Modelling the ecological niche of a treeline species (*Betula utilis*) in the Himalayan region

---

Cumulative Dissertation

Dissertation to obtain a doctoral degree

at the Faculty of Mathematics,

Informatics and Natural Sciences

Department of Geosciences

University of Hamburg

submitted by

Maria Bobrowski

Hamburg, June 2018

Evaluation of doctoral dissertation by:

1. Prof. Dr. Udo Schickhoff

   Institute of Geography, University of Hamburg

2. Prof. Dr. Jürgen Böhner
   Institute of Geography, University of Hamburg

Date of defense: 30.10.2018

qed

5 Jahre und es passierte so vieles: Im Folgenden möchte ich mich bei denen bedanken, die maßgeblich zum Gelingen meiner *Betula*-Trilogie beigetragen haben.

Als erstes gebührt mein Dank Professor Dr. Udo Schickhoff. Ich bedanke mich für die Möglichkeit, diese Arbeit zu schreiben, das Vertrauen und die Freiheit, sie nach meinen Vorstellungen umzusetzen. Die Arbeit hat meine Begeisterung für den Himalaya geweckt und die zahlreichen Exkursionen haben mir die von Hochgebirgen ausgehende Faszination gezeigt.

Weiterhin danke ich Prof. Dr. Jürgen Böhner für die Übernahme der Zweitkorrektur dieser Arbeit.

Der erste Stein im Fundament dieser Arbeit ist von Caroline Stolter gelegt worden. Ich möchte mich an dieser Stelle ganz herzlich für ihren Zuspruch bedanken, der mir das Selbstvertrauen gegeben hat, mit dieser Arbeit anzufangen. Der Wechsel auf die andere Straßenseite brachte nicht nur immer einen Perspektivwechsel, sondern auch zu jeder Zeit einen Motivationsschub.

Die ökologische Promotions-Nische einer Doktorandin ist die Gesamtheit aller abiotischen und biotischen Umweltfaktoren, die eine Doktorandin zum Leben braucht. Diese ökologische Promotions-Nische ist kein Raum, sondern ein Wirkungsgefüge. Die zugrundeliegenden Faktoren, die zum Gelingen dieser Arbeit von Bedeutung waren, lassen sich eindeutig identifizieren. Das Zusammenspiel und die Wechselwirkungen der verschiedenen Elemente im Ökosystem Geomatikum gaben den Rahmen für die optimale Erfüllung aller Dissertationsfunktionen vor. Die Nahrungsgrundlage lieferten schätzungsweise eine Tonne Bonbons und hektoliterweise Kaffee. Die wichtigste biotische Komponente waren Franzi, Niels und Jo, mit denen ich so einige unvergesslichen Erlebnisse inner- und außerhalb des täglichen Aktionsraumes verbracht habe. Desweiteren haben Jule, Vanessa, Tobi, Elke, und Marion einen erheblichen Einfluss auf die Habitatqualität im 8. Stock gehabt. Niels, Birgit und Alina haben für eine entspannte Atmosphäre im Büro-Mikrokosmos gesorgt. Wissenschaftliche Interaktionen mit Lars, Jens und Benni haben mir besonders im Hinblick auf Statistikfragen und Publikationsstrategien geholfen.

Ich bedanke mich auch bei meinen Eltern, die mir die Freiheit gelassen haben, meine Träume zu verfolgen. Bewusst und unbewusst haben meine Geschwister Anna, Johannes und Matthias einen erheblichen Beitrag zum Gelingen dieser Arbeit geleistet.

Zu guter Letzt gilt mein Dank Henny, dem wichtigste Menschen in meinem Leben. Vielen Dank für deine Unterstützung und Geduld, die du in den letzten fünf Jahren aufgebracht hast, als ich mich entschlossen hatte, diesen Weg einzuschlagen.Zusammen mit Bubu und Erbse sorgst du für die nötige Gelassenheit.

Allen genannten sei herzlichst gedankt, ihr habt meine Zeit hier zu einer unvergleichlichen Erfahrung gemacht und schlussendlich zum Fertigstellen meines Opus magnum einen erheblichen Beitrag geleistet.

This thesis consists of three publications, which are provided in the Appendix.

To increase the readability of the thesis, all figures and tables have been renumbered in a consecutive manner.

One section containing all supplementary material is given at the end of the thesis.

Abstract

Mountains are fascinating habitats, characterized by steep ecological vertical gradients and corresponding altitudinal vegetation zonation. Alpine treelines as upper boundaries of more or less contiguous tree stands are the most conspicuous vegetation limits; they have always attracted great research interest. Globally, alpine treeline elevations in mountains are caused by heat deficiency. At landscape and local scales, however, multiple interactions of influencing factors and mechanisms determine treeline position, spatial pattern, and dynamics.

In the course of climate change, it is postulated that treelines will shift to higher elevations. In order to be able to quantify potential shifts, an analysis of the underlying factors and a correct modelling of the treeline ecotone under current climatic conditions are of great importance. For this purpose, statistical models are used to calculate the ecological niche of species based on climatic factors. These models serve as a baseline for models that project the distribution under future climatic conditions.

The Himalayas are the largest mountain range in the world, yet they are often underrepresented in scientific literature. This holds particularly true in relation to modelling studies. Modelling treeline species in remote high altitude regions faces several challenges, especially the availability of occurrence data and high quality environmental variables.

This research aimed at modelling the ecological niche of the Himalayan birch (*Betula utilis*) under present climatic conditions in the Himalayan mountain system. *B. utilis* represents a favourable target species for modelling studies, since it is widespread as a treeline-forming species along the entire Himalayan arch. Due to less distinctive habitat requirements and high adaptation potential, it is gaining importance as a pioneer tree species for possible succession developments at treelines under future climate conditions.

In all three parts of this work, generalized linear models (GLMs) were used to model the ecological niche of *B. utilis*. By evaluating the models on the basis of several quality criteria, statistically valid results were obtained. In order to ensure the transferability of the results to other studies, primarily freely available data were used (except some of the climate data in Article I). In a synergistic approach, a detailed study of the underlying climatic, topographical and plant phenological factors was undertaken in order to model the potential and the actual distribution of the focal species.

In the first part, the climatic factors influencing the distribution of *B. utilis* were determined, followed by modelling the potential distribution under present climatic conditions. In order to classify the results, the modelled distribution was compared with the vegetation map of Schweinfurth (1957), and the deviations were discussed.

In the second part, the ecological niche of *B. utilis* was modelled based on two different climate data sets. In order to investigate the impact of each climate data set we compared model accuracy and prediction of the modelled niche of *B. utilis*.

Furthermore, with regard to possible distortion of the modelled distributional areas, the importance of analysing climate input variables was highlighted. Biased results of current distributions lead to flawed distributions under future scenarios, which may have to far-reaching consequences for the derived climate and nature conservation implications.

In the third part, the focus was on exploring the potential of remote sensing data for modelling the current distribution of *B. utilis*. Topographical and plant phenological data were used to model the realised niche and to identify the underlying factors. Another focus was evaluating an exclusively remote sensing-based approach. For future studies, remote sensing data can provide long-term, high-resolution, and species-adapted variables to model current distributions.

The present research is the first of its kind in the Himalayan region, and it is characterised by its large-scale and comprehensive nature. To date, comparable studies dealing with modelling the ecological niche of *B. utilis* under present climatic conditions along the entire Himalayan arch were not conducted. The present results provide a new starting point for further investigations aimed at modelling the distribution of the species under past or future climate scenarios. Simultaneously, the presented approaches can also be transferred to other treeline species in high mountains.

Zusammenfassung

Gebirge stellen faszinierende Lebensräume dar, die durch steile ökologische Vertikalgradienten und eine entsprechende Vegetationshöhenstufung charakterisiert sind. Alpine Waldgrenzen stellen die auffälligsten Höhengrenzen dar, da sie die obere Grenze mehr oder weniger geschlossener Baumbestände markieren; sie haben seit jeher großes Forschungsinteresse geweckt. Global betrachtet wird die Höhenlage der Waldgrenze in Gebirgen durch Wärmemangel bedingt, lokal hängen Lage, räumliche Strukturen und Dynamik der Waldgrenze von einer Vielzahl sich wechselseitig beeinflussender Faktoren und Prozesse ab.

Im Zuge des Klimawandels wird postuliert, dass sich die Waldgrenze in höhere Lagen verschieben wird. Um diese möglichen Arealverschiebungen quantifizieren zu können, ist eine Analyse der zugrundeliegenden Faktoren und eine korrekte Modellierung des Waldgrenzökotons unter gegenwärtigen klimatischen Bedingungen von großer Bedeutung. Hierfür kann mit Hilfe von statischen Modellen die ökologische Nische von Arten anhand von Umweltvariablen berechnet werden. Diese Ergebnisse dienen dann als Ausgangsbasis für Studien, welche die Verbreitung unter Klimaszenarien der Vergangenheit oder der Zukunft modellieren.

Der Himalaya ist zwar das größte Gebirge der Erde, es ist in der wissenschaftlichen Literatur jedoch oftmals unterrepräsentiert. Dies gilt vor allem in Bezug auf Modellierungsstudien. Grundsätzlich ist die Modellierung von Baumarten an der Waldgrenze mit verschiedenen Herausforderungen konfrontiert, dies bezieht sich vor allem auf die Verfügbarkeit von Vorkommensdaten und qualitativ hochwertigen Umweltvariablen.

Das Ziel dieser Arbeit war die Modellierung der ökologischen Nische der Himalaya-Birke (*Betula utilis*) unter gegenwärtigen klimatischen Bedingungen im Himalaya. *B. utilis* stellt eine für Modellierungen vorteilhafte Zielart dar, da sie als waldgrenzbildende Art entlang des gesamten Himalayabogens verbreitet ist. Aufgrund geringer Standortansprüche und hohem Anpassungspotenzial gewinnt sie als Pionierbaumart im Zuge möglicher Sukzessionsentwicklungen unter zukünftigen klimatischen Verhältnissen an der Waldgrenze an Bedeutung.

In den dieser Arbeit zugrunde liegenden publizierten Studien wurden verallgemeinerte lineare Modelle zur Modellierung der ökologischen Nische von *B. utilis* verwendet. Durch die Evaluation der Modelle anhand von mehreren Gütekriterien wurden Ergebnisse von statistischer Validität erzielt. Um die Übertragbarkeit der Ergebnisse auf andere Gebiete gewährleisten zu können, wurden weitestgehend frei verfügbare Daten verwendet. In einem synergetischen Ansatz wurde eine detaillierte Untersuchung der zugrunde liegenden klimatischen, topographischen und pflanzenphänologischen Faktoren vorgenommen, um sowohl die potenzielle als auch die aktuelle Verbreitung der Art zu modellieren.

Im ersten Teil dieser Arbeit wurden die klimatischen Faktoren bestimmt, die für die Verbreitung von *B. utilis* maßgeblich sind. Es folgte die Modellierung der potenziellen Nische unter gegenwärtigen klimatischen Bedingungen. Um die Ergebnisse einordnen zu können, wurde die modellierte Verbreitung mit der Vegetationskarte von Schweinfurth (1957) verglichen und die Abweichungen erörtert.

Im zweiten Teil wurden Modelle der ökologischen Nische von *B. utilis* basierend auf zwei verschiedenen Klimadatensätzen erstellt. Anschließend wurden Modellgüte und – vorhersage der modellierten ökologischen Nische von *B. utilis* miteinander verglichen. Desweiteren wurde, in Bezug auf mögliche Verzerrungen der modellierten Verbreitungsgebiete, die Bedeutung der Analyse der zugrunde liegenden klimatischen Eingangsvariablen untersucht. Verzerrte Ergebnisse der gegenwärtigen Verbreitung führen zu fehlerhaften Verbreitungen unter Zukunftsszenarien und können weitreichende Folgen für den daraus abgeleiteten Klima- und Naturschutz bedeuten.

Im dritten Teil lag der Fokus auf der Erkundung des Potenzials von Fernerkundungsdaten zur Modellierung der aktuellen Verbreitung von *B. utilis*. Hierfür wurden topographische und pflanzenphänologische Daten herangezogen, um die realistische Nische zu modellieren und die zugrunde liegenden Faktoren zu identifizieren. Ein weiterer Schwerpunkt war die Evaluation eines ausschließlich fernerkundungsbasierten Ansatzes. Für zukünftige Studien können Fern-erkundungsdaten langjährige, hochaufgelöste und auf die Art angepasste Variablen zur Modellierung der gegenwärtigen Verbreitung liefern.

Die vorliegende Arbeit ist mit allen drei Teilen die erste ihrer Art für die Himalaya-Region und zeichnet sich durch ihre großflächige und umfassende Perspektive aus. So gibt es bis dato keine vergleichbaren Studien, die sich mit der Modellierung der ökologischen Nische von *B. utilis* unter gegenwärtigen klimatischen Bedingungen entlang des gesamten Himalayabogens beschäftigen. Die vorliegenden Ergebnisse stellen einen neuen Ausgangspunkt für weitere Untersuchungen dar, die die Modellierung der Verbreitung von *B. utilis* unter vergangenen oder zukünftigen Klimaszenarien zum Ziel haben. Ebenso können die vorgestellten Ansätze auch auf andere Arten in Hochgebirgen übertragen werden.

Outline

List of Figures

List of Tables

List of Supplementary Material

List of Abbreviations

A	Abiotic
AIC	Akaike Information Criterion
AUC	Area Under the Curve
B	Biotic
BAM	Biotic-Abiotic-Mobility diagram
bio7	Temperature Annual Range (bio5–bio6)
bio8	Mean Temperature of Wettest Quarter
bio15	Precipitation Seasonality (Coefficient of Variation)
bio19	Precipitation of Coldest Quarter
CHELSA	Climate based on Karger *et al.*, 2016 (Article II)
CLIMATE	Climate based on Karger *et al.*, 2016 (Article III)
DEM	Digital elevation model
E	Environmental space
EA	Fundamental niche in environmental space
EI	Invadable niche in environmental space
ENM	Ecological Niche Model
EO	Realised niche in environmental space
EP	Biotically reduced niche in environmental space
EVI	Enhanced vegetation index
G	Geographic space
GA	Abiotic suitable conditions in geographic space
GDATA	Species presence and absence data in geographic space
GI	Invadable distributional area in geographic space
GLM	Generalized Linear Model
GO	Occupied distributional area in geographic space
GP	Potentially distributional area in geographic space
Green_max	Onset Greenness Maximum
Green_min	Onset Greenness Minimum
IPCC	International Panel on Climate Change
LST	MODIS Land Surface Temperature data derived from Bechtel, 2015
M	Movement
MAST	Mean annual surface temperature
NCSA	Number of clear-sky acquisitions
PHENO	Phenological traits derived from MODIS LP DAAC, 2012
prec_mam	Average precipitation March, April, May
SDM	Species Distribution Model
THETA	Phase shift relative to spring equinox on the Northern hemisphere
TOPO	Digital Elevation Model based on USGS, 2004
TSS	True Skill Statistics
WORLDCLIM	Climate based on Hijmans *et al.*, 2005
YAST	Mean annual amplitude of land surface temperature

1. Introduction

High mountains feature fascinating vegetation zonations and host considerable biodiversity and endemism. With over 10.000 alpine plant species on a comparatively small part of the Earth' surface (< 3 %), the biodiversity is 3.7 % higher than the global average (Körner, 2003). High mountains are characterized inter alia by recent or past glaciations, geomorphological processes such as frost weathering and solifluction, raising to an elevation above the climatic treeline, and complex altitudinal vegetation zonations (Troll, 1975; Jentsch & Liedtke, 1980).

High elevation treelines can be considered one of the most conspicuous natural vegetation boundaries, and the investigation of underlying factors for treeline formation has generated considerable research interest for a long time (Holtmeier, 2009). The causes and underlying ecological drivers regarding natural treelines are manifold and cross-comparison of studies is often impeded because researchers do not adhere to a universally accepted terminology in defining spatial dimensions of treeline environments (Elliot, 2017). The term 'treeline' is defined as an intentional line connecting the highest patches of tree stands composed of trees at least 3 m in height (Körner, 2012). Below the treeline, the timberline marks the upper distributional limit of closed forests. Above the treeline, by contrast, the tree species line represents the uppermost occurrences of tree species individuals, either seedlings or trees characterised by crippled growth habit, also described as krummholz (Körner, 2012).

Although the terms are described as lines, the demarcation to other vegetation formations may not always appear to be abrupt or even clearly delimited, resulting in a more or less wide treeline ecotone. Ecotones are the result of environmental gradients that determine physiological and reproductive range limits of species, resulting in separate spatial ranges of adjacent vegetation formations which may overlap. Steep environmental gradients result in more constrained ecotones, and a gradual gradient in broader ecotones. Compared to arctic treelines, alpine treelines show rather narrow ecotones due to steep slopes (Elliot, 2017). Depending on the geographical position of the treeline and the tree species, ecological characteristics and requirements may differ at temporal and spatial scales.

Three types of treelines can be distinguished: orographic, climatic and anthropogenic treelines. For orographic treelines, the surrounding topography and geomorphology governs the treeline position, whereas for climatic treelines, climate conditions determine the treeline position, tree establishment and potential treeline advance. At a global scale, the position of natural treelines is mainly controlled by heat deficiency, or in other words, a mean growing season air temperature of 5.5 to 7.0 °C and a mean growing season soil temperature of 6.4±0.7 °C (Körner, 2007; Körner & Paulsen, 2004), whereas at smaller spatial scales, other abiotic and biotic factors such as precipitation, inter-annual climatic variability, wind, radiation, snow, soils, inter-specific competition, insect infestations as well as historical factors influence spatial patterns and physiognomy of treelines (Holtmeier, 2009; Holtmeier & Broll, 2005). At almost all treelines worldwide, human impacts (fire, logging or pastoralism) have influenced treeline positions or even altered species compositions (anthropogenic treeline; Holtmeier, 2009). Although the uppermost occurrences of trees are mainly attributed to cold temperatures, trees may not reach their distributional limit due to interactions of multiple specific site factors, leading to a decline in subalpine forests and thereafter to modified or depressed treeline positions (Schickhoff *et al.*, 2016a).

Since treeline elevations are characterised by low temperatures, high elevation climatic treelines can be considered sensitive indicators of past and recent climate change and variability at local and global scales (Kullman, 1998; Holtmeier, 2009; Smith *et al.,* 2009; Körner, 2012). During recent decades, investigation of climate change-driven treeline dynamics has generated considerable research interest, and results have been widely reported from various treelines around the world (e.g., Randin *et al.,* 2009; Harsch *et al.,* 2009; Paulsen & Körner, 2014, Schibalski *et al.,* 2014; Schickhoff *et al.,* 2015, 2016). Since high mountain environments are subjected to above-average warming rates, treeline dynamics under future climate change scenarios are of particular interest in this respect (Schickhoff, 2011; IPCC, 2014). Global average mean temperature has increased by +0.85 °C between 1880 and 2012 (IPCC, 2013). For the Himalayan mountain system, it is hypothesised that the climate is changing at a faster rate than the global average (Shrestha *et al.,* 2012; Schickhoff *et al.*, 2016b). Since 1989, temperature increases during winter months of up to +0.8 °C per decade have been determined in the eastern Himalayas (Gerlitz *et al.,* 2014), whereas pre-monsoon season temperature increases of up to +1.0 °C per decade have been found for higher elevations along the entire Himalayan arc (Schickhoff *et al.,* 2015). Shrestha *et al.* (2012) found an extended growing season by 4.7 days at average during a 25-year period, with seasonal and regional variations. For the previous century, decreases in annual precipitation (up to 20%) have been identified for the western, but not for the eastern Himalayas (Jain *et al.,* 2013; Schickhoff *et al.*, 2016b). It is to be expected that an increase in temperature and coherently evapotranspiration, combined with a decrease in precipitation, results in amplified drought stress, primarily in the pre-monsoon season (Schickhoff *et al.,* 2015).

One popular hypothesis is that, due to changing regional climatic conditions, ranges of subalpine and alpine species as well as treelines shift upwards along altitudinal, thermally defined gradients (Gottfried *et al.,* 2012; Pauli *et al.,* 2012; Wieser *et al.,* 2014). Furthermore, it is postulated that species respond by altered seasonal phenology (Hughes, 2000; Smith *et al.,* 2012; Anadon-Rosell *et al.,* 2014; Ernakovich *et al.,* 2014; Hart *et al.,* 2014), while some species are threatened by extinction or are already extinct (Parmesan, 2006; Pauli *et al.,* 2012; Alexander *et al.,* 2015; Cotto *et al.,* 2017). Treelines are regarded as particularly responsive to changing temperature regimes, and initial effects of future climate-induced range shifts are expected for species in high altitude treeline ecotones.

Modelling the distributional range of treeline species and predicting changes under future climate scenarios has become an increasingly applied component in investigations of high altitude treelines (e.g., Dullinger *et al.*, 2004; Thuiller *et al.*, 2005; Parolo *et al.*, 2008). In contrast to other mountains of the world, the Himalayan region has largely been neglected and is clearly under-represented in scientific literature on climate change-induced species range shifts (Schickhoff, 2005; Miehe *et al.,* 2007; Telwala *et al.,* 2013, Dutta *et al.,* 2014; Schickhoff *et al.,* 2015). The number of studies which aim to predict species' distribution or forecast species range shifts under climate change scenarios is limited (e.g., Kumar, 2012; Menon *et al.*, 2012; Ranjitkar *et al.*, 2014 on *Rhododendron* spp.; Menon *et al.,* 2010 on *Gymnocladus assamicus*; Jaryan *et al.,* 2013 on *Sapium sebiferum*; Gajurel *et al.,* 2014 on *Taxus wallichiana*; Ranjitkar *et al.*, 2014 on *Oxybaphus himalaicus* and *Boerhavia diffusa;* Shrestha *et al.,* 2014 on *Ophiocordyceps chinensis*).

Moreover, high altitude treeline studies in the Himalayas have investigated coniferous tree species (e.g., *Abies*, *Juniperus*, *Pinus*) and broadleaved evergreen tree species (e.g., *Rhododendron*) while deciduous tree species (e.g., *Betula*) have remained largely out of focus. Presently, only few researchers have addressed the problem of modelling distribution ranges of deciduous treeline species[1] in the Himalayas, whereas conducted studies on *B. utilis* were local studies (e.g., Huo *et al.*, 2010: SW China; Singh *et al.*, 2013: Indian Himalaya, Uttarakhand; Wang *et al.,* 2017: Tibetan Plateau).

The genus *Betula* is known to inhabit a considerably wide ecological niche in the northern hemisphere and can be found in high altitude and high latitude treeline ecotones (Truong *et al.,* 2007; Holtmeier, 2009; Speed *et al.,* 2011). Alpine treelines with *Betula* as conspicuous treeline species can be found in Russia in the Urals (*B. litwinowii*; Hansen *et al.,* 2018), in Kamchatka (*B. ermanii*; Krestov *et al.,* 2008) and in Japan (*B. ermanii*, *B. platyphylla*, *B. maximowicziana*; Koike *et al.,* 2003; Yasaka, 2005). Our target species, *Betula utilis*, is widespread in Himalayan alpine treelines (Schickhoff, 2005; Ashburner & McAllister, 2013), and some authors consider *B. utilis* as an indicator species for climate-driven treeline dynamics (e.g., Liang *et al.,* 2014). *B. utilis* shows many characteristics of a pioneer species, for instance, a high degree of adaptability to altered environmental and climatic conditions. *B. utilis* is able to

---

[1] For reasons of readability, the terms 'treeline' and 'treeline ecotone' will be used synonymous in the presented thesis.

rejuvenate readily under changed light and soil conditions, and facilitates natural reforestation processes and forest edge closure, since it promotes humus accumulation in the course of natural succession.

We selected the treeline-forming species *B. utilis* as a target species because a) underlying environmental factors of the species distribution have not adequately been described, and b) improved accuracy in modelling the current distribution is a precondition for more precisely modelling potential range expansions of treelines under climate change conditions (Schickhoff *et al.,* 2015). The latter applies in particular to a pioneer species such as *B. utilis*, characterized by high adaptability to changing environments. To date, modelling the ecological niche of *B. utilis* covering the entire Himalayan mountain region has remained a major research deficit.

The following sections deal with the theory behind ecological niche models (Chapter 2) and associated challenges and limitations of these models with regard to modelling species in high-altitude ecosystems (Chapter 3). In Chapter 4 information on the target species *B. utilis* is provided. The research objectives of the three articles are outlined in Chapter 5, followed by the methodological approaches in Chapter 6. The cumulative PhD thesis is composed of three articles and the respective abstracts can be found in Chapter 7 (for complete articles please refer to the Appendix section). A synthesis combining the results and interpretation of the respective articles is provided (Chapter 8), followed by an outlook for future studies. Finally, in Chapter 9 the conclusion section is presented.

2. Ecological niche models

Using ecological niche models (ENMs) or species distribution models (SDMs) requires analysing the underlying environmental factors of species distributions, which, in turn, constitutes a basic requirement for understanding current, past and future species distribution ranges. Although ENMs and SDMs are used synonymously in some modelling studies (Peterson & Soberón, 2012), they differ in their approach of the question to be answered. In contrast to SDMs, which aim to characterise the species' occupied distribution (*sensu stricto*), ENMs investigate relationships between known species occurrences and environmental variables in a particular region of interest (Peterson & Soberón, 2012). Hereinafter, the concept of ENMs is applied.

Initially, the term 'niche' was coined by Joseph Grinnell (1917) to refer to the relationship between the ecological properties of a species and its distribution. He described the niche as a region featuring environmental site conditions in which the species can survive and successfully reproduce (Grinnellian niche). Subsequently, Charles S. Elton introduced the term 'niche' (1927) in a similar sense, but focussing primarily on the role of the species in a biocoenosis and its biotic interactions (Eltonian niche).

The term 'ecological niche' dates back to the definition of G. Evelyn Hutchinson (1957). Although his concept expands the ideas of Grinnell and Elton, the main difference is that the niche is not a rigid characteristic of the environment, but rather a characteristic of the species itself.

Today, the conceptual framework behind ecological niche models (ENMs) is the conceptual BAM (Biotic-Abiotic-Movement) framework (Soberón & Peterson, 2005; 2011; Soberón & Nakamura, 2009) (Fig. 1). Although this concept is static, it provides the theoretical background for questions to be answered by means of ENMs. It consists of the region **A** in geographic space **G** where abiotically suitable conditions (e.g., climate and soils), which secure the species' survival and growth, can be found (i.e., $\mathbf{G}_{A}$ the existing fundamental niche). The circle **B** represents the region where biotic conditions enable a stable population of the species, attributed to Eltonian factors. The region **M** represents geographic regions that have been accessible to dispersal or colonisation by the species over a certain period of time (Soberón & Nakamura, 2009).

The intersection $\mathbf{G}_{O} = \mathbf{A} \cap \mathbf{B} \cap \mathbf{M}$ represents the area actually occupied by the species[2] (i.e., the occupied or realised niche). $\mathbf{G}_{I} = \mathbf{A} \cap \mathbf{B} \cap \mathbf{M}^{c}$ ($\mathbf{M}^{c}$ is the complement of **M**) represents a potentially inhabitable area with appropriate abiotic and biotic conditions, but which remains uninhabited (i.e., the invadable niche) (Soberón & Nakamura, 2009). The union of occupied $\mathbf{G}_{O}$ and invadable $\mathbf{G}_{I}$ can be defined as $\mathbf{G}_{P} = \mathbf{G}_{O} \cup \mathbf{G}_{I}$ and represents the potential distributional area[3] (i.e., biotically reduced niche) (Gaston, 2003; Peterson *et al.,* 2011).

---

[2] For reasons of readability, the term 'actual distribution' will be used when referring to the area that the species actually inhabits.

[3] For reasons of readability, the term 'potential distribution' will be used when referring to area that the species can potentially inhabit.

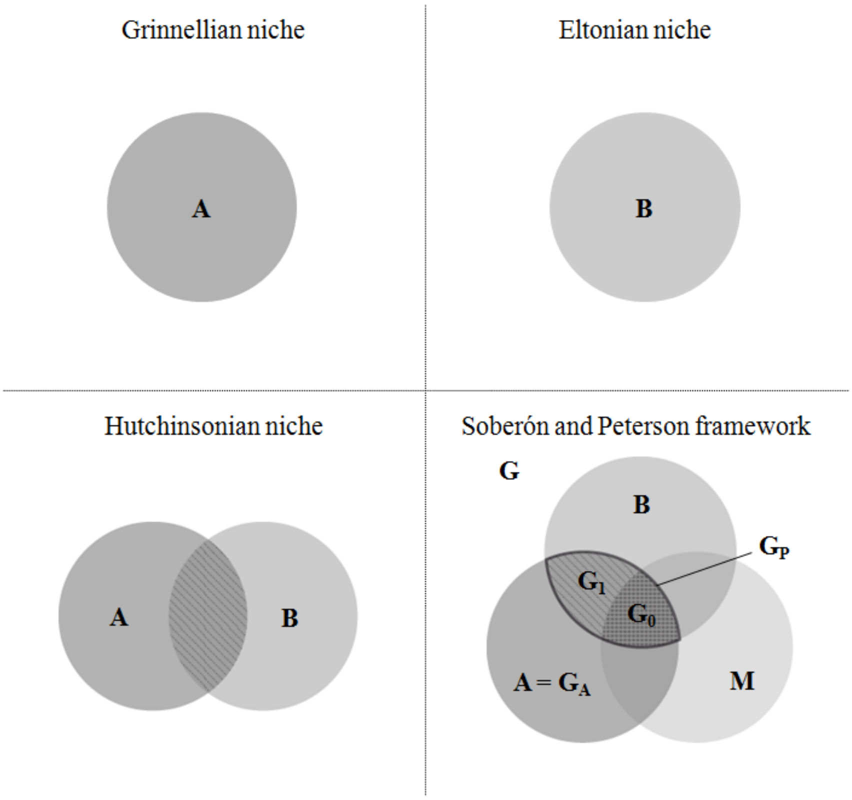


Fig. 1: Overview of niche concepts. Grinnellian niche: **A** represents all abiotic factors (e.g., climate and soils) determining the species occurrences; Eltonian niche: **B** represents all biotic factors (e.g., interactions and competition) determining the species occurrences; Hutchinsonian niche: **A** represents all abiotic factors limiting the species presence (fundamental niche), **B** represents all biotic interactions limiting the species presence and the intersection of **A** and **B** (hatched) represents the realised niche; Soberón and Peterson framework: BAM diagram consisting of **B** (biotic), **A** (abiotic) and **M** (movement) factors in **G** (geographic space) at time **t**. G_A represents abiotically suitable conditions (i.e., the fundamental niche). The intersections of **A** and **B**, $\mathbf{G_I}$ (hatched) represents the invadable distributional area, whereas only a small proportion, the intersecting area with **M**, $\mathbf{G_0}$ (checkered) represents the occupied distributional area (i.e., the realised niche) due to movement limitations (e.g., dispersal strategies or barriers). The union of $\mathbf{G_0}$ and $\mathbf{G_I}$ represents the potentially distributional area $\mathbf{G_P}$ (i.e., biotically reduced niche) of the species (bold line). Concept is based on Peterson *et al.* (2011) and modified after Escobar and Craft (2016).

In order to describe the species distributions statistically, mechanistic and correlative model approaches are frequently used. In the mechanistic approach, the geographic expression of the species fundamental niche **A** is defined using physiological experiments or physical modelling of responses of individuals to temperature, humidity and other physical parameters without consideration of **B** or **M** (Sóberon and Peterson, 2005). The correlative approach defines habitats correlatively (Kearney, 2006). Here the underlying concept is the assumption that the current distribution of the species serves as an effective indicator of ecological requirements (Kearney & Porter, 2009). Correlative model outputs express less than the full fundamental niche **A**, since effects of **B** and **M** are frequently not distinguishable (Peterson *et al.,* 2015). This presumed disadvantage constitutes an advantage at the same time, because effects of numerous causal factors can be incorporated in a single model (Peterson *et al.,* 2015).

The basic principle of modelling ecological niches refers back to the Hutchinson's Duality (Colwell & Rangel 2009), which encompasses the complex linkage of niches (environmental space: $\mathbf{E_A}$, $\mathbf{E_O}$, $\mathbf{E_I}$, or $\mathbf{E_P}$) and spatial distributions (geographic space: $\mathbf{G_A}$, $\mathbf{G_O}$, $\mathbf{G_I}$, or $\mathbf{G_P}$) (Peterson *et al.*, 2011). The species distribution is modelled using abiotic and/or biotic variables with the aim of characterising suitable habitat conditions of the targeted species in accordance with the respective research question.

In the following the workflow of ENMs after Hirzel *et al.* (2002) and Peterson *et al.*, (2011) is summarized (Fig. 2).

In step 1 two types of input data are required: occurrence data and environmental variables. The general approach is to link species occurrences with climatic and topographic variables to estimate the species distribution range, since habitat suitability is considerably influenced by the prevailing climate (Pearson & Dawson, 2003). The study area is represented by raster grids with an extent **G** and specific grid cell size and the dependent variable is the distribution of the species ($\mathbf{G_O}$, $\mathbf{G_P}$, or $\mathbf{G_A}$), concluded on occurrence records and absences $\mathbf{G_{data}}$ (Peterson *et al.*, 2011).

In step 2 input variables are compiled to characterize each cell of the study area in environmental space **E**. With a model algorithm the function $\mu(\mathbf{G_{data}}, \mathbf{E})$ is generated, that characterizes the distribution of the species in terms of the environmental variables, to indicate the degree to which each cell in **G** is suitable for the species. Calibrating the model ensures that the algorithm provides valid results (Peterson *et al.*, 2011).

Step 3 finally comprises the mapping of the prediction in **G** and the evaluation of prediction accuracy on independent data. Depending on the ability of input variables to depict ecological reality (i.e., configuration of the BAM diagram), model results map $\mathbf{G_A}$ (abiotically suitable area), $\mathbf{G_P}$ (potentially occupied area) or even $\mathbf{G_O}$ (occupied distributional area) (Peterson *et al.*, 2011). In order to interpret model predictions, predictive performance must be evaluated. Ideally, data used for evaluation would be collected independently, however, data-split approaches are usually applied to generate subsets for calibrating (training) and evaluation (testing) the model (Peterson *et al.*, 2011).

Step 4 is optional and refers back to the research question, whether modelled niche conditions are intended to be predicted across space and time. In contrast to SDMs, ENMs can be projected in geographic space and time, identifying consistent areas with suitable environmental conditions for the species and assessing distributional changes under climate change (Araújo & Guisan, 2006; Peterson & Soberón, 2012).

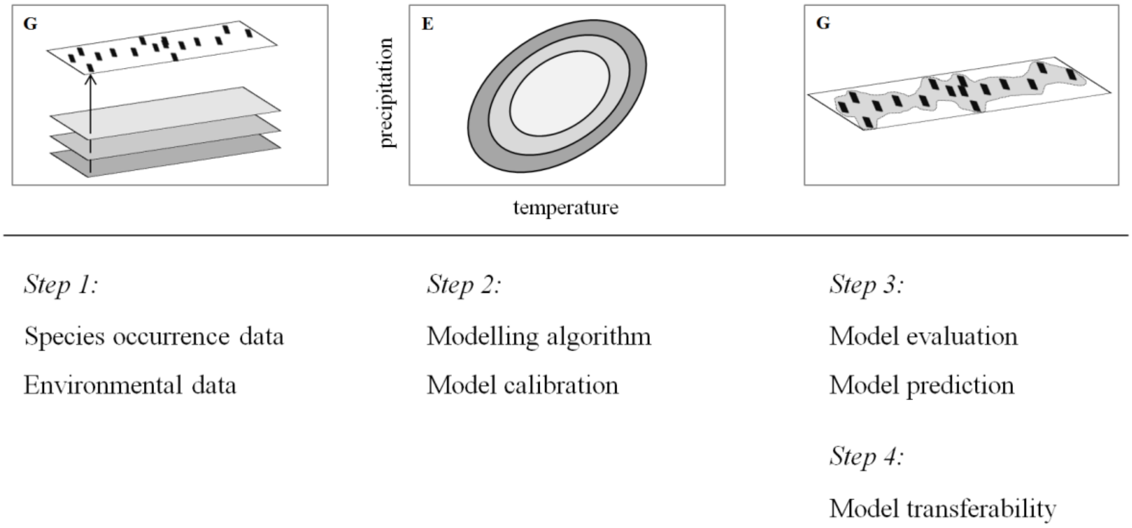


Fig. 2: Workflow of ecological niche models. Ecological niche models consist of 4 steps: 1) input data in geographical space, 2) modelling in environmental space, 3) model evaluation and model prediction in geographic space and 4) model transferability across geographic space or time (concept based on Peterson *et al.,* 2011).

3. Challenges and limitations of ecological niche models

During recent decades, ENMs have become an integral part of biogeography, ecology, evolution and conservation biology. In order to ensure that model results are processed correctly, it is worth noting that within the workflow of modelling species' niches, every step may be afflicted with errors and deficiencies. Investigations include quality and quantity of input data (Franklin, 2009), spatial structure of occurrence data (Araújo & Guisan, 2006; Loiselle *et al.,* 2008, Naimi *et al.,* 2013), choice of modelling algorithm (Elith *et al.,* 2006; Aráujo & New, 2007), multi-collinearity and spatial autocorrelation of predictors (Dirnböck & Dullinger, 2004; Dormann *et al.,* 2007, 2013; Braunisch *et al.,* 2013). Although extensive care is taken in selecting uncorrelated predictor variables, differences in model performance arising from available climate datasets remain largely out of focus in most studies.

The following chapter seeks to summarise challenges and limitations of ENMs in remote high altitude regions.

In order to investigate high altitude treeline dynamics, modelling techniques have become an indispensable method to predict species distributions under current climate conditions, to hindcast distributions under past climate conditions and to forecast changed distributional ranges under future climate scenarios (e.g., Dullinger *et al.,* 2004; Thuiller *et al.,* 2005; Parolo *et al.,* 2008, Schorr *et al.,* 2012). Inherently, the accuracy of models under climate change scenarios depends on their accuracy under current climate conditions, the importance of which should not be underestimated.

Modelling ecological niches across vast distribution ranges in remote, high mountain regions like the Himalayas remains a challenging task. Challenges include, first and foremost, the lack of species occurrence data and fine-scale environmental information of sufficiently high quality (i.e., environmental variables).

In many cases, presence-absence data are not available, and presence-only data are often derived from databases of natural history museums and herbaria, which contain occurrences sampled by numerous researchers and with different techniques, intensities and periods of time (Soberón & Peterson, 2004). Moreover, sampling records often cluster near the centre of climatic conditions under which the species occurs (Loiselle *et al.,* 2008). This leads to species documentations that do not cover the entire range of suitable habitat conditions for respective species. Such geographic sampling bias can lead to sampling bias in environmental space, which represents a major problem for modelling (Veloz, 2009; Anderson & Gonzalez, 2011). This holds particularly true for sampling treeline species in remote areas like the Himalayan region. Due to lower accessibility of treeline sites, the number of available sampling plots is sparse, which demonstrates a reciprocal effect on prediction performance (Aráujo *et al.,* 2005).

This also applies for environmental variables, as most modelling studies use climatic variables for predicting the distribution range of the species. In topo-graphically complex areas like the Himalayas, climate stations are quite rare due to rough terrain and complicated accessibility. This in turn leads to a poor data basis for calculating climate datasets compared to other more accessible terrains. In addition, climate stations are prevailingly located near settlements at lower elevations, where climatic conditions are most suitable for habitation, livestock farming and agriculture. Those climate stations are not representative of climatic conditions at higher elevations. Besides the data basis, the calculation method and bias correction also influences the quality of climate datasets.

The choice of environmental variables used to model species distributions may result in different distribution maps for the same species (Luoto *et al.,* 2007). If important local abiotic or biotic factors that influence the actual species' distribution (i.e., the realised niche) are disregarded, predictions will represent the potential distribution (i.e., the existing fundamental niche), since climate is not the exclusive factor determining habitat suitability (Thuiller, 2004). This will gain extraordinary importance, since an improved understanding and modelling capacity of the current distribution constitutes a precondition for modelling treeline dynamics under climate change scenarios.

4. Target species

The distribution range of the Himalayan birch (*Betula utilis*) extends across the Himalayan range from Afghanistan to southwest China, with the total elevational range extending from 2700 to 4500 m (Polunin & Stainton, 1984). *B. utilis* was selected as a study species due to its status as a principal broadleaved treeline species in the western and central Himalayan ranges (Fig. 3). This species primarily grows in shady locations on north-facing slopes. In the northwest Himalayas, *B. utilis* is widely distributed in the elevational range between 3100 and 3700 m, while the range shifts to higher altitudes towards the east Himalayas (mainly between 3800 and 4300 m). A higher dominance can be found in the western and central part of the mountain system. Over much of its range, *B. utilis* forms a narrow forest belt between evergreen coniferous forests (e.g., *Abies spectabilis*) below and an evergreen broadleaved krummholz belt (e.g., *Rhododendron campanulatum*) above (Schickhoff, 2005; Miehe *et al.,* 2015a). Pure birch stands with *Rhododendron campanulatum* and *Sorbus microphylla* in the understory are often found at the uppermost limit of subalpine forests (Schickhoff *et al.,* 2015).

Fig. 3: Photographs of *Betula* treelines at 3900 m, Manang, Annapurna, Nepal (Schickhoff, 2013).

5. Study objectives

The number of ecological modelling studies in the Himalayas is very limited, even more so the number of studies modelling deciduous treeline species like *Betula utilis* (Huo *et al.,* 2010; Singh *et al.*, 2013; Wang *et al.,* 2017). Previous work on *B. utilis* has focused primarily on dendroecology (Bhattacharyya *et al.,* 2006; Tenca & Carrer, 2010, Dawadi *et al.,* 2013; Gaire *et al.,* 2013; Liang *et al.,* 2014), remote sensing (Singh *et al.,* 2012, 2013; Rai *et al.,* 2013), plant physiology (Xu *et al.,* 2012; Shi *et al.,* 2008) and regeneration (Shrestha *et al.,* 2007).

Since modelling the potential distribution of *B. utilis* and investigating the underlying climatic determinants still represents a major research deficit, one objective of Article I was to model the potential distribution of *B. utilis* in the subalpine and alpine treeline ecotone of the Himalayas on a broad scale. We aimed at identifying key climatic constraints for the current distribution of *B. utilis*, with a special focus on the climatic variables' ability to reflect the climatic gradient from more continental western to more oceanic eastern regions of the Himalayan mountain range. Furthermore, the accuracy of the models was evaluated and the current model prediction compared to the distribution range delineated in the detailed vegetation map of Schweinfurth (1957). The Schweinfurth map summarises the 1950s' state of knowledge of Himalayan vegetation distribution. It is based on an extensive literature survey, and despite including some 'white patches', it is still the most detailed vegetation map of the entire Himalayan region. The following were the questions that Article I strives to answer:

- **What are key climatic constraints for the current distribution of *B. utilis* in the Himalayan region?**
- **Are there differences between the current model predictions and the distribution range delineated in the detailed vegetation map of Schweinfurth (1957)?**

Since underlying climatic determinants may be afflicted with errors, we compared two freely available climate datasets, namely, CHELSA (Karger *et al.,* 2016) and WORLDCLIM (Hijmans *et al.,* 2005), and evaluated the performance of ENMs (Article II). In contrast to multi-collinearity and spatial autocorrelation of predictors (Dirnböck & Dullinger, 2004; Dormann *et al.,* 2007, 2013; Braunisch *et al.,* 2013), differences in model performance arising from available climate datasets has remained largely out of focus in most studies (Soria-Auza *et al.,* 2010). Comparative studies that evaluate the performance of ENMs using different global climate datasets for modelling the potential distribution of Himalayan treeline tree species or other Himalayan vascular plant species do not exist. We compared input parameters and statistical computations methods (i.e., quasi-mechanistical statistical downscaling and interpolation) of the climate datasets with regard to temperature- and precipitation-related variables.

Furthermore, we investigated whether the two climate datasets showed discrepancies in model predictions. Additionally, we emphasized prior modelling examination of input climate predictor datasets when modelling ecological niches of species in remote, high altitude regions. The following were the questions that Article II strives to answer:

- **How do input parameters and computation methods of the climate datasets (CHELSA and WORLDCLIM) influence derived climatic variables in vast mountain systems like the Himalayas?**
- **Are there discrepancies in the prediction of the two climate datasets (CHELSA and WORLDCLIM) when modelling the potential distribution of *B. utilis* under current climate conditions?**

Modelling ecological niches and species distributions in remote, high mountain regions like the Himalayas constitutes a challenging task. Current studies in the field of plant distribution modelling in the Himalayan Mountains primarily use climatic variables to predict species distribution or to forecast species range shifts under climate change scenarios (e.g., Menon *et al.,* 2010, 2012; Kumar, 2012; Jaryan *et al.,* 2013; Singh *et al.,* 2013; Gajurel *et al.,* 2014; Ranjitkar *et al.,* 2014; Shrestha & Bawa, 2014; Schickhoff *et al.,* 2015; Manish *et al.,* 2016). Although reasonable results were obtained using solely climate for predicting the potential distribution of *B. utilis* (Articles I and II), the necessity arose for approximating the actual distribution of *B. utilis*. Therefore, in Article III, different remotely sensed variables were incorporated into the modelling approach, and we evaluated possible improvement to the niche model of *B. utilis* based solely on bioclimatic variables. In this pioneering study, we aimed to bridge the gap between the potential and actual distributions of *B. utilis* for the entire Himalayan mountain system. Furthermore, the potential of a pure remote sensing approach by substituting the modelled bioclimatic variables with remotely sensed land surface temperature data was also explored. The following were the questions that Article III strives to answer:

- **To what extent can niche models built solely on bioclimatic predictor variables be improved by adding different remotely sensed predictor variables?**
- **What is the potential of a purely remote sensing-based approach when modelling the ecological niche of *B. utilis*?**
- **Which variable combinations explain the current distribution of *B. utilis* in the Himalayan region?**

In particular, we investigated the suitability of various predictor sets, including modelled bioclimatic variables (Chelsa; Karger *et al.,* 2016), remotely sensed topography (USGS, 2004), remotely sensed phenological traits derived from MODIS Land Cover Dynamics data (LP DAAC, 2012), remotely sensed annual cycle parameters derived from MODIS Land Surface Temperature data (Bechtel, 2015) and their combinations.

6. Modelling the ecological niche of Betula utilis – a methodological overview

Modelling ecological niches of species contributes to understanding the ecology of species in their habitats. This concept is based on species occurrences and underlying factors characterising the habitat at different spatial and temporal scales. The results constitute a basic requirement for understanding current, past and future species distribution ranges. Since high mountain environments are subjected to above-average warming rates, shifts in the environmental niches of alpine treeline species under future climate change scenarios are of particular interest. The models' accuracy under climate change scenarios (i.e., the species' response to changing climate conditions) depends on the accuracy of models of the current environmental niche of the species. In Article I we investigated the underlying climatic factors for the current distribution of *Betula utilis* and compared differences in model accuracy of two different climatic input data sets (Article II).

It emerges that most studies focus on modelling the potential distribution of the species where favourable abiotic conditions can be found (Araújo & Guisan, 2006). Theoretically, species ranges in geographic space closely coincide with the species niche in environmental space, which can be attributed to prevailing environmental conditions under which the species can occur (Hutchinson, 1978). As these models are based on the theoretical assumption of an equilibrium between the distribution of the species and the prevailing environmental conditions. Resulting distributional ranges may be over- or underestimated, because climate is not the exclusive factor determining habitat suitability (Araújo & Pearson, 2005). Such models may differ considerably from the actual distribution, and special care should be taken to interpret the modelling results. As previously mentioned, modelling species' distributions in high altitude regions faces numerous challenges. The most important constraint is that data availability is often sparse due to poor terrain accessibility. Therefore the aim of Article III was to elucidate the potential of additional remotely sensed data (i.e., topography and phenological traits) for modelling the actual distribution of the focal species under current climatic conditions.

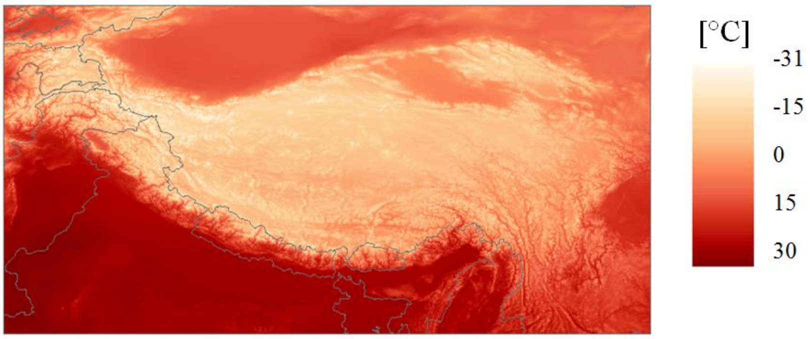
6.1 Study area

The Himalayan mountain range is located between the Indian subcontinent in the south and the Tibetan Highland in the north, and it extends from Afghanistan in the northwest (c. 36°N and 70°E) to Yunnan in the southeast (c. 26°N and 100°E). It is a vast mountain region covering an area of more than 1.000.000 km^2, with a length of c. 3000 km (Pakistan to SW China) and a maximum width of 400 km (Zurick & Pacheco, 2006).

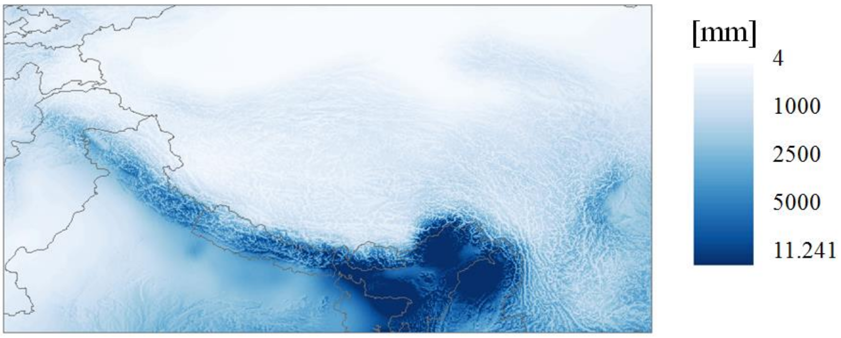
The Himalayan Mountains show a distinct three-dimensional geoecological differentiation, with a high variation of climate, rainfall, altitude and soils (Troll, 1972; Zurick & Pacheco*,* 2006; Miehe *et al.,* 2015b). Generally, the macroclimate is under influence of the monsoon, whereas at local and regional scales complex pattern of climatic, edaphic and biotic site conditions can be found (Schickhoff, 2005). The climate ranges from tropical in the Indian lowlands to permanent ice and snow at the highest elevations, and from more continental in the northwest to more oceanic in the southeast (Fig. 4).

As the climate primarily influences the position of the treeline, treeline elevation becomes evident along two gradients (Schickhoff, 2005). A NW-SE gradient relates to higher temperature sums at the same elevational level, whereas mean temperatures of the warmest month are higher at treelines in the northwestern part. Nevertheless, treelines in the NW are situated at lower elevations, due to a higher degree of continentality, because extreme winter cold, later snow melt, and shorter growing seasons overcompensate the advantage of higher summer temperatures (Schickhoff, 2005). A peripheral-central gradient ranging from the Himalayan south slope to the Great Himalayan range and the Tibetan plateau also relates to higher temperature sums. Increasing elevation of the treeline along this gradient is attributed to effects of continentality and mass-elevation (Schickhoff, 2005).

The amount of annual precipitation increases with increasing monsoonal influence from NW to SE along the southern front of the range (Schickhoff, 2005; Böhner *et al.,* 2015) (Fig. 4). The small-scale heterogeneity of habitats and site conditions supports a high level of diversity of species and communities.



Mean annual temperature



Annual precipitation sums

Fig. 4: Mean annual temperature and annual precipitation sums derived from CHELSA, Karger *et al.* (2016).

6.2 Species data collection

Currently, modelling studies are often based on presence-only species occurrence data instead of presence-absence data, which always constitute a source of uncertainty. Sometimes, areas have been unequally sampled or occurrence data have been inaccurately geo-referenced, resulting in occurrences of species with sampling bias. Such geographic sampling bias can lead to sampling bias in environmental space, which represents a major problem for modelling (Veloz, 2009; for the effects of sampling bias on model evaluation: Anderson & Gonzalez, 2011).

To avoid these pitfalls, presence-only occurrence data of *B. utilis* was gathered from three different sources. A total of 827 records were extracted from freely available satellite images (GoogleEarth™, ver. 7.1.1.1888, Google, 2015). Extractions from GoogleEarth have been demonstrated to be valuable in global treeline research (Paulsen and Körner, 2014; Irl *et al.,* 2016). These occurrence localities were validated through expert knowledge, obtained from numerous field visits in the Himalayan Mountains. Additionally, 215 geo-referenced records (1980–2016) were accessed via the Global Biodiversity Information Facility (gbif.org). Further, 202 records were added from a database compiled from a literature survey (Schickhoff, 2005; unpublished data).

