Synthesis

What we use is not what we know: environmental predictors in plant distribution models

Heidi K. Mod, Daniel Scherrer, Miska Luoto & Antoine Guisan

Mod, H.K. (corresponding author, heidi.mod@helsinki.fi)[a, b, †]

Scherrer, D. (daniel.scherrer@unil.ch)[b, †]

Luoto, M. (miska.luoto@helsinki.fi)[a, ‡]

Guisan, A. (antoine.guisan@unil.ch)[b, c, ‡]

[a] Department of Geosciences and Geography, University of Helsinki, P.O. Box 64 (Gustaf Hällstöminkatu 2a), FI-00014 University of Helsinki, Finland

[b] Department of Ecology and Evolution, University of Lausanne, Biophore, CH-1015 Lausanne, Switzerland

[c] Institute of Earth Sciences, University of Lausanne, Géopolis, CH-1015 Lausanne, Switzerland

[†] Shared first authorship.

[‡] Shared last authorship.

Printed journal page estimate: 9837 words (12.3 pages), 1 table 0.5 pages, 5 figures 2 pages, total 15 pages

ABSTRACT

Questions: The choice of environmental predictor variables in correlative models of plant species distributions (hereafter SDMs) is crucial to ensure predictive accuracy and model realism, as highlighted in multiple earlier studies. Because variable selection is directly related to a model's capacity to capture important species' environmental requirements, one would expect an explicit prior consideration of all ecophysiologically meaningful variables. For plants, these include temperature, water, soil nutrients, light, and in some cases, disturbances and biotic interactions. However, the set of predictors used in published correlative plant SDM studies varies considerably. No comprehensive review exists of what environmental predictors are meaningful, available (or missing), and used in practice to predict plant distributions. Contributing to answer these questions is the aim of this review.

Methods: We carried out an extensive, systematic review of recently published plant SDM studies (years 2010-2015; *n* = 200) to determine the predictors used (and not used) in the models. We additionally conducted an in-depth review of SDM studies in selected journals to identify temporal trends in the use of predictors (years 2000-2015; *n* = 40).

Results: A large majority of plant SDM studies neglected several ecophysiologically meaningful environmental variables, and the number of relevant predictors used in models has stagnated or even declined over the last 15 years.

Conclusions: Neglecting ecophysiologically meaningful predictors can result in incomplete niche quantification and can thus limit the predictive power of plant SDMs. Some of these missing predictors are already available spatially or may soon become available (e.g., soil moisture). However, others are not yet easily obtainable across whole study extents (e.g., soil pH and nutrients), and their development should receive increased attention. We conclude that more effort should be made to build ecologically more sound plant SDMs. This requires a more thorough rationale for the choice of environmental predictors needed to meet the study goal, and the

development of missing ones. The latter calls for increased collaborative effort between ecological and geo-environmental sciences.

Keywords: covariate; environment; habitat suitability; independent variable; model; niche; plant; predictor; species distribution;

Abbreviations: DEM = digital elevation model, GIS = geographic information system, SDM = correlative species distribution modelling, WoS = ISI Web of Science

Running title: Variable selection and species distribution models

INTRODUCTION

Correlative species distribution modelling (SDM; also called ecological niche, habitat suitability, and (bio)climatic envelope modelling as well as various other names, hereafter all included under the acronym 'SDM'; see Guisan et al. 2013) is a topical approach in ecology and biogeography (Franklin 2009, Peterson et al. 2011, Moquet et al. 2015). Over the last decades (Booth et al. 2014), the number of correlative SDM studies has steadily increased, and SDM is currently one of the most popular methods used to study the impact of various threats to biodiversity and to support related conservation decisions (Guisan et al. 2013). In addition to a large number of case studies on species distributions for conservation and risk assessment (Broennimann & Guisan 2008; Araújo et al. 2011; Jiménez-Valverde et al. 2011; Alagador et al. 2014), there is on-going discussion on theoretical and technical issues, including modelling techniques, selection and evaluation of models, handling of spatial autocorrelation and, most importantly, variable selection (Franklin 1995; Austin

2002, 2007; Guisan & Thuiller 2005; Araujo & Guisan 2006; Guisan et al. 2006, Dormann 2007; Elith & Leathwick 2009; Zimmermann et al. 2010; Austin & Van Niel 2011a; Thibaud et al. 2014). As SDMs statistically relate environmental variables to the presence/absence (or presence-only) of a species to predict species distributions (Guisan & Zimmermann 2000), the selection of the most appropriate set of environmental variables as predictors is essential (Dormann 2007).

Many of the SDM (sensu lato) reviews published within the last 20 years have called for the use of more ecologically meaningful predictors (Franklin 1995, 2009; Guisan & Zimmermann 2000; Guisan & Thuiller 2005; Guisan et al. 2006, Elith & Leathwick 2009; Austin & Van Niel 2011a, Peterson et al. 2011). For plants, seven environmental factors are generally considered essential for growth and survival: temperature, water, nutrients, light, disturbances, biotic interactions and CO_2 (Körner 2014, see also Guisan & Zimmermann 2000; Austin & van Niel 2011a and Appendix S1). However, although CO_2 is crucial for plant survival and productivity, it is not a limiting resource under natural growth conditions at current and future atmospheric concentrations (e.g., Körner 2006; Norby & Zak 2011; Inauen et al. 2012; Bader et al. 2013). Under such conditions, the nutrient cycle and climatic constraints control carbon capture, and therefore CO_2 is usually omitted in correlative analyses of species distributions, such as SDMs, and will not be considered further in this review. All of the other factors can be resources (i.e., can be consumed by the species; e.g. nutrients, water, light) or regulators (i.e., can affect metabolic processes; e.g. temperature; see Huston 2002) and can have direct (proximal) and indirect (distal) effects on plants (Austin 2002). Thus, in standard SDMs, where species occurrence (and absence) is modelled principally as a function of environmental conditions, the most realistic and accurate predictions should only be achieved when all factors defining a species' niche and thus driving its distribution are accounted for at the geographic scale considered (Pearson & Dawson 2003; McGill 2010). When considering the environmental factors shaping species distribution from a niche modelling perspective, it is also important to distinguish between bionomic (dynamically altered by the species through being consumed or modified) and scenopoetic (constant, not affected by the species) variables (see Hutchinson 1978; Peterson et al. 2011). In this review, by considering the environmental niche (Grinnell 1917; Hutchinson 1957) of plants (Austin 1980; Austin & Smith 1989) in a wide sense, we include both regulator and resource predictors, but because precise data on the dynamics of environmental variables are scarce, we consider resources to remain constant (i.e. we do not consider what could be consumed by the species itself) over the location and time period of the study.

In addition to the importance of ecological justification for the use of ecophysiologically relevant variables in SDMs, Austin (2002) and later Araujo & Guisan (2006) highlighted the importance of acknowledging the biological significance of the selected variables, despite the diverse automated and mathematically optimized variable selection methods developed for SDMs. Additionally, Petitpierre et al. (in review) showed that selecting variables based on expert knowledge rather than an automated selection from huge numbers of predictors can lead to better predictive performances and be more reflective of biological and ecological understanding, especially for fine-scale studies (see also Pearson & Dawson 2003 for the hypothesized higher importance of non-climatic variables at finer scales; but see Harwood et al. 2014).

Although ecophysiological theory (Lambers et al. 2008; Körner 2014), community assembly experiments (Fukami et al. 2005; Scherber et al. 2010) and biogeographical models (e.g. Franklin 1995; Bertrand et al. 2012; Dubuis et al. 2013; Wisz et al. 2013) stress the importance of various groups of ecophysiologically essential predictors (Fig. 1), it seems that a large majority of SDMs are built without consideration of the ecophysiological relevance and comprehensiveness of the set of predictors (Pearson & Dawson 2003; Guisan & Thuiller 2005; Austin & Van Niel 2011a). The most prominent explanation for this incomplete choice of predictors is the unavailability of some data. It seems that largely available variables are frequently used in models (e.g., WorldClim; Hijmans et al. 2005), while the use of less easily available or lacking environmental data is understandably less frequent or absent in SDMs, respectively. This is however a working hypothesis. Making further progress in SDM science therefore requires understanding the primary causes of incomplete use of environmental information. Species distribution models are potentially powerful tools to analyse and predict plant species and community distributions, but their strength, validity and accuracy depend largely on the input data used. Yet, despite a long-standing knowledge of which predictors should theoretically be used, no study has comprehensively reviewed which ecophysiologically meaningful variables are currently used and not used or missing, so that recommendations can be made on where further development is required to obtain all important predictors in a spatially explicit form.

Here, we evaluate whether the predictors used in correlative plant SDM studies correspond to the known ecophysiological needs of plant species and whether additional constraints, such as biotic factors and disturbances, are included. Simultaneously, we aim to identify which of the ecophysiologically relevant variables are missing and whether their omission is due to the unavailability of data in a mapped format or to other causes. We do not either intend to review exhaustively the literature to exemplify good from bad modelling practices, nor to provide examples from our own analyses. We concentrate on niche-based species distribution models of plants (vascular plants and bryophytes) and mainly consider direct abiotic variables – both regulator and resource (sensu Austin 1980) – as well as biotic and disturbance variables. Plants form the basis of primary production and the food chain and, as such, are important for other species, biodiversity and environmental conservation in general. Focusing solely on plants also allows for a more in-depth review. We acknowledge the importance of other, non-niche processes influencing plant distributions, such as dispersal and (evolutionary) history (Soberón & Peterson 2005), but we do not examine these processes explicitly here, as we consider them to be outside the scope of this review, which centres on environmental niche predictors. Further, although efforts towards incorporating the environmental predictors discussed here are also in progress in the field of mechanistic modelling (see, e.g., D'Amen et al. in press), this review only considers correlative SDMs.

