

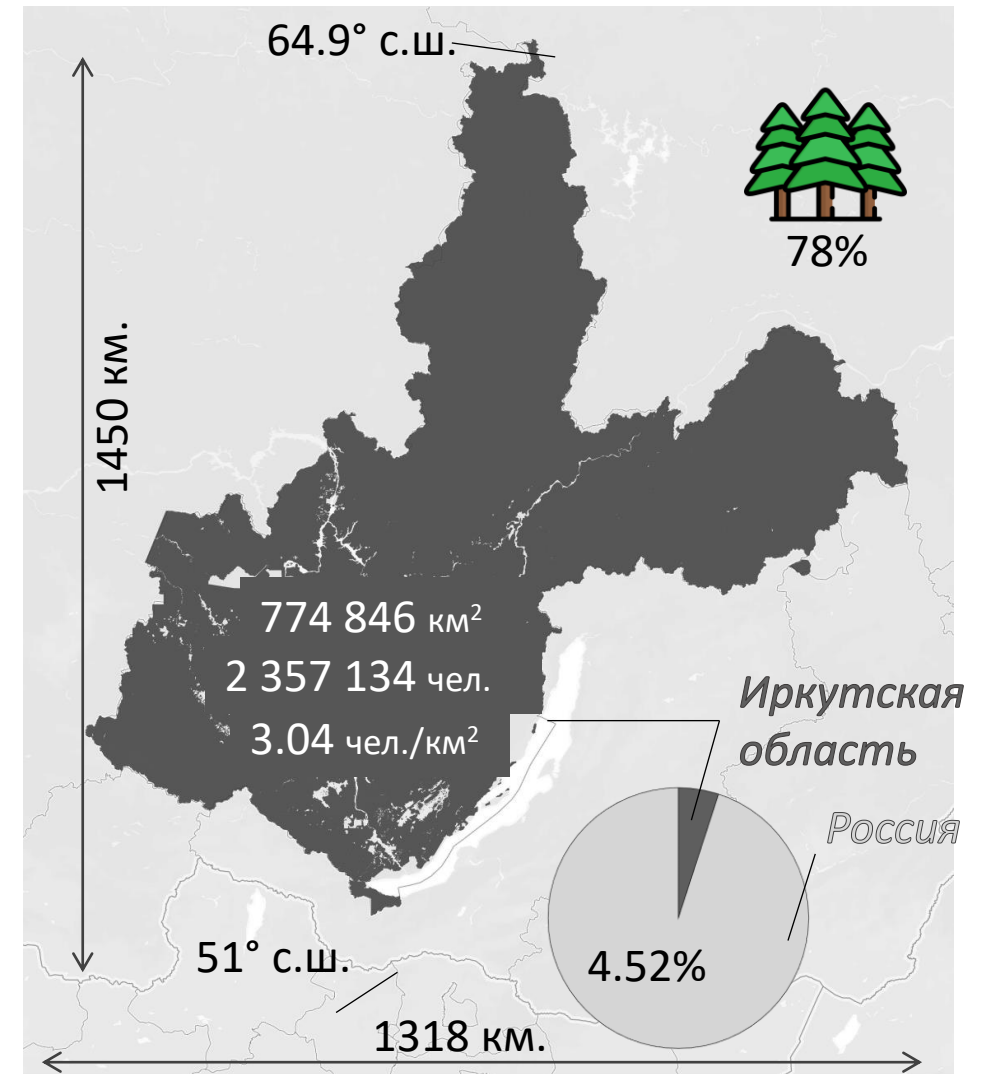
Картографирование пожароопасности Иркутской области с использованием метода машинного обучения Random Forest

Пестова Ю.В. (yupest@icc.ru),
Дородых Н.О.,
Николайчук О.А.,
Юрин А.Ю.

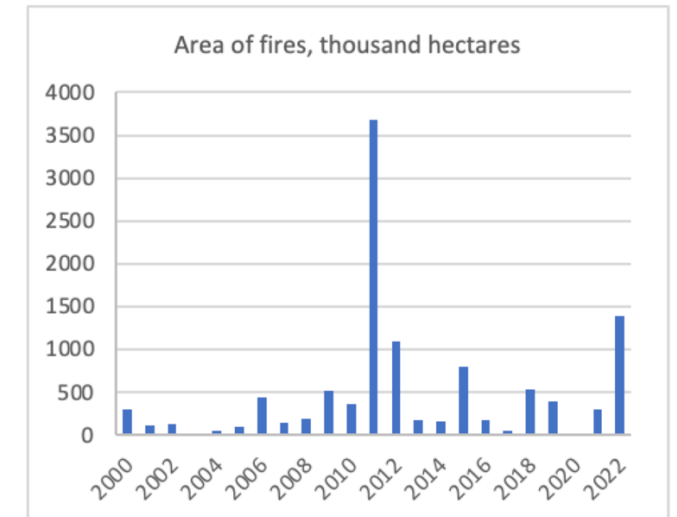
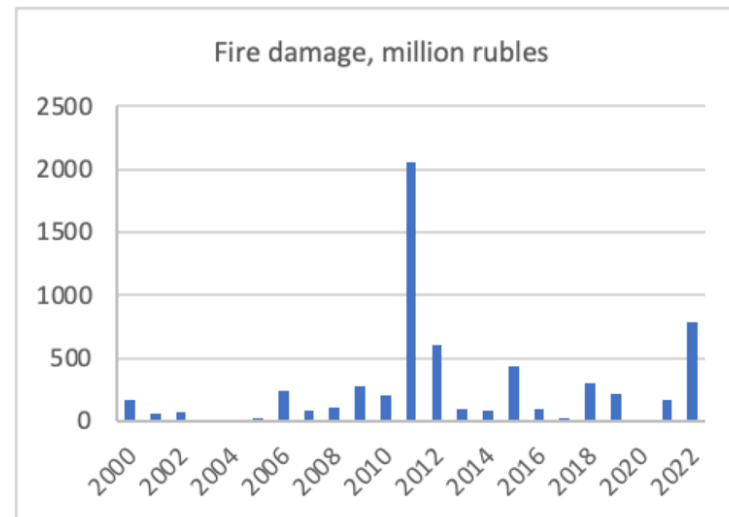
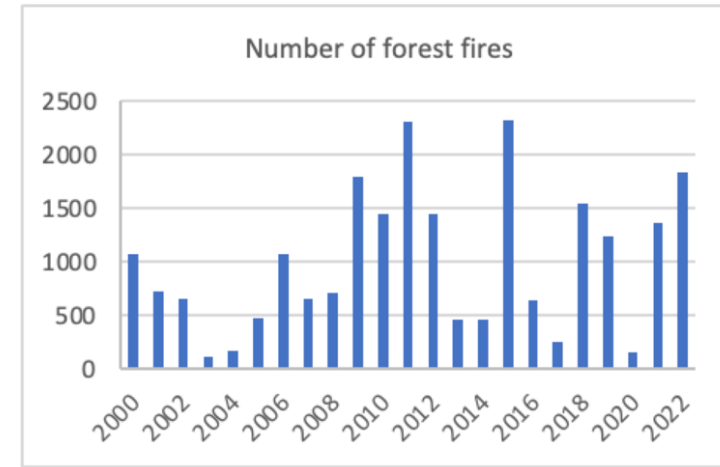
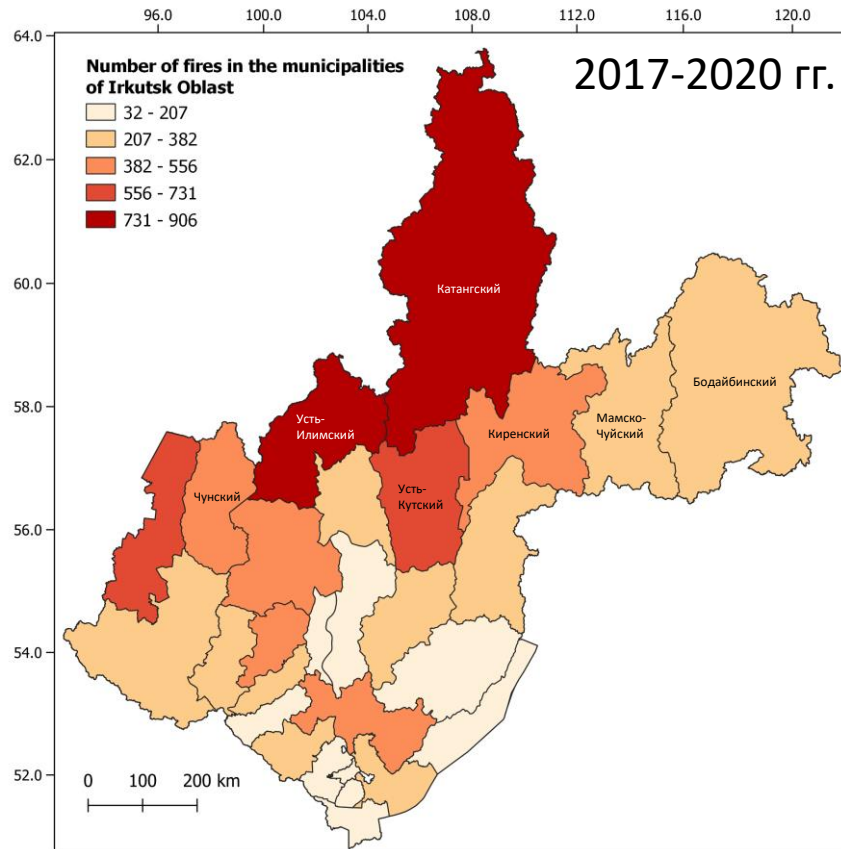


Объект исследования

- Областью исследования является Иркутская область — крупнейшая область Российской Федерации. Его площадь составляет 774 846 км². (4,52% территории России), с населением 2 357 134 человек и плотностью населения 3,04 чел./км² (2022).
- Иркутская область расположена в Восточной Сибири. Крайняя южная точка региона расположена на 51° северная широта, а северная оконечность почти достигает 65-й параллели.
- С севера на юг регион простирается почти на 1450 км, а с запада на восток - на 1318 км. Юго-восточная граница Иркутской области проходит вдоль озера Байкал. Регион занимает юго-восточную часть Центрально-Сибирского плоскогорья, плато и хребты которого имеют высоту от 500 до 1000 м.
- Леса занимают 78% территории.
- В Иркутской области 32 административных района.



