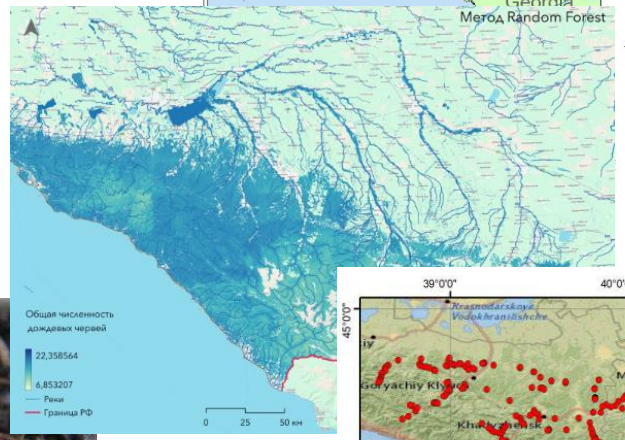
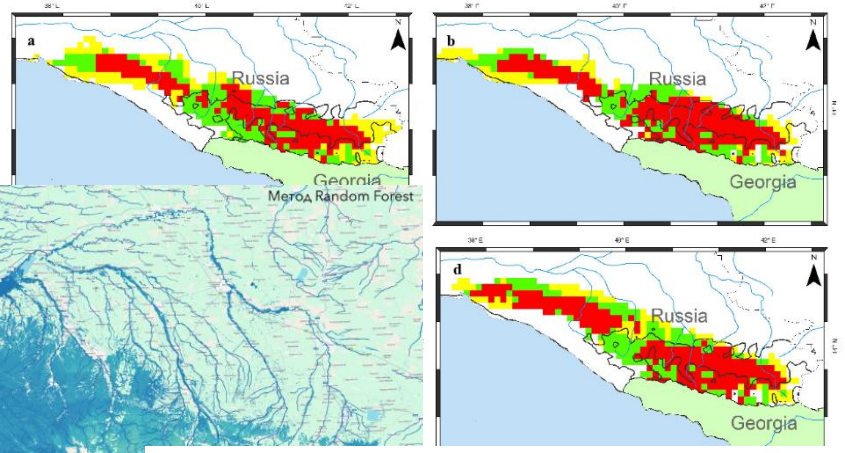
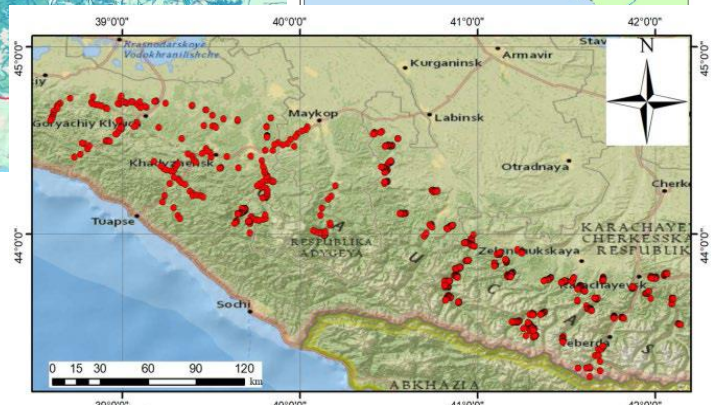
ВШЭ
24 сентября 2025 года

Экосистемные инженеры лесных почв: от полевых исследований к моделям машинного обучения

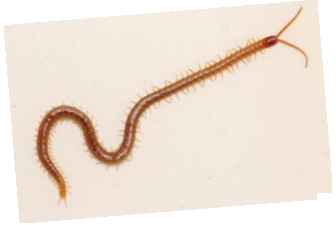
Гераськина Анна

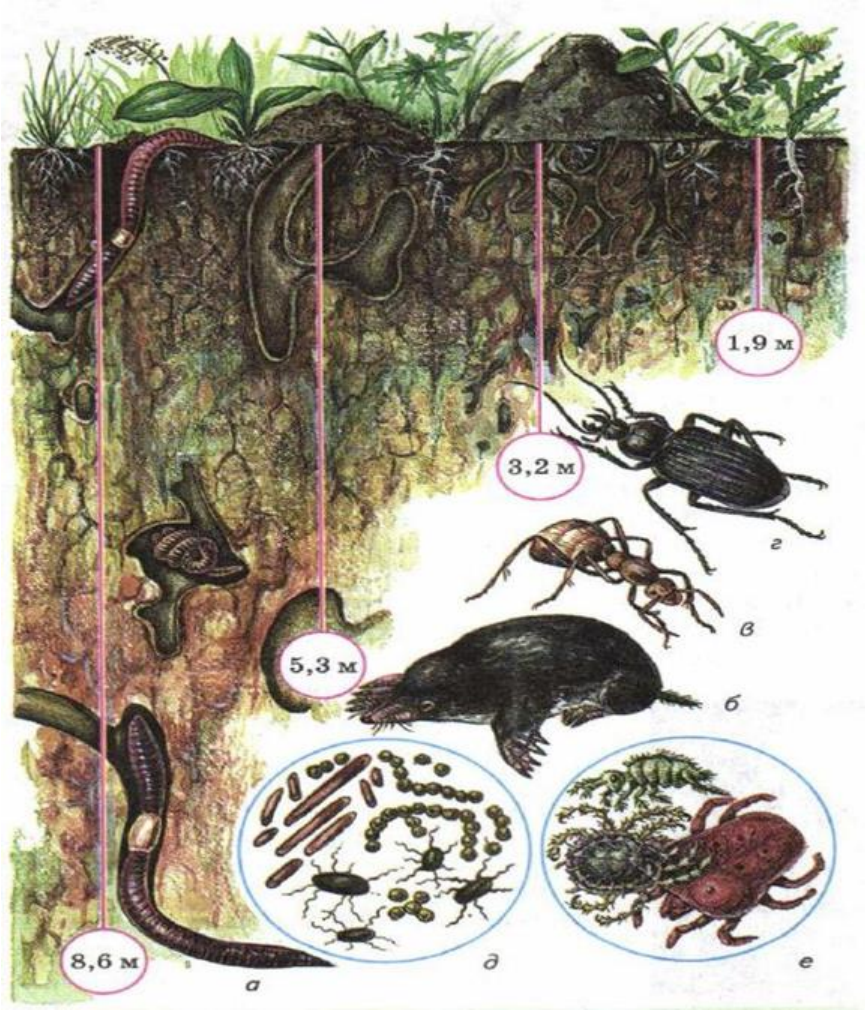






Разнообразие почвенных животных





Коллемболы

Навозный червь – 8 - 12 см, кирпично-красного цвета с неяркими желтыми полосками

Пашенный червь – 8 - 15 см, окраска серая, бледная

Каштановый червь – темно-красный, длина 3-4 см.

Розовая эйзения – 3 - 8 см, окраска серовато-розовая. Иногда окрашен поясок

Сколько весит почвенная фауна?

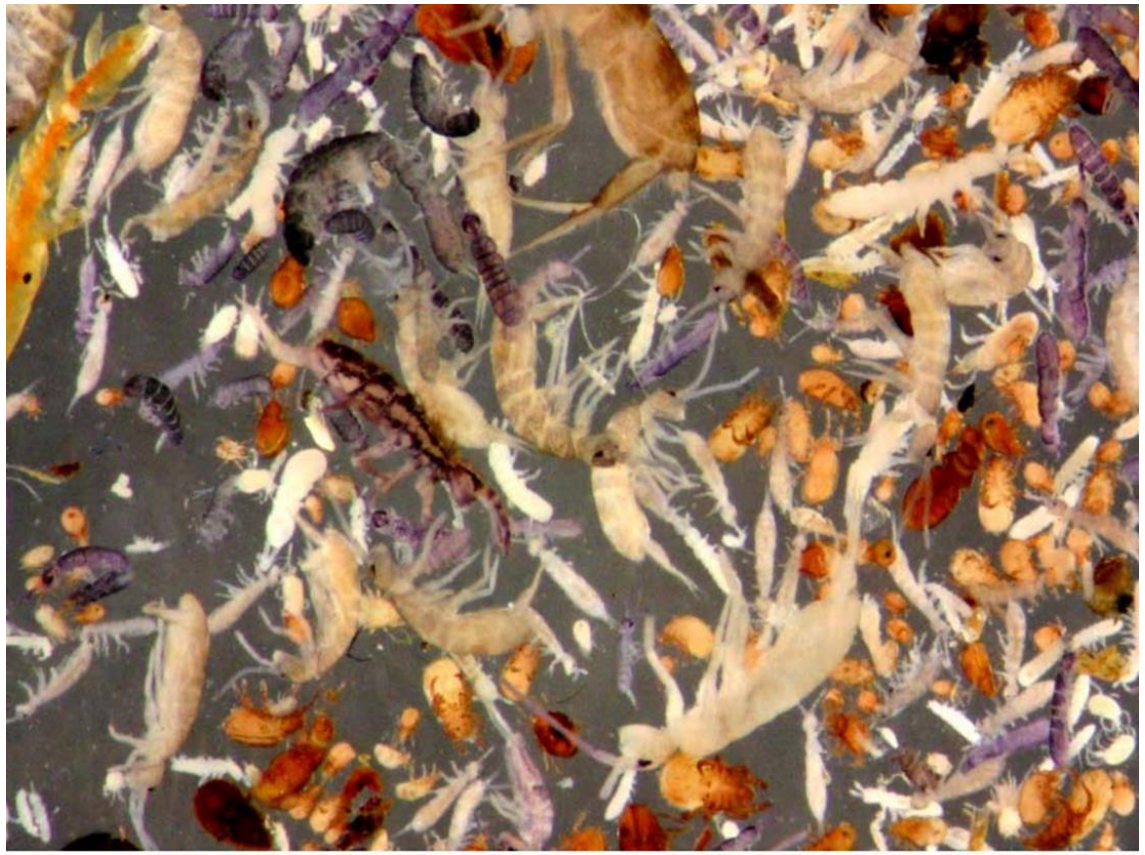
В умеренном климатическом поясе **масса животных в почве:**

