Disentangling the role of remotely sensed spectral heterogeneity as a proxy for North American plant species richness

D. Rocchini¹,²,³, L. Dadalt², L. Delucchi¹, M. Neteler¹ and M.W. Palmer²

¹ Fondazione Edmund Mach, Research and Innovation Centre, Department of Biodiversity and Molecular Ecology, GIS and Remote Sensing Unit, Via E. Mach 1, 38010 S. Michele all’Adige (TN), Italy
² Oklahoma State University, Department of Botany, Stillwater, OK 74078, USA
³ Corresponding author. E-mail: ducciorocchini@gmail.com, duccio.rocchini@fmach.it

Keywords: Area effects, Species diversity, Spectral variation hypothesis, Variance partitioning.

Abstract: Due to the difficulties of field-based species data collection at wide spatial scales, remotely sensed spectral diversity has been advocated as one of the most effective proxies of ecosystem and species diversity. It is widely accepted that the relationship between species and spectral diversity is scale dependent. However, few studies have evaluated the impacts of scale on species diversity estimates from remote sensing data. In this paper we tested the species versus spectral relationship over very large scales (extents) with a varying spatial grain using floristic data of North America. Spectral diversity explained a low amount of variance while spatial extent of the sampling units (floras) explained a high amount of variance based on our variance partitioning analyses. This leads to the conclusion that spectral diversity must be carefully related to species diversity, explicitly taking into account potential area effects.

Abbreviations: MODIS - Moderate Resolution Imaging Spectrometer, NDVI - Normalized Difference Vegetation Index.

1 Introduction

Entropy measured by the spatial variation of remotely sensed spectral signal may be a powerful proxy for species diversity at a number of spatial scales (Rocchini 2007). The entropy of the Earth’s surface is closely related to physical and ecological diversity. Because landscape diversity often relates to biological diversity at several ecological levels from community diversity to genetic diversity, measuring entropy is potentially an efficient and relatively cheap way of providing biodiversity estimates over large geographical areas (Gillespie et al. 2008).

Biodiversity can be modeled at different scales using remote sensing. In particular, remote sensing is used to assess habitat heterogeneity, which is positively correlated with species diversity according to the Spectral Variation Hypothesis (Palmer et al. 2000, 2002, Rocchini et al. 2005). Remote sensing is potentially a powerful tool for estimation of species diversity, especially at broad spatial scales, for which field-based data collection is very difficult (Palmer and White 1994, Stanton and Diggle 2013).

Nowadays, the wide availability of image data makes it feasible to study all terrestrial regions of the globe up to a resolution of few meters. These data include satellite-derived images of the earth’s surface with high spatial resolution such as IKONOS, Orbview-3, BGIS-2000 (Ball’s Global Imaging System-2000, RapidEye); high spectral resolution image data from CHRIS (Compact High Resolution Imaging Spectrometer, Hyperion, GLI (Global Imager), MERIS (Medium Resolution Imaging Spectrometer), and MODIS (Moderate Resolution Imaging Spectrometer)); and medium spatial and spectral resolution (Landsat program).

Because species richness in large regions is expected to be higher, as demonstrated in the seminal papers of Arrhenius (1921) and Gleason (1922) on SAR (species-area relationship), the impact of scale on the relationship between spectral and species diversity should be explicitly taken into account when searching for remotely sensed proxies of biological diversity. While some studies exist that consider scale impact on species diversity estimate from remote sensing data (e.g., Oldeland et al. 2010, Lechner et al. 2012), no studies have directly examined area effects over very large scales (extents) using floristic data. Hence, in this paper we used image data from MODIS and selected floras of North America to evaluate the relationship between spectral diversity and species richness at different spatial scales by explicitly taking into account area effects due to differences in area occupied by each flora.

