

ПРОСТРАНСТВЕННО-РАСПРЕДЕЛЕННЫЙ МОНИТОРИНГ НАЗЕМНЫХ ЭКОСИСТЕМ РОССИИ

Гушан Н.Ю., Кочкаров Р.А.

Финансовый университет при Правительстве РФ, Москва, Россия
249351@edu.fa.ru, rkochkarov@fa.ru

Аннотация. Исследуются методы, модели и алгоритмы пространственно-распределённого мониторинга наземных экосистем России с применением спутниковых данных, ГИС и алгоритмов машинного обучения. Результатом применения представленного инструментария являются карты углеродного баланса с высокой детализацией для последующей оценки состояния экосистем.

Ключевые слова: мониторинг экосистем, пространственный анализ, углеродный цикл, дистанционное зондирование, машинное обучение.

Введение

Наземные экосистемы России играют ключевую роль в глобальном углеродном цикле, аккумулируя более 46 млрд тонн углерода [1]. Их пространственное распределение крайне неоднородно, а воздействие климатических и антропогенных факторов требует регулярного мониторинга. Современные вызовы в области климата и устойчивого природопользования ставят задачу создания масштабируемой системы пространственного мониторинга, учитывающей локальные особенности ландшафта, лесных типов и климатических условий [2].

Использование методов машинного обучения в геопространственном анализе позволяет учитывать широкий спектр факторов, влияющих на процесс депонирования углерода, включая климатические условия, характеристики рельефа, почвенные свойства, продуктивность лесов и степень антропогенного воздействия. Дистанционное зондирование Земли, в свою очередь, предоставляет огромные объемы данных, которые могут быть использованы для построения прогнозных моделей и принятия обоснованных решений в области природопользования и климатической политики. Такой подход не только повышает эффективность мониторинга лесных экосистем, но и способствует разработке стратегий их сохранения и восстановления. Полученные результаты могут быть использованы в практике лесного хозяйства, экологического планирования и оценки углеродных единиц в рамках международных климатических инициатив.

1. Методологический подход

1.1. Цель мониторинга

Методологический подход к решению этой задачи базируется на интеграции космических данных, геоинформационных систем и интеллектуальных алгоритмов обработки. Спутниковое дистанционное зондирование позволяет регулярно получать информацию о состоянии растительности и структуре лесного покрова по всей территории страны [3].

Современные методы машинного обучения, такие как ансамбли решающих деревьев, градиентный бустинг и др. позволяют моделировать взаимосвязи между спектральными признаками, климатическими переменными и запасами углерода [4].

Цель – создание интеллектуальной системы, обеспечивающей регулярное обновление пространственно-распределённых данных о состоянии экосистем и углеродных потоках на территории России.

1.2. Пространственная детализация

Для анализа применяются спутниковые данные (Sentinel-2, Landsat, MODIS) с разрешением от 10 до 30 метров. Это позволяет учитывать мелкомасштабные различия экосистем (типы леса, продуктивность, воздействие пожаров и вырубок).

Космические мультиспектральные служат основой для вычисления вегетационных индексов и прямого оценивания состояния растительности. Например, серию снимков Sentinel-2 (10–20 м разрешения) можно использовать для расчёта NDVI и других индексов, коррелирующих с биомассой и продуктивностью леса [5].

Спутниковые радарные данные дают информацию о структурных характеристиках леса (высота кроны, биомасса), поскольку микроволновые волны проникают в полог и не зависят от облачности.

Особую ценность представляет мультичастотный радар (L- и С-диапазоны), позволяющий оценивать как общий объем биомассы, так и тип древесного покрова [6].

Следующим сервисом, который предоставляет данные со спутника, является Google Earth Engine. У данного сервиса есть как веб-версия, так и специальное приложение для ПК. Их отличие заключается в том, что в версии для ПК есть огромное количество дополнительных возможностей.

