Biodiversity Informatics, 13, 2018, pp. 11-26

A METRIC TO QUANTIFY ANALOGOUS CONDITIONS AND RANK ENVIRONMENTAL LAYERS

PETER LÖWENBERG-NETO
Instituto de Ciências da Vida e da Natureza, UNILA, Av. Tarquínio Joslin dos Santos, 1000 CEP 85870-901, Foz do Iguacu, Parana, Brasil. peter.lowenberg@unila.edu.br

Abstract.—Analogous conditions in environmental variables are expected because environments are spatially autocorrelated and often present similar combinations over geographic space. That similar environmental combinations may be found at different localities provides a crucial basis for correlative species distribution modeling. An absolutely analogous variable is constant, while a non-analogous variable has no-repeating values, yet no current method allows researchers to quantify intermediate degrees of analogous conditions and rank environmental layers. I approached this issue from the perspective of dual-space correspondence, in which (a) variable range and modal frequency have a theoretical inverse relationship ($y \propto x^{-1}$), and (b) modal values of frequency are limited by the number of pixels in a given raster layer. For two geographic extents and two resolutions (2.5' and 10'), I obtained range and modal frequency of 19 bioclimatic variables and 5 reference variables. Then, I measured Euclidean distances from candidate variables to the non-analogous variable as a metric for degree of analogous conditions, which were used to rank variables. Bioclimatic layers were plotted in log-log scatterplots of range *vs.* modal frequency; variables were located inside the upper-right triangle (except for one set), and no layer fit the inverse model. Temperature variables presented higher degrees of analogous conditions than precipitation for South America and the Araucaria Moist Forests ecoregion. Geographic extent and pixel resolution changed the degree of analogous conditions of derived variables (quarterly and monthly); however, a pattern of change was not observed, which suggested *ad hoc* hypotheses on geographic and temporal idiosyncrasies. Variables with high contribution in previous SDM/ENM studies (e.g., temperature seasonality and annual precipitation) showed low degree of analogous conditions. It is expected that heterogeneous layers would generate better correlational geographic distributional predictions than analogous variables, even though this hypothesis remains untested. Ranking layers can provide grounds for selecting variables in distribution and niche modeling, particularly as regards interpreting spatial projection and transferability. Alternatively, ranking can be used to compare degrees of analogous conditions of the same layer in different time spans.

Key words.—analog conditions, bioclimatic variables, environmental space, geographic space

An environmental digital layer is a common object of geographic information systems, in which values of a continuous variable are stored in a spatially referenced matrix (Chang 2017). In biogeography and macroecology, environmental layers include temperature, precipitation, humidity, radiation, soil, and human occupation. They can be used as a background in illustrating maps and as predictors in statistical analysis (Williams et al. 2012). Environmental layers are normally raster format objects, which implies some level of discretization of continuous space (Hijmans and Elith 2017).

Environmental variables are not distributed heterogeneously across space. Variables like temperature are spatially autocorrelated, and show repeated or similar values over space (Legendre 1993). Additionally, combinations of variables can replicate more complex circumstances and cause localities to represent analogous conditions, for example, a monsoon-like climate in South America caused by heating and circulation regimes associated with topography (Zhou and Lau 1998). In the literature, 'analogous' or 'analog' has frequently been employed to designate equivalent climatic conditions through time. Several studies have inferred how populations and communities responded to past (Overpeck et al. 1992, Jackson and Overpeck 2000) and current climate change (Garcia et al. 2014) by comparing to modern climates against analogous past and/or future climate scenarios. They have also estimated new and disappearing climatic combinations in scenarios of change (Ohlemüller et al. 2006, Williams et al. 2007, Ackerly et al. 2010).

Contemporary analogous conditions, such as pixels with equal or similar values, provide grounds for species distribution modeling (Guisan and Zimmermann 2000). In a correlative approach, occurrences of species and digital environmental layers are used to estimate existing or realized niches of species (Peterson et al. 2011). Then, a given algorithm may search (Elith and Graham 2009) across the geographic extent for equivalent conditions (Elith and Leathwick 2009). An interesting topic relevant to correlational modeling is projection over space and the problem of non-analog climates (Fitzpatrick and Hargrove 2009). It is possible that native-range geographic range climate conditions are distinct from conditions in the invaded region (Fitzpatrick et al. 2006, Soberón and Peterson 2011). In these cases, challenging the equilibrium assumption, including invasion stages (Gallien et al. 2012), and controlling for possible niche shifts (Guisan et al. 2014) in spatial projections may overcome the problem.

Correlational modeling procedures are based on the dual-space correspondence (Peterson et al. 2011, Soberón et al. 2017). Dual-space correspondence, or Hutchinson's duality (Colwell and Rangel 2009), occurs when points from an input space (e.g. biotope space, geographic space) are plotted in a feature space (e.g. climatic space), where variables are axes and measures are coordinates (Hutchinson 1978, Colwell and Rangel 2009). Each input point g from geographic space (**G**) has a vector $\mathbf{e}_g$ composed of measures of v variables, such as annual mean temperature, annual precipitation, and so on. The vector $\mathbf{e}_g$ with v elements represents the coordinates of the point in a v-dimensional space, the environmental space (**E**) (Peterson et al. 2011).

When enough variables are used at sufficient precision, points from **G** generally correspond one to one to points in **E** (Apinall and Lees 1994). However, this situation is not necessarily the case, and the same or very similar (analogous) climatic combinations may occur in separated geographic localities.

In any case, the cloud of points can be interpreted as a particular realization of environmental conditions that occur across a given geographic extent at a particular time (i.e., the realized environment) (Jackson and Overpeck 2000). Two additional features from **E** are of particular interest: (i) empty environmental spaces denote combinations of conditions that are missing, such as warm (>30°C) and wet (>3000 mm) climates in California (Ackerly et al. 2010), and (ii) closely-located points that are environmentally analogous localities (de Oliveira et al. 2014).

For a climatic variable, when many localities have the same value, their corresponding points pile up in a kernel in **E**, and represent high frequency regions (Figure 1). In this case, when more than one locality has the same environmental-variable value, if a single point was mapped back, it would represent all pixels in **G** in an asymmetric relationship (one-to-many, a partial reciprocity, Colwell and Rangel 2009). On the other hand, a non-analogous (non-repeating) variable in **G** has corresponding points (1:1) spread over **E**, with a maximum frequency of one.

Let $n = |\mathbf{G}|$ the number of points in the rasterization of a variable; y is the range of values of the variables ($y = y_{\max} - y_{\min}$) and x is the modal frequency of variable y. If the extent and resolution of the discretization of **G** do not change, then an inverse relationship ($y \propto x^{-1}$) between the range of a variable and its modal frequency is to be expected. In other words, if the range of a variable is small, most cells in the raster have values in that small range. If the range is broad, the distribution of frequencies will tend to be flat (Figure 1). This effect occurs because (a) each variable's range determines the span of its axis in **E**, and (b) the number of pixels is constant for a given extent and resolution, which provides a zero-sum scenario. Hence, when the span is low, density will be concentrated in a small region, and kernel modal frequency will be high; when the breadth is high, density is spread over the axis and maximum frequency is low. This point is important because of Hutchinson's Duality: the same niche breadths in regions of contrasting spans of values of environmental variables may predict contrasting sizes of areas of distribution.

