FORUM

Maxent is not a presence–absence method: a comment on Thibaud *et al.*

Gurutzeta Guillera-Arroita*, José J. Lahoz-Monfort and Jane Elith

School of Botany, University of Melbourne, Parkville, Victoria, Australia

Summary

1. Thibaud *et al.* (*Methods in Ecology and Evolution* 2014) present a framework for simulating species and evaluating the relative effects of factors affecting the predictions from species distribution models (SDMs). They demonstrate their approach by generating presence–absence data sets for different simulated species and analysing them using four modelling methods: three presence–absence methods and Maxent, which is a presence-background modelling tool. One of their results is striking: that their use of Maxent performs well in estimating occupancy *probabilities* and even outperforms the other methods on small sample sizes. This result is of concern to us, because it suggests that Maxent directly offers a useful alternative for modelling presence–absence data, which may prompt widespread adoption of this use of Maxent. In this paper, we explore why this would be a mistake.

2. We draw on the theory underlying how the Maxent model operates and on simulations to discover: (i) why Maxent appears to fare as well as it does in their evaluation and (ii) why the best-suited presence–absence method for data analysis (the generating model; a GLM) does not perform as well as we would expect.

3. We demonstrate that (i) the good performance observed for Maxent is largely a coincidence; the simulated species match well the arbitrary default parameter that Maxent applies to map its relative output into a 0–1 scale, but errors are much larger for other species we simulate; (ii) the performance of the GLM is poorer than expected because Thibaud *et al.* do not use model selection and fit a model that is too complex for the amount of data available.

4. Maxent is a presence-background method and only provides estimates of *relative* suitability regardless of how the background sample is specified. When presence–absence data are available, one can transform Maxent's relative estimates into estimates of occupancy probability, and we provide methods to do so. However, this requires the user to post-process Maxent's output. Proper PA methods such as GLMs can perform well under small sample sizes, provided care is taken during modelling to avoid overfitting. We demonstrate an effective method using regularisation with the R package *glmnet*.

Key-words: *glmnet*, relative suitability, overfitting, prevalence, presence-only, presence-background, Maxent, probabilities, presence–absence, species distribution model

Introduction

Thibaud *et al.* (2014) present a useful framework for simulating species and evaluating the relative effects of factors affecting the predictions from species distribution models ('SDMs' hereafter). Here, we address a minor yet important component of the demonstration of their approach. Thibaud *et al.* simulate numerous presence–absence ('PA') data sets and use four methods for fitting species distribution models to the data. They assess their results in terms of root mean square errors (RMSEs) as a measure of how well occupancy *probabilities* are estimated by each method. The program Maxent (Phillips, Anderson & Schapire 2006; Phillips & Dudík 2008) is one of the methods used. However, as we explain next, Maxent was

not developed to analyse PA data. Indeed, Thibaud *et al.* clearly acknowledge that fitting PA data with Maxent is a 'non-standard manner' of use for this tool. Their results for small sample sizes show Maxent outperforming the other methods. Because these results and their subsequent discussion suggest that Maxent as used by Thibaud *et al.* actually performs well in modelling PA data, we expect they may prompt widespread adoption of this use of the already popular program Maxent. Here, we explore why this would be a mistake, both from the viewpoint of how the Maxent model operates and by exploring the simulations to discover why Maxent appeared to fare as well as it did in their evaluation.

The Maxent model

The Maxent that we refer to is a software package developed for modelling species distributions given presence-only species

*Correspondence author. E-mail: gguillera@unimelb.edu.au
All authors contributed equally to the paper.

records and a 'background' sample of environments in the region of interest (Phillips, Anderson & Schapire 2006; Phillips & Dudík 2008). We hereafter refer to this type of data as 'presence-background' (PB) data. Maxent applies the maximum-entropy or 'maxent' principle (Jaynes 1957) for fitting the model so that the estimated species distribution deviates from a uniform distribution as minimally as required to explain the observations. The details of the software and the principles underlying this approach have been extensively explained in the literature (e.g., references above and Elith *et al.* 2011; Merow, Smith & Silander 2013), so we only mention here the aspects that are key for our discussion. Recent work has shown how Maxent relates to other PB modelling methods such as point process and PB generalized linear models (GLMs) (Warton & Shepherd 2010; Fithian & Hastie 2013; Renner & Warton 2013).

