

The limits of direct community modeling approaches for broad-scale predictions of ecological assemblage structure



Chia-Ying Ko^{a,b,c,d,e,*}, Oswald J. Schmitz^d, Walter Jetz^{c,**}

^a Institute of Fisheries Science, National Taiwan University, Taipei, Taiwan

^b Department of Life Science, National Taiwan University, Taipei, Taiwan

^c Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

^d School of Forestry and Environmental Studies, Yale University, New Haven, CT, USA

^e Research Center for Environmental Changes, Academia Sinica, Taipei, Taiwan

ARTICLE INFO

Article history:

Received 9 February 2016

Received in revised form 12 July 2016

Accepted 20 July 2016

Available online xxx

Keywords:

Community assembly

Stacked-species distribution models

Macroecological community model

Dietary traits

Species prevalence

Biodiversity

ABSTRACT

Two distinct modeling approaches are often used when predicting biodiversity patterns: stacking of species models (predict first, assemble later), and directly modeling a characteristic of a community such as species richness (assemble first, predict later). The relative utility of these two approaches for biogeographic, macroecological and global change analyses is uncertain. Here we compared the two approaches by predicting current-day avian dietary guild structure of assemblages worldwide. We found that the stacked-species modeling approach consistently predicted the geographic distribution of observed dietary guilds better than a direct community modeling approach. The exception was for plant-eating birds, especially frugivores, which are expected to have particularly strong climatic constraints on their diversity and distributions. Assemblage-level biodiversity patterns predicted by community-based modeling approaches, such as the stacked-species and direct community modeling approaches in this study, offer a means to help guide conservation decisions for determining environmental suitability and analyzing diversity hotspots. However, our results generally caution against the widespread use of direct community modeling approach at the large spatial extents for predicting species assemblages.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

A major focus in biogeography and macroecology is to accurately model and predict geographic variation in the composition and attributes of biological communities. This pursuit is particularly relevant for projecting how communities and their many functions may be perturbed under impending climate change (Jetz et al., 2007; Dawson et al., 2011; Bellard et al., 2012). The potential constraints shown on community structure arising from assembly rules limiting species' coexistence have long been realized (Weiher and Keddy, 1999; Ackerly and Cornwell, 2007). But, how such constraints impinge on our ability to predict community structure and ecosystem functioning and how consistently they are shown at different spatial scales, especially in a broad-scale climate change future, still remain uncertain (Thuiller et al., 2003; Baselga and Araújo, 2009; Belmaker and Jetz, 2013). If prevalent,

assembly rules may under human-driven global change significantly affect the re-aggregation of communities and thus change the individual responses of species. Such community-level constraints on future species distributions would have significant consequences for the appropriate modeling approaches (Araújo and Luoto, 2007; Heikkinen et al., 2007; Baselga and Araújo, 2009).

Communities and their attributes (such as total species richness, functional indices or prevalence of a given group) are commonly modeled using a *stacked-species* community modeling approach that first models the spatial distributions of each species and then derives community attributes at different locations by aggregating the predicted individual species distributions (Guisan and Zimmermann, 2000; Ferrier and Guisan, 2006; Dubuis et al., 2011). This kind of approach offers detailed information on the exact identity of the species in the predicted assemblages and has often led to accurate predictions of community attributes for small and spatially restricted datasets (Leathwick et al., 2006; Elith and Leathwick, 2007; Algar et al., 2009; Baselga and Araújo, 2009; Chapman and Purse, 2011; Dubuis et al., 2011). However, by assuming that species assemble independently of each other, this approach ignores potential interrelationships among species comprising a community. One simple alternative is a *direct community* modeling

* Correspondence to: C.-Y. Ko, Institute of Fisheries Science, National Taiwan University, Taipei, Taiwan.

** Correspondence to: W. Jetz, Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA.

E-mail addresses: chiayingko@gate.sinica.edu.tw (C.-Y. Ko), oswald.schmitz@yale.edu (O.J. Schmitz), walter.jetz@yale.edu (W. Jetz).

approach that statistically associates aggregate community attributes with environmental variables (Ferrier and Guisan, 2006; Baselga and Araújo, 2009; Dubuis et al., 2011). Following in the footsteps of macroecological models of species richness (Brown, 1981; Jetz and Rahbek, 2002), this approach is straightforward to use with large, multi-species datasets and putatively able to accommodate species interdependencies within communities because it accommodates spatial co-occurrences of species to infer community-level patterns (Baselga and Araújo, 2009; Chapman and Purse, 2011). These factors combined with the approach's ability to quickly synthesize complex data for conservation applications have led to notable popularity (Ferrier and Guisan, 2006; Algar et al., 2009; Chapman and Purse, 2011).

An increasing number of studies are considering community attributes such as total species richness, proportional richness or community functional attributes as a response variable for current-day predictions and projections under climate change (Lemoine and Böhning-Gaese, 2003; Kissling et al., 2007; Steinmann et al., 2009; Kissling et al., 2010; Sommer et al., 2010; Chown et al., 2012; Dubuis et al., 2013; Venevskaia et al., 2013; D'Amen et al., 2015; Distler et al., 2015). One specific suggestion is that geographic patterns of energetic “capacity constraints”, the limits on species richness imposed by available energy, may be re-arranged under climate change and affect future community assemblages (Sommer et al., 2010; Chown et al., 2012). Biotic associations between taxonomic groups show sensitivity when forecasting future biodiversity (Kissling et al., 2007; Kissling et al., 2010). Accounting for trait data through an emerging trait biogeographic perspective reveals heterogeneous taxonomic partitioning of assemblages (Kissling et al., 2012; Ko et al., 2014). So far, community-based models are increasingly used to examine attributes of communities.

Despite this popularity, the appeal of potentially addressing biotic constraints in projecting communities under climate change, and an already existing debate about the choice of the *stacked-species* versus *direct community* modeling approach (Ferrier and Guisan, 2006; Leathwick et al., 2006; Baselga and Araújo, 2009; Chapman and Purse, 2011; Dubuis et al., 2011; Guisan and Rahbek, 2011; Distler et al., 2015), empirical evaluations are rare. Algar et al. (2009) found good performance of *direct community* compared to *stacked-species* modeling approaches for predicting changes in butterfly diversity in Canada. And for a study along a Swiss elevational gradient, Dubuis et al. (2011) showed that, while having slightly less predictive ability, only *direct community* modeling approaches were able to recover the specific shape of the elevation–richness relationship (Bonthoux et al., 2013). More recently, Distler et al. (2015) found that *direct community* modeling approaches could provide more accurate estimates of North American bird species richness than *stacked-species* modeling approaches, particularly during the summer season. However, apart from these case studies, the general suitability and performance of the community-level modeling approaches for broader spatial scales and for ecologically partitioned groups remain unclear.

