

Fluorescence diagnostics of oil pollution in coastal marine waters by use of artificial neural networks

Tatiana A. Dolenko, Victor V. Fadeev, Irina V. Gerdova, Serge A. Dolenko, and Rainer Reuter

We discuss the problems with and the real possibilities of determining oil pollution *in situ* in coastal marine waters with fluorescence spectroscopy and of using artificial neural networks for data interpretation. In general, the fluorescence bands of oil and aquatic humic substance overlap. At oil concentrations in water from a few to tens of micrograms per liter, the intensity of oil fluorescence is considerably lower than that of humic substances at concentrations that typically are present in coastal waters. Therefore it is necessary to solve the problem of separating the small amount of oil fluorescence from the humic substance background in the spectrum. The problem is complicated because of possible interactions between the components and variations in the parameters of the fluorescence bands of humic substances and oil in water. Fluorescence spectra of seawater samples taken from coastal areas of the Black Sea, samples prepared in the laboratory, and numerically simulated spectra were processed with an artificial neural network. The results demonstrate the possibility of estimating oil concentrations with an accuracy of a few micrograms per liter in coastal waters also in cases in which the contribution from other organic compounds, primarily humic substances, to the fluorescence spectrum exceeds that of oil by 2 orders of magnitude and more. © 2002 Optical Society of America

OCIS codes: 300.2530, 300.6280, 010.4450, 100.3190.

1. Introduction

We describe new methods for and uses of spectral analysis that emerged with the application of inverse problem algorithms based on artificial neural networks (ANNs).^{1,2} We take as an example the solution of an extremely important, but also exceedingly complicated, issue in ecology: *in situ* diagnostics of oil pollution (OP) in water.

Monitoring of OP in marine waters remains a topical issue. This is especially true in coastal sea areas, where the consequences of such an adverse influence on the marine ecosystem are particularly hazardous for both the biota and humankind. At the same time, the difficulties involved in OP control in coastal waters are much greater than those en-

countered in open sea areas. In this paper we discuss the possibility of developing an efficient method for *in situ* diagnosis of oil in marine waters. Only *in situ* methods with data available in real time can prove adequate for present-day environmental monitoring of vitally important coastal areas.

In its complete formulation, the problem at hand is exceedingly complicated, and its solution was not to our knowledge reported previously. At present the most effective tool for obtaining information on the presence of OP in water is fluorescence spectroscopy, excited at UV wavelengths and emitted in the UV and in the short-wavelength part of the visible spectrum.^{3,4} From fluorescence data, oil pollution can be detected both in the water column and on the water surface by use of submersible probes and remotely operated lidars.⁵ However, to collect this information quantitatively it will be necessary to overcome the difficulties associated with the facts that, in a real marine environment, transformations of OP may occur and that other organic compounds and complexes are present that are able to fluoresce and interact with OP. An especially complicated situation occurs in coastal waters, where the concentration of natural dissolved organic matter can be as much as several milligrams per liter.

The order of investigations in the present study

T. A. Dolenko, V. V. Fadeev (fadeev@lid.phys.msu.su), and I. V. Gerdova are with the Faculty of Physics, M. V. Lomonosov Moscow State University, Moscow, Russia. S. A. Dolenko is with the Institute of Nuclear Physics, M. V. Lomonosov Moscow State University, Moscow, Russia. R. Reuter is with the Fachbereich Physik, Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg, Germany.

Received 19 November 2001; revised manuscript received 30 April 2002.

0003-6935/02/245155-12\$15.00/0

© 2002 Optical Society of America

and the structure of the material presented in this paper are as follows:

(a) During three expeditions in August–September 1998–2000 to the Black Sea, a large number of water samples taken from various depths at different coastal positions were fluorometrically analyzed (see Section 3 below). It was found that the contribution of OP to the total fluorescence spectrum of the samples is, as a rule, quite small.

(b) A computer simulation (numerical experiment) was carried out with the aims of assessing the potential abilities of an ANN to separate and quantify small contributions of OP to the total fluorescence spectrum and of studying the practical stability of this method (see Section 4 below). We investigated, as a starting material for the computer simulation, fluorescence spectra of fulvic acid solutions, light oil, diesel fuel, and their mixtures at various partial concentrations of the components.

(c) The efficiency of this approach was validated by tests in which the spectra from the seawater samples taken during the Black Sea expeditions were used (Section 5 below).

2. Instruments and Methods

Fluorescence spectra were recorded with two instruments (at different stages of the experiments): a Perkin-Elmer LS50 luminescence spectrometer and a laser spectrometer that uses the 3rd (355-nm) and the 4th (266-nm) harmonics of a Nd:YAG laser or a N₂ laser (337 nm).

A. Luminescence Spectrometer

The Perkin-Elmer LS50 spectrometer was operated in the August 1998 expedition and in laboratory measurements of oil in water emulsions. Emission scans with 230-, 266-, 308-, 337-, and 355-nm excitation wavelengths and excitation scans with 340-, 430-, 600-, 660-, and 680-nm emission wavelengths were recorded. The spectral resolution was set to 7 or 10 nm, depending on the scan type. The scan speed was set to 100 nm/min. The spectra were corrected for the wavelength-dependent sensitivity of the instrument. Corrections of the emission spectra were made (a) at wavelengths below 380 nm with a Spectralon white diffuser set at the position of the water sample volume and with a synchronous excitation–emission scan, making use of the internal calibration of the instrument; (b) in the 380–530-nm range with an emission scan at the 347-nm excitation of quinine sulfate dihydrate (1.28×10^{-6} mol/L) in 0.105 mol/L of HClO₄ (Ref. 6); and (c) at wavelengths above 530 nm with a tungsten halogen lamp with calibrated color temperature.

From these data a correction function for emission spectra was derived for the entire relevant wavelength range. We measured the correction function for excitation spectra by using a solution of Rhodamine B dye as a quantum counter.

Analyses of water samples collected on board ship with Niskin bottles were made in the laboratory of

the P. P. Shirshov Institute in Russia. Unfiltered samples were stored in glass bottles in the dark at room temperature. The time delay between sampling and analysis was typically 10 h and not more than 30 h. Before each measurement the sample bottles were shaken, and the sample was transferred into a quartz flow-through cuvette through Teflon tubes after rinsing. The purity of the tubes and cuvette was checked before each sample run by use of purified Millipore water.

B. Laser Spectrometer

With the laser spectrometer, excitation was limited to the three wavelengths mentioned above. The advantage of this instrument is its high sensitivity, which allows us to record spectra with a polychromator with a narrow entrance slit (25 μm, corresponding to a spectral resolution of 0.4 nm). With this resolution, not only the fluorescence bands but also the narrow band of Raman scattering is well spectrally resolved. The laser spectrometer was operated on board the research vessel *Akvanavt*, which was used for investigations in the present study. Equipped with a light guide cable attached to a submersible probe, it allows the water column to be probed also when the ship is under way.⁷ Using such a spectrometer allows us to perform nonlinear fluorimetry.^{8,9} As a receiver we used an optical multichannel analyzer assembled on the basis of an EG&G polychromator (model 1226), equipped with a UV CCD chamber (DeltaTech, Scientific Park of Moscow University). The spectral sensitivity of the laser spectrometer receiver in the wavelength range 360–550 nm was obtained by comparison of fluorescence spectra of the same anthracene solution sample (concentration, 10 μg/L) measured with the laser spectrometer and with a Hitachi-850 spectrafluorimeter that had a built-in system of automatic spectra correction for spectral sensitivity. (Support for performing measurements with this fluorescence spectrometer was provided by the Biophysics Chair of the Biological Department of Moscow State University.)

