

ПРИМЕНЕНИЕ МОДЕЛИРОВАНИЯ ДЛЯ ОЦЕНКИ И ПРОГНОЗА ИЗМЕНЕНИЙ ЗАПАСОВ УГЛЕРОДА ПОЧВ

УДК 631.4

ФАКТОРЫ ПРОСТРАНСТВЕННОГО РАСПРЕДЕЛЕНИЯ ПОЧВЕННОГО ОРГАНИЧЕСКОГО УГЛЕРОДА В ГОРАХ ЮЖНОГО УРАЛА: ПОДХОД С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

© 2024 г. А. Р. Сулейманов^{a, b, *} (<http://orcid.org/0000-0001-7974-4931>),

Р. Р. Сулейманов^{a, b, c}, Л. Н. Белан^a, И. Г. Асылбаев^{a, d}, И. О. Туктарова^a,

Р. Д. Шагалиев^a, Е. А. Богдан^a, И. И. Файрузов^c, Р. Р. Мирсаяпов^d, А. Н. Давыдычев^b

^aУфимский государственный нефтяной технический университет, ул. Космонавтов, 1, Уфа, 450064 Россия

^bУфимский институт биологии УФИЦ РАН, пр. Октября, 69, Уфа, 450054 Россия

^cУфимский университет науки и технологий, ул. Заки Валиди, 32, Уфа, 450076 Россия

^dБашкирский государственный аграрный университет, ул. 50-летия Октября, 34, Уфа, 450001 Россия

*e-mail: filpip@yandex.ru

Поступила в редакцию 01.02.2024 г.

После доработки 21.06.2024 г.

Принята к публикации 22.06.2024 г.

Проведена оценка взаимосвязей между содержанием $C_{\text{орг}}$ и основными факторами почвообразования, среди которых определены ключевые, объясняющие пространственное распределение $C_{\text{орг}}$. Исследование проводили в южной части Уральских гор на вытянутом участке протяженностью 420 км по территории Республики Башкортостан. Преобладающие типы почв: горные серые лесные (Eutric Retisols (Loamic, Cutanic, Humic)), темно-серые лесные (Luvic Retic Greyzemic Someric Phaeozems (Loamic)) и литоземы серогумусовые (Eutric Leptosols (Loamic, Humic)). Лес представлен преимущественно березами (*Betula pendula*), соснами (*Pinus sylvestris*), елью (*Picea obovata* Ledeb.), пихтами (*Abies sibirica* Ledeb.). Использовали набор данных из 306 почвенных образцов, отобранных из (0–20 см) горизонтов, и провели машинное обучение методом “случайного леса”. В качестве объясняющих переменных применяли 94 пространственные переменные окружающей среды, включая данные дистанционного зондирования, климат (температуру, осадки, облачность и др.), цифровую модель рельефа и ее производные, типы землепользования, биоклиматические зоны и др. Результаты показали, что содержание $C_{\text{орг}}$ варьировало в широком диапазоне от 0.8 до 32%. Прогнозная модель случайного леса объяснила 55% вариации $C_{\text{орг}}$, со среднеквадратичной ошибкой – 1.35%. Ключевыми переменными были температура поверхности, абсолютная высота над уровнем моря, количество осадков и облачность, в совокупности отображающие законы широтной и вертикальной зональности В.В. Докучаева. Полученные результаты подчеркивают важность учета множества факторов окружающей среды при исследованиях пространственного распределения $C_{\text{орг}}$.

Ключевые слова: случайный лес, климат, рельеф, цифровое моделирование, данные дистанционного зондирования

DOI: 10.31857/S0032180X24110142, **EDN:** JNONJT

ВВЕДЕНИЕ

Почвенный органический углерод ($C_{\text{орг}}$) является основным элементом глобального цикла углерода в окружающей среде и играет важную роль в климатических изменениях [17]. Особенно важным аспектом изучения $C_{\text{орг}}$ является его роль в лесных

экосистемах. Зачастую максимальные значения $C_{\text{орг}}$ в нетронутых лесных почвах могут служить эталоном для оценки состояния и устойчивости экосистем [25]. Установление факторов, влияющих на пространственное распределение $C_{\text{орг}}$, представляет собой важный этап в почвенных исследованиях и управлении природными ресурсами. Например,

анализ таких факторов представляет ученым данные для прогнозирования изменений почв в связи с климатическим и антропогенным воздействиями. Более того, в условиях ограниченного времени и ресурсов, установление ведущих факторов может быть более эффективным подходом по сравнению с обширными почвенными обследованиями, так как первое дает информацию для разработки эффективных стратегий сохранения лесных ресурсов и управления углеродным балансом.

Горные регионы представляют собой особенно сложную и уникальную среду, где факторы, влияющие на динамику $C_{орг}$, могут значительно различаться от присутствующих в равнинных зонах. Уральский хребет представляет собой горную систему, простирающуюся от побережья Северного Ледовитого океана до границы северо-западного Казахстана, и характеризующуюся разнообразным почвенным покровом. Почвы Уральских гор изучали в северных широтах [4–6, 15], в средней [8] и южной частях [2, 12, 13]. Ряд работ посвящен изучению органического вещества почвы [10, 23, 32].

