

Spatial non-stationarity, anisotropy and scale: The interactive visualisation of spatial turnover

Shawn W. Laffan^a

^a*School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia, 2052*

Email: shawn.laffan@unsw.edu.au

Abstract: Spatial data analysis is often focussed on the identification of spatial patterns across some geographic extent (Laffan, 2002). Properties of these spatial patterns of particular interest are: (1) the spatial scale of the process, which is essentially the extent of a set of locally related values; (2) spatial anisotropy, where the relationship of a location to other locations changes as a function of direction between them; and (3) spatial non-stationarity, where the scale and anisotropy vary across the geographic extent of the data set. All of these are inter-related and are associated with both abrupt and gradual transitions, often with both occurring within the one geographic region. There are numerous analytical methods of exploring these spatial properties, but core to them is visualisation.

Visualisation is an integral part of the data analysis process. For spatial analyses of compositional turnover, where the data consists of sets of labels or objects at each location rather than singular values, such visualisation has been difficult. This is due to issues of complexity, large number of spatial associations and computational limitations. An additional cause of this difficulty has been a lack of available software, with GIS software focussing on analyses of layers of single values per location, with compositional data stored as one layer per label or object.

This paper contains a description of the interactive visualisation of compositional turnover across geographic data sets, developed as an extension to the *Biodiverse* software (Laffan *et al.*, 2010). This represents an important step towards improved understanding of spatial turnover patterns in compositional data, particularly for spatial non-stationarity, scale and anisotropy. There is also the potential to use this approach in analyses of distributional breaks and gradational transitions, and of anomalies.

Keywords: *Spatial analysis, visualisation, spatial turnover, Biodiverse software*

1. INTRODUCTION

Spatial data analysis is often focussed on the identification of spatial patterns across some geographic extent (Laffan, 2002). Properties of these spatial patterns of particular interest are: (1) the spatial scale of the process, which is essentially the extent of a set of locally related values; (2) spatial anisotropy, where the relationship of a location to other locations changes as a function of direction between them; and (3) spatial non-stationarity, where the scale and anisotropy vary across the geographic extent of the data set. All of these are inter-related and are associated with both abrupt and gradual transitions, often with both occurring within the one geographic region. There are numerous analytical methods of assessing these spatial properties, but core to them is visualisation.

Visualisation is an integral component of data analysis and is used at nearly every stage of the analytical process. For example, before commencing an analysis it is used to develop a provisional understanding of data relationships, derive hypotheses, and choose analysis methods. Each time a set of analysis results is obtained, it is used to assess model performance, revise analytical procedures and develop additional hypotheses.

Spatial phenomena and patterns are typically visualised using maps that plot the data values across geographic space. This is a well proven method with a long history. However, often we are interested in understanding the rate of change of a phenomenon with increasing distance from some index location, a process referred to as spatial turnover. The quantification of spatial turnover is a well-established process in spatial data analysis. It underlies many analyses, for example in geostatistics when average rates of turnover (plotted as a variogram) are used to calibrate interpolation weights.

The visualisation and interpretation of turnover is often implicit for geographic layers representing a single phenomenon, for example when one compares elevation values between sets of locations. In this case one can merely plot a geographic surface and compare the differences based on the symbology used (for example, differences in hue or symbol size). Providing the symbology used is meaningful, then differences between sets of locations should be readily apparent. The rate of change for continuous field surfaces, for example elevation or rainfall, can be quantified by subtracting the value at one location from that at the other. Alternately, one can use metrics such as the semivariance from geostatistics.

However, such a visualisation approach becomes more difficult when one is analysing data consisting of collections of objects at each location. Examples include species distributions (Bickford *et al.*, 2004; Laffan & Crisp, 2003), income classes by census districts, language groups (Jones & Laffan, 2008), and pedodiversity. Continuous field data can also fall into this category, for example when one aggregates a set of spatially discrete observations to some coarser resolution, e.g. rainfall at a 1 km resolution aggregated to collections of observations at a 10 km resolution. To use the single-layer visualisation approach with collections of objects one must summarise the data into a single value per location, for example by using the population sizes of cities or the average age of populations within census polygons. Such an aggregation has the clear limitation that the underlying variability of the data is lost.

