NeoBiota 23: 95–99 (2014)
doi: 10.3897/neobiota.23.7939
http://neobiota.pensoft.net/

COMMENTARY

A peer-reviewed open-access journal
NeoBiota
Advancing research on alien species and biological invasions

Response to Kriticos et al.

Jane Elith[1], Mark A. Burgman[1]

1 *Centre of Excellence for Biosecurity Risk Analysis. School of Botany, The University of Melbourne. Parkville. Australia 3010*

Corresponding author: *Jane Elith* (j.elith@unimelb.edu.au)

Academic editor: *I. Kühn* | Received 22 May 2014 | Accepted 5 June 2014 | Published 2 September 2014

Citation: Elith J, Burgman MA (2014) Response to Kriticos et al. NeoBiota 23: 95–99. doi: 10.3897/neobiota.23.7939

Kriticos et al. (2014) discuss a recent paper of ours (Elith et al. 2013). While we agree with several of the points they raise, in this brief response we focus on clarifying a subset of issues around improving pest risk modelling. We do so because several of their suggestions are based on misunderstandings of technical details and will not, as they hope, lead to improved modelling practice.

Elith et al. (2013) used the *Puccinia psidii* complex as a case study to explore the impact of taxonomic uncertainty on modelled predictions. The work was clear in stating that it did not intend to provide definitive predictions and advice for biosecurity managers. This position is clear in the title and in statements in Elith et al. (2013) such as 'The purpose of this study is to explore the implications of taxonomic uncertainty for the management of a new invasive pathogen. We use the Australian incursion of myrtle/ guava rust as an example, not to argue which taxonomic interpretation, data set or model is correct, but to highlight the impact of taxonomic belief on modelled predictions.' We discussed how to make risk weighted decisions that accommodate this kind of uncertainty. We created 5 datasets of species records that were either the full set of available presence records for the *P. psidii* complex ("Pucc94") or subsets of that, each of which accorded with a different taxonomic interpretation.

The Kriticos et al. (2014) commentary primarily uses for illustration the Pucc94 dataset (94 records) and the Ured27 dataset – a subset of Pucc94 comprising 10 records of *Uredo rangelii* (variously viewed as a separate species or a member of the complex) and an additional 17 records from similar environments (see Elith et al. 2013 for details). Elith et al. (2013) emphasised that "we recognise that the basis for these groupings could be debated and that future evidence might prove them wrong. Our aim ... is not to argue that the specific choices are indisputable, but to create five distinct datasets with reasoning behind each, and to use these to model and predict". In other words, we were not trying to model these entities definitively, but to explore the effect on predictions of different perspectives on the group's taxonomy. We used the modelling software Maxent (Phillips et al. 2006, Phillips and Dudik 2008) and 7 covariates selected for their likely ecological relevance to model these five datasets.

Kriticos et al. (2014) note that the covariate rankings in the five models vary, which they interpret as 'unstable'. They state "instability in the covariate importance rankings in this type of analysis provides an indication that the model may be unsound". In Maxent, variable importance is estimated in two ways – one during model building and the other on permutation tests on the final model. Kriticos et al. (2014) focus on the second and expect similar covariate rankings across all 5 datasets. We agree that repeated random samples of the full distribution of a species are likely to lead to similar covariate rankings, provided there are enough samples to reliably model the species *and* provided the entity is in fact a single species. We emphasise, though, that we were not taking repeated random samples. Many circumstances could lead to different covariate rankings. These include different datasets representing different species or subspecies, a random sample that happened to be biased to one part of covariate space, or highly correlated variables. In the latter case, different covariate rankings may lead to very similar predictions because covariates are largely interchangeable. There is no reason to impose an *a priori* expectation that covariate rankings should be similar across our 5 datasets, not the least because we were exploring the possibility that some subsets display different environmental constraints because they represent different taxa. Kriticos et al.'s arguments are pre-conditioned on the assumption that we are dealing with one species. We held no such presumption.

