

Моделирование ареалов видов в среде статистического программирования R



Шашков М.П.
ПНЦБИ,
лаборатория
Моделирование экосистем

Екатеринбург - 2021

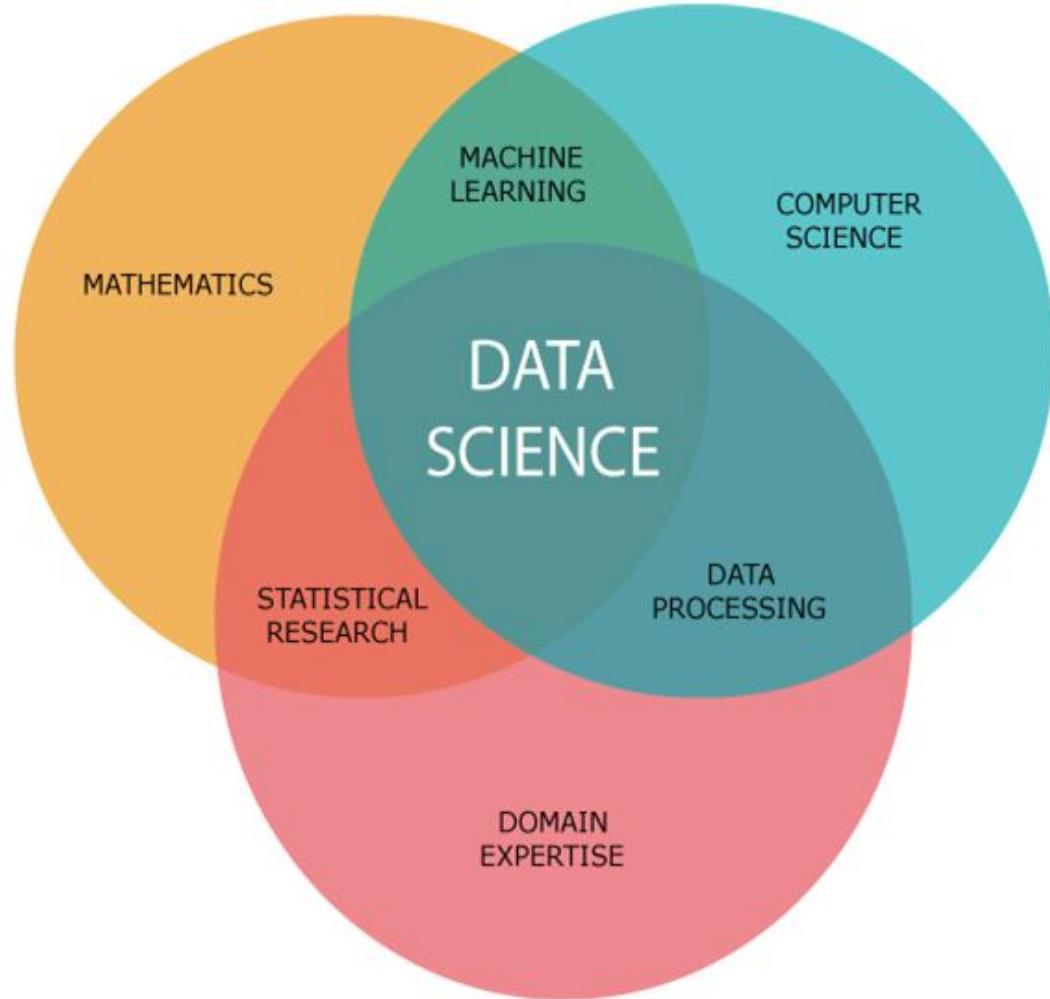
Data Science

Наука о данных (Data Science)

сплав математической статистики, информатики и предметной области.

Междисциплинарная область, направленная на извлечение знаний из наборов данных, обычно большого объема (Big Data)

Machine Learning



Species Distribution Modeling - SDM

Ecological niche modelling

Species niches and distribution

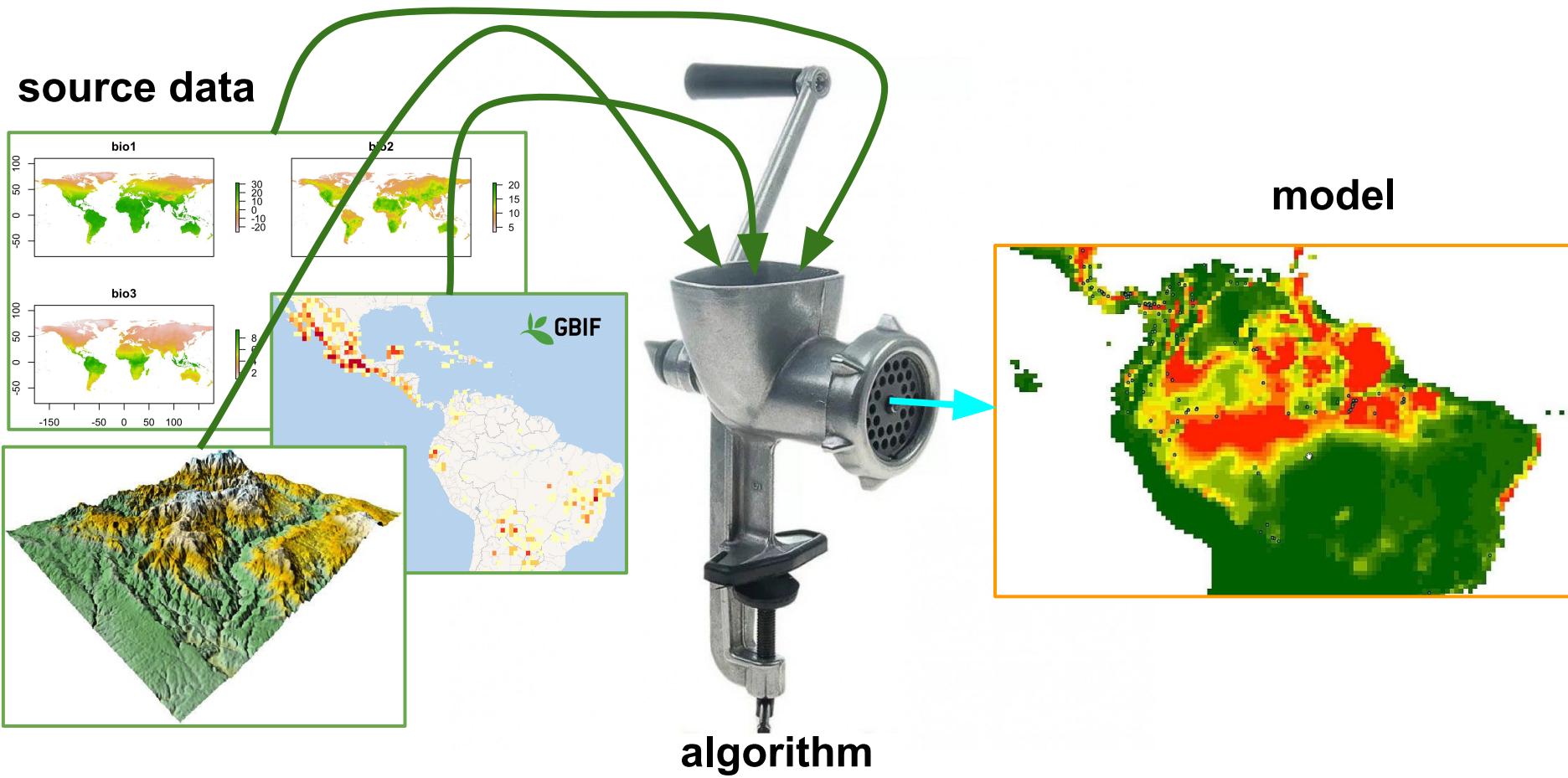
Niche modelling

Bioclimatic envelope models

Habitat suitability models

Nix, H.A. (1986) A biogeographic analysis of Australian elapid snakes. *Atlas of elapid snakes of Australia: Australian flora and fauna series 7* (ed. by R. Longmore), pp. 4–15. Bureau of Flora and Fauna, Canberra

Species Distribution Modelling (SDM)



Источники данных о биоразнообразии

Get data How-to Tools Community About

GBIF | Global Biodiversity Information Facility

Free and open access to biodiversity data

Occurrences Species Datasets Publishers Resources

Search

WHAT IS GBIF? ABOUT GBIF RUSSIAN FEDERATION

Occurrence records 1 887 185 643 Datasets 61 514 Publishing institutions 1 715 Peer-reviewed papers using data 6 087

