Remote Sensing of Environment 113 (2009) 904–906

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Commentary on Krishnaswamy et al. - Quantifying and mapping biodiversity and ecosystem services: Utility of a multi-season NDVI based Mahalanobis distance surrogate

Duccio Rocchini *

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

ARTICLE INFO

Article history:
Received 31 December 2008
Received in revised form 30 January 2009
Accepted 31 January 2009

Keywords:
Biodiversity
Distance decay
Habitat heterogeneity
Quantile regression
Spectral distance

ABSTRACT

Remote sensing is a powerful tool for characterizing, estimating or modelling species diversity. Differences in environmental properties of different habitats should lead to differences of spectral responses, which can be detected by satellite imagery. Hence, spectral distance may be related to species diversity. Based on previous studies, Krishnaswamy et al. [Krishnaswamy, J., Bawa, K. S., Ganeshaiah, K. N., & Kiran, M. C. (2009). Quantifying and mapping biodiversity and ecosystem services: Utility of a multi-season NDVI based Mahalanobis distance surrogate. Remote Sensing of Environment.] used spectral distance to estimate species diversity. Since a noisy scatterplot of species versus spectral diversity is expected, the commonly used Ordinary Least Square regression may fail to detect trends which occur across other quantiles than the mean. Krishnaswamy et al. [Krishnaswamy, J., Bawa, K. S., Ganeshaiah, K. N., & Kiran, M. C. (2009). Quantifying and mapping biodiversity and ecosystem services: Utility of a multi-season NDVI based Mahalanobis distance surrogate. Remote Sensing of Environment.] proposed a quantile–quantile plot method as an alternative to conventional regression based approaches which are inappropriate for dependent pair-wise dissimilarity or similarity data. By this commentary I demonstrate the utility of a quantile regression technique to complement the Krishnaswamy et al. [Krishnaswamy, J., Bawa, K. S., Ganeshaiah, K. N., & Kiran, M. C. (2009). Quantifying and mapping biodiversity and ecosystem services: Utility of a multi-season NDVI based Mahalanobis distance surrogate. Remote Sensing of Environment.] graphical approach in terms of a predictive model.

In a recent paper, Krishnaswamy, Bawa, Ganeshaiah, and Kiran (2009-this issue) investigated the potential of using remote sensing for explaining plant species compositional turnover, also referred to as beta-diversity. In particular, spectral (rather than spatial) distance was used to model the decay of species similarity. It is expected that the higher the spectral distance the higher the species diversity among sites. Hence, the use of spectral distance has enormous advantages.

From a biological point of view, spatial distance generally acts as an ecological limiting factor accounting for the dispersal of both plant and animal species (Chust et al., 2006).

Quoting Rocchini (2007) and Krishnaswamy et al. (2009) "methods based on distance decay do not necessarily account for environmental heterogeneity, especially in heavily fragmented landscapes (Palmer, 2005)". Meanwhile, in interspersed landscapes, spatial distance may not account for species diversity since there will be a high species turnover even at very local spatial scales with low spatial distance.

Krishnaswamy et al. (2009) acknowledge that the achieved pattern dates back to Tuomisto et al. (2003a) and Rocchini (2007) who modeled species similarity decay by increasing spectral distance. Of course, Krishnaswamy et al. (2009) reinforced the idea with new empirical data and habitat types. In particular, in Fig. 1 I simulate the pattern found by Krishnaswamy et al. (2009, see their Fig. 3) with species similarity decaying with spectral distance. This noisy pattern may have several causes: (i) the mismatch between the grain of the Landsat ETM+ image and the field data, (ii) the difference between the time of satellite image acquisition and the field survey period, (iii) the grain of sampling units, etc. Considering the dimension of sampling units, if grain is small enough, one might expect that samples should share no or few species, even if their ecological properties are the same (Nekola & White, 1999) thus provoking a high amount of points within the scatterplot with low species similarity even when spectral distance is small.

Krishnaswamy et al. (2009) provide a useful quantile–quantile plot method as an alternative to conventional regression based approaches which are inappropriate for dependent pair-wise dissimilarity or similarity data. Here I formally demonstrate the utility of a quantile

* Tel.: +39 0577 235415; fax: +39 0577 232896.
E-mail address: rocchini@unisi.it.