- насекомые – до 1 т/га,
- дождевые черви – 500-1000 кг/га,
- нематоды – 50 кг/га,
- ракообразные – 40 кг/га,
- моллюски – 20 кг/га,
- змеи, грызуны – 20 кг/га

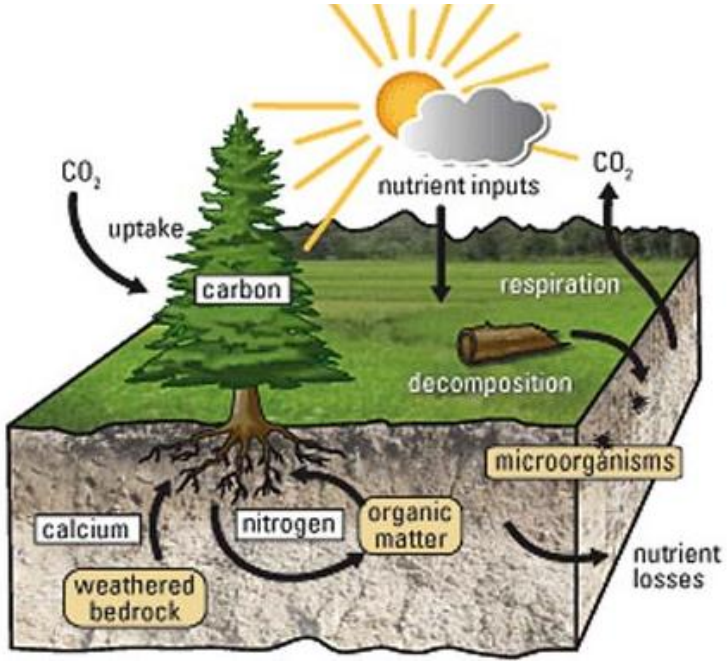
>

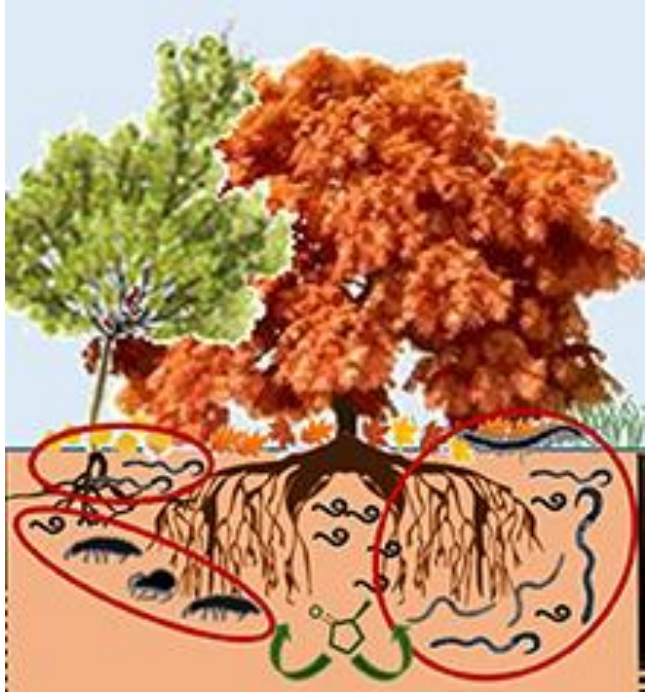
Для сравнения: масса всех **наземных позвоночных** животных составляет **100-200 кг/га**

Зачем такое разнообразие?



Круговорот веществ





Экосистемные функции

- ✓ **Минерализация органического вещества**
- ✓ **Аккумуляция органического вещества**
- ✓ **Гумификация**
- ✓ **Регуляция водного режима**
- ✓ **Регуляция циклов азота и углерода**
- ✓ **Санитарная роль**
- ✓ **Смягчение климатических изменений**

Сапрофаги

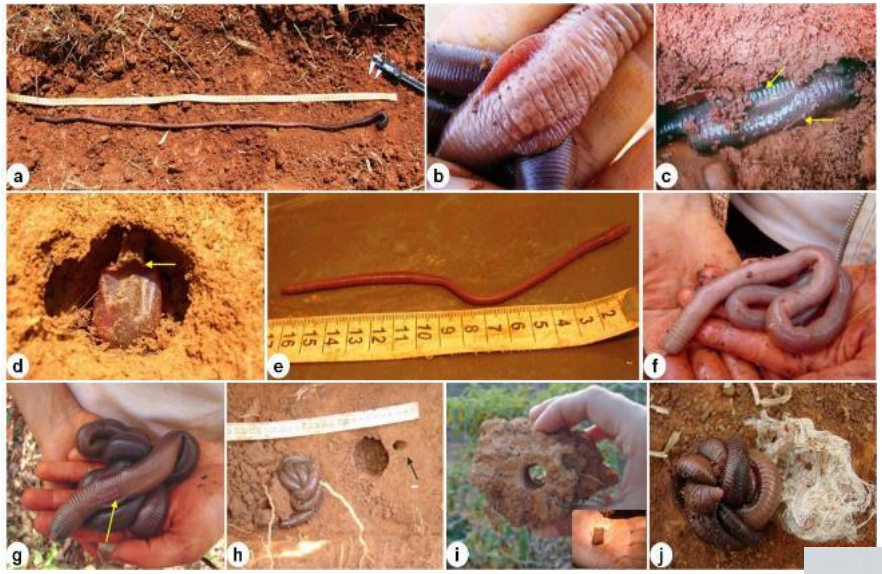
Дождевой червь

Кивсяк Криволуцкого

За год активности, один 15-20-сантиметровый белый кивсяк съедает до пяти ведер гниющих растительных остатков (примерно 10 литров в месяц).

Экосистемные инженеры

Экосистемные инженеры («ecosystem engineer») – организмы, которые прямо или косвенно регулируют доступность ресурсов для других видов, вызывая изменения физического состояния биотических или абиотических компонентов (Jones et al., 1994).



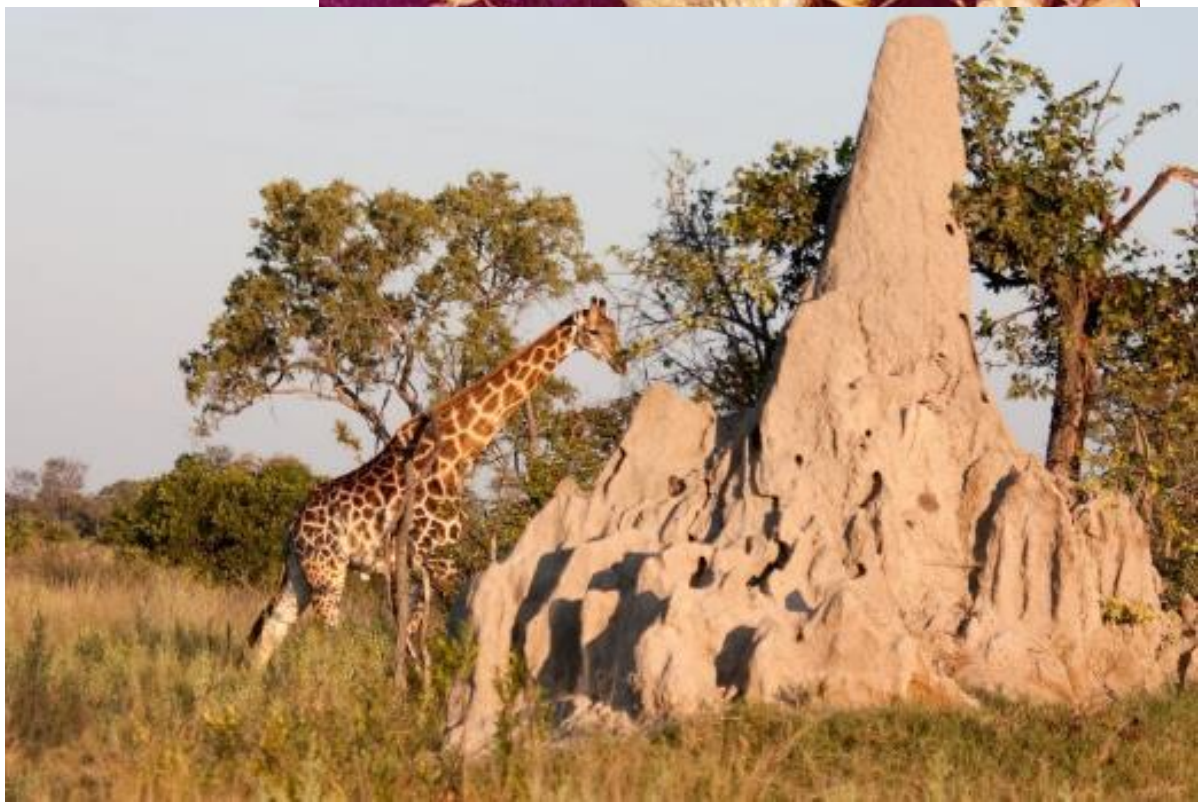
Муравьи

Термиты

Термиты – общественные насекомые с неполным превращением, инфраотряд в составе **тараканообразных**.

