

АЭРОКОСМИЧЕСКИЕ ИССЛЕДОВАНИЯ ЗЕМЛИ, ФОТОГРАММЕТРИЯ / AEROSPACE RESEARCH OF THE EARTH, PHOTOGRAMMETRY

DOI: <https://doi.org/10.60797/IRJ.2024.144.93>

РАЗРАБОТКА МЕТОДИКИ АВТОМАТИЧЕСКОГО МОНИТОРИНГА ПРОЦЕССА УРБАНИЗАЦИИ БАСЕЙНА КРАСНОЙ РЕКИ (ВЬЕТНАМ)

Научная статья

Фам Ч.<sup>1,\*</sup>, Фам Д.<sup>2</sup>

<sup>1</sup> ORCID : 0009-0003-3849-4173;

<sup>2</sup> ORCID : 0009-0000-7429-584X;

<sup>1,2</sup> Московский государственный университет геодезии и картографии, Москва, Российская Федерация

\* Корреспондирующий автор (vietnam.phamtronghai[at]gmail.com)

**Аннотация**

Настоящая статья посвящена разработке инновационной методики автоматической оценки процесса урбанизации бассейна Красной реки во Вьетнаме. Исследование основано на использовании многолетних данных фотоснимков от спутника Sentinel-2 и картографических материалов от департамента экологических ресурсов Вьетнама с 2016 по 2018 год. Для обучения модели был применен алгоритм случайных лесов, который автоматически выявляет жилые районы. Оценка точности модели послужит основой для обнаружения изменений в урбанизации в период с 2019 по 2024 год.

Кроме того, авторы статьи разработали веб-сайт на платформе Google Earth Engine (GEE), предоставляющий возможность мониторинга и статистического анализа изменений в урбанизации бассейна Красной реки. Этот веб-ресурс обеспечивает мониторинг изменений в городских территориях, что способствует более эффективному планированию и управлению урбанизированными областями. Полученные результаты предоставляют ценные инструменты для научного и практического анализа динамики урбанизации, обеспечивая более точное восприятие изменений в городской среде.

**Ключевые слова:** Google Earth Engine (GEE), урбанизации, алгоритм случайных лесов, Красная река.

DEVELOPMENT OF A METHODOLOGY FOR AUTOMATIC MONITORING OF THE URBANIZATION PROCESS IN THE RED RIVER BASIN (VIETNAM)

Research article

Pham T.<sup>1,\*</sup>, Pham D.<sup>2</sup>

<sup>1</sup> ORCID : 0009-0003-3849-4173;

<sup>2</sup> ORCID : 0009-0000-7429-584X;

<sup>1,2</sup> Moscow State University of Geodesy and Cartography, Moscow, Russian Federation

\* Corresponding author (vietnam.phamtronghai[at]gmail.com)

**Abstract**

This article is dedicated to the development of an innovative methodology for automatic assessment of the urbanization process of the Red River Basin in Vietnam. The study is based on the use of multi-year Sentinel-2 satellite image data and mapping materials from the Department of Environmental Resources of Vietnam from 2016 to 2018. A random forest algorithm was applied to train the model, which automatically identifies residential areas. Evaluation of the model accuracy will serve as a basis for detecting changes in urbanization from 2019 to 2024.

In addition, the authors of this article have developed a website on the Google Earth Engine (GEE) platform that provides monitoring and statistical analysis of changes in urbanization in the Red River Basin. This web resource provides monitoring of changes in urban areas, which contributes to better planning and management of urbanized areas. The results provide valuable tools for scientific and practical analysis of urbanization dynamics, providing a more accurate perception of changes in the urban environment.

**Keywords:** Google Earth Engine (GEE), urbanizations, random forest algorithm, Red River.

**Введение**

Современное общество сталкивается с неизбежным вызовом урбанизации, который влияет на экологию, социальные структуры и экономическое развитие. Однако, несмотря на все преимущества этого процесса, урбанизация также создает серьезные вызовы, такие как ухудшение экологии, увеличение плотности населения и деградацию природных ресурсов. В контексте данной статьи ключевым концептом является разработка методики автоматической оценки процесса урбанизации в бассейне Красной реки [1] во Вьетнаме.

