

Метод построения эталонов сезонной динамики вегетационного индекса NDVI для однолетних сельскохозяйственных культур территории России

Хвостиков С.А., Барталев С.А.

khvostikov@d902.iki.rssi.ru



Задача верификации данных о с.-х. культурах

Большой объем информации о культурах на полях сельскохозяйственного назначения, обладающих значительным уровнем погрешности, приводит к необходимости разработки метода верификации этой информации.

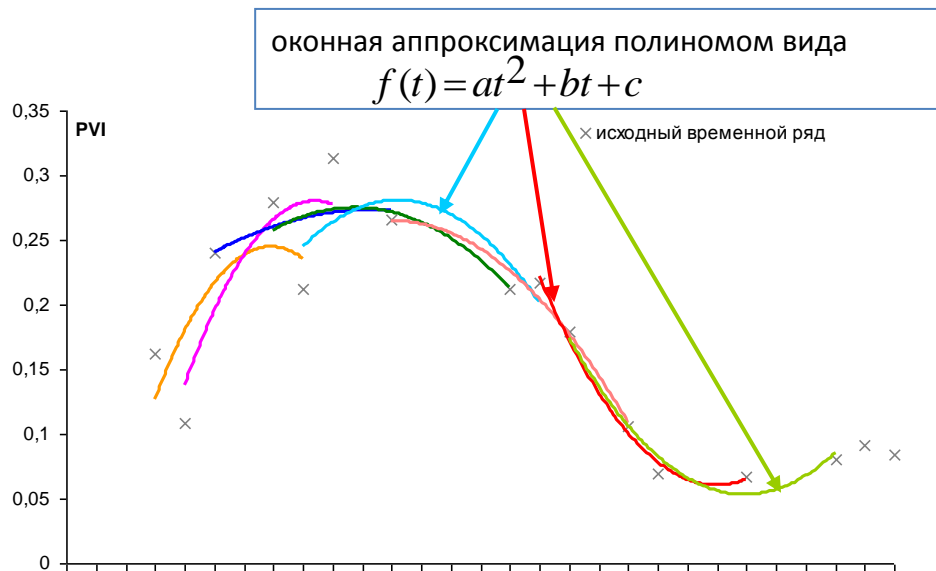
Предложенный в работе метод основан на построении эталонов культур и определении достоверности информации по ним. Метод использует только самую информацию о культурах и данные ДЗЗ.

Для каждого поля есть данные о культуре и его периметре.

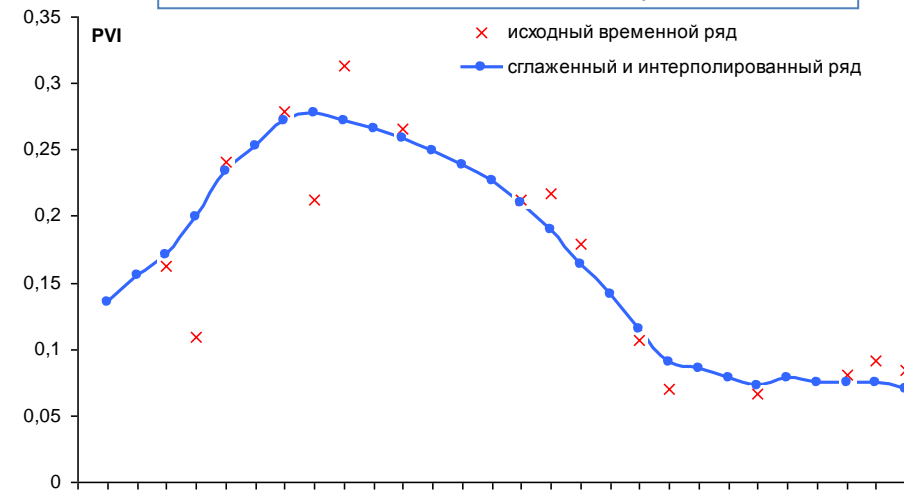
	2016	2017	2018
Всего полей	141258	257576	151740
Пшеница озимая	49433	63312	39439
Подсолнечник	11045	20317	10061
Кукуруза на зерно	8084	19805	3139
Пшеница яровая	4265	15686	20813
Ячмень яровой	5211	11536	3642
Соя	2632	7857	3424

Временной ряд вегетационного индекса NDVI

Для каждого поля были получены еженедельные осредненные данные MODIS о вегетационном индексе NDVI. Полученный временной ряд также был сглажен методом полиномиальной аппроксимации.



