

К. Н. Дубровин^{1}, А. Н. Фролов², Д. А. Батяев¹, Т.Н. Федорова²*

Использование методов дистанционного зондирования Земли и машинного обучения для мониторинга пахотных земель Дальнего Востока

¹ Вычислительный центр ДВО РАН, г. Хабаровск,
Российская Федерация

² Дальневосточный научно-исследовательский институт сельского хозяйства,
с. Восточное, Российская Федерация

* e-mail: nob_keeper_93@mail.ru

Аннотация. На текущий момент существующие федеральные базы данных по землям сельскохозяйственного назначения (ЗСН) содержат неполные данные о пахотных землях. Использование методов дистанционного зондирования Земли (ДЗЗ) и машинного обучения позволяет как восстанавливать отсутствующие сведения о севообороте, так и оценивать площади посевных земель в оперативном режиме (по ходу сельскохозяйственного сезона). В работе предложена методика картографирования пахотных земель с использованием временных рядов NDVI (нормализованного относительного индекса вегетации) и метода гистограммного градиентного бустинга. Предложенная методика применена для построения карт пахотных земель Хабаровского района за 2021–2023 гг. и построения маски сои и залежи в 2019–2020 гг. Точность классификации в конце периода вегетации находилась на уровне 96–97 %, для ранней классификации варьировалась от 88 до 92,2 %. Полученные результаты могут использоваться органами исполнительной власти для комплексной оценки пахотных земель региона: уточнения сведений о севообороте, введения неиспользуемых земель в севооборот, прогнозирования урожайности сельскохозяйственных культур.

Ключевые слова: сельскохозяйственный мониторинг, дистанционное зондирование Земли, машинное обучение

К. Н. Dubrovin^{1}, A. N. Frolov², D. A. Batyaev, T. N. Fedorova²*

Use of remote sensing and machine learning methods for monitoring cropland in the Far East

¹ Computing Center FEB RAS, Khabarovsk, Russian Federation

² Far Eastern Agriculture Research Institute

* e-mail: nob_keeper_93@mail.ru

Abstract. Currently, the existing federal databases on agricultural land contain incomplete data on cropland. The use of remote sensing and machine learning methods allows both to restore missing information on crop rotation and to estimate the area of cultivated land in operational mode (in the course of the vegetation season). The paper proposes a methodology of cropland mapping using NDVI (Normalized Difference Vegetation Index) time series and histogram gradient boosting. The proposed methodology was applied to construct cropland maps of Khabarovsk district for 2021–2023 and a mask of soybean and fallow land in 2019–2020. The classification accuracy was 96–97 %, for early classification varied from 88 to 92,2 %. The obtained results can be used by executive authorities

for comprehensive assessment of cropland in the region: clarification of information on crop rotation, introduction of unused land into crop rotation, forecasting crop yields.

Keywords: agricultural monitoring, remote sensing, machine learning

Введение

Мониторинг сельскохозяйственных земель – система оперативных, периодических и базовых наблюдений за изменением качественного и количественного состояния земель сельскохозяйственного назначения [1]. Точное картографирование пахотных земель может дать важнейшую информацию для оценки производства сельскохозяйственных культур и продовольственной безопасности. Актуальность решения этой задачи напрямую связана как с необходимостью уточнения севооборота на отдельных полях, так и в целом с оценкой использования пахотных земель. Использование данных дистанционного зондирования Земли (ДЗЗ) позволяет снизить затраты и улучшить качество карт внутрихозяйственного землеустройства, инвентаризации земель, оценки общего состояния посевов сельскохозяйственных культур, контролировать выполнение посевных и уборочных работ.

Для исследования растительного покрова используются космические снимки различного масштаба, пространственного и временного разрешения. Основываясь на комбинировании функций яркостных значений объектов на снимках, полученных в определенных спектральных каналах, можно рассчитать индексные значения каждого пикселя исследуемой области, которые необходимы для проведения анализа и оценки исследуемого объекта по его изображениям на снимке. Спектральные индексы, используемые для оценки состояния растительности, называются вегетационными индексами.

