

**С. А. Буриков, А. О. Ефиторов, Т. А. Доленко, В. Р. Широкий, С. А. Доленко**

*Московский государственный университет им. М. В. Ломоносова  
Ленинские горы, д. 1, стр. 2, Москва, ГСП-1, 119991, Россия*

*sergey.burikov@gmail.com*

## **РЕШЕНИЕ ОБРАТНЫХ ЗАДАЧ СПЕКТРОСКОПИИ КОМБИНАЦИОННОГО РАССЕЯНИЯ ВОДНЫХ РАСТВОРОВ СОЛЕЙ С ПРИМЕНЕНИЕМ ВЕЙВЛЕТ-НЕЙРОННЫХ СЕТЕЙ \***

Представлены результаты решения задачи по определению солевого состава многокомпонентных водных растворов по их спектрам комбинационного рассеяния света с использованием искусственных нейронных сетей и метода проекций на латентные структуры. Сравниваются три метода извлечения признаков из спектральных кривых: агрегация каналов, дискретное и непрерывное вейвлет-преобразование. Показано, что для нейросетевого решения обратной задачи определения солевого состава наилучшие результаты обеспечивает непрерывное вейвлет-преобразование входных данных.

*Ключевые слова:* спектроскопия комбинационного рассеяния, водные растворы солей, проекции на латентные структуры, искусственные нейронные сети, вейвлет-анализ.

### **Введение**

Определение солевого состава многокомпонентных водных растворов является чрезвычайно актуальной задачей экологического мониторинга природных и пластовых вод, контроля сбросовых вод промышленных предприятий и т. д. [1]. В стоках промышленных предприятий могут содержаться токсичные вещества – соли аммония, аммиак, сероводород, сернистые соединения, соли тяжелых металлов, галогены и др. Многие из перечисленных неорганических соединений в воде присутствуют в диссоциированной форме, поэтому задача определения их концентрации сводится к определению концентрации ионов.

Решение задачи диагностики солевого состава вод в приложении к экологическому мониторингу водных сред выдвигает определенные требования к разрабатываемому методу. Он должен быть экспрессным (работающим в режиме онлайн) и дистанционным, а также не должен предполагать сложную пробоподготовку [2; 3]. Спектроскопия комбинационного рассеяния (КР) полностью отвечает этим требованиям [4].

Принципиальная возможность определения типа и концентрации ионов в многокомпонентных водных растворах с помощью спектроскопии КР обусловлена тем, что полосы КР водного раствора весьма чувствительны к присутствию в нем растворенных ионов. Сложные по составу ионы

---

\* Работа выполнена при поддержке гранта РФФИ № 17-07-01479.

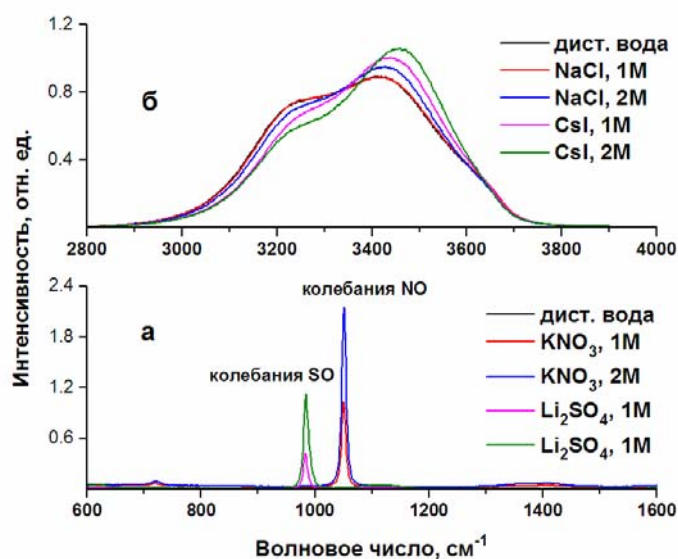


Рис. 1. Спектры КР водных растворов неорганических солей и дистиллированной воды

(нитраты  $\text{NO}_3^-$ , сульфаты  $\text{SO}_4^{2-}$ , фосфаты  $\text{PO}_4^{3-}$  и т. д.) имеют собственные колебательные линии в области  $500\text{--}1700\text{ см}^{-1}$ . Это позволяет однозначно идентифицировать ионы и определять их концентрацию по интенсивности соответствующих линий (рис. 1, а) [5; 6]. Присутствие в растворе простых ионов, не имеющих собственных линий КР, также проявляется в спектрах КР водных растворов. Такие ионы, как  $\text{Cl}^-$ ,  $\text{I}^-$ ,  $\text{Br}^-$ ,  $\text{Na}^+$ ,  $\text{K}^+$  и т. д., влияют на форму и положение наиболее интенсивной полосы спектра – полосы валентных колебаний ОН-групп в области  $3000\text{--}4000\text{ см}^{-1}$  (рис. 1, б) [4; 7]. При этом различные ионы влияют на валентную полосу по-разному, что обеспечивает успешное применение метода спектроскопии КР для определения содержания конкретных ионов в воде – нитратов [8], растворенного  $\text{CO}_2$  [9], солености морской воды [10].

