

**ПОДХОД К ОЦЕНКЕ ЗАПАСОВ УГЛЕРОДА В ХМАО-ЮГРЕ
С ПОМОЩЬЮ УГЛЕРОДНЫХ КАРТ**

Бредихин Арсентий Игоревич

*магистрант Института цифровой экономики
ФГБОУ ВО «Югорский государственный университет»,
Ханты-Мансийск, Россия
E-mail: bredihin.igorr@yandex.ru*

Ханты-Мансийский автономный округ – Югра обладает большой площадью лесных территорий. А лесная растительность, как и любая растительность, естественным образом рано или поздно отмирает, вследствие чего из органического вещества происходит выделение углекислого газа в атмосферу. Данный факт ведет к усилению парникового эффекта и усилению глобального потепления. Для того чтобы не допустить повышения глобальной температуры, необходимо оценивать запас углерода в виде количества растительной биомассы, поскольку более 90 % территории Ханты-Мансийского автономного округа – Югры (ХМАО-Югра) покрыто лесами. Одним из способов оценки растительной биомассы является создание так называемых углеродных карт с применением методов дистанционного зондирования Земли (ДЗЗ) и машинного обучения. Применение полученных с помощью методов ДЗЗ спутниковых снимков и методов их обработки позволит получить карту округа с полным охватом всей территории, а применение моделей машинного обучения позволит разработать модель, с помощью которой будет возможно создавать углеродную карту округа.

В данной работе приведен обзор существующих решений в области ДЗЗ и машинного обучения, направленных на создание углеродных карт. На основании данного обзора предложена программа исследований, которая позволит разработать подход, позволяющий получать цифровую углеродную карту ХМАО с заданной точностью.

Ключевые слова: дистанционное зондирование Земли, спутник, лидар, изображение, углерод, растительность, биомасса, машинное обучение, регрессионная модель

**AN APPROACH TO THE ASSESSMENT OF CARBON RESERVES
IN KHAMAO-YUGRA USING CARBON MAPS**

Arsenty I. Bredihin

*Master Student at the Institute of Digital Economy,
Yugra State University
Khanty-Mansiysk, Russia
E-mail: bredihin.igorr@yandex.ru*

Khanty-Mansi Autonomous Okrug-Yugra has a large area of forest territories. And forest vegetation, like any vegetation, naturally dies sooner or later, as a result of which carbon dioxide is released into the atmosphere from organic matter. This fact leads to an increase in the greenhouse effect and an increase in global warming.

In order to prevent an increase in global temperature, it is necessary to estimate the carbon stock in the form of the amount of plant biomass, since more than 90% of the territory of the Khanty-Mansi Autonomous Okrug-Yugra (KhMAO-Yugra) is covered with forests.

One of the ways to assess plant biomass is to create so-called carbon maps using remote sensing of the Earth (remote sensing) and machine learning methods.

This paper provides an overview of existing solutions in the field of remote sensing and machine learning aimed at creating carbon maps. Based on this review, a research program has been proposed that will allow us to develop an approach that allows us to obtain a digital carbon map of the KhMAO with a given accuracy.

Keywords: Remote sensing, satellite, lidar, image, carbon, vegetation, biomass, machine learning, regression model

Введение

Ханты-Мансийский автономный округ – Югра является крупным нефте- и газодобывающим российским регионом. Он характеризуется большими объемами выбросов парниковых газов в атмосферу. Так, на территории округа имеется большое количество болот, которые являются источником метана (CH_4), а нефте- и газодобыча – источником диоксида углерода (CO_2).

Для недопущения повышения глобальной температуры необходимо вести учет запасов углерода. Помимо выбросов от сжигания топлива, диоксид углерода (CO_2), который является парниковым газом, образуется при разложении мертвой растительности, поскольку растения накапливают и хранят в себе углерод. При этом Ханты-Мансийский автономный округ – Югра (ХМАО-Югра) обладает большой площадью лесов: по состоянию на 2020 год она составляет 503 990 км² при общей площади территории округа в 534,8 тыс. км² [1, с. 608-610], что составляет 94,2 % площади всего округа. Поэтому для учета запаса углерода в рамках округа необходимо оценить данный запас в виде количества растительной (лесной) биомассы.

Одним из способов оценки лесной биомассы округа является применение спутниковых изображений оптического и инфракрасного диапазонов совместно с лидарными снимками. Здесь примечательным является подход к решению данной задачи американскими учеными [2], который основан на применении полевых измерений с совместными данными НАСА (лидар GLAS со спутника ICESat).

Суть подхода заключается в последовательном получении полевых измерений – массы деревьев на определенных участках, определении согласованности данных с сигналами лидара для данных участков, обучении модели «случайный лес» на отдельных участках карты и вычислении с ее помощью биомассы на остальных участках карты.

В работе [2] описана усовершенствованная версия данного подхода из статьи [3]. В обоих случаях тот же коллектив ученых разрабатывал модели в континентальном масштабе – для Африки, Азии и Южной Америки и использовал для этого изображения MODIS с разрешением 500 м. В последней версии подхода ученым удалось объяснить с помощью полученной модели 83 %, 78 % и 71 % различий в биомассе для тропической Америки, Африки и Азии соответственно.

Однако для нашего случая подход необходимо модифицировать. Ведь разрешение изображений в 500 м оказывается недостаточным для построения углеродной карты в рамках региона РФ. В качестве альтернативы необходимо использовать снимки с большим разрешением, например Landsat-7/8 (разрешение 30-60 м), WorldView3 (разрешение 3,7 м). Также следует принять во внимание тот факт, что различия в видах растительности, обусловленные различиями климата, также могут влиять на точность модели (авторы [2] создавали модели для экваториальных и тропических регионов, тогда как ХМАО-Югра расположена в таежной зоне или зоне резко континентального климата).

К тому же данный подход обладает следующими ограничениями: он не учитывает углерод, хранящийся в почве, травянистую и кустарниковую растительность и предполагает полное выделение CO_2 в атмосферу, в то время как на практике это происходит не всегда [2].

Данная статья рассматривает проблему получения углеродной карты. Она разбита на 3 части: в первой части будут подробно рассмотрены методы сбора полевых данных и преоб-

разования их в значения биомассы. Вторая часть будет посвящена анализу сигналов лидара и проверке согласованности полевых данных с метриками сигналов лидара. В третьей части будет уделено внимание построению модели машинного обучения для вычисления значений биомассы различных участков будущей углеродной карты округа. На протяжении всей работы будут приводиться предложения по изменению подхода к созданию углеродных карт для адаптации его к условиям северного региона.

Сбор полевых данных

Сбор полевых данных является первым этапом создания углеродной карты. Для начала выбираются участки, соответствующие «пятнам» лидара GLAS.

«Пятна», т. е. отпечатки лидара GLAS, имеют диаметр около 70 метров и расположены на расстоянии 170 метров друг от друга. Съемка ведется вдоль поверхности Земли с помощью двух каналов с длинами волн 532 и 1064 нм [4]. Также в [2] упоминается, что одно «пятно» GLAS встречается примерно на каждые 6 км² земной поверхности, что для территории ХМАО соответствует ~89 тысячам снимков в среднем. Согласно данным из [5], архив данных лидара GLAS за 3 года (с 10.2018 по 01.2022) содержит примерно 5468 файлов общим размером ~11 Тб для территории ХМАО. Данные загружаются с периодичностью в 3 месяца; количество получаемых для территории ХМАО наборов данных (датасетов) составляет от 4 до 6 в день.

