

Цифровое картографирование показателей, определяющих сорбционные свойства почв по отношению к поллютантам, по данным дистанционного зондирования Земли с применением машинного обучения

К.Г. Гиниятуллин*, И.А. Сахабиев, Е.В. Смирнова, И.А. Уразметов, Р.В. Окунев, К.А. Гордеева
Казанский (Приволжский) федеральный университет, Казань, Россия

По данным дистанционного зондирования Земли (ДЗЗ) проведено сравнение точности пространственного прогноза почвенных показателей, определяющих сорбционные свойства по отношению к поллютантам. Для построения пространственных карт изменения свойств почвы использовались методы машинного обучения на основе моделей регрессии опорных векторов (SVMr – support vector machine regression) и случайного леса (RF – random forest). Показано, что методы машинного моделирования с использованием ДЗЗ могут быть успешно использованы для пространственного прогноза содержания гранулометрических фракций, органического вещества, pH и емкости катионного обмена почв на участках небольшой площади. Выявлено, что пространственный прогноз содержания фракции пыли наилучшим образом моделируется с помощью алгоритма RF, тогда как остальные свойства почв, способные определить их сорбционный потенциал по отношению к поллютантам, лучше моделируются с помощью метода SVMr. В целом, оба метода машинного обучения дают близкие результаты пространственного прогноза.

Ключевые слова: сорбционные свойства почвы, пространственный прогноз, данные дистанционного зондирования Земли, методы машинного обучения

Для цитирования: Гиниятуллин К.Г., Сахабиев И.А., Смирнова Е.В., Уразметов И.А., Окунев Р.В., Гордеева К.А. (2022). Цифровое картографирование показателей, определяющих сорбционные свойства почв по отношению к поллютантам, по данным дистанционного зондирования Земли с применением машинного обучения. *Георесурсы*, 24(1), с. 84–92. DOI: <https://doi.org/10.18599/grs.2022.1.8>

Введение

Важной проблемой функционирования и динамического развития современной горнодобывающей и перерабатывающей минеральное сырье промышленности является необходимость разработки методов объективной оценки результатов систематического или аварийного загрязнения сопредельных сред опасными соединениями. Почва – один из самых уязвимых компонентов окружающей среды, подверженных в первую очередь загрязнению нефтепродуктами и тяжелыми металлами, которые относятся к группе наиболее опасных почвенных поллютантов (Mishra, 2016). Для предприятий данной отрасли построение карт систематического загрязнения и прогнозирование масштабов аварийного загрязнения прилегающих территорий является актуальной задачей.

В последние три десятилетия для объективной пространственной оценки загрязнения почв различными поллютантами широко используются геостатистические подходы (Platenburg et al., 1988; Juang, et al., 2008; Lin et al., 2010; Milillo et al., 2012). Считается, что геостатистические методы представляют полезный инструмент для

исследования пространственной структуры загрязнителей и объективного описания степени и площади загрязнения почв (Einaх et al., 1995). Вместе с тем, интерполяция пространственных данных загрязнения почв с применением традиционных подходов кригинга может иметь ряд недостатков. Объективное разграничение загрязненных и незагрязненных участков на основе кригинга может быть затруднено вариацией оценки содержания поллютанта (например, эффектом сглаживания дисперсии) (Juang et al., 2008; Ha et al., 2014). Риски неправильного определения также могут быть связаны с недостаточной детализацией и необоснованной методикой отбора почвенных проб (Andronikov et al., 2000; Hooda et al., 2005). Недостатки кригинга могут привести как к ложноположительной, так и к ложноотрицательной оценке степени загрязненности в местах предсказания, что создает серьезные финансовые и экологические риски из-за завышенной или заниженной оценки (Cui et al., 2016). Отбор большего количества проб для получения более точных моделей поверхностного распределения загрязнителей требует слишком высоких затрат времени, денег и является часто не оправданным (Andronikov et al., 2000; Cui et al., 2016). Поэтому в последнее время при построении пространственных моделей все большее внимание уделяется методам, позволяющим при ограниченном количестве точек опробования добиться необходимой точности прогноза за счет использования более совершенных и сложных интерполяторов, использующих

* Ответственный автор: Камиль Гашикович Гиниятуллин
e-mail: ginijatullin@mail.ru

© 2022 Коллектив авторов

Контент доступен под лицензией Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

вспомогательные переменные. В работе (Lin et al., 2011) показано, что применение методов, учитывающих дополнительные предикторы при использовании логистической регрессии и регрессионного кригинга, обеспечивает более надежную оценку опасности загрязнения почв тяжелыми металлами для информационного обеспечения мониторинга, чем пространственные вероятностные модели, полученные на основе индикаторного кригинга. Комбинация многомерной статистики и геостатистический анализ также успешно применяются для определения характеристик загрязнения почв тяжелыми металлами (Saby et al., 2009; Yuan et al., 2013).

В последнее время при моделировании и прогнозе загрязнения почв уделяется много внимания использованию методов машинного обучения, а также различных гибридных методов, позволяющих существенно увеличить надежность пространственного прогноза (Lv et al., 2013; Tarasov et al., 2018; Sergeev et al., 2021; Shi et al., 2021; Sakizadeh et al., 2021). В работе (Cho et al., 2011) выполнена оценка предсказательной эффективности четырех различных моделей: множественной линейной регрессии, однокомпонентной регрессии, искусственной нейронной сети и сочетания однокомпонентной регрессии с искусственной нейронной сетью для прогнозирования концентрации мышьяка. Показано, что среди четырех различных моделей точность предсказания последней является наилучшей. В работе (Sergeev et al., 2019) проведено моделирование нелинейного крупномасштабного тренда с использованием искусственной нейронной сети с последующим моделированием остатков геостатистическими методами, что позволило добиться повышения точности прогноза загрязнения приповерхностного слоя почвы тяжелыми металлами по сравнению с базовыми моделями искусственной нейронной сети. В обзорной работе (Al-Ruzouq et al., 2020) подробно рассматриваются вопросы использования классических и современных методов машинного обучения для оценки нефтяного загрязнения различных природных сред, включая почвы, с применением технологий дистанционного зондирования Земли (ДЗЗ). Несмотря на все разнообразие используемых машинных и гибридных методов, в конечном счете, они базируются на использовании предикторов, которые прямо или косвенно влияют на пространственное перераспределение загрязняющих веществ по поверхности Земли, а также в пределах почвенного покрова. В качестве предикторов могут использоваться не только результаты полевого изучения почв или материалы ДЗЗ, но и данные цифровых карт изменения почвенных свойств, в том числе полученные с применением современных геостатистических методов и методов машинного обучения.

Основными первичными показателями, определяющими сорбционные свойства почв по отношению к различным классам поллютантов, являются литологический и гранулометрический состав (ГМС), содержание органического вещества, pH и емкость катионного обмена (ЕКО). Данные показатели являются определяющими для процессов аккумуляции (перемещения) в верхних горизонтах почв тяжелых металлов (Levy et al., 1992; Kabata-Pendias, 2000; Zwolak et al., 2019; Ryazanov et al., 2019; Sakizadeh et al., 2021) и в значительной степени могут определять закономерности их пространственного

распределения (Güler et al., 2010). Несмотря на то, что в понятие органические почвенные поллютанты в настоящее время включают соединения, очень сильно различающиеся по своей химической природе, в целом их поведение в почвах определяется содержанием почвенного органического вещества (ПОВ), литологическими и физико-химическими свойствами. Считается, что содержание ПОВ и, в определенной степени, его качественный состав являются основными факторами, определяющими сорбционную способность почв по отношению к большей части органических загрязняющих веществ, включая углеводороды нефти (Schneckenburger et al., 2020; Shi et al., 2020). Очевидно, что использование данных показателей в качестве предикторов при построении пространственных моделей распределения загрязнителей в пределах верхних гумусовых горизонтов почв должно быть приоритетным и может обеспечить успешное решение задач прогноза загрязнений потенциальными поллютантами.