News News News News

Call for author proposals for new and updated GBIF technical guides

BID programme funds 18 new projects in sub-Saharan Africa

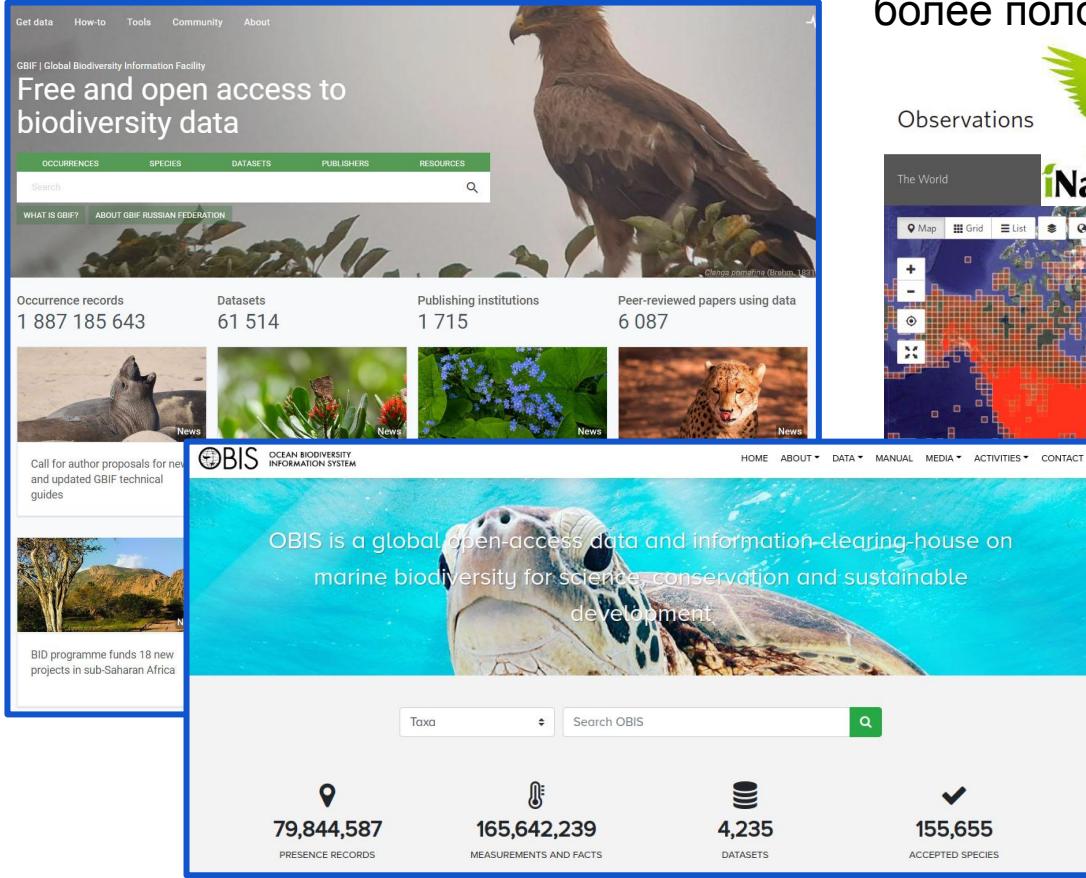
OBIS OCEAN BIODIVERSITY INFORMATION SYSTEM

HOME ABOUT DATA MANUAL MEDIA ACTIVITIES CONTACT

OBIS is a global open-access data and information-clearing-house on marine biodiversity for science, conservation and sustainable development.

Taxa Search OBIS

79,844,587 PRESENCE RECORDS 165,642,239 MEASUREMENTS AND FACTS 4,235 DATASETS ✓ 155,655 ACCEPTED SPECIES



более половины записей в GBIF - Citizen Science

Observations

iNaturalist 36 360 360

The World

Map Grid List Places of Interest

Search in map

Trametes ochracea (Green Marsh Hawk) 33720 Herma, Suom... Today Research Guide 1 21m

Orthetrum sabina (Green Marsh Hawk) Salem district • Sep 11, 2021 Research Guide 1 21m

Veronica persica (Bitter-cress Speedwell) Cesere-Zapovednyi ad... • Sep 11, 2021 Research Guide 1 22m

Papilio polytes (Common Mormon Swallowtail) Salem district • Sep 11, 2021 Research Guide 1 22m



Data Quality

Соответствие стандартам (дата YYYY-MM-DD,
координаты в десятичных градусах, ...)

Целостность (соответствие внутренней логике)

Непротиворечивость

Полнота

....

Соответствие целям и задачам исследования

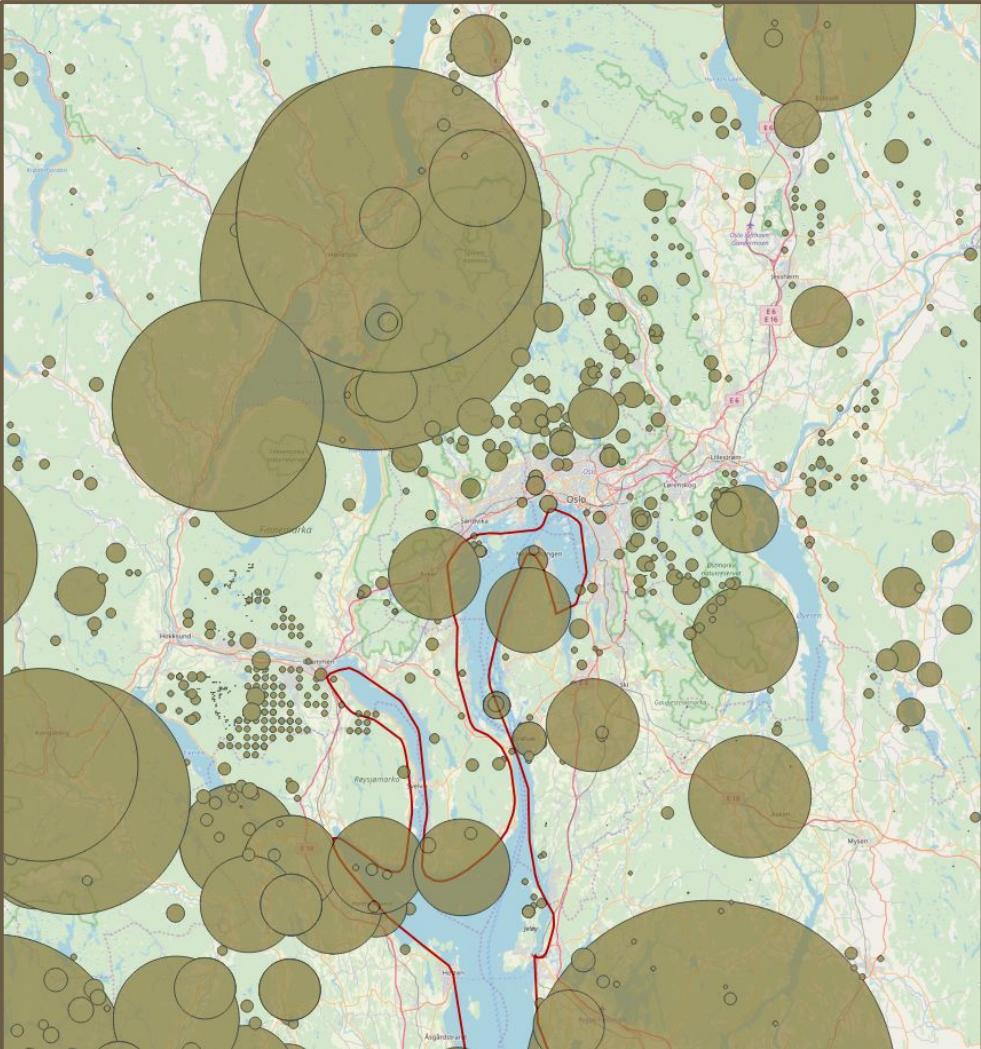
Data Quality

географическая привязка

должна быть указана система координат
(WGS 84, например)

должна сопровождаться указанием
точности (accuracy) / неточности
(uncertainty) и данный показатель
должен соответствовать масштабам
исследования

На схеме представлено
распространение вида *Linnaea borealis*
на юге Норвегии в районе Осло (по
данным GBIF)



Данные среды

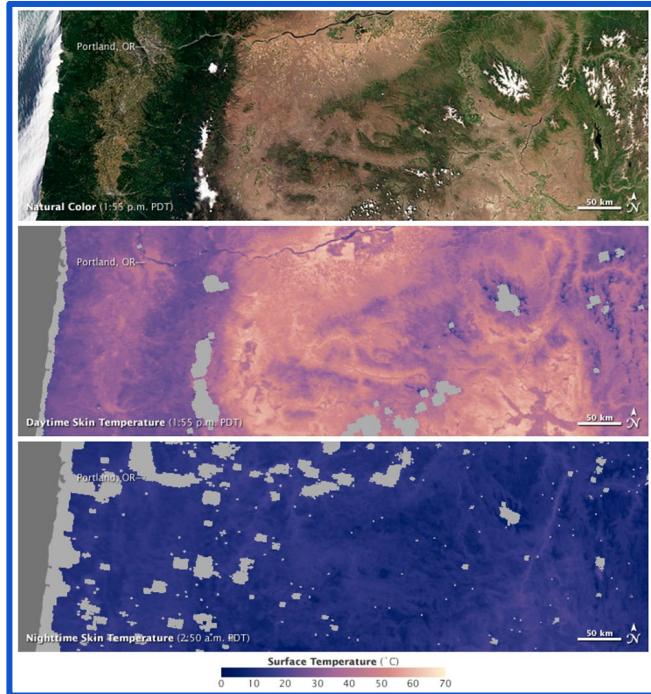
WorldClim (www.worldclim.org)

биоклиматические данные: bio1...19, altitude

EarthEnv (www.earthenv.org)

Global 1-km Consensus Land Cover

Классификация земной поверхности (12 классов)