Актуальность данного исследования обусловлена необходимостью эффективного мониторинга и управления урбанизированными областями. Бассейн Красной реки, как регион с быстро растущей урбанизацией, становится объектом повышенного внимания в свете стремительных изменений, происходящих в городской среде. На сегодняшний момент существует острая потребность в разработке инновационных методик, способных автоматически оценивать и отслеживать динамику урбанизации в данном регионе.

Цель данного исследования заключается в разработке и применении методики на основе снимков Sentinel-2 и данных карт за период с 2016 по 2018 год для автоматического обнаружения жилых районов в бассейне Красной реки.

Алгоритм случайных лесов будет использоваться для обучения модели, а точность её оценки послужит основой для обнаружения изменений в урбанизации в период с 2019 по 2024 год.

Характер проблемы обусловлен динамикой роста населения и увеличением строительной активности в регионе, что делает необходимым создание эффективных инструментов для наблюдения и анализа изменений в городской среде. Несмотря на активные исследования в области дистанционного зондирования земли (ДЗЗ), изученность данной проблемы требует новаторских подходов и инструментов для более точного и автоматизированного мониторинга урбанизации.

В данном контексте предлагаемая методика не только решает актуальные проблемы наблюдения за урбанизацией в бассейне Красной реки, но и открывает перспективы для дальнейшего развития подобных инструментов в других регионах. Кроме того, создание веб-сайта на платформе Google Earth Engine (GEE) [2] совместно с мониторингом позволяет предоставлять статистический анализ изменений, обеспечивая общество и ответственных за урбанизацию данными для принятия более обоснованных решений.

#### **Методы и принципы исследования**

Исследование выполняется с использованием коллекции снимков Sentinel-2 MSI [3] и данных карт за период с 2016 по 2018 [4].

Исследование проводилось по 4 основным этапам (см. рисунок 1):

- Этап 1: Подбор и предварительная обработка исходных данных.
- Этап 2: Обучение модели для автоматического обнаружения жилых районов в соответствии с ландшафтом Вьетнама.
- Этап 3: Оценка точности обученной модели.
- Этап 4: Построение карта-схемы автоматической классификации ландшафта бассейна Красной реки.