The goal of our study is to provide such a general assessment and resolve the relative appropriateness of different methodologies for modeling patterns of functional community attributes over a broad geographic extent (and the typically concomitant coarse spatial grain) and a large range of species and communities. We use a global dataset of bird dietary preferences and geographic distributions to compare model performance of the *stacked-species* modeling approach with two types of *direct community* modeling approaches, one with total and the other with relative (or proportional) richness of different dietary species groups as response variable. Relative guild species richness, or guild prevalence, is a community attribute with important functional implications and has previously been shown to exhibit strong environmental associations (Kissling et al., 2012). We here ask whether the two types of *direct community* modeling approaches are indeed able to predict the variation in guild assemblage along climatic gradients and, by extension, are suited to address their potential perturbation under

climate change. If weak environmental associations of the response variable or overall limited predictive performance constrain the *direct community* modeling approach, is the *stacked-species* modeling approach able to successfully capture the aggregate response and predict its spatial variation? We here assess this for a range of guild groupings that vary in the respective environmental associations. Given the direct consequences of such changes for the functioning and services of ecosystems (Schmitz et al., 2003; Mooney et al., 2009; Kardol et al., 2010), identifying suitable approaches for modeling these and other community attributes is of strong applied importance and useful for biodiversity conservation and management.

2. Methods

2.1. Species data and environmental predictors

We used a comprehensive database of recent expert-based breeding distributions of all 9993 bird species in the world, spanning a latitudinal range of 60°S to 85°N (see Jetz et al., 2012) and mol.org for sources, taxonomic treatment and individual maps). Marine and pelagic species as well as species with less than four occurrence grid cells were excluded from the analysis, leaving a total of 8472 bird species. We intersected these distributions with a global 110 × 110 km equal area grid (approximately 1° near equator), a spatial resolution that according to recent validation of these sorts of expert-based maps offers sufficiently low false presence rates (Hurlbert and Jetz, 2007). This resulted in 11,079 grid cell assemblages for analysis.

We obtained estimates of the proportional use of each of seven food categories (fruits, nectar, plants, seeds, invertebrates, vertebrates, and carcasses) for each bird species from Wilman et al. (2014) to assign each species to a dietary guild. In a “coarse” dietary classification all species were classified as either primary consumers or high-level consumers based on their main proportional use of summed plant (i.e. fruits, nectar, plants, and seeds) and summed animal (i.e. invertebrates, vertebrates, and scavengers) diets. “Fine” dietary guilds were identified by highest proportional use as frugivores, nectarivores, herbivores, granivores, insectivores, carnivores, or scavengers. Those species exhibiting equal use of multiple summed or individual diets were classified as omnivores (named mixed consumers in the coarse dietary guilds and omnivores in the fine dietary guilds). For similar treatment and additional details see also Ko et al. (2014).

We extracted nine environmental predictor variables (all \log_{10} -transformed) including one topographic and eight climatic predictors over the same 110 km grid. All selected variables are known to exert a strong influence on the distributions of individual species as well as overall and guild species richness (Jetz and Rahbek, 2002; Field et al., 2009; Kissling et al., 2012; Ko et al., 2014). Elevational range was derived from GTOPO30 at 30-arc seconds (approximately 1 km), produced by the USGS (<http://eors.usgs.gov/>). Additionally, mean annual temperature, temperature seasonality (standard deviation of monthly means), mean temperatures of the coldest and warmest months, total annual precipitation, seasonality of precipitation (coefficient of variation of monthly precipitation), and total precipitation in the driest and wettest months were obtained for the 1975–2001 period (representing current conditions, i.e. 2000, in this study) from CRU TS 2.1 (Mitchell and Jones, 2005) with an original spatial resolution of 30-arc minutes (approximately 50 km). All environmental predictor variables at the coarse scale of 110 km grid cells (i.e. the spatial resolution of species distributions) were expressed by the area-weighted average values of fine scale data (i.e. topographic and climatic predictors in approximately 1 km and 50 km resolutions, respectively).

2.2. Models

Considering the success of ensembles of models in reducing both false negative and positive errors in predictions of species

distributions (Araújo and New, 2007), we used information from ensemble modeling instead of picking a single ‘best’ model in this study. To model species distributions using the *stacked-species* modeling approach, we applied seven different species distribution modeling algorithms (Elith et al., 2006) to each of the 8472 bird species, including one classification method (mixture discriminant analysis, MDA), three regression methods (generalized linear models, GLM, generalized additive models, GAM and multivariate adaptive regression splines, maximum entropy, MARS), and three machine learning methods (general boosting method, GBM, random forests, RF and maximum entropy, MAXENT). Among these we selected the four models with the highest AUC (area under the curve) values for further use (see below for AUC calculation). We weighted the predictions of all these four models according to their pre-evaluated AUC model performance (Marmion et al., 2009) to generate a final probabilistic distribution map for each species. Species presences based on maps of current species distributions and absences within 4000 km of a species’ current distribution represented suitable and unsuitable species locations, respectively. Because absences far from species occurrences are less informative than those closer (VanDerWal et al., 2009), absence grid cells were assigned weights that exponentially decreased with distance from the nearest species presence grid cell and used in all seven modeling algorithms. As sampling prevalence has been shown to influence the accuracy of species distribution models (McPherson et al., 2004), equal numbers of species presences and absences were assigned for model calibration except for species with <10 presence grid cells, which were assigned 100 absences. We first ran all seven algorithms with the nine environmental predictor variables. The variables with the high contributions averaged across those seven algorithms were then left in the final modeling. For species with ≥ 10 presence cells, the presence cells were divided into four quadrants based on their coordinates above or below the median latitude and median longitude of all presence localities (Peterson et al., 2006). A random subset of three of quadrants of species presence and absence data was used to calibrate the model in the first run while the remaining one for evaluation (Peterson and Shaw, 2003; Peterson et al., 2006), using a threshold independent receiver operating characteristic (ROC) curve along with the associated AUC; 100% of the available species presence and (selected) absence data were used for making final predictions. For species with <10 presence cells, we used a jackknife validation to select the best-performing modeling algorithms (Pearson et al., 2006). Seven modeling algorithms were evaluated using the jackknife procedure and final predictions were obtained by averaging the modeling algorithms which had a corresponding p value <0.01 in the jackknife procedure. Predicted guilds and total richness were then given as the sum of predicted presence probabilities for each species in each cell (Calabrese et al., 2014).