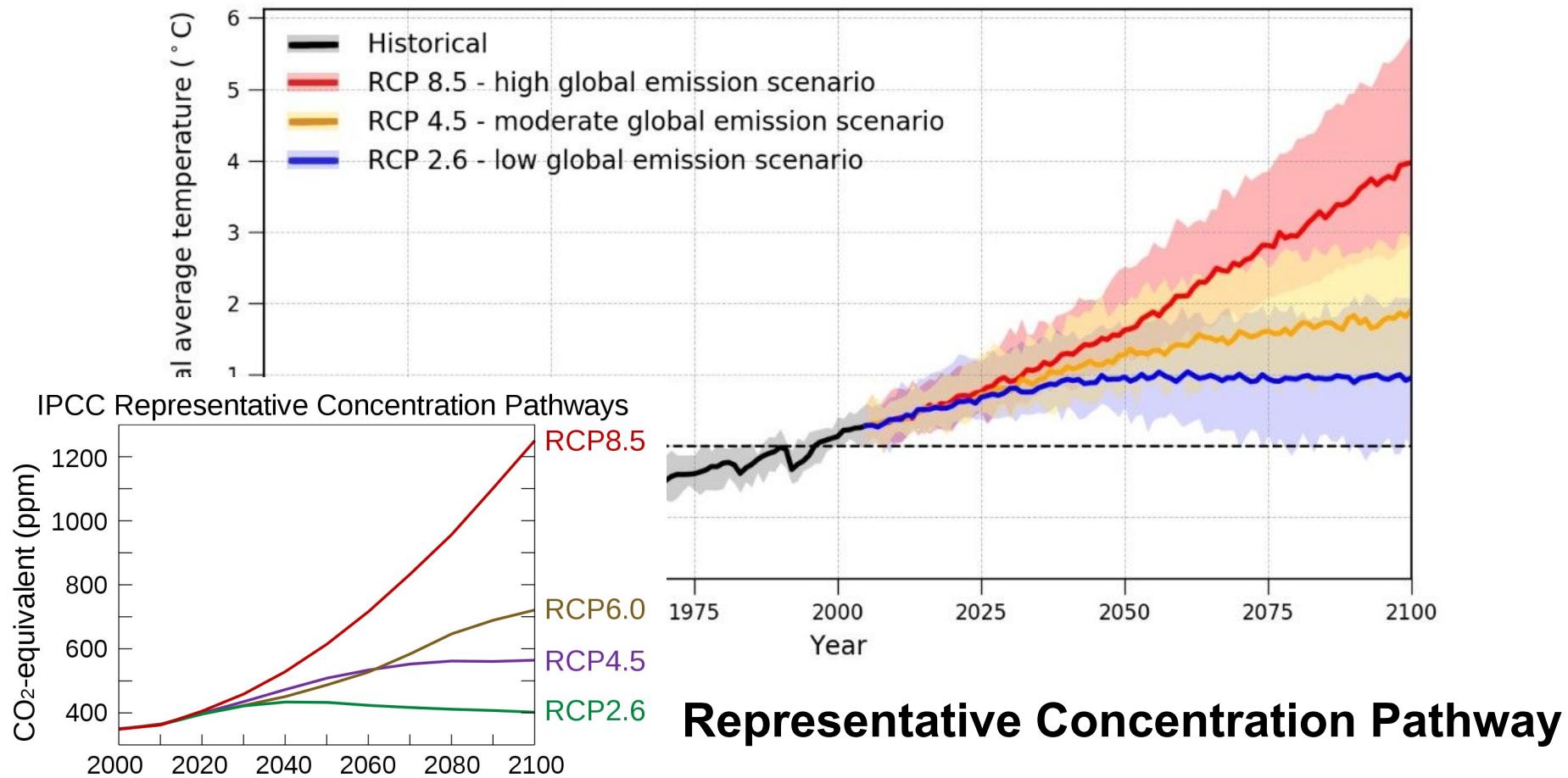
Данные NASA для SDM - результат обработки спутниковых данных

(earthdata.nasa.gov/learn/pathfinders/biodiversity/species-distribution)

температура земной поверхности, влажность почвы, снежный покров

температура и соленость морской поверхности

Сценарии изменения климата



Классы моделей SDM

Минимум 17 основных методов моделирования
(**по данным bccvl**), объединенных в 4 класса:

Географические

Профильные

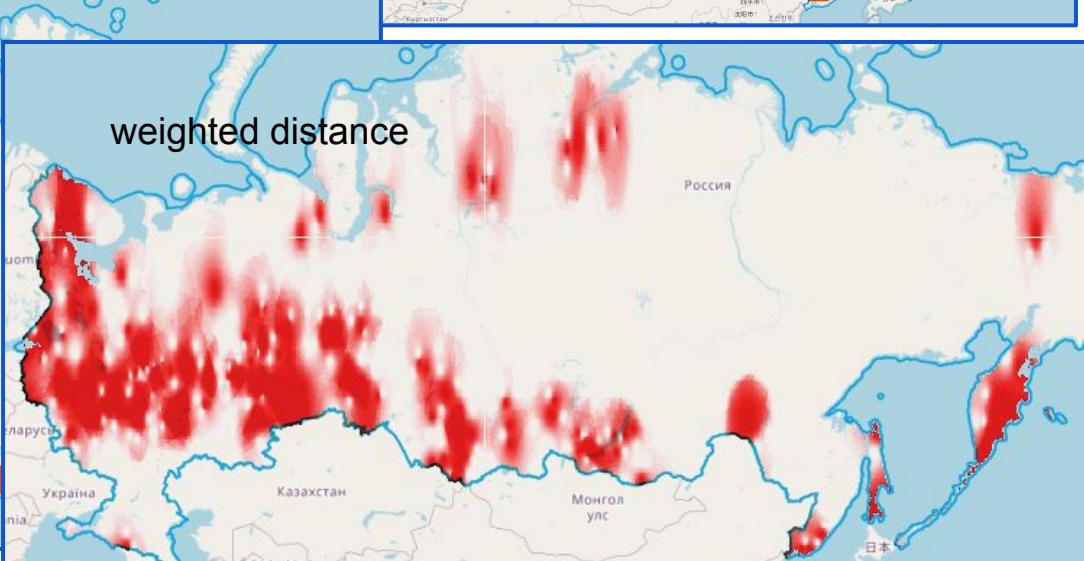
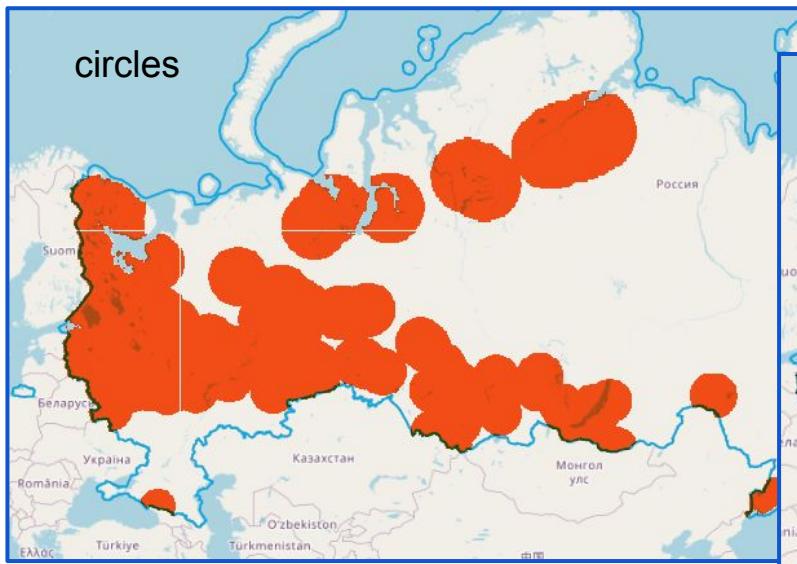
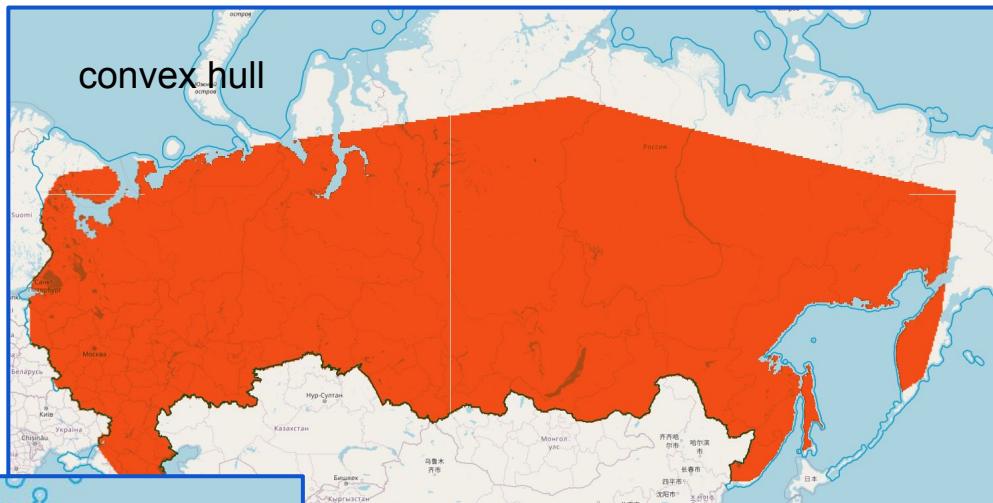
Статистические (регрессионные)

Основанные на машинном обучении (Machine Learning)

