

ГЕНЕЗИС И ГЕОГРАФИЯ ПОЧВ

УДК 630.114.351:528.88

ГЕОПРОСТРАНСТВЕННОЕ МОДЕЛИРОВАНИЕ СОДЕРЖАНИЯ И ЗАПАСОВ АЗОТА И УГЛЕРОДА В ЛЕСНОЙ ПОДСТИЛКЕ НА ОСНОВЕ РАЗНОСЕЗОННЫХ СПУТНИКОВЫХ ИЗОБРАЖЕНИЙ SENTINEL-2

© 2021 г. Е. А. Гаврилюк^а, *, А. И. Кузнецова^а, А. В. Горнов^а^аЦентр по проблемам экологии и продуктивности лесов РАН, ул. Профсоюзная, 84/32, стр. 14, Москва, 117997 Россия

*e-mail: egor@ifi.rssi.ru

Поступила в редакцию 02.04.2020 г.

После доработки 20.06.2020 г.

Принята к публикации 22.06.2020 г.

Показаны возможности использования оптических мультиспектральных спутниковых данных Sentinel-2 для моделирования содержания азота (N) и углерода (C), их отношения (C : N) и запасов в лесной подстилке. Исследование проводили на территории заповедника “Брянский лес” и его охранной зоны. Образцы подстилки отбирали на 33 наземных пробных площадях, заложенных с учетом формационного разнообразия лесов заповедника. Два подгоризонта подстилки (L и FH) рассматривали независимо друг от друга. Основные переменные для геопространственного моделирования получали на основе временной серии из восьми разносезонных изображений Sentinel-2. К ним добавляли базовые характеристики рельефа местности и координаты положения пикселей. В работе использовали алгоритм случайных лесов для построения регрессионных моделей и соответствующий комплекс стандартных методов для оценки их эффективности. Наилучшие результаты получены для величин C : N – коэффициент детерминации $R^2 = 0.71$ при относительной ошибке RMSE = 12.5% в подгоризонте L и $R^2 = 0.83$ при RMSE = 10.6% в подгоризонте FH. Для остальных моделей значения R^2 варьировали от 0.23 до 0.61, а RMSE – от 15.8 до 48.6%, с наименее надежными результатами для показателей запасов. Спутниковые переменные были наиболее информативны при моделировании содержания N и C, особенно C : N. Наиболее значимыми периодами во временной серии были ранняя весна, лето и снежная зима. Спутниковые изображения Sentinel-2 могут быть успешно использованы для оценки и картографирования содержания и запасов N и C в лесной подстилке в качестве доступной и актуальной альтернативы тематическим данным, характеризующим видовой состав и связанные с ним свойства древостоев.

Ключевые слова: лесные почвы, хвойно-широколиственные леса, отношение C : N, данные ДЗЗ, случайные леса, машинное обучение

DOI: 10.31857/S0032180X21020040

ВВЕДЕНИЕ

Лесная подстилка, являясь продуктом функционирования лесных биогеоценозов, регулирует широкий спектр экосистемных процессов. Характеризующие качество лесного опада показатели содержания азота (N) и углерода (C), а также величина их отношения (C : N) – надежные индикаторы скорости процессов минерализации, от которых напрямую зависит продуктивность лесных почв и их способность к депонированию углерода [10, 18, 52]. В условиях глобального изменения климата большое внимание уделяется уточнению оценок запасов почвенного органического вещества. Показано, что доля подстилки в общих запасах углерода может достигать 30% [9, 19].

Данные дистанционного зондирования Земли (ДЗЗ) в настоящее время находят широкое применение в качестве основы для цифрового карто-

графирования типов почв и геопространственного моделирования их количественных характеристик [15]. Концептуальным примером совместного использования спутниковых и наземных данных для получения тематических продуктов базовых свойств почв методами машинного обучения является глобальная информационная система SoilGrids [34]. Стоит отметить, что в большинстве исследований, посвященных геопространственному моделированию почвенных показателей, переменные, полученные на основе данных ДЗЗ, используются совместно с климатическими, орографическими, геологическими и прочими тематическими признаками, и при этом часто не демонстрируют высокой информативности [33, 44, 48].

Потенциал использования спутниковых изображений, как оптических, так и радарных, для

оценки свойств почв, особенно лесных, часто бывает ограничен из-за наличия плотного растительного покрова. Сомкнутость полога более 30% исключает возможность прямого анализа спектральных характеристик подстилающей поверхности [29]. Однако отдельные свойства почвы могут находиться в тесной взаимосвязи с качественными и/или количественными характеристиками растительности, что оставляет возможности для опосредованной оценки таких свойств. В частности, древесные растения доминирующих видов, формирующих леса разных типов, оказывают значительное влияние на состав органического вещества, кислотность, общее содержание N и величину C : N лесных почв, причем как в органогенном, так и в верхнем (0–10 см) минеральном слоях [27, 43]. Поэтому тематические характеристики лесного покрова (преобладающие породы древостоев, доля хвойных и лиственных деревьев, тип леса, сомкнутость полога, наземная фитомасса и др.) часто используются в качестве переменных при геопространственном моделировании различных свойств лесных почв, включая C : N и запасы C в подстилке [21, 24, 25].

В то же время распознавание и картографирование породной структуры лесов — одна из классических задач, решаемая с применением данных ДЗЗ с разной степенью успешности [31], как и оценка биометрических и структурных характеристик древостоев [41]. Соответственно каналы спутниковых изображений и/или их производные потенциально способны в той или иной степени заменить тематические переменные, отвечающие за вариабельность характеристик лесного покрова. Поскольку свойства опада напрямую связаны с видовым составом древостоев, опосредованные взаимосвязи со спектральными признаками древесного полога потенциально должны демонстрировать органогенный горизонт лесных почв. При этом, хотя существуют исследования, направленные на определение площади покрытия [39], влажности [51] и степени разложения [45] опада по данным ДЗЗ, нам не удалось найти похожих работ, посвященных оценкам содержания C и N в лесной подстилке.

Настоящая работа посвящена исследованию возможности использования временной серии оптических мультиспектральных спутниковых изображений высокого пространственного разрешения (совместно с базовыми характеристиками рельефа местности) для геопространственного моделирования содержания C и N в лесной подстилке методами машинного обучения на территории заповедника «Брянский лес» и его охранной зоны. Подгоризонты подстилки L (свежий или слабо разложившийся опад) и FH (слой ферментации и разложения растительных остатков) рассматривали независимо друг от друга. Для каждого из них проанализировали пять количественных показате-

телей: содержание (%) и запас (г/м²) органических C и N, а также величину их отношения по содержанию — C : N.

ОБЪЕКТЫ И МЕТОДЫ

Район исследования. Исследование проводили в южной части Брянского полесья в пределах государственного природного биосферного заповедника «Брянский лес» и его охранной зоны (рис. 1), на 97% покрытых лесами, общая площадь которых составляет более 200 км².

В ботанико-географическом плане район относится к Полесской подпровинции Восточноевропейской провинции Европейской широколиственно-лесной области [14]. Леса заповедника с давних времен подвергались хозяйственной деятельности. В результате современный лесной покров территории представлен, главным образом, раннесукцессионными сообществами с монодоминантным (сосняки, березняки, черноольшаники и осинники) или олигодоминантным (смешанные хвойно-лиственные и мелколиственные леса) составом древостоев [1, 3, 4]. При этом в заповеднике сохранились уникальные участки полидоминантных хвойно-широколиственных и широколиственных лесов [2, 30].

Климат территории умеренно континентальный, с четырьмя выраженными фенологическими сезонами в течение года. Средняя многолетняя температура воздуха для зимних месяцев составляет –5.2°C, для летних — +18.4°C, среднегодовое количество осадков — 556 мм, наибольшая часть которых (33%) приходится на летний период [12].

По почвенно-географическому районированию заповедник входит в Среднерусскую провинцию подзолистых почв [6]. В ландшафтной структуре территории выделяют участки пойменного, террасного, полесского, предполесского ландшафтов [5]. В поймах преобладают серогумусовые органо-аккумулятивные почвы [8] (Umbrisols по WRB [49]) — на аллювиальных отложениях [17]. В пределах террасного, полесского и предполесского ландшафтов преобладают дерново-подзолы [8] (Albic Podzols (Arenic)) — с выраженным гумусовым горизонтом на флювиогляциальных песках (супесях) и глинистом элювии кремнистой опоки [16]. Повсеместно встречаются признаки процессов оглеения. На менее нарушенных территориях заповедника отмечаются признаки буроземообразования [7].

