

# Application of Adaptive Neural Network Algorithms for Determination of Salt Composition of Waters Using Laser Spectroscopy

Tatiana A. Dolenko, Sergey A. Burikov, Alexander O. Efitorov, Sergey A. Dolenko

**Abstract**—In this study, a comparative analysis of the approaches associated with the use of neural network algorithms for effective solution of a complex inverse problem – the problem of identifying and determining the individual concentrations of inorganic salts in multicomponent aqueous solutions by the spectra of Raman scattering of light – is performed. It is shown that application of artificial neural networks provides the average accuracy of determination of concentration of each salt no worse than 0.025 M. The results of comparative analysis of input data compression methods are presented. It is demonstrated that use of uniform aggregation of input features allows decreasing the error of determination of individual concentrations of components by 16-18% on the average.

**Keywords**—Inverse problems, multi-component solutions, neural networks, Raman spectroscopy.

## I. INTRODUCTION

THE Artificial Neural Networks (ANN) demonstrate very high efficiency for solution of various problems of approximation, prediction, evaluation, classification, pattern recognition etc. ANN are used also for solution of inverse problems (IP), where their properties, such as training by examples, high noise stability, stability to contradictory data etc., play a special role [1], [2]. Most often, IP of regression type are under consideration – those with continuous output (for example, determination of temperature of plasma by its proper glow spectrum) [3], [4]. However, more complicated situations can take place, when simultaneous solution of problems of classification or pattern recognition, i.e. IP with discrete output, and traditional IP of regression type for every found component, is required. We call this type of IP the complex inverse problem. In this study, the elaborated method of solution of such complex IP is presented at the example of complex IP of identification and determination of individual concentrations of salts in multi-component water solution by

T.A.Dolenko and S.A.Burikov are with Physical Department, M.V.Lomonosov Moscow State University, Moscow, 119991, Russian Federation (e-mail tdolenko@lid.phys.msu.ru, burikov@lid.phys.msu.ru).

A.O.Efitorov is with Physical Department and with D.V.Skobeltsyn Institute of Nuclear Physics, M.V.Lomonosov Moscow State University, Moscow, 119991, Russian Federation (e-mail sasha.efitorov@yandex.ru).

S.A.Dolenko is with D.V.Skobeltsyn Institute of Nuclear Physics, M.V.Lomonosov Moscow State University, Moscow, 119991, Russian Federation (e-mail dolenko@srd.sinp.msu.ru).

This study has been supported by the grant of the Russian Foundation for Basic Research (project no.12-01-00958-a). The study has been conducted in D.V.Skobeltsyn Institute of Nuclear Physics, M.V.Lomonosov Moscow State University, Moscow, 119991, Russian Federation.

Raman spectra.

The problem of determination of salts composition of multicomponent solutions is very important for oceanology, ecological monitoring, and control of technical and mineral waters. Obviously, to solve this problem, a remote method which can be implemented in the on-line mode is required.

Such properties are possessed by the non-contact radiometric method of measurement of salinity of the sea surface waters, most common in oceanology [5]-[7]. At present, however, it does not provide the required accuracy of salinity determination. That is mainly due to the difficulties of extraction of relatively small changes of thermal radio emission caused by variations in salinity, against the background of larger signals caused by surface waves. In addition, radiometric methods measure the total concentration of dissolved salts (salinity) only in subsurface water layer and do not allow to solve the problem of identification of substances dissolved in water.

The problem of identification and determination of individual concentrations of salts in solution can be solved using spectrophotometry, measuring absorption spectra at several wavelengths [8]. However, to identify salts, it is necessary that the absorption coefficients of the investigated substances differ significantly. In addition, this method is unsuitable for remote sensing.

It is very promising to use Raman spectra for express remote determination of concentrations of inorganic substances dissolved in water. Principle opportunity of using Raman spectra for diagnostics of solutions results from high sensitivity of their characteristics to types and concentrations of salts dissolved in water [9]-[14].

In [9], [10], to identify and determine the concentration of salts with complex ions, such as  $\text{NO}_3^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{PO}_4^{3-}$ ,  $\text{CO}_3^{2-}$  anions, it was suggested to use Raman spectra of these ions in the area of  $1000 \text{ cm}^{-1}$ . In this area, these anions have narrow bands produced by valence vibrations of these anions (Fig. 1). Anion type can be determined by the position of the corresponding band, its concentration by intensity of that band. However, this method can be used only for diagnostics of salts with complex ions. In natural waters there are much less salts with complex anions than, for example, metal halides.