При решении многих реальных задач диагностики водных сред ситуация осложняется тем, что в растворе присутствует большое количество разных ионов, каждый из которых по-разному влияет на спектр КР. Эти ионы сложным нелинейным образом взаимодействуют друг с другом. На рис. 2 приведены спектры некоторых многокомпонентных растворов. Спектральные области, в которых расположены колебательные линии сложных ионов и валентная полоса

КР воды, очевидно, наиболее чувствительны к типу и концентрации растворенных ионов. Области деформационной ( $1600\text{--}1700\text{ см}^{-1}$ ) и ассоциативной ( $2000\text{--}2400\text{ см}^{-1}$ ) полос КР воды гораздо менее информативны. Таким образом, информативность разных спектральных каналов неодинакова.

Ситуация осложняется еще и тем, что спектральные полосы, которые необходимо анализировать одновременно, существенно отличаются друг от друга как по интенсивности, так и по ширине (см. рис. 2). В настоящий момент не существует адекватной физической модели, которая позволяла бы численно получить зависимость спектра КР воды от концентраций растворенных солей, особенно с учетом их нелинейных взаимодействий. Поэтому для решения обратной задачи определения типа и концентрации одновременно многих ионов в воде по спектрам КР необходимо применять современные адаптивные методы анализа данных.

Предыдущие работы авторов данной статьи показали, что подобные многопараметрические обратные задачи лазерной спектроскопии эффективно решаются с помощью искусственных нейронных сетей (ИНС) [11; 12]. ИНС представляют собой математические алгоритмы, которые успешно используются при решении задач обработки и распознавания образов и изображений, прогнозирования, задач, решение которых плохо поддается алгоритмизации и т. д. [13]. Ос-

новными преимуществами нейронных сетей являются высокая устойчивость к шумам, независимость от характера шумов, способность работать на неполных или противоречивых данных.

Помимо ИНС, описанный круг задач способны решать проекционные методы; одним из наиболее эффективных является метод проекций на латентные структуры (ПЛС) [14], успешно применяемый для построения регрессионных моделей и классификаторов [14; 15].

Ранее авторами настоящей статьи с помощью ИНС и метода ПЛС была успешно решена задача идентификации и одновременного определения концентрации каждой из пяти солей – NaCl, NH<sub>4</sub>Br, Li<sub>2</sub>SO<sub>4</sub>, KNO<sub>3</sub>, CsI – в воде по спектрам КР водных растворов [16]. Показано, что наилучших результатов в случаях, когда нелинейность задачи не слишком велика, удается добиться проекционным методом ПЛС. В случаях существенной нелинейности нейросетевой метод решения задачи оказывается более эффективным. В среднем применение обоих методов обеспечило одновременное определение концентрации каждой из 5-ти указанных солей в воде с погрешностью 0,037 М в диапазоне концентраций солей 0–2,5 М [16]. С помощью ИНС была также решена задача идентификации и одновременного определения концентрации каждого из 10-ти ионов в воде: Mg<sup>2+</sup>, Li<sup>+</sup>, NO<sub>3</sub><sup>-</sup>, NH<sub>4</sub><sup>+</sup>, F<sup>-</sup>, SO<sub>4</sub><sup>2-</sup>, K<sup>+</sup>, HCO<sub>3</sub><sup>-</sup>, Na<sup>+</sup>, Cl<sup>-</sup> [12]. Точность определения концентрации составила в среднем около 0,06 М.

Специфика спектроскопических методов с точки зрения обработки и анализа данных состоит в извлечении полезной информации из данных высокой размерности, так как регистрируемые спектры содержат сотни и тысячи каналов. При этом далеко не все каналы являются одинаково информативными. Как показали исследования по анализу весовых коэффициентов НС и возможности отсечения слабых связей [17], меры по уменьшению количества параметров модели могут быть эффективными как для уменьшения количества вычислительных операций без потери точности решения, так и для построения новой модели только на отобранных признаках, и в результате для повышения точности ее работы.

