A new method for broad-scale modeling and projection of plant assemblages under climatic, biotic, and environmental cofiltering

Alessandro Ferrarini\textsuperscript{1} | Yang Bai\textsuperscript{2} | Junhu Dai PhD\textsuperscript{3,4,5} | Juha M. Alatalo\textsuperscript{6}

\textsuperscript{1} Via G. Saragat 4, Parma, Italy
\textsuperscript{2} Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Xishuangbanna 666303, China
\textsuperscript{3} Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China (Email: daijh@igsnrr.ac.cn)
\textsuperscript{4} University of Chinese Academy of Sciences, Beijing, China
\textsuperscript{5} China-Pakistan Joint Research Center on Earth Sciences, CAS-HEC, Islamabad, Pakistan
\textsuperscript{6} Environmental Science Center, Qatar University, Doha, Qatar

Correspondence
Juha M. Alatalo, Environmental Science Center, Qatar University, Doha. Email: jalatalo@qu.edu.qa

Article impact statement: Broad-scale analysis and modeling of plant assemblages under climatic-biotic-environmental cofiltering identifies species for early conservation.

Funding information
Qatar Petroleum; Qatar National Library; National Key R & D Program of China; National Natural Science Foundation of China

Abstract
There is increasing interest in broad-scale analysis, modeling, and prediction of the distribution and composition of plant species assemblages under climatic, environmental, and biotic change, particularly for conservation purposes. We devised a method to reliably predict the impact of climate change on large assemblages of plant communities, while also considering competing biotic and environmental factors. To this purpose, we first used multilabel algorithms in order to convert the task of explaining a large assemblage of plant communities into a classification framework able to capture with high cross-validated accuracy the pattern of species distributions under a composite set of biotic and abiotic factors. We applied our model to a large set of plant communities in the Swiss Alps. Our model explained presences and absences of 175 plant species in 608 plots with >87% cross-validated accuracy, predicted decreases in $\alpha$, $\beta$, and $\gamma$ diversity by 2040 under both moderate and extreme climate scenarios, and identified likely advantaged and disadvantaged plant species under climate change. Multilabel variable selection revealed the overriding importance of topography, soils, and temperature extremes (rather than averages) in determining the distribution of plant species in the study area and their response to climate change. Our method addressed a number of challenging research problems, such as scaling to large numbers of species, considering species relationships and rarity, and addressing an overwhelming proportion of absences in presence–absence matrices. By handling hundreds to thousands of plants and plots simultaneously over large areas, our method can inform broad-scale conservation of plant species under climate change because it allows species that require urgent conservation action (assisted migration, seed conservation, and ex situ conservation) to be detected and prioritized. Our method also increases the practicality of assisted colonization of plant species by helping to prevent ill-advised introduction of plant species with limited future survival probability.

KEYWORDS
classifier chains, climate change, multilabel modeling, plant species conservation, rare species, Swiss Alps

Un Método Nuevo para el Modelado y Proyección a Gran Escala de Ensamblados de Plantas Bajo Cofiltreo Climático, Biótico y Ambiental

Resumen: Hay un creciente interés en el análisis, modelado y predicción a gran escala de la distribución y composición de ensambles de especies de plantas bajo cambio climático, ambiental y biótico, particularmente con fines de conservación. Diseñamos un método para predecir confiablemente el impacto del cambio climático sobre ensambles grandes de comunidades de plantas, considerando también factores bióticos y ambientales. Para ello, primero utilizamos algoritmos multi-etiqueta para convertir la tarea de explicar un
ensamble grande de comunidades vegetales en un marco de clasificación capaz de capturar, con alta precisión validada, el patrón de distribuciones de especies bajo un conjunto de factores bióticos y abióticos. Aplicamos nuestro modelo a un conjunto grande de comunidades vegetales de los Alpes Suizos. Nuestro modelo explicó la presencia y ausencia de 175 especies de plantas en 608 parcelas con una precisión validada de >87%, predijo disminuciones de diversidad α, β, y γ en 2040 bajo escenarios tanto moderados como extremos e identificó especies de plantas potencialmente ventajosas y desventajosas por el cambio climático. La selección de variables multi-ética reveló la importancia primordial de la topografía, suelos y extremos de temperatura (en lugar de los promedios) para determinar la distribución de especies de plantas en el área de estudio y su respuesta al cambio climático. Nuestro método abordó un número de problemas de investigación desafiantes, tal como escalar a un número mayor de especies, considerar las relaciones entre especies y rareza, y atender una proporción enorme de ausencias en las matrices de presencia-ausencia. Al manejar cientos o miles de plantas y parcelas simultáneamente, nuestro método puede informar la conservación a gran escala de especies de plantas bajo el cambio climático porque permite detectar y priorizar especies que requieren acción urgente de conservación (migración asistida, conservación de semillas, conservación ex situ). Nuestro método también incrementa la utilidad de la colonización asistida de especies de plantas al ayudar a prevenir la introducción mal aconsejada de especies de plantas con una probabilidad de supervivencia futura limitada.