Как и все общественные насекомые, термиты живут в колониях, число зрелых особей в которых может достигать от нескольких сотен до нескольких миллионов и состоящих из каст.

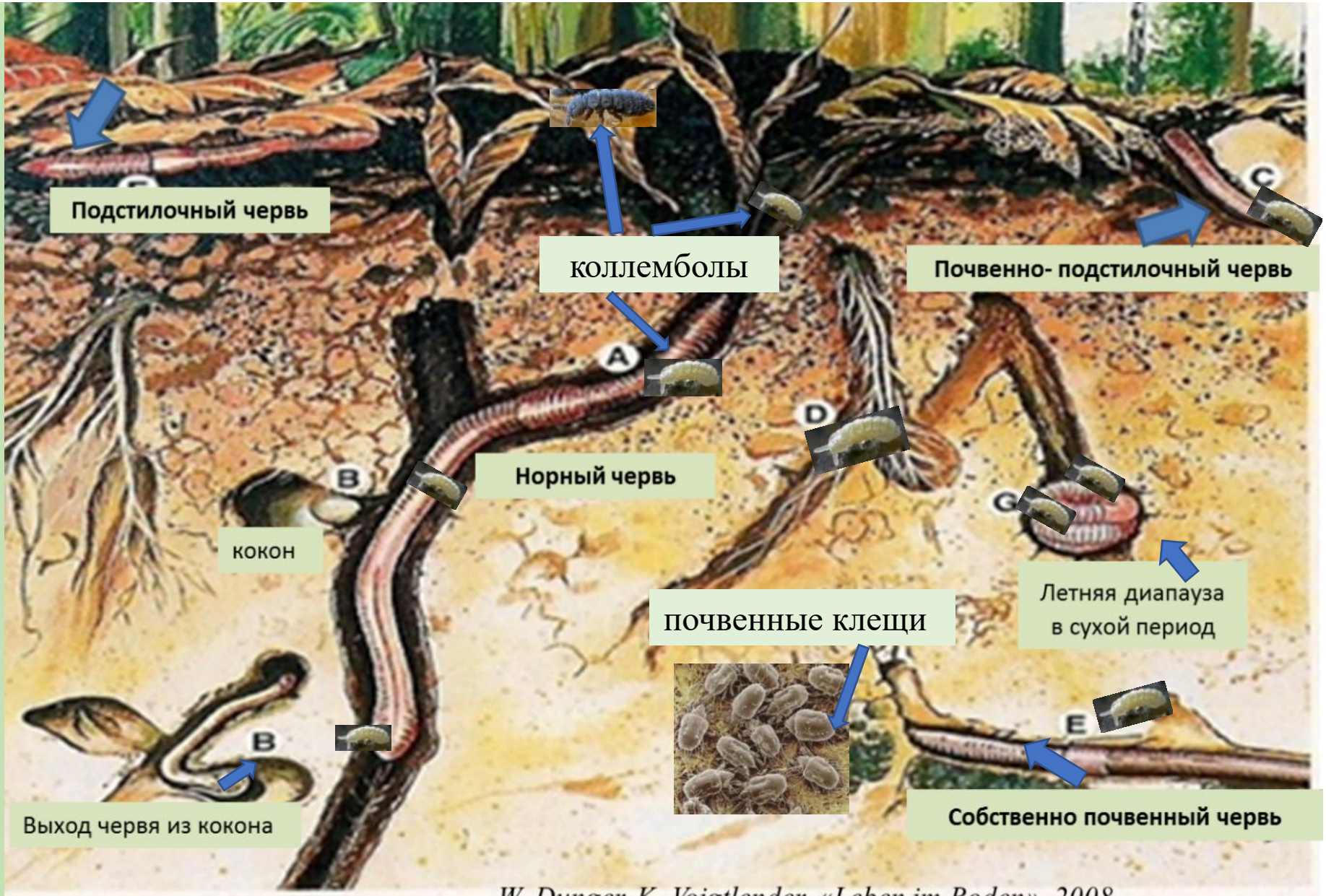
Термиты важны для почв **тропиков**. Термиты из частичек почвы строят термитники таких размеров, что в них может укрыться жираф или слон. В засушливых местах термиты способны на несколько десятков метров проделывать ходы в глубь земли, чтобы достать грунтовые воды. Вместе с водой они извлекают необходимые для питания химические элементы и, доставая их на поверхность почв, увеличивают плодородие.



Дождевые черви

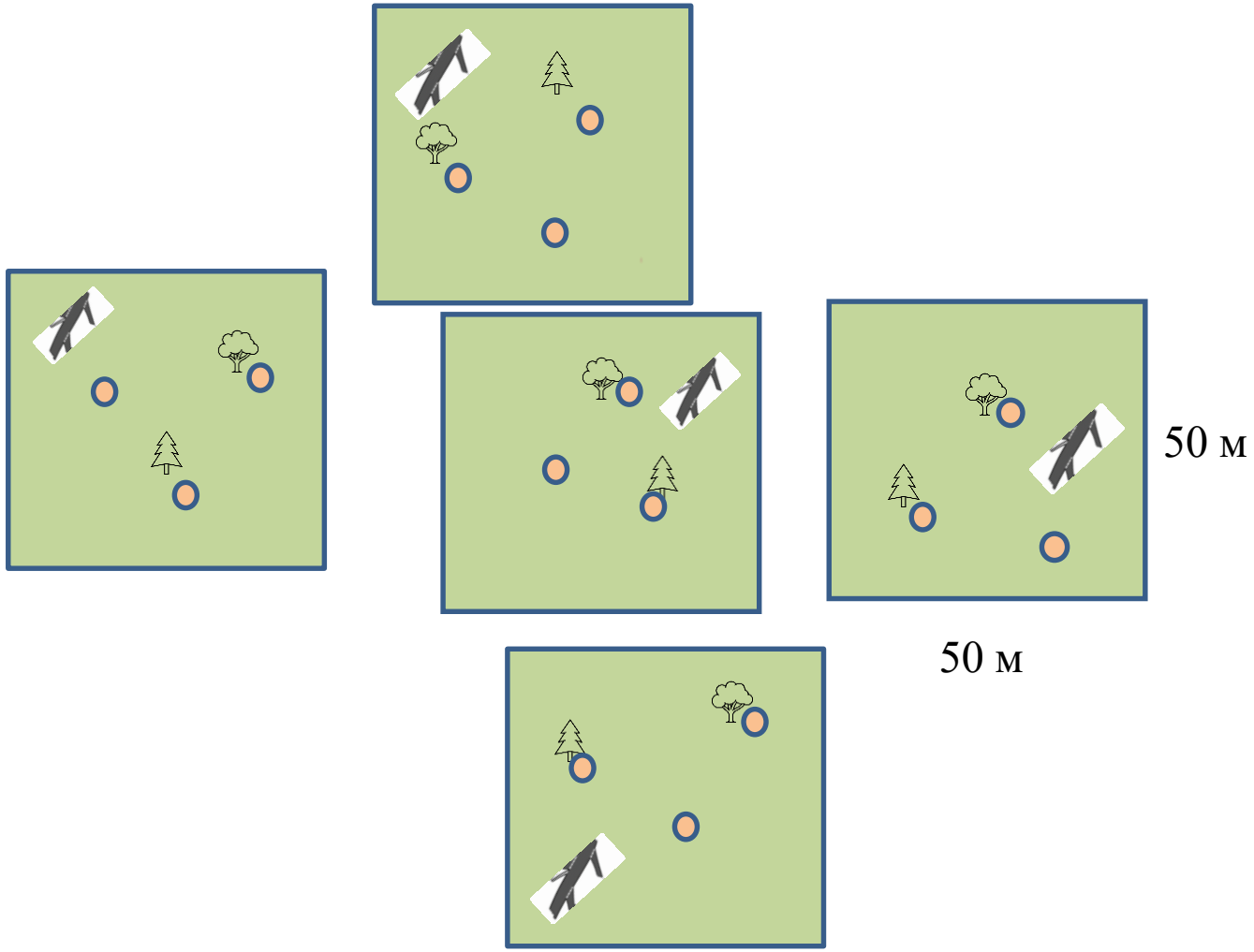
«Биота почвы взаимодействует как большой оркестр, где музыканты из почвы играют вместе на разных инструментах, а дирижеры оркестра - дождевые черви, которые прокладывают почвенные ходы, активируют почвенные бактерии, разносят мицелий и споры грибов, стимулирует рост популяций простейших и нематод»

(Cunha et al., 2016; Eisenhauer, 2012).



