

Мониторинг развития лесных экосистем при помощи Sentinel-2 на примере Красносамарского лесничества

Дарья Быкова*, Анна Денисова*, Виктор Федосеев*,
Людмила Кавеленова**, Евгений Корчиков**,

Самарский национальный исследовательский университет

** кафедра геоинформатики и информационной безопасности*

*** кафедра экологии, ботаники и защиты окружающей среды*

vicanfed@gmail.com

Содержание доклада

- Постановка задачи анализа территории. Её специфика
- Описание опорной технологии тематической классификации, предложенной в 2018-19 гг.
- Модификации технологии классификации. Применение для данных 2022-23 гг.
- Способы использования этой технологии для решения исходной прикладной задачи.

Территория. Данные. Задача

Рассматриваемая территория

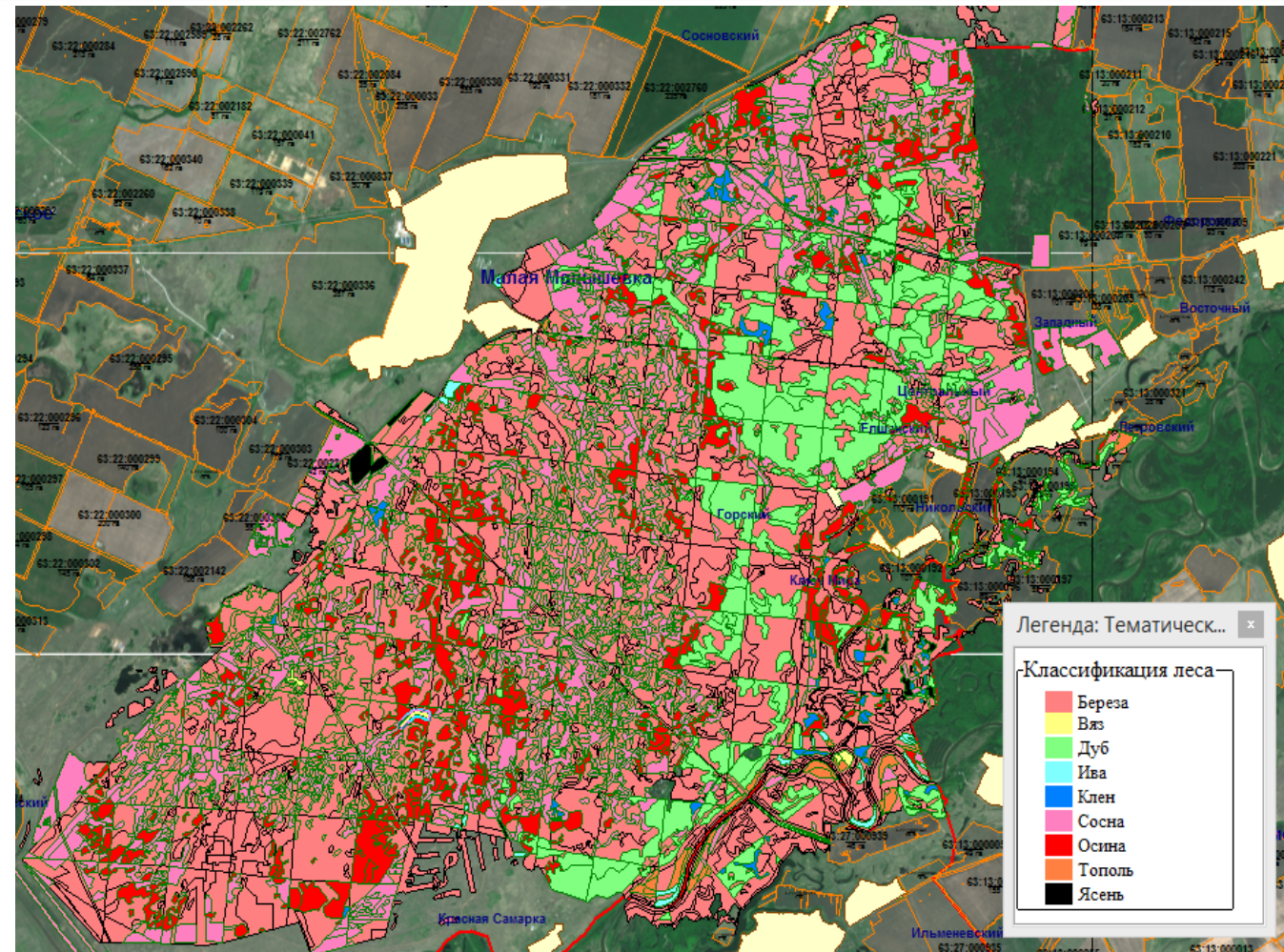
- Красносамарское лесничество, Самарская область, площадь – 35 556 га
- 3800 выделов леса
- Данные таксации леса – 2013-14 гг.
- Таксационная формула: 8Я2С, сомкнутость

