Carpatian Basin 30ka BP Environmental Niche Modelling

Christian Willmes

Scope: Documentation of R based Species Distribution Modelling (SDM) applied to human sites for 30ky BP.

# Data

The modelling of course depends on the quality and quantity of the input data. The more environmental information, as well as archaeological finds for the  probed time period is available to be incorporated into the model, the better are the results.

## Sites

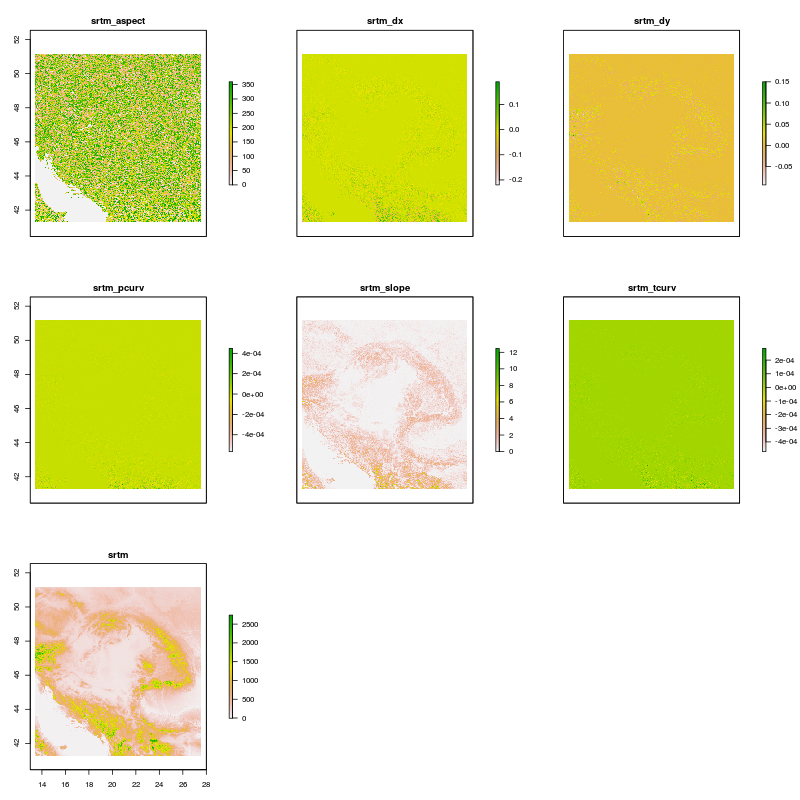
## 



67 Sites for human settlements of around 30ky BP.

## Topography

The topography data used in this study is based on the SRTM DEM (Jarvis et al. 2008).



Topography predictor variables derived from SRTM DEM.

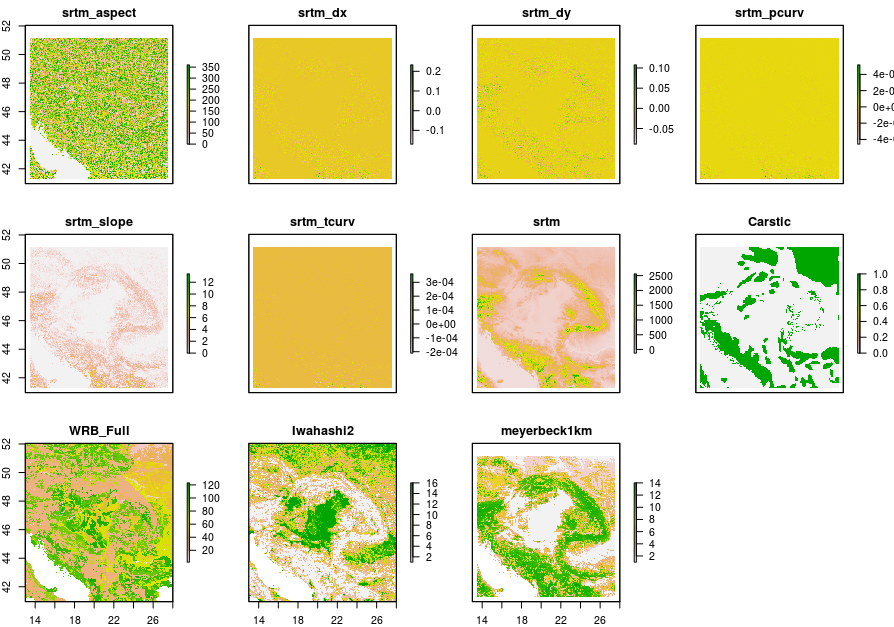
A set of 6 additional predictor varibales (slope, aspect, curvatures and partial derivatives) see fig. ???, were computed from the SRTM DEM data using the GRASS GIS*r.slope.aspect* tool (Shapiro and Waupotitsch 2015)*.*

