

Modeling present and prospective distribution of *Phyteuma* genus in Carpathian region with machine learning techniques using open climatic and soil data

Alexander Mkrтчian

Geographic faculty, Ivan Franko National University of Lviv, Doroshenko st. 41, Lviv, Ukraine

alemkrt@gmail.com

Data preparation

- Get data on taxon occurrences for study area
- Remove data duplicates/dubious data, if needed
- Simulate background data, if needed

- Get data on predictive variables
- Reproject/resample data, if needed

Data preprocessing

Extracting values of predictive variables for locations of occurrences (SWD object constructed with *prepareSWD*)

Modeling

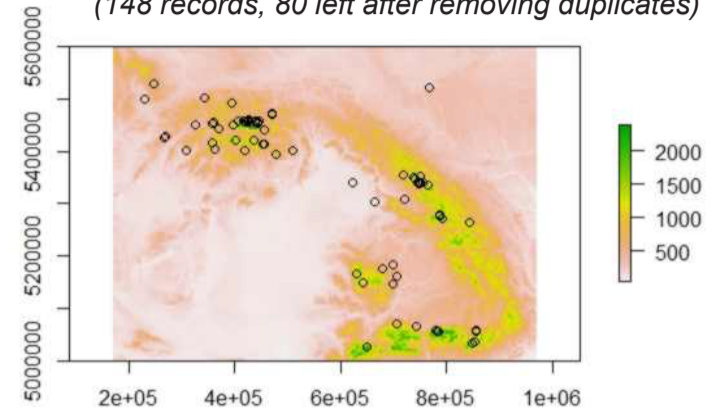
Building a model
(SDMmodel object constructed with *train*)

Making a prediction
(*RasterLayer* object with *predict*)



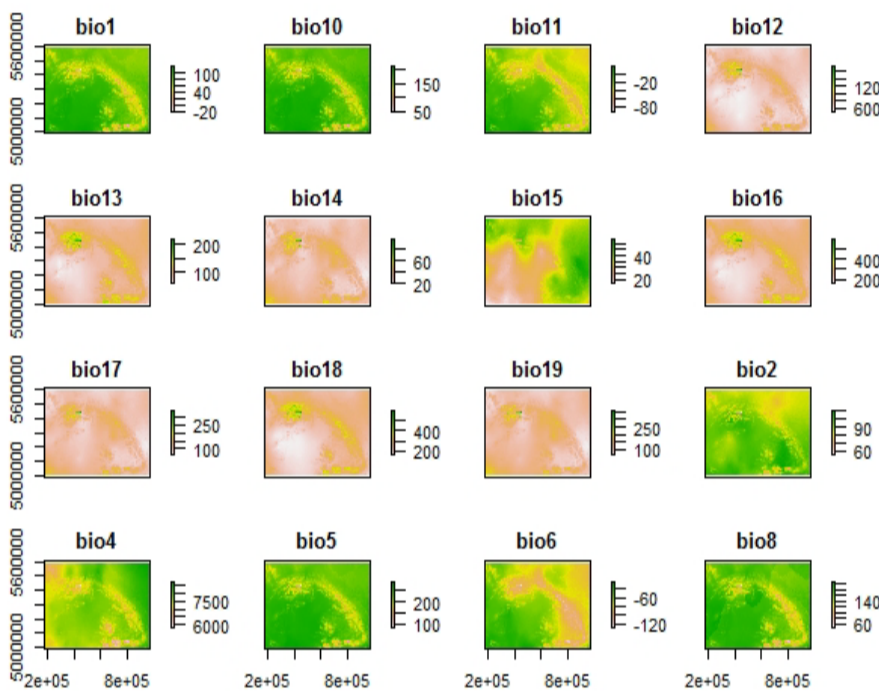
Predictive variables

Phyteuma genus occurrences in GBIF (148 records, 80 left after removing duplicates)



BIOLIM variables

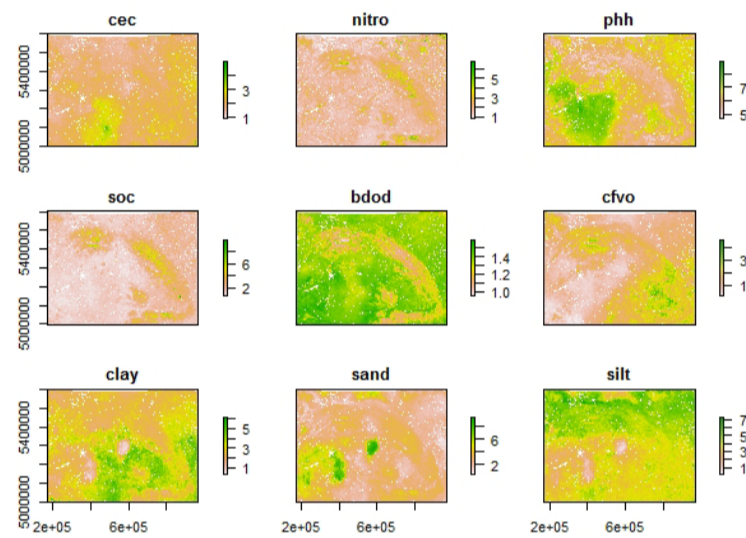
(Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978)



- BIO1 = Annual Mean Temperature
- BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
- BIO4 = Temperature Seasonality (standard deviation × 100)
- BIO5 = Max Temperature of Warmest Month
- BIO6 = Min Temperature of Coldest Month
- 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

SoilGrids variables

(Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, et al. (2017) SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE 12(2): e0169748)