Состав леса

- 9 пород деревьев, которые преобладают хотя бы в одном выделе:
берёза, вяз, дуб, ива, клён, осина, сосна, тополь, ясень

Задача

- **Разработка технологии актуализации данных таксации леса путём анализа космоснимков и наземных обследований**



Карта лесничества с выделением преобладающих пород леса

Как решать поставленную задачу?

- **Ядро решения** – модель тематической классификации территории по космоснимкам (применение БПЛА невозможно)
- **Проблема:** обучать её можно только по несвежим данным, которые и надо обновить

Дополнительные вопросы

- Что до классификации?
 - Как формировать обучающие данные?
 - Как обучать модель?
- Что после неё?
 - Как агрегировать результаты классификации?
 - Как их интерпретировать в контексте таксационных выделов?
- Как осуществлять наземный контроль?
В том числе как минимизировать объём полевых исследований?

Предыдущее наше исследование 2018-2019 гг.

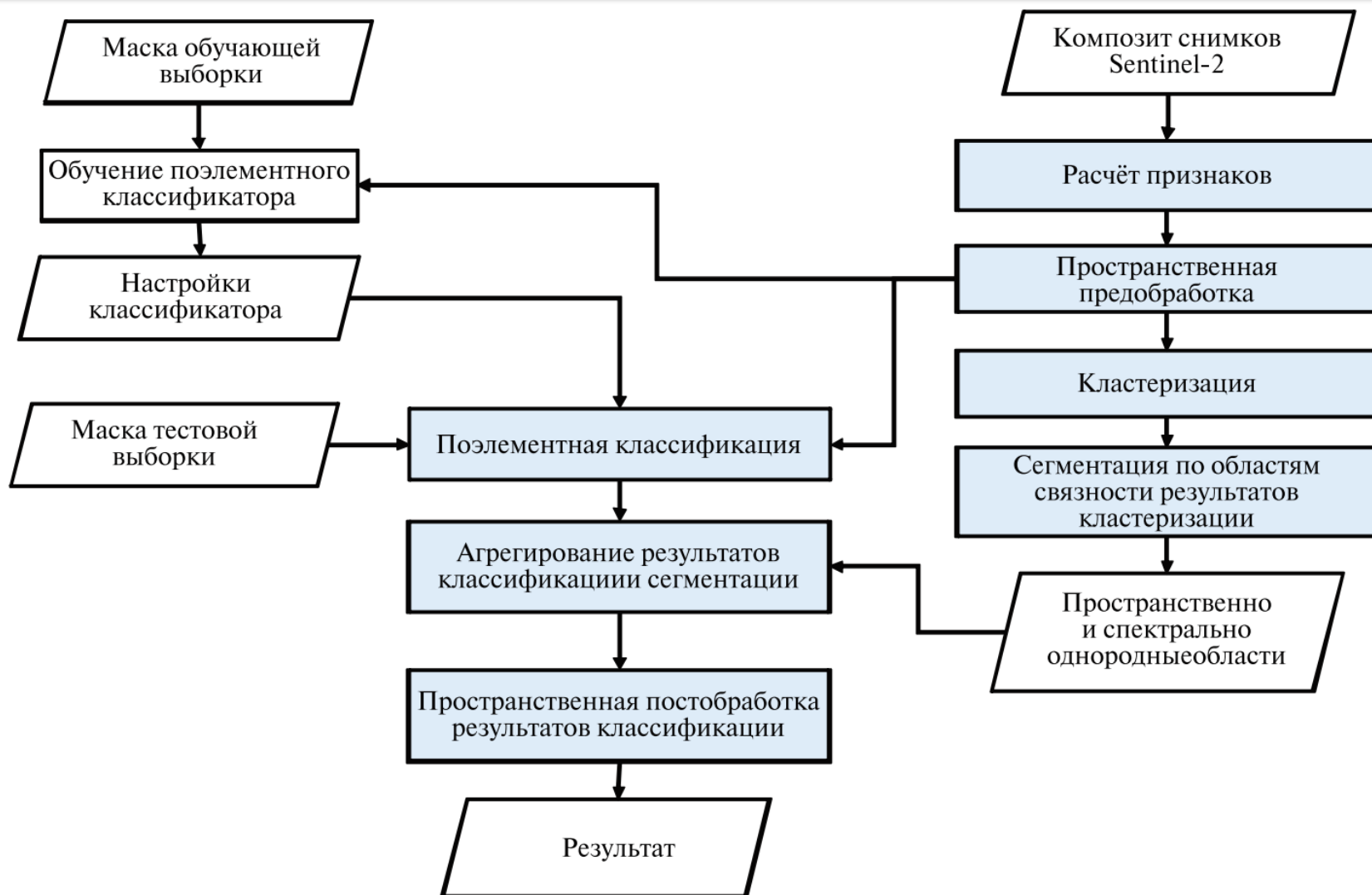
Главное об исследовании

- Основная публикация – [Денисова и др., *Современные проблемы ДЗЗ из космоса*, 2019]
- Цель – построение эффективного классификатора древесных пород при помощи снимков Sentinel-2 и данных таксации леса
- Основное применение – обнаружение определённых пород леса на других территориях
- Основная идея: **использование композитов** из нескольких снимков Sentinel-2 одного вегетационного сезона
- Классификация: **пространственно-спектральная** процедура классификации гиперспектральных снимков ДЗЗ (основа - [Борзов и Потатуркин, *Автометрия*, 2018], [Fedoseev, *JOP: Conf Series*, 2018])

Основные результаты

- Точность классификации на всём Красносамарском лесничестве – **0.8535**
- Предпочтительны данные уровня L2A (с атмосферной коррекцией)
- Композит должен включать серию весенних снимков
- Процедуры пространственной пред- и постобработки повышают точность классификации на 2-10%
- В композиты целесообразно включать 20-метровые каналы, разрешение которых повышено до 10 м.
