it sometimes forms two treelines: an alpine treeline and a xeric treeline towards the arid steppes [80]. Very low values for precipitation variables (bio 17, 18 and 19) and precipitation seasonality (bio 15) define the limits in the core range and the eastern boundary of the species. However, particularly in northern Patagonia, increasingly low precipitation during the spring and summer months negatively affects tree growth [22,81]. In more southerly regions, high spring precipitation is primarily associated with prolonged snow cover, leading to a shortened growing season, which in turn also hinders tree growth [22,82,83]. Between 1900 and 2020, tree growth was found to correlate most strongly with rising temperatures on mesic sites, followed by wetter sites, while growth rates declined on drier sites [25]. Some authors state that all treelines in southern South America have experienced a negative growth trend due to drought since the 1980s and even suggest that the limitation at the treeline has changed from cold-limited to drought-limited [26,84]. Our results of the RF regression model seem plausible in this context, as the precipitation variables bio 15 (precipitation seasonality) and bio 19 (mean monthly precipitation of the coldest quarter) played an important role. Both variables reflect annual precipitation distribution patterns. Bio 15 suggests that *N. pumilio* is unlikely to occur in regions with high precipitation seasonality, which may be associated with periods of drought. Bio 19, on the other hand, may be linked to precipitation in the form of snow, the resulting snow cover, and water availability following thaw.

4.2. Future Distribution Range of N. pumilio

Abrupt treelines respond less to global warming than diffuse treelines, primarily due to higher seedling mortality outside the protecting forest stand climate [14]. The emergence and establishment of seedlings represent the most critical life stage for trees at the treeline, with the availability of species-specific safe sites being the basic precondition for seedling recruitment [85,86]. Seed production, fruit dispersal, seed viability, and seedling establishment all decline with increasing elevation [17]. Higher temperatures and an extension of the growing season can thus facilitate seedling emergence and survival [19,23,24,75,81], a precondition for a future treeline advance. However, the comparatively slow advance or persistence of the treeline in the southern Andes is also linked to multi-faceted interactions with edaphic, topographic, biotic, and other factors, including the development of alpine mats [24,28]. Demographic constraints across different life stages have been highlighted in previous studies examining the relationship between climate change and tree habitats. As treelines shift to higher elevations, trees are exposed to new climate–habitat interactions, and different life stages may respond in distinct ways [87,88]. For instance, tree fertility is primarily influenced by temperature, whereas seedling establishment depends more heavily on moisture availability and soil water content [87]. In fact, seedling survival declines with increasing drought at both high and low elevations, although at some high elevation stands this effect can be mitigated by spring snow cover [75]. A deterioration in growth conditions due to drought is also predicted for adult trees at lower elevations as well as at the treeline, based on studies of radial growth patterns [26,84,89]. However, even if temperature and precipitation conditions are favourable for seedling establishment, local factors such as steep topography and the absence of herbaceous vegetation can inhibit treeline advance [24,28]. While this study focuses on assessing the effects of key climatic drivers on the treeline, future modelling efforts should consider incorporating additional variables such as topography, wind exposure, soil characteristics, and vegetation cover.

The use of CMIP6 SSP scenarios from CHELSA showed an advance to higher elevations in both models, as indicated by spatial predictions and treeline elevation estimates derived from a digital surface model (DSM). The scenarios are based on global circulation models and CO_2 concentration estimates. It should be noted that less periodic variations caused by different phases of ENSO, PDO, and AAO/SAM cannot be fully modelled in the scenarios [90]. While the RF classification model showed a significant decrease in the northern range, lower cover values of the RF regression model remained at higher elevations. The result that *N. pumilio* occurrences decreased at lower elevations is consistent with previous findings [75], highlighting a decrease in the number of seedlings and a reduction in survival at lower-elevated sites. A review study modelling biome-level changes predicts a reduction in area with suitable climatic conditions for temperate deciduous forests of approximately 30% under the RCP8.5 scenario for the period 2040–2070 [16]. Our results seem congruent with these findings, whereas the RF classification model predicts a greater decline than the RF regression model. In particular, the RF classification model predicted a shift towards the wetter western region. It will be necessary to investigate the future competitive relationships with dominant tree species in this area (e.g., *N. betuloides*). In summary, precipitation conditions/drought stress will play a significant role in future competitive relationships and successful regeneration of *Nothofagus* species in the southern Andes.

5. Conclusions

To our knowledge, this is the first SDM study that models the current and future distribution of *N. pumilio* across its entire distribution range in the southern Andes. Even though the distribution range encompasses two extreme climatic gradients, both models were able to comprehensively predict the current potential distribution and its future development. The direct comparison of model approaches highlighted major differences in the model results and the advantages of using more complex, continuous data. Continuous data can provide better insights into suitable bioclimatic conditions for *N. pumilio* occurrence leading to more detailed spatial predictions and meaningful predictors based on variable importance. In contrast to presence–absence data, which can only take values of 0 or 1, continuous cover values can reflect subtle or unknown effects of land cover, topography, and species composition, thereby providing model results of greater ecological value. However, we acknowledge that remote sensing data across a large geographic extent are rarely available, difficult to obtain in very high resolution, and may still contain gaps that introduce bias. Climatic parameters represent the principal limiting factors at the alpine treeline, and bioclimatic variables have already proven effective in capturing the climatic conditions at the treeline of the southern Andes. It will be of great interest to further model the conditions at the treeline with more complex abiotic predictors, like topography, wind and soil variables as well as to embed biotic and anthropogenic variables to model the influences of vegetation composition, fire, grazing, and forestry.