1. каждое измерение получает n аппроксимирующих оценок x_t
2. элемент скорректированной серии имеет значение $\bar{x} = \frac{\sum x_t}{n}$

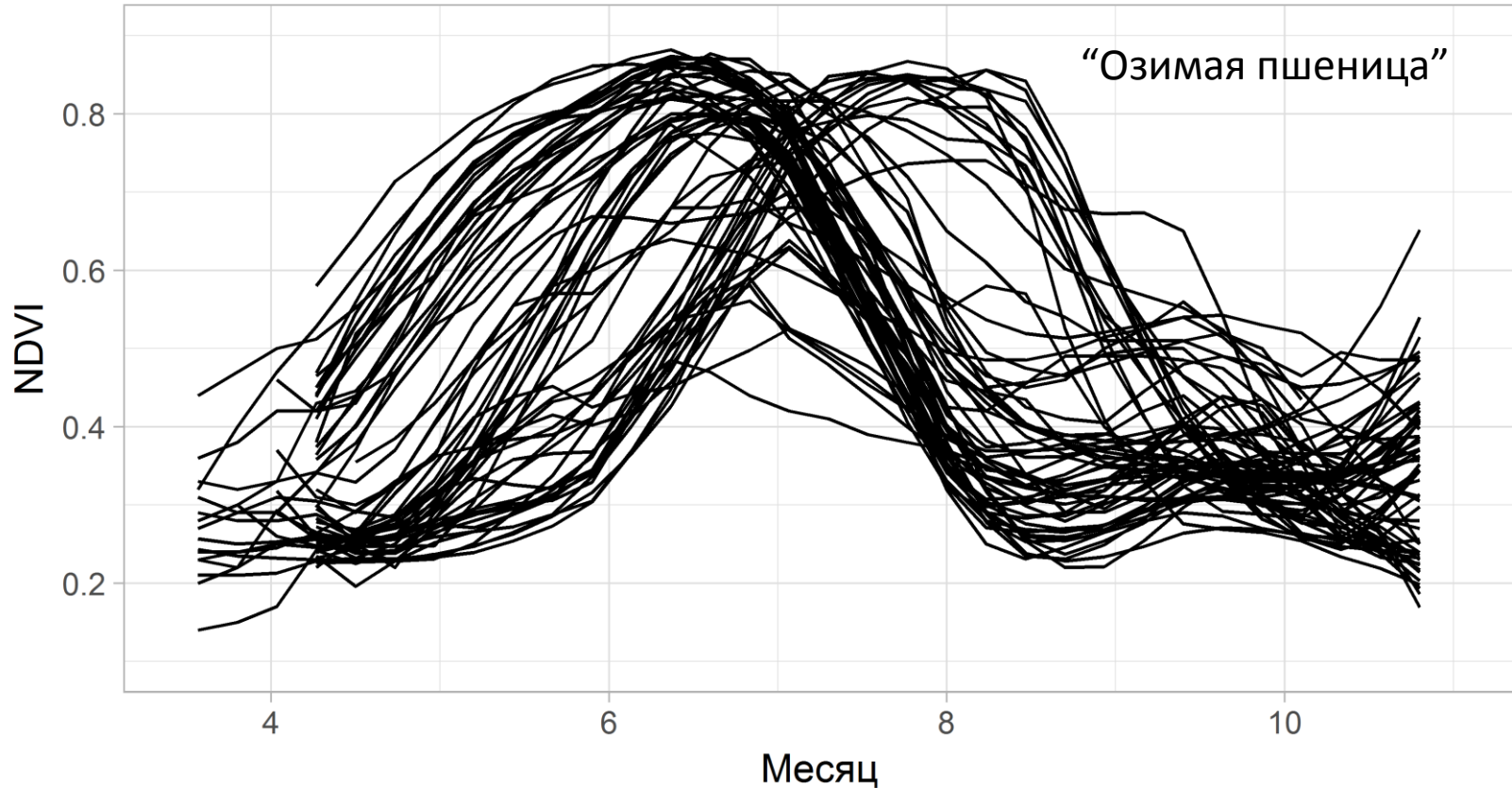


Плотников Д.Е., Миклашевич Т.С., Барталев С.А. Восстановление временных рядов данных дистанционных измерений методом полиномиальной аппроксимации в скользящем окне переменного размера // Современные проблемы дистанционного зондирования Земли из космоса. 2014. Т. 11. №. 2. С. 103-110.

Пример погрешностей данных о культурах

На слайде представлены временные ряды индекса NDVI для полей, относящихся к «озимой пшенице» по данным ЕФИС ЗСН

Горшеченский район, Курская область, 2017 год.