Методики создания цифровых почвенных карт в настоящее время хорошо отработаны и широко используются в практике почвенных исследований (McBratney et al., 2003; Grunwald, 2009; Vincent et al., 2016; Цифровая почвенная картография, 2017). Для создания цифровых карт изменения почвенных показателей, включая содержание ПОВ, ГМС, химические и физико-химические свойства, находят все большее применение методы машинного обучения и глубокого обучения (Beucher et al., 2017; Zhang et al., 2018; Laborcz et al., 2018; Paterson et al., 2018; Caubet et al., 2019). Можно с высокой степенью уверенности предположить, что карты изменения в пространстве показателей, определяющих сорбционные свойства почв, при обеспечении необходимой точности и надежности могут быть использованы как источники предикторов для создания прогнозных моделей распределения загрязнителей в пределах почвенного покрова. В работе (Pahlavan-Rad et al., 2018) интерполированные карты изменения ГМС использованы как предикторы содержания ПОВ. В работе (Were et al., 2015) также для создания прогнозных карт содержания ПОВ с применением методов машинного обучения использовали интерполированные карты содержания фракций ГМС, Ca, Mg, P, K, общего азота и pH.

Вместе с тем, успешный прогноз динамики загрязнения может быть проведен только при описании пространственной неоднородности показателей, определяющих сорбционные свойства почв, которое должно масштабироваться на уровне одного отдельно взятого поля или массива потенциального загрязнения. Площадь обследуемых участков, в данном случае, может составлять всего лишь сотни или даже десятки гектар и должна быть сопоставимой с площадью обычных аварийных загрязнений. Решение подобной задачи имеет определенные трудности, связанные с обеспечением необходимой точности пространственного прогноза в местах предсказания при ограниченном количестве точек опробования (Cui, et. al., 2016), что требует объективной оценки выбора используемых предикторов и методов пространственного прогноза.

Цель настоящей работы состояла в оценке возможности использования в качестве предикторов данных ДЗЗ для пространственного прогноза почвенных показателей,

определяющих сорбционные свойства почв по отношению к поллютантам, с применением методов машинного обучения.

Материалы и методы

В качестве объекта изучения использовали отдельно взятое поле севооборота площадью 254 га, расположенное в Заинском районе Республики Татарстан. Почвенный покров участка представлен выщелоченными черноземами и неоднороден по показателям почвенных свойств. Территория была поделена на элементарные участки размерами 5 га. Всего было отобрано 50 смешанных образцов. Содержание органического углерода (ОУ) определяли сухим сжиганием (CN-анализатор Vario Max Cube Elementar), содержание фракций пыли и глины – методом лазерной седиментографии (анализатор размера частиц Microtrac SDC), pH – ионометрически, ЕКО – химическим методом. Содержание пыли и глины было преобразовано из оригинальных данных в композитные с помощью алгебраической трансформации (Loiseau et al., 2021).

В качестве источников ДЗЗ использовали данные спутника Sentinel 2, полученные с открытого сайта Европейского космического агентства. Для работы использовались изображения поверхности почвы с минимальным влиянием растительности. Этим условиям соответствовал снимок от 12 мая 2019 года. Изображения выбирались с учетом минимального влияния атмосферных возмущений, но, все же, были подвергнуты атмосферной коррекции методом DOS 1. По полученным данным рассчитывались спектральные индексы, которые представлены отношениями отдельных каналов, и индексами, характеризующими открытую поверхность (NDVI, Grain size index, Clay index, MIR index, Bare soil index, Redness index, Saturation index, Coloration index и др.). Всего было использовано 18 индексов, а также каналы 2–8, каналы 11–12 и спутника Sentinel 2. Данные отдельных каналов и спектральных индексов экстрагировались и усреднялись по элементарным участкам отбора почвенных проб. Географическая привязка результатов анализов почвенных проб приводилась к центру элементарного участка.

Для прогноза были использованы модели регрессии на основе опорных векторов (SVMr – support vector machine regression) и модели случайного леса (RF – random forest). Регрессия на основе опорных векторов представляет собой контролируемый непараметрический метод машинного обучения. Благодаря способности обрабатывать нелинейные отношения и эффективности в обобщении, SVMr показал себя многообещающим методом в различных почвенных исследованиях (Cortes et al., 1995; Smola et

al., 2004; Pasolli et al., 2011; Deiss et al., 2020; Taghizadeh-Mehrjardi et al., 2021).

Случайный лес – это древовидный алгоритм машинного обучения, который в основном использовался для задач классификации. Несколько сравнительных исследований доказали, что это один из лучших и доступных в настоящее время методов машинного обучения (Biau et al., 2016; Hengl et al., 2018).

Модели RF и SVMr подвергались процедуре настройки. Проверка моделей производилась с помощью процедуры бутстрапа с учетом оптимизма производительности (Hartrell, 2001). Вначале рассчитывались значения производительности моделей для отдельных выборок бутстрапа. В последующем рассчитывалась производительность к исходным данным. Показатель оптимизма прогнозирующей способности определялся путем вычитания усредненных значений производительности моделей отдельных выборок бутстрапа и моделей на основе исходных данных. Итоговыми значениями производительности считались значения без показателя оптимизма. Мерами оценки моделей являлись средняя абсолютная ошибка (MAE – mean absolute error), среднеквадратичная ошибка (RMSE – root mean square error) и коэффициент детерминации (R^2), формулы расчета которых представлены ниже.

Средняя абсолютная ошибка:

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - o_i|;$$

Среднеквадратичная ошибка:

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2 \right]^{1/2};$$

Коэффициент детерминации:

$$R^2 = 1 - \frac{\sum_{i=1}^n (o_i - p_i)^2}{\sum_{i=1}^n (o_i - \bar{o})^2};$$

где p_i – прогнозируемое значение показателя, o_i – наблюдаемое значение показателя.

Наилучшими признавались модели с минимальным значением RMSE, MAE и максимальным значением R^2 . Работа с растровыми изображениями и моделирование проводилось в среде объектно-ориентированного языка R.

Результаты

В таблице 1 представлены описательные статистические данные изученных почв. Среднее содержание фракции пыли на участке исследования составляет 76.4 %, глины – 9.9 %, варибельность показателей слабая. Среднее содержание органического углерода – 4.0 % при

Почвенный показатель	ГМС, %		Органический углерод, %	pH	ЕКО, ммоль (экв)/100 г
	Пыль (0.002-0.05 мм)	Глина, (<0.002 мм)			
Минимальное значение	72.2	8.6	2.7	4.9	42.0
Максимальное значение	79.3	11.2	5.6	6.4	73.0
Размах варьирования	7.1	2.6	2.9	1.5	31.0
Среднее значение	76.4	9.9	4.0	5.5	59.8
Коэффициент вариации, %	1.5	5.2	18.7	6.1	12.3
Оценка коэффициента вариации	Слабая	Слабая	Средняя	Слабая	Средняя

Табл. 1. Описательные статистические данные свойств почв

средней вариабельности. Среднее значение величины ЕКО – 59.8 ммоль(экв)/100 г, вариабельность средняя. Вариабельность рН оценивается как слабая.

Перед моделированием пространственного распределения почвенных свойств была оценена их взаимосвязь с данными ДЗЗ на основе корреляции по методу Спирмена (рис. 1). При построении диаграммы корреляций были удалены индексы, имеющие полную корреляцию друг с другом (т.е. $r = -1$ или $r = 1$).