Актуальность



Массовые лесные пожары вызвали значительное изменение ландшафтов водосбора озера Байкал и сокращение лесных запасов. Наибольшие площади природных пожаров в среднем за 10 лет отмечены в Катангском, Киренском, Мамско-Чуйском, Бодайбинском, Боханском, Усть-Кутском и Чунском районах.

Основные направления исследований лесных пожаров

- Обнаружение пожаров и их картографирование;
- Исследование метеорологических условий и изменений климата, влияющих на возникновение пожаров;
- Определение вероятности и риска пожара;
- Прогнозирование поведения пожара;
- Анализ последствий пожаров;
- Управление пожарами.

Цель исследования

Целью исследования является картирование пожароопасности (определение вероятности и риска пожара) Иркутской области на основе данных дистанционного зондирования, метеорологии, органов государственного управления в области лесного хозяйства и чрезвычайными ситуациями с использованием метода машинного обучения Random Forest.

Метод Random Forest (RF)

- **Random Forest** — алгоритм машинного обучения, основанный на использовании ансамбля решающих деревьев, каждое из которых даёт очень невысокое качество классификации. Алгоритм применяется для задач классификации, регрессии и кластеризации [Breiman].
- Данный метод активно применяется для решения задачи картирования пожароопасности. В исследованиях было продемонстрировано, что RF обладает высокой точностью прогнозирования и высокой устойчивостью к выбросам и “шуму”. Некоторые примеры достигнутых результатов представлены в таблице.

Территория	Точность	Ссылка на источник
Yunnan Province (China)	84,36	Zhang G., 2019
European Mediterranean region	96,3	Oliveira S., 2012
Global ecosystem	78,33	Luo R., 2013
China’s boreal forest, located in the Daxing’an Mountains of northeastern China	70,1	Guo F., 2016
Yunnan Province (China)	88,3	Cao Y., 2017
Hunan Province (China)	91,68	Chaoxue T., 2023
Ethiopia	67,2	Harris L.B., 2023

Этапы применения метода RF

- Выбор (обоснование) факторов, влияющих на возникновение лесного пожара.
- Сбор данных о пожарах и факторах для заданной территории (класс «наличие пожара»).
- Генерация данных, для формирования класса «отсутствие пожара» с учетом пространственных и временных критериев.
- Подбор параметров метода RF (использование рекомендаций из литературных источников).
- Применение метода RF. Возможно изменение параметров метода.
- Анализ точности решения задачи.
- Оценка риска.
- Визуализация результатов применения метода.
- Анализ результатов применения метода.

Описание исходных данных о пожарах

- База данных включает более 45 000 записей, описывающих информацию о термальных точках, выявленных в результате анализа спутниковых снимков за 2017-2020 годы, полученные со спутников NOAA серии «Алиса-SC» производства ООО «Сканэкс» и в результате анализа на платформе FirePro (ИСЗФ СО РАН).
- Термальная точка - это значительное повышение температуры на поверхности Земли, зарегистрированное во время прохождения спутника по сравнению с соседними участками. Тепловая точка может указывать на сжигание мусора, искусственный процесс, падение, искусственный пожар или лесной пожар. Тепловая точка фиксируется спутником в виде объектов полигонального типа (набор четырехугольников).



Предварительная обработка исходных данных

- Для обработки использован программный модуль GeoAnalytics, разработанный на Python3 с использованием библиотек “shapely”, “pyproj”, “pandas” и “geopandas” для работы с геообъектами, а также ГИС “QGIS”.

- Обработка данных термальных точек:

- Определены термальные точки, расположенные в границах промышленных зон, населенных пунктов и зон добычи полезных ископаемых, которые не являются лесными пожарами, с последующим исключением их из рассмотрения. Оставшиеся термальные точки позже были интерпретированы как лесные пожары.



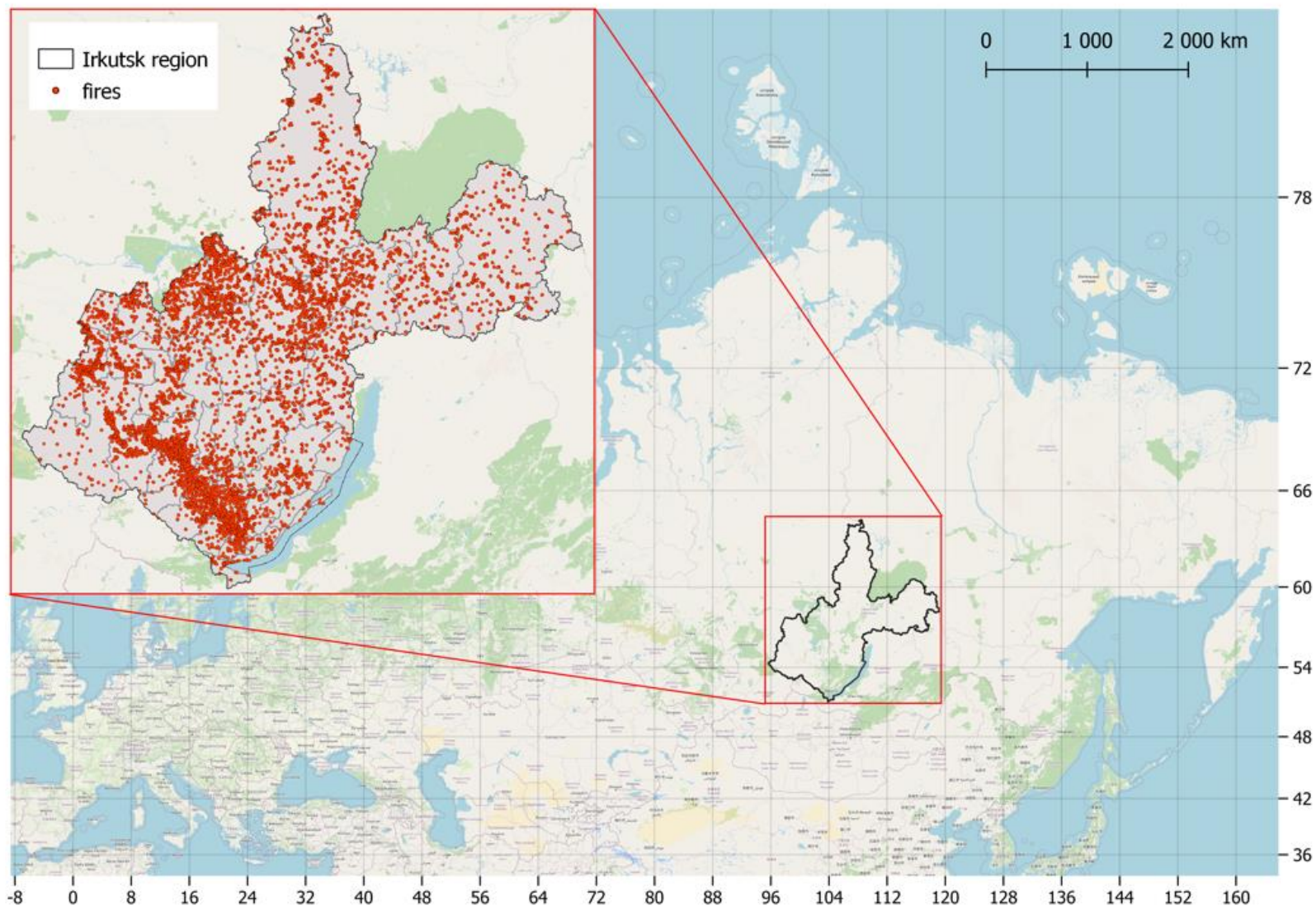
- Сгруппирована (агрегирована) информация о пожарах путем объединения данных, характеризуемых местоположением пожара на пересекающихся полигонах и минимальным интервалом времени, с дальнейшим определением продолжительности, минимальной и максимальной площади обнаруженного пожара.



- Выделены начальные события (возгорания) лесного пожара с определением даты обнаружения, координат и полигона пожара.

Исходные данные о пожарах

9001 пожар
В период с
2017-2020 гг.
март-ноябрь



Факторы, влияющие на лесные пожары

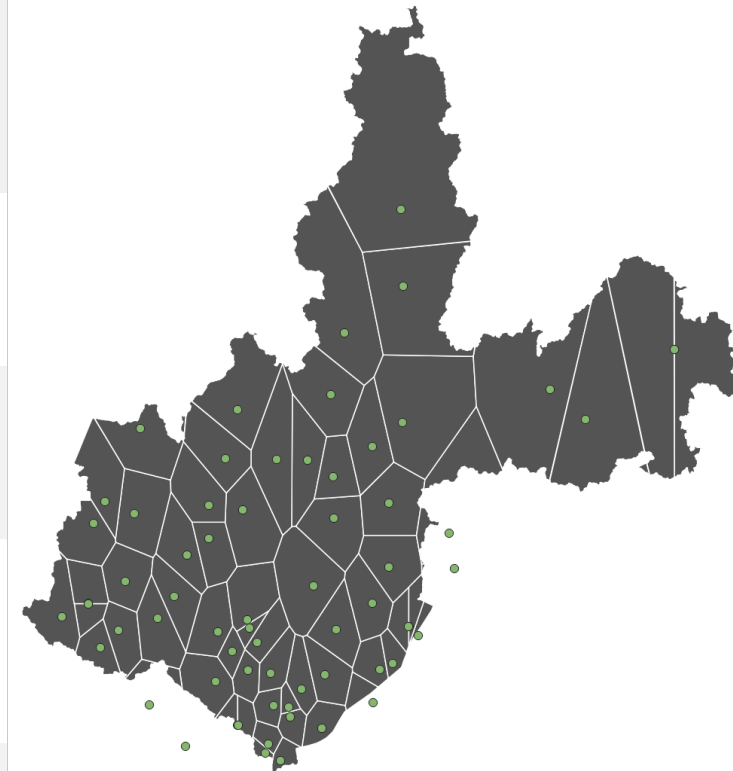
В исследованиях среди факторов влияющих на возникновение пожаров отмечают следующие группы:

- метеорологические,
- топографические,
- характеристики растительности и
- деятельности человека (социальные факторы).

