



How to characterize and predict alien species? A response to Pyšek *et al.* (2004)

In a recent study we attempted to characterize alien species in two regions in Argentina (Prinzing *et al.*, 2002). We abandoned the widespread approach of comparing alien species with local natives. This approach confounds differences between aliens and non-aliens with differences between species from different parts of the world. Instead we used a source-area approach. Within the pool of species in a source region we compared those that became aliens in the Argentinean target regions with those that did not. As a source region we considered central Europe because it is one of the major source pools for the colonization of Argentina (see References in Prinzing *et al.*, 2002). We excluded 'neophytes' from the source pool, i.e. species that have been introduced to (parts of) Central Europe later than c. 1500 AD (Ellenberg, 1996). For such neophytes, Central Europe is unlikely to be the true source region from which they were transported to Argentina. We found clear differences between Argentina's alien and non-alien species: alien species are widely distributed in central Europe and worldwide, they are utilized by humans (crops, ornamental plants, medicine, or spices), they prefer warm, dry and nitrogen-rich conditions, and they show a ruderal life strategy. We also found that, in a cross-validation, these traits permitted us to predict correctly 81% of the alien species and 70% of the non-alien species.

In their paper, Pyšek *et al.* (2004) refer extensively to our study. They basically welcome our source-area approach and aspects of our analyses. But the authors also make three critical comments. First, they suggest that 'archaeophytes', species introduced to the Central European source area before 1500 AD, should also be excluded from the analysis. Second, they suggest analysing species of different biogeographical origin separately. Third, they suggest that the predictive power of the statistical model relating species traits to alien status should be assessed for all species pooled, not for aliens and non-aliens

separately. Whilst we welcome these constructive suggestions, we feel that their criticisms deserve some response. We start with the third criticism because it leads the authors to a particularly harsh conclusion as to the usefulness of our analysis.

WHAT IS A USEFUL PREDICTION?

In our paper we developed a model relating species traits to alien status for a subset of the species pool and then used it to predict the alien status of the remaining species. Pyšek *et al.* evaluate the usefulness of this predictive model by recording how many of the species predicted to be aliens are indeed aliens (true positives), and how many predicted to be non-aliens are indeed so (true negatives). Using our own cut-off value separating predicted aliens and predicted non-aliens, Pyšek *et al.* find a very high percentage of true negatives (97%) but a low percentage of true positives (21%). There is thus a high proportion of false positives. The authors conclude that 'no useful prediction' is possible and that there is hence 'no point in refining *this particular analysis*' (our italics).

We first want to point out that the percentage of false positives could of course be strongly reduced by shifting the cut-off point. But this would be at the cost of very strongly increasing the number of false negatives, of aliens predicted to be non-aliens. From a conservation perspective, these false negatives are more problematic than false positives. False negatives are overlooked aliens, and with some bad luck they may end up as invasive aliens that devastate parts of the native flora. False positives 'only' trigger a false alarm (which admittedly will cause substantial costs when included in screening programs or practical eradication measures). Second we want to remind the reader why there are so many false positives in our analysis. Pyšek *et al.* themselves mention the reason. It is simply because non-aliens make up the bulk of the species in the source pool (91%). Hence non-aliens are bound to make up also a large part of the incorrectly predicted species, of the false positives (except if the prediction is made extremely restrictive, see above).

However, this is by no means 'particular' to our data set. Any source pool contributes only a minority of its species as aliens to a specific target region. Hence, if we apply the criteria of the authors, a satisfactory prediction of alien species will never be feasible. We believe, however, that these criteria are too restrictive. Our cross validation had shown that, despite many false positives, our model precisely (i.e. reproducibly; Zar, 1984) predicts most true positives. The practical usefulness of models with even much larger numbers of false positives has already been proven in epidemiology where they help to fight the spread of diseases (Loong, 2003). Moreover, it has recently become possible to quantify the predictive power of a model without bias from the prevalence of aliens and non-aliens in the source pool, based on the Delta statistic of Martin Andrés & Femia Marzo (2004). The authors demonstrate the superiority of this statistic over classical measures such as Kappa. Delta ranges from 0 (no correct prediction) to 1 (perfect prediction) and our logistic regression model yielded a Delta of 0.57 ± 0.014 . Delta even increased when we reduced the cut-off value separating predicted aliens and non-aliens, but this was at the cost of many false negatives (see above). Hence, by reference to these Delta statistics we reaffirm the usefulness of our model for predicting the alien status of species.