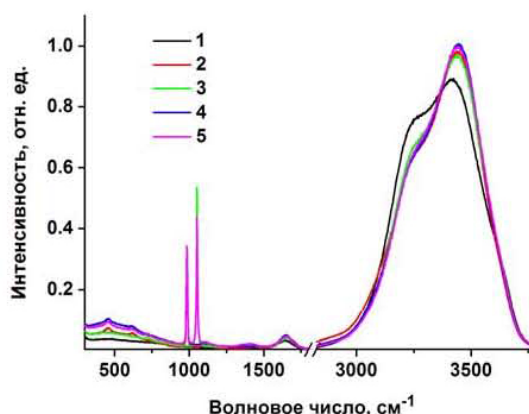


Рис. 2. Спектры многокомпонентных растворов солей и дистиллированной воды: 1 – дистиллированная вода; 2 – 1,85 М NH<sub>4</sub>Br, 0,4 М Li<sub>2</sub>SO<sub>4</sub>; 3 – 0,4 М NaCl, 0,8 М NH<sub>4</sub>Br, 0,55 М KNO<sub>3</sub>; 4 – 0,3 М NaCl, 1 М NH<sub>4</sub>Br, 0,35 М Li<sub>2</sub>SO<sub>4</sub>, 0,45 М CsI; 5 – 0,5 М NaCl, 0,7 М NH<sub>4</sub>Br, 0,4 М Li<sub>2</sub>SO<sub>4</sub>, 0,5 М KNO<sub>3</sub>, 0,3 М CsI

С целью уменьшения входной размерности задачи используют компрессию входных данных [18], в результате чего количество входных признаков уменьшается, остаются наиболее информативные из них. В дальнейшем ИНС и ПЛС модель обучаются на таких сжатых примерах с отобранными признаками. Одним из методов выделения существенных входных признаков является отображение признаков в другое пространство меньшей размерности с использованием некоторого нового базиса (извлечение информативных признаков). Наиболее простой метод такого вида компрессии входных данных – агрегация спектральных каналов. Метод заключается в суммировании интенсивностей в определенном числе соседних спектральных каналов. С практической точки зрения это соответствует уменьшению спектрального разрешения системы регистрации, а из-за усреднения интенсивностей этот метод приводит к снижению шума. Для описанной выше задачи по определению концентраций 5-ти солей агрегация по 8-ми соседним каналам позволила снизить среднюю абсолютную ошибку определения концентраций солей в среднем на 16–18 % [19]. Для более сложной задачи определения 10-ти ионов оптимальной оказалась агрегация всего по 2-м соседним каналам, которая привела к снижению средней абсолютной ошибки на 3,1 % [20].

В данной работе, кроме агрегации, авторы рассматривают возможность использования для извлечения более компактного представления исходных спектров вейвлет-преобразования. Вейвлет-преобразование [21] является широко распространенным спектральным методом анализа сигналов, общая теория которого активно развивалась в 1980-е гг. с введением понятий мультимасштабного анализа [22] и появлением ортогональных вейвлетов с компактным носителем [23]. Благодаря требованию локализованности вейвлет-функции вейвлет-анализ имеет преимущество над классическим спектральным методом – Фурье-анализом – в том, что, помимо извлечения частотной картины, становится возможным также учесть появление и исчезновение той или иной компоненты сигнала во временной области [24].

В данной статье на примере задачи определения концентраций 5-ти неорганических солей в многокомпонентном водном растворе по их спектрам КР была проведена компрессия входных данных путем извлечения информативных признаков с помощью агрегации по соседним каналам, а также дискретного и непрерывного вейвлет-преобразования. Представлены результаты сравнительного анализа использования указанных методов отбора существенных входных признаков для решения поставленной обратной задачи лазерной спектроскопии КР.

### Материалы и методы

Объектами исследований являлись водные растворы солей, содержание которых существенно в природных водах, – NaCl, NH<sub>4</sub>Br, Li<sub>2</sub>SO<sub>4</sub>, KNO<sub>3</sub>, CsI. Концентрация каждой соли в растворах изменялась в диапазоне от 0 до 2,5 М с шагом по концентрациям 0,2–0,25 М. Диапазон концентраций выбран в соответствии со следующими соображениями: содержание солей в морской воде, соответствующее солености 35 ‰, – 0,5 М. В минеральных водах концентрация отдельных солей может составлять до 1 М, в сточных и сбросовых – до 1–2,5 М. Для приготовления растворов использовались бидистиллированная вода и реактивы марки ЧДА (чистый, для анализа).

