

## **Using Crowdsourcing Datasets and Landsat Satellite Data for Cropland Mapping in Different Agrosystems of Global JECAM Network**

*Dmitry E. Plotnikov<sup>1</sup>, Diego de Aballeyra<sup>2</sup>, Santiago R. Veron<sup>2</sup>, Miao Zhang<sup>3</sup>, Vladimir A. Tolpin<sup>1</sup>, Sergey A. Bartalev<sup>1</sup>, Mykola Lavreniuk<sup>4</sup>, Francois Waldner<sup>5</sup>, Aly Ziad<sup>6</sup>*

<sup>1</sup>Space Research Institute, Russian Academy of Sciences, Moscow, Russia

[dmitplot@d902.iki.rssi.ru](mailto:dmitplot@d902.iki.rssi.ru)

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

<sup>3</sup>Institute of Remote Sensing and Digital Earth, Chinese Academy of Science, Beijing, China

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

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

<sup>6</sup>Agriculture and Agri-Food Canada, Ottawa, Canada

### **Abstract**

The study explores the reference information obtained by photo-interpretation of satellite images by an international team of volunteers for accurate cropland mapping over large areas based on Landsat remote sensing data and supervised classification. JECAM teams from Argentina, Belgium, Canada, China, Russia and Ukraine had arranged the crowdsourcing campaign and in situ data collection, used to assess the accuracy of the cropland maps derived. Vega-Geoglam system was set to collect distributed reference crowdsourcing dataset of controlled accuracy over each JECAM site, which was used to map cropland for the target growing season in six contrasted agroregions of the globe.

**Keywords:** cropland mapping, JECAM, Landsat, spectro-temporal features, crowdsourcing, VGI

## КАРТОГРАФИРОВАНИЕ ПАХОТНЫХ ЗЕМЕЛЬ В РАЗЛИЧНЫХ РЕГИОНАХ ГЛОБАЛЬНОЙ СЕТИ JESAM НА ОСНОВЕ СПУТНИКОВЫХ ДАННЫХ LANDSAT И ПОЛУЧЕННОЙ МЕТОДОМ КРАУДСОРСИНГА ОПОРНОЙ ИНФОРМАЦИИ

Д.Е. Плотников<sup>1</sup>, D. de Abelleira<sup>2</sup>, S. Veron<sup>2</sup>, M. Zhang<sup>3</sup>, В.А. Толпин<sup>1</sup>, С.А. Барталев<sup>1</sup>, М. Lavreniuk<sup>4</sup>, F. Waldner<sup>5</sup>, A. Ziad<sup>6</sup>

<sup>1</sup>Институт космических исследований РАН, Москва, Россия

[dmitplot@d902.iki.rssi.ru](mailto:dmitplot@d902.iki.rssi.ru)

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

<sup>3</sup>Institute of Remote Sensing and Digital Earth, Chinese Academy of Science, Beijing, China

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

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

<sup>6</sup>Agriculture and Agri-Food Canada, Ottawa, Canada

В работе рассматривается эффективность опорной информации, полученной с помощью фотоинтерпретации спутниковых снимков международным коллективом волонтеров, для решения задачи устойчивого распознавания используемой пашни на больших территориях на основе данных Landsat и методов обучаемой классификации. В эксперименте участвовали коллективы, осуществляющие поддержку участков JESAM из Аргентины, Бельгии, Канады, Китая, России и Украины, которые также обеспечивали настоящее исследование наземными опорными данными для оценки точности получаемых карт. В результате с помощью системы Vega-Geoglam была собрана репрезентативная опорная выборка с известными характеристиками точности, которая была использована для распознавания пахотных земель текущего сезона вегетации на примере шести различных агроландшафтов в указанных регионах мира.

