

Article

FactorsR: An RWizard Application for Identifying the Most Likely Causal Factors in Controlling Species Richness

Cástor Guisande ^{1,*}, Juergen Heine ², Emilio García-Roselló ², Jacinto González-Dacosta ², Baltasar J. García Perez-Schofield ³, Luis González-Vilas ¹, Antonio Vaamonde ⁴ and Jorge M. Lobo ⁵

¹ Facultad de Ciencias del Mar, Universidad de Vigo, Campus Lagoas-Marcosende s/n, 36310 Vigo, Spain; E-Mail: luisgv@uvigo.es

² Department of Computer Science, Universidad de Vigo, Campus Lagoas-Marcosende s/n, 36310 Vigo, Spain; E-Mails: juergen.heine@gmx.net (J.H.); egrosello@esei.uvigo.es (E.G.-R.); jdacosta@uvigo.es (J.G.-D.)

³ Escuela Superior de Ingeniería Informática, Edificio Politécnico s/n, Campus As Lagoas, Universidad de Vigo, 32004 Orense, Spain; E-Mail: jbgarcia@uvigo.es

⁴ Departamento de Estadística e Investigación Operativa Facultad de CCEE y Empresariales, Universidad de Vigo, Torrecedeira 105, 36208 Vigo, Spain; E-Mail: vaamonde@uvigo.es

⁵ Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales (CSIC), c/José Gutiérrez Abascal 2, 28006 Madrid, Spain; E-Mail: mcnj117@mncn.csic.es

* Author to whom correspondence should be addressed; E-Mail: castor@uvigo.es; Tel.: +34-986-812-586.

Academic Editor: Michael Wink

Received: 29 August 2015 / Accepted: 5 November 2015 / Published: 16 November 2015

Abstract: We herein present FactorsR, an RWizard application which provides tools for the identification of the most likely causal factors significantly correlated with species richness, and for depicting on a map the species richness predicted by a Support Vector Machine (SVM) model. As a demonstration of FactorsR, we used an assessment using a database incorporating all species of terrestrial carnivores, a total of 249 species, distributed across 12 families. The model performed with SVM explained 91.9% of the variance observed in the species richness of terrestrial carnivores. Species richness was higher in areas with both higher vegetation index and patch index, *i.e.*, containing higher numbers of species whose range distribution is less fragmented. Lower species richness than expected was observed in Chile, Madagascar, Sumatra, Taiwan, and Sulawesi.

Keywords: species richness; patch distribution; terrestrial carnivores

1. Introduction

The identification of the main factors affecting species richness or extension risk is an important topic in macroecological, biogeographical, and conservation studies [1,2]. Multiple regression (MR) has been frequently used to identify factors influencing response variables such as species richness or occurrence [3], but predictor variables are frequently significantly intercorrelated (multicollinearity), and the most frequently used stepwise-selection techniques (e.g., Akaike Information Criterion (AIC), quasi-Akaike Information Criterion (QAIC), Bayesian Information Criterion (BIC), *etc.*) are not the most effective way of identifying those variables most likely to influence variation in species richness, because these techniques are focused on finding a single best predictive function and not drawing inferences about the likely causality of variables [4]. Hierarchical partitioning (HP) [2] has been suggested as a possible solution to identify the most likely causal factors in an MR, because it considers all models in an MR setting jointly [5].

The relationship between species richness and some factors may be non-linearly related [6]. Therefore, the selection of the factors significantly affecting species richness would be improved if both untransformed and/or log-transformed factors are taken into account when selecting the factors.

Once the variables have been selected, it may be useful to perform a model for predicting species richness. Techniques such as the Support Vector Machine (SVM) models can model complex and non-linear patterns without making any assumptions. Unlike MR or other classical statistical techniques, SVM uses kernel functions in order to operate in a feature space with a higher dimension than the input space [7]. Therefore, this facilitates overcoming common problems in species distribution modeling such as the autocorrelation [8]. Another difference compared to MR is that SVM is based on the structural risk minimization principle, so it attempts to minimize the generalization error instead of minimizing the observed training error. As a consequence of this theoretical basis, SVM is a more robust technique with a better generalization capability and a lower risk of overfitting [8]. Some comparative studies in environmental applications have shown that SVM outperforms MR or other modeling techniques [9–13]. SVM has been successfully applied to explain the factors affecting species richness of marine elasmobranchs [14] and of freshwater fish species [15].