Возбуждение спектров КР растворов осуществлялось излучением аргонового лазера с длиной волны 488 нм и мощностью около

350 мВт. Для подавления рассеяния на несмещенной частоте использовался интерференционный фильтр (Semrock), позволяющий приблизиться до 200 см<sup>-1</sup> к линии лазерного возбуждения. Регистрация спектров проводилась CCD-камерой (Synapse 1024\*128 BIUV фирмы «Jobin Yvon») после прохождения сигнала через монохроматор (Acton, решетка 900 штр/мм, фокусное расстояние 500 мм) по очереди в двух спектральных диапазонах: 300...1800 и 2600...3800 см<sup>-1</sup> для каждой пробы. Практическое разрешение КР-спектрометра при этом составляло 2 см<sup>-1</sup>, время накопления одного спектра составляло 1 с. Температура проб во время эксперимента поддерживалась равной 21,0 ± 0,2 °С. Спектры нормировались на мощность лазерного излучения и время накопления спектра. Дальнейшая математическая обработка для полос спектров КР заключалась в вычитании пьедестала, обусловленного рассеянием света в кювете с образцом, и нормировке спектра в обоих спектральных диапазонах на площадь валентной полосы КР воды. Отношение сигнал/шум (по валентной полосе КР воды) изменялось в диапазоне от 200 (для многокомпонентных растворов с большими концентрациями) до 700 (для растворов с малыми концентрациями и дистиллированной воды).

Были приготовлены водные растворы всевозможных сочетаний указанных солей – одно-, двух-, трех-, четырех- и пятикомпонентные – в исследуемом диапазоне концентраций 0–2,5 М. Регистрировались спектры КР каждого раствора. Некоторые спектры КР многокомпонентных водных растворов, содержащих ионы неорганических солей NaCl, NH<sub>4</sub>Br, Li<sub>2</sub>SO<sub>4</sub>, KNO<sub>3</sub>, CsI, представлены на рис. 1, 2. Всего было получено 8695 спектров.

Для работы с нейронными сетями и применения метода ПЛС полученный массив данных был разбит на тренировочный, валидационный и экзаменационный наборы в соотношении 70:20:10. Тренировочный набор использовался для подстройки весов в процессе обучения, валидационный – для определения момента прекращения обучения по минимуму ошибки на этом наборе, экзаменационный – для проверки работы полученной сети на независимых данных. Формирование ПЛС-модели останавливалось при достижении сходимости на трени-

ровочном наборе. Для независимой оценки все представленные ниже результаты приведены для экзаменационного набора.

### Компрессия входных данных

Используемые в настоящей работе дискретное и непрерывное вейвлет-преобразование (ДВП и НВП) осуществляют так называемый мультимасштабный анализ. Исходный сигнал (в данном случае спектр) рассматривается как некоторое масштабное пространство с наилучшим разрешением, а для некоторого заданного базиса ортогональных функций находится множество подпространств с меньшей детализацией. И. Добеши в начале 1990-х гг. ввела класс таких функций – вейвлетов (впоследствии названных ее именем), и продемонстрировала возможность осуществления такого разложения с их помощью [25]. В частности, для ДВП принято выбирать разложение по базисным функциям, масштаб которых отличается в  $2, 4, \dots, 2^N$  раз. Функции разного масштаба соответствуют разным уровням разложения. Для каждого уровня разложения результатом преобразования являются наборы из  $N/2$  детальных коэффициентов, описывающих изменчивость сигнала на данном масштабном уровне, и  $N/2$  аппроксимационных коэффициентов, описывающих интегральный характер поведения сигнала на данном масштабном уровне. Последние используются для перехода к следующему уровню детализации вдвое меньшего масштаба.

В качестве сжатого представления сигнала можно взять набор коэффициентов с некоторого заданного уровня ДВП. Совокупность аппроксимационных коэффициентов визуально практически не отличается от исходного сигнала, однако имеет меньшее разрешение – меньшую размерность, соответствующую своему уровню. Чтобы учесть высокочастотную составляющую преобразованного сигнала, можно использовать детальные коэффициенты. Если же рассматривать задачу не понижения размерности, а преобразования исходных входных признаков в другое пространство без потери информации, то можно использовать собранные в единый массив детальные коэффициенты всех уровней, получив таким образом преимущество подхода мультимасштабного анализа и сохранив при этом всю имею-

щуюся информацию. Общее количество входных признаков при этом не изменится.

