



Анализ результатов эксперимента по созданию опорных выборок на больших территориях с помощью краудсорсинга

Ёлкина Е.С.¹, Плотников Д.Е.¹, Толпин В.А.¹, Щербенко Е.В.²
¹ Институт космических исследований РАН, Москва, Россия
² ООО ИКИЗ, Москва, Россия

Актуальность

Для картографирования объектов различного типа землепользования на больших территориях и в различные эпохи, в том числе, используемых пахотных земель, требуется пространственно-распределенная и точная обучающая выборка. Вопрос создания опорных выборок на большие территории особо остро стоит для исторического картографирования, для эпох до запуска спутников семейства Sentinel-2.

В качестве альтернативы получению опорной информации с помощью экспертов/готовых картографических продуктов рассматривается краудсорс-подход: организация фотоинтерпретации спутниковых изображений группой волонтеров.

Потенциальными преимуществами подхода является скорость обработки больших территорий, относительно низкие затраты на организацию и оплату работы, возможность сократить количество экспертов до определённого минимума.

Краудсорсинг (англ. crowdsourcing, от crowd — толпа и sourcing — использование или привлечение ресурсов) — привлечение к решению тех или иных проблем инновационной производственной деятельности широкого круга лиц для использования их творческих способностей, знаний и опыта по типу субподрядной работы на добровольных началах с применением информационных технологий*

*<https://ru.wikipedia.org/wiki/Краудсорсинг>

Цели и задачи эксперимента

Цель эксперимента - оценка применимости подхода к созданию опорных выборок *с известной точностью* с помощью краудсорс-фотоинтерпретации спутниковых изображений на большие территории.

Задачи:

1. Организация эксперимента: создание инструмента для краудсорс-разметки; подготовка инструкций; обучение участников; создание репрезентативного набора точек для обработки (сэмплинг);
2. Сбор итоговых данных, подготовка к анализу;
3. Выбор метрик оценки качества работы волонтеров;
4. Оценка точности волонтеров: средние показатели, распределение точностей внутри группы. Расчёт показателей, позволяющих определить волонтеров с достаточно высокой стабильной (не зависящей от региона) точностью;
5. Оценка пространственных тенденций точности на больших территориях. Оценка возможности определения территорий с высокой точностью большинства участников-волонтеров и территорий, сложных для интерпретации, в которых требуется привлечение экспертов.

Условия эксперимента. Определения

Волонтёры - участники эксперимента без опыта в дешифрировании спутниковых изображений и ДЗЗ.

Волонтёрам проводится краткий инструктаж по фотоинтерпретации, а также по системе, в которой выполняется работа (Bege-science), устанавливаются единые признаки отнесения объекта на снимке к классу. Задача краудсорсеров - по временному ряду изображений определить класс используемости объекта вокруг заданной точки, нарисовать однородный объект и присвоить ему метку класса (пашня или не-пашня).

Эксперты - участники эксперимента с опытом в сфере обработки и дешифрирования спутниковых изображений. Задача экспертов - создание контрольной выборки (истинных меток объектов). Контрольная выборка охватывает лишь часть точек, обработанных волонтёрами, и служит для проведения оценки точности.

Обрабатываемые объекты: точки, распределённые в основном по Европейской части России, попадающие на условную “травянистую растительность”

Классы: “пашня”, “не пашня” и “неизвестно”. Метка “неизвестно” использовалась в случае затруднений, и при оценке точностей не учитывалась (не считалась за ошибку).

Используемая пашня - участок земной поверхности, имеющий в исследуемом сезоне (2018 год) признаки распашки, выращивания сельскохозяйственных культур или обработки с целью сенокошения

Не пашня - участок земной поверхности без признаков использования (распашки/укосов) в исследуемом году.

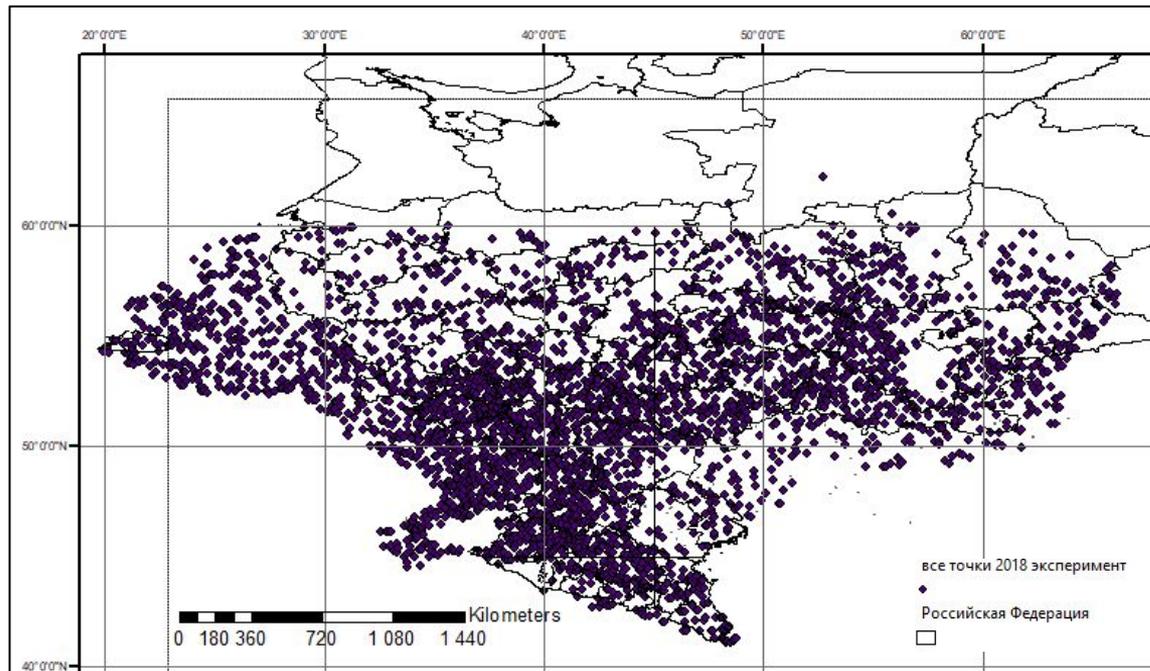
Условия эксперимента

- Задача для участника: волонтерам предлагалось детектировать по спутниковым изображениям объект вокруг заданной точки и отнести его к одному из двух классов: “пашня” или “не-пашня”
- В эксперименте приняли участие 19 волонтеров и несколько экспертов
- Эксперимент проходил в 2018 году, использованы данные 2017 года
- Географический охват: Европейская часть России и прилегающие территории
- Количество точек на человека: около 400 точек
- Время выполнения работы: 2 месяца

Географический охват и расположение точек

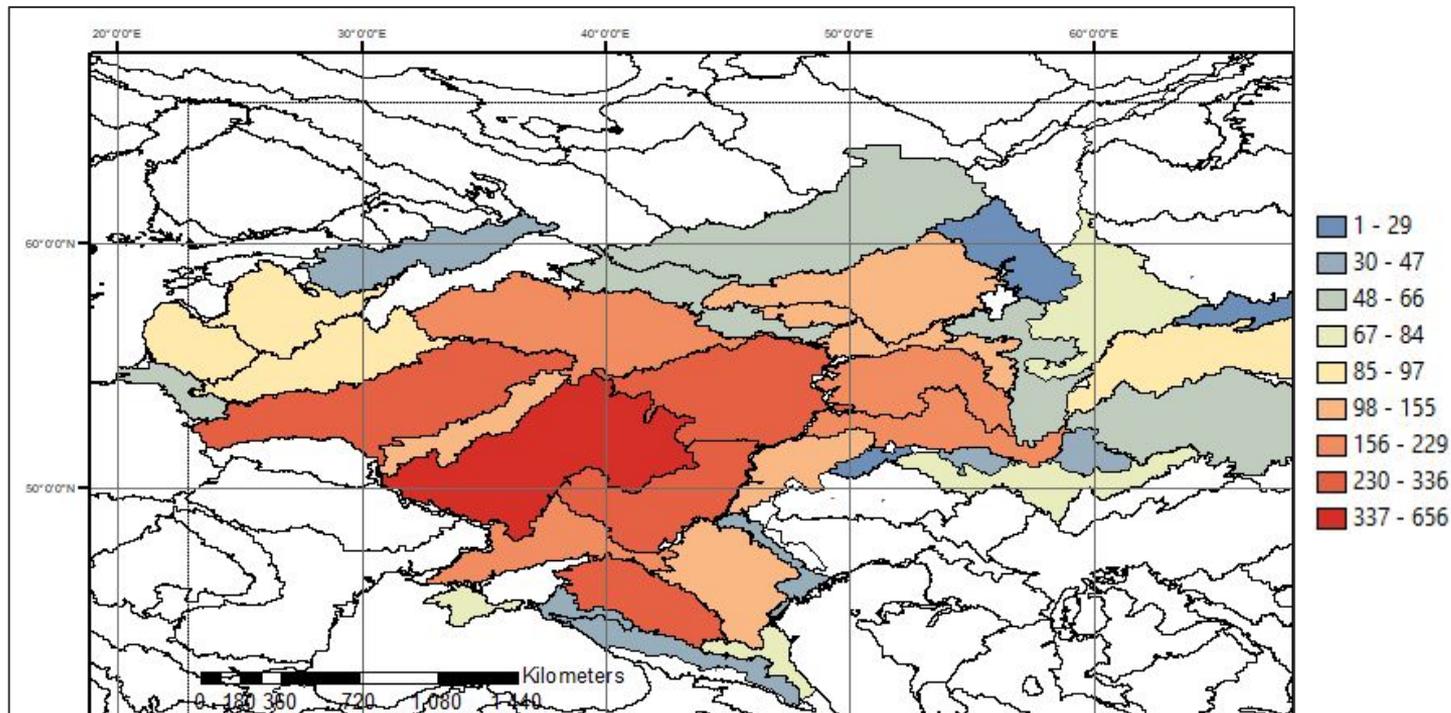
Точки расположены равномерно по территории России, небольшая доля точек расположена на территории прилегающих стран (Украина, Казахстан, Латвия и др.). В дальнейшем при использовании локально-адаптивного классификатора благодаря достаточной большой зоне охвата точками будет возможность обучаться на данных в окрестности участка.