Although the algorithm within Maxent works in geographic space (Phillips, Anderson & Schapire 2006), its behaviour can be explained intuitively from a statistical point of view by looking at the environmental domain (Elith *et al.* 2011); we follow this equivalent interpretation in our explanations hereafter. Maxent uses PB data to obtain a picture of environmental characteristics at presence sites and at background locations (which are a random or regular sample of the landscape, or could be targeted to match known biases in the sampling process). More formally, Maxent estimates the ratio f_1/f, where f_1 is the probability distribution describing characteristics of sites at which the species occurs and f is the probability distribution describing characteristics of the environment as a whole, that is both at sites where the species is present ($y = 1$) and absent, ($y = 0$). This ratio is proportional to the probability of presence given the environmental covariates z, $\Pr(y = 1|z) = \psi(z)$, but the scaling factor (the prevalence of the species over the landscape, $\Pr(y = 1) = \bar{\psi}$) cannot be identified from PB data only (Elith *et al.* 2011; Hastie & Fithian 2013; Phillips & Elith 2013). The '*raw output*' of Maxent is the estimate of f_1/f scaled to sum to 1 over the fitted background and represents the probability distribution over sites x, that is the probability, given the species is present, that it is found at site x: $\Pr(x|y = 1)$ (Phillips, Anderson & Schapire 2006). Hence, as currently implemented, that raw output represents $\psi(z)/(n \cdot \bar{\psi})$ where n is the number of points in the background sample (see Appendix S1a for mathematical details). Being proportional to occupancy probabilities by a factor that is not identifiable without external data, the raw output only represents *relative* suitability.

In order to estimate the ratio f_1/f, Maxent fits an exponential model so that $\log(f_1/f) = \eta(z)$, where $\eta(z)$ is a linear term of a set of features. Features represent an expanded set of transformations of the covariates and in Maxent they can belong to six classes: linear, product, quadratic, hinge, threshold and categorical (Phillips & Dudík 2008). Features provide great flexibility to fit complex environmental relationships, and Maxent uses regularization to control the trade-off between model fit and model complexity in order to avoid overfitting.

Rather than working with Maxent's 'raw output', it is common for users to look at the so-called *logistic output*, which Maxent provides by default. This output maps the raw values into a scale of relative suitability ranging between 0 and 1, using a logistic transformation based on a user-specified parameter 'tau' (Elith *et al.* 2011). This parameter is defined as the prevalence for sites with 'average' environmental conditions under f_1 (see details in Appendix S3 of Elith *et al.* 2011) and is set by default to an arbitrary value ($\tau = 0.5$). Here, we again emphasize that it is not possible to estimate occurrence probabilities from a PB data set and that hence the default logistic output does *not* represent probabilities either.

Thibaud *et al.*'s modelling

Thibaud *et al.* provide comprehensive and well-commented code that allows us to understand and replicate their simulations, model fitting and evaluation. Here, we first briefly summarize the relevant points of their simulation and then describe in detail their use of Maxent.

OVERVIEW

1. *Species simulation* – Species distributions are simulated based on survey data from 10 real species, which are fitted with a generalized linear model (GLM; family = binomial, link = probit) using five environmental covariates with linear and quadratic terms. The modelled distributions, projected to rasters containing 1·127 million grid cells, are assumed to represent the 'true' probabilities of occurrence of the simulated species. Each of these probabilistic 'truths' are then realized 10 times to generate 'true' scenarios of species presence and absence (PA). Thibaud *et al.* include simulations where spatial autocorrelation (SAC) is injected into the distributions, but we do not address that here because the simplest case is sufficient for our purposes.

2. *Sampling* (*training data*) – For each of the PA 'truths', the sampling of $n = 100$ or 500 sites is repeated 5 times, with sampling either randomly placed or biased towards roads. Both the presence/absence status of the species and the values of the covariates are recorded at the sampled sites. We again focus on the simplest random case, as it is sufficient for our purposes.