For the *direct community* modeling approach, we implemented four modeling algorithms: generalized linear models (GLM), generalized additive models (GAM); general boosting method (GBM) and random forests (RF). GLM, GAM, and GBM were implemented as Poisson regressions for observed guild richness as response and as Bernoulli/binomial regression for guild prevalence as response. For GBM and RF we used a maximum of 2000 trees. For each modeling algorithm, beginning with all nine environmental predictors, one predictor variable at a time was sequentially removed using a supervised backward stepwise variable selection procedure. The calibration and evaluation datasets were also obtained by randomly splitting the original dataset into four quadrants, using the same partition way as the *stacked-species* modeling approach. The model with the highest adjusted R^2 and minimum number of predictor variables was selected as the best model for each modeling algorithm. The final *direct community* predictions were obtained by weighting the predictions of the four modeling algorithms according to the R^2 of their best models.

2.3. Statistics

We compared the results provided by the *stacked-species* and *direct community* modeling approaches using predicted prevalence (i.e. relative or proportional richness) of three coarse and eight fine dietary guilds. For this we truncated the richness predictions for each grid cell provided by the *stacked-species* and *community richness* approaches (based on summed probabilities and predicted richness, respectively) to the closest integer values and divided predicted guild richness by predicted total richness. For all modeling approaches, we limited the geographic scope of predictions to cells with non-zero values for the response variable, i.e. those currently occupied by at least one guild member.

We evaluated the relationship between observed and predicted guild prevalence and calculated the t statistic for the difference from a slope equal to 1 (Equation: $t = (S_{op} - 1) / SE$, where S_{op} is the slope of the observed versus predicted regression lines and SE is the standard error of the slope), root mean squared deviation and R^2 value of the observed versus predicted regression lines (Piñeiro et al., 2008). Modeling algorithms were implemented within the “BIOMOD” (Thuiller et al., 2009) and “dismo” R packages and statistical analyses were performed using R version 2.12.0 (Development Core Team, 2011).

3. Results

Among the 8472 species analyzed, the seven individual-species level modeling algorithms, RF, GBM, MAXENT and GAM generally showed the highest predictive accuracy for species with ≥ 10 presences, while GLM, MARS, MDA and GBM performed well for species with <10 presences (Fig. A.1 in Appendix A). Both at the coarse and fine ecological resolutions, the *stacked-species* and *direct community* modeling approaches varied substantially in their ability to predict the observed prevalence, i.e. the relative or proportional richness, of avian dietary guilds worldwide (Table 1; Figs. 1 and 2). The *stacked-species* modeling approach consistently yielded predictions for guild prevalence that were very close to those observed, as indicated by slope estimates very close to 1 (low absolute t values for a deviation from 1) and a low root mean squared deviations (Table 1). The two types of *direct community* modeling approaches using richness and prevalence as response variable showed mostly very weak performance that was only in some cases

Table 1

Variation in the observed and predicted prevalence (proportional richness) of avian dietary guilds, modeled with *stacked-species* vs. *direct community* modeling approaches. Values in the table are t statistic (t) for the difference from a slope of 1 and root mean squared deviation (RMSD) from linear regression of predicted vs. observed prevalence, calculated from the predictions by the *stacked-species*, *community richness* and *prevalence* modeling approaches.

Dietary guild	Stacked-species approach		Direct community approach				
			Richness model		Prevalence model		
	t	RMSD	t	RMSD	t	RMSD	
Coarse	Primary consumers	13.8	0.015	-391.1	0.04	-424.6	0.04
	High-level consumers	7.6	0.017	-320.2	0.046	-318.3	0.047
	Mixed consumers	-1.5	0.008	-226	0.02	-222.1	0.02
Fine	Frugivores	25.4	0.004	-37	0.014	-36.1	0.014
	Nectarivores	-18.5	0.003	-89.4	0.013	-99.8	0.013
	Herbivores	12.7	0.013	-139.4	0.029	-92.5	0.026
	Granivores	37.3	0.012	-150	0.029	-128.7	0.028
	Insectivores	4.2	0.025	-145.4	0.048	-113.2	0.046
	Carnivores	45.6	0.02	-162.3	0.044	-172.3	0.045
	Scavengers	-54.8	0.008	-169.3	0.011	-134.2	0.01
	Omnivores	23.6	0.015	-122	0.042	-131.4	0.041

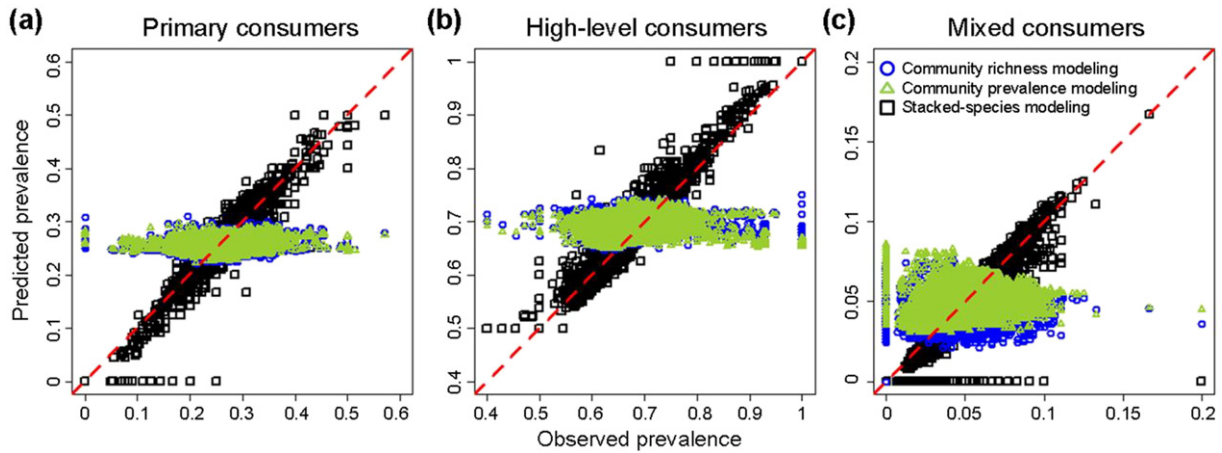


Fig. 1. Relationships between observed and predicted present (2000) relative richness/prevalence (proportions of total) of three coarse dietary guilds, i.e. primary consumers (a), high-level consumers (b) and mixed consumers (c), using *stacked-species* and *direct community* modeling approaches for the global bird fauna (8472 species; 11,079,110 × 110 km grid cells worldwide). Red dashed lines represent a 1:1 relationship. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

close to any of the *stacked-species* results (e.g. frugivores; Table 1). Both *community richness* and *prevalence* modeling approaches were mostly unable to predict the strong spatial variation in guild prevalence and tended to strongly over-predict relative guild richness in low-prevalence assemblages, and under-predict it in high-prevalence assemblages (Figs. 1 and 2).