Samples for the laser spectrometer were taken on board the research vessel *Akvanavt* by use of clean Niskin bottles, from depths selected based on conductivity–temperature–depth probe readings. The sampled water was then poured into the system that pumped the liquid through the analytical volume of the laser spectrometer. Both original seawater samples and (selectively) their hexane extracts, in which OP extracted from the water with hexane was concentrated, were analyzed. No other treatment of the samples (e.g., filtration) was performed.

C. Data Evaluation

We converted data into quantitative spectral information by using the water Raman scatter band to quantify the individual wavelength readings. The Raman signal is a well-defined signal with known quantum efficiency, line shape, and dependence on excitation wavelength, and it is present in all measured spectra. We made our calibrations by divid-

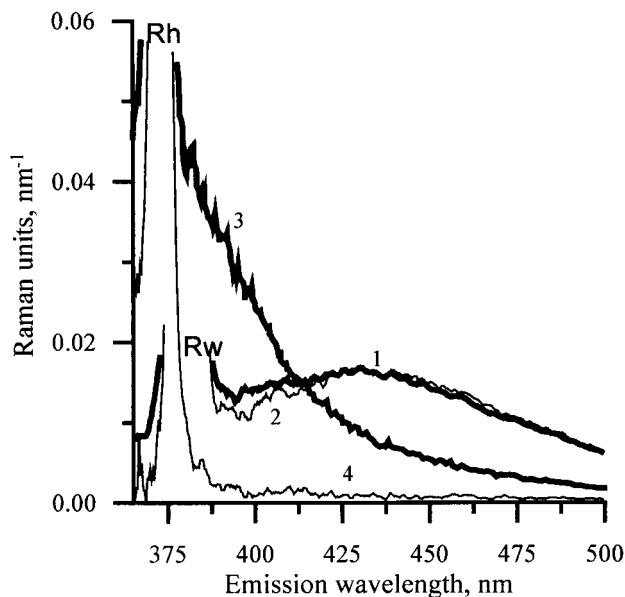


Fig. 1. Typical fluorescence spectra of a Black Sea water sample (expedition 2000) taken at 0.5-m depth. Excitation wavelength, 337 nm. 1, Initial seawater sample; 2, sample after extraction of OP by hexane (ratio of water volume V_w to hexane volume v_h , $V_w/v_h = 40$); 3, a hexane extract, where Rw and Rh are Raman scattering bands of water and hexane, respectively; 4, pure hexane. Spectra were measured with the laser spectrometer.

ing the spectrally corrected fluorescence signals at a given emission wavelength by the water Raman scatter band integrated over its entire bandwidth, after subtraction of its baseline, which is due to fluorescence at the Raman scatter wavelengths. Inasmuch as signal intensities taken at specific wavelengths are normalized to intensities integrated over a wavelength interval, data are thus given in units of inverse wavelengths (nm^{-1}), denoted in Raman units.¹⁰ A comparison of these units with units based on quinine sulfate, which is also commonly used, is found in Refs. 11 and 12.

Figure 1 (curve 1) illustrates a typical spectrum calibrated in this way: the response from a water sample excited by N_2 laser radiation ($\lambda_{\text{exc}} = 337 \text{ nm}$). To find the total fluorescence intensity of a given sample independently of its spectral shape we calculated fluorescence parameter $\Phi_0 = N_f/N_{\text{RS}}$, where N_f and N_{RS} are the numbers of fluorescence and water Raman scatter photons (or of any other solvent), respectively.^{13,14} To calculate this parameter we converted the measured spectra (intensity versus wavelength) to the spectra (intensity versus wave number) for which the fluorescence intensity was integrated over all significant emission wave numbers and normalized to the integral water Raman-scattering band.

Having removed OP from the seawater by hexane extraction, one obtains the fluorescence background, which is largely due to aquatic humic substance (AHS), illustrated by spectrum 2 in Fig. 1. The concentration of OP can be determined from the value of the fluorescence parameter of the hexane extract

(spectrum 3 in Fig. 1) if the necessary calibrating procedures^{8,15} are performed.

3. Properties of the Fluorescence Spectra of Black Sea Water Samples

In the expeditions mentioned above, a large number of fluorescence excitation and emission spectra of seawater samples were recorded. The samples were taken from various depths in the Novorossiysk port areas (Novorossiyskaya or Tsemesskaya Bay), in the near-shore areas of the resort city of Gelendzhik (Gelendzhikskaya and Golubaya Bay), and in the area of the settlement of Yuzhnaya Ozereyevka, where the construction of a big oil terminal is being planned. Samples were also taken in the coastal zone between these areas down to 200-m depth, where the hydrogen sulfide layer begins. Qualitatively, the recorded spectra of the optical response from the seawater samples are quite consistent with previously published ones (see, e.g., Refs. 8 and 16–20). For all the excitation wavelengths that we used, the spectra are dominated by a broad fluorescence band of AHS ($\lambda_{\text{fl}}^{\text{max}} \approx 440 \text{ nm}$, $\Delta\lambda_{\text{fl}} \approx 95 \text{ nm}$) and by the Raman-scattering band of water. At $\lambda_{\text{exc}} = 337, 355 \text{ nm}$ the fluorescence band of OP is superimposed upon that of an AHS. At $\lambda_{\text{exc}} = 230, 266 \text{ nm}$ the fluorescence bands of light fractions of oils and oil products appear in the range 300–350 nm; these bands, however, are superimposed upon the fluorescence bands of proteinlike compounds.²¹ At $\lambda_{\text{exc}} = 308 \text{ nm}$, one observes a transient picture: The short-wavelength fluorescence bands of OP and proteinlike compounds are less pronounced than at $\lambda_{\text{exc}} = 266 \text{ nm}$, and moreover they are superimposed not only upon one another but also on the Raman scattering band.

Thus, in all cases, one has to solve the problem of separating the fluorescence of OP from the background, which is due mainly to AHSs and proteinlike compounds (although the presence of other fluorescent admixtures cannot be excluded). The complexity of this problem depends on the peculiarities encountered in a particular geographic water area. As a rule, for the coastal areas, where the fluorescence background is significantly more intense than in open sea areas, the problem of separating OP from background is extremely complicated at any excitation wavelength.

In this study we investigated the possibility of solving this problem for the spectra obtained at $\lambda_{\text{exc}} = 337 \text{ nm}$ (see Fig. 1). One would expect that separating the OP from the background at $\lambda_{\text{exc}} = 337 \text{ nm}$ (or at $\lambda_{\text{exc}} = 355 \text{ nm}$) would be a good prerequisite for solving this problem also at $\lambda_{\text{exc}} = 308, 266, 230 \text{ nm}$.

