Species Distribution, Abundance and Survival Modeling: New Opportunities and Methods

Karyakin I.V., Knizhov K.I.

Russian Raptor Research and Conservation Network

Distribution

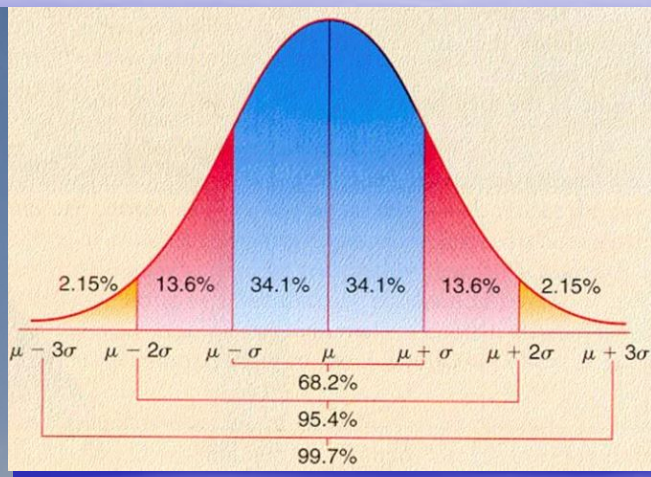
Demography (Productivity + Survival)

Abundance (Population numbers)

Demography (Productivity + Survival)

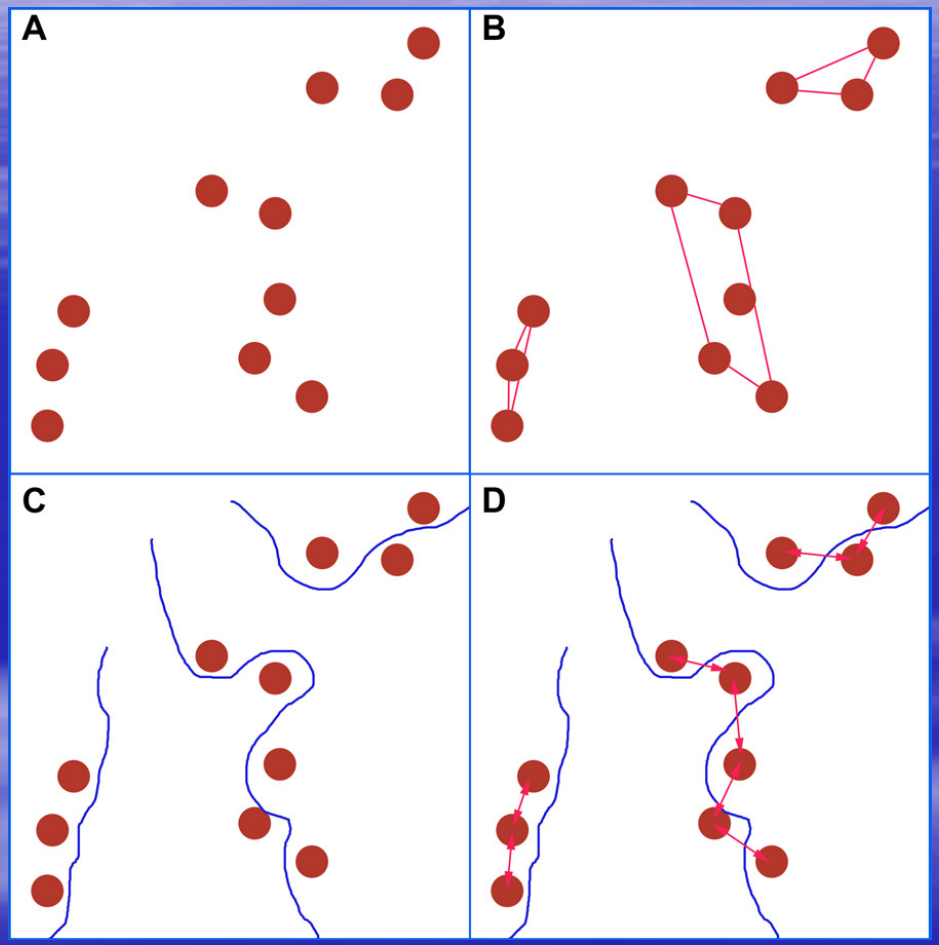
What is the main factor influencing the survival of nestlings?

Abundance (Population numbers)

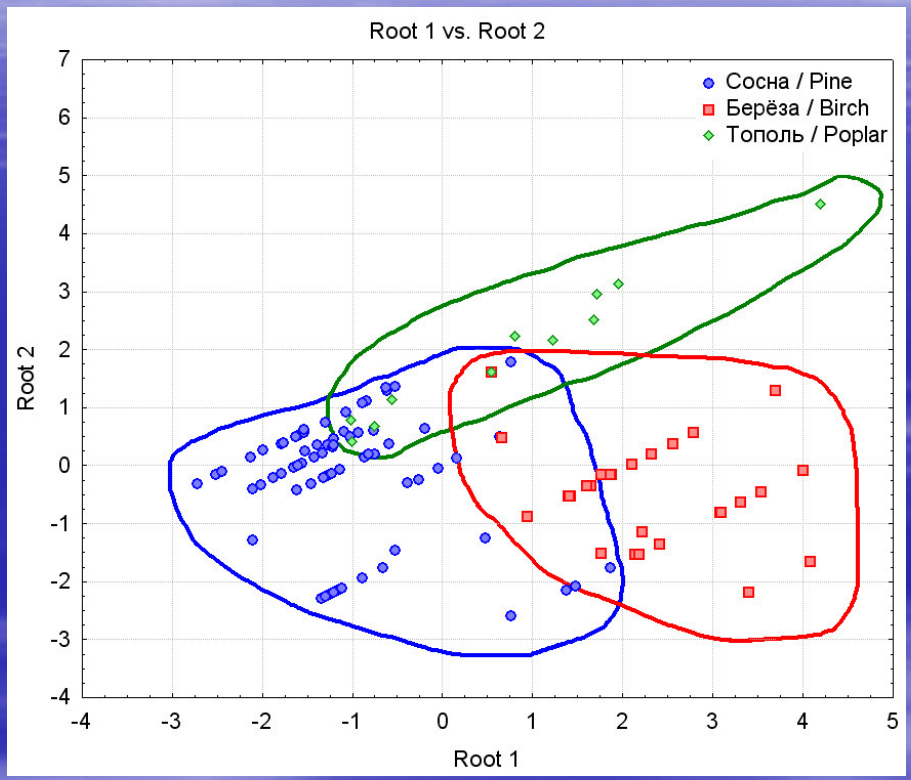


Distribution and occurrence are normal or Gaussian! What if not?

Abundance (Population numbers)