In the *stacked-species* modeling approach predictive performance was similarly high for all guilds (Fig. 3, Fig. A.2 in Appendix A), with only a slight lower goodness of fit for scavengers and the three coarse guilds. This contrasts strongly with the *direct community* modeling approach, in which fit was strongest for frugivores and herbivores, but then deteriorated rapidly for all other guilds and especially the coarse

guild groupings. Fits differed little among the two *direct community* modeling approaches. In no case was their fit superior to that of the *stacked-species* approach, but the *direct community* modeling approach did offer close to comparable strengths for predicting plant-eating birds.

Although greatly variable among the lower-richness assemblages, (“Observation” panels in Figs. 4 and 5), globally observed guild prevalence was surprisingly uniform along the global richness gradient. For example, the global prevalence of primary and higher-level consumers centered around 0.3 and 0.7, respectively. The *stacked-species* modeling approach was able to capture both the variation and magnitude of prevalence along the richness gradient reasonably well, whereas the *direct community* modeling approach misrepresented the observed

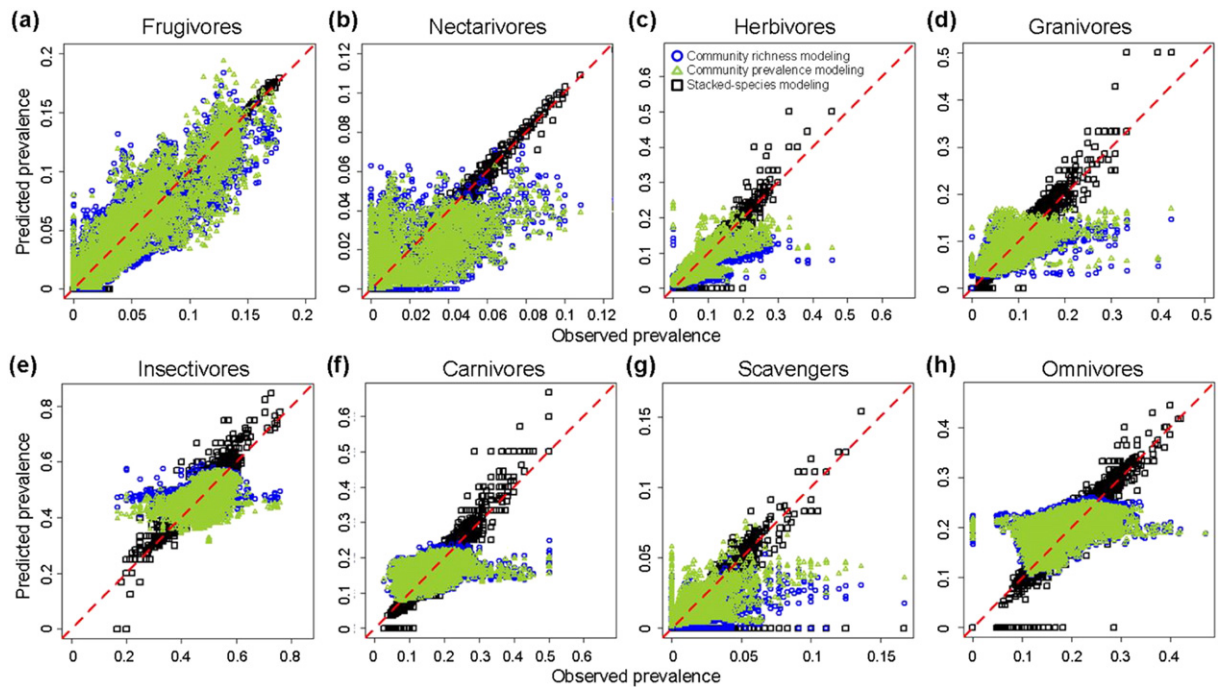


Fig. 2. Relationships between observed and predicted present (2000) relative richness/prevalence (proportions of total) of eight fine dietary guilds, i.e. frugivores (a), nectarivores (b), herbivores (c), granivores (d), insectivores (e), carnivores (f), scavengers (g) and omnivores (h), using *stacked-species* and *direct community* modeling approaches for the global bird fauna (8472 species; 11,079,110 × 110 km grid cells worldwide). Red dashed lines represent a 1:1 relationship. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

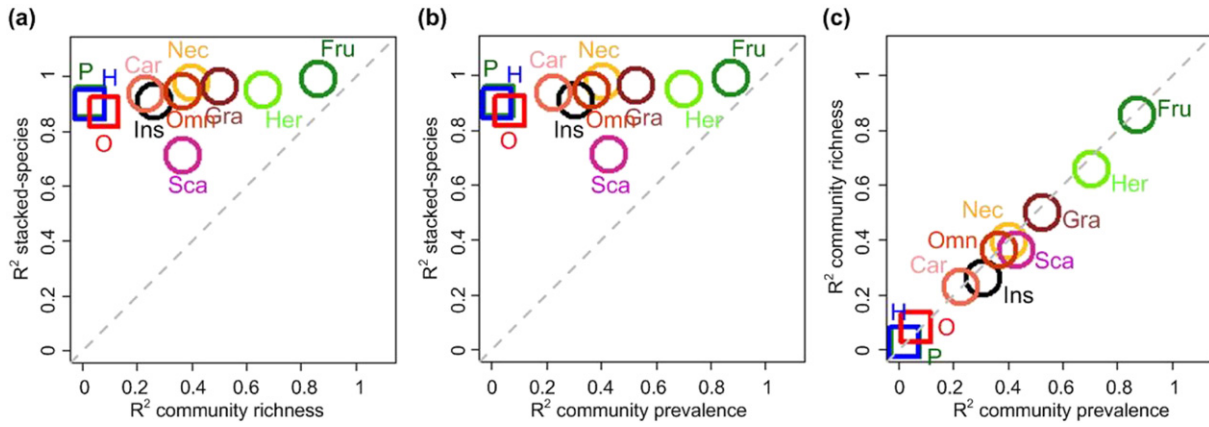


Fig. 3. Scatterplots of R-squared values from linear regression of predicted vs. observed prevalence for three coarse and eight fine dietary guilds using *stacked-species* and *direct community* modeling approaches. Dashed lines represent a 1:1 relationship. Abbreviations of the dietary guilds: P, primary consumers; H, high-level consumers; M, mixed consumers; Fru, Frugivores; Nec, Nectarivores; Her, Herbivores; Gra, Granivores; Ins, Insectivores; Car, Carnivores; Sca, Scavengers; Omn, Omnivores.

