



Agriculture and
Agri-Food Canada



Ministerio de Agroindustria
Presidencia de la Nación



Сравнение разных источников обучающей информации для регионального картографирования пашни в различных регионах мира

Плотников Д.Е.¹, de Abellegra D.², Veron S.², Zhang M.³, Толпин В.А.¹, Ёлкина
Е.С.¹, Самофал Е.В.¹, Барталев С.А.¹, Lavreniuk M.⁴, Waldner F.⁵, Ziad A.⁶

¹ Terrestrial Ecosystems Monitoring Laboratory, Space Research Institute of Russian Academy of Sciences (IKI), Moscow, Russia

² Instituto de Clima y Agua, Instituto Nacional de Tecnología Agropecuaria (INTA), Buenos Aires, Argentina

³ Institute of Remote Sensing and Digital Earth, Chinese Academy of Science, Beijing, China

⁴ Department of Space Information Technologies, Space Research Institute NAS and SSA (SRI), Kyiv, Ukraine

⁵ Earth and Life Institute - Environment, Croix du Sud, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

⁶ Agriculture and Agri-Food Canada, Ottawa, Canada

Цели и задачи эксперимента

- ▶ Разработка методик для создания и сравнения наборов обучающих данных (CS, LC, IS) для картографирования пашни в 6 различных регионах мира (Аргентина, Бельгия, Канада, Китай, Россия, Украина)
 - ▶ Оценка и сравнение наборов напрямую и через карты, созданные на их основе
 - ▶ Исследование и обработка данных краудсорсинга
 - ▶ Оценка эффективности полученных из восстановленных изображений Landsat спектрально-динамических метрик для картографирования пашни
 - ▶ Оценка возможностей локально-адаптивного классификатора в сравнении с классическим для картографирования пашни на большой территории
-

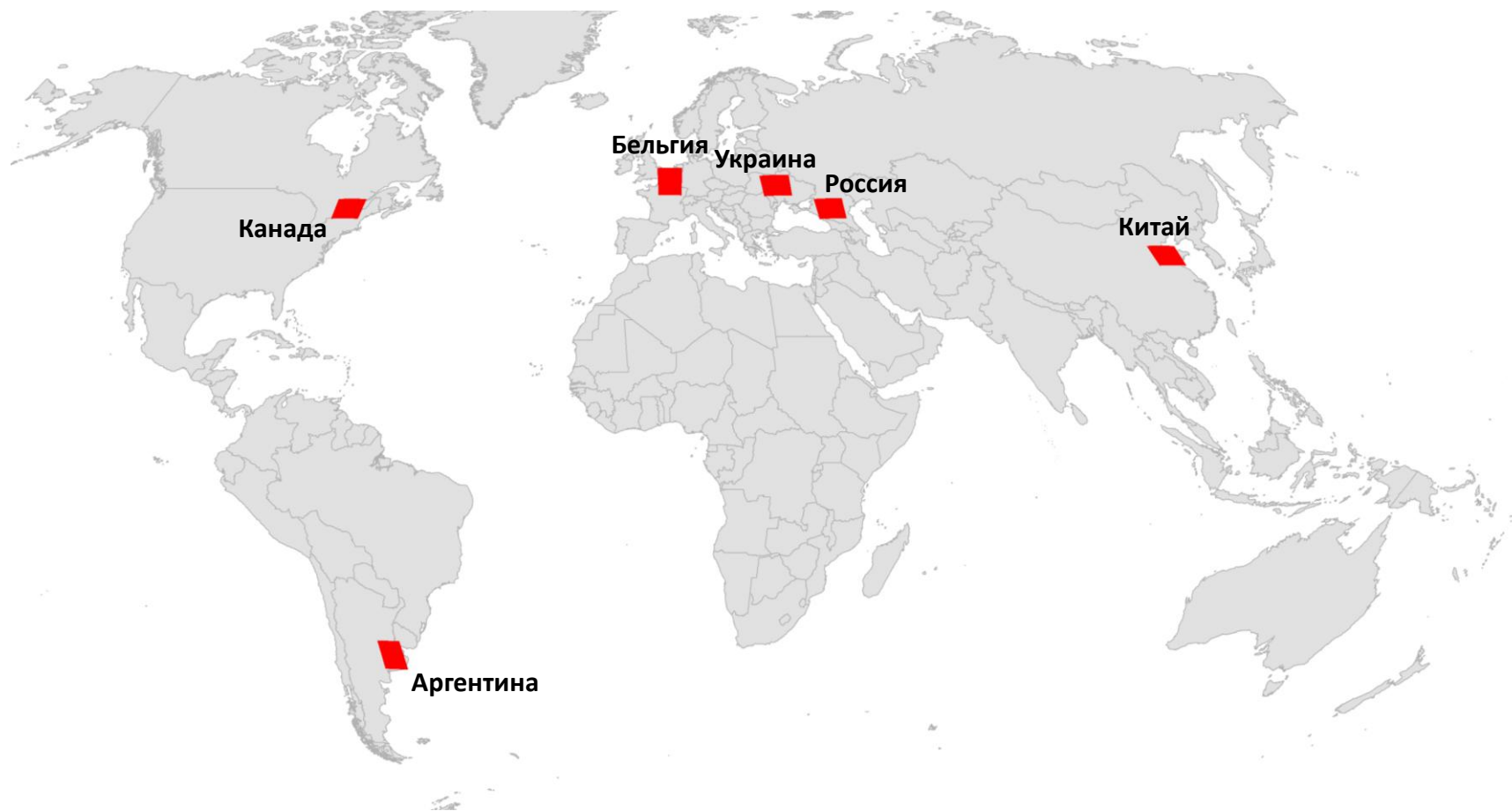
Условия эксперимента

- ▶ Единый сезон вегетации: 2016 год в северном полушарии и 2015-2016 годы в южном (Аргентина)
 - ▶ Одинаковые спутниковые данные:
 - ▶ Восстановленный Landsat-8 для распознавания пашни
 - ▶ Sentinel-2, Landsat-7,8 для визуальной фотоинтерпретации (краудсорсинга)
 - ▶ Использование единого определения JESAM для пахотных земель
 - ▶ *«The annual cropland from a remote sensing perspective is a piece of land of minimum 0.25-ha (minimum width of 30-m) that is sowed/planted and harvestable at least once within the 12 months after the sowing/planting date. The annual cropland produces an herbaceous cover and is sometimes combined with some tree or woody vegetation»*
 - ▶ Территории фиксированных размеров: малая (~50 км) и большая (300 на 500 км)
-

Исследуемые источники обучающих данных

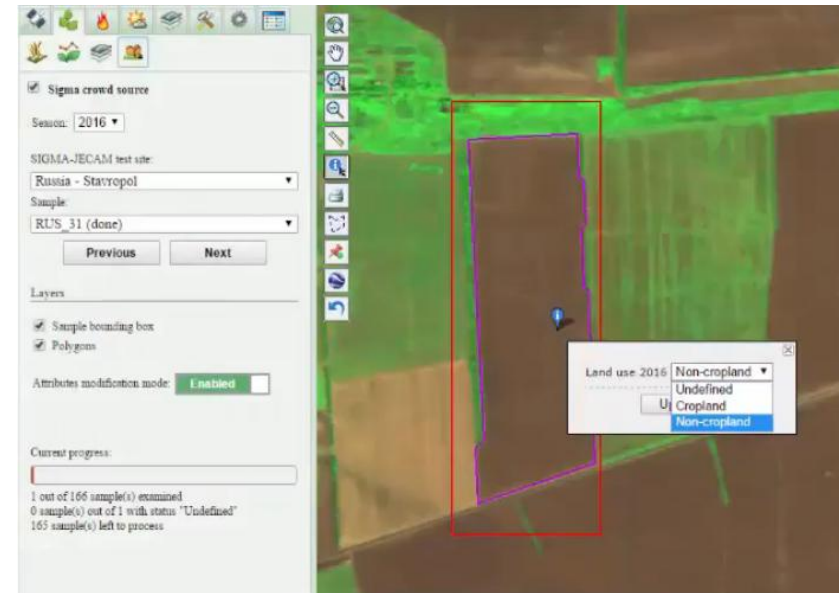
- ▶ In situ данные (IS): получены путем посещения, объезда полей, а также на основе сообщений с мест; класс объекту присваивался в соответствии с определением JESAM
 - ▶ Crowdsourcing (CS): набор создан путём визуальной фотоинтерпретации спутниковых снимков высокого разрешения (10-30 м) за весь сезон вегетации в интерфейсе VEGA-GEOGLAM относительно большим количеством независимых пользователей (экспертов), которые также следовали определению JESAM
 - ▶ Land Cover (LC): классы извлекались из карты GlobeLand30 (Chen et al. 2014)
-

Расположение и размеры участков JESAM



Первый этап CS на малой территории (оценка экспертов)

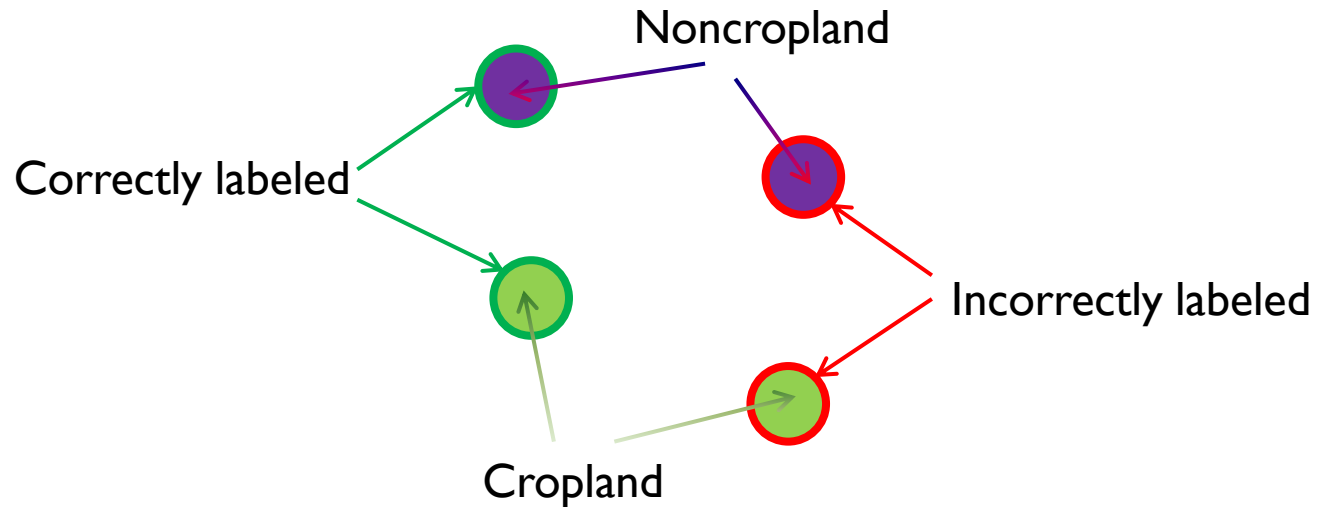
- ▶ Инструмент для краудсорсинга в VEGA-GEOGLAM обеспечивал визуализацию спутниковых данных и навигацию между предварительно заданными объектами



