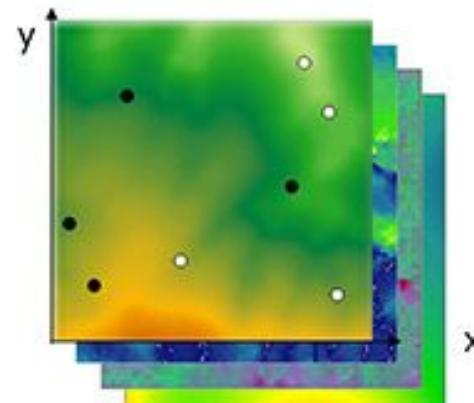


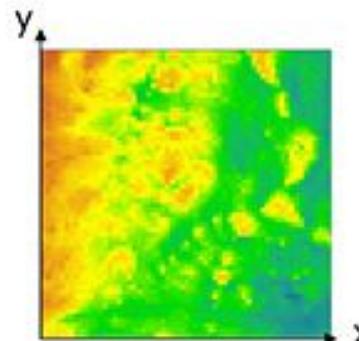


ЛЕКЦИЯ 11

Моделирование ареалов Species Distribution Modelling



Field records and maps of environment



Map of probability species is present

Species Distribution Modeling - SDM

Ecological (Environmental) Niche Modelling

Species niches and distribution

Bioclimatic envelope models

Habitat Suitability

Climate suitability

Установление количественной связи между географическим распределением популяции исследуемого вида и факторами среды с целью моделирования его распространения в настоящих, прошлых или будущих условиях

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

Открытые источники данных

Биоразнообразие



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


Clanga pomarina (Brehm, 1833)


News
Call for author proposals for new and updated GBIF technical guides


News
Call for applications: regional support contractors to develop the community of practice in Asia


News
Call for data papers describing datasets from Russia


News
Namibia joins GBIF as associate participant

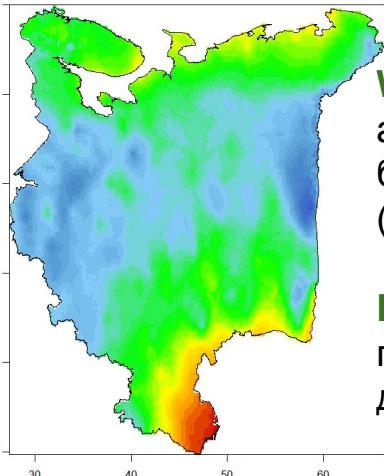
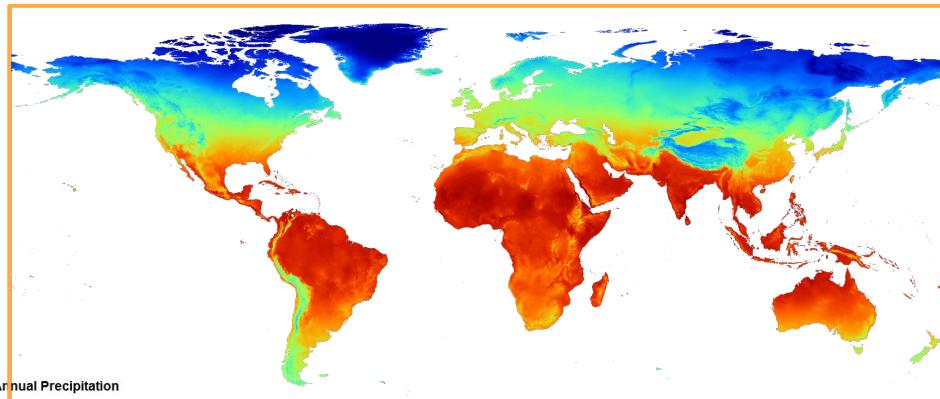

News
BID programme funds 18 new projects in sub-Saharan Africa


News
Community consultation: Converging Digital Specimens and Extended Specimens


News
GBIF forecast: increasing chance of clouds for species occurrence data


Blog
Using GBIF-mediated data with Apache Spark on Amazon Web Services

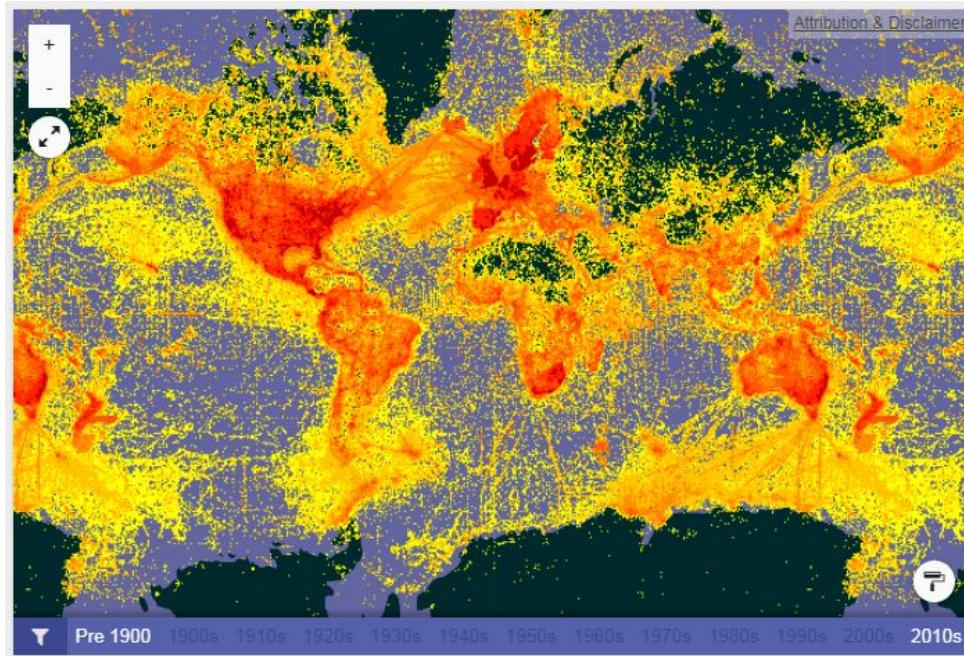
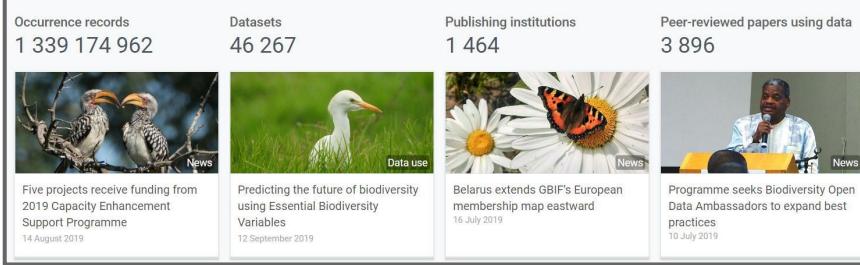
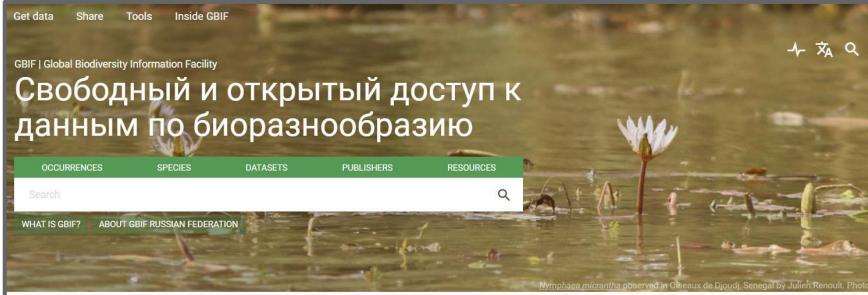
Данные среды



WorldClim - климатические данные: архивные, современное состояние и будущие, согласно разным сценариям (RCP и SRES, например)

EarthEnv - классификация земной поверхности на основе данных дистанционного зондирования

Репозитории - данные о биоразнообразии



Данные среды

WorldClim (www.worldclim.org)

биоклиматические данные: bio1...19, altitude

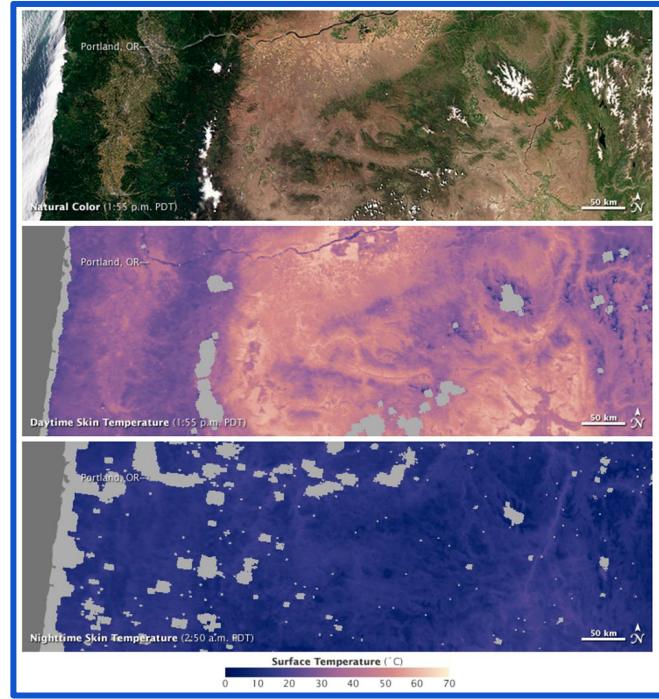
EarthEnv (www.earthenv.org)

Global 1-km Consensus Land Cover

Классификация земной поверхности (12 классов)

SoilGrids250 2.0 - ISRIC (soilgrids.org)

глобальная почвенная картография



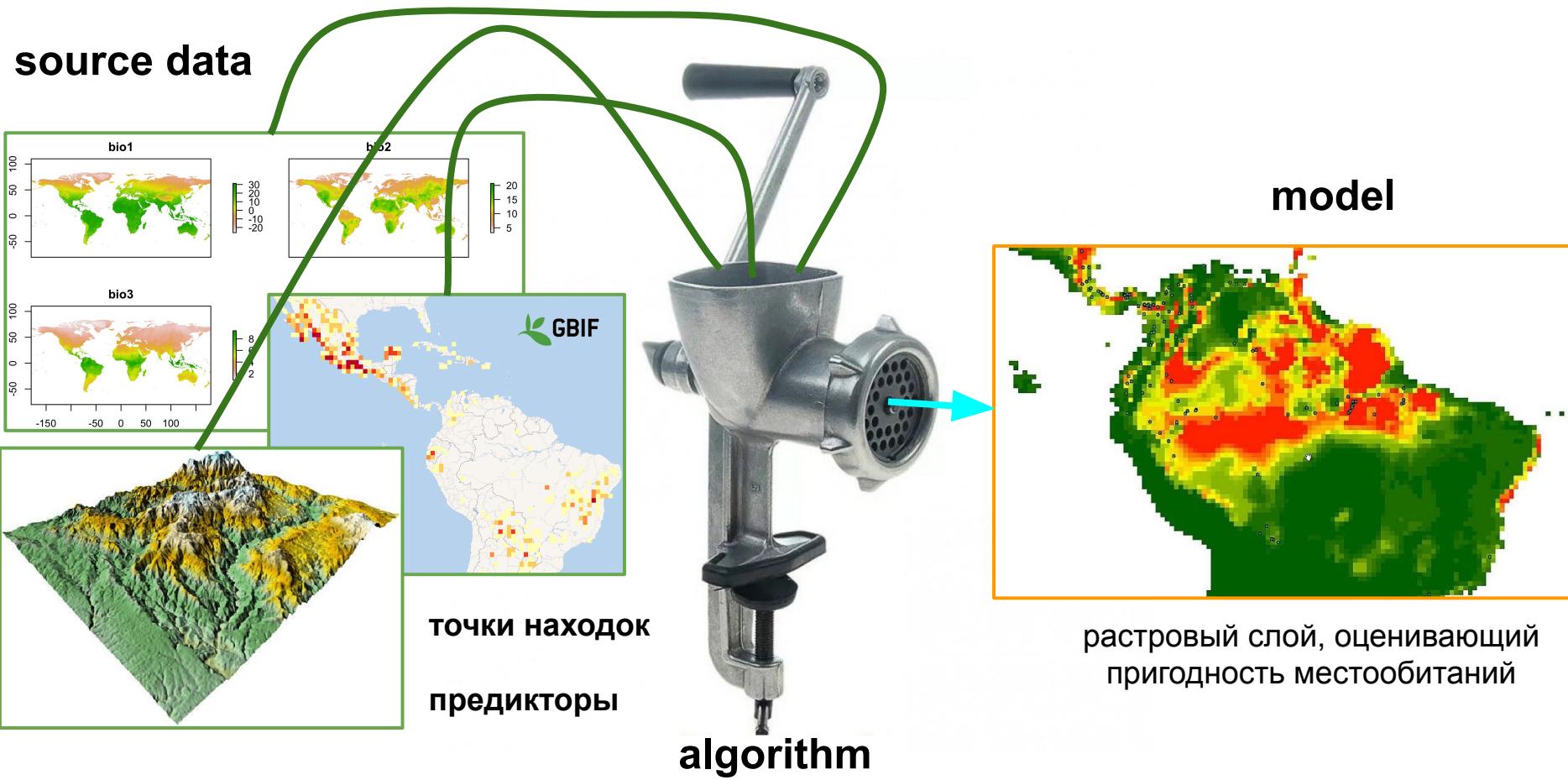
Данные NASA для SDM - результат обработки спутниковых данных

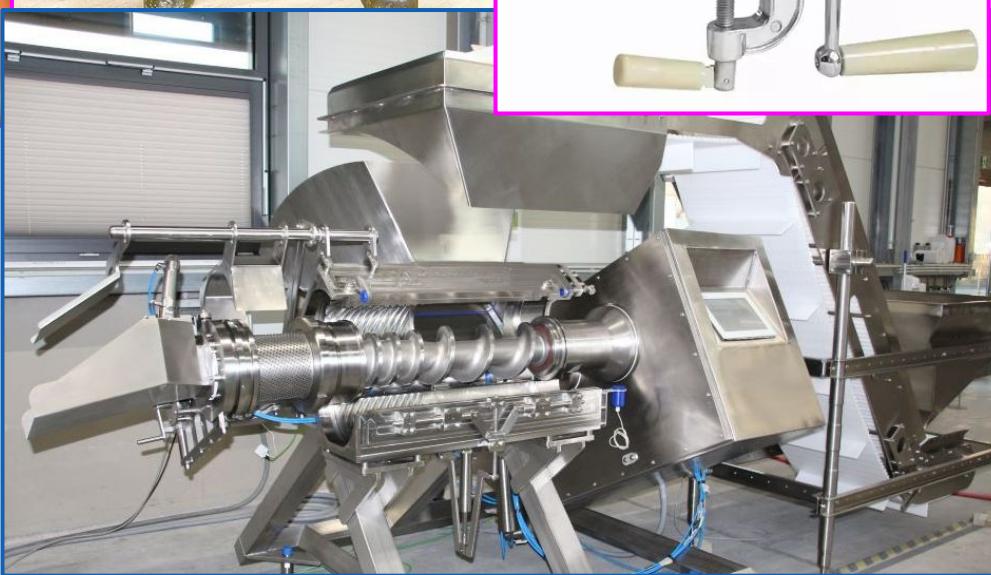
(earthdata.nasa.gov/learn/pathfinders/biodiversity/species-distribution)

температура земной поверхности, влажность почвы, снежный покров

температура и соленость морской поверхности

Species Distribution Modelling - Workflow





Классы моделей SDM

Минимум 17 основных методов моделирования
(**по данным `bccv1`**), объединенных в 4 класса:

