Оптический журнал

Оптическое приборостроение

УДК 631.421

Сравнение эффективности различных методов предварительной обработки данных спектрометрирования для прогнозирования содержания органического углерода почв

© 2018 г. А. В. Чинилин*, аспирант; И. Ю. Савин**, ***, доктор с.-х. наук

*Российский государственный аграрный университет — Московская сельскохозяйственная академия им. К.А. Тимирязева, Москва

**Почвенный институт им. В.В. Докучаева, Москва

***Аграрно-технологический институт РУДН, Москва

E-mail: achinilin@rgau-msha.ru

Поступила в редакцию 25.06.2018

DOI:10.17586/1023-5086-2018-85-12-60-68

Проведено исследование эффективности применения ряда методов предварительной обработки данных спектрометрирования в диапазоне длин волн 325–1075 нм для прогнозирования содержания органического углерода почв. Методы предварительной обработки спектральных данных (фильтрация скользящим средним, сглаживание Савицкого–Голея, расчёт первой и второй производных и масштабирование) были последовательно применены к спектральным данным почв (в естественном сложении и растёртых) для повышения надёжности и результативности моделей. В соответствии с критерием максимального значения коэффициента детерминации и минимального значения корня среднеквадратической ошибки при перекрёстной проверке наилучшим методом прогнозирования органического углерода почв оказался метод регрессии частных наименьших квадратов при вычислении первых производных от исходных спектров ($R^2_{cv} = 0{,}758$, $RMSE_{cv} = 0{,}492$).

Ключевые слова: *спектроскопия почв, спектральная отражательная способность, прогнозирование, регрессия.*

Коды OCIS: *280.4788, 300.6190, 300.6340, 300.6490, 300.6550.*

ВВЕДЕНИЕ

Почвенный органический углерод формирует и поддерживает основные режимы, свойства и функции почв — экосистемные и социально-экономические, агроэкологические и санитарно-защитные, а также придаёт ей уникальные свойства эмерджентной системы [1]. Он является ключевым индикатором качества почв [2] и их экологической устойчивости [3], а также источником энергии и питательных веществ для растений. На сегодняшний день, при переходе к цифровым методам картографирования почв требуется альтернативная количественная оценка содержания органического углерода почв, которая имеет преимущества по сравнению с отбором и лабораторным анализом большого количества образцов и менее трудоёмка [4, 5]. Спектроскопия в видимой и ближней инфракрасной областях спектра — быстрый, точный, воспроизводимый и экономически эффективный аналитический метод бесконтактной оценки свойств почв [5, 6]. Бесконтактную спектроскопию можно рассматривать как альтернативу, направленную на улучшение традиционных и сложившихся методов анализа свойств почв. Этот метод продемонстрировал свою эффективность в прогнозировании содержания органического углерода почв, а также ряда других свойств почв в исследованиях последних лет [7, 8].

Основной проблемой, ограничивающей применение спектроскопии для оценки ряда свойств почв, является поиск подходящих параметров предобработки и трансформации «сырых» спектральных данных [9, 10]. В ряде случаев это связано с тем, что обычные кривые отражения очень похожи и по ним далеко не всегда удаётся вычленить особенности тех или иных почв, или же свойств

почв [11]. Методы предварительной обработки спектральной информации используются для преобразования исходных спектров, удаления шумов (вызванных конструктивными особенностями спектрорадиометра), извлечения наиболее информативных участков спектра для количественных прогностических моделей. Предварительная обработка может включать в себя фильтрацию и сглаживание, расчёт производных, коррекцию рассеивания, ряд методов по нормализации спектров и удалению мультипликативных и аддитивных эффектов [9, 12].

Результативность различных методов предварительной обработки спектральных данных с точки зрения прогнозирования свойств почв может варьировать в зависимости от цели исследования. Так, к примеру, Vasques с соавторами [13] сравнили 30 различных методов предварительной обработки спектральных данных, включая производные Савицкого–Голея и Норриса–Вильямса, трансформацию Кубелка–Мунка, стандартизацию и нормализацию, для прогнозирования содержания органического углерода почв. Авторы пришли к выводу, что расчёт производных Савицкого–Голея последовательно улучшал результативность моделей для всех рассматриваемых методов статистического анализа. Аналогичный результат был получен Peng с соавторами [14], которые исследовали влияние различных методов предобработки спектральных данных для кривых спектральной отражательной способности (СОС) почв из различных регионов Китая. Напротив же, Muñoz и Kravchenko [15] не наблюдали никаких улучшений в точности моделей при использовании методов предобработки спектральных данных (стандартизация и нормализация, расчёт производных) для прогнозирования содержания органического углерода почв. Рассмотренные разнообразные выводы из различных исследований указывают, что на сегодняшний день не существует ни одного метода или комбинации методов предварительной обработки, которые бы работали одинаково хорошо на различных наборах данных в различных почвенно-географических зонах. Это послужило мотивом данного исследования, направленного на сравнение эффективности разных методов предобработки спектральных данных для прогнозирования содержания органического углерода чернозёмных почв.

