*Ecology*, 94(9), 2013, pp. 1913–1919


# To mix or not to mix: comparing the predictive performance of mixture models vs. separate species distribution models

Francis K. C. Hui,[1,2,5] David I. Warton,[1,3] Scott D. Foster,[2,4] and Piers K. Dunstan[4]

[1]*School of Mathematics and Statistics, The University of New South Wales, Sydney, Australia*
[2]*CSIRO Mathematics, Informatics and Statistics, Canberra, Australia*
[3]*Evolution and Ecology Research Centre, The University of New South Wales, Sydney, Australia*
[4]*CSIRO Wealth from Oceans Flagship, Floreat Park, Australia*

*Abstract*. Species distribution models (SDMs) are an important tool for studying the patterns of species across environmental and geographic space. For community data, a common approach involves fitting an SDM to each species separately, although the large number of models makes interpretation difficult and fails to exploit any similarities between individual species responses. A recently proposed alternative that can potentially overcome these difficulties is species archetype models (SAMs), a model-based approach that clusters species based on their environmental response. In this paper, we compare the predictive performance of SAMs against separate SDMs using a number of multi-species data sets. Results show that SAMs improve model accuracy and discriminatory capacity compared to separate SDMs. This is achieved by borrowing strength from common species having higher information content. Moreover, the improvement increases as the species become rarer.

*Key words: community level modeling; cross validation; generalized linear models; mixture models; species archetypes; species distribution models.*

## Introduction

Species distribution models (SDMs), which relate the observed occurrence of species to their environment, are an important tool in conservation planning and biodiversity management (Elith and Leathwick 2009). Over the past two decades, there has been substantial development in the statistical methodology behind single-species SDMs, from generalized linear models (GLMs; McCullagh and Nelder 1989) and generalized additive models (GAMs; Hastie and Tibshirani 1990) to recent developments in Bayesian modeling techniques (Chakraborty et al. 2010). For multi-species data sets, these methods require fitting a large number of models, which is inefficient and makes interpretation challenging. This has motivated advances in community-level approaches, such as multivariate regression trees (MVPART; De'ath 2002) and community-level multivariate adaptive regression splines (MARS; Leathwick et al. 2006).

Recently, Dunstan et al. (2011) proposed using finite mixture of regression models (McLachlan and Peel 2000) to analyze community data, which we call species archetypes models (SAMs). The key idea is to assume that all species can be classified into a small number of responses to the environment, referred to as *archetypal responses*. By clustering based on environmental response, we can borrow strength across species. Therefore, compared to fitting a large number of separate SDMs, SAMs allow responses for rarer species to be estimated with greater precision by grouping them with prevalent species having statistically similar responses. Another advantage SAMs have over separate species models is that management of a large assemblage is simplified to the handling of a (much) smaller number of archetypal units.

SAMs differ from other clustering approaches such as hierarchical agglomerative clustering using distance measures (Legendre and Legendre 1998) in that it is a model-based approach. Such an approach allows for formal inference about which environmental covariates are associated with species distributions. Other benefits include: data-driven methods for determining the number of archetypes, a probabilistic (soft) classification of each species into different archetypes, and predicted distributions for each species individually. This last point is particularly important: it makes possible a comparison of the predictive performance between SAMs and separate SDMs.

In this work, we compare the predictive capacity between methods that separately estimate the environmental response of species (separate SDMs) and SAMs, which cluster these responses across species, over a number of real data sets. The methodology we use for SAMs is given in Dunstan et al. (2013), but this paper is distinct in that it is the first to specifically compare SAMs to separate SDMs. In doing so, we evaluate whether the *act of mixing* (act of clustering) species

Manuscript received 31 July 2012; revised 1 January 2013; accepted 17 April 2013. Corresponding Editor: E. E. Holmes.
[5] E-mail: fhui28@gmail.com

TABLE 1. Data sets used in this study.