Географические

Используют точки находок

Данные среды не используют

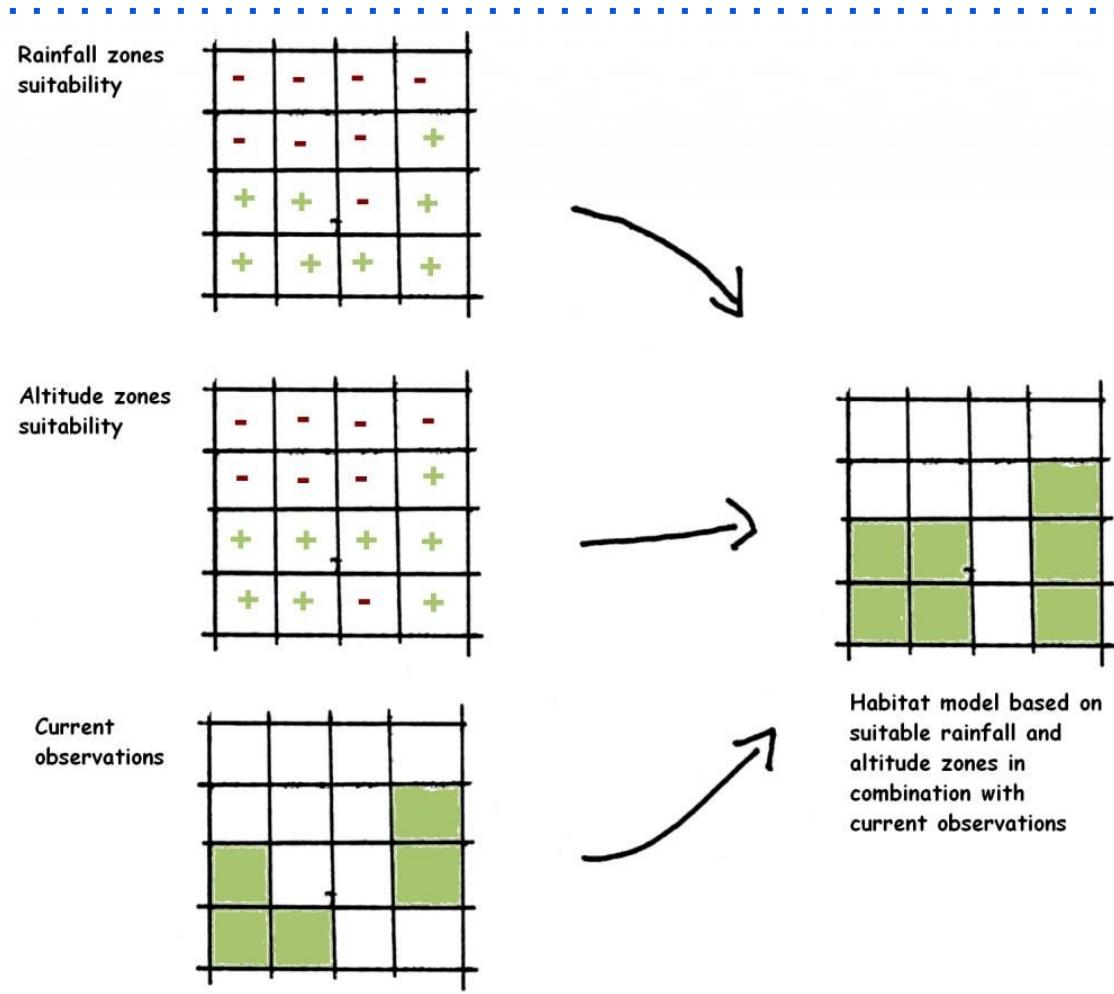


Профильные

с чего всё началось:
многомерное пространство
факторов среды в диапазоне
пригодном для обитания

BIOCLIM (1986)

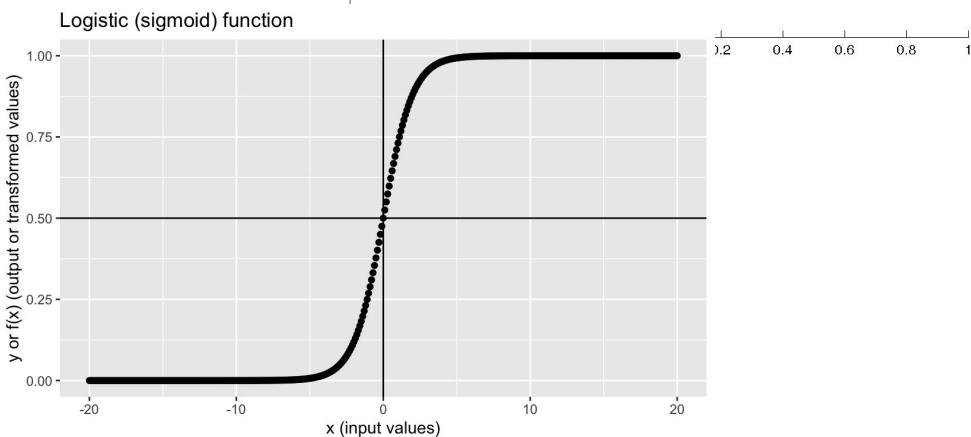
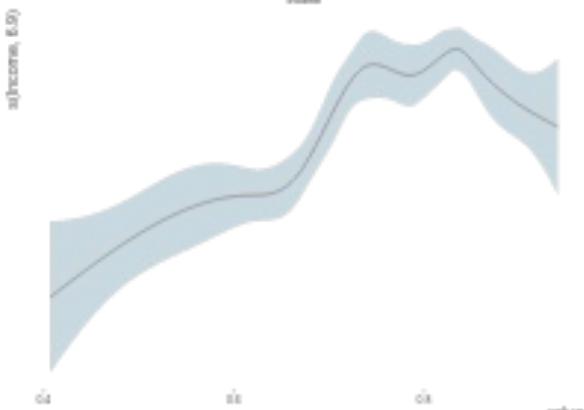
Используются точки находок
и переменные условий среды
не подходят для работы
с изменением климата



Статистические

Используют данные среды

Используют точки находок
и отсутствия целевого вида



Машинное обучение

данные среды

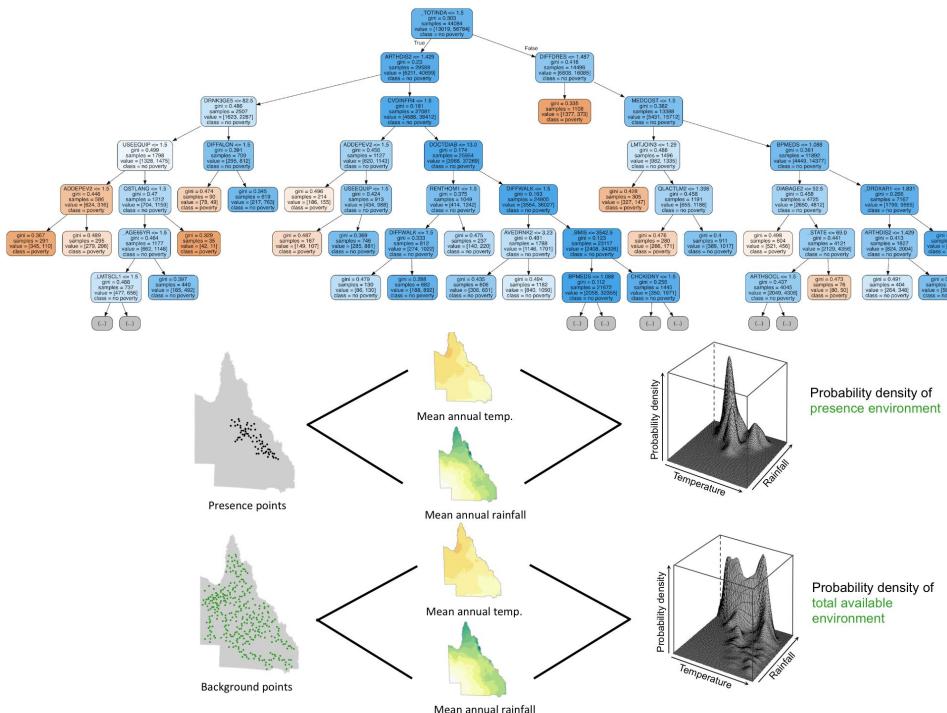
точки присутствия и отсутствия

кросс-валидация

Random Forest

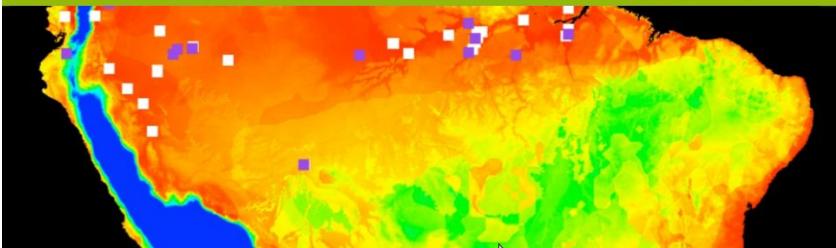
MaxEnt - классика жанра

Boosted Regression Tree



Adapted from Elith et al. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17, 43-57.

