

УДК 528.854.4

ДЕШИФРИРОВАНИЕ СОВРЕМЕННОЙ СТРУКТУРЫ ЗЕМЕЛЬ НА ТЕРРИТОРИИ СРЕДНЕГО ПОВОЛЖЬЯ РФ ПО ДАННЫМ LANDSAT

М. А. Иванов

*Институт экологии и природопользования, Казанский федеральный университет,
ул. Товарищеская. 5, 420097, Казань, Россия, maximko-87@mail.ru*

В рамках исследования проведено распознавание современной структуры земель на территории крупного региона России по данным Landsat с применением алгоритма RandomForest. Особое внимание уделено подготовке обучающей выборки. Точность распознавания, как общая, так и по отдельным классам превышает 96 %. В результате получена пространственная модель землепользования/земного покрова, включающая 5 классов, рассчитаны их площади. Отработанная методика и полученные данные в дальнейшем будут использованы для детектирования заброшенных пахотных земель за период 1984-2022 гг.

Ключевые слова: землепользование; земной покров; Landsat; классификация; Random Forest.

RECOGNIZING THE MODERN LAND USE STRUCTURE IN THE TERRITORY OF THE MIDDLE VOLGA REGION OF THE RUSSIA USING LANDSAT DATA

M. A. Ivanov

*Institute of Environmental Sciences, Kazan Federal University,
Tovarisheskaya str., 5, 420097, Kazan, Russia, maximko-87@mail.ru*

The modern land use structure in the territory of a large region of Russia was mapped on Landsat data using the Random Forest algorithm. Particular attention is paid to preparing the training sample. The recognition accuracy, both overall and for individual classes, exceeds 96 %. As a result, a spatial model of land use/land cover was obtained, including 5 classes, and their areas were calculated. The methodology and the data obtained will be used in the future to detect abandoned cropland for the period 1984-2022.

Keywords: land use; land cover; Landsat; classification; Random Forest.

Основной целью исследования является распознавание современной структуры землепользования/земного покрова на территории Среднего Поволжья РФ по данным Landsat 8 и 9. Для этого был решен ряд задач:

- подготовка композитов безоблачных данных Landsat 8, 9 за бесснежные сезоны 2018-2022 гг.;
- подготовка обучающей выборки основных классов землепользования/земного покрова;
- распознавание земель алгоритмом RandomForest и оценка достоверности результатов.

Территория исследования включает в себя 6 субъектов Российской Федерации: Республику Марий-Эл, Республику Татарстан, Чувашскую Республику, Ульяновскую, Самарскую и Саратовскую области. Общая площадь составляет 301733 км².

Существует ряд открытых глобальных моделей современной структуры землепользования и земного покрова, однако большинство из них создано без использования эталонов на территории России. Кроме того, полученные данные будут использоваться для выявления заброшенных пахотных земель за период 1984-2022 гг., которые целесообразно вернуть в сельскохозяйственный оборот. Для этого необходимо соблюдать единую методику при подготовке и анализе данных за предыдущие периоды.

В качестве исходных данных были использованы снимки Landsat 8 и 9 OLI (Collection 2, Level-2 DataProducts) за бесснежный период (май-октябрь) 2018-2022 гг. Средствами GoogleEarthEngine были рассчитаны двухмесячные композиты (май-июнь, июль-август, сентябрь-октябрь) данных, каждый из которых включает медианные значения спектральных каналов 2-7, и статистические метрики (среднее, медиана, максимум, стандартное отклонение) по 3 вегетационным индексам: NDVI (Normalized Difference Vegetation Index), EVI2 (Enhanced Vegetation Index 2), TSAVI (Transformed Soil Adjusted Vegetation Index).

Далее была подготовлена обучающая выборка для классификации основных типов землепользования и земного покрова. Согласно литературе, в идеале необходимо использовать в качестве обучающей выборки данные наземных наблюдений (GroundTruth). Однако в силу обширного пространственного охвата исследуемой территории, а также труднодоступности ряда участков допускается использование данных ДЗЗ с разрешением на порядок выше данных, по которым проводится распознавание [1, 2]. Также опорная информация должна быть близкой по дате к данным, используемых при производстве карт, чтобы избежать влияния изменения ландшафта [3], а также должна быть независима от данных, используемых при дешифрировании [4, 5].

Таким образом, в качестве опорных данных для создания обучающей выборки для дешифрирования современной структуры земель были использованы разновременные (за период 2018-2023 гг.) спутниковые снимки сверхвысокого разрешения (не хуже 1 м) из каталога

MaxarTechnologies (GeoEye-1,2; WorldView-2,3; Ikonos 1,2 и др.), представленные в программе GoogleEarth.

