Niche modeling of a freshwater fish species: Estimating and mapping the uncertainty among modeling methods and freshwater variables.

Micael Rosa Parreira[1]* (PG), João Carlos Nabout[1] (PQ), Geiziane Tessarolo[1] (PQ), Fabrício Barreto Teresa[1] (PQ). *E-mail: micael_rp@hotmail.com

[1]Universidade Estadual de Goiás, Campus de Ciências Exatas e Tecnológicas (CCET), BR-153, nº 3.105, CEP 75132-903, Anápolis, GO, Brasil.

ENMs are increasingly accurate, which give us more reliability to define environmentally suitable areas to species. Although, it is necessary to remember that even the predicted suitable areas are suffused with uncertainty and thus, may not indicate the "real" effect of the environment over the species dynamics. Considering this, the aim of this work is to assess which factors create more uncertainties on freshwater large fish species modelling: methods or climate variables. To predict the species distribution, we collected the occurrence sites of *Brachyplathystoma filamentosum* and 1 km freshwater-specific variables available in climate and environmental groups. The suitability maps showed that the areas with great environmental suitability values were the Amazon large rivers, and nearby areas. Moreover, the Amazon basin presented high uncertainty values for the methods component, while for the variables the uncertainty mapped in this area was lower. The methods and variables were responsible for 46% and 40% of the uncertainty. Therefore, due to the predictions' uncertainty presented, it is necessary to be caution in choosing the variables and methods to model a specie distribution. Moreover, we emphasize the importance of using uncertainty analysis to verify the accuracy of ENMs in future works.

Keywords: ENMs. Hierarchical ANOVA. *Piraiba*. Freshwater variables. Variability.

Introduction

Methodological uncertainties on ENMs are a topic widely tested and discussed due to variable number and characteristics of databases and methods used to niche modelling (Buisson et al. 2010). The ENMs uncertainty can be addressed to various factors, some of them tested in recent papers, such as methods (Diniz-Filho et al. 2009), climate and environmental variables (Stoklosa et al. 2015), biases in databases (Tessarolo et al. 2014), emission scenarios/AOGCMs (Diniz-Filho et al. 2009; Buisson et al. 2010).

By the increasing of researches' interest on niche modelling, there was consequently, a great expansion in the development of modelling methods and climate variables. All methods have benefits and disadvantages depending on the data and aim of the work (Araújo e Guisan 2006). All of them, however, generate methodological uncertainties which in most of the papers are ignored or often only partially considered (Buisson et al. 2010; Beale and Lennon 2010). Similar to the methods, climate variables have also been focus with increased development, whether datasets based on present measurements (Hijmans et al. 2005), compilation of multiple climate models – focusing on more accurate past models (see Lima-Ribeiro et al. 2015) or climate and environmental-specific variables (see Domisch et al. 2015 for freshwater data).

Because most papers addressing species distribution lack uncertainties analysis in their predictions, there is a need of papers to address the requirement of test for uncertainties in order to obtain more reliable models for prediction. Therefore, the aim of this paper is to determine which factors deal more uncertainties on ecologic niche modelling of species of large freshwater fishes: the modelling methods (algorithms) or each group of freshwater variables.

Material and Methods

For modeling purposes, we used basically two online datasets, one regarding the species distribution which was collected in three online databases: [*Species Link, Gbif* and *Fishbase*]. Second, we used freshwater climate and environmental layers available in 1km resolution (*http://www.earthenv.org*). For modeling procedures, more details are available in the figure 1.

For data analysis, we used a hierarchical ANOVA (Diniz-Filho et al. 2009), using two factors (methods and variables) to assess uncertainty in the prediction models. This ANOVA calculates the variance between each cell of the spatial grid. After the analysis, we calculated the proportion of the sum of mean squares explained by each component. Lastly, we had those proportions values plus the coordinates for each cell and thus, we converted into a raster stack to map the uncertainties to the extension map for each factor using the uncertainty percentages as shown in the figure 2b.

