SIMULTANEOUS DETERMINATION OF TEMPERATURE AND SALINITY OF NATURAL WATERS BY RAMAN SPECTRA USING ARTIFICIAL NEURAL NETWORKS DATA ANALYSIS

Alexey Vervald, Ernest Mazurin, and Ivan Plastinin

Lomonosov Moscow State University, Physics Department, Moscow, Russia; {amver / plastinin_ivan}(at)mail.ru

ABSTRACT

A method of simultaneous determination of temperature and salinity of water by Raman spectra was tested and validated on natural waters from the White Sea area. The basis of the method is the solution of this multi-parametric inverse problem by a modern technique of pattern recognition: artificial neural networks. Testing of the presented method was carried out on natural waters of unique meromictic lakes of the White Sea coast. The accuracy of determination of natural water parameters is 0.1°C for temperature and 0.2 psu for salinity in laboratory conditions in the investigated ranges of variation of parameters.

INTRODUCTION

The problem of remote determination of such key parameters of natural waters as temperature and salinity has become very important in recent years. This is due to the fact that salinity (*S*) and temperature (*T*) can indicate the transfer of energy and mass in conterminous layers of oceans, rivers and lakes, they determine the degree of desalination of the surface layer of water reservoirs because of global warming, melting of ice and other reasons.

For real-time determination of temperature and salinity of water in field conditions, non-contact radiometric methods of determination of either salinity or temperature of the water surface currently widely used (1,2). However, due to the surface roughness, and changes of the emissivity of the water surface under the influence of wind, cloudiness, presence of oil films, the radiometric measurement of seawater salinity and temperature, based on the dependence of the emissivity of the water surface on the concentration of salts, allows us to determine *S* with an error not better than a few tenths of the practical salinity unit (psu) and *T* with an error of 1°C (3). As opposed to this, methods of laser vibrational and fluorescent spectroscopy are free from these shortcomings and they can provide information about parameters of water from a distance of some metres or from any point of the water column by using of waveguide cable.

The possibility of using Raman and fluorescence spectroscopy for the diagnosis of aquatic environments is owing to the fact that the quantitative characteristics of the spectral bands are highly sensitive to the type and concentration of dissolved organic matter (DOM) in water, as well as to changes of the water temperature (4,5) and salinity. That is why laser spectroscopy is the basis of methods for determining the presence and amount of inorganic and organic compounds in water and determining water temperature and salinity.

Thus, the method of determining water temperature and salinity in the presence of DOM is based on the influence of these parameters on the water Raman spectrum, in the first place, on the valence band of water (2800 - 3800 cm^{-1}). Such multi-parametric inverse problems can be solved using pattern recognition methods. The method of simultaneous measurement of temperature and salinity in natural waters using the technique of artificial neural networks (ANN) was elaborated in (6,7,8,9). In addition, a large experimental database was created and the optimal ANN was trained (see the section “ANN technique”) in these works.

DOI: 10.12760/02-2015-1-04

In this work, the method of simultaneously determining the temperature and salinity of natural waters by Raman spectroscopy was carried out on water samples from natural sources. Due to the required wide range of changes of salinity, meromictic lakes were chosen for water sampling. The characteristic feature of meromictic lakes is the absence of mixing between layers with different mineralization. Moreover, a huge diversity of dissolved organic matter is present in the different layers. Thus, for quality testing of our method, samples were taken from seven meromictic lakes near the White Sea Biological Station of Moscow State University (WSBS, Kandalakshsky Bay, White Sea) (see Figure 1). Information about these lakes is given below (see the section Results, Figure 4).

Figure 1: Area map of the expedition 2014 in the area of the White Sea Biological Station with marked lakes where the samples were taken from. Lakes: 1 - *Upper Yershovskoye, 2 - Lower Yershovskoye, 3 - lake in the Green Cape, 4 - Kislo-sladkoye, 5 - Verkhneye, 6 - Vodoprovodnoye, 7 - Krugloye.*

METHODS

The method of simultaneously determining the temperature and salinity of natural waters is based on the effect of these parameters on the Raman spectrum of water. To solve this inverse problem of the recognition of *T* and *S* by changes in spectra, the artificial neural networks technique was used.