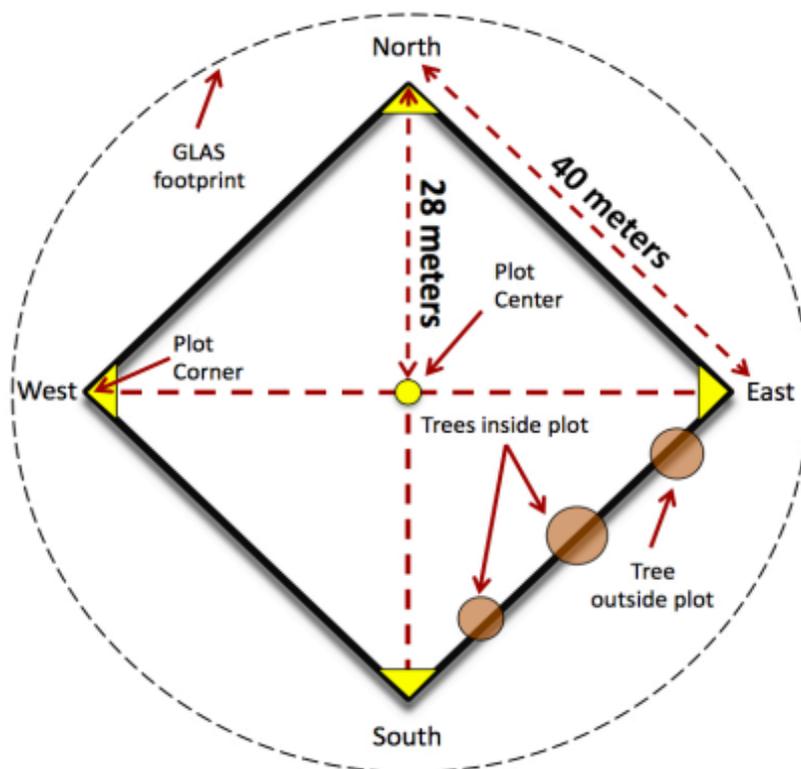


Рисунок 1. Схема выбора участка в «пятне» лидара GLAS [3]

Замечание. В работе [3] используются датасеты, в которых диаметр «пятен» равен 65 м; участок имеет форму квадрата со сторонами 40 м.

Затем выбираются участки лесных массивов, целиком лежащие в «пятнах» лидара GLAS. После чего в границах каждого участка измеряется диаметр ствола всех деревьев на высоте 1,3 м от земли. При вычислении биомассы деревьев используется среднее значение удельного веса дерева (плотность) на основе вида.

В работе [2] для оценки биомассы предлагается регрессионная модель следующего вида:

$$\ln(AGB) = \alpha + \beta_1 \ln(D) + \beta_2 \ln(H) + \beta_3 \ln(\rho), \quad (1)$$

где AGB – значение биомассы (англ. above-ground biomass), D и H – диаметр и высота дерева, а ρ – плотность древесины. Кроме того, здесь же выделяются следующие 6 видов данной модели:

1. Все 4 параметра модели (1) настраиваются независимо для различных типов лесов.
2. Аналогично предыдущей модели с предположением, что параметры модели не меняются в зависимости от типа леса.

Выделяются и модели, зависящие только от массы отдельных деревьев (переменная $\rho D^2 H$):

1. Модель вида $\ln(AGB) = \alpha + \beta_1 \ln(\rho D^2 H)$.
2. Аналогично предыдущей модели с тем же предположением.
3. Модель вида $\ln(AGB) = \alpha + \ln(\rho D^2 H)$.
4. Аналогично предыдущей модели с тем же предположением.

Однако на практике высота дерева не всегда доступна. Поэтому, как считают авторы [2], необходимо отойти от ее использования в модели:

$$\ln(AGB) = \alpha + b \ln(D) + c(\ln(D))^2 + d(\ln(D))^3 + \beta_3 \ln(p). \quad (2)$$

При логарифмировании данных окончательная оценка будет смещаться, и, таким образом, нескорректированные оценки биомассы будут недооценивать реальные значения. Поэтому оценку необходимо умножить на поправочный коэффициент:

$$CF = \exp\left(\frac{RSE^2}{2}\right), \quad (3)$$

где RSE – метрика полученной модели (англ. root square error – корень квадратичной ошибки).

Авторы [2] утверждают, что разновидности данной модели в основном применяются для оценки биомассы широколиственных и тропических деревьев. Что же касается лесов ХМАО (тайга, тундра), то в данном случае необходимо провести собственные измерения и определить более точные регрессионные модели для данного случая. В случае приемлемого значения ошибки полученной регрессионной модели ее можно будет применять для вычисления массы дерева в зависимости от диаметра и плотности, а также для проверки согласованности полученных полевых данных с метриками сигналов лидара GLAS.

Лидары применяются для оценки биомассы и в аналогичных исследованиях. Так, в работе [6] авторы применяют гиперспектральные и лидарные данные для оценки биомассы кукурузы. Однако вместо спутниковых лидаров используется бортовой лидар Leica с дискретным возвратом, предназначенный для съемок на малой высоте. В данной работе упоминается об ограничении применения лидарных данных в случае густой растительности, об ограничении применения самих лидаров дискретного возврата, поскольку они регистрируют только один возврат для каждого испускаемого импульса в районах с низкорослой растительностью, а также о невозможности использования лидара для анализа спектральных характеристик растительного покрова, поскольку лидары работают на одной длине волны.

При этом в работе [6] процедура полевых измерений схожа с таковой в работе [3]. Для этого на участках $4 \times 4 \text{ м}^2$ измерялись высота и масса стеблей кукурузы, а также индекс площади проективного покрытия листьев (LAI) с помощью портативного прибора LAI-2200. При этом биомасса кукурузы вычислялась с помощью регрессионного уравнения, полученного опытным путем. Для оценки способности регрессионных моделей к обобщению применяется метод машинного обучения под названием «кросс-валидация без исключения» (leave-one-out), чего нет в подходе из работы [3].

Об ограниченности применения данных лидаров дискретного возврата упоминается и в работе [7], которая посвящена оценке биомассы хвойных лесов. Как утверждают авторы, с помощью индексов проникновения лазера не может быть точно оценен покров полога.

Способ получения полевых измерений в работе [7] полностью аналогичен основной работе ([3]), с тем отличием, что дополнительно измеряется высота дерева с помощью лазерного гипсометра. Там же приведены аллометрические уравнения, которые учитывают еще и

высоту дерева. При этом уравнения предназначены для определения биомассы ствола, ветвей, листьев и плодов (sic!) дерева.

В работе [8] применяются лидарные данные и гиперспектральные изображения для определения влияния размера выбранного участка на точность определения биомассы. При этом оценивается не только надземная, но еще и подземная биомасса.