PALABRAS CLAVE
Cadenas de clasificación, cambió climático, conservación de especies de plantas, modelado multi-ética

**INTRODUCTION**

Many studies about the potential effects of climate change on vegetation have been performed at the species level, often under the assumption that species respond to climate change in individual ways (Baselga & Araújo, 2009). However, biotic interactions may influence the wide-scale effect of climate on species distributions (Kissling et al., 2012). Complex feed-backs among species suggest that species-specific predictions of climate-change impacts are not necessarily applicable to the overall plant community (Suttle et al., 2007). For example, competition tends to be severe under low environmental stress (Meier et al., 2011), and there is evidence that species interactions can shift from facilitation in cold environments to competition in warm environments (Callaway et al., 2002).
A more complete understanding of climate-change impacts on vegetation requires scaling from individual species to the entire interaction network (McCann, 2007). This could potentially be accomplished at the community level (Franklin et al., 2016), but producing a climate-species predictive model at this level would require precise estimates of the strength of species interactions. An approach in which all the necessary data (species coverages, intraspecific parameters, and colonization coefficients) are known with no or little uncertainty is feasible for small plant communities, but application to large communities seems almost impracticable (Ferrarini et al., 2017a).

Plant species are also influenced by soil type, topography, and human disturbance. Edaphic factors can compensate locally for climate factors and influence the response of plants to climate change (Kardol et al., 2010). Steeper slopes may act as barriers to upward dispersal of plants (Ferrarini et al., 2014), and acclivity also influences soil wetness, erosion, wind impacts, solar input, and snow persistence (Theurillat & Guisan, 2001). Land cover is another important covariate of plant species presences and absences (Luoto et al., 2007).

Accordingly, several modeling techniques have been proposed for predicting the distribution and composition of species assemblages under environmental, climatic, and biotic filtering. The apparent and logical solution is to overlay individual species distributions to obtain assemblage predictions, in so-called stacked-species distribution modeling (S-SDM) (Dubuis et al., 2011). Guisan and Rahbek (2011) proposed a refinement of S-SDM (SESAM) that starts with predictions from S-SDM and then progressively refines these predictions by applying macroecological models, dispersal filters, and ecological assembly rules. The joint species distribution modeling approach (J-SDM) (Clark et al., 2014) incorporates species co-occurrence data into S-SDM by simultaneously estimating the species–environment relationship of multiple species and the residual correlation among those species that could be indicative of biotic interactions.

However, these existing approaches have several weaknesses (for an elevated number of species) (Zurell et al., 2019). Therefore, we devised a novel method: broad-scale analysis and modeling of plant assemblages under climatic-biotic-environmental cofiltering (BAM-PACC). This method permits broad-scale modeling of assemblages of plant communities and handling of data that stem from heterogeneous, autonomous sources, such as digital environmental maps (topography, soils, and land cover), current and projected climate scenarios, and species occurrence and interaction data, to predict the impact of climate change while considering the complex interplay of biotic and abiotic variables. The method BAM-PACC makes use of machine-learning algorithms for multilabel variable selection and multilabel classification and is designed to address a number of the challenging research problems mentioned above.

We applied BAM-PACC to a large assemblage of plant communities in the Swiss Alps, tested the degree to which it could explain current species presences and absences, projected species distributions under future climate scenarios, and considered how BAM-PACC can overcome some difficulties inherent to species- and community-level climate-change projections.
**METHODS**

**Study area and plot data**

Canton de Vaud (46°10′–46°30′N, 6°50′–7°10′E) is in the western Swiss Alps. It covers approximately 700 km² and has a steep elevation gradient from 406 to 2486 m asl (Figure 1).

The study area was exhaustively inventoried in 912 plots of 4 m² from 2002 to 2009 (Dubuis et al., 2011). We chose a subset of 608 plots for which soil data were also available. The average distance between plots was 454 m (SD 304), and the largest distance between plots was 40,859 m (two plots).

We organized the plant data set as a binary presence–absence matrix with 175 columns (species [Appendix S1]) and 608 rows (plots). A 1 indicated presence and a 0 absence of a particular species in a plot.