2 Methods

2.1 Study area and the FloraS of North America dataset

North American floras with differing geographic extents were selected for this study from the FloraS of North America project database (http://botany.okstate.edu/floras/, see also Palmer 2005, 2006, Qian et al. 2007, Denslow et al. 2010, Palmer et al. 2012). Based on the FloraS Project, we define
a flora as a list of the vascular plant species documented to occur in a determined area, which is intended to be representative of the area. In other words, a flora represents a concerted effort to record all major groups of vascular plants of the study area in all seasons. Although few authors of floras would regard their studies as 100% complete, their efforts are almost certainly more complete and representative than casual explorations of herbarium databases. A total of 2738 floras were selected for this study from FloraS and we recorded the total number of species, area and geographic location for each one. We used the four extreme geographic coordinates of each flora to draw a bounding box/polygon that was later used to extract the spectral data by means of a routine in GRASS GIS (Neteler et al. 2012).

The study area covered was identified based on the available FloraS data and MODIS Normalized Difference Vegetation Index (NDVI) tiles (Figure 1, Appendix 1 and see section ‘2.2 Modeling’) ranging from ~20°N to ~80°N in latitude and from ~55°W to ~160°W in longitude (Figure 1).

2.2 Modeling

Plant species richness was related to the following macroecological predictors, considering as a unit each polygon containing floras (Fig. 1): the logarithm of the area, the latitude and the spectral diversity. In order to measure spectral diversity we relied on 16-day composite MODIS (Moderate Resolution Imaging Spectroradiometer) Normalized Difference Vegetation Index (NDVI) tiles (date: July 12th-27th 2010) covering the whole study area with a spatial resolution of 250 meters (Fig. 1; rely on http://pymodis.fem-environment.eu/ for MODIS image pre-processing code). NDVI is generally derived as a normalized difference between reflectance values in the near infrared and red bands as:

$$ NDVI = \frac{\lambda_{NIR} - \lambda_{RED}}{\lambda_{NIR} + \lambda_{RED}} $$

where $\lambda_{NIR}$ = reflectance in the NIR part of the spectrum and $\lambda_{RED}$ = reflectance in the red part of the spectrum, NDVI can vary from a theoretical minimum of -1 (minimum reflectance in the NIR and maximum in the red, low biomass) and a theoretical maximum of 1 (maximum reflectance in the NIR and minimum in the red, high biomass). Strictly speaking, NDVI increases at high biomass due to (i) the high reflectance by vegetation in the NIR which is linked to scattering at the leaf scale and (ii) the low reflectance in the red spectrum due to the absorption by chloroplasts for photosynthesis (Huete et al. 2002). MODIS NDVI values are rescaled to a 16-bit radiometric resolution (https://lpdaac.usgs.gov/products/, 216 potential values). Concerning spectral diversity, using the 16-bit coded MODIS NDVI imagery, we relied on ln(spectral richness), namely on the natural logarithm of the number of different NDVI values per each polygon (see Ricotta 2005). Remote sensing data were processed by the free and open source software GRASS GIS (Neteler and Mitasova 2008, 2012).
Neteler et al. 2012; see also Rocchini and Neteler 2012). Appendix 1 provides the complete code used to calculate spectral richness per each (overlapping) floras polygon.

The relationship among predictors was shown by hexagon binning relying on R software (R Development Core Team 2012) using the `hexbin` library (Carr et al. 2011). Refer to Appendix 2 for R based code. In hexagon binning the scatterplot containing the set y versus x is tessellated by a regular grid of hexagons (Lewin-Koh 2011). The main strength of hexagon binning is its potential to show the relation between pairs of variables but also to display the relative density (higher frequency) of particular values, with a gray-scaled legend of hexagons based on the number of points falling within each hexagon in the scatterplot.

We performed univariate correlations between species diversity and each spatial predictor. Hexagon binning was also performed to explore the relationship between each predictor and species diversity.

