Incorporating biotic interactions in the distribution models of African wild silk moths (*Gonometa* species, Lasiocampidae) using different representations of modelled host tree distributions

Short running title:

Incorporating biotic interactions into SDMs

Morgan Jade Raath[1,2], Peter Christiaan le Roux[1], Ruan Veldtman[2] & Michelle Greve[1]

[1] Department of Plant and Soil Sciences, University of Pretoria, Private Bag X20, Pretoria 0002, South Africa

[2] South African National Biodiversity Institute, Kirstenbosch Research Centre, Private Bag X7, Claremont, 7735, South Africa

Corresponding author: Morgan J. Raath

Postal address: Department of Plant and Soil Sciences, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa

Email address: morganj.r@hotmail.com

Telephone number: +27 83 468 0043

Abstract Biotic interactions influence species niches and may thus shape distributions. Nevertheless, species distribution modelling has traditionally relied exclusively on environmental factors to predict species distributions, while biotic interactions have only seldom been incorporated into models. This study tested the ability of incorporating biotic interactions, in the form of host plant distributions, to increase model performance for two host-dependent lepidopterans of economic interest, namely the African silk moth species, *Gonometa postica* and *Gonometa rufobrunnea* (Lasiocampidae). Both species are dependent on a small number of host tree species for the completion of their life cycle. We thus expected the host plant distribution to be an important predictor of *Gonometa* distributions. Model performance of a species distribution model trained only on abiotic predictors was compared to four species distribution models that additionally incorporated biotic interactions in the form of four different representations of host plant distributions as predictors. We found that incorporating the moth-host plant interactions improved *G. rufobrunnea* model performance for all representations of host plant distribution, while for *G. postica* model performance only improved for one representation of host plant distribution. The best-performing representation of host plant distribution differed for the two *Gonometa* species. While these results suggest that incorporating biotic interactions into species distribution models can improve model performance, there is inconsistency in which representation of the host tree distribution best improves predictions. Therefore, the ability of biotic interactions to improve species distribution models may be context-specific, even for species which have obligatory interactions with other organisms.

Key words: species interactions, niche modelling, Maxent.

INTRODUCTION

Biotic interactions contribute to determining the realised niche of species (Araújo & Guisan 2006; Godsoe & Harmon 2012), and may strongly influence species distributions, either limiting or expanding the conditions, and total area in which a species can occur (Leathwick & Austin 2001; Araújo & Luoto 2007; Schweiger *et al*. 2008; Meier *et al*. 2010; le Roux *et al*. 2012). While climate has been proposed to be the main factor driving species distributions at broad (e.g. continental) scales, with biotic interactions expected to be more important at local scales (Pearson & Dawson 2003), several studies have documented the importance of biotic interactions in shaping species distributions at broad scales (Araújo & Luoto 2007; Heikkinen *et al*. 2007; Gilman *et al*. 2010; Meier *et al*. 2010; Schweiger *et al*. 2012; Wisz *et al*. 2013). However, SDMs, a widely utilised tool for predicting species distributions, often model species distributions based only on abiotic factors (usually climatic variables), overlooking the role of biotic interactions in shaping species distributions, particularly at broad spatial scales (Guisan *et al*. 2006; Wisz *et al*. 2013). Species distribution models that are trained only on environmental variables model the fundamental niche of a species, i.e. the niche where the climate and edaphic conditions are suitable for an organism to occur (Guisan & Zimmermann 2000; Pulliam 2000; Wiens *et al*. 2009). However, the actual distribution of a species (i.e. as represented by the realised niche) is a complex expression of several overlapping factors (Pearson & Dawson 2003; Soberón & Peterson 2005; Soberón 2007; Kearney 2006; Jiménez-Valverde *et al*. 2008), which in addition to abiotic variables, also include biotic interactions, dispersal ability and the evolutionary capacity of populations of species to adapt to novel conditions (see BAM diagram [B = biotic, A = abiotic, M = accessibility] of Soberón & Peterson 2005 and Soberón 2007). Including the effects of biotic interactions in SDMs may thus significantly improve the

predictive abilities of SDMs. Indeed, studies that have examined the influence of biotic interactions on species distributions and richness patterns at broad scales have generally shown that explicitly considering inter-specific interactions in SDMs increases model accuracy, indicating the importance of biotic interactions influencing species distributions at broad scales (e.g. Araújo & Luoto 2007; Heikkinen *et al*. 2007; Gilman *et al*. 2010; Meier *et al*. 2010; Bateman *et al*. 2012; Boulangeat *et al*. 2012; Schweiger *et al*. 2012; Wisz *et al*. 2013).

Several approaches have been used to model the influence of biotic interactions on species distributions, mainly at local scales (Kissling *et al*. 2012; Silva *et al*. 2014; Anderson 2017). The most widely used approach involves adding the occurrence of purportedly interacting species as a proxy for biotic interactions (or by adding the cover or abundance of a species as a proxy for the intensity or frequency of an interaction) (Leathwick & Austin 2001; Kissling *et al*. 2012). A related approach is to use abiotic variables to model the distributions of interacting species individually, and then to restrict the distribution of the one species based on the modelled distribution of the other species (Schweiger *et al*. 2008). Biotic interactions have also been incorporated into SDMs by combining classic SDM approaches with process-based models (for example, modelling physiological representations of resource competition; Lischke *et al*. 2006; Hickler *et al*. 2012). With all of these approaches, the effects of biotic interactions on species distributions have frequently been limited to investigating just the effect of a single species or the influence of one type of interaction (Van der Putten *et al*. 2010). More novel approaches involve incorporating biotic interactions into SDMs within multispecies systems (Kissling *et al*. 2012). However, this kind of modelling often requires detailed knowledge of all interactions, which is unavailable for most species.

The objective of this study is therefore to assess the ability of four fairly simple methods of accounting for biotic interactions on SDM performance. We consider these methods as they are simple to implement and can be used for species for which only fairly basic life history information is available. We make use of two economically-important African wild silk moth species (*Gonometa* species, Lasiocampidae), that are dependent on specific tree hosts for food, as a model system. More specifically, we compare the performance of models run using only climatic (i.e. abiotic) variables as predictors with the performance of models run with both climate predictors and different representations of the occurrence of host tree species as a proxy for moth-host plant interactions (i.e. abiotic and biotic predictors; see Silva *et al*. 2014 for comparison).

METHODS

Study species

The moths *Gonometa postica* Walker 1855 and *Gonometa rufobrunnea* Aurivillius 1922 occur in southern and East Africa and are of economic value due to the high-quality silk derived from their cocoons (Akai *et al*. 1997; Veldtman *et al*. 2002; Veldtman *et al*. 2004; Delport 2006). However, the extent and drivers of the distribution of these moths are poorly understood (Chikwenhere 1992; Hartland-Rowe 1992; Veldtman *et al*. 2007). Indeed, these species exhibit large inter-annual population fluctuations, and outbreaks of these moths are patchily distributed (Veldtman *et al*. 2007), varying across space (both locally and regionally) and time.

Gonometa larvae feed only on a small number of host tree species, and usually also pupate on these species. The main *G. postica* host plant in southern Africa is *Vachellia erioloba* (camel thorn, Fabaceae), while other host plants include *Vachellia tortilis* (umbrella thorn acacia,

Fabaceae), *Senegalia mellifera* (blackthorn, Fabaceae) and *Burkea africana* (wild seringa, Caesalpiniaceae).

Gonometa rufobrunnea forages predominantly on *Colophospermum mopane* (Fabaceae) which occupies dry savannas in southern Africa (Huntley & Walker 1982). Other plants such as *Parinari curatellifolia* (mobola-plum, Chrysoalanaceae) and *Brachystegia* species have also been reported as hosts of *G. rufobrunnea* (Kroon 1999). However, only *Brachystegia spiciformis* (zebrawood, Fabaceae) was considered in addition to *C. mopane* as a host plant for *G. rufobrunnea* in this study because it is an important host plant for *G. rufobrunnea* in East Africa (Kroon 1999) where *C. mopane* does not occur.

Locality records

The majority of the locality records of *G. postica* (n = 302) and *G. rufobrunnea* (n = 89) were obtained from museums (Afromoths; Denver Museum of Nature and Science; Ditsong Museum; Mississippi Entomological Museum; Natural History Museum of Zimbabwe; Natural History Museum of London and Oxford University Museum), private collections (of W. Mey; A. Prozorov; H. Staude and M. Ströhle) and the literature. Additional records were obtained from personal observations made by amateur lepidopterists (J. Boomker; A. Curle; J. Groenewald and M. Maclean). The locality records for the host tree species were obtained from databases (Greve *et al.* 2012; PRECIS; Flora of Namibia; Flora of Zimbabwe) and herbaria (H.G.W.J. Schweickerdt Herbarium and Buffelskloof Nature Reserve Herbarium). Records of both moth and host tree species for which locality descriptions, but not co-ordinates, were supplied were georeferenced using the gazetteer ACME Mapper 2.1.

Species distribution models are frequently built using occurrence records that are spatially biased towards areas that are more accessible and easy to survey (Phillips *et al.* 2009). This leads

to spatial clumping of occurrence records and therefore over-representation of some environmental features which can result in poor model predictions (Kramer-Schadt *et al*. 2013). Therefore, the 'thin' function in the SpThin package (see Aiello-Lammens *et al.* 2015) was used to spatially thin the both moth and host tree occurrence records prior to modelling (Table S1 and Fig. S1-S3 in Appendix S1).

Environmental predictor variables

Seven climate variables were downloaded from the Worldclim database (v.1, Hijmans *et al*. 2005) at 2.5' spatial resolution: annual mean temperature (BIO1), maximum temperature of the warmest month (BIO5), mean temperature of the driest quarter (BIO9), annual precipitation (BIO12), precipitation seasonality (BIO15), precipitation of the driest quarter (BIO17) and precipitation of the coldest quarter (BIO19). BIO9 and BIO19 were subsequently removed because of their high correlation ($r > 0.8$) with BIO1 and BIO17 respectively. Five soil variables, namely soil organic carbon content, pH, cation exchange capacity, percentage sand content and bulk density were obtained from the SoilGrids database, downloaded at 1 km spatial resolution (Hengl *et al*. 2014) and resampled to 2.5' spatial resolution. These soil variables did not display high collinearity. All map conversions were conducted in ArcMap v. 10.1.

