

ИЗМЕНЧИВОСТЬ ДЫХАНИЯ ПОЧВЫ: ОЦЕНКА ВКЛАДА ПРОСТРАНСТВА И ВРЕМЕНИ С ПОМОЩЬЮ АЛГОРИТМА RANDOM FOREST

© 2022 г. И. А. Сморкалов*

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

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

Поступила в редакцию 25.01.2022 г.

После доработки 09.02.2022 г.

Принята к публикации 16.02.2022 г.

Моделирование дыхания почвы (эмиссии углекислого газа с ее поверхности) важно для анализа и прогнозирования изменений цикла углерода в наземных экосистемах. Наряду с классическими регрессионными моделями в последнее время используют методы машинного обучения, основанные на нейросетях или ансамблях регрессионных деревьев. Однако построенные на основе этих методов модели часто применяют лишь в качестве “черного ящика”, что препятствует анализу механизмов процессов. В работе показаны возможности алгоритма Random Forest для оценки влияния разных факторов на дыхание почвы на основе измерения важности предикторов. Используя разделение дисперсий, предикторы были отнесены либо к пространственным (тип биотопа, тип почвы, характер растительности, влажность почвы), либо к временным (температура почвы и воздуха, NDVI, LAI, FPAR, SPEI). Модели построены на материале 5670 измерений дыхания в течение пяти вегетационных сезонов (2012–2016 гг.) на 30 пробных площадях в сосновых лесах и на лугах южной тайги, различающихся характером растительности и почв, но расположенных в пределах небольшой территории. Модели включали разные наборы предикторов (все, только временные, только пространственные, только температура и влажность), их точность достигала $R^2 = 0.88$ ($MSE = 0.47$). Показано, что дыхание сильнее всего зависело от временных факторов (76–91% важности); пространственные факторы сильнее влияли на дыхание в лесах, чем на лугах.

Ключевые слова: эмиссия CO₂, цикл углерода, машинное обучение, лесные экосистемы, температура, влажность, экологические факторы

DOI: 10.31857/S0367059722040059

В анализе и прогнозировании изменений цикла углерода важную роль играет математическое моделирование. Для улучшения качества моделей важно знать ключевые факторы, влияющие на интенсивность основных компонентов цикла в разных временных и пространственных масштабах [1].

Вопросы математического моделирования дыхания почвы разрабатывают уже более 50 лет [2]. В качестве предикторов чаще всего используют температуру [3, 4] и/или влажность почвы [3, 5–7], поскольку эти параметры легко определить непосредственно в момент измерения дыхания. Характер выявленных зависимостей между дыханием и этими предикторами может быть линейным [8], квадратичным [9], экспоненциальным [10] или степенным [11].

При использовании только температуры и влажности в качестве предикторов в каждом конкретном случае авторы приводят модели, кото-

рые удовлетворительно описывают практически только их данные и “работают” в масштабе десятков метров. Даже в рамках одного исследования для отдельных точек характер зависимостей может различаться [3]. Это связано с тем, что не учитываются многие факторы, обуславливающие пространственную и временную изменчивость потоков CO₂ [12].

Достоинством классических методов множественной регрессии следует признать возможность напрямую оценить важность разных предикторов на основе коэффициентов модели и выбрать наиболее значимые из них, например при пошаговой регрессии [12]. Однако эти методы имеют существенные ограничения, связанные с требованиями к используемому типу данных (сложно оперировать перемерными в номинальной шкале) и связям между отдельными факторами (важно отсутствие корреляции между предикторами). Поэтому в настоящее время активно

развивается альтернативный подход — моделирование на основе машинного обучения.

Один из самых известных и популярных методов машинного обучения — искусственные нейронные сети. При их использовании задействуют от нескольких [16] до нескольких десятков переменных [17]. Такие модели описывают дыхание почвы как на локальном [18], так и на глобальном [19] уровнях. Однако авторы практически никогда не раскрывают архитектуру сетей, ограничиваясь лишь перечислением переменных, от которых строится зависимость, и указанием числа узлов (блоков) в нейросети [19, 20]. По этим причинам такими моделями, кроме самих авторов, никто не может воспользоваться, в том числе для верификации на другом материале.

Для того чтобы результатом работы нейросети был не только некий “черный ящик”, но и понимание механизмов процессов, разработана специальная группа методов искусственного интеллекта — объяснительный искусственный интеллект (explainable artificial intelligence). Рекомендовано даже отказаться от использования “черных ящиков” в пользу изначально интерпретируемых моделей [21].

Одним из таких типов интерпретируемых моделей является алгоритм “случайного леса” (Random Forest, далее в тексте — RF), который находит все более широкое применение в задачах регрессии, классификации и выявления наиболее информативных признаков [22]. Алгоритм кардинально отличается от подхода, используемого в нейросетях. Он основан на построении большого числа (ансамбля) деревьев решений, каждое из которых строится на подвыборке, получаемой из исходной выборки с помощью случайного выбора и данных, и части из имеющихся предикторов [22, 23]. Его высокая точность обеспечивается усреднением ответов многих деревьев. К основным достоинствам метода можно отнести: 1) относительную защищенность от переобучения, даже если число признаков превышает количество наблюдений; 2) для настройки достаточно всего двух параметров (количество деревьев и максимальное число признаков, используемых для разделения); 3) возможность использовать признаки, измеренные в разных шкалах (отношений, порядковой и номинальной); 4) возможность оценивать важность используемых предикторов для точности модели (feature importances); 5) устойчивость к выбросам в данных. Основными недостатками метода следует считать неспособность к экстраполяции, тяжеловесность моделей и плохую работу с линейными зависимостями.

Для моделирования дыхания почвы случайный лес только начинают использовать [24–31], поэтому он пока еще уступает в популярности классической регрессии. Хотя благодаря RF до-

стигают большей точности по сравнению с классическими моделями, в выполненных работах задействованы не все возможности алгоритма для анализа влияния экологических факторов на дыхание почвы. Другими словами, в большинстве работ не использован инструментарий, позволяющий интерпретировать результат моделирования.

Цель данной работы — не столько построить очередную модель эмиссии CO_2 для конкретной ситуации, сколько продемонстрировать возможности алгоритма RF для оценки влияния экологических факторов на интенсивность дыхания почвы. Были подобраны контрастные по характеру растительности варианты типичных для южной тайги лесных и луговых биотопов в пределах относительно небольшой территории, что позволило исключить влияние различий в климате, неизбежное в макромасштабе. Основная идея работы заключалась в анализе роли двух групп факторов — более переменных либо в пространстве, либо во времени. Оценка их вклада в изменчивость дыхания почвы выполнена двумя путями: сопоставлением важности отдельных предикторов и сравнением качества моделей с разными их комбинациями.

МАТЕРИАЛ И МЕТОДЫ

Район исследований расположен в подзоне южной тайги, в 30 км юго-восточнее г. Екатеринбург. Работы выполнены в разных вариантах лесных и луговых биотопов (табл. 1). Выбрано 10 участков — 7 лесных и 3 луговых (рис. 1). Участки расположены в границах треугольника с размером стороны около 1 км (координаты вершин: 56.6072 с.ш., 61.0480 в.д.; 56.6072 с.ш., 61.0682 в.д.; 56.5998 с.ш., 61.0602 в.д.). Расстояние между соседними участками составляло 30–150 м. На каждом участке заложили по 3 пробные площади (всего 30); в сосняках (кроме двух вариантов) их размер равен 10×10 м, в кисличном и мертвопокровном сосняках — 2×4 м, на лугах — 5×5 м.

Участки в сосновых лесах заложены на сходных по строению почвах (за исключением мертвопокровного и молодого), но значительно различаются по разнообразию и обилию травяно-кустарничкового яруса — от мертвопокровных до разнотравных (см. табл. 1). Кроме того, кипрейный сосняк имеет признаки недавнего низового пожара, а почвы молодого сосняка, формирующегося на старопашотных землях, механически нарушены и лишены напочвенного растительного покрова. Все три луговых участка заложены в различных эдафических условиях: от суходольного скашиваемого луга (M1) до переувлажненного лабазникового. Все почвы имели слабо кислую реакцию (pH 4.7–6.0).

