

МЕТОДЫ И СРЕДСТВА ОБРАБОТКИ И ИНТЕРПРЕТАЦИИ КОСМИЧЕСКОЙ ИНФОРМАЦИИ

ЦИФРОВОЕ КАРТОГРАФИРОВАНИЕ СОДЕРЖАНИЯ И ЗАПАСОВ ОРГАНИЧЕСКОГО УГЛЕРОДА В ПОЧВАХ ПРЕДСАЛАИРЬЯ С ИСПОЛЬЗОВАНИЕМ ОНЛАЙН-ПЛАТФОРМЫ GOOGLE EARTH ENGINE И АЛГОРИТМА СЛУЧАЙНОГО ЛЕСА

© 2025 г. Н. В. Гопп¹*, Т. В. Нечаева¹

¹ФГБУН Институт почвоведения и агрохимии СО РАН, Новосибирск, Россия

*E-mail: gopp@issa-siberia.ru

Поступила в редакцию 22.10.2024 г.

Для ключевого участка в Предсалаирье проведено цифровое картографирование содержания почвенного органического углерода (ПОУ) в пахотном слое (0–30 см) почв алгоритмом случайного леса, реализованного на облачной онлайн-платформе Google Earth Engine (GEE). В качестве предикторов в модели случайного леса использовались: 19 биоклиматических переменных WorldClim; 5 климатических переменных, рассчитанных на основе WorldClim и почвенно-климатического атласа; 8 вегетационных индексов, рассчитанных на основе Landsat 8; 26 морфометрических характеристик рельефа, рассчитанных на основе цифровой модели высот ALOS; 2 переменные, характеризующие пространственное положение (долгота и широта). Коэффициенты корреляции (R) между содержанием ПОУ и значениями предикторов учтены при формировании следующих наборов предикторов: 1) BIO11+RVI; 2) Longitude+CNBL; 3) SAT10+CC+Texture; 4) 60 предикторов; 5) 42 (без кривизн рельефа, вегетационных индексов и предикторов с нулевыми значениями); 6) 37 (все с $R > \pm 0.5$); 7) 32 (все с $R > \pm 0.3$ без вегетационных индексов); 8) 27 (все с $R > \pm 0.5$ без вегетационных индексов); 9) 23 (без BIO1–19, кривизн рельефа, вегетационных индексов и предикторов с нулевыми значениями). Результат моделирования содержания ПОУ на основе 32 предикторов и обучающего набора данных ($n = 42$) с более низким RMSE (0.72) выбран в качестве лучшего. С применением педотрансферной функции проведено моделирование плотности сложения почв, которая совместно с картой содержания ПОУ была использована для составления карты запасов ПОУ. Содержание ПОУ в пахотном слое (0–30 см) варьировало от 1.3 до 6.1% по фактическим данным, а запасы ПОУ – от 84 до 203 т/га. Наибольшее содержание и запасы ПОУ выявлены в почвах верхней части склона, тогда как вниз по склону отмечено постепенное снижение значений этих показателей. Плотность сложения почв по расчетным данным варьировала в диапазоне от 1.20 до 1.36 г/см³ и увеличивалась вниз по склону, т.е. имела обратный тренд распределения по сравнению с содержанием и запасами ПОУ. Общие запасы ПОУ в пахотном слое (0–30 см) почв исследуемой территории площадью 225 га составили 28.7 кт.

Ключевые слова: климат, рельеф, черноземы, серые лесные почвы, SAGA GIS, Random Forest, машинное обучение

DOI: 10.31857/S0205961425020015, **EDN:** EJDWQO

ВВЕДЕНИЕ

Тематические карты являются основным источником получения информации о пространственном распределении содержания и запасов почвенного органического углерода (ПОУ), которые помогают улучшить наши знания о плодородии почв, последствиях изменения климата, а также в почвенно-агроэкологическом мониторинге и моделировании. Цифровое почвенное картографирование (ЦПК) базируется на идеях В.В. Докучаева и Х. Дженни о связи свойств почв с факторами почвообразования. Геопространственные данные о факторах почвообразования в ЦПК принято называть предикторами или ковариатами окружающей среды (Hengl, 2007; McBratney et al., 2003),

которые необходимы для геопространственного моделирования и создаются на основе данных дистанционного зондирования Земли, архивных источников и др. Установление основных факторов, влияющих на качественные и количественные характеристики почв, дает ценную информацию для понимания их пространственного распределения (Suleymanov et al., 2024).

Особое внимание в ЦПК уделяется изучению взаимосвязи между предикторами и различными свойствами почв, необходимой для обоснования выбора того или иного метода картографирования, которые в настоящее время отличаются большим разнообразием. К ним относятся обобщенные линейные модели, деревья классификации и ре-

грессии, нейронные и нечеткие нейронные сети, геостатистика (Hengl, 2007). Алгоритм случайного леса является перспективным и более надежным с точки зрения уменьшения шума в данных и способен обрабатывать как количественные, так и качественные наборы геопространственных данных (Breiman, 2001; Gandhi, 2024). Потенциал алгоритма случайного леса в цифровом картографировании содержания и запасов ПОУ был продемонстрирован в следующих исследованиях (Cutler et al., 2007; Dharumarajan et al., 2017; Grimm et al., 2008; FAO and ITPS, 2020; Sreenivas et al., 2016; Suleymanov et al., 2024; Vågen et al., 2016).

Несмотря на широкое распространение различных алгоритмов машинного обучения, используемых в ЦПК, в Российской Федерации количество исследований в этом направлении недостаточно. В этом контексте цель исследования — провести цифровое картографирование содержания и запасов почвенного органического углерода (ПОУ), а также плотности сложения в пахотном слое (0–30 см) почв Предсалаирской дренированной равнины с использованием набора геопространственных данных, алгоритма случайного леса, реализованного на облачной онлайн-платформе Google Earth Engine (GEE), и педотрансферной функции. В настоящей работе проведено сравнение влияния предикторов на результаты картографирования содержания ПОУ, относящихся к 4-м факторам почвообразования (климат, рельеф, растительность, пространственное положение). Коэффициенты корреляции (R) между содержанием ПОУ и значениями предикторов учтены при формировании следующих наборов предикторов: 1) BIO11+RVI; 2) Longitude+CNBL; 3) SAT10+CC+Texture; 4) 60 предикторов; 5) 42 (без кривизн рельефа, вегетационных индексов и предикторов с нулевыми значениями); 6) 37 (все с $R > \pm 0.5$); 7) 32 (все с $R > \pm 0.3$ без вегетационных индексов); 8) 27 (все с $R > \pm 0.5$ без вегетационных индексов); 9) 23 (без BIO1–19, кривизн рельефа, вегетацион-

ных индексов и предикторов с наличием нулевых значений в растровой карте).

ОБЪЕКТЫ И МЕТОДЫ

Исследования проведены на территории Предсалаирской дренированной равнины (Предсалаирье) в лесостепной зоне юго-восточной части Западной Сибири (Новосибирская область, Тогучинский район, с. Усть-Каменка, 55°02'20" с.ш., 83°50'00" в.д.).

Участок пахотного угодья площадью 225 га и протяженностью 4 км занимает территорию водосборных бассейнов рек Ирба и Хайрузовка (рис. 1). На исследуемой территории распространены агрочерноземы, агротемно-серые и агросерые почвы (табл. 1).

Согласно группировке почв по содержанию гумуса (Шишов и др., 2004), агрочерноземы в верхней части склона относятся к сильногумусированным (5–8%); агрочерноземы, агротемно-серые и агросерые почвы в средней и нижней частях склона — к среднегумусированным (3–5%). Почвы характеризуются слабокислой и близкой к нейтральной реакцией среды (рН водной суспензии) и тяжелосуглинистым гранулометрическим составом.

Отбор индивидуальных почвенных проб ($n = 57$) произведен буром из пахотного слоя (0–30 см) по нерегулярной сетке в начале июля 2013 года. Координаты точек опробования определены с помощью системы геопозиционирования (GPS, Garmin eTrex Vista). Почвенные пробы проанализированы на содержание ПОУ методом мокрого озоления по Тюрину (Аринушкина, 1970). Данные по содержанию ПОУ умножены на коэффициент 1.15, который, согласно исследованиям (FAO, 2021), позволяет учесть ошибку, связанную с неполным окислением органического углерода, определяе-

Таблица 1. Преобладающие почвы на территории исследования

Название по классификации почв России (Шишов и др., 2004)	Название по международной классификации почв WRB (IUSS, 2015)
Агрочернозем глинисто-иллювиальный элювиированный среднепахотный сильногумусированный тяжелосуглинистый	Luvic Greyzemic Chernozems (Siltic, Aric, Pachic)
Агрочернозем глинисто-иллювиальный элювиированный среднепахотный среднегумусированный тяжелосуглинистый	Luvic Greyzemic Chernozems (Siltic, Aric, Pachic)
Агрочернозем глинисто-иллювиальный темнаяязыковый среднегумусированный тяжелосуглинистый	Haplic Chernozems (Siltic, Aric, Pachic)
Агротемно-серая типичная среднепахотная среднегумусированная тяжелосуглинистая	Luvic Greyzemic Phaeozems (Siltic, Aric)
Агросерая типичная среднепахотная среднегумусированная тяжелосуглинистая	Greyzemic Phaeozems (Siltic, Aric)

