

Исследование возможностей распознавания преобладающих пород леса с использованием регрессионной оценки долевого участия пород по запасу на основе данных Sentinel-2 и материалов пробных площадей

¹Богодухов М.А.

¹Барталев С.А.

¹Жарко В.О.

(1) ИКИ РАН

Двадцать вторая международная конференция
«СОВРЕМЕННЫЕ ПРОБЛЕМЫ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ ИЗ КОСМОСА»
11-15 ноября, 2024
Москва, ИКИ РАН

Актуальность:

- Для формирования по спутниковым данным Дистанционного Зондирования Земли (ДЗЗ) карт характеристик леса, в том числе необходимых для оценки их запасов углерода, в национальном масштабе, используются данные ДЗЗ со средним пространственным разрешением (100-500 метров на пиксель). Однако, наиболее актуальные данные наземных обследований лесов, доступные для их валидации, характеризуют лесной покров на уровне Пробных Площадей (ПП) диаметром около 25-30 метров, что затрудняет их прямое сопоставление. Следовательно, актуальным является создание карт характеристик лесов высокого разрешения с размером пикселя 10-30 метров, основанных на материалах ПП и ДЗЗ, для возможности валидации продуктов обработки данных ДЗЗ среднего пространственного разрешения.

Цель работы:

- Исследование возможностей распознавания преобладающих пород леса на основе данных Sentinel-2 и материалов ПП

Используемые данные:

- Материалы наземных обследований лесов на уровне регулярной сети ПП, заложенных на территории более 100 Тестовых Полигонов (ТП);
- Сформированные на основе автоматизированной обработки безоблачные композитные изображения Sentinel-2 с временным разрешением 5 дней.

Территориальное расположение ТП:

- ТП распределены по всей территории России

Территориальное расположение ТП



Используемые данные

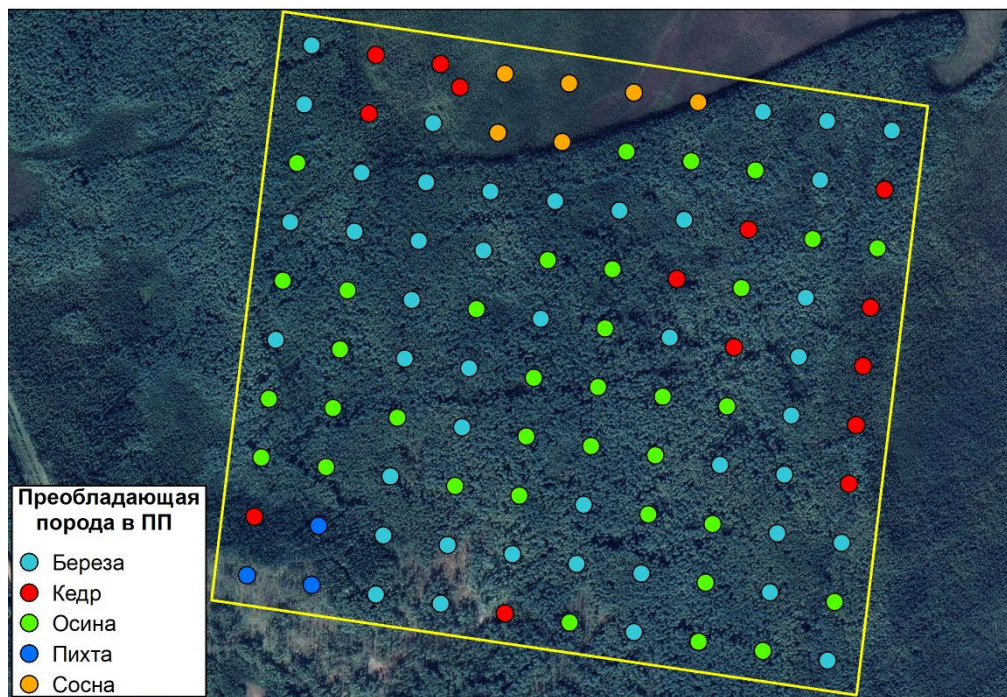
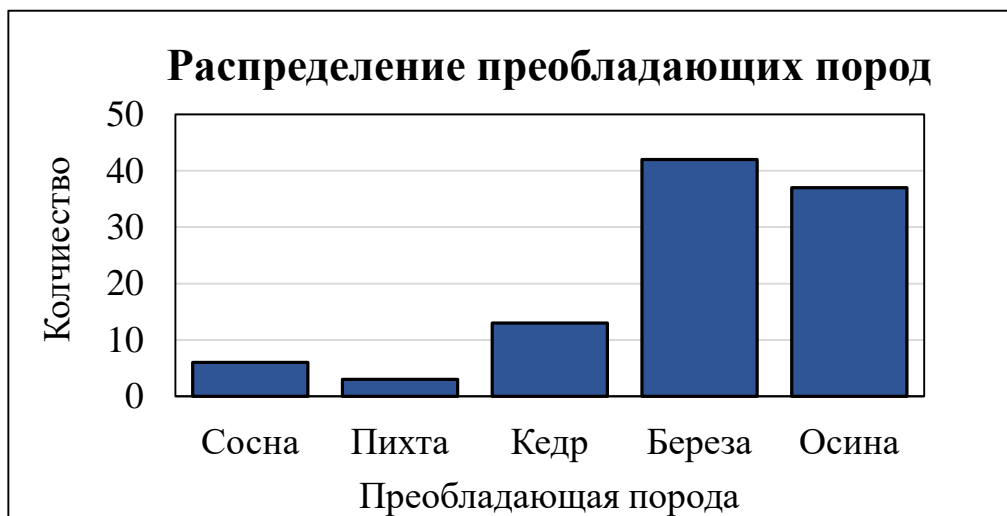