Географические

Профильные

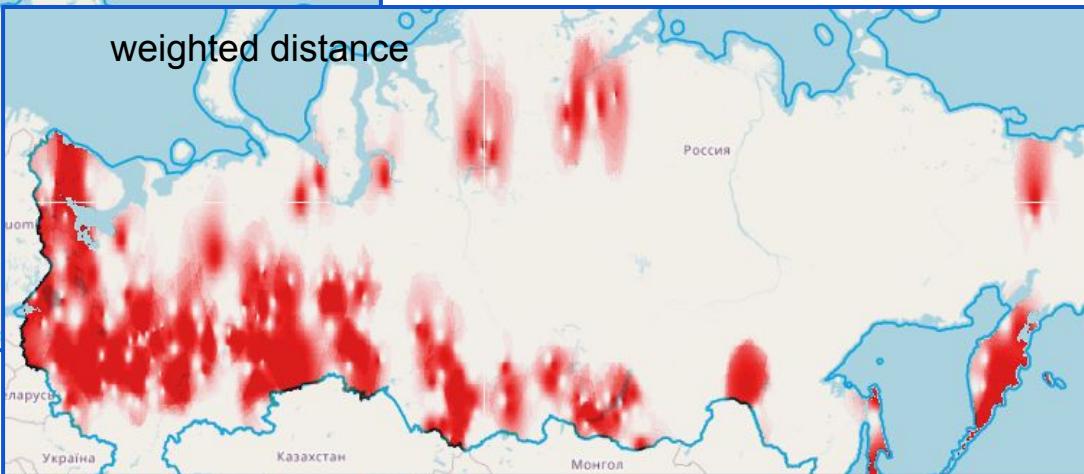
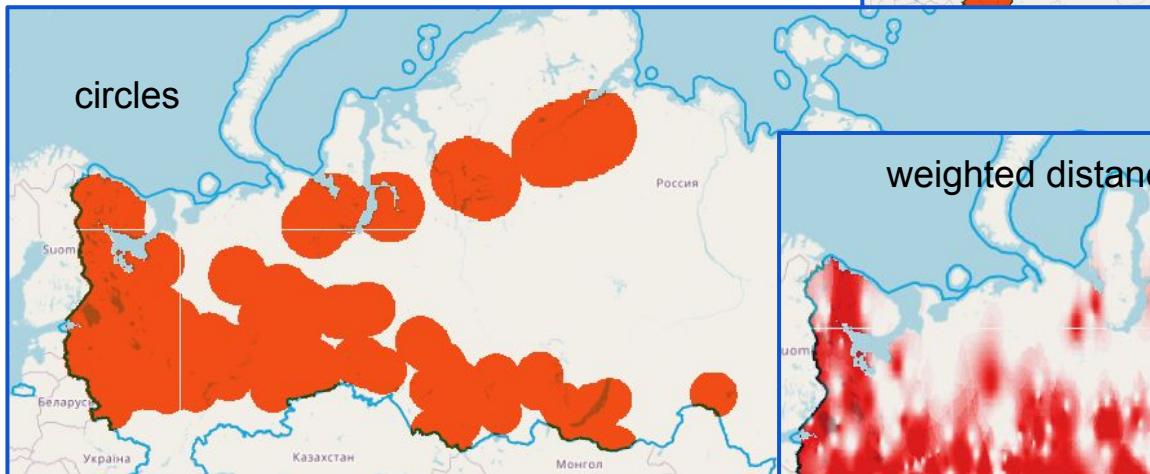
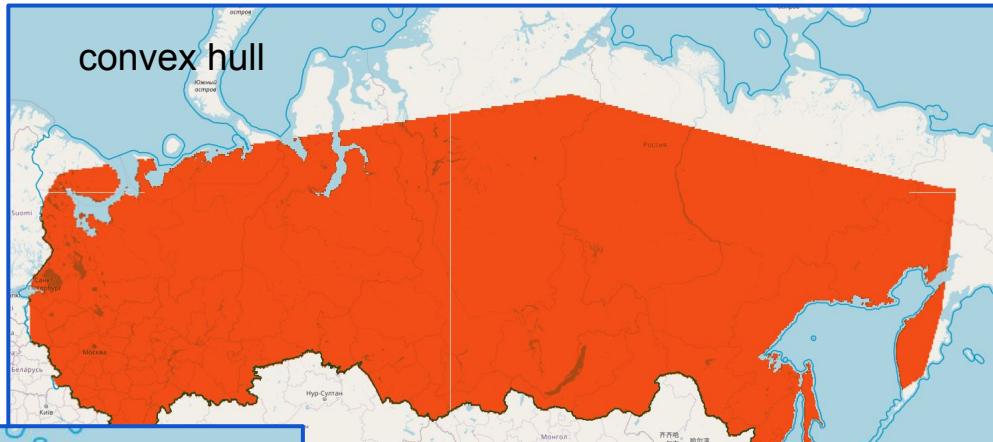
Статистические (регрессионные)

Основанные на машинном обучении (Machine Learning)

Географические

Используют точки находок

Данные среды не используют



Профильные

*climate envelope models
bioclimatic models*

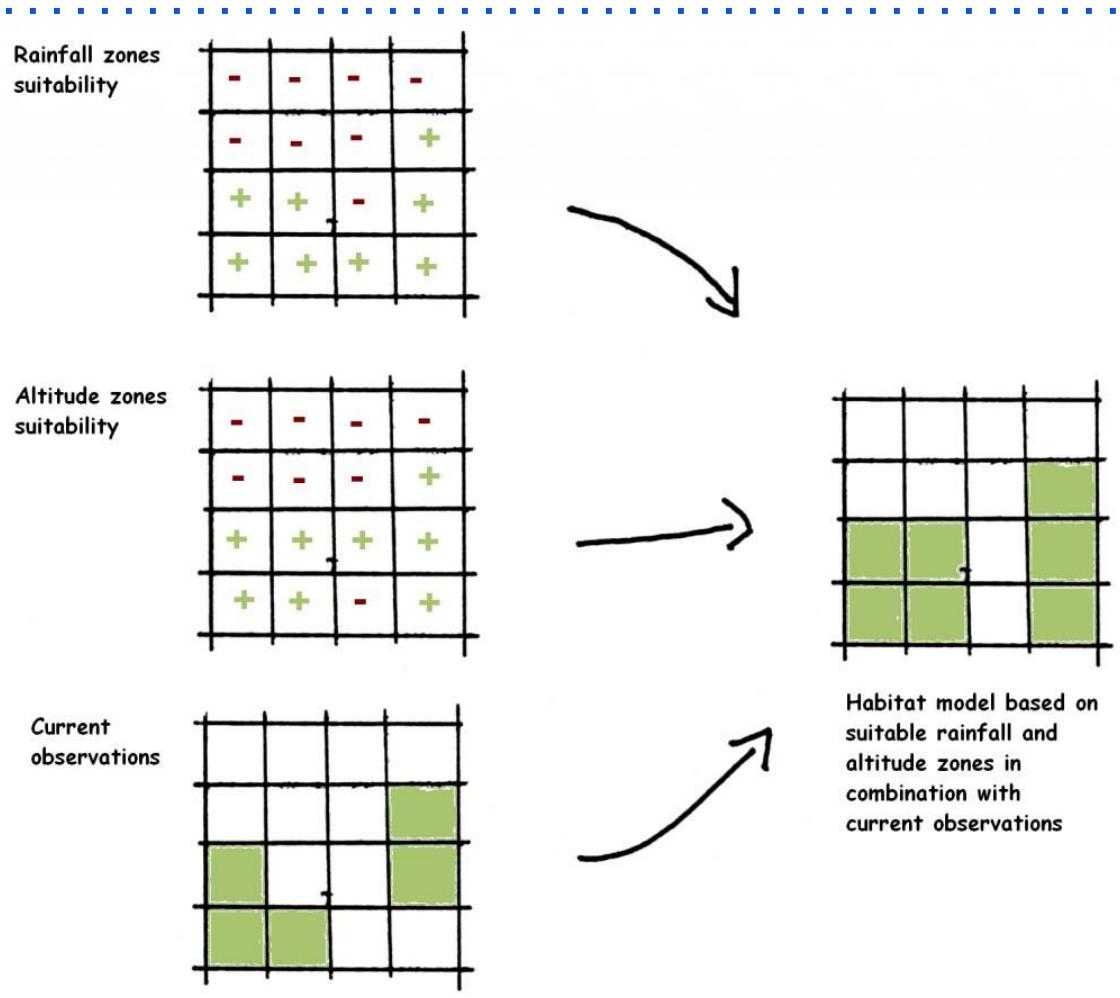
с чего всё началось:

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

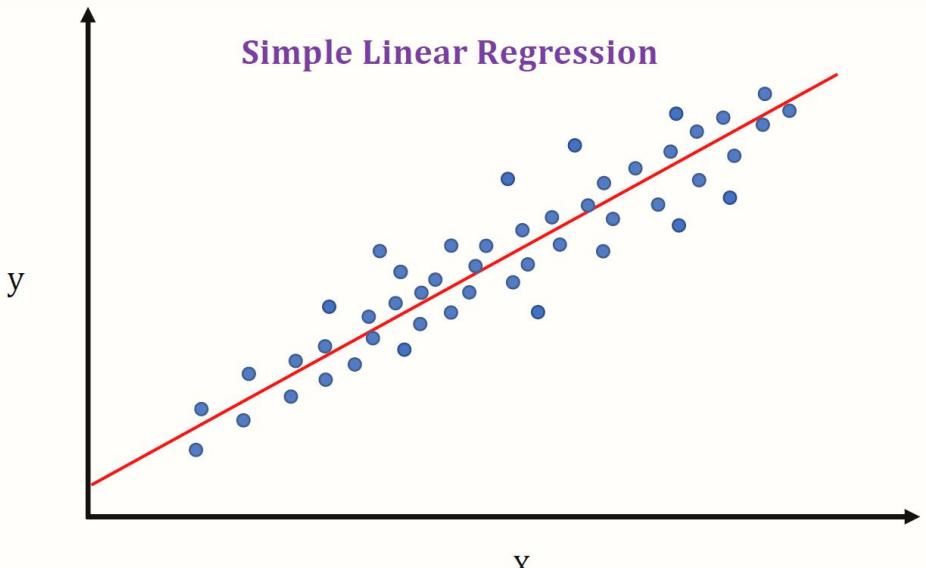


Статистические - регрессионные

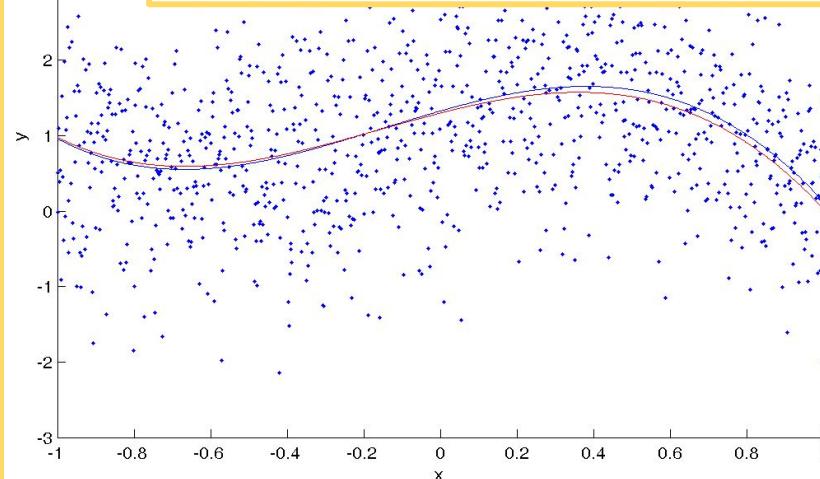
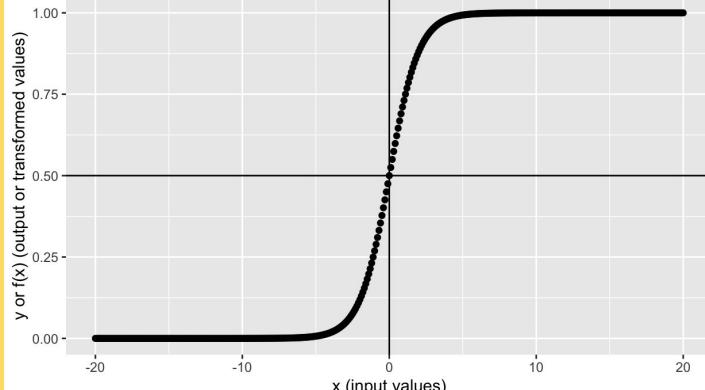
Используют данные среды

Используют точки находок
и отсутствия целевого вида

Simple Linear Regression



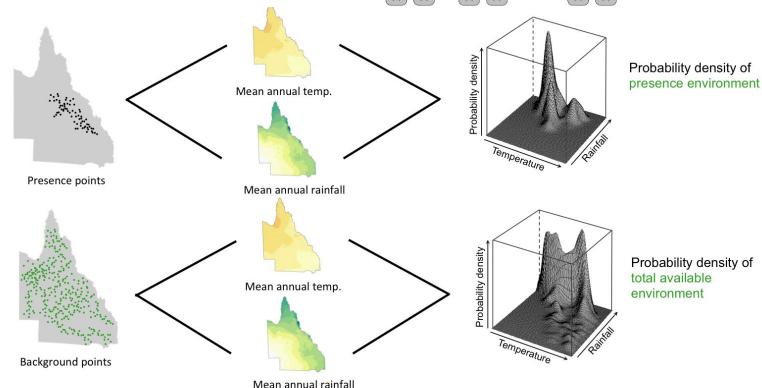
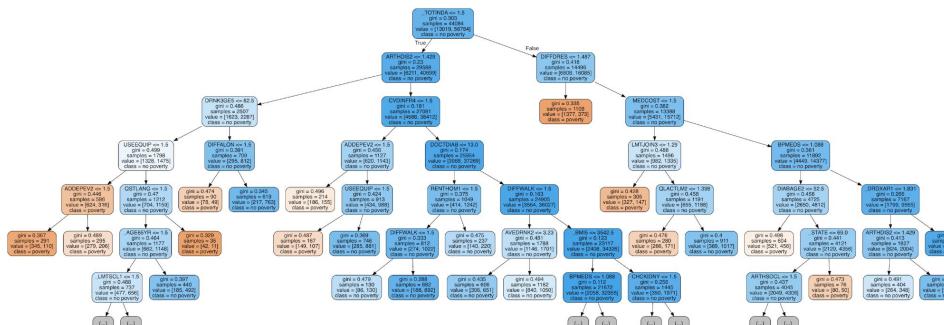
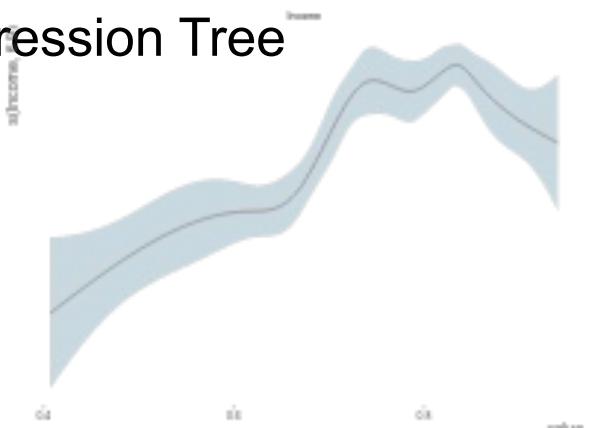
Logistic (sigmoid) function



Машинное обучение

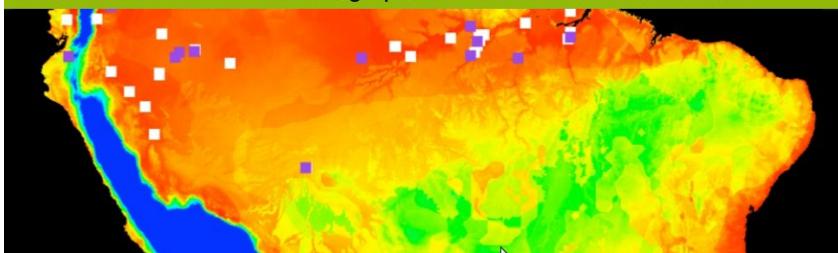
данные среды
точки присутствия и отсутствия
кросс-валидация

Random Forest
MaxEnt - классика жанра
Boosted Regression Tree

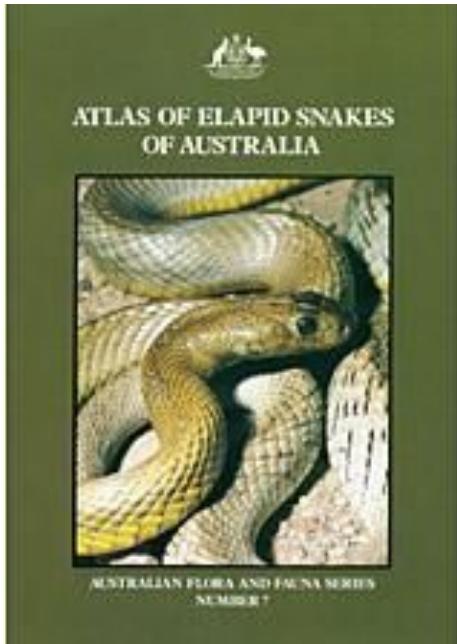


Adapted from Elith et al. (2011) *A statistical explanation of MaxEnt for ecologists*. *Diversity and Distributions*, 17, 43–57.