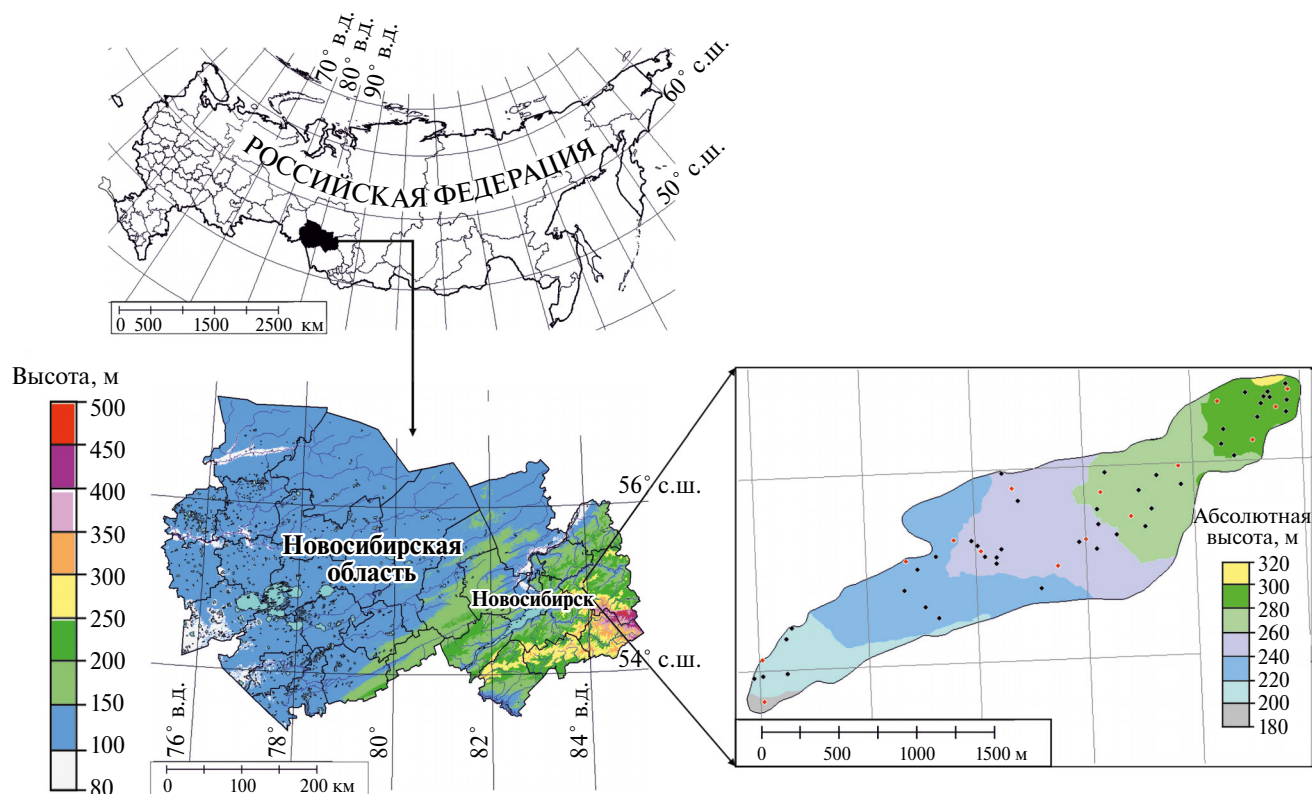


Рис. 1. Карта высот территории исследования и схема отбора почвенных проб. Примечание для рис. 1, 3, 6: черными точками выделен обучающий набор данных ($n = 42$), красными – валидационный ($n = 15$).

мого методом Тюринга. Карта плотности сложения почв рассчитана на основе растровой карты содержания ПОУ с использованием педотрансферной функции (Abdelbaki, 2018):

$$BD = 1.44 \times \exp(-0.03 \times \text{ПОУ}),$$

где BD (г/см^3) – плотность сложения почв; ПОУ (%) – содержание почвенного органического углерода.

Запасы ПОУ (ЗПОУ) рассчитаны по следующей формуле:

$$\text{ЗПОУ} = \text{ПОУ} (\%) \times BD (\text{г/см}^3) \times \text{МС} (\text{см}),$$

где ПОУ (%) – содержание почвенного органического углерода; BD (г/см^3) – плотность сложения почв; МС – мощность слоя (см).

В качестве предикторов в модели случайного леса использовали: 19 биоклиматических переменных WorldClim (Fick, Hijmans, 2017); 5 климатических переменных, рассчитанных на основе WorldClim и данных и почвенно-климатического атласа (Сляднев, 1978); 8 вегетационных индексов, рассчитанных на основе Landsat 8 OLI (снимок от 7 июля 2013 г.); 26 морфометрических характеристик рельефа, рассчитанных на основе цифровой модели высот ALOS World 3D ver. 4.1; 2 переменные, характеризующие пространственное положение (табл. 2). В общем было использовано

60 предикторов, характеризующих такие факторы почвообразования как климат, растительность, рельеф и пространственное положение. Все растровые карты приведены к разрешению 30×30 м.

Обучающий набор данных (ОНД) состоял из 42 точек лабораторно-полевого обследования почв, валидационный (ВНД) – из 15 (рис. 1). Описательная статистика наборов данных по содержанию ПОУ приведена в таблице 3.

Оценку эффективности моделей проводили по следующим показателям: R^2 – коэффициент детерминации; RMSE (англ. Root Mean Square Error) – корень из среднеквадратической ошибки; MAPE (англ. Mean Absolute Percentage Error) – средняя абсолютная ошибка в процентах, MAE (англ. Mean Absolute Error) – средняя абсолютная ошибка.

Корень из среднеквадратической ошибки (RMSE, Root Mean Square Error) рассчитан следующим образом (Hengl, 2007):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=0}^n [\hat{z}(sj) - z(sj)]^2},$$

где n – количество наблюдений в анализируемом наборе данных; $\hat{z}(sj)$ – предсказанные значения; $z(sj)$ – фактические значения в точках проверки.

Таблица 2. Предикторы, используемые для моделирования содержания ПОУ

Название на английском языке	Название на русском языке
Предикторы, характеризующие климат	
BIO1 (Annual Mean Temperature)	Среднегодовая температура
BIO2 (Mean Diurnal Range (Mean of monthly (max temp – min temp)))	Среднесуточная разность (среднее значение за месяц (макс. температура – мин. температура))
BIO3 (Isothermality (BIO2/BIO7) ($\times 100$))	Изотермичность (BIO2/BIO7) ($\times 100$)
BIO4 (Temperature Seasonality (stand. dev. $\times 100$))	Сезонность температуры (стандартное отклонение $\times 100$)
BIO5 (Max Temperature of Warmest Month)	Максимальная температура самого теплого месяца
BIO6 (Min Temperature of Coldest Month)	Минимальная температура самого холодного месяца
BIO7 (Temperature Annual Range (BIO5-BIO6))	Годовой диапазон температур (BIO5-BIO6)
BIO8 (Mean Temperature of Wettest Quarter)	Средняя температура самого влажного квартала
BIO9 (Mean Temperature of Driest Quarter)	Средняя температура самого сухого квартала
BIO10 (Mean Temperature of Warmest Quarter)	Средняя температура самого теплого квартала
BIO11 (Mean Temperature of Coldest Quarter)	Средняя температура самого холодного квартала
BIO12 (Annual Precipitation)	Годовое количество осадков
BIO13 (Precipitation of Wettest Month)	Количество осадков в самый влажный месяц
BIO14 (Precipitation of Driest Month)	Количество осадков в самый засушливый месяц
BIO15 (Precipitation Seasonality (Coefficient of Variation))	Сезонность осадков (коэффициент вариации)
BIO16 (Precipitation of Wettest Quarter)	Количество осадков в самом влажном квартале
BIO17 (Precipitation of Driest Quarter)	Количество осадков в самом сухом квартале
BIO18 (Precipitation of Warmest Quarter)	Количество осадков в самом теплом квартале
BIO19 (Precipitation of Coldest Quarter)	Количество осадков в самом холодном квартале
SradMean	Средняя солнечная радиация (сумма средних значений по месяцам (с 1970 по 2000 гг.), деленная на 12)
SAT10 (Sum of Active Temperatures above 10°C)	Сумма активных температур выше 10°C
CC (Coefficient of Continentality of Climate)	Коэффициент континентальности климата
Evaporation	Испаряемость с поверхности пресных озер
CH (Humidity coefficient)	Коэффициент увлажнения
AgPot (Agroecological potential)	Агроэкологический потенциал (Гопп, 2023)
Предикторы, характеризующие рельеф (Conrad et al., 2015)	
Elevation	Высота над уровнем моря
Slope	Крутизна склонов
LS-Factor	Коэффициент соотношения длины и крутизны склона
Aspect	Экспозиция склонов
CrosSecCurv (Cross-Sectional Curvature)	Кривизна поперечного сечения
FlowLineCurv (Flow Line Curvature)	Кривизна линии потока
GenCurv (General Curvature)	Главная кривизна
LongCurv (Longitudinal Curvature)	Продольная кривизна
MinCurv (Minimal Curvature)	Минимальная кривизна
MaxCurv (Maximal Curvature)	Максимальная кривизна
ProfCurv (Profile Curvature)	Профильная кривизна
TanCurv (Tangential Curvature)	Тангенциальная кривизна
ConvIndex (Convergence Index)	Индекс конвергенции
Texture	Текстурность

