Experimenting with Modelling via a Virtual Laboratory: Evaluating pseudo-absence strategies to refine a species distribution model

S. Low-Choy $^{\rm a}$ and C. Huijbers $^{\rm b}$

^aGriffith Social & Behavioural Research College, Griffith University, Australia ^bBiodiversity and Climate Change Virtual Laboratory Email: S.Low-Choy@Griffith.edu.au

Abstract: Virtual laboratories (VLs) are fast becoming realities in many fields of enquiry. For instance, the Biodiversity and Climate Change Virtual Laboratory (BCCVL) provides users with a high-performance computational platform to enable more efficient investigation of biological systems. This kind of VL is more than a mere portal to dispersed data sources and a diverse range of modelling options; it also reduces computational overheads and tedium required to implement models. In this way, a VL allows users to explore each model to more fully apply scientific method in model development. Here we explore how the BCCVL can be used to support an iterative process of investigating and refining models, through experimentation.

The BCCVL supports many kinds of modelling for biodiversity, measured by species presence, traits or aggregate measures such as species richness. Here we narrow our focus to consider species distribution modelling (SDM), and in particular, the source of absence data. Absences in SDM provide a useful case study for exploring models in VLs, as there are many potential settings, known to substantially impact SDM results. When absence of the species has not been explicitly recorded, several strategies are available to impute 'pseudo-absences'. New users may inadvertently specify pseudo-absences in a way that leads to issues such as 'naughty noughts' or pseudo-replication. It is possible to identify those issues during SDM, and this process can be accelerated through a VL. Additionally after initial exploration in a VL, it is easy to export data for analysis into a statistical package, such as R, and continue to refine SDMs.

Here we show how the SDM for the Golden bowerbird is sensitive to the strategy for generating pseudoabsences, as defined by settings that can be altered within the BCCVL. A sequence of well-defined experiments gradually helps refine the options defining this strategy. We begin with the study region, which implicitly delimits search effort, and potentially defined by: the continent, a bioregion or a convex hull delimited by the farthest occurrences. At the same time BCCVL makes it easy to compare SDM algorithms. We consider regression (GLM), tree (CTA) and machine learning (MaxEnt) algorithms. Next we undertake separate experiments to further refine selection of pseudo-absences. The sampling strategy may be: completely random; constrained by a disc centred at occurrences; or defined by a Surface Range Envelope, comprising locations that fall outside the usual range of predictors evaluated at occurrences. In comparison to the number of occurrences, the intensity of pseudo-absences may be set to be equal or any other ratio. We export model results for out-of-VL analysis, and apply recursive partitioning trees in R to investigate naughty noughts.

The Golden bowerbird is similar to many specialist species in Australia: generating pseudo-absences across the continent gave a large contrast between occurrence and absence, as evidenced by the distribution of predicted probability of presences. Constraining pseudo-absences to a bioregion, we were able to choose an SDM algorithm that permitted examination of gradients from absence to presence, whilst retaining high accuracy. Further experimentation assessed sensitivity to the sampling strategy of pseudo-absences to R, tree modelling identified uninhabited climates (with high mean temperature of the warmest quarter). When omitted, the estimates of climate effects on this species' presence were greatly sharpened. This demonstrates how a VL may be used to refine modelling, evaluating sensitivity to settings via performance measures relevant at each stage. In this case the choice of pseudo-absence strategy to support SDM for the Golden bowerbird might have been discarded using a 'one-off' modelling approach that focussed on a single indicator.

Keywords: Iterative modelling, computer experiments, naughty noughts, probability of presence

1 INTRODUCTION

Many Virtual Laboratories (VLs) provide a platform for 'virtual' experiments, as a substitute for physical experiments. Here we consider a VL that streamlines and facilitates experimentation with modelling. The Biodiversity and Climate Change Virtual Laboratory (BCCVL, Hallgren et al., 2016) includes a platform for species distribution modelling (SDM). In landscape ecology, SDMs predict geographic distribution of a species, current or future, by modelling the relationship between species' occurrence with predictors, of environment and climate (Franklin, 2010). Species distribution can inform a wide range of policy and decisions, including regional planning, conservation efforts, managing weeds and preparedness for global threats such as climate change or exotic pest incursions (Franklin, 2010; Belbin and Williams, 2016).