Lowermost occurrences (e.g., in avalanche paths) were removed, since they do not represent the 'zonal' climatic conditions of the treeline birch belt. To reduce sampling bias and spatial auto-correlation, we kept only one occurrence point per grid cell (i.e., 1 x 1 km), resulting in 590 occurrences (Article I and II) and 1041 occurrences (Article III) for modelling the current distribution of *B. utilis* (Fig. 5). To the knowledge of the authors, the compiled occurrence dataset of *B. utilis* is the most comprehensive dataset available.

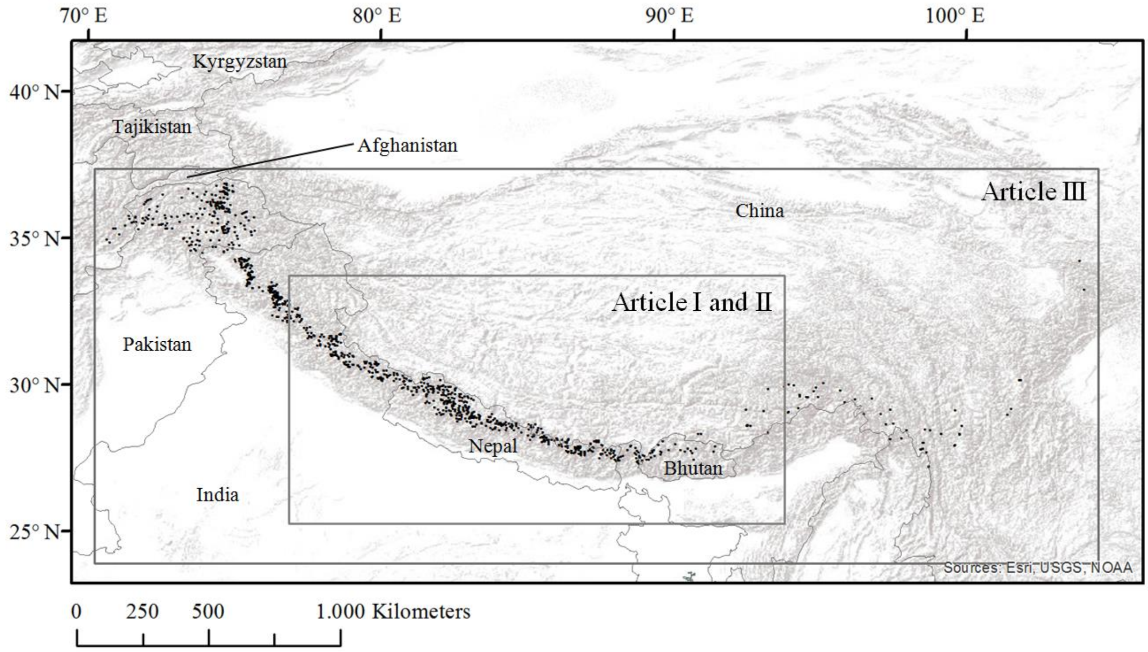


Fig. 5: Occurrences of *Betula utilis* along the Himalayan arc
(Article I and II N= 590; Article III N= 1041).

6.3 Environmental predictors

As previously mentioned performance of different predictor variable sets for modelling the ecological niche of *B. utilis* was evaluated. Although the predictor variable sets are described in detail in the three articles, a short overview is provided here.

Generally, all environmental predictors had a spatial resolution of 1 x 1 km and were long-time (climate) datasets. Based on gridded monthly fields of temperature and precipitation, 19 bioclimatic variables were generated, which are widely used in ecological niche modelling and represent annual characteristics (Hijmans *et al.,* 2005). In addition, the average precipitation of May and of March, April and May were calculated in order to account for potential pre-monsoon drought stress (Liang *et al.,* 2014; Schickhoff *et al.,* 2015, 2016).

Environmental predictors were chosen to reflect the species' physiological needs. We checked for multi-collinearity among the predictor variables using Spearman's rank correlation, since high collinearity might lead to low model performance and wrong interpretations (Dormann *et al.,* 2013). Only ecologically meaningful variables representing general patterns and annual variability were included for modelling the potential distribution of *B. utilis* (Table 1).

For Articles I and II, five climatic variables were included out of 24 potential predictor variables. For Article III, five climatic variables, two topographic variables, four land cover metrics and four land surface temperature variables were selected out of 40 potential predictor variables.

Table 1: Overview of the environmental predictor variables used for modelling the ecological niche of *Betula utilis*, for all three articles respectively.

			Article I	Article II		Article III
Input set	Label	Variable		CHELSA	WORLDCLIM	
Bioclimatic variables	bio7	Temperature Annual Range (bio5-bio6)	X	X	X	X
	bio8	Mean Temperature of Wettest Quarter	X	X	X	X
	bio15	Precipitation Seasonality (Coefficient of Variation)	X	X	X	X
	bio19	Precipitation of Coldest Quarter	X	X	X	X
	prec_mam	Average Precipitation March, April, May	X	X	X	X
Topography	Northness	Northness				X
	Slope	Slope angle				X
MODIS Land Cover Dynamics	Green_Inc	Onset Greenness Increase				X
	Green_Max	Onset Greenness Maximum				X
	Green_Dec	Onset Greenness Decrease				X
	EVI_Area	NBAR EVI Area				X
	Dym_QC	Dynamics QC				
MODIS Land Surface Temperature	MAST	Mean annual land surface temperature				X
	YAST	Mean annual amplitude of land surface temperature				X
	THETA	Phase shift relative to spring equinox on the Northern hemisphere				X
	RMSE	Inter-diurnal and inter-annual variability (Root Mean Squared Error of fit)				X
	NCSA	Number of clear-sky acquisitions				X

In Article I, we merged two climate datasets (one for temperature and one for precipitation, respectively), explicitly accounting for important topo-climatic processes, in order to examine the underlying factors and model the ecological niche of *B. utilis*. The monthly mean temperature data were derived from ERA-interim reanalysis (Gerlitz *et al.,* 2014), and the monthly precipitation sums were downloaded freely from the Chelsa climate dataset (Karger *et al.,* 2016) (Fig. 6). Furthermore, a digital elevation model was extracted from GLOBE (ngdc.noaa.gov; Global Land One-Kilometer Base Elevation).

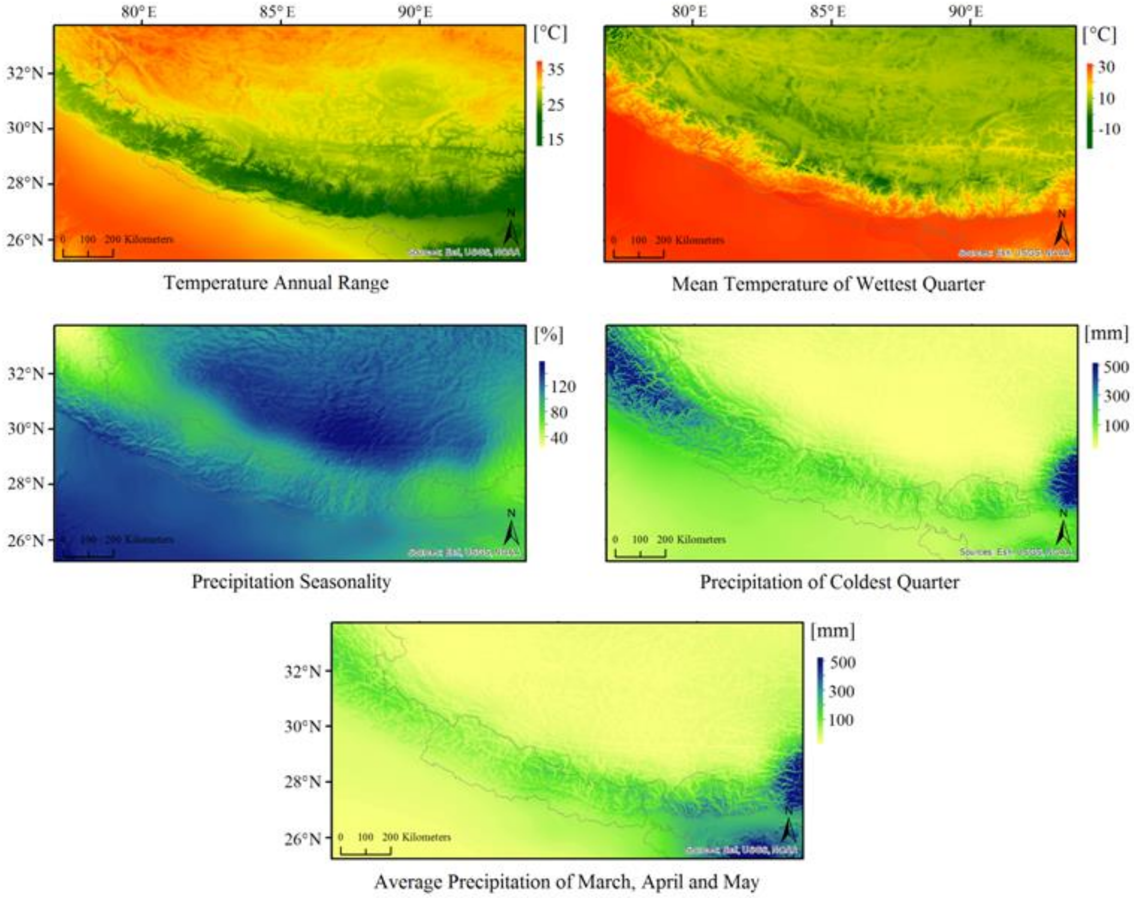


Fig. 6: Bioclimatic variables for modelling the ecological niche of *Betula utilis* in Article I (temperature-related variables obtained from Gerlitz *et al.* (2014) and precipitation-related variables obtained from Chelsa, Karger *et al.* (2016)).

In Article II, we evaluated the performance of two climate datasets, namely, Chelsa climate dataset (Karger *et al.,* 2016) and Worldclim climate dataset (Hijmans *et al.,* 2005). We compared the variables identified in Article I (Fig. 7) for modelling the potential distribution of *B. utilis* at the same spatial and temporal scales. Whereas WORLDCLIM consists of interpolated climate data, CHELSA climate data are based on a quasi-mechanistical statistical downscaling of the ERA-interim global circulation model with Global Precipitation Climatology Centre (GPCC) and Global Historical Climatology Network (GHCN) bias correction (for details, see Karger *et al.,* 2016).

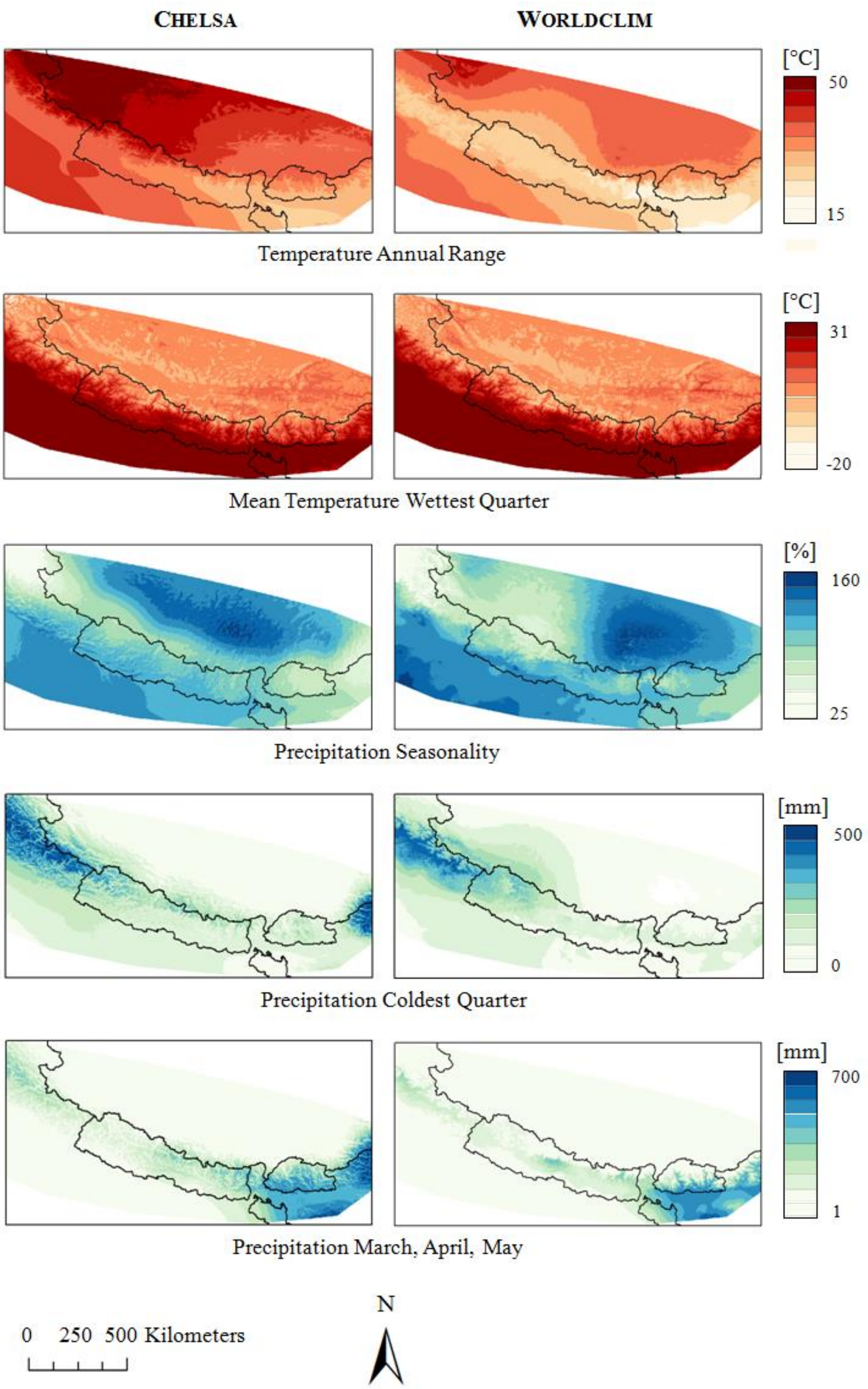


Fig. 7: Comparison of climatic variables for modelling the ecological niche of *Betula utilis* in Article II (derived from CHELSA (Karger *et al.,* 2016) and WORLCLIM (Hijmans *et al.,* 2005) climate datasets, respectively).

In Article III, we investigated the potential of combining climate data with remotely sensed data to bridge the gap between the potential and actual distribution of *B. utilis*. To this end, we evaluated 1) a quasi-mechanistical statistically downscaled Chelsa CLIMATE dataset (Karger *et al.,* 2016), 2) topographical variables based on a remotely sensed Digital Elevation Model (TOPO) (USGS, 2004), 3) phenological traits derived from MODIS Land Cover Dynamics data (PHENO) (LP DAAC, 2012), and 4) annual cycle parameters derived from MODIS Land Surface Temperature data (LST) (Bechtel, 2015). To test the potential of surface temperature in order to substitute downscaled climate data, predictor sets TOPO and PHENO were combined with either CLIMATE or LST (Fig. 8).

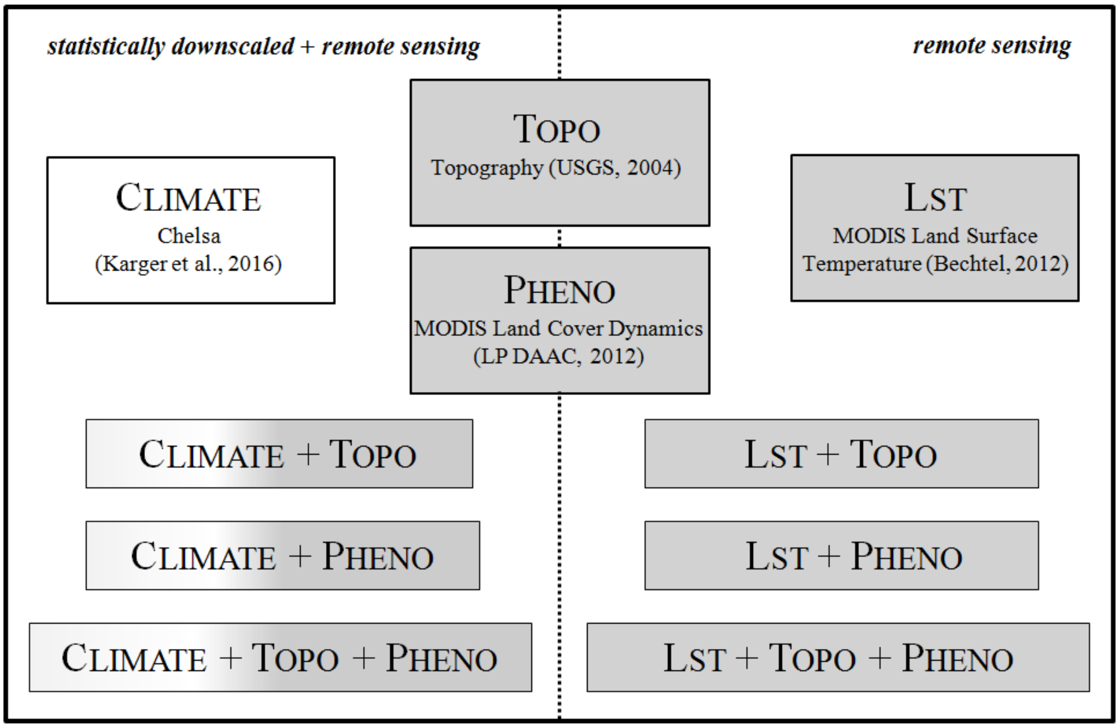


Fig. 8: Overview of the predictor sets used in the modelling procedure for estimating the ecological niche of *Betula utilis* in Article III.

6.4 Modelling procedure

Similar statistical analyses were used for all three articles in order to ensure comparability between the approaches. In all articles, Generalized Linear Models (GLMs) were applied since they represent a classical and robust approach to analyse presence and absence data (Nelder & Wedderburn, 1972; McCullagh & Nelder, 1989). The main advantages of GLMs over more complex machine-learning algorithms (e.g., random forest) are that they are easily interpretable and not 'black box' predictions. We calculated GLMs with binomial distribution, logit-link function and polynomial terms of second order (Austin, 1999), but did not include interaction terms among predictor variables. Prior to the modelling, step-wise variable selection in both directions (i.e., forward and backward) was applied using the Akaike Information Criterion (AIC) (Akaike, 1974), resulting in the model possessing the lowest AIC value (Burnham & Anderson, 2002; Guisan *et al.,* 2002).

As GLMs require presence and absence points, pseudo-absence points were generated. For study area selection, we utilised a convex hull, covering the full extent of the known occurrences of *B. utilis* distribution in the Himalayan region. By limiting the study area, large regions where the species cannot occur were excluded in further statistical analyses, so as to prevent overpredicting the distribution range of the species (VanderWal *et al.,* 2009). For random selection of pseudo-absences, the limits were set as 5 km from the nearest occurrence, resulting in a total of 6.000 (Articles I and II) and 10.000 (Article II) pseudo-absences (following the pseudo-absence selection procedure for GLMs described by Barbet-Massin *et al.,* 2012).

For model validation, all presence and pseudo-absence points were split into training and testing data samples with a ratio of 80:20 % using random sample splitting (Kuhn & Johnson, 2013). For each predictor variable set, we repeated this procedure five times, resulting in five versions of the model and accuracy metrics, which were finally averaged. Due to the lack of a universally valid model evaluation measurement, we applied several performance evaluation metrics. In order to assure comparability between the three articles, calculated evaluation measures included, among others, explained variance of the test dataset, threshold-independent AUC (Area Under the Curve) and threshold-dependent measures TSS (True Skill Statistics). Moreover, visual inspection of the predictions' spatial patterns was also conducted, since evaluation parameters may perform well in the model's environmental space, but not in geographic space.

We calculated variable importance in order to evaluate variable contribution in the final models for each predictor variable set in the respective articles.

All statistical analyses were performed using the programming language R (R Core Team, 2015, version: 3.1.3), and maps were produced using ArcGIS (Version 10.1; Esri Inc. Redlands, CA, USA).

7. Overview of original publications

7.1 Article I

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

Abstract: Developing sustainable adaptation pathways under climate change conditions in mountain regions requires accurate predictions of treeline shifts and future distribution ranges of treeline species. Here, we model for the first time the potential distribution of *Betula utilis*, a principal Himalayan treeline species, to provide a basis for the analysis of future range shifts. Our target species *Betula utilis* is widespread at alpine treelines in the Himalayan mountains, the distribution range extends across the Himalayan mountain range. Our objective is to model the potential distribution of *B. utilis* in relation to current climate conditions. We generated a data set of 590 occurrence records and used 24 variables for ecological niche modelling. We calibrated generalized linear models using the Akaike Information Criterion (AIC) and evaluated model performance using threshold-independent the Area Under the Curve (AUC), and threshold-dependent (TSS, True Skill Statistics) characteristics as well as visual assessments of projected distribution maps. We found two temperature-related (Mean Temperature of the Wettest Quarter, Temperature Annual Range) and three precipitation-related variables (Precipitation of the Coldest Quarter, Average Precipitation of March, April and May and Precipitation Seasonality) to be useful for predicting the potential distribution of *B. utilis*. All models had high predictive power (AUC ≥ 0.98 and TSS ≥ 0.89). The projected suitable area in the Himalayan mountains varies considerably, with most extensive distribution in the western and central Himalayan region. A substantial difference between potential and real distribution in the eastern Himalaya points to decreasing competitiveness of *B. utilis* under more oceanic conditions in the eastern part of the mountain system. A comparison between the vegetation map of Schweinfurth (1957) and our current predictions suggests that *B. utilis* does not reach the upper elevational limit in vast areas of its potential distribution range due to anthropogenically caused treeline depressions. This study underlines the significance of accuracies of current environmental niche models for species distribution modelling under climate change scenarios. Analysing and understanding the environmental factors driving the current distribution of *B. utilis* is crucial for the prediction of future range shifts of *B. utilis* and other treeline species, and for deriving appropriate climate change adaptation strategies.

Maria Bobrowski: Study design, climate data compilation, data analysis, modelling, writing and editing

Lars Gerlitz: Climate data compilation, discussion on interpretation of the results

Udo Schickhoff: Species occurrence data compilation, discussion on interpretation of the results and editing

7.2 Article II

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. doi: 10.1016/j.ecolmodel.2017.05.021.

Abstract: *Betula utilis* is a major constituent of alpine treeline ecotones in the western and central Himalayan region. The objective of this study is to analyse for the first time the performance of different climatic predictors in modelling the potential distribution of *B. utilis* in the subalpine and alpine belts of the Himalayan region. Using generalized linear models (GLM) we aim at examining climatic factors controlling the species distribution under current climate conditions. We evaluate the predictive ability of climate data derived from different statistical methods GLMs were created using least correlated bioclimatic variables derived from two different climate data sets: 1) interpolated climate data (i.e., WORLDCLIM; Hijmans *et al.,* 2005), and 2) quasi-mechanistical statistical downscaling (i.e., CHELSA; Karger *et al.,* 2016). Model accuracy was evaluated using threshold-independent (Area Under the Curve) and threshold-dependent (True Skill Statistics) measures. Although there were no significant differences between the models in AUC, we found highly significant differences ($p \leq 0.01$) in TSS. We conclude that models based on variables of CHELSA climate data had higher predictive power, whereas models using WORLDCLIM climate data consistently overpredicted the potential suitable habitat for *B. utilis*. Although climatic variables of WORLDCLIM are widely used in modelling species distribution, our results suggest to treat them with caution when topographically complex regions like the Himalaya are in focus. Unmindful usage of climatic variables for environmental niche models potentially causes misleading projections.

Maria Bobrowski:	Study design, climate data compilation, data analysis, modelling, writing and editing
Udo Schickhoff:	Discussion on interpretation of the results and editing

7.3 Article III

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; doi:10.3390/rs10060814.

Abstract: Modelling ecological niches across vast distribution ranges in remote, high mountain regions like the Himalayas faces several data limitations, in particular nonavailability of species occurrence data and fine-scale environmental information of sufficiently high quality. Remotely sensed data provide key advantages such as frequent, complete, and long-term observations of land surface parameters with full spatial coverage. The objective of this study is to evaluate modelled climate data as well as remotely sensed data for modelling the ecological niche of *Betula utilis* in the subalpine and alpine belts of the Himalayan region covering the entire Himalayan arc. Using generalized linear models (GLMs), we aim at testing factors controlling the species distribution under current climate conditions. We evaluate the additional predictive capacity of remotely sensed variables, namely remotely sensed topography and vegetation phenology data (phenological traits), as well as the capability to substitute bioclimatic variables from downscaled numerical models by remotely sensed annual land surface temperature parameters. The best performing model utilized bioclimatic variables, topography, and phenological traits, and explained over 69% of variance, while models exclusively based on remotely sensed data reached 65% of explained variance. In summary, models based on bioclimatic variables and topography combined with phenological traits led to a refined prediction of the current niche of *B. utilis*, whereas models using solely climate data consistently resulted in overpredictions. Our results suggest that remotely sensed phenological traits can be applied beneficially as supplements to improve model accuracy and to refine the prediction of the species niche. We conclude that the combination of remotely sensed land surface temperature parameters is promising, in particular in regions where sufficient fine-scale climate data are not available.

Maria Bobrowski:	Study design, climate data compilation, data analysis, modelling, writing and editing
Benjamin Bechtel:	Land Surface Temperature data compilation, discussion on model outputs, interpretation and editing
Jürgen Böhner:	Editing
Jens Oldeland:	Editing and statistical advise
Johannes Weidinger:	Modis Land Cover data compilation
Udo Schickhoff:	Discussion on interpretation of the results and editing

8. Synthesis

The presented PhD thesis consists of three articles modelling the ecological niche of *Betula utilis* in the Himalayan region. Article I aims at identifying the underlying factors influencing the current potential distribution of *B. utilis*, whereas Article II evaluates the performance of two different climate datasets when modelling the potential distribution of *B. utilis*. In Article III, the synergistic combination of modelled climate data and remotely sensed data were investigated for bridging the gap between the potential and actual distribution of *B. utilis*. Additionally, the potential of a purely remotely sensed approach was studied. In the following section, the key findings are briefly summarized on the basis of the articles and according to the initial research questions.

Furthermore, in this section, the results of the three articles are subsequently discussed in light of existing challenges for modelling ecological niches in high altitude regions. The new findings are embedded in the context of current and future treeline research. Finally, limitations of the applied modelling approach are considered, along with how the application of remotely sensed data may be beneficially incorporated into further studies on modelling treeline species and treeline dynamics under future climate change.

8.1 Modelling the potential distribution of *Betula utilis* in the Himalayan region

The potential distribution of *B. utilis* under current climate conditions was modelled and thermal and precipitation-related factors as predictive variables identified. The calculation of explained variance revealed that the model explained 77 % of the variance in the test dataset. Although the distribution range matches the distributional patterns of several vegetation maps (Troll, 1939; Schweinfurth, 1957; Miehe, 1991; Schickhoff, 1994; Braun, 1996; Nüsser & Dickoré, 2002; Eberhardt *et al.,* 2007), the distribution in the eastern Himalayas was overestimated. Given the respective climatically and topographically complex terrain, models based on climate variables alone predict only the potential distribution of species, since land cover characteristics are omitted. However, to estimate the actual distribution of species, other variables need to be incorporated (e.g., species' ecology data and biotic interactions) into the model. Nevertheless, this study investigated climatic conditions in treeline ecotones, and the results provide insights into underlying climatic factors and may serve as a baseline for altitudinal treeline shifts under future climate conditions.

What are key climatic constraints for the current distribution of *B. utilis* in the Himalayan region?

We found Mean Temperature of the Wettest Quarter, Temperature Annual Range, Precipitation of the Coldest Quarter, Average Precipitation of March, April and May, and Precipitation Seasonality to be important for modelling the potential distribution of *B. utilis*. Generally, the position of treelines is controlled by low temperatures during the growing season, which our models confirmed, as Mean Temperature of the Wettest Quarter was most influential. Due to the climatic gradient in the Himalayan Mountains ranging from more maritime in the East to more continental in the West, the position of the treeline and the composition of treeline species varies. The higher degree of maritime climate in the Eastern Himalayas lowers the competitive capacity of *B. utilis* and favours the competitiveness of *Rhododendron* spp. in subalpine forests and at treelines (Schickhoff, 2005).

Furthermore, the importance of Precipitation of the Coldest Quarter was also highlighted, since precipitation-related variables potentially limit the climatic space of treeline tree species. Winter precipitation in the form of snow represents the annual second precipitation maximum in high elevation areas of the Himalayan mountain system (Böhner *et al.,* 2015). Higher winter snowfall in the western Himalaya favours *B. utilis* occurrences compared to eastern Himalayas. The importance of snowpack as a site factor was already stressed by Schweinfurth (1957), Champion and Seth (1968), Stainton (1972), Herzhoff and Schnitzler (1981), Puri *et al.* (1989), Schickhoff (1993), Dickoré and Nüsser (2000), and Eberhardt (2004). A higher preoccupation of the potential distribution range of *B. utilis* is linked to a higher proportion of winter snowfall.

Are there differences between the current model predictions and the distribution range delineated in the detailed vegetation map of Schweinfurth (1957)?

The comparison between the prediction and the Schweinfurth map (Schweinfurth, 1957) revealed remarkable differences, especially in the central part of the Himalayan Mountains (Fig. 9). Although some inaccuracies in the Schweinfurth map cannot be ruled out, the differences in the distributional ranges are most likely related to anthropogenic land use effects. The vast majority of Himalayan treelines are considered to be anthropogenic due to animal husbandry, timber logging and fuelwood collection (Schickhoff *et al.,* 2015). The main agents lowering treelines have been overgrazing and fire, and treeline depressions on north-facing slopes can be up to 300 m (Schickhoff, 1995; Beug & Miehe, 1999). The difference between the modelled distribution range and the Schweinfurth map occurrence of *B. utilis* reflects large-scale, long-term anthropogenic interferences in Himalayan treeline landscapes. Even if the effects of climate warming trigger upslope shifts of *B. utilis* in coming decades (Schickhoff *et al.,* 2015), this might be primarily attributed to the cessation of land use. The few remaining near-natural krummholz treelines in the Himalayas can be regarded as still relatively unresponsive to climate warming, suggesting treeline advance to higher elevation to be a medium- to long-term process (Schickhoff *et al.*, 2016a; Schwab *et al.,* 2016).

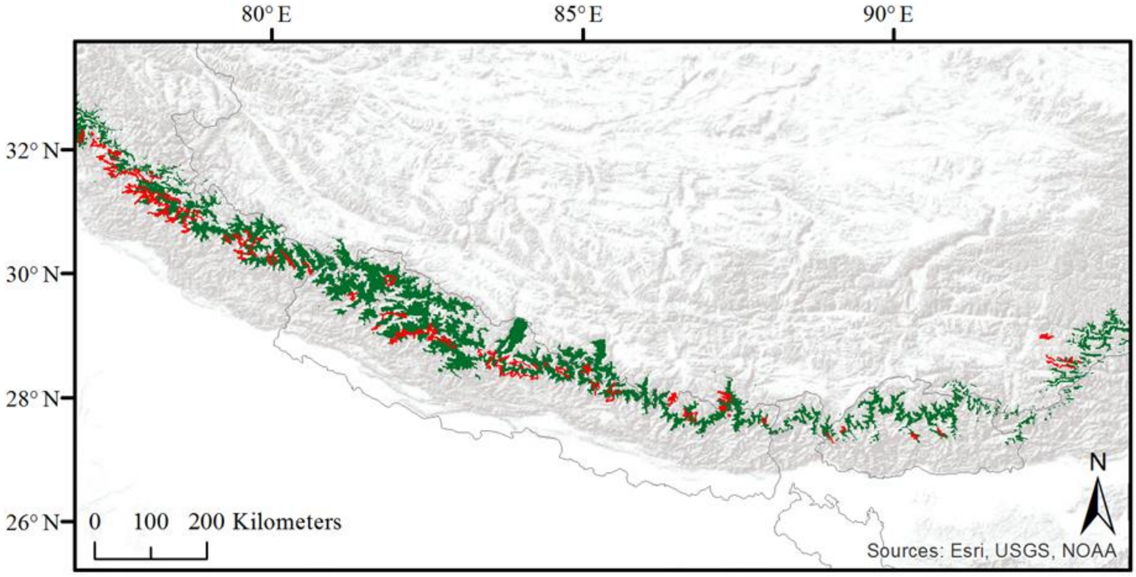


Fig. 9: Geographic extent of regions predicted as the potential distribution of *Betula utilis* (green) thresholded at True Skill Statistics (TSS ≥ 0.89) and the distribution of *B. utilis* according to the Schweinfurth vegetation map (red) (Schweinfurth, 1957).

8.2 Selection of climate datasets for modelling the potential distribution of a *Betula utilis* in the Himalayan region

Modelling species distribution in high altitude regions faces challenges due to limited data availability, which especially holds true for high quality climate data. Biased climate data can lead to distorted models (Heikkinen *et al.,* 2006), affecting accuracy when modelling potential range expansions of treeline trees under climate change conditions. Global climate datasets should not be used to model ecological niches without critically scrutinising the origin of climate data and the computation method of the climate dataset, and without being aware of the afflicted limitations. This may be particularly true for modelling studies in the Himalayan mountain system, and comparative studies are needed, since unmindful use of WORLDCLIM in heterogeneous landscapes like the Himalayan region could be misleading. The consequences of using misevaluated results as a baseline will severely affect model projections under future climate conditions.

How do input parameters and computation methods of the climate datasets (CHELSA and WORLDCLIM) influence derived climatic variables in vast mountain systems like the Himalayas?

Apart from different calculation methods, differences between the two climate datasets can be attributed to different input variables, resulting in different modelled temperature and precipitation data. Temperature is strongly negatively correlated with altitude (except for cold air inversions in winter months), allowing little room for variation in regional-scale climate datasets (Soria-Auza *et al.,* 2010, Karger *et al.,* 2016). Correspondingly, the comparison of temperature-related variables between the two climate datasets demonstrated a high degree of consistency. This was not the case with regard to precipitation patterns, which are influenced by a variety of factors, such as wind currents, topography and the diurnal cycle of solar radiation, which changes air pressure and therefore causes differential precipitation along altitudinal gradients. Dissimilarities in patterns of cloud formation became apparent between CHELSA and WORLDCLIM (Karger *et al.,* 2016). In summary, CHELSA demonstrated a more consistent relation between terrain and the resulting precipitation patterns, whereas the algorithm of WORLDCLIM produces flawed correlations between elevation and precipitation (Karger *et al.,* 2016). The WORLDCLIM climate dataset neglects local-scale atmospheric processes, which cause site-specific topo-climatic conditions in high mountain environments.

Are there discrepancies in the prediction of the two climate datasets (CHELSA and WORLDCLIM) when modelling the potential distribution of *B. utilis* under current climate conditions?

We compared CHELSA and WORLDCLIM climate data for the first time to model the potential distribution of *B. utilis* in the Himalayan region, using the same variables as in Article I to allow for direct comparison (Fig. 10). Unlike CHELSA, which only recently became available (Karger *et al.,* 2016), WORLDCLIM climate data (Hijmans *et al.,* 2005) have been widely used with reliable results (Elith *et al.,* 2006; Pearman *et al.,* 2008; Casalegno *et al.,* 2010) in Europe, North America and some tropical areas. Nevertheless, we found the ecological niche modelled based on CHELSA data to be in closer correspondence to the authors' field knowledge, and the model predictions match the actually existing distribution range of *B. utilis* to a vast extent, whereas models using WORLDCLIM climate data consistently overpredicted the potential suitable habitat for *B. utilis*. Due to different methods of computation, CHELSA and WORLDCLIM contain disparities, some of which may be of considerable importance for vast mountain systems like the Himalayas. We conclude that CHELSA climate data, which reflect topo-climatic conditions, yield more accurate results, particularly in terms of precipitation-related variables. In regard to model evaluation metrics, CHELSA significantly outperformed WORLDCLIM with 80.6 % explained variance of the test dataset compared to 67.7 %.

We provide a better understanding of how different climate datasets affect modelling species' potential distributions in regions with limited data availability. The accuracy of current potential distributions could gain considerable importance with regard to range shifts under climate change scenarios.

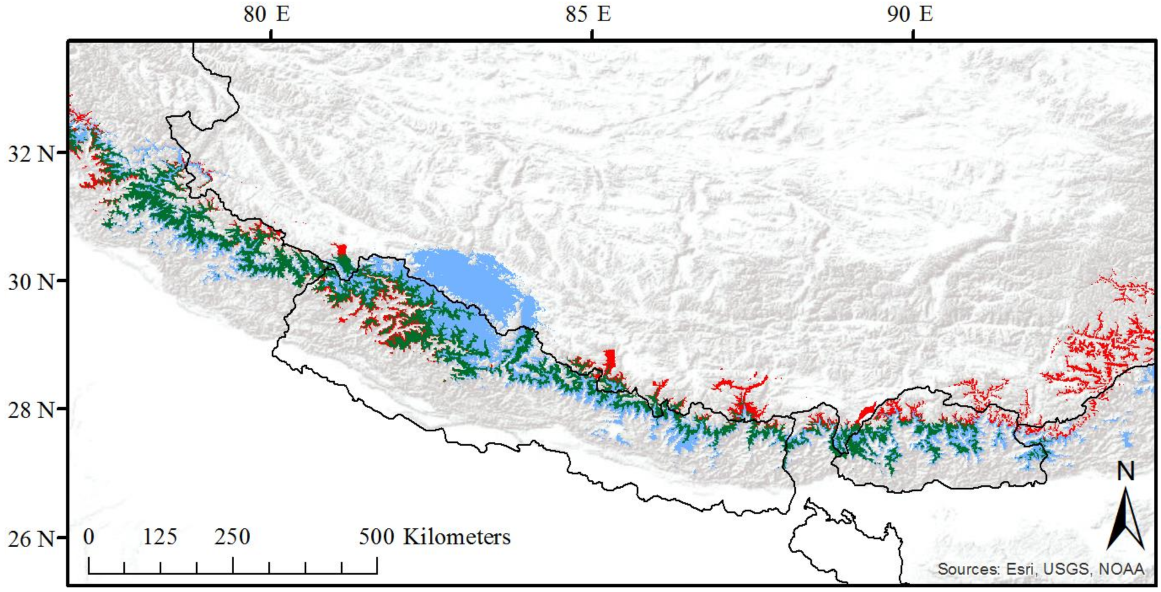


Fig. 10: Intersection of binary projections by both climate datasets (areas predicted by CHELSA (Karger *et al.*, 2016) climate data = red; areas predicted by WORLDCLIM (Hijmans *et al.,* 2005) climate data = blue; areas predicted by both climate datasets = green).

8.3 Application of thermal and phenological land surface parameters for improving ecological niche models of *Betula utilis* in the Himalayan region

Globally, climate governs global patterns of land cover (Dale, 1997), but land cover and climate are not fully independent (Thuiller *et al.,* 2004). Our results confirm topo-climatic variables as the main drivers behind the distribution range, whereas phenological traits substantially contribute to narrowing the modelled ecological niche of *B. utilis* (i.e., more realistic distribution). Similar results were obtained by Parra *et al.* (2004) and Buermann *et al.* (2008) when predicting species distribution across the Amazonian and Andean region. Remotely sensed data made a valuable contribution when modelling the current distribution of *B. utilis*, since they provide long-term, fine-scale and freely available observations. The approach of substituting bioclimatic variables with remotely sensed land surface temperature variables is promising, particularly in regions where sampling efforts are low and sufficient fine-scale climate data are not available.

To what extent can niche models built solely on bioclimatic predictor variables be improved by adding different remotely sensed predictor variables?

Given the respective climatically and topographically complex terrain of the Himalayan Mountains, models based on climate variables alone predict the potential distribution of species only when land cover characteristics are omitted. When modelling the ecological niche of a species, the major concern is the quality and content of environmental input data to generate the models. We presented the first detailed comparisons utilising remote sensing data for modelling the actual distribution of *B. utilis*. We conclude that the addition of remotely sensed topography and phenological traits as predictor variables led to improved model performance, resulting in a more constrained predicted niche compared to models solely built on climatic variables (Fig. 11).

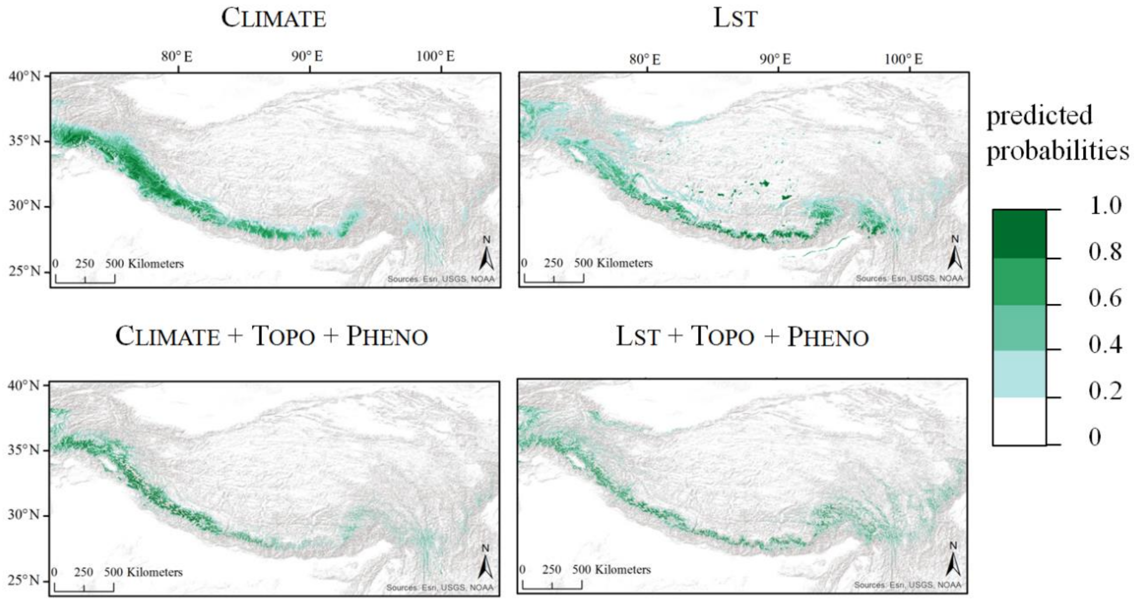


Fig. 11: Continuous predictions of the models using four different predictor variable sets: Chelsa climate data (CLIMATE); Land Surface Temperature (LST); and both combined with Topography (TOPO) and Land Cover Dynamics data (PHENO).

The results underline the relevance of additional remotely sensed environmental variables for reducing the gap between the potential and actual distribution of *B. utilis*. It becomes apparent that the core distribution of *B. utilis* was predicted in the western part of the Himalayan mountain system, whereas only the LST model predicted a principal distribution in the central part of the mountains. All models showed a uniform distribution along the Himalayan arc. The habitat predicted by CLIMATE tends to be wider in range compared to the other predictions.

To emphasize these findings (Fig. 11), I included a reanalysis of the model results (Fig. 12) with regard to the BAM diagram (Fig. 1). According to the BAM-diagram (Chapter 2), CLIMATE and LST models revealed the potential distribution $\mathbf{G_P}$ of *B. utilis*, respectively (Fig. 12a,b).

By contrast, in the CLIMATE + TOPO + PHENO and LST + TOPO + PHENO models, remotely sensed predictors supplemented information on **M** (TOPO: topographical barriers) and **B** (PHENO: distinction between phenological divergent vegetation formations), leading to the modelled actual distribution $\mathbf{G_O}$ of *B. utilis*, respectively (Fig. 12a,b). However, the actual distribution of *B. utilis* might be smaller than predicted, since topo-climate variables and phenological traits are not the only factors determining habitat suitability.

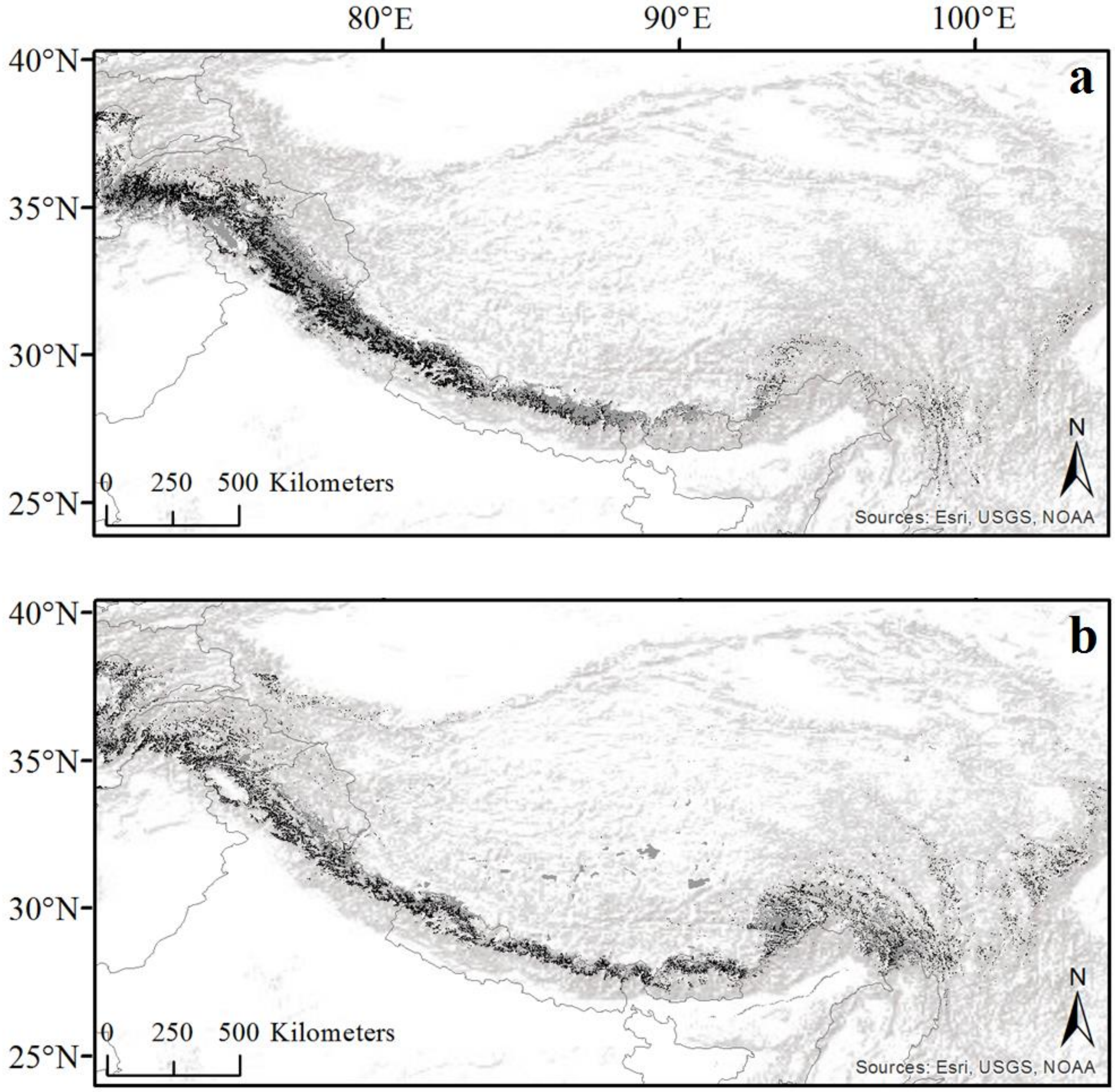


Fig. 12: Reanalysis of the model predictions for modelling the ecological niche of *Betula utilis*. **G$_P$** the potential distribution (grey) was modelled solely based on climate-related variables (a) Climate and b) Lst), and **G$_O$** the actual distribution (black) was modelled with additional remotely sensed variables like topography and phenological traits (a) Climate + Topo + Pheno and b) Lst + Topo + Pheno).

Furthermore, I included a detailed excerpt of the reanalysis displaying differences in the model predictions (Fig. 13), where the overall appearance (Fig. 11 and 12) is better elucidated. The model solely based on climate predictor variables (Climate) roughly met the lower limit of occurrences compared to the Climate + Topo + Pheno model, but overpredicted the uppermost limits of *B. utilis* (**G$_P$** in Fig. 13a) Overall, the broadleaved deciduous treeline could not be distinguished from other vegetation formations. The same picture emerged for Lst models (**G$_P$** in Fig. 13b).

The Climate + Topo + Pheno model differentiates between slope structures, and clearly delimits the lower distributional range of *B. utilis*. At higher altitudes, occurrence probability decreases (**G_O** in Fig. 13a). Comparing Climate + Topo + Pheno models with Lst + Topo + Pheno-based models, similar patterns were observed, whereas the distributional range is predicted more constrained in the latter, leaving a smaller distribution range of *B. utilis* (**G_O** in Fig. 13b).

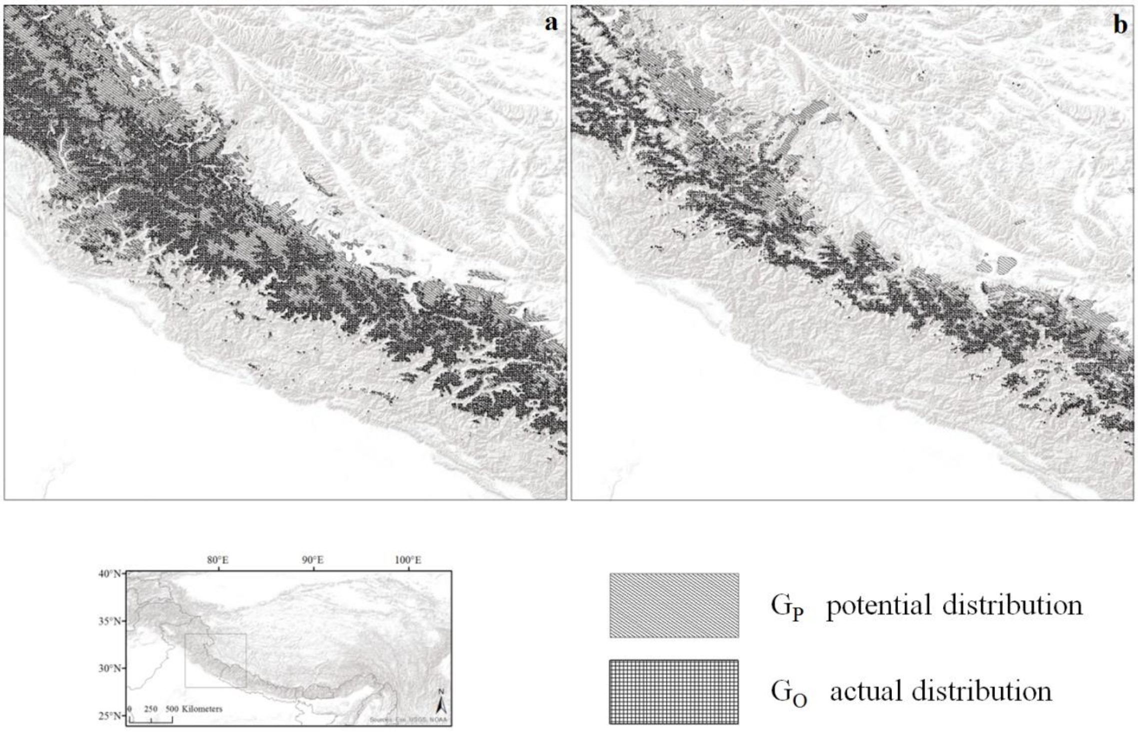


Fig. 13: Detailed excerpt of model predictions of the ecological niche of *Betula utilis* with regard to the BAM diagram. The potential distribution **G_P** was modelled solely based on climate-related variables and to estimate the actual distribution **G_O** additional remotely sensed variables such as topography and phenological traits were included. (a) **G_P**: Climate and **G_O**: Climate + Topo + Pheno b) **G_P**: Lst and **G_O**: Lst + Topo + Pheno).

What is the potential of a purely remote sensing-based approach when modelling the ecological niche of *B. utilis*?

Although explained variance was highest for Climate + Topo + Pheno with 69 %, Lst + Topo + Pheno obtained 65 %. The model solely based on downscaled climate parameters (Climate) explained 56 %, while the model solely based on remotely sensed land surface temperature (Lst) explained 41 % of the variance. Generally, Climate always performed better than Lst, except for explained variance for the combination with phenological traits (Climate + Pheno: 0.63; Lst + Pheno: 0.64). However, LST benefitted more from the addition of further predictors than CLIMATE (increase of 24 % compared to 13 % explained variance). In combination, LST + TOPO + PHENO performed better than CLIMATE. Our results emphasize the need for high-resolution data when modelling the actual distribution of treeline species in order to account for the heterogeneous terrain and microclimate. The incorporation of remotely sensed temperature derivatives expands the classical approach (i.e., bioclimatic variables and topography) and demonstrates great potential to derive more tailored variables (e.g., temperature of the growing season, precipitation amounts and snow cover) for ecological niche modelling.