| Name | Organism | Reference | Type | $N$ | $S$ | Area | DNN | Covariates |
|---|---|---|---|---|---|---|---|---|
| Butterfly | butterflies | Oliver et al. (2006) | Ab | 66 | 33 | 157 | 0.309 | habitat, percentage building in surrounding area, percentage of urban vegetation in surrounding area† |
| North West Shelf (NWS) | fish | Young and Sainsbury (1985) | Ab | 464 | 319 | 48 968 | 4.290 | depth, distance along the coast, means and intra-annual standard deviations of salinity, nitrate, oxygen, silicone, and sea surface temperature |
| Great Barrier Reef (GBR) | invertebrates | Pitcher et al. (2007) | PA | 1189 | 1562 | 447 944 | 7.711 | depth, bottom stress, percent gravel, percentage mud, percentage carbonate, means and intra-annual standard deviations of temperature, oxygen, salinity and K490‡ |
| Blue Mountains (BM) | Myrtaceae trees/shrubs | NSW Office of Environment of Heritage (2010) | PA | 3682 | 90 | 78 184 | 0.858 | fire history, annual rainfall, mean minimum temperature, mean maximum temperature§ |

*Notes:* Features listed include organisms of interest; data type (Abundance [Ab] or presence–absence [PA]); number of sites $N$; number of species $S$; the area surveyed (km$^2$); mean distance to nearest site (DNN; km); and covariates used in the analysis.

† Habitat includes hayfield, mixed-grass prairie, short-grass prairie, and tallgrass prairie.

‡ K490 is a measure of turbidity.

§ Fire history is a count of previous wildfires and prescribed burns.

based on their environmental responses can improve predictive performance.

## DATA SETS AND METHODS

### *Description of data sets*

Four data sets were used for comparison, as summarized in Table 1. Both presence–absence and abundance data sets are used to assess how clustering performs for two commonly recorded response types. The four data sets vary in the number of sites $N$ and species $S$. The environmental predictors used for each data set, also listed in Table 1, were selected based on a priori biological assumptions and a simplistic screening method. The latter involves fitting separate GAMs to each species and choosing covariates that were significant ($P < 0.1$) for $\geq 25\%$ of the models. Continuous covariates were modeled using quadratic terms (Austin 2002). Finally, for the North West Shelf (NWS) data set around 2% of the records for abundance were missing, and we chose to impute the response on these based on biomass data available for the corresponding sites and species (see Appendix A for details).

### *Separate species SDMs*

Separate SDMs were modeled as an independent GLM for each species. For $j = 1, \ldots, S$, the likelihood function is given as

$$\prod_{i=1}^{N} f(y_{ij}; \mu_{ij}, \phi_j); \quad g(\mu_{ij}) = \boldsymbol{\beta}_{0j} + \mathbf{x}_i' \boldsymbol{\beta}_j \tag{1}$$

where the response of species $j$ at site $i$ has distribution $f(\cdot)$ with mean $\mu_{ij}$. Eq. 1 is the usual formulation for a GLM, where we assume a distribution on the responses values and relate the mean of this distribution to a vector $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ of environmental covariates ($p$) using a link function $g(\cdot)$ (Zuur et al. 2007). For presence–absence data, we assumed a Bernoulli distribution with logit link function. For abundance data, we used a negative binomial distribution with log link function to account for the mean–variance overdispersion, which often displays a quadratic relationship $Var(y_{ij}) = \mu_{ij} + \phi_j \mu_{ij}^2$ where $\phi_j$ is the dispersion parameter (see, for instance, Warton 2005). We constructed mean–variance plots for the Butterfly and NWS data sets (not shown), and verified the presence and quadratic form of the overdispersion.

An important aspect of our comparison between SAMs and separate SDMs is that, for each data set, the same mean structure is used for fitting each separate GLM and for each archetype of the SAM. This is necessary to ensure that any difference in predictive performance observed could be attributed solely to mixing (see Baselga and Araújo [2009] for a similar approach).