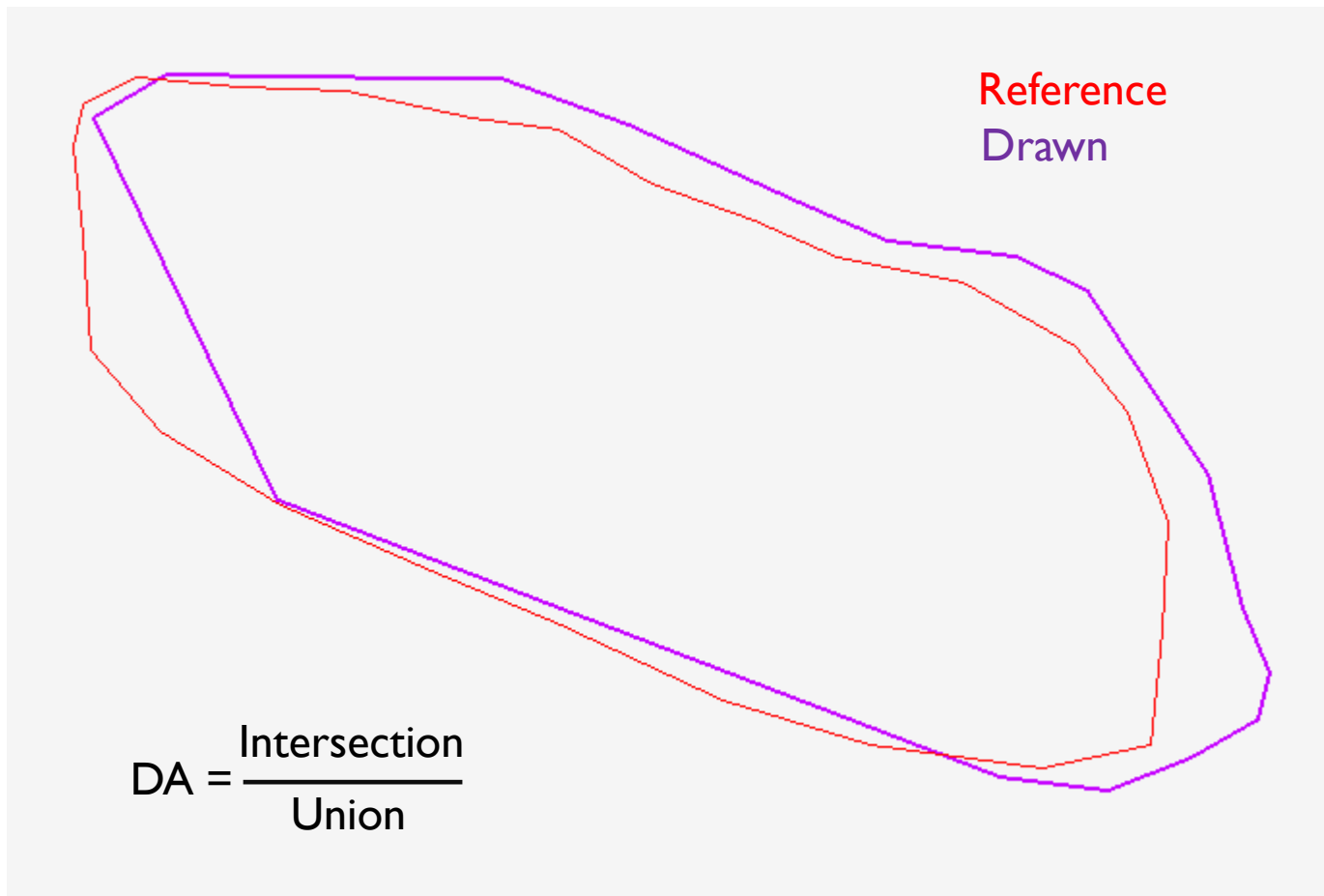
- ▶ Пользователи были проинструктированы как пользоваться интерфейсом с помощью обучающего видео
- ▶ Задачей было рисование границ однородного объекта внутри заданной прямоугольной области
- ▶ Пользователи руководствовались определением JECAM при классификации

Метрики оценки точности: Labeling Accuracy

$$LA = \frac{\text{Number of correctly labeled samples}}{\text{Number of all samples}}$$