The most important point is that false positives, non-aliens predicted to be aliens, may not be such a bad thing overall. If we recall that our model correctly predicts most true aliens something may be special about the non-aliens misclassified as aliens. These false positives might in fact be not-yet-alien species that have a high risk of becoming aliens in the future. For instance, all five species which are most notoriously misclassified as aliens in our cross validation analysis (*Solanum nigrum*, *Echium vulgare*, *Descurainia sophia*, *Chelidonium majus*, *Polygonum lapathifolium*) have indeed already become aliens in other places of South America, some even within Argentina (Söyrinki, 1991; Brako & Zarucchi, 1993; Villamil & Zalba, 2002; Corra, 1969–86; Meusel *et al.*, 1978–92). They hence have a high chance of becoming aliens also in the

Argentinean regions we investigated. For conservation purposes, this is exactly the kind of information needed. These false positives deserve attention if we want to prevent any further increase of the already existing alien flora.

HOW TO DELIMITATE A SPECIES POOL?

We now come to the first and second criticism of Pyšek *et al.* that archaeophytes should be excluded from the source pool and species of different biogeographical origin should be analysed separately. The justification for both points is essentially the same: indigenous species and archaeophytes, as well as species from different biogeographical regions, differ in their chance to get into contact with humans travelling to Argentina. Hence, they differ in their chance of becoming aliens. At the same time these groups of species differ in several ecological traits. Most archaeophytes, for instance, are agricultural weeds, and hence most of them are therophytes and display a CR or R strategy. Therefore, observed correlations between ecological traits of species and their alien status in Argentina might in essence be 'pseudocorrelations'.

Pyšek *et al.* specify how these groups of species may differ in their chance to get into contact with humans. First, many archaeophytes are/have been cultivated by humans or are associated with the cultivars, which obviously increased the chance of contact. Second, many species originating in biogeographical regions far outside central Europe may only be narrowly distributed within central Europe, which would decrease the chance of contact within Central Europe. Finally, archaeophytes and species originating far outside central Europe may get into contact with humans also outside central Europe. In our analysis we were indeed aware of the first and the second mechanisms and tried to take them into account. We included the variables *use by humans*, and *occurrence in anthropogenic vegetation* as a measure of the degree to which species are cultivated or associated with cultivars; and we included the *grid frequency* as a measure of the species' distribution within central Europe. We did, admittedly, not account for the third possibility, contact with humans outside central Europe. This may have introduced a bias into our analysis.

To rectify this bias Pyšek *et al.* suggest dropping archaeophytes from the analysis and analysing species from different biogeographical regions separately. However, this implies that, for instance, an archaeophyte became an alien because it is an archaeophyte

and not because it is characterized by certain ecological traits. But how shall we know this a priori? We believe that a more adequate way would be to include the *status* (indigenous or archaeophyte) and the *biogeographical origin* as additional variables into the model. With this information we can check whether or not the ecological traits still have an effect on the capacity of species to become aliens. We can also assess which are the most important determinants of the capacity of species to become aliens: the ecology of species, their status, or their biogeographical origin.

We have therefore re-calculated our multiple regression model across species as data points, and included the archaeophytic vs. indigenous status of the species, and we have included a rough indicator of their biogeographical origin (European endemics vs. non-endemics as given in Meusel *et al.*, 1978–92; based, for each species, on inspections of dozens to hundreds of regional floras worldwide). We found that despite the large sample size *biogeographical origin* had no effect at all ($P = 0.93$). Hence, further refinement of the definition of *biogeographical origin* is unlikely to give fundamentally different results. *Status*, in contrast, was highly important. It was indeed the most important variable, together with the *grid frequency* of the species ($t_{734} = 5.41$ and 5.59 , $P < 0.001$). Nevertheless, all other variables that significantly contributed to the original model also did so in the new model (P -values changed for less than 0.034). Only the variable that was least significant in the original model now failed: *ruderal life-strategy* ($P = 0.29$ instead of 0.003). This is the variable that is most strongly correlated with species' *status* according to Pyšek *et al.* Of course, a variable may fail as an artifact of multicollinearity, but this was probably not the case. The tolerance value of *ruderal life-strategy* was rather high (0.6). Moreover, *ruderal life-strategy* remained completely non-significant even in a ridge regression analysis, a procedure that efficiently reduces multicollinearity (StatSoft, 1999).