### *Species archetype models*

The SAM is formulated as follows: for $j = 1, \ldots, S$, the likelihood function is given as

$$\sum_{g=1}^{G} \pi_g \prod_{i=1}^{N} f(y_{ij}; \mu_{ijg}, \phi_j) \qquad g(\mu_{ijg}) = \boldsymbol{\beta}_{0j} + \mathbf{x}_{i'} \boldsymbol{\beta}_g \tag{2}$$

where $G \ll S$ is the number of archetypes, $f(\cdot)$ is the distribution for the $g$th archetype (Bernoulli or negative binomial for presence–absence and abundance data respectively), and $\pi_g$, with $\sum_{g=1}^{G} \pi_g = 1$ are the overall proportion of species whose mean response is governed by archetype $g$.

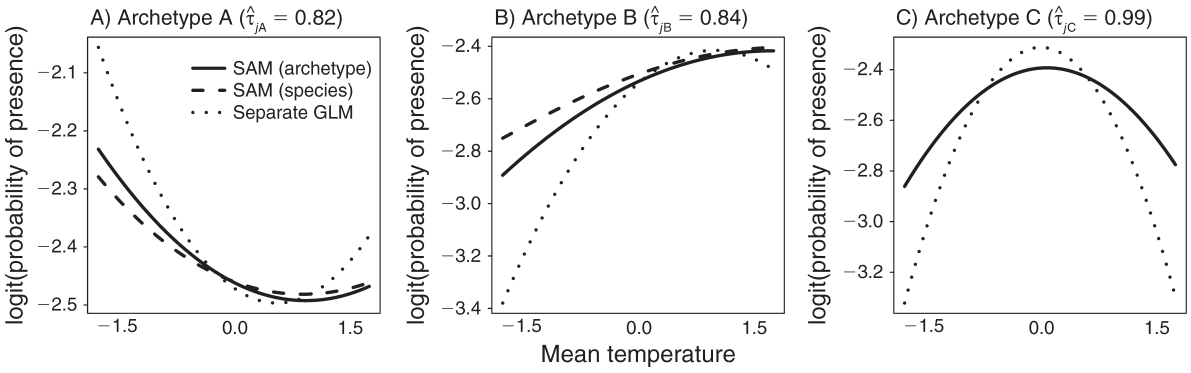


FIG. 1. Plots depicting the archetypal responses ($\hat{\mu}_{ijg}$; solid line) to standardized intra-annual mean temperature for three archetypes ($g$, denoted as archetypes A, B, and C) for species $j$ at site $i$ in the GBR data set. All other covariates are held fixed at their respective means. Also plotted are the predicted responses from SAMs (dashed line) of three species (*Musculus cumingianus*, *Lissocarcinus polybioides*, and *Choerodon monostigma*) that characterize each of these archetypes, in the sense that each species has a large posterior probability $\hat{\tau}_{jg}$ of belonging to their respective archetype. Predicted responses from SAMs are based on a weighted average across all archetypes, and hence they are not exactly equal to the archetypal responses (solid line), although they are share similar shapes due to the large posterior probability $\hat{\tau}_{jg}$. Note also that the predicted species responses from SAMs are considerably flatter than from fitting a separate SDMs to each species alone (dotted line).

Although both Eqs. 1 and 2 fit a mean model based on the $\mathbf{x}_i$'s, the form of the environmental response from the two methods are not equivalent. For separate SDMs, the environmental response is estimated independently of other species (hence the subscript $j$ on $\beta_j$). For SAMs, the environmental response is based on the estimated coefficients from $G \ll S$ archetypes (hence the subscript $g$ on $\beta_g$). It is this estimation of clustered responses in SAMs that captures the idea of borrowing strength across species. Furthermore, unlike other methods such as agglomerative clustering (Legendre and Legendre 1998), for SAMs the species are classified in a probabilistic (soft) manner. Finally, note the intercepts and dispersion parameters in Eq. 2 are species specific, since we only want to be clustering on the *form* of the environmental response (see Dunstan et al. [2013] for a detailed explanation).