A. Magnitude of the Contribution of Oil Pollution to the Aquatic Humic Substance Fluorescence Band

In the expeditions mentioned above the fluorescence parameter Φ_0 varied in the range from 5 to 12, and the concentration of OP (from the results of selective analyses made by the extraction method) mostly varied from 1 to 20 $\mu\text{g/L}$, with rare exceptions when this concentration reached 40–50 $\mu\text{g/L}$. (In Russia the

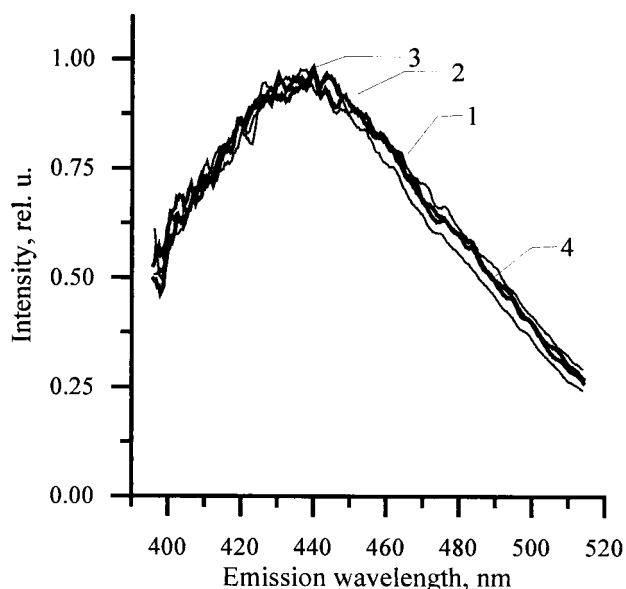


Fig. 2. Fluorescence bands of seawater samples after extraction of OP by hexane. The samples were taken from Blue Bay at 0.5-m depth. Distances of sample positions from the coastline, in meters: 1, 100; 2, 300; 3, 600; 4, 1000. Excitation wavelength, 337 nm. Spectra were measured with the laser spectrometer.

maximum permissible concentration of OP in water is $50 \mu\text{g/L}$).²² At typical OP concentrations of less than $10 \mu\text{g/L}$ the OP fluorescence is barely noticeable in the fluorescence spectra of the coastal waters that we studied so the contribution of OP to the total fluorescence band does not exceed a few percent (see Fig. 1). This fact makes the problem of separating the OP contribution, i.e., *in situ* determination of the parameter Φ_0^{OP} , extremely difficult.

B. Variations of Width and Position of the AHS Fluorescence Band

Accuracy in determining small contributions by OP to the total fluorescence spectrum by use of the ANN technique depends on the accuracy of the information concerning the shape and position of the component bands at the time when a measurement is made.

Obtaining such information directly when one is determining the parameter Φ_0^{OP} *in situ* is extremely difficult. Accordingly, an evaluation of the degree of stability (or variability) of the shape and position of the fluorescence bands of the components from point to point in the area under study is of great importance. This evaluation was made for the AHS band (i.e., the seawater fluorescence after OP extraction with hexane). Figure 2 illustrates the degree of variability of the shape of this band. As can be seen, the bandwidth at half-height varies within 10%; the position of the maximum shifts by 5 nm.

These variations in the width and position of the fluorescence band of AHS may be caused by a number of factors: a change in the AHS type, some contribution from other organic impurities that are not extracted from water by hexane, the effect of insolation, and the time lapse between the moment of sam-

pling and that of recording the spectrum. Specially designed tests have revealed that certain weak dependences on these factors were indeed observed.

In Section 4 the data that we obtained on the stability of the background fluorescence, against which the OP contribution must be determined, are discussed in the context of the practical stability of the solution of the inverse problem with ANN.

4. Determination of the Contribution of Oil Pollution to the Fluorescence Bands of Model Samples by Use of an Artificial Neural Network

Our main objective in this paper is to demonstrate the new methods for and uses of spectroscopy that have emerged with the advent of ANNs. The mathematical algorithms and procedures based on ANNs enable one to solve effectively a large variety of problems of diagnosing, evaluating, classifying, and recognizing images of widely different natures.¹ They have also come to be applied in laser spectroscopy.^{23,24} It has been shown that the ANN technique is indeed effective in solving the inverse problems of saturation fluorimetry.^{9,25}

As is known,^{1,2} a training set of data is used to train a network. In the present case the training set is a data set of spectra. To control the quality of the network in the process of training we use a test set; to estimate the accuracy of parameter reproduction by the trained network we use a checking (examination) set.

We can reduce the influence of the input data noise on the quality of the algorithm to a significant extent by training the network with data degraded by noise. We do so by feeding the network (in the course of training) with patterns of the training set with added random noise in a given range of amplitudes.

These procedures were used in the present study.

A. Preparation of Model Samples

So-called base solutions of AHS and OP were prepared (in the latter case, the term "solution" denotes the emulsion and the dissolved fraction taken together). As a model of AHS, fulvic acid (FA) extracted from soil was used. FA is one of the fractions of AHS, and in seawater FA prevails over the other AHS fraction humic acid; as was found in a number of studies,²⁶ the shape of the band of FA practically coincides with the band shape of the entire AHS. Because FA was available as a powder, we prepared base solutions of FA in purified water with known concentrations in the range 0.15–2.8 mg/L, corresponding to $\Phi_0^{\text{FA}} = 1\text{--}23$ at $\lambda_{\text{exc}} = 337 \text{ nm}$.

We used heavy crude oil (HO), light crude oil (LO), and diesel fuel (DF) as the oil pollutant. Base model samples were prepared with a few drops of LO or DF added to the water in a beaker and placed in an ultrasonic bath for 30 min. We prepared mixtures by adding base solutions in given proportions. As the theoretical weight coefficients for the components of the mixtures, we used the calculated partial values of the parameters Φ_0 . Figure 3 illustrates an exam-

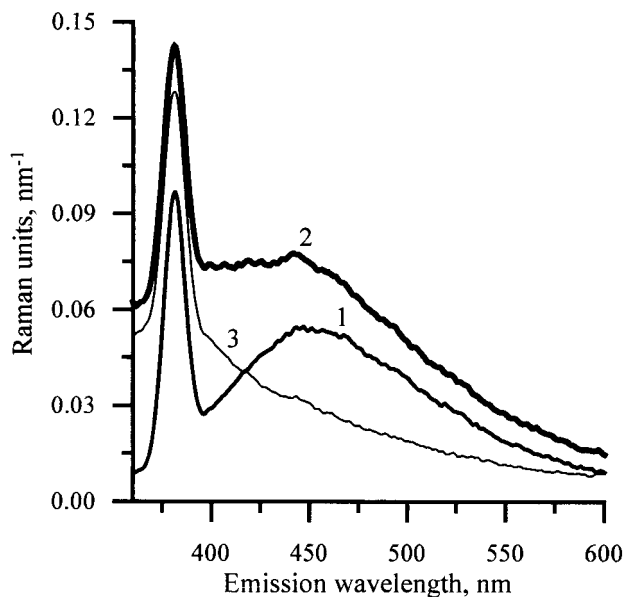


Fig. 3. Fluorescence spectra of 1, FA (0.9 mg/L); 2, LO; and 3, their mixture in distilled water. Excitation wavelength, 337 nm. Spectra were measured with the Perkin-Elmer LS50 luminescence spectrometer.

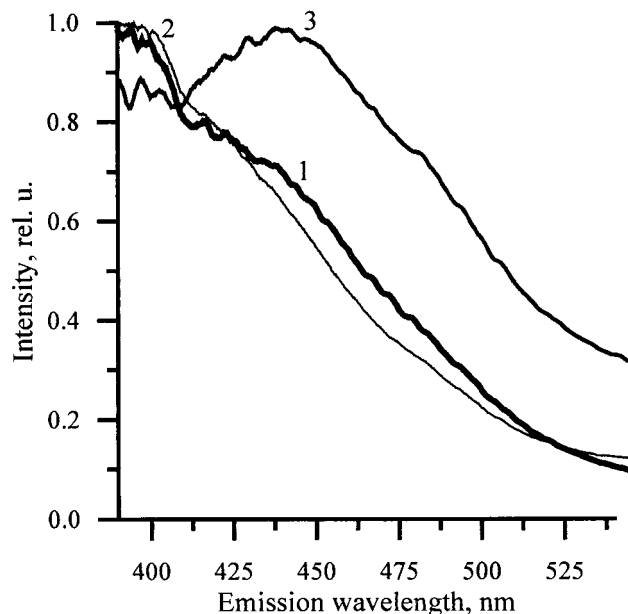
ple of the spectra of base solutions and of their mixture for LO.