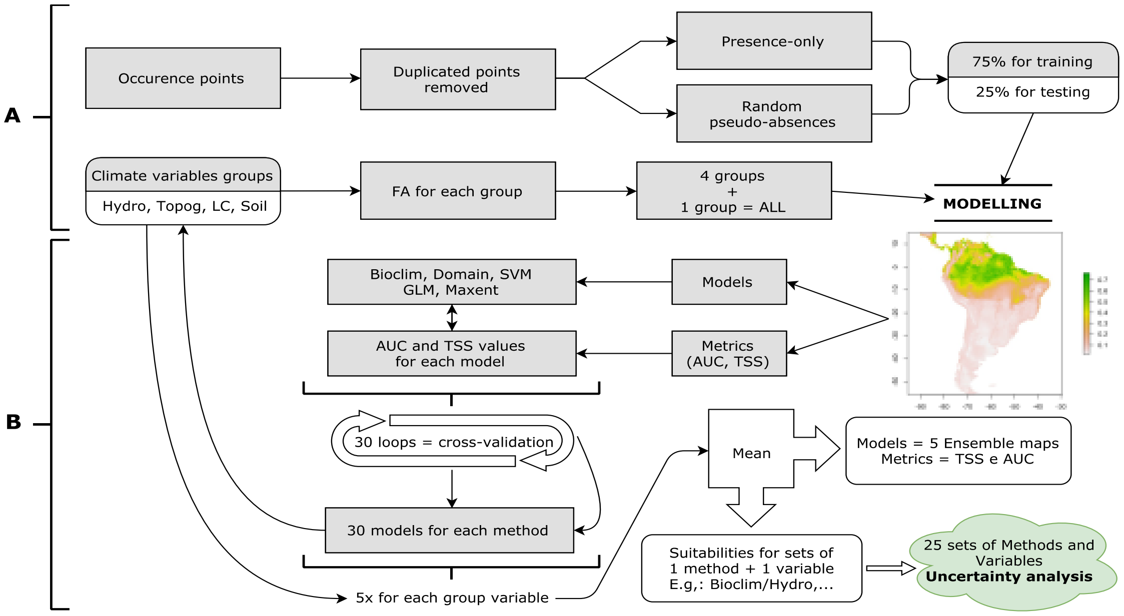

Figure 1- Flowchart representing the modelling methodology divided into two steps: A) Representing data preparation and B) Modeling procedures.

Results and Discussion

In total, we built 5 ensemble maps (weighted maps by the metrics values of all predictions) each one for a variable group (Figure 2a). These maps show distinct predictions depending on which variable group the researcher uses to predict the potential distribution of a freshwater species. Although, all of them seem to interpolate well for the species core distribution, they vary considerably to extrapolate (areas off the species core distribution).

The results of the ANOVA using the suitability values for each grid cell were plotted in gradient maps of each component in order to assess the predictions uncertainties (Figure 2b). Individually, the methods presented higher uncertainties on predictions located in the South American's western side (Andes). This component also presented relative high uncertainties in the core area of the currently species distribution. The second component (variables) assessed low uncertainties values in the core area of the species distribution. Thus, both components used acts like complements for each other in uncertainties assessments.

The methods component had higher uncertainty proportion in the niche models (~46%). Nevertheless, there was not significant difference between both components, due to the high uncertainty of the variables as well (~40%). This result corroborates the map visual analysis of the geographic uncertainty variation, i.e., that both components are complementary in the variation of the models.