Иллюстрация регулярной сети 101 ПП в пределах ТП

Данные Sentinel-2			
Дата	Используемые каналы	Resolution	Description
Временные ряды композитных изображений с шагом 5 дней: 2022-04-10 — 2022-10-27	B2	10 m	Blue
	B3	10 m	Green
	B4	10 m	Red
	B8	10 m	Visible and Near Infrared (VNIR)
	B11	20 m	Short Wave Infrared (SWIR)
Усредненные зимние композиты за 2019-2024 гг.	B12	20 m	Short Wave Infrared (SWIR)



Характеристики древостоя на ПП						
	Количество деревьев на 1 га	Диаметр, см	Возраст, лет	Высота, м	Запас, м ³ /га	Долевое участие
Береза	631	7	55	7.3	20	4.65
Сосна	526	14	90	8.3	21	4.88
Кедр	105	10	55	6	2	0.47

Распознавание преобладающих пород леса на основе использования случайных лесов в режиме классификации

- Для данного эксперимента были отобраны следующие породы:
сосна, ель, пихта, лиственница, кедр, дуб, береза, осина, ольха, липа, ясень;
- В качестве признаков обучения модели используются временные ряды композитных изображений с шагом 5 дней и усредненные зимние композиты за 2019-2024 гг. Признаки отфильтровываются по условию наличия артефактов в окрестностях ПП.
- Для ПП ставится в соответствие среднее значение яркости в окне 3x3 выбранных признаков.
- Оценка точности полученной модели определяется на основе подхода out-of-bag

Алгоритм распознавания преобладающей породы леса в режиме классификации в пределах каждого ТП:

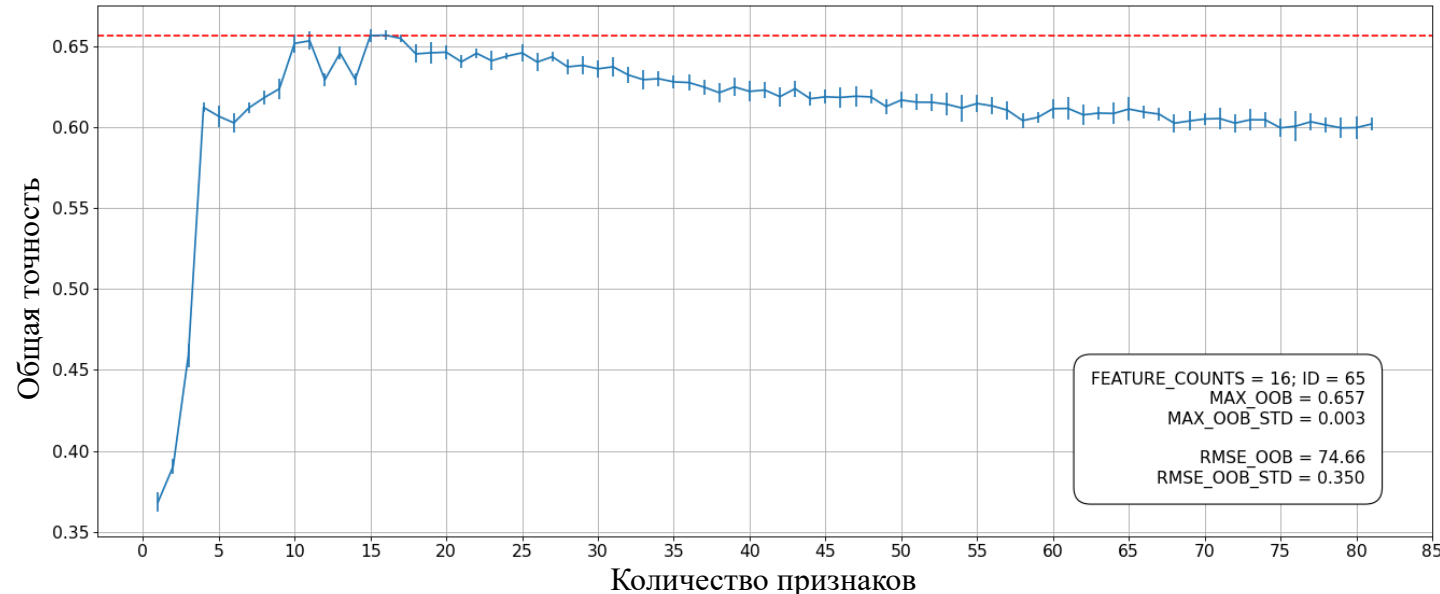
1. Формирование обучающей выборки;
 - 1.1. Для каждого элемента выборки извлекается информация о преобладающей породе и её доле участия в ПП;
 - 1.2. Отбрасываются элементы с долей преобладания < 50 .
 - 1.3. Если количество представителей какой-либо породы < 2 , элементы с данной породой отбрасываются;
2. Подбор оптимального набора признаков обучения модели;
3. Обучение модели с использованием сформированной выборки и оптимального набора признаков
4. Рассчитывается матрица перепутывания и показатели точности при оценке преобладающих пород

Подбор оптимального набора признаков для обучения модели

Алгоритм поиска оптимального набора признаков (На основе **Recursive Feature Elimination**) :