Experiment

The Raman scattering signal was excited using an argon laser (wavelength 488 nm, power 450 mW, linear polarization). Integral spectra (polarized and depolarized) were obtained in 90° geometry using a monochromator (Acton 2500i, grating 900 lines/mm, focal length 500 mm), and a CCD camera (Jobin Yvon, Synapse1024 128 BIUV-SYN). To suppress the elastic scattering signal an interference edge filter (Semrock) was used, which made it possible to approach the excitation laser line down to 100 cm^{-1}. The practical resolution of the Raman spectrometer was 2 cm^{-1}. The thermostabilizational system provides a control and measurement of water temperature with an accuracy of 0.1°C.

Measurement conditions were the following: the camera exposure time was 5 s for valence bands and 10 s for low-frequency bands (average of 10 spectra). All spectra were corrected by the spectral sensitivity of the spectrometer measured using a light source with a continuous spectrum based on a temperature lamp operating at 2850 K. The obtained spectra were also normalised by the laser power and by the spectrum accumulation time.

Figures 2 and 3 from (9) show the Raman valence bands and low-frequency Raman bands of aqua solutions with the different sets of *T*, *S* and *DOM* (Humic acid sodium salt H16752, Aldrich) values recorded under identical conditions. These spectra are part of the training set for the ANN (see the section "ANN technique").

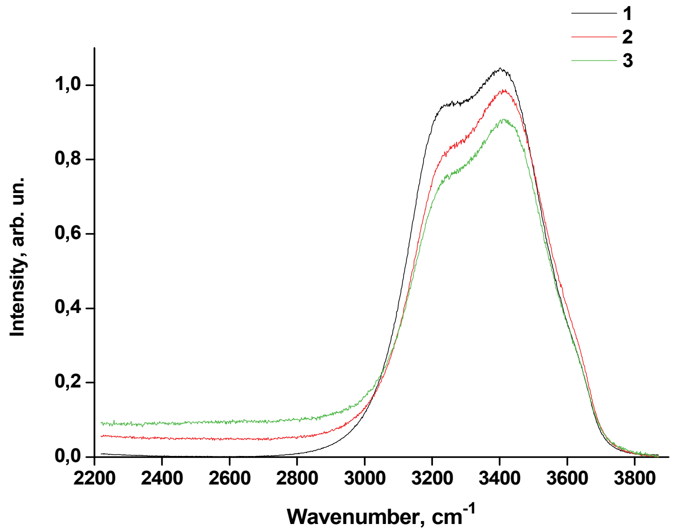

Figure 2: Raman valence bands of aqua solutions with the following set (T, S, DOM) of parameter values: 1 – (0°C, 25 psu, 0 mg/l); 2 - (25°C, 15 psu, 175 mg/l); 3 - (15°C, 45 psu, 350 mg/l), from (9).

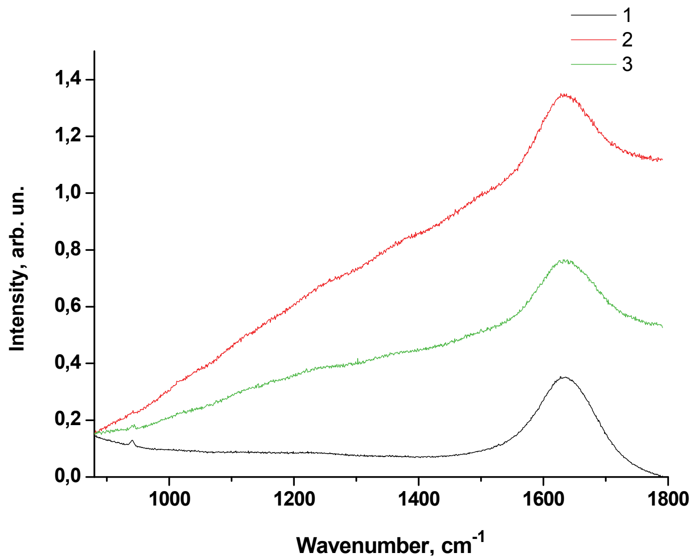

Figure 3: Low-frequency Raman bands of aqua solutions with the following set (T, S, DOM) of parameter values: 1 – (15°C, 40 psu, 0 mg/l); 2 - (25°C, 15 psu, 175 mg/l); 3 - (21°C, 15 psu, 58.3 mg/l), from (9).