B. Fluorescence Spectra of Water, Hexane Extracts, and Hexane Solutions

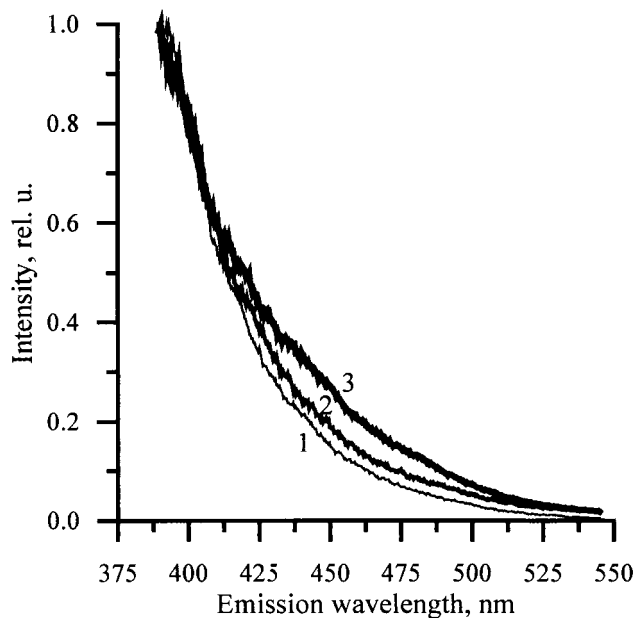
As was mentioned above, to solve the problem of separating out the contribution from OP to the fluorescence band of a marine environment one should know the fluorescence band shapes of OP and of the background. Information on the background, which is formed mainly by the AHS, was obtained from the analysis of natural seawater samples (see Section 3). Information on the properties of the fluorescence band of OP can be obtained only from model experiments.

The shape of the spectrum of the hexane extract can be used at least as a crude indicator for the identification of OP. We verified this by comparing the fluorescence bands from two types of OP (DF and HO) in water, in hexane extract, and in hexane solution. The results are presented in Fig. 4. We conclude that the band shape of OP in water retains the features of the original oil product and correlates well with the hexane-extract band shape, which, in turn, coincides with the fluorescence band shape of the solution of the given oil product in hexane. The most likely reason for this property is that the main contribution to the fluorescence of the oil product in water comes from the oil emulsion, which appears to retain, to a significant extent, the properties of the original OP (as distinct from the dissolved fraction).²⁷ Thus a crude identification of OP from the spectrum of the hexane extract is possible, from which one can choose the most likely fluorescence band of the given OP in water and use it as a model in ANN training.

Of course, the method of oil identification *in situ*



(a)



(b)

Fig. 4. Fluorescence spectra of (a) HO and (b) DF in 1, hexane; 2, distilled water; and 3, a hexane extract from water. Excitation wavelength, 337 nm. Spectra were measured with the laser spectrometer.

(without extraction) is preferable. Some hope for the solution of the *in situ* identification problem is given by the results of Ref. 28, where the possibility of such identification by use of four or more (as many as ten) specially selected spectral channels was demonstrated.

C. Results of Numerical Experiments

To determine the possibility of separating the small amount of fluorescence from one substance against

the background formed by another substance by using artificial neural networks, we performed several numerical experiments.

A neural network that is capable of separating small components from background requires a representative set of data for training. The set should include the fluorescence spectra of mixtures that contain both components, over the entire range of variation of the fluorescence parameters Φ_0 to correspond to the concentration range of interest of each component.

There are two fundamentally different methodological approaches to working with ANNs, one from experiment and the other from a model.

In the first, experimental fluorescence spectra of the mixtures are used to train an ANN. The disadvantage of this approach is the necessity for having a large number of quality experimental spectra with the *a priori* known values of parameter Φ_0 for the components of the mixtures over the entire concentration range. The spectra must be recorded in identical conditions and by use of the same experimental setup, conditions that are difficult to ensure in practice.

The second approach is based on the assumption that one can use modeled spectra for training. Construction of such a model is rather difficult, because of the complexity of the object and because it has been little studied so far.

For these reasons we decided for the present study to adopt an intermediate approach, based on so-called indirect simulation. With this approach the model fluorescence spectrum of a mixture was calculated as a linear combination of the base experimental fluorescence spectra of the solutions, each of which contains only one of the components under study.

As the base spectra for indirect simulation we used laboratory data from aqueous solutions of LO and FA (see Subsection 4.A). The shape thus formed of the spectrum of the mixture describes a linear model, which neglects any influences exerted by the spectra of the components (water, LO, FA) on one another. The Raman scattering band of the mixture is the same as that of purified water, and the fluorescence is a linear superposition of the fluorescence bands of LO and FA with weight coefficients equal to Φ_0^{LO} and Φ_0^{FA} , respectively.

Using this linear model, we calculated the training, test, and examination sets with the following variation ranges: Φ_0^{FA} , 0.01–20.0; Φ_0^{LO} , 0.01–20.0. This range of fluorescence parameter Φ_0^{LO} corresponds to a LO concentration range of 0.05–100 $\mu\text{g/L}$. Various architectures of neural networks were trained. In this way a versatile ANN was prepared that was suitable for analysis not only of the fluorescence spectra of coastal waters (with $\Phi_0^{\text{OP}} \leq 0.01$ Φ_0^{AHS}) but also of spectra for which the ratio of the fluorescence parameters for OP and AHS in a wider range, corresponding to any seawater.

As the input data for the ANN the intensities in the corrected fluorescence spectra at 481 wavelengths (i.e., in 481 spectral channels) were used. As the

output data of the ANN the fluorescence parameters Φ_0^{OP} and Φ_0^{AHS} were chosen. From a practical point of view, direct determination of the component concentrations is more attractive. However, making such a determination would require special attention to calibration procedures. Here we shall merely note that in the range of Φ_0^{OP} and Φ_0^{AHS} values that we used the relation between concentration and Φ_0 is linear with high precision. Thus, in our approach, we operate with Φ_0 parameters, which are more widely used and are evident from a spectroscopic point of view. The procedure for calibration of their concentration is beyond the scope of this paper, though possibly it somewhat reduces the precision of the final result, which is the determination of concentration.

All experiments with an ANN were conducted with the software package NeuroShell 2 from Ward Systems Group, Inc. Three numerical experiments were performed.

1. Numerical Experiment 1

We trained the ANN on model (simulated) spectra of a mixture of FA and LO and applied them to spectra, which were also model (simulated) ones, with various noise levels. To train and examine a network we used a set of simulated spectra that had been calculated from a linear model of a mixture in which the partial fluorescence parameters Φ_0^{FA} and Φ_0^{LO} varied in the range 0.01–20.

Random noise with an amplitude of 1–10% was added to the spectra that made up the main examination set to produce additional, noise-degraded examination sets. This procedure was repeated 10 times for each spectrum of the main set.

Using an ANN, we restored the parameters Φ_0^{FA} and Φ_0^{LO} and determined their mean relative deviations from the true values, denoted as errors in reproducing, ϵ_{FA} and ϵ_{LO} . Several ANN architectures were tested, and the lowest values of ϵ were obtained by use of a five-layer perceptron.¹ The results are illustrated in Fig. 5. Analysis of the results obtained has revealed the following: In a perfect case (in the absence of noise in the input data) the ANN has proved capable of determining the contribution from LO down to a value of $\Phi_0^{\text{LO}} = 0.02$ against a fluorescence background of FA with $\Phi_0^{\text{FA}} = 20.0$. In that case the error in determination of the parameter Φ_0^{LO} was less than 10%. With increasing Φ_0^{LO} the error decreased, so its value, averaged over the entire range of variation of Φ_0^{LO} and Φ_0^{FA} , was $\sim 2\%$ (both for Φ_0^{LO} and Φ_0^{FA}). The minimum detectable value of $\Phi_0^{\text{LO}} = 0.02$ corresponds to a LO concentration of $\sim 0.1 \mu\text{g/l}$.