This re-analysis of our data shows that the archaeophytic vs. indigenous status of species is indeed an important trait that should be considered when characterizing and predicting alien species. However, we suggest that this trait should not be treated as a filter to exclude certain species. Instead it should be included as an additional covariable. Moreover, it should be kept in mind that any approach aimed at taking into account *status* will introduce new complications. Many of the species ranked as archaeophytes in floristic data bases are classed as such only in certain parts of the reference area, in other parts these

'archaeophytes' may well be indigenous. This is especially true for modern databases representing floras of large regions covering a whole source pool from which aliens in another part of the world have originated.

Overall, we believe that this exchange between Pyšek *et al.* does not require a substantial change to our original conclusion that aliens differ from non-aliens in several traits: they are common within their source region, utilized by humans, preadapted to the abiotic conditions in the target region and ecologically versatile. The discussion has, however, provided additional insight into the role of archaeophytic status and has down-rated the importance of ruderal life-strategy for the chance of species becoming an alien. We hope that this discussion will contribute to an optimization of the protocol of how to characterize and predict alien species.

ACKNOWLEDGEMENTS

This research was supported through a European Community Marie Curie Individual Fellowship to AP (contract number MCFI-2001–01539). The manuscript benefited from comments by two anonymous referees.

ANDREAS PRINZING^{1,†},
WALTER DURKA², STEFAN KLOTZ² and
ROLAND BRANDL³

¹University of Nijmegen; Dept. of Ecology & Alterra Institute Wageningen; Dept. of Ecology and the Environment; Postal address: Alterra Institute; Dept. of Ecology and the Environment; Postbus 47; 6700 AA Wageningen; The Netherlands, ²UFZ — Centre for Environmental Research Leipzig-Halle; Department of Community Ecology; Theodor-Lieser-Straße 4; 06120 Halle (Saale); Germany,

³Philipps-University Marburg; Faculty of Biology, Department of Animal Ecology; Karl-von-Frisch-Strasse 1; 35032 Marburg; Germany

[†]Present address: Université de Rennes 1; UMR-CNRS 6553-Ecobio; Campus de Beaulieu, Bâtiment 14A; 263 Avenue du Général Leclerc, CS 74205; 35042 Rennes Cedex; France, E-mail: andreas.prinzing@wur.nl

REFERENCES

- Brako, L. & Zarucchi, J.L. (1993) Catalogue of the flowering plants and gymnosperms of Peru. *Monographs in Systematic Botany from the Missouri Botanical Garden*, **45**, 1–1286.
- Correra, M.N. (1969–86) *Flora patagonica I–VII*. Colección científica del INTA, Buenos Aires.

- Ellenberg, H. (1996) *Vegetation mitteleuropas mit den alpen in ökologischer, dynamischer und historischer sicht*, 5th edn. Ulmer, Stuttgart.
- Loong, T.W. (2003) Understanding sensitivity and specificity with the right side of the brain. *British Medical Journal*, **327**, 716–719.
- Martin Andrés, A. & Femia Marzo, P. (2004) Delta: a new measure of agreement between two ratios. *British Journal of Mathematical and Statistical Pathology*, **57**, 1–19. Software available at <http://www.ugr.es/~bioest/Delta.pdf>.
- Meusel, H., Jäger, E.J., Rauschert, S. & Weinert, E. (1978–92) *Vergleichende chorologie der zentraleuropäischen Flora I–II*. Fischer Verlag, Jena.
- Prinzing, A., Durka, W., Klotz, S. & Brandl, R. (2002) Which species become aliens? *Evolutionary Ecology Research*, **4**, 385–405.
- Pyšek, P., Richardson, D.M. & Williamson, M. (2004) Predicting and explaining plant invasions through analysis of source area floras: some critical considerations. *Diversity and Distributions*, **10**, 179–187.
- Söyrinki, N. (1991) On the alien flora of the province of Buenos Aires, Argentina. *Annales Botanici Fennici*, **28**, 59–79.
- StatSoft (1999) *STATISTICA 5.5*. StatSoft Inc, Tulsa, USA.
- Villamil, C.B. & Zalba, S.M. (2002) Proyecto. *Red de información sobre especies exóticas invasoras -i3n-Argentina. final report*. Universidad Nacional del Sur, Bahía Blanca, Argentina. http://www.iabin-us.org/projects/i3n/i3n_documents/final_reports/final_argentina_report.doc.
- Zar, J.H. (1984) *Biostatistical analyses*, 2nd edn. Prentice Hall International, Inc., Englewood Cliffs, New Jersey.