Таблица 2. Окончание

Название на английском языке	Название на русском языке
Convexity	Индекс выпуклости
MassBalans (Mass Balance Index)	Индекс баланса массы
AnalitHill (Analytical Hillshading)	Аналитическая затененность холмов
ClosDepr (Closed Depressions)	Бессточные депрессии
VallDepth (Valley Depth)	Глубина долин
TWI (Topographic Wetness Index)	Топографический индекс влажности
TCA (Total Catchment Area)	Общая площадь водосбора
RSP (Relative Slope Position)	Относительное положение на склоне
TPI (Topographic Position Index)	Топографический индекс положения
TRI (Terrain Ruggedness Index)	Индекс расчлененности рельефа
CND (Channel Network Distance)	Расстояние до водотоков (дренажной сети)
CNBL (Channel Network Base Level)	Базовый уровень водотоков
Предикторы, характеризующие пространственное положение	
Longitude	Долгота
Latitude	Широта
Предикторы, характеризующие растительность	
NDVI (Normalized Difference Vegetation Index)	Нормализованный разностный вегетационный индекс
CTVI (Corrected Transformed Vegetation Index)	Скорректированный трансформированный вегетационный индекс
DVI (Difference Vegetation Index)	Разностный вегетационный индекс
RVI (Ratio Vegetation Index)	Относительный вегетационный индекс
NRVI (Normalized Ratio Vegetation Index)	Нормализованный относительный вегетационный индекс
SAVI (Soil Adjusted Vegetation Index)	Вегетационный индекс с коррекцией по почве
TTVI (Tiam's Transformed Vegetation Index)	Трансформированный растительный индекс Тиамы
TVI (Transformed Vegetation Index)	Трансформированный вегетационный индекс

Таблица 3. Описательная статистика по содержанию ПОУ для общего, обучающего и валидационного наборов данных

Набор данных	Стат. показатель	ПОУ, %	Набор данных	Стат. показатель	ПОУ, %	Набор данных	Стат. показатель	ПОУ, %
Общий (n = 57)	Среднее	3.3	Обучающий (n = 42)	Среднее	3.4	Валидационный (n = 15)	Среднее	3.1
	SD	1.3		SD	1.4		SD	1.3
	Минимум	1.2		Минимум	1.3		Минимум	1.2
	Медиана	3.0		Медиана	3.0		Медиана	2.6
	Максимум	6.1		Максимум	6.1		Максимум	5.5
	CV, %	40.2		CV, %	40.6		CV, %	40.1
	Ассиметрия	0.56		Ассиметрия	0.58		Ассиметрия	0.43
	Экспесс	-0.78		Экспесс	-0.80		Экспесс	-0.78

Сокращения: SD – стандартное отклонение; CV – коэффициент вариации.

RMSE представлен в единицах измерения почвенного показателя. Значения RMSE, близкие к нулю, указывают на идеальное соответствие модели фактическим данным. Значение RMSE, меньше половины стандартного отклонения (SD) измеренных

данных, указывает на высокую точность прогноза (Singh et al., 2004).

Средняя абсолютная ошибка в процентах (MAPE, Mean Absolute Percentage Error) рассчитана следующим образом (Афанасьев, Цыпин, 2008):

$$\text{MAPE}(\%) = \frac{1}{n} \sum \frac{|y_t - \bar{y}_t|}{\bar{y}_t} \times 100,$$

где y_t – фактическое значение, $\bar{y}_t$ – предсказанное значение, n – количество наблюдений в анализируемом наборе данных. $|y_t - \bar{y}_t|$ – показывает абсолютное значение (берется без отрицательного знака).

MAPE показывает отклонение предсказанного значения от фактического в среднем по выборке. Точность моделирования оценивается по значениям MAPE следующим образом (Афанасьев, Цыпин, 2008): <10% – высокая; 10–20% – хорошая; 20–50% – удовлетворительная; и >50% – неудовлетворительная точность.

Средняя абсолютная ошибка (MAE, Mean Absolute Error) рассчитана по следующей формуле:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \bar{y}_i|,$$

где y_i – фактическое значение, $\bar{y}_i$ – предсказанное значение, n – количество наблюдений в анализируемом наборе данных. $|y_i - \bar{y}_i|$ – показывает абсолютное значение (берется без отрицательного знака). MAE представлена в единицах измерения исследуемого показателя.

Математические операции с геопривязанными растровыми картами проведены на облачной онлайн-платформе Google Earth Engine (Gorelick et al., 2017) и в программном обеспечении SAGA GIS (Conrad et al., 2015). Карты содержания ПОУ составлены с использованием алгоритма случайного леса (*англ.* Random Forest) на облачной онлайн-платформе Google Earth Engine (200 деревьев, значения гиперпараметров по умолчанию). Карта запасов ПОУ, а также карта плотности сложения почв составлены в SAGA GIS.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Оценку важности предикторов для цифрового картографирования содержания ПОУ провели с использованием следующих подходов: корреляционный и регрессионный анализы, а также расчет показателя важности отдельных переменных для модели случайного леса с различными наборами предикторов.

Корреляционный анализ между шестидесятью предикторами и содержанием ПОУ показал, что тридцать пять предикторов с коэффициентом корреляции больше 0.5 оказались наиболее важными для моделирования содержания ПОУ (табл. 4).

Таблица 4. Коэффициенты корреляции (R) Спирмена между содержанием почвенного органического углерода (ПОУ) и значениями предикторов, характеризующих факторы почвообразования ($p < 0.05$)

Показатели	R	Показатели	R	Показатели	R
BIO1	0.75	SAT10	0.75	MassBalans	–
BIO2	–	CC	-0.58	AnalitHill	–
BIO3	0.34	Evaporation	-0.45	ValDepth	-0.70
BIO4	-0.59	CH	0.73	TWI	–
BIO5	-0.69	AgPot	0.60	TCA	–
BIO6	-0.80	Elevation	0.73	RSP	0.73
BIO7	-0.51	Slope	–	TPI	-0.31
BIO8	-0.72	LS-factor	–	TRI	–
BIO9	-0.81	Aspect	–	CND	0.58
BIO10	-0.72	CrosSecCurv	–	CNBL	0.65
BIO11	-0.83	FlowLineCurv	–	Longitude	0.73
BIO12	0.73	GenCurv	–	Latitude	0.70
BIO13	–	LongiCurv	–	NDVI	0.63
BIO14	0.79	MinCurv	–	CTVI	0.62
BIO15	-0.54	MaxCurv	–	DVI	0.66
BIO16	0.59	ProfCurv	-0.29	NRVI	0.63
BIO17	0.79	TanCurv	–	RVI	0.66
BIO18	0.59	ConvIndex	–	SAVI	0.63
BIO19	–	Texture	0.56	TTVI	0.62
SradMean	-0.48	Convexity	0.41	TVI	0.62

Сокращения в обозначении показателей приведены в таблице 1.

Положительные корреляции заметной (R от 0.5 до 0.7) и высокой (R от 0.7 до 0.9) силы связи выявлены между содержанием ПОУ и следующими предикторами: BIO1, BIO12, BIO14, BIO16–BIO18, SAT10, CH, AgPot, Elevation, Convexity, Texture, RSP, CND, CNBL, Longitude, Latitude, NDVI, CTVI, DVI, NRVI, RVI, SAVI, TTVI, TVI. Отрицательные корреляции заметной (R от -0.5 до -0.7) и высокой (R от -0.7 до -0.9) силы связи выявлены между содержанием ПОУ и следующими предикторами: BIO4–BIO11, BIO15, SradMean, CC, Evaporation, ValDepth. Регрессионный анализ методом пошагового введения переменных с использованием 35-ти предикторов показал, что наибольший коэффициент детерминации ($R^2 = 0.75$) получен при использовании в модели всего двух переменных (BIO11 и RVI). Остальные переменные исключены в связи с эффектом мультиколлинеарности (высокая сила связи между предикторами) или статистической незначимостью коэффициентов в уравнениях регрессии. После исключения переменных BIO11 и RVI из анализа с использованием аналогичного подхода выбраны следующие наборы предикторов: Longitude+CNBL ($R^2=0.65$); SAT10+CC+Texture ($R^2 = 0.69$).

При построении модели случайного леса нет ограничения на количество используемых предикторов, поэтому для сравнительного анализа подготовили еще 6 наборов предикторов: 60 предикторов; 42 (без кривизн рельефа, вегетационных индексов и предикторов с нулевыми значениями); 37 (все с $R > \pm 0.5$); 32 (все с $R > \pm 0.3$ без вегетационных индексов); 27 (все с $R > \pm 0.5$ без вегетационных индексов); 23 (без BIO1–19, кривизн рельефа, вегетационных индексов и предикторов с нулевыми значениями) (табл. 2, рис. 2).