We herein present FactorsR (Supplementary Material: Help FactorsR), a freely available RWizard application which provides tools for the identification of the most likely causal factors (working with untransformed and/or log-transformed variables) significantly correlated with species richness and for depicting on a map the species richness predicted by an SVM model. The algorithm of FactorsR follows these steps: (1) quantification of the severity of multicollinearity among the factors; (2) estimation of the contribution of factors to species richness by using hierarchical partitioning; (3) selection of the factors with higher contribution among auto-correlated factors; (4) selection of those factors significantly correlated with species richness using the Akaike Information Criterion (AIC) in a stepwise multiple regression; and finally (5) species richness predictions and residuals of the SVM model performed with

the factors selected are shown on a map. The application also performs statistical analyses to test for normality and homoscedasticity of residuals and autocorrelation in the MR.

2. Methods

2.1. FactorsR

2.1.1. Installation and Platform Availability

FactorsR can be installed and started as a regular R add-on package in both Linux and Mac OS installing the file PlotsR.tar.gz (Package source) and in RGUI (R Graphical User Interface) for Windows installing the file PlotsR.ZIP (Windows binaries), both files are available at the website [16], in the downloads section, *i.e.*, FactorsR can be used without the need of RWizard. However, it is only when it is used in conjunction with RWizard that all its capabilities are truly revealed. To use FactorsR in RWizard, it is necessary to install the file Setup. FactorsR.v1.0.RAR is available on the website mentioned above. In this way, FactorsR will be added as an RWizard application. FactorsR is not available through Comprehensive R Archive Network (CRAN) because its large size is not acceptable according to CRAN policies. This large size is due to figures and a full explanation of the algorithm included in the help manual to facilitate the work of users.

RWizard is an open source interface under GNU General Public License that has been developed in C# on the Net platform. It is designed as an interface to facilitate the interaction with R (see video demonstration at the website [16]). The only requirement for the installation of RWizard is that R and Net Framework 4.0 must be already installed, and then the file Setup.RWizard.V1.1.exe (also available as RAR file) must be installed.

Geographic data about country, region, and province, *etc.* borders were obtained from the web page [17].

2.1.2. Software Design

It has a friendly menu to easily import a file (CSV format), to transform the variables, to select the statistical analyses to be performed, to export plots and maps to JPEG files or windows device, and to select the administrative areas to be included in the maps (from the entire world to any specific country, region, province, *etc.*). Any modification on this menu is passed to the script. For further details, see the user's manual of RWizard only available at the website [16] in the documentation section.

2.1.3. Functions of R Used in FactorsR

The goodness-of-fit measures for the entire hierarchy of models using all combinations of N independent variables is calculated using the package hier.part [18]. Data is analyzed using multiple regressions and the SVM for regression. In both models there is the possibility to use log-transformed variables. Stepwise multiple regressions were performed with the stats R package [19]. The relative contribution of each variable in the regressions was also estimated with the different methods available in the R package relaimpo [20,21]. The Kolmogorov-Smirnov test with Lilliefors correction was used to test for normality of residuals and performed with the package nortest [22]. Residuals from a linear

regression or multiple regressions must be independent and the Durbin–Watson statistic is used to detect the presence of autocorrelation violating this assumption, which was performed with the package *lmtest* [23,24]. The Breusch-Pagan test was used to test for homoscedasticity of residuals and performed with the package *lmtest* [23,24].

The presence of a multicollinearity—linear relationship among regressors—may be a major problem due to causing instability of estimates. The variance inflation factor (VIF) quantifies the severity of multicollinearity, thus providing an index that measures how much the variance of an estimated regression coefficient is increased because of collinearity [25], and this analysis was performed with the package *usdm* [26].

SVMs are supervised learning models with associated learning algorithms that analyze data and recognize patterns, and are used for classification and regression analysis. The SVM has excellent theoretical properties and a good performance under very general conditions and no strict assumptions are needed [10–13]. We used the function *ksvm* of the package *kernlab* [27,28], which supports the well-known C-svc, nu-svc, (classification) one-class-svc (novelty) eps-svr, and nu-svr (regression) formulations along with native multi-class classification formulations and the bound constraint SVM formulations. SVM is a discriminative classifier formally defined by a separating hyperplane, so given labeled training data the algorithm outputs an optimal hyperplane, which categorizes new examples. The operation of the SVM algorithm is based on finding the hyperplane that gives the largest minimum distance to the training examples. The optimal separating hyperplane maximizes the margin of the training data.

2.2. Usage of *FactorsR* by Applying It to Terrestrial Carnivores

2.2.1. Species Distribution

The range maps of terrestrial carnivores were developed using *ModestR* [29,30] functionalities by importing shape files downloaded from the website of The International Union for Conservation of Nature [31] and geographic records from The Global Biodiversity Information Facility [32]. This included a total of 249 species of terrestrial carnivores distributed across 12 families (Table S1. List of species). All families of the order Carnivora without marine species were included, therefore excluding the families Odobenidae, Otariidae and Phocidae. A new species of the olinguito *Bassaricyon neblina* [32] was not included.