Всего участников	11
Всего точек	4866
Контрольных	2437



Географический охват

Пространственный анализ проводился по агроэкологическим зонам (стратам) GAEZ* - Global Agro-Ecological Zones, которые отражают преобладание определенных экосистем, агропрактик, а также климатические условия и агроклиматические ресурсы территории.

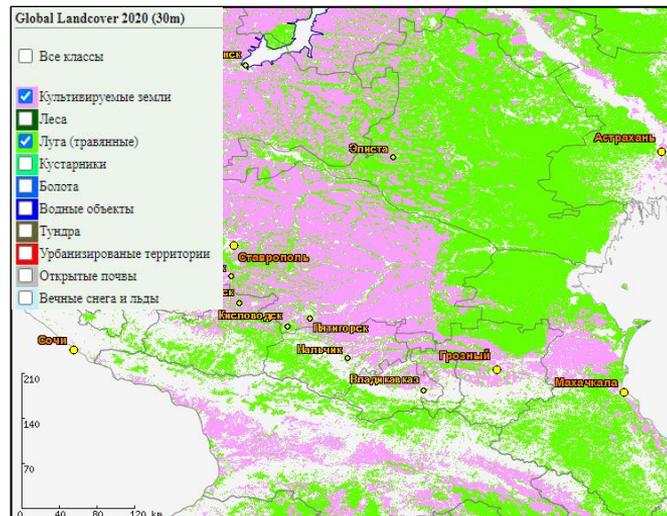


Распределение количества точек по агрозонам

Географический охват: сэмплинг

С целью обеспечения репрезентативности и полноты получаемой выборки точки по исследуемой территории распределялись методом случайного стратифицированного сэмплинга в несколько этапов:

1. Карта наземного покрова GlobeLand30 с пространственным разрешением 30 метров (Chen et al., 2015 [1]) была использована для извлечения данных, соответствующих классу “культивируемые земли” и “луга”. По этим данным создана маска, внутри которой проводился сэмплинг;
2. Внутри маски была проведена сегментация временных серий данных MODIS, в результате были получены однородные сегменты растительного покрова (Плотников и др., 2018 [2]);
3. Для каждого сегмента определен центроид - потенциально будущая опорная точка;
4. Проведено прореживание точек таким образом, чтобы на каждого участника эксперимента приходилось не менее 200 и не более 500 уникальных точек, расположенных в различных агрозонах страны.



Карта наземного покрова GlobeLand30

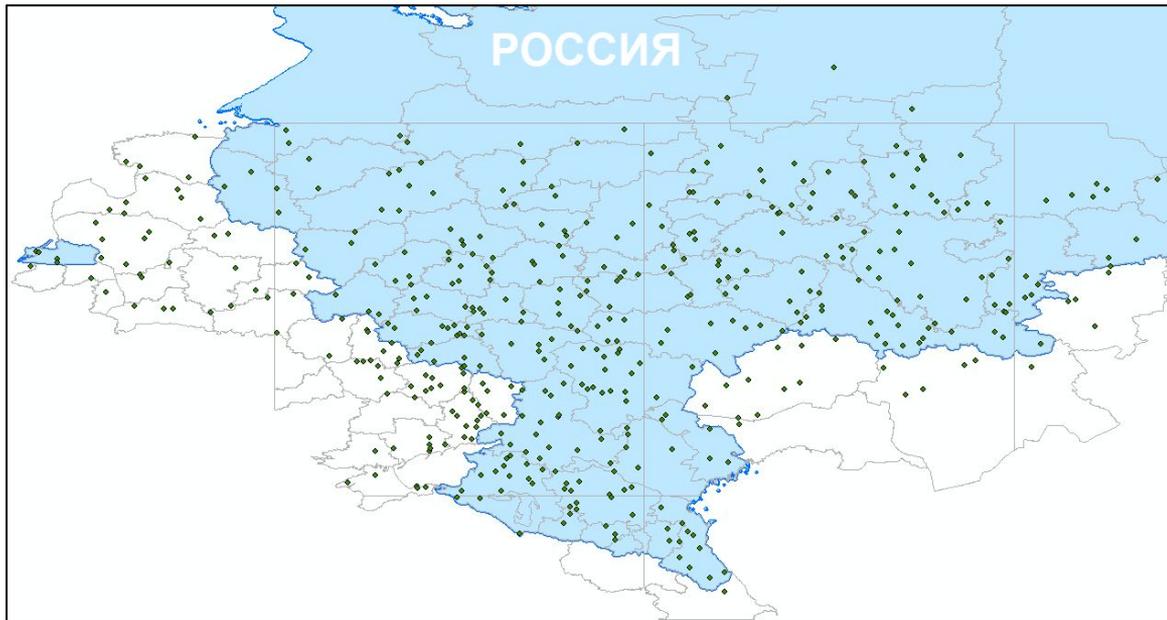
[2] Плотников Д.Е., Колбудаев П.А., Барталев С.А. Выделение сезонно-однородных областей на основе анализа временных серий спутниковых изображений // Компьютерная оптика. 2018. Т. 42. № 3. С. 447-456. DOI: 10.18287/2412-6179-2018-42-3-447-456. URL: <http://www.computeroptics.smr.ru/KO/PDF/KO42-3/420313.pdf>

[1] Jun Chen, Jin Chen, A. Liao, X. Cao, Lijun Chen, Xuehong Chen, C. He, G. Han, S. Peng, M. Lu, W. Zhang, X. Tong, J. Mills. Global land cover mapping at 30m resolution: A POK-based operational approach. ISPRS Journal of Photogrammetry and Remote Sensing, Volume 103, 2015, Pages 7-27, ISSN 0924-2716, <https://doi.org/10.1016/j.isprsjprs.2014.09.002>.

Доля контрольных точек

Половина всех точек участника - контрольные (то есть для них известен истинный класс). Точки, которым участники поставили класс "неизвестно" (не смогли определить класс), были исключены из анализа точностей. Доля контрольных точек среди всех точек, обработанных волонтером, в среднем составляет 49%

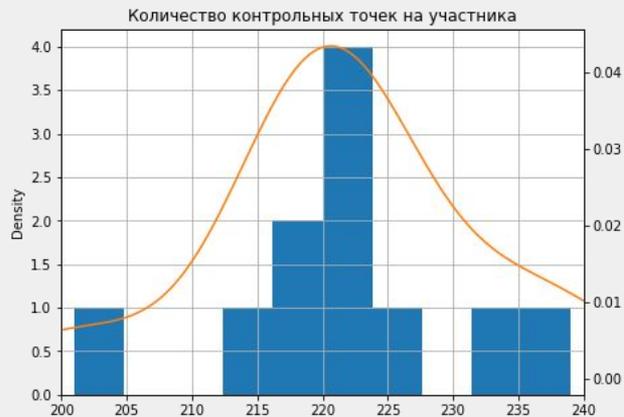
Среднее количество точек на участника	454
Среднее кол-во контрольных точек на участника	221



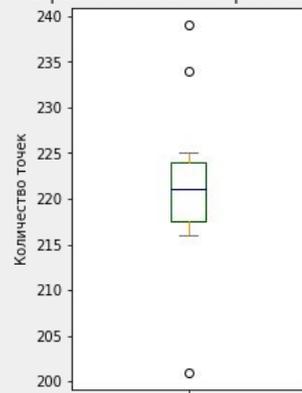
Пример распределения точек для одного участника