Взаимосвязи между почвенными свойствами и влияющими на них факторами сложны и зачастую нелинейные. Алгоритмы машинного обучения являются эффективным инструментом для пространственного прогнозирования почвенных свойств и классов, а также установления ведущих причин, ответственных за их пространственное распределение [9]. Методы машинного обучения позволяют проводить нелинейные взаимосвязи между компонентами окружающей среды (факторами почвообразования) и почвенными свойствами. Так, ряд исследователей предпринимали попытки объяснить влияние условий окружающей среды на пространственную варьирование $C_{орг}$ методами машинного обучения. Например, в работе [38] использовали методы машинного обучения для моделирования содержания $C_{орг}$ в Центральной Европе. Авторы показали, что высота над уровнем моря и растительность были ведущими предикторами, ответственными за пространственное распределение $C_{орг}$. В исследовании в северной части Уральских гор было показано, что растительность, климат и геология являлись наиболее важными переменными для прогнозирования содержания $C_{орг}$ [31].

Проведение почвенных изысканий в горах часто сталкивается с трудностями, связанными с недоступностью территории. Сложный рельеф, высокогорные условия и удаленность от инфраструктуры могут затруднять доступ к местам исследований [31]. Таким образом, установление ведущих факторов окружающей среды в горных почвах может значительно облегчить проведение исследований и создать основу для разработки целенаправленных стратегий управления углеродным циклом в горных ландшафтах.

Цель работы — установить факторы окружающей среды, ответственные за пространственное распределение содержания $C_{орг}$ в лесах горного Южного Урала.

ОБЪЕКТЫ И МЕТОДЫ

Исследование проводили в южной части Уральских гор на вытянутом участке, протяженностью 420 км с севера на юг, на территории Республики Башкортостан (рис. 1). Климат умеренно теплый, засушливый. Средняя температура июля +18–20°C, января — –17°C. Количество осадков 270–308 мм [14]. Лес представлен преимущественно березами (*Betula pendula*), соснами (*Pinus sylvestris*), елью (*Picea obovata* Ledeb.), пихтами (*Abies sibirica* Ledeb.) [18]. Преобладающие типы почв — горные серые лесные (Eutric Retisols (Loamic, Cutanic, Humic)), темно-серые лесные (Luvic Retic Greyzemic Someric Phaeozems (Loamic)) и литоземы серогумусовые (Eutric Leptosols (Loamic, Humic)) [3]. Западная часть гор включает осадочные (карбонатные) породы, восточная часть — магматические [11].

В работе использовали набор данных из 306 почвенных образцов, собранных из разных источников. Часть из них собрана авторами, другая часть взята из открытых источников, включая опубликованную литературу. Большинство почвенных точек расположено под лесами, южнее обследования проводили на нетронутых лугах и залежах. Использовали только почвенные прикопки или профили с географической привязкой. Во всех образцах содержание $C_{орг}$ определяли по Тюрину с окончанием по Орлову и Гриндель [1]. В моделирование включали образцы, отобранные с высокогорных органогенных почв. Стремилась представить полную картину анализа, поскольку эти образцы взяты с уникальных участков в отношении как высоты рельефа, так и остальных факторов почвообразования. Так как почвенные образцы отбирали преимущественно из генетических горизонтов, содержание $C_{орг}$ для всей базы унифицировали для поверхностного слоя (0–20 см). Для этой цели применяли функцию “spline” с использованием библиотек *aqp* и *ithir* в среде программирования R.

В качестве объясняющих переменных (ковариат) использовали 94 пространственные переменные окружающей среды, включающие основные факторы почвообразования (табл. 1). Аналогичный набор ковариат применяли для создания всемирной цифровой почвенной картографической базы SoilGrids 2.0 [29]. Ковариаты представлены данными дистанционного зондирования, климатическими переменными, цифровой моделью рельефа и ее производными, классами рельефа и литологии, типами землепользования, гидрологией, биоклиматическими зонами и др. Пространственное разрешение всех переменных составляло 250 м на пиксель.

