

2016

Do Community-Level Models Account for the Effects of Biotic Interactions? A Comparison of Community-Level and Species Distribution Modeling of Rocky Mountain Conifers

Paige E. Copenhaver-Parry
George Fox University, pparry@georgefox.edu

Shannon Albeke

Daniel B. Tinker

Follow this and additional works at: http://digitalcommons.georgefox.edu/bio_fac

 Part of the [Biology Commons](#)

Recommended Citation

Originally published in *Plant Ecology*, 2016, 217, 533-547. Posted with permission. See it here: <https://link.springer.com/article/10.1007/s11258-016-0598-5>

This Article is brought to you for free and open access by the Department of Biology and Chemistry at Digital Commons @ George Fox University. It has been accepted for inclusion in Faculty Publications - Department of Biology and Chemistry by an authorized administrator of Digital Commons @ George Fox University. For more information, please contact arolfe@georgefox.edu.

Do community-level models account for the effects of biotic interactions? A comparison of community-level and species distribution modeling of Rocky Mountain conifers

Paige E. Copenhaver-Parry · Shannon E. Albeke · Daniel B. Tinker

Abstract Community-level models (CLMs) aim to improve species distribution modeling (SDM) methods by attempting to explicitly incorporate the influences of interacting species. However, the ability of CLMs to appropriately account for biotic interactions is unclear. We applied CLM and SDM methods to predict the distributions of three dominant conifer tree species in the U.S. Rocky Mountains and compared CLM and SDM predictive accuracy as well as the ability of each approach to accurately reproduce species co-occurrence patterns. We specifically evaluated the performance of two statistical algorithms, *MARS* and *CForest*, within both CLM and SDM frameworks. Across all species, differences in SDM and CLM predictive accuracy were slight and can be attributed to differences in model structure rather than accounting for the effects of biotic interactions. In addition, CLMs generally over-predicted species co-occurrence, while SDMs under-predicted co-

occurrence. Our results demonstrate no real improvement in the ability of CLMs to account for biotic interactions relative to SDMs. We conclude that alternative modeling approaches are needed in order to accurately account for the effects of biotic interactions on species distributions.

Keywords Conditional random forests · Co-occurrence · Douglas-fir · Lodgepole pine · Multivariate adaptive regression splines · Ponderosa pine

Introduction

There is increasing evidence that species rarely occur in complete equilibrium with climate (e.g., Araujo et al. 2005a; Worth et al. 2014; Blois et al. 2014). Historical factors, dispersal, and biotic interactions have all been found to exert substantial constraints on species distributions and range movement of a variety of species (Araujo and Luoto 2007; Leathwick 2009; Meier et al. 2010; Boulangeat et al. 2012; Meineri et al. 2012; HilleRisLambers et al. 2013; Blois et al. 2014). Consequently, many recent developments in species distribution modeling have focused on incorporating non-climatic factors and community- and population-level processes into distribution predictions (e.g., Meier et al. 2010; Boulangeat et al. 2012; Kissling et al. 2012; Clark et al. 2014; Normand et al.

2014). These efforts are particularly important when the intended application of a model is to project to new environments or future climate scenarios where climate equilibrium assumptions are likely to break down (Klanderud and Totland 2005; Suttle et al. 2007; Araújo and Luoto 2007; Swab et al. 2015).

Biotic interactions among species, in particular, have been the subject of significant focus. Both positive and negative interactions may affect species distributions by either inhibiting or facilitating establishment, individual growth, and population growth (Holt 2009; HilleRisLambers et al. 2013; Svenning et al. 2014). At a macroecological scale, the effects of such interactions may be manifested as non-random species co-occurrence patterns (Araújo et al. 2011; Blois et al. 2014; Clark et al. 2014). The information from such co-occurrence patterns may be formally incorporated into models and modeling frameworks that predict the distributions of multiple species. One such group of models, referred to as community-level models (CLMs), are intended to improve the performance of species distribution models (SDMs) using co-occurrence as a proxy for biotic interactions (Ferrier and Guisan 2006). CLM strategies may first combine distributions into community types, which are then predicted by the model ('assemble first, predict later'), or they may predict species independently and use a variety of ad hoc methods to combine SDM predictions ('predict first, assemble later'). Alternatively, the influences of interacting species can be accounted for directly within the modeling framework by predicting species distributions simultaneously ('assemble and predict together') (Ferrier and Guisan 2006; Baselga and Araújo 2010).

An obvious shortcoming of CLMs is their inability to explicitly account for the underlying processes driving species co-occurrence patterns (Baselga and Araújo 2010). This shortcoming is shared by many community ecology approaches, which regularly attribute non-random co-occurrence patterns to biotic interactions (Webb et al. 2002; Hardy 2008). Co-occurrence patterns can, however, be generated by a variety of alternative processes, including shared environmental responses of sympatric species, opposing environmental responses of parapatric species, or dispersal limitation (Boulangeat et al. 2012; Pollock et al. 2014; Morueta-Holme et al. 2015). When environmental responses are the primary driver of species co-occurrence patterns, simple SDMs may be

sufficient modeling tools as they quantify only species–environment correlations. Nevertheless, where biotic interactions act as an important constraint on species distributions, their effects will be implicitly represented in the datasets used to fit distribution models (Olden et al. 2006; Baselga and Araújo 2009; Godsoe and Harmon, 2012). In such a case, an SDM will likely suffer from poor predictive accuracy when applied to new environments where fitted climate equilibrium relationships may break down (Araújo et al. 2005b; Godsoe and Harmon 2012; Wisz et al. 2013). Additionally, combined SDM predictions will likely fail to accurately reproduce co-occurrence patterns due to their reliance on simple climate-occurrence relationships (Guisan and Rahbek 2011; Pellissier et al. 2012). By more explicitly accounting for the influence of interacting species in the model fitting process, CLMs may have the potential to address the shortcomings of SDMs. In particular, 'assemble and predict together' approaches, which model species simultaneously, may capture additional influences on species distributions. Specifically, if CLMs are able to account for effects of biotic interactions undetected by SDMs, we may expect systematic differences in model predictions that result in two primary outcomes: (1) if interspecific competition acts to constrain species distributions, CLMs should predict a lesser degree of spatial overlap among species (i.e., co-occurrence) than SDMs and (2) if facilitation among species acts to expand distributions beyond environmental tolerances, CLMs should predict a greater degree of spatial overlap than SDMs. In either instance, a model that accounts for biotic interactions should predict different co-occurrence patterns than SDMs, and these co-occurrence patterns should more accurately represent observed co-occurrence patterns (i.e., greater model performance and predictive accuracy). In this way, CLMs may have the potential to improve understanding of biotic constraints on species distributions despite their inability to explicitly address processes underlying co-occurrence patterns.

While CLMs and SDMs have been compared in previous studies, the combined results are inconclusive and thus appropriate applications for CLMs remain unclear (Baselga and Araújo 2009). CLMs have been shown to outperform SDMs in some cases (Elith et al. 2006; Olden et al. 2006), perform worse in other cases (Baselga and Araújo 2009), and to perform

similarly (Leathwick et al. 2006). Differences in performance of the two modeling approaches have been largely attributed to differences in species prevalence and range size (Elith et al. 2006; Leathwick et al. 2006; Chatfield 2008; Baselga and Araújo 2009), or major statistical differences in the SDM and CLM models being compared (Elith et al. 2006; Baselga and Araújo 2010). In light of these inconsistencies, there is a need to assess these two approaches using consistent statistical methods and species data with similar prevalence. Further, the comparative approaches used thus far have relied primarily on measures of model performance (e.g., classification accuracy, correlation between observed and fitted values), and have not compared differences in predicted co-occurrence patterns among the two approaches, which may provide greater insight into the underlying relationships captured by SDMs and CLMs (Guisan and Rahbek 2011).