The data was filtered using the following data cleaning facilities available in *ModestR* [33]: (1) data cleaning when downloading records from Global Biodiversity Information Facility (GBIF) and (2) habitat data cleaning.

For model estimation purposes, the available data resolution was downscaled to $30' \times 30'$. As the entire world was used, this encompassed areas in which species richness was zero.

2.2.2. Extent of Occurrence

Extent of occurrence (EOO) is defined as the area contained within the shortest continuous imaginary boundary, which can be drawn to encompass all the known, inferred or projected sites of present occurrence of a taxon [31].

One possibility to calculate the EOO in ModestR is to use the convex hull or the α -shape of the set of present occurrences of a species [30]. In this study, the convex hull was used.

2.2.3. Area of Occupancy

Area of occupancy (AOO) is the spatial distribution of known, inferred or projected sites of occurrence expressed in km² and was estimated with ModestR as described by [34]. We estimated AOO for each species with ModestR using rasterized maps with a resolution of 1' × 1' cells, in which the occupied area was estimated using the following equation:

$$1.852 \times \frac{12756.2 \times \pi}{21600} \cos\left(\text{latitude} \times \frac{\pi}{180}\right) \quad (1)$$

The value 1.852 is a nautical mile in km, 12,756.2 is 2× the radius of the earth in km and finally 21,600 in the value used to transform 1 min of arc. The latitude in the equation is the central value of the individual considered pixels. The equation was applied to all the pixels occupied by the species and all the obtained values were summed to calculate AOO.

2.2.4. Patch Distribution

The patch distribution of the species was estimated using the following equation, which ranges from 0 to 1, and is an indicator of habitat fragmentation. A value close to 1 indicates that AOO is equal to EOO, so the species occupies all areas (or habitats) within the geographical limits to its occurrence. A value close to 0 indicates that the species is patchy distributed.

$$PD = \frac{AOO}{EOO \text{ estimated with convex hull}} \quad (2)$$

2.2.5. Area of Occupancy Index and Patch Index

Area of occupancy index (AOO index) is the mean AOO of all species present in each cell and was also estimated with ModestR [30]. A lower AOO index indicates a higher number of species with a small geographic range size in the cell.

Patch index is the mean patch distribution of all species present in each cell and was also estimated with ModestR. A high patch index indicates a higher number of species with a range distribution less fragmented in the cell.

2.2.6. Environmental Variables

All the environmental variables used in this study are shown in Table S2. Environmental variables and were described in detail by [15].

3. Results

The factors selected with higher contribution (using hierarchical partitioning), without collinearity among them (a maximum VIF value of 9.4), and significantly correlated with species richness (using a stepwise regression) explained the 60.1% of the variance observed in species richness of terrestrial

carnivores (Figure 1). Area of rivers in the cell, BIO3 and AOO index were negatively correlated with species richness. Patch index was the factor with the highest contribution to species richness followed by vegetation index (Figure 1). Figure 2 shows the relationships between species richness and those factors with a higher contribution.