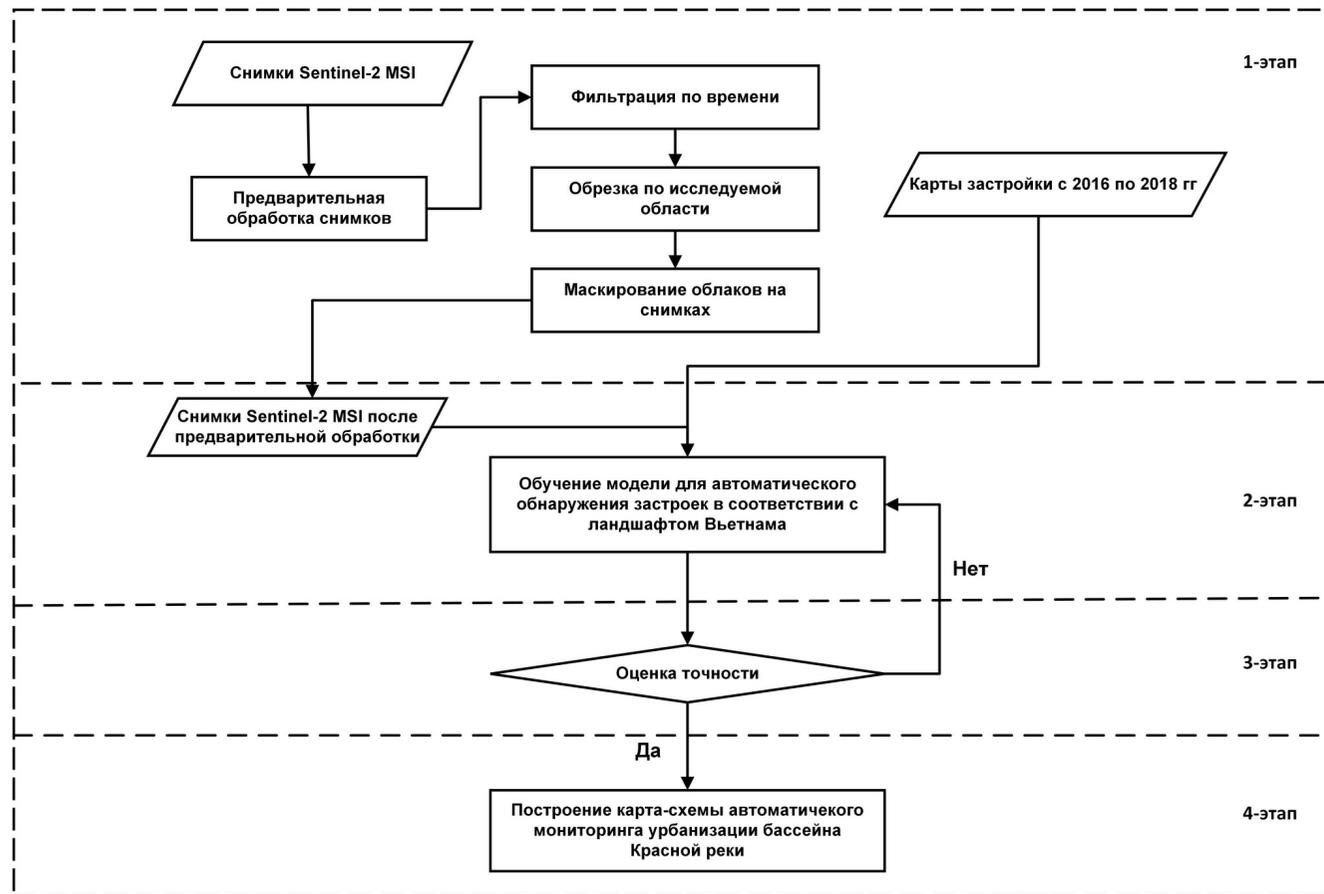


Рисунок 1 - Общая схема исследования  
DOI: <https://doi.org/10.60797/IRJ.2024.144.93.1>

### 2.1. Этап 1. Подбор и предварительная обработка исходных данных

В рамках данного исследования предварительная обработка снимков Sentinel-2, осуществлена непосредственно на платформе Google Earth Engine (GEE). Процесс предварительной обработки [5] включает в себя последовательное проведение следующих этапов:

1. Фильтрация снимков по времени: данная стадия предполагает отбор снимков в соответствии с временными параметрами, обеспечивая релевантность данных для конкретного временного интервала, существенного для анализируемого периода.

2. Обрезка снимков по исследуемой области: производится пространственное выделение и обрезка изображений, ограничивая анализ только территориями, представляющей интерес для исследования, что повышает эффективность анализа и уменьшает вычислительную нагрузку.

3. Маскирование облаков на снимках: этот этап включает в себя автоматизированное создание масок, исключающих области изображения, заслоненные облаками. Данный процесс обеспечивает точность и надежность последующего анализа, исключая искажения, вызванные воздушными облаками на изображениях.

### 2.2. Этап 2. Обучение модели для автоматического обнаружения жилых районов в соответствии с ландшафтом Вьетнама

На втором этапе исследования осуществляется обучение модели с целью автоматического обнаружения жилых районов, адаптированного к особенностям ландшафта Вьетнама. Данный этап включает в себя следующие ключевые шаги:

1. Выбор обучающих данных: подготовка набора данных, на котором будет проводиться обучение модели. Важно учесть разнообразие жилых районов в соответствии с характеристиками ландшафта Вьетнама, чтобы обеспечить модели максимальную обобщенность.

2. Выбор алгоритма обучения: определение оптимального алгоритма для обучения модели, учитывая особенности ландшафта и задачи обнаружения жилых районов. В данном случае, использование алгоритма случайных лесов может обеспечить высокую точность и способность к обобщению на различные типы местности. Алгоритм случайного леса представляет собой ансамбль методов машинного обучения, состоящий из множества решений деревьев, каждое из которых обучается на случайной подвыборке исходных данных. Итоговое решение принимается путем голосования всех деревьев в ансамбле. Алгоритм особенно эффективен для задач классификации и регрессии, так как снижает вероятность переобучения и повышает точность модели. В нашем исследовании использовалось 100 деревьев, и модель обучалась на множестве признаков, таких как спектральные характеристики изображений Sentinel-2 [6], [7], [8].

