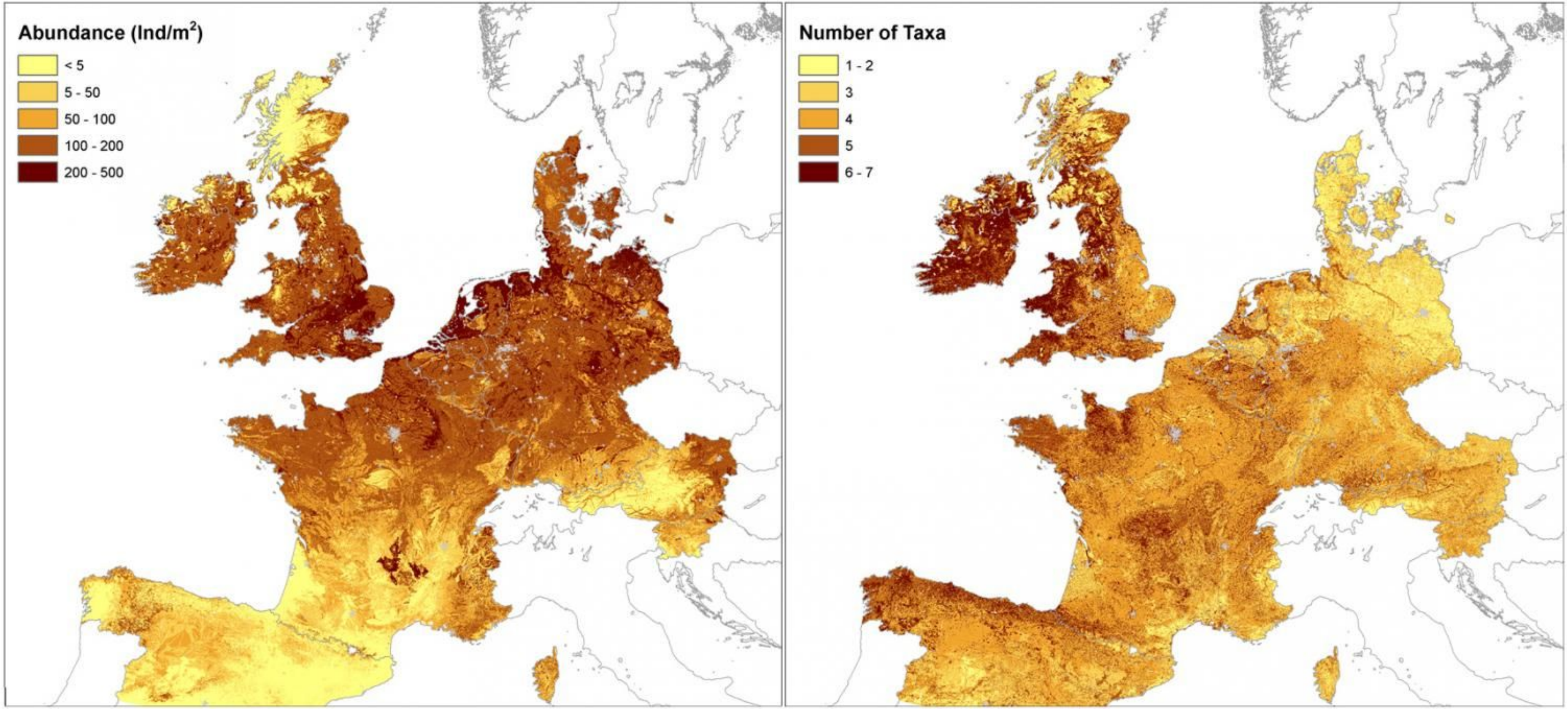
Earthworms of European Russia: literature heritage and citizen science for modelling

Maxim Shashkov, Natalya Ivanova, Sergey Ermolov

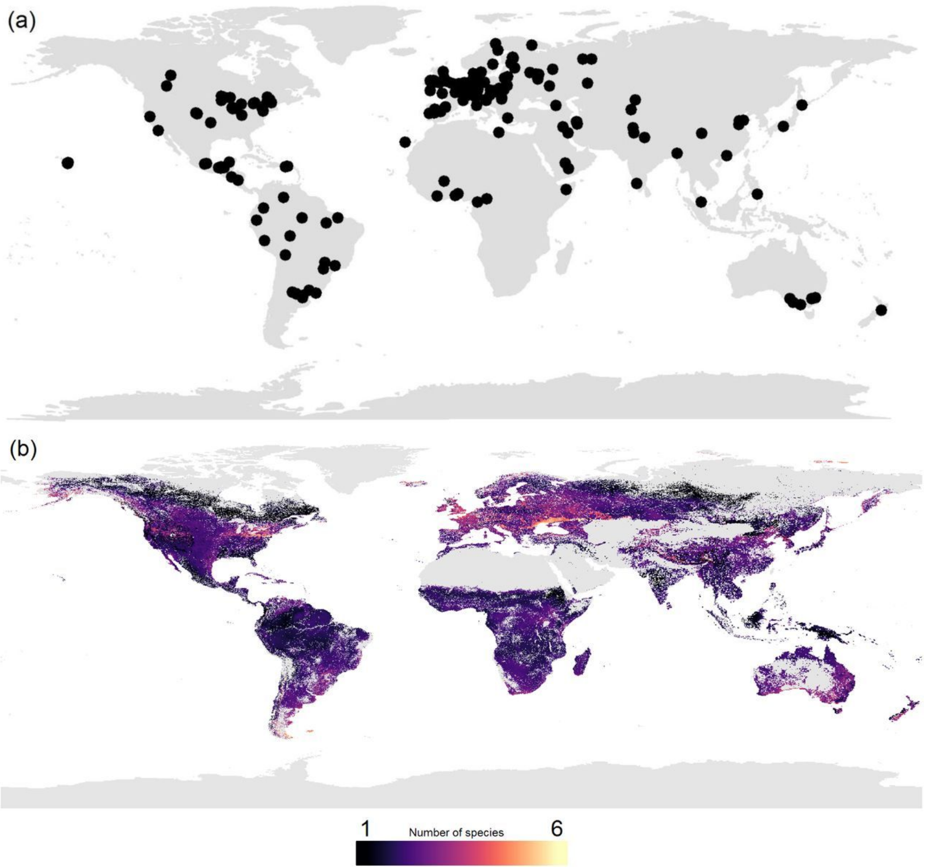
Karaganda, Kazakhstan, 2025

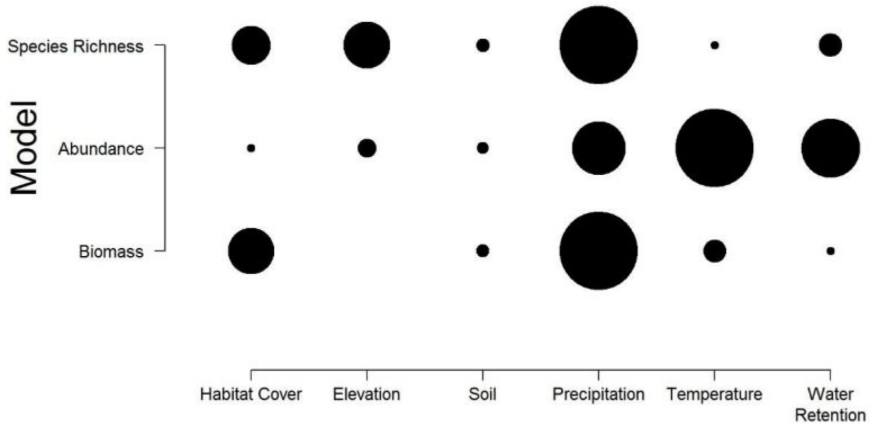
Regional Spatial Model (Rutgers et al., 2016)



Global Distribution of Earthworm Diversity

Phillips et al., 2019

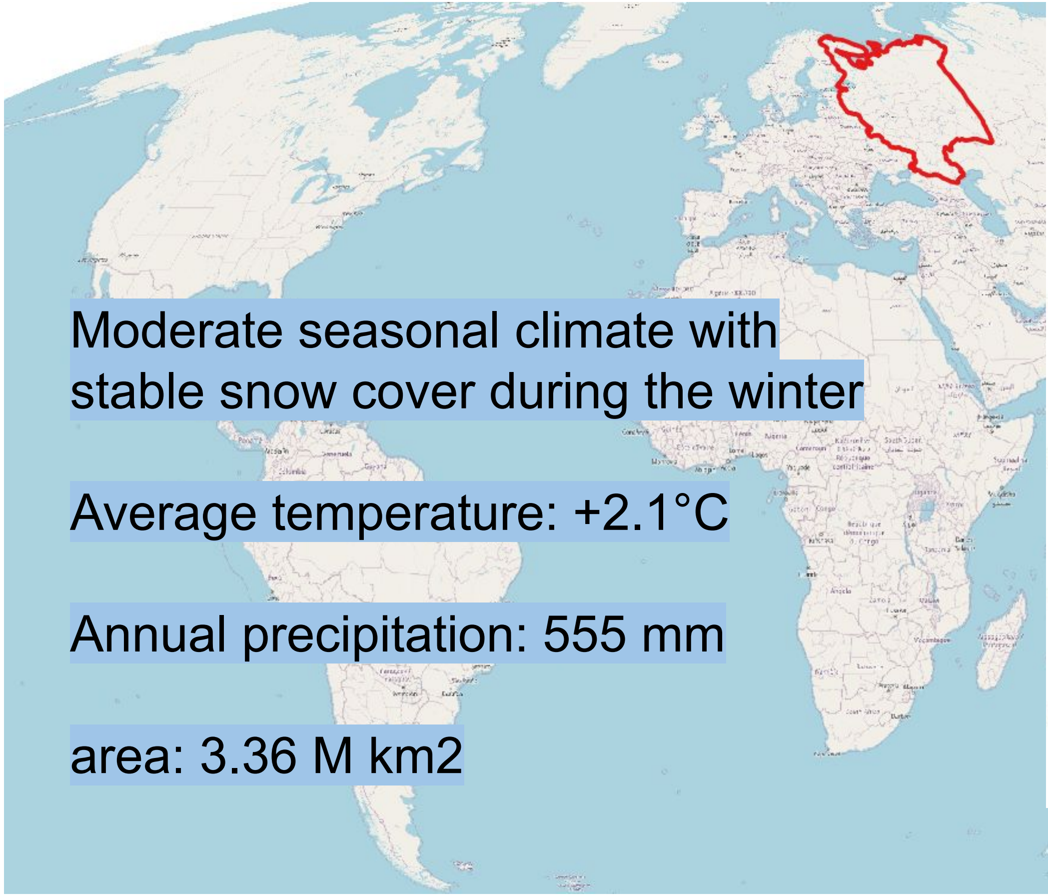




The objective of the study:

quantitative estimation of the factors limiting the distribution of the common earthworm species of the territory of the European part of Russia

Territory of interest

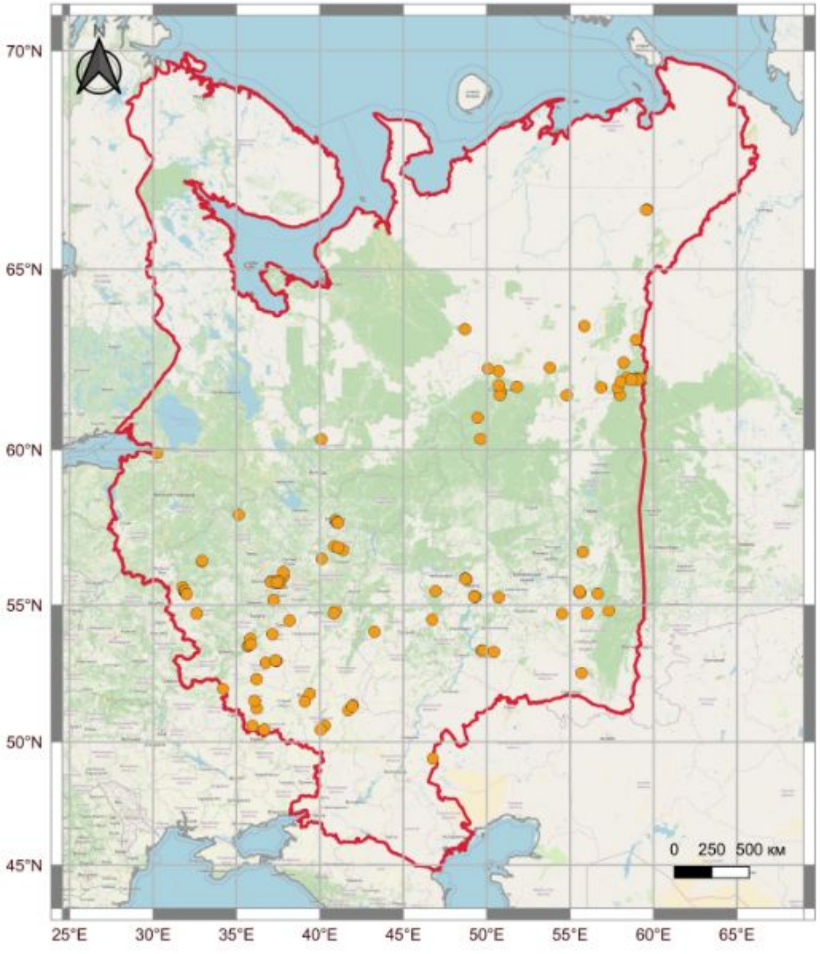


Moderate seasonal climate with stable snow cover during the winter

Average temperature: +2.1°C

Annual precipitation: 555 mm

area: 3.36 M km2



Focal species

Aporrectodea caliginosa (Savigny, 1826)
endogeic

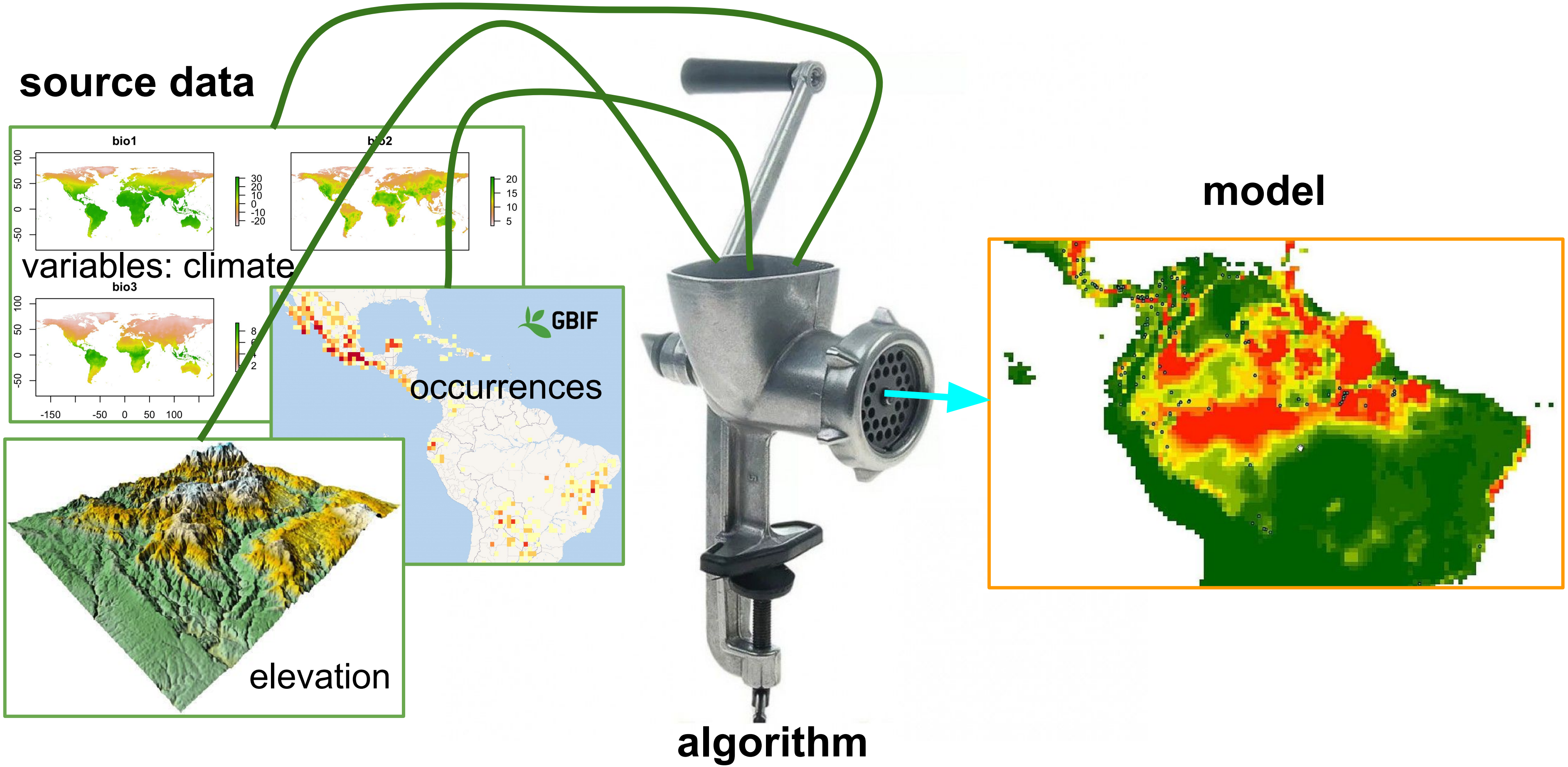
Dendrobaena octeadra (Savigny, 1826)
epigeic

