Ecography

ECOG-00403
Godsoe, W. 2013. Inferring the similarity of species distributions using Species' Distribution Models. – Ecography 36: xxx–xxx.

Supplementary material

APPENDIX 1:

Derivation for ESP (Equation 2)

The expected number of presences of species 1 in cells where species 2 is present is:

$$\sum_j P_{1,j} P_{2,j}$$

Just as the expected number of presences of species 2 in cells where species 1 is present is:

$$\sum_j P_{1,j} P_{2,j}$$

Thus the total number of presences in which 1 of the species will be in the same location as the other is:

$$\sum_j P_{1j} P_{2j} + \sum_j P_{1j} P_{2j}$$
$$= 2\sum_j P_{1j} P_{2j}$$

We can also compute the expected number of presences for species 1 and 2 as: ΣP_{1j}, ΣP_{2j} respectively. Noting this, the expected total number of presences across both species is $\Sigma(P_{1j}+P_{2j})$. Thus, the proportion of all presences from locations with both species (shared presences) is:

$$= \frac{2\sum_j P_{1,j} P_{2,j}}{\sum_j \left(P_{1,j} + P_{2,j}\right)}.$$

This is equation 2 in the main text.

Proof that Equation (2) is an extension of Sørensen similarity

If we eliminate uncertainty, such that the probability of presences in cell j are either 0 or 1 then $P_{1j} \in (0,1)$ and $P_{2j} \in (0,1)$. When this is true, $P_{1j}P_{2j}=1$ if $P_{1j}=1$ and $P_{2j}=1$. $P_{1j}P_{2j}=0$ otherwise. As a result, $2\Sigma P_{1j}P_{2j}$ is twice the sum of locations in which both species are predicted to be present, or in traditional terminology $2a$. P_{1j} will be equal to 1 either if both species are present (a in a confusion matrix) or if species 1 is present (b in a confusion matrix). P_{2j} will be equal to 1 either if both species are present (a in a confusion matrix) or if only species two is present (c) in a confusion matrix. As a result $\Sigma(P_{1j}+P_{2j})=\Sigma P_{1j}+\Sigma P_{2j}=a+b+a+c=2a+b+c$

Substituting these expressions into equation (2) we obtain 2a/(2a+b+c) or Sørensen's similarity index.

Proof that equation (2) varies between 0 and 1:

Since $0 \leq P_{ij} \leq 1$ the numerator in equation (2) is a sum of multiples of two positive numbers, and hence is never negative. Likewise the denominator is the sum of two sums of positive numbers and hence never negative. Thus, the minimum value of this equation will be 0. This occurs if either P_{1j} or P_{2j} are equal to 0 for each j. Note that if the probability of presence were equal to 0 for all P_{ij} then the denominator would be 0 this metric would be undefined.

The denominator of this metric is always less than the numerator, except in the special case that $P_{1j}=P_{2j}=1$. Thus this metric reaches a maximum value when $P_{1j}=P_{2j}$

=1. Substituting these values into equation (2), we obtain the maximum value of this metric, 1.

APPENDIX 2:

```
#function to compute the expected number of shared presences
#this function requires two vectors, each the predicted
probability of presence for a species across j cells within
the same biogeographic region


E_shared_presences<-function(p1,p2){
    gdvalue<-sum(2*p1*p2)/sum(p1+p2)
    return(gdvalue)
    }
```