

# Исследование возможностей повышения точности распознавания преобладающих пород леса на основе данных Sentinel-2 и материалов пробных площадей

<sup>1</sup>Богодухов М.А.

<sup>1</sup>Жарко В.О.

<sup>1</sup>Барталев С.А.

(1) ИКИ РАН

XXI международная конференция «СОВРЕМЕННЫЕ ПРОБЛЕМЫ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ ИЗ КОСМОСА  
(Физические основы, методы и технологии мониторинга окружающей среды, природных и антропогенных объектов)»

13-17 ноября 2023 г.

Москва, ИКИ РАН

# Введение

## **Используемые данные:**

- Материалы наземных обследований лесов на уровне пробных площадей (ПП);
- Временные ряды данных Sentinel-2 (для исследуемых участков №1, №2, №3);
- Летние композитные изображения Sentinel-2 (для исследуемых участков №2 и №3).

## **Территориальное расположение исследуемых участков:**

- №1: Ленинградская область;
- №2: Республика Башкортостан;
- №3: Томская область.

Распознавание преобладающих пород леса проводилось с использованием классификации методом случайных лесов.

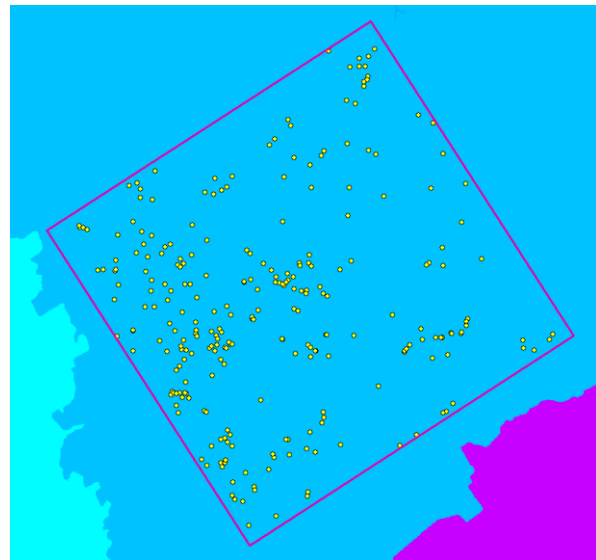
# Исследуемый участок №1 в пределах Ленинградской области



*Территориальное расположение исследуемого участка*

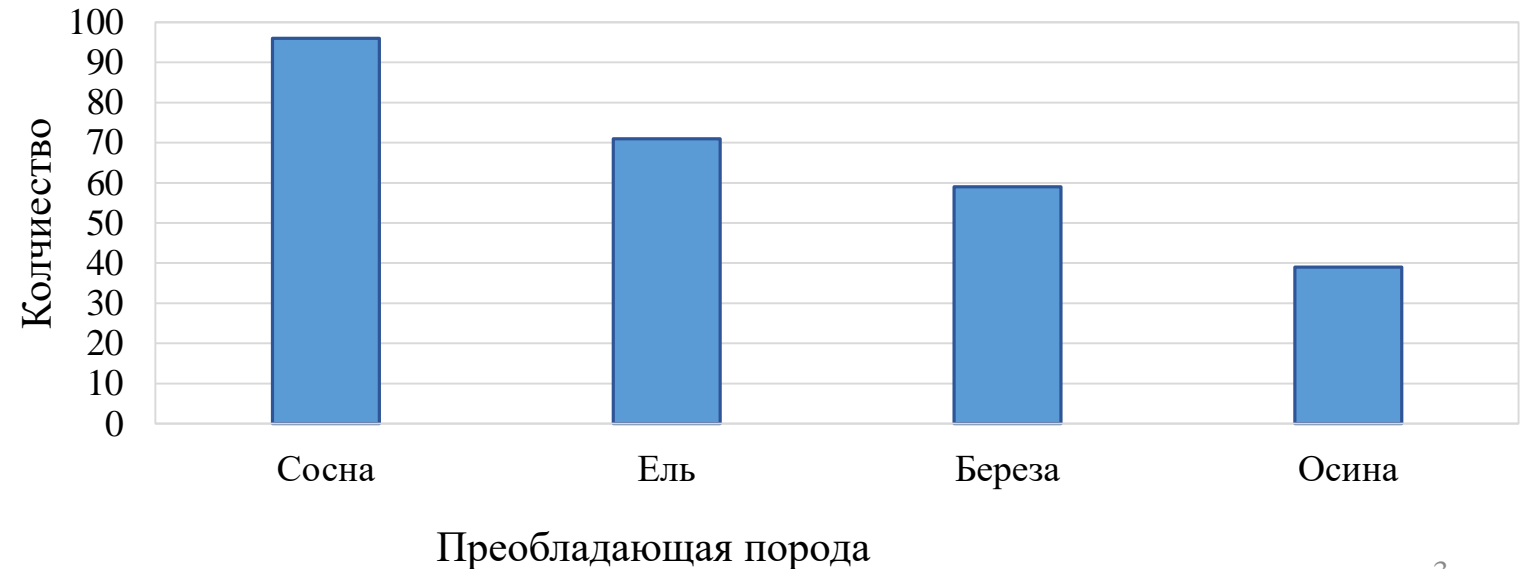
Дата сцены	Используемые каналы
Зима 2023-03-16	B3, B4, B8
Весна 2018-05-06	B2, B3, B4, B8, B11, B12
Лето 2022-06-29	
Осень 2019-08-29	
2021-10-07	

