

FIDEGAM: A novel method based on Generalized Additive Models for quantifying the sampling effort in biodiversity databases

M.P. Pata^a; I. Pardo^a M.B. García-González^a and D. Gómez^a

^aDepartment of Biodiversity Conservation

Pyrenean Institute of Ecology (CSIC)

Zaragoza, Spain

marpata@ipe.csic.es; iker.pardo@ipe.csic.es; mariab@ipe.csic.es; dgomez@ipe.csic.es

Keywords: monitoring of biodiversity; measures of biodiversity; survey design and analysis

Abstract: Biodiversity databases are the main source of information for a fast growing body of research in macroecology and conservation biology. However, such information is often severely biased (typically some areas, species or years have been more intensely sampled than others), compromising the accuracy of results and conclusions provided by classic and advanced analysis of biodiversity. The quantification of sampling effort plays a key role in the process for correcting this bias, and different methods based on species accumulation curves are the most widely adopted. Despite the availability of diverse completeness measures for describing the accumulative-sampling effort relationship, we still lack of an objective criteria to measure the capability for discriminating between well and bad sampled units. We explore the statistical pitfalls of the available methods, evaluate for the first time their discrimination capability, propose a new method, named FIDEGAM, based on the first derivative of the curve obtained from a fitted Generalized Additive Model (GAM, Hastie and Tibshirani 1990; Wood 2006)), and illustrate how our approach can account for sampling effort in biodiversity analyses. The discrimination capability of different methods is evaluated with ROC curves (Pepe 2003) and applied to databases with different degree of exhaustiveness: (a) the database of vascular flora of a mountain Spanish National Park (Ordesa-Monte Perdido; Pyrenees) and (b) a simulated database. The novel method FIDEGAM outperformed other methods in terms of discriminatory capability and regardless of the sampling exhaustiveness. Therefore, we propose the use of this method to improve the reliability of biodiversity analyses based on databases. We show an application of this technique by incorporating FIDEGAM for modelling the alpha diversity of vascular plants in Ordesa-Monte Perdido National Park.

References

Hastie T.J. and Tibshirani R.J. (1990) *Generalized Additive Models*. New York: Chapman and Hall.

Pepe M.S. (2003) *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford: Oxford University Press.

Wood S.N. (2006) *Generalized Additive Models: An Introduction with R*. Boca Raton: Chapman and Hall.