Science of the Total Environment 671 (2019) 1086–1093

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Redefining the climate niche of plant species: A novel approach for realistic predictions of species distribution under climate change

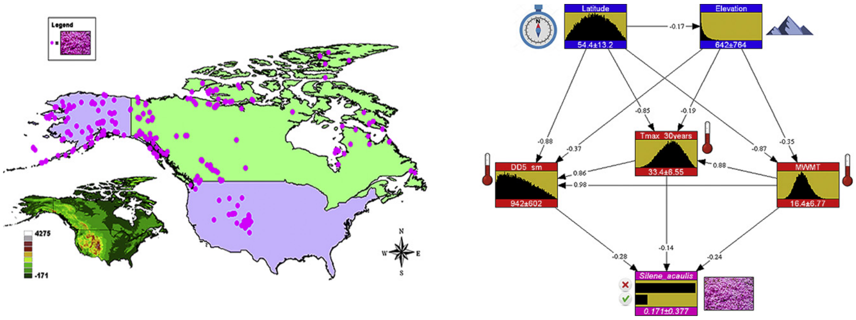
Alessandro Ferrarini [a], Junhu Dai [b], Yang Bai [c], Juha M. Alatalo [d,e,*]

[a] Via G. Saragat 4, I-43123 Parma, Italy
[b] Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
[c] Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla, Yunnan, 666303, China
[d] Department of Biological and Environmental Sciences, College of Arts and Sciences, Qatar University, P.O. Box 2713, Doha, Qatar
[e] Environmental Science Center, Qatar University, P.O. Box: 2713, Doha, Qatar

HIGHLIGHTS

- Climate change is increasingly affecting plant species distributions.
- We developed the relevant climate niche (RCN) of plant species.
- Using RCN, we projected the fate of *Silene acaulis* in North America.
- Within 20 years, >50% of current locations will be outside its climate hypervolume.
- We detected the minimal geographical shifts required for climate change compensation.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 27 January 2019
Received in revised form 21 March 2019
Accepted 22 March 2019
Available online 27 March 2019

Editor: Elena Paoletti

Keywords:
Climate change compensation
Network-like climate niche
Non-parametric Bayesian network
North America
Reverse climate simulations
Silene acaulis L

ABSTRACT

Climate change is increasingly affecting plant species distributions, in ways that need to be predicted. Here, in a novel prediction approach, we developed the relevant climate niche (RCN) of plants, based on thorough selection of climate variables and implementation of a non-parametric Bayesian network for climate simulations. The RCN was conditionalized to project the fate of *Silene acaulis* in North America under moderate (Representative Concentration Pathway 4.5; RCP4.5) and extreme (RCP8.5) short-term (2011–2040) climate scenarios. We identified a three-variable climate hypervolume for *S. acaulis*. Within 20 years >50% of current locations of the species will be outside the defined climate hypervolume. It could compensate for climate change in 2011–2040 through a poleward shift of 0.97 °C or an upshift of 138 m in the RCP4.5 scenario, and 1.29 °C or 184 m in the RCP8.5 scenario. These results demonstrate the benefits of redefining the climate niche of plant species in the form of a user-defined, data-validated, hierarchical network comprising only variables that are consistent with species distribution. Advantages include realism and interpretability in niche modeling, and new opportunities for predicting future species distributions under climate change.

1. Introduction

Climate is the primary factor regulating the geographical distribution of plant species (Davis and Shaw, 2001; Doak and Morris, 2010; Loarie et al., 2009; Pearson and Dawson, 2003). There is already

---

[*] Corresponding author at: Department of Biological and Environmental Sciences, College of Arts and Sciences, Qatar University, P.O. Box 2713, Doha, Qatar.
E-mail addresses: sgtpm@libero.it (A. Ferrarini), jalatalo@qu.edu.qa (J.M. Alatalo).

https://doi.org/10.1016/j.scitotenv.2019.03.353
0048-9697/

evidence that the changing climate of the past century has resulted in a globally consistent fingerprint of poleward and/or upward shifts in species distributions (Parmesan and Yohe, 2003; Root et al., 2003; Walther et al., 2002).

Understanding the climate niches of plant species and projecting their potential future shifts in spatial distributions are therefore necessary to assess the vulnerability of plant species under a rapidly changing climate (Bakkenes et al., 2006; Ferrarini et al., 2019) and to develop adaptive management strategies, including assisted migration (Ferrarini et al., 2016; Isaac-Renton et al., 2014). Climate niches constitute the climate component of Hutchinson's fundamental niche (Hutchinson, 1957; Pearson and Dawson, 2003). As they are built upon the observed distribution of plant species, niche models reflect an actual climate niche (Holt, 2009), as opposed to the fundamental niche, which is based on the environmental requirements of plant species (Hutchinson, 1957). Models of this form are typically referred to as climate niche models, bioclimatic envelope models, or ecological niche models.

The main theoretical assumptions behind climate niche models have widely been discussed (Nogués-Bravo, 2009; Pearman et al., 2008). With regard to the methodological issues, several authors have criticized and proposed solutions to the choice of climate-relevant variables (Ferrarini et al., 2019; Thuiller et al., 2005) and the most appropriate niche model (Ferrarini et al., 2019; Wiens et al., 2009).

We believe that definition of the climate niche of plant species could be further upgraded to offer new desirable information for interpretative and projective purposes. First, it could encompass a user-defined graphical structure to explicitly represent causes, and test assumptions, in a series of conditional relationships that can be easily understood. Second, it could thoroughly and strictly select only those climate variables that actually govern (i.e. most influence) the presence/absence of plant species in a certain region (the n-dimensional climate hypervolume; Hutchinson, 1957), thus excluding all climate variables that add unwanted noise in modeling and simulation work. Third, besides forward climate simulations (i.e., anticipating climate change effects on species locations), it could permit meaningful reverse simulations. This would provide the opportunity to detect a) the climate conditions that can lead to species extinction and b) the magnitude of poleward and elevation shifts that can compensate for climate change.

To this end, we propose a new theoretical and operative definition of the climate niche of plant species, which we call the relevant climate niche (RCN). In this study, we applied it in the case of *Silene acaulis* L. in North America to: 1) project the fate of this species under climate change, 2) detect the minimal amount of climate change that will likely lead to species extinction; and 3) detect the minimal amount of geographical shifts that can compensate for the effects of climate change in the short term (2011–2040).