Which variable combinations explain the current distribution of *B. utilis* in the Himalayan region?

Furthermore, phenological patterns will gain importance under climate change, since species-specific changes (e.g., earlier flowering and green-up date) will become apparent at an early stage (Xu *et al.,* 2009; Panday & Ghimire, 2012). The inclusion of topographic variables like Slope led to an improved prediction of climate-only models (both CLIMATE and LST). In highly topographically complex regions like the Himalayas, areas can be predicted as climatically suitable, but they might be inaccessible due to the slope angle. *Betula* forests thrive on humid, shady slopes with deeply weathered podzolic soils, and are more or less absent from south-facing slopes, particularly in the more continental western Himalaya (Schickhoff, 1995; Miehe *et al.,* 2015a). Regarding temperature-related remotely sensed variables, all variables proved to be of great importance, but mean annual temperature amplitude (YAST) demonstrated highest variable importance. Generally, temperature is strongly negatively correlated with altitude (except for cold air inversions in winter months), allowing little room for variation in regional-scale climate datasets (Soria-Auza *et al.,* 2010; Karger *et al.,* 2016). This, in turn, leads to similar results for LST + TOPO + PHENO compared to CLIMATE + TOPO + PHENO models, where additional information on phenological traits refined the predicted niche of *B. utilis* (Fig. 11). Interestingly, the number of cloud free acquisitions (NCSA) also demonstrated considerable variable importance, indicating that it contains some information about precipitation. Furthermore, MAST (mean annual temperature) and THETA (phase shift in days relative to spring equinox) were revealed to be important when modelling the ecological niche of *B. utilis*. We conclude that THETA, as it represents heat accumulation, indicates sensitivity to seasonal snow cover and thickness. Furthermore the duration of snow cover with sufficient soil moisture at the beginning and at the end of the growing season, is one of the principal factors controlling the distribution of *B. utilis* forests (Schickhoff, 1995, 2000, 2005).

8.4 Modelling treeline dynamics under climate change

Global and regional treeline responsiveness to climate change is highly complex and influenced by a variety of abiotic and biotic factors and their interrelations. Given the fact that the elevational position of treelines is attributed to prevailing thermal conditions, worldwide treeline ecotones constitute sensitive indicators to changing climate conditions (Körner, 2012).

During the Holocene, treeline fluctuations were caused by climate variability. Evidence was found for upslope movement during warmer periods and recession during cooler periods (Alps: Schwörer *et al.,* 2014, Himalaya: Schickhoff *et al.*, 2016a). More specifically, after the Pleistocene-Holocene transition period (11.7 kyr. BP) the treeline position in the Himalayas was situated several hundred meters higher than today. Highest elevational positions of treelines in the early Holocene can be attributed to warm and moist climate conditions resulting from a reinforced Asian monsoon regime (Schickhoff *et al.*, 2016a). In the mid-Holocene (5 kyr BP) treeline positions shifted to somewhat lower elevations due to decreasing temperatures (Schickhoff *et al.*, 2016a). In recent millennia, human impact has become the dominant driver of treeline elevational positions. In the Himalayas, Holocene treeline history was not uniform due to regional and local particularities.

Under future climate change scenarios, treeline positions are postulated to advance to higher elevations. To date, treeline responses do not show consistent patterns at global and local scales (Dullinger *et al.,* 2004; Harsch *et al.,* 2009; Körner, 2012; Schickhoff *et al.,* 2015). The degree of treelines' susceptibility to being significantly affected by changing climate depends on treeline type and form (Schickhoff *et al.,* 2015). Climatic treelines are highly susceptible to climate warming (Holtmeier & Broll, 2007; Körner, 2012), whereas orographic treelines do not show significant changes (Schickhoff *et al.*, 2016a). In terms of their responsiveness to climate warming, anthropogenic treelines can be compared to climatic treelines. Based on treeline types, four treeline forms with different responsiveness patterns can be distinguished (i.e., diffuse, abrupt, island and krummholz treeline forms; Harsch & Bader, 2011). Only diffuse treelines exhibit a strong response signal, whereas the other forms remain rather unreactive in terms of elevational shifts.

In the Himalayas, explicit differences can be found between treelines on north- and south-facing slopes. Whereas human impact transformed treelines on south-facing slopes to a large extent, north-facing slopes have a limited use potential, and near-natural treeline ecotones, including an intact krummholz belt, still exist (Schickhoff *et al.,* 2015). Krummholz treelines usually show a lagged response to climate warming, and treeline shifts may occur only in the long term (Schickhoff *et al.,* 2015, 2016). It is assumed that the strong competition within the krummholz belt and the occurrence of dense dwarf scrub heaths located above hinders upward migration of tree species

(Schickhoff *et al.,* 2015). However, stand densification and prolific regeneration within the treeline ecotone indicates beneficial preconditions for treeline advance in the future (Schickhoff *et al.,* 2015).

The analyses of treelines responding to changed climatic conditions and differentiation of small- and broad-scale underlying mechanisms and factors remains a pending question. Recent studies have focused on ecological, dendroecological, forest-ecological, and remote sensing aspects (Liang *et al.,* 2011; Gaire *et al.,* 2011, 2014; Bharti *et al.,* 2012; Rai *et al.,* 2013; Shrestha, 2013; Müller *et al.*, 2016a,b; Schwab *et al.*, 2016; Bürzle *et al.*, 2018). Increasingly, modelling approaches have been applied to gain a better understanding of treeline dynamics, particularly to investigate the underlying process-based relationships and identify potential range shifts of species in response to changed climatic conditions and altered land use regimes (Dullinger *et al.,* 2004; Wallentin *et al.,* 2008; Paulsen & Körner, 2014; Schickhoff *et al.,* 2015, 2016). However, the number of studies modelling treeline dynamics in the Himalayas remains sparse. A particular challenge for modelling studies involves the lack of natural treeline sites, since the treeline position is almost everywhere depressed to lower altitudes due to human impact (Miehe *et al.,* 2015b; Schickhoff *et al.,* 2015, 2016). At anthropogenically depressed treelines, it is a challenge to disentangle the climatic signal and anthropogenic land-use impacts as the driver behind treeline dynamics. Upslope shifts of treeline species in the short term might be attributed to changes in land use regimes rather than to climate change. For near natural treelines, it is postulated that changes in their elevational position will be a medium- to long-term process (Schickhoff *et al.,* 2015, 2016). The comparison of the vegetation map of Scheinfurth (1957) with the results of Articles I revealed deviations, possibly attributed to land use change rather than changed climatic conditions (Fig. 9). In the reanalysis the same pattern can be observed for the entire Himalayan Mountains, where similar discrepancies between the Schweinfurth map (1957) and predictions of the CLIMATE + TOPO + PHENO model (Article III) were found (Fig. 14).

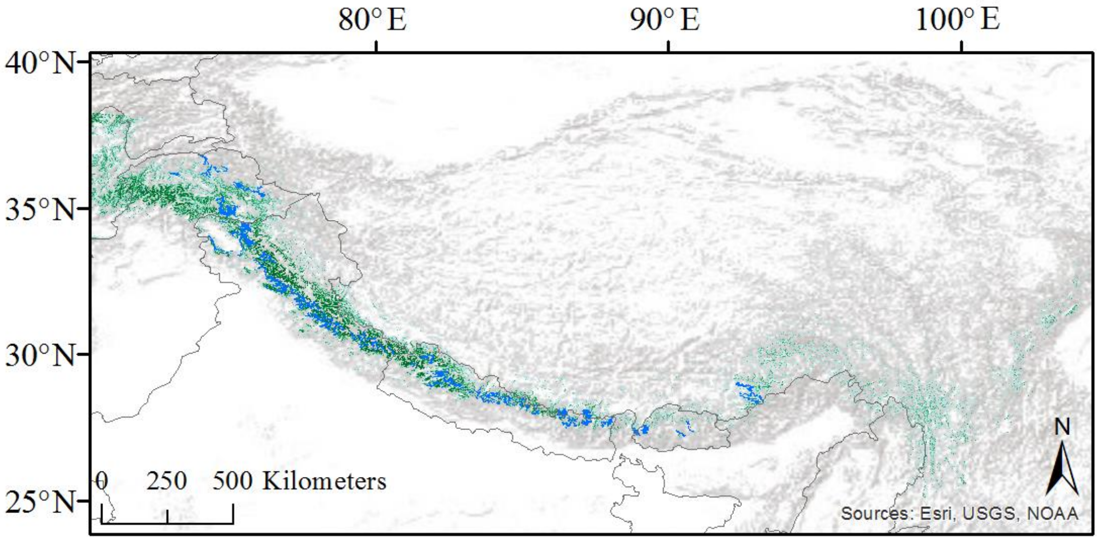


Fig. 14: Reanalysis of the modelled ecological niche of *Betula utilis* based on CLIMATE + TOPO + PHENO models and *B. utilis* forests according to the vegetation map of Schweinfurth (1957).

However, changing species' distributional patterns and phenology are responses to recent climate change that will modify the structure, composition and position of the treeline in the Himalayan mountain system. Remotely sensed data of plant phenological seasonal variations can be used to track changes in vegetation phenology (Beck *et al.,* 2007), since shifts in seasonal phenological events are among the first responses at plant and ecosystem levels to climate change (Badeck *et al.,* 2004). Shifts of flowering dates have been reported for *Rhododendron* species (Xu *et al.,* 2009), and earlier green-up data resulting in an extension of the growing season (Panday & Ghimire, 2012; Shrestha *et al.,* 2012) have been reported for the Himalayas. However, responses to above-average warming rates projected for the 21th century will most likely be associated with biodiversity loss and a decrease of ecosystem functions (Schickhoff *et al.*, 2016a).

Therefore, investigations of underlying climatic factors and the quantification of changing plant phenological traits provide the basis for efficient nature conservation management, expansion of protected areas and appropriate habitat restoration strategies. With the results presented, we provide a stepping stone for further investigations of treeline dynamics in the Himalayan mountain system.

8.5 Limitations of ecological niche models and potentials of remote sensing data

The investigation of factors driving the current distribution of treeline species is a *conditio sine qua non* of factors behind treeline dynamics. In order to obtain meaningful modelling results, a synthesis of various ecology-related disciplines would be desirable. Since the availability and quality of input parameters determine model performance, complete high-resolution long-term data is advisable. Besides species occurrence data, information ranging from plant-specific characteristics and responsiveness to changing climatic conditions and inter- and intra-specific competition to succession experiments would enhance modelling procedures.

However, in reality, things are different. By using correlative modelling approaches, limitations and errors may occur at any step of the procedure. Far-reaching consequences can be traced back to the input parameters. The model is only as precise as the quality and relevance of the biotic and abiotic parameters used to build the model for the targeted species. In an extensive literature review, He *et al.* (2015) presented numerous applications of remotely sensed data for modelling species' distributions. They demonstrated the adaptability of remote sensing products for modelling marine and terrestrial biota, and how they can be customised in accordance with specific research questions.

As already mentioned in Chapter 3, modelling studies are often based on presence-only species occurrence data instead of presence-absence data, which represent a major source of uncertainty. They are often derived from databases of natural history museums and herbaria, whereby sampling techniques, intensities and periods of time may differ (Soberón & Peterson, 2004). Sampling bias in geographic space leads to sampling bias in environmental space, which must certainly be considered problematic (Veloz, 2009). Spatial filtering (i.e., only one point per 1 x 1 km grid cell) of the occurrence points was applied to decrease sampling bias and spatial auto-correlation. No assurance can be provided regarding afflicted biases of museum- and literature-based occurrences.

In many studies, the investigation and evaluation of input parameters is neglected. The impact of the results' implementation may have severe consequences. Abiotic and biotic data derived from remote sensing may open up new opportunities for analysing and modelling species' distributions, since they provide response and predictor variables.

In all articles, the benefits of remotely sensed data in deriving tree species occurrences could be highlighted. Almost 80 % (Article III) and 55 % (Article I and II) of the *Betula utilis* occurrence points were extracted from freely available satellite imagery (Google, 2015). The potential of remote sensing data for future studies lies in the generation of presence and absence datasets, which are highly required in ENMs (Fithian *et al.,* 2015). Due to unique biophysical properties, hyperspectral sensors can detect subtle differences in reflectance based on unique plant chemistries, which is beneficial for identifying plant species (Buermann *et al.,* 2008). Not only clearly noticeable vegetation structures like treeline ecotones can be distinguished, but technologies may also be applicable for detecting rare or invasive species at the plant species level (He *et al.,* 2015 and references therein). Another advantage is the possibility to incorporate biotic interactions into the models, which are often disregarded due to data limitations (Kissling *et al.,* 2012; Dormann *et al.,* 2018). The inclusion of biotic interactions of tree species treeline species associated with *B. utilis* would be beneficial for modelling possible future range shifts.

Often, standardised statistically derived parameters do not fully reflect the species' physiological needs and habitat requirements, and therefore lead to poor modelling results. As illustrated in Article II, evaluating and comparing the performance of climate datasets remains a challenging task. In Article III, thermal metrics were tested, which may be beneficially incorporated into further treeline studies in remote mountainous regions, as they provide freely accessible, complete and long-term data. The main advantages of LST-related variables are continuous observations without interpolation and geographical bias, and therefore with less uncertainty (He *et al.,* 2015). Recent studies have revealed how LST data could improve species modelling studies (e.g., Buermann *et al.,* 2008; Bisrat *et al.,* 2012; Still *et al.,* 2014). These parameters offer numerous possibilities, such as tailored predictors in high-resolution. As time series data of vegetation characteristics (i.e., phenological metrics) are becoming more and more readily available, changing habitat suitability can be estimated and incorporated into model approaches. In this way, knowledge can be generated that is particularly important for modelling spatial expansion of invasive species, extinction risk assessment and range shifts under future climate change (He *et al.,* 2015 and references therein). In mountainous areas, the resolution of climate data (i.e., 1 x 1 km) is often too coarse for models to distinguish between north- and south-facing slopes. With high-resolution remote sensing data, however, the heterogeneity of the terrain can be taken into account, leading to more precise modelling results.

We conclude that although the available data derived from remote sensing technology is rather short-term, the presented results of Article III may serve as a baseline for future studies. Restrictions in the practical applicability arise from the fact that high resolution satellite imagery is still often very expensive. On the other hand, the free of cost imagery and software is already available and will become more customary in the future. The results of Article III show that, even with freely available data, model performances could be improved, indicating the potential for future modelling studies. Airborne technology is a continually expanding field, and high resolution remotely sensed data will provide more insights into spatial patterns and underlying factors in future modelling studies.

9. Conclusions

The aim of the thesis was to model for the first time the ecological niche of *Betula utilis* in the subalpine and alpine belts in the Himalayan Mountains on a broad scale.

The special focus was on challenges involved in modelling the ecological niches of *B. utilis* in remote, high elevation treeline ecotones along the Himalayan mountain range. *B. utilis* occurrences in the treeline ecotone along the Himalayan arc were compiled primarily from freely available satellite images. To the knowledge of the authors, the presented data set is the most comprehensive data set available.

The results of this study provide a comprehensive analysis of the underlying environmental factors (climatic patterns, topography and phenological traits) determining the ecological niche of *B. utilis* in the Himalayan region under current climate conditions. Using GLMs, potential suitable habitats of the species were successfully predicted as a function of climatic variables that characterise current climatic conditions at treeline locations. The identified environmental factors were discussed in light of species-specific ecological habitat requirements. We conclude that ecological niche modelling presents a valuable predictive tool for analysing the distribution of treeline species when the existing complexity of remote high altitude regions is denoted in climate input variables.

We highlight that global climate data sets should not be used to model ecological niches without critically scrutinizing the origin of climate data and the computation method of the climate data set, and without being aware of potentially afflicted limitations. The obtained results could be particularly misleading, when modelling ecological niches in heterogeneous landscapes like the Himalayan region. We emphasize the use of high resolution (< 1 km^2) local climate data sets for future modelling studies.

We expanded the solely climate-based approach with freely available remotely sensed variables to model the actual distribution of the species. The inclusion of variables characterising spatial variation in environmental variables, such as remotely sensed vegetation indices, provided key inputs. *Betula* is a conspicuous broadleaved deciduous tree species at treelines, allowing for a clear separation on the basis of phenological traits from adjacent vegetation types (evergreen coniferous and evergreen broadleaved species). The incorporation of remote sensing data led to a more refined modelled distribution since, based on real information of the Earth's surface, as they account for non-climatic dimensions (i.e., anthropogenic impacts), leading in turn to a more realistic actual distribution. Although the model predictions are in general agreement with several vegetation maps, the actual distribution might be smaller than indicated by the models. Our results point to the need for further investigations of microclimatic conditions with parameters related to soil properties or solar radiation, as well as investigations of biotic interactions or dispersal limitations.

The presented synergetic modelling approach can be transferred to any other species in conspicuous vegetation formations, such as treeline ecotones, and the applied environmental predictors are transferable without severe modification due to global coverage. Further research may include additional remotely sensed metrics, such as solar radiation, precipitation amounts and snow cover.

Transferability to other deciduous treelines remains to be examined, but promising results and insights can be expected. The obtained insights may serve as a role model for other climatic treelines, and especially for other *Betula* treelines in mountain regions. Our findings may serve as a baseline for further investigations of treeline dynamics under future climate change scenarios in regions with limited data availability.

10. Bibliography

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Alexander, J.M.; Diez, J.M.; Levine, J.M. (2015) Novel competitors shape species' responses to climate change. *Nature*, **525**, 515–518.

Anadon-Rosell, A.; Rixen, C.; Cherubini, P.; Wipf, S.; Hagedorn, F.; Dawes, M.A. (2014) Growth and phenology of three dwarf shrub species in a six-year soil warming experiment at the alpine treeline*. PLoS One*, **9**, e100577.

Anderson, R.P.; Gonzalez, Jr, I. (2011) Species-specific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. *Ecological Modelling*, **222**, 2796–2811.

Araújo, M.; Pearson, R.G.; Thuiller, W.; Erhard, M. (2005) Validation of species - climate impact models under climate change. *Global Change Biology*, **11**, 1504–1513.

Araújo, M.; Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.

Araújo, M.B.; New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42–47.

Ashburner, K.; McAllister, H.A. (2013) *The genus Betula – a taxonomic revision of birches*. Botanical Magazine Monograph 5. Royal Botanic Gardens, Kew.

Austin M.P. (1999) A silent clash of paradigms: some inconsistencies in community ecology. *Oikos*, **86**, 170–178.

Barbet-Massin, M.; Jiguet, F.; Albert, C.H.; Thuiller, W. (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, **3**, 327–338.

Bechtel, B. (2015) A new global climatology of annual land surface temperature. *Remote Sensing*, **7**, 2850–2870.

Beug, H.J.; Miehe, G. (1999) *Vegetation history and human impact in the eastern central Himalaya (Langtang and Helambu, Nepal)*. Dissertationes Botanicae 318, Borntraeger, Berlin-Stuttgart.

Bhattacharyya, A.; Shah, S.K.; Chaudhary, V. (2006) Would tree-ring data of *Betula utilis* have potential for the analysis of Himalayan glacial fluctuations? *Current Science*, **91**, 754–761.

Bisrat, S.A.; White, M.A.; Beard, K.H.; Richard Cutler, D. (2012) Predicting the distribution potential of an invasive frog using remotely sensed data in Hawaii. *Diversity and Distributions*, **18**, 648–660.

Böhner, J.; Miehe, G.; Miehe S.; Nagy, L. (2015) Climate and Weather. *Nepal: an introduction to the natural history, ecology and human environment of the Himalayas* (ed. by G. Miehe; C.A. Pendry and R. Chaudhary). pp. 23–89. Royal Botanic Garden, Edinburgh.

Braun, G. (1996) *Vegetationsgeographische Untersuchungen im NW–Karakorum (Pakistan). Kartierung der aktuellen Vegetation und Rekonstruktion der aktuellen Waldverbreitung auf der Basis von Satellitendaten, Gelände– und Einstrahlungsmodellen*. Bonner Geographische Abhandlungen 93, Dümmlers, Bonn.

Braunisch, V.; Coppes, J.; Arlettaz, R.; Suchant, R.; Schmid, H.; Bollmann, K. (2013) Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. *Ecography*, **36**, 971–983.

Buermann, W.; Saatchi, S.; Smith, T.B.; Zutta, B.R.; Chaves, J.A.; Milá, B.; Graham, C.H. (2008) Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeograph*, **35**, 1160–1176.

Burnham, K.P.; Anderson, D.R. (2002) *Model selection and multimodel inference: a practical information–theoretic approach*, 2nd ed., Springer, New York.

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

Casalegno, S.; Amatulli, G.; Camia, A.; Nelson, A.; Pekkarinen, A. (2010) Vulnerability of *Pinus cembra* L. in the Alps and the Carpathian mountains under present and future climates. *Forest Ecology and Management*, **259**, 750–761.

Champion, H.G.; Seth, S.K. (1968) *A revised survey of the forest types of India*. Government of India, New Delhi.

Colwell, R.K.; Rangel, T.F. (2009) Hutchinson's duality: The once and future niche. *Proceedings of the National Academy of Sciences USA*, **106**, 19651–19658.

Cotto, O.; Wessely, J.; Georges, D.; Klonner, G.; Schmid, M.; Dullinger, S. Thuiller, W.; Guillaume, F. (2017): A dynamic eco-evolutionary model predicts slow response of alpine plants to climate warming. *Nature Communications*, **8**, 1–9.

Dale. V.H. (1997) The relationship between land-use change and climate change. *Ecological Applications*, **7**, 753–769.

Dawadi, B.; Liang, E.; Tian, L.; Devkota, L.P.; Yao, T. (2013) Pre-monsoon precipitation signal in tree rings of timberline *Betula utilis* in the central Himalayas. *Quaternary International*, **283**, 72–77.

Dickoré, W.B.; Nüsser, M. (2000) *Flora of Nanga Parbat (NW Himalaya, Pakistan). An annotated inventory of vascular plants with remarks on vegetation dynamics*. Englera 19, Veröffentlichungen aus dem Botanischen Garten und Botanischen Museum Berlin-Dahlem.

Dirnböck, T.; Dullinger, S. (2004) Habitat distribution models, spatial autocorrelation, functional traits and dispersal capacity of alpine plant species. *Journal of Vegetation Science*, **15**, 77–84.

Dormann, C.F.; McPherson, J.M.; Araujo, M.B.; Bivand, R.; Bolliger, J.; Carl, G.; Davies, R.G.; Hirzel, A.; Jetz, W.; Kissling, W.D.; Kühn, I.; Ohlemüller, R.; Peres-Neto, P.R.; Reineking, B.; Schröder, B.; Schurr, F.M.; Wilson, R.(2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.

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.; Lautenbach, S. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 27–46.

Dormann, C.F.; Bobrowski, M.; Dehling, M.; Harris, D.J.; Hartig, F.; Lischke, H.; Moretti, M.D.; Pagel, J.; Pinkert, S.; Schleuning, M.; Schmidt, S.; Sheppard, C.; Steinbauer, M.J.; Zeuss, D.; Kraan, C. Biotic interactions in species distribution modelling: ten questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography*, **2018**, (in press).

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.

Dutta, P.K.; Dutta, B.K.; Das, A.K.; Sundriyal, D R.C. (2014) Alpine timberline research gap in Himalaya: a literature review. *Indian Forester*, **140**, 419–427.

Eberhardt, E. (2004) *Plant life of the Karakorum. The vegetation of the upper Hunza catchment (Northern Areas, Pakistan). Diversity, syntaxonomy, distribution*. Dissertationes Botanicae 387, Borntraeger, Stuttgart.

Eberhardt, E.; Dickoré, W.B.; Miehe, G. (2007) Vegetation map of the Batura Valley (Hunza Karakorum, North Pakistan). *Erdkunde*, **61**, 93–112.

Elith, J.; Graham, C.H.; Anderson, R.P.; Dudík, M.; Ferrier, S.; Guisan, A.; Hijmans, R.J.; Huettmann, F.; Leathwick, J.R.; Lehmann, A.; Li, J.; Lohmann, L.G.; Loiselle, B.A.; Manion, G.; Moritz, C.; Nakamura, M.; Nakazawa, Y.; McC. M. Overton, J.; Peterson, A.T.; Phillips, S.J.; Richardson, K.; Scachetti-Pereira, R.; Schapire, R.E.; Soberón, J.; Williams, S.; Wisz, M.S.; Zimmermann N.E. (2006) Novel methods improve prediction of species distributions from occurrence data. *Ecography*, **29**, 129–151.

Elliot, G.P. (2017): Treeline ecotones. *International encyclopedia of geography. People, the earth, environment, and technology.* (ed. by D. Richardson; N. Castree; M. F. Goodchild; A. Kobayashi; W. Liu und R. A. Marston). pp.1–10. John Wiley & Sons.

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

Ernakovich, J.G.; Hopping, K.A.; Berdanier, A.B.; Simpson, R.T.; Kachergis, E.J.; Steltzer, H.; Wallenstein, M.D. (2014) Predicted responses of arctic and alpine ecosystems to altered seasonality under climate change. *Global Change Biology*, **20**, 3256–3269.

Escobar, L.E.; Craft, M.E. (2016) Advances and Limitations of Disease Biogeography Using Ecological Niche Modeling. *Frontiers in Microbiology*, **7**, 1–21.

ESRI, 2012. ArcGIS Desktop: Release 10.1. Environmental Systems Research Institute, Redlands, CA.

Fithian, W.; Elith, J.; Hastie, T.; Keith. D.A. (2015) Bias correction in species distribution models: pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, **6**, 424–438.

Franklin. J. (2009) *Mapping species distributions: spatial inference and prediction*. Ecology, Biodiversity and Conservation. Cambridge University Press, Cambridge.

Gaire, N.P.; Bhuju, D.R.; Koirala, M. (2013) Dendrochronological studies in Nepal: Current status and future prospects. FUUAST *Journal of Biology*, **3**, 1–9.

Gaire, N.P.; Koirala, M.; Bhuju, D.R.; Borgaonkar, H.P. (2014) Treeline dynamics with climate change at Central Nepal Himalaya. *Climate of the Past*, **9**, 5941–5976.

Gajurel, J.P.; Werth, S.; Shrestha, K.K.; Scheidegger, C. (2014) Species distribution modeling of *Taxus wallichiana* (Himalayan Yew) in Nepal Himalaya. Asian *Journal of Conservation Biology*, **3**, 127–134.

Gaston, K.J. (2003) *The Structure and Dynamics of Geographic Ranges.* Oxford University Press, Oxford.

gbif.org: Biodiversity occurrence data provided by: Missouri Botanical Garden*, Royal Botanic Garden Edinburgh and T*he Himalayan Uplands Plant database*, Accessed through GBIF Data Portal, http://*www.gbif.org*, Last accessed 28 April 2015.*

Gerlitz, L., Conrad, O.; Böhner, J. (2014) Warming patterns over the Tibetan Plateau and adjacent lowlands derived from elevation- and bias-corrected ERA-Interim data. *Climate Research*, **58**, 235–246.

Google Earth, ver. 7.1.1.1888, Google, 2015.

Gottfried, M.M.; Pauli, H.; Futschik, A.; Akhalkatsi, M.; Barančok, P.; Benito A., José L.; Coldea, G.; Dick, J.; Erschbamer, B.; Fernández Calzado, M.R.; Kazakis, G.; Krajči, J.; Larsson, P.; Mallaun, M.; Michelsen, O.; Moiseev, D.; Moiseev, P.; Molau, U.; Merzouki, A.; Nagy, L.; Nakhutsrishvili, G.; Pedersen, B.; Pelino, G.; Puscas, M.; Rossi, G.; Stanisci, A.; Theurillat, J.-P.; Tomaselli, M.; Villar, L.; Vittoz, P.; Vogiatzakis, I.; Grabherr, G. (2012) Continent-wide response of mountain vegetation to climate change. *Nature Climate Change*, **2**, 111–115.

Grinnell, J. (1917). The niche-relationships of the California Thrasher. The Auk, **34**, 427–433.

Guisan, A.; Edwards Jr., T.C.; Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**, 89–100.

Hansen, W.; Magiera, A.; Theissen, T.; Waldhardt, R.; Otte, A.; Rocchini, D. (2018): Analysing *Betula litwinowii* encroachment and reforestation in the Kazbegi region, Greater Caucasus, Georgia. *Journal of Vegetation Science*, **29**, 110–123.

Harsch, M.A.; Hulme, P.E.; McGlone, M.S.; Duncan, R.P. (2009) Are treelines advancing? A global meta-analysis of treeline response to climate warming. *Ecology Letters*, **12**, 1040–1049.

Hart, R.; Salick, J.; Ranjitkar, S.; Xu, J. (2014) Herbarium specimens show contrasting phenological responses to Himalayan climate. *Proceedings of the National Academy of Sciences USA*, **111**, 10615–10619.

Heikkinen, R.K.; Luoto, M.; Araujo, M.B.; Virkkala, R.; Thuiller, W.; Sykes, M.T. (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, **30**, 751–777.

Herzhoff, B.; Schnitzler, H. (1981) Die subalpinen Birken-Rhododendron-Wälder im Dagwan-Tal/West-Himalaya. *Mitteilungen der Deutschen Dendrologischen Gesellschaft*, **72**, 171–186.

Hijmans, R.J.; Cameron, S.E.; Parra ,J.L.; Jones, P.G.; Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hirzel, A.H.; Hausser, J.; Chessel, D.; Perrin, N. (2002) Ecological-niche factor analysis: how to compute habitat-suitablility maps without absence data? *Ecology*, **83**, 2027–2036.

Holtmeier, F.-K.; 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**, 395–410.

Holtmeier, F.-K.; Broll, G. (2007): Treeline advance - driving processes and adverse factors. *Landscape Online*, **1**, 1–33.

Holtmeier, F.-K. (2009) *Mountain timberlines – Ecology, patchiness and dynamics*. Advances in Global Change Research 36, Springer, Berlin.

Hughes, L. (2000) Biological consequences of global warming: is the signal already apparent? *Trends in Ecology & Evolution*, **15**, 56–61.

Huo, C.; Cheng, G.; Lu, X.; Fan, J. (2010) Simulating the effects of climate change on forest dynamics on Gongga Mountain, Southwest China. *Journal of Forest Research*, **15**, 176–185.

Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology.* **22**, 415–427.

Hutchinson, G.E. (1978) *An Introduction to Population Ecology*. Yale University Press, New Haven.

IPCC (2013) Summary for Policymakers. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by T.F. Stocker; D. Qin; G.-K. Plattner; M. Tignor; S.K. Allen; J. Boschung; A. Nauels; Y. Xia; V. Bex and P.M. Midgley). pp.1–30. Cambridge University Press, Cambridge and New York.

IPCC (2014) *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by Core Writing Team, R.K. Pachauri and L.A. Meyer). International Panel on Climate Change, Switzerland, Geneva.

Irl, S.D.H.; Anthelme, F.; Harter, D.E.V.; Jentsch, A.; Lotter, E.; Steinbauer, M.J,; Beierkuhnlein, C. (2016) Patterns of island treeline elevation - a global perspective. *Ecography*, **39**, 1–10.

Jain, S. K.; Kumar, V.; Saharia, M. (2013) Analysis of rainfall and temperature trends in northeast India, *International Journal of Climatology*, **33**, 968– 978.

Jaryan, V.; Datta, A.; Uniyal, S.K.; Kumar, A.; Gupta, R.C.; Singh, R.D. (2013) Modelling potential distribution of *Sapium sebiferum* - an invasive tree species in western Himalaya. *Current Science*, **105**, 1282–1288.

Jentsch, C.; Liedtke, H. (1980) Höhengrenzen in Hochgebirgen - Einleitende Bemerkungen zum Rundgespräch. *Höhengrenzen in Hochgebirgen*. (ed. by C. Jentsch and H. Liedtke). pp.29–33. Arbeiten aus dem Geographischen Institut der Universität des Saarlandes Band 29, Saarbrücken.

Karger, D.N.; Conrad, O.; Böhner, J.; Kawohl, T.; Kreft, H.; Soria-Auza, R.W.; Zimmermann, N.; Linder, H.P.; Kessler, M. (2016) Climatologies at high resolution for the earth land surface areas. arXiv:1607.00217 [physics].

Kearney, M. (2006). Habitat, environment and niche. What are we modelling? *Oikos*, **115**, 186–191.

Kearney, M.; Porter, W. (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334–350.

Kissling, W.D.; Dormann, C.F.; Groeneveld, J.; Hickler, T.; Kühn, I.; McInerny, G.J.; Montoya, J.M.,; Römermann, C.; Schiffers, K.; Schurr, F M.; Singer, A.; Svenning, J.; Zimmermann, N.E.; O'Hara, R.B. (2012) Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–2178.

Koike, T.; Kitao, M.; Quoreshi, A. M.; Matsuura, Y. (2003): Growth characteristics of root-shoot relations of three birch seedlings raised under different water regimes. *Plant and Soil*, **255**, 303–310.

Körner, C.; Paulsen, J. (2004): A world-wide study of high altitude treeline temperatures. *Journal of Biogeography*, **31**, 713–732.

Körner, C. (2007): The use of 'altitude' in ecological research. *Trends in Ecology & Evolution,* **22**, 569–574.

Körner, C. (2012) *Alpine treelines – functional ecology of the global high elevation tree limits*. Springer, Berlin-Heidelberg.

Krestov, P.; Omelko, A.; Nakamura, Y. (2008) *Vegetation and natural habitats of Kamchatka.* Berichte der Reinhold-Tüxen-Gesellschaft 20, Hannover.

Kuhn, M.; Johnson, K. (2013) *Applied predictive modeling*. Springer: New York.

Kullman, L. (1998) Tree-limits and montane forests in the Swedish Scandes: sensitive biomonitors of climate change and variability. *Ambio*, **27**, 312–321.

Kumar, P. (2012) Assessment of impact of climate change on Rhododendrons in Sikkim Himalayas using Maxent modelling: limitations and challenges. *Biodiversity and Conservation*, **21**, 1251–1266.

Liang, E.; Dawadi, B.; Pederson, N.; Eckstein, D. (2014) Is the growth of birch at the upper timberline in the Himalayas limited by moisture or by temperature? *Ecology*, **95**, 2453–2465.

Loiselle, B.A.; Jørgensen, P.M.; Consiglio, T.; Jiménez, I.; Blake, J.G.; Lohmann, L.G.; Montiel, O.M. (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes?. *Journal of Biogeography*, **35**, 105–116.

LP DAAC (2012) NASA Land Processes Distributed Active Archive Center, USGS/Earth Resources Observation and Science (EROS) Center (https://lpdaac.usgs.gov/data_access/data_pool).

Luoto, M.; Virkkala, R.; Heikkinen, R.K. (2007) The role of land cover in bioclimatic models depends on spatial resolution. *Global Ecology and Biogeography*, **16**, 34–42.

Manish, K.; Telwala, Y.; Nautiyal, D.C.; Pandit, M.K. (2016) Modelling the impacts of future climate change on plant communities in the Himalaya: a case study from eastern Himalaya, India. *Modeling Earth Systems and Environment*, **2**, 1–12.

McCullagh, P.; Nelder, J.A. (1989) *Generalized linear models*. Chapman and Hall, London.

Menon, S.; Choudhury, B.I.; Khan. M.L.; Peterson, A.T. (2010) Ecological niche modelling and local knowledge predict new populations of *Gymnocladus assamicus*, a critically endangered tree species. *Endangered Species Research*, **11**, 175–181.

Menon, S., Khan, M.L., Paul, A.; Peterson, A.T. (2012) Rhododendron species in the eastern Himalayas: new approaches to understanding rare plant species distributions. Grand Valley State University, Biology Department. *Peer Reviewed Publications*, **38**, 78–84.

Miehe, G. (1991) Die Vegetationskarte des Khumbu Himal (Mt. Everest-Südabdachung) 1:50.000. Gefügemuster der Vegetation und Probleme der Kartierung. *Erdkunde*, **45**, 81–94.

Miehe, G.; Miehe, S.; Vogel, J.; Co, S.; La, D. (2007): Highest Treeline in the Northern Hemisphere Found in Southern Tibet. *Mountain Research and Development*, **27**, 169–173.

Miehe, G.; Miehe, S.; Böhner, J.; Bäumler, R.; Ghimire, S.K.; Bhattarai, K.; Chaudhary, R.P.; Subedi, M.; Jha, P.K.; Pendry, C. (2015a) Vegetation ecology. *Nepal: an introduction to the natural history, ecology and human environment of the Himalayas.* (ed. by G. Miehe; C.A. Pendry and R. Chaudhary R). pp. 385–472. Royal Botanic Garden Edinburgh, Edinburgh.

Miehe, G.; Pendry ,C.A.; Chaudhary, R. (2015b) *Nepal: an introduction to the natural history, ecology and human environment of the Himalayas*. Royal Botanic Garden Edinburgh, Edinburgh.

Müller, M., Schickhoff, U., Scholten, T., Drollinger, S., Böhner, J. & Chaudhary, R.P. (2016a) How do soil properties affect alpine treelines? General principles in a global perspective and novel findings from Rolwaling Himal, Nepal. *Progress in Physical Geography* **40**, 1-26.

Müller, M., Schwab, N., Schickhoff, U., Böhner, J. & Scholten, T. (2016b) Soil temperature and soil moisture patterns in a Himalayan alpine treeline ecotone. *Arctic, Antarctic, and Alpine Research* **48,** 501-521.

Naimi, B.; Hamm, N.A.S.; Groen, T.A.; Skidmore, A.K.; Toxopeus, A.G. (2013) Where is positional uncertainty a problem for species distribution modelling? *Ecography*, **36**, 1–13.

Nelder, J.A.; Wedderburn, R.W.M. (1972) Generalized linear models. *Journal of the Royal Statistic Society A*, **135**, 370–384.

Nüsser, M.; Dickoré, W.B. (2002) A tangle in the triangle: Vegetation map of the eastern Hindukush (Chitral, northern Pakistan). *Erdkunde*, **56**, 37–59.

Panday, P.K.; Ghimire, B. (2012) Time-series analysis of NDVI from AVHRR data over the Hindu Kush-Himalayan region for the period 1982–2006. In *International Journal of Remote Sensing*, **33**, 6710–6721.

Parmesan, C. (2006): Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review* of *Ecology*, *Evolution, and Systematics*, **37**, 637–669.

Parolo, G.; Rossi, G.; Ferrarini, A. (2008) Toward improved species niche modelling: *Arnica montana* in the Alps as a case study. *Journal of Applied Ecology*, **45**, 1410–1418.

Parra, J.L.; Graham, C.C.; Freile, J.F. (2004) Evaluating alternative data sets for ecological niche models of birds in the Andes. *Ecography*, **27**, 350–360.

Pauli, H.; Gottfried, M.; Dullinger, S.; Abdaladze, O.; Akhalkatsi, M.; Benito Alonso, J.L.; Coldea, G.; Dick, J.; Erschbamer, B.; Fernández Calzado, R.; Ghosn, D.; Holten J.I.; Kanka, R.; Kazakis, G.; Kollár, J., Larsson, P.; Moiseev, P.; Moiseev, D.; Molau, U.; Molero Mesa, J.; Nagy, L.; Pelino, G.; Puşcaş, M.; Rossi, G.; Stanisci, A.; Syverhuset, A.O.; Theurillat, J.P.; Tomaselli, M.; Unterluggauer, P.; Villar, L.; Vittoz, P.; Grabherr, G. (2012) Recent plant diversity changes on Europe´s mountain summits. *Science*, **336**, 353–355.

Paulsen, J.; Körner, C. (2014) A climate-based model to predict potential treeline position around the globe. *Alpine Botany*, **124**, 1–12.

Pearman, P.B.; Guisan, A.; Broennimann, O.; Randin, C.F. (2008) Niche dynamics in space and time. *Trends in Ecology and Evolution*, **23**, 149–158.

Pearson, R.G.; Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?. *Global Ecology and Biogeography*, **12**, 361–371.

Peterson A.T.; Soberón, J.; Pearson, R.G.; Anderson, R.P.; Martínez-Meyer, E.; Nakamura, M.; Araújo, M.B. (2011*) Ecological niches and geographic distibutions*. Princeton University Press, Princeton.

Peterson, A.T.; Soberón, J. (2012) Species distribution modeling and ecological niche modeling: getting the concepts right. *Naturaleza and Conservacao*, **10**, 1–6.

Peterson, A.T.; Papeş, M.; Soberón, J. (2015) Mechanistic and Correlative Models of Ecological Niches. *European Journal of Ecology*, **1**, 28–38.

Polunin, O.; Stainton, A. (1984) *Flowers of the Himalaya*. Oxford University Press, New Delhi.

Puri, G.S.; Gupta, R.K.; Meher-Homji, V.M.; Puri, S. (1989) *Forest ecology*. Vol. II: *Plant form, diversity, communities and succession*. Oxford & IBH Publishing, New Delhi-Bombay-Calcutta.

R Core Team, 2015, version: 3.1.3 R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/.

Rai, I.D.; Bharti, R.R.; Adhikari B.S.; Rawat, G.S. (2013) Structure and functioning of timberline vegetation in the Western Himalaya: a Case Study. *High–Altitude Rangelands and their interfaces in the Hindu Kush Himalayas* (ed. by W. Ning; G.S. Rawat; S. Joshi; M. Ismail and E. Sharma), pp. 91–106. ICIMOD Nepal, Kathmandu.

Randin C.F.; Engler, R.; Normand, S.; Zappa, M.; Zimmermann, N.E.; Pearman, P.B.; Vittoz, P.; Thuiller, W.; Guisan, A. (2009) Climate change and plant distribution: local models predict high-elevation persistence. *Global Change Biology*, **15**, 1557–1569.

Ranjitkar, S.; Kindt, R.; Sujakhu, N.M.; Hart, R.; Guo, W.; Yang, X.; Shrestha, K.K., Xu, J.; Luedeling, E. (2014) Separation of the bioclimatic spaces of Himalayan tree rhododendron species predicted by ensemble suitability models. *Global Ecology and Conservation*, **1**, 2–12.

Schibalski, A.; Lehtonen, A.; Schröder, B. (2014) Climate change shifts environmental space and limits transferability of treeline models. *Ecography*, **37**, 321–335.

Schickhoff, U. (1993) *Das Kaghan-Tal im Westhimalaya (Pakistan). Studien zur landschaftsökologischen Differenzierung und zum Landschaftswandel mit vegetationskundlichem Ansatz.* Bonner Geographische Abhandlungen 87, Dümmlers, Bonn.

Schickhoff, U. (1994) Die Verbreitung der Vegetation im Kaghan-Tal (Westhimalaya, Pakistan) und ihre kartographische Darstellung im Maßstab 1:150.000. *Erdkunde*, **48**, 92–110.

Schickhoff, U. (2005) The upper timberline in the Himalaya, Hindu Kush and Karakorum: a review of geographical and ecological aspects. *Mountain ecosystems. Studies in treeline ecology* (ed. by G. Broll and B. Keplin), pp. 275–354. Springer, Berlin-Heidelberg.

Schickhoff, U. (2011) Dynamics of mountain ecosystems. *Handbook of biogeography* (ed. by A. Millington; M. Blumler and U. Schickhoff), pp. 313–337. Sage Publications, London.

Schickhoff, U.; Bobrowski, M.; Böhner, J.; Bürzle, B.; Chaudari, R.P.; Gerlitz, L.; Heyken, H.; Lange, J.; Müller, M.; Scholten, T.; Schwab, N.; Wedegärtner, R. (2015) Do Himalayan treelines respond to recent climate change? An evaluation of sensitivity indicators. *Earth System Dynamics*, **6**, 245–265.

Schickhoff, U.; Bobrowski, M.; Böhner, J.; Bürzle, B.; Chaudhary, R.P.; Gerlitz, L.; Lange, J.; Müller, M.; Scholten, T.; Schwab, N. (2016a) Climate change and treeline dynamics in the Himalaya. *Climate Change, Glacier Response, and Vegetation Dynamics in the Himalaya* (ed. by R.B. Singh; U. Schickhoff and S. Mal), pp. 271–306. Springer, Switzerland.

Schickhoff, U.; Singh, R.B.; Mal, S. (2016b) Climate change and dynamics of glaciers and vegetation in the Himalaya: an overview. *Climate Change, Glacier Response, and Vegetation Dynamics in the Himalaya* (ed. by R.B. Singh; U. Schickhoff and S. Mal), pp. 1–26. Springer, Switzerland.

Schorr, G.; Holstein, N.; Pearman, P.B.; Guisan, A.; Kadereit, J.W. (2012) Integrating species distribution models (SDMs) and phylogeography for two species of Alpine *Primula*. *Ecology and Evolution*. **2**, 1260–1277.

Schwab, N.; Schickhoff, U.; Bobrowski, M.; Böhner, J.; Bürzle, B.; Chaudhary, R.P.; Gerlitz, L.; Müller, M.; Scholten, T. (2016) Treeline responsiveness to climate warming: Insights from a krummholz treeline in Rolwaling Himal, Nepal. *Climate Change, Glacier Response, and Vegetation Dynamics in the Himalaya* (ed. by R.B. Singh, U. Schickhoff and S. Mal), pp. 307–345. Springer, Switzerland.

Schweinfurth, U. (1957) *Die horizontale und vertikale Verbreitung der Vegetation im Himalaya*. Bonner Geographische Abhandlungen 20, Dümmlers, Bonn.

Schwörer, C.; Kaltenrieder, P.; Glur, L.; Berlinger, M.; Elbert, J.; Frei, S.; Gilli, A.; Hafner, A.; Anselmetti, F.S.; Grosjean, M.; Tinner, W. (2014) Holocene climate, fire and vegetation dynamics at the treeline in the Northwestern Swiss Alps. *Vegetation History and Archaeobotany*, **23**, 479–496.

Shi, P.; Körner, C.; Hoch, G. (2008) A test of the growth-limitation theory for alpine tree line formation in evergreen and deciduous taxa of the eastern Himalayas. *Functional Ecology*, **22**, 213–220.

Shrestha, B.B.; Ghimire, B.; Lekhak, H.D.; Jha, P.K. (2007): Regeneration of treeline birch (*Betula utilis* D. Don) forest in a Trans-Himalayan dry valley in Central Nepal. *Mountain Research and Development*, **27**, 259–267.

Shrestha, U.B; Gautam, S.; Bawa, K.S. (2012): Widespread climate change in the Himalayas and associated changes in local ecosystems. *PLoS ONE*, **7**, e36741.

Shrestha, K.B.; Hofgaard, A.; Vandvik, V. (2014) Recent treeline dynamics are similar between dry and mesic areas of Nepal, Central Himalaya. *Journal of Plant Ecology*, **8**, 347–358.

Shrestha, U.B.; Bawa, K.S. (2014) Impact of climate change on potential distribution of Chinese Caterpillar Fungus (*Ophiocordyceps chinensis*) in Nepal Himalaya. *PLoS ONE*, **9**, e106405.

Singh, C.P.; Panigrahy, S.; Thapliyal, A.; Kimothi, M.M.; Soni, P.; Parihar, J. S. (2012) Monitoring the alpine treeline shift in parts of the Indian Himalayas using remote sensing. *Current Science*, **12**, 559–562.

Singh, C.P.; Panigrahy, S.; Parihar, J.S.; Dharaiya N (2013) Modeling environmental niche of Himalayan birch and remote sensing based vicarious validation. *Tropical Ecology*, **54**, 321–329.

Smith, William K.; Germino, M. J.; Johnson, D. M.; Reinhardt, K. (2009): The Altitude of Alpine Treeline: A Bellwether of Climate Change Effects. *The Botanical Review,* **75**, 163–190.

Smith, J.; Sconiers, W.; Spasojevic, M.; Ashton, I.; Suding, K. (2012) Phenological changes in alpine plants in response to increased snowpack, temperature, and nitrogen. *Arctic, Antarctic, and Alpine Research*, **44**, 135–142.

Soberón, J.; Nakamura, M. (2009) Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences*, **106**, 19644–19650.

Soberon, J.; Peterson, A.T. (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 689–698.

Soberon, J.; Peterson, A.T. (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2**, 1–10.

Soberón, J.; Peterson, A.T. (2011). Ecological niche shifts and environmental space anisotropy: a cautionary note. *Revista Mexicana de Biodiversidad*, **82**, 1348–1355.

Soria-Auza, R.; Kessler, M.; Bach, K.; Barajas-Barbosa, P.; Lehnert, M.; Herzog, S.; 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**, 1221–1229.

Speed, J.D.M.; Austrheim, G.; Hester, A.J.; Mysterud, A. (2011) Growth limitation of mountain birch caused by sheep browsing at the altitudinal treeline. *Forest Ecology and Management*, **261**, 1344–1352.

Stainton, J.D.A. (1972) *Forests of Nepal*. Hafner, New York.

Still, C.J.; Pau, S.; Edwards, E.J. (2014) Land surface skin temperature captures thermal environments of C3 and C4 grasses. *Global Ecology and Biogeography*, **23**, 286–296.

Telwala, Y.; Brook, B.W.; Manish, K.; Pandit M.K. (2013) Climate-induced elevational range shifts and increase in plant species richness in a Himalayan biodiversity epicentre. *PLoS ONE*, **2**, e57103.

Tenca, A.; Carrer, M. (2010) Growth climate response at high elevation: comparing Alps and Himalayas. *TRACE – Tree Rings in Archaeology, Climatology and Ecology*, Vol. 8 (ed. by T. Levanic; J. Gricar; P. Hafner; R. Krajnc; S. Jagodic; H. Gärtner; I. Heinrich and G. Helle), pp. 89–97. Scientific Technical Report 10/05, GFZ German Research Centre for Geosciences, Potsdam.

Thuiller, W. (2004) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, **10**, 2220–2227.

Thuiller W.; Araújo, M.B.; Lavorel, S. (2004) Do we need land-cover data to model species distributions in Europe*? Journal of Biogeography*, **31**, 353–361.

Thuiller,W.; Lavorel, S.; Araujo, M.B. (2005) Niche properties and geographical extent as predictors of species sensitivity to climate change. *Global Ecology and Biogeography*, **14**, 347–357.

Troll, C. (1939) Das Pflanzenkleid des Nanga Parbat. Begleitworte zur Vegetationskarte der Nanga-Parbat-Gruppe (Nordwest-Himalaja) 1:50.000. *Wissenschaftliche Veröffentlichungen des Deutschen Museums für Länderkunde zu Leipzig N.F.*, **7**, 151–193.

Troll, C. (1972) The three-dimensional zonation of the Himalayan system. Geoecology of the high-mountain regions of Eurasia. *Erdwissenschaftliche Forschung IV* (ed. by C. Troll), pp. 264–275. Franz Steiner Verlag, Wiesbaden.

Troll, C. (1975) Vergleichende Geographie der Hochgebirge der Erde in landschaftsökologischer Sicht. *Geographische Rundschau,* **27**, 185–198.

Truong, C.; Palmé, A.E.; Felber, F. (2007) Recent invasion of the mountain birch *Betula pubescens ssp. tortuosa* above the treeline due to climate change: genetic and ecological study in Northern Sweden. *Journal of Evolutionary Biology*, **20**, 369–380.

USGS (2004) Shuttle Radar Topography Mission, 1 Arc Second scene SRTM_u03_n008e004, Unfilled Unfinished 2.0, Global Land Cover Facility, University of Maryland, College Park, Maryland, February 2000.

VanderWal, J.; Shoo, L.P.; Graham, C.; William S.E. (2009) Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecological Modelling*, **220**, 589–594.

Veloz, S.D. (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, **36**, 2290–2299.

Wang Q.; Fan X.; Wang M. (2014) Recent warming amplification over high elevation regions across the globe. *Climate Dynamics*, **43**, 87–101.

Wang, X.; Gao, Q.; Wang, C.; Yu, M. (2017): Spatiotemporal patterns of vegetation phenology change and relationships with climate in the two transects of East China. *Global Ecology and Conservation*, **10**, 206–219.

Wieser, G.; Holtmeier, F.-K.; Smith, W.K. (2014) Treelines in a changing global environment. *Trees in a Changing Environment* (ed. by M. Tausz and N. Grulke). pp. 221–263. Springer, Dordrecht.

Xu, J.; Grumbine, R.E.; Shrestha, A.; Eriksson, M.; Yang, X.; Wang, Y.U.N.; Wilkes, A. (2009) The Melting Himalayas: Cascading Effects of Climate Change on Water, Biodiversity, and Livelihoods. *Conservation Biology*, **23**, 520–530.

