

УДК 631.433.3+519.2:004.8+528.88

ОЦЕНКИ ГОДОВЫХ ПОТОКОВ УГЛЕКИСЛОГО ГАЗА ИЗ ПОЧВЫ ЕЛОВЫХ ЛЕСОВ КАРБОНОВОГО ПОЛИГОНА «УРАЛ-КАРБОН» ПО НЕПОЛНЫМ ВРЕМЕННЫМ РЯДАМ С ПРИМЕНЕНИЕМ КЛАССИЧЕСКИХ РЕГРЕССИОННЫХ ПОДХОДОВ И МАШИННОГО ОБУЧЕНИЯ

© 2025 г. И. А. Сморкалов^{a, b}

^aУральский федеральный университет им. Первого президента России Б.Н. Ельцина, Россия 620002
Екатеринбург, ул. Мира, 19

^bИнститут экологии растений и животных УрО РАН,
Россия 620144 Екатеринбург, ул. 8 Марта, 202

*e-mail: ivan.a.smorkalov@gmail.com

Поступила в редакцию 25.11.2024 г.

После доработки 26.02.2025 г.

Принята к публикации 25.03.2025 г.

Годовой поток углекислого газа из почв различных биомов играет ключевую роль в создании глобальных климатических моделей и при анализе циклов углерода в наземных экосистемах. Тем не менее существуют значительные пробелы в подобных исследованиях регионального масштаба. Из-за высокой трудоемкости получения ежедневных показателей почвенного дыхания используются различные методы моделирования. В данной работе на основе 2760 измерений дыхания почвы в еловых лесах карбонового полигона «Урал-Карбон» (Средний Урал), выполненных осенью 2021 г. и с апреля по октябрь 2022 г., с применением классических регрессионных подходов и машинного обучения были оценены годовые показатели дыхания почвы. Мы также исследовали зависимость результатов от сложности модели (числа предикторов) и используемых методов (экстраполяции моделью случайного леса и комбинированными подходами для оценки зимних потоков CO₂). «Упрощенная» модель с 7 предикторами показала лишь незначительное снижение точности по сравнению с полной моделью с 21 предиктором ($R^2 = 0.89$, $MSE = 0.31$ против $R^2 = 0.92$, $MSE = 0.22$). Предикторы, основанные на дистанционном зондировании, оказались более значимыми для точности модели, чем данные, полученные в полевых условиях. Хотя первоначальные результаты разных подходов различались, добавление значений зимнего дыхания, взятых из литературы, в модель случайного леса и усреднения значения комбинированных подходов позволило добиться схожих величин годового дыхания почвы: 830.3 ± 6.4 и 851.6 ± 8.0 г C/м² · год соответственно.

Ключевые слова: эмиссия CO₂, цикл углерода, машинное обучение, лесные экосистемы, карбоновые полигоны, экологические факторы

DOI: 10.31857/S0367059725030059 **EDN:** TCDWHO

Точные оценки потоков климатически активных газов важны для анализа и прогнозирования изменений цикла углерода [1]. Величины этих потоков могут значительно варьировать в зависимости от типа ландшафта. Однако для территорий многих регионов России до сих пор отсутствуют данные о годовых величинах эмиссии углекислого газа.

Для оценки эмиссий углекислого газа из почвы применяют два основных метода: камерный и, косвенно, метод микровихревых пульсаций. Камерный метод позволяет проводить измерения в большем масштабе, однако он не обеспечивает возможность ежедневного покрытия данными в течение всего года, особенно в случае исполь-

зования нестационарных камер. Дополнительную сложность представляет оценка зимней эмиссии углекислого газа, которая может достигать четверти годового объема [2].

Чтобы преодолеть проблему недостатка данных, прибегают к математическому моделированию. Для перехода от разрозненных временных измерений к годовому потоку применяются как традиционные регрессионные модели, так и методы машинного обучения, например алгоритм случайного леса (random forest (RF)), который демонстрирует лучшие результаты по сравнению с классической регрессией и нейронными сетями. Однако модели на базе машинного обучения зачастую включают

избыточное число предикторов, что усложняет их интерпретацию. Поэтому важно создавать простые модели с максимальной точностью.

Основной недостаток алгоритма RF заключается в плохой способности к экстраполяции за пределы обучающего диапазона, например когда отсутствуют зимние измерения. Для решения этой задачи предлагаются модели, рассчитывающие годовой поток на основании данных теплого периода года [3, 4]. Такие модели, основанные на классических регрессионных подходах, демонстрируют высокие коэффициенты детерминации (0.81–0.95), но они создавались на разных наборах данных, и мы не нашли сравнительных исследований их работы, проведенных на одних и тех же данных.

Таким образом, с одной стороны, существует актуальная задача получения оценок годовых потоков парниковых газов из почв различных биомов, а с другой – проблема отсутствия полных рядов данных в течение года и разные способы ее решения. Поэтому цель нашей работы – с применением разных подходов (в том числе машинного обучения) оценить величину годовой эмиссии CO_2 из почвы еловых лесов карбонового полигона «Урал-Карбон» (Средний Урал) в 2022 г.

МАТЕРИАЛ И МЕТОДЫ

Район исследования

Для налаживания системы мониторинга и получения адекватных оценок потоков парниковых газов в стране стартовали два федеральных проекта по мониторингу баланса углерода: «Карбоновые полигоны» и ВИП ГЗ (Важнейший инновационный проект государственного значения) «Российская система климатического мониторинга». Программа создания сети карбоновых полигонов была запущена в 2021 г. Каждый такой полигон представляет собой участок земли с характерными для региона рельефом, растительным и почвенным покровом. Эти полигоны предназначены для проведения научных исследований, развития инфраструктуры и тестирования технологий контроля за балансом климатических активных газов в природных экосистемах [5]. Карбоновый полигон «Урал-Карбон» находится в Свердловской области и состоит из двух участков, расположенных на противоположных склонах Урала (западном и восточном) и включающих типичные для этих макросклонов леса (еловые и сосновые соответственно). Район исследований (координаты центра 57.036389° с.ш., 59.552222° в.д.) расположен в южной тайге, в пределах гряды останцовых

гор осевой части Среднего Урала и ее западного склона, на водораздельной, надпойменной террасе с террасовидным склоном и участком долины р. Чусовая. По физико-географическому районированию Свердловской области территория относится к природному району низкогорий Среднего Урала с преобладанием темнохвойных лесов. Почвенный покров представлен дерново-подзолистыми, горно-бурыми лесными, дерново-луговыми почвами [6]. Климат района исследования континентальный, умеренно холодный, влажный. Средняя температура воздуха 0.3 °С, снежный покров устанавливается в конце октября–начале ноября, разрушается – с середины апреля; средняя продолжительность вегетационного периода 109 дней (с 20 мая по 7 сентября) [7].