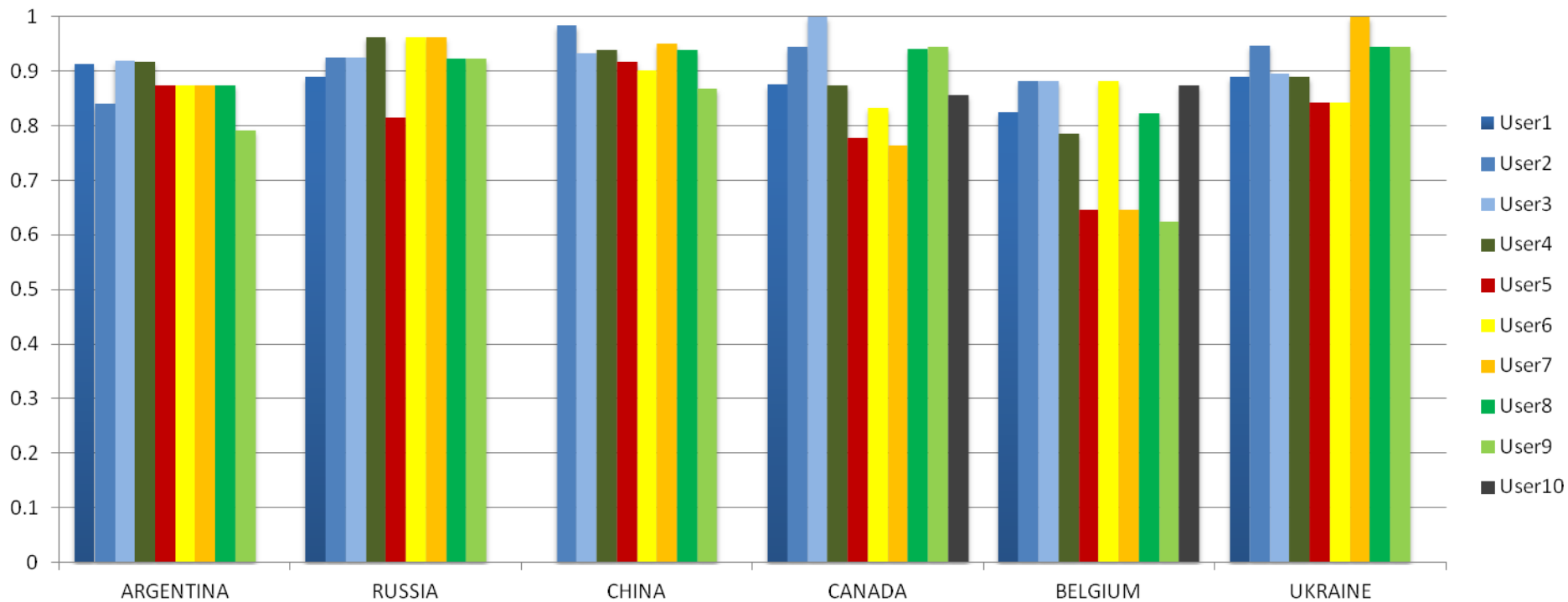


Метрики оценки точности: Drawing Accuracy



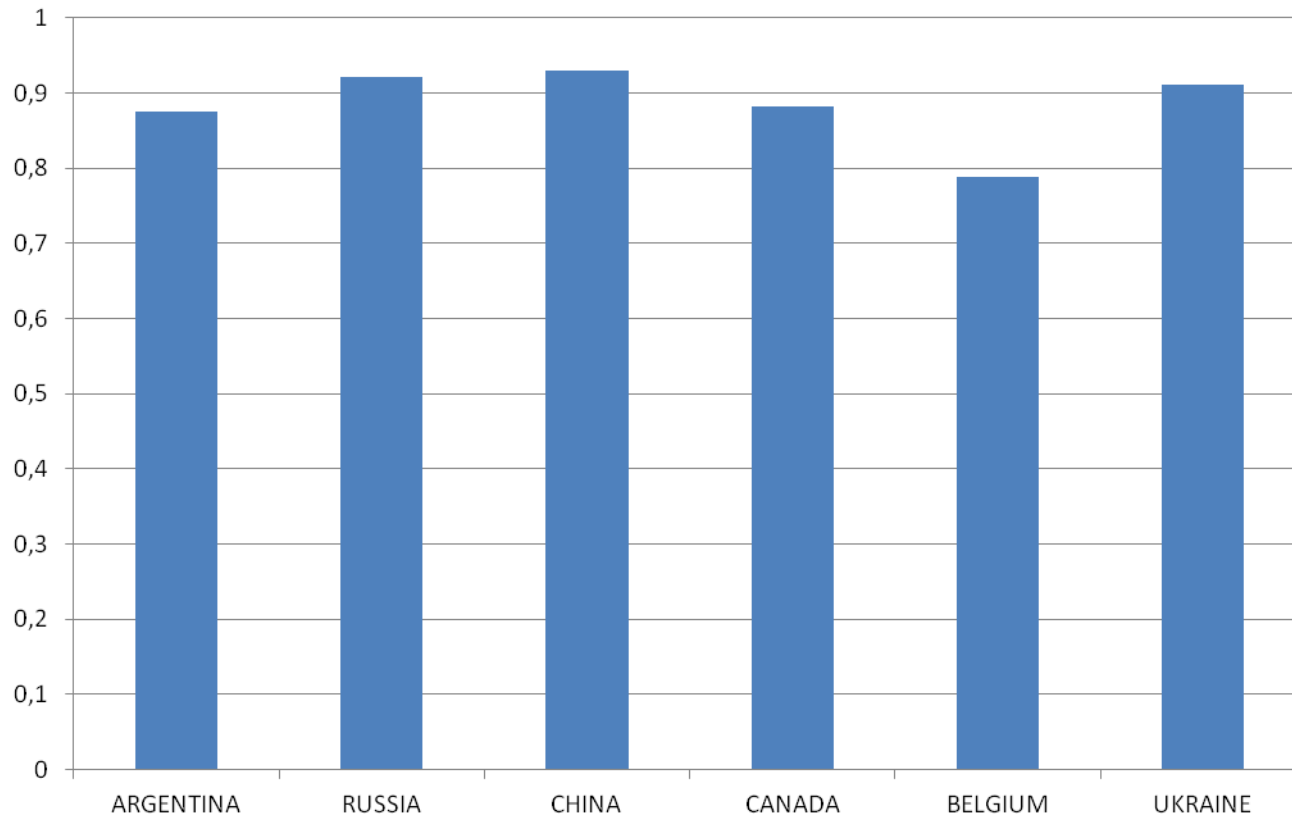
Результаты первого этапа

Labeling Accuracy



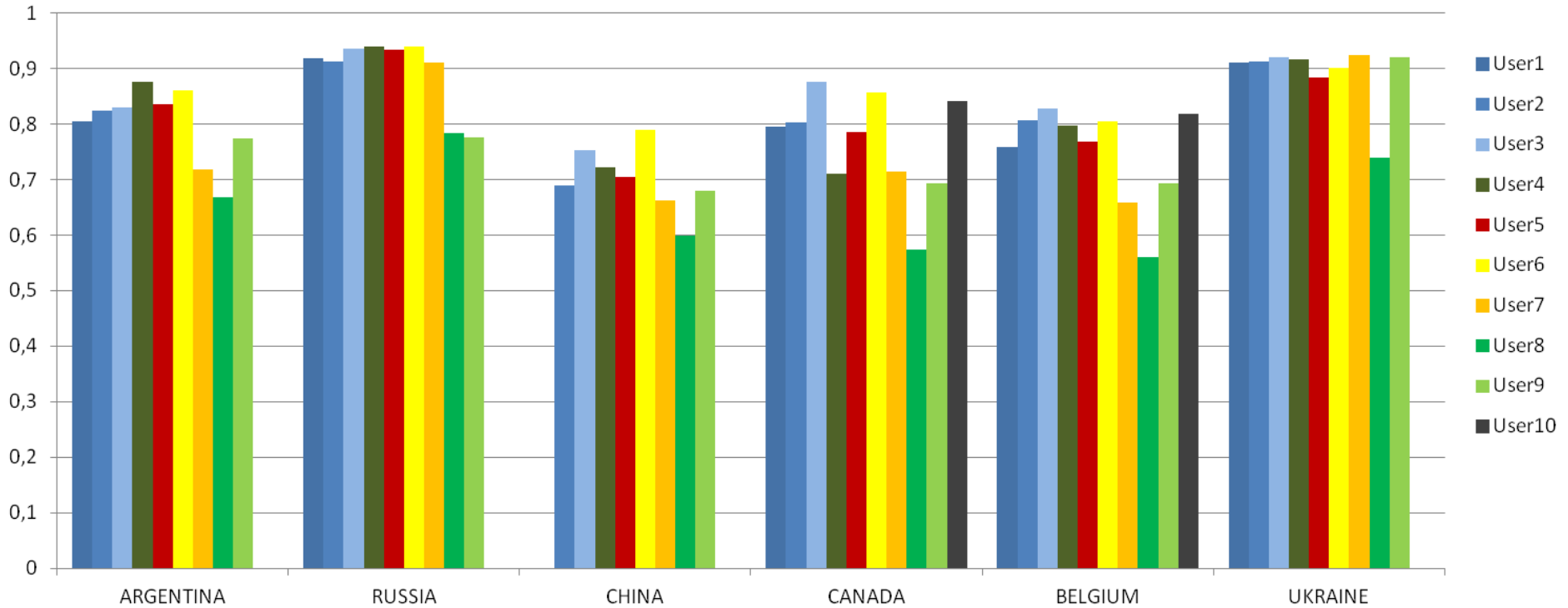
Результаты первого этапа

Mean Labeling Accuracy



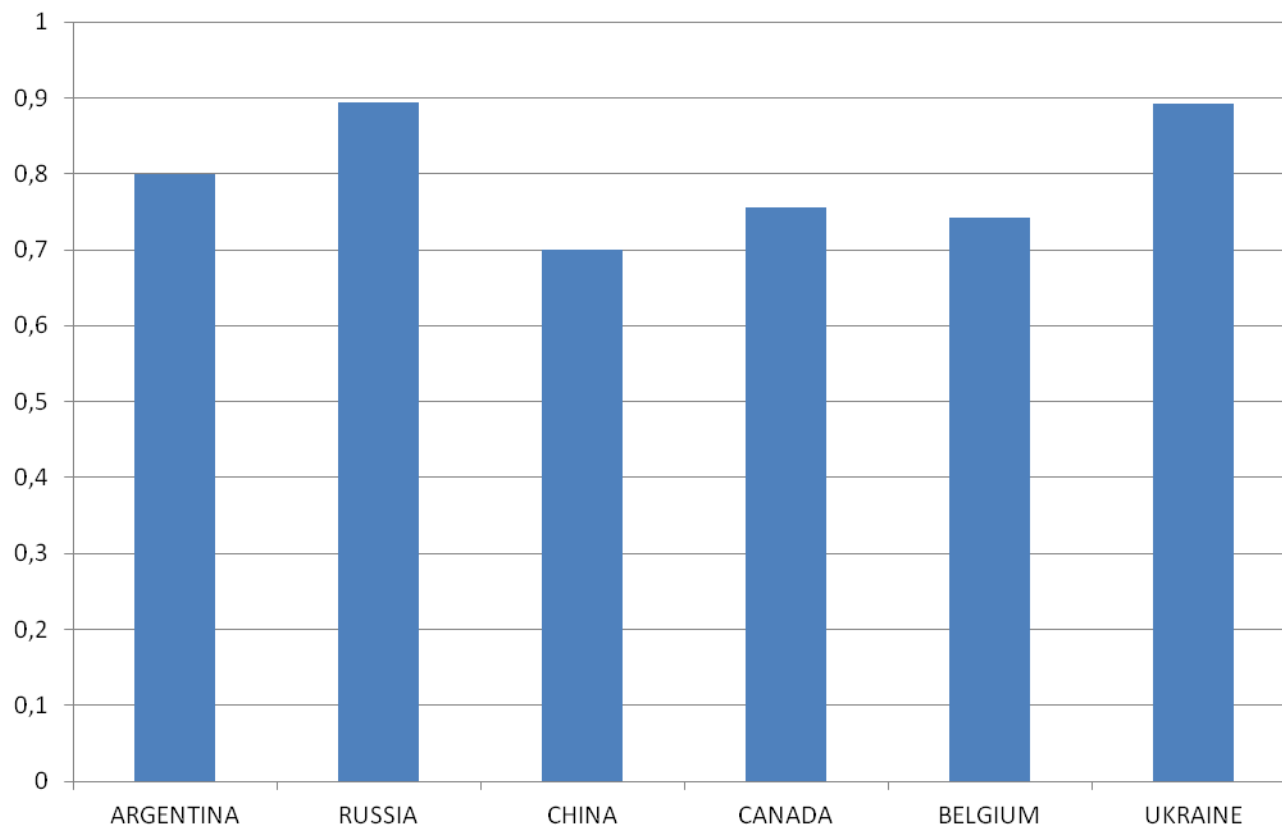
Результаты первого этапа

Drawing Accuracy



Результаты первого этапа

Mean Drawing Accuracy



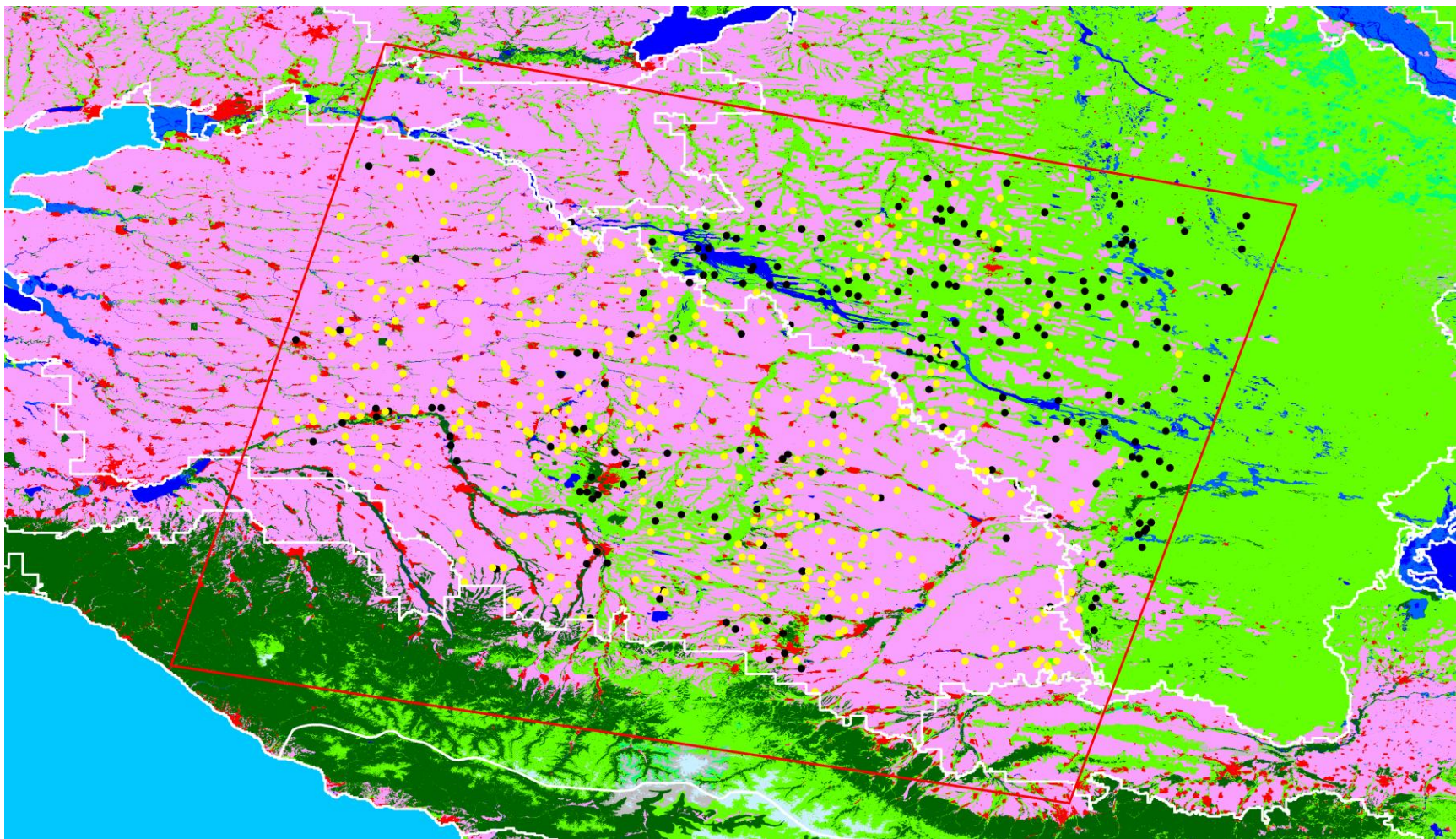
Второй этап: создание опорных точек для CS на большой территории

- ▶ Использование сезонной временной серии снимков Landsat-8 на каждый участок (6 тайлов)
 - ▶ Сегментация временной серии снимков (AAFC, Канада)
 - ▶ Использование центроидов сегментов в качестве кандидатов опорных точек
 - ▶ Использование зон GAES в качестве страт
 - ▶ Для каждой страты внутри участка JESAM рассчитывается соотношение между точками разных классов с использованием GlobeLand30
 - ▶ Точки расставляются случайным стратифицированным образом (Random Stratified Sampling), однако точки одного класса ставятся на расстоянии не менее 2 км друг от друга
-

Сегментация Landsat-8, GlobeLand30 и результат сэмплинга для обоих классов