 Modelling and model evaluation

Species distributions were modelled for Africa south of the equator using a maximum entropy approach in Maxent (Phillips *et al*. 2006). Maxent is amongst one of the best performing SDM methods for presence-only data (Elith *et al*. 2006; Hernandez *et al*. 2006; Phillips *et al*. 2006; Elith *et al*. 2011) because it is not sensitive to sample size and can use both categorical and continuous predictors. Maxent models produce a probability of occurrence distribution over the

defined study area (Phillips & Dudík 2008) and the target distribution is estimated by finding the distribution that is closest to uniform (i.e. the distribution of maximum entropy) under a set of environmental constraints (or features) derived from species' occurrence records (Phillips *et al*. 2006).

Initially, the current potential distributions of *G. postica* and *G. rufobrunnea* were modelled using only the climatic variables (excluding the soil variables) mentioned above (hereafter referred to as "Abiotic models"; Fig. 1). These variables were selected based on the seasonal nature of *Gonometa* species' outbreaks and their sensitivity to rainfall and drought periods (Hartland-Rowe 1992; Delport 2006). In southern Africa, *G. postica* and *G. rufobrunnea* have two generations per year coinciding with rainfall seasonality (Hartland-Rowe 1992). It is therefore expected that temperature and precipitation will have an effect on *Gonometa* distribution, which is in line with other SDM studies on Lepidoptera (Beaumont & Hughes 2002; Tóth *et al*. 2013).

It is expected that host tree distributions (i.e. biotic predictors) will play a role in shaping *Gonometa* distributions because *Gonometa* species are largely dependent on a limited number of host plant species for the completion of their life cycles (Hartland-Rowe 1992; Fening *et al*. 2008; Fening *et al*. 2010). Therefore, a second set of models was run that incorporated moth-host plant interactions, in the form of host plant distributions, when predicting *Gonometa* species distributions. The distributions of the host plants *V. erioloba*, *S. mellifera*, *V. tortilis* and *B. africana* (for *G. postica*) and *C. mopane* and *B. spiciformis* (for *G. rufobrunnea*) were modelled in Maxent using the five bioclimatic and the five soil variables as predictors, all of which have been shown to be important drivers of savanna tree distributions (Huntley & Walker 1982; Smith & Goodman 1986).
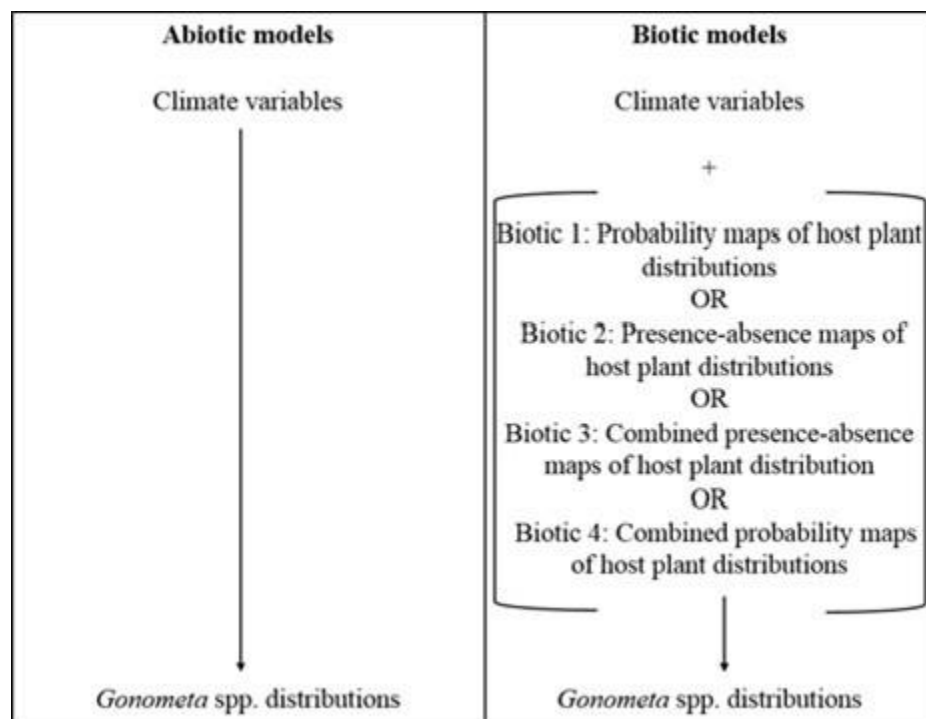
Fig. 1. Diagram illustrating the different types of models run to predict Gonometa species distributions.

To model *Gonometa* distributions that include biotic interactions, four different models (hereafter referred to as Biotic 1, Biotic 2, Biotic 3 and Biotic 4) incorporating different representations of modelled host plant distributions as predictors, in addition to the climatic variables, were run for each *Gonometa* species.

The different representations of host plant distributions (Fig. 1) were: the probability distribution maps of the individual host trees (Biotic 1), the presence-absence distribution maps of the individual host trees (Biotic 2), the presence-absence distribution maps of all host trees combined into a single host tree presence-absence distribution map (Biotic 3) and the probability distribution maps of all host trees combined into a single host tree probability distribution map (Biotic 4). In the Biotic 1 models, all individual host tree probability maps, as produced by the Maxent output, were included in the moth SDMs as predictors (Fig. 1). In Biotic 2 models, these probability maps were converted to presence-absence maps by applying the "maximum training sensitivity plus specificity" threshold (Liu *et al*., 2005), and the presence-absence map of each host tree species included as a separate predictor in the moth SDMs. In Biotic 3 models, the individual presence-absence maps were combined into a single presence-absence map of where any host tree species of the modelled moth species occur (i.e. all areas where at least one of the host trees occurred was considered a presence). This map was included as a predictor in SDMs. To combine the individual host tree probability distribution maps into a single map of host tree probability (Biotic 4), the individual host tree probability maps were summed to produce one map that was included as a predictor in the moth SDM. Because the two different moth species utilize different host trees, the host tree maps used in models of the two moth species differed.

For all SDMs (host tree and *Gonometa*), localities were divided into five random training (80%) and test (20%) datasets. The model was built using the training dataset, and statistical

model evaluation (see below) was based on the test dataset (Phillips & Dudík 2007). To account for the effects of sampling bias for the models predicting *Gonometa* species distributions, pseudo-absences were chosen randomly from the locality records for all Lepidoptera in southern and East Africa available from GBIF. These pseudo-absence data were selected based on the assumption that sampling biases for Lepidoptera overall will be a good proxy for sampling biases in *Gonometa* species (see Elith & Leathwick 2007; Barbet-Massin *et al*. 2012). For modelling host tree distributions, pseudo-absences were taken from an extensive database of locality records of all African acacia species (Greve *et al*. 2012), as these data were considered to be representative of the tree sampling bias in African savannas. Although this method does not guarantee that pseudo-absences are true absences, as collectors often target specific species and not all species in an area, this method will reduce bias with respect to selecting pseudo-absences from the landscape at random (see Elith *et al*. 2011). The maximum number of background points (i.e. pseudo-absences) was specified to be 5000. Presence-absence maps for the modelled moth distributions were generated using the "maximum training sensitivity plus specificity" threshold (Liu *et al*. 2005).

The performance of our models was assessed using the area under the ROC curve (AUC), true skills statistic (TSS) and test gain. The AUC is a threshold-independent measure of model performance (Thuiller 2003) and its values range from 0.5 to 1, where model performance is excellent if the AUC is > 0.9, good if 0.8 < AUC < 0.9, fair if 0.7 < AUC < 0.8, poor if 0.6 < AUC < 0.7, bad if 0.5 < AUC < 0.6 and no better than random if AUC < 0.5 (Swets 1988). These AUC values represent the ability of a model to discriminate between areas where a species is predicted to be present and those where it is predicted to be absent (Hanley & McNeil 1982). Model performance was also assessed using the TSS (Allouche *et al*. 2006) because the AUC

can be a misleading measure of model performance (Lobo *et al*. 2008). Both omission and commission errors are accounted for by the TSS. Furthermore, TSS is unaffected by prevalence and the size of a validation set (Allouche *et al*. 2006). The TSS values range between -1 and +1, where TSS > 0.75 is excellent, 0.40 < TSS < 0.75 is good, TSS < 0.40 is poor and TSS < 0 indicates that model performance is no better than random (Allouche *et al*. 2006; but see also Landis & Koch 1977). Biotic and abiotic model performances were statistically compared from ten cross-validated replicate models (per abiotic/biotic model) using one-way analysis of variance (ANOVAs). This method of comparison is not particularly robust, as there are a limited number of presence-absence records to test this and tests are not independent of one another; nevertheless, they were run to provide an additional comparison. We also report the test gain, a likelihood statistic which maximises the probability of presence in relation to background data for the data used to evaluate how well Maxent predicts distributions.

To understand which variables were most important for the prediction of *Gonometa* distributions, analysis of variable contribution to the model and estimates of variable importance (using Jackknife tests) were extracted from the Maxent output for each *Gonometa* species.

RESULTS

Model performance and predicted distribution of *Gonometa* species

The models predicting host tree distributions (Fig. 2) had a fair to high predictive power (AUC > 0.70 and TSS > 0.4; Table S2) (except *V. tortilis*, AUC = 0.67, TSS = 0.38) and predicted host tree distributions overlapped fairly well with the *Gonometa* species' distributions (Fig. S4-S5).

