METHODS

Structural water self-organization in river flow

O. M. Rozental[1] and V. Kh. Fedotov[2]*

[1]Water Problems Institute, Russian Academy of Sciences, Moscow, Russia
[2]Chuvash State University, Cheboksary, Russia

*Correspondence:
V. Kh. Fedotov,
fvh@inbox.ru

Application of neural-network methods revealed relationships between hydrological and hydrochemical characteristics of water flow, suggesting structural self-organization of substances dissolved in water in the form of micro layering. In particular, the coefficient of correlation between the concentrations of such substances in some cases reaches its nearly maximal value (0.99), combining with the high weights of neural network edges. This can be supposed to be due to the mechanical and chemical interactions in river flow with the participation of Van der Waals forces, hydration, and sorption. Other factors, not taken into account, can also have their effect, in particular those responsible for the fluctuations of the parameters of order, determining the singular contributions to the dynamic characteristics of the non-linear system under consideration. Such can be the cyclic oscillations of the characteristics under control with an amplitude decreasing with a decrease in the intensity of the pollution/self-purification processes in water medium and increasing with an increase in this intensity. The obtained information, in addition to its direct purpose as a means to study the nature and properties of fresh water, is a necessary condition for the effective control of water resource quality and water management activity.

Keywords: artificial neural networks, industrial waste discharges, neural-network analysis, pair and multiple correlations, water quality, water-monitoring site

Introduction

The quasiperiodic variations of river water quality characteristics (1) suggest the structural self-organization taking place in it, i.e., the process of reordering of the parameters of water composition and properties due to internal factors. The importance of further studies into this process is determined not only by scientific and informative objectives but also by the need to more reliably manage the economic activity under the stricter requirements of water use and the maintenance of water-environmental safety.

There is no single and easily comprehensive model of the structural self-organization of impurities in water (2, 3). However, for water management, it is quite enough to answer the question of whether there exist pair or multiple correlations between hydrodynamics and hydrochemical characteristics for any individual water object. This study is focused on the Iset River (a tributary of the Tobol River) in Sverdlovsk oblast territory, where large-scale water consumption and disposal imply the processes of either formation or destruction of mutual correlations between the monitored characteristics at different sites of hydrochemical and hydrodynamic monitoring.

Experimental data

The study was based on the results of monthly monitoring of river water in 1990–2010 performed by the Ural Department of the Federal Agency for Hydrometeorology and Environmental Monitoring.

The selected data have been collected at the following stations (**Figure 1**):

– 5.2 km upstream Ekaterinburg (Palkino Vil. station no. 1),
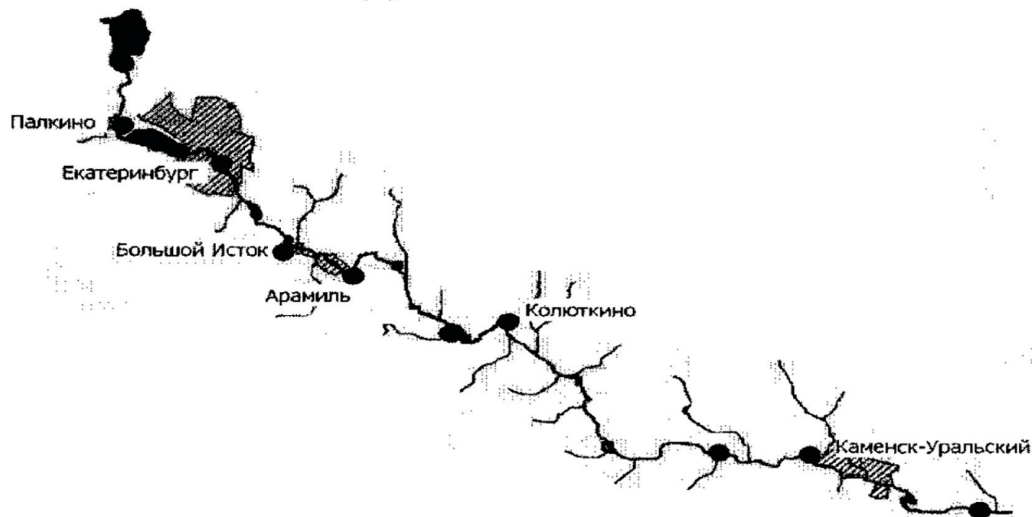– within city boundaries (2),