Multiparametric inverse problem

The basis of the method is the effect of temperature and salinity on the Raman spectrum of water. As can be seen from Figure 2, this effect changes the shape and position of the water Raman valence band: When the water temperature and/or the concentration of dissolved salts is increased, the intensity of the high-frequency region of the valence band is increased and the intensity of the low-frequency region is decreased. At the same time, the band narrows and shifts to higher frequencies. These effects on the shape of the water Raman valence band were described and their use for determining the temperature was shown in (10,11,12) and for determining the salt concentration in (13).

However, trends in the behaviour of the valence Raman band of water under the influence of temperature and various salts are quite similar. Nevertheless, temperature and different ions cause different quantitative changes in the position and shape of the valence Raman band of water, and it enables us to successfully solve the inverse problem, i.e., the measurement of temperature and salinity of the samples *via* Raman spectroscopy using a special technique. As can be seen in Figures 3 and 4, a fluorescence pedestal (200 - 3500 cm^{-1}) due to the presence of *DOM* in water exerts some influence on the shape of the valence band of water (9). Therefore, the problem of determining *T* and *S* must take into account the presence of *DOM* in waters. Such multi-parametric inverse problems are successfully solved using pattern recognition methods.

ANN technique

To solve the announced multi-parametric inverse problem, the technique of artificial neural networks (ANN) has been used. ANN represent a powerful data analysis algorithm that provides an efficient solution of inverse problems and problems of pattern recognition, including those in optical spectroscopy (14). In this study, the most widespread type of artificial neural networks - the multilayer perceptron (MLP) - was used. Every layer of the MLP consists of *N* elementary units that are so-called formal neurons, units with several inputs and single output. Each of the inputs is characterized by its own weight coefficient, w_j. The value y at the output of a neuron is calculated as a weighted sum of the values x_j at its inputs, put through a non-linear monotonically increasing step function *F* called activation function:

$$y = F\left(\sum_{j=0}^{N} w_j x_j\right), \quad x_0 \equiv 1 \qquad (1)$$

N identical neurons, different only by their weights, are combined to layers. Each neuron is connected with each of the neurons of the preceding and the next layers (fully connected scheme). So, every neuron takes for the inputs every value of all neurons from the previous layer. The signal is fed, according to Eq. (1), from the input of the first layer of the MLP to the output of the last one.

The first layer of an MLP is called the input layer, as it is used to feed data into the MLP, and serves as the input of the whole network. There are as many neurons in the input layer as there are features describing each of the processed data samples. The neurons of the input layer perform no calculations; they simply distribute the input signal to all the neurons of the next layer. The last layer of an MLP is called the output layer; it serves as the output of the whole network. There are as many neurons in the output layer as values that are to be simultaneously determined for each data sample. One or several layers between the input and the output layers are called hidden layers. The numbers of neurons in the hidden layer(s) determine the complexity of the network; it can be estimated by some half-empirical equations, but often it is determined by trial and error.

When the MLP is applied, its output depends on the values of the weight coefficients *w* in Eq. (1). Therefore, to provide correct answers of an MLP, it should be trained prior to its application. This is done by analysing the answers of the MLP on the samples of the training set of data, for which the correct (desired) answers are known, with subsequent tuning of the weight coefficients. The goal of training is to minimize the average error of the MLP over the training set, by changing the values of the weight coefficients of all neurons. Mostly, this is performed by the so-called error backpropagation algorithm, which performs gradient descent in the space of weight coefficients (14).

In the final stage, the Raman spectra of a water sample is presented to the trained ANN as input parameters, and the search parameters are obtained as output values of ANN. The input layer is fed by all spectral points of one sample when the output layer consists of only two neurons which present *T* and *S* of this sample.

