

ОЦЕНКА СОДЕРЖАНИЯ ОРГАНИЧЕСКОГО УГЛЕРОДА В ПОЧВАХ РОССИИ С ПОМОЩЬЮ АНСАМБЛЕВОГО МАШИННОГО ОБУЧЕНИЯ

А.В. Чинилин¹, И.Ю. Савин²

^{1,2} Почвенный институт имени В.В. Докучаева, отдел генезиса, географии, классификации и цифровой картографии почв

² Российский университет дружбы народов, Институт экологии, Департамент рационального природопользования

¹Канд. биол. наук; e-mail: andrey.chinilin@yandex.ru

²Д-р с.-х. наук, проф., академик РАН; e-mail: savigory@gmail.com

Исследование показывает современную количественную оценку содержания органического углерода почв России с учетом их огромного разнообразия, а также отражает понимание отдельных факторов, регулирующих и контролирующих содержание органического углерода почв в пределах страны. В работе приводятся результаты трехмерного моделирования содержания органического углерода почв с пространственным разрешением 500 м по ряду стандартных глубин (0–5, 5–15, 15–30, 30–60, 60–100 см) на территории Российской Федерации с помощью ансамблевого машинного обучения. Автоматизированное прогнозирующее картографирование основывалось на данных по 4961 почвенному горизонту из 863 профилей почв, а также на обширном наборе пространственной информации, включающем биоклиматические переменные, цифровую модель рельефа и ее производные, долгосрочные усредненные временные ряды данных MODIS. Для построения моделей латеральной и вертикальной дифференциации использовался ансамблевый алгоритм машинного обучения (стекинг, стековое обобщение, стековая регрессия). Оценку точности полученных картографических моделей определяли с помощью пространственной перекрестной проверки. Результаты пространственной кросс-валидации показывают меньшую точность: коэффициент детерминации 0,46, CCC – 0,63, logRMSE – 0,88 (RMSE – 1,41 г/кг) по сравнению с традиционной перекрестной (R^2_{cv} – 0,68, CCC – 0,81, logRMSE – 0,68 (RMSE – 0,97 г/кг)). Предлагаемая количественная оценка полностью автоматизирована и позволяет воспроизводить моделирование и уточнять результаты по мере получения новых данных о почвах.

Ключевые слова: почвенный покров, пространственное моделирование, стековая регрессия, пространственная кросс-валидация

DOI: 10.55959/MSU0579-9414-5-2022-6-49-63

ВВЕДЕНИЕ

Почва является крупнейшим резервом органического углерода, в котором содержится около 1460,5 Пг (10^{15} грамм или миллиард тонн) углерода в метровой толще [Scharlemann et al., 2014]. Почвенный органический углерод формирует и поддерживает основные режимы, свойства и функции почв – агроэкологические и экосистемные, санитарно-защитные и социально-экономические; является ключевым индикатором качества почв и их устойчивости. Вместе с этим почвенный органический углерод достаточно чувствителен к антропогенному воздействию, и незначительные изменения в его содержании в глобальном масштабе могут оказать заметное влияние на потоки парниковых газов, а затем и на концентрацию CO_2 , CH_4 в атмосфере и, следовательно, на изменение климата. Повышенное внимание и ожидания в отношении потенциала почвенного органического углерода почв для содействия смягчению последствий изменения климата отражены в ряде международных проектов, таких

как проект Глобального почвенного партнерства по рекарбонизации почв [Recarbonization..., 2020], инициатива «4 на 1000» [Minasny et al., 2017], проекты устойчивого развития [Griscom et al., 2017].

Российская Федерация – крупнейшая по площади страна в мире, значительная часть территории которой находится в пределах между 60° и 70° с. ш., что обуславливает значительные запасы органического углерода в почвах [FAO, 2018].

Последние разработки в области автоматизированного прогнозирующего почвенного картографирования и достижения в области данных дистанционного зондирования позволяют специалистам в области наук о почвах переосмыслить методы картографирования почв на основе «факторного» моделирования [Савин и др., 2019]. В этом отношении McBratney с соавторами [McBratney et al., 2003] предложили новую парадигму почвенного картографирования, основанную на количественном моделировании. Эта парадигма основана на модели *scorpan*, включающей пространственные данные по факторам почвообра-

зования. Тем не менее одних пространственных данных зачастую бывает недостаточно, чтобы описать дифференциацию свойств почвы, которую необходимо рассматривать как трехмерное тело.

В последнее время ряд исследователей сосредоточили внимание на картографировании содержания органического углерода почв в национальном масштабе, например в Венгрии [Szatmári et al., 2021], Китае [Liang et al., 2019], Бразилии [Gomes et al., 2019], Швейцарии [Zhou et al., 2021]. Работы по оценке и картографированию содержания органического углерода почв проводились и для территории РФ [Бирюкова, Бирюков, 2011а, 2011б; Чернова и др., 2021; Щепаченко и др., 2013; Rozhkov et al., 1996; Stolbovoi, 2002], где в качестве основы использовалась контурная часть базовых мелкомасштабных почвенных карт и каждый полигон получал экспертное присвоение того или иного среднего значения содержания органического углерода для всех почв, входящих в состав того или иного контура. Но природное варьирование содержания органического углерода почв на таких тематических и схематических картах не всегда будет сопоставимо с выделами на почвенной карте, что приведет к накоплению ошибок. К тому же контуры почв являются неоднородными в почвенном отношении, и степень неоднородности изменяется от контура к контуру. Часть контуров на используемых мелкомасштабных почвенных картах является сложной и может содержать до трех сопутствующих почв.

Известны примеры глобальных оценок содержания и запасов органического углерода. К примеру, глобальный проект SoilGrids 2.0 [Poggio et al., 2021], представляющий пространственную почвенную информацию, основан на большом количестве разнокачественных национальных и локальных данных, комбинация которых влечет за собой ошибки в получаемом картографическом материале. Карта глобального распределения органического углерода почв (GSOC) [FAO, 2018] содержит информацию только о верхнем 30-сантиметровом слое, а содержание на других глубинах не рассматривается.

Учитывая вышесказанное, целью данного исследования является получение информации о латеральном и вертикальном распределении содержания органического углерода почв Российской Федерации. Для достижения поставленной цели решался ряд задач: 1) оценка эффективности применения ансамблевого метода машинного обучения и пространственной кросс-валидации; 2) количественная оценка ошибки предсказания; 3) оценка изменчивости содержания органического углерода почв с глубиной; 4) определение основных факторов, влияющих на пространственное распределение органического углерода почв.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

В качестве основы для моделирования выступал Единый государственный реестр почв России (ЕГРПР) и, в частности, семантическая его часть (база данных) по свойствам почв. База данных состоит из обширной информации по морфологическим признакам почв, физическим и химическим свойствам по 4961 почвенному горизонту из 863 профилей почв.

В качестве объясняющих переменных при моделировании содержания органического углерода почв использовался обширный набор пространственной информации, включая:

- набор биоклиматических переменных CHELSA Climate (среднегодовое количество осадков, среднегодовая температура воздуха, количество снежных дней, сезонность температур, сезонность осадков);
- цифровую модель рельефа (MERIT DEM) и производные от нее (крутизна склонов, кривизна поверхности, топографический индекс влажности);
- долгосрочные усредненные ежедневные значения коэффициента отражения для 2-го (NIR) и 7-го (MIR) каналов спектрорадиометра MODIS (MCD43A4) за период с 01.2015 по 12.2019 г.;
- долгосрочные усредненные 8-дневные значения температуры (ночной и дневной) подстилающей поверхности (MODIS MOD11A2) за период с 01.2010 по 12.2019 г.;
- долгосрочные усредненные 16-дневные значения вегетационного индекса EVI (MODIS MOD13A2) за период с 01.2010 по 12.2019 г.;
- долгосрочные усредненные 4-дневные значения индекса листовой поверхности (LAI) и фотосинтетически активной радиации (FPAR) за период с 01.2010 по 12.2019 г. (MODIS MCD15A3H);
- долгосрочные усредненные 8-дневные значения валовой первичной продукции (GPP) за период с 01.2010 по 12.2019 г. (MODIS MOD17A2H);
- долгосрочные усредненные ежедневные значения проективного покрытия снежного покрова (MODIS MOD10A1) за период с 01.2015 по 12.2019 г.;
- карту наземного покрова по данным Европейского космического агентства¹.

Ковариаты подбирались исходя из предположения о том, что все эти слои пространственной информации могут обеспечивать определенную степень контроля содержания органического углерода почв.

Продукты MODIS были получены с помощью облачной вычислительной платформы Google Earth Engine (GEE) [Gorelick et al., 2017], статистической среды R и пакета rgee.

¹ 300 m annual global land cover time series from 1992 to 2015, URL: <https://www.esa-landcover-cci.org/?q=node/175>.