MATERIALS AND METHODS

We performed two web searches to extract original articles (excluding reviews, opinions and perspectives) dealing with SDMs of vascular plants and bryophytes. The target of the first search was to record recently published (2010-2015) articles in high-quality ecological journals (see Appendix S2 for the journals used), while the target of the second search was to examine the temporal changes in the variables used in the SDMs. The first search was performed using the query (*"species distribution model*" OR "habitat model*" OR "ecological niche model" OR "niche model*" OR "habitat distribution model*" OR "habitat suitability model*" OR "niche-based model*" OR "bioclimatic envelope model*") AND (vegetation OR plant* OR vascular OR bryophyte*)* following Guisan et al. (2013) in the ISI Web of Science (WoS), restricting the time range and journals to meet the filters specified above. This search resulted in 745 papers (hereafter called the 'recent search'). The second WoS search used the same search words, but the results were limited to two journals, *Journal of Vegetation Science* and *Journal of Biogeography*, after preliminary queries showed the high number of plant SDM studies published in these journals, accounting for the years 2000-2015. The second search was also repeated in other search engines to increase the number of articles and to complement missing years, resulting in a total of 171 articles (hereafter called the 'temporal search').

For all of the selected articles, we recorded the environmental predictors that were used in the SDMs. To standardize the results, we divided the predictors into eight variable categories, partially following Austin and Van Niel (2011a, see also Appendix S1): temperature, water, substrate (including nutrients but not moisture), radiation, biotic interactions, disturbance (including anthropogenic factors), topography and land use (Table 1, see detailed list of different variables in Appendix S3). The temperature and water categories were further divided into mean, extreme and seasonality variables, and the water category had two additional sub-classes: water balance and soil moisture. The substrate-related category was divided into two classes: bedrock/pH and nutrients. The category of biotic variables accounted for all variables expressing the influence of other biological agents (e.g., cover of vegetation or certain plant species, species richness, and presence or abundance of animal species). The disturbance category accounted for processes that primarily destroy vegetation, such as fire, geomorphological disturbance and human activities, although these processes can also have a positive impact on certain species (e.g., ruderals; Grime 1977). Topographic and land-use related variables do not represent direct or resource variables for plants, but because these are regularly included in SDMs (Franklin 1995) and have an indirect impact on plant distribution through altering the distribution of temperature, moisture, nutrients and light, they were also recorded here (Moeslund et al. 2013). All generally ecophysiologically meaningful predictor variables could be assigned to 16 classes (Table 1). Predictors that were meaningful for the target of the original study but not for our review (such as fragmentation and distance to waterbodies) were not recorded but are included in the total number of predictors.

From each selected SDM study, we further recorded the taxonomic group of species of interest and the resolution of the input/environmental data. Only studies that used species distribution data (presence-absence or presence-only) were included in further analyses, i.e. studies on species richness or abundance were not considered. To avoid bias in our analyses due to the tendency to highlight the use of climate variables only, we restricted our searches to studies conducted up to a resolution of 1 km^2 (~30 arc seconds). Studies at coarser resolution (and often larger scale) effectively tend to include only climatic variables due to data availability and the scale-dependence of different predictors (Pearson & Dawson 2003, Thuiller et al. 2004; but see Harwood et al. 2014). From the 745 ‘recent’ articles found in the WoS, 182 met our requirements (that is, they involved actual SDMs concerning plants and had a maximum 1 km^2 resolution). Hereafter, however, our analyses include 200 studies due to some articles using distinct sets of predictors for different species or different spatial resolutions. Each of these studies were divided into separate studies. Of the ‘temporal’ articles, forty pertained to plants and were conducted at a maximum resolution of 1 km^2. The resulted dataset was used to examine the number and type of predictors included in the models. Especially, this was done in order to distinguish which predictors are frequently used in the SDMs, and on the other hand, which predictors are not used and might require further developing.

To account for environmental and spatial coverage, we recorded the continent and biome of origin of the data. The articles included study areas from all continents. Most studies were from Europe (n = 84) and North America (n = 53), with fewer studies from Australia (n = 25), Africa (n = 20), Latin America (n = 15) and Asia (n = 12). All biomes were covered with an expected bias towards European and North American biomes (temperate, boreal, Mediterranean, alpine, arctic) where more studies have been conducted overall.

RESULTS

In the ‘recent’ articles, the average number of predictors included in the models was eleven (Fig. 2). The number of predictors considered in the models varied from one to 75. The different classes of variables covered in the models varied from one to thirteen (out of the 16 defined in this study), with only two studies covering all eight of our categories (Fig. 2). Several variables under one class and/or category were often simultaneously included as predictors. Variables from the five most essential categories (temperature, water, substrate, radiation, biotic interactions) were included in seven studies, with all of these also including disturbance, topography and/or land-use related variables. Overall, the reviewed studies represent considerable variability in the different variables used. In particular, the ‘water balance’ and ‘biotic’ classes included various sets of different types of factors (see Appendix S3).

Most of the ‘recent’ studies included temperature- and water-related variables (both were included in 88.5 % of studies). Each of the temperature sub-classes appeared in more than half of the SDMs. The most frequently included water-related variables were monthly or annual mean precipitation (68.5 %), with extreme and seasonal precipitation and water balance appearing in approximately one third of the studies (Fig. 3). Approximately one third of the studies included only climatic variables (derived from temperature and/or precipitation). Measurements or approximations of actual or potential soil water or soil moisture were incorporated in 15 studies.

Substrate-related variables were used in ~ 40 % of the studies, and variables directly representing bedrock/pH or nutrients were included in approximately one quarter of the studies. Only 60 studies involved variables representing light. One fifth of the studies included some biotic component as a predictor variable. Variables representing natural disturbances were included in 17 studies. Variables related to human activity were included in 19 studies.

After climatic variables, topographic factors were most commonly included in the SDMs screened in this study (44.5 %). Land use was included in 32 studies, with one study using land use as a mask to exclude certain areas.

There were no significant differences in the number of variable classes used among the continents (Fig. 4). Only Latin America (LAm) had a significantly lower number of variable categories compared with the other continents.

The ‘temporal search’ showed no increase in the number of categories accounted for in the SDMs through time (2000-2015). On the contrary, the number of variables from different categories showed a decreasing trend (Spearman’s rank correlation -0.40*; Fig. 5). Exceptions were the SDM studies from 2011 (by Austin and Van Niel (2011b), Meier et al., Mellert et al. and Ohmann et al.), which increased the number of categories included; all studies discussed the importance of selecting variables on an ecological basis or the impacts of omitting meaningful predictors in the models and thus included variables from multiple categories.

DISCUSSION

Ecological theory, supported by experimental and correlative studies, stresses that multiple environmental factors drive the distribution of species (e.g., Larcher 1975, Fitter & Hay 2002, Schulze et al. 2005, see also e.g., Guisan & Zimmermann 2000; Elith & Leathwick 2009; Franklin 2009; Austin & Van Niel 2011a; Bertrand et al. 2012; Dubuis et al. 2013; le Roux et al. 2013a, b), particularly temperature, water, nutrients, light, biotic interactions and disturbances (see Appendix S1). In recently published SDM studies, many of these factors were omitted or replaced with rough surrogates (e.g., precipitation for plant available water). Indeed, more than half (53 %) of the plant SDM studies reviewed here based their predictions solely on the categories of temperature and water or on those two categories plus one additional variable, thus potentially neglecting several other ecophysiologically relevant aspects (e.g., substrate, radiation and/or biotic interactions. Although it is important to highlight that not all of these categories might be meaningful for all SDMs; see the next paragraph). While data availability is likely a potential reason for the omission of ecophysiologically meaningful predictors, the wide range of variables used in some exemplar studies (see next sections and Appendix S3) indicates that some influential and available predictors may tend to be neglected. Furthermore, there was no difference in the number of predictor classes used in studies from the “data rich” continents (Europe, North America) and the “data poor” continents (Fig. 4), suggesting that data availability may not be a sufficient explanation for the absence of important predictors in the models.

The intentional use of an ecophysiologically incomplete set of predictors in correlative modelling is acceptable, for instance, if the study deliberately focuses on the climatic niche or climatic range only, provided that this is clearly acknowledged. Therefore, it is important to distinguish here between two classes of studies according to their ultimate goal: studies which aim would require including all potentially important variables (e.g. fine-scale predictions for conservation, or addressing aspects of species’ ecology in general), and studies which aim does not necessarily require more than one type of predictors (e.g. climate-change studies only interested in fitting species’ climatic niches and climatic ranges). Also, in some other cases, a comprehensive set of meaningful predictors may not be essential in SDMs (e.g., when illustrating the development of new methodologies, or if models representing a specific aspect of the niche are explicitly desired; Thuiller et al. 2005). Nevertheless, in all type of SDMs, it is important to justify the choice of predictors, and interpret the results in accordance with used predictors. Indeed, only few of the studies reviewed here acknowledged the ecophysiologically incomplete set of environmental drivers used as predictors (e.g., Bertrand et al. 2012; Aguirre-Gutiérrez et al. 2013; Ikeda et al. 2014; Riordan & Rundel 2014, Petitpierre et al. in review), and many studies provided no ecological rationale for the choice of predictors. In the next sections, focusing our discussion on SDMs aiming to comprehensively capture species ecological niche, we aim to provide such rationale, discuss ways to account for the needed predictors in SDMs, and identify missing predictors for which development and mapping are still needed at a fine scale. However, we do not provide any estimates of an adequate number of predictors, which depends on the number and distribution of species occurrences and the algorithm or approach used (see e.g., Wisz et al. (2008) and Franklin (2009)).