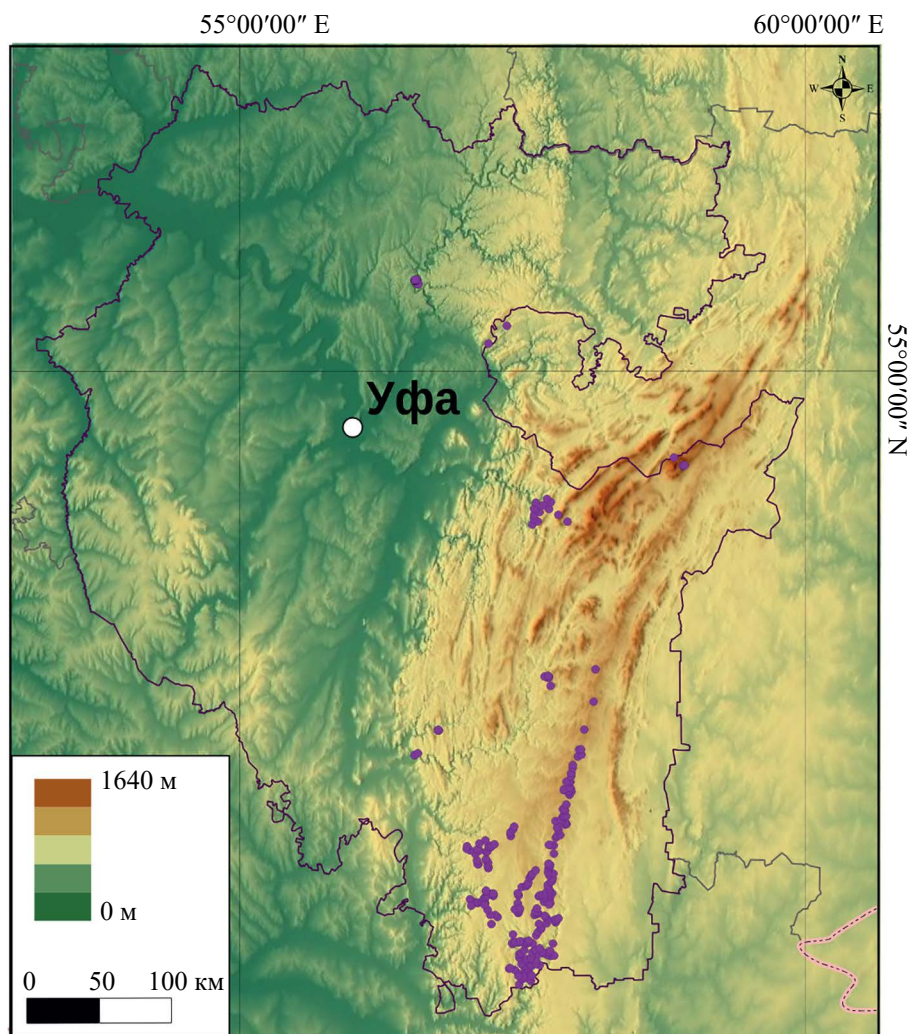


Рис. 1. Расположение исследуемого участка и почвенных образцов (фиолетовые точки).

Подготовка ковариат включала несколько этапов. На первом этапе все переменные перепроецировали в систему координат EPSG:4326. Категориальные переменные такие, как типы землепользования, представлены в виде фиктивных переменных (*dummy variable*). Такие переменные принимают значения 0 или 1 и указывают на наличие или отсутствие каждой категории (например, определенного геологического класса или типа землепользования) соответственно. После этого общее количество предикторов увеличилось до 136 шт.

Дальнейшие шаги включали отсев переменных. Алгоритмы деревьев решений позволяют использовать коррелированные переменные без снижения точности и переобучения моделей [19]. Тем не менее, это может исказить интерпретацию важности переменных, поскольку относительная важность каждой из них становится ниже [26]. Так как основной целью исследования была оценка

наиболее важных ковариат, на втором этапе исключили сильно коррелированные между собой переменные. Данный этап был реализован в пакете *removeCollinearity* в среде программирования R. Ковариаты с $r > 0.85$ исключали из анализа. Такими переменными являлись данные дистанционного зондирования и климатические ковариаты, отображающие одно и то же явление. Например, солнечная радиация или индекс EVI для соседних месяцев. Приоритет включения для дальнейшего анализа отдавали переменным, приуроченным к летним месяцам, подразумевая схожесть с периодом полевых обследований.

Далее применяли алгоритм рекурсивного исключения ковариат (*recursive elimination feature*) (**РИК**). Данный метод отбора применяется в машинном обучении для улучшения производительности модели путем итеративного удаления наименее значимых признаков [24]. РИК начинается с

Таблица 1. Список используемых пространственных переменных

Тип	Ковариаты
Климат	Количество осадков, суточный диапазон температуры, температура поверхности, солнечная радиация, давление водяного пара, вероятность снегопада, облачность, скорость ветра, биоклиматические зоны
Типы землепользования и гидрология	Классы землепользования, глубина залегания грунтовых вод
Данные дистанционного зондирования	Ближний инфракрасный диапазон со спутника MODIS и Landsat (NIR band 4), улучшенный индекс растительности (enhanced vegetation index (EVI))
Рельеф	Абсолютная высота, локальная кривизна подъема (local upslope curvature), кривизна спуска (downslope curvature), число изрезанности Мелтона (melton ruggedness number), отрицательная топографическая открытость (negative topographic openness), положительная топографическая открытость (positive topographic openness), уклон местности (terrain slope), индекс топографического положения (topographic position index), индекс влажности SAGA (SAGA wetness index), мультимасштабный индекс плоскостности дна долин (multiresolution index of valley bottom flatness – MRVBF), глубина долины (valley depth)
Геология и почвы	Классы рельефа и литологии, слои осадочных отложений, средняя мощность почвенных горизонтов и осадочных отложений

обучения модели на всем наборе ковариат и оценки их важности. Затем переменные с наименьшей важностью удаляются, и модель повторно обучается на уменьшенном наборе переменных. Этот процесс рекурсивно повторяется до тех пор, пока не будет достигнут определенный критерий, например, улучшение метрик качества модели. В качестве метрики использовали корень среднеквадратичной ошибки (RMSE).

В качестве алгоритма машинного обучения применяли случайный лес (Random Forest), наиболее популярный алгоритм в цифровой почвенной картографии [34]. Он представляет собой метод машинного обучения, который объединяет несколько деревьев решений для улучшения обобщающей способности модели и снижения переобучения [20]. В процессе построения модели случайного леса создается множество деревьев решений. Каждое дерево строится на основе бутстрэп-выборки, случайного подмножества обучающих данных. При построении узлов деревьев также происходит случайный выбор подмножества признаков, что способствует уменьшению корреляции между деревьями и делает модель более устойчивой. Финальное предсказание получается путем усреднения предсказаний всех деревьев. Этот подход уменьшает переобучение, повышает обобщающую способность модели и позволяет оценить важность объясняющих переменных. Поскольку случайный лес относительно нечувствителен к параметризации [30], использовали параметры модели по умолчанию с использованием библиотеки *randomForest* в R.

