MAPPING OF ALGAE RICHNESS USING SPATIAL DATA INTERPOLATION

F.O. Tapia-Silva, [a, *] , O.E. Hernández-Cervantes [b], M.I. Vilchis-Alfaro [b], A. Sentíes [a], K.M. Dreckmann [a]

[a] Departamento de Hidrobiología, Universidad Autónoma Metropolitana-Iztapalapa. A.P. 55-535, México, D.F. 09340, México. (otapia, asg, tuna)@xanum.uam.mx
[b] Maestría en Biología. Universidad Autónoma Metropolitana-Iztapalapa. A.P. 55-535, México, D.F. 09340, México. (anfibio_io, isa_bell80)@hotmail.com,

THEME: Marine and Coastal Environment, resources and dynamics

KEY WORDS: Algae richness, world wide coastal mapping, kriging interpolation, Gulf of Mexico

Abstract

This work describes the generation of a database of algal species richness at two spatial scales – regional (Gulf of Mexico and the Caribbean) and global (coastal zones). As a first approach to the definition of the temporal variability, and to produce the corresponding maps, a previously published decision tree is used in order to select the best spatial interpolation technique according to the characteristics of the spatial data. The methods presented are ordinary Kriging (since no relationship exists with any environmental variable that could function as an external variable) and inverse distance squared, for comparative purposes. The methods to generate the spatial layers are evaluated using the leave-one-out cross validation technique. Although the evaluation did not find a large correspondence (in terms of linear regression) between the interpolated and measured values, it was possible to capture the spatial variability of the process and produce the cartography of this variable, with which future ecological analyses can be performed..

Introduction

Species richness can be considered to be the simplest way to measure biodiversity (Krebs, 1978) and the results can be analyzed using several methods. One such method is the generation of maps, which have been important to describe both populations and species found in habitats and areas. Maps have also contributed to interpreting heterogeneous data related to the specific presence along diverse environmental gradients. Marine algae are no exception to this approach, since the mapping of this parameter has greatly helped to identify the possible environmental and historical factors that explain the patterns of algal diversity both regionally and globally.

Investigations about the mapping of species richness are scarce and have been dedicated to identifying variations in the measurements of diversity based on an environmental gradient (eg. latitudinal gradient) (Santelices & Marquet, 1998). Meanwhile, the development of geographic information systems has led to the implementation of certain predictors in order to develop algal diversity maps. One method which has been used to predict richness is to determine whether a correlation exists between environmental variables and macroalgae (Keith et al 2014). Another important method to generate maps of the distribution of algal diversity is to obtain the distribution intervals of algae genera or species and thereby determine the value of the total number of intervals that are overlapped in each quadrant or site, using the inverse distance weighted interpolator (IDW) to estimate this parameter (Kerswell, 2006).

The classical statistical methods that have been used to estimate species richness ignore the fact that the spatial distribution of species richness presents a continuity and a dependency pattern at the spatial level. This is due to the presumptions about stationarity in space and time, independency among data and the identical distribution of the parameters. Nevertheless, these presumptions are not always accurate.

Geostatistical methods have been applied to these types of cases, which consider dependency among contiguous geographic units based on the value of the species richness (Crúz-Cárdenas et al. 2013).

Therefore, even though some methodological approaches exist to map patterns of algal species richness, alternatives that are based on different principles are needed. One example is the kriging method pertaining to geostatistics, given that the application of geographic information systems to the study of the diversity of these organisms is fully developed.

One of the ways in which this tool is useful to biology is to identify areas that are of interest because of their high or low diversity. As a consequence of the identification of hotspots present in the maps generated, potential areas for the conservation of species or endemism can be evaluated (Escalante, 2003), as well as the potential distribution of a particular species. Important information can also be obtained to determine the coastal zones that need a better evaluation of algal diversity. It is worth mentioning another factor to be considered when conducting bio-geographical investigations—the distribution patterns of species richness (Morrone 2009).

It is important to mention that these methods may operate differently depending on the geographic scale and the taxonomic level of the study (Kerswell, 2006). Thus, the scales and biological models used in the present work are: (1) on a global scale, the Laurencia genus which includes the red algal species and presents a tropical, subtropical and template distribution interval (Sentíes & Fujii, 2002); and (2) on a regional scale, large macroalgae on the coasts of the Gulf of Mexico and the Mexican Caribbean (GMC), which is a biological group that may represent a potentially important resource for the country (Dreckmann & Sentíes, 2013).

The use of ordinary kriging (OK) enables taking into account local variations in the mean, limiting the stationarity domain to a local area Ω around the position x where the variable will be estimated.