Та же проблема существует и с выбором соответствующих методов статистического анализа, использующих спектральные данные для оценки ряда свойств почв. По-видимому, выбор метода многомерного анализа зависит от выбора метода предварительной обработки спектральных данных [16].

Для прогнозирования содержания почвенного органического углерода успешно применялись несколько методов многомерного анализа. Так, регрессия частных наименьших квадратов (альтернативное название — метод проекции на скрытые структуры) является наиболее распространённым методом калибровки моделей связи спектральных данных с исследуемым свойством почв. Этот метод применялся для прогнозирования содержания почвенного органического углерода в ряде исследований последних лет [6, 17–19]. Кроме того, другие методы, такие как регрессия на главные компоненты (второй по применяемости метод) и множественная линейная регрессия показали значительные результаты в прогнозировании содержания органического углерода почв [20–22]. В дополнении к перечисленным методам рассматриваются методы нелинейного анализа данных: ансамбль деревьев решений, метод опорных векторов, которые относительно недавно получили распространение в целях прогнозирования содержания почвенного органического углерода [6, 18, 23].

В данной статье изложены результаты оценки влияния методов предварительной обработки спектральных данных видимой и ближней инфракрасной областей спектра на результативность прогнозирования содержания органического углерода чернозёмных почв.

ОБЪЕКТ И МЕТОДЫ ИССЛЕДОВАНИЯ

Объектом исследования выступает почвенный покров тестовых участков сельскохозяйственного предприятия «Белогорье» Закрытого акционерного общества «Агрофирма Апротек — Подгоренская» Подгоренского района Воронежской области. Территория исследования расположена в южной части Среднерусской возвышенности на правобережье р. Дон в пределах Калитвянского волнисто-балочного южно-лесостепного района.

Согласно почвенно-географическому районированию [24] территория объекта исследования входит в Центральную лесостепную и степную область, зону обыкновенных и южных чернозёмов, Южно-Русскую провинцию. На большей части территории исследования распространены чернозёмы обыкновенные. В пределах тестовых участков установлено формирование следующих родов обыкновенных чернозёмов: обычные, карбонатные, остаточно-карбонатные и бескарбонатные. По пониженным элементам рельефа (неглубоким ложбинам и депрессиям водоразделов, выположенным участкам вогнутых склонов) формируются лугово-чернозёмные почвы (названия почв даны в соответствии с «Классификацией и диагностикой почв СССР» [25]).

Почвообразующие и подстилающие породы представлены покровными отложениями, элюводелювием коренных меловых пород, неогеновыми

песчаными и неогеновыми глинистыми отложениями. Особо выделяются почвы на двучленных породах, где покровные отложения подстилаются неогеновыми песчаными отложениями (глубина подстилания варьирует от 40 до 80 см), неогеновыми глинистыми отложениями (глубина подстилания от 150 см и более).

В ходе почвенно-ландшафтного обследования было заложено 22 опорных разреза и на местах заложения разрезов были отобраны образцы поверхностного слоя почв (0–5 см). Всего для анализа было отобрано 22 образца поверхностного пахотного горизонта почв. Далее в образцах проводился анализ на определение содержания органического углерода в почвах по методу И.В. Тюрина в модификации В.Н. Симакова. В табл. 1 представлена описательная статистика по исследуемому свойству.

Как видно из табл. 1, для почв тестовых участков характерным является довольно большая разница в содержании органического углерода. Это связано, в первую очередь, со сложным геологическим строением и маломощностью чехла четвертичных отложений. В целом, по данным Dotto и соавторов [16] достаточный разброс данных в лучшую сторону влияет на результативность моделей. Следует сказать, что согласно глобальной спектральной библиотеке почв [8] среднее значение органического углерода почв из 17928 образцов почв мира — 2,16%.

После отбора образцов они были высушены, и была проведена съёмка СОС воздушно-сухих образцов в их естественном сложении спектрорадиометром FieldSpec® HandHeld 2, который проводит измерение в диапазоне длин волн от 325 до 1075 нм (видимая и ближняя инфракрасная области спектра) при точности ±1 нм и спектральном разрешении менее 3 нм в рассматриваемом диапазоне. Во время съёмки спектрорадиометр располагался в надир от поверхности образца на высоте 10 см. Съёмка проводилась в дневное время в солнечную погоду в апреле 2016 г. Перед началом съёмки СОС была проведена калибровка прибора с помощью белой контрольной панели Spectralon®, состоящей из политетрафторэтилена и имеющей отражение 99%. При съёмке образцов калибровка проводилась также каждые 10 мин ввиду изменения условий освещения. Нужно отметить, что для каждого образца полученные кривые СОС представляли собой усреднённые кривые 10 измерений. Большее значение повторности вызвало бы увеличение времени съёмки, что повлекло бы за собой возможное появление шумов вследствие изменяющегося спектрального состава солнечного света и других внешних условий.