Xu, X.; Piao, S.; Wang, X.; Chen, A.; Ciais, P.; Myneni, R.B. (2012) Spatio-temporal patterns of the area experiencing negative vegetation growth anomalies in China over the last three decades. *Environmental Research Letters*, **7**, 1–9.

Yasaka, M. (2005) The Pollen Production and Dispersal of *Betula platyphylla var. japonica* and *B. ermanii*. *Journal of the Japanese Forest Society*, **87**.

Zurick, D.; Pacheco, J. (2006) *Illustrated atlas of the Himalaya.* The University Press of Kentucky, Lexington.

List of publications

Original publications in the framework of the PhD thesis

Bobrowski, M.; Gerlitz, L.; Schickhoff, U. (2017) Modelling the potential distribution of *Betula utilis* in the Himalaya. *Global Ecology and Conservation*, **11**, 69–83. doi: 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. doi: 10.1016/j.ecolmodel.2017.05.021.

Bobrowski, M.; Bechtel, B.; Böhner, J.; Oldeland, J.; Weidinger, J.; Schickhoff, U. (2018) Upgrading ecological niche models with phenological traits: Refinement of the predicted distribution range of *Betula utilis* in the Himalayan region. *Remote Sensing*, **10**, 814; doi:10.3390/rs10060814.

Additional publications

Gillich, B.; Bobrowski, M.; Stolter, C. Different effects on the habitat utilization of red deer (*Cervus elaphus*) and roe deer (*Capreolus capreolus*) in winter (*in prep.*).

Heider, K.; Weinzierl, T.; Schwab, N.; Bobrowski, M.; Schickhoff, U. Future agricultural conditions in the Himalayan Region - A fuzzy logic approach using high resolution climate scenarios. Submitted to *Die Erde* (*under review*).

Dormann, C.F.; Bobrowski, M.; Dehling, M.; Harris, D.J.; Hartig, F.; Lischke, H.; Moretti, M.D.; Pagel, J.; Pinkert, S.; Schleuning, M.; Schmidt, S.; Sheppard, C.; Steinbauer, M.J.; Zeuss, D.; Kraan, C. (2018) Biotic interactions in species distribution modelling: ten questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography*. doi: 10.1111/geb.12759. (in press).

Gerlitz L.; Bechtel B.; Böhner J.; Bobrowski M.; Bürzle B.; Müller M.; Scholten, T.; Schickhoff, U.; Schwab, N.; Weidinger J. (2016) Analytic comparison of temperature lapse rates and precipitation gradients in a Himalayan treeline environment: Implications for statistical downscaling. In: Singh, R. B.; Schickhoff, U. & Mal, S. (Eds.): Climate Change, Glacier Response, and Vegetation Dynamics in the Himalaya. Springer, Switzerland; 49–64. doi: 10.1007/978-3-319-28977-9_3.

Schickhoff, U.; Bobrowski, M.; Böhner, J.; Bürzle, B.; Chaudhary, R.P.; Gerlitz, L.; Lange, J.; Müller, M.; Scholten, T.; Schwab, N. (2016): Climate change and treeline dynamics in the Himalaya. In: Singh, R.B.; Schickhoff, U. and Mal, S. (Eds.): Climate Change, Glacier Response, and Vegetation Dynamics in the Himalaya. Springer, Switzerland: 271–306. doi: 10.1007/978-3-319-28977-9_15.

Bobrowski, M.; Gillich, B.; Stolter, C. (2015) Modelling browsing of deer on beech and birch in northern Germany. *Forest Ecology and Management*, **358**; 212–221. doi: 10.1016/j.foreco.2015.08.031.

Schickhoff, U.; Bobrowski, M.; Böhner, J.; Bürzle, B.; Chaudhary, R.P.; Gerlitz, L.; Heyken, H.; Lange, J.; Müller, M.; Scholten, T.; Schwab, N.; Wedegärtner, R. (2015): Do Himalayan treelines respond to recent climate change? An evaluation of sensitivity indicators. *Earth System Dynamics*, **6**, 245–265. doi: 10.5194/esd-6-245-2015.

List of oral presentations

Arbeitskreis Hochgebirge, 2018: Bobrowski, M.; Schickhoff, U. Improving ecological niche models with remotely sensed data for modelling the current distribution of *Betula utilis* in the Himalayan region. Innsbruck.

Arbeitskreis Biogeographie, 2017: Bobrowski, M.; Schickhoff, U. Why inputs matter: Selection of climatic variables for species distribution modelling in the Himalayan region. Erlangen.

Gesellschaft für Ökologie, 2016: Bobrowski, M.; Gillich, B.; Stolter, C. Modelling feeding habitat selection by cervids in regard to forest conversion. Marburg.

Arbeitskreis Biogeographie, 2015: Bobrowski, M.; Gerlitz, L.; Schickhoff, U. Why inputs matter: Selection of climatic variables for species distribution modelling in the Himalayan region. Basel.

88th Annual Meeting of the German Society for Mammalian Biology, 2014: Bobrowski, M.; Gillich, B.; Stolter, C. Modelling feeding habitat selection by cervids in regard to forest conversion. Gießen. doi: 10.13140/2.1.1425.2481.

Conference Paper

European Geosciences Union, 2017: Bobrowski, M.; Schickhoff, U. Why inputs matter: Selection of climatic variables for species distribution modelling in the Himalayan region. Wien.

Gesellschaft für Ökologie, 2016: Bobrowski, M.; Gerlitz, L.; Schickhoff, U. Modelling the potential distribution of *Betula utilis* in the Himalyan mountains. In: Röder J, Brandl R (eds.): 150 years of Ecology - lessons for the future. 46th Annual Meeting of the Ecological Society of Germany, Austria and Switzerland. 5–9 September 2016. Marburg, Germany: 59. ISSN 0171-1113.

Tagung der Vereinigung der Wildbiologien und Jagdwissenschaftler Deutschlands, 2016: Stolter, C.; Gillich, B.; Bobrowski, M. Reh- und Rotwild in der Lüneburger Heide: Habitatnutzung und Nahrungswahl unter menschlichem Einfluss - Habitat Utilisation and Food Selection of Roe and Red Deer under Consideration of Human Impacts. Trippstadt.

International Biogeography Society, 2015: Bobrowski, M. Modelling potential distribution of *Betula utilis* in the Himalayan region under climate change. In: Gavin D, Beierkuhnlein, C.; Holzheu, S.; Thies, B.; Faller, K.; Gillespie, R.; Hortal, J. (Eds.): Conference program and abstracts. International Biogeography Society 7th Biennial Meeting. 8–12 January 2015, Bayreuth. Frontiers of Biogeography Vol. 6, suppl. 1: 108. doi: 10.13140/2.1.4880.1600.

Arbeitskreis Hochgebirge, 2015: Bobrowski, M.; Schickhoff, U. Modelling distribution ranges of *Betula utilis* in the Himalayan region. Hamburg.

88th Annual Meeting of the German Society for Mammalian Biology, 2014: Gillich, B.; Bobrowski, M.; Stolter, C. Determination of habitat choice by red (*Cervus elaphus*) and roe deer (*Capreolus capreolus*) using faecal counts. Gießen.

Gesellschaft für Ökologie, 2013: Gillich, B.; Bobrowski, M.; Stolter, C. Faecal pellet count as a tool for estimating habitat choice of red and roe deer. Potsdam.

Appendix

Original publications

Article I

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

Abstract

Developing sustainable adaptation pathways under climate change conditions in mountain regions requires accurate predictions of treeline shifts and future distribution ranges of treeline species. Here, we model for the first time the potential distribution of *Betula utilis*, a principal Himalayan treeline species, to provide a basis for the analysis of future range shifts. Our target species *Betula utilis* is widespread at alpine treelines in the Himalayan mountains, the distribution range extends across the Himalayan mountain range . Our objective is to model the potential distribution of *B. utilis* in relation to current climate conditions. We generated a dataset of 590 occurrence records and used 24 variables for ecological niche modelling. We calibrated Generalized Linear Models using the Akaike Information Criterion (AIC) and evaluated model performance using threshold-independent (AUC, Area Under the Curve) and threshold-dependent (TSS, True Skill Statistics) characteristics as well as visual assessments of projected distribution maps. We found two temperature-related (Mean Temperature of the Wettest Quarter, Temperature Annual Range) and three precipitation-related variables (Precipitation of the Coldest Quarter, Average Precipitation of March, April and May and Precipitation Seasonality) to be useful for predicting the potential distribution of *B. utilis*. All models had high predictive power ($AUC \geq 0.98$ and $TSS \geq 0.89$). The projected suitable area in the Himalayan mountains varies considerably, with most extensive distribution in the western and central Himalayan region. A substantial difference between potential and real distribution in the eastern Himalaya points to decreasing competitiveness of *B. utilis* under more oceanic conditions in the eastern part of the mountain system. A comparison between the vegetation map of Schweinfurth (1957) and our current predictions suggests that *B. utilis* does not reach the upper elevational limit in vast areas of its potential distribution range due to anthropogenically caused treeline depressions. This study underlines the significance of accuracies of current environmental niche models for species distribution modelling under climate change scenarios. Analysing and understanding the environmental factors driving the current distribution of *B. utilis* is crucial for the prediction of future range shifts of *B. utilis* and other treeline species, and for deriving appropriate climate change adaptation strategies.

Keywords: climatic space; ecological niche modeling; habitat, range shift; treeline dynamics; treeline ecotone

Introduction

Ecological niche models (ENMs) and species distribution models (SDMs) based on presence-only occurrence data are significant research tools in biogeography, ecology, evolution and conservation biology (Guisan & Thuiller, 2005 and references therein). Using modelling techniques requires the analysis of underlying climatic factors of species distributions, which is, in turn, a basic requirement for the understanding of current and future species distribution ranges. Shifts of the environmental niches of alpine treeline species under future climate change scenarios are of particular interest in this respect since high mountain environments are subjected to above-average warming rates (Schickhoff, 2011; IPCC, 2014), and treelines are expected to advance to higher elevations (Harsch *et al.*, 2009; Wieser *et al.*, 2014).

Linking the occurrence or abundance of a species to environmental and/or geographical variables has been shown to be a valuable tool to predict species distribution ranges or potential suitable habitats (Franklin, 1995; Austin, 2002; Barry & Elith, 2006).

In this study, we follow the principles of ecological niche modelling regarding the conditions suitable for a target species based on calibration, evaluation and interpretation (Peterson *et al.*, 2011; Anderson, 2012). In contrast to SDMs, which intend to characterize the species' occupied distribution (*sensu stricto*), ENMs investigate relationships between known species occurrences and abiotic environmental (frequently climatic) variables in a particular region of interest (the species' abiotically suitable distribution; see Peterson & Soberón, 2012, for the distinction of ENM and SDM terminology). A variety of statistical approaches are incorporated in ENMs (for details see Guisan & Zimmermann, 2000; Thuiller *et al.*, 2008). Basically, the model estimates the environmental conditions that a species can live in (the species' existing realized niche). Subsequently the model can be projected in geographic space, identifying consistent areas with suitable environmental conditions for the species (Araújo & Guisan, 2006; Peterson & Soberón, 2012).

Accuracies of the models under climate change scenarios (i.e., response of the species to changing climate conditions) depend on the accuracies of models of the current environmental niche of the species, the importance of which should not be underestimated. Model predictions may then be used for biodiversity conservation in order to implement appropriate ecosystem management strategies.

At local and global scales, high-elevation climatic treelines can be considered as sensitive indicators of past and recent climate change and variability (Kullman, 1998; Holtmeier, 2009; Körner, 2012). During the last decade, investigation of climate change-driven treeline dynamics has generated considerable research interest, and results have been widely reported from various treelines of the world (e.g., Randin *et al.*, 2009; Harsch *et al.*, 2009; Paulsen & Körner, 2014, Schibalski *et al.*, 2014; Schickhoff *et al.*, 2015, 2016). One widespread hypothesis is that global warming will shift the climatic ranges of treeline and subalpine/alpine species upward along altitudinal, thermally defined gradients (Gottfried *et al.*, 2012; Pauli *et al.*, 2012).

Modelling the distributional range of treeline species and predicting changes under future climate scenarios has become an increasingly applied component in investigations of high altitude treelines (e.g., Dullinger *et al.*, 2004; Thuiller *et al.*, 2005; Parolo *et al.*, 2008). Climate-induced future range shifts have been postulated for both high-altitude and high-latitude treeline species as a consequence of rapid warming rates (Holtmeier & Broll, 2010; Dufour-Tremblay *et al.*, 2012; Gaire *et al.*, 2014; Shrestha *et al.*, 2014). In contrast to other mountains of the world, the Himalayan region has been largely neglected in this respect and is clearly under-represented in scientific literature on climate change-induced species range shifts (Schickhoff, 2005; Telwala *et al.*, 2013, Dutta *et al.*, 2014; Schickhoff *et al.*, 2015).

Moreover, most high altitude and high latitude treeline studies investigated coniferous tree species (e.g., *Abies*, *Picea*, *Pinus*, *Larix*), while deciduous tree species (e.g., *Betula*) remained largely out of focus. The genus *Betula* is known to be widely distributed in the Northern Hemisphere and birches are among the conspicuous broad-leaved high altitude and high latitude treeline species (Truong *et al.*, 2007; Speed *et al.*, 2011). *Betula* spp. can be found at polar treelines (subarctic and boreal zone) as well as at alpine treelines (cool and warm temperate zones) (Holtmeier, 2009). In Europe, *Betula* species were target species in several modelling studies (e.g., Thuiller, 2003 on *B. nana*; Svenning *et al.*, 2008 on *B. pendula* and *B. pubescens*). Our target species *Betula utilis* is widespread at alpine treelines in the Himalayan mountains (Ashburner & McAllister, 2013), and is considered to be an indicator species for climate-driven treeline dynamics (Liang *et al.*, 2014). Presently, only few researchers have addressed the problem of modelling distribution ranges of deciduous treeline species. This applies in particular to the Himalaya. Up to date, modelling studies were either local studies on *B. utilis* (e.g., Singh *et al.*, 2013: Indian Himalaya, Uttarakhand; Huo *et al.*, 2010: SW China) or focused on other species (Kumar, 2012 on *Rhododendron* spp.; Ranjitkar *et al.*, 2014 on *Oxybaphus himalaicus* and *Boerhavia diffusa*). Previous work on *B. utilis* focused mainly on dendroecology (Bhattacharyya *et al.*, 2006; Tenca & Carrer, 2010, Dawadi *et al.*, 2013; Gaire *et al.*, 2013; Liang *et al.*, 2014), remote sensing (Singh *et al.*, 2012, 2013; Rai *et al.*, 2013), plant physiology (Xu *et al.*, 2012; Shi *et al.*, 2008), and regeneration (Shrestha *et al.*, 2007).

Thus, modelling the distribution of treeline species such as *B. utilis*, a predominant species in western and central Himalayan treeline ecotones (Schweinfurth, 1957; Schickhoff, 2005), is still a major research deficit. This study models for the first time the potential distribution of *B. utilis* in the subalpine and alpine treeline ecotone of the Himalayan mountains on a broad scale, based on underlying climatic determinants of its distribution range. We aim at assessing the performance of Generalized Linear Models (GLM), implementing a multi-step evaluation approach: i) calibration of the model, ii) evaluation of spatial prediction success and accuracy of the models,

iii) projecting the distribution of *B. utilis* under current climate conditions. Using this approach, we seek to answer the following key questions:

1. **Are selected climatic variables able to reflect the climatic gradient from more continental western to more oceanic eastern regions of the Himalayan mountain range?**
2. **What are key climatic constraints for the current distribution of *B. utilis* in the Himalayan region?**
3. **Which areas are predicted as potentially suitable for *B. utilis*?**
4. **Are there differences between the current model predictions and the distribution range delineated in the detailed vegetation map of Schweinfurth (1957)?**

Material and Methods

Study area

The Himalayan mountain range is located between the Indian Subcontinent in the south and the Tibetan Highland in the north, and extends from Afghanistan in the northwest (c. 36°N and 70°E) to Yunnan in the southeast (c. 26°N and 100° E). It is a vast mountain region, covering an area of more than 1.000.000 km^2, with a length of c. 3000 km (Pakistan to SW China) and a maximum width of 400 km.

The Himalayan mountains show a distinct three-dimensional geoecological differentiation, with a high variation of climate, rainfall, altitude, and soils (Troll, 1972; Zurick *et al.*, 2006; Miehe *et al.*, 2015a). The climate ranges from tropical in the Indian lowlands to permanent ice and snow at the highest elevations, and from more continental in the NW to more oceanic in the SE. The amount of annual precipitation increases with increasing monsoonal influence in the same direction along the southern front of the range (Schickhoff, 2005; Böhner *et al.*, 2015). The small-scale heterogeneity of habitats and site conditions supports a high diversity of species and communities.

Target species

The distribution range of *Betula utilis* extends across the Himalayan range from Afghanistan to SW China, with the total elevational range extending from 2700 to 4500 m (Polunin & Stainton, 1984). *B. utilis* was selected as a study species due to its status as a principal broadleaved treeline species in the western and central Himalayan mountains (for photographs see Supplementary material S1, Fig. S1a-c). This species mainly grows on north-facing slopes in shady locations. In the NW Himalaya, *B. utilis* is widely distributed in the elevational range between 3100 and 3700 m, while the range shifts to higher altitudes towards the E Himalaya (mainly between 3800 and

4300 m; own database). *B. utilis* grows in mixed forests with conifers and rhododendrons and forms a narrow forest belt between coniferous forests below and a krummholz belt above (for associated tree species see Schickhoff, 2005, Miehe *et al.*, 2015b). Pure birch stands with Rhododendron campanulatum and Sorbus microphylla in the understory are often found at the uppermost limit of subalpine forests (Schickhoff *et al.*, 2015).

Species data collection

Presence-only occurrence data of *B. utilis* were accessed via the Global Biodiversity Information Facility (gbif.org). The database hosts 215 geo-referenced records (1980-2016) without any known coordinate issues for our study region. 202 records were added from a database compiled from a literature survey (Schickhoff, 2005; unpublished data). Additionally, we extracted 327 records from freely available satellite images (GoogleEarth™, ver. 7.1.1.1888, Google, 2015) and added them to the dataset. This method has been proven to be useful in global treeline research (Paulsen & Körner, 2014; Irl *et al.*, 2015). These occurrence localities were validated through expert knowledge, obtained from numerous field visits in the Himalayan region. We used a total of 744 records for further analysis.

Prior to building the distribution model of a species, error sources have to be minimized, in particular those arising from spatial autocorrelation of species occurrence data. Only one occurrence point per grid cell (1 km x 1 km) was considered, furthermore only the uppermost occurrences in the treeline ecotone were used, resulting in 590 occurrence points of *B. utilis* as a basis for modelling the current distribution of this species (Fig. 15). Lowermost occurrences (e.g., in avalanche paths) were omitted since they do not represent the 'zonal' climatic conditions of the treeline birch belt.

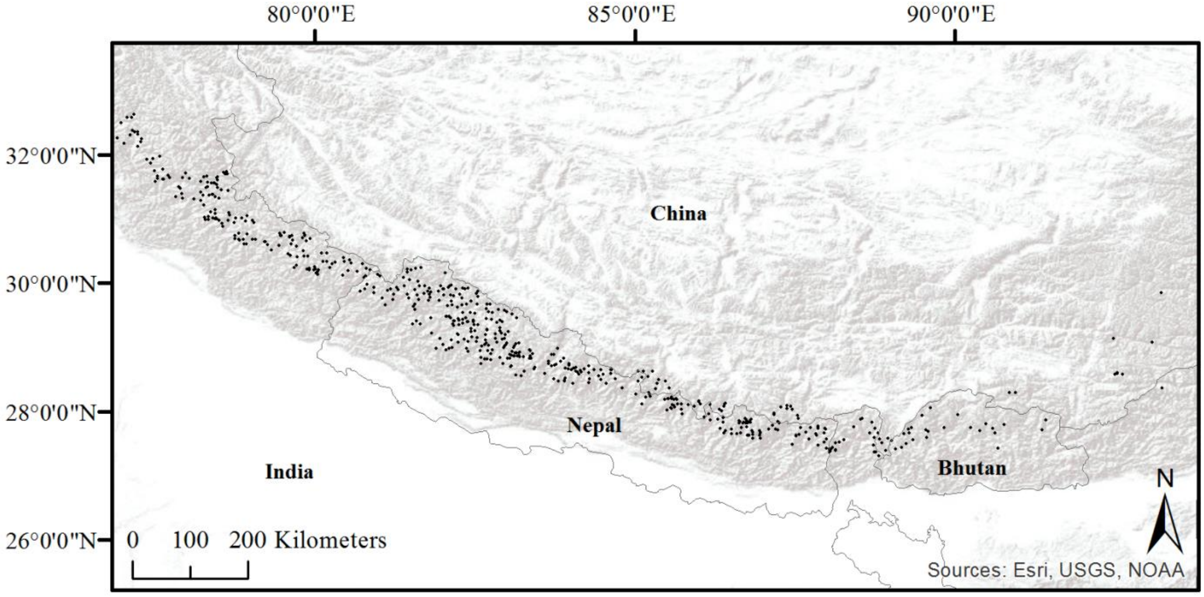


Fig. 15: Occurrences of *Betula utilis* in the Himalayan mountains (*N* = 590).

Climatic variables and statistical analyses

Ecological modelling applications frequently apply simple interpolated climate data sets such as ‘worldclim’ (Hijmans *et al.*, 2005), which regionalize monthly observations of precipitation and temperature based on a weighted linear regression approach, using latitude, longitude and elevation as statistical predictor variables. Despite of the high spatial resolution of approximately 1 km^2, these data sets neglect local scale atmospheric processes which are crucial for the formation of site specific topo-climatic conditions in high mountain environments. Many studies show that local scale atmospheric conditions are highly influenced by the underlying terrain. Anisotropic heating at different slope positions as well as cold air drainage and pooling in mountain valleys during autochthonous weather conditions result in a complex temperature pattern, which distinctly modifies the distribution of plant communities. Likewise, the spatial pattern of precipitation is influenced by wind- and leeward slope positions, resulting in hyper-humid climate conditions at the southern declivity of the Himalayan range and semi-arid to arid conditions in the Trans-Himalayan valleys. Freely available gridded climate data sets often do not satisfy the requirements of ecological climate impact studies and impede the interdisciplinary investigation of climate ecosystem interactions (Soria-Auza *et al.*, 2010).

In this study all statistical analyses were performed using the programming language R (R Core Team, 2015). All maps were created using ArcGIS (Version 10.1; Esri Inc. Redlands, CA, USA).

In order to sufficiently incorporate local scale climate variations in the framework of the presented modelling approach, we make use of two gridded long-time climate data sets (one for temperature and one for precipitation respectively), which explicitly account for important topo-climatic processes. Gridded monthly mean temperature fields with a spatial resolution of 1x1 km were derived from the ERA-Interim reanalysis by means of an elevation and bias correction approach, which considers the climate model internal stratification of the atmosphere (Gerlitz *et al.*, 2014). Compared with simple linear elevation adjustment techniques the approach was shown to considerably improve the quality of high resolution temperature fields in high mountain environments (Gerlitz *et al.*, 2014).

Monthly precipitation sums with the same resolution were downloaded from the freely available ‘chelsa’ climate dataset (Karger *et al.*, 2016). This climate data is based on a quasi-mechanistical statistical downscaling of the ERA-interim global circulation model with GPCC and GHCN bias correction. Precipitation amounts, which were estimated under consideration of orographic factors such as wind fields, valley exposition and boundary layer height showed high preciseness compared to precipitation data from other climate datasets (Karger *et al.*, 2016).

Based on the gridded monthly fields of temperature and precipitation, we generated 19 climatic variables using the R-package ‘dismo’ (Hijmans *et al.*, 2011). These climatic variables are widely used in species distribution modelling and represent annual characteristics (e.g., mean annual temperature), seasonality (e.g., annual range in precipitation) and extreme environmental factors (e.g., precipitation of driest month) (Nix, 1986, Hijmans *et al.*, 2005). In addition, average precipitation of May and of March, April and May was calculated in order to account for potential pre-monsoon drought stress (Liang *et al.*, 2014; Schickhoff *et al.*, 2015, 2016). Furthermore, a digital elevation model was extracted from GLOBE (ngdc.noaa.gov; Global Land One-Kilometer Base Elevation). We applied ArcGIS (Version 10.1; Esri Inc. Redlands, CA, USA) to calculate slope angle and slope aspect using the 'surface' function of the Spatial Analyst toolbox. The DEM-elevation as well as its derivates slope and aspect were utilized as additional predictor variables. The resolution of climatic raster grids and the DEM is 30 arc seconds, which equals 1 km x 1 km. All processing of climate data was done with R-package 'raster' (Hjimans, 2015).

Climatic and topographic predictors were chosen to reflect the species' physiological needs. Prior modelling, we examined possible correlations between all variables, since they often show high collinearity, resulting in poor model performance and misleading interpretations (Dormann *et al.*, 2013). Spearman's rank correlations were calculated using R-package 'usdm' (Naimi, 2015) to detect multi-collinearity between environmental variables, with thresholding Spearman's rank correlation coefficient $r_s \leq 0.7$ (Elith *et al.*, 2006). Correlations were visualized with R-package 'corrplot' (Wei & Simko, 2016, see Supplementary material S2 in Fig. S2). For all pairs of correlated variables, decisions were based on performance of univariate models and the results of hierarchical partitioning using the R- package 'hier.part' (Walsh & Mac Nally, 2013). Several test runs were calculated and out of 24 predictor variables only five predictors were included in the modelling procedure (Table 2). Only ecologically meaningful variables, which represent general pattern and annual variability of the climate, were included for modelling the potential distribution of *B. utilis*. We used variables (e.g., temperature of growing season and winter temperatures) which are known to reflect prevailing climate conditions at the treeline. Some variables (e.g., Temperature of Warmest Month and Temperature of Warmest Quarter) were highly correlated, in this case we decided to use variables, which cover more than one month (i.e., quarter which equals a period of three month). In all cases usage of 'Quarter' variables yielded to models with higher discriminatory power and higher variable importance.

Descriptive statistics for all variables were calculated using the R-package 'pastecs' (Grosjean & Ibanez, 2014, see Supplementary material S3, Table S3a and S3b).

Table 2: Climatic variables used for modelling the potential distribution of *Betula utilis*.

Label	Variable	Scaling Factor	Units	Used for modelling
bio1	Annual Mean Temperature	1	Degree Celsius	
bio2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	1	Degree Celsius	
bio3	Isothermality (bio2/bio7)	1	Dimensionless	
bio4	Temperature Seasonality (Standard Deviation)	100	Degree Celsius	
bio5	Max Temperature of Warmest Month	1	Degree Celsius	
bio6	Min Temperature of Coldest Month	1	Degree Celsius	
bio7	Temperature Annual Range (bio5-bio6)	1	Degree Celsius	X
bio8	Mean Temperature of Wettest Quarter	1	Degree Celsius	X
bio9	Mean Temperature of Driest Quarter	1	Degree Celsius	
bio10	Mean Temperature of Warmest Quarter	1	Degree Celsius	
bio11	Mean Temperature of Coldest Quarter	1	Degree Celsius	
bio12	Annual Precipitation	1	Millimetre	
bio13	Precipitation of Wettest Month	1	Millimetre	
bio14	Precipitation of Driest Month	1	Millimetre	
bio15	Precipitation Seasonality (Coefficient of Variation)	100	Percentage	X
bio16	Precipitation of Wettest Quarter	1	Millimetre	
bio17	Precipitation of Driest Quarter	1	Millimetre	
bio18	Precipitation of Warmest Quarter	1	Millimetre	
bio19	Precipitation of Coldest Quarter	1	Millimetre	X
prec_may	Average Precipitation May	1	Millimetre	
prec_mam	Average Precipitation March, April, May	1	Millimetre	X
alt	Altitude	1	Meters	
aspect	Aspect	1	Degree	
slope	Slope	1	Percentage	

Modelling procedure

Model algorithm

Several methods are available for constructing SDMs, and they have been comprehensively compared in terms of performance (Austin *et al.*, 2006; Tsoar *et al.*, 2007). We used multiple regression techniques (i.e., Generalized Linear Models) as they represent a classical and robust approach to analyse presence and absence data (Nelder & Wedderburn, 1972; McCullagh & Nelder, 1989), with the function glm() implemented in the R-package 'stats' (R Core Team, 2015). GLMs were calculated with binomial distribution, logit-link function and polynomial terms of second order (Austin, 1999), but did not include terms of interactions among variables. To calculate the final GLMs step-wise variable selection in both directions (i.e., forward and backward) was applied, using the Akaike Information Criterion (AIC, Akaike, 1974), thereby the final and best model possesses the lowest value for AIC (Burnham & Anderson, 2002, Guisan *et al.*, 2002).

Pseudo-absence selection

As GLMs need presence and absence points, pseudo-absence points were generated. For study area selection, a convex hull was used, covering the full extent of the known occurrences of *B. utilis* distribution in the Himalayan region. By limiting the study area, large regions where the species cannot occur were excluded in further statistical analyses, in order to prevent over-predicting the distribution range of the species (VanderWal *et al.*, 2009).

For random selection of pseudo-absences, the limits were set as 5 km from the nearest occurrence, resulting in total 6000 pseudo-absences (following the pseudo-absence selection procedure for GLMs described by Barbet-Massin *et al.*, 2012, using the 'random points' function of Data Management toolbox in ArcGIS, ESRI, 2012)

Model calibration and evaluation

Independent data collected from sites other than those used to train the model are essential to evaluate the predictive performance of the model. In order to obtain reliable estimates of model performance, models are tested with independent data (Araújo *et al.* 2005). Prior modelling, all presence and pseudo-absence points were randomly split into training and testing subsets, using the R-package 'caret' (Kuhn *et al.*, 2016), whereas training data represented 80% of the original dataset and testing data 20%. Statistical re-sampling techniques such as cross-validation (Stone, 1974) are used to decrease bias in measuring the predictive performance of the model (Pearce & Ferrier, 2000). The calibrated GLMs were validated using 10-fold cross-validation function of the R-package 'ecospat' (Broennimann *et al.*, 2015). In cross-validation, all presences and pseudo-absences are divided in *K* equal groups. Each group consists of two subsets (i.e., training set and testing set). The predicted values are used to assess the accuracy of the prediction on the withheld test data (Pearce & Ferrier, 2000). In order to reduce variability, this procedure was repeated 5 times, resulting in 5 versions of the potential species distribution that were averaged.

The final projected distribution map only contains potential presences, predicted by all 5 model versions.

Due to the lack of a universally valid model evaluation criterion, it is essential to use more than one performance evaluation metric. We chose a multi-step evaluation approach to assess the performance of the modelling algorithms: 1) threshold-independent (AUC), 2) threshold-dependent (TSS), explained variance, Pearson's correlation coefficient (r_p), explained deviance, slope and intercept of the calibration curve. Finally we conducted a visual assessment of projected distribution maps.

All models were evaluated using functions and arguments incorporated in the 'PresenceAbsence' library (Freeman & Moisen, 2008a, b).

For threshold-independent evaluation, we used the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve (Fielding & Bell, 1997; Elith & Burgman, 2002). The AUC is a non-parametric measure (i.e., no assessment of the goodness-of-fit of the model) and varies according to the suitability proportion of the study region. The AUC constitutes a measure of overall model performance (continuous discrimination between suitable and unsuitable habitat) irrespective of threshold. The AUC obtains values between 0 and 1, whereas values of > 0.9 are considered high, 0.7-0.9 moderate, 0.5-0.7 low and < 0.5 no better than random (Wiley *et al.*, 2003; Phillips *et al.*, 2006).

As threshold-dependent measure, we used True Skill Statistics (TSS) (Flueck, 1987; introduced to SDM by Allouche *et al.*, 2006). In contrast to the AUC measure, the TSS is used to convert a continuous prediction into a binary one (i.e., those areas predicted as suitable versus not suitable for the species) and provides a measure of map veracity.

TSS is defined as 1- maximum (sensitivity + specificity) where sensitivity and specificity are calculated on the probability threshold for which their sum is maximized (Table 2 and eqn. 1). The TSS value can accommodate values between 0 and 1. This has been suggested as an alternative to Cohen's Kappa (Allouche *et al.*, 2006) when a threshold-dependent measure of performance is needed. TSS responds to species prevalence differently than Cohen's Kappa, but nonetheless, the statistic has been shown to be negatively related to prevalence (Allouche *et al.*, 2006). This method outperformed other techniques in recent comparisons of threshold methods (Jiménez-Valverde & Lobo, 2007). TSS was calculated using the MaxSens+Spec argument incorporated in the 'PresenceAbsence' library (Freeman &Moisen, 2008b). Sensitivity and Specificity are derived from the confusion matrix (Table 3, Eqn. 1). Consequently, TSS assesses both omission and commission errors. In detail, sensitivity is the proportion of observed presences that are predicted as such, and therefore quantifies omission errors (false negative error rate). Specificity is the proportion of observed absences that are predicted as such, and therefore quantifies commission errors (false positive error rate). Sensitivity and specificity are independent of each other when compared across models, and are independent of prevalence ($(a + c)/n$, where n equals the total number of sites in which the species was recorded as present; Allouche *et al.*, 2006).

Table 3: A confusion matrix used to evaluate the predictive performance of presence-absence models, showing the cross-tabulation of observed and predicted values as a two-by-two contingency table.

Validation dataset

		Presence	**Absence**
Model	**Presence**	TP (true positive)	FP (false negative)
	Absence	FN (false negative	TN (true negative)

For a 2 x 2 confusion matrix TSS is defined as:

$$TSS = \frac{TP * TN - FP * FN}{(TP + FN)(FP + TN)} = Sensitivity + Specificity - 1$$

Eqn. 1

Furthermore we calculated explained variance of the training dataset. With Pearson's correlations the coefficient (r_P), between the predicted and observed values for *B. utilis* were calculated. To account for the goodness-of-fit of the models, we used explained deviance, which is calculated as the quotient of the residual deviance and the deviance of the null model subtracted from 1. For models performing better than the null model, values range from 0 to 1, whereas 1 is the value of the best model. Furthermore slope and intercept of the calibration curve were calculated to examine overfitting of the models. Optimally calibrated models feature a calibration curve with intercept 0 and slope 1 (Reineking & Schröder, 2006).

All performance measures were calculated on training and testing datasets and averaged respectively. Results yielding any presumptions of the models are shown in Table 3.

Model comparison

To compare the geographic extent and frequency distribution of the predictions in the GLMs, the consensus maps of the predicted probability of species presence for each raster cell of the study region were converted into binary "presence-absence"-maps, above their respective TSS threshold. Finally, we compared our current projection with the landmark work of Schweinfurth (1957), who provided the first detailed vegetation map of the Himalayan mountains (hereafter Schweinfurth map), in order to quantify differences between the current potential and previous real distribution of *B. utilis*. The Schweinfurth map summarizes the 1950s' state of knowledge of Himalayan vegetation distribution. It is based on an extensive literature survey, viz. on available published records up to 1956, complemented by all other reliable sources available such as unpublished diaries, personal communications, photographs, maps, etc. Despite showing some "white areas" it is still the most detailed vegetation map of the entire Himalayan region.

Results

Climatic Variables

After prior-modelling examinations (i.e., multi-collinearity), five variables were selected out of twenty four potential predictor variables for modelling the potential distribution of *B. utilis*. These variables were Temperature Annual Range, Mean Temperature of the Wettest Quarter, Precipitation Seasonality, Precipitation of Coldest Quarter and Average Precipitation of March, April and May (Fig. 16).

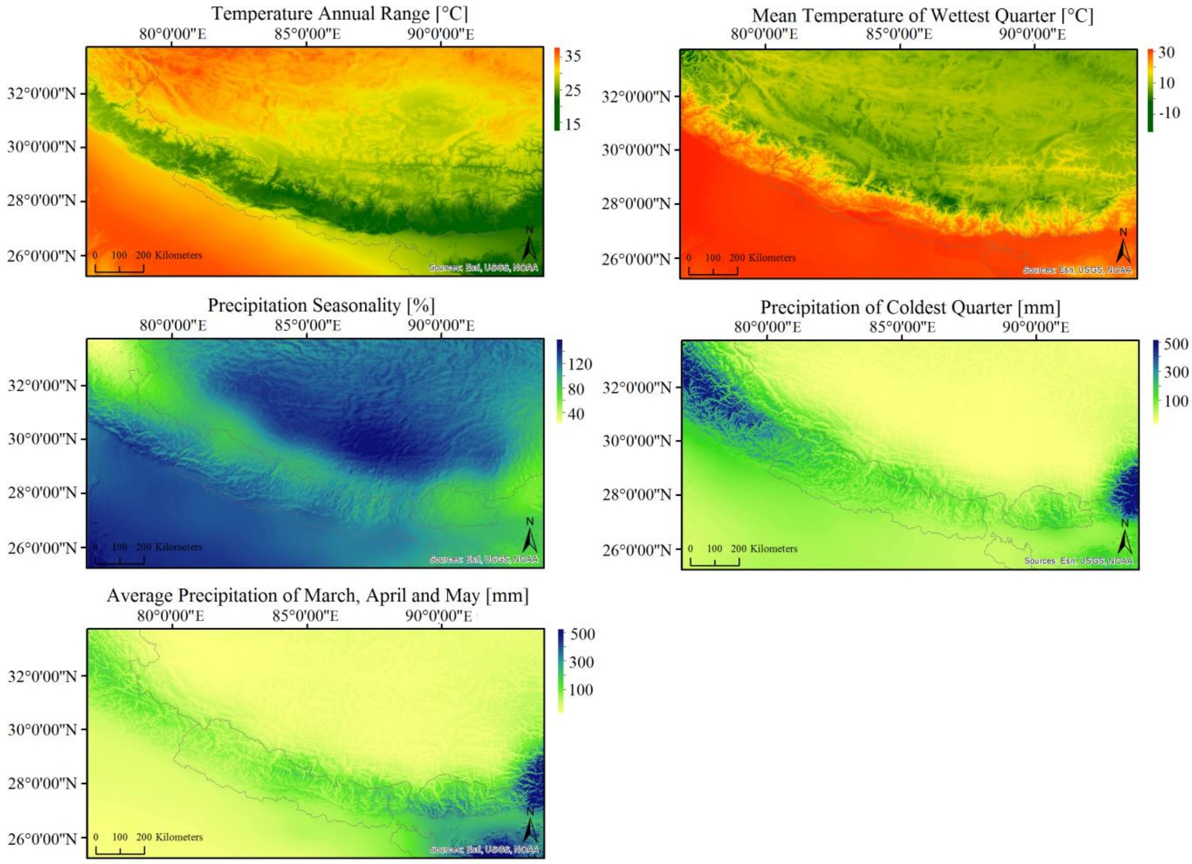


Fig. 16: Climatic variables used for modelling the potential distribution of *Betula utilis*.

The results of the final, averaged model revealed that variable importance varied between the predictor variables. Response curves of selected predictor variables were used to quantify the climatic niche of *B. utilis* (Fig. 17). Highest relative variable importance was found for Mean Temperature of Wettest Quarter, followed by Precipitation of Coldest Quarter. High relative variable importance was also ensued for Temperature Annual Range, followed by the Average Precipitation March-May, whereas lowest value was found for Precipitation Seasonality. Mean Temperature of Wettest Quarter showed a close-to-normal with a bell-shaped distribution. Precipitation Seasonality and Precipitation of Coldest Quarter show unimodal response curves, whereas Temperature Annual Range and Average Precipitation of March, April and May show sigmoidal response curves.

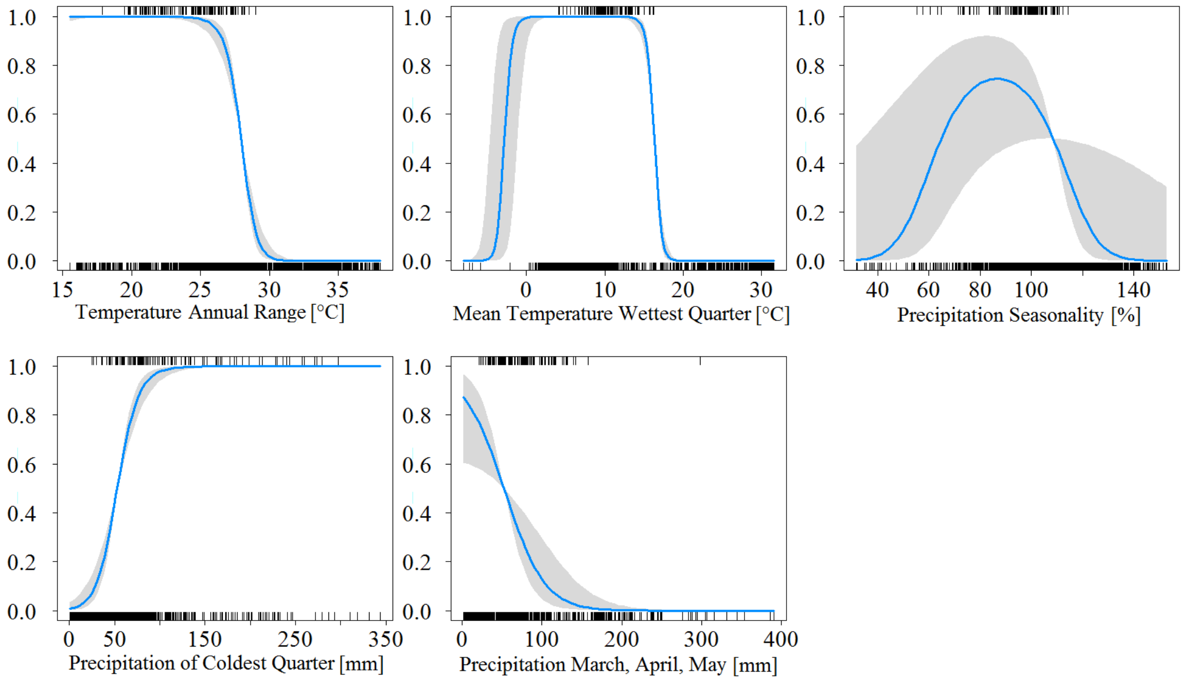


Fig. 17: Averaged response curves of the variables used in the generalized linear models to model the potential distribution of *Betula utilis*. Grey-shaded areas represent the confidence interval of the predicted probabilities. Response curves do not account for interactions between the variables.

Model calibration and evaluation

The results of all model performance measures are given in Table 4. The calculation of explained variance revealed that our model explained 84 % of the variance in the training dataset. Training and Testing models showed a clear correlation of the predicted and observed values ranging from r_p 0.85 to 0.88 respectively. Explained deviance was moderately high for the subsets. Calculation of slope and intercept of the calibration curve revealed an optimal fit of the models.

Table 4: Evaluation results for generalized linear models (averaged from 5 runs) on the selected predictor variables. Following evaluation metrics were calculated: The Explained Variance on the training dataset [%]), Pearson's correlation coefficient (r_p) between observed and predicted values, the Area under the Curve (AUC), True Skill Statistics (TSS), Explained Deviance, the Slope and Intercept of the calibration curve. The results for training and test data are displayed respectively (training 80% and testing 20% of the data).

	Explained variance	**Correlation r_p**		**AUC**		**TSS**		**Explained Deviance**		**Slope**		**Intercept**	
Model	Train	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
GLM	0.84	0.88	0.85	0.99	0.98	0.92	0.89	0.76	0.73	0.97	0.83	-0.01	-0.08

The threshold-independent "Area Under the Curve" (AUC) of the receiver-operating characteristic (ROC) plot approach was very high (AUC = 0.99 AUC_{SD} = 0.00163) for the training dataset as well as for the testing dataset (AUC = 0.98, AUC_{SD} = 0.0041). Continuous predictions of the current distribution of *B. utilis* are displayed in Fig. 18.

In order to assess model quality, the continuous maps were converted to binary distribution maps, using the threshold-dependent measure TSS. In order to obtain binary distribution maps, continuous maps were thresholded at the averaged optimal threshold (0.088). Confusion matrices of the final model (Supplementary material S4, Table S4) were then used to calculate the True Skill Statistics. The calculated TSS showed high value (TSS ≥ 0.89). The binary distribution map is shown in Fig. 19.

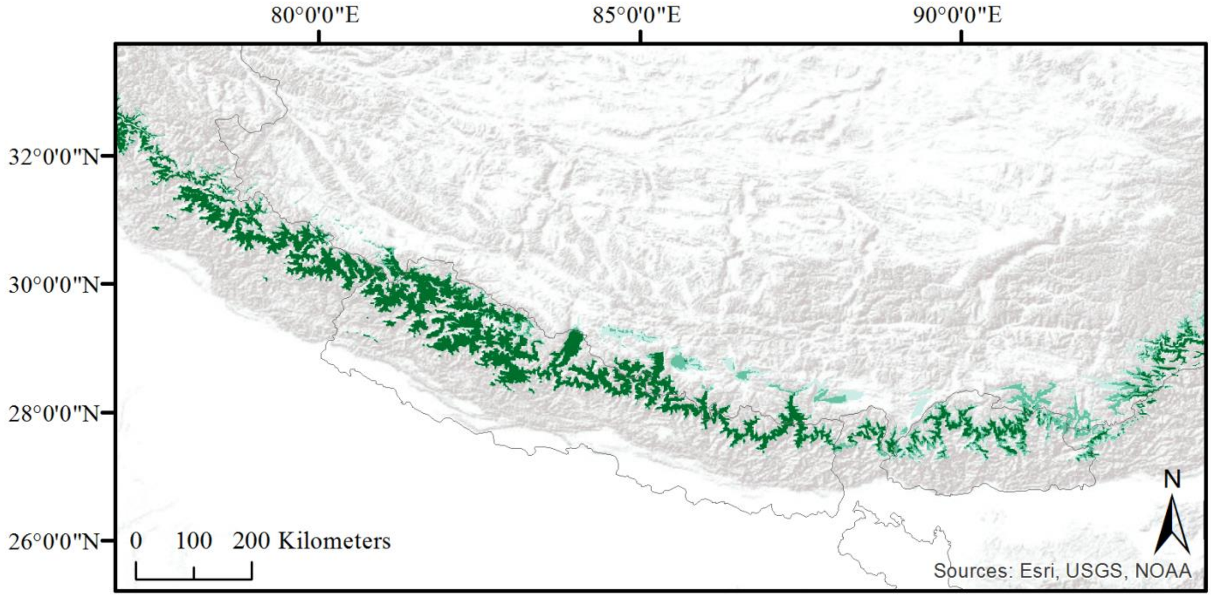


Fig. 18: Continuous predictions of the current distribution of *Betula utilis* predicted by all 5 model runs. For larger maps see Supplementary material S5, Fig. S5.

Model comparison

After converting the predictions from continuous to binary, the suitable habitat area was predicted to be much larger in the western and central region of the Himalayan mountain system compared to the eastern region.

In a final step we compared the Schweinfurth map (Schweinfurth, 1957) with our current prediction to detect differences between the real distribution of *B. utilis* in the 1950s and the current potential distribution (Fig. 19). The Schweinfurth map shows patches (red) where *B. utilis* is the dominant treeline species in the uppermost subalpine forests. In the Indian western Himalaya (Himachal Pradesh, Uttarakhand), the uppermost limit of the distribution range of *B. utilis* (red) predicted by the models roughly coincides with the vegetation map. In western and central Nepal and in Bhutan, however, the predicted high altitude occurrences are considerably higher than the distribution depicted in the Schweinfurth map. In the eastern Himalayan mountains (eastern Nepal and south-east Tibet), the Schweinfurth map shows patches of *B. utilis* as a prominent treeline species which are located in some instances above the predicted distribution range of the model.

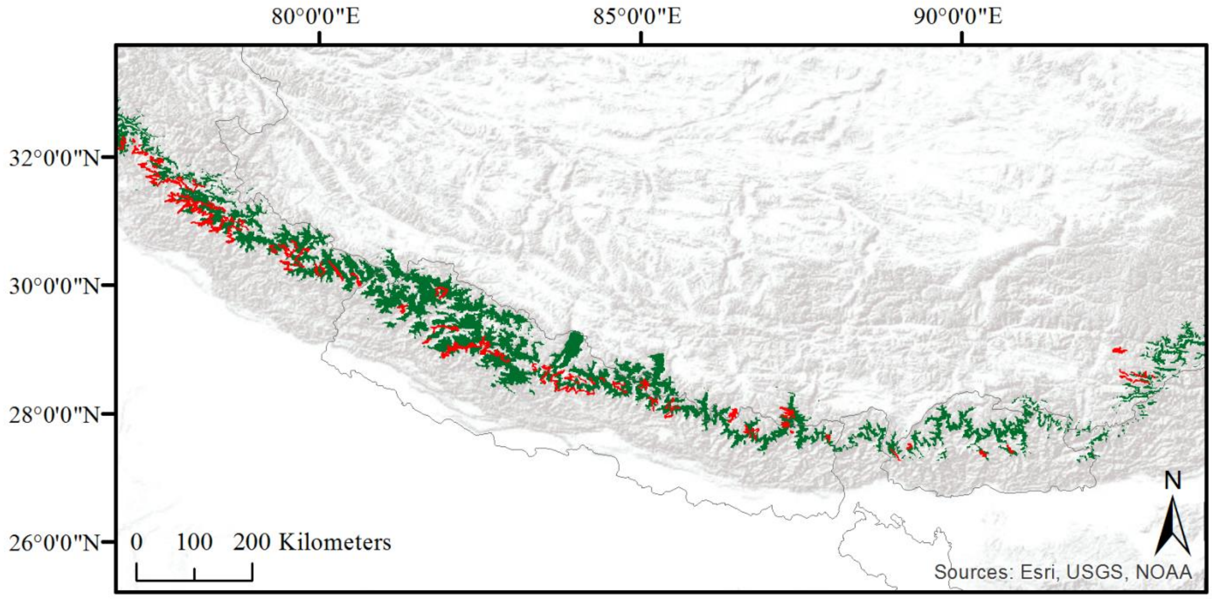


Fig. 19: Geographic extent of regions predicted as the potential distribution of *Betula utilis* (green) thresholded at True Skill Statistics (TSS ≥ 0.89) and distribution of *B. utilis* according to the Schweinfurth vegetation map (red) (Schweinfurth, 1957). For larger maps see Supplementary material S6, Fig. S6.

Discussion

Climatic space of B. utilis in the Himalayan region

We found three precipitation- and two temperature-related variables most important for predicting the potential distribution of *B. utilis* (Fig. 17). Mean Temperature of the Wettest Quarter and Temperature Annual Range were most significant among the temperature-related variables. Much lower growing season temperatures (i.e., Mean Temperature of the Wettest Quarter) in the E Himalayan mountains contribute to the lower seasonal temperature variation which favours evergreen *Rhododendron* treeline species and obviously constrains the capabilities of *B. utilis* to preoccupy its potential distribution range (Fig. 16). Notwithstanding the higher competitiveness of evergreen treeline tree species in the more oceanic E Himalaya, both evergreen and deciduous tree species are affected by direct growth limitation (sink limitation) due to low growing season temperatures at treeline elevations (Shi *et al.*, 2008).

Furthermore, the amplitude of the highest average summer temperature (June) and the lowest average winter temperature (January) for species occurrence locations were of great importance. This reflects seasonal temperature variation having a significant effect on the occurrence of *B. utilis* (Fig. 16 and 17). The seasonal temperature amplitude decreases towards the more oceanic eastern Himalaya where we detected striking differences between the potential and real distribution of *B. utilis* (Fig. 18). In the E Himalaya, the birch becomes less competitive and is ousted by coniferous and in particular by evergreen broadleaved species (*Rhododendron* spp.) as principal treeline species (Schickhoff, 2005). Higher degree of continentality in the W and NW of the Himalayan mountain system with higher mean temperatures of warmest months and severe winter coldness at treeline elevations lowers the competitive capacity of *Rhododendron* spp. and favours the competitiveness of *B. utilis* and the realisation of its climatic niche (Fig. 16). Although mean summer temperatures are distinctly higher in the more continental W and NW Himalaya, the treeline birch belt is located at much lower elevations (3600-3800 m) compared to treelines in the E Himalaya with uppermost birch occurrences around 4500 m (Schickhoff, 2005; Schickhoff *et al.*, 2015). This has to be attributed to very low winter temperatures (-13 °C mean temperature of coldest month at 3900 m in the Karakorum compared to -4 to -5 °C at 3900 m in SE Tibet and Yunnan; Schickhoff, 2005), long duration of massive snow cover, and shorter growing seasons which more than compensate for the effects of higher summer mean temperatures.

While growing season air and soil temperatures are considered key factors controlling tree growth at treelines and elevational position of treelines at the global scale (Holtmeier, 2009; Körner, 2012), precipitation and related factors such as soil moisture and soil nutrient availability can be significant at regional and local scales (e.g., Müller *et al.*, 2016 a, b). Thus, precipitation-related variables potentially limit the climatic space of treeline tree species.

