**Title:**

Predictability and transferability of local biodiversity-environment relationships across the tree of life

**Authors:**

Martin Jung1

1 International Institute for Applied Systems Analysis (IIASA) - Schlossplatz 1 - A-2361 Laxenburg, Austria [jung@iiasa.ac.at](mailto:jung@iiasa.ac.at)

**Abstract:**

Here I combine

**Keywords:** Spectral-diversity, productivity, transferability, PREDICTS, remote-sensing, extrapolation

**Introduction:**

- Introduction to species-environment relationships & Mechanisms

Biodiversity is known to vary along environmental gradients (Hillebrand 2004; Holt *et al.* 2017).

This variation

Ecologists have long postulated a number of hypotheses as possible source of this variation. These include, among others, the widely tested species-energy hypothesis (Hurlbert 2004; Evans *et al.* 2005; Duncan *et al.* 2015), the species spectral-heterogeneity hypotheses (Oldeland *et al.* 2010; Rocchini *et al.* 2010) and the species-geodiversity relationship (Theobald *et al.* 2015; Alahuhta *et al.* 2020).

And while a number of empirical studies exist that investigated these hypotheses using local biodiversity measures, such as species richness or abundance, rarely has been investigated whether those hypotheses hold across multiple taxonomic groups and/or geographic regions.

- Previous studies on the relationships

Spatial and temporal

- Predictability and transferability

According to Yates *et al.* (2018) ‘transferability’ is broadly defined as the capacity of a model to produce *predictions for a new set of predictors that differ from those on which the model was trained*.

In spatial ecological niche models (Petchey *et al.* 2015; Moon *et al.* 2017)

Predictability

For a number of species-environment relationships there is increasing evidence that environmental covariates often fail to accurately predict local biodiversity in new areas (Parmentier *et al.* 2011; Schmidtlein & Fassnacht 2017). With even pressures such as land use been found to be highly variable in geographic space and local setting (Jung *et al.* 2017; Phillips *et al.* 2017).

This is especially relevant, since a number of studies spatially extrapolated local biodiversity estimates, e.g. species richness or abundance, to unsampled areas (Newbold *et al.* 2016; König *et al.* 2017; Phillips *et al.* 2019; van den Hoogen *et al.* 2019). These approaches thus assume that local biodiversity-environment relationships are transferable to new environments, which might not necessarily be the case.

The question of transferability has also long been investigated in the context of species distribution modelling (Mesgaran *et al.* 2014; Regos *et al.* 2019). Tools have been developed to quantify and visualize the uncertainty introduced by projecting trained models to novel environments in space and time (Zurell *et al.* 2012; Mesgaran *et al.* 2014).

A better understanding of the transferability of ecological models to novel context is empirical, if the aim is to produce forecasts of the impact of likely future impacts of global change.

- Remote sensing as a means to assess this relationship

Advances it remote sensing and cloud-processing increasingly enable the robust characterization of environmental heterogeneity for biodiversity science (Gorelick *et al.* 2017; Randin *et al.* 2020). Using repeated satellite measurements it has become possible to robustly quantify measures of environmental heterogeneity, such as differences in vegetation availability or spectral dissimilarity (Rocchini *et al.* 2010; Radeloff *et al.* 2019). Subsequently, these measures have been incorporated in statistical models for the prediction of species distributions (Cord *et al.* 2013; He *et al.* 2015) and also differences in local biodiversity (Oldeland *et al.* 2010; Rocchini *et al.* 2015; Jung *et al.* 2019a, b).

Environmental heterogeneity (Stein *et al.* 2014)

While spatial cross-validation can be used to robustly estimate uncertainty in predictor space (Roberts *et al.* 2017), few macroecological studies based on local biodiversity data report or map uncertainty, which can ultimately hinder their application (Rapacciuolo 2019).

Biodiversity-environment relationships are likely also affected by differences in scale

Across taxonomic

Many tested only for few selected ones

It is therefore an important question, whether biodiversity-environment relationships are transferable across taxonomic groups and geographic regions.