3. Подготовка признаков: выделение и подготовка характеристик обучающего набора данных, необходимых для эффективного обучения модели. Это может включать в себя спектральные характеристики изображений Sentinel-2, а также другие параметры, характерные для жилых зон.

4. Разделение данных: разбиение обучающего набора данных на тренировочные и тестовые подмножества для оценки производительности модели и её обобщающей способности.

5. Обучение модели: процесс обучения, включающий подачу обучающих данных модели, настройку параметров, итеративное уточнение весовых коэффициентов и оценку модели на тестовых данных.

Данный этап направлен на разработку модели, способной автоматически выявлять жилые районы в ландшафте Вьетнама, что представляет собой ключевой вклад в анализ процесса урбанизации в указанном регионе.

### 2.3. Этап 3. Оценка точности обученной модели

На третьем этапе проводится оценка точности обученной модели, что является критическим шагом для обеспечения надежности и репрезентативности результатов автоматического обнаружения жилых районов в ландшафте Вьетнама.

Для оценки точности модели используются тестовые данные, которые не были использованы в процессе обучения. Эти данные представляют собой изображения, соответствующие различным временным отрезкам после 2018 года. Модель применяется к этим тестовым данным, и ее результаты сравниваются с уже известными фактическими жилыми районами.

Для количественной оценки точности применяются метрики, такие как точность, чувствительность и специфичность. Эти параметры позволяют оценить эффективность модели в правильном выделении жилых зон и избегании ошибок. Оценка точности модели является ключевым фактором для утверждения ее пригодности к использованию в практических приложениях, таких как мониторинг урбанизации в бассейне Красной реки.

Итоговые результаты оценки точности модели получили 85,65 %. Они служат основой для доверительного применения её в процессе обнаружения изменений в урбанизированных областях с 2019 по 2024 год, обеспечивая надежность и достоверность полученных данных [9], [10].

### 2.4. Этап 4. Построение карта-схемы автоматической классификации ландшафта бассейна Красной реки

Для облегчения процесса мониторинга, авторами был разработан и размещен веб-сайт «Карта-схема автоматического мониторинга урбанизации» на платформе Google Earth Engine (см. рисунок 2).