Maxent software for modeling species niches and distributions



Первые работы



Atlas of Elapid Snakes of Australia
1986

Аспидовые змеи (Epalidae). Австралия - древний
центр разнообразия группы

BIOCLIM 1984 - первая компьютерная программа
одна из первых работ - распространение 78 видов
ядовитых змей

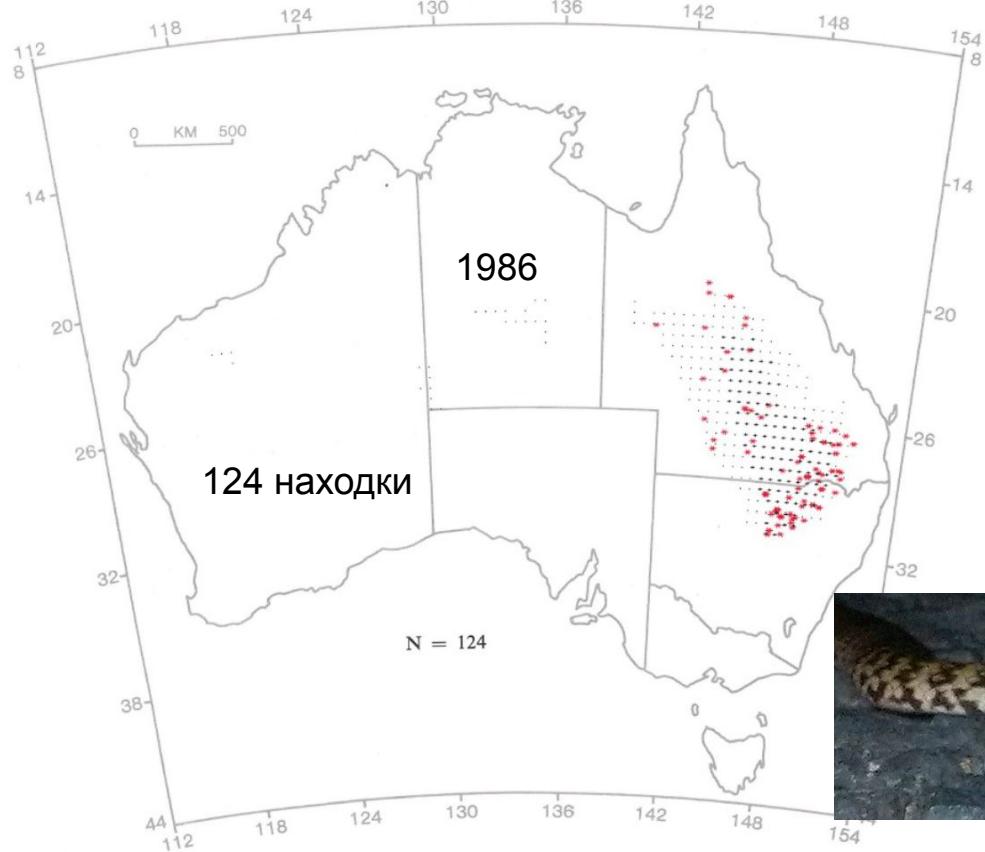
Nix, H.A. (1986) **A Biogeographic Analysis
of Australian Elapid Snakes**

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

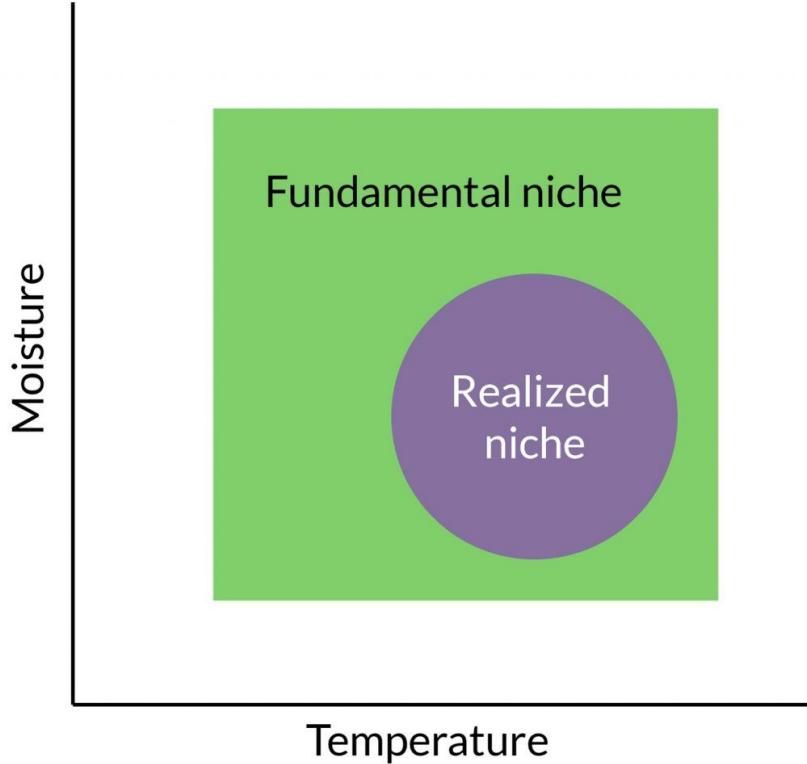


Commonwealth
Scientific and
Industrial
Research
Organisation

Ареал *Denisonia devisi*



Фундаментальная ниша



Реализованная ниша

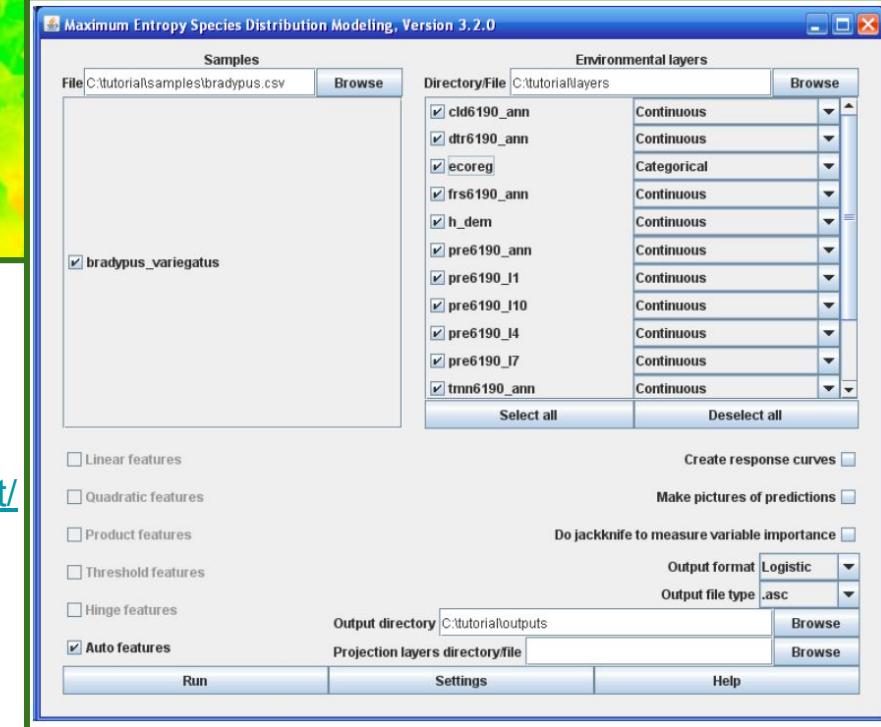
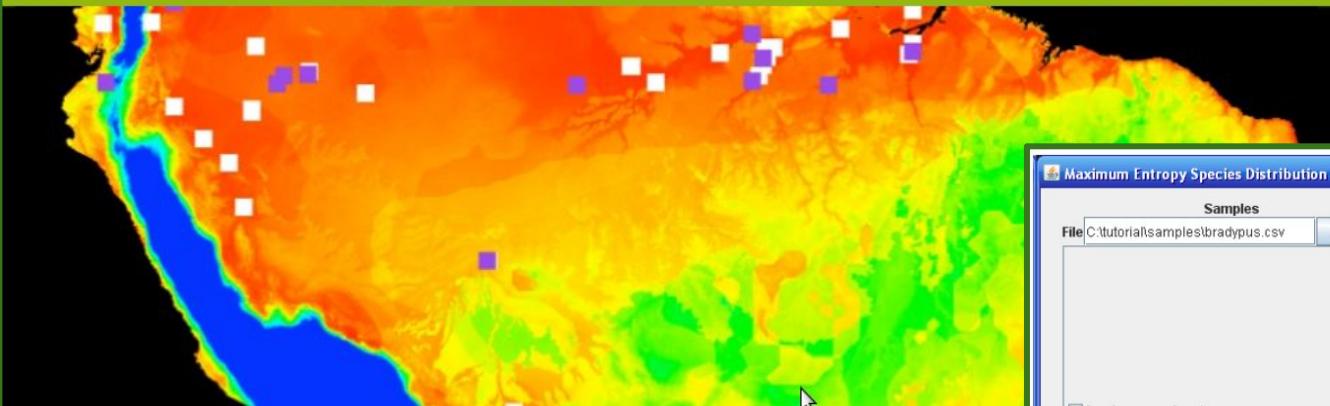
Ограничена:

- физическими барьерами
- конкуренцией
- антропогенным прессом

Оценка реализованной ниши - один из главных результатов

MaxEnt - приложение, 2006

Maxent software for modeling species niches and distributions



<https://github.com/mrmaxent/Maxent>

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

Руководство, учебные наборы данных и проч.

[1382 публикации, связанные с GBIF](#)

R - язык статистического программирования

с 1993 года, среди множества библиотек:

ggplot2 - продвинутая графика

dplyr - разнообразные манипуляции с таблицами

RPostgreSQL - для связи с СУБД на языке SQL

для работы с пространственными данными:

rgdal - работа географические проекциями

raster - обработка растровых данных

rgeos - операции с геометрией и топологией

sp - базовые методы для работы с пространственными данными

dismo - моделирование ареалов



Как получить данные через GBIF

1. Через портал

Чтобы загрузить на свой компьютер данные, доступные через GBIF, необходимо создать учетную запись.

Инструкция Как создать учетную запись на портале GBIF.org

После этого вы можете формировать запросы, загружать данные и использовать их в соответствии с лицензией и соблюдением правил цитирования.

Инструкция Как загрузить данные, опубликованные в GBIF

2. Через API

пакет для статистической среды R - `rgbif`

docs.ropensci.org/rgbif



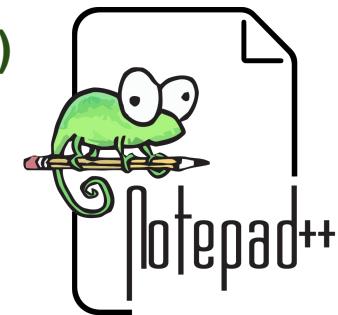
Необходимое программное обеспечение

Notepad ++ notepad-plus-plus.org/downloads

LibreOffice www.libreoffice.org

QGIS qgis.org

RStudio www.rstudio.com/products/rstudio (R не ниже 4.0)



PostgreSQL



Anaconda



Области применения SDM

Поиск новых местообитаний редких видов

Реинтродукция

Отслеживание инвазионных видов

Изучение сезонных миграций

Предсказание изменения ареала в условиях меняющегося климата

Реконструкция ареалов в прошлом

...

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)

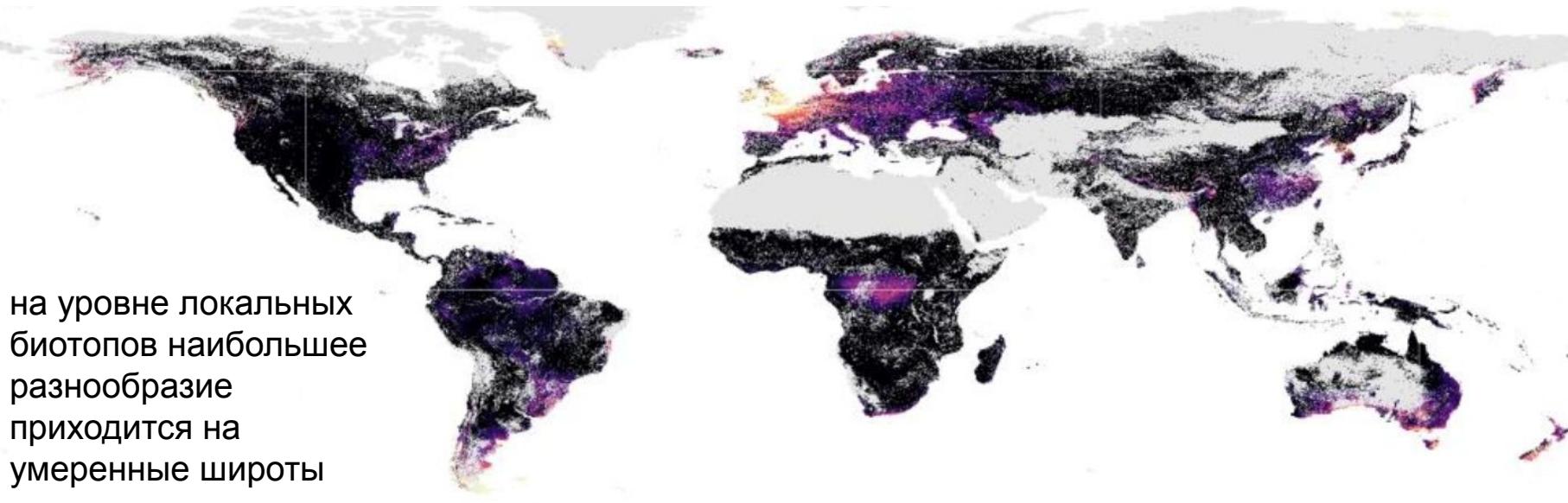
Number Of Sites

- 1
- 100
- 200
- 300
- 400
- 500
- 600



Предикторы:
Habitat cover
Elevation
Soil
Precipitation
Temperature
Water retention