Maxent software for modeling species niches and distributions



Этапы моделирования

1. Формируется набор данных находок (и отсутствия вида)
2. Собираются значения-предикторы параметров окружающей среды
3. Входные данные комбинируются, исключаются скоррелированные факторы
4. Строится (обучается) модель - Model Fitting
5. Оценка качества полученной модели (техническая)
6. Выяснение значимых переменных (среды)
7. Оценка качества полученных результатов, биологическая, согласно поставленной гипотезе, выявление артефактов.
8. Модель используется для предсказания
9. В зависимости от полученных результатов возможен возврат на один из предыдущих шагов

Глобальное распределение разнообразия дождевых червей

6928 точек сбора в 57 странах

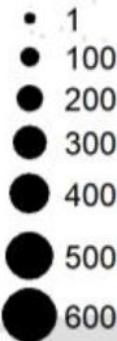
Three **generalized linear mixed-effects models**

A

1. Видовое богатство
2. Численность
3. Биомасса

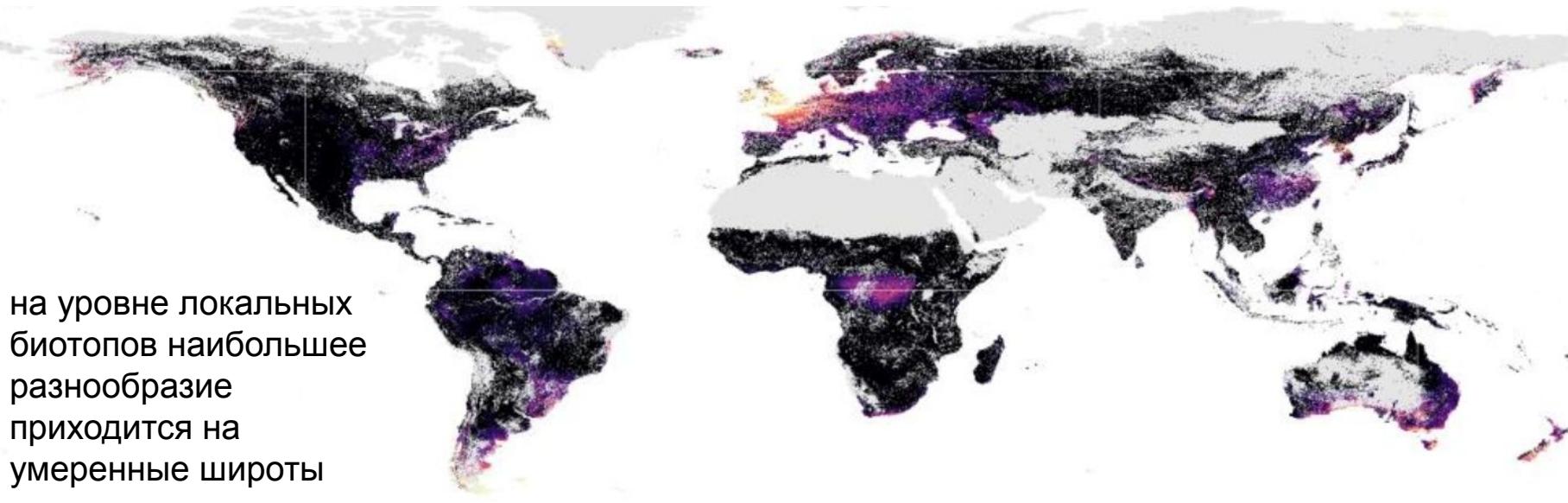
DOI: [10.1126 /science aax4851](https://doi.org/10.1126/science.aax4851)

Number Of Sites



Предикторы:
Habitat cover
Elevation
Soil
Precipitation
Temperature
Water retention