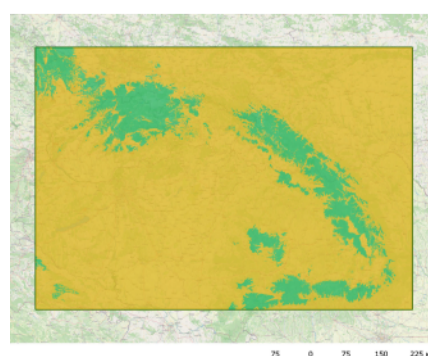
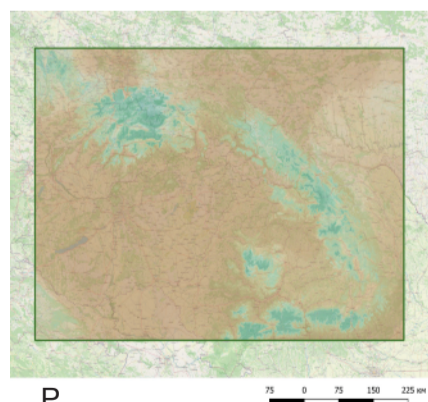


- CEC = Cation Exchange Capacity of the soil
- nitro = Total nitrogen (N)
- phh = Soil pH
- soc = Soil organic carbon content in the fine earth fraction
- bdod = Bulk density of the fine earth fraction
- cfvo = Volumetric fraction of coarse fragments (> 2 mm)
- clay = Proportion of clay particles (< 0.002 mm) in the fine earth fraction
- sand = Proportion of sand particles (> 0.05 mm) in the fine earth fraction
- silt = Proportion of silt particles (≥ 0.002 mm and ≤ 0.05 mm) in the fine earth fraction

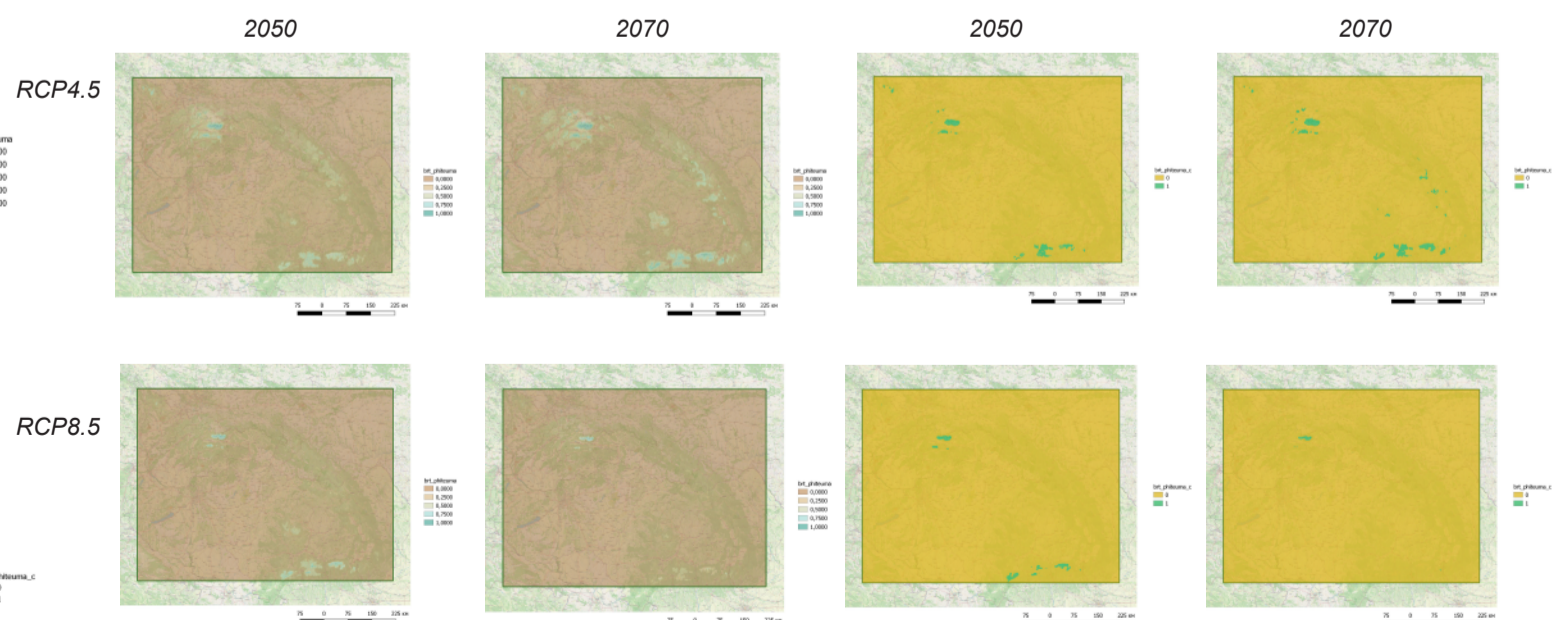
Model diagnostics (6-fold cross-validation)

Model type	AUC	COR	TSS	Deviance
General linear model	0.91	0.73	0.75	0.86
Support vector machines	0.92	0.75	0.77	0.73
Random forest	0.96	0.81	0.84	0.51
Boosted regression trees	0.92	0.75	0.76	0.78

Present distribution of taxon



Prospective distribution of taxon (calculated with BRT model)



Changes in Phyteuma range

Area, present, km ²	Area, prospective, km ²		
	RCP	2050	2070
98733	RCP4.5	3325	5272
	PCP8.5	2034	379

Supported by the International Visegrad Fund