Наибольшую связь со спектральными индексами имеет содержание ОУ. Коэффициент корреляции с индексом ВІ составляет $r = -0.71$, а с индексом открытой поверхности почвы BSI 2 – $r = 0.78$. Содержание глины лучше коррелирует с индексом NDVI $r = -0.52$, а содержание пыли – с индексом SI $r = -0.50$, тогда как показатель рН наиболее тесно связан с индексом BSI 2 ($r = -0.44$). Из исследуемых показателей ЕКО менее всего коррелирует с данными ДЗЗ, наибольшее значение коэффициента корреляции приходится на спектральный канал 6 и составляет $r = 0.33$. Таким образом, исследуемые свойства имеют разную взаимосвязь со спектральными индексами.

Из двух моделей наилучший результат для содержания глины получен при использовании метода RF, так как этот алгоритм дает более низкое значение MAE равное 0.02 и более высокое значение коэффициента детерминации ($R^2 = 0.67$). Для содержания фракции пыли показатель MAE для модели RF ниже и соответствует 0,03, однако показатель RMSE ниже при применении SVMr ($RMSE = 0.61$).

Тем не менее, по значению коэффициента детерминации обе модели показали одинаковый результат ($R^2 = 0.72$). Показатель рН почвы наилучшим образом моделируется при применении SVMr, здесь значение R^2 существенно выше ($R^2 = 0.84$).

Пространственное распределение ОУ хорошо прогнозируется с использованием данных ДЗЗ. В этом случае обе модели показали высокие значения коэффициента детерминации ($R^2_{SVMr} = 0.83$ и $R^2_{RF} = 0.81$). Все же лучшим являлось применение модели SVMr, поскольку в этом случае остальные метрики имеют более низкие значения. Пространственная изменчивость ЕКО при использовании спектральных индексов прогнозируется существенно хуже, чем ОУ, поскольку в отличие от содержания ОУ показатель ЕКО не имеет прямого влияния на спектральные характеристики отражательной способности поверхности почв. Содержание ПОВ непосредственно определяет цветовые характеристики почв и имеет тесную связь со значениями спектрального отражения их поверхности. Для прогнозирования изменчивости ЕКО больше подходит алгоритм SVMr с более низким значением RMSE равным 4.25 и более высоким коэффициентом детерминации ($R^2 = 0.67$). По данным оценок RMSE, MAE, R^2 можно сделать общий вывод, что применение алгоритмов SVMr и RF дает близкие показатели надежности пространственного прогноза. Все же модель RF может быть признана лучшей для прогнозирования содержания глины, в остальных случаях лучшей можно признать модель SVMr (рис. 2, табл. 2).

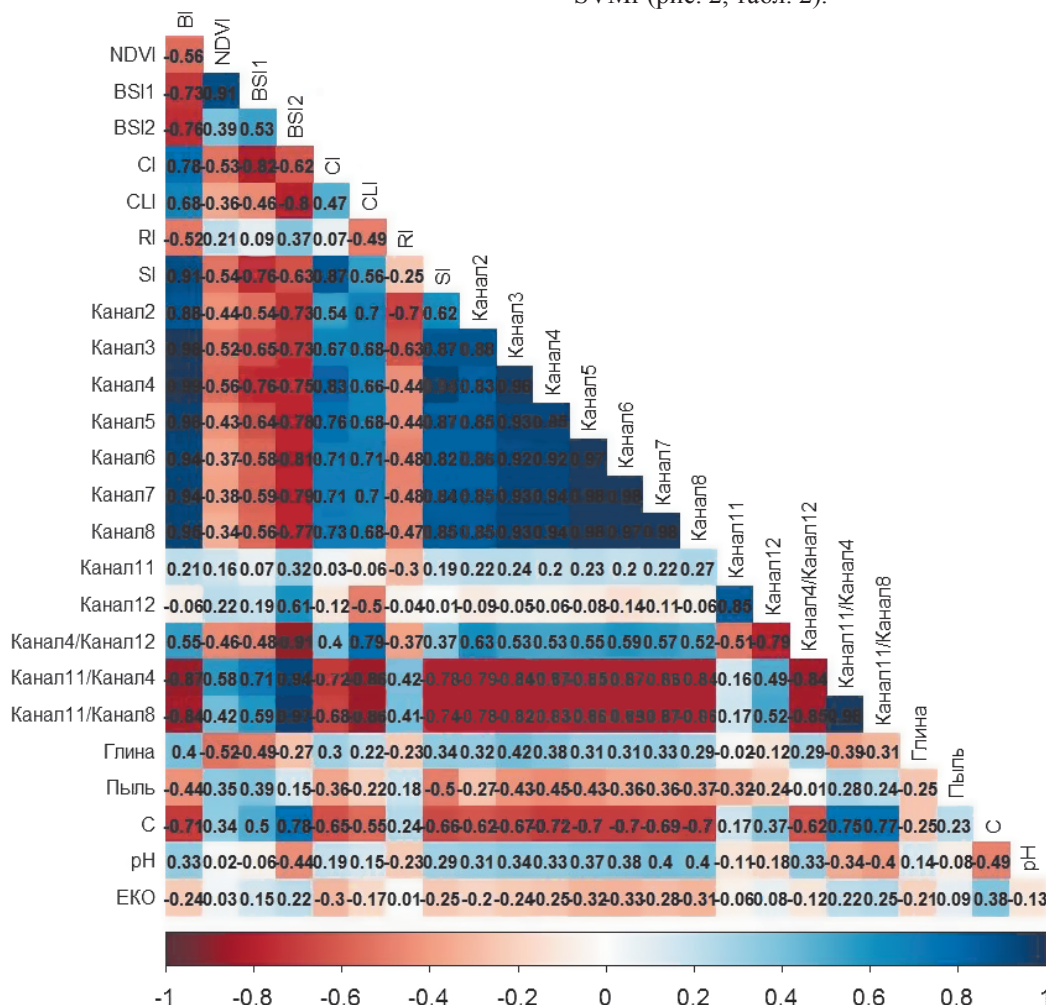


Рис. 1. Диаграмма корреляций свойств почв и данных дистанционного зондирования

Полученные результаты в целом сопоставимы с данными других исследований. В работе (Mahmoudzadeh et al., 2021) проведено сравнение пяти различных методов машинного обучения для создания карты содержания ПОВ, получены близкие результаты пространственного прогноза, но все же лучший результат показал алгоритм случайного леса. В работе (Matinfara et al., 2021) показано, что использование гибридных моделей машинного обучения при комбинации с ковариатами геоморфометрии и дистанционного зондирования позволяет моделировать и прогнозировать содержание ПОВ с приемлемой

точностью в масштабах одного поля. В работе (Were et al., 2015) применение моделей регрессии опорных векторов, искусственных нейронных сетей (ANN – Artificial neural network) и случайных лесов также показали одинаковую пространственную структуру изменения в почвенном покрове запасов ПОВ.