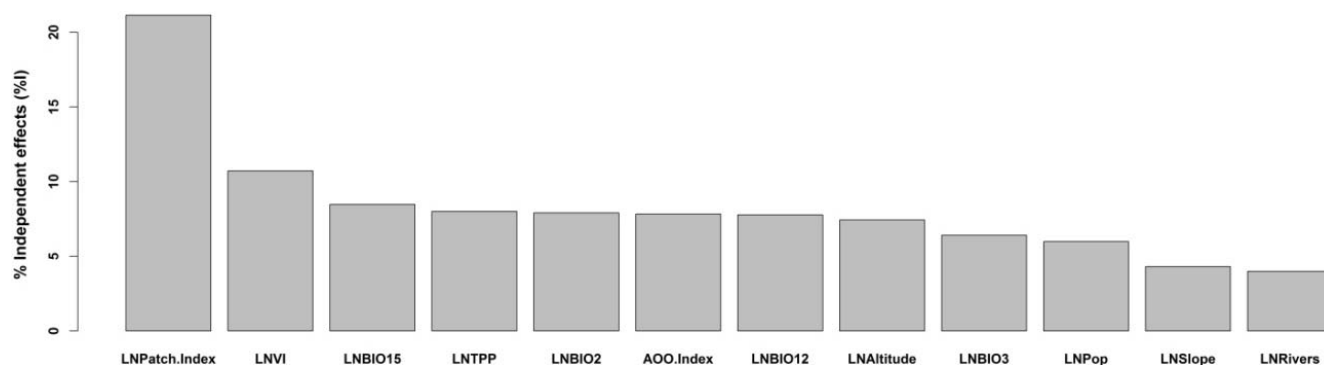


Figure 1. Relative contribution of the variables significantly correlated with species richness estimated by using hierarchical partitioning. Abbreviations: VI (vegetation index, in $\text{g C m}^{-2}\cdot\text{d}^{-1}$), BIO15 (precipitation seasonality, coefficient of variation), TPP (terrestrial primary production in $\text{g C m}^{-2}\cdot\text{d}^{-1}$), BIO2 (mean diurnal range in $^{\circ}\text{C}$), BIO12 (annual precipitation in mm), altitude (in m), BIO3 (isothermality in $^{\circ}\text{C}$), Pop (human population density in number of people per km^2), slope (topographic slope in degrees) and rivers (area of the river basins in km^2). LN means that the variable is ln transformed.

It was not possible to test the significance of the regression because the standardized residuals did not have a normal distribution (Shapiro–Wilk, $p < 0.001$), there was autocorrelation (Durbin–Watson, $p < 0.001$), and there was no homoscedasticity in the residuals (Breusch–Pagan test, $p < 0.001$).

The model performed with SVM explained 91.9% of the variance observed in the species richness of terrestrial carnivores. Figure 3 shows the actual richness, the richness predicted by SVM model, and the residuals of the SVM model. There are well-known and widely used estimators and species accumulation techniques directed to predict species richness and to identify well surveyed spatial units (WS_{sus}).

The residuals of the model performed with SVM are also useful for discriminating the location of WS_{sus} . The residuals showed a uniform green color in the map, indicating that most of the averaged residuals were close to zero (Figure 3, lower panel).

Negative residuals, which may be areas with undiscovered and/or unregistered species, or areas with lower species richness due to the negative effect of anthropogenic factors, were observed in Chile and other areas of South America, in Madagascar, Sumatra, Taiwan, and Sulawesi (Figure 3).