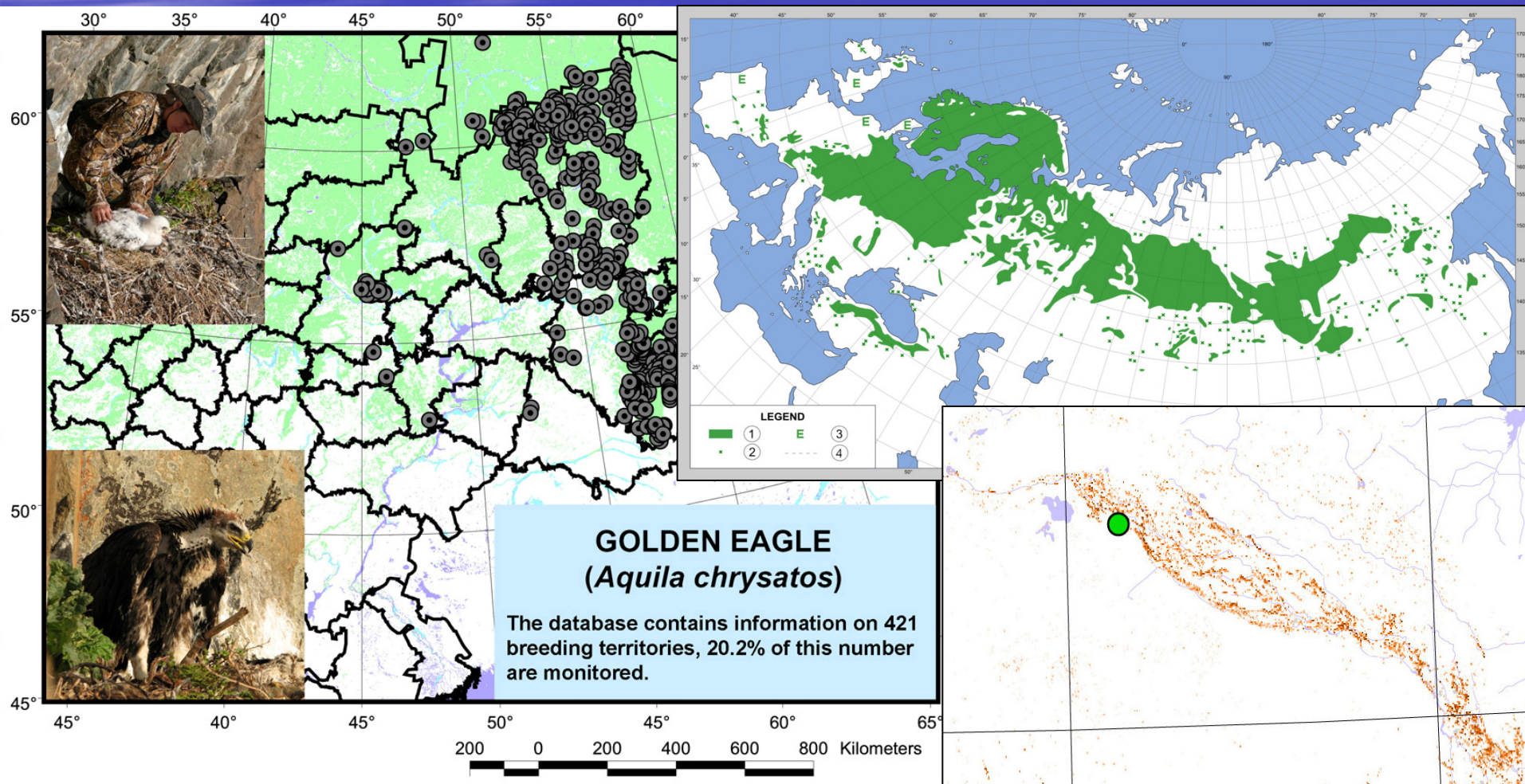


Let's add rivers and the distribution of observations along them will become normal



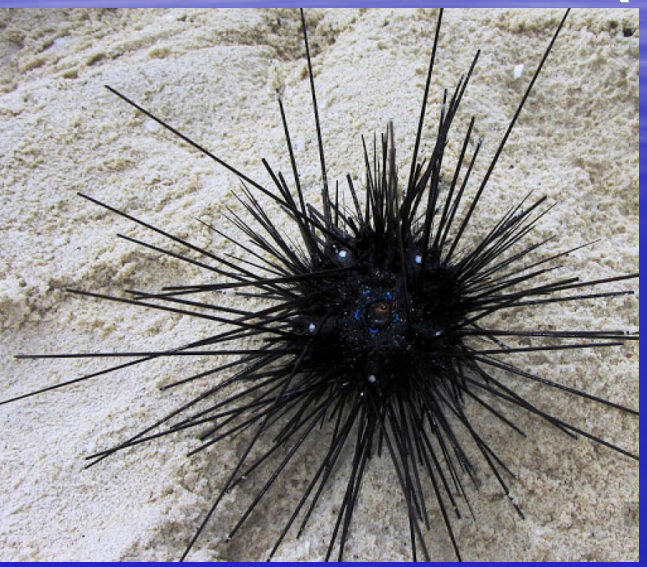
Let us analyze the distribution of our observations regarding two factors - and here is the result!

How to determine the range in which the species occurs for which it is possible to extrapolate the obtained accounting data?



What is species range (distribution) modeling?

Central to understanding species distributions is the niche theory (Hutchinson 1957)

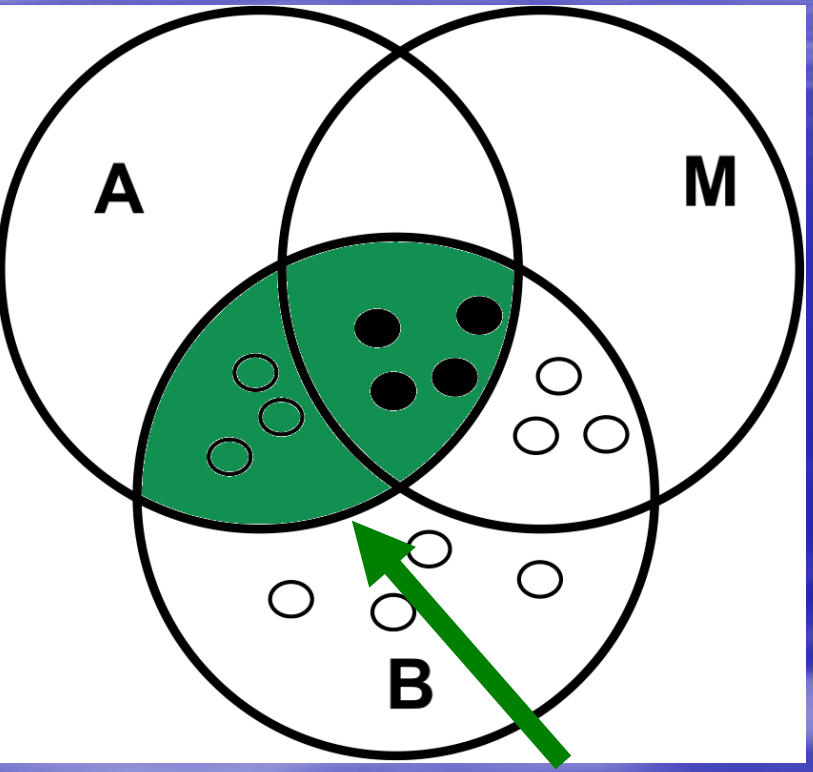


Fundamental niche comprises all abiotic environmental conditions where a species can survive indefinitely, meaning where it has a positive population growth

Realised niche smaller than the fundamental niche due to negative interspecific interactions

Central to understanding species distributions is the niche theory (Hutchinson 1957)

Central to understanding species distributions is the niche theory (Hutchinson 1957)

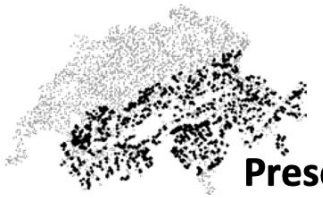


Biotic-abiotic-movement diagrams emphasise the complex interplay between these three factors

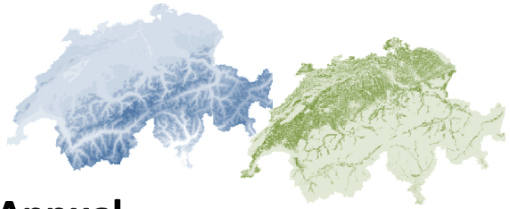
A species can only survive in geographic areas where both the abiotic environmental conditions (A) and the prevailing biotic interactions (B) allow positive population growth. The intersection of A and B thus represents the potential distribution of the species, or the realised niche. The movement capacity (M) of a species will determine which geographic area is accessible now. The intersection of A, B and M represents the geographic area that is actually occupied by the species and where we can find source populations (filled circles).