Исходные данные. В работе использовали данные трех типов: результаты наземных обследований, спутниковые изображения и цифровую модель рельефа (ЦМР) местности с ее производными. Кроме того, пространственные координаты положения центров пикселей изображений (в виде порядковых номеров по горизонтали и верти-

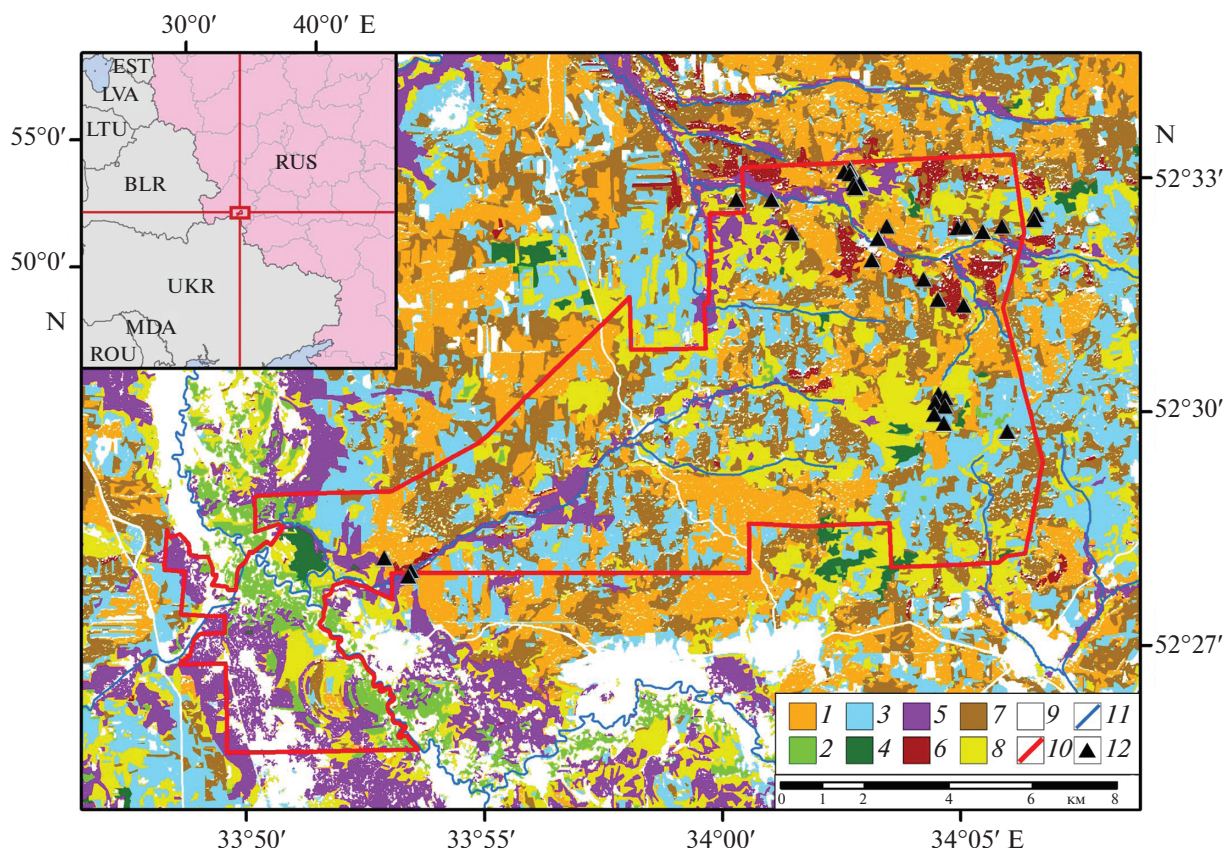


Рис. 1. Район исследования. Типы древостоев по видовому составу (из [1]): 1 – сосняки, 2 – широколиственные, 3 – березняки, 4 – осинники, 5 – черноольшаники, 6 – смешанные хвойные, 7 – смешанные хвойно-лиственные, 8 – смешанные лиственные. Прочие обозначения: 9 – безлесные территории, 10 – граница заповедника “Брянский лес”, 11 – водотоки, 12 – места закладки наземных пробных площадей.

кали) добавляли в качестве двух вспомогательных независимых переменных для геопространственного моделирования, что является распространенной практикой. Стоит отметить, что компактность района исследований делает нецелесообразным привлечение климатических данных в качестве дополнительных переменных, а актуальные карты почв достаточной детальности (масштаба порядка 1 : 100 000), которые могли бы быть использованы в работе, на территорию заповедника отсутствуют.

Полевые данные и их предварительная обработка. В ходе полевых исследований 2016–2017 гг. в лесах заповедника и его охранный зоны было заложено 33 наземных пробных площадки (НПП) по 400 м². Места для закладки НПП (рис. 1) выбирали в однородных по видовой структуре древесного яруса участках лесов таким образом, чтобы в выборке был представлен опад из всех характерных для территории исследования типов древостоев. Для оценки породного состава насаждений использовали карту (далее карта пород), полученную в результате тематической обработки спутниковых данных Landsat [1], и материалы лесо-

устройства заповедника 2006 г. Для каждого из восьми тематических классов карты пород закладывали от трех до шести НПП. На каждой НПП выполняли геоботаническое описание с выявлением полного флористического состава с учетом ярусной структуры леса. В каждом ярусе определяли участие видов по шкале обилия-покрытия Браун-Бланке [11]. На всех НПП производили отбор лесной подстилки с использованием рамки размером 25 × 25 см в трех- или четырехкратной повторности (всего 108 образцов).

В лабораторных условиях отобранные образцы высушивали до абсолютно-сухого состояния при 105°C, взвешивали и оценивали содержание N и C на элементном анализаторе EA1110 (CHNS-O). При расчете запасов углерода пользовались методическими указаниями по количественному определению объема поглощения парниковых газов [13]. В итоге оценено по пять показателей для L- и FH-подгоризонтов лесной подстилки: содержание N (N_%), содержание C (C_%), отношение C : N, запас N (N_{stk}) и запас C (C_{stk}). Базовый статистический анализ полученных измерений (табл. 1) свидетельствует о достаточной однородности (коэф-

Таблица 1. Описательная статистика для наземных данных, над чертой — значения для подгоризонта L, под чертой — для подгоризонта FH

Показатель	Среднее	v	КВ	Минимум	1-й квартиль	Медиана	3-й квартиль	Максимум	Мощность*
$N_{\%}$	<u>2.2</u>	<u>0.4</u>	<u>19.5</u>	<u>1.1</u>	<u>1.9</u>	<u>2.2</u>	<u>2.4</u>	<u>3.1</u>	<u>88.4</u>
	2.0	0.5	25.3	1.0	1.6	1.9	2.3	3.5	85.4
$C_{\%}$	<u>44.4</u>	<u>3.1</u>	<u>6.9</u>	<u>32.8</u>	<u>42.9</u>	<u>44.9</u>	<u>46.3</u>	<u>49.5</u>	<u>92.3</u>
	37.5	7.1	18.9	18.1	32.0	39.0	43.1	49.8	80.8
$C : N$	<u>21.3</u>	<u>5.4</u>	<u>25.4</u>	<u>14.5</u>	<u>17.5</u>	<u>20.3</u>	<u>23.5</u>	<u>39.4</u>	<u>82.6</u>
	19.5	5.2	26.8	12.1	15.9	17.2	22.2	36.8	84.3
$N_{stk}, \text{г/м}^2$	<u>8.7</u>	<u>6.0</u>	<u>68.4</u>	<u>1.2</u>	<u>5.0</u>	<u>6.9</u>	<u>11.2</u>	<u>30.3</u>	<u>59.3</u>
	20.8	11.3	54.2	3.1	11.3	19.9	28.4	58.8	87.2
$C_{stk}, \text{г/м}^2$	<u>185.4</u>	<u>127.2</u>	<u>68.6</u>	<u>21.1</u>	<u>88.9</u>	<u>145.5</u>	<u>268.3</u>	<u>712.8</u>	<u>92.1</u>
	415.3	262.7	63.3	46.3	196.6	344.8	583.0	1425.1	90.4

Примечание. v — стандартное отклонение, КВ — коэффициент вариации показателя (%). * Мощности выборки (%), оцененная по стандартному t -тесту при уровне значимости $\alpha = 0.1$ и величине эффекта, равной 10% от разброса значений.

Таблица 2. Фенологические периоды, для которых формировались разносезонные спутниковые композитные изображения Sentinel-2

Период	Границы периода	Длина периода, дни
Зима	26 октября—30 марта	166
Начало весны	31 марта—30 апреля	31
Середина весны	1—15 мая	15
Конец весны	16—30 мая	15
Лето	31 мая—24 июля	45
Начало осени	25 июля—3 сентября	41
Середина осени	4 сентября—4 октября	31
Конец осени	5 октября—25 октября	21

коэффициент вариации $<30\%$) и представительности (мощность более 80%) выборок для показателей $N_{\%}$, $C_{\%}$ и $C : N$. В отношении запасов N и C выборки неоднородны, что может быть следствием недостаточного числа повторностей на пробных площадях, но, тем не менее, представительны для всех показателей, кроме N_{stk} в подгоризонте L (мощность около 60%).

Спутниковые данные и их предварительная обработка. В качестве основного источника данных для геопространственного моделирования использовали разносезонные мультиспектральные изображения Sentinel-2 [28]. По результатам анализа ежегодных глобальных данных MODIS о динамике наземного покрова (MCD12Q2) версии 6 [32] для района исследования выделяли восемь последовательных фенологических периодов (табл. 2), для каждого из которых формировали композитные изображения из сцен Sentinel-2 2016–2018 гг.

MCD12Q2 — это набор ежегодных глобальных тематических изображений пространственным разрешением 500 м за период с 2000 по 2017 гг.,

которые содержат информацию на попиксельном уровне о семи ключевых датах в динамике хода кривой спектрального индекса EVI2 [35] и степени достоверности их определения. Ключевые даты соответствуют началу и середине роста, выходу на плато, пику, началу, середине и окончанию убывания значений индекса. Для локальных территорий, относительно однородных по климатическим условиям, как в нашем случае, медианные значения этих дат, оцененные по всем валидным пикселям изображения, могут быть достаточно надежно соотнесены с последовательной сменой фенологических фаз зеленой растительности. Для определения границ четырех основных периодов года (зима, весна, лето, осень) использовали данные 2017 г., после чего в пределах весны и осени аналитически выделяли по три дополнительных периода (начало, середина и конец). Границы дополнительных периодов подбирали таким образом, чтобы они располагались симметрично относительно исходных дат середины роста и убывания значений индекса EVI2, при этом длина каждого из периодов составляла не менее 15 дней.