In this study I investigate the predictability and transferability of two commonly found environmental relationships with biodiversity measures. The expectation is that (a) predictability is generally larger than transferability, (b) species-environmental relationships are less transferable between different environmental spaces than between taxonomic groups, and that (c) any unexplained variation is predominantly linked to differences in functional and taxonomic groups as well as study design, e.g. spatial scale and sampling duration.

To test this, I combine local biodiversity data from multiple taxonomic groups from the PREDICTS database (Hudson *et al.* 2014, 2017) with remotely-sensed environmental predictors quantifying vegetation availability (Evans *et al.* 2005; Stein *et al.* 2014; Duncan *et al.* 2015) and spectral variability (Rocchini *et al.* 2010). Using hierarchical Bayesian models, I assess the predictability, quantified as overall and within-study reduction in prediction error, and transferability, quantified as reduction in prediction error between different studies of comparable study design, within the same and different environmental spaces and taxonomic groups (Figure 1).

I expect that the results of this work provide further insights into the generality of local biodiversity-environment relationships at a global scale, which hopefully stimulates a debate on whether local biodiversity measures, such as abundance or richness, can accurately be predicted or even extrapolated in unsampled regions.

<Figure 1: Schematic of approach. Show missing and permutated data. Link to remote sensing data and spatial differences in assemblages. Use a globe illustration with phylopic icons to illustrate prediction within and among studies>

**Methods:**

*Biodiversity data preparation*

I took species assemblage data from the global **P**rojecting **R**esponses of **E**cological **D**iversity **I**n **C**hanging **T**errestrial **S**ystems (PREDICTS) database (Hudson *et al.* 2017), which contains records of species occurrence and abundance at spatial-explicit sites ‘sites’ as reported in published ‘studies’. Only studies which differ in ‘land-use’ and have spatial and temporal information associated with them, e.g. sampling extent and date of sampling, were included in the PREDICTS database (Hudson *et al.* 2014).

Studies in the PREDICTS database vary widely in scope, notably in taxonomic coverage (e.g. invertebrates, plants, birds, … ), spatial grain (0.05 – 39,150m, median = 60m), sampling start (1984 - 2013) and effort (>0 – 4,382 days, median = 91 days) or methodology (flight traps, transects, …). Because of these differences, previous models utilizing PREDICTS data have exclusively relied on relative comparisons between studies (Purvis *et al.* 2018), where site-based biodiversity measures are typically compared against a (discrete) reference site within the same study (Newbold *et al.* 2015; Gray *et al.* 2016; Jung *et al.* 2019b). Comparisons of relative differences in biodiversity have obvious limitations, in that they require baselines, which often vary with local factors (Jung *et al.* 2017), and furthermore easily cause misinterpretations of the magnitude of effect.

In this study I aim to investigate whether environmental covariates can help to predict site-based biodiversity measures within and between studies. I created groups of comparable studies based on the taxonomic information and study methodology, for instance birds that were counted through point counts.

To account for differences in spatial grain and sampling effort, I used the Mobr package (REF) to create cross-validated species accumulation curves. Previous studies using PREDICTS data assumed that within-study abundance of species individuals increases linearly with sampling effort (Newbold *et al.* 2015). Finally, I calculated for each site *j* two commonly used measures of local biodiversity, a rarefied estimate of the total number of species r and the rarefied log10-transformed abundance r of all species.

*Environmental predictors*

For each PREDICTS site I calculated two different remotely-sensed predictors that reflect the overall heterogeneity of environmental conditions. First, 16-day time series of atmospherically corrected spectral data (MCD43A v006, [Schaaf *et al.* 2002]) from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor on board the Terra and Aqua satellites were downloaded for each PREDICTS site from Google Earth Engine (Gorelick *et al.* 2017). From the obtained time series, I then selected the study and site specific first year of spectral data before biodiversity sampling.

Gap filling

No gap filling of missing spectral data was conducted as overall proportion of missing data was low (mean: 21.9% ± 22.7 SD) and aggregated in sub-sequential steps of the analysis. Nevertheless we removed sites with more than half of remote sensing measures missing, assuming that aggregated annual statistics are less precise, resulting in a total of 19954 sites for analyses.