In this work, the optimal ANN that was trained in (9) - five-layer perceptron - was used. To train ANN a large base of experimental data - Raman spectra of water from 500 cm^{-1} to 4000 cm^{-1} at different temperatures, salinities and different content of *DOM* (482 spectra) - was obtained. The spectra for training, testing ANN sets (9) and for the validation set [this work] were obtained under identical conditions (see the section "Experiment"). The values of *T*, *S* and *DOM* concentration changed in the following ranges: *T*: 0,5,…,30,35°C; *S*: 0,5,…, 40, 45 psu., *DOM*: 0,…,350 mg/l. The preparation and characterisation of the samples for training are described in work (9). It was demonstrated that the method of determining *T* and *S* of seawater using low- and high-frequency regions of water Raman spectra and ANN (five-layer perceptron) enables *T* and *S* to be determined in a wide DOM concentration range with mean absolute errors of 0.8°C and 1.1 psu, respectively.

Validation of the method

The validation of the method of simultaneously determining *T* and *S* comprises the following steps:

- Collecting a diverse set of water samples from lakes of the White Sea area.
- Contact measurements of samples' salinity using a conductometer YSI-85 in the field conditions (accuracy 2%). Contact measurements of the sample temperature using a calibrated probe based on a thermistor in the laboratory (accuracy 0.05°C).
- Spectral measurement of water samples. Application of the trained ANN (five-layer perceptron) to the water Raman spectra.
- Comparison of results of *T* and *S* measurements. Estimation of experimental errors.

RESULTS

Natural water samples were collected from the meromictic lakes of the coast of the White Sea on the peninsula Kindo during a student expedition of Moscow State University, which took place near the White Sea Biological Station of Moscow State University (WSBS) (Kandalakshsky Bay, White Sea) from 26 January to 6 February 2014,. Natural waters of the White Sea area were taken from seven meromictic lakes: Lower and Upper Yershovskoye (31.01.2014), lake in the Green Cape (30.01.2014), Kislo-sladkoye (29.01.2014), Verkhneye (01.02.2014), Vodoprovodnoye (01.02.14) and Krugloye (01.02.2014). Some lake characteristics lakes are presented in Figure 4. In this figure, *h* means the maximum depth of the lake. Position of the blue rectangle presents the altitude of the lake above sea level. *S* is for the area of the lake and *F* is for the area of the lake basin. In accordance with their altitude, Verkhneye, Vodoprovodnoye and Krugloye lakes appear to be limnetic, while the other four lakes appear to have salt in some layers.