- Добавление производных каналов (например, NDVI) не повышают точность классификации

Общая схема процедуры классификации (2018-19 гг.)



Детали формирования выборки

- Для обучения/тестирования используются участки с долей 8/10 преобладающей породы
- 8 каналов: 2-4, 8 (10 м) + 5-7, 8A (20 м → 10 м) + {NDVI}
- Выбранные классы (13): 9 пород леса, водоёмы, постройки, гари/прогалины, смешанный лес

Методы и алгоритмы, используемые в процедуре классификации

Расчёт признаков

- нормализация – приведение к диапазону $[-1,1]$
- преобразование к главным компонентам (PCA)

Пространственная предобработка

- медианная фильтрация

Кластеризация

- k-means++ или обобщенный EM алгоритм

Сегментация

- выделение областей связности по результатам кластеризации

Поэлементная классификация

- SVM-RBF

Агрегирование результатов классификации и сегментации

- голосование большинства по сегменту или поэлементная классификация медиан сегментов
- постобработка скользящим окном.

Модификации по сравнению с исследованием 2018-19 гг.

- Актуальные снимки 2022-23 гг.
- Убрали процедуру сегментации, поскольку нужно отталкиваться от имеющейся сегментации данных – деления леса на выделы
 - *«Таксационный выдел представляет собой ограниченный участок лесного фонда, относительно однородный по почвенно-грунтовым условиям и по составу произрастающей на нем растительности или иных поверхностных структур, в котором изменчивость его качественных и количественных характеристик не превышает нормативных допусков»*
- Убрали переход к главным компонентам
- Убрали класс «смешанный лес»
- Добавили каналы B11-B12, не использовавшиеся в предыдущем исследовании
- Добавили некоторый учёт сомкнутости крон – в обучающей выборке используются только участки с сомкнутостью 0.5 и выше
- Исследовали разные модели классификаторов
- Исследовали эффект фильтрации аномалий методом IsolationForest (пока не надежды не оправдались)
- Сократили число обучающих данных для повышения производительности: $\min(70\%, 10000)$ per class