Работа [9] посвящена разработанному авторами методу оценки объема ствола под названием Outer Hull Model – ОНМ. Данный метод применяется для оценки биомассы отдельных деревьев, что в нашем случае может быть слишком затратно. В данной работе используется надземный лидар с разрешением 6 мм. Здесь в полевых измерениях измеряются не использовавшиеся ранее характеристики отдельного дерева (сосны): высота достижения диаметра основного ствола в 4 дюйма (10,16 см), корона (расстояние между первой ветвью и высоты достижения диаметра ствола). Метод измерения характеристик дерева, в отличие от остальных работ, здесь связан со сбором компонентов дерева (ветви) и их дальнейшей обработкой (высушивание). Согласно полученным в [9] результатам, для хвойных деревьев биомасса хвои и ветвей дерева может достигать до 20 % биомассы всего дерева.

В работе [10], посвященной оценке лесной биомассы, используются полученные в ходе миссии GEDI лидарные данные. Миссия GEDI была начата сравнительно недавно на момент публикации данной статьи – 5 декабря 2018 г. Согласно данным из [11], лидар миссии GEDI установлен на Международной космической станции (МКС) и собирает данные в диапазоне с 51.6 ю.ш. до 51.6 с.ш. В нашем случае данный лидар не подойдет, поскольку самая южная точка ХМАО имеет координату ~58 35' с.ш. К тому же разрешение лидара GEDI слишком грубо для нашего случая – 1 км. Упоминается и об оптимальном разрешении снимков для создания углеродных карт – оно принято равным 100 м.

В работе [12] процедура сбора полевых данных также ничем не отличается от таковых в рассмотренных работах. В ней упоминается об использовании значения диаметра дерева на уровне груди в 10 см. Авторы [12] утверждают, что это значение может быть слишком велико, поскольку на деревья с меньшим диаметром может приходиться значительная доля биомассы (произведена оценка биомассы в Кении). Здесь также приводятся аллометрические уравнения для разнородных лесных участков, указаны справочные материалы для определения плотности древесины, а также упоминается о способе определения положения центра участка (лидарного «пятна») с помощью GNSS-приемника (Trimble GeoXH).

Работа [13] посвящена оценке лесной биомассы на острове Калимантан. В данной работе представляет интерес необычный подход к определению биомассы на участке – он основан на методе Монте-Карло. Авторы работы вносят погрешности в характеристики деревьев (высота, диаметр, плотность), предполагая, что в ходе измерений данных характеристик могут возникать ошибки, связанные с неточностью измерения или отсутствием необходимых справочных данных (таксономическая информация, плотность древесины). Для приемлемого уровня надежности, как утверждают авторы [13], достаточно по 100 оценок на каждое дерево. Здесь метод Монте-Карло применяется и для учета возможных ошибок при определении координат участка в «пятне» лидара.

Работа [14] в этом списке стоит особняком: здесь описывается подход к оценке биомассы травянистой растительности, растущей вблизи болот. Процедура сбора полевых данных отличается от таковой для деревьев: трава срезается с определенной площади, высушивается при 55 °С и затем взвешивается, после чего определяется плотность отдельного вида болотной растительности. В отличие от вычисления биомассы лесов здесь не требуется использование лидарных данных.

Работа [15] также отличается от остальных: в ней описывается подход оценки биомассы в урбанизированных территориях. Особенность данного подхода заключается в использовании высокоточных спутниковых снимков (разрешение 5 м), а для измерения высоты деревьев при сборе полевых данных используется лазерный высотомер. Подход включает в себя метод стратификации лидарных снимков, т. е. деление снимков на слои по высоте.

Исходя из представленного обзора методов сбора полевых данных, можно видеть, что во всех работах получение данных о растительности связано либо с измерениями на местах (деревья), либо со сбором и обработкой образцов растительности (трава). При агрегировании данных о биомассе наиболее подходящим в нашем случае способом агрегации данных является вычисление среднего значения биомассы по классам растительности аналогично предельным действиям в работах [3, 15].

Анализ сигналов лидара

Следующий этап создания углеродной карты – анализ формы сигнала лидара GLAS. Это необходимо для установления статистической взаимосвязи между полевыми оценками плотности надземного углерода (англ. aboveground carbon density, ACD). В работе [2] предполагается, что отношение массы углерода к общей биомассе составляет 1:2.

Здесь же подробно изложен процесс анализа «пятен» лидара GLAS. Так, перед непосредственной обработкой данных авторы отбросили «пятна», которые:

- имеют менее 2 пиков;
- имеют максимальную высоту, превышающую шум менее чем в 2 раза;
- имеют расхождение со значениями высоты от SRTM более 25 м.

Подробное описание алгоритма оценки биомассы по снимкам лидара (LVIS) дано в работе [16].

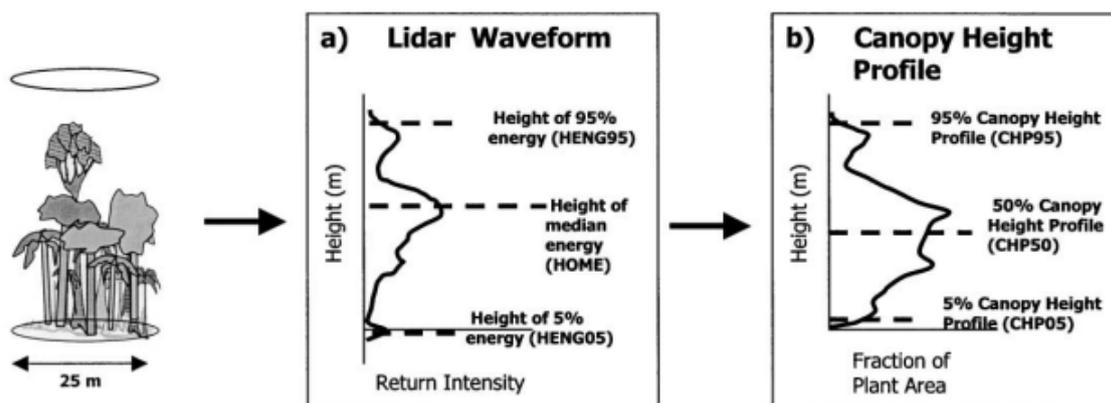


Рисунок 2. Схема оценки биомассы по снимку лидара [16]

Как видно из рисунка 2, сигнал с лидара (а) переводится в профиль высоты покрова (b). При этом используется методика из статьи [17]. По сути, дальнейший анализ «пятна» лидара связан с анализом гистограммы профиля покрова.

Методика перевода сигнала с лидара в профиль высоты покрова из статьи [17] состоит из следующих шагов:

1. Определение и отсечение шума. Сводится к последовательному вычислению значений среднего арифметического и дисперсии шума в части сигнала, которая соответствует «отрицательной высоте», генерации сигнала с заданными значениями среднего и дисперсии и последующему его вычитанию из исходного сигнала.

2. Сглаживание функции и вычисление кумулятивной «функции покрова».

3. Преобразование кумулятивной «функции покрова» по методологии МакАртура-Хорна (применение формулы $-\ln[1-closure]$, где $closure$ – полученная на предыдущем шаге «функция покрова»).

4. Нормализация функции и обратное преобразование ее в плотность вероятности.