The abiotic model predicting *G. postica* distribution performed fairly well (AUC > 0.70 and TSS > 0.5; Table 1). Occurrence records of *G. postica* showed good overlap with the predicted

Table 1. Model performance for the SDMs used to predict *Gonometa* distributions. Abiotic models were run using climatic predictor variables only. Biotic models were run using host tree maps in addition to climatic predictors. In Biotic 1 models individual probability maps of each host plant distribution were used; in Biotic 2 models individual host plant presence-absence maps were used; in Biotic 3 models the combined presence-absence maps of host plant distributions were used; and in Biotic 4, host plant probability maps were summed and used as predictors of moth distribution

Model[†]	Test AUC[‡]	Training AUC[§]	SD AUC[††]	Test gain[‡‡]	TSS[§§]
G. postica (Abiotic)	0.7811	0.7816	0.0634	0.4861	0.5417
G. postica (Biotic 1)	0.7193	0.8473	0.0443	0.2883	0.5917
G. postica (Biotic 2)	0.768	0.8058	0.0724	0.411	0.5878
G. postica (Biotic 3)	0.8097	0.8139	0.068	0.6369	0.5467
G. postica (Biotic 4)	0.7574	0.7951	0.0792	0.4463	0.5864
G. rufobrunnea (Abiotic)	0.7643	0.6906	0.086	0.3062	0.5545
G. rufobrunnea (Biotic 1)	0.8346	0.7387	0.028	0.25	0.7883
G. rufobrunnea (Biotic 2)	0.8513	0.7576	0.0169	0.3991	0.8795
G. rufobrunnea (Biotic 3)	0.8514	0.7529	0.0206	0.8514	0.8488
G. rufobrunnea (Biotic 4)	0.9245	0.7314	0.0151	0.7789	0.9266

[†]Different models used to predict *Gonometa* species distributions. The AUC is the area under the curve and is used as a measure of model performance for [‡]test data (i.e. data set aside for statistical analyses), §training data (i.e. data used to build the model)ˑ [††]Standard deviation of the measure of model performance. [‡‡]Gain of the model: a likelihood statistic (maximising the probability of *Gonometa* presence in relation to background data) for the test data (i.e. data used to evaluate how well Maxent predicts independent data).

[§§] The true skill statistic (TSS) is takes omission and commission errors into account and is unaffected by prevalence and the size of the validation set.
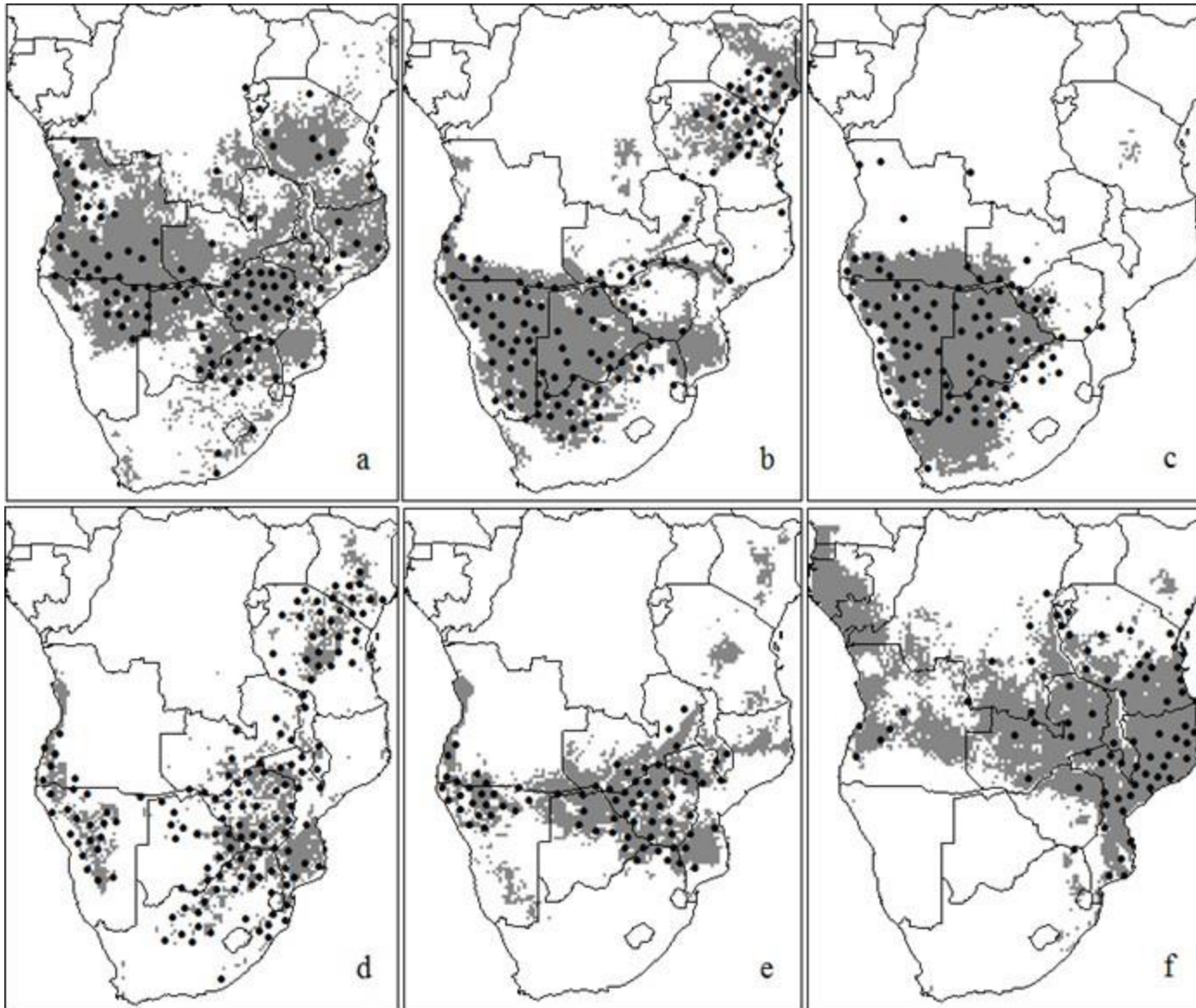
Fig. 2. Predicted presence (grey)-absence (white) maps of the current distribution of (a) *Burkea africana,* (b) *Senegalia mellifera,* (c) *Vachellia erioloba,* (d) *Vachellia tortilis,* (e) C*olophospermum mopane* and (f) *Brachystegia spiciformis.* Known occurrence records (thinned) are illustrated in black. Maps were generated using the "maximum training sensitivity plus specificity" threshold.
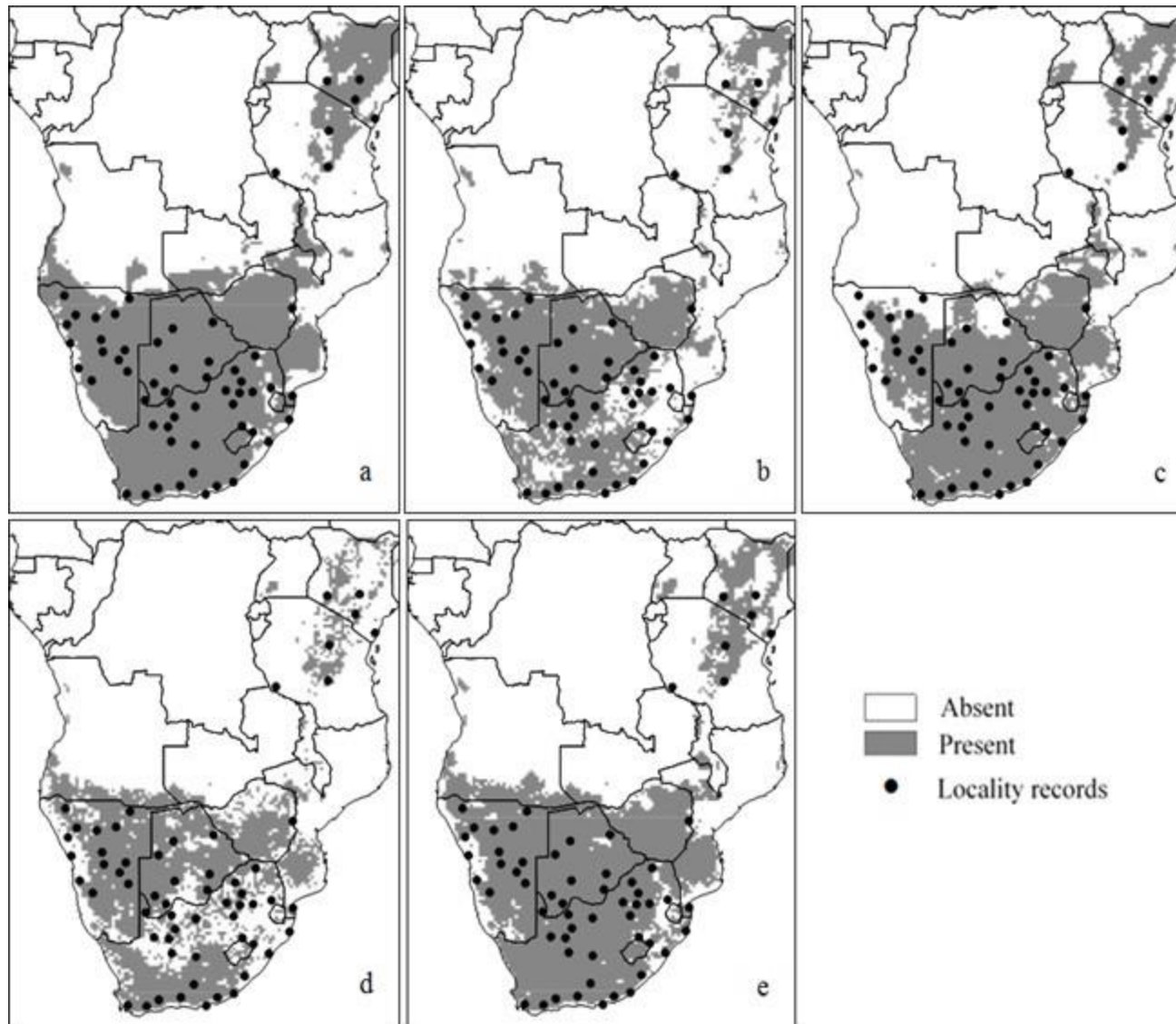
Fig. 3. The abiotic (a) and biotic (b−e) presence-absence maps generated for the current predicted distribution of 622 *Gonometa postica*. In (a) only climatic data were used as predictors (Abiotic model). In (b), the individual probability maps of host plant distributions were used in addition to climate data as predictor variables (Biotic 1). In (c), the individual presence-absence maps of host plant distributions were used in addition to climatic data as predictor variables (Biotic 2). In (d) the combined presence-absence maps of host plant distributions were used in conjunction with climate data as predictor variables (Biotic 3), and in (e), the combined probability maps of host plant distributions were used in conjunction with climate data as predictor variables (Biotic 4). Thinned known occurrence records of *G. postica* are illustrated in black. Presence-absence maps were generated using the "maximum training sensitivity plus specificity" threshold.

distribution from the abiotic model (Fig. 3a). When host plant distributions were incorporated into the model as biotic predictors of *G. postica* distribution (Figs. 3b-e), the predicted range of the species was more restricted than when they were excluded (Fig. 3a). Incorporating biotic interactions into the SDMs of *G. postica* only resulted in improved model performance when the combined presence-absence map of host plant distributions was incorporated as additional predictor, i.e. in the Biotic 3 model (Table 1). Model performances for Biotic 1, Biotic 2 and Biotic 4 were weaker compared to the abiotic model, although these model performances remained fairly good (AUC > 0.70 and TSS > 0.5).

