

УТИЛИЗАЦИЯ И ЗАХОРОНЕНИЕ ОТХОДОВ

УДК 543.08+628.4

АНАЛИЗ СТРУКТУРЫ И СОСТОЯНИЯ ПОЛИГОНОВ И СВАЛОК ТВЕРДЫХ БЫТОВЫХ ОТХОДОВ. ХЕМОМЕТРИЧЕСКИЙ ПОДХОД

© 2009 г. Е. В. Михайлов*, О. Е. Родионова**, А. Л. Померанцев***,****

*Самарский государственный технический университет

**Институт химической физики РАН

*** Государственный южный научно-исследовательский полигон РАН

Поступила в редакцию 16.08. 2007 г.

После исправления 21.01.2008 г.

Хемометрические методы анализа многомерных данных применяются для оценки структуры и состояния полигонов складирования отходов (свалок). Анализ проводится для трех объектов: несанкционированная свалка, плохо организованный полигон и усовершенствованный полигон. Образцы массива свалок, полученные пробным бурением, характеризуются измеренными параметрами, такими как глубина, температура, объемный вес, влажность, зольность. Для тех же образцов рассматриваются и оцениваемые свойства – возраст, принадлежность к слою или линзе, полученные традиционными методами исследования свалок. Хемометрические методы дают возможность установить структуру свалки; выявить специфичные области. Предсказание таких показателей как стабильность и возраст, может быть выполнено с помощью регрессии, базирующейся только на измеряемых, объективных данных.

1. ВВЕДЕНИЕ

В этой статье мы продолжаем исследование такого необычного аналитического объекта, как полигон складирования бытовых и промышленных отходов, или, проще говоря, свалка. Существующие свалки – как нелегальные, так и организованные, современные полигоны, – представляют гетерогенную, эволюционирующую массу, в теле которой часто можно выделить обширные участки неоднородности: линзы, слои и т.п. Первая задача при изучении таких объектов – выявление этих пространственных неоднородностей, установление их природы. Не менее важна и вторая задача – определение временной неоднородности, т.е. оценка возраста различных участков тела свалки и прогнозирование перспектив ее эволюции. Как известно, в массиве свалки проходят разнообразные физико-химические и биохимические процессы, приводящие к изменению ее состава и структуры. Изучение этих явлений позволяет оценить стабильность объекта и выработать направления по его рекультивации.

В первой части работы [4] для решения поставленных задач был применен кинетический подход. Используя данные, полученные в ходе компостирования [6], мы построили модель кинетики разложения компостного вещества в форме уравнения реакции первого порядка. Рассчитанная константа скорости оказалась общей для образцов, извлеченных из различных объектов.

Предполагая, что процессы разложения органики при компостировании и в условиях свалки подобны, мы сумели оценить константы скорости разложения в условиях свалки, которые оказались разными для каждого рассмотренного объекта. Опираясь на эти индивидуальные константы, мы рассчитали ориентировочный возраст для каждого образца свалочного тела. Кроме того, нам удалось выявить принадлежность некоторых образцов к специфичным областям свалки, таким как линзы осадков сточных вод, промышленные отходы и т.п.

Существующие подходы к математическому моделированию физико-химических процессов, происходящих в теле свалки, основаны на построении сложных кинетических схем [10, 15]. Эти модели содержат большое количество неопределенных кинетических, структурных и физических параметров, оценка которых представляет непростую задачу. В рамках этого подхода используются сложные, дорогостоящие биореакторы, призванные имитировать реальные процессы. Появившиеся недавно работы по распределенным моделям [20, 21], еще более громоздки. Сейчас стало совершенно очевидно, что такой подход является, по сути, тупиковым, так как он никогда не сможет учесть главный фактор, влияющий на поведение свалки – ее неоднородность.

В этой статье представлен другой, хемометрический подход к исследованию свалок, для которого некоторые предварительные результаты

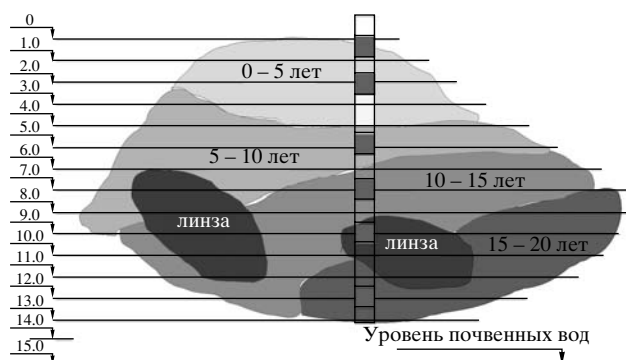


Рис. 1. Вертикальный разрез свалки Безенчук с распределением отходов по возрасту. Показаны линзы осадков сточных вод, положение шурфа № 14 и глубины слоев в метрах.

были опубликованы в [14]. Этот подход основан на анализе многомерных данных методом главных компонент (МГК) и с помощью проекций на латентные структуры (ПЛС). Цель работы – последовательное сравнение выводов, полученных традиционными кинетическими методами, с результатами, вытекающими из хемометрического анализа. Основной вопрос, на который мы должны ответить, можно сформулировать следующим образом. Позволяет ли хемометрический подход получить результаты, схожие с теми, которые достигаются традиционными (кинетическими) методами? Положительный ответ на этот вопрос откроет возможность заменить длительные, дорогостоящие опыты по компостированию быстрым математическим расчетом.

В последние годы методы анализа многомерных данных с успехом применяются в самых разных областях, в том числе и к задачам экологического мониторинга. Так, в работах [18, 19] они использовались для определения стабильности отходов с помощью ИК спектроскопии. Аналогичная задача решалась в работе [9], где исследовались данные по содержанию летучей золы, образующейся при сжигании отходов. Метод многомерного статистического контроля [12] использовался для управления процессами компостирования [23, 24] и анаэробной деградации [11].

В этой статье статистический анализ применяется к данным, полученным в ходе исследовательского бурения свалок с последующим лабораторным анализом образцов, полученных с разных глубин. Эксперименты проводились в несколько стадий. Геологические исследования объектов дали такие показатели, как размер, объем и последовательность заполнения свалок. Последний показатель – только для организованных полигонов. Результаты этих исследований представлены в разделе 2.1. На втором этапе проводили бурение скважин, отбор проб и лабораторные исследова-

ния. Были измерены различные физические химические показатели, такие как зольность, влажность, pH для образцов, отобранных с различных глубин. Подробнее этот этап описан в разделе 2.2.

Собранные данные затем анализировались с помощью традиционных методов, которые кратко описаны в разделе 3.1. Методы анализа многомерных данных представлены в разделах 3.2 и 3.3. В разделе 4 приводятся полученные результаты, дается их обсуждение.

