



АНАЛИЗ ГОРОДСКИХ ЛЕСОВ: КЛАССИФИКАЦИЯ ПОРОД С ПРИМЕНЕНИЕМ МАШИННОГО ОБУЧЕНИЯ И ДАННЫХ ДЗЗ

М.В. Платонова¹, А.В. Кухарский¹, Е.Б. Таловская^{1,2}, Г.И. Лазоренко¹

¹Новосибирский государственный университет, ул. Пирогова, 1, Новосибирск, 630090, Россия,

²Центральный сибирский ботанический сад СО РАН, ул. Золотодолинская, 101, Новосибирск, 630090, Россия,

e-mail: gumoznaya@gmail.com

Эффективное управление городскими лесами требует комплексного подхода, начиная с полной инвентаризации их биоразнообразия. На сегодняшний момент данные о флористическом составе городских лесов в сибирских городах либо ограничены, либо фрагментарны. Цель данного исследования заключается в классификации городских лесов по породам и определение их онтогенетического состояния по материалам данных дистанционного зондирования. Данное исследование нацелено на глубокий анализ структуры городских лесов с использованием данных дистанционного зондирования, в частности использованию беспилотного летательного аппарата.

Машинное обучение, БПЛА, классификация леса

URBAN FOREST ANALYSIS: SPECIES CLASSIFICATION USING MACHINE LEARNING AND REMOTE SENSING DATA

M.V. Platonova¹, A.V. Kukharskii¹, E.B. Talovskaya^{1,2}, G.I. Lazorenko¹

¹Novosibirsk State University, Pirogova Str., 1, Novosibirsk, 630090, Russia,

²Central Siberian Botanical Garden of the Siberian Branch of the Russian Academy of Sciences,

Zolotodolinskaya Str., 101, Novosibirsk, 630090, Russia,

e-mail: gumoznaya@gmail.com

Effective management of urban forests requires an integrated approach, starting with a complete inventory of their biodiversity. At the moment, data on the floristic composition of urban forests in Siberian cities is either limited or fragmentary. The purpose of this study is to classify urban forests by species and determine their ontogenetic state using remote sensing materials. This study aims to deeply analyze the structure of urban forests using remote sensing data, in particular the use of unmanned aerial vehicles.

Machine learning, UAV, forest classification

ВВЕДЕНИЕ

Городские леса стали неотъемлемой частью городского окружения, принося много преимуществ в жизнь жителей. Однако городская среда, с ее агрессивностью, и чрезмерная рекреационная активность оказывают негативное воздействие на лесные массивы, влияя на их биологическое разнообразие и способствуя разрушению лесной экосистемы [Ontl et al., 2020].

Эффективное управление городскими лесами требует комплексного подхода, который начинается с полной инвентаризации биоразнообразия городских лесов. На сегодняшний момент данные о

флористическом составе городских лесов в сибирских городах либо ограничены, либо фрагментарны. Экологическая классификация леса становится ключевым шагом в создании базы данных, необходимой для долгосрочного мониторинга и устойчивого планирования ухода за городскими лесами.

Исследование экологических характеристик растений и их использование в ландшафтной архитектуре приобретают весомое значение в контексте стремления к углеродно-нейтральной среде в городах. Формирование зеленых насаждений, устойчивых к городским условиям, требует глубокого понимания экологических и биологических особенностей растений, особенно в условиях сибирского климата. Решение этой проблемы представляет собой не только вопрос эстетики, но и важный шаг в направлении устойчивого управления городской средой. Также автоматизация мониторинга леса – создаст фундамент для разработки моделей эмиссии и секвестрирования парниковых газов в лесных системах [Johnston, Withey, 2017; Cunliffe et al., 2020; Huang et al., 2021].

В современном контексте использование передовых методов машинного обучения и анализа данных дистанционного зондирования становится неотъемлемой составляющей в изучении городских лесов. Часто в таких задачах после сбора данных проводится анализ с применением методов машинного обучения для выделения и классификации различных пород деревьев [Johnston, Withey, 2017].

Актуальность данной задачи обусловлена не только необходимостью создания базы данных по биоразнообразию городских лесов, но и острой потребностью в создании геоинформационной системы, интегрирующей современные технологии с науками о биологии, экологии и т. д. [Chen, Guestrin, 2016; Zeng et al., 2020].