Lumbricus rubellus Hoffmeister, 1843
epi/endogeic

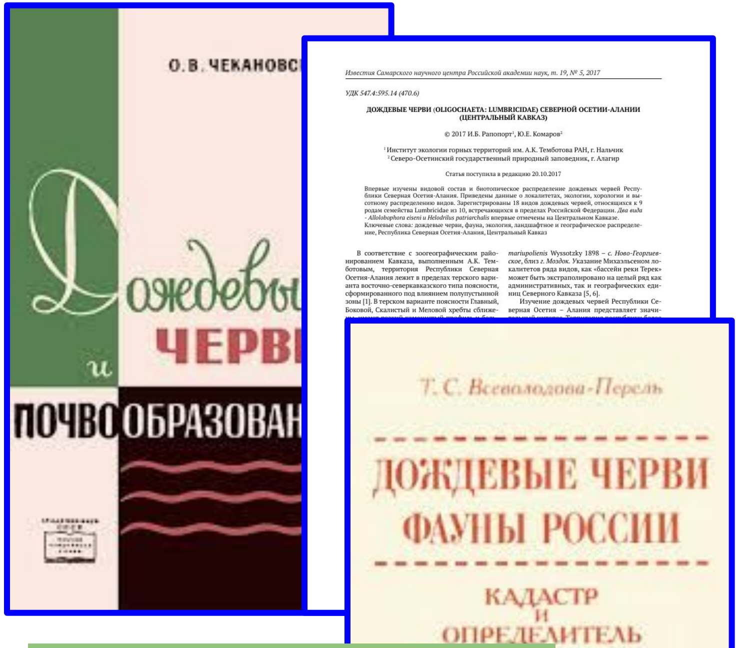
Lumbricus terrestris Linnaeus, 1758
anecic

Octolasion lacteum (Örley, 1881)
endogeic

Method: Species Distribution Modelling based on MaxEnt



Species occurrences



verbatimIdentification	organismQuantity	organismQ	fieldNur	habitat	stateProvince
Aporrectodea caliginosa				Дерново-подзолистая окультуренная удобрям	Moscow Oblast
Aporrectodea caliginosa (Savigny, 1826)				Ольшаник сырой крапивно-разнотравный, почв	Voronezh Oblast
Aporrectodea caliginosa				Ботанический сад МГУ им. М.В. Ломоносова н	Moscow
Aporrectodea caliginosa caliginosa (Savigny 1826)	36.5±5.5	ind./sq m		Выпасаемый суходольный луг	Komi Republic
Allolobophora caliginosa (Savigny, 1826)				в парке Екатерингоф (водоем вокруг стадиона	Saint Petersburg
Allolobophora caliginosa (Savigny, 1826)				на Крестовском острове, в 20 м от Морского ях	Saint Petersburg
Aporrectodea caliginosa Sav.				футбольное поле	Moscow Oblast
Aporrectodea caliginosa (Savigny, 1826)					Murmansk Oblast
Aporrectodea caliginosa (Savigny, 1826)					Murmansk Oblast
Aporrectodea caliginosa (Savigny, 1826)					Murmansk Oblast
Aporrectodea caliginosa				Контрольный и техногенно загрязненный участ	Komi Republic
Aporrectodea caliginosa caliginosa	3.2±0.8	ind./sq m		Ельник черничный	Komi Republic
Aporrectodea caliginosa caliginosa	1.5±0.6	ind./sq m		Суходольный луг	Komi Republic
Aporrectodea caliginosa caliginosa	16.5±1.6	ind./sq m		Поле с многолетними травами	Komi Republic
Aporrectodea caliginosa caliginosa	0.2±0.1	ind./sq m		Елово-березовый разнотравный лес	Komi Republic
Aporrectodea caliginosa caliginosa	1.6±0.8	ind./sq m		Ельник черничный	Komi Republic
Aporrectodea caliginosa caliginosa	48.6±4.8	ind./sq m		Пойменный разнотравно-злаковый луг	Komi Republic
Aporrectodea caliginosa caliginosa	6.4±1.6	ind./sq m		Осинник разнотравный I класса возраста	Komi Republic
Aporrectodea caliginosa caliginosa	1.2±0.4	ind./sq m		Осинник разнотравный VII-VIII классов возраст	Komi Republic
Aporrectodea caliginosa caliginosa	13.3±2.4	ind./sq m		Разнотравный луг	Komi Republic
Aporrectodea caliginosa caliginosa (Savigny, 1826)	84±23	ind./sq m		Бор высокотравно-папоротниковый	Novosibirsk Oblast
Aporrectodea caliginosa caliginosa (Savigny, 1826)	80±33	ind./sq m		Бор высокотравно-папоротниковый	Novosibirsk Oblast

159 literature sources
5,304 occurrences
time coverage: 1868-2022
110 species, 27 countries