Общая схема построения эталонов и последующей верификации

Построение эталонов на основе данных о индексе NDVI и культурах на полях (выполняется на уровне районов):

- Выделение кластеров в динамике индекса NDVI;
- Выбор наиболее представительного кластера;
- Построение эталонов (многомерного нормального распределения) по данному кластеру
- Сравнение эталонов соседних районов, замена «нетипичных» эталонов

Верификация данных о культурах на полях на основе эталонов:

- Сравнение эталонов всех культур в одном районе, определение «неразличимых» эталонов
- Сравнение поля со всеми эталонами района для оценки корректности заявленной информации о культуре

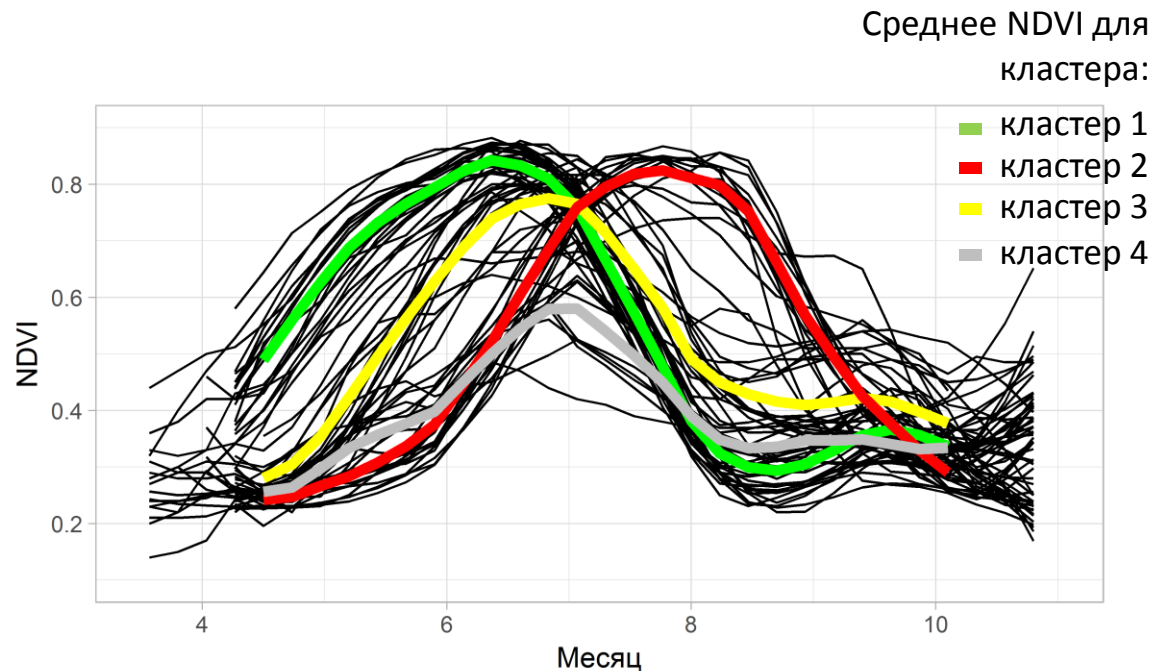
Построение эталонов динамики сельскохозяйственных культур

Построение эталонов динамики выполняется на уровне районов.

В случае, если в районе менее 300 полей данной культуры, осуществляется добор данных из соседних районов, удаленных не более чем на 300 км от целевого района. Если после агрегации данных не удалось найти 300 полей, то эталон данной культуры не строится в данном районе.

Выполняется кластеризация временных рядов NDVI методом k-средних.

Для определения числа кластеров оценивается минимальное среднее расстояние между центрами всех кластеров. Если оно окажется ниже порога 0,1, то берется предыдущий вариант кластеризации.



Построение эталонов динамики с.х. культур

В работе считается, что эталонное поведение культуры можно описать многомерным нормальным распределением, каждый элемент которого соответствует значению NDVI в заданную неделю.

Ряды NDVI полей кластера с наибольшим числом элементов используется для оценки вектора средних и матрицы ковариации многомерного нормального распределения эталона.

Вводится понятие неразличимых эталонов, на основе расстояния Бхаттачария :