Данное исследование является не только хорошим примером автоматизации методов мониторинга леса, но и является частью инструмента для экологического планирования и управления городскими лесами в долгосрочной перспективе (этап предобработки данных).

МАТЕРИАЛЫ И МЕТОДЫ

Целью данного исследования является классификация городских лесов по породам и определение их онтогенетического состояния по материалам данных дистанционного зондирования (ДЗЗ). Эти лесные участки, несмотря на свою важность, подвергаются сильной антропогенной нагрузке, что отражается на их структуре и биоразнообразии. Данное исследование нацелено на глубокий анализ структуры городских лесов с использованием ДЗЗ, в частности при помощи использования беспилотного летательного аппарата (БПЛА). На первом этапе проводится сбор данных, охватывающих различные аспекты лесного покрытия. Далее эти данные используются для формирования детальных карт ландшафтных биоконплексов, отражающих многообразие и структуру растительности в городской среде. Однако, с увеличением объемов данных возникает необходимость в эффективных методах их обработки. Здесь ключевую роль играет машинное обучение – с его помощью можно автоматизировать процесс классификации биоконплексов. В современном мире применение методов машинного обучения, в том числе технология машинного зрения, становится неотъемлемой частью мониторинга лесов.

Использование технологий машинного зрения вносит существенный вклад в повышение точности при проведении классификации ландшафтных биоконплексов, это особенно ценно в контексте обработки информации о растительности в условиях городской среды, где точность и эффективность процесса классификации играют важную роль.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Для проведения анализа была выбрана территория леса в городе Новосибирске, тесно связанная с городской средой, и представляющая собой участок природного ландшафта в границах агломерации. Этот участок леса охватывает площадь в 17 гектар, представляя собой обширный фрагмент природы, взаимодействующей с городской инфраструктурой (тропы, окружен автомобильными дорогами).

Данные наблюдений были собраны при использовании БПЛА Geoscan 401 10S, оснащенного лазерным сканером Lidar AGM Systems. Лазерный сканер Lidar, работающий на основе принципа времени пролета лазерного излучения, позволяет создать точное трехмерное изображение поверхности и объектов. Параллельно с проведением лазерного сканирования, была использована камера Sony Alpha A6000 для создания фотографий лесного участка. Эти этапы сбора материалов необходимы для последующего создания ортофотоплана и генерации цифровых моделей рельефа (ЦМР).

Проводя пролеты на высоте 170 метров с помощью БПЛА, была охвачена вся территория леса, выделенная для проведения эксперимента. Такая высота обеспечила получение детальных данных о рельефе и структуре леса. Собранные данные представляют собой первичный материал – ряд фотографий с геокоординатной привязкой. Этот материал не является окончательным результатом, а представляет собой сырой базовый набор фотографий. Таким образом, фотографии, сделанные с определенной периодичностью, не предоставляют полного представления об объекте, так как некоторая информация на них может дублироваться с соседними изображениями. Более того, части территорий, запечатленные на различных фотографиях, могут накладываться друг на друга, что требует дополнительного анализа для корректного объединения и интерпретации данных.

Набор ряда фотографий с геокоординатной привязкой не только обеспечивают необходимую информацию для создания ортофотоплана, а также предоставляют дополнительную информацию, позволяя фиксировать одно и то же дерево или лесной участок на нескольких изображениях с разных точек, под разным углом. Такой подход обогащает процесс анализа, делая его более всесторонним и детализированным. Часто решение подобных задач осуществляется с применением технологий компьютерного зрения.



Рис. 1. Пример фотоснимка с БПЛА Geoscan 401 10S

В результате выполненного этапа сбора информации было получено 499 фотографий, пример одной фотографии представлен на рис. 1. Каждая из этих фотографий имеет геокоординатную привязку, и информацию о высоте, на которой сделан фотоснимок. Каждая фотография представляет собой детальное изображение территории размером примерно 160 метров на 110 метров.

На следующем этапе предобработки данных была применена процедура фрагментации изображения. Примерная сетка предполагающая разбиение фотографии на квадратные участки изображения показана на рис. 2. Важно отметить, что данная операция не включает в себя выделение индивидуальных объектов, она направлена на разделение изображения на части без привязки к конкретным объектам.