Схема эксперимента

Работы проводили в темнохвойных лесах, образованных елью и пихтой, возрастом 80–110 лет (6ЕЗП1С+Б+Л) [8] (участок полигона «Коуровка») на трех кластерах (Cluster) (рис. 1). Каждый кластер включал 2 участка (Site), расположенных на двух уровнях приречных террас (Position) – нижней и верхней (разница высот 25–27 м). На каждом участке было заложено по три пробные площади (Plot) 5×15–20 м (65–100 м²), отходящих в разные стороны от условного центра участка. Всего заложили 18 пробных площадей, на каждой из которых измерения проводили в 10 случайных точках. Точки старались располагать не ближе 1 м от крупных стволов деревьев и избегая окон древостоя [9].

Измерения эмиссии выполнили трижды в 2021 г. (26 августа (только 1 кластер), 22 сентября и 12 октября) и 13 раз с 18 мая по 1 ноября 2022 г. каждые 10–16 дней (в основном 14 дней). Всего провели 16 туров. В каждый тур работы вели в интервале 9:00–17:00 ч. Порядок посещения участков меняли в разные туры, чтобы избежать систематической погрешности, вызванной суточной динамикой скорости эмиссии углекислого газа. Всего выполнено 2760 измерений (по 10 измерений на каждой из 18 пробных площадей за 15 полных туров (+один неполный)).

Измерение эмиссии CO_2

Скорость потока CO_2 с поверхности почвы измеряли закрытым динамическим камерным методом [10] с помощью полевого респирометра Li-8100A (Li-Cor biosciences, США). За 5–10 мин до измерения устанавливали кольца из нержавеющей стали диаметром 10.5 см и высотой 5 см,

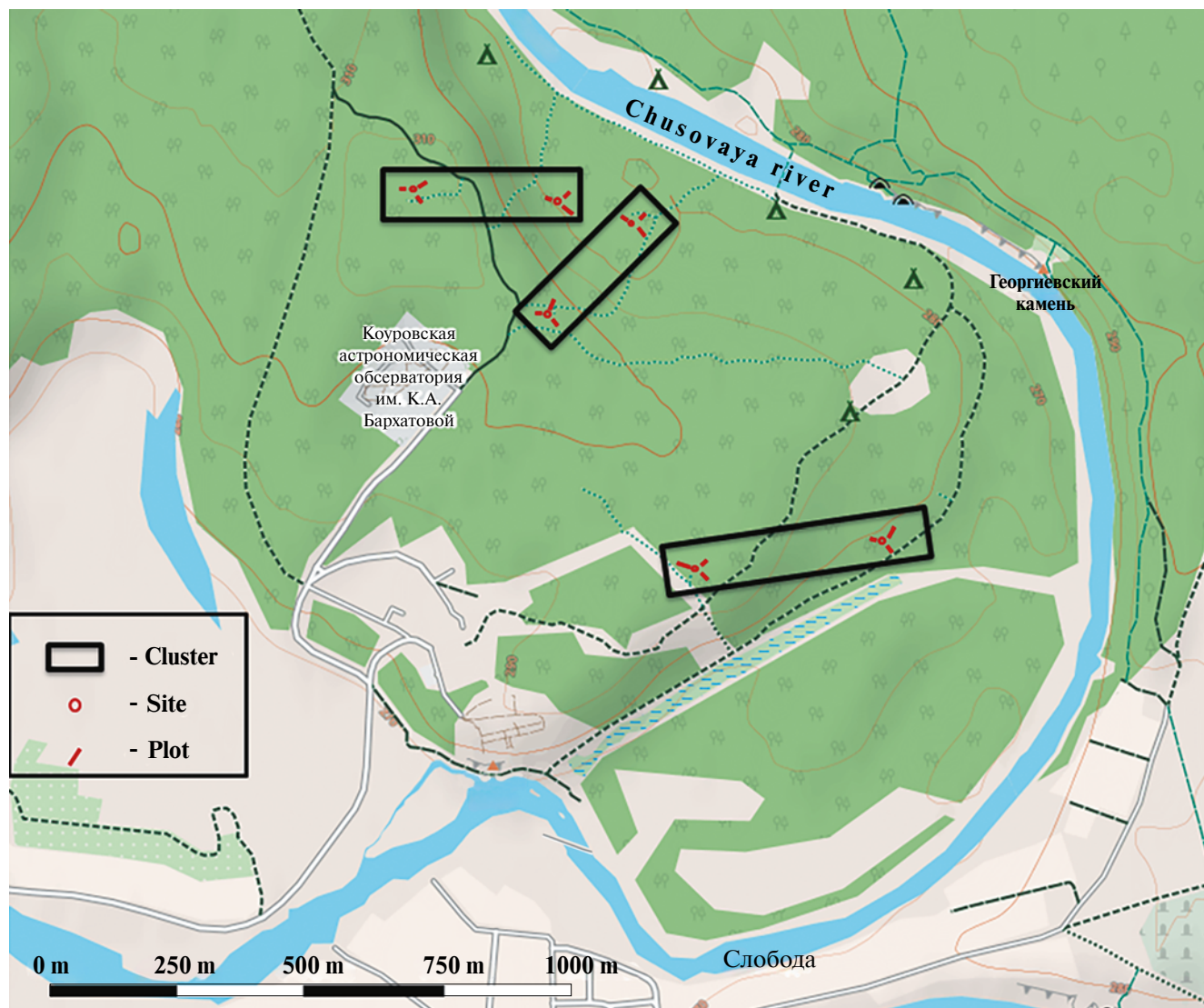


Рис.1. Схема расположения пробных площадей: Cluster – кластер, Site – участок, Plot – пробная площадь.

которые заглубляли в почву на 3 см; зеленые части растений были предварительно срезаны. Далее на кольцо помещали камеру прибора. Время одного измерения составляло 1 мин.

Одновременно измеряли температуру почвы почвенным термометром Omega 88311E, входящим в комплект респирометра, с точностью 0.1 °C (OMEGA Engineering, Великобритания), объемную влажность почвы – датчиком ThetaProbe ML2 (Delta-T devices, Великобритания), соединенным с блоком управления респирометра, с точностью 0.1%. Измерения влажности и температуры почвы проводили в непосредственной близости к точке измерения дыхания на глубине 5 см. Также измеряли температуру воздуха термодатчиком, встроенным в камеру респирометра.

Анализ данных

Статистическую обработку проводили в среде R v. 4.1.2. Статистической единицей во всех случаях считали пробную площадь, т.е. среднее значение по 10 измерениям на пробной площади за один тур. Это значение использовали как среднесуточное, так как за счет значительной микромасштабной изменчивости в лесу при большом количестве точек измерения эмиссия, определенная в конкретный момент, значимо не отличается от среднесуточной [11].

Подготовка данных. Для моделирования использовали предикторы, определенные непосредственно на месте измерений и данные дистанционного зондирования Земли. Дистанционные данные по температуре почвы и воздуха и влаж-

ности почвы, а также NDVI, основанные на продуктах спектрометра MODIS со спутников Terra и Aqua за 2022 год, получили с помощью сервиса ВЕГА-Science [12] (табл. 1).