Пример расположения сэмплов на большой территории



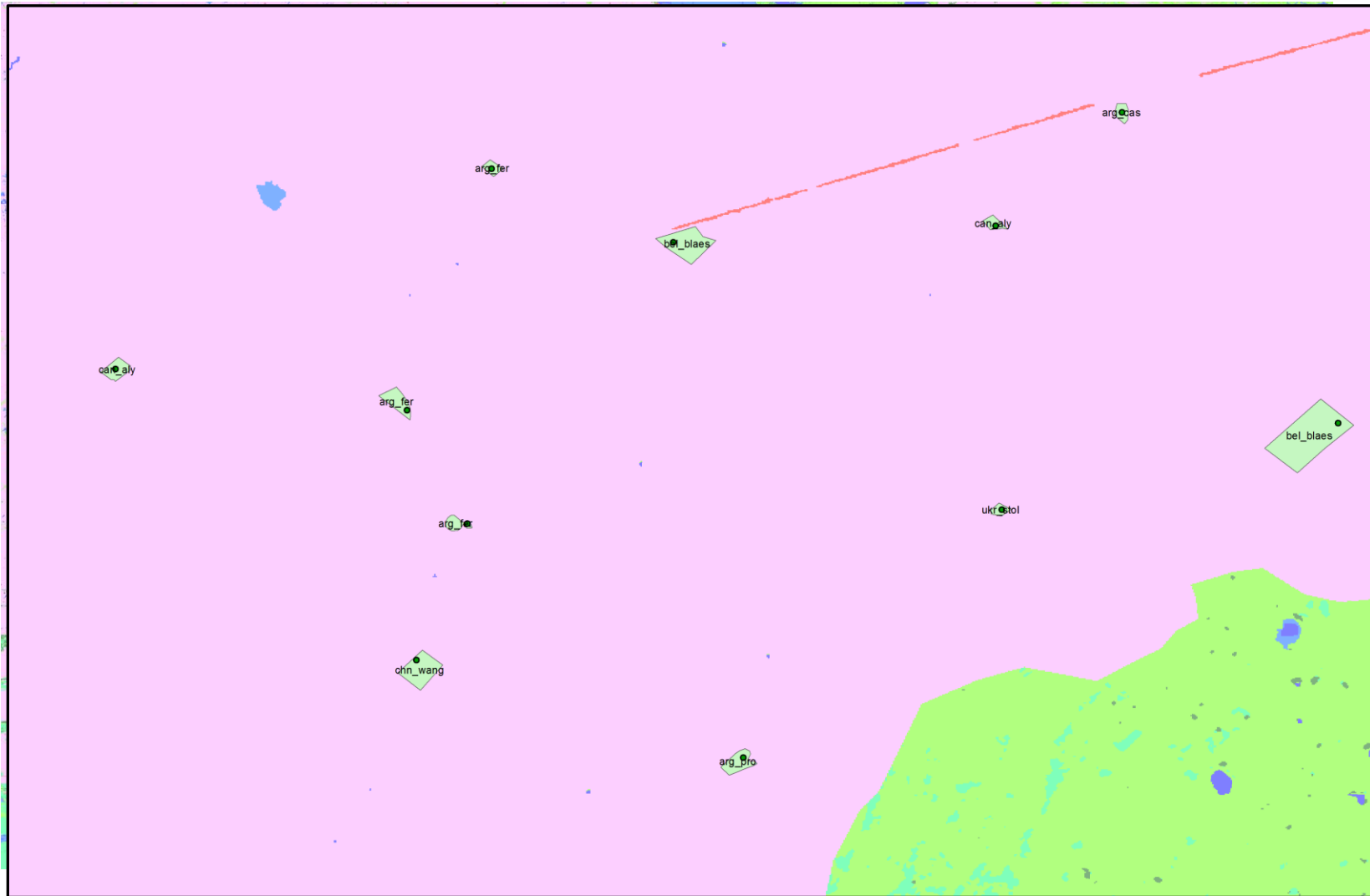
Характеристики набора

- ▶ Всего было создано 3594 точки для 6 участков и 10 экспертов
 - ▶ Для каждого участка набор разбивался на равные части (примерно по 60 точек на участок) случайным образом между участниками
 - ▶ 3,5% всех точек являлись контрольными, то есть доставались каждому пользователю и проходили через многократную интерпретацию
 - ▶ В результате 25% точек каждого пользователя были контрольными
-

Рисование полигонов CS на большой территории

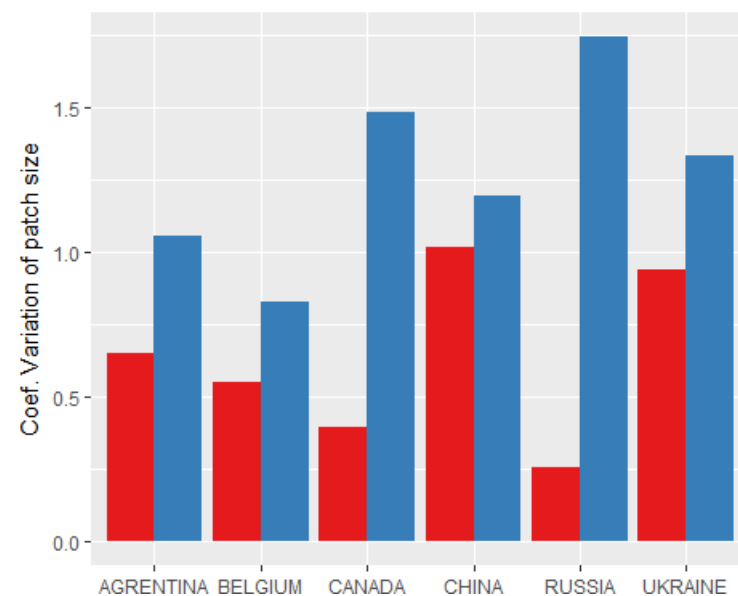
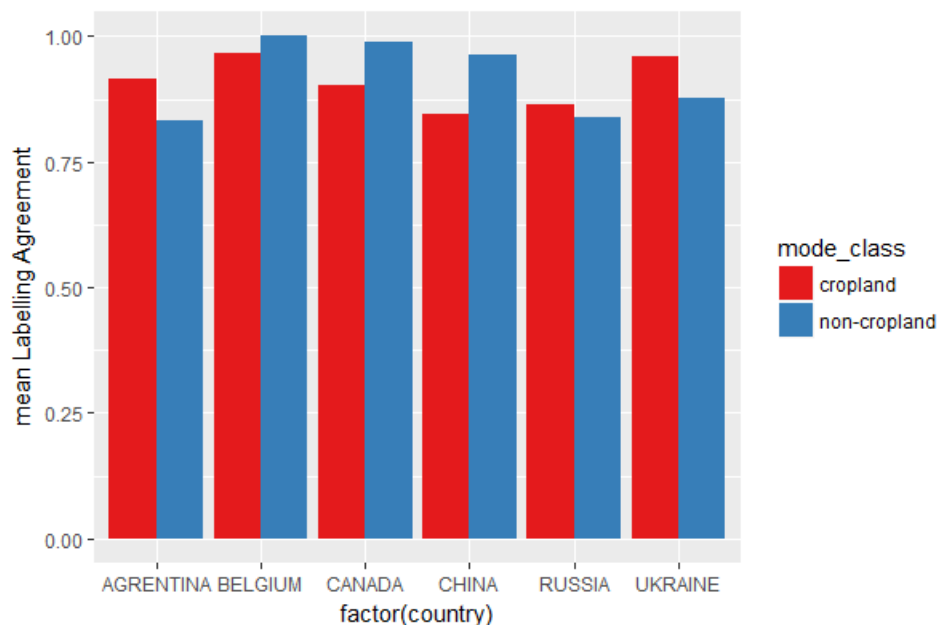
- ▶ Пользователи использовали временную серию снимков и инструменты навигации интерфейса VEGA-GEOGLAM для классификации растительности вокруг контрольной точки
 - ▶ Задачей пользователей было нарисовать замкнутый полигон, являющийся максимально однородным с точки зрения его сезонной фенологии, и который содержал бы в себе текущую опорную точку
 - ▶ При классификации полигона пользователи руководствовались определением пахотных земель JESAM
-

