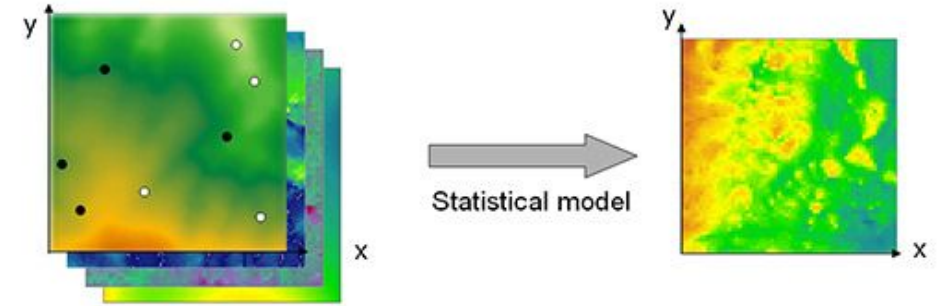
Open Data Sources for Species Distribution Modeling:

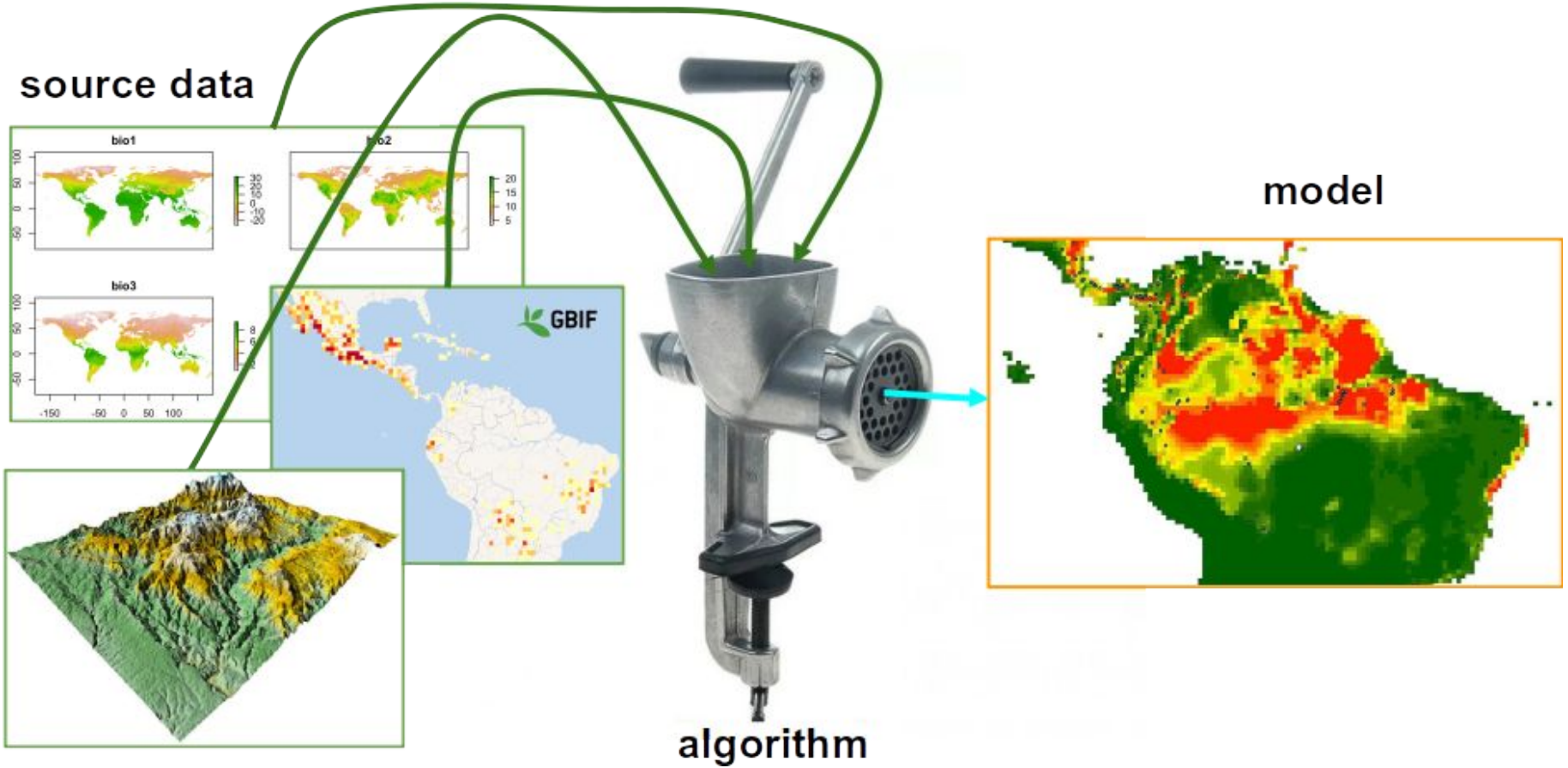
Species Occurrences and Environmental Predictors



Maxim Shashkov
Karaganda Buketov University

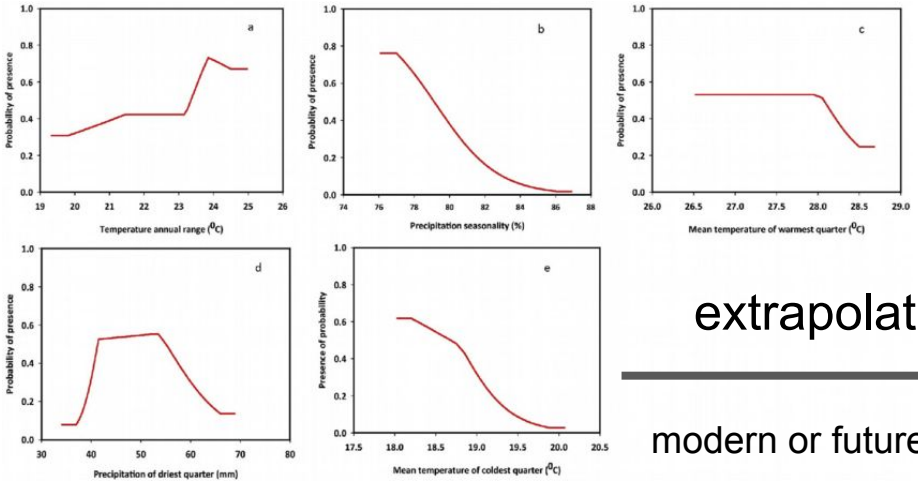
Almaty - 2023

Species Distribution Modelling, the first approach



Species Distribution Modelling

Explanation *response curves from MaxEnt*



extrapolation

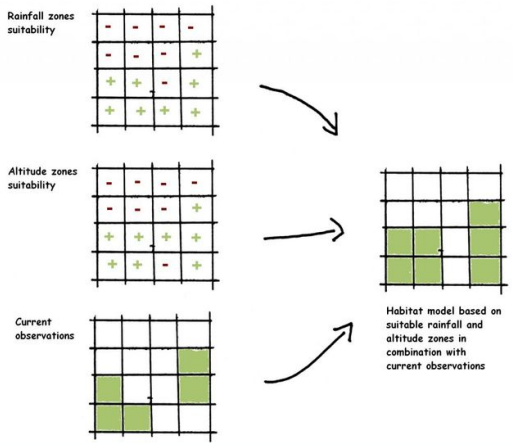
modern or future climate

Quantifying the relationship between the geographical distribution of a population of a species under study and environmental factors in order to model its distribution under present, past or future conditions