В мире существует успешный опыт создания масштабных карт пахотных земель. Например, в США агентство NASS еще в 2008 г. создало карту пахотных земель с выделенными на ней отдельными культурами CDL (Cropland Data Layer) [2]. В Евросоюзе была разработана информационная система сельскохозяйственных полей LPIS [3]. В России функционирует единая федеральная система о землях сельскохозяйственного назначения (ЕФИС ЗСН), предоставляющая информацию о севообороте на сельскохозяйственных полях муниципальных образований Российской Федерации [4]. Выявлено, что в ЕФИС ЗСН полностью отсутствуют данные о культурах, произрастающих на полях Хабаровского края в 2021 и 2022 гг., а данные о севообороте за 2018–2020 гг. содержат значительное количество неточностей. При этом контуры сельскохозяйственных полей давно не актуализировались и в настоящее время их использование для задач мониторинга земель сельскохозяйственного назначения проблематично.

Исследования, посвященные созданию масок пахотных земель и сельскохозяйственных культур с использованием данных ДЗЗ, производились преимущественно в муниципальных образованиях европейской части РФ [5–8]. Отличные от западных регионов сортовой состав, сроки сева и уборки урожая затрудняют

использование результатов этих работ для регионов Дальнего Востока. Значительные пробелы в сведениях о сельскохозяйственных полях требуют создания моделей для построения региональных карт пахотных земель, учитывающих особенности климата и роста сельскохозяйственных культур на Дальнем Востоке.

Методы и материалы

Схема картографирования пахотных земель, предложенная в текущем исследовании, представлена на рис. 1. Основными этапами построения карт пахотных земель являются: получение и предобработка спутниковой информации, сбор информации о сельскохозяйственных угодьях, создание непрерывных временных рядов индекса вегетации, формирование датасетов, машинное обучение, формирование масок сельскохозяйственных культур, оценка точности классификации.

Разработанная схема была применена для распознавания культур в Хабаровском районе Хабаровского края. Расчет значений NDVI для построения временных рядов осуществлялся по снимкам спутников Sentinel-2A/2B с разрешением 10 метров. Для восстановления отсутствующих значений (ввиду облачных явлений) использовалась аппроксимация временных рядов с использованием двух первых членов ряда Фурье. В итоге все временные ряды композитов NDVI включали по 27 значений (с конца апреля по конец октября).

Для построения карт пахотных земель (с распознаванием посевов сои, зерновых, гречихи, кукурузы, а также неиспользуемых земель) в 2021–2023 гг. создано по два датасета: полный сезонный и сокращенный для решения задачи раннего (в июле, по окончании 29 календарной недели) распознавания культур. В качестве метода машинного обучения использовался гистограммный градиентный бустинг [9]. Для оценки качества распознавания культур производилась 10-кратная кросс-валидация.

Затем с использованием классификатора, обученного по данным за 3 года (2021–2023 годы), осуществлено ретроспективное распознавание сельскохозяйственных культур в два предшествующие года (2019 и 2020 год, когда достоверная информация о произрастающих культурах отсутствовала) в целях восстановления севооборота. Результаты распознавания применялись для оценки площади, занимаемых посевами сои, а также площади неиспользуемых в ходе сезона земель. Полученные результаты сравнивались с данными официальной статистики и значениями, внесенными в ЕФИС ЗСН.



Рис. 1. Схема построения карт пахотных земель с использованием данных ДЗЗ

Результаты

Общая точность кросс-валидации разработанной модели распознавания культур Хабаровского района с использованием сезонных рядов NDVI на основе гистограммного градиентного бустинга оказалась высокой: в 2021 г. она составила 97,3 %, в 2022 и 2023 г. – 96,3 %.

На рисунке 2 представлена созданная при использовании полных временных рядов NDVI карта пахотных земель для исследуемых полей Хабаровского края (Хабаровский район и группа полей в районе имени Лазо) в 2023 г.

Общая точность классификации при уменьшении количества наблюдений в рядах хоть и снизилась на несколько процентов, но тем не менее осталась на достаточно высоком уровне: 92,2 % в 2021 г., 88,6 % в 2022 г., 88 % в 2023 г.

