

УДК 631.4

DOI: 10.19047/0136-1694-2022-111-77-96



Ссылки для цитирования:

Прудникова Е.Ю., Савин И.Ю., Виндекер Г.В. Распознавание пахотных почв по фотографиям, получаемым в рамках краудсорсинговых технологий // Бюллетень Почвенного института имени В.В. Докучаева. 2022. Вып. 111. С. 77-96. DOI: 10.19047/0136-1694-2022-111-77-96

Cite this article as:

Prudnikova E.Yu., Savin I.Yu., Vindeker G.V., Recognition of arable soils from photographs obtained as part of crowdsourcing technologies, Dokuchaev Soil Bulletin, 2022, V. 111, pp. 77-96, DOI: 10.19047/0136-1694-2022-111-77-96

Благодарность:

Исследования выполнены при финансовой поддержке проекта РФ в лице Минобрнауки (соглашение № 075-15-2022-321).

Acknowledgments:

The studies were carried out with the financial support of the Ministry of Education and Science of the Russian Federation in the framework of the scientific project (Agreement No. 075-15-2022-321).

Распознавание пахотных почв по фотографиям, получаемым в рамках краудсорсинговых технологий

© 2022 г. Е. Ю. Прудникова^{1,2*}, И. Ю. Савин^{1,2**},
Г. В. Виндекер^{1***}

¹ФИЦ “Почвенный институт им. В.В. Докучаева”, Россия,
119017, Москва, Пыжевский пер, 7, стр. 2,

* <https://orcid.org/0000-0001-7743-8607>, e-mail: prudnikova_eyu@esoil.ru,

** <https://orcid.org/0000-0002-8739-5441>, e-mail: savin_iyu@esoil.ru,

*** <https://orcid.org/0000-0002-0463-42411>, e-mail: gretelericka@gmail.com.

²Институт экологии, Российский университет дружбы народов, Россия,
117198, Москва, ул. Миклухо-Маклая, 6.

Поступила в редакцию 19.04.2022, после доработки 17.05.2022,
принята к публикации 27.09.2022

Резюме: В статье рассматриваются возможности использования фотографий, получаемых при использовании краудсорсинговых

технологий для оперативной инвентаризации пахотных почв. Объектом исследования выступает спектральная отражательная способность открытой поверхности пахотных почв тестовых участков, измеренная с помощью спектрорадиометра HandHeld-2, регистрирующего отражение в диапазоне 325–1 075 нм, и их изображение на фотографиях, полученных обычными фотокамерами. Тестовые участки расположены в Тульской, Московской и Тверской областях. Почвы тестовых участков – дерново-подзолистые, серые лесные, черноземы выщелоченные. На основе анализа фотографий поверхности и информации, полученной с помощью спектрорадиометра, был рассчитан набор спектральных параметров в цветовых системах RGB, YMC и HSI, а также их соотношения (45 параметров). Данные параметры использовались для разделения анализируемых типов почв с помощью деревьев классификации. Точность классификации по результатам валидации варьирует в пределах 63–100%. При этом параметры цветовых систем HSI и YMC оказались более информативны, чем параметры цветовой системы RGB. Установленные правила классификации в дальнейшем могут применяться для определения классификационного положения почв по изображениям, собранным с помощью краудсорсинговых технологий.

Ключевые слова: инвентаризация почв, краудсорсинг, дистанционные данные, деревья принятия решений.

Recognition of arable soils from photographs obtained as part of crowdsourcing technologies

© 2022 E. Yu. Prudnikova^{1,2*}, I. Yu. Savin^{1,2**},
G. V. Vindeker^{1***}

¹Federal Research Centre “V.V. Dokuchaev Soil Science Institute”,
7 Bld. 2 Pyzhevskiy per., Moscow 119017, Russian Federation,

*<https://orcid.org/0000-0001-7743-8607>, e-mail: prudnikova_eyu@esoil.ru,

**<https://orcid.org/0000-0002-8739-5441>, e-mail: savin_iyu@esoil.ru,

***<https://orcid.org/0000-0002-0463-42411>, e-mail: gretelericka@gmail.com.

²Institute of Ecology, RUDN University,
8/2 Miklukho-Maklaya Str., Moscow 117198, Russian Federation.

Received 19.04.2022, Revised 17.05.2022, Accepted 27.09.2022

Abstract: The study focuses on the possibilities of using photographs obtained using crowdsourcing technologies for the operational inventory of arable soils. The object of the study is the spectral reflectance of the open surface of arable soils of the test plots, measured using a HandHeld-2 spectroradiometer

operating in the range of 325–1 075 nm, and their image in photographs taken with conventional cameras. Test sites are located in the Tula, Moscow and Tver regions. The soils of the test plots are sod-podzolic, gray forest, and leached chernozems. Based on the analysis of photographs of the surface and information obtained using a spectroradiometer, a set of spectral parameters in the RGB, YMC and HSI color systems, as well as their ratios (45 parameters), was calculated. These parameters were used to separate the analyzed soil types using classification trees. The accuracy of classification based on the results of validation varies from 63–100%. At the same time, the parameters of the HSI and YMC color systems turned out to be more informative than the parameters of the RGB color system. The established classification rules can later be used to determine the classification position of soils from images collected using crowdsourcing technologies.