We found significant influence of three precipitation-related variables (Precipitation of the Coldest Quarter, Average Precipitation of March, April and May, and Precipitation Seasonality) for estimating the climatic space of *B. utilis*. The results of our study highlight the importance of the variable Precipitation of the Coldest Quarter for the distribution of *B. utilis*. Averaged over all occurrence locations, the period from November to January was identified as the coldest quarter. Winter precipitation in form of snow represents the annual second precipitation maximum in high elevation areas of the Himalayan mountain system (Böhner *et al.*, 2015). Higher winter snowfall in the more continental western parts of the Himalayan region obviously contributes to increased birch occurrences compared to the eastern Himalaya (Fig. 16). The importance of snowpack as site factor for treeline birch forests was already stressed by Schweinfurth (1957), Champion & Seth (1968), Stainton (1972), Herzhoff & Schnitzler (1981), Puri *et al.* (1989), Dickoré & Nüsser (2000), and Eberhardt (2004). Troll (1939, 1967) highlighted the physiognomic adaptation of *B. utilis* to thick snowpack (snow-deformed trunks). Based on vegetation analyses in the W Himalaya and Karakorum, Schickhoff (1993, 2002) concluded that the distribution pattern of *B. utilis* forests is primarily controlled by thickness and duration of snow cover providing sufficient soil moisture at the beginning and at the end of the growing season, and that monsoonal summer rains, on the other hand, are of much less significance. Miehe *et al.* (2015b) also emphasize the association with a long-lasting snow cover, but indicate very different proportions of winter snowfall and monsoonal rainfall *B. utilis* forests receive across their entire distribution range. Nevertheless, a higher preoccupation of the potential distribution range of *B. utilis* is obviously linked to a higher proportion of winter snowfall. Our findings suggest Average Precipitation March, April and May to be a useful predictor variable to model the potential distribution of *B. utilis*. This accentuates the results of dendroecological studies (Dawadi *et al.*, 2013; Liang *et al.*, 2014; Gaire *et al.*, 2014; Schickhoff *et al.*, 2015), which emphasized low moisture availability in the pre-monsoon season as a significant site factor restricting the growth performance of *B. utilis*. The coincidence of low precipitation, higher temperatures, and higher evaporation results in potential pre-monsoon drought stress for *B. utilis*, that will amplify in coming decades due to high temperature trends in the pre-monsoon season for most of the Himalayan regions (Gerlitz *et al.*, 2014). Although March-April-May precipitation is low throughout the distribution range of *B. utilis*, a certain level of pre-monsoon precipitation is obviously needed to ensure sufficient soil moisture availability for the more hygrophilous birch, in particular in regions where winter snowfall is lower and less melt water is available.

The significance of higher humidity levels and sufficient soil moisture availability for the climatic space of *B. utilis* is evident from its small-scale distribution patterns. *Betula* forests thrive on humid, shady slopes with deeply weathered podzolic soils, and are more or less absent from south-facing slopes, in particular in the more continental W Himalaya (Schickhoff, 1993, 2002; Miehe *et al.*, 2015b). Only occasionally birch individuals or patches of *Betula* trees are found in sunny exposures, most notably in the more humid E Himalaya, and when azonal site conditions (water surplus habitats) can overcompensate the effects of high irradiation and soil moisture deficits. In semi-arid to semi-humid Himalayan regions with long-lasting snow cover, north- and south-facing slopes show sharp contrasts with regard to humidity and soil moisture conditions (Schickhoff, 2005). Considerable differences in irradiation intensity may result in southern aspects below 3200 m becoming free of snow already in January, while snow cover lasts until May at same elevations on north-facing slopes (Cramer 1997). We consider sufficient soil moisture availability to be crucial for defining the topo-climatically suitable habitat of *B. utilis*. The role of soil moisture for distribution and growth performance is often underestimated, but was shown to be significant for other tree genera and species near the warmer edge of their suitable range (Bonan & Sirois, 1992 for *Picea mariana*; Fang & Lechowicz, 2006 for *Fagus*; Ranjitkar *et al.*, 2014 for *Rhododendron*).

The last precipitation-related variable we found having an impact on the potential distribution of *B. utilis* is Precipitation Seasonality, which refers to the variation in monthly precipitation sums over the course of the year, expressed as percentage. Species distribution can be strongly influenced by the variability of precipitation. The median value of 97 indicates a high degree of precipitation variability due to the seasonal contrasts of a monsoon-influenced climate (Fig. 15; Supplementary material S3, Table S3a and S3b).

Summing up, the climatic space for *B. utilis* was well identified by our environmental niche model, and the projection to geographic space coincides closely with the current distribution of the species throughout the Himalayan mountains. Since this distribution largely represents climatic conditions in treeline ecotones, the results of our study provide insights into underlying climatic factors which might to a certain extent also delimit suitable habitats for other Himalayan treeline-forming tree species.

Performance of the modelling procedure

The predictive uncertainty associated with one evaluation metric can be reduced by comparing a variety of metrices. In order to define the environmental niche of *B. utilis* we modelled the climatic space of the species and projected the results in geographic space to predict the potential species distribution. We assessed model accuracy and prediction success by threshold-independent (i.e., AUC) and threshold-dependent (i.e., TSS) model performance measures. We found very high AUC values (AUC ≥ 0.98) for continuous projections (Fig. 18). Recent research suggested AUC being a questionable measure of model performance (Lobo *et al.*, 2008; Warren & Seifert, 2011), as it does not reflect absolute, but relative model performance. Thus, it is only valid and useful for comparison among models for a single study species in a single study region (Peterson *et al.*, 2011). Subsequently, the maps showing the predicted probability of species presence for each raster cell were converted into binary ‘presence - absence’-maps, using the threshold-dependent measure of the TSS. The TSS value for the final model was high (TSS = 0.89), assuming a high predictive performance.

Geographic projection of B. utilis

We conducted visual inspections of the predictions of the model to observe whether the model successfully predicted the species’ realized distribution throughout the entire Himalayan mountain system. According to the authors’ local field knowledge, the model predictions match the actual existing distribution range of *B. utilis* to the greatest possible extent. To interpret these findings, we investigated the binary species distribution maps (Fig. 19). *B. utilis* covers a notable environmental niche width across the entire Himalayan arc. The projections of the modelling results show that the most favourable conditions are located in the upper subalpine belt in the western and central Himalayan region where climatic conditions show a comparatively higher degree of continentality. This is consistent with the distribution pattern documented in several vegetation maps showing a narrow band of birch forests forming the upper treeline in northern exposures (Troll, 1939; Schweinfurth, 1957; Miehe, 1991; Schickhoff, 1994; Braun, 1996; Nüsser & Dickoré, 2002; Eberhardt *et al.*, 2007). Towards the eastern Himalaya, where more maritime climatic conditions favour the competitiveness of evergreen *Rhododendron* spp., *B. utilis* becomes a less frequent companion in subalpine forests and at treelines (Schickhoff, 2005).

The comparison of our predictions with the Schweinfurth map showed remarkable deviations in the distribution range of *B. utilis*, especially in the central part of the Himalayan mountains (Fig. 19), where the *B. utilis* is now predicted to occur at higher elevations compared to the real occurrence in the first half of the twentieth century. As the Schweinfurth map is based on an extensive literature survey, it cannot completely ruled out that there is some noise in the underlying data due to inaccurate altimetry (in particular in the E Himalaya with high elevation *Betula* occurrences) or sampling location errors. However, a much more likely explanation is that this finding has to be attributed to land use effects. Animal husbandry, timber logging, fuelwood collection and the like have been integral parts of village economies for centuries, if not millennia. Thus, treeline ecotones have been transformed to a large extent, and the far majority of Himalayan treelines is considered to be anthropogenic (Schickhoff *et al.*, 2015). Treeline depressions on north-facing slopes can be in the order of 300 m, with overgrazing and fire having been the main agents for lowering treelines (Schickhoff, 1995; Beug & Miehe, 1999). The difference between the modelled distribution range and the Schweinfurth map occurrence of *B. utilis* reflects large-scale, long-term anthropogenic interferences in Himalayan treeline landscapes. Effects of climate warming could also play a role since the distribution range of *B. utilis* is projected to shift upslope in coming decades (Schickhoff *et al.*, 2015). Upslope movement of anthropogenic treelines is, however, primarily a response to the cessation of land use. The few remaining near-natural krummholz treelines in the Himalaya are still relatively unresponsive to climate warming, suggesting treeline advance to higher elevation to be a medium- to long-term process (Schickhoff *et al.*, 2016; Schwab *et al.*, 2016).

Problems/ Limitations of our approach

Currently, modelling studies are often based on presence-only species occurrence data instead of presence-absence data, which are always a source of uncertainty. These presence-only data are often derived from databases of natural history museums and herbaria, which contain occurrences from numerous researchers, sampled with different techniques, intensities and periods of time (Soberón & Peterson, 2004). Sometimes, areas have been unequally sampled or occurrence data have been inaccurately georeferenced, resulting in occurrences of species with sampling bias. Such geographic sampling bias can lead to sampling bias in environmental space, which represents a major problem for modelling (Veloz, 2009; for the effects of sampling bias on model evaluation: Anderson & Gonzalez, 2011). We tried to overcome this problem by spatially filtering occurrence points (i.e., only one point per 1 x 1km grid cell). But not only presences are a source of uncertainties, also the absence points of species have to be carefully inspected. Although presence-absence data from planned surveys are highly requested, high quality presence-absence data are available only for very few species worldwide. Models using presence-absence data have proven to be of great value in predicting species distributions (Guisan *et al.*, 2002; Thuiller *et al.*, 2008). If true absences are not available, pseudo-absences must be generated. The selection of pseudo-absences involves a certain degree of uncertainty, which may lead to over- or underestimated predictions. Based on our results, we do stress the need for incorporating true absences in ENMs. Furthermore, biased climate data can lead to distorted models (Heikkinen *et al.*, 2006). These geographic and environmental biases are contrary to the assumption of many modelling techniques, viz. that the localities represent a random sample from the area being modelled (Phillips *et al.*, 2006).

Finally, it should be noted that the actual distribution range might presumably be smaller than indicated by our models, because climatic factors are not the only factors determining habitat suitability. Although not considered in this study, interactions of a whole array of site factors such as topo-climate (radiation, temperature, precipitation, length of growing season, and snow cover), topography (slope inclination, relief forms), ecology of tree species (regeneration, seed dispersal, succession stages), site history (climate oscillations, fire, human impacts, insect attacks), current biotic (browsing, trampling, diseases and insect pests) and anthropogenic influences (burning, logging,grazing, recreation and tourism) affect treeline species spatial distributions (Schickhoff, 2005; Holtmeier, 2009). In mountainous areas the resolution of climate data (i.e., 1 x 1km) is often too coarse for models to distinguish between north- and south-facing slopes. We hereby emphasise the need for fine-scale climate data in order to account for the heterogeneous terrain, leading to more precise modelling results.

Conclusions

The distributional range of *B. utilis* was found to be primarily associated with thermal (Mean Temperature of the Wettest Quarter and Temperature Annual Range) and precipitation-related factors (Precipitation of the Coldest Quarter, Average Precipitation of March, April and May, and Precipitation Seasonality).

The application of our environmental niche modelling approach with appropriate model evaluation measures for assessment of model performance provided the current distribution range of *B. utilis*. The robustness of the results increased significantly using multi-faceted approach of model performance evaluation.

The results of our study, which account for underlying climatic factors in mountain ecosystems, may serve as a baseline for the exploration of potential impacts of climate change on future treeline species distribution in regions with limited data availability. The results will enhance the quantification of altitudinal shifts and other analyses of spatial pattern dynamics in mountain ecotones, which will be of increasing significance with regard to future climate change. We expect to provide a new starting point for modelling future treeline dynamics and treeline shifts in the Himalaya under novel climate conditions.

Our results point to the need for further investigations of relevant abiotic and biotic factors for species' occurrences and range shifts such as soil parameters and anthropogenic impacts at different temporal and spatial scales. In conclusion, a better understanding of potential ranges of alpine treeline species is supported by our model results which may also provide insights for decision makers in climate change adaptation and biodiversity conservation.

Acknowledgements

We would like to thank Kim Stolle (University of Hamburg) for georeferencing and digitalizing the Schweinfurth vegetation map. We also express our gratitude to Himalayan colleagues, guides and local people who accompanied us on numerous field trips to *Betula* treelines. Additionally we would like to thank two anonymous reviewers for their comments and suggestions. The study was conducted under the framework TREELINE project and partially supported by a specific grant from the German Research Foundation (SCHI 436/14-1).

References

*Akaike, H. (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control, **19**, 716-723.*

Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223-1232.

Anderson, R.P. & Gonzalez Jr, I. (2011) Species-specific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. *Ecological Modelling*, **222**, 2796-2811.

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.

Araújo, M., Pearson, R.G., Thuiller, W. & Erhard, M. (2005) Validation of species - climate impact models under climate change. *Global Change Biology*, **11**, 1504-1513.

Araújo, M. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677-1688.

Ashburner, K. & McAllister, H.A. (2013) *The genus Betula - a taxonomic revision of birches*. Botanical Magazine Monograph 5. Royal Botanic Gardens, Kew.

Austin M.P. (1999) A silent clash of paradigms: some inconsistencies in community ecology. *Oikos*, **86**, 170-178.

Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101-118.

Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D. & Luoto, M. (2006) Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. *Ecological Modelling*, **199**, 197-216.

Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, **3**, 327-338.

Barry, S. & Elith, J. (2006) Error and uncertainty in habitat models. *Journal of Applied Ecology*, **43**, 413-423.

Beug, H.J. & Miehe, G. (1999) *Vegetation history and human impact in the eastern central Himalaya (Langtang and Helambu, Nepal)*. Dissertationes Botanicae 318, Borntraeger, Berlin-Stuttgart.

Bhattacharyya, A., Shah, S.K. & Chaudhary, V. (2006) Would tree-ring data of *Betula utilis* have potential for the analysis of Himalayan glacial fluctuations? *Current Science*, **91**, 754–761.

Böhner, J., Miehe, G., Miehe S. & Nagy, L. (2015) Climate and Weather. *Nepal: an introduction to the natural history, ecology and human environment of the Himalayas* (ed. by G. Miehe, C.A. Pendry and R. Chaudhary). pp. 23-89. Royal Botanic Garden, Edinburgh.

Bonan, G.B. & Sirois, L. (1992) Air temperature, tree growth, and the northern and southern range limits to *Picea mariana*. *Journal of Vegetation Science*, **3**, 495-506.

Braun, G. (1996) *Vegetationsgeographische Untersuchungen im NW-Karakorum (Pakistan). Kartierung der aktuellen Vegetation und Rekonstruktion der* *aktuellen Waldverbreitung auf der Basis von Satellitendaten, Gelände- und Einstrahlungsmodellen*. Bonner Geographische Abhandlungen 93, Dümmlers, Bonn.

Broennimann, O., Petipierre, B., Randon, C., Engler, R., Di Cola, V., Breier, F., D'Amen, M., Pellissier, L., Pottier, J., Pio, D., Gracia Mateo, R., Hordijk, W., Dubuis, A., Scherrer, D., Salamin, N. & Guisan, A. (2015) ecospat: spatial ecology miscellaneous methods. R-package version 1.1, http://CRAN.R-project.org/package=ecospat.

Burnham, K. P. & Anderson, D. R. (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd ed., Springer, New York.

Champion, H.G. & Seth, S.K. (1968) *A revised survey of the forest types of India*. Government of India, New Delhi.

Cramer, T. (1997) Climatic gradients in the Karakorum and their effects on the natural vegetation. *Perspectives on History and Change in the Karakorum, Hindukush and Himalaya* (ed. by I. Stellrecht and M. Winiger), pp. 265-276. Culture Area Karakorum Scientific Studies 3, Köppe, Köln.

Dawadi, B., Liang, E., Tian, L., Devkota, L.P. & Yao, T. (2013) Pre-monsoon precipitation signal in tree rings of timberline *Betula utilis* in the central Himalayas. *Quaternary International*, **283**, 72-77.

Dickoré, W.B. & Nüsser, M. (2000) *Flora of Nanga Parbat (NW Himalaya, Pakistan). An annotated inventory of vascular plants with remarks on vegetation dynamics*. Englera 19, Veröffentlichungen aus dem Botanischen Garten und Botanischen Museum Berlin-Dahlem.

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. & Lautenbach, S. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 27-46.

Dufour-Tremblay, G., De Vriendt, L., Lévesque, E. & Boudreau, S. (2012) The importance of ecological constraints on the control of multi-species treeline synamics in eastern Nunavik, Québec. *American Journal of Botany*, **99**, 1638-1646.

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.

Dutta, P.K., Dutta, B.K., Das, A.K. & Sundriyal, R.C. (2014) Alpine treeline research gap in Himalaya: a literature review. *Indian Forester*, **4**, 419-427.

Eberhardt, E. (2004) *Plant life of the Karakorum. The vegetation of the upper Hunza catchment (Northern Areas, Pakistan). Diversity, syntaxonomy, distribution*. Dissertationes Botanicae 387, Borntraeger, Stuttgart.

Eberhardt, E., Dickoré, W.B. & Miehe, G. (2007) Vegetation map of the Batura Valley (Hunza Karakorum, North Pakistan). *Erdkunde*, **61**, 93-112.

Elith, J. & Burgmann, M. (2002) Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. *Predicting species occurrences: issues of accuracy and scale* (ed. by J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall & F.B. Samson), pp. 303–313. Island Press, Washington, DC.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann N.E. (2006) Novel methods improve prediction of species distributions from occurrence data. *Ecography*, **29**, 129-151.

ESRI, 2012. ArcGIS Desktop: Release 10.1. Environmental Systems Research Institute, Redlands, CA.

Fang, J. & Lechowicz, M.J. (2006) Climatic limits for the present distribution of beech (*Fagus* L.) species in the world. *Journal of Biogeography*, **33**, 1804-1819.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38-49.

Flueck, J.A. (1987) A study of some measures of forecast verification. *Preprints: Tenth conference on probability and statistics in the atmospheric sciences* (Edmonton, Alta., Canada), pp. 69-73. American Meteorological Society, Boston.

Franklin, J. (1995) Predictive vegetation mapping - geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, **19**, 474-499.

Freeman, E.A. & Moisen, G.G. (2008a) A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling*, **217**, 48-58.

Freeman, E. A. & Moisen, G.G. (2008b) PresenceAbsence: An R package for presence-absences model analysis. *Journal of Statistical Software*, **23**, 1-31.

Gaire, N.P., Bhuju, D.R. & Koirala, M. (2013) Dendrochronological studies in Nepal: Current status and future prospects. FUUAST *Journal of Biology*, **3**, 1–9.

Gaire, N.P., Koirala, M., Bhuju, D.R., Borgaonkar, H.P. (2014) Treeline dynamics with climate change at Central Nepal Himalaya. *Climate of the Past*, **9**, 5941-5976.

gbif.org: Biodiversity occurrence data provided by: Missouri Botanical Garden*, Royal Botanic Garden Edinburgh and T*he Himalayan Uplands Plant database*, Accessed through GBIF Data Portal, http://*www.gbif.org*, Last accessed 28 April 2015.*

Gerlitz, L., Conrad, O. & Böhner, J. (2014) Warming patterns over the Tibetan Plateau and adjacent lowlands derived from elevation- and bias-corrected ERA-Interim data. *Climate Research*, **58**, 235-246.

Gerlitz, L. Conrad, O., Thomas, A & Böhner, J. (2015) Large-scale atmospheric forcing and topographic modification of precipitation rates over High Asia - a neural-network-based approach. *Earth System Dynamics*, **6**, 61-81.

Gottfried, M., Pauli, H., Futschik, A., *et al.* (2012) Continent-wide response of mountain vegetation to climate change. *Nature Climate Change*, 2, 111-115.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.

Guisan, A., Edwards Jr., T.C. & Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**, 89-100

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.

Grosjean, P. & Ibanez, F. (2014) pastecs: Package for analysis of space-time ecological series. R-package version 1.3.18, http://cran.r-project.org/web/packages/pastecs/index.html.

Harsch, M.A., Hulme, P.E., McGlone, M.S. & Duncan, R.P. (2009) Are treelines advancing? A global meta-analysis of treeline response to climate warming. *Ecology Letters*, **12**, 1040-1049.

Heikkinen, R.K., Luoto, M., Araujo, M.B., Virkkala, R., Thuiller, W. & Sykes, M.T. (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, **30**, 751-777.

Herzhoff, B. & Schnitzler, H. (1981) Die subalpinen Birken-Rhododendron-Wälder im Dagwan-Tal/West-Himalaya. *Mitteilungen der Deutschen Dendrologischen Gesellschaft*, **72**, 171-186.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Hijmans, R.J., Phillips, S.J., Leathwick, J. & Elith, J. (2011) dismo: species distribution modeling. R-package version 1.0-12, http://cran.r-project.org/web/packages/dismo/index.html.

Hijmans, R.J. (2015) raster: geographic data analysis and modeling. R-package version 2.5-2, http://cran.r-project.org/package=raster.

Holtmeier, F.-K. (2009) *Mountain timberlines - Ecology, patchiness and dynamics*. Advances in Global Change Research 36, Springer, Berlin.

Holtmeier, F.-K. & Broll, G. (2010) Altitudinal and polar treelines in the northern hemisphere - causes and response to climate change. *Polarforschung*, **79**, 139-153.

Huo, C., Cheng, G., Lu, X. & Fan, J. (2010) Simulating the effects of climate change on forest dynamics on Gongga Mountain, Southwest China. *Journal of Forest Research*, **15**, 176-185.

IPCC (2014) *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by Core Writing Team, R.K. Pachauri & L.A. Meyer). IPCC, Geneva, Switzerland, 151 pp.

Irl, S.D.H., Anthelme, F., Harter, D.E.V., Jentsch, A., Lotter, E., Steinbauer M.J. & Beierkuhnlein, C. (2015) Patterns of island treeline elevation – a global perspective. *Ecography*, **38**, 1-10.

Jimenez-Valverde, A.& Lobo, J.M. (2007) Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica*, **31**, 361-369.

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N., Linder, H.P. & Kessler, M. (2016) Climatologies at high resolution for the earth land surface areas. arXiv:1607.00217 [physics].

Körner, C. (2012) *Alpine treelines - functional ecology of the global high elevation tree limits*. Springer, Berlin Heidelberg.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang Y. & Candan, C. (2016) caret: Classification and Regression Training. R-package version 6.0-68. http://CRAN.R-project.org/package=caret.

Kullman, L. (1998) Tree-limits and montane forests in the Swedish Scandes: sensitive biomonitors of climate change and variability. *Ambio*, **27**, 312-321.

Kumar, P. (2012) Assessment of impact of climate change on rhododendrons in Sikkim Himalayas using Maxent modelling: limitations and challenges. *Biodiversity and Conservation*, **21**, 1251-1266.

Liang, E., Dawadi, B., Pederson, N. & Eckstein, D. (2014) Is the growth of birch at the upper timberline in the Himalayas limited by moisture or by temperature? *Ecology*, **95**, 2453-2465.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145-151.

McCullagh, P. & Nelder, J. A. (1989) *Generalized linear models*. Chapman and Hall, London.

Miehe, G. (1991) Die Vegetationskarte des Khumbu Himal (Mt. Everest-Südabdachung) 1:50.000. Gefügemuster der Vegetation und Probleme der Kartierung. *Erdkunde*, **45**, 81-94.

Miehe, G., Pendry, C.A. & Chaudhary, R. (2015a) *Nepal: an introduction to the natural history, ecology and human environment of the Himalayas*. Royal Botanic Garden, Edinburgh.

Miehe, G., Miehe, S., Böhner, J., Bäumler, R., Ghimire, S.K., Bhattarai, K., Chaudhary, R.P., Subedi, M., Jha, P.K. & Pendry, C. (2015b) Vegetation ecology. *Nepal: an introduction to the natural history, ecology and human environment of the Himalayas* (ed. by G. Miehe, C.A. Pendry and R. Chaudhary), pp. 385-472. Royal Botanic Garden, Edinburgh.

Müller, M., Schickhoff, U., Scholten, T., Drollinger, S., Böhner, J. & Chaudhary, R.P. (2016a) How do soil properties affect alpine treelines? General principles in a global perspective and novel findings from Rolwaling Himal, Nepal. *Progress in Physical Geography* **40**, 1-26.

Müller, M., Schwab, N., Schickhoff, U., Böhner, J. & Scholten, T. (2016b) Soil temperature and soil moisture patterns in a Himalayan alpine treeline ecotone. *Arctic, Antarctic, and Alpine Research* **48,** 501-521.

Naimi, B. (2015) usdm: uncertainty analysis for species distribution models. R-package version 1.1-15. http://CRAN.R-project.org/package=usdm.

Nelder, J.A. & Wedderburn, R.W.M. (1972) Generalized linear models. *Journal of Royal Statistical Society*, Series A, **135**, 370-384.

Nix, H.A. (1986) A biogeographic analysis of Australian Elapid snakes. *Atlas of Australian Elapid Snakes, Australian Flora and Fauna*, Series 8. (ed. by R. Longmore), pp. 4-15. Australian Government Publishing Service, Canberra.

Nüsser, M. & Dickoré, W.B. (2002) A tangle in the triangle: Vegetation map of the eastern Hindukush (Chitral, northern Pakistan). *Erdkunde*, **56**, 37-59.

Parolo, G., Rossi, G. & Ferrarini, A. (2008) Toward improved species niche modelling: *Arnica montana* in the Alps as a case study. *Journal of Applied Ecology*, **45**, 1410-1418.

Pauli, H., Gottfried, M., Dullinger, S., *et al.* (2012) Recent plant diversity changes on Europe´s mountain summits. *Science*, **336**, 353-355.

Paulsen, J. & Körner, C. (2014) A climate-based model to predict potential treeline position around the globe. *Alpine Botany*, **124**, 1-12.

Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225-245.

Peterson, A.T., Papes, M. & Eaton, M. (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, **30**, 550-560.

Peterson A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011*) Ecological niches and geographic distibutions*. Princeton University Press, Princeton, NJ.

Peterson, A.T. & Soberón, J. (2012) Species distribution modeling and ecological niche modeling: getting the concepts right. *Naturaleza and Conservacao*, **10**, 1-6.

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

Polunin, O. & Stainton, A. (1984) *Flowers of the Himalaya*. Oxford University Press, New Delhi.

Puri, G.S., Gupta, R.K., Meher-Homji, V.M. & Puri, S. (1989) *Forest ecology*. Vol. II: *Plant form, diversity, communities and succession*. Oxford & IBH Publ., New Delhi-Bombay-Calcutta.

R Core Team, 2015, version: 3.1.3 R Core Team (2013). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/.

Rai, I.D. Bharti, R.R Adhikari B.S. & Rawat, G.S. (2013) Structure and functioning of timberline vegetation in the Western Himalaya: a Case Study. *High-Altitude Rangelands and their interfaces in the Hindu Kush Himalayas* (ed. by W. Ning, G.S. Rawat, S. Joshi, M. Ismail and E. Sharma), pp. 91-106. ICIMOD, Kathmandu, Nepal.

Randin C.F., Engler, R., Normand, S., Zappa, M., Zimmermann, N.E., Pearman, P.B., Vittoz, P., Thuiller, W. & Guisan, A. (2009) Climate change and plant distribution: local models predict high-elevation persistence. *Global Change Biology*, **15**, 1557-1569.

Ranjitkar, S., Kindt, R., Sujakhu, N.M., Hart, R., Guo, W., Yang, X., Shrestha, K.K., Xu, J. & Luedeling, E. (2014) Separation of the bioclimatic spaces of Himalayan tree rhododendron species predicted by ensemble suitability models. *Global Ecology and Conservation*, **1**, 2-12.

Reineking, B. & Schröder, B. (2006) Constrain to perform: Regularization of habitat models. *Ecological Modelling.* **193**, 675-690.

Schibalski, A., Lehtonen, A. & Schröder, B. (2014) Climate change shifts environmental space and limits transferability of treeline models. *Ecography*, **37**, 321-335.

Schickhoff, U. (1993) *Das Kaghan-Tal im Westhimalaya (Pakistan). Studien zur landschaftsökologischen Differenzierung und zum Landschaftswandel mit vegetationskundlichem Ansatz*. Bonner Geographische Abhandlungen 87, Dümmlers, Bonn.

Schickhoff, U. (1994) Die Verbreitung der Vegetation im Kaghan-Tal (Westhimalaya, Pakistan) und ihre kartographische Darstellung im Maßstab 1:150.000. *Erdkunde*,**48**, 92-110.

Schickhoff, U. (1995) Himalayan forest-cover changes in historical perspective. A case study in the Kaghan Valley, Northern Pakistan. *Mountain Research and Development*, **15**, 3-18.

Schickhoff, U. (2002) *Die Degradierung der Gebirgswälder Nordpakistans. Faktoren, Prozesse und Wirkungszusammenhänge in einem regionalen Mensch-Umwelt-System*. Erdwissenschaftliche Forschung 41, Steiner, Stuttgart.

Schickhoff, U. (2005) The upper timberline in the Himalaya, Hindu Kush and Karakorum: a review of geographical and ecological aspects. *Mountain ecosystems. Studies in treeline ecology* (ed. by G. Broll and B. Keplin), pp. 275-354. Springer, Berlin - Heidelberg.

Schickhoff, U. (2011) Dynamics of mountain ecosystems. *Handbook of biogeography* (ed. by A. Millington, M. Blumler and U. Schickhoff), pp. 313-337. Sage Publications, London.

Schickhoff, U., Bobrowski, M., Jürgen Böhner, J., Bürzle, B., Chaudari, R.P., Gerlitz, L., Heyken, H., Lange, J., Müller, M., Scholten, T., Schwab, N. & Wedegärtner, R. (2015) Do Himalayan treelines respond to recent climate change? An evaluation of sensitivity indicators. *Earth System Dynamics*, **6**, 245-265.

Schickhoff, U., Bobrowski, M., Böhner, J., Bürzle, B., Chaudhary, R.P., Gerlitz, L., Lange, J., Müller, M., Scholten, T. & Schwab, N. (2016) Climate change and treeline dynamics in the Himalaya. *Climate Change, Glacier Response, and Vegetation Dynamics in the Himalaya* (ed. by R.B. Singh, U. Schickhoff and S. Mal), pp. 271-306. Springer, Switzerland.

Schwab, N., Schickhoff, U., Bobrowski, M., Böhner, J., Bürzle, B., Chaudhary, R.P., Gerlitz, L., Müller, M. & Scholten, T. (2016) Treeline responsiveness to climate warming: Insights from a krummholz treeline in Rolwaling Himal, Nepal. *Climate Change, Glacier Response, and Vegetation Dynamics in the Himalaya* (ed. by R.B. Singh, U. Schickhoff and S. Mal), pp. 307-345. Springer, Switzerland.

Schweinfurth, U. (1957) Die horizontale und vertikale Verbreitung der Vegetation im Himalaya. *Bonner Geographische Abhandlungen* 20, Dümmlers, Bonn.

Shi, P., Körner, C. & Hoch, G. (2008) A test of the growth-limitation theory for alpine tree line formation in evergreen and deciduous taxa of the eastern Himalayas. *Functional Ecology*, **22**, 213-220.

Shrestha, B.B., Ghimire, B., Lekhak, H.D. & Jha, P.K. (2007): Regeneration of treeline birch (*Betula utilis* D. Don) forest in a Trans-Himalayan dry valley in Central Nepal. *Mountain Research and Development*, **27**, 259-267.

Shrestha, K.B., Hofgaard, A. & Vandvik, V. (2014) Recent treeline dynamics are similar between dry and mesic areas of Nepal, Central Himalaya. *Journal of Plant Ecology*, **8**, 347-358.

Singh, C.P., Panigrahy, S., Thapliyal, A., Kimothi, M.M., Soni, P. & Parihar, J. S. (2012) Monitoring the alpine treeline shift in parts of the Indian Himalayas using remote sensing. *Current Science*, **12**, 559-562.

Singh, C.P., Panigrahy, S., Parihar, J.S. & Dharaiya, N. (2013) Modeling environmental niche of Himalayan birch and remote sensing based vicarious validation. *Tropical Ecology*, **54**, 321-329.

Soberon, J. & Peterson, A.T. (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 689-698.

Soria-Auza, R., Kessler, M., Bach, K., Barajas-Barbosa, P., Lehnert, M., Herzog, S. & 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**, 1221-1229.

Speed, J.D.M., Austrheim, G., Hester, A.J. & Mysterud, A. (2011) Growth limitation of mountain birch caused by sheep browsing at the altitudinal treeline. *Forest Ecology and Management*, **261**, 1344-1352.

Stainton, J.D.A. (1972) *Forests of Nepal*. Hafner, New York.

Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B: Methodological*, **36**, 111-147.

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

Telwala, Y., Brook, B.W., Manish, K. & Pandit M.K. (2013) Climate-induced elevational range shifts and increase in plant species richness in a Himalayan biodiversity epicentre. *PLoS ONE*, **2**, e57103.

Tenca, A. & Carrer, M. (2010) Growth climate response at high elevation: comparing Alps and Himalayas. *TRACE - Tree Rings in Archaeology, Climatology and Ecology*, Vol. 8 (ed. by T. Levanic, J. Gricar, P. Hafner, R. Krajnc, S. Jagodic, H. Gärtner, I. Heinrich and G. Helle), pp. 89-97. Scientific Technical Report 10/05, GFZ German Research Centre for Geosciences, Potsdam.

Thuiller, W. (2003) BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*, **9**, 1353-1362.

Thuiller, W., Lavorel, S. & Araujo, M.B. (2005) Niche properties and geographical extent as predictors of species sensitivity to climate change. *Global Ecology Biogeography*, **14**, 347-357.

Thuiller, W., Albert, C., Araujo, M.B., Berry, P.M., Cabeza, M., *et al.* (2008) Predicting global change impacts on plant species' distributions: future challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 137-52.

Troll, C. (1939) Das Pflanzenkleid des Nanga Parbat. Begleitworte zur Vegetationskarte der Nanga-Parbat-Gruppe (Nordwest-Himalaja) 1:50.000. *Wissenschaftliche Veröffentlichungen des Deutschen Museums für Länderkunde zu Leipzig* N.F. 7: 151-193.

Troll, C. (1967) Die klimatische und vegetationsgeographische Gliederung des Himalaya-Systems. *Khumbu Himal 1. Ergebnisse des Forschungsunternehmens Nepal Himalaya* (ed. by W. Hellmich), pp. 353-388. Springer, Berlin-Heidelberg-New York.

Troll, C. (1972) The three-dimensional zonation of the Himalayan system. Geoecology of the high-mountain regions of Eurasia. *Erdwissenschaftliche Forschung IV* (ed. by C. Troll), pp. 264-275. Franz Steiner Verlag, Wiesbaden.

Truong C., Palmé, A.E. & Felber, F. (2007) Recent invasion of the mountain birch *Betula pubescens ssp. tortuosa* above the treeline due to climate change: genetic and ecological study in Northern Sweden. *Journal of Evolutionary Biology*, **20**, 369-380.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. & Kadmon, R. (2007) A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, **13**, 397-405.

VanderWal, J., Shoo, L.P., Graham, C. & William S.E. (2009) Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecological Modelling*, **220**, 589-594.

Veloz, S.D. (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, **36**, 2290-2299.

Walsh, C. & Mac Nally, R. (2013) hier.part: Hierarchical Partitioning. R-package version 1.04, http://CRAN.R-project.org/package=hier.part.

Wang, W., Anderson, B.T., Phillips, N. & Kaufmann, R.K. (2006) Feedbacks of vegetation on summertime climate variability over the North American grasslands. Part I: Statistical Analysis. *Earth Interactions*, **10**, 1-27.

Warren, D.L. & Seifert, S.N. (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, **21**, 335-342.

Wei, T. & Simko, V. (2016) corrplot: Visualization of a correlation matrix. R-package version 0.77, http:// https://cran.r-project.org/web/packages/corrplot/index.html.

Wieser, G., Holtmeier, F.K. & Smith, W.K. (2014) Treelines in a changing global environment. *Trees in a Changing Environment* (ed. by M. Tausz & N. Grulke), pp. 221-263. Springer, Dordrecht.

Wiley, E.O., McNyset, K.M., Peterson, A.T., Robins, C.R. & Stewart, A.M. (2003) Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanography*, **16**, 120-127.

Xu, X., Piao, S., Wang, X., Chen, A., Ciais, P. & Myneni, R.B. (2012) Spatio-temporal patterns of the area experiencing negative vegetation growth anomalies in China over the last three decades. *Environmental Research Letters*, **7**, 1-9.

Zurick, D. & Pacheco, J. (2006) *Illustrated atlas of the Himalaya.* The University Press of Kentucky, Lexington.

Article II

<u>Bobrowski, M.</u>; 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. doi: 10.1016/j.ecolmodel.2017.05.021.

Abstract

Betula utilis is a major constituent of alpine treeline ecotones in the western and central Himalayan region. The objective of this study is to analyse for the first time the performance of different climatic predictors in modelling the potential distribution of *B. utilis* in the subalpine and alpine belts of the Himalayan region. Using Generalized Linear Models (GLM) we aim at examining climatic factors controlling the species distribution under current climate conditions. We evaluate the predictive ability of climate data derived from different statistical methods. GLMs were created using least correlated bioclimatic variables derived from two different climate data sets: 1) interpolated climate data (i.e., WORLDCLIM; Hijmans *et al.*, 2005), and 2) quasi-mechanistical statistical downscaling (i.e., CHELSA; Karger *et al.*, 2016). Model accuracy was evaluated using threshold-independent (Area Under the Curve) and threshold-dependent (True Skill Statistics) measures. Although there were no significant differences between the models in AUC, we found highly significant differences ($p \leq 0.01$) in TSS. We conclude that models based on variables of CHELSA climate data had higher predictive power, whereas models using WORLDCLIM climate data consistently overpredicted the potential suitable habitat for *B. utilis*.

Although climatic variables of WORLDCLIM are widely used in modelling species distribution, our results suggest to treat them with caution when topographically complex regions like the Himalaya are in focus. Unmindful usage of climatic variables for environmental niche models potentially causes misleading projections.

Keywords: CHELSA, *Betula utilis*, model evaluation, predictive modeling, WORLDCLIM

Introduction

The aim of modelling species potential distribution is to characterize suitable habitat conditions, based on climatological, environmental and biotic correlates (Soberón & Namakura, 2009). The general approach is to link species occurrences with climatic and topographic variables to estimate the species distribution range, since habitat suitability is considerably influenced by the prevailing climate (Pearson & Dawson, 2003). It is assumed that a species occurs within a climatic range determined by its climatic needs within a range of spatial scales (Trivedi *et al.*, 2008).

Within the scope of modelling species niches or distribution, modelling studies face numerous challenges. Not only the choice of modelling algorithm is subject to numerous sources of uncertainties (Elith *et al.*, 2006; Aráujo & New, 2007), but also the data used for modelling. Models using presence-absence data have proven to be of great value in predicting species distributions (Guisan *et al.*, 2002; Thuiller *et al.*, 2008), but this data is often not available. Other elements of uncertainties in modelling species distribution are attributed to sample design, sample size, species prevalence, sample resolution, study area extent and the like (for detailed discussion see Franklin, 2009). Further challenges arise from the spatial structure of species occurrence data that may be collinear with environmental data (Araújo & Guisan, 2006; Loiselle *et al.*, 2008, Naimi *et al.*, 2013). Sometimes, areas have been unequally sampled due to differential accessibility of a study area, resulting in occurrences of species with sampling bias. Sampling records often cluster near the centre of climatic conditions under which the species occurs (Loiselle *et al.*, 2008). This leads to species documentations that do not cover the whole range of suitable habitat conditions for respective species. Such geographic sampling bias can lead to sampling bias in environmental space, which represents a major problem for modelling (Veloz, 2009; for the effects of sampling bias on model evaluation see Anderson & Gonzalez, 2011). This holds particularly true for sampling treeline species in remote areas like the Himalayan region. Due to lower accessibility of treeline sites, the number of available sampling plots is sparse, which has a reciprocal effect on prediction performance (Aráujo *et al.*, 2005). Araújo & Guisan (2006) found that models tend to predict species occupying a narrow niche better than species with a wider niche.

The underlying concept of most modelling studies is the prediction of species distribution ranges using climatic variables. The choice of environmental variables used to model species distributions may result in different distribution maps for the same species (Luoto *et al.*, 2007). Whereas multi-collinearity and spatial autocorrelation of predictors are subject in numerous studies (Dirnböck & Dullinger, 2004; Dormann *et al.*, 2007, 2013; Braunisch *et al.*, 2013), and extensive care is taken in selecting uncorrelated predictor variables, differences in model performance arising from available climate data sets remains largely out of focus in most studies.

Biased climate data can lead to distorted models (Heikkinen *et al.*, 2006). Geographic and environmental biases are contrary to the assumption of many modelling techniques that localities represent a random sample from the area being modelled (Phillips *et al.*, 2006). In many cases, freely available gridded climate data sets do not satisfy the requirements of ecological climate impact studies, and complicate the investigation of climate ecosystem interactions (Soria-Auza *et al.*, 2010).

In the last decade, WORLDCLIM (Hijmans *et al.*, 2005) has been the most prominent global climate data set. Especially in Europe and Northern America, WORLDCLIM shows high accuracy (Hijmans *et al.*, 2005), and is used in numerous biogeographical studies (Elith *et al.*, 2006; Hijmans & Graham, 2006; Broennimann *et al.*, 2012). WORLDCLIM has also been used to model species distributions in the Himalayan region (Forrest *et al.*, 2012; Liu *et al.*, 2017), the accuracy, however, needs to be evaluated. Bobrowski *et al.* (2017) pointed out some drawbacks, related to the usage of WORLDCLIM.

WORLDCLIM represents a simple interpolated climate data set, which regionalizes monthly observations of precipitation and temperature based on a weighted linear regression approach, using latitude, longitude and elevation as predictor variables. Despite the high spatial raster resolution (i.e., 1 x 1 km), WORLDCLIM ignores atmospheric processes at local scale which are essential for the formation of site-specific topoclimatic conditions in high mountain environments. Many studies show that local-scale atmospheric conditions are highly influenced by the underlying terrain. Anisotropic heating at different slope positions as well as cold air drainage and pooling in mountain valleys during autochthonous weather conditions result in a complex temperature pattern, which distinctly modifies the distribution of plant communities (Bobrowski *et al.*, 2017). The spatial pattern of precipitation is affected by wind- and leeward slope positions, with hyper-humid climate conditions at the southern declivity of the Himalayan range and semi-arid to arid conditions in the Trans-Himalayan valleys.

Since 2016, a new fine-scale (i.e., 1 x 1 km), long-term climate raster data set with global coverage called CHELSA (Climatologies at high resolution for the earth' land surface areas) is available (Karger *et al.*, 2016). CHELSA was compared and evaluated with three climate data sets (i.a. WORLDCLIM), and showed similar performance for temperature, but higher performance for prediction of orographic precipitation patterns (Karger *et al.*, 2016). Both climate data sets use the same raw data to produce the same bioclimatic raster-layers. However, CHELSA represents the first global climate data set based on statistical downscaling, whereas WORLDCLIM is based on interpolation.

To date, there are only very few studies aiming at comparing and evaluating modelling results obtained by different (e.g., climate data sets (comparison of SAGA and WORLDCLIM in Soria-Auza *et al.*, 2010 using Böhner, 2006 and Hijmans *et al.*, 2005).

Comparative studies, which evaluate the performance of ecological niche models using different global climate data sets for modelling the potential distribution of Himalayan treeline tree species' or other Himalayan vascular plant species' do not exist. We selected the treeline-forming species *Betula utilis* as a target species since an improved accuracy in modelling the current distribution is a precondition for a more precise modelling of potential range expansions of treeline trees under climate change conditions (Schickhoff *et al.*, 2015).

In order to investigate the impact of each climate data set we compared the predicted current distribution of *B. utilis* in the Himalayan region. We applied Generalized Linear Models, using each climate data set respectively, to model the distribution range and compare and evaluate projected distribution range maps. We hypothesized that there will be discrepancies in the predictions of the two climate data sets. We assume a higher prediction accuracy of CHELSA because of its capability to reflect mountain-specific climatic conditions, in particular in terms of precipitation-related variables.

Material and Methods

Study area and species data

The Himalayan mountain system is located between the Tibetan Highland in the north and the Indo-Gangetic plains in the south, and extends from Afghanistan in the northwest (c. 36 °N and 70 °E) to Yunnan in the southeast (c. 26 °N and 100 °E). It is a vast mountain region, covering an area of more than 1.000.000 km^2, with a length of c. 3000 km (Pakistan to SW China) and a maximum width of 400 km.

The Himalayan mountains show a distinct three-dimensional geoecological differentiation, with complex topography and high variation of climatic and edaphic conditions. The climate ranges from tropical in the Indo-Gangetic plains to permanent ice and snow at highest elevations, and from more continental in the NW to more oceanic in the SE (Troll, 1972; Zurick & Pacheco, 2006). The distribution range of *Betula utilis* extends across the Himalayan range from Afghanistan to SW China, with the total elevational range extending from 2700 to 4500 m (Polunin & Stainton, 1984). *B. utilis* was selected as a study species due to its status as a principal broadleaved treeline species in the western and central Himalaya. The Himalayan birch mainly grows on north-facing slopes in shady locations. In the NW Himalaya, *B. utilis* is widely distributed between 3100 and 3700 m, while the range shifts to higher elevations towards the E Himalaya (mainly between 3800 and 4300 m; own database). *B. utilis* grows in mixed forests with conifers and rhododendrons and forms a narrow forest belt between coniferous forests below and a krummholz belt above (for associated tree species see Schickhoff, 2005, Miehe *et al.*, 2015). Pure birch stands with *Rhododendron* *campanulatum* and *Sorbus microphylla* in the understory are often found at the uppermost limit of subalpine forests (Schickhoff *et al.*, 2015).

Presence-only occurrence data of *B. utilis* were accessed via the Global Biodiversity Information Facility (gbif.org). The database hosts 215 geo-referenced records (1970-2016) without any known coordinate issues for our study region. 202 records were added from a database compiled from a literature survey (Schickhoff, 2005; unpublished data). Additionally, we extracted 327 records from freely available satellite images (GoogleEarth™, ver. 7.1.1.1888, Google, 2015) and added them to the dataset. Extractions from GoogleEarth have been shown to be useful in global treeline research (Paulsen & Körner, 2014; Irl *et al.*, 2015). These occurrence localities were validated through expert knowledge, obtained from numerous field visits in the Himalayan region. To reduce spatial auto-correlation, only one occurrence point per grid cell (i.e., 1 x 1 km) was considered, resulting in 590 occurrence points of *B. utilis* as a basis for modelling the current distribution of this species (Fig. 20). Lowermost occurrence records (e.g., in avalanche paths) were omitted since they do not represent the 'zonal' climatic conditions of the treeline birch belt. 6000 pseudo-absence points were generated (following the procedure for Generalized Linear Models described by Barbet-Massin *et al.*, 2012).

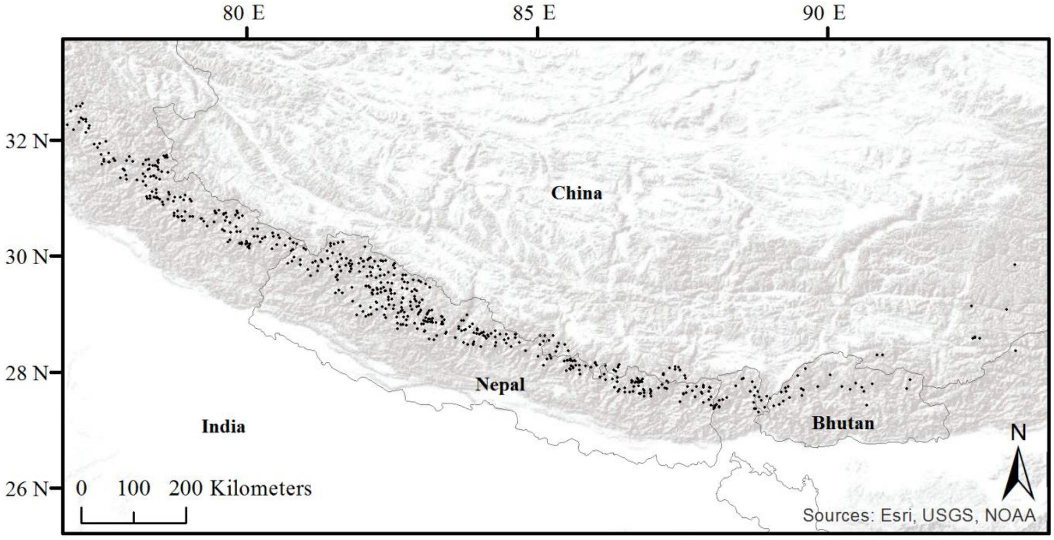


Fig. 20: Occurrences of *Betula utilis* in the Himalayan mountain system (*N*=590)

Selection of climatic variables and climate data sets

The aim of the study was to compare the performance of two different climate data sets: 1) interpolated climate data (i.e., WORLDCLIM, Hijmans *et al.*, 2005), and 2) quasi-mechanistical statistical downscaling (i.e., CHELSA, Karger *et al.*, 2016). Based on the gridded monthly fields of temperature and precipitation at a resolution of 30 arc seconds, we generated 19 climatic variables for each climate data set respectively. These climatic variables are widely used in species distribution modelling and represent annual characteristics (e.g., Mean Annual Temperature), seasonality (e.g., Annual Range in Precipitation) and extreme environmental factors (e.g., Precipitation of Driest Month) (Nix, 1986, Hijmans *et al.*, 2005). In addition, the variables Average Precipitation of May and Average Precipitation of March, April and May were calculated in order to account for potential pre-monsoon drought stress (Liang *et al.*, 2014; Schickhoff *et al.*, 2015, 2016). Furthermore, we used a DEM (ngdc.noaa.gov; Global Land One-Kilometer Base Elevation) to calculate slope angle and slope aspect using ArcGIS (Version 10.1; Esri Inc. Redlands, CA, USA). In each climate data set we checked for multi-collinearity among the variables using Spearman's rank correlation, since high collinearity might lead to low model performance and wrong interpretations (Dormann *et al.*, 2007; 2013). We calculated pairwise correlations, resulting in a small set of predictor variables ($r_s \leq 0.7$ according to Dormann *et al.*, 2013) (see Fig. 21 for selected variables, which represent general patterns and annual climatic variability, and which we used for modelling).

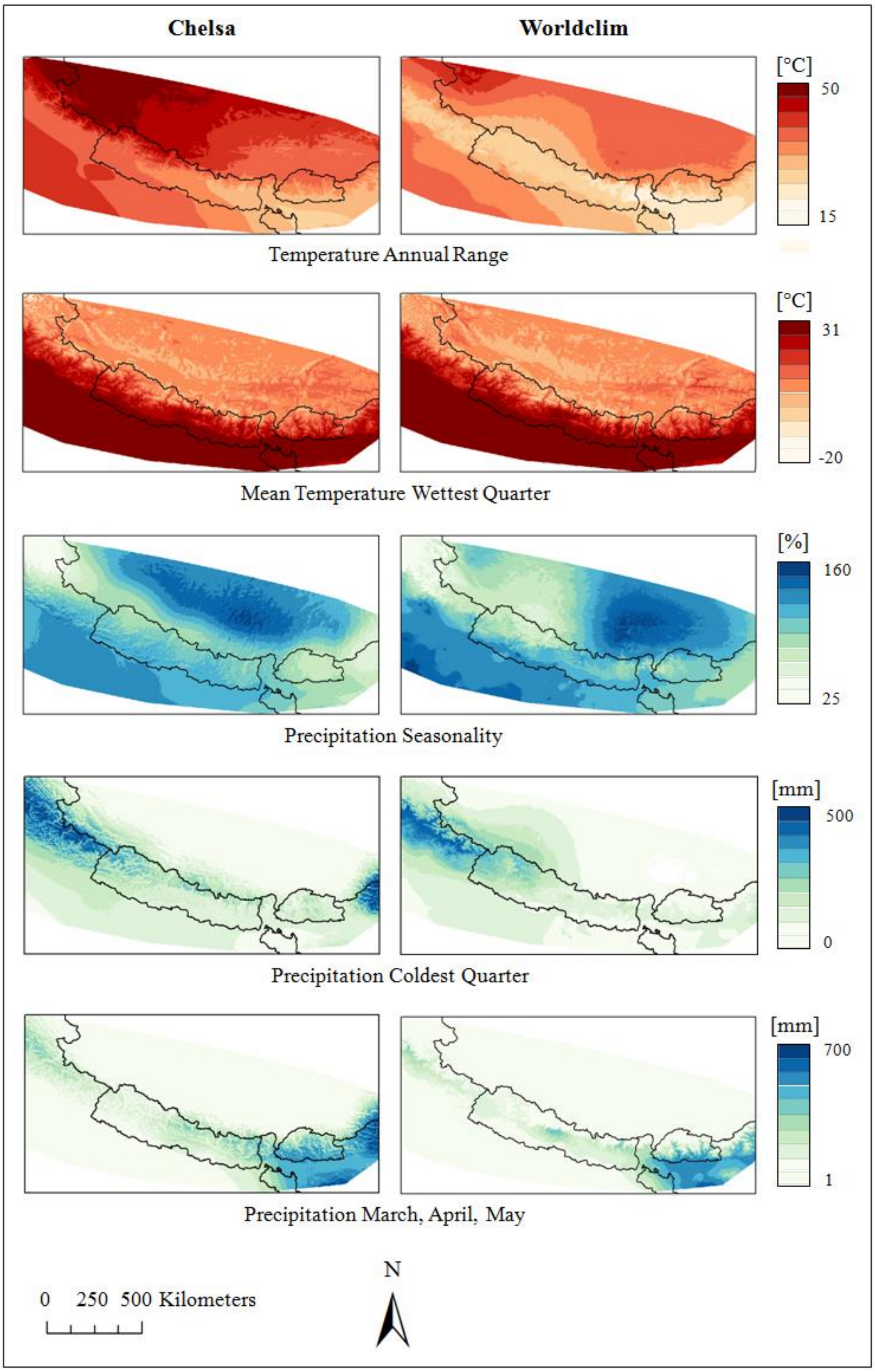


Fig. 21: Comparison of climatic variables used to model the potential distribution of *B. utilis* in the Himalayan region for each climate data set respectively.