Таблица 1. Характеристика участков

№	Биотоп	Код	Доминанты древесного и травяно-кустарничкового яруса	Тип почвы по WRB	Тип почвы КиДПП
Сосновые леса					
1	Сосняк черничный	Sch	<i>Pinus sylvestris</i> , в подросте единичные <i>Betula</i> spp. и <i>Sorbus aucuparia</i> ; <i>Vaccinium myrtillus</i> , злаки	Albic Retisols (Differential)	Дерново-подзолистая типичная
2	Сосняк орляковый	Spar	<i>P. sylvestris</i> , в подросте единичные <i>S. aucuparia</i> ; <i>Pteridium aquilinum</i> , <i>Calamagrostis arundinaceae</i> , <i>V. myrtillus</i>	Albic Retisols (Differential) и Pretic Luvisols (Siltic)	Дерново-подзолистая, типичная и бурозем оподзоленный
3	Сосняк разнотравный (кипрейный)	SDG	<i>P. sylvestris</i> , в подросте единичные <i>Alnus</i> spp.; <i>Chamaenerion angustifolium</i>	Albic Luvisols (Loamic)	Бурозем оподзоленный
4	Сосняк злаковый	SZL	<i>P. sylvestris</i> , в подросте единичные <i>Betula</i> spp.; <i>C. arundinaceae</i> , <i>Brachypodium pinnatum</i>	Albic Retisols (Clayic)	Дерново-подзолистая оподзоленная
5	Сосняк, формирующийся на старопашотных землях	SSG	Подрост <i>P. sylvestris</i> (10–15 лет)	Skeletal Combisols (Densic, Turbic)	Бурозем в комплексе с турбоземом, турбированный
6	Сосняк мертвopoкровный	SDC	<i>P. sylvestris</i>	Eutric Retisols (Densic)	Дерново-подзолистая, типичная
7	Сосняк кисличный	SA	<i>P. sylvestris</i> , в подросте единичные <i>S. aucuparia</i> ; <i>Oxalis acetosella</i> , <i>Rubus saxatilis</i>	Albic Luvisols (Loamic)	Бурозем оподзоленный
Луга					
8	Суходольный луг разнотравный	M1	<i>Poa</i> spp., <i>Geum rivale</i> , <i>Arctium tomentosum</i> , <i>Trifolium</i> spp.	Haplic Luvisols (Densic)	Бурозем оподзоленный
9	Пойменный луг разнотравный	M2	<i>Carex</i> spp., <i>Aegopodium podagraria</i> , <i>Vicia cracca</i>	Eutric Fluvisols (Siltic)	Аллювиальная серогумусовая, типичная
10	Пойменный луг лабазниковый	M3	<i>Filipendula ulmaris</i>	Gleyic Phaeozems (Clayic)	Аллювиальная серогумусовая, глееватая

Измерение эмиссии CO₂. Скорость потока CO₂ с поверхности почвы измеряли по стандартному варианту закрытого динамического камерного метода (Closed dynamic chamber method) [32] полевым респирометром Li-8100A (Li-Cor biosciences, США) с использованием заранее установленных в почву на глубину 3 см полипропиленовых колец с внутренним диаметром 105 мм. На каждой пробной площади измерения проводили в 10 случайно выбранных постоянных точках (по 5 – в мертвopoкровном и кисличном сосняках).

Работы выполнены в вегетационные сезоны 2012–2016 гг. (с мая по октябрь), всего 21 тур измерений: 2012 г. – 23–24 августа и 1–2 октября; 2013 г. – 5–6 мая, 28–29 мая, 20–21 июня, 24–25 июля, 27–28 августа, 24–25 сентября и 22–23 октября; 2014 г. – 10–11 мая, 27–28 мая, 26–27 июня, 30–31 июля, 3–4 сентября и 1–2 октября; 2015 г. – 27–28 мая и 29–30 июня; 2016 г. – 30–31 мая, 5–6 июля, 2–3 августа и 5–6 октября. Измерения проводили в

светлое время суток в период с 10 до 16 ч. Специальные исследования показали, что в этот промежуток дыхание значимо не отличается от среднесуточных показателей, поэтому можно сравнивать результаты, полученные в разное время на разных участках [33]. Всего выполнено 5670 измерений.

Анализ данных. Статистическую обработку проводили в среде R v. 3.6.2. Статистической единицей во всех случаях считали пробную площадь, т.е. среднее значение по 10 (или 5) измерениям.

Для моделирования использовали предикторы двух типов – определенные непосредственно на месте измерений и данные дистанционного зондирования Земли (табл. 2). Температуру почвы (с точностью 0.1°C) и объемную влажность почвы (с точностью 0.1 об. %) измеряли датчиками, соединенными с блоком управления респирометра: термометром Omega 88311E (OMEGA Engineer-

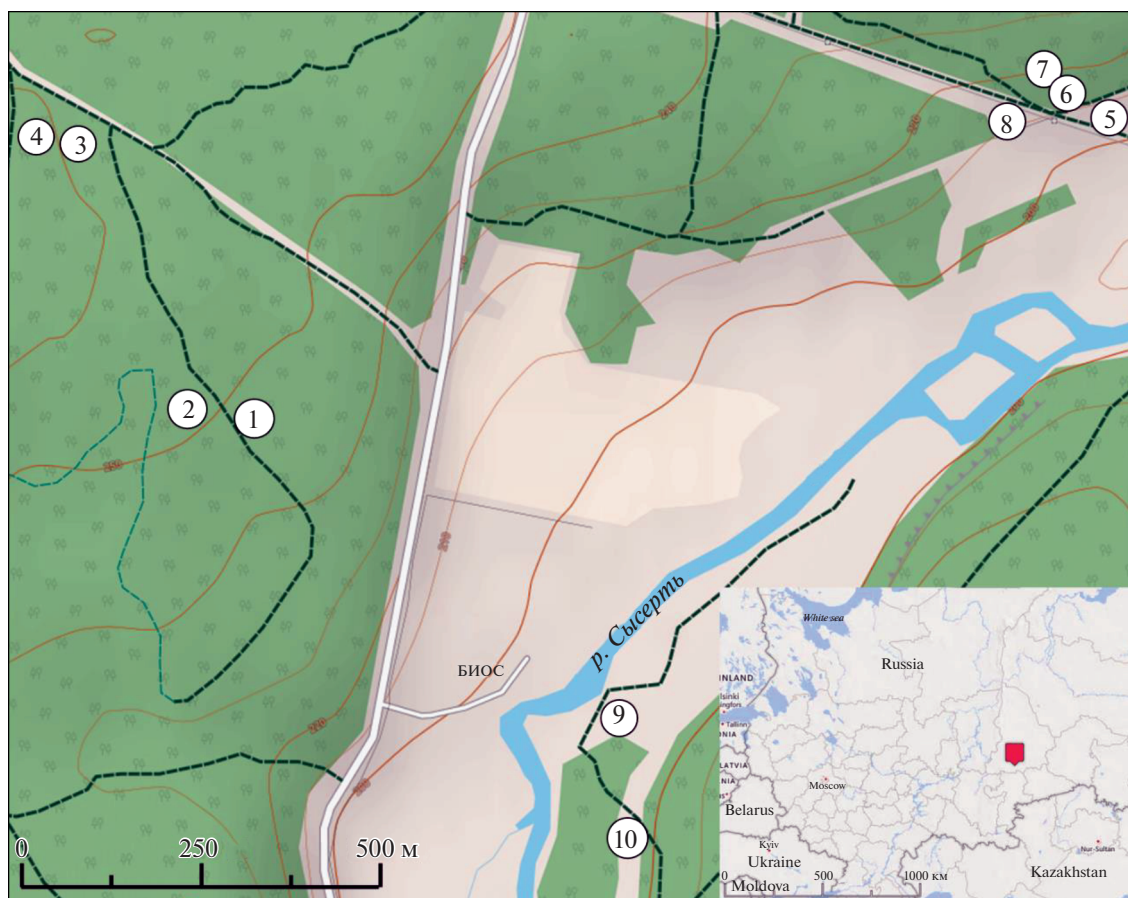


Рис. 1. Схема расположения участков.

Сосняки: 1 – черничный (Sch), 2 – орляковый (SP), 3 – разнотравный (SDG), 4 – злаковый (SZ), 5 – самозарастания (SSG), 6 – мертвопокровный (SDC), 7 – кисличный (SA); луга: 8 – суходольный разнотравный (M1), 9 – пойменный разнотравный (M2), 10 – пойменный лабазниковый (M3). Карта взята с www.openstreetmap.org.

ing, Великобритания) и датчиком влажности ThetaProbe ML2 (Delta-T devices, Великобритания); температуру воздуха – термодатчиком, встроенным в камеру респирометра.