Для оценки важности пространственных переменных применяли метрику %IncMSE (permu-

tation), указывающую на увеличение среднеквадратической ошибки при случайной перестановке данной переменной. Для более достоверной оценки важности ковариат рассчитывали среднее значение %IncMSE после 100 повторений модели случайного леса.

Эффективность модели оценивали с использованием метода перекрестной проверки с 10 частями (fold) в ходе 10 повторений. Этот подход к оценке производительности модели включает разделение набора данных на 10 частей и многократное обучение модели на 9 из них, с последующей проверкой на оставшейся части. Процесс повторяли 10 раз, обеспечивая более надежные оценки производительности модели на различных подмножествах данных.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Содержание $C_{орг}$. Содержание $C_{орг}$ варьировало в широком диапазоне от 0.8 до 32% со средним значением 3.2%. Содержание $C_{орг}$ в образцах характеризовалось высоким коэффициентом вариации (59.8%), когда как стандартное отклонение составляло 2.6%.

Корреляция $C_{орг}$ и переменных окружающей среды. В табл. 2 представлены результаты корреляции Пирсона между содержанием $C_{орг}$ пространственными переменными. Установлено, что 11 переменных показали статистически значимую взаимосвязь ($p < 0.001$) с содержанием $C_{орг}$ в слое 0–20 см. Среди них высота над уровнем моря ($r = 0.61$) и скорость ветра ($r = -0.48$) показали наибольшую корреляцию. Среди других переменных,

Таблица 2. Корреляция $C_{орг}$ с переменными окружающей среды (показаны только переменные с $r > |0.4|$)

Параметр	r	p
Дневная температура поверхности в мае	0.44	(<0.001)
Ночная температура поверхности в апреле	0.44	(<0.001)
Ночная температура поверхности в августе	0.42	(<0.001)
Годовой диапазон температур	-0.41	(<0.001)
Средняя температура самого влажного квартала	-0.46	(<0.001)
Общее количество осадков в апреле	0.44	(<0.001)
Общее количество осадков в мае	0.45	(<0.001)
Годовая солнечная радиация	-0.39	(<0.001)
Скорость ветра	-0.48	(<0.001)
Высота	0.61	(<0.001)
Индекс EVI июль–август	0.40	(<0.001)

солнечная радиация, индекс растительности, количество осадков, температура поверхности и воздуха имели взаимосвязь с содержанием $C_{орг}$.

Отсев переменных окружающей среды. На рис. 2 представлен результат отсева переменных с использованием алгоритма РИК. Видно, что при использовании 61 переменной достигаются лучшие результаты модели, согласно RMSE значению. Кривая резко падает до около 40–41 переменных, а затем остается на плато.

Точность модели. Согласно перекрестной проверки с 10-кратным повторением, модель объяснила 55% дисперсии $C_{орг}$ ($R^2 = 0.55$) с ошибкой $RMSE = 1.35\%$. На рис. 3 изображена диаграмма рассеяния наблюдаемых и предсказанных значений.

Важность пространственных переменных. На рис. 4 представлен анализ важности ковариат после 100 повторений. Поскольку количество предикторов в модели машинного обучения составило 61, показаны только 15 наиболее важных переменных. Согласно результатам, наиболее важной переменной являлась дневная температура поверхности в ноябре. На втором месте расположилась высота над уровнем моря, далее сезонность осадков и облачность. Также три переменные растительности (MODIS NIR для мая и октября, MODIS EVI для июля–августа) были включены в топ 15 переменных. Таким образом, преимущественно климатические факторы и высота над уровнем моря оказались наиболее влиятельными переменными для

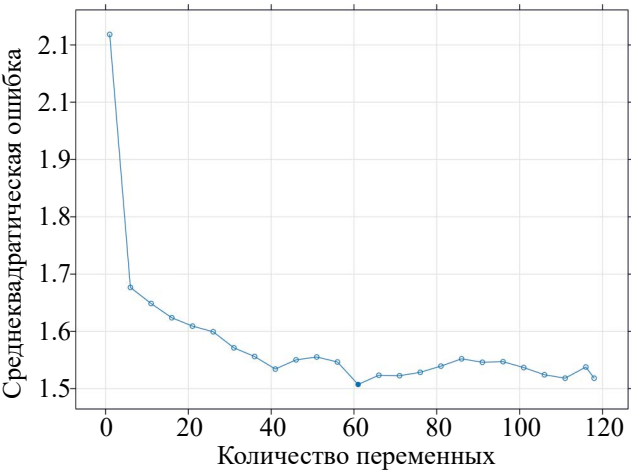


Рис. 2. График снижения значения RMSE при использовании различного количества переменных в алгоритме случайного леса.