The abiotic model of *G. rufobrunnea* distribution performed fairly well (AUC > 0.75 and TSS > 0.55; Table 1). In comparison, all biotic models of *G. rufobrunnea* distribution performed better than the abiotic model, with higher AUC scores and TSS values (AUC > 0.83 and TSS > 0.78; Table 1). Incorporating the summed host plant probability predictor of *G. rufobrunnra* distribution (Biotic 4) resulted in the best performing model (AUC = 0.92 and TSS = 0.90; Table 1). Again, the abiotic model (Fig. 4a) predicted a broader range of suitable conditions for *G. rufobrunnea* than the biotic models (Fig. 4b-e).

The TSS values did not significantly differ between different model types (*G. postica*: $F_{(45)}$ = 0.08711, P = 0.986; *G. rufobrunnea*: $F_{(45)}$ = 0.5025, P = 0.734), though we did not consider these tests to be particularly robust, as they were built on subsets of the same dataset (see Methods).

Analyses of variable importance (based on Jackknife tests) and variable contribution (obtained as outputs from Maxent) revealed that annual mean temperature, annual precipitation and precipitation seasonality were the most important variables for defining *Gonometa* species distributions in the abiotic models (Table S3, Table S4). When the moth-host plant interactions were incorporated into the models, host tree distributions were important predictors of *G.*
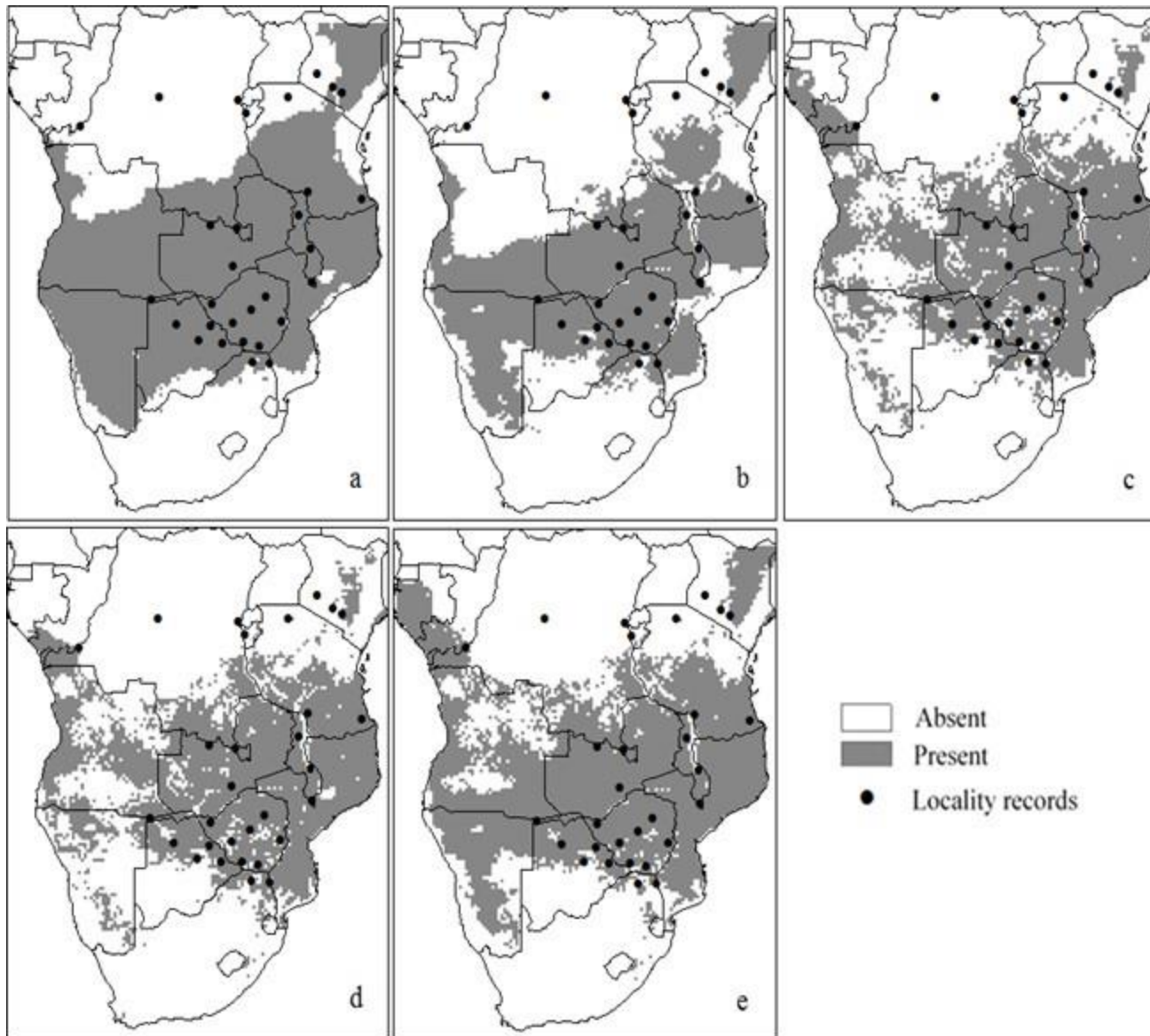
Fig. 4. The abiotic (a) and biotic (b−e) presence-absence maps generated for the current predicted distribution of *Gonometa rufobrunnea*. In (a) only climatic data were used as predictors (Abiotic models). In (b), the individual probability maps of host plant distributions were used in addition to climate data as predictor variables (Biotic 1). In (c), the individual presence-absence maps of host plant distributions were used in addition to climatic data as predictor variables (Biotic 2). In (d) the combined presence-absence maps of host plant distributions were used in conjunction with climate data as predictor variables (Biotic 3), and in (e), the combined probability maps of host plant distributions were used in conjunction with climate data as predictor variables (Biotic 4). Thinned known occurrence records of *G. rufobrunnea* are illustrated in black. Presence-absence maps were generated using the "maximum training sensitivity plus specificity" threshold.

rufobrunnea distribution and contributed strongly to the models predicting *G. postica* distribution, although climatic variables remained most important in biotic models of *G. postica*.

DISCUSSION

This study assessed the performance of SDMs incorporating different types of biotic predictors (along with abiotic variables) against models built using only abiotic predictors, examining *Gonometa* as a study group. Because the association between *Gonometa* species and their host plants is necessary for the completion of the moths' life cycles (Hartland-Rowe 1992; Fening *et al*. 2008; Fening *et al*. 2010), we expected that host plants would strongly shape *Gonometa* distributions, and that incorporating the moth-host plant interaction into the modelling process would improve model performance.

We found that incorporating the moth-host plant biotic interaction into the model improved model performance for *G. rufobrunnea* in all cases; the combined host plant probability maps resulted in the best performing *G. rufobrunnea* model. However, incorporating the moth-host plant biotic interaction only improved *G. postica* model performance in one instance, i.e. when the combined host plant presence-absence map was included in the model. Given the two *Gonometa* species differ in terms of host plant species range (Veldtman *et al*. 2002; Veldtman *et al*. 2007), it could be expected that the biotic interaction would have a stronger impact on *G. rufobrunnea* given its more limited host species range. Thus, although at least one of biotic models performed better than the abiotic model for both moth species (in agreement with e.g. Araújo & Luoto 2007; Giannini *et al*. 2013), which representation of the biotic interaction gave the best outcome was inconsistent between the species, despite them being closely related (Veldtman *et al*. 2004; Veldtman *et al*. 2007).

Other studies have shown that incorporation of biotic interactions can improve model explanatory power at both local (Leathwick & Austin 2001; Leathwick 2002; Pellissier *et al*. 2010) and broader (Araújo & Luoto 2007; Giannini *et al*. 2013) spatial scales, although there are exceptions (McPherson & Jetz 2007; Bateman *et al*. 2012; Silva *et al*. 2014). Studies which used similar methods to ours to model species distributions also showed that host plant occurrence in conjunction with climatic predictors improved model performances significantly compared to climate-only models (i.e. Araújo & Luoto 2007; Preston *et al*. 2008; Bateman *et al*. 2012), although in some studies this did not improve predictive ability of models and even reduced model performances (e.g. Silva *et al*. 2014).

The abiotic models for both *G. postica* and *G. rufobrunnea* predicted larger moth ranges than the biotic models, suggesting that host plant distributions may constrain the distribution of *Gonometa*, likely due to the association between *Gonometa* species and their host plants for development and survival (Veldtman *et al*. 2007; Fening *et al*. 2008; Fening *et al*. 2010). A similar result was observed by Silva *et al*. (2014) who demonstrated that including biotic variables (i.e. host plant distributions) into distribution models of a bee species (*Lithurgus huberi*) did not improve the models' predictive abilities; however, host plant species' distributions constrained the distribution of the bee.

Annual precipitation, annual mean temperature and precipitation seasonality were important predictors of *Gonometa* species distribution irrespective of the inclusion of host plant data, in broad agreement with studies showing that *Gonometa* species are dependent on both temperature and precipitation cues for outbreak initiation and emergence from cocoons (Hartland-Rowe 1992, Delport 2006).