Метеорологические факторы

ФАКТОРЫ	ЕДИНИЦЫ ИЗМЕРЕНИЯ	СПОСОБ ПОЛУЧЕНИЯ: данные метеостанций
Дальность горизонтального обзора (VV)	Километры	определяется на уровне станции
Относительная влажность (U)	%	определяется на высоте 2 метра над поверхностью земли
Атмосферное давление (P)	миллиметры ртутного столба	определяется на уровне станции
Скорость ветра (Ff)	метры в секунду	определяется на высоте 10-12 метров над земной поверхностью
Погодные явления (WW_code)	гроза, туман, дождь, дымка, снег, облачно, морось, пыль, град	данные метеостанций

<https://rp5.ru/>

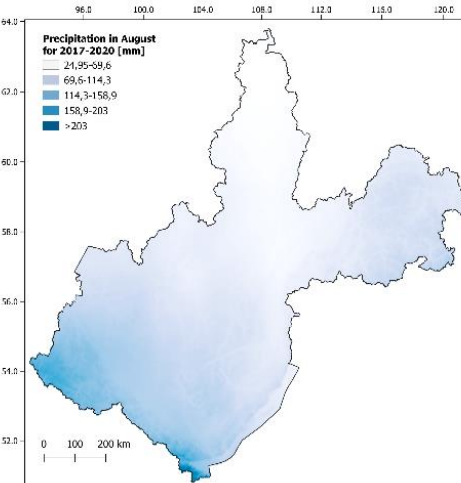
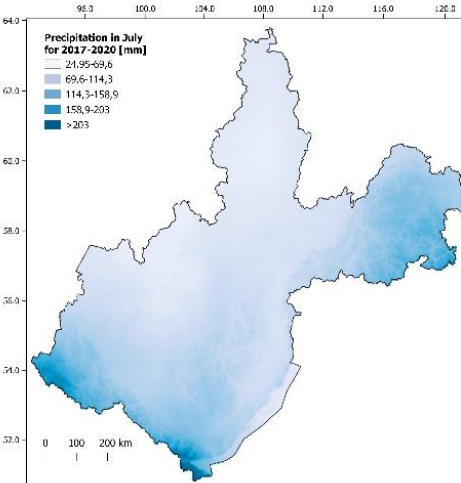
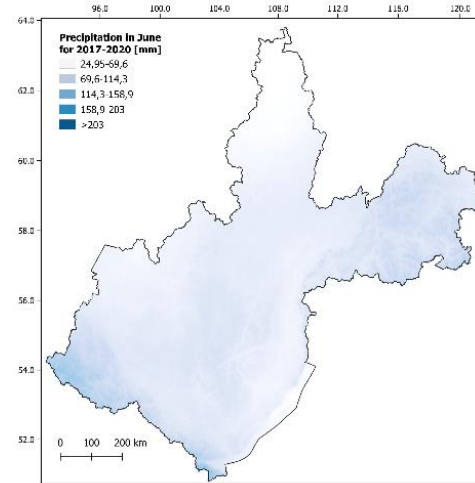


Метеорологические факторы: среднемесячные температура и осадки

<https://worldclim.org/>

**Количество
выпавших
осадков
(RRR)**

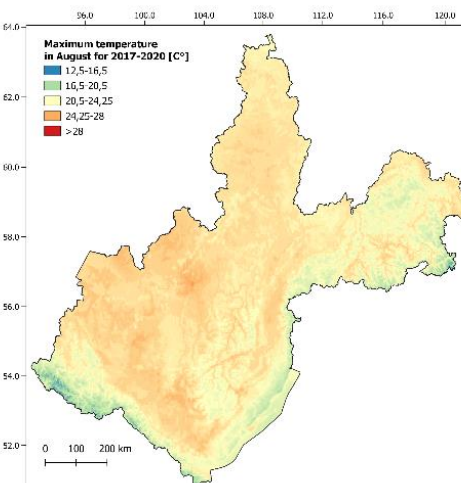
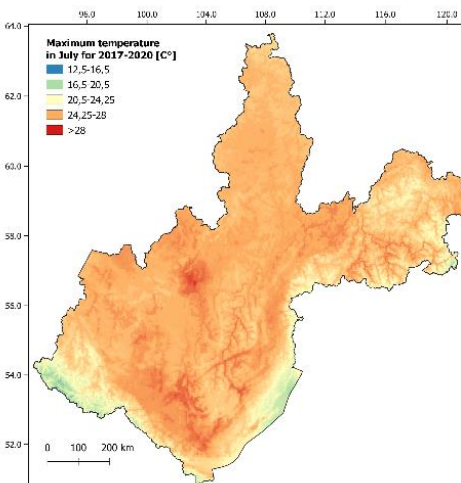
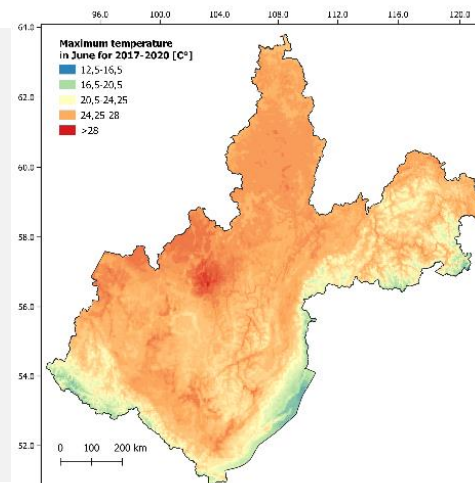
миллиметры данные
метеостанций