Temperature

Temperature and water-related variables were the most commonly used predictors among the reviewed studies (Fig. 3). While temperature is frequently accounted for in the models and plays an indisputable role in regulating plant species growth and thus, distribution (see Appendix S1), two noteworthy issues concerning temperature were identified from our literature analyses. First, there is a large variety of temperature data products available, with the class of temperature variable used having an impact on model performance (Barbet-Massin & Jetz 2014; Slavich et al. 2014). For example, the impact of mean temperature on plants differs from that of extremes or seasonality in both ecological meaning and modelling performance (Zimmermann et al. 2009). In seasonally variable environments especially, annual mean temperature does not represent the growing season or over-wintering conditions, which potentially play a more central role in governing the distribution of plants (Aerts et al. 2006; Paulsen & Körner 2014). One solution to choose between different temperature-related variables might be to include multiple variables in a model, as exemplified by many studies using climatic data provided by WorldClim (Hijmans et al. 2005). However, this raises problems of multicollinearity (Graham 2003; Dormann et al. 2013) and conflicts with the objective of parsimony (Mac Nally 2000). Ultimately, the environmental conditions of the study area and the requirements of the species should determine the most suitable temperature-related variable(s) – a viewpoint only rarely considered or tested in the modelling studies.

Second, while there is a multitude of temperature data readily available for modelling, their resolution and accuracy can be coarse compared with the species data (Dingman et al. 2013; Franklin et al. 2013; Potter et al. 2013; Pradervand et al. 2014). Temperature measurements are typically obtained by interpolating sparse measurements and neglecting the impact of local topography, land cover or water bodies on local temperatures experienced by plants (Scherrer & Körner 2011; Franklin et al. 2013; Aalto et al. 2014; Slavich et al. 2014). Alternatively, improved temperature maps could be obtained by a combination of increased field measurements (e.g., thermal loggers), predictive methods, high-resolution digital elevation models (DEMs) and thermal remote sensing rather than spatial interpolations (Scherrer and Körner 2010, Dingman et al. 2013; Pradervand et al. 2014). Thus, while the availability of temperature data is not a primary problem, their usability and ecological significance in SDMs could be improved by increasing their resolution and accuracy.

Water

Predictors representing water availability for plants are often derived from precipitation, a class of climatic predictors inheriting similar challenges to those discussed for temperature. In addition, precipitation is a poor surrogate for plant available water, especially in high-resolution studies that cover small areas, due to the effects of local topography and soil substrate on the amount and distribution of soil moisture (le Roux et al. 2013c; Piedallu et al. 2013). Therefore, while water as a category of predictors is almost always acknowledged in the models, the ecophysiological significance of the water predictors being used might be poor in many cases. Some studies have used water balance (precipitation minus evapotranspiration), which represents a more accurate measure of plant available water compared with precipitation. Some soil moisture indices derived from climate data and geographic information systems (GIS) modelling are available (e.g., Trabucco & Zomer 2010), but these proxies also neglect the impact of terrain on plant available moisture. Using high-resolution topographic information in combination with climate and soil measurements could provide a more promising basis for modelling high-resolution soil moisture data (Aalto et al. 2013; Pradervand et al. 2014).

Ideally, soil moisture measurements taken in the field should most accurately represent the water available to plants. Studies that incorporate field-quantified soil moisture values in their models have improved predictive power, especially at high spatial resolutions (le Roux *et al.*, 2013c). However, collecting these high-resolution and accurate soil moisture data over large areas is rarely feasible. Remote sensing combined with GIS provides ready-to-use (coarse-scale) indices of moisture or wetness (e.g., the surface saturation degree of ASCAT soil wetness indices, see Brocca et al. 2010; Lakshmi 2013; Wagner et al. 2013), and other recent developments such as Synthetic Aperture Radars (Elbialy et al. 2014), hyperspectral aerial images (Pottier et al. 2014) and spatial modelling (Aalto et al. 2013) show promise in estimating actual soil moisture at higher resolutions. To conclude, although often accounted for in SDMs with distal predictors, water-related variables could be improved through combined approaches mixing refined field measures, GIS modelling and remote sensing.

Nutrients

The role of soil and its nutrients on plant performance is acknowledged by most ecologists (Epstein & Bloom 2005; see also Appendix S1) as well as their role on model performance by many modellers (almost half in our study; see also Coudun et al. 2006; Coudun & Gégout 2007; Bertrand et al. 2012; Dubuis et al. 2013). It seems hardly feasible to obtain high-resolution field measurements of nutrient content and geo-chemical properties of soils across a whole study area.

Thus, most studies that included substrate variables used either geological or geomorphological surrogates such as bedrock, pH or landforms, or factors related to soil structure, such as texture or soil depth (Bertrand et al. 2012; Dubuis et al. 2013). This highlights the need for more sophisticated indices of soil nutrient content, analogous to those being developed for soil moisture. The use of soil ecological indicator values (e.g., Ellenberg) also highlights such a need (Coudun et al. 2006). Improved spatial predictors of soil characteristics are thus still required, such as those derived from remote sensing (Parviainen et al. 2013) or potentially from statistical modelling (Lagacherie 1992), to further improve plant SDMs (Dubuis et al. 2013).

Light

The importance of light for plants and its use as a predictor in SDMs were previously discussed by Austin and Van Niel (2011a). Solar radiation can be calculated using DEM and, if available, canopy cover in efficient GIS tools (McCune & Keon 2002). However, light-related variables were only included in less than one third of the studies we reviewed, meaning that more than two thirds of the reviewed studies neglected an important factor controlling plant distributions, especially at local scales. In the studies accounting for light, it was mostly represented by the sum of (potential) solar radiation over various seasons. In these cases, the radiation variable actually expresses heat rather than photosynthetically active radiation (PAR) and therefore acts similarly to temperature. To obtain a real measure of PAR, light must be measured specifically, and the effects of cloud cover and canopy interception must be taken into account (Aguiar et al. 2012; Wang et al. 2014). Nevertheless, inclusion of a solar radiation variable often improves model prediction by adding information on fine-scale energy input, especially in topographically heterogeneous areas (Austin & Van Niel 2011a). At a given elevation, slopes with different aspects can have very different soil and vegetation temperatures (Scherrer & Körner 2010; Gunton et al. 2015). In contrast to average temperatures based mostly on adiabatic lapse rates, solar radiation can include information regarding aspect, relief shading and daylight period (Kumar et al. 1997; Austin & Van Niel 2011a). However, as mentioned before, the use of solar radiation as a predictor can lead to misleading interpretations, as its impact on plants might strongly depend on season, canopy structure and cloud cover. Thus, the radiation variables should firstly be incorporated into SDMs, seasonal variations should be accounted for, and the effects of canopy and cloud cover should be included when studying understory vegetation (Nieto-Lugilde et al. 2015).

Biotic interactions

Biotic interactions play a role in altering the potential environmental niche, for example, through competition, facilitation and herbivory (Brooker & Callaghan 1998; Callaway et al. 2002; Araújo & Luoto 2007; Pellissier et al. 2010; Mod et al. 2014). As the importance of biotic interactions and how to measure their importance (Godsoe & Harmon 2012) and account for them in SDMs are still under discussion (Kissling et al. 2012; Wisz et al. 2013), many SDMs do not include biotic factors. Implicitly, these SDMs assume that the important biotic interactions (in a given area or habitat) are already indirectly accounted for at the sampling stage (when gathering observations) because biotic interactions influence the realized distribution of the species (McGill et al. 2006) and are thus captured in the realized environmental niche (Araùjo & Guisan 2006). Nonetheless, biotic components were used in approximately one-fifth of the studies, indicating their increasing importance in SDMs. However, explicit information on biological interactions remains difficult to obtain in a spatially explicit form, as the biotic factors governing the assemblage of individual species into communities are still largely unknown (Kissling et al. 2012, Wisz et al. 2013), and associated assembly rules remain to be developped (Guisan & Rahbek 2011). However, surrogates such as dominant species cover have been shown to provide some measure of biotic interactions (le

Roux et al. 2014), and incorporating these surrogates has improved both the explanatory and predictive power of SDMs (Meier et al. 2010; Pellissier et al. 2010). Various methods to account for biotic interactions in SDMs are presented in Kissling et al. (2012), Wisz et al. (2013) and Pollock et al. (2014).

Disturbance

The type and necessity of including disturbance variables in models are highly environment-specific. Frost-related disturbances can strongly impact vegetation in arctic and alpine areas by destroying some species and subsequently, creating space for other species (le Roux et al. 2013a; le Roux & Luoto 2014). In dryer areas, fire may play such a role (Tucker et al. 2012, but see Crimmins et al. 2013). Disturbance has been incorporated in some models, for example, as the proportion of the area that is disturbed (le Roux et al. 2013a), as an index of geomorphic disturbances (Randin et al. 2009a), or as time elapsed since the last fire (Moretti et al. 2006). The use of predictors related to natural disturbances in SDMs may be particularly important when analysing the potential impacts of changing climate because changes in the intensity of these processes associated with climatic shifts may represent key mechanisms by which changes in temperature and rainfall patterns affect vegetation assemblages (le Roux & Luoto 2014, although see Crimmins et al. 2013). Similar to other disturbances, the use of anthropogenic predictors is situational, depending on the study environment, species and study target. For semi-natural or urban landscapes and/or species highly associated with humans, the use of anthropogenic predictors might be crucial to obtain reasonable predictions (Kouba et al. 2011; Senan et al. 2012).

Topography and land use

Variables representing topography are often included in plant distribution models (see also Franklin 1995). Including these variables has been demonstrated to improve plant SDMs (e.g., Sormunen et al. 2011), but interpreting the actual drivers of plant distributions related to these variables can be difficult. Because the effects of topographic variables on plant distributions are distal (i.e., they do not directly impact plants, but they do alter light, moisture, temperature and nutrient conditions; Moeslund et al. 2013), it is not possible to interpret the causal relationships between these variables and the target species (Austin 2007). Correlation between indirect gradients and species distribution results only from location dependence (Austin 2002). Despite the demonstrated ability of topographic variables to improve local models, the use of these indirect variables hampers understanding of proximal species-environment relationships and reduces transferability (Randin et al. 2006). Field quantification of environmental variables or the use of purely proximal variables (sensu Austin 2002) would assist in identifying the actual environmental factors that species respond to and would thus provide more detailed understanding of species distributions and ultimately, yield more realistic SDMs. Therefore, using in-situ measured direct and resource variables instead of indirect gradients (such as elevation, aspect and topographic position) would be advisable (Austin 2002; Pradervand et al. 2014), especially when SDMs are also used to explain species distributions. Land use was occasionally included in the models we reviewed. Its inclusion usually improves the explanatory and predictive power of SDMs (Von Holle & Motzkin 2007) but only for predicting species abundances in some cases (Randin et al. 2009b). However, interpreting the proximal impact of land-use predictors on plant distributions suffers the same problems discussed for topographic variables (i.e., being often not proximal).