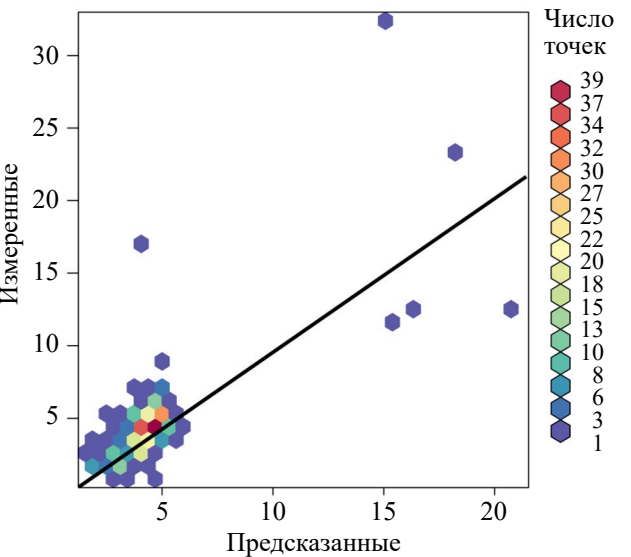


Рис. 3. Диаграмма рассеяния измеренных и предсказанных значений $C_{орг}$.

пространственного распределения содержания $C_{орг}$ в горах Южного Урала.

Следовательно, на исследуемом участке четко проявляется закон вертикальной (горной) и горизонтальной зональности почв В.В. Докучаева [11]. Высота исследуемого участка снижается с севера на юг, что ведет к изменению климатических условий от более холодного и влажного на севере к более сухому на юге. На юге Уральских гор климат становится более засушливым, что, соответственно, обуславливает изменение основных климатических показателей и почвообразовательных процессов. К югу уменьшается высота гор, что делает

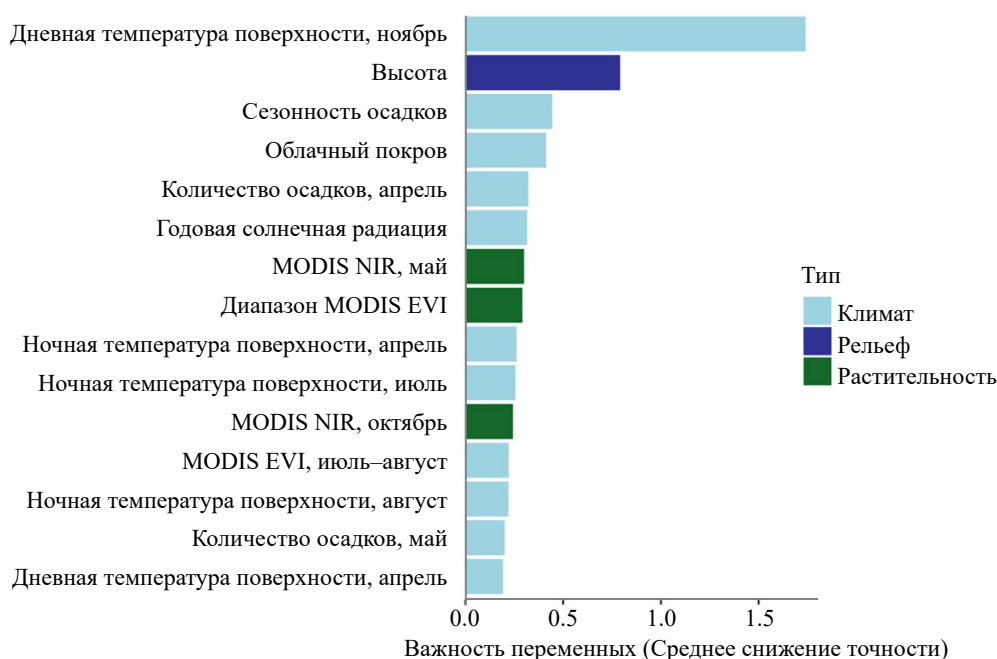


Рис. 4. Топ 15 важнейших пространственных переменных в прогнозировании содержания $C_{орг}$

данную переменную важным фактором при моделировании $C_{орг}$. Таким образом, наблюдается изменение климатических и растительных условий, и это влияет на процессы разложения органического материала [36]. Например, температура почвы и поверхности рассматриваются Всемирной метеорологической организацией в качестве основных климатических параметров [37], определяющих скорость разложения органического вещества. Несмотря на наличие в наборе данных органогенных образцов с высоким содержанием $C_{орг}$ (от 11.9 до 32%), их исключение из процесса машинного обучения не повлияло как на точность модели, так и на важность переменных. Это объясняется их малым количеством (2% от выборки), а также согласованием с общей выявленной тенденцией ключевых переменных, поскольку данные высокогорные образцы почв имеют наивысшие отметки высот и минимальную температуру.

Полученные результаты согласуются с результатами других исследований, проведенных в региональных и национальных масштабах. Например, при моделировании содержания $C_{орг}$ на территории всей России, показатель изменения температуры в течение года (сезонность температуры) был важнейшей переменной в ансамблевом алгоритме машинного обучения [22]. Аналогично, переменная количества осадков внесла наибольший вклад в точность моделирования содержания и запасов $C_{орг}$ с использованием методов машинного обучения на территории Дании [16]. Таким образом, климатические переменные вносят наибольший

вклад в объяснение пространственного распределения $C_{орг}$ и других свойств в региональном и глобальном масштабах, что было также показано в обзорной статье [21].