Statistical selection of variables for species distribution modeling frequently aims at controlling variable collinearity and ranking variables (Negrão and Löwenberg-Neto in prep.). Current metrics for analogy of conditions

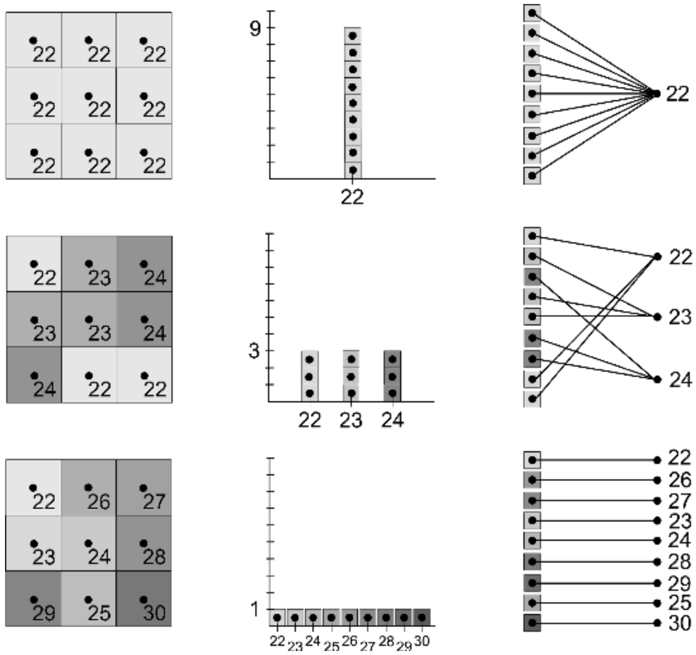

Figure 1. Inverse relationship between variable range in **G** (geographic space) and modal frequency in **E** (environmental space). Top row represents an absolutely analogous variable, zero-ranged, with modal frequency equal to number of pixels, and bipartite network asymmetric. Middle row represents a layer with intermediate degrees. Bottom row is a non-analogous layer: range equals the number of pixels minus one, modal frequency equals one; bipartite network is symmetric.

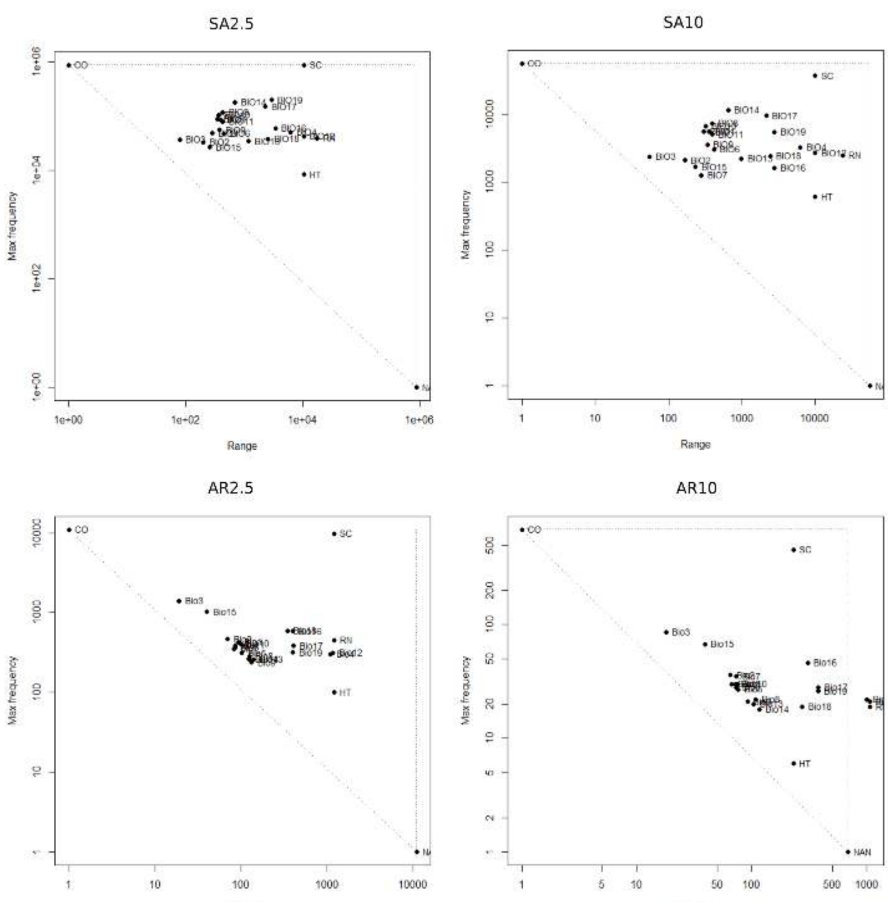

Figure 2. Log-transformed scatterplots of variable range *versus* maximum frequency for two extents (South America, SA; Araucaria Moist Forests, AR), and two pixel resolutions (2.5’, 10’). Bioclimatic variables are labeled following Hijmans et al. (2006), and referential variables as CO = constant, SC = semi-constant and wide-range, RN = random normal, HT = heterogeneous, and NAN = non-analogous.

are available only in temporal frameworks and only for comparing pixels within layers (Garcia et al. 2014), which does not allow comparison among layers. In the present paper, I present a layer-scoped metric that quantifies overall degree of analogy of environmental layers under Hutchinson's duality. I have then used the measurements to rank variables, and discuss the importance of these tools and ideas in the broader field of distributional ecology.

METHODS

I obtained 19 bioclimatic variables (Hijmans et al. 2005), and calculated their ranges and modal values. Measurements were developed for variables at two geographic extents: all of South America (SA) and the Araucaria Moist Forests ecoregion (AR) in southern Brazil (Olson et al. 2002). For both extents, I analyzed bioclimatic variables at two resolutions: 2.5' and 10' (Guisan and Thuiller 2005).

For each combination of extent and resolution, I created 5 reference variables: (1) Constant variable (CO) is a homogeneous variable, with a modal frequency equal to the number of pixels and zero for range. For the log-transformed distance (see below), I assigned variable range to one. (2) Semi-constant, wide range (SC), is the second most homogeneous variable has a single, high modal frequency and a wide range with low frequencies. This variable is important variable because it controls for variables that are very homogeneous but that may mislead interpretation or metric quantification owing to its wide range of values. (3) Random normal (RN) is a heterogeneous variable drawn from a normal distribution; its range is similar to the bioclimatic variable with the broadest range in all combinations, annual precipitation. (4) Heterogeneous (HT) is the most heterogeneous variable, with a range similar to that of annual precipitation; it has repeating values with the lowest maximum frequency. Finally, (5) Non-analogous (NAN) is the absolutely heterogeneous variable, with no repeating values, range equal to the number of pixels, and a modal frequency of one.