3. *Test data* – Each training data set is accompanied by a test data set of size 5000 sites. Both the values corresponding to the probabilistic and PA 'truths' are stored.

4. *Model fitting* – Four methods are used to fit models to the data. These include Maxent and three true PA methods: GLMs (probit regression), generalized additive models (GAMs, with probit link) and Random Forests (RF). For references see Thibaud *et al.* Here, we focus on their use of Maxent (detailed below) and compare it to a GLM. Since the generating model is a GLM, and hence, the GLM should be the best-suited PA model for fitting the samples, we do not pursue the results for GAM and RF further.

5 *Performance evaluation* – Thibaud *et al.* assert that all the methods they use predict probabilities of occurrence and consequently evaluate their results in terms of root mean squared error (RMSE), comparing true probabilities with predicted values. They also provide code to compute the area under the

receiver operating characteristic curve (AUC) and point-biserial correlation (COR) by comparing predicted values with the PA realisations of truth. Following Thibaud *et al.*, we focus on RMSE. RMSE reflects the calibration of the predicted values (i.e. whether the predicted values are well-scaled estimates of the true probability of occurrence), while in contrast, AUC and COR only measure performance in terms of discrimination ability (Lawson *et al.* 2014).

USE OF MAXENT

Thibaud *et al.* specify that they run Maxent with default settings and that they provide the absence data as a background sample. Default settings imply that they use Maxent's logistic output (i.e. *output format* = logistic) produced with the arbitrary default value for *tau* (i.e. *default prevalence* = 0·5). A second important implication is that Maxent adds presence records to the background sample because by default, the *add samples to background* option is invoked, whereby any presence sample with a combination of environmental values not present in the existing background sample is added to it. Therefore, in practice, we can think of the background used by Thibaud *et al.* as the full set of sites in the PA data set, rather than only the absence data as stated by the authors (technically it might contain less than the full set of sites if there are presence and absence sites that happen to be identical environmentally but this is unlikely given that their covariates are continuous).

We note that using the set of sites sampled (i.e. the PA data set) as background does *not* bring Maxent any closer to estimating occupancy probabilities. Maxent still just compares the characteristics of the sites where the species is detected to those in the background set, to provide estimates of *relative* suitability. In fact, assuming that sampling is unbiased, using the PA data set as background is very similar to using an independent random sample of the environment: we provide Maxent data to characterize the environment in general, that is the characteristics at sites regardless of whether the species is present or not (i.e. the distribution f). There is a distinction though that makes this background specification more attractive when working with PA data than providing an independent sample of sites (as done when only PB data are available). The background now represents exactly the places that have been surveyed, so this improves the estimation of the ratio f_1/f by compensating for potential sampling biases, such as tendency to sample close to roads. In such cases, it is similar to the idea of target group background (Phillips *et al.* 2009) for the analysis of PB data, but here the 'target group' is only the species in question. Even when sampling is truly random, some performance improvement can be expected, as the set of sites selected for sampling may over/under-represent some environments just by chance, especially when the sample size is not large (see an illustration of this effect in Appendix S2). Given these advantages, we have used this background specification in what follows. However, we emphasize again that the PA data are being used to create a PB data set, and Maxent is still operating as a PB method.

Our concerns

We understand that the choice of modelling methods was not the main priority of Thibaud *et al.*, however, one of their results was striking: that their use of Maxent outperformed the other methods (had lower RMSE) on small sample sizes ($n = 100$). This result is of concern to us, because it suggests that Maxent might directly be a useful alternative for modelling presence–absence data. However, we do not believe this is a good idea because Maxent does *not* estimate probabilities and hence falls one step short of what actual PA methods can achieve (i.e. methods that model PA records as such and hence directly provide estimation of probabilities). Here, we explore (i) why Maxent's logistic output appears well scaled in Thibaud *et al.*'s simulation even though we assert that Maxent only estimates *relative* suitability; (ii) why Maxent performed better than a GLM on small sample sizes. We use the code and data supplied by Thibaud *et al.* to address these two points. As a baseline, we repeat their sample sizes and testing regime and only deviate from that where necessary for clarifying issues.