Band	Resolution	Central Wavelength	Description
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B8	10 m	842 nm	Visible and Near Infrared (VNIR)
B11	20 m	1610 nm	Short Wave Infrared (SWIR)
B12	20 m	2190 nm	Short Wave Infrared (SWIR)



*Распределение 265 ПП в пределах исследуемого участка*

## Распределение преобладающих пород



# Серия проведенных экспериментов для исследуемого участка №1:

1. Непосредственная классификация преобладающих пород;
  2. Добавление промежуточного этапа классификации на вечнозеленые/хвойные и листопадные/лиственные группы пород;
    - 2.1. Использование маски лиственные/хвойные как признак при обучении модели;
    - 2.2. Использование маски лиственные/хвойные как АОI.
  3. Расширение обучающей выборки на основе кластеризации данных ДЗЗ и совместного использования материалов наземного обследования на уровне ПП и таксационных выделов;
  4. Регрессионная оценка доли участия каждой породы в насаждении по запасу для определения породного состава и преобладающей породы лесов;
- Для всех экспериментов выборка из 265 элементов разбивалась следующим образом: 100 – обучающая выборка, 165\* – контрольная выборка;
  - Разбиение выборки и обучение модели выполнялось несколько раз для оценки средней достижимой точности.

# 1. Непосредственная классификация преобладающих пород

Все измерения в опорной выборке

	Результат классификации						
		Сосна	Ель	Береза	Осина	Recall	F1-score
Контрольная выборка	Сосна	4633	551	401	135	0.810	<b>0.786</b>
	Ель	867	2575	853	265	0.565	<b>0.584</b>
	Береза	524	720	1985	541	0.527	<b>0.509</b>
	Осина	51	413	790	1196	0.488	<b>0.522</b>
	Precision	0.763	0.605	0.493	0.560	<b>0.630</b>	

Отфильтрованы измерения с коэффициентом преобладания породы < 50% (231 из 265 элементов)

	Результат классификации						
		Сосна	Ель	Береза	Осина	Recall	F1-score
Контрольная выборка	Сосна	4541	364	247	78	0.868	<b>0.862</b>
	Ель	362	1967	387	154	0.685	<b>0.660</b>
	Береза	370	478	1485	557	0.514	<b>0.520</b>
	Осина	26	277	703	1104	0.523	<b>0.538</b>
	Precision	0.857	0.637	0.526	0.583	<b>0.694</b>	

## 2. Добавление промежуточного этапа классификации на вечнозеленые/хвойные и листопадные/лиственные группы пород

### 2.1. Использование маски лиственные/хвойные как признак при обучении модели

- Выбранные для обучения 100 ПП используются на первом этапе для классификации исследуемого участка №1 на хвойные/лиственные на основе только красного, зеленого и ближнего ИК каналов зимней сцены Sentinel-2;
- Полученная маска лиственные/хвойные используется как признак для классификации совместно с полным набором всех используемых каналов/сцен; при этом классифицируются все породы вместе (лиственные и хвойные).

		Результат классификации					
		Сосна	Ель	Береза	Осина	Recall	F1-score
Контрольная выборка	Сосна	4380	739	420	101	0.777	<b>0.733</b>
	Ель	937	2618	419	246	0.620	<b>0.539</b>
	Береза	753	1242	1329	366	0.360	<b>0.42</b>
	Осина	223	895	385	947	0.387	<b>0.462</b>
Precision		0.696	0.477	0.521	0.570	<b>0.58</b>	

## 2. Добавление промежуточного этапа классификации на вечнозеленые/хвойные и листопадные/лиственные группы пород

### 2.2. Использование маски лиственные/хвойные как АОІ

- Выбранные для обучения 100 ПП используются на первом этапе для классификации исследуемого участка №1 на хвойные/лиственные на основе только красного, зеленого и ближнего ИК каналов зимней сцены Sentinel-2;
- Полученная маска лиственные/хвойные используется как АОІ лиственных и хвойных лесов, в пределах которых выполняется классификация береза/осина и сосна/ель соответственно на основе полного набора всех используемых каналов/сцен;

