

# **Biodiversity modelling reveals a significant gap between diversity hotspots and protected areas for Iranian reptiles**

Sajad Noori, Oliver Hawlitschek, Jens Oldeland, Hossein Rajaei, Martin Husemann & Marianna Simões

Supplementary Information: S2

## **Materials and Methods**

### *Occurrence data*

A dataset downloaded initially yielded 12,735 records (GBIF, 06 August 2019). The GBIF dataset includes the comprehensive geo-referenced dataset of 8,525 records of 146 species of lizards by Šmíd et al., (2014). After taxonomic and geographic cleaning of the records (by excluding duplications and missing values), in total 5,687 records were retained. The dataset includes coordinate data for 245 species of reptiles, of which 235 species are Squamata (more than 98% of the occurrence records), nine Testudines (ca. 1.6% occurrence records), and one Crocodylia (< 0.2% occurrence records). All analyses were conducted at the species level. There were 24 families of reptiles in the dataset, with Gekkonidae and Lacertidae represented by each 51 species, followed by Colubridae (29 spp.), Agamidae (26 spp.), Scincidae (19 spp.), Viperidae (11 spp.), Phyllodactylidae (10 spp.); the other families have less than six species (Appendix S1). Over half of the included species (53%) had more than five records in the dataset. Approximately 20% of the species (46 spp.) were endemic, all of which belong to Squamata. Phyllodactylidae had the highest number of endemic species (80%), followed by Lacertidae (ca. 30%) and Gekkonidae (ca. 24%; Fig. 1). In total, 15 species were assessed as threatened and 16 species were considered as “Data Deficient” by the IUCN Red List (IUCN, 20 April 2020), meaning that insufficient data is available to assess the risk status of these species.

### *Richness-Environment models*

The model selection was conducted using the information-theoretical framework (Burnham & Anderson, 2002) based on the Akaike Information Criterion (AIC; Akaike, 1973). We calculated the AIC for each of the 13 candidate models. From the AIC values, we derived delta-AIC (dAIC), as well as Akaike Weights (AW; scaled between 0 and 1) to identify a single best model. The best model received the lowest AIC value; subsequent models were required to have a dAIC value larger than two, and an AW substantially larger than the second model. We further calculated the deviance-squared for each model, which represents the (adjusted) proportion of deviance, accounted for the model. This was required to better interpret the fit of the best model to the data. For model selection, we used the R packages bbmle (Bolker, 2020) and MuMIn (Barton, 2009).

The selected best models for residual model structure were evaluated using the standard diagnostic plots (e.g., QQ-plots, histograms, and leverage plots), the applicability of the number of knots, and their effects on the estimated degrees of freedom. To interpret, models were visualized using univariate and spatial partial regression

plots. To check for potential effects of spatial autocorrelation (SAC), we analysed the model residuals by plotting the residuals on a map and calculated isotropic, as well as anisotropic semi-variograms for the directions (0, 45, 90, and 135). Finally, we calculated a Moran's I spatial correlogram to verify whether SAC had a strong influence at a certain distance (*lag*). *Lags* were measured in kilometres and distributed in 50 bins across the spatial extent of the study area. To conduct the model evaluation, we used the R packages *ncf* (Bjornstad & Cai, 2020), *gstat* (Pebesma, 2020), and *spdep* (Bivand et al., 2017). Model visualization was done with the *gratia* package (Simpson, 2020) and *ggplot2* (Wickham, 2016); the summary table was built with *sjPlot* (Lüdecke, 2017).