WHY ARE MAXENT'S OUTPUTS APPARENTLY WELL SCALED FOR THESE SIMULATED SPECIES?

As stated above, Maxent is a PB method and does not provide estimates of occupancy probability. It only estimates *relative* suitability. However, the results of Thibaud *et al.* could at a first glance suggest that Maxent is able to estimate probabilities directly: in some simulations, its RMSE is lower than that obtained with PA methods such as GLMs. Here, we show that this behaviour is just a coincidence due to a good agreement between the prevalence of the particular simulated species and the arbitrary parameter value ($\tau = 0·5$) that Maxent uses by default to produce the logistic output. To do this, we simulate new species that have lower and higher prevalence values by scaling the 10 original species, subtracting or adding a constant ($-0·5/+1$) to the original species values on the probit scale. We use Thibaud *et al.*'s code for creating data sets and fitting models. We follow them in creating samples of 100 or 500 sites, but increase the replicates: where they used 10 realisations and 5 samples, we use 20 realisations and 25 samples – simply to clarify the trends. Our results in Fig. 1 compare GLMs (warm colours) and Maxent models (blues) for Thibaud *et al.*'s modelling and ours, with each panel showing data for a species (rows) and its rescaled variants (centre and right columns). First consider the top two boxplots in each panel (Thibaud *et al.*'s methods, red and darker blue). These show high RMSE values for Maxent among several of the scaled species, which is a reflection that Maxent is unable to estimate probabilities. While for some of our new simulated species Maxent's relative output might by chance be well calibrated (even better calibrated than for Thibaud *et al.*'s species) with Maxent's default settings, for others, RMSEs are large. For all species, the true PA method (a GLM, red) is relatively unaffected by scaling (i.e. in one row, RMSE's for the GLMs remain similar across the columns). Results for 500 samples are consistent (Appendix S4).
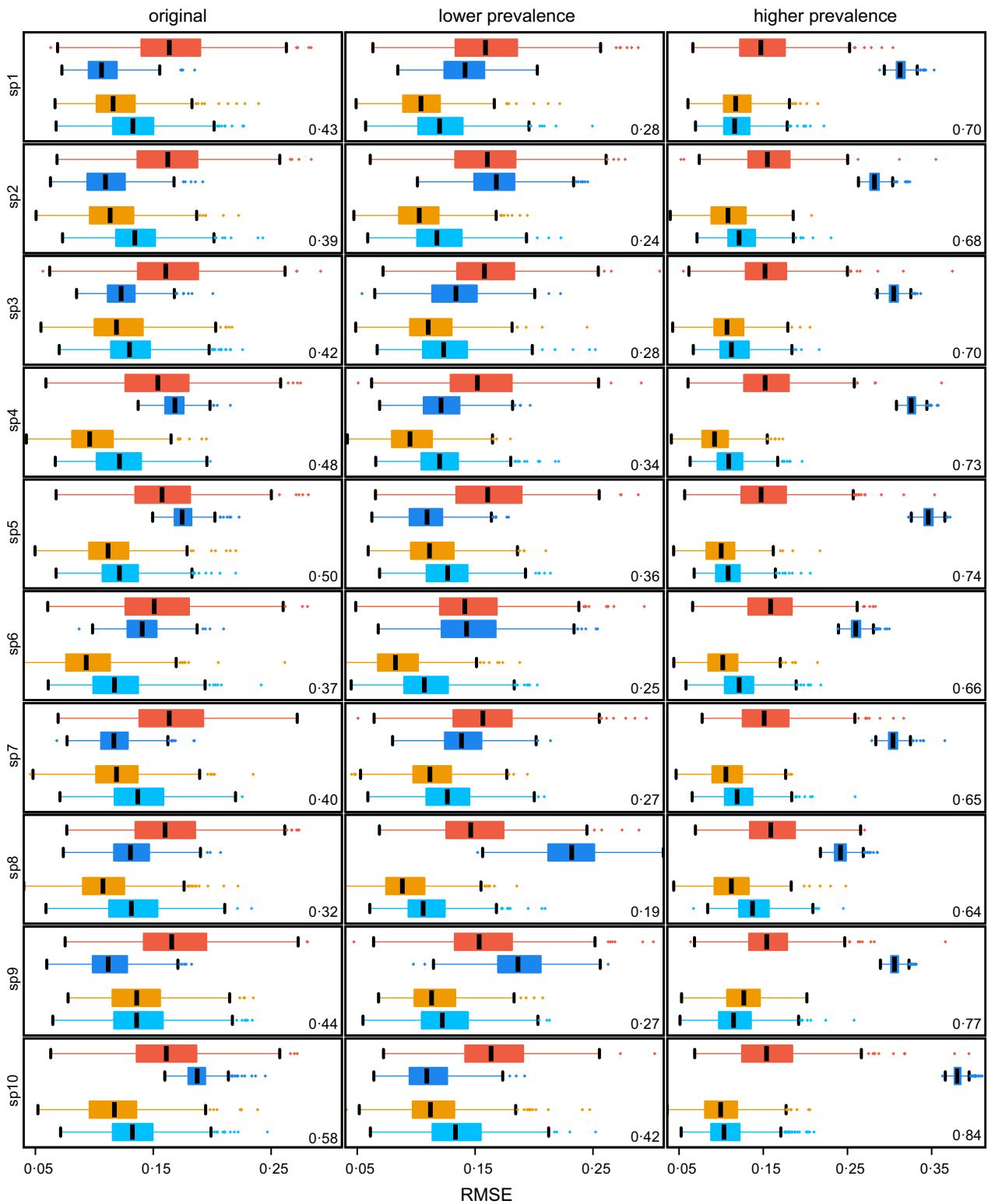
Fig. 1. Root mean square error (RMSE, *x*-axis) for the simulated species (sample size = 100, boxplots summarize 500 repeats as detailed in text). Each panel presents (from the top) boxplots for Thibaud *et al*.'s GLM (red) and Maxent (deep blue), then our regularized GLM (orange) and post-processed Maxent (light blue). Prevalence for the simulated truth is bottom right. The left column contains species simulated by Thibaud *et al*., and to their right two rescaled variants. Species are in rows. Note that the right column's x axes have a different range to the other two columns.