Глубина – как независимая переменная (ковариата). Недавние исследования показали, что изменения тех или иных свойств почв с глубиной могут быть отображены с помощью моделей, включающих эту самую глубину в качестве переменной-предиктора наряду с пространственными переменными [Zhang et al., 2020]. Некоторые работы описывают использование комбинации отдельных моделей, которые строятся для различных фиксированных интервалов глубин, охватывающих различные части профиля почв. Как правило, используется стандартный набор из шести интервалов глубин (0–5, 5–15, 15–30, 30–60, 60–100, 100–200 см), предложенных инициативой GlobalSoilMap. Такой подход 2,5D-моделирования получил применение в методологии цифровой почвенной картографии [Malone et al., 2009]. Однако 3D-моделирование, учитывающее в более явном виде влияние глубины на изменение свойств почв, позволяет охарактеризовать почву как истинное трехмерное природное тело.

В этом исследовании мы используем подход, предложенный T. Hengl и R. MacMillan [Hengl, MacMillan, 2019], где в качестве глубины присваивается срединное значение, нижний и верхний пределы (при условии, что мощность генетического горизонта более 15 см) почвенного горизонта. Используя такой подход и включая глубину в качестве ковариаты, становится возможным использовать одну единственную модель для прогнозирования содержания углерода почв на любой произвольной глубине.

Рабочий процесс моделирования. Моделирование содержания органического углерода почв включало следующие этапы:

- подготовку исходных данных;
- пересечение (соотнесение) существующих точек наблюдения и значений предикторов в этих точках (создание регрессионной матрицы);
- моделирование с помощью ансамбля методов машинного обучения и использования пространственной кросс-валидации;
- оценку точности полученной модели;
- пространственное предсказание содержания органического углерода почв для каждого элемента (пиксела) раstra.

Подготовка исходных данных. Несмотря на то что в случае использования нелинейных методов анализа данных логарифмическое преобразование целевой переменной не требуется, мы решили, что подобное преобразование даст пропорционально больший вес низким значениям содержания органического углерода почв. Поэтому мы моделировали логарифмически преобразованные значения ($\log_e(x + 1)$, где x – заданное значение) содержания углерода почв, а затем использовали обратную

трансформацию для пространственного прогноза ($e^x - 1$).

Ансамблевое машинное обучение. В этом исследовании мы используем ансамблевый алгоритм машинного обучения, называемый стекингом (стековым обобщением, стековой регрессией) [Breiman, 1996; Laan van der et al., 2007]. Архитектура такого алгоритма содержит две или более базовые модели, а также метамодель, которая объединяет прогнозы базовых моделей.

В роли базовых или нулевых алгоритмов в данном исследовании мы использовали «ансамбль деревьев решений» (Random Forest), бустинг (Gradient boosting machine), обобщенную линейную модель с регуляризацией (ElasticNet). В качестве метамодели – линейную регрессию, которая будет вычислять взвешенную сумму прогнозов, сделанных вышеперечисленными базовыми алгоритмами. Для создания ансамблевого алгоритма машинного обучения мы использовали пакет mlr в статистической среде R.

При использовании одиночных методов, основанных на построении деревьев решений, получают надежные прогнозы только на подпространстве признаков, определяемых обучающей выборкой, а в области экстраполяции (т. е. прогнозировании за пределами обучающей выборки) точность может резко снижаться. Стековое обобщение, сочетающее в себе различные модели (и линейные, и нелинейные), дает более реалистичные прогнозы, особенно в областях экстраполяции. Стекинг учитывает характеристики различных алгоритмов машинного обучения, уменьшает дисперсию единичных моделей и обеспечивает лучшие и более стабильные результаты. Ранее проведенные исследования с использованием стекинга показывают потенциал в повышении точности получаемых карт ряда свойств почв [Taghizadeh-Mehrjardi et al., 2021].

Пространственная перекрестная проверка. Обработка пространственных наборов данных (где к каждому наблюдению прикреплен информация о пространственной привязке) как непространственных приводит к чрезмерно оптимистичным результатам [Brenning, 2005]. Это связано с лежащей в основе данных пространственной автокорреляцией. Чем ближе наблюдения расположены друг к другу, тем больше они похожи.

При выполнении традиционной процедуры кросс-валидации данные обучающего и тестового наборов данных располагаются непосредственно рядом друг с другом (в географическом пространстве). Следовательно, тестовый набор данных будет содержать данные, которые несколько похожи (из-за пространственной автокорреляции) на наблюдения в обучающем наборе данных. Это приведет к тому, что модель, которая была обучена на обучаю-

шей выборке, довольно хорошо работает с тестовыми данными.

Для 5-кратной пространственной перекрестной проверки мы использовали ID пространственных блоков (страт) размером 150×150 км, созданных в конической равнопромежуточной проекции с помощью пакета blockCV в статистической среде R (рис. 1). Это гарантирует, что точки, которые лежат в непосредственной близости (менее 150 км), используются либо для обучения, либо для провер-

ки, что в итоге обеспечивает более объективную оценку точности. Для сравнения нами также была проведена процедура традиционной перекрестной проверки, исключая использование пространственных блоков.

Точность полученных моделей оценивалась с помощью коэффициента детерминации (доля дисперсии, объясняемая метамоделью) R^2_{cv} , корня среднеквадратичной ошибки $RMSE_{cv}$, коэффициента конкордации CCC.

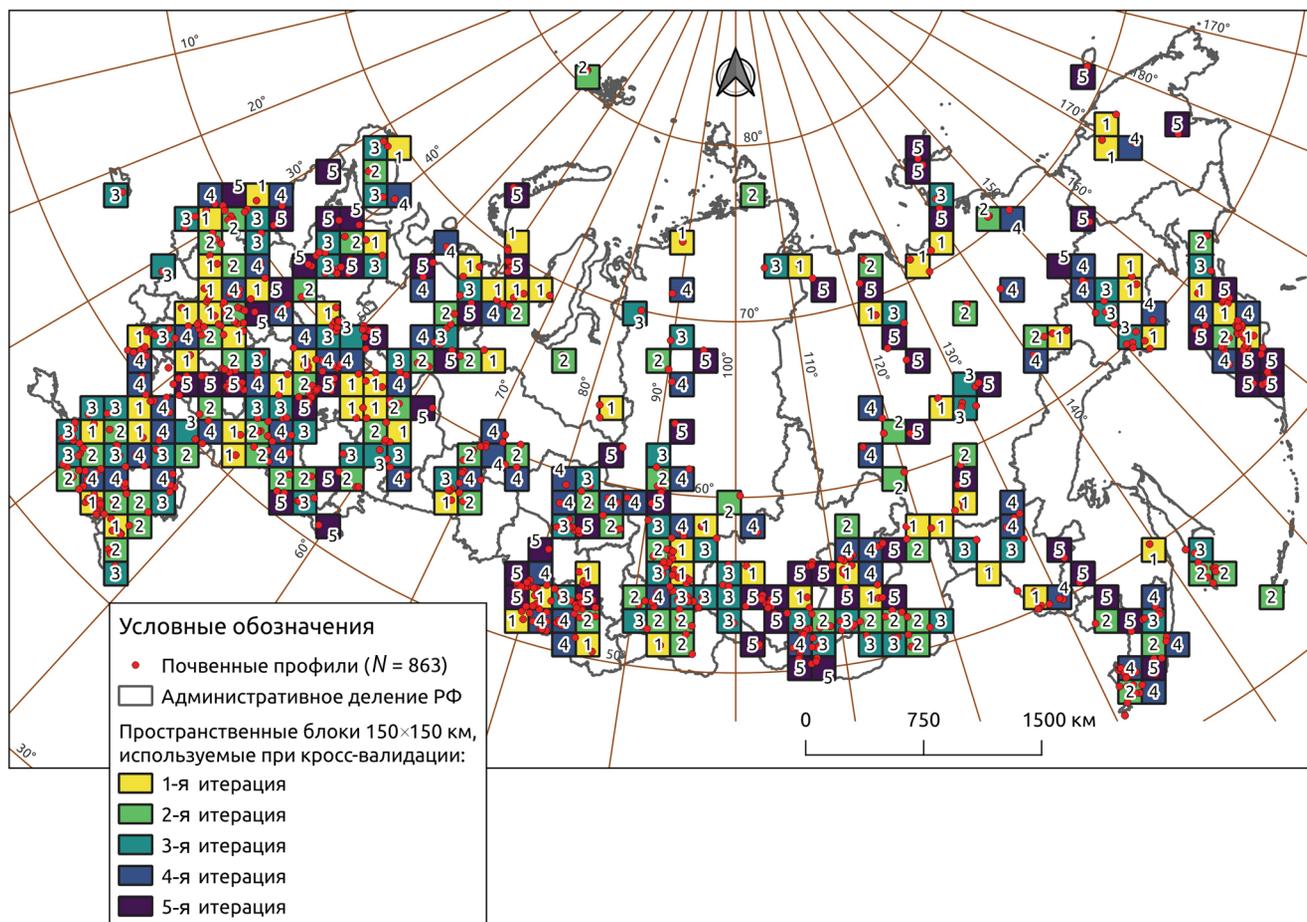


Рис. 1. Пространственная перекрестная проверка.

Для 5-кратной пространственной перекрестной проверки были использованы блоки размером 150×150 км, случайным образом распределенные таким образом, чтобы каждое из пяти подмножеств содержало около 20% наблюдений. Затем проводилась 5-кратная процедура обучения и тестирования ансамблевого алгоритма, каждый раз оставляя одно из подмножеств для валидации

Fig. 1. Spatial cross validation.