GBIF

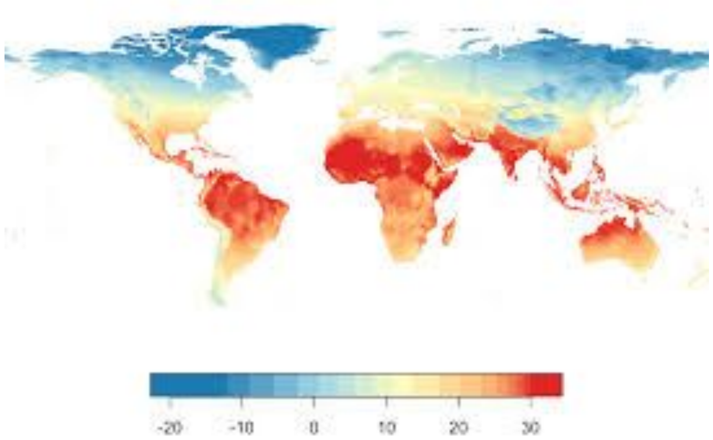
iNaturalist

Species occurrences extracted from literature

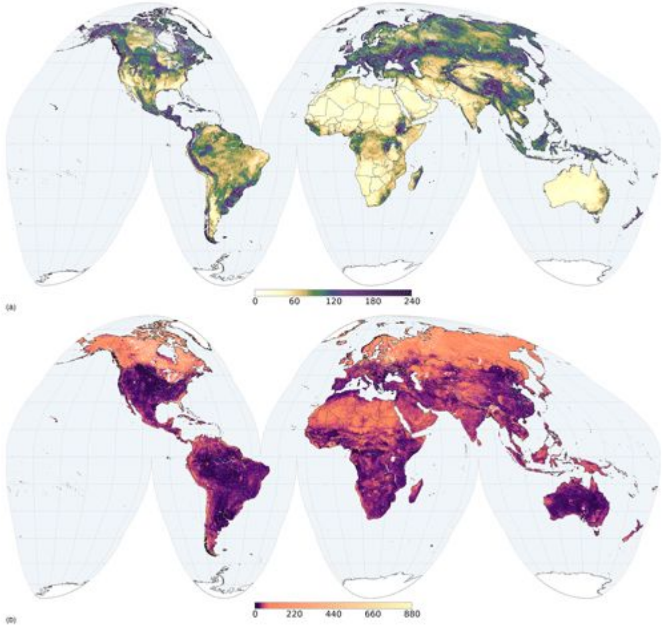
GBIF

Environmental Variables

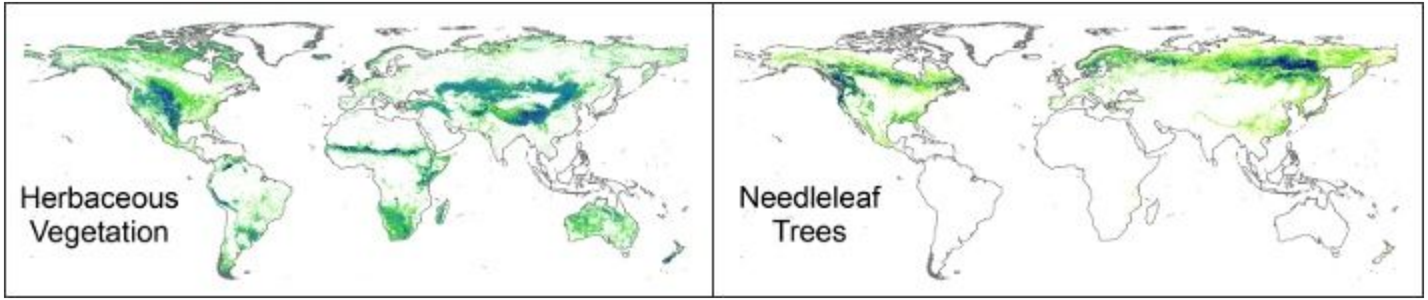
WorldClim



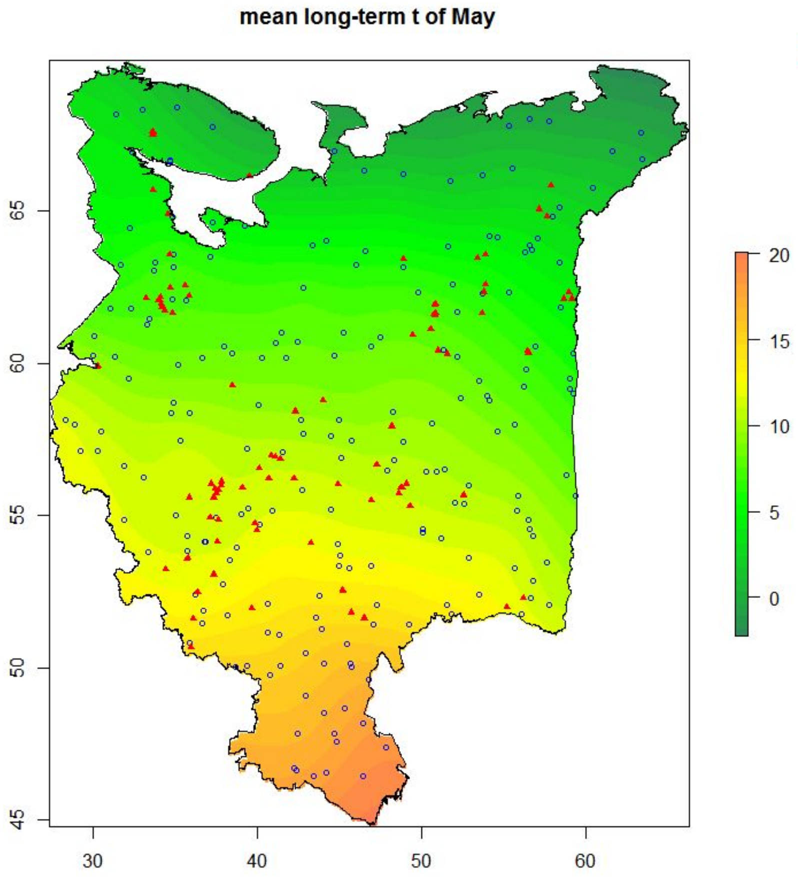
Soil Grid 250

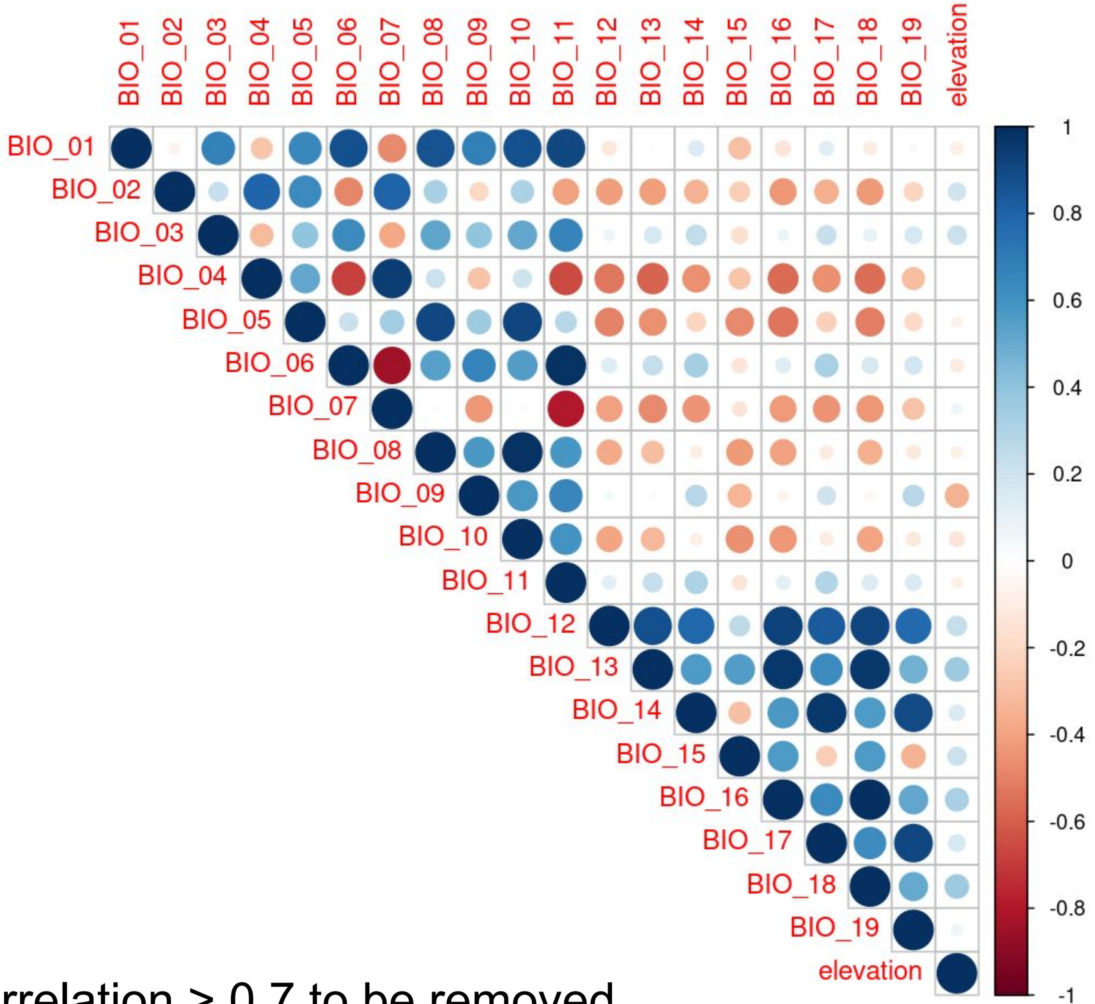


Global 1-km Consensus Land Cover



Collinearity



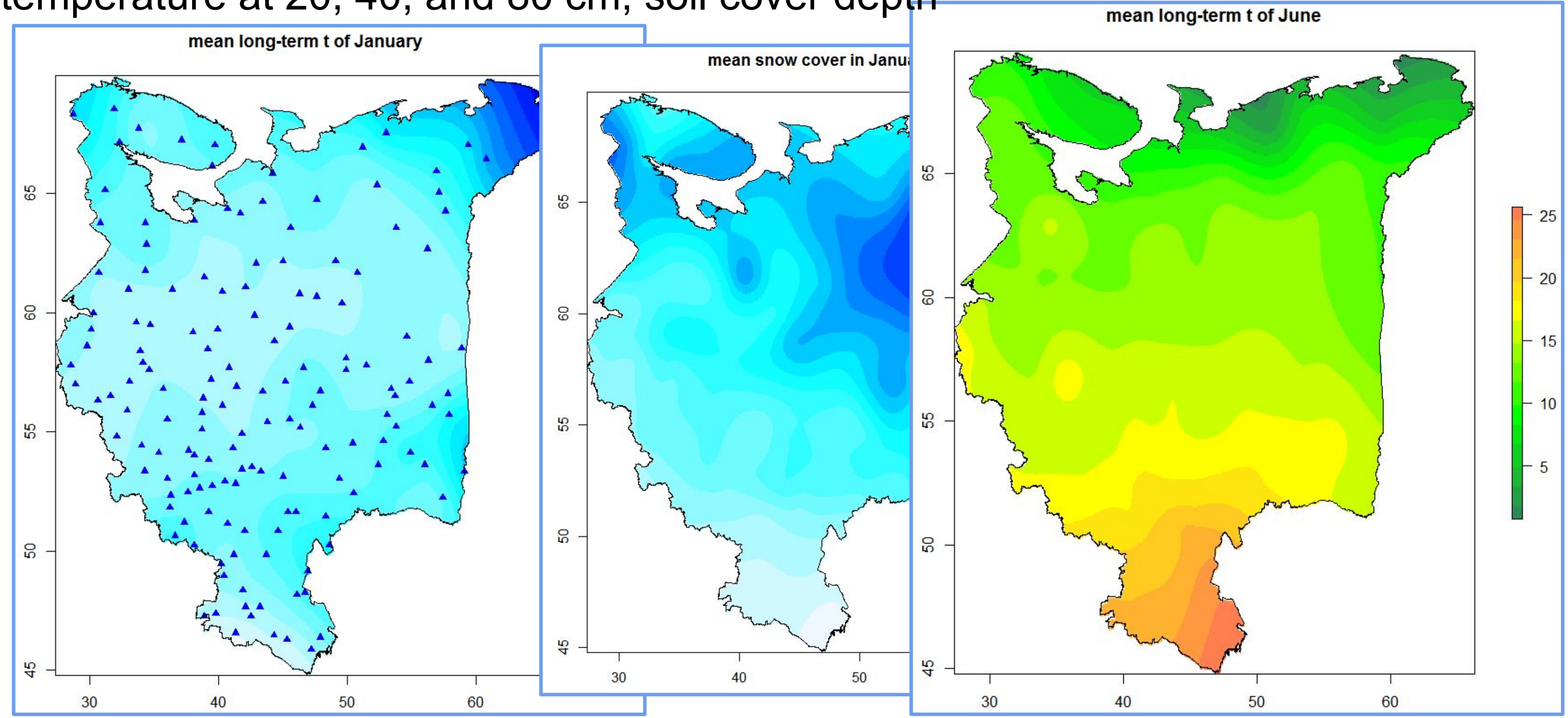


one of the layers with a pair correlation > 0.7 to be removed

Soil Climate

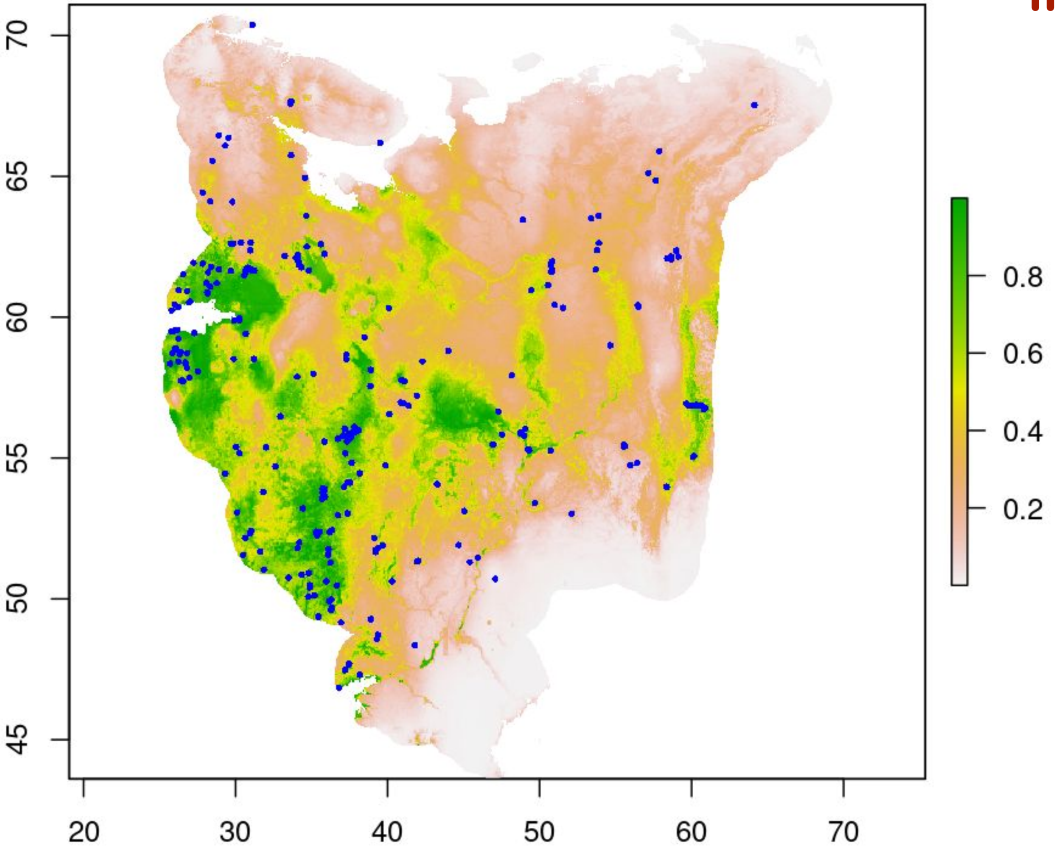
weather stations archive

soil temperature at 20, 40, and 80 cm; soil cover depth

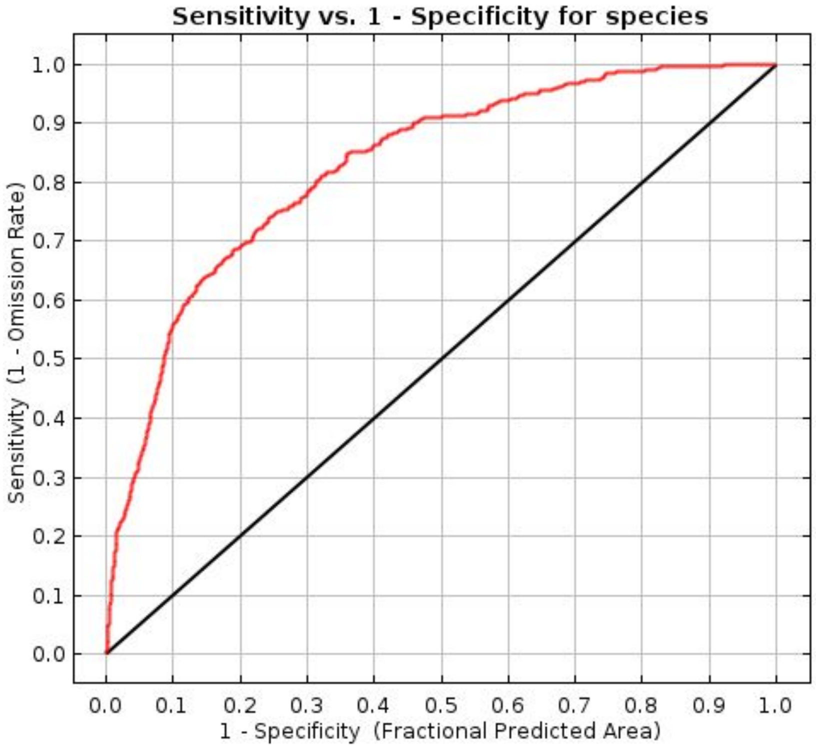


Modelling results: WorldClim

Aporrectodea caliginosa



model performance: AUC = 0.827



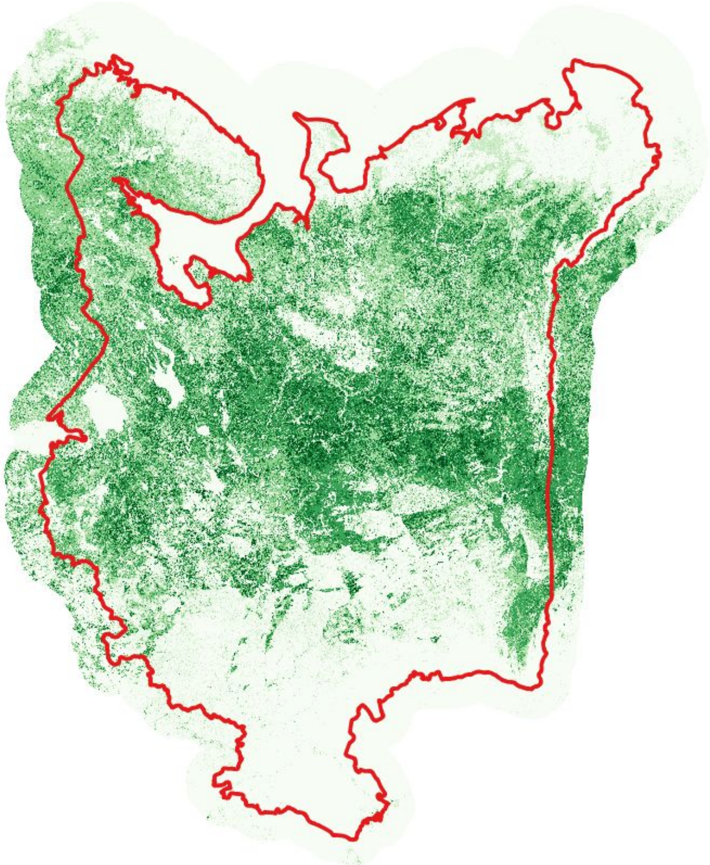
World Climate: Variable Importance

	Ac	Do	Lr	Lt	Ol
t diurnal range	18.4-24.3%	3-57.4%	3.1-5.5%	16.3-25.8%	5.0-13.2%
max t or warmest month	5.2-13.4%	57.4-61.4%	27.7-34.3%	10.2-14.8%	23.6-30.7%
annual precipitation	18.1-32.7%	1.4-6.4%	24.7-35.7%	3.8-42.6%	12.5-24.8%
elevation	5.2-8.8%	4.8-14.6%	12.7-15.5%	1.2-4.1%	0.1-1.9%

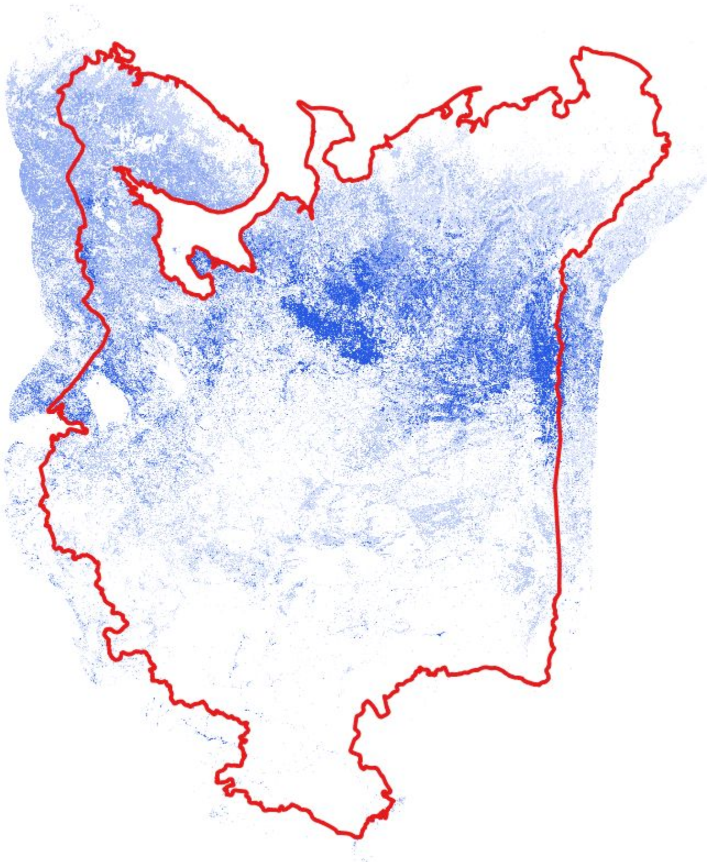
18.1-32.7%	- high importance of variable for model
3-57.4%	- low to high importance
12.5-24.8%	- high importance with failed permutation test