## 

## Geology and Soil

ToDo: get Geology and (Paleo-)Soil information for the Region.

## 



This is a caption

## Landuse

ToDo: Include recent Landuse classifications, because most finds are found on open farm land.

## Landform classifications

Describe Iwahshi and Meyerbeck data

## Visiblity range score

ToDo: create a dataset, that computes a score for best viewshed and visibility ranges. Something similar to cellphone tower placement GIS applications.

## Hydrology

ToDo: Create a paleo DEM for the Carpatian Basin (huge Task), and compute the Hydrology (River network) of the CB.

## Climate

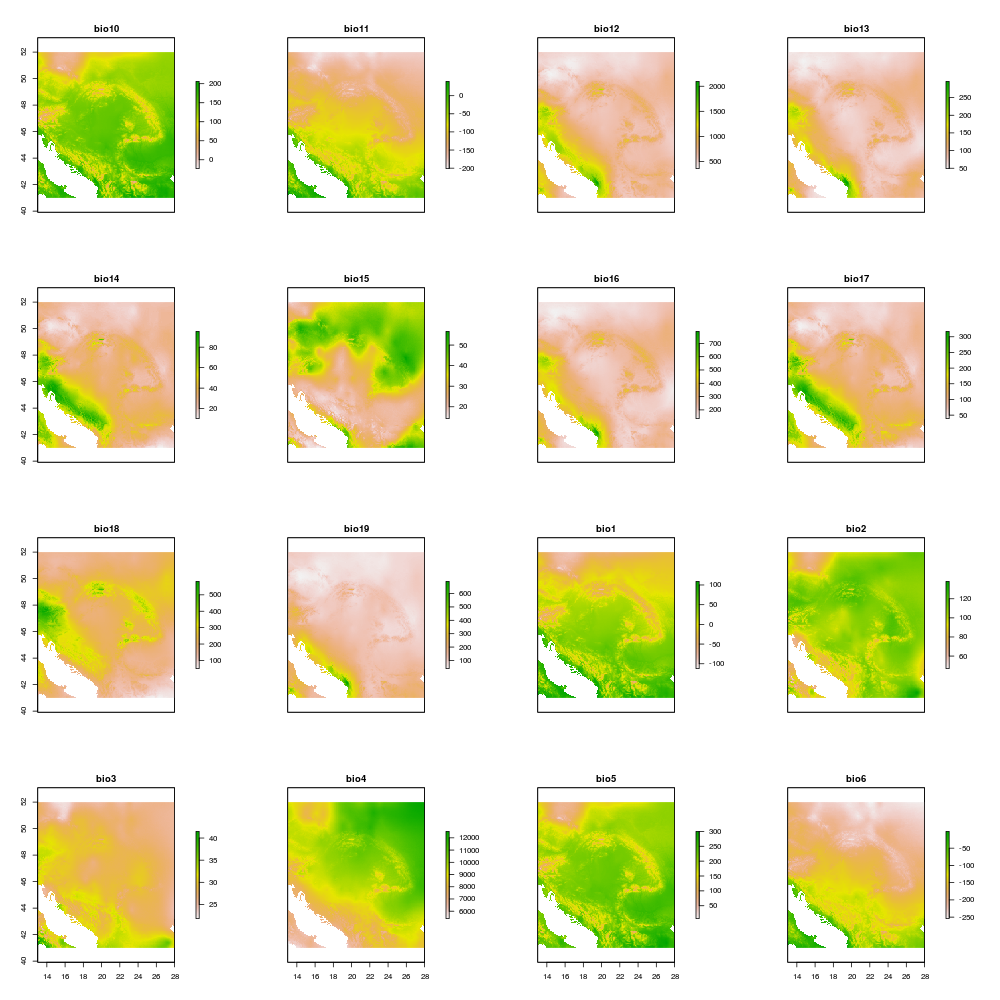
## 

### Climate proxy data

To have a reliable climate model, it is requested to have paleoenvironmental information from climate proxies, like sedimentological (e.g. from loess), palynological or other analysis, available for validating the climate model. Without a regular validation of the data, we can’t use it for producing valid estimates or assertions based on that model.

### Climate Model

The climate data applied for this study is based on a new approach for spatio-temporal climate data interpolation (Willmes et al. 2017). For the application with SDM models, the 19 bio climatic variables are computed using the R dismo package (Hijmans and Elith 2016), see Fig. ???.