**Ключевые слова:** используемые пахотные земли, JESAM, Landsat, спектрально-динамические признаки, краудсорсинг, VGI

### Введение

Точная и актуальная опорная информация является ключевым условием для надёжного распознавания используемой пашни на основе спутниковых данных и методов обучаемой классификации. Международная сеть JESAM (Joint Experiment for Crop Assessment and Monitoring, [jesam.org](http://jesam.org)) использует единые стандарты для сбора и обмена наземной научной информацией и включает в себя более трёх десятков участков в различных регионах мира, поддерживая исследования в области дистанционного мониторинга сельскохозяйственной

растительности [1-4]. На базе участков JESAM в Аргентине, Бельгии, Канаде, Китае, России и Украине был проведен эксперимент по сравнению эффективности различных источников опорной информации для распознавания используемых пахотных земель, а именно, данных, собранных путём наземных обследований (*in situ*), и полученных методами интерпретации спутниковых изображений экспертами. Несмотря на то, что данные наземных обследований являются наиболее подробными, актуальными и часто используются как эталон точности, их сбор для регулярного картографирования пашни является дорогостоящим, времязатратным и неосуществим в глобальном масштабе. В настоящем эксперименте альтернативой *in situ* являлись данные, полученные путем визуальной фотоинтерпретации спутниковых изображений экспертами-добровольцами из стран-участников. Данные краудсорсинга были собраны с помощью платформы Vega-Geoglam [5] по результатам интерпретации снимков высокого пространственного разрешения Landsat-7,8 (OLI) и Sentinel-2 (MSI) участниками эксперимента, которые пользовались определением пашни JESAM для установления класса объектов. Целевой сезон вегетации был одинаков для всех участков в обоих полушариях. Целью исследования являлась, в том числе, оценка возможности высокодетального картографирования используемой пашни в этих регионах мира на субнациональном уровне пространственного охвата (несколько сотен километров) на основе спутниковых данных Landsat.

### Сбор *in situ* данных

Сбор наземной информации происходил на небольшой части целевого региона картографирования. Несколько сотен элементов выборки (сэмплов) было собрано для каждого из шести участков JESAM методом визуального обследования вдоль дорог с последующим уточнением их границ на основе данных высокого пространственного разрешения (Таблица 1). Размеры обеспеченного наземной информацией полигона различались между участками JESAM и составляли в среднем около 50 км в поперечнике. Время полевых обследований соответствовало вегетационному периоду 2015 года в северном полушарии и периоду 2014-2015 годов в южном полушарии. В процессе сбора наземной информации составлялось краткое описание сэмпла, которое позволяло установить его класс для текущего сезона вегетации на основе легенды «пашня» – «не пашня» (или «остальное»). Для исключения неопределенности на всех этапах интерпретации наземной и дистанционной информации использовалось единое для всех участков определение, принятое для пахотных земель в рамках сети JESAM: “с точки зрения дистанционного зондирования, пашней текущего сезона вегетации называется участок земной поверхности площадью минимум 0.25 га (с шириной минимум 30 метров), который засеян или засажен растениями и с которого собирается урожай хотя бы один раз в течение 12 месяцев после даты сева или посадки; при этом пашня текущего сезона вегетации должна быть занята травянистой растительностью, которая в некоторых случаях может сочетаться с древесной или кустарниковой растительностью”. Используя априорные знания и дополнительную картографическую информацию о регионе исследования, экспертами было установлено соотношение между классами «пашня» и «не пашня», которое служило ориентиром при сборе ими наземной информации для обеспечения необходимой репрезентативности классов.

Таблица 1. Распределение числа обследованных полей обоих классов для JESAM участков в Аргентине, Бельгии, Канаде, Китае, России и Украине

Участок	Общее число полей	Полей класса «пашня»	Полей класса «не пашня»
Аргентина	293	226	67
Бельгия	201	131	70
Канада	216	81	135
Китай	731	578	153
Россия	319	259	60
Украина	224	170	54

## Организация сбора волонтерской информации: инструменты и этапы

Интерпретация спутниковой информации происходила с использованием интерфейса и инструментов системы Vega-Geoglam ([www.vega-geoglam.ru](http://www.vega-geoglam.ru)), которая даёт возможность проверить принятое определение пахотных земель текущего сезона вегетации применительно к исследуемому объекту. Навигационные инструменты позволяли участникам устанавливать последовательность обработки сэмплов и участков JECAM на всех этапах сбора опорных данных (Рис.1).