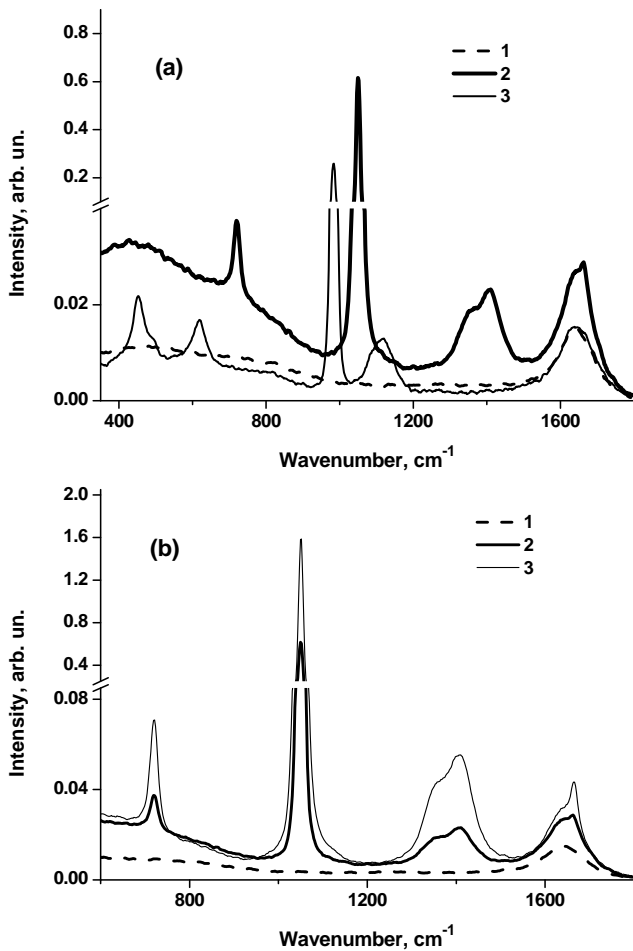


Fig. 1 The low-frequency region of Raman spectra of aqueous solutions of salts with complex anions. (a): 1 – distilled water; 2 –  $\text{KNO}_3$ , 1 M; 3 –  $\text{Li}_2\text{SO}_4$ , 1 M. (b): 1 – distilled water; 2 –  $\text{KNO}_3$ , 1 M; 3 –  $\text{KNO}_3$ , 3 M

As it was found out in [11]-[14], the position and shape of water Raman valence band ( $2700 - 4000 \text{ cm}^{-1}$ , Fig. 2) also significantly depends on the type of salt and its concentration in the solution. This influence is manifested, for example, in the dependence of such spectrum characteristics as the position of the valence band maximum, its width, the ratio of intensities of high-frequency and low-frequency fields of valence band - the parameter  $\chi_{21} = I(\nu_2)/I(\nu_1)$ , on the salt type and its concentration.

The authors of [15]-[17] elaborated a method of determination of concentration of dissolved salts by water Raman valence band. In [15], [16], concentration of the single salt in the solution was determined using the dependence  $\chi_{21}(C)$  of water Raman valence band. In [17], [13], two-parameter inverse problem was solved: the temperature  $T$  and salinity  $S$  of sea water were determined simultaneously using the dependence  $\chi_{21}(T,S)$ .