Bioclimatic variables for 30ky BP.

The bioclimatic variables computed using the dismo R package, are defined as follows:

BIO1 = Annual Mean Temperature

BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))

BIO3 = Isothermality (BIO2/BIO7) (\* 100)

BIO4 = Temperature Seasonality (standard deviation \*100)

BIO5 = Max Temperature of Warmest Month

BIO6 = Min Temperature of Coldest Month

BIO7 = Temperature Annual Range (BIO5-BIO6)

BIO8 = Mean Temperature of Wettest Quarter

BIO9 = Mean Temperature of Driest Quarter

BIO10 = Mean Temperature of Warmest Quarter

BIO11 = Mean Temperature of Coldest Quarter

BIO12 = Annual Precipitation

BIO13 = Precipitation of Wettest Month

BIO14 = Precipitation of Driest Month

BIO15 = Precipitation Seasonality (Coefficient of Variation)

BIO16 = Precipitation of Wettest Quarter

BIO17 = Precipitation of Driest Quarter

BIO18 = Precipitation of Warmest Quarter

BIO19 = Precipitation of Coldest Quarter

## GIS data set

The GIS data set is published in the CRC806-Database (Willmes 2017).

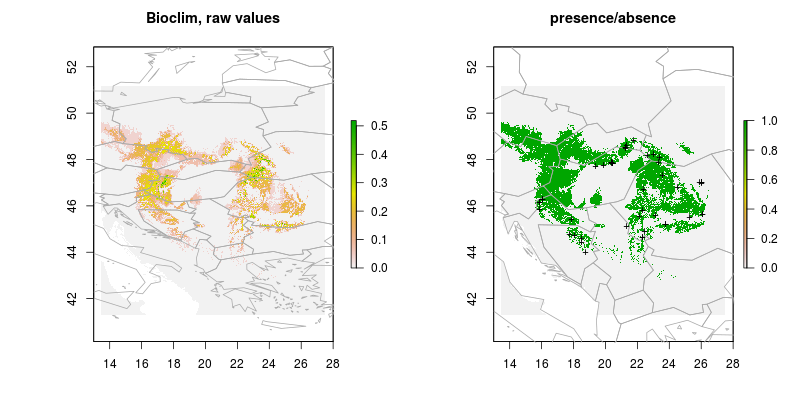
# Modelling Results

Application of several SDM approaches to the data presented above.