W. Dunger, K. Voigtlender. «Leben im Boden», 2008

Пример дизайна отбора почвенно-зоологических проб



Почвенно-зоологическая проба (25х25 см, глубина 30 см)

Фрагмент валежа (диаметр не менее 10 см, длина 30 см)

ЭТАПЫ РАБОТЫ

1

Отбор почвенно-зоологических проб (25х25х30 см)

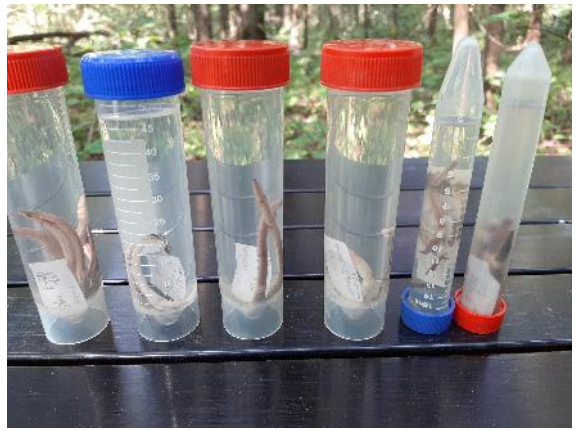
2

Ручной разбор почвенных проб

3

Ручной разбор валежа

4

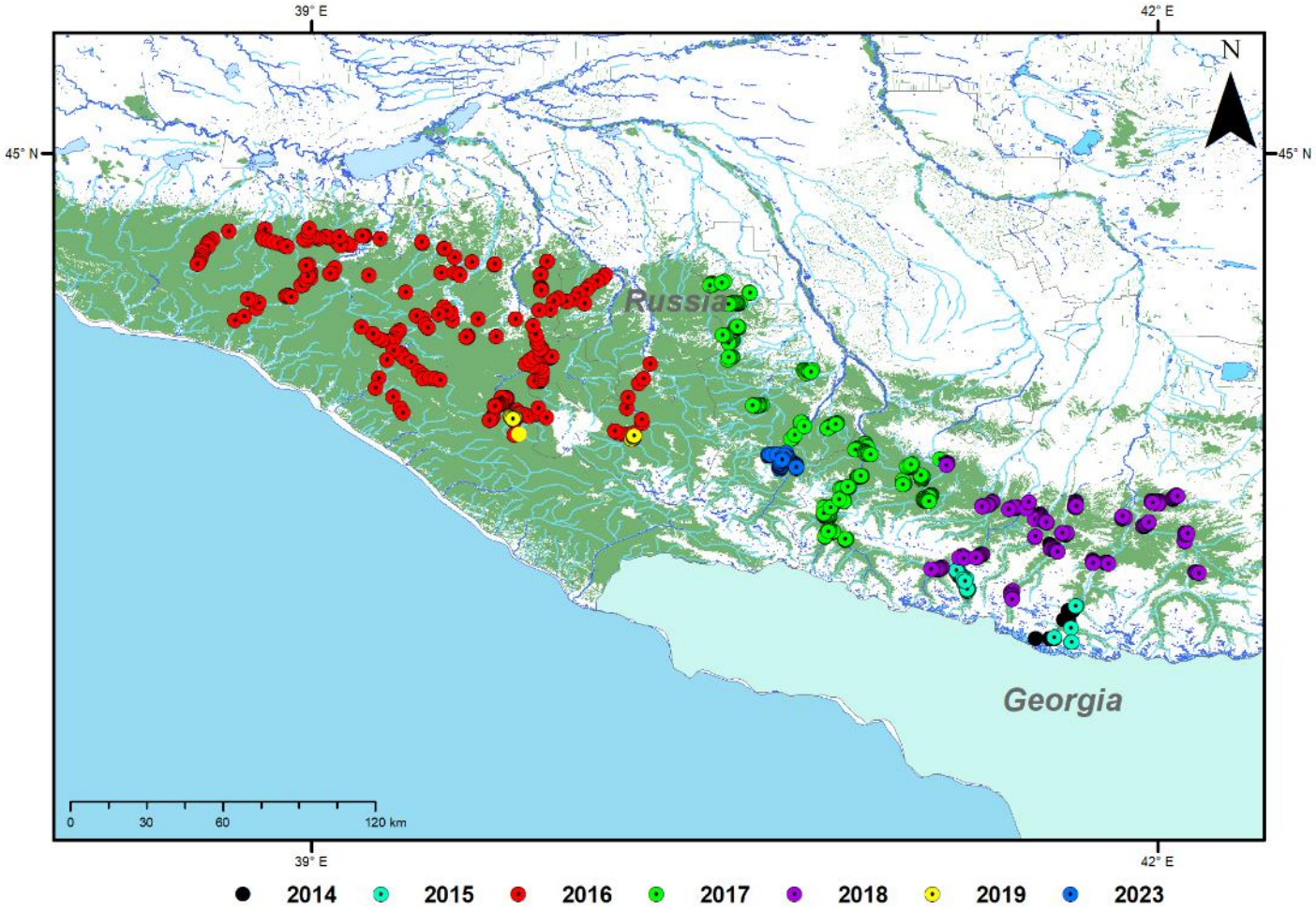


Фиксация материала

5

Идентификация видов, групп и определение биомассы

Исследования дождевых червей на Северо-Западном Кавказе (Гераськина, Шевченко)



Map of points (N. Shevchenko)

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2024623421

Видовой состав, численность и биомасса дождевых червей Северо-Западного Кавказа

Правообладатель: *Федеральное государственное бюджетное учреждение науки Центр по проблемам экологии и продуктивности лесов Российской академии наук (RU)*

Авторы: *Гераськина Анна Петровна (RU), Шевченко Николай Евгеньевич (RU)*

Заявка № 2024622949

Дата поступления 09 июля 2024 г.

Дата государственной регистрации в Реестре баз данных *06 августа 2024 г.*

Руководитель Федеральной службы по интеллектуальной собственности

Ю.С. Зубов

Видовой состав, численность и биомасса дождевых червей Северо-Западного Кавказа

В рамках исследования биологического разнообразия Северо-Западного Кавказа разработана база данных «Видовой состав, численность и биомасса дождевых червей Северо-Западного Кавказа». База данных содержит информацию о точках отбора почвенно-зоологических проб (**1044 точки**), характеристиках растительности, почвы, видовом составе дождевых червей, их численности и биомассе. Количественные показатели (численность, биомасса) приведены для 17-ти видов дождевых червей (один из которых представлен тремя формами), а также для 4-х морфо-экологических групп (подстилочные, почвенно-подстилочные, собственно почвенные, норные) и 5-ти хорологических групп (космополиты, крымско-кавказские эндемики, средиземноморские, восточноевро-азиатские, палеарктические). Данные собраны в весенне-летние сезоны с 2016 по 2019 гг.

Морфо-экологические группы дождевых червей

Epigeic

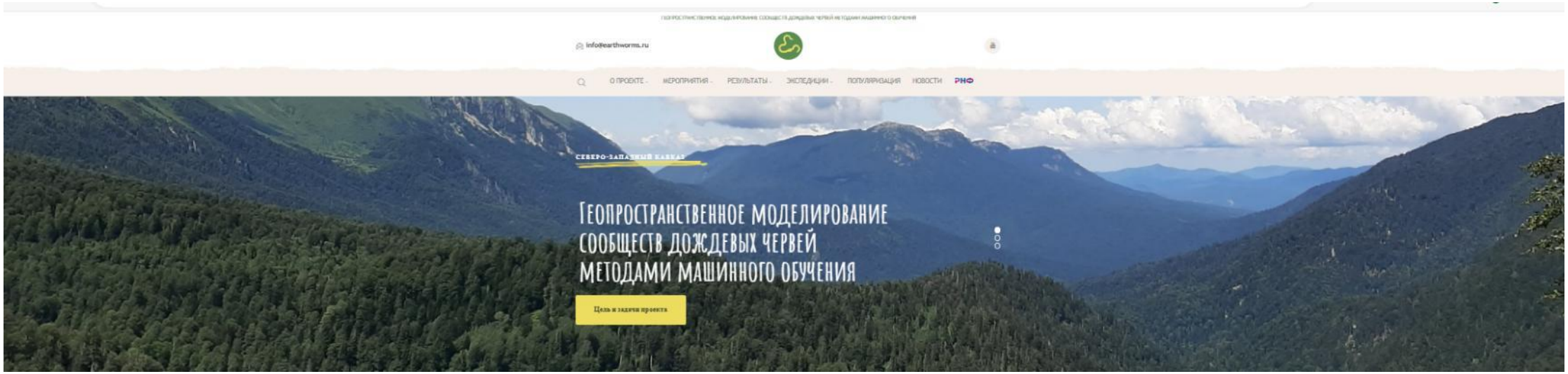
Endogeic

Epi-endogeic

Anecic

ГЕОПРОСТРАНСТВЕННОЕ МОДЕЛИРОВАНИЕ СООБЩЕСТВ ДОЖДЕВЫХ ЧЕРВЕЙ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ (2024-2025) № 23-24-00543.