Полученные модели были использованы для создания карт пространственного распределения почвенных свойств (рис. 2). Почва в юго-западной части поля содержит больше фракции глины и меньше пыли. В целом, участок имеет неоднородное распределение

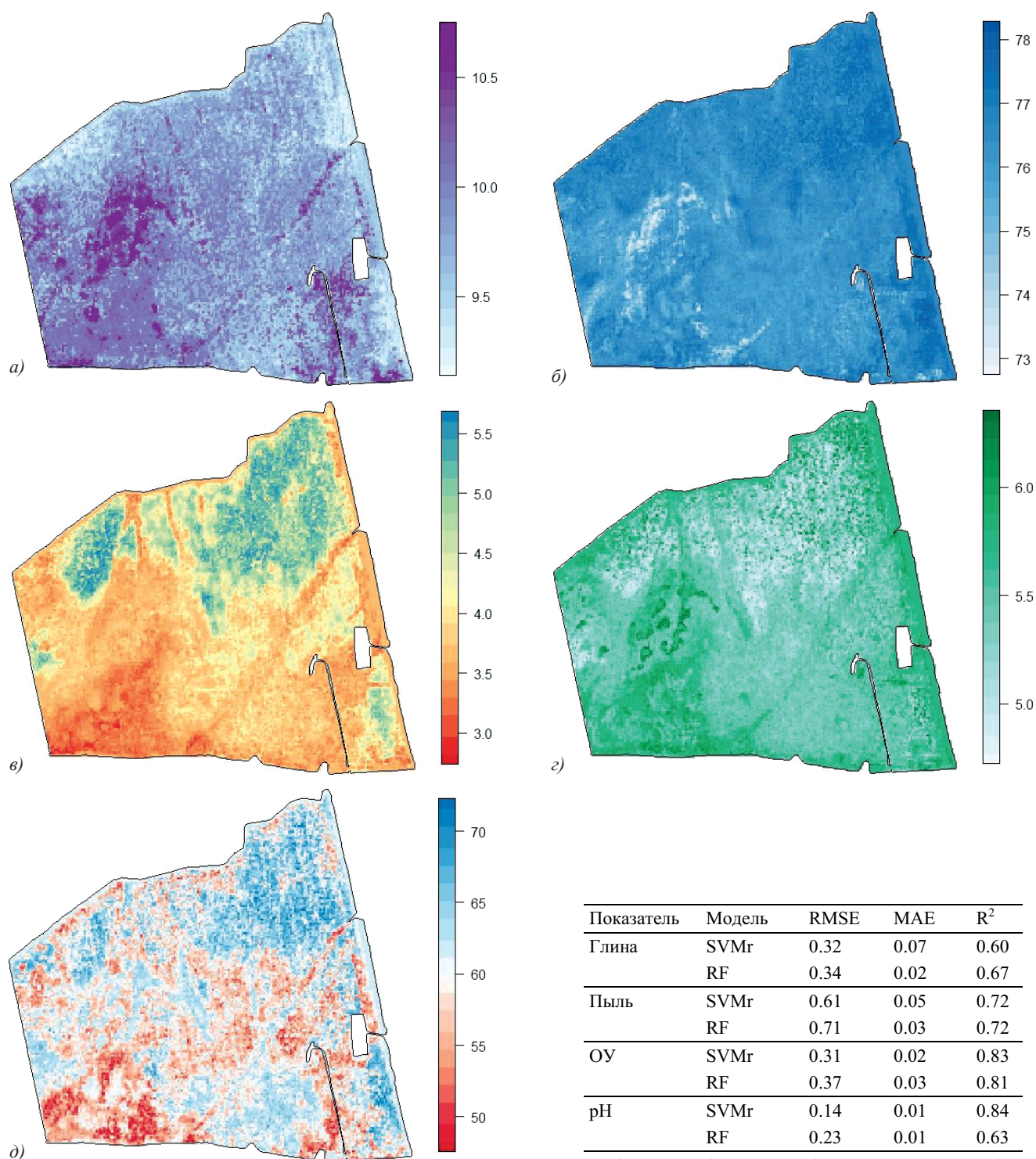


Рис. 2. Итоговые карты пространственного распределения почвенных свойств. а) Фракция глины, %; б) Фракция пыли, %; в) ОУ, %, г) pH, ед., д) ЕКО, ммоль (экв)/100 г.

Показатель	Модель	RMSE	MAE	R ²
Глина	SVMr	0.32	0.07	0.60
	RF	0.34	0.02	0.67
Пыль	SVMr	0.61	0.05	0.72
	RF	0.71	0.03	0.72
ОУ	SVMr	0.31	0.02	0.83
	RF	0.37	0.03	0.81
pH	SVMr	0.14	0.01	0.84
	RF	0.23	0.01	0.63
ЕКО	SVMr	4.25	0.32	0.67
	RF	5.31	0.21	0.60

Табл. 2. Показатели оценки точности моделей

гранулометрических фракций. Наиболее контрастным показателем является содержание ОУ. Карта позволяет условно разделить территорию на две зоны: северную с содержанием ОУ от 4.5 % до 5.5 % и южную с изменением значений от 2.7 % до 4.5 %. Карта рН показывает небольшие пространственные изменения показателя, которые составляют всего одну единицу. Влияние выше названных показателей дает контрастность участка по величине ЕКО.

Таким образом, пространственное распределение почвенных свойств на участке, несмотря на его небольшие размеры, характеризуется достаточно существенной пространственной неоднородностью, которая должна учитываться при определении рисков потенциального загрязнения поллютантами.

Можно сделать вывод, что методы машинного обучения при использовании данных ДЗЗ, полученных из спутника Sentinel 2 с пространственным разрешением 10 м, для территорий с ограниченной площадью обеспечивают построение пространственных карт с относительно точным прогнозом распределения показателей почвы, которые определяют сорбционные свойства по отношению к поллютантам. Карты, полученные с использованием различных алгоритмов (RF и SVMr), дают близкие результаты прогноза и могут быть эффективно использованы как предикторы при создании прогнозных карт загрязнения.

Заключение

Задача создания прогнозных карт загрязненности территорий имеет очень важный аспект, связанный с необходимостью объективного определения источников поступления поллютантов в почвы (Güler et. al., 2010; Lv et. al., 2013). Наряду с промышленными предприятиями, важными источником загрязнения почв тяжелыми металлами в настоящее время является сельское хозяйство за счет внесения минеральных удобрений, пестицидов/гербицидов, орошения загрязненными водами (Güler et. al., 2010). Значительное поступление отдельных тяжелых металлов связывают с урбанизацией территорий (Shi et. al., 2020). Поступление органических поллютантов также может быть связано с сельским хозяйством, с полигонами отходов, городской деятельностью (Mishra, 2016). Например, существенное поступление в окружающую среду полиароматических углеводородов может быть наряду с загрязнением нефтью и нефтепродуктами связано с естественными источниками (лесными и степными пожарами, вулканизмом), а также сжиганием древесины, нефтепродуктов, угля (Цибарт и др., 2013). Создание карт пространственного распределения поллютантов на основе данных полевых исследований и применения методов машинного обучения, учитывающих вспомогательные предикторы, включая показатели пространственной неоднородности сорбционных свойств почв, позволяет существенно повысить их точность и надежность (Shi et al., 2020). Применение данных карт может обеспечить объективность определения источников загрязнения и ответственности за причиняемый экологический ущерб (Saby et al., 2009; Yuan et al., 2013; Shi et al., 2020).

Финансирование

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, проект № 19-29-05061-мк.