Schema of the species distribution modelling concept

Biodiversity & environmental data



Presence/ Absence

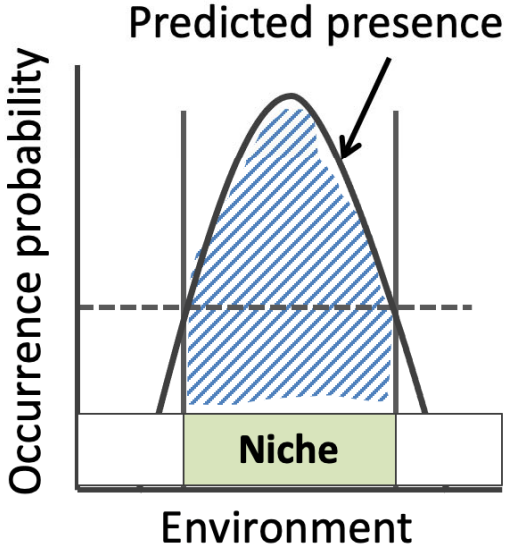


Annual temperature

Grassland cover

SDMs

$$\text{logit}(\pi_i) = \alpha + \sum_{j=1}^{p} \beta_j X_{ij}$$



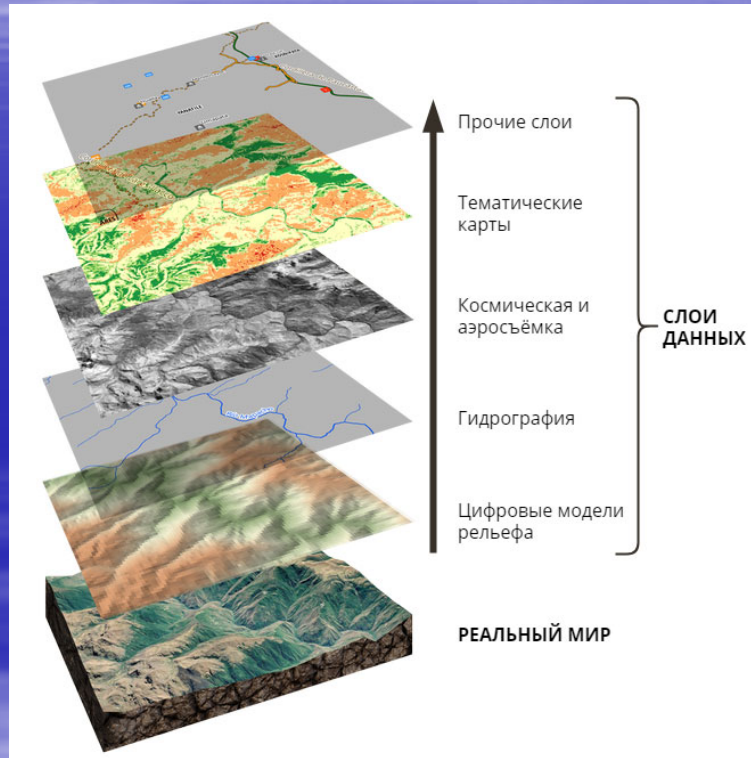
Predicting potential distribution

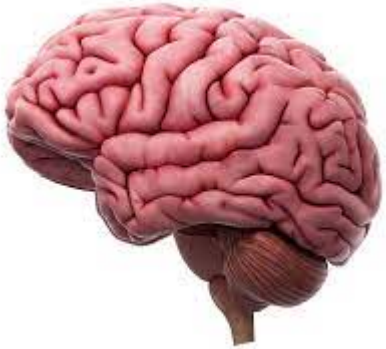
2001

2100

Zurell, 2020

What do we need for SDM?





Back in the 90s. ...

First in ArcView 3x, then in ArcGIS 8~, 9~, 10~

What do we need for SDM?

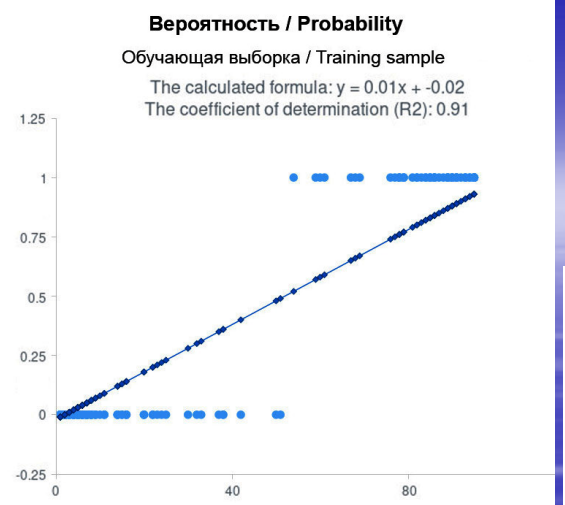
Linear regression methods:

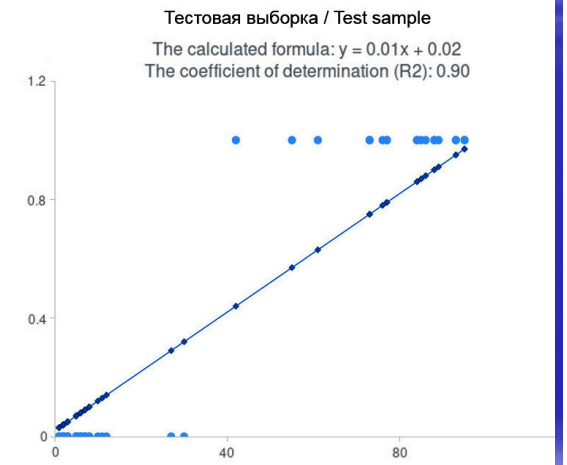
- Generalized linear model (GLM) (Nelder, Wedderburn, 1972),
- Generalized additive model (GAM) (Hastie, Tibshirani, 1990);

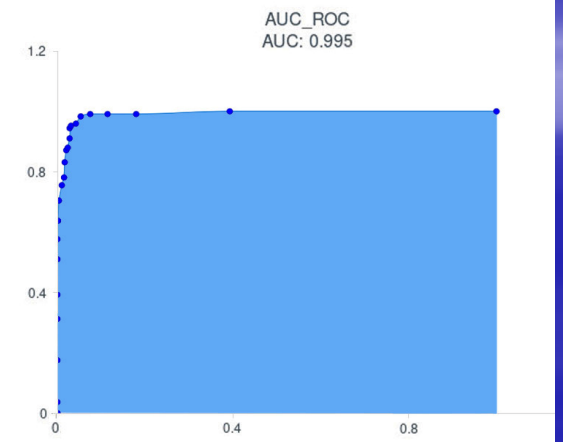
Machine learning methods:

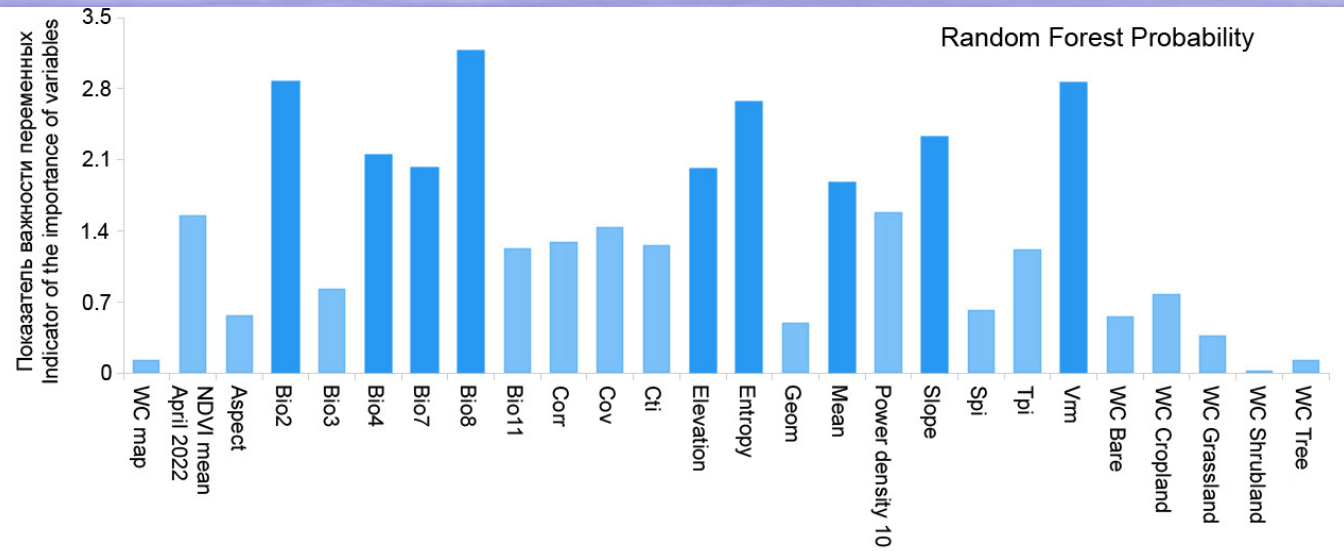
- Maximum entropy method implemented in the MaxEnt program (Soberson, Peterson, 2005; Phillips et al., 2006; Phillips, Dudik, 2008),
- Random Forest (RF) is an ensemble learning method for classification and regression that works by constructing multiple decision trees during training (Breiman, 2001),
- Boosted Regression Trees (BRT),
- Convolutional Neural Networks (CNN) (LeCun *et al.*, 1989),
- Genetic algorithm for Rule Set Production (GARP) (Stockwell, 1999; Stockwell, Peters, 1999),
- Machine learning supporting vector networks (Support Vector Machines, SVM) (Cortes, Vapnik, 1995; Vapnik *et al.*, 1997),
- XGBoost (eXtreme Gradient Boosting, XGB) (Chen, Guestrin, 2016).

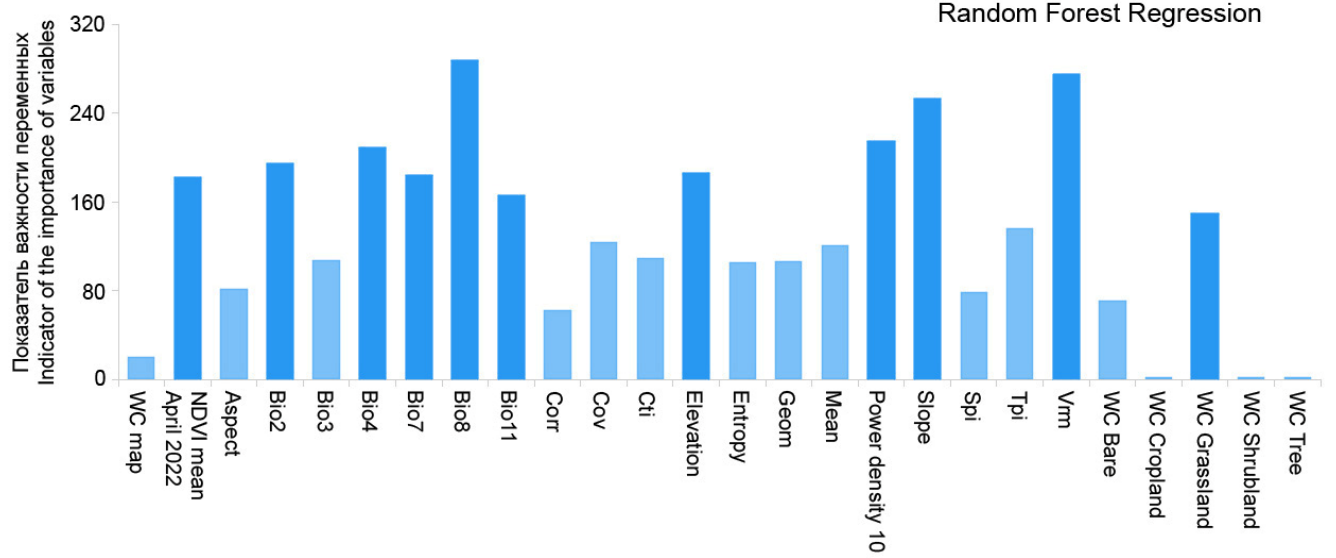
What do we need for SDM?