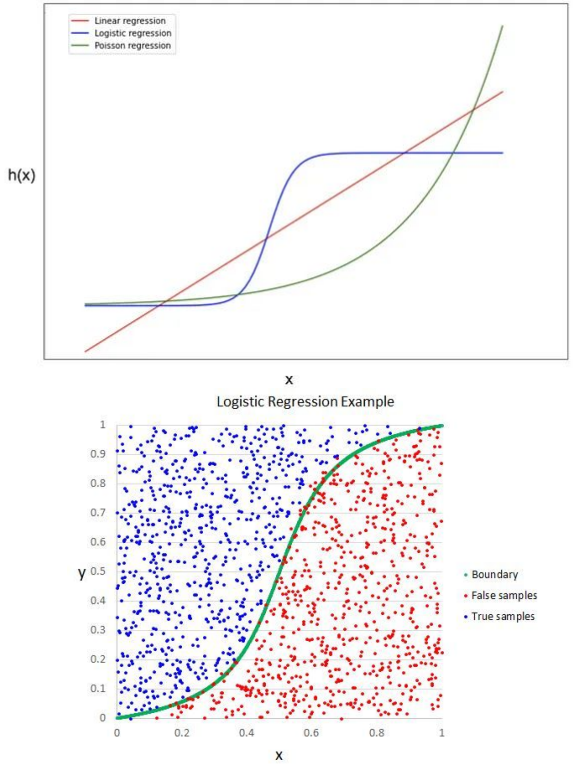
Prediction

Modelling methods

climatic envelope

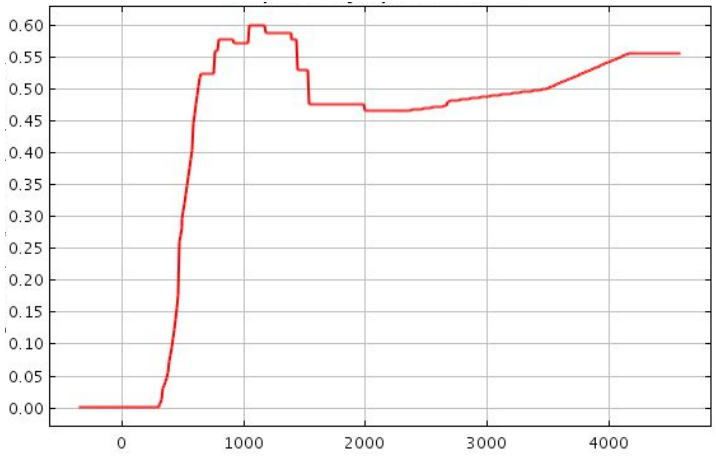


regression analysis



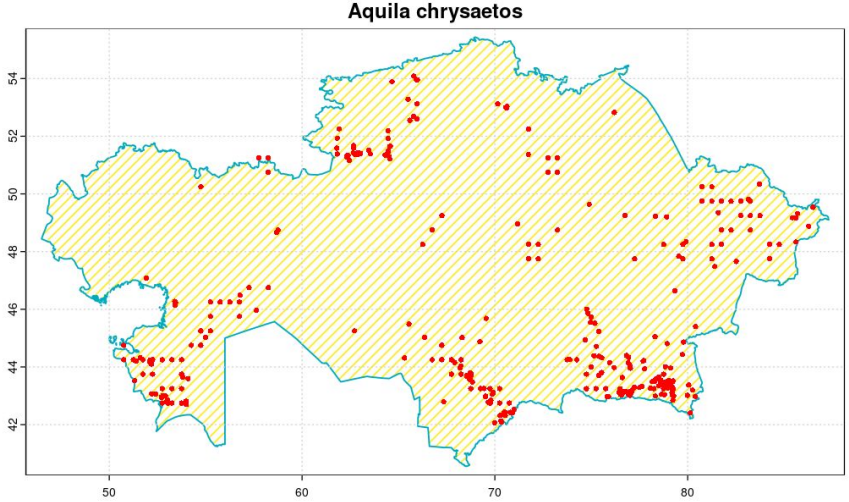
machine learning

MaxEnt
Boosted Regression Trees
Random Forest



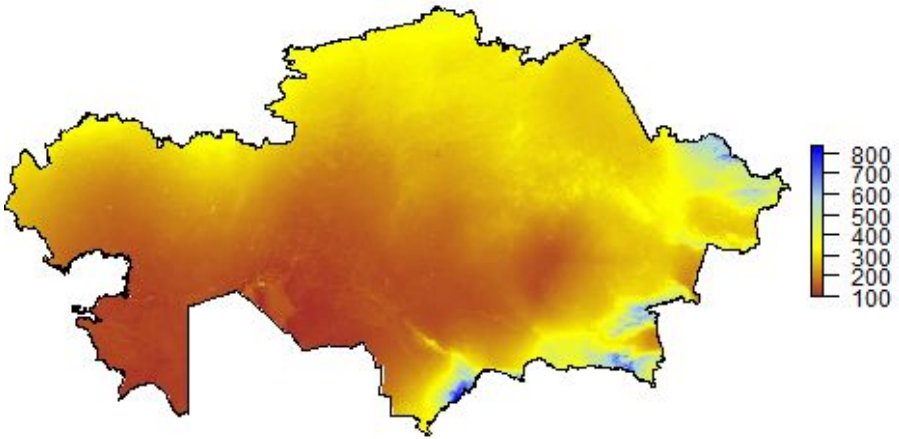
Input Data: occurrence points and environmental layers

Occurrences of focal species - vector layer



response variable - what we trying to explain in the model output - probability of occurrence

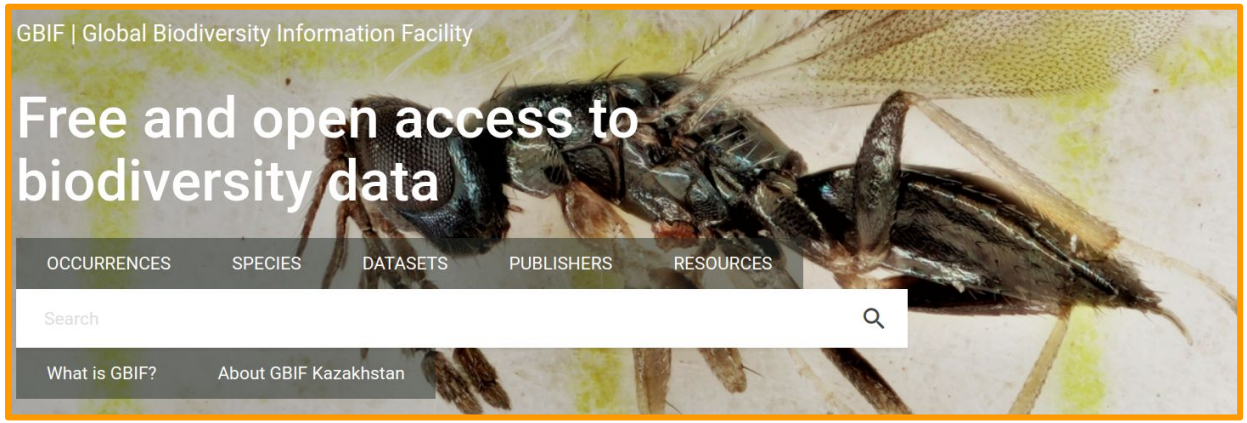
Environmental variables - raster layer



predictors - independent (explanatory) variables

There are plenty of open source to obtain data both on species occurrences and environmental variables

Global Biodiversity Information Facility - GBIF.org

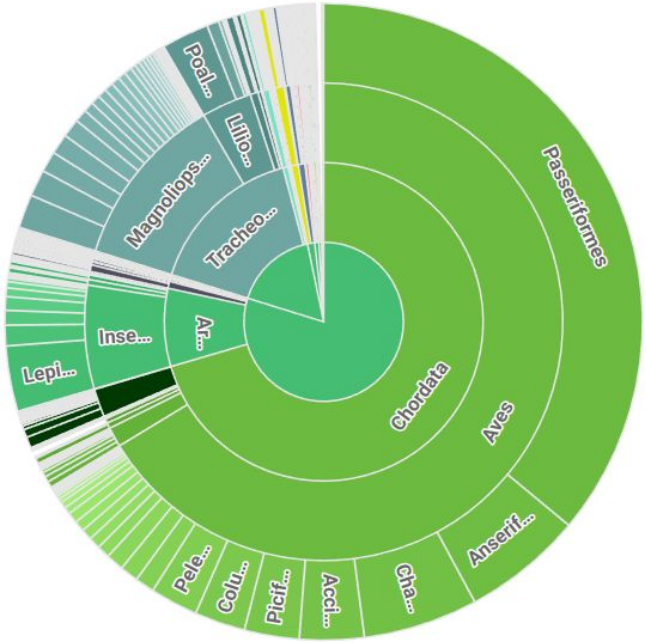


2,578,743,049
Occurrence records

90,136
Datasets

2,128
Publishing institutions

9,348
Peer-reviewed papers using data

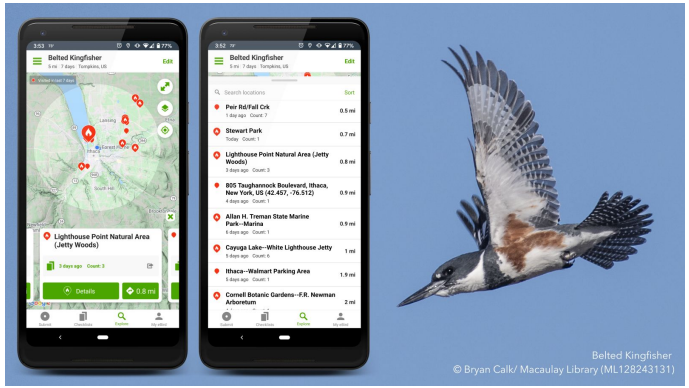


1,710,054,371 records on **birds** including:
collection specimens ~ 8.5 млн
machine observation > 9 млн