После съёмки кривых СОС образцы были очищены от остатков растительности, мелких камней и просеяны через сито с отверстиями диаметром 1 мм. После этого была проведена съёмка (июль 2016) кривых СОС растёртых образцов таким же образом, как это было описано выше.

В ходе первичного анализа были отброшены зашумлённые участки спектра (искусственный шум, вызванный конструктивными особенностями спектрорадиометра) от 325 до 400 нм и от 905 до 1075 нм.

Первый этап анализа спектральных данных часто состоит из предварительной обработки, о чём уже было сказано выше. В целом, методы предварительной обработки могут быть разделены на 4 группы: сглаживание кривых спектрального отражения, их трансформация, масштабирование и нормирование [26]. Первая категория — сглаживание, используется для уменьшения количества шума (случайной погрешности измерения). Сглаживание Савицкого–Голея [27] и фильтрация скользящим средним — одни из самых популярных методов усреднения спектральных данных. Одними из лучших методов по трансформации исходных кривых являются вычисление первой и второй производных, направленных на устранение смещения исходных спектров и устранение линейного тренда из спектральных данных. Масштабирование диапазона спектральных данных является следующим методом предварительной обработки. Среди различных методов масштабирования наиболее распространённым является метод стандартной нормальной переменной (standart normal variate). Другими словами, после масштабирования каждый спектр будет иметь среднее значение, равным нулю, и стандартное отклонение, равным единице.

Рассмотренные методы предобработки последовательно были применены к спектральным данным почв (в естественном сложении и растёртых) в диапазоне от 400 до 905 нм. Для начала к «сырым» спектральным данным осуществили фильтрацию скользящим средним (MA) с размером окна, равным 11 нм. Затем применили сглаживание Савицкого–Голея (SG) с аналогичным размером окна и полиномом второй степени. После — расчёт первой (FD) и второй производных (SD) также с полиномом второй степени и размером окна, равным

Таблица 1. Описательная статистика по анализируемому свойству

Свойство	Число наблюдений	Минимальное значение	1ый квартиль	Медиана	Среднее	3ий квартиль	Максимальное значение
Орг. углерод, %	22	0,74	1,91	2,67	2,49	3,20	3,71