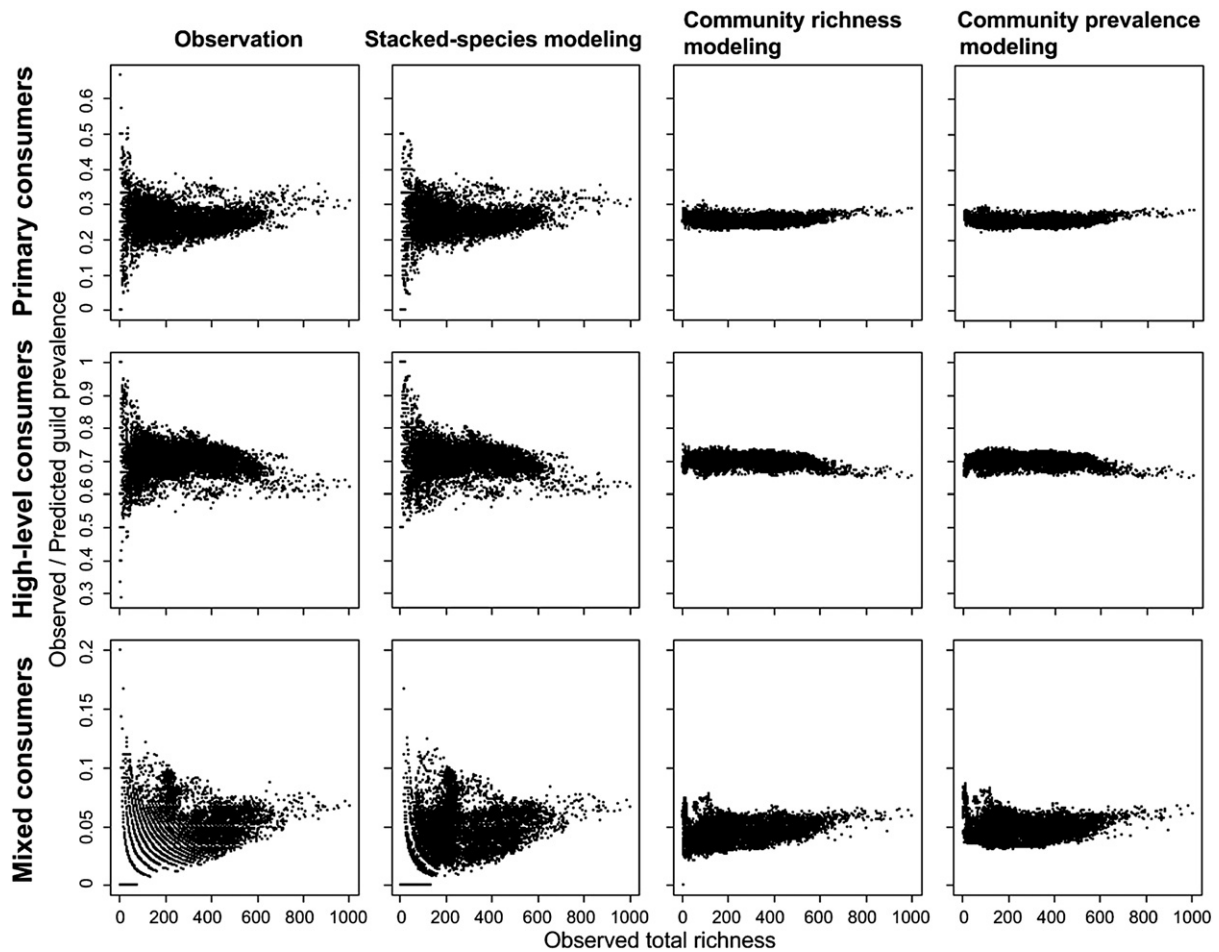
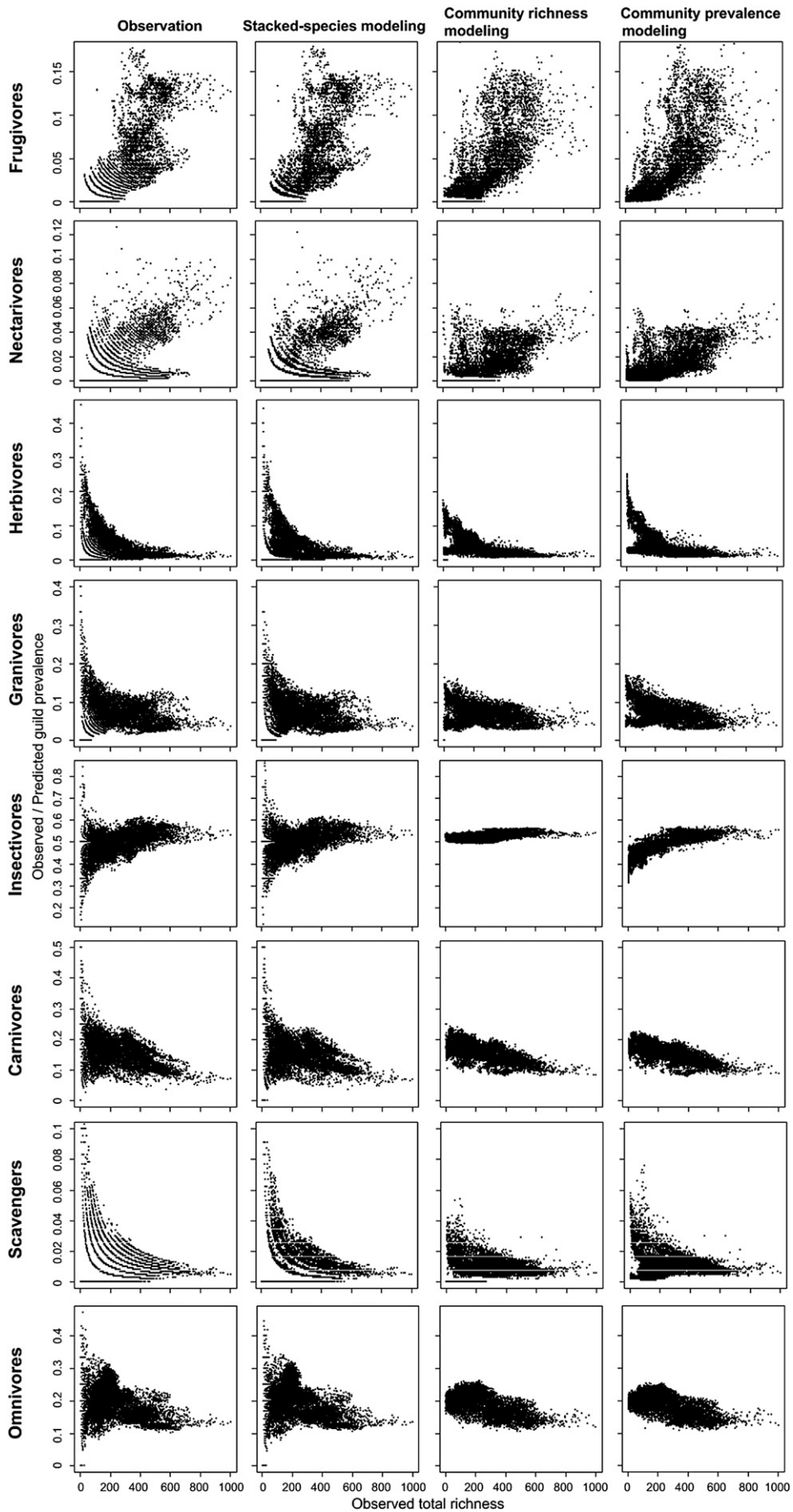