Расчет показателя важности отдельных переменных для модели случайного леса с различными наборами предикторов представлен на рис. 2. Анализ графиков показал, что ранжирование важности предикторов определяется не в соответствии с большими значениями коэффициентов корреляции между исследуемыми показателями (табл. 2, рис. 2). Это может указывать на наличие более значимых зависимостей в отдельных деревьях модели случайного леса, однако эксперту невозможно определить силу их корреляционной зависимости.

Результаты моделирования алгоритмом случайного леса с использованием различных наборов предикторов представлены на рис. 3.

Оценка точности моделирования содержания ПОУ показала, что у моделей с количеством предикторов 37 и 60 коэффициент детерминации ($R^2 = 0.88$) выше по сравнению с другими моделями, однако они не выбраны в качестве лучших в связи с более низкими коэффициентами детерминации, рассчитанными по валидационному набору данных (ВНД), и более высокими значениями RMSE (табл. 5, рис. 4).

Самый высокий R^2 (0.67) по ВНД отмечен у модели с 32 предикторами без вегетационных индексов (рис. 4ж). По всей видимости, использование вегетационных индексов приводит к снижению R^2 в моделях с 37 и 60 предикторами (рис. 4е, з).

Полевые наблюдения и анализ снимка, по которому рассчитаны вегетационные индексы, показали, что на изучаемом участке имелись ареалы с разреженной растительностью (плохие всходы, вредители и др.), поэтому использование этих предикторов в моделях с 37 и 60 предикторами ухудшило качество моделирования содержания ПОУ. Кроме этого, на карте содержания ПОУ, состав-

Таблица 5. Показатели эффективности модели (ЭМ) случайного леса для цифрового картографирования содержания ПОУ

Набор переменных в модели случайного леса, количество предикторов (рис. 2)	Показатели ЭМ			
	R^2	RMSE, %	MAPE, %	MAE, %
BIO11+RVI	0.83	0.96	33.5	0.77
Longitude + CNBL	0.82	0.78	22.8	0.61
SAT10+CC+Texture	0.82	0.85	28.4	0.67
60	0.88	0.78	27.2	0.67
42	0.86	0.72	22.2	0.58
37	0.88	0.83	27.9	0.68
32	0.87	0.72	22.1	0.57
27	0.87	0.74	22.3	0.58
23	0.85	0.78	24.1	0.62

Сокращения: R^2 – коэффициент детерминации (рассчитано по ОНД); RMSE – корень из среднеквадратической ошибки; MAPE – средняя абсолютная ошибка в процентах; MAE – средняя абсолютная ошибка. RMSE, MAPE и MAE рассчитаны по ВНД.

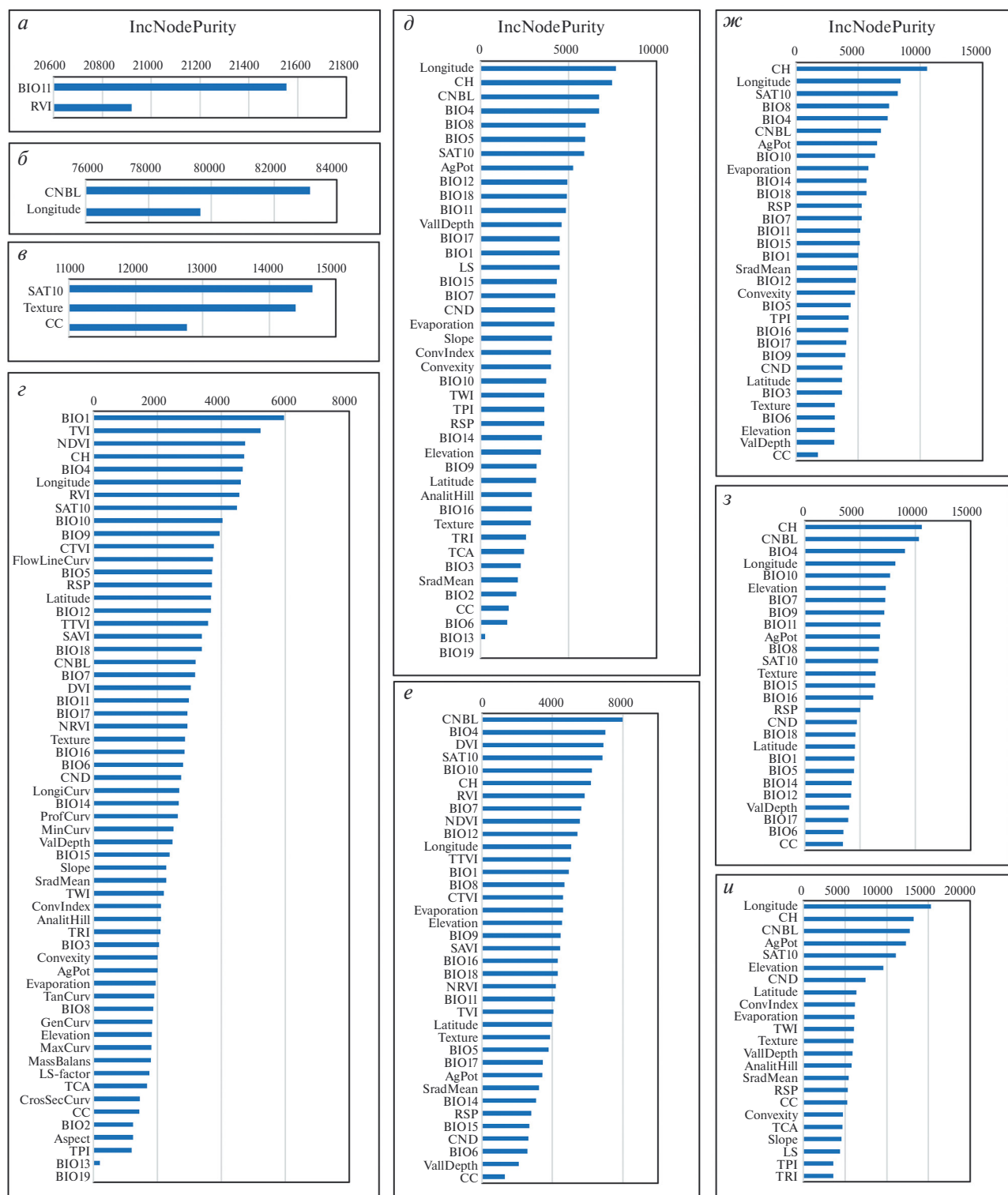


Рис. 2. Показатель важности (IncNodePurity) отдельных переменных в модели случайного леса с наборами предикторов: *а* – BIO11+RVI; *б* – Longitude+CNBL; *в* – SAT10+CC+Texture; *г* – 60 предикторов; *д* – 42 (без кривизн рельефа, вегетационных индексов и предикторов с нулевыми значениями); *е* – 37 (все с $R > \pm 0.5$); *ж* – 32 (все с $R > \pm 0.3$ без вегетационных индексов); *з* – 27 (все с $R > \pm 0.5$ без вегетационных индексов); *и* – 23 (без BIO1–19, кривизн рельефа, вегетационных индексов и предикторов с нулевыми значениями).