Распределение точек между участниками

Нормальное распределение контрольных точек по участникам эксперимента свидетельствует о равномерном и репрезентативном покрытии “контролями” и гарантирует сравнимость результатов оценки точности участников

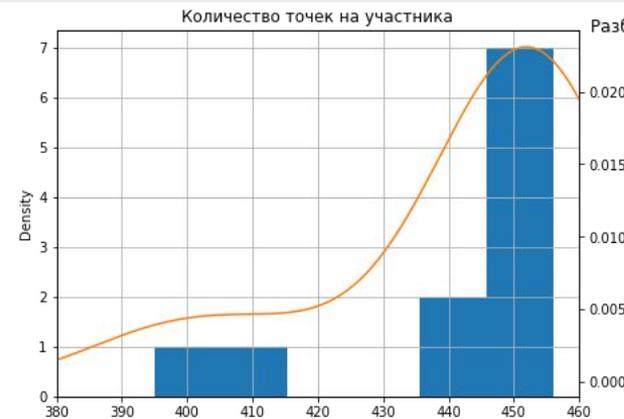


Разброс количества контрольных точек

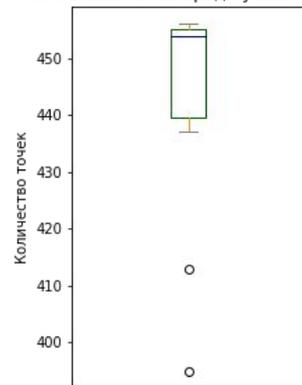


Случаи, когда доля контрольных точек среди всех точек участника составила менее 50% - исключение.

Значительное, возможно даже избыточное, количество контролей позволяет достоверно определить точность участника и оценить минимально необходимую долю контролей для достоверных оценок точности.



Разброс количества всех точек среди участников эксперимента

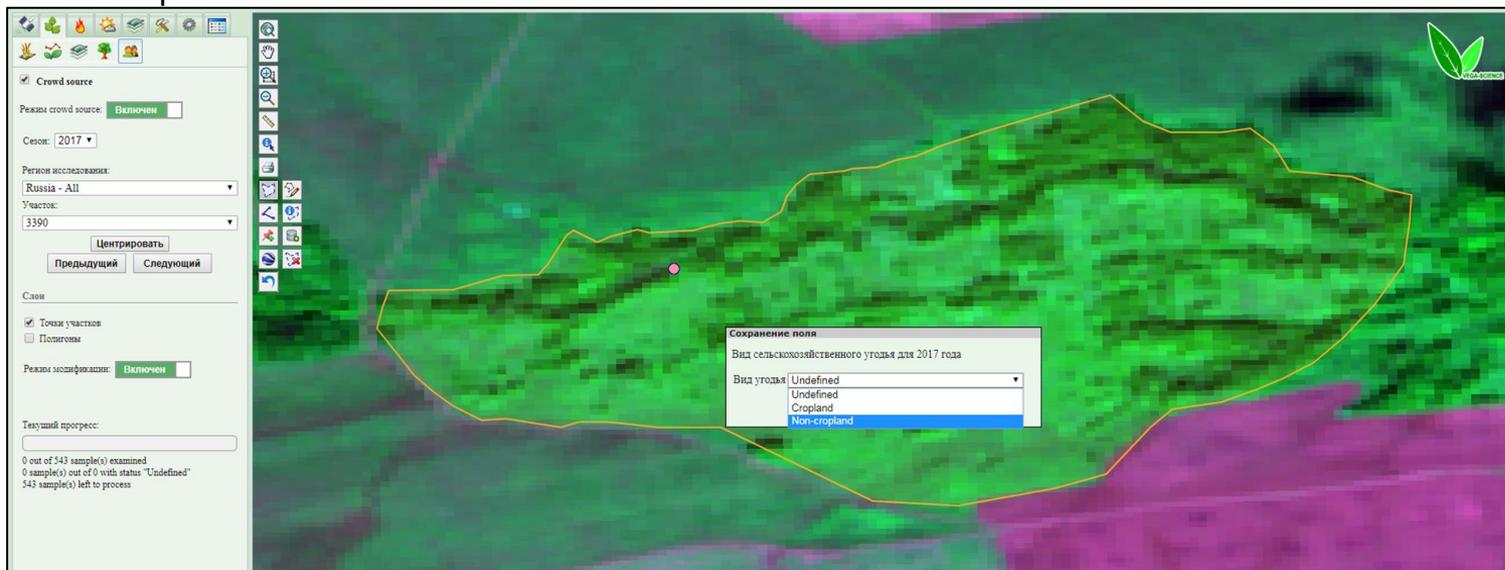


Каждый участник обработал не менее 395 точек. В среднем на волонтера пришлось 454 точек/объектов (+/- 20) для классификации.

Диаграмма размаха по количеству точек показывает, что лишь у двух участников число обработанных точек выходило за средние значения.

Организация эксперимента

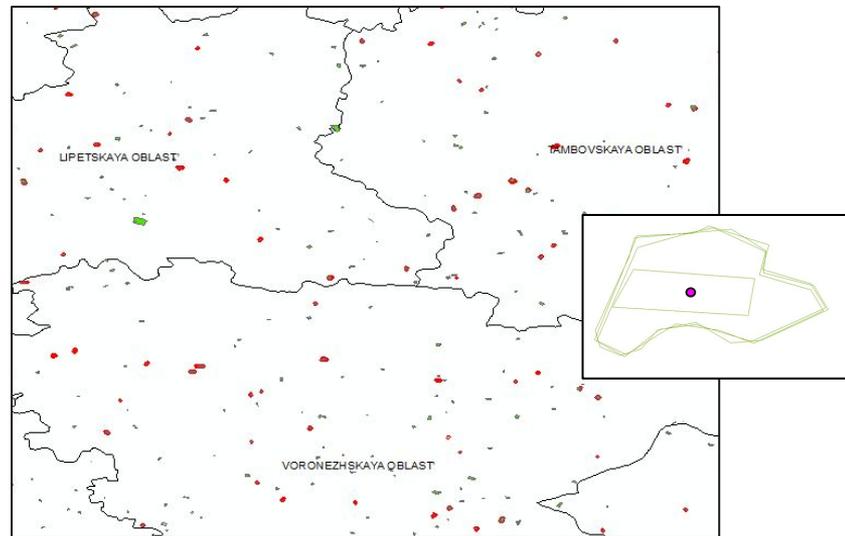
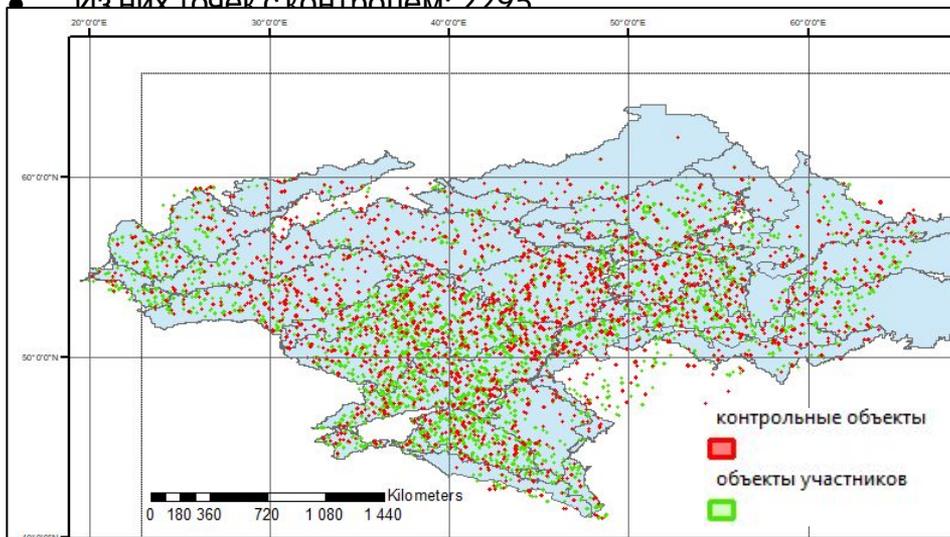
- Был разработан и использован специальный инструментарий для ручного аннотирования объектов на базе Vega-science (sci-vega.ru).
- В интерфейсе Веги участники выбирали из выпадающего списка точку для обработки, затем подбирали спутниковые изображения на исследуемый период (2017 год), анализировали их, определяя признаки “пашни” / “не пашни”, рисовали вокруг точки однородный объект и сохраняли его вместе с присвоенным классом.