В настоящем исследовании использовали большой набор переменных окружающей среды для моделирования содержания $C_{орг}$. Тем не менее, для достижения более точных результатов рекомендуется тестировать и интегрировать дополнительные переменные, например карты почвенных классов и свойств, которые являются ценным источником для объяснения пространственной вариации $C_{орг}$ [21]. Биота также является ключевым фактором для качественных и количественных показателей $C_{орг}$ [7]. Установлено, что спектральные индексы, отображающие состояние вегетации, также являлись важными предикторами $C_{орг}$. Ранее авторы [27] показали, что запасы углерода в лесных почвах сильно зависят от типа леса, вида и качества опада. Установлено, что наибольшие запасы углерода обнаружены в почвах смешанных лесов, характеризующихся разнообразием видов и, следовательно, лесного опада. Ранее в горах Башкортостана с применением метода случайного леса отмечено, что спектральные индексы были наиболее важными переменными для цифрового картографирования почвенного органического вещества и pH [33]. Важно отметить, что пространственное разрешение переменных в данном исследовании составляло 250 м. Предполагаем, что использование более детальных карт, представляющих основные факторы почвообразования, повысит точность

прогнозирования содержания $C_{орг}$. Одним из лимитирующим фактором рассматриваем факт, что содержание и распределение почвенных свойств зависят от множества факторов, действие которых происходит в течение значительных периодов времени. Поэтому важно понимать, что используемые ковариаты преимущественно отражают только определенную моментальную ситуацию, которая может быть не полностью репрезентативной для всех аспектов динамики $C_{орг}$. По этой причине в прогноз также следует интегрировать факторы возраста и времени, представляющие структуру *scorpan* модели [28].

Следует отметить, что оценка важности переменных с использованием методов машинного обучения, как показано в настоящей работе, является интерпретацией самой созданной модели, когда как природные взаимосвязи между переменными окружающей среды и почвенными свойствами являются другим аспектом [35]. Таким образом, дальнейшими исследованиями в данной области могут быть применение различных методов интерпретации сложных моделей, таких значения Шепли, суррогатное моделирование и др. Например, авторы [35] продемонстрировали применения значений Шепли для изучения важности факторов, контролирующих изменение запасов $C_{орг}$ на территории Франции. Авторы показали, что помимо общей оценки вклада окружающих факторов в прогноз, метод также позволил оценить вклада ковариатов в зависимости пространственного положения.

ЗАКЛЮЧЕНИЕ

Изучение ведущих факторов окружающей среды, объясняющие пространственное распределение $C_{орг}$ в лесных экосистемах, способствуют более глубокому пониманию роли лесов в почвенном углеродном цикле. В исследовании использовали набор данных по содержанию $C_{орг}$ 306 почвенных образцов и 94 пространственные переменные для выявления взаимосвязей. С помощью линейных и нелинейных методов показано, что климатические факторы и высота над уровнем моря были наиболее сильными предикторами для прогнозирования содержания $C_{орг}$ в горных почвах Южного Урала. Полученные результаты демонстрируют законы вертикальной и горизонтальной зональности почв, сформулированные В.В. Докучаевым. Показана важность учета множества почвообразующих факторов для пространственного моделирования почвенного углерода. Исследования в горах Урала позволяют расширить знание о влиянии географических особенностей на аккумуляцию и циркуляцию $C_{орг}$, что в свою очередь может привести к разработке более точных и эффективных методов сохранения природных ресурсов в данных регионах.

БЛАГОДАРНОСТЬ

Авторы выражают благодарность Международному почвенному справочно-информационному центру (the International Soil Reference and Information Centre, ISRIC) за предоставление данных об окружающей среде. Часть почвенных данных была собрана и проанализирована в рамках государственного контракта с Министерством земельных и имущественных отношений Республики Башкортостан и программы “Выполнение работ по почвенному обследованию, оцифровке и корректировке почвенных карт, формированию экспликаций почвенных разновидностей и угодий земельных участков сельскохозяйственного назначения на территории муниципальных районов Республики Башкортостан”.

ФИНАНСИРОВАНИЕ РАБОТЫ

Исследование выполнено в рамках программы Министерства науки и высшего образования Российской Федерации “Приоритет 2030” (национальный проект “Наука и университет”).

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

В данной работе отсутствуют исследования человека или животных.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Агрохимические методы исследования почв. М.: Наука, 1976. 656 с.
2. Бурангулова М.Н., Мукастанов А.Х., Курчиев П.А. Горные почвы Башкирии // Почвы Башкирии. Уфа, 1973. Т. 1. С. 405–437.
3. Габбасова И.М., Гарипов Т.Т., Сулейманов Р.Р., Комиссаров М.А., Хабиров И.К., Сидорова Л.В., Назырова Ф.И., Простякова З.Г., Котлугалямова Э.Ю. Влияние низовых пожаров на свойства и эрозию лесных почв Южного Урала (Башкирский государственный природный заповедник) // Почвоведение. 2019. № 4. С. 412–421.
4. Дымов А.А., Жангуров Е.В. Морфолого-генетические особенности почв кряжа Енганэпэ (Полярный Урал) // Почвоведение. 2011. № 5. С. 515–524.
5. Жангуров Е.В., Королёв М.А., Дубровский Ю.А., Шамрикова Е.В. Почвы горного хребта Рай-Из (Полярный Урал) // Почвоведение. 2023. № 4. С. 417–432.
6. Жангуров Е.В., Старцев В.В., Дубровский Ю.А., Дегтева С.В., Дымов А.А. Морфолого-генетические