Author Contributions: Conceptualization, M.W., J.B., J.O., U.S., J.W. and M.B.; methodology, M.W., J.O., J.W. and M.B.; validation, M.W., J.O. and M.B.; formal analysis, M.W.; investigation, M.W.; data curation, M.W.; writing—original draft preparation, M.W.; writing—review and editing, M.W., J.B.,

J.O., U.S. and M.B.; visualisation, M.W.; supervision, J.B. and U.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The species input data used in this study are online available at the free data provider of University of Hamburg: http://doi.org/10.25592/uhhfdm.16239, accessed on 7 June 2025.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A.1

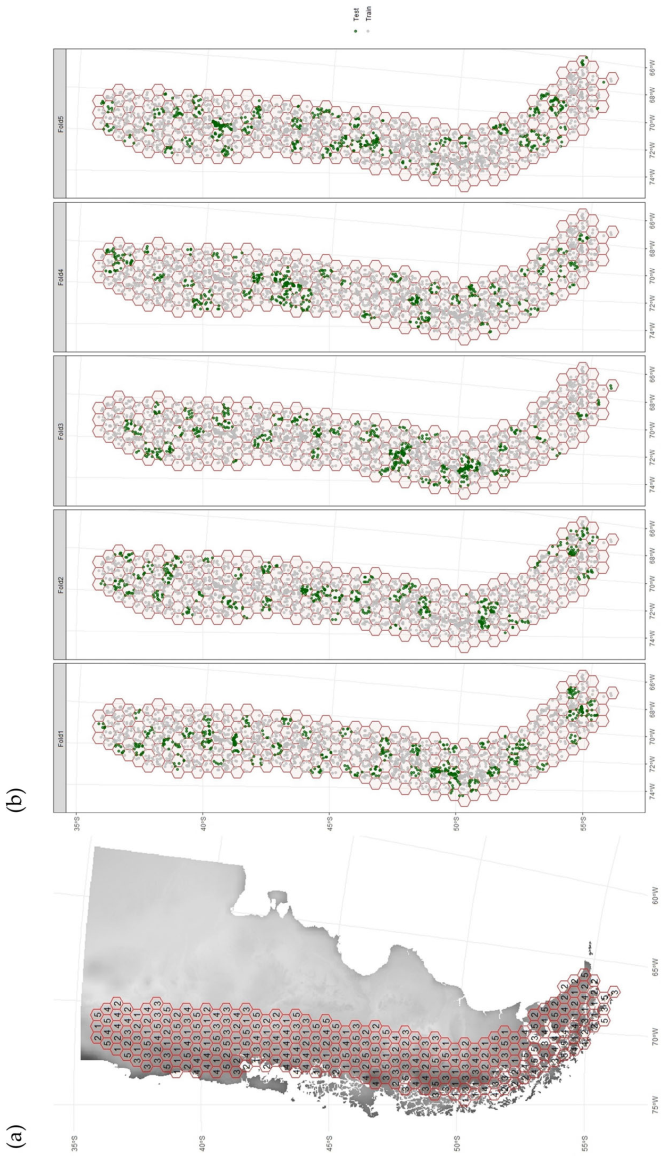


Figure A1. (**a**) Hexagonal spatial blocks with the allocation of folds (1–5), created using the "blockCV" R package, and (**b**) the five folds showing test data points (dark green) and training data points (grey) for the Random Forest classification model.