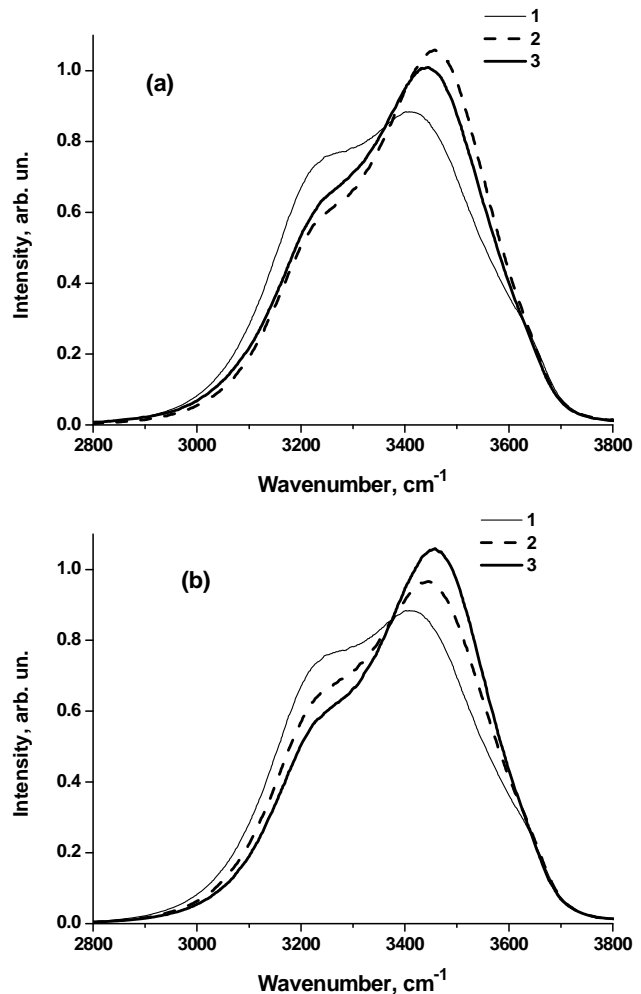


Fig. 2 Valence Raman bands of water. (a): 1 – distilled water; 2 –  $\text{CsI}$ , 2 M; 3 –  $\text{NaCl}$ ,  $\text{NH}_4\text{Br}$ ,  $\text{Li}_2\text{SO}_4 - 0.4 \text{ M}$ ,  $\text{KNO}_3$  и  $\text{CsI} - 0.6 \text{ M}$ . (b): 1 – distilled water; 2 –  $\text{CsI}$ , 1 M; 3 –  $\text{CsI}$ , 2 M.

However, the increase of IP dimension, for example, the attempt to determine the concentrations of three salts dissolved in water, reveals that the suggested methods are not sufficiently robust and fail to achieve acceptable accuracy of measurement of concentrations of three salts. It became obvious that it was necessary to pass from the analysis of individual spectral parameters to the analysis of the whole shape of a spectrum, for example, with the help of ANN. The use of ANN allowed the authors of this work to develop a method for solution of IP of identification and determination of individual concentrations of salts by Raman water valence band in three-component solutions [18], [19].

This paper presents the results of the application of ANN for solving multiparameter IP: identification and determination of the concentration of each of salt in five-component aqueous solutions by their Raman spectra. The increase of number of search parameters became possible due to the fact that ANN performs the analysis of valence bands and low-frequency bands of Raman spectra simultaneously.

This improves the sensitivity of the method for determination the concentration of complex ions significantly

Intermediate results of the work were presented in [20], [21]. In the present study, the previously developed techniques are summarized, and the results of the final stage of the development of neural network method for solving multi-parameter IP - identification and determination of the concentration of each salt in five-component aqueous solutions by their Raman spectra – are presented.

## II. METHODOLOGICAL ASPECTS OF USING ANN

### A. ANN Training

Training of ANN for solution of IP requires a representative dataset, i.e. a set of data that would reflect all characteristic aspects of behavior of the object. This dataset can be obtained in different ways [22]. If an adequate analytical model of solution of the direct problem is available, it can be used for generation of arrays of data with necessary representativity. However, the Raman spectra of aqueous solutions cannot be described analytically. Therefore, to solve the pointed IP, "experiment-based" and quasi-model approaches were used.

#### 1) "Experiment-Based" Approach

The data used for ANN training are obtained in experiment. (In this paper, the data are Raman spectra of various solutions with different combinations and concentrations of components, obtained using laser Raman spectrometer). This approach does not require a model to be available, and it inherently takes the non-linear properties of the object into account. However, obtaining a representative data set can be a non-trivial experimental problem.

#### 2) "Quasi-Model" Approach

If no adequate physically grounded analytical model of solution of the direct problem is available, one can replace it by a parametrical quasi-model based on experimental data. This model formally describes the dependence of the observed data on the sought-for parameters. In the present case, a quasi-model describes the dependence of intensity in every spectral channel on concentrations of components. The simplest quasi-model is a linear one when it is supposed that the intensity in each spectral channel is a linear combination of concentrations of the components. To construct more complicated quasi-models, which can describe the desired dependence better, one can use more efficient adaptive methods of model construction.

### B. Used Architectures of ANN

The considered complex IP naturally breaks down into two problems – the problem of identification/classification of components of the solution, and the problem of determination of their individual concentrations. These problems may be solved separately [20]. The method applied in this study provides simultaneous solution of these two problems, when a single ANN is used. The values at the outputs of the ANN are treated as estimated concentrations of components. If the value at some output is lower than a pre-defined threshold, it is

considered that the corresponding component is not present in the solution.

The ANN architectures used in this study to solve the complex IP were perceptrons with one and three hidden layers, trained by error back-propagation method, General Regression NN (GRNN) [23] and Group Method of Data Handling (GMDH) [24]. To construct the quasi-models, perceptrons and GRNN were used.

### C. Criteria of Model Quality Evaluation

To evaluate the quality of models, four basic criteria were used in this study. The following designations will be used:  $y$  – estimation of the value of an output variable made by the model (neural network);  $d$  – the desired value of this output variable;  $\bar{d}$  – average of the desired output value over the whole concerned dataset;  $N$  – the number of patterns in the concerned dataset. Summation is carried out over all the patterns of the dataset for which the criterion is calculated (from 1 to  $N$ ).