Следует отметить, что ДВП является лишь частной реализацией более общей практики вейвлет-анализа. В общем случае можно построить полный вейвлет-спектр, т. е. результат осуществления исходного сигнала в каждой точке его реализации (с параметром сдвига, равным значению шага дискретного представления сигнала) на всех возможных масштабах (путем построения сетки с исчезающе малым шагом различных значений параметра масштаба вейвлета). Однако очевидно, что подобное представление нереализуемо на практике и избыточно, поскольку вейвлеты обладают ограниченной неопределенностью в частотно-временной области. Учитывая последнее обстоятельство, можно выбрать такой шаг сетки параметров сдвига и масштаба, при котором будет получен вейвлет-спектр практически без потери информации относительно избыточного представления. Классическим подходом в этом случае является линейная зависимость масштаба и сдвига: при меньшей ширине вейвлет-функции используется меньшее значение шага. Для простоты выбора параметров этой связи и гарантирования одинакового поведения вейвлет-спектра на разных масштабах в данной работе будем использовать понятие степени перекрытия вейвлет-окон, а параметр масштаба будет подбираться автоматически. В работе мы использовали реальную часть вейвлета Морле со степенью перекрытия окон 50 %.

Вычисления ДВП проводились на языке R с использованием библиотеки *wavethresh: Wavelet Statistics and Transforms*<sup>1</sup>. Использовались вейвлеты семейства Добеши 10. Данные исходных спектров, описанные интенсивностями сигнала в 1 535 каналах спектра, дополнялись нулями до 2 048 значений. НВП вычислялось с помощью собственной реализации кода на языке Python, поддерживающем параллельные вычисления на GPU за счет использования функций библиотеки *tensorflow*<sup>2</sup>. Вычислительные эксперименты с обучением НС проводились

<sup>1</sup> Библиотека *Wavelet Statistics and Transforms*: <https://cran.r-project.org/package=wavethresh/>.

<sup>2</sup> Библиотека для эффективных математических операций с большими массивами данных *tensorflow*: <https://www.tensorflow.org/>.