Рис.1. Общий вид вкладки краудсорсинга интерфейса Vega-Geoglam с панелью навигации (слева) и инструментами рисования и классификации сэмплов

Сбор получаемой путём краудсорсинга опорной информации происходил в два этапа. На первом этапе эксперимента происходила оценка участников-волонтеров с точки зрения понимания ими поставленной задачи и навыков использования геоинформационных инструментов и спутниковой информации. Оценка осуществлялась на основе данных *in situ* и сопровождалась расчетом количественных метрик для каждого участника. На рис.2 приведена агрегированная по участникам информация, которая позволяет обнаружить различия мировых агроландшафтов с точки зрения особенностей их интерпретации международным коллективом экспертов. В частности, участок JECAM в Китае оказался самым сложным при установлении точных границ объектов из-за малых средних размеров полей в стране, а участок в Бельгии характеризуется минимальной точностью классификации ввиду обилия сенокосов и сложностей с их интерпретацией по спутниковым данным из-за частой облачности в регионе.

Второй этап предусматривал создание экспертами-волонтерами набора опорной информации для целевого региона картографирования субнационального уровня охвата (размером примерно 500 на 300 км) для каждого участка JECAM. Требуемые рассмотрения и интерпретации волонтерами фрагменты территории были заранее определены с помощью случайно стратифицированного сэмплинга, созданного с использованием карты GlobeLand30[6] и результатов глобальной агростратификации GAES[7]. Всего для интерпретации второго этапа было создано 3594 сэмпла для 6 участков и 10 волонтеров. При этом некоторые случайно выбранные сэмплы обоих классов являлись контрольными, т.е. проходили многократную интерпретацию всеми участниками и использовались, в том числе, для оценки полученного набора на втором этапе.

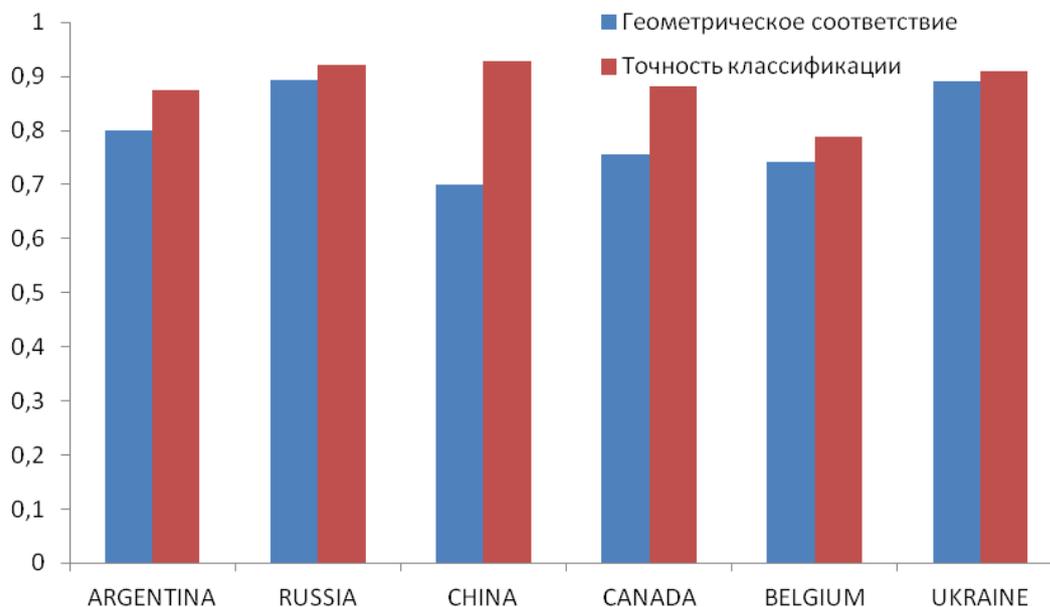


Рис.2. Средние по участкам значения метрик геометрического соответствия (точности рисования) и точности классификации, полученные путем сопоставления данных краудсорсинга с *in situ* данными

### Спектрально-динамические признаки и локально-адаптивная классификация

Для распознавания используемых пахотных земель использовались атмосферно-скорректированные измерения коэффициентов спектральной яркости (КСЯ) прибора OLI (Landsat-8) с пространственным разрешением 30 метров. Временные серии измерений прибора в красном, ближнем ИК и среднем ИК были очищены от влияния облачности и теней от неё, а образованные в результате пропуски были восстановлены с помощью ансамблей нейронных сетей Кохонена[8]. С использованием восстановленных серий были построены признаки  $\text{Band}(\text{NDVI}_{\max})$ ,  $\text{Band}(\text{NDVI}_{\min})$  [9], где Band означает один из перечисленных каналов, а в скобках указано условие выбора значения внутри текущего сезона (наступление максимального или минимального значения). Кроме этого, в качестве признаков также были использованы средние за целевой сезон вегетации значения КСЯ в этих каналах.