Whereas WORLDCLIM consists of interpolated climate data, CHELSA climate data are based on a quasi-mechanistical statistical downscaling of the ERA-interim global circulation model with GPCC and GHCN bias correction and a resolution of 30 arc s (for details see Karger *et al.*, 2016). Precipitation amounts, estimated under consideration of orographic factors such as wind fields, valley exposition and boundary layer height, showed high preciseness compared to precipitation data from other climate datasets (for a comprehensive description see Karger *et al.*, 2016).

Contrasts between selected variables of each climate data set are obvious from Fig. 21. For instance, a comparison of the Temperature Annual Range between the two climate data sets reveals a higher temperature amplitude for CHELSA, whereas WORLDCLIM fails to predict the striking differences between the lowlands, the Himalayan arc and the Tibetan plateau (e.g., exemplified within the borders of Nepal). Moreover, Karger *et al.* (2016) found the same pattern for Mean Annual Temperature.

For the variable Mean Temperature of Wettest Quarter the predictions of both climate data sets were quite uniform (Table 5). Comparing Precipitation Seasonality, the same pattern of artefacts in the WORLDCLIM predictions as in the case of Temperature Annual Range is obvious. The distinction between the Himalayan range and the Tibetan plateau (e.g., western Nepal) is missing. Although distributions of predicted values differ in environmental space, maximum and minimum values show the same range in climate space. According to the predicted Precipitation of the Coldest Quarter (i.e., December, January and February), both climate data sets show low values for most parts of the Himalaya. Nevertheless, CHELSA predicts higher values and larger areas with winter precipitation (i.e., snow), whereas WORLDCLIM fails to predict winter precipitation in the eastern parts of the arc. Average Precipitation of March, April and May appears optically uniform, but CHELSA shows higher values for the eastern part of the Himalayan region (Table 5).

Table 5: Descriptive statistics for the variables of both climate data sets respectively.

	$\text{CHELSA}_{\text{Max}}$	$\text{WORLDCLIM}_{\text{Max}}$	$\text{CHELSA}_{\text{Min}}$	$\text{WORLDCLIM}_{\text{Min}}$
Annual Temp. Range [°C]	47.8	41.3	24.1	16.2
Mean Temp. Wettest Quarter [°C]	30.2	30.3	19.7	12.9
Precipitation Seasonality	158.1	156	28.4	32
Precipitation of Coldest Quarter [mm]	494.3	416	0.5	0
Precipitation March, April and May [mm]	678	420	1.3	2

Comparison of climate data sets

To test for correlation between the variables of the two climate data sets, values for each presence and pseudo-absence point were extracted and examined using Spearman's rank correlation. To account for the effects of spatial autocorrelation, we applied Dutilleul (1993) correction for adjusted *p* values. To visualize geographical discordance between the two climate data sets, variables were intersected performing cell by cell subtraction of CHELSA *minus* WORLDCLIM.

Modelling algorithm and evaluation

We used Generalized Linear Models (GLM) as they represent a classical and robust approach to analyse presence and absence data (Nelder & Wedderburn, 1972; McCullagh & Nelder, 1989). The iterative weighted linear regression technique is used to derive the maximum likelihood estimates of the parameters, with observations distributed according to an exponential family and systematic effects, which can be made linear by suitable transformation (Nelder & Wedderburn, 1972). GLMs were calculated with binomial distribution, logit-link function and polynomial terms of second order (Austin, 1999), but did not include interaction terms among predictor variables. To calculate the final GLMs, step-wise variable selection in both directions (i.e., forward and backward) was applied, using the Akaike Information Criterion (AIC, Akaike, 1974), resulting in the final and best model possessing the lowest value for AIC (Burnham & Anderson, 2002, Guisan *et al.*, 2002). For model validation, all presence and absence points were split into training and testing data samples with a ratio of 80:20 percent using random stratified sampling (Kuhn & Johnson, 2013). This procedure was repeated five times, resulting in five versions of the model, which were finally averaged. The trained GLMs were validated using statistical re-sampling techniques such as 10-fold cross-validation (Stone, 1974; Arlot & Celisse, 2010) in order to decrease bias in measuring the predictive performance (Pearce & Ferrier, 2000). The validated models were evaluated with the withheld testing data. In addition, we calculated pseudo-R^2 to account for explained variance in the dataset (Nagelkerke, 1991).

Due to the lack of an universally valid model evaluation measurement, it is essential to use more than one performance evaluation metric. We chose the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve (Fielding & Bell, 1997; Elith & Burgman, 2002) as a threshold-independent evaluation metric. As threshold-dependent measure, we calculated True Skill Statistics (TSS) (Flueck, 1987; introduced to SDM by Allouche *et al.*, 2006) to convert continuous predictions, above their respective TSS threshold (i.e., threshold that maximizes the sum of sensitivity and specificity, for details see Freeman & Moisen, 2008) to binary distribution maps.

Final projected maps include only grid cells predicted by all five model versions for both climate data sets respectively).

To quantify differences between predicted values of the climate data sets, we applied Wilcoxon test statistics. To verify differences between the areas predicted as suitable between the two climate data sets, final maps were intersected (i.e., cell by cell subtraction of CHELSA *minus* WORLDCLIM) and the TSS_{MAP} was calculated (Fig. 24).

In this study all statistical analyses were performed using the programming language R (R Core Team, 2015, version: 3.1.3).

All maps were created using ArcGIS (Version 10.1; Esri Inc. Redlands, CA, USA).

Results

Comparison between climate data sets

Correlations between corresponding climate variables yielded partially high correlation coefficients (Table 6). As for temperature-related variables, the highest correlation coefficient was found for Temperature of the Wettest Quarter (r_S= 0.98, $p \leq 0.001$). For Temperature Annual Range the correlation coefficient was moderate, but still highly significant (r_S= 0.62, $p \leq 0.001$). Regarding precipitation-related variables, Average Precipitation of March, April and May yielded the highest correlation coefficient (r_S= 0.76, $p \leq 0.001$), followed by Precipitation of the Coldest Quarter (r_S= 0.73, $p \leq 0.001$) and Precipitation Seasonality (r_S= 0.62, $p \leq 0.001$).

Table 6: Bivariate spearman's rank correlation between the corresponding variables of the two climate data sets. Degrees of freedom were corrected according to Dutilleul's criterion (Dutilleul, 1993).

Climatic Variable	r_S	*p* value	DF (Dutilleul's corrections)
Temperature Annual Range	0.62	≤0.001	21.67
Temperature Wettest Quarter	0.98	≤0.001	13.72
Precipitation Seasonality	0.62	≤0.001	16.00
Precipitation of Coldest Quarter	0.73	≤0.001	14.25
Precipitation March, April, May	0.76	≤0.001	31.85

To identify the areas of discrepancies, we produced maps of disagreement between the climate variables (Fig. 22). For the variable Temperature Annual Range, CHELSA predicts almost everywhere higher values than WORLDCLIM. We detected extraordinary disagreement in the western parts of the study area and in western parts of Nepal. Mean Temperature of Wettest Quarter represents the variable that is almost identical between the two climate data sets (i.e., no major discrepancies (yellow)). Precipitation Seasonality shows two contrasting patterns with high values for the western part of the Himalayas and central part of the Tibetan plateau (CHELSA) and higher values throughout the eastern Himalayan region (WORLDCLIM). Precipitation of the Coldest Quarter exhibits quite large areas with uniform agreement between the climate data sets. However, disagreement is obvious at the margins of the study area, especially along the arc in the eastern part, with CHELSA predicting higher values for this variable. The intersected map of Average Precipitation of March, April and May shows a similar pattern.. CHELSA predicts higher values at the margins of the study area and eastern part of Nepal, whereas WORLDCLIM predicts higher values on the Tibetan plateau and below the range in the western parts of the study area.

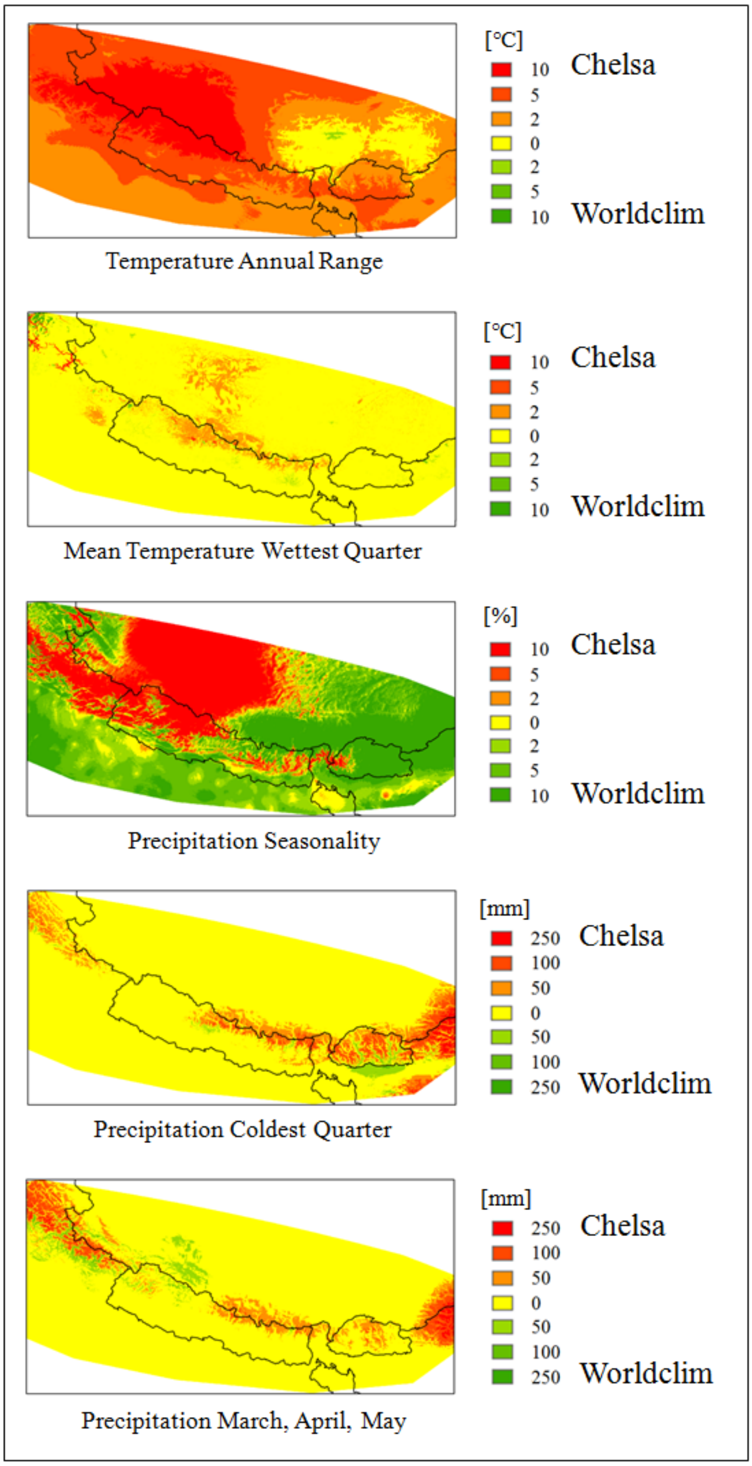


Fig. 22: Maps of disagreement between Chelsa and Worldclim for each variable used in the modelling approach. Reddish colours represent higher values for Chelsa compared to Worldclim, and greenish colours higher values in the Worldclim data set. Yellow implies no difference between the climate data sets. For colour-blind friendly figure please see Supplementary material S7 Fig., S7.

Modelling species distribution

The calculation of pseudo-R^2 revealed for the CHELSA model 80.6 % explained variance, whereas for WORLDCLIM 67.7 % variance could be explained in the test dataset.

Due to high disagreement of some of the variables between the two climate data sets (Fig. 21) and significant differences in explained variance of the models, high discrepancies in response curves of the models can be assumed. A comparison of response curves of selected predictor variables of the two climate data sets showed distinctive differences for most variables (Fig. 23).

For the variable Temperature Annual Range both climate data sets show concave curves, whereas response types differed. CHELSA predicted a unimodal response, contrary to the sigmoidally switched response by WORLDCLIM. In addition, temperature ranges differed. Both climate data sets showed a close-to-normal distribution for Mean Temperature of Wettest Quarter, but CHELSA showed an optimum around 10°C, whereas WORLDCLIM predicted a wider range of temperature with an optimum around 15°C. Striking differences in predicted probabilities were observed for Precipitation Seasonality and Precipitation of the Coldest Quarter. CHELSA predicted a close-to-normal distribution and a sigmoid response curve, respectively, whereas predictions of WORLDCLIM were inverted. The curve of CHELSA for Average Precipitation of March-May showed a sigmoidal response, while WORLDCLIM predicted a unimodal response curve.

For models using CHELSA climatic variables, highest relative variable importance was found for Mean Temperature of Wettest Quarter, followed by Precipitation of Coldest Quarter, whereas lowest variable importance was found for Average Precipitation March-May. By contrast, WORLDCLIM variables showed highest variable importance for Temperature Annual Range, followed by Mean Temperature of Wettest Quarter. All precipitation-related variables had much lower values for variable importance, whereas the lowest value was found for Precipitation Seasonality.

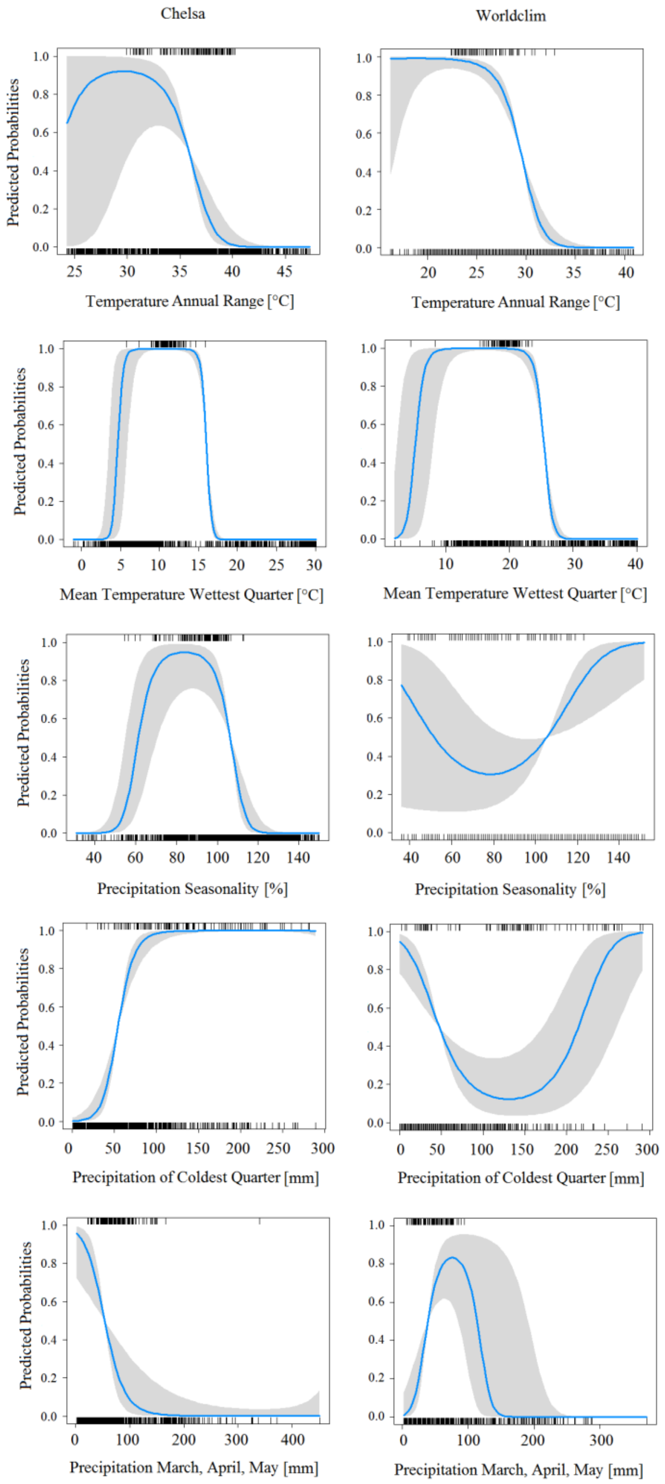


Fig. 23: Response curves for the climatic variables used for modelling the potential distribution of *Betula utilis*, averaged over 5 model runs for each climate data set (CHELSA left, and WORLDCLIM right) respectively. Grey-shaded areas represent the confidence interval of the predicted probabilities. Response curves do not account for interactions between the variables.

Model evaluation and comparison

The threshold-independent “Area Under the Curve” (AUC) of the receiver-operating characteristic (ROC) plot approach was very high for both climate data sets (Chelsa: $AUC = 0.98$, $AUC_{SD} = 0.0041$; Worldclim: $AUC = 0.97$, $AUC_{SD} = 0.0093$). Wilcoxon test statistics revealed no significant differences in AUC values between the two climate data sets ($z = 2.0226$, $p = 0.0625$).

Due to the fact, that the AUC does not represent a standardized measure of predictive accuracy, threshold dependent measures were calculated to assess model quality. In addition, the threshold-dependent measure TSS was used to evaluate the prediction accuracy of the models and to convert the continuous maps into binary distribution maps. In order to obtain binary distribution maps, continuous maps were thresholded at the averaged optimal threshold (0.15). Confusion matrices were then used to calculate the True Skill Statistics. Despite high values of the calculated TSS for both climate data sets, Chelsa gained a higher averaged TSS ($TSS \geq 0.93$, $TSS_{SD} = -0.99$) than Worldclim ($TSS \geq 0.87$, $TSS_{SD} = -0.99$), binary distribution maps show noticeable differences (Fig. 24). Wilcoxon test statistics revealed high significant differences in the averaged TSS value between the models ($z = -2.5236$, $p \leq 0.01$).

It is apparent, that the suitable habitat predicted by Worldclim tends to be much larger in the western and central region of the Himalayan mountain system compared to the predictions of Chelsa, which shows a uniform distribution along the Himalayan arc (Fig. 24). However, areas predicted as potential habitat of *B. utilis* in the eastern part of the study area are practically absent in the Worldclim projection (Fig. 24). The calculated TSS for the intersected area revealed excellent model prediction accuracy ($TSS_{MAP} = 0.9$).

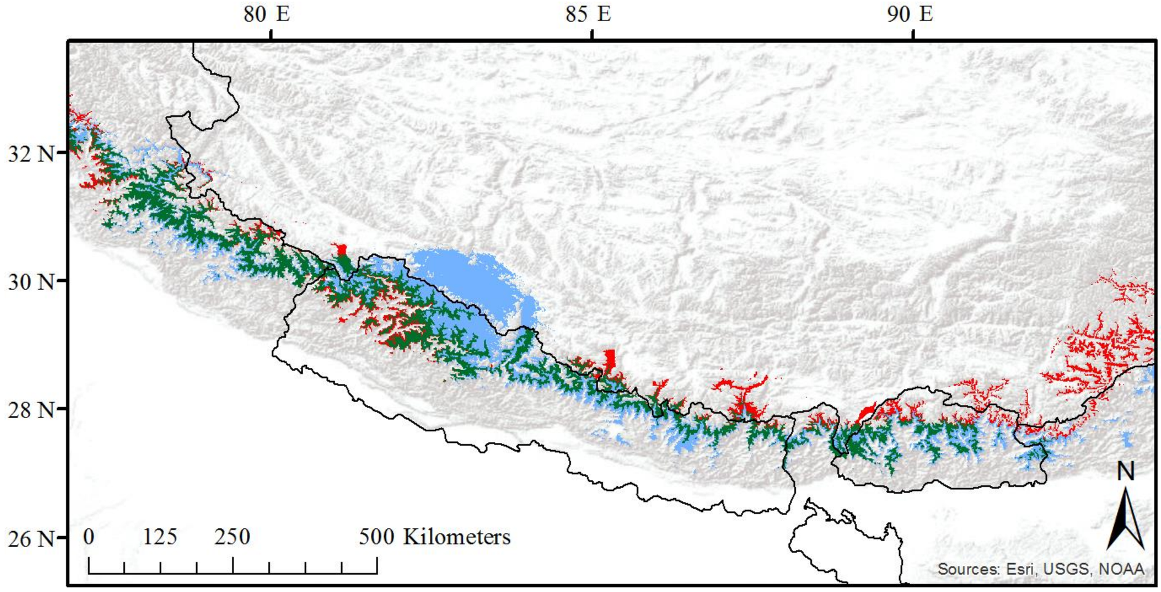


Fig. 24: Intersection of binary projections by both climate data sets (areas predicted by CHELSA climate data = red; areas predicted by WORLDCLIM climate data= blue; areas predicted by both climate data sets= green). For colour-blind friendly figure please see Supplementary material S8 Fig. S8.

Discussion

Comparison of the climate data sets

In ecological niche modelling, the evaluation of model performance using different climate data sets has rarely been addressed so far. This study compares for the first time CHELSA and WORLDCLIM climate data to model the potential distribution of *B. utilis* in the Himalayan region. Unlike CHELSA, which only recently became available (Karger *et al.*, 2016), WORLDCLIM climate data (Hijmans *et al.*, 2005) have been widely used (6284 citations in ISI Web of Knowledge in February 2017), usually yielding reliable results (Elith *et al.*, 2006; Pearman *et al.*, 2008; Casalegno *et al.*, 2010). Especially in Europe, North America and in some tropical areas, where the density of climate stations is higher than in South America, Africa and Asia, the accuracy of WORLDCLIM has been evaluated (see Fig 1a-c in Hijmans *et al.*, 2005). Despite different methods of computation, CHELSA and WORLDCLIM contain disparities, some of which may be of extraordinary importance for vast mountain systems like the Himalaya.

In topographically complex areas like the Himalayas, climate stations are quite rare due to rough terrain and complicated accessibility. This leads in turn to a poor data basis for calculating climate data sets compared to other more accessible terrain. In addition, climate stations are prevailingly located near settlements at lower elevations, where climatic conditions are most suitable for habitation, livestock farming and agriculture. Those climate stations are not representative for climatic conditions at higher elevations. Despite the high spatial raster resolution of approximately 1km^2, the WORLDCLIM climate data set neglects local-scale atmospheric processes which cause site-specific topoclimatic conditions in high mountain environments. For instance, anisotropic heating at different slope positions as well as cold air drainage and pooling in mountain valleys during autochthonous weather conditions result in a complex temperature pattern, which may distinctly modify the distribution of vegetation types.

A comparison of temperature data sets yielded the result that CHELSA can keep up with competing data sets (i.a. WORLDCLIM) (Karger *et al.*, 2016). Correspondingly, the comparison of temperature-related variables (e.g., Temperature of Wettest Quarter) between the two climate data sets shows high consistency (Fig. 22). Temperature is strongly negatively correlated with altitude (except for cold air inversions in winter months), allowing little room for variation in regional-scale climate data sets (Soria-Auza *et al.*, 2010, Karger *et al.*, 2016).

A different picture emerges for precipitation patterns, which are influenced by a variety of factors like wind currents, topography, the diurnal cycle of solar radiation that changes air pressure and therefore causes differential precipitation along altitudinal gradients (forcing factors in regional circulation models). Such complex interactions cause a more irregular spatial pattern of precipitation in mountain regions that is not directly correlated to either latitude or altitude. Dissimilarities in patterns of cloud formation become apparent between CHELSA and WORLDCLIM (Karger *et al.,* 2016). In complex topographic terrain, a high correlation between the formation of clouds and precipitation patterns is assumed by Karger *et al.* (2016), who claim that CHELSA is capable of capturing the topographic heterogeneity of precipitation at small spatial scales. They also compared the mean annual precipitation sums between numerous climate data sets. In summary, CHELSA shows a more consistent relation between the terrain and the resulting precipitation patterns, whereas the algorithm of WORLDCLIM produces flawed correlations between elevation and precipitation (Karger *et al.*, 2016). Generally, CHELSA predicts higher precipitation amounts compared to WORLDCLIM (cf. Fig. 3 in Karger *et al.*, 2016). Regarding precipitation-related variables in our study (e.g., Precipitation of the Coldest Quarter and Average Precipitation of March, April and May) higher values could be confirmed for CHELSA compared to WORLDCLIM (Fig. 22). Throughout the Himalayan region, precipitation is highest during the summer monsoon, when moisture-laden air currents are forced to rise at the Himalayan arc and release rainfall until they reach the Tibetan plateau. Clear differences were visible at microscale: WORLDCLIM predicted wet valleys and dry slopes, whereas CHELSA predicted dry valleys and orographic precipitation at higher-elevated wind-exposed slopes (Karger *et al.*, 2016). However, CHELSA predictions corresponded closely with the observed distribution of clouds in the area, which shows lower cloud cover in the isolated mountain valleys compared to the wind-exposed slopes of the southern declivity (Karger *et al.*, 2016).

Modelling the ecological niche of Betula utilis

The distributional range of *B. utilis* was found to be primarily associated with thermal (Temperature Annual Range and Mean Temperature of the Wettest Quarter) and precipitation-related factors (Precipitation Seasonality, Precipitation of the Coldest Quarter and Average Precipitation of March, April and May). For detailed discussion on variable interpretation see Bobrowski *et al.*, 2017.

Although WORLDCLIM climate data have been used to model species in the Himalayan region (Forrest *et al.*, 2012), no comparative studies exist yet. Furthermore, there are hardly any studies investigating the current distribution of *B. utilis*. Singh *et al.* (2013) modelled the distribution of *B. utilis* in Uttrakhand, India, using WORLDCLIM climate data. The performance of different modelling techniques was evaluated for modelling i.a. *B. platyphylla* in China with WORLDCLIM climate data (Duan *et al.*, 2014). Song *et al.* (2004) modelled current and future distribution of *B. platyphylla* on the Tibetean plateau using PRISM climate data (Daly *et al.*, 2000; Daly *et al.*, 2002). To date, hardly any modelling studies focusing on treeline tree species have been conducted in the Himalayan region, although investigating the response of Himalayan treelines to climate change is a very active research field (Schickhoff *et al.* 2015, 2016).

Apart from a lower fit of the GLMs using WORLDCLIM climate data compared to CHELSA, variable contribution differed between the final models (Fig. 23). For CHELSA-related models, Mean Temperature of the Wettest Quarter and Precipitation of Coldest Quarter were the most important variables when modelling *B. utilis*, followed by Annual Temperature Range and Precipitation Seasonality. In contrast, WORLDCLIM showed highest importance for Temperature Annual Range, followed by Mean Temperature of Wettest Quarter, whereas all remaining precipitation-related variables played a minor role for variable importance. Generally, response curves of models using CHELSA climate were better able to distinguish between presences and pseudo-absences (showed a higher model fit). For the variable Temperature Annual Range, CHELSA showed higher values of temperature amplitude than WORLDCLIM. For Mean Temperature of the Wettest Quarter, curve shapes are similar, but models disagree in predicted probability ranges. CHELSA pointed to a narrower temperature range for the ecological niche of *B. utilis* than WORLDCLIM. For the variables Precipitation Seasonality and Precipitation of the Coldest Quarter, differences between the two models are striking. The response curve of predicted probability of WORLDCLIM completely fails to distinguish between presences and pseudo-absences of *B. utilis* (Fig. 23). Regarding Average Precipitation of March, April and May, CHELSA predicted lower values compared to WORLDCLIM for estimating the climatic space of *B. utilis*.

After calibrating the models, they were evaluated with threshold-independent (i.e., AUC) and threshold-dependent performance measures (i.e., TSS). Admitting that recent research criticises the use of AUC to evaluate model performance (Lobo *et al.*, 2008), as it does not reflect absolute, but relative model performance, we applied the AUC as a threshold-independent evaluation measure. According to Peterson *et al.* (2011), it is appropriate when models for a single study species in a single study region are compared. We found no significant differences in AUC values; they were very high ($\text{AUC} \geq 0.97$) for both models. Using the threshold-dependent measure TSS, continuous maps were converted to binary distribution maps. Notwithstanding high values ($\text{TSS}_{\text{CHELSA}} \geq 0.93$; $\text{TSS}_{\text{WORLDCLIM}} \geq 0.87$), we found significant differences between the models. Models using CHELSA climate data gained a higher model fit and higher evaluation metrics (i.e., TSS and Pseudo R^2). Also binary distribution maps show noticeable differences between the model projections (Fig. 24).

We conducted visual inspections of the predictions of the models to observe whether the model successfully predicted the species' realized distribution throughout the entire Himalayan mountain system. *B. utilis* covers a notable environmental niche width along the Himalayan arc. The projections of the modelling results show that most favourable conditions are located in the upper subalpine belt in the western and central Himalayan region where climatic conditions show a comparatively higher degree of continentality. This is consistent with the distribution pattern documented in several vegetation maps showing a narrow band of birch forests forming the upper treeline on north-facing slopes (Troll, 1939; Schweinfurth, 1957; Miehe, 1991; Schickhoff, 1994; Braun, 1996; Nüsser & Dickoré, 2002; Eberhardt *et al.*, 2007). Good agreement between the two projections can be found in the western part of the study area (i.e., Himachal Pradesh and Uttarakhand, India), whereas eye-catching differences can be found in Far West, Midwest and West Nepal (i.e., Humla, Jumla, Mugu, Dolpa (Karnali region), Mustang (Dhaulagiri region) and Tibet (autonomous region of China)). These differences may be attributed to artefacts in the WORLDCLIM climate data (Fig. 21). Towards the eastern Himalaya, where more maritime climatic conditions favour the competitiveness of evergreen *Rhododendron* spp., *B. utilis* becomes a less frequent companion in subalpine forests and at treelines (Schickhoff, 2005). In these areas, further disagreements between the predictions arise. In contrast to CHELSA, WORLDCLIM fails to predict any climatic suitable areas for *B. utilis* in Arunachal Pradesh and SE Tibet, leading to the assumption of non-availability of suitable climate in the eastern Himalaya (Fig. 24).

Conclusions

CHELSA located the ecological niche of *B. utilis* at higher elevations. In addition, the modelled niche tends to be less diffuse compared to WORLDCLIM (Fig. 24). In fact, the ecological niche modelled by CHELSA are in closer correspondence to the authors' field knowledge, and the model predictions match the actual existing distribution range of *B. utilis* to a vast extent.

These findings expand on former research on *B. utilis* (Bobrowski *et al.*, 2017), confirming that climate data, which reflect topoclimatic conditions, yield more accurate results.

Although the need for comparative studies of model performance using different climate data sets gained potential interest over the last few years (e.g., Kriticos *et al.*, 2012; Watling *et al.*, 2014; Wang *et al.*, 2016), up to date WORLDCLIM still remains the most often used climate data set on regional and global scale. Global climate data sets should not be used to model ecological niches without critically scrutinizing the origin of climate data and the computation method of the climate data set, and without being aware of the afflicted limitations revealed in this study. This may be particularly true for modelling studies in the Himalayan mountain system, since, as we pointed out, the naïve application of WORLDCLIM in heterogeneous landscapes like the Himalayan region could be misleading. Further challenges arise from the scale of the raster climate data. Especially for European high mountain regions, several studies exist where local climate sets were used to model species distribution ranges (e.g., Climate Research Unit (CRU) by Thuiller 2004; Thuiller *et al.*, 2005; Austria climate station data by Dullinger *et al.*, 2004), which points to the need of small scale climate data for modelling alpine species (i.e., smaller than $1km^2$).

Our results provide compelling evidence for the need of more critical, reflective and responsible investigations on the effect of climatic variables in ecological niche modelling, especially in remote mountain regions. In conclusion, a better understanding of the impact of different climate data sets in modelling species potential ranges is supported by our model results that may trigger follow-up studies on the exploration of potential impacts of different climate data sets on modelling species potential distribution in regions with limited data availability. This could gain extraordinary importance with regard to range shifts under climate change scenarios.

Acknowledgements

We would like to express our gratitude to Himalayan colleagues, guides and local people who accompanied us on numerous field trips to *Betula* treelines. This study was carried out in the framework of the TREELINE project and partially funded by the German Research Foundation (DFG-SCHI 436/14-1). We would like to thank two anonymous reviewers for their diligent work and thoughtful suggestions on the earlier version of the manuscript.

References

Akaike, H. (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, 716-723.

Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43, 1223-1232.

Anderson, R.P. & Gonzalez Jr, I. (2011) Species-specific tuning increases robustness to sampling bias in models of species distributions: an implementation with Maxent. *Ecological Modelling*, 222, 2796-2811.

Araújo, M. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33, 1677-1688.

Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, 22, 42-47.

Araújo, M.B., Pearson, R.G., Thuiller, W., & Erhard, M. 2005. Validation of species-climate impact models under climate change. *Global Change Biology*, 11, 1504-1513.

Arlot, S. & Celisse, A. (2010) A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 40-79.

Austin M.P. (1999) A silent clash of paradigms: some inconsistencies in community ecology. *Oikos*, 86, 170-178.

Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution*, 3, 327-338.

Bobrowski, M., Gerlitz, L. & Schickhoff, U. Modelling the potential distribution of *Betula utilis* in the Himalaya. *Global Ecology and Conservation*, in press.

Böhner, J. (2006) General climatic controls and topoclimatic variations in Central and High Asia. *Boreas*, 35, 279-295.

Braun, G. (1996) *Vegetationsgeographische Untersuchungen im NW-Karakorum (Pakistan). Kartierung der aktuellen Vegetation und Rekonstruktion der aktuellen Waldverbreitung auf der Basis von Satellitendaten, Gelände- und Einstrahlungsmodellen*. Bonner Geographische Abhandlungen 93, Dümmlers, Bonn.

Braunisch, V., Coppes, J., Arlettaz, R., Suchant, R., Schmid, H. and Bollmann, K. (2013) Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. *Ecography*, 36, 971-983.

Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., Thuiller, W., Fortin, M.-J., Randin, C., Zimmermann, N. E., Graham, C. H. and Guisan, A. (2012) Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, 21, 481-497.

Burnham, K. P. & Anderson, D. R. (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd ed., Springer, New York.

Casalegno, S., Amatulli, G., Camia, A., Nelson, A. & Pekkarinen, A. (2010) Vulnerability of *Pinus cembra* L. in the Alps and the Carpathian mountains under present and future climates. *Forest*

Ecology and Management, 259, 750-761.

Daly, C., Gibson, W.P., Hannaway, D., Taylor, G.H. (2000) Development of new climate and plant adaptation maps for China. In: *Proceedings of the 12th Conference on Applied Climatology*, Ashville, NC, 8-11 May 2000. Boston, MA: American Meteorological Society.

Daly, C., Gibson, W.P., Taylor, G.H., Johnson, G.L. & Pasteris, P. (2002) A knowledge-based approach to the statistical mapping of climate. *Climate Research*, 22, 99-113.

Dirnböck, T. & Dullinger, S. (2004) Habitat distribution models, spatial autocorrelation, functional traits and dispersal capacity of alpine plant species. *Journal of Vegetation Science*, 15, 77-84.

Dormann, C.F., McPherson, J.M., Araujo, M.B., Bivand, R., Bolliger, J., Carl, G. *et al.* (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30, 609-628.

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. & Lautenbach, S. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27-46.

Duan, R.-Y., Kong, X.-Q., Huang, M.-Y., Fan, W.-Y., & Wang, Z.-G. (2014). The Predictive Performance and Stability of Six Species Distribution Models. *PLoS ONE*, 9, 11.

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.

Dutilleul, P. (1993) Modifying the t test for assessing the correlation between two spatial processes. *Biometrics*, 49, 305-314.

Eberhardt, E., Dickoré, W.B. & Miehe, G. (2007) Vegetation map of the Batura Valley (Hunza Karakorum, North Pakistan). *Erdkunde*, 61, 93-112.

Elith, J. & Burgmann, M. (2002) Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. *Predicting species occurrences: issues of accuracy and scale* (ed. by J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall & F.B. Samson), pp. 303-313. Island Press, Washington, DC.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann N.E. (2006) Novel methods improve prediction of species distributions from occurrence data. *Ecography*, 29, 129-151.

ESRI, 2012. ArcGIS Desktop: Release 10.1. Environmental Systems Research Institute, Redlands, CA.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24, 38-49.

Flueck, J.A. (1987) A study of some measures of forecast verification. *Preprints: Tenth conference on probability* *and statistics in the atmospheric sciences* (Edmonton, Alta., Canada), pp. 69-73. American Meteorological Society, Boston.

Freeman, E. A. & Moisen, G.G. (2008b) PresenceAbsence: An R package for presence-absences model analysis. *Journal of Statistical Software*, **23**, 1-31.

Forrest, J.L., Wikramanayakea, E., Shrestha R., Areendrand, G., Gyeltshen, K., Maheshwari, A., Mazumdar, S., Naidoo, R., Thapa, G.J. & Thapa, K. (2012) Assessing the vulnerability of snow leopard habitat to treeline shift in the Himalaya. *Biological Conservation*, 150, 129-135.

Franklin, J. (1995) Predictive vegetation mapping - geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, 19, 474-499.

gbif.org: Biodiversity occurrence data provided by: Missouri Botanical Garden*, Royal Botanic Garden Edinburgh and T*he Himalayan Uplands Plant database*, Accessed through GBIF Data Portal, http://www.gbif.org*.

Guisan, A., Edwards Jr., T.C. & Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157, 89-100.

Heikkinen, R.K., Luoto, M., Araujo, M.B., Virkkala, R., Thuiller, W. & Sykes, M.T. (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, 30, 751-777.

Hijmans, R. J. & Graham, C. H. (2006) The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology*, 12, 2272-2281.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.

Irl, S.D.H., Anthelme, F., Harter, D.E.V., Jentsch, A., Lotter, E., Steinbauer M.J. & Beierkuhnlein, C. (2015) Patterns of island treeline elevation – a global perspective. *Ecography*, 38, 1-10.

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N., Linder, H.P. & Kessler, M. (2016) Climatologies at high resolution for the earth land surface areas. arXiv:1607.00217 [physics].

Kriticos, D. J., Webber, B. L., Leriche, A., Ota, N., Macadam, I., Bathols, J. and Scott, J. K. (2012) CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution*, 3, 53-64.

Kuhn, M. & Johnson, K. (2013) *Applied predictive modeling*. Springer.

Liang, E., Dawadi, B., Pederson, N. & Eckstein, D. (2014) Is the growth of birch at the upper timberline in the Himalayas limited by moisture or by temperature? *Ecology*, 95, 2453-2465.

Liu, L., Zhao, Z., Zhang, Y. & Wu, X. (2017) Using MaxEnt model to predict suitable habitat changes for key protected species in Koshi Basin, Central Himalayas. *Journal of Resources and Ecology*, 8, 77-87.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17, 145-151.

Loiselle, B. A., Jørgensen, P. M., Consiglio, T., Jiménez, I., Blake, J. G., Lohmann, L. G. & Montiel, O. M. (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes?. *Journal of Biogeography*, 35, 105-116.

Luoto, M., Virkkala, R. & Heikkinen, R.K. (2007) The role of land cover in bioclimatic models depends on spatial resolution. *Global Ecology and Biogeography*, 16, 34-42.

McCullagh, P. & Nelder, J. A. (1989) *Generalized linear models*. Chapman and Hall, London.

Miehe, G. (1991) Die Vegetationskarte des Khumbu Himal (Mt. Everest-Südabdachung) 1:50.000. Gefügemuster der Vegetation und Probleme der Kartierung. *Erdkunde*, 45, 81-94.

Miehe, G. (2015) Landscapes of Nepal. *Nepal: an introduction to the natural history, ecology and human environment of the Himalayas* (ed. by G. Miehe, C.A. Pendry and R. Chaudhary), pp. 7-16. Royal Botanic Garden, Edinburgh.

Miehe, G., Miehe, S., Böhner, J., Bäumler, R., Ghimire, S.K., Bhattarai, K., Chaudhary, R.P., Subedi, M., Jha, P.K. & Pendry, C. (2015) Vegetation ecology. *Nepal: an introduction to the natural history, ecology and human environment of the Himalayas* (ed. by G. Miehe, C.A. Pendry and R. Chaudhary), pp. 385-472. Royal Botanic Garden, Edinburgh.

Nagelkerke, N. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, 78, 691-692.

Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K. & Toxopeus, A.G. (2013) Where is positional uncertainty a problem for species distribution modelling? *Ecography*, 36, 001-013.

Nelder, J.A. & Wedderburn, R.W.M. (1972) Generalized linear models. *Journal of Royal Statistical Society*, Series A, 135, 370-384.

Nix, H.A. (1986) A biogeogaphic analysis of Australian Elapid snakes. *Atlas of Australian Elapid Snakes, Australian Flora and Fauna*, Series 8. (ed. by R. Longmore), pp. 4-15. Australian Government Publishing Service, Canberra.

Nüsser, M. & Dickoré, W.B. (2002) A tangle in the triangle: Vegetation map of the eastern Hindukush (Chitral, northern Pakistan). *Erdkunde*, 56, 37-59.

Paulsen, J. & Körner, C. (2014) A climate-based model to predict potential treeline position around the globe. *Alpine Botany*, 124, 1-12.

Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133, 225-245.

Pearman, P.B., Guisan, A., Broennimann, O. & Randin, C.F. (2008) Niche dynamics in space and time. *Trends in Ecology and Evolution*, 23, 149-158.

Pearson, R. G. & Dawson, T. P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?. *Global Ecology and Biogeography*, 12, 361-371.

Peterson A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011*) Ecological niches and geographic distibutions*. Princeton University Press, Princeton, NJ.

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

Polunin, O. & Stainton, A. (1984) *Flowers of the Himalaya*. Oxford University Press, New Delhi.

R Core Team, 2015, version: 3.1.3 R Core Team (2013). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/.

Schickhoff, U. (1994) Die Verbreitung der Vegetation im Kaghan-Tal (Westhimalaya, Pakistan) und ihre kartographische Darstellung im Maßstab 1:150.000. *Erdkunde*, 48, 92-110.

Schickhoff, U. (2005) The upper timberline in the Himalaya, Hindu Kush and Karakorum: a review of geographical and ecological aspects. *Mountain ecosystems. Studies in treeline ecology* (ed. by G. Broll and B. Keplin), pp. 275-354. Springer, Berlin - Heidelberg.

Schickhoff, U., Bobrowski, M., Böhner, J., Bürzle, B., Chaudhary, R.P., Gerlitz, L., Lange, J., Müller, M., Scholten, T. & Schwab, N. (2016) Climate change and treeline dynamics in the Himalaya. *Climate Change, Glacier Response, and Vegetation Dynamics in the Himalaya* (ed. by R.B. Singh, U. Schickhoff and S. Mal), pp. 271-306. Springer, Switzerland.

Schickhoff, U., Bobrowski, M., Jürgen Böhner, J., Bürzle, B., Chaudari, R.P., Gerlitz, L., Heyken, H., Lange, J., Müller, M., Scholten, T., Schwab, N. & Wedegärtner, R. (2015) Do Himalayan treelines respond to recent climate change? An evaluation of sensitivity indicators. *Earth System Dynamics*, 6, 245-265.

Schweinfurth, U. (1957) Die horizontale und vertikale Verbreitung der Vegetation im Himalaya. *Bonner Geographische Abhandlungen* 20, Dümmlers, Bonn.

Singh, C.P., Panigrahy, S., Parihar, J.S. & Dharaiya, N. (2013) Modeling environmental niche of Himalayan birch and remote sensing based vicarious validation. *Tropical Ecology*, 54, 321-329.

Soberón, J. & Nakamura, M. (2009) Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences*, 106, 19644-19650.

Song, M., Zhou, C. & Ouyang, H. (2004) Distributions of dominant tree species on the Tibetan Plateau under current and future climate scenarios. *Mountain Research and Development*, 24, 166-173.

Soria-Auza, R., Kessler, M., Bach, K., Barajas-Barbosa, P., Lehnert, M., Herzog, S. & 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, 1221-1229.

Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B: Methodological*, 36, 111-147.

Thuiller, W. (2004) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, 10, 2220-2227.

Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T., Prentice, I.C. (2005) Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences*, USA, 102, 8245-8250.

Thuiller, W., Albert, C., Araujo, M.B., Berry, P.M., Cabeza, M., *et al.* (2008) Predicting global change impacts on plant species' distributions: future challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, 9, 137-52.

Trivedi, M.R., Berry, P.M., Morecroft, M.D. & Dawson, T.P. (2008) Spatial scale affects bioclimate model projections of climate change impacts on mountain plants. *Global Change Biology*, 14, 1089-1103.

Troll, C. (1972) The three-dimensional zonation of the Himalayan system. Geoecology of the high-mountain regions of Eurasia. *Erdwissenschaftliche Forschung IV* (ed. by C. Troll), pp. 264-275. Franz Steiner Verlag, Wiesbaden.

Veloz, S.D. (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, 36, 2290-2299.

Wang, T., Hamann, A., Spittlehouse, D. & Carroll, C. (2016) Locally Downscaled and Spatially Customizable Climate Data for Historical and Future Periods for North America. *PLoS ONE*, 11,6.

Watling, J.I., Fletcher, R.J. Jr., Speroterra, C., Bucklin, D.N., Brandt, L.A., Romañach, S.S., Pearlstine, L.G., Escribano, Y. & Mazzotti, F.J. (2014) Assessing effects of variation in global climate data sets on spatial predictions from climate envelope models. *Journal of Fish and Wildlife Management* 5, 14–25.

Zurick, D. & Pacheco, J. (2006) *Illustrated atlas of the Himalaya.* The University Press of Kentucky, Lexington.

Article III

Bobrowski, M.; Bechtel, B.; Böhner, J.; Oldeland, J.; Weidinger, J.; Schickhoff, U. (2018) Upgrading ecological niche models with phenological traits: Refinement of the predicted distribution range of *Betula utilis* in the Himalayan region. *Remote Sensing*. **10**, 814; doi:10.3390/rs10060814.

Abstract

Modelling ecological niches across vast distribution ranges in remote, high mountain regions like the Himalayas faces several data limitations, in particular nonavailability of species occurrence data and fine-scale environmental information of sufficiently high quality. Remotely sensed data provide key advantages such as frequent, complete, and long-term observations of land surface parameters with full spatial coverage. The objective of this study is to evaluate modelled climate data as well as remotely sensed data for modelling the ecological niche of *Betula utilis* in the subalpine and alpine belts of the Himalayan region covering the entire Himalayan arc. Using generalized linear models (GLM), we aim at testing factors controlling the species distribution under current climate conditions. We evaluate the additional predictive capacity of remotely sensed variables, namely remotely sensed topography and vegetation phenology data (phenological traits), as well as the capability to substitute bioclimatic variables from downscaled numerical models by remotely sensed annual land surface temperature parameters. The best performing model utilized bioclimatic variables, topography, and phenological traits, and explained over 69% of variance, while models exclusively based on remotely sensed data reached 65% of explained variance. In summary, models based on bioclimatic variables and topography combined with phenological traits led to a refined prediction of the current niche of *B. utilis*, whereas models using solely climate data consistently resulted in overpredictions. Our results suggest that remotely sensed phenological traits can be applied beneficially as supplements to improve model accuracy and to refine the prediction of the species niche. We conclude that the combination of remotely sensed land surface temperature parameters is promising, in particular in regions where sufficient fine-scale climate data are not available.

Keywords: *Betula utilis*; Chelsa; ecological niche model; Enhanced Vegetation Index; Himalaya; MODIS (Moderate-resolution Imaging Spectroradiometer) Land Cover Dynamics; MODIS Land Surface Temperature; plant phenology; remote sensing; treeline ecotone

Introduction

As high-elevation treelines can be considered indicators of past and recent climate change and variability [1,2], ecological niche modelling studies frequently use climate variables from numerical models to predict the current and future potential distribution of treeline species [3–5]. However, these predictions potentially disregard important local abiotic or biotic factors, which influence the actual species' distribution (i.e., the realized niche) because climate is not the exclusive factor determining habitat suitability [6]. To date, the number of studies modelling distribution ranges of deciduous treeline species' in the entire Himalayan mountains remains very limited (e.g., [7,8]). Studies which evaluate the performance of ecological niche models comparing many different predictor sets for modelling the current distribution of a treeline species covering the entire Himalayan arc do not exist. As pointed out by Bobrowski and Schickhoff [9], modelling species' distributions in high-altitude regions faces numerous challenges, most importantly the sparse data availability due to poor accessibility of the terrain. This applies in particular to species occurrence data (often obtained from herbaria) as well as to environmental predictors such as climate variables, which are often spatially interpolated and afflicted with errors [9].

Therefore, remotely sensed data can provide additional, spatially contiguous information in higher resolution and accuracy in particular in remote areas like the Himalayan mountains, giving insight into vegetation characteristics and spatial patterns, surface temperatures, and topographical features. Since the availability of multispectral satellite images in the early 1970s, biophysical mapping of the earth's surface has contributed to ecological studies [10–12]. Previous studies have documented the merit of incorporating remote sensing data in species distribution models such as land cover data [13–15] and topographical information from the Shuttle Radar Topography Mission (SRTM) (for examples, see [16]). Today, remote sensing data play an increasing role in modelling species distributions [17–20]. Compared to models using solely climatic/topographical predictors [18,21–23], the inclusion of remotely sensed variables as predictors in species distribution models improves prediction accuracy and refines the mapped distribution range of the species. The spectral measurements and their derivates are directly linked to biophysical properties of the land surface which, in turn, are linked to the primary environmental regimes and to habitat quality (productivity, vegetation structure, land cover type) [24]. Thus, the use of remotely sensed variables may also help to improve the understanding of the interactions of driving factors for complex species composition and vegetation zonation, e.g., at alpine treelines. This includes phenological traits, which represent species characteristics of recurring seasonal biological events in the life cycles [25,26]. Remote-sensing-based phenological analyses yielded notable results for modelling species' distributions [27–29].

For this study in the Himalayan mountains, characterized by distinct vertical climatic gradients and respective altitudinal zonations of vegetation, we selected the treeline-forming Himalayan Birch (*Betula utilis*) as a target species. Birch forests feature prominently within this zonation, and *B. utilis* constitutes an ideal study organism due to its status as a principal deciduous broadleaved treeline species. The distribution extends across the Himalayan arc with higher dominance in the western and central part of the mountain system. Over much of its range, *B. utilis* forms a narrow forest belt on north-facing slopes between evergreen coniferous forests (e.g., Abies spectabilis) below and an evergreen broadleaved krummholz belt (e.g., Rhododendron campanulatum) above (for more associated tree species, see [30,31]).

Bobrowski *et al.* [32] and Bobrowski and Schickhoff [9] modelled the potential distribution of *B. utilis* at a smaller spatial extent in the Himalayan region using modelled climate-related predictor variables only. In this study, we aim to bridge the gap between species' potential and actual distributions by deriving the first comprehensive ecological niche models for *B. utilis* for the entire Himalayan mountains, as well as by supplementing and substituting the standard predictors with remotely sensed land surface temperature, vegetation phenology, and topography parameters. In particular, we investigate the suitability of various predictor sets including bioclimatic variables (Chelsa [33]), topography [34], phenological traits derived from MODIS Land Cover Dynamics data [35], annual cycle parameters derived from MODIS Land Surface Temperature data [36], and their combinations. In light of the above, we (1) analyse possibilities to improve the niche model of *B. utilis* based solely on bioclimatic variables by adding different remotely sensed variables, and (2) explore the potential of a pure remote sensing approach by substituting the bioclimatic variables with remotely sensed land surface temperature data.