Here, we compare the abilities of CLMs and SDMs to accurately predict the individual occurrence patterns and co-occurrence patterns of species of relatively similar prevalence using two ‘assemble and predict together’ CLM methods and their SDM counterparts: multivariate adaptive regression splines (MARS) and conditional random forests (CForest). We apply these models to predict current and future distributions of three dominant Rocky Mountain conifer tree species: *Pinus contorta* var. *latifolia* [(Engelm.), lodgepole pine], *Pinus ponderosa* var. *scopulorum* [(Engelm.), ponderosa pine], and *Pseudotsuga menziesii* var. *menziesii* [(Mirb.), Douglas-fir]. These canopy species form primarily parapatric distributions with distinct elevational zonation in the montane zone of the Rocky Mountains and co-occur with few other tree species (Fig. 1). Specifically, ponderosa pine dominates on dry, low-elevation sites (>1700 m), while Douglas-fir tends to occupy more xeric sites at mid-elevations (~2000 m). Lodgepole pine forms primarily monospecific stands on more mesic and higher elevation slopes ranging in elevation from 2400 to 3000 m, interacting with subalpine fir and Engelmann spruce at its upper elevational edge (Peet 1981). Lodgepole pine’s distribution is constrained to more northern latitudes than either ponderosa pine or Douglas-fir, yet focal species interact along mid-elevation ecotonal bands in the Northern and Central Rockies and form mixed-species stands in portions of their ranges (Bartlein et al. 1997). These

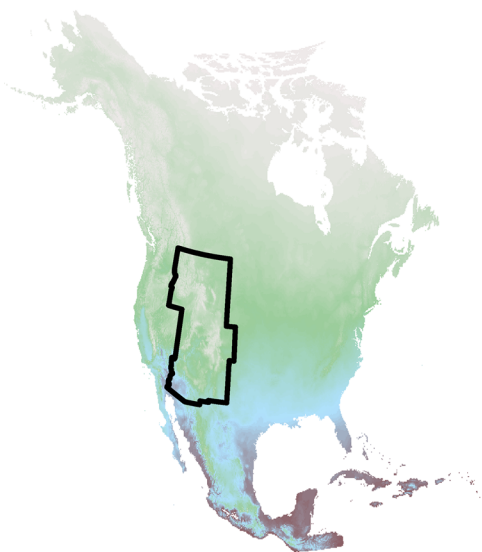
species exhibit somewhat divergent climate envelopes, particularly with regard to precipitation (Bell et al. 2014). It is unclear what role interactions between these species play in shaping current distributions, but differences in competitive ability (Copenhaver-Parry and Cannon 2016) and dispersal (McCaughy et al. 1985) suggest that current distributions may differ substantially from climatic equilibrium. To evaluate the ability of each approach to account for the effects of biotic interactions, we first compare the predictive accuracies of CLMs and SDMs from the same families of models fitted to current distribution data. We then assess accuracy of predicted co-occurrence patterns with particular emphasis on regions of known species overlap. We predict that (1) CLMs should exhibit improved predictive accuracy over SDMs by explicitly accounting for the effects of biotic interactions, and (2) CLMs and SDMs should demonstrate systematic differences in predictions of species co-occurrence patterns, diverging most strongly in regions of known species overlap.

Materials and methods

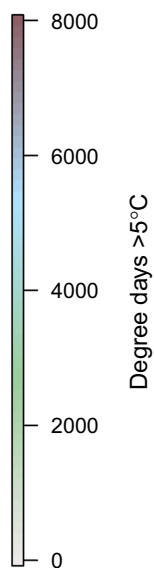
Occurrence data

Occurrence data for lodgepole pine, ponderosa pine, and Douglas-fir were extracted from the U.S. Forest Service’s Forest Inventory and Analysis (FIA) database. The FIA database consists of plot-level forest data from a comprehensive survey of forest conditions across the United States. Forests are surveyed every 5–10 years and data are provided at several spatial resolutions, based either on remote sensing (Phase 1), or field-level observations (Phase 2 and 3). The FIA has established 125,000 Phase 2 plots per 6000 acres of forested land and 8000 Phase 3 plots, or one for every 95,000 acres of forested land (Smith 2002). Plots are stratified based on landscape homogeneity in an attempt to represent the full range of forest conditions and to reduce spatial autocorrelation (Woudenberg et al. 2010). FIA data provides the most comprehensive source of presence/absence data on tree species available in the United States. FIA data does introduce a limitation to the spatial resolution of predictive models: coordinates of most plot locations are perturbed slightly within a 0.8 km radius of actual plot locations (Woodall et al. 2010). However,

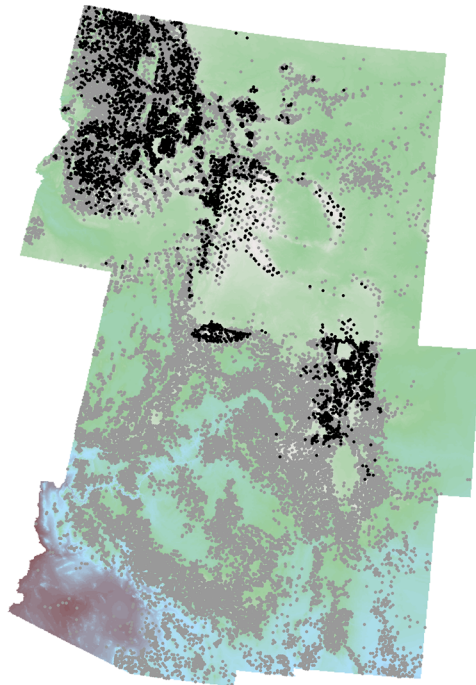
(a) Study area



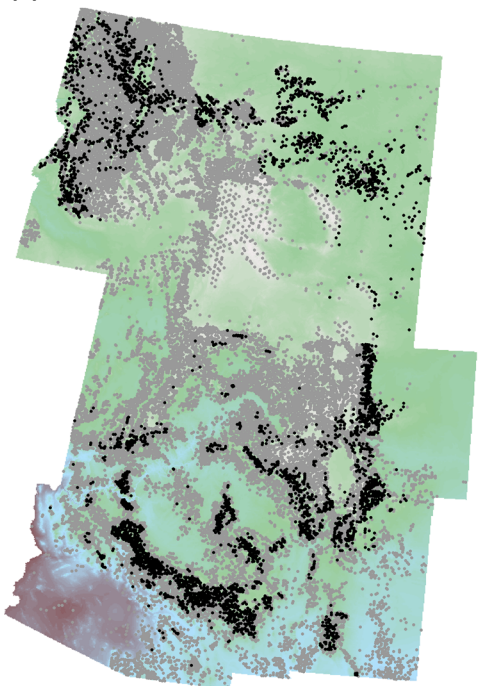
(b) Lodgepole pine



Degree days >5°C



(c) Ponderosa pine



(d) Douglas-fir

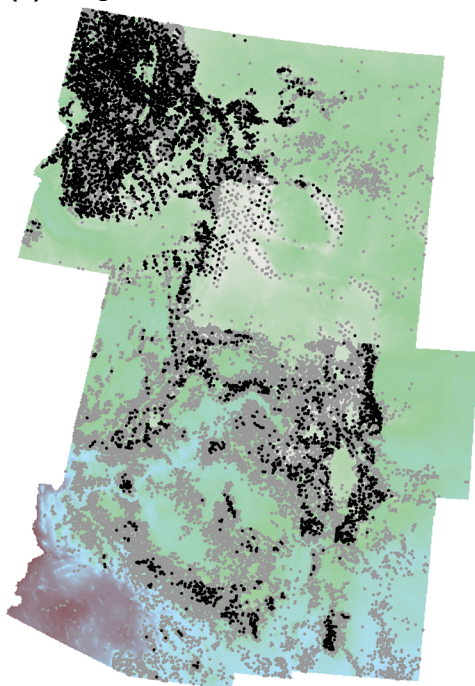


Fig. 1 The study area (a) encompassed the U.S. states of Idaho, Montana, Utah, Wyoming, Colorado, Arizona, and New Mexico. FIA data used to model species occurrence demonstrate the general distribution of lodgepole pine (b), ponderosa pine (c), and Douglas-fir (d) across the study region. Presence locations

for each species are shown in *black*, while absence locations are shown in *gray*. Points are superimposed on a gradient of growing degree days >5 °C, which is a covariate that was consistently selected for in the models developed in this study

perturbed coordinates used in SDMs have resulted in similar performance to SDMs using precise coordinates (Gibson et al. 2014), and the uncertainty in resolution is deemed acceptable for integration with 1 km resolution climate data (Woodall, personal communication).