Для всех периодов, кроме зимы, отбирали по три сцены Sentinel-2 соответствующего временного диапазона с минимальным облачным покровом из всех доступных в архиве. Композитные изображения периодов формировали из медианных значений пикселей отобранных сцен, рассчитанных независимо для каждого спектрального канала. Для периода снежной зимы использовали одиночную, полностью безоблачную сцену. Все отобранные изображения Sentinel-2 предварительно преобразовывали в продукты уровня L2A (значения коэффициентов отражения на уровне земной поверхности) с использованием программного модуля Sen2Cor [40].

ЦМР и ее производные. Базовые орографические характеристики местности получали из ЦМР пространственным разрешением 10 м, сформированной в результате интерполяции значений высот по горизонталям топографической карты масштаба 1 : 50 000 (высота сечения рельефа 5 м). В качестве геопространственных переменных использовали показатели абсолютной высоты, крутизны, экспозиции (как синус и косинус ее угловых значений) и общей кривизны склонов [46], а также топографического индекса влажности [22], рассчитанные стандартными средствами ГИС SAGA [26].

Формирование исходного набора переменных. Принимая во внимание высокую корреляцию между каналами Sentinel-2, последовательно расположенными в видимом, ближнем инфракрасном и среднем инфракрасном участках спектра, в работе использовали классический анализ главных компонент [36] для сокращения пространства признаков. В нашем случае, первые две главных компоненты, в зависимости от периода съемки, описывали от 88 (конец осени) до 98% (снежная зима) вариации, содержащейся в десяти основных (с порядковыми номерами 2-8, 8А, 11 и 12) каналах Sentinel-2, что позволяет добиться пятикратного сжатия данных без значительных потерь в информативности. Шесть каналов с пространственным разрешением 20 м (5-7, 8А, 11 и 12) предварительно приводили к разрешению 10 м методом ближайшего соседа.

В итоге после преобразования изображений методом главных компонент, 16 признаков, полученных на основе спутниковых данных (по две главных компоненты для каждого из восьми фенологических периодов), в сочетании с шестью орографическими характеристиками местности и двумя координатами положения пикселей составили исходный набор геопространственных переменных для моделирования. Учитывая уровень точности привязки как спутниковых, так и полевых данных (средняя погрешность порядка 10 м), измерения с НПП сопоставляли с медианными значениями переменных в окне размером 3×3 пикселя (30×30 м) вокруг точки заложения площадок.

Оптимизация набора переменных. Исходный набор переменных был оптимизирован сначала с применением корреляционного анализа, а затем на основе метода рекурсивного исключения признаков. При корреляционном анализе для всех переменных оценивали попарную корреляцию Пирсона (r), после чего из пар со значением $r > 0.95$ отбрасывали переменные с более высоким средним значением r , рассчитанным для каждой переменной по всем ее парам. Процедура рекурсивного исключения признаков подразумевает последовательное построение регрессионных моделей (с фиксированными параметрами алгоритма) для анализируемого показателя с поэтапным отсеиванием наименее информативных переменных. При этом на каждом этапе отсеивания проводится оценка эффективности модели, что позволяет сформировать оптимальный набор наиболее значимых признаков для каждого из показателей, обеспечивающий наилучшую точность с точки зрения выбранного формального критерия. В качестве такого критерия использовали величину средней квадратической ошибки (СКО) модели, оцениваемой методом повторной кросс-валидации (25 повторов с разбиением исходной выборки на 4 части). По результатам процедуры из дальнейшего анализа исключали только те переменные, которые не попали в число оптимальных ни для одного из рассматриваемых показателей, таким образом, для всех моделей использовался единый набор признаков.

Построение регрессионных моделей. В работе использовали алгоритм машинного обучения “случайные леса” [23] для построения регрессионных моделей, а также весь сопутствующий ему комплекс методов для автоматического подбора параметров алгоритма, оценки информативности переменных, качества обучения и эффективности предсказаний, реализованный в программной среде R в пакетах *caret* [38] и *ranger* [50].

Случайные леса — статистический метод для задач классификации и регрессии, основанный на использовании большого числа (ансамбля) деревьев решений, каждое из которых строится по неполной выборке, получаемой из исходной с помощью бутстрепа (случайной выборки с возвращением), а для расщепления вершин используется фиксированное число переменных, случайно отбираемых из полного набора. В базовом варианте алгоритма классификация осуществляется с помощью простого голосования классификаторов, определяемых отдельными деревьями, а регрессионное моделирование — путем осреднения результатов по всем деревьям. В настоящее время случайные леса являются одним из наиболее популярных методов машинного обучения, так как сочетают в себе (относительную) универсальность, простоту настройки и быстроту работы с

высокими показателями эффективности получаемых моделей.

Использовали ансамбль из 1000 деревьев, а подбор прочих параметров алгоритма, в частности — числа случайных признаков при каждом расщеплении дерева, метода расщепления и минимального размера узла, — осуществляли индивидуально для каждого моделируемого показателя простым перебором вариантов аналогично процедуре рекурсивного исключения признаков.

Оценка эффективности и статистической значимости моделей. Эффективность полученных моделей оценивали по стандартным статистическим метрикам — коэффициенту детерминации (R^2) и квадратному корню из СКО (RMSE), а также относительным величинам RMSE — процентным значениям от среднего (RMSE_{AVG}) и разброса (RMSE_{RNG}) моделируемых показателей. При этом использовали как оценки, полученные на основе полной исходной выборки (33 измерения) интегрированным в случайные леса методом out-of-bag (OOB), так и предварительное разбиение выборки на обучающую и контрольную (в соотношении 25/8). OOB-метод подразумевает формирование отдельных бутстреп-выборок для обучения каждого из деревьев ансамбля, что позволяет использовать не вошедшие в них измерения для оценки эффективности отдельных деревьев, а затем и всей модели путем осреднения результатов. Измерения для контрольной выборки подбирали аналитически таким образом, чтобы избежать их попадания на края распределения для всех анализируемых показателей и обеспечить представительство всех типов древостоев.

Статистическую значимость моделей в целом (p -значение) оценивали методом, описанным в [20], который представляет собой множественный пермутационный тест (в нашем случае из 200 итераций), для зависимой переменной (величины, для которой выполняется моделирование). Процедура пермутации подразумевает случайную перестановку значений переменной, после чего происходит подгонка модели и оценка эффективности (в нашем случае по СКО). Доля случаев, когда пермутационные модели оказываются эффективнее исходной, характеризует искомое p -значение.

Выбор наилучшего типа прогнозирования. На основе абсолютной величины RMSE оценивали три типа результатов прогнозирования моделей — среднее, скорректированное среднее или медиану, — с целью определения наиболее эффективного варианта для каждого моделируемого показателя. Первый тип, который является стандартным для регрессий, построенных с помощью случайных лесов, — это осреднение результатов по всем деревьям ансамбля. Второй тип получается из значений первого путем применения к ним простой

линейной модели (предсказанные значения подгоняются к исходным измерениям), призванной скорректировать характерный эффект завышения низких и занижения высоких значений. Для построения этой дополнительной модели использовали метод повторных медиан [47], реализованный в R-пакете mblm [37]. Третий тип — это медианные значения (в 50-м перцентиле), получаемое при построении квантильной регрессии в ее реализации для случайных лесов [42].

Оценка информативности переменных. В качестве меры информативности переменных в процессе обучения использовали стандартный для случайных лесов показатель среднего уменьшения общей точности модели (MDA — Mean Decrease in Accuracy) после случайной перестановки значений оцениваемой независимой переменной. В отличие от похожего пермутационного теста для статистической значимости, описанного выше, здесь перестановка значений, переподгонка модели и оценка точности происходят однократно для каждой переменной. Критерием точности модели была выбрана величина СКО.

Геопространственное моделирование. Для получения тематических продуктов, характеризующих распределение анализируемых показателей по территории исследования, к оптимизированному набору переменных на попиксельном уровне применяли соответствующие модели с наиболее эффективным типом прогнозирования, обученные по полной выборке. Одновременно с помощью квантильной регрессии оценивали границы доверительного интервала шириной 80% — значения в 10- и 90-м перцентилях распределения предсказаний всех деревьев ансамбля. Разница этих значений относительно результата прогнозирования характеризует величины предельных ошибок моделирования, соответственно в сторону занижения и завышения, в рамках заданного доверительного интервала. В качестве пространственной меры неопределенности моделирования для полученных тематических продуктов использовали отношение ширины доверительного интервала (разность значений на его границах) к результату прогнозирования, выраженное в процентах.