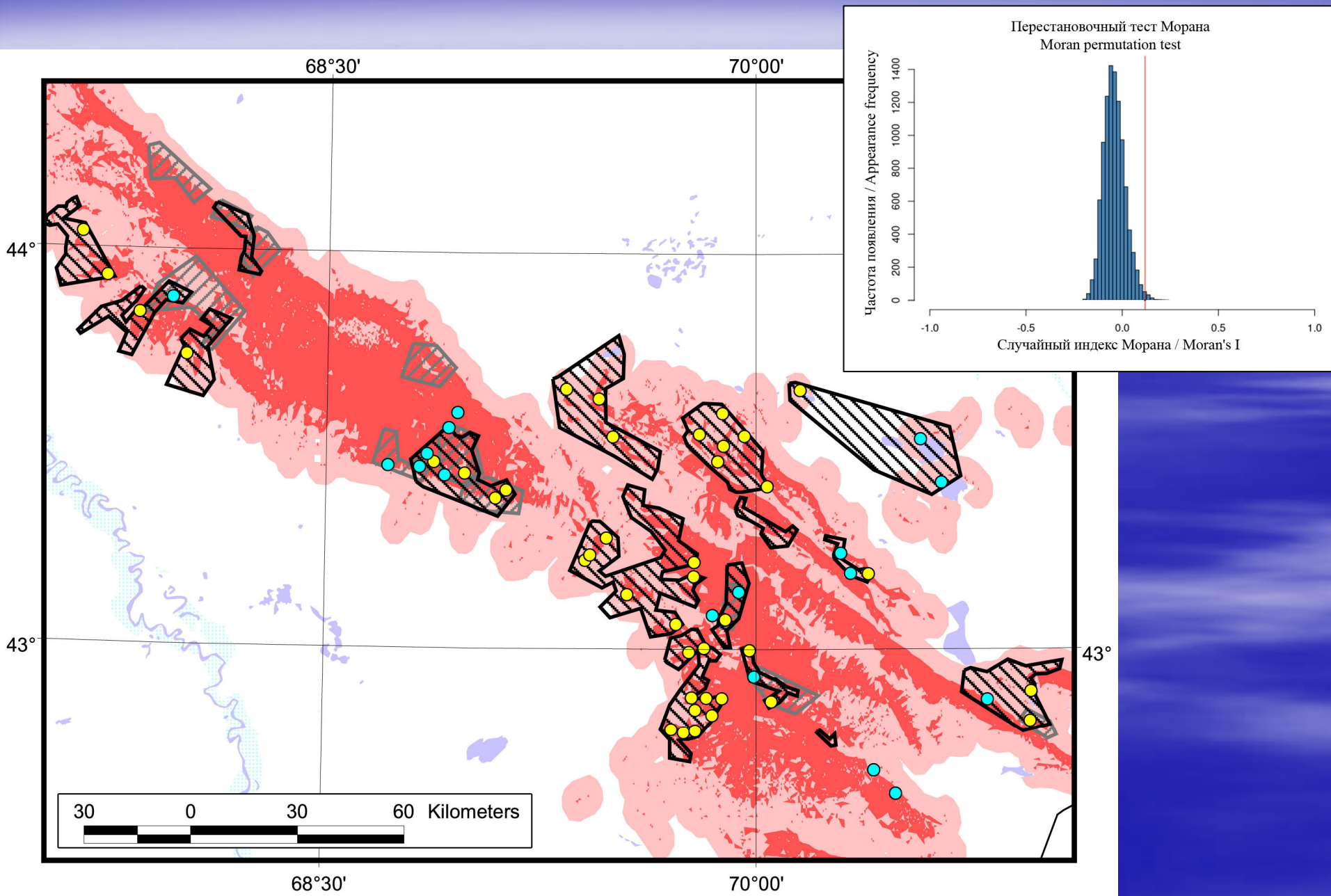






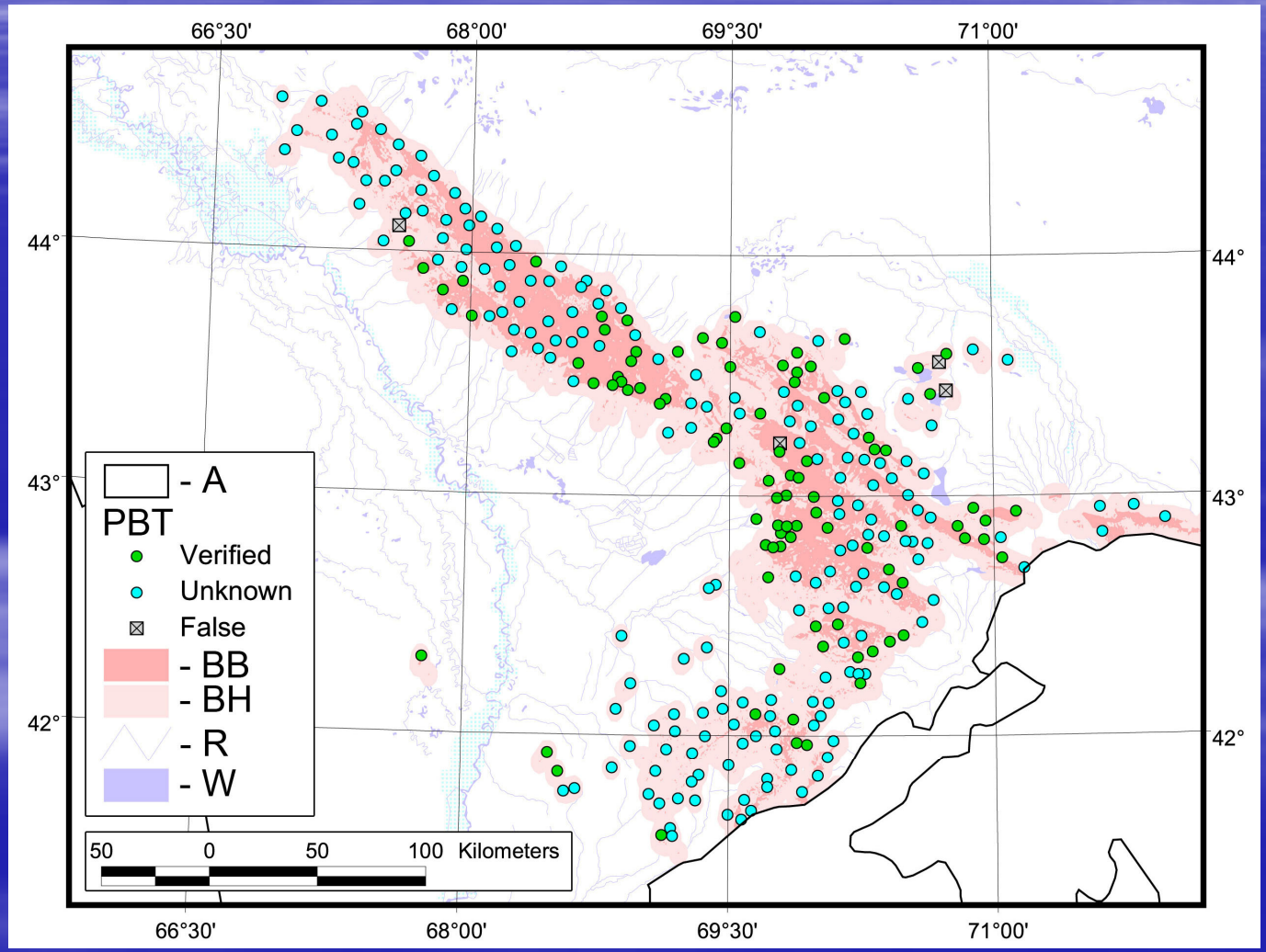


What do we need for SDM?