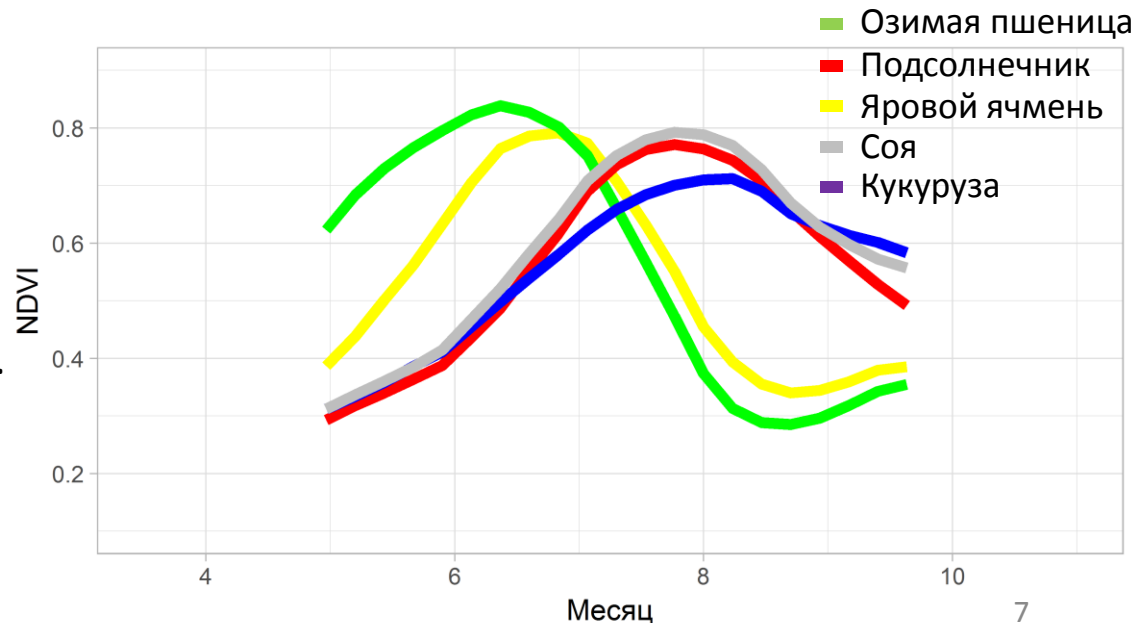
$$D = \frac{1}{8} (\mu_1 - \mu_2) \Sigma^{-1} (\mu_1 - \mu_2)^T + \frac{1}{2} \ln \left(\frac{\det \Sigma}{\det \Sigma_1 \det \Sigma_2} \right)$$

$\mu_{1,2}, \Sigma_{1,2}$ - параметры норм. распр.

$$\Sigma = \Sigma_1 + \Sigma_2$$

Если оно меньше 2,5, то эталоны неразличимы.

Средние значения эталонов основных культур

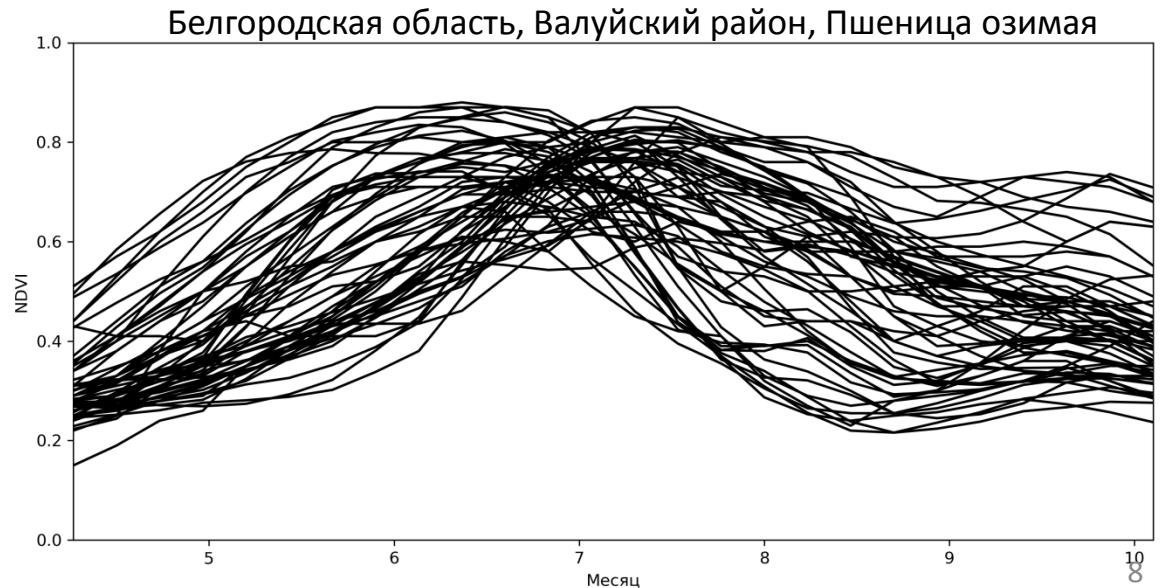


Сравнение эталонов соседних районов

В отдельных районах поля с неверной информацией о культуре могут преобладать. Эталоны в таких районах заменяются посредством сравнения с эталонами соседних районов.

Берутся эталоны соседних районов, удаленные не более чем на 150 километров. Для 10 ближайших эталонов оценивается их взаимная неразличимость.

Также для эталонов вводится понятие веса, как число полей данной культуры в районе деленое на порог 300 полей.