the main source of open biodiversity data

Species Occurrence

eBird



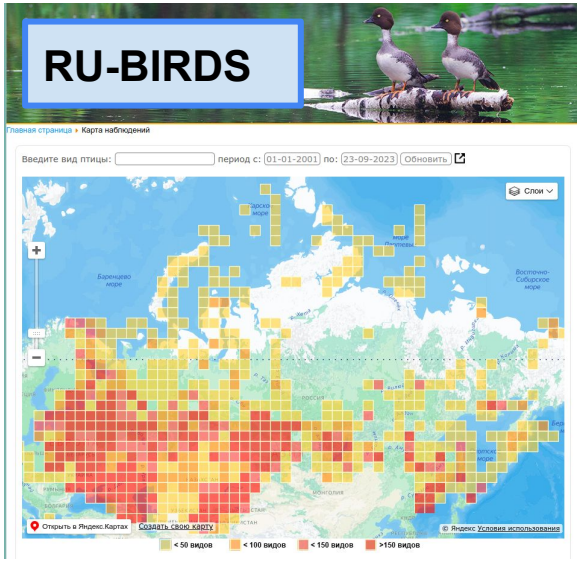
iNaturalist

15 M records

820 000 orservers
1 277.5 M records

BIRDA

observations.org

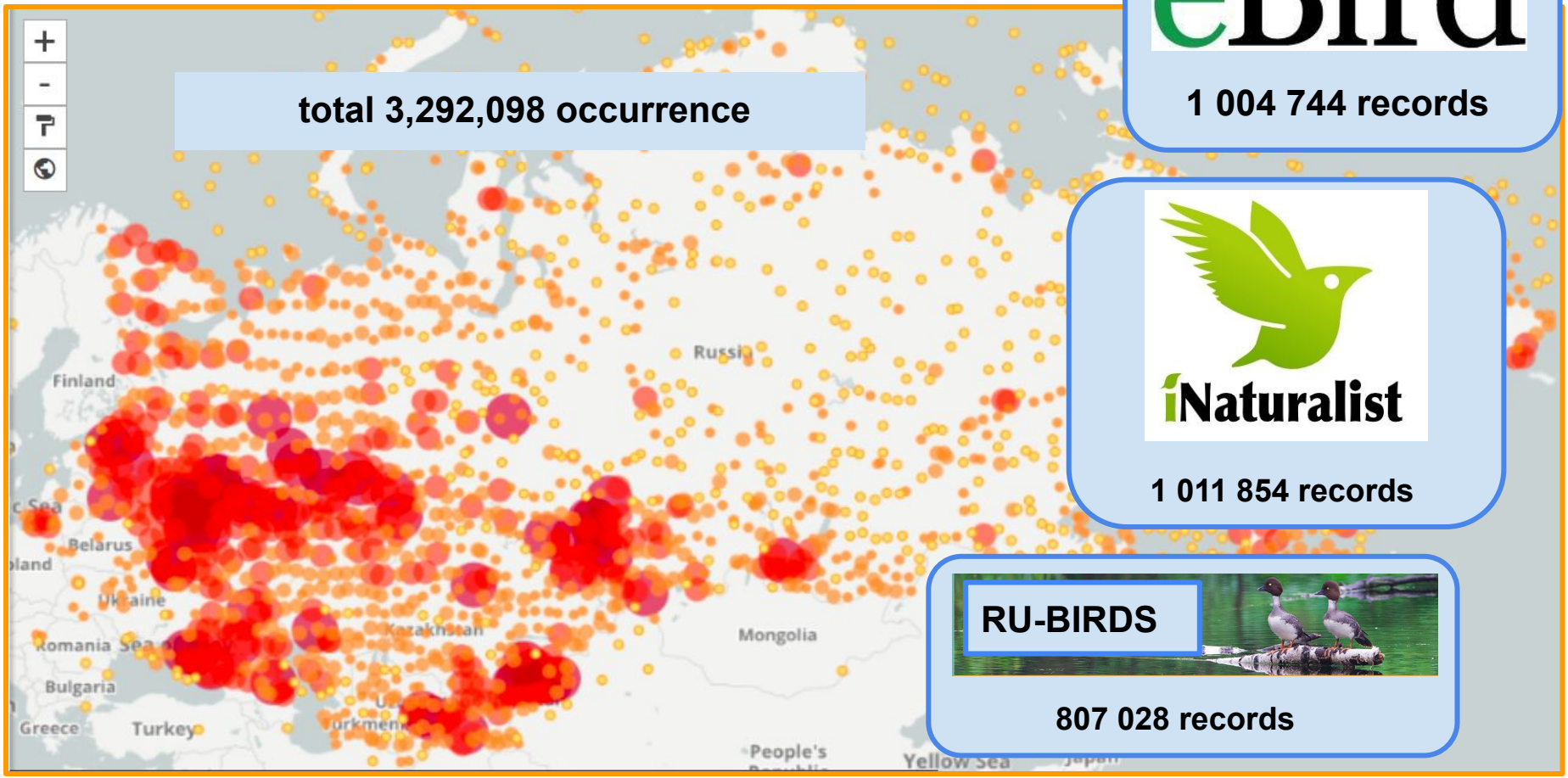


Birds of Northern Eurasia

RRRCN

Российская сеть
изучения и охраны
ПЕРНАТЫХ ХИЩНИКОВ

GBIF Birds Occurrences: North Eurasia



eBird

1 004 744 records

iNaturalist

1 011 854 records

RU-BIRDS

807 028 records

Occurrences: Data quality

out of 1 712 578 362 bird occurrences 21 685 716 supplied with images, which allow you to verify identification