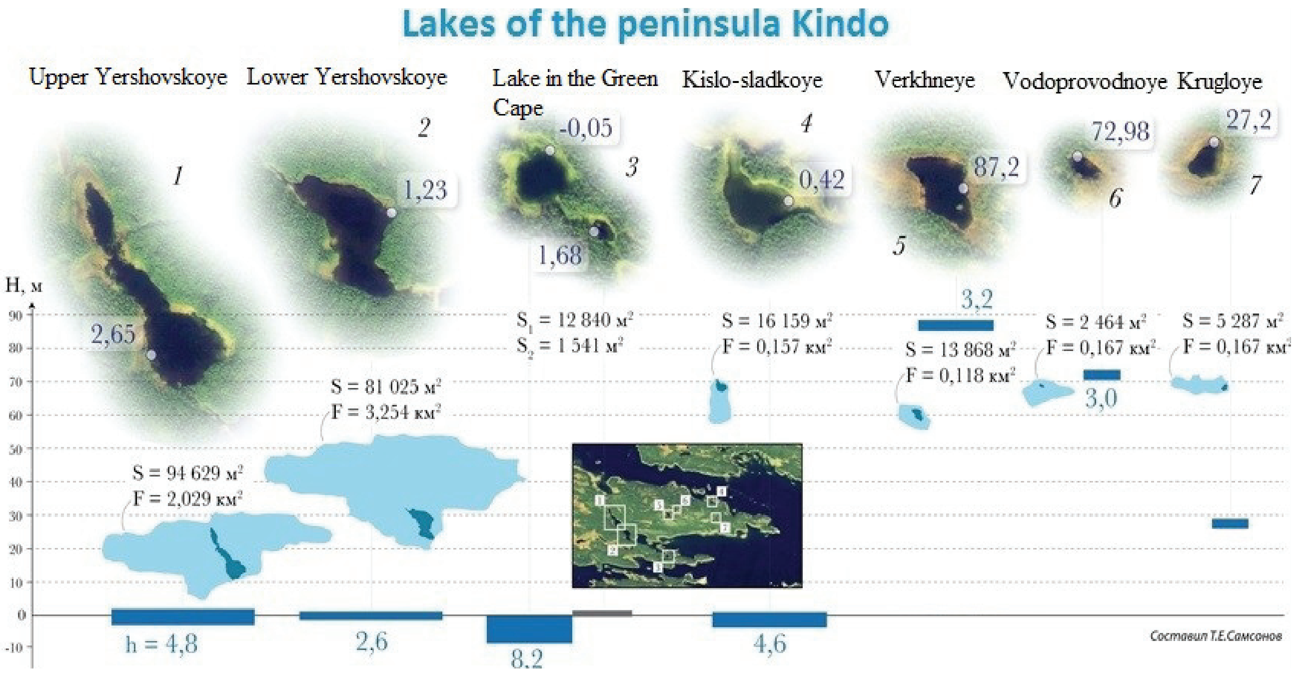

Figure 4: Description of the lakes of the peninsula Kindo.

Water samples were collected with a Whale Premium Submersible Pump GP1352 from every upper to every lower layer of the lakes at steps of 0.5 m. This procedure allowed us to avoid mixing and collect a diverse set of data. Salinity was measured *in situ* using a field conductometer YSI-85. Following collection, the samples were carefully packed to prevent evaporation and pollution. Due to this, it was assumed that salinity had not changed from the time of sampling to the time of measurements in the laboratory.

Our targets were to test and verify the method on most diverse samples. Due to small differences between limnetic lakes and between Upper and Lower Yershovskoye lakes the samples were selected from only four lakes for the validation. In addition, the properties of the lakes vary strongly between different layers. According to these arguments, 16 samples were chosen (Table 1). Figures 5 and 6 show the obtained Raman spectra of selected samples. The graphs obtained from a given sample are shown by the same colour. The samples 1, 26, 36, 41 and distilled water are marked by a strait line. They are chosen to demonstrate the most different salinities of the set. As can be seen, it is impossible to understand immediately, which sample has a higher *S*. The difference in appearance of the low-frequency pedestal is caused by the different content of *DOM* (see Figures 2 and 3).

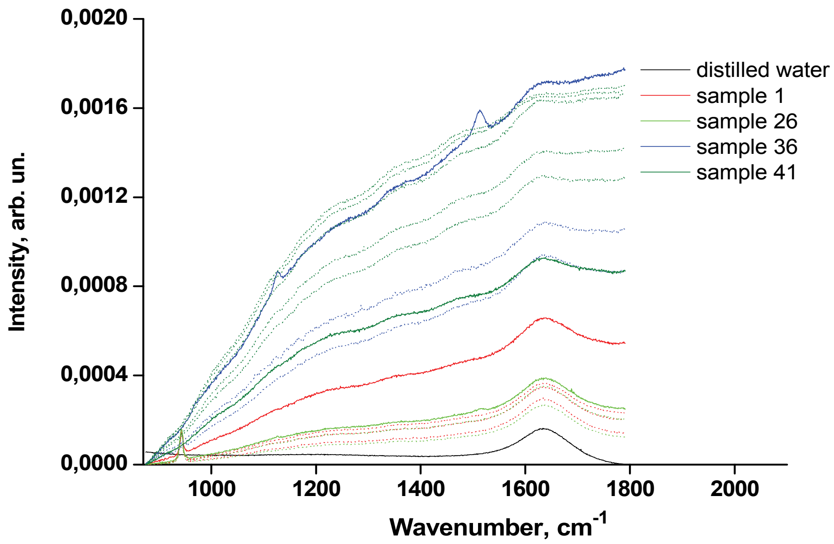

Figure 5: Low-frequency Raman bands of natural waters. Lake colours: red - Kislo-sladkoye, green - Lake in the Green Cape, blue - Lower Yershovskoye, dark green - Vodoprovodnoye.