Keywords: soil inventory, crowdsourcing, remote data, decision trees.

ВВЕДЕНИЕ

Краудсорсинг (англ. crowdsourcing, от crowd – толпа и sourcing – использование или привлечение ресурсов) – привлечение к решению тех или иных проблем инновационной производственной деятельности широкого круга лиц на добровольных началах с применением информационных технологий. Среди краудсорсинговых проектов встречаются проекты для голосования и сбора идей, по обработке научных материалов, конкурсные платформы, краудсорсинговые агрегаторы, справочные ресурсы, проекты по совместному использованию данных, различные формы совместной работы с контентом и поддержки коллективных инноваций. Одним из примеров известных краудсорсинговых проектов является Википедия.

Также фотографии, размещаемые в Google Earth¹ пользователями, являются примером краудсорсинговой базы данных географически привязанных изображений различных объектов (в том числе почв), которые можно использовать для их анализа и оценки.

Изображение пахотных почв и посевов является источником информации об их свойствах. Изображения, получаемые со спутников или с помощью БПЛА, достаточно широко используются

¹ <https://www.google.ru/intl/ru/earth/>

для оценки и мониторинга почв и их свойств ([Chen et al., 2000](#); [Савин, Симакова, 2012](#); [Dou et al., 2019](#)). Фотографии поверхности, сделанные в поле, применяются для оценки проективного покрытия почвы растительностью (посевами, сорняками) и состояния открытой поверхности пахотных почв ([Prudnikova, Savin, 2021](#)).

По цифровым фотографиям Пузаченко и др. ([2004](#)) анализировали строение почвенного профиля. Zhang и Hartemink ([2019](#)) использовали цифровые изображения для выделения почвенных горизонтов. Ху и др. ([2019](#)) применяли фотографии для определения засоления и неоднородности поверхности почв. Levin и др. ([2005](#)) изучали возможность применения цифровых фото для оценки содержания оксида железа и илестых и глинистых частиц в песчаных почвах полупустынь. Persson ([2005](#)) показал возможность применения цифровых фото для определения влажности почв. V. Rossel и др. ([2008](#)) оценивали возможность применения цифровых фото для определения содержания органического углерода и оксида железа.

Также ведутся исследования по направлению использования мобильных телефонов и их фотокамер для оперативной оценки свойств почв. В работе Chodium и др. ([2013](#)) изучалась возможность применения iPhone для количественного анализа почвы на содержание взрывчатых веществ. Aitkenhead (2013)² продемонстрировал приложение для смартфона, связывающее камеру, анализ изображений и их обработку, для оценки содержания углерода в почве. Нап и др. ([2016](#)) предложили новый сенсор оценки почвенного цвета, основанный на смартфоне, для определения типа почв.

Несмотря на перспективность использования цифровых фотографий почв для анализа их различных характеристик, на сегодняшний день отсутствуют подходы, позволяющие использовать для этих целей краудсорсинговые данные. Поскольку получаемые в рамках краудсорсинговых технологий данные отличаются высокой неоднородностью (разные фотокамеры, разные условия съем-

² <http://www.hutton.ac.uk/research/groups/information-and-computational-sciences/esmart>

ки (освещенность, угол съемки)), необходима разработка подхода, учитывающего эту неоднородность.

Целью исследования на данном этапе было установить информативные параметры спектральной отражательной способности почв (СОС), которые инвариантны к разным условиям съемки, но позволяют наиболее точно определять тип почвы, а также установить пороговые значения информативных параметров для трех типов почв (дерново-подзолистой, серой лесной и чернозема) на основе компьютерного анализа фотографий их поверхности.

ОБЪЕКТЫ И МЕТОДЫ

Объект исследования

Объектом исследования выступает спектральная отражательная способность открытой поверхности пахотных почв тестовых участков и их изображение на фотографиях, полученных обычными фотокамерами. Тестовые участки расположены в Тульской, Московской и Тверской областях. Почвы тестовых участков – дерново-подзолистые, серые лесные, черноземы выщелоченные.

Стратегия исследования

В основе краудсорсинговой технологий оперативной инвентаризации пахотных почв лежит характер их изображения на фотографиях, а также их СОС в оптическом диапазоне, которые связаны с их свойствами ([Rossel et al., 2006](#)).