		Результат классификации					
		Сосна	Ель	Береза	Осина	Recall	F1-score
Контрольная выборка	Сосна	7997	1083	1907	193	0.718	<b>0.695</b>
	Ель	1742	4778	1348	572	0.570	<b>0.494</b>
	Береза	1555	2745	2314	526	0.332	<b>0.338</b>
	Осина	534	2226	687	1393	0.295	<b>0.373</b>
Precision		0.685	0.448	0.385	0.557	<b>0.522</b>	

### **3. Расширение обучающей выборки на основе кластеризации данных ДЗЗ и совместного использования материалов наземного обследования на уровне ПП и таксационных выделов**

- Были выбраны все таксационные выдела в пределах 50м от каждой ПП; территория, охватываемая этими выделами, использовалась как АОI для дальнейшей обработки;
- В пределах выбранного АОI проводилась кластеризация полного набора используемых каналов/сцен на 1000 кластеров;
- Далее были выбраны те кластеры, в пределах которых попали только ПП одной и той же породы (на основе анализа 100 ПП, отобранных для обучения); выбранные кластеры были сгруппированы в маски соответствующих преобладающих пород;
- Полученная маска была пересечена с маской таксационных выделов, растеризованных по преобладающей породе в выделе, т.е. пиксель кластера отфильтровывался, если преобладающая порода попавших в данный кластер ПП не совпадала с преобладающей породой выдела, в котором находится пиксель;
- Таким образом была сформирована расширенная обучающая выборка размером ~23000 пикселей Sentinel-2, использованная на следующем этапе для обучения классификатора; точность оценивалась по неиспользованным ~160 ПП.



### 3. Расширение обучающей выборки на основе кластеризации данных ДЗЗ и совместного использования материалов наземного обследования на уровне ПП и таксационных выделов

Оценка точности на основе ПП

Результат классификации						
	Сосна	Ель	Береза	Осина	Recall	F1-score
Сосна	500	20	40	0	0.893	<b>0.803</b>
Ель	98	182	67	43	0.467	<b>0.486</b>
Береза	77	100	146	37	0.406	<b>0.433</b>
Осина	10	57	61	162	0.559	<b>0.609</b>
<b>Precision</b>	0.730	0.507	0.465	0.669	<b>0.619</b>	

Оценка точности на основе ПП с фильтрацией по доле преобладающей породы  $\geq 70\%$

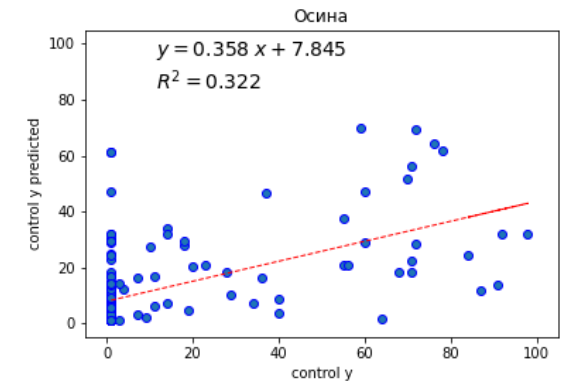
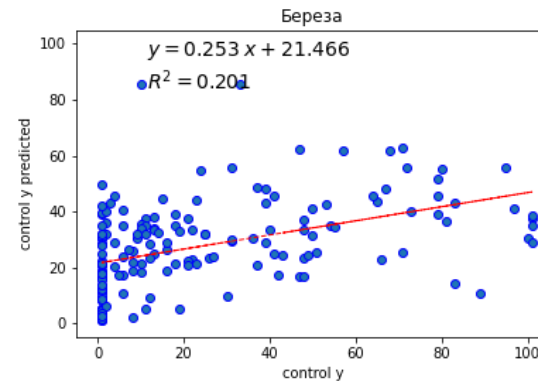
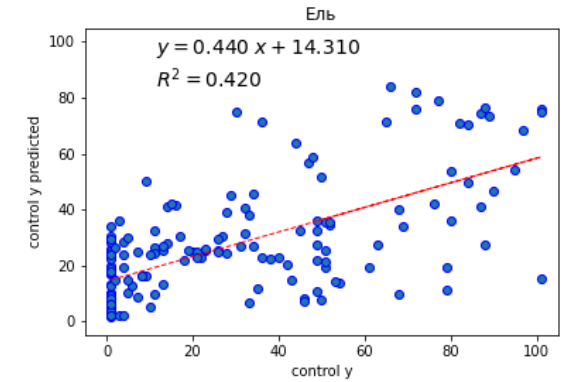
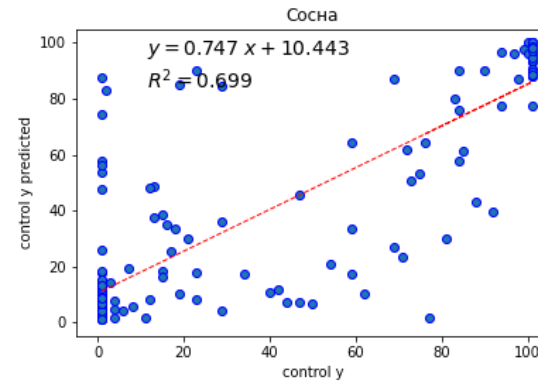
Результат классификации						
	Сосна	Ель	Береза	Осина	Recall	F1-score
Сосна	430	0	0	0	1.000	<b>0.89</b>
Ель	40	120	0	30	0.632	<b>0.614</b>
Береза	56	50	77	27	0.367	<b>0.487</b>
Осина	10	31	29	70	0.500	<b>0.525</b>
<b>Precision</b>	0.802	0.597	0.726	0.551	<b>0.719</b>	