Поскольку сельскохозяйственная растительность зачастую характеризуется неодномодальным распределением внутриклассовых значений, выявление пахотных земель осуществлялось на основе непараметрического классификатора Random Forest [10]. Кроме этого, изменчивость почвенно-климатических и агротехнических условий возделывания пашни внутри регионов субнационального уровня пространственного охвата потребовала использования локально-адаптивного метода распознавания LAGMA [11].

Полученная в результате краудсорсинга выборка имела заданный тип сэмпинга (случайно стратифицирована), однако эксперты не были ограничены в размерах создаваемых ими сэмплов. Для восстановления статистического баланса между классами была выполнена эквализация (выравнивание) размера каждого сэмпла выборки до типичного (медианного) размера, вычисленного для каждого участка JESAM как число занимаемых объектом пикселей. Прореживание выполнялось, если размеры сэмпла были выше медианного, при этом в первую очередь исключались наиболее нетипичные по отношению к исходному среднему в пространстве используемых признаков значению пиксели. В случае если размеры сэмпла были меньше медианного, делалась попытка увеличить его размер, присоединяя прилежащие пиксели, которые отклонялись от исходного среднего не более чем на заданное число стандартных отклонений.

## Результаты распознавания используемой пашни на больших территориях различных агроландшафтов

Сбалансированная опорная выборка была использована совместно с признаками распознавания для получения карт пахотных земель на больших территориях субнационального уровня пространственного охвата вокруг участков JESAM (рис. 3). Валидация полученных региональных карт высокого пространственного разрешения осуществлялась на основе наземной информации с вычислением метрик общей точности и F-score [12]. F-score является класс-ориентированной метрикой, которая может быть рассчитана отдельно для каждого класса, минимизируя влияние со стороны остальных классов.