При этом используются фотографии, полученные в разных условиях съемки и для разного состояния поверхности: при ярком освещении, при темном освещении, под разными углами, для трансформированной поверхности, для свежеспаханной поверхности. Поскольку при съемке СОС в поле прибор калибруется на условия освещения с помощью специальной эталонной панели, спектральные кривые не зависят от условий освещенности в отличие от фотографий.

Фотографии используются для расчетов различных параметров изображения для почв и посевов. Те же параметры рассчитываются для кривых СОС. Затем с помощью дерева классификации выбираются параметры, информативные для определения классификационного положения почв, а также устанавливаются

правила классификации. При этом параметры, полученные по фотографиям и по кривым СОС, анализируются совместно. Таким образом выявляются информативные спектральные параметры, которые инвариантны к разным условиям съемки, но позволяют наиболее точно определять тип почвы.

Полевые работы

В 2021 г. в ходе полевых выездов собиралась следующая информация: спектры открытой поверхности пахотных почв и фотографии поверхности (табл. 1).

Таблица 1. Собранные полевые данные
Table 1. Collected field data

Тип почвы	Количество точек с фото	Количество спектров	Общее количество точек
Дерново-подзолистая почва	13	25	38
Серая лесная почва	10	35	45
Чернозем	10	10	20

Съемка СОС проводилась спектрорадиометром HandHeld-2, работающим в диапазоне 325–1 075 нм. Прибор располагался перпендикулярно измеряемой поверхности на расстоянии около 150 см. Съемка в каждой точке проводилась в 5-кратной повторности. Измерения СОС при использовании данного прибора проводятся с использованием яркости, отраженной от образца, и яркости, отраженной от белой эталонной панели со 100%-ной отражающей способностью.

Съемка поверхности проводилась разными фотокамерами в разных условиях (под разными углами и разным направлением падения солнечных лучей, в разных условиях освещения, с разного расстояния), для разного состояния поверхности почв (свежевспаханная, с посевами, со стерней) для того, чтобы максимально полно имитировать именно краудсорсинговый путь сбора исходных данных.

Анализ фотографий поверхности пахотных почв

Полученные изображения анализировались в программе ILWIS Academic 3.3. Сначала был проведен визуальный анализ фотографий, по результатам которого были замаскированы участки с посторонними объектами, тенью от людей и посторонних предметов (при их наличии). Далее с помощью операции Color separation изображения были разложены на каналы в цветовых системах RGB (RED, GREEN, BLUE), YMC (YELLOW, MAGENTA, CYAN), HSI (HUE, SATURATION, INTENSITY), и канал GRAY.

Пересчет из системы RGB в YMC, HSI и канал GRAY проводился по следующим формулам:

$$\text{Yellow} = 255 - \text{Blue} \quad (1)$$

$$\text{Magenta} = 255 - \text{Green} \quad (2)$$

$$\text{Cyan} = 255 - \text{Red} \quad (3)$$

$$\text{Hue} = (255/2\pi) * \arctan2(1/2 \sqrt{3} * (\text{Green} - \text{Blue}), \quad (4)$$

$$\text{Red} - (\text{Green} + \text{Blue}) / 2) * 240/255$$

$$\text{Saturation} = \sqrt{(\text{Red}^2 + \text{Green}^2 + \text{Blue}^2 - \text{Red} * \text{Green} - \text{Red} * \text{Blue} - \text{Green} * \text{Blue})} * 240/255 \quad (5)$$

$$\text{Intensity} = 1/3 * (\text{Red} + \text{Green} + \text{Blue}) * 240/255 \quad (6)$$

$$\text{Gray} = 0.3 * \text{Red} + 0.59 * \text{Green} + 0.11 * \text{Blue} \quad (7)$$

После этого каждое изображение было классифицировано на 2 основных класса: “почва” и “другое” (посевы, листья и прочее) с помощью классификации с обучением.

Выбор комбинации каналов для классификации каждого изображения осуществлялся на основе визуальной оценки качества разделения классов при сочетании разных каналов. Чаще всего для классификации использовалась комбинация каналов RED, GREEN, BLUE.

По завершении классификации для каждого изображения определялось среднее значение отражения в каждом канале для класса “почва”. Далее полученные значения переносились в MS

Excel, где дополнительно рассчитывались отношения между каналами (45 параметров).

Предварительная обработка и анализ данных о СОС

Для каждой точки на основе 5 повторностей рассчитывалась средняя спектральная кривая. После чего проводилась фильтрация данных о СОС методом Савицкого-Голея. Фильтрация осуществлялась в программе R с помощью функции “*savitzkyGolay*” пакета “*prospectr*”³.

Затем СОС пересчитывалась в те же параметры, которые рассчитывались на основе анализа фотографий, включая 45 соотношений между каналами. Сначала проводился пересчет в отражение в RGB с помощью пакетов “*pavo*”⁴, “*grDevices*”⁵ программы R, а затем расчет остальных параметров по формулам, представленным выше (формулы 1–7).