Распределение видового разнообразия

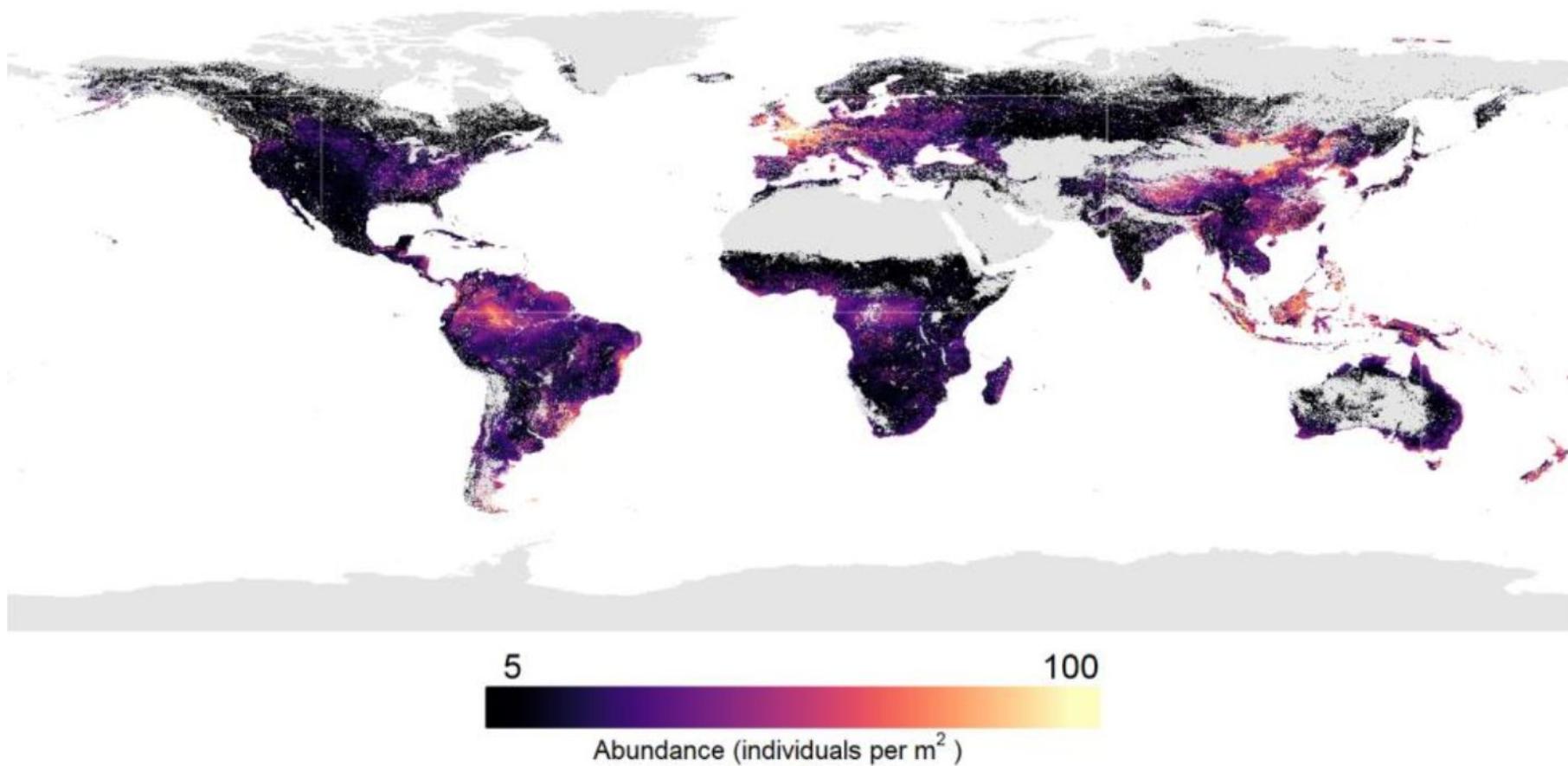


1

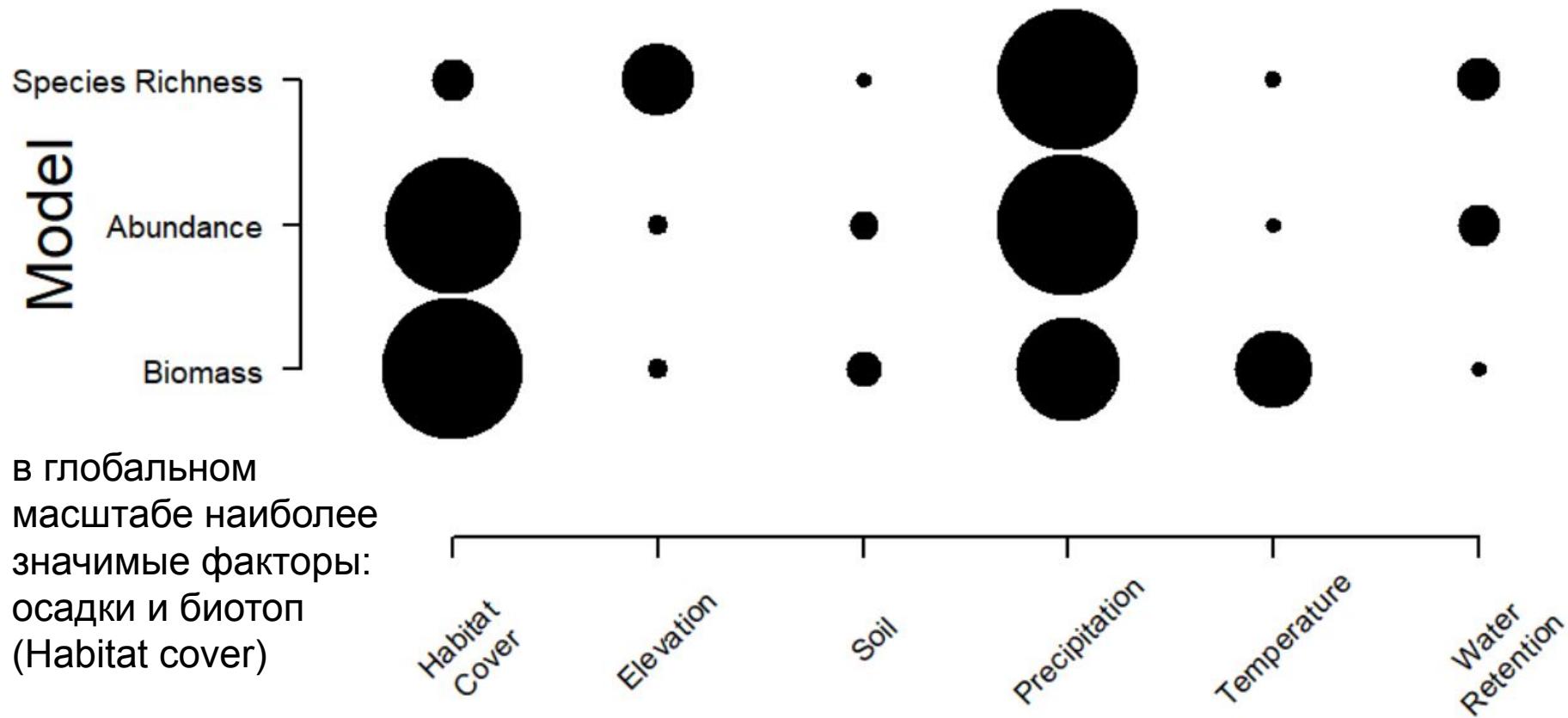
4

Number of species

Распределение обилия, в экз. на кв. м



Влияние факторов на население дождевых червей

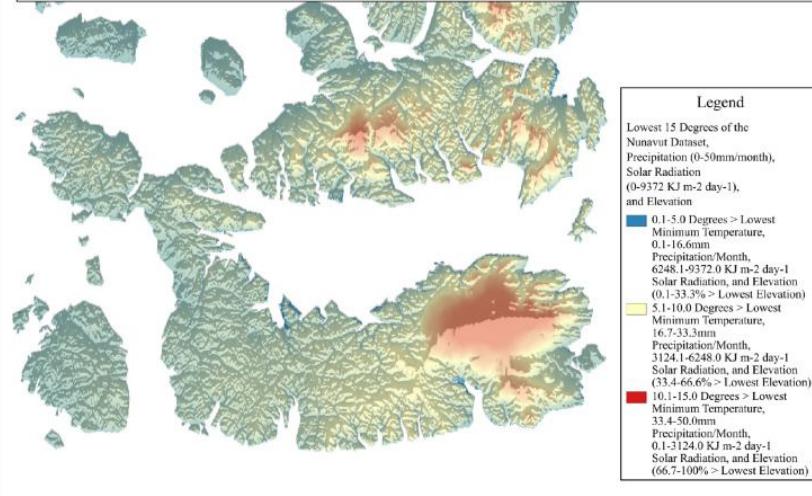


И на Марсе будут мятлики цвести

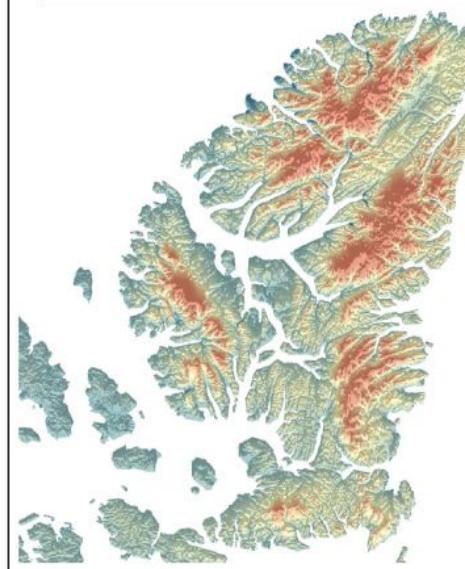
Eric Vaz, Elissa Penfound (2020) Mars Terraforming: A Geographic Information Systems Framework. Life Sciences in Space Research

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

Devon Island: Merged Monthly Minimum Temperature (Lowest 15 Degrees of the Nunavut Dataset), Monthly Precipitation (0-50mm/month), Monthly Solar Radiation (0-9372 KJ m⁻² day⁻¹), and Elevation with Digital Elevation Model Hillshade Overlay

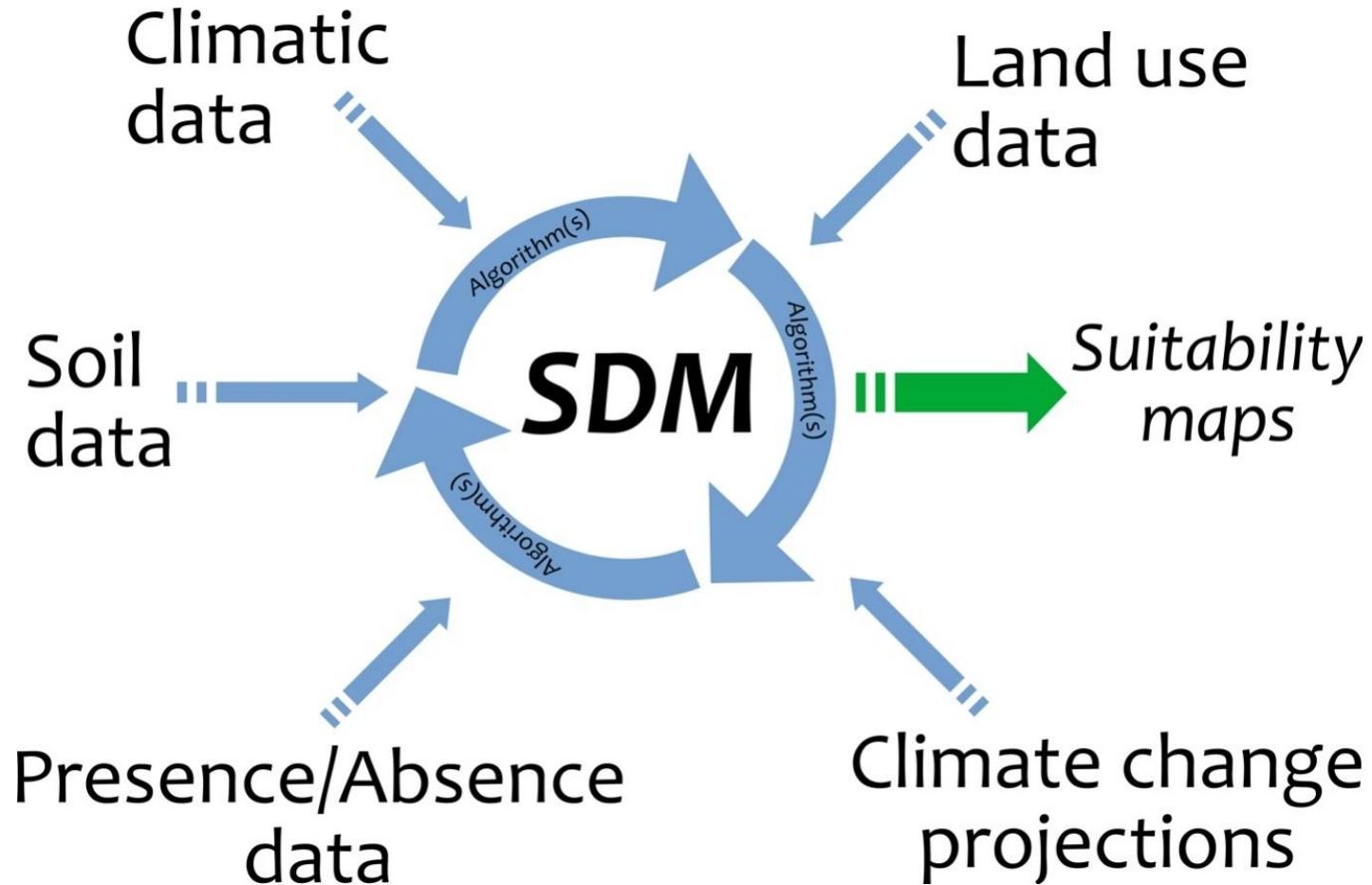


Ellesmere Island: Merged Monthly Minimum Temperature (Lowest 15 Degrees of the Nunavut Dataset), Monthly Precipitation (0-50mm/month), Monthly Solar Radiation (0-9372 KJ m⁻² day⁻¹), and Elevation with Digital Elevation Model Hillshade Overlay



Этапы моделирования - Modeling Workflow

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



Выбор находок

GBIF

iNaturalist

коллекции

литература

...



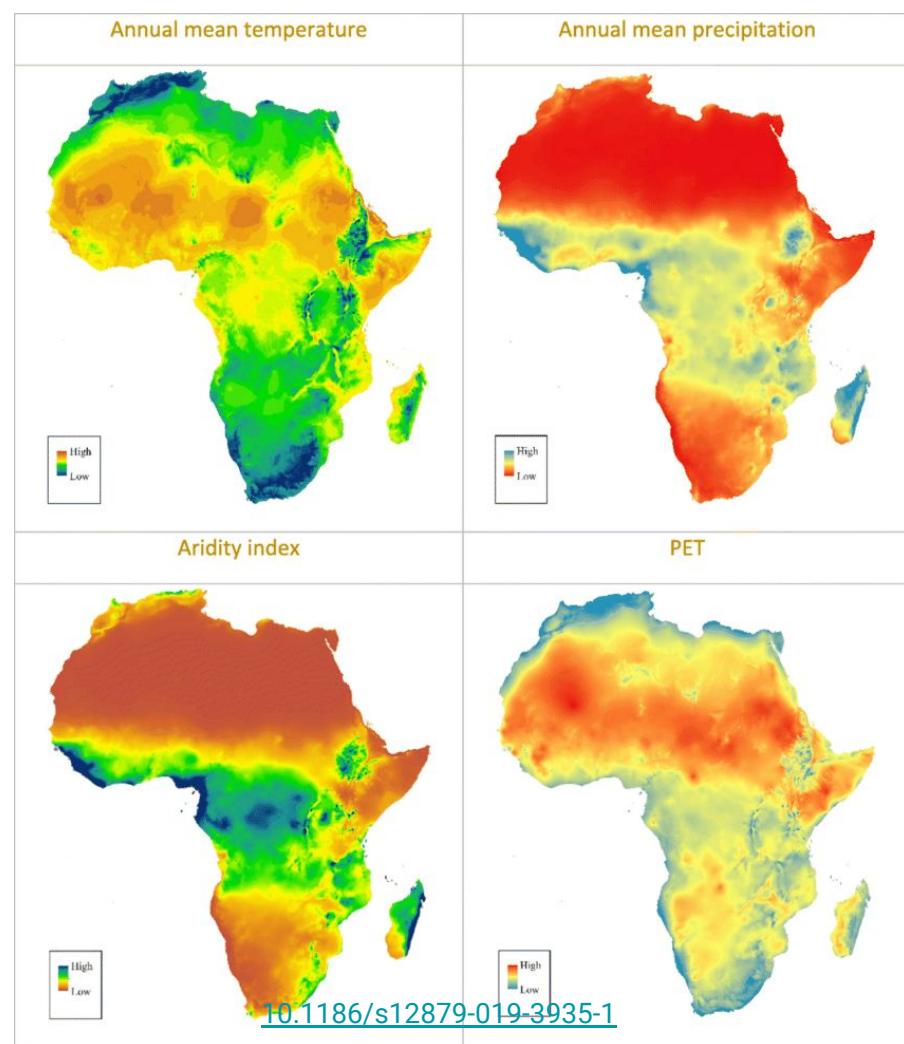
Факторы среды

Климатические данные

Высота над уровнем моря

Биотоп

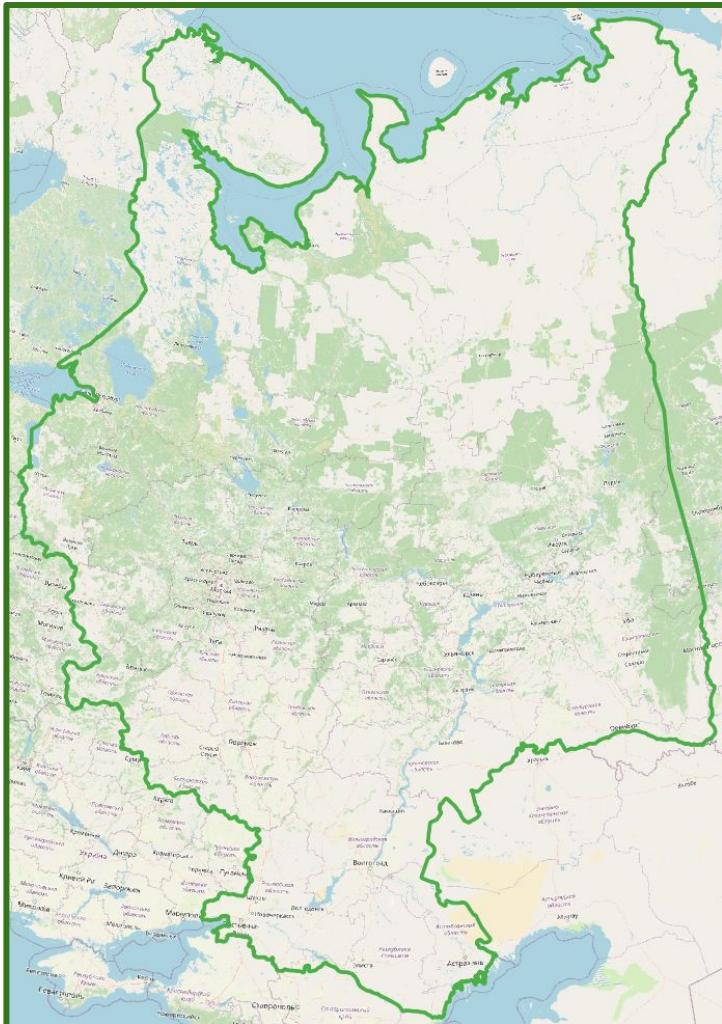
Прочая информация



Территория интереса

Обычно модель строится на локальном или региональном масштабе

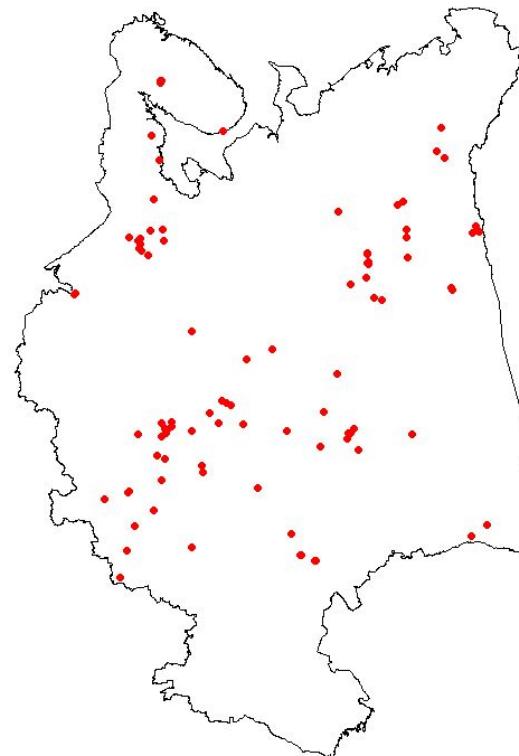
Надо учитывать не только горизонтальных, но и вертикальный масштаб