Biotopic / land use classification: Global 1 km Consensus Land Cover

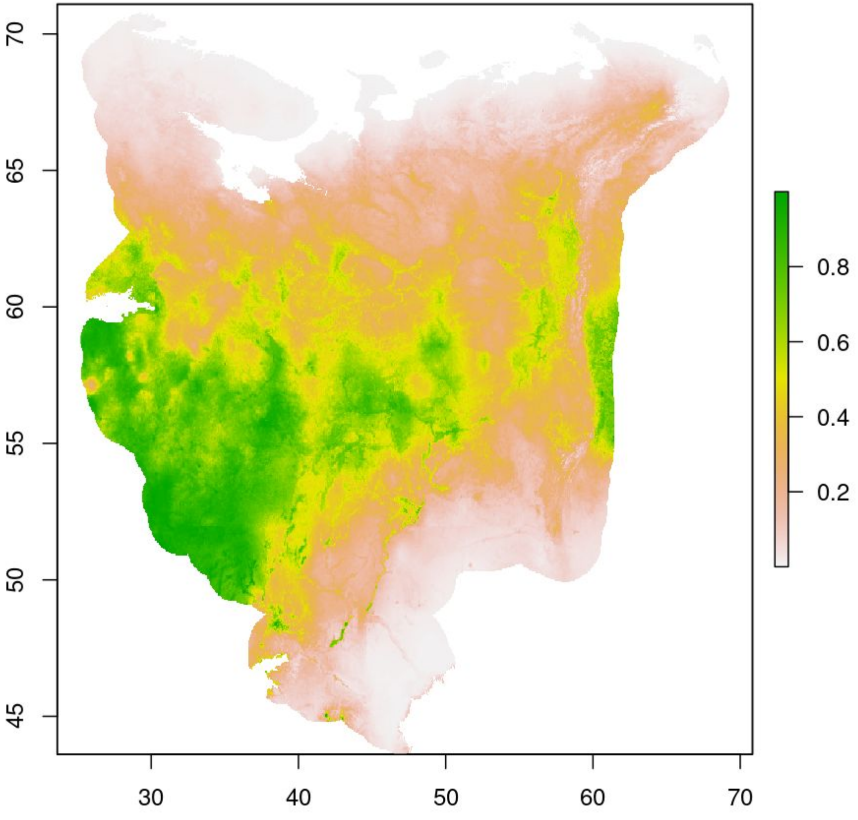
Mixed forests

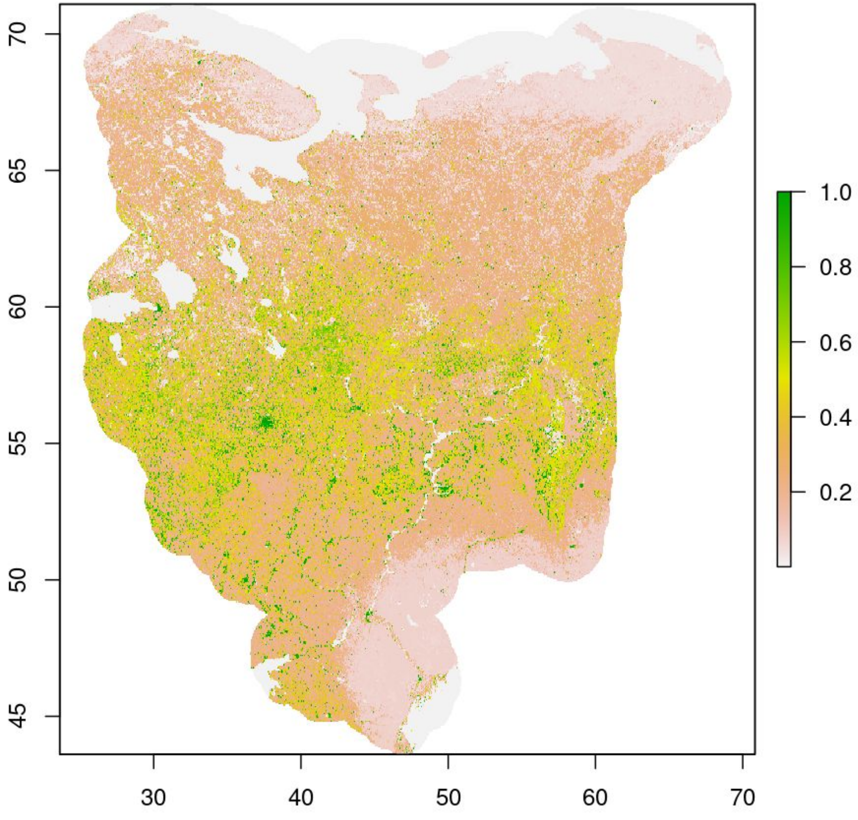


Coniferous forests



World Climate vs Consensus layers

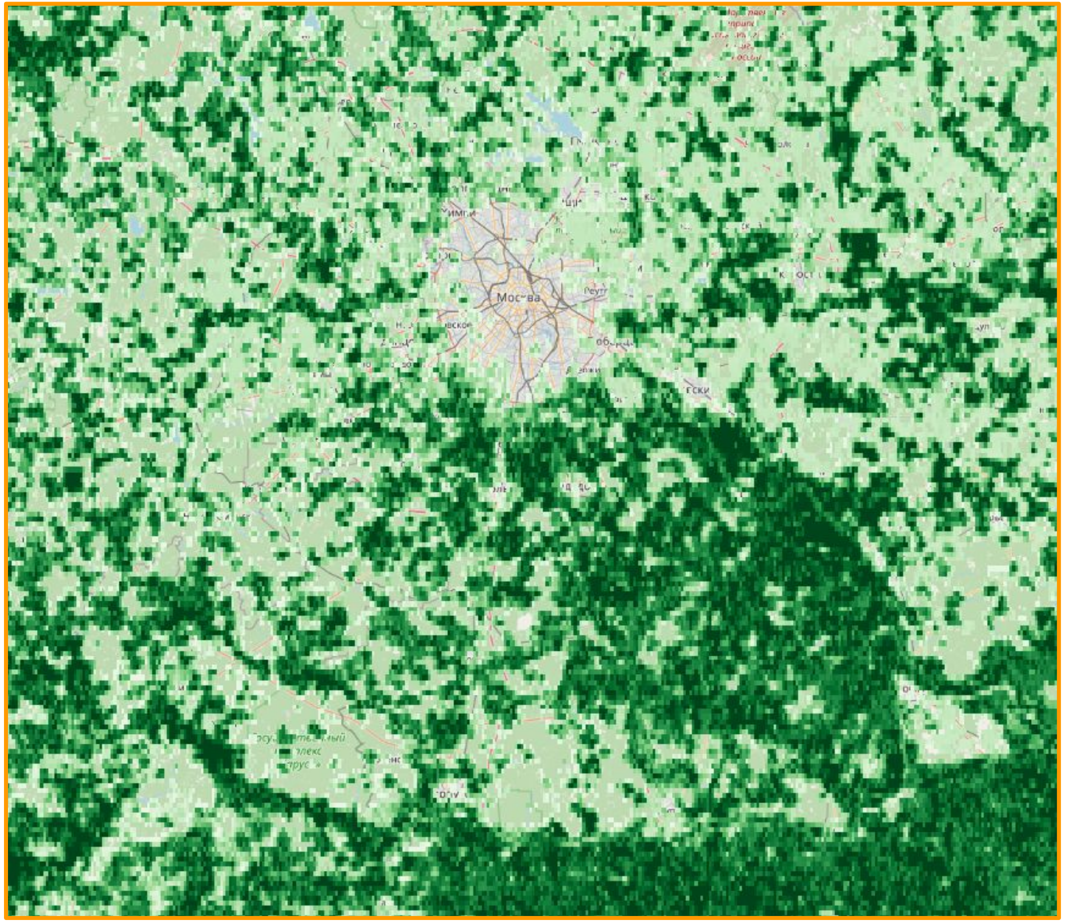




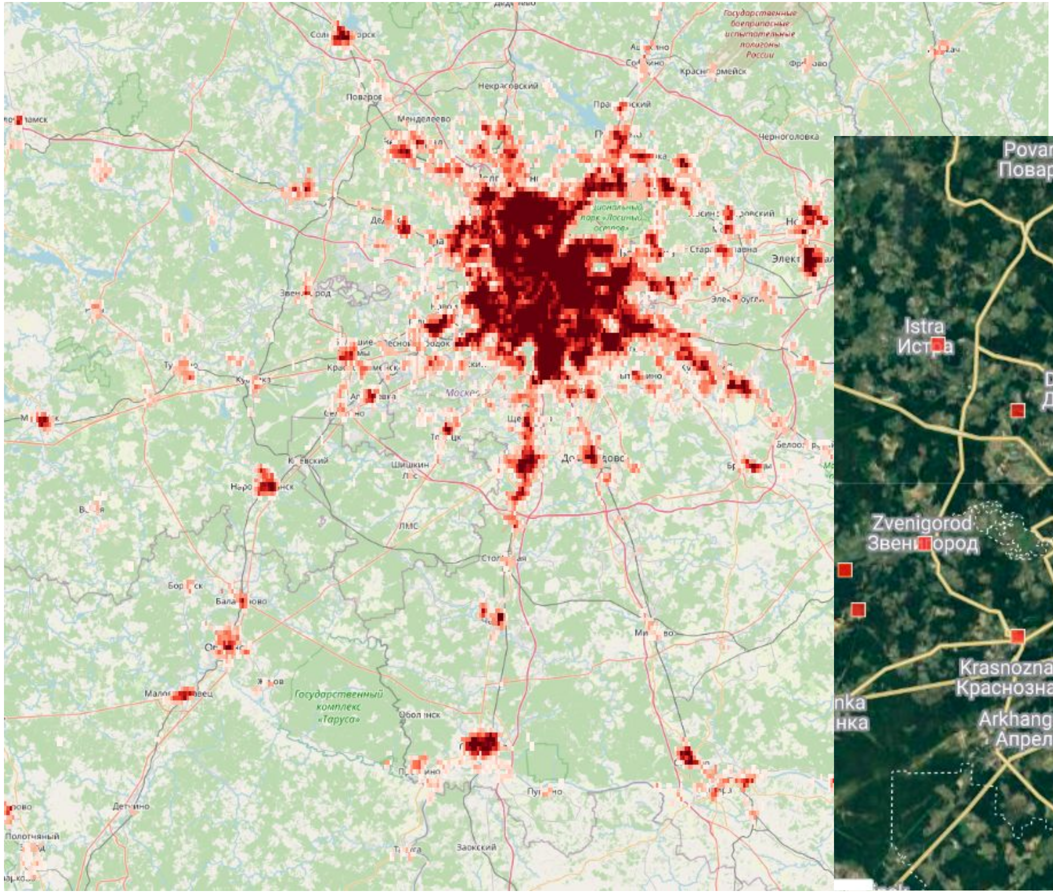
The Most Important Land Cover Variables

	Ac	Do	Lr	Lt	Ol
cultivated area	2.5-5.2%	19.7-27.5%	6.7-10.9%	10.6-32.7%	7.0-13.4%
deciduous forests	6.9-31.6% 4.7-12.7%	13.2-17.3%	2.5-12.4%	1.9-10.8%	24.4-52.6%
needleleaf forests	4.6-8.8%	9.9-18.1%	8.9-16.8%	1.6-3.5%	5.0-8.3%
mixed forests	5.1-11.5%	4.1-5.5%	8.6-24.6%	5.3-15.6%	3.6-5.4%
herbaceous biotopes	2.5-4.6%	5.8-10.2%	2.7-5.1%	2.6-4.6%	1.3-3.0%
urban territories	31.6- 59.8% 12.6-12.8%	7.9-22.4% 3.2-9.2%	18.0-56.2% 4.9-17.6%	25.1-68.2% 8.0-32.0%	10.4-28.6% 3.3-9.2%
water	2.2-5.3%	3.0-6.5%	5.2-5.9%	2.9-12.5%	10.5-14.5%

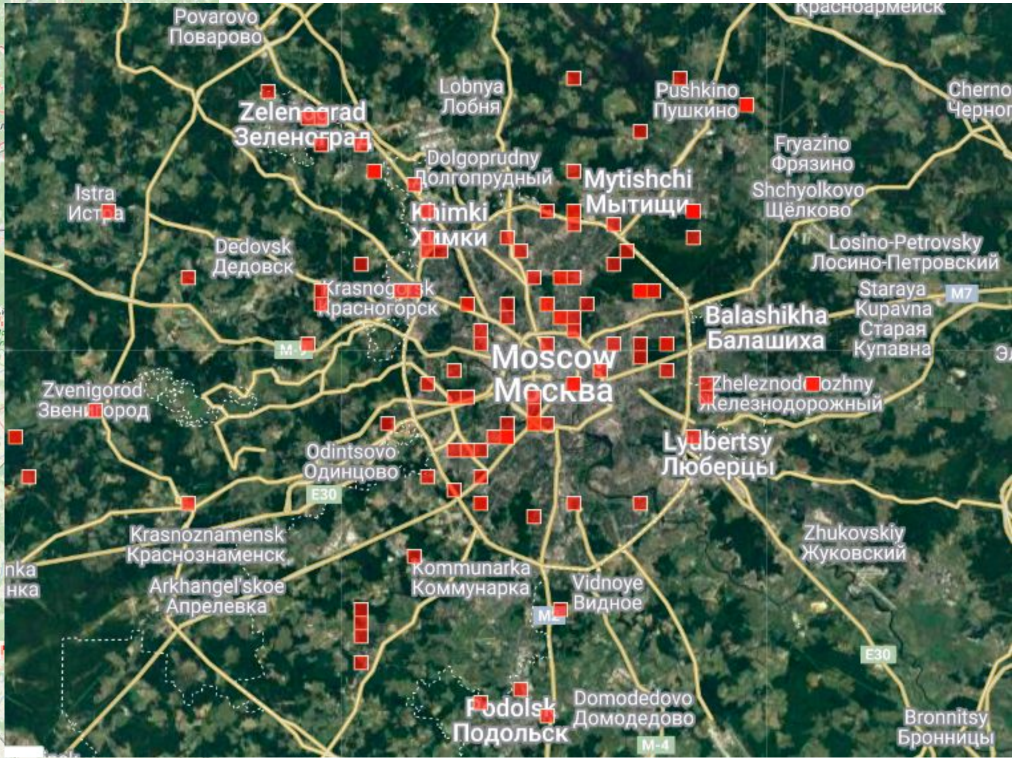
Interpretation issues: "Cultivated / Managed" class



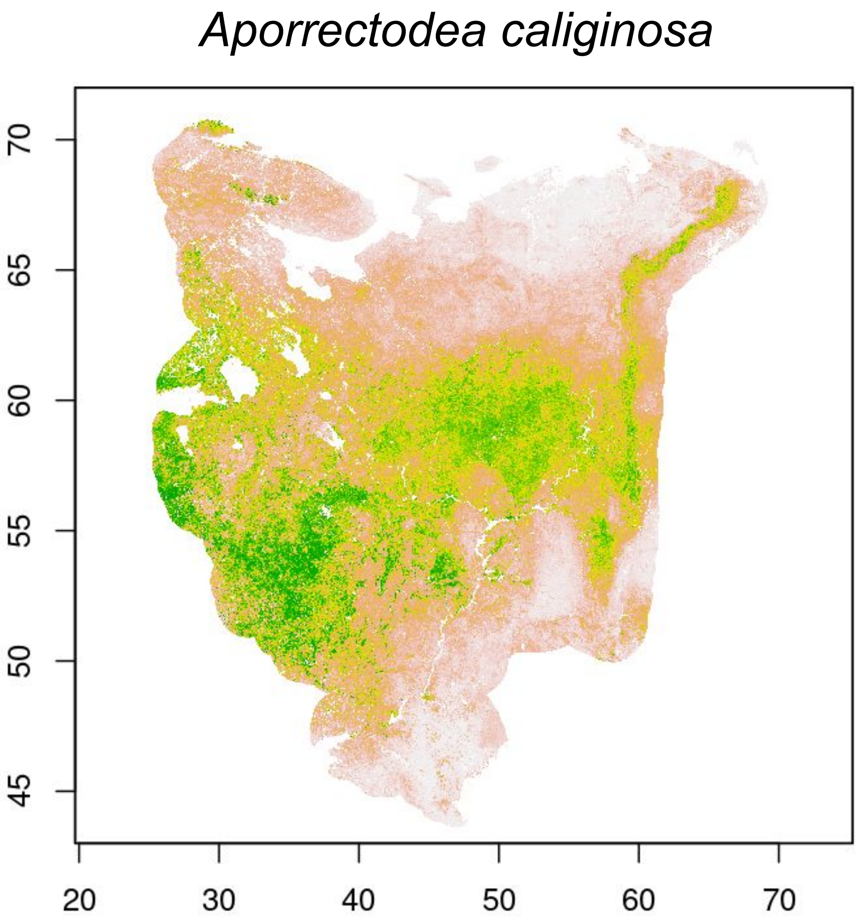
Land cover interpretation issues: "Urban" class