## Bioclim

Application of the Bioclim model (Booth et al. 2013)

# 



Results of Bioclim model.

class          : ModelEvaluation

n presences    : 13

n absences     : 16

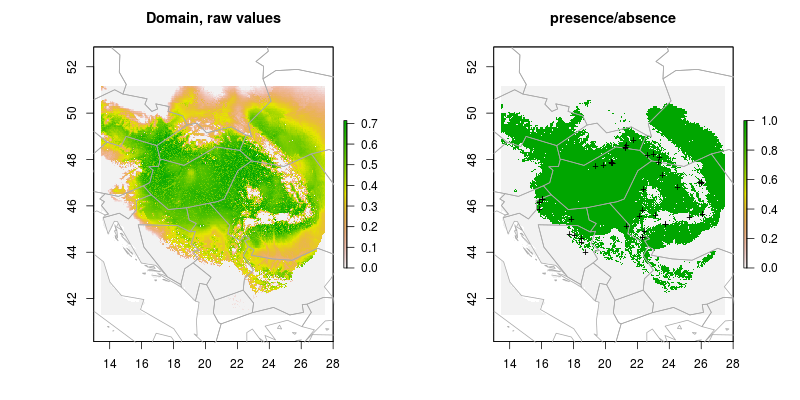
AUC            : 0.8990385

cor            : 0.6514319

max TPR+TNR at : 0.03693704

## Domain

Application of the Domain model (Carpenter, Gillison, and Winter 1993).



Results of domain model.

class          : ModelEvaluation

n presences    : 13

n absences     : 16

AUC            : 0.7331731

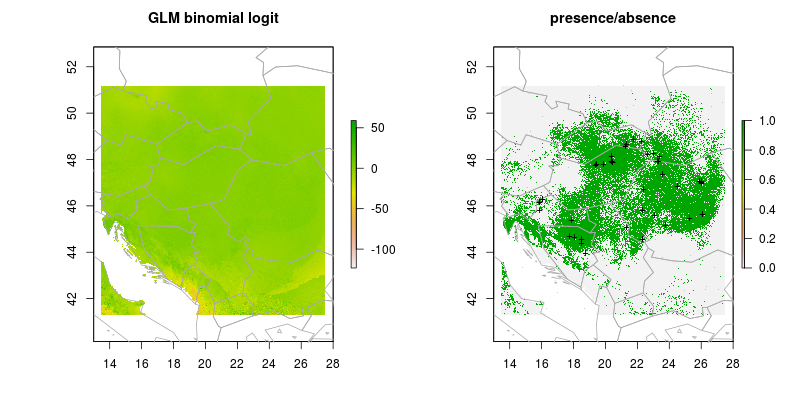
cor            : 0.4003718

max TPR+TNR at : 0.3602315

## Generalized Linear Models (GLM)

Application of the GLM method (Guisan, Edwards, and Hastie 2002).

### Binomial



GLM binomial logit result.

class          : ModelEvaluation

n presences    : 13

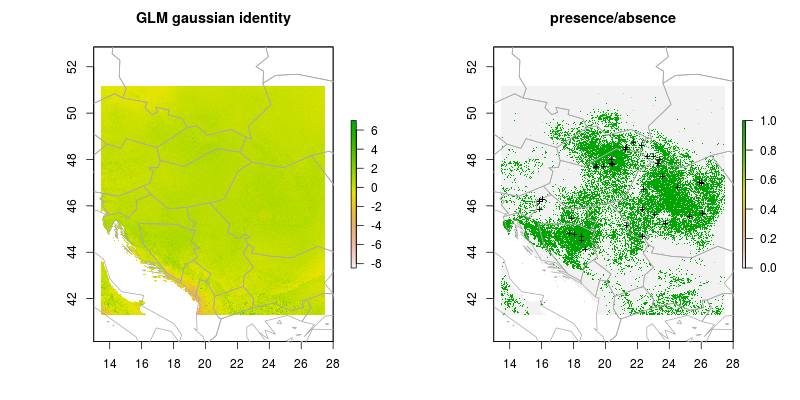
n absences     : 20

AUC            : 0.7807692

cor            : 0.3320665

max TPR+TNR at : -0.7914661

### Gaussian



GLM gaussian identity result.