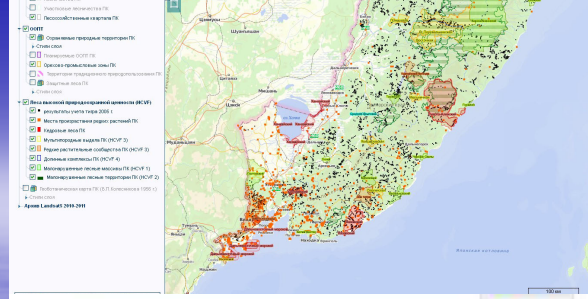
What do we need for SDM?

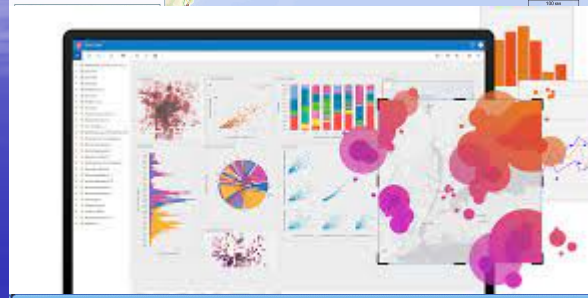
Verification in the field using a network of random points generated according to a species distribution model with a pixel probability higher than 50%