Details regarding the estimation procedure for a SAM (with fixed $G$) may be found in Dunstan et al. (2013). To choose the number of archetypes, we fitted SAMs for a range of $G$ and chose the one that minimized the Bayesian information criterion (BIC; Schwarz 1978):

$$\text{BIC} = -2 \times \text{log-likelihood function} + \ln S \times (\text{number of parameters})$$

with the number of parameters set to $(G-1)+S+G\times p$. If the $\phi_j$'s need to be estimated as well (for abundance data), then $S$ is replaced with $2S$. Based on extensive simulations conducted in forthcoming work, we found that BIC worked quite well for prediction purposes.

Having selected an optimal $G$, predicted values can be obtained from the fitted SAM. Note that having $\beta_{0j}$ means predictions are obtained *individually* for each species. We used a weighted average of the predictions from each archetype as the final predicted values. Specifically, the predictions for species $j = 1, \ldots, S$, are:

$$\hat{\mu}_{ij} = \sum_{g=1}^{\hat{G}} \hat{\tau}_{jg}\hat{\mu}_{ijg}$$

where $\hat{\tau}_{jg}$ is the posterior probability of species $j$ belonging to archetype $g$. The $\tau_{jg}$'s differ from the $\pi_g$'s defined in Eq. 2 in two ways: (1) $\tau_{jg}$ is a species-specific probability, whereas $\pi_g$ are the overall proportions of species, belonging to archetype $g$, (2) $\tau_{jg}$ are functions of parameters, whereas $\pi_g$ are parameters themselves. Furthermore, the vector $\boldsymbol{\tau}_j = (\tau_{j1}, \tau_{j2}, \ldots, \tau_{jG})$ are precisely the probabilities characterizing the soft classification feature of SAMs. This is illustrated in Fig. 1 where the predicted species responses ($\hat{\mu}_{ij}$; dashed lines) and the archetypal responses ($\hat{\mu}_{ijg}$; solid lines) are slightly different from each other. Sometimes, as in Fig. 1C, the two sets of responses are indistinguishable due to $\hat{\tau}_{jg}$ being almost 1.

### *Model comparison and evaluation*

For each of the four data sets, we consider two cases: the first cases implements a "number of parameters plus 1" rule, where species with presences less than or equal to the number of parameters entering the model are discarded prior to analysis. This rule is based on the concept of degrees of freedom (see Zuur et al. [2007] for an explanation), although with many covariates it tends to remove a large fraction of species. The second case implements a "one-third" rule, where species with presences less than one-third the number of parameters entering the model are discarded. Both rules remove extremely rare species, e.g., singletons and species with zero recorded presences, which is recommended since these have insufficient information to be modeled well regardless of the method used (see, for example, Leathwick et al. [2006]). The one-third rule includes many more rare species for analysis compared to the

TABLE 2. Mean differences in predictive log-likelihood per testing site (log $L_p$) and area under the curve (AUC), along with 95% confidence intervals in parentheses, when applying (A) "number of parameters + 1" rule and (B) "one-third" rule.

| Rule and data set | $S_{remain}$ | Summary of selected $G$ | Difference in log $L_p$ | Difference in AUC |
|---|---|---|---|---|
| A) Number of parameters + 1 | | | | |
| Butterfly | 14 | 2 (2, 2.25) | 0.005 (−0.033, 0.043) | |
| NWS | 122 | 6 (6, 7) | 0.898 (0.862, 0.933) | |
| GBR | 137 | 11 (11, 12) | 0.018 (0.015, 0.021) | 0.011 (0.007, 0.015) |
| BM | 50 | 11 (10, 12) | 0.010 (0.005, 0.014) | 0.001 (−0.006, 0.007) |
| B) One-third | | | | |
| Butterfly | 22 | 2 (2, 2) | 0.745 (0.498, 0.993) | |
| NWS | 223 | 11 (10, 11) | 1.074 (1.048, 1.100) | |
| GBR | 675 | 17 (16, 18) | 0.089 (0.084, 0.094) | 0.052 (0.049, 0.055) |
| BM | 76 | 13 (11, 13) | 0.046 (0.042, 0.049) | 0.048 (0.039, 0.057) |

