International Journal of Remote Sensing
Vol. 30, No. 10, 20 May 2009, 2705–2711

Letter

Relating spectral and species diversity through rarefaction curves

D. ROCCHINI*†‡, C. RICOTTA§, A. CHIARUCCI†‡, V. DE DOMINICIS†,
I. CIRILLO† and S. MACCHERINI†‡
†Dipartimento di Scienze Ambientali 'G. Sarfatti', Università di Siena, via P.A. Mattioli
4, 53100 Siena, Italy
‡TerraData *environmetrics*, Dipartimento di Scienze Ambientali 'G. Sarfatti', Università
di Siena, via P.A. Mattioli 4, 53100 Siena, Italy
§Dipartimento di Biologia Vegetale, Università La Sapienza di Roma, Piazzale Aldo
Moro 5, 00185 Rome, Italy

(*Received 21 July 2008; in final form 16 December 2008*)

Rarefaction represents a powerful analytical approach in ecology for estimating the expected number of species within a given study area from local (α-diversity) to regional (γ-diversity) scales. From a landscape perspective, rarefaction curves are directly related to the environmental heterogeneity of the area sampled. The greater the landscape heterogeneity, the greater the expected species diversity. Therefore, remotely sensed images may potentially be used for predicting species diversity through the indirect method of analysing local spectral variation. The aim of this study was to test whether spectral variability can be used as a proxy for species diversity, from local to regional spatial scales. A total of 977 sampling units, each 50 m × 50 m, were selected within the Asciano district (Central Italy) following a stratified random sampling. Each sampling unit was manually classified according to the first level of the Corine Land Cover classification legend. Data on plant species composition were collected in 10 m × 10 m plots located within 98 random sampling units. The normalized difference vegetation index (NDVI) was calculated from a QuickBird image, and quantized into 8-bit data (256 digital numbers, DNs) for building spectral rarefaction curves. Only those plots falling within the QuickBird image were used, which had the effect of reducing the thematic legend to two classes: crops and seminatural vegetation. Species and spectral rarefaction curves were then constructed for each land cover class. Rarefaction curves based on species and spectral properties showed similar results, that is a significantly different number of accumulated values given the same sampling effort for the two classes considered. The results of this study suggest that the shape of the spectral rarefaction curves may be an indirect indicator of environmental diversity, and thus may have potential for predicting biodiversity from local to landscape scales.

1. Introduction

In ecology, accumulation curves have long been used for estimating the expected number of species within a given study area (see Chiarucci *et al.* 2008 for a review). From a statistical viewpoint, as the number of individuals sampled increases, the chance of encountering additional species also increases. Sample-based rarefaction

---

*Corresponding author. Email: rocchini@unisi.it

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2009 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160902755312

curves provide a framework for quantifying expected species richness as a function of sampling effort (Gotelli and Colwell 2001). A series of *N* plots is established across the study area and the presence of all species within each plot is recorded. A rarefaction curve is then obtained by repeatedly resampling the pool of *N* plots at random without replacement and plotting the average number of species represented by 1, 2, …, *N* plots. Thus, sample-based rarefaction generates the expected number of accumulated species as the number of sampled plots increases from 1 to *N*.

As the ordering of samples added to an accumulation curve accounts for its shape (Rocchini *et al.* 2005), an order-free curve should be adopted, derived by means of an analytical solution or Monte-Carlo permutations of samples (e.g. Fattorini 2007, Chiarucci *et al.* 2008). In general, the steeper the curve, the greater the increase in species as the sample size increases (e.g. Ricotta *et al.* 2002).

From a landscape perspective, rarefaction curves are directly related to the environmental heterogeneity of the area sampled. In fact, it is expected that the greater the landscape heterogeneity, the greater the species diversity, including both fine-scale and coarse-scale species richness (i.e. α- and γ-diversity, respectively), and compositional turnover, or β-diversity (Palmer *et al.* 2002). In addition, from a remote sensing perspective, it has long been recognized that digital imagery is a powerful tool for relating environmental heterogeneity to species diversity, as the variation in pixel reflectance of remotely sensed images is generally assumed as a meaningful proxy for both α- and β-diversity (see, for example, Foody and Cutler 2006, Rocchini 2007*a*,*b*). The aim of this study was therefore to test by rarefaction theory whether spectral variability could be used as a proxy for species diversity, from local to regional spatial scales.