РЕЗУЛЬТАТЫ

Оптимальные переменные для моделирования. По результатам корреляционного анализа исходных переменных были отброшены главные компоненты конца весны и начала осени, которые были сильно коррелированы с аналогичными переменными прилегающих к ним фенологических периодов. После процедуры рекурсивного исключения признаков были отсеяны еще шесть переменных: главные компоненты конца осени, пока-

Таблица 3. Сводная статистика эффективности регрессионных моделей, над чертой — значения для подгоризонта L, под чертой — для подгоризонта FH

Показатель	N _%		C _%		C : N		N _{stk} , г/м ²		C _{stk} , г/м ²	
	тип оценки*									
	ООВ	тест	ООВ	тест	ООВ	тест	ООВ	тест	ООВ	тест
Тип прогноза**	<u>Ср.</u>	<u>Ср.</u>	<u>Мед.</u>	<u>ЛМ</u>	<u>Мед.</u>	<u>Мед.</u>	<u>Мед.</u>	<u>Ср.</u>	<u>Мед.</u>	<u>ЛМ</u>
	Мед.	Ср.	Мед.	Мед.	Ср.	ЛМ	Мед.	Мед.	Ср.	Ср.
R ²	<u>0.48</u>	<u>0.57</u>	<u>0.46</u>	<u>0.46</u>	<u>0.71</u>	<u>0.77</u>	<u>0.58</u>	<u>0.49</u>	<u>0.61</u>	<u>0.54</u>
	0.44	0.12	0.56	0.44	0.83	0.95	0.23	0.30	0.37	0.24
RMSE	<u>0.3</u>	<u>0.3</u>	<u>1.8</u>	<u>1.8</u>	<u>2.7</u>	<u>1.9</u>	<u>3.4</u>	<u>3.5</u>	<u>72.3</u>	<u>77.8</u>
	0.4	0.4	4.2	3.8	2.1	0.8	9.2	9.1	190.9	229.8
RMSE _{RNG} , %	<u>15.8</u>	<u>24.4</u>	<u>16.8</u>	<u>35.2</u>	<u>12.5</u>	<u>15.1</u>	<u>15.5</u>	<u>25.5</u>	<u>14.3</u>	<u>28.1</u>
	17.7	38.6	17.5	25.6	10.6	8.6	18.4	29.6	20.7	31.3
RMSE _{AVG} , %	<u>12.5</u>	<u>11.5</u>	<u>4.0</u>	<u>4.0</u>	<u>12.8</u>	<u>9.2</u>	<u>39.8</u>	<u>43.0</u>	<u>40.3</u>	<u>45.4</u>
	18.1	19.1	11.1	9.8	10.9	4.3	46.0	42.9	48.6	54.9
p <	<u>0.005</u>	<u>0.005</u>	<u>0.005</u>	<u>0.005</u>	<u>0.005</u>	<u>0.005</u>	<u>0.005</u>	<u>0.010</u>	<u>0.005</u>	<u>0.005</u>
	0.005	0.005	0.005	0.005	0.005	0.005	0.010	0.015	0.005	0.010

* Тип оценки эффективности моделей: ООВ — методом out-of-bag при обучении по полной выборке, тест — с предварительным разбиением выборки на обучающую и контрольную.

** Наилучший по величине RMSE тип прогнозирования модели: Ср. — простое среднее, ЛМ — среднее, скорректированное линейной моделью, Мед. — медиана.

затели ориентации и кривизны склонов, а также топографический индекс влажности.

Таким образом, оптимальный набор данных для моделирования показателей содержания и запасов N и C в лесной подстилке был составлен из четырнадцати геопространственных переменных:

— десять переменных на основе спутниковых изображений — по две главные компоненты для периодов снежной зимы, начала и середины весны, середины лета и середины осени;

— две характеристики на основе ЦМР — абсолютная высота рельефа и крутизна склонов;

— две координаты пространственного положения пикселей.

Эффективность моделей. В табл. 3 приведены показатели эффективности полученных регрессионных моделей для содержания и запасов N и C в лесной подстилке. Оценки для моделей с обучением по полной выборке (на основе ООВ-метода) более согласованы в аспекте использованных метрик (высокие значения R^2 сопровождаются низкими относительными значениями RMSE и наоборот), чем для моделей с предварительным разбиением выборки на обучающую и контрольную, однако абсолютные величины RMSE, в целом, близки для обоих подходов. Исключением являются модели для величин C : N, для которых значения RMSE по контрольной выборке в 1.4 (подгоризонт L) и в 2.6 (подгоризонт FH) раза ниже аналогичных ООВ-показателей, что может быть

случайной особенностью измерений, отбравшихся для контроля вручную.

Наилучшие результаты продемонстрировали модели для величин отношения C : N — $R^2 = 0.71$ при $RMSE_{RNG} = 12.5\%$ для подгоризонта L, и $R^2 = 0.83$ и $RMSE_{RNG} = 10.6\%$ для подгоризонта FH, — что было ожидаемо с учетом известной тесноты взаимосвязи данных показателей с соотношением хвойных и лиственных пород в древостоях. Для абсолютных показателей содержания N_% и C_% в обоих подгоризонтах R^2 находится в пределах от 0.44 до 0.56 при $RMSE_{RNG}$ от 15.8 до 17.7%, что можно оценить как умеренную степень эффективности моделей. Наихудшие результаты получены для показателей запасов N_{stk} и C_{stk} в обоих подгоризонтах — R^2 варьирует от 0.23 до 0.61 при $RMSE_{AVG}$ от 39.8 до 48.6%, что, в первую очередь, может быть следствием неоднородности исходной выборки. Тем не менее, метод случайных лесов, как правило, дает надежные результаты даже при малом объеме и неоднородности обучающих данных, в том числе и при большом количестве независимых переменных, поэтому все полученные модели статистически значимы — большинство на уровне $p < 0.005$, в худшем случае — на уровне $p < 0.015$ (для N_{stk} в подгоризонте FH). При этом невозможно однозначно отдать предпочтение ни одному из трех рассмотренных типов прогнозирования: для десяти моделей из двадцати лучшим было значение медианы, для семи — про-

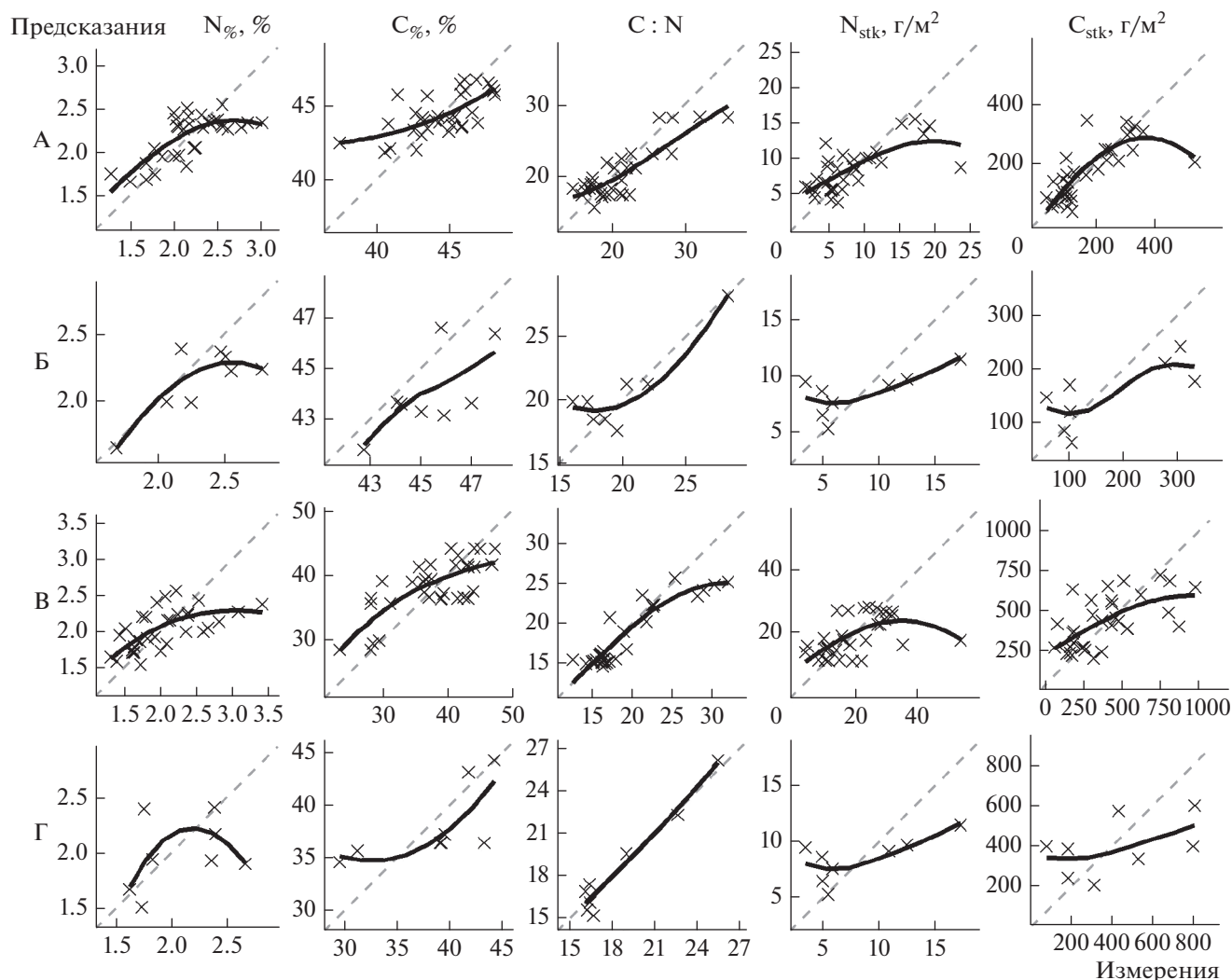


Рис. 2. Диаграммы рассеяния измеренных и предсказанных значений показателей содержания и запасов N и C в лесной подстилке для подгоризонтов L (А, Б) и FH (В, Г), оцененных по полной выборке ООВ-методом (А, В) и по отдельной контрольной выборке (Б, Г).

стое среднее, для трех оставшихся — скорректированное среднее.