При решении задачи восстановления сведений о севообороте в 2019–2020 гг. моделировалась ситуация ранней классификации пахотных земель. Была получена оценка для общего объема посевных площадей сои в Хабаровском районе – в оба исследуемые года она составляла около 9220 га. Эта площадь меньше оценки Росстата на 14 % и значительно меньше оценки ЕФИС (почти на 30 %). Оценочная площадь неиспользуемой пашни в Хабаровском районе составила 5132 га в 2019 г. и 4659 га в 2020 г. Несмотря на незначительное улучшение си-

туации в 2020 г., объем залежи превышает 25 %, что подтверждает проблемы с актуальностью существующих карт пахотных земель для Хабаровского района.

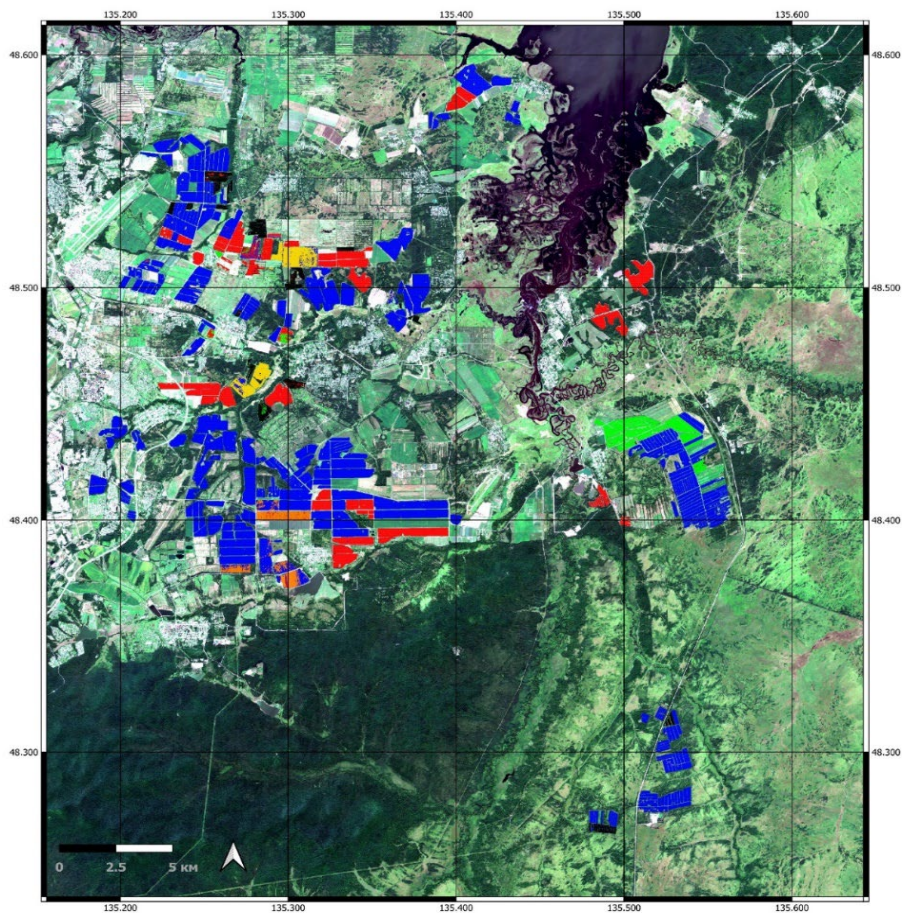


Рис. 2. Карта исследуемых сельскохозяйственных полей в 2023 г. с указанием распознанных культур (автоматизированная классификация с использованием полных сезонных временных рядов NDVI)

Заключение

Данные ДЗЗ обладают рядом преимуществ, среди которых можно выделить высокую достоверность и объективность получаемой информации, оперативность предоставления снимков, высокую периодичность получения информации на заданную территорию. Методы машинного обучения на основе деревьев решений позволяют строить быстрые и гибкие модели, позволяющие эффективно осуществлять распознавание сельскохозяйственных культур. Применение гистограммного градиентного бустинга в качестве метода классификации и сезонных рядов NDVI в качестве входных данных способствовало построению детальных, высокоточных (точность во все три года превысила 96 %) карт пахотных земель. Также доказана возможность раннего распознавания сельскохозяйственных культур в целях оперативного уточнения площадей посевов сельскохозяйственных культур – точность распознавания культур в июле находилась в районе 90 %.