FIGURE 1 | Hydrographic scheme of the Iset River within Sverdlovsk oblast (the circles show hydromonitoring stations). Translation of geographical names in the diagram: Палкино: Palkino; Екатеринбург: Ekaterinburg; Большой: Bol'shoi Istok; Арамиль: Aramil; Колюткино: Kolyutkino; Каменск: Kamensk-Ural'skii.

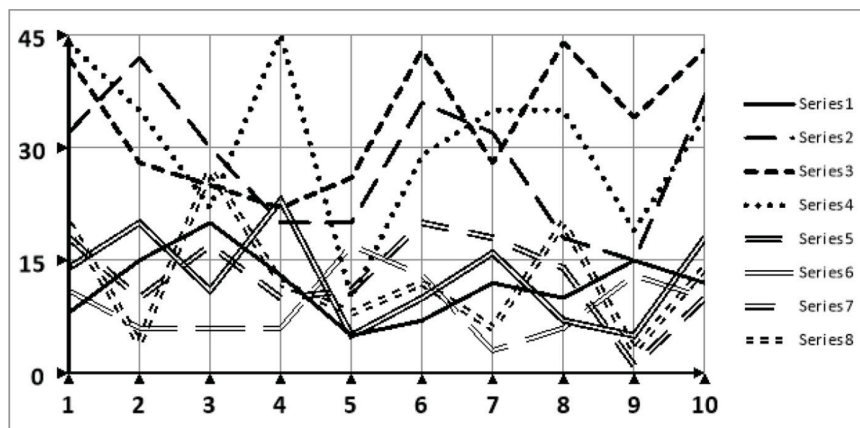**FIGURE 2 |** Copper concentration in the first 10 months of 2008 at stations with numbers corresponding to those of data series.
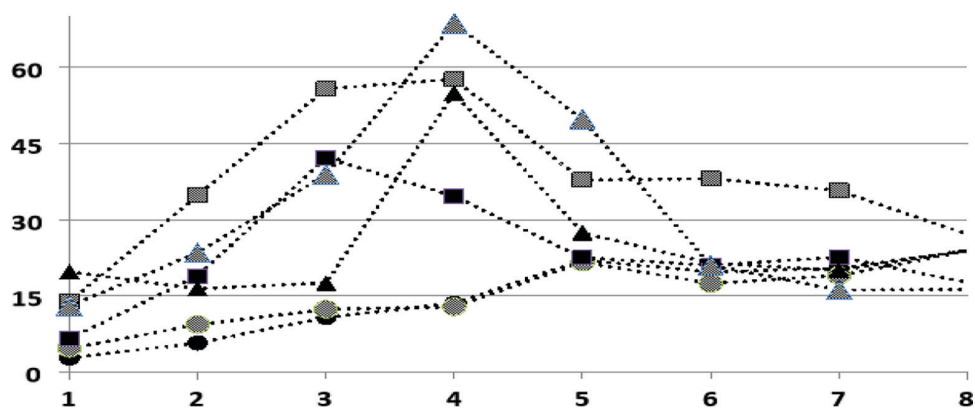

FIGURE 3 | Mean copper concentration (black triangular markers), suspended matter concentration (squares), and water discharge (circles), as well as the root-mean-square deviations of these characteristics, given by appropriate gray markers at each of the eight observation points.

- 7 km downstream (Bol'shoi Istok Settl., 3),
- 19.1 km downstream (Aramil Town, 4),
- further downstream (Kolyutkino Vil., 5),
- 21.3 km upstream of Kamensk-Ural'skii C. (6),
- 5.3 km upstream of this city (7), and
- 9.3 km downstream (8).

TABLE 1 | A list of the neural networks studied for the observation sites (a fragment).