Fig. 4. Relationships between current-day observed total richness and predicted guild prevalence from different modeling approaches across three coarse dietary guilds. Each point is the species richness and the proportion of a guild in a grid cell.

prevalence, particularly in the low richness assemblages. These patterns were similar for most fine dietary guilds (Fig. 5), with the exception of plant-eating guilds, especially frugivores, where the guild prevalence strongly increased with increasing total richness. The exceptional

pattern for frugivores was also evident in the direct assessment of predicted vs. observed prevalence (Table 1; Figs. 2 and 3), where the two types of *direct community* modeling approaches, i.e. *community richness* and *prevalence* models, also offered a surprisingly reliable prediction.

Fig. 5. Relationships between current-day observed total richness and predicted guild prevalence from different modeling approaches across eight fine dietary guilds. Each point is the species richness and the proportion of a guild in a grid cell. Both in the observations and predictions, frugivores show increasing guild prevalence with the increasing total richness while most fine dietary guilds maintain prevalence across the total richness gradient especially under high species richness.



4. Discussion

Statistical approaches to predict species assemblages in communities are numerous and highly varied. Comparisons between two prominent approaches—*stacked-species* versus *direct community* modeling—have produced mixed conclusions about their relative predictive ability (Ferrier and Guisan, 2006; Leathwick et al., 2006; Baselga and Araújo, 2009; Dubuis et al., 2011; Guisan and Rahbek, 2011; Distler et al., 2015). Our comparison found that the *direct community* modeling approach showed as good performance as the *stacked-species* modeling approach only for frugivores. Previous analyses have concluded that *direct community* modeling approaches would be favored when species share similar environmental associations and face strong assembly constraints (Baselga and Araújo, 2009; Chapman and Purse, 2011; Dubuis et al., 2011). Our analysis does not support this conclusion, despite the fact that the ecological group, i.e. dietary guilds, analyzed in this study share similar environmental conditions and assembly constraints. Instead, the *stacked-species* modeling approach predicted the prevalence of guilds observed today equally or better than the two types of *direct community* modeling approaches. It has been claimed that a drawback of *stacked-species* modeling approach is that it overpredicts species richness, as in the case of analyses of Swiss plants and New World birds at resolutions of 25 m and 1° latitude/longitude, respectively (Guisan and Rahbek, 2011), because the thresholding approach used to relate predicted probabilities to occurrence may lead to bias (Guillera-Aroita et al., 2015).

We offer some reasons why the two types of *direct community* modeling approach may not be as suitable as the *stacked-species* modeling approach. First, the *direct community* modeling approach only estimates one aggregate response – environment association for all species in a grouping or guild. It therefore cannot appropriately capture the many variations in environmental associations of each species with similar statistical power as the *stacked-species* modeling approach, which fits many individual response curves. Second, in the *direct community* modeling approach, the response variables are strongly influenced by wide-ranging species, as they dominate grid cell records (Jetz and Rahbek, 2002). This overwhelms the statistical signal attribute able to species with more narrow ranges. The *stacked-species* modeling approach instead represents the environmental associations of all species more uniformly and appropriately, which in turn may contribute to their stronger predictive performance. Third, the co-occurrence of species in communities and resulting patterns of absolute or relative richness may be attributed more to regional factors or habitat filters than species interactions (Belmaker and Jetz, 2012; Jetz and Fine, 2012), especially at grains coarser than those at which species interactions may play out (Cooper and Belmaker, 2010), as is likely the case in our study. As a consequence, environmental correlates may not capture anything meaningful about species' co-occurrences, or at least may not do so equally for all species. We note that the best predictive performance of the *direct community* modeling approach was found for guilds that are known to face strong environmental constraints in their regional diversification or local coexistence; e.g. for obligate frugivores the requirement for fruits has obvious strong links with year-round high temperature and productivity (Kissling et al., 2009). Fourth, the ecological guild characterizations used here may yet be too broad to appropriately capture groups of interacting species suited for the one-group perspective of a *direct community* modeling approach. Our results did suggest better performance (i.e. lower absolute *t* values) for a few of the fine-grained guilds than the coarse-grain guilds.

Interestingly, we found relatively uniform guild prevalence along the gradient from low to high richness assemblages for most guilds, and especially so for high-richness assemblages. This might suggest there are “natural” restrictions on co-occurrence among environments, i.e. there is a maximum species richness attainable in a particular

environment (environmental load). This may imply that the prevalence of different dietary guilds is maintained with a dynamic balance between the immigration of new species into the same or different guild and the extinction of species already present in a guild. This also may suggest that low richness assemblages in today's environment may not have yet reached a putative maximum environmental load.