**Температура
воздуха (T)**

Градусы
Цельсия

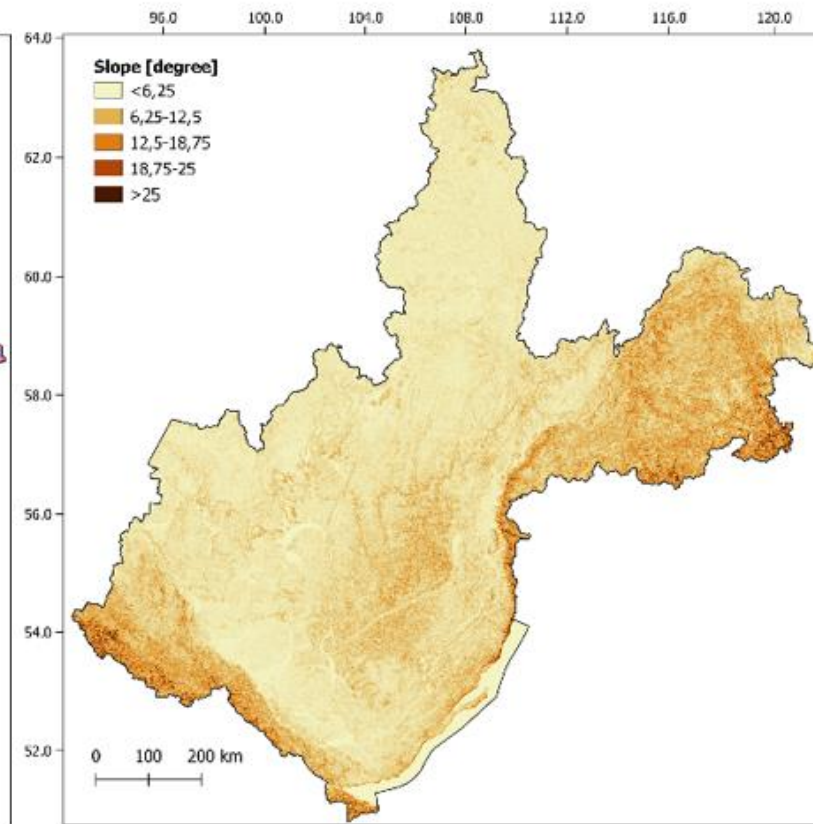
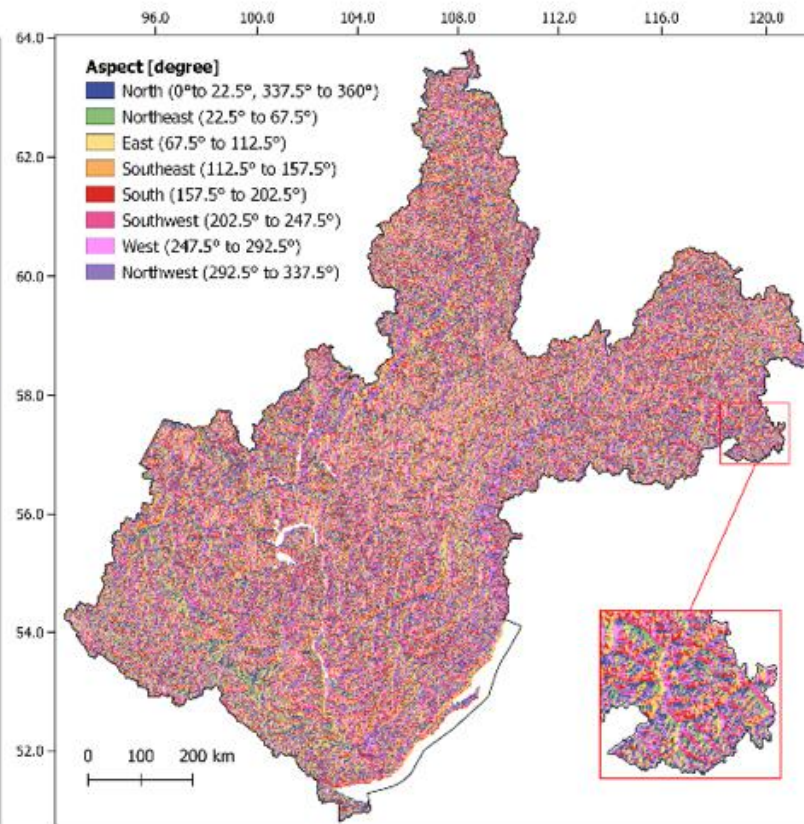
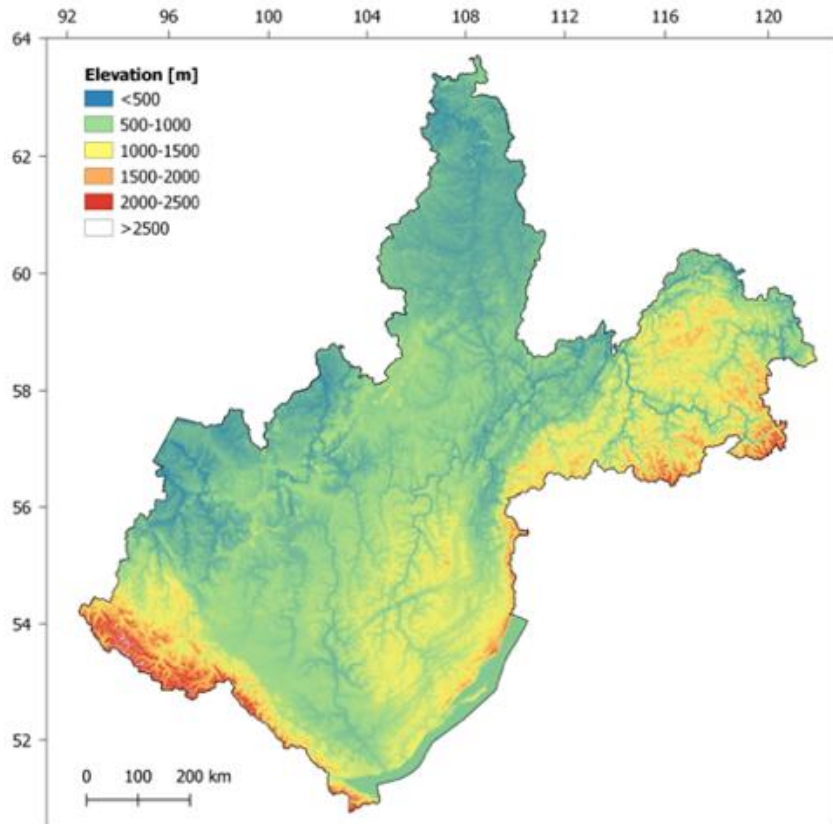
данные
метеостанций,
определяется на
высоте 2 метра
над поверхностью
земли



Топографические факторы

ФАКТОРЫ	ОПИСАНИЕ, ВОЗМОЖНЫЕ ЗНАЧЕНИЯ, ЕДИНИЦЫ ИЗМЕРЕНИЯ	СПОСОБ ПОЛУЧЕНИЯ																
Высота над уровнем моря	метры	SRTM (https://worldclim.org/)																
Аспект	градусы (0-360) <table border="0" data-bbox="1184 578 1834 1078"> <tr> <td>Север</td> <td>337.5-22.5</td> </tr> <tr> <td>Северо-восток</td> <td>22.5-67.5</td> </tr> <tr> <td>Восток</td> <td>67.5-112.5</td> </tr> <tr> <td>Юго-восток</td> <td>112.5-157.5</td> </tr> <tr> <td>Юг</td> <td>157.5-202.5</td> </tr> <tr> <td>Юго-запад</td> <td>202.5-247.5</td> </tr> <tr> <td>Запад</td> <td>247.5-292.5</td> </tr> <tr> <td>Северо-запад</td> <td>292.5-337.5</td> </tr> </table>	Север	337.5-22.5	Северо-восток	22.5-67.5	Восток	67.5-112.5	Юго-восток	112.5-157.5	Юг	157.5-202.5	Юго-запад	202.5-247.5	Запад	247.5-292.5	Северо-запад	292.5-337.5	Анализ цифровой модели рельефа в QGIS
Север	337.5-22.5																	
Северо-восток	22.5-67.5																	
Восток	67.5-112.5																	
Юго-восток	112.5-157.5																	
Юг	157.5-202.5																	
Юго-запад	202.5-247.5																	
Запад	247.5-292.5																	
Северо-запад	292.5-337.5																	
Уклон	градусы	Анализ цифровой модели рельефа в QGIS																