I compiled variable ranges and modal frequencies into data matrices. For each dataset, I measured the Euclidean distance matrix between rows using *dist* command and method = "euclidean" $sqrt(sum((x_i - y_i)^2))$ in R version 3.5.0. The distance between a given variable to the non-analogous variable (NAN) was used to quantify variable's degree of analogous conditions; therefore, longer distances to NAN denoted more homogeneous (analogous) variables. The same measurement procedure was done for a log-transformed data matrix (log distance). Pearson's correlations were calculated among distance, log distance, and secondary metrics, which included range, maximum frequency, and the Shannon-Weaver diversity index (Shannon 2001). The last metric was calculated using the command *diversity* in the 'vegan' R package (Oksanen et al. 2007).

RESULTS

Variable histograms used to measure range and modal frequency are presented in Appendix A; measurements are in Appendix B. For each variable, range and modal frequency were log-transformed and plotted (Figure 2). Correlation analyses showed that distance and log distance were strongly positively correlated; log distance was negatively correlated with variable range (Appendix C).

For each combination of geographic extent and resolution, distances to the NAN variable were used to compare variables. Ranking showed that reference variables CO and SC had the highest degree of analogous conditions while RN, HT and NAN the least (Figure 3A). Statistical variables showed disparate degrees of analogous conditions (Figure 3B): mean diurnal range, temperature annual range, and precipitation seasonality had their degree of analogy of conditions affected by geographic extents, whereas isothermality and temperature seasonality, which are standardized variables, were less affected. Temperature variables presented higher degrees of analogy of conditions than precipitation in both regions (Figure 3C and 3D).

DISCUSSION

A metric that quantifies overall degree of analogy of conditions for individual environmental layers was presented. By creating a non-analogous variable in which the range equals the number of pixels and modal frequency equals one, it was possible to plot and measure

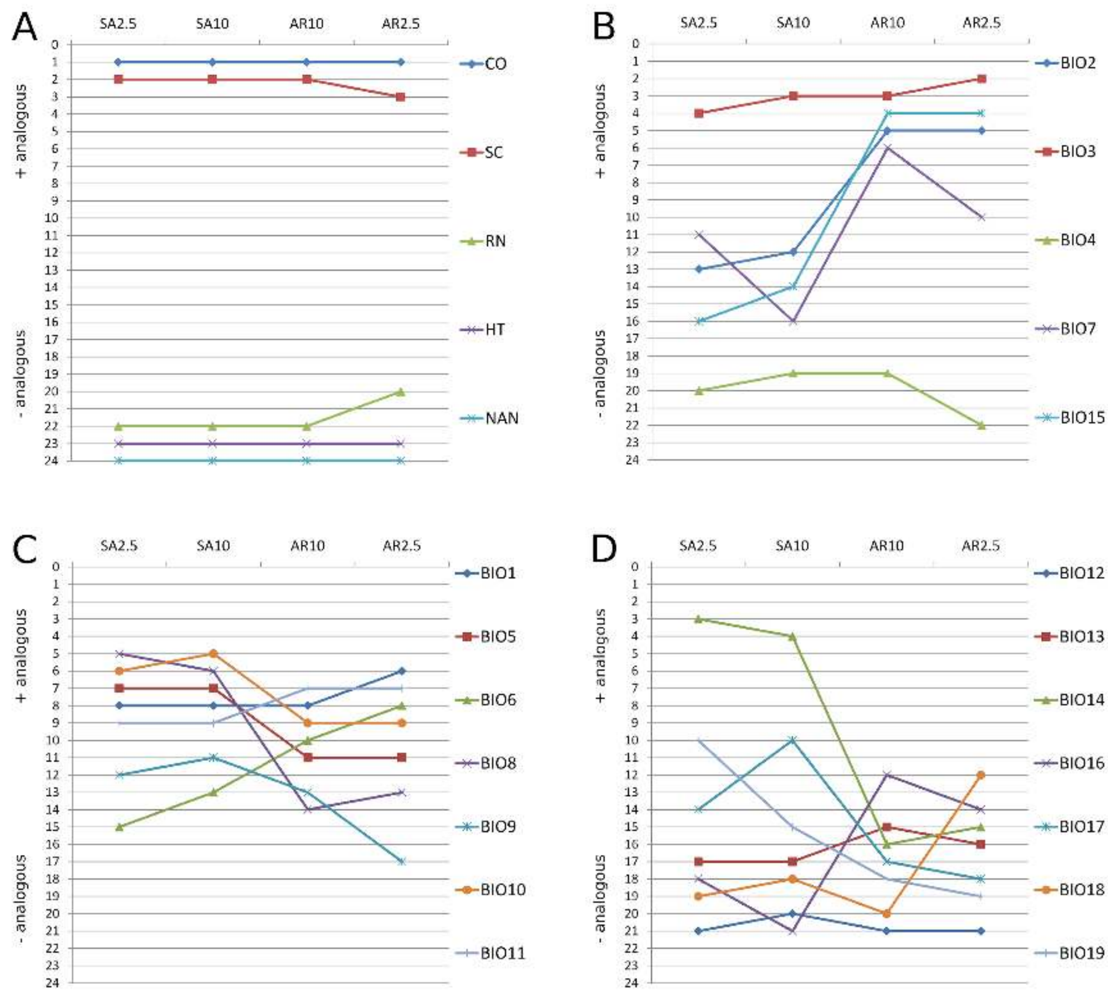

Figure 3. Ranking environmental variables by their Euclidean distance to the non-analogous (NAN) variable in a line graph for four combinations of extent and resolution (South America at 2.5', South America at 10', Araucaria Moist Forests at 2.5', Araucaria Moist Forests at 10'). Variables were displayed in four subsets: (a) reference, (b) statistical, (c) temperature, and (d) precipitation. Bioclimatic variables were labeled following Hijmans et al. (2006), and reference variables are as follows: CO = constant, SC = semi-constant and wide-range, RN = random normal, HT = heterogeneous, and NAN = non-analogous.

the Euclidean distance to the candidate variable as an index of dissimilarity to the non-analogous variable. In this sense, higher distances to NAN denote that a variable has a high degree of analogous conditions.

Log-transformed plots provided a better visualization of the variables and their spatial positions in the kernel, and allow visualization of expected upper limits, which are based on the number of pixels the expected inverse relationship between variable range and modal frequency (Figure 2). Bioclimatic layers were located inside the right-angled triangle, and no layer fit the inverse model, which was expected for climatic variables.

Reference variables, especially SC and HT, showed consistent positions in all plots, providing internal references for the bioclimatic variables; conversely, RN was very close to realistic variables. Three variables were placed outside the triangle envelope, beyond the vertical axis. This effect occurred because variables had a wider range than the numbers of pixels available. The AR10 treatment comprised 684 pixels, and ranges were above 1000 for temperature seasonality, annual precipitation, and RN.

Secondary metrics were not tested formally because I intended to provide a metric based on the duality ontology. Nevertheless, their correlations with Euclidean distance provided some information. For example, variable range was strongly negatively correlated with distance, which supports an exploratory approximation to analogous degree with no need for developing scatterplot. Modal frequencies were less correlated with distance in treatments with finer pixels; diversity index showed a weak relation to distance; the index did not discern the semi-constant wide-range variable well.