*Notes:* Listed also is the number of species remaining for comparison following application of either rule ($S_{remain}$) and the modal number of archetypes $G$ selected for SAMs across the 40 cross-validation sets (first and third quartiles in parentheses). Difference is defined as log $L_p$ of SAMs minus log $L_p$ of separate GLMs, and similarly for AUC. Key to data set abbreviations: NWS, North West Shelf; GBR, Great Barrier Reef; BM, Blue Mountains.

number of parameters plus 1 rule, motivating a greater need for methods that can borrow strength across species, i.e., SAMs.

We assessed predictive performance by measuring how well models predict to independent "test" data via cross validation (Hastie et al. 2009). For the Butterfly, Great Barrier Reef (GBR), and NWS data sets, we used random cross-validation, whereby 10% of the sites were randomly sampled out to act as test data, while the model was fitted to the remaining 90% training sites. Given the sparse sampling intensity (Table 1; mean nearest site) for GBR and NWS, and the lack of residual correlation in relevant diagnostic plots for Butterfly (plot not shown), then spatial autocorrelation for these three data sets was negligible. For the Blue Mountains (BM) data set, however, the high density of sites (Table 1) motivated us to use block cross validation to (approximately) ensure independence between training and test data sets. To perform block cross validation, we divided the area surveyed into 50 × 50 km blocks (38 blocks in total). Each block was then regarded as a sampling unit, and we randomly sampled 10 blocks for testing with the model fitted to the remaining 28 training blocks. For each data set, we considered 40 different splits of training/test data for the cross-validation procedure, allowing the construction of confidence intervals and hypothesis testing for our measures of predictive performance.

We evaluated performance using predictive log-likelihood per testing site (abbreviated to log $L_p$) as a measure of model calibration and predictive accuracy; a higher log $L_p$ meant a better fit of the testing data using the proposed model. For binary data, area under the ROC curve (AUC; Fielding and Bell 1997) was also used to evaluate a model's ability to discriminate between presences and absences. To assess if SAMs outperformed separate GLMs for rare species in particular (borrowing strength), we plotted species differences in log $L_p$ (or AUC), averaged over the 40 cross-validation splits, against species prevalence in the full data set. A difference greater than 0 in log $L_p$ (or in AUC) meant greater prediction accuracy (discriminatory capacity) for SAMs.

## RESULTS

### *Comparable performance for prevalent species*

We first considered when each species had sufficient prevalence to effectively parameterize the environmental responses relationship separately (number of parameters plus 1 rule). For the Butterfly data set, this led to 14 species, allowing us to assess predictive performance for a small assemblage. The mean difference in log $L_p$ was close to zero (Table 2A), although there was some evidence that SAMs were predicting better for rarer species (see Appendix B). In the three larger data sets, we observed slight evidence in favor of SAMs improving predictions (Table 2A). For both presence–absence data sets GBR and BM, SAMs offered small but statistically significant improvements over separate GLMs in log $L_p$ and AUC. Plots of individual species differences in log $L_p$ (and AUC) vs. prevalence presented some evidence that rarer species were been better predicted by SAMs compared to separate GLMs (see Appendix B).

The biggest difference between the two methods occurred for the NWS data set, where log $L_p$ was substantially higher for SAMs (Table 2A). This result, however, was largely due to the chosen mean structure overfitting many of the separate GLMs. A further analysis comparing SAMs to separate intercept-only GLMs showed a much smaller but nevertheless statistically significant difference in log $L_p$ in favor of SAMs: 0.022 (0.015, 0.030) (mean and 95% CL). A plot of species differences in log $L_p$ against prevalence revealed SAMs outperformed separate intercept-only GLMs by successfully modeling the responses present in the most abundant species, while performing comparably for rarer species (see Appendix B).