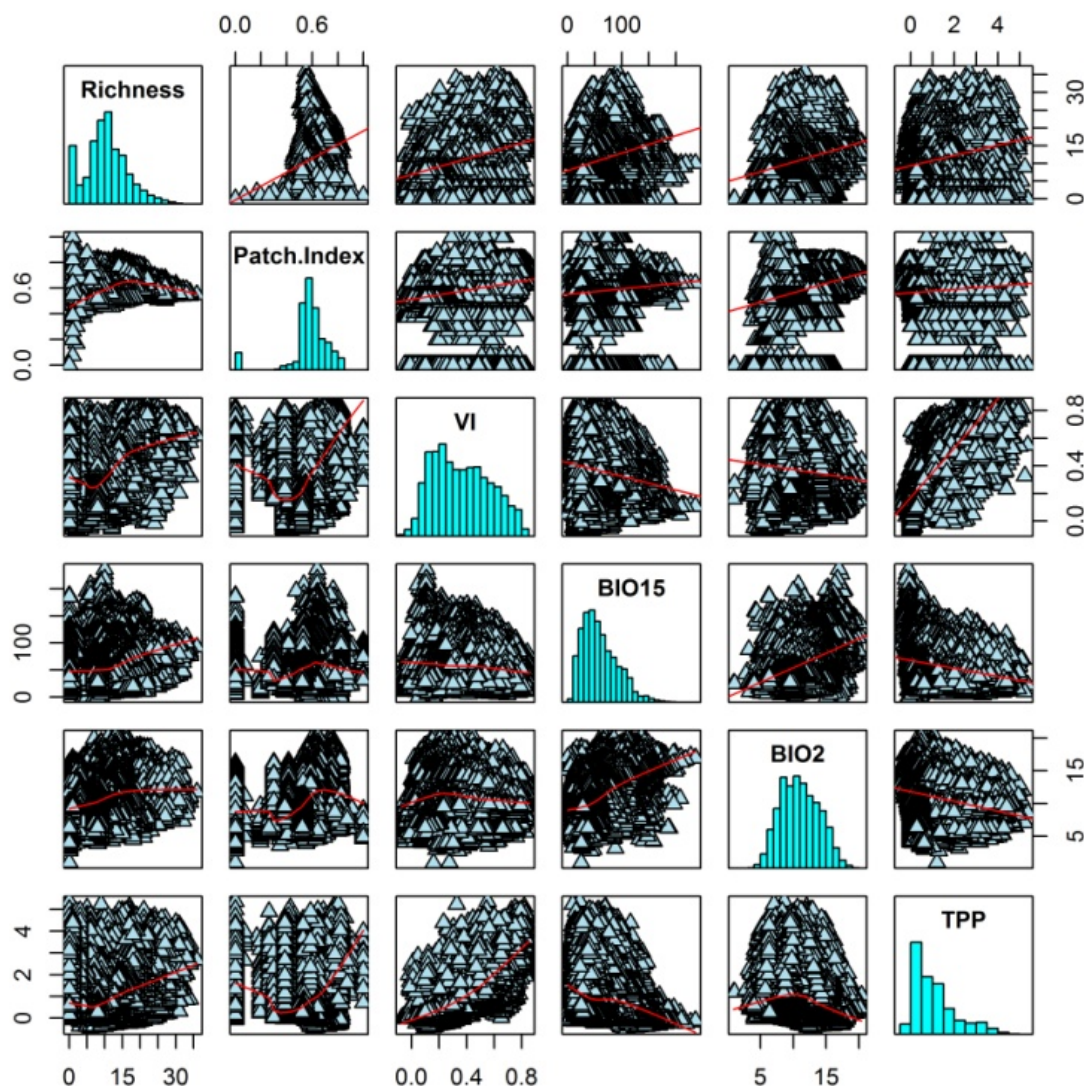


Figure 2. Relationships between species richness and some of the variables with higher relative contribution to species richness. Abbreviations as shown in Figure 1. The diagonal shows the frequency histograms of the variables, and the lower left panel indicates the smooth regression among variables and the upper right panel indicates the regression among variables.

4. Discussion

We demonstrated that the algorithm of FactorsR, combining classical statistical techniques for testing multicollinearity and the significance of factors with techniques focused on drawing inferences about the likely causality of factors, is useful for the identification of the factors affecting species richness. The novel contribution of the FactorsR's algorithm is to select among auto correlated factors (using untransformed and log-transformed factors) the most likely causal factors in controlling species richness, before to perform the stepwise multiple regression. Therefore, the algorithm is focused on drawing inferences about the likely causality of the significant variables, which is a more ecological rather than statistical approach.

