

ИСПОЛЬЗОВАНИЕ МАССИВОВ БОЛЬШИХ ДАННЫХ ДЛЯ ОПРЕДЕЛЕНИЯ ПОКАЗАТЕЛЕЙ СВОЙСТВ ПОЧВ

Д. А. Азиков

Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М.В. Ломоносова», Москва, Россия, azi-inna@yandex.ru

Аннотация

Цифровая революция затронула методологическую и методическую базы всех проводимых в настоящее время научных исследований. Интерес к созданию спектральных библиотек во всём мире увеличивается, так как спектральный анализ может быть использован в качестве ёмкого и не деструктивного метода, результаты которого коррелируют со многими почвенными свойствами. Спектральная библиотека (SSL - Soil Spectra libraries) представляет собой базу данных, в которой представлены эталонные образцы спектров почв с описанием их химических и физических свойств.

Ключевые слова: почвенный покров, почвенные горизонты, почвенный профиль, цифровая почвенная морфометрия, базы данных, спектральные характеристики, цвет почв, CIELAB, SSL, LUCAS.

USING BIG DATA TO DETERMINE SOIL PROPERTIES INDICATORS

D. A. Azikov

Abstract

The digital revolution has affected the methodological and methodological bases of all ongoing scientific research. Interest in the creation of spectral libraries around the world is increasing, since spectral analysis can be used as a capacious and non-destructive method, the results of which correlate with many soil properties. The Soil Spectra Libraries (SSL) is a database that contains reference samples of soil spectra with a description of their chemical and physical properties.

Keywords: soil scape, soil horizons, soil profile, digital soil morphometry, data bases, spectral characteristics, soil color, CIELAB, SSL, LUCAS.

Введение. Современный этап развития почвоведения связан с формированием баз данных, которые позволяют систематизировать как уже накопленные ранее сведения, так и организовать сбор новой информации по установленным протоколам и единой схеме. Интерес к созданию спектральных библиотек во всем мире увеличивается, что связано как с развитием дистанционных методов исследования почв, так и с привлекательностью спектрального анализа в качестве не деструктивного, информационно и ёмкого метода, результаты которого коррелируют со многими почвенными свойствами.

Примером такой базы данных служит европейская спектральная библиотека (The European spectral library) LUCAS, которая содержит данные 23 стран Европы о спектрах более, чем 20 тыс. образцов верхнего слоя почвы и связанных с ними основными почвенными свойствами. Данные в этой спектральной библиотеке получены за два периода: 2009 и 2015 годы [16]. Это позволяет прово-

дить анализ с учетом времени и регистрировать тенденции в изменении состояния почвенных свойств. Эти опубликованные данные находятся в открытом доступе и представляют возможность работать с ними зарегистрированным пользователям.

Анализ данной библиотеки проводится по трем главным направлениям. Первое – определение математических методов (моделей), наиболее эффективных, с точки зрения предсказательной способности почвенных свойств, для глобального массива данных. Второе – установить, насколько будут меняться модели при переходе от глобальных данных к локальным. Третье направление, специфичное для России, связано с тем, что в нашей стране в результате многолетней работы почвоведов накопилось большое количество изображений почвенных спектров отражения, которые удалось собрать в базу данных “Спектры отражения и цифровые показатели цвета почв” [4]. Первые полученные результаты анализа данных показали, что применение некоторых показателей цветности, не входящих пока в международный научный оборот, позволяет проводить диагностику почвенных свойств с большей эффективностью, чем применяемые в настоящее время стандартные показатели.

Материалы и методы. На первом этапе производилась предварительная обработка графических файлов. Сканированные или сфотографированные изображения спектральных кривых подвергали предварительной очистке, с целью получения только одного изображения спектра в файле *.bmp, для чего использовали программу Paint.Net. Полученное изображение сохраняли в формате *.bmp с размерами, точно соответствующими внутреннему полю первоначального файла, ограниченному осями X (длина волны) и Y (спектральное отражение, %).