Исследование качества классификации для разных композитов

№ композита	1	2	3	4	5	6	7
Состав композита	2022: апрель - сентябрь	2022: апрель - июнь	2022: апрель - май	2022: апрель, май, сентябрь	2023: апрель - май	2023: апрель - август	2022-2023
Число каналов	200	80	60	70	140	180	380
Пространственная предобработка + Поэлементная классификация	0.8629	0.8299	0.8229	0.8301	0.8510	0.8594	0.8776
+ пространственная постобработка	0.8667	0.8360	0.8298	0.8362	0.8550	0.8627	0.8802

Выводы

- Весеннего сезона достаточно для обеспечения приемлемого качества классификации
- При увеличении числа снимков / сезонов в составе композита качество возрастает, но в пределах 5%

Исследование альтернативных моделей классификации

- Модели: Случайный лес, Деревья решений, бустинг, kNN...
- Опции: Учитывать / не учитывать априорные вероятности классов,
- Производилась оптимизация гиперпараметров среди лучших вариантов

Классификатор	SVM-RBF фиксированный		SVM-RBF оптимизированный		Random Forest оптимизированный	
	Равные	Пропорционально обучающей выборке	Равные	Пропорционально обучающей выборке	Равные	Пропорционально обучающей выборке
Пространственная предобработка + Поэлементная классификация	0.8499	0.8594	0.8658	0.8702	0.8322	0.8484
+ пространственная постобработка	0.855	0.8627	0.8742	0.8755	0.8399	0.8542

Выводы

- Мультиклассовая модель SVM-RBF среди классических моделей так и осталась лучшей
- Random Forest уступает на 2%
- Оптимизация гиперпараметров позволяет повысить точность на 1%

При обучающей выборке в 25% размеченной части снимка точность составила **0.9048**

Ошибки как ключевой источник информации

- Причины ошибок классификации в отдельных пикселях
 - Неточная привязка снимка / неточные координаты выделов в данных таксации
 - Сомкнутость не 100%
 - Смешивание в пределах одного пикселя 10×10 м.
 - Смешанный тип леса – деревья растут попеременно
 - Плохо обучен классификатор
 - Выдел неоднороден: 8С1Б1КЛ, и текущий пиксель соответствует не основной породе для данного выдела
 - Реальные изменения на местности – **то, что мы ищем**
- **Ошибки классификации – наш источник информации**, а обучающие данные не точны. Поэтому строго говоря нет смысла сводить к минимуму к ошибки при обучении классификатора и настройке обучающей процедуры
- Нужно построить автоматизированную процедуру так, чтобы **минимизировать ошибки при последующей наземной верификации**

Процедура уточнения данных таксации леса

Два проверенных варианта:

1. Ориентация на лесные выделы (проверено 15 точек)
 1. Применяем классификатор
 2. Агрегируем в пределах выделов
 3. Выявляем ошибки классификации
 4. Производим наземное обследование
2. Ориентация на ошибочно классифицированные области (проверено 13 точек)
 1. Применяем классификатор
 2. Ищем ошибки
 3. Выявляем крупные сегменты в поле ошибок
 4. Производим наземное обследование