Пример полигонов CS



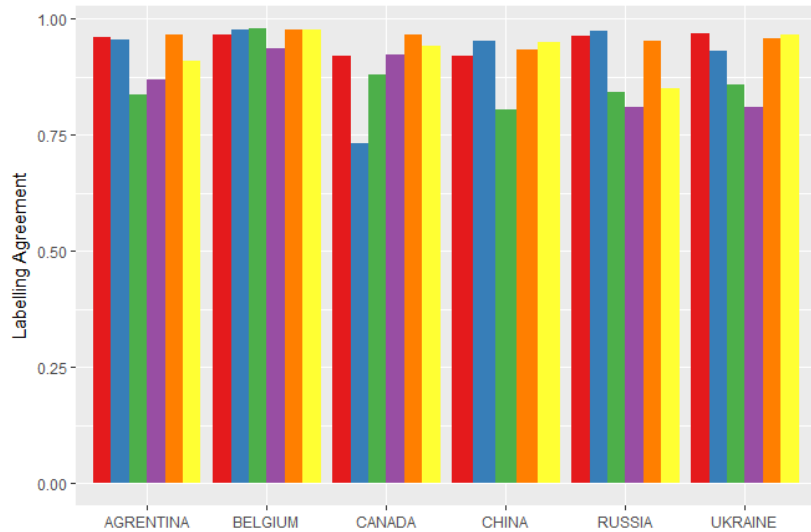
Оценка полученного набора

- ▶ Согласованность класса и размеров нарисованного полигона на **контрольных точках**

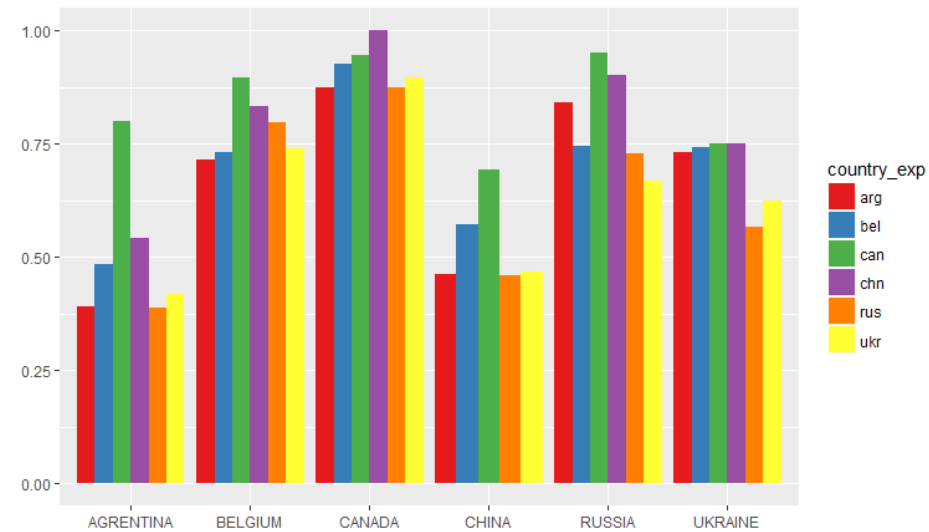


- ▶ Низкая вариабельность размеров полей для России и Канады (границы полей легко отличить), высокая вариабельность для Китая и Украины
- ▶ Как правило, высокая вариабельность размеров объектов класса «не пашня»

Сравнение CS с LC (GlobeLand30)



пашня



не пашня

- ▶ Высокая согласованность классов и отсутствие значимых тенденций для класса «пашня»
- ▶ Очевидная связь уровня согласованности классов с конкретным участком JESAM для класса «не пашня»

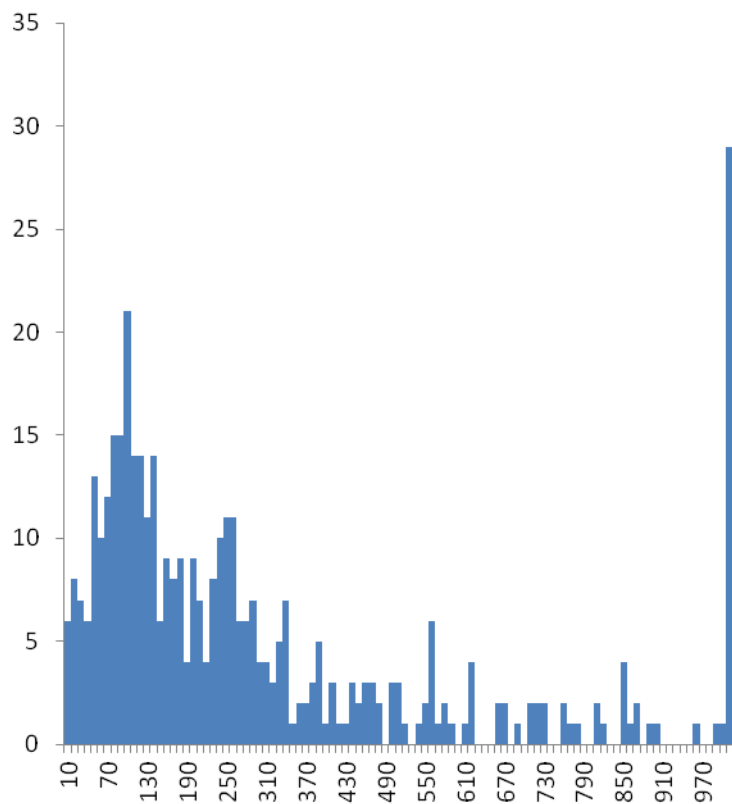
Оценка CS путём картографирования пашни и валидации результатов с наземными данными

Для создания опорной карты:

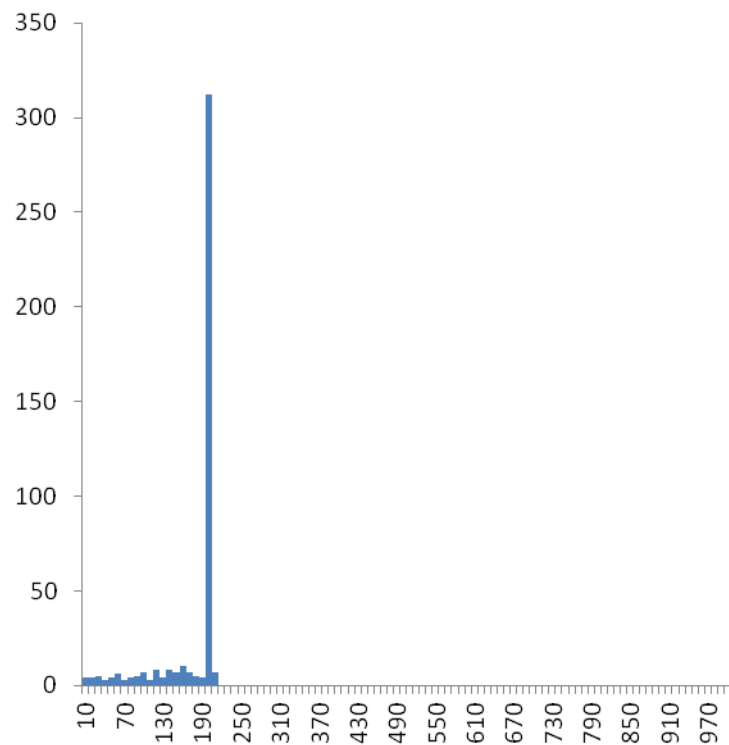
- ▶ Использование спектрально-динамических признаков $\text{Band}(\text{NDVI}_{\text{max}})$, $\text{Band}(\text{NDVI}_{\text{min}})$, Band_Mean извлеченных из восстановленных временных серий Landsat-8 для каналов Red, NIR, SWIR1 прибора OLI (восстановление с помощью NN Кохонена, SRI, Украина)
 - ▶ Использование локально-адаптивного подхода LAGMA и классификатора Random Forest
 - ▶ Эквализация (выравнивание) размеров объектов обучающей выборки для обеспечения корректного баланса между классами для НП классификатора
-

Эквализация размеров сэмплов выборки, полученной по данным CS (Аргентина)