Second, I then calculated from the acquired time series, as proxy of overall vegetation availability, the arithmetic mean of the two-band Enhanced Vegetation Index (EVI, Jiang *et al.* 2008). Variations in vegetation availability have previously been shown to reflect continuous gradients in land cover (Huete *et al.* 2002; Radeloff *et al.* 2019) and directly influence local biodiversity measures and life history (Pettorelli *et al.* 2005; Oldeland *et al.* 2010; Jung *et al.* 2019a, b). Furthermore, I also calculated a measure of overall spectral heterogeneity from the spectral data (Rocchini *et al.* 2010, 2015; Randin *et al.* 2020). Spectral heterogeneity is expected to give a more nuanced view on habitat structure and condition than any single vegetation index. To capture spectral heterogeneity I first conducted a principal component analysis of all spectral bands (1-7) and then from the first two bands, which on average explained 91.3% ± 5.93 SD of the variation, calculated the centroid of the resulting scatter plot. Spectral heterogeneity per site was then summarized as the mean Euclidean distance to this centroid. Both environmental predictors were reasonably uncorrelated (SI Figure 1).

*Statistical analysis*

For the statistical analysis I relied on hierarchical Bayesian linear models using a

with intercepts varying per study and/or sampling methodology and taxonomic grouping.

3 scenarios

Individual

Partial pooling per biome (SS|biome)

Pooling per

Absolute measure of error

Run through another model to detect most important variables

Bayesian

Need to vary random intercepts

Study level intercepts makes prediction impossible across studies. Find replacement through accounting for sampling methodology.

Or study as random slope! Nested within habitat type

Cross-validation

In the context of this study, I define ‘Predictability’ as the ability to accurately infer *y* based on environmental predictors *x* within a PREDICTS study and ‘Transferability’ as the ability to predict *y* with *x* across studies with the same sampling methodology and taxonomic group (Figure 1). The explained variance and precision of both predictions was assed using Bayesian R² and Root-Mean-Square Error (RMSE) metrics and compared them between taxonomic groups and biomes.

Classification tree on residuals -> find most influential variables on errors

Variables: Habitat type dominant, climate, traits (body size, taxonomic group), scale, time period, sampling duration

All Bayesian statistical models were fitted constructed using the ‘brms’ package (Bürkner 2017) and

MCMC stats, parameters

**Results:**

Figure 2: Overall correlation strength within study and across studies

Figure 3: Across taxonomic groups with their R-square

Figure 4: Path Model? Or classification tree of residuals

Figure 5: MESS or similar globally for all predictors. Insert points. Below the 3 maps show a simple boxplot (vertical) highlighting the proportion of terrestrial land area outside the environmental surface. Or scatterplot of environmental surface with the points highlightedg

* Show average fitted R2 across combinations
* Show cross-validated prediction

**Discussion:**

Do not quantify land-use as used by many common models, however the situation should be similar

Ecological models can and should be used for predictions (Houlahan *et al.* 2017), however caveats and limitations should be better identified, communicated and hopefully build upon. According to Yates *et al.* (2018) ‘transferability’ is broadly defined as the capacity of a model to produce *predictions for a new set of predictors that differ from those on which the model was trained*.

Especially recently predicted global relationships of local data across scales can be problematic. Uncertainty unknown!

Data was not specifically gathered to test these hypotheses

Species environment relationships do not equate relationships to pressures.

Another limitation of the presented results is that preceding changes in remotely-sensed environmental predictors have generally been ignored. In previous work I have found that observed differences in local biodiversity tend – to a measurable degree – be influenced by preceding changes in remotely-sensed vegetation availability (Jung *et al.* 2019b, a).

**Conclusion:**

Improve existing models to account for underlying hierarchies

Ensure covariates are based on globally replicable and transparent way, e.g. no

**References:**

Alahuhta, J., Toivanen, M. & Hjort, J. (2020). Geodiversity–biodiversity relationship needs more empirical evidence. *Nat. Ecol. Evol.*, 4, 2–3.

Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw.*, 80.

Cord, A.F., Meentemeyer, R.K., Leitão, P.J. & Václavík, T. (2013). Modelling species distributions with remote sensing data: bridging disciplinary perspectives. *J. Biogeogr.*, 40, 2226–2227.