### *Borrowing strength for the rarer species*

When more rare species were included in the model (the one-third rule), strong evidence of SAM's superior

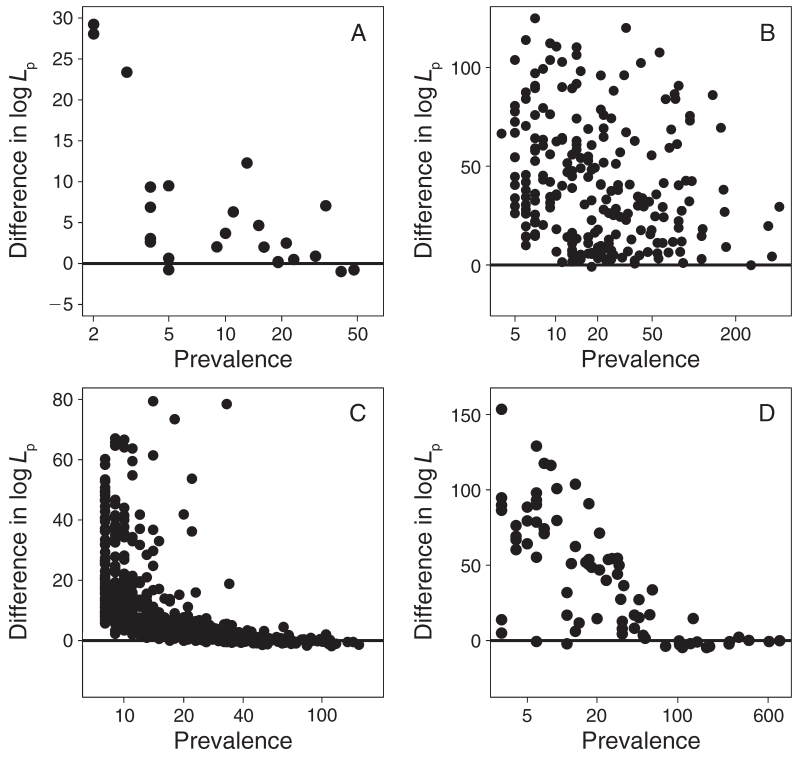


FIG. 2. Individual species differences in predictive log-likelihood per testing site ($\log L_p$) (averaged across the 40 cross-validation splits) vs. prevalence, applying the "one-third" rule to (A) Butterfly, (B) North West Shelf (NWS), (C) Great Barrier Reef (GBR), and (D) Blue Mountains (BM) data sets. Difference here is defined as $\log L_p$ of SAMs minus $\log L_p$ of separate species GLMs. For all four data sets, there are very few points less than zero (below the solid horizontal line), indicating that SAMs consistently outperforms separate GLMs in terms of predictive accuracy. Also, we see that the rarer a species is, the better SAMs perform relative to separate GLMs. Similar trends were observed for species differences in AUC plotted against prevalence (see Appendix B).

performance emerged in all four data sets (Table 2B). Across the four data sets, plots of species differences in $\log L_p$ against prevalence showed strong evidence in the ability of SAMs to borrow strength across species: the rarer a species was, the greater the improvement in predictive performance by mixing over the coefficients compared to estimating them independently (Fig. 2). Similar trends were observed for species differences in AUC (see Appendix B). Further evidence of the SAMs borrowing strength across species was given by the number of archetypes chosen in Table 2B, which were not much more compared to when only the prevalent species were included in Table 2A. This showed that rare species had insufficient information to justify an entirely new response group, and it was more beneficial to class them with a prevalent species having similar response.