We used 150×150 km blocks for 5-fold spatial cross-validation, randomly distributed so that each of the 5 subsets contained about 20% of observations. Then a 5-fold training and testing procedure of the ensemble algorithm was carried out, each time leaving one of the subsets for validation

Карты пространственного распределения неопределенности оценки содержания органического углерода почв были получены с использованием подхода квантильных регрессионных ансамблей деревьев решений (Quantile Regression Forests), предложенного N. Meinshausen [Meinshausen, 2006]

и реализованного в пакете ranger. Этот подход позволяет количественно оценить достоверность прогноза, используя интервалы прогнозирования. Интервал прогнозирования – это оценка интервала, который с заданной вероятностью попадут предсказанные наблюдения. В данном подходе вместо

среднего значения переменной отклика в каждом листе дерева мы будем получать все наблюдаемые отклики на листе. Таким образом, модель предсказывает не только среднее значение переменной, но и полное условное распределение значений.

Таким образом, мы получили верхний (Q_{84}) и нижний (Q_{16}) квантили, соответствующие доверительному интервалу в 68% (т. е. $[\mu - 1\sigma; \mu + 1\sigma]$). Карты неопределенности были получены как половина разницы между этими квантилями: ошибка равна $(Q_{84} - Q_{16})/2$.

Сравнение с другими картами. Одним из важных этапов составления карт является оценка ее качества. Качество получаемых карт оценивалось с помощью сравнения, во-первых, с цифровой почвенной картой РФСР масштабом 1 : 2 500 000, лежащей в основе ЕГРПР; во-вторых, с картами, полученными в рамках проекта SoilGrids 2.0² [Poggio et al., 2021]. Почвенная карта РФСР (под редакцией В.М. Фридланда) является наиболее подробной почвенной картой, созданной по единой системе на всю территорию страны. Эта карта основана на географо-генетических представлениях о почвенном покрове [Конюшков и др., 2020] и является уникальным по полноте информации о почвах страны картографическим материалом.

Основой для сравнения выступала геометрическая часть цифровой почвенной карты РФСР, представленная выделами почв. К определенным выделам (единым по почвенной разности) попиксельно извлекались значения сравниваемых карт (SoilGrids 2.0) и описываемых в данном исследовании. Далее проводился расчет описательной статистики для каждой из сравниваемых карт, и полученные значения содержания органического углерода почв сравнивались с эталонными, в качестве которых выступала информация о разностях почв, лежащая в основе ЕГРПР. В качестве критерия оценки качества карт использовалась близость к эталонным значениям.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Точность моделей и значимость переменных.

Результаты пространственной и рандомизированной (непространственной) перекрестной проверок показаны на рис. 2. Оба графика отражают наличие статистически достоверной зависимости между наблюдаемыми и полученными в процессе моделирования значениями. Результаты пространственной кросс-валидации (см. рис. 2, справа) показывают меньшую точность: коэффициент детерминации

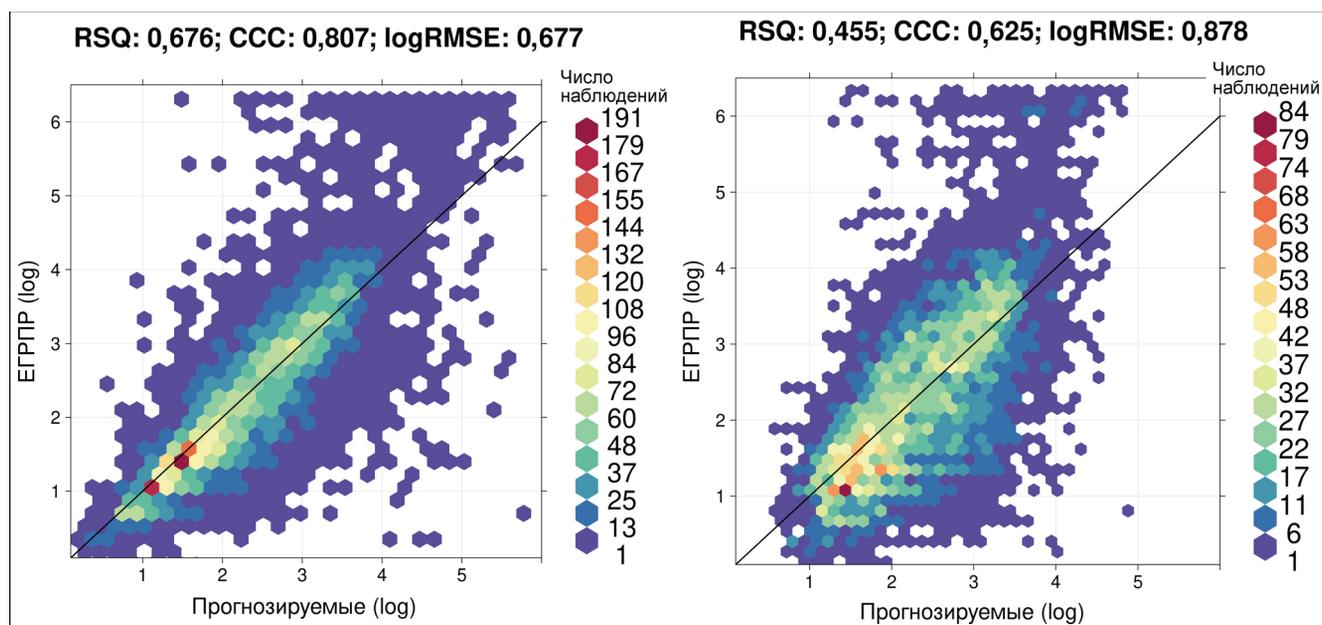


Рис. 2. Результаты рандомизированной (слева) и пространственной (справа) перекрестной проверок. Точность моделей оценивалась с использованием коэффициента детерминации (RSQ), коэффициента конкордации (CCC) и корня среднеквадратичной ошибки прогнозирования (logRMSE). Значения по осям представлены с логарифмическим преобразованием

Fig. 2. Results of randomized (left) and spatial (right) cross-validation.

The model accuracy was assessed using determination coefficient (RSQ), Lin's concordance correlation coefficient (CCC) and root mean square error (logRMSE). The values along the axes are represented with a logarithmic transformation

² ISRIC – World Soil Information, URL: <https://files.isric.org/soilgrids/latest/data/>.

0,46 ($p < 0,01$), CCC – 0,63, logRMSE – 0,88 (RMSE $\pm 1,41$ г/кг) в отличие от рандомизированной, где $R^2_{cv} = 0,68$ ($p < 0,01$), CCC – 0,81, logRMSE – 0,68 (RMSE $\pm 0,97$ г/кг). Достаточный контраст между результатами перекрестной проверки предостерегает от чрезмерно оптимистичной оценки точности прогнозирования с помощью традиционно используемой рандомизированной (непространственной) перекрестной проверки. Следует также добавить, что средняя точность моделирования (RMSE) содержания органического углерода почв в рамках проекта SoilGrids 2.0 составила $\pm 3,96$ г/кг почвы. На приведенных графиках можно наблюдать, что прогнозируемые низкие значения содержания органического углерода почв оказываются слегка завышенными, в то время как прогнозируемые высокие значения оказываются заниженными по сравнению с истинными.

Эффективность применения ансамблевого машинного обучения и пространственной кросс-валидации превосходит эффективность применения одиночного метода моделирования. Так, при сравнении стековой регрессии и ансамбля дере-

вьев решений для последнего получены менее точные результаты: $R^2_{cv} = 0,44$ ($p < 0,01$), CCC – 0,60, logRMSE – 0,90 (RMSE $\pm 1,55$ г/кг). По результатам стековой регрессии получено, что в целом из трех используемых нулевых алгоритмов больший вес при обобщении отдается ансамблю деревьев решений, за которым следуют бустинг и линейная модель. Стековое обобщение не позволяет оценить важность используемых независимых переменных, так как представлено совокупностью различных моделей. В этом случае мы обучили одиночную модель на основе ансамбля деревьев решений (в совокупности с 5-кратной пространственной проверкой), так как ему отдавалось большее предпочтение при объединении.

На рис. 3 отражен ряд независимых переменных, использованных в процессе моделирования. Ось абсцисс на графике показывает значимость: чем выше значение, тем важнее переменная. Глубина является важнейшей переменной в моделировании содержания органического углерода почв. Это объясняется довольно высоким его варьированием с глубиной для различных почв.