The BCCVL guides the user in designing an experiment via a simple workflow, starting with specification of inputs on species occurrence and absence, environment and climate. A key advantage is easy access to an up-to-date integrated Geographic Information System (GIS) comprising over 4000 environmental and climate predictors at multiple resolutions and scales. The VL allows the user to upload their own data on species presence/absence or to source occurrences from online data warehouses, such as the Australian Living Atlas (Belbin and Williams, 2016), which contains over 70 million records on over 150 thousand native or invasive species. SDM outputs comprise a predicted map of species distribution and nearly two dozen predictive performance measures. The BCCVL streamlines comparison of 17 different SDM algorithms.

Altogether, this leads to an extremely broad choice of inputs, outputs, algorithms and their respective settings. In this way the BCCVL empowers end-users, but may also overwhelm them with the rich range of options. For users new to SDM, this VL makes it easier to get started, with default settings for most SDM algorithms. For simplicity, it would be tempting to adhere to 'default' settings or heuristics (e.g. Barbet-Massin et al., 2012). However, it has been established that the settings of SDM algorithms do matter (Wisz and Guisan, 2009; Lobo et al., 2010; Merow and Silander, 2014), and need to be tailored to the situation. BCCVL provides a responsive modelling environment that allows the user to exercise their expertise, by experimenting with SDM settings.

Here we investigate how a set of modelling options need not be investigated all at once, but can manageably be investigated via a sequence of experiments, to methodically refine an SDM (Wisz and Guisan, 2009). This reflects a modern view of statistical analysis as an iterative cycle of modelling (Greenland et al., 2016).

In particular, the strategy for generating pseudo-absences provides a useful basis for demonstrating the use of a VL for experimenting with models. Poor choice of pseudo-absence strategy may adversely affect SDM, producing naughty noughts or pseudo-replication. Choosing a study region that is too large may include 'naughty noughts' in regions uninhabitable by the species, e.g. deserts or waterbodies (Pirathiban et al., 2015). The resulting high contrast between inhabitable and uninhabitable zones has high predictive performance, but does not examine the gradient of inhabitability within the species' envelope (Austin and Meyers, 1996). Pseudo-replication is a risk when studies of different kinds are combined (Witman et al., 2015). Consider a behavioural study of a highly mobile species. When combined with a cross-sectional study, the repeated observations may over-emphasize those individuals, unless aggregated or re-weighted. For this reason pseudo-replication can arise from inadequate checking after uploading from a data warehouse. Pseudo-replication is also related to high spatial auto-correlation, e.g. when too many pseudo-absences are generated to 'fill' the region, thus violating conditional independence assumptions that occurrence at a site is not solely explained by its environment but also by its neighbours (Mateo et al., 2010).

The BCCVL provides a platform that supports new or experienced users in experimenting with models, reducing the need to use 'default' settings or a 'one-size-fits-all' approach to modelling. To illustrate this, we outline the way in which BCCVL allows users to specify and experiment with the sampling strategy used to generate pseudo-absences (Section 2), introduce a case study on SDM for the Golden bowerbird (Section 3), and then apply the experiment to the case study (Section 4) before concluding (Section 5).

2 DESIGNING EXPERIMENTS: PSEUDO-ABSENCES IN SDM

We conducted a sequence of experiments to refine the pseudo-absence strategy.

Phase 1. Ecological scale Fundamentally, SDM is influenced by the choice of study region. Essentially, if there is a high chance of detecting the species using a search method, but the species is undetected, then we may conclude the species is plausibly absent from the area searched. Hence it is important to delineate where search effort has been adequate, and where pseudo-absences may sensibly be generated. So the first modelling decision concerns ecological scale of search effort: e.g. continental, bioregion, landscape, or watershed.

Phase 2. SDM algorithm Pseudo-absence sampling strategies affect inferences in GLM (Wisz and Guisan, 2009) and other SDM algorithms (Mateo et al., 2010; Barbet-Massin et al., 2012). Here we focus on 3 of 17 algorithms available in BCCVL (Hallgren et al., 2016). The most popular algorithm, Maximum Entropy (MaxEnt), constructs 'feature functions' that describe the likely environment where the species occurs, compared to the whole environment. A regression or Generalized Linear Model (GLM) trades off the effects of environmental predictors: each 'adding' to or 'subtracting' from a score of habitat suitability. Classification tree analysis (CTA) identifies one 'best' tree that divides sites into environmental profiles. A sequence of branching rules split sites into profiles using a threshold on one environmental predictor at a time (Franklin, 2010). GLM is better at representing how gradual changes in environmental gradients affect the likelihood of presence *vs* absence, whereas CTA can accommodate abrupt changes. MaxEnt describes the likelihood of environmental gradients for occurrence sites *vs* the whole study region.