with geographical coordinates: 1 688 421 712

with estimation of georeference accuracy: 316 174 456 < 20 %

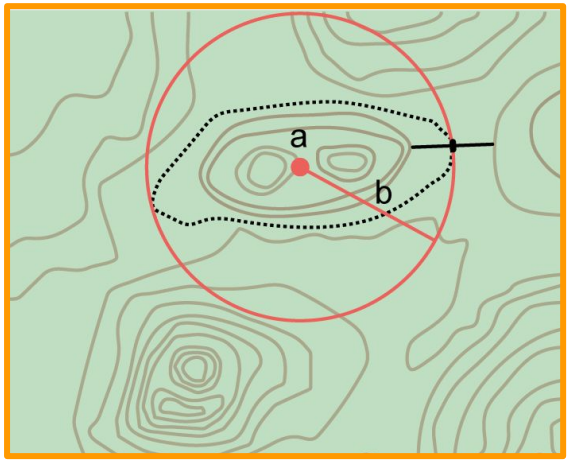
within 10 km: ~ 311 M

5 km: ~ 289 M

1 km: ~ 178 M

500 m ~ 142 M

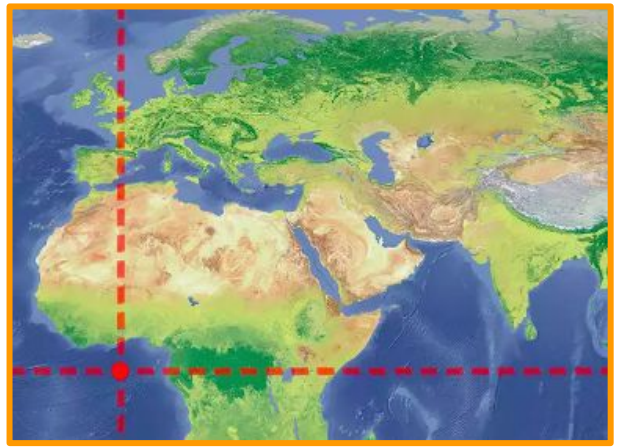
100 m ~ 67 M

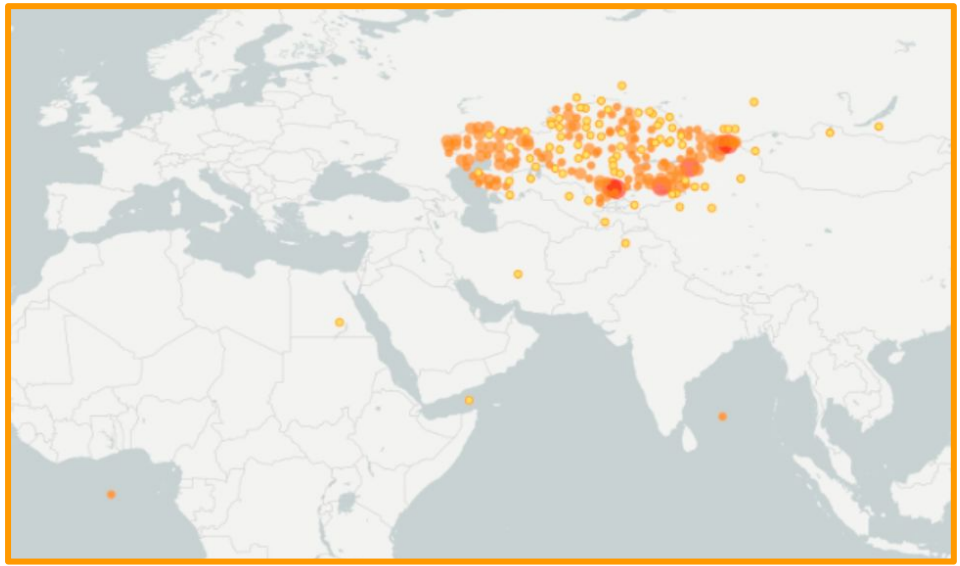


DOI: 10.35035/e09p-h128

Data Cleaning and Filtering

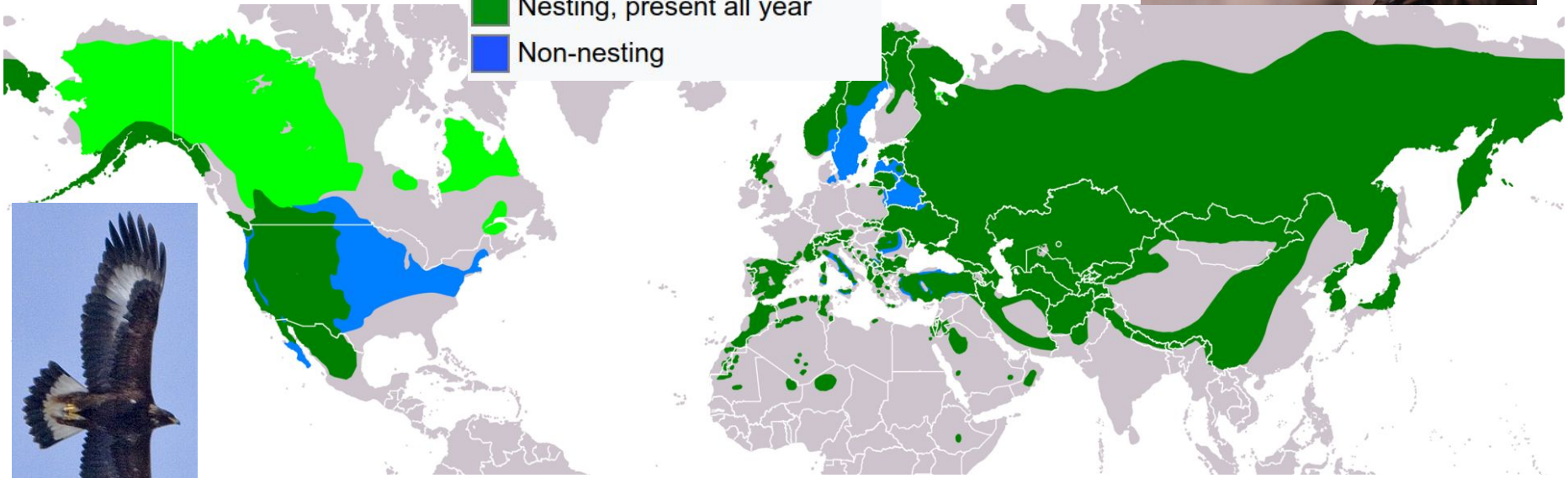
1. remove occurrence without geographic coordinates
2. remove occurrence with zero coordinates
3. remove occurrence for which the coordinates do not match the specified country
4. remove occurrence with duplicated coordinates
5. remove occurrence without uncertainty assessment
6. remove occurrence with georeference uncertainty more than NN km
7. remove occurrence without clear date
8. remove occurrence before 19XX
9. remove …





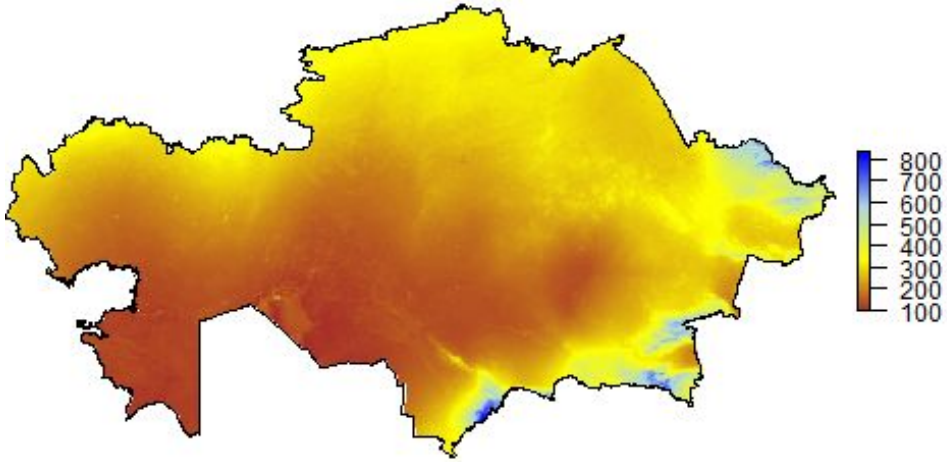
Different zones of the range

Golden Eagle



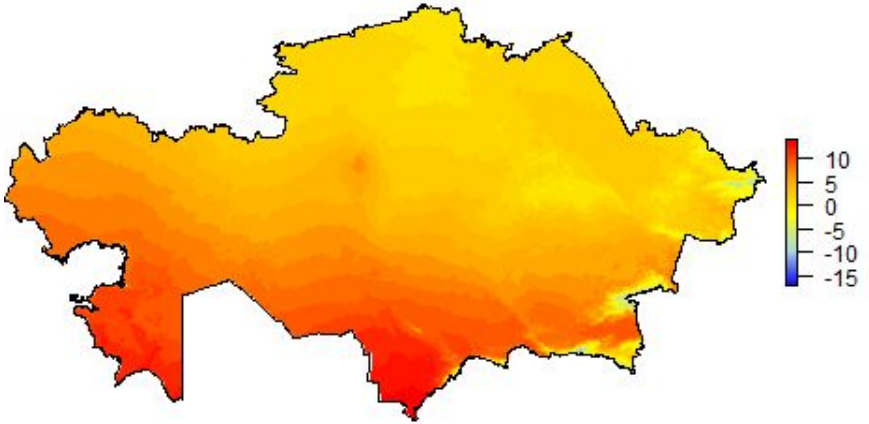
WorldClim - the most used spatial dataset for SDM

at the regional scale we are expected distribution to be according to climatic conditions



year sum of precipitation

mean year temperature



at the local scale we need to take into account habitats, land use, urbanisation, etc