Схема методики показана ниже.

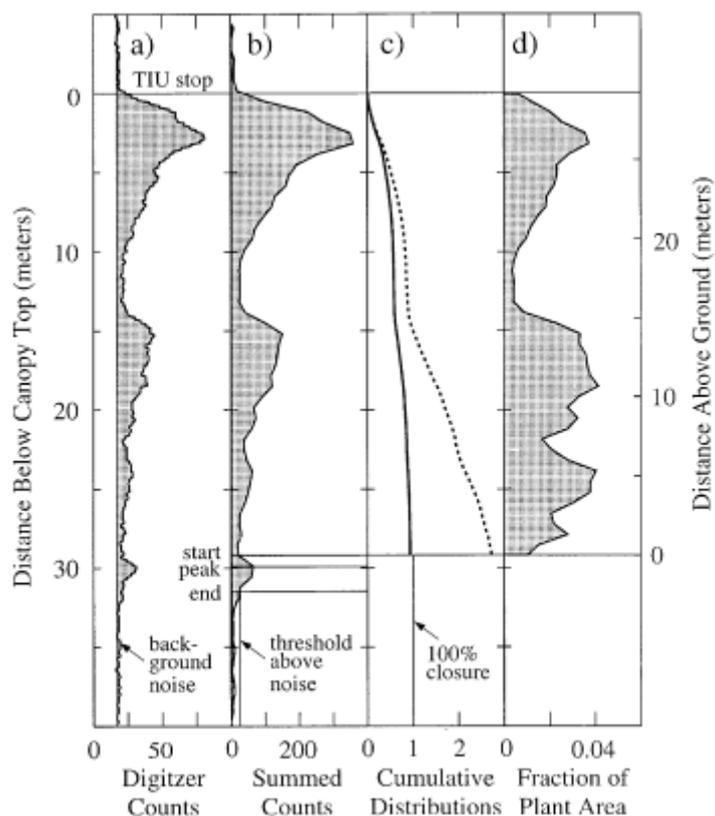


Рисунок 3. Схема перевода сигнала лидара в профиль высоты покрова [17]

Профиль высоты покрова является не чем иным, как функцией плотности распределения высот в «пятне» лидара. Далее из профиля высоты извлекаются следующие значения:

- H_p – значения p -го перцентиля функции распределения высот. Используемые значения p : 10, 25, 30, 40, 60, 75, 90;
- H_{OME} – медианное значение высоты;
- $MAXPEAKHT$ – максимальное значение высоты;
- $CANOPY_ENE$ – интеграл функции высоты от начала сигнала (поверхность земли) до $MAXPEAKHT$.

Данные значения применяются для построения регрессионной модели, которая имеет следующий изначальный вид: $(0 + H_{OME} + HEIGHT2 + H_{10} + H_{25} + H_{30} + H_{40} + H_{60} + H_{75} + H_{90} + MAXPEAKHT + CANOPY_DEP + CANOPY_ENE)$.

Замечание. Авторы [5] не посчитали нужным указать смысл независимых переменных $HEIGHT2$ и $CANOPY_DEP$, поскольку в их регрессионной модели значимыми оказались только переменные H_{OME} , H_{10} , H_{25} , H_{60} , $CANOPY_ENE$.

Применение вышеописанных алгоритмов в работе [3] позволило получить по 2 значения биомассы на каждый участок, по которым была построена линейная регрессионная модель. Проверка модели с помощью статистического теста на несоответствие, основанного на локальной взвешенной регрессии (LOWESS), подтвердила правильность выбора именно линейной модели и показала хорошую согласованность модели с данными.

В других работах по составлению карты плотности углерода проводились аналогичные исследования по установлению связи между полевыми данными биомассы и метриками формы сигнала лидара. В работе [7], в отличие от подхода из [3], при анализе лидарных данных для каждого «пятна» вместо профиля высоты создается цифровая модель земной поверхности с помощью воксельного подхода. Пространство, в котором находятся формы сигналов, разделяется на воксели, а затем каждому вокселю присваивается максимальное значение амплитуды среди всех амплитуд в пределах одного и того же вокселя.

А в работе [8] для получения цифровой модели земной поверхности используется метод интерполяции триангулированной нерегулярной сети. Цифровая модель земной поверхности, так же как и в аналогичных работах, используется для калибровки сигналов лидара, поскольку для получения модели высоты растительного покрова необходимо использование двух источников данных: карты высот земной поверхности и самих лидарных снимков. Более подробно подход составления цифровой модели высоты покрова описан в [21].

Также в [8] рассматриваются различные метрики показателей лидара. Упоминается об обычно используемых лидарных показателях для оценки биомассы растительности: процентилях, максимумы, средние значения и стандартные отклонения высоты лидара и индекс лазерного перехвата (LH). Проведено исследование, в котором протестирован ряд лидарных показателей для получения оптимальной модели оценки биомассы.

Упоминается в работе [8] и о влиянии размера участка с лидара на точность оценки параметров растительности. Поэтому было проведено еще одно соответствующее исследование, результаты которого показали, что использование переменных размеров участков для извлечения лидарных показателей может быть полезным для точной оценки параметров растительности (использованы размеры участков от 11 до 30 м). Также определен порог высоты для отделения отдачи от полога от отдачи от земли.

В работе [9] для оценки объема ствола хвойного дерева так же, как и в работе [7], применяется воксельный подход при анализе сигналов лидара. Для обработки данных лидара применялось ПО «Faro SCENE».

При анализе сигналов лидара в работе [10] используется метод случайного объема над землей (Random Volume over Ground – RvOG), который применяется к полученным с помощью InSAR изображениям – спутник TanDEM-X (TDX) InSAR, используемый для измерения высот. Получаемые в ходе RVoG статистические значения затем применяются для настройки линейных регрессионных моделей. Авторы [10] утверждают, что метод RVoG может быть применен к любым данным «wall-to-wall».

Работа [22] представляет скорее теоретический интерес: в ней описывается трехмерная модель формы волны, которая использовалась для моделирования сигналов лидара в зависимости от структуры древостоя и характеристик датчика. В контексте данной модели крона деревьев не что иное, как рассеивающая среда, параметризованная индексом плотности листьев, коэффициентом отражения листьев, G-фактором Росса – Нильсона и коэффициентом отражения от земли.

В [22] предлагается алгоритм обработки «сырых» (raw) лидарных данных с применением формы волны сигнала и цифровой модели земной поверхности. Форма обратной волны лидара моделируется как сумма отражений в пределах «пятна» лидара диаметром 25 м, затем формы сигналов нормализуются до их максимального пика и сворачиваются ядром Гаусса.

Одним из этапов алгоритма обработки сигнала лидара является применение поисковой таблицы из 100 тыс. различных образцов сигнала лидара. Определение объема биомассы сводится к выбору наиболее близкого ко входному сигналу образца с помощью функции дистанции.

Интерес представляет сам алгоритм преобразования «сырых» лидарных данных, в то время как применение поисковой таблицы является устаревшим приемом, который в настоящий момент заменили методы машинного обучения.