Входные данные с находками



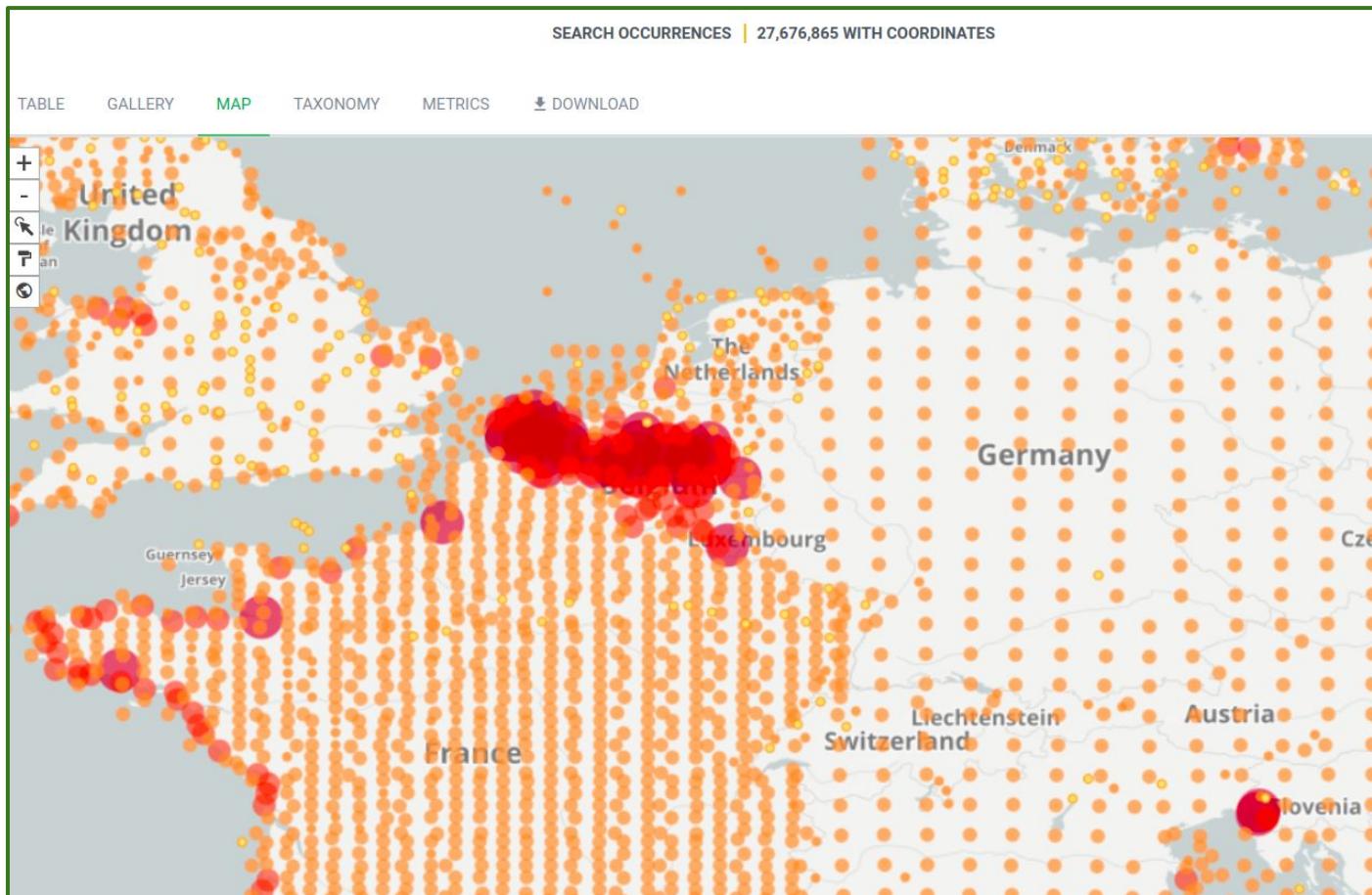
Малевич, Перель, Рапопорт, Зенкова, Гераськина, Шашков, ...
+ прочие находки из GBIF
= 110 уникальных точек находок



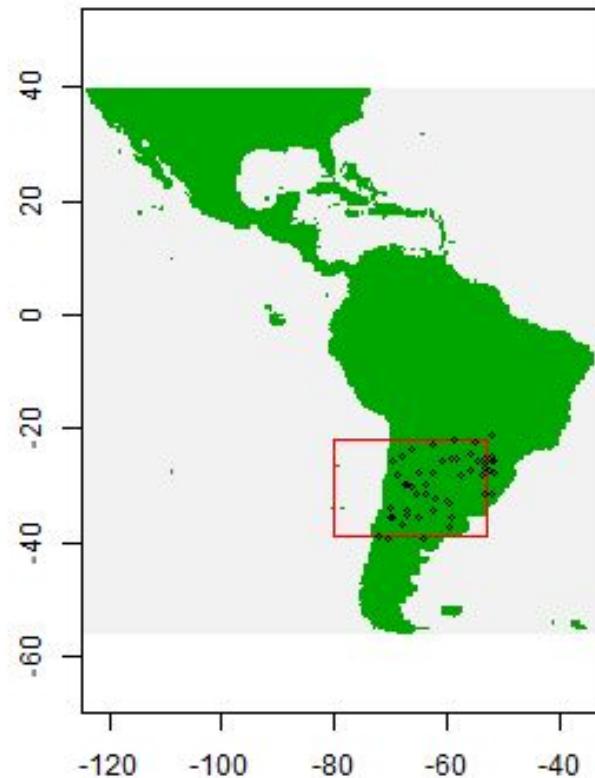
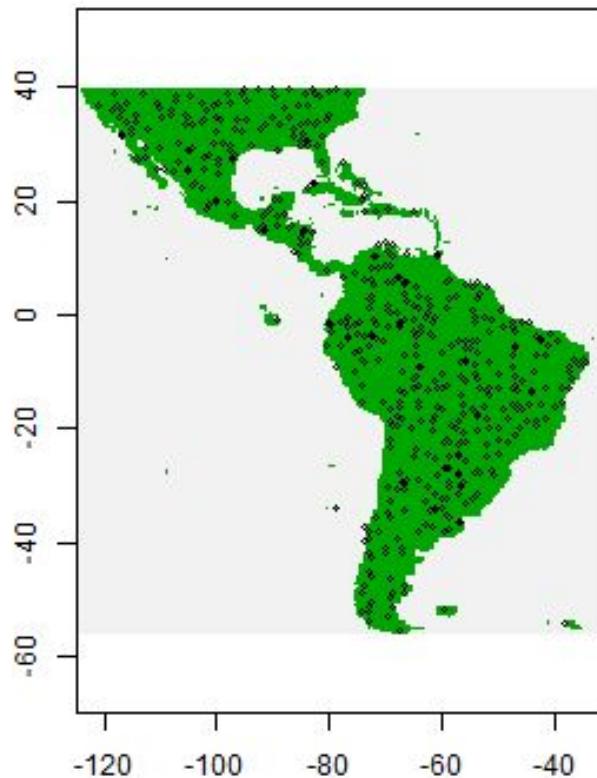
<https://www.gbif.org/dataset/e52aacbf-c701-462b-b68f-346502fbc270>

Точки отсутствия

В GBIF около 28 млн точек - 1.3 % от общего количества



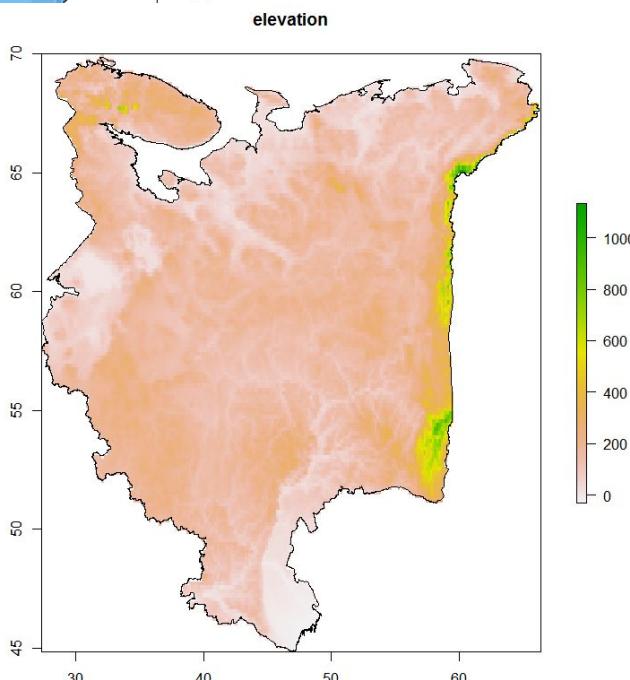
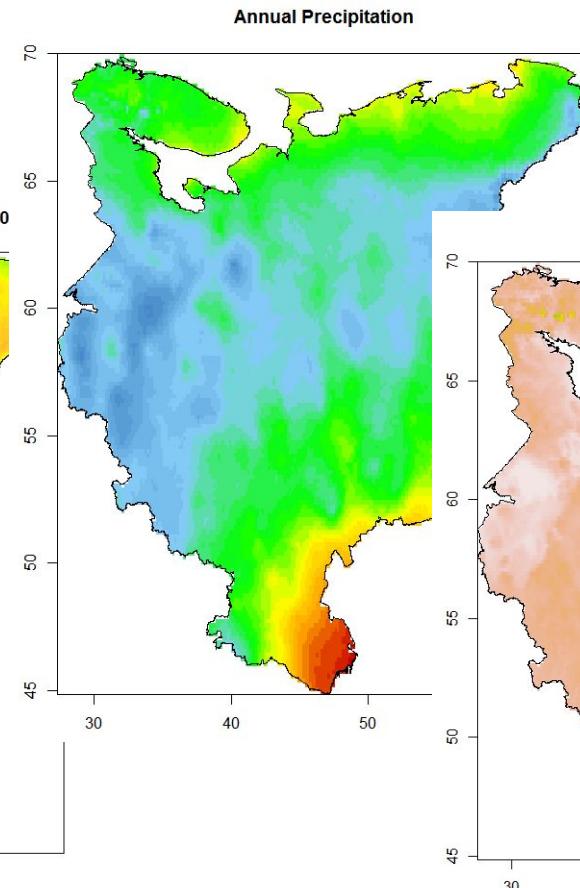
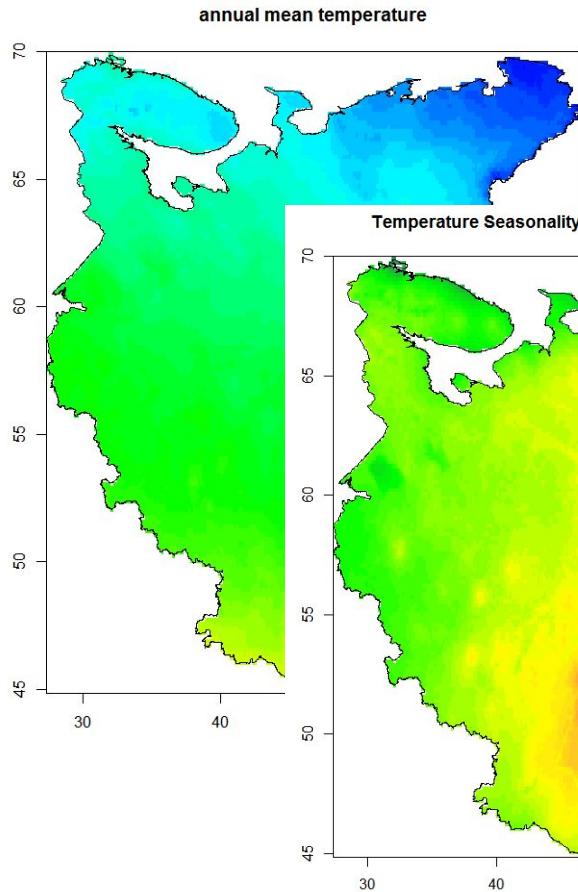
Точки отсутствия - Pseudo Absence



R Statial

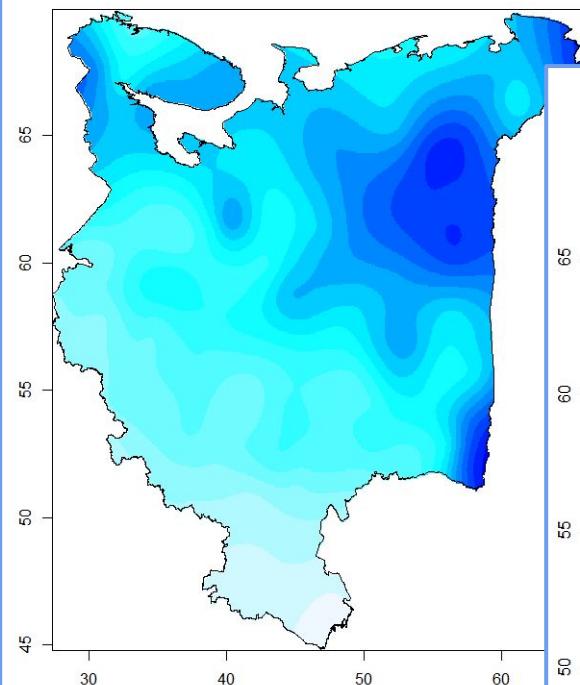
https://r-spatial.org/raster/sdm/3_sdm_absence-background.html