## 4. Регрессионная оценка доли участия каждой породы в насаждении по запасу для определения породного состава и преобладающей породы лесов

- Для каждой породы выполнялась независимая регрессионная оценка методом случайных лесов ее доли участия в насаждении по запасу древесины

Порода	$R^2$	Среднеквадратическая ошибка, %
Сосна	$0.706 \pm 0.03$	$22.49 \pm 1.28$
Ель	$0.398 \pm 0.11$	$22.73 \pm 1.66$
Береза	$0.213 \pm 0.04$	$26.26 \pm 0.99$
Осина	$0.296 \pm 0.078$	$21.46 \pm 1.29$

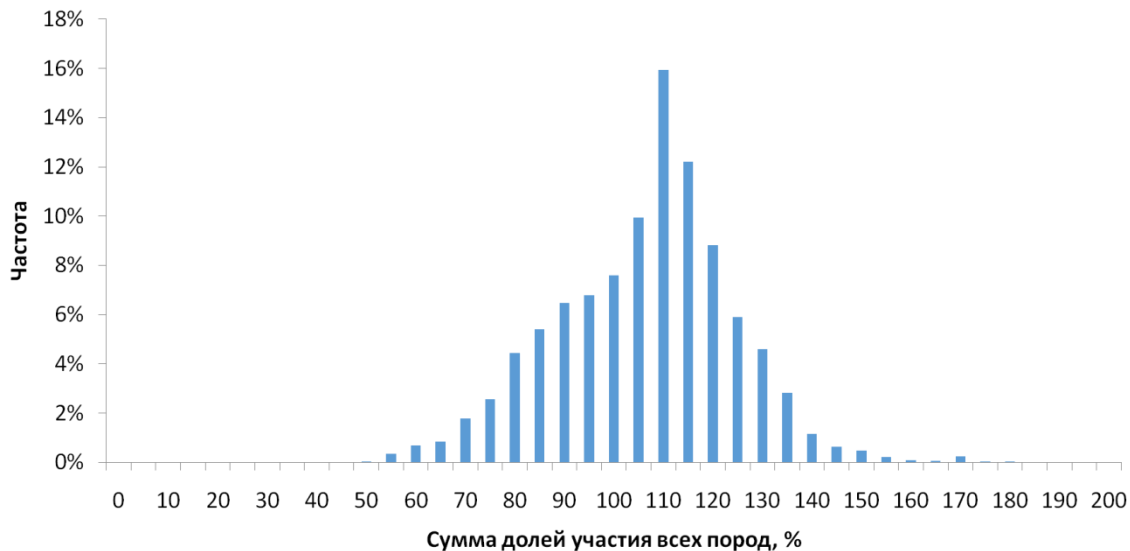
*Оценка точности определения долей участия пород в насаждении по результатам 100 экспериментов*



*Примеры диаграмм рассеяния по результатам оценки долей участия пород в насаждении по запасу древесины*

## 4. Регрессионная оценка доли участия каждой породы в насаждении по запасу для определения породного состава и преобладающей породы лесов

- Помимо высокого уровня неопределенности оценки, на представленной ниже гистограмме продемонстрировано также наличие систематической погрешности, поскольку пик гистограммы находится в области более 100%;
- Для каждой ПП на основе полученных оценок определялась преобладающая порода (древесная порода с максимальной долей) и анализировалась точность определения преобладающей породы;