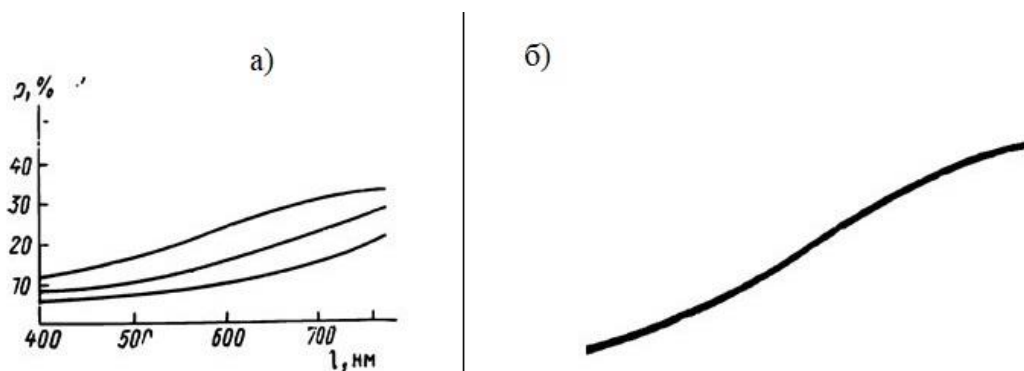


Рис. 10. Спектральные кривые гумусных горизонтов

5*

131

Рисунок 1 – Изображение до (а) и после обработки в графическом редакторе (б)

Рисунок сохраняли с новым названием. Для примера, рисунок 1 имел название 00079_00_50_400_750.bmp. Формирование названия файла соответствовало алгоритму, в соответствии с которым первые 5 знаков (00079) обозначали уникальный номер спектра, следующие 4 знака – точки начала (00) и конца координат (50) по оси Y, затем начала (400) и конца (750) координат по оси X (длина волны).

Информацию о каждом спектре заносили в таблицу (таблица 1, фрагмент).

Таким образом, формировали два потока информации – графический (файл в формате *.bmp) и тематический (таблица в *.xls), связанные друг с другом по уникальному названию файла.

Таблица 1 – Информация о спектральных кривых, вошедшая в базу данных (фрагмент)

Файл	Источник литературы	Стр.	Рис.	Линия	Сведения о горизонте	Сведения о таксономии	Инструмент измерения	Геометрия
0007 9_00 _50_ 400_ 750.b mp	Д. С. Орлов и др. Почвенно-экологический мониторинг. 1994	131	10	1	Гумусный горизонт 1	-	Спектрофотометр	Собирающая сфера
0008 0_00 _50_ 400_ 750.b mp	Д. С. Орлов и др. Почвенно-экологический мониторинг. 1994	131	10	2	Гумусный горизонт 2	-	Спектрофотометр	Собирающая сфера
0008 1_00 _50_ 400_ 750.b mp	Д. С. Орлов и др. Почвенно-экологический мониторинг. 1994	131	10	3	Гумусный горизонт 3	-	Спектрофотометр	Собирающая сфера
0008 2_00 _70_ 400_ 750.b mp	Д. С. Орлов и др. Почвенно-экологический мониторинг. 1994	131	11	1	Осветленный горизонт 1	-	Спектрофотометр	Собирающая сфера
0008 3_00 _70_ 400_ 750.b mp	Д. С. Орлов и др. Почвенно-экологический мониторинг. 1994	131	11	2	Осветленный горизонт 2	-	Спектрофотометр	Собирающая сфера
0008 4_00 _70_ 400_ 750.b mp	Д. С. Орлов и др. Почвенно-экологический мониторинг. 1994	131	11	3	Осветленный горизонт 3	-	Спектрофотометр	Собирающая сфера

0008 5_00 _50_ 400_ 750.b mp	Д. С. Орлов и др. Почвенно-экологический мониторинг. 1994	132	12	1	Ожелезненный горизонт 1	-	Спектрофотометр	Собирающая сфера
0008 6_00 _50_ 400_ 750.b mp	Д. С. Орлов и др. Почвенно-экологический мониторинг. 1994	132	12	2	Ожелезненный горизонт 2	-	Спектрофотометр	Собирающая сфера
0008 7_00 _50_ 400_ 750.b mp	Д. С. Орлов и др. Почвенно-экологический мониторинг. 1994	132	12	3	Ожелезненный горизонт 3	-	Спектрофотометр	Собирающая сфера

На втором этапе проводилось векторно-цифровое преобразование изображений в программе SAGA GIS. Целью данного этапа было получение цифровой информации о спектре в формате *.csv. Каждая таблица содержала данные в 3 полях: n – номер точки на графике; x – длина волны; y – отражение, %. Процесс обработки в геоинформационной системе SAGA GIS [13] был подразделен на несколько этапов, которые будут последовательно описаны на примере файла 00079_00_50_400_750.bmp.