An example of such data aggregation is species diversity, where one considers the number of unique species occurring at each location. For such data one can count the number of objects at each location (the variety or richness), but this does not give a sense of the rate of change of objects between locations (Figure 1). However, for biogeographical and ecological purposes one is more often interested in the rate of change of species composition with increasing geographic distance in different directions, as opposed to the rate of change of the absolute number of species. This is known as compositional turnover.

2. CALCULATING AND VISUALISING COMPOSITIONAL TURNOVER

Compositional turnover can be calculated in a myriad of ways (Tuomisto, 2010; Legendre & Legendre, 2000; Laffan *et al.*, 2010). Examples for sets of objects or labels include Sørensen, Jaccard and S2 (Eq 1-3), while numerical data can be assessed using measures such as the mean absolute difference (MAD; Eq 4). However, the approach is fundamentally one of assessing the number of shared objects between a pair of locations relative to the total number of objects across both locations. Such compositional turnover forms the basis of simple correlogram plots (Figure 2) and analytical methods such as Generalised Dissimilarity Modelling (Rosauer *et al.*, 2009a; Ferrier *et al.*, 2007) where one relates the turnover of species composition between locations to the associated changes in environmental phenomena.

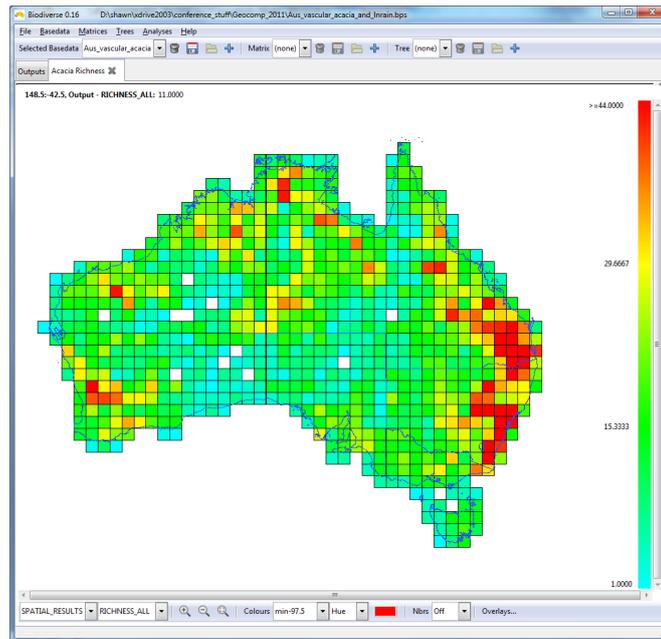


Figure 1. Species richness (variety) of Acacia species across Australia at a 1° resolution. The rate of change of richness is simple to interpret, but not the relative changes in species composition. The colour scale progresses from cyan (one species) to red (more than 44 species). White cells contain no data. Data are derived from Laffan and Crisp (2003).

$$Sørensen = 1 - \frac{2A}{2A+B+C} \quad (1)$$

$$Jaccard = 1 - \frac{A}{A+B+C} \quad (2)$$

$$S2 = 1 - \frac{A}{A+\min(B+C)} \quad (3)$$

where A is the number of objects (labels) shared between sites i and j , B is the number of objects unique to site i , and C is the number unique to site j .

$$MAD = \frac{\sum_{l_{1i} \in L_1} \sum_{l_{2j} \in L_2} abs(l_{1i} - l_{2j})(w_{1i} \times w_{2j})}{n_1 \times n_2} \quad (4)$$

where L_1 and L_2 are the sets of labels (values) in samples 1 and 2 respectively and n_1 and n_2 are the overall abundances of labels in label sets 1 and 2. w_{1i} and w_{2j} are the abundances of labels l_{1i} and l_{2j} in samples 1 and 2 respectively.

An important consideration when dealing with compositional turnover is that, for sets of locations where each location has the same number of objects, the rate of turnover of objects is not likely to be constant between each pair of locations. For example, locations i, j and k might have the same number of objects but, as the sets of their respective objects differ, the compositional change between these locations will not be consistent (i.e. $\delta_{ij} \neq \delta_{jk} \neq \delta_{ik}$). In such cases the composition cannot be directly displayed or stored as a single surface, making conventional GIS based approaches inappropriate.

For compositional turnover the simplest data structure to use is a matrix of turnover values, where the $n \times n$ rows and columns represent the set of locations used, and each of the values δ_{ij} represents the turnover between each pair of locations i and j . It is from these matrices that correlogram plots are derived (albeit in many implementations the full matrix is never actually populated, as only the elements needed for an iteration are stored during processing).