A different matter is that, given that PA data are available, one could post-process the *relative* output that Maxent produces in order to obtain estimates of occupancy probabilities.

One approach to do this is to obtain an estimation of prevalence from the PA data set available for model fitting ($\hat{\bar{\psi}}$ = number of detections/total number of sampled sites) and

then use this and the total number of points in the background (n) to scale Maxent's raw output as follows:

$$raw = \frac{\psi(z)}{(n \cdot \bar{\hat{\psi}})} \rightarrow \psi(z) = raw \cdot n \cdot \bar{\hat{\psi}}.$$

Our results (Fig. 1 for sample size of 100 and Appendix S4 for 500) most commonly show reductions in RMSE when this approach is applied and Maxent's relative output is appropriately mapped into probability estimates (compare our results, light blue boxplots, with the darker blue ones).

An alternative way of conducting this calibration involves transforming the raw output through a *complementary log–log* ('cloglog') transformation (S. Phillips, pers. comm.), so that $\psi_i = 1 - \exp\left(-\exp(\eta(z_i) + c)\right)$. The value of the parameter c can be selected doing a binary search so that the mean of the transformed values approximates the prevalence estimated from the PA data: $\sum_i \{1 - \exp(-\exp(\eta(z_i) + c))\} = \bar{\hat{\psi}}$, or alternatively as the intercept of a fitted GLM with a cloglog link function using $\eta(z_i)$ as an offset (see code for both options in Appendix S5). The cloglog transformation links with the theory of point process models (Baddeley *et al.* 2010; Aarts, Fieberg & Matthiopoulos 2012; Fithian & Hastie 2013) and is more appropriate when records represent point data (i.e. records represent specific point locations rather than referring to sites of a given area). The transformation links estimation in terms of density to the probability that at least one individual occupies a site. Direct scaling of the raw output is coherent with records that represent sites, as in these simulations, where simulated probabilities and binary data are generated at the site level. Finally, one could also adjust the logistic output, tuning the parameter *tau* defined in our description of Maxent (note that the *defaultprevalence* argument in Maxent refers to *tau* and is therefore not equal to the prevalence across the landscape, thus the observed prevalence in the data set cannot be directly inserted to the *defaultprevalence* argument in Maxent; we provide a correct method for performing this tuning in Appendix S5).