1. В результате открытия файла в программе образуется растровое изображение в виде грида с расширением *.sgrd - внутренний формат растров в SAGA (00079_00_50_400_750.sgrd).

2. Поводится векторизация растрового изображения (Shapes -> Shapes -> Grid Tool -> Vectorization -> Grid Values to Points). В результате образуется векторный файл формата *.shp. Так как этот этап в преобразованиях промежуточный, файлу присваивается название с префиксом temp (temp_GV1.shp). Данный файл состоит из совокупности белых (фон) и черных (полезный сигнал – спектральные значения) точек. Они различаются по полю атрибута (Z) – это значение, извлеченное из грида. У белых точек Z>0, у черных точек Z=0.

3. На следующем этапе в файле temp_GV1.shp выделяются все фоновые точки, у которых Z>0: Shapes -> Selection -> Select by Attributes (Numerical Expression) -> Z > 0. Затем происходит удаление точек фона: Shapes -> Selection -> Delete Selection from Shapes Layer.

4. Далее в файле temp_GV1.shp проводится преобразование значений контуров спектральных кривых по координате X с помощью вычисления по формуле

$$400 + [X] * (750 - 400)/990, \quad (1)$$

где 400 – начальная координата X (длина волны); X – текущая координата; 750 – конечная координата X; 990 – ширина файла *.bmp в пикселях.

В результате создаются новые данные по точкам 02. Grid Values [Nodes] и файл temp_GV2.shp. Изначальные данные по точкам 01. Grid Values [Nodes] могут быть по усмотрению удалены.

5. Затем в файле temp_GV2.shp проводится преобразование значений контуров спектральных кривых по координате Y с помощью вычисления по формуле

$$0 + [Y] * (50 - 0)/577, \quad (2)$$

где 0 – начальная координата Y (длина волны); Y – текущая координата; 50 – конечная координата Y; 577 – ширина файла *.bmp в пикселях.

В результате вычислений создаются новые данные по точкам 05. Grid Values [Nodes] и файл temp_GV5.shp. Предыдущие данные по точкам 02. Grid Values [Nodes] могут быть по усмотрению удалены.

6. Потом в файле temp_GV5.shp проводится округление значений контуров спектральных кривых по координате X с помощью вычисления по формуле

$$\text{int}([X] + 0.5), \quad (3)$$

где int – функция возвращающая целочисленное значение дробной переменной; X – текущая координата;

В процессе вычисления создаются новые данные по точкам 07. Grid Values [Nodes] и файл temp_GV7.shp. Предыдущие данные по точкам 05. Grid Values [Nodes] могут быть по усмотрению удалены.

7. Далее в файле temp_GV7.shp проводится агрегация значений точек спектральных кривых с помощью атрибутов по полю 1 и статистическому полю 2: Table -> Calculus -> Aggregate Values by Attributes [Table] по координате X. В полях отмечается только значение среднего и создается новая таблица 01. Grid Values [Nodes] [Aggregated: X]. Потом предыдущие данные по точкам 07. Grid Values [Nodes] могут быть по усмотрению удалены. Затем таблица переименовывается в Spec_01nm: Window -> Show Object Properties Window и сохраняется в формате dbf. Результатом этого этапа является получение цифровой информации о спектре объекта с разрешением по длине волны 1 нм.

8. Затем на основе таблицы Spec_01nm.dbf создаётся новая таблица Spec_10nm.dbf: Shapes -> Table -> Field Calculator [Shapes] по столбцу X10. Потом новая созданная таблица Spec_01nm.dbf переименовывается в Spec_10nm.dbf: Window -> Show Object Properties Window. Далее проводится агрегация значений таблицы с помощью атрибутов, при этом происходит добавление столбца X10, по статистическому столбцу Y: Table -> Calculus -> Aggregate Values by Attributes [Table]. В полях отмечается только значение среднего и создается новая таблица 06. Spec_10nm.dbf [Aggregated: X10]. Потом предыдущая таблица 05. Spec_10nm.dbf может быть по усмотрению удалена. И далее название таблицы 06. Spec_10nm.dbf [Aggregated: X10] сокращается до 06. Spec_10nm.dbf: Window -> Show Object Properties Window и сохраняется в формате dbf с заменой предыдущей таблицы с тем же названием. Результатом этого этапа является получение цифровой информации о спектре объекта с разрешением по длине волны 10 нм.