2. Methods

2.1. Study area and study species

The study area was North America (USA and Canada), which extends over 14,300 km in longitude and 6600 km in latitude (Fig. S1). Locations of *Silene acaulis* (moss campion) were determined from the online Global Biodiversity Information Facility atlas. We chose locations recorded and confirmed in the period 1981–2010, which were consistent with the baseline climate conditions available for North America. Overall, 401 *S. acaulis* locations were available. A digital elevation model (1-km resolution) of the study area was used to assign the elevation (m above sea level, m a.s.l.) to each *S. acaulis* location (Fig. S1).

Silene acaulis L. (Caryophyllaceae) is a herbaceous cushion plant characteristic of alpine and arctic-tundra habitats. It has a circumpolar distribution and is found in Asia, Europe, Greenland, and North America. In Europe, the southernmost populations are found in the Alps, Balkans, British Isles, Carpathian mountains, Cantabrian Mountains, and Pyrenees. Cushion plants such as *S. acaulis* play a key role in alpine ecosystems, where they function as facilitator species and nurse plants (Antonsson et al., 2009). They also play an important function in determining the resilience of alpine and high-latitude ecosystems to global climate change (Reid and Lortie, 2012).

2.2. Creation of pseudo-absence points

We noted a clear relationship between latitude and elevation, which were significantly and inversely correlated, at the 401 *S. acaulis* locations (Fig. S2). However, we found no clear relationship between longitude and elevation, or between longitude and latitude, at the 401 *S. acaulis* locations (Figs. S3 and S4). This suggested that two (latitude and elevation) distal variables out of three influence *S. acaulis* presence in the study area, and that these two variables have a combined effect on this species.

Our strategy for creation of pseudo-absence points had three goals: 1) to systematically sample the study area; 2) to place pseudo-absences with probability directly proportional to their distance from the mean distal (latitude and elevation) attributes of *S. acaulis* presence; and 3) to take into account the variances/covariances of distal variables, which was necessary as they showed a joint effect on *S. acaulis* presence in the study area (Fig. S2).

To fulfill these goals, we used Mahalanobis distance (Mahalanobis, 1936) and random number generation. Mathematically, the Mahalanobis distance was computed as:

$$D^2 = (\underline{x}-\underline{m})^T \mathbf{C}^{-1} (\underline{x}-\underline{m})$$

where $\underline{m}$ is the mean < *latitude*, *elevation* > vector of *S. acaulis* locations in North America, $\underline{x}$ is the <*latitude*, *elevation* > vector of the generic location in the study area, C^{-1} is the inverse covariance matrix of latitude and altitude calculated at *S. acaulis* locations, and the T superscript denotes the transpose operator.

Mahalanobis distance is approximated by the χ^2 distribution with n-1 degrees of freedom ($n = 2$ in our study). Using this approximation, it was possible to assign corresponding p-values to any point in the study area. The p-value reflects the probability of obtaining a Mahalanobis distance as large or larger than the actual Mahalanobis distance. P-values close to 0 reflect high Mahalanobis distance for areas that are therefore very dissimilar to the mean < *latitude*, *elevation* > vector of *S. acaulis* locations in North America, while p-values close to 1 reflect low Mahalanobis distance for areas very similar to the mean < *latitude*, *elevation* > vector of *S. acaulis*.

In order to place pseudo-absence points in the study area, we first generated one sampling point for every degree latitude and longitude. We then assigned a random number in the [0–1] interval to each point, which was considered as pseudo-absence if, and only if, the random number was higher than the Mahalanobis distance p-value associated with that point. Thus, parcels of land with p-values close to 0 (i.e. far distant from the mean vector of *Silene* presence) had a very high probability of becoming pseudo-absence points, while parcels with Mahalanobis distance p-values close to 1 had a very low probability.

2.3. Baseline climate conditions

To represent baseline climate conditions, we used climatological data for 1981–2010, calculated using the ClimateNA v5 package (Wang et al., 2016), which extracts and downscales climate data from PRISM (Daly et al., 2008) and WorldClim (Hijmans et al., 2005). The downscaling is achieved through a combination of bilinear interpolation and dynamic local elevation adjustment. In total, the available climate dataset consisted of 247 climate variables (Table S1) for each of the *S. acaulis* presence/pseudo-absence points.

2.4. Selecting the relevant climate variables

We employed a stringent variable selection procedure to a) reduce the amount of data needed to achieve learning, b) improve predictive accuracy, c) produce more compact and easily understood information, and d) reduce execution time.

To predict *S. acaulis* presence/pseudo-absence, we used the correlation-based feature selector (CFS) method (Hall, 1999), which seeks a subset that contains variables highly correlated with the class, yet uncorrelated with each other. Given a starting set S of K climate variables, the algorithm seeks the variable subset $s \subset S$ that maximizes

$$CFS = k * \bar{r}_1 / \sqrt{k + k * (k-1) * \bar{r}_2}$$

where $k < K$ is the number of variables in the subset s, $\bar{r}_1$ is the average variable-class correlation, and $\bar{r}_2$ is the average variable-variable intercorrelation. The numerator can be considered as providing an indication of the predictive accuracy of a set of predictors (here climate variables) for the class (here *S. acaulis* presence/pseudo-absence). The denominator indicates the level of redundancy among the variables.

Because the CFS algorithm is computationally very demanding, we coupled it with genetic algorithms (Holland, 1992) in order to shorten its search for the best subset of climate variables. We first applied CFS separately to annual, seasonal, and monthly climate variables. We then used a second round of selection, applied to the three (annual, seasonal, and monthly) reduced datasets of the first round, as a means to detect the overall optimal subset of climate variables.

To validate the reduced subset of climate variables, we used six classification algorithms: naïve Bayes (Bayes et al., 1763), classification trees (Breiman et al., 1984), logistic regression (Cox, 1958), MLP neural network (Rumelhart et al., 1985), random forests (Breiman, 2001), and SVM (Cortes and Vapnik, 1995). If the six algorithms accurately classified *S. acaulis* presence/pseudo-absence using the CFS subset as input, the highly reduced dataset of climate variables would be correct, i.e., it would contain the relevant climate drivers of *S. acaulis* in the study area. We measured classification performance using area under the curve (AUC), i.e., the probability that, when a site with the species present and a site with the species absent are drawn at random, the former will have a higher predicted value than the latter. We also used classification accuracy (CA), i.e., the proportion of correct predictions (true positives plus true negatives, divided by the total number of instances). Models were validated using stratified 10-fold cross-validation (out-of-sample datasets used 10 times). Stratification ensured that each class was (approximately) equally represented across each test fold, and was necessary because the two categories to be recognized (*S. acaulis* presence and pseudo-absence) were unbalanced. We also employed kernel density plots (Rosenblatt, 1956) for visual inspection of the discriminative power of the CFS dataset, choosing the bandwidth that minimized the mean integrated squared error (Silverman, 2018).

