**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2015

Spatio-phylogenetic multispecies distribution models

Kaldhusdal, Arne ; Brandl, Roland ; Müller, Jörg ; Möst, Lisa ; Hothorn, Torsten

Abstract: 1. Ecologists increasingly consider phylogenetic relatedness in both community composition and spatial arrangements in communities. Here we considered both the phylogenetic correlation between multiple species and the spatial correlation induced by unobserved spatial heterogeneity on multiple plots. For this analysis, we introduced phylogenetic spatial generalised linear mixed models (PSGLMMs), which are an extension of phylogenetic generalised linear mixed models (PGLMMs). 2. We used the framework of generalised linear array models to simultaneously model species and plot dimension. Such models have the potential to explain the correlation of the phylogenetic relationship of the observed species and of the spatial proximity of the plots, or both. We proposed model selection strategies based on proper scores and empirically evaluated them in a case study using bird count data. In our analysis, we focused on two special cases: the community composition model and the environmental sensitivity model. 3. Our simulation study indicated that it might be difficult to correctly identify phylogenetic signals when the phylogenetic correlation is rather low and when studying presence-absence or count data of rare or pervasive species.

DOI: <https://doi.org/10.1111/2041-210X.12318>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-115524>

Journal Article

Supplemental Material

Originally published at:

Kaldhusdal, Arne; Brandl, Roland; Müller, Jörg; Möst, Lisa; Hothorn, Torsten (2015). Spatio-phylogenetic multispecies distribution models. *Methods in Ecology and Evolution*, 6(2):187-197.

DOI: <https://doi.org/10.1111/2041-210X.12318>

Supplementary Material: Spatio-Phylogenetic Multi-Species Distribution Models

Arne Kaldhusdal¹, Roland Brandl², Jörg Müller^{3,4}, Lisa Möst¹
and Torsten Hothorn⁵

¹Institut für Statistik, LMU München, Germany

²Fachbereich Biologie, Philipps-Universität Marburg, Germany

³Sachgebiet Forschung und Dokumentation,
Nationalparkverwaltung Bayerischer Wald, Grafenau, Germany

⁴Lehrstuhl für Terrestrische Ökologie, Technische Universität
München, Germany

⁵Institut für Epidemiologie, Biostatistik und Prävention,
Universität Zürich, Switzerland

A Online Supplement

A.1 Comparison of PGLMMs and PSGLMMs

Figure 1: Comparison of the (top) log-likelihood, (middle) random effect variance parameter estimate $\hat{\sigma}_{\text{spp}}^2$ of the species and (bottom) plot random effect variance estimate $\hat{\sigma}_{\text{plot}}^2$ for a PGLMM and a PSGLMM for 72 simulated data sets of binary occurrence data for $S = 16$ species in $P = 30$ plots. The y-axis for $\hat{\sigma}_{\text{plot}}^2$ is trimmed to ease visibility. The red dots mark the mean values for each model.

Figure 2: Comparison of the fixed effects estimates for species 1 through 16 for a PGLMM and a PSGLMM for 72 simulated data sets of binary occurrence data for $S = 16$ species in $P = 30$ plots. The red dots mark the mean values for each model.

A.2 Simulation Study

Figure 3: Simulation study for the CCM with binary responses: the distribution of the variance parameter estimates for each random effect component (rows) and setting (columns) subject to the phylogenetic correlation ρ between species 1 and 2. The true parameter to be estimated was $\hat{\sigma}_\gamma^2 = 1$ in settings (a) and (b) and $\hat{\sigma}_\gamma^2 = 1 \cdot 1 = 1$ in setting (c). Outliers are not drawn.

Figure 4: Simulation study for the CCM with Poisson responses: the distribution of the variance parameter estimates for each random effect component (rows) and setting (columns) subject to the phylogenetic correlation ρ between species 1 and 2. The true parameter to be estimated was $\hat{\sigma}_\gamma^2 = 1$ in settings (a) and (b) and $\hat{\sigma}_\gamma^2 = 1 \cdot 1 = 1$ in setting (c). Outliers are not drawn.

Figure 5: Simulation study for the ESM with Gaussian responses: the distribution of the variance parameter estimates for each random effect component (rows) and setting (columns) subject to the phylogenetic correlation ρ between species 1 and 2. The true parameter to be estimated was $\hat{\sigma}_\gamma^2 = 2$ in settings (a) and (b) and $\hat{\sigma}_\gamma^2 = 2 \cdot 2 = 4$ in setting (c). Outliers are not drawn.

Figure 6: Simulation study for the ESM with binary responses: the distribution of the variance parameter estimates for each random effect component (rows) and setting (columns) subject to the phylogenetic correlation ρ between species 1 and 2. The true parameter to be estimated was $\hat{\sigma}_\gamma^2 = 1$ in settings (a) and (b) and $\hat{\sigma}_\gamma^2 = 1 \cdot 1 = 1$ in setting (c). Outliers are not drawn.

Figure 7: Simulation study for the ESM with Poisson responses: the distribution of the variance parameter estimates for each random effect component (rows) and setting (columns) subject to the phylogenetic correlation ρ between species 1 and 2. The true parameter to be estimated was $\hat{\sigma}_\gamma^2 = 1$ in settings (a) and (b) and $\hat{\sigma}_\gamma^2 = 1 \cdot 1 = 1$ in setting (c). Outliers are not drawn.

A.3 Bird Abundance Data

To illustrate the application of the generalised linear mixed models, we used data collected by Kaluza (2010) across 371 forests plots of $1ha$ in southeastern Germany. Overall, the sites cover an elevational gradient of approximately 280 up to $1,420m$ above sea level. On these sites, point counts (see also Bässler et al., 2009) were used to estimate the abundance of 47 bird species. To characterise the forest structure, numerous environmental variables were recorded. For the present exercise, we used only elevation, coverage (in %) of the middle forest layer by broadleaved trees (referred to as broadleaved trees) and by rejuvenating broadleaved trees (rejuvenation), the maximum diameter (in cm) at breast height (DBH; at $130cm$ height) and the total number of tree holes within an area of $0.1ha$.

Figure 8: Map of sampling plots with the four main transects located inside the Bavarian Forest National Park.

A.4 Phylogenetic Tree

For the 47 species shown in Figure 9, we retrieved 200 phylogenies from www.birdtree.org using the Ericson all species data set (10,000 trees covering 9,993 species each (Jetz et al., 2012)). For each of these phylogenies, we calculated a distance matrix using the function `distTips` in the R package `adephylo` (Jombart & Dray, 2010). Finally, we averaged these 200 distance matrices to estimate the phylogenetic structure and used this average distance matrix in a UPGMA cluster algorithm.

Figure 9: Ultrametric tree estimating the phylogeny of the 47 species recorded across the plots (Figure 8).

References

- Bässler, C.; Förster, B.; Moning, C. & Müller, J. (2009) The BIOKLIM project: Biodiversity research between climate change and wilding in a temperate montane forest – the conceptual framework, *Waldökologie, Landschaftsforschung und Naturschutz*, (7), 21–34.
- Jetz, W.; Thomas, G.H.; Joy, J.B.; Hartmann, K. & Mooers, A.O. (2012) The global diversity of birds in space and time, *Nature*, 491, 444–448.
- Jombart, T. & Dray, S. (2010) *adephylo*: Exploratory analyses for the phylogenetic comparative method, *Bioinformatics*, 26, 1907–1909.
- Kaluza, B. (2010) *Funktionale Beziehungen von Vogelmerkmalen und Habitatstruktur entlang eines Mittelgebirgsgradienten*, Master’s thesis, Julius-Maximilians-University Würzburg.