1. *Coefficient of Multiple Determination*:  $R^2$  is calculated according to the following formula:

$$R^2 = 1 - \frac{\sum (d - y)^2}{\sum (d - \bar{d})^2}$$

This criterion compares the error of the constructed model with the error of trivial reference model (the estimate provided by this reference model is the mean value of the estimated variable over all the patterns of the dataset). When the estimate is absolutely accurate,  $R^2$  equals 1. If the accuracy of the estimate is worse than the accuracy of the trivial model,  $R^2$  is negative. For many kinds of problems,  $R^2$  is the most substantial universal criterion for evaluation of model quality. This criterion is dimensionless.

2. *Root Mean Squared Error (RMSE)* is calculated according to the following formula:

$$RMSE = \sqrt{\frac{\sum (d - y)^2}{N}}$$

This criterion has the same dimension as the estimated variable. In our case (determination of concentration), it is M (molarity – quantity of moles of substance per liter of solution).

3. *Mean Absolute Error (MAE)*, calculated according to:

$$MAE = \frac{\sum |d - y|}{N}$$

This criterion also has the same dimension as the estimated variable.

4. *Mean Relative Error (MRE)*, calculated according to:

$$MRE = \frac{1}{N} \sum \frac{|d - y|}{d} \cdot 100\%$$

This criterion has no dimension and it is usually expressed in percent.

### III. EXPERIMENT

#### A. Raman Spectrometer

All the Raman spectra of aqueous solutions were obtained by the Raman spectrometer [20]. Excitation of Raman spectra was performed by argon laser with wavelength 488 nm and output power 350 mW. In order to remove elastic scattering signal, edge-filter was used. It allowed approaching laser line to  $100\text{ cm}^{-1}$ . Registration of spectra was performed by double monochromator (lattice 900 gr/mm, focal length 500 mm) and by CCD-camera. Spectra were measured in two spectral bands:  $200\text{--}2300\text{ cm}^{-1}$  and  $2300\text{--}4000\text{ cm}^{-1}$  for each sample. Practical resolution of the Raman spectrometer was  $2\text{ cm}^{-1}$ . The temperature of the sample during the experiment was stabilized at  $22.0\pm 0.2^\circ\text{C}$ . Spectra were normalized to laser power, spectral sensitivity of the detector and the time of spectrum accumulation.

#### B. Preparation of Solutions

The objects of research were aqueous solutions of the salts NaCl,  $\text{NH}_4\text{Br}$ ,  $\text{Li}_2\text{SO}_4$ ,  $\text{KNO}_3$ , CsI. These salts are present in natural waters at significant concentrations. Concentration of every salt in the solutions changed from 0 to 2.5 M with increment 0.2-0.25 M. The range of concentrations was chosen according to the following considerations. Normal salinity of ocean water 35‰ corresponds to concentration 0.5 M of the most widespread salt NaCl. In mineral waters, concentration of certain salts can reach 1 M, in waste waters – up to 1-2 M.

To prepare solutions, bidistilled water and analytically pure reagents were used.

#### C. Obtaining Experimental Datasets for ANN Training

As it was pointed above, each of the two bands of Raman spectrum was recorded in 1024 spectral channels, in the wavenumber range  $200\text{--}2300\text{ cm}^{-1}$  for the low-frequency band and  $2300\text{--}4000\text{ cm}^{-1}$  for the valence band. For further processing, more narrow informative ranges were extracted: 766 channels in the range  $281\text{--}1831\text{ cm}^{-1}$  for the low-frequency band (Fig. 1) and 769 channels in the range  $2700\text{--}3900\text{ cm}^{-1}$  for the valence band (Fig. 2).

For each of the bands separately, the pedestal produced by elastic scattering in the cuvette with sample, was subtracted. Then the spectra were normalized to the area of the valence band in the pointed region.

The obtained data array (1535 features, 9144 samples) was randomly divided into training set (for ANN training), test set (used to prevent overtraining), and examination set (for out-of-sample testing) in the ratio of 70:20:10. All the results presented below are for the examination set. The results on all the three sets were close, confirming sufficient data representativity.

### IV. RESULTS AND DISCUSSION

#### A. Results of IP Solution, "Experiment-Based" Approach

The "experiment-based" approach to ANN solving of the IP was as follows. The inputs of ANN were fed with all the 1535

features chosen according to the routine described above. Each feature separately was normalized into 0...1 range over the whole data array.

The ANN had five outputs according to the maximum number of components in a solution. Each output corresponded to one of the considered salts, and its desired value corresponded to the concentration of this salt in the solution.