Набор факторов среды - WorldClim



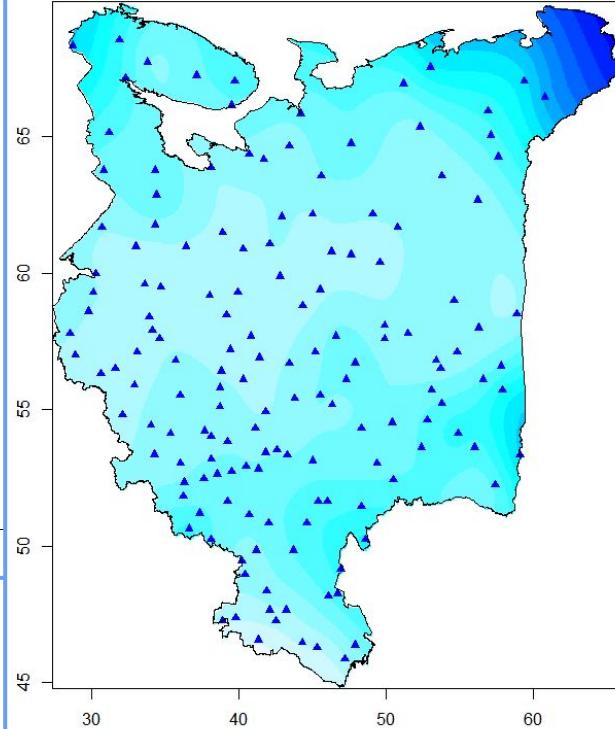
Набор факторов среды - почвенный климат

mean snow cover in January

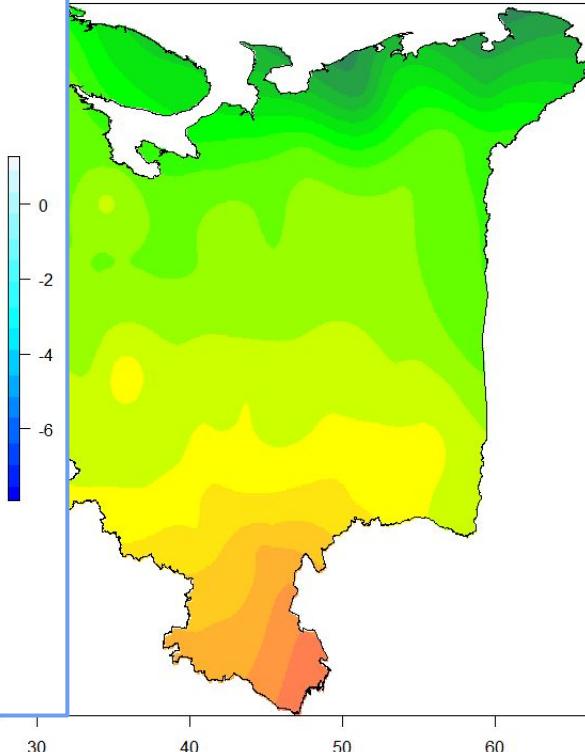


температура почвы на глубине 20 см

mean long-term t of January

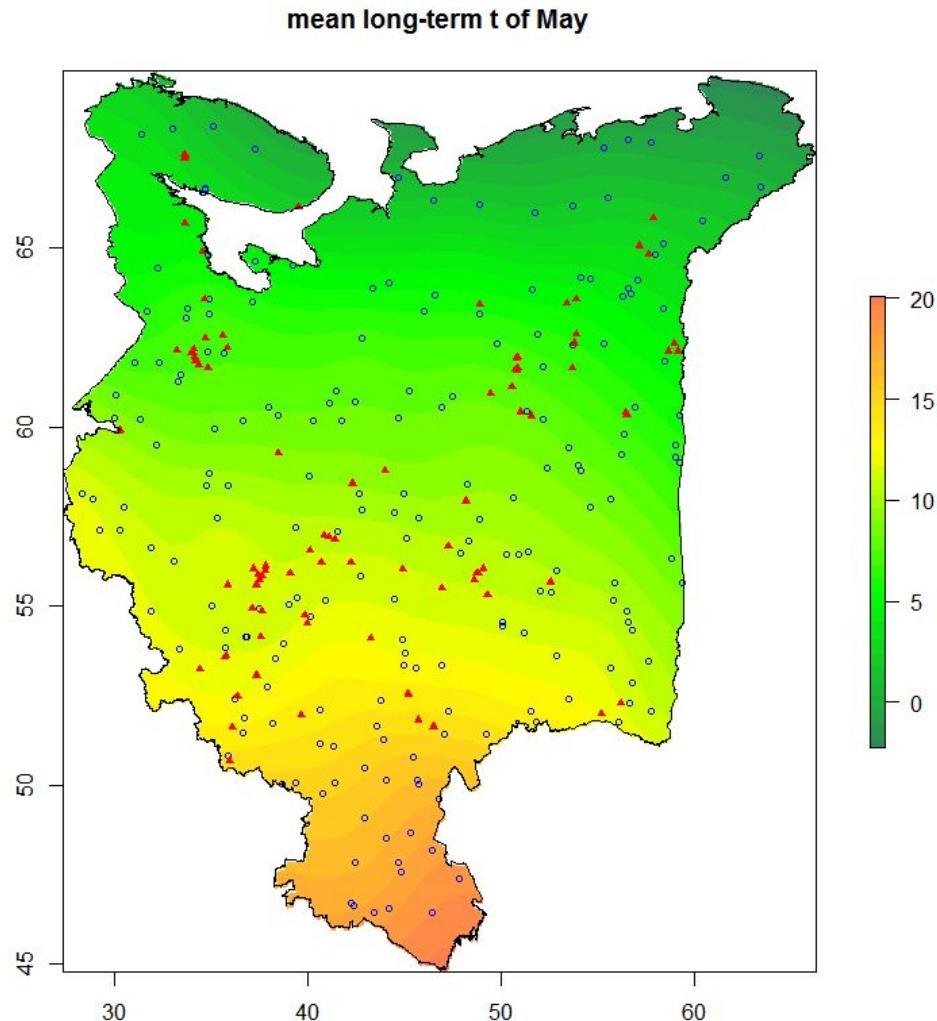


mean long-term t of June

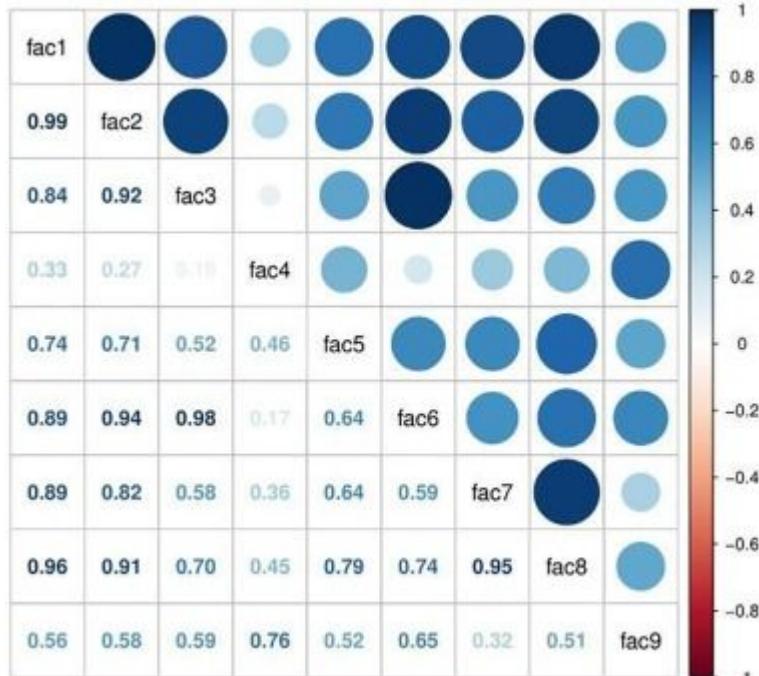


Объединение источников данных

из каждого предиктора
(растрового слоя)
берутся значения переменной
соответствующие координатам точек



Значения факторов среды - коллинеарность



Расчёт корреляции идёт по значениям переменных среды для тех точек в пространстве, для которых есть находки целевого вида

Из пары скоррелированных переменных одну убирают

обычно берут порог коэф. - 0.7

Dorman et al. 2013 Collinearity: a review of methods to deal with it and a simulation study evaluating their performance

doi: 10.1111/j.1600-0587.2012.07348.x

Выбор метода

Все моделируют MaxEnt'ом - и мы тоже будем

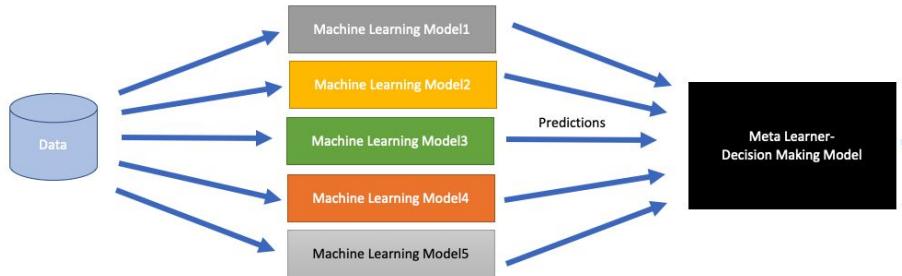
Можно моделировать несколькими и сравнивать результаты

Можно использовать ансамблевый подход

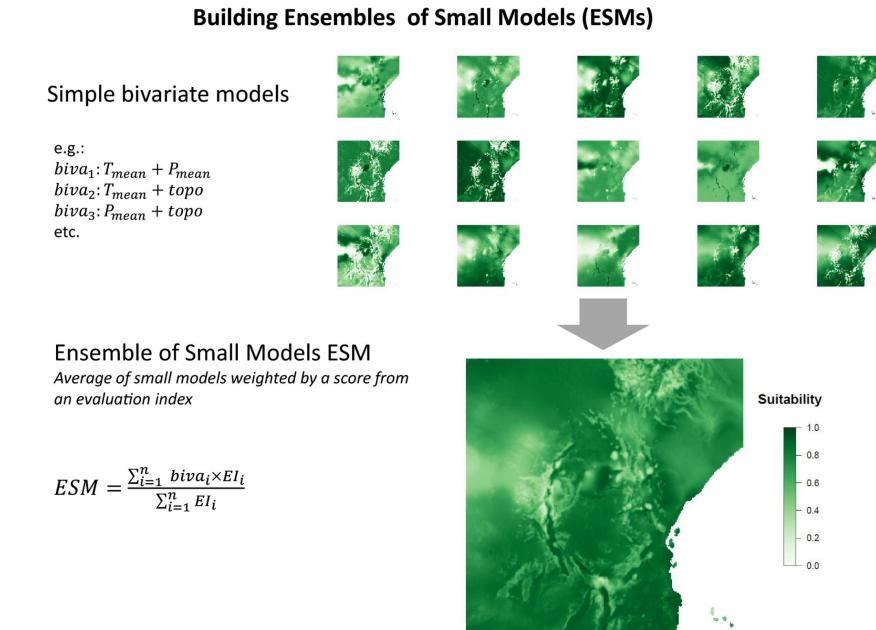
Для освоения технологий - начинайте с регрессионных моделей, затем переходите на MaxEnt

Ансамблевое моделирование

Разные методы



Разные наборы предикторов

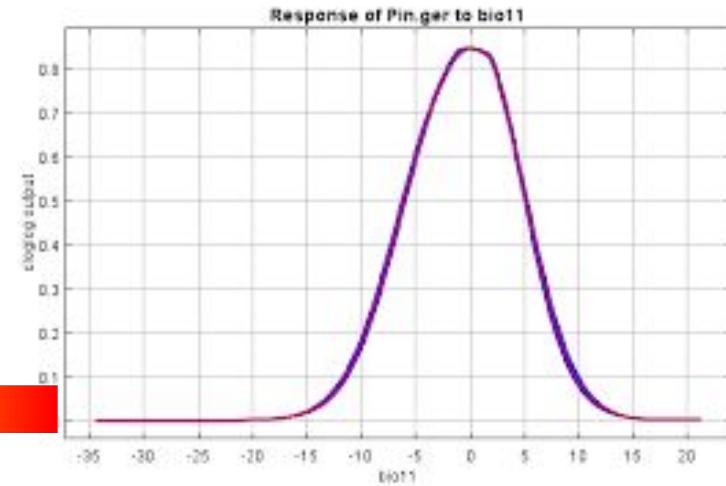
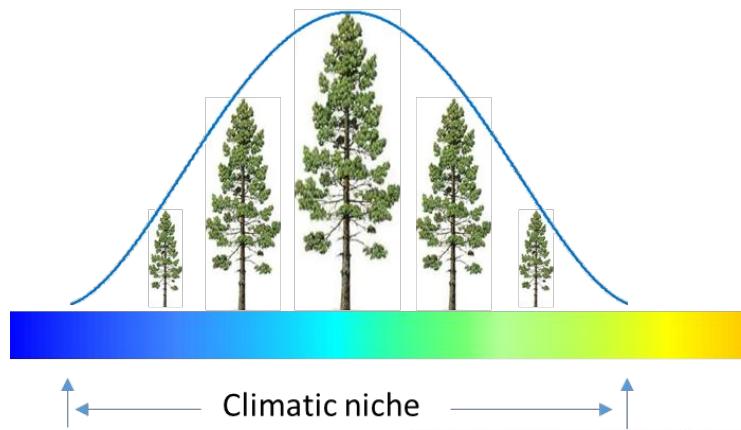


DOI: 10.1111/2041-210X.12957

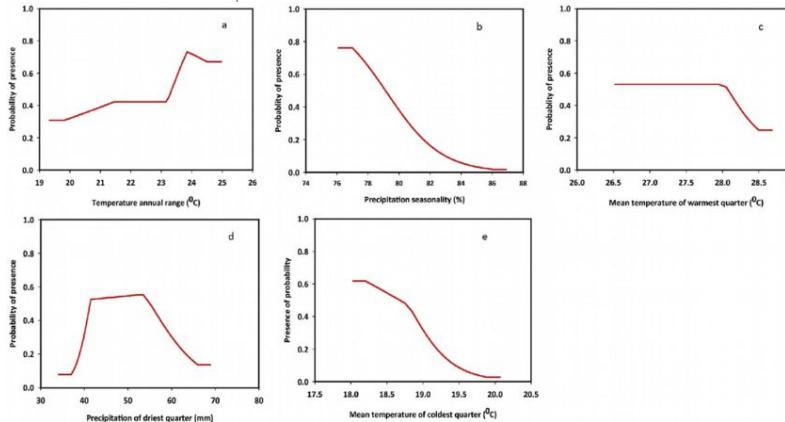
<https://onlinelibrary.wiley.com/doi/full/10.1111/ddi.12893>

Кривые отклика

Регрессионная модель



MaxEnt



Результат моделирования

на примере *Patrinia sibirica* (L.) Juss.