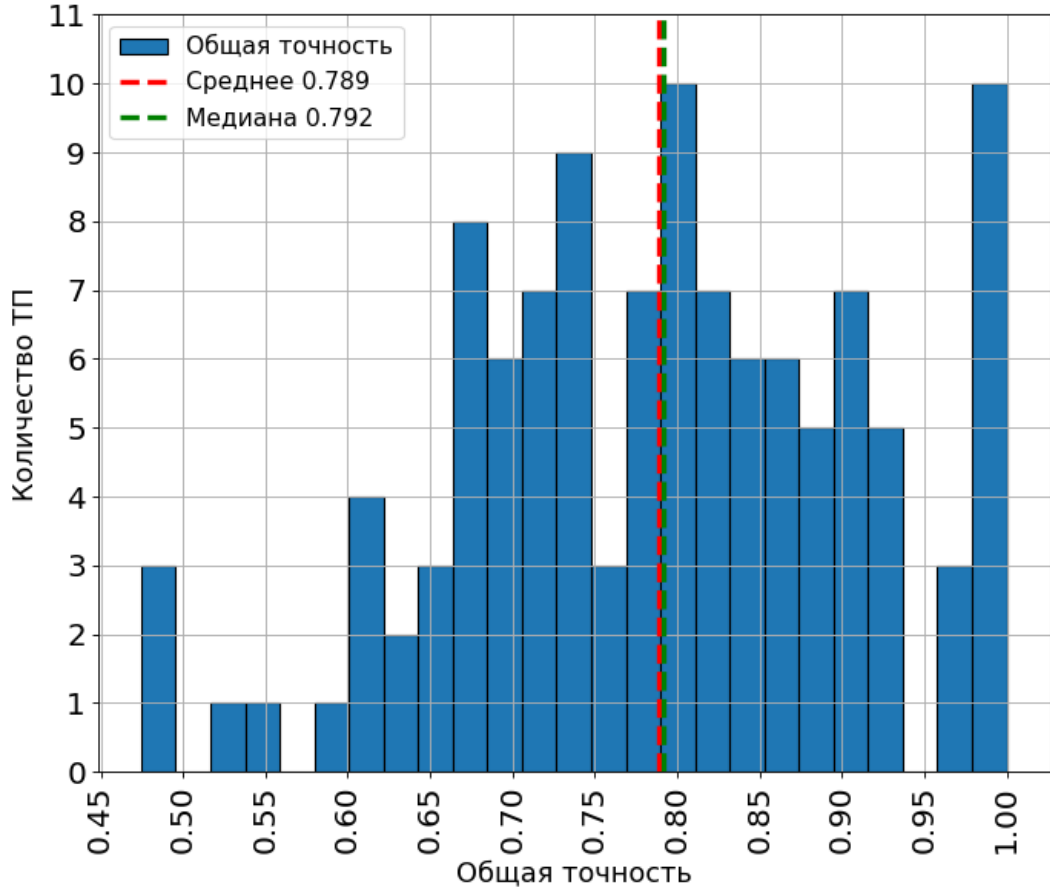
1. Подается обучающая выборка и полный набор признаков обучения модели;
 2. Рассчитывается показатель точности оценки выбранной характеристики леса методом случайных лесов в выбранном режиме с использованием подхода out-of-bag;
 3. Для каждого признака при обучении рассчитывается параметр *feature_importances_*, который указывает, насколько этот признак улучшает производительность модели;
 4. Из набора признаков отбрасывается наименее информативный признак согласно п.3;
 5. Повторяются пп.1 – 4, пока не останется 1 признак.
- Для поиска оптимального набора признаков используются оценки точности, полученные из п.2 представленного выше алгоритма. Оптимальным набором признаков считается тот, для которого достигается максимальный показатель точности.

Иллюстрация подбора оптимального набора признаков



Преобладающие породы леса на основе использования случайных лесов в режиме классификации по всем ТП

Матрица перепутывания и показатели точности при оценке преобладающих пород



		Порода по данным ДЗЗ											Recall		F1-score	
		Сосна	Ель	Пихта	Лиственница	Кедр	Дуб	Береза	Осина	Ольха	Липа	Ясень				
Порода на ТП	Сосна	1995	88	1	29	4	0	125	19	1	0	0	0.88	0.87		
	Ель	117	1403	8	27	1	0	130	36	2	0	0	0.81	0.81		
	Пихта	3	15	29	6	2	0	6	0	0	0	1	0.47	0.54		
	Лиственница	17	13	3	975	0	2	45	3	0	0	0	0.92	0.88		
	Кедр	7	3	1	1	54	0	5	1	0	2	0	0.73	0.73		
	Дуб	0	0	0	1	0	99	8	0	0	1	1	0.90	0.92		
	Береза	159	172	2	90	6	1	1564	97	18	12	2	0.74	0.73		
	Осина	18	44	0	17	0	0	171	466	9	9	0	0.63	0.68		
	Ольха	2	4	0	0	0	1	49	8	94	0	2	0.58	0.65		
	Липа	1	1	0	1	4	0	27	10	0	46	2	0.5	0.56		
	Ясень	0	0	1	0	2	2	3	0	1	1	11	0.52	0.55		
Precision		0.86	0.81	0.64	0.85	0.73	0.93	0.73	0.73	0.73	0.65	0.58	0.798			

Распознавание преобладающих пород леса с использованием регрессионной оценки долевого участия пород по запасу