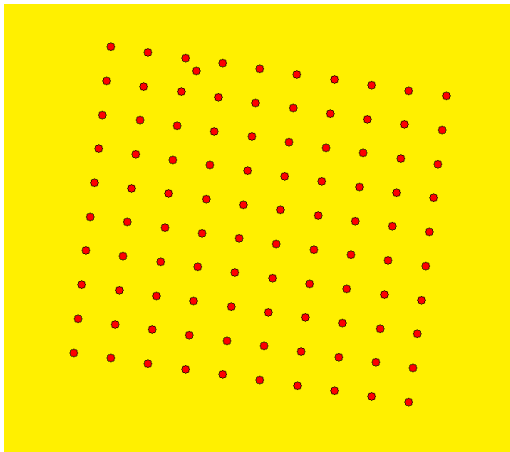
*Распределение суммы долей всех пород по результатам серии экспериментов*

	Результат классификации				Recall	F1-score	
		Сосна	Ель	Береза			Осина
Контрольная выборка	Сосна	4620	742	460	58	0.786	<b>0.754</b>
	Ель	848	2577	623	202	0.606	<b>0.574</b>
	Береза	677	913	1884	396	0.487	<b>0.486</b>
	Осина	215	496	864	925	0.370	<b>0.444</b>
	Precision	0.726	0.545	0.492	0.585	<b>0.606</b>	

# Исследуемый участок №2 в пределах Республики Башкортостан



Территориальное расположение исследуемого участка

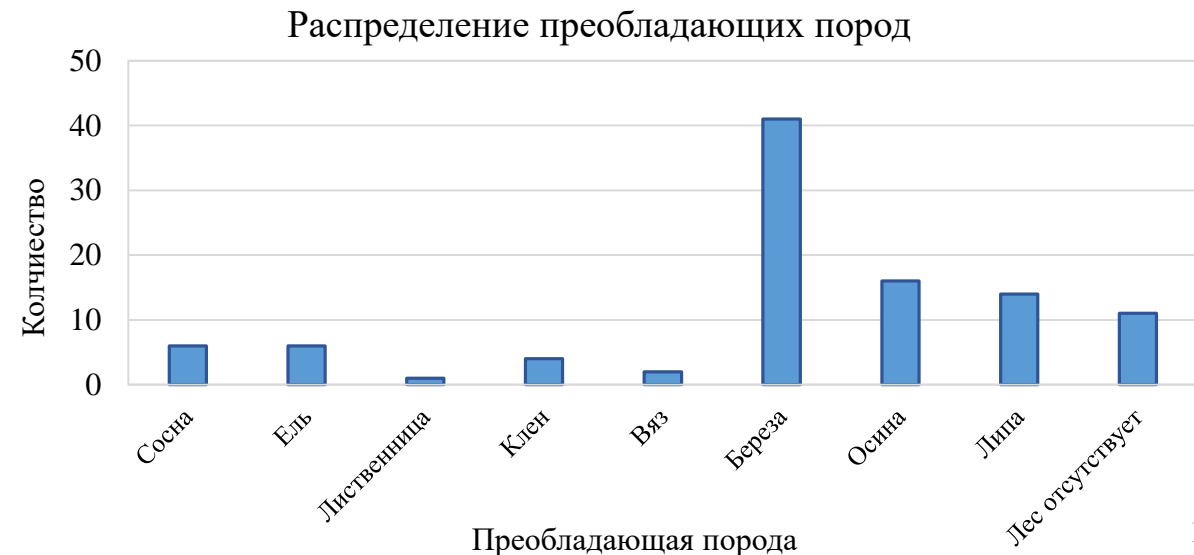


Распределение 101 ПП в пределах исследуемого участка

Дата сцены	Используемые каналы
Зима 2023-03-24	B2, B3, B4, B5, B6, B7, B8
Весна 2022-05-08 2021-05-28	B2, B3, B4, B5, B6, B7, B8, B11, B12
Лето 2022-06-27 2021-07-14 2021-07-24 2022-08-23	
Осень 2022-09-22 2021-10-07	

Летние 5-дневные композитные изображения	
Период	Используемые каналы
2022-05-15 - 2022-10-07	B2, B3, B4, B8, B11, B12

Band	Resolution	Central Wavelength	Description
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B5	20 m	705 nm	Visible and Near Infrared (VNIR)
B6	20 m	740 nm	Visible and Near Infrared (VNIR)
B7	20 m	783 nm	Visible and Near Infrared (VNIR)
B8	10 m	842 nm	Visible and Near Infrared (VNIR)
B11	20 m	1610 nm	Short Wave Infrared (SWIR)
B12	20 m	2190 nm	Short Wave Infrared (SWIR)