Site	Type	Inputs	Hidden	Output	TError	VError	TeError	Training
1	Linear	3	–	4	1.03	1.08	1.24	PI
1	MLP	1	13	4	1.05	1.05	1.25	BP16b
1	RBF	3	5	4	0.65	0.74	0.82	KM,KN,PI
2	Linear	3	–	4	2.15	3.48	2.14	PI
2	MLP	1	58	4	2.20	3.53	2.14	BP50,CG2b
2	RBF	2	4	4	1.01	1.02	0.95	KM,KN,PI
3	Linear	3	–	4	2.35	1.61	1.65	PI
3	MLP	3	8	4	8.23	7.86	7.88	BP0b
3	RBF	3	5	4	1.32	0.95	1.15	KM,KN,PI
4	Linear	3	–	4	8.23	13.58	42.29	PI
4	MLP	3	4	4	4.34	14.30	15.00	BP0b
4	RBF	3	13	4	1.94	5.36	2.66	KM,KN,PI
5	Linear	3	–	4	1.96	2.03	3.30	PI
5	MLP	1	8	4	5.39	5.06	5.85	BP0b
5	RBF	3	20	4	0.87	1.45	2.57	KM,KN,PI
6	Linear	3	–	4	1.32	1.12	2.44	PI
6	MLP	1	12	4	4.34	4.30	5.00	BP0b
6	RBF	2	2	4	0.90	0.88	1.47	KM,KN,PI
7	Linear	3	–	4	2.19	2.15	2.40	PI
7	MLP	1	25	4	2.27	2.06	2.36	BP6b
7	RBF	3	2	4	1.03	0.81	1.10	KM,KN,PI
8	Linear	3	–	4	2.19	1.93	1.93	PI
8	MLP	1	88	4	2.82	2.48	2.53	BP0b
8	RBF	3	6	4	0.92	0.87	0.71	KM,KN,PI

Denotations of the columns and networks. Type is network architecture type (Lin, linear; MLP, multilayer perceptron; RBF, radial; Inputs: number of neurons at the network entry; Hidden, number of neurons in the intermediate (hidden) network layer; Output, number of neurons at the exit from the network; TError, error on the learning set; VError, error on the verification set; TeError, error on the test set; Training, training algorithms (PI, pseudo invert; BP, back propagation). For example, MLP «1–13–4» means a three-layer MLP-network, including 1 input neuron, 13 hidden neurons, and 4 output neurons.

The geochemical character of the provinces where the river's channel lies causes a higher natural pollution of river water. For example, the mean concentrations of iron, copper, nickel, and zinc that are standardized by their maximal allowable concentrations reach 10 units and more at all stations. At the same time, the chaotic character of the time series, which is partly due to wastewater discharges, hampers the identification of regularities in the variations of the monitored characteristics. As a result, the only conclusion that can be derived, for example, from **Figure 2** is that at all hydromonitoring stations in the industrial region under consideration, the concentration of copper compounds is far in excess of the maximal allowable value for water bodies used in the economy, one of which is the Iset River.

It is more difficult to identify any regularities in large data bodies for water-polluting substances over longer time intervals. As can also be seen from **Figure 3**, variations in the arithmetic mean concentration of impurities at observation sites in some cases disagree with the values of the root-mean-square deviation of the respective characteristics. This indicates the heteroscedasticity of the system, which is due to the heterogeneity and non-stationarity of the data series. That is why numerical technologies of intellectual analysis are required to achieve the goal.

Neural-network method of study

The numerical analysis of data was based on artificial neural networks (ANN) in the specialized neural-network software package Statistica Neural Networks (SNN) (4). For the detailed study of the available time series, neural regressions were constructed with the use of networks with four types of architecture (topologies): linear two-layer networks (Linear), non-linear three-layer and four-layer perceptrons [multi-layer perceptron (MLP)], non-linear networks with a radial basis function [radial basis function (RBF)], and non-linear generalized regressive networks [generalized regressive neural network (GRNN)].

The data were prepared by separating each time series into subsets:

- *Training*, the major portion, not less than 50%,
- *Verification*, confirming the result by its comparison with experimental data, up to 25%,
- *Test*, showing how well the system can analyze new data, up to 25%.