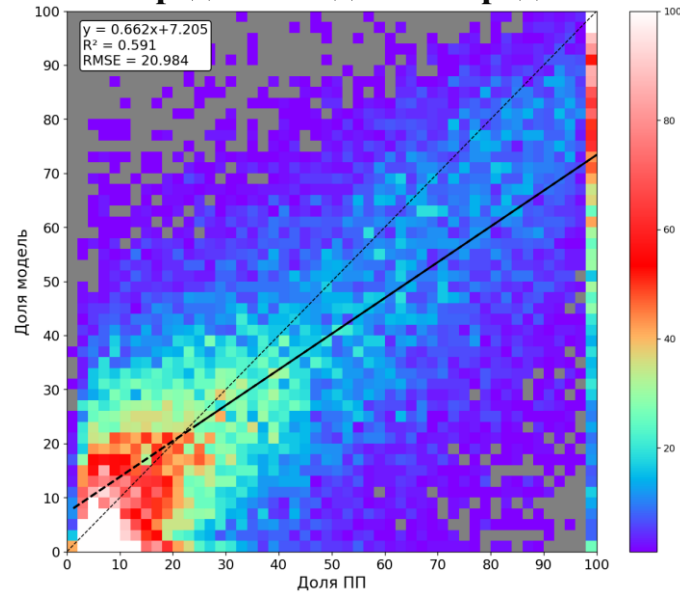
- Для данного эксперимента были отобраны следующие породы:
сосна, ель, пихта, лиственница, кедр, дуб, береза, осина, ольха, липа, ясень;
- В качестве признаков обучения модели используются временные ряды композитных изображений с шагом 5 дней и усредненные зимние композиты за 2019-2024 гг. Признаки отфильтровываются по условию наличия артефактов в окрестностях ПП.
- Для ПП ставится в соответствие среднее значение яркости в окне 3x3 выбранных признаков.
- Оценка точности полученной модели определяется на основе подхода out-of-bag и на тех же ПП, что и в режиме классификации

Алгоритм распознавания преобладающей породы леса предложенным методом в пределах каждого ТП:

1. Из общего списка выбранных пород отбираются породы, имеющие доленое участие как минимум в 2 ПП;
2. Для каждой отобранной породы формируется обучающая выборка, с информацией о доленом участии породы в ПП;
3. Если порода не представлена в ПП, но представлена в ТП, то в обучающей выборке фиксируется информация об её отсутствии;
4. Для каждой обучающей выборки выполняется независимый поиск оптимального набора признаков обучения модели;
5. Выполняется оценка доленого участия по запасу каждой из пород на основе использования случайных лесов в режиме регрессии;
6. Выбирается преобладающая порода в насаждении по наибольшему значению полученных оценок доленого участия

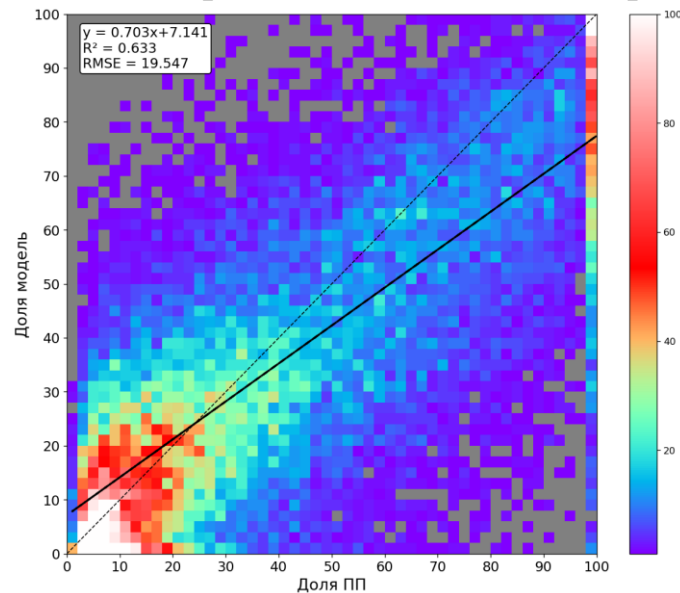
Иллюстрация распределения долей преобладания пород леса по данным ПП и результатам оценок