Кроме того, использовали стандартизированный индекс эвапотранспирации осадков (the Standardized Precipitation Evapotranspiration index, SPEI) [13] из глобальной базы (<http://spei.csic.es>) [14]. Этот индекс показывает, насколько засушливым был изучаемый период относительно нормы за несколько последних десятилетий (в данной базе – с 1950 г. до 2022 г.). Величина индекса $[-0.99, 0.99]$ – показатель нормального увлажнения, $[1.00, 1.49]$ / $[-1, -1.49]$ – умеренное переувлажнение/засуха, >1.50 / <-1.50 – сильное переувлажнение/засуха.

Дистанционные данные по температурам и влажности предоставляются с шестичасовым интервалом. Однако эти значения являются продуктом расчетов алгоритма по двум непосредственным ежедневным измерениям со спутника, поэтому мы использовали их среднесуточные значения.

В случае NDVI мы имели набор значений из 1-, 4- и 8-дневных композитных снимков. Многодневные композитные снимки призваны уменьшить влияние облачности на определение NDVI, однако в исследуемом районе количество пасмурных дней может быть больше 200, поэтому в данных все равно случаются пропуски. Для получения ряда ежедневных значений NDVI мы заполнили пропущенные данные средними из соседних значений и затем усреднили все три ряда (1-, 4- и 8-дневные).

Вне вегетационного сезона (до 20 апреля и после 30 сентября) в связи с неадекватно сильными колебаниями NDVI (из-за большого количества облачных дней) использовали самые низкие значения на границе вегетации – около 0.69. Это минимальная величина, которая была в нашей обучающей выборке, поэтому значения ниже не влияли на результат из-за принципа работы деревьев решений, на которых основана работа алгоритма RF.

Выбор предикторов проводили с помощью пакета “Boruta” [15] и анализа корреляций. Этот алгоритм основан на алгоритме RF и оценивает вклад конкретного предиктора в точность модели. Для этого он создает по несколько копий каждого предиктора и случайным образом перемешивает значения в каждой копии. Такие предикторы называются «теневыми». Далее алгоритм строит множество моделей, оценивает точность моделей с теневыми предикторами и разницу с начальной моделью с реальными предикторами. По этим разностям рассчитываются безразмерные Z-оценки (Z-scores), определяются значимости предикторов и их границы (importance).

Расчет годовых потоков. На первом этапе для оценки годовых потоков на основе полевых измерений дыхания почвы с помощью алгоритма random forest (“randomForest” package [16]) мы создали две модели: «полную» (с использованием всех переменных) и «упрощенную» (только с отобранными предикторами). Параметры моделирования (количество деревьев, максимальное количество признаков, рассматриваемых для

Таблица 1. Список используемых предикторов

Предиктор	Описание	Примечание
Soil temperature	Температура почвы на глубине 5 см	Измерены в непосредственной близости от точки измерения
Air temperature	Температура воздуха	
Soil moisture	Объемная влажность почвы на глубине 5 см	
Cluster	Номер кластера	
Position	Верхняя или нижняя террасы	
Soil temperature (remote)	Температура почвы на глубине 10 см	По данным дистанционного зондирования, разрешение 1 км
Air temperature (remote)	Температура воздуха	
Soil moisture (remote)	Влажность почвы на глубине 10 см	
NDVI	Нормализованный дифференциальный вегетационный индекс, обычно хорошо коррелирует с общей наземной фитомассой	Продукт MYD09Q1, разрешение 250 м
SPEI 1–12	Индекс сухости за 1–12 месяцев, предшествующий измерениям	Из базы SPEI Global Drought Monitor, пространственное разрешение 0.5°

разделения) подобрали поиском по сетке, качество оценивали путем кросс-валидации с 5 фолдами в пакете “caret” [17]. Качество модели оценивали по коэффициенту детерминации R^2 и среднему квадрату отклонений расчетных значений от измеренных (MSE (Mean of squared errors)).

Полученные модели использовали для оценки годовых потоков CO_2 из почвы двумя способами:

1. Экстраполяция на основании ежедневных данных о величинах выбранных предикторов.

2. Комбинированный подход использования модели случайного леса и одной из двух моделей:

2.1) модели расчета годовой эмиссии на основе определения вклада дыхания почвы за летний период и среднегодовой температуры [4];

2.2) модели на основании расчета эмиссии за период с температурами выше $5\text{ }^\circ\text{C}$ [3].

При оценке годовых потоков статистической единицей служил «участок», т.е. каждым из способов мы получали 6 значений годовых потоков, что позволило нам оценить пространственный размах результатов.

В качестве референсных значений использовали величины годового потока из глобальной базы дыхания почвы (Soil respiration data base v. 5.0 (SRBD) [18]). Выбрали значения, полученные для лесов севернее 55° с.ш. с применением инфракрасных газоанализаторов (IRGA): во всем северном полушарии ($n = 170$) и только в России ($n = 39$). Для оценки влияния зимней эмиссии на величину рассчитанного годового потока мы взяли значения зимнего дыхания из исследования ельника в Канаде [19], потому что он был близок по климату, и в статье была приведена формула, по которой мы смогли рассчитать скорость эмиссии для температур почвы $0, -1, -3$ и $-5\text{ }^\circ\text{C}$ ($0.13-0.16\text{ г (C-CO}_2)/\text{м}^2 \cdot \text{сут}$) (при дальнейшем понижении температуры показатели не менялись). Эти рассчитанные величины мы добавили в обучающую выборку и оценили годовой поток.

Данные о скорости эмиссии углекислого газа, полевым температурам почвы и воздуха, влажности почвы, использованные в статье, доступны в депозитории [20].

РЕЗУЛЬТАТЫ

Полевые измерения

Дыхание почвы лежало в диапазоне $1.52-9.81\text{ г (C-CO}_2)/\text{м}^2 \cdot \text{сут}$ и имело выраженную сезонную динамику с максимальными значениями в середине июля. В связи с засушливой погодой,

установившейся во второй половине лета, мы наблюдали относительно низкие значения скорости выделения CO_2 из почвы, даже при температуре почвы выше, чем в начале июля.

Температура почвы во время измерений лежала в диапазоне $3.2-24.2\text{ }^\circ\text{C}$, воздуха $-4.2-29.0\text{ }^\circ\text{C}$, влажность почвы $-8.3-61.7\%$ (рис. 2). Данные температуры, полученные дистанционно, хорошо соответствовали полевым, однако температура воздуха, измеренная в полевых условиях, была стабильно выше среднесуточной, вычисленной по данным дистанционного зондирования. В отличие от температуры полевые данные по влажности почвы хуже соответствовали дистанционным.

Выбор предикторов

Анализ важности предикторов с использованием метода Boruta показал, что значимость фактора «пробная площадь» (Plot) оказалась ниже, чем у теневых предикторов (рис. 3а). Значимость факторов «Position» и «Cluster» была немного выше, но пересекалась с границами значимости теневых предикторов. Поэтому мы приняли решение исключить эти переменные из «упрощенной» модели. Наибольшее влияние на точность модели оказали температура воздуха и почвы, а также влажность почвы. Причем данные дистанционного зондирования имели большую значимость, чем те, что были получены непосредственно на месте измерения дыхания почвы. Кроме того, полевые и дистанционные данные сильно коррелировали друг с другом (рис. 4). Поэтому мы исключили результаты полевых измерений температур и влажности для создания модели с меньшим числом предикторов. Меньшее влияние на точность оказывал индекс NDVI, а значимость индекса SPEI находилась примерно на одном уровне, за исключением SPEI10. После исключения полевых наблюдений различия в значимости индексов SPEI стали более заметными (рис. 3б). Для упрощенной модели из 12 предикторов мы оставили три показателя SPEI: за 1, 3 и 10 месяцев. Таким образом, для «упрощенной» модели мы выбрали только предикторы, основанные на данных дистанционного зондирования: температуру воздуха и почвы, влажность почвы, NDVI и индексы SPEI1, SPEI3 и SPEI10.