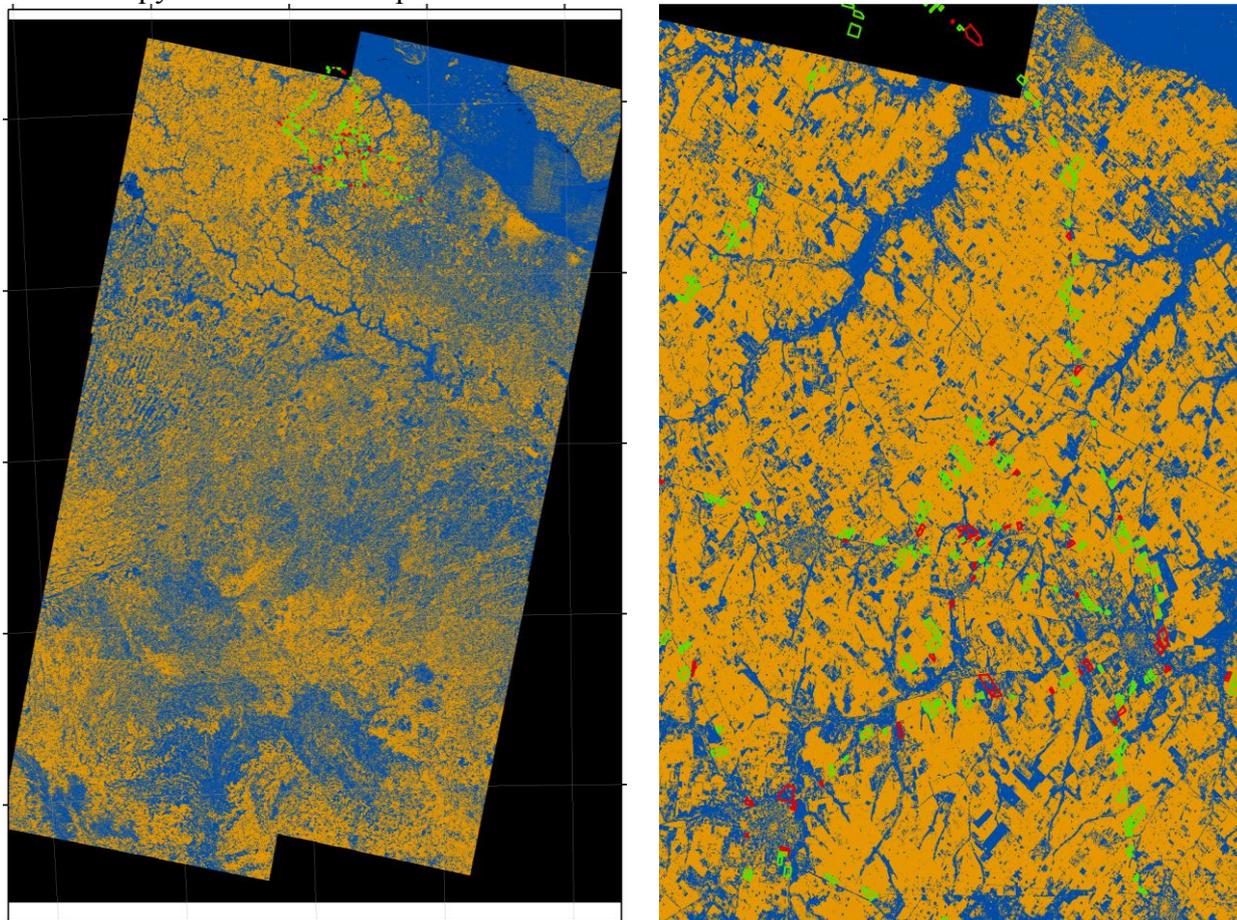


Рис. 3. Результат распознавания используемых пахотных земель Аргентины на субнациональном уровне пространственного охвата (слева), оранжевым цветом обозначен класс «пашня», синим – «остальное»; а также увеличенный участок полученной карты на регион наземных обследований (справа), зелеными контурами обозначены обследованные поля класса «пашня», красным – класса «остальное»

Сравнение полученных карт с данными полевых обследований указывает на относительно высокие точности выявления пашни в рамках эксперимента (рис. 4). Можно увидеть, что повышенная сложность фотоинтерпретации, выявленная на первом этапе краудсорсинга, также приводит к понижению точности распознавания пахотных земель, что проявляется на примере участка JESAM в Бельгии. Повышение числа одновременных наблюдений благодаря использованию дополнительных данных высокого пространственного разрешения при фотоинтерпретации волонтерами может решить такого рода проблемы.

Работа была выполнена в рамках темы «Мониторинг» (госрегистрация № 01.20.0.2.00164) с использованием ресурсов Центра коллективного пользования «ИКИ-Мониторинг» [13].

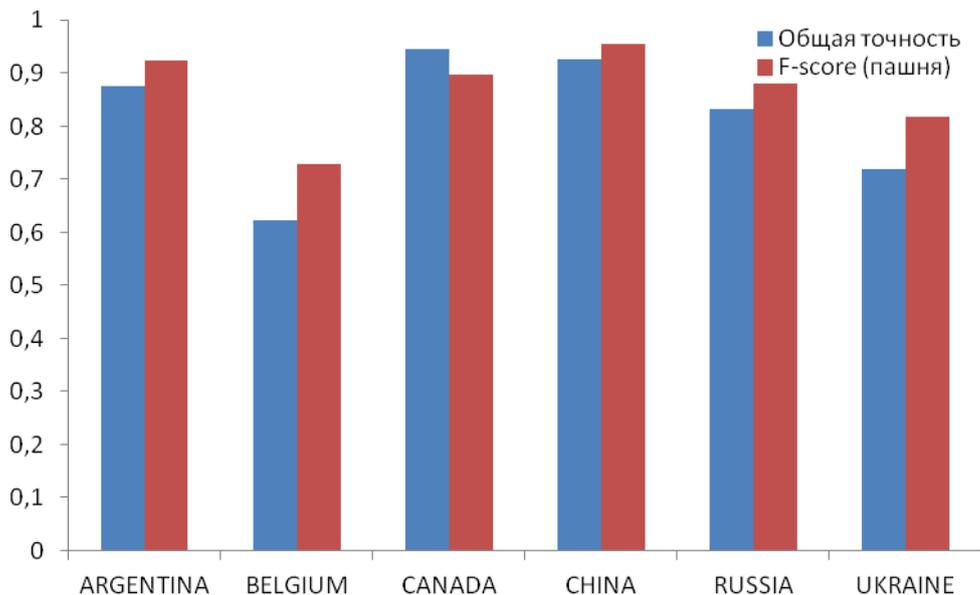


Рис.4. Точности полученных региональных карт используемой пашни в сравнении с результатами наземных обследований в терминах значений общей точности и F-score для класса «пашня»