**Analytical phase**

Within geographic information system, we assigned 93 (topographical, soil, land-cover, and climate) variables to each plot, one each for slope aspect, slope degree, land-cover class, nutrient storage capacity, soil permeability, waterlogging, stone content, root penetration depth, and water storage capacity, and 84 climate variables (Appendix S2).

A GoogleEarth image of the study area was used to derive an accurate digital terrain model (DTM). We considered 1 point every 10 m in both the x and y directions, and calculated slope aspect and slope degree for each position in the DTM by taking local derivatives of elevation in the x and y directions.

A land-cover map of the study area (updated to 2012) was obtained from the website of the Swiss Federal Institute for Forest, Snow and Landscape Research (2019). Soil maps of the study area were obtained from the mapping platform of the Swiss Confederation (Swiss Geoportal, 2019). Nutrient storage capacity specifies how many cation equivalents can be stored in the soil. Permeability was measured using saturated soil samples of the least permeable horizon in the top 50 cm of soil. Waterlogging indicates whether external water runoff is present in the soil. Stone content refers to mineral components larger than 2 mm in the top 50 cm of soil. Root penetration depth measures the depth of soil that can be penetrated by roots. Water storage capacity refers to water retained in the soil that is thus easily available to plants.

To represent baseline climate conditions (84 climate variables for each plot), we used climatological data for 1991–2009, calculated through the ClimateEU version 4.63 package, which applies the PRISM method for precipitation and ANUSPLIN for temperature to latitude, longitude, and elevation of each plot location.

We analyzed distribution of plant plots in different topographical, soil, and land cover classes; species richness in different topographical and land cover categories; and the most common plant associations in different elevation ranges.

**Modeling phase**

The initial matrix had 608 plots in the rows, 93 predictor variables, and 175 plant species in the columns. We connected the plant species to the explanatory variables as \( P = f(T, S, L, C, B) \), where \( P \) is the (608 rows × 175 columns) matrix of species presences and absences, \( T \) is the (608 × 2) matrix of topographical features (slope aspect and slope degree), \( S \) is the (608 × 6) matrix of soil features, \( L \) is the (608 × 1) column matrix of land-cover classes, \( C \) is the (608 × 84) matrix of climate variables, \( B \) is the (175 × 175) matrix of biotic interactions among the plant species, and \( f \) is the mathematical function linking \( P \) to the explanatory variables in the 608 plots. Our goal was to identify \( f \) in order to explain accurately, and subsequently project forward, species presences and absences (Appendix S3).

The available set of predictor variables was most likely oversized (one or more variables may have had little predictive power), overfitted (number of predictors may have been too high), and redundant (some predictors may have been correlated in a significant manner, but not correlated strongly enough with the set of plant species to be predicted). Accordingly, we employed the ReliefF-ML algorithm (Reyes et al., 2015), a recent machine-learning approach that performs variable ranking on data sets with an arbitrary number of explanatory and dependent variables. ReliefF-ML starts with a vector \( W \) of 0s for all the predictors. Within the data set, continuous \( X_j \) are scaled to the interval [0, 1]. At each iteration, ReliefF-ML takes one random plot \( k \) and the weight \( W_j \) associated with \( X_j \) is updated as:

\[
\Delta W_j = -(x_{jk} - x_{j\mu})^2 + (x_{jk} - x_{jd})^2,
\]

where \( x_{jk} \) is the value assumed by \( X_j \) at plot \( k \), \( x_{j\mu} \) is the average value of \( X_j \) for \( \epsilon \) plots (\( \epsilon = 10 \) in our case study) of the same class (i.e., with at least 80% of plant species in common) with the closest value of \( X_j \), and \( x_{jd} \) is the average value of \( X_j \) at the \( \epsilon \) plots of different class (i.e., with <20% of plant species in common) with the closest value of \( X_j \). For categorical variables (e.g., soil features), ReliefF-ML sets \( (x_{jk} - x_{j\mu}) = d_1 \), where \( d_1 \) is the proportion of the \( \epsilon \) plots of the same class with different categories with respect to \( x_{jk} \), and set \( (x_{jk} - x_{jd}) = d_2 \), where \( d_2 \) is the proportion of the \( \epsilon \) plots of different classes with different categories with respect to \( x_{jk} \).