Дисперсионный анализ влияния типа почв на значения отражения в различных цветовых системах

Для определения существования статистически значимых различий между параметрами отражения, рассчитанными в разных цветовых системах, для разных типов почв использовался однофакторный дисперсионный анализ. Для дисперсионного анализа использовалась функция “*aov*” пакета “*stats*”⁶ программы R.

В случае наличия статистически значимых различий далее проводился апостериорный анализ. Использовался Критерий Тьюки достоверно значимой разности, который рассчитывался с помощью функции “*TukeyHSD*” пакета “*stats*” программы R.

В результате для каждого параметра отражения в разных цветовых системах устанавливалось наличие или отсутствие статистически значимого влияния типа почв, а также при наличии такого влияния – типы почвы, значимо различающиеся между собой по величине данного параметра.

³ <https://cran.r-project.org/web/packages/prospectr/index.html>

⁴ <https://cran.r-project.org/web/packages/pavo/index.html>

⁵ <https://www.rdocumentation.org/packages/grDevices/versions/3.6.2>

⁶ <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>

Определение типа почв с помощью дерева классификации

При определении типа почв с помощью дерева классификации в основе лежали параметры отражения почв, рассчитанные в разных цветовых системах. Для построения дерева классификации использовался пакет *“rpart”*⁷ программы R.

Предварительно до построения дерева данные были разделены на обучающую и тестовую выборки в соотношении 70/30. Для этого использовалась функция *“createDataPartition”* пакета *“caret”*⁸. Дерево строилось на обучающей выборке и проверялось на тестовой.

Для каждого дерева классификации устанавливался свой параметр сложности *ср*. Параметр *ср* определялся с помощью функции *“train”* пакета *“caret”* в диапазоне от 0.01 до 0.5 с шагом 0.01. Функция *“train”* устанавливает сетку параметров настройки для классификации, подгоняет модель и вычисляет качество модели на основе ресамплинга. В результате для каждого значения *ср* рассчитывается точность модели. В итоге выбирается такое значение параметра сложности, при котором получается модель с наибольшей точностью. Определение значения оптимального параметра сложности проводится на обучающей выборке.

Точность классификации оценивалась как отношение количества точек тестовой выборки с правильно определенным типом почв к общему числу точек тестовой выборки, выраженное в процентах. Также строилась матрица ошибок классификации, которая показывает какое количество точек каждого типа было классифицировано верно и какое количество было отнесено к другим типам.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Спектральная отражательная способность почв в разных цветовых системах

⁷ <https://www.rdocumentation.org/packages/rpart/versions/4.1-15/topics/rpart>.

⁸ <https://cran.r-project.org/web/packages/caret/vignettes/caret.html>.

В целом, наибольшие значения отражения отмечаются в системе YMC, а наименьшие в канале SATURATION системы HSI (рис. 1).

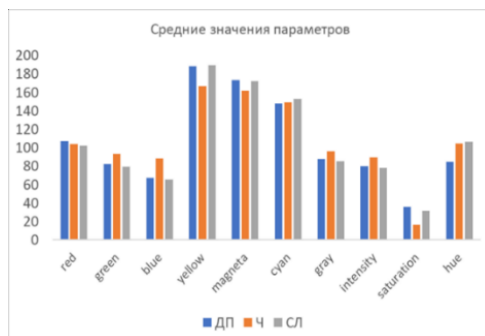
При этом идет снижение средних значений отражения в ряду RED – GREEN – BLUE системы RGB, YELLOW – MAGNETA – CYAN системы YMC, HUE – INTENSITY – SATURATION системы HSI. Отражение в канале GRAY близко к отражению в канале INTENSITY.

В каналах RED, GREEN и GRAY в случае спектров наибольшим отражением характеризуется дерново-подзолистая почва, в то время как по результатам анализа фотографий в этом канале сильнее всего отражает чернозем. В канале BLUE по спектрам три анализируемых типа почвы имеют очень близкое среднее отражение, в то время как по фотографиям прослеживается достаточно сильная разница. При этом наибольшее среднее отражение также отмечается для чернозема.

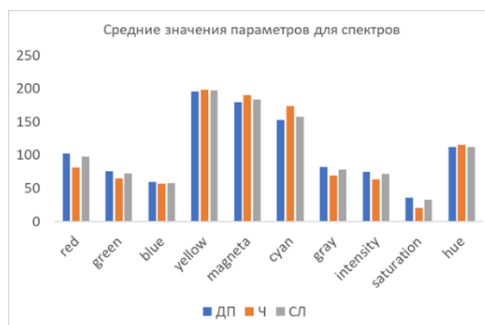
В цветовой системе YMC при анализе фото чернозем имеет в основном наименьшее среднее значение отражения, а дерново-подзолистая почва – наибольшее, за исключением канала MAGNETA, в котором наименьшим средним отражением характеризуется серая лесная почва. При анализе спектров в канале YELLOW все три анализируемых типа почв характеризуются близким отражением, в остальных каналах системы YMC выщелоченный чернозем имеет наибольшее среднее значение отражения, а среднее отражение дерново-подзолистой и серой лесной почвы практически одинаковое.