In this study, we made use of all available field-observation (Phase 2 and 3) FIA plots within the U.S. states of Montana, Idaho, Utah, Wyoming, Colorado, New Mexico, and Arizona (Fig. 1). This study area was selected in order to capture the U.S. distributions of inland varieties of the focal species. For each plot, we extracted presence/absence data for lodgepole pine, ponderosa pine, and Douglas-fir from the most recent plot surveys. Only live, mature trees were considered due to greater uncertainty in identification of seedlings and the possibility of sink (non-equilibrium) populations at plots with only seedlings of a given species. In total, 21,950 presence/absence observations were retained for analysis. Prevalence (proportion of plots where species occurs) was relatively similar for all species: 16 % for lodgepole pine, 18 % for ponderosa pine, and 28 % for Douglas-fir.

Climate data

Current climate estimates were extracted from the U.S. Forest Service's Moscow Forestry Sciences Laboratory (MFSL) down-scaled climate dataset. MFSL data are provided at a 30 arc second (~ 1 km) resolution as 30-year normals (1961–1990) with coverage spanning Western North America. The MFSL dataset provides plant-relevant climate variables for integration with ecological data (Rehfeldt 2006). We utilized only a subset of available climate variables to reduce model dimensionality and minimize overfitting, a problem common to both modeling methods used here (Hothorn et al. 2006a; Leathwick et al. 2006). While model complexity may also contribute to overfitting, we evaluated possible overfitting of each method by validating models on geographically stratified data (see “MARS model” section). Climate variables were selected to represent seasonality of temperature and precipitation, which are known controls on Rocky Mountain tree distributions (Bell et al. 2014). To capture topographic relationships that may not be well represented by climate data, we extracted elevation (m.a.s.l.) from a USGS 30 m digital elevation model (DEM),

re-sampled to a 1 km grid using bilinear interpolation to remain consistent with the resolution of the MFSL climate data. From this DEM, we derived an index of topographic radiation based on a continuous transformation of circular aspect (TRASP; Roberts and Cooper 1989; Evans et al. 2014). The ability of both modeling methods used here to accommodate collinearity has been questioned (Leathwick et al. 2006; Murphy et al. 2010), and thus we omitted highly correlated variables ($r > 0.7$) to ensure independence among covariates (Dormann et al. 2013). Final variables included growing degree days $> 5^{\circ}\text{C}$ (dd5), TRASP, growing season precipitation (gsp), and summer precipitation balance (smprb).

MARS model

Both single-species and CLM implementations of the *MARS* algorithm were used to fit climate and topography metrics to a subset of the occurrence data. Utilizing the same algorithm for both SDM and CLM implementations ensured that the only major differences between the two approaches were related to the number of species being modeled, leaving inclusion of co-occurrence information as the most parsimonious explanation for substantial differences in SDM and CLM predictive performance. *MARS* uses piecewise parametric fitting of basis functions based on recursive partitioning regression with a back-fitting algorithm to maintain model parsimony (Friedman 1991). This back-fitting approach removes basis functions that no longer contribute substantially to model fit, thus minimizing overfitting problems that are common to many other recursive partitioning approaches. The *MARS* algorithm is designed to reduce computational complexity and increase analytical speed and greatly reduces the computational costs associated with alternative recursive partitioning methods (Friedman 1993). In the multiple-species implementation of *MARS*, basis functions are optimized simultaneously across all species (Friedman 1991). Because *MARS* is designed to accommodate continuous responses, we adopted the approach of Leathwick et al. (2006) to model probability of presence based on binary occurrence data. Basis functions generated by the *MARS* algorithm were used to fit a GLM with a logit link function. For the multiple-species implementation, GLM coefficients were fitted separately for each species.

Because we lacked a large independent dataset for model validation, we partitioned our data into calibration and validation datasets using a spatially segregated splitting approach (Bahn and McGill 2013). This approach ensures greater independence between calibration and validation data and provides more realistic assessments of model predictive ability (Peterson et al. 2007). Following (Bahn and McGill 2013), we quadri-sected our data longitudinally. Quarters one and three were combined and used as calibration data, and quarters 2 and 4 were used as validation data. It should be noted that we found no spatial autocorrelation among our species occurrence data or our model covariates (Moran's $I = 0$, effective spatial range $\phi = 0$), likely due to the stratified sampling approach used for FIA data and the complex, heterogeneous topography of our study region. However, the spatially segregated splitting approach ensured that our models were validated on climatic and geographic conditions that were not fully represented in the calibration data, thus improving our ability to evaluate predictive accuracy (Arau'jo et al. 2005b; Bahn and McGill 2013). Single-species *MARS* models were developed for each species (SDMs), and a multiple-species *MARS* model was fitted to all species simultaneously CLM. Both additive models and two-way interaction models were fitted and compared. Models were evaluated for fit and parsimony based on the Generalized Cross Validation (GCV) criterion (Craven and Wahba 1979), and the model with the lowest GCV in each pair was retained for subsequent analysis. In all cases, two-way interaction models were retained. All *MARS* models were fitted with the 'earth' package (version 4.2.0; Milborrow 2015) in R (R Core Team 2014). Fitted models were then predicted back to climate and topography grids to spatially display probability of occurrence across the entire study region. Probability of occurrence was converted to predicted presence/absence using a prevalence-based threshold (Liu et al. 2005). Prevalence-based thresholds have been shown to outperform other threshold approaches, including approaches that maximize model performance criterion, and provide consistency when comparing predictions across species with similar prevalence (Liu et al. 2005). Maintaining consistency among species and models was a primary goal in our selection of methods, as our evaluation relies on the relative differences between modeling approaches and algorithms, rather than on the predictive accuracy of each approach on its own.

CForest model

The *CForest* algorithm generates an ensemble of conditional classification trees using recursive partitioning and is designed to overcome some of the biases associated with the more commonly used Random Forest algorithm. Conditional trees differ from the standard classification trees in that the variable selection process is separated from the splitting value selection, thus minimizing bias towards variables with many splits. Both variable selection and split determination are accomplished by permutation tests that measure the association between covariates and responses based on a P value. Stopping criteria based on statistical significance are incorporated into the *CForest* algorithm to halt recursion when additional splits do not contribute significantly to model fit. This approach maintains model parsimony and reduces overfitting. In the multiple-species CLM implementation, all response variables are transformed to log-rank scores for use in the permutation test, and the association between covariates and the log-ranked responses is tested (Hothorn et al. 2006a).

We fit *CForest* models to individual species occurrence data (SDM) and to co-occurrence data for all species simultaneously (CLM). For each model, 128 trees were grown with 4 variables evaluated at each split. Models were fitted to the same calibration data that the *MARS* models were constructed with, and evaluated on the same remaining validation data. All *CForest* models were fitted using the 'party' package (Hothorn et al. 2006b) in R (R Core Team 2014). As with the *MARS* models, probability of presence across the study region was modeled by predicting the fitted *CForest* models back to climate and topography grids, and predicted presence/absence was evaluated using a prevalence-based threshold.