Индексы NDVI, LAI и FPAR были определены по 8-дневным композитным снимкам среднего разрешения спектрометра MODIS со спутников Terra и Aqua, предоставленных сервисом BEGA-Science [34] (см. табл. 2). Эти параметры часто используют для моделирования дыхания почвы, продукции и газообмена растительности [35–39].

Индекс сухости SPEI (Standardized Precipitation Evapotranspiration index) [40] получен из глобальной базы (<http://spei.csic.es>) [41]. Он показывает, насколько засушливым был изучаемый период относительно нормы за несколько последних десятилетий (в данной базе – с 1950 г.). Величина индекса в интервале от 0.99 до – 0.99 свидетельствует о нормальном увлажнении, от 1.00 до 1.49 – умеренном переувлажнении, более 1.50 – сильном переувлажнении, от –1 до –1.49 – умеренной засухе, менее –1.50 – сильной засухе.

Для отнесения предиктора к группе пространственных или временных факторов использовали разложение дисперсии на компоненты в двухфакторном дисперсионном анализе (пакет VCA [42]). Предиктор относили к группе пространственных факторов, если больший вклад в дисперсию вносил тип биотопа, временных – дата измерения.

Для реализации алгоритма RF весь массив данных разделили на обучающую (80%) и тестовую (20%) выборки (пакет caret [43]). Регрессию строили в пакете randomForest [44] со следующими параметрами: количество деревьев – 500, максимальное количество признаков, используемых для разделения, – 3–5. При обучении модели использовали четыре варианта: 1) полный набор предикторов; 2) только пространственные предикторы; 3) только временные предикторы; 4) “стандартный набор”, используемый в большинстве регрессионных моделей: температура почвы, температура воздуха, влажность почвы. Качество модели оценивали на тестовой выборке

Таблица 2. Список используемых предикторов

Предиктор	Описание	Примечание
T_{soil}	Температура почвы на глубине 5 см	Измерены в непосредственной близости от точки измерения
T_{air}	Температура воздуха	
Hum	Объемная влажность почвы на глубине 5 см	См. табл. 1
Soil type	Тип почвы по WRB	
Vegetation	Характер растительности	Лес/луг
Biotop	Тип биотопа	Участок – см. табл. 1
NDVI	Нормализованный дифференциальный вегетационный индекс, обычно хорошо коррелирует с общей надземной фитомассой	Продукт MYD09Q1, разрешение 250 м
LAI	Индекс листовой поверхности (площадь листьев, приходящаяся на единицу поверхности почвы)	Продукт MCD15A2, разрешение 500 м
FPAR	Доля поглощенной фотосинтетически активной радиации (доля солнечного излучения, теоретически доступного для фотосинтеза)	
SPEI-1	Индекс сухости за один месяц, предшествующий измерениям	Из базы SPEI Global Drought Monitor, пространственное разрешение 0.5°
SPEI-12	Индекс сухости за 12 месяцев, предшествующих измерениям	

по коэффициенту детерминации R^2 и среднему квадрату отклонений расчетных значений от измеренных (Mean of squared errors, MSE).

Для оценки вклада временных и пространственных факторов определяли важность предикторов (Features importance) по двум показателям: $\%IncMSE$ (увеличение MSE в ходе случайных перестановок каждой переменной) и $IncNodePurity$ (мера среднего увеличения “чистоты узла” дерева; вычисляется на основе MSE , определенной перед каждым разбиением узлов в каждом дереве). Для оценки важности обычно используют только $\%IncMSE$. Однако для моделей с полным набором предикторов мы использовали также и $IncNodePurity$, поскольку эта характеристика обладает свойством аддитивности, соответственно позволяет оценить относительный вклад переменных. Абсолютные величины важности для удобства выражали в процентах. Эти показатели не тождественны друг другу, но различаются не сильно (не более единиц процентов).

Полученные модели и код размещены по адресу: github.com/IASmorkalov/Respiration_RF_2022.

РЕЗУЛЬТАТЫ

Дыхание почвы за все время лежало в диапазоне 0.5–9.9 мкмоль $\text{CO}_2/\text{м}^2 \text{ с}$ (рис. 2а). Максимальные значения наблюдались в летние месяцы, минимальные – в начале и после завершения вегетационного сезона. Максимальная разница средних значений между типами биотопов составила 1.9 раза. Дыхание почвы определялось главным образом временной изменчивостью: дата измерений объясняла 60.9% дисперсии. Влияние простран-

ственной изменчивости проявлялось в заметно меньшей степени – 10.4% дисперсии. Большинство предикторов можно отнести к группе временных факторов (вклад даты измерений в дисперсию более 75%): температура почвы, температура воздуха, SPEI, NDVI, LAI, FPAR (рис. 3). Влажность почвы была отнесена к группе пространственных факторов (49.7% объясненной дисперсии), хотя существенная часть ее дисперсии была связана со временем измерений (36.8%). Характер растительности, тип биотопа, тип почвы – это полностью пространственные факторы, не менявшиеся за время исследований.

Рассмотренные факторы мало различались между биотопами, за исключением влажности почвы (рис. 2): самым влажным был лабазниковый луг, самым сухим – сосняк мертвопокровный (см. рис. 2г), разница между ними составила 3.3 раза.

Во всех случаях лучшая аппроксимация на тестовых выборках достигалась при использовании всего набора предикторов (рис. 4–6). Временные факторы для всего набора данных и отдельно для сосняков составляли 76–77% важности (по $IncNodePurity$), для лугов – 91%. Важность “стандартного набора” предикторов (температура почвы, температура воздуха, влажность почвы) составляла 37, 26 и 52% для всего набора данных, сосняков и лугов соответственно.

В разных вариантах моделей порядок отдельных предикторов по важности различался. Так, для всего набора данных наибольшая важность была у температуры почвы (рис. 4). Из пространственных факторов наиболее важным был тип биотопа, наименее – характер растительности. При ис-