Implications for future studies

As hypothesized, limited data availability could be one justification for omitting potentially influential ecophysiologically predictors in SDMs despite their demonstrated advantages for the explanatory and predictive power (e.g., Austin & Van Niel 2011b, Bertrand et al. 2012, le Roux et al. 2014). The other hypothesized explanation was the intended omission, e.g., in studies of climatic niches and ranges (e.g., Thuiller 2005, Petitpierre et al. 2012). However, data unavailability and intended omission can hardly explain all instances (especially in data-rich areas of Europe, North-America and Australia, Fig. 4) where important non-climatic factors were excluded (see similar statement made 20 years previously by Franklin 1995). Indeed, many of the studies provided no justification for the choice of predictors or only provided a reference to another study relying on a similar set of predictors without considering the influence of the study area or the ecophysiological requirements of the studied species to determine a meaningful set of predictors. Furthermore, despite increasing recognition of the importance of a variety of environmental variables for predicting plant distributions (e.g. Austin & Van Niel 2011a, Dubuis et al. 2013) and the increasing availability of numeric data (including from remote sensing), the number of ecophysiologically significant variable categories considered in SDMs seems rather to have decreased during the 21st century. Therefore, we argue that in the future, an ecologically sound reasoning for the choice of predictors in the SDMs should become common practice, and the models and predictions should always be interpreted in perspective of the set of predictors used.

In addition, our literature review highlighted that some variable classes are poorly represented in terms of data quantity (e.g. global coverage) and quality (e.g. resolution). More attention should be paid to ensure that all relevant environmental predictors are made available for modelling at the scale investigated. Although measuring or deriving proximal predictors over large areas can be difficult for single researchers, large international efforts are increasingly developed to use remote sensing products for such purpose (Zimmermann et al. 2007, Estes et al. 2010). More research should also be dedicated to produce finer-scale and more proximal data to improve our understanding of the factors driving species distributions (Gunton et al. 2015) and therefore, the production of more realistic predictions. Here too, remote sensing and GIS can produce promising data products (Bradley et al. 2012, Pottier et al. 2014, He et al. 2015), and ecologists and ecological modellers should give more attention to collaborative research within the geo-environmental sciences.

CONCLUSIONS

Our study reveals that the rationale, selection and use of environmental predictors in many plant species distribution models do not systematically match established ecophysiological theory, perspectives on ecologically meaningful variable selection or demonstrated improvements in SDMs, and therefore calls for the need to add several meaningful variables in SDMs. Except for the pure climatic niche studies and methodological experiments, many plant SDMs so far have omitted important environmental variables, and the number of predictors representing the essential ecophysiological aspects pertaining to plants has not increased during the 21st century, despite increased numerical data availability. In particular, nutrients, actual light, disturbance and biotic interactions should be incorporated more systematically into SDMs, together with the most commonly used temperature and water variables. Furthermore, the type of temperature and water variables to be used should also be given more careful attention. The development of new environmental variables will require improved collaborative research between ecological and geo-environmental sciences as well as access to advanced technology, such as remote sensing and GIS modelling approaches. Developing new sets of ecophysiologically more meaningful predictors provides the basis for a paradigm change in SDM research.

ACKNOWLEDGEMENTS

H.M. received funding from the Research Foundation of the University of Helsinki and the Kordelin foundation. M.L. was partly funded by the Academy of the Finland project nr 286950. A.G. received financial support from the Swiss National Science Foundation (Grant 31003A-152866/1).

AUTHOR CONTRIBUTIONS

A.G. and M.L. conceived the idea and outline for this manuscript; H.M. and D.S. performed the literature review; H.M. and D.S. led the writing, with A.G. and M.L.

REFERENCES

Aalto, J., le Roux, P.C. & Luoto, M. 2013. Vegetation mediates soil temperature and moisture in arctic-alpine environments. *Arctic, Antarctic, and Alpine Research* 45: 429-439.

Aalto, J., le Roux, P.C. & Luoto, M. 2014. The meso-scale drivers of temperature extremes in high-latitude fennoscandia. *Climate Dynamics* 42: 237-252.

Aerts, R., Cornelissen, J.H.C. & Dorrepaal, E. 2006. Plant performance in a warmer world: General responses of plants from cold, northern biomes and the importance of winter and spring events. *Plant Ecology* 182: 65-77.

Aguiar, L.J.G., Fischer, G.R., Ladle, R.J., Malhado, A.C.M., Justino, F.B., Aguiar, R.G. & Costa, J.M.N. 2012. Modeling the photosynthetically active radiation in South West Amazonia under all sky conditions. *Theoretical and Applied Climatology* 108: 631-640.

Aguirre-Gutiérrez, J., Carvalheiro, L.G., Polce, C., van Loon, E.E., Raes, N., Reemer, M. & Biesmeijer, J.C. 2013. Fit-for-purpose: Species distribution model performance depends on evaluation criteria – Dutch hoverflies as a case study. *PLoS One* 8: e63708.

Alagador, D., Cerdeira, J.O. & Araújo, M.B. 2014. Shifting protected areas: Scheduling spatial priorities under climate change. *Journal of Applied Ecology* 51: 703-713.

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

Araújo, M.B. & Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography* 16: 743-753.

Araújo, M.B., Alagador, D., Cabeza, M., Nogués-Bravo, D. & Thuiller, W. 2011. Climate change threatens european conservation areas. *Ecology Letters* 14: 484-492.

Austin, M.P. 1980. Searching for a model for use in vegetation analysis. *Vegetatio* 42: 11-21.

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. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling* 200: 1-19.

Austin, M.P. & Smith, T.M. 1989. A new model for the continuum concept. *Vegetatio* 83: 35-47.

Austin, M.P. & Van Niel, K.P. 2011a. Improving species distribution models for climate change studies: Variable selection and scale. *Journal of Biogeography* 38: 1-8.

Austin, M.P. & Van Niel, K.P. 2011b. Impact of landscape predictors on climate change modelling of species distributions: A case study with *Eucalyptus fastigata* in southern New South Wales, Australia. *Journal of Biogeography* 38: 9-19.

Bader, M.K.F., Leuzinger, S., Keel, S.G., Siegwolf, R.T.W., Hagedorn, F., Schleppi, P. & Körner, C. 2013. Central European hardwood trees in a high-Co_2 future: Synthesis of an 8-year forest canopy Co_2 enrichment project. *Journal of Ecology* 101: 1509-1519.

Barbet-Massin, M. & Jetz, W. 2014. A 40-year, continent-wide, multispecies assessment of relevant climate predictors for species distribution modelling. *Diversity and Distributions* 20: 1285-1295.

Bertrand, R., Perez, V. & Gegout, J.C. 2012. Disregarding the edaphic dimension in species distribution models leads to the omission of crucial spatial information under climate change: The case of *Quercus pubescens* in France. *Global Change Biology* 18: 2648-2660.

Booth, T. H., Nix, H. A., Busby, J. R. & Hutchinson, M. F. 2014. BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Diversity and Distributions* 20: 1-9.

Bradley, B.A., Olsson, A.D., Wang, O., Dickson, B.G., Pelech, L., Sesnie, S.E. & Zachmann, L.J. 2012. Species detection vs. Habitat suitability: Are we biasing habitat suitability models with remotely sensed data? *Ecological Modelling* 244: 57-64.

Brocca, L., Melone, F., Moramarco, T., Wagner, W. & Hasenauer, S. 2010. Ascat soil wetness index validation through in situ and modeled soil moisture data in central italy. *Remote Sensing of Environment* 114: 2745-2755.

Broennimann, O. & Guisan, A. 2008. Predicting current and future biological invasions: Both native and invaded ranges matter. *Biology Letters* 4: 585-589.

Brooker, R.W. & Callaghan, T.V. 1998. The balance between positive and negative plant interactions and its relationship to environmental gradients: A model. *Oikos* 81: 196-207.

Callaway, R.M., Brooker, R.W., Choler, P., Kikvidze, Z., Lortie, C.J., Michalet, R., Paolini, L., Pugnaire, F.I., Newingham, B., Aschehoug, E.T., Armas, C., Kikodze, D. & Cook, B.J. 2002. Positive interactions among alpine plants increase with stress. *Nature* 417: 844-848.

Coudun, C. & Gégout, J.-C. 2007. Quantitative prediction of the distribution and abundance of *Vaccinium myrtillus* with climatic and edaphic factors. *Journal of Vegetation Science*, 18: 517-524.

Coudun, C., Gégout, J.-C., Piedallu, C. & Rameau, J.-C. 2006. Soil nutritional factors improve models of plant species distribution: An illustration with *Acer campestre* (l.) in France. *Journal of Biogeography* 33: 1750-1763.

Crimmins, S.M., Dobrowski, S.Z., Mynsberge, A.R. & Safford, H.D. 2013. Can fire atlas data improve species distribution model projections? *Ecological applications* 24: 1057-1069.

D'Amen, M., Rahbek, C., Zimmermann, N.E. & Guisan A. in press. Spatial predictions at the community level: from current approaches to future frameworks. *Biological Reviews* 10.1111/brv.12222.

Dormann, C. 2007. Promising the future? Global change projections of species distributions. *Basic and Applied Ecology* 8: 387-397.

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.

Dubuis, A., Giovanettina, S., Pellissier, L., Pottier, J., Vittoz, P. & Guisan, A. 2013. Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. *Journal of Vegetation Science* 24: 593-606.