Как упоминалось ранее, в работе [21] достаточно подробно описан процесс получения цифровой модели покрова, которая используется для проверки связи таких характеристик деревьев, как высота и диаметр, на высоте груди с общей биомассой. Общая модель покрова строится на основе цифровой модели поверхности (digital surface model – DSM; суть – лидарные снимки) и цифровой модели земной поверхности (digital terrain model – DTM; суть – карта высот). Строились 2 вида моделей DSM – на основе снимков лидаров с дискретным возвратом и на основе снимков лидаров полной волны. Общая модель покрова получается путем вычитания DTM из DSM; по необходимости модели DTM и DSM приводятся к единому разрешению с помощью интерполяции по методу кригинга.

Также в [21] представляет интерес процесс получения итоговой регрессионной модели для оценки объема ствола дерева: коэффициенты получены путем усреднения соответствующих коэффициентов, полученных в ходе перекрестной проверки моделей. Перекрестная проверка производилась по методу LOOCV. Общее количество деревьев на участках составило 61, а общее количество полученных моделей – 61.

В работе [12] заслуживают внимания приемы обработки лидарных данных. В ней рассказывается об обработке лидарных снимков: здания, линии электропередач и выбросы (высокие точки) фильтруются с помощью ПО Terrascan, LAStools (Rapidlasso GmbH) и ручного редактирования. Приводятся аллометрические уравнения для разнородных лесных участков, а также указаны справочные материалы для определения плотности древесины. Поиск наиболее точных регрессионных моделей производится путем перебора различных комбинаций из 1-3 предикторов (функция «regsubsets» пакета «leaps» языка программирования R).

Также заслуживают внимания метрики лидарных сигналов из работы [13]. Среди метрик лидарного сигнала используется необычная метрика – высота верхушки покрова (top canopy height – TCH). Суть метрики – отношение количества пикселей, лежащих на высоте не меньше определенной (в работе [13] – 20 м) к общему количеству пикселей. Рассматриваются модели для определения базальной площади и плотности деревьев. Для оптимизации уравнений моделей используется процедура нелинейной оптимизации L-BFGSB, реализация которой имеется в Python.

В работе [14] выполняется классификация травянистой растительности с помощью ПО eCognition и на основании объектно ориентированного подхода. В данной работе требуется другая карта для болотистой растительности, чтобы в дальнейшем свести данные из нее к разрешению спутниковых снимков. Для создания детализированной карты болотистой растительности использовались снимки NAIP с разрешением 1 м, после чего данные биомассы с первой карты были использованы для сведения полученных значений к разрешению спутниковых снимков Sentinel (разрешение 30 м). Причем вторая карта является фракционной, т. е. в пределах каждого пикселя Landsat-изображения были определены доли того или иного вида болотистой растительности.

В работе [15], как уже говорилось ранее, обработка лидарных снимков включает в себя стратификацию лидарных снимков, т. е. деления снимков на слои по высоте. В качестве моделей регрессии для оценки объема биомассы используются «бустинговые регрессионные деревья» (BRT) и пространственное авторегрессионное моделирование (spatial autoregressive modelling) для объяснения наиболее значимых переменных. Для вычисления общей биомассы используются усредненные значения биомассы по классам. Полученные в работе [15] значения коэффициентов корреляции Пирсона высоты покрова с биомассой возрастают с ростом высоты слоя: для высот 2-5 м коэффициент корреляции равен 0,3, для 5-10 м – 0,5, а для высот более 10 м – 0,7.

Исходя из представленного обзора методов обработки лидарных данных, видно, что во всех случаях применение характеристик сигналов лидара приводит к повышению точности оценки биомассы.

В ряде рассмотренных работ для определения формы растительности используются бортовые лидары [9, 12, 13] или лазерные высотомеры [15]. При этом в работе [13] утверждается, что модели оценки биомассы на основе данных бортовых лидаров показывают более высокую точность, чем модели оценки биомассы на основе данных спутниковых лидаров. В нашем случае, скорее всего, не будет возможности применения бортовых лидаров или лазерных высотомеров, поэтому в нашей задаче придется задействовать данные спутниковых лидаров, прежде всего GLAS.

Создание карты биомассы

В работе [3] создание карты плотности углерода сводится к созданию карты лесной биомассы. Разработанные для создания карт биомассы методы включают в себя:

- классификацию типов растительного покрова, где каждому классу присваивается среднее значение плотности биомассы на основе справочных оценок или данных инвентаризации весов;
- определение взаимосвязей между плотностью биомассы и характеристиками пикселей спутниковых изображений, которые могут быть отображены на больших пространствах карты.

Полученные с использованием подобного подхода карты имеют то преимущество, что обеспечивают пространственно согласованные и непрерывные значения количества биомассы, присутствующие в любой точке карты.

В работах [3, 16] для создания непрерывных карт использованы данные об отражающей способности поверхности Земли, полученные с датчиков спектрорадиометра MODIS, т. е. изображения среднего разрешения (500 м). Здесь также упоминается о том, что применение только высококачественных снимков приводит к получению изображения со значительным количеством пропусков данных. Для преодоления этого ограничения авторы [3] составили композицию из изображений, взятых за период в 2 полных календарных года.

Затем полученные на основе показателей формы сигнала лидара GLAS оценки количества углерода используются для обучения модели машинного обучения, которая оценивает биомассу как функцию показателей спектрального отражения для каждого пикселя композитного изображения.

После получения свободного от облаков композитного изображения для всех его пикселей были вычислены индексы растительности EVI2 и NDII:

$$EVI2 = 2.5 * \frac{NIR-Red}{NIR+2.4Red+1}, \quad (4)$$

где NIR соответствует каналу инфракрасного диапазона, а Red – каналу красного цвета.

$$NDII = \frac{\rho_{NIR}-SWIR}{\rho_{NIR}+SWIR}, \quad (5)$$

где ρ_{NIR} и $SWIR$ соответствуют 8 (842 нм) и 13 каналам (2190 нм) спутникового изображения от Sentinel-2.

Кроме того, для тех же наблюдений извлечены данные о температуре поверхности суши (MODIS LST) и вычислены средние значения. Также применены данные о высоте (SRTM) с разрешением 90 м.

Таким образом, для обучения модели машинного обучения «случайный лес», которая генерирует оценки плотности углерода в каждом пикселе изображения, использовались следующие переменные:

- каналы 1, 2, 4, 5, 6, 7 спутникового изображения MODIS;
- значения отклонений в 7 канале MODIS;
- средняя температура и значения ее отклонений (LST);
- EVI2;
- NDII;
- SRTM.

В работе [2] установлено, что каналы 1 (620-670 нм) и 7 (2105-2155 нм) MODIS, скорее всего (sic!), являются наиболее важными переменными для различий в плотности углерода в наземной части. В частности, канал красного цвета позволяет отличать области с растительностью от областей без нее, а канал 7 позволяет идентифицировать регионы с высокой плотностью углерода.

Также в [2] рассмотрен вопрос применения изображений из коротковолнового ИК-диапазона для приблизительной оценки объема растительности. Здесь говорится о том, что короткие ИК-волны (каналы 6 и 7 спутника MODIS, 1628-1652 нм и 2105-2155 нм соответственно) особенно чувствительны к таким структурным параметрам растительности, как высота полога, диаметр ствола, диаметр кроны, тип растительности.