В качестве эталонных участков выступали векторные квадратные полигоны со стороной 90 м, что соответствует участку 3x3 пикселя на 30-метровых каналах Landsat. В качестве эталонных выбирались однородные участки, точно соответствующие тому или иному типу земель, для избегания проблемы смешанных пикселей. Обучающая выборка создавалась для 5 типов землепользования и земного покрова. Каждому полигону в таблице атрибутов присваивался соответствующий код: 1 – Водные объекты, 2 – Лесопокрытые территории, 3 – Луга, 4 – Используемая пашня, 5 – Антропогенные объекты. Для каждого класса были определены дешифровочные признаки. Наиболее простым классом являются водоемы, отображающиеся на снимках темно-синим, почти черным цветом. Подавляющая часть полигонов выборки была выбрана на водохранилищах Волги и Камы. Леса хорошо распознаются по зеленому цвету (разного оттенка в зависимости от породного состава), характерному рисунку (зернистой текстуре), создаваемой кронами деревьев, падающей тени. В осенний и зимний период также важен цвет (желтеющая листва в лиственных и смешанных лесах, зеленые кроны хвойных пород зимой).

Особое внимание стоит уделить классу травянистой растительности (лугам). В силу того, что этот класс является не целевым для распознавания, а также по причине схожести спектральных характеристик в этот класс были включены гигрофитные пойменные луга, мезо- и ксерофитные луга, приуроченные к склонам речных долин. В это же класс были отнесены участки заброшенной пашни, которая в зависимости от давности забрасывания сначала зарастает травостоем одного-двух видов, а далее разнотравьем, которое на снимках хорошо распознается по характерному рисунку, создаваемому ассоциациями, несмотря на то, что во многих случаях сохраняются границы полей и даже следы распашки. Особенно это характерно для южной части Саратовской области. Кроме того, в это класс были отнесены сфагновые болота (сплавины). Поскольку подготовка обучающей выборки — процесс итеративный, ее создание сопровождалось проведением тестовых классификаций. Изначально болота были отнесены в отдельный класс, но результаты тестового распознавания показали низкую достоверность их выделения. Более того, из-за этого снижалась точность детектирования лугов. Поэтому было принято решение объединить этот класс с луговой растительностью.

Используемая пашня на снимках сверхвысокого разрешения легко детектируется по характерным ортогональным границам полей, цветам почв после первичной обработки, ранним всходам озимых культур (когда

естественный травостой еще не взошел), стерне и пожнивным остаткам после уборки урожая, и текстуре, создаваемой бороздами от распашки.

В качестве эталонных участков для распознавания антропогенных объектов выступали крупные здания (склады и цеха на территории промышленных предприятий, торговые центры), бетонные и заасфальтированные участки промплощадок, парковок, взлетно-посадочных полос аэропортов, а также карьеры и отсыпка строительных площадок.

Изначально было создано по 200 эталонных полигонов для каждого из пяти классов (т. е. 1800 пикселей данных Landsat), насколько возможно равномерно распределенных по исследуемой территории. После ряда тестовых классификаций и последующих корректировок на основании выявляемых ошибок, общий объем обучающей выборки составил 1328 полигонов, т. е. 11952 пикселя Landsat. Не смотря на различное количество эталонов для разных классов, обучающая выборка балансируется на этапе классификации и составляет 1000 пикселей на каждый класс.