Топографические факторы: высота, аспект, уклон

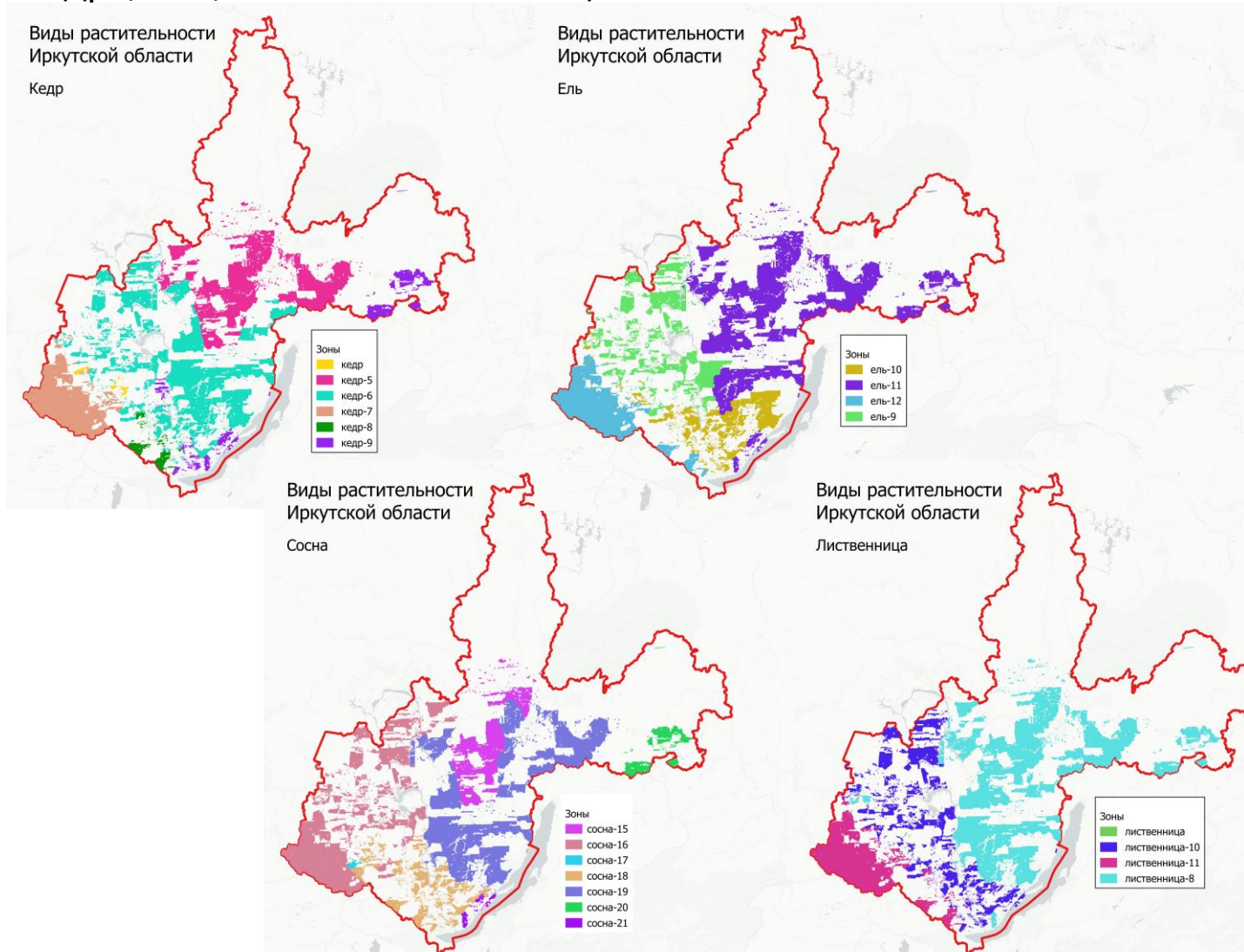


Факторы, описывающие тип поверхности

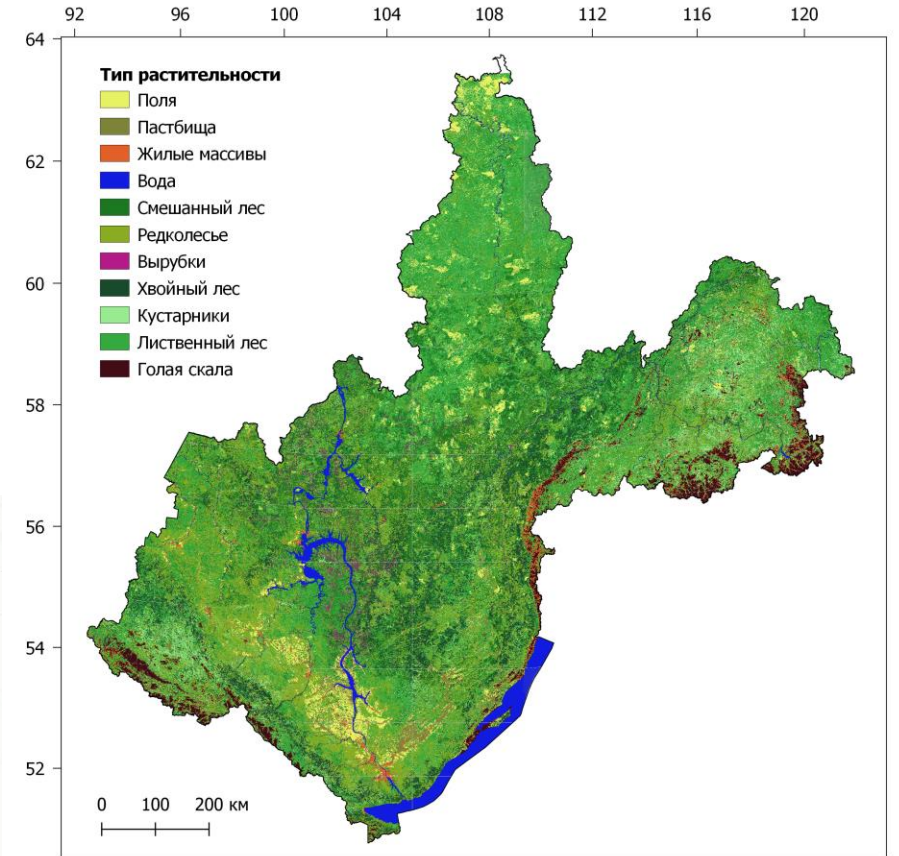
ФАКТОРЫ	ОПИСАНИЕ, ВОЗМОЖНЫЕ ЗНАЧЕНИЯ	СПОСОБ ПОЛУЧЕНИЯ
Зона лесосеменного районирования	сосна, ель, лиственница, кедр	определяется по данным лесных регламентов
Лесорастительная зона	лесостепная, таежная, Южно-Сибирская горная	определяется по данным лесных регламентов
Класс поверхности	11 классов поверхности: вода, облака, жилая зона, смешанный лес, хвойный лес, лиственный лес, редколесье, голая скала, вырубки, пастбище, сельскохозяйственные угодья	Классификация типов поверхности земли выполнена на основе сверточной нейронной сети архитектуры ResNet-50. Использовались размеченные космоснимки Иркутской области летнего периода за 2018-2020 гг. [Бычков И.В. и др.]

Факторы, описывающие тип поверхности

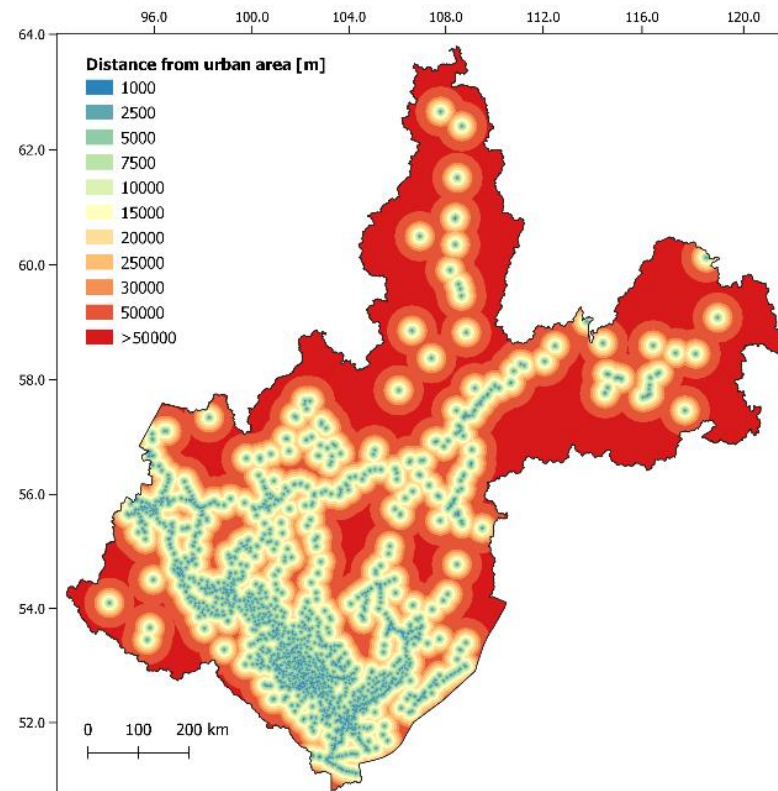
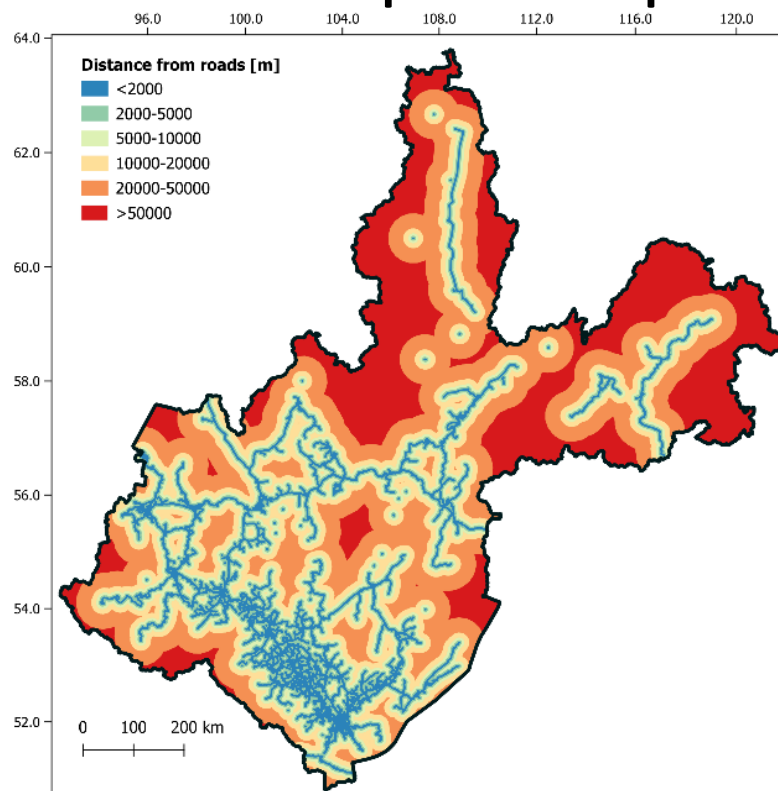
Разделение территории на зоны эволюции популяций:
кедра, ели, сосны и лиственницы



Классы поверхности территории
Иркутской области



Социальные факторы



ПАРАМЕТРЫ/СВОЙСТВА ПРЕЦЕДЕНТ	ОПИСАНИЕ, ВОЗМОЖНЫЕ ЗНАЧЕНИЯ, ЕДИНИЦЫ ИЗМЕРЕНИЯ	СПОСОБ ПОЛУЧЕНИЯ
расстояние до ближайшего населенного пункта	километры	определяется на
расстояние до ближайшей автомобильной дороги	километры	основе ГИС-данных

Коллинеарность переменных

VIF – коэффициент инфляции дисперсии
– измеряет силу корреляции между
независимыми переменными.
Значение, равное 1 – отсутствие
корреляции.
Значение 5 и более – сильная
корреляция, такие переменные
исключаются из модели.

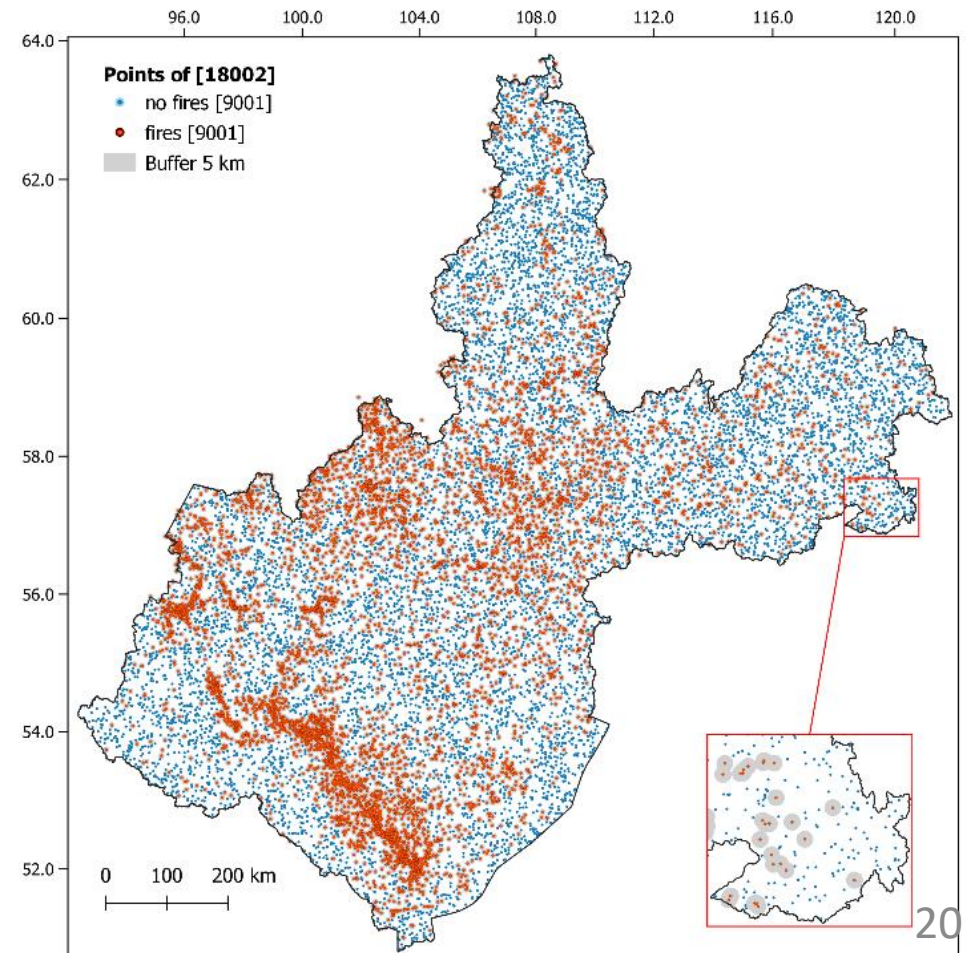
В результате исключена характеристика –
давления (P).