The degradation of the input data with 3% noise makes the result considerably worse: With a 10% error, Φ_0^{LO} could be determined down to a value of ~ 1 , which nevertheless testifies to a fairly high accuracy and sensitivity of the algorithm used. The influence of noise in the input data can be strongly reduced if the network is trained on noise-degraded spectra (cf. curve 3 with curve 2), as we discuss below

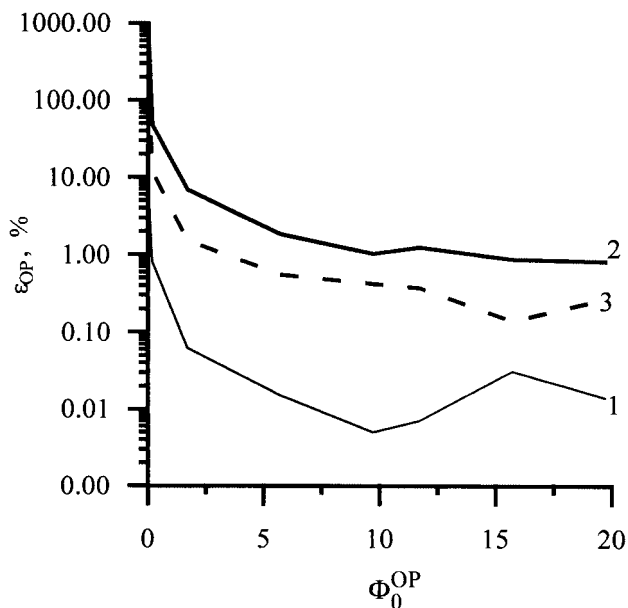


Fig. 5. Mean relative error of the Φ_0^{OP} parameter estimation (ϵ_{OP}) versus true Φ_0^{OP} values ($\Phi_0^{\text{FA}} = 19.7$): 1, noise-free data presented to the network trained on noise-free data; 2, data with 3% noise presented to the network trained on noise-free data; 3, data with 3% noise presented to the network trained on noisy data.

in the context of analyzing the practical stability of solutions of the inverse problem.

It is worth noting that the high sensitivity and accuracy of the method are explained by adoption of the model approach. The network was trained on the basis of the adopted model, which allowed a large training set of data to be formed, and data that completely satisfy the model were presented to the network. Then the sensitivity and accuracy of the method are limited only by the accuracy of the network itself. This is a kind of instrumental limit. Recall that the fluorescence band of OP is rather wide (approximately the same as that of an AHS) and structureless; the bands almost completely overlap (Fig. 3).

2. Numerical Experiment 2

The ANN trained on simulated spectra (according to a linear model) was applied to the experimental spectra of the model samples. In this experiment the best results were obtained with the three-layer perceptron ANN architecture.¹ A set of 85 spectra was presented to the network, in which the parameter Φ_0

varied within the ranges $\Phi_0^{\text{LO}} = 0-8.86$ and $\Phi_0^{\text{FA}} = 0-23.3$.

A comparison of the results of the three numerical experiments is given in Table 1. As one can see from the table, error ϵ_{LO} increases by more than an order of magnitude when an ANN trained on the linear model is applied to the experimental spectra. This means that the fluorescence spectrum of a real mixture does not completely correspond to the model for which the network was trained. There may be several reasons for this: large errors present in the input data, violation of linearity because of interactions between the components, the influence of organic compounds on the Raman band of water, etc.

3. Numerical Experiment 3

An ANN was trained on experimental spectra of model samples and applied to spectra of the same kind. The way in which this experiment differs from experiments 1 and 2 is that here the approach from experiment 1 is adopted, because the ANN was trained by use of an array of actual (rather than simulated) spectra. Accordingly, the results of solving the problem should improve, because the effects of aging and the interaction between the components are automatically taken into account. However, some new difficulties arise: One should have a reasonably large array of training spectra with known Φ_0 values of the components, which uniformly cover the range of interest of their variation. In this experiment, that requirement was not met to the needed extent: To train the network we used 54 spectra, for which the values of Φ_0 varied in the ranges $\Phi_0^{\text{LO}} = 0-8.86$ and $\Phi_0^{\text{FA}} = 0-23.3$. As the examination array, only seven spectra were available. Therefore, as can be seen from Table 1, the accuracy of restoring parameter Φ_0^{LO} , although it is improved compared with that in experiment 2, failed by far to reach the level attained in experiment 1.

Noting the reasons for the considerable increase in errors in experiments 2 and 3 compared with experiment 1, we have not indicated one source of error that is common to them all and is associated with errors in preparing the mixtures or, to put it more exactly, in determining the true values of parameters Φ_0^{LO} and Φ_0^{FA} in these mixtures. However, in experiments 2 and 3 (unlike in experiment 1), error ϵ depends substantially on how accurately the true (expected) values of parameters Φ_0 coincide with those

Table 1. Results of the Numerical Experiments

Architecture	Experiment 1 ^a		Experiment 2 ^b		Experiment 3 ^c	
	ϵ_{FA} (%)	ϵ_{LO} (%)	ϵ_{FA} (%)	ϵ_{LO} (%)	ϵ_{FA} (%)	ϵ_{LO} (%)
Three-layer perceptron	2.1	1.7	8.8	25.7	9.6	12.8
Five-layer perceptron	0.1	0.2	8.5	28.9	12.2	14.6

^aNumerical experiment 1: ANN trained on the linear model; applied to a set of calculated spectra.

^bNumerical experiment 2: ANN trained on the linear model; applied to a set of experimental spectra.

^cNumerical experiment 3: ANN trained on experimental spectra; applied to a set of experimental spectra.

that actually exist in the real mixture at the moment of measurement.

Thus the error in experiment 2 depends on the accuracy of preparation of the mixtures and on the changes of the shapes of the fluorescence bands of the components compared with those in base solutions at the moment when the spectra of the mixtures are recorded, i.e., on the difference of the real spectra from the model ones or, which is the same, on the difference of the model from the real situation. The sensitivity of the result to such changes, as well as to noise in the input data, is determined by the practical stability, which is ensured by the character of the inverse problem that is being solved and by the algorithm that is used to solve it.

D. Stability of the Solution of the Inverse Problem

The problem of determining the contribution from OP to the total fluorescence band of seawater belongs to the class of inverse problems. One of the criteria of such problems' being well posed on a chosen manifold of solutions²⁹ is the stability of the solution on this manifold. The theoretical stability of our problem is ensured by the limited number of parameters to be determined. However, a theoretically stable solution may become unstable at certain finite (rather than small) deviations of input data from an exact dependence that describes the chosen model of the process. In this case the solution may significantly deviate from the true solution. Such instability is referred to as practical instability.²⁹

Here we consider successively the stability of the solution with respect to the noise level in the input data and to the variation of the model parameters. This methodology follows that described in Ref. 30 as applied to the inverse problem of nonlinear fluorimetry.

E. Practical Stability of the Solution versus Input Data Noise

As was mentioned above, noise in the spectra presented to the ANN leads to a sharp increase in the error in determining parameter Φ_0^{LO} . By training the ANN on noise-degraded spectra one can reduce this error to some extent. This is a manifestation of the practical instability but also an illustration of one of the methods of overcoming this problem. Figure 6 illustrates this method in more detail. As can be seen, for the chosen mixture of LO and FA the error in determining Φ_0^{LO} by the network trained on noise-free spectra is more than 10% with a 6–8% noise level in the input data. This value seems acceptable from a practical standpoint. It can be accepted as a conventional threshold of practical instability. ANN training on noise-degraded data considerably raises this threshold of practical instability (Fig. 6).