Model comparison

SDM and CLM predictions were compared using non-spatial metrics based on validation data and comparisons of mapped model predictions. The ability of models to discriminate between presences and absences was assessed with AUC, a threshold-independent metric that indicates both the sensitivity (correctly classified presences) and specificity (correctly classified absences) of the model (Manel et al. 2001). AUC is calculated as the area under the

receiver-operating characteristic curve, which is generated by plotting sensitivity against the false-positive rate for all possible threshold values. We also evaluate discriminatory and predictive ability using the true skill statistic (TSS), a prevalence-dependent criterion based on sensitivity and specificity as determined by the prevalence-based threshold (Allouche et al. 2006). TSS is used to indicate improvement of a model from random prediction, which is assessed at a TSS value of 0.5. Sensitivity and specificity, again estimated using a prevalence-based threshold, were also evaluated independently to pinpoint underlying differences in model discrimination. To assess geographic overlap of mapped predictions for both current and future conditions, we used Schoener's D statistic (D). D represents the proportional geographic overlap of two distribution predictions as an index ranging from 0 to 1 (Renkonen 1938; Warren et al. 2008; Roder and Engler 2011). Differences in mapped predictions between CLMs and SDMs were also assessed by comparing the percent difference in total area predicted to be occupied by a given species ($Adiff$) based on a prevalence-based threshold. $Adiff$ was calculated as the percent difference in CLM-predicted occurrence relative to SDM-predicted occurrence; thus, a positive value indicates a greater area of occurrence under the CLM. We emphasize that these criteria are utilized as a comparative tool to assess differences between SDM and CLM approaches, rather than individual model performance, in an attempt to evaluate whether CLMs capture the effects of biotic interactions.

Species co-occurrence based on SDM and CLM predictions was evaluated by assessing the number of species predicted to be present at each site. For the SDM approach, we stacked SDM predictions for our three focal species and calculated the sum of predicted presences at each site, ranging from zero to three. For the CLM predictions, we simply summed the number of species predicted present at each site from the simultaneous CLM prediction of all focal species. Similarities between predicted species occurrence and underlying data were evaluated by comparing the overall classification accuracy, sensitivity, and specificity for each species richness classification (0–3), and Cohen's Kappa (κ ; Cohen 1960). κ was used because of its ability to evaluate classification accuracy on more than two categories.

For additional illustrative purposes, we introduced a small independent dataset of species co-occurrence

to further validate predicted species co-occurrence. These data classify the number of focal species present at sampling plots across four ecotones in the Northern and Central U.S. Rocky Mountains and were collected independently from FIA data (Copenhaver-Parry and Cannon 2016). These ecotones represent regions of known species overlap, and provide additional insight into the ability of SDMs and CLMs to capture true co-occurrence patterns. For this small dataset, we visually compared the predicted number of species present at each site to measured values.

Results

MARS

The predictive accuracy of SDMs and CLMs predicted to spatially segregated validation data varied across species and discrimination metrics (Table 1). While the CLM approach slightly outperformed the SDM approach for lodgepole pine, predictive accuracy was higher for ponderosa pine and Douglas-fir under the SDM. Overall, Douglas-fir, the most prevalent species, was predicted with the lowest accuracy, performing no better than random according to the TSS statistic. Nevertheless, differences between SDM and CLM predictive accuracy are slight across all species. The SDMs and CLMs both predicted very similar geographic distributions for Douglas-fir and ponderosa pine, while geographic distributions for lodgepole pine differed more substantially, despite similar predictive accuracy across species (Table 1). This suggests that incorrectly classified locations for lodgepole pine were counterbalanced by the two modeling approaches, i.e., many locations predicted inaccurately by the SDM were predicted more accurately by the CLM and vice versa. This points to a difference in the underlying relationship captured by the two modeling approaches for lodgepole pine. This is consistent with the $Adiff$ statistic, which identifies a larger difference in the area of predicted presence for lodgepole pine by the SDM versus the CLM relative to other species (26.8 % increase in predicted area of occurrence by the CLM). Also consistent with D , $Adiff$ was substantially smaller and negative for ponderosa pine (−1.71 %) and Douglas-fir (−3.32 %), indicating that both modeling approaches classified sites similarly for these two species. The CLM predicted a

Table 1 Predictive accuracy of CLM and SDM models fitted under both the MARS and CForest algorithms along with geographic similarity in mapped predictions (*D*), and differences in area of predicted occurrence (*Adiff*)

	AUC		TSS		Sens.		Spef.		<i>D</i>	<i>Adiff</i> (%)
	SDM	CLM	SDM	CLM	SDM	CLM	SDM	CLM		
MARS										
Lodgepole pine	0.862	0.876	0.574	0.630	0.852	0.884	0.759	0.746	0.742	26.8
Ponderosa pine	0.833	0.833	0.507	0.488	0.696	0.680	0.811	0.808	0.866	−1.71
Douglas-fir	0.803	0.790	0.451	0.399	0.809	0.773	0.642	0.627	0.943	−3.32
CForest										
Lodgepole pine	0.812	0.861	0.571	0.629	0.767	0.853	0.804	0.776	0.814	10.2
Ponderosa pine	0.836	0.850	0.481	0.558	0.639	0.814	0.842	0.744	0.749	29.9
Douglas-fir	0.796	0.800	0.401	0.415	0.757	0.824	0.644	0.591	0.886	17.9

slightly smaller area of occurrence for ponderosa pine and Douglas-fir than the SDM.

In general, the SDM approach under-predicted species co-occurrence, while the CLM over-predicted co-occurrence (Table 2; Fig. 2). These differences are mostly due to differences in the predicted geographical extent of lodgepole pine, which was under-predicted by the SDM and over-predicted by the CLM. Classification accuracy for the number of species present was similar and poor for both approaches, indicating that neither modeling approach appropriately captures co-occurrence. Consistent with our hypothesis, co-occurrence predictions diverge most strongly in mid-elevation zones in the Northern and Central Rockies, where species are most likely to interact (Fig. 2). When compared to independent field data along ecotones of known species overlap (Copenhaver-Parry and Cannon 2016), SDMs generally under-predict the number of species present along ecotones (Fig. 2b–e), particularly for the two southernmost ecotones (Fig. 2d, e). The CLM, however, also fails to predict many sites of three-species co-occurrence in ecotones (Fig. 2g–j).

CForest

Similar to the MARS results, predictive accuracy for the CForest algorithm differed only slightly between SDM and CLM models, and varied across species (Table 1). Predictive accuracy was slightly higher for CLM models for all species, though TSS values indicate that predictions were often little better, and sometimes worse, than random. Conversely, AUC values indicate fair to good predictive accuracy, suggesting that an alternative threshold approach

Table 2 Classification accuracy for predicted species co-occurrence patterns

		Accuracy ^a	Kappa ^b	Percent area ^c			
				0	1	2	3
MARS							
SDM	0.524	0.293	51.9	16.4	28.3	3.36	
CLM	0.505	0.272	54.6	16.1	23.2	6.00	
CForest							
SDM	0.585	0.369	57.1	18.6	20.1	4.26	
CLM	0.544	0.331	49.9	16.1	27.4	6.66	

^a The proportion of co-occurrences that were correctly predicted by the model

^b A measure of agreement between true classified values and predicted classified values; 1 = perfect agreement; 0 = agreement equivalent to chance

^c The percentage of the study area predicted to be occupied by the specified number of species

may have produced improved classification accuracy. Douglas-fir was predicted with the lowest accuracy under both SDM and CLM approaches. In general, lodgepole pine was predicted with the greatest accuracy, though ponderosa pine data generated a higher AUC and specificity in the SDM.