Whether the inclusion of biotic variables in lepidopteran SDMs will consistently improve model predictions remains to be determined. Many studies have concluded that climate is the main driver of lepidopteran distributions (Araújo & Luoto 2007; Menèndez *et al*. 2007; Schweiger *et al*. 2012). However, biotic variables can increase or decrease model performance (Araújo & Luoto 2007), and while several studies find improved predictions if biotic variables are included (Araújo & Luoto 2007; Heikkinen *et al*. 2007; Gilman *et al*. 2010; Meier *et al*. 2010; Bateman *et al*. 2012; Boulangeat *et al*. 2012; Schweiger *et al*. 2012; Wisz *et al*. 2013), the type of model and the species used may determine whether model performance is improved by biotic predictors (Araújo & Luoto 2007; Giannini *et al*. 2013). Furthermore, biotic interactions (e.g. competition) may impact on species distributions at spatial scales that are much smaller than the course-grained scales that are typically used in ecological niche modelling (Eltonian Noise Hypothesis, Soberón & Nakamura 2009). In addition, the particular species (e.g. the life history traits of the particular species) being modelled may also determine the importance of biotic variables in predicting species ranges, e.g. the influence of biotic interactions may vary with the degree of specialisation of species (Hernandez *et al*. 2006; Araújo & Luoto 2007; Menèndez *et al*. 2007; Preston *et al*. 2008; Giannini *et al*. 2013), the direct effect of climate on species biology or the level of environmental stress experienced by species (Hanspach *et al*. 2014). Furthermore, the strength of the interaction (i.e. obligate versus facultative interaction) between a modelled organism and the species that it interacts with may influence the contribution of biotic variables to the distribution of a species. Clearly, the influence of inter-specific interactions on species distributions is species- and context-specific (Schweiger *et al*. 2012; le Roux *et al*. 2012; Eskildsen *et al*. 2013).

Due to the simple nature of the SDMs used here, some aspects of the study species' biology were not included in the models (see also Text S1). *Gonometa* species, which are characterised by eruptive, unpredictable population cycles and display large inter-annual fluctuations in population size over different spatial and temporal scales (Delport 2006; Veldtman *et al*. 2007), are probably influenced by stochasticity within their environments, which could influence model performance (Pöyry *et al*. 2008), but can be difficult to incorporate into SDMs at broad scales. Such population fluctuations are characteristic of many moth species (Alalouini *et al*. 2013; Tamburini *et al*. 2013) and can be attributed to both density-dependent factors (e.g. presence of host trees and natural enemies), or density-independent factors (e.g. climate). Both *Gonometa* species occur in regions that display high inter-annual variation in precipitation and display high geographic and temporal variation in population outbreaks. Therefore, information on when and where *Gonometa* outbreaks have occurred in the past, and how these outbreaks relate to annual variation in climate would be ideal for modelling *Gonometa* distributions. Currently, this information is not available.

In addition, the modelled species is could be influenced by a range of biotic interactions (e.g. predation and parasitism) that were not included in our models (Case *et al*. 2005; Schweiger *et al*. 2012; Anderson 2017). Some of these interactions can be labile, and their effects can range from antagonistic to mutualistic. More sophisticated modelling techniques, e.g. where error matrices are used in multivariate regression models to incorporate biotic interactions between multiple species, are available (Kissling *et al*. 2012), but to accurately model distributions based on all mutualistic, competitive or consumer-resource interactions, a large body of information about the ecological networks surrounding them (e.g. the impact of natural enemies) and the natural history of the modelled species are needed; this is information which is often not

available and may be complex to interpret (Soberón & Peterson 2005). In addition, some spatial mismatches between the distributions of host trees and moth species occurs (see Schweiger *et al*. 2008; Mateo *et al*. 2010). For example, although *Gonometa* are mainly found on the host tree species mentioned here, there are isolated cases of them occurring on other species, e.g. the invasive *Acacia mearnsii* in the Western Cape. Equally, the host trees occur in areas where the lepidopterans have not been recorded (Fig. S4, S5). Therefore, a better understanding of *Gonometa* interactions with other species (e.g. examining possible regional preference of *Gonometa* for certain host plant species over others and investigating the effect of natural enemies on *Gonometa* distributions and dynamics) may improve predictions of *Gonometa* distributions.

Conclusions

This study is comparable to previous research that has found improved SDM performance after incorporating consumer-resource interactions into models when predicting broad-scale species distributions. However, the incorporation of biotic variables did not always improve model performance, and there was no consistency in which biotic model most improved model predictions. Nonetheless, our biotic models resulted in more constrained distributions compared to climate-only models (e.g. Silva *et al*. 2014). Therefore, the impact of biotic interactions on species distributions may be context-specific, even when the species involved have obligatory interactions with other organisms.

REFERENCES

Aiello-Lammens M. E., Boria R. A., Radosavljevic A., Vilela B. & Anderson R. P. (2015) spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecology* **38,** 541–545.

Akai H. (1997) Anti-bacteria function of natural silk materials. *International Journal of Wild Silkmoth Silk* **3,** 79–81.

Allouche O., Tsoar A. & Kadmon R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43,** 1223–1232.

Anderson R. P. (2017) When and how should biotic interactions be considered in models of species niches and distributions? *Journal of Biogeography* **44,** 8–17.

Araújo M. B. & Guisan A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography* **33,** 1677–1688.

Araújo M. B. & Luoto M. (2007) The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography* **16,** 743–753.

Barbet-Massin M., Jiguet F., Albert C. H. & Thuiller W. (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* **3,** 327–338.

Bateman B. L., VanDerWal J., Williams S. E. & Johnson C. N. (2012) Biotic interactions influence the projected distribution of a specialist mammal under climate change. *Diversity and Distributions* **18,** 861–872.

Beaumont L. J. & Hughes L. (2002) Potential changes in the distributions of latitudinally restricted Australian butterfly species in response to climate change. *Global Change Biology* **8,** 954–971.

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

Case T. J., Holt R. D., McPeek M. A. & Keitt T. H. (2005) The community context of species' borders: ecological and evolutionary perspectives. *Oikos* **108,** 28–46.

Chikwenhere G. P. (1992) Distribution and abundance of the African silkmoth (*Gonometa rufobrunnea*) in Matebeleland Province of Zimbabwe. *Zimbabwe Journal of Agricultural Research* **29,** 133–137.

Delport W. (2006) Population genetic inference of demographic processes in the African Wild Silk Moth, *Gonometa postica* (Lasiocampidae). PhD thesis, University of Pretoria, South Africa.

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. McC., Peterson A. T., Phillips S. J., Richardson K. S., 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.

Elith J. & Leathwick J. (2007) Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions* **13,** 265–275.

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

Elith J., Phillips S. J., Hastie T., Dudík M., Chee Y. E. & Yates C. J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17,** 43−57.

Elith J. & Franklin J. (2013) Species distribution modelling. Encyclopaedia of Biodiversity (ed.by S. Levin), pp. 692–705. Academic Press, Waltham, MA.

Eskildsen A., le Roux P. C., Heikkinen R. K., Høye T. T., Kissling W. D., Pöyry J., Wisz M. S. & Luoto M. (2013) Testing species distribution models across space and time: high latitude butterflies and recent warming. *Global Ecology and Biogeography* **22,** 1293–1303.

Fening K. O., Kioko E. N., Raina S. K. & Mueke J. M. (2008) Monitoring wild silkmoth, *Gonometa postica* Walker, abundance, host plant diversity and distribution in Imba and Mumoni woodlands in Mwingi, Kenya. *International Journal of Biodiversity Science and Management* **4,** 104–111.

Fening K. O., Kioko E. N., Raina S. K. & Mueke J. M. (2010) Effects of seasons and larval food plants on the quality of *Gonometa postica* cocoons. *Phytoparasitica* **38,** 111–119.

Giannini T. C., Chapman D.S., Saraiva A. M., Alves-dos-Santos I. & Biesmeijer J. C. (2013) Improving species distribution models using biotic interactions: a case study of parasites, pollinators and plants. *Ecography* **36,** 649–656.

Gilman S. E., Urban M. C., Tewksbury J., Gilchrist G. W. & Holt R. D. (2010) A framework for community interactions under climate change. *Trends in Ecology and Evolution* **25,** 325–331.

Godsoe W. & Harmon L. J. (2012) How do species interactions affect species distribution models? *Ecography* **35,** 811–820.

Greve M., Lykke A. M., Fagg C. W., Bogaert J., Friis I., Marchant R., Marshall A. R., Ndayishimiye J., Sandel B. S., Sandom C., Schmidt M., Timberlake J. R., Wieringa J. J.,

Zizka G. & Svenning J-C. (2012) Continental-scale variability in browser diversity is a major driver of diversity patterns in acacias across Africa. *Journal of Ecology* **100,** 1093–1104.

Guisan A. T. & Zimmermann N. E. (2000) Predictive habitat distribution models in ecology. *Ecological modelling* **135,** 146–186.

Guisan A. T., Lehmann A., Ferrier S., Austin M., Overton J. MC. C., Aspinall R. & Hastie T. (2006) Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology* **43,** 386–392.

Hanley J. A. & McNeil B. J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143,** 29–36.

Hanspach J., Schweiger O., Kühn I., Plattner M., Pearman P. B., Zimmermann N. E. & Settele J. (2014) Host plant availability potentially limits butterfly distributions under cold environmental conditions. *Ecography* **37,** 301-308.

Hartland-Rowe R. (1992) The biology of the wild silkmoth *Gonometa rufobrunnea* Aurivillius (Lasiocampidae) in northeastern Botswana, with comments on its potential as a source of wild silk. *Botswana Notes and Records* **24,** 123–133.

Hengl T., Mendes de Jesus J., MacMillan R. A., Batjes N. H., Heuvelink G.B. M., Ribeiro E., Samuel-Rosa A., Kempen B., Leenaars J. G. B., Walsh M. G. & Gonzalez M. R. (2014) SoilGrids1km — Global Soil Information Based on Automated Mapping. *PLoS ONE* **9,** 1–17.

Heikkinen R. K., Luoto M., Virkkala R., Pearson R. G. & Köber J-H. (2007) Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography* **16,** 754–763.

Hernandez P. A., Graham C. H., Master L. L. & Albert D. L. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* **29,** 773–785.

Hickler T., Vohland K., Feehan J., Miller P. A., Smith B., Costa L., Giesecke T., Fronzek S., Carter T. R., Cramer W., Kühn I. & Sykes M. T. (2012) Projecting the future distribution of European potential natural vegetation zones with a generalized, tree species-based dynamic vegetation model. *Global Ecology and Biogeography* **21,** 50-63.