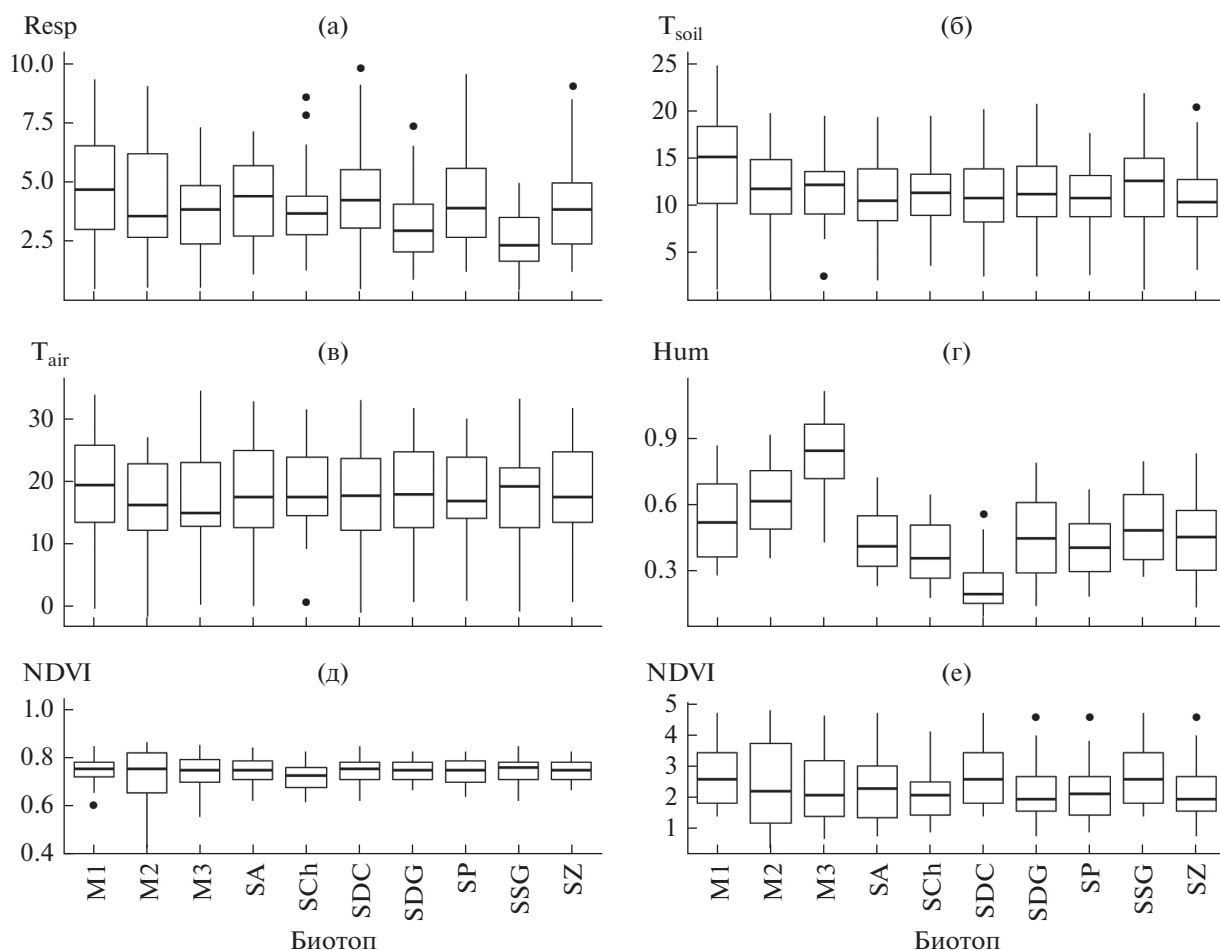


Рис. 2. Значения параметров на участках за весь период наблюдений: а – дыхание почвы (мкмоль $\text{CO}_2/\text{м}^2 \text{с}$), б – температура почвы ($^{\circ}\text{C}$), в – температура воздуха ($^{\circ}\text{C}$), г – объемная влажность почвы, $\text{м}^3/\text{м}^3$, д – NDVI, е – LAI. По горизонтальной оси – тип биотопа (обозначения – см. на рис. 1). Горизонтальная линия – медиана, прямоугольник – 25 и 75% квантили, планки – $1.5 \times$ межквартильный размах, точки – выбросы.

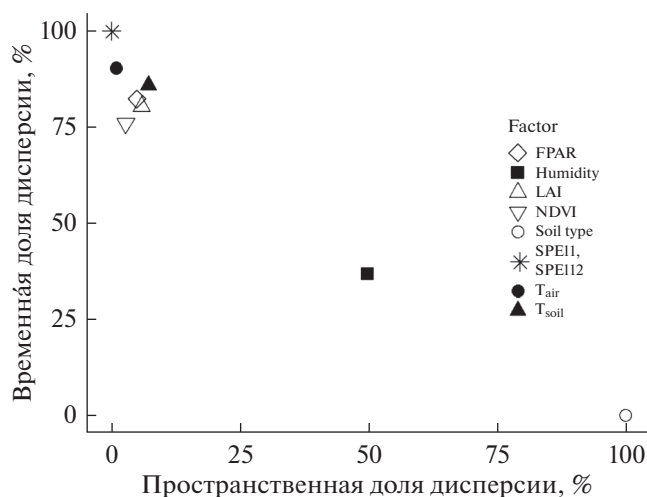


Рис. 3. Вклад пространственной и временной составляющих в дисперсию предикторов. Обозначение предикторов – см. табл. 2.

пользовании усеченного набора факторов коэффициент детерминации был меньше, а *MSE* – больше. Самое низкое качество аппроксимации было при использовании только временных предикторов.

В сосняках самыми важными были NDVI и температура почвы (рис. 5). Важность температуры воздуха была значительно ниже, чем температуры почвы. Даже при использовании всех предикторов модель для сосняков была менее точной, чем для всего массива данных.

Лучшая аппроксимация была достигнута для лугов (рис. 6), причем результаты при использовании всех предикторов и только временных почти не различались, а важность температуры воздуха была больше, чем температуры почвы. При использовании только пространственных факторов для лугов самым важным оказалась влажность почвы.

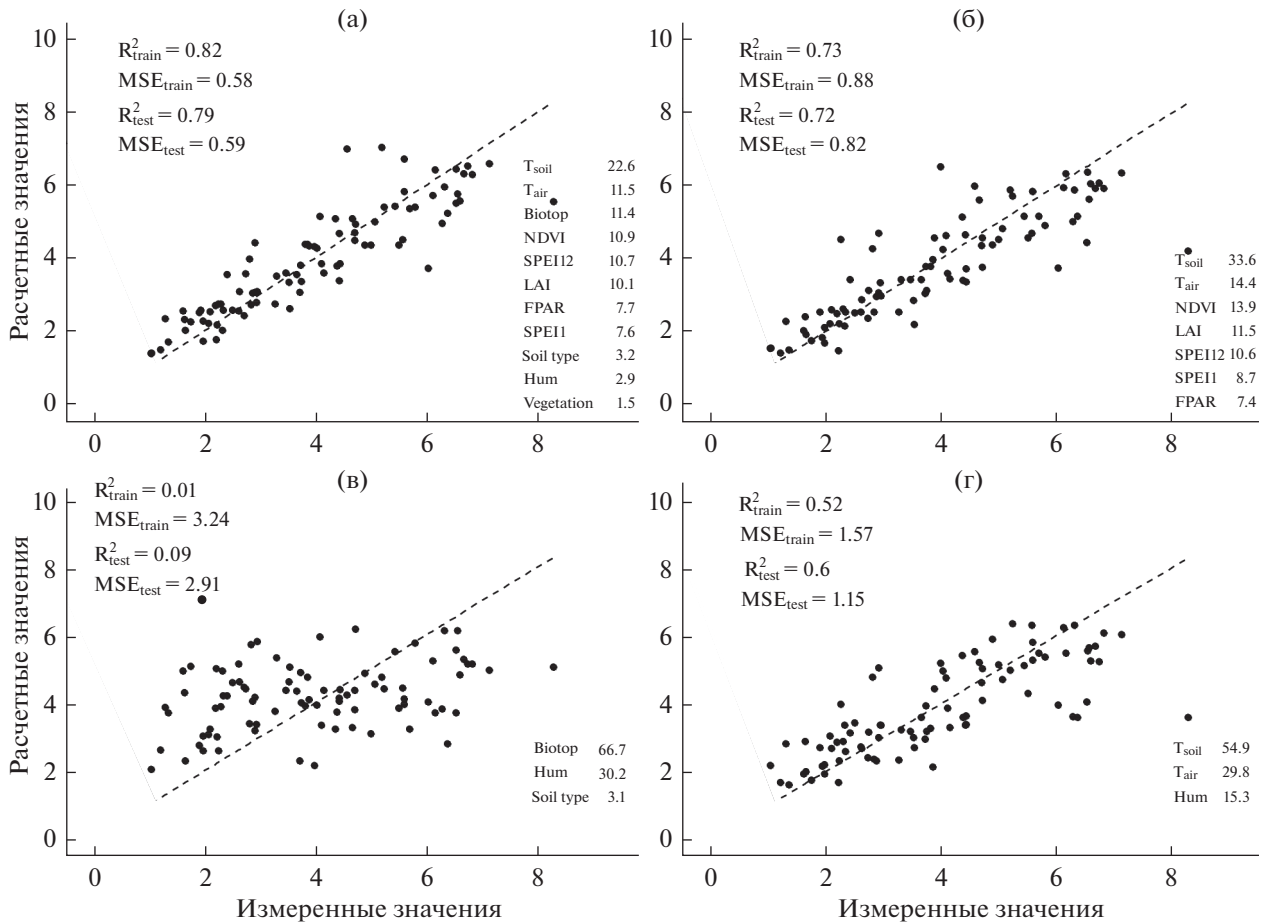


Рис. 4. Результаты аппроксимации данных по дыханию почвы (мкмоль CO₂/м² с) во всех биотопах алгоритмом RF. Здесь и на рис. 5, 6: а – все предикторы, б – временные предикторы, в – пространственные предикторы, г – температура и влажность. Приведены R² и MSE для обучающей (train) и тестовой (test) выборки, важность %IncMSE (в %). Пунктирная линия – линия идеального прогноза.