WHY DOES MAXENT APPEAR TO DO BETTER THAN A GLM?

Maxent appeared to do well on small sample sizes in relation to GLMs, so we also need to understand the performance of the GLMs. Why could not a GLM fit the samples well, particularly since the generating model was a GLM? The 10 original simulated species varied in prevalence from 0·32 to 0·58, meaning that small samples (100 sites) would on average contain 32–58 presence records. Specifically, Thibaud *et al.* noted 20–80% of presence samples in training data over their simulations. Since the simulated species were based on linear and quadratic terms fitted to five covariates, 11 parameters (including the intercept) have to be estimated for the full model. For GLMs, Thibaud *et al.* fitted full models with no model selection. In small data sets, it is likely that many of the coefficients were poorly estimated, since rules of thumb indicate 10 observations in the least frequent class (presence or absence) are needed to accurately estimate each parameter (Harrell, Lee &

Mark 1996; Steyerberg *et al.* 2000; Harrell 2001, pp. 60–61). This implies that in samples with only 20–80 observations of presence or absence the full model will have been poorly estimated, and some sort of regularisation or model selection is required to avoid overfitting.

Maxent uses L1 regularisation for model parameterization and selection (Elith *et al.* 2011). This is a modern model selection method that penalizes (shrinks) model parameters to values that balance model fit and complexity, so the model fits known data while generalising well (Hastie, Tibshirani & Friedman 2009). Here, we test a comparable method for model selection in regression, to explore whether the GLM performance improves. We use the R package *glmnet* (v.1.9–5), which fits GLMs via penalized maximum likelihood and allows a range of penalty types, from the lasso (L1-regularisation, *alpha* = 1) to ridge regression (L2-regularisation, *alpha* = 0), with the elastic-net penalty (Zou & Hastie 2005; Hastie, Tibshirani & Friedman 2009) covering all the intermediate cases. Following some experimentation on the data, we use the ridge penalty with alpha = 0. We also use prevalence-stratified 10-fold cross-validation to select the shrinkage parameter (*lambda*) that, for each fitted model, gives lowest prediction error. For details of the implementation and code, including methods for making orthogonal linear and quadratic terms, and impacts of our choices for model settings vs defaults, see Appendices S3 and S5. The probit link is not supported in this package, so we use a logistic regression instead. Tests confirmed that the impact of probit vs logistic regression on the GLMs as fitted by Thibaud *et al.* was minimal (results not shown).

Our results (Fig. 1, orange boxplots) show that regularized GLMs have comparable performance to rescaled Maxent, with a small but consistent improvement in median RMSE performance for GLMs over Maxent across all species. That is, once a GLM is implemented in ways more comparable to Maxent, the supposed advantage of Maxent – even in small samples – disappears. The results for sample sizes of 500 (Appendix S4) show a larger advantage for GLMs, even though we controlled Maxent to only fit linear, quadratic and hinge features (likely to give best performance since the generating species only had linear and quadratic terms).

Further clarifications

Some may wonder whether one could make program Maxent estimate probabilities directly by using only absence records as background, which would require toggling the option *addsamplestobackground* to 'false' (using flag '-d'). The answer is no. When used in this manner, Maxent would still only estimate a relative quantity: its raw output would provide an estimation of f_1/f_0 where f_0 is the probability distribution describing the characteristics of sites where the species is absent; this ratio is proportional to the odds of species occupancy rather than to occupancy probability (see details in Appendix S1b). Furthermore, not being the usage for which the tool was developed, Maxent is numerically unstable when used in this manner (see example plot in

Appendix S1b). In summary, regardless of how we specify the background in program Maxent, the tool does not become a PA method; Maxent cannot exploit all the information available in the PA data set but rather treats it as PB data, requiring post-processing to obtain estimates of occurrence probability. We note as an aside that the fact that the program Maxent implements a PB method does not imply that a tool cannot be developed to model PA data directly based on maximum-entropy principles.