В цветовой системе HSI при анализе спектров чернозем характеризуется наименьшим значением среднего отражения в каналах INTENSITY и SATURATION, в то время как дерново-подзолистая и серая лесная почвы имеют близкие значения отражения. В канале HUE анализируемые типы почв имеют близкие средние значения отражения. Что касается фото, в каналах INTENSITY и HUE чернозем характеризуется наибольшим средним отражением, а дерново-подзолистая почва – наименьшим. В канале SATURATION отмечаются те же закономерности, что и в случае спектров.

а)



б)



в)



Рис. 1. Средние значения параметров в разных цветовых системах для разных типов почв: ДП – дерново-подзолистая почва, СЛ – серая лесная почва, Ч – чернозем. По оси ОУ – значения отражения.

Fig. 1. Average values of parameters in different color systems for different soil types: ДП – sod-podzolic soil, СЛ – grey forest soil, Ч – leached chernozem. Along OY axe – spectral reflectance values.

Дисперсионный анализ показал, что статистически значимо СОС анализируемых типов почв различается в канале BLUE цветовой системы RGB, в канале YELLOW цветовой системы YMC и каналах HUE и SATURATION системы HSI (табл. 2).

Таблица 2. Результаты дисперсионного анализа по оценке влияния типа почв на их СОС в разных цветовых системах

Table 2. Results of analysis of variance on the assessment of the influence of soil type on their spectral reflectance in different color systems

Параметры отражения	Df	Sum_sq	Mean_sq	F	p-value
RED	2	562	281.1	0.445	0.642
GREEN	2	1 965	982.4	1.483	0.232
BLUE	2	6 695	3347	5.762	0.004
YELLOW	2	6 578	3289	5.578	0.005
MAGNETA	2	1 240	620.2	0.798	0.453
CYAN	2	470	235.2	0.372	0.69
HUE	2	11 263	5631	6.458	0.023
SATURATION	2	5 162	2581	51.46	<0.0001
INTENSITY	2	1 221	610.4	1.09	0.34
GRAY	2	1 063	531.3	0.852	0.429

Примечание. Df – число степеней свободы; Sum_sq – сумма квадратов (разница между средним значением для группы и общим средним); Mean_sq – среднее значение суммы квадратов, вычисленное делением суммы квадратов на степень свободы для каждого параметра; F – статистика теста из F-теста; p-value – уровень статистической значимости F-статистики.

Note. Df – degrees of freedom; Sum_sq – sum of squares (the difference between group and total means); Mean_sq – mean sum of squares, calculated as a ratio of sum of squares to the degrees of freedom for each parameter; F – statistics for F-test; p-value – the level of statistical significance for F-statistics.

Апостериорный тест показал, что в случае отражения в канале BLUE системы RGB и отражения в канале YELLOW цветовой системы YMC статистически значимые различия наблюдались между дерново-подзолистой почвой и черноземом (p-value 0.01 в обоих случаях) и между черноземом и серой лесной почвой (p-

value 0.004 и 0.005 соответственно), в то время как серая лесная и дерново-подзолистые почвы не различались между собой (p-value 0.94 и 0.95 соответственно).

В случае отражения в канале HUE цветовой системы HSI статистически значимые различия наблюдались между дерново-подзолистой почвой и черноземом (p-value 0.04) и между дерново-подзолистой и серой лесной почвой (p-value 0.003), в то время как чернозем и серые лесные почвы не различались между собой (p-value 0.98).

Для канала SATURATION цветовой системы HSI наблюдались те же закономерности, что и в случае отражения в канале BLUE системы RGB и отражения в канале YELLOW цветовой системы YMC. Статистически значимые различия наблюдались между дерново-подзолистой почвой и черноземом (p-value < 0.0001) и между черноземом и серой лесной почвой (p-value < 0.0001), в то время как серая лесная и дерново-подзолистые почвы не различались между собой (p-value 0.05).

Определение типа почв на основе их СОС в разных цветовых системах с помощью дерева классификации

Для разделения дерново-подзолистой почвы и чернозема информативным оказалось отношение SATURATION/HUE. Классификация осуществляется за один шаг: в случае если данное отношение меньше 0.26, то почвы относят к черноземам; если больше или равно – к дерново-подзолистым. По результатам валидации точность классификации составила 100%.

Для разделения чернозема и серой лесной почвы информативным было соотношение YELLOW/MAGNETA. Классификация осуществляется за один шаг: в случае если данное соотношение меньше 1.035, то почвы относят к черноземам, если больше – к серым лесным.