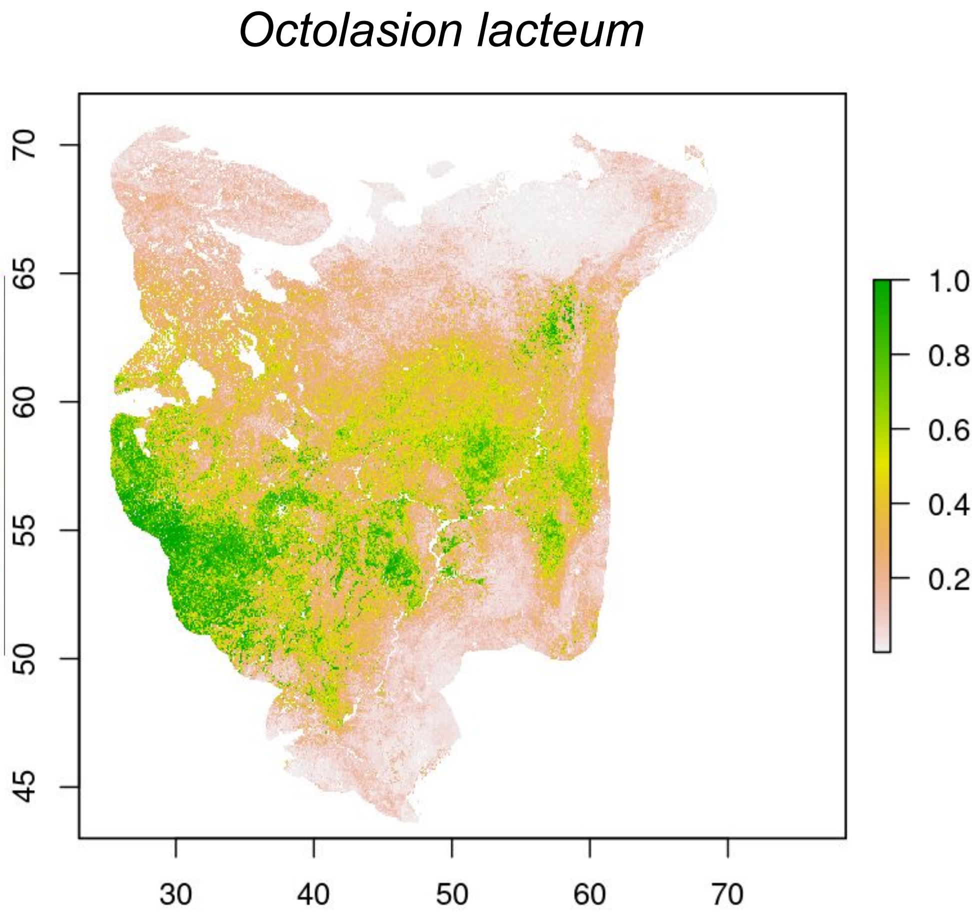


iNaturalist: *Lumbricus terrestris*



Soil Grid 250: soil properties with high precision

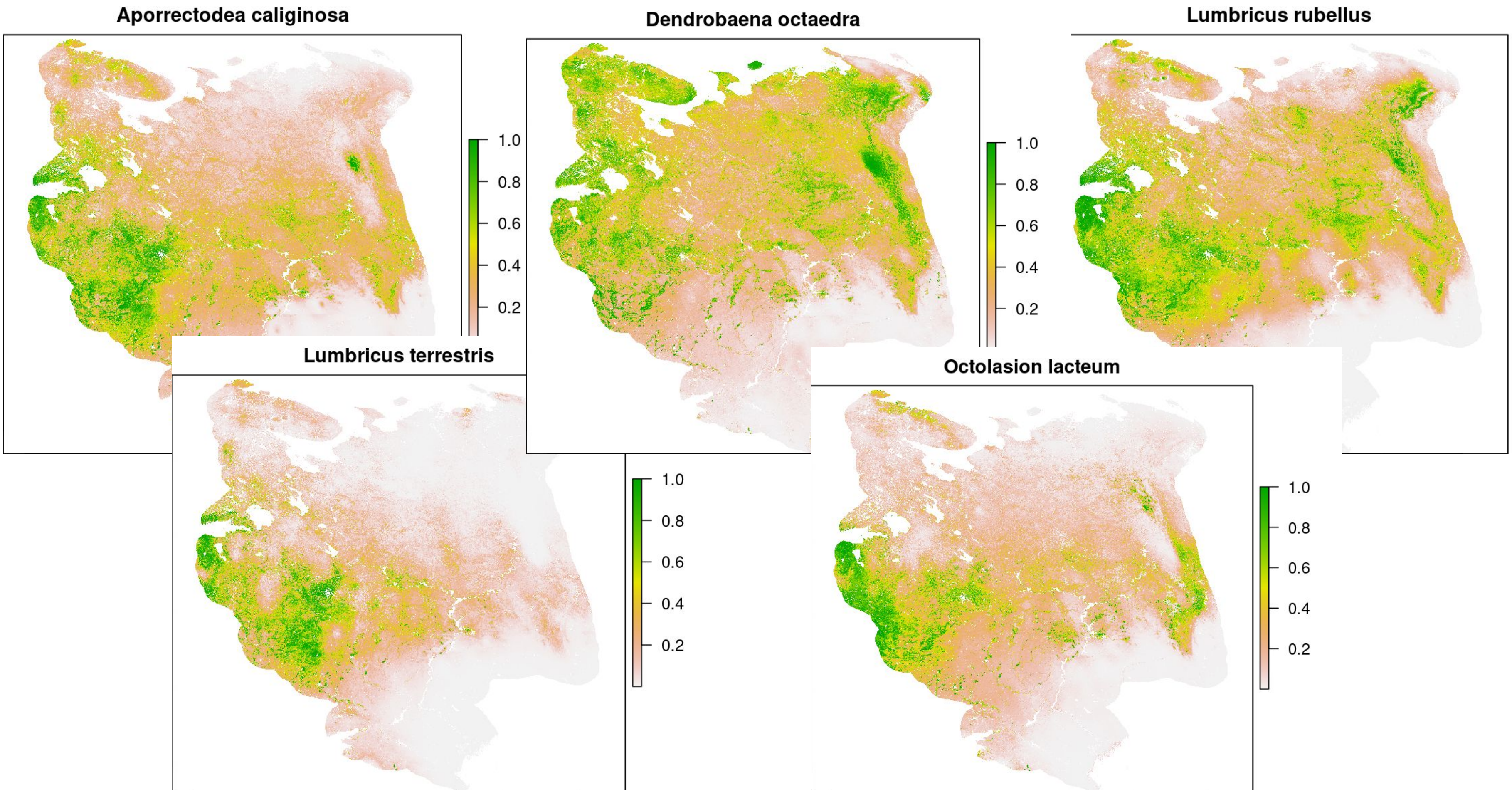




The Most Valuable Soil Properties Variables

	Ac	Do	Lr	Lt	Ol
Nitrogen content: 0-5 cm	24.7-34.4%	33.5-37.5%	27.6-47.0%	19.5-29.7%	13.9-18.0%
bulk density: 0-5 cm	32.9-34.4%	14.5-17.1%	25.1-38.0%	24.0-31.5%	24.9-27.3%
Cation exchange capability 5-15 cm	4.1-9.7%	1.3-8.7%	9.0-11.1%	8.3-17.5%	23.9-31.9%
Organic Carbon density: 30-60 cm	10.1-11.5%	3.2-4.3%	1.8-2.7%	6.3-8.8%	13.3-16.0%
Nitrogen content: 5-15 cm	0.5-1.1%	1.1-2.9%	1.1-1.6%	1.0-9.7%	0.6-1.1%
Organic Carbon density: 5-15 cm	2.0-11.6%	5.3-12.1%	3.5-4.8%	0.9-3.9%	1.7-3.7%
sand content	3.1-3.3%	2.7-12.0%	2.0-5.9%	8.7-13.7%	2.8-3.9%

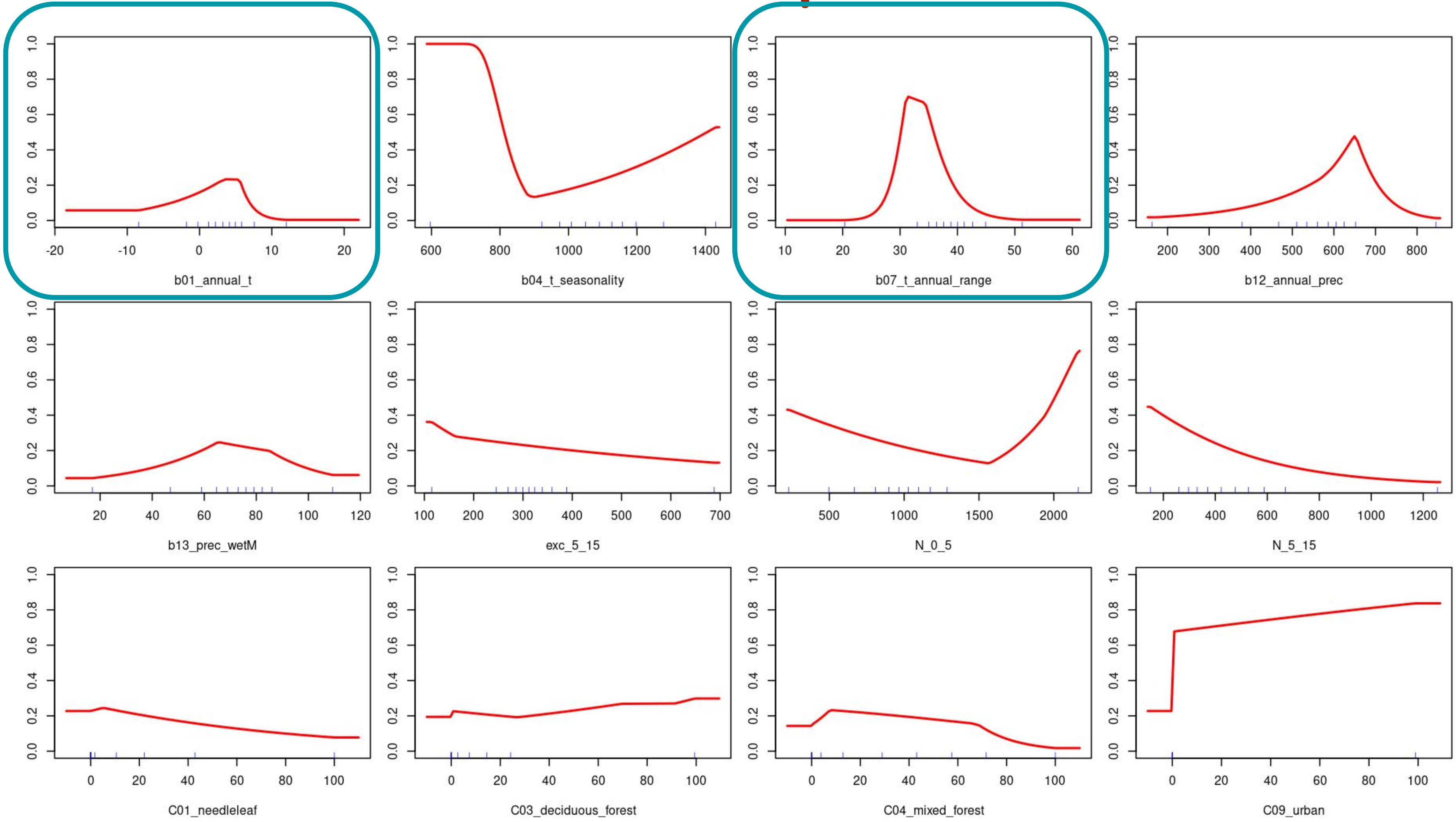
Resulting Models



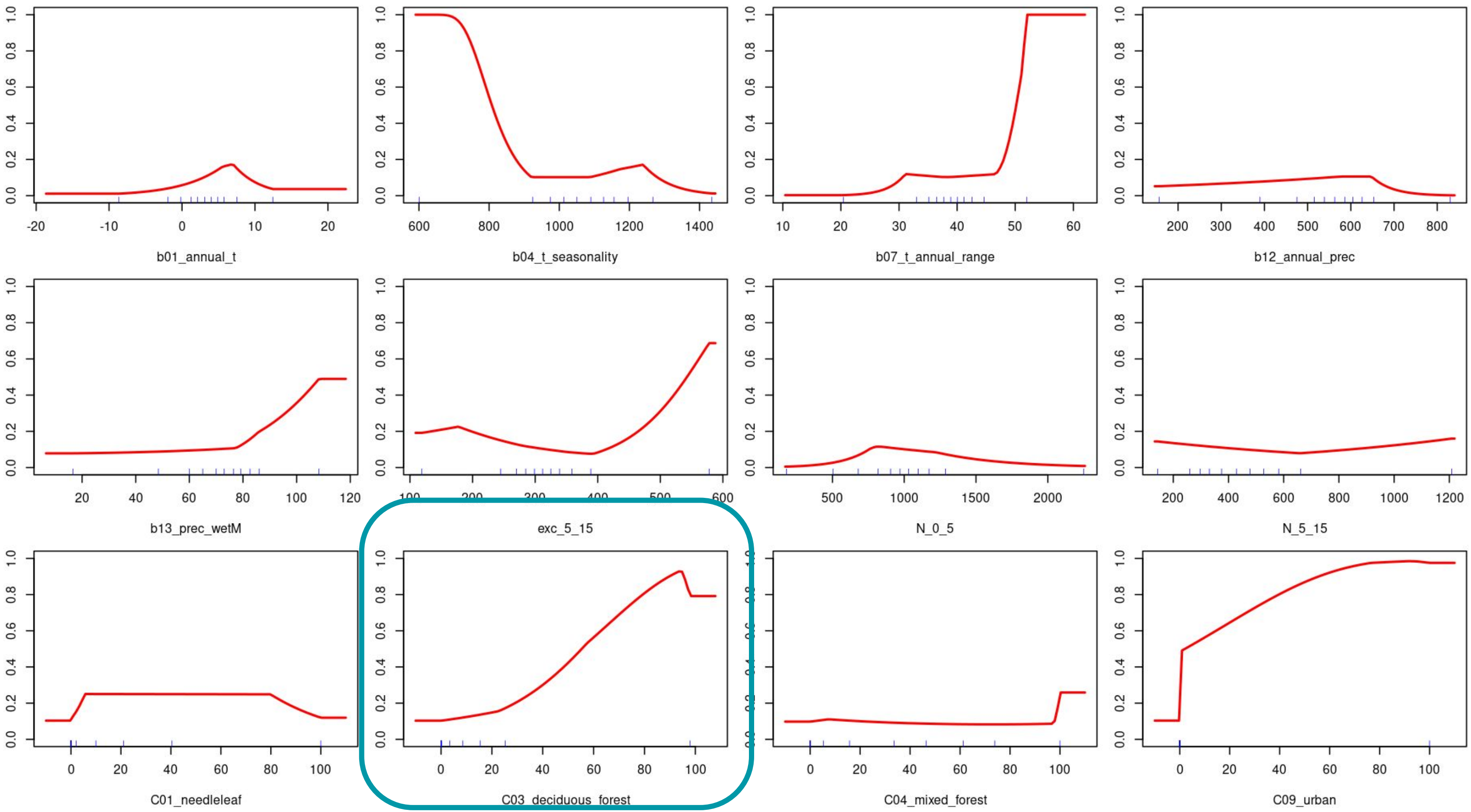
The Importance of Factors from the Overall Set

	Ac	Do	Lr	Lt	Ol
t annual range	14.1-22.1%	4.7-8.7%	4.6-13.0%	26.1-35.2%	1.4-7.6%
mean annual t	3.5-11.7%	3.4-9.5%	3.6-7.5%	14.6-23.0%	4.9-7.6%
ubran sites	18.0-22.0%	9.3-11.3%	24.8-31.1%	25.1-26.3%	16.1-16.2%
deciduous forest biotops	13.0-18.6%	11.5-14.4%	4.0-5.5%	2.4-6.8%	23.8-29.8%
t seasonality	2.8-5.2%	5.1-13.7%	1.8-7.3%	1.2-3.6%	2.7-5.9%
annual precipitation	8.4-12.1%	0.3-1.4%	12.9-22.1%	1.3-3.8%	8.4-11.5%
needleleaf forests biotops	2.0-5.6%	16.9-22.8%	2.7-11.2%	0.8-3.6%	2.7-7.2%
Nitrogen content: 0-5 cm	1.9-2.5%	13.6-15.7%	5.3-13.2%	0.5-2.6%	2.0-5.2%

Lumbricus terrestris: response to variables



Octolasion lacteum



Conclusions

1. The World Climate dataset is not enough as a set of environmental variables.

2. High model performance does not ensure reliable spatial prediction. The results should be evaluated according to your expert opinion.

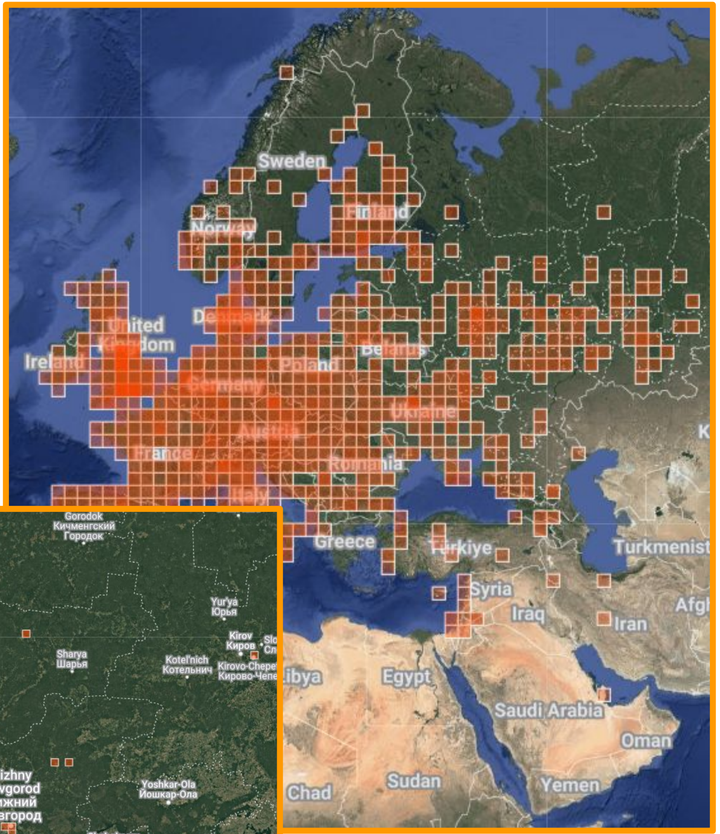
3. For EW modelling, alongside the climatic variables, we should use biotopic / land use classification and soil property layers. This is likely to be considered for other groups of soil macrofauna as well.