Заключение. В настоящее время в научной литературе сложилась противоречивая ситуация. С одной стороны, существует огромное количество информации, которую трудно бывает не только осмыслить, но даже прочитать. С другой стороны, некоторые, известные ранее факты и закономерности, или забыты, или исключены из научного оборота, поскольку приоритет отдается англоязычным публикациям, а их авторы практически не знакомы с русскоязычной литературой. Поэтому происходит “переоткрытие” того, что было открыто ещё в 60–80-ые годы прошлого века.

Показательна в этом отношении работа китайских исследователей, которые предложили в 2010 году подход, который они назвали “схема динамической обратной связи”. Они полагали, что разрабатывали новую переменную среды, которая изменяется в зависимости от пространственного изменения почвы. В их исследовании изучалась идея о том, что изменяющиеся типы поверхности земли, такие как те, которые ежедневно снимаются с помощью изображений дистанционного зондирования в течение короткого периода (6–7 дней) после сильного дождя, могут использоваться для дифференциации типов почвы. Для измерения динамики обратной связи они использовали спектральное расхождение информации (spectral information divergence, SID). Результаты показали, что существует значительная разница в значениях SID между парами пикселей в одной и той же подгруппе почвы и в разных подгруппах. Это указывало на то, что области с разными типами почв (уровень подгруппы) демонстрировали существенно разные динамические схемы обратной связи, а области в пределах одного и того же типа почвы имеют сходные динамические схемы обратной связи. Также было обнаружено, что чем больше сходных типов почв, тем больше сходных моделей обратной связи. Авторы считали, что полученные результаты могут привести к разработке нового экологического ковариата, который может быть использован для повышения точности картирования почв в районах с низким рельефом [25].

Обобщенные формы связи коэффициента яркости ρ поверхности почвы в оранжево-красной части спектра с влажностью поверхностного горизонта почв w были описаны Б.В. Виноградовым еще в 80-ые годы прошлого века [1]. Изменение яркости пикселей на аэрокосмическом изображении при иссушении почвы полностью соответствует термину «схема динамической обратной связи» Zhu с соавторами [25]. Для измерения динамики обратной связи Виноградов использовал эмпирические уравнения, которые давали возможность с высокой точностью предсказывать гранулометрический состав почв по формам математически описанных кривых. На этом примере видно, что работа 2010 года «переоткрывает» найденные ранее закономерности, а также уходит в сторону их математического формального описания, не используя функциональные зависимости.

Созданная нами база данных поможет избежать именно такой ситуации, поскольку объединяет в себе «старые» и «новые» знания и дает возможность не потерять ранее найденные закономерности.

Выводы:

1. Собрана информация по спектральным характеристикам 223 образцов почв из 22 литературных источников.
 2. Проведена графическая обработка и векторно-цифровое преобразование изображений в программе SAGA GIS.
 3. Для всех спектров собраны метаданные: таксономическая принадлежность, почвенные свойства в систему таблиц.
 4. Проведено объединение векторизованных графических материалов с данными по спектрам отражения в цифровом формате.
 5. Создана база данных, которая включает 705 спектров отражения, сведения о принадлежности образцов к 19 группам генетических горизонтов. 556 спектров имеют таксономическую характеристику на уровне типа, 488 - подтипа, 416 – разряда (почвообразующая порода). Для 349 образцов указан гранулометрический состав.
 6. Показана возможность использования базы данных для сравнительной характеристики цветовых различий на уровне спектров отражения и цветовых показателей основных типов почв, в том числе с привлечением цветовых коэффициентов, разработанных ранее в отечественном почвоведении.
- Благодарности.** Автор данной работы выражает искреннюю благодарность доктору биологических наук Кирилловой Наталии Петровне, под руководством которой данная работа была выполнена.