The expression $Z(x)=Y(x)+m(x)$ represents a stochastic process with a variable mean $m(x)$ and covariance function $C(h)$. As such, $Y(x)$ is a stochastic process with a null mean. A linear estimator is a linear combination of the measurements $Z(x_1),\ Z(x_2),\ldots,\ Z(x_n)$ at positions $x_1,\ x_2,\ldots,x_n\in\Omega$. Specifically:

$$\widehat{Y}(x)=\sum_{k=1}^{n}\lambda_k(x)Y(x_k)$$

or:

$$\widehat{Z}(x)=m(x)+\sum_{k=1}^{n}\lambda_k(x)\big(Z(x_k)-m(x_k)\big)$$

If the mean is constant in domain Ω, then the above equation can be eliminated by forcing the sum of the kriging weights λ_k to equal 1. In such a case, the estimator is called ordinary kriging and is expressed as:

$$\widehat{Z}_{KO}(x)=\sum_{k=1}^{n}\lambda_k(x)Z(x_k)$$

with

$$\sum_{k=1}^{n}\lambda_k(x)=1$$

The optimal weights that minimize the variance of the estimation error are obtained using the Lagrange multiplier method (Goovaerts, 1997, pp. 133), which results in the following system of equations:

$$\begin{cases}\displaystyle\sum_{k=1}^{n}\lambda_k(x)C(x_j-x_k)+\mu(x)=C(x_j-x),\quad j=1,\ldots,\\[2mm]\displaystyle\sum_{k=1}^{n}\lambda_k(x)=1\end{cases}$$

where μ denotes the Lagrange multiplier. Alternatively, if the relation between the covariance function and the semivariogram function $\gamma(h)$ is considered that is, $C(h)=C(0)-\gamma(h)$, . This system can be written as:

$$\begin{cases}\displaystyle\sum_{k=1}^{n}\lambda_k(x)\gamma(x_j-x_k)+\mu(x)=\gamma(x_j-x),\quad j=1,\ldots,n\\[2mm]\displaystyle\sum_{k=1}^{n}\lambda_k(x)=1\end{cases}$$

Method

The species richness values of the Laurencia genus were obtained at the global level by consulting the primary literature that has reported on the species of this genus (eg. taxonomic monographs, floristics catalogs, species checklist and isolated records) as well as the AlgaeBase online (Guiry & Guiry 2015). The sites registered were georeferenced and the taxonomic

validity of the species was verified. The compilation of these reports resulted in the specific richness value (number of species) for each one of the localities. A total of 130 species were reported in a total of 501 localities worldwide, with values from 1 to 29 per locality.

With respect to the regional scale (GMS), a checklist was made of the large macroalgae species (macroalgae sizes 10 to 100 cm) present in the study area. These sizes were selected based on the Littler & Littler (2000) identification guide. A database was built with these data and the localities where these organisms have been reported, their georeferencing and the specific richness of each one. This process was primarily based on specialized catalogues of the study region produced by Ortega et al. (2002) and Dreckmann (1998) and complemented with information from registries reported in the Algaebase database (Guiry & Guiry, 2015) and various recent publications. A total of 110 species were included in the database, which represented large macroalgae with valid taxonomies pertaining to the classes Rhodophyta, ChlAropyta and Ochrophyta. They were distributed throughout the 118 localities and their values ranged from 1 to 87 species in each one.

A decision tree was used to select the spatial analysis technique (Hengl, 2009). First, a determination was made as to whether the variable of interest had a linear relationship with an environmental variable. The variables tested included surface temperature obtained from a MODIS sensor. When no relationship was found, an interpolation procedure was performed using ordinary kriging, since it was possible to estimate a semivariogram of the empirical data. A mechanical interpolation method was also performed using the inverse distance squared method for comparative purposes. The results were evaluated with a leave-one-out cross validation.

Results and Discussion

Next, the results from the two exercises performed at different scales are presented. As mentioned, the decision tree model by Hengl (2009) was followed. No significant correlation was found between the species richness data corresponding to the Laurencia genus and marine surface temperature (annual, January and July). The same findings were obtained with the species richness data for the GMC region. Nevertheless, when fitting the semivariogram to an experimental model (Figures 1a, 1b), a spatial correlation of the data was identified for both cases.
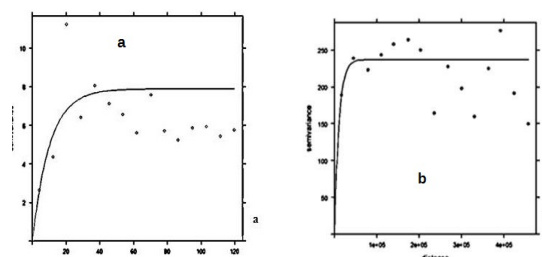

Figure 1. Semivariogram fitted to the total species richness for a) Laurencia genus and b) large macroalgae in the Gulf of Mexico and the Caribbean..