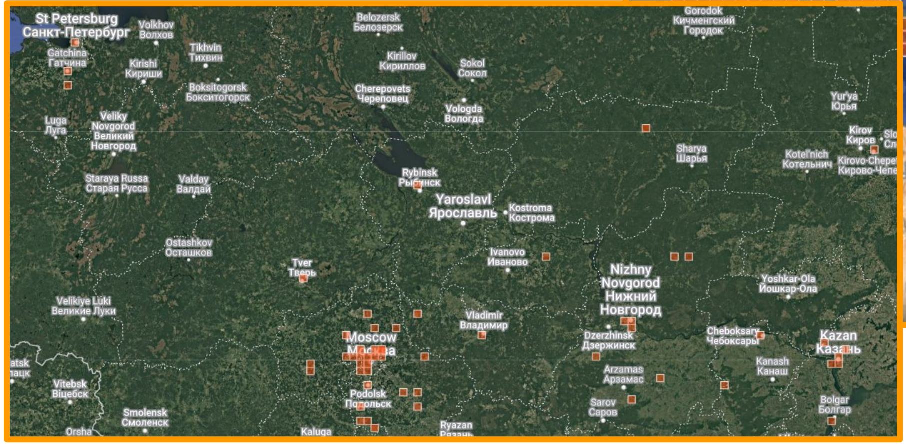
4. Different important predictors for particular EW species apparently show their different ecological properties.

Focal species - *Lumbricus terrestris* Linnaeus, 1758

iNaturalist - the global citizen science system

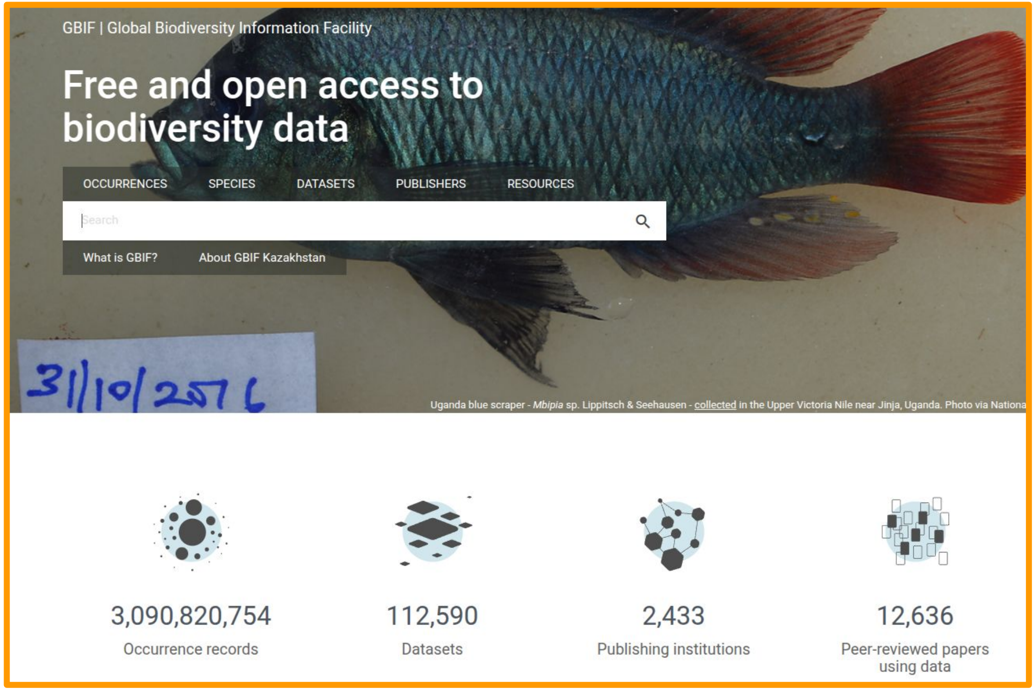
> 3.5 M observers
> 230 M observations
> 500 k species
Crassiclitellata
~112 k observations
163 species
Lumbricus terrestris
> 36 k observations





Citizen Science Data Prevailed

observations
> 63 %



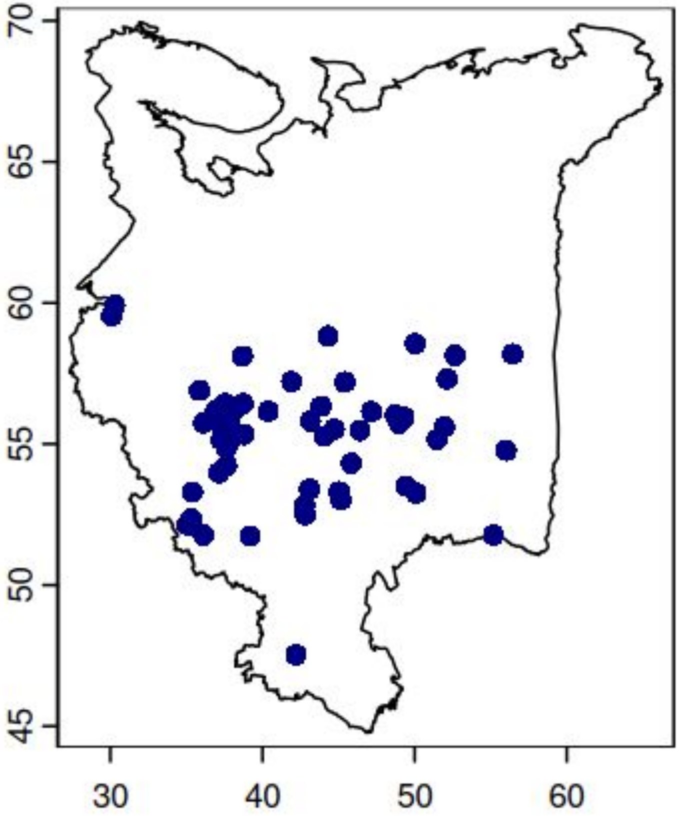
biological
collection
~ 8.6%

Research goal

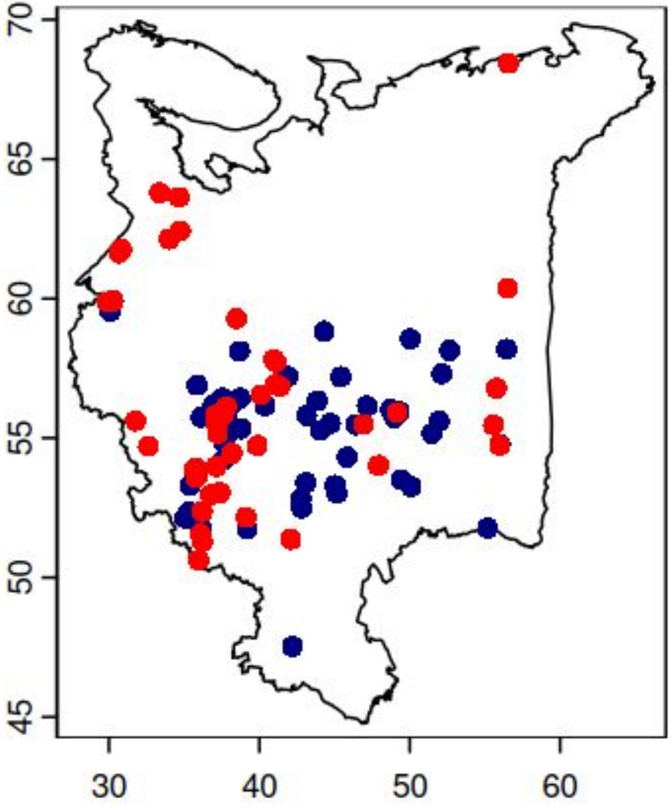
Using the example of the common earthworm species *Lumbricus terrestris*, to examine how data from the citizen science platform iNaturalist influence the results of species distribution modelling

Lumbricus terrestris occurrences

iNaturalist: 146 points

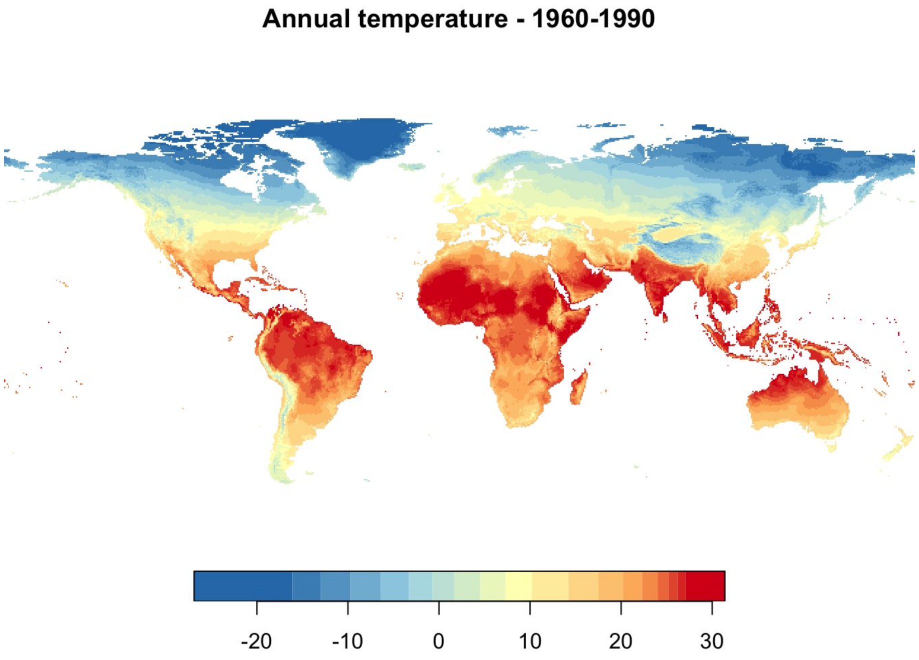


+ literature and collections 64 points



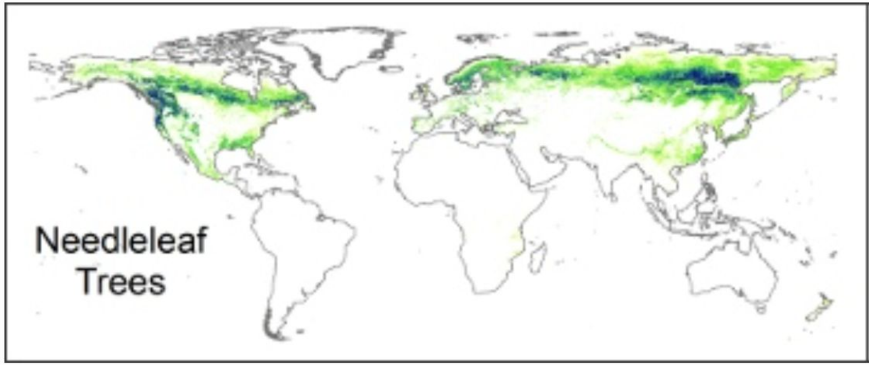
Data sources - environmental variables

climatic conditions
WorldClim



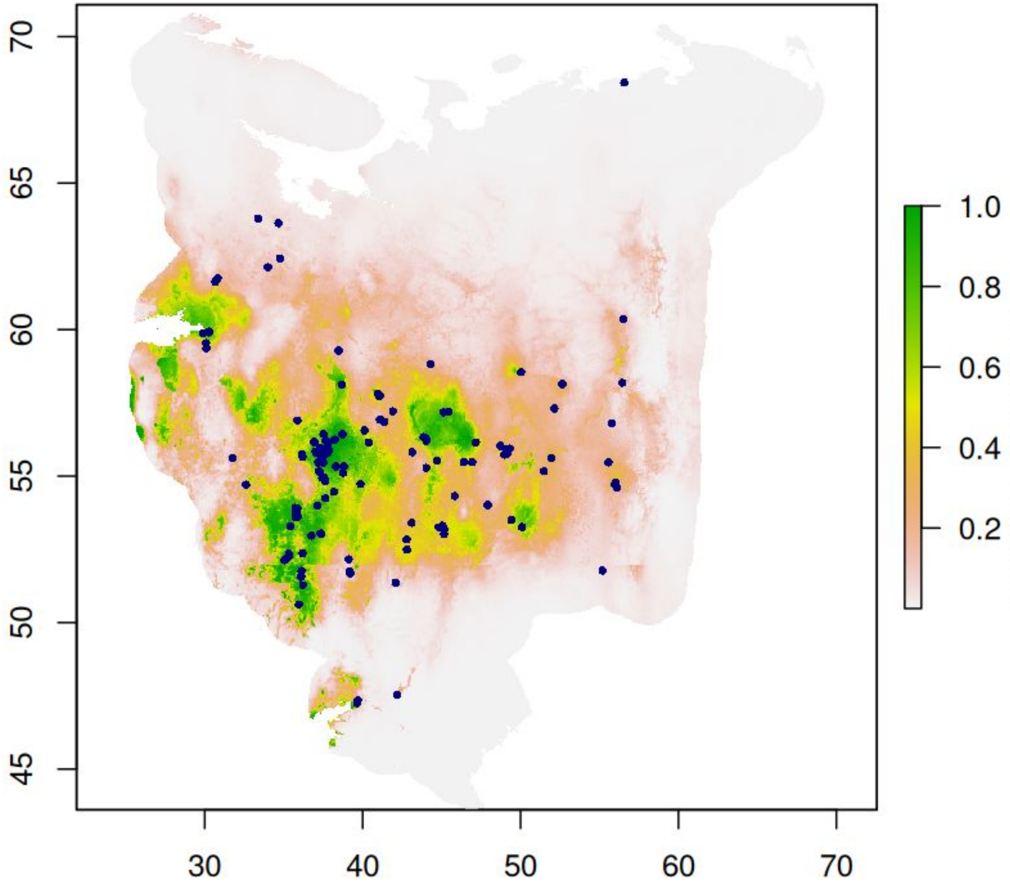
19 climatic layers + BHYM SRTM

Land cover classification
Consensus 1km Global Land Cover

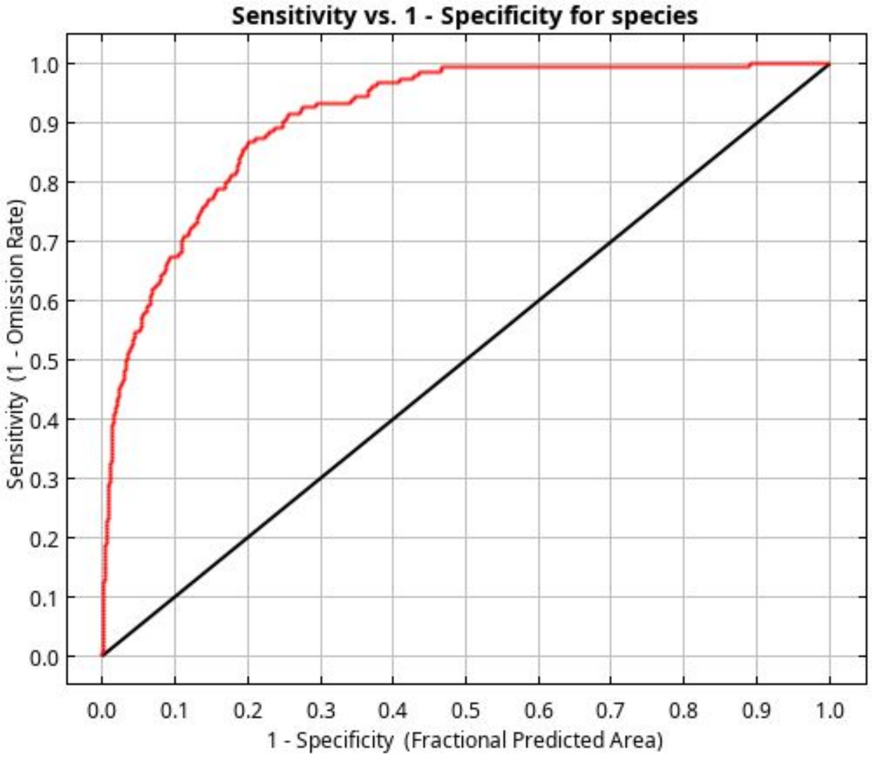


a set of raster layers reflecting the quantitative representation of 12 types of biotopes, including cultivated land and urbanised areas