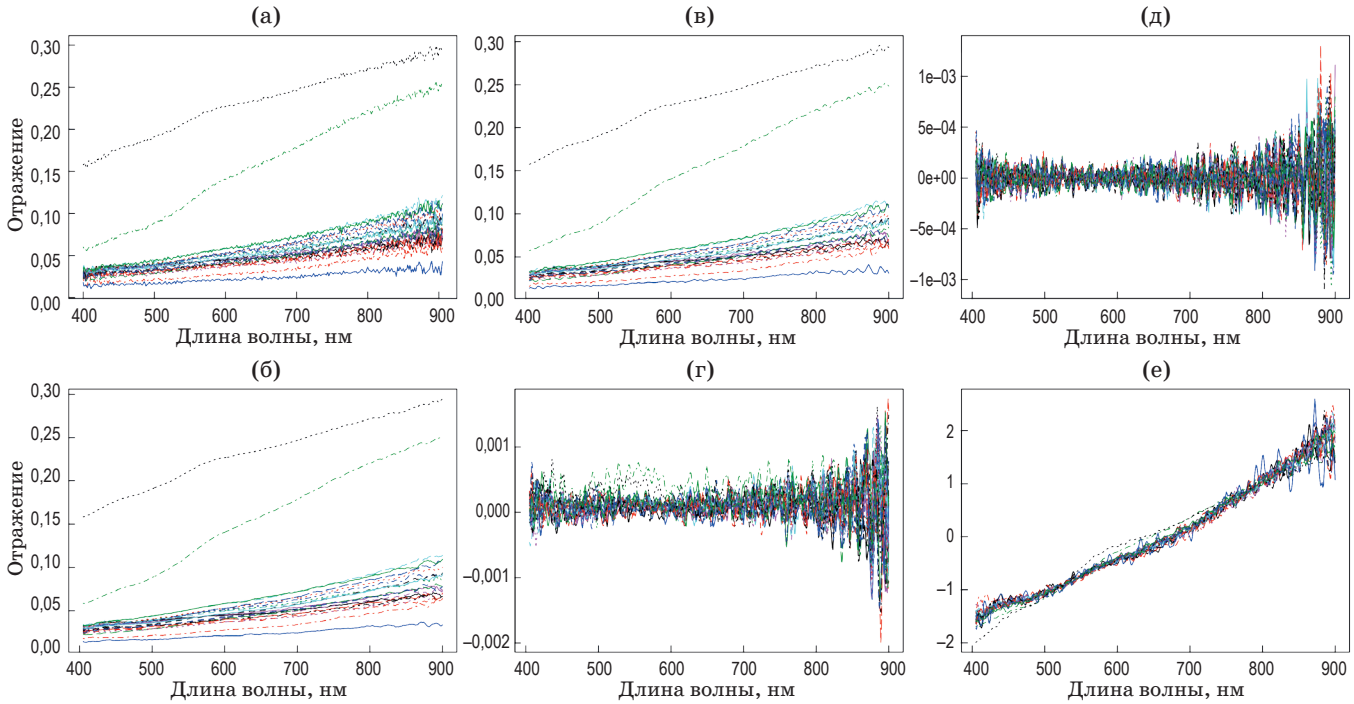

Рис. 1. Предобработка кривых СОС для всех образцов (в естественном сложении): «сырые» спектры (а), фильтрация скользящим средним (MA) (б), сглаживание Савицкого–Голея (SG) (в), первая производная (FD) (г), вторая производная (SD) (д), SNV-масштабирование (е).

11 нм, и масштабирование (SNV). Следует сказать, что масштабирование выполнялось после сглаживания Савицкого–Голея. Таким образом, сравнивали влияние 5 методов предобработки спектральных данных на результативность прогнозирования наряду с «сырыми» данными (рис. 1).

На следующем этапе проводился анализ связи между СОС и исследуемым свойством и определялись потенциально наиболее информативные длины волн для прогнозирования содержания почвенного органического углерода. Для этого применяли несколько распространенных методов статистического анализа: пошаговую множественную регрессию в сочетании с методом главных компонент (StepMLR), регрессию на главные компоненты (PCR), регрессию частных наименьших квадратов (PLSR), лассо/гребневую регрессию (Lasso/Ridge), ансамбль деревьев решений (RF). Работы по предварительной обработке спектральных данных, моделированию и проверке моделей на устойчивость проводились в свободной вычислительной среде R [28] при помощи пакетов «prospectr», «pls», «caret», «doParallel». Этапы обработки и трансформации данных, моделирования, проверки моделей на устойчивость, функции визуализации, скрипт и спектральные данные расположены в открытом доступе на интернет-странице github-аккаунта[1].

[1] https://github.com/chinilin/Spectra

При сравнении результативности и устойчивости моделей использовалась n-кратная перекрёстная проверка, реализованная в пакете «caret» [29]:

$$\mathrm{R} > \mathrm{ctrl1} = \mathrm{trainControl}(\mathrm{method} =$$

$$= \text{“repeatedcv”}, \mathrm{number} = 5, \mathrm{repeats} = 10).$$

Это означает, что модель подбирается 5 раз с использованием 90% значений и предсказанные значения от полученной модели сравниваются с оставшимися 10%. Этот процесс повторяется 10 раз для получения стабильного результата. Для каждой модели были получены коэффициенты детерминации (R^2_{cv} — доля дисперсии, объясняемая моделью), корень среднеквадратической ошибки ($RMSE_{cv}$), как показатели результативности. Значение $RMSE_{cv}$ — легко интерпретируемая величина, так как имеет те же единицы измерения, что и исследуемое свойство. Модель с наименьшим значением $RMSE_{cv}$ выбиралась как оптимальная.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Разница в кривых спектральной отражательной способности почв каждого из образцов объясняется изменчивостью в свойствах почв. Рис. 1а показывает кривые СОС образцов в их естественном сложении, характеризующихся по большей части низкими значениями отражения, что характерно для чернозёмных почв. В связи с достаточной

разницей в содержании почвенного органического углерода наблюдается разница в наклоне и высоте над осью абсцисс у некоторых кривых СОС. Также некоторые кривые отчётливо выделяются по своей форме и наличию экстремумов, что, по-видимому, связано с различным минералогическим составом почв.

Таблица 2 показывает результаты прогнозирования по рассматриваемым многомерным методам с использованием различных стратегий предобработки спектральных данных образцов в их естественном сложении. Из таблицы следует, что предобработка кривых спектрального отражения методами MA (фильтрация скользящим средним), SG (сглаживание Савицкого–Голея) и SNV-масштабирования при использовании любого из статистических подходов к прогнозированию даёт результаты, которые почти не отличаются от результатов, получаемых по «сырым» данным. При использовании вторых производных (SD) качество получаемых моделей хуже, чем для «сырых» (RAW) данных, что, по-видимому, связано с добавлением дополнительного шума и переобучением моделей. Самые хорошие результаты получены при использовании вычислений первой производной от исходных спектров (FD). При этом результативность и устойчивость моделей изменяется в зависимости от применяемого метода. Так, коэффициент детерминации изменяется от 0,531 до 0,63, $RMSE_{cv}$ от 0,623 до 0,688 (табл. 2). Наиболее низкие значения $RMSE_{cv}$ получены для двух методов — регрессии на главные компоненты (PCR) и ансамбля деревьев решений (RF). Оба метода дают схожие результаты, но последний представляет собой некий «чёрный ящик» и является менее интерпретируемым, чем регрессия на главные компоненты. При этом, переменные, внёсшие вклад в лучшие модели, различаются в зависимости от применяемого метода анализа. Так, к примеру, наиболее значимыми переменными (длинами волн), внесшими вклад в PCR модель (в порядке убывания значимости) являются 450, 643, 451, 463, 449, 565, 879, 588, 501, 562 нм. Наиболее значимыми переменными, внесшими вклад в RF модель (в порядке убывания значимости), — 565, 564, 645, 562, 496, 516, 495, 643, 424, 452 нм (рис. 2). Переменные 530, 562, 565, 643, 644 нм являются информативными для обоих рассматриваемых методов. Также, следует сказать, что наиболее значимые переменные, в целом, относятся к видимой (400–700 нм) области спектра.