По результатам валидации точность классификации составила 89.5% – две точки с серой лесной почвой были ошибочно отнесены к черноземам (табл. 3).

Таблица 3. Матрица ошибок по результатам валидации для дерева классификации для чернозема и серой лесной почвы

Table 3. Error matrix based on validation results for the classification tree for chernozem and gray forest soil

		Предсказанные	
		Ч	СЛ
Наблюдаемые	Ч	4	2
	СЛ	0	13

Для разделения серой лесной и дерново-подзолистой почвы информативными оказались два параметра: отражение в канале RED и соотношение MAGNETA/HUE. Классификация осуществляется в два шага. Сначала используется соотношение MAGNETA/HUE. В случае если оно меньше 1.55, то почва классифицируется как серая лесная, если больше или равно 1.55, то классификация продолжается дальше на основе отражения в канале RED. Если оно больше или равно 83.5, то почву относят к дерново-подзолистой, если меньше – к серой лесной.

Точность классификации по результатам валидации составила 71% (табл. 4). Ошибки возникают на обоих этапах классификации. На первом шаге при анализе соотношения MAGNETA/HUE три точки с дерново-подзолистой почвой были ошибочно отнесены к серой лесной почве. Все ошибки сделаны при классификации спектров. На втором этапе при анализе параметра RED четыре точки с серой лесной почвой были ошибочно отнесены к дерново-подзолистой: из них три ошибки возникли при классификации спектров, и одна – при классификации фото.

Для разделения трех типов почв информативными оказались четыре параметра: соотношение SATURATION/HUE, RED, RED/GREEN и CYAN/HUE. Полученное дерево классификации на первом шаге использует соотношение SATURATION/HUE (рис. 2). В случае если данное соотношение <0.25 , для дальнейшего разделения используется параметр RED, если ≥ 0.25 – соотно-

шение CYAN/HUE. В случае если $RED \geq 94.5$, то почвы классифицируют как черноземы, если данный параметр < 94.5 , – как серые лесные. Если соотношение $CYAN/HUE < 1.3$, то почвы также классифицируют как серые лесные, если оно ≥ 1.3 , то для дальнейшего разделения используется параметр $RED/GREEN$. Если он больше 1.4, то почвы классифицируются как серые лесные, если меньше – как дерново-подзолистые.

Таблица 4. Матрица ошибок по результатам валидации для дерева классификации для дерново-подзолистой и серой лесной почвы

Table 4. Error matrix based on validation results for the classification tree for sod-podzolic and gray forest soil

		Предсказанные	
		ДП	СЛ
Наблюдаемые	ДП	8	3
	СЛ	4	9

Точность классификации по результатам валидации составила 63% (табл. 5). Пять точек с серой лесной почвой (три фото и два спектра) были классифицированы как дерново-подзолистые, а три точки с дерново-подзолистой почвой (спектры) были классифицированы как серая лесная почва. Кроме того, три точки с черноземом (спектры) были ошибочно отнесены к серой лесной почве. Таким образом, черноземы и дерново-подзолистые почвы разделяются без ошибок, как и в случае построения дерева классификации только для этих двух типов почв. В то же время в процессе классификации серая лесная почва смешивается с дерново-подзолистой, а дерново-подзолистая и чернозем – смешиваются с серой лесной.

Таким образом, при построении деревьев классификации для двух (в разных комбинациях) или всех трех анализируемых типов почв чаще всего использовался параметр HUE цветовой системы HSI в составе разных спектральных отношений.

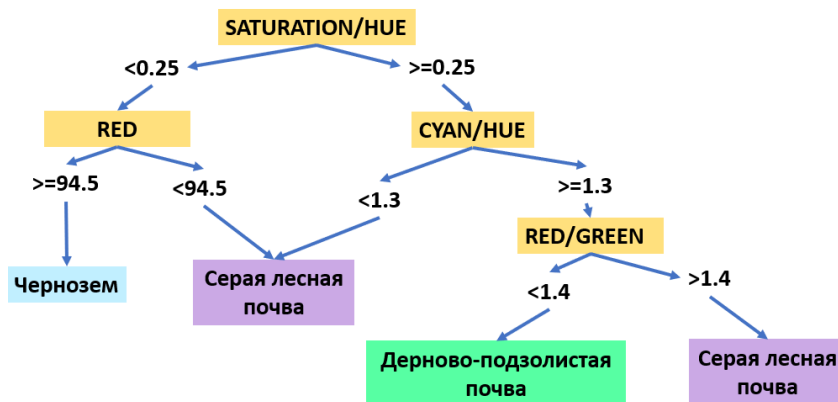


Рис. 2. Дерево классификации для разделения трех анализируемых типов почв на основе рассчитанных параметров отражения в разных цветовых системах.

Fig. 2. Classification tree for three analysed soil types based on spectral parameters in different color systems.