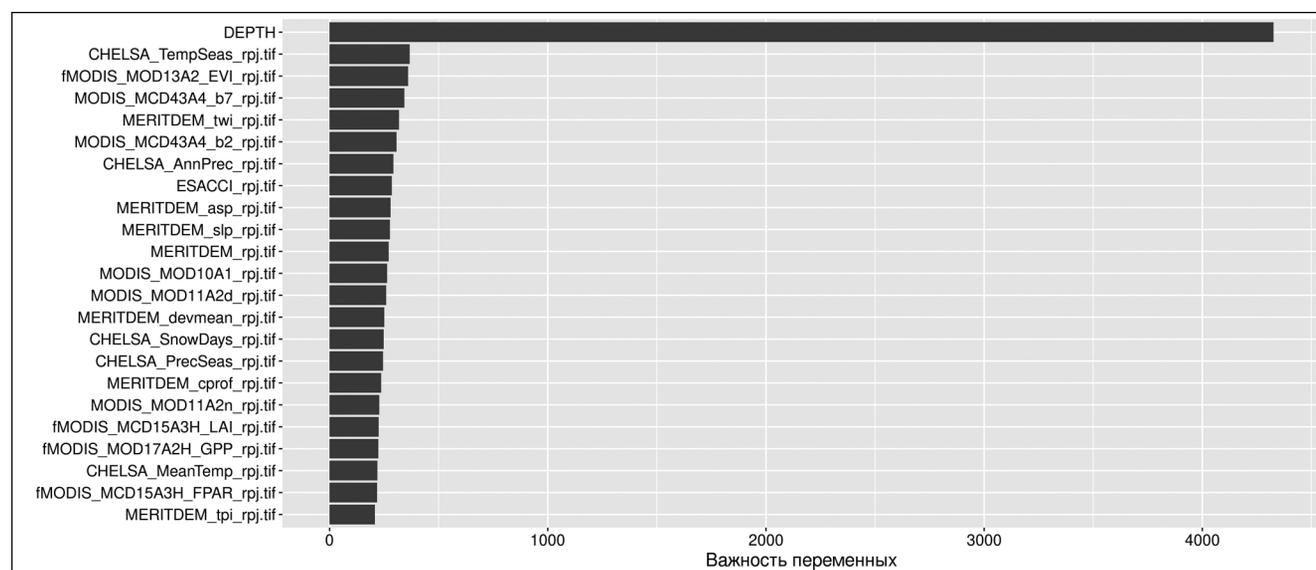


Рис. 3. График важности переменных (предикторов) при моделировании содержания органического углерода почв.

Ось абсцисс на графике отражает значимость: чем выше значение, тем важнее переменная (CHELSA_* – биоклиматические переменные; MERITDEM_* – переменные рельефа; MODIS_* – данные дистанционного зондирования и производные от них; ECASSI – характер наземного покрова; DEPTH – глубина профиля почв)

Fig. 3. Variable importance of predictors during modeling of organic carbon content in soils.

The abscissa displays the importance: the higher the value, the more important is a covariate (CHELSA_* – bioclimatic variables; MERITDEM_* – terrain variables; MODIS_* – remote sensing data; ECASSI – land cover; DEPTH – soil profile depth)

Следующей по степени важности следует одна из биоклиматических переменных CHELSA Climate, а именно сезонность температур, которая отражает величину изменения температуры на основе стандартного отклонения средних месячных темпера-

тур (показатель изменения температуры в течение года). Климат оказывает непосредственное воздействие на накопление углерода в почвах через их увлажнение осадками, промачивание почвы, нагревание и охлаждение, но и также оказывает косвенное

воздействие на растительный и животный мир, т. е. обуславливает тип растительности, количество первичной продукции, темпы создания и разложения органического вещества, состав и интенсивность деятельности почвенной микрофлоры, которая участвует в процессах разложения растительных остатков, гумификации и консервации компонентов опада. Как правило, более низкие температуры способствуют накоплению органического углерода почв, так как в этих условиях скорость аккумуляции выше, чем скорость разложения или миграции.

Далее, в порядке ранжирования переменных по важности следует долгосрочный усредненный вегетационный индекс EVI, отражающий биомассу растительного покрова. Определенные растительные формации характеризуются своими особенностями в составе органического вещества, в поступлении его в почву, в процессах его разложения и во взаимодействии продуктов распада с минеральной частью почвы. Интенсивность процессов аккумуляции / разложения зависит от состава растительности и масштабов ее продуктивности, природных условий трансформации растительных остатков (климата, рельефа, гидрологических условий), что и обуславливает формирование почв с различным содержанием органического углерода.

В десятку наиболее важных переменных при моделировании также входят: 2-й (NIR) и 7-й (MIR)

каналы спектрорадиометра MODIS и ряд производных от цифровой модели рельефа (топографический индекс влажности, экспозиция и крутизна склонов).

Изменчивость содержания органического углерода почв с глубиной. Распределение содержания органического углерода почв в интервале глубин 0–5 см представлено на рис. 4. В целом среднее значение содержания органического углерода почв закономерно увеличивается с юга на север, вплоть до северо-таежной подзоны. Максимум содержания органического углерода почв в указанном интервале приходится на перегнойно-темногумусовые (АН), грубогумусовые (АО), гумусово-слаборазвитые (W), подстильно-торфяные (O, Ooa, Oh) и торфяные (Т, ТО, ТЕ, ТJ) горизонты со значениями от 100 до 301 г/кг. Рассматривая зональное распределение, можно сказать, что наибольшее содержание органического углерода почв в интервале глубин 0–5 см характерно для почв подзоны средней тайги, почв тундровой зоны и почв северо-таежной подзоны. При этом для азиатской части страны, по сравнению с европейской, пространственное распределение содержания органического углерода почв несколько сложнее. Связано это с неблагоприятным проявлением экстроконтинентального климата, наличием вечной мерзлоты и неудовлетворительными условиями

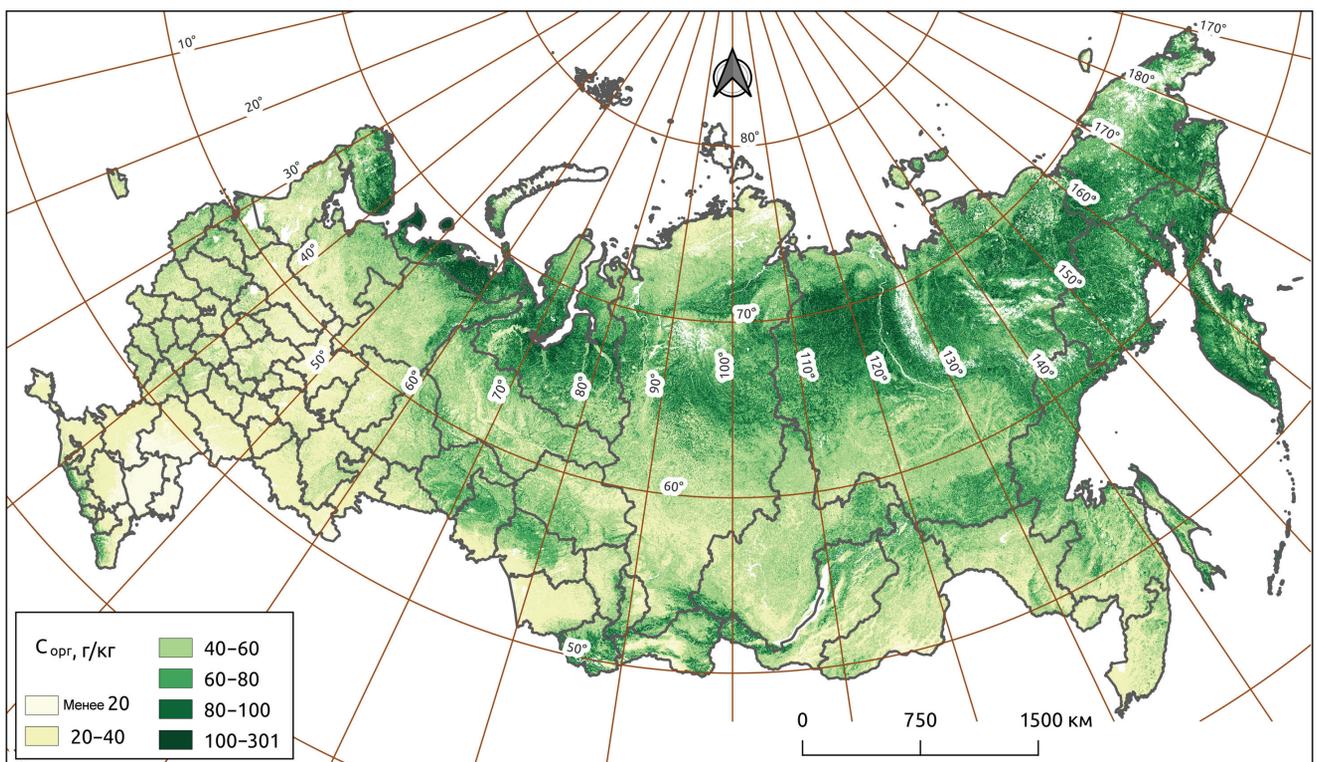


Рис. 4. Содержание органического углерода почв в интервале глубин 0–5 см

Fig. 4. Content of soil organic carbon in the depth range of 0–5 cm

разложения растительного опада. Эти факторы приводят к аккумуляции и консервации органического вещества на поверхности почвы. Европейская часть страны характеризуется более мягким и теплым климатом, влажность и температура оптимальнее для активного течения биологических процессов, что проявляется в ежегодном поступлении в почву растительных остатков и быстром их разложении. В районах проявления вертикальной зональности – в горных провинциях (Большой Кавказ, горные области Сибири и Дальнего Востока) из-за изменений климатических условий проявляется замедленность процессов гумификации и накопление слабогумифицированных соединений, придающих грубый характер органическому веществу. Содержание органического углерода в поверхностных горизонтах таких почв (некоторые типы литоземов (грубогумусированные) и слабо-развитых почв) может достигать 100 и более г/кг.