Таблица 2. Результаты прогнозирования по рассматриваемым многомерным методам с использованием различных стратегий предобработки спектральных данных образцов в их естественном сложении

Метод	Предобработка	R^2_{cv}	RMSE (%)
StepMLR	RAW	0,449	1,511
	MA	0,448	1,525
	SG	0,448	1,518
	FD	0,592	0,688
	SD	0,356	0,879
	SNV	0,483	0,808
PCR	RAW	0,558	0,800
	MA	0,558	0,799
	SG	0,558	0,799
	FD	0,630	0,623
	SD	0,392	0,815
	SNV	0,588	0,723
PLSR	RAW	0,556	0,793
	MA	0,556	0,792
	SG	0,556	0,792
	FD	0,621	0,671
	SD	0,428	0,823
	SNV	0,540	0,753
Lasso/Ridge	RAW	0,497	0,727
	MA	0,509	0,718
	SG	0,522	0,721
	FD	0,531	0,671
	SD	0,274	0,878
	SNV	0,540	0,821
RF	RAW	0,472	0,772
	MA	0,473	0,778
	SG	0,468	0,778
	FD	0,639	0,670
	SD	0,401	0,801
	SNV	0,519	0,713

Совсем иная картина складывается для результатов прогнозирования с использованием различных стратегий предобработки спектральных данных растёртых образцов. Из табл. 3 следует, что предобработка кривых спектрального отражения методами MA (фильтрация скользящим средним), SG (сглаживание Савицкого–Голея) при использовании методов регрессии на главные компоненты (PCR), регрессии частных наименьших квадратов (PLSR) и лассо/гребневой регрессии (Lasso/Ridge) улучшает качество моделей. Так, к примеру, для PLSR модели значение $RMSE_{cv}$ изменяется от 0,698 для «сырых» спектров до 0,571 для данных, прошедших этап сглаживания Савицкого–Голея. Для PCR модели значение $RMSE_{cv}$ изменяется от 0,762 для «сырых» спектров до 0,573 для данных, прошедших этап сглаживания Савицкого–Голея.

Таблица 3. Результаты прогнозирования по рассматриваемым многомерным методам с использованием различных стратегий предобработки спектральных данных растертых образцов

Метод	Предобработка	R^2_{cv}	RMSE (%)
StepMLR	RAW	0,671	1,272
	MA	0,671	1,275
	SG	0,670	1,273
	FD	0,738	0,533
	SD	0,302	0,848
	SNV	0,407	0,816
PCR	RAW	0,681	0,762
	MA	0,688	0,604
	SG	0,705	0,573
	FD	0,751	0,502
	SD	0,374	0,804
	SNV	0,488	0,717
PLSR	RAW	0,685	0,698
	MA	0,618	0,651
	SG	0,703	0,571
	FD	0,758	0,492
	SD	0,383	0,822
	SNV	0,498	0,708
Lasso/Ridge	RAW	0,749	0,613
	MA	0,719	0,604
	SG	0,721	0,604
	FD	0,669	0,537
	SD	0,314	0,894
	SNV	0,583	0,632
RF	RAW	0,626	0,609
	MA	0,604	0,628
	SG	0,613	0,621
	FD	0,698	0,600
	SD	0,312	0,855
	SNV	0,492	0,736

Как и в предыдущем случае, самые хорошие результаты получены при использовании вычислений первой производной (FD) от исходных спектров (независимо от метода многомерного анализа). При этом наиболее низкие значения $RMSE_{cv}$ получены для двух методов — регрессии частных наименьших квадратов (PLSR) — 0,492 и регрессии на главные компоненты (PCR) — 0,502, коэффициенты детерминации — 0,758 и 0,751 соответственно. При использовании вторых производных (SD), а также SNV-масштабирования качество получаемых моделей хуже, чем для «сырых» данных (RAW).

Переменные, внесшие вклад в лучшие модели, различаются в зависимости от применяемого метода анализа, что показано на рис. 3. Следует сказать, что переменные 478 и 710 нм являются информативными для обоих рассматриваемых методов, причём последняя является наиболее значимой в обоих случаях.