Such a distribution was made taking into account the structure and completeness of the available source data (up to 250 for each characteristic involved in each observation section).

The cause-and-effect relationships were studied by observing the effect of three input (Input) characteristics [water discharge (m^3/s), water temperature (°C), and suspended matter concentration (mg/dm^3)] on four output (Output) water chemistry characteristics (the concentrations of iron, copper, zinc, and nickel, all in mg/dm^3).

The network architectures were formed by superimposing experimental data arrays, considering their preparation, on all types of network architectures that correspond to the idea of the problem.

Network training was made with the use of only training and verification samples by more than 10 algorithms of linear and non-linear optimization, including back propagation of error, pseudo invert, and so on.

The test samples intended for forecasting were not used in training. The setting of detailed parameters for these algorithms, such as the number of training epochs, noise, accuracy, etc., was taken as standard by default.

In the estimation of ANN training, an algorithmic characteristic (criterion) of good quality was taken to be the small enough absolute errors of training (TError),

TABLE 2 | Neural regression statistics for linear «3–4» network at site 6.

	Tr. Fe	Ve. Fe	Te. Fe	Tr. Cu	Ve. Cu	Te. Cu	Tr. Zn	Ve. Zn	Te. Zn	Tr. Ni	Ve. Ni	Te. Ni
Data mean	0.47	0.44	0.57	19.70	18.54	26.32	30.26	26.61	31.55	9.84	9.75	9.37
Data S.D.	0.46	0.44	0.73	13.49	12.77	36.23	23.87	17.90	33.16	13.80	9.01	6.96
Error mean	0	0.01	−0.13	0	0.68	−5.96	0	2.35	−0.28	0	−0.34	0.38
Error S.D.	0.44	0.36	0.70	13.28	13.22	35.45	23.45	18.97	31.93	13.71	8.53	6.92
Abs E. mean	0.33	0.25	0.39	9.58	10.31	16.67	15.78	15.33	17.83	7.01	6.31	5.71
S.D. ratio	0.94	0.81	0.96	0.98	1.03	0.97	0.98	1.05	0.96	0.99	0.94	0.99
Correlation	0.32	0.70	0.31	0.17	−0.28	0.27	0.18	−0.21	0.28	0.11	0.46	0.15

Data Mean is the mean concentration at a site. Data Standard Deviation (S.D.) is the standard deviation of the concentration at the site from the value Data Mean. Error Mean is the mean error, i.e., the mean residual difference between the measured and calculated (by neural network) values at the site. Error S.D. is the standard deviation of the source data from those calculated by the neural network. Abs E. Mean is the absolute mean error (the arithmetic mean of the modules of differences between the source and calculated values at the site). S.D. Ratio: Error S.D./Data S.D. is the relative error of neural regression. Correlation is multiple correlation coefficient.

verification (VError), and testing (TeError). The main statistical characteristics of ANN training were Pearson's multiple correlation coefficients and the values of the weight coefficients of neural network links.

The analytical means for network operation estimation included the results of calculating correlation-regression statistics, optimal weights of links, sensitivity analysis (testing the response to small variations of weights), and analysis of *what-if* type (testing the response to small variations of source data). The graphical means included illustrations of network architecture and its response surface.

The issue of the choice of the network architecture that can give an acceptable model for a problem remained open (it has no unique solution), and, in practice, it was solved by the search of variants. Therefore, more than a hundred different acceptable network architectures were examined for each set of data collected at observation stations.

The number of neurons at the entry and in the inner layer of the network was varied.

The number of neurons at the exit of the network (Output) remained constant and equal to the number of metals involved in the study.

The teaching parameters were taken as standard (by default).

Several best networks with minimal training and testing errors were chosen for detailed study and making conclusions. The first approximation was taken in the form of linear networks. Non-linear networks were used to test and improve the obtained neural-network relationships, as shown in **Table 1**.

It was found that, at all hydromonitoring gages, linear and non-linear neural networks, with the use of different training algorithms, showed similar and relatively high quality in revealing possible relationships between network elements, which makes the results of neural network studies given below quite reliable.