Важность применения именно ИК-изображений для анализа лесного покрова подтверждается и в других работах [18, 19, 20]. В последней [20] есть замечательная иллюстрация, демонстрирующая разделение наземной растительности от хвойных деревьев:

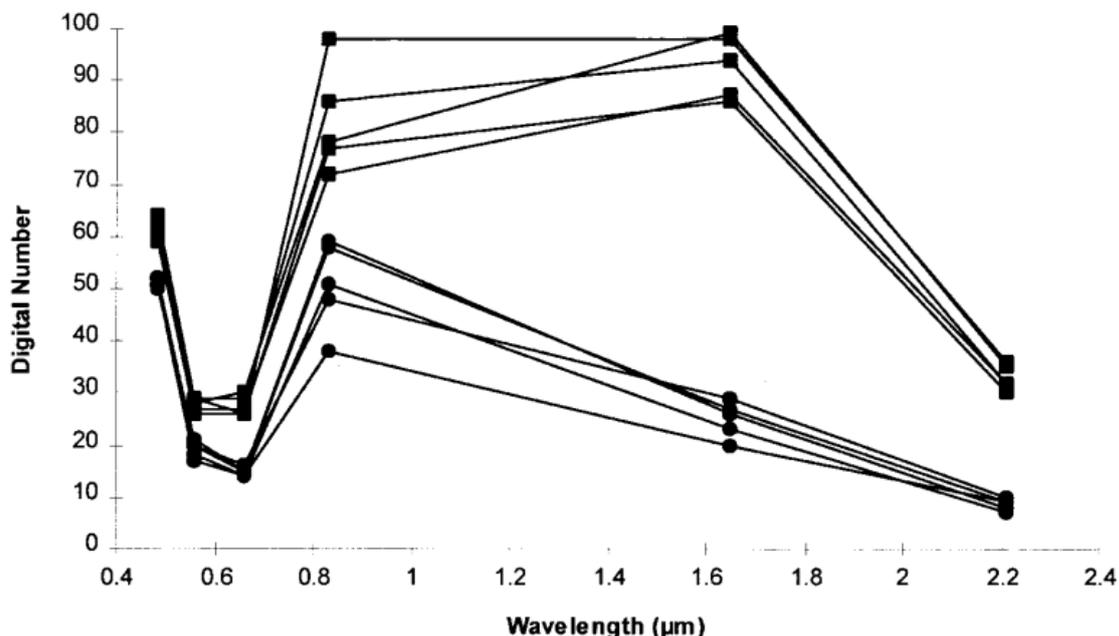


Рисунок 4. Разделимость наземной растительности от хвойных деревьев на ИК-каналах [20]

На рисунке 4 видно, что наилучшая разделимость хвойной и наземной растительности проявляется на длинах волн ~1650 нм. Как утверждается в работе [20], это обусловлено отсутствием спектральной изменчивости пикселей на ИК-изображениях, которая, однако, имеется на каналах видимого диапазона волн.

Помимо прочего, авторы [2] указали общие требования к участкам, на которых проводятся полевые измерения. Так, для всех участков, где проводились измерения, необходимо знать:

- размеры, форму и точные GPS-координаты каждого участка;
- измерения деревьев (высота стебля, диаметр);
- информацию об участках, которые располагаются более чем на одном пикселе изображения.

Здесь также говорится, что чем больше изменчивость плотности углерода в пикселе, тем больше участков необходимо взять в пределах данного пикселя. Это актуально для пикселей с изменяющимся ландшафтом (рисунок 5).

Упоминается и про необходимость использования карт с одинаковым разрешением, поскольку изменение разрешения карт может приводить к ухудшению результатов. Однако на практике доступность карт одного и того же разрешения зависит от их наличия.

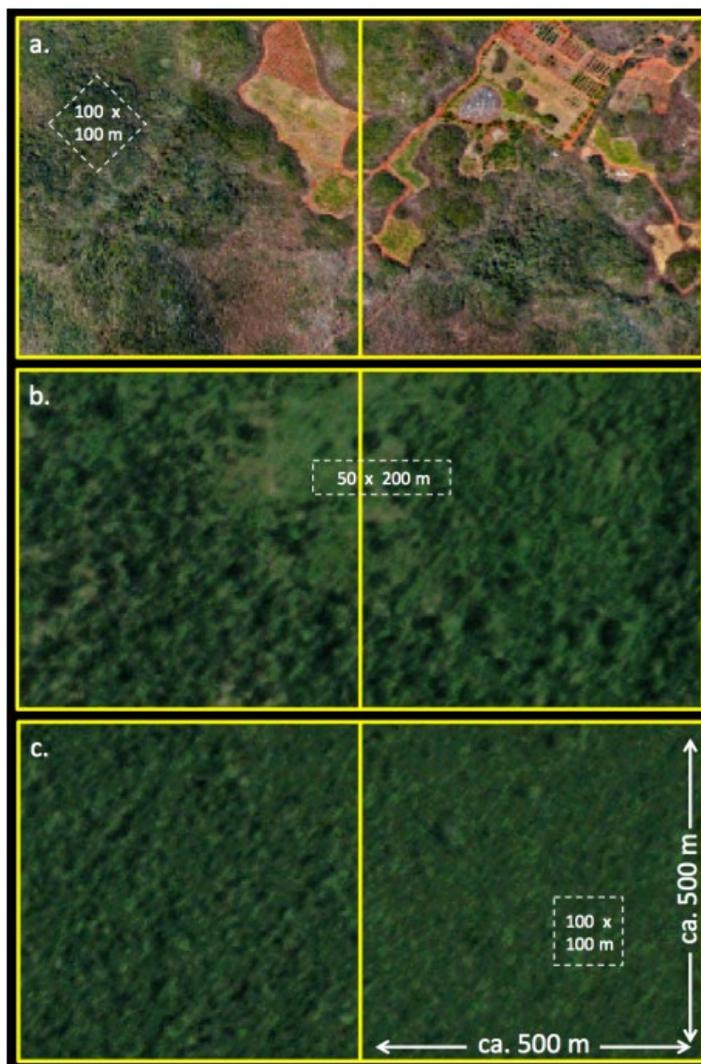


Рисунок 5. 3 различных варианта лесного ландшафта (сверху вниз): деградированный лес, деградированный густой лес, нетронутый густой лес [2]

Результаты. Как сообщается в приложении к статье [2], полученная модель случайного леса позволила объяснить 83 %, 78 % и 71 % различий в ACD на тестовых данных для тропической Америки, Африки и Азии соответственно.

В работе [6] (оценка биомассы кукурузы) проведено исследование по определению индекса растительности, дающего наибольшее значение R^2 итоговой регрессионной модели, в котором показаны очень высокие значения корреляции индексов растительности друг с другом. Также подтверждается надежность применения лидаров для оценки биомассы ($R^2 = 0,835$), а сочетание лидарных снимков с индексами растительности повышает значение R^2 ($R^2 = 0,883$).

И в работе [7] (оценка биомассы хвойных лесов) сочетание различного рода характеристик сигналов лидара приводит к повышению R^2 : так, на характеристиках лидаров дискретного возврата наибольшее достигнутое значение R^2 равно 0,702, на характеристиках лидара полной волны – 0,760, а на сочетании характеристик лидаров дискретного возврата и полной волны – 0,815.