Appendix A.2

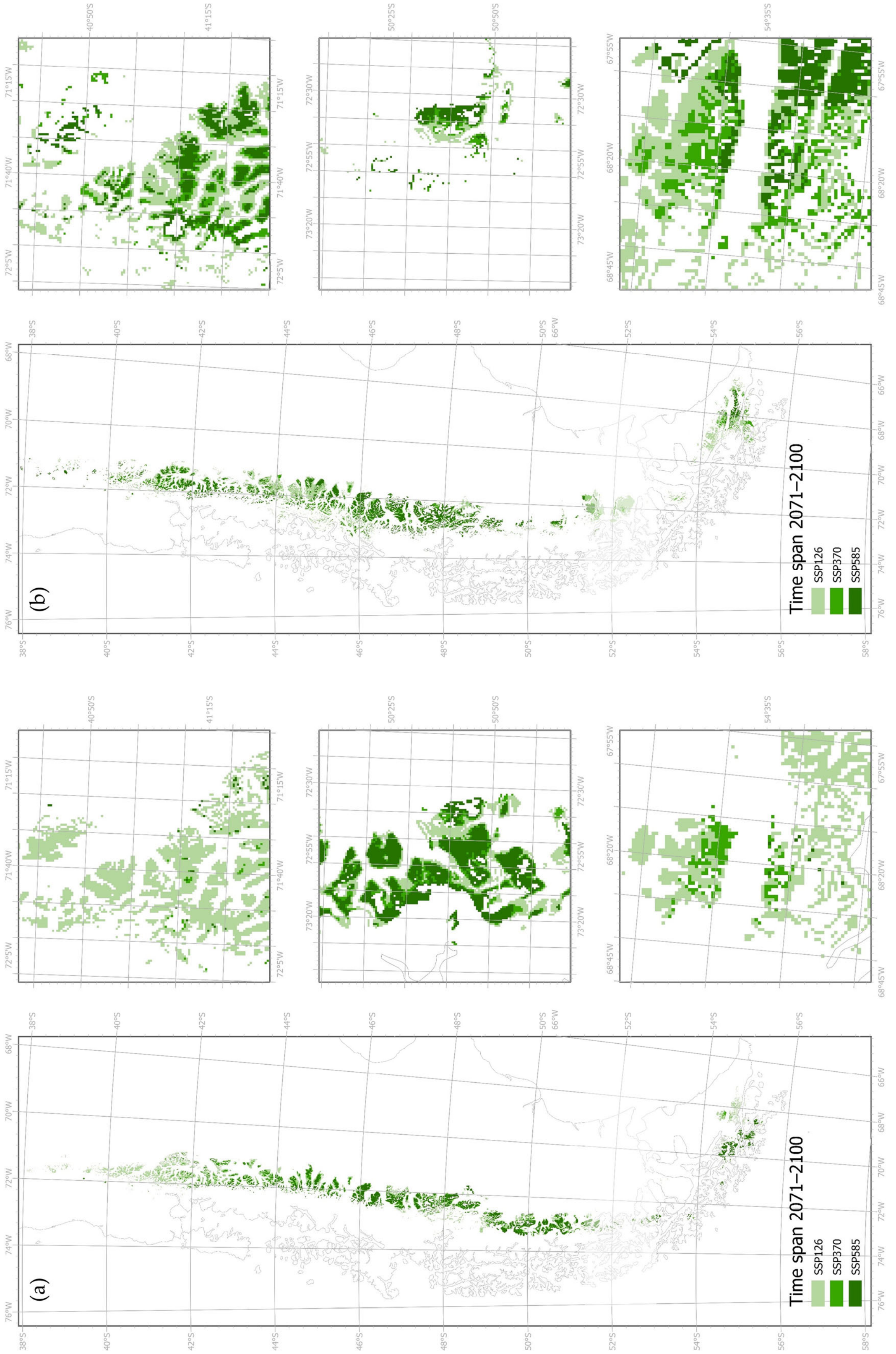


Figure A2. Potential future distribution of *Nothofagus pumilio* modelled by (**a**) the Random Forest classification and (**b**) RF regression model using CMIP6 SSP Scenarios for CHELSA Bioclim variables for the time span of 2071 to 2100.

Appendix A.3

Table A1. Results of the 5-fold spatial cross-validation of the Random Forest classification model. The table presents the mean values across all five models for the various hyperparameter settings (45 models in total). The hyperparameter setting for the final model, highlighted in bold, was selected based on the highest mean AUC.

No.	mtry	ntree	AUC Mean	AUC 95% CI [1]	Acc. Mean	Acc. 95% CI [1]	TSS Mean	TSS 95% CI [1]
1	2	100	0.9270	0.8908–0.9632	0.8460	0.7782–0.9139	0.6125	0.4169–0.8081
2	3	100	0.9204	0.8785–0.9623	0.8407	0.7803–0.9011	0.5968	0.4164–0.7772
3	4	100	0.9192	0.8821–0.9564	0.8454	0.7761–0.9147	0.6106	0.4025–0.8187
4	2	300	0.9268	0.8943–0.9592	0.8441	0.7767–0.9115	0.6053	0.4063–0.8043
5	3	300	0.9237	0.8884–0.9590	0.8473	0.7827–0.9118	0.6163	0.4212–0.8114
6	4	300	0.9228	0.8898–0.9557	0.8461	0.7831–0.9092	0.6127	0.4265–0.7988
7	**2**	**500**	**0.9279**	**0.8960–0.9599**	**0.8466**	**0.7799–0.9132**	**0.6148**	**0.4183–0.8112**
8	3	500	0.9233	0.8881–0.9585	0.8438	0.7765–0.9111	0.6025	0.4015–0.8035
9	4	500	0.9224	0.8837–0.9610	0.8429	0.7756–0.9102	0.6050	0.4082–0.8019

[1] Confidence intervals.

Table A2. Results of the 5-fold spatial cross-validation of the Random Forest regression model. Hyperparameter for the final model, highlighted in bold, were selected based on the highest mean R^2.

No.	mtry	ntree	R^2 Mean	95% CI [1]
1	2	100	0.3910	0.3419–0.4400
2	3	100	0.3868	0.3366–0.4369
3	4	100	0.3835	0.3323–0.4348
4	2	300	0.3933	0.3432–0.4433
5	3	300	0.3892	0.3378–0.4407
6	4	300	0.3873	0.3362–0.4384
7	**2**	**500**	**0.3933**	**0.3425–0.4441**
8	3	500	0.3898	0.3386–0.4410
9	4	500	0.3869	0.3357–0.4381

[1] Confidence intervals.