Phase 3. Spatial extent and intensity of pseudo-absences The boundary may be defined in geographic space, as a perimeter, or in environmental space, as an envelope. BCCVL provides four options. The simple convex hull sets a hard boundary that joins the outermost occurrence sites (Barbet-Massin et al., 2012). This can lead to abrupt exclusion of sites that are immediately adjacent to known occurrences. Such discontinuities can be reduced through an extended convex hull, which expands the outer perimeter by a fixed distance. The second option is to select a pre-defined geographical region, such as a local government area or bioregion. Alternatively, users can use the entire extent of the input data or elect to draw their own polygon on the map.

The boundary of the study region may be modified via expansions or exclusions. The convex hull or geographic regions can be expanded by the same distance in every direction, by defining a disc. This may be unrealistic in some situations, e.g. in the Wet Tropics. Alternatively, inhabitable regions can be excluded using a Surface Range Envelope (SRE, Thuiller et al., 2009), which like BIOCLIM, sets an inner boundary at the outer percentiles (e.g. 2.5% and 97.5%) of predictor values corresponding to presences. Given the study region, the BCCVL user can generate pseudo-absences randomly in the modified region.

Bigger is not necessarily better when specifying the number of pseudo-absences. Too many can be inefficient or induce pseudo-replication, especially with high resolution data (Mateo et al., 2010). BCCVL allows specification of the intensity of pseudo-absences, relative to the number of occurrences. Popular choices are an equal ratio, or 10,000 pseudo-absences for regression (Barbet-Massin et al., 2012).

Phase 4. Investigating naughty noughts At this stage pseudo absences have been sufficiently refined, as a basis for evaluating naughty noughts. Uninhabitable regions can be identified (and then excluded) using a kind of CTA, where misclassification of negatives is penalised more than for positives (as detailed in Pirathiban et al., 2015). This can be achieved in R (R Core Team, 2017), by exporting BCCVL data comprising species presence/absence and predictors across the study region, then fitting CTA via recursive partitioning, with a penalty on false negatives using rpart (see code in Pirathiban et al., 2015).

2.1 Outputs: model diagnostics

For simplicity, we focus on three main diagnostics. The primary diagnostic is a map showing the predicted geographic distribution (e.g. Fig. 1).

Another key diagnostic shows the predictions: the distribution of the predicted probability of presence (PPP) for absences vs occurrences (Pirathiban et al., 2015). As a smoothed version of a histogram, these plots show the relative frequency (y-axis) of each possible predicted probability of presence (PPP) (x-axis). Ideally PPP should be low for absences and high for known occurrences. The quality of predictions is reflected by the width or range of PPP for absences and occurrences. For instance a 'tight' or 'narrow' distribution of predictions for absences covers a small range of values (on the x-axis), which means that, for the sites corresponding to absences in the input data, the predicted probability that the species is present is relatively consistent. When predictions have a tighter range for absences compared to presences, this clearly indicates that the SDM is better at predicting absence, e.g. Fig. 2 (CTA, right).

The predictions plot relies on an ability to interpret a histogram, a fundamental plot useful for all statistical analyses. It provides richer information than a single summary statistic, such as the popular AUC. The AUC refers to the ROC, which graphs the True Positive Rate (TPR) against the False Positive Rate (FPR), evaluated at different values of the predicted probability of presence (PPP). The predictions plot is also more relevant than the AUC, which does not evaluate how well occurrences are discriminated from absences (Lobo et al., 2010). Instead, discriminative ability is more visible with the prediction plot. Here FNR, the proportion of

occurrences with low PPP, equates to the area of the blue curve overlapping the red. Similarly, FPR, the proportion of absences with high PPP, is the area of the red curve overlapping the blue. The PPP curve accentuates that FPR and FNR compete (one increases as the other decreases).

For GLM, we use 95% confidence intervals, based on rescaled predictors (subtracting the mean, and dividing by the standard deviation) to assess contributions of each environmental predictor (Greenland et al., 2016).