Распределение долей пород леса

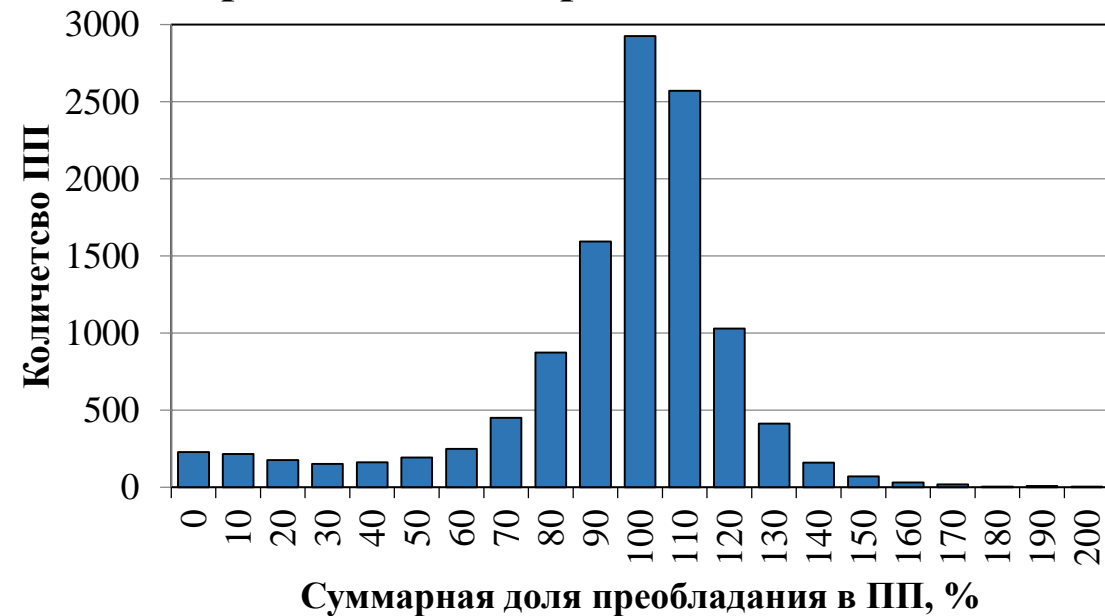


Порода	Доля преобладания		
	По модели	Нормализованная по модели	На ПП
Ель	«0»	«0»	«0»
Пихта	«0»	«0»	«0»
Лиственница	40.88	50.95	56
Береза	27.08	33.75	23
Осина	12.27	15.29	21
Сумма	80.23	100	100