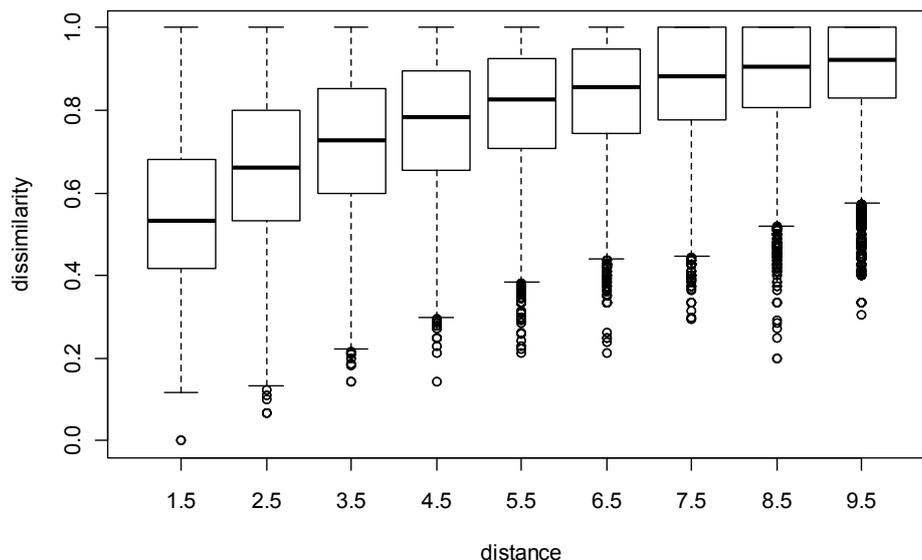


Figure 2. Correlogram of turnover for Australian acacias showing the rate of change of turnover as distance between pairs of locations increases. The x-axis is distance between pairs of locations in decimal degrees, aggregated to 1° increments. The y-axis is the Sorenson dissimilarity metric, where a value of 0 represents identical composition and 1 complete compositional turnover. Data are derived from Laffan and Crisp (2003).

The visualisation of turnover matrices can be complex because, for each location stored in the $n \times n$ matrix, there will be $n-1$ possible values that could be plotted, each of which relates to a different location in the data set to which it is paired. To plot all pairs of values in the matrix simultaneously is difficult. One approach is to plot lines radiating from each location i to each neighbour j , symbolised by their respective value v_{ij} . However, such a diagram rapidly becomes cluttered as n increases, more so when the spatial density of locations is high. Subsetting the lines by direction of most important values is possible, but could potentially exclude multi-directional relationships, while subsetting to distance classes can result in interpreters missing insights where relationships span distance classes.

The simplest solution to the visualisation problem in this case is to create n geographic layers, one per location using values extracted from row n and column n in the matrix. These can then be plotted separately. This is comparatively simple to implement as a post-hoc process and is an established method in biogeography (Kreft & Jetz, 2010). However, an interactive approach is far more user-friendly, and working directly from the matrix is potentially more computationally efficient. This approach also allows one to retain the link with the underlying data, thus allowing one to readily access the set of objects in each location and any associated traits, for example plant height and seed size.

Such an interactive approach has been implemented as an extension to the *Biodiverse* software (Laffan *et al.*, 2010; <http://www.purl.org/biodiverse>). In this system a matrix is calculated and then displayed for one index location, with the remaining locations plotted using their turnover values relative to the index location. When the user clicks on a new location on the map it is set as the index location, and the display is updated to show the turnover from that location to all other locations using values extracted dynamically from the matrix.

An example plot showing four index locations is given in Figure 3, where one can observe the rate of change of *Acacia* species composition at a one degree resolution for a west to east transect through central Australia. Of particular note is that one can readily visualise, for each index location, the spatial scale, anisotropy and non-stationarity of the turnover from that location. This is perhaps most clearly understood by observing how the spatial distribution of lower turnover values (plotted as yellow to cyan) changes as the index location moves through the arid zone and then to the east coast. These patterns can be used to aid in an understanding of the aggregate measures of turnover based on these data, for example in interpreting the correlogram in Figure 2.