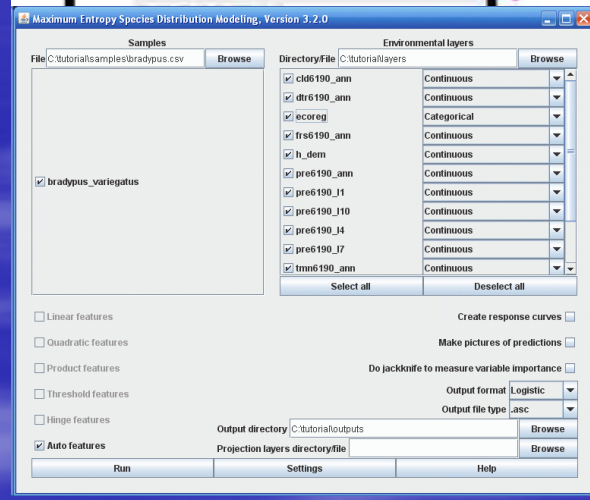
ArcGIS

Confusing interface





English language



Difficulty understanding and learning

http://www.gis.altaiproject.org

ru

ЭКСПОРТ ДАННЫХ | RANDOM FOREST

Добавить точки

Выберите файл | Файл не выбран

csv

latitude_key

longitude_key

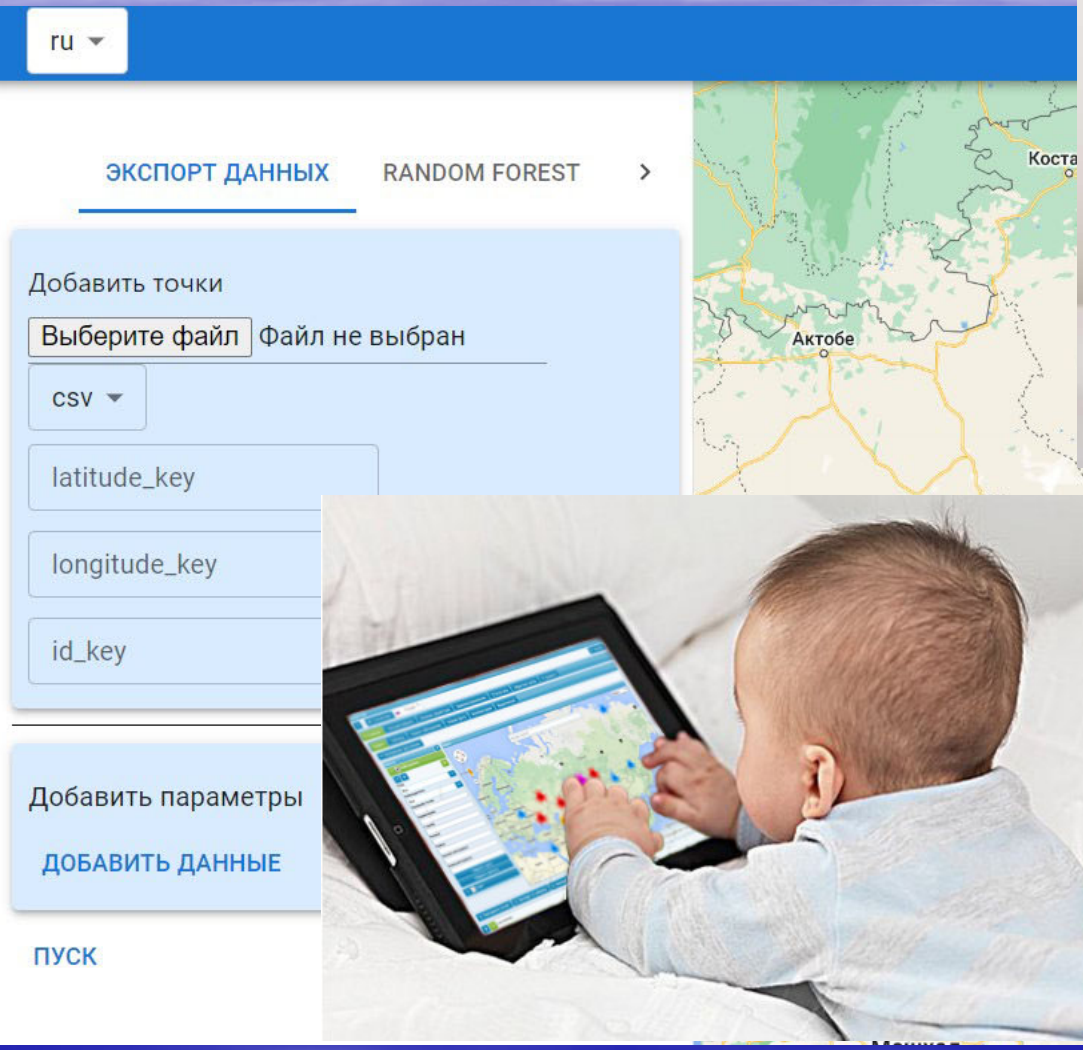
id_key

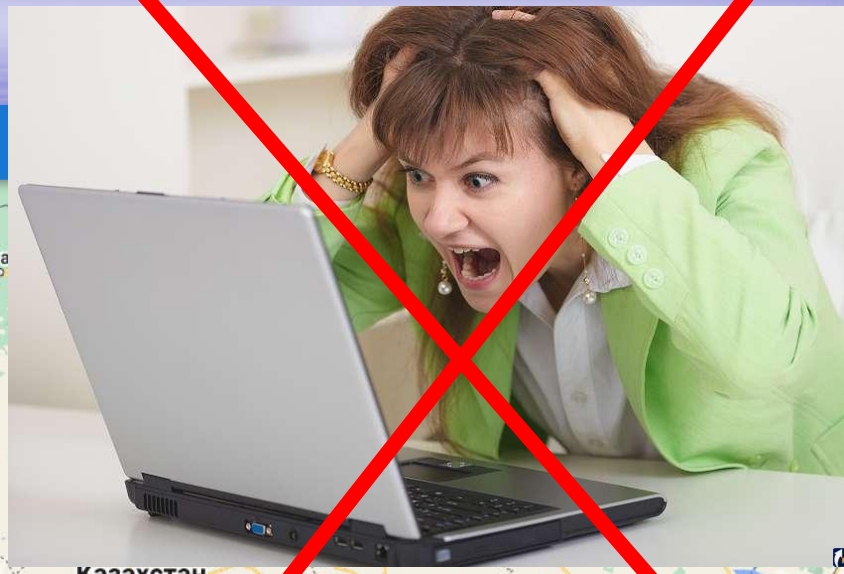
Добавить параметры

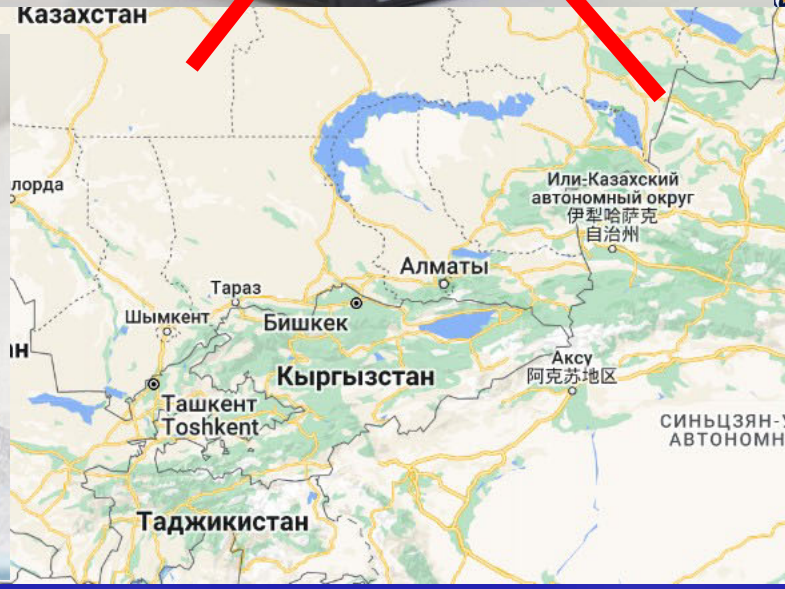
ДОБАВИТЬ ДАННЫЕ

ПУСК

http://www.gis.altaiproject.org



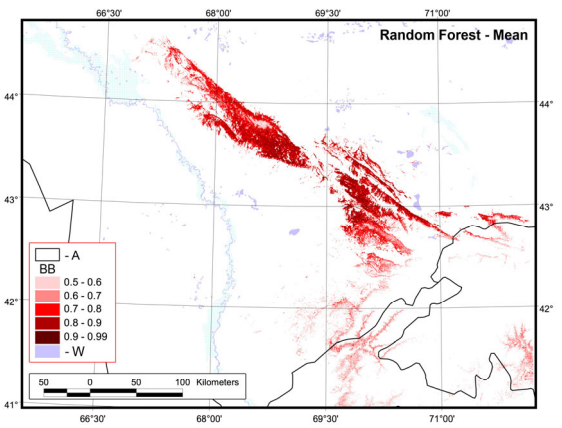


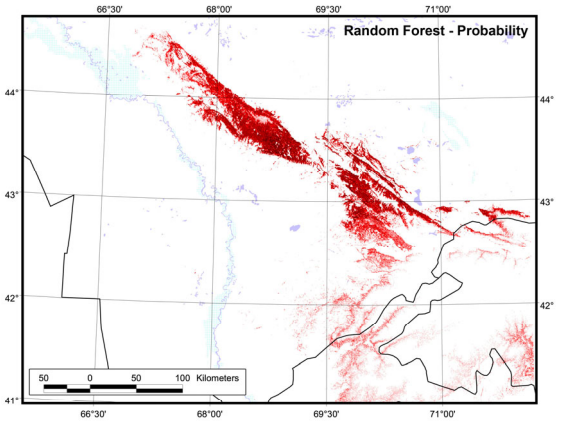


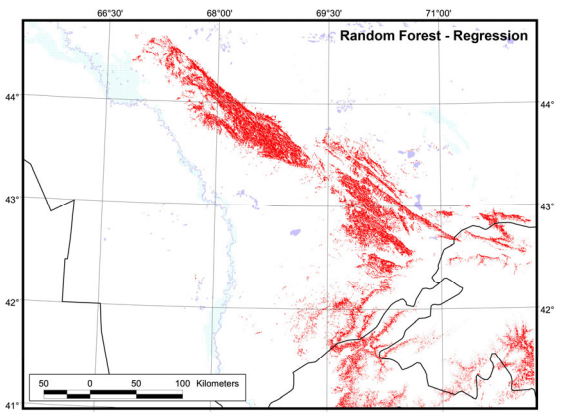
csv, shp, geojson

To run algorithms in which it is necessary to add data from GEE rasters, a selection field is provided from the list of available earth remote sensing (ERS) products: NASADEM, MOD13A1.061 Terra Vegetation Indices 16-Day Global 500m, Geomorpho90m, Global Habitat Heterogeneity, Global Wind Atlas, World Clim, ERA5-Land Monthly Aggregated – ECMWF Climate Reanalysis, ESA WorldCover 10m v100, Dynamic World V1, unclassified satellite data such as surface reflectivity (SR) collection 2 Landsat 8 atmospheric-corrected (blue, red, green, near-infrared and shortwave infrared 1 bands with 30 m spatial resolution) and ALOS-2 PALSAR L-band dual-polarization (HH and HV) SAR data, and NDVI and EVI calculation data from Landsat 8 images using the GEE (normalizedDifference) function.

http://www.gis.altaiproject.org





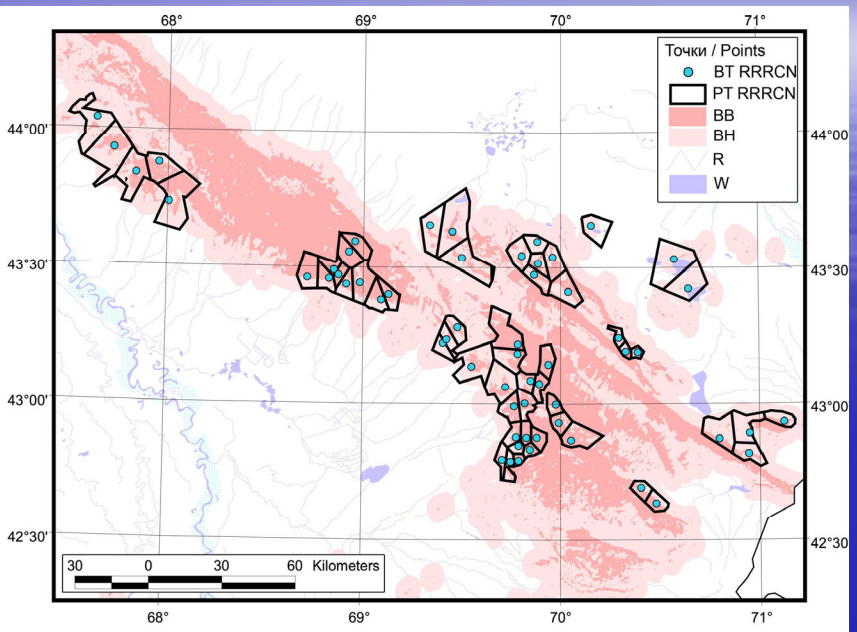


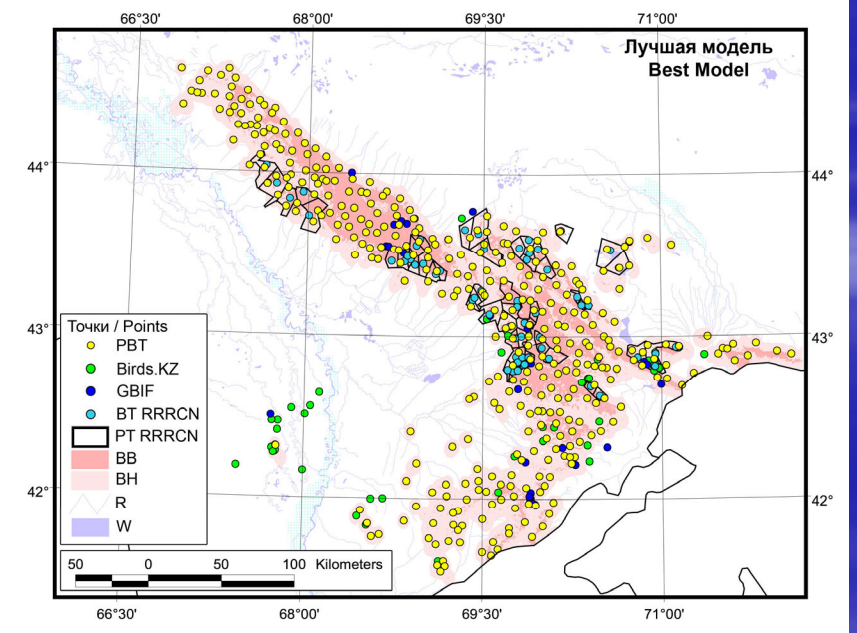
At the current stage, the product includes the following modules:

1) Obtaining data from GEE rasters for given points (result presented in a table with data selected for points from rasters included in the GEE collection);