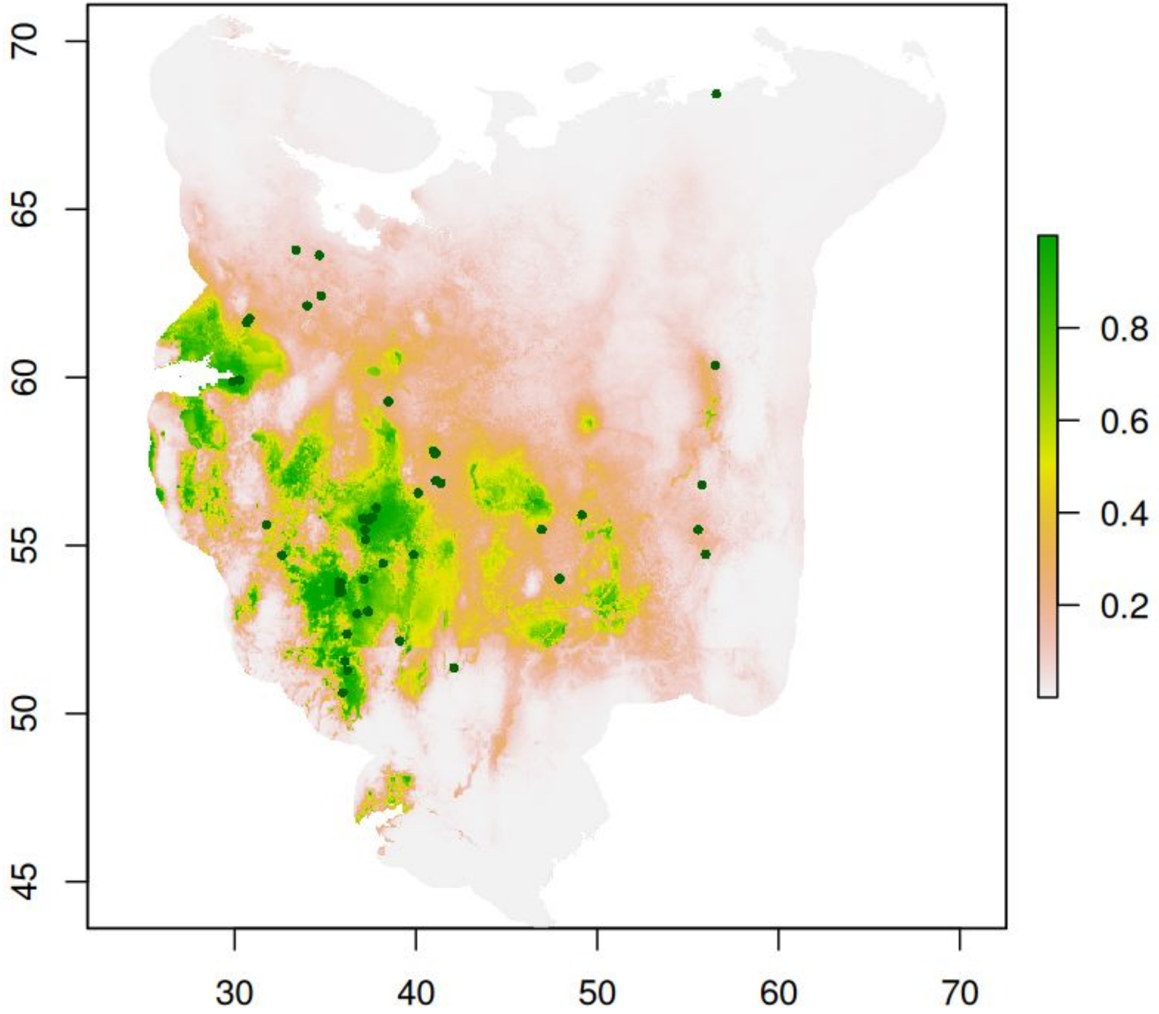
Total number of occurrences against WorldClim



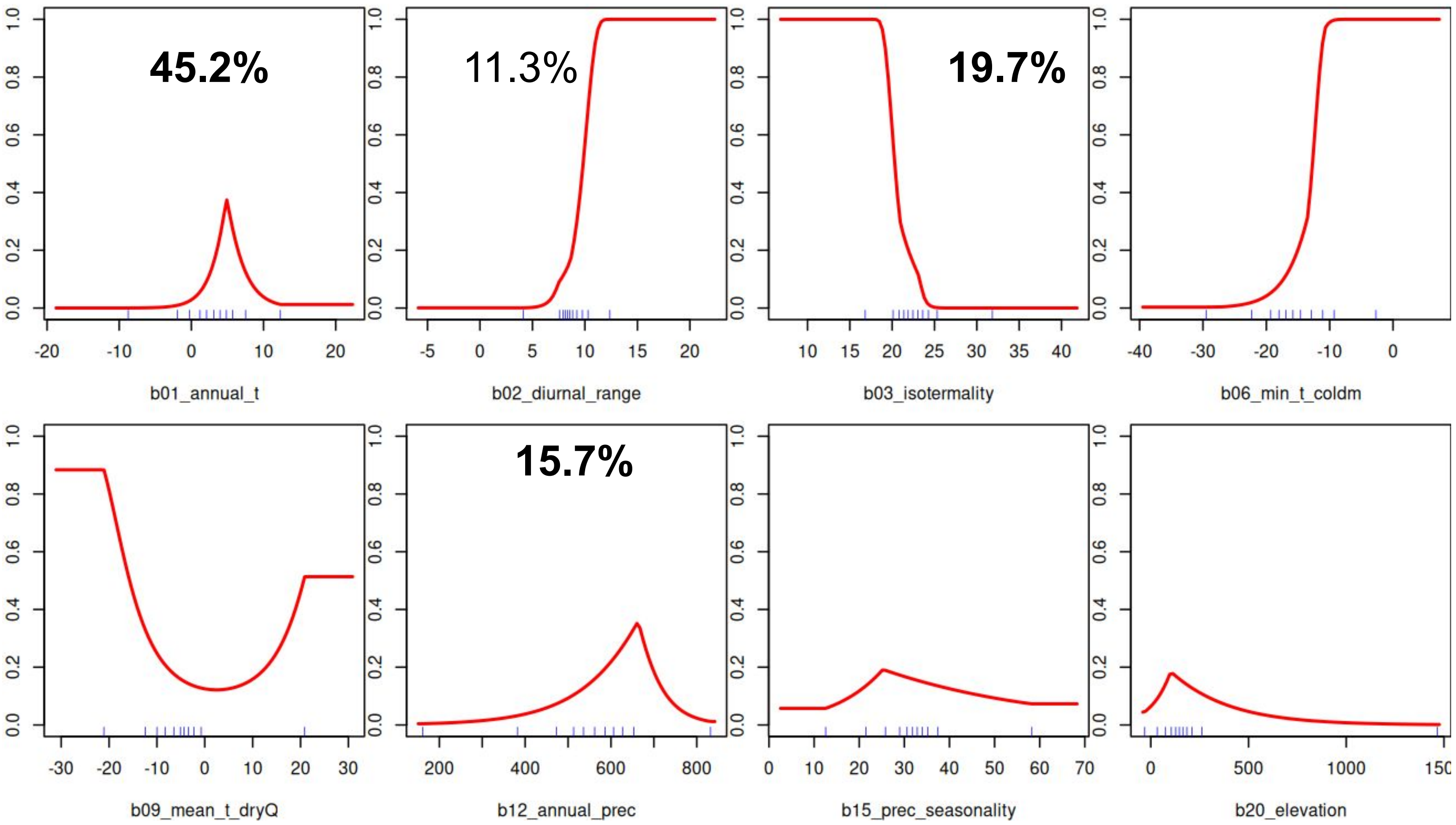
AUC = 0.909



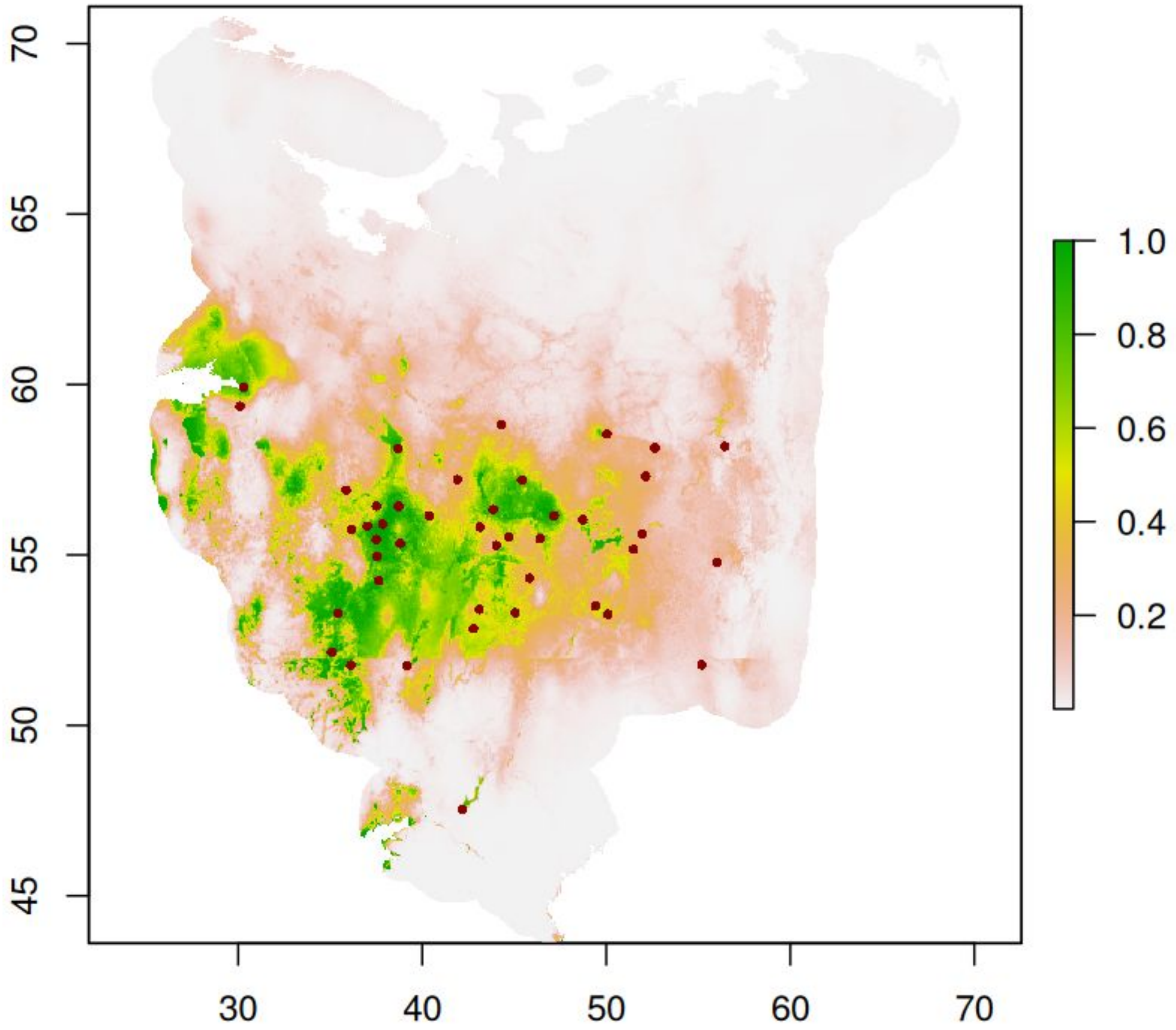
Occurrences without iNaturalist against WorldClim



Response of habitat suitability for particular predictors



iNaturalist only occurrences against WorldClim



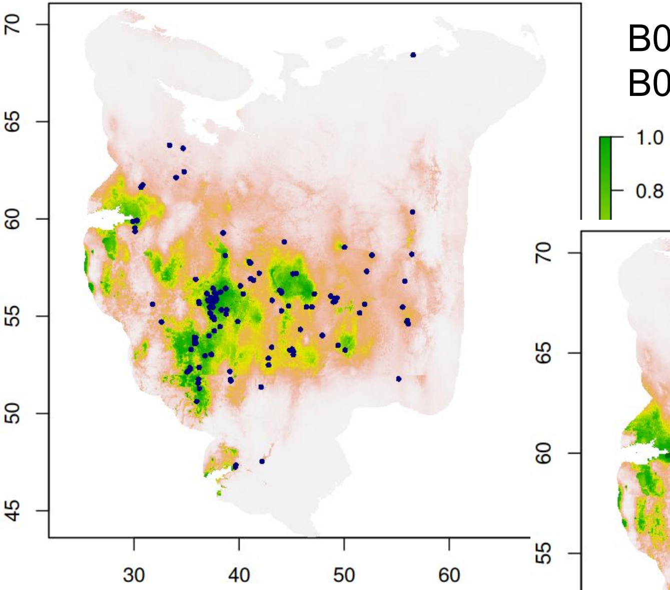
point were thinned with tolerance of 30 km