Duncan, C., Chauvenet, A.L.M., Brown, M.E. & Pettorelli, N. (2015). Energy availability, spatio-temporal variability and implications for animal ecology. *Divers. Distrib.*, 21, 290–301.

Evans, K.L., Warren, P.H. & Gaston, K.J. (2005). Species–energy relationships at the macroecological scale: a review of the mechanisms. *Biol. Rev.*, 80, 1–25.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.*, 202, 18–27.

Gray, C.L., Hill, S.L.L., Newbold, T., Hudson, L.N., Börger, L., Contu, S., *et al.* (2016). Local biodiversity is higher inside than outside terrestrial protected areas worldwide. *Nat. Commun.*, 7, 12306.

He, K.S., Bradley, B.A., Cord, A.F., Rocchini, D., Tuanmu, M.-N., Schmidtlein, S., *et al.* (2015). Will remote sensing shape the next generation of species distribution models? *Remote Sens. Ecol. Conserv.*, 1, 4–18.

Hillebrand, H. (2004). On the Generality of the Latitudinal Diversity Gradient. *Am. Nat.*, 163, 192–211.

Holt, B.G., Costa, G.C., Penone, C., Lessard, J.-P., Brooks, T.M., Davidson, A.D., *et al.* (2017). Environmental variation is a major predictor of global trait turnover in mammals. *J. Biogeogr.*, 1–13.

van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D.A., *et al.* (2019). Soil nematode abundance and functional group composition at a global scale. *Nature*, 572, 194–198.

Houlahan, J.E., McKinney, S.T., Anderson, T.M. & McGill, B.J. (2017). The priority of prediction in ecological understanding. *Oikos*, 126, 1–7.

Hudson, L.N., Newbold, T., Contu, S., Hill, S.L.L., Lysenko, I., De Palma, A., *et al.* (2014). The PREDICTS database: a global database of how local terrestrial biodiversity responds to human impacts. *Ecol. Evol.*, 4, 4701–4735.

Hudson, L.N., Newbold, T., Contu, S., Hill, S.L.L., Lysenko, I., De Palma, A., *et al.* (2017). The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. *Ecol. Evol.*, 7, 145–188.

Huete, a., Didan, K., Miura, T., Rodriguez, E.., Gao, X. & Ferreira, L.. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.*, 83, 195–213.

Hurlbert, A.H. (2004). Species-energy relationships and habitat complexity in bird communities. *Ecol. Lett.*, 7, 714–720.

Jiang, Z., Huete, A., Didan, K. & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.*, 112, 3833–3845.

Jung, M., Hill, S.L.L., Platts, P.J., Marchant, R., Siebert, S., Fournier, A., *et al.* (2017). Local factors mediate the response of biodiversity to land use on two African mountains. *Anim. Conserv.*, 20, 370–381.

Jung, M., Rowhani, P., Newbold, T., Bentley, L., Purvis, A. & Scharlemann, J.P.W. (2019a). Local species assemblages are influenced more by past than current dissimilarities in photosynthetic activity. *Ecography (Cop.).*, 42, 670–682.

Jung, M., Rowhani, P. & Scharlemann, J.P.W. (2019b). Impacts of past abrupt land change on local biodiversity globally. *Nat. Commun.*, 10, 5474.

König, C., Weigelt, P. & Kreft, H. (2017). Dissecting global turnover in vascular plants. *Glob. Ecol. Biogeogr.*, 26, 228–242.

Mesgaran, M.B., Cousens, R.D. & Webber, B.L. (2014). Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Divers. Distrib.*, 20, 1147–1159.

Moon, J.B., Dewitt, T.H., Errend, M.N., Bruins, R.J.F., Kentula, M.E., Chamberlain, S.J., *et al.* (2017). Model application niche analysis: assessing the transferability and generalizability of ecological models. *Ecosphere*, 8, e01974.

Newbold, T., Hudson, L.N., Arnell, A.P., Contu, S., De Palma, A., Ferrier, S., *et al.* (2016). Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science (80-. ).*, 353, 288–291.

Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Lysenko, I., Senior, R.A., *et al.* (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, 520, 45–50.