Table A3. Treeline elevation estimates based on the highest raster cells from the model outputs (Random Forest classification and RF regression) under current climatic conditions and SSP scenarios for the future period 2071–2100. NA = not available (no data recorded).

Coordinates		Treeline Elevation [m] Current Climate		Treeline Elevation [m] SSP126 (2071–2100)		Treeline Elevation [m] SSP370 (2071–2100)		Treeline Elevation [m] SSP585 (2071–2100)	
X	Y	RF Class.	RF Reg.	RF Class.	RF Reg.	RF Class.	RF Reg.	RF Class.	RF Reg.
−71.00	−35.36	NA	NA	NA	NA	NA	NA	NA	NA
−71.11	−37.27	1988	1949	2328	NA	NA	NA	NA	2530
−71.33	−38.42	1854	1789	2071	1700	1780	1871	2460	2035
−72.15	−40.42	1591	1437	1679	1636	2026	1971	2026	2026
−72.19	−41.48	1500	1201	1555	1510	1917	1743	1917	1730
71.45	−43.07	1839	1440	1955	1545	2059	1955	2059	2059
−71.42	−44.39	1320	1216	1703	1595	1952	1593	1952	1427
−72.24	−47.12	1361	1197	1439	1346	1742	1651	1901	1840
−72.30	−48.30	1522	1074	1399	1098	1578	1171	1698	1340
−72.54	−50.57	1176	956	1317	1000	1457	1229	1537	1287
−71.00	−53.00	543	560	592	728	NA	NA	NA	NA
−68.45	−54.17	544	520	648	616	615	757	NA	NA
−67.30	−54.57	610	610	609	614	NA	614	NA	614