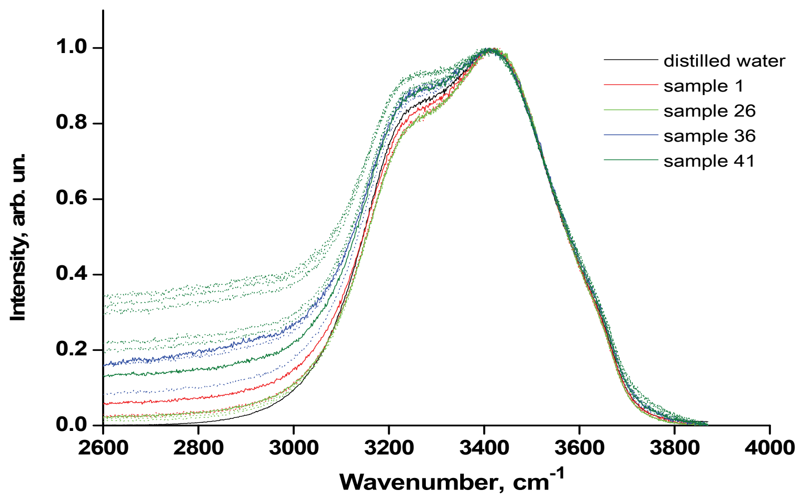

Figure 6: High-frequency Raman bands of natural waters. All spectra are normalized to the maximum intensity. Lake colours are the same as in the Figure 5.

Table 1 presents the results of measuring temperature and salinity of natural water samples by contact methods and the results of using ANN. As can be seen, the ANN technique provides a determination of *T* and *S* without knowing the *DOM* content. It is due to training ANN on the samples with different *DOM* concentration.

Table 1: The results of measuring temperature and salinity of natural water samples using remote and contact methods.

Sample	Date/lake/depth, m	Contact methods		ANN	
		T, °C	*S*, psu.	*T*, °C	*S*, psu.
1	29.01.14 / Kislo-sladkoye / 0	26.8	19.3	27.1	19.3
2	29.01.14 / Kislo-sladkoye / 0,5	27.1	26.9	27.1	27.0
4	29.01.14 / Kislo-sladkoye / 1,5	26.3	27.4	26.5	27.6
9	29.01.14 / Kislo-sladkoye / 4	26.8	27.8	26.6	27.7
14	30.01.14 / Lake in the Green Cape / 0	27.0	27.9	26.8	27.7
21	30.01.14 / Lake in the Green Cape / 3,5	27.1	28.0	27.0	28.0
26	30.01.14 / Lake in the Green Cape / 5,8	27.1	28.8	27.1	28.3
31	31.01.14 / Lower Yershovskoye / 0	25.9	0.06	25.8	0.0
35	31.01.14 / Lower Yershovskoye / 2	26.3	1.0	26.2	1.4
36	31.01.14 / Lower Yershovskoye / 2,5	27.8	4.6	27.7	4.5
41	01.02.14 / Vodoprovodnoye / 0	27.4	0.025	27.5	0.0
42	01.02.14 / Vodoprovodnoye / 0,5	26.5	0.025	26.5	0.0
43	01.02.14 / Vodoprovodnoye / 1	27.0	0.035	26.8	0.0
45	01.02.14 / Vodoprovodnoye / 2	27.1	0.04	27.0	0.0
46	01.02.14 / Vodoprovodnoye / 2,5	26.9	0.045	26.8	0.0
47	01.02.14 / Vodoprovodnoye / 2,7	27.4	0.053	27.2	0.0

Thus, it is clear that the ANN method provides results quite similar to those provided by the contact method.