Ranking bioclimatic layers showed that geographic extent and pixel resolution both affect the degree of analogy of conditions. It was expected that a change in grain size would not severely affect the degree of analogy of conditions (Guisan et al. 2007). For the same extent, ranking showed that pixel resolution modified a variable's ranking position, even though it showed no pattern of increasing analogy of conditions with decreasing resolution or *vice-versa*. Regarding geographic extent, it was expected that extents limited the ranges of environmental variables (Thuiller et al. 2004, Randin et al. 2006), and that the smaller geographic extent would present higher degrees of analogy of conditions (Anderson and Raza 2010). This effect was observed only for a few variables, including statistical ones constructed by consideration of ranges of values (temperature diurnal and annual ranges) and coefficients of variation (precipitation seasonality).

A third expectation was that variables arranged in temporal slices (quarterly and monthly) would show increasing degrees of analogy of conditions when compared to annual variables. In fact, annual mean temperature showed consistent ranking positions (8, 8, 8, 6, Fig. 3C), whereas time-sliced variables showed changeable positions (e.g., mean temperature of wettest quarter, ranks 5, 6, 14, 13). The same effect was observed for precipitation variables (e.g., annual precipitation, ranks 21, 20, 21, 21; precipitation of wettest quarter, ranks 18, 21, 12, 14); however, I did not observe any general trend of increasing degree by decreasing temporal slice size. In sum, geographic extent and pixel resolution changed the degree of analogous conditions of derived variables whereas annual variables tended to maintain their rankings. No consistent trend of change between extension/resolution and increasing degree of analogous conditions was recognized, which suggested *ad hoc* hypotheses for geographic and temporal idiosyncrasies.

For the purpose of species distribution modeling (SDM), variables showing high degrees of analogy of conditions tend to estimate broad geographic ranges and few values are frequent across geographic space (Peterson et al. 2011). For a given range or niche-breadth, SDM models using variables close to the upper-left side of the triangle would predict larger geographic expanses than those in the lower-right part of the triangle. This observation thus offers a cautionary note for studies relating niche breadth to distributional area without controlling for variable degrees of analogy of conditions (Slatyer et al. 2013).

In fact, this statement depends on each band of the variable histogram having a correlation with occurrences of species (Guisan and Thuiller

2005). In any case, a study that summarized environmental variables that most contributed to estimating species' geographic distributions showed that, for the WorldClim dataset, temperature seasonality, annual precipitation, and precipitation of the driest month had highest mean contributions (Bradie and Leung 2017). Interestingly, in this study, the former two variables were consistently ranked as showing low degree analogy for both extents and resolution; precipitation of the driest month showed an atypical trend, with high analogy for South America and low analogy for Araucaria Moist Forests (Fig. 3D), perhaps owing to odd contrasts in this variable in homogeneity across the two regions (Appendix A).

Further, it is common in processing raster layers to transform decimal values into integers by multiplying by 10, 100, or 1000 (Hijmans et al. 2005), which produces increasing variable heterogeneity. It is also common to use raster layers arranged into categorical, nominal, or ordinal classes (Peterson et al. 2011), which dramatically homogenizes variables. Increasing or decreasing numbers of bins on the environmental axis affects modal frequencies in the **E**-space kernel and therefore the degree of analogy of conditions of environmental layers.

In this paper, I have focused on quantification of analogy of conditions within the same variable layer as 'contemporary' analogous conditions. This quantification approach can be used to compare degrees of analogy of conditions in different time spans (Garcia et al. 2014). For a given variable with constant extent and resolution, the variable can be plotted for different spans, and degrees of analogous conditions in temporal scenarios of change can be ranked.

By analyzing how analogous conditions were coded in G and E, I used two parameters to characterize degree of analogy of conditions. The Euclidean distance between the candidate layer and the non-analogous layer provided a metric of dissimilarity used to rank and compare variables by their degree of analogy. The resulting information may be used to select layers and interpret results in species distribution and ecological niche modeling.

ACKNOWLEDGMENTS

I am grateful to L.R.R. Faria Jr., T. Vasconcelos, J. Soberón, and Town Peterson, for suggestions that improved the manuscript. This research was conducted in the Biogeography and Macroecology Lab (ILACVN/UNILA/Brazil).

REFERENCES

Ackerly, D.D., S.R. Loarie, W.K. Cornwell, S.B. Weiss, H. Hamilton, R. Branciforte, and N.J.B. Kraft. 2010. The geography of climate change: Implications for conservation biogeography. Diversity and Distributions, 16:476-487.

Aspinall, R. and B.G. Lees. 1994. Sampling and analysis of spatial environmental data. Proceedings of the Sixth International Symposium on Spatial Data Handling, vol 2. Edinburgh.

Anderson, R. P., and Raza, A. 2010. The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: Preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. Journal of Biogeography, 37:1378-1393.

Bradie, J., and B. Leung. 2017. A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. Journal of Biogeography, 44:1344-1361.

Chang, K.-T. 2017. Geographic Information System. The International Encyclopedia of Geography.[1]

Colwell, R.K. and T.F. Rangel. 2009. Hutchinson's duality: The once and future niche. Proceedings of the National Academy of Sciences USA 106:19651-19658.

Elith, J., and Graham, C. H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32:66-77.

Elith, J., and J.R. Leathwick. 2009. Species distribution models: Ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40:677-697.

Fitzpatrick, M.C., and W.W. Hargrove. 2009. The projection of species distribution models and the problem of non-analog climate. Biodiversity and Conservation 18:2255-2261.

Fitzpatrick, M.C., J.F. Weltzin, N.J. Sanders, and R.R. Dunn. 2007. The biogeography of prediction error: Why does the introduced range of the fire ant over-predict its native range? Global Ecology and Biogeography 16:24-33.

Gallien, L., R. Douzet, S. Pratte, N.E. Zimmermann, and W. Thuiller. 2012. Invasive species

[1] DOI: 10.1002/9781118786352.wbieg0152.

distribution models—how violating the equilibrium assumption can create new insights. Global Ecology and Biogeography 21:1126-1136.

Garcia, R. A., M. Cabeza, C. Rahbek, and M.B. Araújo. 2014. Multiple dimensions of climate change and their implications for biodiversity. Science 344:1247579.

Guisan, A., C.H. Graham, J. Elith, and F. Huettmann. 2007. Sensitivity of predictive species distribution models to change in grain size. Diversity and Distributions 13:332-340.

Guisan, A., B. Petitpierre, O. Broennimann, C. Daehler, and C. Kueffer. 2014. Unifying niche shift studies: Insights from biological invasions. Trends in Ecology and Evolution 29:260-269.

Guisan, A., and N.E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147-186.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: Offering more than simple habitat models. Ecology Letters 8:993-1009.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965-1978.

Hijmans, R. and J. Elith. 2017. Species Distribution Modeling with R.[2]

Hutchinson, G.E. 1978. An introduction to population ecology. Yale University Press, New Haven.

Jackson, S.T. and J.T. Overpeck. 2000. Responses of plant populations and communities to environmental changes of the late Quaternary. Paleobiology 26:194-220.