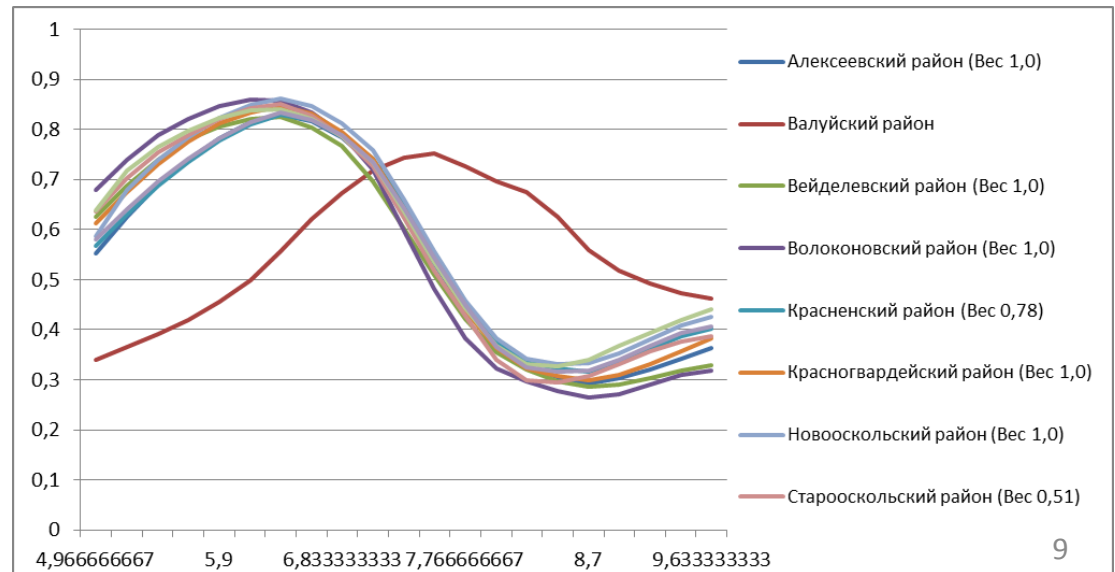
Сравнение эталонов соседних районов

Если эталон неотличим от 66% эталонов соседних районов (по их весам), то эталон признается корректным.

В противном случае осуществляется поиск другого ближайшего эталона, неотличимого от 66% остальных эталонов (по весам). Если такой эталон найден, то происходит замена на него в текущем районе.

Если не удастся достичь консенсусного значения 66%, то в список эталонов добавляется еще один эталон (удаленный не более чем на 150 км) и процедура повторяется.

Если были проанализированы все районы в области 150 км, и консенсус так и не был достигнут, то эталон также считается корректным и не изменяется.



Верификация наземных данных о культурах

Для верификации используется расстояние Махаланобиса между временным рядом NDVI поля и эталоном культуры:

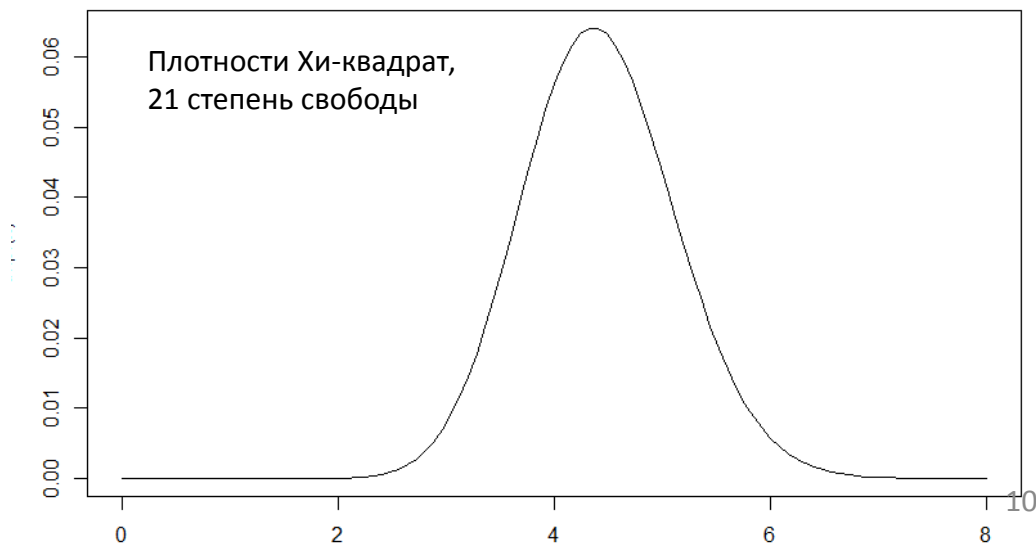
$$D = \sqrt{(x - \mu)\Sigma^{-1}(x - \mu)^T}$$

Где x - временной ряд NDVI на поле, μ, Σ - параметры нормального распределения эталона

На основе распределения Хи-квадрат задается порог, в пределах которого находится 95% элементов нормального распределения. Если расстояние Махаланобиса превышает этот предел, то на поле точно не может быть данная культура.