2) Obtaining a classified raster for a given area and a set of points of presence and absence of a view (training points) using the RF and MaxEnt classifiers based on GEE (both classifiers allow, for a given area of interest, a set of training points and selected remote sensing products from GEE, to obtain a classified one with using appropriate GEE raster methods of the area of interest. It is possible to cross-validate the selected models and evaluate their predictive effectiveness);

http://www.gis.altaiproject.org

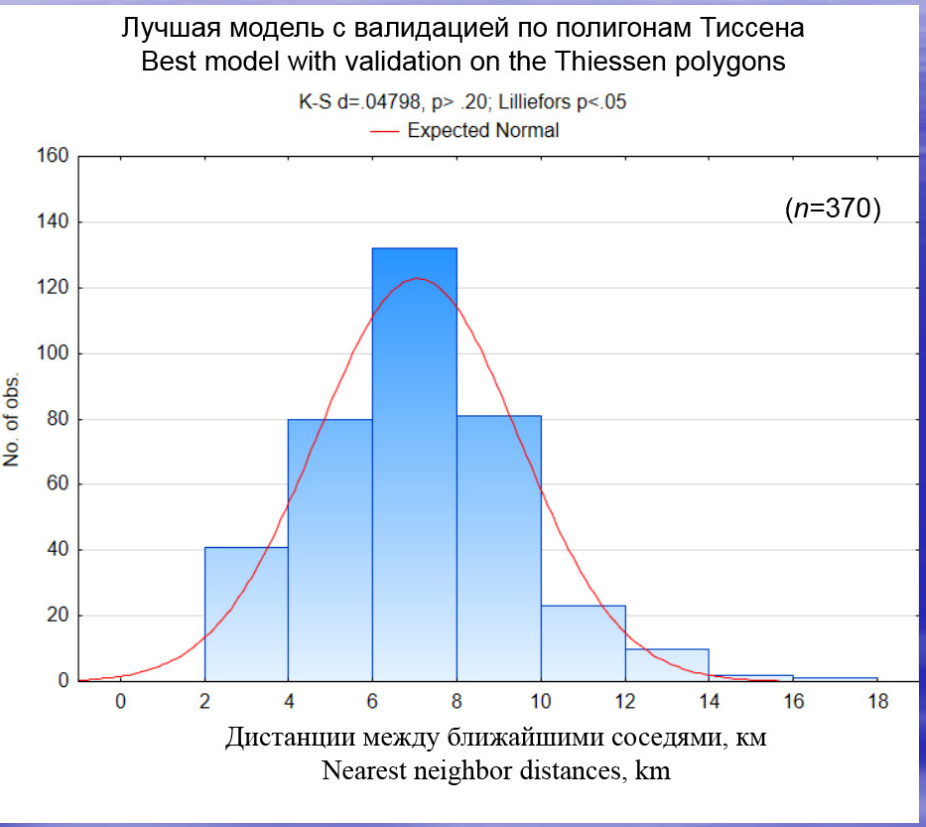




3) Three different methods to stimulate population size:

3.1) Generation of random points in a regular network – a heuristic algorithm that, based on data on the points of presence of the species and on the studied areas, generates random points, simulating species' distribution in the general area of interest;

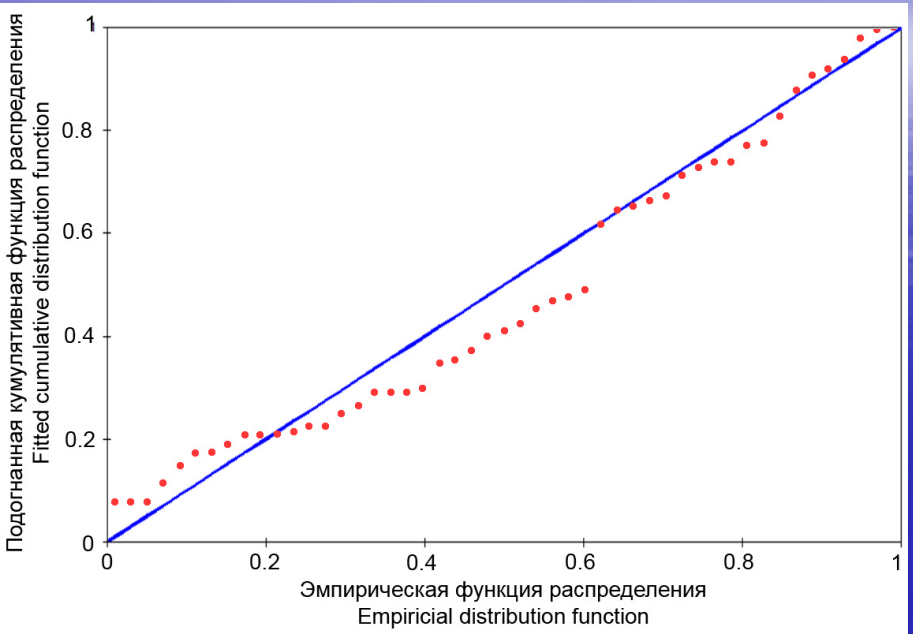
http://www.gis.altaiproject.org

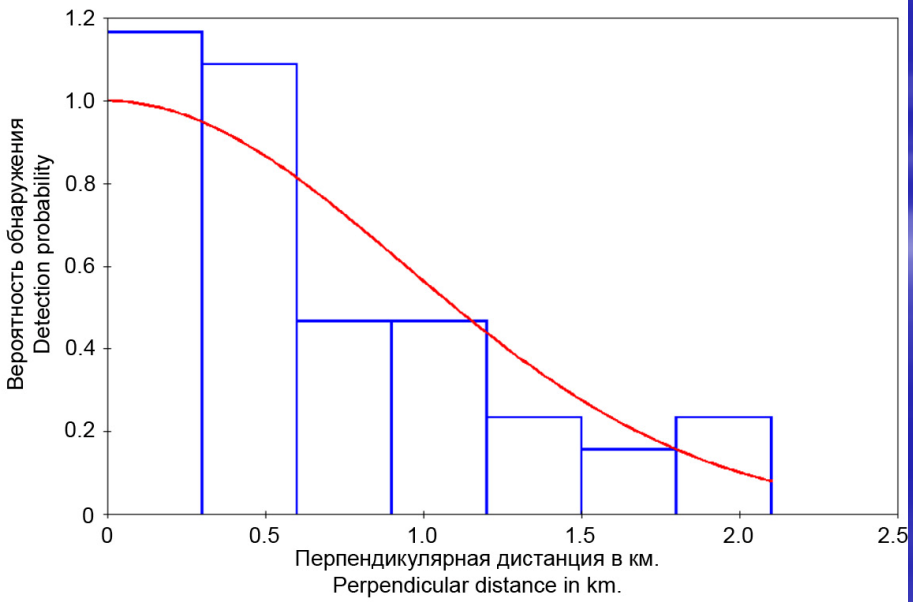


3) Three different methods to stimulate population size:

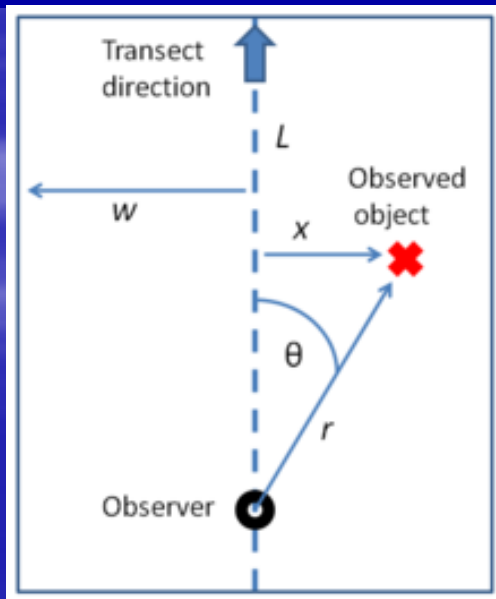
3.1) Generation of random points in a regular network – a heuristic algorithm that, based on data on the points of presence of the species and on the studied areas, generates random points, simulating species' distribution in the general area of interest;

http://www.gis.altaiproject.org

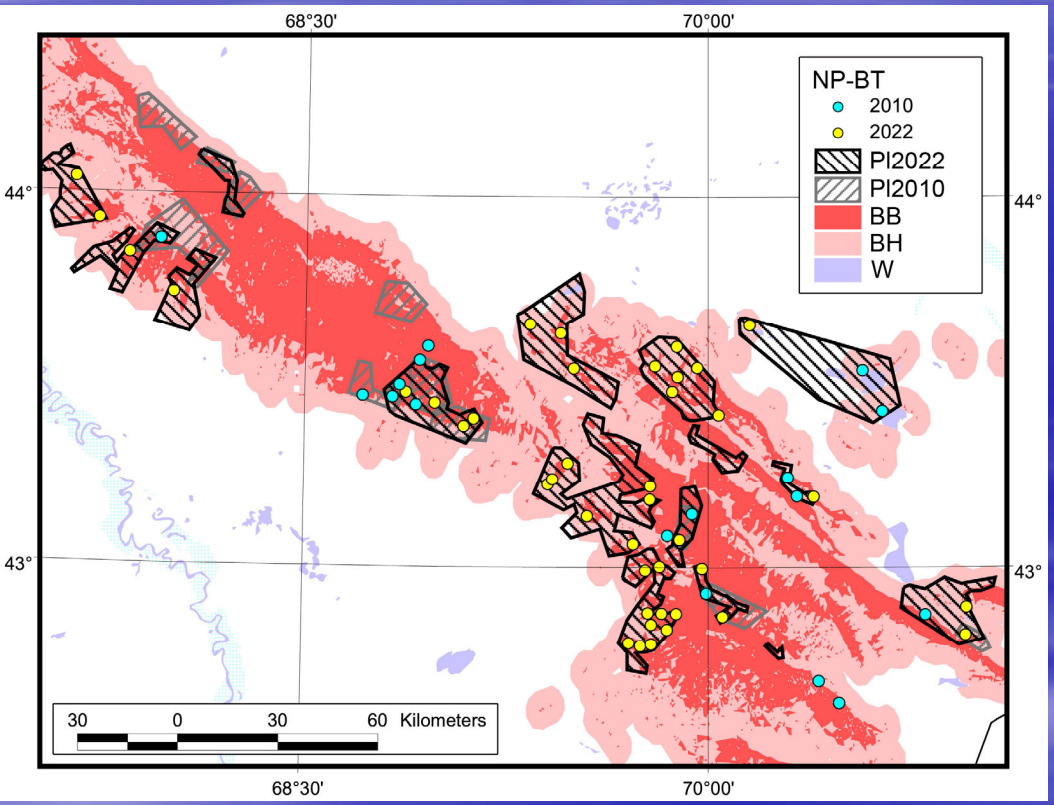




3.2) Distance – a method based on the Distance Sampling model (Buckland *et al.*, 2015), that accepts input of a file with the necessary variables for points and areas and displays detailed statistics as a result;