After \( p \) iterations (\( p = 400 \) in our case study), ReliefF-ML divides \( W_j \) by \( p \), thus assigning a weight ranging in the interval [-1, 1]. A high value of \( W_j/p \) indicates that \( X_j \) has different values on plots with dissimilar values of species presences and absences and similar values on plots with similar presences and absences (i.e., elevated discriminating power). Low weight for \( X_j \) indicates the opposite (i.e., low discriminating power). This procedure was repeated for all 93 predictor variables. Based on \( W_j/p \) weights, the predictor variables were ranked from most to least discriminating.

To build an accurate \( f \), we employed a machine-learning solution in which the independent explanatory variables can
be used to explain an arbitrary number of binary dependent variables (multilabel classification). We used the ensembles of classifier chains (ECC) (Read et al., 2011), an algorithm that transforms multilabel classification into a chain of single-label classifications (Appendix S4). Using ECC, we incorporated co-occurrences among species in the classification process by means of chains of plant species as predictors for other species as follows:

\[
P_i = g_i(X_1 \ldots X_m) \\
P_s = g_s(X_1 \ldots X_m, P_1) \\
\vdots \\
P_p = g_p(X_1 \ldots X_m, P_1 \ldots P_{p-1})
\]

where \( P_i \) (\( 1 \leq i \leq m \); \( n = 175 \) in our case) is the column vector of presences and absences of the \( i \)th species in the 608 plots, \( X_j \) (\( 1 \leq j \leq m \); \( m = 93 \) in our case) is the column vector of the \( j \)th explanatory variable in the 608 plots, and \( g \) (\( 1 \leq i \leq n \)) is the mathematical function that links \( P_i \) to its predictors. The vector \( g = \langle g_1 \ldots g_n \rangle \) is the classifier chain and represents the reduced function \( f_R \) described above. Each mathematical function \( g \) belonging to the vector \( g \) was tested as a base classifier. After several attempts with different functions, we chose the multi-perceptron neural network as the base classifier (Appendix S4).

To solve problems linked to the order of the species in the data set, we used ECC to train several classifier chains with random order of species on a random subset of the data set. After several trial-and-error attempts, we opted for \( 10^{12} \) classifier chains with random order of species. For each classifier chain, we used a random 80% of subsets of the 608 plots to speed up calculations. Each species was forced to be the dependent variable in at least \( 10^2 \) classifier chains and \( 10^3 \) classifier chains in the case of rare species (i.e., present in 10% of plots or fewer) because rare species are known to be more difficult to classify (Zurell et al., 2019). Species presence or absence in each plot was predicted by each classifier chain separately. The total number of predictions (0 or 1) was counted, and the most common was chosen (majority rule) for each species.

To retain only the predictor variables actually governing species presences and absences in the assemblage of plant communities, we applied ECC to the most relevant predictor variable (i.e., that with the highest ReliefF-ML score) and then added one variable at a time following the ranking order. We stopped this stepwise filtering procedure when the ECC classification accuracy decreased when further predictor variables were added. This filtering procedure allowed us to obtain the reduced model \( P \approx f_R(T_R, S_R, L_R, C_R, B) \), where \( T_R \) is the reduced matrix of topographical features, \( S_R \) is the reduced matrix of soil features, \( L_R \) is the reduced matrix of land cover classes, \( C_R \) is the reduced matrix of climate variables, and \( f_R \) is the reduced function linking \( P \) to the reduced set of explanatory variables.

We measured model predictivity through overall accuracy (OA), sensitivity (SE), specificity (SP), true skill statistic (TSS), and balanced accuracy (BA) (Appendix S4). We used stratified 10-fold cross-validation because stratification ensured that each class (presence or absence) was approximately equally represented for each species across each test fold, which was necessary because presences and absences were unbalanced for many species.

**Simulation phase**

Once the ECC was built, we used it to project the fate of the 175 plant species in each plot under moderate (RCP4.5) and extreme (RCP8.5) 2011–2040 climate scenarios as \( \Delta P = f_R(T, S, L, \Delta C_R, B) \), where \( \Delta P \) is the projected change (from presences to absences, from absences to presences, and unchanged) in the matrix of species presences or absences and \( \Delta C_R \) is the change in the reduced matrix of climate variables under climate scenarios.

We gathered values only for the relevant climate variables selected through the variable selection procedure above. We used climate data sets from the recent Intergovernmental Panel on Climate Change Fifth Assessment Report (2013) and employed ensemble projections averaged across 15 CMIP5 models (Appendix S5) to account for the uncertainty in the climate predictions.