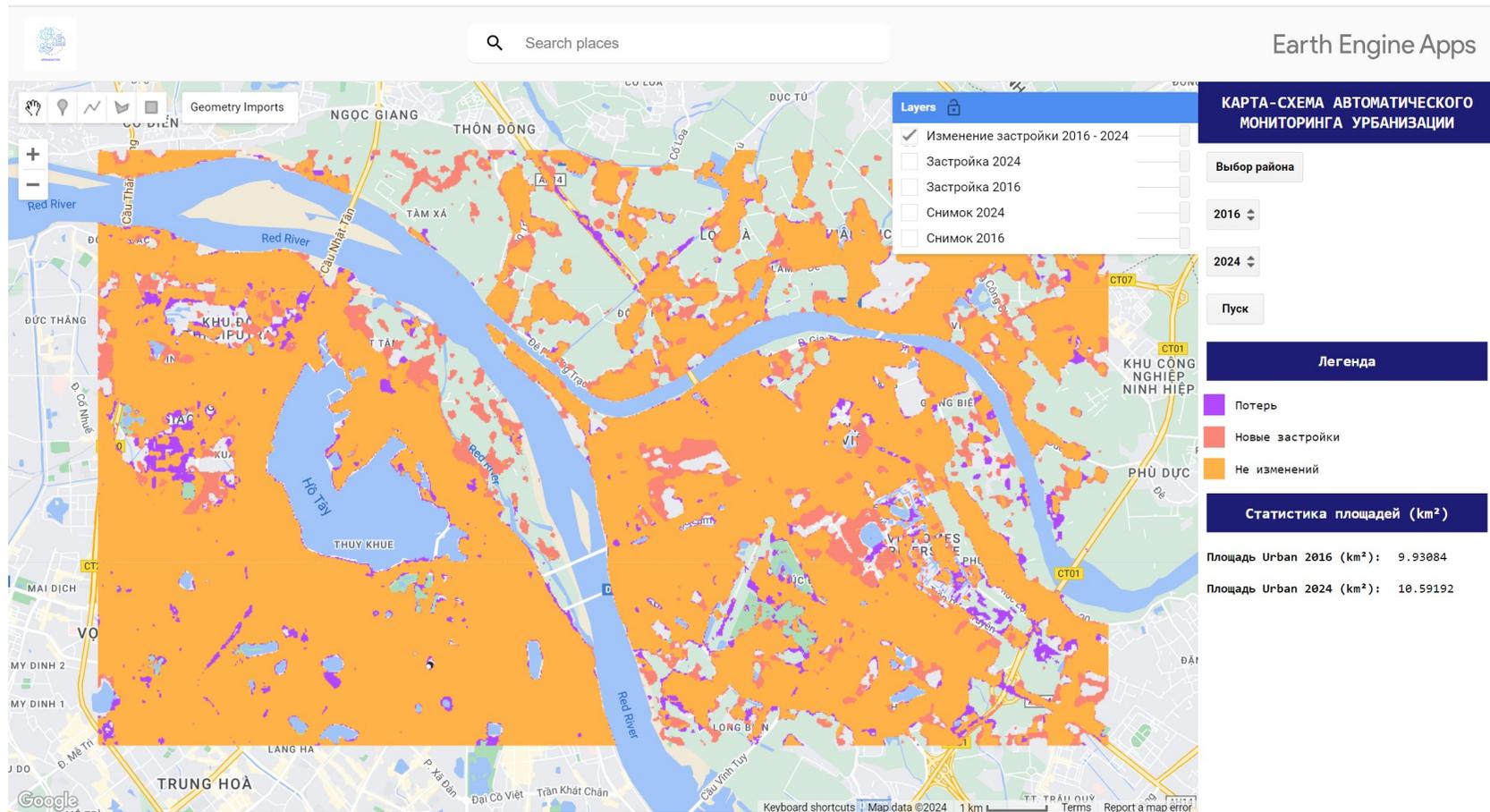


Рисунок 2 - Карта-схема автоматического мониторинга урбанизации  
DOI: <https://doi.org/10.60797/IRJ.2024.144.93.2>

### Основные функции веб-сайта:

1. Выбор области исследований: веб-сайт предоставляет пользователю возможность выбора конкретной области для проведения исследований, что обеспечивает индивидуализацию анализа в соответствии с интересующим регионом.

2. Выбор 1-го и 2-го года: пользователь может выбрать интересующие года, и веб-сайт автоматически загружает и вырезает снимки Sentinel-2 за эти периоды в соответствии с выбранной областью. Веб-сайт также проводит автоматическое обнаружение застроек на данных снимках.

3. Отображение районов изменений: веб-сайт автоматически отображает области с изменениями застройки, включая потери застроек, новые застройки и отсутствие изменений. Это предоставляет наглядное представление о динамике урбанизации в выбранной области.

4. Отображение 4 слоев карты: включает в себя отображение снимков и застроек за два выбранных года, а также изменений в застройке между двумя годами. Это предоставляет комплексный визуальный анализ изменений в урбанистической среде.

5. Статистика площадей застройки: веб-сайт предоставляет статистическую информацию о площади застроенных территорий за два выбранных года, обеспечивая количественные данные для более глубокого понимания динамики урбанизации в исследуемой области.

### Обсуждение

В данном исследовании была разработана методика автоматического мониторинга процесса урбанизации бассейна Красной реки во Вьетнаме, основанная на использовании данных спутниковых снимков Sentinel-2 и картографических материалов за период с 2016 по 2018 год. Применение алгоритма случайных лесов позволило автоматически выявлять жилые районы с высокой точностью, что стало основой для последующего обнаружения изменений в урбанизации с 2019 по 2024 год.

#### 3.1. Сравнение с предыдущими исследованиями

Сравнивая полученные результаты с предыдущими исследованиями в области мониторинга урбанизации, можно отметить несколько ключевых аспектов. Во-первых, использование алгоритма случайных лесов в контексте анализа спутниковых снимков для выявления жилых районов демонстрирует высокую эффективность данного подхода, что согласуется с выводами других авторов [5], [6], [7]. Во-вторых, создание веб-сайта на платформе Google Earth Engine для мониторинга и анализа изменений урбанизации представляет собой инновационный подход, который расширяет возможности использования открытых данных и инструментов для исследований в данной области.