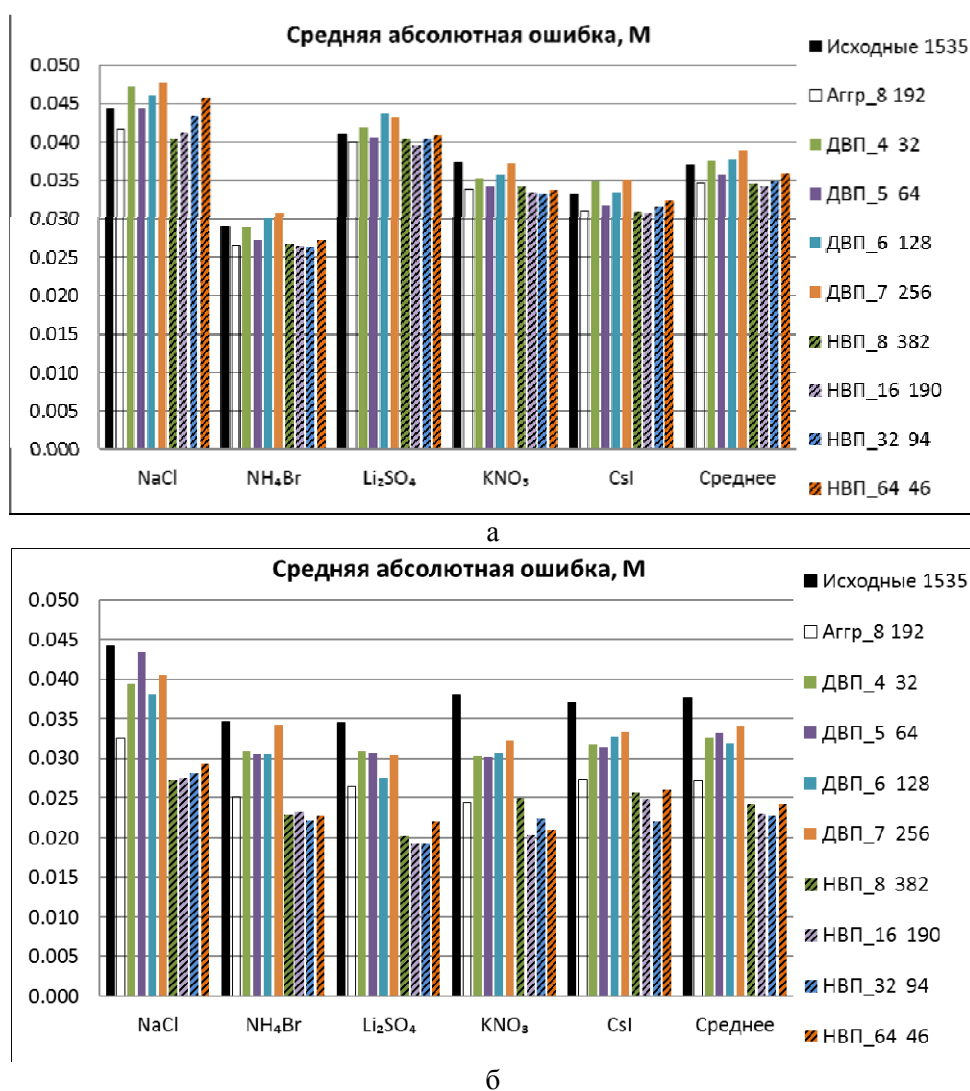


Рис. 3. Сравнение результатов применения метода ПЛС (а) и ИНС (б) к данным с различной компрессией входных признаков: средняя абсолютная ошибка на экзаменационном наборе данных. Методы компрессии в легенде: Аггр – агрегация, ДВП – дискретное вейвлет-преобразование, НВП – непрерывное вейвлет-преобразование, через пробел указано количество входных признаков

средствами языка Python на базе библиотеки машинного обучения scikit-learn<sup>3</sup> и tensorflow.

В предыдущих работах авторов [26] было установлено, что перцептроны с 1 и 3 скрытыми слоями показывают результаты хуже, чем перцептроны с 2 скрытыми слоями. Во всех случаях среди рассмотренных архитектур наилучшие результаты показал перцептрон со 120 нейронами в первом скрытом слое и 60 во втором. Поэтому в настоящей работе использовалась именно эта архитектура ИНС. Для исключения зависимости

результатов от начальной инициализации весов каждая нейронная сеть тренировалась по 5 раз с разными начальными весами; результаты применения всех 5 сетей усреднялись. Везде ниже приведены именно такие усредненные результаты – значения среднеквадратичной ошибки (СКО) на экзаменационных наборах данных.

## Результаты

### Применение метода проекций на латентные структуры

Рассмотрим результаты работы метода проекций на латентные структуры (ПЛС) в применении к исходным входным данным и

<sup>3</sup> Библиотека машинного обучения scikit-learn: <http://scikit-learn.org/stable/index.html>

данным, обработанным различными методами компрессии. В качестве компрессии входных данных использовались агрегация по 8-ми входным признакам, ДВП для 4, 5, 6 и 7 уровней и НВП с шириной свертки по 8, 16, 32 и 64 каналам.

На рис. 3, а представлены результаты применения различных методов компрессии в сравнении с исходными данными, при использовании сжатых данных в качестве входных для метода ПЛС. Как можно видеть, при использовании некоторых методов компрессии ПЛС дает лучший результат, чем без использования компрессии. В случае ДВП наилучший результат достигается при использовании 5 уровня, что соответствует 32 аппроксимационным и 32 детальным коэффициентам. Увеличение, как и уменьшение числа используемых коэффициентов, приводит к ухудшению результатов. Использование агрегации по 8 признакам показывает результат лучший, чем ДВП. Наилучший результат достигается при использовании НВП с окном в 16 значений, что соответствует 190 входным признакам. При увеличении и уменьшении окна ошибка монотонно растет, что подтверждает предположение об оптимальном числе входных признаков. В среднем наилучшая точность определения солей – 0,034 М.

#### *Применение искусственных нейронных сетей*

Рассмотрим результаты использования ИНС на аналогичных исходных данных.

На рис. 3, б представлены аналогичные результаты для многослойного персептрона. Как видно из графика, результаты применения нейронных сетей без компрессии входных данных показывают большую ошибку, чем в результате применения ДВП. При этом ДВП дает большую ошибку, чем агрегация, а наилучший результат получен при использовании НВП. Наименьшая ошибка (в среднем 0,023 М) достигается при использовании НВП с окном в 32 значения, что соответствует 94 входным признакам.

Таким образом, из трех рассмотренных методов извлечения информативных признаков наиболее эффективным оказалось непрерывное вейвлет-преобразование, обеспечивающее наименьшие значения средней

абсолютной погрешности на экзаменационном наборе данных при использовании как многослойного персептрона, так и метода ПЛС. При этом ИНС показывает существенно лучшие результаты, чем ПЛС, что свидетельствует о значительной нелинейности решаемой задачи.

#### **Заключение**

Спектроскопия комбинационного рассеяния света в сочетании с применением искусственных нейронных сетей, обученных на преобразованных входных признаках, продемонстрировала высокую эффективность при решении задачи идентификации и определения концентраций 5 неорганических солей, растворенных в воде.