Community-based conservation as a concept is relatively new and has burst into the center of the global conservation discourse in the past three decades. This concept espouses the integration of ecosystem management and human well-being and the requirements for linking protection and institutions at multiple levels and scales (Berkes, 2007; Gruber, 2011; Ruiz-Mallen and Corbera, 2013). Assemblage-level biodiversity patterns predicted by community-based modeling approaches, such as the *stacked-species* and *direct community* modeling approaches in this study, offer a means to help guide conservation decisions for determining environmental suitability and analyzing diversity hotspots (Ferrier and Guisan, 2006; Guisan and Rahbek, 2011; Guisan et al., 2013). Moreover, incorporating community functional attributes, e.g. dietary guilds, with environmental responses in the modeling approaches reveals additional ecological insights. Besides, when species assemblages co-occur in range, habitat, and management tolerance, it is more able to use indicator species to monitor environmental changes, assess the efficacy of management, and provide warning signals for any impending ecological shifts, which further enhances the effectiveness of conservation projects (Carignan and Villard, 2002). In summary, our results highlight limitations for the use of *direct community* modeling approach for predicting geographic patterns in biodiversity. They support *stacked-species* modeling approach or methods combining *direct* and *stacked* elements, such as an emerging new generation of joint distribution models that combine co-occurrence with community signals (Pollock et al., 2014). The weak performance of *direct community* modeling approaches for current-day predictions calls into question their additional application for global biodiversity projections under global change. For some specific groups with apparent strong climatic constraints, the *direct community* modeling approach, including community richness and prevalence modeling, does appear to offer good predictive ability while requiring much fewer parameters than the *stacked-species* modeling approach. And at spatial grains much finer than those analyzed here (e.g. sub-kilometer), constraints arising from resource diversity or species interactions may favor methods using community attributes as response. But overall, our findings suggest caution against the utility of *direct community* modeling approach at the large spatial extents usually necessary to fully capture the environmental niches of groups of species.

Acknowledgements

We thank Jonathan Belmaker, Petr Keil, and Morgane Barbet-Massin for constructive suggestions of our work on modeling approaches and Antoine Guisan, Jui-Yi Hsieh and Ming-Fai Chow for helpful comments. We also thank two anonymous reviewers that made lots of insightful comments which improved the contents of this manuscript. We greatly appreciate financial support by multiple postdoc research fellowships funded by the Delta Electronics Foundation, Yu Chi-Chung Culture and Educational Foundation, and Taiwanese National Science Council (100IP999900127) to C.-Y. Ko and NSF grants DBI0960550 and DEB1026764 and NASA Biodiversity Grant NNX11AP72G to W. Jetz.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.biocon.2016.07.026>.