- особенности почв горных лиственных лесов и редколесий Приполярного Урала // Почвоведение. 2019. № 12. С. 1415–1429.
7. Лукина Н.В., Кузнецова А.И., Гераскина А.П., Смирнов В.Э., Иванова В.Н., Тебенкова Д.Н., Горнов А.В., Шевченко Н.Е., Тихонова Е.В. Неучтенные факторы, определяющие запасы углерода в лесных почвах // Метеорология и гидрология. 2022. № 10. С. 92–110.
 8. Самофалова И.А., Лузянина О.А. Горные почвы Среднего Урала (на примере ГПЗ “Басеги”). Пермь: Прокрость, 2014. 154 с.
 9. Сморгалов И.А. Изменчивость дыхания почвы: оценка вклада пространства и времени с помощью алгоритма Random Forest // Экология. 2022. № 4. С. 299–311.
 10. Старцев В.В., Мазур А.С., Дымов А.А. Содержание исостав органического вещества почв Приполярного Урала // Почвоведение. 2020. № 12. С. 1478–1488.
 11. Хазиев Ф.Х., Мукастанов А.Х., Хабиров И.К., Кольцова Г.А., Габбасова И.М., Рамазанов Р.Я. Почвы Башкортостана. Эколого-генетическая и агропроизводственная характеристика. Уфа: Гилем, 1995. Т. 1. 384 с.
 12. Халитов Р.М., Абакумов Е.В., Сулейманов Р.Р., Котлугалимова Э.Ю. Горные почвы Южного Урала (на примере Национального парка “Башкирия”) // Известия Самарского НЦ РАН. 2011. Т. 13. № 5-2. С. 128–130.
 13. Халитов Р.М., Перова Е.Н., Абакумов Е.В., Сулейманов Р.Р. Минералогический состав коренной горной породы торфянисто-подзолистой почвы горного массива Иремель, Южный Урал // Почвоведение. 2017. № 8. С. 992–1001.
 14. Чибилев А.А. Природа Оренбургской области. Ч. 1. Физико-географический и историко-географический очерк. Оренбург, 1995. 128 с.
 15. Шамрикова Е.В., Жангуров Е.В., Кулюгина Е.Е., Королев М.А., Кубик О.С., Туманова Е.А. Почвы и почвенный покров горно-тундровых ландшафтов Полярного Урала на карбонатных породах: разнообразие, классификация, распределение углерода и азота // Почвоведение. 2020. № 9. С. 1053–1070.
 16. Adhikari K., Hartemink A.E., Minasny B., Bou Kheir R.M., Greve B., Greve M.H. Digital mapping of soil organic carbon contents and stocks in Denmark // PloS One. 2014. V. 9(8). P. e105519. <https://doi.org/10.1371/journal.pone.0105519>
 17. Batjes N.H. Total carbon and nitrogen in the soils of the world // Eur. J. Soil Sci. 1996. V. 47(2). P. 151–163. <https://doi.org/10.1111/j.1365-2389.1996.tb01386.x>
 18. Belan L., Suleymanov A., Bogdan E., Volkov A., Gay-sin I., Tuktarova I., Shagaliev R. Assessing and Mapping Changes in Forest Growing Stock Volume over Time in Bashkiriya Nature Reserve, Russia // Forests. 2023. V. 13(11). P. 2251. <https://doi.org/10.3390/f14112251>
 19. Biau G., Scornet E. A random forest guided tour // TEST. 2016. V. 25(2). P. 197–227. <https://doi.org/10.1007/s11749-016-0481-7>
 20. Breiman L. Random Forests // Machine Learning. 2001. V. 45(1). P. 5–32.
 21. Chen S., Arrouays D., Leatitia Mulder V., Poggio L., Minasny B., Roudier P., Libohova Z., Lagacherie P., Shi Z., Hannam J., Meersmans J., Richer-de-Forges A.C., Walter C. Digital mapping of GlobalSoilMap soil properties at a broad scale: A review // Geoderma. 2022. V. 409. P. 115567. <https://doi.org/10.1016/j.geoderma.2021.115567>
 22. Chinilin A., Savin I.Yu. Combining machine learning and environmental covariates for mapping of organic carbon in soils of Russia // The Egypt. J. Remote Sensing Space Sci. 2023. V. 26(3). P. 666–675. <https://doi.org/10.1016/j.ejrs.2023.07.007>
 23. Dymov A.A., Startsev V.V., Milanovsky E.Y., Valdes-Korovkin I.A., Farkhodov Y.R., Yudina A.V., Donnerhack O., Guggenberger G. Soils and soil organic matter transformations during the two years after a low-intensity surface fire (Subpolar Ural, Russia) // Geoderma. 2021. V. 404. P. 115278. <https://doi.org/10.1016/j.geoderma.2021.115278>
 24. Guyon I., Weston J., Barnhill S., Vapnik V. Gene Selection for Cancer Classification using Support Vector Machines // Machine Learning. 2022. V. 46(1). P. 389–422. <https://doi.org/10.1023/A:1012487302797>
 25. Johnson D.W., Curtis P.S. Effects of forest management on soil C and N storage: meta analysis // Forest Ecology and Management. 2001. V. 140(2-3). P. 227–238.
 26. Kuhn M., Johnson K. Applied Predictive Modeling. N.Y.: Springer, 2013.
 27. Kuznetsova A.I., Geraskina A.P., Lukina N.V., Smirnov V.E., Tikhonova E.V., Shevchenko N.E., Gornov A.V., Ruchinskaya E.V., Tebenkova D.N. Linking Vegetation, Soil Carbon Stocks, and Earthworms in Upland Coniferous–Broadleaf Forests // Forests. 2021. V. 12(9). P. 1179. <https://doi.org/10.3390/f12091179>
 28. McBratney A.B., Mendonça Santos M.L., Minasny B. On digital soil mapping // Geoderma. 2003. V. 117(1–2). P. 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
 29. Poggio L., de Sousa L.M., Batjes N.H., Heuvelink G.B.M., Kempen B., Ribeiro E., Rossiter D. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty // Soil. 2021. V. 7(1). P. 217–240. <https://doi.org/10.5194/soil-7-217-2021>
 30. Rodriguez-Galiano V.F., Ghimire B., Rogan J., Chica-Olmo M., Rigol-Sanchez J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification // ISPRS J. Photogrammetry Remote Sensing. 2012. V. 67. P. 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>
 31. Suleymanov A., Abakumov E., Alekseev I., Nizamutdinov T. Digital mapping of soil properties in the high