F. Practical Stability of the Solution versus Variable Model Parameters

As elements of the model whose change may lead to instability of the solution, the width and relative position, i.e., the distance separating their maxima

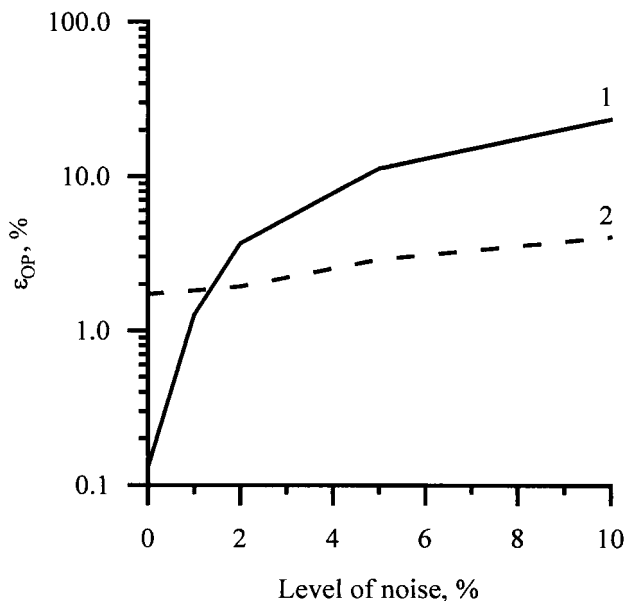


Fig. 6. Mean relative error of the Φ_0^{OP} parameter estimation (ϵ_{OP}) versus noise level in the input data ($\Phi_0^{OP} = 1.7$, $\Phi_0^{FA} = 9.7$): 1, network trained on noise-free data; 2, network trained on noisy data.

$\Delta\lambda_{max}$, of the OP and AHS fluorescence bands are considered.

It is stated in Section 3 that the width of the AHS band in Black Sea coastal waters, where the field studies were made, varied by $\pm 5\%$ with respect to the average value of 95 nm. Thus it is necessary to elucidate whether such changes in the width of the FA (and the AHS) fluorescence band could lead to practical instability.

The check was made as follows: The experimental Raman spectrum of purified water was subtracted from the base experimental FA spectrum, which allowed its fluorescence spectrum to be derived in pure form. The shape of this pure spectrum was approximated with an analytical polynomial model by use of the Group Method of Data Handling (GMDH).³¹ Next, we deformed the band shape by multiplying the argument of the found functional dependence by a variable factor such that its width at half-height, $\Delta\lambda$, changed while the found functional dependence on the wavelength was retained.

Based on each of the deformed spectra with a step of relative change in width of $\delta(\Delta\lambda)/\Delta\lambda_0 = 0.05$, where $\Delta\lambda_0$ is the average width, an examination set of 48 items was calculated that was analogous to the examination set for the starting problem, except for the use of the deformed spectrum instead of the base experimental spectrum. To each examination set we applied the ANN and calculated the average relative errors ϵ . Note that the widths of the original (undeformed) spectra of FA and LO were $\Delta\lambda_{FA} = 125$ nm and $\Delta\lambda_{LO} = 86$ nm; the maximum of the latter was shifted to the short-wavelength side from the FA maximum by $\Delta\lambda_{max}^0 = 60$ nm (Fig. 3).

Figure 7 illustrates the dependence of ϵ on the

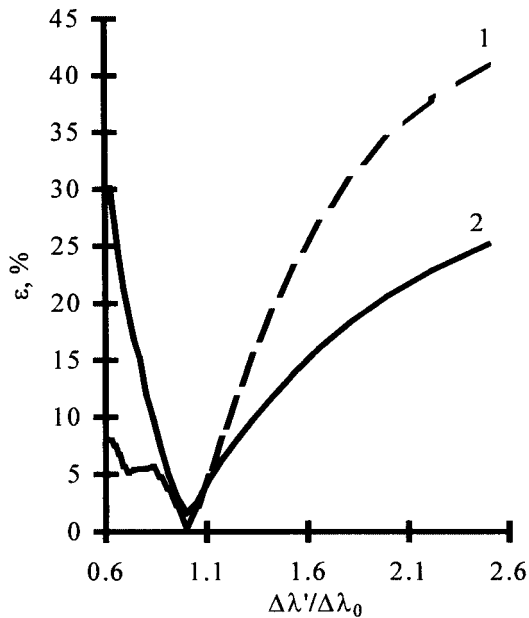


Fig. 7. Dependence of mean relative errors of 1, the Φ_0^{OP} parameter estimation ε_{OP} and 2, the Φ_0^{FA} parameter estimation ε_{FA} on relative bandwidth $\Delta\lambda'/\Delta\lambda_0$.

examination set as a function of the relative bandwidth $[\Delta\lambda'/\Delta\lambda_0, \text{ where } \Delta\lambda' = \Delta\lambda_0 + \delta(\Delta\lambda)]$. The very high stability of the ANN to the change in the width of the FA spectrum is a remarkable feature. The conditional threshold of practical instability in changing the FA bandwidth (when errors of the Φ_0^{LO} and Φ_0^{FA} estimation increase from 1% to $\sim 10\%$) constitutes 20–30% of the change in the width, which is much higher than the experimentally registered variations in this parameter. A check of stability of the solution to the change in the LO bandwidth was carried out similarly to that described above. The results are both qualitatively and quantitatively close to those presented in Fig. 7.

The stability of the solution to the change of another model parameter, the distance between the maxima of the OP and AHS fluorescence bands, $\Delta\lambda_{\text{max}}$, was also investigated. We did this by presenting to the network trained with the original model (see numerical experiments 1 and 2) the spectral bands of mixtures simulated as in numerical experiment 1 but with different values of $\Delta\lambda_{\text{max}} = \lambda_{\text{max}}^{\text{FA}} - \lambda_{\text{max}}^{\text{LO}}$. The value $\Delta\lambda_{\text{max}}^0 = 60 \text{ nm}$, which corresponds to the original bands, was increased by 15 nm and decreased by 60 nm (to $\Delta\lambda_{\text{max}} = 0$). The results are presented in Fig. 8. As can be seen, the increase in the distance between the band maxima within the indicated limits did not result in any substantial increase in ε : The errors with which both parameters were determined did not exceed 7%. However, the problem obviously became more complicated when the fluorescence bands of the components approached each other, leading to decreased accuracy. The relative error ε of Φ_0^{FA} and Φ_0^{LO} estimations increased to $\sim 12.5\%$ as the bands came closer by 10 and 30 nm, respectively.

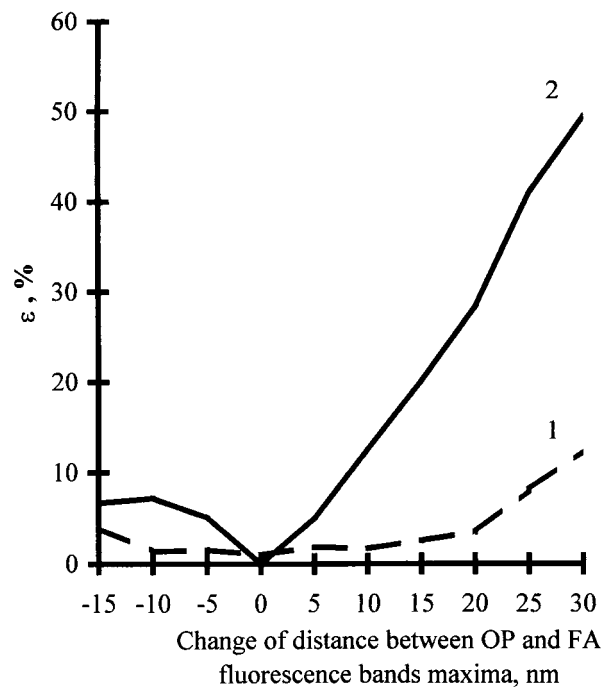


Fig. 8. Dependence of mean relative errors 1, ε_{OP} and 2, ε_{FA} on the change of distance between OP and FA band maxima $\Delta\lambda_{\text{max}}$. This change is equal to 0 for $\Delta\lambda_{\text{max}} = 60 \text{ nm}$ ($\Delta\lambda_{\text{max}} = 60 \text{ nm}$ was used during ANN training).