For the Butterfly data set, performance of separate GLMs worsened due to overfitting of the 10 rare species newly included. Overfitting with the selected mean model was also an issue in the NWS data set. For both data sets, additional comparisons between SAMs and separate intercept-only GLMs found that while differences in $\log L_p$ dropped considerably, they remained significant in favor of SAMs for both the Butterfly (0.041 (0.013, 0.070)) and NWS (0.129 (0.119, 0.140)) data sets.

Greater insight into the improvement in predictive performance made by SAMs can be gained by visualizing how the predicted mean response differed between SAMs and separate GLMs. This is illustrated in Fig. 1 for three selected archetypes (denoted for simplicity as archetypes A, B, and C) in the GBR data set. Each plot shows how the species predicted response from SAMs (dashed line) and separate GLMs (dotted line) varied as a function of standardized mean temperature, while holding the other covariates fixed at their respective means. Each plot was constructed by choosing, for each archetype, a species having large posterior probability $\hat{\tau}_{jg}$ of belonging to that archetype. Compared to the predicted responses from separate SDMs (Fig. 1, dotted line), the archetypal responses (solid line) and the SAM predicted species' response (dashed line) are much flatter in curvature. This is because the SAM predictions, both archetypal and species, are based on "average" responses across a group of species with similar behavior. These points led to the significant improvements in predictive performance.

## Discussion

In this paper, we compared the performance of separate SDMs and a mixture-model approach on a number of real data sets. Results showed SAMs outperformed GLMs both in prediction accuracy and discriminatory capacity by borrowing strength across species. Furthermore the rarer a species was, the greater the improvement made by mixing the coefficients. These findings complement those found in Leathwick et al. (2006) and Ovaskainen and Soininen (2011), both of which found strong evidence that community models outperformed univariate counterparts when many recorded species were sparse. Comparing across the four data sets in Table 2A and B, we observed an interesting trend: the larger the ratio $S/N$ was, the better SAMs perform relative to single species GLMs. This is an important finding, since, in many ecological surveys, the number of species is often a significant fraction of the number of sites (sometimes even greater than 1). These results thus present a strong argument for adopting a model-based clustering approach for such data sets.

One drawback of our comparison was the use of GLMs as the basic modeling tool, as they are limited in their ability to model the often complex relationships in ecology (Leathwick et al. 2006). To see if relaxing assumptions on the mean model made any difference, we fitted separate GAMs to the same 40 cross-validation splits used for the Butterfly and BM data sets, and compared the predictions made against SAMs (mixture of GLMs). We found that while separate GAMs predicted better than SAMs for abundant species, SAMs continued to strongly outperform GAMs for rarer species (see Appendix C). From this, we conclude that mixture modeling is a powerful technique that could, in principle, be applied to any modeling tool. A topic of future research will be to extend SAMs to encompass mixture of GAMs, to further improve predictive performance.

SAMs can be considered as a type of hierarchical model, meaning they share similarities with Generalized Linear Mixed Models (GLMMs). Both SAMs and GLMMs borrow strength in order to better model rare species. A key difference however is that the latent variable (random effect) in SAMs is categorical, while in GLMMs it is assumed to be normally distributed (e.g., Ives and Helmus 2011). Importantly, it is the categorical latent variable in SAMs that facilitates the clustering of species into archetypes. SAMs also differ from the hierarchical model proposed by Ovaskainen and Soininen (2011) in two ways: (1) SAMs are fitted via maximum likelihood, which is computationally less intensive compared to the Bayesian approach of Ovaskainen and Soininen (2011) that requires elicitation of prior distributions, (2) we develop SAMs in a broader framework to handle abundance data and perform model selection (see Dunstan et al. 2013).

Species archetype modeling is an "assemble and predict together" strategy (Ferrier and Guisan 2006), and such an approach has a number of strengths reflected in this paper. Better parameterization allows greater power at uncovering environmental responses relationships in rare species. Consequently, SAMs may provide more reliable extrapolations to new locations in geographic and environmental space. The very idea of a "species archetype" embodies the notion of borrowing strength across species, allowing biodiversity management to be simplified into strategies for a smaller number of archetypal units. At the same time, individual species distributions are easily obtained. This capacity to easily switch focus from the community level to species level and vice versa is an important advantage of SAMs.