Следует сказать, что результаты прогнозирования содержания почвенного органического углерода по спектральным данным растёртых образцов лучше (независимо от стратегии предобработки), чем по спектральным данным образцов в их естественном сложении. В абсолютном большинстве работ, в которых анализируется связь спектральных данных со свойствами почв исследователи используют именно высушенные и растёртые образцы, которые представляют собой «идеальную» среду. Подобная пробоподготовка входит в стандарты и протоколы по использованию спектроскопии [8].

По полученным наиболее информативным переменным (рис. 2 и 3) (участкам спектра) теорети-

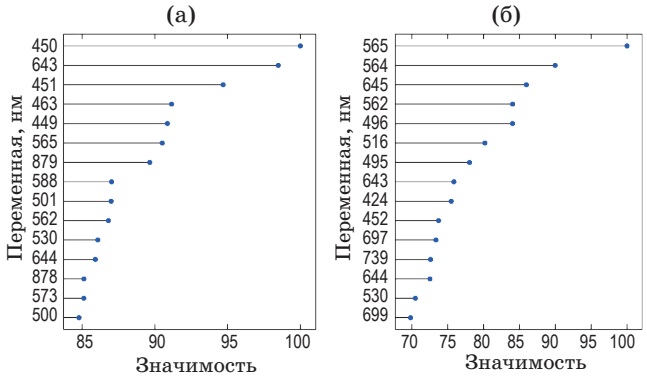

Рис. 2. Топ-15 наиболее значимых переменных (в порядке убывания значимости) для метода регрессии на главные компоненты (PCR) (а), ансамбля деревьев решений (RF) (б).

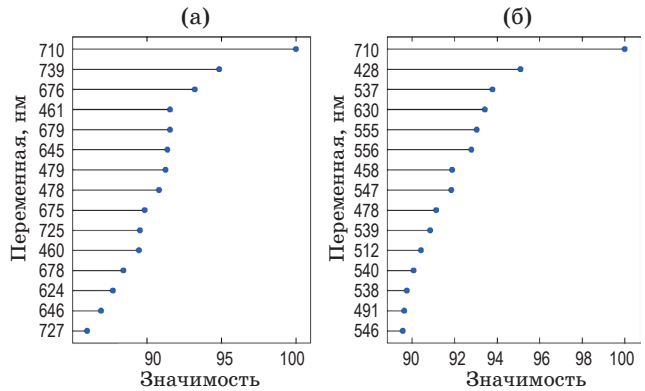

Рис. 3. Топ-15 наиболее значимых переменных (в порядке убывания значимости) для метода регрессии частных наименьших квадратов (PLSR) (а), для метода регрессии на главные компоненты (PCR) (б).

чески можно переходить к картографическим моделям содержания органического углерода почв, используя, к примеру, данные гиперспектральной съёмки, выбирая те каналы съёмки, в которые попадают информативные переменные. Однако зачастую в подобных исследованиях картографические модели теряют в точности, что связано с несколькими причинами – влиянием влажности, условий освещённости и высоты Солнца, рельефа, состояния поверхности почв и прочих [30–32].

Результаты эффективности по рассматриваемым методам многомерного анализа показывают, что регрессия на главные компоненты (PCR), регрессия частных наименьших квадратов (PLSR) и ансамбль деревьев решений (RF) являются наилучшими. Первые два метода, как уже было сказано выше, являются наиболее часто используемыми методами калибровки моделей связи спектральных данных со свойствами почв. Оба метода используются тогда, когда в данных присутствует большой набор скоррелированных переменных-предикторов. Оба метода схожи в том, что генерируют новые переменные-предикторы, которые представляют собой линейную комбинацию исходных переменных. Метод ансамблей деревьев решений является эффективным инструментом прогнозирования, он универсальный и гибкий, работает как с малыми, так и с большими наборами данных, однако, он представляет собой «чёрный ящик» и трудно интерпретируемый.

ВЫВОДЫ

1. Предварительная обработка не во всех случаях увеличивает качество прогнозных моделей содержания органического углерода почв региона исследования. Эффективность метода предварительной обработки спектральных данных зависит от используемого метода многомерного анализа данных.

2. Результаты прогнозирования содержания почвенного органического углерода по спектральным данным растёртых образцов лучше, чем по спектральным данным образцов в их естественном сложении независимо от стратегии предобработки.

3. В случае использования предобработки спектральных данных образцов почв в их естественном сложении лучшие результаты получены при вычислении первых производных от исходных спектров ($RMSE_{cv}$ для 0,623). При этом качество и устойчивость моделей изменяется в зависимости от применяемого метода. Предобработка кривых спектрального отражения методами фильтрации, сглаживания и масштабирования при использовании любого из статистических подходов в данном конкретном случае не улучшает точность прогностических моделей.