We compared \( \alpha \) diversity (average number of species), \( \beta \) diversity (difference in species composition among plots), and \( \gamma \) diversity (total number of species) for the current and projected assemblage of plant communities under the RCP4.5 and RCP8.5 2011–2040 climate scenarios. We computed \( \beta \) diversity with the standard equation by Routledge (1977) (Appendix S4).

**RESULTS**

**Broad-scale analyses**

The 608 plots were unevenly distributed with respect to all the explanatory variables (sample vs. expected \( \chi^2 \) test, \( p < 0.01 \)). The plots had predominantly northward slope aspect, slope degree from 20° to 30°, and herbaceous land cover (Appendix S6). The prevailing nutrient storage capacity of soils was very low, and soil permeability was average. Root penetration depth was generally very superficial, soil was very stony, waterlogging was low, and water storage capacity was average (Appendix S6).

Only seven plant species were present in at least 50% of the plots, and nine were present in 10 or fewer plots. On average, the 608 plots had 24.9 (SD 9.9) plant species, and 604 plots out of 608 (i.e., 99.34%) showed a unique combination of species. Species richness was unevenly distributed with respect to elevation, slope, slope aspect, and land cover (Appendix S7).

The plots contained common plant associations (Appendix S8); 38 pairs of species co-occurred in at least 33.3% of the 608 plots. The most common pairwise aggregations were between *Alchemilla vulgaris* and *Trifolium pratense* and between *A. vulgaris* and *Festuca rubra*. Below 1000 asl, the most common plant association was between *Trifolium repens* and *Dactylis glomerata* (75.4% of plots). Between 1000 and 2000 m asl, the most common co-occurrence was between *A. vulgaris* and *T. pratense* (58.4% of plots). Above 2000 m asl, *Polygonum viviparum* and *Campanula scheuchzeri* co-occurred in 59.0% of plots.
Broad-scale modeling

The variable selection procedure selected 23 meaningful variables out of 93 (75.26% reduction). The topographical variables were the most influential variables, followed by four temperature-related variables and three soil-related variables (Figure 2).

Using these variables, the 175 plant species were classified with average cross-validated accuracy of 87.04%, and BAM-PACC scored OA as 86.63%, SE as 80.29%, SP as 87.69%, TSS as 67.98%, and BA as 83.99% (Appendix S9); thus, accuracy was not biased due to the prevalence of absences in the presences–absences matrix (91,275 absences and 15,125 presences). Average cross-validated accuracy decreased by 19.58% when biotic interactions were not included in the model. Accuracy ranged from 60.4% for *Lotus corniculatus* to 100% for *Poa minor*. Accuracies >90% were achieved for 75 species out of 175. An additional 64 plant species were classified with cross-validated accuracy of 80–90% (Appendix S10). Classification accuracies showed a quadratic decay going from more rare to more common species (Appendix S11), and for the nine plant species present in 10 or fewer plots, accuracies ranged from 98.1% (*Salix reticulata*) to 100% (*P. minor*) (Appendices S12 & S13).

Broad-scale projections

In both the RCP4.5 and RCP8.5 scenarios, the 23-variable BAM-PACC simulations predicted a decrease in $\alpha$, $\beta$, and $\gamma$ diversity by 2040 (Figure 3). Most species (142/175; 81.14%) were predicted to be disadvantaged (i.e., to have the number of suitable plots decrease) in the moderate climate scenario, whereas 18.86% of species were predicted to acquire a greater number of suitable plots. Under the extreme climate scenario, these values were 84.30% and 15.70%, respectively.

The BAM-PACC predicted the most advantaged and most disadvantaged plant species by 2040 (Appendix S14). *Cirsium spinosissimum* was predicted to have a gain-loss balance in suitable plots of +187 in the RCP4.5 scenario and +189 in RCP8.5. *Cerastium fontanum* was predicted to be the most disadvantaged species, with a balance of −176 suitable plots in both scenarios. For the 10 plant species currently most frequent in the study area (Appendix S15), different patterns of projected change emerged. Four species (*F. rubra*, *Anthoxanthum odoratum*, *L. corniculatus*, and *Leontodon hispidus*) were predicted to decrease continuously in a number of suitable plots. *Ranunculus acris* was projected to lose suitable plots to the same degree in both scenarios, *Agrostis capillaris* only in the RCP8.5 scenario. Two species (*T. pratense* and *D. glomerata*) were predicted to have an increase in the number of suitable plots, and two species (*A. vulgaris* and *T. repens*) to maintain the same number of suitable plots. For the 10 plant species currently least frequent in the study area (Appendix S15), two patterns emerged. Nine species were projected to lose all suitable plots, whereas *Saxifraga moschata* was predicted to gain a considerable number of suitable plots.