Hijmans R. J., Cameron S. E., Parra J. L., Jones P. G. & Jarvis A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25,** 1965–1978.

Huntley B. J. & Walker B. H. (1982) *Ecology of Tropical Savannas*. New York: Springer-Verlag, Berlin.

Jiménez-Valverde A., Lobo J. M. & Hortal J. (2008) Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions* **14,** 885-890.

Kearney M. (2006) Habitat, environment and niche: what are we modelling? *Oikos* **155,** 186–191.

Kissling W. D., Dormann C. F., Groeneveld J., Hickler T., Kühn I., McInerny G. J., Montoya J. M., Römermann C., Schaffers K., Schurr F. M., Singer A., Svenning J-C., Zimmermann N. E. & O'Hara R.B. (2012) Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography* **39,** 2163–2178.

Kramer-Schadt S., Niedballa J., Pilgrim J. D., Schröder B., Lindenborn J., Reinfelder V., Stillfried M., Heckmann I., Scharf A. K., Augeri D. M., Cheyne S. M., Hearn A. J., Ross J., Macdonald D. W., Mathai J., Eaton J., Marshall A. J., Semiadi G., Rustam R., Bernard H.,

Alfred R., Samejima H., Duckworth J. W., Breitenmoser-Wuersten C., Belant J. L., Hofer H. & Wilting A. (2013) The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions* **19,** 1366–1379.

Kroon D. M. (1999) Lepidoptera of Southern Africa: Host–Plants and Other Associations, A catalogue. Lepidopterists' Society of Africa, Salisburgtd.

Landis J. R. & Koch G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* **33,** 159–174.

le Roux P. C., Virtanen R., Heikkinen R. K. & Luoto M. (2012) Biotic interactions affect the elevational ranges of high-latitude plant species. *Ecography* **35,** 1048–1056.

Leathwick J. R. & Austin M. P. (2001) Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology* **82,** 2560–2573.

Leathwick J. R. (2002) Intra-generic competition among *Nothofagus* in New Zealand's primary indigenous forests. *Biodiversity and Conservation* **11,** 2117–2187.

Lischke H., Zimmermanna N. E., Bolligera J., Rickebuscha S. & Löfflerb T. J. (2006) TreeMig: A forest-landscape model for simulating spatio-temporal patterns from stand to landscape scale. *Ecological Modelling* **119,** 409–420.

Liu C., Berry P. M., Dawson T. P. & Pearson R. G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* **28,** 385–393.

Lobo J. M., Jiménez-Valverde A., Real R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17,** 145–151.

Mateo R. G., Felicísimo Á. M. & Muñoz J. (2010) Effects of the number of presences on reliability and stability of MARS species distribution models: the importance of regional niche variation and ecological heterogeneity. *Journal of Vegetation Science* **21**, 908–922.

McPherson J. M. & Jetz W. (2007) Effects of species' ecology on the accuracy of distribution models. *Ecography* **30,** 135–151.

Meier S. M., Kienast F., Pearman P. B., Svenning J-C., Thuiller W., Araújo M. B, Guisan A. & Zimmermann N. E. (2010) Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography* **33,** 1038–1048.

Pearson R. G. & Dawson T. P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* **12,** 361–371.

Pellissier L., Bråthen K. A., Pottier J., Randin C. F., Vittoz P., Dubuis A., Yoccoz N. G., Alm T., Zimmermann N. E. & Guisan A. (2010) Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography* **33,** 1004–1014.

Phillips S. J., Anderson R. P. & Schapire R. E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190,** 231–259.

Phillips S. & Dudík M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31,** 161–175.

Phillips S. J., Dudík M., Elith J., Graham C. H., Lehmann A., Leathwick J. & Ferrier S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* **19,** 181–197.

Pöyry J., Luoto M., Heikkinen R. K. & Saarinen K. (2008) Species traits are associated with the quality of bioclimatic models. *Global Ecology and Biogeography* **17,** 403–414.

Preston K. L., Rotenberry J. T., Redak R. A. & Allen M. F. (2008) Habitat shifts of endangered species under altered climate conditions: importance of biotic interactions. *Global Change Biology* **14,** 2501–2515.

Pulliam H. R. (2000) On the relationship between niche and distribution. *Ecology Letters* **3,** 349–361.

Schweiger O., Settele J., Kudrna O., Klotz S. & Kühn I. (2008) Climate change can cause spatial mismatch of trophically interacting species. *Ecology* **89,** 3472–3479.

Schweiger O., Heikkinen R. K., Harpke A., Hickler T., Klotz S., Kudrna O., Kühn I., Pöyry J. & Settele J. (2012) Increasing range mismatching of interacting species under global change is related to their ecological characteristics. *Global Ecology and Biogeography* **21,** 88–99.

Silva D. P., Gonzalez V. H., Melo G. A. R., Lucia M., Alvarez L. J. & De Marco Jr. P. (2014) Seeking the flowers for the bees: Integrating biotic interactions into niche models to assess the distribution of the exotic bee species *Lithurgus huberi* in South America. *Ecological Modelling* **273,** 200–209.

Smith T. M. & Goodman P. S. (1986) Successional dynamics in an *Acacia nilotica-Euclea divinorum* savannah in southern Africa. *Journal of Ecology* **75,** 603–610.

Soberón J. & Peterson A. T. (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* **2,** 1–10.

Soberón J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* **10,** 1115–1123.

Soberón J. & Nakamura M. (2009) Niches and distributional areas: Concepts, methods and assumptions. *PNAS* **106,** 19644–19650.

Swets J. A. (1988). Measuring the accuracy of diagnostic systems. *Science* **240,** 1285–1293.

Tamburini G., Marini L., Hellrigl K., Salvadori C. & Battiste A. (2013) Effects of climate and density-dependent factors on population dynamics of the pine processionary moth in the Southern Alps. *Climate Change* **121,** 701–712.

Thuiller W. (2003) BIOMOD: optimising predictions of species distributions and projecting potential future shifts under global change. Global Change Biology, 9, 1353–1362

Tóth J. P., Varga K., Végvári Z. & Varga Z. (2013) Distribution of the Eastern knapweed fritillary (*Melitaea ornata* Cristoph, 1893) (Lepidoptera: Nymphalidae): past, present and future. *Journal of Insect Conservation* **17,** 245–255.

Van der Putten W. H., Macel M. & Visser M. E. (2010) Predicting species distribution and abundance responses to climate change: why it is essential to include biotic interactions across trophic levels. *Philosophical Transactions of the Royal Society B* **365,** 2025–2034.

Veldtman R., McGeoch M. A. & Scholtz C. H. (2002) Variation in cocoon size in southern African wild silk moths: implications for sustainable harvesting. *African Entomology* **10,** 127–136.

Veldtman R., McGeoch M. A. & Scholtz C. H. (2004) Parasitoids of the southern African wild silk moths (Lepidoptera). *African Entomology* **12,** 117–122.

Veldtman R., McGeoch M. A. & Scholtz C. H. (2007) Can life-history and defence traits predict the population dynamics and natural enemy responses of insect herbivores? *Ecological Entomology* **32,** 662–673.

Wiens J. A., Stralberg D., Jongsomjit D., Howell C. A. & Snyder M. A. (2009) Niches, models, and climate change: Assessing the assumptions and uncertainties. *PNAS* **106,** 19729 –19736.

Wisz M. S., Pottier J., Kissling W. D., Pellissier L., Lenoir J., Damgaard C. F., Dormann C. F., Forchhammer M. C., Grytnes J-A., Guisan A., Heikkinen R. K., Høye T. T., Kühn I., Luoto M., Maiorano L., Nilsson M-C., Normand S., Öckinger E., Schmidt N. M., Termansen M., Timmermann A., Wardle D. A., Aastrup P. & Svenning J-C. (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews* **88,** 15–30.

Appendix S1.

Table S1. Table showing the original number of occurrence records for each species and the number of occurrence records retained for each species following 20 repetitions of thinning using a thinning distance of 100 Km between records (see main text for more details).

Species	Original no. occurrences	Max. no. occurrences retained
G. postica	302	60
G. rufobrunnea	89	30
V. erioloba	1276	101
S. mellifera	1396	138
V. tortilis	1372	153
B. africana	1051	126
C. mopane	830	77
B. spiciformis	409	79

The AUC is the area under the receiver operating characteristic (ROC) curve and is used as a measure of model performance for [†]test data (i.e. data set aside for statistical analyses), [‡]training data (i.e. data used to build the model). [§]Standard deviation of the measure of model performance. [††]Gain of the model: a likelihood statistic (maximising the probability of host tree presence in relation to background data) for the test data (i.e. data used to evaluate how well Maxent predicts independent data), [§§] The true skill statistic (TSS) is takes omission and commission errors into account and is unaffected by prevalence and the size of the validation set.