Годовые потоки CO_2

Полная и упрощенная модели были близки по характеру годовой динамики и показателям точности (рис. 5, табл. 2); годовые величины эмиссии CO_2 для каждого подхода не отличались при использовании

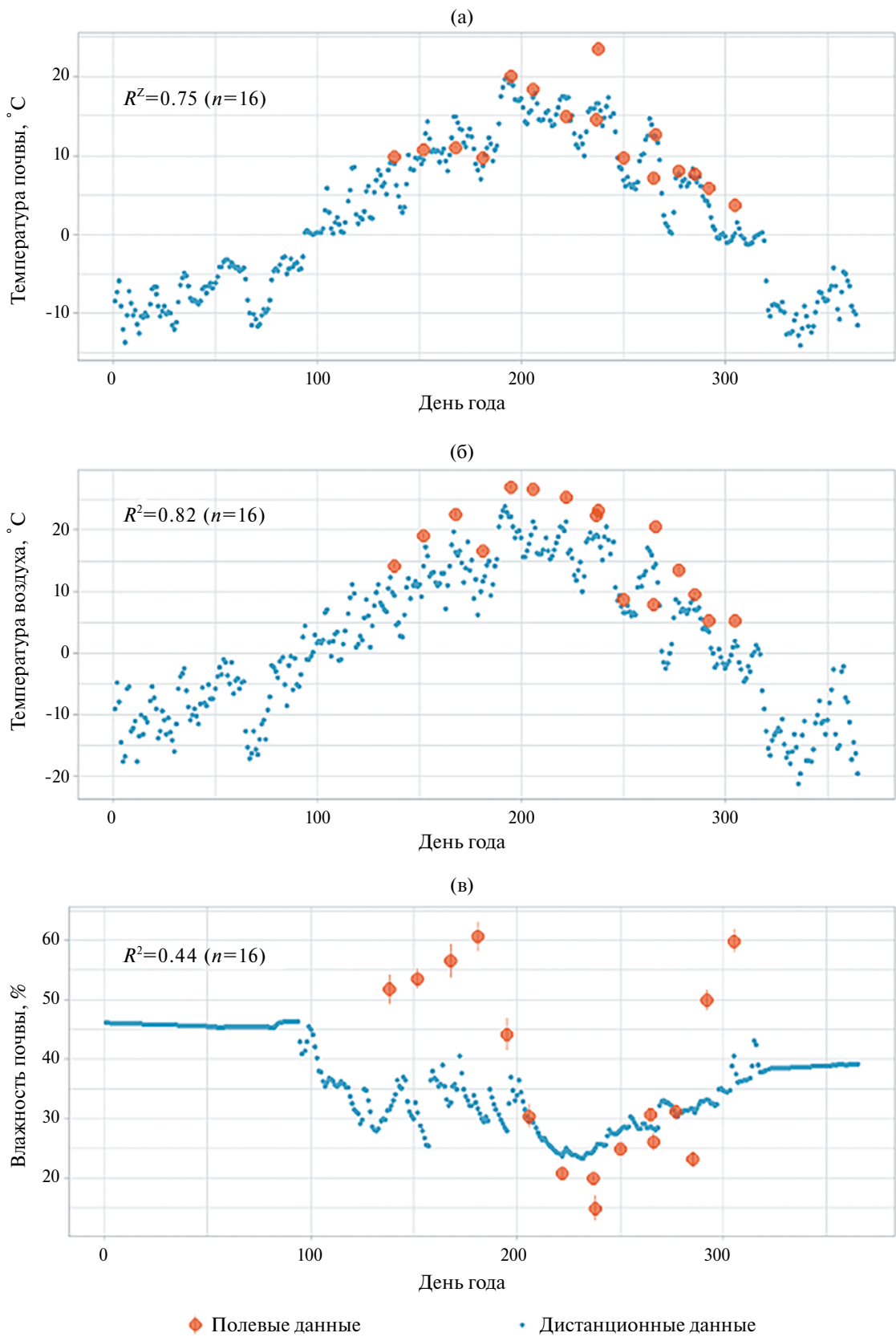


Рис. 2. Динамика температуры почвы (а) и воздуха (б) и влажности почвы (в) в течение года.