#### 3.2. Значение результатов

Разработанная методика и полученные результаты имеют важное значение для научного и практического анализа динамики урбанизации. Они позволяют не только точно оценить текущее состояние урбанизированных территорий, но и прогнозировать будущие изменения, что крайне важно для планирования и управления урбанизационными процессами. Кроме того, предложенный подход может быть адаптирован и применен для исследования других регионов, что делает его универсальным инструментом в области мониторинга урбанизации.

#### 3.3. Ограничения исследования

Несмотря на высокую эффективность предложенной методики, исследование имеет ряд ограничений. Во-первых, анализ ограничен данными за период с 2016 по 2018 год, что может не полностью отражать текущие тенденции урбанизации. Во-вторых, точность модели зависит от качества исходных данных, в частности, от разрешения спутниковых снимков и актуальности картографических материалов. В-третьих, алгоритм случайных лесов, хотя и показывает высокую эффективность, может требовать дополнительной настройки и оптимизации для конкретных условий и типов ландшафта.

#### 3.4. Предложения по будущим исследованиям

Для устранения ограничений и повышения точности исследований в будущем предлагается расширить временной диапазон анализа, включив более свежие данные. Также целесообразно исследовать возможности применения других алгоритмов машинного обучения и глубокого обучения для улучшения качества автоматического выявления жилых районов. Кроме того, важным направлением будущих исследований является разработка методик для интеграции и анализа данных из различных источников, включая социально-экономические показатели, что позволит получить более комплексное представление о процессах урбанизации.

### Заключение

В ходе данного исследования была разработана инновационная методика автоматической оценки урбанизации в бассейне Красной реки во Вьетнаме, основанная на комплексном анализе многолетних данных Sentinel-2 и картографических материалов с 2016 по 2018 год. Применение алгоритма случайных лесов на этапе обучения модели позволило автоматически выявлять жилые районы с высокой точностью. Результаты этого исследования становятся основой для последующего обнаружения изменений в урбанизации в период с 2019 по 2024 год.

Кроме того, созданный веб-сайт на платформе Google Earth Engine (GEE) предоставляет уникальные возможности мониторинга и статистического анализа эволюции урбанизации в бассейне Красной реки. Этот веб-ресурс не только следит за текущими тенденциями в развитии городских территорий, но также предоставляет ценные инструменты для эффективного планирования и управления урбанизированными областями.

Методология исследования, представленная на 4 этапах, начиная от подбора и предварительной обработки данных до построения карты-схемы автоматической классификации ландшафта, обеспечивает системный и всесторонний подход к изучению динамики урбанизации. Полученные результаты и разработанный веб-сайт представляют собой

ценный вклад в область дистанционного зондирования земли и обеспечивают научное и практическое понимание изменений в городской среде бассейна Красной реки во Вьетнаме.

### Конфликт интересов

Не указан.

### Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

### Conflict of Interest

None declared.

### Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

### Список литературы / References

1. Luu T N.M. Hydrological regime and water budget of the Red River Delta (Northern Vietnam) / T.N.M. Luu, J. Garnier, G. Billen [et al.] // *Journal of Asian Earth Sciences*. — 2010. — 37(3). — P. 219-228. — DOI: 10.1016/j.jseae.2009.08.004
2. Earth Engine guides. — URL: <https://earthengine.google.com/> (accessed: 13.03.2024)
3. Claverie M. The Harmonized Landsat and Sentinel-2 surface reflectance data set / M. Claverie, J. Ju, J.G. Masek [et al.] // *Remote sensing of environment*. — 2018. — 219. — P. 145-161.
4. Ứng dụng công khai thông tin đất đai thành phố Hà Nội. — URL: <http://qhkhsdd.hanoi.gov.vn/> (accessed: 13.03.2024)
5. Würsch L. Google Earth Engine image pre-processing tool: user guide / L. Würsch, K. Hurni, A. Heinimann // *Centre for Development and Environment*. — 2017. — URL: <https://clck.ru/3BGD9d> (accessed 13.03.2024)
6. Belgiu M. Random forest in remote sensing: A review of applications and future directions / M. Belgiu, L. Drăguț // *ISPRS journal of photogrammetry and remote sensing*. — 2016. — 114. — P. 24-31. — DOI: 10.1016/j.isprsjprs.2016.01.011
7. Svoboda J. Random forest classification of land use, land-use change and forestry (LULUCF) using sentinel-2 data / J. Svoboda, P. Štych, J. Laštovička [et al.] // *A case study of Czechia. Remote sensing*. — 2022. — 14(5). — P. 1189. — DOI: 10.3390/rs14051189
8. Pal M. Random forest classifier for remote sensing classification / M. Pal // *International journal of remote sensing*. — 2005. — 26(1). — P. 217-222. — DOI: 10.1080/01431160412331269698
9. Rodriguez-Galiano V.F. An assessment of the effectiveness of a random forest classifier for land-cover classification / V.F. Rodriguez-Galiano, B. Ghimire, J. Rogan [et al.] // *ISPRS journal of photogrammetry and remote sensing*. — 2012. — 67. — P. 93-104. — DOI: 10.1016/j.isprsjprs.2011.11.002
10. Shi D. An assessment of algorithmic parameters affecting image classification accuracy by random forests / D. Shi, X. Yang // *Photogrammetric Engineering & Remote Sensing*. — 2016. — 82(6). — P. 407-417. — DOI: 10.14358/PERS.82.6.407

### Список литературы на английском языке / References in English

1. Luu T N.M. Hydrological regime and water budget of the Red River Delta (Northern Vietnam) / T.N.M. Luu, J. Garnier, G. Billen [et al.] // *Journal of Asian Earth Sciences*. — 2010. — 37(3). — P. 219-228. — DOI: 10.1016/j.jseae.2009.08.004
2. Earth Engine guides. — URL: <https://earthengine.google.com/> (accessed: 13.03.2024)
3. Claverie M. The Harmonized Landsat and Sentinel-2 surface reflectance data set / M. Claverie, J. Ju, J.G. Masek [et al.] // *Remote sensing of environment*. — 2018. — 219. — P. 145-161.
4. Ứng dụng công khai thông tin đất đai thành phố Hà Nội [Application of Public Land Information Hanoi City]. — URL: <http://qhkhsdd.hanoi.gov.vn/> (accessed: 13.03.2024) [in Vietnamese]
5. Würsch L. Google Earth Engine image pre-processing tool: user guide / L. Würsch, K. Hurni, A. Heinimann // *Centre for Development and Environment*. — 2017. — URL: <https://clck.ru/3BGD9d> (accessed 13.03.2024)
6. Belgiu M. Random forest in remote sensing: A review of applications and future directions / M. Belgiu, L. Drăguț // *ISPRS journal of photogrammetry and remote sensing*. — 2016. — 114. — P. 24-31. — DOI: 10.1016/j.isprsjprs.2016.01.011
7. Svoboda J. Random forest classification of land use, land-use change and forestry (LULUCF) using sentinel-2 data / J. Svoboda, P. Štych, J. Laštovička [et al.] // *A case study of Czechia. Remote sensing*. — 2022. — 14(5). — P. 1189. — DOI: 10.3390/rs14051189
8. Pal M. Random forest classifier for remote sensing classification / M. Pal // *International journal of remote sensing*. — 2005. — 26(1). — P. 217-222. — DOI: 10.1080/01431160412331269698
9. Rodriguez-Galiano V.F. An assessment of the effectiveness of a random forest classifier for land-cover classification / V.F. Rodriguez-Galiano, B. Ghimire, J. Rogan [et al.] // *ISPRS journal of photogrammetry and remote sensing*. — 2012. — 67. — P. 93-104. — DOI: 10.1016/j.isprsjprs.2011.11.002
10. Shi D. An assessment of algorithmic parameters affecting image classification accuracy by random forests / D. Shi, X. Yang // *Photogrammetric Engineering & Remote Sensing*. — 2016. — 82(6). — P. 407-417. — DOI: 10.14358/PERS.82.6.407