Таблица 5. Матрица ошибок по результатам валидации для дерева классификации для анализируемых типов почв

Table 5. Error matrix based on validation results for the classification tree for analysed soil types

		Предсказанные		
		ДП	Ч	СЛ
Наблюдаемые	ДП	8	0	3
	ДП	0	3	3
	ДП	5	0	8

По значениям данного параметра между собой статистически достоверно различаются дерново-подзолистые и черноземы, а также дерново-подзолистые и серые лесные (табл. 2). Спектральное отношение SATURATION/HUE оказалось информативным для двух деревьев из четырех построенных: для разделения чернозема и дерново-подзолистой почвы и для разделения всех анализируемых типов почв. По значениям параметра SATURATION достоверно различаются между собой дерново-подзолистые почвы и черноземы, а также черноземы и серые лесные почвы. Таким образом, чаще всего в процессе классификации использовались именно параметры цветовой системы HSI (кроме INTENSITY).

На втором месте по частоте использования параметры цветовой системы YMC (спектральное отношение YELLOW/MAGNETA), в том числе в сочетании с параметром HUE цветовой системы HSI (спектральные отношения MAGNETA/HUE, CYAN/HUE). При этом статистически значимые отличия между анализируемыми типами почв (дерново-подзолистыми почвами и черноземами, черноземами и серыми лесными) среди параметров цветовой системы YMC отмечались только для параметра YELLOW.

Параметры цветовой системы RGB использовались при классификации реже всего, из них чаще всего использовалось отражение в канале RED: в двух случаях как отдельный параметр, и в одном случае в составе спектрального отношения RED/GREEN. При этом ни для параметра RED, ни для параметра GREEN не отмечалось статистически достоверных различий между анализируемыми типами почв. Данные параметры использовались в деревьях классификации на втором или третьем шаге классификации.

Параметры цветовых систем HSI и YMC оказались более информативными для разделения анализируемых типов почв, по сравнению с параметрами цветовой системы RGB, которые в основном и используются для оценки почв по фотографиям их поверхности. При этом следует отметить, что параметры цветовых систем HSI и YMC рассчитываются на основе параметров цветовой системы RGB (формулы 1–7), то есть являются ее производными. Таким образом, осуществляемые математические преобразования значений отражения в каналах RED, GREEN и BLUE для

получения значения параметров в других цветовых системах, а также последующий расчет спектральных отношений позволили усилить разницу в спектральных характеристиках между анализируемыми типами почв и тем самым повысить их информативность для дальнейшей классификации.

Точность детектирования почв, по-видимому, можно повысить и путем задания строгих условий съемки (например, снимать только в надир, только с высоты 1 метр, только в полдень, только на фотокамеру с широкоформатным объективом и т. п.). Но это сделает сбор подобных данных более ограниченным и не позволит включать в анализ фотографии, полученные в прошлые годы и в других странах мира.

ВЫВОДЫ

На данном этапе разработки предлагаемый подход к распознаванию пахотных почв на основе краудсорсинговых данных позволяет определять типовую принадлежность почв по изображениям их поверхности с точностью 63–100%.

Информативными параметрами для классификации могут выступать такие параметры как соотношения SATURATION/HUE, YELLOW/MAGNETA, CYAN/HUE, MAGNETA/HUE, RED/GREEN и отражение в канале RED.

Лучше всего разделяются дерново-подзолистая почва и чернозем, достаточно высока точность классификации и при разделении чернозема и серой лесной почвы. Наименее точная классификация отмечается при разделении трех типов почв. При этом чаще всего серые лесные почвы смешиваются с дерново-подзолистыми.

Привлечение большего количества фотографий и определение наиболее благоприятных для классификации условий съемки, а также фиксирование типа камеры и условий съемки, скорее всего, позволят повысить точность разделения почв различных типов.

Установленные правила классификации в дальнейшем могут применяться для определения классификационного положения почв по изображениям, собранным с помощью краудсорсинговых технологий. Полученная при этом информация может быть использована для обновления почвенных карт.