При использовании “стандартного набора” предикторов во всех случаях наиболее важной была температуры почвы, наименее – ее влажность (рис. 4–6г).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Полученные абсолютные величины дыхания почвы в разгар вегетационного сезона в сосняках (1.3–10.7 мкмоль CO₂/м² с) близки к значениям, которые обычно регистрируют для лесов умеренных широт: до 4.3 ± 0.75 [45], 1.9–8.8 [46], 2.8–6.7 [47], 6.3–9.5 [48], 3.9–6.9 [49], 4.4–11.4 [50], 3.5–4.4 [51]. На лугах диапазон значений был несколько уже (1.5–9.9 мкмоль CO₂/м² с), чем в сосняках, но шире, чем в других исследованиях травяных экосистем: 2.5–8.8 [52], 3.2–5.1 [53], 3.2–5.1 [54], 5.1–6.3 [55]. Сезонная динамика с максимумом эмиссии CO₂ в летние месяцы сходна с динамикой, описанной для лесов южной тайги [52], се-

верной тайги в Восточной Сибири [51] и Центральной Сибири [56].

Характер растительности влияет на дыхание, но не всегда однозначно. Так, при анализе базы “Дыхание почвы” почти в половине случаев тип растительности значимо влиял на величину потока CO₂ [57], в частности различалось дыхание почвы луговых и лесных ценозов [58]. По данным разных авторов [52, 59], при одинаковом типе почвы луга отличаются от лесов большей интенсивностью дыхания, а также отмечается значимое влияние типа почвы на дыхание [60]. Однако показано [61], что большое влияние на поток углекислого газа из почвы имеет и продолжительность вегетационного периода. В нашей работе влияние типа биотопа было максимальным при рассмотрении всего массива данных или только сосняков. Раннее весеннее прогревание и осеннее остывание лугов, видимо, нивелировало разницу между средним дыханием почвы на лугах и в сосняках за весь вегетационный период, т.е. в це-

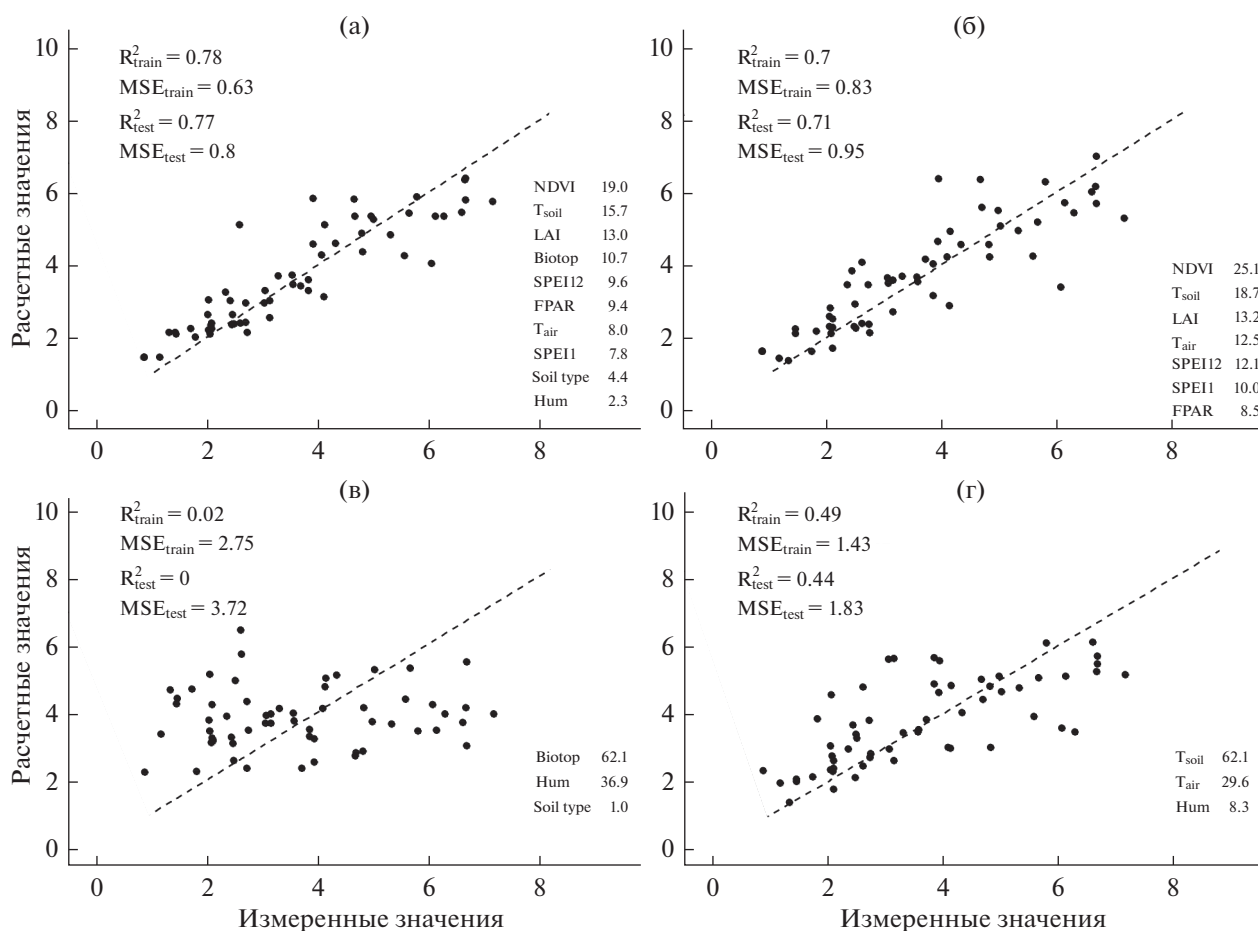


Рис. 5. Результаты аппроксимации данных по дыханию почвы в сосняках ($\text{мкмоль CO}_2/\text{м}^2 \text{ с}$) алгоритмом RF.

лом наши результаты не противоречат литературным данным: существует разница между разными биотопами, но ведущую роль играют климатические факторы.

Построенные нами модели показали высокую точность: R^2 достигал 0.88 ($MSE = 0.47$) для лугов и 0.79 ($MSE = 0.59$) для всех биотопов. Это выше точности моделей на основе алгоритма RF для горных лесов в южных Скалистых горах ($R^2 = 0.44$, $MSE = 0.8$) [25] и сопоставимо с моделями для плантаций сахарного тростника ($R^2 = 0.8$) [62], в целом лесов северного полушария (R^2 до 0.86, $MSE = 2.16$) [28] и дыхания в глобальном масштабе ($R^2 = 0.89$) [27].

Полученные значения коэффициента детерминации выше, чем у большинства классических регрессионных моделей. Например, даже при использовании отдельного уравнения для каждой конкретной локации R^2 для дыхания почв г. Курска не превышал 0.6 [3]. Более высокая точность классических регрессионных моделей достигается в экстремальных условиях: например, в горных тундрах R^2 достигал 0.9 [63], на трансекте с боль-

шим градиентом влажности в горах — 0.8 [64]. В более сложных (имитационных) моделях R^2 составлял 0.34 и 0.77 для сухих и влажных лет соответственно [1]. Точность моделей на основе нейросетей в целом не превышает точность классической регрессии: R^2 колеблется от 0.3–0.4 в локальном масштабе [18] до 0.6 в глобальном [19]. Другими словами, случайный лес в отличие от методов традиционной параметрической регрессии и нейросетей позволяет хорошо интерполировать данные даже при отсутствии сильных средовых градиентов. Строго говоря, R^2 — не лучший вариант для сравнения эффективности разных алгоритмов. Но, к сожалению, в публикациях редко приводятся другие способы оценки качества моделей, поэтому была использована именно эта метрика.