Elbialy, S., Mahmoud, A., Pradhan, B. & Buchroithner, M. 2014. Application of spaceborne synthetic aperture radar data for extraction of soil moisture and its use in hydrological modelling at gottleuba catchment, Saxony, Germany. *Journal of Flood Risk Management* 7: 159-175.

Elith, J. & Leathwick, J.R. 2009. Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40: 677-697.

Epstein, E. & J., B.A. (2005) *Mineral nutrition of plants: principles and perspectives*, 2 edn. Sinauer Associates, Sunderland.

Estes, L. D., P. R. Reillo, Mwangi, A. G., Okin, G. S. & Shugart, H. H. 2010. Remote sensing of structural complexity indices for habitat and species distribution modeling. *Remote Sensing of Environment* 114: 792-804.

Fitter, A.H. & Hay, R.K.M. (2002) *Environmental physiology of plants*, 3 ed. Academic Press, London & San Diego.Franklin, J. 1995. Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19: 474-499.

Franklin, J. 2009. *Mapping species distributions: Spatial inference and prediction*. Cambridge University Press Cambridge.

Fukami, T., Martijn Bezemer, T., Mortimer, S.R. & van der Putten, W.H. 2005. Species divergence and trait convergence in experimental plant community assembly. *Ecology Letters* 8: 1283-1290.

Godsoe, W. & Harmon, L.J. 2012. How do species interactions affect species distribution models? *Ecography* 35: 811–820.

Graham, M.H. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology* 84: 2809-2815.

Grime, J.P. 1977. Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *The American naturalist* 111: 1169-1194.

Grinnell, J. 1917. The niche-relationships of the california thrasher. *The Auk*: 427-433.

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

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

Guisan, A. & Rahbek, C. 2011. SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography* 38: 1433-1444.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C., Aspinall, R. & Hastie, T. 2006. Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology* 43: 386-392.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P. & Buckley, Y.M. 2013. Predicting species distributions for conservation decisions. *Ecology Letters* 16: 1424-1435.

Gunton, R.M., Polce, C. & Kunin, W.E. 2015. Predicting ground temperatures across european landscapes. *Methods in Ecology and Evolution* 6: 532–542.

Harwood, T.D., Mokany, K. & Paini, D.R. 2014. Microclimate is integral to the modeling of plant responses to macroclimate. *Proceedings of the National Academy of Sciences of the United States of America*, 111: E1164-E1165.

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

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.

Huston, M.A. (2002) Introductory essay: Critical issues for improving predictions. *Predicting species occurrences: Issues of accuracy and scale* (eds. by J.M. Scott, P.J. Heglund and M.L. Morrison), pp. 7-21.

Hutchinson, G.E. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22: 415-427.

Hutchinson, G.E. 1978. *An introduction to population ecology*. Yale University Press, New Haven and London.

Ikeda, D.H., Grady, K.C., Shuster, S.M. & Whitham, T.G. 2014. Incorporating climate change and exotic species into forecasts of riparian forest distribution. *Plos One* 9: e107037.

Inauen, N., Körner, C. & Hiltbrunner, E. 2012. No growth stimulation by CO_2 enrichment in alpine glacier forefield plants. *Global Change Biology* 18: 985-999.

Jiménez-Valverde, A., Peterson, A.T., Soberón, J., Overton, J.M., Aragón, P. & Lobo, J.M. 2011. Use of niche models in invasive species risk assessments. *Biological Invasions* 13: 2785-2797.

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.-C., 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.

Kouba, Y., Alados, C.L. & Bueno, C.G. 2011. Effects of abiotic and anthropogenic factors on the spatial distribution of *Quercus faginea* in the spanish central pyrenees. *Plant Ecology* 212: 999-1007.

Kumar, L., Skidmore, A.K. & Knowles, E. 1997. Modelling topographic variation in solar radiation in a gis environment. *International Journal of Geographical Information Science* 11: 475-497.

Körner, C. 2006. Plant CO_2 responses: An issue of definition, time and resource supply. *New Phytologist* 172: 393-411.

Körner, C. (2014) Pflanzen im lebensraum. *Strasburger - lehrbuch der pflanzenwissenschaften* (eds. by J.W. Kadereit, C. Körner, B. Kost and U. Sonnewald), pp. 759-810. Springer Spektrum, Berlin.

Lagacherie, P. 1992. *Formalisation des lois de distribution des sols pour automatiser la cartographie pédologique à partir d'un secteur pris comme référence*. PhD Thesis, Université de Montpellier, Montpellier.

Lakshmi, V. 2013. Remote sensing of soil moisture. *ISRN Soil Science* 33, 1-33.

Lambers, H., Chapin III, F.S. & Pons, T.L. 2008. *Plant water relations*. Springer, New York, US.

Larcher, W. 1975. *Physiological plant ecology.* 2nd ed. Springer, London, UK.

le Roux, P.C. & Luoto, M. 2014. Earth surface processes drive the richness, composition and occurrence of plant species in an arctic–alpine environment. *Journal of Vegetation Science* 25: 45-54.

le Roux, P.C., Virtanen, R. & Luoto, M. 2013a. Geomorphological disturbance is necessary for predicting fine-scale species distributions. *Ecography* 36: 800-808.

le Roux, P.C., Aalto, J. & Luoto, M. 2013c. Soil moisture's underestimated role in climate change impact modelling in low-energy systems. *Global Change Biology* 19: 2965-2975.

le Roux, P.C., Pellissier, L., Wisz, M.S. & Luoto, M. 2014. Incorporating dominant species as proxies for biotic interactions strengthens plant community models. *Journal of Ecology* 102: 767-775.

le Roux, P.C., Lenoir, J., Pellissier, L., Wisz, M.S. & Luoto, M. 2013b. Horizontal, but not vertical, biotic interactions affect fine-scale plant distribution patterns in a low energy system. *Ecology* 94: 671-682.

Mac Nally, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: The distinction between – and reconciliation of – 'predictive' and 'explanatory' models. *Biodiversity & Conservation* 9: 655-671.

McCune, B. & Keon, D. 2002. Equations for potential annual direct incident radiation and heat load. *Journal of Vegetation Science* 13: 603-606.

McGill, B.J. 2010. Matters of scale. *Science* 328: 575-576.

McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. 2006. Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution* 21: 178-185.

Meier, E.S., Edwards Jr, T.C., Kienast, F., Dobbertin, M. & Zimmermann, N.E. 2011. Co-occurrence patterns of trees along macro-climatic gradients and their potential influence on the present and future distribution of *fagus sylvatica* l. *Journal of Biogeography* 38: 371-382.

Meier, E.S., Kienast, F., Pearman, P.B., Svenning, J.-C., Thuiller, W., Araújo, M.B., Guisan, A. & Zimmermann, N.E. 2010. Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography* 33: 1038-1048.

Mellert, K.H., Fensterer, V., Kuechenhoff, H., Reger, B., Koelling, C., Klemmt, H.J. & Ewald, J. 2011. Hypothesis-driven species distribution models for tree species in the Bavarian Alps. *Journal of Vegetation Science* 22: 635-646.

Mod, H.K., le Roux, P.C. & Luoto, M. 2014. Outcomes of biotic interactions are dependent on multiple environmental variables. *Journal of Vegetation Science* 25: 1024-1032.

Moeslund, J.E., Arge, L., Bøcher, P.K., Dalgaard, T. & Svenning, J.-C. 2013. Topography as a driver of local terrestrial vascular plant diversity patterns. *Nordic Journal of Botany* 31: 129-144.

Moretti, M., Conedera, M., Moresi, R. & Guisan, A. 2006. Modelling the influence of change in fire regime on the local distribution of a mediterranean pyrophytic plant species (*Cistus salviifolius*) at its northern range limit. *Journal of Biogeography* 33: 1492-1502.

Mouquet, N., Lagadeuc, Y., Devictor, V., Doyen, L., Duputié, A., Eveillard, D., Faure, D., Garnier, E., Gimenez, O., Huneman, P., Jabot, F., Jarne, P., Joly, D., Julliard, R., Kéfi, S., Kergoat, G.J., Lavorel, S., Le Gall, L., Meslin, L., Morand, S., Morin, X., Morlon, H., Pinay, G., Pradel, R., Schurr, F.M., Thuiller, W. & Loreau, M. 2015. Predictive ecology in a changing world. *Journal of Applied Ecology* 52: 1293-1310.

Nieto-Lugilde D., Lenoir J., Abdulhak S., Aeschimann D., Dullinger S., Gégout J.-C., Guisan A., Pauli H., Renaud J., Theurillat J.-P. Thuiller, W., Van Es, J., Vittoz, P., Willner, W., Wohlgemuth, T., Zimmermann, N.E. & Svenning, J.-C. 2015. Tree cover at fine and coarse spatial grains interacts with shade tolerance to shape plant species distributions across the Alps. *Ecography* 37: 1-12.

Norby, R.J. & Zak, D.R. 2011. Ecological lessons from free-air Co_2 enrichment (face) experiments. *Annual Review of Ecology, Evolution, and Systematics* 42: 181.

Ohmann, J.L., Gregory, M.J., Henderson, E.B. & Roberts, H.M. 2011. Mapping gradients of community composition with nearest-neighbour imputation: extending plot data for landscape analysis. *Journal of Vegetation Science* 22: 660-676.

Parviainen, M., Zimmermann, N., Heikkinen, R. & Luoto, M. 2013. Using unclassified continuous remote sensing data to improve distribution models of red-listed plant species. *Biodiversity and Conservation* 22: 1731-1754.

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

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.

Pellissier, L., Bråthen, K.A., Pottier, J., Randin, C.F., Vittoz, P., Dubuis, A., Yoccoz, N.G., Alm, T., Zimmermann, N.E. & Guisan, A. 2010. Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography* 33: 1004-1014.