В данной работе рассмотрены три метода преобразования входных признаков: агрегация смежных спектральных каналов, дискретное и непрерывное вейвлет-преобразование. Показано, что все они позволяют понизить как размерность входных данных, так и погрешность определения концентраций солей в рассматриваемой задаче. Наиболее эффективным методом для рассмотренной задачи оказалось непрерывное вейвлет-преобразование. Обученный на преобразованных этим методом входных признаках многослойный персептрон обеспечил среднюю по всем 5 солям погрешность определения концентрации 0,023 М, что на 38 % меньше погрешности, получаемой ИНС без компрессии данных.

#### **Список литературы**

1. Шевцов М. Н. Водно-экологические проблемы и использование водных ресурсов. Хабаровск: Изд-во Тихоокеан. гос. ун-та, 2015. 197 с. ISBN 978-5-7389-1817-9.
2. Andersson M., Edner H., Johansson J., Svanberg S., Wallinder E., Weibring P. Remote sensing of the environment using laser radar techniques // Atomic Physics Methods in Modern Research. 1997. Vol. 499. P. 257–269. Springer. DOI: 10.1007/BFb0104329.
3. Harsdorf S., Janssen M., Reuter R., Toeneboen S., Wachowicz B., Willkomm R. Submarine lidar for seafloor inspection // Meas Sci Tech. 1999. Vol. 10. No. 12. P. 1178–1184.
4. Dolenko T. A., Churina I. V., Fadeev V. V., Glushkov S. M. Valence band of liquid water



Raman scattering: some peculiarities and applications in the diagnostics of water media // *J. Raman Spectroscopy*. 2000. Vol. 31. P. 863–870.

5. *Kauffmann T. H., Fontana M. D.* Inorganic salts diluted in water probed by Raman spectrometry: Data processing and performance evaluation // *Sensors and Actuators B*. 2015. Vol. 209. P. 154–161.

6. *Rudolph W. W., Irmer G.* Raman and Infrared Spectroscopic Investigation on Aqueous Alkali Metal Phosphate Solutions and Density Functional Theory Calculations of Phosphate-Water Clusters // *Appl. Spectroscopy*. 2007. Vol. 61. No. 12. P. 274A–292A.

7. *Rull F., De Saja J. A.* Effect of electrolyte concentration on the Raman spectra of water in aqueous solutions // *J. Raman Spectroscopy*. 1986. Vol. 17. No. 2. P. 167–172.

8. *Sadate S., Kassu A., Farley C. W., Sharma A., Hardisty J., Lifson Miles T. K.* Standoff Raman measurement of nitrates in water // *Proc. SPIE Remote Sensing and Modeling of Ecosystems for Sustainability VIII*. 2011. P. 81560D-1-6.

9. *Somekawa T., Tani A., Fujita M.* Remote Detection and Identification of CO<sub>2</sub> Dissolved in Water Using a Raman Lidar System // *Applied Physics Express*. 2011. Vol. 4. P. 112401-1-3.

10. *Mernagh T. P., Wilde A. R.* The use of the laser Raman microprobe for the determination of salinity in fluid inclusions // *Geochimica et Cosmochimica Acta*. 1989. Vol. 53. P. 765–771.

11. *Burikov S. A., Dolenko S. A., Dolenko T. A., Persiantsev I. G.* Application of Artificial Neural Networks to Solve Problems of Identification and Determination of Concentration of Salts in Multi-Component Water Solutions by Raman Spectra // *Optical Memory and Neural Networks (Information Optics)*. 2010. Vol. 19. No. 2. P. 140–148.

12. *Efitorov A., Burikov S., Dolenko T., Laptinskiy K., Dolenko S.* Significant Feature Selection in Neural Network Solution of an Inverse Problem in Spectroscopy // *Procedia Computer Science*. 2015. Vol. 66. P. 93–102.

13. *Haykin S. S.* Neural networks and learning machines. 3<sup>rd</sup> ed. Upper Saddle River. NJ. USA: Pearson, 2009.

14. *Esbensen K. H.* Multivariate Data Analysis // *Practice, an Introduction to Multivariate Data Analysis and Experimental Design*. 5<sup>th</sup> ed., CAMO Software AS. 2006, 599 p.

15. *Rumondor A. C., Taylor L. S.* Application of partial least-squares (PLS) modeling in quantifying drug crystallinity in amorphous solid dispersions // *Int. J. Pharm.* 2010. Vol. 398. Is. 1–2. P. 155–160.

16. *Gushchin K. A., Burikov S. A., Dolenko T. A., Persiantsev I. G., Dolenko S. A.* Data dimensionality reduction and evaluation of clusterization quality in the problems of analysis of composition of multi-component solutions // *Optical Memory and Neural Networks*. 2015. Vol. 24. Is. 3. P. 218–224.