Литература

- Цибарт А.С., Геннадиев А.Н. (2013). Полициклические ароматические углеводороды в почвах: источники, поведение, индикационное значение (обзор). *Почвоведение*, 7, с. 788–802. <https://doi.org/10.7868/S0032180X13070125>
- Цифровая почвенная картография (2017). М.: Изд-во РУДН, 152 с.
- Al-Ruzouq R., Gibril M. A., Abdallah S., Kais A., Hamed O., Saeed Al-M., Mohamad K. (2020). Sensors, Features, and Machine Learning for Oil Spill Detection and Monitoring: A Review. *Remote Sensing*, 12, 3338. <https://doi.org/10.3390/rs12203338>
- Andronikov S.V., Davidson D.A., Spiers R.B. (2000). Variability in Contamination by Heavy Metals: Sampling Implications. *Water, Air, & Soil Pollution*, 120, pp. 29–45. <https://doi.org/10.1023/A:1005261522465>
- Beucher A., Adhikari K., Breuning-Madsen H., Greve M.B., Österholm P., Frøjdø S., Jensen N.H., Greve M.H. (2017). Mapping potential acid sulfate soils in Denmark using legacy data and LiDAR-based derivatives. *Geoderma*, 308, pp. 363–372. <https://doi.org/10.1016/j.geoderma.2016.06.001>
- Biau G., Scornet E. (2016). A random forest guided tour. *Test*, 25, pp. 197–227. <https://doi.org/10.1007/s11749-016-0481-7>
- Caubet M., Dobarco R. M., Arrouays D., Minasny B., Saby N. (2019). Merging country, continental and global predictions of soil texture: Lessons from ensemble modelling in France. *Geoderma*, 337, pp. 99–110. <https://doi.org/10.1016/j.geoderma.2018.09.007>
- Cho K.H., Sthiannopkao S., Pachepsky Y.A., Kim K.W., Kim J.H. (2011). Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network. *Water Res.*, 45(17), pp. 5535–5544. <https://doi.org/10.1016/j.watres.2011.08.010>
- Cortes C., Vapnik V. (1995). Support-vector networks. *Mach. Learn.*, 20, pp. 273–297. <https://doi.org/10.1007/BF00994018>
- Cui Y.-Q., Yoneda M., Shimada Y., Matsui Y. (2016). Cost-Effective Strategy for the Investigation and Remediation of Polluted Soil Using Geostatistics and a Genetic Algorithm Approach. *Journal of Environmental Protection*, 07(01), pp. 99–115. <https://doi.org/10.4236/jep.2016.71010>
- Deiss L., Margenot A.J., Culman S.W., Demyan M.S. (2020). Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. *Geoderma*, 365, 114227. <https://doi.org/10.1016/j.geoderma.2020.114227>
- Einax J., Soldt U., Geostatistical investigations of polluted soils. (1995). *Fresenius' Journal of Analytical Chemistry*, 351, pp. 48–53. <https://doi.org/10.1016/j.envpol.2012.06.006>
- Grunwald S. (2009). Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152, pp. 195–207. <https://doi.org/10.1016/j.geoderma.2009.06.003>
- Güler C., Alpaslan M., Kurt M.A. (2010). Deciphering factors controlling trace element distribution in the soils of Karaduvar industrial-agricultural area (Mersin, SE Turkey). *Environ Earth Sci*, 60, pp. 203–218. <https://doi.org/10.1007/s12665-009-0180-8>
- Ha H., Olson J.R., Bian L., Rogerson P.A. (2014). Analysis of Heavy Metal Sources in Soil Using Kriging Interpolation on Principal Components. *Environmental Science & Technology*, 48, pp. 4999–5007. <https://doi.org/10.1021/es405083f>
- Harrell F.E.Jr. (2001). Regression Modeling Strategies. With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer, 507 p. <https://doi.org/10.1007/978-1-4757-3462-1>
- Hengl T., Nussbaum M., Wright M.N., Heuvelink G.B.M., Gräler B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6, e5518. <https://doi.org/10.7717/peerj.5518>
- Hooda, P.S., Glavinandp R.J. (2005). A Practical Examination of the Use of Geostatistics in the Remediation of a Site with a Complex Metal Contamination History. *Soil and Sediment Contamination*, 14, pp. 155–169. <https://doi.org/10.1080/15320380590911814>
- Juang K.-W., Liao W.-J., Liu T.-L., Tsui L., Lee D.-Y. (2008). Additional sampling based on regulation threshold and kriging variance to reduce the probability of false delineation in a contaminated site. *Science of the Total Environment*, 389, pp. 20–28. <https://doi.org/10.1016/j.scitotenv.2007.08.025>