Geographic predictions from SDMs and CLMs were relatively similar across all species, with ponderosa pine showing the greatest difference (lowest *D*; Table 1). Sensitivity and specificity between SDMs and CLMs also show the greatest difference for ponderosa pine, indicating that both modeling approaches classified many locations differently for this species. This is reflected by the *Adiff* statistic, which demonstrates a large difference in the area of

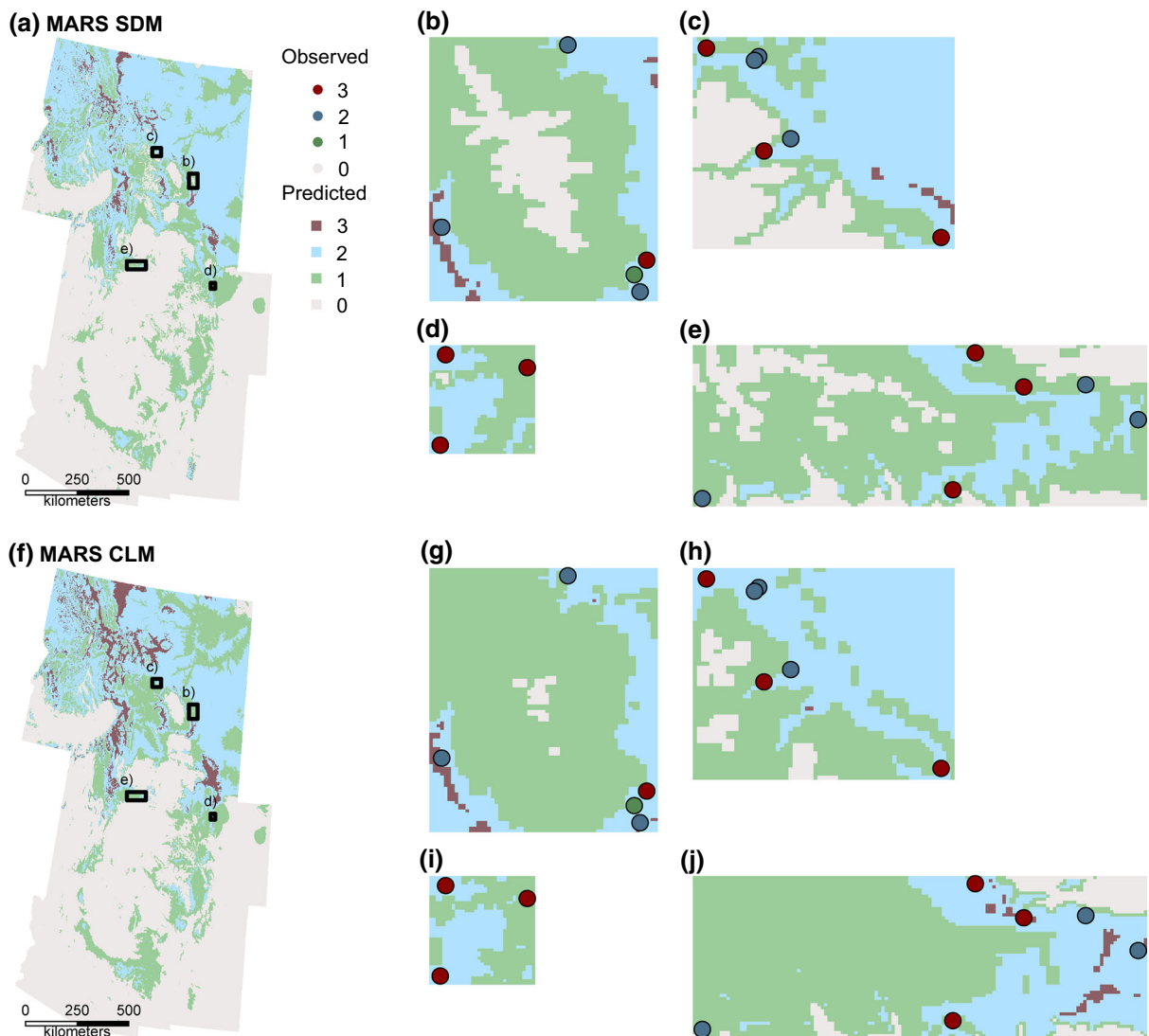


Fig. 2 Co-occurrence predictions from the *MARS* SDMs (a–e) and the *MARS* CLMs (f–j). Across ecotones where all species are known to interact, both the SDM models (b–e) and the CLM

models (g–j) generally failed to accurately reproduce observed species co-occurrence patterns (colored points; Color figure online)

predicted presence locations for ponderosa pine between the SDM and the CLM, and a smaller difference for lodgepole pine and Douglas-fir. Across all species, the CLM approach resulted in a greater area of predicted occurrence than the SDM approach. Differences in co-occurrence prediction accuracy were slight between the two modeling approaches, and both demonstrated relatively poor agreement with observed co-occurrence (Table 2). Across the study region, SDMs under-predicted species co-occurrence, while the CLM over-predicted co-occurrence. Both approaches indicate substantial overlap of all three

focal species in the Northern and Central Rockies and absence of lodgepole pine in the Southern Rockies, which is consistent with data (Fig. 3a, f). When predictions are evaluated against independent data from ecotonal regions, differences in species overlap between SDMs and the CLM appear to be slight. In general, both approaches reasonably agree with data in the two northernmost ecotones (Fig. 3b, c, g, h), while the CLM more accurately captures co-occurrence in the two southernmost ecotones (Fig. 3d, e, i, j). The SDM approach does not accurately capture the southern limit of lodgepole pine's distribution in the

Rocky Mountains, and inaccurately limits the distribution of this species to a more northern extent.

MARS and CForest comparison

Differences in predictive accuracy between *MARS* and *CForest* models are similar in magnitude to differences between SDMs and CLMs within the same modeling approach (Table 1). In general, the SDM

implementation of the *MARS* model demonstrated improved predictive accuracy over the *CForest* SDMs, while the *CForest* CLM generally predicted data more accurately than the *MARS* CLM. However, we note several important exceptions. First, when comparing CLMs, lodgepole pine was predicted more accurately by the *MARS* algorithm. Additionally, the *CForest* SDM model showed a higher AUC for ponderosa pine relative to the *MARS* SDM, yet all