3. LIMITATIONS AND COMPUTATIONAL ISSUES

A key limitation of the interactive visualisation is that many matrices will exceed the memory capacity of 32 bit computer architectures, and these are currently the most common platform in desktop computing. The size of the matrix is $n(n-1)$ cells, as the matrices are symmetric so only one value per pair need be stored (typically represented as lower left or upper right of the diagonal). However, sizes still increase rapidly even though half the values need not be stored when using a sparse matrix format, as is the case in *Biodiverse*. For the Acacia example given in Figure 3 there are 708 locations at a 1° resolution and thus a matrix of 310,078 elements. For Australia at a 10 km resolution there will be approximately 4.9×10^9 values in the matrix. The extension to global extents at such resolutions is clearly beyond the capacity of typical desktop systems.

Biodiverse has been developed for complex spatial analyses, so one can also construct spatially constrained matrices such that the sets of neighbours considered for each location can be restricted to a subset of locations around each location, or within a biome, or indeed any arbitrarily complex spatial condition. This can be used to reduce the size of the matrix and thus memory requirements, and is readily stored using the sparse matrix format. An example of such a spatial constraint is given in Figure 4, plotting the rate of turnover of rainfall for Australia derived from the WorldClim data set (Hijmans *et al.*, 2005), with values aggregated at a 1° resolution and scaled using the natural logarithm. In this case the matrix of turnover values has been restricted to include only those pairs of locations that are within a 5° radius of each other. Note the changing orientation of the highest rates of rainfall dissimilarity associated with the tropical and arid zones.

