Testing the optimal proportion of pseudo-absences/presences in combination with spatial thinning methods

– ODMAP Protocol –

2025-05-21

## Overview

#### Authorship

<Contact >

<Study link>

#### Model objective

Model objective: Inference and explanation

#### Focal Taxon

Focal Taxon: virtual species

#### Location

Location: Australia

#### Scale of Analysis

Spatial extent: 113, 154, -44, -10 (xmin, xmax, ymin, ymax)

Spatial resolution: 30 sec/1 km

Temporal extent: climate data: 1970-2000; VS – not specified

Temporal resolution: climate data: average per 30 year period; VS – not specified

Boundary: natural

#### Biodiversity data

Observation type: standardised monitoring data

Response data type: presence/absence

#### Predictors

Predictor types: climatic

#### Hypotheses

Hypotheses: Preprogrammed dependence on two climate variables which include precipitation and temperature measurement

#### Assumptions

Model assumptions: XXXXX

#### Algorithms

Modelling techniques: gam, glm, maxent, randomForest

Model complexity: XXXX

Model averaging: yes XXXX

#### Workflow

Model workflow: XXX

#### Software

Software: RStudio XXX

Code availability: <https://github.com/UP-macroecology/Kuznetsova_VirtualSp_SDM_pseudoabsences_and_thinning_2024/tree/main>

Data availability: <https://www.worldclim.org/data/worldclim21.html>

## Data

#### Biodiversity data

Taxon names: Virtual species (inspired by Eastern Water Dragon)

Taxonomic reference system: –

Ecological level: individuals

Data sources: Artificially simulated

Sampling design: Random sampling

Sample size: Presence sample sizes (20, 50, 100, 500, 1000) and applied varying pseudo-absence ratios (x1, x3, x5, x10)

Clipping: Australia

Scaling: Spatial thinning using checkerboard and spThin method

Cleaning: None

Absence data: –

Background data: Buffer of 200 km around presence points

Errors and biases: –

#### Predictor variables

Predictor variables: 19 bioclim variables

Data sources: <https://www.worldclim.org/data/worldclim21.html>; Accession Date: May 2025; WorldClim version 2.

Spatial extent: 113, 154, -44, -10 (xmin, xmax, ymin, ymax)

Spatial resolution: 1 km/ 30 sec

Coordinate reference system: WGS 84

Temporal extent: 1970-2000

Temporal resolution: 30 year average

Data processing: –

Errors and biases: –

Dimension reduction: –

#### Transfer data

<Spatial extent>

Spatial resolution: –

Temporal extent: –

Temporal resolution: –

## Model

#### Multicollinearity

<Multicollinearity>

#### Model settings

<gam>

<glm>

<maxent>

<randomForest>

#### Model estimates

<Coefficients>

#### Analysis and Correction of non-independence

<Spatial autocorrelation>

## Assessment

#### Performance statistics

<Performance on training data>

#### Plausibility check

<Response shapes>