TABLE 3 | Weights of the links in linear «3–4» neural network at site 6.

	Fe	Cu	Zn	Ni
Water discharge	0.49	−0.11	−0.05	0.06
Water temperature	0.06	−0.01	−0.06	−0.01
Concentration of suspended particles	−0.13	0.26	0.22	0.02

Results of neural-network study of the data

The results mentioned in the head of this section include:

1. Statistics of linear and non-linear neuro-regressions of the chosen type of network architecture with successive specification of.

   – the number of input characteristics, such as water discharge, suspended particle concentration, and the temperature;
   – the number of intermediate (hidden) neurons; and
   – the numbers of output parameters.

   Thus, the record of the type linear «3–4» means that the network used in the study was a two layer perceptron, containing 3 input neurons (discharge, temperature, and suspended particles), 0 hidden neurons, and 4 output neurons (iron, copper, nickel, and zinc).

2. The weights of the links of linear/non-linear neural networks with specified numbers of input and output characteristics.

Tables 2 and **3** exemplify the results of the study obtained with the use of the data at an observation site.

The value of Error Mean (**Table 2**) was close to zero at all sites, thus demonstrating that the neural network, trained on the results of preliminary measurements, reproduced the source data almost without errors, at least taking into
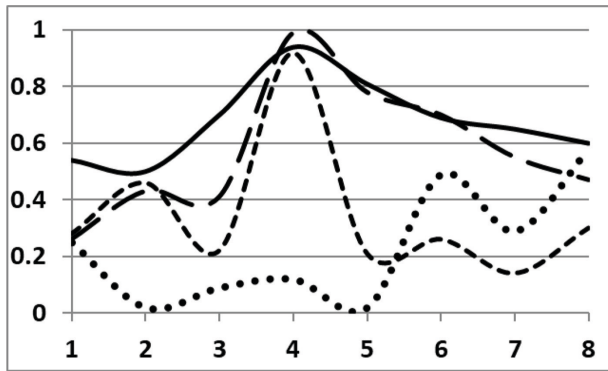
FIGURE 4 | The largest detected values of the coefficients of pair linear correlations for concentration of metals (the full line), neuro correlations of the type «3–4» (long-dash line), and the weights of neuro correlations in the systems of suspended particles-metal (short-dash line), and water discharge-metal (dotted line).

account the measurement errors in the source data. At the same time, the scatter of the calculated values of Error standard deviation (S.D.) was commonly less than the scatter of the original values. In addition, the level of Abs E. Mean remained higher than the Error Mean, i.e., the signs of the differences between the measured and calculated values alternated almost uniformly and, in the calculation of Error Mean, they compensated one another. In addition, the total relative error of neural regression S.D. Ratio was most often less than unit. This was most typical of copper and iron concentrations, such that the standard deviations for these two metals were much less than the standard deviations for the source data, thus suggesting the high accuracy of the neural regressions.

The effect of temperature on metal concentrations was found to be negligible for all sites, as can be seen in **Table 3**.

The key results of studies by neural network method are given in **Figure 4**. One can distinctly see the high correlation of the metals with one another and with the concentration of suspended particles at the fourth site, likely after active discharges of these substances near the observation sites 2 and 3, where many industrial plants-water users are concentrated. The weight of the neural link

between water discharge and metals, i.e., the hydrodynamic and hydrochemical characteristics, also somewhat increases here, although the maximal values of this characteristic were recorded in the zones near sites 6 and 8, which can be naturally attributed to the least effect of the factor of concentration scatter here because of the decrease in the rate of industrial waste discharge in this part of the industrial region.

Overall, **Figure 4** indicates the existence of processes of reordering (self-organization) of water composition and properties in river flow, lower in the zone of intense technogenic impact and higher at its decrease. This can be seen from multiple and pair correlations, which, depending on the level of impact mentioned above, vary from weak (with a regression coefficient of 0.2−0.4) to very strong (0.99), synchronous with the values of weights of neural network edges.