The cross validation analysis of the values obtained by the two interpolation methods showed a significant correlation (Table 1) between the known and predicted richness values. Although the

values of R2 were low for a spatial analysis (Table 1) they were similar for the two interpolators at both scales, having achieved a good capture of the spatial variability of the species richness. This can be seen in the definition of the richness pattern in the maps generated (Fig. 2,3).

	Global scale			Regional Scale		
	Signifi-cance	R^2	RMSE	Signifi-cance	R^2	RMSE
IDW	$p<0.05$	0.04	2.65	$p<0.05$	0.08	14.23
OK	$P<0.05$	0.04	3.14	$p<0.05$	0.08	14.12

Table 1. Significance values, R^2 and RMSE for the IDW and OK interpolators at both scales.

With the spatial autocorrelation according to the decision tree by Hengl (2009), it would have been necessary to estimate the species richness using the OK interpolation. Nevertheless, the results were very similar when applying the IDW method. At both the global and regional levels, the two methods also identified the same locations as having the largest number of species (Fig. 2,3).

One difference between the results from the two interpolators is seen in the estimated size of the areas that contain the highest species richness. For example, for the GMC, the area estimated using the IDW interpolator was larger than with the OK (Fig. 3a, 3b), whereas the opposite occurred with the interpolation at the global scale (Fig. 2a,2b). For the GMS, differences in the IDW and the OK estimations of low species richness also exist, in which the former estimated large areas with values near zero (Fig. 3b) and the latter resulted in values near the middle interval (8 – 15 species) for these areas (Fig. 3a). This may indicate a possible underestimation of the diversity on the part of the IDW and an overestimation by the OK.

Did the analyses using OK and IDW function in this study, even though the values were low for a spatial analysis? The answer to this question is "yes" given that the maps obtained adequately represent distribution patterns that can be interpreted by ecological (environmental) and historical (geological) factors. Nonetheless, it is necessary to take into account that a characteristic of the distribution of these organisms is that they inhabit coastal zones, which can influence the analysis and indicates that this distribution is represented more by a coastline than by an area. This disadvantage can be added to the discussion about the geographic unit used in the spatial analysis of diversity. According to Murguía (2005), the scale can create a problem for defining the geographic unit. In light of this study, it can be said that the distribution of the organisms is highly important to the definition of the geographic unit.

Conclusions

The estimators were determined to be significant for both scales and interpolators (IDW and OK). They were similar in terms of capturing the distribution of the species richness. Therefore, in this case, OK was as effective as IDW. Geostatistical methods can continue to be tested by studies similar to the present work through improvements such as generating a more detailed database of algae richness. Other environmental variables can also be tested that may be associated with the variable of interest, such as: phosphate concentrations, turbidity, salinity, pH, nitrates,

oxygen levels and dissolved carbon dioxide levels, type of sediment and calcite concentrations. It is very important to evaluate correlations with these variables based on temporal scales or average levels.
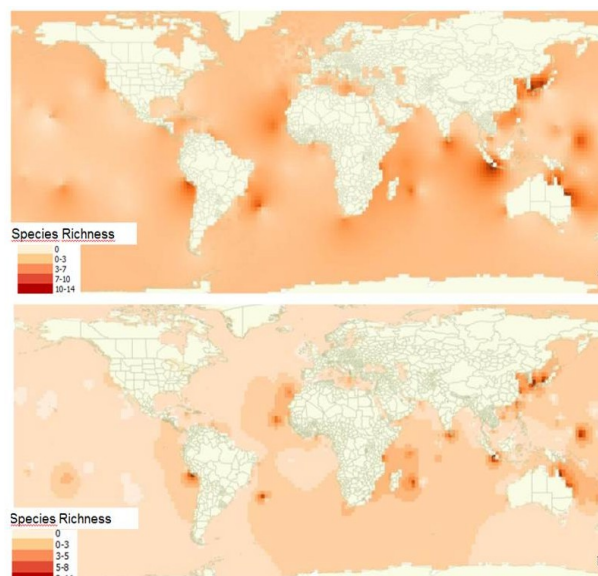