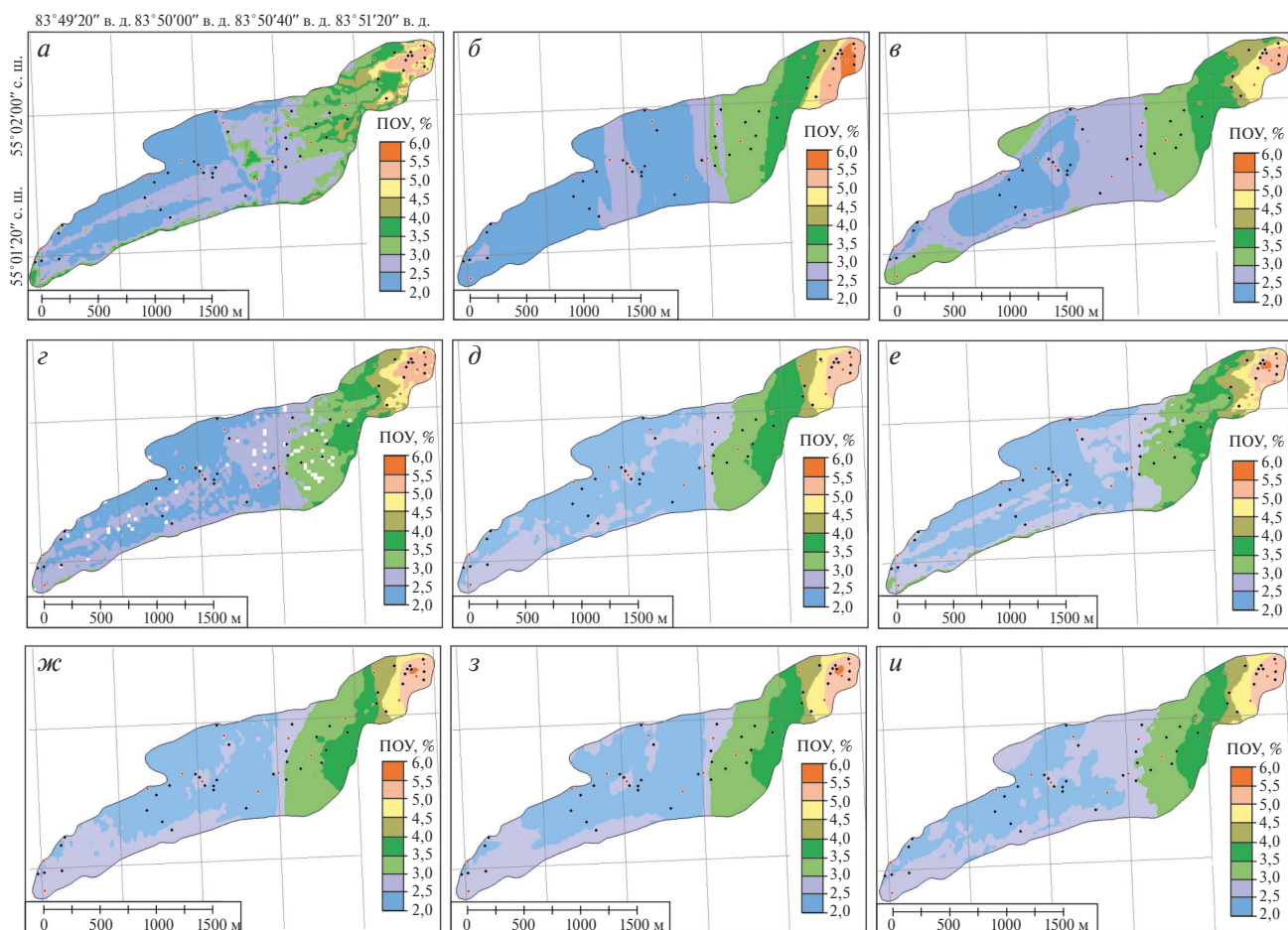


Рис. 3. Карты пространственной изменчивости содержания ПОУ, составленные алгоритмом случайного леса на основе различных наборов предикторов: *a* – BIO11+RVI; *б* – Longitude+CNBL; *в* – SAT10+CC+Texture; *г* – 60 предикторов; *д* – 42 (без кривизн рельефа, вегетационных индексов и предикторов с нулевыми значениями); *е* – 37 (все с $R > \pm 0.5$); *ж* – 32 (все с $R > \pm 0.3$ без вегетационных индексов); *з* – 27 (все с $R > \pm 0.5$ без вегетационных индексов); *и* – 23 (без BIO1–19, кривизн рельефа, вегетационных индексов и предикторов с наличием нулевых значений в растре).

ленной на основе 60 предикторов, присутствовали белые пиксели (неизвестные значения), появившиеся в результате использования различных кривизн рельефа и экспозиции склонов, в растровых картах которых присутствовали нулевые значения (рис. 3г).

Различие средних значений между предсказанным и фактическим содержанием ПОУ в обучающем и валидационном наборах данных статистически незначимо (рис. 5а, б).

Сравнение фактического содержания ПОУ с предсказанным моделью случайного леса на различных высотных ступенях показало удовлетворительное совпадение по средним значениям (рис. 6а, б). Предсказанные значения характеризовались меньшим диапазоном варьирования. В почвах первой высотной ступени содержание ПОУ было больше в 1.4–2.1 раза по сравнению с осталь-

ными (рис. 6а). Наибольшие значения содержания ПОУ выявлены в почвах верхней части склона с постепенным их снижением вниз по склону (рис. 6).

Карта содержания ПОУ, составленная с использованием 32 предикторов, с более высоким коэффициентом детерминации (0.67) по ВНД и более низкими значениями RMSE (0.72) и MAE (0.57) выбрана в качестве лучшей для дальнейшего моделирования запасов ПОУ (табл. 4, рис. 3ж, рис. 4ж). Согласно MAPE (22.1%), точность моделирования содержания ПОУ удовлетворительная (табл. 3).

В связи с отсутствием достаточного количества фактических данных по плотности сложения почв, геопространственное моделирование этого показателя выполнено с использованием карты содержания ПОУ и педотрансферной функции (рис. 7а).

Различия между предсказанной и фактической плотностью сложения агротемно-серой почвы

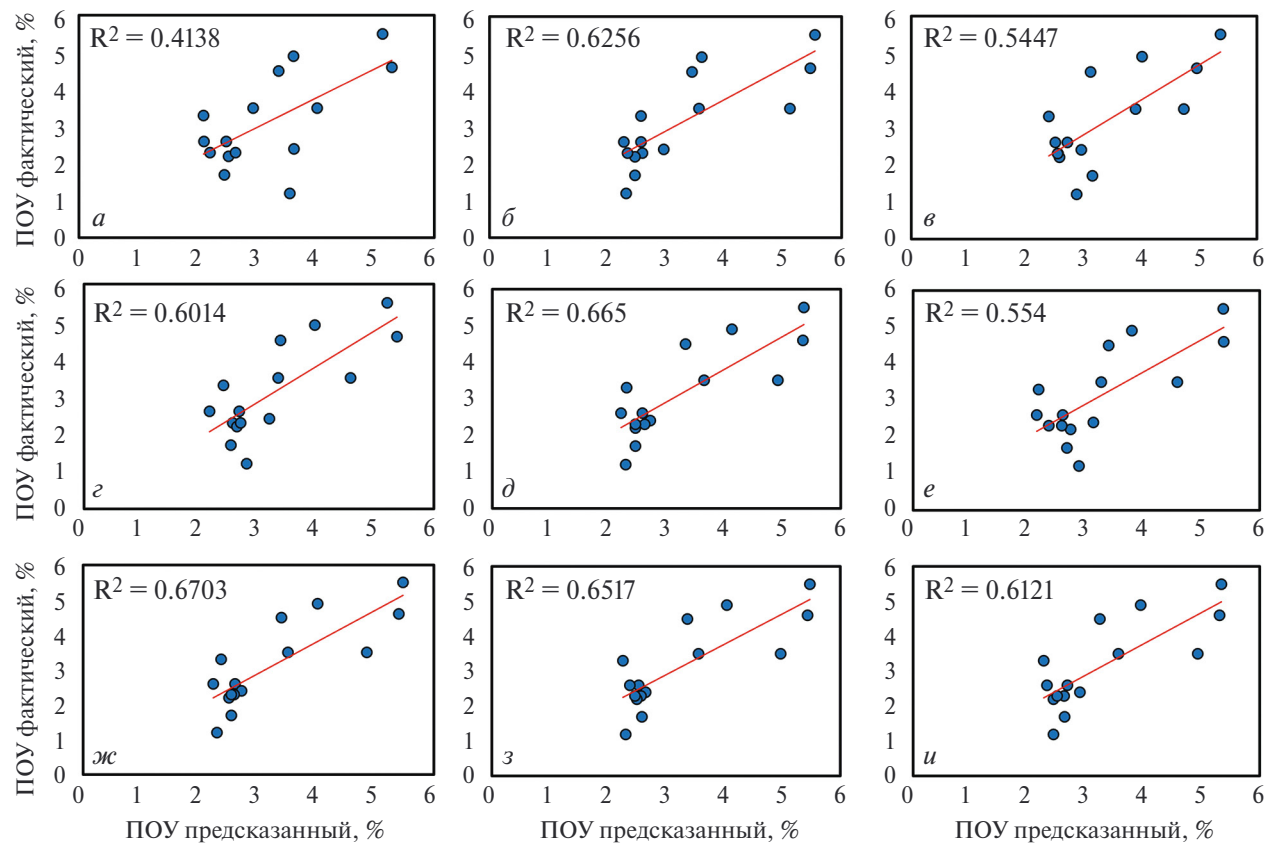


Рис. 4. Зависимость между фактическим и предсказанным содержанием ПОУ в валидационном наборе данных (n = 15) для наборов предикторов: а – BIO11+RVI; б – Longitude+CNBL; в – SAT10+CC+Texture; г – 60 предикторов; д – 42 (без кривизн рельефа, вегетационных индексов и предикторов с нулевыми значениями); е – 37 (все с $R > \pm 0.5$); ж – 32 (все с $R > \pm 0.3$ без вегетационных индексов); з – 27 (все с $R > \pm 0.5$ без вегетационных индексов); и – 23 (без BIO1–19, кривизн рельефа, вегетационных индексов и предикторов с нулевыми значениями).