In the use of the ANN, it was taken into account that each layer of the neural network depends not only on the state of the entry neurons but also on an additional (Threshold) neuron, the state of which determines the threshold of activation of the input neuron (4). In this study, the input and output neurons correspond to the dynamic and chemical characteristics incorporated in the model. The threshold neurons do not correspond to any of them; therefore, in the protocols of studying the edge weights of neural networks, they were denoted as factors not taken into account. As can be seen from **Figure 5**, the values of these factors for all observation sites are negative, meaning that they accelerated the activation of the output neuron. As to their absolute values, it was found that, for these factors, they show periodic oscillations. As can be seen from **Figure 5**, such oscillations are relatively low for the concentrations of iron and zinc, and they are higher for nickel after the third observation site. The nature of this phenomenon requires additional studies; however, we can suppose that, for the non-linear system under consideration, i.e., the water matrix, which contains admixture particles, this is a consequence of the dissipation of the energy of microstructure self-organization and regular restructuring of the system, i.e., bifurcation, caused by this restructuring.
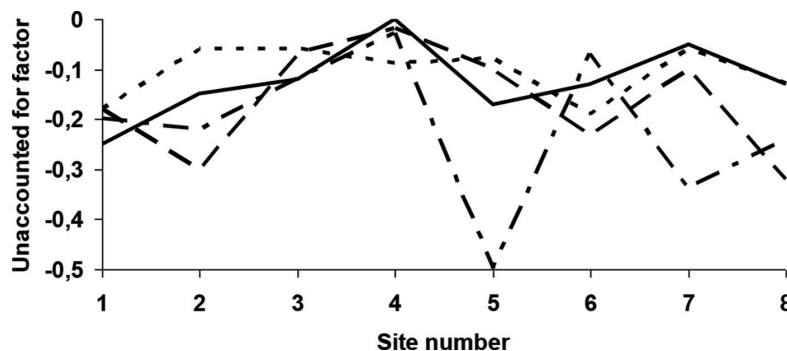

FIGURE 5 | Unaccounted for factors that arise at evaluating neural network weight for the concentrations of iron (full line), copper (long-dash line), nickel (short-dash line), and zinc (dotted line).

Developing in time, this process resembles cyclic oscillations of the characteristics of the composition and properties of natural water with an amplitude decreasing at moderate rates of pollution/self-purification and increasing when these processes accelerate (5).

Conclusion

The problems with recording non-linear correlations between varying chemical and dynamic water characteristics in river flow required the use of intellectual information technologies (neural networks) as a way to generalize classical correlation-regression methods. This made it possible to reveal hidden non-linear relationships between seemingly unordered data arrays without a priori analytical specification of the form of the expected relationships.

Experiments revealed a correlation between the concentrations of dissolved substances and their relationship with water discharges at the observation sites with a narrower scatter of the examined characteristics, which forms under the effect of heavy use of water in the industrial region under consideration. This indicates the structural self-organization of water in the flow and ordering in the form of micro layering of foreign particles.

The self-organization processes can be assumed to proceed under the effect of forces of physical, chemical, and mechanical nature (hydration, sorption, and hydraulics). In particular, the elastic forces caused by deformations of the hydrogen bond grid in water push particles in water toward grid defects, thus forming micro layering, which counteracts the diffusion leveling of the concentration (2). At the same time, the hydrodynamic structure of the water flow, examined by the analysis of the complete system of equations describing the dynamics of fluid flow, revealed its structural elements, such as waves, vortices, and high-gradient layers (6, 7), which contribute to the structural self-organization.

Discussion

The factors that counteract the self-organization described above include the technogenic activity, in particular wastewater discharge and diffusion, which tend to level the concentrations of admixture particles. The effects that appear at the predominance of some tendency, i.e., admixture micro layering or its concentration leveling, can be seen at the joining of water volumes with different compositions and properties:

1. If the differences are not large, the diffusion does not cope with the layering, and, after joining in a single channel, the water flows with different compositions run over a long distance without mixing (8). Similar are examples of river water which spreads over the

ocean surface, forming haloclines in tidal estuaries (9). Also known are other stratifications, for example, at jumps in temperature (thermoclines) or density (pycnocline) (10).

