A MAXENT PREDICTIVE MODEL FOR GRASSLAND HUNTER-GATHERER SITES

– ODMAP Protocol –

Daniel Rafuse

2021-01-11

## Overview

#### Authorship

Contact : [drafuse@soc.unicen.edu.ar](mailto:drafuse@soc.unicen.edu.ar)

Study link: <https://github.com/drafuse>

#### Model objective

Model objective: Forecast and transfer

Target output: The main target output are archaeology sites.

#### Focal Taxon

Focal Taxon: Hunter-gatherer archaeology sites with associated radiocarbon dates.

#### Location

Location: Southeastern Pampas, Argentina

#### Scale of Analysis

Spatial extent: -57.5, -63, -36.5, -39 (xmin, xmax, ymin, ymax)

Spatial resolution: 124,740 km square.

Temporal extent: 12,240 to 470 radiocarbon years BP.

Temporal resolution: Approximately 40 years of professional archaeological survey.

Boundary: rectangle

#### Biodiversity data

Observation type: field survey

Response data type: presence-only

#### Predictors

Predictor types: habitat, topographic

#### Hypotheses

Hypotheses: If human behavior is modeled with respect to the natural and social environment, therefore we can predict the location of hunter-gatherer archaeological sites based on physical factors that remained stable over extended periods of time.

#### Assumptions

Model assumptions: The enviromental predictors are measured (or estimated) without error.

#### Algorithms

Modelling techniques: maxent

Model complexity: MaxEnt models were built with linear and quadratic features only.

Model averaging: 30 replicates were run.

#### Workflow

Model workflow: All enviormental predictors were processed at a 3 arc second (approximately 90-meter) cell size.

#### Software

Software: Maxent Version 3.4.3, November 2020; ENMTools Version 1.3, August 2020; QGIS Version 3.16.0-Hannover, October 2020.

Code availability: <https://github.com/drafuse>

Data availability: <https://github.com/drafuse>

## Data

#### Biodiversity data

Taxon names: Homo sapiens.

Taxonomic reference system: Hunter-gatherers.

Ecological level: populations

Data sources: See references in database.

Sampling design: Stratified (only sites with radiocarbon dates).

Sample size: 66 sites.

Clipping: Southeastern Pampas, Argentina.

Scaling: All data (raster and vector) was clipped to study area.

Cleaning: Redundant records that occur in a single cell were randomly eliminated (total = 10 sites).

Absence data: No absence data was used in this model.

Background data: Bias grid file was built by deriving a kernel density estimation map.

Errors and biases: Some geo-referencing errors are likely in sites recorded before the regular use of GPS.

#### Data partitioning

Training data: 2 to 3 presence records.

Validation data: Crossvalidate.

Test data: Known field survey data from three localities.

#### Predictor variables

Predictor variables: Digital Elevation Model (DEM); Slope; Aspect; Topographic Wetness Index (TWI); Watercourse; Bodies of water; Tool source.

Data sources: Digital Elevation Model (DEM): STRM 3 arc second void-filled raster (<https://earthexplorer.usgs.gov/>). Slope, aspect and TWI derived from DEM. Watercourses, bodies of water from IGN (<https://www.ign.gob.ar/>). Tool source from SegemAR (<http://www.segemar.gov.ar/institucional/>) and DEM.

Spatial extent: -57.5000, -63.0000, -36.5000, -39.000 (xmin, xmax, ymin, ymax)

Spatial resolution: 3 arc second.

Coordinate reference system: WGS 1984, EPSG:4326.

Temporal extent: Most recent data was downloaded.

Temporal resolution: For raster data (3 arc second).

Data processing: For watercourses, bodies of water, and tool source; a Euclidean Distance was calulated in QGIS and normalized (matrix 0-1).

Errors and biases: Kernel density estimation map (Eligh et al. 2010) of the archaeological site locations, with a maximum radius of 10 km.

Dimension reduction: No dimension reduction.

#### Transfer data

Data sources: Not applicable.

Spatial extent: Not applicable., Not applicable., Not applicable., Not applicable. (xmin, xmax, ymin, ymax)

Spatial resolution: Not applicable.

Temporal extent: Not applicable.

Temporal resolution: None.

Models and scenarios: Not applicable.

Data processing: None.

Quantification of Novelty: Not applicable.

## Model

#### Variable pre-selection

Variable pre-selection: Physical factors that remained stable over extended periods of time.

#### Multicollinearity

Multicollinearity: Multicollinearity between predictors was investigated using ENMTools Correlation (Pearson correlation coefficient). All variables were uncorrelated (r < 0.99).

#### Model settings

maxent: featureSet (linear and quadratic), regularizationMultiplierSet (0.75 ), convergenceThresholdSet (0.00001), samplingBiasRule (kernel density estimation map), background points (10,000), replicates (30), maximum iterations (500), default prevalence (0.5), random test percentage (0, cross-validation)

Model settings (extrapolation): No extrapolation.

#### Model estimates

Coefficients: Cross-validation.

Parameter uncertainty: No quantification of uncertainty.

Variable importance: Percent contribution and Permutation importance

#### Model selection - model averaging - ensembles

Model selection: Parameter with the lowest AICc value.

Model averaging: point-wise mean and standard deviation

#### Analysis and Correction of non-independence

Spatial autocorrelation: None

#### Threshold selection

Threshold selection: Not applicable.

## Assessment

#### Performance statistics

Performance on training data: AIC, AUC

Performance on validation data: AIC, AUC

Performance on test data: AIC, AUC

#### Plausibility check

Response shapes: Response curves and inspecting the mapped predictions.

Expert judgement: Mapped predictions.

## Prediction

#### Prediction output

Prediction unit: Point-wise mean and standard deviation of the 30 output grids.

Post-processing: None.

#### Uncertainty quantification

Algorithmic uncertainty: None.

Input data uncertainty: None.

Parameter uncertainty: None.

Scenario uncertainty: None.

Novel environments: None.