Исходное распределение размеров



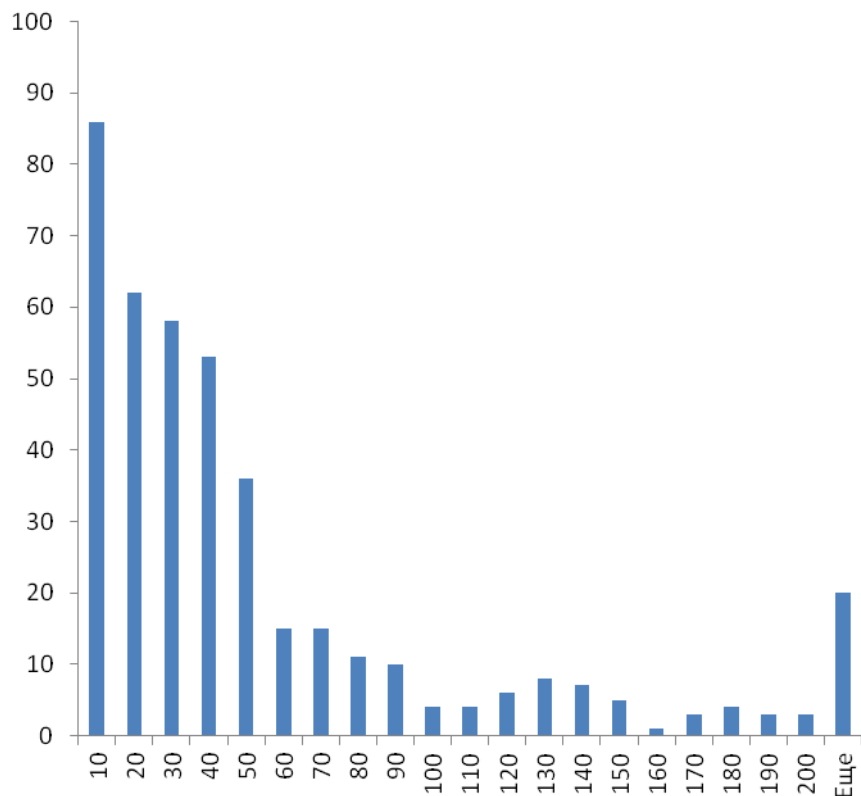
Распределение размеров модифицированной выборки



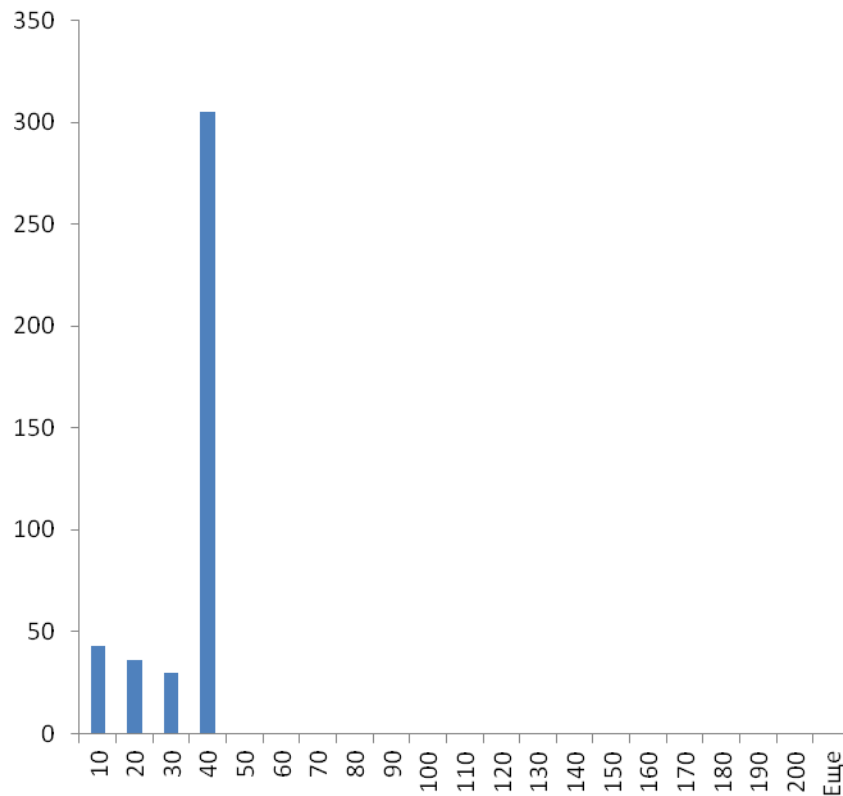
Осуществлялось приведение к медианному размеру сэмпла (в данном случае 200 пикселей)

Эквализация размеров сэмплов выборки, полученной по данным CS (Китай)

