

РАНГОВЫЙ КЛАССИФИКАТОР СОСТОЯНИЙ ПРИРОДНОЙ СРЕДЫ

А.В. Скатков, А.А. Брюховецкий, Д.В. Моисеев

ФГАОУ ВО «Севастопольский государственный университет»,

РФ, г. Севастополь, ул. Университетская, 33

E-mail: dmitriymoiseev@mail.ru

В работе рассматриваются основные особенности, связанные с разработкой и исследованием устройства на основе методов интеллектуальной технологии оценивания состояний природной среды. Предлагаемый подход базируется на методах непараметрической статистики с использованием ранговых критериев и позволит осуществить в ходе мониторинга интеллектуальный анализ данных ключевых показателей природной среды, например, таких как гидрометеорологические данные об уровне загрязнения и составе воздуха, почвы, предельно допустимых выбросов вредных веществ, экологического контроля аномалий и другие. Представлены статическая, динамическая, интегральная и обобщенная модели классификации состояний природной среды.

Ключевые слова: системы мониторинга, статическая модель, динамическая модель, интегральная модель, обобщенная модель, ранговые критерии, интеллектуальная технология.

Поступила в редакцию: 04.02.2021. После доработки: 05.03.2021.

Введение. Исследования антропогенной трансформации природной среды предполагают использование измерительных устройств, предназначенных для решения задач контроля параметров объектов и оценки их динамического состояния [1, 2].

Применение в информационно-измерительных системах искусственных нейронных сетей в настоящее время активно внедряется не только в нашей стране, но и по всему миру [3–6]. Особый интерес применение нейронных сетей, которые являются относительно новым инструментом обработки информации, имитирующим способности человеческого мозга к обучению и запоминанию, а также его чрезвычайно разветвленную архитектуру, обеспечивающую многоканальное параллельное функционирование, при выполнении классификации состояний природной среды [7].

Необходимо отметить, что природотехнические объекты, а также процессы в них протекающие, характеризуются высокой сложностью и динамичностью, а значительная часть данных процессов до настоящего времени изучена и формализована далеко не полностью. Поэтому одно из важнейших направлений анализа данных в рассматриваемой области – применение в информационно-

измерительных системах искусственных нейронных сетей. В работах ученых из различных стран [8–10] показана высокая эффективность применения аппарата искусственных нейронных при решении отдельных задач обработки данных при классификации состояний природной среды.

Известные существующие измерители характеристик природной среды выполняют точечные, выборочные измерения. Предлагаемые решения базируются на интеллектуальных методах и отличаются от известных в использовании малых объемов выборок; низкой трудоемкостью и высоким быстродействием; критериев для оценки информационных ситуаций изменения состояния контролируемых объектов и отнесения их к заданному классу, комплексно оценивают состояние объектов природной среды на основе статической, динамической, интегральной и обобщенной моделях. Поэтому повышение степени интеллектуализации в разрабатываемых устройствах для оценивания состояний объектов природной среды представляет научный и практический интерес.

Постановка задачи. Целью работы является разработка классификатора состояний природной среды на основе ранговых критериев для повышения интел-

лектуальности и уровня достоверности применяемых методов классификации. В качестве ранговых критериев могут применяться критерии Спирмена, Уилкоксона, знаков и другие [11]. Основными особенностями применяемых методов является то, что каждый из них оценивает различия двух выборок, представленных гистограммами распределений контролируемого параметра. В зависимости от критерия проверяется гипотеза:

- о наличии корреляции между случайными величинами X и Y ,
- имеют ли распределения двух выборок одинаковую форму и отличаются только сдвигом на константу,
- плотность распределения симметрична относительно своего центра.

Обозначим $O = \{O_1, \dots, O_j, \dots, O_m\}$ – множество контролируемых объектов природной среды. Определим контролируемые характеристики объектов, по которым будем оценивать их состояние:

D_j – значение характеристики j -го объекта, векторная величина с компонентами элементов множеств O_j ; V_j – скорость изменения значения; D_j – векторная величина с компонентами элементов множеств O_j , где $V = \Delta V / \Delta t$; $\Delta V = D(t_i) - D(t_{i-1})$, изменение значения D_j за время Δt ; $\Delta t = t_i - t_{i-1}$ – промежуток между двумя соседними моментами времени; t_i – i -ый момент времени измерения характеристик, $t_i \in [0; T]$; Σ_j – сумма значений накопленных измерений

j -го объекта, векторная величина с компонентами элементов множеств O_j .

Значения характеристик D_j, V_j, Σ_j нормированы, определены в диапазоне $[0; 1]$. Структуры $\{D, V, \Sigma\}$ являются векторными матричными функциями операционных моментов времени принятия решений, т.е. при фиксированном t_i образуют числовые матрицы.