## References

- [1] Waldner F., De Abelleira D., Santiago V., Zhang M., Wu B., Plotnikov D.E., Bartalev S.A., Lavreniuk M., Skakun S., Kussul, N.N., Le Maire G., Dupuy S., Jarvis I., Defourny, P. Towards a set of agrosystem-specific cropland mapping methods to address the global cropland diversity, *International Journal of Remote Sensing*, 2016, Vol. 37. Issue 14. DOI:10.1080/01431161.2016.1194545.
- [2] Plotnikov D.E., Kolbudaev P.A., Bartalev S.A., Loupian E.A. Automated annual cropland mapping from reconstructed time series of Landsat data, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2018, Vol. 15, No. 2, pp. 112–127. DOI: 10.21046/2070-7401-2018-15-2-112-127. (in Russian)
- [3] Bartalev S.A., Plotnikov D.E., Loupian E.A. Mapping of arable land in Russia using multiyear time series of MODIS data and the LAGMA classification technique, *Remote Sensing Letters*, 2016, Vol. 7. No. 3, pp. 269-278. DOI: 10.1080/2150704X.2015.1130874.
- [4] Plotnikov D.E., Kolbudaev P.A., Bartalev S.A. Identification of dynamically homogeneous areas with time series segmentation of remote sensing data, *Computer Optics*, 2018, Vol. 42, No. 3, pp. 447-456. DOI: 10.18287/2412-6179-2018-42-3-447-456. (in Russian).
- [5] Tolpin V.A., Balashov I.V., Bartalev S.A., Loupian E.A., Plotnikov D.E., Ladonina N.N., Stytsenko F.V. Information System for Global Satellite Monitoring of Agriculture VEGA-GEOGLAM // *XIII All-Russian Open Conference "Contemporary Problems of Remote Sensing of the Earth from Space"*, Book of Abstracts, Moscow, IKI RAS, 16-20 November 2015, P. 431. (in Russian)
- [6] 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, pp 7-27, ISSN 0924-2716, <https://doi.org/10.1016/j.isprsjprs.2014.09.002>.
- [7] Fischer G., Nachtergaele F., Prieler S., Teixeira E., Tóth G., Velthuizen H., Verelst L., Wiberg D. Global Agro-Ecological Zones (GAEZ v3.0) - Model Documentation, 2012.
- [8] Skakun, S.V.; Basarab, R.M. Reconstruction of Missing Data in Time-Series of Optical Satellite Images Using Self-Organizing Kohonen Maps, *Journal of Automation and Information Sciences*, 2014, 46.

- [9] Matton, N., G. S. Canto, F. Waldner, S. Valero, D. Morin, J. Inglada, M. Arias, S. Bontemps, B. Koetz, and P. Defourny. An Automated Method for Annual Cropland Mapping along the Season for Various Globally-Distributed Agrosystems Using High Spatial and Temporal Resolution Time Series, *Remote Sensing*, 2015, 7 (10): 13208–13232. doi:10.3390/rs71013208.
- [10] Breiman L. Random Forests, *Machine Learning*, 2001, Vol. 45, No. 1, DOI:10.1023/A:1010933404324.
- [11] Bartalev S.A., Egorov V.A., Loupian E.A., Khvostikov S.A. A new locally-adaptive classification method LAGMA for large-scale land cover mapping using remote-sensing data, *Remote Sensing Letters*, 2014, Vol. 5, No. 1, pp. 55-64, DOI: 10.1080/2150704X.2013.870675.
- [12] Stehman S.V. Selecting and Interpreting Measures of Thematic Classification Accuracy. *Remote Sensing of Environment*, 1997, Vol.62, pp. 77–89, doi:10.1016/S0034-4257(97)00083-7.
- [13] Loupian EA, Proshin AA, Burtsev MA, Balashov IV, Bartalev SA, Efremov VYu, Kashnitskiy AV, Mazurov AA, Matveev AM, Sudneva OA, Sychugov IG, Tolpin VA, Uvarov IA. IKI center for collective use of satellite data archiving, processing and analysis systems aimed at solving the problems of environmental study and monitoring, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2015, 12(5), pp. 263-284.(in Russian)