2. ЭКСПЕРИМЕНТ

2.1. Объекты исследования

Объектами исследований являлись несанкционированная свалка твердых бытовых (ТБО) и сельскохозяйственных отходов пос. Безенчук, неорганизованная свалка ТБО (г. Отрадный), усовершенствованный полигон ТБО и промышленных отходов (г. Кинель). Исследуемые свалки находятся на территории Самарской области и рассматриваются как наиболее типичные объекты размещения отходов. Подробно о том, что представляют собой массивы, и как они были сформированы, рассказано в работе [4].

2.2. Образцы

Исследование объектов проводилось в течение 2000–2004 гг. с бурением шурфов и отбором проб по стандартным методикам [1]. На каждой свалке пробурили следующее число шурфов: Безенчук – 21; Отрадный – 12; Кинель – 12. Исследования объектов Кинель и Отрадный проводились только в теплое время года, а на свалке Безенчук еще и зимой. Глубину каждой скважины устанавливали в ходе бурения, учитывая особенности исследуемого объекта. Таким образом, глубина скважин составила: Безенчук – 0,3–13,0 м; Отрадный – 0,3–7,0 м; Кинель – 0,5–10,0 м. Образцы извлекали с шагом 1 м, в виде целостных массивов ненарушенной структуры массой 1,5–3,0 кг. Температуру образцов измеряли в поле, остальные показатели, такие как зольность, влажность и объемный вес, определяли в лаборатории Научного центра СамГТУ, используя стандартные аналитические методы [2, 17]. Для объекта Отрадный дополнительно определяли pH из водной вытяжки. На рис. 1 приведен вертикальный разрез свалки Безенчук с распределением отходов по возрасту. Показаны линзы осадков сточных вод и положение шурфа № 14, по которому проводилось кинетическое исследование в режиме компостирования [4].

3. МЕТОДЫ

3.1. Кинетический метод исследования

Кинетический подход к анализу состояния свалок основан на экспериментах, в которых образцы свалочного грунта подвергаются ускоренному разложению (компостированию) [13] на экспериментальных площадках в контролируемых условиях в теплое время года (май – июль). В течение 80 дней, один раз в декаду, производился отбор проб компостируемого материала и их анализ. Содержание органического вещества в свалочном грунте определяли прокаливанием предварительно высушенных образцов при температуре $600 \pm 5^\circ\text{C}$. Относительное содержание органического вещества определяли как отношение потерянной массы к массе образца до прокалывания. Этот показатель, обозначаемый далее как СОВ, использовался в [4]. Некоторые результаты этих экспериментов показаны на рис. 2, на котором отмечены экспериментальные (точки) и предсказанные (кривые) значения. Эти образцы были извлечены из одной и той же скважины № 14, с различных глубин, величины которых обозначены на рис. 2.

Изменение СОВ, обозначаемого как $C(t)$, для i -го образца в процессе компостирования может быть описано уравнением реакции первого порядка, т.е.

$$C(t) = \frac{(C_i - B_i)e^{-kt} + B_i(1 - C_i)}{(C_i - B_i)e^{-kt} + (1 - C_i)}. \quad (1)$$

В работе [4] показано, что константа скорости K является общей для всех исследованных свалок. Параметры C_i и B_i , представляющие начальное и квазиравновесное значения СОВ, индивидуальны для каждого образца i .

Для расчета возраста образцов мы предположили, что разложение отходов в условиях свалки также идет по уравнению реакции первого порядка, т.е.

$$A(t) = \frac{(A_0 - B_i)e^{-kt} + B_i(1 - A_0)}{(A_0 - B_i)e^{-kt} + (1 - A_0)}, \quad (2)$$

где A – величина СОВ в процессе естественного разложения, а A_0 – начальное значение СОВ. Константа скорости естественной деградации k отличается от константы скорости при компостировании K . Кроме того, мы предположили, что предельное значение B_i одинаково в уравнениях (1) и (2). Отсюда можно оценить возраст T_i каждого образца –

$$T_i = \frac{1}{k} \ln \left[\frac{(A_0 - B_i)(1 - C_i)}{(C_i - B_i)(1 - A_0)} \right]. \quad (3)$$

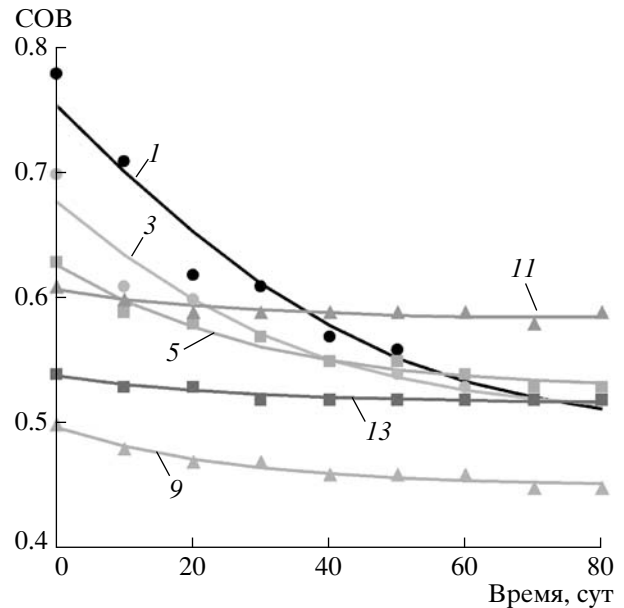


Рис. 2. Экспериментальные и предсказанные значения кинетики компостирования образцов свалки Безенчук. Числа указывают глубину, с которой были извлечены образцы.

Предполагая, что начальное значение СОВ на свалке (A_0) известно, можно рассчитать отношение T_i/T_j для каждой пары образцов. Если имеется “стандартный образец”, для которого известен возраст, например $T_1 = 1$, то можно рассчитать возраст для других образцов. В частности, возраст образцов, показанных на рис. 2, составляет 2, 4.4, 6.8, 14.5, 11.7, 13.5 лет (в соответствии с глубиной).

Помимо возрастов из данных по компостированию можно извлечь и другую полезную информацию. Очевидно, что начальное значение СОВ A_0 может изменяться в объеме свалки, например, оно специфично для линз осадков сточных вод. На рис. 2 видно, что кинетические кривые для образцов, извлеченных с глубины 9, 11 и 13 м, сильно отличаются от других, соседних кривых. Можно предположить, что эти образцы были отобраны из такой линзы, в которой начальное содержание органики было более высоким. Принимая это допущение, можно не только пересмотреть возраст образца, но и определить местоположение линзы осадков сточных вод.

Подробное описание кинетического подхода приведено в [4].