В работе [8] (оценка надземной и подземной биомассы лесов) наибольшие значения R^2 для подземной/надземной/общей биомассы составили 0,742/0,874/0,860, 0,513/0,545/0,552 и 0,785/0,893/0,882 для только лидарных, только гиперспектральных и лидарных + гиперспектральных данных соответственно. Как видно из результатов, подземная биомасса различается хуже. Дополнительно показано, что логарифмирование данных полевых измерений при использовании регрессионных моделей не приводит к улучшению результатов.

Результаты работы [10] подтверждают результаты предыдущих работ: применение данных TDХ (мультиспектральных) и GEDI (лидар) на трех участках в Калифорнии (США), Нью-Гемпшире (США) и Коста-Рике позволило получить значения R^2 от 0,82 до 0,90, в то время как только для данных GEDI – от 0,58 до 0,90, что еще раз показывает важность совместного применения лидарных и мультиспектральных снимков для оценки биомассы.

В работе [21] полученные результаты (R^2 -коэффициент) следующие: для уравнений объема ствола с использованием только общей модели покрова, общей модели покрова с DSM (дискретный возврат) и общей модели покрова с DSM (полная волна) значения составили 0,93, 0,94 и 0,95 соответственно. А для уравнений объема биомассы – 0,87, 0,88 и 0,91 соответственно. Показано, что наиболее надежными характеристиками для оценки биомассы оказались интегральные характеристики сигналов лидара.

В работе [12] рассказывается об используемом методе классификации объектов карты по классам покрытия. При определении списка классов голая земля и застроенные территории относятся к одному и тому же классу. Применение сегментации с переменным значением масштаба (ПО eCognition (Trimble)) и алгоритма классификации kNN позволило получить значения R^2 -коэффициентов для моделей оценки биомассы в горных районах и низинах, равные 0,93 и 0,89 соответственно.

В работе [13] точность полученных моделей оценки биомассы, по утверждению авторов, не является высокой (значения R^2 -коэффициентов при этом не приведены). Они же утверждают, что в сравнении с данными бортовых лидаров данные спутниковых лидаров (работы других авторов, на которые ссылаются авторы [13]) показывают более худшие результаты, точнее углеродные карты на их основе (в данной работе использовался лидар Leica на борту самолета).

А для оценки биомассы травянистой растительности на болотах в работе [14] используется модель «случайный лес», которая обучается с применением подхода перекрестной проверки K-Fold. Для распространения стандартных ошибок средней плотности углерода применяется многомерный дельта-метод, а для вычисления стандартных ошибок по группам пикселей и типам болотистых угодий (классификация) – матрица смежности. Результаты на нескольких участках болот в Америке показали, что плотность травянистой растительности в 95 % случаев не превышает $1,56 \text{ кг/м}^2$ (разумеется, для северного региона данная цифра может отличаться). Значения общей точности классификации растительности составили от 80,5 % до 98 %. Однако для задачи регрессии полученные значения R^2 невысокие: от 0,36 до 0,61. Причем рассматривались различные комбинации индексов растительности (SAVI, NDVI, WDRVI5) и наборов данных (спутники Sentinel, Landsat).

В работе [15] для оценки биомассы растительности на городских территориях использовались BRT-модели. Значения R^2 -коэффициентов для 300 BRT-моделей составили 0,77-0,89 для данных ландшафта и 0,42-0,65 для моделей формы города.

Исходя из представленных результатов аналогичных работ, видно, что совместное применение мультиспектральных данных и данных лидара при создании карт плотности углерода приводит к наиболее точным оценкам биомассы, а углеродные карты на основе данных бортовых лидаров имеют более высокую точность, чем углеродные карты на основе спутниковых лидаров. Различий в результатах при использовании различных способов обработки сигналов лидара (построение профиля высоты или воксельный подход) и моделей регрессии (случайный лес, BRT, методы классификации и сегментации) не замечено, поэтому в нашем случае будет необходима проверка этих двух способов на более высокую точность оценки биомассы.

При оценке биомассы лесов практически во всех работах применяются регрессионные модели для оценки согласованности сигналов лидара с полевыми измерениями.

Адаптация подхода к северному региону. Как говорилось ранее, разрешение снимка в 500 метров оказывается слишком грубым для анализа региона. Протяженность ХМАО с запада на восток составляет около 1400 км (2 800 пикселей), а с севера на юг – около 820 км (1

640 пикселей). Поэтому были рассмотрены (с учетом ограничений по длине волны) альтернативные варианты спутниковых снимков, которые могут подойти в нашем случае:

Таблица 1 – Альтернативные варианты спутниковых снимков

| Спутник | Канал (длина волны, нм) | Разрешение, м |
|---------------------------|-------------------------|---------------|
| Landsat-7 | band 5 (1550-1750) | 30 |
| | band 7 (2080-2350) | 30 |
| Landsat-8 | band 6 (1560-1660) | 30 |
| | band 7 (2100-2300) | 60 |
| WorldView3 (коммерческий) | SWIR-3 (1640-1680) | 3,7 |
| | SWIR-5 (2145-2185) | 3,7 |
| | CAVIS Snow (1620-1680) | 30 |
| | Aerosol-3 (2105-2145) | 30 |
| Sentinel 2 | B4 (665) | 10 |
| | B11 (1610) | 20 |
| | B12 (2190) | 20 |

Данные каналы удовлетворяют предложенному в работе [10] оптимальному значению разрешения спутниковых изображений в 100 м.

Также следует иметь в виду отличие видов растительности северного региона от видов растительности экваториальных районов Земли, для которых, собственно, и проводились описанные в работах [2, 3] исследования.

Что касается бортовых лидаров, то возможности их применения в нашей работе, скорее всего, не будет, хотя углеродные карты на основе данных бортовых лидаров и показывают большую точность по сравнению со спутниковыми лидарами.

Предлагаемая программа исследований

Исходя из анализа представленной литературы, предлагается выполнение следующих исследований по созданию углеродной карты ХМАО:

1. Исследование лидарных данных GLAS и мультиспектральных данных представленных в таблице 1 спутников. Цель – определение наиболее подходящих и одновременно доступных спутниковых изображений.
2. Определение классов земного покрова и участков, соответствующих пятнам лидара. Проведение полевых измерений. Цель – определение средних значений биомассы на участках для разных классов земного покрова.
3. Исследование различных способов обработки сигналов лидара и регрессионных моделей перевода характеристик сигнала в значения биомассы. Цель – получение способа обработки сигналов лидара, дающего наибольшую точность вычисления биомассы.
4. Исследование моделей классификации земного покрова по мультиспектральным спутниковым изображениям. Цель – получение модели классификации земного покрова с наибольшей точностью и создание на ее основе углеродной карты ХМАО.

Заключение

В данной работе проанализирован один из подходов к построению карт плотности углерода, которые предназначены для оценки объема растительности и содержащегося в ней углерода. Подход основан на применении биометрических данных (измерение высоты деревьев и диаметра ствола), спутниковых снимков (оптических, инфракрасных и лидарных), расчете и анализе групп пикселей с помощью множественной линейной регрессии. Данный подход применен для построения карты плотности углерода для экваториальных районов Африки, Азии, Южной Америки.