In order to test the potential of spectral diversity to explain plant species richness, mutually conditional to area and latitude effects (intersection of effects), we performed a variance partitioning analysis using the vegan library (Oksanen et al. 2012) within R (see Appendix 2 for the complete code). Variance partitioning allows calculating the relative single and joint contribution of predictors to the explained variance of a response variable. Figure 2 presents an example with two predictors (X1, e.g., area, and X2). Once a model is generated (e.g., in our case a partial regression analysis) to explain the variability of a response variable Y (e.g., in our case species richness) the final output variance will equal the exclusive variance explained by each predictor together with the variance jointly explained by X1 and X2 and the residual unexplained variance.

Refer to Borcard et al. (1992) or Legendre and Legendre (1998) for the theoretical basis of variance (also said variation) partitioning and to Lobo et al. (2002) for an empirical application.

---

**Figure 2.** Conceptual example of the variance partitioning approach. The variability of a response variable Y (e.g., in our case species richness) is explained by the single and joint contribution of different predictors (X1, area in this case, and X2). The final output variance will equal the exclusive variance of each predictor together with the variance jointly explained by X1 and X2 and the residual unexplained variance.

**Figure 3.** Hexagon binning of the relationship among predictors. Area was correlated with both latitude and spectral diversity, while spectral diversity and latitude were not correlated. Refer to the main text for univariate statistics.
3 Results

Concerning the relationship among predictors, area was correlated with latitude (Pearson R=0.26 using ln(area), p<0.01, N=2738), which is not surprising because the common subjects of floras, parks, provinces and islands are typically larger in the far north (Fig. 1). Spectral diversity was highly correlated with the area of each floras bounding box (Pearson R=0.887 using ln(area), p<0.01, Fig. 3), because more habitats are expected to be present in larger areas, on average (which is analogous to the species-area relationship). Finally, spectral diversity was related to some extent with latitude (Pearson R=0.205, p<0.01, Fig. 3) mainly because of the aforementioned area/latitude relationship.

Species richness was positively correlated with spectral diversity (Pearson R=0.645, p<0.001, N=2738, relying on ln(species richness)) while a high correlation was found with area (Pearson R=0.709, p<0.001, relying on ln(species richness) and ln(area)) following an Arrhenius-based law relation pattern (Fig. 4), with no evident univariate correlation with latitude (Pearson R=0.014, p=0.4695). This is graphically evident in Fig. 4 where hexagon bins are shown (refer to the Methods section for information about hexagon binning and to Appendix 2 for the complete R code).

Applying a variance partitioning approach, the effect of area versus spectral diversity in explaining species richness variation throughout the North American region became apparent (Fig. 5). The amount of variation in richness jointly explained by area and spectral diversity was 43%, demonstrating that it is difficult to disentangle the effects of those variables. The unique contribution of spectral diversity to species richness was only 0.05%.

4 Discussion

Variance partitioning allowed us to estimate the percentage of species richness variance explained by pure spectral diversity and that explained by its combined effects with both area and latitude. Broad-scale biogeographic patterns correlated with latitude have been detected and discussed in previous studies (e.g., Brown and Lomolino 1998, Currie et al. 2004), but were not detected by our variance partitioning

![Figure 4](image)

**Figure 4.** Hexagon binning of the relationship between floras species richness and area (A), latitude, (B), spectral diversity (C).

![Figure 5](image)

**Figure 5.** Venn diagram showing the percentage of variance in species richness explained by the three considered predictors (area, latitude and spectral diversity). The unique contribution of spectral diversity explained only 0.05% of total percentage of variance in species richness. However, 43% of the variation in species richness is jointly explained by heterogeneity and area. Dashes represent values of explained variance lower than 0.001. Notice that ln(area) and ln(species richness) has been used during analysis, see Appendix 2 for complete R code.
analysis. Absence of a latitude effect is probably due to the fact that the latitudinal gradient is only prominent north of ca. 50 degrees in North America where there are fewer floras, Fridley et al. (2006) also discuss this weak latitudinal gradient with the same data.

Concerning the relationship between spectral and species diversity, a low amount of unique variance was explained by spectral diversity per se while a high amount of variance was explained once accounting for area effects. In this view, it is worth noting that the spectral diversity power in predicting species diversity has been proven to depend on the scale (grain) at which it is applied (Palmer et al. 2002).