Гераськина А.П., Плотникова А.С., Нарыкова А.Н., Шевченко Н.Е., Архипцева Е.А. (ЦЭПЛ РАН)



Сайт проекта:
https://earthworms.ru/

От поля к модели

Из точки в пространство

Моделирование методами машинного обучения даёт возможность решать сложные задачи и делать прогнозы на основе больших объёмов данных. Вместо того чтобы следовать строго запрограммированным инструкциям, алгоритмы учатся выявлять закономерности, прогнозировать результаты и адаптироваться к изменениям. Цель машинного обучения - увеличить точность решений и скорость их принятия.

Методы машинного обучения

Naive Bayes

Random Forest

GradientTreeBoost

Maxent

Random Forest, NaiveBayes, GradientTreeBoost

Для проведения регрессионного моделирования используется 21 пространственный предиктор, классифицированные по категориям.

- ✓ **Климат**
- ✓ **Почва**
- ✓ **Рельеф**
- ✓ **Растительность**
- ✓ **Пространственное положение объекта исследования**

Предикторы получены из общедоступного архива данных Google Earth Engine Catalog и обработаны посредством облачной платформы для анализа геопространственных данных Google Earth Engine (https://developers.google.com/earth-engine/datasets/catalog).

Рельеф

- **Aspect -** экспозиция
- **B1** – высота над уровнем моря
- **Slope** – уклон

Пространственное разрешение: *30 м/пиксель*

Источник данных: Цифровая модель рельефа ASTER DEM (https://asterweb.jpl.nasa.gov/gdem.asp)

Почва

- **Clay_mean** - среднее значение содержания глины в слое 0-30 см
- **Nitrogen_mean** – среднее значение содержания азота в слое 0-30 см
- **Sand_mean** – среднее значение содержания песка в слое 0-30 см

Пространственное разрешение: *250 м/пиксель*

Источник данных: Системы глобального цифрового картографирования почвы — SoilGrids (англ. System for global digital soil mapping, https://www.isric.org)

Растительность

- NDVI_AUG (OCT) – нормализованный разностный вегетационный индекс за август (октябрь). Индекс вычисляется на основе ближнего инфракрасного (NIR) и красного (Red) каналов

10 м/пиксель

Спутниковые данные Sentinel-2 (https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2)

Были исследованы различные комбинации каналов Landsat-8. (https://www.usgs.gov/landsat-missions/landsat-8)

Наиболее информативными оказались:

SR_B7 – медианное значение в канале SWIR-2.

ClayIndex – соотношение коротковолновых инфракрасных каналов SWIR1 и SWIR2 (Dogan, 2009).

OxideGB – соотношение зеленого (Green) и синего (Blue) каналов (Segal, 1982).

NDMI - нормализованный разностный индекс увлажнения растительности. Использует значения ближнего инфракрасного (NIR) и коротковолнового инфракрасного (SWIR) каналов (Wilson and Sader, 2002).

30 м/пиксель

Растительность

- Текстурные показатели Gray Level Co-occurrence Matrix (GLCM), позволяющие обнаружить связи между окружающими пикселями на спутниковом снимке. GLCM метрики были построены для индекса OxideGB, полученного на основе каналов Landsat-8.

Из построенных показателей GLCM информативными для моделирования сообществ дождевых червей оказались:

- CORR - корреляция между парами пикселей (OxideGB_corr),
- DISS – различие (OxideGB_diss),
- DVAR - разностная дисперсия (**OxideGB_dvar**)

30 м/пиксель

Спутниковые данные Landsat-8 (https://www.usgs.gov/landsat-missions/landsat-8)

Климат

- LST_MED – медианное значение температуры поверхности за 5-ти летний (2015-2019 гг.) период по данным канала TIR-1 (Thermal Infrared)

100 м/пиксель

Спутниковые данные Landsat-8 (https://www.usgs.gov/landsat-missions/landsat-8)

Биоклиматические предикторы - показатели, которые используются для характеристики климатических условий

- **Осадки самого сухого месяца**
- **Осадки самой сухой четверти года**
- **Сезонность выпадения осадков (коэффициент вариации)**
- **Максимальная температура наиболее теплого месяца**
- **Среднегодовое количество осадков**
- **Годовой размах температур**
- **Среднегодовая температура воздуха**
- **Температурная сезонность**
- **Средняя дневная разница температур**
- **Средняя температура самой холодной четверти года**
- **Средняя температура самой теплой четверти года**
- **Минимальная температура наиболее холодного месяца**
- **Изотермичность**
- **Осадки самой холодной четверти года**
- **Осадки самой теплой четверти года**
- **Осадки самого влажного месяца**
- **Осадки самой влажной четверти года**

Пространственное положение

- **Latitude** - широта
- **Longitude** - долгота

30 м/пиксель

Метрики оценки качества моделирования:

- коэффициент детерминации (R^2),
- средняя абсолютная ошибка (Mean Absolute Error, MAE),
- корень из средней квадратичной ошибки (Root Mean Squared Error, RMSE).

Коэффициент детерминации R^2 показывает насколько зависимая переменная объясняется или предсказывается независимыми переменными (пространственными предикторами) в модели.

MAE - это метрика, которая характеризует среднюю величину отклонений предсказанных значений модели от фактических значений. MAE измеряет точность модели, суммируя и усредняя абсолютные разности между предсказанными и реальными значениями.

Метрика RMSE используется для оценки общей точности прогноза, где меньшее значение означает более высокую точность прогноза модели. RMSE имеет одинаковую размерность с исходными данными.

Naive Bayes

- Naive Bayes - это алгоритм классификации в машинном обучении, основанный на **теореме Байеса**. Он предполагает, что все переменные в наборе данных «наивные», **то есть не коррелируют** друг с другом. Теорема Байеса - одна из основных теорем элементарной теории вероятностей, которая позволяет определить вероятность события при условии, что произошло другое статистически взаимозависимое с ним событие.
- Naive Bayes лучше работает с **категориальными** переменными, поскольку считает каждое значение признака отдельной категорией. Существуют варианты наивного байесовского классификатора, которые могут обрабатывать **непрерывные переменные с условием того, что данные имеют нормальное распределение**.
- В задаче построения геопространственных моделей сообществ дождевых червей данные обучающей выборки **не подчиняются закону и предикторы являются непрерывными.** Ввиду перечисленных ограничений использование метода Naive Bayes для создания регрессионных моделей сообществ дождевых червей невозможно.

Random Forest

- ✓ **Random Forest -** алгоритм машинного обучения, который основан на ансамбле деревьев решений. Каждое дерево строится независимо друг от друга на разных подвыборках обучающих данных, при этом используются разные комбинации признаков (характеристик) объектов.

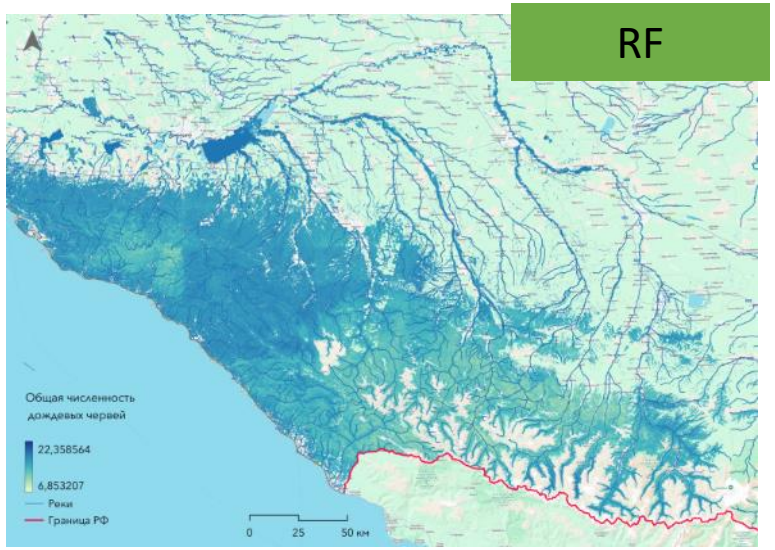
Преимущества:

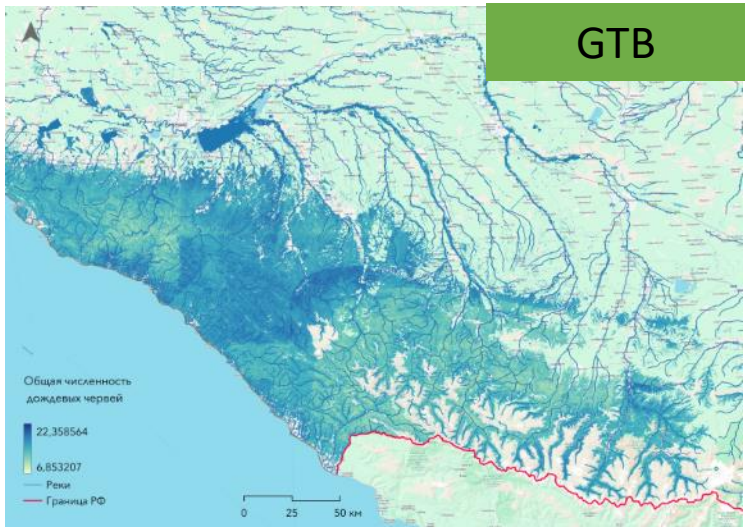
- ✓ устойчивость к переобучению (overfitting) благодаря использованию подвыборок и случайных признаков;
- ✓ способность обрабатывать большое количество признаков и работать с данными различных типов (например, числовыми, категориальными, текстовыми);
- ✓ возможность оценить важность каждого признака для задачи.