The estimation of the error in parameter determination for chosen samples showed that the application of ANN provides average values of the absolute error of determination equal to 0.1°C for temperature and 0.2 psu for salinity. It should be mentioned that the errors were estimated under the changes of the parameters *T* from 25.9 to 7.8 ºC and *S* from 0.025 to 28.8 psu. Even so, a combination of laser Raman spectroscopy and ANN technique appears to provide a high accuracy of *T* and *S* measurement in water.

DISCUSSION

The average values of the absolute error of determination are 0.1°C for temperature and 0.2 psu for salinity for natural waters from meromictic lakes. For comparison, the mean absolute errors in the examination set obtained for the modelling probes were 0.8°C and 1.1 psu for the same ANN. ANN was trained in the ranges T = 0…35°C, S = 0…45 psu in steps of 5°C and 5 psu. On the other hand, the variations of these parameters were smaller in the lake waters.

ANN is a technique that provides a good measurement accuracy of parameters in the centre of the trained range of parameters with increasing errors in the edges. In addition, results should be more precise near parameters values of the trained set. From this point of view, such precious measurements are caused by the temperature range: The validation was carried out under changing temperatures from 25.9 to 27.8°C which is too far from the edges.

Furthermore, we should mark the results for the Vodoprovodnoye Lake that appears to be limnetic. Despite the fact that the salinity value in this lake is quite near the edge, ANN provides a good accuracy of salinity and temperature. The reason of such high accuracy is the fact that a temperature of 0°C belongs to the training set. Thus, the mean error for other ranges can be greater than that for our range (0.1°C and 0.2 psu), but it cannot be greater than the errors for the examination set of ANN (0.8°C and 1.1 psu).

Sources of errors and the characteristic time of getting a result are of interest. The main problem of the ANN technique in this field is the noise in the spectra. The common way to obtain more precise results from ANN is to provide a larger training set with more frequent points of parameters. On the other hand, some errors can occur, when the spectrum of the sample of interest is measured under conditions that differ from the spectra of the training set. Moreover, the accuracy of ANN cannot exceed the accuracy of the parameters by which it was trained.

For getting a result by the method under investigation, two steps should be taken: The spectrum of the sample must be obtained and this normalized spectrum must be presented to ANN. The time required for a result from ANN after presentation does not exceed some seconds. Thus, the main time is spent on obtaining the spectrum and the total time of measurement is one or two minutes. Hence such characteristics of the presented method, contactless and rapid, enable us to work in real time aboard.

The method under investigation provides T and S data simultaneously. The authors do not know any other works where this has been demonstrated before. At least a comparison can be performed for the temperature measurements. As an example, we can consider recent results of Lidar applications. Based on Brillouin scattering an information about temperature can be obtained with an accuracy of about 0.1°C from a range of 100 m in water with a 1 m distance resolution (15). However, a disadvantage is the need of a constant known salinity. So it cannot be used, for example, in estuaries and inland waters where salinity can vary greatly and where fluorescence from dissolved organics may distort the backscattering spectra.

CONCLUSIONS

The validation of the method of simultaneously determining water temperature and salinity on natural water samples from the White Sea area by Raman spectroscopy demonstrated the efficiency of using the ANN technique. The accuracy of determination of natural water parameters is 0.1°C for temperature and 0.2 psu for salinity in laboratory conditions. The errors were estimated under changes of the parameters T from 25.9 to 7.8°C and S from 0.025 to 28.8 psu. The high stability of ANN to noise was demonstrated for the determination of T and S in water samples from meromictic lakes with different *DOM* content and S between layers.

This study represents one of numerical stages of elaborating the method of remote determination of T and S in seawater in results for solving a two-parameter problem using the “experiment-based” approach of training ANN. It is planned to describe the dependence of the Raman spectra intensity in each point on the relevant parameters by a polynomial “quasi-model”. It is also planned to increase the representativeness of the data sets for ANN training by modelling the necessary amount of Raman spectra using the “quasi-model” approach.