References

1. Holtmeier, F.-K.; Broll, G. Treeline Research—From the Roots of the Past to Present Time. A Review. *Forests* **2020**, *11*, 38. [CrossRef]
2. Schickhoff, U.; Bobrowski, M.; Mal, S.; Schwab, N.; Singh, R.B. The World's Mountains in the Anthropocene. In *Mountain Landscapes in Transition*; Schickhoff, U., Singh, R.B., Mal, S., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 1–144. ISBN 978-3-030-70237-3.
3. Körner, C. Climatic Controls of the Global High Elevation Treelines. In *Encyclopedia of the World's Biomes*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 275–281. ISBN 9780128160978.
4. Paulsen, J.; Körner, C. A climate-based model to predict potential treeline position around the globe. *Alp. Bot.* **2014**, *124*, 1–12. [CrossRef]
5. Qiu, J.; Feng, S.; Yuan, W. Upward-moving mountain treelines: An indicator of changing climate. *Glob. Change Biol.* **2023**, *29*, 6832–6833. [CrossRef]
6. Schickhoff, U.; Bobrowski, M.; Böhner, J.; Bürzle, B.; Chaudhary, R.P.; Gerlitz, L.; Heyken, H.; Lange, J.; Müller, M.; Scholten, T.; et al. Do Himalayan treelines respond to recent climate change? An evaluation of sensitivity indicators. *Earth Syst. Dynam.* **2015**, *6*, 245–265. [CrossRef]
7. Schickhoff, U.; Bobrowski, M.; Böhner, J.; Bürzle, B.; Chaudhary, R.P.; Müller, M.; Scholten, T.; Schwab, N.; Weidinger, J. The Treeline Ecotone in Rolwaling Himal, Nepal: Pattern-Process Relationships and Treeline Shift Potential. In *Ecology of Himalayan Treeline Ecotone*; Singh, S.P., Reshi, Z.A., Joshi, R., Eds.; Springer Nature: Singapore, 2023; pp. 95–145. ISBN 978-981-19-4475-8.
8. Treml, V.; Veblen, T.T. Does tree growth sensitivity to warming trends vary according to treeline form? *J. Biogeogr.* **2017**, *44*, 1469–1480. [CrossRef]
9. Harsch, M.A.; Hulme, P.E.; McGlone, M.S.; Duncan, R.P. Are treelines advancing? A global meta-analysis of treeline response to climate warming. *Ecol. Lett.* **2009**, *12*, 1040–1049. [CrossRef]
10. Hansson, A.; Dargusch, P.; Shulmeister, J. A review of modern treeline migration, the factors controlling it and the implications for carbon storage. *J. Mt. Sci.* **2021**, *18*, 291–306. [CrossRef]
11. Hansson, A.; Shulmeister, J.; Dargusch, P.; Hill, G. A review of factors controlling Southern Hemisphere treelines and the implications of climate change on future treeline dynamics. *Agric. For. Meteorol.* **2023**, *332*, 109375. [CrossRef]
12. Lu, X.; Liang, E.; Wang, Y.; Babst, F.; Camarero, J.J. Mountain treelines climb slowly despite rapid climate warming. *Glob. Ecol. Biogeogr.* **2021**, *30*, 305–315. [CrossRef]
13. Harsch, M.A.; Bader, M.Y. Treeline form—A potential key to understanding treeline dynamics. *Glob. Ecol. Biogeogr.* **2011**, *20*, 582–596. [CrossRef]
14. Bader, M.Y.; Llambí, L.D.; Case, B.S.; Buckley, H.L.; Toivonen, J.M.; Camarero, J.J.; Cairns, D.M.; Brown, C.D.; Wiegand, T.; Resler, L.M. A global framework for linking alpine-treeline ecotone patterns to underlying processes. *Ecography* **2021**, *44*, 265–292. [CrossRef]
15. Körner, C. *Alpine Treelines*; Springer: Basel, Switzerland, 2012; ISBN 978-3-0348-0395-3.
16. Tovar, C.; Carril, A.F.; Gutiérrez, A.G.; Ahrends, A.; Fita, L.; Zaninelli, P.; Flombaum, P.; Abarzúa, A.M.; Alarcón, D.; Aschero, V.; et al. Understanding climate change impacts on biome and plant distributions in the Andes: Challenges and opportunities. *J. Biogeogr.* **2022**, *49*, 1420–1442. [CrossRef] [PubMed]
17. Cuevas, J.G. Tree recruitment at the *Nothofagus pumilio* alpine timberline in Tierra del Fuego, Chile. *J. Ecol.* **2000**, *88*, 840–855. [CrossRef]
18. Holtmeier, F.-K. *Mountain Timberlines: Ecology, Patchiness, and Dynamics*; Springer: Dordrecht, The Netherlands, 2009; ISBN 978-1-4020-9704-1.
19. Fajardo, A.; Piper, F.I. An experimental approach to explain the southern Andes elevational treeline. *Am. J. Bot.* **2014**, *101*, 788–795. [CrossRef] [PubMed]
20. Bozkurt, D.; Rojas, M.; Boisier, J.P.; Valdivieso, J. Projected hydroclimate changes over Andean basins in central Chile from downscaled CMIP5 models under the low and high emission scenarios. *Clim. Change* **2018**, *150*, 131–147. [CrossRef]
21. Pessacg, N.; Flaherty, S.; Solman, S.; Pascual, M. Climate change in northern Patagonia: Critical decrease in water resources. *Theor. Appl. Clim.* **2020**, *140*, 807–822. [CrossRef]
22. Álvarez, C.; Veblen, T.T.; Christie, D.A.; González-Reyes, Á. Relationships between climate variability and radial growth of *Nothofagus pumilio* near altitudinal treeline in the Andes of northern Patagonia, Chile. *For. Ecol. Manag.* **2015**, *342*, 112–121. [CrossRef]
23. Srur, A.M.; Villalba, R.; Rodríguez-Catón, M.; Amoroso, M.M.; Marcotti, E. Establishment of *Nothofagus pumilio* at Upper Treelines Across a Precipitation Gradient in the Northern Patagonian Andes. *Arct. Antarct. Alp. Res.* **2016**, *48*, 755–766. [CrossRef]
24. Srur, A.M.; Villalba, R.; Rodríguez-Catón, M.; Amoroso, M.M.; Marcotti, E. Climate and *Nothofagus pumilio* Establishment at Upper Treelines in the Patagonian Andes. *Front. Earth Sci.* **2018**, *6*, 57. [CrossRef]
25. Brand, R.; Srur, A.M.; Villalba, R. Contrasting growth trends in *Nothofagus pumilio* upper-elevation forests induced by climate warming in the Southern Andes. *Agric. For. Meteorol.* **2022**, *323*, 109083. [CrossRef]