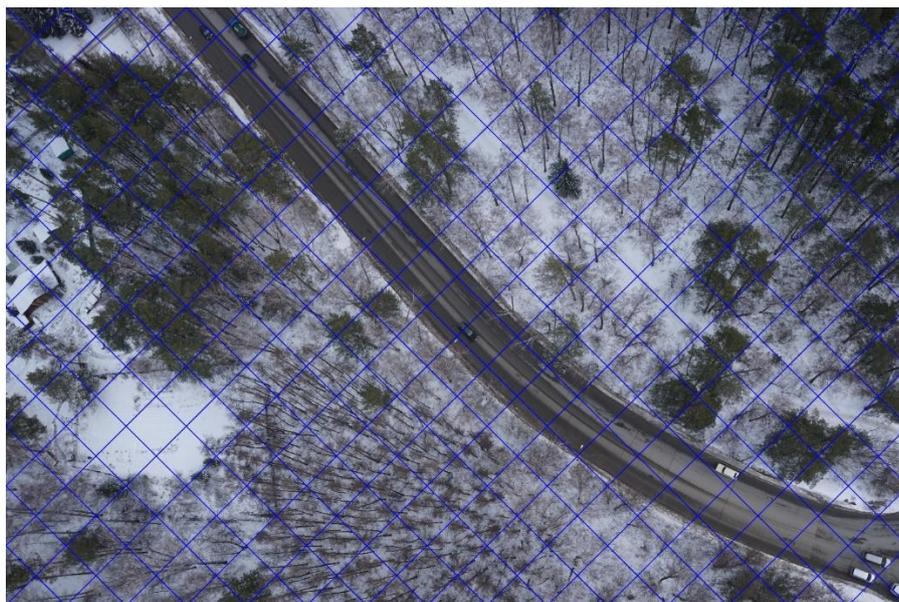


Рис. 2. Пример условной разметки, используемой для фрагментации изображения на этапе предобработки

Следующий этап – подготовка данных для обучения нейронной сети. На этом этапе происходит выбор и разметка отдельных участков большого изображения, выделенных в соответствии с различными породами деревьев и их онтогенетическим состоянием. Используя данные по онтогенезу различных видов деревьев: сосна *Pinus sylvestris* L., ель *Picea obovata* Ledeb., береза *Betula pendula* Roth [Работнов, 1956; Чистякова, 1989; Жукова, 2013], специалистами в области биоморфологии и популяционной биологии растений были выделены четыре группы деревьев. Отметим, что при выделении групп ученые опирались на возможность визуально различить данные группы деревьев по снимкам, информацию о породах деревьев и на их онтогенетическое состояние.

В рамках данного эксперимента были выделены следующие группы деревьев:

1. Зрелые, старые и отмирающие сосны.

Категории сосен, включая зрелые, старые и находящиеся в состоянии отмирания, представлены различными состояниями согласно ботанической классификации [Работнов, 1956]. Это обеспечивает ясность определения данной группы деревьев. Такой подход позволяет нам эффективно осуществлять анализ данных в рамках нашего исследования, сохраняя в то же время понимание разнообразных онтогенетических состояний сосен. Пример зрелых и отмирающих сосен представлен на рис. 3.



Рис. 3. Пример группы: Отмирающие и зрелые сосны

2. Молодые сосны.

Молодые сосны представляют собой группу деревьев, находящихся на ранних стадиях своего развития. В рамках данного исследования мы рассматриваем эту группу, включая в нее подрост и молодые генеративные растения. Пример данной группы представлен на рис. 4.