- Kabata-Pendias A. (2000). Trace Elements in Soils and Plants. CRC Press, 403 p. <https://doi.org/10.1201/9781420039900>
- Laborci A., Gábor S., Kaposi A., László P. (2018). Comparison of soil texture maps synthesized from standard depth layers with directly compiled products. *Geoderma*, 352, pp. 360–372. <https://doi.org/10.1016/j.geoderma.2018.01.020>
- Levy D.B., Barbarrick K.A., Siemer E.G., Sommers L.E. (1992). Distribution and partitioning of trace metals in contaminated soils near Leadville, Colorado. *J Environ Qual*, 21, pp. 185–195. <https://doi.org/10.2134/jeq1992.004724250002100020006x>
- Lin Y.-P., Cheng B.-Y., Chu H.-J., Chang T.-K., Yu H.-L. (2011). Assessing how heavy metal pollution and human activity are related by using logistic regression and kriging methods. *Geoderma*, 163(3–4), pp. 275–282. <https://doi.org/10.1016/j.geoderma.2011.05.004>
- Lin Y.-P., Chu H.-J., Huang Y.-L., Cheng B.-Y., Chang T.-K. (2010). Modeling Spatial Uncertainty of Heavy Metal Content in Soil by Conditional Latin Hypercube Sampling and Geostatistical Simulation. *Environmental Earth Sciences*, 62, pp. 299–311. <https://doi.org/10.1007/s12665-010-0523-5>
- Loiseau T., Arrouays D., Richer-de-Forges A., Lagacherie P., Ducommun C., Minasny B. (2021). Density of soil observations in digital soil mapping: A study in the Mayenne region, France. *Geoderma Reg.*, 24, e00358. <https://doi.org/10.1016/j.geodrs.2021.e00358>
- Ly J., Yang L., Zhang Z., Dai J. (2013). Factorial kriging and stepwise regression approach to identify environmental factors influencing spatial multi-scale variability of heavy metals in soils. *Journal of Hazardous Materials*, 261(15), pp. 387–397. <https://doi.org/10.1016/j.jhazmat.2013.07.065>
- Mahmoudzadeh H., Matinfar H.R., Taghizadeh-Mehrjardi R., Kerry R. (2020). Spatial prediction of soil organic carbon using machine learning techniques in western Iran. *Geoderma Reg.*, 21, e00260. <https://doi.org/10.1016/j.geodrs.2020.e00260>
- Matinfar H. R., Maghsodi Z., Mousavi S. R., Rahmani A. (2021). Evaluation and Prediction of Topsoil organic carbon using Machine learning and hybrid models at a Field-scale. *Catena*, 202, pp. 105258. <https://doi.org/10.1016/j.catena.2021.105258>
- McBratney A.B., Mendonça Santos M.L., Minasny B. (2003). On digital soil mapping. *Geoderma*, 117, pp. 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Milillo T.M., Sinha G., Gardella J.A. (2012). Use of Geostatistics for Remediation Planning to Transcend Urban Political Boundaries. *Environmental Pollution*, 170, pp. 52–62. <https://doi.org/10.1016/j.envpol.2012.06.006>
- Mishra R., Naseer M., Nilanjan R. (2016). Soil pollution: Causes, effects and control. *Van Sanyan*, 3, pp. 1–14.
- Pahlavan-Rad M.R., Dahmardeh K., Brungard C. (2018). Predicting soil organic carbon concentrations in a low relief landscape, eastern Iran. *Geoderma Reg.*, 15, e00195. <https://doi.org/10.1016/j.geodrs.2018.e00195>
- Pasolli L., Notarnicola C., Bruzzone L. (2011). Estimating soil moisture with the support vector regression technique. *IEEE Geosci. Remote Sens. Lett.*, 8, pp. 1080–1084. <https://doi.org/10.1109/LGRS.2011.2156759>
- Paterson S., Minasny B., Mcbratney A. (2018). Spatial variability of Australian soil texture: A multiscale analysis. *Geoderma*, 309, pp. 60–74. <https://doi.org/10.1016/j.geoderma.2017.09.005>
- Platenburg R.J.P.M., Tuinhof H., Bot A.P., Iwaco B.V. (1988). Geostatistics in Soil Pollution Research. Contaminated Soil '88. Springer, Dordrecht, pp. 209–211. https://doi.org/10.1007/978-94-009-2807-7_32
- Ryazanov S. S., Ivanov D. V., Kulagina V. I. (2019). Heavy metals in topsoils of the Republic of Tatarstan. *Russian Journal of Ecosystem Ecology*, 4(3), pp. 1–14. <https://doi.org/10.21685/2500-0578-2019-3-4>
- Saby N., Thioulouse J., Jolivet C., Ratie C., Boulonne L., Bispo A., Arrouays D. (2009). Multivariate analysis of the spatial patterns of 8 trace elements using the French monitoring network data. *Science of the Total Environment*, 407, pp. 5644–5652. <https://doi.org/10.1016/j.scitotenv.2009.07.002>
- Sakizadeh M., Martín J.A.R. (2021). Spatial methods to analyze the relationship between Spanish soil properties and cadmium content. *Chemosphere*, 268, 129347. <https://doi.org/10.1016/j.chemosphere.2020.129347>
- Schneckenburger T., Thiele-Bruhn S. (2020). Sorption of PAHs and PAH derivatives in peat soil is affected by prehydration status: the role of SOM and sorbate properties. *J Soils Sediments*, 20, pp. 3644–3655. <https://doi.org/10.1007/s11368-020-02695-z>
- Sergeev A.P., Buevich A.G., Baglaeva E.M., Shichkin A.V. (2019). Combining spatial autocorrelation with machine learning increases prediction accuracy of soil heavy metals. *Catena*, 174, pp. 425–435. <https://doi.org/10.1016/j.catena.2018.11.037>
- Shi B., Ngueleu S.K., Rezanezhad F., Slowinski S., Pronk G.J., Smeaton C.M., Stevenson K., Al-Raoush R.I., Van Cappellen P. (2020) Sorption and Desorption of the Model Aromatic Hydrocarbons Naphthalene and Benzene: Effects of Temperature and Soil Composition. *Front. Environ. Chem*, 1, 581103. <https://doi.org/10.3389/fenvc.2020.581103>
- Shi T., Yang C., Liu H., Wu C., Wang Z., Li H., Zhang H., Guo L., Wu G., Su F. (2021). Mapping lead concentrations in urban topsoil using proximal and remote sensing data and hybrid statistical approaches. *Environmental Pollution*, 272, 116041. <https://doi.org/10.1016/j.envpol.2020.116041>
- Smola A.J., Scholkopf B. (2004). A tutorial on support vector regression. *Stat. Comput.*, 14, pp. 199–222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
- Taghizadeh-Mehrjardi R., Schmidt K., Toomanian N., Heung B., Behrens T., Mosavi A., Band S.S., Amirian-Chakan A., Fathabadi A., Scholten T. (2021). Improving the spatial prediction of soil salinity in arid regions using wavelet transformation and support vector regression models. *Geoderma*, 383, 114793. <https://doi.org/10.1016/j.geoderma.2020.114793>
- Tarasov D.A., Buevich A.G., Sergeev A.P., Shichkin A.V. (2018). High variation topsoil pollution forecasting in the Russian Subarctic: Using artificial neural networks combined with residual kriging. *Applied Geochemistry*, 88, Part B, pp. 188–197. <https://doi.org/10.1016/j.apgeochem.2017.07.007>
- Vincent S., Lemerrier B., Berthier L., Walte C. (2018). Spatial disaggregation of complex Soil Map Units at the regional scale based on soil-landscape relationships. *Geoderma*, 311, pp. 130–142. <https://doi.org/10.1016/j.geoderma.2016.06.006>
- Were K., Bui D.T., Dick Ø.B., Singh B.R. (2015). A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape. *Ecol. Indic.*, 52, pp. 394–403. <https://doi.org/10.1016/j.ecolind.2014.12.028>
- Yuan G., Sun T., Han P., Li J. (2013). Environmental geochemical mapping and multivariate geostatistical analysis of heavy metals in topsoils of a closed steel smelter: Capital Iron and Steel Factory, Beijing, China. *Journal of Geochemical Exploration*, 130, pp. 15–21. <https://doi.org/10.1016/j.gexplo.2013.02.010>
- Zhang L., Liu Y., Li X., Huang L., Yu D., Shi X., Chen H., Xing S. (2018). Effects of soil map scales on simulating soil organic carbon changes of upland soils in Eastern China. *Geoderma*, 312, pp. 159–169. <https://doi.org/10.1016/j.geoderma.2017.10.017>
- Zwolak A., Sarzyńska M., Szpyrka E., Stawarczyk K. (2019). Sources of Soil Pollution by Heavy Metals and Their Accumulation in Vegetables: a Review. *Water, Air, & Soil Pollution*, 230(164). <https://doi.org/10.1007/s11270-019-4221-y>

Сведения об авторах

Камиль Гашикович Гиниятуллин – канд. биол. наук, доцент, Казанский (Приволжский) федеральный университет Россия, 420008, Казань, ул. Кремлевская, д. 18 e-mail: ginijatullin@mail.ru

Ильназ Алимович Сахабиев – старший преподаватель Казанский (Приволжский) федеральный университет Россия, 420008, Казань, ул. Кремлевская, д. 18

Елена Васильевна Смирнова – канд. биол. наук, зав. кафедрой, доцент, Казанский (Приволжский) федеральный университет Россия, 420008, Казань, ул. Кремлевская, д. 18

Ильдар Анварович Уразметов – канд. пед. наук, доцент Казанский (Приволжский) федеральный университет Россия, 420008, Казань, ул. Кремлевская, д. 18

Родион Владимирович Окунев – канд. биол. наук, доцент, Казанский (Приволжский) федеральный университет Россия, 420008, Казань, ул. Кремлевская, д. 18

Карина Андреевна Гордеева – аспирант Казанский (Приволжский) федеральный университет Россия, 420008, Казань, ул. Кремлевская, д. 18

Статья поступила в редакцию 22.07.2021;

Принята к публикации 12.10.2021; Опубликована 30.03.2022

Digital mapping of indicators that determine the sorption properties of soils in relation to pollutants, according to remote sensing data of the Earth using machine learning

K.G. Giniyatullin*, I.A. Sahabiev, E.V. Smirnova, I.A. Urazmetov, R.V. Okunev, K.A. Gordeeva

Kazan Federal University, Kazan, Russia

*Corresponding author: Kamil G. Giniyatullin, e-mail: giniyatullin@mail.ru

Abstract. According to the data of remote sensing of the Earth, the accuracy of the spatial prediction of soil indicators determining sorption properties in relation to pollutants was compared. To build spatial maps of changes in soil properties, machine learning methods based on support vector regression models (SVMr) and random forest (RF) were used. It was shown that the methods of machine modeling using remote sensing can be successfully used for spatial prediction of the content of particle size fractions, organic matter, pH and the capacity of cation exchange of soils in small areas. It is shown that the spatial prediction of the content of silt fraction is best modeled using the RF algorithm, while the other properties of soils that can determine their sorption potential in relation to pollutants are better modeled using the SVMr method. In general, both machine learning methods have similar spatial prediction results.