Three models were used to solve the considered IP: perceptrons with three hidden layers, GRNN and GMDH. The results (MAE on examination set, M) are presented in Table I.

TABLE I  
COMPARISON OF THE RESULTS OF DETERMINATION OF SALTS CONCENTRATIONS BY RAMAN SPECTRA IN MULTI-COMPONENT SOLUTIONS (MAE ON EXAMINATION SET, M)

Salt	Perceptron	GRNN	GMDH
NaCl	0.029	0.102	0.059
$\text{NH}_4\text{Br}$	0.024	0.047	0.046
$\text{Li}_2\text{SO}_4$	0.020	0.064	0.032
$\text{KNO}_3$	0.019	0.058	0.033
CsI	0.023	0.050	0.046

As can be seen from Table I, the best results were demonstrated by perceptron with three hidden layers. The hidden layers contained 40, 20 and 10 neurons. Linear activation function was used in the output layer, and logistic activation function was used in the hidden layers. The following values of training parameters were used: learning rate – 0.01; moment – 0.5; stop training criterion – 1000 epochs after minimum of error on test dataset. Statistical results of determination of salts concentrations with help of perceptron with three hidden layers are presented in Table II.

TABLE II  
STATISTICAL RESULTS OF DETERMINATION OF SALTS CONCENTRATIONS IN MULTI-COMPONENT SOLUTIONS BY RAMAN SPECTRA USING PERCEPTRON WITH THREE HIDDEN LAYERS (ON EXAMINATION SET)

Salt	$R^2$	RMSE, M	MAE, M	MRE, %
NaCl	0.990	0.041	0.029	9.2
$\text{NH}_4\text{Br}$	0.993	0.036	0.024	5.8
$\text{Li}_2\text{SO}_4$	0.996	0.029	0.020	5.6
$\text{KNO}_3$	0.996	0.031	0.019	4.3
CsI	0.994	0.036	0.023	5.9

#### B. Results of IP Solution, "Quasi-Model" Approach

One of the main reasons causing decrease of accuracy of solution of this IP using the "experiment-based" approach is the unfavorable ratio of the number of input variables of the problem (1535) and the number of patterns in the training set (8229 in training and test sets together). One can overcome it in one of two ways: by further reduction of the number of input variables (this way will be considered in this study further) or by increasing the number of patterns. When additional experiments cannot be conducted, one can try to get additional patterns by spectra interpolation using quasi-models – parametrical or adaptive models of solution of the direct problem (see II A). In our case, this is the model of dependence of intensity in each channel of the spectrum on concentrations of the components in the solution. However, if

a quasi-model is not adequate enough, using such quasi-model approach can lead not to improvement but to degradation of the quality of solution of the IP. In particular, for this reason, a linear quasi-model is unsuitable for modeling of multi-component mixtures, in which non-linear interaction of components takes place [18].

In this study, two types of quasi-models were considered: based on perceptron and based on GRNN. To evaluate the applicability of the quasi-model approach for solution of this IP, it was necessary first to choose the best of the constructed quasi-models (i.e. the quasi-model providing the smallest error of solution of the direct problem on the examination set).

In Table III, statistics, their minimal, maximal and average values calculated over all 1535 modeled channels are presented for quasi-models based on perceptron and on GRNN. From Table III, one can see that average statistics are close, so at this stage there was no reason to prefer either of these two quasi-models.

TABLE III  
THE STATISTICS OF CONSTRUCTED QUASI-MODELS BASED ON PERCEPTRON AND ON GRNN (ON THE EXAMINATION SET)

Criterion, model	R <sup>2</sup> , perceptron	R <sup>2</sup> , GRNN	MRE, %, perceptron	MRE, %, GRNN
average	0.933	0.933	9.9	7.1
RMSE	0.048	0.041	27.0	11.6
min	0.681	0.768	0.5	0.4
max	0.983	0.974	272.9	75.0

To create the quasi-model datasets on the base of experimental spectra, the spectra of solutions with concentrations corresponding to a grid with 0.15 M increment for all salts were calculated. Only spectra with total concentration of salts less than 2.5 M were considered. There were 53130 such spectra with different concentrations of salts, i.e. more than 6 times as much as experimental spectra with different concentrations. The obtained array of 53130 quasi-model spectra was randomly divided into training, test, and examination datasets in the 70:20:10 ratio.