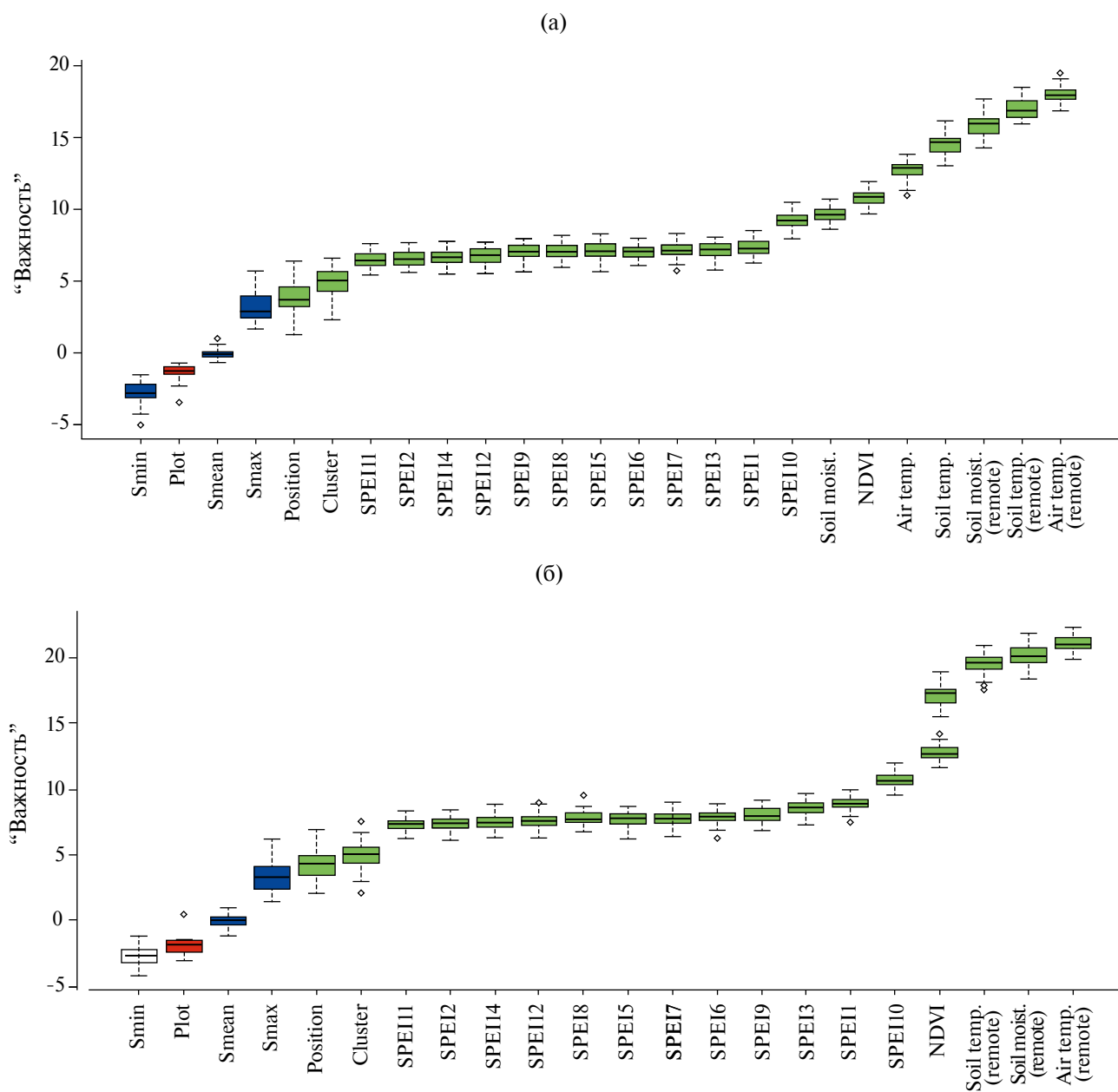


Рис. 3. “Важность” предикторов, оцененных алгоритмом Boruta: а – все предикторы, б – исключены предикторы, измеренные в полевых условиях. Smin, Smean, Smax – теневые предикторы.

обеих моделей. Однако результаты разных подходов значительно отличались друг от друга (см. табл. 2). При оценке модели, дополненной литературными данными о зимнем дыхании, величина годовой эмиссии при прямой экстраполяции составила 830.3 ± 6.4 г С/м² · год, т.е. начальная модель завышала годовые результаты почти в 1.4 раза. При усреднении результатов подходов, основанных на классической регрессии, величина годового потока составила 851.6 ± 8.0 г С/м² · год. Мы не обнаружили значимых различий между результатами расчетов и данными SRBD.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Полевые измерения

Полученные абсолютные величины дыхания почвы ($1.52\text{--}9.81$ г (С-СО₂)/м² · сут) близки к значениям, которые обычно регистрируют для лесов умеренных широт: до 4.5 ± 0.8 [21], 2.9 ± 0.7 [22], $4.6\text{--}11.7$ [23], $3.6\text{--}4.6$ [24], $1.3\text{--}14.4$ [9, 25]. Сезонная динамика, характеризующаяся максимальными выбросами СО₂ в летние месяцы, сходна с динамикой, описанной для южно-таежных [26] и северо-таежных лесов Восточной Сибири [24] и Центральной

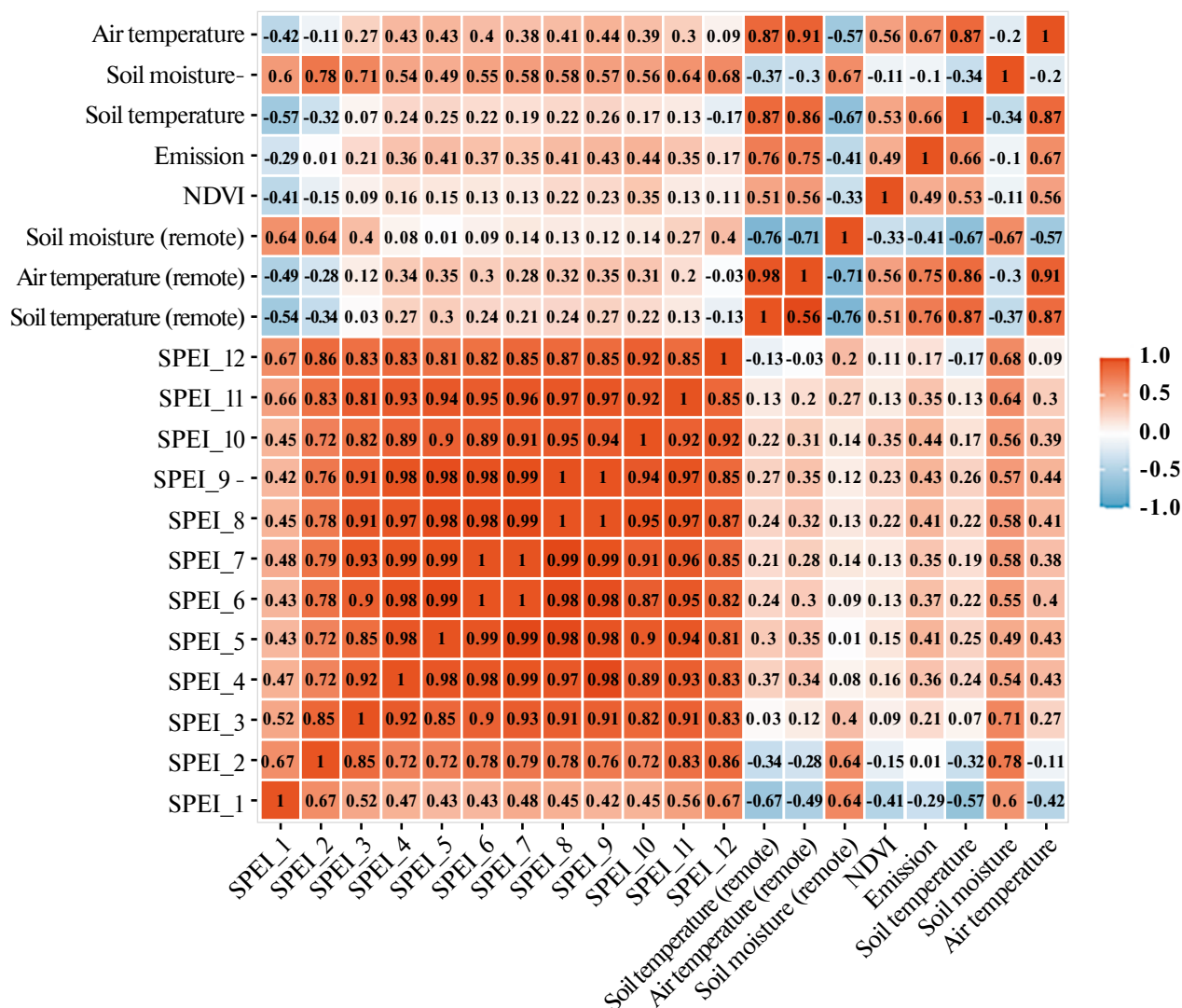


Рис. 4. Корреляции предикторов друг с другом и со скоростью дыхания почвы.