Table S2. Model performance for the predictions of the *Gonometa* host tree distributions: *Burkea africana*, *Senegalia mellifera*, *Vachellia erioloba* and *V. tortilis* (*G. postica* hosts), and *Colophospermum mopane* and *Brachystegia spicifomis* (*G. rufobrunnea* hosts). Models were run using climatic and soil predictor variables

Model	Test AUC†	Training AUC‡	SD AUC§	Test gain††	‡‡TSS
B. africana	0.7646	0.7407	0.0538	0.4066	0.42993
S. mellifera	0.7299	0.8008	0.0318	0.3554	0.44561
V. erioloba	0.8259	- 0.8499	0.0266	0.7462	0.66672
V. tortilis	0.6745	0.8024	0.0315	0.2427	0.38337
C. mopane	0.8485	0.8546	0.0307	0.8634	0.64899
B. spiciformis	0.9377	0.8467	0.0352	0.9377	0.65092

Table S3. Important predictor variables in Maxent models of *Gonometa postica* distribution. Displayed variables had the highest contribution to the models and also yielded the highest training gain and AUC scores based on Jackknife tests of variable importance, which was the basis for variable selection and variable representation in the table. Variables marked with asterisks did not contribute highly to the models based the analysis of variable contribution, but were important predictors of moth distribution based on the Jackknife tests of variable importance (training gain and AUC columns)

Model	Variable	Contribution to model (%)[†]	Permutation importance[‡]	Training gain[§] (without variable)	Training gain[††] (variable only)	AUC[‡‡] (without variable)	AUC[§§] (variable only)
G. postica (Abiotic)	Annual Precipitation	69.0116	73.9499	0.2582	0.3439	0.8289	0.8223
	Annual Mean Temperature*	15.7964	7.0404	0.3942	0.1509	0.7426	0.7655
	Vachellia tortilis	30.7355	16.3532	0.4375	0.188	0.7212	0.6265
G. postica (Biotic 1)	*Vachellia erioloba*	26.9526	6.2948	0.4641	0.2248	0.7384	0.8171
	Annual Precipitation*	15.483	47.7332	0.4293	0.337	0.7656	0.8164
	Annual Precipitation	40.819	55.3198	0.3436	0.317	0.7762	0.8199
G. postica (Biotic 2)	Annual Mean Temperature	27.437	0	0.4191	0.1897	0.7466	0.64
	*Vachellia erioloba**	13.9353	0	0.4143	0.0908	0.7488	0.6849
	Annual Precipitation	50.7669	61.5029	0.3621	0.337	0.8236	0.8164
G. postica (Biotic 3)	Host trees	23.2795	10.2607	0.4116	0.1347	0.7323	0.7549
	Annual Mean Temperature*	13.1277	7.7395	0.4492	0.1474	0.7877	0.7625
	Annual Precipitation	62.4444	66.4212	0.3288	0.337	0.1038	0.8164
G. postica (Biotic 4)	Annual Mean Temperature*	12.6867	7.7637	0.419	0.1474	0.1038	0.7625
	Host plant distribution*	10.6005	6.8206	0.4116	0.1038	0.7323	0.7065

[†] Percentage contribution of each of the variables to the models. [‡]Permutation importance is determined by randomly permuting values of the variable in question among the training points. The gain given in the table starts at 0 during a model run, increasing towards an asymptote and gives the maximum likelihood ratios of average presences to average background points. The [§]training gain without the variable represents the model gain once the variable in question has been removed from the model, whereas the [††]training gain with the variable only, represents the gain of the model when the variable in question is considered in isolation from all the other variables. The [‡‡]AUC (without variable) represents the predictive power of the model when the variables in question are excluded from the model, and the [§§]AUC (variable only) represents AUC from a model containing only the predictor variable.

Table S4. Important predictor variables in Maxent models of *Gonometa rufobrunnea* distribution. Displayed variables had the highest contribution to the models and also yielded the highest training gain and AUC scores based on Jackknife tests of variable importance, which was the basis for variable selection and variable representation in the table.

Species model	Variable	Contribution to model (%)[†]	Permutation importance[‡]	Training gain[§] (without variable)	Training gain[††] (variable only)	AUC[‡‡] (without variable)	AUC[§§] (variable only)
G. rufobrunnea (Abiotic)	Precipitation Seasonality	99.9848	99.4354	0.041	0.1206	0.6747	0.7628
G. rufobrunnea (Biotic 1)	*Colophospermum mopane*	68.7705	8.9111	0.1751	0.2035	0.8383	0.84
	Precipitation Seasonality	30.8968	78.9111	0.2288	0.1751	0.2035	0.8383
G. rufobrunnea (Biotic 2)	*Colophospermum mopane*	63.1869	28.5087	0.1861	0.1659	0.878	0.4605
	Precipitation Seasonality	9.3046	52.5995	0.2447	0.1751	0.8043	0.8383
G. rufobrunnea (Biotic 3)	Host trees	96.1712	52.0275	0.1751	0.3038	0.8383	0.8431
	Precipitation Seasonality	3.7054	40.7638	0.3135	0.1751	0.8784	0.8383
G. rufobrunnea (Biotic 4)	Host trees	67.1476	19.6506	0.1751	0.1775	0.8383	0.9512
	Precipitation Seasonality	32.8524	80.3494	0.1902	0.1751	0.9437	0.8383

[†] Percentage contribution of each of the variables to the models. [‡]Permutation importance is determined by randomly permuting values of the variable in question among the training points. The gain given in the table starts at 0 during a model run, increasing towards an asymptote and gives the maximum likelihood ratios of average presences to average background points. The [§]training gain without the variable represents the model gain once the variable in question has been removed from the model, whereas the [††]training gain with the variable only, represents the gain of the model when the variable in question is considered in isolation from all the other variables. The [‡‡]AUC (without variable) represents the predictive power of the model when the variables in question are excluded from the model, and the [§§]AUC (variable only) represents AUC from a model containing only the predictor variable.
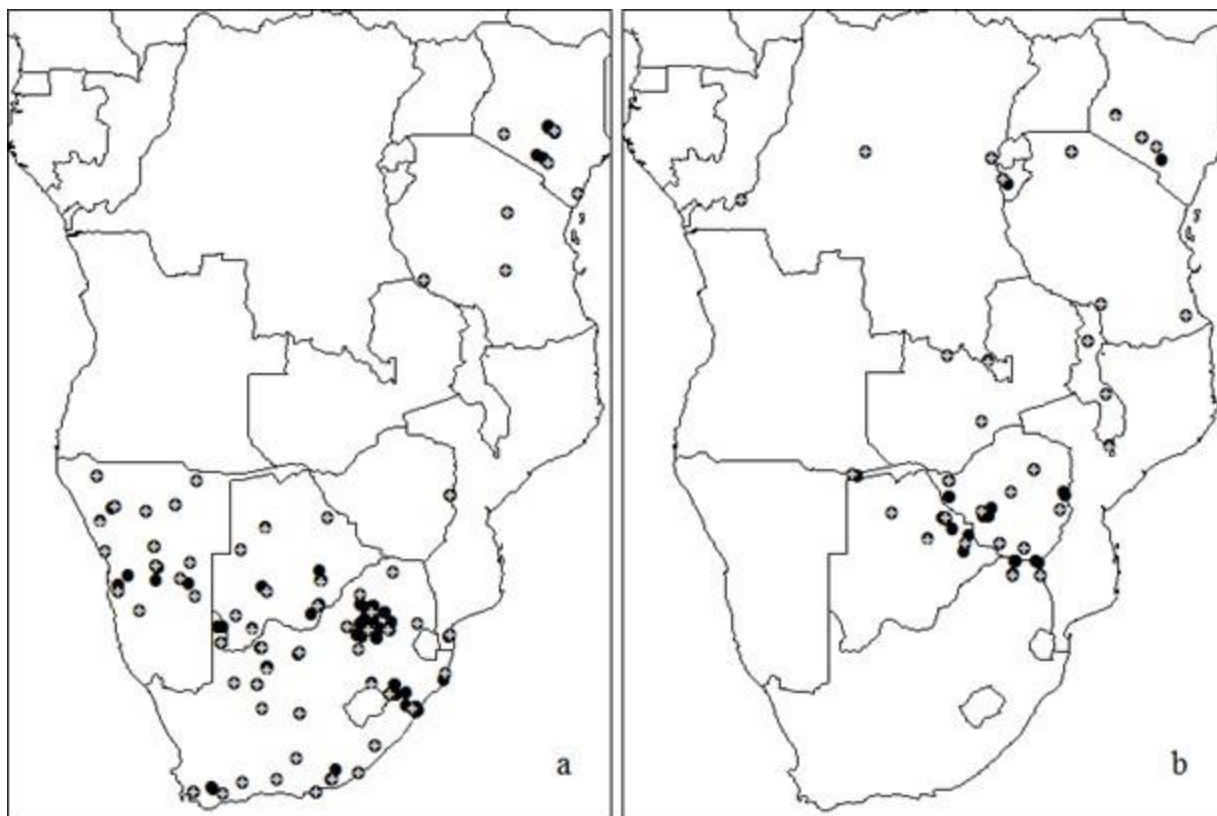
Fig. S1. Locality records of a) *Gonometa postica* and b) *Gonometa rufobrunnea* showing retained (grey crosses) and dropped (black dots) occurrence records following thinning.
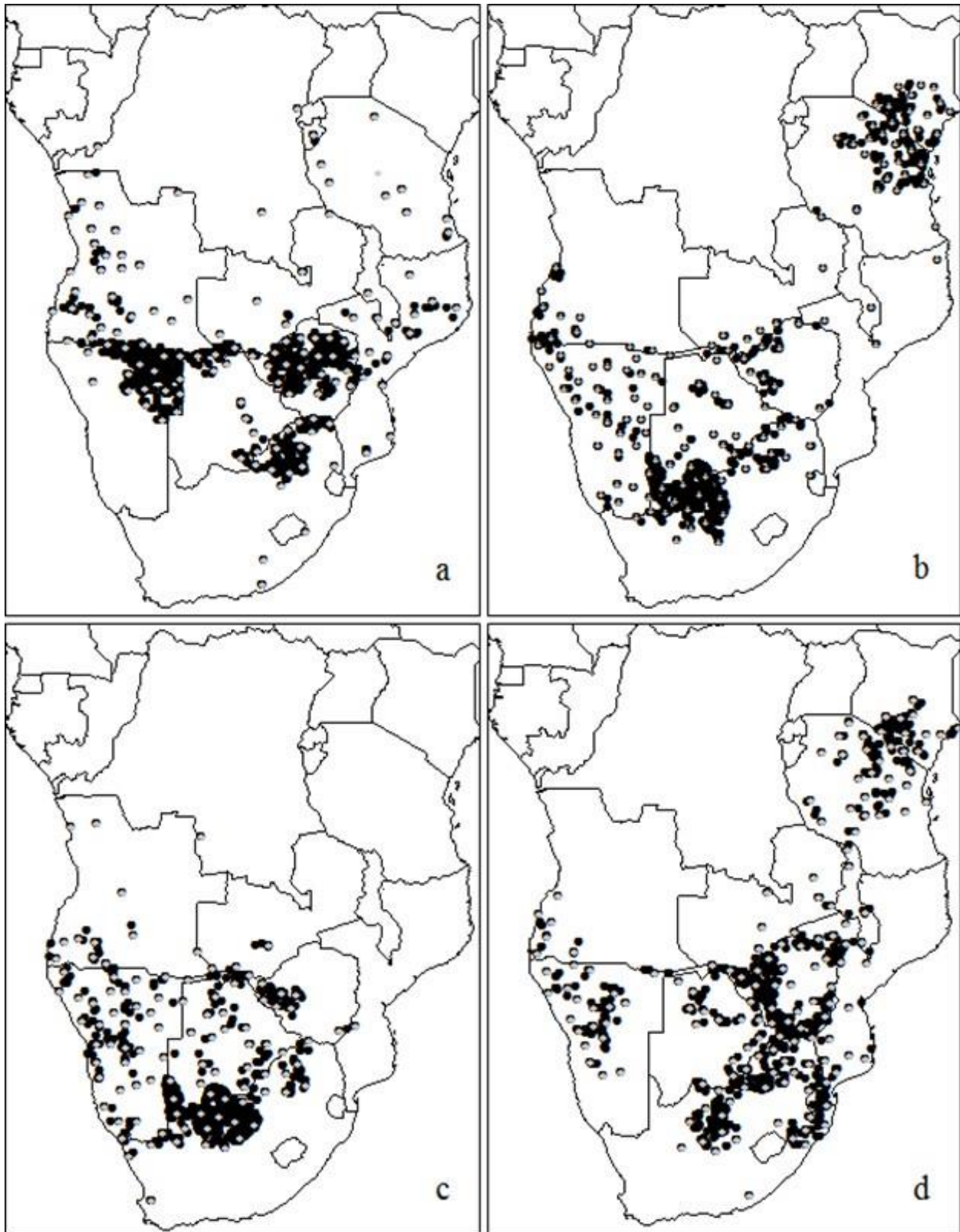
Fig. S2. Locality records of a) *Burkea africana*, (b) *Senegalia mellifera*, (c) *Vachellia erioloba*, and (d) *Vachellia tortilis* showing retained (grey crosses) and dropped (black dots) occurrence records following thinning.
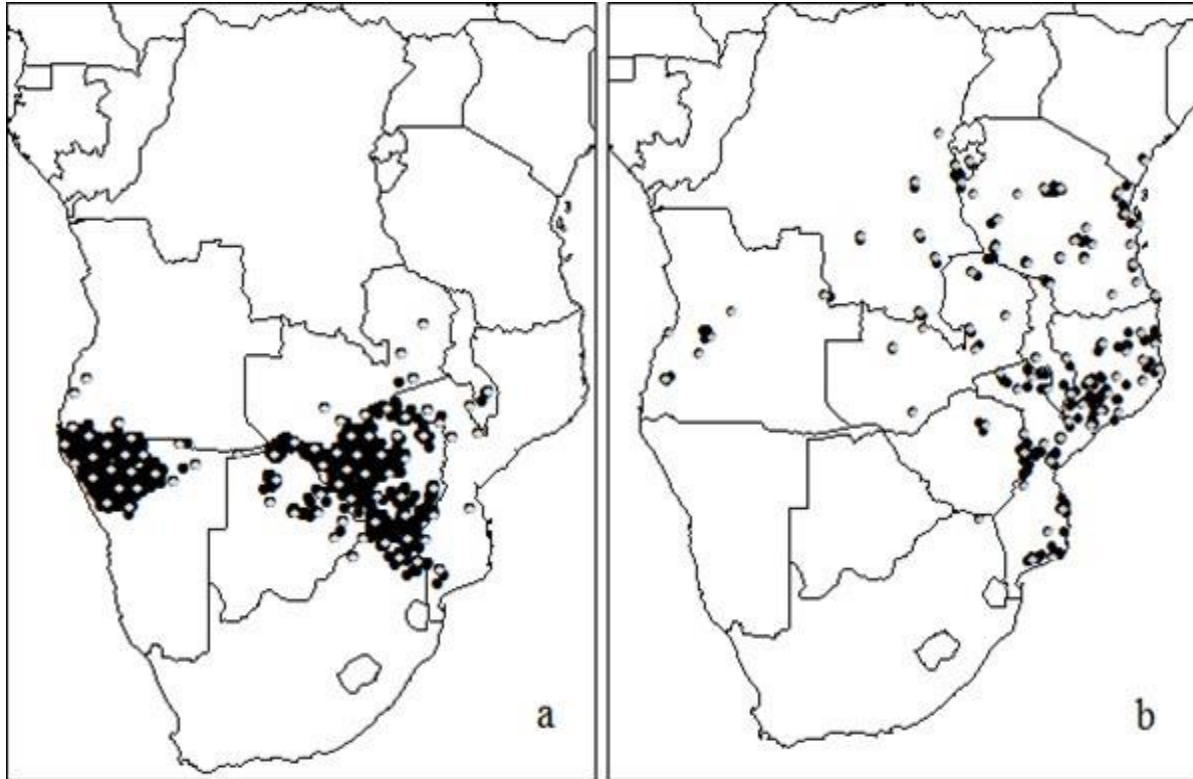
Fig. S3. Locality records of a) *Colophospermum mopane* and (b) *Brachystegia spiciformis* showing retained