After the division of the array of quasi-model spectra, the thus obtained training and test sets were supplemented with training and test datasets from the experimental array. The examination datasets were left separate. Thus, the following datasets were used for the quasi-model approach: training set (43593 patterns), test set (12454 patterns), examination quasi-model set (5313 patterns) and examination experimental set (914 patterns). Two such complete sets have been obtained: one by calculation of model spectra with the quasi-model based on perceptron, the other one – with the quasi model based on GRNN. The thus obtained training sets were used to train identical perceptrons, with architecture and training parameters identical to those used to solve the IP within the "experiment-based" approach.

Table IV displays comparison of the results obtained within the "experiment-based" approach on the examination (experimental examination) dataset, with the results obtained within the quasi-model approach on quasi-model examination dataset and on experimental examination dataset, for two

kinds of quasi-model (based on perceptron and on GRNN).

TABLE IV  
SOME STATISTICS FOR ANN DETERMINATION OF CONCENTRATIONS OF SALTS IN MULTI-COMPONENT SOLUTIONS BY RAMAN SPECTRA WITHIN THE "EXPERIMENT-BASED" AND "QUASI-MODEL" APPROACHES

Approach Dataset	EXP	QMP	QMGR	QMP	QMGR
	EE	EE	EE	QE	QE
R <sup>2</sup>					
NaCl	0.990	0.969	0.980	0.989	0.990
NH <sub>4</sub> Br	0.993	0.985	0.986	0.995	0.992
Li <sub>2</sub> SO <sub>4</sub>	0.996	0.983	0.992	0.990	0.991
KNO <sub>3</sub>	0.996	0.986	0.991	0.991	0.991
CsI	0.994	0.987	0.989	0.996	0.992
MAE, M					
NaCl	0.029	0.050	0.042	0.030	0.029
NH <sub>4</sub> Br	0.024	0.037	0.034	0.021	0.025
Li <sub>2</sub> SO <sub>4</sub>	0.020	0.041	0.029	0.027	0.027
KNO <sub>3</sub>	0.019	0.035	0.029	0.028	0.027
CsI	0.023	0.037	0.031	0.019	0.025
MRE, %					
NaCl	9.2	17.4	12.5	10.4	11.3
NH <sub>4</sub> Br	5.8	9.3	8.8	7.2	9.5
Li <sub>2</sub> SO <sub>4</sub>	5.6	10.4	7.7	8.6	9.5
KNO <sub>3</sub>	4.3	8.2	6.8	8.9	9.5
CsI	5.9	9.3	7.5	6.2	9.1

Notation for the datasets: EE – experimental examination, QE – quasi-model examination. Notation for the approaches: Exp – "experiment-based" approach, QMP and QMGR – quasi-model approach with quasi-model based on perceptron and GRNN, respectively.

The presented results allow making the following conclusions.

1. The quasi-model based on GRNN provided the results on the experimental examination dataset that outperformed the results provided by the quasi-model based on perceptron.
2. The results obtained within the quasi-model approach on the quasi-model examination dataset, are comparable to those obtained within the "experiment-based" approach on the experimental examination set, sometimes outperforming them. This is the manifestation of higher representativity of datasets in the quasi-model approach.
3. The quasi-model approach failed to meet expectations compared to the "experiment-based" approach. In all cases, the results on the experimental examination dataset, obtained within the quasi-model approach, turned out to be worse or significantly worse than the results obtained within the "experiment-based" approach. This is an evidence of low adequateness of the quasi-models used.

#### C.Methods of Input Data Compression, "Experiment-Based" Approach

To increase the accuracy of solving IP (that is the accuracy of determination of individual concentrations of salts) by reducing the dimension of space of input features (see IV (B)), several methods of input data compression were used:

- Selection by the absolute value of standard deviation (StD) - selection of input features based on the value of StD of intensity in each channel, as StD is proportional to the entropy and therefore to the volume of information

that each channel brings. The 704 features selected corresponded to channels with StD exceeding some threshold value.

- Selection by GMDH [24]. The subsets of the whole set of features, which are the most significant ones for determination of concentration of each salt, were selected with the help of Group Method of Data Handling. Union of such subsets for the 5 salts gave 314 features.
- Selection by the values of cross-correlation (CC) and cross-entropy (CE). The input features for which CC exceeded the sum of average CC and its StD were selected. This procedure was repeated for each output and for CE instead of CC. The association of the thus obtained 10 sets of input features gave 1134 inputs, which did not include some of the 704 input features selected by the value of StD.
- Aggregation by channels. That is creation of new features by summation of intensity values over several adjacent channels of a spectrum. Several variants of uniform aggregation were tested, with different number of channels aggregated into each single feature – from 8 to 32. The values given in Table V were obtained for optimal uniform aggregation, by each 12 channels.

The results obtained by the methods listed above are presented in Table V.