4. В случае использования предобработки спектральных данных образцов растёртых почв методами фильтрации и сглаживания качество моделей повышается при использовании ряда статистических подходов. Лучшие результаты опять же получены при вычислении первых производных от исходных спектров ($RMSE_{cv}$ для 0,492) в независимости от применяемого подхода. При использовании вторых производных, а также масштабирования, точность моделей хуже, чем для исходных спектральных данных.

5. По результатам исследования наиболее хорошие результаты показали следующие методы многомерного анализа:

- регрессия на главные компоненты — $RMSE_{cv}$ для 0,623, R^2_{cv} для 0,630 в случае использования предобработки спектральных данных образцов почв в их естественном сложении; $RMSE_{cv}$ для 0,502, R^2_{cv} для 0,751 в случае использования предобработки спектральных данных образцов растёртых почв,
- регрессия частных наименьших квадратов — $RMSE_{cv}$ для 0,492, R^2_{cv} для 0,758 в случае использования предобработки спектральных данных образцов растёртых почв,
- ансамбль деревьев решений — $RMSE_{cv}$ для 0,670, R^2_{cv} для 0,639 в случае использования предобработки спектральных данных образцов почв в их естественном сложении.

6. В случае использования предобработки спектральных данных образцов почв в их естественном сложении наиболее значимыми переменными (длинами волн), внесшими вклад в PCR модель (в порядке убывания значимости) являются 450, 643, 451, 463, 449, 565, 879, 588, 501, 562 нм. В случае же использования предобработки спектральных данных образцов растёртых почв наиболее значимыми переменными (длинами волн), внесшими вклад в PCR модель (в порядке убывания значимости) являются 710, 739, 676, 461, 679, 645, 479, 478, 675, 725 нм. Полученные информативные участки спектра потенциально могут использоваться для картографирования содержания органического углерода почв региона исследования по данным дистанционного зондирования.

Спектроскопия имеет хороший потенциал для целей анализа почвенных свойств. Тем не менее, почвенная спектроскопия до сих пор широко не используется, тем более в нашей стране. Наиболее вероятная причина этого заключается в том, что, несмотря на все перспективные результаты, методы прогнозирования все ещё громоздки и иногда ненадежны.

Исследования выполнены при поддержке гранта РФФИ 18-016-00052 с использованием оборудования Центра коллективного пользования: «Функции и свойства почв и почвенного покрова» Почвенного института им. В.В. Докучаева.