17. *Kamruzzaman S. M., Ahmed Ryadh Hasan.* Pattern Classification using Simplified Neural Networks. arXiv:1009.4983v1 [cs.NE]

18. *Peng H., Ding C., Long F.* Minimum redundancy maximum relevance feature selection // *IEEE Intelligent Systems*. 2005. Vol. 20. No. 6. P. 70–71

19. *Dolenko S., Burikov S., Dolenko T., Efitorov A., Gushchin K., Persiantsev I.* Neural Network Approaches to Solution of the Inverse Problem of Identification and Determination of Partial Concentrations of Salts in Multi-component Water Solutions // *Lecture Notes in Computer Science*. 2014. Vol. 8681. P. 805–812. DOI 10.1007/978-3-319-11179-7\_101.

20. *Efitorov A., Dolenko T., Burikov S., Laptinskiy K., Dolenko S.* Solution of an Inverse Problem in Raman Spectroscopy of Multi-component Solutions of Inorganic Salts by Artificial Neural Networks // *Lecture Notes in Computer Science*. 2016. Vol. 9887. P. 355–362. DOI 10.1007/978-3-319-44781-0\_42.

21. *Strang G., Nguyen T.* Wavelets and filter banks. 2<sup>nd</sup> ed. Wellesley-Cambridge Press, 1996. 520 p.

22. *Mallat S. G.* A theory for multiresolution signal decomposition: the wavelet representation // *IEEE Transactions on Pattern Recognition and Machine Intelligence*. 1989. Vol. 11. No. 7. P. 674–693.

23. *Daubechies I.* Orthonormal bases of compactly supported wavelets // *Comm. Pure & Appl. Math.* 1988. Vol. 41. No. 7. P. 909–996.

24. *Jean-Philippe Lachaux, Antoine Lutz, David Rudrauf, Diego Cosmelli, Michel Le Van Quyen, Jacques Martinerie, Francisco Varela.* Estimating the time-course of coherence between single-trial brain signals: an introduction to wavelet coherence // *Neurophysiol Clin.* Elsevier. 2002. Vol. 32. P. 157–174.

25. *Daubechies I.* Ten Lectures on Wavelets // *SIAM*. 1992. 356 p.



26. Efitorov A., Dolenko T., Burikov S., Laptinskiy K., Dolenko S. Neural Network Solution of an Inverse Problem in Raman Spectroscopy of Multi-component Solutions of In-

organic Salts // *Advances in Intelligent Systems and Computing*. 2016. Vol. 449. P. 273–279.

*Материал поступил в редколлегию 09.07.2018*

**S. A. Burikov, A. O. Efitorov, T. A. Dolenko, V. R. Shirokiy, S. A. Dolenko**

*Lomonosov Moscow State University  
1, build. 2, Leninskie Gory, Moscow, GSP-1, 119991, Russian Federation*

*sergey.burikov@gmail.com*

**SOLUTION OF INVERSE PROBLEMS OF RAMAN SPECTROSCOPY  
OF AQUEOUS SALT SOLUTIONS  
WITH THE APPLICATION OF WAVELET NEURAL NETWORKS**

This paper presents the results of solving the problem of determining the salt composition of multicomponent aqueous solutions by their Raman spectra using artificial neural networks and the method of projections on latent structures. Three methods of input feature transformation are considered: aggregation of adjacent spectral channels, discrete and continuous wavelet transforms. It is shown that all of them can reduce both the dimension of the input data and the error of determination of the salt concentrations in the problem under consideration. The most effective method for the solution of the considered problem was a continuous wavelet transform. The multilayer perceptron trained on the transformed by this method input features provided the average error of determination of concentration for all 5 salts 0.023 M that is 38 % less than the error obtained by artificial neural network used without data compression. Thus, Raman spectroscopy combined with the use of artificial neural networks trained on transformed input features demonstrated high efficiency in solving the problem of identification and determination of concentrations of 5 inorganic salts dissolved in water.

*Keywords:* Raman spectroscopy, aqueous solutions of salts, projections on latent structures, artificial neural networks, wavelet analysis.

*For citation:*

Burikov S. A., Efitorov A. O., Dolenko T. A., Shirokiy V. R., Dolenko S. A. Solution of Inverse Problems of Raman Spectroscopy of Aqueous Salt Solutions with the Application of Wavelet Neural Networks. *Siberian Journal of Physics*, 2018, vol. 13, no. 3, p. 101–109. (in Russ.)

DOI 10.25205/2541-9447-2018-13-3-101-109