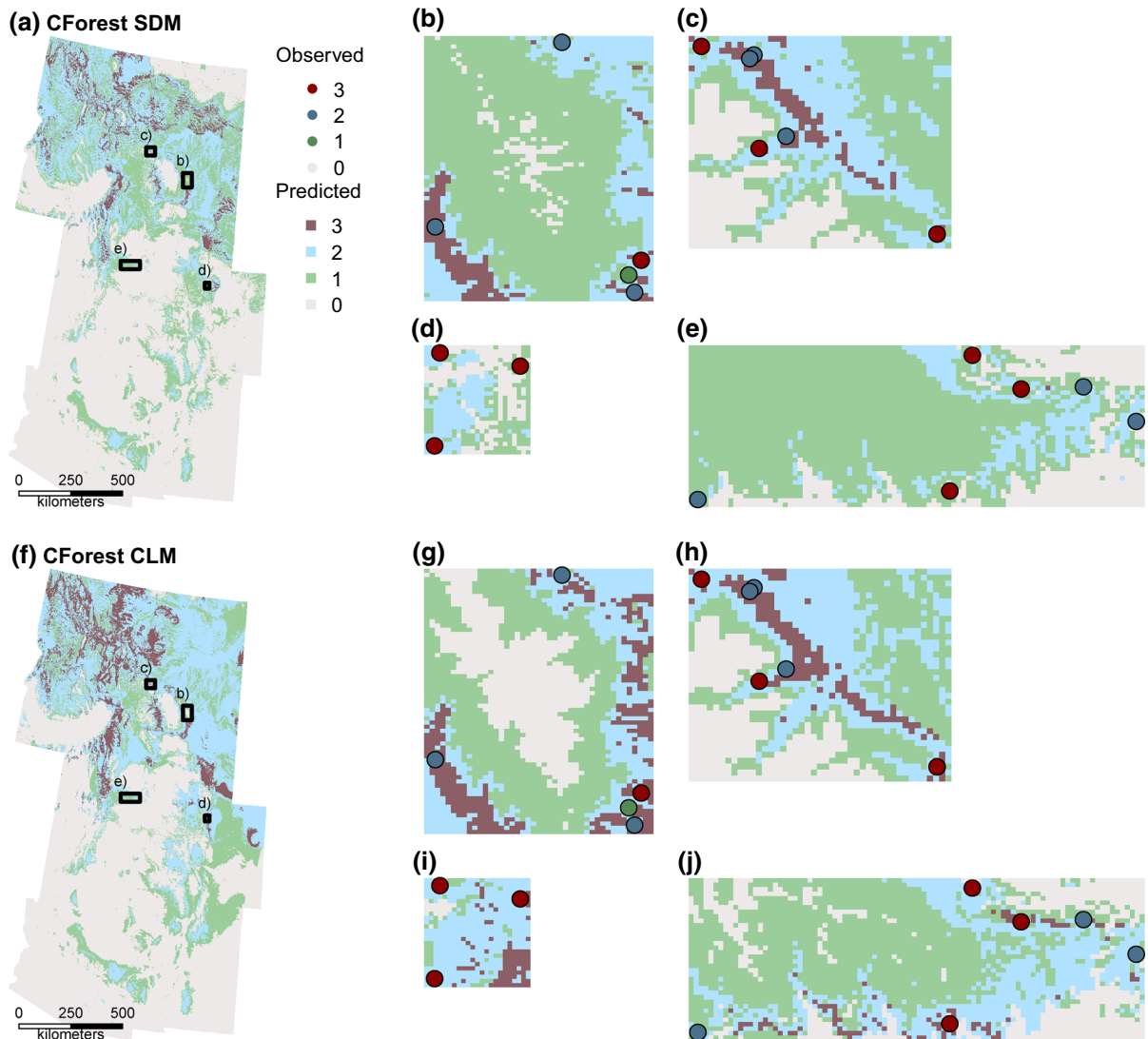


Fig. 3 Co-occurrence predictions from the *CForest* SDMs (a–e) more accurately classify species occurrence in more northern ecotonal regions (b, c), but fail to model known regions of three-species occurrence (red points) in more southern ecotones (d, e). Co-occurrence predictions from the *CForest* CLM (f–j) show a

greater area of species overlap than SDM predictions, particularly for regions of three-species overlap (red), and show slightly improved classification of species overlap along ecotones (g–j) when compared to independent observation data (colored points; Color figure online)

threshold-based statistics demonstrated improved classification for the *MARS* model. However, we emphasize that differences in predictive accuracy between approaches are slight.

Neither modeling approach was able to reproduce species co-occurrence with good accuracy, yet *CForest* models demonstrated slightly improved classification agreement over *MARS* models (Table 2). In general, CLM predictions from the two modeling approaches were more similar than were SDM predictions (Table 3). Predicted area of occurrence was most similar for Douglas-fir across the two modeling approaches, and most dissimilar for lodgepole pine under the SDM and ponderosa pine under the CLM. We evaluated the *Adiff* of these two approaches as *CForest* relative to *MARS*; thus, a negative value indicates a larger predicted area of occurrence by the *MARS* model than the *CForest* model. Across all species, *MARS* predicted greater regions of occurrence than *CForest*, with the notable exception of the lodgepole pine SDM, where *CForest* predicted a larger area of occurrence than *MARS*. Differences were most pronounced for ponderosa pine when comparing SDM predictions, and lodgepole pine when comparing CLM predictions.

Discussion

A species' distribution not only reflects its climate-induced physiological tolerances, but may also be shaped by interactions with other species (Case et al. 2005). At broad scales, biotic interactions are expected to generate non-random co-occurrence patterns and to alter species–environment relationships from these occurring in isolation (Wisiz et al. 2013). Both of these expectations are fundamental assumptions of the CLM

approaches evaluated in this study. CLMs are intended to more accurately model species–environment relationships by explicitly accounting for other species when assigning statistical correlations (Ferrier and Guisan 2006). However, the ability of CLMs to appropriately account for the effects of biotic interactions has remained uncertain, and thus their utility as a modeling tool has been questioned (Baselga and Araujo 2009). We predicted that CLMs should predict systematically different species distributions than SDMs by incorporating biotic information. In particular, CLMs and SDMs should differ in the spatial overlap, or co-occurrence predicted for multiple species. We further predicted that if CLMs do in fact account for biotic interactions, the distributions and co-occurrence patterns predicted by CLMs should more accurately reproduce observed patterns than SDMs, indicated by an improvement in predictive accuracy.

Our results do demonstrate some systematic differences between CLM and SDM predictions, although trends are not entirely consistent across species. In general, CLMs predict larger areas of occurrence than SDMs (Table 1), and a greater degree of co-occurrence (Table 2). However, we note several important exceptions: in the case of the *MARS* algorithm, the SDM approach predicted slightly larger areas of occurrence for both ponderosa pine and Douglas-fir, although differences were small. The general trends observed suggest that inclusion of biotic information generally acts to alter the extent of the climatic niche of modeled species. This finding is consistent with the model fitting approaches used by both the *MARS* algorithm and the *CForest* algorithm, and may reflect model structure more than biological phenomena.

In the *MARS* algorithm, multispecies predictions are generated by optimizing basis functions simultaneously across all species (Friedman 1991; Leathwick et al. 2006). This essentially amounts to averaging environmental responses across all species. In the *CForest* CLM fitting process, a permutation test based on log-rank scores allows each species to have varying amounts of influence at each split, yet the model is still fit across all species simultaneously (Hothorn et al. 2006a, b). Thus, in both algorithms, an averaging effect is imposed. For species that exhibit opposing environmental responses or parapatric distributions, this averaging effect may alter the predicted climatic niche and geographical extent of each species,

Table 3 Comparison of *CForest* and *MARS* predictions

	<i>D</i>		<i>Adiff</i> (%)	
	SDM	CLM	SDM	CLM
Lodgepole pine	0.705	0.804	2.80	−19.30
Ponderosa pine	0.746	0.759	−67.40	−15.40
Douglas-fir	0.820	0.860	−45.40	−15.60

D represents the geographic similarity of *MARS* and *CForest* predictions, and *Adiff* represents the percent difference in predicted area of occurrence (a positive value indicates a greater area of occurrence under the *CForest* model)

specifically by broadening the niche/extent of narrowly distributed or climatically constrained species, and constraining the niche/extent of more broadly distributed species (Madon et al. 2013). Our findings reflect the effects of this averaging process. For example, lodgepole pine is the most narrowly distributed species in our dataset, with a distribution that is constrained to a more northern extent of the study region (Fig. 1). Additionally, lodgepole pine has a smaller climatic niche than either ponderosa pine or Douglas-fir, particularly with regard to summer precipitation (Bell et al. 2014). When modeled with the *MARS* CLM, the distribution of lodgepole pine is expanded relative to SDM predictions, while the distributions of ponderosa pine and Douglas-fir are constrained (Table 2; *Adiff*). Additionally, ponderosa pine and Douglas-fir distributions are more similar between SDM and CLM predictions, while lodgepole pine differs more substantially (Table 2; *D*). This may reflect differences in species prevalence, with more prevalent species (i.e., ponderosa pine and Douglas-fir) exerting greater influence in the averaging process. The *CForest* algorithm may overcome some of the limitations of a pure averaging process by allowing species to have varying degrees of influence at each split (Hothorn et al. 2006a, b). In our *CForest* analysis, all species distributions were expanded in the CLM relative to the SDMs (Table 2). The greater consistency across species indicates that the *CForest* algorithm may minimize the influence of species prevalence and geographic extent on fitted environmental responses.

Despite systematic differences in model predictions related to the inclusion of biotic information, we found no consistent improvement in CLM predictive accuracy relative to that of SDMs. Differences in predictive accuracy between CLMs and SDMs were similar in magnitude to differences between *MARS* and *CForest* models (Table 3), again indicating that model structure, rather than the effects of biotic interactions, explains much of the variation in model output. Additionally, both approaches failed to accurately reproduce patterns of species co-occurrence. CLMs generally predicted too great an area of three-species co-occurrence, while SDMs failed to predict many regions of three-species co-occurrence (Table 2). Further, neither approach sufficiently captured species co-occurrence along ecotones, particularly in the more southern portions of lodgepole pine's distribution

(Figs. 2, 3). We do note that the *CForest* algorithm demonstrated higher classification accuracy for species co-occurrence than *MARS*, again indicating that the *CForest* algorithm overcomes some of the limitations associated with *MARS*'s pure averaging approach (Hothorn et al. 2006a, b).