Распределение видового разнообразия

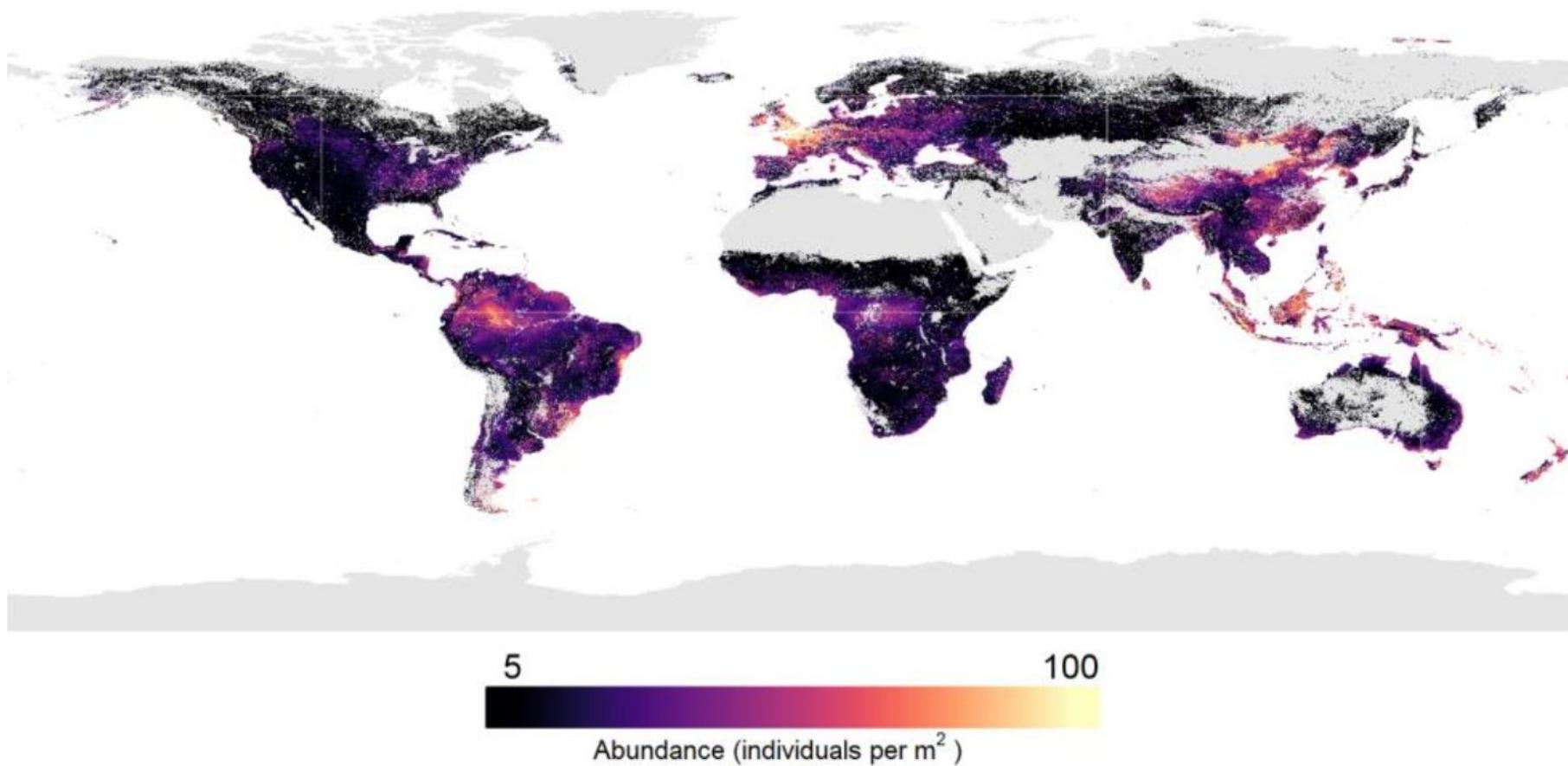


1

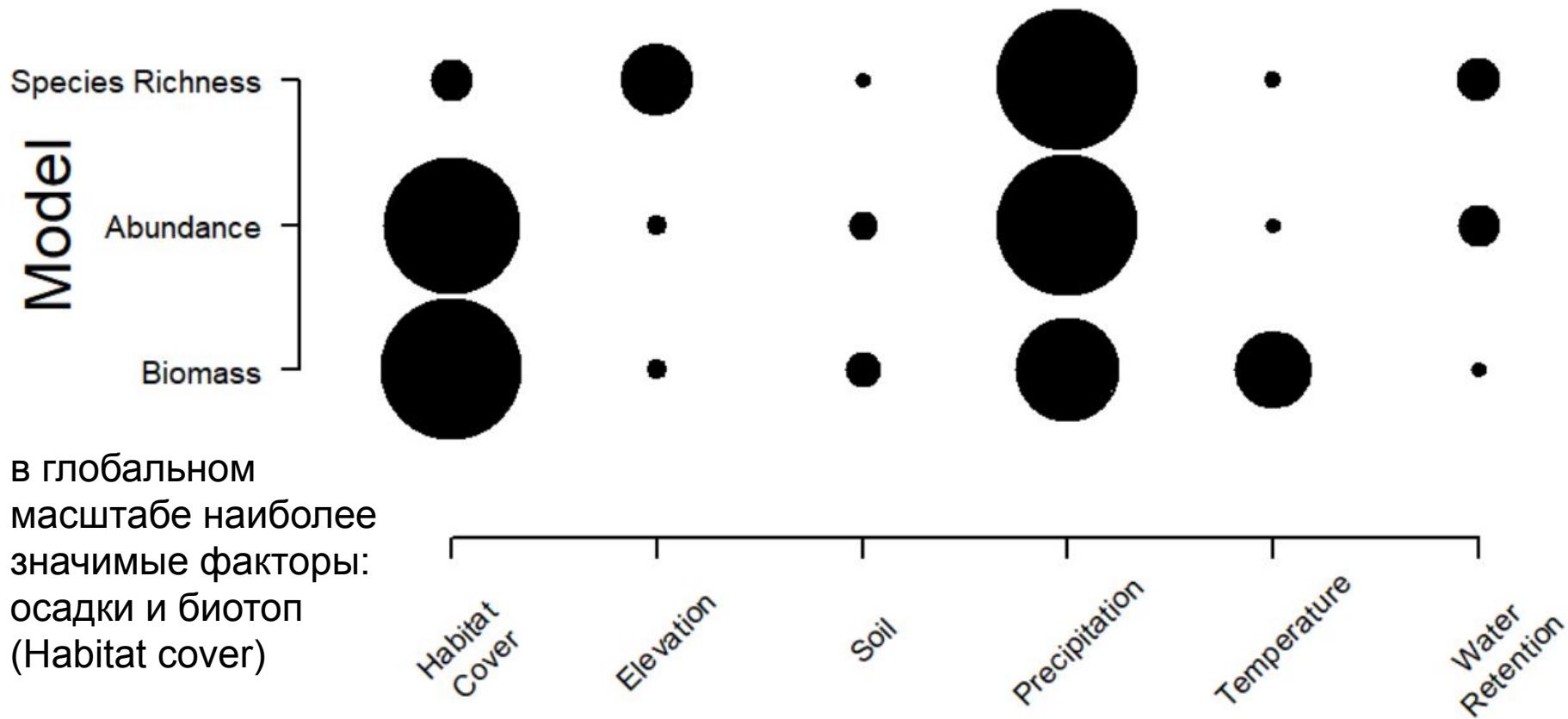
4

Number of species

Распределение обилия, в экз. на кв. м



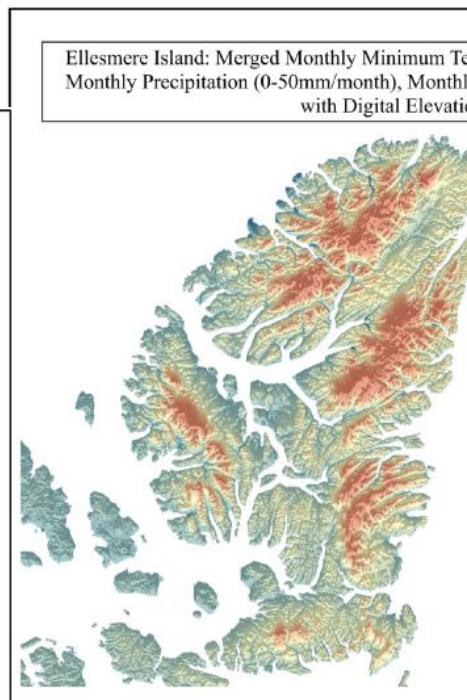
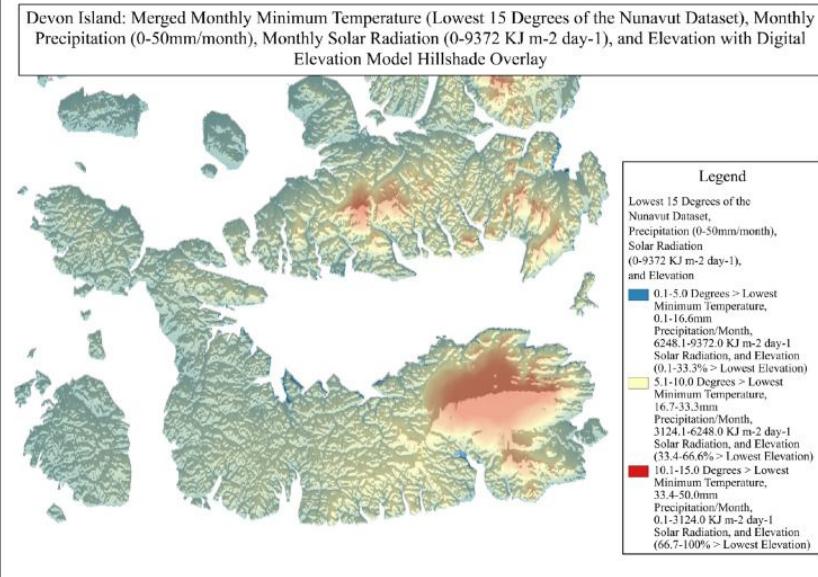
Влияние факторов на население дождевых червей



И на Марсе будут мятелики цвести

Eric Vaz, Elissa Penfound (2020) Mars Terraforming: A Geographic Information Systems Framework. Life Sciences in Space Research

DOI: [10.1016/j.issr.2019.12.001](https://doi.org/10.1016/j.issr.2019.12.001)



Системы облачных вычислений *ecocloud* и *bccvl*

 **ecocloud**
beta

[Dashboard](#) [Workspace](#) [Explorer](#) [Tools](#) [Support](#)

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.	 Integration Connect easily with other analyses hosted externally on GitHub repositories.
 Multi-code language We offer either Python or R (and RStudio) platforms to write, edit and run your code.	 Use existing data Take your analyses further by integrating publicly available datasets to your collected data.
 Easy access Login through your Australian University account. Otherwise, it's free to make a new account.	 Ongoing support We are a helpful team here to support you with your code, analysis and teaching materials.

 **bccvl**

[About](#) [Data Portal](#) [Features](#) [Training](#) [Blog](#) [Contact](#) [Login/Register](#)



Modelling at your fingertips
Your complete biodiversity and climate impact modelling platform

[Get Started](#)

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](https://doi.org/10.1016/j.envsoft.2015.10.025)

4121

Climate Data Layers

306

Environmental Data Layers

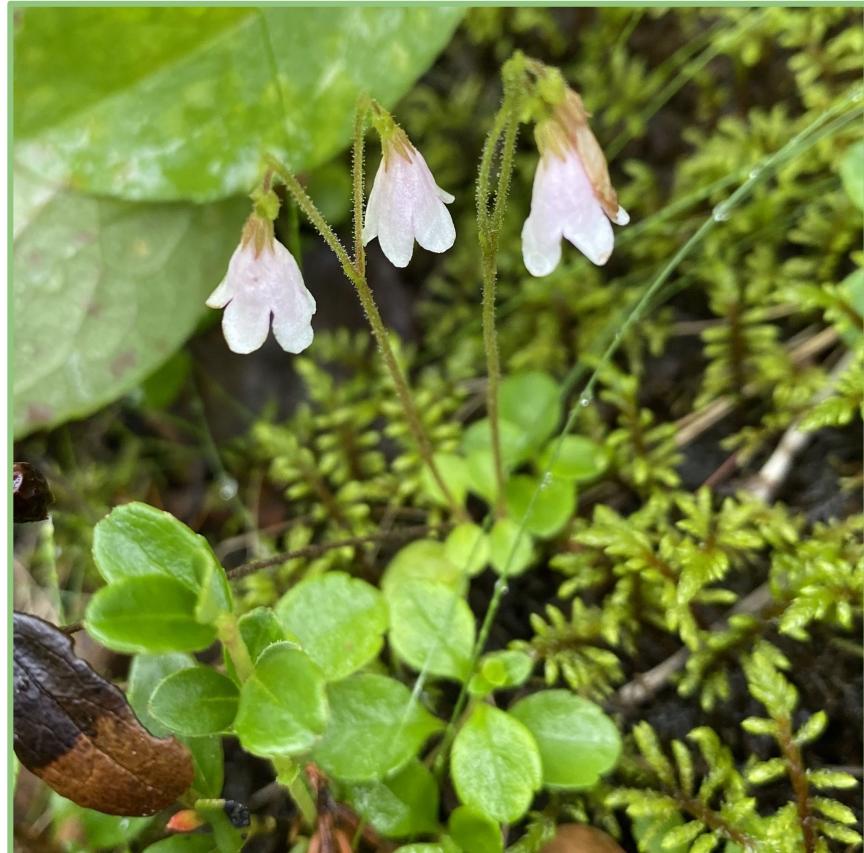
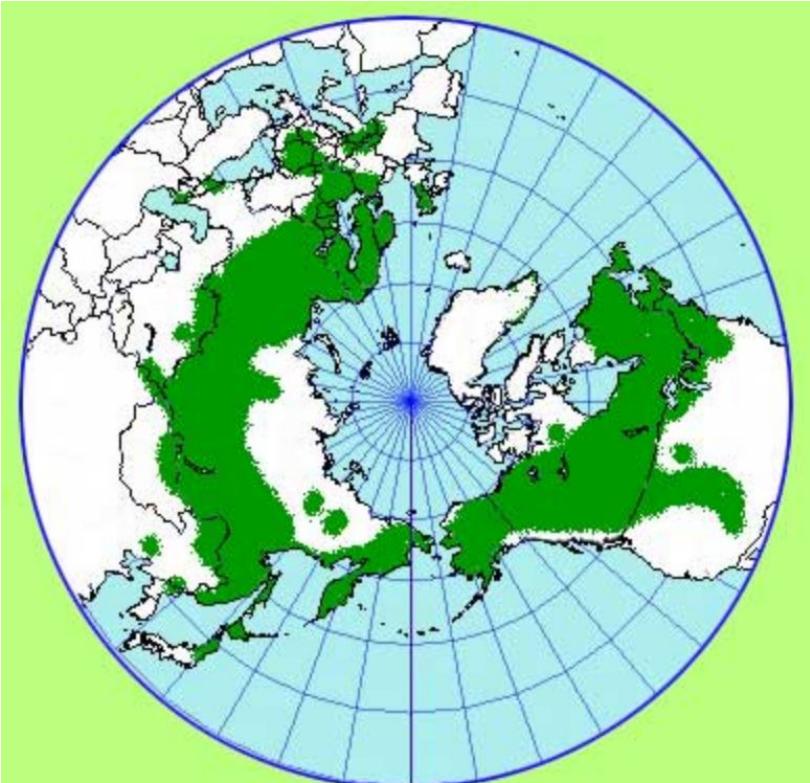
58

Average model runtime (sec)

Linnaea borealis L.

e-flora for Central Yukon

<http://www.flora.dempstercountry.org/>



Наблюдения *Linnaea borealis* в iNaturalist

на 15 августа 2020 года, с точностью привязки не более 2500 м
всего - 5008 наблюдений,
из них в России - 516



Набор данных

 bccvl BETA

Datasets Experiments Training Support Share Maxim Shashkov

New Species Distribution Modelling Experiment

Do you have a dataset you would like to see available in the BCCVL? - Contact us and we'll try and make it happen.