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

Petitpierre, B., C. Kueffer, Broennimann, O., Randin, C., Daehler, C. & Guisan, A. 2012. Climatic niche shifts are rare among terrestrial plant invaders. *Science* 335: 1344-1348.

Petitpierre, B., Broenniman, O., Kueffer, C., Daehler, C. and Guisan A. in review. Selecting the best set of predictors to build transferable species distribution models. *Global Ecology and Biogeography*.

Piedallu, C., Gegout, J.-C., Perez, V. & Lebourgeois, F. 2013. Soil water balance performs better than climatic water variables in tree species distribution modelling. *Global Ecology and Biogeography* 22: 470-482.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A. & McCarthy, M.A. 2014. Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution* 5: 397-406.

Potter, K.A., Woods, H.A. & Pincebourde, S. 2013. Microclimatic challenges in global change biology. *Global Change Biology* 19: 2932-2939.

Pottier, J., Malenovský, Z., Psomas, A., Homolová, L., Schaepman, M.E., Choler, P., Thuiller, W., Guisan, A. & Zimmermann, N.E. 2014. Modelling plant species distribution in alpine grasslands using airborne imaging spectroscopy. *Biology Letters* 10

Pradervand, J.-N., Dubuis, A., Pellissier, L., Guisan, A. & Randin, C. 2014. Very high resolution environmental predictors in species distribution models: Moving beyond topography? *Progress in Physical Geography* 38: 79-96.

Randin, C.F., Vuissoz, G., Liston, G.E., Vittoz, P. & Guisan, A. 2009a. Introduction of snow and geomorphic disturbance variables into predictive models of alpine plant distribution in the western swiss alps. *Arctic Antarctic and Alpine Research* 41: 347-361.

Randin, C.F., Jaccard, H., Vittoz, P., Yoccoz, N.G. & Guisan, A. 2009b. Land use improves spatial predictions of mountain plant abundance but not presence-absence. *Journal of Vegetation Science* 20: 996-1008.

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. 2006. Are niche-based species distribution models transferable in space? *Journal of Biogeography* 33: 1689-1703.

Riordan, E.C. & Rundel, P.W. 2014. Land use Compounds habitat losses under projected climate change in a threatened california ecosystem. *Plos One* 9: e86487.

Scherber, C., Eisenhauer, N., Weisser, W.W., Schmid, B., Voigt, W., Fischer, M., Schulze, E.-D., Roscher, C., Weigelt, A., Allan, E., Beszler, H., Bonkowski, M., Buchmann, N., Buscot, F., Clement, L.W., Ebeling, A., Engels, C., Halle, S., Kertscher, I., Klein, A.-M., Koller, R., Konig, S., Kowalski, E., Kummer, V., Kuu, A., Lange, M., Lauterbach, D., Middelhoff, C., Migunova, V.D., Milcu, A., Muller, R., Partsch, S., Petermann, J.S., Renker, C., Rottstock, T., Sabais, A., Scheu, S., Schumacher, J., Temperton, V.M. & Tscharntke, T. 2010. Bottom-up effects of plant diversity on multitrophic interactions in a biodiversity experiment. *Nature* 468: 553-556.

Scherrer, D. & Körner, C. 2010. Infra-red thermometry of alpine landscapes challenges climatic warming projections. *Global Change Biology* 16: 2602-2613.

Scherrer, D. & Körner, C. (2011). Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. *Journal of Biogeography* 38: 406-416.

Schulze, E., Beck, E. & Muller-Hohenstein, K. 2005. *Plant Ecology*, 1st ed. Springer-Verlag, Berlin.Senan, A.S., Tomasetto, F., Farcomeni, A., Somashekar, R.K. & Attorre, F. 2012. Determinants of plant species invasions in an arid island: Evidence from Socotra island (Yemen). *Plant Ecology* 213: 1381-1392.

Slavich, E., Warton, D.I., Ashcroft, M.B., Gollan, J.R. & Ramp, D. 2014. Topoclimate versus macroclimate: how does climate mapping methodology affect species distribution models and climate change projections? *Diversity and Distributions*, 20: 952-963.

Soberón, J. & Peterson, A.T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* 2: 1-10.

Sormunen, H., Virtanen, R. & Luoto, M. 2011. Inclusion of local environmental conditions alters high-latitude vegetation change predictions based on bioclimatic models. *Polar Biology* 34: 883-897.

Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A.C. & Guisan, A. 2014. Measuring the relative effect of factors affecting species distribution model predictions. *Methods in Ecology and Evolution* 5: 947-955.

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., Richardson, D.M., Pysek, P., Midgley, G.F., Hughes, G.O. & Rouget, M. 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology* 11: 2234-2250.

Trabucco, A. & Zomer, R.J. 2010. *Global Soil Water Balance Geospatial Database.* CGIAR Consortium for Spatial Information. Published online, available from the CGIAR-CSI GeoPortal at: http://www.cgiar-csi.org

Tucker, C.M., Rebelo, A.G. & Manne, L.L. 2012. Contribution of disturbance to distribution and abundance in a fire-adapted system. *Ecography* 35: 348-355.

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., de Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Blöschl, G., Eitzinger, J., Steinnocher, K., Zeil, P. & Rubel, F. 2013. The ascat soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift* 22: 5-33.

Wang, L., Gong, W., Ma, Y., Hu, B. & Zhang, M. 2014. Photosynthetically active radiation and its relationship with global solar radiation in Central China. *International Journal of Biometeorology* 58: 1265-77.

Wisz, M.S., Hijmans, R. Li, J., Peterson, A.T., Graham, C.H. & Guisan, A. 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14: 763-773.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., Heikkinen, R.K., Hoye, T.T., Kuhn, I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Ockinger, E., Schmidt, N.M., Termansen, M., Timmermann, A., Wardle, D.A., Aastrup, P. & Svenning, J.C. 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews* 88: 15-30.

Von Holle, B. & Motzkin, G. 2007. Historical land use and environmental determinants of nonnative plant distribution in coastal southern new england. *Biological Conservation* 136: 33-43.

Zimmermann, N. E., Edwards, T. C. Moisen, G. G., Frescino, T. S. &Blackard, J. A. 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *The Journal of Applied Ecology* 44: 1057-1067.Zimmermann, N.E., Edwards, T.C., Graham, C.H., Pearman, P.B. & Svenning, J.C. 2010. New trends in species distribution modelling. *Ecography* 33: 985-989.

Zimmermann, N.E., Yoccoz, N.G., Edwards, T.C., Meier, E.S., Thuiller, W., Guisan, A., Schmatz, D.R. & Pearman, P.B. 2009. Climatic extremes improve predictions of spatial patterns of tree species. *Proceedings of the National Academy of Sciences of the United States of America* 106 Suppl 2: 19723-19728.

Supporting Information

Online Supporting Information may be found in the online version of this article:

Appendix S1 Ecophysiological meaning of different categories of variables for plant species

Appendix S2 Journals and numbers of studies included in the paper.

Appendix S3 Variables included in the different classes and categories

TABLES

Table 1. Classification of predictors into eight categories and 16 classes (see Appendix 3 for details of the variables). The five first columns represent the most important categories, which we refer to as ‘the five most essential categories’ in the text.

Cate-gories	**Temperature**	**Water**	**Substrate**	**Radiation**	**Biotic inter-actions**	**Disturbance**	**Topo-graphy**	**Land use**
Classes	mean (annual, seasonal, monthly) temperature	mean / summed (annual, seasonal, monthly) precipitation	pH, bedrock	radiation, clouds	variables related to other organisms	geomorpho-logical processes, fire	slope, aspect, elevation,	land-use classes
	extreme temperatures	extreme precipitation	nutrients			anthropo-genic variables		
	seasonality	seasonality						
		water balance						
		soil moisture						