Анализируя распределение содержания органического углерода почв в интервале 5–15 см (рис. 5), можно сказать, что прослеживается четкая закономерность изменения содержания исследуемого свойства в широтном ряду автоморфных почв, что особенно отчетливо выражено для европейской части РФ, представленной, главным образом, равнинными территориями. Содержание органического углерода растет от подзолистых почв северной и средней тайги (10–20 г/кг) до черноземов глинисто-иллювиальных (40–60 г/кг). Далее, при движении на юг содержание органического углерода почв постепенно уменьшается до минимального уровня у бурых аридных почв (5–10 г/кг), но при движении к побережью Черного моря вновь повышается. Подобная картина может нарушаться в регионах с широким распространением гидрометаморфических почв, торфоземов и торфяных почв с высоким содержанием органического углерода почв.

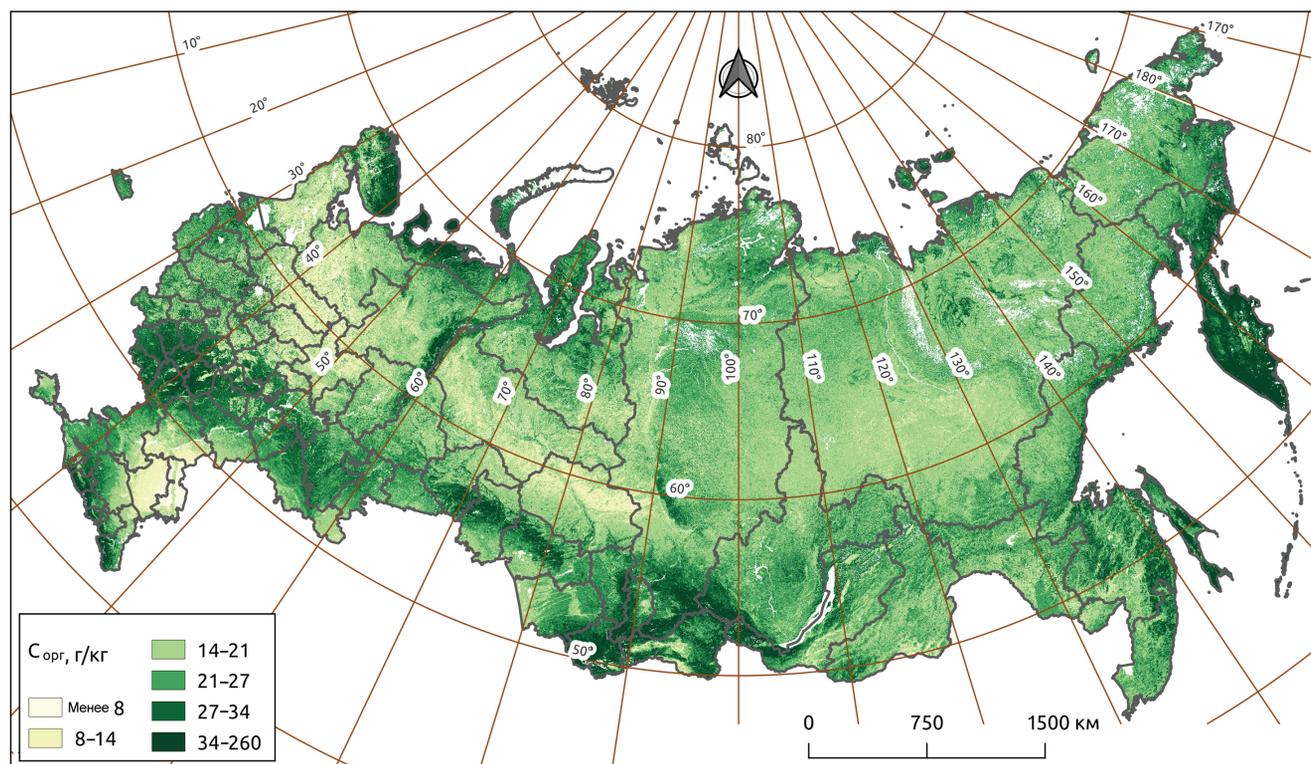


Рис. 5. Содержание органического углерода почв в интервале глубин 5–15 см

Fig. 5. Content of soil organic carbon in the depth range of 5–15 cm

При анализе азиатской части РФ не наблюдается подобной четко выраженной закономерности широтного распределения содержания органического углерода почв. В первую очередь это связано с влиянием экстремально континентального климата, во вторую – с наличием горных областей и проявлением вертикальной зональности. В южной части Западно-Сибирской равнины возможно проследить за-

кономерное увеличение содержания органического углерода почв равнинных территорий (в пределах верхних дерновых или гумусово-аккумулятивных горизонтов) от подзолистых (10–20 г/кг) и дерново-подзолистых (20–35 г/кг) почв до черноземов (40–50 г/кг), и затем уменьшение содержания в каштановых почвах (15–20 г/кг) на границе с северным Казахстаном. Для Среднесибирского плоскогорья

и территории, простирающейся далее на восток, характерны как высокие значения содержания органического углерода почв (50–200 г/кг, Колымская низменность с глееземами, побережье Охотского и Японского морей с буроземами, вулканические почвы Камчатки), так и низкие значения (менее 10 г/кг) – палевые почвы Центральной Якутии. Нельзя не отметить и широкое распространение почв горных провинций, которые занимают значительную территорию азиатской части РФ (Алтай, Саяны, нагорья Забайкалья, хребты и нагорья Якутии и Чукотки, Сихотэ-Алинь) и могут характеризоваться высоким содержанием органического углерода (более 100 г/кг).

Анализируя распределение содержания органического углерода почв по профилю, нужно отметить его неодинаковое распределение. Для одних почв характерно его максимальное содержание в верхней части профиля и резкое уменьшение с глубиной (подзолистые, дерново-подзолистые, темногумусово-глеевые почвы). Для некоторых почв характерно постепенное убывание содержания углерода почв (черноземы, каштановые почвы). Подобная разница в распределении связана не столько с масштабами поступления в почву растительного опада, сколько с характером его локализации. В первую очередь это проявляется при сравнении лесных и травянистых формаций. Также встречаются почвы с бимодаль-

ным распределением содержания органического углерода – с двумя максимумами. Второй максимум связан с процессами иллювиирования водорастворимых органических соединений из верхнего органогенного горизонта. Подобное распределение свойственно, например, иллювиально-гумусовым подзолам Средней и Восточной Сибири, Карелии. Встречаются почвы с полимодальным распределением содержания органического углерода почв по профилю, в которых наблюдается своеобразная «этажность» – залегание серии погребенных профилей. Примером таких почв являются вулканические почвы Камчатки и Курильских островов.

Ошибка прогнозирования. На рис. 6 показан пример карты, иллюстрирующей ошибку прогнозирования. Средняя ошибка прогнозирования в интервале глубин 0–5 см составляет $\pm 3,1$ г/кг. Области с наименьшим количеством точек наблюдений обладают наибольшей неопределенностью. Ошибка прогнозирования также больше в областях наибольшего удаления от точек наблюдений, т. е. в областях экстраполяции. Западно-Сибирская равнина, Средне-Сибирское плоскогорье, Приленское плато являются крупными районами, для которых доступно малое количество профилей и данных по структуре почв, поэтому здесь ошибка выше из-за большего интерквартильного размаха прогнозирования, чем в европейской части РФ. Большая ошибка