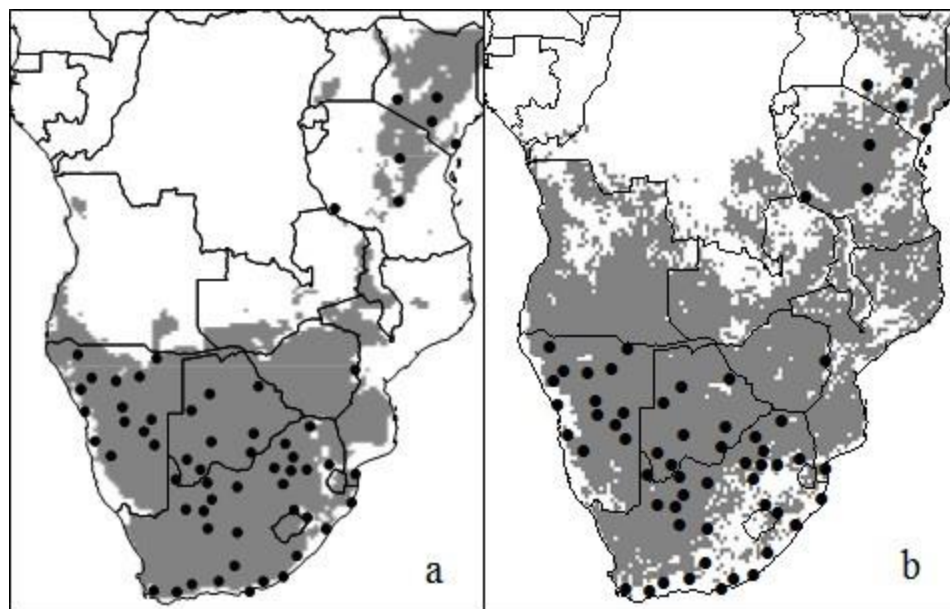(grey crosses) and dropped (black dots) occurrence records following thinning.

Fig. S4. Comparison between the distribution of a) *Gonometa postica* (from model using only abiotic variables) and b) the combined presence-absence maps of *G. postica* host plants (*Burkea africana*, *Senegalia mellifera*, *Vachellia erioloba* and *Vachellia tortilis*). *Gonometa postica* occurrence records are indicated by black dots on the maps.
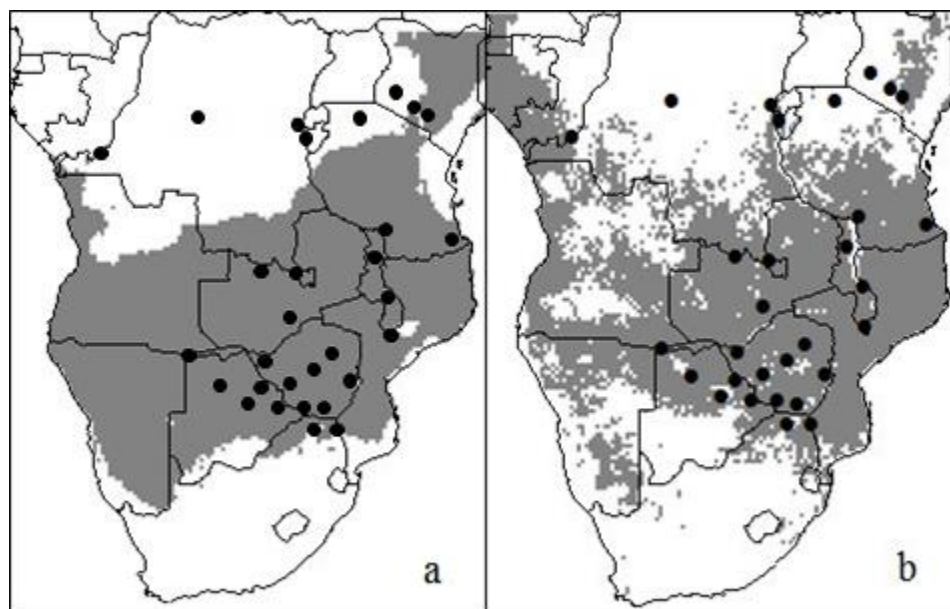

Fig. S5. Comparison between the distribution of a) *Gonometa rufobrunnea* (from model using only abiotic variables) and b) the combined presence-absence maps of *G. rufobrunnea* host plants (*Colophospermum mopane* and *Brachystegia spiciformis*). *Gonometa rufobrunnea* occurrence records are illustrated in black on both maps.

Text S1. Potential shortcomings of SDM performance that could have influenced model performance.

A number of issues can influence SDM model performance. While we included abiotic and biotic interactions in the models, dispersal limitation, which was not considered here, may prevent species from occurring in areas where the climate is suitable (Blach-Overgaard *et al*. 2010), leading to contingent absences (Lobo *et al*. 2010). However, *Gonometa* have the ability to disperse over fairly large distances (Veldtman *et al*. 2002; Veldtman *et al*. 2007); therefore, it is unlikely that many false presences were predicted due to dispersal limitation. In addition, because of the eruptive life cycle of *Gonometa* populations, a locality recorded as a presence during one time period may be recorded as an absence in another time period. Such sampling bias would be easily accounted for if the location of collection could be matched to seasonal weather patterns and so incorporates into models. However, this information is difficult to come by. This scarcity of knowledge on species distributions that exists is referred to as the Wallacean shortfall (Lomolino 2004) and is a reality for many organisms, especially invertebrates (Beck *et al*. 2013).