3.2. Метод главных компонент

Данные, с которыми мы работаем, представлены матрицей X , число строк в которой (I) равно числу исследованных образцов, а число столбцов (J) равно числу свойств (глубина, зольность, темпе-

ратура, и т.д.) этих образцов. Для анализа подобных матриц в хемометрике применяют методы *сжатия данных*. Этот подход позволяет представить данные в компактном виде, удобном для визуализации и интерпретации. С его помощью можно извлечь полезную информацию о внутренних связях и структурах в матрице \mathbf{X} .

Метод главных компонент (МГК) [5, 8, 16, 22] – один из популярных способов сжатия данных. С математической точки зрения МГК представляет разложение исходной матрицы данных \mathbf{X} , т.е. представление ее в виде произведения двух матриц \mathbf{T} и \mathbf{P}

$$\mathbf{X} = \mathbf{T}\mathbf{P}' + \mathbf{E}. \quad (4)$$

В этом уравнении \mathbf{X} – это исходная матрица данных размерностью $I \times J$, \mathbf{T} – матрица *счетов* размерностью $I \times N$, \mathbf{P} – матрица *нагрузок* размерностью $J \times N$, а \mathbf{E} – матрица *остатков* размерностью $I \times J$. Величина N называется *числом главных компонент* (ГК). МГК можно интерпретировать как проекцию исходных данных \mathbf{X} на пространство меньшей размерности N . Набор \mathbf{t}_i из I векторов-строк (размерностью N), составляющих матрицу \mathbf{T} , представляет проекции исходных образцов на новое подпространство. Набор из J векторов-строк \mathbf{p}_j (размерностью N), составляющих матрицу \mathbf{P} , представляет проекции исходных координатных векторов-переменных. Набор \mathbf{t}_n из N векторов-столбцов матрицы \mathbf{T} представляет проекцию всех образцов на n -ую ось главных компонент.

В результате преобразования (4) появляется возможность представить сложные данные в более простом виде так, чтобы исследователь смог “увидеть” результаты экспериментов в простой графической интерпретации. На *графике счетов* каждый образец изображается в координатах $(\mathbf{t}_k, \mathbf{t}_j)$. Чаще всего используют первые две ГК $(\mathbf{t}_1, \mathbf{t}_2)$. Близость двух точек-образцов на таком графике означает их схожесть, а отдаленность – различие. *График нагрузок* применяется для исследования роли переменных. На нем каждая исходная переменная отображается точкой (вектором) в координатах $(\mathbf{p}_k, \mathbf{p}_j)$, например $(\mathbf{p}_1, \mathbf{p}_2)$. Анализируя его аналогично графику счетов, можно понять, какие переменные связаны, а какие независимы. Так, близость точек на графике нагрузок означает положительную корреляцию соответствующих переменных. Точки, расположенные под прямым углом, соответствуют независимым переменным, а расположенные диаметрально противоположно-переменным, имеющим отрицательную корреляцию.

Перед выполнением декомпозиции (4), матрица \mathbf{X} должна быть преобразована с помощью центрирования и/или шкалирования [5]. Цель такой предварительной подготовки – преобразование

исходных данных в форму, наиболее удобную для анализа. *Центрирование* – это вычитание из исходной матрицы \mathbf{X} матрицы \mathbf{M} , т.е. $\mathbf{X}' = \mathbf{X} - \mathbf{M}$. Обычно центрирование проводится по столбцам: для каждого вектора \mathbf{x}_j вычисляется среднее значение

$$m_j = \frac{1}{I} \sum_{i=1}^I x_{ij},$$

тогда $\mathbf{M} = (m_1\mathbf{1}, \dots, m_J\mathbf{1})$, где $\mathbf{1}$ – это вектор из единиц размерности I .

Второе простейшее преобразование данных – шкалирование. *Шкалирование* по столбцам – это умножение матрицы \mathbf{X} справа на диагональную матрицу $\mathbf{W} = \text{diag}(w_{jj})$, т.е. $\mathbf{X}' = \mathbf{X}\mathbf{W}$. При этом размерность матрицы \mathbf{W} равна $J \times J$, а ее элементы w_{jj} – это обратные значения стандартных отклонений столбцов матрицы \mathbf{X} . Шкалирование данных применяют для того, чтобы уравнивать вклад в модель от различных переменных. В нашем случае X -переменные имеют различные единицы измерения (метры, концентрации, градусы и т.п.), и для их совместной обработки необходимо шкалирование.

Число главных компонент (N) выбирается с использованием различных критериев. Одним из них является величина объясненной дисперсии E_n –

$$E_n = 100 \left(1 - \frac{\sum_{i=1}^I \sum_{j=1}^J e_{ij}^2}{\sum_{i=1}^I \sum_{j=1}^J x_{ij}^2} \right), \quad (5)$$

$$n = 1, \dots, N,$$

которая вычисляется по матрице остатков $\mathbf{E} = \{e_{ij}\}$. Величина E_n обычно изображается на графике в зависимости от числа n ; тогда ее резкое изменение указывает на нужное значение числа главных компонент N .

3.3. Метод проекций на латентные структуры

Метод *проекций на латентные структуры* (ПЛС) [5, 8], в отличие от МГК, работает с двумя матрицами – \mathbf{X} и \mathbf{Y} . Матрица \mathbf{Y} состоит из *зависимых* переменных (отклики), в то время как \mathbf{X} содержит соответствующие *независимые* переменные (предикторы). В методе ПЛС строится многомерная регрессионная (калибровочная) модель $\mathbf{Y} = \mathbf{X}\mathbf{B}$, т.е. определяются неизвестные параметры \mathbf{B} . Далее полученная модель используется для определения новых значений \mathbf{y} по новым измерениям \mathbf{x} . Смысл этой модели состоит в том, чтобы получить возможность в дальнейшем не измерять значения \mathbf{Y} , а определять их по значениям \mathbf{X} .

Чтобы в первом приближении понять, как работает метод ПЛС, проще всего его представить как два МГК-анализа, производимых одновре-

менно – для \mathbf{X} и \mathbf{Y} . Ниже представлены соответствующие разложения, аналогичные (4):

$$\mathbf{X} = \mathbf{TP}^t + \mathbf{E}, \quad \mathbf{Y} = \mathbf{UQ}^t + \mathbf{F}. \quad (6)$$

Здесь \mathbf{T} и \mathbf{P} – это матрицы ПЛС счетов и ПЛС нагрузок для \mathbf{X} , а \mathbf{U} и \mathbf{Q} – это ПЛС счета и ПЛС нагрузки для \mathbf{Y} . Проекция строится согласованно – так, чтобы максимизировать корреляцию между соответствующими векторами \mathbf{t}_n и \mathbf{u}_n . Эта величина, $\text{cor}(\mathbf{t}_n, \mathbf{u}_n)$, играет важную роль при выборе числа главных ПЛС компонент, N . Помимо нее, для оценки точности ПЛС модели используют еще две характеристики. Это – *среднеквадратичный остаток калибровки (RMSEC)*, который вычисляется по формуле

$$RMSEC = \sqrt{\sum_{i=1}^I (y_i - \hat{y}_i)^2 / F}, \quad (7)$$

где F – это число степеней свободы, y_i и \hat{y}_i , соответственно, измеренные и предсказанные значения откликов для образцов обучающего набора, т.е. того набора данных, который использовался для построения модели. Величина *RMSEC* характеризует точность моделирования (калибровки).