3 CASE STUDY

To illustrate use of the VL for experimenting with models, we investigate pseudo-absences in SDM for the Golden bowerbird, which is Australia's smallest bowerbird. A total of 1136 occurrence records were imported from the Atlas of Living Australia into the BCCVL. These reduced to 281 'cleaned' records, after removing outliers, duplicates, records before 1950 or with an uncertainty > 1km. We adopted the BIOCLIM climate variables (Booth et al., 2014) chosen by Hilbert et al. (2004): annual mean temperature (B01), minimum temperature of the coldest month (B06), mean temperature of the warmest (B10) and coldest (B11) quarter, annual precipitation (B12), precipitation of the wettest (B16) and driest (B17) quarter. We used 30 arcsec (\sim 1km) resolution climate data available in the BCCVL, generated from aggregating monthly data from the Australia Water Availability Project over the years 1976–2005 (VanDerWal, 2012).

Conservation is a priority for the Golden bowerbird, as one of twelve bird species that are endemic to the Wet Tropics in north Queensland, Australia. Restricted to high altitude rainforest areas, it is increasingly scattered due to forest clearing and global warming. Like many highland species, Golden bowerbirds are intolerant of the high temperatures of the lowland tropics (Hilbert et al., 2004). Thus, we expect temperature of the warmest quarter (B10) to be crucial in its SDM. Moreover, climate change is likely to shift the bird's suitable habitat to higher altitudes, forcing it to move upwards and ultimately face extinction (Hilbert et al., 2004).

4 RESULTS

4.1 Experiment with ecological scale and study region

We start by comparing SDMs at the continental scale, throughout Australia, and the bioregional scale, within the Wet Tropics bioregion. Initially, 2810 pseudo-absences were generated, at a ratio of 10:1 occurrences.

Predictions at the continental scale clearly highlight the Wet Tropics bioregion and a small area further south as prime habitat for the Golden bowerbird (Fig. 1, top). Focussing on the Wet Tropics gave greater finesse in PPP, which varies more within the region: the species is more likely away from the coast (Fig. 1, bottom). For the next phase of experimentation, we chose bioregional scale, as it reveals gradients in habitat quality within its envelope, although continental scale helps locate the envelope.

At continental scale, CTA did not identify a 'spot' of occurrence in the far north-east (Oyala Thumotang National Park). MaxEnt and CTA predict the species is more likely on the coast (near Mackay) south of the Wet Tropics. Since this occurs beyond the range of the data, local ecological knowledge may refine this choice.



Figure 1. Golden bowerbird: Pseudo-absences were generated (column 2) for study regions (column 3) specified at the continental scale of Australia (top) or bioregional scale of the Wet Tropics (bottom). Mapped predictions (darker red for more likely presence) resulting from fitting SDMs using 3 algorithms (right).



Figure 2. Distribution of the predicted probability of presence (PPP) (x-axis), showing relative proportion (y-axis) of actual presences (blue) and absences (red), by SDM algorithm and study region (columns).

At the continental scale, all algorithms provided near perfect prediction of presence and absence, with very strong contrast, as depicted in the predictions plot (Fig. 2): the predicted probability of presence for true presences is extremely high (near 0) and for pseudo-absences is extremely low (near 1), with negligible misclassification (flat in the mi ddle). At the bioregional scale, SDMs were able to differentiate habitats due to a wider spread in PPP (Fig. 2, right). Thus for the next phase, we chose bioregional scale to investigate environmental gradients within the envelope.

4.2 Experiment with SDM Algorithms in bioregion

In the Wet Tropics, MaxEnt provided the best prediction of absences (narrowest, red), but the worst prediction of presence (widest, blue). The tree model (CTA) misclassified a 'heap' of presences with low PPP (blue overlap, left), and a heap of absences with high PPP (red overlap, right). The regression model (GLM) gave the best and most gradual separation in the PPP for presences *vs* absences. Other statistical and machine learning algorithms provided similar results to these three shown. Further model exploration focused on GLM as the simplest (most parsimonious) of these models.

These apparent differences in predictive performance among algorithms could be explained in different ways. Misclassification of presences could occur if the study region has been poorly defined, or search effort has limited coverage. If so, then the tree algorithms map areas that would benefit from extra search effort. Alternatively it could be that the regression algorithms reflect the underlying ecological processes better: tradeoffs among habitat characteristics may be better than decision rules for describing Golden bowerbird habitat.