Пример выделения однородного объекта вокруг точки и присвоения ему класса (undefined/cropland/non-cropland) в системе Vega-science

Результаты

- В результате эксперимента был получен набор полигонов с присвоенным классом (пашня/не-пашня) по всем обработанным участником точкам (рис.1). Обработанная точка - точка, вокруг которой нарисован объект и объекту присвоен класс.
- Была проведена оценка точности каждого волонтера и даны средние оценки точности. Для оценки точности использовался контрольный набор, созданный экспертами.
- Были проанализированы распределения точности по волонтерам, а также распределения точности по агрозонам.
- Всего точек, обработанных участниками: 4604
- Из них точек с контролем: 2295



Полигоны - объекты вокруг точек, нарисованные и классифицированные участниками. Красным обведены контрольные объекты (для которых есть экспертная аннотация)

Метрики точности

Для оценки точности были выбраны следующие метрики:

- F1-score для каждого класса(пашня и не-пашня)
$$F_1 = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$
$$= \frac{2 \text{tp}}{2 \text{tp} + \text{fp} + \text{fn}}$$
- Overall Accuracy: $\text{Overall Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{P} + \text{N})}$
- Метрики были рассчитаны по матрицам ошибок (confusion matrix) для каждого волонтера / агрозоны

	$y = 1$	$y = 0$
$\hat{y} = 1$	True Positive (TP)	False Positive (FP)
$\hat{y} = 0$	False Negative (FN)	True Negative (TN)

\hat{y} — ответ волонтера

y — истинная метка
(класс)

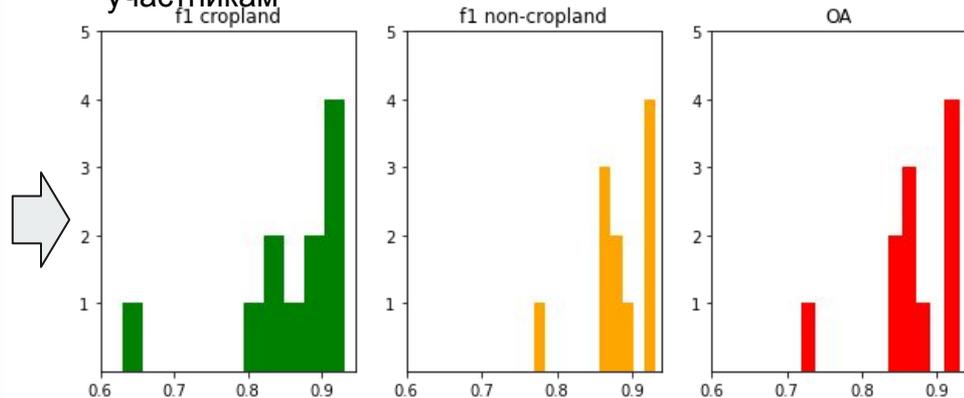
Результаты - оценка точности волонтеров

Была проведена оценка точности распознавания пашни для каждого участника эксперимента. В таблице:

- id_person - идентификатор волонтера
- OA - общая точность
- f1_crp - f1-score для класса пашня
- f1_noncrp - f1-score для класса не-пашня
- total points, control points - количество точек всего и из них контрольных

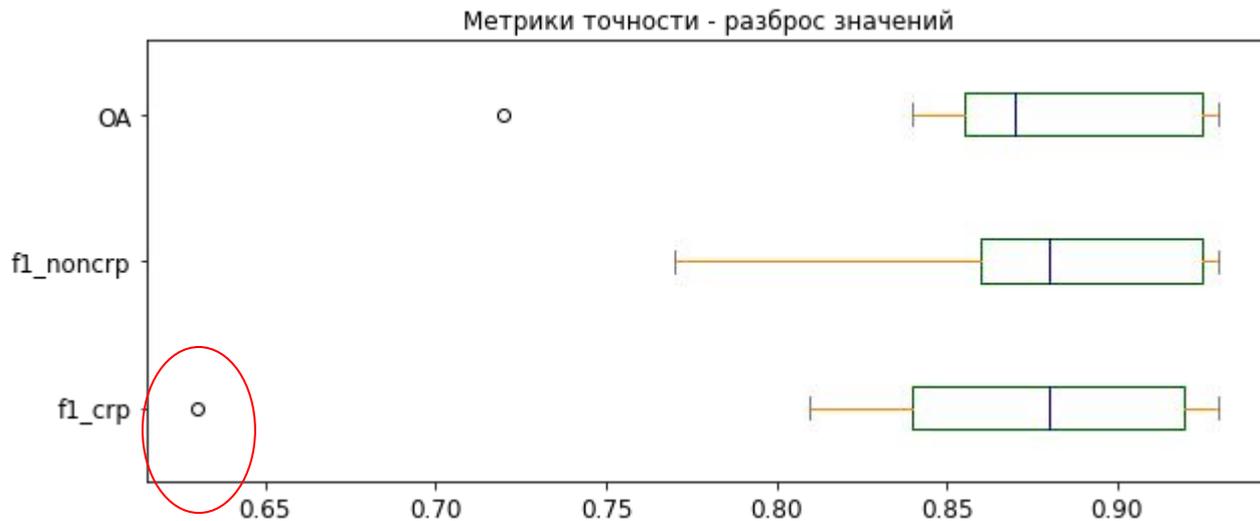
id_person	f1_crp	f1_noncrp	OA	total_points	control_points
14	0.63	0.77	0.72	456	239
16	0.81	0.86	0.84	395	201
1	0.84	0.86	0.85	456	225
12	0.84	0.88	0.86	449	221
6	0.86	0.88	0.87	413	218
17	0.88	0.86	0.87	454	221
0	0.89	0.89	0.89	454	223
10	0.91	0.92	0.92	454	216
3	0.93	0.93	0.93	442	234
4	0.93	0.93	0.93	437	222
13	0.93	0.93	0.93	456	217

Распределение метрик точности в среднем по всем участникам



Точность распознавания не-пашни в среднем выше точности распознавания пашни. Общая точность не опускается ниже 0.65.

Результаты - оценка точности волонтеров



Наблюдается выброс в виде “неточного” волонтера, метрика f1-score которого по классу пашня оказалась значительно ниже средней (0.63 против 0.88), что вносит вклад в оценку общей точности.

Результаты - оценка точности волонтеров

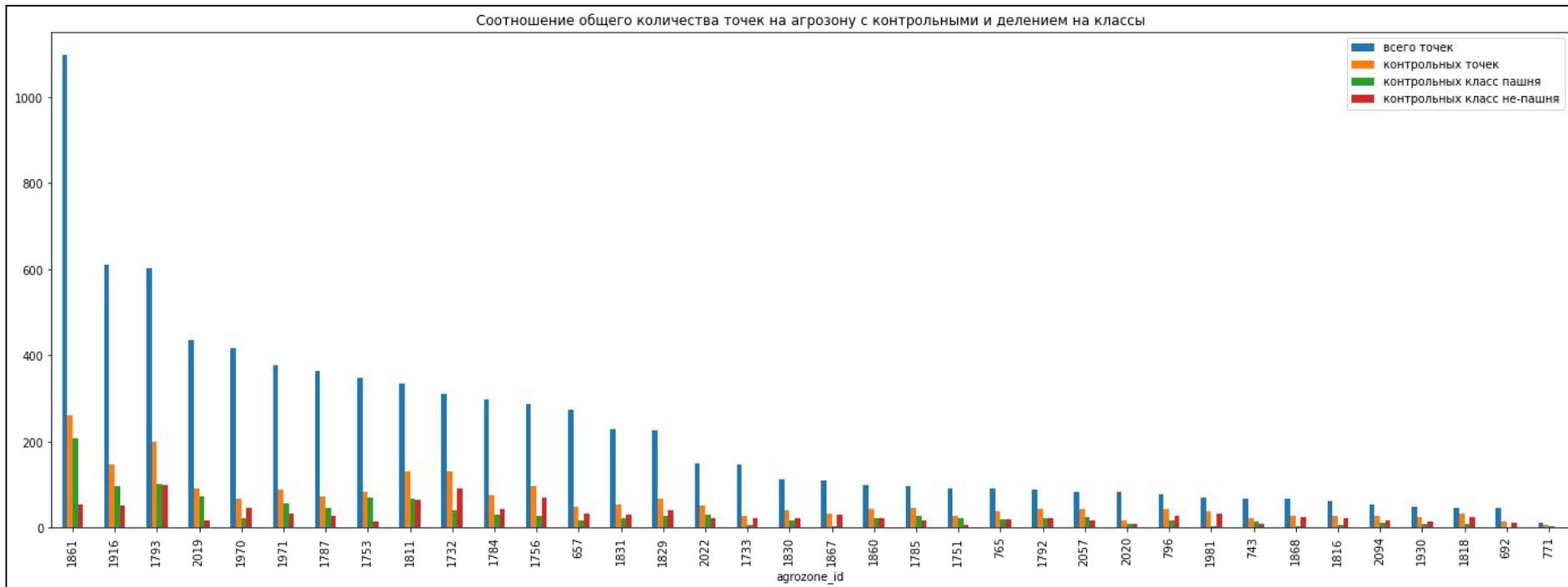
- Средняя общая точность классификации участниками краудсорс-эксперимента составила **0.87** со стандартным отклонением 0.06
- Точность распознавания “пашни” - 0.86
- Точность распознавания “не-пашни” - 0.88

	ПАШНЯ	НЕ ПАШНЯ
f1-score	0.86	0.88
std	0.09	0.05

Выводы:

- В среднем участники эксперимента хорошо справились с поставленной задачей : по сравнению с константной точностью (0.5) показатели точности краудсорс-дешифровщиков достаточно высоки, отклонения в точностях среди волонтеров не высок. Полученные данные свидетельствуют о применимости подхода создания опорных выборок методом краудсорсинга.
- Распознавать не-пашню (не используемые в исследуемом сезоне под сельское хозяйство земли) участникам оказалось проще, чем искать следы использования земли. Об этом говорит и более высокое стандартное отклонение точности класса “пашня” - точность менялась от участника к участнику сильнее, чем по классу не-пашня.