Чтобы сравнить точность параметрических моделей с точностью случайного леса, мы провели также моделирование средствами множественной регрессии для всего массива данных. Для тестовой выборки мы получили следующие оценки модели: $R^2 = 0.55$, $MSE = 1.3$ для всего на-

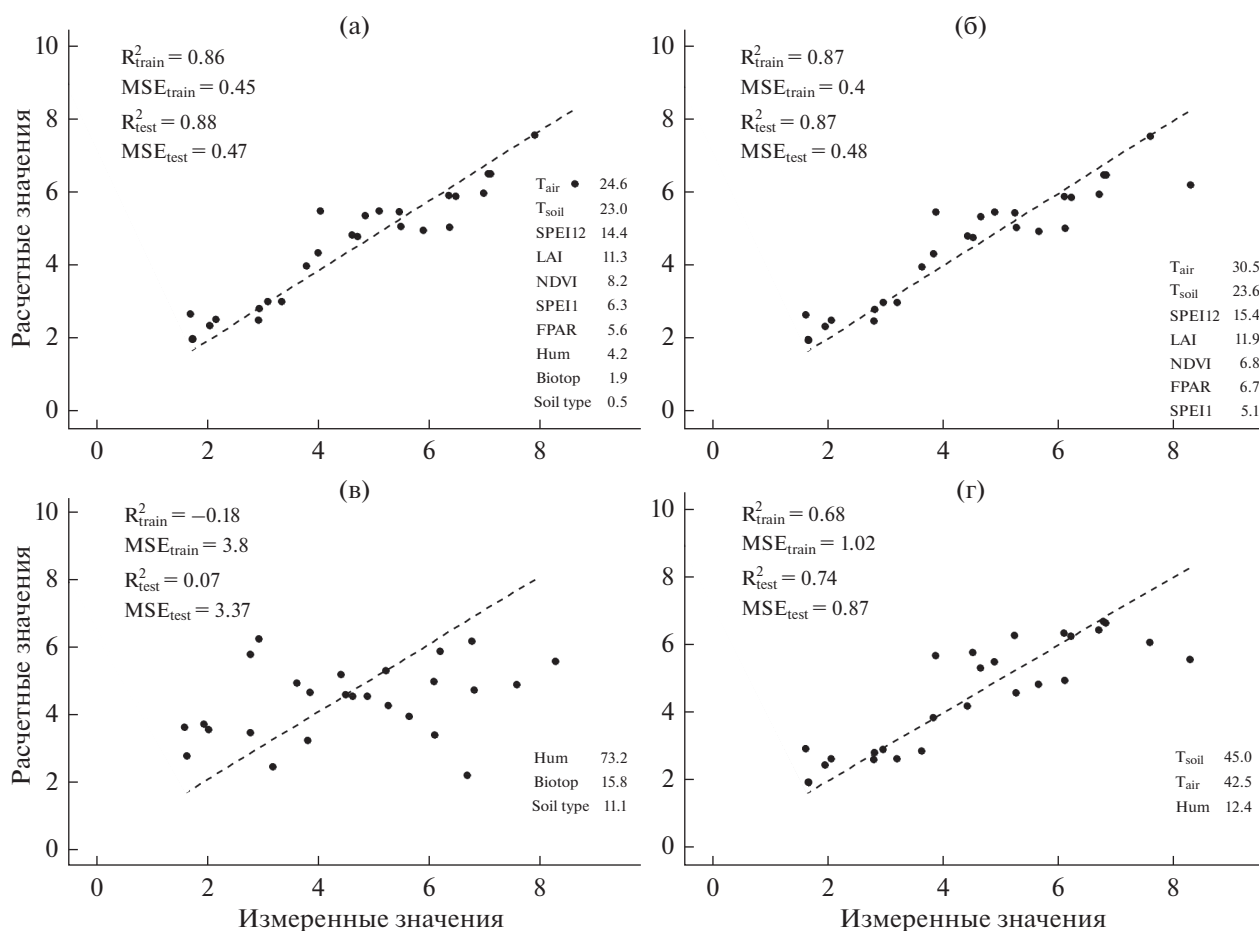


Рис. 6. Результаты аппроксимации данных по дыханию почвы на лугах (мкмоль CO₂/м² с) алгоритмом RF.

бора предикторов; $R^2 = 0.25$, $MSE = 2.21$ – при использовании температуры почвы, воздуха и влажности почвы. При реализации пошаговой регрессии значимыми предикторами оказались только температура почвы, NDVI и SPEI1, а точность модели была ниже, чем при использовании всех предикторов ($R^2 = 0.33$, $MSE = 1.92$). Таким образом, использование случайного леса для одних и тех же данных позволило значительно лучше описать их по сравнению с классическими регрессионными моделями. По эффективности RF может уступать параметрическим регрессионным моделям и даже отдельному дереву решений при незначительном количестве данных, но если их достаточно, то он показывает результаты лучше [30].

Сходная с нашей схема разделения факторов на пространственные и временные описана в работе [12], в которой критерием служило сравнение средней дисперсии по различным датам с дисперсией между точками. Результаты обеих работ оказались близки, за исключением того, что в цитируемой публикации, помимо температуры почвы и воздуха, к временным была отнесена и объ-

емная влажность почвы, которая в данной работе оказалась в группе пространственных факторов.

По результатам обработки алгоритмом RF наиболее важны были временные переменные, что более характерно для масштаба единиц–десятков метров. Возможно, это связано с тем, что пространственное разрешение данных дистанционного зондирования было сопоставимо с размерами рассматриваемой территории, поэтому пространственный разброс этих показателей не был значительным. Однако влажность почвы очень сильно различалась между участками, но не оказала сильного влияния на интенсивность дыхания почвы, хотя данный параметр традиционно считают основным детерминантом эмиссии [32].

Пространственные факторы оказывают значимое влияние на дыхание почвы в региональном или глобальном масштабах, например тип биомы [19, 65]. Учет пространственных факторов существенно улучшает точность моделей и в меньшем масштабе, но при наличии сильных средовых градиентов. Например, включение в модели влажности почвы на крутых горных трансектах в масшта-

бе десятков метров повышало R^2 с 0.5 до 0.8 [64], а включение информации об уровне загрязнения почв в масштабе десятков километров – с 0.2 до 0.74 [66].

Наиболее важной из рассмотренных пространственных переменных оказался тип биотопа, хотя это справедливо только для всего массива данных и сосняков, но не для лугов. Следует подчеркнуть, что тип биотопа – это комплексный предиктор, включающий в себя все учтенные и неучтенные особенности участков. Поэтому даже в масштабе сотен метров при создании моделей почвенной эмиссии CO_2 следует учитывать не только временные, но и пространственные факторы. На лугах важность всех пространственных факторов меньше, чем временных, даже при учете значительной разницы по влажности почвы. В ходе предварительного анализа в модель включали также химические показатели почв (рН, содержание органического углерода, обменного Са и Mg, подвижного Fe), но это не увеличивало точность, поэтому в окончательный анализ эти переменные не вошли.

Данные дистанционного зондирования значительно повышают точность модели, что очевидно при сравнении со “стандартным набором” предикторов (рис. 4–6б, г). Аналогично они улучшали модели и в региональном масштабе [39].

Качество моделей может снижаться из-за “эффекта памяти” почв от предыдущих воздействий в разных временных масштабах [67]. В рекуррентных нейросетях для уменьшения этого эффекта можно реализовать “кратковременную память” [68]. При других способах моделирования необходимо вводить переменные, которые характеризуют предыдущие состояния системы. В нашей работе такой переменной был SPEI, который характеризует условия увлажнения, предшествовавшие измерениям. Оказалось, что условия предыдущих 12 мес. важнее, чем месяца, непосредственно предшествующего измерению. Это отчасти соответствует обнаруженному влиянию суммы осадков за период весна–лето на суммарную годовую эмиссию CO_2 [69].

В целом результаты оценки важности параметров соответствуют ожидаемым: ведущую роль играет температура, что было многократно продемонстрировано ранее [32]. Включение пространственных предикторов улучшало качество модели, особенно в лесах, что подтверждает необходимость учитывать не только временную, но и пространственную составляющие изменчивости дыхания почвы [12]. Большие значения важности NDVI подтвердили также необходимость использовать характеристики растительности в моделях потоков углекислого газа из почвы. Следует уточнить, что сделанные выводы распространяются

только на вегетационный сезон, так как в зимнее время исследования не проводились.

ЗАКЛЮЧЕНИЕ

Несмотря на значительное биотопическое разнообразие, включавшее две группы биотопов (леса и луга), скорость эмиссии CO_2 из почвы сильнее всего зависела от временных факторов. Для лугов точность модели была идентичной как при использовании только временных, так и при добавлении пространственных факторов, несмотря на большую разницу в условиях увлажнения. Тем не менее в сосняках один из пространственных факторов (тип биотопа) имел большое значение. Это показывает, что даже для масштаба сотен метров модели дыхания почвы должны включать не только временные, но и пространственные факторы.

Оценка важности переменных при использовании алгоритма RF позволяет содержательно интерпретировать результаты моделирования. Возможность одновременного использования переменных разных типов, в том числе сильно скоррелированных между собой, делает алгоритм RF более гибким по сравнению с классическими регрессионными методами. Но не только это позволяет рекомендовать алгоритм RF к широкому использованию в моделировании потоков углекислого газа из почвы: несомненен его потенциал в оценке вклада разных факторов в вариабельность дыхания почвы, что позволяет лучше понять механизмы, лежащие в основе цикла углерода в наземных экосистемах.