GradientTreeBoost

- Метод машинного обучения, использующий ансамбль деревьев решений.
- Метод отличается от Random Forest тем, что вместо усреднения решений отдельных элементов ансамбля, происходит последовательное добавление дополнительных элементов, чтобы получить наиболее точную оценку моделируемой переменной.
- По сравнению с RF градиентный бустинг более подвержен переобучению. Эффективность моделирования при использовании градиентного бустинга может зависеть от настройки гиперпараметров.

Random Forest, GradientTreeBoost





Во всех моделях сообществ дождевых червей «случайный лес» продемонстрировал более высокую точность моделирования по сравнению с «градиентным бустингом», что подтверждается метриками точности.

Среднее значение R2 по всем моделям RF варьирует от 0.271 до 0.447, в то время как у моделей BRT - от 0.172 до 0.44.

Средняя ошибка RMSE при моделировании численности дождевых червей методом RF находится в интервале от 4.3 до 19.7 (ind/m^2) , методом BRT - от 3.6 до 23.7 (ind/m^2). Средняя ошибка RMSE при моделировании биомассы дождевых червей методом RF находится в интервале от 3.5 до 7.9 (g/m^2) , методом BRT - от 3.8 до 8.9 (g/m^2).

Для повышения точности моделирования методом GTB необходима более детальная настройка гиперпараметров.

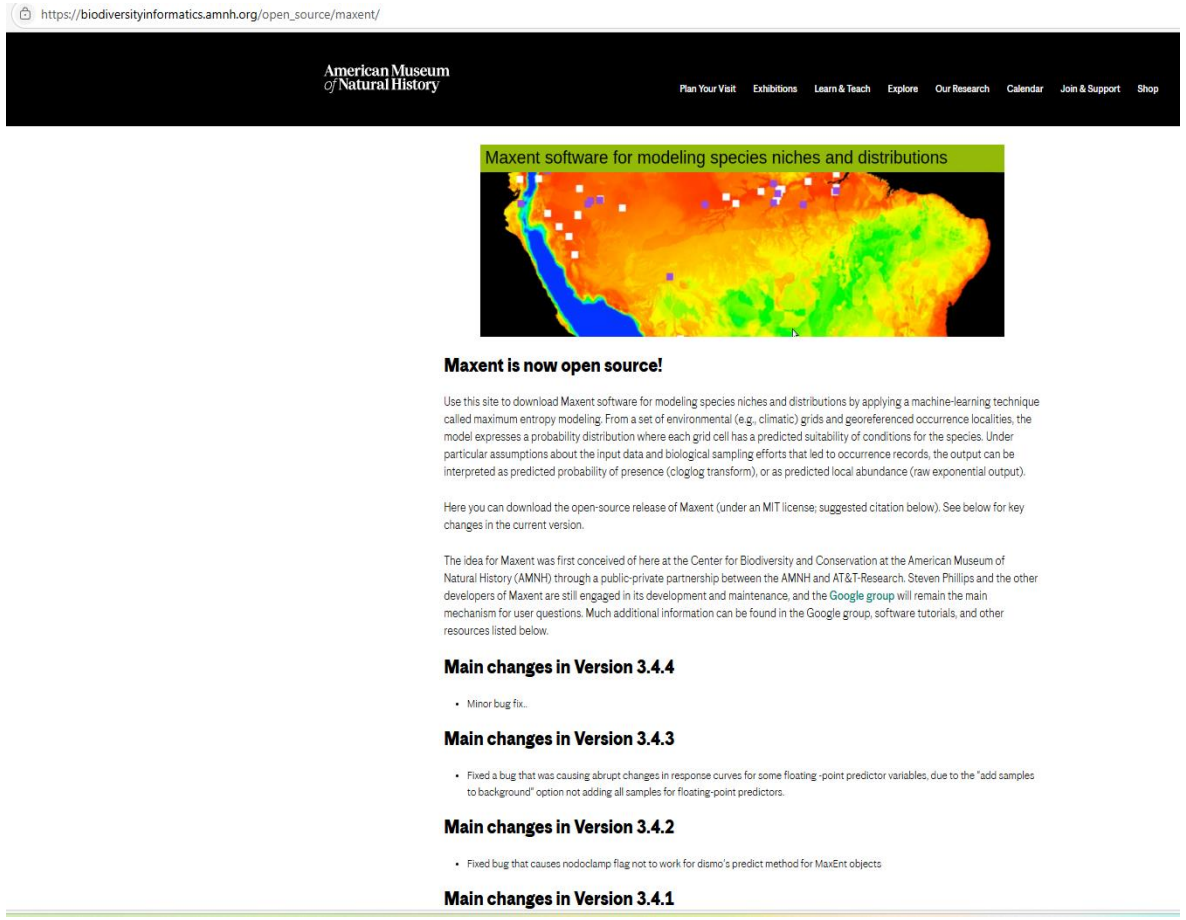
Информативные предикторы регрессионных моделей

Суммарно по двум моделям (RF и GTB) сообществ дождевых червей Северо-Западного Кавказа наиболее информативными предикторами оказались:

- ✓ пространственное положение – **долгота,** широта
- ✓ почвенные характеристики – **содержание песка,** содержание глины и азота
- ✓ характеристики растительности – текстурные показатели, вычисленные по спутниковым данным, степень увлажнения (соотношение каналов SWIR 1 и SWIR 2)
- ✓ климат – многолетняя температура поверхности, осадки самой сухой четверти года, максимальная температура наиболее теплого месяца

Maxent

Maxent это алгоритм машинного обучения, который применим для прогнозирования области вероятности присутствия вида. Метод основывается только на точках присутствия видов (presence-only), без учета мест документированного отсутствия. Суть метода сводится к поиску закономерностей распределения значений экологических параметров в точках, где доказано обитание вида. Maxent выбирает область вероятности присутствия, которая больше всего похожа на среду, из которой она была выведена, уменьшая при этом все другие предположения (или максимизируя ее энтропию): «*она согласуется со всем, что известно, но тщательно избегает допущения чего-либо неизвестного*» (Джейнс, 1990).



Main changes in Version 3.4.0

- Released under a MIT license**
- Threshold features are now turned off by default
- A cloglog transform has been added and now constitutes the default transformation for model output (formerly the logistic transform)

Download

Current version 3.4.4

Please tell us a little about yourself!

Name:

Institution:

Email:

Comment/Intended Use*:
*Optional

Submit

I prefer to download without providing this information

Citation

If you use the application for analyses that result in a publication, report, or online posting, the following represents a proper citation of the software itself:

Steven J. Phillips, Miroslav Dudík, Robert E. Schapire. [Internet] Maxent software for modeling species niches and distributions (Version 3.4.1). Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2025-8-20.

***For information about earlier versions, please refer to the readme file on github or contact the developers mrmaxent@gmail.com*

Биоклиматические предикторы - показатели, которые используются для характеристики климатических условий

- **Осадки самого сухого месяца**
- **Осадки самой сухой четверти года**
- **Сезонность выпадения осадков (коэффициент вариации)**
- **Максимальная температура наиболее теплого месяца**
- **Среднегодовое количество осадков**
- **Годовой размах температур**
- **Среднегодовая температура воздуха**
- **Температурная сезонность**
- **Средняя дневная разница температур**
- **Средняя температура самой холодной четверти года**
- **Средняя температура самой теплой четверти года**
- **Минимальная температура наиболее холодного месяца**
- **Изотермичность**
- **Осадки самой холодной четверти года**
- **Осадки самой теплой четверти года**
- **Осадки самого влажного месяца**
- **Осадки самой влажной четверти года**