SRTM - Shuttle Radar Topography Mission

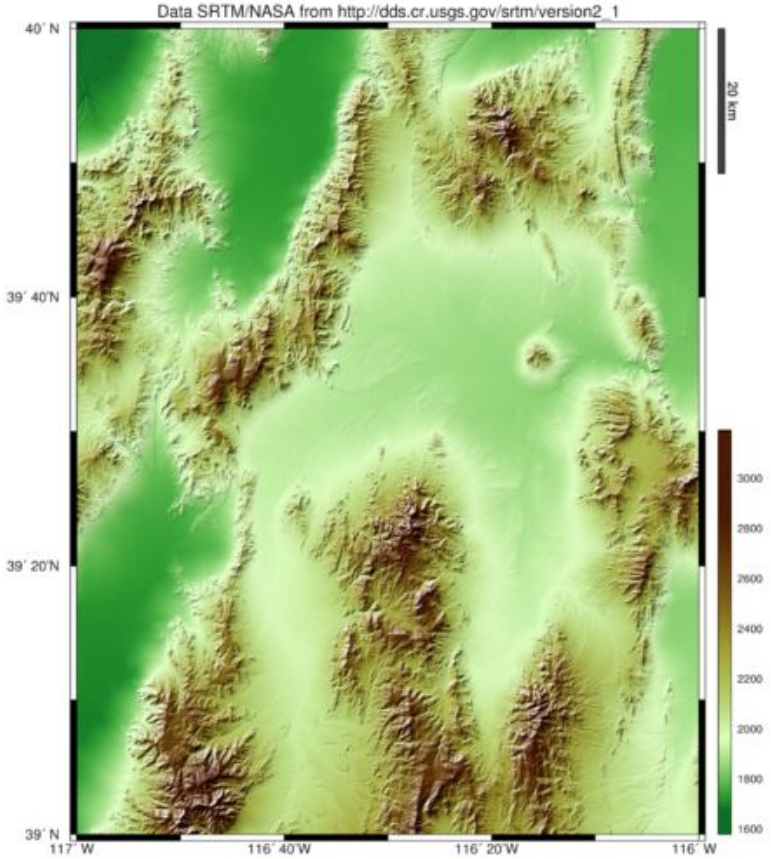
the main variable is elevation

resolution: 90m, 250m, 500m, 1km

v 1.0
2003

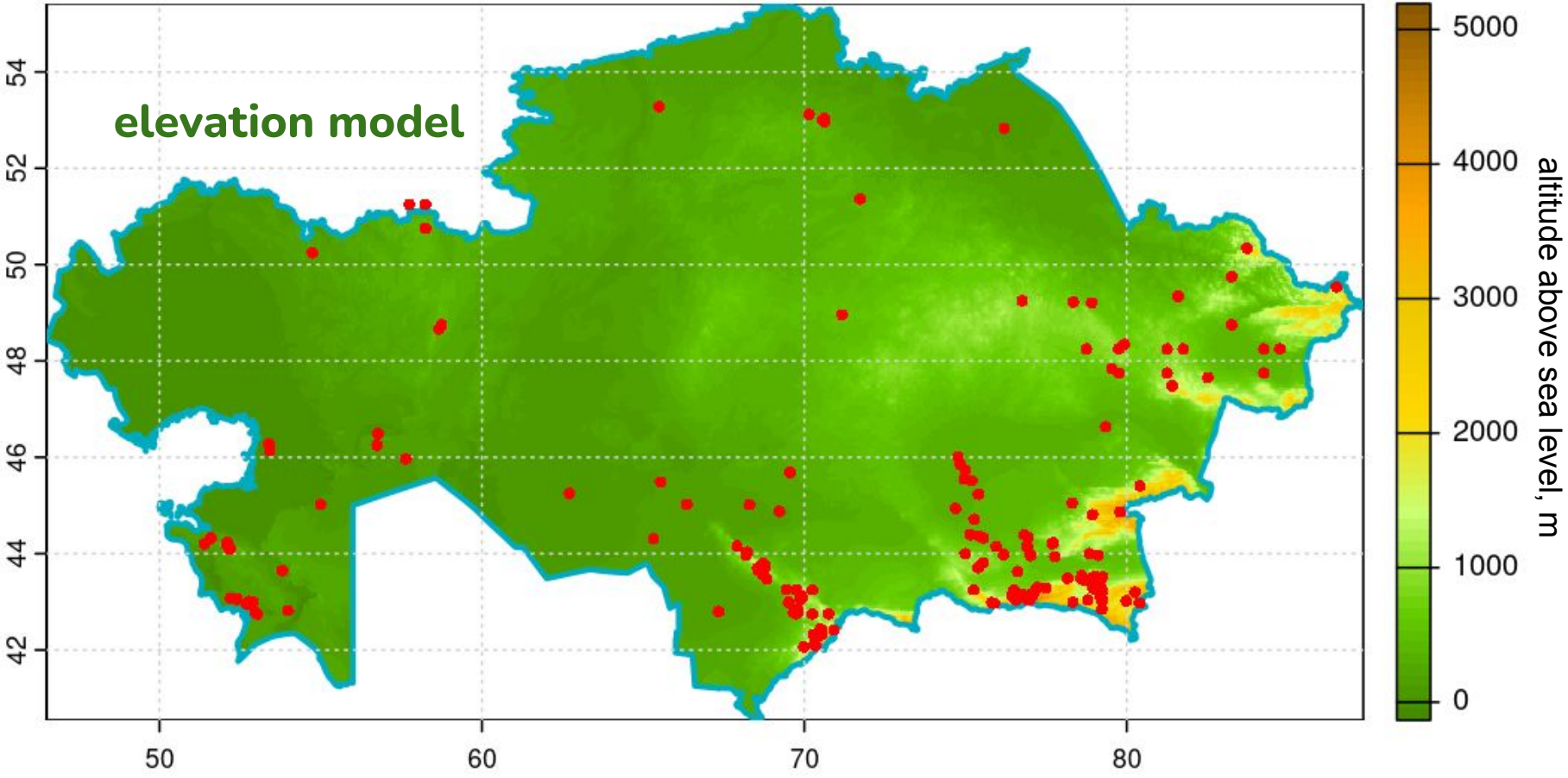
v 3.0
2019-11

v 4.1
actual



https://csidotinfo.wordpress.com/data/srtm-90m-digital-elevation-database-v4-1/

Observations of Golden Eagle *Aquila chrysaetos*



Remote Sensing Data: Landsat and Sentinel satellites

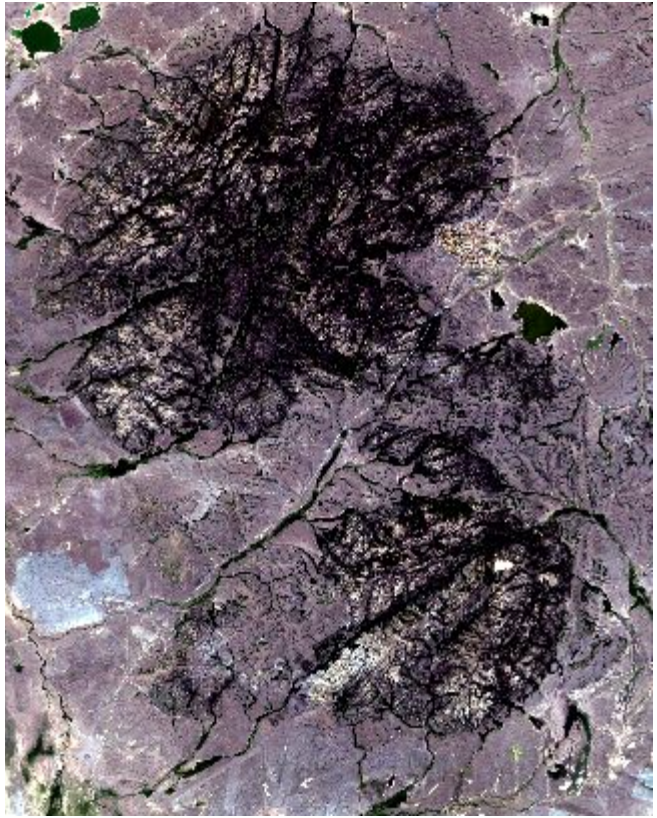
multispectral imaging of land cover

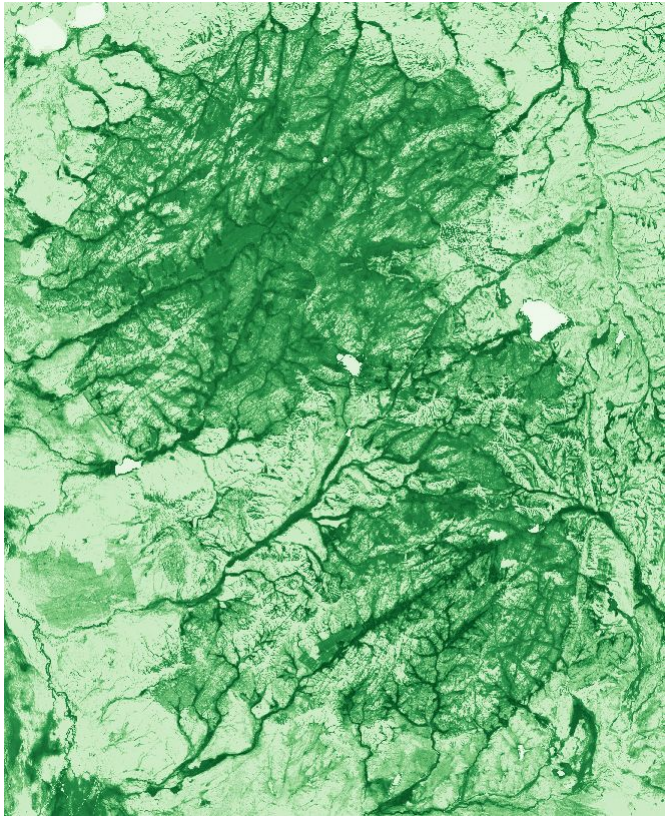
Landsat-5 since 1984 (30 m resolution),

Landsat-9 from beginning of 2022

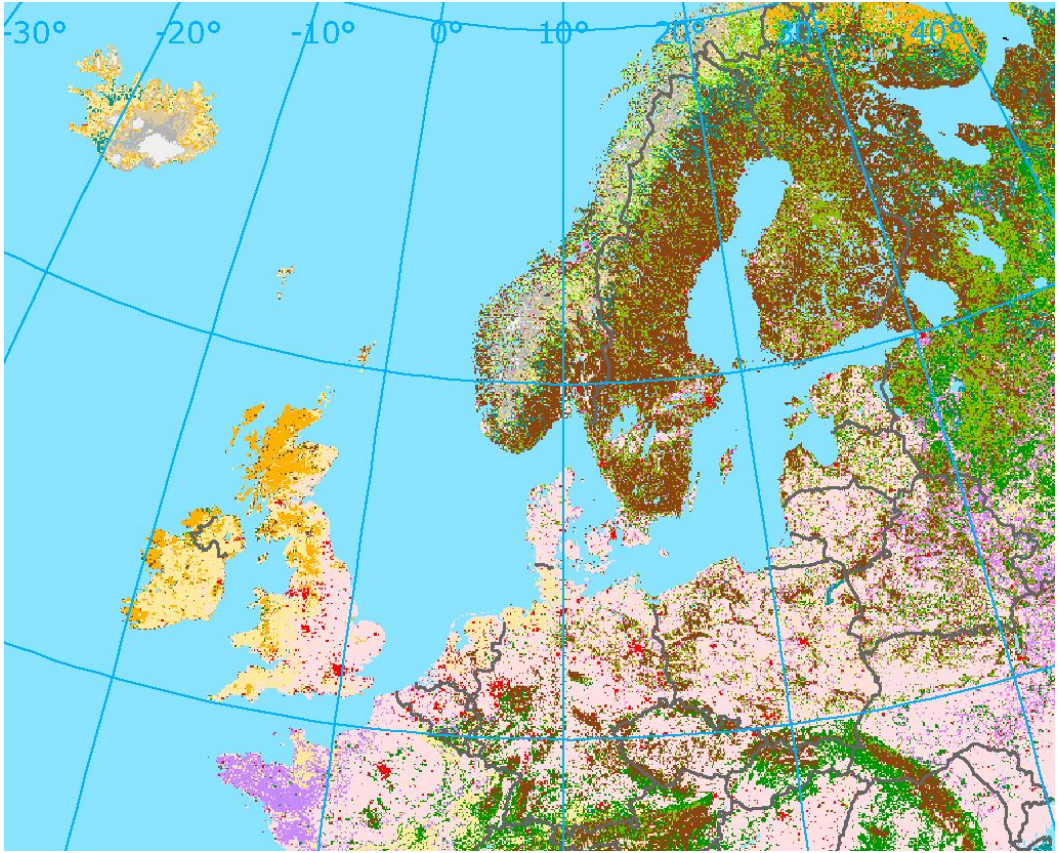
Sentinel since 2014 (10 m resolution)

NDVI - Normalized difference vegetation index





Habitat Types: Global Land Cover 2000



Global land cover 2000, 250m

- Tree Cover, broadleaved, evergreen
- Tree Cover, broadleaved, deciduous, dosed
- Tree Cover, broadleaved, deciduous, open
- Tree Cover, neddle-leaved, evergreen
- Tree Cover, neddle-leaved, deciduous
- Tree Cover, mixed leaf type
- Tree Cover, regularly flooded, fresh water
- Tree Cover, regularly flooded, saline water
- Mosaic: Tree cover/Other natrual vegetation
- Tree Cover, bunt
- Shrub Cover, dosed-open, evergreen
- Shrub Cover, dosed-open, deciduous