Modelling results based on WorldClim

all occurrences

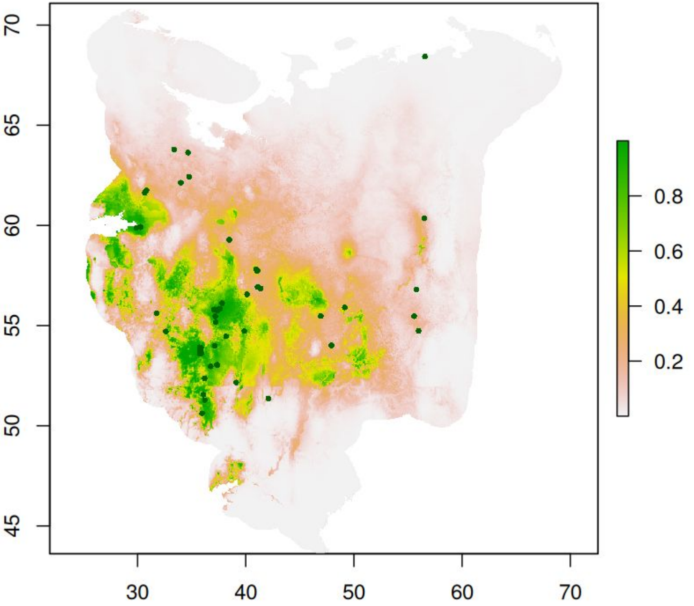
without iNaturalist

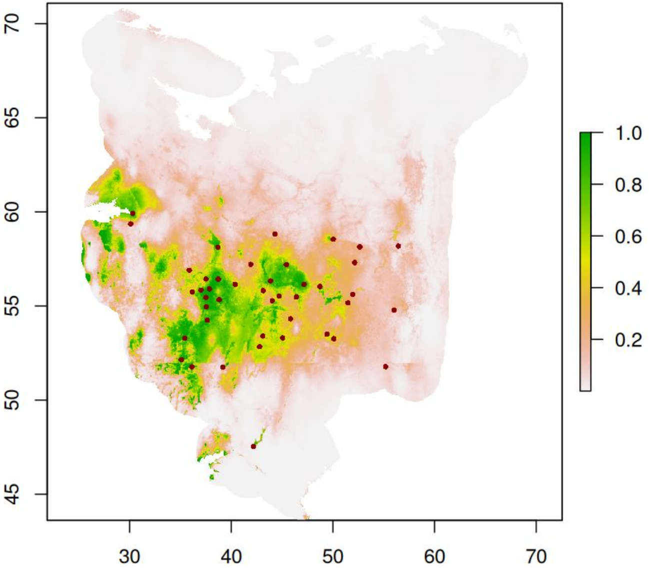
iNaturalist only



B05 - Max Temperature of Warmest Month

B01 - Annual Mean Temperature
B03 - Isothermality

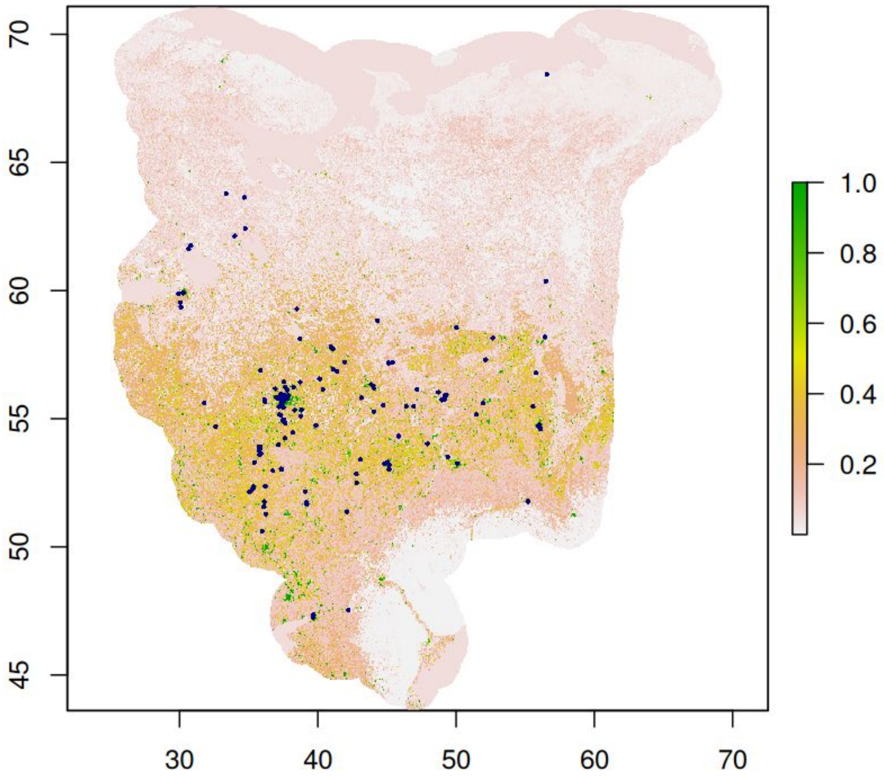




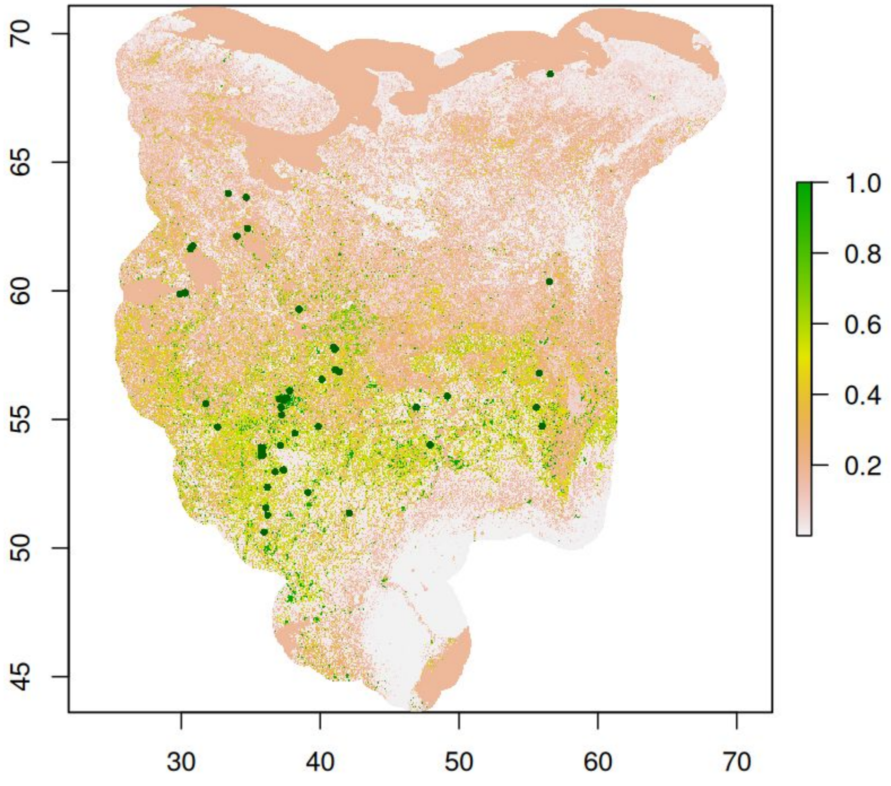
B02 - Mean Diurnal Range
B03 - Isothermality
B05 - Max Temperature of Warmest Month

Modelling results based on Consensus Land Cover

all points

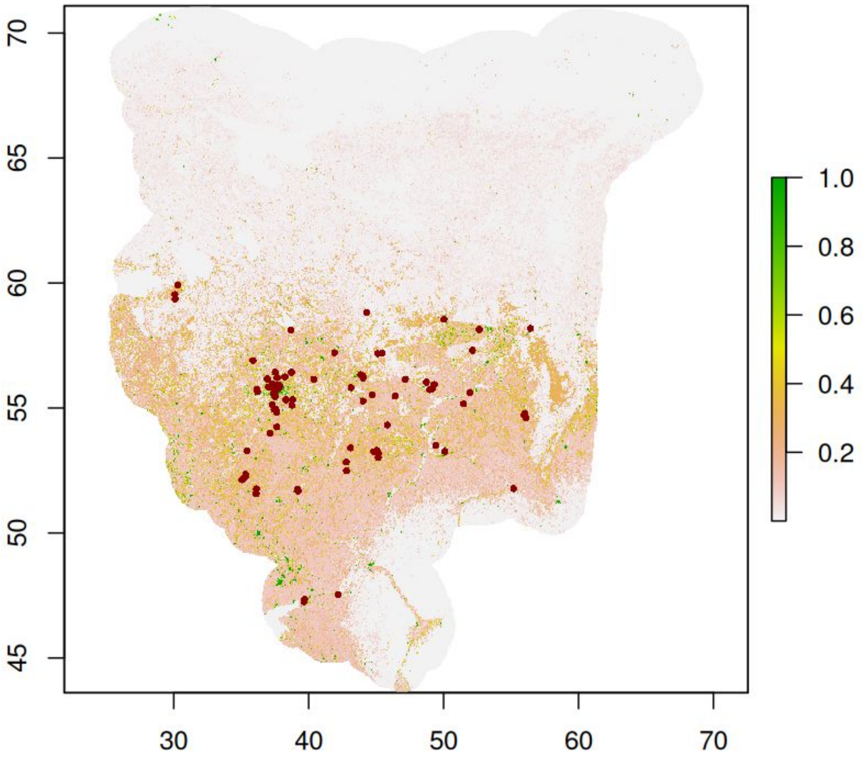


without iNaturalist

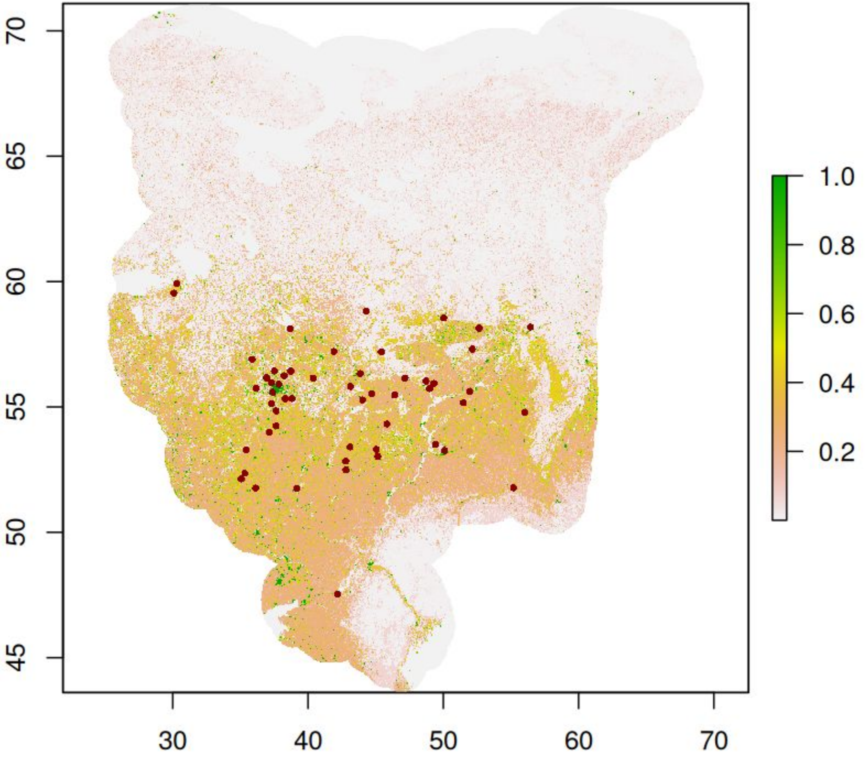


Modelling results based on Consensus Land Cover

all iNat occurrences

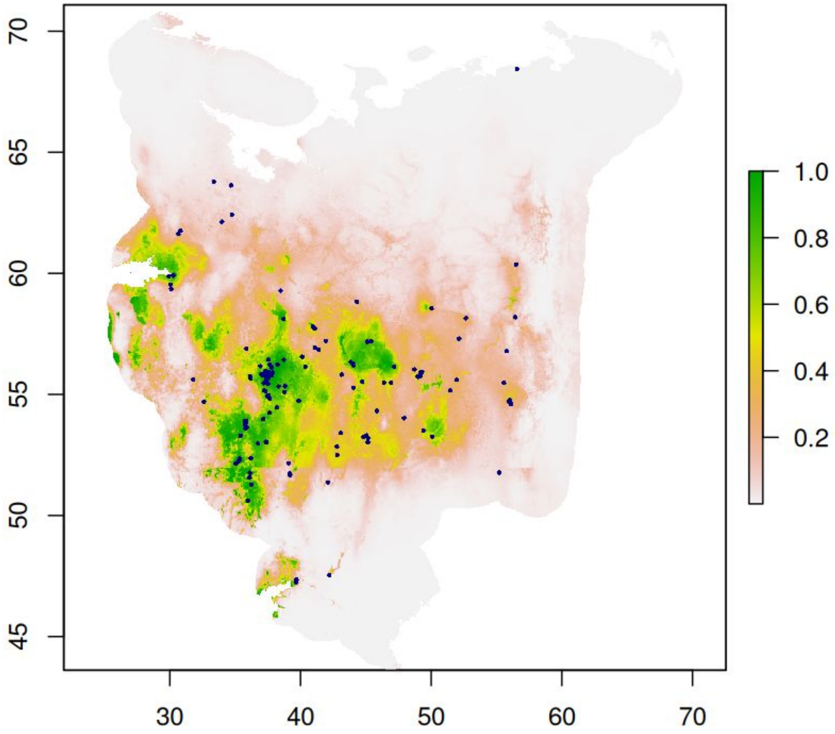


thinned with threshold of 30 km

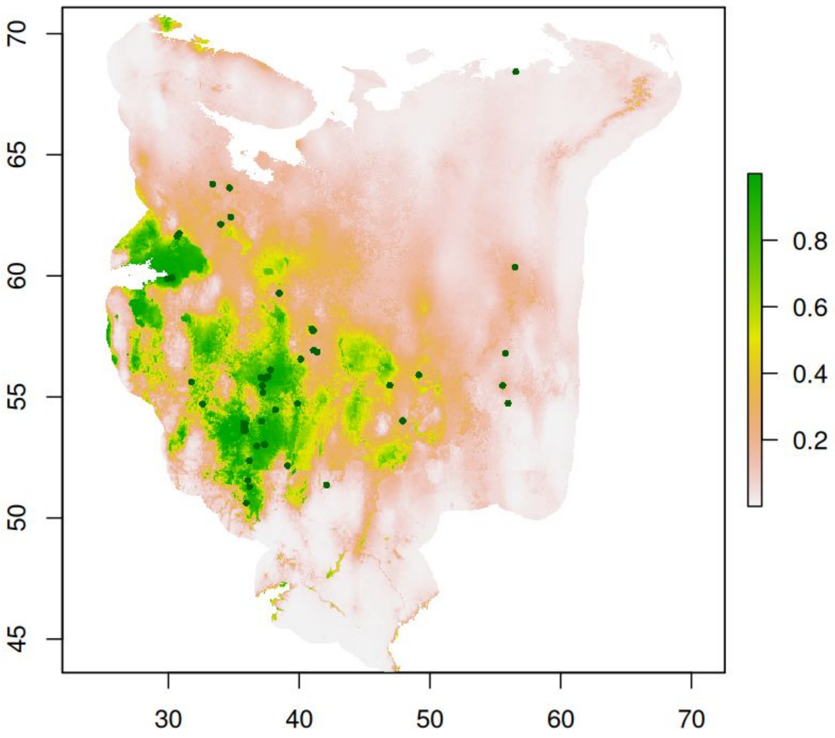


Modelling results - WorldClim + Consensus

all points

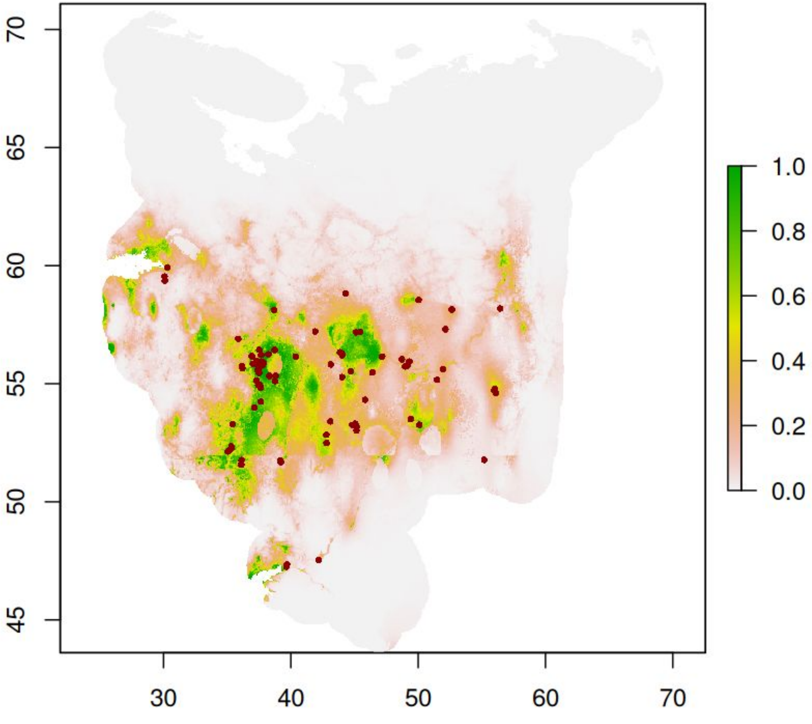


without iNaturalist

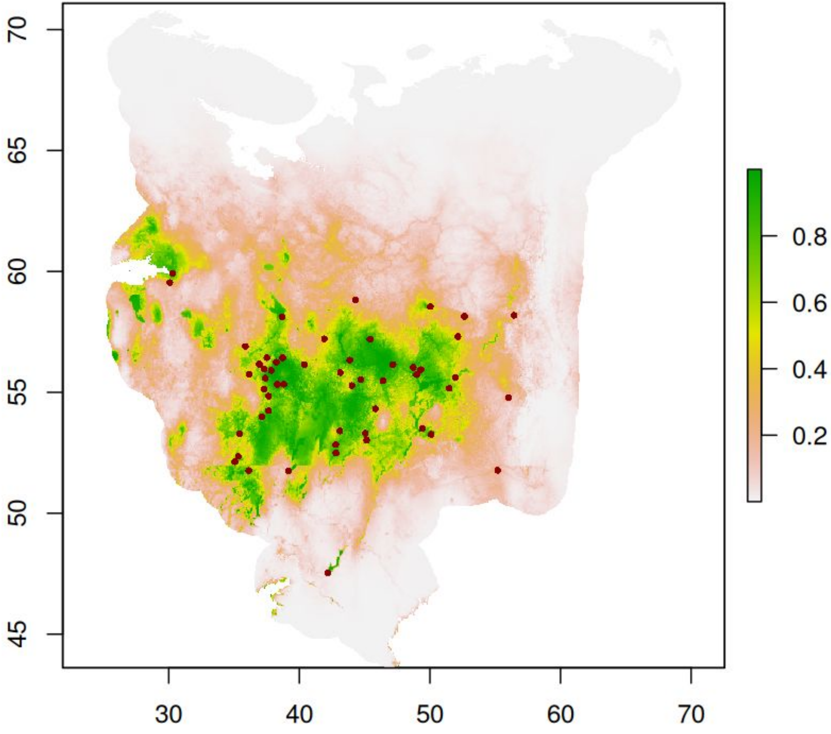


Modelling results - WorldClim + Consensus

all iNat occurrences



thinned with threshold of 30 km



Conclusions

Climatic variables that have the greatest impact on habitat suitability are **average annual air temperature** (bio01), **average daily temperature range** (bio02), **isothermality** (bio03), and **maximum temperature of the warmest month** (bio05)

Precipitation within this territory does not have a strong influence; in some cases, the contribution to the model > 10% is made by the **sum of precipitation in the coldest quarter** (bio19)

Of the biotopic conditions, urbanized areas (as expected) and agricultural land are the most important

There are enough points from iNaturalist to build a reliable model with proper processing of the input data

Thank you for your attention!

All input datasets used are available online. The whole modelling process was implemented with the R programming language, and the scripts can be shared by request.