◀ Previous Description Occurrences Absences Climate & Environmental Data Constraints Algorithms Run Next ▶

Select the occurrences.

Species Occurrence Datasets

Linnaea borealis (GBIF) 250 m selective (+) (x)
Linnaea borealis L.
Rows: 2631

+ Select A Dataset

Need to import or upload new occurrence datasets? Visit the datasets page.

You must select an occurrence dataset for this experiment type.



Данные среды: климатические, топографические и прочие

Если набор о среде обитания включает несколько растровых слоёв, то можно выбрать необходимые.

Так как результатом моделирования представляет собой растровый слой, вычисления производятся на определенном масштабе, к которому должны быть приведены все входные слои

Select the climate and other environmental variables

- Select common resolution
- Scale to finest resolution
- Scale to coarsest resolution

Climate & Environmental Datasets

WorldClim, current climate (1950-2000), 5 arcmin (~10 km)

Select All / Select None

bccvl BETA

Datasets Experiments Training Support Share

Maxim Shashkov

New Species Distribution Modelling Experiment

Do you have a dataset you would like to see available in the BCCVL? - Contact us and we'll try and make it happen.

◀ Previous Description Occurrences Absences Climate & Environmental Data Constraints Algorithms Run Next ▶

Select the climate and other environmental variables.

Select common resolution
 Scale to finest resolution
 Scale to coarsest resolution

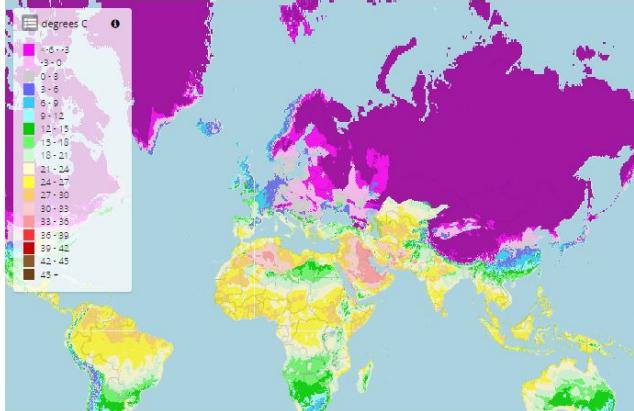
Climate & Environmental Datasets
You must select at least one.

WorldClim, current climate (1950-2000), 5 arcmin (~10 km)
Select All / Select None

B01 - Annual Mean Temperature
 B02 - Mean Diurnal Range (Mean of monthly (max temp - min temp))
 B03 - Isothermality (BIO2/BIO7)
 B04 - Temperature Seasonality (standard deviation)
 B05 - Max Temperature of Warmest Month
 B06 - Min Temperature of Coldest Month
 B07 - Temperature Annual Range (BI05-BI06)
 B08 - Mean Temperature of Wettest Quarter
 B09 - Mean Temperature of driest Quarter
 B10 - Mean Temperature of Warmest Quarter
 B11 - Mean Temperature of Coldest Quarter
 B12 - Annual Precipitation
 B13 - Precipitation of Wettest Month
 B14 - Precipitation of driest Month
 B15 - Precipitation Seasonality (Coefficient of Variation)
 B16 - Precipitation of Wettest Quarter
 B17 - Precipitation of driest Quarter
 B18 - Precipitation of Warmest Quarter
 B19 - Precipitation of Coldest Quarter

degrees C

0 - 3
3 - 6
6 - 9
9 - 12
12 - 15
15 - 18
18 - 21
21 - 24
24 - 27
27 - 30
30 - 33
33 - 36
36 - 39
39 - 42
42 - 45
45 -



You must select at least one.

WorldClim Altitude at 5 arcmin (~10 km)
Select All / Select None

Altitude (elevation above sea level)

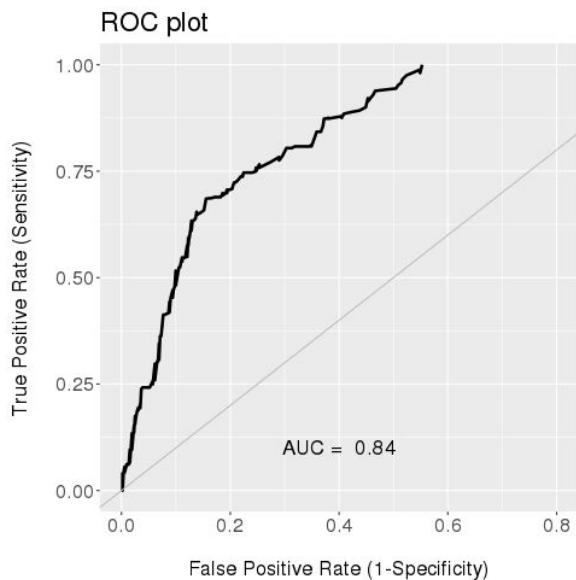
Select Available Datasets

Результат моделирования ареала *Linnaea borealis* методом maxEnt

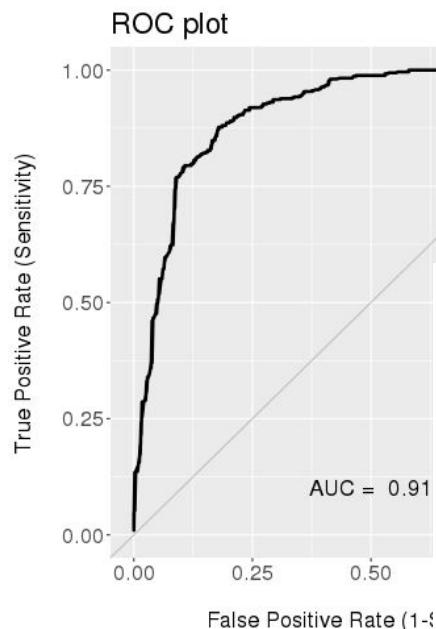


Оценка качества модели ROC - AUC

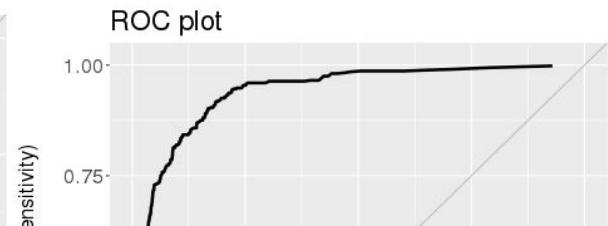
bioclim



Generalized
Linear Model



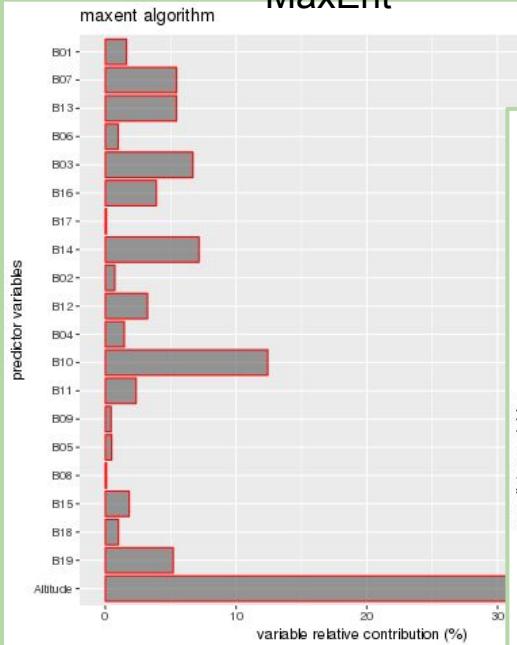
MaxEnt



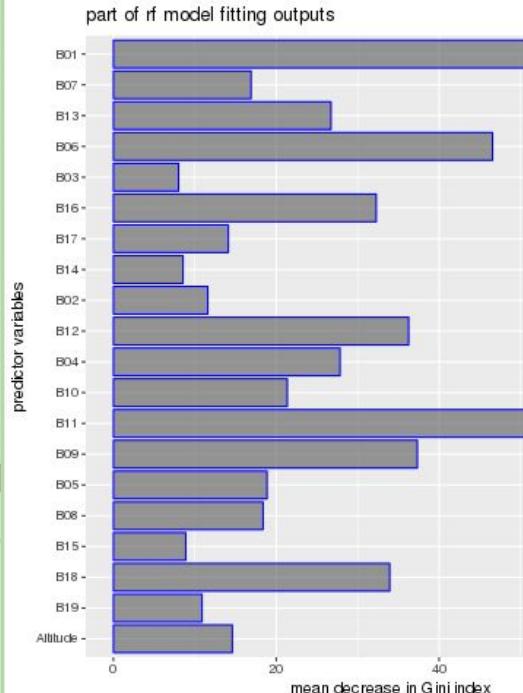
Observation			
		Present	Absent
Prediction	Present	True positive	False positive
	Absent	False negative	True negative