## Серия проведенных экспериментов для исследуемого участка №2:

1. Непосредственная классификация преобладающих пород;
  2. Предварительная коррекция опорной выборки с последующей классификацией преобладающих пород
- Были отфильтрованы ПП по значению коэффициента главной породы  $<50\%$ . Кроме того, были исключены некоторые породы (лиственница, клен, вяз), которые были недостаточно представлены в опорной выборке, а также ПП, имеющие выбивающиеся значения характеристик леса или с отсутствием информации о наличии лесных насаждениях. В результате фильтрации в экспериментальной версии в опорной выборке было оставлено 51 из 101 возможных ПП.
  - Разбиение опорной выборки на обучающую и контрольную выборки не проводилось из-за её достаточно небольшого размера. Оценка точности выполнялась с использованием стандартного метода out-of-bag;
  - Обучение модели выполнялось несколько раз для оценки средней достижимой точности.

# 1. Непосредственная классификация преобладающих пород:

Признаки: временные ряды данных Sentinel-2

Результат классификации							
	Сосна	Ель	Береза	Осина	Липа	Recall	F1-score
Сосна	1	29	10	0	0	0.025	0.032
Ель	15	25	8	2	0	0.500	0.467
Береза	7	3	243	3	4	0.935	0.783
Осина	0	0	52	17	11	0.213	0.333
Липа	0	0	48	0	32	0.400	0.504
Precision	0.043	0.439	0.673	0.773	0.681	<b>0.624</b>	

Признаки: летние композитные изображения Sentinel-2

Результат классификации							
	Сосна	Ель	Береза	Осина	Липа	Recall	F1-score
Сосна	0	30	10	0	0	0	<b>0</b>
Ель	17	23	10	0	0	0.460	<b>0.430</b>
Береза	0	10	244	0	6	0.938	<b>0.786</b>
Осина	0	0	64	6	10	0.075	<b>0.118</b>
Липа	0	0	75	5	0	0	<b>0</b>
Precision	0	0.404	0.676	0.273	0	<b>0.535</b>	

## 2. Предварительная коррекция опорной выборки с последующей классификацией преобладающих пород

- Были выбраны все пиксели в пределах 20м от каждой ПП и добавлены в опорную выборку для соответствующей породы;
- Опорная выборка итеративно разбивалась по всем ПП на обучающую и контрольную выборки таким образом, что в контрольной выборке оставались только те пиксели, которые находятся в пределах текущей ПП. То есть обучение модели происходило по всем пикселям, не принадлежащим окрестностям текущей ПП;
- Обученная таким образом модель применялась к контрольной выборке и оценивалась вероятность (с использованием *out-of-bag decision function*) отнесения каждого пикселя контрольной выборки к породе, которая была изначально указана в этой ПП. Значениям характеристик леса текущей ПП ставились в соответствие значения яркости пикселя, у которого полученная вероятность оказывалась максимальной;
- При использовании такого подхода формируется скорректированная опорная выборка, которая позволяет учесть погрешности привязки центров ПП и данных ДЗЗ;
- Полученная скорректированная опорная выборка использовалась для обучения новой модели и классификации преобладающих пород. Оценка точности выполнялась с использованием стандартного метода *out-of-bag*.

## 2. Предварительная коррекция опорной выборки с последующей классификацией преобладающих пород

Признаки: временные ряды данных Sentinel-2

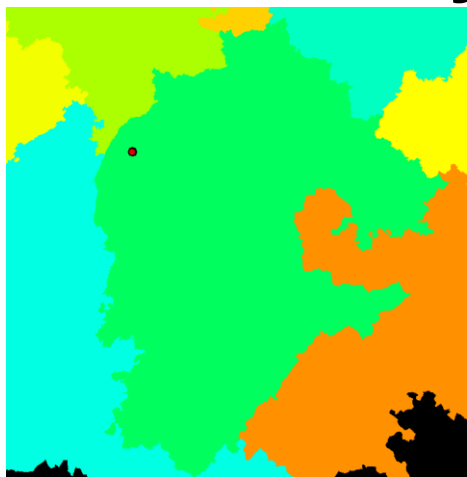
Результат классификации							
	Сосна	Ель	Береза	Осина	Липа	Recall	F1-score
Сосна	19	14	3	0	4	0.475	0.481
Ель	20	30	0	0	0	0.600	0.638
Береза	0	0	260	0	0	1.000	0.967
Осина	0	0	14	52	14	0.650	0.684
Липа	0	0	1	20	59	0.738	0.752
<b>Precision</b>	0.487	0.682	0.935	0.722	0.766	<b>0.824</b>	