Таблица 2. Сравнение годовой почвенной эмиссии CO₂ лесов карбонового полигона «Урал-Карбон» (г С/м²·год), оцененной с помощью разных подходов и литературных данных (среднее ± ст.откл., n = 6)

Подход	Использованная модель	
	Полная <i>R</i> ² = 0.92, <i>MSE</i> = 0.22	Упрощенная <i>R</i> ² = 0.89, <i>MSE</i> = 0.31
Прямая экстраполяция	1187.31 ± 45.03	1152.86 ± 7.83
На основе летней эмиссии	966.99 ± 49.32	965.99 ± 10.35
На основе эмиссии при <i>T</i> > 5 °С	737.11 ± 27.73	737.15 ± 5.55
По данным SRBD	Все леса севернее 55° с.ш. 751.85 ± 477.29 (n = 170)	Леса РФ севернее 55° с.ш. 687.35 ± 273.42 (n = 39)

Примечание. SRBD – глобальная база по почвенному дыханию [18].

Сибири [27]. Во второй половине лета (между 220 и 240 днями), несмотря на самые высокие в году температуры почвы (более 15 °С), низкая влажность была причиной того, что значения дыхания почвы

почти не поднимались выше 5 г (С-CO₂)/м² · сут, хотя даже при низкой температуре, но более высокой влажности почвы скорость дыхания достигала этой величины и даже превышала. Известно, что

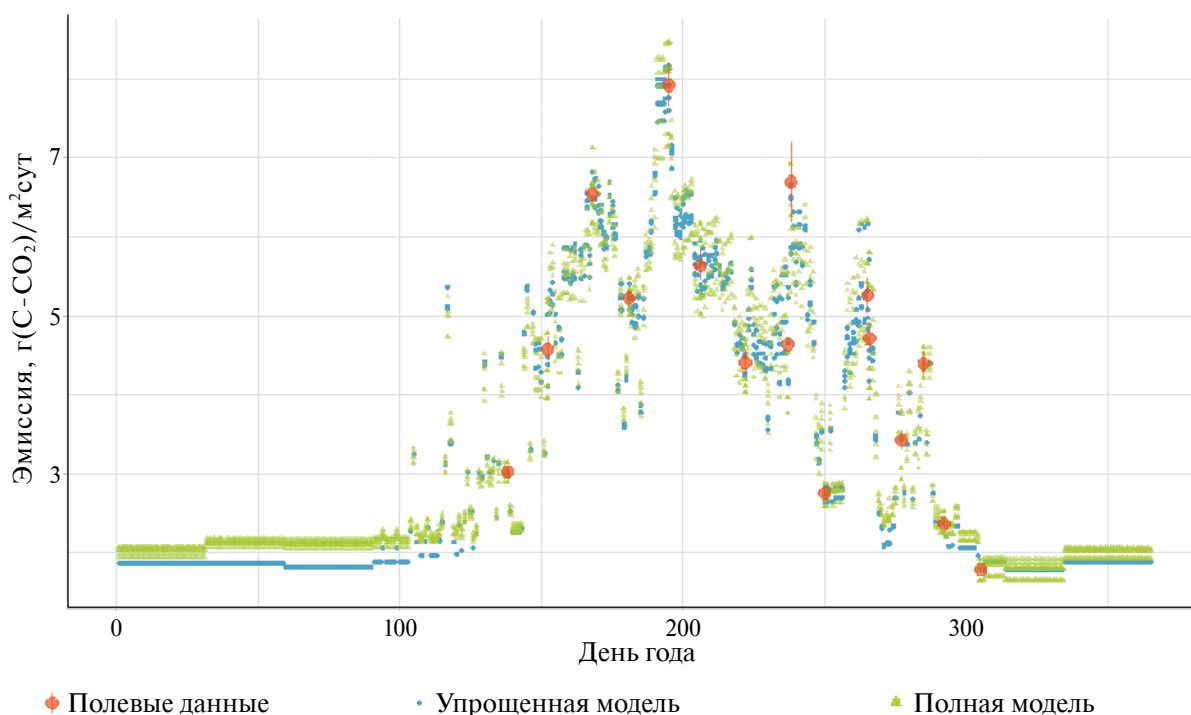


Рис. 5. Динамика дыхания почвы в течение года.

при экстремальных значениях влажности почвы (слишком низких или высоких) уменьшаются абсолютные показатели дыхания почвы и изменяется его температурная чувствительность [10, 28]. Кроме того, несмотря на низкую корреляцию со скоростью эмиссии углекислого газа, влажность почвы имеет высокую «значимость» для точности модели. Это можно интерпретировать как факт ее сильного влияния на дыхание почвы в отдельные моменты времени. Поэтому мы делаем вывод, что дыхание почвы во время летней засухи 2022 г. лимитировалось именно ее влажностью, а не температурой, что отмечено и в других исследованиях [28].

Очевидно, что в связи с засушливым летом пространственная вариабельность эмиссии CO_2 оказалась меньше временной: пространственные предикторы («Cluster», «Position», «Plot») были наименее значимыми для точности модели.

Подбор предикторов

В упрощенной модели мы оставили практически только временные предикторы, при выборе которых руководствовались не только численными показателями (вклад в точность модели и степень взаимокорреляции), но и потенциальной интерпретируемостью механизмов их действия на скорость почвенной эмиссии CO_2 . В этом плане

SPEI1 соответствует ближайшим предшествующим состояниям влажности почвы в экосистеме, что может отчасти нивелировать непропорциональные скачки величины дыхания при повышении влажности после некоторого засушливого периода [10]. Важность SPEI3 вполне согласуется с данными о том, что влажность весенних месяцев значительно влияет на годовые потоки углекислого газа из почвы [2]. Пока мы использовали в модели данные только за один год, поэтому возможно, что при анализе многолетних данных (или при анализе суммарных годовых потоков) мы выявим большой вклад SPEI3 именно для мая каждого года, а не для конкретной даты измерения. Также вероятно, что и высокая важность SPEI10 связана с анализом данных по сути одного года, а при анализе многолетней динамики его заменит SPEI12 (т.е. какой по влажности был год, предшествующий измерениям), подобно тому, как было нами показано для сосняков Среднего Урала [25].

Годовые потоки

И полная, и упрощенная модели зависимости скорости почвенной эмиссии от факторов среды на основе случайного леса оказалась выше, чем точность модели для сосняков Среднего Урала с сопоставимым набором предикторов ($R^2 = 0.77$, $MSE = 0.8$) [25]. Их качество также превосходило

локальную модель на основе алгоритма RF для горных лесов в южных Скалистых горах ($R^2 = 0.44$, $MSE = 0.8$) [29] и модель глобального масштаба лесов северного полушария с более высоким R^2 (до 0.86, $MSE = 2.16$) [30].