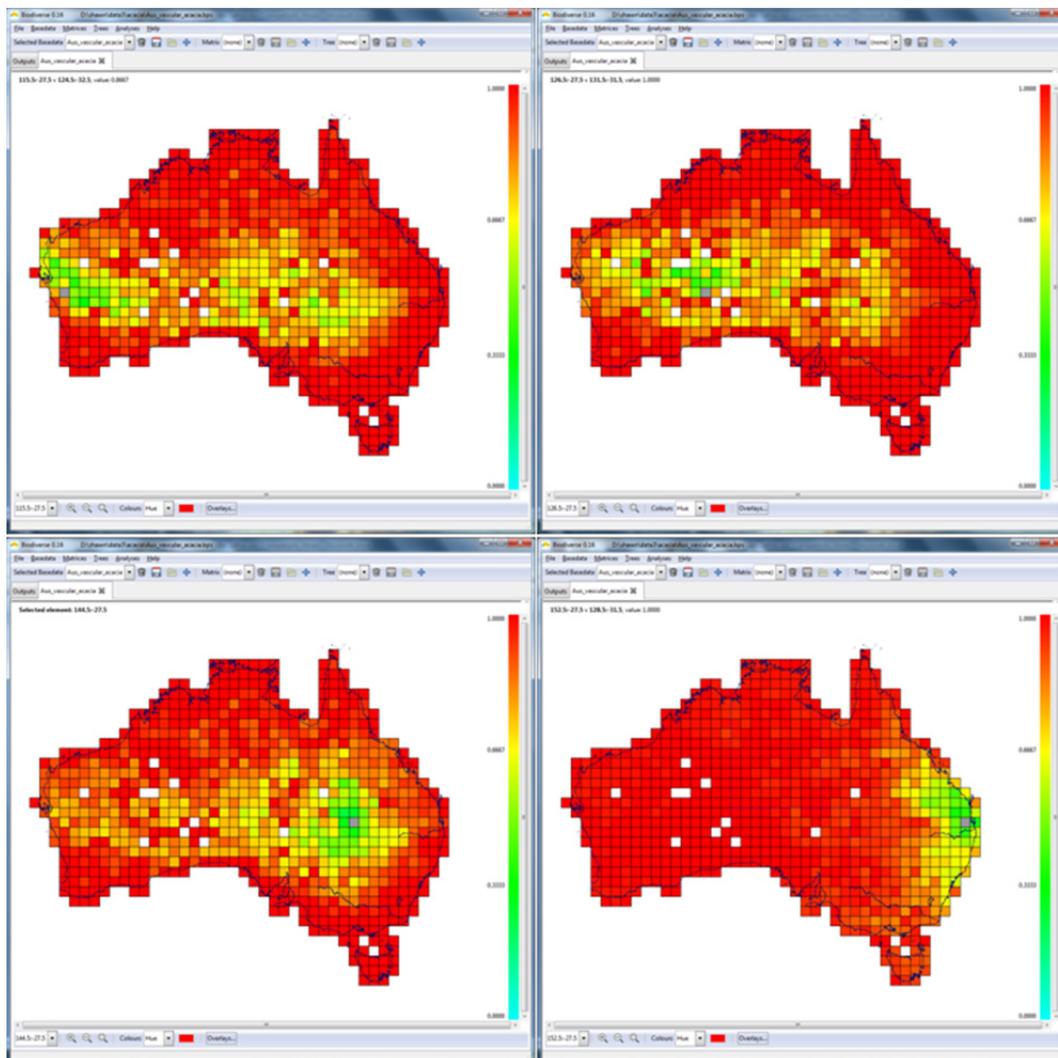


Figure 3. Spatial turnover plots for four cells along a west-east transect for Acacia species in Australia. Turnover is measured using the Sorenson dissimilarity metric. Index cells are denoted in grey. The colour scale progresses from cyan (most similar, value 0) to red (most dissimilar, value 1). White cells contain no data. Data are derived from Laffan and Crisp (2003).