ЛИТЕРАТУРА

1. *Семенов В.М., Когут Б.М.* Почвенное органическое вещество. М: ГЕОС, 2015. 233 с.
2. *Борисов Б.А., Ганжара Н.Ф.* Органическое вещество почв (генетическая и агрономическая оценка). М: Изд-во РГАУ-МСХА, 2015. 214 с.
3. *McBratney A.B., Field D.J., Koch A.* The dimensions of soil security // Geoderma. 2014. Т. 213. С. 203–213. DOI: 10.1016/j.geoderma.2013.08.013.
4. *Minasny B., McBratney A.B.* Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy // Chemom. Intell. Lab. Syst. 2008. Т. 94. № 1. С. 72–79. DOI: 10.1016/j.chemolab.2008.06.003.
5. *Viscarra Rossel R.A., 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. Т. 131. № 1–2. С. 59–75. DOI: 10.1016/j.geoderma.2005.03.007.
6. *Rossel R.A.V., Behrens T.* Using data mining to model and interpret soil diffuse reflectance spectra // Geoderma. 2010. Т. 158. № 1–2. С. 46–54. DOI: 10.1016/j.geoderma.2009.12.025.
7. *Dor E. Ben, Ong C., Lau I.C.* Reflectance measurements of soils in the laboratory: Standards and protocols // Geoderma. 2015. Т. 245–246. DOI: 10.1016/j.geoderma.2015.01.002.
8. *Viscarra Rossel R.A.A., Behrens T., Ben-Dor E., Demattê J.A.M., Adamchuk V., Bayer A.D.D.* A global spectral library to characterize the world's soil // Earth-Science Rev. 2016. Т. 155. С. 198–230. DOI: 10.1016/j.earscirev.2016.01.012.
9. *Dotto A.C., Dalmolin R.S.D., Grunwald S., ten Caten A., Pereira Filho W.* Two preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopy // Soil Tillage Res. 2017. Т. 172. С. 59–68. DOI: 10.1016/j.still.2017.05.008.
10. *Gholizadeh A., Borůvka L., Saberioon M.M., Kozák J., Vašát R., Němeček K.* Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features // Soil Water Res. 2016. Т. 10. № No. 4. С. 218–227. DOI: 10.17221/113/2015-SWR.
11. *Орлов Д.С., Суханова Н.И., Розанова М.С.* Спектральная отражательная способность почв и их компонентов. М: Изд-во Моск. ун-та, 2001. 176 с.
12. *Volkan Bilgili A., van Es H.M., Akbas F., Durak A., Hively W.D.* Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey // J. Arid Environ. 2010. Т. 74. № 2. С. 229–238. DOI: 10.1016/j.jaridenv.2009.08.011.
13. *Vasques G.M., Grunwald S., Sickman J.O.* Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra // Geoderma. 2008. Т. 146. № 1–2. С. 14–25. DOI: 10.1016/j.geoderma.2008.04.007.
14. *Peng X., Shi T., Song A., Chen Y., Gao W.* Estimating soil organic carbon using VIS/NIR spectroscopy with SVMR and SPA methods // Remote Sens. 2014. Т. 6. № 4. С. 2699–2717. DOI: 10.3390/rs6042699.
15. *Muñoz J.D., Kravchenko A.* Soil carbon mapping using on-the-go near infrared spectroscopy, topography and aerial photographs // Geoderma. 2011. Т. 166. № 1. С. 102–110. DOI: 10.1016/j.geoderma.2011.07.017.
16. *Dotto A.C., Dalmolin R.S.D., ten Caten A., Grunwald S.* A systematic study on the application of scatter-correctiveand spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra // Geoderma. 2018. Т. 314. С. 262–274. DOI: 10.1016/j.geoderma.2017.11.006.
17. *Conforti M., Castrignanò A., Robustelli G., Scarciglia F., Stelluti M., Buttafuoco G.* Laboratory-based Vis–NIR spectroscopy and partial least square regression with spatially correlated errors for predicting spatial variation of soil organic matter content // CATENA. 2015. Т. 124. С. 60–67. DOI: 10.1016/j.catena.2014.09.004.
18. *Knox N.M., Grunwald S., McDowell M.L., Bruland G.L., Myers D.B., Harris W.G.* Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy // Geoderma. 2015. Т. 239–240. С. 229–239. DOI: 10.1016/j.geoderma.2014.10.019.
19. *Kuang B., Tekin Y., Mouazen A.M.* Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content // Soil Tillage Res. 2015. Т. 146. С. 243–252. DOI: 10.1016/j.still.2014.11.002.
20. *Bayer A., Bachmann A., Muller A., Kaufmann H.* A comparison of feature-based MLR and PLS regression techniques for the prediction of three soil constituents in a degraded South African Ecosystem // Appl. Environ. Soil Sci. 2012. Т. 2012. С. 1–20. DOI: 10.1155/2012/971252.
21. *Chang C.-W., Laird D.A., Mausbach M.J., Hurburgh C.R.* Near-infrared reflectance spectroscopy – principal components regression analyses of soil properties // Soil Sci. Soc. Am. J. 2001. Т. 65. № 2. С. 480. DOI: 10.2136/sssaj2001.652480x.
22. *Wang Y., Huang T., Liu J., Lin Z., Li S.* Soil pH value, organic matter and macronutrients contents prediction using optical diffuse reflectance spectroscopy // Comput. Electron. Agric. 2015. Т. 111. С. 69–77. DOI: 10.1016/j.compag.2014.11.019.
23. *Terra F.S., Demattê J.A.M.M., Viscarra Rossel R.A.* Spectral libraries for quantitative analyses of tropical Brazilian soils: comparing vis–NIR and mid-IR reflectance data // Geoderma. 2015. Т. 255–256. С. 81–93. DOI: 10.1016/j.geoderma.2015.04.017.

24. *Добровольский Г.В., Урусевская И.С.* География почв. М: МГУ, 2015. 458 с.

25. Классификация и диагностика почв СССР. М: Колос, 1977. 221 с.

26. *Xie X.-L., Pan X.-Z., Sun B.* Visible and near-infrared diffuse reflectance spectroscopy for prediction of soil properties near a copper smelter // Pedosphere. 2012. Т. 22. № 3. С. 351–366. DOI: 10.1016/S1002-0160(12)60022-8.

27. *Savitzky A., Golay M.J.E.* Smoothing and differentiation of data by simplified least squares procedures // Anal. Chem. 1964. Т. 36. № 8. С. 1627–1639.

28. *R. Core Team.* R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria // 2018. URL https://www.R-project.org/.

29. *Kuhn M.* Building predictive models in R using the caret package // J. Stat. Softw. 2008. Т. 28. № 5. DOI: 10.18637/jss.v028.i05.

30. *Прудникова Е.Ю., Савин И.Ю.* Исследование оптических свойств открытой поверхности почв // Оптический журнал. 2016. Т. 83. № 10. С. 79–86.

31. *Савин И.Ю., Прудникова Е.Ю.* Об оптимальном сроке спутниковой съемки для картографирования пахотных почв // Бюллетень Почвенного института им. В.В. Докучаева. 2014. № 74. С. 66–77.

32. *Lagacherie P., Baret F., Feret J-B., Madeira Netto J., Robbez-Masson J.M.* Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements // Remote Sens. Environ. 2008. Т. 112. № 3. С. 825–835. DOI: 10.1016/j.rse.2007.06.014.