GEE – это платформа облачных вычислений, которая может извлекать параметры (например, нормализованный разностный индекс растительности или NDVI) из спутниковых изображений. Затем GEE может визуализировать данные, отображать их в виде графиков временных рядов, а данные можно загружать для внешней обработки [7].

Лидарные (LiDAR) данные – спутниковые или воздушные/беспилотные лазерные сканирования обеспечивают трехмерную структуру леса: высоты деревьев, вертикальный профиль. Эти данные дают наиболее точную оценку биомассы на участке, дополняя спутниковые снимки. Успешные задачи построения карт биомассы часто опираются на ограниченные LiDAR-измерения для калибровки моделей.

Полевые данные лесного инвентаря – статистика и наблюдения по лесоустроительным пробным площадям, которую собирает Всероссийский лесной кадастровый. Эти данные содержат измерения запасов древесины, возраста насаждений и других характеристик леса, служат в качестве опорных (учебных) для обучения моделей. Например, анализ первого цикла Всероссийского учёта показал более точную оценку углеродного запаса ($\pm 1\%$) на основе выборочных полевых данных.

Станции экосистемного обмена (eddy-covariance) – точечные измерения потоков CO_2 , CH_4 , H_2O между лесом и атмосферой. Эти данные не масштабируются напрямую на всю территорию, но ценные для калибровки и валидации моделей потоков углерода на уровне биоценозов.

Климатические и географические данные – температура, осадки (метеостанции, ERA5-реконструкция), рельеф (цифровая модель высот), почвенные карты и другие переменные используются в сочетании с лесными показателями в моделях. Эти данные учитываются как вспомогательные факторы при обучении моделей ML.

2. Источники и обработка данных

Используются:

- **Оптические спутниковые снимки:** Sentinel-2, Landsat 8;
- **Радарные данные:** Sentinel-1, ALOS PALSAR;
- **Лидарные измерения:** GEDI, ICESat-2 (локальная калибровка);
- **Платформы:** Google Earth Engine, QGIS, SNAP.

Предобработка включает калибровку, облачную маску, расчёт NDVI, EVI, NIRv, а также извлечение рельефа и почвенных карт.

Методы количественной оценки углерода и потоков парниковых газов позволяют разделить на несколько категорий.

Измерение изменения запасов углерода – косвенный метод, при котором по разности оценок биомассы (и/или почвенного углерода) в разное время судят об абсорбции/выбросах. Этот подход базируется на лесном инвентаре, спутниковых оценках биомассы и моделях накопления. Он широко используется в учёте углеродных стоков (метод IPCC). Основное достоинство – связь с лесным хозяйством и отчётностью, однако точность сильно зависит от качества картирования биомассы и учёта почвенного углерода.

Прямые измерения потоков используют станций eddy-covariance для регистрации потоков CO_2 , CH_4 , водяных паров на выбранных участках. Эти данные позволяют глубоко исследовать экосистемные процессы, но имеют ограниченную пространственную представленность (несколько десятков станций на всю Россию) и не дают интегральных карт без дополнительного масштабирования.

Вегетационные индексы представляют собой простейший метод косвенной оценки биомассы и продуктивности: сочетания отражательной способности растительности в разных каналах превращаются в индексы, коррелирующие с LAI/биомассой. Широко используются из-за простоты вычислений и устойчивости к помехам. Недостаток – быстрое насыщение при высокой лиственной биомассе и влияние фоновых факторов (почва, дым). Тем не менее, множество работ показывают достаточную корреляцию этих индексов с полевыми измерениями углерода.

Модели светопроизводительности базируются на концепции, что валовая первичная продукция (GPP) пропорциональна поглощённой фотосинтетически активной радиации, умноженной на эффективность использования света. Классический пример – глобальная карта GPP по алгоритму

MOD17 (NASA MODIS) в режиме реального времени. Такие модели учитывают сезонность LAI/NDVI и данные об освещенности и погоде, но требуют параметризации по типам экосистем и часто не дают детализации по видам леса.