Thus the practical stability of solution of the inverse problem of fluorescence parameter Φ_0^{LO} and Φ_0^{FA} estimations of the model parameters has proved to be unexpectedly high. This stability significantly lowers the requirements for the accuracy of the *a priori* information on the width and the relative position of the bands of the components.

The observed high stability of ANN responses might be explained by the properties of the ANN as an algorithm of data processing. First, to solve the inverse problem the ANN makes use of a multitude of characteristics, identified in the course of training, rather than of a few evident and easily formulated notions (the position of a spectrum, its half-width, etc.). Second, when it is trained, the ANN assimilates information that is contained not only in a single curve under study but in the entire training data set. In the process, it singles out the essential and reproducible information but discards the unessential and noise-related changes in the input data. Such behavior advantageously distinguishes the ANN-based approach from those based on other algorithms.

At the same time it can be expected that, as a consequence of the properties of the ANN described above, ensuring high stability of the solution to the change in $\Delta\lambda$ and $\Delta\lambda_{\text{max}}$, there would be as high a sensitivity (i.e. a low stability) of the solution to a change in the functional dependence on $\Delta\lambda$ that describes the shape of the fluorescence band of each component. These expectations have been confirmed by a numerical experiment, as follows: To a

network trained on the bands of mixtures of FA and LO (as in numerical experiments 1 and 2), fluorescence bands of mixtures of FA and DF were presented (Fig. 3). The values of the DF half-width and the distance between the maxima lay in the stability range for the mixture of FA and LO. The results of this numerical experiment showed that the error of the fluorescence parameter estimations increased by 1.5–2 orders of magnitude. Thus solving the problem requires information on the shape of the fluorescence band of the OP, whereas the width and the position of the OP may vary in a wide range.

An investigation of the variability of the functional dependence that describes the shape of the component bands (at fixed bandwidths) and of the practical stability of the inverse problem solution to changes in this shape becomes a key task in the problem considered. It is also important from the instrumental point of view: Depending on that solution it may be possible to determine the minimum number of spectral channels of the spectrometer receiver that will provide an acceptable precision in determination of the Φ_0 parameters of the components instead requiring that the quasi-continuous spectra be measured. Such precision could determine the possibilities of simplifying future instruments designed for detecting oil in water.

This issue will be the subject of further studies. *A priori* it could be said that small changes in the shape of the AHS band (Fig. 2) most probably will not cause any difficulties, unlike possible changes in the shape of the OP band, which need to be studied in more detail than has yet been done (from the point of view of quantifying OP in coastal waters).

5. Field Verification of the Method: Determination of the Contribution of Oil Pollution to the Fluorescence Band of Coastal Seawater

A field verification of fluorescence spectroscopy coupled with the artificial neural network developed to determine *in situ* the contribution of OP to the fluorescence band of coastal seawater was carried out in the following manner: Three water samples that had a fluorescence band shape sufficiently close to that of the band of a DF solution in hexane were selected. The experimental spectra were similar to those presented in Fig. 1. These spectra were presented to the ANN trained with the spectra that represent a linear superposition of the band of a seawater sample from which OP had been removed by double extraction, and the band of a DF solution–emulsion. The first of the base bands of mixtures to train the ANN was regarded as the fluorescence band of the AHS, although, strictly speaking, this band could also be contributed by other organic impurities that are not extractable from water with hexane. In construction of the total spectra the values of fluorescence parameters Φ_0^{AHS} and Φ_0^{DF} served as weight coefficients. The true value of Φ_0^{AHS} was determined from the spectrum of water samples after OP extraction. The true value of Φ_0^{DF} was found from

Table 2. Results of Application of ANN to Field Experimental Spectra to Determine the Parameters Φ_0^{OP} for which the Suspected OP is DF and $\Phi_0^{\text{AHS}\alpha}$

Sample Number	Φ_0^{AHS}		Φ_0^{OP}	
	Experiment	Network Response	Experiment	Network Response
B32000	1.4	1.3	0.14 ± 0.03	0.10
B34000	1.2	1.2	0.15 ± 0.03	0.13
B21000	1.9	1.6	0.26 ± 0.03	0.21

^aThe seawater was sampled from a depth of 0.5 m in the Golubaya Bay area.

the formula $\Phi_0^{\text{DF}} = C/a$, where C is the DF concentration in water as measured by the extraction technique^{8,15} and a is the coupling coefficient obtained in a calibrating experiment. The ANN was trained in a considerably narrower range of variation of the parameters Φ_0^{AHS} (6–13) and Φ_0^{DF} (0.2–2) than in the numerical experiment. The above ranges corresponded to the scatter of the experimental values of parameters Φ_0^{AHS} and Φ_0^{DF} obtained in Golubaya Bay during the measurement period. The results are presented in Table 2. (The errors in determining experimental values of Φ_0^{OP} that are listed in the table are associated with those of the extraction procedure, which was separately determined.) As can be seen, these results, to which the following circumstances contributed, are quite satisfactory: The samples were taken from Golubaya Bay near the pier, where the research vessel *Akvanavt* and other small-tonnage boats were moored. To train the ANN we used the fluorescence band of DF taken from those vessels.

In general, a water sample may contain some OP for which there is no such definite information about the fluorescence spectrum as for the samples in this study or may even contain a mixture of several OPs. Therefore, in real situations, the procedure for determining Φ_0^{OP} proposed in this paper should be preceded by identification of the OP. Bearing in mind the *in situ* diagnostics (without using extraction), one can do this by using principal component analysis²⁸ or analysis of total luminescence spectra with help of an ANN.

6. Conclusions

The studies presented in this paper give evidence that there is a real possibility of determining the concentration of oil pollution in coastal seawater *in situ* with a precision of micrograms per liter against the fluorescence background of aquatic humic substance with a concentration of milligrams per liter by using an artificial neural network to solve the inverse problem. Numerical and experimental simulation revealed a unique ability of the ANN technique to solve the problem at hand. At the same time a number of problems were brought to light that need to be solved to bring the method to the status of a practical tool. Field verification of the method in coastal areas of the Black Sea confirmed, on the whole, the

results of the numerical model experiments. The results led to an optimistic perspective regarding the development of a truly effective method for *in situ* OP determinations with high sensitivity and accurate results even in coastal seawater.

The next stage in improving the method should involve a detailed study of the photophysical and hydrochemical processes that are responsible for AHS and OP bands in natural conditions; the elaboration of a more-precise method for identifying the types of OP present in water and the selection on this basis of some base spectra to train the ANN; and studies of the possibility of reducing the number of spectral channels and of other ways that can lead to simplification (and therefore to cost reduction) of the laser (or LED) spectrometer, and so to the opportunity for its wide use in systems for monitoring of coastal water areas.

This study was performed with support from the International Association for Cooperation with Scientists from the former Soviet Union (project 96-2063), NATO Collaborative Linkage Grant EST. CLG.978263, and the federal special-purpose programs "Integration" (projects 475 and Center of Fundamental Optics and Spectroscopy), Fundamental Spectroscopy, and World Ocean (project 6.10).