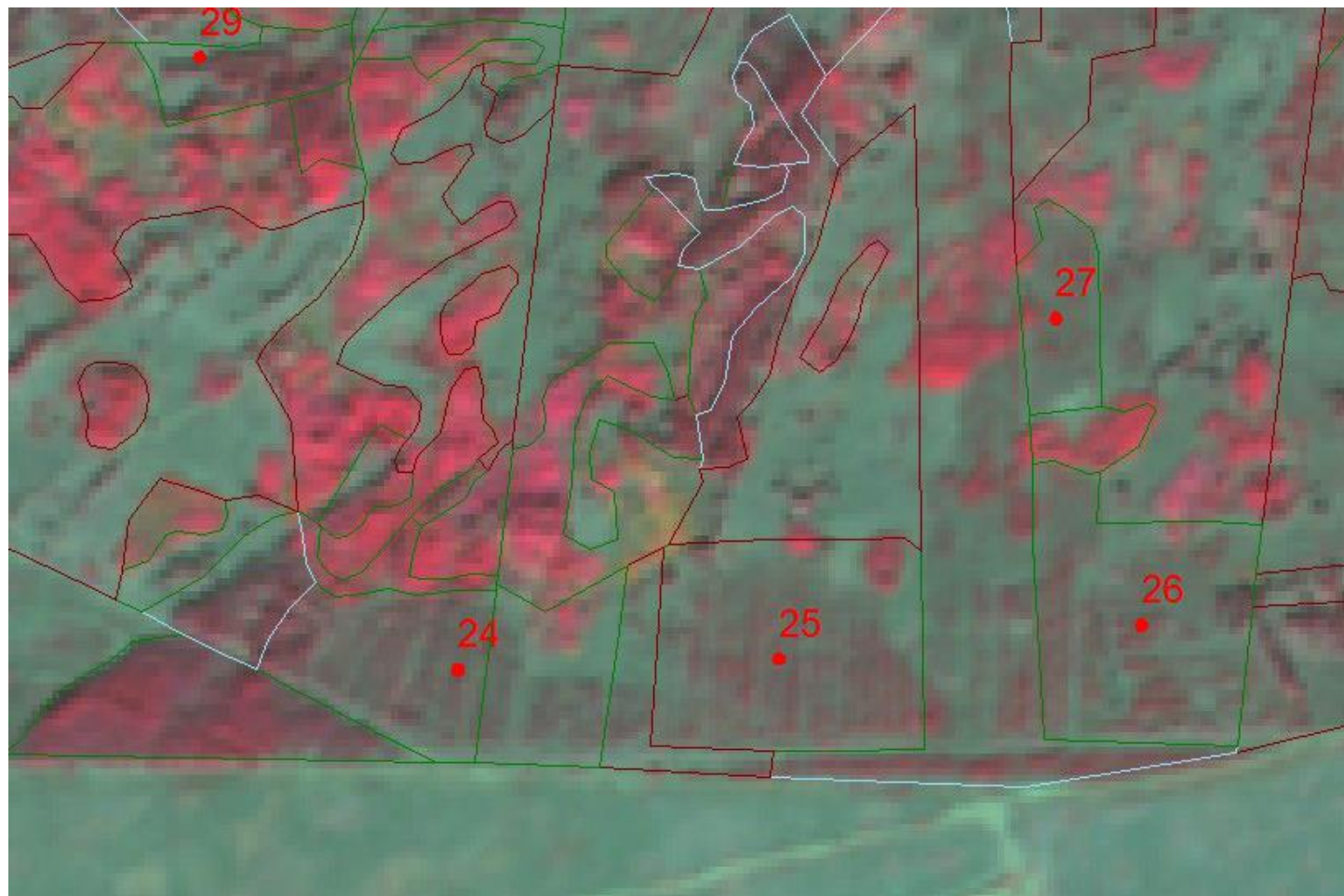
Процедура уточнения данных таксации леса

Реальные изменения на местности:

- Разрастание дубовых и сосновых сообществ на территории бывших осиновых участков
- Разрастание сосновых участков на территорию березняков
- Гари
- Молодые посевы сосны

Ложные точки:

- Классификатор ошибся на распространённых классах при низкой сомкнутости
- Выдел неоднороден. В окрестности тестовой точки реально оказался не преимущественный класс
- Центр масс кластера оказался вне выдела
- Смещение координат между снимком и данными ГИС



Молодые посевы сосны ошибочно классифицируются как прогалы

Выводы по итогам полевого обследования

Сопоставление методов

- 1) Ориентация на лесные выделы – около 50% неинформативных точек
- 2) Ориентация на ошибочно классифицированные области – около 25% неинформативных точек
- **Следствие:** нужно пересечь два разбиения: таксационные выделы и участки с однородной классификацией – и ошибки выделять по этому новому разбиению

Дополнительные следствия

- Возможно, требуется ручная допривязка снимков
- Проверять выбранные точки (центры областей) – попадают ли в выбранный сегмент
- Просмотр нескольких сезонов: чем больше сезонов, тем больше вероятность реальных изменений

Заключение

- «Реанимирован» подход к классификации пород леса на основе композитов Sentinel-2
- Проведены исследования и скорректированы детали выбранного подхода
- Проанализированы подходы к решению задачи обновления данных таксации леса
- Проверены два подхода, проведена верификация на местности

- Работа продолжается, мы постараемся сделать больше и лучше

Спасибо за внимание!