Процессно-ориентированные экосистемные модели (Biome-BGC, Century, ORCHIDEE и др.) – вычисляют углеродный баланс на основе биофизических уравнений, имитируя фотосинтез, дыхание, разложение опада. Эти модели могут прогнозировать динамику углерода, но чувствительны к многим параметрам и не всегда способны учесть земную неоднородность без калибровки данными ДЗЗ.

Статистические и ML-модели используют эмпирический подход: по данным наземных учётов и спутниковым признакам строятся регрессионные или обучаемые модели (линейная регрессия, случайный лес, градиентный бустинг и пр.) для предсказания биомассы/углерода. Преимущества машинного обучения – высокая гибкость и точность при наличии репрезентативной обучающей выборки. Недостатки – «чёрный ящик» без физической интерпретации, зависимость от качества и репрезентативности тренировочных данных, а также риски переобучения. Используются методы, включая «случайные леса» (Random Forest), градиентный бустинг и искусственные нейронные сети. Эти модели позволяют анализировать большие объемы данных, включая спутниковые снимки, полевые измерения и климатические данные.

3. Модели и алгоритмы

Пространственное моделирование реализовано с использованием [8]:

- **XGBoost** – для прогнозирования углеродных запасов по признакам вегетации и окружающей среды;
- **Random Forest** – для классификации типов лесов;
- **ГИС-алгоритмы** – для мозаики данных, ресэмплинга и агрегации.

Валидация выполнялась по данным инвентаризации и флюкс-станциям. R^2 моделей достигает 0,75 – 0,80 при оценке надземной биомассы.

Учитывая обзор существующих методик, оптимальный подход для пространственного мониторинга углерода в лесах России строится на машинном обучении с применением спутниковых данных и индикаторов вегетации [9]. В частности, ансамблевые алгоритмы (Random Forest, XGBoost) показали высокую эффективность в оценке биомассы и структурных характеристик леса. Был выбран XGBoost как широко применяемый в экологических задачах инструмент (включающий регуляризацию для предотвращения переобучения).

В качестве входных признаков предлагается использовать многоканальные космические данные (с красным, NIR, SWIR-диапазонами), рассчитать серию вегетационных индексов, а также учесть дополнительные факторы (рельеф, климат). Такой многофакторный набор показателей позволит модели учесть неоднородность лесных условий. Данные обучающей выборки будут формироваться из сочетания полевых измерений (инвентарных и флюкс-станций) и существующих карт лесной биомассы [10].

Для практической реализации желательно использовать облачные вычисления и инструменты Big Data. Такой подход обеспечивает масштабируемость и возможность регулярного обновления карт углерода. В сумме выбранный метод – машинное обучение на основе многоканальных спутниковых и геопризнаковых данных – обещает оптимальное соотношение точности, устойчивости и вычислительной эффективности для поставленной задачи.

4. Результаты и применения

Изучение показателей количества фотосинтетически активной биомассы имеет важное значение для оценки состояния растительности. Спутниковые данные предоставляют уникальную возможность наблюдения за биомассами на больших территориях. Однако данные, полученные из космоса, часто содержат искажения, вызванные различными факторами, такими как атмосферные условия, угол съёмки, топографические особенности и т. д.

Минимальные значения индекса растительности с высокой вероятностью являются ложными показаниями. Этот факт вызван перекрытием облаками, тенями от них, низкой освещённостью, проблемами, связанными с перепадами в насыщенности, контрастностью фотоснимков и т. п.

Метод сглаживания медианным фильтром и скользящим средним зачастую применяется для устранения шума. Медианный фильтр заменяет каждую точку данных медианой из набора соседних точек, данный метод сглаживания является нелинейным процессом, а значит хорошо сохраняет края и структуру сигнала при подавлении шума. Метод скользящего среднего подразумевает замену каждого

значения во временном ряду средним значением соседних данных внутри заданного промежутка. Применение медианного фильтра в первую очередь поможет устраниить импульсный шум и сохранить важные характеристики сигнала, такие как резкие перепады или пики. После этого использование скользящего среднего может помочь сгладить значения временного ряда, уменьшив влияние случайных колебаний и выделив основную тенденцию временного ряда.