- latitudes of Russia using sparse data // *Geoderma Regional*. 2024. V. 36. P. e00776.
<https://doi.org/10.1016/j.geodrs.2024.e00776>
32. Suleymanov A., Gabbasova I., Suleymanov R., Abakumov E., Polyakov V., Liebelt P. Mapping soil organic carbon under erosion processes using remote sensing // *Hungarian Geographical Bulletin*. 2021. V. 70(1). P. 49–64.
<https://doi.org/10.15201/hungeobull.70.1.4>
 33. Suleymanov A., Tuktarova I., Belan L., Suleymanov R., Gabbasova I., Araslanova L. Spatial prediction of soil properties using random forest, k-nearest neighbors and cubist approaches in the foothills of the Ural Mountains, Russia // *Modeling Earth Systems and Environment*. 2023. V. 9(3). P. 3461–3471.
<https://doi.org/10.1007/s40808-023-01723-4>
 34. Wadoux A.M.J.-C., Minasny B., McBratney A.B. Machine learning for digital soil mapping: Applications, challenges and suggested solutions // *Earth-Sci. Rev.* 2020. V. 210. P. 103359.
<https://doi.org/10.1016/j.earscirev.2020.103359>
 35. Wadoux A.M.J.-C., Saby N.P.A., Martin M.P. Shapley values reveal the drivers of soil organic carbon stock prediction // *Soil*. 2023. V. 9. P. 21–38.
<https://doi.org/10.5194/soil-9-21-2023>
 36. Wang Q., Zhao X., Chen L., Yang Q., Chen S., Zhang W. Global synthesis of temperature sensitivity of soil organic carbon decomposition: latitudinal patterns and mechanisms // *Functional Ecology*. 2019. V. 33. P. 514–523.
<https://doi.org/10.1111/1365-2435.13256>
 37. WMO, World Meteorological Organization (WMO). State of the Global Climate 2022 (WMO-No. 1316). 2023. WMO, Geneva.
 38. Zhou T., Geng Y., Chen J., Pan J., Haase D., Lausch A. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms // *Sci. Total Environ.* 2020. V. 729. P. 138244.
<https://doi.org/10.1016/j.scitotenv.2020.138244>

Drivers of Soil Organic Carbon Spatial Distribution in the Southern Ural Mountains: A Machine Learning Approach

A. R. Suleymanov^{1, 2, *}, R. R. Suleymanov^{1, 2, 3}, L. N. Belan¹, I. G. Asylbaev^{1, 4},
 I. O. Tuktarova¹, R. D. Shagaliev¹, E. A. Bogdan¹, I. I. Fairuzov³,
 R. R. Mirsayapov⁴, and A. N. Davydychev²

¹Ufa State Petroleum Technological University, Ufa, 450064 Russia

²Ufa Institute of Biology UFRS RAS, Ufa, 450054 Russia

³Ufa University of Science and Technology, Ufa, 450076 Russia

⁴Bashkir State Agrarian University, Ufa, 450001 Russia

*e-mail: filpip@yandex.ru

This study aims to assess the relationships between SOC content and main soil-forming factors and identify key factors explaining the spatial distribution of SOC. The research was conducted in the Southern Ural Mountains throughout 420 km from north to south in the Republic of Bashkortostan. The predominant soil types are mountainous gray forest (Eutric Retisols (Loamic, Cutanic, Humic)), dark gray forest (Luvic Retic Greyzemic Someric Phaeozems (Loamic)) soils, and gray-humus lithozems (Eutric Leptosols (Loamic, Humic)). Forest stands are mainly composed of birch (*Betula pendula*), pine (*Pinus sylvestris*), spruce (*Picea obovata* Ledeb.), and fir (*Abies sibirica* Ledeb.). A data set of 306 soil samples taken from the top layer (0–20 cm) was studied using the “random forest” machine learning method. Ninety four spatial environmental covariates were used as explanatory variables, including remote sensing data, climate (temperature, precipitation, cloudiness, etc.), digital elevation model and its derivatives, land uses, bioclimatic zones, etc. The results showed that the SOC content varied widely from 0.8 to 32%. The random forest predictive model explained 55% of SOC variation (R^2) with a root mean squared error (RMSE) of 1.35%. Key variables included surface temperature, absolute elevation, precipitation, and cloudiness, which together reflect the Dokuchaev vertical and horizontal zonality laws. The findings emphasize the importance of considering multiple environmental factors in subsequent research focused on assessing the spatial distribution of SOC.

Keywords: random forest, climate, relief, digital soil mapping, remote sensing data