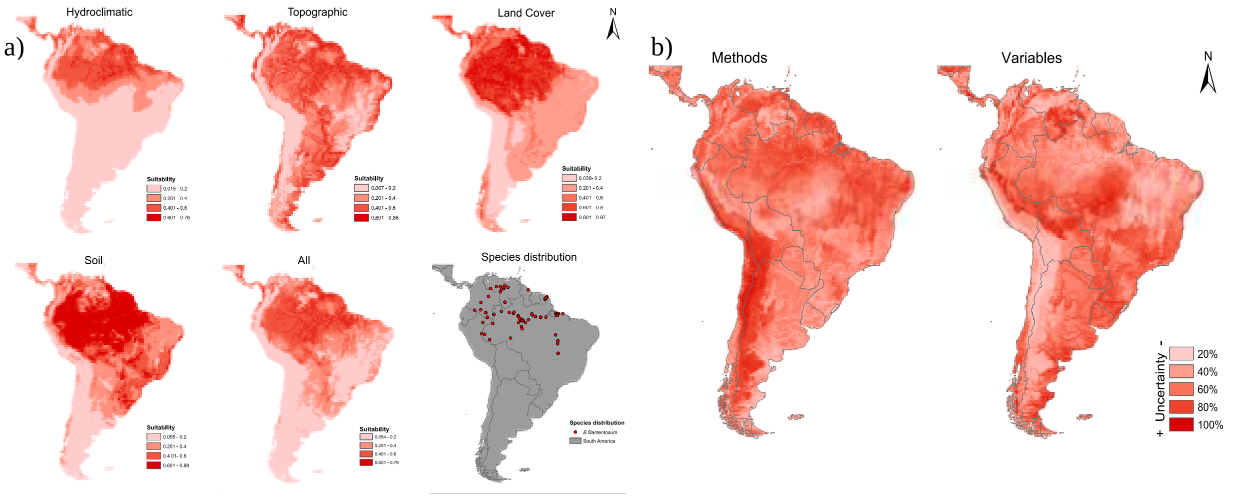

Figure 2 – a) Suitability maps for each variable group available online plus all variables together. The actual distribution of *Brachyplatystoma filamentosum* is plotted in the last map. b) Uncertainty maps for each component of the ENMs: Methods/algorithms and Climate variable groups.

Final Considerations

In this work, the uncertainty analysis showed that both methods and variables have similar contribution to the models' uncertainty based on the hierarchical ANOVA output. Additionally, both components act as complement in the mapped uncertainty, i.e. areas that show high uncertainty due to the methods are low for the variables and vice versa. Moreover, the uncertainty analysis can be a substantial addition to models' validation based-only on the accuracy metrics.

Therefore, due to the predictions' uncertainty presented, it is necessary to be caution in choosing the variables and methods to model a specie distribution. Moreover, we emphasize the importance of using uncertainty analysis to verify the accuracy of models in future ENM's articles.

Acknowledgements

We acknowledge the CAPES financial support to MRP regarding his master's program.

References

ARAÚJO, M. B.; GUISAN, A. Five (or so) challenges for species distribution modelling. **Journal of Biogeography**, v. 33, n. 10, p. 1677–1688, out. 2006.

BEALE, C. M.; LENNON, J. J. Incorporating uncertainty in predictive species distribution modelling. **Philosophical Transactions of the Royal Society B**, v. 367, n. 1586, p. 247-258, jan. 2012.

BUISSON, L.; THUILLER, W.; CASAJUS, N.; LEK, S. and GRENOUILLET, G. (2010), Uncertainty in ensemble forecasting of species distribution. **Global Change Biology**, v. 16, n. 4, p. 1145–1157, abril. 2010.

DINIZ-FILHO, J. A. F.; BINI, L. M.; RANGEL, T. F.; LOYOLA, R. D.; HOF, C.; NOGUÉS-BRAVO, D.; ARAÚJO, M. B. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. **Ecography**, v. 32, n. 6, p. 897–906, dez. 2009.

DOMISCH, S.; AMATULLI, G.; JETZ, W. Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. **Scientific Data,** v. 2, n. 150073, dez. 2015.

HIJMANS, R. J.; CAMERON, S. E.; PARRA, J. L.; JONES P.G.; JARVIS, A. Very high resolution interpolated climate surfaces for global land areas. **International Journal of Climatology**, v. 25, n. 15, p. 1965-1978, dez. 2005.

LIMA-RIBEIRO, M. S.; VARELA, S.; GONZÁLEZ-HERNÁNDEZ, J.; OLIVEIRA, G.; DINIZ-FILHO, J.A.F.; TERRIBILE, L.C. ecoClimate: a database of climate data from multiple models for past, present, and future for Macroecologists and Biogeographers. **Biodiversity Informatics**, v. 10, p. 1-21, ago. 2015.

STOKLOSA, J.; DALY, C.; FOSTER, S. D.; ASHCROFT, M. B.; WARTON, D. I. A climate of uncertainty: accounting for error in climate variables for species distribution models. **Methods in Ecology and Evolution**, v. 6, n. 4, p. 412–423, abril. 2015.

TESSAROLO, G.; RANGEL, T. F.; ARAÚJO, M. B.; HORTAL, J. Uncertainty associated with survey design in Species Distribution Models. **Diversity and Distributions**, v. 20, n. 11, p. 1258–1269, nov. 2014.