2. Study area

The study area is the Asciano district (centroid coordinates: longitude 11°31′03″ E, latitude 43°15′03″ N, datum ED50), an area of 215 km^2 situated in the Crete Senesi landscape, near Siena, Italy (figure 1). Arable lands, dominated by *Triticum aestivum*

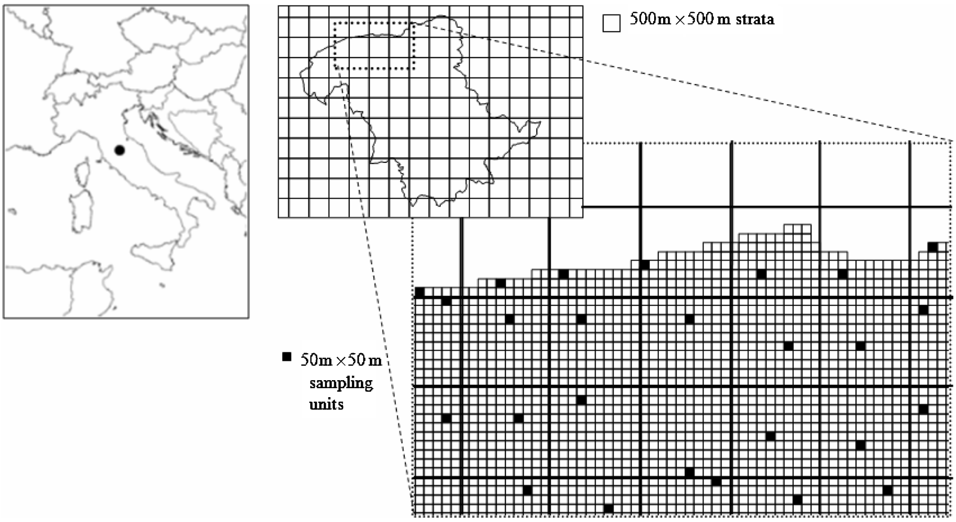


Figure 1. The study area in the Asciano district, Italy, and the stratified random design applied for field sampling.

(common wheat), are interspersed with dissected Pliocene claystone landscapes, also known as badlands, dominated by seminatural vegetation characterized by woods (e.g. *Acer campestre*, hedge maple), shrubs (e.g. *Rubus ulmifolius*, elm-leaved bramble) and herbaceous vegetation (e.g. *Artemisia cretacea*, artemisia).

3. Data and methods

3.1 *Acquiring and preprocessing satellite data*

A QuickBird image of the study site was acquired on 19 July 2004. The image has 2.88 m multispectral pixels (blue, 0.45–0.52 μm; green, 0.52–0.60 μm; red, 0.63–0.69 μm; near-infrared (NIR), 0.76–0.90 μm) and 0.61 m panchromatic pixels. The image was obtained with an 8-bit radiometric depth. The image was orthorectified with a root mean squared error (RMSE) of 0.1 m.

3.2 *Sampling design and field data collection*

A total of 977 sampling units, each 50 m × 50 m, were selected using a stratified random sampling within 500 m × 500 m strata, derived from a UTM (ED50) grid (figure 1). Each 50 m × 50 m unit was attributed with a land cover class based on a photointerpretation of the QuickBird image. The photointerpretation legend was based on the Corine Land Cover first level, and included the following classes: built-up areas (i.e. urban settlements and isolated housing, commercial and industrial buildings, roads), crops (i.e. arable land, vineyards, olive groves, timber plantations), seminatural vegetation (i.e. woodlands, herbaceous vegetation and shrublands, pastures, grasslands, landforms with bare soil such as ‘calanchi’ and ‘biancane’), and water bodies (e.g. lakes, natural or artificial basins, streams and channels).

A random selection of 98 (~10%) of the 50 m × 50 m sampling units were visited during a field survey. For this purpose, one 10 m × 10 m unit (hereafter referred to as a plot) was randomly selected for each of the 98 50 m × 50 m sampling units selected. Vascular plant species composition was recorded within the 98 10 m × 10 m selected plots during June 2004.

To obtain comparable results between species and spectral-based rarefaction curves, only the 58 plots that fell within the satellite image were used in the subsequent analysis. These plots were associated only with the land cover classes ‘crops’ and ‘seminatural vegetation’ (43 and 15 plots, respectively). Thus, the land cover classes ‘artificial areas’ and ‘water bodies’ were not considered in the analysis.

3.3 *Rarefaction curves*

Species rarefaction curves for each land cover class were computed with a Monte-Carlo permutation of samples based on 1000 orderings of samples by means of R software (vegan package; Oksanen *et al.* 2007). Confidence intervals at 99% were calculated to check for statistical differences of rarefaction curves between the classes considered.