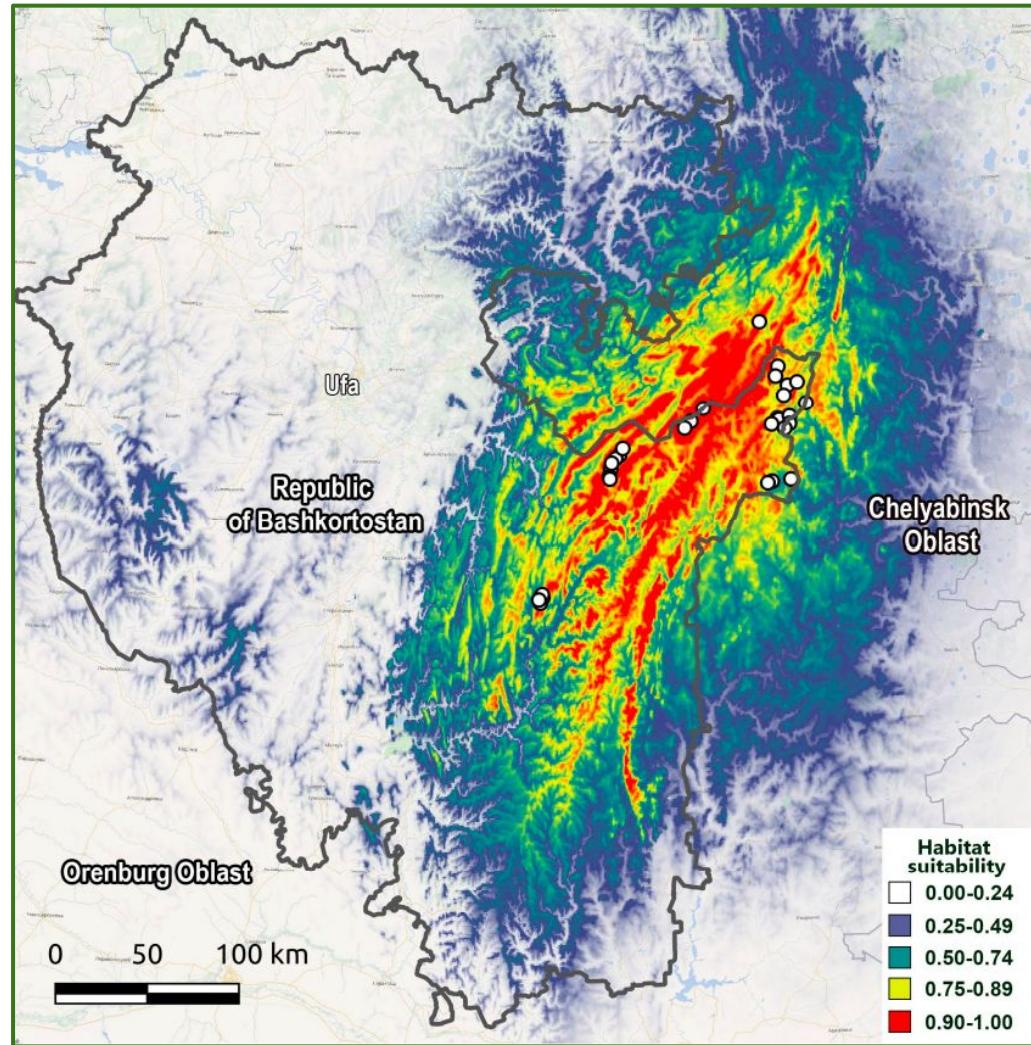
Kutueva et al., 2020

doi: 10.3897/ap.2.e57082

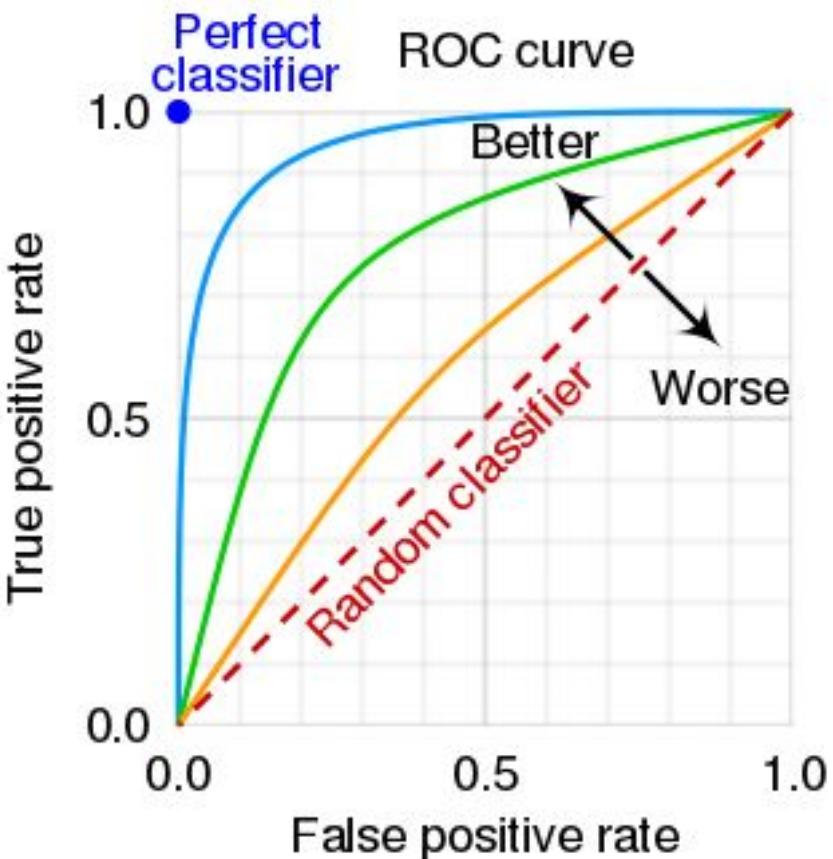


urgamal CC-BY

<https://www.inaturalist.org/observations/127434277>



Оценка качества модели



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

Чувствительность (Sensitivity)

доля истинно положительных случаев в общем числе положительных наблюдений

Специфичность (Specify)

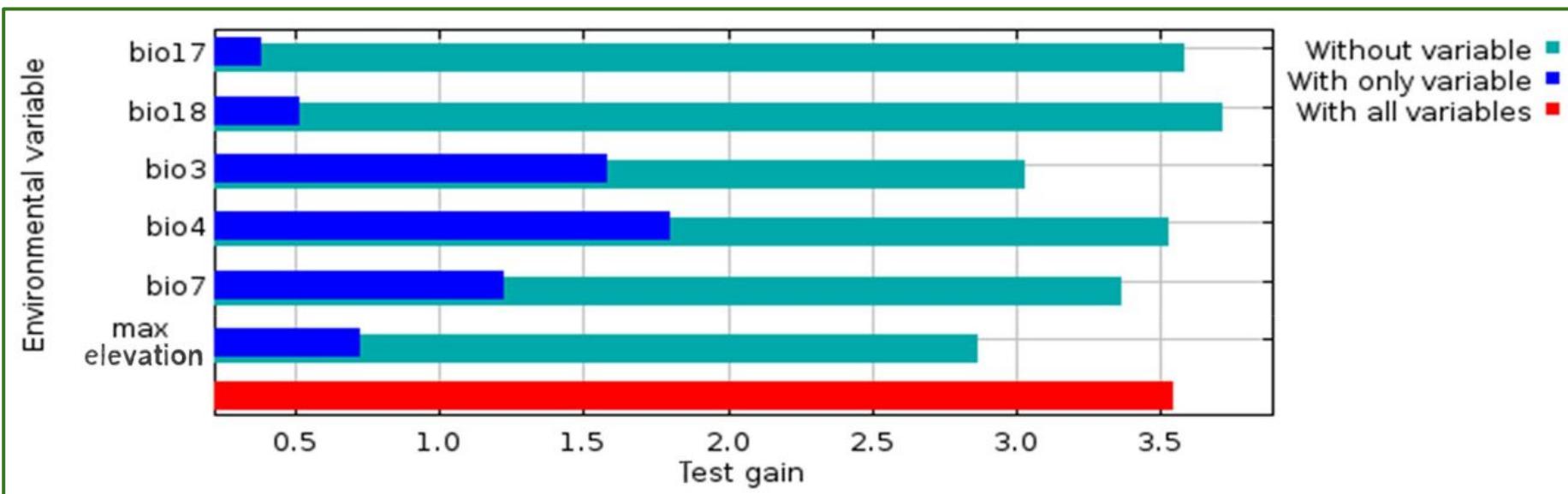
доля истинно-отрицательных случаев среди всех отрицательных наблюдений

Количественная оценка:

AUC - area under curve

Подробнее: <https://wiki.loginom.ru/articles/roc-analysis.html>

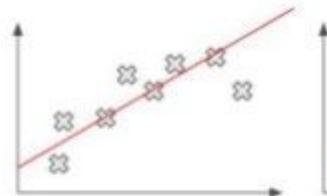
Вклад предикторов в распространение вида



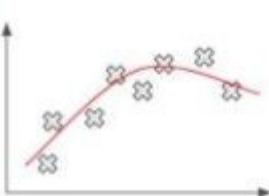
на примере *Patrinia sibirica*
Kutueva et al., 2020
doi: 10.3897/ap.2.e57082

Переобучение модели

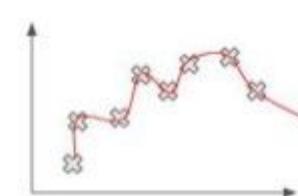
модель не должна быть излишне подробной



Underfitting

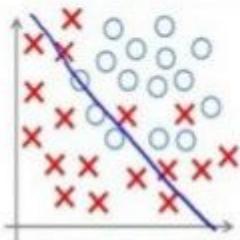


Optimal

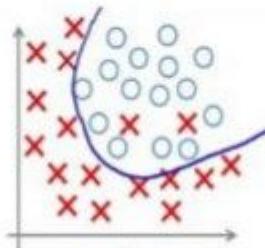


Overfitting

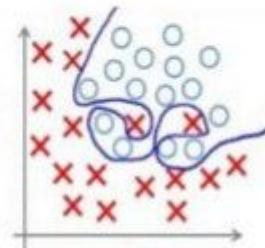
если модель объясняет каждую точку на входных данных, скорее всего она не будет



Under-fitting



Appropriate-fitting

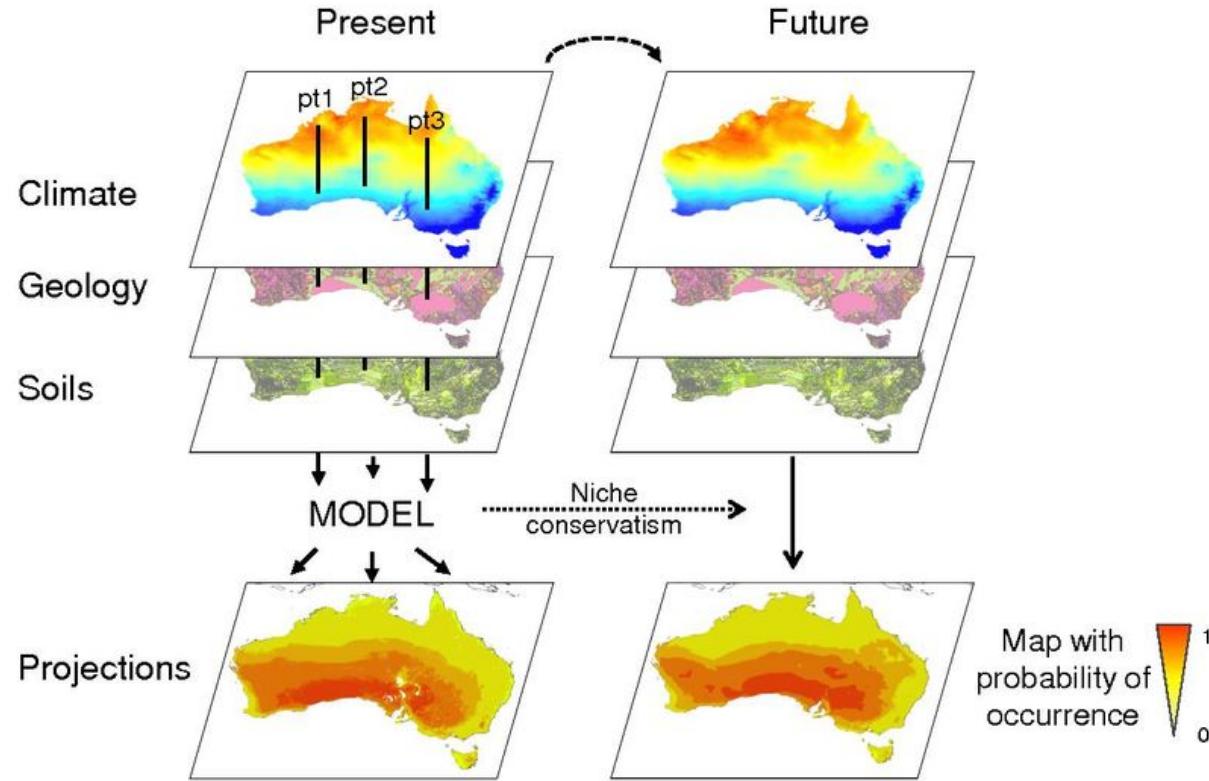


Over-fitting

Кросс-валидация

Model Projection

применение готовой
модели в изменяющихся
условиях



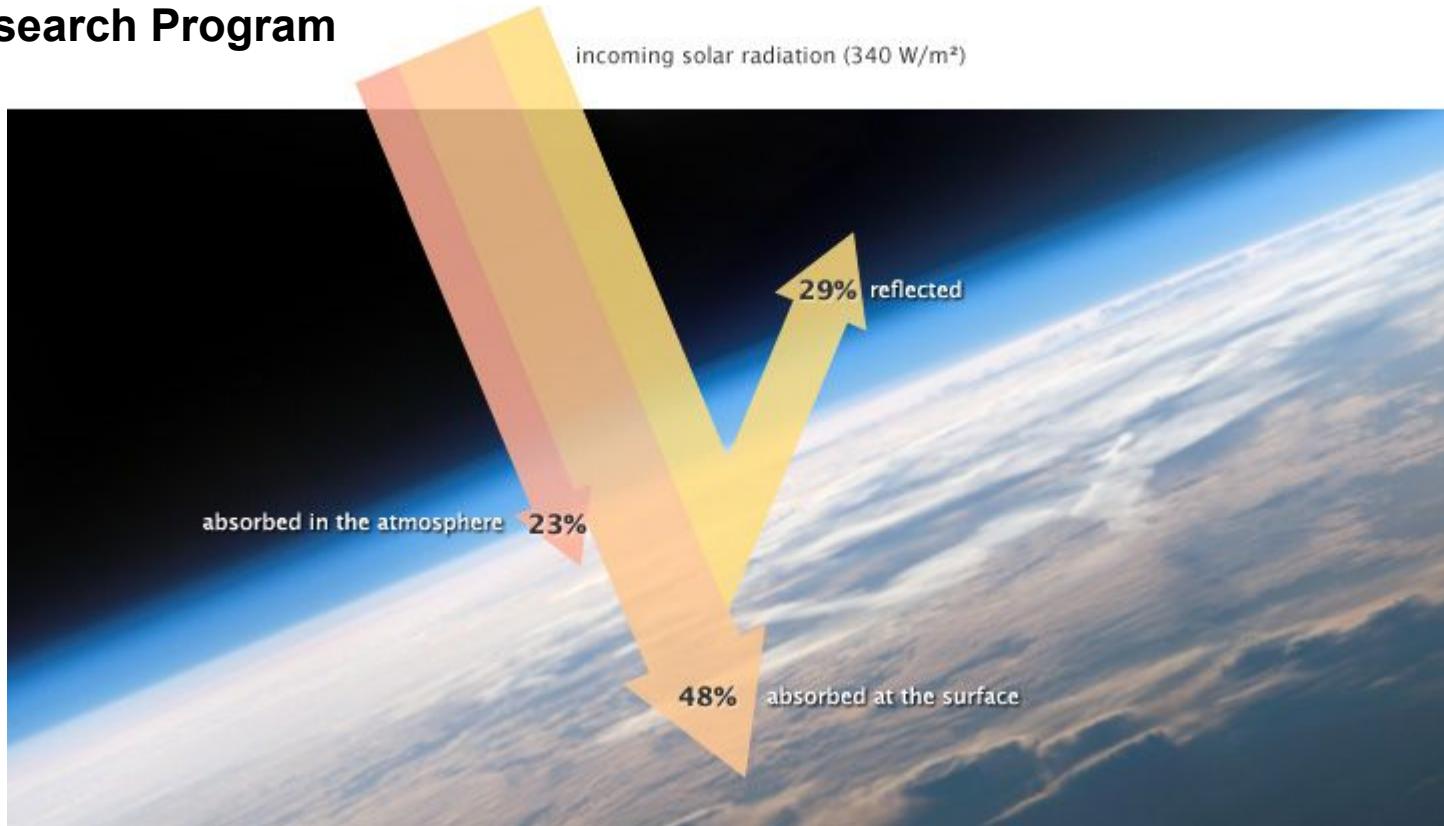
<https://conservationbytes.com/2020/07/21/history-of-species-distribution-models/>