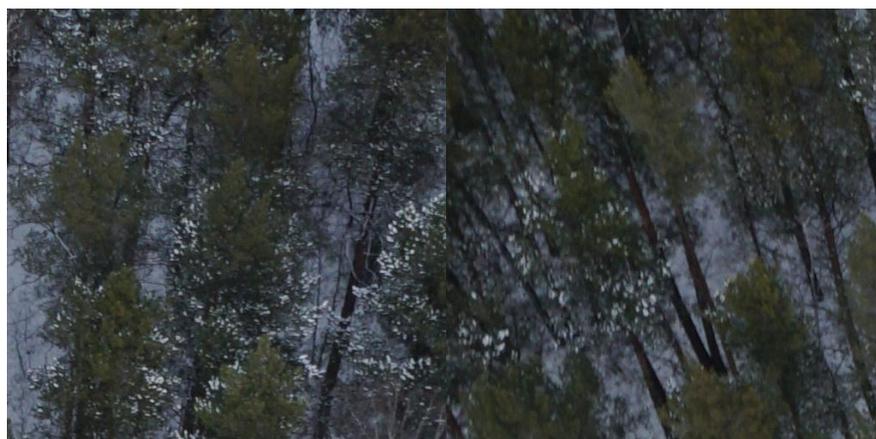


Рис. 4. Пример группы: Молодая сосна

3. Ель. В данной группе учтены деревья, соответствующие виду "ель". Пример приведен на рис. 5.

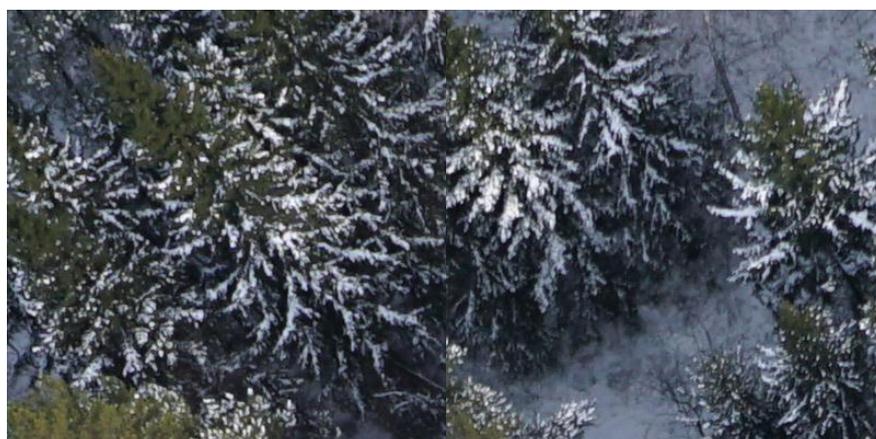


Рис. 5. Пример группы: Ель

4. Береза: Группа деревьев "береза". Пример приведен на рис. 6.



Рис. 6. Пример группы: Береза

Эти условные обозначения были введены исключительно в рамках данной задачи для удобства идентификации пород и состояния деревьев, отраженных на изображениях. Исходным материалом для обучения являлись изображения размером 400 на 400 пикселей.