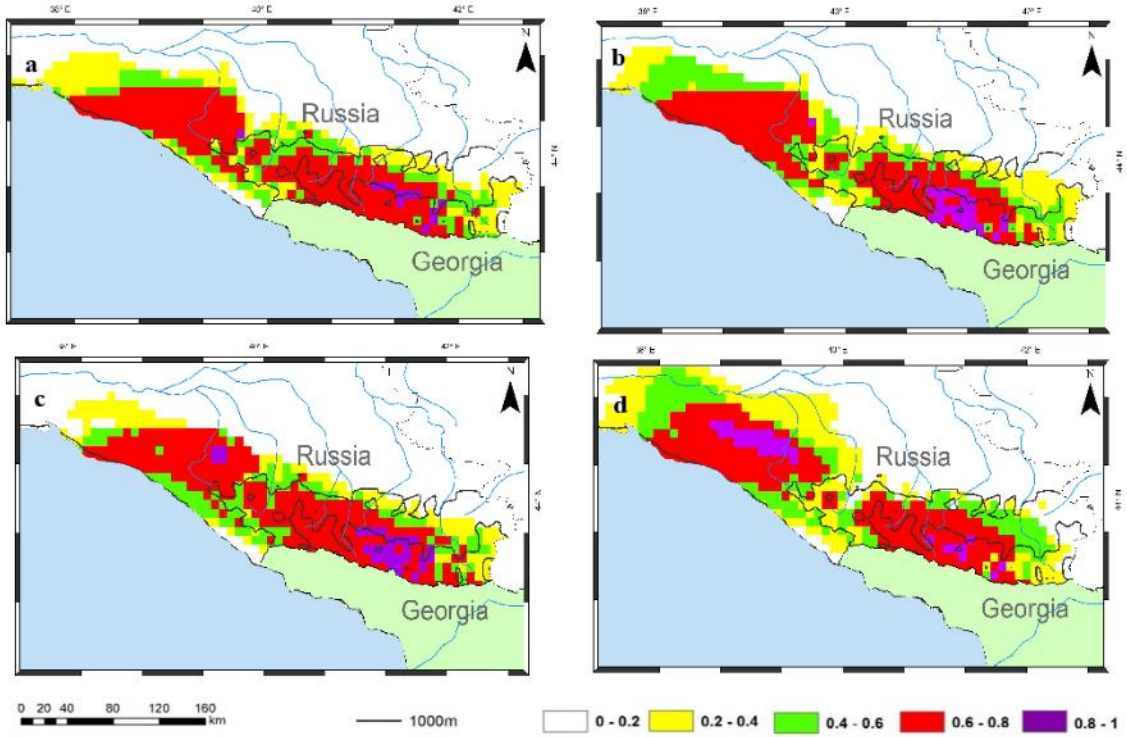
Maxent моделирование

Наиболее значимые биоклиматические предикторы для дождевых червей Кавказа:

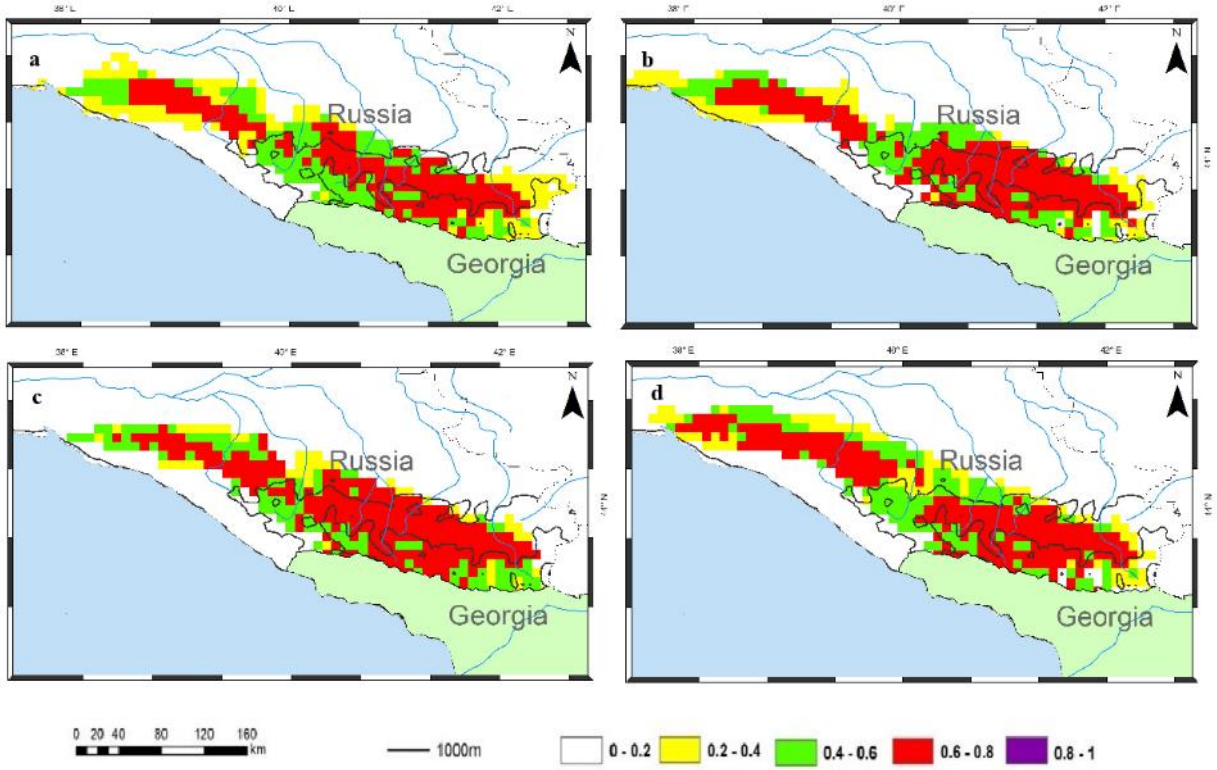
- **Осадки самого сухого месяца**
- **Сезонность выпадения осадков (коэффициент вариации)**
- **Максимальная температура наиболее теплого месяца**

Дождевые черви чувствительны к повышению температуры в сочетании со снижением количества осадков, а также к неравномерному их выпадению, что находит отражение в сокращении площади потенциальных ареалов в наиболее пригодных условиях обитания для этих видов дождевых червей по разным климатическим сценариям, в особенности наиболее «жестким» – RCP 6.0 и RCP 8.5. Наиболее чувствительны к изменениям климата **эндемичные/субэндемичные крымско-кавказские виды**, а также **средиземноморские**, для которых сокращается область распространения в средне- и высокогорных лесах. Наименее чувствительны – космополиты.

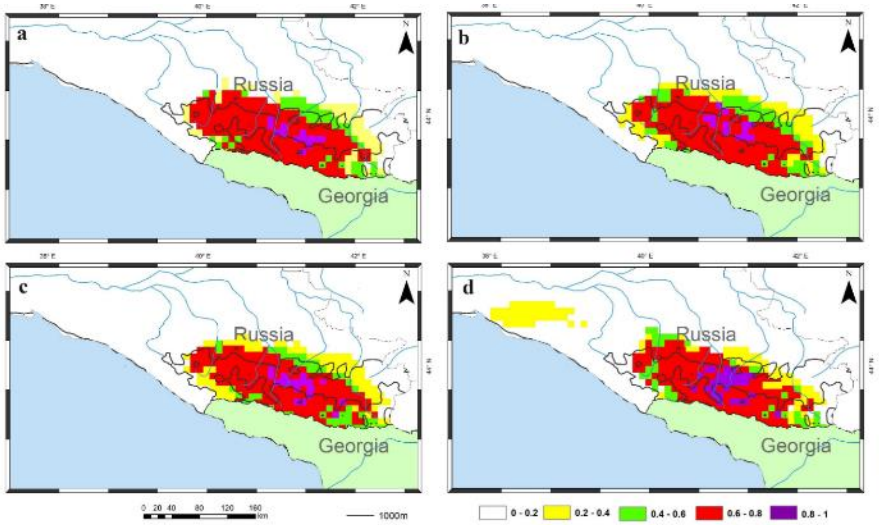
Range models for earthworm species were constructed using the maximum entropy method (Maxent) for 4 probable climate scenarios (rcp): 2.6, 4.5, 6.0 and 8.5 up to 2070.



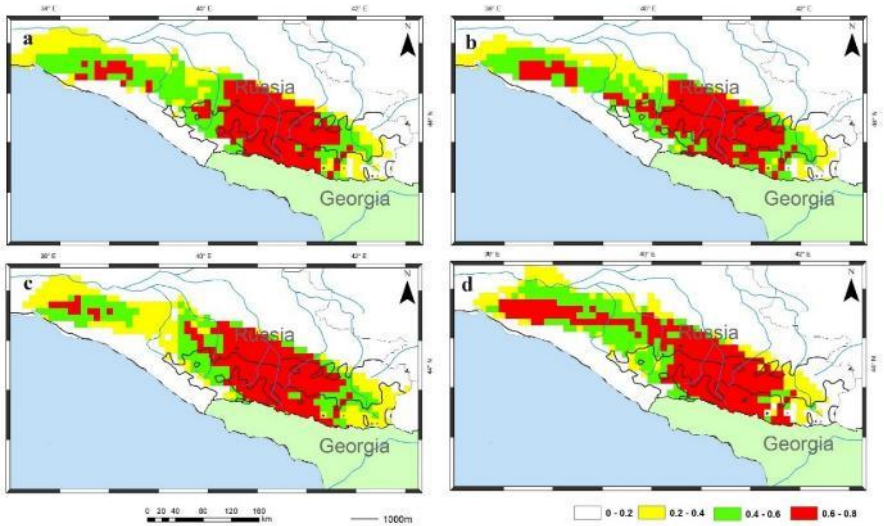
Dendrobaena attemsi (epigenic, Mediterranean)