2.5. Modeling the relevant climate niche (RCN)

The RCN was conceived as an interacting network comprising distal and proximal variables actually ruling the presence/absence of the plant species under study. These variables have statistical (marginal) distributions that can be computed from GIS data. Changes to each variable propagate through the network and produce (direct and indirect) effects on all the other variables. The strength of the effects depends on how much the variables influence each other. These influences can be computed from GIS data as well.

For modeling, we used the non-parametric Bayesian network (NP-BN) methodology (Hanea et al., 2006, 2010). A NP-BN is a directed acyclic graph that combines a graphical structure of the variables involved in a system and the relationships between them. The nodes represent univariate random variables (X_1, X_2, …, X_n) which can be discrete, continuous, or both. The arcs represent direct influences, i.e., cause-effect assumptions articulated in a series of conditional correlations among variables.

When using a NP-BN, an empirical marginal distribution for each variable and a conditional correlation for each arc must be specified (Hanea et al., 2015). Here, the marginal distribution of each variable was computed from GIS data using pre-defined statistical distributions, while the conditional correlation was calculated using the normal copula (Nelsen, 2007). A copula is the joint distribution of n uniform variables. The heart of the copula application in dependence modeling lies in Sklar's theorem (Sklar, 1959), which states that any joint cumulative distribution function (here denoted $F_{1 \dots n}$) of variables $X_1 \dots X_n$ can be rewritten in terms of the corresponding copula C as:

$$F_{1 \dots n}(x_1 \dots x_n) = C(F_1(X_1) \dots F_n(X_n))$$

where $F_i(X_i)$ denotes the marginal distribution of the i-th variable.

The normal copula is expressed as:

$$C_R(u_1 \dots u_n) = \Phi_R\left(\Phi^{-1}(u_1) \dots \Phi^{-1}(u_n)\right)$$

where Φ^{-1} denotes the inverse cumulative distribution of an univariate standard normal distribution, and Φ_R denotes the joint cumulative distribution of a multivariate normal distribution with zero mean and correlation matrix R.

We manually built the NP-BN, using arcs between nodes to indicate potential relationships between variables. We then populated the model with GIS data on *S. acaulis* presence and pseudo-absence points. The model was validated by testing whether joint normal copulas adequately represented the original data by comparing the empirical correlation structures to correlation structures using the normal copula (Hanea et al., 2015). The empirical correlation structures were calculated using Spearman's *rho* correlation coefficient:

$$rho_s = 1 - \left(6 * \sum_{i=1}^{n} d_i^2 / (n^3 - n)\right)$$

where n is the number of *S. acaulis* presence and pseudo-absence points, and d_i is the rank of the i_{th} point in the first variable minus the rank of the i_{th} point in the second variable. The Spearman's *rho* correlation coefficient was deemed more appropriate than Pearson's *r* correlation coefficient, since the variables did not follow a normal distribution.

Climate simulations were performed by conditionalization, i.e., by setting the value of one or more variables to infer how it/they affect the state of other variables. Conditionalizing one (or more) variables allowed the other (unconditionalized) variables to vary, based on the influences defined within the NP-BN. We used the UninetEngine COM library (Cooke et al., 2017) for Matlab.

2.6. Climate projections

The RCN was conditionalized to project the fate of *S. acaulis* in North America under moderate (Representative Concentration Pathway (RCP) 4.5) and extreme (RCP8.5) short-term (2011–2040) climate scenarios. We gathered values (using the ClimateNA v5 package) of only the relevant climate variables that were selected through the filtering procedure described above. We used climate datasets from the recent Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (IPCC, 2013) and employed ensemble projections averaged across 15 Coupled Model Intercomparison Project Phase 5 (CMIP5) models (CanESM2, ACCESS1.0, IPSL-CM5AMR, MIROC5, MPI-ESM-LR, CCSM4, HadGEM2-ES, CNRM-CM5, CSIRO Mk 3.6, GFDL-CM3, INM-CM4, MRI-CGCM3, MIROC-ESM, CESM1-CAM5, GISS-E2R) to account for the uncertainty in the climate predictions.

Besides the standard forward simulation (i.e., anticipating climate change effects on *S. acaulis* locations), the RCN allowed for two kinds of reverse climate simulations, identifying: a) the minimal amount of climate change that likely leads to extinction of *S. acaulis* in North America and b) the minimal geographical (poleward and elevation) shifts that can compensate for the effects of climate change on *S. acaulis*.

3. Results

The procedure used to generate the pseudo-absence points (Fig. 1; Fig. S5) led to the creation of 1938 points (Fig. S6). The Mahalanobis distance from the mean <*latitude*, *elevation*> vector of *S. acaulis* locations in North America was significantly higher (unequal variance independent two-sample $t = -19.94$; $p < 0.001$) for pseudo-absence points (mean ± S.D. 6.74 ± 7.06, $n = 1938$) than for presence points (1.95 ± 1.39, $n = 401$).

The first-round variable selection isolated 20 climate variables (out of 247; 8.09% selection) that strongly influence the presence of *S. acaulis* in North America (Table S2). Nineteen of these 20 variables (all except Hargreaves reference evaporation) relate to temperature. The presence of *S. acaulis* in the study area proved to be highly influenced by maxima of temperature and by degree-days above 5 °C (*DD5*; growing degree-days). Both these variables act at all temporal scales studied, i.e., annual, seasonal, and monthly (Table S2).

The second-round variable selection detected three climate variables (out of 20; 15% selection) that rule *S. acaulis* presence in the study area. These were: summer DD5 (*DD5_sm*), mean warmest month temperature (*MWMT*), and extreme maximum temperature over 30 years (*Tmax_30y*) (Table S2).

The three-variable climate dataset was successfully validated by the six classification algorithms (Table S3). The 10-fold cross-validated accuracy was higher than 0.8 for all algorithms employed, and for both area under the curve (AUC) and classification accuracy (CA).