The values of the average alpha diversity ($\bar{\alpha}$) and γ-diversity for both land cover classes were derived directly from the rarefaction curve as the minimum and maximum accumulated number of species, respectively. The β-diversity values were derived according to the additive model of diversity partitioning (Lande 1996) as:

$$\beta = \gamma - \bar{\alpha} \qquad (1)$$

Hence, β-diversity could be viewed as the difference between global (γ) and local ($\bar{\alpha}$) diversity.

A similar procedure was used for deriving spectral rarefaction curves for each land cover class (see also the worked example in the next section). It is worth noting that traditional rarefaction methods work only with single variables. Thus, the researcher is forced to select a single band or to reduce the multispectral data set, either by conventional ordination methods such as principal component analysis or by using band combinations. The normalized difference vegetation index (NDVI) was therefore used as the single variable because it provides a potentially optimal discrimination between vegetation types. The NDVI data were quantized into 256 digital numbers (DNs) by the Rmcdr R-package (Fox *et al.* 2007). This 8-bit range was chosen empirically as the best compromise between noise reduction and image feature preservation (see Le Hégarat-Mascle *et al.* 1997). Based on the NDVI DNs in all 10 m × 10 m plots, spectral rarefaction curves for both crops and seminatural vegetation land cover classes were then constructed. To ensure the comparability between the two classes considered, the same number of samples, (n=15), was used for calculating the diversity measures for each land cover class.

3.4 *Worked example of spectral rarefaction*

Consider a sampling unit with pixels from a single band of a satellite image. Two contrasting areas are shown in figure 2, a heterogeneous area (left side of figure 2)

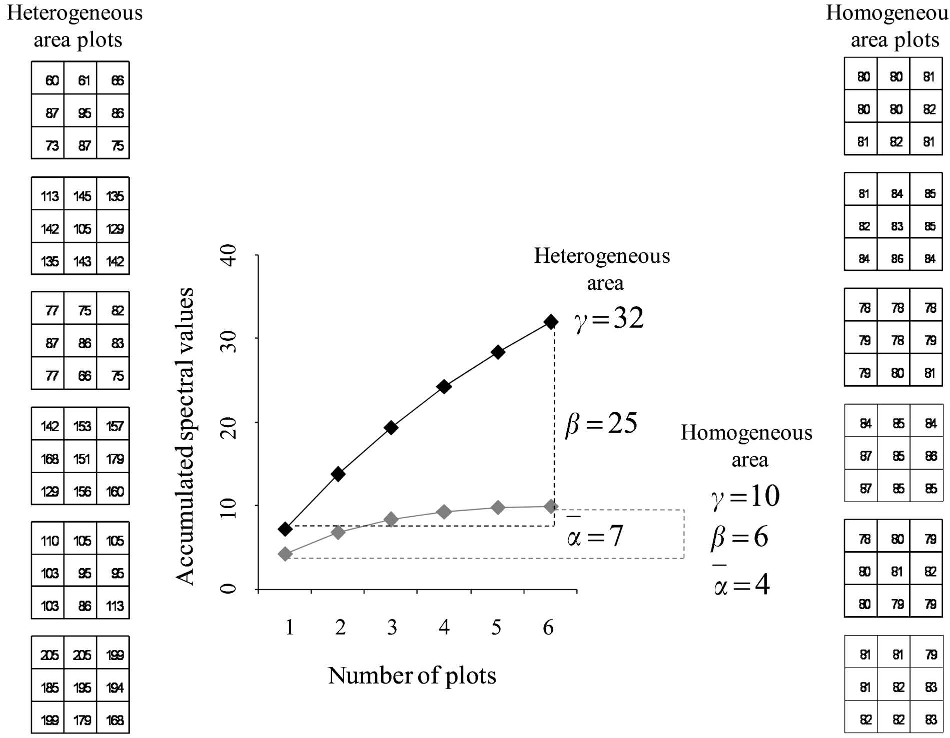
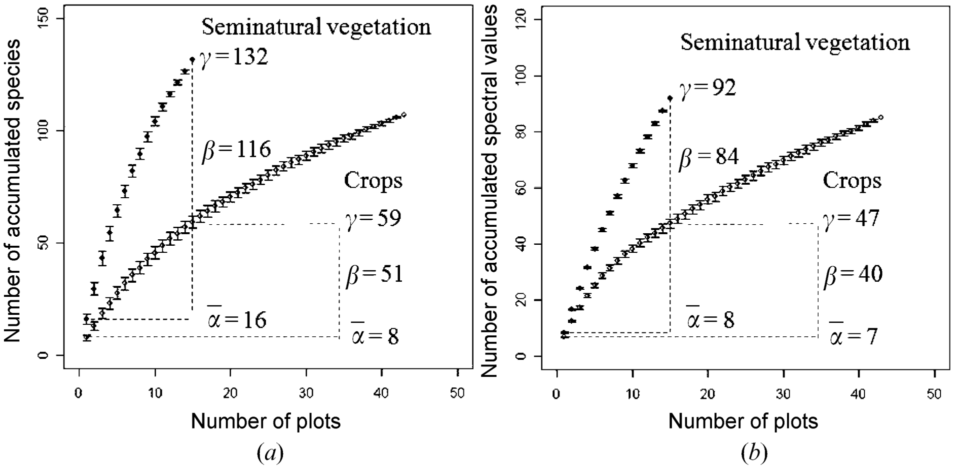