26. Reiter, E.J.; Weigel, R.; Walentowski, H.; Loguercio, G.A.; Fierke, J.; Winter, A.F.N.; Simon, A.; Kotowska, M.M.; Joelson, N.Z.; Caselli, M.; et al. Climate vulnerability of *Nothofagus pumilio*, *Nothofagus dombeyi* and *Austrocedrus chilensis* in northern Patagonia's temperate forests. *For. Ecol. Manag.* **2024**, *572*, 122261. [CrossRef]
27. Rodriguez-Souilla, J.; Cellini, J.M.; Roig, F.A.; Lencinas, M.V.; Chaves, J.E.; Peri, P.L.; Martínez Pastur, G.J. Climate influence seed production in managed and unmanaged *Nothofagus pumilio* forests of Southern Patagonia. *NZJFS* **2024**, *54*, 1–14. [CrossRef]
28. Pissolito, C. Sucesión Primaria de la Vegetación en Bosques de *Nothofagus* spp. en los Andes Patagónicos: Patrones Espacio Temporales de Establecimiento de Plantas e Implicancias Para la Restauración. Ph.D. Thesis, Universidad Nacional de Cuyo, Mendoza, Argentina, 2016.
29. Feng, X.; Park, D.S.; Walker, C.; Peterson, A.T.; Merow, C.; Papeş, M. A checklist for maximizing reproducibility of ecological niche models. *Nat. Ecol. Evol.* **2019**, *3*, 1382–1395. [CrossRef]
30. Karger, D.N.; Conrad, O.; Böhner, J.; Kawohl, T.; Kreft, H.; Soria-Auza, R.W.; Zimmermann, N.E.; Linder, H.P.; Kessler, M. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **2017**, *4*, 170122. [CrossRef] [PubMed]
31. Karger, D.N.; Wilson, A.M.; Mahony, C.; Zimmermann, N.E.; Jetz, W. Global daily 1 km land surface precipitation based on cloud cover-informed downscaling. *Sci. Data* **2021**, *8*, 307. [CrossRef] [PubMed]
32. Pocock, M.J.O.; Tweddle, J.C.; Savage, J.; Robinson, L.D.; Roy, H.E. The diversity and evolution of ecological and environmental citizen science. *PLoS ONE* **2017**, *12*, e0172579. [CrossRef] [PubMed]
33. Aceves-Bueno, E.; Adeleye, A.S.; Feraud, M.; Huang, Y.; Tao, M.; Yang, Y.; Anderson, S.E. The Accuracy of Citizen Science Data: A Quantitative Review. *Bull. Ecol. Soc. Am.* **2017**, *98*, 278–290. [CrossRef]
34. Beck, J.; Böller, M.; Erhardt, A.; Schwanghart, W. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecol. Inform.* **2014**, *19*, 10–15. [CrossRef]
35. Marcer, A.; Chapman, A.D.; Wieczorek, J.R.; Xavier Picó, F.; Uribe, F.; Waller, J.; Ariño, A.H. Uncertainty matters: Ascertaining where specimens in natural history collections come from and its implications for predicting species distributions. *Ecography* **2022**, *2022*, e06025. [CrossRef]
36. Chauvier, Y.; Zimmermann, N.E.; Poggiato, G.; Bystrova, D.; Brun, P.; Thuiller, W. Novel methods to correct for observer and sampling bias in presence-only species distribution models. *Glob. Ecol. Biogeogr.* **2021**, *30*, 2312–2325. [CrossRef]
37. Rocchini, D.; Boyd, D.S.; Féret, J.-B.; Foody, G.M.; He, K.S.; Lausch, A.; Nagendra, H.; Wegmann, M.; Pettorelli, N. Satellite remote sensing to monitor species diversity: Potential and pitfalls. *Remote Sens. Ecol. Conserv.* **2016**, *2*, 25–36. [CrossRef]
38. Reddy, C.S. Remote sensing of biodiversity: What to measure and monitor from space to species? *Biodivers. Conserv.* **2021**, *30*, 2617–2631. [CrossRef]
39. Wang, L.; Diao, C.; Lu, Y. The role of remote sensing in species distribution models: A review. *Int. J. Remote Sens.* **2025**, *46*, 661–685. [CrossRef]
40. Werner, M.; Weidinger, J.; Böhner, J.; Schickhoff, U.; Bobrowski, M. Instagram data for validating *Nothofagus pumilio* distribution mapping in the Southern Andes: A novel ground truthing approach. *Biogeography* **2024**, *17*, e140606. [CrossRef]
41. Amigo Vázquez, F.J.; Rodríguez Guitián, M.A. Bioclimatic and phytosociological diagnosis of the species of the *Nothofagus* genus (Nothofagaceae) in South America. *IJGR* **2011**, *1*, 1–20. [CrossRef]
42. Soliani, C.; Tsuda, Y.; Bagnoli, F.; Gallo, L.A.; Vendramin, G.G.; Marchelli, P. Halfway encounters: Meeting points of colonization routes among the southern beeches *Nothofagus pumilio* and N. antarctica. *Mol. Phylogenet. Evol.* **2015**, *85*, 197–207. [CrossRef]
43. Garreaud, R.; Lopez, P.; Minvielle, M.; Rojas, M. Large-Scale Control on the Patagonian Climate. *J. Clim.* **2013**, *26*, 215–230. [CrossRef]
44. Meyer, C.; Weigelt, P.; Kreft, H. Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol. Lett.* **2016**, *19*, 992–1006. [CrossRef]
45. Ward, D.F. Understanding sampling and taxonomic biases recorded by citizen scientists. *J. Insect Conserv.* **2014**, *18*, 753–756. [CrossRef]
46. Bird, T.J.; Bates, A.E.; Lefcheck, J.S.; Hill, N.A.; Thomson, R.J.; Edgar, G.J.; Stuart-Smith, R.D.; Wotherspoon, S.; Krkosek, M.; Stuart-Smith, J.F.; et al. Statistical solutions for error and bias in global citizen science datasets. *Biol. Conserv.* **2014**, *173*, 144–154. [CrossRef]
47. Zizka, A.; Antonelli, A.; Silvestro, D. Sampbias, a method for quantifying geographic sampling biases in species distribution data. *Ecography* **2021**, *44*, 25–32. [CrossRef]
48. Phillips, S.J.; Dudík, M.; Elith, J.; Graham, C.H.; Lehmann, A.; Leathwick, J.; Ferrier, S. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecol. Appl.* **2009**, *19*, 181–197. [CrossRef] [PubMed]
49. Barbet-Massin, M.; Jiguet, F.; Albert, C.H.; Thuiller, W. Selecting pseudo-absences for species distribution models: How, where and how many? *Methods Ecol. Evol.* **2012**, *3*, 327–338. [CrossRef]
50. Hijmans, R.J.; Cameron, S.E.; Parra, J.L.; Jones, P.G.; Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Intl. J. Climatol.* **2005**, *25*, 1965–1978. [CrossRef]

51. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Intl. J. Climatol.* **2017**, *37*, 4302–4315. [CrossRef]
52. Soria-Auza, R.W.; Kessler, M.; Bach, K.; Barajas-Barbosa, P.M.; Lehnert, M.; Herzog, S.K.; Böhner, J. Impact of the quality of climate models for modelling species occurrences in countries with poor climatic documentation: A case study from Bolivia. *Ecol. Model.* **2010**, *221*, 1221–1229. [CrossRef]
53. Bobrowski, M.; Weidinger, J.; Schickhoff, U. Is New Always Better? Frontiers in Global Climate Datasets for Modeling Treeline Species in the Himalayas. *Atmosphere* **2021**, *12*, 543. [CrossRef]
54. Jentsch, H.; Weidinger, J.; Bobrowski, M. *ClimDatDownloadR: Downloads Climate Data from Chelsa and WorldClim*; Zenodo: Geneva, Switzerland, 2023. [CrossRef]
55. Jentsch, H. ClimDatDownloadR. GitHub, Inc., 2025. Available online: https://github.com/HelgeJentsch/ClimDatDownloadR (accessed on 7 June 2025).
56. Bobrowski, M.; Gerlitz, L.; Schickhoff, U. Modelling the potential distribution of Betula utilis in the Himalaya. *Glob. Ecol. Conserv.* **2017**, *11*, 69–83. [CrossRef]
57. Genuer, R.; Poggi, J.-M.; Tuleau-Malot, C. VSURF: An R Package for Variable Selection Using Random Forests. *R J.* **2015**, *7*, 19–33. [CrossRef]
58. Franklin, J. *Mapping Species Distributions*; Cambridge University Press: Cambridge, UK, 2012; ISBN 9780521876353.
59. Miller, J. Species Distribution Modeling. *Geogr. Compass* **2010**, *4*, 490–509. [CrossRef]
60. Li, X.; Wang, Y. Applying various algorithms for species distribution modelling. *Integr. Zool.* **2013**, *8*, 124–135. [CrossRef]
61. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
62. Breiman, L. Statistical Modeling: The Two Cultures. *Stat. Sci.* **2001**, *16*, 199–215. [CrossRef]
63. R Core Team. A Language and Environment for Statistical Computing. Available online: https://www.R-project.org/ (accessed on 11 September 2024).
64. Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; Wehberg, J.; Wichmann, V.; Böhner, J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* **2015**, *8*, 1991–2007. [CrossRef]
65. ESRI. ArcGIS Pro, Version 2.7.0. Available online: https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview (accessed on 12 February 2024).
66. Legendre, P. Spatial Autocorrelation: Trouble or New Paradigm? *Ecology* **1993**, *74*, 1659–1673. [CrossRef]
67. Dormann, C.F.; McPherson, J.M.; Araújo, M.B.; Bivand, R.; Bolliger, J.; Carl, G.; Davies, R.G.; Hirzel, A.; Jetz, W.; Daniel Kissling, W.; et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography* **2007**, *30*, 609–628. [CrossRef]
68. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
69. Valavi, R.; Elith, J.; Lahoz-Monfort, J.J.; Guillera-Arroita, G. blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **2019**, *10*, 225–232. [CrossRef]
70. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
71. Greenwell, B. *Fastshap: Fast Approximate Shapley Values*; R Package Version 0.1.1; The R Foundation: Vienna, Austria, 2024.
72. Li, M.; Sun, H.; Huang, Y.; Chen, H. Shapley value: From cooperative game to explainable artificial intelligence. *Auton. Intell. Syst.* **2024**, *4*, 2. [CrossRef]
73. Mohr-Bell, D.; Díaz, G.; Príncipe, R.; Gonzalez, C.; Bono, J.; Ciuffoli, L.; Strada, M.; Parmuchi, G.; Chomnalez, F.; Montenegro, C.; et al. *Monitoreo de la Superficie de Bosque Nativo de la República Argentina, Región Forestal Bosque Andino Patagónico*; Tomo I Informe; Secretaría de Ambiente y Desarrollo Sustentable de la Nación: Esquel, CHU, Argentina, 2019; p. 87.
74. Lara, A.; Villalba, R.; Wolodarsky-Franke, A.; Aravena, J.C.; Luckman, B.H.; Cuq, E. Spatial and temporal variation in *Nothofagus pumilio* growth at tree line along its latitudinal range (35°40′–55° S) in the Chilean Andes. *J. Biogeogr.* **2005**, *32*, 879–893. [CrossRef]
75. Aschero, V.; Srur, A.M.; Guerrido, C.; Villalba, R. Contrasting climate influences on *Nothofagus pumilio* establishment along elevational gradients. *Plant Ecol.* **2022**, *223*, 369–380. [CrossRef]
76. Toro-Manríquez, M.D.; Huertas Herrera, A.; Soler, R.M.; Lencinas, M.V.; Martínez Pastur, G.J. Combined effects of tree canopy composition, landscape location, and growing season on *Nothofagus* forest seeding patterns in Southern Patagonia. *For. Ecol. Manag.* **2023**, *529*, 120708. [CrossRef]
77. Elith, J.; Franklin, J. Species Distribution Modeling. In *Encyclopedia of Biodiversity*; Elsevier: Amsterdam, The Netherlands, 2013; pp. 692–705. ISBN 9780123847201.
78. Stecconi, M.; Puntieri, J.G.; Barthélémy, D. An architectural approach to the growth forms of *Nothofagus pumilio* (Nothofagaceae) along an altitudinal gradient. *Botany* **2010**, *88*, 699–709. [CrossRef]
79. Körner, C. *Alpine Plant Life*; Springer International Publishing: Cham, Switzerland, 2021; ISBN 978-3-030-59537-1.