Важным этапом в обработке данных является работа с пропущенными значениями. Пропуски могут существенно исказить результаты моделирования, и поэтому их необходимо корректно обработать. Заполнение пропущенных значений средними значениями, медианой или использование более сложных методов множественной импутации может обеспечить сохранение полноты данных.

Нормализация числовых признаков также играет важную роль. Она позволяет уравновесить влияние различных параметров, создавая стандартные условия для обучения модели. Min – Max нормализация является одним из распространенных методов, приводящих значения признаков к определенному диапазону.

Обработка выбросов – еще один важный шаг. Выбросы могут существенно повлиять на обучение модели, их выявление и коррекция с использованием таких показателей, как межквартильный диапазон, помогает обеспечить стабильность модели.

Учет взаимосвязей между признаками – это еще одна важная часть процесса подготовки данных. Анализ корреляции между различными характеристиками древостоя позволяет выявить зависимости, что помогает предотвратить мультиколлинеарность и лучше понять влияние различных параметров.

И, наконец, полезным может быть увеличение объема данных. Аугментация данных или создание синтетических данных помогает улучшить обобщающую способность модели, что особенно важно при ограниченном объеме исходных данных.

Также можно использовать системы компьютерного зрения для исключения облачных снимков из статистики. В таком случае лучший результат показывают модели на базе сверточных нейронных сетей.

Таким образом, внимательная обработка и подготовка данных представляют собой неотъемлемые этапы в построении надежных и точных моделей для исследований древостоев, обеспечивая более глубокий анализ и лучшее понимание экосистем.

5. Заключение

Пространственно-распределенный мониторинг экосистем России – ключевой элемент климатической политики и природоохранного планирования. Интеграция спутниковых данных, машинного обучения и ГИС позволяет создать надежную систему оценки и прогноза углеродного состояния территорий. Предложенный подход может стать основой для национальной платформы мониторинга экосистем и углеродного баланса.

Использование перечисленных технологий обеспечивает надежную информационную среду: от сбора исходных данных до вычислительной платформы для моделей в рамках существующей методологии пространственно-распределенных информационных систем мониторинга [11-17]. Такой технологический стек соответствует мировым практикам в области дистанционного мониторинга лесов и позволяет реализовать задуманный интеллектуальный подход.

Литература

1. *Filipchuk A.N. Carbon stock in living biomass of Russian forests: new quantification based on data from the first cycle of the State Forest Inventory / A.N. Filipchuk, N.V. Malysheva, T.A. Zolina, A.A. Seleznev // Central European Forestry Journal. – 2023. – Vol. 69, № 4. – P. 248–261.*
2. *Лукина Н.В. Роль старовозрастных лесов в аккумуляции и хранении углерода / Н.В. Лукина, С.А. Барталев, А.П. Гераськина [и др.] // Известия Российской академии наук. Серия географическая. – 2023. – Vol. 87, № 4. – P. 536–557.*
3. *Liu P. Combining Sentinel-2 and diverse environmental data largely improved aboveground biomass estimation in China's boreal forests / P. Liu, C. Ren, X. Yang et al // Scientific Reports. – 2024. – № 14. – Article number: 27528.*
4. *Illarionova S. Advancing Forest carbon stocks' mapping using a hierarchical approach with machine learning and satellite imagery / S. Illarionova, P. Tregubova, I. Shukhratov et al // Scientific Reports. – 2024. – № 14. – Article number: 21032.*
5. Статья 2 Парижского соглашения https://unfccc.int/sites/default/files/russian_paris_agreement.pdf (дата обращения 03.06.2025).
6. UNEP. <https://www.unep.org/explore-topics/environmental-rights-and-governance/what-we-do/monitoring-environment> (дата обращения 03.06.2025).