Точность прогноза по построенной модели характеризуют *среднеквадратичным остатком предсказания (RMSEP)*, вычисляемым аналогично (7) по формуле

$$RMSEP = \sqrt{\sum_{l=1}^L (y_l - \hat{y}_l)^2 / L}. \quad (8)$$

Для ее вычисления можно использовать независимый проверочный набор, и тогда L – это число образцов в таком наборе, но можно прибегнуть к процедуре перекрестной самопроверки [8]. Именно последний алгоритм применялся в работе.

Алгоритмы МГК и ПЛС приведены в [3]; там же опубликованы и Matlab процедуры для их осуществления.

3.4. Анализ данных

В нашем исследовании матрицы \mathbf{X} формировались для каждой свалки. Строка матрицы \mathbf{X} соответствует одному образцу, отобранному на свалке. Столбцы матрицы \mathbf{X} представляют измеряемые и оцениваемые свойства образцов. Они разделялись на два блока. Первый блок, обозначаемый $\mathbf{X1}$, состоит из измеряемых свойств, таких как зольность (COB), температура, объемный вес, рН, влажность и глубина. Второй блок, обозначаемый $\mathbf{X2}$, включает переменные, оцениваемые с помощью традиционных методов, описанных в работе [4]. Эти переменные являются индексами, обозначающие то, что образец отно-

сится к некоторой специфичной области, такой как линзы осадков сточных вод, слои промышленных отходов, основание свалки и т.д.

Другая оцениваемая переменная – это возраст образца T . Однако этот показатель можно определить только для тех образцов, которые представляют массив отходов. Для образцов, полученных из основания свалки, возраст не определяем. Для того, чтобы преодолеть эту трудность, предлагается заменить возраст T на другую характеристику, названную *стабильностью*, M . Она рассчитывается по формуле

$$M_i = 1 - e^{-mT_i}, \quad (9)$$

где M_i – стабильность образца i , T_i – его возраст, а m – это константа. Такое преобразование можно интерпретировать как некоторую модификацию (предобработку) данных, при которой исходная переменная T заменяется новой переменной M . Константа m выбирается так, чтобы линеаризовать ПЛС модель (т.е. увеличить корреляцию между \mathbf{t} и \mathbf{u}), имея в виду, что ПЛС моделирование предполагает наличие линейной зависимости между счетами \mathbf{T} и \mathbf{U} . Использовались следующие значения m : для свалки Безенчук $m = 1/15$, для Отрадного $m = 1/5$. Очевидно, что величина m связана с константой скорости разложения k (2), оценки которой приведены в [4], однако для установления этой связи нам потребуется большее число объектов, нежели те два, с которыми мы работали.

Из формулы (9) видно, что стабильность образцов, отобранных из основания свалки, равняется единице ($T = \infty$), в то время как стабильность новых отходов равна нулю. Переменная *стабильность* (обозначаемая в дальнейшем на графиках как *Maturity*) использовалась для данных, полученных со свалок Безенчук и Отрадный, потому что эти объекты не имеют системы защиты окружающей среды, и образцы, собранные с них, могут принадлежать не только телу свалки, но и основанию. В то же время полигон Кинель имеет такую защиту: слои отходов укладывались с послойной пересыпкой. Основание этого полигона было покрыто полимерной пленкой; все образцы с полигона Кинель заведомо принадлежат телу свалки и им могут быть приписаны значения возраста. Поэтому для этого объекта нет необходимости в преобразовании возраста в стабильность. При ПЛС моделировании возраст или стабильность представлялись как отклик \mathbf{Y} .

Данные по каждому объекту анализировались в два этапа. Сначала строилась МГК модель с использованием только блока $\mathbf{X1}$. Эта модель позволила выявить группирование образцов и специфичные области, которые могут быть связаны с блоком оцениваемых переменных $\mathbf{X2}$. На втором этапе строились две ПЛС регрессии для предска-

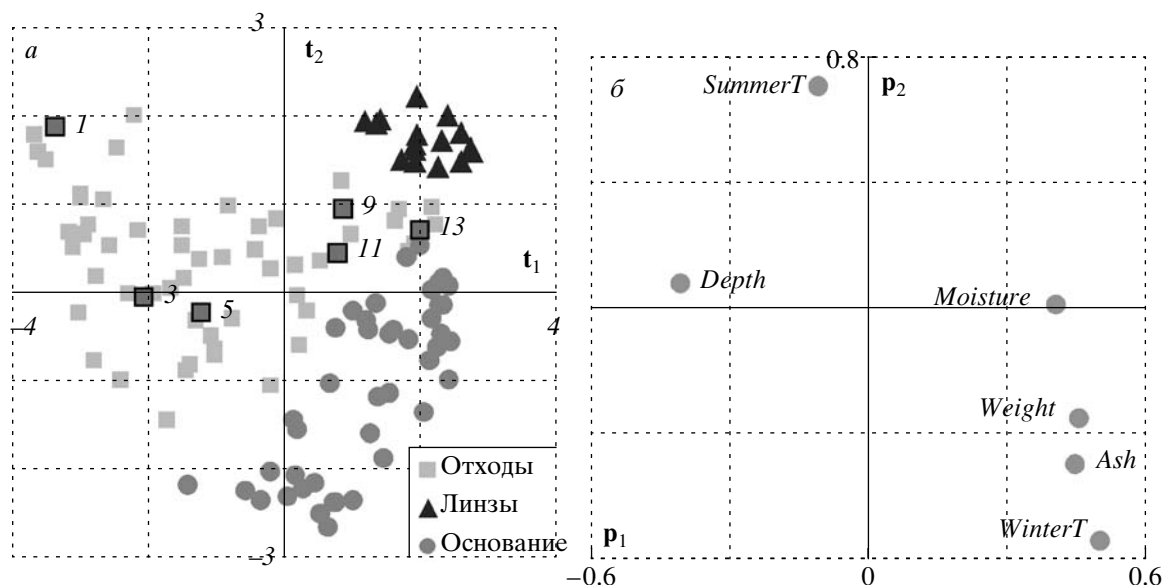


Рис. 3. Безенчук. МГК модель для матрицы X1, первые 2 компоненты: а – график счетов, выделены образцы, представленные на рис. 2; б – график нагрузок.