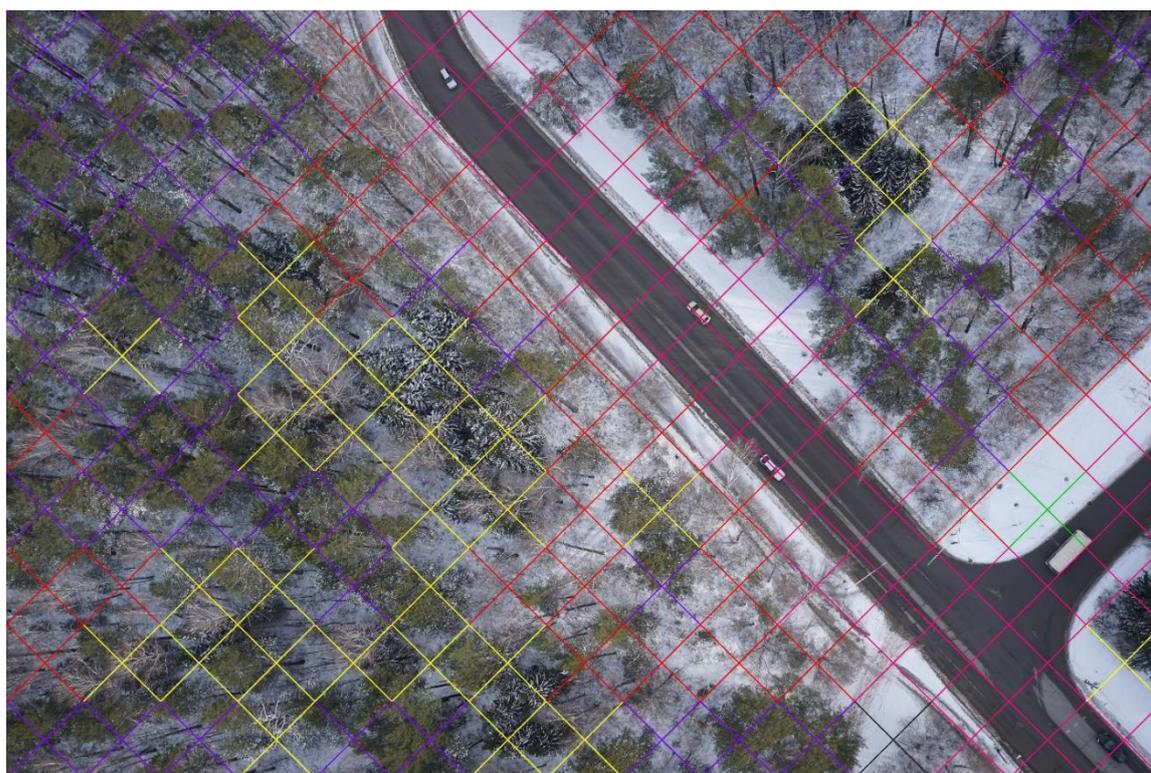


Рис. 7. Результат разметки исходной фотографии по породам

Следующим этапом проведения эксперимента было переобучение нейронной сети VGG16. Обучение проводилось на наборе из 100 изображений, с валидацией на 50 изображениях. Этап валидации был особенно важен, поскольку он позволил оценить точность модели на данных, не использованных в процессе обучения. Результаты валидации свидетельствуют о высокой точности модели на уровне 90 %.

После успешного обучения модели мы применили ее к неразмеченным изображениям, полученным с помощью БПЛА. Результаты обработки представлены на рис. 7. В соответствии с заложенными в эксперимент группами деревьев, на изображениях были выделены разным цветом группы деревьев, что наглядно разбивает фотоснимок по видам. В частности, береза обозначена зеленым цветом, отмирающая сосна – синим, молодая и зрелая сосна – фиолетовым, а ель – желтым. Важно отметить, что на изображениях также выделена дорога, обозначенная розовым цветом.

Также был проведен сравнительный эксперимент на том же изображении, но с измененным размером сетки фрагментаций фотоснимков. Результаты эксперимента с более детальным разбиением представлены на рис. 8.