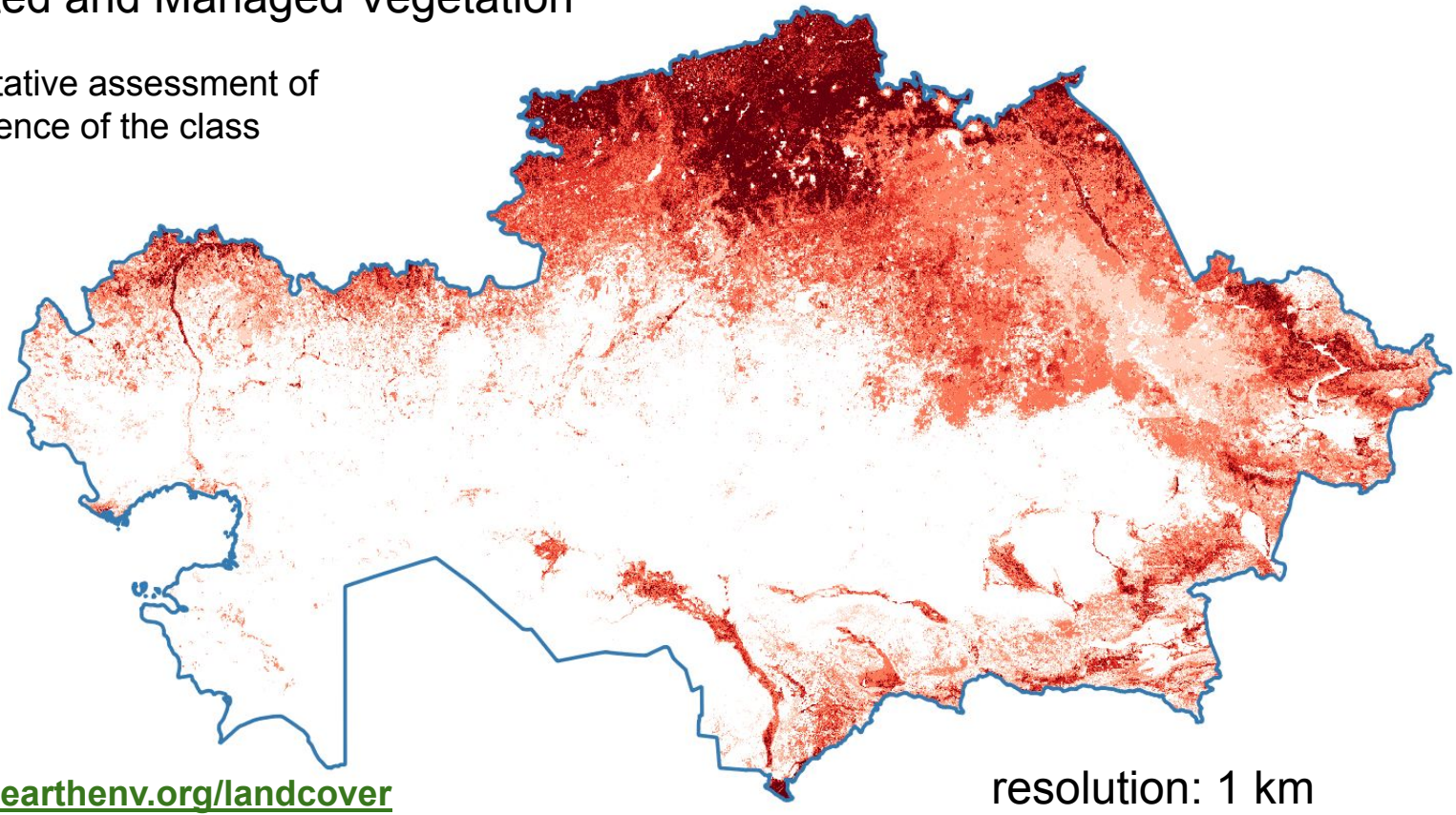
Categorical variable

forobs.jrc.ec.europa.eu/glc2000

Habitat Types: Global 1-km Consensus Land Cover

Cultivated and Managed Vegetation

quantitative assessment of prevalence of the class

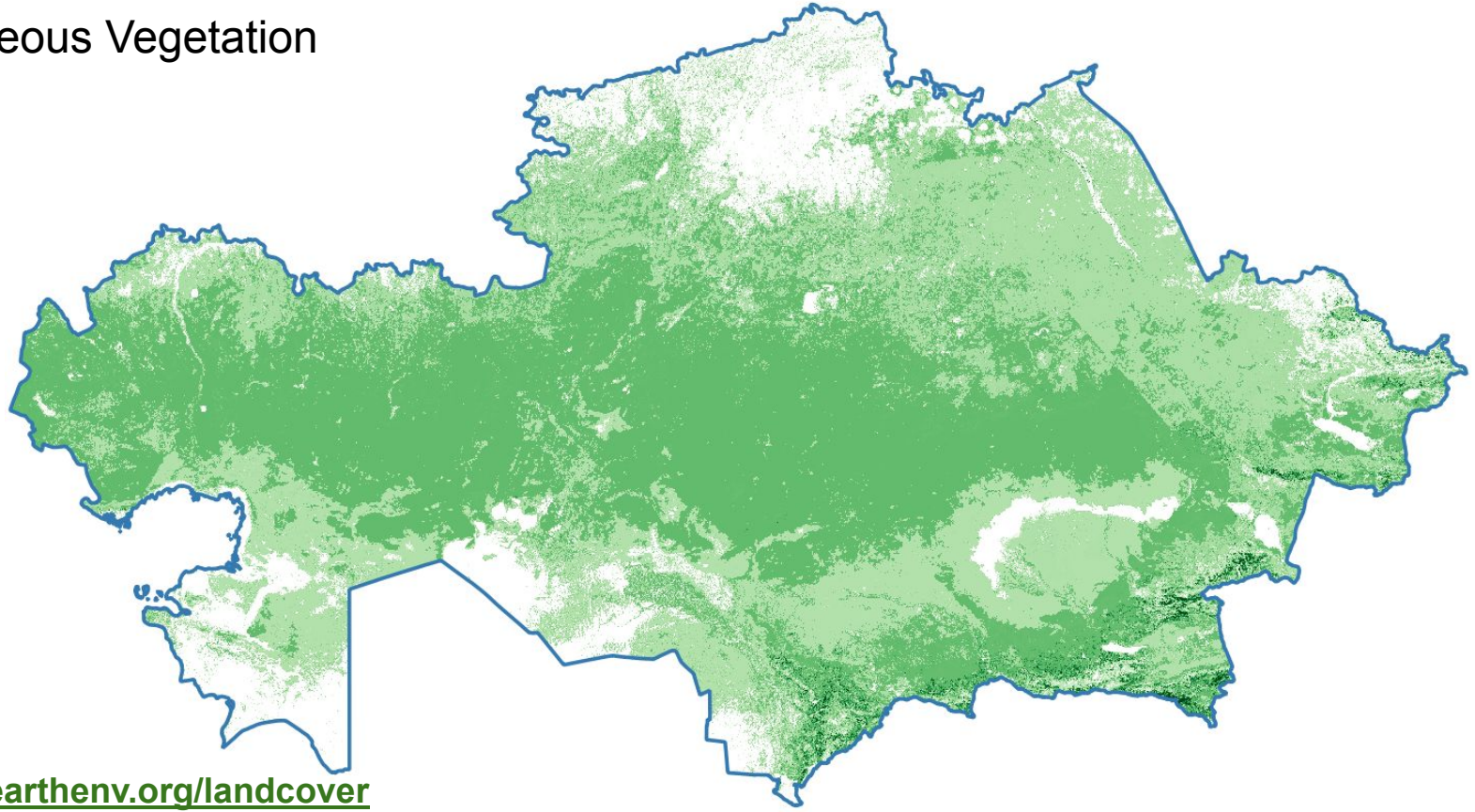


https://www.earthenv.org/landcover

resolution: 1 km

Habitat Types: Global 1-km Consensus Land Cover

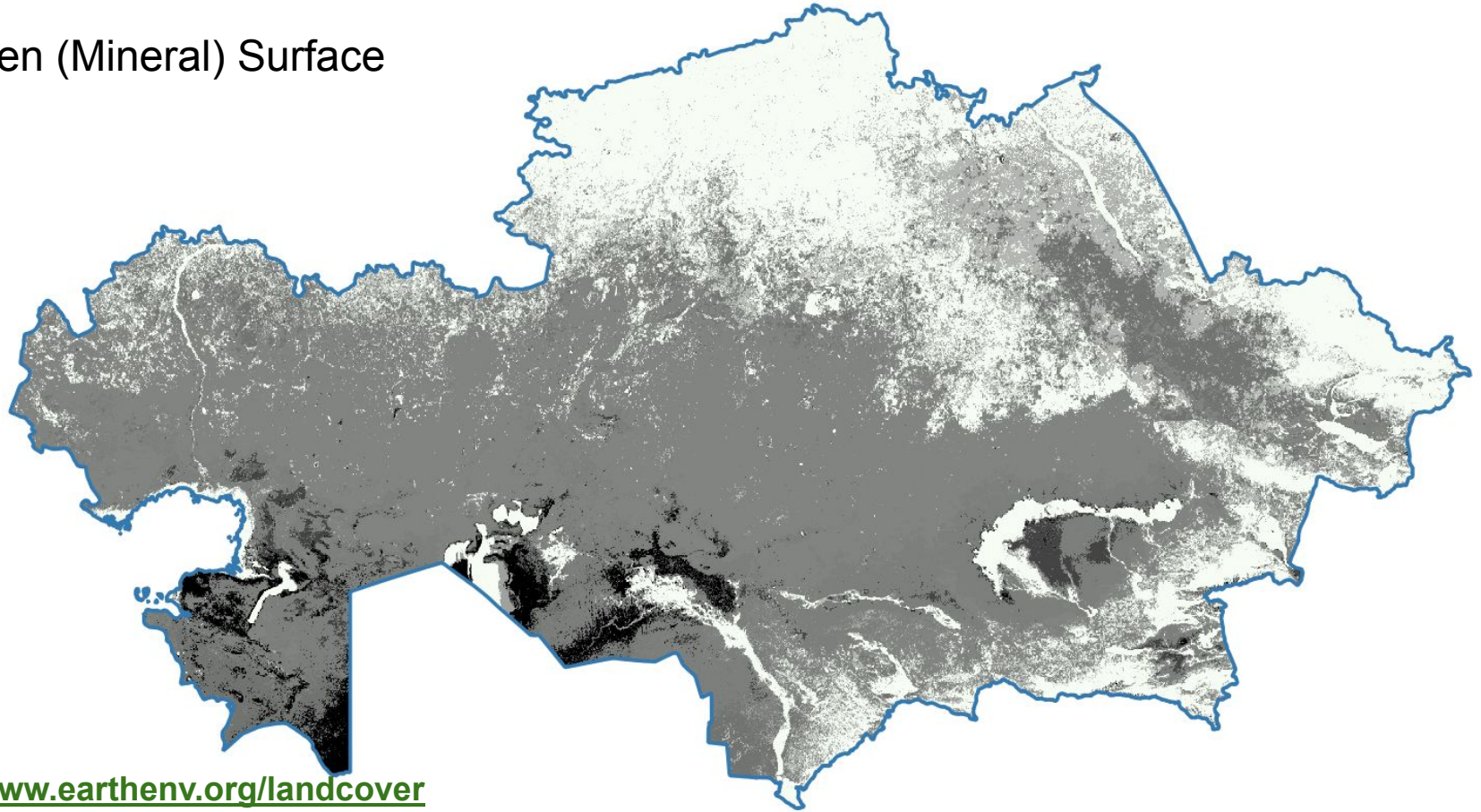
Herbaceous Vegetation



https://www.earthenv.org/landcover

Habitat Types: Global 1-km Consensus Land Cover

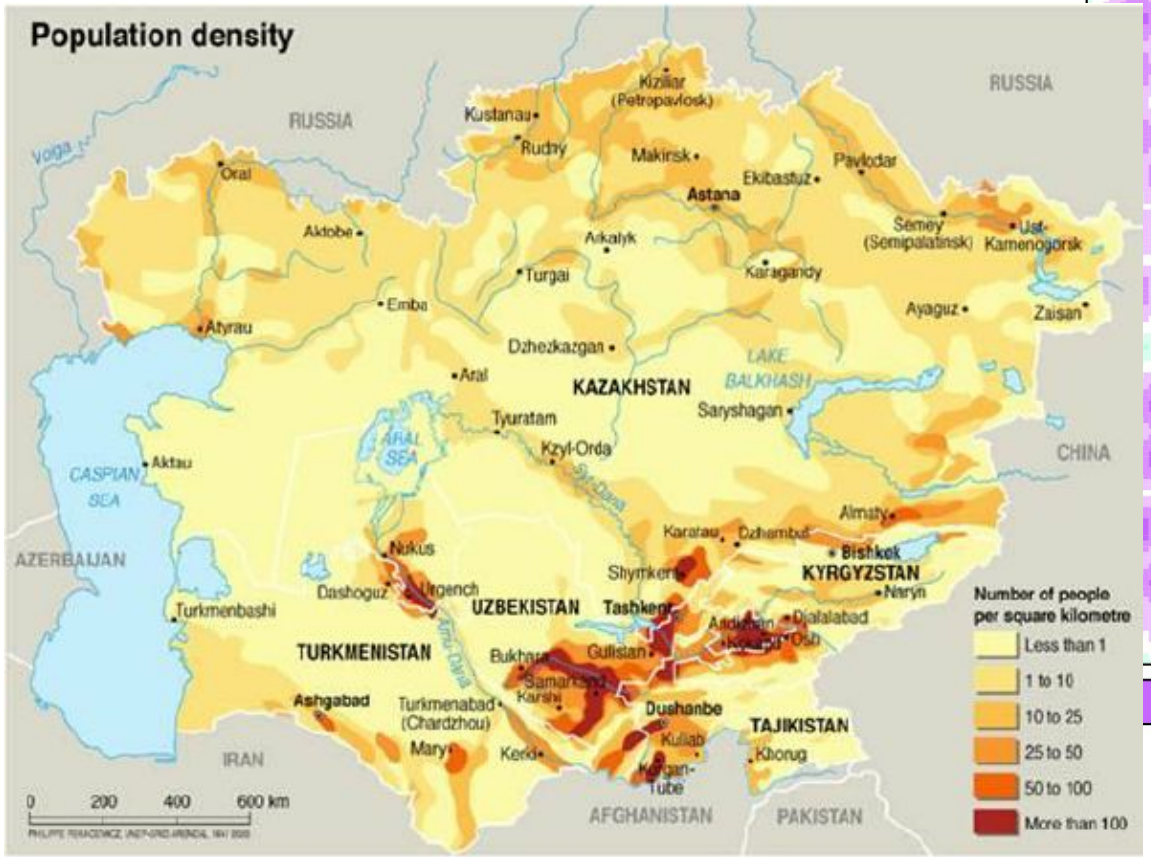
Barren (Mineral) Surface

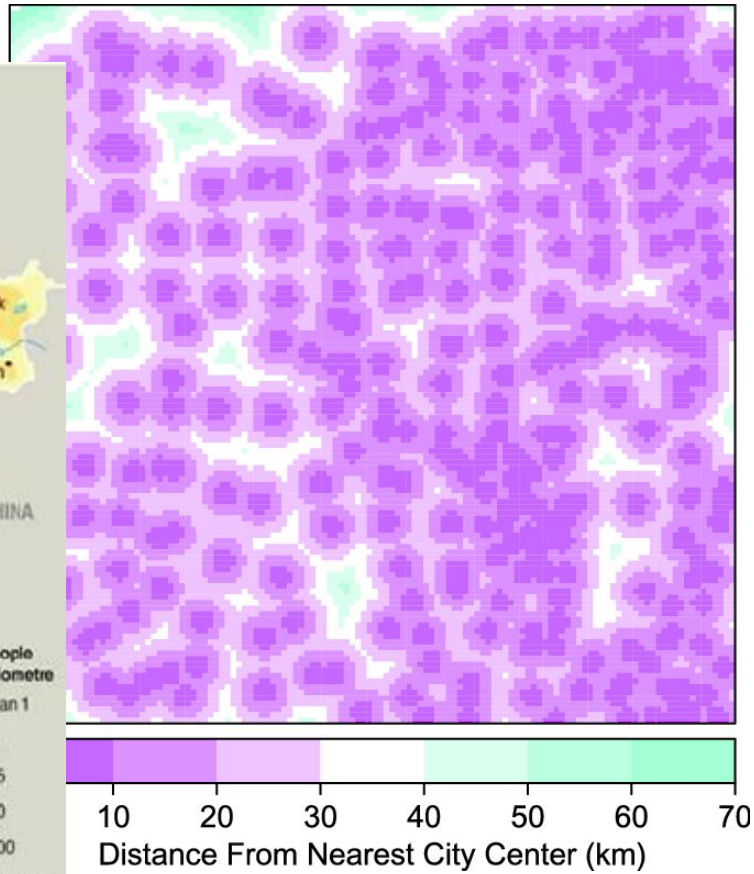


https://www.earthenv.org/landcover

Anthropogenic Influence and Infrastructure

humal population density

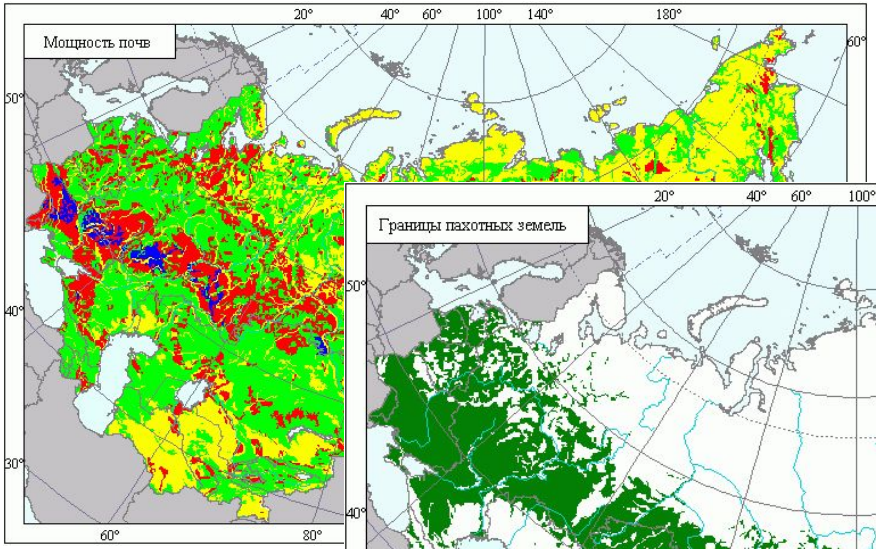




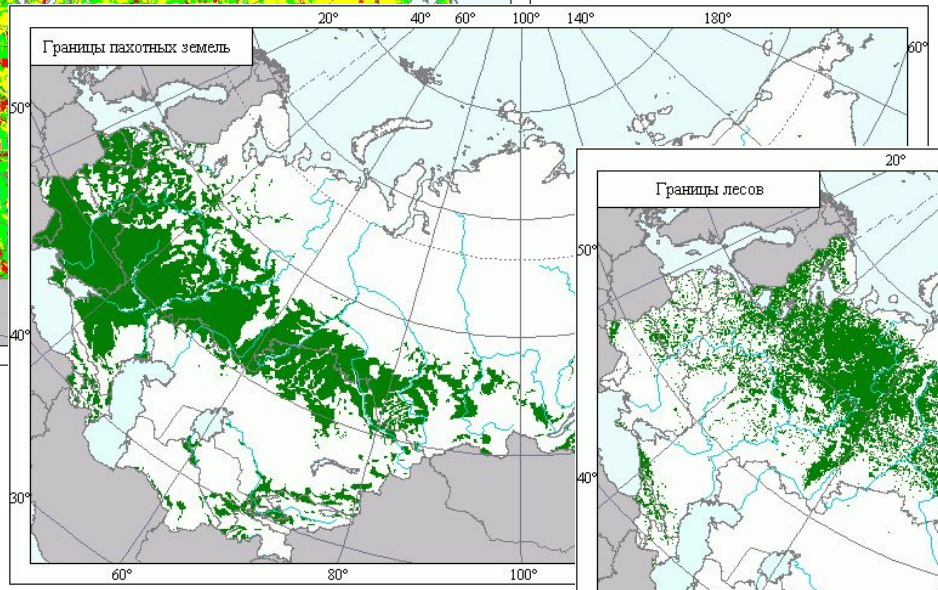
https://www.researchgate.net/figure/Population-density-in-Central-Asia_fig1_246547400

10.1175/WCAS-D-12-00040.1

Agroecological atlas of Russia and neighbouring countries



Soil thickness
quantitative discrete variable



Arable lands
qualitative binary variable

agroatlas.ru

Forests
qualitative binary variable

GADM - Global ADMinistrative database

Download GADM data (version 4.1)

Country

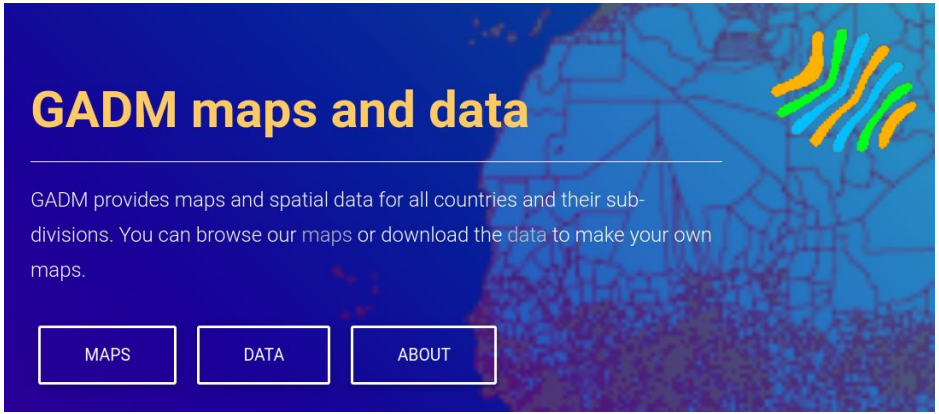
Kazakhstan

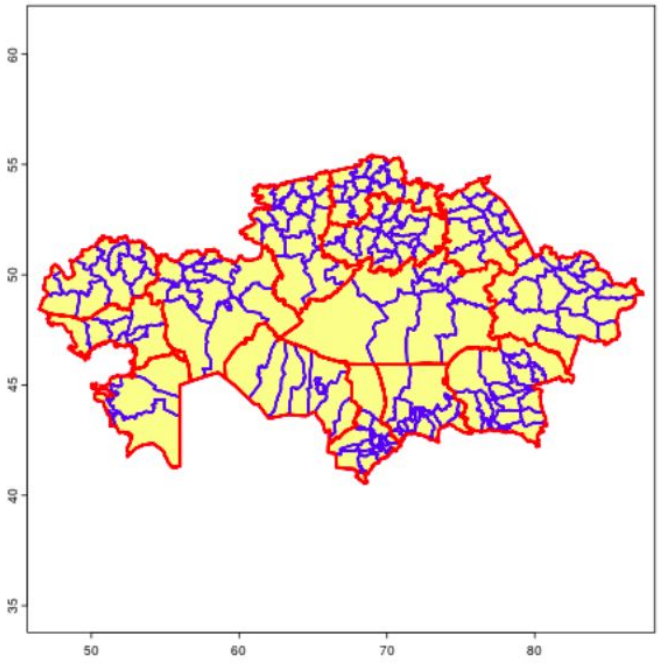
Geopackage

Shapefile

GeoJSON: level-0, level1, level2

KMZ: level-0, level1, level2





Software for Species Distribution Modelling

as conclusion