зания возраста или стабильности образцов. Короткая ПЛС модель использовала блок X1 как независимые переменные. Объединенная матрица $X = (X1, X2)$ использовалась при построении полной ПЛС модели. Сравнение результатов, полученных с помощью обеих моделей, позволяет понять полезность использования добавочных переменных X2, например, таких как индексация слоев.

4. РЕЗУЛЬТАТЫ

4.1. Безенчук

Набор данных Безенчук состоит из 123 образцов. Блок данных X1 (123×6) включает в себя шесть измеряемых переменных: глубина (*Depth*), зольность (*Ash*), объемная масса (*Weight*), влажность (*Moisture*), температура летом (*SummerT*) и температура зимой (*WinterT*). Блок X2 включает две переменные-индексы: принадлежность образца к основанию свалки (*Base*) и к существующим линзам осадков сточных вод (*Lens*). В скобках приведены условные обозначения переменных, которые используются далее на рисунках. МГК модель с двумя главными компонентами ($N = 2$), построенная на матрице данных X1, объясняет 88% общей дисперсии (5).

При МГК анализе данных Безенчук ставилась задача выделения специфичных областей, таких как линзы осадков сточных вод и основание свалки, используя только измеряемые переменные, представленные матрицей данных X1. На рис. 3а видно как группируются образцы на графике счетов (t_1 – t_2), где отмечены образцы, принадлежа-

щие к линзам осадков сточных вод (треугольники) и основанию свалки (кружки). Образцы, показанные на рис. 2, выделены цифрами. Видно, что специфические образцы 9, 11 и 13 действительно близки к группе образцов, представляющих линзу осадков сточных вод. Также видно, что образец 1, взятый с поверхности свалки, сильно отличается от других образцов. Об этом свидетельствует его экстремальное положение на графике счетов (левый верхний угол).

Совместный анализ графиков счетов и нагрузок (рис. 3а и б) помогает понять, какие переменные наиболее сильно влияют на разделение образцов на группы. В исследуемом примере переменная, отвечающая за летнюю температуру (*SummerT*), разделяет образцы вдоль оси t_2 , что и характеризует различные группы. Кроме того, можно заметить, что переменные зольность (*Ash*) и объемный вес (*Weight*) расположены очень близко друг к другу на графике нагрузок (рис. 3б). Это означает, что между этими показателями имеется сильная корреляция. Непосредственные вычисления показывают, что $\text{cor}(\text{Weight}, \text{Ash}) = 0.89$.

Таким образом, видно, что МГК позволяет визуально исследовать набор данных, выделить специфичные области в теле массива, используя только объективную информацию (т.е. непосредственно измеряемые переменные), и подтвердить результаты, полученные в работе [4] с помощью кинетического подхода.

Второй задачей было предсказание стабильности. Для этого использовались две ПЛС модели. Первая, короткая модель, использовала матрицу данных X1 (123×6) как предикторы и стабиль-

Таблица 1. Короткая и полная ПЛС модели свалки Безенчук

ГК	Короткая модель			Полная модель		
	<i>RMSEC</i>	<i>RMSEP</i>	cor(t , u)	<i>RMSEC</i>	<i>RMSEP</i>	cor(t , u)
1	0.171	0.175	0.86	0.111	0.115	0.95
2	0.135	0.14	0.61	0.092	0.095	0.56
3	<i>0.120</i>	<i>0.126</i>	<i>0.46</i>	<i>0.078</i>	<i>0.083</i>	<i>0.53</i>
4	0.119	0.126	0.14	0.077	0.082	0.14

Примечание. Курсивом выделена выбранная ГК.

ность **M** как отклик **Y**. В полной модели матрица предикторов включала в себя также и оцененные величины, т.е. еще блок **X2**. Обе регрессионные модели использовали три главные ПЛС компоненты. В таблице 1 представлены основные характеристики этих регрессионных моделей.

Можно отметить, что использование субъективных переменных (блок **X2**) дает некоторый выигрыш в точности прогноза. Предсказание стабильности с помощью короткой модели представлено на рис. 4а. Незакрашенными кружками обозначены образцы, приведенные на рис. 2.

Результаты ПЛС анализа на полной модели помогают более детально исследовать взаимосвязи различных переменных. График ПЛС нагрузок для первых двух ГК (рис. 4б) показывает, что стабильность (*Maturity*) имеет сильную отрицательную корреляцию с глубиной (*Depth*), а переменная *Base* (принадлежность к основанию массива) коррелирует с летней температурой (*SummerT*). Непосредственные вычисления

показывают, что $cor(Depth, Maturity) = -0.92$, $cor(SummerT, Base) = -0.72$.

Хемометрический подход помогает проанализировать полученные данные. МГК позволяет выделить образцы линз и основания массива, используя только измеряемые значения (матрица данных **X1**). ПЛС регрессия, построенная только с использованием матрицы **X1** в качестве предикторов, позволяет с удовлетворительной точностью предсказывать стабильность образцов. Исследование графиков нагрузок позволяет выявлять скрытые зависимости между различными переменными.

4.2. Кинель

Набор данных Кинель состоит из 105 образцов. Матрица данных **X1** (105×4) включает в себя следующие переменные: зольность, объемный вес, летняя температура и глубина. Блок **X2** состоит только из одной переменной-индекса, которая обозначает принадлежность образца к одно-