Доля выборки, расстояния до которой меньше чем D

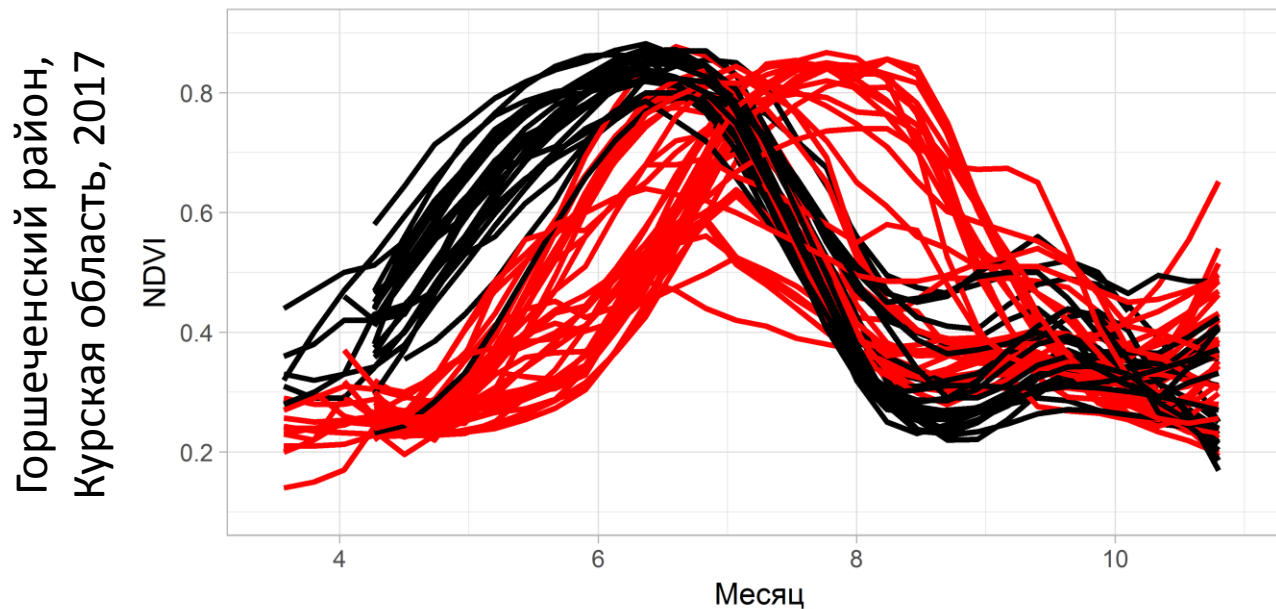
% выборки	D
0,1	3,64
0,3	4,14
0,66	4,8
0,9	5,44
0,95	5,72
0,99	6,24



Верификация наземных данных о культурах

Оценка корректности данных о культурах происходит на уровне районов. Сначала все эталоны в районе сравниваются попарно на неразличимость.

Затем каждое поле сравнивается с эталоном каждой культуры. Если ближайший к полю эталон соответствует заявленной на поле культуре (или неотличим от нее), то данные на поле признаются корректными (верифицированными).