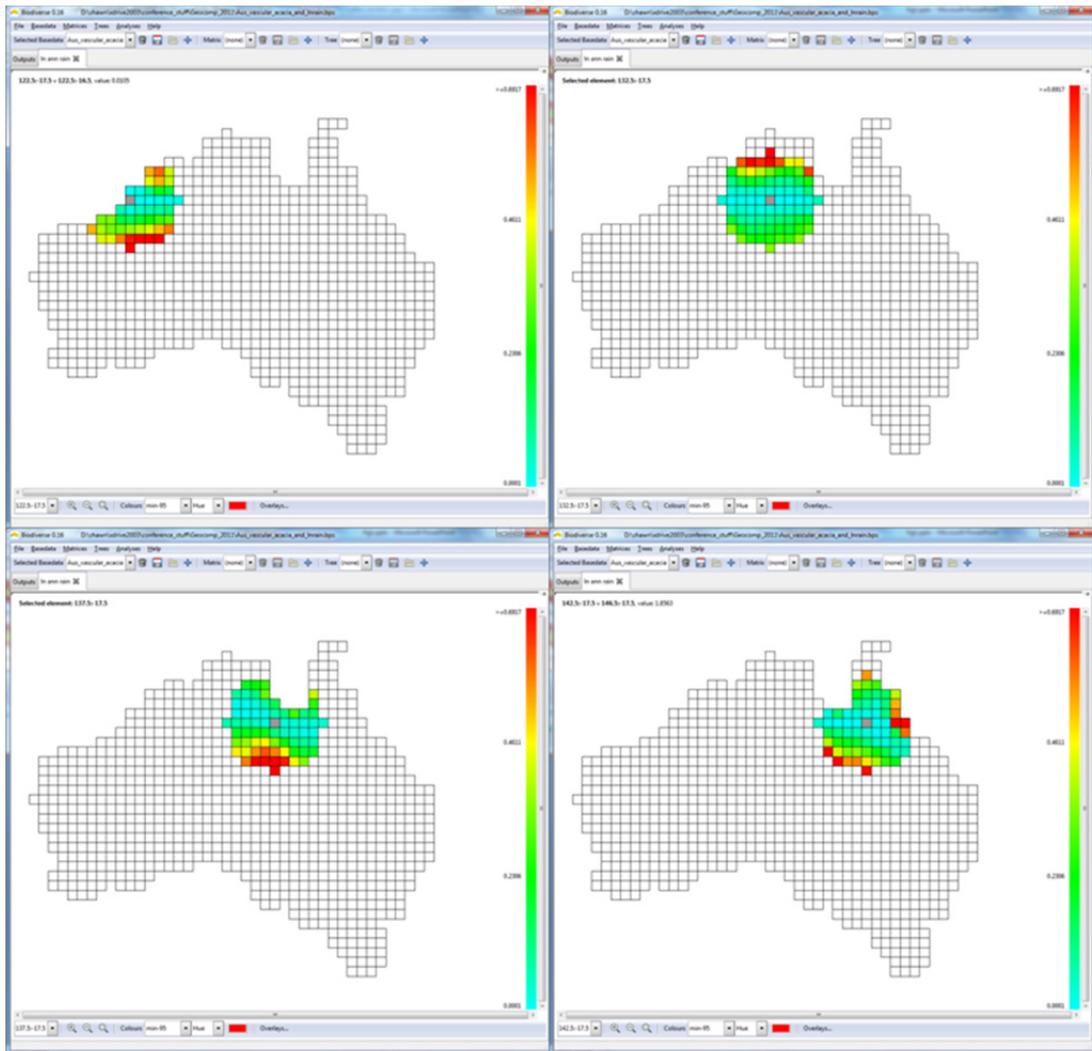


Figure 4. Spatial turnover (mean absolute difference, eq 4) plots for four cells along a west-east transect for the natural logarithm of annual rainfall in Australia. Turnover is measured using the mean absolute difference between each pair of observations in each pair of locations. Index cells are denoted in grey. The colour scale progresses from cyan (most similar) to red (most dissimilar). Only those locations within 5° of the index location are stored and plotted, thus reducing memory requirements. Data are derived from Hijmans *et al.* (2005) and aggregated at a 1° resolution.

4. CONCLUSIONS

Visualisation is an integral part of the data analysis process. For spatial analyses of compositional turnover such visualisation has been difficult due to complexity, large number of spatial associations and computational limitations. An additional cause of this difficulty has been a lack of available software, with GIS software focussing on analyses of individual layers for each object type rather than layers that represent collections of objects. The advancement of the visualisation functionality within *Biodiverse* described here is an important step towards the understanding of spatial turnover patterns in compositional data, particularly for spatial non-stationarity, scale and anisotropy. There is also the potential to use this approach in analyses of distributional breaks and gradational transitions, and of anomalies.

A further advantage of *Biodiverse* software for this application, is that one can apply any one of more than 150 currently supported scalar indices, so one can assess more than simply species turnover, for example species endemism (Laffan & Crisp, 2003), phylogenetic endemism (Rosauer *et al.*, 2009b), phylogenetic turnover, trait data, continuous fields, etc. Hence, one can also visualise relationships underlying assessments of complementarity, which is the additional contribution of locations to an aggregate species richness measure and is an important component of reserve design and assessment (Zhang *et al.*, 2011).

REFERENCES

- Bickford, S. A., Laffan, S. W., de Kok, R. & Orthia, L. (2004) Spatial analysis of taxonomic and genetic patterns and their potential for understanding evolutionary histories. *Journal of Biogeography*, **31**, 1715-1733.
- Ferrier, S., Manion, G., Elith, J. & Richardson, K. (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, **13**, 252-264.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.
- Jones, C. & Laffan, S. W. (2008) Lexical similarity and endemism in historical wordlists of Australian Aboriginal languages of the greater Sydney region. *Transactions of the Philological Society*, **106**, 456-486.
- Kreft, H. & Jetz, W. (2010) A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, **37**, 2029-2053.
- Laffan, S. W. (2002) Using process models to improve spatial analysis. *International Journal of Geographical Information Science*, **16**, 245-257.
- Laffan, S. W. & Crisp, M. D. (2003) Assessing endemism at multiple spatial scales, with an example from the Australian vascular flora. *Journal of Biogeography*, **30**, 511-520.
- Laffan, S. W., Lubarsky, E. & Rosauer, D. F. (2010) Biodiverse: a tool for the spatial analysis of biological and other diversity. *Ecography*, **33**, 643-647.
- Legendre, P. & Legendre, L. (2000) *Numerical ecology*, 2nd English edn. Elsevier Science BV, Amsterdam.
- Rosauer, D., Ferrier, S., Manion, G., Laffan, S. W. & Williams, K. (2009a) Nice weather for frogs - using environmental data to model phylogenetic turnover. *10th International Conference on GeoComputation* (ed. by B.G. Lees & S.W. Laffan), Sydney, Australia.
- Rosauer, D. F., Laffan, S. W., Crisp, M. D., Donnellan, S. C. & Cook, L. G. (2009b) Phylogenetic endemism: a new approach to identifying geographical concentrations of evolutionary history. *Molecular Ecology*, **18**, 4061-4072.
- Tuomisto, H. (2010) A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography*, **33**, 2-22.
- Zhang, K., Laffan, S. W., Ramp, D. & Webster, E. (2011) Incorporating a distance cost in systematic reserve design. *International Journal of Geographical Information Science*, **25**, 393-404.