Признаки: летние композитные изображения Sentinel-2

Результат классификации							
	Сосна	Ель	Береза	Осина	Липа	Recall	F1-score
Сосна	25	10	5	0	0	0.625	0.649
Ель	12	28	10	0	0	0.560	0.629
Береза	0	1	259	0	0	0.996	0.826
Осина	0	0	49	27	4	0.338	0.482
Липа	0	0	44	5	31	0.388	0.539
<b>Precision</b>	0.676	0.718	0.706	0.844	0.886	<b>0.725</b>	



# Исследуемый участок №3 в пределах Томской области

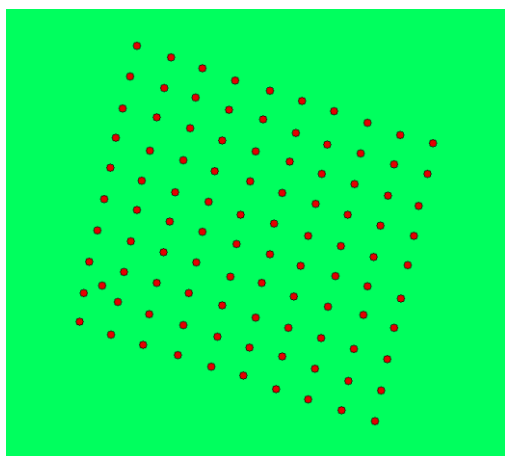


Территориальное расположение исследуемого участка

Дата сцены	Используемые каналы
Зима 2023-03-14	B2, B3, B4, B5, B6, B7, B8
Весна 2022-05-08 2022-05-25 Лето 2022-06-14 2021-07-04 2021-07-24 2020-08-11 осень 2020-09-12 2022-10-10	B2, B3, B4, B5, B6, B7, B8, B11, B12

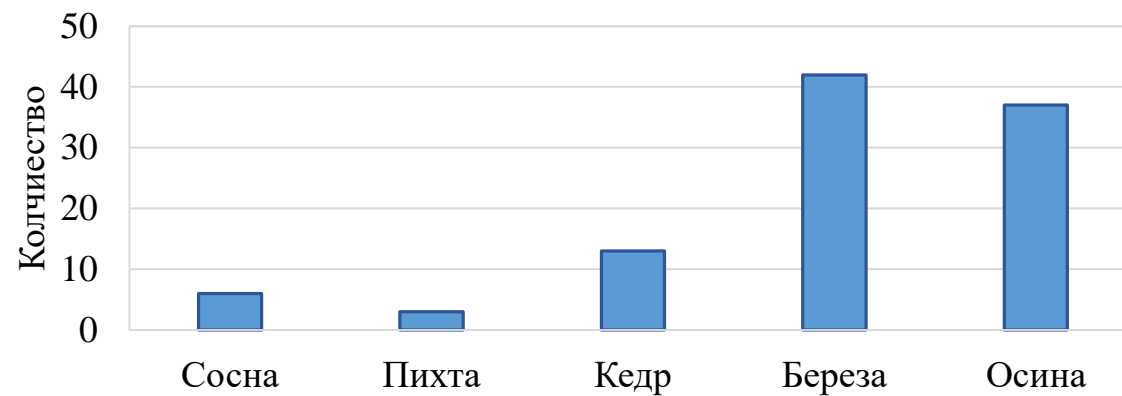
Band	Resolution	Central Wavelength	Description
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B5	20 m	705 nm	Visible and Near Infrared (VNIR)
B6	20 m	740 nm	Visible and Near Infrared (VNIR)
B7	20 m	783 nm	Visible and Near Infrared (VNIR)
B8	10 m	842 nm	Visible and Near Infrared (VNIR)
B11	20 m	1610 nm	Short Wave Infrared (SWIR)
B12	20 m	2190 nm	Short Wave Infrared (SWIR)

Летние 5-дневные композитные изображения	
Период	Используемые каналы
2022-05-10 - 2022-09-07	B2, B3, B4, B8, B11, B12



Распределение 101 ПП в пределах исследуемого участка

Распределение преобладающих пород



## Серия проведенных экспериментов для исследуемого участка №3:

1. Непосредственная классификация преобладающих пород;
  2. Предварительная коррекция опорной выборки с последующей классификацией преобладающих пород
- Были отфильтрованы ПП по значению коэффициента главной породы  $<50\%$ . Кроме того, была исключена пихта, которая была недостаточно представлена в опорной выборке, а также ПП, имеющие выбивающиеся значения характеристик леса. В результате фильтрации в экспериментальной версии в опорной выборке было оставлено 76 из 101 возможных ПП.
  - Разбиение опорной выборки на обучающую и контрольную выборки не проводилось из-за её достаточно небольшого размера. Оценка точности выполнялась с использованием стандартного метода out-of-bag;
  - Обучение модели выполнялось несколько раз для оценки средней достижимой точности.