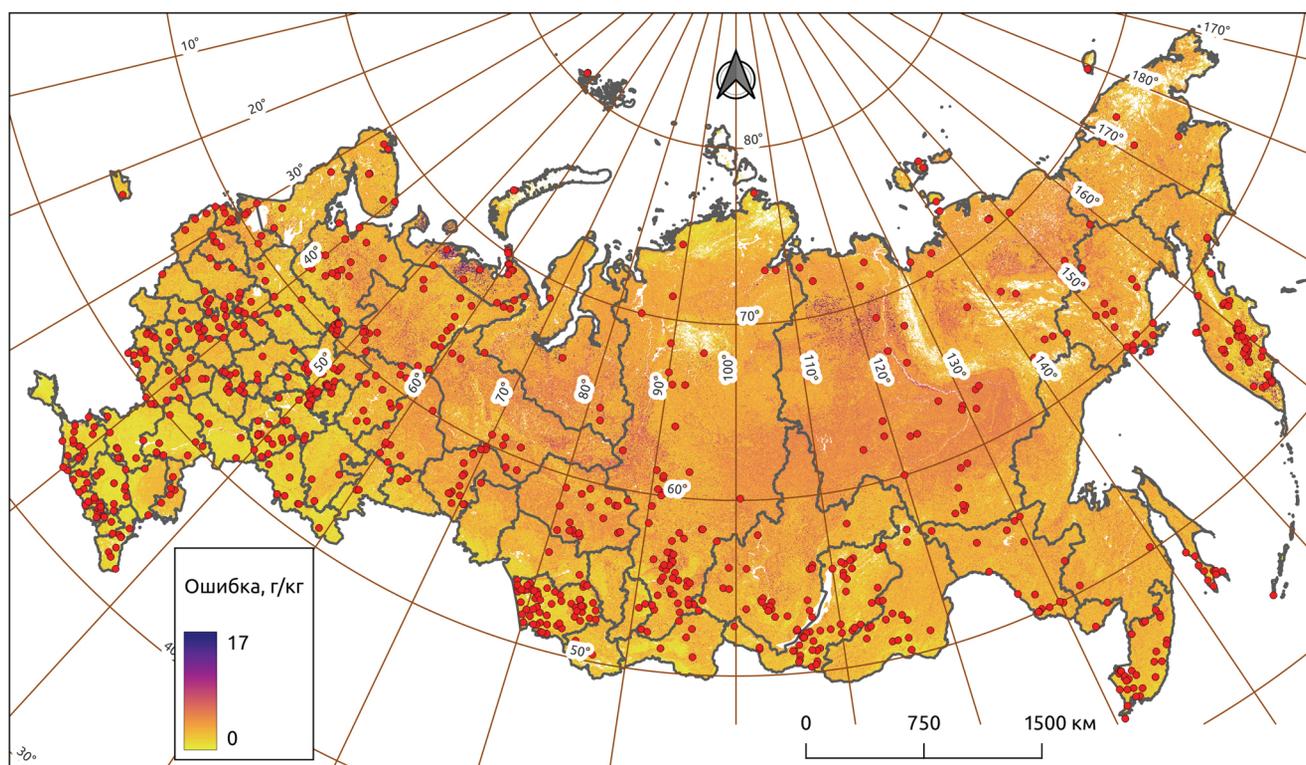


Рис. 6. Карта ошибки прогнозирования содержания органического углерода почв в интервале глубин 0–5 см
Fig. 6. Map of prediction error for soil organic carbon content in the depth range of 0–5 cm

также характерна для областей, где высокие значения единичных наблюдений не могут быть объяснены переменными-предикторами (северная часть Ненецкого автономного округа, к примеру). Интересным также представляется тот факт, что неопределенность прогнозирования больше для регионов с большим диапазоном содержания органического углерода. Здесь лучшие результаты моделирования получены для регионов с более узким диапазоном исследуемого свойства.

Сравнение с существующими почвенными картами. В таблице отражены результаты сравнения карт. Эталонные значения представлены в процентах содержания гумуса, поэтому значения сравниваемых карт были пересчитаны в идентичные величины. Сравнивались карты на одинаковых интервалах глубин, при этом для дерново-подзолистых и серых лесных почв анализировался интервал 5–15 см (так как серогумусовый горизонт АУ указанных типов почв может иметь в верхней своей части подгоризонт с примесью слаборазложившихся растительных остатков, за счет чего общее содержание углерода почв может достигать 80–100 г/кг), для всех остальных почв интервал 0–5 см. По результатам можно отметить, что и средние, и медианные значения содержания гумуса почв (по всем рассматриваемым разностям) составленных нами карт значительно ближе к эталонным. Карты, полученные в рамках проекта SoilGrids 2.0, показывают значительное смещение в сторону увеличения содержания гумуса почв. По-видимому, это объясняется тем, что в рамках проекта используется большое количество разнокачественных данных национального и локального уровней, комбинация которых приводит к подобным ошибкам.

Необходимо отметить также значительное расхождение максимальных значений по сравнению с эталонными. В случае с полученными картами это можно объяснить тем, что использованная для сравнения карта мелкого масштаба составлялась путем пространственной генерализации почвенного покрова, при которой ареалы почв, определяющие специфичность территории, не выдерживали масштабного уменьшения. К тому же, как уже отмечалось ранее, выделы использованной для сравнения карты могут быть сложными и содержать до трех сопутствующих почв, содержание органического углерода в которых может быть действительно большим. Так, к примеру, в пределах азиатской части страны (горная Алтайско-Саянская провинция (e₂)) среди каштановых почв сухих и опустыненных степей, занимающих межгорные котловины, на склонах хребтов (в зависимости от абсолютных высот) встречаются горные черноземы (выщелоченные и типичные), горно-луговые и горные лу-

гово-степные почвы, которые характеризуются действительно большими значениями содержания органического углерода и определяют специфичность территории. С большим разбросом значений для серых лесных почв может наблюдаться подобная картина. Это может объясняться переходным положением данных почв между таежно-лесной и лесостепной зонами. Поэтому в структуре почвенного покрова могут встречаться сложные контрастные сочетания и комплексы серых лесных, серых лесных глеевых, луговых и лугово-болотных почв, солодей.

Нами также было проведено сравнение плотности распределения (probability density function – PDF) для различных карт. Для этого из каждой точки наблюдения извлекались значения содержания органического углерода почв ряда сравниваемых карт в интервале 0–100 см. На рис. 7 приведены графики плотности распределения не только сравниваемых карт, но и графики исходных и модельных значений. Исходя из полученных графиков плотности распределения, можно отметить заметный сдвиг распределения значений органического углерода почв согласно данным SoilGrids 2.0 в сторону больших значений. Распределение значений органического углерода почв тестируемых карт ближе к истинному, хотя можно отметить, что низкие значения содержания органического углерода оказываются слегка завышенными, в то время как высокие значения – заниженными.

Вместе с тем полученные нами карты не лишены недостатков. Один из самых значимых – значительно заниженные результаты содержания органического углерода почв Западно-Сибирской низменности с большими ареалами органогенных торфяных почв (междуречье Оби и Иртыша, Васюганская равнина). Ожидается, что содержание органического углерода в таких почвах будет превышать 350 г/кг на всех рассматриваемых интервалах глубин (и даже глубже). Что, однако, не находит отражение на полученных нами картах. Это связано с недостаточным и ограниченным количеством профилей органогенных торфяных почв в составе исходной базы данных и, следовательно, недостатком их в составе обучающей выборки при моделировании. Следовательно, для этих областей результаты прогнозирования имеют ограниченное использование. Заниженными также оказываются результаты прогнозирования для верхних перегнойно-темногумусовых (АН), грубогумусовых (АО), гумусово-слаборазвитых (W), подстилочно-торфяных (O, Ooa, Oh) горизонтов почв, для которых содержание органического углерода должно быть выше 350 г/кг, но на полученных картах не превышает 301 г/кг. Аналитических данных по профилям почв севернее

Таблица

Результаты сравнения полученных карт с картами проекта SoilGrids 2.0

Наименование почв в терминах ЕГРПР	Содержание гумуса согласно ЕГРПР (эталонные значения), %	Содержание гумуса согласно полученным картам (статистика), %						Содержание гумуса согласно картам SoilGrids 2.0 (статистика), %					
		Минимум	1-й кв.	Медиана	Среднее	3-й кв.	Максимум	Минимум	1-й кв.	Медиана	Среднее	3-й кв.	Максимум
Дерново-подзолистые	3-7 (целина), 1-2,5 (пашня)	0,50	2,85	3,52	3,72	4,34	28,28	2,64	9,1	11,77	13,4	16,9	42,19
Серые лесные	4-8	1,24	3,69	4,39	4,45	5,12	27,93	2,49	6,72	7,83	8,17	9,26	28,76
Черноземы типичные	8-12	2,07	5,52	6,21	6,21	6,90	15,69	5,16	12,22	15,00	15,03	17,50	28,03
Черноземы южные	3-6	1,21	3,79	4,48	4,47	5,00	13,45	5,41	9,67	10,95	11,46	12,64	33,27
Каштановые	2-4	1,03	2,93	3,62	4,17	4,83	23,45	4,12	9,15	11,60	11,98	14,05	39,31
Бурые полупустынные	0,7-1,4	0,69	1,21	1,38	1,60	2,07	4,48	4,31	6,79	7,55	7,89	8,67	24,71

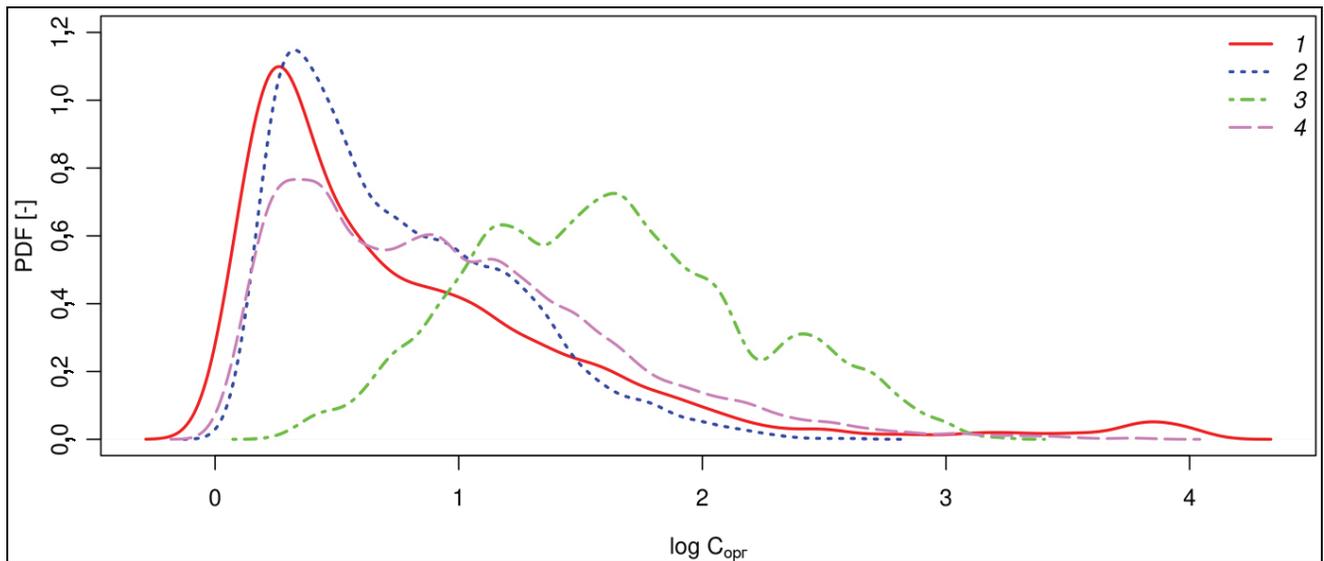


Рис. 7. Сравнение графиков плотности распределения значений органического углерода почв: 1 – наблюдаемые значения ЕГРПР; 2 – полученная модель; 3 – данные SoilGrids 2.0 (0–100 см) [Poggio et al., 2021]; 4 – тестируемые карты (0–100 см)