MaxEnt 3.4.4 https://biodiversityinformatics.amnh.org/open_source/maxent

Краткое введение в MaxEnt (GIS-lab)

R environment https://rspatial.org/raster/sdm (tutorial)

A curated list of R packages for species distribution modelling

Introduction to species distribution modelling (SDM) in R

Macroecology
University of Potsdam

Системы облачных вычислений ecocloud и bccvl

ecocloud beta

Dashboard Workspace Explorer

Welcome to *ecocloud*

ecocloud is a free online environment that requires no setup and runs entirely in the cloud. You can write, edit and run code, save and share your analyses through GitHub or Google Drive, and access powerful computing resources, all for free from your browser!

Sign in to get started

Cloud based processing
Big data needs big processing power. We use cloud based processing so you can save space.

Multi-code language
We offer either Python or R (and RStudio) platforms to write, edit and run your code.

Easy access
Login through your Australian University account. Otherwise, it's free to make a new account.

bccvl

About Data Portal Features Training Blog Contact Login/Register

ecocloud.org.au
bccvl.org.au
Hallgren et al., 2016, The Biodiversity and Climate Change Virtual Laboratory: Where ecology meets big data.
DOI: 10.1016/j.envsoft.2015.10.025

High performance computing from your laptop

Access and visualise a large set of biological, environmental, and climate (including future scenario) datasets and concurrently run statistical analyses on your data without burdening your personal computer.