Figure 2. Spatial distribution maps of species richness for the Laurencia genus at the global level. A) Estimation with OK interpolator and B) with IDW.
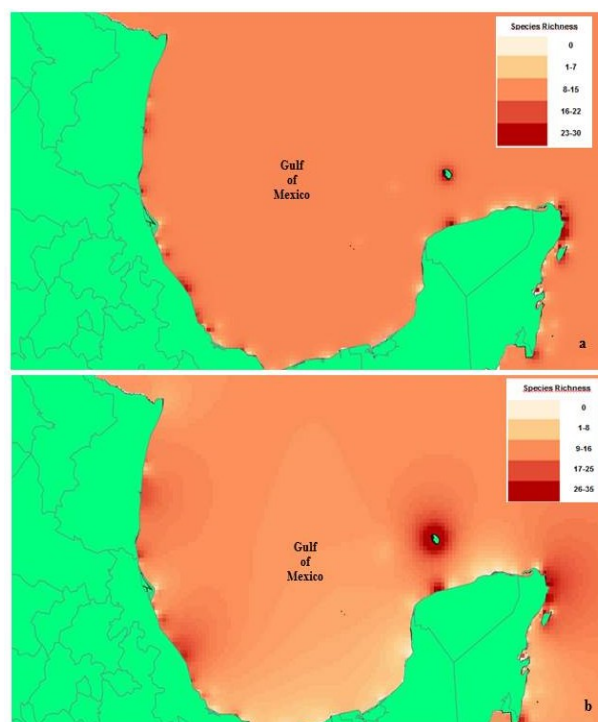

Figure 2. Spatial distribution maps of species richness for large macroalgae in the Gulf of Mexico and the Mexican Caribbean. A) Estimation with OK interpolator and B) with IDW.

It is worth mentioning that the design of the analysis should not only consider the scale of the geographic unit but also the way in which the organisms are distributed in an area. Furthermore, how the biological model in question is distributed must also be considered since they can be spread over areas or along coastlines, as in the case of microalgae. Continuing to perform floristic studies will enable improving estimations of the spatial variability of the species richness of these organisms. In turn, this will contribute to future bio-geographic studies and will serve as a useful tool to make decisions about the conservation of algae diversity.

Bibliography

Cruz-Cárdenas, G., J.L. Villaseñor, López-Mata L. & E. Ortíz, 2013. Distribución espacial de la riqueza de especies de plantas vasculares en México. Revista Mexicana de Biodiversidad 84: 1189-1199.

Dreckmann, K.M. 1998. Clasificación y Nomenclatura de las Algas Marinas Bentónicas del Atlántico Mexicano. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. México, D.F.

Dreckmann, K.M. & A. Sentíes, 2013. Los arribazones de algas marinas en el caribe mexicano: evento biológico natural o basura en las playas. CONABIO. Biodiversitas 107: 7-11.

Escalante, T., 2003. Determinación de prioridades en las áreas de conservación para los mamíferos terrestres de México, empleando criterios biogeográficos. Anales del Instituto de Biología. Universidad Nacional Autónoma de México. Serie Zoología 74(2): 211-237

Guiry M.D. & G.M. Guiry, 2015. Algaebase. World-wide electronic publication. National University of Ireland, Galway. http://www.algaebase.org

Goovaerts P. 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press. Pp. 133

Hengl, T., 2009. A practical guide to geostatistical mapping. Office for Official Publications of the European Communities, Luxembourg.

Keith, S.A., A.P. Karswell & S.R. Connolly, 2014. Global diversity of marine macroalgae: environmental condition explain less variation in the tropics. Global Ecology and Biogeography 23: 517-529.

Kerswell, A.P., 2006. Global biodiversity patterns of benthic marine algae. Ecology 87 (10): 2479-2488.

Krebs, C.J., 1985. Ecologia. Estudio de la distribución y la abundancia. Harper & Row Publisher Inc. New York.

Littler, D.S. & M.M. Littler. 2000. Caribbean reef plants. An identification guide to the reef plants of the Caribbean, Bahamas, Florida and Gulf of Mexico. Offshore Graphics. Washington.

Morrone, J.J., 2009. Evolutionary Biogeography. An integrative approach with case studies. Columbia University Press, New York.

Murguía, M., 2005. La estadística espacial como herramienta de análisis de la biodiversidad. pp. 63-71. In: Halffer, G., J. Soberón, P. Koleff & A. Melic (eds). Sobre diversidad biológica: significado de las diversidades Alfa, Beta y Gamma. m3m-Monografías 3ercer Milenio, vol.4. SEA, CONABIO, Grupo Diversitas & CONACyT, Zaragoza.

Ortega, M.M., J.L. Godínez and G. Garduño Solórzano. 2001. Catálogo de algas bentónicas de las costas mexicanas del Golfo de México y Mar Caribe. Cuadernos del Instituto de Biología 34: 1-594.

Santelices B. & P. Marquet, 1998. Seaweeds, latitudinal diversity, and Rapoport's rule. Diversity and Distributions 4: 71-75.

Sentíes G., A. & M.T. Fujii. 2002. El complejo *Laurencia* (Rhodomelaceae, Rhodophyta) en el Caribe Mexicano. pp. 119-192. In: Sentíes G., A. and K.M. Dreckmann (eds). Monografías Ficológicas 1. UAM-Iztapalapa and Red Latinoamericana de Botánica, México, D.F.