На рис. 2 представлены диаграммы рассеяния измеренных и предсказанных значений (по наилучшему типу прогнозирования) для всех проанализированных показателей. Как видно из диаграмм, характерные тенденции к завышению низких и занижению высоких значений в разной степени выражены для всех моделей, причем, исходя из количественных оценок эффективности, корректировка результатов предсказания дополнительной линейной моделью не вносит заметных улучшений в данном аспекте. Наиболее вероятной причиной данной ситуации может быть низкая чувствительность использованных переменных к вариабельности анализируемых свойств подстилки по достижении ими определенных пороговых значений на обоих концах распределения (то есть на-

блюдается эффект насыщения). В целом наибольшая неопределенность предсказаний большинства моделей (с тенденцией к занижению) характерна для высоких значений оцениваемых показателей, и только для содержания C наблюдается обратная ситуация, когда для низких значений завышение выражено более явно.

Информативность переменных. Итоговая информативность отобранных переменных для каждого из моделируемых показателей приведена на рис. 3. Для величин содержания N и C, особенно для отношения C : N, можно констатировать значительное превосходство признаков, полученных на основе спутниковых изображений, над признаками, полученными на основе ЦМР, что косвенно отражает высокую взаимосвязь данных показателей с видовой структурой древостоев. Для показателей запасов N и C (кроме запаса C в

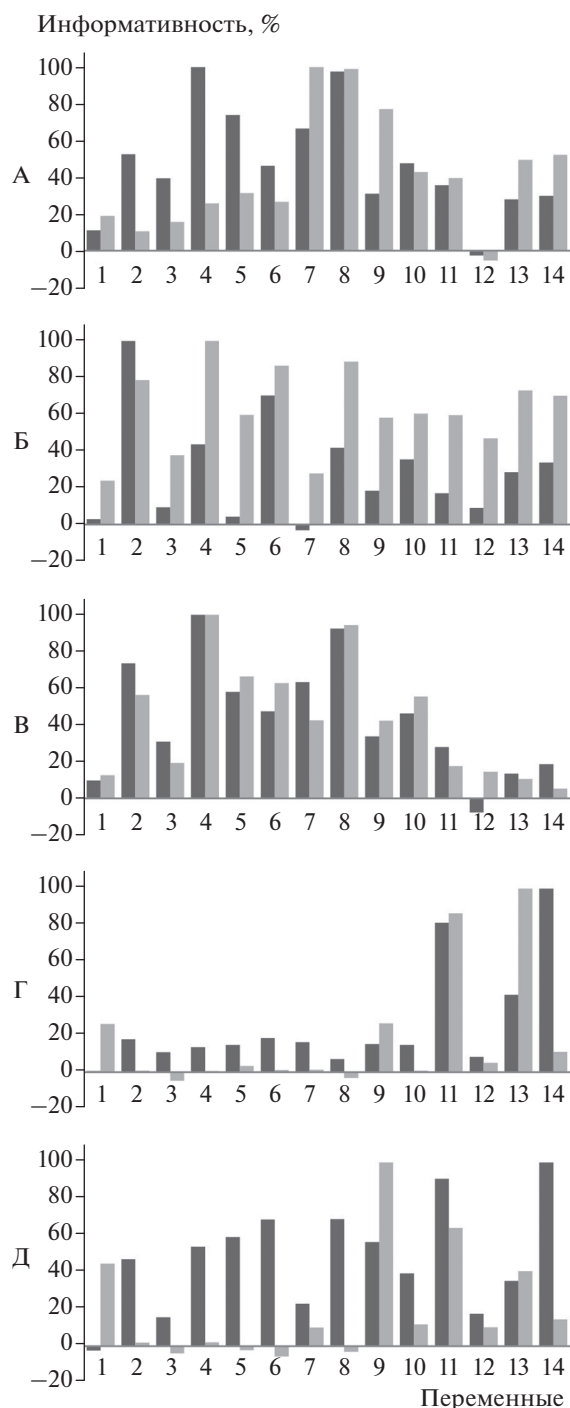


Рис. 3. Относительная информативность переменных при регрессионном моделировании показателей: $N_{\%}$ (А), $C_{\%}$ (Б), $C:N$ (В), N_{stk} (Г), C_{stk} (Д). Исходные значения отнормированы по абсолютному максимуму. Темно-серым даны значения для подгоризонта L, светло-серым — для подгоризонта FH. Обозначения переменных: 1 — зима I (первая главная компонента зимнего изображения), 2 — зима II, 3 — начало весны I, 4 — начало весны II, 5 — середина весны I, 6 — середина весны II, 7 — лето I, 8 — лето II, 9 — середина осени I, 10 — середина осени II, 11 — абсолютная высота рельефа, 12 — крутизна склона, 13 — координата пикселя по оси X, 14 — координата пикселя по оси Y.

подгоризонте FH) высота рельефа и координаты положения пикселей превосходят все спутниковые переменные по информативности. Это можно трактовать как следствие более низкой взаимосвязи данных показателей с видовым составом древостоев, с одной стороны, и более высокой зависимости от их положения в ландшафте, с другой. Кроме того, запасы N и C в подстилке связаны не только с качеством опада, но и с его массой, зависящей в первую очередь от фитомассы и продуктивности древостоев, определяемой условиями местообитаний, которые не всегда явно отражаются на спектральных свойствах лесного полога.

Наиболее значимыми периодами года при моделировании, в целом, оказались ранняя весна, лето и снежная зима. Однако степень информативности отдельных переменных сильно варьирует от показателя к показателю. Интересно, что изображения осеннего периода, несмотря на высокий потенциал распознавания видовой структуры вследствие изменения окраски листвы, оказались малоинформативными для моделирования характеристик подстилки (не считая запаса C в подгоризонте FH). Также необходимо отметить, что в большинстве случаев, вторые главные компоненты периодов, которые отражают межканальную вариабельность значений спектральной яркости, закономерно более информативны, чем первые, которые показывают различия в суммарной интенсивности отраженного излучения. Тем не менее, полностью исключить первые компоненты из моделей без значительной потери их эффективности не представляется возможным.

Результаты геопространственного моделирования. Исходя из анализа эффективности моделей и информативности переменных, наиболее надежные результаты геопространственного моделирования на основе спутниковых данных могут быть получены для отношения $C:N$ (рис. 4). Визуальное сопоставление с картой пород (рис. 1) показывает четкую пространственную согласованность предсказанных величин $C:N$ с видовым составом древостоев, что обусловливается высокой информативностью спутниковых переменных в процессе обучения. Самые низкие значения $C:N$ соответствуют широколиственным лесам, чуть более высокие — мелколиственным, самые высокие — хвойным, смешанные хвойно-лиственные древостои характеризуются промежуточными значениями. Для результатов моделирования остальных показателей (не приводятся) также характерна пространственная дифференциация по видовому составу древостоев. Однако вариабельность значений в пределах однородных массивов леса выражена в меньшей степени, чем для отношения $C:N$, — сказывается более низкий уровень взаимосвязей с использованными переменными (что отражено в более умеренных значениях коэффициента детерминации моделей). Кроме то-

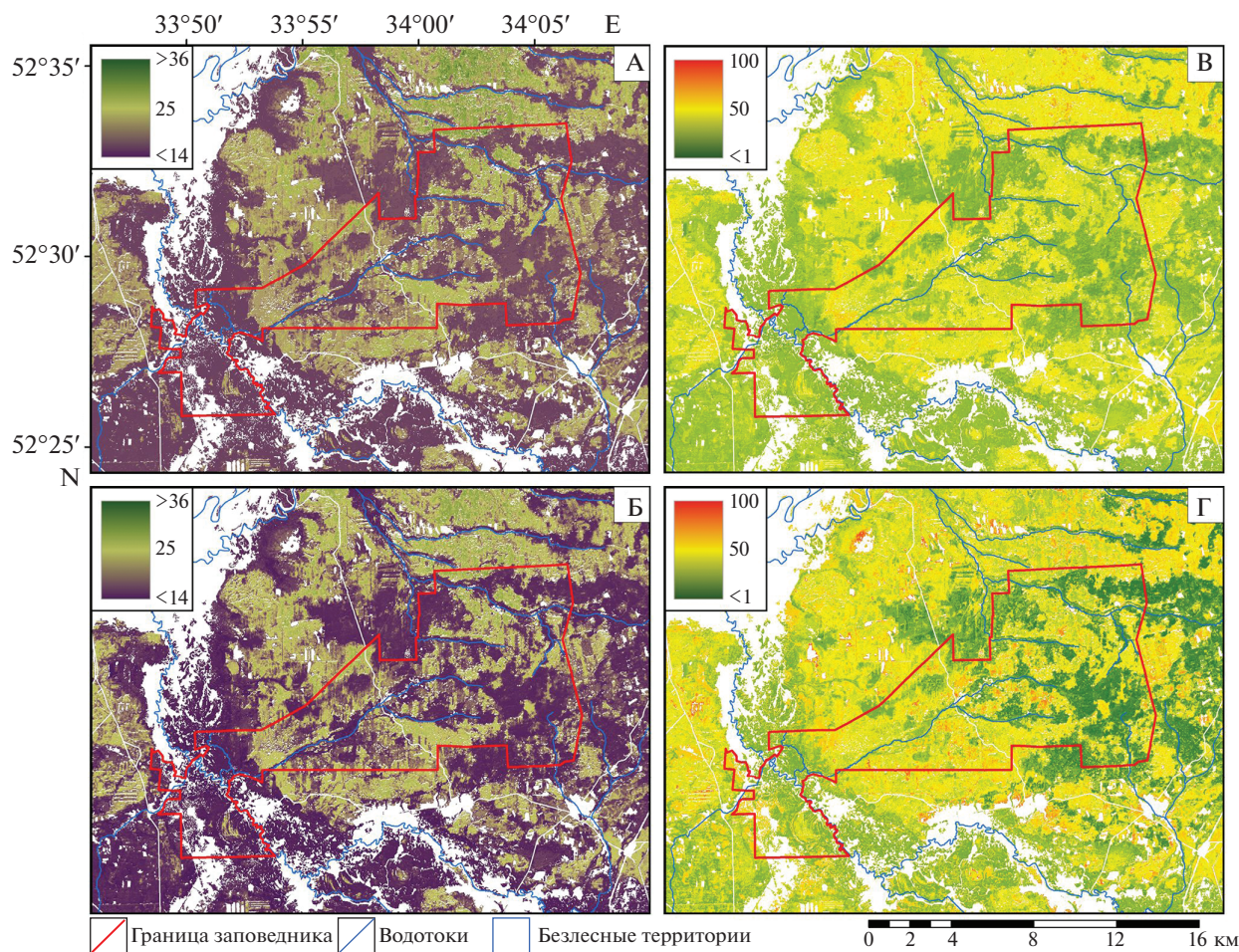


Рис. 4. Результаты геопространственного моделирования показателей С : N лесной подстилки в подгоризонтах L (A) и FH (Б) для района исследований с попиксельными оценками неопределенности (В, Г, выражены в %).