0034-4257/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2009.01.014

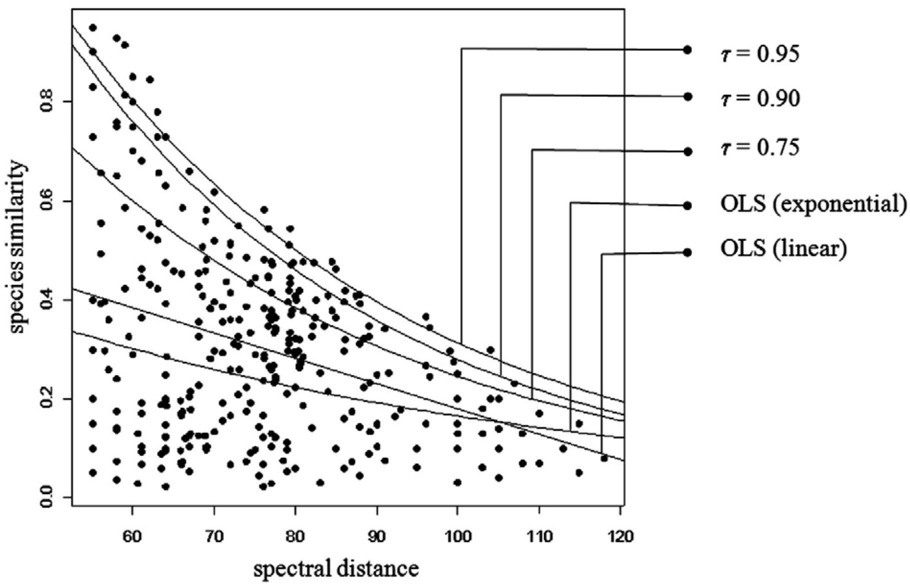

Fig. 1. Decay of species similarity versus spectral distance as simulated from Krishnaswamy et al. (2009), who used species dissimilarity versus spectral distance. Here I am using species similarity in order to better show distance decay, being consistent with previous work on the matter (see Tuomisto, 2003b and Rocchini, 2007).

regression technique to complement the Krishnaswamy et al. (2009) graphical approach in terms of a predictive model.

In Least Square Regression analysis, the residual sum of squares is minimized within the regression model. However, ecological datasets, and in particular those related to plant communities as in this case, are often characterized by a high amount of noise whose causes have been previously disentangled.

In these cases, attention should be focused on other quantiles than the mean in order to detect trends which may be lost by using OLS. By this commentary I provide for an alternative regression-based method for accounting for the noise in a distances—scatterplot, based on quantile regression (Koenker & Hallock, 2001; Koenker & Bassett, 1978; Cade & Noon, 2003; Cade, Noon, & Flather, 2005).

Fig. 1 shows a scatterplot of species similarity as measured by the Sørensen coefficient (as in Krishnaswamy et al., 2009, hereafter ϖ_y) versus spectral distance (hereafter ϖ_x), referred to as eco-climatic distance by Krishnaswamy et al. (2009). ϖ_x and ϖ_y appear to be related by a negative exponential model.

However, when fitting with an OLS-based model, the best fitting is found to be a linear model (adjusted $R^2=0.1195$, slope $=-0.005$, $p<0.01$) rather than an exponential one (adjusted $R^2=0.068$, slope $=-0.014$, $p=0<0.01$). Conclusions from the use of the "mean" of the input data would therefore be that: (i) variables are weakly related to each other with a very low slope, and (ii) variables are related to each other by a linear rather than an exponential function.

Nonetheless, when considering the upper trend of values (considering different quantiles τ), the exponential relation between ϖ_x and ϖ_y become apparent (slope $=-0.022$, $p<0.01$, $\tau=0.75$) reaching higher values at $\tau=0.90$ and 0.95, with slope equalling -0.024 and -0.025 ($p<0.01$), respectively.

Thus, a variable ϖ_y may show different response patterns with respect to ϖ_x when considering different subsets $S\varpi_y$. A method considering different responses with respect to each $S\varpi_y$ is strongly recommended to fully understand the phenomenon under study. Obviously the same reasoning may be applied considering lower trends in the data.