Список цитированных источников

1. Виноградов, Б. В. Аэрокосмический мониторинг экосистем. – М. : Наука, 1984. – 320 с.
2. Джадд Д., Вышецки Г. Цвет в науке и технике. Пер. с англ. М. : Мир, 1978. – 154 с.
3. Караванова, Е. И. Оптические свойства почв и их природа. – М. : Издательство Московского университета, 2003. – 151 с.
4. Кириллова, Н. П. Спектральные почвенные базы данных / Н. П. Кириллова, Д. М. Хомяков, Е. И. Караванова, Д. А. Азиков, Д. А. Жулидова // Вестник Московского Университета. Серия 17. Почвоведение. – 2020. – № 2. – С. 11–17.
5. Михайлова, Н. А. Оптические свойства почв и почвенных компонентов / Н. А. Михайлова, Д. С. Орлов // М. : Наука, 1986. – 118 с.
6. Обухов, А. И. Спектральная отражательная способность главнейших типов почв и возможность использования диффузного отражения при почвенных исследованиях / А. И. Обухов, Д. С. Орлов // Почвоведение. – 1964. – № 2. – С. 83–93.
7. Орлов, Д. С. Химия почв. – М. : МГУ, 1992. – 400 с.
8. Орлов, Д. С. Количественные параметры спектральной отражательной способности почв / Д. С. Орлов, Л. А. Воробьева, Н. И. Суханова // Вестник МГУ, сер. Почвоведение. – 1995. – № 4. – С. 35–42.
9. Орлов, Д. С. Спектральная отражательная способность почв и их компонентов / Д. С. Орлов, Н. И. Суханова, М. С. Розанова // М. : Издательство Московского университета, 2001. – С. 6–146.

10. Barrett L. R., 2002. Spectrophotometric color measurement in situ in well drained sandy soils // *Geoderma*. V. 108. P. 49–77.
11. California Institute of Technology. Jet Propulsion Laboratory. ECOSTRESS Spectral Library - <https://speclib.jpl.nasa.gov/ecospeclibdata/>
12. Colorimetry, CIE Techn. Rep., Vienna: CIE Central Bureau, 2004, no. 15:2004
13. Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Boehner, J. 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.*, 8, 1991-2007.
14. Demattê J.A.M. et al., 2019. The Brazilian Soil Spectral Library (BSSL): a general view, application and challenges // *Geoderma*. V. 354. P. 1-20.
15. Demattê J.A.M. et al., 2020. Bare Earth's surface spectra as a proxy for soil resource monitoring // *Nature: Scientific Reports* - www.nature.com/scientificreports
16. European Soil Data Centre (ESDAC). URL: <https://esdac.jrc.ec.europa.eu/>
17. Hardeberg, J.Y. 2003. Colorimetric scanner characterisation. *ActaGr*. v. 15. P. 89–104.
18. Lindbloom B., 2020. Useful Colour Equations. URL: <http://www.brucelindbloom.com>; (дата обращения: 15.04.2020).
19. Optical_spectrometer Wikipedia - Optical spectrometer, https://en.wikipedia.org/wiki/Optical_spectrometer
20. Shepherd K.D., Walsh M.G., 2002. SOIL & PLANT ANALYSIS Development of Reflectance Spectral Libraries for Characterization of Soil Properties. P. 988–998.
21. Viscarra Rossel R.A., Behrens T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. // *Geoderma*. V. 158. P. 46–54.
22. Viscarra Rossel R.A. et al., 2016. A global spectral library to characterize the world's soil // *Earth-Science Reviews*. V. 155. P. 198-230.
23. Viscarra Rossel R.A., Minasny B., Roudier P., McBratney A.B., 2006. Colour space models for soil science // *Geoderma*. V. 133. P. 320–337.
24. Viscarra Rossel R.A., Webster R., 2012. Predicting soil properties from the Australian soil visible-near infrared spectroscopic database. // *European Journal of Soil Science*. V. 63 (6). P. 848-860.
25. Zhu, A.X., Liu, F., Li, B.L., Pei, T., Qin, C.Z., Liu, G.H., Wang, Y.J., Chen, Y.N., Ma, X.W., Qi, F., and Zhou, C.H., 2010. Differentiation of soil conditions over flat areas using land surface feedback dynamic patterns extracted from MODIS // *Soil Sci. Soc. Am. J.* V. 74, P. 861–869.

УДК 528.1

ИНТЕРАКТИВНАЯ КАРТА РЕК И ОЗЕР РЕСПУБЛИКИ БЕЛАРУСЬ

О. А. Акулова, И. Н. Розумец, А. А. Лисицкая, Е. В. Горбачук

УО «Брестский государственный технический университет», Брест, Беларусь,
akylovaolya@yandex.by

Аннотация

Применение современных геоинформационных технологий является важным и актуальным направлением исследований. Они позволяют создавать ин-