Conclusion

PA methods provide estimates of absolute occupancy probabilities. The program Maxent was developed to analyse PB data and, as such, it only provides estimates of *relative* suitability, that is Maxent does *not* estimate occupancy probabilities (unlike stated by Thibaud *et al.*). When PA data are available, one can transform Maxent's relative estimates into estimates of occupancy probability. However, this requires the user to post-process Maxent's output, for instance using one of the methods we demonstrate. In our simulations, we found Maxent to perform well when this was done but we caution that the approach could potentially be open to unknown problems given that Maxent was not programmed to be applied this way. Furthermore, using Maxent for PA data seems circuitous when proper PA methods exist. We have shown that, unlike implied by some of Thibaud *et al.*'s results, proper PA methods such as GLMs can perform well under small sample sizes, provided care is taken with modelling methods to avoid overfitting. Our application of regularisation (with package *glmnet*) in this context will be of interest to those wanting to explore new methods for fitting regression models.

Data accessibility

Species data: simulated, using code of Thibaud *et al.* (2014) and variations described in text. R scripts: uploaded as online supporting information.

References

Aarts, G., Fieberg, J. & Matthiopoulos, J. (2012) Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods in Ecology and Evolution*, **3**, 177–187.

Baddeley, A., Berman, M., Fisher, N.I., Hardegen, A., Milne, R.K., Schuhmacher, D., Shah, R. & Turner, R. (2010) Spatial logistic regression and change-of-support in Poisson point processes. *Electronic Journal of Statistics*, **4**, 1151–1201.

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

Fithian, W. & Hastie, T. (2013) Finite-sample equivalence in statistical models for presence-only data. *Annals of Applied Statistics*, **7**, 1917–1939.

Harrell, F.E. (2001) *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer, New York.

Harrell, F.E., Lee, K.L. & Mark, D.B. (1996) Multivariate prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, **15**, 361–387.

Hastie, T. & Fithian, W. (2013) Inference from presence-only data; the ongoing controversy. *Ecography*, **36**, 864–867.

Hastie, T., Tibshirani, R. & Friedman, J.H. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn. Springer-Verlag, New York.

Jaynes, E.T. (1957) Information theory and statistical mechanics. *Physical Review*, **106**, 620–630.

Lawson, C.R., Hodgson, J.A., Wilson, R.J. & Richards, S.A. (2014) Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution*, **5**, 54–64.

Merow, C., Smith, M.J. & Silander, J.A. (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, **36**, 1058–1069.

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

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

Phillips, S.J. & Elith, J. (2013) On estimating probability of presence from use-availability or presence-background data. *Ecology*, **94**, 1409–1419.

Phillips, S.J., Dudik, 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.

Renner, I.W. & Warton, D.I. (2013) Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, **69**, 274–281.

Steyerberg, E.W., Eijkemans, M.J.C., Harrell, F.E. & Habbema, J.D.F. (2000) Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Statistics in Medicine*, **19**, 1059–1079.

Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A.C. & Guisan, A. (2014) Measuring the relative effect of factors affecting species distribution model predictions. *Methods in Ecology and Evolution*, **5**, 947–955.

Warton, D.I. & Shepherd, L.C. (2010) Poisson point process models solve the "Pseudo-Absence Problem" for presence-only data in ecology. *Annals of Applied Statistics*, **4**, 1383–1402.

Zou, H. & Hastie, T. (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society B*, **67**, 301–320.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Interpretation of Maxent's raw output.

Appendix S2. Implications of using the presence-absence sample as background.

Appendix S3. GLMnet analysis.

Appendix S4. Results for larger sample size ($n = 500$).

Appendix S5. Computer code.