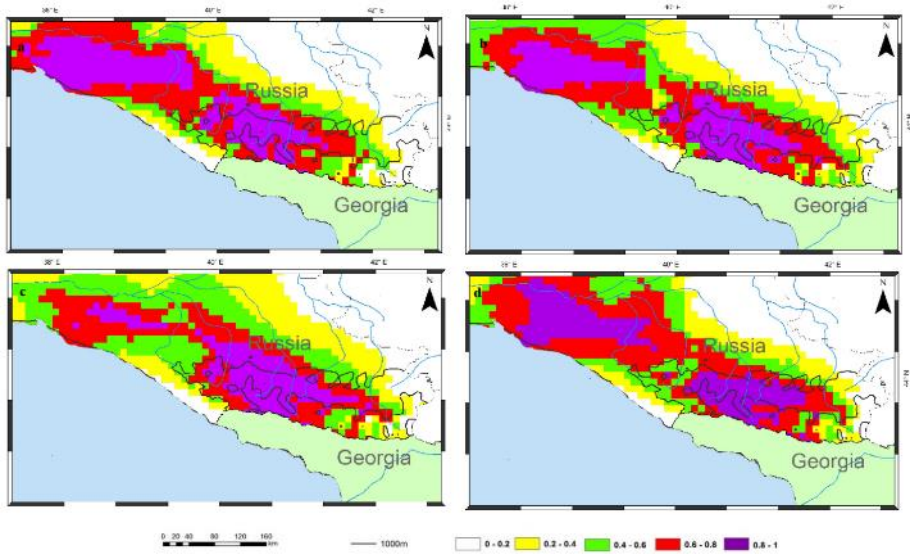
Dendrobaena octaedra (epigenic, cosmopolitan)



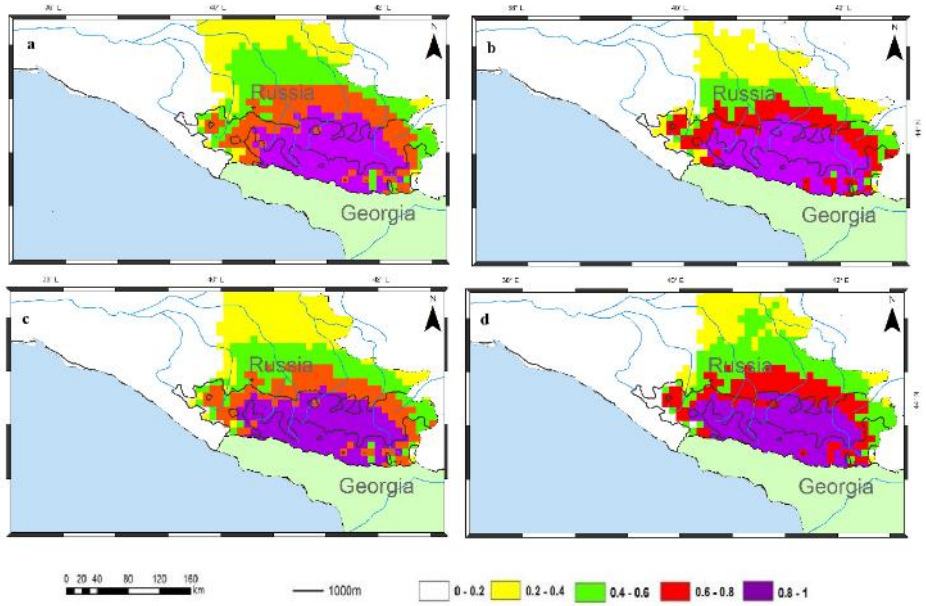
Dendrobaena schmidti (epi-endogenic, (sub)endemic)



Dendrobaena schmidti (endogenic, (sub)endemic)



Eisenia fetida (epi-endogeic, cosmopolitan)



Dendrobaena veneta (epi-endogeic, средиземноморский)

Основные условия успеха в моделировании:

- Достоверность первичной информации, полученной в полевых условиях (верная идентификация видов, четкая географическая привязка).
- Грамотное формирование набора первичных данных. Если точки присутствия распределены агрегировано, это может оказать существенное влияние на результаты прогнозирования. Большинство авторов предлагает устранять агрегации находок видов (исключение может быть для редких видов с узкими ареалами).
- Знания и умения работы в ГИС для формирования корректного пространственного материала для анализа.
- Обучение и тестирование модели должно проводиться на независимых наборах данных.
- Для анализа влияния факторов среды на формирование ареала, необходимо минимизировать корреляции предикторов.

Вместо заключения…

«Методы моделирования распространения видов не являются “волшебной палочкой”, позволяющей получать оптимальные результаты при минимальных усилиях. Напротив, от исследователя требуется внимание и специальные знания на разных этапах разработки модели».

(Лисовский, Дудов, 2020).

Благодарности: команда проекта

к.б.н., зав. лабораторией ЦЭПЛ РАН

АННА ГЕРАСЬКИНА

Полевые исследования. Количественный учет дождевых червей. Идентификация видов дождевых червей. Экспертная оценка результатов моделирования

к.т.н., с.н.с. ЦЭПЛ РАН

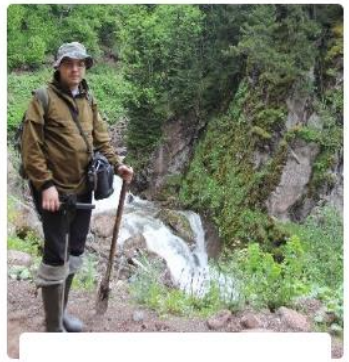
АЛЕКСАНДРА ПЛОТНИКОВА

Подготовка предикторов геопространственного моделирования. Техническое сопровождение моделирования. Разработка сайта проекта

аспирант ЦЭПЛ РАН

АННА НАРЫКОВА

Геопространственное моделирование. Выявление предикторов, определяющих состав сообществ дождевых червей. Создание WEB-ГИС проекта



к.б.н., ученый секретарь ЦЭПЛ РАН

НИКОЛАЙ ШЕВЧЕНКО

Полевые исследования. Геоботанические описания. Максент-моделирование. Экспертная оценка результатов моделирования

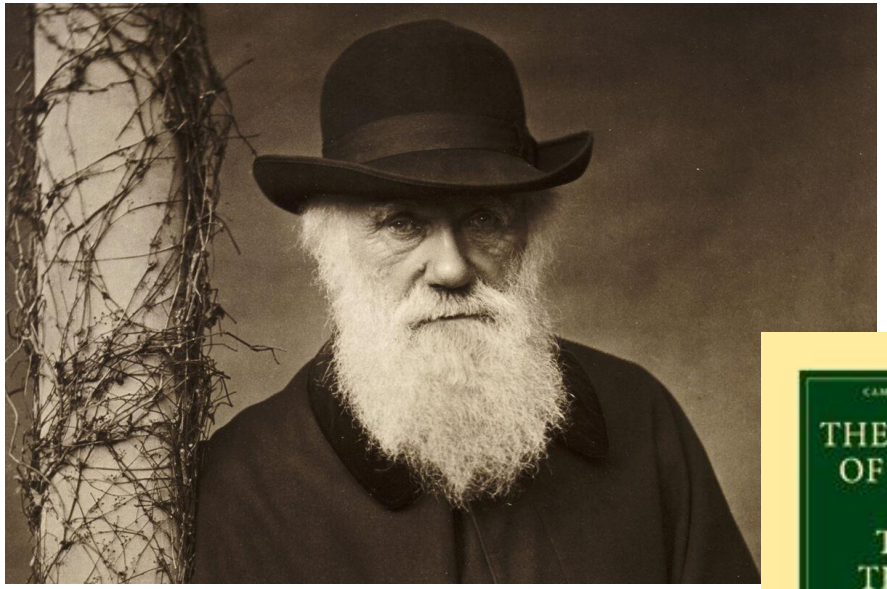
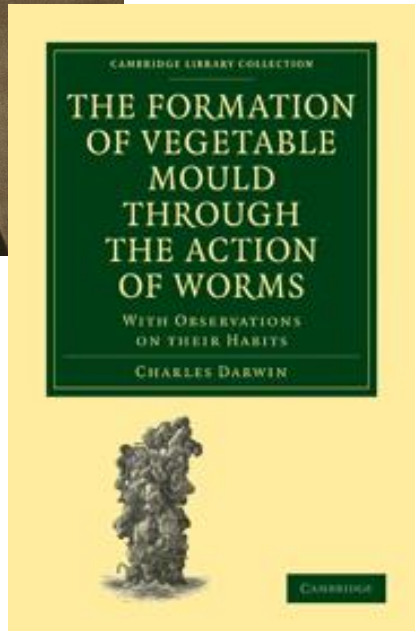
инженер ЦЭПЛ РАН

ЕЛЕНА АРХИПЦЕВА

Создание WEB-ГИС и сайта проекта

https://earthworms.ru/

РИТМ УГЛЕРОДА

ЦЭПЛ

Спасибо за внимание!