Visual inspection, through kernel density plots, of the discriminative power of the reduced climate dataset (Figs. S7–S9) confirmed that *S. acaulis* presence-absence in the study area was well discriminated by the selected climate variables. Absence of *S. acaulis* was most common in the study area for *DD5_sm* = (mean ± S.D) 1042.8 °C ± 597.5, *MWMT* = 17.48 °C ± 6.69, and *Tmax_30y* = 34.60 °C ± 6.51. Presence of *S. acaulis* was most common for *DD5_sm* = 388.4 °C ± 189.1, *MWMT* = 10.21 °C ± 2.43, and *Tmax_30y* = 27.08 °C ± 2.66.

The zero-order Spearman's correlation matrix revealed: a) strong negative correlations between latitude and the selected climate variables, b) weaker negative correlations between elevation and climate variables, c) negative correlations between *S. acaulis* presence/absence and climate variables, and d) positive correlations between *S. acaulis* presence/absence and geographical variables (i.e., latitude, elevation) (Fig. 2).

The resulting non-parametric Bayesian network (NP-BN) (Fig. 3) shows marginal distributions fitted from data (Table S4) and conditional correlations among variables calculated using normal copulas. As correlation matrices revealed maximum (element-wise) differences in the order of ±0.06, we concluded that the correlation matrix using the normal copula assumptions provided an adequate approximation of the empirical (Spearman's) correlation matrix (Fig. 2).

The resulting RCN (Fig. 3) presents the distal variables (latitude and elevation) on the first layer (root nodes), the selected climate variables (*DD5_sm*, *MWMT* and *Tmax_30y*) on the second one (parent nodes), and the species under study on the third one (child node). Latitude and elevation resulted negatively correlated, as an increase/decrease in latitude of species locations has to be compensated for by a decrease/increase in elevation, and vice versa, in order to keep suitable climate conditions for the species unchanged. The distal variables resulted negatively correlated with the selected climate variables, as an increase/decrease in the elevation or latitude of species locations is expected to produce a correspondent decrease/increase in the T° values. The selected climate variables resulted to act synergistically, as an increase/decrease in *MWMT* is expected to determine an increase/decrease in *DD5_sm* and *Tmax_30y*, which in turn act upon *S. acaulis* presence/absence. All the climate variables resulted to act negatively upon the species, because *S. acaulis* presence is negatively correlated to T°. However the selected climate variables resulted unequally important. *DD5_sm* showed to be

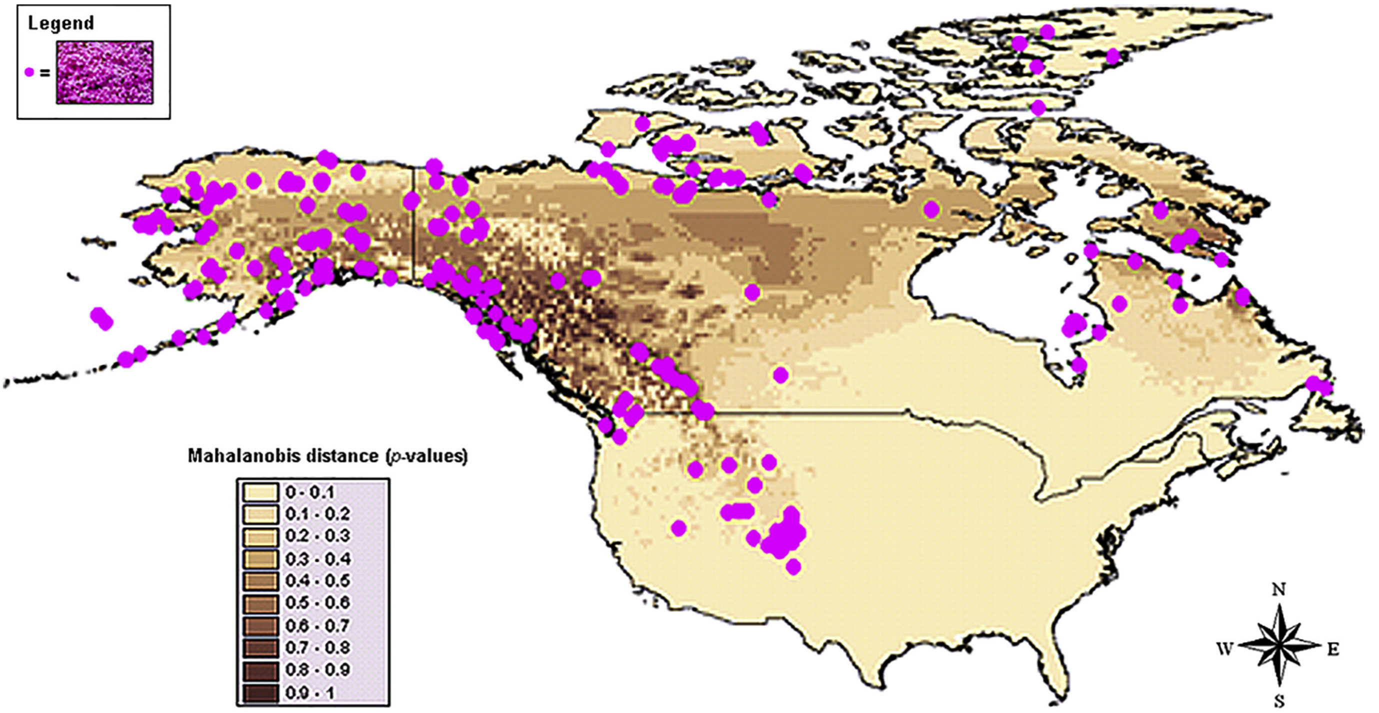


Fig. 1. Probability (p-) values associated with the Mahalanobis distance of each portion of the study area from the mean <*latitude*, *elevation*> vector of *Silene acaulis* locations. Spots in magenta indicate recorded locations of *S. acaulis* in the period 1981–2010.