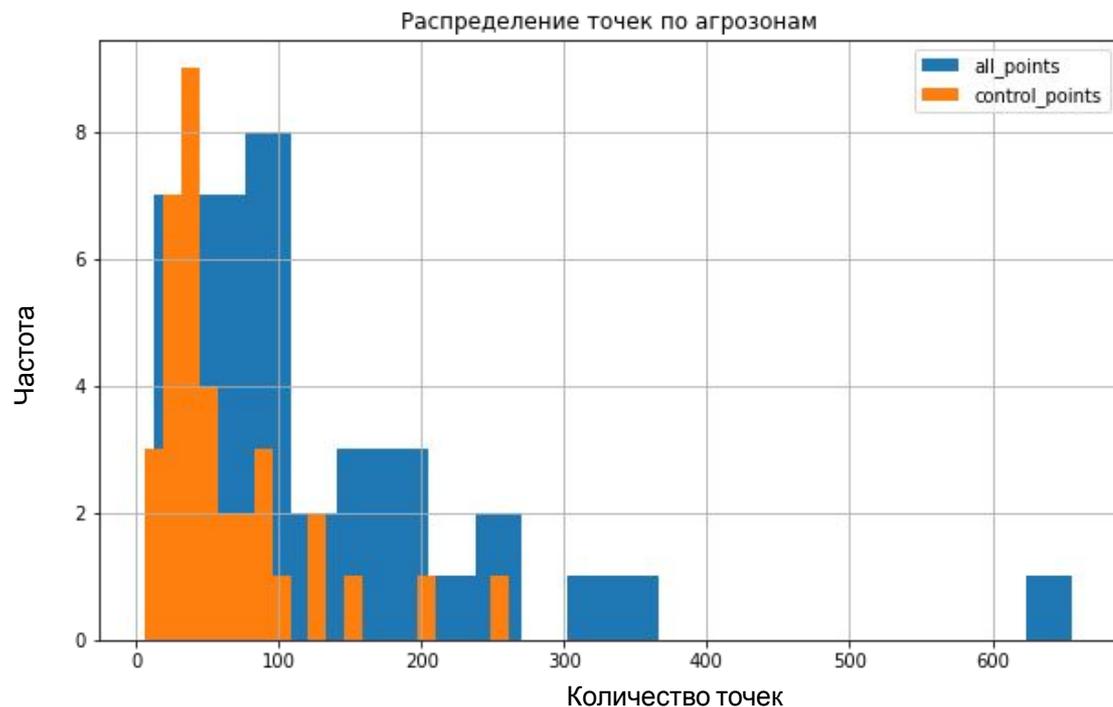
Результаты - оценка по агрозонам



Количество агрозон в исследовании: 36 (изначально было 44, убрали страты с количеством точек менее десяти)

Результаты - оценка по агрозонам

- Среднее количество точек на агрозону 86, стандартное отклонение 123
- Максимум - 656, минимум - 12
- Среднее количество контрольных точек на агрозону 44, стандартное отклонение 53
- Соотношение контрольных точек - почти 50%



Результаты - оценка точности по агрозонам

- Средняя общая точность классификации внутри страты составила **0.88** со стандартным отклонением 0.09
- Точность распознавания “пашни” - 0.85
- Точность распознавания “не-пашни” - 0.88

	ПАШНЯ	НЕ ПАШНЯ
f1-score	0.86	0.87
std	0.22	0.1

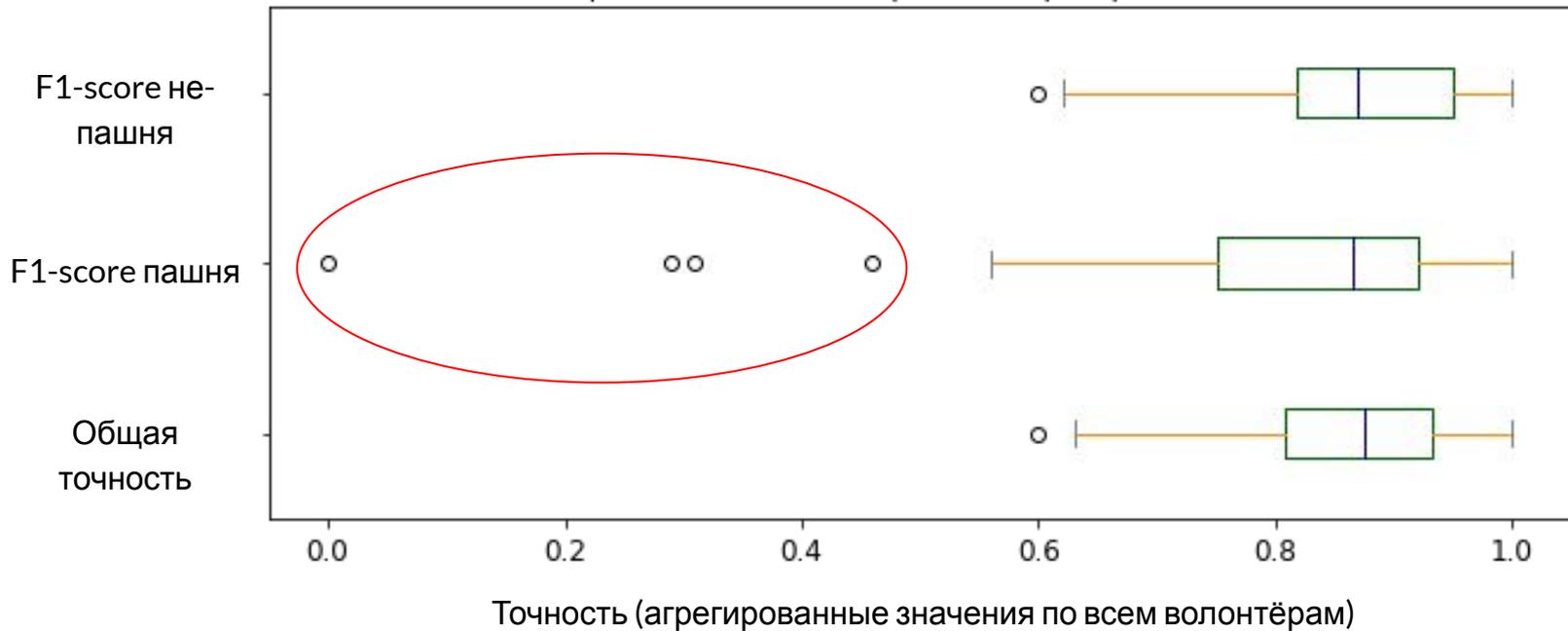
Выводы:

- Средняя точность по стратам отличается более высоким стандартным отклонением по сравнению с общей точностью
- Точность распознавания пашни в среднем по агрозоне ожидаемо (после анализа точностей в общем для всех страт вместе) ниже точности распознавания не-пашни
- Примечательно высокое стандартное отклонение (0.22) для f1-score по классу пашня - это говорит о сильных различиях в точностях от агрозоны к агрозоне; необходимо установить пространственные различия.