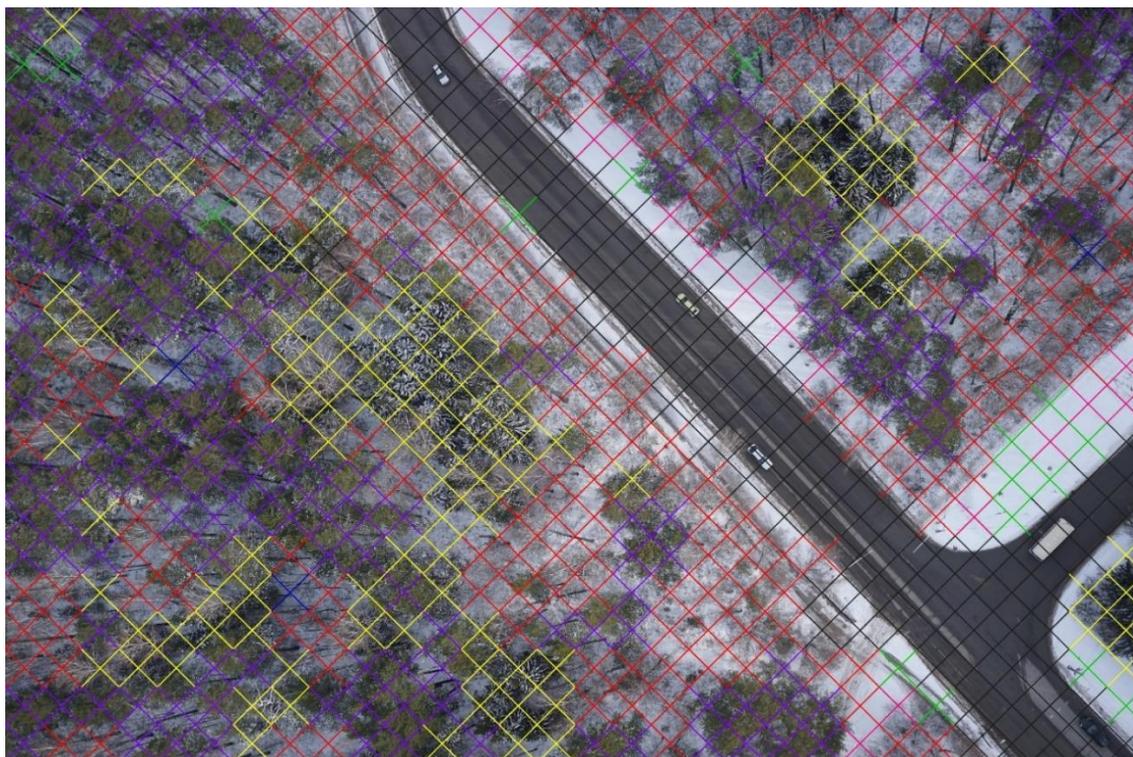


Рис. 8. Применение модели к снимку с разбиением по более мелкой сетке

ВЫВОДЫ

В ходе экспериментов был проведен сравнительный анализ результатов использования разных масштабов сеток разбиений в процессе обработки изображений. При этом стоит отметить, что при сборе данных необходимо подбирать высоту пролета БПЛА, что это дает возможность снимать с определенным масштабом. Фактически, выбор высоты съемки является настройкой точности и подробности изображений на снимках. Это позволяет выбирать оптимальную сетку разбиений в зависимости от имеющихся фотографий для эксперимента.

Степень разбиения изображений имеет значительное влияние на производительность и точность модели. Наши эксперименты показали, что при подходящем подборе параметров, включая высоту пролета БПЛА, можно добиться оптимальной сетки разбиений, обеспечивающей наилучшие результаты классификации пород деревьев. Это предоставляет дополнительные инструменты для настройки методики в зависимости от конкретных условий съемки и целей исследования.

ЗАКЛЮЧЕНИЕ

Важно отметить, что подобный подход к оптимизации сетки разбиений открывает перспективы для дальнейшего повышения эффективности модели и ее применимости в различных условиях городской среды. Наши результаты подчеркивают важность тщательного анализа и подбора параметров для достижения оптимальных результатов в задачах машинного обучения на основе данных ДЗЗ в области городской экологии.

Эти результаты подтверждают не только высокую точность модели, но и ее способность эффективно классифицировать разнообразные породы деревьев и выделять особенности окружающей среды. Данная методика предоставляет ценные инструменты для мониторинга и управления городскими лесами, а также может быть использована в различных экологических и научных исследованиях.

Работа была выполнена в рамках государственного задания № FSUS-2023-0001 «Создание геоинформационной системы изучения ландшафтных биоконплексов на основе данных дистанционного зондирования Земли с использованием методов машинного обучения».

СПИСОК ЛИТЕРАТУРЫ

- Жукова Л.А.** Ontogenesis of *Pinus sylvestris* L. Онтогенетический атлас растений. – Йошкар-Ола: Марийский гос. ун-т., 2013. – Том 7. – С. 26–65.
- Работнов Т.А.** Жизненный цикл многолетних травянистых растений в луговых ценозах // Труды БИН АН СССР. – М., Л., 1950. – Вып. 6. – С. 179–196.
- Чистякова А.А.** Ontogenesis of *Betula pendula* Roth. Диагнозы и ключи возрастных состояний лесных растений // Деревья и кустарники / Под ред. О.В. Смирновой. – М.: Прометей, 1989. – С. 89–96.
- Chen T., Guestrin C.** Xgboost: A scalable tree boosting system // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – 2016. – P. 785–794, doi: 10.1145/2939672.2939785.
- Cunliffe A.M., Assmann J.J., Daskalova G.N., Kerby J.T., Myers-Smith I.H.** Aboveground biomass corresponds strongly with drone-derived canopy height but weakly with greenness (NDVI) in a shrub tundra landscape // Environmental Research Letters. – 2020. – Vol. 15. – Article 125004, doi: 10.1088/1748-9326/aba470.
- Huang S., Tang L., Hupy J.P., Wang Y., Shao G.** A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing // Journal of Forestry Research. – 2021. – Vol. 32 (5). – P. 1–6, doi: 10.1007/s11676-020-01155-1.
- Johnston C.M.T., Withey P.** Managing forests for carbon and timber: a Markov decision model of uneven-aged forest management with risk // Ecological Economics. – 2017. – Vol. 138. – P. 31–39, doi: 10.1016/j.ecolecon.2017.03.023.
- Ontl T.A., Janowiak M.K., Swanston C.W., Daley J., Handler S., Cornett M., Hagenbuch S., Handrick C., McCarthy L., Patch N.** Forest management for carbon sequestration // Journal of Forestry. – 2020. – Vol. 118 (1). – P. 86–101, doi: 10.1093/jofore/fvz062.
- Zeng J., Matsunaga T., Tan Z.-H., Saigusa N., Shirai T., Tang Y., Peng S., Fukuda Y.** Global terrestrial carbon fluxes of 1999–2019 estimated by upscaling eddy covariance data with a random forest // Scientific Data. – 2020. – Vol. 7 (1). – Article 313, doi: 10.1038/s41597-020-00653-5.