Variable	VIF
Ff	1.206742
RRR	1.169508
T	1.138275
VV	1.094837
U	1.568197
WW_code	1.111928
distance_to_road	3.036646
distance_to_river	1.036036
distance_to_set	3.123470
elevation	1.872036
slope	1.846820
aspect	1.010684
vegetation	1.096116

Генерация данных, для формирования класса «отсутствие пожара»

- Для формирования класса «отсутствие пожара» сгенерированы случайные точки, случайность определена во времени и пространстве.
- Сгенерированные координаты событий расположены вне буферной зоны вокруг пожаров размером 5 км, вне населенных пунктов и техногенных объектов. Количество сгенерированных точек на каждую область соответствовало количеству пожаров на этой территории:
класс «наличие пожара» - 9001,
класс «отсутствие пожара» - 9001.

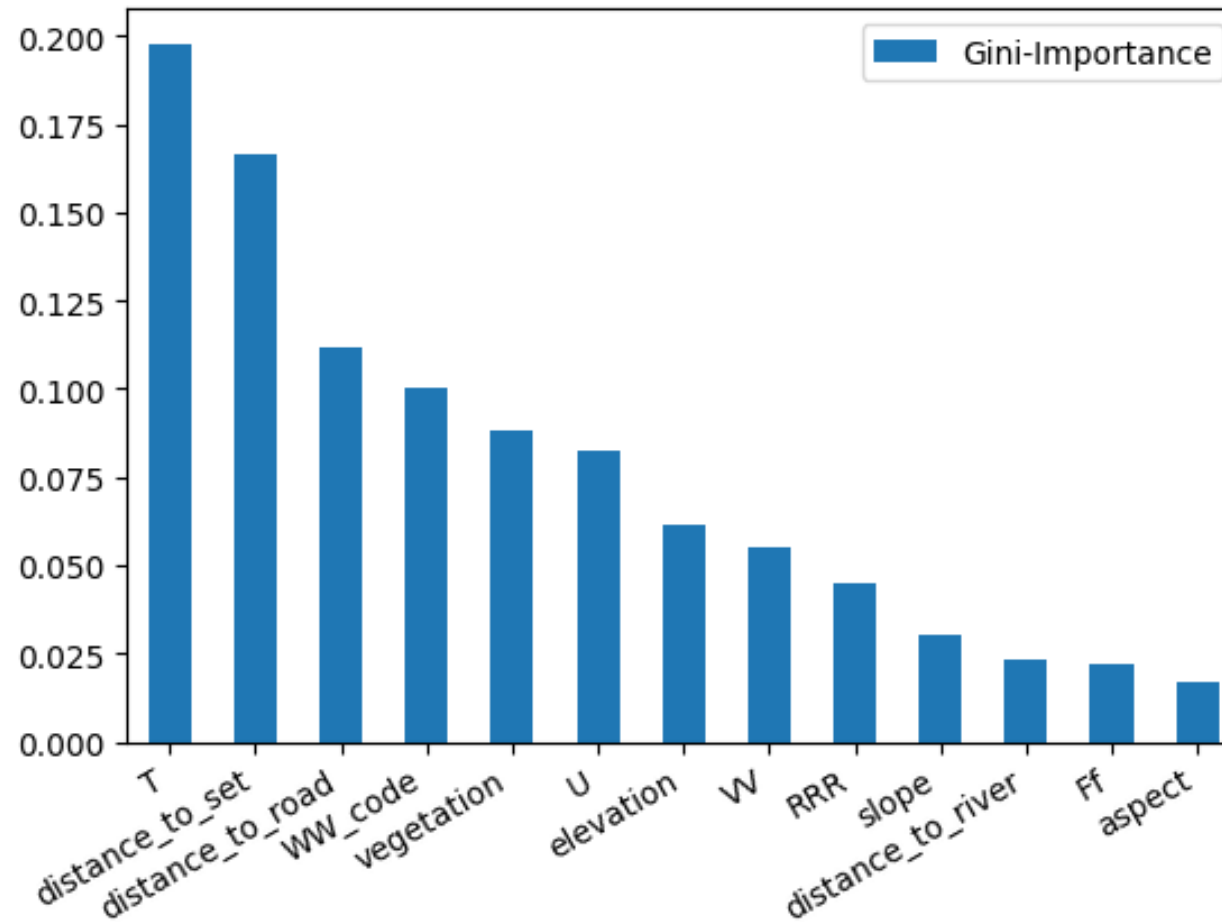
Исторические пожары (точки «наличие пожара») и сгенерированные (точки «отсутствие пожаров») с удаленностью 5 км. от пожаров



Подбор параметров метода RF

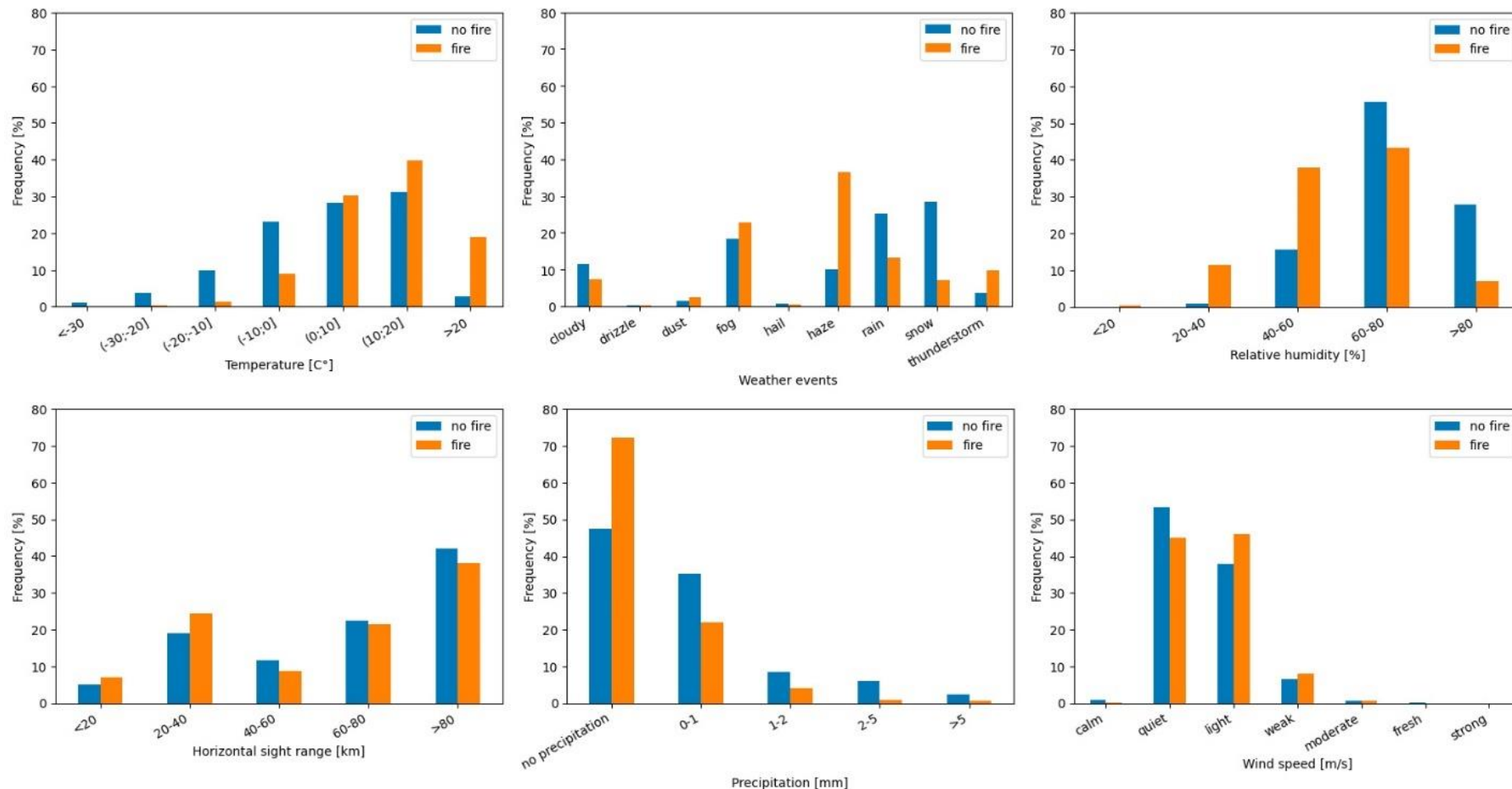
- На производительность RF влияют два важных параметра: количество деревьев в лесу и количество случайных величин на узел разделения. С помощью метода `RandomizedSearchCV()` и входных данных (модель `Random Forest Classifier` с параметрами по умолчанию, количество итераций, число кросс-валидаций) оптимизированы и выбраны гиперпараметры модели.
- В данной работе использовались следующие параметры:
 - количество деревьев (`n_estimators`) - **600**,
 - количество величин в узле дерева (`min_samples_leaf`) – **2**,
 - минимальное количество выборок, необходимое для разделения внутреннего узла (`min_samples_split`) – **23**,
 - максимальная глубина дерева (`max_depth`) - **15**.