Legendre, P. 1993. Spatial autocorrelation: Trouble or new paradigm? Ecology 74:1659-1673.

de Oliveira, G., T.F. Rangel, M.S. Lima-Ribeiro, L.C. Terribile, and J.A.F. Diniz-Filho. 2014. Evaluating, partitioning, and mapping the spatial autocorrelation component in ecological niche modeling: A new approach based on environmentally equidistant records. Ecography 37:637-647.

Oksanen, J., F.G. Blanchet, M. Friendly, P. Kindt, P. Legendre, D. McGlinn, P.R. Minchin, R.B. O'Hara, G.L. Simpson, P. Solymos, M.H.H. Stevens, E. Szoecs, and H. Wagner. 2007. Community Ecology Package: Ordination, Diversity and Dissimilarities, the vegan package for R, version 2.4.[3]

Olson, D.M., E. Dinerstein, E.D. Wikramanayake, N.D. Burgess, G.V. Powell, E.C. Underwood, J.A. D'Amico, I. Itoua, H.E. Strand, J.C. Morrison, C.J. Loucks, T.F. Allnutt, T.H. Ricketts, J.F. Lamoreux, W.W. Wettengel, P. Hedao, and K.R. Kassem. 2001. Terrestrial ecoregions of the world: a new map of life on Earth. BioScience 51:933-938.

Overpeck, J.T., R.S. Webb, and T. Webb. 1992. Mapping eastern North American vegetation change of the past 18 ka: No-analogs and the future. Geology 20:1071-1074.

Peterson, A.T., J. Soberón, R.G. Pearson, R.P. Anderson, E. Martínez-Meyer, M. Nakamura, and M.B. Araújo. 2011. Ecological niches and geographic distributions. Princeton University Press, Princeton.

Randin, C.F., T. Dirnböck, S. Dullinger, N.E. Zimmermann, M. Zappa, and A. Guisan. 2006. Are niche-based species distribution models transferable in space? Journal of Biogeography 33:1689-1703.

Shannon, C. E. 2001. A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review 5:3-55.

Slatyer, R.A., M. Hirst, and J.P. Sexton. 2013. Niche breadth predicts geographical range size: A general ecological pattern. Ecology Letters 16:1104-1114.

Soberón, J. and M. Nakamura. 2009. Niches and distributional areas: Concepts, methods, and assumptions. Proceedings of the National Academy of Sciences USA 106:19644-19650.

Soberón, J., L. Osorio-Olvera, and A.T. Peterson. 2017. Diferencias conceptuales entre modelación de nichos y modelación de áreas de distribución. Revista Mexicana de Biodiversidad 88:437-441.

Soberón, J., and A.T. Peterson. 2011. Ecological niche shifts and environmental space anisotropy: A cautionary note. Revista Mexicana de Biodiversidad 82:1348-1355.

Thuiller, W., L. Brotons, M.B. Araújo, and S. Lavorel. 2004. Effects of restricting environmental range of data to project current and future species distributions. Ecography 27:165-172.

Williams, K.J., L. Belbin, L., M.P. Austin, J.L. Stein, and S. Ferrier. 2012. Which environmental variables should I use in my biodiversity model? International Journal of Geographical Information Science 26:2009-2047.

Zhou, J. and K.-M. Lau. 1998. Does a monsoon climate exist over South America? Journal of Climate 11:1021-1040.

[2] https://cran.r-project.org/web/packages/dismo/vignettes/sdm.pdf.

[3] https://cran.r-project.org/web/packages/vegan/vegan.pdf.

APPENDIX 1: HISTOGRAMS OF BIOCLIMATIC VARIABLES.

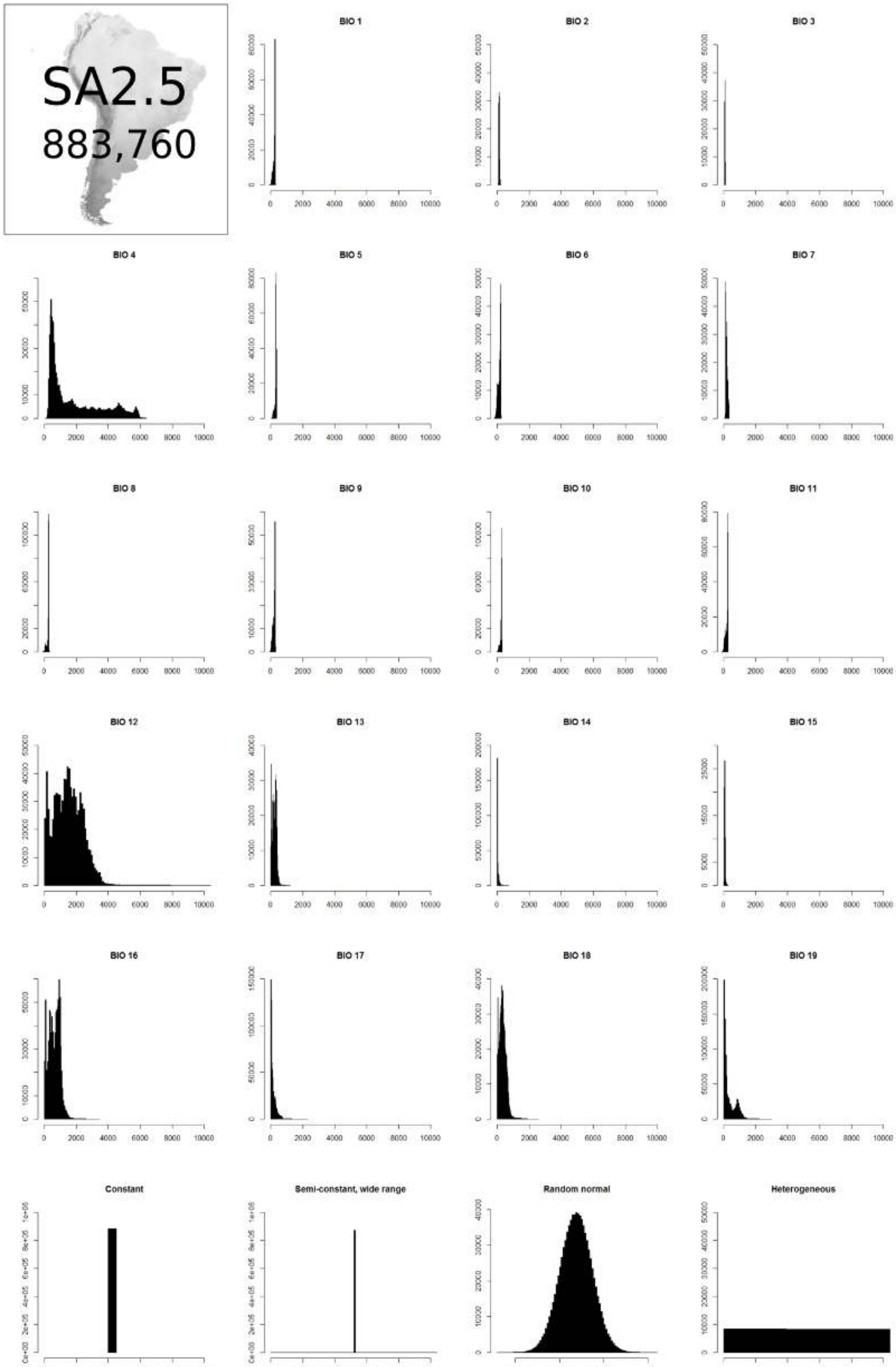

Figure A.1 Bioclimatic variables and reference variables presented in histograms for the extent of all of South America at 2.5’ resolution, with 883,760 pixels.