It is possible that our findings also indicate a limited role for biotic interactions in defining the distributions of our focal species. While mounting evidence from other plant systems has demonstrated improved pre-dictions of tree distributions after accounting for biotic interactions (Rouget et al. 2001; Meier et al. 2010; Boulangeat et al. 2012; Clark et al. 2014), the strong elevational zonation in our study system may simply be a function of sharp physiological limitations that interact with climate to determine local distribution edges. This is consistent with the findings of Copen-haver-Parry and Cannon (2016), which identify climate as the primary driver of growth trends at distribution edges of our focal species. However, our SDM models generally failed to predict the distributional limit of lodgepole pine, and under-predicted co-occurrence, suggesting that factors in addition to the climatic factors evaluated here contribute to these species' distribution patterns. Our results indicate that it will require improved modeling approaches to determine the precise role of biotic interactions in structuring these species' distributions.

Both CLMs and SDMs have a variety of additional limitations related to their ability to account for biotic interactions that were not directly highlighted in our analyses. For example, both CLMs and SDMs are unable to account for changes in species interactions over time. The magnitude and direction of plant interactions have been shown to be altered by past environmental change, suggesting that changes in interactions will also be observed under future conditions (Tylianakis et al. 2008; Blois et al. 2014). Climatic changes may drive reversals in competitive hierarchies, or even result in novel species assemblages (Jackson et al. 2009). Specifically, large environmental changes may reduce the competitive advantage of more specialized species and favor generalist species (Schubert and Bottjer 1995; Sahney and Benton 2008). Additionally, changes in interactions with other taxa across a variety of trophic levels may also have large impacts on future distributions. Of particular relevance in our study region, bark beetle outbreaks associated with climate warming have had a

comparatively larger impact on *Pinus ponderosa* and *Pinus contorta* in the Rocky Mountains than on *Pseudotsuga menziesii* (Meddens et al. 2012). This competitive advantage is independent of climatic tolerance, and may result in range expansion of Douglas-fir and contraction of lodgepole pine and ponderosa pine that cannot be predicted by static CLMs or SDMs (Wisz et al. 2013). Neither CLMs nor SDMs can differentiate between the contributions of environmental tolerances and biotic interactions to co-occurrence patterns, making direct quantifications of the factors underlying species distributions impossible (Wisz et al. 2013; Pollock et al. 2014).

Our results suggest that CLMs offer no real improvement over SDMs in accounting for the effects of biotic interactions. Therefore, CLMs are unlikely to generate accurate predictions of species whose distributions are influenced by biotic interactions. This is highlighted in our results by similar predictive accuracy of SDMs and CLMs, poor classification accuracy for co-occurrence patterns across both SDMs and CLMs, and over-prediction of species co-occurrence by CLMs. Further, CLMs cannot be used to identify the causes of climate-distribution disequilibria, which may be due to true biotic interactions or simply due to environmental factors that remain unaccounted for in the model. However, CLMs may find utility in modeling assemblages of regularly co-occurring and strongly overlapping species that demonstrate shared environmental responses and similar climatic niches (see Chatfield 2008; Baselga and Araújo 2009; Madon et al. 2013 for a more complete discussion of CLMs in this context). While our results do not directly evaluate the utility of CLMs in such contexts, we do demonstrate a slight improvement in classification accuracy for the *CForest* CLM over the *MARS* CLM, and note improvement in the species averaging process in the *CForest* algorithm. Thus, in contexts where a CLM may be appropriate, *CForest* is likely to produce more accurate predictions than *MARS*. Overall, we conclude that alternative methods to CLMs may provide more useful approaches to account for the effects of biotic interactions and, consequently, provide more reliable predictions of species distributions.

Acknowledgments The authors would like to thank Chris Woodall and Brian Walters for providing the data used in this study, as well as the students of the 2014 Advanced Spatial Analysis course at the University of Wyoming for their feedback and assistance. We are also grateful to two anonymous

reviewers whose comments substantially improved the quality of this manuscript. P. Copenhagen-Parry was supported by a National Science Foundation Fellowship (G-K12 Project #0841298) and a fellowship from the Wyoming NASA Space Grant Consortium during the development of this manuscript.

References

- Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol* 43:1223–1232
- Araújo MB, Luoto M (2007) The importance of biotic interactions for modelling species distributions under climate change. *Glob Ecol Biogeogr* 16:743–753
- Araújo MB, Pearson RG, Rahbek C (2005a) Equilibrium of species' distributions with climate. *Ecography* 28:693–695
- Araújo MB, Pearson RG, Thuiller W, Erhard M (2005b) Validation of species-climate impact models under climate change. *Glob Change Biol* 11:1504–1513
- Araújo MB, Rozenfeld A, Rahbek C, Marquet PA (2011) Using species co-occurrence networks to assess the impacts of climate change. *Ecography* 34:897–908
- Bahn V, McGill BJ (2013) Testing the predictive performance of distribution models. *Oikos* 122:321–331
- Bartlein PJ, Whitlock C, Shafer SL (1997) Future climate in the Yellowstone National Park region and its potential impact on vegetation. *Conserv Biol* 11:782–792
- Baselga A, Araújo MB (2009) Individualistic vs community modelling of species distributions under climate change. *Ecography* 32:55–65
- Baselga A, Araújo MB (2010) Do community-level models describe community variation effectively? *J Biogeogr* 37:1842–1850
- Bell DM, Bradford JB, Lauenroth WK (2014) Early indicators of change: divergent climate envelopes between tree life stages imply range shifts in the western United States. *Glob Ecol Biogeogr* 23:168–180
- Blois JL, Gotelli NJ, Behrensmeyer AK, Faith JT, Lyons SK, Williams JW, Amatangelo KL, Bercovici A, Du A, Eronen DJ, Graves GR, Jud N, Labandeira C, Looy CV, McGill B, Patterson D, Potts R, Riddle B, Terry R, Tóth A, Villaseñor A, Wing S (2014) A framework for evaluating the influence of climate, dispersal limitation, and biotic interactions using fossil pollen associations across the late Quaternary. *Ecography* 37:1095–1108
- Boulangeat I, Gravel D, Thuiller W (2012) Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecol Lett* 15:584–593
- Case TJ, Holt RD, McPeck MA, Keitt H (2005) The community context of species' borders: ecological and evolutionary perspectives. *Oikos* 108:28–46
- Chatfield BS (2008) How to find the one that got away. Predicting the distribution of temperate demersal fish from environmental variables. PhD Thesis, School of Earth and Geographical Sciences, University of Western Australia, Perth
- Clark JS, Gelfand AE, Woodall CW, Zhu K (2014) More than the sum of the parts: forest climate response from joint species distribution models. *Ecol Appl* 24:990–999