При отсутствии данных о севообороте для обучения классификаторов могут использоваться данные за предыдущие/последующие сезоны. Созданные с использованием многолетних спутниковых данных маски позволяют уточнять количество задействованных в севооборотах полей и посевные площади сельскохозяйственных культур, определять состояние пахотных земель, оценивать их зарастание сорной растительностью. Поэтому предложенная методика и результаты ее применения могут эффективно применяться региональными министерствами в задачах сельскохозяйственного мониторинга.

Благодарности

Исследования выполнены при финансовой поддержке РФФ, проект № 23-76-00007 на тему «Разработка научных методов и подходов устойчивого управления почвенными ресурсами на основе технологий дистанционного зондирования Земли (на примере юга Дальнего Востока)».

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Земельный кодекс Российской Федерации. № 136-ФЗ от 25 октября 2001 года: [федер. закон: принят Гос. Думой 28 сент. 2001 г.] – Собрание законодательства Российской Федерации. – №44. – Ст. 4147.
2. Boryan, C. Monitoring US Agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program [Text] / C. Boryan, Z. Yang, R. Mueller, M. Craig // GeoCarto International. – 2011. – Vol. 26 (5). – P. 341–358.
3. Şimşek, F. Land cover classification using Land Parcel Identification System (LPIS) data and open source Eo-Learn library [Text] / F. F. Şimşek, S. S. Durduran // Geocarto International. – 2022. – Vol. 38(1). – P. 1–18.
4. Буланов, К.А. Блок работы с данными дистанционного зондирования Земли Единой федеральной информационной системы о землях сельскохозяйственного назначения [Текст] / К.А. Буланов, П.В. Денисов, Е.А. Лупян, А.С. Мартыанов, И.И. Середя, К.А. Трошко, В.А. Толпин, С.А. Барталев, С.А. Хвостиков // Современные проблемы дистанционного зондирования Земли из космоса. – 2019. – Т. 16. – № 3. – С. 171–182.
5. Балдина, Е.А. Картографирование современного состояния и многолетних изменений в использовании сельскохозяйственных земель в дельте Волги [Текст] / Е.А. Балдина, К.А. Трошко // Геодезия и картография. – 2016. – № 11. – С. 39–46.
6. Терехин, Э.А. Анализ спектрально-отражательных свойств залежных земель Среднерусской лесостепи по данным Sentinel-2 [Текст] / Э.А. Терехин // Компьютерная оптика. – 2023. – Т. 47, № 2. – С. 306–313.
7. Береза, О. В. О возможности прогнозирования урожайности озимой пшеницы в Среднем Поволжье на основе комплексирования наземных и спутниковых данных [Текст] / О. В. Береза, А. И. Страшная, Е. А. Лупян // Современные проблемы дистанционного зондирования Земли из космоса. – 2015. – Т. 12, № 1. – С. 18–30.
8. Кашницкий, А.В. Анализ возможности использования данных различного пространственного разрешения при проведении мониторинга объектов [Текст] / А. В. Кашницкий, Е. А. Лупян, Д. Е. Плотников, В. А. Толпин // Современные проблемы дистанционного зондирования Земли из космоса. – 2023. – Т. 20. – № 2. – С. 60–74.
9. Tanha, J. Boosting methods for multi-class imbalanced data classification: an experimental review [Text] / J. Tanha, Y. Abdi, N. Samadi, [et al] // Journal of Big Data. – 2020. – Vol. 7. – 70.

© К. Н. Дубровин, А. Н. Фролов, Д. А. Батяев, Т. Н. Федорова, 2024