го, для запасов наблюдаются локальные эффекты “блочности” (выраженные вертикальные и горизонтальные границы между участками изображения) из-за высокой информативности переменных, отвечающих за пространственные координаты пикселей.

Пространственная согласованность результатов моделирования (рис. 5) между двумя подгоризонтами лесной подстилки явно выражена для содержания N (коэффициент корреляции Пирсона $r = 0.87$) и отношения С : N ($r = 0.97$), для остальных показателей характерна умеренная положительная корреляция (r от 0.41 до 0.65). Величины неопределенности моделирования растут пропорционально предсказанным значениям для отношения С : N в обоих подгоризонтах ($r = 0.86$ для L и $r = 0.79$ для FH), а также для содержания N в подгоризонте FH ($r = 0.78$), для остальных показателей ярко выраженной зависимости не наблюдается (r от -0.43 до 0.66). В целом наибольшие значения неопределенности характерны для участков с относительно низкой сомкнутостью полога и/или

находящихся на границе леса. Очевидно, что пиксели изображения, для которых характерна “смешанная” спектральная яркость, сформированная из сигналов от подстилающих поверхностей разного типа, сложно однозначно интерпретировать (если не закладывать целенаправленно для таких участков НПП), что соответствующим образом отражается на работе моделей.

ОБСУЖДЕНИЕ

Двумя основными факторами, не позволяющими сделать однозначный вывод о стабильности и универсальности полученных в работе результатов, являются локальность района исследования и относительно небольшой объем наземных данных, использованных для обучения регрессионных моделей. В частности, локальность накладывает ограничения на выводы об относительной информативности переменных различного типа, поскольку очевидно, что роль орографических и/или климатических данных (которые не рас-

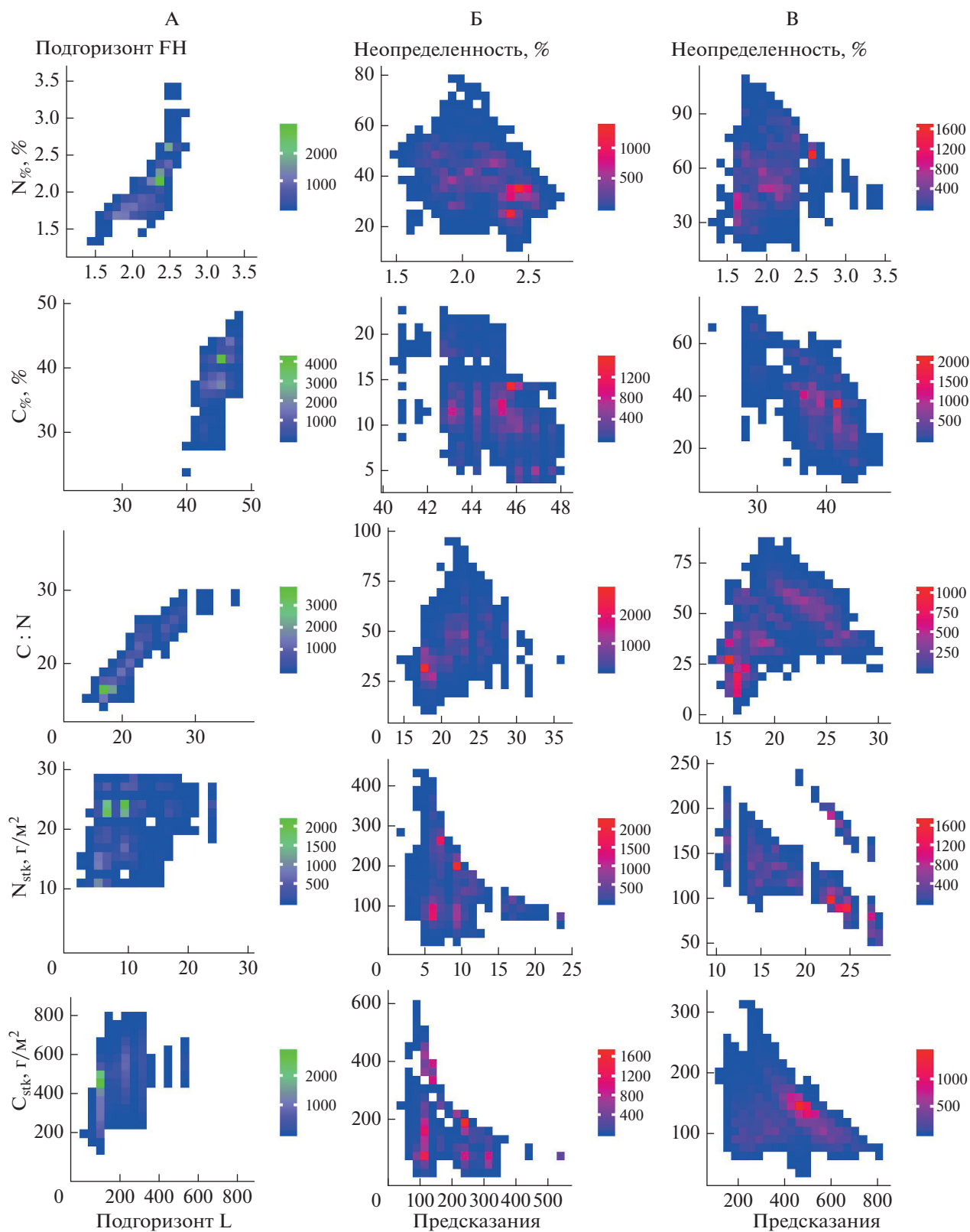


Рис. 5. Двумерные гистограммы распределения площади района исследования (га) по показателям содержания и запасов N и C в лесной подстилке: результаты моделирования по двум подгоризонтам (А) и неопределенность моделирования относительно предсказанных значений для подгоризонтов L (Б) и FH (В).

сма тривали в работе из соображений рациональности) может значительно возра стать по мере увеличения пространственного охвата и разнообразия условий произрастания лесов территории, для которой проводится моделирование. В свою очередь, малый объем наземных данных ограничивает возможности для надежной верификации результатов предсказания моделей, поэтому полученные в работе оценки эффективности носят в большей мере сравнительный, нежели точностной характер. Другими словами, определенно можно сделать вывод о том, что показатели содержания N и C, а также отношения C : N, моделируются точнее, чем показатели запасов, однако чтобы судить о том, насколько количественные оценки эффективности моделей близки к реальной точности полученных с их помощью тематических продуктов, необходимо сформировать независимый набор контрольных наземных данных, сопоставимый по объему с использованным для обучения.

Помимо объема выборки, не меньшую роль играет и пространственное расположение НПП. В данной работе при планировании наземных обследований учитывался, главным образом, видовой состав древостоев, и этого оказалось достаточно для успешного моделирования отношения C : N и удовлетворительных результатов для показателей содержания N и C. Однако показатели запасов N и C оказались в значительной степени чувствительны к положению древостоев в ландшафте, что находит подтверждение в ряде исследований. Так, запасы подстилок автоморфных и полугидроморфных почв могут различаться почти в два раза [19]. Кроме того, по результатам геопространственного моделирования низкая сомкнутость древесного полога оказалась заметным источником неопределенности для предсказаний. Таким образом, для дальнейших исследований по данному направлению при выборе мест закладки НПП необходимо учитывать как видовые и морфоструктурные характеристики древостоев, так и ландшафтные особенности территории.