7. *Jin H.* Assessing the effect of ensemble learning algorithms and validation approach on estimating forest aboveground biomass: a case study of natural secondary forest in Northeast China / H. Jin, Y. Zhao, U. Pak, Z. Zhen, K. So // *Geo-Spatial Information Science*. – 2024. – Vol. 27, № 1. – Article number: 231126.
8. *Cheng F.* Remote Sensing Estimation of Forest Carbon Stock Based on Machine Learning Algorithms / F. Cheng, G. Ou, M. Wang, C. Liu // *Forests*. – 2024. – Vol. 15, № 4. – Article number: 681.
9. *Justice C.O.* An overview of MODIS Land data processing and product status / C.O. Justice, J. Townshend, E. Vermote, Masuoka Ed et al // *Remote Sensing of Environment*. – 2002. – Vol. 83, № 1–2. – P. 3–15.
10. *Gorelick N.* Google Earth Engine: Planetary-scale geospatial analysis for everyone / N. Gorelick, Matt Hancher, Mike Dixon et al // *Remote Sensing of Environment*. – 2017. – Vol. 202. – P. 18–27.
11. *Кочкаров Р.А.* Модель пространственно-распределенной информационной системы непрерывного мониторинга с предфрактальной динамической структурой в условиях воздействия дестабилизирующих факторов / Р.А. Кочкаров, Д.С. Чиров, А.В. Тимошенко, А.М. Казанцев // *T-Comm: Телекоммуникации и транспорт*. – 2025. – Т. 19, № 1. – С. 4–12.
12. *Кочкаров Р.А.* Алгоритм оценки структурно-функциональной устойчивости и целостности гетерогенной сети передачи данных пространственно-распределенной системы мониторинга / Р.А. Кочкаров, М.Т. Балдычев, А.М. Казанцев [и др.] // *Труды МАИ*. – 2024. – № 137.
13. *Шевцов В.А.* Показатель структурной эффективности управления информационным взаимодействием в гетерогенной сети передачи данных пространственно-распределенной системы мониторинга / В.А. Шевцов, А.М. Казанцев, А.В. Тимошенко [и др.] // *Вестник Воронежского государственного технического университета*. – 2024. – Т. 20, № 2. – С. 124–131.
14. *Шевнина Ю.С.* Подходы к прогнозированию изменения состояния обеспечивающих компонентов информационно-управляющей системы / Ю.С. Шевнина, П.Е. Рябов, С.В. Прокопчина, Р.А. Кочкаров // *Моделирование, оптимизация и информационные технологии*. – 2024. – Т. 12, № 2(45). – DOI: 10.26102/2310-6018/2024.45.2.023.
15. *Кочкаров Р.А.* Теоретико-графовый алгоритм динамического назначения средств системы непрерывного мониторинга / Р.А. Кочкаров, А.А. Кочкаров // *Успехи современной радиоэлектроники*. – 2023. – Т. 77, № 9. – С. 44–50.
16. *Казанцев А.М.* Некоторые подходы к оценке процесса функционирования структурно-динамических систем мониторинга в условиях внешних воздействий / А.М. Казанцев, Р.А. Кочкаров, А.В. Тимошенко, А.А. Сычугов // *Моделирование, оптимизация и информационные технологии*. – 2021. – Т. 9, № 4(35). – DOI: 10.26102/2310-6018/2021.35.4.005.
17. *Chernoritskii S.S.* Assessment of tree biomass using remote sensing techniques: validation at the territory of the ‘forest lake’ educational and recreational complex / S.S. Chernoritskii, N.Y. Gushan, K.M. Khuranova // *Environmental Studies and Protection Issues – 2024: Proceedings of the International Youth Scientific and Academic Conference, Moscow, 19 апреля 2024 года*. – Moscow: P. Lumumba Peoples' Friendship University of Russia, 2024. – P. 74–81.