Figure 2. Worked example of spectral rarefaction curves. Black diamonds: heterogeneous area; grey diamonds: homogeneous area. The values reported in the cells represent the input DNs used for building the curves.

and a homogeneous area (right side of figure 2), each covered by six sampling units. For figure 2, $\bar{\alpha}_{\mathrm{DN}}$ equals 7 and 4 for the heterogeneous and homogeneous areas, respectively, that is there are on average 7 and 4 distinct reflectance values for each sampling unit. γ_{DN} is equal to 32 and 10 for the heterogeneous and homogeneous areas, respectively, and the corresponding spectral value turnover, β_{DN}, as calculated by equation (1), is 25 and 6. The diversity measures, $\bar{\alpha}_{\mathrm{DN}}$, γ_{DN} and β_{DN}, can be observed on the rarefaction curves as the minimum, the maximum and the turnover (i.e. maximum – minimum) on each curve (figure 2). In general, the higher the minimum value, the higher the local variability within a plot, and, given the same local variability, the steeper the slope of the curve, the higher the variability across the different plots within the area. However, if the slope is low and the curve rapidly reaches an asymptote, the coarse-scale accumulated spectral values are similar to the fine-scale spectral values, thus indicating a high degree of homogeneity of the area.

4. Results

The species-based rarefaction curves show that, given the same sampling effort, seminatural vegetation hosted a significantly greater ($p<0.01$) number of accumulated vascular plant species than the crops (figure 3(*a*)). Seminatural vegetation hosted double the number of species compared to the crops at the three levels of diversity considered, α, β and γ. The spectral-based rarefaction curves showed similar results in discriminating between the spectral variability of crops and seminatural vegetation (figure 3(*b*)). A notable difference between the species and spectral rarefaction curves is that, at the scale of the single plots, the spectral α-diversity was only slightly different between classes, although the values are significantly different at $p<0.01$. However, starting from the second sampled plot, the two spectral curves diverged rapidly, such that, as for the species-based rarefaction curves, seminatural vegetation was associated with double the β- and γ-diversity compared to crops.



(*a*) (*b*)

Figure 3. (*a*) Species and (*b*) spectral rarefaction curves for seminatural vegetation and crops. Bars: 99% confidence intervals.

5. Discussion

The results of this pilot study suggest that spectral-based rarefaction may be a useful tool to characterize biodiversity across spatial scales. This is because spectral heterogeneity may be an indicator of environmental diversity (Palmer *et al.* 2002, Rocchini 2007*b*).

An important distinction between species and spectral rarefaction curves is that the spectral curve implicitly assumes an upper threshold of coarse-scale diversity ($\gamma=256$ values with 8-bit images). For species rarefaction, no theoretical or empirical asymptote can be assumed *a priori*. However, from an operational viewpoint, this does not hamper spectral rarefaction from potentially identifying those areas or habitats hosting a relatively higher species diversity in terms of both species richness and turnover.

When dealing with diversity estimates, the spatial resolution of remotely sensed images is of great importance. As stressed by Woodcock and Strahler (1987), the image scale should be appropriate for the theory being tested. For instance, in this study local spectral diversity ($\bar{\alpha}_{\mathrm{DN}}$) showed only minor differences between the selected classes. This is possibly due to the poor spatial resolution of the imagery (2.88 m) as compared to the grain adopted for sampling plant species (10 m). A resolution of roughly 3 m implies that, while each field sampling unit can potentially host a large number of species, it cannot contain more than 9 pixels, and thus 9 DNs. This constrains the potential difference between different classes. In fact, when a larger amount of area was accumulated (i.e. more plots were added to the curve), differences between the two land cover classes became apparent. Pan-sharpening, that is resampling multispectral lower resolution data to match panchromatic finer resolution data, may be used to enhance the spatial resolution of satellite imagery. Nonetheless, pan-sharpening should be applied with care because it inevitably leads to a change in the histogram of multispectral channels, thus affecting the statistical properties of an image (van der Meer 1997).