## Acknowledgments

F. K. C. Hui was supported by a University of New South Wales Research Excellence Award and CSIRO Ph.D. Scholarship. D. I. Warton is supported by Australian Research Council Discovery Projects and Future Fellow funding schemes (project number DP130102131 and FT120100501). P. K. Dunstan and S. D. Foster are supported through the National Environment Research Program (NERP), an Australian Government initiative, and in particular by the NERP Marine Biodiversity Hub. Thanks to David Peel and two anonymous reviewers for useful comments.

## Literature Cited

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling 157:101–118.

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

Chakraborty, A., A. E. Gelfand, A. M. Wilson, A. M. Latimer, and J. A. Silander, Jr., 2010. Modeling large scale species abundance with latent spatial processes. Annals of Applied Statistics 4:1403–1429.

De'ath, G. 2002. Multivariate regression trees: a new technique for modeling species–environment relationships. Ecology 83: 1105–1117.

Dunstan, P. K., S. D. Foster, and R. Darnell. 2011. Model based grouping of species across environmental gradients. Ecological Modelling 222:955–963.

Dunstan, P. K., S. D. Foster, D. I. Warton, and F. K. C. Hui. 2013. Finite mixture of regression modelling for high-dimensional count and biomass data in ecology. Journal of Agricultural, Biological and Environmental Sciences. In press.

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

Ferrier, S., and A. Guisan. 2006. Spatial modelling of biodiversity at the community level. Journal of Applied Ecology 43:393–404.

Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24:38–49.

Hastie, T., and R. J. Tibshirani. 1990. Generalized additive models. Chapman and Hall, London, UK.

Hastie, T., R. J. Tibshirani, and J. H. Friedman. 2009. The elements of statistical learning: data mining, inference, and prediction. Second edition. Springer-Verlag, New York, New York, USA.

Ives, A. R., and M. R. Helmus. 2011. Generalized linear mixed models for phylogenetic analyses of community structure. Ecological Monographs 81:511–525.

Leathwick, J. R., J. Elith, and T. Hastie. 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecological Modelling 199:188–196.

Legendre, P., and L. Legendre. 1998. Numerical ecology. Elsevier Science, Amsterdam, The Netherlands.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Chapman and Hall, London, UK.

McLachlan, G. J., and D. Peel. 2000. Finite mixture models. Wiley, New York, New York, USA.

NSW Office of Environment of Heritage. 2010. Bionet atlas of NSW wildlife. http://www.bionet.nsw.gov.au/

Oliver, J. C., K. L. Prudic, and S. K. Collinge. 2006. Boulder County open space butterfly diversity and abundance. Ecology 87:1066.

Ovaskainen, O., and J. Soininen. 2011. Making more out of sparse data: hierarchical modeling of species communities. Ecology 92:289–295.

Pitcher, C., et al. 2007. Seabed biodiversity on the continental shelf of the Great Barrier Reef World Heritage Area. Technical report. CSIRO Marine and Atmospheric Research, Brisbane, Australia.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6:461–464.

Warton, D. I. 2005. Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. Envirometrics 16:275–289.

Young, P. C., and K. J. Sainsbury. 1985. CSIRO's north west shelf program. Australian Fisheries 44:16–20.

Zuur, A. F., E. N. Ieno, and G. M. Smith. 2007. Analysing ecological data. Springer Science + Business Media, New York, New York, USA.

## Supplemental Material

**Appendix A**

Imputation of the response for the NWS data set (*Ecological Archives* E094-174-A1).

**Appendix B**

Additional results for comparison between species archetype models (SAMs) and separate species distribution models (SDMs) (*Ecological Archives* E094-174-A2).

**Appendix C**

Comparing the performance of SAMs and separate species generalized additive models (GAMs) (*Ecological Archives* E094-174-A3).