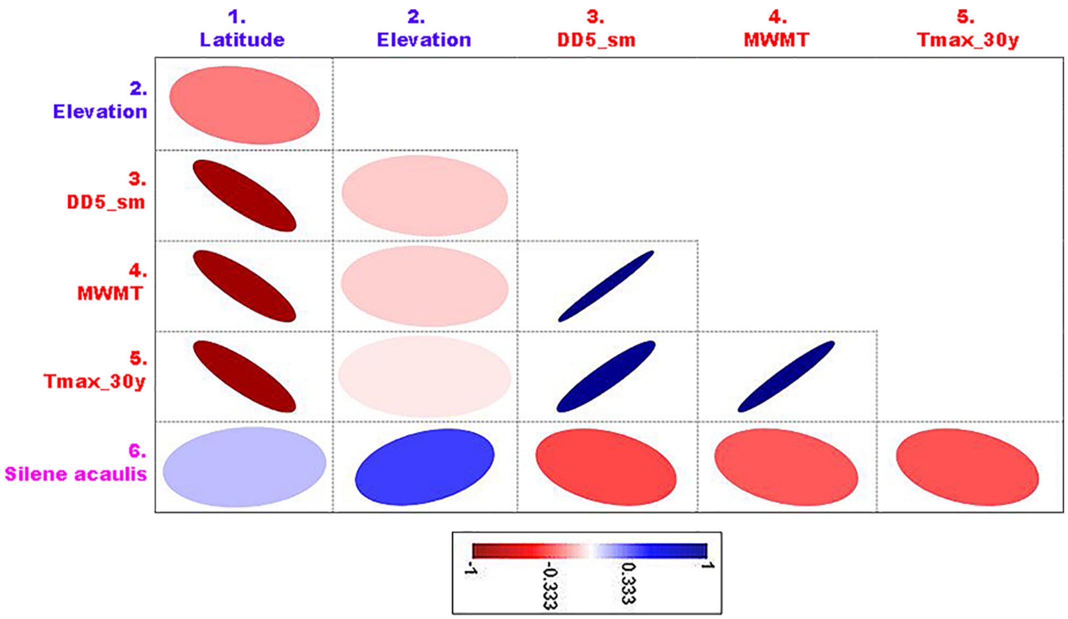


Fig. 2. Correlations (Spearman's *rho*) between distal (geographical) and proximal (climate) variables and *Silene acaulis* presence/absence in North America. Correlations refer to 2339 points (401 presences, 1938 pseudo-absences), with negative correlations in reddish tones and positive correlations in bluish tones. The narrower the ellipse, the higher the correlation between variables. Perfect positive correlations are represented by diagonal lines.

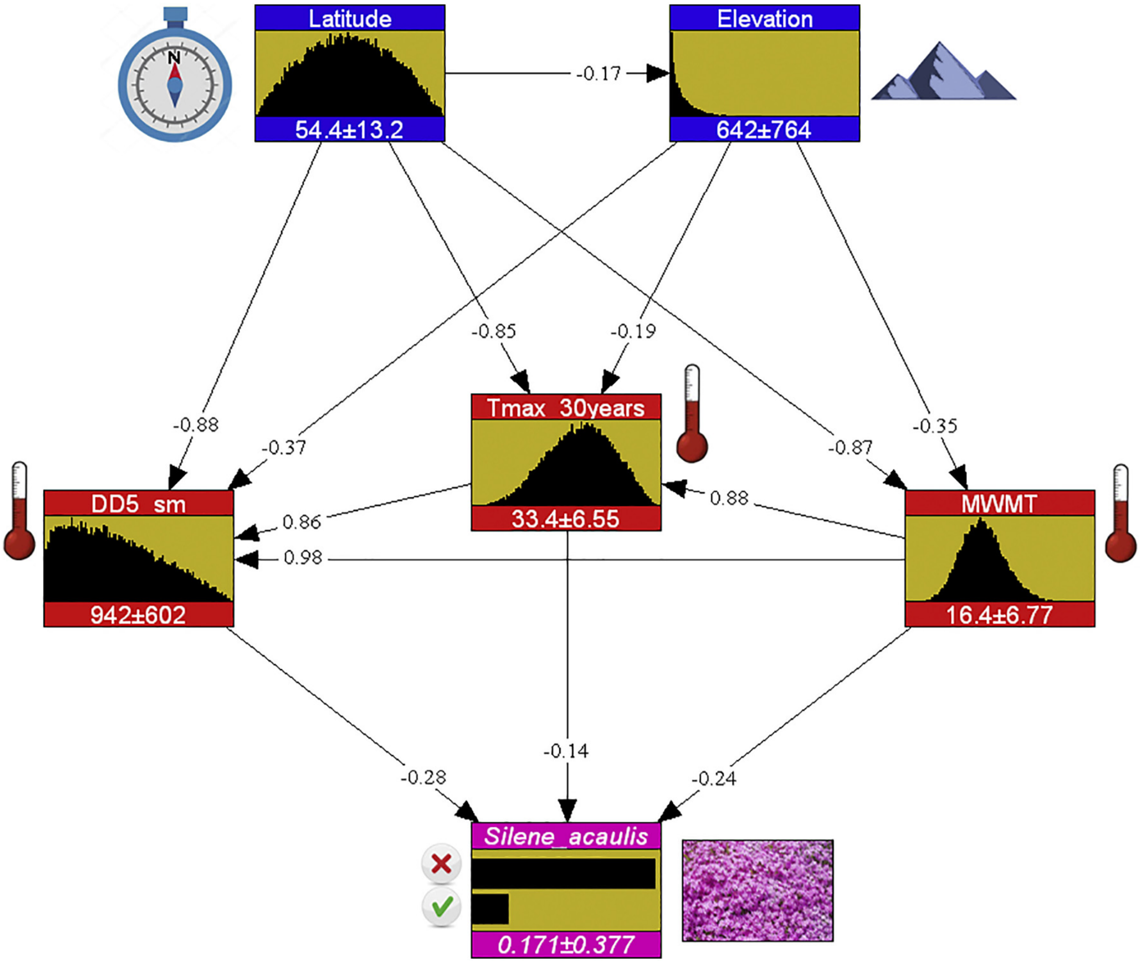


Fig. 3. Relevant climate niche (RCN) of *Silene acaulis* in North America (2339 points; 401 presences and 1938 pseudo-absences). Distal (geographical) variables are in blue, proximal (climate) variables are in red. For each variable, marginal distribution (nodes) and influences (arrows) on the other variables were computed from GIS data. The variable *S. acaulis* is discrete, with presence probability equal to 17.1% (= 401/2339). *DD5_sm* = summer degree-days above 5 °C; *MWMT* = mean warmest month temperature; *Tmax_30years* = maximum temperature over 30 years.

the most influential variable (influence = −0.28), followed by *MWMT* (influence = −0.24) and then *Tmax_30y* (influence = −0.14).