Исходное распределение размеров

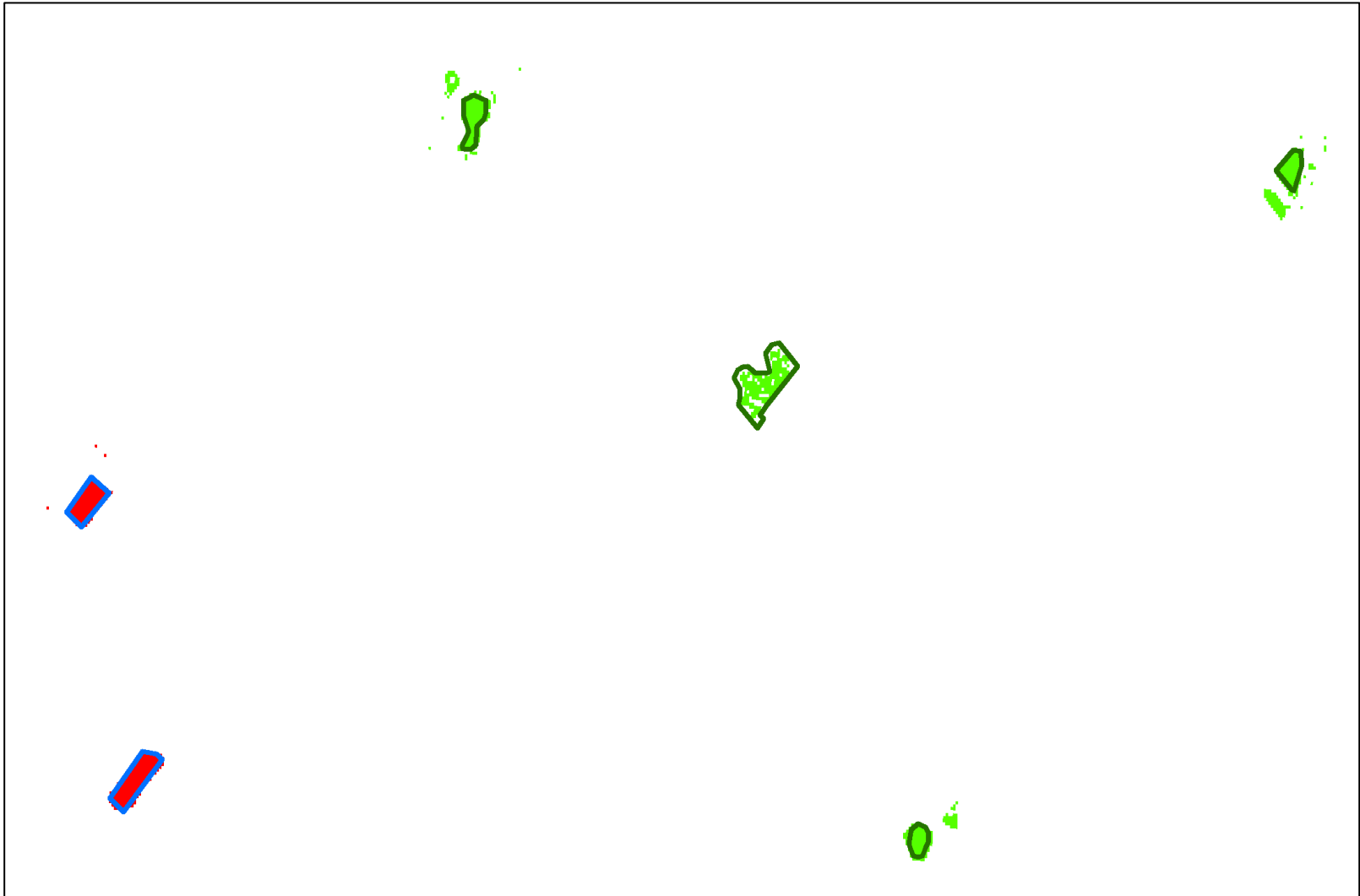


Распределение размеров модифицированной выборки

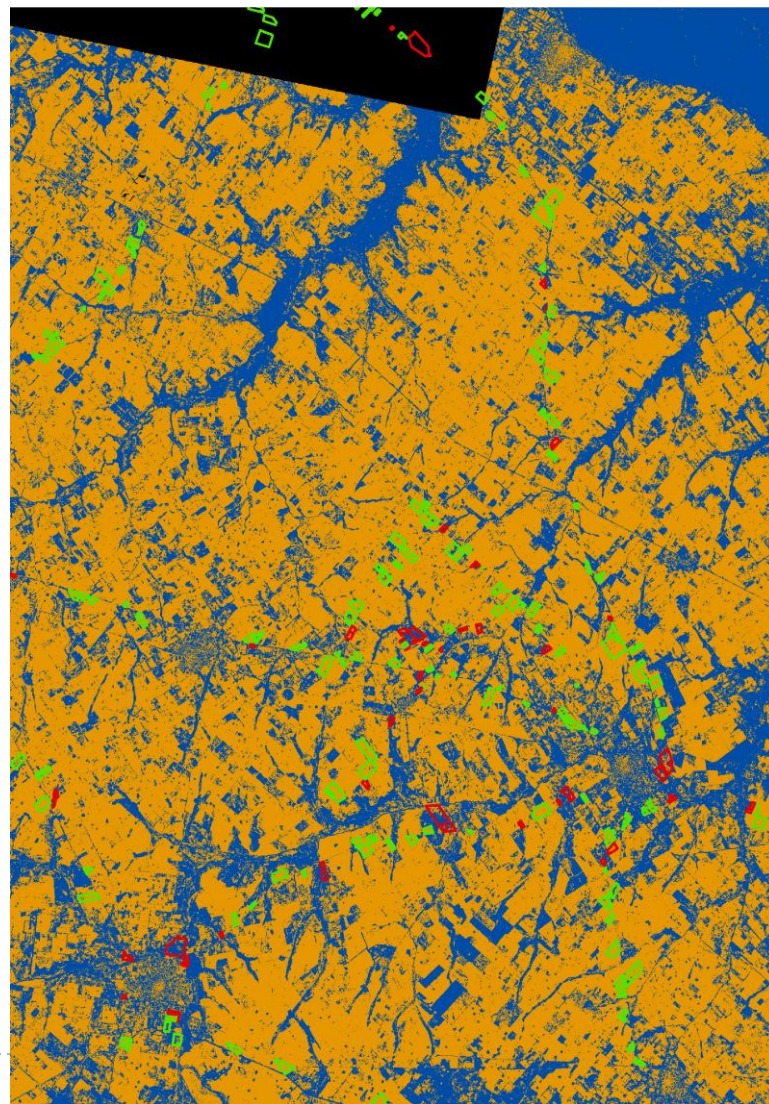
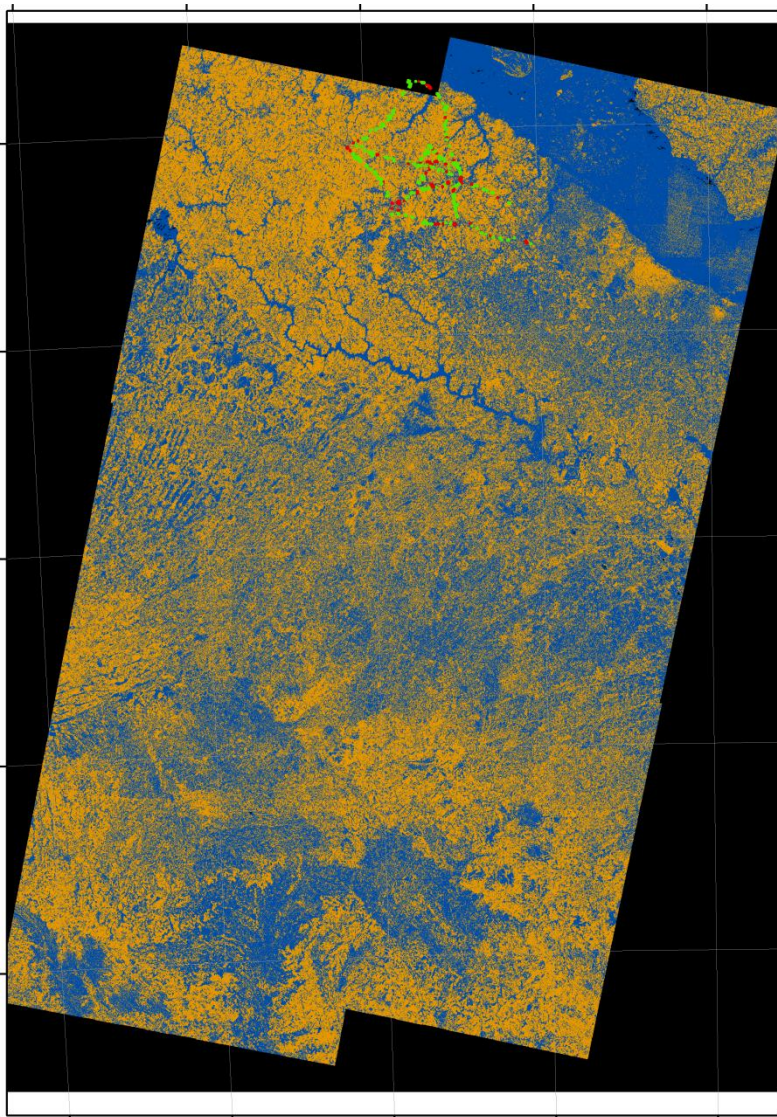


Медианный размер сэмпла в исходном распределении равен 31 пиксель

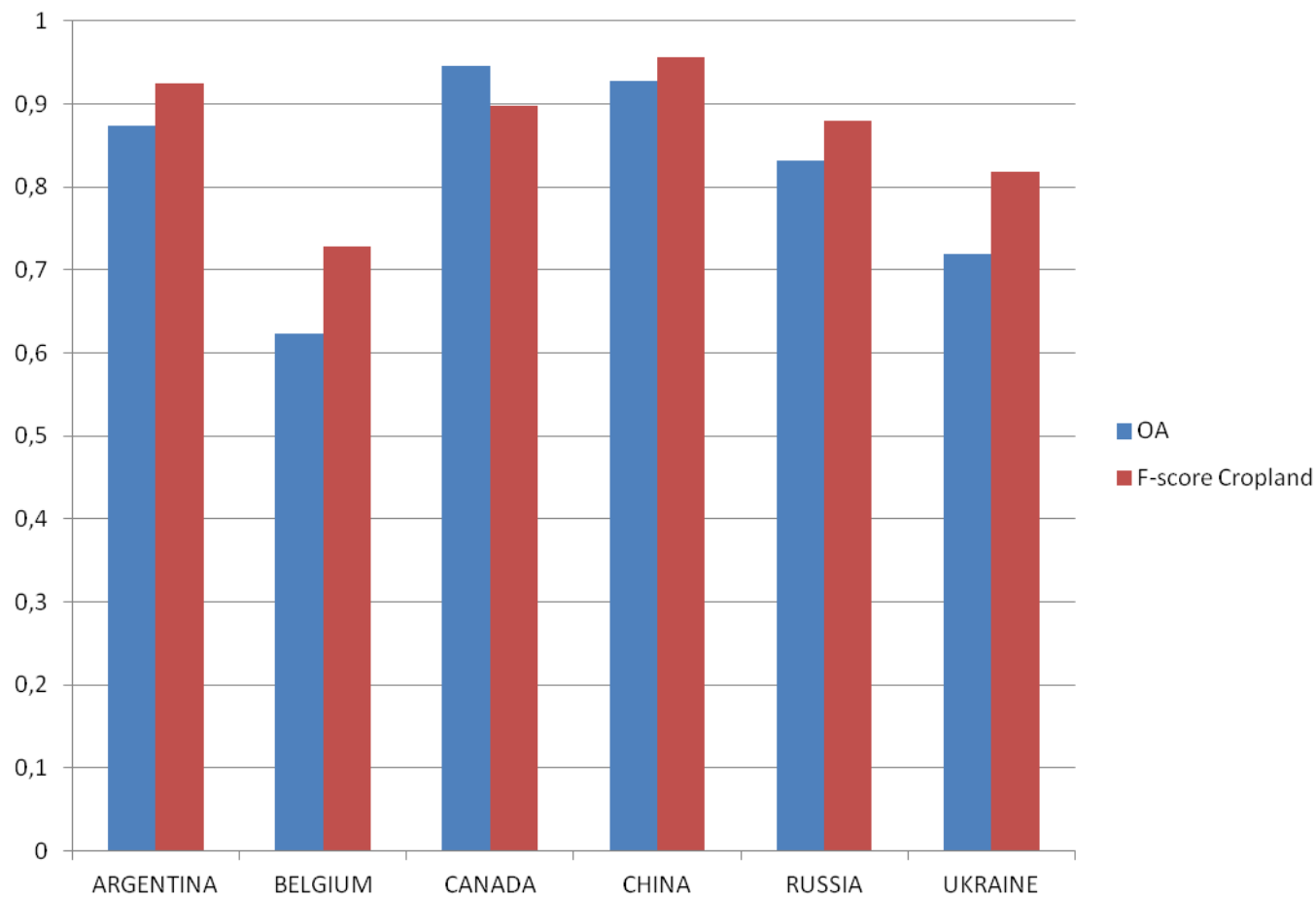
Эквализация размеров сэмплов



Опорная карта и наземные данные (IS) для её валидации (Аргентина)



Результаты валидации опорных карт пашни по данным наземных обследований (набор IS)