- Cohen J (1960) A coefficient of agreement for nominal scales. *Education and Psychological Measurement* 20:37–46
- Copenhaver-Parry PE, Cannon E (2016) The relative influences of climate and competition on tree growth along montane ecotones in the Rocky Mountains. *Oecologia*. doi:10.1007/s00442-016-3565-x
- Craven P, Wahba G (1979) Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer Math* 31:317–403
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, García Márquez JR, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:027–046
- Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Wisz S, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151
- Evans JS, Oakleaf J, Cushman SA, Theobald D (2014) An ArcGIS toolbox for surface gradient and geomorphometric modeling, version 2.0-0. Available <http://evansmurphy.wix.com/evansspatial>. Accessed 2015
- Ferrier S, Guisan A (2006) Spatial modelling of biodiversity at the community level. *J Appl Ecol* 43:393–404
- Friedman JH (1991) Multivariate adaptive regression splines. *Ann Stat* 19:1–67
- Friedman JH (1993) Fast MARS. Technical Report No. 110: Department of Statistics, Stanford University
- Gibson J, Moisen G, Frescino T, Edwards TC Jr (2014) Using publicly available forest inventory data in climate-based models of tree species distribution: examining effects of true versus altered location coordinates. *Ecosystems* 17:43–53
- Godsoe W, Harmon LJ (2012) How do species interactions affect species distribution models? *Ecography* 35:811–820
- Guisan A, Rahbek C (2011) SESAM- a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *J Biogeogr* 38:1433–1444
- Hardy OJ (2008) Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *J Ecol* 96:914–926
- HilleRisLambers J, Harsch MA, Ettinger AK, Ford KR, Theobald EJ (2013) How will biotic interactions influence climate change-induced range shifts? *Ann N Y Acad Sci* 1297:112–125
- Holt RD (2009) Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. *Proc Natl Acad Sci* 106:19659–19665
- Hothorn T, Buehlmann P, Dudoit S, Molinaro A, Van Der Laan M (2006a) Survival ensembles. *Biostatistics* 7:355–373
- Hothorn T, Hornik K, Zeileis A (2006b) Unbiased recursive partitioning: a conditional inference framework. *J Comput Graph Stat* 15:651–674
- Jackson ST, Betancourt JL, Booth RK, Gray ST (2009) Ecology and the ratchet of events: climate variability, niche dimensions, and species distributions. *Proc Natl Acad Sci* 106:19685–19692
- Kissling WD, Dormann CF, Groenvelde J, Hickler T, Kühn I, McNemy GJ, Montoya JM, Römermann C, Schifffers K, Schurr FM, Singer A, Svenning JC, Zimmermann NE, O'Hara RB (2012) Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *J Biogeogr* 39:2163–2178
- Klanderud K, Totland Ø (2005) Simulated climate change altered dominance hierarchies and diversity of an alpine biodiversity hotspot. *Ecology* 86:2047–2054
- Leathwick J (2009) Are New Zealand's Nothofagus species in equilibrium with their environment? *J Veg Sci* 9:719–732
- Leathwick JR, Elith J, Hastie T (2006) Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol Model* 199:188–196
- Liu C, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28:385–393
- Madon B, Warton DI, Araújo MB (2013) Community-level vs species-specific approaches to model selection. *Ecography* 36:1291–1298
- Manel S, Williams HC, Ormerod SJ (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *J Appl Ecol* 38:921–931
- McCaughy WW, Schmidt WC, Shearer RC (1985) Seed dispersal characteristics of conifers of the inland Mountain West. In: Shearer RC. *Proceedings: Conifer Tree Seed in the Inland Mountain West Symposium*. April 1985, Missoula. General Technical Report INT-203. USDA Forest Service Intermountain Research Station, Ogden, pp 50–62
- Meddens AJH, Hicke JA, Ferguson CA (2012) Spatiotemporal patterns of observed bark beetle-caused tree mortality in British Columbia and the western United States. *Ecol Appl* 22:1876–1891
- Meier ES, Kienast F, Pearman PB, Svenning JC, Thuiller W, Araújo MB, Guisan A, Zimmermann NE (2010) Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography* 33:1038–1048
- Meineri E, Skarpaas O, Vandvik V (2012) Modeling alpine plant distributions at the landscape scale: do biotic interactions matter? *Ecol Model* 231:1–10
- Milborrow S (2015) Derived from mda:mars by T. Hastie and R. Tibshirani. Uses A. Miller's Fortran utilities with T. Lumley's leaps wrapper. Earth: multivariate adaptive regression splines. R package version 4.2.0. <http://CRAN.R-project.org/package=earth>
- Morueta-Holme N, Blonder B, Sandel B, McGill B, Peet RK, Ott JE, Violle C, Enquist BJ, Jorgensen PM, Svenning JC (2015) A network approach for inferring species association from co-occurrence data. *Ecography*. doi:10.1111/ecog.01892
- Murphy MA, Evans JS, Storfer A (2010) Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology* 91:252–261
- Normand S, Zimmermann NE, Schurr FM, Lischke H (2014) Demography as the basis for understanding and predicting range dynamics. *Ecography* 37:1149–1154

- Olden JD, Joy MK, Death RG (2006) Rediscovering the species in community-wide predictive modeling. *Ecol Appl* 16:1449–1460
- Peet RK (1981) Forest vegetation of the Colorado front range: composition and dynamics. *Vegetatio* 45:3–75
- Pellissier L, Pradervand JN, Pottier J, Dubuis A, Maiorano L, Guisan A (2012) Climate-based empirical models show biased predictions of butterfly communities along environmental gradients. *Ecography* 35:684–692
- Peterson TA, Papes M, Eaton M (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* 30:550–560
- Pollock LJ, Tingley R, Morris WK, Golding N, O'Hara RB, Parris KM, Vesk PA, McCarthy MA (2014) Understanding co-occurrence by modeling species simultaneously with a joint species distribution model (JSDM). *Methods Ecol Evol* 5:397–406
- R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, URL <http://www.R-project.org/>
- Rehfeldt GL (2006) A spline model of climate for the Western United States. General Technical Report RMRS-GTR-165. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins
- Renkonen O (1938) Statistisch-ökologische Untersuchungen über die terrestrische Käferwelt der finnischen Bruchmoore. *Annales Zoologici Societatis Zoologicae-Botanicæ Fennicae 'Vanamo'* 6:1–231
- Roberts DW, Cooper SV (1989) Concepts and techniques of vegetation mapping. In: Land classifications based on vegetation: applications for resource management. USDA Forest Service GTR INT-257, Ogden, pp 90–96
- Rödder D, Engler JO (2011) Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. *Glob Ecol Biogeogr* 20:915–927
- Rouget M, Richardson DM, Lavorel S, Vayreda J, Gracia C, Milton SJ (2001) Determinants of distribution of six *Pinus* species in Catalonia, Spain. *J Veg Sci* 12:491–502
- Sahney S, Benton MJ (2008) Recovery from the most profound mass extinction of all time. *Proc R Soc B* 275:759–765
- Schubert JK, Bottjer DJ (1995) Aftermath of the Permian-Triassic mass extinction event: paleoecology of lower triassic carbonates in the western USA. *Palaeogeogr Palaeoclimatol Palaeoecol* 116:1–39
- Smith WB (2002) Forest inventory and analysis: a national inventory and monitoring program. *Environ Pollut* 116:S233–S242
- Suttle KB, Thomsen MA, Power ME (2007) Species interactions reverse grassland responses to changing climate. *Science* 315:640–642
- Svenning JC, Gravel D, Holt RD, Schurr FM, Thuiller W, Münkemüller T, Schiffers KH, Dullinger S, Edwards TC Jr, Hickler T, Higgins SI, Nabel JEMS, Pagel J, Normand S (2014) The influence of interspecific interactions on species range expansion rates. *Ecography* 37:1198–1209
- Swab RM, Regan HM, Matthies D, Becker U, Bruun HH (2015) The role of demography, intra-species variation, and species distribution models in species' projections under climate change. *Ecography* 38:221–230
- Tylianakis JM, Didham RK, Bascompte J, Wardle DA (2008) Global change and species interactions in terrestrial ecosystems. *Ecol Lett* 11:1351–1363
- Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62:2868–2883
- Webb CO, Ackerly DD, McPeck MA, Donoghue MJ (2002) Phylogenies and community ecology. *Annu Rev Ecol Syst* 33:475–505
- Wisn MS, Pottier J, Kissling WD, Pellissier L, Lenoir J, Damgaard CF, Dormann CF, Forchhammer MC, Grytnes JA, Guisan A, Heikkinen RK, Høye TT, Kühn I, Luoto M, Maiorano L, Nilsson MC, Normand S, Öckinger E, Schmidt NM, Termansen M, Timmermann A, Wardle DA, Aastrup P, Svenning JC (2013) The role of biotic interactions in shaping distributions and realized assemblages of species: implications for species distribution modelling. *Biol Rev* 88:15–30
- Woodall CW, Oswalt CM, Westfall JA, Perry CH, Nelson MD, Finley AO (2010) Selecting tree species for testing climate change migration hypotheses using forest inventory data. *For Ecol Manag* 259:778–785
- Worth JRP, Williamson GJ, Sakaguchi S, Nevill PG, Jorden GJ (2014) Environmental niche modelling fails to predict Last Glacial Maximum refugia: niche shifts, microrefugia or incorrect palaeoclimate estimates? *Glob Ecol Biogeogr* 23:1186–1197
- Woudenberg SW, Conkling BL, O'Connell BM, LaPoint EB, Turner JA, Waddell KL (2010) The forest inventory and analysis database: database description and user manual version 4.0 for phase 2. General Technical Report RMRS-GTR-245. USDA Forest Service Rocky Mountain Research Station, Fort Collins