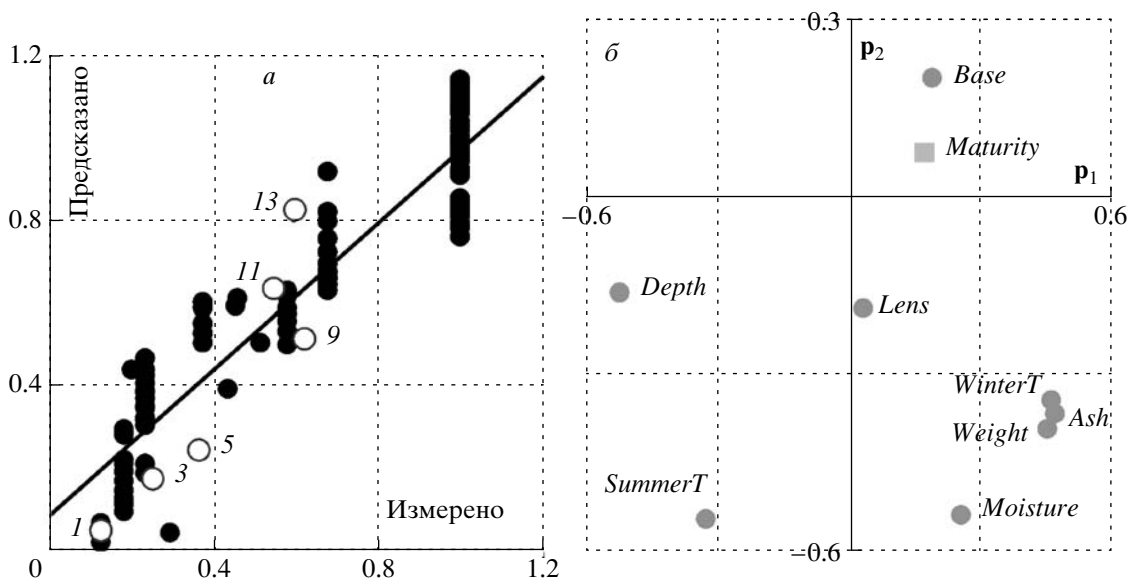


Рис. 4. Моделирование объекта Безенчук: а – короткая ПЛС модель с 3 ГК – предсказанные значения против измеренных, выделены образцы, представленные на рис. 2; б – полная ПЛС модель, график нагрузок.

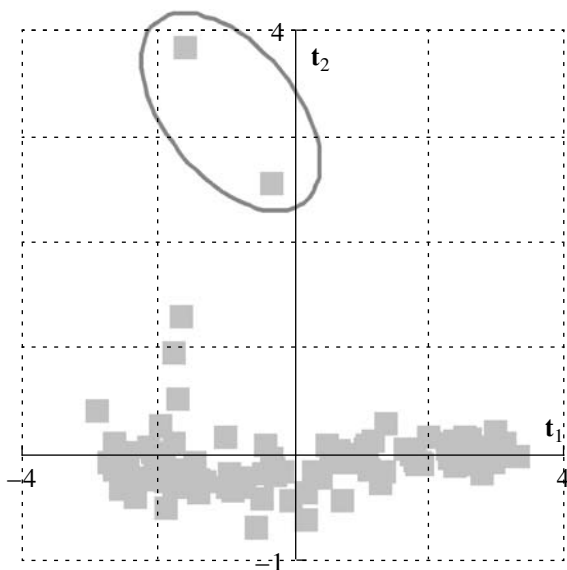


Рис. 5. Моделирование объекта Кинель. МГК модель блока $X1$. График счетов. Отмечены два выброса – предположительно промышленные отходы.

му из четырех слоев отходов (*Layer*). Так как Кинель – это усовершенствованный полигон, значения глубин и число слоев известно априори до обработки данных. По этой причине нет необходимости в использовании преобразования возраст/стабильность, и предсказываемая переменная Y – это возраст (*Age*).

Декомпозиция матрицы $X1$ с помощью МГК показала наличие двух выпадающих образцов (рис. 5). Известно, что в этом массиве имеются участки, где размещались промышленные отходы. По всей видимости, эти два образца принадлежат к такому участку. Аналогичный вывод был нами сделан в работе [4], где рассматривался один из этих образцов. Таким образом, в этом примере МГК позволил разделить промышленные и бытовые отходы. Дальнейший анализ данных полигона Кинель проводился уже без этих образцов.

Для исследования взаимосвязи между различными переменными была построена ПЛС модель с матрицей $X1$, используемой как блок независи-

мых переменных, и матрицей $X2$ в качестве отклика. Две ПЛС компоненты объясняют 97% вариации данных $X1$ и 91% вариации данных $X2$. Совместный анализ графиков $X1$ - и $X2$ -нагрузок показывает, что переменная *Layer* имеет сильную корреляцию с зольностью и глубиной.

Для предсказания возраста Y были построены две ПЛС модели. Короткая модель использовала матрицу данных $X1$ как независимые переменные, в то время как полная модель – объединенную матрицу данных $X1$ и $X2$. Обе регрессионные модели описываются двумя ПЛС компонентами. В предыдущем примере мы предсказывали значение стабильности, которое изменялось от 0 до 1. Чтобы облегчить сравнение этих двух наборов данных, мы отшкалировали переменную *Age* для полигона Кинель так, чтобы она менялась в том же диапазоне. В табл. 2 представлены некоторые важные характеристики короткой и полной моделей. Сравнивая табл. 1 и 2, можно заметить, что нет явных преимуществ в предсказании возраста для усовершенствованного полигона Кинель над несанкционированной свалкой Безенчук. Из табл. 2 также видно, что дополнительная переменная *Layer* – принадлежность образца к слою – существенно улучшает предсказание возраста. На рис. 6б видна большая отрицательная корреляция между переменными *Age* и *Layer*, равная – 0.99. Это можно объяснить тем, что полигон Кинель имеет четкую вертикальную слоистую структуру. График ПЛС счетов (рис. 6а) показывает, как распределяются образцы. Четыре группы образцов, принадлежащие различным слоям, также выделяются с помощью короткой ПЛС модели, где переменная *Layer* не использовалась.

Предсказание возраста кажется не столь необходимым для усовершенствованного полигона, когда сроки создания и заполнения массива известны заранее. С другой стороны, ввиду гетерогенности состава отходов процессы биоразложения могут идти неоднобразно и не непрерывно [10]. Исследование различий между известным и предсказанным возрастом дает возможность точнее определить стабильность и помочь при дальнейшем исследовании и эксплуатации массива отходов.

Таблица 2. Короткая и полная ПЛС модели полигона Кинель

ГК	Короткая модель			Полная модель		
	<i>RMSEC</i>	<i>RMSEP</i>	$\text{cor}(\mathbf{t}, \mathbf{u})$	<i>RMSEC</i>	<i>RMSEP</i>	$\text{cor}(\mathbf{t}, \mathbf{u})$
1	0.088	0.090	0.96	0.069	0.070	0.98
2	0.081	0.085	0.41	0.041	0.043	0.80
3	0.079	0.083	0.17	0.035	0.037	0.53
4	0.079	0.083	0.01	0.033	0.035	0.33

Примечание. Курсивом выделена выбранная ГК.