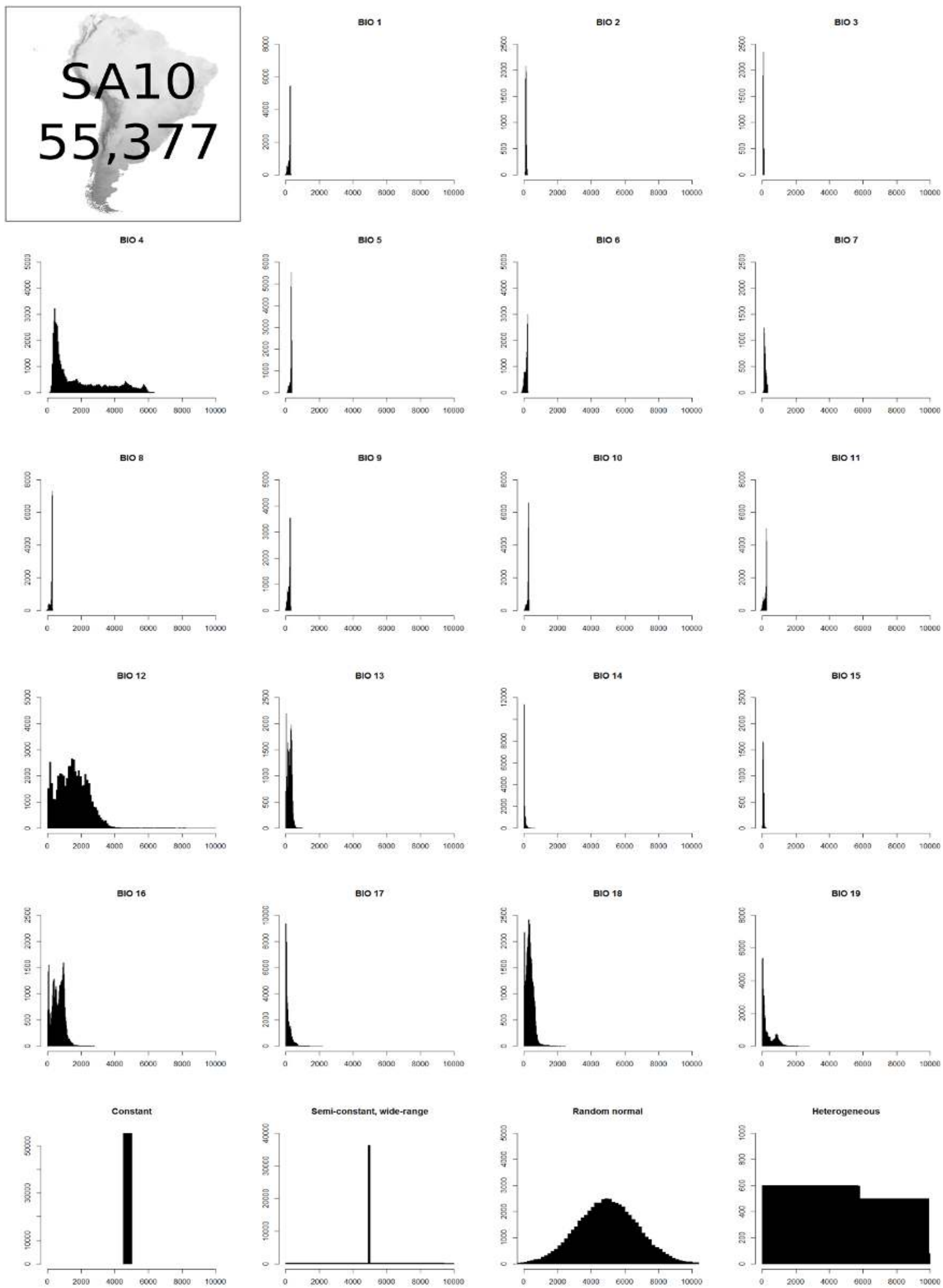

Figure A.2 Bioclimatic variables and reference variables presented in histograms for the extent of all of South America at 10' resolution, with 55,377 pixels.