Распределение нормализованных долей пород леса

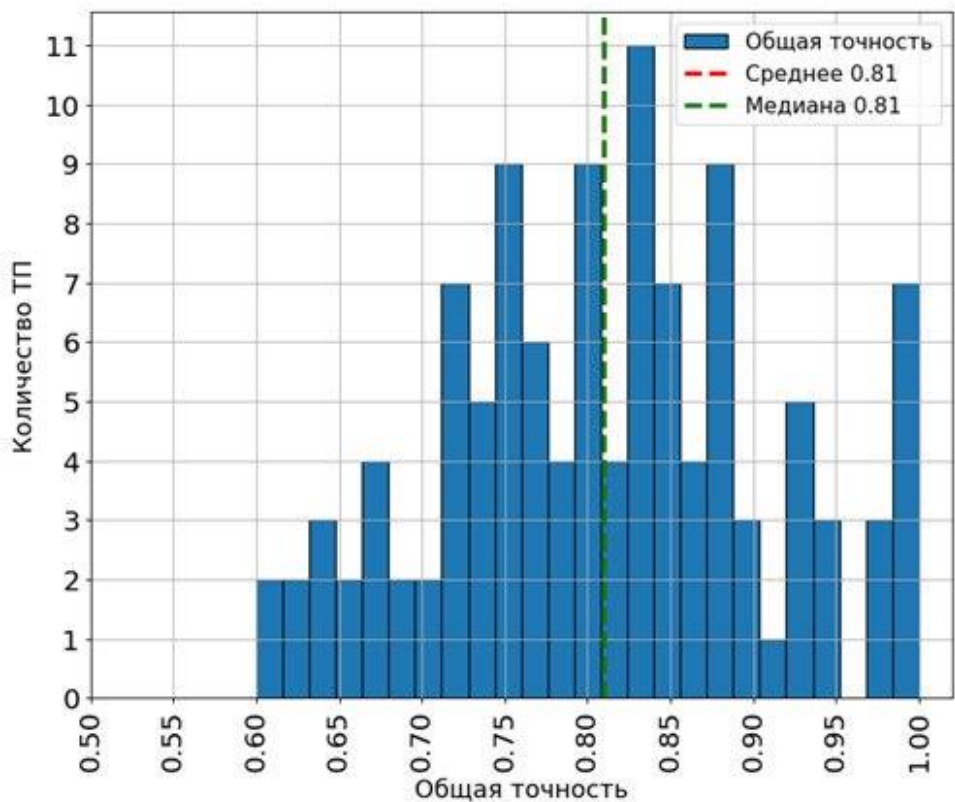


Распределение суммы долей преобладания пород в ПП по всем ТП



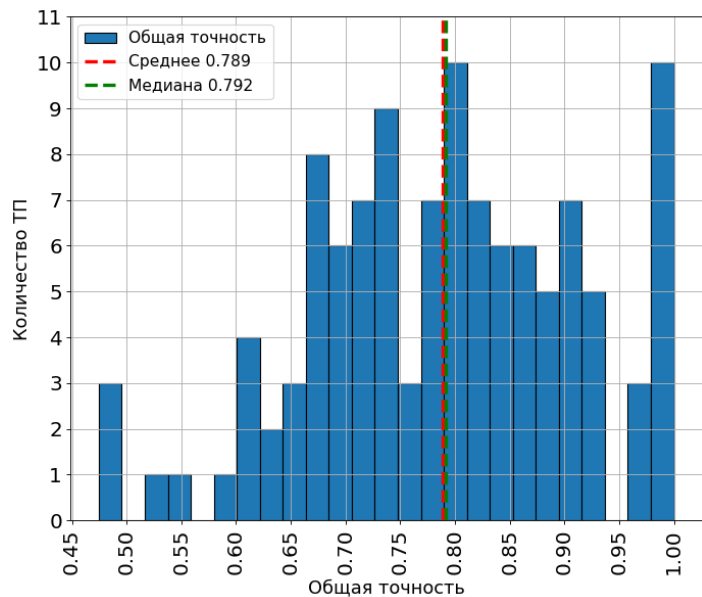
Преобладающие породы леса на основе использования случайных лесов в режиме регрессии долевого участия по запасу по всем ТП