FIGURES

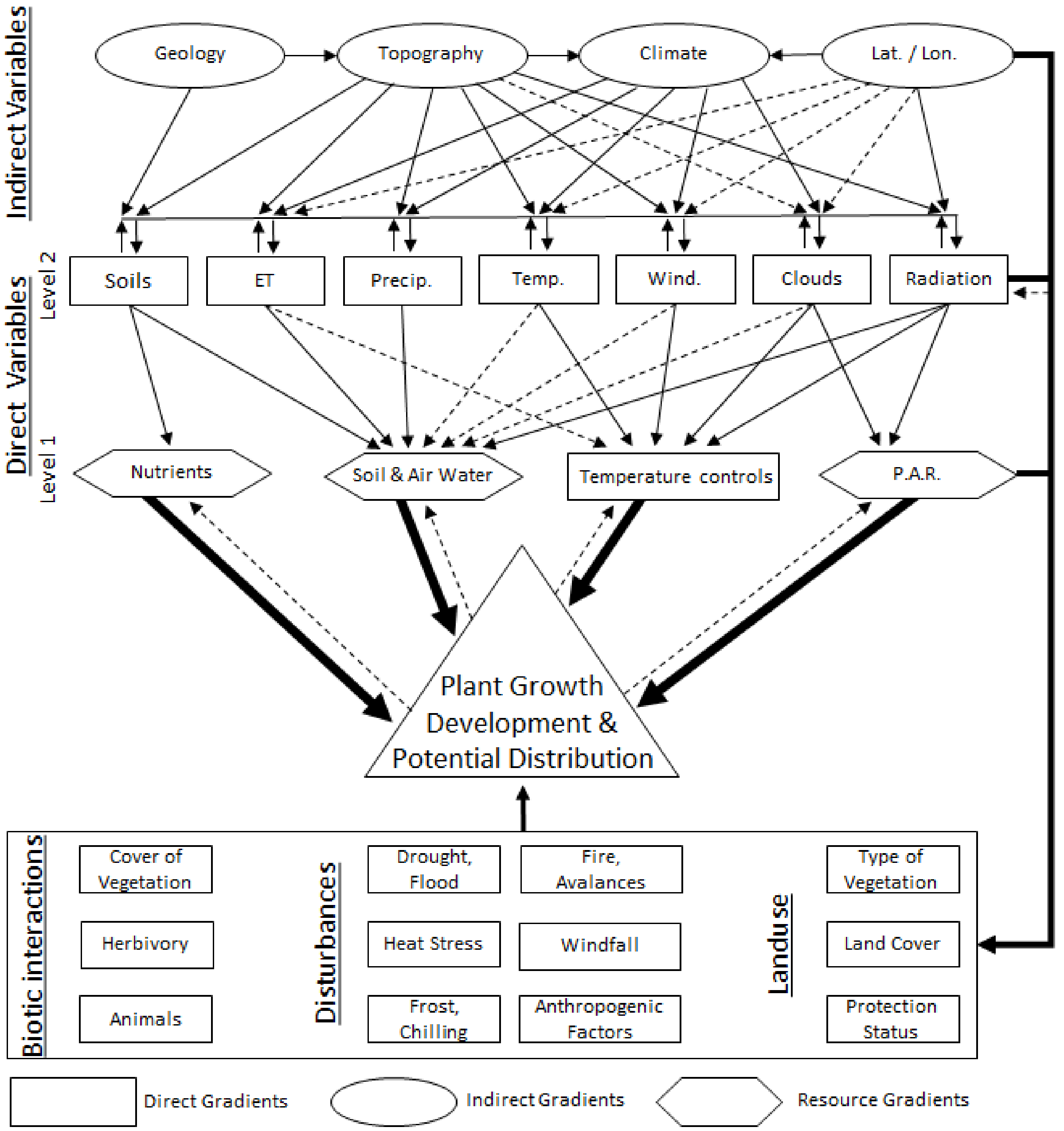

Fig. 1. Example of a conceptual framework of relationships between resources, direct and indirect environmental gradients and their influence on the growth, performance, and geographical distribution of vascular plants and vegetation. ET = Evapotranspiration, P.A.R = Photosynthetically active radiation. Adapted from Guisan & Zimmermann 2000.

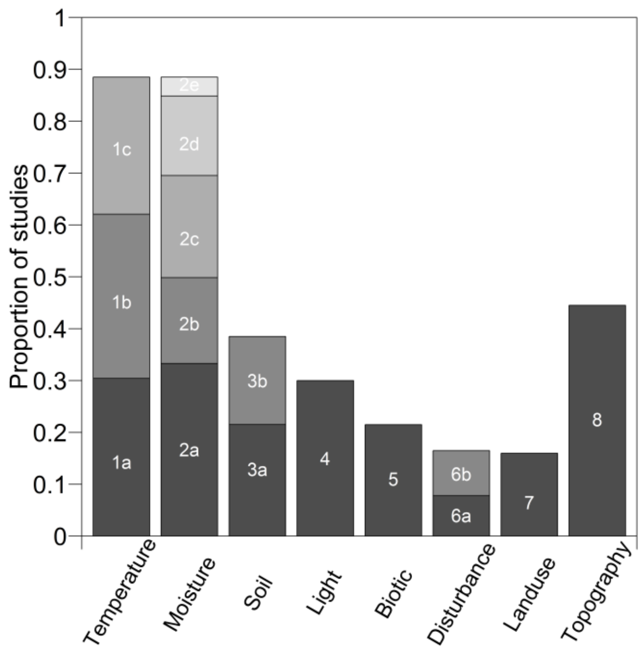

Fig. 3. Proportion of studies in which each predictor class was used: 1a mean temperature; 1b extreme temperature; 1c seasonality of temperature; 2a mean precipitation; 2b extreme precipitation; 2c seasonality of precipitation; 2d water balance; 2e soil moisture; 3a pH/bedrock; 3b nutrients; 4 radiation; 5 biotic interactions; 6a natural disturbances; 6b human disturbances; 7 land use; 8 topography.

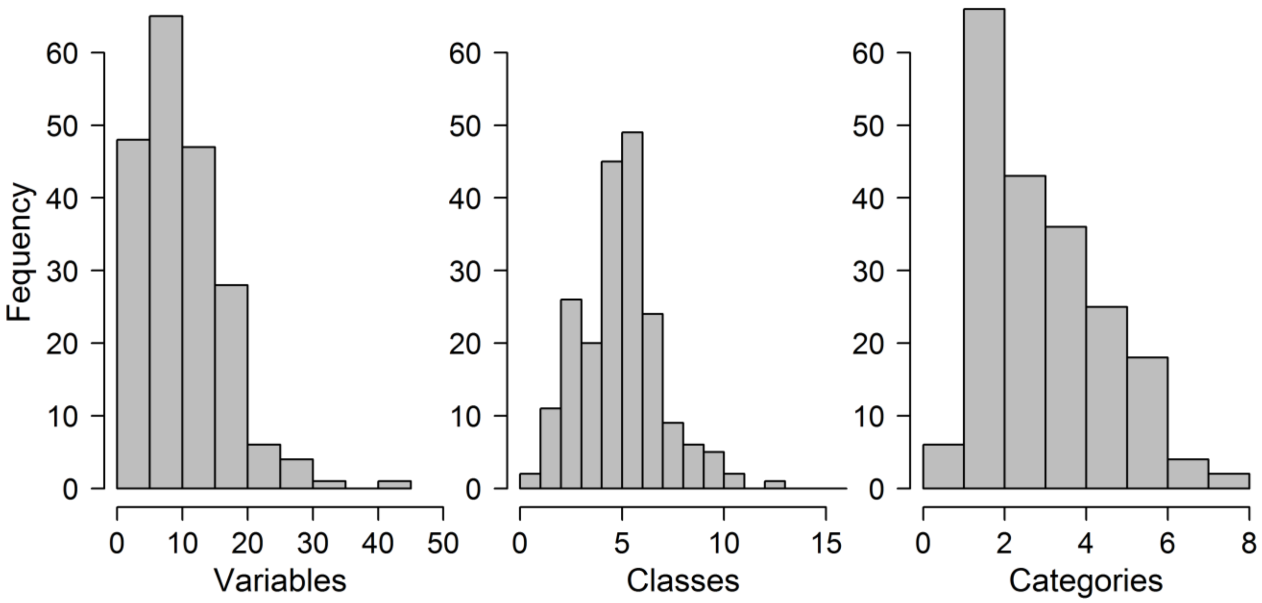

Fig. 2. Frequency of the number of variables, classes (16) and categories (see Table 1) accounted for in the plant species distribution modelling studies. One outlier value (75) was removed from the histogram representing the number of variables in the SDMs.

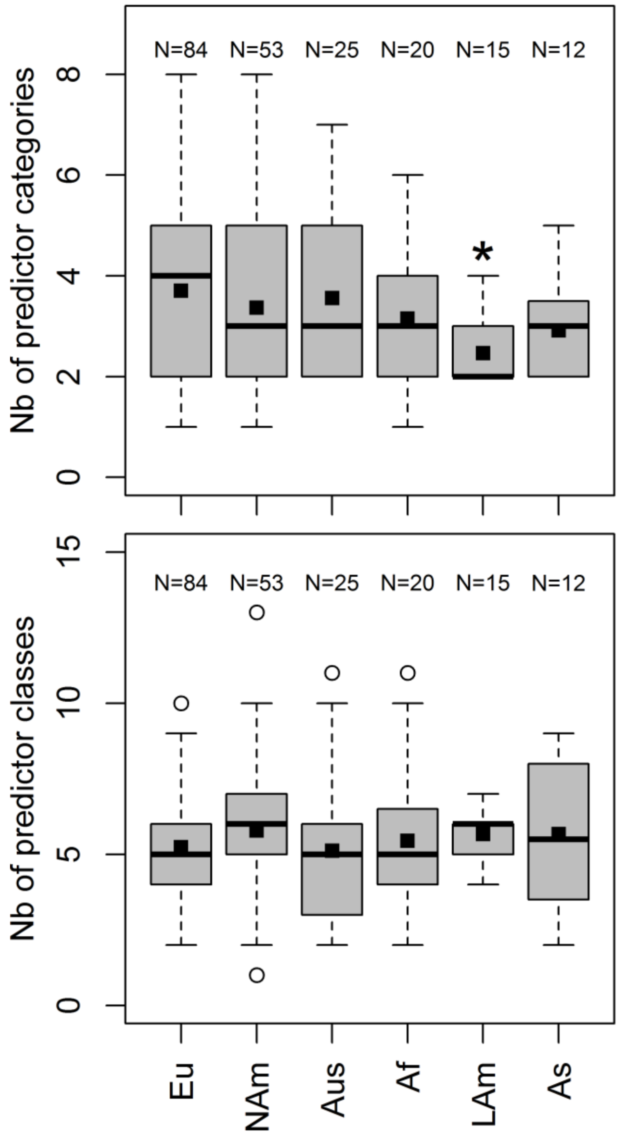

Fig. 4. The number of categories and classes accounted for in the plant species distribution models (SDMs) using data from different continents. The boxes represent the median and the 25/75 percentile, and the whiskers are 2 SD. The mean is indicated by a black square, and significant differences are marked with an asterisk.

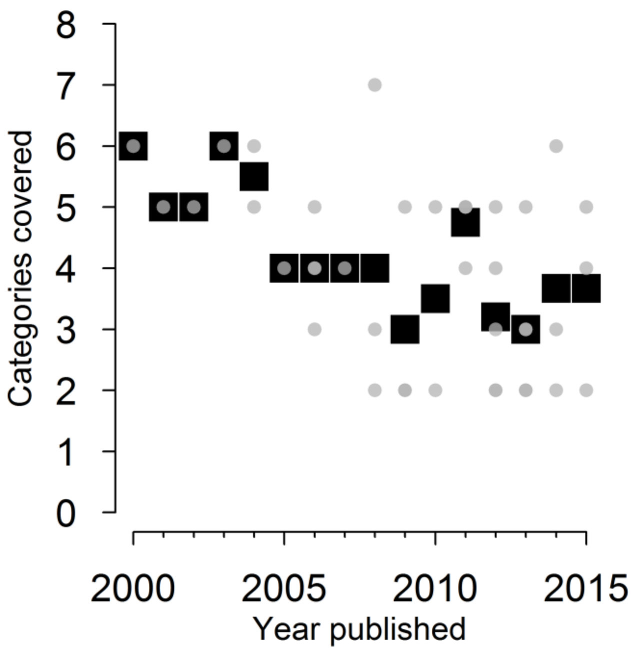

Fig. 5. Number of variable categories (as presented in Table 1) used in the SDM studies published in two journals from 2000-2015. Spearman’s rank correlation between the years and categories included is -0.40*. Black squares indicate the mean values of all studies published within a year, and the grey dots indicate individual studies.