Oldeland, J., Wesuls, D., Rocchini, D., Schmidt, M. & Jürgens, N. (2010). Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity? *Ecol. Indic.*, 10, 390–396.

Parmentier, I., Harrigan, R.J., Buermann, W., Mitchard, E.T.A., Saatchi, S., Malhi, Y., *et al.* (2011). Predicting alpha diversity of African rain forests: models based on climate and satellite-derived data do not perform better than a purely spatial model. *J. Biogeogr.*, 38, 1164–1176.

Petchey, O.L., Pontarp, M., Massie, T.M., Kéfi, S., Ozgul, A., Weilenmann, M., *et al.* (2015). The ecological forecast horizon, and examples of its uses and determinants. *Ecol. Lett.*, 18, 597–611.

Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.-M., Tucker, C.J. & Stenseth, N.C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends Ecol. Evol.*, 20, 503–510.

Phillips, H.R.P., Guerra, C.A., Bartz, M.L.C., Briones, M.J.I., Brown, G., Crowther, T.W., *et al.* (2019). Global distribution of earthworm diversity. *Science (80-. ).*

Phillips, H.R.P., Newbold, T. & Purvis, A. (2017). Land-use effects on local biodiversity in tropical forests vary between continents. *Biodivers. Conserv.*, 26, 2251–2270.

Purvis, A., Newbold, T., De Palma, A., Contu, S., Hill, S.L.L., Sanchez-Ortiz, K., *et al.* (2018). Modelling and projecting the response of local terrestrial biodiversity worldwide to land use and related pressures: The PREDICTS project. In: *Advances in Ecological Research*. pp. 201–241.

Radeloff, V.C., Dubinin, M., Coops, N.C., Allen, A.M., Brooks, T.M., Clayton, M.K., *et al.* (2019). The Dynamic Habitat Indices (DHIs) from MODIS and global biodiversity. *Remote Sens. Environ.*, 222, 204–214.

Randin, C.F., Ashcroft, M.B., Bolliger, J., Cavender-Bares, J., Coops, N.C., Dullinger, S., *et al.* (2020). Monitoring biodiversity in the Anthropocene using remote sensing in species distribution models. *Remote Sens. Environ.*, 239, 111626.

Rapacciuolo, G. (2019). Strengthening the contribution of macroecological models to conservation practice. *Glob. Ecol. Biogeogr.*, 28, 54–60.

Regos, A., Gagne, L., Alcaraz-Segura, D., Honrado, J.P. & Domínguez, J. (2019). Effects of species traits and environmental predictors on performance and transferability of ecological niche models. *Sci. Rep.*, 9, 4221.

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., *et al.* (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography (Cop.).*, 40, 913–929.

Rocchini, D., Balkenhol, N., Carter, G. a., Foody, G.M., Gillespie, T.W., He, K.S., *et al.* (2010). Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecol. Inform.*, 5, 318–329.

Rocchini, D., Hernández-Stefanoni, J.L. & He, K.S. (2015). Advancing species diversity estimate by remotely sensed proxies: A conceptual review. *Ecol. Inform.*, 25, 22–28.

Schaaf, C.B., Gao, F., Strahler, A.H., Lucht, W., Li, X., Tsang, T., *et al.* (2002). First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sens. Environ.*, 83, 135–148.

Schmidtlein, S. & Fassnacht, F.E. (2017). The spectral variability hypothesis does not hold across landscapes. *Remote Sens. Environ.*, 192, 114–125.

Stein, A., Gerstner, K. & Kreft, H. (2014). Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecol. Lett.*, 17, 866–880.

Theobald, D.M., Harrison-Atlas, D., Monahan, W.B. & Albano, C.M. (2015). Ecologically-Relevant Maps of Landforms and Physiographic Diversity for Climate Adaptation Planning. *PLoS One*, 10, e0143619.

Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., *et al.* (2018). Outstanding Challenges in the Transferability of Ecological Models. *Trends Ecol. Evol.*, 33, 790–802.

Zurell, D., Elith, J. & Schröder, B. (2012). Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. *Divers. Distrib.*, 18, 628–634.