СПИСОК ЛИТЕРАТУРЫ

1. Савин И.Ю., Симакова М.С. Спутниковые технологии для инвентаризации и мониторинга почв в России // Современные проблемы дистанционного зондирования Земли из космоса. 2012. Т. 9. № 5. С. 104–115.
2. Chen F., Kissel D.E., West L.T., Adkins W. Field-scale mapping of surface soil organic carbon using remotely sensed imagery // Soil Science Society of America Journal. 2000. Vol. 64. No. 2. P. 746–753.
3. Choodum A., Kanatharana P., Wongniramaikul W., Nic Daeid N. Using the iPhone as a device for a rapid quantitative analysis of trinitrotoluene in soil // Talanta. 2013. Vol. 115. P. 143–149.
4. Dou X., Wang X., Liu H., Zhang X., Meng L., Pan Y., Yu Z., Cui Y. Prediction of soil organic matter using multi-temporal satellite images in the Songnen Plain, China // Geoderma. 2019. Vol. 356. P. 113896.
5. Han P., Dong D., Zhao X., Jiao L., Lang, Y. A smartphone-based soil color sensor: For soil type classification // Computers and Electronics in Agriculture. 2016. Vol. 123. P. 232–241.
6. Levin N., Ben-Dor E., Singer A. A digital camera as a tool to measure colour indices and related properties of sandy soils in semi-arid environments // International Journal of Remote Sensing. 2005. Vol. 26. P. 5475–5492.
7. Persson M. Estimating surface soil moisture from soil color using image analysis // Vadose Zone Journal. 2005. Vol. 4. P. 1119–1122.
8. Prudnikova E., Savin I. Some peculiarities of arable soil organic matter detection using optical remote sensing data // Remote Sensing. 2021. Vol. 13. No. 12. P. 2313.
9. Puzachenko Yu.G., Aleshchenko G.M., Puzachenko M.Yu., Kozlov D.N. Soil structure analysis with the use of digital color images // Eurasian Soil Science. 2004. Vol. 37. No. 2. P. 109–121.
10. Rossel R.V., Fouad Y., Walter C. Using a digital camera to measure soil organic carbon and iron contents // Biosystems Engineering. 2008. Vol. 100. P. 149–159.
11. Rossel R.V., Walvoort D.J.J., McBratney A.B., Janik L.J., Skjemstad J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties // Geoderma. 2006. Vol. 131. No. 1–2. P. 59–75.
12. Xu L., Zheng C., Wang Z., Nyongesah M.J. A digital camera as an alternative tool for estimating soil salinity and soil surface roughness // Geoderma. 2019. Vol. 341. P. 68–75.
13. Zhang Y., Hartemink A.E. A method for automated soil horizon delineation using digital images // Geoderma. 2019. Vol. 343. P. 97–115.

REFERENCES

1. Savin I. Yu., Simakova M. S., Sputnikovye tekhnologii dlya inventarizatsii i monitoringa pochv v Rossii (Satellite technologies for inventory and monitoring of soils in Russia), *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2012, Vol. 9, No. 5, pp. 104–115.
2. Chen F., Kissel D.E., West L.T., Adkins W., Field-scale mapping of surface soil organic carbon using remotely sensed imagery, *Soil Science Society of America Journal*, 2000, Vol. 64, No. 2, pp. 746–753.
3. Choodum A., Kanatharana P., Wongniramaikul W., Nic Daeid N., Using the iPhone as a device for a rapid quantitative analysis of trinitrotoluene in soil, *Talanta*, 2013, Vol. 115, pp. 143–149.
4. Dou X., Wang X., Liu H., Zhang X., Meng L., Pan Y., Yu Z., Cui Y., Prediction of soil organic matter using multi-temporal satellite images in the Songnen Plain, China, *Geoderma*, 2019, Vol. 356, p. 113896.
5. Han P., Dong D., Zhao X., Jiao L., Lang, Y., A smartphone-based soil color sensor: For soil type classification, *Computers and Electronics in Agriculture*, 2016, Vol. 123, pp. 232–241.
6. Levin N., Ben-Dor E., Singer A., A digital camera as a tool to measure colour indices and related properties of sandy soils in semi-arid environments, *International Journal of Remote Sensing*, 2005, Vol. 26, pp. 5475–5492.
7. Persson M., Estimating surface soil moisture from soil color using image analysis, *Vadose Zone Journal*, 2005, Vol. 4, pp. 1119–1122.
8. Prudnikova E., Savin I., Some peculiarities of arable soil organic matter detection using optical remote sensing data, *Remote Sensing*, 2021, Vol. 13, No. 12, p. 2313.
9. Puzachenko Yu.G., Aleshchenko G.M., Puzachenko M.Yu., Kozlov D.N., Soil structure analysis with the use of digital color images, *Eurasian Soil Science*, 2004, Vol. 37, No. 2, pp. 109–121.
10. Rossel R.V., Fouad Y., Walter C., Using a digital camera to measure soil organic carbon and iron contents, *Biosystems Engineering*, 2008, Vol. 100, pp. 149–159.
11. Rossel R. V., Walvoort D. J. J., McBratney A. B., Janik L. J., Skjemstad J. O., Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties, *Geoderma*, 2006, Vol. 131, No. 1–2, pp. 59–75.
12. Xu L., Zheng C., Wang Z., Nyongesah M.J., A digital camera as an alternative tool for estimating soil salinity and soil surface roughness, *Geoderma*, 2019, Vol. 341, pp. 68–75.
13. Zhang Y., Hartemink A.E., A method for automated soil horizon delineation using digital images, *Geoderma*, 2019, Vol. 343, pp. 97–115.