In order to conditionalize the NP-BN for climate projections, we fitted the statistical distributions of the three climate variables for both the moderate and extreme IPCC scenarios (IPCC, 2013) (Tables S5 and S6) and entered the fitted distributions into the NP-BN. Under the moderate climate scenario (RCP4.5), average increments of 97 °C in *DD5_sm*, 1.4 °C in *MWMT*, and 0.9 °C in *Tmax_30y* at the 401 *S. acaulis* occurrence points are predicted (Fig. S10). Under these climate conditions, *S. acaulis* was predicted to lose 56.10% of its occurrence points (i.e., 225 out of 401) due to unfavorable climate conditions. Under the extreme climate scenario (RCP8.5), average increments of 106 °C in *DD5_sm*, 1.5 °C in *MWMT*, and 1.0 °C in *Tmax_30y* at the 401 *S. acaulis* occurrence points are predicted. Under these climate conditions, *S. acaulis* was predicted to lose 60.84% of its occurrence points (i.e., 244 out of 401) due to unfavorable climate conditions (Fig. S10).

The climate conditions that would lead to local extinction of *S. acaulis* in North America were very different from those expected for 2011–2040 (Fig. S11). Even with respect to the extreme scenario (RCP8.5), further increments of 50 °C in *DD5_sm*, 0.9 °C in *MWMT*, and 0.6 °C in *Tmax_30y* would be required. These increments would cause a further 39.16% decrease in *S. acaulis* presence in the study area, after the 60.84% reduction expected due to the RCP8.5 scenario.

In the RCP4.5 scenario, expected climate change in the period 2011–2040 could still be compensated for at the current *S. acaulis* locations by: a) an average increment of 0.97 °C in latitude (with no increment in elevation) or b) an average increment of 138 m in elevation (with no increment in latitude) (Fig. S11). Of course, combined increments in latitude and elevation are also possible (e.g., 100 m in elevation and 0.266° in latitude; or 50 m in elevation and 0.615° in latitude). In the RCP8.5 scenario, the expected climate change in 2011–2040 could be compensated for by: a) an average increment of 1.29 °C in latitude (with no increment in elevation); b) an average increment of 184 m in elevation (with no increment in latitude); or c) combined increments in latitude and elevation (e.g. 100 m in elevation and 0.587° in latitude; or 50 m in elevation and 0.937° in latitude) (Fig. S11).

4. Discussion

Silene acaulis proved to have a three-dimensional climate hypervolume (sensu Hutchinson, 1957). Our simulation results, validated through six classification algorithms and visual inspection of kernel plots, indicated that the spatial distribution of this species in North America is largely influenced by temperature-related variables. Among these, our results demonstrated the crucial importance of temperature extremes and growing-degree days (accumulated warmth) for the species. In contrast, climate factors linked to rainfall, humidity, frost period, chilling days, solar radiation, and evaporation proved to be irrelevant. We expect this outcome to be species-specific, with other plant species influenced by different sets of climate variables.

Our projections suggest that, in the next 20 years, *S. acaulis* will lose >50% of its climatically suitable presence points, even in a moderate climate change scenario. This is in good agreement with findings for the British Isles (Ferrarini et al., 2019), although our projections here were less extreme and also more reliable, because they relate to the exact locations of the species, rather than to 10 km × 10 km squares as in that study. Our results are also in line with previous empirical findings that southern populations of *S. acaulis* in North America have lower survival rates than northern populations, and that mortality increases in the warmest years (Doak and Morris, 2010). Experimental work has shown that, after an initial positive response to experimental warming and nutrient addition, *S. acaulis* has declined over the medium term in sub-arctic Sweden (Alatalo and Little, 2014). Experimental warming has been shown to increase seed production by *S. acaulis*, but this seems to be insufficient to compensate for increased mortality due to increased frequency of warmer years (Alatalo and Totland, 1997). Heat damage has been observed in *S. acaulis* under natural conditions at an air temperature of only 21 °C (Gauslaa, 1984). Several studies have found that the leaf temperature of *S. acaulis* cushions is 15–24.5 °C warmer than the air temperature (Salisbury and Spomer 1964). This heat-trapping feature was later observed to cause the leaf temperature of an *S. acaulis* cushion to reach 45 °C at an Arctic site (Körner and De Moraes, 1979).

The three-variable climate change vector detected in this study (<*DD5_sm* = 156 °C, *MWMT* = 2.4 °C, *Tmax_30y* = 1.6 °C>) represents the set of climate changes that exclude the presence of *S. acaulis* in the study area, i.e., the minimal amount of climate change that likely leads to negative intrinsic growth rate in all current locations. This outcome can thus be interpreted as the climate bounds of the three-dimensional climate hypervolume of this species in North America. However, loss of suitable climate conditions does not necessarily mean immediate disappearance of this species, as *S. acaulis* is long-lived, with individuals potentially living >200 years (Morris and Doak, 1998). Therefore loss of suitable climate conditions will more likely involve a decrease in reproductive success and recruitment, which will bring the species towards local extinction over time. In addition, although evidence of evolution of plant species in response to climate change remains sparse at best (Franks et al., 2007), phenotypic plasticity may temporarily counteract the negative effect of climate change on *S. acaulis*, but may not confer long-term persistence as conditions fall outside its climate bounds (Nicotra et al., 2010).

Under the implicit assumption that the current climate hypervolume limits will not change for future climates, the minimal latitude and elevation shifts identified here suggest that this species requires an upshift velocity of at least 4.5 m yr^{-1} in the moderate scenario and 6.1 m yr^{-1} in the extreme scenario in order to compensate for climate change in the period 2011–2040. These results are in good agreement with predicted compensation velocities of <10 m yr^{-1} on steep slopes (Loarie et al., 2009). Alternatively, compensation can be achieved through a poleward velocity of 0.033° yr^{-1} in the moderate scenario and 0.043° yr^{-1} in the extreme scenario. Combined latitude-elevation shifts are also possible, through a simple linear combination of latitude and elevation shifts. An increment in elevation (with no increment in latitude) is possible in all mountain areas where *S. acaulis* is currently present. A poleward shift (with no increment in elevation) is possible in northernmost areas where *S. acaulis* is present in almost flat areas at sea level.

The question is whether plant species can actually keep up with the predicted velocity of climate change (Loarie et al., 2009). Human-assisted colonization could help plant species in achieving combined latitude-elevation shifts, and also poleward-only or elevation-only compensations, for which the required degree of magnitude can be computed using RCN.

In this study, we used the western hemisphere populations of this circumboreal species for analyses and modeling. The application of RCN to the eastern hemisphere could be useful to further support our results on this species.

4.1. Methodological issues and conclusions

Our relevant climate niche (RCN) of plant species approach involves identifying a network of interacting (distal and proximal) variables that are the relevant drivers of species presence/absence in an area. This network allows the static and flat representation of the zero-degree Spearman's correlations to be disentangled, and the overall influences among variables to be split into (direct and indirect) hierarchical influences. This in turn provides the opportunity to dynamically simulate changes to any variable and predict the direct and indirect effects expected for all the other variables.