Materials and Methods

Study Area and Species Data

The Himalayan mountains are located between the Indian Subcontinent in the south and the Tibetan Highland in the north, extending from Afghanistan in the northwest (c. 36°N and 70°E) to Yunnan in the southeast (c. 26°N and 100° E), and covering an area of more than 1,000,000 km2. Due to a distinct three-dimensional geoecological differentiation, the Himalayas show a high variation of climate, rainfall, altitude, and soils [37,38]. The climate ranges from tropical in the Indian lowlands to permanent ice and snow at the highest elevations, and from more continental in the northwest to more oceanic in the southeast [39].

The distribution range of *B. utilis* extends across the Himalayan arc from Afghanistan to southwest China. Towards the eastern Himalayas, where more maritime climatic conditions favour the competitiveness of evergreen Rhododendron spp., *B. utilis* becomes a less frequent companion in subalpine forests and at treelines [30]. The total elevational range of *B. utilis* extends from 2700 to 4500 m [40]. In the northwest Himalayas, *B. utilis* is widely distributed between 3100 and 3700 m, while the range shifts to higher elevations towards the eastern Himalayas (mainly between 3800 and 4300 m). Pure birch stands with Rhododendron campanulatum and Sorbus microphylla in the understory and are often found at the uppermost limit of subalpine forests [41].

Presence-only occurrence data of *B. utilis* were compiled from three different sources: 215 geo-referenced records (1980–2016) were accessed via the Global Biodiversity Information Facility [42]. Further, 202 records were added from a database compiled from a literature survey ([30], unpublished data). Additionally, 827 records were extracted from freely available satellite images (GoogleEarthTM [43]) and added to the dataset. Extractions from GoogleEarth have been shown to be valuable in global treeline research [44,45]. These occurrence localities were validated through expert knowledge, obtained from numerous field visits in the Himalayan mountains.

Lowermost occurrences (e.g., in avalanche paths) were removed since they do not represent the “zonal” climatic conditions of the treeline birch belt. To reduce spatial autocorrelation, we kept only one occurrence point per grid cell (i.e., 1 km × 1 km), resulting in 1041 occurrence points for modelling the current distribution of *B. utilis* (Fig. 25).

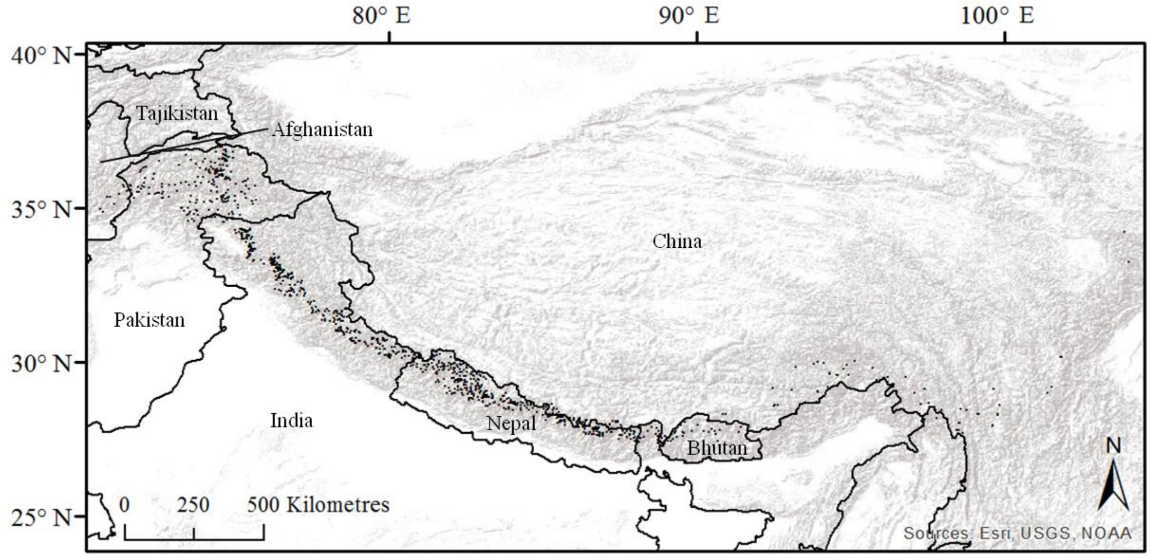


Fig. 25: Occurrences of *Betula utilis* along the Himalayan arc ($N = 1041$).

Predictor Variable Sets

The aim of the study was to evaluate the performance of four different predictor variable sets: (1) a quasi-mechanistical statistically downscaled Chelsa CLIMATE data set [33]; (2) topographical variables based on a remotely sensed Digital Elevation Model (TOPO) [34]; (3) phenological traits derived from MODIS Land Cover Dynamics data (PHENO) [35]; and (4) annual cycle parameters derived from MODIS Land Surface Temperature data (LST) [36]. Furthermore, we assessed the suitability of combinations of these predictor variable sets using an additive procedure, which resulted in 11 final models (Figure 26). First, every predictor set was used separately, followed by the combination of two and three predictor sets. Subsequently, two approaches were followed: (1) combinations of statistically downscaled variables with remotely sensed variables; and (2) combinations of exclusively remotely sensed variables.

To test the potential of surface temperature to substitute downscaled climate data, predictor sets TOPO and PHENO were combined with either CLIMATE or LST (Figure 26).

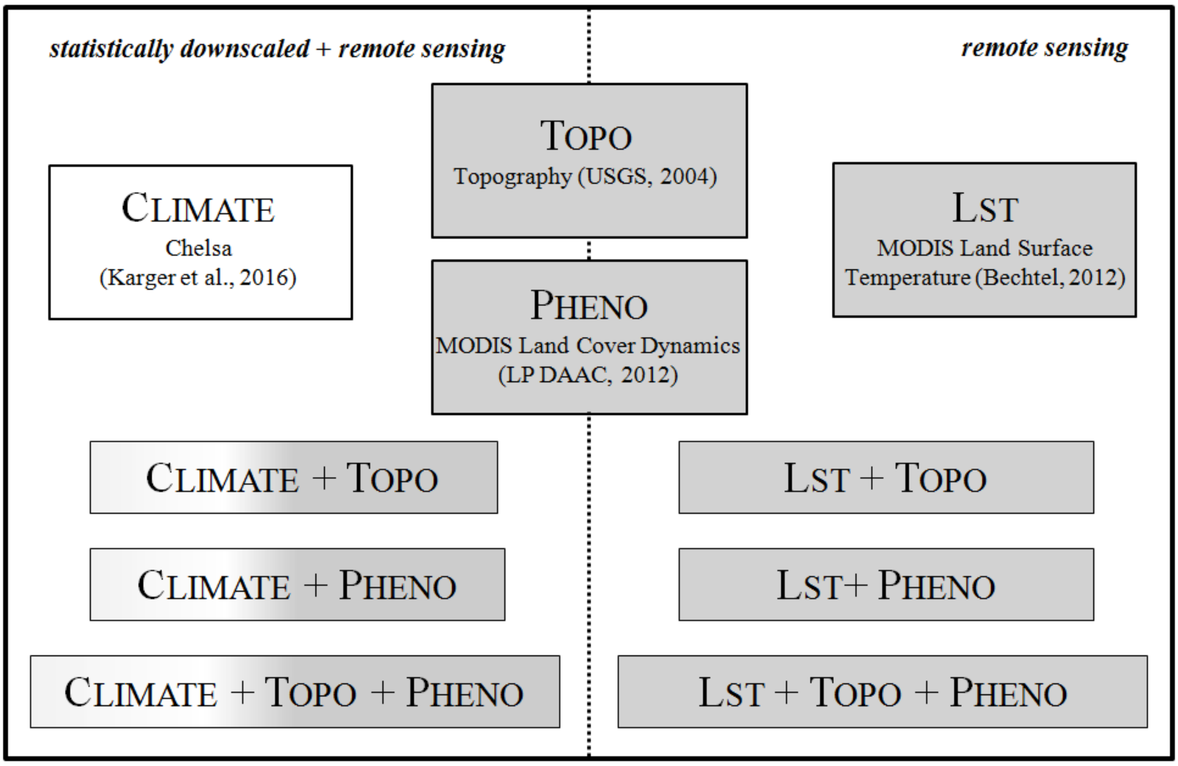


Fig. 26: Overview of the predictor sets used in the modelling procedure for estimating the ecological niche of *Betula utilis*.

All predictor variables were tested for multicollinearity using Spearman's rank correlation, since high collinearity might lead to low model performance and wrong interpretations [46]. Regarding climate variables, only ecologically relevant variables were included, which represent general patterns and annual climatic variability in the Himalayan mountains. We calculated pairwise correlations, resulting in a small set of predictor variables ($rs \leq 0.7$ according to [46]) (Table 7).

Table 7: Predictor sets with variables used for modelling the ecological niche of *Betula utilis*.

Input Set	Label	Variable	Scaling Factor	Units	Used for Modelling
CLIMATE	bio1	Annual Mean Temperature	1	Degree Celsius	
Chelsa	bio2	Mean Diurnal Range (Mean of monthly (max temp–min temp))	1	Degree Celsius	
	bio3	Isothermality (bio2/bio7)	1	Dimensionless	
	bio4	Temperature Seasonality (Stand. Dev.)	100	Degree Celsius	
	bio5	Max Temperature of Warmest Month	1	Degree Celsius	
	bio6	Min Temperature of Coldest Month	1	Degree Celsius	
	bio7	Temperature Annual Range (bio5–bio6)	1	Degree Celsius	X
	bio8	Mean Temperature of Wettest Quarter	1	Degree Celsius	X
	bio9	Mean Temperature of Driest Quarter	1	Degree Celsius	
	bio10	Mean Temperature of Warmest Quarter	1	Degree Celsius	
	bio11	Mean Temperature of Coldest Quarter	1	Degree Celsius	
	bio12	Annual Precipitation	1	Millimetre	
	bio13	Precipitation of Wettest Month	1	Millimetre	
	bio14	Precipitation of Driest Month	1	Millimetre	
	bio15	Precipitation Seasonality (Coefficient of Variation)	100	Percentage	X
	bio16	Precipitation of Wettest Quarter	1	Millimetre	
	bio17	Precipitation of Driest Quarter	1	Millimetre	
	bio18	Precipitation of Warmest Quarter	1	Millimetre	
	bio19	Precipitation of Coldest Quarter	1	Millimetre	X
	prec_may	Average Precipitation May	1	Millimetre	
	prec_mam	Average Precipitation March, April, May	1	Millimetre	X
TOPO	Alt	Altitude	1	Meters	
Topography	Northness	Northness	1	Radians	X
	Eastness	Eastness	1	Radians	
	Slope	Slope angle	1	Percentage	X
PHENO	Green_Inc	Onset Greenness Increase	1	Days	X
MODIS Land	Green_Max	Onset Greenness Maximum	1	Days	X
Cover Dynamics	Green_Dec	Onset Greenness Decrease	1	Days	X
	Green_Min	Onset Greenness Minimum	1	Days	
	EVI_Min	NBAR EVI Onset Greenness Min	0.0001	EVI value	
	EVI_Max	NBAR EVI Onset Greenness Max	0.0001	EVI value	

	EVI_Area	NBAR EVI Area	0.01	EVI area	X
	Dym_QC	Dynamics QC	1	Concatenated flags	
LST MODIS Land Surface Temperature	MAST	Mean annual land surface temperature	1	K	X
	YAST	Mean annual amplitude of land surface temperature	1	K	X
	THETA	Phase shift relative to spring equinox on the Northern hemisphere	1	days	X
	RMSE	Inter-diurnal and inter-annual variability (Root Mean Squared Error of fit)	1	K	X
	NCSA	Number of clear-sky acquisitions	1	--	X
	Max	Daytime mean maximum annual surface temperature	1	K	
	Min	Daytime mean minimum annual surface temperature	1	K	

Chelsa CLIMATE Data

Chelsa climate data (CLIMATE) are based on quasi-mechanistical statistical downscaling of the ERA-interim global circulation model with Global Precipitation Climatology Centre and Global Historical Climatology Network bias correction, and a resolution of 30 arc seconds (for details, see [33]). Precipitation amounts, estimated under consideration of orographic factors such as wind fields, valley exposure, and boundary layer height, showed high precision compared to precipitation data from other climate data sets [33].

Based on the gridded monthly fields of temperature and precipitation at a resolution of 30 arc seconds, we generated 19 bioclimatic variables (Table 7), which are widely used in species distribution modelling and represent annual characteristics (e.g., Temperature Annual Range), seasonality (e.g., Precipitation Seasonality) and extreme environmental factors (e.g., Precipitation of Driest Month) [47]. In addition, the variables Average Precipitation of May and Average Precipitation of March, April, and May were calculated in order to account for potential premonsoon drought stress [41,48,49]. The selected bioclimatic variables have proven to be suitable for modelling the potential distribution of *B. utilis* at smaller spatial scales in the Himalayan mountains [32].

Digital Elevation Model

The topographic data is based on the Digital Elevation Model obtained by the Shuttle Radar Topography Mission (SRTM) [34], which employed interferometric synthetic aperture radar, and which is considered here as a remote sensing dataset as well. The data was aggregated to a 1 km grid to calculate Slope angle and Slope aspect using SAGA GIS [50], complementing the set of potential explanatory variables. Since Slope aspect is a circular variable, it was converted into two separate continuous quantitative variables (i.e., Northness and Eastness).

MODIS Land Cover Dynamics

We chose MODIS Land Cover Dynamics (PHENO) product MCD12Q2 as it provides eight parameters that can be linked to phenological events and plants' phenology [35]. The data was obtained from the NASA Land Processes Distributed Active Archive Center [35]. We compiled time series with a spatial resolution of 500 m × 500 m from 2000 to 2013 using the 16-day composite. Long term means of the annual metrics were used to reduce the effect of the interannual variability. Four variables providing cardinal phenophase transition dates at annual time steps were selected (Onset Greenness Increase, Onset Greenness Maximum, Onset Greenness Decrease, and Onset Greenness Minimum). These dates correspond to the timing of vegetation green-up, maturity, senescence, and dormancy, respectively. Furthermore, we selected the

Enhanced Vegetation Index (EVI [51]). The first two EVI metrics (EVI Onset Greenness Min and EVI Onset Greenness Max) correspond to the EVI value of green-up and dormancy onset dates. The third metric records the sum of fitted daily EVI values during the identified vegetation cycle (i.e., Onset Greenness Increase to Onset Greenness Minimum)- EVI Area. For further information on the calculation of each PHENO predictor, see Zhang *et al.* [52,53] and Ganguly *et al.* [54].

MODIS Land Surface Temperature

MODIS Land Surface Temperature (LST) offers a unique archive of daily LST observations in 1 km resolution [55]. However, LST is a spatially and temporally highly variable quantity, and moreover affected by plenty of gaps resulting from cloud coverage. Therefore, we use long time series rather than few observations to model the long-term seasonality of LST. A model is fitted for the annual temperature cycle resulting in parameters that describe the annual temperature variation [36,56]. Here, we use a model with three parameters, which is sufficient for regions outside the tropics [57]. MAST is the mean annual land surface temperature for 2003–2014, YAST is the mean annual amplitude of the land surface temperature for 2003–2014, and THETA is the phase shift in days relative to spring equinox on the northern hemisphere. Additionally, we used NCSA, which represents the number of clear-sky acquisitions. These parameters represent a very robust estimate of LST and its annual dynamics but can also be used to estimate LST for any day of the year. The parameters were derived globally based on collection 5 of the MODIS daily level 3 global 1 km grid product from EOS Terra and Aqua (MOD11A1 and MYD11A1)—separately for night time and for day time (see [36] for details on calculation). For this study, these four variables were reprojected to the target grid and directly considered as predictor variables.

Modelling Algorithm

Generalized linear models (GLM) were applied since they represent a classical and robust approach to analyse presence and absence data [58]. Major advantages of GLMs over more complex machine-learning algorithms (e.g., random forest) include predictions which are easily interpretable and not "black box" predictions. We used the iterative weighted linear regression technique to derive maximum likelihood estimates of the response variable, with observations distributed according to exponential family and systematic effects [59]. We calculated GLMs with binomial distribution, logit-link function, and polynomial terms of second order [60], but did not include interaction terms among predictor variables. Prior to the modelling, stepwise variable selection in both directions (i.e., forward and backward) was applied using the Akaike Information Criterion (AIC) [61], resulting in the model possessing the lowest AIC value [62,63] for each predictor variable set respectively.

Pseudo-Absence Selection

As GLMs need presence and absence points, pseudo-absence points were generated. For study area selection, we used a convex hull, covering the full extent of the known occurrences of *B. utilis* distribution in the Himalayan region. By limiting the study area, large regions where the species cannot occur were excluded in further statistical analyses in order to prevent overpredicting the distribution range of the species [64]. For random selection of pseudo-absences, the limits were set as 5 km from the nearest occurrence, resulting in total 10,000 pseudo-absences (following the pseudo-absence selection procedure for GLMs described by Barbet-Massin *et al.* [65]).

Model Evaluation

For model validation, all presence and pseudo-absence points were split into training and testing data samples with a ratio of 80:20 percent using random sample splitting [66]. For each predictor variable set respectively, we repeated this procedure five times, resulting in five versions of the model and accuracy metrics, which were finally averaged. Due to the lack of a universally valid model evaluation measurement, we applied several performance evaluation metrics. Calculated evaluation measures included the AIC, the Area Under the Curve (AUC), Cohen's Kappa, and the coefficient of determination for explained variance in the data set (Pseudo-R2; [67]). We calculated the root mean square error (RMSE) in order to account for overfitting of the data (i.e., RMSE should be very similar between training and testing data sets), indicating a good fit of all models. Moreover, visual inspection on spatial patterns of the predictions was conducted since evaluation parameters may perform well in climatic space of the model, but not in geographic space (i.e., spatial prediction).

We calculated variable importance to evaluate variable contribution in the final models for each predictor variable set respectively. All statistical analyses were performed using the programming language R [68], maps were produced using ArcGIS [69].

Results

Model Evaluation and Comparison

Using an additive approach with four predictor sets resulted in 11 final models (Table 8) which showed substantial differences with regard to evaluation measurements. Regarding all different evaluation metrics and predictor variables, CLIMATE + TOPO + PHENO performed best, followed by CLIMATE + TOPO and LST + TOPO + PHENO, while the PHENO model exhibited the poorest performance in predicting the distribution of *Betula utilis*. CLIMATE + TOPO + PHENO possessed the lowest AIC with 535, followed by CLIMATE + TOPO (577) and LST + TOPO + PHENO with 594. High AUC values could be observed for all models ranging between 0.96 and 0.92, whereas the PHENO model showed a lower AUC value with 0.77. Performance ranks for Cohen’s Kappa were as follows: CLIMATE + TOPO + PHENO: 0.72, CLIMATE + TOPO: 0.66, CLIMATE + PHENO: 0.66, and LST + TOPO + PHENO: 0.66, whereas PHENO ranked last (0.01).

The explained variance was highest for the CLIMATE + TOPO + PHENO model with 69%, followed by CLIMATE + TOPO and LST + TOPO + PHENO with 65%. The model solely based on downscaled climate parameters (CLIMATE) explained 56%, while the model solely based on remotely sensed land surface temperature (LST) explained 41% of the variance in the test sets and the PHENO model 15%. Generally, CLIMATE always performed better than LST, except for explained variance for the combination with phenological traits (CLIMATE + PHENO: 0.63; LST + PHENO: 0.64). However, LST benefitted more from the addition of further predictors than CLIMATE (increase of 24% compared to 13% explained variance). In combination, LST + TOPO + PHENO performed better than CLIMATE.

Values for RMSE revealed no overfitting between the training and testing data sets for all models.

Table 8: Model evaluation results with regard to several performance measures for five averaged generalized linear model runs based on the four predictor variable sets and their combinations: Topography (TOPO); Chelsa climate data (CLIMATE); Land Cover Dynamics data (PHENO); Land Surface Temperature (LST). The results for training and test data are displayed (training 80% and testing 20%, means of 5 runs).

Model	Akaike Information Criterion	Area Under the Curve	Cohen's Kappa	Pseudo R2 Explained Variance		RMSE	
	Test	Test	Test	Train	Test	Train	Test
TOPO	778	0.92	0.41	0.48	0.48	0.26	0.26
CLIMATE	688	0.93	0.59	0.58	0.56	0.23	0.24
CLIMATE + TOPO	577	0.96	0.66	0.65	0.65	0.21	0.21
CLIMATE + PHENO	625	0.94	0.66	0.64	0.63	0.21	0.22
CLIMATE + TOPO + PHENO	535	0.96	0.72	0.70	0.69	0.19	0.19
LST	868	0.91	0.31	0.41	0.41	0.26	0.27
LST + TOPO	642	0.95	0.60	0.59	0.60	0.23	0.22
LST + PHENO	755	0.92	0.54	0.51	0.64	0.24	0.23
LST + TOPO + PHENO	594	0.96	0.66	0.64	0.65	0.21	0.21
PHENO	1148	0.77	0.01	0.18	0.15	0.30	0.31
PHENO + TOPO	722	0.93	0.51	0.55	0.54	0.24	0.24

Variable Importance

We selected 5 climatic variables, 2 topographic variables, 4 land cover metrics, and 4 land surface temperature variables out of 40 potential predictor variables.

In the following, only CLIMATE, CLIMATE + TOPO + PHENO, LST, and LST + TOPO + PHENO will be considered (Fig. 27; see Supplementary material S9, Fig. S9 for variable importance of all models). Relative variable importance varied among the four subsets of predictor variables. However, a few general characteristics became evident. Regarding climatic variables (i.e., CLIMATE), Precipitation of Coldest Quarter (bio19) was the most important predictor, followed by Temperature of Wettest Quarter (bio8, i.e., temperature of growing season), Temperature Annual Range (bio7), and Average Precipitation of March, April, and May (prec_mam), whereas Precipitation Seasonality (bio15) had lowest variable importance. Among the topographical variables, the highest importance was found for Slope. Phenological traits derived from PHENO data were always lower in variable importance, whereas temporal metrics, e.g., Onset Greenness Increase (Green_Inc), were superior to the Enhanced Vegetation Index (EVI) metric.

For the LST model, differences in variable importance were low between all variables, whereas highest variable importance was found for YAST (mean annual temperature). Differences in importance were striking between the predictor variable sets of the LST + TOPO + PHENO model. The highest value of variable importance was found for Slope, followed by YAST (mean annual temperature), NCSA (number of clear-sky acquisitions), MAST (mean annual amplitude of temperature), and THETA (phase shift in days relative to spring equinox on the northern hemisphere), whereas differences in variable importance were rather low between LST-related variables. For phenological traits, EVI showed highest variable importance.

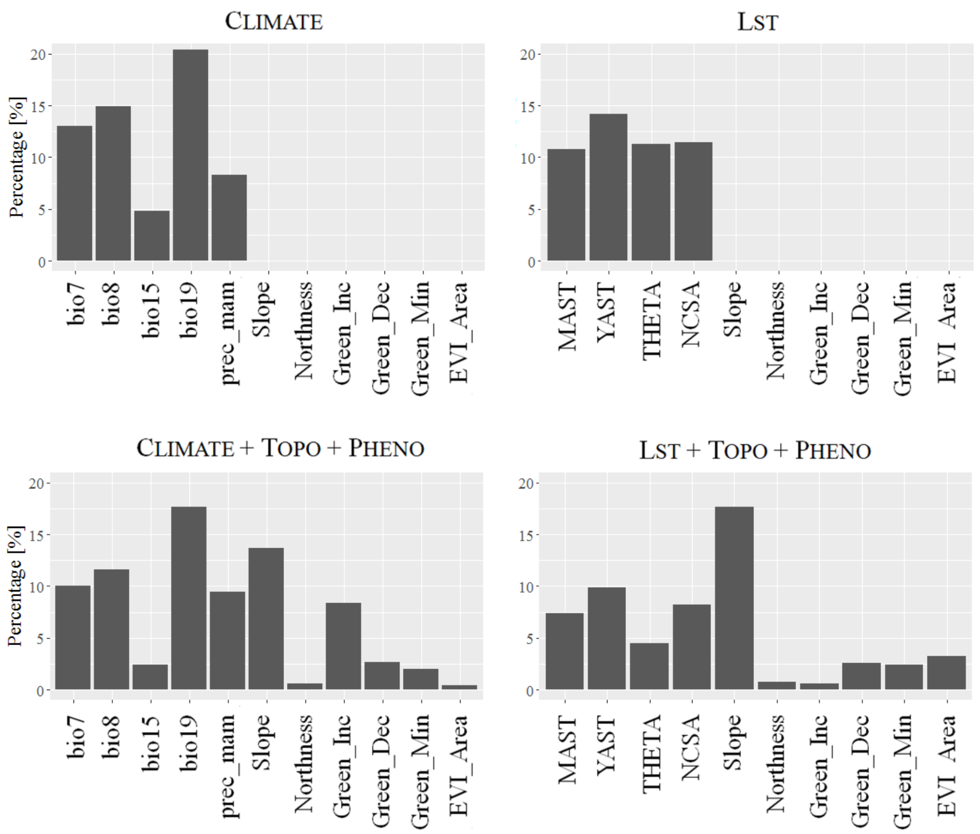


Fig. 27: Variable importance of the models using four different predictor variable sets: Chelsa climate data and Land Cover Dynamics data (Climate + Pheno); Chelsa climate data (Climate); Land Surface Temperature and Land Cover Dynamics data (Lst + Pheno); Land Cover Dynamics data (Pheno) and Land Surface Temperature (Lst) For variable description, see Table 1 and for variable importance of all Supplementary material S9, Fig. S9.

Ecological Niche Models

The continuous predictions of the models showed noticeable differences (Fig. 28). According to the model evaluation measurements, Climate + Topo + Pheno performed best, followed by Lst + Topo + Pheno, whereas Climate was superior to Lst.

It becomes apparent that the core distribution of *B. utilis* was predicted in the western part of the Himalayan mountain system, whereas only the Lst model predicted a principal distribution in the central part of the mountains. All models showed a uniform distribution along the Himalayan arc. The habitat predicted by Climate tends to be wider in range compared to the other predictions.

Displaying differences in the model predictions in detail (Fig. 29), the overall appearance (Fig. 28) is better elucidated. The model solely based on climate predictor variables (Climate, Fig. 29a) roughly met the lower limit of occurrences compared to the Climate + Topo + Pheno model (Fig. 29c), but overpredicted the uppermost limits of *B. utilis*. Overall, the prediction appears blurry and the broadleaved deciduous treeline could not be distinguished from other vegetation formations (Fig. 29a). The CLIMATE + TOPO + PHENO model differentiates between slope structures, and clearly delimits the lower distributional range of *B. utilis*. At higher altitudes, occurrence probability decreases and diminishes (Fig. 29c). Comparing CLIMATE + TOPO + PHENO models with LST + TOPO + PHENO-based models, similar patterns were observed, whereas the distributional range is predicted more constrained in the latter, leaving a smaller distribution range of *B. utilis* (Fig. 29c,d).

For the model built with remotely sensed temperature-related variables only (LST, Fig. 29b), similar patterns compared to CLIMATE (Fig. 28a) could be found. The distribution appears coarse and fuzzy, compared to CLIMATE + TOPO + PHENO and LST + TOPO + PHENO models (Fig. 29c,d).

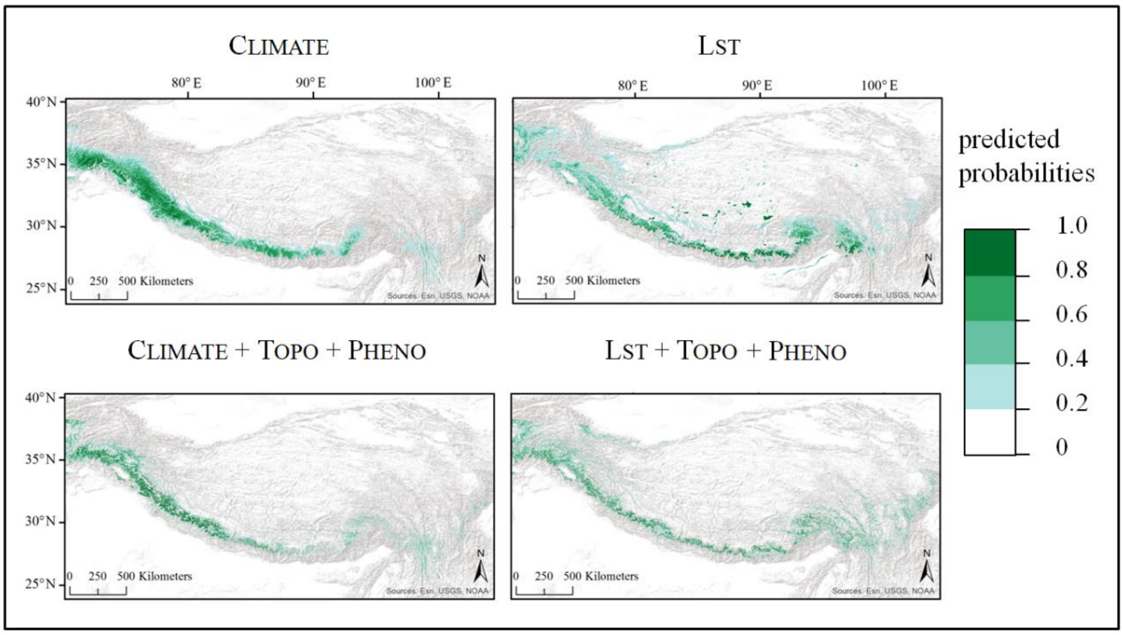


Fig. 28: Continuous predictions of the models using four different predictor variable sets: Chelsa climate data (CLIMATE); Land Surface Temperature (LST); and both combined with Topography (TOPO) and Land Cover Dynamics data (PHENO). For continuous prediction of all models, see S10, Fig. S10.

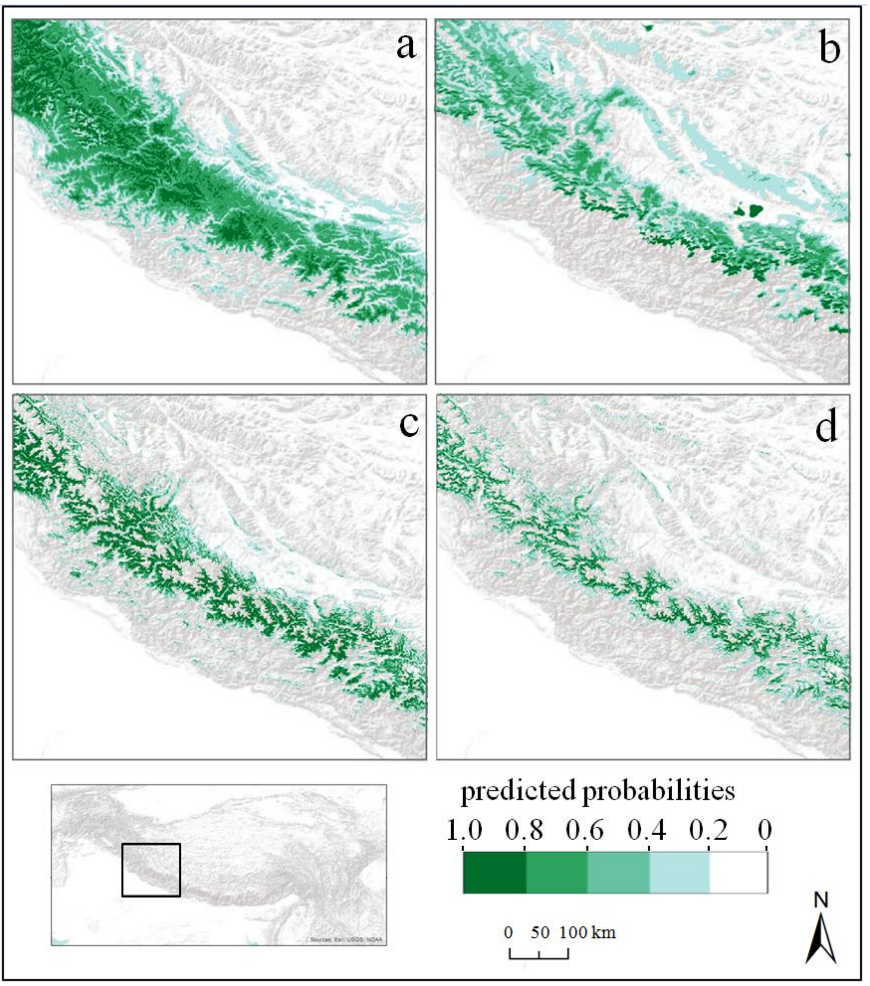


Fig. 29: Detailed excerpt of continuous predicted occurrence probability using four different predictor variable sets: Chelsa climate data (CLIMATE); (b) Land Surface Temperature (LST); (c) Chelsa climate data, Topography and Land Cover Dynamics data (CLIMATE + TOPO + PHENO); (d) Land Surface Temperature, Topography and Land Cover Dynamics data (LST + TOPO + PHENO). For detailed continuous prediction of all models, see Supplementary material S11, Fig. S11.

Discussion

Modelling ecological niches and species distributions in remote high mountain regions like the Himalayas are challenging tasks. Current studies in the field of plant species distribution modelling in the Himalayan mountains mainly use climatic variables to predict species' distribution or to forecast species range shifts under climate change scenarios (e.g., [7,32,41,70–77]). With regard to *Betula utilis*, reasonable results were obtained using solely climate for predicting the potential distribution [9,32]. However, the necessity arose for approximating the actual distribution of *B. utilis*, addressed in this study by evaluating the benefits of incorporating remotely sensed data into the modelling approach.

Modelling the Ecological Niche of Betula utilis

We found that models combining climate-related variables with remotely sensed topography and phenological traits (CLIMATE + TOPO + PHENO and LST + TOPO + PHENO) outperformed models using only one predictor variable set (CLIMATE and LST). The latter revealed no concise distinction between the vegetation formations (Fig. 28 and 29a,b), which indicates the need for further explanatory variables when modelling a species' realized niche on a large scale, i.e., the entire Himalayan mountain system.

We found that all models predicted suitable habitats for *B. utilis* as a more or less narrow line stretching along the Himalayan arc, featuring prominently in the western parts. All models predicted lower suitability in the eastern parts of the mountain system, a result which coincides with the occurrence data, and which is in line with the fact that due to the more maritime climate in the east, *B. utilis* loses its dominance as a treeline species in favour of evergreen Rhododendron species [30]. Model evaluation revealed highest performance for models built on bioclimatic variables combined with remotely sensed topography and phenology data (CLIMATE + TOPO + PHENO), emphasizing the improvement achieved by incorporating remotely sensed land cover data for predicting the distribution of *B. utilis* in the Himalayan mountains. Information on the phenological traits (Greenness Increase, and Greenness Decrease) turned out to be beneficial in refining the ecological niche of *B. utilis* (e.g., demarcation of other vegetation formations and exposed rocks).

In addition, models built solely on remotely sensed data (LST + TOPO + PHENO) suggested that phenological traits are highly relevant for modelling a more realistic distribution range of the species. The findings indicate that models built on remote sensing data yield promising predictions, which may be of interest in remote areas like the Himalayan mountains due to free availability, global coverage, and fine-scale resolution (<1 km). However, temperature variables alone (LST) are not fully capable of predicting the distribution range, as they predicted primarily suitable habitats in the central part of the mountain system and some artefacts (i.e., lakes) on the Tibetan plateau. This is to be attributed to the fact that surface temperature is, apart from atmospheric conditions, also affected by land surface characteristics. Moreover, LST only has a weak proxy for precipitation, namely the number of cloudy/cloud-free days. Consequently, the results using LST +TOPO + PHENO show lower model performance compared to the CLIMATE + TOPO + PHENO model, which incorporates precipitation-related and more plant growth specific variables. However, interestingly the LST model benefits much more from additional predictor sets (i.e., TOPO), indicating a certain redundancy between bioclimatic and topographic variables. This is plausible since topographic features are used in the downscaling process of the Chelsa data [33].

Globally, climate governs global patterns of land cover [78], but land cover and climate are not fully independent [13]. Our results confirm topoclimatic variables as the main drivers behind the distribution range, whereas phenological traits substantially contribute to narrow the modelled ecological niche of *B. utilis* (i.e., more realistic distribution). Similar results were obtained by Parra *et al.* [79] and Buermann *et al.* [17] when predicting species distribution across the Amazonian and Andean region. In a hierarchical scheme of environmental controls on species distributions, climatic variables are large-scale determinants, followed by geology, topography, and land cover, which moderate many of the effects of macroclimatic variables [13,16]. Our predictions of the CLIMATE + TOPO + PHENO model are consistent with the distribution pattern documented in several vegetation maps showing a narrow band of birch forests forming the upper treeline on north-facing slopes (e.g., [80,81]). However, the actual distribution might be smaller, since topoclimate variables and phenological traits are not the only factors determining habitat suitability. Although not considered in this study, interactions of a whole array of site factors such as ecology of tree species, site history, current biotic interactions, and anthropogenic influences affect treeline species' spatial distributions [1,30].

Ecological Interpretation of Predictor Variables

We found two temperature- and three precipitation-related variables (CLIMATE) among the most important for predicting the potential distribution of *B. utilis*. Mean Temperature of the Wettest Quarter and Temperature Annual Range were most relevant among the temperature-related variables. Growing season air and soil temperatures are considered key factors controlling tree growth at treelines and elevational position of treelines at global scales [1,2]. Much lower growing season temperatures (i.e., Mean Temperature of the Wettest Quarter) and higher average winter temperature (January) in the eastern Himalayas result in lower seasonal temperature variation which favours evergreen Rhododendron species over the deciduous species *B. utilis* at treelines. In the eastern Himalayas, *B. utilis* becomes less competitive and evergreen broadleaved species (Rhododendron spp.) are the principal treeline species [30]. By contrast, higher degrees of continentality with higher mean temperatures of warmest months and severe winter coldness lower the competitiveness of Rhododendron spp. at treeline elevations in the western and northwestern Himalayas, and contribute to the higher competitive strength of *B. utilis* on north-facing slopes.

Precipitation Seasonality, Precipitation of the Coldest Quarter, and Average Precipitation of March, April, and May were the precipitation-related variables, with Precipitation of the Coldest Quarter having the highest variable importance. Precipitation and related factors such as soil moisture and soil nutrient availability can be significant for treeline formation and dynamics [82,83]. Thus, precipitation-related variables potentially limit the climatic space of treeline tree species. The period from November to January was identified as the coldest quarter. Higher winter snowfall in the more continental western parts of the Himalayan region obviously contributes to the competitiveness of *B. utilis*. Monsoonal summer rains, on the other hand, are of less significance.

Regarding temperature-related remotely sensed variables, all variables proved to be of great importance, but mean annual temperature amplitude (YAST) revealed highest variable importance. Generally, temperature is strongly negatively correlated with altitude (except for cold air inversions in winter months), allowing little room for variation in regional-scale climate data sets [33,84]. This, in turn, leads to similar results for LST + TOPO + PHENO, compared to CLIMATE + TOPO + PHENO models, where additional information on phenological traits refined the predicted niche of *B. utilis* (Fig. 28). Interestingly, the number of cloud-free acquisitions NCSA also showed considerable variable importance, indicating that it, despite being a rather poor proxy, contains some information about precipitation.

Furthermore, MAST (mean annual temperature) and THETA (phase shift in days relative to spring equinox) were revealed to be important when modelling the ecological niche of *B. utilis*. We conclude that THETA, as it represents heat accumulation, shows sensitivity to seasonal snow cover. Vegetation analyses [30,85,86] showed that thickness and duration of snow cover providing sufficient soil moisture at the beginning and at the end of the growing season is one of the principal factors controlling the distribution of *B. utilis* forests.

The inclusion of topographic variables like Slope led to an improved prediction of climate-only models (both CLIMATE and LST). This is not surprising, since in topographically highly complex regions like the Himalayas, areas can be predicted as climatically suitable but might be inaccessible due to the slope angle. Betula forests thrive on humid, shady slopes with deeply weathered podzolic soils, and are more or less absent from south-facing slopes, in particular in the more continental W Himalaya [31,32,85].

Due to the steep gradient of hydrothermal conditions in the Himalayas, ranging from subtropical at lower to alpine conditions at higher elevations, diverse vegetation formations are formed, characterized by changing phenological traits. Our results provide compelling evidence that the climate-only models (both CLIMATE and LST) could be improved with remotely sensed phenological traits. Differences in variable importance were found between CLIMATE + TOPO + PHENO and LST + TOPO + PHENO models (Figure 3), the most important phenological traits were Onset Greenness Increase, Onset Greenness Decrease, and the Enhanced Vegetation Index (EVI). Based on these variables, a distinction between evergreen coniferous forests below the birch belt and the evergreen broadleaved krummholz belt and Rhododendron dwarf thickets above could be achieved (Fig. 29c,d).

The CLIMATE + TOPO + PHENO model revealed the importance of two temporal metrics. The timing of increasing greenness (Onset Greenness Increase) of the *B. utilis* belt differed significantly from the vegetation formations located below (Kruskal–

Wallis Test; $p \leq 0.05$). The median green-up date for *B. utilis* was at 141 Julian days after the snow melt, reflected in the predictor variable Average Precipitation March, April, and May, when sufficient soil moisture for foliation is available. The evergreen coniferous vegetation below had a much earlier green-up date (125 Julian days), whereas the evergreen krummholz belt above had a similar green-up date as the *B. utilis* belt (141 Julian days). Comparing the averaged dates of vegetation senescence (Decrease of Greenness), no significant differences could be found between the three vegetation formations (Kruskal–Wallis Test; $p \geq 0.05$). Nevertheless, the *B. utilis* belt undergoes senescence earlier (217 Julian days) than the evergreen vegetation formations (after 222 Julian days).

The LST + TOPO + PHENO model showed considerable variable importance of the EVI, which is a convenient measure of plant phenology computed from MODIS surface-reflectance data. The EVI identifies vegetation growth, maturity, and senescence, and thus marks seasonal cycles [51]. The applied EVI identifies the vegetation cycle and records sums of the onset of greenness and of minimum greenness. Hence, at the start of the growing season of evergreens, coniferous vegetation in lower altitudes is earlier, while the growing season ends up later compared to deciduous broadleaved vegetation. This results in the highest median EVI value (37) for vegetation below the treeline, whereas the EVI values for the *B. utilis* belt (24) and for vegetation formations above the treeline (17) were lower. The amount of light reflected from leaves at visible and infrared wavebands is determined by leaf traits and physiological performance [87]. The information of distinctly different phenological characteristics between the vegetation types was causal to the refinement of the niche models. Furthermore, these seasonal variations can be used to track changes in vegetation phenology [88], whereas shifts in seasonal phenological events are among the first responses at plant and ecosystem levels to climate change [89]. Shifts of flowering to earlier dates have been reported for Rhododendron species [90], and earlier green-up data resulting in an extension of the growing season [91,92] have been reported for the Himalayas. In this context, investigations of underlying climatic factors and quantification of changing plant phenological traits may provide the basis for efficient nature conservation management, expansion of protected areas, and appropriate habitat restoration strategies.

Application of Remote Sensing Data for Modelling Species' Distributions

The availability and the quality of input parameters determine model performance. Often, standardized statistically derived parameters do not fully reflect the species' physiological needs and habitat requirements. Abiotic and biotic data derived from remote sensing may open up new opportunities in analysing and modelling species' distributions, since they provide response and predictor variables. In an extensive literature review, He *et al.* [23] and references therein present countless applications of remotely sensed data for modelling species' distributions. In general, many remote sensing products are adaptable for modelling biota and can be customized in accordance with the study aims.

In the present study, we highlight the benefits of remotely sensed data in deriving tree species occurrences and predictor variables. The potential for future studies lies in the generation of presence and absence data sets, which are highly required in ENM [93]. Due to unique biophysical properties, hyperspectral sensors can detect subtle differences in reflectance based on unique plant chemistries, which is beneficial for identification of plant species [23]. Another advantage is the possibility to incorporate biotic interactions into the models, which are often disregarded due to data limitations [94,95].

Our results emphasize thermal metrics, which could be beneficially incorporated into further treeline studies in remote mountainous regions as they provide freely accessible, complete, and long-term data. Major advantages of LST-related variables include continuous observations without interpolation and geographical bias and therefore fewer uncertainties [23]. Recent studies showed how LST data could improve species modelling studies (e.g., [17,96,97]). These parameters offer numerous possibilities, such as tailored predictors in high resolution. As time series data of vegetation characteristics (i.e., phenological metrics) are becoming more and more available, changing habitat suitability can be estimated and incorporated into the model approach. In this way, knowledge can be generated, which is particularly important for modelling spatial expansion of invasive species, extinction risk assessment, and range shifts under future climate change [23].

We conclude that the current state of information may serve as a baseline for future studies, even though the available data derived from remote sensing technology is rather short term. Restrictions in the practical applicability arise from the fact that high resolution satellite imagery is still often very expensive. On the other hand, the free of cost imagery and software is already available and will become more customary in the future. Our results show that, even with freely available data, software model performance could be improved, indicating the potential for future modelling studies.

Airborne technology is a continually expanding field, and high resolution remotely sensed data will provide more insights into spatial patterns and underlying factors in future modelling studies.

Conclusions

Modelling ecological niches in remote mountain regions like the Himalayas remains a challenging task due to severe data limitations such as availability of high-quality environmental information. Given the respective climatically and topographically complex terrain, models based on climate variables alone predict the potential distribution of species only, since land cover characteristics are omitted. Nevertheless, climatic gradients determine floristic gradients in high mountain regions, and in turn, phenological traits represent niche proxies. We conclude that the addition of remotely sensed topography and phenological traits as predictor variables leads to improved model performances for the current distribution of *B. utilis*. It is a conspicuous broad-leaved deciduous tree species at the treeline in the Himalayan mountains, allowing a clear separation on the basis of phenological traits from adjacent vegetation types which consist mainly of evergreen coniferous and evergreen deciduous species in the tree layer. It becomes obvious that the inclusion of remotely sensed topography and phenological traits results in a more constrained predicted niche, compared to solely climate based models. Our results underline the relevance of phenological traits to reduce the gap in modelling studies between potential and actual distributions of species over vast, remote, and heterogeneous mountain regions. However, the actual distribution of *B. utilis* might be smaller than predicted, since topoclimate variables and phenological traits are not the only factors determining habitat suitability, and the resolution of 1 km^2 is often too coarse to account for small-scale landscape characteristics such as varying aspects.

We conclude that the approach of substituting bioclimatic variables with annual temperature cycles from remotely sensed Lst data is promising, in particular in regions where sampling efforts are low and sufficient fine-scale climate data are not available.

We further conclude that remotely sensed data make a valuable contribution when modelling the current distribution of *B. utilis*, since they provide long-term, fine-scale, and freely available observations. Our results emphasize the need for high-resolution data when modelling the actual distribution of treeline species in order to account for the heterogeneous terrain and microclimate. The incorporation of remotely sensed temperature derivates expands the classical approach (i.e., bioclimatic variables and topography) and shows great potential to derive more tailored variables (e.g., temperature of the growing season, precipitation amounts, snow cover) for ecological niche modelling. In future studies, further improvement of ecological niche models could be achieved by incorporating high-resolution remote sensing data.

Funding This study was carried out in the framework of the TREELINE project and partially funded by the German Research Foundation (DFG, SCHI 436/14-1; BO 1333/4-1) as well as the Cluster of Excellence CliSAP (EXC177) of the University of Hamburg.

Acknowledgments

We would like to express our honest gratitude to Himalayan colleagues, guides and local people who accompanied us on numerous field trips to Betula treelines. Furthermore we would like to thank the three anonymous reviewers for their diligent work and thoughtful suggestions on the earlier version of the manuscript.

References

1. Holtmeier, F.-K. Mountain Timberlines—Ecology, Patchiness and Dynamics; Advances in Global Change Research; Springer: Berlin/Heidelberg, Germany, 2009; Volume 36.
2. Körner, C. Alpine Treelines—Functional Ecology of the Global High Elevation Tree Limits; Springer: Berlin/Heidelberg, Germany, 2012.
3. Dullinger, S.; Dirnböck, T.; Grabherr, G. Modelling climate-change driven treeline shifts: Relative effects of temperature increase, dispersal and invasibility. J. Ecol. 2004, 92, 241–252, doi:10.1111/j.0022-0477.2004.00872.x.
4. Thuiller, W.; Lavorel, S.; Araujo, M.B. Niche properties and geographical extent as predictors of species sensitivity to climate change. Glob. Ecol. Biogeogr. 2005, 14, 347–357, doi:10.1111/j.1466-822X.2005.00162.x.
5. Parolo, G.; Rossi, G.; Ferrarini, A. Toward improved species niche modelling: Arnica montana in the Alps as a case study. J. Appl. Ecol. 2008, 45, 1410–1418, doi:10.1111/j.1365-2664.2008.01516.x.
6. Thuiller, W. Patterns and uncertainties of species’ range shifts under climate change. Glob. Chang. Biol. 2004, 10, 2020–2027, doi:10.1111/j.1365-2486.2004.00859.x.
7. Singh, C.P.; Panigrahy, S.; Parihar, J.S.; Dharaiya, N. Modeling environmental niche of Himalayan birch and remote sensing based vicarious validation. Trop. Ecol. 2013, 54, 321–329, ISSN 0564-3295.
8. Huo, C.; Cheng, G.; Lu, X.; Fan, J. Simulating the effects of climate change on forest dynamics on Gongga Mountain, Southwest China. J. For. Res. Jpn. 2010, 15, 176–185, doi:10.1007/s10310-009-0173-1.
9. Bobrowski, M.; Schickhoff, U. Why input matters: Selection of climate data sets for modelling the potential distribution of a treeline species in the Himalayan region. Ecol. Model. 2017, 359, 92–102, doi:10.1007/s10310-009-0173-1.
10. Tucker, C.J. Red and infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 1979, 8, 127–150, doi10.1016/0034-4257(79)90013-0.
11. Strahler, A.H. Stratification of natural vegetation for forest and rangeland inventory using Landsat digital imagery and collateral data. Int. J. Remote Sens. 1981, 2, 15–41, doi:10.1080/01431168108948338.
12. Hutchinson, C.F. Techniques for combining Landsat and ancillary data for digital classification improvement. Photogramm. Eng. Remote Sens. 1982, 48, 123–130.
13. Thuiller, W.; Araújo, M.B.; Lavorel, S. Do we need land-cover data to model species distributions in Europe? J. Biogeogr. 2004, 31, 353–361, doi:10.1046/j.0305-0270.2003.00991.x.
14. Luoto, M.; Virkkala, R.; Heikkinen, R.K. The role of land cover in bioclimatic models depends on spatial resolution. Glob. Ecol. Biogeogr. 2007, 16, 34–42, doi:10.1111/j.1466-822x.2006.00262.x.

15. Zimmermann, N.E.; Edwards, T.C.; Moisen, G.G.; Frescino, T.S.; Blackard, J.A. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. J. Appl. Ecol. 2007, 44, 1057–1067, doi:10.1111/j.1365-2664.2007.01348.x.
16. Franklin, J. Mapping Species Distributions: Spatial Inference and Prediction; Ecology, Biodiversity and Conservation; Cambridge University Press: Cambridge, UK, 2009.
17. Buermann, W.; Saatchi, S.; Smith, T.B.; Zutta, B.R.; Chaves, J.A.; Milá, B.; Graham, C.H. Predicting species distributions across the Amazonian and Andean regions using remote sensing data. J. Biogeogr. 2008, 35, 1160–1176, doi:10.1111/j.1365-2699.2007.01858.x.
18. Feilhauer, H.; He, K.S.; Rocchini, D. Modeling Species Distribution Using Niche-Based Proxies Derived from Composite Bioclimatic Variables and MODIS NDVI. Remote Sens. 2012, 4, 2057–2075, doi:10.3390/rs4072057.
19. Braunisch, V.; Patthey, P.; Arlettaz, R. Where to Combat Shrub Encroachment in Alpine Timberline Ecosystems. Combining Remotely-Sensed Vegetation Information with Species Habitat Modelling. PLoS ONE 2016, 11, e0164318, doi:10.1371/journal.pone.0164318.
20. West, A.M.; Evangelista, P.H.; Jarnevich, C.S.; Young, N.E.; Stohlgren, T.J.; Talbert, C.; Talbert, M.; Morisette, J.; Anderson, R. Integrating Remote Sensing with Species Distribution Models, Mapping Tamarisk Invasions Using the Software for Assisted Habitat Modeling (SAHM). J. Vis. Exp. 2016, 116, doi:10.3791/54578.
21. Peterson, A.T.; Nakazawa, Y. Environmental data sets matter in ecological niche modelling: An example with Solenopsis invicta and Solenopsis richteri. Glob. Ecol. Biogeogr. 2007, 17, 135–144, doi:10.1111/j.1466-8238.2007.00347.x.
22. Cord, A.F.; Klein, D.; Mora, F.; Dech, S. Comparing the suitability of classified land cover data and remote sensing variables for modeling distribution patterns of plants. Ecol. Model. 2014, 272, 129–140, doi:10.1016/j.ecolmodel.2013.09.011.
23. He, K.S.; Bradley, B.A.; Cord, A.F.; Rocchini, D.; Tuanmu, M.-N.; Schmidtlein, S.; Turner, W.; Wegmann, M.; Pettorelli, N. Will remote sensing shape the next generation of species distribution models? Remote Sens. Ecol. Conserv. 2015, 1, 4–18, doi:10.1002/rse2.7.
24. Franklin, J. Predictive vegetation mapping—Geographic modelling of biospatial patterns in relation to environmental gradients. Prog. Phys. Geogr. 1995, 19, 474–499, ISSN 0309-1333.
25. Forrest, J.; Miller-Rushing, A.J. Toward a synthetic understanding of the role of phenology in ecology and evolution. Philos. Trans. R. Soc. Lond. B Biol. Sci. 2010, 365, 3101–3112, doi:10.1098/rstb.2010.0145.
26. Polgar, C.A.; Primack, R.B. Leaf-out phenology of temperate woody plants: From trees to ecosystems. New Phytol. 2011, 191, 926–941, doi:10.1111/j.1469-8137.2011.03803.x.