Keywords: sorption properties of soil, spatial prediction, remote sensing data of the Earth, machine learning methods

Acknowledgements

This research was funded by the Russian Foundation for Basic Research, research project No. 19-29-05061-mk.

Recommended citation: Giniyatullin K.G., Sahabiev I.A., Smirnova E.V., Urazmetov I.A., Okunev R.V., Gordeeva K.A. (2022). Digital mapping of indicators that determine the sorption properties of soils in relation to pollutants, according to remote sensing data of the Earth using machine learning. *Georesursy = Georesources*, 24(1), pp. 84–92. DOI: <https://doi.org/10.18599/grs.2022.1.8>

References

- Al-Ruzouq R., Gibril M. A., Abdallah S., Kais A., Hamed O., Saeed Al-M., Mohamad K. (2020). Sensors, Features, and Machine Learning for Oil Spill Detection and Monitoring: A Review. *Remote Sensing*, 12, 3338. <https://doi.org/10.3390/rs12203338>
- Andronikov S.V., Davidson D.A., Spiers R.B. (2000). Variability in Contamination by Heavy Metals: Sampling Implications. *Water, Air, & Soil Pollution*, 120, pp. 29–45. <https://doi.org/10.1023/A:1005261522465>
- Beucher A., Adhikari K., Breuning-Madsen H., Greve M.B., Österholm P., Fröjdö S., Jensen N.H., Greve M.H. (2017). Mapping potential acid sulfate soils in Denmark using legacy data and LiDAR-based derivatives. *Geoderma*, 308, pp. 363–372. <https://doi.org/10.1016/j.geoderma.2016.06.001>
- Biau G., Scornet E. (2016). A random forest guided tour. *Test*, 25, pp. 197–227. <https://doi.org/10.1007/s11749-016-0481-7>
- Caubet M., Dobarco R. M., Arrouays D., Minasny B., Saby N. (2019). Merging country, continental and global predictions of soil texture: Lessons from ensemble modelling in France. *Geoderma*, 337, pp. 99–110. <https://doi.org/10.1016/j.geoderma.2018.09.007>
- Cho K.H., Sthiannopkao S., Pachepsky Y.A., Kim K.W., Kim J.H. (2011). Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network. *Water Res.*, 45(17), pp. 5535–5544. <https://doi.org/10.1016/j.watres.2011.08.010>
- Cortes C., Vapnik V. (1995). Support-vector networks. *Mach. Learn.*, 20, pp. 273–297. <https://doi.org/10.1007/BF00994018>
- Cui Y.-Q., Yoneda M., Shimada Y., Matsui Y. (2016). Cost-Effective Strategy for the Investigation and Remediation of Polluted Soil Using Geostatistics and a Genetic Algorithm Approach. *Journal of Environmental Protection*, 07(01), pp. 99–115. <https://doi.org/10.4236/jep.2016.71010>
- Deiss L., Margenot A.J., Culman S.W., Demyan M.S. (2020). Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. *Geoderma*, 365, 114227. <https://doi.org/10.1016/j.geoderma.2020.114227>
- Digital soil cartography (2017). Moscow: RUDN University, 152 p. (In Russ.)
- Einax J., Soldt U., Geostatistical investigations of polluted soils. (1995). Fresenius' Journal of Analytical Chemistry, 351, pp. 48–53. <https://doi.org/10.1016/j.envpol.2012.06.006>
- Grunwald S. (2009). Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152, pp. 195–207. <https://doi.org/10.1016/j.geoderma.2009.06.003>
- Güler C., Alpaslan M., Kurt M.A. (2010). Deciphering factors controlling trace element distribution in the soils of Karaduvar industrial-agricultural area (Mersin, SE Turkey). *Environ Earth Sci*, 60, pp. 203–218. <https://doi.org/10.1007/s12665-009-0180-8>
- Ha H., Olson J.R., Bian L., Rogerson P.A. (2014). Analysis of Heavy Metal Sources in Soil Using Kriging Interpolation on Principal Components. *Environmental Science & Technology*, 48, pp. 4999–5007. <https://doi.org/10.1021/es405083f>
- Harrell F.E.Jr. (2001). Regression Modeling Strategies. With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer, 507 p. <https://doi.org/10.1007/978-1-4757-3462-1>
- Hengl T., Nussbaum M., Wright M.N., Heuvelink G.B.M., Gräler B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6, e5518. <https://doi.org/10.7717/peerj.5518>
- Hooda, P.S., Glavinandp R.J. (2005). A Practical Examination of the Use of Geostatistics in the Remediation of a Site with a Complex Metal Contamination History. *Soil and Sediment Contamination*, 14, pp. 155–169. <https://doi.org/10.1080/15320380590911814>
- Juang K.-W., Liao W.-J., Liu T.-L., Tsui L., Lee D.-Y. (2008). Additional sampling based on regulation threshold and kriging variance to reduce the probability of false delineation in a contaminated site. *Science of the Total Environment*, 389, pp. 20–28. <https://doi.org/10.1016/j.scitotenv.2007.08.025>
- Kabata-Pendias A. (2000). Trace Elements in Soils and Plants. CRC Press, 403 p. <https://doi.org/10.1201/9781420039900>
- Laborezi A., Gábor S., Kaposi A., László P. (2018). Comparison of soil texture maps synthesized from standard depth layers with directly compiled products. *Geoderma*, 352, pp. 360–372. <https://doi.org/10.1016/j.geoderma.2018.01.020>
- Levy D.B., Barbarrick K.A., Siemer E.G., Sommers L.E. (1992). Distribution and partitioning of trace metals in contaminated soils near Leadville, Colorado. *J Environ Qual*, 21, pp. 185–195. <https://doi.org/10.2134/jeq1992.00472425002100020006x>
- Lin Y.-P., Cheng B.-Y., Chu H.-J., Chang T.-K., Yu H.-L. (2011). Assessing how heavy metal pollution and human activity are related by using logistic regression and kriging methods. *Geoderma*, 163(3–4), pp. 275–282. <https://doi.org/10.1016/j.geoderma.2011.05.004>
- Lin Y.-P., Chu H.-J., Huang Y.-L., Cheng B.-Y., Chang T.-K. (2010). Modeling Spatial Uncertainty of Heavy Metal Content in Soil by Conditional Latin Hypercube Sampling and Geostatistical Simulation. *Environmental Earth Sciences*, 62, pp. 299–311. <https://doi.org/10.1007/s12665-010-0523-5>
- Loiseau T., Arrouays D., Richer-de-Forges A., Lagacherie P., Ducommun C., Minasny B. (2021). Density of soil observations in digital soil mapping:

A study in the Mayenne region, France. *Geoderma Reg.*, 24, e00358. <https://doi.org/10.1016/j.geodrs.2021.e00358>

Ly J., Yang L., Zhang Z., Dai J. (2013). Factorial kriging and stepwise regression approach to identify environmental factors influencing spatial multi-scale variability of heavy metals in soils. *Journal of Hazardous Materials*, 261(15), pp. 387–397. <https://doi.org/10.1016/j.jhazmat.2013.07.065>

Mahmoudzadeh H., Matinfar H.R., Taghizadeh-Mehrjardi R., Kerry R. (2020). Spatial prediction of soil organic carbon using machine learning techniques in western Iran. *Geoderma Reg.*, 21, e00260. <https://doi.org/10.1016/j.geodrs.2020.e00260>

Matinfar H. R., Maghsodi Z., Mousavi S. R., Rahmani A. (2021). Evaluation and Prediction of Topsoil organic carbon using Machine learning and hybrid models at a Field-scale. *Catena*, 202, pp. 105258. <https://doi.org/10.1016/j.catena.2021.105258>

McBratney A.B., Mendonça Santos M.L., Minasny B. (2003). On digital soil mapping. *Geoderma*, 117, pp. 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)

Milillo T.M., Sinha G., Gardella J.A. (2012). Use of Geostatistics for Remediation Planning to Transcend Urban Political Boundaries. *Environmental Pollution*, 170, pp. 52–62. <https://doi.org/10.1016/j.envpol.2012.06.006>

Mishra R., Naseer M., Nilanjan R. (2016). Soil pollution: Causes, effects and control. *Van Sangyan*, 3, pp. 1–14.