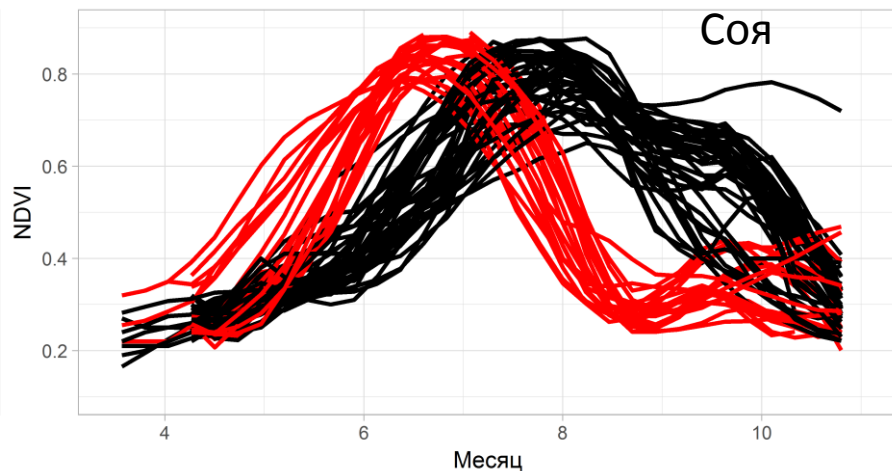
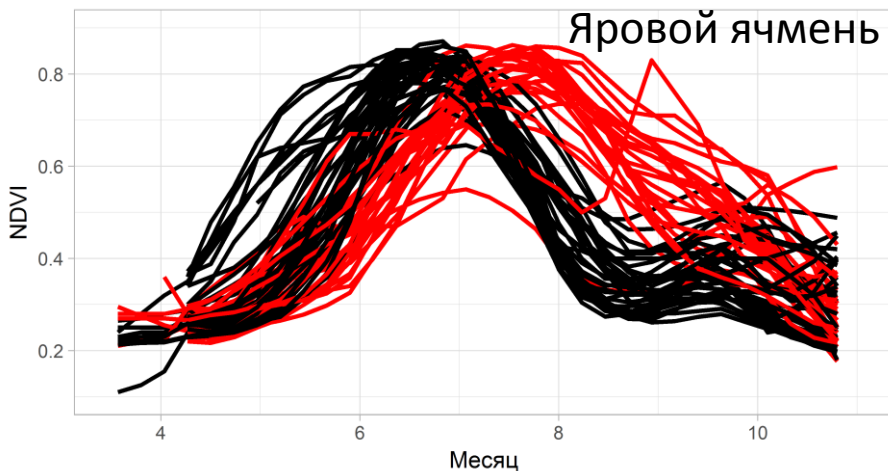
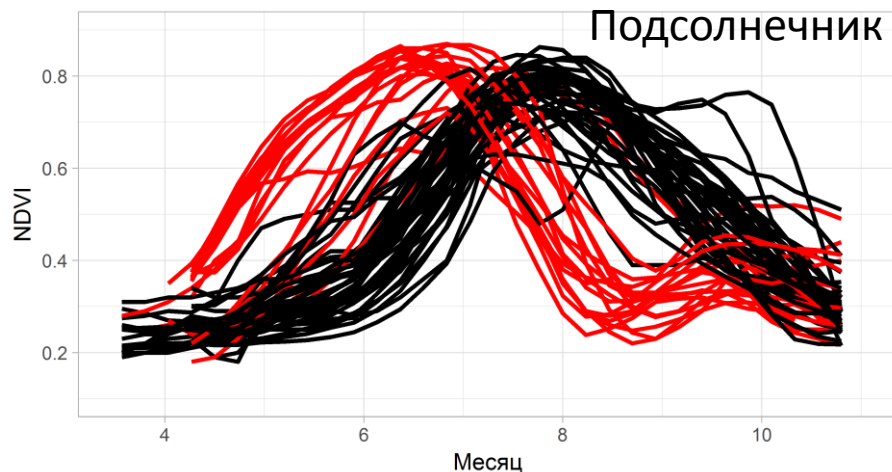
■ - поле прошло верификацию

■ - поле не прошло верификацию ¹¹

Пример результатов верификации

В 2017 году около 20% полей не проходит верификацию.

Также 5% полей не удастся верифицировать из-за отсутствия статистических данных (нет 300 полей культуры в радиусе 300 км).