80. Hertel, D.; Therburg, A.; Villalba, R. Above- and below-ground response by *Nothofagus pumilio* to climatic conditions at the transition from the steppe–forest boundary to the alpine treeline in southern Patagonia, Argentina. *Plant Ecol. Divers.* **2008**, *1*, 21–33. [CrossRef]
81. Daniels, L.D.; Veblen, T.T. Spatiotemporal Influences of Climate on Altitudinal Treeline in Northern Patagonia. *Ecology* **2004**, *85*, 1284–1296. [CrossRef]
82. Villalba, R.; Boninsegna, J.A.; Veblen, T.T.; Schmelter, A.; Rubulis, S. Recent Trends in Tree-Ring Records from high elevation Sites in the Andes of Northern Patagonia. *Clim. Change* **1997**, *36*, 425–454. [CrossRef]
83. Lara, A.; Aravena, J.C.; Villalba, R.; Wolodarsky-Franke, A.; Luckman, B.; Wilson, R. Dendroclimatology of high-elevation *Nothofagus pumilio* forests at their northern distribution limit in the central Andes of Chile. *Can. J. For. Res.* **2001**, *31*, 925–936. [CrossRef]
84. Fajardo, A.; Gazol, A.; Mayr, C.; Camarero, J.J. Recent decadal drought reverts warming-triggered growth enhancement in contrasting climates in the southern Andes tree line. *J. Biogeogr.* **2019**, *46*, 1367–1379. [CrossRef]
85. Batllori, E.; Camarero, J.J.; Ninot, J.M.; Gutiérrez, E. Seedling recruitment, survival and facilitation in alpine Pinus uncinata tree line ecotones. Implications and potential responses to climate warming. *Glob. Ecol. Biogeogr.* **2009**, *18*, 460–472. [CrossRef]
86. Bürzle, B.; Schickhoff, U.; Schwab, N.; Wernicke, L.M.; Müller, Y.K.; Böhner, J.; Chaudhary, R.P.; Scholten, T.; Oldeland, J. Seedling recruitment and facilitation dependence on safe site characteristics in a Himalayan treeline ecotone. *Plant Ecol.* **2018**, *219*, 115–132. [CrossRef]
87. Qiu, T.; Sharma, S.; Woodall, C.W.; Clark, J.S. Niche Shifts from Trees to Fecundity to Recruitment That Determine Species Response to Climate Change. *Front. Ecol. Evol.* **2021**, *9*, 719141. [CrossRef]
88. Lett, S.; Dorrepaal, E. Global drivers of tree seedling establishment at alpine treelines in a changing climate. *Funct. Ecol.* **2018**, *32*, 1666–1680. [CrossRef]
89. Rodríguez-Catón, M.; Villalba, R.; Morales, M.; Srur, A. Influence of droughts on *Nothofagus pumilio* forest decline across northern Patagonia, Argentina. *Ecosphere* **2016**, *7*, e01390. [CrossRef]
90. Karger, D.N.; Schmatz, D.R.; Dettling, G.; Zimmermann, N.E. High-resolution monthly precipitation and temperature time series from 2006 to 2100. *Sci. Data* **2020**, *7*, 248. [CrossRef]

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Sofern im Zuge der Erstellung der vorliegenden Dissertationsschrift generative Künstliche Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

Ich versichere weiterhin, dass das gebundene Exemplar und das in elektronischer Form eingereichte Exemplar der Dissertationsschrift identisch sind.

Kiel, 12.02.2026, M. Werne ______________________________ (Melanie Werner)

Auflistung der Nutzung von generativer KI

Zur Verbesserung der Satzstruktur und Grammatik einzelner Textpassagen wurde ChatGPT in den Versionen GPT-4o und GPT-5 eingesetzt. Dabei kam folgender Prompt zur Anwendung:

„Guten Morgen, ich schreibe gerade meine Dissertation. Kannst Du mir bitte helfen, indem Du die Sätze, die ich einfüge im British Scientific English verbesserst? Dabei ist es wichtig, dass die Aussage meines Satzes unverändert und nah am Original bleibt, sodass keine Bedenken im Sinne des Plagiats entstehen können, wenn ich Sätze komplett oder teilweise übernehme. Du kannst gerne auch Sätze zusammenfügen oder die Struktur leicht verbessern. Vielen Dank im Voraus!“

Der vollständige Chatverlauf wurde dokumentiert und archiviert, um die Nachvollziehbarkeit der Bearbeitungsschritte zu gewährleisten.