According to Le Hégarat-Mascle *et al.* (1997), any measure of image spatial variability, such as spectral rarefaction curves, depends on the selected coding. For instance, 8-bit images should lead to higher variability than 7-bit images. Nevertheless, as stated previously, the current study focused on the relative differences in variability considering crops *vs.* seminatural vegetation. Performing the same analysis with a 7-bit coding (i.e. 128 DNs, data not shown here) led to similar results.

In this study, satellite images were shown to offer a potentially powerful tool for investigating environmental variability, strictly related to species richness and turnover. Of course, the proposed experiment is preliminary, in that two very different land cover classes were compared. In future work it would be useful to compare multiple land cover classes with a higher degree of uncertainty. Moreover, as a cautionary remark, remotely sensed approaches applied to biodiversity estimates cannot replace field surveys, but are instead tools for supporting more efficient field-sampling strategies.

Acknowledgements

We thank three anonymous referees for insights on a previous version of this manuscript. The support of the Editor Timothy Warner was crucial during the reviewing process. Marco Ferretti, Federica Baffetta, Lorenzo Fattorini and Michela Marignani helped us in developing and managing the field sampling. Giovanni Bacaro provided useful comments.

References

Chiarucci, A., Bacaro, G., Rocchini, D. and Fattorini, L., 2008, Discovering and rediscovering the rarefaction formula in ecological literature. *Community Ecology*, **9**, pp. 121–123.

Fattorini, L., 2007, Statistical inference on accumulation curves for inventorying forest diversity: a design-based critical look. *Plant Biosystems*, **141**, pp. 231–242.

Foody, G.M. and Cutler, M.E.J., 2006, Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks. *Ecological Modelling*, **195**, pp. 37–42.

Fox, J., Ash, M., Boye, T., Calza, S., Chang, A., Grosjean, P., Heiberger, R., Kerns, G.J., Lancelot, R., Lesnoff, M., Messad, S., Maechler, M., Putler, D., Ristic, M. and Wolf, P., 2007, *Rcmdr: R Commander*, R package version 1.2-9. www.r-project.org, http://socserv.socsci.mcmaster.ca/jfox/Misc/Rcmdr/ (accessed 28 November 2008).

Gotelli, N. and Colwell, R.K., 2001, Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, pp. 379–391.

Lande, R., 1996, Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos*, **76**, pp. 5–13.

Le Hégarat-Mascle S., Vidal-Madjar, D., Taconet, O. and Zribi, M., 1997, Application of Shannon information theory to a comparison between L- and C-band SIR-C polarimetric data versus incidence angle. *Remote Sensing of Environment*, **60**, pp. 121–130.

Oksanen, J., Kindt, R., Legendre, P. and O'Hara, R.B., 2007, *Vegan: Community Ecology Package version 1.8-6*. http://cran.r-project.org (accessed 28 November 2008).

Palmer, M.W., Earls, P., Hoagland, B.W., White, P.S. and Wohlgemuth, T., 2002, Quantitative tools for perfecting species lists. *Environmetrics*, **13**, pp. 121–137.

Ricotta, C., Carranza, M.L. and Avena, G.C., 2002, Computing β-diversity from species area curve. *Basic and Applied Ecology*, **3**, pp. 15–18.

Rocchini, D., 2007*a*, Distance decay in spectral space in analysing ecosystem β-diversity. *International Journal of Remote Sensing*, **28**, pp. 2635–2644.

Rocchini, D., 2007*b*, Effects of spatial and spectral resolution in estimating ecosystem α-diversity by satellite imagery. *Remote Sensing of Environment*, **111**, pp. 423–434.

Rocchini, D., Andreini Butini, S. and Chiarucci, A., 2005, Maximizing plant species inventory efficiency by means of remotely sensed spectral distances. *Global Ecology and Biogeography*, **14**, pp. 431–437.

van der Meer F., 1997, What does multisensor image fusion add in terms of information content for visual interpretation? *International Journal of Remote Sensing*, **18**, pp. 445–452.

Woodcock, C.E. and Strahler, A.H., 1987, The factor of scale in remote sensing. *Remote Sensing of Environment*, **21**, pp. 311–332.