2. Another picture is formed when the minimal admissible concentration difference between the touching water objects with different compositions is high enough for the interphase tension to be higher than the elasticity coefficient of the grid of hydrogen bridges. In such cases, judging by laboratory data (11, 12), turbulent mass transport forms near the phase interface, accompanied by density fluctuations, protuberance formation, opalescence, and an increase in the megahertz dielectric permittivity. Such phenomena should be expected in the zones near sites where high-pollution wastewater is discharged into river flow.

Energy exchange in multicomponent water medium, containing various foreign substances, includes several processes, some of which are slow dissipative (diffusion) and some are relatively fast processes controlled by ion-molecular interactions (5) and hydrodynamic forces in forced flows (6). The dynamics of fluid flows can be described by a system of equations of transport of matter, density, momentum, and total energy–analogs of the conservation laws for isolated systems (13). The high rank of the system of fundamental equations for weakly dissipative media is related to the existence of fine-structure components of periodic and non-steady-state flows (7). The simplest leveling of the concentration works only in a stationary phase. Accordingly, the hypothesis of an inert admixture has no grounds, except for J. Taylor's declaration (14).

According to the general scientific knowledge, the factors that have an effect on the revealed non-linearity in the examined characteristics of natural waters are fluctuations of the parameters of order (15), which determine the singular contributions to the dynamic characteristics of water flow.

Author contributions

OR: Analysis of water problems. VF: Neural network data analysis.

References

1. Rosenthal OM, Tambieva DA. Cyclic variations of quality characteristics in river water in an industrial region. *Water Sci Technol.* (2021) 83:854–61.

2. Rodnikova MN. On the elasticity of a spatial network of hydrogen bonds in liquids and solutions. In: Tsivadze AY editor. *Structural self-organization in solutions and at the phase boundary.* Moskow: LKI (2008). p. 151–98.

3. Danilov-Danilyan VI, Rosenthal OM. The properties of natural waters determined by their microstructural self-organization. *Water Resour.* (2021) 48:254–62.

4. Borovikov VP. *Neural networks statistica neural networks: methodology and technology of modern data analysis*. Moskow: Hotline–Telecom (2008).

5. Danilov-Danilyan VI, Rosenthal OM. Nonlinear effects in the formation of water quality. *Doklady Earth Sci.* (2021) 497:345–7.

6. Chashechkin Y. Singularly perturbed components of flows – linear precursors of shock waves. *J Math Model Nat Phenom.* (2018) 13:17.

7. Chashechkin YD, Rozental OM. River flow structure and its effect on pollutant distribution. *Water Resour.* (2019) 46:910–8.

8. Tripadvisor. *Who has not seen the confluence of the waters of the Amazon and the Rio Negro – he has not seen anything.* (2023). Availale online at: https://www.tripadvisor.ru/ShowUserReviews-g303235-d554183-r229755948, (accessed December 8, 2023).

9. Surinati D, Kusmanto E. Stratification of water mass in Lasolo Bay, Southeast Sulawesi. *Oseanol Limnol Indonesia.* (2016) 1:17.

10. Giering SLS, Sanders R, Adrian M, Henson S, Riley JS, Marsay C, et al. Particle fux in the oceans: challenging the steady state assumption. *Glob Biogeochem.* (2017) 31:159–71.

11. Rozental OM, Podkin YG. Methods and effectiveness of dielectric measurements of water-saturated systems. *Meas Tech.* (2014) 57:941–6.

12. Aleksandrov VS, Badenko LA, Snegov VS. Macroscopic fluctuations in the density of water. *Meas Tech.* (2004) 47:295–9.

13. Landau LD, Lifshits EM. *Theoretical physics. Volume VI. Hydrodynamics.* Moskow: University Science Books (1986).

14. Taylor JR. *An Introduction to error analysis: the study of uncertainties in physical measurements.* Sausalito: University Science Books (1997).

15. Nicolis G, Prigogine I. *Self-Organization in Nonequilibrium systems: from dissipative structures to order through fluctuations.* New York, NY: Wiley (1977).