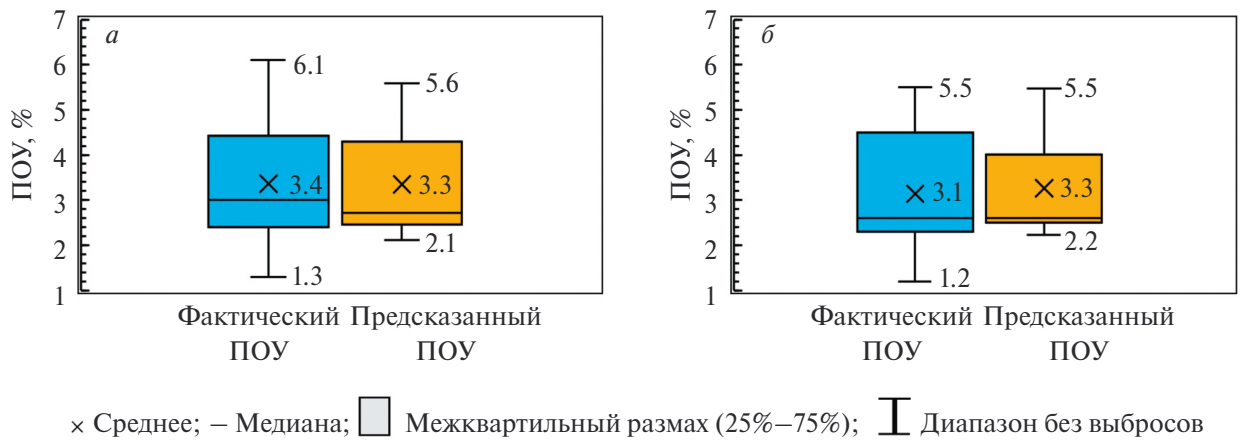


Рис. 5. Сравнение средних значений ПОУ (фактических и предсказанных) по обучающему (а) и валидационному (б) наборам данных (карта содержания ПОУ на основе 32 предикторов).

с различной степенью смывости почвы представлены в таблице 6.

RMSE составил 0.067 г/см³, а MAPE – 5%, что соответствует высокой точности моделирования плотности сложения почв. Используемая в настоящем исследовании педостранферная функция

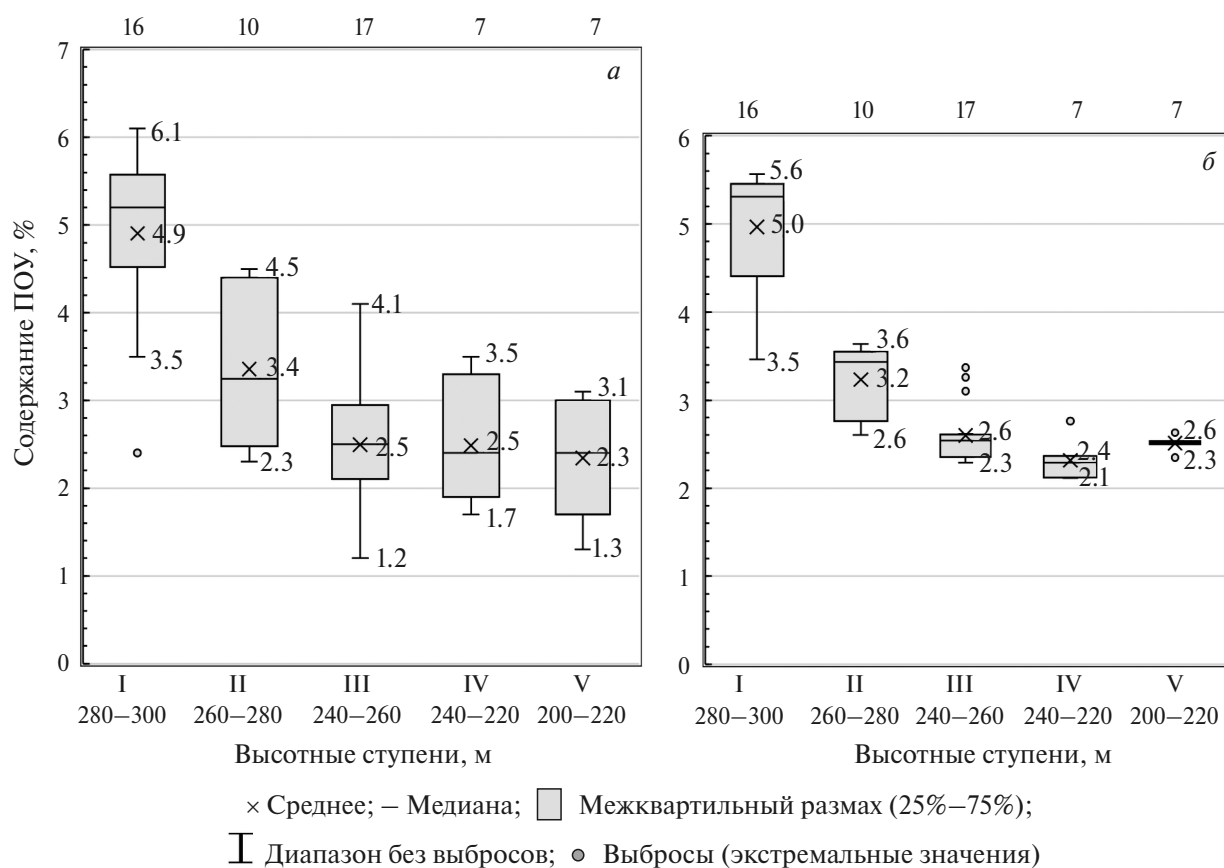


Рис. 6. Фактическое содержание ПОУ (а) и предсказанное моделью случайного леса (б) на различных высотных ступенях исследуемой территории (рис. 1). Примечание: объем выборки указан сверху над диаграммами размаха.

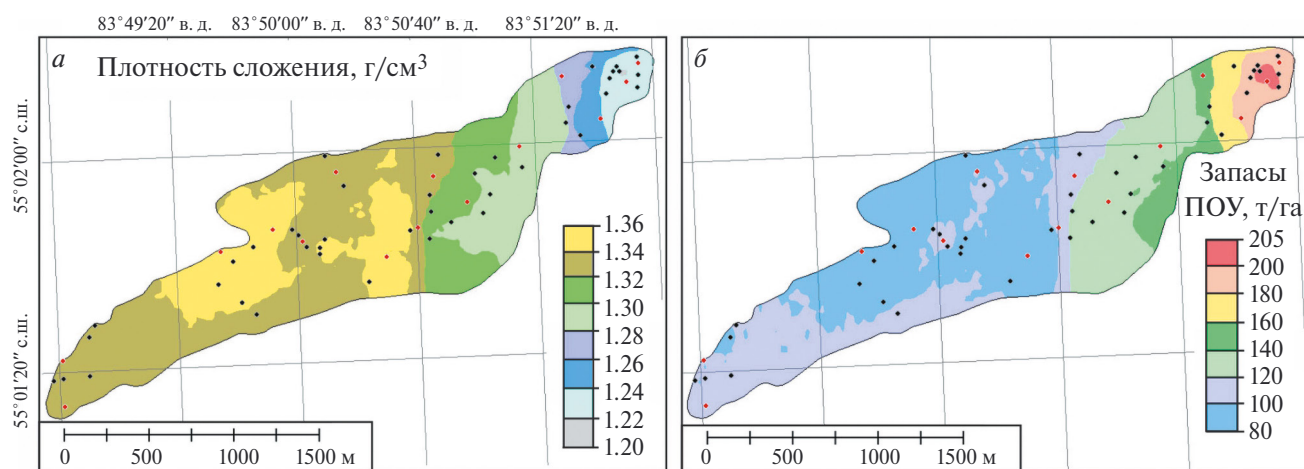


Рис. 7. Карты пространственной изменчивости: а – плотности сложения почв в слое 0–30 см; б – запасов ПОУ в слое 0–30 см.

Таблица 6. Фактическая и предсказанная плотность сложения агротемно-серой типичной почвы разной степени смытости

Степень смытости почв	Фактическая плотность, г/см ³	Предсказанная плотность, г/см ³
Несмытая	1.30	1.33
Слабосмытая	1.26	1.33
Среднесмытая	1.25	1.33

была разработана на основе географической базы данных обследований почв США (Abdelbaki, 2018). Однако расчет ошибки моделирования показал, что применение этой функции для почв исследуемой территории можно считать приемлемым.

На основе карт содержания ПОУ и плотности сложения почв проведено картографирование запасов ПОУ (рис. 7б), закономерности пространственного варьирования которого были аналогичны содержанию ПОУ (рис. 8а, рис. 6а). Значения по плотности сложения почв (предсказанные) варьировали в диапазоне от 1.22 до 1.35 г/см³ и увеличивались вниз по склону, т.е. имели обратный тренд распределения по сравнению с содержанием и запасами ПОУ (рис. 8б).

Растровая карта исследуемой территории площадью 225 га состоит из 2768 пикселей размером 30 × 30 м с определенным значением запасов ПОУ в каждом элементе в расчете на гектар. Расчеты показали, что общие запасы ПОУ в пахотном слое (0–30 см) почв исследуемой территории площадью 225 га составили 28721.14 т или 28.72 кт.

ЗАКЛЮЧЕНИЕ

Предсказанные алгоритмом случайного леса значения по содержанию ПОУ характеризовались меньшим диапазоном варьирования по сравнению с фактическими данными. Использование алгоритма случайного леса, реализованного на онлайн-платформе Google Earth Engine, а также набора геопространственных данных, состоящих из 32 предикторов без вегетационных индексов, позволило установить специфику пространственной изменчивости таких показателей как содержание и запасы ПОУ, а также плотности сложения почв. Отличие средних значений между предсказанным и фактическим содержанием ПОУ в обучающем и валидационном наборах данных статистически незначимо.