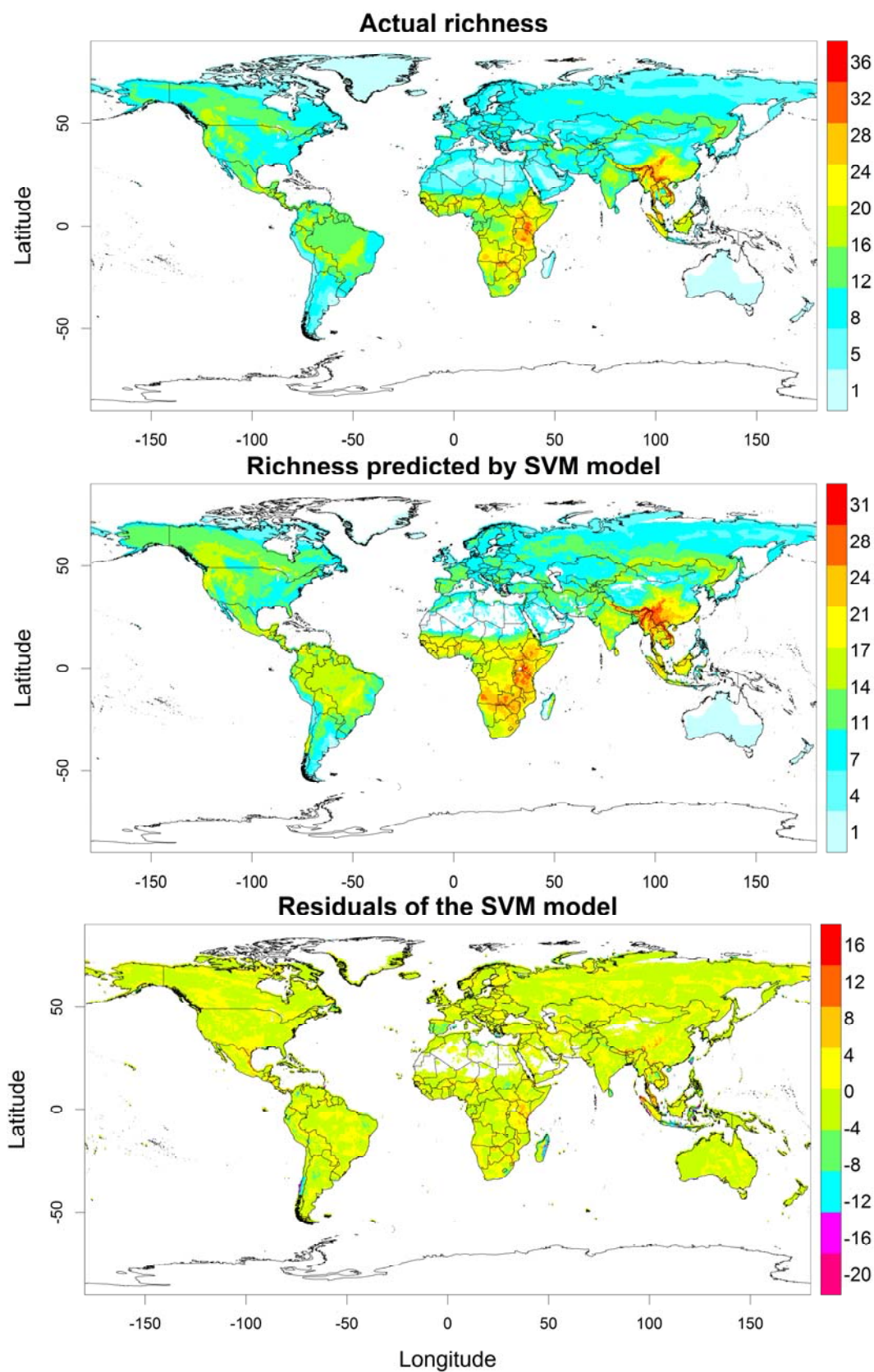


Figure 3. Actual richness, the richness predicted by the Support Vector Machine (SVM) model and the residuals of the SVM in cells of 30' × 30'. Negative residuals indicate a number of observed species in the cell lower than that predicted by the model.

In the example used in this study to illustrate the usage of FactorsR, the variables selected corroborate many previous findings that terrestrial carnivores are vulnerable to habitat fragmentation and require landscape connectivity [36]. Species richness was higher in areas with a higher vegetation index, high patch index, and low AOO index, *i.e.*, areas containing a higher number of species confined to a specific, relatively small geographical area and/or species whose range distribution is not fragmented. Instead of measuring the fragmentation of the habitat, we approached the study by analyzing the degree of fragmentation of the species' range distribution. It is necessary to point out, however, that since habitat loss is assumed to be typically accompanied by habitat fragmentation, the results of empirical studies of habitat fragmentation are difficult to interpret if the measure fragmentation does not distinguish between habitat loss and habitat fragmentation *per se*, *i.e.*, the breaking apart of habitat after controlling for habitat loss [37]. Fragmentation of the species' range distribution may be the final product of habitat loss, habitat fragmentation *per se*, and/or other causes.

The other contribution of FactorsR's algorithm is to perform an SVM model for predicting species richness instead of using MR. The problem is that MR makes several strong assumptions that cannot be violated: no or little multicollinearity, multivariate normality, no auto-correlation, and homoscedasticity of residuals [38–40]. It is easy to resolve the problem of multicollinearity, by just leaving only one variable among auto correlated variables [39]. The problem resulting from residuals without normal distribution is that there can be no assurance that probability of the model, the degree of significance, is the correct one [39]. If the assumption of no auto-correlation is not met, it is not possible to know exactly how much the variance is explained by the independent variables [39], so the r^2 is not correct. The fact of not fulfilling the homoscedasticity of the residuals means that the model is not as predictive for the entire range of values of the dependent variable [39].

It is difficult to fulfill all these assumptions of MR in ecological data, particularly the common problem of spatial autocorrelation [41]. SVM models, however, can model complex and non-linear patterns without making any assumptions [7]. In the example used in this study, it was not possible to test the significance of the regression because the standardized residuals did not have a normal distribution, and there was autocorrelation and there was no homoscedasticity in the residuals, common violations of the MR's assumptions in macroecological, biogeographical and conservation studies [39]. Moreover, the variance explained by SVM (91.9%) was much higher than the one obtained with MR (60.1%); thereby corroborating many previous findings that SVM outperforms MR or other modeling techniques [10–13].

5. Conclusions

In summary, the novel statistical approach of the FactorsR's algorithm focusing on the identification of most likely causal factors significantly correlated with species richness, may be a useful tool in biodiversity research. Moreover, with its possibility of depicting negative residuals of the SVM model on a map, it may also be useful for identifying areas with undiscovered and/or unregistered species, or areas with lower species richness due to the negative effect of anthropogenic factors.

Supplementary Materials

Supplementary materials can be accessed at: <http://www.mdpi.com/1424-2818/7/4/385/s1>.