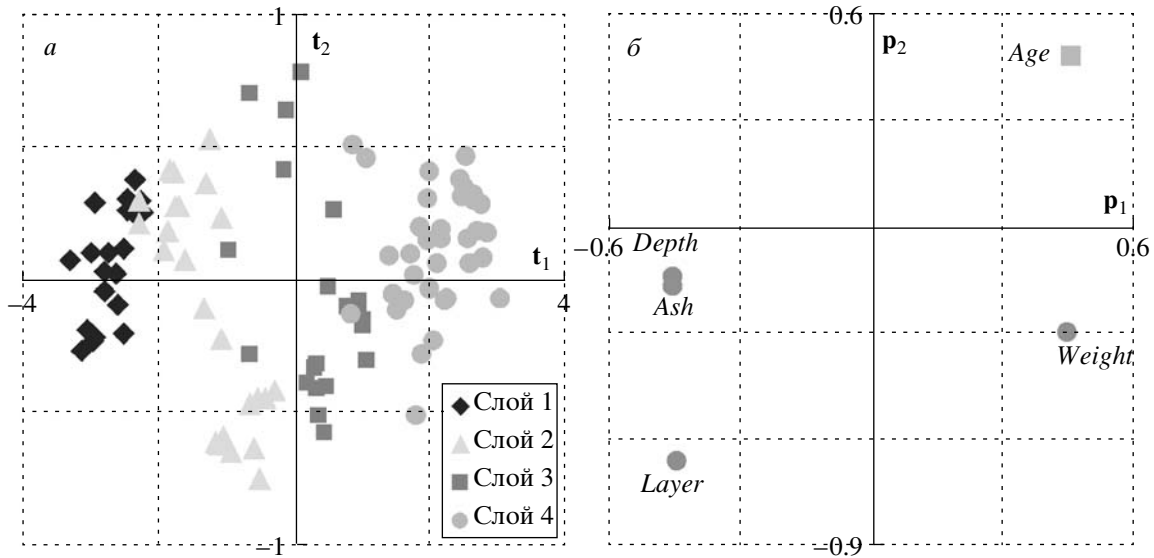


Рис. 6. Моделирование объекта Кинель: *a* – короткая ПЛС модель, график счетов. Четыре группы представляют различные слои; *б* – полная ПЛС модель, график нагрузок.

4.3. Отрадный

Набор данных состоит из 84 образцов. Матрица данных **X1** (84×5) включает в себя следующие параметры: зольность, объемная масса, глубина, влажность и рН. Блок данных **X2** состоит из одной переменной-индекса (*Layer*), которая обозначает принадлежность образца к слоям; откликом **Y** является стабильность.

Эта свалка имеет неправильную структуру, так как в процессе ее формирования постоянно изменялись как виды отходов, так и технология их складирования. Таким образом, мы не можем доверять переменной *Layer*. Модель МГК, построенная на матрице **X1**, подтверждает это, так как мы не наблюдаем разделения по слоям на графике счетов. ПЛС анализ **X1** → **X2**, который был полезен на примере полигона Кинель, также не дал интересных результатов.

В этом примере мы вновь используем показатель стабильности *M*, вычисляемый по формуле (9). Были построены две ПЛС регрессии: короткая модель (матрица данных **X1**) и полная

модель (объединенные матрицы **X1** и **X2**). Обе модели описываются двумя главными компонентами. Основные характеристики моделей представлены в табл. 3. Видно, что значения *RMSEP* для этих моделей практически совпадают.

На рис. 7 представлены результаты для полной ПЛС модели. На графике нагрузок (*б*) видно, что переменные *Maturity* и *Layer* ортогональны. Это объясняет, почему короткая и полная модели дают схожие результаты. Как видно из графика счетов на рис. 7*a*, здесь нет четкого разделения образцов на отдельные слои.

5. ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

Стабилизация свалки – это длительный процесс. Кроме того, время, за которое отходы полностью разложатся, и свалка станет стабильной, не может быть четко определено. Возможность детального исследования и мониторинга свалок особенно важны в свете новых технологий, которые повышают скорость биологического разло-

Таблица 3. Короткая и полная ПЛС модели свалки Отрадный

ГК	Короткая модель			Полная модель		
	<i>RMSEC</i>	<i>RMSEP</i>	<i>cor(t, u)</i>	<i>RMSEC</i>	<i>RMSEP</i>	<i>cor(t, u)</i>
1	0.107	0.111	0.90	0.115	0.120	0.88
2	0.098	0.104	0.41	0.097	0.104	0.53
3	0.097	0.105	0.14	0.094	0.105	0.26
4	0.097	0.105	0.01	0.093	0.104	0.11

Примечание. Курсивом выделена выбранная ГК.

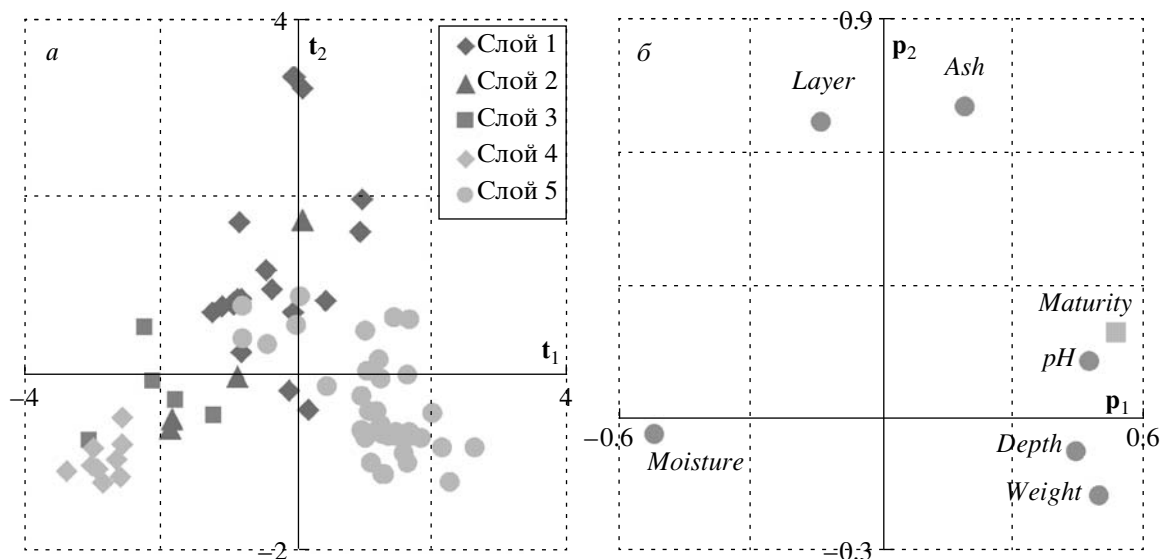


Рис. 7. Моделирование объекта Отрадный. Полная ПЛС модель: а – график ПЛС счетов с выделенными слоями; б – график ПЛС нагрузок.

жения и помогают сократить время стабилизации свалок [7]. Аналитические и кинетические методы совместно с анализом многомерных данных помогают установить состояние свалки и оценить степень разложения органического вещества.

Исследование трех массивов различных типов подтверждает пригодность применения многомерного анализа для целей экологического мониторинга. Хемометрические методы дают возможность установить структуру свалки; выявить специфичные области. Предсказание таких показателей, как стабильность и возраст, может быть выполнена с помощью ПЛС регрессии, базирующейся только на измеряемых, объективных данных. Предсказанные значения близки к результатам, полученным с помощью традиционного кинетического подхода, описанного в работе [4].