Pahlavan-Rad M.R., Dahmardeh K., Brungard C. (2018). Predicting soil organic carbon concentrations in a low relief landscape, eastern Iran. *Geoderma Reg.*, 15, e00195. <https://doi.org/10.1016/j.geodrs.2018.e00195>

Pasolli L., Notarnicola C., Bruzzone L. (2011). Estimating soil moisture with the support vector regression technique. *IEEE Geosci. Remote Sens. Lett.*, 8, pp. 1080–1084. <https://doi.org/10.1109/LGRS.2011.2156759>

Paterson S., Minasny B., Mcbratney A. (2018). Spatial variability of Australian soil texture: A multiscale analysis. *Geoderma*, 309, pp. 60–74. <https://doi.org/10.1016/j.geoderma.2017.09.005>

Platenburg R.J.P.M., Tuinhof H., Bot A.P., Iwaco B.V. (1988). Geostatistics in Soil Pollution Research. Contaminated Soil '88. Springer, Dordrecht, pp. 209–211. https://doi.org/10.1007/978-94-009-2807-7_32

Ryazanov S. S., Ivanov D. V., Kulagina V. I. (2019). Heavy metals in topsoils of the Republic of Tatarstan. *Russian Journal of Ecosystem Ecology*, 4(3), pp. 1–14. <https://doi.org/10.21685/2500-0578-2019-3-4>

Saby N., Thioulouse J., Jolivet C., Ratie C., Boulonne L., Bispo A., Arrouays D. (2009). Multivariate analysis of the spatial patterns of 8 trace elements using the French monitoring network data. *Science of the Total Environment*, 407, pp. 5644–5652. <https://doi.org/10.1016/j.scitotenv.2009.07.002>

Sakizadeh M., Martín J.A.R. (2021). Spatial methods to analyze the relationship between Spanish soil properties and cadmium content. *Chemosphere*, 268, 129347. <https://doi.org/10.1016/j.chemosphere.2020.129347>

Schneckenburger T., Thiele-Bruhn S. (2020). Sorption of PAHs and PAH derivatives in peat soil is affected by prehydration status: the role of SOM and sorbate properties. *J Soils Sediments*, 20, pp. 3644–3655. <https://doi.org/10.1007/s11368-020-02695-z>

Sergeev A.P., Buevich A.G., Baglaeva E.M., Shichkin A.V. (2019). Combining spatial autocorrelation with machine learning increases prediction accuracy of soil heavy metals. *Catena*, 174, pp. 425–435. <https://doi.org/10.1016/j.catena.2018.11.037>

Shi B., Ngueleu S.K., Rezanezhad F., Slowinski S., Pronk G.J., Smeaton C.M., Stevenson K., Al-Raoush R.I., Van Cappellen P. (2020) Sorption and Desorption of the Model Aromatic Hydrocarbons Naphthalene and Benzene: Effects of Temperature and Soil Composition. *Front. Environ. Chem.*, 1, 581103. <https://doi.org/10.3389/fenvc.2020.581103>

Shi T., Yang C., Liu H., Wu C., Wang Z., Li H., Zhang H., Guo L., Wu G., Su F. (2021). Mapping lead concentrations in urban topsoil using proximal and remote sensing data and hybrid statistical approaches. *Environmental Pollution*, 272, 116041. <https://doi.org/10.1016/j.envpol.2020.116041>

Smola A.J., Schölkopf B. (2004). A tutorial on support vector regression. *Stat. Comput.*, 14, pp. 199–222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>

Taghizadeh-Mehrjardi R., Schmidt K., Toomanian N., Heung B., Behrens T., Mosavi A., Band S.S., Amirian-Chakan A., Fathabadi A., Scholten T. (2021). Improving the spatial prediction of soil salinity in arid regions using wavelet transformation and support vector regression models. *Geoderma*, 383, 114793. <https://doi.org/10.1016/j.geoderma.2020.114793>

Tarasov D.A., Buevich A.G., Sergeev A.P., Shichkin A.V. (2018). High variation topsoil pollution forecasting in the Russian Subarctic: Using artificial neural networks combined with residual kriging. *Applied Geochemistry*, 88, Part B, pp. 188–197. <https://doi.org/10.1016/j.apgeochem.2017.07.007>

Tsibart A.S., Gennadiev A.N. (2013). Polycyclic aromatic hydrocarbons in soils: sources, behavior, and indication significance (a review). *Eurasian Soil Science*, 46(7), pp. 728–741. <https://doi.org/10.1134/S1064229313070090>

Vincent S., Lemerrier B., Berthier L., Walte C. (2018). Spatial disaggregation of complex Soil Map Units at the regional scale based on soil-landscape relationships. *Geoderma*, 311, pp. 130–142. <https://doi.org/10.1016/j.geoderma.2016.06.006>

Were K., Bui D.T., Dick Ø.B., Singh B.R. (2015). A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecol. Indic.*, 52, pp. 394–403. <https://doi.org/10.1016/j.ecolind.2014.12.028>

Yuan G., Sun T., Han P., Li J. (2013). Environmental geochemical mapping and multivariate geostatistical analysis of heavy metals in topsoils of a closed steel smelter: Capital Iron and Steel Factory, Beijing, China. *Journal of Geochemical Exploration*, 130, pp. 15–21. <https://doi.org/10.1016/j.gexplo.2013.02.010>

Zhang L., Liu Y., Li X., Huang L., Yu D., Shi X., Chen H., Xing S. (2018). Effects of soil map scales on simulating soil organic carbon changes of upland soils in Eastern China. *Geoderma*, 312, pp. 159–169. <https://doi.org/10.1016/j.geoderma.2017.10.017>

Zwolak A., Sarzyńska M., Szpyrka E., Stawarczyk K. (2019). Sources of Soil Pollution by Heavy Metals and Their Accumulation in Vegetables: a Review. *Water, Air, & Soil Pollution*, 230(164). <https://doi.org/10.1007/s11270-019-4221-y>

About the Authors

Kamil G. Giniyatullin – PhD (Biology), Associate Professor, Kazan Federal University
18, Kremlevskaya st., Kazan, 420008, Russian Federation
e-mail: giniyatullin@mail.ru

Il'nas A. Sahabiev – Senior Lecturer, Kazan Federal University
18, Kremlevskaya st., Kazan, 420008, Russian Federation

Elena V. Smirnova – PhD (Biology), Associate Professor, Kazan Federal University
18, Kremlevskaya st., Kazan, 420008, Russian Federation

Ildar A. Urazmetov – PhD (Pedagogic), Associate Professor, Kazan Federal University
18, Kremlevskaya st., Kazan, 420008, Russian Federation

Rodion V. Okunev – PhD (Biology), Associate Professor, Kazan Federal University
18, Kremlevskaya st., Kazan, 420008, Russian Federation

Karina A. Gordeeva – PhD student, Kazan Federal University
18, Kremlevskaya st., Kazan, 420008, Russian Federation

Manuscript received 22 July 2021;

Accepted 12 October 2021; Published 30 March 2022