Интересна разница в полученных разными подходами оценках годовых потоков CO_2 из почвы. Мы можем уверенно утверждать, что использование прямой экстраполяции модели RF на значения предикторов вне пределов обучающей выборки привело к завышению итоговой величины. Дело в самом принципе работы деревьев решений, на которых основан этот алгоритм: при расщеплении по крайнему значению предиктора в обучающей выборке алгоритм присваивает одинаковую величину целевой переменной всему множеству, находящемуся за этой крайней точкой [31]. Например, в нашей модели скорость дыхания почвы была одинаковой при всех температурах почвы ниже $3^\circ C$, что естественно совершенно не соответствует действительности, потому что при низких значениях температуры почвы ее дыхание особенно к ней чувствительно и быстро уменьшается при понижении [19, 32]. Скорость эмиссии CO_2 зимой может различаться на порядок величин, но все равно иметь очень низкие значения: например, $0.5 \text{ г } (C-CO_2)/\text{м}^2 \cdot \text{сут}$ [33] или $0.2-0.3$ — при $0^\circ C$, около $0.05 \text{ г } (C-CO_2)/\text{м}^2 \cdot \text{сут}$ при $-3^\circ C$ [34]. Поэтому даже использование литературных данных о зимней эмиссии значительно улучшило точность нашей модели при прямой экстраполяции.

Сложнее ответить на вопрос, какой из двух комбинированных подходов дает значения ближе к реальным? Каждый из них по сути представляет собой классическую регрессионную модель, учитывающую только температуру, игнорирующую изменения других факторов и имеющую свои источники неопределенности. Они основаны на большом объеме данных: «Дыхание почв России» [4] и глобальной базе данных SRBD [3], но имеют не абсолютную точность ($R^2 = 0.95$ и $R^2 = 0.81$ соответственно), т.е. в каждый конкретный год результаты расчетов этих подходов могут отклоняться от идеальной регрессионной зависимости, которой они описываются. Поэтому усреднение результатов разных подходов может иметь эффект ансамблирования, когда среднее значение разных прогнозов находится ближе к истинным значениям, чем каждый отдельный прогноз.

Высокая сходимость полевых и дистанционных данных по температурам почвы и воздуха и больший вклад дистанционных данных в точность модели позволяют рекомендовать использование

именно спутниковых данных для создания моделей дыхания почвы, что позволит унифицировать получение этих величин и избежать погрешностей, связанных с оборудованием и человеческим фактором при моделировании и определении потоков в разных регионах.

ЗАКЛЮЧЕНИЕ

Таким образом, в 2022 г. поток углекислого газа из почвы участка темнохвойных лесов карбонового полигона был незначимо выше средних величин, регистрируемых для лесов севернее 55° с.ш. Однако эти результаты еще необходимо уточнять. Поэтому, помимо получения этих первых оценок, важными результатами работы мы считаем выводы методического характера:

1. Пространственная изменчивость в пределах полигона мала, и ее учет почти не влияет на точность модели.

2. Спутниковые данные по температурам почвы и воздуха хорошо соответствуют полевым измерениям. В целом данные дистанционного зондирования имеют большее влияние на точность модели и позволяют унифицировать способ получения этих данных разными исследователями.

3. Прямая экстраполяция случайным лесом дает завышенные значения годовых потоков. Для устранения данной погрешности необходимо проводить хотя бы единичные зимние измерения либо использовать сочетание комбинированных подходов.

4. За счет отбора наиболее значимых предикторов можно получить более интерпретируемую и практически не уступающую в точности модель дыхания почвы.

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена в рамках государственного задания Министерства науки и высшего образования (тема FEUZ-2024-0011).

КОНФЛИКТ ИНТЕРЕСОВ

Автор заявляет об отсутствии конфликта интересов.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая работа не содержит исследований с участием людей или животных в качестве объектов изучения.

СПИСОК ЛИТЕРАТУРЫ

1. *Yang M., Yu G.R., He N.P.* et al. A Method for Estimating Annual Cumulative Soil/Ecosystem Respiration and CH₄ Flux from Sporadic Data Collected Using the Chamber Method // *Atmosphere*. 2019. V. 10. № 10. Art. 623.
<https://doi.org/10.3390/atmos10100623>
2. *Курганова И.Н., Лопес де Гереню В.О., Мяскина Т.Н.* и др. Эмиссия CO₂ из почв различных экосистем южно-таежной зоны: анализ данных непрерывных 12-летних круглогодичных наблюдений // Доклады РАН. 2011. Т. 436. № 6. С. 843–846.
3. *Mukhortova L., Schepaschenko D., Moltchanova E.* et al. Respiration of Russian soils: Climatic drivers and response to climate change // *Science of the Total Environment*. 2021. V. 785. Art. 147314.
<https://doi.org/https://doi.org/10.1016/j.scitotenv.2021.147314>
4. *Кудряков В.Н., Курганова И.Н.* Дыхание почв России: анализ базы данных, многолетний мониторинг, общие оценки // *Почвоведение*. 2005. № 9. С. 1112–1121.
5. Карбоновые полигоны (официальный сайт) / URL: <https://carbon-polygons.ru/>
6. *Гафуров Ф.Г.* Почвы Свердловской области. Екатеринбург: Изд-во Урал. ун-та, 2008. 417 с.
7. *Усольцев В.А., Воробейчик Е.Л., Бергман И.Е.* Биологическая продуктивность лесов Урала в условиях техногенного загрязнения: исследование системы связей и закономерностей. Екатеринбург: Изд-во УГЛТУ, 2012. 365 с.
8. *Фомин В.В., Рогачев В.Е., Аганитов Е.М.* и др. Депонирование углерода основными лесообразующими древесными породами карбонового полигона Свердловской области // *Леса России и хозяйство в них*. 2024. № 4(91). С. 4–16.
<https://doi.org/10.51318/FRET.2024.91.4.001>
9. *Сморкалов И.А., Воробейчик Е.Л.* Влияние отдельных деревьев на дыхание почвы лесных экосистем в условиях промышленного загрязнения // *Почвоведение*. 2023. № 9. С. 1116–1127.
10. *Luo Y., Zhou X.* Soil respiration and the environment. Burlington: Acad. Press, 2006. 316 p.
11. *Карелин Д.В., Почикалов А.В., Замолодчиков Д.Г., Гитарский М.Л.* Факторы пространственно-временной изменчивости потоков CO₂ из почв южно-таежного ельника на Валдае // *Лесоведение*. 2014. № 4. С. 56–66.
12. *Луян Е.А., Савин И.Ю., Барталев С.А.* и др. Спутниковый сервис мониторинга состояния растительности («Вега») // *Современные проблемы дистанционного зондирования Земли из космоса*. 2011. Т. 8. № 1. С. 190–198.
13. *Vicente-Serrano S.M., Beguería S., López-Moreno J.I.* A multiscale drought index sensitive to global warming: The standardized precipitation evapotranspiration index // *Journal of Climate*. 2010. V. 23. № 7. P. 1696–1718.
<https://doi.org/10.1175/2009jcli2909.1>
14. *Beguería S., Vicente-Serrano S.M., Reig F., Latorre B.* Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring // *International Journal of Climatology*. 2014. V. 34. № 10. P. 3001–3023.
<https://doi.org/10.1002/joc.3887>
15. *Kursa M.B., Rudnicki W.R.* Feature selection with the Boruta package // *Journal of Statistical Software*. 2010. V. 36. № 11. P. 1–3.
<https://doi.org/10.18637/jss.v036.i11>
16. randomForest: Breiman and Cutler's Random Forests for classification and regression version 4.6-14 / URL: <https://cran.r-project.org/package=randomForest>
17. caret: Classification and Regression Training version 6.0-86 / URL: <https://cran.r-project.org/package=caret>
18. *Jian J., Vargas R., Anderson-Teixeira K.J.* et al. A global database of soil respiration data. Version 5.0. 2021.
<https://doi.org/10.3334/ORNLDAAC/1827>
19. *Mavrovic A., Sonnentag O., Lemmetyinen J.* et al. Environmental controls of winter soil carbon dioxide fluxes in boreal and tundra environments // *Biogeosciences*. 2023. V. 20. № 24. P. 5087–5108.
<https://doi.org/10.5194/bg-20-5087-2023>
20. *Smorkalov I.A.* Soil respiration of spruce-fir forests at the carbon supersite “Ural-Carbon”, Russia, Sverdlovsk region on 2021–2022. 2023.
<https://doi.org/10.17632/8ns8zprx9c.1>
21. *Осипов А.Ф.* Влияние межгодовых различий метеорологических характеристик вегетационного периода на эмиссию CO₂ с поверхности почвы среднетаежного сосняка бруснично-лишайникового (Республика Коми) // *Почвоведение*. 2018. Т. 12. С. 1455–1463.
<https://doi.org/10.1134/S0032180X18120080>
22. *Кадулин М.С., Концик Г.Н.* Эмиссия CO₂ почвами в зоне влияния горно-металлургического комбината “Североникель” в Кольской Субарктике // *Почвоведение*. 2013. № 11. С. 1387–1396.
<https://doi.org/https://doi.org/10.7868/S0032180X13110063>
23. *Kozlov M.V., Zvereva E.L., Zverev V.E.* Impacts of point pollutants on terrestrial biota: Comparative analysis of 18 contaminated areas. Dordrecht: Springer, 2009. 466 p.
24. *Takakai F., Desyatkin A.R., Lopez C.M.L.* et al. Influence of forest disturbance on CO₂, CH₄ and N₂O fluxes from larch forest soil in the permafrost taiga region of eastern Siberia // *Soil Science and Plant Nutrition*. 2008. V. 54. № 6. P. 938–949.
<https://doi.org/https://doi.org/10.1111/j.1747-0765.2008.00309.x>
25. *Сморкалов И.А.* Изменчивость дыхания почвы: оценка вклада пространства и времени с

