fuzzySim: an R package for analysing fuzzy similarity in species occurrence patterns

A. Márcia Barbosa [a]

[a] CIBIO/InBIO
Universidade de Évora
7004-516 Évora (Portugal)
barbosa@uevora.pt

Mots clefs : Statistique; Ecologie; Distribution des espèces; Similitude; Logique floue.

Spatial associations among species distributions, translated into chorotypes and biotic regions, provide insights into the processes that drive biodiversity patterns and identify natural units to optimize management and conservation planning. Such associations, even when analysed in the light of fuzzy logic, rely primarily on comparing species distributions via similarity indices commonly based on categorical presence/absence data [1,2]. Proximity between species' occurrence sites is thus disregarded, such that the distributions of species living at adjacent, sometimes interspersed, but not strictly coincident localities are considered as different as those of species occurring very far from each other (Figure 1). This amplifies the effects of data deficiencies and the scale-dependence of distributional relationships, particularly for species with small distribution ranges [3]. The *fuzzySim* R package, now available on R-Forge, converts binary to fuzzy occurrence data on multiple-species data tables, using either trend surface analysis [4] or inverse distance to presence [5]; and implements fuzzy versions of the binary similarity indices that are commonly used for analysing distributional relationships and for determining chorotypes and biotic regions [6,7]. The method is robust to data source disparities, gaps or other errors in species occurrence data, even for restricted species for which slight inaccuracies can affect substantial parts of their range. The package is provided with sample data and an illustrated tutorial suitable for R beginners.

Références
[1] Olivero, J., Real, R., Márquez, A.L. (2011). Fuzzy chorotypes as a conceptual tool to improve insight into biogeographic patterns. Systematic Biology, 60, 645-660
[2] Olivero, J., Márquez, A.L., Real, R. (2013). Integrating fuzzy logic and statistics to improve the reliable delimitation of biogeographic regions and transition zones. Systematic Biology, 62, 1-21
[3] Barbosa, A.M., Estrada, A., Márquez, A.L., Purvis, A., Orme, C.D.L. (2012). Atlas versus range maps: robustness of chorological relationships to distribution data types in European mammals. Journal of Biogeography, 39, 1391-1400
[4] Legendre P., Legendre L. (1998). Numerical Ecology. Amsterdam: Elsevier
[5] Shepard D. (1968). A two-dimensional interpolation function for irregularly-spaced data. Proceedings of the 1968 23rd ACM National Conference, 517-524
[6] Jaccard P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles, 37, 547-579
[7] Baroni-Urbani C., Buser M.W. (1976). Similarity of Binary Data. Systematic Zoology, 25, 251
[8] Mitchell-Jones A.J., Amori G., Bogdanowicz W., Krystufek B., Reijnders P.J.H., Spitzenberger F., Stubbe M., Thissen J.B.M., Vohralík V., Zima J. (1999). The Atlas of European Mammals. Academic Press.
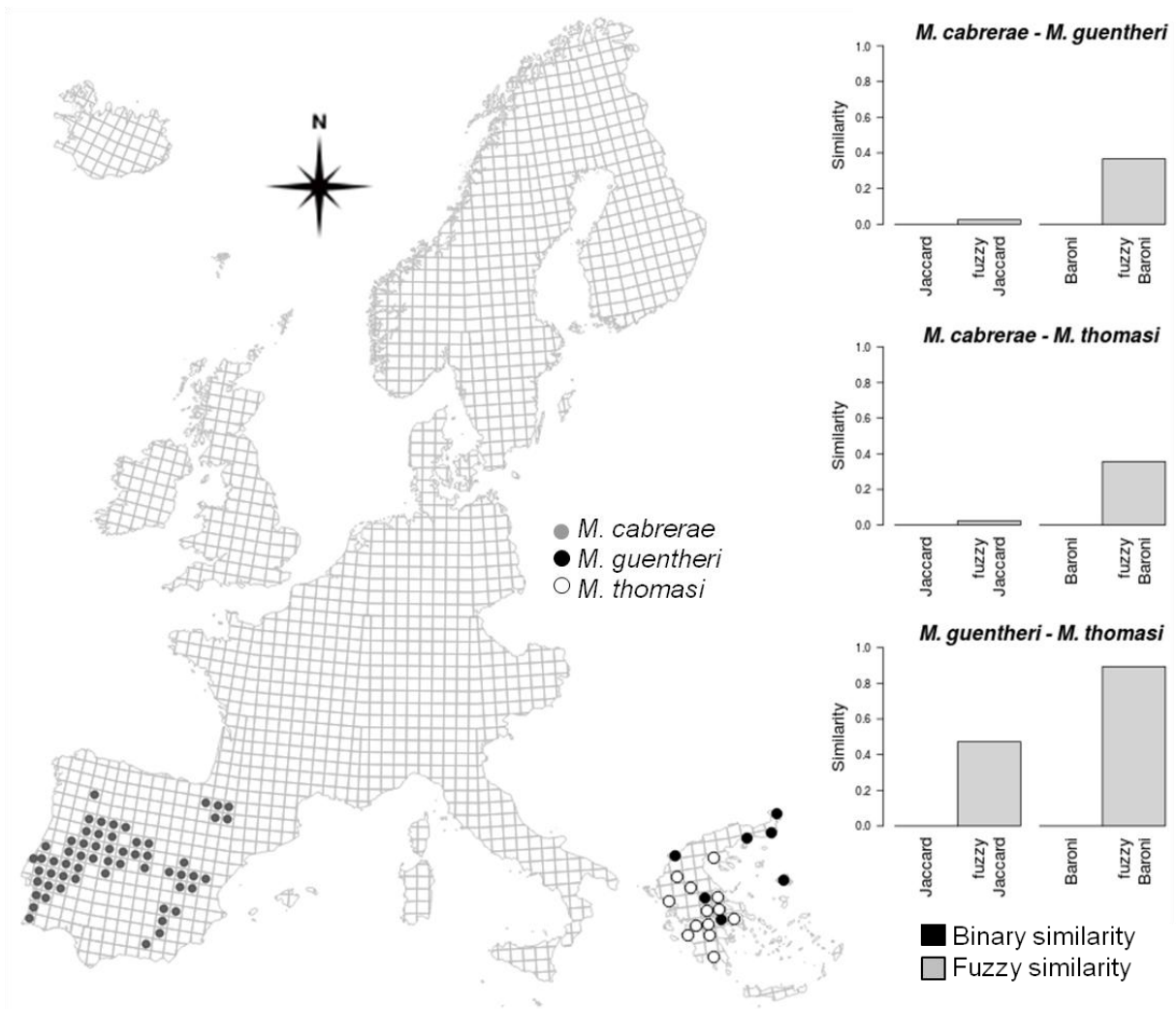
Figure 1. Left: recorded occurrences of three vole (*Microtus*) species in Western Europe on a UTM 50 km × 50 km grid [8]. Right: pair-wise similarities among these occurrence patterns using the traditional binary and the proposed fuzzy versions of Jaccard's [6] and Baroni-Urbani & Buser's [7] similarity indices. Note that binary similarity is zero in all cases, while fuzzy similarity detects that *M. guentheri* and *M. thomasi* are more similar to each other (even with no overlapping occurrences) than to *M. cabrerae*.