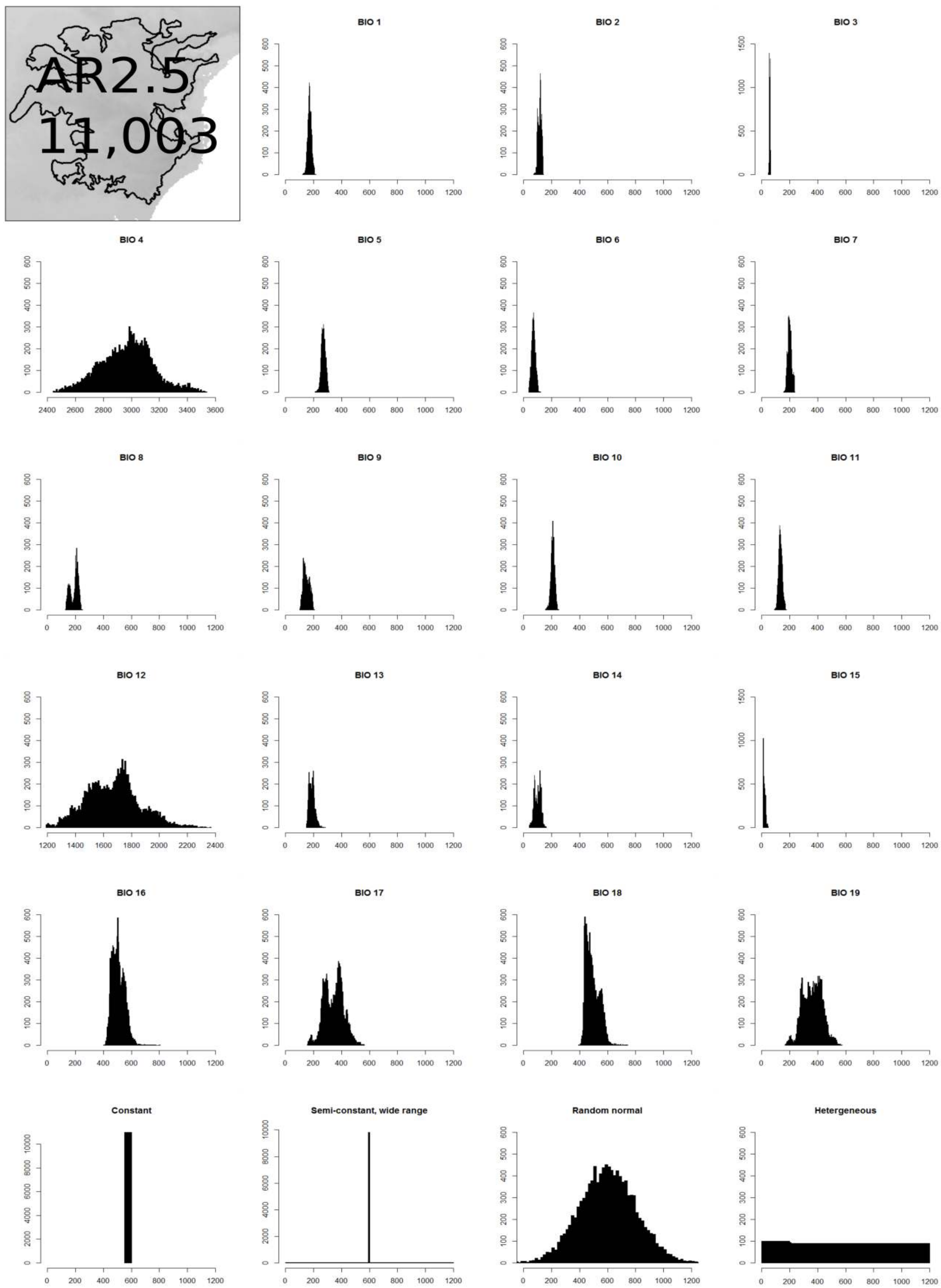

Figure A.3 Bioclimatic variables and reference variables presented in histograms for the extent of the Araucaria Moist Forest ecoregion at 2.5’ resolution, with 11,033 pixels.

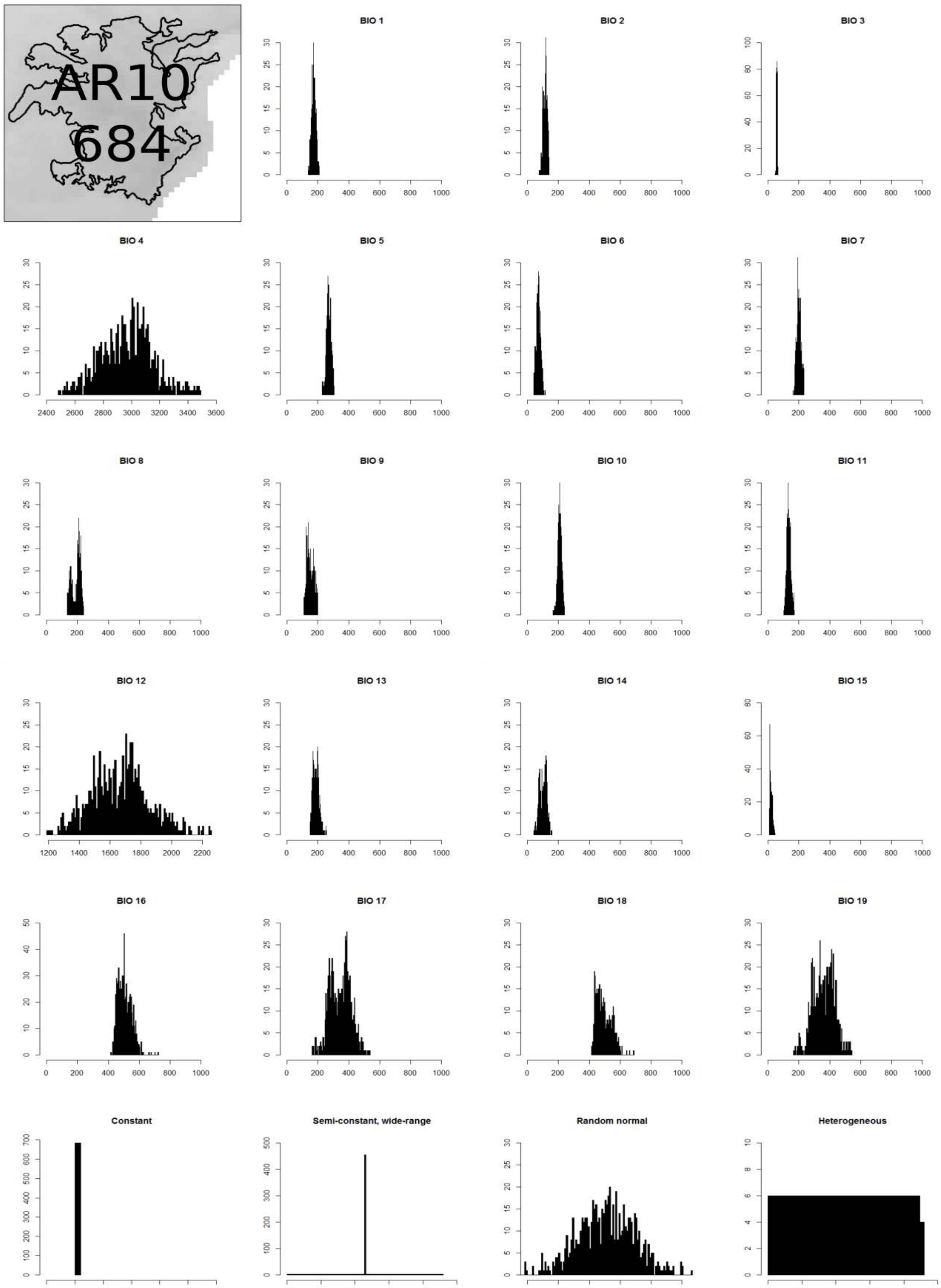

Figure A.4 Bioclimatic variables and reference variables presented in histograms for the extent of the Araucaria Moist Forest ecoregion at 10' resolution, with 684 pixels.

APPENDIX 2: VARIABLE MEASUREMENTS.

Table B.1. Measurements for the extent of all of South America at a 2.5' spatial resolution. Bioclimatic and reference variables: CO = constant, SC = semi-constant and wide range, RN = random normal, HT = heterogeneous, NAN = non-analogous.

Variable	Range	Modal frequency	Diversity	Distance to NAN	Log distance
BIO1	399	87311	13.6680	1087163.5	7.3081
BIO2	202	33077	13.6732	1082891.1	7.1080
BIO3	80	37357	13.6732	1083249.1	7.4754
BIO4	6283	51005	13.3247	1076499.4	6.3375
BIO5	355	88750	13.6736	1087391.9	7.3503
BIO6	458	47793	13.6567	1083401.9	7.0026
BIO7	286	48816	13.6397	1083680.7	7.1585
BIO8	434	118219	13.6706	1091494.6	7.4176
BIO9	380	55925	13.6552	1084081.0	7.1281
BIO10	371	105921	13.6729	1089675.5	7.4149
BIO11	435	79219	13.6589	1086189.6	7.2395
BIO12	10577	42407	13.5000	1070686.8	6.1368
BIO13	1197	34742	13.5276	1081751.6	6.5779
BIO14	697	182027	13.0964	1104264.7	7.4798
BIO15	259	26691	13.5697	1082557.0	6.9363
BIO16	3450	59923	13.5211	1080650.0	6.5528
BIO17	2319	149545	13.1747	1094966.9	7.0824
BIO18	2574	38118	13.4865	1080237.2	6.4129
BIO19	2962	198539	13.1451	1105818.1	7.1615
CO	1	883760	13.6919	1530715.5	10.2994
SC	10600	873260	13.6896	1512443.2	7.6473
RN	17787	39148	13.4992	1061679.1	5.9963
HT	10600	8400	13.4977	1069447.7	5.3512
NAN	883760	1	13.4988	0.0	0.0000

Table B.2. Measurements for the extent of all of South America at a 10' spatial resolution. Bioclimatic and reference variables: CO = constant, SC = semi-constant and wide range, RN = random normal, HT = heterogeneous, NAN = non-analogous.