В работах [2-3] приведено подробное описание используемых математических методов перевода биометрических данных в значения плотности углерода, указаны требования к ис-

пользуемым данным, а также приведены результаты тестирования алгоритма и другие рекомендации по использованию данных.

Помимо указанных работ проведен детальный анализ аналогичных существующих подходов к оценке биомассы с помощью дистанционного зондирования Земли. На основе анализа определены основные методы и подходы, используемые как для оценки биомассы лесов, так и для оценки биомассы травяной растительности болот (которых достаточно много на территории ХМАО) и оценки биомассы растительности в городских территориях. Полученный обзор методов и подходов позволил составить программу исследований именно для нашего случая.

Также спланированы шаги по адаптации подхода из работы [3] к одному из регионов Крайнего Севера России – ХМАО. Так, выяснилось, что для составления карты углерода в масштабе региона необходимо использовать изображения большего разрешения. Поэтому был произведен поиск действующих на январь 2022 года спутников, а также подходящих нам каналов съемки. Что касается лидарных данных, то для ХМАО имеется более 5000 файлов со спутника ICESat-2 за более чем 3 года. Общий объем – около 11 Тб.

Также необходимо учитывать различие в видах растительности для экваториальных лесов и тайги (тундры), в зоне которой расположен ХМАО. Для них придется обучать новые регрессионные модели, которые, вероятнее всего, будут отличаться набором переменных и коэффициентами при них.

Список литературы

1. Государственный доклад «О состоянии и об охране окружающей среды Российской Федерации в 2020 году». – Текст : электронный // Министерство природных ресурсов и экологии Российской Федерации. – 2021. – URL: https://www.mnr.gov.ru/docs/gosudarstvennye_doklady/gosudarstvennyu_doklad_o_sostoyanii_i_ob_okhrane_okruzhayushchey_sredy_rossiyskoy_federatsii_v_2020/?PAGEN_2=2 (дата обращения: 14.01.2022).
2. Tropical forests are a net carbon source based on aboveground measurements of gain and loss / A. Baccini, W. Walker, L. Carvallo [et al.] // *Science*. – 2017. – Vol. 358, № 6360. – P. 230–234.
3. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps / A. Baccini, S. J. Goetz, W. S. Walker [et al.] // *Nature climate change*. – 2012. – Vol. 2, № 3. – P. 182–185.
4. ICESat. Cryospheric Sciences Lab // NASA. – 2021. – URL: <https://icesat.gsfc.nasa.gov/icesat/glas.php> (date of application: 20.01.2022).
5. ATLAS/ICESat-2 L2A Global Geolocated Photon Data, Version 5 // National Snow and Ice Data Center. – 2021. – URL: <https://nsidc.org/data/ATL03/versions/5> (date of application: 20.01.2022).
6. Estimating the biomass of maize with hyperspectral and LiDAR data / C. Wang, S. Nie, X. Xiaohuang [et al.] // *Remote Sensing*. – 2017. – Vol. 9, № 1. – P. 11.
7. Above-ground biomass estimation using airborne discrete-return and full-waveform LiDAR data in a coniferous forest / S. Nie, C. Wang, H. Zeng [et al.] // *Ecological Indicators*. – 2017. – Vol. 78. – P. 221–228.
8. Fusion of airborne LiDAR data and hyperspectral imagery for aboveground and below-ground forest biomass estimation / S. Luo, C. Wang, X. Xiaohuang [et al.] // *Ecological Indicators*. – 2017. – Vol. 73. – P. 378–387.
9. Non-destructive aboveground biomass estimation of coniferous trees using terrestrial LiDAR / A. Stovall, A. Voster, R. Anderson [et al.] // *Remote Sensing of Environment*. – 2017. – Vol. 200. – P. 31–42.
10. Forest biomass estimation over three distinct forest types using TanDEM-X InSAR data and simulated GEDI lidar data / W. Qi, S. Saarela, J. Armston [et al.] // *Remote Sensing of Environment*. – 2019. – Vol. 232. – P. 111283.
11. Global Ecosystem Dynamics Investigation // Wikipedia. – 2022. – URL: https://en.wikipedia.org/wiki/Global_Ecosystem_Dynamics_Investigation (date of application: 11.02.2022).

12. Impact of land cover change on aboveground carbon stocks in Afromontane landscape in Kenya / P. K. E. Pellikka, V. Heikinheimo, J. Hietanen [et al.] // *Applied Geography*. – 2018. – Vol. 94. – P. 178–189.
13. Estimating aboveground carbon density and its uncertainty in Borneo's structurally complex tropical forests using airborne laser scanning / T. Jucker, G. A. Asner, M. Dalponte [et al.] // *Biogeosciences*. – 2018. – Vol. 15, № 12. – P. 3811–3830.
14. A remote sensing-based model of tidal marsh aboveground carbon stocks for the conterminous United States / K. B. Byrd, L. Ballanti, N. Thomas [et al.] // *ISPRS Journal of Photogrammetry and Remote Sensing*. – 2018. – Vol. 139. – P. 255–271.
15. Identification of fine scale and landscape scale drivers of urban aboveground carbon stocks using high-resolution modeling and mapping / M. G. E. Mitchell, K. Johansen, M. Maron [et al.] // *Science of the total Environment*. – 2018. – Vol. 622. – P. 57–70.
16. Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest / J. B. Drake, R. Dubayah, R. G. Knox [et al.] // *Remote Sensing of Environment*. – 2002. – Vol. 81, № 2-3. – P. 378–392.
17. Laser altimeter canopy height profiles: Methods and validation for closed-canopy, broad-leaf forests / D. J. Harding, M. A. Lefsky, G. Parker, J. B. Blair // *Remote Sensing of Environment*. – 2001. – Vol. 76, № 3. – P. 283–297.
18. Cohen, W. B. Estimating structural attributes of Douglas-Fir / W. B. Cohen, T. A. Spies // *Remote sensing of environment*. – 1992. – Vol. 41, № 1. – P. 1–17.
19. Gemmell, F. M. Effects of forest cover, terrain, and scale on timber volume estimation with Thematic Mapper data in a Rocky Mountain site / F. M. Gemmell // *Remote Sensing of Environment*. – 1995. – Vol. 51, № 2. – P. 291–305.
20. Puhr, C. B. Remote sensing of upland conifer plantations using Landsat TM data: a case study from Galloway, south-west Scotland / C. B. Puhr, D. N. M. Donoghue // *International Journal of Remote Sensing*. – 2000. – Vol. 21, № 4. – P. 633–646.
21. Stem volume and above-ground biomass estimation of individual pine trees from LiDAR data: Contribution of full-waveform signals / T. Allouis, S. Durrieu, C. Vega [et al.] // *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. – 2012. – Vol. 6, № 2. – P. 924–934.
22. Inversion of a lidar waveform model for forest biophysical parameter estimation / B. Koetz, F. Morsdorf, G. Sun [et al.] // *IEEE Geoscience and Remote Sensing Letters*. – 2006. – Vol. 3, № 1. – P. 49–53.