References

- Ackerly, D.D., Cornwell, W.K., 2007. A trait-based approach to community assembly: Partitioning of species trait values into within- and among-community components. *Ecol. Lett.* 10, 135–145.
- Algar, A.C., Kharouba, H.M., Young, E.R., Kerr, J.T., 2009. Predicting the future of species diversity: Macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography* 32, 22–33.
- Araújo, M.B., Luoto, M., 2007. The importance of biotic interactions for modelling species distributions under climate change. *Glob. Ecol. Biogeogr.* 16, 743–753.
- Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42–47.
- Baselga, A., Araújo, M.B., 2009. Individualistic vs community modelling of species distributions under climate change. *Ecography* 32, 55–65.
- Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., Courchamp, F., 2012. Impacts of climate change on the future of biodiversity. *Ecol. Lett.* 15, 365–377.
- Belmaker, J., Jetz, W., 2012. Regional pools and environmental controls of vertebrate richness. *Am. Nat.* 179, 512–523.
- Belmaker, J., Jetz, W., 2013. Spatial scaling of functional structure in bird and mammal assemblages. *Am. Nat.* 181, 464–478.
- Berkes, F., 2007. Community-based conservation in a globalized world. *PNAS* 104, 15188–15193.
- Bonthoux, S., Baselga, A., Balent, G., 2013. Assessing community-level and single-species models predictions of species distributions and assemblage composition after 25 years of land cover change. *PLoS One* 8, e54179.
- Brown, J.H., 1981. Two decades of homage to Santa Rosalia: toward a general theory of diversity. *Am. Zool.* 21, 877–888.
- Calabrese, J.M., Certain, G., Kraan, C., Dormann, C.F., 2014. Stacking species distribution models and adjusting bias by linking them to macroecological models. *Glob. Ecol. Biogeogr.* 23, 99–112.
- Carignan, V., Villard, M.-A., 2002. Selecting indicator species to monitor ecological integrity: review. *Environ. Monit. Assess.* 78, 45–61.
- Chapman, D.S., Purse, B.V., 2011. Community versus single-species distribution models for British plants. *J. Biogeogr.* 38, 1524–1535.
- Chown, S.L., le Roux, P.C., Ramaswila, T., Kalwij, J.M., Shaw, J.D., McGeoch, M.A., 2012. Climate change and elevational diversity capacity: do weedy species take up the slack? *Biol. Lett.* 9, 20120806.
- Cooper, N., Belmaker, J., 2010. Opinion: Habitat data resolution and the detection of species interactions. *Front. Biogeogr.* 2, 46.
- D'Amen, M., Dubuis, A., Fernandes, R.F., Pottier, J., Pellissier, L., Guisan, A., 2015. Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. *J. Biogeogr.* 42, 1255–1266.
- Dawson, T.P., Jackson, S.T., House, J.J., Prentice, I.C., Mace, G.M., 2011. Beyond predictions: biodiversity conservation in a changing climate. *Science* 332, 53–58.
- Development Core Team, R., 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Distler, T., Schuetz, J.G., Velasquez-Tibata, J., Langham, G.M., 2015. Stacked species distribution models and macroecological models provide congruent projections of avian species richness under climate change. *J. Biogeogr.* 42, 976–988.
- Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P., Guisan, A., 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Divers. Distrib.* 17, 1122–1131.
- Dubuis, A., Rossier, L., Pottier, J., Pellissier, L., Vittoz, P., Guisan, A., 2013. Predicting current and future spatial community patterns of plant functional traits. *Ecography* 36, 1158–1168.
- Elith, J., Leathwick, J.R., 2007. Predicting species distribution from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Divers. Distrib.* 13, 265–275.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J., M., M.C., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151.
- Ferrier, S., Guisan, A., 2006. Spatial modelling of biodiversity at the community level. *J. Appl. Ecol.* 43, 393–404.
- Field, R., Hawkins, B.A., Cornell, H.V., Currie, D.J., Diniz-Filho, J.A.F., Guégan, J.-F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M., Turner, J.R.G., 2009. Spatial species-richness gradients across scales: A meta-analysis. *J. Biogeogr.* 36, 132–147.
- Graham, R.W., Lundelius Jr., E.L., Graham, M.A., Schroeder, E.K., Toomey III, R.S., Anderson, E., Barnosky, A.D., Burns, J.A., Churcher, C.S., Grayson, D.K., Guthrie, R.D., Harington, C.R., Jefferson, G.T., Martin, L.D., McDonald, H.G., Morlan, R.E., Semken Jr., J.A., Webb, S.D., Werdelin, L., Wilson, M.C., 1996. Spatial response of mammals to late quaternary environmental fluctuations. *Science* 272, 1601–1606.
- Gruber, J.S., 2011. Perspectives of effective and sustainable community-based natural resource management: an application of Q methodology to forest projects. *Conserv. Soc.* 9, 159–171.
- Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., McCarthy, M.A., Tingley, R., Wintle, B.A., 2015. Is my species distribution model fit for purpose? Matching data and models to applications. *Glob. Ecol. Biogeogr.* 24, 276–292.
- 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.
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135, 147–186.
- Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P., Buckley, Y.M., 2013. Predicting species distributions for conservation decisions. *Ecol. Lett.* 16, 1424–1435.
- Heikkinen, R.K., Luoto, M., Virkkala, R., Pearson, R.G., Korber, J.-H., 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Glob. Ecol. Biogeogr.* 16, 754–763.
- Hurlbert, A.H., Jetz, W., 2007. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *PNAS* 104, 13384–13389.
- Jackson, S.T., Overpeck, J.T., 2000. Responses of plant populations and communities to environmental changes of the late quaternary. *Paleobiology* 26, 194–220.
- Jetz, W., Fine, P.V., 2012. Global gradients in vertebrate diversity predicted by historical area-productivity dynamics and contemporary environment. *PLoS Biol.* 10, e1001292.
- Jetz, W., Rahbek, C., 2002. Geographic range size and determinants of avian species richness. *Science* 297, 1548–1551.
- Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K., Mooers, A.O., 2012. The global diversity of birds in space and time. *Nature* 491, 444–448.
- Jetz, W., Wilcove, D.S., Dobson, A.P., 2007. Projected impacts of climate and land-use change on the global diversity of birds. *PLoS Biol.* 5, 1211–1219.
- 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, 767–778.
- Kissling, W.D., Böhning-Gaese, K., Jetz, W., 2009. The global distribution of frugivory in birds. *Glob. Ecol. Biogeogr.* 18, 150–162.
- Kissling, W.D., Böhning-Gaese, K., Jetz, W., 2010. Woody plants and the prediction of climate-change impacts on bird diversity. *Philos. Trans. R. Soc. B* 365, 2035–2045.
- Kissling, W.D., Rahbek, C., Böhning-Gaese, K., 2007. Food plant diversity as broad-scale determinant of avian frugivore richness. *Proc. R. Soc. Lond. B Biol. Sci.* 274, 799–808.
- Kissling, W.D., Sekercioglu, C.H., Jetz, W., 2012. Bird dietary guild richness across latitudes, environments and biogeographic regions. *Glob. Ecol. Biogeogr.* 21, 328–340.
- Ko, C.-Y., Schmitz, O.J., Barbet-Massin, M., Jetz, W., 2014. Dietary guild composition and disaggregation of avian assemblages under climate change. *Glob. Change Biol.* 20, 790–802.
- Leathwick, J.R., 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.
- Lemoine, N., Böhning-Gaese, K., 2003. Potential impact of global climate change on species richness of long-distance migrants. *Conserv. Biol.* 17, 577–586.
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2009. Evaluation of consensus methods in predictive species distribution modelling. *Divers. Distrib.* 15, 59–69.
- McPherson, J.M., Jetz, W., Rogers, D.J., 2004. The effects of species' range sizes on the accuracy of distribution models: Ecological phenomenon or statistical artefact? *J. Appl. Ecol.* 41, 811–823.
- Mitchell, T.D., Jones, P.D., 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.* 25, 693–712.
- Mooney, H., Larigauderie, A., Cesario, M., Elmquist, T., Hoegh-Guldberg, O., Lavorel, S., Mace, G.M., Palmer, M., Scholes, R., Yahara, T., 2009. Biodiversity, climate change, and ecosystem services. *Curr. Opin. Environ. Sustain.* 1, 46–54.
- Pearson, R.G., Raxworthy, C.J., Nakamura, M., Peterson, A.T., 2006. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *J. Biogeogr.* 34, 102–117.
- Peterson, A.T., Shaw, J.J., 2003. *Lutzomyia* vectors for cutaneous leishmaniasis in southern Brazil: ecological niche models, predicted geographic distributions, and climate change effects. *Int. J. Parasitol.* 33, 919–931.
- Peterson, A.T., Papeş, M., Eaton, M., 2006. Transferability and model evaluation in ecological niche modeling: A comparison of GARP and Maxent. *Ecography* 30, 550–560.
- Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J., 2008. How to evaluate models: observed vs. predicted or predicted vs. observed? *Ecol. Model.* 216, 316–322.
- Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Veski, P.A., McCarthy, M.A., 2014. Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods Ecol. Evol.* 5, 397–406.
- Ruiz-Mallen, I., Corbera, E., 2013. Community-based conservation and traditional ecological knowledge: implications for social-ecological resilience. *Ecol. Soc.* 18, 12.
- Schmitz, O.J., Post, E., Burns, C.E., Johnston, K.M., 2003. Ecosystem responses to global climate change: moving beyond color-mapping. *Bioscience* 53, 1199–1205.
- Sommer, J.H., Kreft, H., Kier, G., Jetz, W., Mutke, J., Barthlott, W., 2010. Projected impacts of climate change on regional capacities for global plant species richness. *Proc. R. Soc. Lond. B Biol. Sci.* 277, 2271–2280.
- Steinmann, K., Linder, H.P., Zimmermann, N.E., 2009. Modelling plant species richness using functional groups. *Ecol. Model.* 220, 962–967.
- Thuiller, W., Araújo, M.B., Lavorel, S., 2003. Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. *J. Veg. Sci.* 14, 669–680.

- Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography* 32, 369–373.
- VanDerWal, J., Shoo, L.P., Graham, C., Williams, S.E., 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecol. Model.* 220, 589–594.
- Venevskaja, I., Venevsky, S., Thomas, C.D., 2013. Projected latitudinal and regional changes in vascular plant diversity through climate change: Short-term gains and longer-term losses. *Biodivers. Conserv.* 22, 1467–1483.
- Weihner, E., Keddy, P.A., 1999. *Ecological Assembly Rules : Perspectives, Advances, Retreats*. Cambridge University Press, Cambridge.
- Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M.M., Jetz, W., 2014. *EltonTraits 1.0: species-level foraging attributes of the world's birds and mammals*. *Ecology* 95, 2027.