27. Morisette, J.T.; Richardson, A.D.; Knapp, A.K.; Fisher, J.I.; Graham, E.A.; Abatzoglou, J.; Wilson, B.E.; Breshears, D.D.; Henebry, G.M.; Hanes, J.M.; *et al.* Tracking the rhythm of the seasons in the face of global change: Phenological research in the 21st century. Front. Ecol. Environ. 2009, 7, 253–260, doi:10.1890/070217.
28. Wilfong, B.N.; Gorchov, D.L.; Henry, M.C. Detecting an invasive shrub in deciduous forest understories using remote sensing. Weed Sci. 2009, 57, 512–520, doi:10.1614/WS-09-012.1.
29. Tuanmu, M.-N.; Viña, A.; Bearer, S.; Xu, W.; Ouyang, Z.; Zhang, H.; Liu, J. Mapping understory vegetation using phenological characteristics derived from remotely sensed data. Remote Sens. Environ. 2010, 114, 1833–1844, doi:10.1016/j.rse.2010.03.008.
30. Schickhoff, U. The upper timberline in the Himalaya, Hindu Kush and Karakorum: A review of geographical and ecological aspects. In Mountain Ecosystems. Studies in Treeline Ecology; Broll, G.B., Keplin, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 275–354.
31. Miehe, G. Landscapes of Nepal. In Nepal: An Introduction to the Natural History, Ecology and Human Environment of the Himalayas; Miehe, G., Pendry, C.A., Chaudhary, R.P., Eds.; Royal Botanic Garden Edinburgh: Edinburgh, UK, 2015; pp. 7–16.
32. Bobrowski, M.; Gerlitz, L.; Schickhoff, U. Modelling the potential distribution of *Betula utilis* in the Himalaya. Glob. Ecol. Conserv. 2017, 11, 69–83, doi:10.1016/j.gecco.2017.04.003.
33. Karger, D.N.; Conrad, O.; Böhner, J.; Kawohl, T.; Kreft, H.; Soria-Auza, R.W.; Zimmermann, N.; Linder, H.P.; Kessler, M. Climatologies at high resolution for the earth land surface areas. arXiv 2016, arXiv:1607.00217.
34. USGS. Shuttle Radar Topography Mission, 1 Arc Second Scene SRTM_u03_n008e004, Unfilled Unfinished 2.0; Global Land Cover Facility, University of Maryland: College Park, MD, USA, 2004.
35. LP DAAC. NASA Land Processes Distributed Active Archive Center, USGS/Earth Resources Observation and Science (EROS) Center, 2012. Available online: https://lpdaac.usgs.gov/data_access/data_pool (accessed onJune, 14 2017).
36. Bechtel, B. A new global climatology of annual land surface temperature. Remote Sens. 2015, 7, 2850–2870, doi:10.3390/rs70302850.
37. Troll, C. The three-dimensional zonation of the Himalayan system. In Geoecology of the High-Mountain Regions of Eurasia; Troll, C., Ed.; Erdwissenschaftliche Forschung, Franz Steiner Verlag: Wiesbaden, Germany, 1972; Volume 4, pp. 264–275.
38. Miehe, G.; Miehe, S.; Böhner, J.; Bäumler, R.; Ghimire, S.K.; Bhattarai, K.; Chaudhary, R.P.; Subedi, M.; Jha, P.K.; Pendry, C. Vegetation ecology. In Nepal: An Introduction to the Natural History, Ecology and Human Environment of the Himalayas; Miehe, G., Pendry, C.A., Chaudhary, R.P., Eds.; Royal Botanic Garden Edinburgh: Edinburgh, UK, 2015; pp. 385–472.

39. Zurick, D.; Pacheco, J. Illustrated Atlas of the Himalaya; The University Press of Kentucky: Lexington, KY, USA, 2006.
40. Polunin, O.; Stainton, A. Flowers of the Himalaya; Oxford University Press: New Delhi, India, 1984.
41. Schickhoff, U.; Bobrowski, M.; Jürgen Böhner, J.; Bürzle, B.; Chaudari, 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. Dyn. 2015, 6, 245–265, doi:10.5194/esd-6-245-2015.
42. GBIF.org: Biodiversity Occurrence Data provided by: Missouri Botanical Garden, Royal Botanic Garden Edinburgh and The Himalayan Uplands Plant Database, Accessed through GBIF Data Portal. Available online: http://www.gbif.org (accessed on January, 10 2016).
43. Google Earth, ver. 7.1.1.1888; Google LLC („Google“), Mountain View, CA, USA, 2015.
44. Paulsen, J.; Körner, C. A climate-based model to predict potential treeline position around the globe. Alp. Bot. 2014, 124, 1–12, doi:10.1007/s00035-014-0124-0.
45. Irl, S.D.H.; Anthelme, F.; Harter, D.E.V.; Jentsch, A.; Lotter, E.; Steinbauer, M.J.; Beierkuhnlein, C. Patterns of island treeline elevation—A global perspective. Ecography 2015, 38, 1–10, doi:10.1111/ecog.01266.
46. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; *et al.* Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography 2013, 36, 27–46, doi:10.1111/j.1600-0587.2012.07348.x.
47. Hijmans, R.J.; Cameron, S.E.; Parra, J.L.; Jones, P.G.; Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 2005, 25, 1965–1978, doi:10.1002/joc.1276.
48. Liang, E.; Dawadi, B.; Pederson, N.; Eckstein, D. Is the growth of birch at the upper timberline in the Himalayas limited by moisture or by temperature? Ecology 2014, 95, 2453–2465, doi:10.1890/13-1904.1.
49. Schickhoff, U.; Bobrowski, M.; Böhner, J.; Bürzle, B.; Chaudhary, R.P.; Gerlitz, L.; Lange, J.; Müller, M.; Scholten, T.; Schwab, N. Climate change and treeline dynamics in the Himalaya. In Climate Change, Glacier Response, and Vegetation Dynamics in the Himalaya; Singh, R.B., Schickhoff, U., Mal, S., Eds.; Springer: Cham, Switzerland, 2016; pp. 271–306.
50. 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, doi:10.5194/gmd-8-1991-2015.
51. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 2002, 83, 195–213, doi:10.1016/S0034-4257(02)00096-2.
52. Zhang, X.; Friedl, M.A.; Schaaf, C.B.; Strahler, A.H.; Hodges, J.C.F.; Gao, F.;

Reed, B.C.; Huete, A. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 2003, 84, 471–475, doi:10.1029/2004JD005263.

53. Zhang, X.; Friedl, M.A.; Schaaf, C.B. Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements. J. Geophys. Res. 2006, 111, doi:10.1029/2006JG000217.
54. Ganguly, D.; Rasch, P.J.; Wang, H.; Yoon, J.-H. Climate response of the South Asian monsoon system to anthropogenic aerosols. J. Geophys. Res. 2012, 117, doi:10.1029/2012JD017508.
55. Wan, Z. New refinements and validation of the MODIS Land-Surface Temperature/Emissivity products. Remote Sens. Environ. 2008, 112, 59–74, doi:10.1016/j.rse.2006.06.026.
56. Bechtel, B. Robustness of annual cycle parameters to characterize the urban thermal landscapes. IEEE Geosci. Remote Sens. Lett. 2012, 9, 876–880, doi:10.1109/LGRS.2012.2185034.
57. Bechtel, B.; Sismanidis, P. Time series analysis of moderate resolution land surface temperatures. In Remote Sensing: Time Series Image Processing; Weng, Q., Ed.; Taylor & Francis: Abingdon, UK, 2017.
58. McCullagh, P.; Nelder, J.A. Generalized Linear Models; Chapman and Hall: London, UK, 1989.
59. Nelder, J.A.; Wedderburn, R.W.M. Generalized linear models. J. R. Stat. Soc. A 1972, 135, doi:10.2307/2344614.
60. Austin, M.P. A silent clash of paradigms: Some inconsistencies in community ecology. Oikos 1999, 86, 170–178, doi:10.2307/3546582.
61. Akaike, H. A new look at the statistical model identification. IEEE Trans. Autom. Control 1974, 19, 716–723, doi:10.1109/TAC.1974.1100705.
62. Burnham, K.P.; Anderson, D.R. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed.; Springer: New York, NY, USA, 2002.
63. Guisan, A.; Edwards, T.C., Jr.; Hastie, T. Generalized linear and generalized additive models in studies of species distributions: Setting the scene. Ecol. Model. 2002, 157, 89–100, doi:10.1016/S0304-3800(02)00204-1.
64. VanderWal, J.; Shoo, L.P.; Graham, C.C.; William, S.E. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? Ecol. Model. 2009, 220, doi:10.1016/j.ecolmodel.2008.11.010.
65. 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, doi:10.1111/j.2041-210X.2011.00172.x.
66. Kuhn, M.; Johnson, K. Applied Predictive Modeling; Springer: New York, NY, USA, 2013.
67. Nagelkerke, N. A note on a general definition of the coefficient ofdetermination. Biometrika 1991, 78, 691–692, doi:10.1093/biomet/78.3.691.

68. R Core Team, version: 3.1.3, 2015, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available online: http://www.R-project.org/ (accessed on May, 01 2015).
69. ESRI. ArcGIS Desktop: Release 10.1.; Environmental Systems Research Institute: Redlands, CA, USA, 2012.
70. Menon, S.; Choudhury, B.I.; Khan; M.L.; Peterson, A.T. Ecological niche modelling and local knowledge predict new populations of Gymnocladus assamicus, a critically endangered tree species. Endanger. Species Res. 2010, 11, 175–181, doi:10.3354/esr00275.
71. Menon, S.; Khan, M.L.; Paul, A.; Peterson, A.T. Rhododendron species in the Indian eastern Himalayas: New approaches to understanding rare plant species distributions. JARS 2012, 38, 78–84.
72. Kumar, P. Assessment of impact of climate change on Rhododendrons in Sikkim Himalayas using Maxent modelling: Limitations and challenges. Biodivers. Conserv. 2012, 21, 1251–1266, doi:10.1007/s10531-012-0279-1.
73. Jaryan, V.; Datta, A.; Uniyal, S.K.; Kumar, A.; Gupta, R.C.; Singh, R.D. Modelling potential distribution of Sapium sebiferum—An invasive tree species in western Himalaya. Curr. Sci. 2013, 105, 1282–1288.
74. Gajurel, J.P.; Werth, S.; Shrestha, K.K.; Scheidegger, C. Species distribution modeling of Taxus wallichiana (Himalayan Yew) in Nepal Himalaya. Asian J. Conserv. Biol. 2014, 3, 127–134, ISSN 2278-7666.
75. Ranjitkar, S.; Kindt, R.; Sujakhu, N.M.; Hart, R.; Guo, W.; Yang, X.; Shrestha, K.K.; Xu, J.; Luedeling, E. Separation of the bioclimatic spaces of Himalayan tree rhododendron species predicted by ensemble suitability models. Glob. Ecol. Conserv. 2014, 1, 2–12, doi:10.1016/j.gecco.2014.07.001.
76. Shrestha, U.B.; Bawa, K.S. Impact of climate change on potential distribution of Chinese Caterpillar Fungus (Ophiocordyceps chinensis) in Nepal Himalaya. PLoS ONE 2014, 9, e106405, doi:10.1371/journal.pone.0106405.
77. Manish, K.; Telwala, Y.; Nautiyal, D.C.; Pandit, M.K. Modelling the impacts of future climate change on plant communities in the Himalaya: A case study from eastern Himalaya, India. MESE 2016, 2, 1–12, doi:10.1007/s40808-016-0163-1.
78. Dale, V.H. The relationship between land-use change and climate change. Ecol. Appl. 1997, 7, 753–769, doi:10.1890/1051-0761.
79. Parra, J.L.; Graham, C.C.; Freile, J.F. Evaluating alternative data sets for ecological niche models of birds in the Andes. Ecography 2004, 27, 350–360, doi:10.1111/j.0906-7590.2004.03822.x.
80. Schweinfurth, U. Die horizontale und vertikale Verbreitung der Vegetation im Himalaya. In Bonner Geographische Abhandlungen 20; Dümmlers: Bonn, Germany, 1957.
81. Schickhoff, U. Die Verbreitung der Vegetation im Kaghan-Tal (Westhimalaya, Pakistan) und ihre kartographische Darstellung im Maßstab 1:150.000. Erdkunde 1994, 48, 92–110.

82. Müller, M.; Schwab, N.; Schickhoff, U.; Böhner, J.; Scholten, T. Soil temperature and soil moisture patterns in a Himalayan alpine treeline ecotone. Arct. Antarct. Alp. Res. 2016, 48, 501–521, doi:10.1657/AAAR0016-004.
83. Müller, M.; Schickhoff, U.; Scholten, T.; Drollinger, S.; Böhner, J.; Chaudhary, R.P. How do soil properties affect alpine treelines? General principles in a global perspective and novel findings from Rolwaling Himal, Nepal. Prog. Phys. Geogr. 2016, 40, 135–160, doi:10.1177/0309133315615802.
84. Soria-Auza, R.; Kessler, M.; Bach, K.; Barajas-Barbosa, P.; Lehnert, M.; Herzog, S.; 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, doi:10.1016/j.ecolmodel.2010.01.004.
85. Schickhoff, U. Himalayan forest-cover changes in historical perspective. A case study in the Kaghan Valley, Northern Pakistan. Mt. Res. Dev. 1995, 15, 3–18, doi:10.2307/3673697.
86. Schickhoff, U. The impact of the Asian summer monsoon on forest distribution patterns, ecology and regeneration north of the main Himalayan range (E Hindukush, Karakorum). Phytocoenologia 2000, 30, 633–654, doi:10.1007/3-540-27365-4_12.
87. Price, J. How unique are spectral signatures? Remote Sens. Environ. 1994, 49, 181–186, doi:10.1016/0034-4257(94)90013-2.
88. Beck, P.S.A.; Jönsson, P.; Høgda, K.-A.; Karlsen, S.R.; Eklundh, L.; Skidmore, A.K. A ground-validated NDVI dataset for monitoring vegetation dynamics and mapping phenology in Fennoscandia and the Kola Peninsula. Int. J. Remote Sens. 2007, 28, 4311–4330, doi:10.1080/01431160701241936.
89. Badeck, F.-W.; Bondeau, A.; Böttcher, K.; Doktor, D.; Lucht, W.; Schaber, J.; Sitch, S. Responses of spring phenology to climate change. New Phytol. 2004, 162, 295–309, doi:10.1111/j.1469-8137.2004.01059.x.
90. Xu, J.; Grumbine, R.E.; Shrestha, A.; Eriksson, M.; Yang, X.; Wang, Y.; Wilkes, A. Melting Himalayas: Cascading effects of climate change on water, biodiversity, and livelihoods. Conserv. Biol. 2009, 23, 520–530, doi:10.1111/j.1523-1739.2009.01237.x.
91. Panday, P.K.; Ghimire, B. Time-series analysis of NDVI from AVHRR data over the Hindu Kush–Himalayan region for the period 1982–2006. Int. J. Remote Sens. 2012, 33, 6710–6721, doi:10.1080/01431161.2012.692836.
92. Shrestha, U.B.; Gautam, S.; Bawa, K.S. Widespread Climate Change in the Himalayas and Associated Changes in Local Ecosystems. PLoS ONE 2012, 7, e36741, doi:10.1371/journal.pone.0036741.
93. Fithian, W.; Elith, J.; Hastie, T.; Keith, D.A. Bias correction in species distribution models: Pooling survey and collection data for multiple species. Methods Ecol. Evol. 2015, 6, 424–438.

94. Kissling, W.D.; Dormann, C.F.; Groeneveld, J.; Hickler, T.; Kühn, I.; McInerny, G.J.; Montoya, J.M.; Römermann, C.; Schiffers, K.; Schurr, F.M.; *et al.* Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. J. Biogeogr. 2012, 39, 2163–2178, doi:10.1111/j.1365-2699.2011.02663.x.

95. Dormann, C.F.; Bobrowski, M.; Dehling, M.; Harris, D.J.; Hartig, F.; Lischke, H.; Moretti, M.D.; Pagel, J.; Pinkert, S.; Schleuning, M.; *et al.* Biotic interactions in species distribution modelling: Ten questions to guide interpretation and avoid false conclusions. Glob. Ecol. Biogeogr. 2018, doi:10.1111/geb.12759.

96. Bisrat, S.A.; White, M.A.; Beard, K.H.; Richard Cutler, D. Predicting the distribution potential of an invasive frog using remotely sensed data in Hawaii. Divers. Distrib. 2012, 18, 648–660, doi:10.1111/j.1472-4642.2011.00867.x.

97. Still, C.J.; Pau, S.; Edwards, E.J. Land surface skin temperature captures thermal environments of C3 and C4 grasses. Glob. Ecol. Biogeogr. 2014, 23, 286–296, doi:10.1111/geb.12121.

Supplementary Material

S1 Fig. S1: Photographs of *Betula utilis* at treelines in India and Nepal.

Fig. S1a: *Betula* treeline at 3800 m, Valley of Flowers, Uttarakhand, India (Schickhoff, 2006).

Fig. S1b: *Betula* treelines at 3900 m, Manag, Annapurna, Nepal (Schickhoff, 2013).

Fig. S1c: *Betula* treeline at 3950 m, Langtang, Rasuwa, Nepal (Schickhoff, 2010).

S2 Fig. S2: Correlation matrix of predictor variables

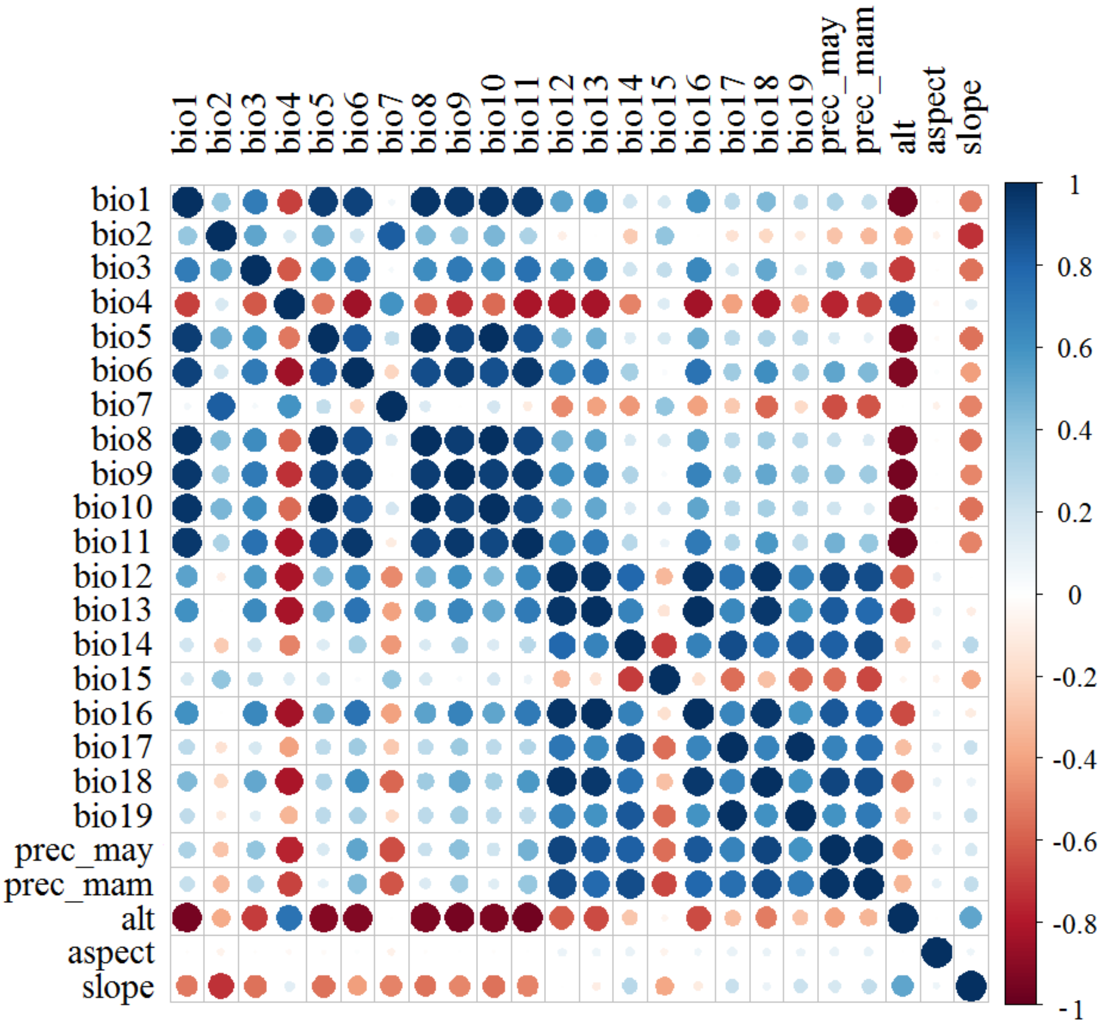


Fig. S2: Correlation matrix of predictor variables, measured by Spearman's rank correlation coefficient (r_S) ranging from -1 to 1. Size of circle displays strength of correlation coefficient. For details on abbreviations of predictor variables see Table 2.

S3 Table S3: Descriptive statistics for presences and pseudo-absences of *Betula utilis*.

For details on abbreviations of predictor variables see Table 2.

S3 Table S3a: Descriptive statistics for presences of *B. utilis* ($N_{\text{presences}}$= 590)

	Median	Std.dev	Minimum	Maximum
bio1	2.75	2.25	-3.14	10.65
bio2	6.65	0.61	5.23	8.08
bio3	27.79	1.56	24.59	34.15
bio4	649.49	81.35	448.23	829.83
bio5	12.43	2.78	5.75	21.08
bio6	-11.12	2.38	-17.70	-4.29
bio7	23.75	2.73	17.57	30.19
bio8	10.42	2.44	4.08	18.22
bio9	-2.69	2.78	-9.15	6.35
bio10	10.42	2.54	4.08	18.76
bio11	-5.14	2.33	-11.68	2.13
bio12	1339.18	603.10	371.42	3652.11
bio13	320.28	156.36	83.41	829.22
bio14	11.54	4.07	2.27	47.52
bio15	96.96	11.99	48.35	119.05
bio16	793.78	390.15	213.91	2066.47
bio17	56.79	24.14	9.75	228.51
bio18	790.78	388.58	213.91	2066.47
bio19	84.71	60.81	11.96	336.38
prec_may	91.00	55.70	24.39	405.46
prec_mam	69.16	34.12	19.39	298.27
alt	3970.00	327.85	2644.00	6077.00
aspect	218.39	115.91	0.00	359.34
slope	90.00	0.00	89.97	90.00

S3 Table S3b: Descriptive statistics for pseudo-absences of *B. utilis* ($N_{\text{pseudo-absences}}$= 6000)

	Median	Std.dev	Minimum	Maximum
bio1	4.06	13.20	-17.70	26.43
bio2	8.17	2.43	3.21	14.59
bio3	27.94	5.30	21.10	42.90
bio4	688.05	154.09	359.48	1072.85
bio5	15.47	13.61	-6.21	43.57
bio6	-10.19	13.39	-33.41	12.87
bio7	28.67	5.23	14.19	38.05
bio8	12.26	10.72	-16.01	31.69
bio9	-3.03	14.29	-24.66	29.66
bio10	12.37	11.40	-8.10	34.59
bio11	-4.17	14.09	-26.27	19.04
bio12	919.48	784.83	53.39	5979.20
bio13	234.18	173.14	13.82	1004.33
bio14	5.12	7.11	0.12	85.81
bio15	114.84	22.55	31.67	154.83
bio16	597.20	440.60	33.84	2597.38
bio17	28.17	33.76	0.76	402.40
bio18	431.04	410.37	32.86	2597.38
bio19	39.36	55.90	0.77	534.17
prec_may	36.27	96.97	1.63	711.22
prec_mam	24.79	62.66	1.59	547.33
alt	4071.00	2279.39	20.00	6816.00
aspect	176.89	98.58	-1.00	359.93
slope	89.99	5.69	0.00	90.00

S4 Table S4: Confusion matrix of model predictions averaged over 5 runs for *Betula utilis* in the Himalayan mountains (N_{Test}= 1317), to calculate the True Skill Statistics (TSS = 0.89), thresholded at 0.088.

		observed	
		1	0
predicted	1	107	71
	0	5	1134

S5 Fig. S5: Continuous predictions of the current distribution of *Betula utilis.*

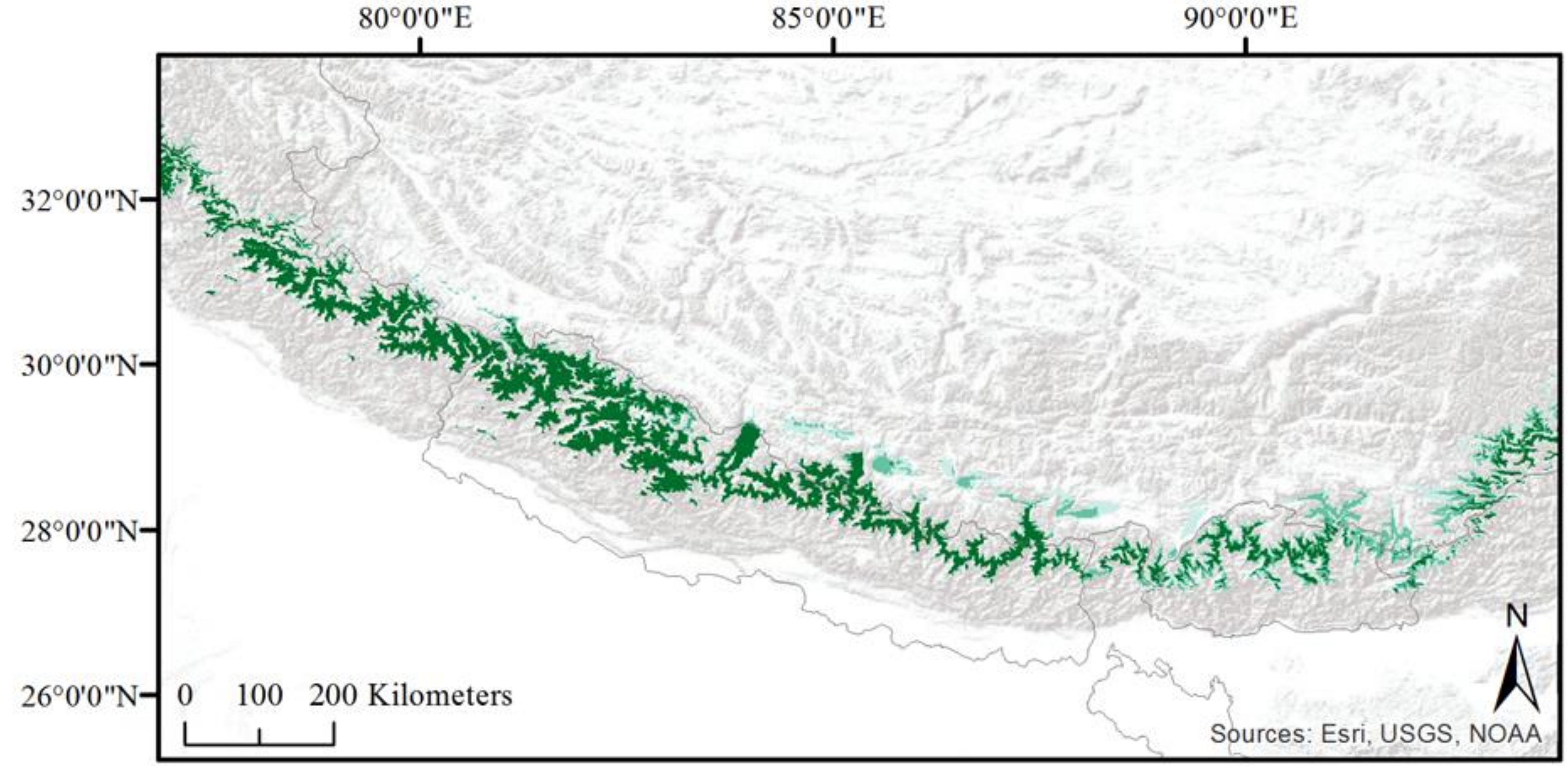


Fig. S5: Continuous predictions of the current distribution of *Betula utilis* predicted by all 5 model runs (larger in size).

S6 Fig. S6: Geographic extent of regions predicted as the potential distribution of *Betula utilis*

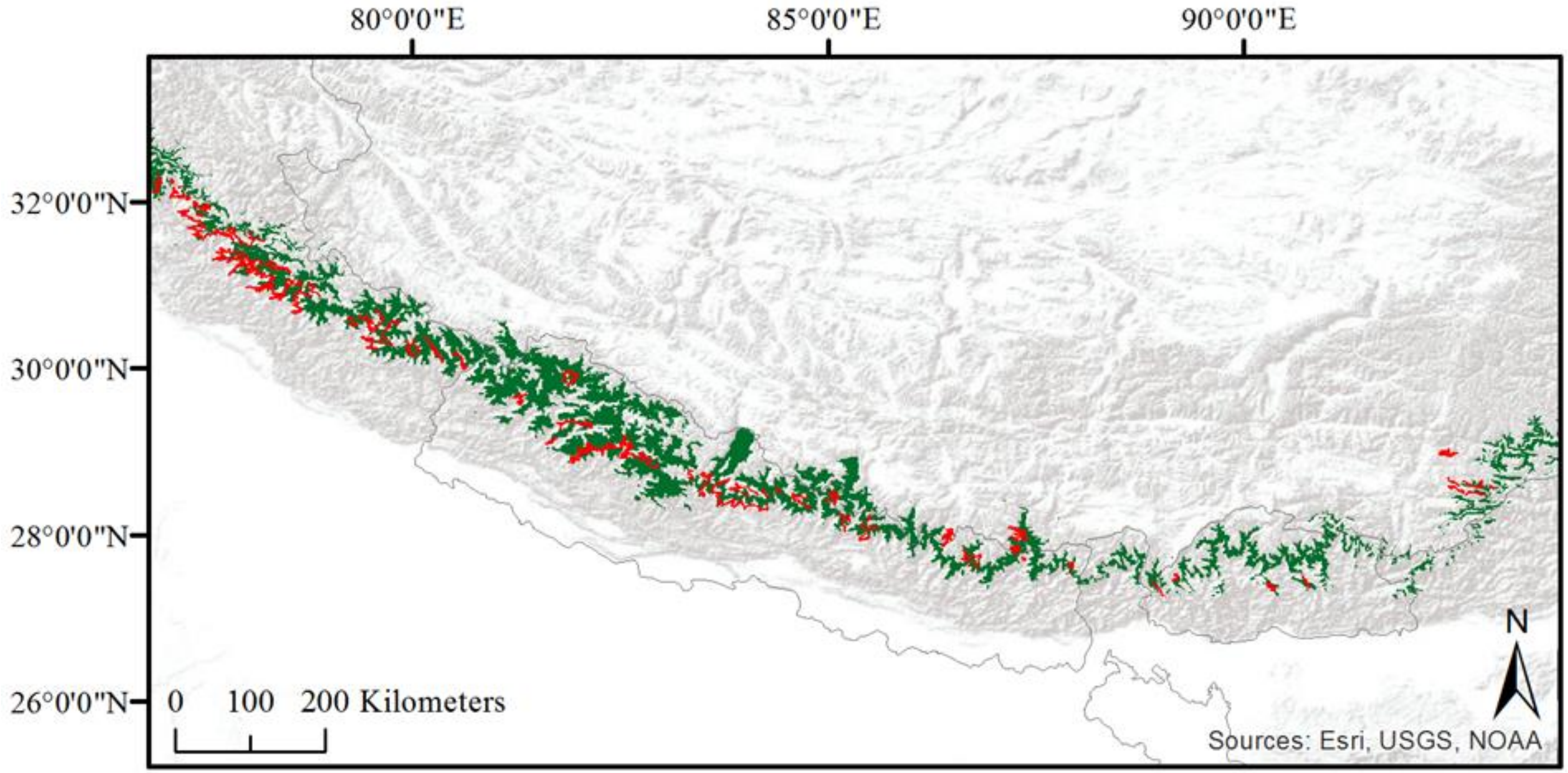


Fig. S6: Geographic extent of regions predicted as the potential distribution of *Betula utilis* (green) thresholded at True Skill Statistics (TSS ≥ 0.89) and distribution of *B. utilis* according to the Schweinfurth vegetation map (red) (Schweinfurth, 1957) (larger in size).

S7 Fig. S7: Maps of disagreement between CHELSA and WORLDCLIM.

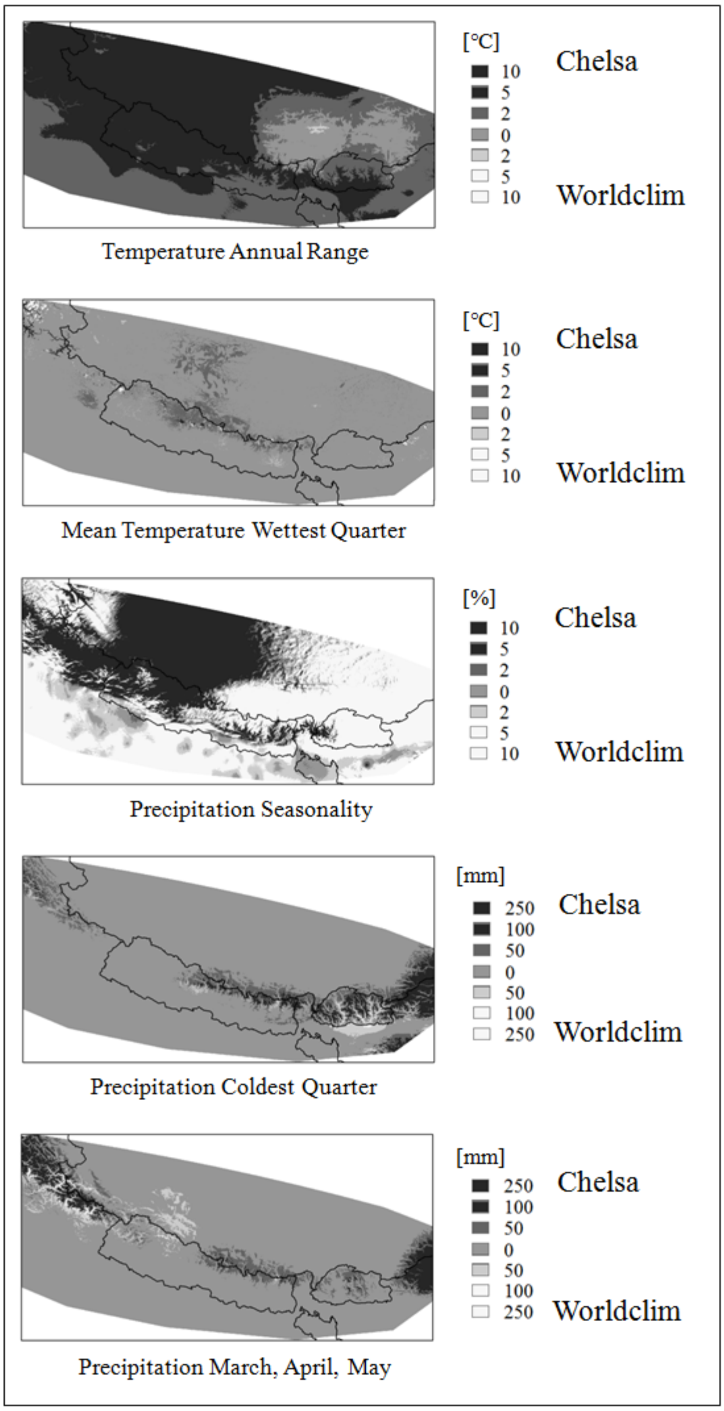


Fig. S7: Maps of disagreement between CHELSA and WORLDCLIM for each variable used in the modelling approach. Black colours represent higher values for CHELSA compared to WORLDCLIM, and white colours higher values in the WORLDCLIM data set. Grey implies no difference between the climate data sets.

S8 Fig. S8: Intersection of binary projections by both climate data sets.

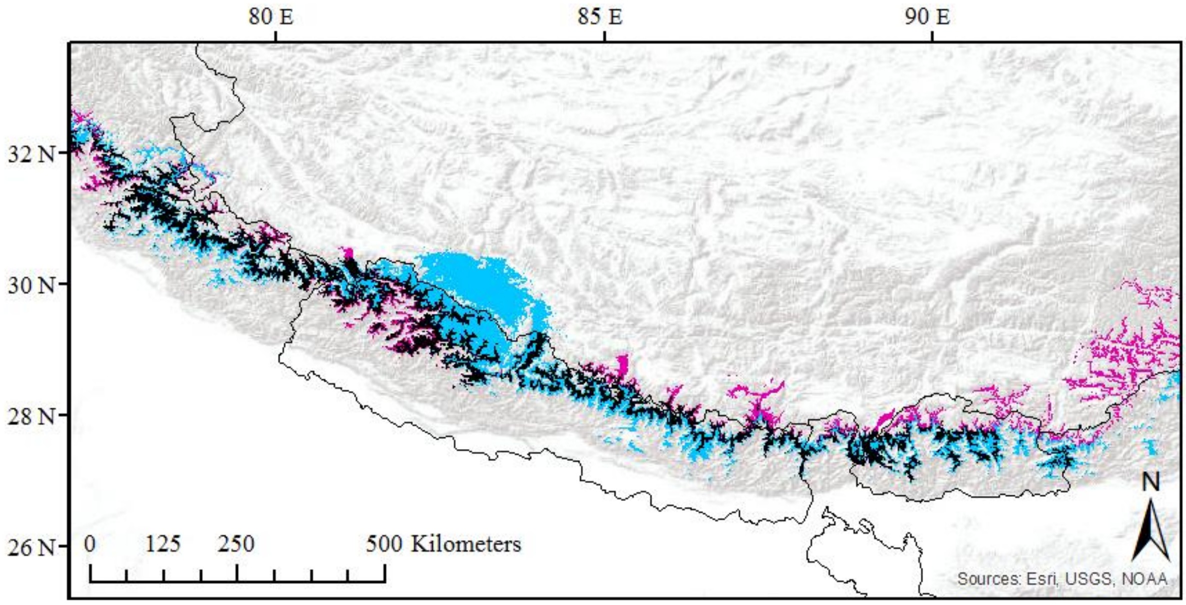
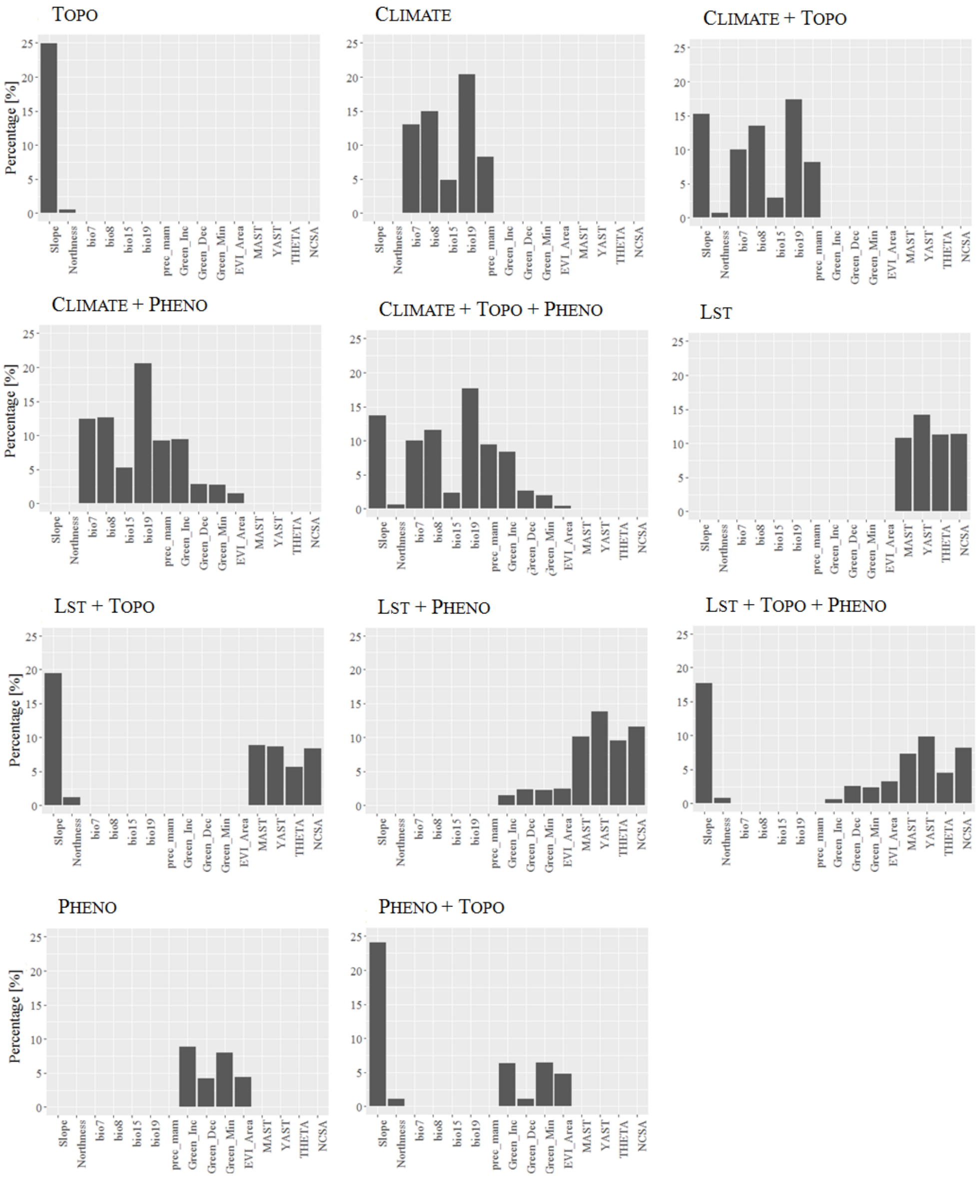


Fig. S8: Intersection of binary projections by both climate data sets (areas predicted by CHELSA climate data = purple; areas predicted by WORLDCLIM climate data= blue; areas predicted by both climate data sets= black).

S9 Fig. S9: Variable importance for all models based on four predictor variable sets and their combinations.



TOPO

CLIMATE

CLIMATE + TOPO

CLIMATE + PHENO

CLIMATE + TOPO + PHENO

LST

LST + TOPO

LST + PHENO

LST + TOPO + PHENO

PHENO

PHENO + TOPO

Fig. S9: Variable importance for all models based on four predictor variable sets and their combinations. Topography (TOPO); Chelsa climate data (CLIMATE); Chelsa climate data and topography (CLIMATE + TOPO); Chelsa climate data and Land Cover Dynamics data (CLIMATE + PHENO); Chelsa climate data, topography and Land Cover Dynamics data (CLIMATE + TOPO + PHENO); Land Surface Temperature (LST); Land Surface Temperature and topography (LST + TOPO); Land Surface Temperature and Land Cover Dynamincs data (LST + PHENO); Land Surface Temperature, topography and Land Cover Dynamics data (LST + TOPO + PHENO); Land Cover Dynamics data (PHENO) and Land Cover Dynamics data and topography (PHENO + TOPO) The results for training and test data are displayed (training 80% and testing 20%).

S10 Fig. S10: Continuous predictions of all models based on four predictor variable sets and their combinations.

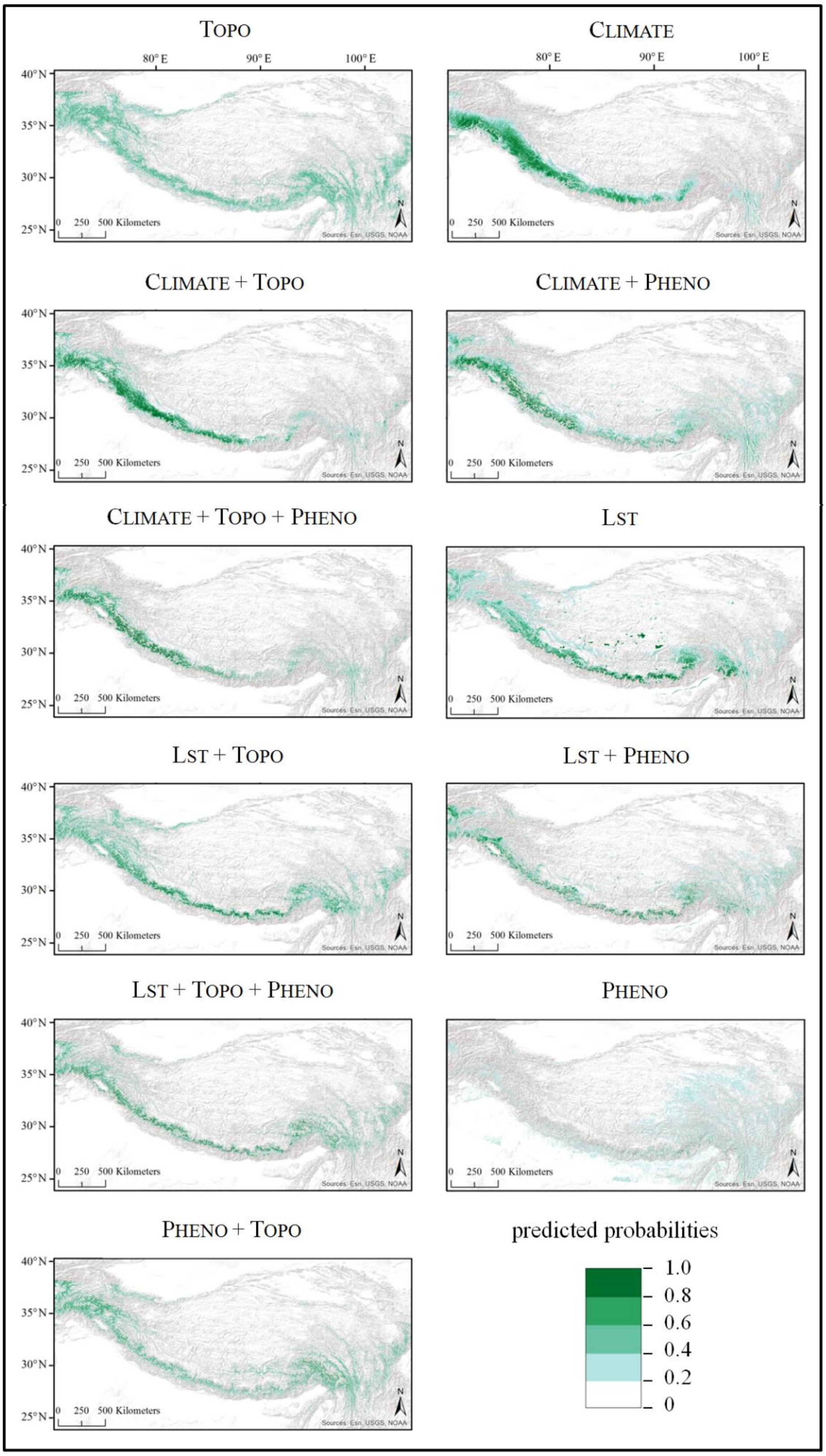


Fig. S10: Continuous predictions of all models based on four predictor variable sets and their combinations. Topography (TOPO); Chelsa climate data (CLIMATE); Chelsa climate data and topography (CLIMATE + TOPO); Chelsa climate data and Land Cover Dynamics data (CLIMATE + PHENO); Chelsa climate data, topography and Land Cover Dynamics data (CLIMATE + TOPO + PHENO); Land Surface Temperature (LST); Land Surface Temperature and topography (LST + TOPO); Land Surface Temperature and Land Cover Dynamincs data (LST + PHENO); Land Surface Temperature, topography and Land Cover Dynamics data (LST + TOPO + PHENO); Land Cover Dynamics data (PHENO) and Land Cover Dynamics data and topography (PHENO + TOPO) The results for training and test data are displayed (training 80% and testing 20%).

S11 Fig. S11: Detailed excerpt of continuous predictions of all models based on four predictor variable sets and their combinations.

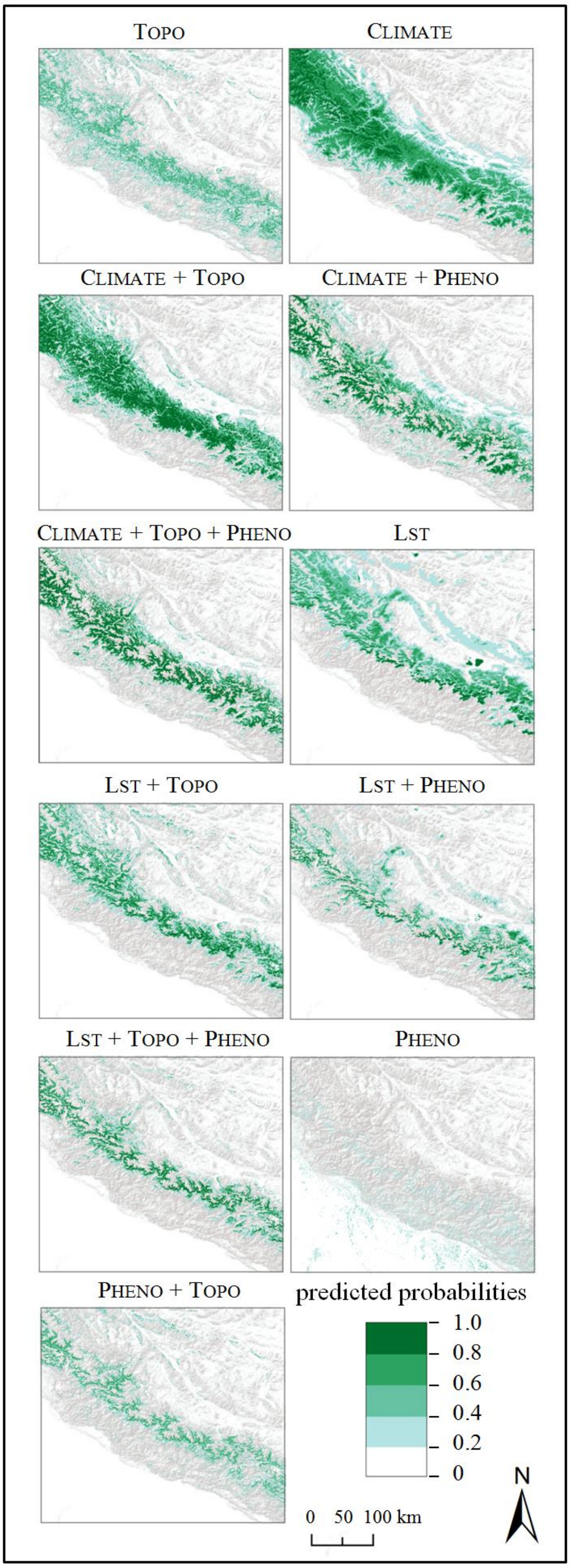


Fig. S11: Detailed excerpt of continuous predictions of all models based on four predictor variable sets and their combinations. Topography (TOPO); Chelsa climate data (CLIMATE); Chelsa climate data and topography (CLIMATE + TOPO); Chelsa climate data and Land Cover Dynamics data (CLIMATE + PHENO); Chelsa climate data, topography and Land Cover Dynamics data (CLIMATE + TOPO + PHENO); Land Surface Temperature (LST); Land Surface Temperature and topography (LST + TOPO); Land Surface Temperature and Land Cover Dynamincs data (LST + PHENO); Land Surface Temperature, topography and Land Cover Dynamics data (LST + TOPO + PHENO); Land Cover Dynamics data (PHENO) and Land Cover Dynamics data and topography (PHENO + TOPO) The results for training and test data are displayed (training 80% and testing 20%).

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, im Juni ____