REFERENCES

- Chen T., Guestrin C.** Xgboost: A scalable tree boosting system // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – 2016. – P. 785–794, doi: 10.1145/2939672.2939785.
- Chistyakova A.A.** Ontogenesis of *Betula pendula* Roth. Diagnosis and keys of age condition of forest plants // Trees and bushes [in Russian]. – Prometei, Moscow, 1989. – P. 89–96.
- Cunliffe A.M., Assmann J.J., Daskalova G.N., Kerby J.T., Myers-Smith I.H.** Aboveground biomass corresponds strongly with drone-derived canopy height but weakly with greenness (NDVI) in a shrub tundra landscape // Environmental Research Letters. – 2020. – Vol. 15. – Article 125004, doi: 10.1088/1748-9326/aba470.
- Huang S., Tang L., Hupy J.P., Wang Y., Shao G.** A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing // Journal of Forestry Research. – 2021. – Vol. 32 (5). – P. 1–6, doi: 10.1007/s11676-020-01155-1.
- Johnston C.M.T., Withey P.** Managing forests for carbon and timber: a Markov decision model of uneven-aged forest management with risk // Ecological Economics. – 2017. – Vol. 138. – P. 31–39, doi: 10.1016/j.ecolecon.2017.03.023.
- Ontl T.A., Janowiak M.K., Swanston C.W., Daley J., Handler S., Cornett M., Hagenbuch S., Handrick C., McCarthy L., Patch N.** Forest management for carbon sequestration // Journal of Forestry. – 2020. – Vol. 118 (1). – P. 86–101, doi: 10.1093/jofore/fvz062.
- Rabotnov T.A.** Life cycle of perennial grasses in meadow coenoses // Proceedings BIN AN SSSR [in Russian]. – Moscow, Leningrad, 1950. – Vol. 6. – P. 179–196.
- Zeng J., Matsunaga T., Tan Z.-H., Saigusa N., Shirai T., Tang Y., Peng S., Fukuda Y.** Global terrestrial carbon fluxes of 1999–2019 estimated by upscaling eddy covariance data with a random forest // Scientific Data. – 2020. – Vol. 7 (1). – Article 313, doi: 10.1038/s41597-020-00653-5.
- Zhukova L.A.** Ontogenesis of *Pinus sylvestris* L. Ontogenetic atlas of plants [in Russian]. – Yoshkar-Ola, 2013. – Vol. 7. – P. 26–65.

КОРОТКО ОБ АВТОРАХ

ПЛАТОНОВА Марина Владимировна – младший научный сотрудник НГУ. Основные научные интересы – анализ данных ДЗЗ.

КУХАРСКИЙ Александр Витальевич – младший научный сотрудник НГУ. Основные научные интересы – машинное обучение.

ТАЛОВСКАЯ Евгения Борисовна – кандидат биологических наук, старший научный сотрудник НГУ. Основные научные интересы: биоморфология и популяционная биология растений.

ЛАЗОРЕНКО Георгий Иванович – кандидат технических наук, научный сотрудник НГУ. Основные научные интересы: науки о Земле и смежные экологические науки.

Статья поступила 14 декабря 2023,
принята к печати 22 декабря 2023