Kriticos et al. (2014) used Rodda et al. (2011) as a support for their argument regarding the stability of covariate rankings. We fail to see the connection with our work. Rodda et al. (2011) discussed rote use of Maxent with default settings and the commonly available suite of 19 Worldclim variables. They specifically stated that their remarks did not apply to "execution of Maxent with different (i.e. customised) settings". We did not use default settings and 19 variables, and even explored the effects of our choices on the modelled outcomes.

Kriticos et al. (2014) note that the choice of background affects model output. They see this – together with the other issues they address - as such a problem that they suggest instead using presence-only methods that do not require background points. It is well established that choice of background affects model outcome (e.g. Elith et al. 2010, Elith et al. 2011, Elith 2014). In Elith et al. (2013), we described our approach to selecting the background and mapped examples of the background extent. We used a strategy consistent with the approach of an informed user (Elith et al. 2013 Fig 1, grey areas). We did not mention the extent of mapping in the figure legend. We agree that this omission was an oversight, and that generally authors should be specific about their choice of background. We tested the effect of varying the background on model results during model construction and found that it was insubstantial. We agree that it would be better to include such results in appendices.

However, the solution to this issue is not to do as Kriticos et al. (2014) suggest and resort to presence-only methods that do not require background points. Methods that do not take a background sample are ignorant of environmental conditions in the region in which the model is trained. This confounds the frequency of environments at occupied locations with the frequency of environments in the region, an issue that is difficult to overcome and that leads to decreased predictive performance, at least in a number of tested equilibrium situations (Elith et al. 2006). We are left with the problem that choice of background is to some extent subjective, and we suggest that good practice includes exploring the effect of choices on modeled results.

Kriticos et al. (2014) question covariate values and extrapolation, in particular the process of clamping predictions in novel environmental domains. We agree that clamping will affect predictions in novel space. That is the intent. Clamping is the default choice and is quite commonly used, including by one of the authors of Kriticos et al. (2014) in Thompson et al. (2011). It is a sensible default because it ensures that predictions outside the sampled range are at least consistent with those made at the most similar sampled environment.

Nevertheless, we agree that predictions in novel environmental space should be treated very cautiously. In Elith et al. (2013), our interpretation and interest was focused on areas of relatively high predictions in Australasia and none of these were in novel space. To guard against unintended uses of our results, we could have masked out all regions with any amount of extrapolation, to make it clearer that (1) we were not focusing on any of these areas, and (2) we do not trust predictions in extrapolated areas. Instead, we took the more conventional path of including numerous messages throughout the manuscript and its appendices that showed the reader that extrapolations are inherently uncertain and should be treated cautiously.

Rather than dealing with every issue in the Kriticos et al. (2014) commentary, we now address a remaining important point regarding areas and extents. As Kriticos et al. (2014) note, Maxent (and other methods for modeling presence-background data) outputs relative probabilities or relative intensities. We agree that it would have been useful in Elith et al. (2013) to remind readers of this point specifically, in case they were inclined to misinterpret them as probabilities of occurrence. Nevertheless, the treatment in Elith et al. (2013) was consistent because it focused on relatively suitable locations, those with the highest relative predictions for each taxon. We have discussed this with Kriticos et al. in person, so it is disappointing to find that they persist in stating that Elith et al. (2013) focused on extent.

Regarding "Predicted area", Kriticos et al. (2014) appeal to "both set theory and ecological reasoning" when discussing "Area" and "Habitat". They then apply their reasoning – largely posed in environmental space – to our results, which they interpret in geographic space. They state "the broader the range of environmental tolerances encompassed by an organism in its native range, the broader the range of conditions we might suppose it is at least capable of inhabiting in an introduced range". The discussion in Kriticos et al.

(2014) on this point seems at least unclear if not wrong. In fact, range in geographic space depends on the environments available in each region, and that the tolerances of taxa, even closely related ones, are not necessarily nested in environmental space.