Полевые работы были выполнены при поддержке УрО РАН (проект № 12-П-4-1057); при поддержке Министерства науки и высшего образования РФ выполнены анализ данных (проект № 122021000076-9) и подготовка текста статьи (проект № FEUZ-2021-0014). Автор благодарит Д.А. Бедина за консультации по вопросам машинного обучения, Е.Л. Воробейчика – за ценные замечания и большую работу по улучшению качества рукописи, С.Ю. Кайгородову и Т.Ю. Габерштейн – за диагностику и анализ почв, Н.О. Садыкову – за обсуждение и комментарии к тексту.

Автор заявляет об отсутствии конфликта интересов и подтверждает, что настоящая работа не содержит исследований с участием людей или животных в качестве объектов исследования.

СПИСОК ЛИТЕРАТУРЫ

1. Припутина И.В., Быховец С.С., Фролов П.В. и др. Применение математических моделей ROMUL и Romul_Num для оценки эмиссии CO_2 и динамики органического вещества в серой лесной почве под лиственным лесом в Южном Подмосковье // Почвоведение. 2020. № 10. С. 1262–1275.

2. *Reiners W.A.* Carbon dioxide evolution from the floor of three minnesota forests // *Ecology*. 1968. V. 49. № 3. P. 471–483.
<https://doi.org/10.2307/1934114>
3. *Неведров Н.П., Саржанов Д.А., Проценко Е.П., Васенев И.И.* Сезонная динамика эмиссии CO₂ из почв города Курска // *Почвоведение*. 2021. № 1. С. 70–79.
4. *Chimner R.A.* Soil respiration rates of tropical peatlands in Micronesia and Hawaii // *Wetlands*. 2004. V. 24. № 1. P. 51.
[https://doi.org/10.1672/0277-5212\(2004\)024\[0051:SRROTP\]2.0.CO;2](https://doi.org/10.1672/0277-5212(2004)024[0051:SRROTP]2.0.CO;2)
5. *Jia B., Zhou G., Wang F., Wang Y.* A comparative study on soil respiration between grazing and fenced typical *Leymus chinensis* steppe, Inner Mongolia // *Chinese Journal of Applied Ecology*. 2004. V. 15. № 9. P. 1611–1615.
6. *Orchard V.A., Cook F.J.* Relationship between soil respiration and soil moisture // *Soil Biology and Biochemistry*. 1983. V. 15. № 4. P. 447–453.
[https://doi.org/10.1016/0038-0717\(83\)90010-X](https://doi.org/10.1016/0038-0717(83)90010-X)
7. *Rubio V.E., Detto M.* Spatiotemporal variability of soil respiration in a seasonal tropical forest // *Ecology and Evolution*. 2017. V. 7. № 17. P. 7104–7116.
<https://doi.org/10.1002/ece3.3267>
8. *Webster K.L., Creed I.F., Skowronski M.D., Kaheil Y.H.* Comparison of the performance of statistical models that predict soil respiration from forests // *Soil Science Society of America Journal*. 2009. V. 73. № 4. P. 1157–1167.
<https://doi.org/10.2136/sssaj2008.0310>
9. *Sotta E.D., Meir P., Malhi Y., Donato A.* et al. Soil CO₂ efflux in a tropical forest in the central Amazon // *Global Change Biology*. 2004. V. 10. № 5. P. 601–617.
<https://doi.org/10.1111/j.1529-8817.2003.00761.x>
10. *Tufekcioglu A., Raich J.W., Isenhardt T.M., Schultz R.C.* Soil respiration within riparian buffers and adjacent crop fields // *Plant and Soil*. 2001. V. 229. № 1. P. 117–124.
<https://doi.org/10.1023/A:1004818422908>
11. *Lloyd J., Taylor J.A.* On the temperature dependence of soil respiration // *Functional Ecology*. 1994. V. 8. № 3. P. 315–323.
<https://doi.org/10.2307/2389824>
12. *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.
13. *Карелин Д.В., Замолодчиков Д.Г., Исаев А.С.* Малоизвестные импульсные составляющие почвенной эмиссии диоксида углерода в таёжных лесах // *Докл. РАН*. 2017. Т. 75. № 4. С. 473–476.
14. *Zeng X.H., Song Y.G., Zeng C.M.* et al. Partitioning soil respiration in two typical forests in semi-arid regions, North China // *Catena*. 2016. V. 147. P. 536–544.
<https://doi.org/10.1016/j.catena.2016.08.009>
15. *Kuzakov Y., Gavrichkova O.* Time lag between photosynthesis and carbon dioxide efflux from soil: A review of mechanisms and controls // *Global Change Biology*. 2010. V. 16. № 12. P. 3386–3406.
<https://doi.org/10.1029/2005JG000063>
16. *Richardson A.D., Braswell B.H., Hollinger D.Y.* et al. Comparing simple respiration models for eddy flux and dynamic chamber data // *Agricultural and Forest Meteorology*. 2006. V. 141. № 2–4. P. 219–234.
<https://doi.org/10.1016/j.agrformet.2006.10.010>
17. *Farhate C.V.V., Souza Z.M., Oliveira S.R.M.* et al. Use of data mining techniques to classify soil CO₂ emission induced by crop management in sugarcane field // *PLoS ONE*. 2018. V. 13. № 3. P. e0193537.
<https://doi.org/10.1371/journal.pone.0193537>
18. *Jarvi L., Nordbo A., Junninen H.* et al. Seasonal and annual variation of carbon dioxide surface fluxes in Helsinki, Finland, in 2006–2010 // *Atmospheric Chemistry and Physics*. 2012. V. 12. № 18. P. 8475–8489.
<https://doi.org/10.5194/acp-12-8475-2012>
19. *Zhao Z.Y., Peng C.H., Yang Q.* et al. Model prediction of biome-specific global soil respiration from 1960 to 2012 // *Earth Future*. 2017. V. 5. № 7. P. 715–729.
<https://doi.org/10.1002/2016ef000480>
20. *Song X.Z., Peng C.H., Zhao Z.Y.* et al. Quantification of soil respiration in forest ecosystems across China // *Atmos. Environ.* 2014. V. 94. P. 546–551.
<https://doi.org/10.1016/j.atmosenv.2014.05.071>
21. *Rudin C.* Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead // *Nature Machine Intelligence*. 2019. V. 1. № 5. P. 206–215.
<https://doi.org/10.1038/s42256-019-0048-x>
22. *Чистяков С.П.* Случайные леса: обзор // *Труды Карельского научного центра РАН*. 2013. № 1. С. 117–136.
23. *Hastie T., Tibshirani R., Friedman J.* *Random Forests // The elements of statistical learning: data mining, inference, and prediction*. NY: Springer New York, 2009. P. 587–604.
24. *Tang X.L., Fan S.H., Du M.Y.* et al. Spatial and temporal patterns of global soil heterotrophic respiration in terrestrial ecosystems // *Earth Syst. Sci. Data*. 2020. V. 12. № 2. P. 1037–1051.
<https://doi.org/10.5194/essd-12-1037-2020>
25. *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>
26. *Jian J.S., Steele M.K., Thomas R.Q.* et al. Constraining estimates of global soil respiration by quantifying sources of variability // *Global Change Biology*. 2018. V. 24. № 9. P. 4143–4159.
<https://doi.org/10.1111/gcb.14301>
27. *Lu H.B., Li S.H., Ma M.N.* et al. Comparing machine learning-derived global estimates of soil respiration and its components with those from terrestrial ecosystem models // *Environmental Research Letters*. 2021. V. 16. № 5. Article 054048. P. 14.
<https://doi.org/10.1088/1748-9326/abf526>
28. *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*. 2022. V. 29. № 11. P. 16123–16137.
<https://doi.org/10.1007/s11356-021-16501-x>