Furthermore, at broad spatial scales, the geographic extent of the spectral data may overwhelm the role of local heterogeneity patterns in explaining species diversity. From a statistical viewpoint, saturation of spectral diversity for large regions may occur due to rescaling of input imagery, which reduces the uncertainty and information content retained in the image. As an example, Rocchini et al. (2011), working in highly heterogeneous environments (alpine regions) of Switzerland with Landsat ETM+ images (rescaled at 8-bit, 256 values), demonstrated that a saturation of spectral diversity is reached after a certain amount of pixels is accumulated with a lower capability of discriminating among different areas. This is particularly true when considering alpha diversity (e.g., local richness, which was the available data for the present study). On the contrary, beta diversity patterns at local (Feilhauer and Schmidtlein 2009) and broad (He and Zhang 2009) scales have been convincingly linked to spectral turnover.

The decay of local effects (e.g., local heterogeneity) on species richness at broad geographic scales (Legendre et al. 2002, Orlóci et al. 2002) seems to recapitulate the biological reality that parameters such as temperature conditions (Lobo et al. 2002), light/insolation conditions (Hardtle et al., 2003), topographical determinants (Geri et al. 2010), and presence of extreme environments (Culver et al. 2004) are primary drivers of biodiversity on regional and continental scales.

Moreover, it is worth stressing that the area used for measuring spectral variation is generally based on the bounding box containing each flora, namely it is generally larger than the real area of the flora. Further, the bounding box may be extremely spectrally variable outside of the boundaries of the flora itself, because it may include roads, parks, fields, cities, pastures, etc. while the flora itself may cover a spectrally homogeneous area such as a continuous forest or a grassland park. Future work will be done to allow a better matching of spectral signatures with the true boundaries of each flora, possibly including higher image spatial, spectral (including different bands) and temporal resolutions (including images at different seasons).

The spatial variability of NDVI has been used as a proxy of species diversity in a number of i) different habitat types, including tropical dry (Gillespie 2005) and wet (Hernandez-Stefanoni et al. 2009) forests, low-shrub tundra (Gould 2000), boreal forests (Parviainen et al. 2009) and ii) different taxa, from lichens (Waser et al. 2004) to vascular plants (He et al. 2009) to birds (Bino et al. 2008) and mammals (Oindo and Skidmore 2002). However, no comprehensive theory about the so called geometry of heterogeneity yet exists mainly because it is a function of scale with no predictable patterns (but see Palmer 2007, for an attempt at such a theory).

In this paper, we showed that inferences of species richness from spectral diversity may be misleading if area effects are not accounted for, demonstrating an increase in the percentage of variance explained in the estimate of species richness, once heterogeneity and area effects are coupled. Hence, relying only on spectral diversity without considering area effects would be a peculiar case of artifiction (sensu Palmer et al. 2008) in which patterns arise due to incorrect calculations.

Acknowledgements: This paper is part of the FloraS of North America project, partly funded by the US National Science Foundation (NSF, code EPS-0919466). We thank all the persons involved (refer to http://botany.okstate.edu/flowas/). We are grateful to the handling Editor and to two anonymous reviewers who provided useful insights on a previous version of the manuscript. A.J. Harris also provided very helpful suggestions for improving the text. M. Metz provided useful comments about MODIS data while M. Marcantonio provided comments on the code being published. DR is supported by the EU BON (Building the European Biodiversity Observation Network) project, funded by the European Union under the 7th Framework programme, Contract No. 308454 and by the ICT COST Action TD1202 “Mapping and the citizen sensor”, funded by the European Commission.

References


Received May 4, 2013
Revised May 22, 2012
Accepted July 17, 2013

**Supplementary material**

**Appendix 1.** Calculating spectral richness for each floras polygon into GRASS GIS

**Appendix 2.** R code for data processing

The file may be downloaded from the web site of the publisher at www.akademiai.com.