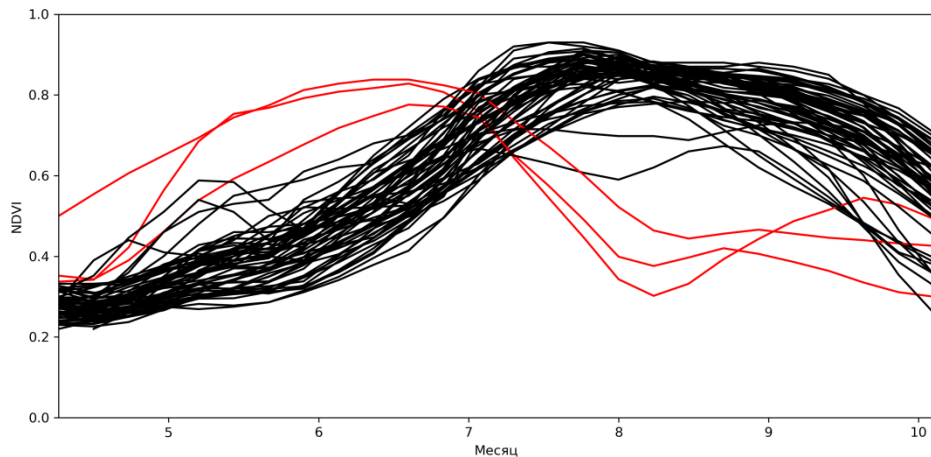


■ - поле прошло верификацию

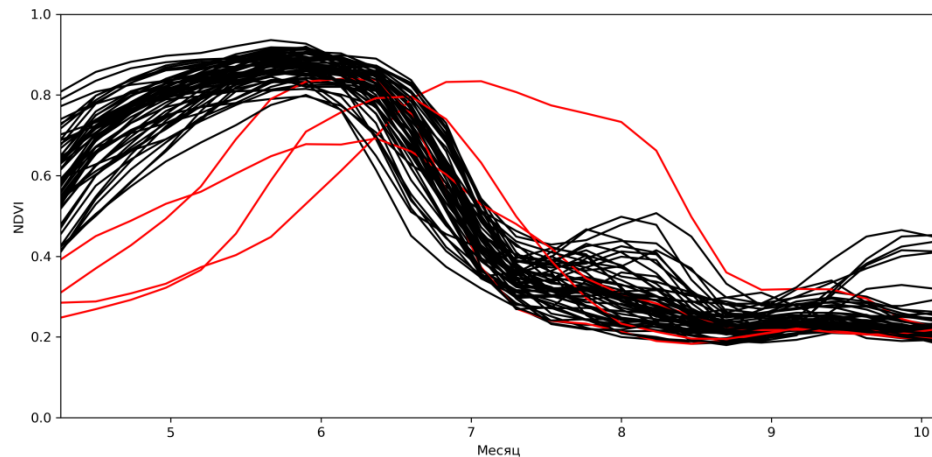
■ - поле не прошло верификацию

Пример результатов верификации

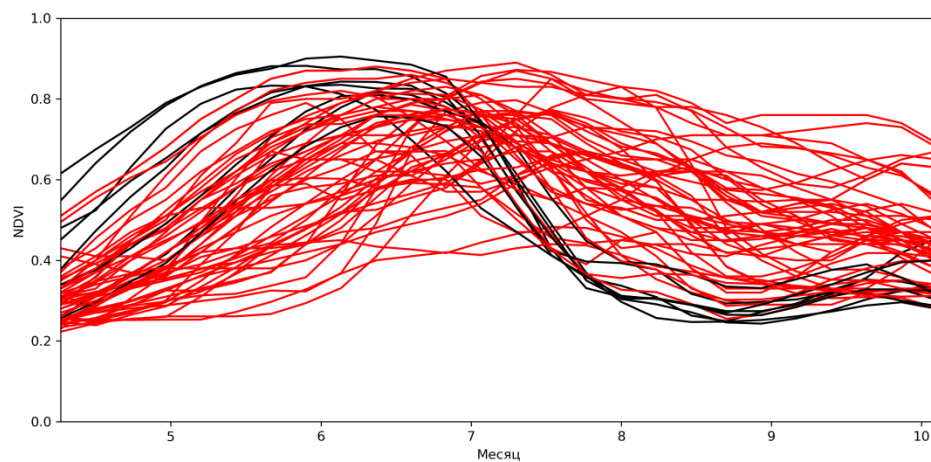
Обоянский район, Курская область, Кукуруза



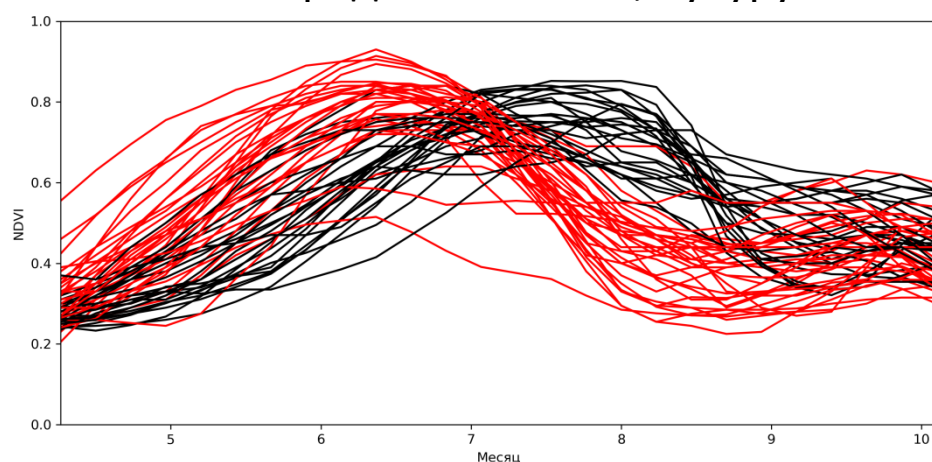
Ейский район, Краснодарский край, пшеница озимая



Белгородская область, Валуйский район, Пшеница озимая



Красногвардейский район, Белгородская область, Кукуруза



■ - поле прошло верификацию

■ - поле не прошло верификацию 13

Подготовка данных для классификации

В задаче классификации может быть нежелательно наличие в обучающей выборке полей с неверно указанной культурой.

Для подготовки данных для классификации предлагается внести следующие изменения в метод и пороги:

- Убрать порог на неразличимость эталона;
- Ввести дополнительный порог на близость ближайшего эталона (квантиль).
- Или порог на разницу расстояний Махаланобиса

Величина порога позволяет варьировать размер выборки:

Квантиль	95%	80%	ΔD
Пшеница озимая	3162	4011	4905
Кукуруза	72	198	310
Подсолнечник	125	268	323
Соя	53	144	291
Ячмень озимый	294	427	531
Ячмень яровой	119	260	460
Пшеница яровая	150	320	410

Дальнейшие шаги

Оценка возможности построения карт на основе полученной в результате верификации выборки.

Использование методов интеграции Монте-Карло по пересечения нормальных распределений эталонов для оценки аналогов ошибки 1 и 2 рода.

Использование методов оценки начала и конца сезона вегетации при построении эталонов и верификации.