DISCUSSION

At continental and planetary scale, the spatial distributions of plant species are mainly determined by climate factors...
At finer scales, climate factors can be overridden by local factors, such as pedology, topography, and land cover (Parolo et al., 2008). Therefore, it is important to extend research on the effects of climate change on plants by adding biotic and abiotic factors that can heavily strengthen or dampen these effects (Lawler et al., 2015). Accordingly, our method simultaneously considered the complex interplay that results from competing effects of biotic and abiotic variables.

The variable-selection phase revealed the importance of topography and soils in determining the distribution of plant species in the study area, and their response to climate change. Variable selection also showed the overriding importance of temperature extremes, rather than averages. This has been reported for individual plant species (Ferrarini et al., 2019a; Ferrarini et al., 2019b), but was demonstrated here for an ensemble of 175 interacting plant species, in some cases occurring in very different biotic and abiotic conditions. This is further evidence that climate extremes are more important than the commonly used monthly or annual climate means in determining plant distributions, even when considered in combination with other biotic and abiotic predictors.

The results suggested that the assemblage of plant communities in the study area will likely undergo a general decrease in plant diversity by 2040; most species were disadvantaged by climate change and only limited subset of species were advantaged. Because BAM-PACC captured the joint effects of biotic and abiotic factors, our results should be interpreted as the outcome of the overall set of drivers influencing species presences and absences. For instance, a species could be highly favored by climate change but if it is strictly linked to a soil type (or land cover class) that is very uncommon in the study area, BAM-PACC will not predict significant gains in the number of suitable plots. For a species highly disadvantaged by climate change, if all other predictors are favorable and the relevant climate predictors do not exceed its climate hypervolume, BAM-PACC will not predict significant losses in suitable plots.

To date, species distributions have mostly been modeled while ignoring the effects of biotic interactions (Baselga & Araujo, 2010). We detected frequent co-occurrences among the plant species in the study area. Patterns of species aggregation may reflect positive biotic interactions, such as mutualism and commensalism or shared environmental requirements and historical driving forces, whereas negative species associations may
reflect negative biotic interactions or distinctive environmental niches and historical factors (Boulangeat et al., 2012). Whatever the reasons determining plant associations, BAM-PACC allowed inclusion of this information in climate projections.

Methodological issues

The BAM-PACC framework introduced the use of multilabel algorithms for prediction of climate change impacts on plant species assemblages. Unlike single-label classification, multilabel algorithms learn from a set of examples associated with an arbitrary number of labels (presences and absences of plant species). This enabled us to convert the task of explaining a large assemblage of plant communities into a classification framework able to capture with high cross-validated accuracy the pattern of species distributions under a composite set of biotic and abiotic factors.

The BAM-PACC overcomes a number of challenging research problems, such as exploiting species relationships, scaling to large number of species and plots, and dealing with species rarity. With regard to species relationships, BAM-PACC differs from S-SDM in that the latter approach is based on the assumption that species respond individually to variation in environmental conditions (Guisan & Rahbek, 2011). The J-SDM predicts the occurrences of all species in a community simultaneously and includes a covariance structure to disentangle species co-occurrence patterns into shared environmental responses and residual correlations (Clark et al., 2014). This could be expected to be superior in predicting composition and richness for those species assemblages for which biotic interactions are a strong driver of local coexistence (Wisz et al., 2013). However, the residual correlations could be explained by many nonbiological reasons, such as missing environmental variables or poor model fit, rather than reflecting ecological or evolutionary processes (Zurell et al., 2019). This could be one reason that J-SDM does not necessarily improve species assemblage predictions compared with S-SDM (Zurell et al., 2019). The BAM-PACC adopts a different approach to incorporate statistical co-occurrence in species distribution modeling, based on the use of some species as predictors for others (Meier et al., 2011). This seems particularly appropriate when some species play disproportionately large roles in the lives of others (Kissling et al., 2012). However, in large species assemblages, this leads to the problem of multiple testing, which has been counteracted so far by including only the most abundant species as predictors. The BAM-PACC overcomes this issue by using a chain of single-label classifications involving all the species as predictors for others.

With regard to species rarity, accurate models for rare species may not be feasible without borrowing information from other species (Guisan & Rahbek, 2011). Our method minimizes this problem; in fact, classification accuracies for the nine plant species present in 10 or fewer plots were close to 100% and were not biased by high prevalence of absences. Use of a chain of single-label classifications involving all the species as predictors for others could be the reason for this relevant result in that ECC intensively exploits the label interdependencies to improve classification accuracies. A recent study by Zurell et al. (2019) does not support the idea that J-SDM generally yields improved predictions for rare species.