Применение метода RF. Важность факторов



Применение метода RF. Анализ факторов

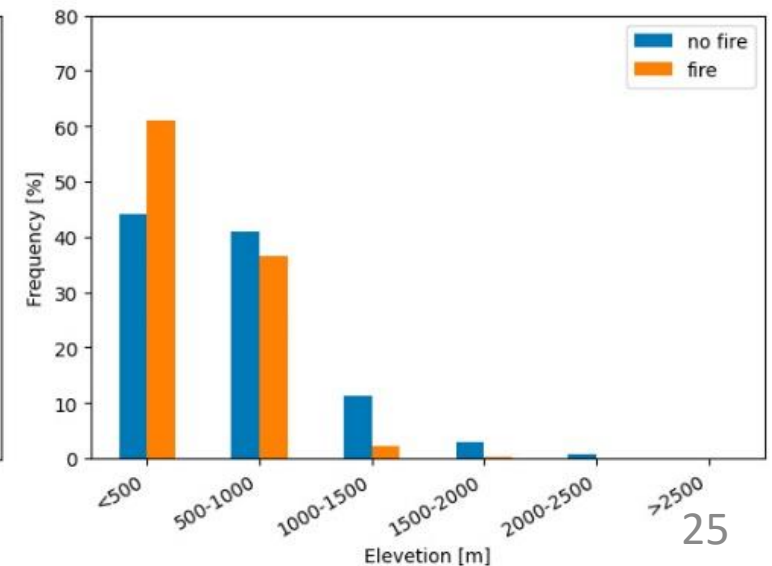
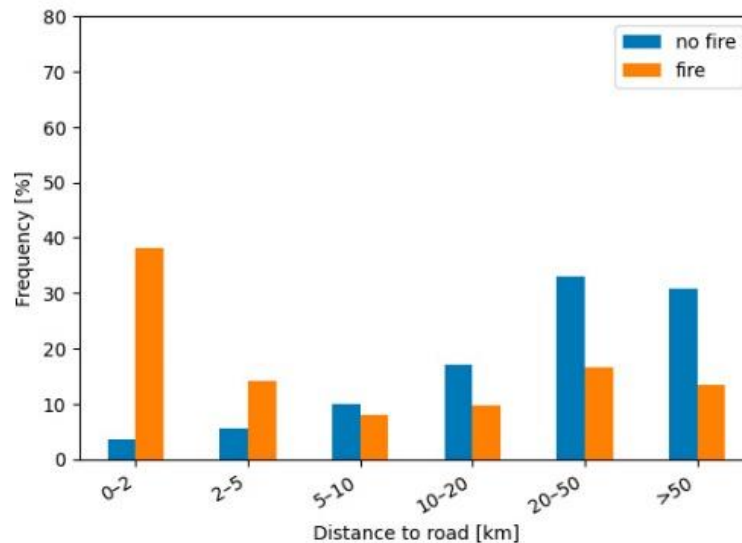
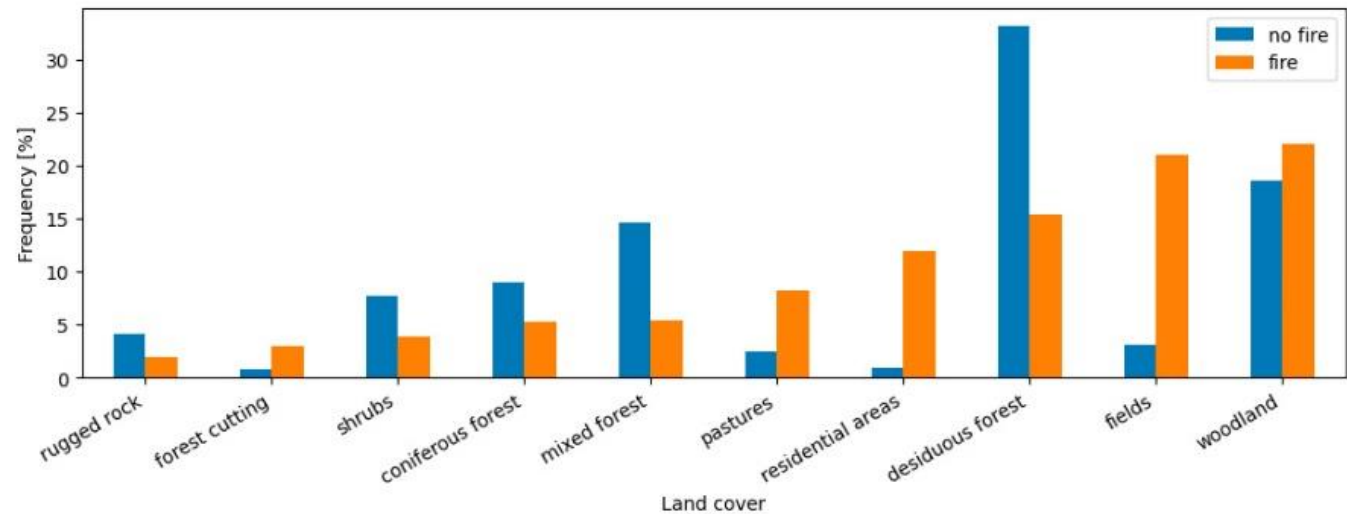
Распределение количества записей по классам «наличие пожара» и «отсутствие пожара» по категориям климатических переменных



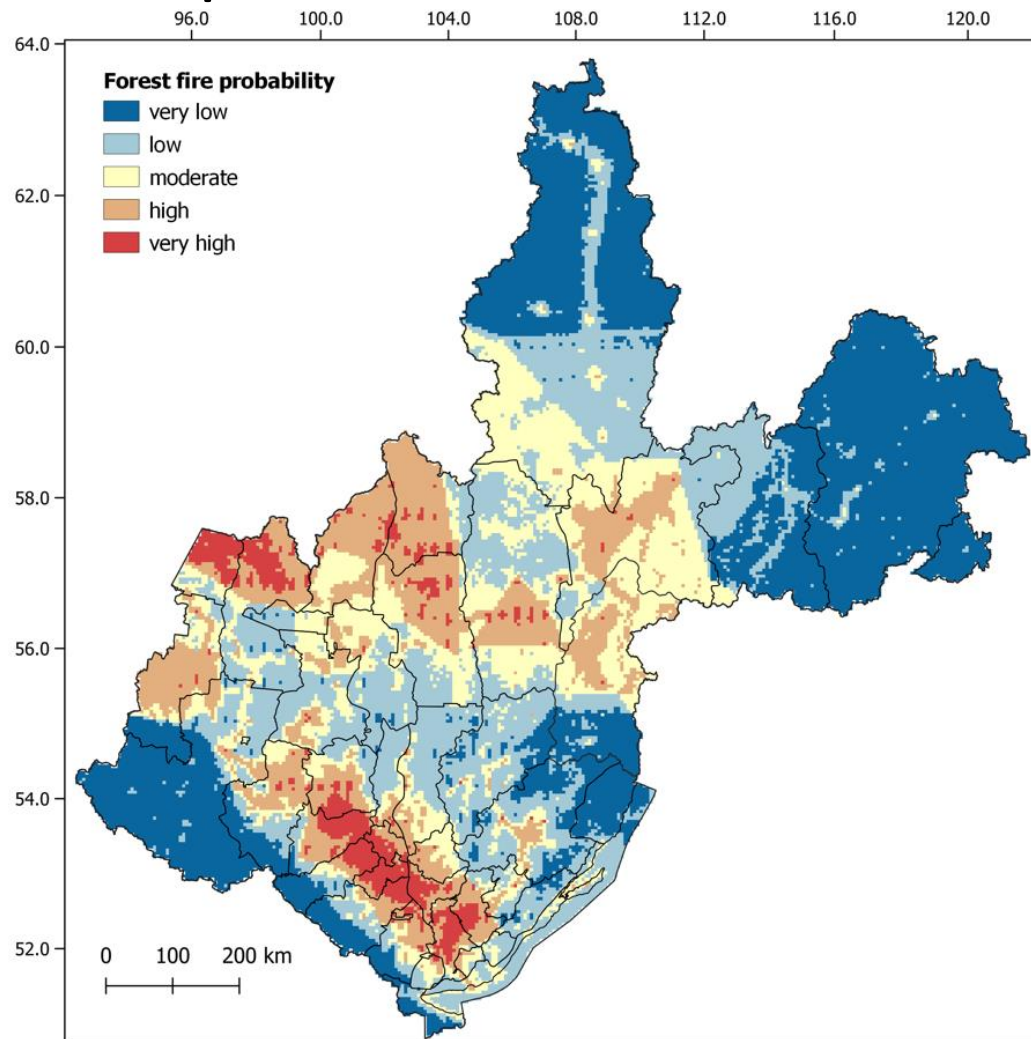
Применение метода RF. Анализ факторов

Распределение количества записей по классам «пожар» и «отсутствие пожара» по категориям:

- растительности,
- расстояния до дорог
- и высот рельефа



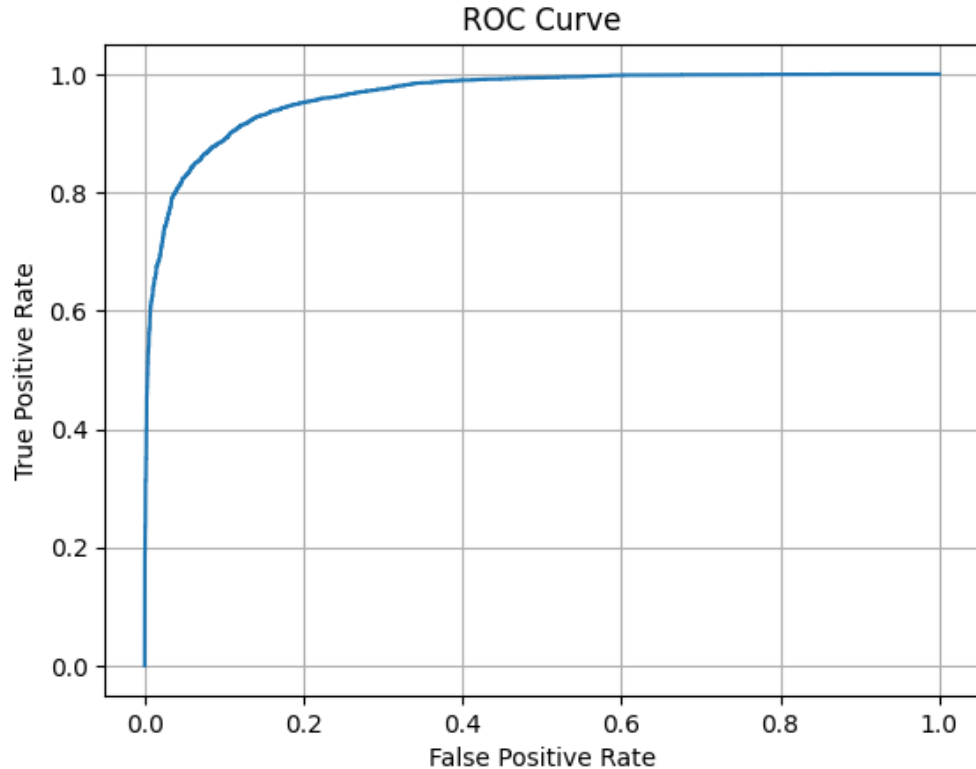
Применение метода RF. Оценка риска



Для создания карт пожароопасности территории была создана сетка точек с удаленностью 1 км. Для каждой точки определены необходимые значения входных факторов и получена вероятность принадлежности к классу «наличие пожара». Для данных точек задана интерполяция по атрибуту «риск», где площадь ячейки 150x150 км. Далее был применен метод реклассификации, в результате которого каждая конечная ячейка карты классифицируется на пять категорий:

- I очень высокий (вероятность 0,8-1),
- II высокий (вероятность 0,6-0,8),
- III средний (вероятность 0,4-0,6),
- IV низкий (вероятность 0,2-0,4),
- V очень низкий (вероятность 0-0,2).

Анализ точности решения задачи



Метрика	Значение
Accuracy	0.90
F1-score	0.88
AUC	0.96

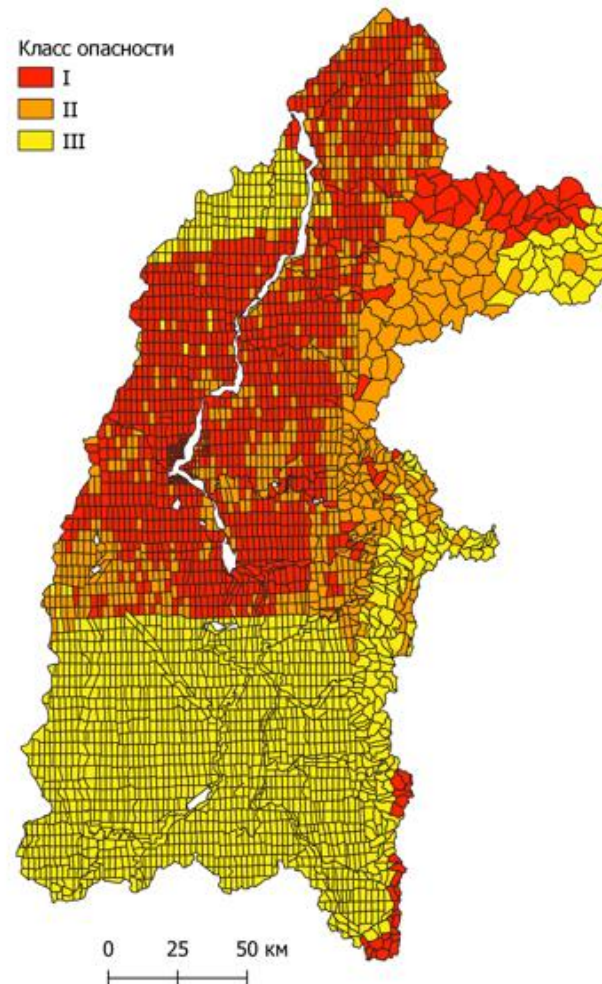
Фактические	Прогнозируемые	
	«отсутствие пожара»	«наличие пожара»
«отсутствие пожара»	4332 (93%)	312 (7%)
«наличие пожара»	534 (15%)	3079 (85%)

Сравнение результатов оценки риска методами CBR и RF

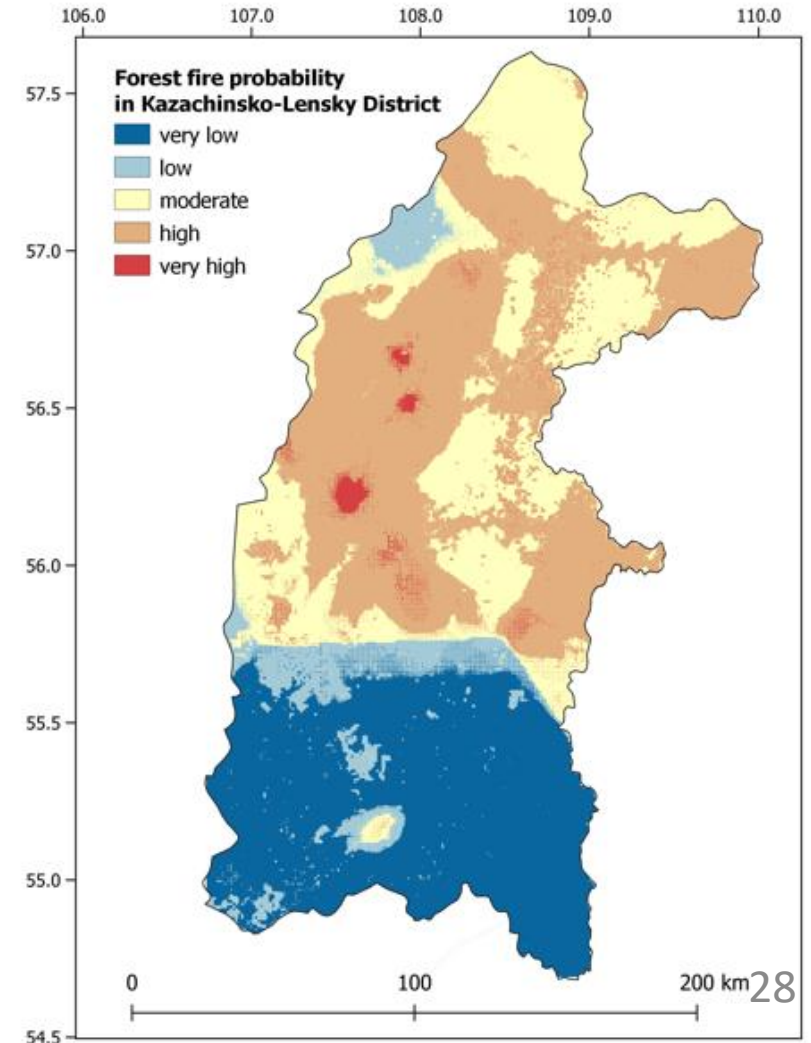
Результаты построения карт пожароопасности для Казачинско-Ленского муниципального образования:

*Dorodnykh, N.; Nikolaychuk, O.; Pestova, J.; Yurin, A. Forest Fire Risk Forecasting with the Aid of Case-Based Reasoning. *Appl. Sci.* **2022**, *12*, 8761.

(a) Метод CBR*



(b) Метод RF



Результаты

- Реализовано прогнозирование риска лесных пожаров. Основным результатом является оценка риска пожара для территории Иркутской области. **Оценка точности составила 0,9**. Для Иркутской области впервые было проведено исследование прогнозирования лесных пожаров на основе Random Forest.
- Результаты картирования риска территории показали, что наиболее **подвержены риску** лесного пожара южные наиболее **густонаселенные территории** области, а также территории, **расположенные вдоль дорог**, что может косвенно подтверждать утверждение органов государственного управления, что основной причиной пожаров является человеческий фактор.
- Сравнение подходов CBR и RF показало сходство оценок вероятности рисков возникновения пожаров. При использовании RF оценка формируется для пикселей, что является более детализированным подходом на основе большего числа факторов, характеризующих местность. Планируется пересмотреть подход к моделированию с учетом растровых данных.

Заключение

- В дальнейшем планируется исследование дополнительных факторов для включения их в модель и получения более точного решения, в частности, снежность зимы, праздничные дни, валовый внутренний продукт, более точные данные о грозовой активности. А также создание и обучение моделей, учитывающих индивидуальные особенности лесничеств.
- Разработанный подход станет основой создания информационной технологии прогнозирования лесных пожаров на заданной территории, предназначенной как для научных исследований, так и управления. Предлагается в рамках технологии реализовать функции оценки риска территорий лесных кварталов, лесничеств, муниципальных образований и др. территориальных структур. Создание карт риска лесных пожаров позволит обеспечить органы государственного управления дополнительной информацией для принятия решений по обеспечению мер по снижению риска и смягчению последствий лесных пожаров.

Картографирование пожароопасности Иркутской области с использованием метода машинного обучения Random Forest

Пестова Ю.В. (yupest@icc.ru),
Дородых Н.О.,
Николайчук О.А.,
Юрин А.Ю.