Моделирование в изменяющемся климате

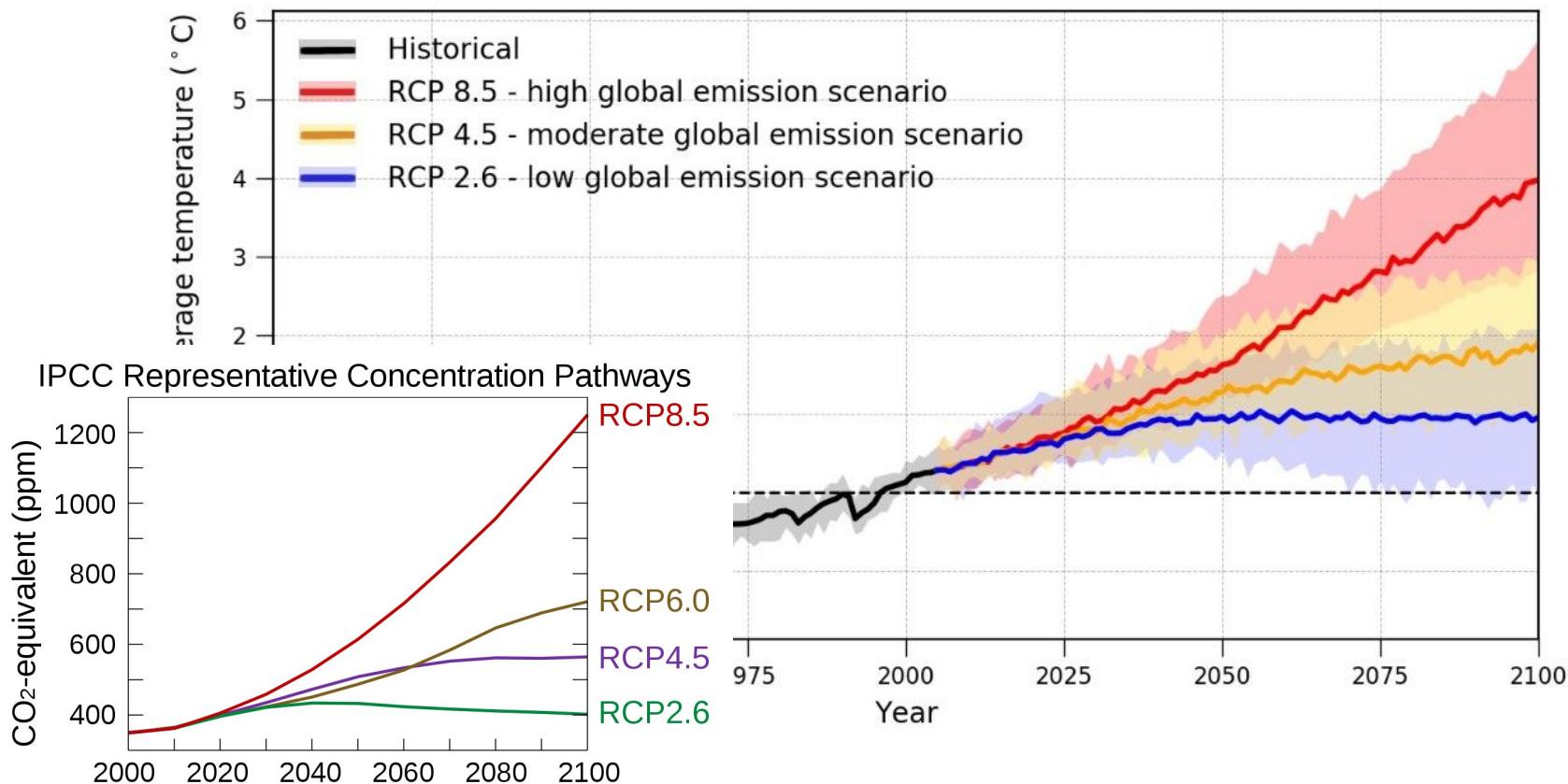
World Climate Research Program
1980

Coupled Model
Intercomparison Project
Phase 6

CMIP6

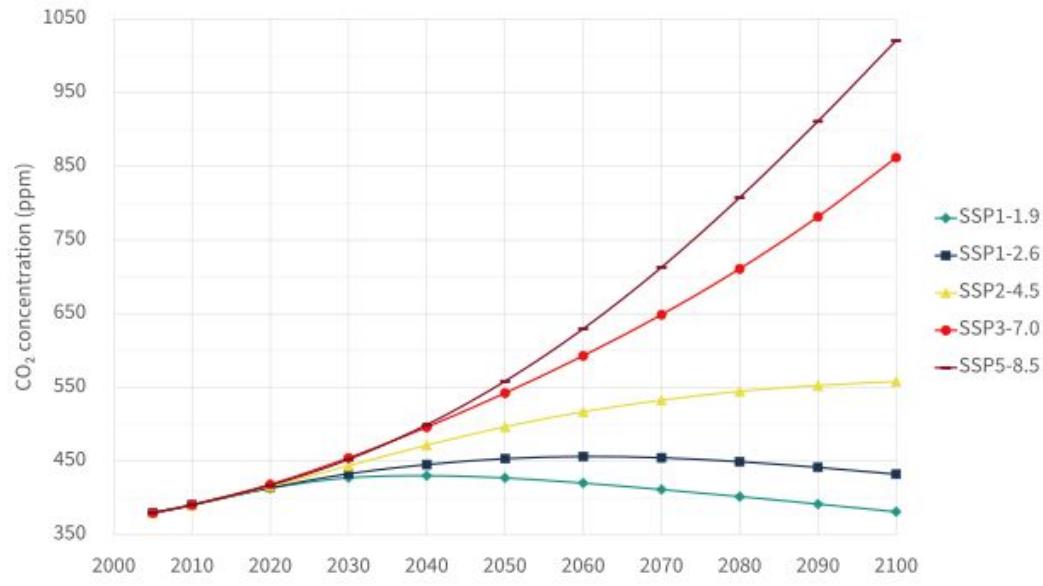


Сценарии изменения климата



IPCC Representative Concentration Pathway

Shared Socio-economic Pathways (SSPs)



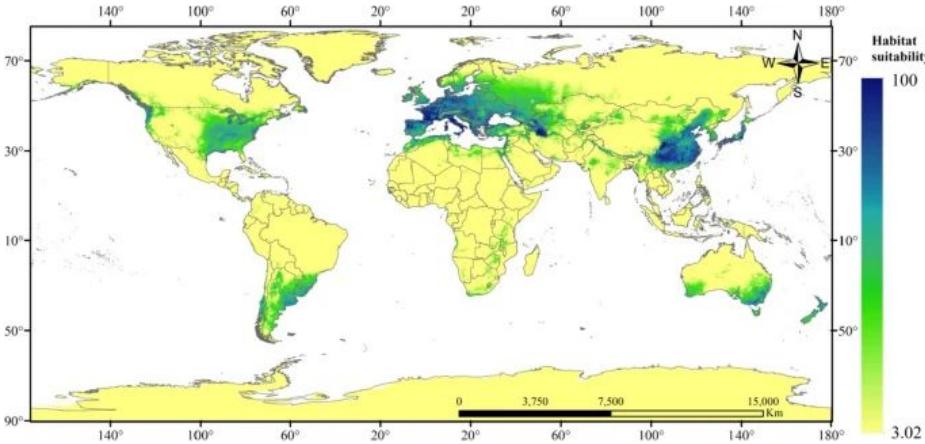
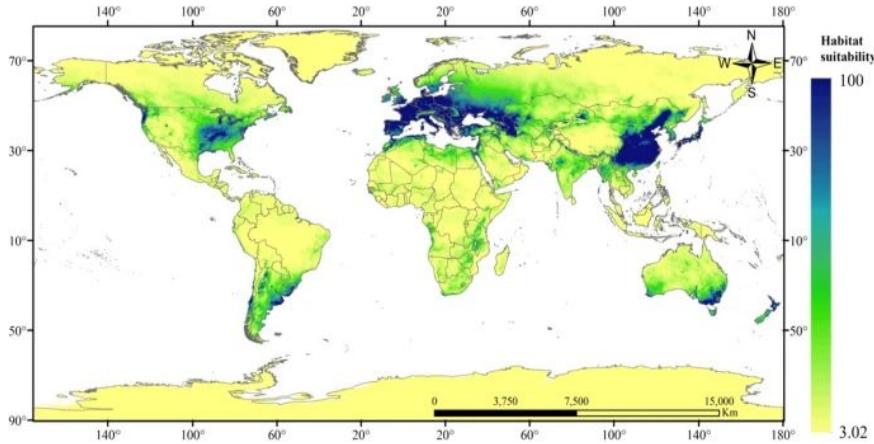
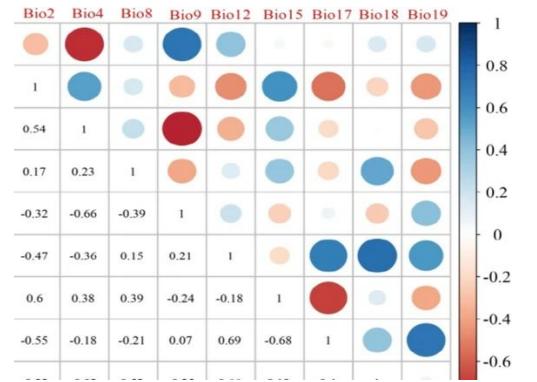
Моделирование в изменяющемся климате

Zhang et. al (2021) Global potential distribution prediction of *Xanthium italicum* based on Maxent model

336 находок из GBIF

WorldClim пространственное
разрешение 5 км

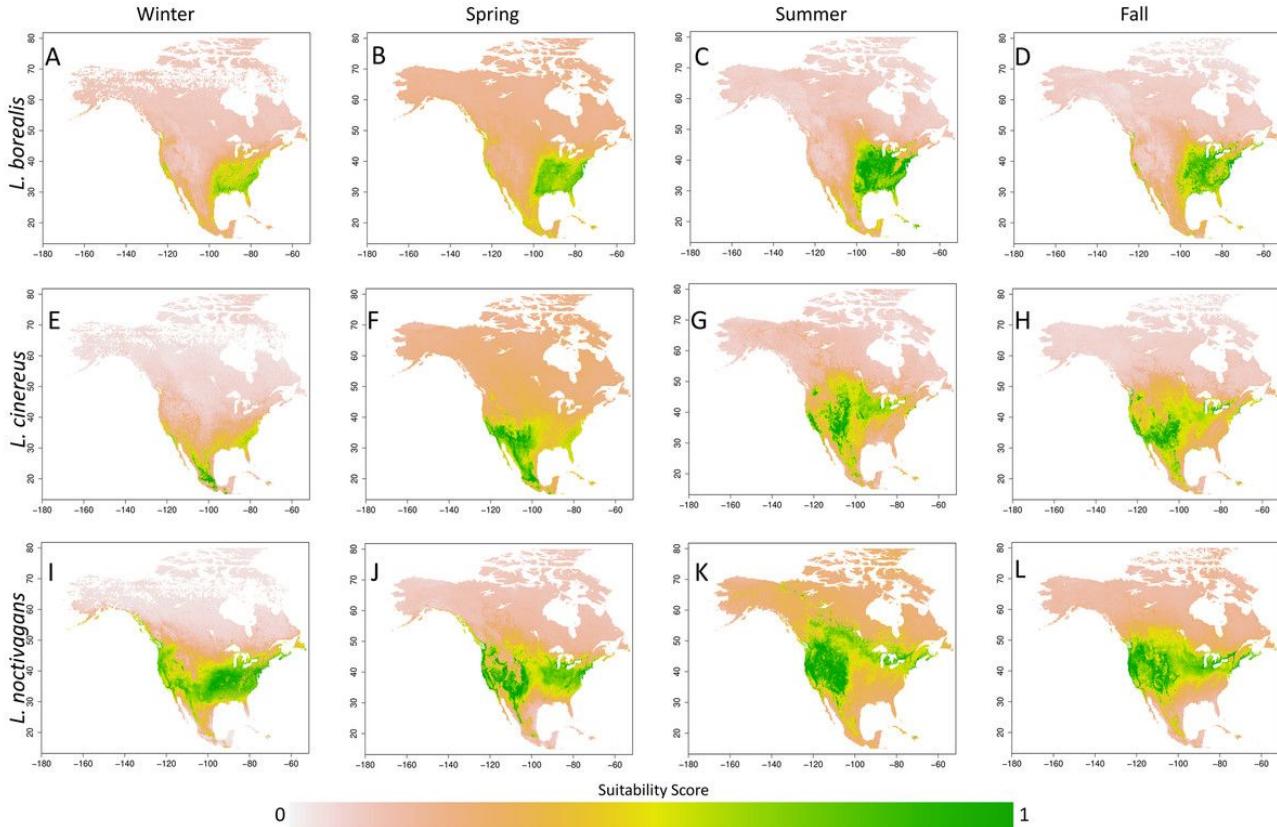
Сценарии изменения климата SSP



Моделирование сезонных миграций

Сведения о находках и
предикторы группируются
по месяцам или сезонам

для каждого строится
своя модель



Wieringa et al. 2021 Predicting migration routes for three species of migratory
bats using species distribution models DOI: [10.7717/peerj.11177](https://doi.org/10.7717/peerj.11177)

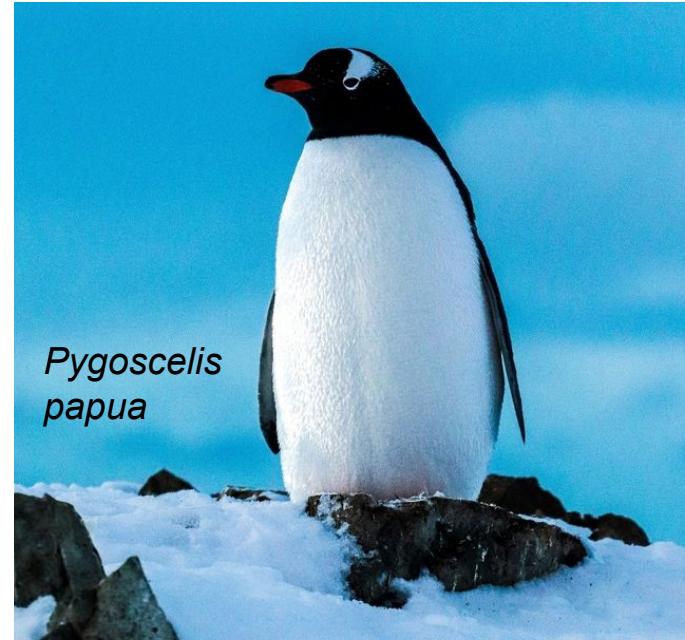
Интерпретация результатов

В любом случае SDM - это еще один инструмент анализа и моделирования
Последнее слово по интерпретации данных остаётся за экспертом.



aleksey_loginov CC-BY

<https://www.inaturalist.org/observations/67694709>



vasyliuk1 CC-BY

<https://www.inaturalist.org/observations/134769105>

Необходимо учитывать весь комплекс факторов

Лес шагает в горы: фотомониторинг изменения климата



1929 г.



2012 г.

Институт экологии растений и животных УрО РАН, Уральский государственный лесотехнический университет

<https://citizen-science.ru/projects/les-idet-v-gory-fotomonitoring-izmeneniya-klimata.html>

Modelling (SDM) - литература

WILEY

1998, 5.717



Springer 2005, 3.906

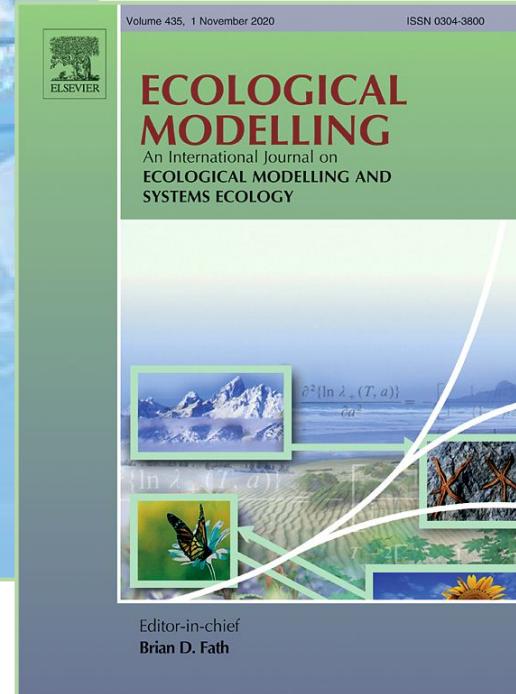
Biological
Invasions

Springer



ELSEVIER

1975, 3.512



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)

Полезные статьи

Санданов Д.В. 2019. Современные подходы к моделированию разнообразия и пространственному распределению видов растений: перспективы их применения в России DOI: 10.17223/19988591/46/5

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

Руководства и дополнительные материалы

Robert J. Hijmans and Jane Elith

Species distribution modeling

(R package raster)

A Brief Tutorial of MaxEnt

By Steven J. Phillips,

https://biodiversityinformatics.amnh.org/open_source/maxent/Maxent_tutorial2017.pdf

Biodiversity Climate Change Virtual Laboratory

<https://support.bccvl.org.au/support/solutions/articles/6000083199-overview-of-sdm-methods-in-bccvl>

EcoCommons

<https://www.ecocommons.org.au/educational-material/>

EcoCloud

<https://ecocloud.org.au/>

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 cloud analyses hosted externally to 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 data 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 that 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

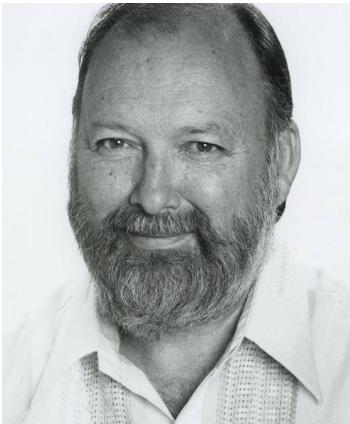
306

Environmental Data Layers

58

Average model runtime (sec)

Personalia



[Henry Nix](#)
1937-2022



Jane Elith



Steven J. Phillips



Trevor H. Booth



Damaris Zurell

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 !!!