Матрица перепутывания и показатели точности при оценке преобладающих пород

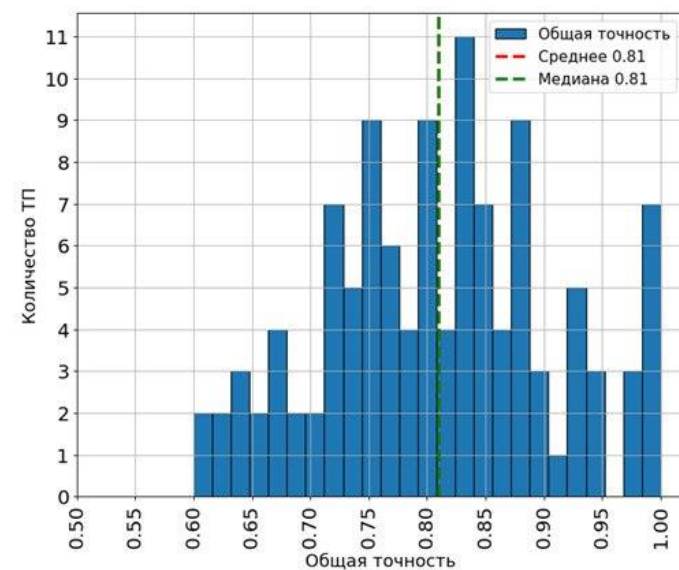


	Порода по данным ДЗЗ											Recall	F1-score	
	Сосна	Ель	Пихта	Лиственница	Кедр	Дуб	Береза	Осина	Ольха	Липа	Ясень			
Порода на ТП	Сосна	2005	84	0	26	3	0	128	16	1	0	0	0.89	0.88
	Ель	97	1451	3	22	0	0	122	23	4	1	1	0.84	0.82
	Пихта	1	16	36	5	1	0	2	1	0	0	0	0.58	0.69
	Лиственница	14	23	2	972	0	1	43	1	2	0	0	0.92	0.90
	Кедр	5	5	0	1	55	0	3	1	0	2	2	0.74	0.78
	Дуб	0	0	0	2	3	100	3	0	0	1	1	0.91	0.93
	Береза	151	178	2	92	4	5	1592	79	11	7	3	0.75	0.75
	Осина	6	46	0	17	0	0	154	492	11	9	0	0.67	0.72
	Ольха	1	6	0	0	0	0	45	7	98	1	3	0.61	0.69
	Липа	1	1	0	1	2	0	24	12	0	50	1	0.54	0.61
	Ясень	0	1	1	0	1	2	0	0	1	1	14	0.67	0.62
	Precision	0.88	0.80	0.82	0.85	0.80	0.93	0.75	0.78	0.77	0.69	0.56	0.815	

		Порода по данным ДЗЗ												
Классификация		Сосна	Ель	Пихта	Лиственница	Кедр	Дуб	Береза	Осина	Ольха	Липа	Ясень	Recall	F1-score
		Порода на ИП	Сосна	1995	88	1	29	4	0	125	19	1	0	0
Ель	117		1403	8	27	1	0	130	36	2	0	0	0.81	0.81
Пихта	3		15	29	6	2	0	6	0	0	0	1	0.47	0.54
Лиственница	17		13	3	975	0	2	45	3	0	0	0	0.92	0.88
Кедр	7		3	1	1	54	0	5	1	0	2	0	0.73	0.73
Дуб	0		0	0	1	0	99	8	0	0	1	1	0.90	0.92
Береза	159		172	2	90	6	1	1564	97	18	12	2	0.74	0.73
Осина	18		44	0	17	0	0	171	466	9	9	0	0.63	0.68
Ольха	2		4	0	0	0	1	49	8	94	0	2	0.58	0.65
Липа	1		1	0	1	4	0	27	10	0	46	2	0.5	0.56
Ясень	0		0	1	0	2	2	3	0	1	1	11	0.52	0.55
Precision		0.86	0.81	0.64	0.85	0.73	0.93	0.73	0.73	0.73	0.65	0.58	0.798	



		Порода по данным ДЗЗ												
Долевое участие		Сосна	Ель	Пихта	Лиственница	Кедр	Дуб	Береза	Осина	Ольха	Липа	Ясень	Recall	F1-score
		Порода на ИП	Сосна	2005	84	0	26	3	0	128	16	1	0	0
Ель	97		1451	3	22	0	0	122	23	4	1	1	0.84	0.82
Пихта	1		16	36	5	1	0	2	1	0	0	0	0.58	0.69
Лиственница	14		23	2	972	0	1	43	1	2	0	0	0.92	0.90
Кедр	5		5	0	1	55	0	3	1	0	2	2	0.74	0.78
Дуб	0		0	0	2	3	100	3	0	0	1	1	0.91	0.93
Береза	151		178	2	92	4	5	1592	79	11	7	3	0.75	0.75
Осина	6		46	0	17	0	0	154	492	11	9	0	0.67	0.72
Ольха	1		6	0	0	0	0	45	7	98	1	3	0.61	0.69
Липа	1		1	0	1	2	0	24	12	0	50	1	0.54	0.61
Ясень	0		1	1	0	1	2	0	0	1	1	14	0.67	0.62
Precision		0.88	0.80	0.82	0.85	0.80	0.93	0.75	0.78	0.77	0.69	0.56	0.815	



Заключение

- Общая точность оценки преобладающих пород леса на уровне пикселя 10 м, рассчитанная по матрице перепутывания на основе подхода out-of-bag для всех ТП с использованием случайных лесов в совокупности составила:
 - **79,8%** в режиме классификации
 - **81.5%** в режиме регрессионной оценки долевого участия пород леса по запасу с последующим определением преобладающей породы
- Значения меры F1-score, характеризующей точность распознавания отдельных пород леса, также оказались устойчиво выше для всех представленных преобладающих пород леса по сравнению с результатами классификации
- На отдельных ТП минимальная точность улучшилась с ~45 % до ~60%
- Полученные результаты свидетельствуют о более высокой эффективности определения преобладающих пород леса на основе регрессионной оценки доли их участия в насаждении по запасу

Территориальное расположение ТП