Результаты - оценка точности по агрозонам

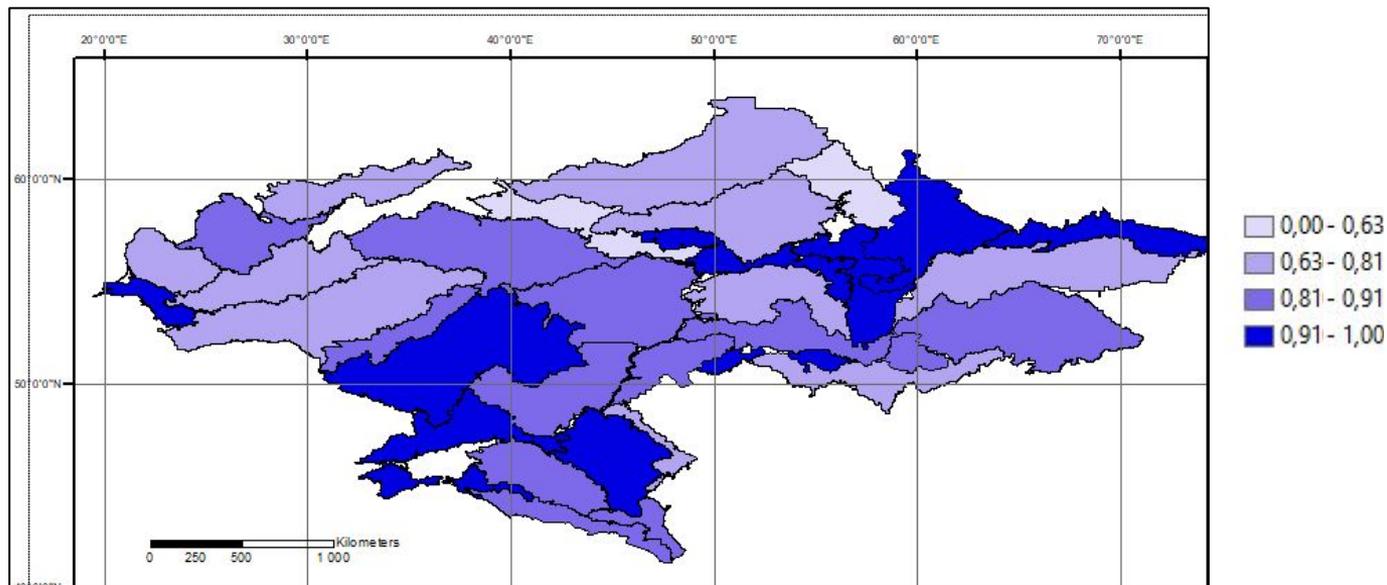
Выделяется несколько агрозон, точность распознавания на которых ниже, чем средняя (ниже 0.8). Видимо, эти агрозоны являются сложными для фотоинтерпретации. Определение данных “сложных” зон позволит грамотно распределить усилия экспертов и волонтеров: следует направлять силы экспертов туда, где не могут справиться волонтеры.

Метрики точности по агрозонам - разброс значений



Результаты - оценка точности по агрозонам

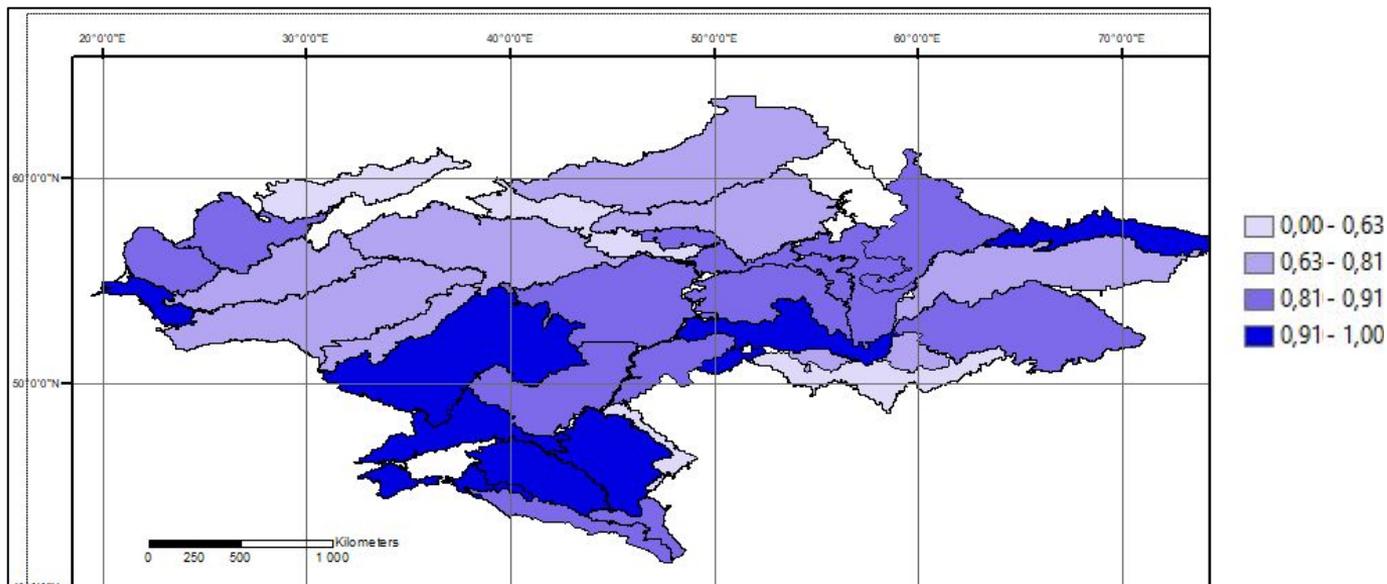
Построены картограммы по значениям метрик точности для оценки пространственных тенденций ошибок. Общая точность классификации волонтерами возрастает от севера страны к югу, что объяснимо увеличением размеров полей и их выраженной геометрией. Высокие точности на востоке (Башкортостан, Свердловская область) объясняются успешным распознаванием не пахотных земель.



Общая точность

Результаты - Пашня точность

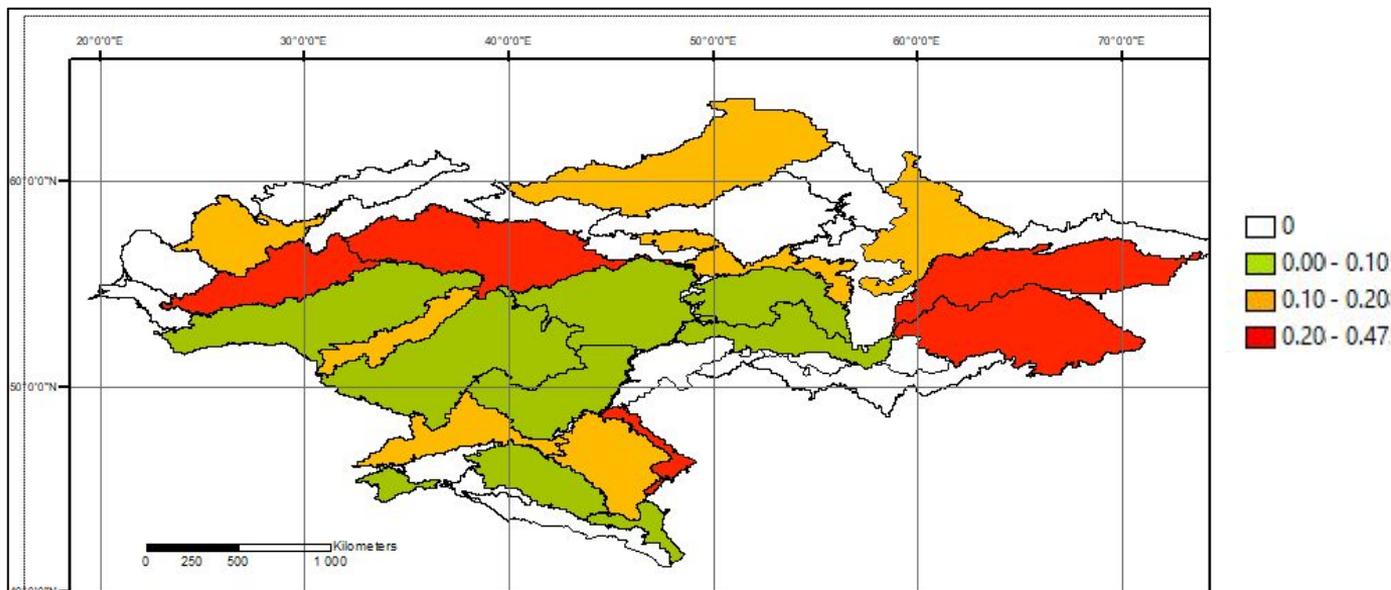
Выделяются агрозоны, в которых участникам проще распознавать пашню: таким образом подтверждается возможность определения территорий с высокой точностью большинства участников-волонтёров; участие экспертов при создании опорных выборок в этих зонах может быть сведено к минимуму.