Особое внимание стоит обратить на полезность использования дополнительных переменных, представленных блоком X2. Основная хемометрическая идея – любая априорная информация помогает понять структуру данных, и во

время анализа она должна приниматься в расчет [5]. С другой стороны, пользоваться этой информацией нужно с осторожностью, так как во многих случаях она не является достоверной. В первом примере (Безенчук) априорная информация о структуре свалки (линзы осадков сточных вод) полностью подтвердилась результатами, полученными с помощью многомерного анализа. Поэтому эта дополнительная информация дала некоторое улучшение при предсказании стабильности (табл. 4). То же самое мы получили и во втором примере (Кинель). Однако в третьем случае (Отрадный) априорная информация о структуре свалки не дала никакого выигрыша при прогнозе. Это объясняется тем, что многие свалки не имеют определенной структуры, слои с определенным возрастом могут располагаться не по порядку, и обнаружение специфичных областей, как на примере свалки Безенчук, является большой удачей. В табл. 4 представлены результаты, полученные при использовании различного набора переменных в ПЛС анализе.

Несмотря на то, что каждый образец характеризуется только 5–6 переменными, они очень сильно взаимосвязаны, вот почему многомерный подход дал такие интересные результаты. С другой стороны, в будущем мы планируем провести более детальное исследование и получить больше характеристик образцов. На данном этапе мы считаем основным результатом работы подтверждение применимости хемометрических методов для анализа состояния свалок.

Таблица 4. Итоговая таблица. Сравнение результатов ПЛС моделирования и предсказания для трех рассмотренных объектов

Объект	X1 переменные		X1 + X2 переменные	
	RMSEC	RMSEP	RMSEC	RMSEP
Безенчук	0.120	0.126	0.078	0.083
Кинель	0.081	0.085	0.041	0.043
Отрадный	0.098	0.104	0.097	0.104

СПИСОК ЛИТЕРАТУРЫ

1. Инженерно-экологические изыскания для строительства: СП 11–102–97. М.: ПНИИИС Госстроя России, 1997. 41 с.
2. Методика выполнения измерений массовой доли золы, влаги (влажности) в твердых отходах гравиметрическим методом. М.: МПР РФ, 2002. 16 с.
3. Михайлов Е.В., Померанцев А.Л. MatLab. Руководство для начинающих, [On line] <http://www.chemometrics.ru/materials/textbooks/matlab.htm> (1 июня 2008).
4. Михайлов Е.В., Померанцев А.Л. Анализ структуры и состояния полигонов и свалок твердых бытовых отходов. Кинетический подход // Геоэкология. 2009. №. 1. С. 61–70.
5. Родионова О.Е., Померанцев А.Л. Хемометрика: достижения и перспективы // Успехи химии. 2006. Т. 75. С. 302–321.
6. Тупицина О.В. Рекультивации массивов органоминеральных отходов с использованием компостирования // Автореф. дис. ... канд. техн. наук. Самара: Самарский гос. технич. ун-т, 2006.
7. Чертес К.Л., Быков Д.Е. Рекультивация карьеров отходами. Самара: Самарский гос. техн. ун-т, 2005. 292 с.
8. Эсбensen К. Анализ многомерных данных. Сокр. пер. с англ. Под ред. О.Е. Родионовой. Черногловка: Изд-во ИПХФ РАН, 2005 [Esbensen K.H. Multivariate Data Analysis – In Practice 4-th Ed., CAMO, 2000].
9. Ecke H. Sequestration of metals in carbonated municipal solid waste incineration (MSWI) fly ash // Waste Management. 2003. V. 23. P. 631–640.
10. Haarstrick A., Hempel D.C., Ostermann L., Ahrens H., Dinkler D. Modelling of the biodegradation of organic matter in municipal landfills // Waste Management & Research. 2001. V. 19. P. 320–331.
11. Hansson M., Nordberg A., Mathisen B. On-line NIR monitoring during anaerobic treatment of municipal solid waste // Water Sci. Technol. 2003. V. 48(4). P. 9–13.
12. Hoskuldsson A., Rodionova O., Pomerantsev A. Path modeling and process control // Chemom. Intell. Lab. Syst. 2007. V. 88 (1). P. 84–99.
13. Komilis D.P. A kinetic analysis of solid waste composting at optimal conditions // Waste Management. 2006. V. 26. P. 82–91.
14. Mikhailov E.V., Tupicina O.V., Bykov D.E., Chertes K.L. et al. Ecological assessment of landfills with multivariate analysis – a feasibility study // Chemom. Intell. Lab. Syst. 2007. V. 88 (1). P. 3–10.
15. Mora-Naranjo N., Meima J.A., Haarstrick A., Hempel D.C. Modelling and experimental investigation of environmental influences on the acetate and methane formation in solid waste // Waste Management. 2004. V. 24. P. 763–773.
16. Næs T., Isaksson T., Fearn T., Davies T. Multivariate Calibration and Classification. Chichester: NIR Publications, 2002.
17. Smidt E., Lechner P. Study on the degradation and stabilization of organic matter in waste by means of thermal analyses, // Thermochimica Acta. 2005. V. 438. P. 22–28.
18. Smidt E., Meissl K., Tintner J. Investigation of 15-year-old municipal solid waste deposit profiles by means of FTIR spectroscopy and thermal analysis // J. Environ. Monit. 2007. V. 9. P. 1387–1393.
19. Tesar M., Prantl R., Lechner P. Application of FT-IR for assessment of the biological stability of landfilled municipal solid waste (MSW) during in situ aeration // J. Environ. Monit. 2007. V. 9. P. 10–118.
20. Vavilin V.A., Rytov S.V., Lokshina L.Ya., Pavlostathis S.G., Barlaz M.A. Distributed model of solid waste anaerobic digestion. Effect of leachate recirculation and pH adjustment // Biotechnol. Bioengn. 2003. V. 81(1). P. 66–73.
21. Vavilin V.A., Rytov S.V., Pavlostathis S.G., Jokela J., Rintala J. A distributed model of solid waste anaerobic digestion: sensitivity analysis // Wat. Sci. Technol. 2003. V. 48(4). P. 147–154.
22. Wold S., Esbensen K., Geladi P. Principal component analysis // Chemom. Intell. Lab. Syst. 1987. V. 2. P. 37–42.
23. Zbytniewski R., Buszewski B. Characterization of natural organic matter (NOM) derived from sewage sludge compost. Part 1: Chemical and spectroscopic properties // Bioresour. Technol. 2005. V. 96. P. 471–478.
24. Zbytniewski R., Buszewski B. Characterization of natural organic matter (NOM) derived from sewage sludge compost. Part 2: Multivariate techniques in the study of compost maturation // Bioresour. Technol. 2005. V. 96. P. 479–484.