In our projections, we simulated changes in the relevant climate variables. However, soil and land-cover properties could also change as a result of climate change. For example, the climate-constrained tree limit is expected to shift upslope rapidly under climate change, replacing herbaceous open areas above the treeline (Alatalo & Ferrarini, 2017; Ferrarini et al., 2017b). The BAM-PACC allows climate projections to be generalized by simulating changes to land cover and soil properties as \( \Delta P = f_R(T, \Delta S, \Delta L, \Delta C, B) \), where \( \Delta S_R \) represents expected changes in the reduced matrix of soil properties and \( \Delta L_R \) represents expected changes in the reduced matrix of land-cover categories.

We applied BAM-PACC to species occurrences (i.e., binary dependent variables). The J-SDM can also be applied to continuous or discrete species abundances. However, multilabel learning is closely related to multitarget regression, which aims at predicting multiple real-valued target variables instead of binary variables. Thus, BAM-PACC could be readily applied to abundance data (counts, ordinal abundances, proportions of coverage, biomass, basal area, etc.), and ECC could be replaced with the similar ensemble of regressor chains (ERC) algorithm (Borchani et al., 2015). Additional discussion of methodological issues is given in Appendix S16.

Management implications

Ongoing climate change has placed many plant species worldwide at an elevated risk of extinction. Conserving at-risk species is complicated by the urgent need to identify and prioritize those species from among hundreds or thousands that require early conservation actions (e.g., assisted migration, seed conservation, conservation in botanical gardens, and herbaria). This task requires a decision framework capable of handling hundreds to thousands of plants and plots simultaneously over large areas and cannot be accomplished when plant species are studied individually. Our integrative approach, which combines in a single predictive framework most of the driving forces that compete to determine species distributions, provides support for broad-scale conservation of plant species under climate change, in particular in the presence of a large number of species, common plant interactions, numerous rare species, and overwhelming proportion of absences in the presence–absence matrix.

Species conservation increasingly involves interventionist forms of conservation, such as assisted colonization (Ferrarini et al., 2016), to reverse local and global extinctions (McLachlan et al., 2007; Seddon et al., 2007). However, the scale at which thousands of plant species would have to be moved to notice-
ably counteract climate change, and the costs of doing so, are not negligible. The BAM-PACC can increase the practicality of assisted colonization of plant species by helping to prevent ill-advised introduction of plant species whose survival probability is limited in a certain area. Therefore, BAM-PACC can offer a useful solution to preserve plant species in the face of climate change, based on largely available data on coarse-scale environmental variables, plot surveys, and climate-change scenarios.

ACKNOWLEDGMENTS

We are grateful to the handling editor and two anonymous reviewers for their valuable comments, which improved this manuscript. J.M.A. was supported by Qatar Petroleum (QUEX-CAS-QP-RD-18/19). J.D. was funded by National Key R & D Program of China (grant 2018YFA0606100) and National Natural Science Foundation of China (grant 41771056). Open access funding was provided by Qatar National Library.

ORCID

Alessandro Ferrarini https://orcid.org/0000-0001-6290-1988
Yang Bai https://orcid.org/0000-0002-0079-3552
Junhu Dai PhD https://orcid.org/0000-0001-6401-1735
Juha M. Alatalo https://orcid.org/0000-0001-5084-850X

LITERATURE CITED

Alatalo, J. M., & Ferrarini, A. (2017). Braking effect of climate and topography on global change-induced upslope forest expansion. International Journal of Biometeorology, 61, 541–548.

Baselga, A., & Araújo, M. B. (2009). Individualistic vs community modelling of species distributions under climate change. Ecography, 32, 55–65.

Baselga, A., & Araújo, M. B. (2010). Do community models fail to project community variation effectively? Journal of Biogeography, 37, 1842–1850.

Borchani, H., Varando, G., Bielza, C., & Larrañaga, P. (2015). A survey on multi-output regression. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 5, 216–233.

Boulangeat, I., Gravel, D., & Thuiller, W. (2012). Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. Ecology Letters, 15, 584–593.

Callaway, R. M., Brooker, R. W., Choler, P., Kikvidze, Z., Lortie, C. J., Michalet, R., Paolino, 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.

Clark, J. S., Gelfand, A. E., Woodall, C. W., & Zhu, K. (2014). More than the sum of the parts: Forest climate response from joint species distribution models. Ecological Applications, 24, 990–999.