class          : ModelEvaluation

n presences    : 13

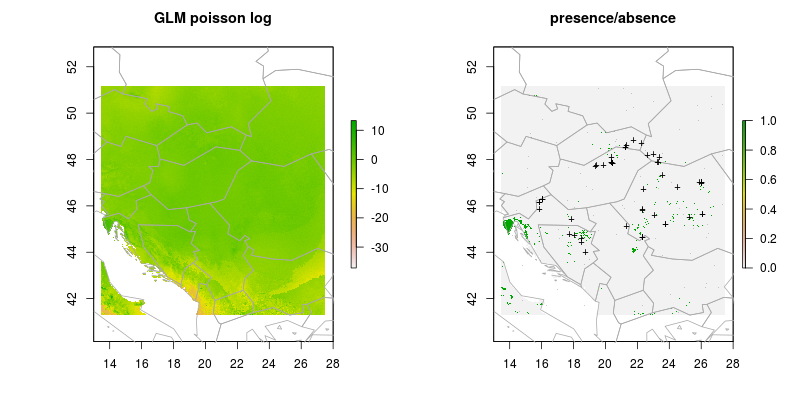
n absences     : 20

AUC            : 0.7807692

cor            : 0.4246963

max TPR+TNR at : 0.5096578

### Poisson



GLM poisson log.

class          : ModelEvaluation

n presences    : 13

n absences     : 20

AUC            : 0.8192308

cor            : 0.3797152

max TPR+TNR at : -1.040075

## Boosted Regression Trees

ToDo

(Elith, Leathwick, and Hastie 2008)

## Random Forest

(Breiman 2001)

## Support Vector Machines

(Vapnik 1998)

# Combining the model predictions

## Average score

## Weighted mean

# References

Jarvis, Andy, Hannes Isaak Reuter, Andrew Nelson, and Edward Guevara. 2008. “Hole-Filled Seamless SRTM Data V4, International Centre for Tropical Agriculture (CIAT)”.

Shapiro, Michael, and Olga Waupotitsch. 2015. “GRASS GIS R.slope.aspect Tool”. U.S.Army Construction Engineering Research Laboratory. <https://grass.osgeo.org/grass70/manuals/r.slope.aspect.html.>

Willmes, C., E. Hölzchen, C. Sommer, and J. Rodriguez. 2017. “Spatio-Temporal Interpolation of Paleoclimate Data Based on δ18O​ Climate Variability Observations (in Prep.)”. *Quarternary International*. <https://www.authorea.com/155404/-9NUD9aEaZHN8y6vcNvRiA.>

Hijmans, Robert J, and Jane Elith. 2016. “Species Distribution Modeling with R”. *Dismo R Package Version 1.1-4*. <https://cran.r-project.org/web/packages/dismo/index.html.>

Willmes, Christian. 2017. “Carpatian Basin 30ky Bp GIS Data Set”. Dataset. CRC806-Database. doi:10.5880/SFB806.37.

Booth, Trevor H., Henry A. Nix, John R. Busby, and Michael F. Hutchinson. 2013. “Bioclim: the First Species Distribution Modelling Package Its Early Applications and Relevance to Most CurrentMaxEntstudies”. Edited by Janet Franklin. *Diversity and Distributions* 20 (1). Wiley-Blackwell: 1–9. doi:10.1111/ddi.12144.

Carpenter, G., A. N. Gillison, and J. Winter. 1993. “DOMAIN: a Flexible Modelling Procedure for Mapping Potential Distributions of Plants and Animals”. *Biodiversity & Conservation* 2 (6): 667–80. doi:10.1007/BF00051966.

Guisan, Antoine, Thomas C Edwards, and Trevor Hastie. 2002. “Generalized Linear and Generalized Additive Models in Studies of Species Distributions: Setting the Scene”. *Ecological Modelling* 157 (2-3). Elsevier BV: 89–100. doi:10.1016/s0304-3800(02)00204-1.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. “A Working Guide to Boosted Regression Trees”. *Journal of Animal Ecology* 77 (4). Wiley-Blackwell: 802–13. doi:10.1111/j.1365-2656.2008.01390.x.

Breiman, Leo. 2001. “Random Forests”. *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Vapnik, Vladimir N. 1998. *Statistical Learning Theory*. Wiley.