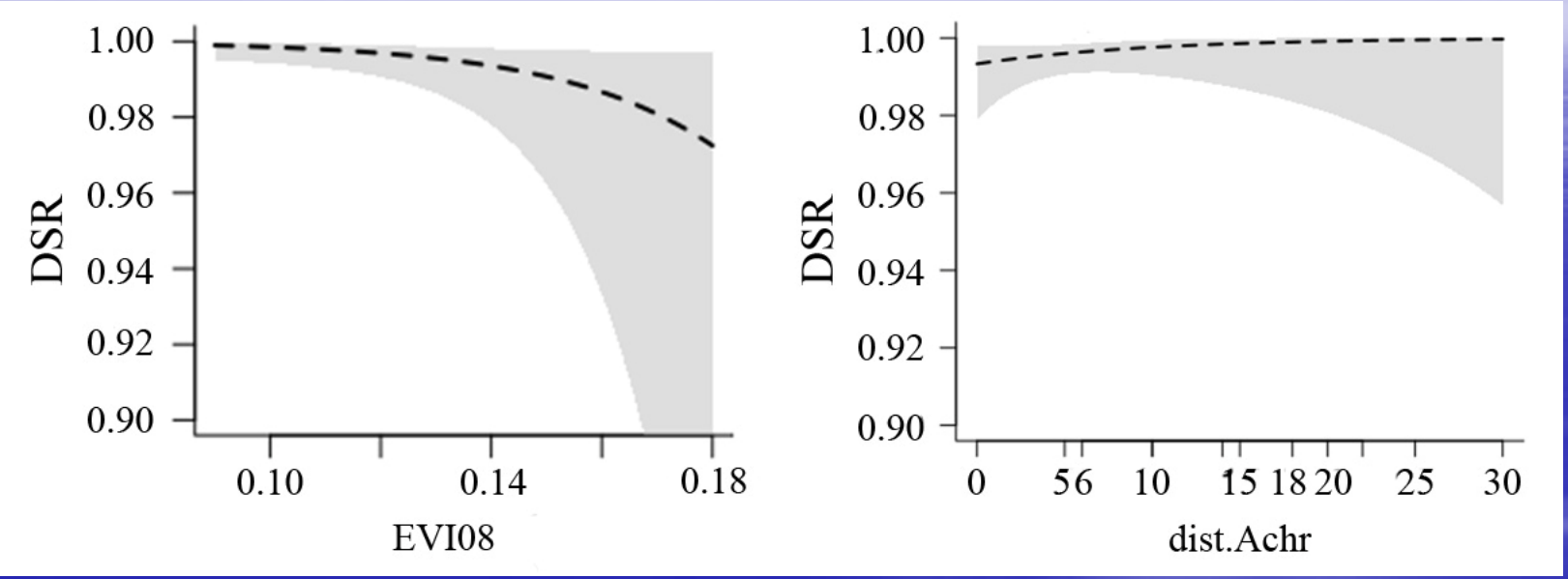


http://www.gis.altaiproject.org



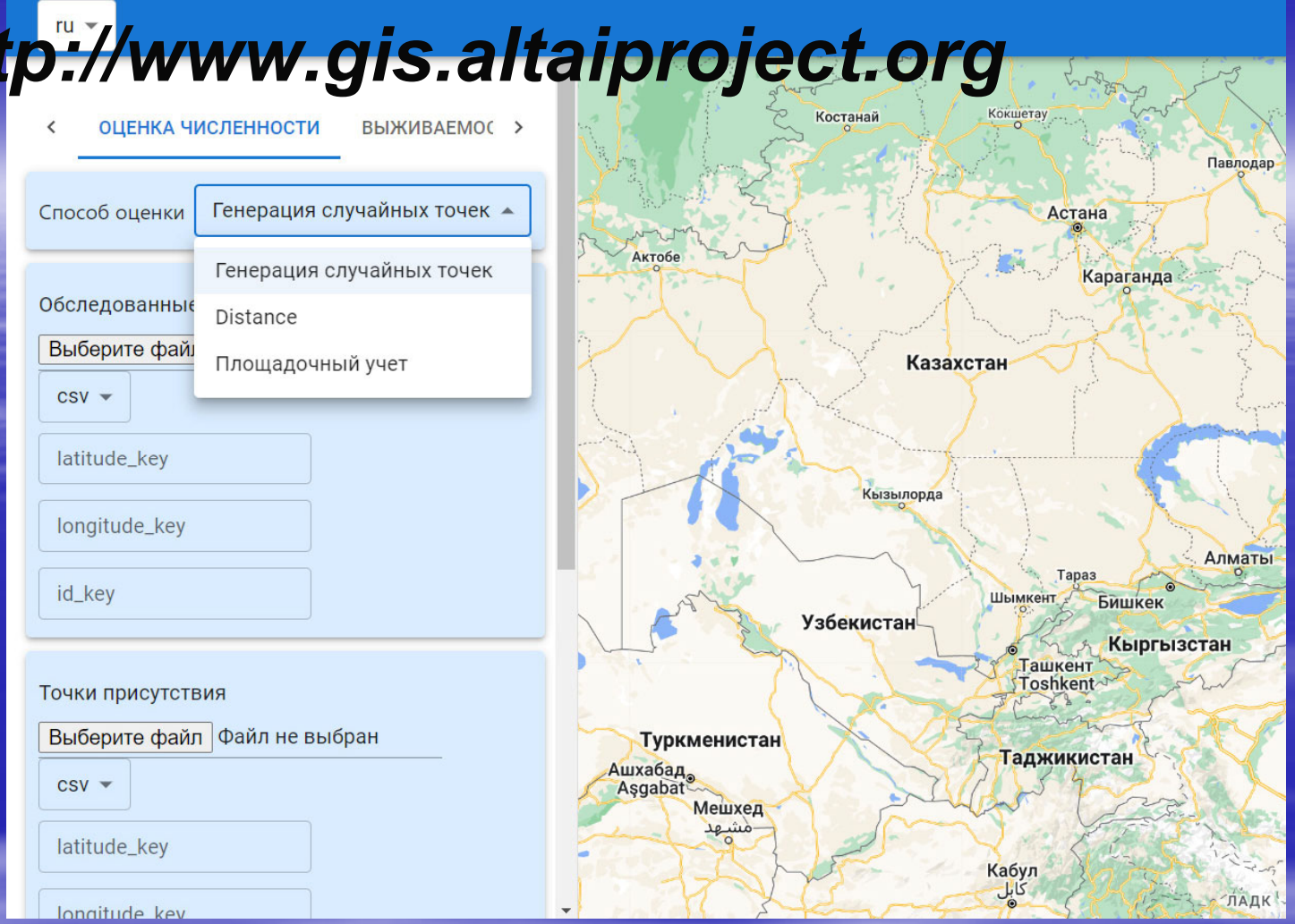
3.3) Simple site surveys using calculation of a weighted average indicator for species distribution density (Karyakin, 2004) with an calculation of asymmetric confidence interval (Ravkin, Chelintsev, 1990);

http://www.gis.altaiproject.org



4) Estimation of nest survival based on the RMARK library (Laake, 2013). The survival calculation module includes processing of nest survival data using the nest method of the RMARK library, which can account for various variables in remote sensing data and infers the importance of variables for nest survival.

http://www.gis.altaiproject.org



The software product is hosted on the servers of organizations recognized as undesirable in Russia, access to which is blocked by Roskomnadzor. The authors are considering options, including creating a clone on a Russian internet resource.

RRRCN BRCC

THE ALTAI PROJECT

CRITICAL ECOSYSTEM PARTNERSHIP FUND

WWF

Спасибо за внимание!

Thank you for your attention!

This work is carried out with financial support from the Critical Ecosystem Partnership Fund (CEPF) within the framework of the project "Endangered Raptors Conservation on the Indo-Palearctic Flyway". Palearctic Migration Flyway").