F1-score пашня

Результаты - Пашня дисперсия

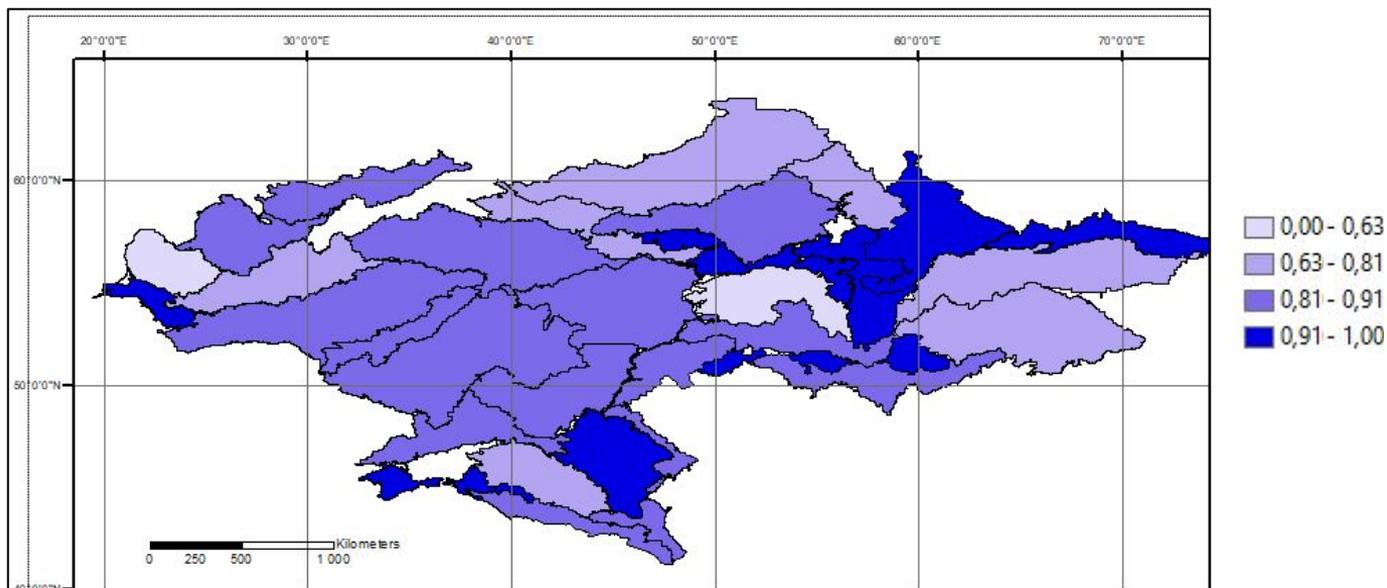
Была рассчитана дисперсия значений метрик для каждой агрозоны, которая свидетельствует о степени разброса значений точности внутри агрозоны и может быть интерпретирована как степень согласованности, надежности результатов. Агрозоны, где дисперсия высока, имеют низкую степень надежности рассчитанных точностей, и наоборот.



Стандартное отклонение от среднего для F1-score по классу пашня

Результаты - Не-пашня точность

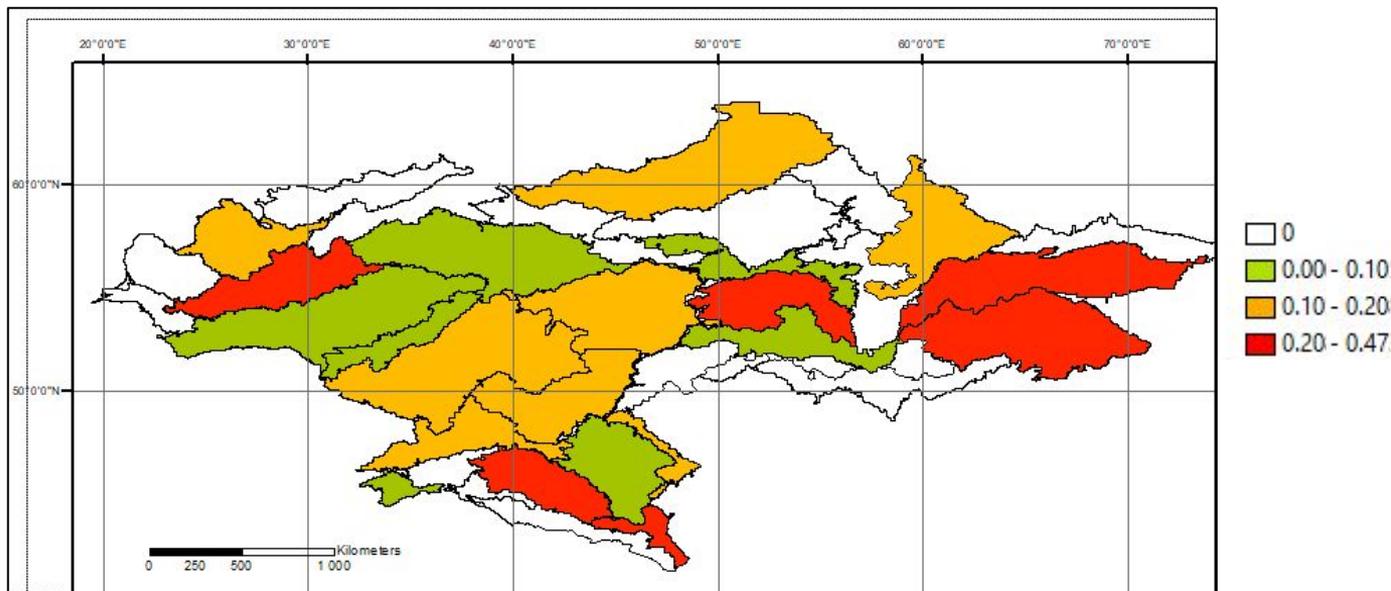
Эксперимент показал, что с распознаванием не-пашни на высоком уровне справляется большинство участников в большинстве агрозон.



F1-score не-пашня

Результаты - Не-пашня дисперсия

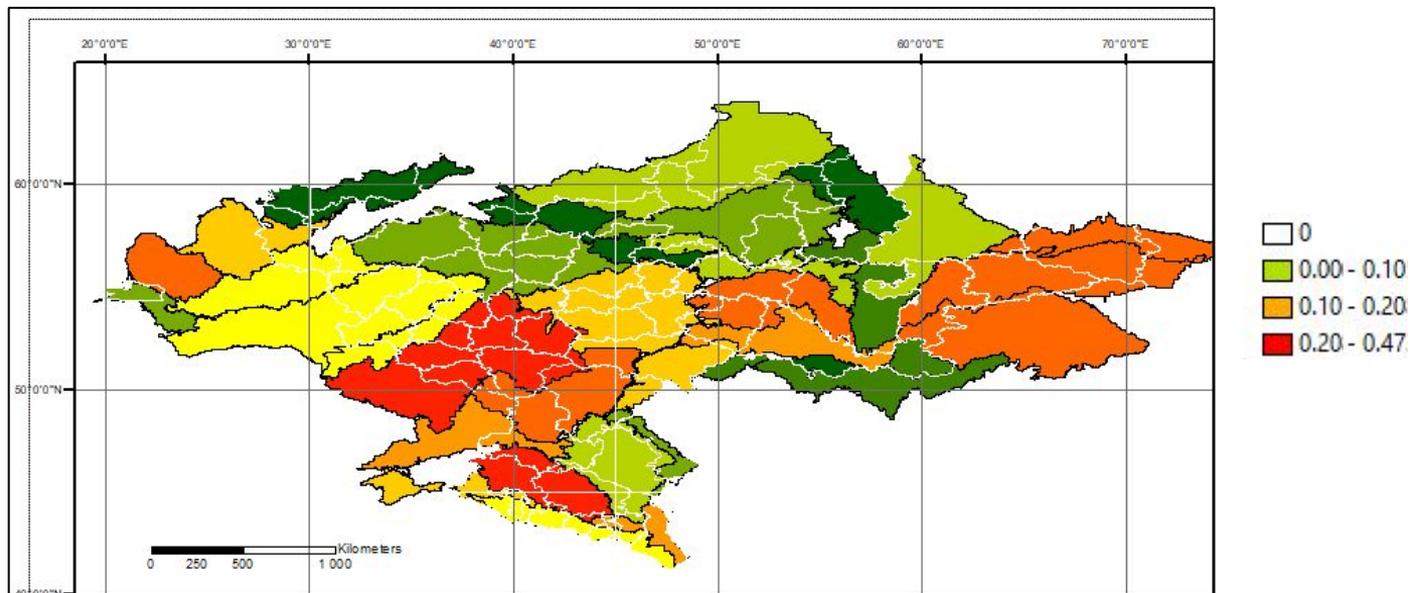
Картограмма стандартного отклонения для класса не-пашня. Отсутствие значений следует интерпретировать как недостаточность информации (малое количество точек рассматриваемого класса на агрозону)