The authors acknowledge support from the administration of the Southern Branch of the P. P. Shirshov Oceanology Institute (Gelendjik) and from the crew of the R/V *Akvanavt* in conducting the nature experiments.

References

1. R. P. Lippman, "An introduction to computing with neural nets," *IEEE Trans. Acoust. Speech Signal Process.* **4**(2), 4–22 (1987).
2. D. Specht, "A general regression neural network," *IEEE Trans. Neural Netw.* **2**, 568–589 (1991).
3. J. Lakovich, *Principles of Fluorescence Spectroscopy* (Plenum, New York, 1986).
4. A. W. Hornig, "Identification, estimation and monitoring of petroleum in marine waters by luminescence methods," in *Marine Pollution Monitoring*, NBS Spec. Publ. **409**, 135–144 (1974).
5. R. M. Measures, *Laser Remote Sensing. Fundamentals and Applications* (Wiley, New York, 1984).
6. R. A. Velapoldi and K. D. Mielenz, *A Fluorescence Standard Reference Material: Quinine Sulfate Dihydrate*, NBS Spec. Publ. SP-260 64 (National Bureau of Standards, Washington, DC, 1980).
7. D. V. Maslov, V. V. Fadeev, and A. I. Lyashenko, "A shore-based lidar for coastal seawater monitoring," in *European Association of Remote Sensing Laboratories (EARSeL) eProceedings (Workshop on LIDAR on Land and Sea)*, R. Reuter, ed. (EARSeL, Paris, 2001), pp. 46–52.
8. E. M. Filippova, V. V. Chubarov, and V. V. Fadeev, "New possibilities of laser fluorescence spectroscopy for diagnostics of petroleum hydrocarbons in natural water," *Can. J. Appl. Spectrosc.* **38**, 139–144 (1993).
9. V. V. Fadeev, T. A. Dolenko, E. M. Filippova, and V. V. Chubarov, "Saturation spectroscopy as a method for determining the photophysical parameters of complicated organic compounds," *Opt. Commun.* **166**, 25–33 (1999).
10. S. Determann, R. Reuter, P. Wagner, and R. Willkomm, "Fluorescent matter in the eastern Atlantic Ocean. 1. Method of measurement and near-surface distribution," *Deep-Sea Res. I* **41**, 659–675 (1994).
11. S. Determann, R. Reuter, and R. Willkomm, "Fluorescent matter in the eastern Atlantic Ocean. 2. Vertical profiles, and relation to water masses," *Deep-Sea Res. I* **43**, 345–360 (1996).
12. B. Nieke, R. Reuter, R. Heuermann, H. Wang, M. Babin, and J. C. Therriault, "Light absorption and fluorescence properties of chromophoric dissolved organic matter (CDOM) in the St. Lawrence Estuary (Case 2 waters)," *Cont. Shelf Res.* **17**, 235–252 (1997).
13. D. N. Klyshko and V. V. Fadeev, "Remote determination of the admixture concentrations in water the method of laser spectroscopy using Raman scattering as an internal standard," *Sov. Phys. Dokl.* **23**, 55–57 (1978).
14. V. V. Fadeev, "Possibilities of standardization of normalized fluorescent parameter as a measure of organic admixtures concentration in water and atmosphere," in *Environmental Sensing and Applications*, M. Carleer, M. Hilton, T. Lamp, R. Reuter, G. M. Russwurm, K. Schafer, K. Weber, K. Weitkamp, J.-P. Wolf, and L. Woppowa, eds., *Proc. SPIE* **3821**, 458–466 (1999).
15. Intergovernmental Oceanographic Commission/United Nations Environmental Programme, *Manual for Monitoring Oil and Dissolved/Dispersed Petroleum Hydrocarbons in Marine Waters and on Beaches*, Manuals and Guides No. 13 UN Educational, Scientific, and Cultural Organization, Paris, 1984).
16. P. Coble, S. A. Green, N. V. Blough, and R. B. Gagosian, "Characterization of dissolved organic matter in the Black Sea by fluorescence spectroscopy," *Nature* **348**, 432–435 (1990).
17. R. F. Chen and J. L. Bada, "The fluorescence of dissolved organic matter in seawater," *Mar. Chem.* **37**, 191–221 (1992).
18. K. Mopper and C. A. Schultz, "Fluorescence as a possible tool for studying the nature and water column distribution of DOC components," *Mar. Chem.* **41**, 229–238 (1993).
19. M. M. De Souza Sierra, O. F. X. Donard, M. Lamotte, C. Belin, and M. Ewald, "Fluorescence spectroscopy of coastal and marine waters," *Mar. Chem.* **47**, 127–144 (1994).
20. P. G. Coble, "Characterization of marine and terrestrial DOM in seawater using excitation–emission matrix spectroscopy," *Mar. Chem.* **51**, 325–346 (1996).
21. S. Determann, J. Lobbes, R. Reuter, and J. Rullkötter, "UV fluorescence excitation and emission spectroscopy of marine algae and bacteria," *Mar. Chem.* **62**, 137–156 (1998).
22. A. I. Simonov and V. I. Mikhailov, "Chemical pollution of thin surface layer of the World Ocean," *Tr. Gos. Okeanogr. Inst.* **149**, 5–16 (1979).
23. S. Babichenko, L. Poryvkina, and S. Kaitala, "Multiple-wavelength remote sensing of phytoplankton," *EARSeL Adv. Remote Sens.* **3**, 78–83 (1995).
24. K. Hennig, T. de Vries, R. Paetzold, K. Jantos, E. Voss, and A. Anders, "Multi sensor system for fast analyses in environmental monitoring with application in waste water treatment," in *European Association of Remote Sensing Laboratories (EARSeL) eProceedings*, R. Reuter, ed. (EARSeL, Paris, 2001), Vol. 1, pp. 61–67.
25. S. A. Dolenko, T. A. Dolenko, V. V. Fadeev, E. M. Filippova, O. V. Kozyreva, and I. G. Persiantsev, "Solution of inverse problem in nonlinear laser fluorimetry of organic compounds with the use of artificial neural networks," *Pattern Recognition Image Anal.* **9**, 510–515 (1999).
26. E. M. Filippova, V. V. Fadeev, and V. V. Chubarov, "The origin and structure of fluorescence band from aquatic humic substances," in *5th International Conference on Laser Application in Life Sciences*, P. A. Apanasevich, N. I. Koroteev,

- S. G. Kruglik, and V. N. Zadkov, eds., Proc. SPIE **2370**, 651–655 (1994).
27. A. G. Abroskin, S. E. Nol'de, V. V. Fadeev, and V. V. Chubarov, "Laser fluorimetry determination of emulsified-dissolved oil in water," Sov. Phys. Dokl. **33**, 215–217 (1988).
28. T. Hengstermann and R. Reuter, "Laser remote sensing of pollution of the sea: a quantitative approach," *EARSeL Adv. Remote Sens.* **1**, 52–60 (1992).
29. A. N. Tikhonov and V. Ya. Arsenin, *Methods of Solving Ill-Posed Problems*, Scripta Series in Mathematics (Scripta Mathematica, New York, 1977).
30. I. V. Boychuk, T. A. Dolenko, A. R. Sabirov, V. V. Fadeev, and E. M. Filippova, "Study of the uniqueness and stability of the solution of inverse problem in saturation fluorimetry," *Quantum Electron.* **30**, 611–616 (2000).
31. H. R. Madala and A. G. Ivakhnenko, *Inductive Learning Algorithms for Complex Systems Modeling* (CRC Press, Boca Raton, Fla., 1994).