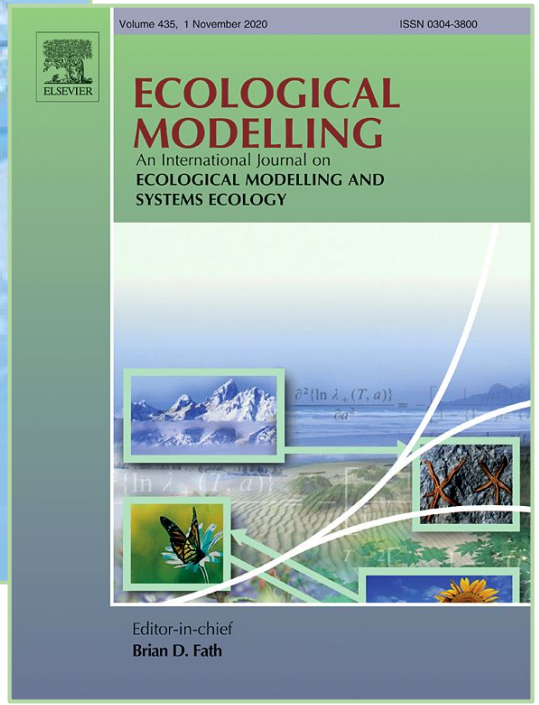
4121	306	58
Climate Data Layers	Environmental Data Layers	Average model runtime (sec)

Modelling (SDM) - литература

WILEY 1998, 5.717

Springer 2005, 3.906

ELSEVIER 1975, 3.512



Ecography

Ecological Indicators

Ecological Informatics

Remote Sensing of Environment

Environmental Modelling & Software

Global Ecology and Conservation

Journal of Applied Ecology

Literature

Phillips et al. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190:231-259. DOI: **10.1016/j.ecolmodel.2005.03.026**

Zurell et al., 2020, A standard protocol for reporting species distribution models DOI: **10.1111/ecog.04960**

Лисовский А.А., Дудов С.В., Оболенская Е.В. 2020. Преимущества и ограничения методов экологического моделирования ареалов. 1. Общие подходы. DOI: 10.31857/S0044459620020037

Лисовский А.А., Дудов С.В. 2020. Преимущество и ограничения методов экологического моделирования вреалов. 2. MaxEnt. DOI: 10.31857/S0044459620020049

Arthur D. Chapman & John R. Wieczorek (2020) Georeferencing Best Practices. Copenhagen: GBIF Secretariat. DOI: 10.15468/doc-gg7h-s853

Arthur D. Chapman (2020) Current Best Practices for Generalizing Sensitive Species Occurrence Data. Copenhagen: GBIF Secretariat. https://doi.org/10.15468/doc-5jp4-5g10.

THANK YOU FOR YOUR ATTENTION !!!

iNaturalist observation: **https://www.inaturalist.org/observations/176569028**