4.3 Experiment with pseudo-absence strategy in bioregion, for GLM

Using random generation of pseudo-absences throughout the whole Wet Tropics, we found very little improvement (and hence results are not shown) when changing the strategy or sampling intensity. Setting an outer perimeter to a Convex Hull gave similar performance to using the Wet Tropics. Intensities 1:1, 10:1 or 60:1 gave similar results in the Wet Tropics. Starting with the whole bioregion (right to left, Fig. 3), we considered strategies that eliminate more sites based on their distance from occurrences, in geographic space (i.e. using discs) or in environmental space (i.e. using SRE). This gradually reduced the differentiation between presences and absences, as to be expected, but also reduced contrasts to better reveal gradients (Section 4.1).

For the last phase of experimentation we chose a strategy with good discrimination (well separated PPP for absence *vs* occurrence in Fig. 3) whilst retaining some information about the gradient (non-zero probability in the middle). This corresponded to a ratio of 10:1 pseudo-absences to occurrences, which were randomly generated in the Wet Tropics with an inner disc boundary of 10km.



Figure 3. Comparison of sampling strategies for pseudo-absences: predicted probability of presence for absence (red) and presence (blue) obtained from separate modelling runs in BCCVL.

4.4 Assess naughty noughts and variable importance

CTA, with penalty for false negatives set equal to false positives, isolated largely uninhabited landscapes using simple splitting rules (Fig. 4, 5). Golden bowerbirds were rarely found (< 2%) where the average temperature of the warmest quarter exceeds¹ 23.75°C. For the three strategies examined at each of three intensities (the four most different shown here), we found that the mean temperature of the warmest quarter was consistently involved in identifying climates with no known occurrences of the Golden bowerbird.

Intensity	PA Strategy	Rules for low presence		% area	PPP in area
10:1	Random	Temp Y \geq 20.25	Temp.WarmQ ≥ 23.35	25.17	0.0272
10:1	SRE	Temp.WarmQ≥ 24.05	Temp.WarmQ ≥ 26.55	19.83	0.002017
10:1	Disk, min=10km	Temp.WarmQ ≥ 23.75	Temp.WarmQ \geq 24.05	26.29	0.01598
60:1	Random	Temp.WarmQ ≥ 23.25	Temp.ColdQ \geq 15.65	25.64	0.005135

Figure 4. Evaluation of naughty noughts for different pseudo-absence strategies and intensities in the Wet Tropics: % area corresponding to naughty noughts; and their predicted probability of presence (PPP).



Figure 5. Comparison of effect sizes estimated for predictors in a GLM, when naughty noughts were omitted (left) or retained (right). Pseudo-absence strategy: Wet Tropics, 10:1 intensity, inner boundary disc at 10km.

These results show that retaining rather than omitting naughty noughts (Fig. 5, right vs left): 'dampens' the relative climate effects (estimates between -0.5 to +0.5) and 'muffles' them (confidence intervals are longer). Conversely, omitting naughty noughts leads to larger relative effects (from -10 to +5) and fewer confidence intervals support zero effect sizes. This also affected the relative importance of variables, especially the interaction between temperature in the coldest and warmest quarters. Climate variables associated with less chance of Golden bowerbird are, in order of influence, annual measures (precipitation and mean temperature), and high annual mean temperatures in conjunction with low mean temperature of the coldest quarter (or vice versa). Variables associated with more chance of Golden bowerbirds are: precipitation in the wettest or driest quarter, and temperature of the warmest quarter.

5 CONCLUSIONS

In this case study, the VL provided a massive reduction—in the amount of data preparation, GIS, coding and hence skills—required to run all analyses directly in R (R Core Team, 2017) or even in a library like biomod (Thuiller et al., 2009). Time was instead dedicated to experimenting with pseudo-absence strategies, to investigate: study region, SDM algorithm, sampling extent and intensity, and impact of naughty noughts on variable contributions. We first confirmed that the Golden bowerbird's distribution was highly sensitive to the scale of the study region, which in turn affected whether pseudo-absences were generated across the continent or within the bioregion. Of the SDM algorithms considered, GLM best satisfied our preference to control misclassification of presences. We chose a pseudo-absence sampling strategy to balance risk of pseudo-replication with discrimination of presence vs absence. Naughty noughts could easily be identified using mean temperature of the warmest quarter, and omitting these, we found that the predicted distribution of the Golden bowerbird was affected by annual and extreme measures of temperature and precipitation. This

¹The threshold of 23.75 is apparently precise, since it reflects the best performing split based on interpolated climate surfaces.

paper demonstrates the steps involved in an iterative approach to modelling, that contrasts with more usual approaches reported as a single model or a single experiment, comparing several models. We hope that users new to iterative modelling will find it useful.