Содержание ПОУ в слое (0–30 см) варьировало от 1.3 до 6.1% по фактическим данным, а запасы ПОУ — от 84 до 203 т/га. Наибольшее содержание и запасы ПОУ выявлены в почвах верхней части склона (первая высотная ступень (280–300 м)), тогда как вниз по склону отмечено постепенное снижение значений этих показателей в 1.4–2.1 раза. Плотность сложения почв (предсказанные значения) варьировала в диапазоне от 1.20 до 1.36 г/см³

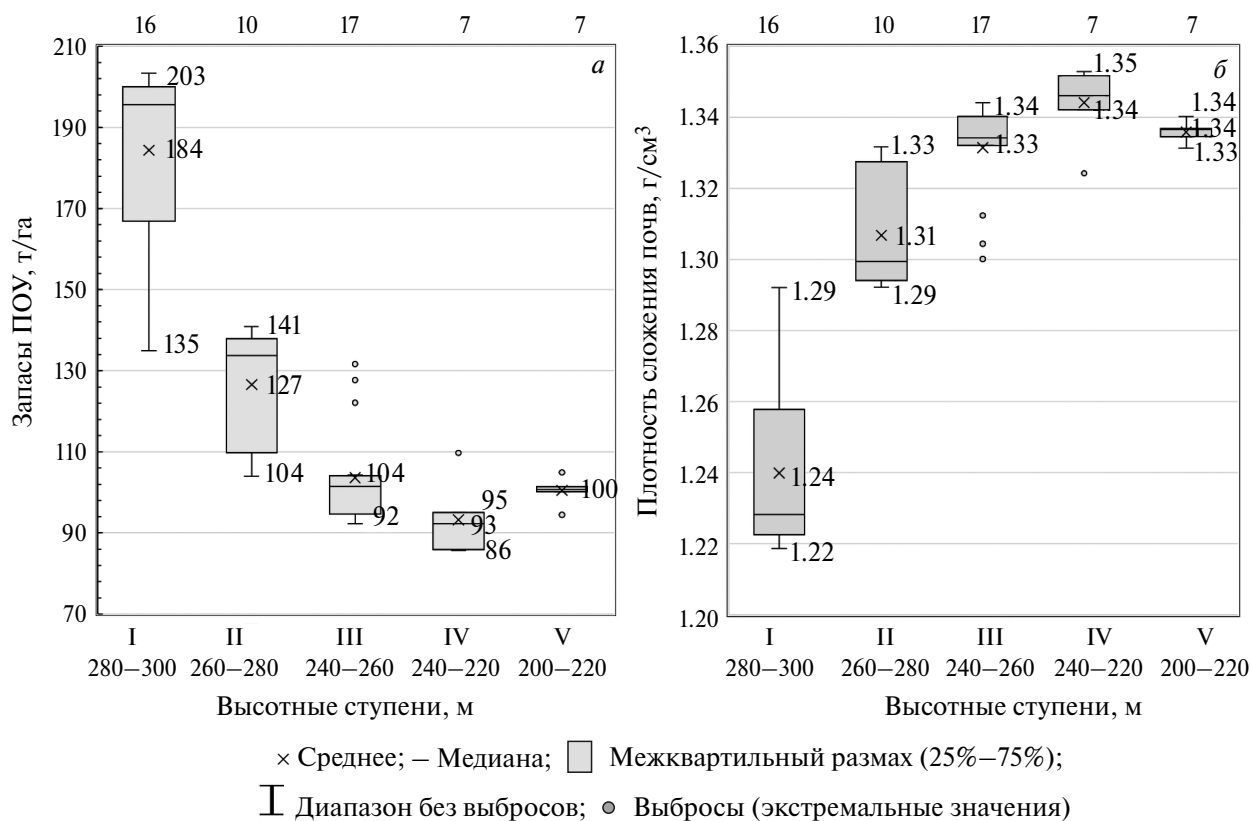


Рис. 8. Запасы ПОУ (а) и плотность сложения почв (б) на разных высотных ступенях исследуемой территории (см. рис. 1). Примечание: объем выборки указан сверху над диаграммами размаха.

и увеличивалась вниз по склону, т.е. имела обратный тренд распределения по сравнению с содержанием и запасами ПОУ. Общие запасы ПОУ в пахотном слое (0–30 см) почв исследуемой территории площадью 225 га составили 28.7 кт.

Специфику пространственной изменчивости содержания и запасов ПОУ необходимо учитывать при разработке адаптивно-ландшафтных систем земледелия и агротехнологий с целью эффективного использования почвенных ресурсов, а также в исследованиях по изменению климата.

БЛАГОДАРНОСТИ

Авторы выражают благодарность к.б.н. Савенкову О.А. и к.б.н. Смирновой Н.В. за помощь в проведении полевых работ, Черепяхиной Л.Д. за определение содержания ПОУ в почвенных образцах.

ИСТОЧНИК ФИНАНСИРОВАНИЯ

Работа выполнена по государственному заданию ИПА СО РАН при финансовой поддержке Министерства науки и высшего образования Российской Федерации.

СПИСОК ЛИТЕРАТУРЫ

- Ариушикина Е.В. Руководство по химическому анализу почв. М.: Изд-во МГУ, 1970. 487 с.
- Афанасьев В.Н., Цыпин А.П. Эконометрика в пакете STATISTICA: учебное пособие по выполнению лабораторных работ. Оренбург: ГОУ ОГУ, 2008. 204 с.
- Гонн Н.В. Агроэкологический потенциал западной части Кузнецко-Салаирской геоморфологической провинции: методика цифрового картографирования, геопрограммный анализ, корреляция с содержанием органического углерода в почвах // Почвы и окружающая среда. 2023. 6(3). e224. <https://doi.org/10.31251/pos.v6i3.224>
- Сляднев А.П. Почвенно-климатический атлас Новосибирской области. Новосибирск: Наука, 1978. 122 с.
- Шишов Л.Л., Тонконогов В.Д., Лебедева И.И., Герасимова М.И. Классификация и диагностика почв России. Смоленск: Ойкумена, 2004. 342 с.
- Abdelbaki A.M. Evaluation of pedotransfer functions for predicting soil bulk density for U.S. soils. *Ain Shams Engineering Journal*. 2018. 9(4). P. 1611–1619. <https://doi.org/10.1016/j.asej.2016.12.002>
- Breiman L. Random forests. *Machine learning*. 2001. 45(1). P. 5–32. <https://doi.org/10.1023/A:1010933404324>
- Conrad O., Bechtel B., Bock M., Dietrich H., Fischer E., Gerlitz L., Wehberg J., Wichmann V., Böhner J. System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*. 2015. 8(7). P. 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Cutler D.R., Edwards T.C. Jr., Beard K.H., Cutler A., Hess K.T., Gibson J., Lawler J.J. Random forests for classification in Ecology. *Ecology*. 2007. 88. P. 2783–2792. <https://doi.org/10.1890/07-0539.1>
- Dharumarajan S., Hegde Rajendra, Singh S.K. Spatial prediction of major soil properties using Random Forest techniques — A case study in semi-arid tropics of South India // *Geoderma Regional*. 2017. 10. P. 154–162. <https://doi.org/10.1016/j.geodrs.2017.07.005>
- FAO. Standard operating procedure for soil organic carbon: Tyurin spectrophotometric method. Rom: FAO, 2021. <https://www.fao.org/3/cb4757en/cb4757en.pdf>
- FAO and ITPS. Global Soil Organic Carbon Map V1.5: Technical report. Rome, FAO, 2020. 169 p.
- Fick S.E., Hijmans R.J. WorldClim 2: new 1 km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*. 2017. 37(12). P. 4302–4315. <https://doi.org/10.1002/joc.5086>
- Gandhi U. JavaScript and the Earth Engine API. In: Cardille J.A., Crowley M.A., Saah D., Clinton N.E. (eds) *Cloud-Based Remote Sensing with Google Earth Engine*. Springer, Cham, 2024. https://doi.org/10.1007/978-3-031-26588-4_1
- Gorelick N., Hancher M., Dixon M., Ilyushchenko S., Thau D., Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. 2017. 202. P. 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Grimm R., Behrens T., Märker M., Elsenbeer H. Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis // *Geoderma*. 2008. 146(1–2). P. 102–113. <https://doi.org/10.1016/j.geoderma.2008.05.008>
- Hengl T.A. Practical guide to geostatistical mapping of environmental variables. EC JRC, Ispra (Italy), 2007. 165 p.
- IUSS. Working Group WRB. World Reference Base for Soil Resources 2014. International soil classification system for naming soils and creating legends for soil maps. Update 2015. *World Soil Resources Reports No. 106*. FAO, Rome, 2015. <https://www.fao.org/3/i3794en/i3794en.pdf>
- McBratney A.B., Mendonça Santos M.L., Minasny B. On digital soil mapping. *Geoderma*. 2003. 117. P. 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Singh J., Knapp H.V., Demissie M. Hydrologic modelling of the Iroquois River watershed using HSPF and SWAT. *Illinois State Water Survey Contract Report 2004-08*. Illinois State Water Survey, Champaign. 2004. <http://www.isws.illinois.edu/pubdoc/CR/ISWSCR2004-08.pdf>
- Sreenivas K., Dadhwal V.K., Kumar S., Harsha G.Sri, Mitran T., Sujatha G., Janaki Rama Suresh G., Fyze M.A., Ravisankar T. Digital mapping of soil organic and inorganic carbon status in India // *Geoderma*. 2016. 269. P. 160–173. <https://doi.org/10.1016/j.geoderma.2016.02.002>
- Suleymanov A., Abakumov E., Alekseev I., Nizamutdinov T. Digital mapping of soil properties in the high latitudes of Russia using sparse data. *Geoderma Regional*. 36. 2024. e00776. <https://doi.org/10.1016/j.geodrs.2024.e00776>
- Vägen T.G., Winowiecki L.A., Tondoh J.E., Desta L.T., Gumbricht T. Mapping of soil properties and land degradation risk in Africa using MODIS reflectance // *Geoderma*. 2016. 263. P. 216–225. <https://doi.org/10.1016/j.geoderma.2015.06.023>