Acknowledgments

We thank Empresa Nacional de Electricidad (ENDESA) for technical and financial support.

Author Contributions

Cástor Guisande designed and developed the package FactorsR, Cástor Guisande, Juergen Heine, Emilio García-Roselló, Jacinto González-Dacosta and Baltasar J. García Perez-Schofield developed RWizard, Luis González Vilas organized the data base of the administrative areas included in RWizard, Antonio Vaamonde helped with the statistic of FactorsR and Cástor Guisande and Jorge M. Lobo wrote and revised the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Gaston, K.J. Global patterns in biodiversity. *Nature* **2000**, *405*, 220–227.
2. Di Marco, M.; Buchanan, G.M.; Szantoi, Z.; Holmgren, M.; Grottolo Marasini, G.; Gross, D.; Tranquilli, S.; Boitani, L.; Rondinini, C. Drivers of extinction risk in African mammals: The interplay of distribution state, human pressure, conservation response and species biology. *Philos. Trans. R. Soc. B* **2014**, doi:10.1098/rstb.2013.0198.
3. Mac Nally, R. Regression and model-building in conservation biology, biogeography and ecology: The distinction between—And reconciliation of—“Predictive” and “explanatory” models. *Biodivers. Conserv.* **2000**, *9*, 655–671.
4. Chevan, A.; Sutherland, M. Hierarchical partitioning. *Am. Stat.* **1991**, *45*, 90–96.
5. Mac Nally, R. Multiple regression and inference in ecology and conservation biology: Further comments on identifying important predictor variables. *Biodivers. Conserv.* **2002**, *11*, 1397–1401.
6. Zuur, A.; Ieno, E.N.; Smith, G.M. *Analysing Ecological Data*; Springer: New York, NY, USA, 2007.
7. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*; Cambridge University Press: New York, NY, USA, 2000.
8. Drake, J.M.; Randin, C.; Guisan, A. Modelling ecological niches with support vector machines. *J. Appl. Ecol.* **2006**, *43*, 424–432.
9. Basak, D.; Pal, S.; Patranabis, D.C. Support vector regression. *Neural Inf. Process. Lett. Rev.* **2007**, *11*, 203–224.
10. Yao, X.J.; Panaye, A.; Doucet, J.P.; Zhang, R.S.; Chen, H.F.; Liu, M.C.; Hu, Z.D.; Fan, B.T. Comparative study of QSAR/QSPR correlations using support vector machines, radial basis function neural networks, and multiple linear regression. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1257–1266.
11. Ribeiro, R.; Torgo, L. A comparative study on predicting algae blooms in Douro River, Portugal. *Ecol. Model.* **2008**, *212*, 86–91.
12. Al-Anazi, A.F.; Gates, I.D. Support vector regression for porosity prediction in a heterogeneous reservoir: A comparative study. *Comput. Geosci.* **2010**, *36*, 1494–1503.