Beyond SDM, we demonstrate more broadly the potential of virtual laboratories to facilitate an experimental approach rather than a prescriptive approach to modelling. This aligns with guidelines advocating an iterative cycle of statistical modelling (Greenland et al., 2016). Without a VL, there would be much less time available for producing, interpreting and refining results. In fact, the VL provided more support for the collaborative aspects of a modelling project, here between a quantitative specialist and an ecologist.

ACKNOWLEDGEMENT

The authors would like to acknowledge the BCCVL project team, who are supported by the National eResearch Computational Tools and Resources Project (NecTaR), an initiative of the Australian Commonwealth Government, Griffith University, Aust. National Data Services (ANDS), and Research Data Services (RDS).

REFERENCES

- Austin, M. P. and J. A. Meyers (1996). Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *Forest Ecology and Management* 85, 95–106.
- Barbet-Massin, M., F. Jiguet, C. H. Albert, and W. Thuiller (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution 3*, 327–338.
- Belbin, L. and K. J. Williams (2016). Towards a national bio-environmental data facility: experiences from the Atlas of Living Australia. *International Journal of Geographical Information Science* 30(1), 108–125.
- Booth, T. H., H. A. Nix, J. R. Busby, and M. F. Hutchinson (2014). BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Diversity and Distributions* 20(1), 1–9.
- Franklin, J. (2010). *Mapping Species Distributions: Spatial Inference and Prediction*. Ecology, biodiversity and conservation. Cambridge University Press: Cambridge.
- Greenland, S., S. J. Senn, K. J. Rothman, J. B. Carlin, C. Poole, S. N. Goodman, and D. G. Altman (2016). Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 31, 337–350.
- Hallgren, W., L. Beaumont, A. Bowness, L. Chambers, E. Graham, H. Holewa, S. Laffan, B. Mackey, H. Nix, J. Price, J. Vanderwal, R. Warren, and G. Weis (2016). The biodiversity and climate change virtual laboratory: Where ecology meets big data. *Environmental Modelling & Software* 76, 182–186.
- Hilbert, D., M. Bradford, T. Parker, and D. Westcott (2004). Golden bowerbird (*prionodura newtonia*) habitat in past, present and future climates: predicted extinction of a vertebrate in tropical highlands due to global warming. *Biological Conservation* 116(3), 367–377.
- Lobo, J. M., A. Jiménez-Valverde, and J. Hortal (2010). The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33, 103–114.
- Mateo, R. G., T. B. Croat, Á. M. Felicsimo, and J. Muñoz (2010). Profile or group discriminative techniques? generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. *Diversity and Distributions 16*(1), 84–94.
- Merow, C. and J. A. J. Silander (2014). A comparison of maxlike and maxent for modelling species distributions. *Methods in Ecology and Evolution* 5, 215–225.
- Pirathiban, R., K. J. Williams, and S. J. Low-Choy (2015, December). Delineating environmental envelopes to improve mapping of species distributions, via a hurdle model with CART &/or MaxEnt. In T. Weber, M. McPhee, and R. Anderssen (Eds.), 21st International Congress on Modelling and Simulation, Gold Coast, Australia, 29 Nov to 4 Dec 2015, pp. 1455–1461. Modelling and Simulation Society of Australia and New Zealand.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Thuiller, W., B. Lafoucade, R. Engler, and M. B. Araújo (2009). BIOMOD: a platform for ensemble forecasting of species distributions. *Ecography 32*, 369–373.
- VanDerWal, J. J. (2012). All future climate layers for Australia 5km resolution. Dataset, James Cook University. https://researchdata.ands.org.au/future-climate-layers-5km-resolution/10856.
- Wisz, M. S. and A. Guisan (2009). Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecology* 9(8), 13pp.
- Witman, J. D., R. W. Lamb, and J. E. Byrnes (2015). Towards an integration of scale and complexity in marine ecology. *Ecological Monographs* 85(4), 475–504.