Digital Mapping of Organic Carbon Content and Stocks in Soils of Cis-Salair Drained Plain Using the Google Earth Engine Online Platform and the Random Forest Algorithm

N. V. Gopp¹, T. V. Nechaeva¹

¹*Institute of Soil Science and Agrochemistry SB RAS, Novosibirsk, Russia*

For a key site on the Cis-Salair drained plain, digital mapping of the content of soil organic carbon (SOC) in the topsoil (0–30 cm) was conducted using the random forest algorithm implemented on the Google Earth Engine cloud platform. The following were used as predictors in the random forest model: 1) 19 bioclimatic variables from WorldClim; 2) 5 climatic variables calculated based on WorldClim data and soil-climate atlas data; 3) 8 vegetation indices calculated based on Landsat 8 OLI images; 4) 26 morphometric characteristics of the terrain calculated based on the ALOS DEM; 5) 2 variables describing the spatial location. The correlation coefficients (R) between the content of SOC and the values of the predictors were taken into account when forming sets of predictors: 1) BIO11+RVI; 2) Longitude+CNBL; 3) SAT10+CC+Texture; 4) 60 predictors; 5) 42 (without relief curvatures, vegetation indices and predictors with zero values); 6) 37 (all with $R > \pm 0.5$); 7) 32 (all with $R > \pm 0.3$ without vegetation indices); 8) 27 (all with $R > \pm 0.5$ without vegetation indices); 9) 23 (without BIO1–19, relief curvatures, vegetation indices and predictors with zero values). The result of modeling the content of SOC based on 32 predictors and a training dataset ($n=42$) with a lower RMSE (0.72) was chosen as the best. Based on this model, a soil bulk density map was compiled using a pedotransfer function. This data, together with a map of the SOC content, was used to create a map of SOC stock. The SOC content in the arable layer (0–30 cm) varied from 1.3 to 6.1%, according to the actual data. The SOC stocks ranged from 84 t/ha to 203 t/ha. The highest levels of SOC content and stocks were found in the soils in the upper part of the slope. A gradual decrease in these values was noted as one moved downhill. The soil bulk density ranged from 1.20 g/cm³ to 1.36 g/cm³ and increased as one moved downhill, indicating a reverse trend compared to the SOC content and stocks. The total SOC stock in the arable layer (0–30 cm) of the soils of the studied territory with an area of 225 hectares amounted to 28.7 kt.

Keywords: climate, relief, chernozems, gray forest soils, SAGA GIS, Random Forest, machine learning

REFERENCES

- Abdelbaki A.M.* Evaluation of pedotransfer functions for predicting soil bulk density for U.S. soils. *Ain Shams Engineering Journal*. 2018. 9(4). P. 1611–1619. <https://doi.org/10.1016/j.asej.2016.12.002>
- Arinushkina E.V.* Rukovodstvo po himicheskomu analizu pochv [Guidelines for chemical analysis of soils]. M.: Izd-vo MGU, 1970. 487 p. (In Russian).
- Afanas'ev V.N., Tsypin A.P.* Ekonometrika v pakete STATISTICA: uchebnoe posobie po vypolneniyu laboratornyh rabot [Econometrics in STATISTICA Software: Manual for Laboratory Practicum]. Orenburg, Orenburg State Univ., 2008 204 p. (In Russian).
- Breiman L.* Random forests. *Machine learning*. 2001. 45(1). P. 5–32. <https://doi.org/10.1023/A:1010933404324>
- Conrad O., Bechtel B., Bock M., Dietrich H., Fischer E., Gerlitz L., Wehberg J., Wichmann V., Böhner J.* System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*. 2015. 8(7). P. 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Cutler D.R., Edwards T.C. Jr., Beard K.H., Cutler A., Hess K.T., Gibson J., Lawler J.J.* Random forests for classification in Ecology. *Ecology*. 2007. 88. P. 2783–2792. <https://doi.org/10.1890/07-0539.1>
- Dharumarajan S., Hegde Rajendra, Singh S.K.* Spatial prediction of major soil properties using Random Forest techniques — A case study in semi-arid tropics of South India // *Geoderma Regional*. 2017. 10. P. 154–162. <https://doi.org/10.1016/j.geoder.2017.07.005>
- FAO. Standard operating procedure for soil organic carbon: Tyurin spectrophotometric method. Rom: FAO, 2021. <https://www.fao.org/3/cb4757en/cb4757en.pdf>
- FAO and ITPS. Global Soil Organic Carbon Map V1.5: Technical report. Rome, FAO, 2020. 169 p.
- Fick S.E., Hijmans R.J.* WorldClim 2: new 1 km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*. 2017. 37(12). P. 4302–4315. <https://doi.org/10.1002/joc.5086>
- Gandhi U.* JavaScript and the Earth Engine API. In: *Cardille J.A., Crowley M.A., Saah D., Clinton N.E. (eds) Cloud-Based Remote Sensing with Google Earth Engine*. Springer, Cham, 2024. https://doi.org/10.1007/978-3-031-26588-4_1
- Gopp N.V.* Agroekologicheskij potencial zapadnoj chasti Kuznetso-Salairskoj geomorfologicheskoy provincii: metodika cifrovogo kartografirovaniya, geoprostanstvennyj analiz, korrelyaciya s sodержaniem organicheskogo ugleroda v pochvah [Agroecological potential of the western part of the Kuznetsk-Salair geomorphological province: digital mapping methodology, geospatial analysis, correlation with soil organic carbon content] // *Počvy i okružajúša sreda*. 2023. 6(3). e224. <https://doi.org/10.31251/pos.v6i3.224> (In Russian).
- Gorelick N., Hancher M., Dixon M., Ilyushchenko S., Thau D., Moore R.* Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. 2017. 202. P. 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>

- Grimm R., Behrens T., Märker M., Elsenbeer H.* Soil organic carbon concentrations and stocks on Barro Colorado Island – Digital soil mapping using Random Forests analysis // *Geoderma*. 2008. 146(1–2). P. 102–113. <https://doi.org/10.1016/j.geoderma.2008.05.008>
- Hengl T.A.* Practical guide to geostatistical mapping of environmental variables. EC JRC, Ispra (Italy), 2007. 165 p.
- IUSS. Working Group WRB. World Reference Base for Soil Resources 2014. International soil classification system for naming soils and creating legends for soil maps. Update 2015. World Soil Resources Reports No. 106. FAO, Rome, 2015. <https://www.fao.org/3/i3794en/I3794en.pdf>
- McBratney A.B., Mendonça Santos M.L., Minasny B.* On digital soil mapping. *Geoderma*. 2003. 117. P. 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Shishov L.L., Tonkonogov V.D., Lebedeva I.I., Gerasimova M.I.* Klassifikaciya i diagnostika pochv Rossii [Classification and Diagnostic System of Russian Soils]. Smolensk, Oikumena, 2004. 342 p. (In Russian).
- Singh J., Knapp H.V., Demissie M.* Hydrologic modelling of the Iroquois River watershed using HSPF and SWAT. Illinois State Water Survey Contract Report 2004-08. Illinois State Water Survey, Champaign. 2004. <http://www.isws.illinois.edu/pubdoc/CR/ISWSCR2004-08.pdf>
- Slyadnev A.P.* Pochvenno-klimaticheskij atlas Novosibirskoj oblasti [Soil and climate atlas of the Novosibirsk region]. Novosibirsk: Nauka, 1978. 122 p. (In Russian).
- Sreenivas K., Dadhwal V.K., Kumar S., Harsha G.Sri, Mitran T., Sujatha G., Janaki Rama Suresh G., Fyze M.A., Ravisankar T.* Digital mapping of soil organic and inorganic carbon status in India // *Geoderma*. 2016. 269. P. 160–173. <https://doi.org/10.1016/j.geoderma.2016.02.002>
- Suleymanov A., Abakumov E., Alekseev I., Nizamutdinov T.* Digital mapping of soil properties in the high latitudes of Russia using sparse data. *Geoderma Regional*. 36. 2024. e00776. <https://doi.org/10.1016/j.geodrs.2024.e00776>
- Vågen T.G., Winowiecki L.A., Tondoh J.E., Desta L.T., Gumbrecht T.* Mapping of soil properties and land degradation risk in Africa using MODIS reflectance // *Geoderma*. 2016. 263. P. 216–225. <https://doi.org/10.1016/j.geoderma.2015.06.023>