Davies, T. J., Barraclough, T. G., Savolainen, V., & Chase, M. W. (2004). Environmental causes for plant biodiversity gradients. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359, 1645–1656.

Dubuis, A., Potier, J., Rion, V., Pellissier, L., Theurel, J.-P., & Guitan, A. (2011). Predicting spatial patterns of plant species richness: A comparison of direct macroecological and species stacking modelling approaches. Diversity and Distributions, 17, 1122–1131.

Ferrarini, A., Alatalo, J. M., Gervasoni, D., & Foggi, B. (2017a). Exploring the compass of potential changes induced by climate warming in plant communities. Ecological Complexity, 29, 1–9.

Ferrarini, A., Alatalo, J. M., & Gustin, M. (2017b). Climate change will seriously impact bird species dwelling above the treeline: A prospective study for the Italian Alps. Science of the Total Environment, 590, 686–694.

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

Ferrarini, A., Dai, J., Bai, Y., & Alatalo, J. M. (2019b). Redefining the climate niche of plant species: A novel approach for realistic predictions of species distribution under climate change. Science of the Total Environment, 671, 1086–1093.

Ferrarini, A., Rossi, G., Mondoni, A., & Orsenigo, S. (2014). Prediction of climate warming impacts on plant species could be more complex than expected. Evidence from a case study in the Himalaya. Ecological Complexity, 20, 307–314.

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. Scientific Reports, 6, 28542.

Franklin, J., Serra-Díaz, J. M., Syphard, A. D., & Regan, H. M. (2016). Global change and terrestrial plant community dynamics. Proceedings of the National Academy of Sciences of the United States of America, 113, 3725–3734.

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.

Intergovernmental Panel on Climate Change (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: Cambridge University Press.

Kardol, P., Cregger, M. A., Campany, C. E., & Classen, A. T. (2010). Soil ecosystem functioning under climate change: Plant species and community effects. Ecology, 91, 767–781.

Kissing, W. D., Dormann, C. F., Groenewald, J., Hickler, T., Kuehn, I., McNerny, G. J., Montoya, J. M., Römermann, C., Schiffer, 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.

Lawler, J. J., Ackert, D. D., Albano, C. M., AndersonMG, D. S. Z., Gill, J. L., Heller, N. E., Pressey, R. L., Sanderson, E. W., & Weiss, S. B. (2015). The theory behind, and the challenges of, conserving nature’s stage in a time of rapid change. Conservation Biology, 29, 618–629.

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

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

McCann, K. (2007). Protecting biostriucture. Nature, 446, 29.

Meier, E. S., Thomas, C. E. Jr., 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.

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

Read, J., Pfähringer, B., Holmes, G., & Frank, E. (2011). Classifier chains for multi-label classification. Machine Learning, 85, 333–359.

Reyes, O., Morell, C., & Ventura, S. (2015). Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context. Neurocomputing, 161, 168–182.

Roulston, R. D. (1977). On Whitaker’s components of diversity. Ecology, 38, 1120–1127.

Seddon, P. J., Armstrong, D. P., & Maloney, R. F. (2007). Developing the science of reintroduction biology. Conservation Biology, 21, 303–312.

Suttle, K. B., Thomsen, M. A., & Power, M. E. (2007). Species interactions reverse grassland responses to changing climate. Science, 315, 640–642.
Swiss Geoportal. (2019). Retrieved from https://map.geo.admin.ch
The Swiss Federal Institute for Forest, Snow and Landscape Research. (2019).
Swiss Land Cover Mapping. Retrieved from https://www.wsl.ch/en.html
Theurillat, J.-P., & Guisan, A. (2001). Potential impact of climate change on vegetation in the European Alps: A review. *Climatic Change*, 50, 77–109.
Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., Forchhammer, M. C., Grytnes, J. A., Guisan, A., Heikkinen, R. K., Høye, T. T., Kühn, I., Luoto, M., Maier, L., Nilsson, M. C., Normand, S., Öckinger, E., Schmidt, N. M., Termansen, M. . . . , & 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.
Zurell, D., Zimmermann, N. E., Gross, H., Baltensweiler, A., Sattler, T., & Wüest, R. O. (2019). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, 47, 101–113.

**SUPPORTING INFORMATION**
Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Ferrarini, A, Bai, Y, Dai, J, Alatalo, JM. (2022). A new method for broad-scale modeling and projection of plant assemblages under climatic, biotic, and environmental cofiltering. *Conservation Biology*. 2021;36:e13797.
https://doi.org/10.1111/cobi.13797