TABLE V  
MEAN ABSOLUTE ERROR OF DETERMINATION OF CONCENTRATIONS OF SALTS  
IN 5-COMPONENT SOLUTIONS (M, ON THE EXAMINATION DATA SET)  
WITHIN THE "EXPERIMENT-BASED" APPROACH, WITH DIFFERENT DATA  
COMPRESSION METHODS

INITIAL DIMENSIONALITY OF THE INPUT DATA: 1535				
Method of compression	Selection by StD	Selection by GMDH	Selection by CC, CE	Uniform aggregation
Number of inputs after compression	704	314	1134	192
NaCl	0.029	0.031	0.028	0.024
NH <sub>4</sub> Br	0.024	0.025	0.022	0.020
Li <sub>2</sub> SO <sub>4</sub>	0.019	0.018	0.019	0.016
KNO <sub>3</sub>	0.017	0.018	0.018	0.016
CsI	0.023	0.023	0.024	0.020

It can be seen that aggregation is the most effective method of input data compression. Use of aggregation of input features provided decreasing the error of determination of concentrations of salts on the average by 16%. In particular, this means that to solve the desired IP, one can use much less expensive experimental equipment with spectral resolution 8-12 fold worse than that of the equipment used in this study (i.e., equipment with resolution about 20 cm<sup>-1</sup> would be sufficient).

## V. CONCLUSIONS

1. Unique experimental material has been obtained. This material is the array of Raman spectra (in the frequency range from 200 cm<sup>-1</sup> to 4000 cm<sup>-1</sup>) of water solutions of inorganic salts (NaCl, NH<sub>4</sub>Br, Li<sub>2</sub>SO<sub>4</sub>, KNO<sub>3</sub>, CsI) in the range of total concentrations from 0 to 2.5 M (mole per

liter of solution). The array includes 8695 spectra for 4268 solutions with various concentrations of salts.

2. The complex IP of identification of salts and determination of their individual concentrations in 5-component aqueous solution by Raman spectra was solved using perceptron with three hidden layers, within "experiment-based" and "quasi-model" approaches. The "experiment-based" approach provides the maximum precision: the mean absolute error for different salts was from 0.019 to 0.029 M in the concentration range from 0 to 2.5 M.
3. The "quasi-model" approach failed to meet expectations compared to the "experiment-based" approach. In all cases, the results on the experimental examination dataset, obtained within the "quasi-model" approach, turned out to be worse or significantly worse than the results obtained within the "experiment-based" approach. This is an evidence of low adequateness of the quasi-models used.
4. The comparative analysis of input data compression methods was performed. It was demonstrated that use of uniform aggregation of input features by each 12 channels allows decreasing the error of determination of individual concentrations of components by 16-18% on the average. This means that to solve the studied IP, one can use much less expensive experimental equipment with spectral resolution 8-12 fold worse than that of the equipment used in this study (i.e., equipment with resolution about 20 cm<sup>-1</sup> would be sufficient).

## ACKNOWLEDGMENT

ANN computations in this study were performed with NeuroShell 2 software package by Ward Systems Group, Inc. (<http://www.wardsystems.com/neuroshell2.asp>).

## REFERENCES

- [1] M. H. Hassoun, *Fundamentals of Artificial Neural Networks*. Massachusetts, Cambridge: MIT Press, 1995.
- [2] S. S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall, 1999.
- [3] S. A. Dolenko, A. V. Filippov, A. F. Pal, I. G. Persiantsev, and A. O. Serov, "Use of neural network based auto-associative memory as a data compressor for pre-processing optical emission spectra in gas thermometry with the help of neural network," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment (NIMA A)*, vol. 502, no. 2-3, pp. 523-525, 2003.
- [4] S. A. Burikov, A. M. Vervalde, I. I. Vlasov, S. A. Dolenko, K. A. Laptinskiy, and T. A. Dolenko, "Use of neural network algorithms for elaboration of fluorescent biosensors on the base of nanoparticles," *Optical Memory and Neural Networks (Information Optics)*, vol. 22, no. 3, pp. 156-165, 2013.
- [5] M. Talone, R. Sabia, A. Camps, M. Vall-Ilossera, C. Gabarró, and J. Font, "Sea surface salinity retrievals from HUT-2D L-band radiometric measurements," *Remote Sens. Environ.*, no. 114, pp. 1756-1764, 2010.
- [6] A. Camps, J. Font, M. Vall-Ilossera, I. Corbella, N. Duffo, F. Torres, S. Blanch, A. Aguasca, R. Villarino, C. Gabarró, L. Enrique, J. Miranda, R. Sabia, and M. Talone, "Determination of the sea surface emissivity at L-band and application to SMOS salinity retrieval algorithms: Review of the contributions of the UPC-ICM," *Radio Science*, vol. 43, p.RS3008, 2008.
- [7] J. Font, J. Boutin, N. Reul, P. Spurgeon, J. Ballabrera-Poy, A. Chuprin, et al., "SMOS first data analysis for sea surface salinity determination," *International J. of Remote Sensing*, no. 1, pp. 1-17, 2013.