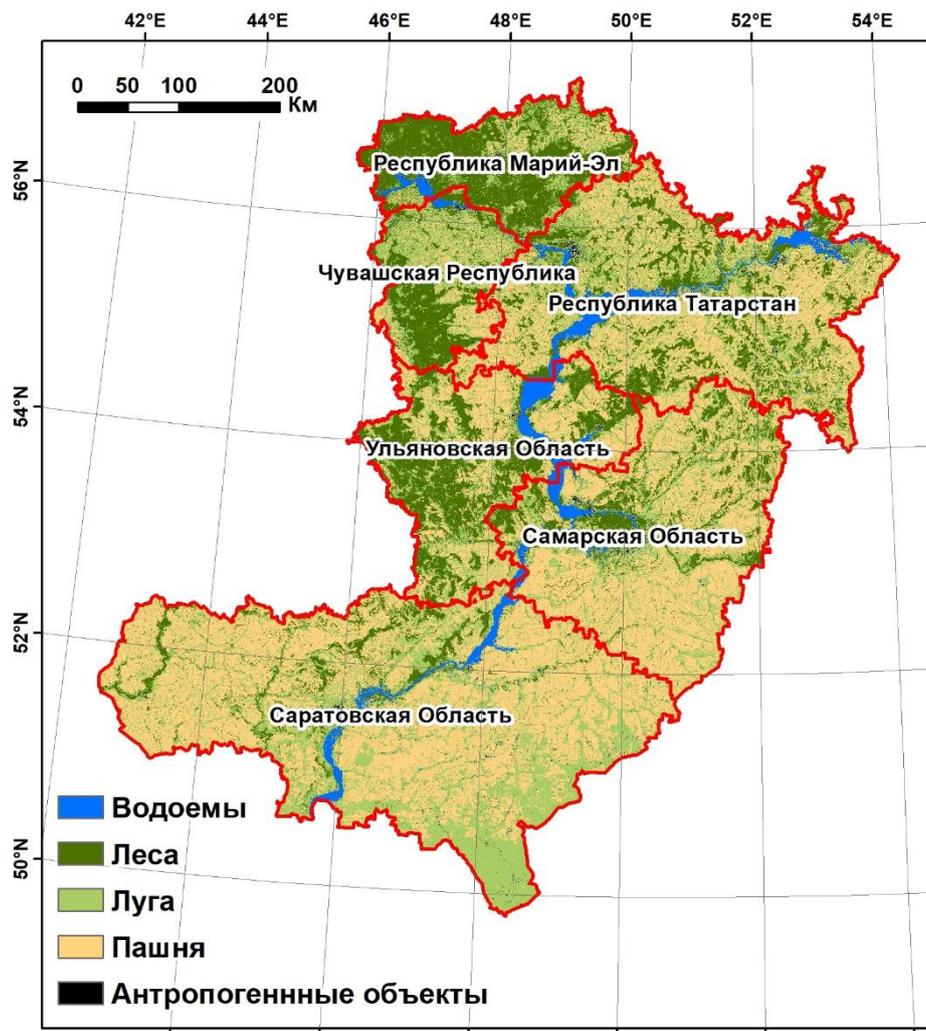
Классификация походилась алгоритмом RandomForest с 30 деревьями. Для оценки точности использовались 500 эталонных пикселей на каждый класс, не участвовавшие в обучении модели. Бала построена матрица ошибок. Рассчитаны ошибки оmissии и комиссии, F1-меры, коэффициент Каппа, общая точность (табл. 1).

Таблица 1

Точность распознавания

| Класс | Validation F1 |
|---|----------------------|
| Водоемы | 1 |
| Леса | 0,998 |
| Луга | 0,964 |
| Пашня | 0,963 |
| Антропогенные объекты | 0,995 |
| Validation kappa: 0,98 | |
| Validation overall accuracy: 0,984 | |

В результате получена растровая модель структуры земель с пространственным разрешением 30 м (рисунок) и рассчитаны площади основных классов землепользования/земного покрова (табл. 2).



Построенная модель землепользования/земного покрова

Таблица 2

Площадь основных типов землепользования и земного покрова

| Класс | Площадь, км ² | Доля, % |
|-----------------------|--------------------------|---------|
| Водоемы | 10312,87 | 3,4 |
| Леса | 66586,38 | 22,1 |
| Луга | 99952,32 | 33,1 |
| Пашня | 122900,93 | 40,7 |
| Антропогенные объекты | 2105,02 | 0,7 |
| Всего | 301857,51 | 100 |

Исследование выполнено при финансовой поддержке Российского научного фонда, проект № 23-27-00292, <https://rscf.ru/project/23-27-00292/>.

Библиографические ссылки

1. *Biging G., Colby D., Congalton R.* Sampling Systems for Change Detection Accuracy Assessment // *Remote Sensing Change Detection Environmental Monitoring Methods and Applications*. 1999. P. 281–308.
2. *Congalton R., Biging G.* A Pilot Study Evaluating Ground Reference Data Collection Efforts for Use in Forest Inventory // *Photogrammetric Engineering and Remote Sensing*. 1992. P. 1669–1671.
3. *Congalton R. G., Green K.* Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Third Edition, CRC Press, 2019.
4. *Hammond T. O., Verbyla D. L.* Optimistic Bias in Classification Accuracy Assessment // *International Journal of Remote Sensing*. 1996. Vol. 17, iss. 6. P. 1261–1266.
5. *Stehman S. V.* Basic Probability Sampling Designs for Thematic Map Accuracy Assessment // *International Journal of Remote Sensing*. 1999. Vol. 20, iss. 12. P. 2423–2441.