Несмотря на обозначенные ограничения, полученные результаты в достаточной мере характеризуют потенциал предложенного подхода, подразумевающего использование временной серии спутниковых изображений в качестве основных переменных для геопространственного моделирования содержания и запасов N и C в лесной подстилке. Как уже отмечалось, нам не удалось найти работ, которые бы в полной мере соответствовали нашей по анализируемым показателям и использованным исходным данным. Однако несколько исследований, близких в тематических и методических аспектах, все же могут рассматриваться в качестве аналогов для сравнения. Все они используют материалы Национальной инвентаризации лесов для моделирования свойств лесной подстилки на субконтинентальном уровне. Бегуин

с соавт. использовали материалы с 500 НПП и геопространственные данные о долях хвойных и лиственных пород в составе древостоев в сочетании с абсолютной высотой рельефа для моделирования отношения C : N в органическом горизонте почв бореальных лесов Канады [21]. Сравнивались различные методы машинного обучения для построения регрессионных моделей, включая случайные леса. Показатели эффективности наилучших из рассмотренных моделей были на уровне $R^2 = 0.4$ и $RMSE = 30\%$ (авторы приводят только графические данные без точных цифр). Карре с соавт. для аналогичной задачи в лесах Европы использовали 739 НПП и около 40 переменных, включая карты преобладающих пород древостоев и форм рельефа, а также климатические данные [25]. Моделирование выполнялось с помощью простого крикинга и нейронных сетей. Результаты с использованием крикинга были наиболее точны — $R^2 = 0.6$, $RMSE = 4.91$ (относительные значения ошибок не оценивались). Као с соавт. использовали 3303 НПП и более 30 различных переменных, включающих тематические характеристики почв, климата, лесов, рельефа и материнской породы, а также разносезонные значения спектрального индекса NDVI, для моделирования запасов C в подстилке и верхнем минеральном слое почв лесов США [24]. Рассматривались три метода обучения моделей, лучшие результаты были у случайных лесов — $R^2 = 0.2$, $RMSE = 923 \text{ г/м}^2$ (даются только абсолютные значения). Как видно из приведенных примеров, результаты моделирования свойств лесной подстилки для больших территорий, даже с использованием значительных по объему обучающих выборок и широкого спектра типов геопространственных переменных, характеризуются относительно низкими показателями точности, в том числе и в сравнении с нашим локальным исследованием. Поэтому вопросы оценки возможности масштабирования предложенного подхода и определения условий, при которых могут быть получены достаточно надежные результаты моделирования, являются наиболее приоритетными для дальнейших исследований в данной области.

ЗАКЛЮЧЕНИЕ

Выполненная работа демонстрирует возможности использования оптических разносезонных спутниковых изображений для геопространственного моделирования содержания и запасов N и C в лесной подстилке, без необходимости предварительного распознавания видовой структуры древостоев, которая является их главным предиктором. Такой подход позволяет автоматически учитывать вариабельность долевого участия деревьев разных видов и/или групп видов в составе (а также ряда сопутствующих характеристик лесного покрова, таких как сомкнутость, санитарное со-

стояние и др.) при построении регрессионных моделей за счет различий во внутригодовой динамике спектральных свойств древесного полога. При этом очевидно, что тематические продукты качественных и/или количественных характеристик лесного покрова, полученные на основе данных ДЗЗ или из материалов лесоустройства, могут быть успешно использованы в качестве переменных для моделирования свойств подстилки. Однако, учитывая широкую доступность оптических спутниковых изображений, которые, как правило, сразу готовы к тематической обработке, для задач оценки характеристик непосредственно органогенного горизонта почв более рационально обойтись без промежуточных продуктов, которые могут иметь разную степень генерализации, актуальности и достоверности.

Для дальнейшего развития данного исследования планируется расширить базу наземных обследований на территории заповедника “Брянский лес” с учетом полученных тематических продуктов для более надежной валидации и калибровки моделей, а также проверить достоверность наших выводов и работоспособность использованных подходов для лесов таежного биома в европейской части России.

БЛАГОДАРНОСТЬ

Авторы выражают признательность директору ЦЭПЛ РАН, д. б. н., чл.-корр. РАН Н.В. Лукиной за предложенную идею данного исследования и содействие на всех этапах его проведения, а также администрации заповедника “Брянский лес” за помощь в организации полевых работ и коллективу экоаналитической лаборатории ЦКП “Хроматография” (регистрационный номер 3297) ИБ Коми НЦ УрО РАН (аттестат аккредитации № РОСС RU.0001.511257) за выполнение количественного анализа углерода и азота.

ФИНАНСИРОВАНИЕ РАБОТЫ

Исследование выполнено в рамках темы ГЗ ЦЭПЛ РАН “Методические подходы к оценке структурной организации и функционирования лесных экосистем” № АААА-А18-118052590019-7 (тематическая обработка данных), а также работ по Программе Президиума РАН № АААА-А18-11802199063-2 “Биоразнообразие природных систем. Рациональное использование биологических ресурсов России” (статистический анализ результатов). Формирование баз данных наземных обследований организовано при финансовой поддержке гранта Российского научного фонда (проект № 16-17-10284).

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Гаврилюк Е.А., Горнов А.В., Ершов Д.В. Оценка пространственного распределения видов деревьев заповедника “Брянский лес” и его охранной зоны на основе разносезонных спутниковых данных Landsat // Бюл. Брянского отделения РБО. 2018. № 3(15). С. 13–23. <https://doi.org/10.22281/2307-4353-2018-3-13-23>
2. Горнов А.В., Горнова М.В., Тихонова Е.В., Шевченко Н.Е., Кузнецова А.И., Ручинская Е.В., Тебенькова Д.Н. Оценка сукцессионного статуса хвойно-широколиственных лесов европейской части России на основе популяционного подхода // Лесоведение. 2018. № 4. С. 243–257. <https://doi.org/10.1134/S0024114818040083>
3. Евстигнеев О.И. Механизмы поддержания биологического разнообразия лесных биогеоценозов. Дис. ... докт. биол. наук. Нижний Новгород, 2010. 48 с.
4. Евстигнеев О.И. Неруссо-Деснянское полесье: история природопользования. Брянск: Десяточка, 2009. 139 с.
5. Евстигнеев О.И., Федотов Ю.П. Оценка разнообразия растительного покрова российско-украинской трансграничной экологической сети (на примере Неруссо-Деснянского полесья) // Перспективы развития экологической сети и создания трансграничных охраняемых территорий в бассейне Десны. М., 1999. С. 27–43.
6. Карта почвенно-географического районирования. М-б 1 : 1 5000 000 // Национальный атлас почв Российской Федерации / Гл. ред. С.А. Шоба. М.: Астрель, АСТ, 2011. С. 198–201.
7. Киселева Ю.А. Особенности формирования почв полесий на примере заповедника “Брянский лес” (вновь к вопросу о буроземо- и подзолотобразовании) // Роль почв в биосфере. Тр. Ин-та почвоведения МГУ-РАН. 2002. Вып. 1. С. 56–78.
8. Классификация и диагностика почв России. Смоленск: Ойкумена, 2004. 342 с.
9. Кузнецова А.И., Лукина Н.В., Тихонова Е.В., Горнов А.В., Горнова М.В., Смирнов В.Э., Гераскина А.П., Шевченко Н.Е., Тебенькова Д.Н., Чумаченко С.И. Аккумуляция углерода в песчаных и суглинистых почвах равнинных хвойно-широколиственных лесов в ходе послерубочных восстановительных сукцессий // Почвоведение. 2019. № 7. С. 803–816. <https://doi.org/10.1134/S0032180X19070086>
10. Меняйло О.В., Матвиенко А.И., Макаров М.И., Ченг Ш.-К. Роль азота в регуляции цикла углерода в лесных экосистемах // Лесоведение. 2018. № 2. С. 143–159. <https://doi.org/10.7868/S0024114818020067>
11. Миркин Б.М., Розенберг Л.Г., Наумова Л.Г. Словарь понятий и терминов современной фитоценологии. М., 1989. 224 с.
12. Раздел “Климат” на официальном сайте заповедника “Брянский лес” <http://www.bryansky-les.ru/naturalconditions/klimat/>
13. Распоряжение Министерства природных ресурсов и экологии РФ от 30 июня 2017 г. № 20-р “О методических указаниях по количественному определению