- [8] T. G. Balicheva, and O. A. Lobaneva, *Electronic and vibrational spectra of inorganic and coordination compounds*. Leningrad: Leningrad State University, 1983, pp. 9-81. (Russian)
- [9] S. F. Baldwin, and C. W. Brown, "Detection of ionic water pollutants by laser excited Raman spectroscopy," *Water Research*, vol. 6, pp. 1601-1604, 1972.
- [10] W. W. Rudolph, and G. Irmer, "Raman and infrared spectroscopic investigation on aqueous alkali metal phosphate solutions and density functional theory calculations of phosphate-water clusters," *Applied spectroscopy*, vol. 61, no 12, pp. 274A-292A, 2007.
- [11] F. Rull, and J. A. De Saja, "Effect of electrolyte concentration on the Raman spectra of water in aqueous solutions," *J. Raman Spectroscopy*, vol. 17, no. 2, pp.167-172, 1986.
- [12] G. E. Walrafen, "Raman studies of the effects of temperature on water and electrolyte solutions," *J. Chem. Phys.*, vol. 44, no. 4, pp. 1546-1558, 1966.
- [13] T. A. Dolenko, I. V. Churina, V. V. Fadeev, and S. M. Glushkov, "Valence band of liquid water Raman scattering: some peculiarities and applications in the diagnostics of water media," *J. Raman Spectroscopy*, vol. 31, pp. 863-870, 2000.
- [14] T. A. Gogolinskaia, S. V. Patsaeva, and V. V. Fadeev, "The regularities of change of the 3100-3700  $\text{cm}^{-1}$  band of water Raman scattering in salt aqueous solutions," *Doklady Akademii Nauk SSSR*, vol. 290, no. 5, pp. 1099-1103, 1986.
- [15] G. M. Georgiev, T. K. Kalkanjiev, V. P. Petrov, and Zh. Nickolov, "Determination of salts in water solutions by a skewing parameter of the water Raman band," *Applied Spectroscopy*, vol. 38, no. 4, pp. 593-595, 1984.
- [16] K. Furic, I. Ciglencecki, and B. Cosovic, "Raman spectroscopic study of sodium chloride water solutions," *J. Molecular Structure*, vol. 6, pp. 225-234, 2000.
- [17] A. Yu. Bekkiev, T. A. Gogolinskaia, and V. V. Fadeev, "Simultaneous determination of temperature and salinity of sea water by the method of laser Raman spectroscopy," *Doklady Akademii Nauk SSSR*, vol. 271, no. 4, pp. 849-853, 1983.
- [18] S. A. Burikov, T. A. Dolenko, V. V. Fadeev, and A. V. Sugonyaev, "Identification of inorganic salts and determination their concentrations in water solutions above the water Raman valence band using artificial neural networks," *Pattern Recognition and Image Analysis*, vol. 15, no. 2, pp. 520-523, 2005.
- [19] S. A. Burikov, T. A. Dolenko, and V. V. Fadeev, "Identification of inorganic salts and determination of their concentrations in multicomponent aqueous solutions by the Raman valence band of water using artificial neural networks," *Neurocomputers: Development, Application*, no. 5, pp. 62-72, 2007.
- [20] S. A. Burikov, S. A. Dolenko, T. A. Dolenko, and I. G. Persiantsev, "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)*, vol. 19, no. 2, pp. 140-148, 2010.
- [21] S. A. Dolenko, S. A. Burikov, T. A. Dolenko, and I. G. Persiantsev, "Adaptive Methods for Solving Inverse Problems in Laser Raman Spectroscopy of Multi-Component Solutions," *Pattern Recognition and Image Analysis*, vol. 22, no. 4, pp. 551-558, 2012.
- [22] I. V. Gerdova, S. A. Dolenko, T. A. Dolenko, I. V. Churina, and V. V. Fadeev, "New opportunity solutions to inverse problems in laser spectroscopy involving artificial neural networks," *Izvestiya Akademii Nauk Seriya Fizicheskaya*, vol. 66, no. 8, pp. 1116-1124, 2002.
- [23] D. Specht, "A General Regression Neural Network," *IEEE Trans. on Neural Networks*, vol. 2, no. 6, pp. 568-576, Nov. 1991.
- [24] H. R. Madala, and A. G. Ivakhnenko, *Inductive Learning Algorithms for Complex Systems Modeling*. CRC Press, 1994.