Оценка вклада факторов среды

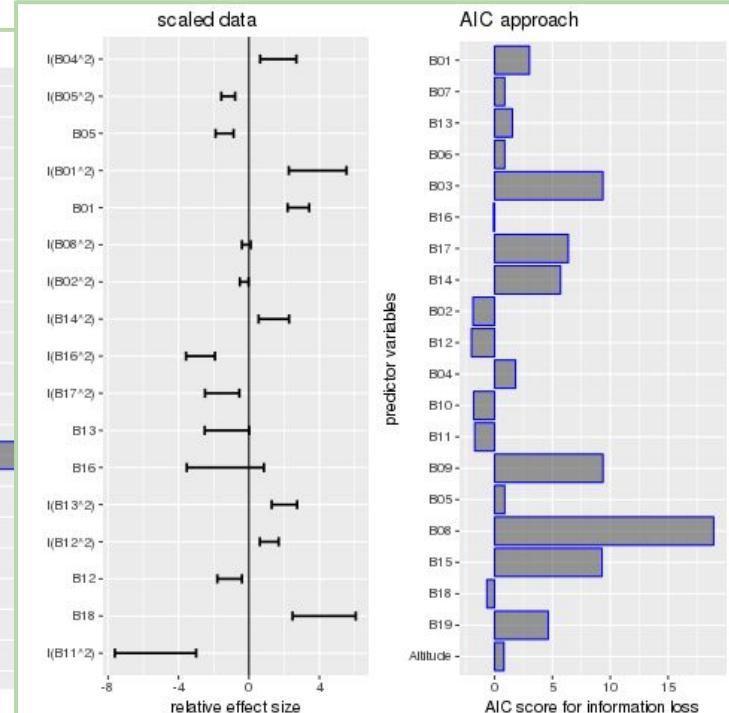
MaxEnt



Random Forest

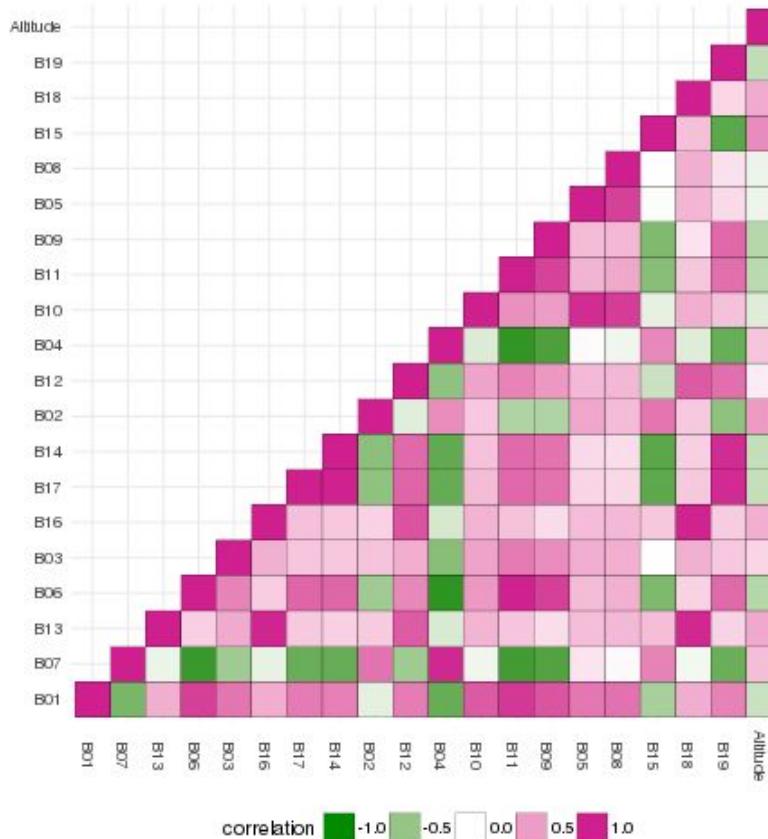


Generalized Linear Model

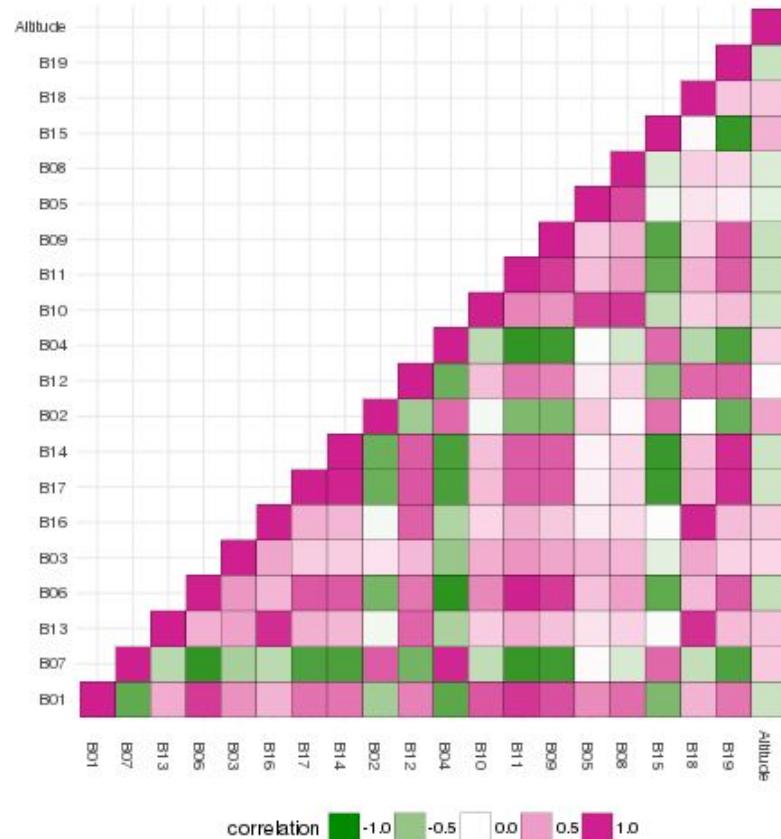


Корреляция факторов среды (переменных)

MaxEnt



Generalized Linear Model



Моделирование в среде R

R Studio

основная библиотека - dismo

дополнительные для подготовки и обработки пространственных
данных:

sp - базовые функции

raster - растровые данных

rgdal - форматы данных

rgeos - обработка геометрии

fields - для интерполяции растров

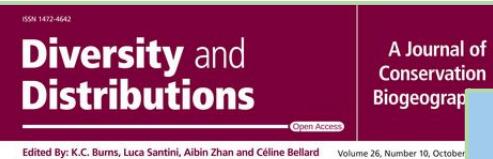
rgbif - данные находок из GBIF (могут быть точки отсутствия)



Modelling (SDM) - литература



1998, 4.092

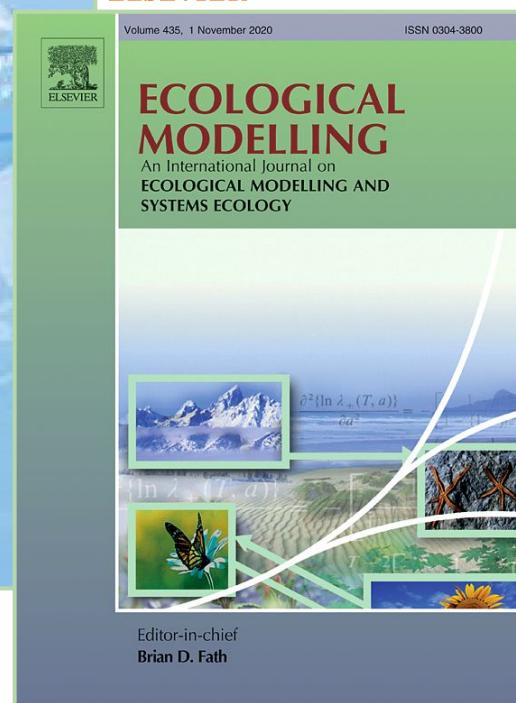


Springer 2005, 3.137



ELSEVIER

1975, 2.978



Ecography

Ecological Indicators

Ecological
Informatics

Remote Sensing of
Environment

Environmental
Modelling & Software

Global Ecology and
Conservation

Journal of Applied
Ecology

Полезные статьи

Zurell et al., 2020, A standard protocol for reporting species distribution models

DOI: [10.1111/ecog.04960](https://doi.org/10.1111/ecog.04960)

Araujo et al. (2019) Standards for distribution models in biodiversity assessments

DOI: [10.1126/sciadv.aat4858](https://doi.org/10.1126/sciadv.aat4858)

Booth et al. (2014) BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies

Elith et al. (2010) A statistical explanation of MaxEnt for ecologists

Phillips, Anderson, Schapire (2006) Maximum entropy modeling of species geographic distributions DOI: [10.1016/j.ecolmodel.2005.03.026](https://doi.org/10.1016/j.ecolmodel.2005.03.026)

Благодарности



Chantal Huijbers



Евгения Брагина



Wildlife Conservation Society

*Стать мастером нетрудно, трудно оставаться мастером.
Масса литературы, масса новых методов, новых приложений,
за всем надо следить, надо непрерывно экспериментировать,
исследовать, и надо непрерывно следить за смежными
областями ...*

*А. и Б. Стругацкие
1965*

Спасибо за внимание !!!