Fig. 7. Comparison of probability density functions of soil organic carbon: 1 – the observed values; 2 – the resulting model; 3 – the SoilGrids 2.0 data (0–100 cm) [Poggio et al., 2021]; 4 – the tested maps (0–100 cm)

60° с. ш. оказывается недостаточно для адекватного прогнозирования содержания органического углерода почв в пределах указанных горизонтов. Решение данных проблем кроется в получении новых репрезентативных профильных данных, в первую очередь для недостаточно представленных регионов. Вероятно, при моделировании следует использовать и другие ковариаты, которые могут помочь объяснить и спрогнозировать пространственные изменения содержания органического углерода почв.

ВЫВОДЫ

Результаты пространственной кросс-валидации показали меньшую (при сравнении с рандомизированной) точность: коэффициент детерминации – 0,46, CCC – 0,63, logRMSE – 0,88 (RMSE – 1,41 г/кг).

Благодарности. Исследования выполнены при поддержке гранта РФФИ 19-05-50063, а также при поддержке Программы стратегического академического лидерства РУДН.

СПИСОК ЛИТЕРАТУРЫ

- Бирюкова О.Н., Бирюков М.В. Содержание органического углерода в верхних горизонтах почв // Национальный атлас почв Российской Федерации. Москва: Астрель-АСТ, 2011а. С. 230–231.
- Бирюкова О.Н., Бирюков М.В. Запасы органического углерода в почвах // Национальный атлас почв Российской Федерации. М.: Астрель-АСТ, 2011б. С. 242–243.
- Савин И.Ю., Жоголев А.В., Прудникова Е.Ю. Современные тренды и проблемы почвенной картографии // Почвоведение. 2019. Т. 52. № 5. С. 517–528. DOI: 10.1134/S1064229319050107.
- Конюшков Д.Е., Ананко Т.В., Герасимова М.И., Лебедева И.И. Актуализация содержания почвенной карты РСФСР масштаба 2,5 млн в формате классификации почв России для создания новой цифровой карты //

- Бюллетень Почвенного института имени В.В. Докучаева. 2020. № 102. С. 21–48. DOI: 10.19047/0136-1694-2020-102-21-48.
- Чернова О.В., Голозубов О.М., Алябина И.О., Щепаченко Д.Г. Комплексный подход к картографической оценке запасов органического углерода в почвах России // Почвоведение. 2021. Т. 54. № 3. С. 325–336. DOI: 10.1134/S1064229321030042.
- Щепаченко Д.Г., Мухортова Л.В., Швиденко А.З., Ведрова Э.Ф. Запасы органического углерода в почвах России // Почвоведение. 2013. Т. 46. № 2. С. 107–116. DOI: 10.1134/S1064229313020129.
- Breiman L. Stacked regressions, *Mach. Learn.*, 1996, vol. 24, no. 1, p. 49–64, DOI: 10.1007/bf00117832.
- Brenning A. Spatial prediction models for landslide hazards: review, comparison and evaluation, *Nat. Hazards Earth Syst. Sci.*, 2005, vol. 5, no. 6, p. 853–862, DOI: 10.5194/nhess-5-853-2005.
- FAO. Global Soil Organic Carbon Map (GSOC map) Technical Report, 2018, FAO, Rome, 162 p.
- Gomes L.C., Faria R.M., Souza de E., Veloso G.V., Schaefer C.E.G.R., Filho E.I.F. Modelling and mapping soil organic carbon stocks in Brazil, *Geoderma*, 2019, vol. 340, p. 337–350, DOI: 10.1016/j.geoderma.2019.01.007.
- Gorelick N., Hancher M., Dixon M., Ilyushchenko S., Thau D., Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone, *Remote Sens. Environ.*, 2017, vol. 202, p. 18–27, DOI: 10.1016/j.rse.2017.06.031.
- Griscom B.W., Adams J., Ellis P.W., Houghton R.A., Lomax G., Miteva D.A., Schlesinger W.H., Shoch D., Siikamäki J.V., Smith P. et al. Natural climate solutions, *Proc. Natl. Acad. Sci.*, 2017, vol. 114, no. 44, p. 11645–11650.
- Laan van der M.J., Polley E.C., Hubbard A.E. Super Learner, *Stat. Appl. Genet. Mol. Biol.*, 2007, vol. 6, no. 1, p. 1–23.
- Liang Z., Chen S., Yang Y., Zhou Y., Shi Z. High-resolution three-dimensional mapping of soil organic carbon in China: Effects of SoilGrids products on national modeling, *Sci. Total Environ.*, 2019, vol. 685, DOI: 10.1016/j.scitotenv.2019.05.332.
- Malone B.P., McBratney A.B., Minasny B., Laslett G.M. Mapping continuous depth functions of soil carbon storage and available water capacity, *Geoderma*, 2009, vol. 154, no. 1–2, p. 138–152, DOI: 10.1016/j.geoderma.2009.10.007.
- McBratney A.B., Mendonça Santos M.L., Minasny B. On digital soil mapping, *Geoderma*, 2003, vol. 117, no. 1–2, p. 3–52.
- Meinshausen N. Quantile Regression Forests, *J. Mach. Learn. Res.*, 2006, vol. 7, p. 983–999.
- Minasny B., Malone B.P., McBratney A.B., Angers D.A., Arrouays D., Chambers A., Chaplot V., Chen Z.S., Cheng K., Das B.S. et al. Soil carbon 4 per mille, *Geoderma*, 2017, vol. 292, p. 59–86, DOI: 10.1016/j.geoderma.2017.01.002.
- 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, vol. 7, no. 1, p. 217–240, DOI: 10.5194/soil-7-217-2021.
- Rozhkov V.A., Wagner V.B., Kogut B.M., Konyushkov D.E., Nilsson S., Sheremet V.B., Shvidenko A.Z. Soil Carbon Estimates and Soil Carbon Map for Russia, *Analysis*, 1996, p. 1–44.
- Scharlemann J.P.W., Tanner E.V.J., Hiederer R., Kapos V. Global soil carbon: understanding and managing the largest terrestrial carbon pool, *Carbon Manag.*, 2014, vol. 5, no. 1, p. 81–91, DOI: 10.4155/cmt.13.77.
- Stolbovoi V. Carbon in Russian soils, *Clim. Change*, 2002, vol. 55, no. 1–2, p. 131–156.
- Szatmári G., Pásztor L., Heuvelink G.B.M. Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics, *Geoderma*, 2021, vol. 403, DOI: 10.1016/j.geoderma.2021.115356.
- Taghizadeh-Mehrjardi R., Hamzehpour N., Hassanzaadeh M., Heung B., Ghebleh Goydaragh M., Schmidt K., Scholten T. Enhancing the accuracy of machine learning models using the super learner technique in digital soil mapping, *Geoderma*, 2021, vol. 399, p. 115108, DOI: 10.1016/j.geoderma.2021.115108.
- Zhang Y., Ji W., Saurette D.D., Easher T.H., Li H., Shi Z., Adamchuk V.I., Biswas A. Three-dimensional digital soil mapping of multiple soil properties at a field-scale using regression kriging, *Geoderma*, 2020, vol. 366, p. 114253, DOI: 10.1016/j.geoderma.2020.114253.
- Zhou T., Geng Y., Ji C., Xu X., Wang H., Pan J., Bumberger J., Haase D., Lausch A. Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images, *Sci. Total Environ.*, 2021, vol. 755, p. 142661, DOI: 10.1016/j.scitotenv.2020.142661.

Электронный ресурс

Hengl T., MacMillan R.A. Predictive Soil Mapping with R. OpenGeoHub foundation, Wageningen, the Netherlands, 2019, URL: <http://www.soilmapper.org> (дата обращения 30.10.2021).

Recarbonization of global soils – A tool to support the implementation of the Koronivia Joint Work on Agriculture, 2020, URL: <https://www.fao.org/documents/card/en/c/ca6522en/> (дата обращения 30.10.2021).