Consider this situation: species A and B exist as native species in a certain region. A might be widespread, and B more narrowly distributed. Only a subset of environments that species A occupies in its native range might exist in a new region, whereas all the environments occupied by species B might exist – in fact, in the new region *more* environments suitable for B might exist than were available in the native range. Thus it is possible (though completely dependent on the relationships between environmental and geographic space) that species A may tolerate a wider range of environmental conditions than B in their native ranges, but species B may have a wider geographic range in a new region. Thinking further about relationships between environmental and geographic space (see the excellent discussion of these issues by Colwell and Rangel 2009): the same suitable environments might be repeated many times in geographic space (the biotope) in a new region, implying that geographic areas are difficult to predict conceptually from environmental thinking. We will not take this argument further since it is difficult to discuss fully in a short reply. Our main point is that whilst Kriticos et al. (2014) might want to interpret our results in terms of areas, we did not make inferences about areas, and their arguments regarding ranges in environmental and geographic space are neither necessary nor sufficient.

Predicted area is not necessarily identifiable from presence-only data. Some authors use 'thresholding' (setting all values above some predicted value to 1 and all below to zero). However, this is not a remedy, since this merely serves to decrease the amount of information available. In Elith et al. (2013) we discussed the effect of small sample sizes (generally predictions will be less well differentiated with small samples) and interpreted the results accordingly.

Lastly, Kriticos et al. (2014) raise a general theme that the results in Elith et al. (2013) will be (mis-)used for other ends. This of course is possible. Elith et al. (2013) focused on the influence of taxonomic uncertainty on the predicted location of the most suitable sites for a set of putative taxa. Those wanting to use the results for other purposes should do so thoughtfully. Our hope is that people wanting to use the predictions in other contexts would either know about or inform themselves about the modelling methods sufficiently that they can make competent interpretations or contact the authors of the original work and ask about their interpretation.

In conclusion, while we appreciate the motivation of Kriticos et al. (2014) was to improve modelling practice and we agree with several of the points they raise about clarity and completeness of descriptions, we find that their main methodological complaints are based on misunderstanding of the technical details of the methods themselves. Our concern and motivation in responding is that their advice is broadly misguided. We agree with some of their concerns regarding the limitations of using correlative methods to model distributions of invasive species (see Elith 2014 for a detailed discussion). The difficulties in predicting species distributions in novel environments remain open and important questions.

References

Colwell RK, Rangel TF (2009) Hutchinson's duality: The once and future niche. Proceedings of the National Academy of Sciences 106: 19651–19658. doi: 10.1073/pnas.0901650106

Elith J (2014) Predicting distributions of invasive species. Available from arXiv. http://arxiv.org/abs/1312.0851

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson KS, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129–151. doi: 10.1111/j.2006.0906-7590.04596.x

Elith J, Kearney M, Phillips SJ (2010) The art of modelling range-shifting species. Methods in Ecology and Evolution 1: 330–342. doi: 10.1111/j.2041-210X.2010.00036.x

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17: 43–57. doi: 10.1111/j.1472-4642.2010.00725.x

Elith J, Simpson J, Hirsch M, Burgman MA (2013) Taxonomic uncertainty and decision making for biosecurity: spatial models for myrtle/guava rust. Australasian Plant Pathology 42: 43–51. doi: 10.1007/s13313-012-0178-7

Kriticos DJ, Morin L, Webber BL (2014) Taxonomic uncertainty in pest risks or modelling artefacts? Implications for biosecurity policy and practice. NeoBiota 23: 81–93. doi: 10.3897/neobiota.23.7496

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190: 231–259. doi: 10.1016/j.ecolmodel.2005.03.026

Phillips SJ, Dudik M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31: 161–175. doi: 10.1111/j.0906-7590.2008.5203.x

Rodda GH, Jarnevich CS, Reed RN (2011) Challenges in Identifying Sites Climatically Matched to the Native Ranges of Animal Invaders. PLoS ONE 6: e14670. doi: 10.1371/journal.pone.0014670

Thompson GD, Robertson MP, Webber BL, Richardson DM, Le Roux JJ, Wilson JRU (2011) Predicting the subspecific identity of invasive species using distribution models: Acacia saligna as an example. Diversity and Distributions 17: 1001–1014. doi: 10.1111/j.1472-4642.2011.00820.x