13. Chen, S.-T.; Yu, P.-S.; Tang, Y.-H. Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *J. Hydrol.* **2010**, *385*, 13–22.
14. Guisande, C.; Patti, B.; Vaamonde, A.; Manjarrés-Hernández, A.; Pelayo-Villamil, P.; García-Roselló, E.; González-Dacosta, J.; Heine, J.; Granado-Lorencio, C. Factors affecting species richness of marine elasmobranchs. *Biodivers. Conserv.* **2013**, *22*, 1703–1714.
15. Pelayo-Villamil, P.; Guisande, C.; Vari, R.P.; Manjarrés-Hernández, A.; García-Roselló, E.; González-Dacosta, J.; Heine, J.; González Vilas, L.; Patti, B.; Quinci, E.M.; *et al.* Global diversity patterns of freshwater fishes—Potential victims of their own success. *Divers. Distrib.* **2015**, *21*, 345–356.
16. RWizard Software, Version 1.1. Available online: <http://ipez.es/RWizard> (accessed on 15 June 2015).
17. OpenStreetMap Foundation Wiki Contributors. <http://www.openstreetmap.org> (accessed on 3 January 2015).
18. Walsh, C.; Mac Nally, R. Hierarchical Partitioning. R Package, ver. 1.0–4. Available online: <https://cran.r-project.org/web/packages/hier.part> (accessed on 7 January 2015).
19. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.
20. Grömping, U. Relative Importance for Linear Regression in R: The Package relaimpo. *J. Stat. Softw.* **2006**, *17*, 1–27.
21. Grömping, U. Relative Importance of Regressors in Linear Models. R Package, ver. 2.2–2. Available online: <https://cran.r-project.org/web/packages/relaimpo> (accessed on 30 March 2015).
22. Gross, J. Five Omnibus Tests for the Composite Hypothesis of Normality. R Package, ver. 1.0–4. Available online: <https://cran.r-project.org/web/packages/nortest> (accessed on 14 April 2015).
23. Zeileis, A.; Hothorn, T. Diagnostic Checking in Regression Relationships. *R News* **2002**, *2*, 7–10.
24. Hothorn, T.; Zeileis, A. Testing Linear Regression Models. R package, ver. 0.9–34. Available online: <https://cran.r-project.org/web/packages/lmtest> (accessed on 2 February 2015).
25. Fox, J.; Weisberg, S. *An R Companion to Applied Regression*; Sage Publications: Thousand Oaks, CA, USA, 2011.
26. Naimi, B. Uncertainty Analysis for Species Distribution Models. R Package, ver. 1.1–15. Available online: <https://cran.r-project.org/web/packages/usdm/index.html> (accessed on 2 June 2015).
27. Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. Kernlab—An S4 Package for Kernel Methods in R. *J. Stat. Softw.* **2004**, *11*, 1–20.
28. Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. Kernel-Based Machine Learning Lab. R Package, ver. 0.9–22. Available online: <https://cran.r-project.org/web/packages/kernlab> (accessed on 5 May 2015).
29. Pelayo-Villamil, P.; Guisande, C.; González-Vilas, L.; Carvajal-Quintero, J.D.; Jiménez-Segura, L.F.; García-Roselló, E.; Heine, J.; González-Dacosta, J.; Manjarrés-Hernández, A.; Vaamonde, A.; *et al.* ModestR: Una herramienta informática para el estudio de los ecosistemas acuáticos de Colombia. *Actual. Biol.* **2012**, *34*, 225–239.
30. García-Roselló, E.; Guisande, C.; González-Dacosta, J.; Heine, J.; Pelayo-Villamil, P.; Manjarrés-Hernández, A.; Vaamonde, A.; Granado-Lorencio, C. ModestR: A software tool for managing and analyzing species distribution map databases. *Ecography* **2013**, *36*, 1202–1207.

31. The IUCN Red List of Threatened Species. Version 2015.1. Available online: <http://www.iucnredlist.org> (accessed on 15 June 2015).
32. GBIF Data Portal. Available online: <http://data.gbif.org> (accessed on 30 June 2015).
33. Helgen, K.M.; Pinto, C.M.; Kays, R.; Helgen, L.E.; Tsuchiya, M.T.N.; Quinn, A.; Wilson, D.E.; Maldonado, J.E. Taxonomic revision of the olingos (*Bassaricyon*), with description of a new species, the Olinguito. *ZooKeys* **2013**, *324*, 1–83.
34. García-Roselló, E.; Guisande, C.; Heine, J.; Pelayo-Villamil, P.; Manjarrés-Hernández, A.; González-Vilas, L.; González-Dacosta, J.; Vaamonde, A.; Granado-Lorencio, C. Using ModestR to download, import and clean species distribution records. *Methods Ecol. Evol.* **2014**, *5*, 708–713.
35. García-Roselló, E.; Guisande, C.; Manjarrés-Hernández, A.; González-Dacosta, J.; Heine, J.; Pelayo-Villamil, P.; González-Vilas, L.; Vari, R.P.; Vaamonde, A.; Granado-Lorencio, C.; *et al.* Can we derive macroecological patterns from primary GBIF data? *Glob. Ecol. Biogeogr.* **2015**, *24*, 335–347.
36. Crooks, K.R.; Burdett, C.L.; Theobald, D.M.; Rondinini, C.; Boitani, L. Global patterns of fragmentation and connectivity of mammalian carnivore habitat. *Philos. Trans. R. Soc. B* **2011**, *366*, 2642–2651.
37. Fahrig, L. Effects of habitat fragmentation on biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **2003**, *34*, 487–515.
38. Guisande, C.; Barreiro, A.; Maneiro, I.; Riveiro, I.; Vergara, A.R.; Vaamonde, A. *Tratamiento de Datos*; Díaz de Santos: Madrid, Spain, 2006.
39. Guisande, C.; Vaamonde, A.; Barreiro, A. *Tratamiento de Datos con R, Statistica y SPSS*; Díaz de Santos: Madrid, Spain, 2011.
40. Guisande, C.; Vaamonde, A. *Gráficos Estadísticos y Mapas con R*; Díaz de Santos: Madrid, Spain, 2012.
41. Kissling, W.D.; Carl, G. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Glob. Ecol. Biogeogr.* **2008**, *17*, 59–71.