Formally, let $\{\varpi y_1, \varpi y_1, \ldots, \varpi y_n\}$ denote the y values of a set of points of the variable ϖ_y. OLS regression minimises residuals by solving:

$$\text{residual} = min \sum (\varpi y_i - \hat{\varpi} y_i)^2 \quad (1)$$

where $\hat{\varpi} y_i$ = estimated value. Once the residual of each ϖy_i value with its corresponding $\hat{\varpi} y_i$ value is calculated, the model is fitted according to the minimisation of symmetrical values.

Giving different weights to positive and negative residuals leads to an asymmetric minimisation of residuals such that:

$$\text{residual} = min \sum |\varpi y_i - \hat{\varpi} y_i| T \quad (2)$$

where T is a multiplier term equalling τ (the quantile value) for positive deviations (ϖy_i–$\hat{\varpi} y_i$) and $(1-\tau)$ for negative deviations. This asymmetric minimisation fits a regression model through the upper data for high τ and through the lower data for small τ.

It should be noted that the quantile minimisation of residuals (Eq. (2)) is based on absolute values rather than on squared deviations as in OLS regression, thus reducing outlier effects. For a more detailed description of quantile-based fitting, see Koenker & Hallock, (2001) and Gotelli & Ellison (2004).

Of course, quoting Krishnaswamy et al. (2009), "As [the] pair-wise distances are not independent of each other, and the number of data points are very large, conventional statistical methods (e.g. regression, linear or non-linear with specific parametric distributional assumptions for the errors) are not possible". However, permutation (Cade, Richards, & Mielke, 2006) or bootstrap (Koenker, 2007) based methods should solve the problem.

I think that the Krishnaswamy et al. (2009) paper will be of great interest to the remote sensing and biodiversity research community. In this view, I acknowledge their claim:

> "Overall, we have demonstrated that we can use multi-date remotely sensed data to quantify forest type variability on a single numerical scale and use this approach to detect broad scale patterns of bio-diversity and ecosystem services that can be used in conservation planning".

References

Cade, B. S., & Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, *1*, 412−420.

Cade, B. S., Noon, B. R., & Flather, C. H. (2005). Quantile regression reveals hidden bias and uncertainty in habitat models. *Ecology*, *86*, 786−800.

Cade, B. S., Richards, J. D., & Mielke, P. W., Jr. (2006). Rank score and permutation testing alternatives for regression quantile estimates. *Journal of Statistical Computation and Simulation*, *76*, 331−355.

Chust, G., Chave, J., Condit, R., Anguilar, S., Lao, S., & Pérez, R. (2006). Determinants and spatial modeling of tree β-diversity in a tropical forest landscape in Panama. *Journal of Vegetation Science*, *17*, 83−92.
Gotelli, N., & Ellison, A. (2004). *A Primer Of Ecological Statistics.* Sunderland, Massachusets, USA: Sinauer Associates.
Koenker, R. (2007). quantreg: Quantile Regression. R package version 4.10. http://www.r-project.org
Koenker, R., & Bassett, G., Jr. (1978). Regression quantiles. *Econometrica*, *46*, 33−50.
Koenker, R., & Hallock, K. (2001). Quantile regression. *Journal of Economic Perspectives*, *15*, 143−156.
Krishnaswamy, J., Bawa, K. S., Ganeshaiah, K. N., & Kiran, M. C. (2009). Quantifying and mapping biodiversity and ecosystem services: Utility of a multi-season NDVI based Mahalanobis distance surrogate. *Remote Sensing of Environment*, *113*, 857−867.
Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, *26*, 867−878.
Palmer, M. W. (2005). Distance decay in an old-growth neotropical forest. *Journal of Vegetation Science*, *16*, 161−166.
Rocchini, D. (2007). Distance decay in spectral space in analysing ecosystem β-diversity. *International Journal of Remote Sensing*, *28*, 2635−2644.
Tuomisto, H., Poulsen, A. D., Ruokolainen, K., Moran, R. C., Quintana, C., Celi, J., et al. (2003a). Linking floristic patterns with soil heterogeneity and satellite imagery in Ecuadorian Amazonia. *Ecological Applications*, *13*, 352−371.
Tuomisto, H., Ruokolainen, K., & Yli-Halla, M. (2003b). Dispersal, environment, and floristic variation of Western Amazonian forests. *Science*, *299*, 241−244.