Variable	Range	Modal frequency	Diversity	Distance to NAN	Log distance
BIO1	355	5460	10.86252	67718.76	5.3070
BIO2	167	2083	10.90318	67666.22	5.1040
BIO3	55	2352	10.90312	67816.49	5.5296
BIO4	6214	3221	10.55476	60341.14	4.4511
BIO5	303	5531	10.90393	67790.77	5.3560

BIO6	414	2999	10.75527	67415.71	4.9916
BIO7	280	1243	10.86969	67496.91	4.7193
BIO8	395	7300	10.85650	67929.69	5.4130
BIO9	340	3527	10.84392	67544.47	5.1200
BIO10	318	6600	10.88888	67915.83	5.4236
BIO11	394	5013	10.80168	67619.34	5.2398
BIO12	9916	2665	10.73032	55773.64	4.2942
BIO13	980	2199	10.75779	66676.81	4.6217
BIO14	652	11355	10.32861	68451.50	5.4999
BIO15	232	1657	10.79936	67569.00	4.9018
BIO16	2787	1592	10.75131	64438.80	4.2316
BIO17	2159	9357	10.40633	66178.05	5.1607
BIO18	2427	2418	10.71755	64917.76	4.4653
BIO19	2787	5381	10.37613	64745.49	4.8381
CO	1	55377	10.92192	95914.04	8.2157
SC	9915	36308	10.85571	71256.49	5.6593
RN	23713	2443	10.72992	38895.48	4.1738
HT	9915	600	10.71872	55684.18	3.5234
NAN	55377	1	10.72878	0.00	0.0000

Table B.3. Measurements for the extent of the Araucaria Moist Forest ecoregion at a 2.5' spatial resolution. Bioclimatic and reference variables: CO = constant, SC = semi-constant and wide range, RN = random normal, HT = heterogeneous, NAN = non-analogous.

Variable	Range	Modal frequency	Diversity	Distance to NAN	Log distance
BIO1	93	426	9.3029	13372.10	4.1008
BIO2	70	464	9.3002	13402.14	4.2311
BIO3	19	1397	9.3044	13560.81	5.1271
BIO4	1092	302	9.3039	12144.04	3.2765
BIO5	102	312	9.3044	13356.38	3.9409
BIO6	85	365	9.2870	13379.19	4.0669
BIO7	83	351	9.3035	13381.08	4.0589
BIO8	125	284	9.2944	13327.28	3.8341
BIO9	133	237	9.2950	13316.11	3.7384
BIO10	98	409	9.3039	13365.19	4.0666
BIO11	86	387	9.3014	13378.89	4.0870
BIO12	1176	313	9.2993	12041.63	3.2797
BIO13	139	259	9.3007	13309.38	3.7606
BIO14	124	261	9.2798	13327.80	3.8017
BIO15	40	1023	9.2325	13485.09	4.7450
BIO16	397	587	9.3020	13009.46	3.8236
BIO17	406	385	9.2853	12987.14	3.6204

BIO18	347	590	9.3013	13070.80	3.8596
BIO19	402	317	9.2873	12989.29	3.5329
CO	1	11003	9.3059	19056.02	7.0001
SC	1200	9813	9.2849	16987.09	5.0290
RN	1200	453	9.1950	12018.93	3.4600
HT	1200	100	9.1034	12006.79	2.7183
NAN	11003	1	9.1128	0.00	0.0000

Table B.4. Measurements for the extent of the Araucaria Moist Forest ecoregion at a 10' spatial resolution. Bioclimatic and reference variables: CO = constant, SC = semi-constant and wide range, RN = random normal, HT = heterogeneous, NAN = non-analogous.

Variable	Range	Modal frequency	Diversity	Distance to NAN	Log distance
BIO1	72	30	6.5251	750.38	2.1695
BIO2	65	36	6.5223	759.33	2.2804
BIO3	18	86	6.5265	822.30	3.0589
BIO4	1002	22	6.5260	390.32	1.6566
BIO5	76	27	6.5265	745.33	2.1069
BIO6	73	28	6.5100	749.05	2.1349
BIO7	73	35	6.5255	749.48	2.2344
BIO8	108	22	6.5167	705.92	1.9150
BIO9	92	21	6.5170	725.46	1.9393
BIO10	75	30	6.5261	746.71	2.1576
BIO11	67	30	6.5237	756.50	2.1909
BIO12	1062	21	6.5215	463.74	1.6362
BIO13	104	20	6.5229	710.73	1.8822
BIO14	117	18	6.5021	694.74	1.8016
BIO15	39	67	6.4550	794.09	2.7061
BIO16	307	46	6.5242	465.01	2.0806
BIO17	379	28	6.5075	375.01	1.8000
BIO18	274	19	6.5234	502.63	1.6400
BIO19	377	26	6.5096	377.24	1.7617
CO	1	684	6.5280	1182.99	4.9105
SC	230	456	6.4635	787.22	3.3077
RN	1062	19	6.3412	463.48	1.5835
HT	230	6	6.3369	556.07	1.1155
NAN	684	1	6.3355	0.00	0.0000

APPENDIX 3: PEARSON CORRELATION COEFFICIENTS.

Table C.1. Correlation coefficients for the parameters of two extents (South America SA, and Araucaria Moist Forests ecoregion AR) and two spatial resolutions (2.5’, 10’). NAN = non-analogous variable.

SA2.5	Modal frequency	Diversity	Distance to NAN	Log distance
Range	-0.1226	-0.0497	-0.8786	-0.8591
Modal frequency	1	0.1363	0.5759	0.4805
Diversity	0.1363	1	0.1488	0.1713
Distance to NAN	0.5759	0.1488	1	0.9259
SA10				
Range	-0.1218	-0.0818	-0.9207	-0.8089
Modal frequency	1	0.1520	0.4938	0.6196
Diversity	0.1520	1	0.1674	0.1884
Distance to NAN	0.4938	0.1674	1	0.9445
AR2.5				
Range	-0.0782	-0.6920	-0.7546	-0.8806
Modal frequency	1	0.1317	0.6270	0.5536
Diversity	0.1317	1	0.6466	0.6317
Distance to NAN	0.6270	0.6466	1	0.9278
AR10				
Range	-0.204	-0.1892	-0.6936	-0.4737
Modal frequency	1	0.0608	0.5563	0.8129
Diversity	0.0608	1	0.4542	0.4736
Distance to NAN	0.5563	0.4542	1	0.8559