Поступила в редакцию 21.11.2021

После доработки 05.03.2022

Принята к публикации 13.05.2022

ESTIMATION OF ORGANIC CARBON CONTENT IN RUSSIAN SOILS USING ENSEMBLE MACHINE LEARNING

A.V. Chinilin¹, I.Yu. Savin²

¹FRC "V.V. Dokuchaev Soil Science Institute", Department of soil genesis, geography and classification

²Institute of Environmental Engineering, Peoples Friendship University of Russia (RUDN University),
Department of environmental management

¹Ph.D. in Biology; email: andrey.chinilin@yandex.ru

²D.Sc. in Agriculture, Professor, Academician of the RAS; email: savigory@gmail.com

The study presents a modern quantitative assessment of the content of organic carbon in Russian soils, taking into account their huge variety, and reflects the understanding of individual factors regulating and controlling the content of organic carbon in soils of the country. The paper gives the results of three-dimensional modeling of organic carbon content in soils at several standard depths (0–5, 5–15, 15–30, 30–60, 60–100 cm) for the territory of the Russian Federation with 500 m spatial resolution using the ensemble machine learning. Automated predictive mapping was based on 4 961 soil horizons from 863 soil profiles, as well as on the extensive set of spatial information, including bioclimatic variables, digital elevation model and its derivatives, and the long-term averaged time series of MODIS data. An ensemble machine learning algorithm (stacking, stacked generalization and stacked regression) was used to build models of spatial and vertical distribution. The accuracy of obtained cartographic models was assessed using spatial cross-validation. The results of spatial cross-validation show lower accuracy: the coefficient of determination is 0,46, CCC – 0,63, logRMSE – 0,88 (RMSE – 1,41 g/kg) compared to randomize (R_{cv}^2 – 0,68, CCC – 0,81, logRMSE – 0,68 (RMSE – 0,97 g/kg)). The proposed quantitative assessment is fully automated and makes it possible to reproduce the modeling and refine the results as new soil data are obtained.

Keywords: soil cover, spatial modeling, stacked regression, spatial cross-validation

Acknowledgments. The study was financially supported by the Russian Foundation for Basic Research (project 19-05-50063), and by the RUDN University Strategic Academic Leadership Program.

REFERENCES

- Biryukova O.N., Biryukov M.V. [Soil organic carbon stocks], *Natsional'nyi atlas pochv Rossiiskoi Federatsii* [National Atlas of Soils of the Russian Federation], Moscow, Astrel', AST Publ., 2011b, p. 242–243. (In Russian)
- Biryukova O.N., Biryukov M.V. [The content of organic carbon in the upper horizons of soils], *Natsional'nyi atlas pochv Rossiiskoi Federatsii* [National Atlas of Soils of the Russian Federation], Moscow, Astrel', AST Publ., 2011a, p. 230–231. (In Russian)
- Breiman L. Stacked regressions, *Mach. Learn.*, 1996, vol. 24, no. 1, p. 49–64. DOI: <https://doi.org/10.1007/bf00117832>.
- Brenning A. Spatial prediction models for landslide hazards: review, comparison and evaluation, *Nat. Hazards Earth Syst. Sci.*, 2005, vol. 5, no. 6, p. 853–862, DOI: <https://doi.org/10.5194/nhess-5-853-2005>.
- Chernova O.V., Golozubov O.M., Alyabina I.O., Schepaschenko D.G. Integrated Approach to Spatial Assessment of Soil Organic Carbon in the Russian Federation, *Eurasian Soil Sci.*, 2021, vol. 54, no. 3, p. 325–336, DOI: <https://doi.org/10.1134/S1064229321030042>.
- FAO. *Global Soil Organic Carbon Map (GSOC map)*, Technical Report, 2018, FAO, Rome, 162 p.
- Gomes L.C., Faria R.M., Souza de E., Veloso G.V., Schaefer C.E.G.R., Filho E.I.F. Modelling and mapping soil organic carbon stocks in Brazil, *Geoderma*, 2019, vol. 340, p. 337–350, DOI: <https://doi.org/10.1016/j.geoderma.2019.01.007>.
- Gorelick N., Hancher M., Dixon M., Ilyushchenko S., Thau D., Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone, *Remote Sens. Environ.*, 2017, vol. 202, p. 18–27, DOI: <https://doi.org/10.1016/j.rse.2017.06.031>.
- Griscom B.W., Adams J., Ellis P.W., Houghton R.A., Lomax G., Miteva D.A., Schlesinger W.H., Shoch D., Sii-kamäki J.V., Smith P. et al. Natural climate solutions, *Proc. Natl. Acad. Sci., U.S.A.*, 2017, vol. 114, no. 44, p. 11645–11650.
- Konyushkov D.E., Ananko T.V., Gerasimova M.I., Lebedeva I.I. Actualization of the contents of the soil map of Russian Federation (1 : 2,5 M scale) in the format of the classification system of Russian soils for the development of the new digital map of Russia, *Dokuchaev Soil Bull.*, 2020, no. 102, p. 21–48, DOI: <https://doi.org/10.19047/0136-1694-2020-102-21-48>.
- Laan van der M.J., Polley E.C., Hubbard A.E. Super Learner, *Stat. Appl. Genet. Mol. Biol.*, 2007, vol. 6, no. 1, p. 1–23.
- Liang Z., Chen S., Yang Y., Zhou Y., Shi Z. High-resolution three-dimensional mapping of soil organic carbon in China: Effects of SoilGrids products on national modeling, *Sci. Total Environ.*, 2019, vol. 685, DOI: <https://doi.org/10.1016/j.scitotenv.2019.05.332>.
- Malone B.P., McBratney A.B., Minasny B., Laslett G.M. Mapping continuous depth functions of soil carbon storage and available water capacity, *Geoderma*, 2009, vol. 154, no. 1–2, p. 138–152, DOI: <https://doi.org/10.1016/j.geoderma.2009.10.007>.

- McBratney A.B., Mendonça Santos M.L., Minasny B. On digital soil mapping, *Geoderma*, 2003, vol. 117, no. 1–2, p. 3–52.
- Meinshausen N. Quantile Regression Forests, *J. Mach. Learn. Res.*, 2006, vol. 7, p. 983–999.
- Minasny B., Malone B.P., McBratney A.B., Angers D.A., Arrouays D., Chambers A., Chaplot V., Chen Z.S., Cheng K., Das B.S. et al. Soil carbon 4 per mille, *Geoderma*, 2017, vol. 292, p. 59–86, DOI: <https://doi.org/10.1016/j.geoderma.2017.01.002>.
- 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, vol. 7, no. 1, p. 217–240, DOI: <https://doi.org/10.5194/soil-7-217-2021>.
- Rozhkov V.A., Wagner V.B., Kogut B.M., Konyushkov D.E., Nilsson S., Sheremet V.B., Shvidenko A.Z. Soil Carbon Estimates and Soil Carbon Map for Russia, *Analysis*, 1996, p. 1–44.
- Savin I.Y., Zhogolev A.V., Prudnikova E.Y. Modern Trends and Problems of Soil Mapping, *Eurasian Soil Sci.*, 2019, vol. 52, no. 5, p. 471–480, DOI: <https://doi.org/10.1134/S1064229319050107>.
- Scharlemann J.P.W., Tanner E.V.J., Hiederer R., Kapos V. Global soil carbon: understanding and managing the largest terrestrial carbon pool, *Carbon Manag.*, 2014, vol. 5, no. 1, p. 81–91, DOI: 10.4155/cmt.13.77.
- Schepaschenko D.G., Shvidenko A.Z., Mukhortova L.V., Vedrova E.F. The pool of organic carbon in the soils of Russia, *Eurasian Soil Sci.*, 2013, vol. 46, no. 2, p. 107–116, DOI: 10.1134/S1064229313020129.
- Stolbovoi V. Carbon in Russian soils, *Clim. Change*, 2002, vol. 55, no. 1–2, p. 131–156.
- Szatmári G., Pásztor L., Heuvelink G.B.M. Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics, *Geoderma*, 2021, vol. 403, DOI: <https://doi.org/10.1016/j.geoderma.2021.115356>.
- Taghizadeh-Mehrjardi R., Hamzeshpour N., Hassanzadeh M., Heung B., Ghebleh Goydaragh M., Schmidt K., Scholten T. Enhancing the accuracy of machine learning models using the super learner technique in digital soil mapping, *Geoderma*, 2021, vol. 399, p. 115108, DOI: <https://doi.org/10.1016/j.geoderma.2021.115108>.
- Zhang Y., Ji W., Saurette D.D., Easher T.H., Li H., Shi Z., Adamchuk V.I., Biswas A. Three-dimensional digital soil mapping of multiple soil properties at a field-scale using regression kriging, *Geoderma*, 2020, vol. 366, p. 114253, DOI: <https://doi.org/10.1016/j.geoderma.2020.114253>.
- Zhou T., Geng Y., Ji C., Xu X., Wang H., Pan J., Bumberger J., Haase D., Lausch A. Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images, *Sci. Total Environ.*, 2021, vol. 755, p. 142661, DOI: <https://doi.org/10.1016/j.scitotenv.2020.142661>.

Web-source

- Hengl T., MacMillan R.A. Predictive Soil Mapping with R. OpenGeoHub foundation, Wageningen, the Netherlands, 2019, URL: <http://www.soilmapper.org> (access date 30.10.2021).

Received 21.11.2021

Revised 05.03.2022

Accepted 13.05.2022