The main advantage of this stringent operational definition of the climate niche of plant species is that it makes it possible to: a) formulate a clear hypothesis on the network of relevant variables that actually determine the spatial distribution of a certain species, b) build a custom network-like model to represent this hypothesis, and c) validate or reject the hypothesis by comparing empirical (i.e. data-based) correlation structures with correlation structures of the non-parametric Bayesian network.

The RCN also provides new opportunities for climate predictions, in terms of reverse simulations. Here, we exploited the opportunity to determine: a) the bounds of the plant climate hypervolume, and b) the minimal rate of geographical shifts required for climate change compensation. These outcomes are important in achieving fitter conservation of plant species in the face of climate change.

In addition, while niche modeling is commonly conceived in flat terms (i.e. two variables levels in the form $X_1 \ldots X_n \rightarrow Y$), our definition of a relevant climate niche allows the researcher to build hierarchical niche models with potentially an arbitrary number of levels (i.e. $L_1 \rightarrow L_2 \rightarrow \ldots \rightarrow L_n \rightarrow Y$) and influences among variables. The simplest network structure is the three-levels form used here, with distal variables in the first layer (root nodes), proximal variables in the second (parent nodes) and the study species in the third (child nodes). This network structure follows an easily defensible rationale and resulted in an adequate approximation of the empirical (Spearman's *rho*) correlation matrix. However, in theory RCN can allow for more complex network structures if the candidate network does not properly fit real-world data. For example, plant communities with strong interactive effects among species might partly overcome direct effects induced by climate change (Alatalo et al., 2016; Ferrarini et al., 2017). At the same time, experimental work has shown that competitive interactions may change, and this will likely affect the community assemblies and dominance hierarchies (Alatalo et al., 2014; Baruah et al., 2017). Biotic interactions with other plant species could be included into the RCN as covariates at the same network level of the plant species under study.

The non-parametric Bayesian network (NP-BN) methodology proved to be well-suited to our purposes. Bayesian networks with only discrete random variables (D-BNs) have recently become quite a popular modeling tool (Sander et al., 2017). They can be seen as the simplest case of probabilistic networks, but are associated with considerable drawbacks, such as subjectivity in the discretization of continuous variables and high computational load, which grows exponentially with the number of network variables. In addition, D-BNs cannot handle ecological models in which some or all random variables can take values in a continuous range. The NP-BN used here allowed these serious methodological limitations to be surmounted.

Construction of the RCN for *S. acaulis* in North America required stringent selection of meaningful climate predictors. Climate data and projections are increasingly available, but results about the major climate drivers governing species presence/absence and their degree of influence are still lacking. Moreover, while good data availability represents a substantial advantage for niche modeling purposes, it also leads to serious modeling drawbacks such as redundancy, overfitting, and oversize (Ferrarini et al., 2019; Parolo et al., 2008). In our study, the available climate dataset (247 climate variables) was highly oversized (many climate variables had little predictive power), overfitted (the number of climate predictors was too high), and redundant (many climate predictors were correlated in a significant manner, and not correlated enough with the class to predict). Overfitting lowers model performance (Guyon and Elisseeff, 2003), redundancy causes multicollinearity (Tabachnick and Fidell, 2007), and oversize makes it awkward to build realistic niche models because too many combinations of variables and interactions are possible. Here we employed a solution based on dimensionality reduction, which transformed highly multidimensional data into a reduced climate hypervolume (sensu Hutchinson, 1957) actually governing the presence/absence of *S. acaulis* in North America. Our approach focused on strongly reducing the overall number of variables in input to the RCN, so as to produce an automatic reduction in the overall correlation in the niche model. This reduced model oversize and overfitting by 98.78% (from 247 to only three variables) and model redundancy by 99.98% (from 247^2 to 3^2 correlations, when also considering self-correlations), while simultaneously improving cross-validation accuracy.

This study demonstrates the advantages of redefining the climate niche of plant species in terms of user-defined, data-validated, hierarchical and interactive networks including only variables relevant for species distribution. The parsimony and realism of the niche model, the reliability and interpretability of results, and the new opportunities provided for climate predictions are the main rewards of this new approach.

Acknowledgements

Qatar Petroleum (QUEX-ESC-QP-RD-18/19) supported J.M.A. J.D. was supported by National Key R & D Program of China (2018YFA0606100). We thank two anonymous reviewers for their helpful comments that improved this manuscript.

Statement of authorship

JMA and AF conceived the paper. AF developed the methodology and performed the modeling work. AF and JMA wrote the paper. JD and YB commented on the paper. All authors edited and reviewed the manuscript.

Appendix A. Supplementary data

Supplementary data, tables and figures related to this article can be found at: https://doi.org/10.1016/j.scitotenv.2019.03.353

References

Alatalo, J.M., Little, C.J., 2014. Simulated global change: contrasting short and medium term growth and reproductive responses of a common alpine/Arctic cushion plant to experimental warming and nutrient enhancement. SpringerPlus 3, 157. https://doi.org/10.1186/2193-1801-3-157.

Alatalo, J.M., Totland, Ø., 1997. Response to simulated climatic change in an alpine and subarctic pollen-risk strategist, Silene acaulis. Glob. Chang. Biol. 3, 74–79. https://doi.org/10.1111/j.1365-2486.1997.gcb133.x.

Alatalo, J.M., Little, C.J., Jägerbrand, A.K., Molau, U., 2014. Dominance hierarchies, diversity and species richness of vascular plants in an alpine meadow: contrasting short and medium term responses to simulated global change. PeerJ 2, e406. https://doi.org/10.7717/peerj.406.

Alatalo, J.M., Jägerbrand, A.K., Molau, U., 2016. Impacts of different climate change regimes and extreme climatic events on an alpine meadow community. Sci. Rep. 6, 21720. https://doi.org/10.1038/srep21720.

Antonsson, H., Björk, R.R.G., Molau, U., 2009. Nurse plant effect of the cushion plant Silene acaulis (L.) Jacq. in an alpine environment in the subarctic Scandes, Sweden. Plant Ecolog. Divers. 2, 17–25.

Bakkenes, M., Eickhout, B., Alkemade, R., 2006. Impacts of different climate stabilisation scenarios on plant species in Europe. Glob. Environ. Chang. 16, 19–28.