ACKNOWLEDGEMENTS

The authors wish to thank Nikolai Pertsov White Sea Biological Station (WSBS MSU) and the Faculty of Geography of Moscow State University for organizing the expedition. In addition, we wish to thank our supervisors Tatiana Dolenko and Sergei Burikov (Physics Department, Moscow State University) for their help and guidance.

REFERENCES

1 Font J, J Boutin, N Reul, P Spurgeon, J Ballabrera-Poy, A Chuprin, C Gabarró, J Gourrion, S Guimbard, C Hénocq, S Lavender, N Martin, J Martínez, M McCulloch, I Meirold-Mautner, C Mugerin, F Petitcolin, M Portabella, R Sabia, M Talone, J Tenerelli, A Turiel, J-L Vergely, P Waldteufel, X Yin, S Zine & S Delwart, 2013. SMOS first data analysis for sea surface salinity determination. International Journal of Remote Sensing, 34(9-10): 3654-3670

2 Torres F, I Corbella, L Wu, N Duffo, J Gourrion, J Font & M Martin-Neira, 2012. Minimization of image distortion in SMOS brightness temperature maps over the Ocean. IEEE Geoscience and Remote Sensing Letters, 9(1): 18-22

3 Lagerloef G & J Font. 2010. SMOS and Aquarius/SAC-D Missions: The Era of Spaceborne Salinity Measurements Is About to Begin. In: Oceanography from Space, edited by V Barale, J F R Gower & L Alberotanza (Springer, Heidelberg) 35-58

4 Walrafen G E, 1967. Raman studies of the effects of temperature on water and electrolyte solutions. Journal of Chemical Physics, 47(1): 114-126

5 Leonard DA, C H Chang & L A Young, 1974. Remote measuring of fluid temperature by Raman scattered radiation./ U.S. Patent 3.986.775, Dec.26, 1974

6 Bekkiev A, T Gogolinskaya (Dolenko) & V Fadeev, 1983. Simultaneous determination of temperature and salinity of seawater by the method of laser Raman spectroscopy. Soviet Physics Doklady, 271(4): 849-853

7 Burikov S A, I V Churina, S A Dolenko, T A Dolenko & V V Fadeev, 2004. New approaches to determination of temperature and salinity of seawater by laser Raman spectroscopy. EARSeL eProceedings, 3(3): 298-305

8 Dolenko S, T Dolenko,. S Burikov, V Fadeev, A Sabirov & I Persiantsev, 2012. Comparison of Input Data Compression Methods in Neural Network Solution of Inverse Problem in Laser Raman Spectroscopy of Natural Waters. In: Artificial Neural Networks and Machine Learning – ICANN 2012, edited by Alessandro E. P. Villa, Włodzisław Duch, Péter Érdi, Francesco Masulli & Günther Palm (Springer: Berlin, Heidelberg) 443-450

9 Dolenko T, S Burikov, A Sabirov & V Fadeev, 2011. Remote determination of temperature and salinity in presence of dissolved organic matter in natural waters using laser spectroscopy. EARSeL eProceedings, 10(2): 159-165

10 Leonard D A, B Caputo, R L Johnson & F E Hoge, 1977. Experimental remote sensing of subsurface temperature in natural ocean water. Geophysical Research Letters, 4(7): 279-281

11 Becucci M, S Cavalieri, R Eramo, L Fini & M Materazzi, 1999. Raman spectroscopy for water temperature sensing. Laser Physics, 9(1): 422-425

12 Raimondi V & G Cecchi, 1995. Lidar field experiments for monitoring sea water column temperature. EARSeL Advances in Remote Sensing, 3(3): 84-89

13 Furić K, I Ciglenečki & B Ćosović, 2000. Raman spectroscopic study of sodium chloride water solutions. Journal of Molecular Structure, 550-551: 225-234

14 Hassoun M H, 1995. Fundamentals of Artificial Neural Networks (MIT Press, Cambridge, Massachusetts) 511 pp.

15 Fry E S, 2012. Remote sensing of sound speed in the ocean via Brillouin scattering. Proceedings of SPIE, 8372: 837207-1