Nature’s Contribution to People and biodiversity mapping in Switzerland: spatial patterns and environmental drivers

– ODMAP Protocol –

Nathan Külling, Antoine Adde, Audrey Lambiel, Sergio Wicki, Antoine Guisan, Adrienne Grêt-Regamey, Anthony Lehmann

2023-12-04

## Overview

#### Authorship

Contact: Antoine Adde (antoine.adde@eawag.ch)

Study link: [doi]

#### Model objective

Model objective: Mapping

Target output: current maps of continuous habitat suitability values for 7,041 species across Switzerland at 25m resolution

#### Focal Taxon

Focal Taxon: 1,833 species. Complete species list (list\_species\_total.csv) available on github at : https://github.com/NKulling/SWISS\_NCP\_ASSESSMENT/tree/main/BD

#### Location

Location: Switzerland

#### Scale of Analysis

Spatial extent: 2480000, 2840000, 1070000, 1300000 (xmin, xmax, ymin, ymax)

Spatial resolution: 25m

Temporal extent: 1981-2010 for the current period

Temporal resolution: 30-year periods

Boundary: political, Switzerland

#### Biodiversity data

Observation type: field collection

Response data type: presence / background absence data

#### Predictors

Predictor types: Eight main categories (bioclimatic, edaphic, hydrologic, land use and cover, population, topographic, transportation, and vegetation); see companion covariate list for details)

#### Hypotheses

Model assumptions: 1) candidate covariates potentially drive the distribution of the species and key covariates are included in the models, 2) The observed distributions of the species are in pseudo-equilibrium with the environment, 3) The entire realised niches of the species are encompassed by the modelling data, 4) The species maintain their niches across time; 5) The correlation structure between covariates does not change over time

#### Algorithms

Modelling techniques: five modelling algorithms were used – Generalized Linear Model (GLM) (McCullagh & Nelder 1989), Generalized Additive Model (GAM) (Hastie 2017), Maxnet (MAX) (Phillips et al. 2017), Random Forest (RF) (Breiman 2001), and light Gradient Boosted Machine (GBM) (Ke et al. 2017).

Model complexity: hyperparameter grid search using N-SDM default settings

Model averaging/ensemble modelling: Combination of the predictions from the three modelling algorithms into unweighted ensemble projections (average)

#### Workflow

Model workflow: See N-SDM workflow (<https://github.com/N-SDM/N-SDM>)

#### Software

Software: N-SDM v1.0.1 (Adde et al., 2023) and R v4.2.2

Code availability: <https://github.com/N-SDM/N-SDM>

Data availability: N-SDM is available from the GitHub Repository, https://github.com/N-SDM/N-SDM, where complementary instructions for installation and example data are provided. In addition, for each version of SDM-Swiss, the specific setting file used for running N-SDM can be found in the companion nsdm\_settings.csv file, available on the study GitHub repository: https://github.com/NKulling/SWISS\_NCP\_ASSESSMENT/tree/main/BD.

## Data

#### Biodiversity data

Taxon names: 1,833 species (see companion species list for details)

Ecological level: species

Data sources: Two sets of species occurrences records for inside and outside of Switzerland were used. For Switzerland, data were provided by the Swiss Species Information Center InfoSpecies (www.infospecies.ch) on August 23, 2021 (<https://doi.org/10.15468/htjezm>). Occurrence records for outside of Switzerland for matching species were obtained from the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/) on October 27, 2021 (https://doi.org/10.15468/dl.zwp3dx).

Sampling design: Data were compiled from various sources, each having specific sampling design.

Sample size: Only species with more than 50 occurrence records after spatial disaggregation were retained for modelling.

Cleaning: Spatial disaggregation

Background data: For each species and occurrence set, 10,000 background absences were randomly generated

#### Data partitioning

Split-sample approach repeated 100 times with 30% of the data kept for validation. No selection of truly independent test data was done.

#### Predictor variables

Predictor variables: Suite of 1,865 candidate covariates from eight main categories See companion covariate list for details.

Data sources: SWECO25 (v.1.0) (Külling, & Adde et al., in review) and CHELSA (Karger et al. 2017).

Spatial extent: 2480000, 2840000, 1070000, 1300000 (xmin, xmax, ymin, ymax)

Spatial resolution: 25 m

Coordinate reference system: CH1903+ / LV95

Temporal extent: 1981-2010

## Model

#### Selected predictors

Covariate selection: Automated covariate selection procedure included in the N-SDM workflow (see Adde et al., 2023)

#### Multicollinearity

Multicollinearity: The covariate selection procedure included in N-SDM allows filtering for multicollinearity (Adde et al., 2023)

#### Model settings

Hyperparameter settings were defined using N-SDM automatic grid search (default settings)

#### Model estimates

Assessment of variable importance: variable importance was assessed with the N-SDM function nsdm.respcurve.

#### Threshold selection

Threshold selection: Maximum True Skill Statistic (maxTSS)

## Assessment

#### Performance statistics

Performance on training data: Model accuracy was evaluated using a split-sample approach repeated 100 times with 30% of the data kept for validation. For each model, the best combination of hyperparameters was identified using the average “Score” of three evaluation metrics including the Area Under the Curve′ (AUC′) (or Somers’ D, such as AUC^'=AUC\*2-1), the maximized True Skill Statistic (maxTSS), and the Continuous Boyce Index (CBI).

Performance on validation data: same as for training data

#### Plausibility check

Response shapes: Response curves (nsdm.respcurve function).

## Prediction

#### Prediction output

Prediction unit: Pixel-level habitat suitability values

#### Uncertainty quantification

Algorithmic uncertainty: Algorithmic uncertainty in predictions was measured by using the coefficient of variation.

## References

Adde, A., Rey, P.L., Brun, P., Külling, N., Fopp, F., Altermatt, F. et al. (2023b). N‐SDM: a high‐performance computing pipeline for Nested Species Distribution Modelling. Ecography, e06540.

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W. et al. (2017). Climatologies at high resolution for the earth’s land surface areas. Scientific Data, 4, 170122.

Külling, N. and Adde, A., Fopp, F., Schweiger, A., Broennimann, O., Rey, P-L., Giuliani, G., Goicolea, T., Petitpierre, B., Zimmermann, N.E., Pellissier, L., Altermatt, F., Lehmann, A., Guisan, A. SWECO25: a cross-thematic raster database for ecological research in Switzerland.