Baruah, G., Molau, U., Bai, Y., Alatalo, J.M., 2017. Community and species-specific responses of plant traits to 23 years of experimental warming across subarctic tundra plant communities. Sci. Rep. 7, 2571.

Bayes, T., Price, R., Canton, J., 1763. An essay towards solving a problem in the doctrine of chances. Philos. Trans. 24, 370–418.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth & Brooks, Belmont.

Cooke, R.M., Kurowicka, D., Hanea, A.M., Morales, O., Ababei, D.A., Ale, B., Roelen, A., 2017. Continuous/Discrete non-parametric Bayesian belief nets with UNICORN and UNINET. Technical report. TU Delft.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Cox, D.R., 1958. The regression analysis of binary sequences. J. R. Stat. Soc. Ser. B Methodol. 215–242.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. 28, 2031.

Davis, M.B., Shaw, R.G., 2001. Range shifts and adaptive responses to Quaternary climate change. Science 292, 673–679.

Doak, D.F., Morris, W.F., 2010. Demographic compensation and tipping points in climate-induced range shifts. Nature 467, 959–962. https://doi.org/10.1038/nature09439.

Ferrarini, A., Selvaggi, A., Abeli, T., Alatalo, J.M., Orsenigo, S., Gentili, R., Rossi, G., 2016. Planning for assisted colonization of plants in a warming world. Sci. Rep. 6, 28542. https://doi.org/10.1038/srep28542.

Ferrarini, A., Alatalo, J.M., Gervasoni, D., Foggi, B., 2017. Exploring the compass of potential changes induced by climate warming in plant communities. Ecol. Complex. 29, 1–9.

Ferrarini, A., Alsafran, M.H., Dai, J., Alatalo, J.M., 2019. Improving niche projections of plant species under climate change: Silene acaulis on the British Isles as a case study. Clim. Dyn. 52, 1413–1423.

Franks, S.J., Sim, S., Weis, A.E., 2007. Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. Proc. Natl. Acad. Sci. 104, 1278–1282.

Gauslaa, Y., 1984. Heat resistance and energy budget in different Scandinavian plants. Ecography 7, 5–6.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Hall, M.A., 1999. Correlation-based Feature Selection for Machine Learning.

Hanea, A.M., Kurowicka, D., Cooke, R.M., 2006. Hybrid method for quantifying and analyzing Bayesian belief nets. Qual. Reliab. Eng. Int. 22, 709–729.

Hanea, A.M., Kurowicka, D., Cooke, R.M., Ababei, D.A., 2010. Mining and visualising ordinal data with non-parametric continuous BBNs. Comput. Stat. Data Anal. 54, 668–687.

Hanea, A., Napoles, O.M., Ababei, D., 2015. Non-parametric Bayesian networks: improving theory and reviewing applications. Reliab. Eng. Syst. Saf. 144, 265–284.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., others, 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978.

Holland, J.H., 1992. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press.

Holt, R.D., 2009. Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. Proc. Natl. Acad. Sci. 106, 19659–19665.

Hutchinson, G.E., 1957. Concluding remarks. Cold Spring Harb. Symp. Quant. Biol. 22, 415–427.

IPCC, 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

Isaac-Renton, M.G., Roberts, D.R., Hamann, A., Spiecker, H., 2014. Douglas-fir plantations in Europe: a retrospective test of assisted migration to address climate change. Glob. Chang. Biol. 20, 2607–2617.

Körner, C., De Moraes, J., 1979. Water potential and diffusion resistance in alpine cushion plants on clear summer days. Oecologia Plantarum 14, 109–120.

Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B., Ackerly, D.D., 2009. The velocity of climate change. Nature 462, 1052.

Mahalanobis, P.C., 1936. On the generalized distance in statistics. Proceedings of the National Institute of Sciences (Calcutta). vol. 2, pp. 49–55.

Morris, W.F., Doak, D.F., 1998. Life history of the long-lived gynodioecious cushion plant Silene acaulis (Caryophyllaceae), inferred from size-based population projection matrices. Am. J. Bot. 85, 784.

Nelsen, R.B., 2007. An Introduction to Copulas. Springer Science & Business Media, New York.

Nicotra, A.B., Atkin, O.K., Bonser, S.P., Davidson, A.M., Finnegan, E.J., Mathesius, U., Poot, P., Purugganan, M.D., Richards, C.L., Valladares, F., 2010. Plant phenotypic plasticity in a changing climate. Trends Plant Sci. 15, 684–692.

Nogués-Bravo, D., 2009. Predicting the past distribution of species climatic niches. Glob. Ecol. Biogeogr. 18, 521–531.

Parmesan, C., Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421, 37–42.

Parolo, G., Rossi, G., Ferrarini, A., 2008. Toward improved species niche modelling: Arnica Montana in the Alps as a case study. J. Appl. Ecol. 45, 1410–1418.

Pearman, P.B., Guisan, A., Broennimann, O., Randin, C.F., 2008. Niche dynamics in space and time. Trends Ecol. Evol. 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? Glob. Ecol. Biogeogr. 12, 361–371.

Reid, A.M., Lortie, C.J., 2012. Cushion plants are foundation species with positive effects extending to higher trophic levels. Ecosphere 3, art96. https://doi.org/10.1890/1ES12-00106.1.

Root, T.L., Price, J.T., Hall, K.R., Schneider, S.H., Rosenzweig, C., Pounds, J.A., 2003. Fingerprints of global warming on wild animals and plants. Nature 421, 57–60.

Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. Ann. Math. Stat. 832–837.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1985. Learning Internal Representations by Error Propagation. California Univ San Diego La Jolla Inst for Cognitive Science.

Sander, E.L., Wootton, J.T., Allesina, S., 2017. Ecological network inference from long-term presence-absence data. Sci. Rep. 7, 7154.

Silverman, B.W., 2018. Density Estimation for Statistics and Data Analysis. Routledge.

Sklar, M., 1959. Fonctions de repartition an dimensions et leurs marges. Publ. inst. statist. univ. Paris 8, 229–231.

Tabachnick, B.G., Fidell, L.S., 2007. Using Multivariate Statistics. Allyn & Bacon/Pearson Education.

Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T., Prentice, I.C., 2005. Climate change threats to plant diversity in Europe. PNAS 102, 8245–8250. https://doi.org/10.1073/pnas.0409902102.

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J., Fromentin, J.-M., Hoegh-Guldberg, O., Bairlein, F., 2002. Ecological responses to recent climate change. Nature 416, 389.

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, e0156720.

Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. Proc. Natl. Acad. Sci. 106, 19729–19736.