Стандартное отклонение от среднего для F1-score по классу не-пашня

Результаты - баланс классов в регионах

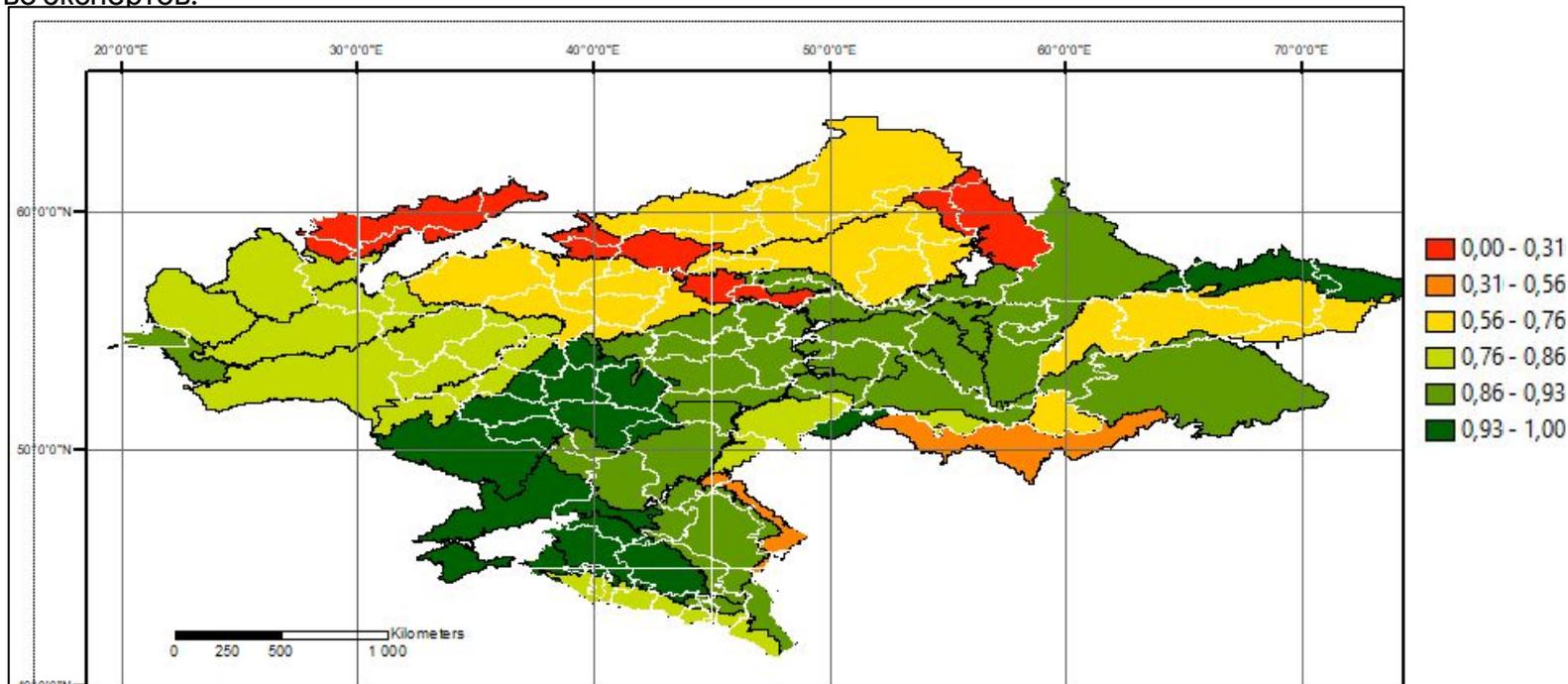
Данные по количеству объектов с меткой “пашня” и “не-пашня” позволяют оценить баланс классов в регионах. Картограмма по доли пашни в агрозоне (с наложением административных границ) может быть использована для сопоставления с картами пахотных земель из других источников.



Доля пашни в агрозоне по результатам краудсорс-интерпретации

Результаты - оценка точности по агрозонам

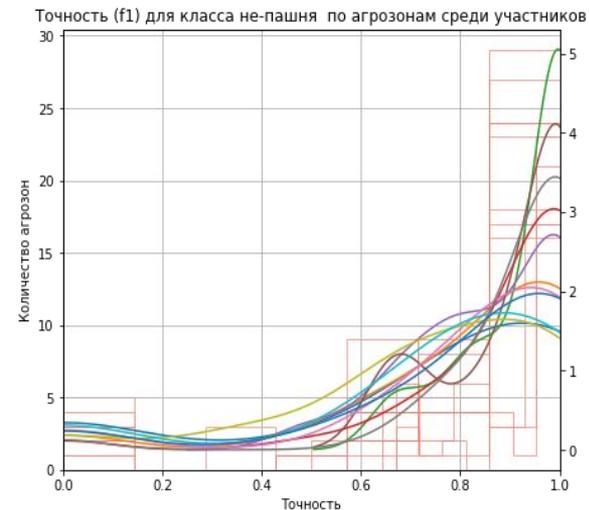
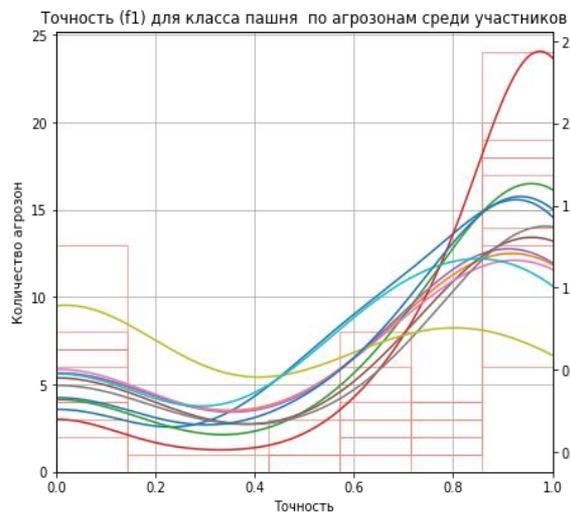
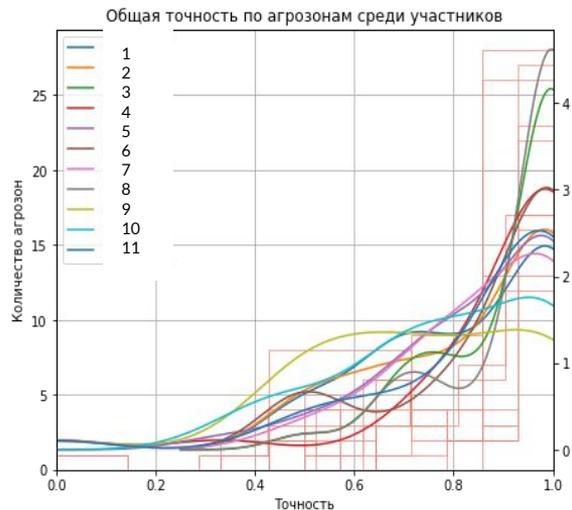
Одной из важных задач эксперимента являлась оценка возможности определения территорий с высокой точностью большинства участников и территорий, сложных для интерпретации, в которых требуется привлечение экспертов. Картограмма точности по классу “пашня” может служить прокси для разделения “лёгких” и “сложных” зон для волонтеров. Также можно выделить переходные зоны, где, например, можно привлекать только самых точных участников и небольшое количество экспертов.



F1-score пашня

Результаты - распределение точности по агрономам

Для оценки стабильности результата волонтёра было рассчитано распределение точности по агрономам в среднем для волонтёра (по трём метрикам точности), а затем гистограммы всех участников совместили, что позволило выявить людей, у которых показатели точности от агрононе к агрононе ведут себя нетипично:



Гистограммы и огибающие их кривые, отражающие распределение точности по агрономам для каждого участника эксперимента

Выводы

Эксперимент показал применимость подхода к созданию опорных выборок с помощью краудсорсинга с точки зрения:

- а) реализуемости подхода на базе инструментария системы Vega-science;
- б) возможности объективной оценки точности итогового набора данных;
- в) возможности получения репрезентативной выборки для бинарной классификации на большую территорию;
- г) возможности оценки пространственных тенденций точности и выявления зон, в которых можно минимизировать участие экспертов.

Итоги:

- Общая точность классификации участниками краудсорс-эксперимента составила 0.87 со стандартным отклонением 0.06;
- Участники эксперимента хорошо справились с поставленной задачей: по сравнению с константной точностью (0.5) показатели краудсорс-дешифровщиков достаточно высоки, отклонения в точностях невысоки;
- Точность распознавания не-пашни в среднем выше точности распознавания пашни;
- Волонтеров с высокой точностью можно привлекать к работам на территориях с малым количеством экспертных данных;
- Точность распознавания пашни меняется от страты к страте - наблюдаются пространственные тенденции точности на больших территориях (на севере, где мало пахотных земель, их распознавание затруднено и требуется привлечение экспертов);
- Предложенные подходы к организации создания выборки с привлечением волонтеров с последующей оценкой точности можно использовать для получения опорных выборок на большие территории; на основе полученных предварительных оценок можно планировать более масштабные работы.



Спасибо за внимание!

elkina@d902.iki.rssi.ru