- помощью алгоритма random forest // Экология. 2022. № 4. С. 299–311.
<https://doi.org/https://doi.org/10.31857/S0367059722040059>
26. Kurganova I., De Gerenyu V.L., Rozanova L. et al. Annual and seasonal CO₂ fluxes from Russian southern taiga soils // Tellus B: Chemical and Physical Meteorology. 2003. V. 55. № 2. P. 338–344.
<https://doi.org/10.3402/tellusb.v55i2.16724>
 27. Shibistova O., Lloyd J., Zrazhevskaya G. et al. Annual ecosystem respiration budget for a *Pinus sylvestris* stand in central Siberia // Tellus, Series B: Chemical and Physical Meteorology. 2002. V. 54. № 5. P. 568–589.
 28. Yan J., Feng Y., Li J. et al. Response of soil respiration and Q₁₀ to temperature and moisture in naturally regenerated and bare lands based on an 11-year observation period // CATENA. 2022. V. 208. Art. 105711.
<https://doi.org/https://doi.org/10.1016/j.catena.2021.105711>
 29. Berryman E.M., Vanderhoof M.K., Bradford J.B. et al. Estimating soil respiration in a subalpine landscape using point, terrain, climate, and greenness data // J. Geophys. Res.-Biogeosci. 2018. V. 123. № 10. P. 3231–3249.
<https://doi.org/10.1029/2018jg004613>
 30. Shiri N., Shiri J., Kazemi M.H., Xu T.R. Estimation of CO₂ flux components over northern hemisphere forest ecosystems by using random forest method through temporal and spatial data scanning procedures // Environmental Science and Pollution Research. 2021. P. Published online.
<https://doi.org/10.1007/s11356-021-16501-x>
 31. Hengl T., Nussbaum M., Wright M.N. et al. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables // PeerJ. 2018. V. 6. Art. e5518.
<https://doi.org/10.7717/peerj.5518>
 32. Kurganova I., Lopes de Gerenyu V., Khoroshaev D. et al. Temperature sensitivity of soil respiration in two temperate forest ecosystems: the synthesis of a 24-year continuous observation // Forests. 2022. V. 13. № 9. Art. 1374.
 33. Reinmann A.B., Templer P.H. Increased soil respiration in response to experimentally reduced snow cover and increased soil freezing in a temperate deciduous forest // Biogeochemistry. 2018. V. 140. № 3. P. 359–371.
<https://doi.org/10.1007/s10533-018-0497-z>
 34. Monson R.K., Lipson D.L., Burns S.P. et al. Winter forest soil respiration controlled by climate and microbial community composition // Nature. 2006. V. 439. № 7077. P. 711–714.
<https://doi.org/10.1038/nature04555>

ESTIMATES OF ANNUAL CARBON DIOXIDE FLUXES FROM SOIL IN SPRUCE FORESTS OF THE «URAL-CARBON» CARBON MEASUREMENT SUPERSITE BASED ON INCOMPLETE TIME SERIES WITH CLASSICAL REGRESSION APPROACHES AND MACHINE LEARNING

I. A. Smorkalov^{a, b}

^aUral Federal University, Russia 620002 Yekaterinburg

^bInstitute of Plant and Animal Ecology, Ural Branch, Russian Academy of Sciences, Russia 620144 Yekaterinburg

*e-mail: ivan.a.smorkalov@gmail.com

Abstract — Annual carbon dioxide flux from soils of different biomes plays a key role in creating global climate models and in analyzing carbon cycles in terrestrial ecosystems. However, there are significant gaps in such studies at a regional scale. Due to the high complexity of obtaining daily soil respiration values, various modeling methods are used. In this work, based on 2760 soil respiration measurements in spruce forests of the «Ural-Carbon» carbon measurement supersite (Middle Urals) carried out in autumn 2021 and from April to October 2022, annual soil respiration values were estimated using classical regression approaches and machine learning. We also investigated the dependence of the results on the complexity of the model (number of predictors) and the methods used (extrapolation by the random forest model and combined approaches to estimate winter CO₂ fluxes). The «simplified» model with 7 predictors showed only a slight decrease in accuracy compared to the full model with 21 predictors ($R^2 = 0.89$, $MSE = 0.31$ vs. $R^2 = 0.92$, $MSE = 0.22$). Predictors based on remote sensing turned out to be more significant for the accuracy of the model than data measured in the field. Although the initial results of different approaches varied, adding winter respiration values taken from the literature to the random forest model and averaging the values of the combined approaches allowed us to achieve similar values of annual soil respiration: 830.3 ± 6.4 and 851.6 ± 8.0 g C/m²year, respectively.

Keywords: CO₂ emissions, carbon cycle, machine learning, forest ecosystems, carbon polygons, environmental factors