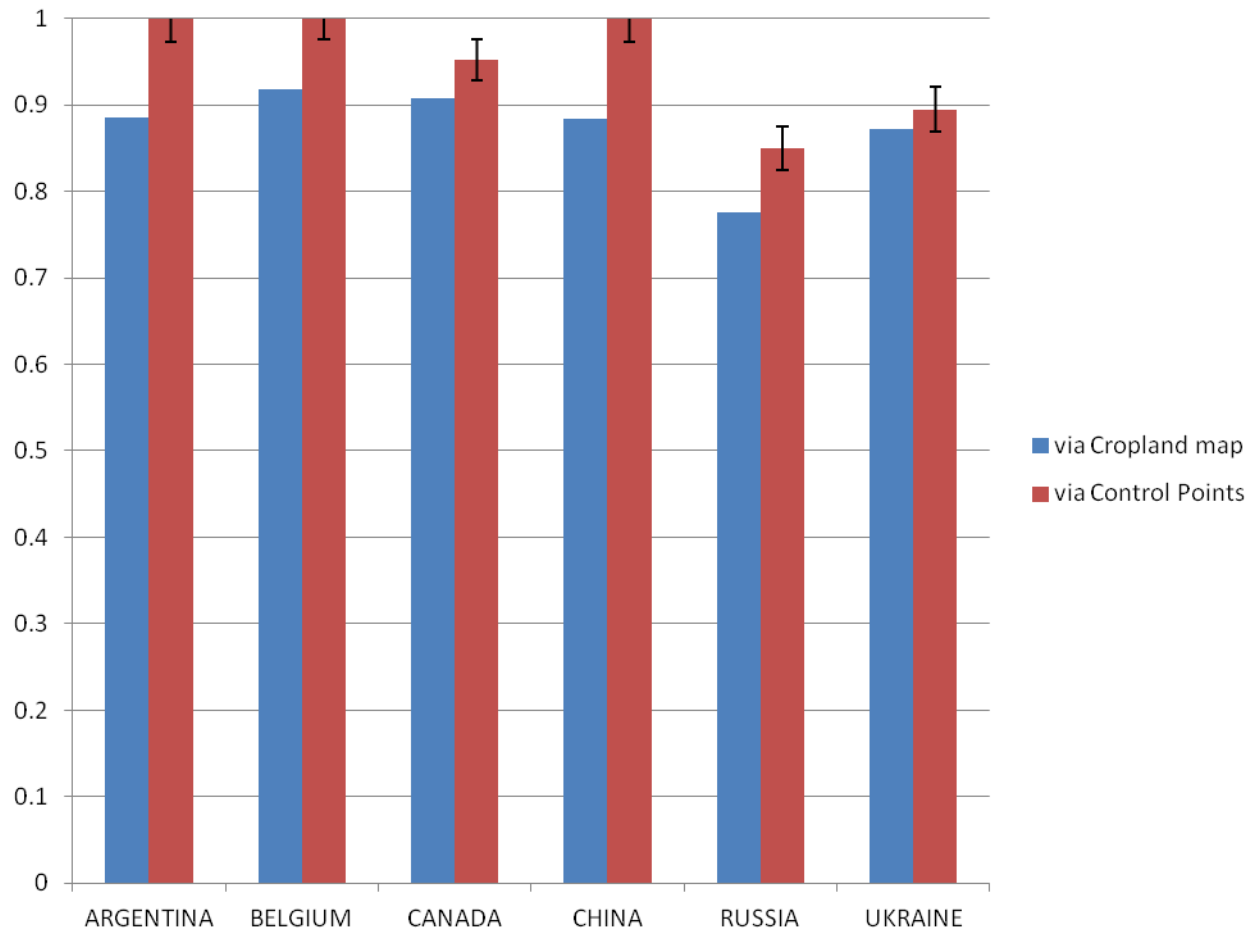
Оценка «новых» CS пользователей

- ▶ Изначально точность пользователей оценивалась с помощью *in situ* данных, однако это не всегда удобно
- ▶ Если опорная карта пахотных земель:
 - ▶ получена на основе CS выборки, созданной проверенными пользователями
 - ▶ имеет высокую точность, подтвержденную валидацией по наземным данным

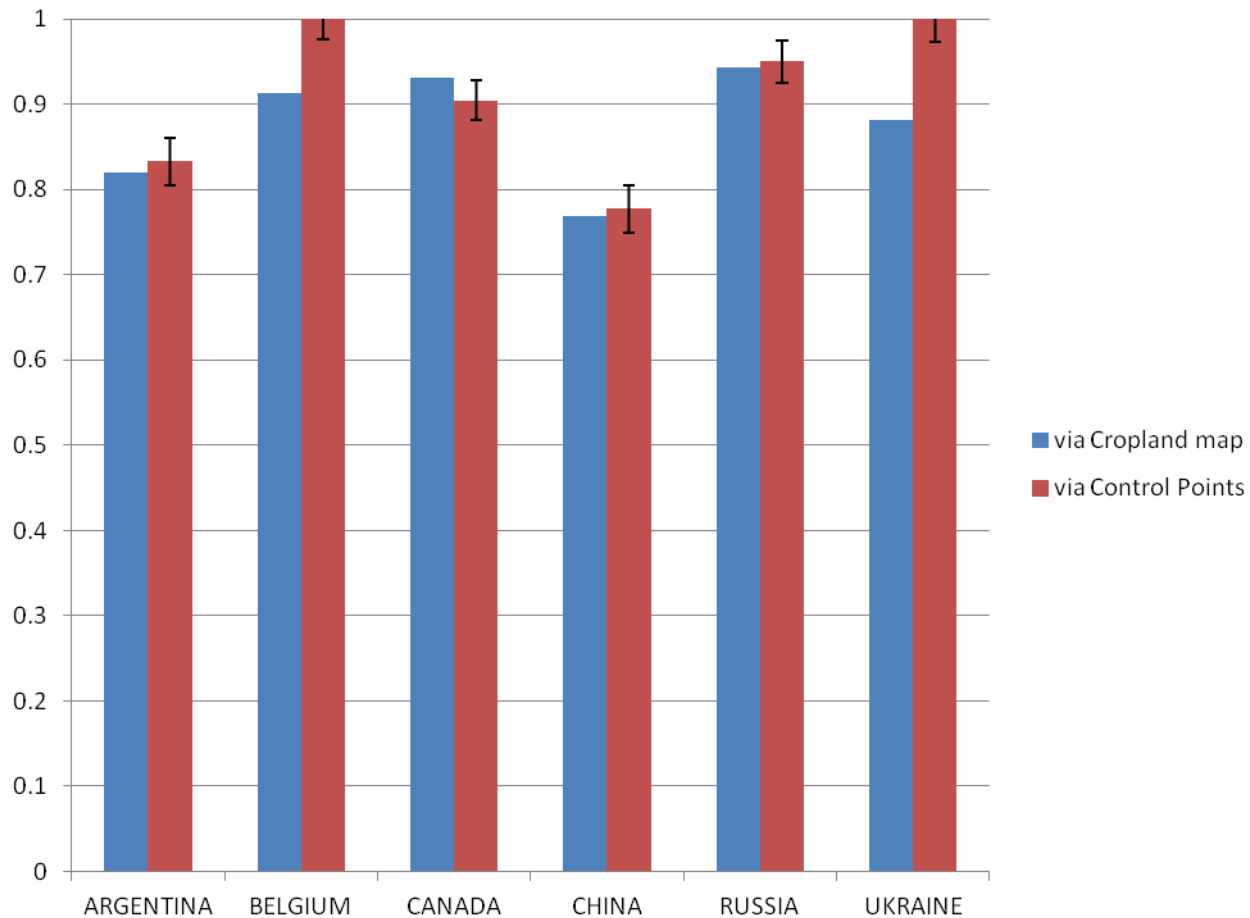


Контрольные точки проверенных пользователей могут быть использованы для оценки новых пользователей

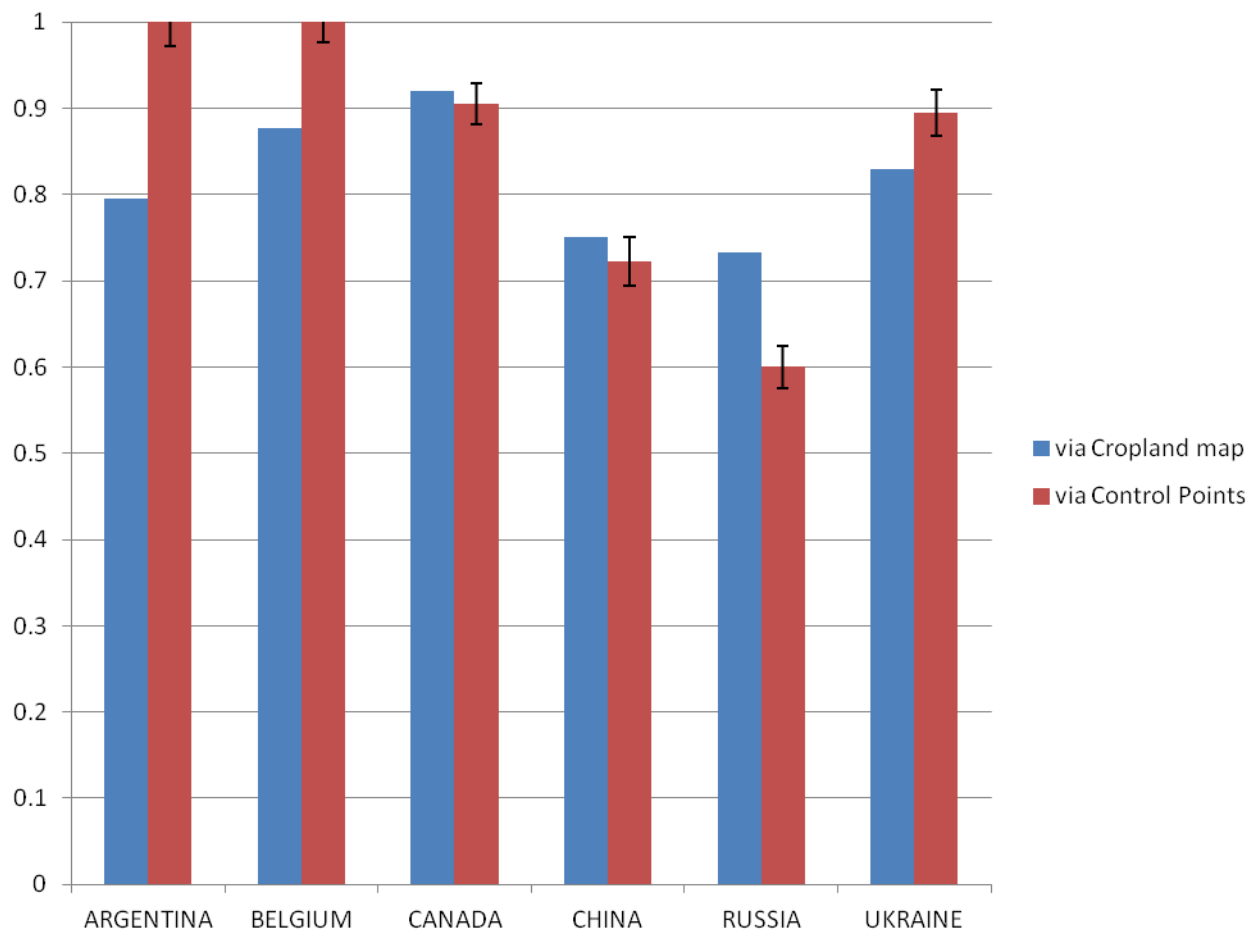
Оценка общей точности New User 1



Оценка общей точности New User2



Оценка общей точности New User3



Промежуточные выводы

- ▶ Данные из GlobeLand30 (LC) можно использовать для картографирования пахотных земель, обязательно учитывая определение пашни
 - ▶ Используемые спектрально-динамические признаки достаточно эффективны для картографирования пашни в ряде регионов мира, однако необходимо учитывать ограничения методов восстановления изображений высокого разрешения
 - ▶ Данные краудсорсинга (CS) позволяют достаточно точно выявлять используемую пашню в ряде регионов мира
 - ▶ Можно оценивать точность новых CS пользователей «на лету», т.е. без использования наземных данных для контроля
 - ▶ *В ближайшем будущем будет сделана оценка эффективности локально-адаптивного подхода и эквализации выборки, а также рассчитаны все необходимые метрики для сравнения источников обучающих данных*
-