Supporting information to the paper
Heidi K. Mod, Daniel Scherrer, Miska Luoto & Antoine Guisan. What we use is not what we know: environmental predictors in plant distribution models. *Journal of Vegetation Science.*

Appendix S1. Ecophysiologically relevant variables for plant distribution

Seven environmental factors are generally considered as essential for plant growth and survival: light, water, temperature, nutrients, biotic interactions, disturbance and CO_2 (Guisan & Zimmermann, 2000, Kadereit *et al.*, 2014). All these factors can have direct and indirect effects on plants and in combination with dispersal and historical factors, they define the abundance and distribution of plant species (Soberon & Peterson, 2005).

Temperature is the most common regulatory factor considered in SDM's. Temperature directly effects the speed of growth and in case of strong seasonality defines the growing season length. Additionally, minimum and maximum temperatures can reflect physiological thresholds for plants by frost or heat resistance.

Water has several essential functions in plants including photosynthesis, cooling by transpiration and maintaining turgor. In SDMs "water" is usually reflected by either precipitation alone or in combination with evapotranspiration (e.g. water balance). These environmental variables are considered a proxy for plant available water. However, this might not be the case if soils and topography are heterogeneous, as plant available water is strongly influence by both soil type and topographic position. The seasonality of available water/precipitation might lead to temporal flooding, drought or snow cover and thus requires special adaptations by the present plant species.

Nutrients are taken up with water by roots (often with the help of mycorrhiza). Many micronutrients are essential for plant survival including potassium, calcium, magnesium, sulphur, boron, chlorine, manganese, molybdenum and zinc but most significant for productivity are usually the contents of nitrogen and phosphorus. Nutrients in a wider sense can also influence the pH of the soils, whereas bedrock together with living organism are the primary regulators of available nutrients in soils. Therefore, while deriving nutrient content of the soils might not be effective, bedrock, soil pH and soil texture are often used as surrogates in the SDMs.

Light is often expressed as global radiation and therefore energy (W/m2) driving temperature (air, leaf, and soil) and evapotranspiration. However, for plants light reflects also photo active radiation (PAR) and is thereby directly related to photosynthesis. While radiation can be easily modelled and is relatively independent of the vegetation, PAR is strongly affected by the canopy structure of the vegetation. Therefore, the available light for photosynthesis might be very different in a forest compared to open grassland at otherwise similar global radiation (energy). Additionally, light might contain important signals for plant development (e.g. germination and photoperiodism).

Biotic interactions act among and between species, and have both positive and negative impact by prohibiting or ameliorating growth. Impact of other species can be direct (e.g. competition, herbivory) or indirect (e.g. ameliorating harsh microclimatic conditions, shading, nutrient addition by manure). Biotic interactions have been included to the SDMs as e.g. presence or cover of dominant species, remote sensed vegetation index or interaction matrices for multispecies co-occurrence datasets.

Disturbance's impact is mainly negative for species as soil, water, air or snow movement, fire or anthropogenic activities destroy vegetation. However, some ruderal species benefit from disturbances indirectly as they decrease competition and create space by destroying dominant species, and some specialist species require disturbances, as fire and water-logging for germination. Disturbances have also secondary impact on vegetation, by indirectly impacting soil properties: e.g. cryoturbation bring nutrients closer to soil surface.

CO_2 the carbon source for plants and therefore essential for their survival and productivity. However, the levels of CO_2 among sites don't vary enough to be limiting or having a significant influence on species composition and therefore are ignored in correlative models such as SDM's.

Topography and land use do not have a direct impact on plants, but they affect the distribution of ecophysiolosically meaningful factors (e.g. temperature, light). Topography and land use related variables are easily available and incorporating them often improve SDMs.

References:

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

Kadereit, J.W., Körner C., Kost B., Sonnewald U. 2014. *Strasburger - Lehrbuch der Pflanzenwissenschaften,* Berlin, Springer Spektrum.

Soberón, J. & Peterson, A.T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* 2: 1-10.

Appendix S2. Journals and number of studies included by the query (and subsequent analyses).

Recent search:

Ambio 3 (1)

American Naturalist 4 (1)

Annals of Botany 9 (2)

Applied Vegetation Science 9 (5)

Biodiversity and Conservation 22 (5)

Biological Conservation 49 (10)

Biology Letters 2 (2)

Climatic Change 7 (3)

Conservation Biology 9 (1)

Conservation Letters 1 (0)

Diversity and Distribution 62 (19)

Ecography 52 (21)

Ecological Applications 20 (5)

Ecological Modelling 50 (20)

Ecological Monographs 3 (3)

Ecology 9 (2)

Ecology Letters 11 (1)

Ecosystems 1 (0)

Functional Ecology 3 (0)

Global Change Biology 58 (16)

Global Ecology and Biogeography 45 (11)

Journal of Applied Ecology 13 (1)

Journal of Biogeography 59 (17)

Journal of Ecology 14 (2)

Journal of Vegetation Science 24 (8)

Landscape Ecology 12 (1)

Methods in Ecology and Evolution 15 (1)

Nature Communications 1 (0)

New Phytologist 5 (2)

Oecologia 1 (0)

Oikos 5 (1)

Perspectives in Plant Ecology 7 (0)

Plant Ecology 8 (7)

Plos One 113 (29)

Proceedings of National Academy of Sciences 10 (1)

Proceedings of Royal Society B 14 (2)

Science 2 (0)

Trends in Ecology and Evolution 1 (0)

Temporal search

Journal of Vegetation Science 39 (12)

Journal of Biogeography 122 (28)

Appendix S3. Variables included in different classes and categories.

TEMPERATURE

mean temperature

- (annual / monthly) mean temperature (of coldest / warmest / driest / wettest quarter / summer / winter)
- soil temperature
- warmth index (the annual sum of positive differences between monthly mean temperatures and e.g. 5 degrees, i.e. a measure of the effective warmth for plants)

extreme temperature

- (annual) min / max temperature (of coldest / warmest driest / wettest quarter / month / season)
- mean temperature of coldest / warmest / driest / wettest month
- mean daily max / min temperature (for DJF / MAM / JJA / SON)

temperature seasonality

- seasonality, annual / diurnal range
- growing degree days (all thresholds) / freezing degree days (FDD) (soil / air) / non-FDD / chilling degree days
- isothermality
- heat units (annual sum of daily temperatures exceeding X degrees)
- frost duration
- winter / summer cold / heat wave duration

WATER

mean precipitation

- (annual / monthly) mean / summed precipitation (of coldest / warmest / driest / wettest quarter / season)
- days with rain > 1 mm
- rainfall intensity

extreme precipitation

- mean / summed / min / max precipitation of coldest / warmest / driest / wettest month
- highest 5-day precipitation

precipitation seasonality

- seasonality, annual range
- snow (cover duration, annual snowfall)

- dry / wet season /day length / intensity / frequency
- % of annual precipitation falling during the growing season
- average flood duration
- the standard deviation of hydrographs

water-balance

- (annual / seasonal / monthly) water balance
- (annual / seasonal) evapo-transipiration, vapour pressure
- (mean / annual / seasonal / soil) water / moisture deficit / surplus / availability /stress
- (annual / seasonal / plant available) water/ wetness / moisture / aridity index
- water content
- flow accumulation
- average water level
- soil moisture (days; days when soil moisture - air temperature ratio is favourable for plant growth)
- waterlogging index

soil water capacity

- soil water capacity, measured soil moisture
- soil drainage class
- hydraulic soil presence class

SUBSTRATE

bedrock / ph

- bedrock, lithology, rock type
- pH
- surface geology, geological substrate

nutrients

- nutrients, fertility, Cation-exchange capacity, calcareous
- soil material / depth / order / quality / texture / type
- organic matter, loaminess, alluvial, clay / silt / sand content, salt, gypsum
- soil grain size, bulk density
- FAO soil group
- remote sensed Normalized difference soil index, soil production index

- water regime (ordered classes from dry to waterlogged)

LIGHT

- solar radiation (daily, annual, seasonal)
- most / least radiated quarter
- mean hours of sunshine
- clouds

BIOTIC

- NDVI, Landsat bands, Enhanced Vegetation Indices, remote sensed vegetation (indices / classes)
- vegetation height / density / volume/ cover
- canopy / forest / tree cover
- productivity, Net Primary Production
- ecological classification, succession time
- pollinators
- litter
- distance to moorland, moorland presence / absence
- stand basal area
- % of sparsely / dense vegetated brownfield
- % of brownfield with low / high vegetation

DISTURBANCE

natural

- fire, volcanic ash
- geomorphological disturbance
- trampling, grazing
- % area of disturbed terrain

anthropogenic

- population / settlement / building density
- distance to urban areas / roads / harbour / roads
- agriculture, afforestation, soil drainage, roads, human perturbation, forest / etc. management

- human footprint, anthropization degree
- brick rubble
- ownership status (measure of land management)
- predominance of exotic species

TOPOGRAPHY

- altitude (range), terrain curvature, topographic position, slope, flatness, meso-topography, % of steep topography, slope type
- aspect, eastness, northness
- rockiness, ruggedness, topographic wetness index,
- topographic diversity

LAND-USE

- Corine, land-use classes (if only "biotic" land-use -> ‘biotic’ class)
- distance to potential forest, age of forest