29. Warner D.L., Bond-Lamberty B., Jian J. et al. Spatial predictions and associated uncertainty of annual soil respiration at the global scale // *Global Biogeochemical Cycles*. 2019. V. 33. № 12. P. 1733–1745. <https://doi.org/10.1029/2019gb006264>
30. 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. P. 147314. <https://doi.org/10.1016/j.scitotenv.2021.147314>
31. Jian J.S., Frissell M., Hao D.L. et al. The global contribution of roots to total soil respiration // *Global Ecology and Biogeography*. 2022. <https://doi.org/10.1111/geb.13454>
32. Luo Y., Zhou X. Soil respiration and the environment. Burlington: Acad. Press, 2006. 316 p.
33. Сморкалов И.А. Соотношение пространственной и временной вариабельности дыхания почвы // *Экология: факты, гипотезы, модели: Мат-лы Всероссийской конф. молодых ученых, посвященной 170-летию В.В. Докучаева*. Екатеринбург, 2016. С. 120–123.
34. Лупян 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.
35. Wu J.S., Fu G. Modelling aboveground biomass using MODIS FPAR/LAI data in alpine grasslands of the Northern Tibetan Plateau // *Remote Sensing Letters*. 2018. V. 9. № 2. P. 150–159. <https://doi.org/10.1080/2150704x.2017.1407045>
36. Liu S., Cheng F., Dong S. et al. Spatiotemporal dynamics of grassland aboveground biomass on the Qinghai-Tibet Plateau based on validated MODIS NDVI // *Sci. Rep.* 2017. V. 7. № 1. P. 4182. <https://doi.org/10.1038/s41598-017-04038-4>
37. Wang X.F., Wang H.B., Li X., Ran Y.H. Photosynthesis (NPP, NEP, Respiration) // *Observation and Measurement* / Eds. Li X., Vereecken H. Berlin, Heidelberg, 2017. P. 1–30.
38. Reynolds L.L., Johnson B.R., Pfeifer-Meister L., Bridgham S.D. Soil respiration response to climate change in Pacific Northwest prairies is mediated by a regional Mediterranean climate gradient // *Global Change Biology*. 2015. V. 21. № 1. P. 487–500. <https://doi.org/10.1111/gcb.12732>
39. Ding S.C., Yao X.W., Wang J. et al. Relationships between soil respiration and hyperspectral vegetation indexes and crop characteristics under different warming and straw application modes // *Environmental Science and Pollution Research*. 2021. P. 15. <https://doi.org/10.1007/s11356-021-13612-3>
40. 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>
41. 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 J of Climatology*. 2014. V. 34. № 10. P. 3001–3023. <https://doi.org/10.1002/joc.3887>
42. VCA: Variance Component Analysis. R package version 1.4.2 / URL: <https://CRAN.R-project.org/package=VCA>
43. caret: Classification and Regression Training version 6.0-86 / URL: <https://cran.r-project.org/package=caret>
44. randomForest: Breiman and Cutler’s Random Forests for classification and regression version 4.6-14 / URL: <https://cran.r-project.org/package=randomForest>
45. Осипов А.Ф. Влияние межгодовых различий метеорологических характеристик вегетационного периода на эмиссию CO₂ с поверхности почвы среднетаежного сосняка бруснично-лишайникового (Республика Коми) // *Почвоведение*. 2018. №. 12. С. 1455–1463.
46. Глухова Т.В., Вомперский С.Э., Ковалев А.Г. Эмиссия CO₂ с поверхности олиготрофных болот южно-таежной зоны европейской территории России с учетом микрорельефа // *Почвоведение*. 2014. № 1. С. 48–57.
47. Кадулин М.С., Коцик Г.Н. Эмиссия CO₂ почвами в зоне влияния горно-металлургического комбината “Североникель” в Кольской Субарктике // *Почвоведение*. 2013. № 11. С. 1387–1396. <https://doi.org/10.7868/S0032180X13110063>
48. Машица А.В. Эмиссия диоксида углерода с поверхности подзолистой почвы // *Почвоведение*. 2006. № 12. С. 1457–1463.
49. Сморкалов И.А., Воробейчик Е.Л. Почвенное дыхание лесных экосистем в градиентах загрязнения среды выбросами металлургических заводов // *Экология*. 2011. № 6. С. 429–435.
50. Kozlov M.V., Zvereva E.L., Zverev V.E. Impacts of point polluters on terrestrial biota: Comparative analysis of 18 contaminated areas. Dordrecht: Springer, 2009. 466 p.
51. 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.
52. Лопес де Гереню В.О., Курганова И.Н., Розанова Л.Н., Кудяров В.Н. Годовые потоки диоксида углерода из некоторых почв южно-таежной зоны России // *Почвоведение*. 2001. № 9. С. 1045–1059.
53. Wang M., Liu X.T., Li X.J. et al. Soil respiration dynamics and its controlling factors of typical vegetation communities on meadow steppes in the western Songnen Plain // *Chinese J of Applied Ecology*. 2014. V. 25. № 1. P. 45–52. <https://doi.org/10.1029/2009GB003538>
54. Bahn M., Rodeghiero M., Anderson-Dunn M. et al. Soil respiration in European grasslands in relation to climate and assimilate supply // *Ecosystems*. 2008. V. 11. № 8. P. 1352–1367. <https://doi.org/10.1029/2005JG000063>
55. Perez-Quezada J.F., Saliendra N.Z., Akshalov K. et al. Land use influences carbon fluxes in northern Kazakhstan // *Rangeland Ecology and Management*. 2010. V. 63. № 1. P. 82–93.

56. *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.
57. *Курганова И.Н., Кудеяров В.Н.* Оценка потоков диоксида углерода из почв таежной зоны России // *Почвоведение*. 1998. № 9. С. 1058–1070.
58. *Кудеяров В.Н., Курганова И.Н.* Дыхание почв России: анализ базы данных, многолетний мониторинг, общие оценки // *Почвоведение*. 2005. № 9. С. 1112–1121.
59. *Raich J.W., Tufekcioglu A.* Vegetation and soil respiration: Correlations and controls // *Biogeochemistry*. 2000. V. 48. № 1. P. 71–90.
60. *Курганова И.Н., Лопес де Гереню В.О., Хорошаев Д.А.* и др. Анализ многолетней динамики дыхания почв в лесном и луговом ценозах Приокско-террасного биосферного заповедника в свете современных климатических трендов // *Почвоведение*. 2020. № 10. С. 1220–1236.
61. *Кудеяров В.Н., Хакимов Ф.И., Деева Н.Ф.* и др. Оценка дыхания почв России // *Почвоведение*. 1995. № 1. С. 33–42.
62. *Tavares R.L.M., Oliveira S.R.D., de Barros F.M.M.* et al. Prediction of soil CO₂ flux in sugarcane management systems using the Random Forest approach // *Sci. Agric*. 2018. V. 75. № 4. P. 281–287. <https://doi.org/10.1590/1678-992x-2017-0095>
63. *Magnani M., Baneschi I., Giamberini M.* et al. Drivers of carbon fluxes in Alpine tundra: a comparison of three empirical model approaches // *Science of the Total Environment*. 2020. V. 732. <https://doi.org/10.1016/j.scitotenv.2020.139139>
64. *Lecki N.A., Creed I.F.* Forest soil CO₂ efflux models improved by incorporating topographic controls on carbon content and sorption capacity of soils // *Biogeochemistry*. 2016. V. 129. № 3. P. 307–323. <https://doi.org/10.1007/s10533-016-0233-5>
65. *Huang N., Wang L., Song X.P.* et al. Spatial and temporal variations in global soil respiration and their relationships with climate and land cover // *Sci. Adv*. 2020. V. 6. № 41. <https://doi.org/10.1126/sciadv.abb8508>
66. *Сморкалов И.А., Бедин Д.А.* Учитывание промышленного загрязнения в моделях дыхания почвы // *Математическое моделирование в экологии: Мат-лы 6-й национ. научн. конф. с международным участием. Пушино, 2019. С. 191–193.*
67. *Parolari A.J., Sizemore J., Katul G.G.* Multiscale legacy responses of soil gas concentrations to soil moisture and temperature fluctuations // *J. Geophys. Res.-Biogeosci.* 2021. V. 126. № 2. Article e2020JG005865. P. 16. <https://doi.org/10.1029/2020jg005865>
68. *Besnard S., Carvalhais N., Arain M.A.* et al. Memory effects of climate and vegetation affecting net ecosystem CO₂ fluxes in global forests // *PLoS ONE*. 2019. V. 14. № 2. P. e0211510. <https://doi.org/10.1371/journal.pone.0211510>
69. *Курганова И.Н., Лопес де Гереню В.О., Мякшина Т.Н.* и др. Эмиссия CO₂ из почв различных экосистем южно-таежной зоны: анализ данных непрерывных 12-летних круглогодичных наблюдений // *Докл. РАН*. 2011. Т. 436. № 6. С. 843–846.