- лению объема поглощения парниковых газов” <https://www.garant.ru/products/ipo/prime/doc/71612096/>
14. Растительность европейской части СССР / Под ред. С.А. Грибова, Т.И. Исаченко, Е.М. Лавренко. Л.: Наука, 1980.
 15. Савин И.Ю., Жоголев А.В., Прудникова Е.Ю. Современные тренды и проблемы почвенной картографии // Почвоведение. 2019. № 5. С. 517–528. <https://doi.org/10.1134/S0032180X19050101>
 16. Соколов Л.А. К вопросу классификации почвообразующих и подстилающих горных пород Брянского лесного массива // Вклад ученых и специалистов в национальную экономику. Брянск, 1998. Т. 2. 125 с.
 17. Стефуришин М.В. Оценка почвенно-экологических условий водно-ледниковых ландшафтов Брянского лесного массива // Вопросы лесоведения и лесоводства. Сб. науч. тр. Брянск: БГИТА, 2000. Вып. 10. С. 48–50.
 18. Федорец Н.Г., Бахмет О.Н. Экологические особенности трансформации соединений углерода и азота в лесных почвах. Петрозаводск: Карельский научный центр РАН, 2003. 240 с.
 19. Чернова О.В., Рыжова И.М., Подвезенная М.А. Оценка запасов органического углерода лесных почв в региональном масштабе // Почвоведение. 2020. № 3. С. 340–350. <https://doi.org/10.31857/S0032180X20030028>
 20. Altmann A., Tolosi L., Sander O., Lengauer T. Permutation importance: a corrected feature importance measure // Bioinformatics. 2010. V. 26. P. 1340–1347.
 21. Beguin J., Fuglstad G.-A., Mansuy N., Pare D. Predicting soil properties in the Canadian boreal forest with limited data: Comparison of spatial and non-spatial statistical approaches // Geoderma. 2017. V. 306. P. 195–205. <https://doi.org/10.1016/j.geoderma.2017.06.016>
 22. Beven K.J., Kirkby M.J. A physically-based variable contributing area model of basin hydrology // Hydrology Sci. Bull. 1979. V. 24. № 1. P. 43–69.
 23. Breiman L. Random forests // Machine Learning. 2001. V. 45. № 1. P. 5–32.
 24. Cao B., Domke G.M., Russell M.B., Walters B.F. Spatial modeling of litter and soil carbon stocks on forest land in the conterminous United States // Sci. Total Environ. 2019. V. 654. P. 94–106. <https://doi.org/10.1016/j.scitotenv.2018.10.359>
 25. Carre F., Jeanne N., Casalegno S., Lemarchand O., Reuter H.I., Montanarella L. Mapping the CN ratio of the forest litters in Europe – lessons for global digital soil mapping // Digital Soil Mapping. / Eds. J.L. Boettinger et al. Dordrecht: Springer, 2010. P. 217–225.
 26. Conrad O., Bechtel B., Bock M., Dietrich H., Fischer E., Gerlitz L., Wehberg J., Wichmann V., Böhner J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4 // Geosci. Model Dev. 2015. V. 8. P. 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
 27. Cools N., Vesterdal L., De Vos B., Vanguelova E., Hansen K. Tree species is the major factor explaining C : N ratios in European forest soils // Forest Ecology and Management. 2014. V. 311. P. 3–16. <https://doi.org/10.1016/j.foreco.2013.06.047>
 28. ESA Sentinel-2 http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2 (дата обращения 20.03.2019).
 29. Escribano P., Schmid T., Chabrilat S., Rodríguez-Caballero E., García M. Optical remote sensing for soil mapping and monitoring // Soil mapping and process modelling for sustainable land use management / Eds. P. Pereira et al. Amsterdam: Elsevier, 2017. P. 87–125. <https://doi.org/10.1016/B978-0-12-805200-6.00004-9>
 30. Evstigneev O.I., Korotkov V.N. Pine Forest Succession on Sandy Ridges within Outwash Plain (Sandur) in Nerussa-Desna Polesie // Russ. J. Ecosystem Ecol. 2016. V. 1. № 3. <https://doi.org/10.21685/2500-0578-2016-3-2>
 31. Fassnacht F., Latifi H., Stereńczak K., Modzelewska A., Lefsky M., Waser L., Straub C., Ghosh A. Review of studies on tree species classification from remote sensed data // Remote Sensing of Environment. 2016. V. 186 P. 64–87. <https://doi.org/10.1016/j.rse.2016.08.013>
 32. Friedl M., Gray J., Sulla-Menashe D. MCD12Q2 MODIS/Terra + Aqua Land Cover Dynamics Yearly L3 Global 500m SIN Grid V006 [Data set]. NASA EOS-DIS Land Processes DAAC. 2019. <https://doi.org/10.5067/MODIS/MCD12Q2.006>
 33. Gallo B., Demattê J., Rizzo R., Safanelli J., Mendes W., Lepsch I., Sato M., Romero D., Lacerda M. Multi-Temporal Satellite Images on Topsoil Attribute Quantification and the Relationship with Soil Classes and Geology // Remote Sens. 2018. V. 10. P. 1571. <https://doi.org/10.3390/rs10101571>
 34. Hengl T., Mendes de Jesus J., Heuvelink G.B.M., Ruiperez Gonzalez M., Kilibarda M., Blagotić A., Shangquan W. et al. SoilGrids250m: Global gridded soil information based on machine learning // PLoS ONE. 2017. V. 12. № 2. P. e0169748. <https://doi.org/10.1371/journal.pone.0169748>
 35. Jiang Z., Huete A.R., Didan K., Miura T. Development of a two-band enhanced vegetation index without a blue band // Remote Sens. Environ. 2008. V. 112. P. 3833–3845. <https://doi.org/10.1016/j.rse.2008.06.006>
 36. Jolliffe I.T. Principal Component Analysis, second edition. N.Y.: Springer-Verlag, 2002. 488 p. <https://doi.org/https://doi.org/10.1007/b98835>
 37. Komsta L. mbml: Median-Based Linear Models. R package version 0.12.1. <https://CRAN.R-project.org/package=mbml>
 38. Kuhn M. Classification and Regression Training. R package version 6.0–84. <https://CRAN.R-project.org/package=caret>
 39. Li Q., Ma L., Liu S., Wufu A., Li Y., Yang S., Yang X. Plant litter estimation and its correlation with sediment concentration in the Loess Plateau // PeerJ Preprints. 2019. P. e27891v1 <https://doi.org/10.7287/peerj.preprints.27891v1>
 40. Louis J., Debaecker V., Pflug B., Main-Knorn M., Bieniař J., Mueller-Wilm U., Cadau E., Gascon F. SENTINEL-2 SEN2COR: L2A Processor for Users // Proceedings “ESA Living Planet Symposium 2016”. SP-740. Prague, Czech Republic. Spacebooks Online, 2016. P. 1–8.

41. *Matasci G., Hermosilla T., Wulder M.A., White J.C., Coops N.C., Hobart G.W., Zald H.S.* Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots // *Remote Sens. Environ.* 2018. V. 209. P. 90–106.
<https://doi.org/10.1016/j.rse.2017.12.020>
42. *Meinshausen N.* Quantile regression forests // *J. Machine Learning Res.* 2006. V. 7. P. 983–999.
43. *Quan Q., Wang C., He N. Zhang Zh., Wen X., Su H., Wang Q., Xue J.* Forest type affects the coupled relationships of soil C and N mineralization in the temperate forests of northern China // *Scientific Reports.* 2014. V. 4. P. 6584.
<https://doi.org/10.1038/srep06584>
44. *Ramcharan A., Hengl T., Nauman T., Brungard C., Waltman S., Wills S., Thompson J.* Soil property and class maps of the conterminous US at 100 meter spatial resolution based on a compilation of national soil point observations and machine learning // *Soil Sci. Soc. Am. J.* 2018. V. 82. P. 186–201.
<https://doi.org/10.2136/sssaj2017.04.0122>
45. *Sabetta L., Zaccarelli N., Mancinelli G., Mandrone S., Salvatori R., Costantini M.L., Zurlini G., Rossi L.* Mapping litter decomposition by remote-detected indicators // *Annals of Geophysics.* 2006. V. 49. № 1. P. 219–226.
46. *Shary P.A.* Land surface in gravity points classification by complete system of curvatures // *Mathematical Geology.* 1995. V. 27. № 3. P. 373–390.
47. *Siegel A.F.* Robust Regression Using Repeated Medians // *Biometrika.* 1982. V. 69. № 1. P. 242–244.
48. *Wang S., Adhikari K., Wang Q., Jin X., Li H.* Role of environmental variables in the spatial distribution of soil carbon (C), nitrogen (N), and C : N ratio from the north-eastern coastal agroecosystems in China // *Ecol. Indic.* 2018. V. 84. P. 263–272.
<https://doi.org/10.1016/j.ecolind.2017.08.046>
49. World Reference Base for Soil Resources 2014. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports / IUSS Working Group. Rome: FAO, 2015. 203 p.
50. *Wright M.N., Ziegler A.* A Fast Implementation of Random Forests for High Dimensional Data in C++ and R // *J. Statistical Software.* 2017. V. 77. № 1. P. 1–17.
<https://doi.org/10.18637/jss.v077.i01>
51. *Yang X., Yu Y., Hu H., Sun L.* Moisture content estimation of forest litter based on remote sensing data // *Environ. Monitoring Assessment.* 2018. V. 190. № 7. P. 421.
<https://doi.org/10.1007/s10661-018-6792-2>
52. *Yang Y., Luo Y., Finzi A.C.* Carbon and nitrogen dynamics during forest stand development: a global synthesis // *New Phytologist.* 2011. V. 190. P. 977–989.
<https://doi.org/10.1111/j.1469-8137.2011.03645.x>

Geospatial Modeling of Nitrogen and Carbon Content and Stock in the Forest Soil Organic Horizon Based on Sentinel-2 Multi-Seasonal Satellite Imagery

E. A. Gavriluk¹ *, A. I. Kuznetsova¹, and A. V. Gornov¹

¹Center for Forest Ecology and Productivity, Russian Academy of Sciences, Moscow, 117997 Russia

*e-mail: egor@ifi.rssi.ru

The capabilities of Sentinel-2 optical multispectral satellite data for modeling nitrogen (N) and carbon (C) contents, their ratio (C : N), and stocks in the forest soil organic horizon were assessed. The study was conducted in the Bryansk Forest Nature Reserve and its buffer zone. The organic horizon samples were taken at 33 plots selected with due account for the tree species diversity of the reserve's forests. Two layers of the organic horizon—L and FH—were sampled separately. The main variables for geospatial modeling were derived from a time series of eight Sentinel-2 multi-seasonal satellite images. Basic terrain characteristics and pixel position coordinates were also added to variables' stack. We used random forest to build regression models and the corresponding standard methods to assess their performance. The best results were obtained for the C : N ratio: the coefficient of determination $R^2 = 0.71$ with a scaled root-mean-square error RMSE = 12.5% in the L layer, and $R^2 = 0.83$ with RMSE = 10.6% in the FH layer. For other models, the values of R^2 ranged from 0.23 to 0.61, and the RMSE ranged from 15.8 to 48.6% with the least reliable results for the N and C stocks. Satellite-based variables were most informative for the contents of N and C, and, notably, for the C : N ratio. The most significant periods in the time series were early spring, summer, and snowy winter. To conclude, Sentinel-2 satellite imagery can be successfully used for estimation and mapping the contents and stocks of N and C in the forest soil organic horizon as a free and relevant alternative to thematic data on the species composition and related properties of stands.

Keywords: forest soil, coniferous–broadleaved forests, C : N ratio, remote sensing data, random forest, machine learning