# 1. Непосредственная классификация преобладающих пород:

Признаки: временные ряды данных Sentinel-2

Результат классификации						
	Сосна	Кедр	Береза	Осина	Recall	F1-score
Сосна	20	10	10	0	0.500	0.519
Кедр	7	52	19	2	0.650	0.675
Береза	10	12	188	110	0.588	0.574
Осина	0	0	118	202	0.631	0.637
<b>Precision</b>	0.541	0.703	0.561	0.643	<b>0.608</b>	

Признаки: летние композитные изображения Sentinel-2

Результат классификации						
	Сосна	Кедр	Береза	Осина	Recall	F1-score
Сосна	8	10	10	12	0.200	0.205
Кедр	20	52	0	8	0.650	0.642
Береза	10	10	214	86	0.669	0.669
Осина	0	10	96	214	0.669	0.669
<b>Precision</b>	0.211	0.634	0.669	0.669	<b>0.642</b>	

## 2. Предварительная коррекция опорной выборки с последующей классификацией преобладающих пород

- Были выбраны все пиксели в пределах 20м от каждой ПП и добавлены в опорную выборку для соответствующей породы;
- Опорная выборка итеративно разбивалась по всем ПП на обучающую и контрольную выборки таким образом, что в контрольной выборке оставались только те пиксели, которые находятся в пределах текущей ПП. То есть обучение модели происходило по всем пикселям, не принадлежащим окрестностям текущей ПП;
- Обученная таким образом модель применялась к контрольной выборке и оценивалась вероятность (с использованием *out-of-bag decision function*) отнесения каждого пикселя контрольной выборки к породе, которая была изначально указана в этой ПП. Значениям характеристик леса текущей ПП ставились в соответствие значения яркости пикселя, у которого полученная вероятность оказывалась максимальной;
- При использовании такого подхода формируется скорректированная опорная выборка, которая позволяет учесть погрешности привязки центров ПП и данных ДЗЗ;
- Полученная скорректированная опорная выборка использовалась для обучения новой модели и классификации преобладающих пород. Оценка точности выполнялась с использованием стандартного метода *out-of-bag*.

## 2. Предварительная коррекция опорной выборки с последующей классификацией преобладающих пород

Признаки: временные ряды данных Sentinel-2

Результат классификации						
	Сосна	Кедр	Береза	Осина	Recall	F1-score
Сосна	37	0	1	2	0.925	0.779
Кедр	0	65	15	0	0.813	0.897
Береза	18	0	299	3	0.934	0.909
Осина	0	0	23	297	0.928	0.955
<b>Precision</b>	0.673	1.000	0.885	0.983	<b>0.918</b>	

Признаки: летние композитные изображения Sentinel-2

Результат классификации						
	Сосна	Кедр	Береза	Осина	Recall	F1-score
Сосна	12	10	16	2	0.300	0.393
Кедр	0	66	0	14	0.825	0.846
Береза	9	0	294	17	0.919	0.880
Осина	0	0	38	282	0.881	0.888
<b>Precision</b>	0.571	0.868	0.845	0.895	<b>0.861</b>	

# Выводы

- Проведены исследования возможностей повышения точности распознавания преобладающих пород на различных независимых участках;
- На данный момент некоторые подходы, такие как предварительный этап классификации типа лесных насаждений, расширение обучающей выборки на основе кластеризации данных ДЗЗ и регрессионная оценка доли участия каждой породы по запасу не показали повышения точности распознавания преобладающих пород;
- Подходы, основанные на коррекции опорной выборки, позволяют минимизировать эффект погрешности привязки центров ПП и данных ДЗЗ, а также пространственную неоднородность лесного покрова по породной структуре в пределах ПП. В серии проведенных экспериментов они показали заметное повышение точности распознавания преобладающих пород;
- В дальнейшем планируется проведение экспериментов по определению других характеристик лесов, в том числе количественных характеристик (таких как запас стволовой древесины) на основе непараметрических регрессионных моделей.

# Финансирование

Работа выполнена в рамках реализации важнейшего инновационного проекта государственного значения «Разработка системы наземного и дистанционного мониторинга пулов углерода и потоков парниковых газов на территории Российской Федерации, обеспечение создания системы учета данных о потоках климатически активных веществ и бюджете углерода в лесах и других наземных экологических системах» (рег. № 123030300031-6). Обработка данных ДЗЗ проводилась с использованием ресурсов ЦКП «ИКИ-Мониторинг», включая УНУ «Вега -Science» (Лупян и др., 2019), развиваемого и поддерживаемого в рамках темы «Мониторинг» (№ 122042500031-8).