Будем полагать, что состояние объекта O_j в заданный момент времени t зависит от значений контролируемых характеристик D_j, V_j, Σ_j . Пусть эти состояния обозначены $S_j^t = \{S_{j0}^t, S_{j1}^t, S_{j2}^t, S_{j3}^t\}$ – множество возможных состояний объекта O_j в момент времени t . Значения состояний определены и нормированы в диапазоне $[0; 1]$. Пусть состояния определены на интервалах $[R_k; R_{k+1}]$, где R_k – порог для задания области определения состояния объекта, $k=0, 2$. Для примера без потери общности будем полагать, что мы имеем четыре интервала, на которых определяются возможные состояния S_j^t каждого объекта: $S_0^t \in [0; R_0]$ – нормальное – 0, $S_1^t \in (R_0; R_1]$ – близкое к нормальному состоянию – 1, $S_2^t \in (R_1; R_2]$ – предкритическое – 2, $S_3^t \in (R_2; 1]$ – аномальное – 3.

Рассмотрим в пространстве координат OD_j, V_j области состояния S_j^t объекта O_j по значениям (D_j, V_j) . Отобразим точку с координатами (D_j, V_j) на плоскости (рис. 1), на которой выделены четыре области значений, соответствующие состояниям S_j^t объекта O_j .

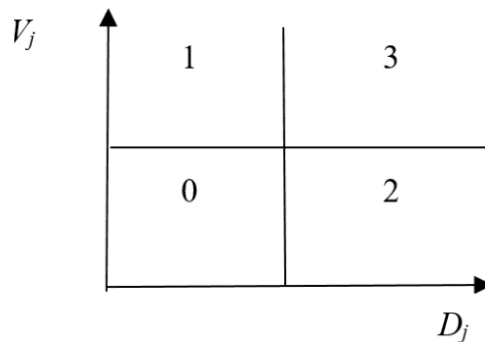


Рис. 1. Области $\{0, 1, 2, 3\}$ различия состояний S_j^t в пространстве параметров D_j, V_j , где D_j – значение характеристики j -го объекта, V_j – скорость изменения значения D_j

Fig. 1. Region $\{0, 1, 2, 3\}$ the differences of States of S_j^t in the parameter space D_j, V_j , where D_j is the characteristic value of the j -th object, V_j – speed change value D_j

Соответственно для трех характеристик D_j , V_j , Σ_j область различий будет представлять собой куб.

Для примера области разделены прямыми, хотя в реальных условиях это могут быть кривые линии, форма которых определяется экспертом для конкретных условий различия состояний объектов.

Методы и результаты. Предлагается использовать следующие модели: статическая, динамическая, интегральная и обобщенная. В каждой из первых трех моделей осуществляется сравнение двух выборок: текущей – содержащей результаты измерений за заданный промежуток времени (апостериорная информация) и эталонной – соответствующей значениям, заданным ЛПР (априорная информация). Статическая модель содержит в каждом интервале количество измерений контролируемого параметра. В этой модели представлены первичные измерения контролируемого параметра. В динамической модели каждый интервал гистограммы содержит значения изме-

нений (производных). Указанная модель позволяет оценивать динамику состояния объекта. Интегральная модель содержит сумму значений накопленных измерений, например, полученных с разных датчиков. Наконец, обобщенная модель представляет собой нейронную трехканальную сеть, на вход которой поступают результаты статической, динамической, интегральной моделей, а на выходе – обобщенный показатель состояния объекта. Настройка нейронной сети выполняется на основе принципа обратного распространения ошибки. Структурная схема устройства представлена на рис. 2.

Преимуществами предлагаемого подхода оценки состояния объектов на основе ранговых критериев являются: универсальная обработка данных, представленных в виде гистограмм, чувствительность к изменениям состояния объектов, получение обобщающей оценки показателя, адаптивность к состоянию природной среды.

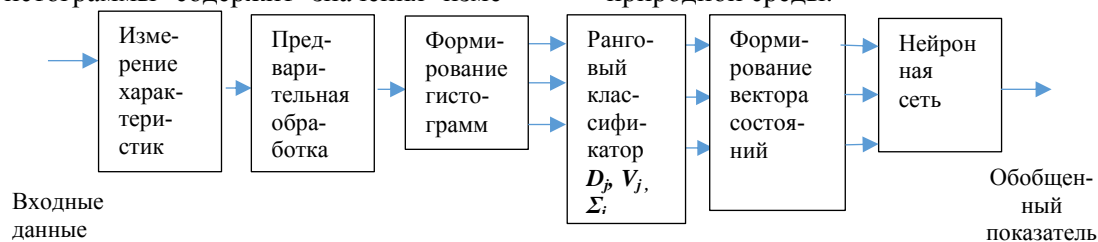


Рис. 2. Структурная схема рангового классификатора
Fig 2. Block diagram of the rank classifier

Целевое назначение программной системы – исследование применения и реализация нового подхода для процессов обнаружения изменения состояния объектов природной среды. Подход базируется на использовании непараметрических методов статистической оценки, методов имитационного моделирования, методов принятия решений.

Рассмотрим функционирование устройства на примере использования рангового критерия Спирмена. Суть эффекта, на основании которого распознается состояния объектов, достигается при анализе дисперсии случайной ошибки в уравнениях регрессии. Допущение о по-

стоянстве дисперсии остатков известно, как допущение о гомоскедастичности [12]. Если это допущение нарушено и дисперсия остатков не является постоянной, то принимают решение, что оценки гетероскедастичны, имеет место G -эффект.

G -эффект – это следствие неоднородности наблюдений. Характеризуется тем, что не выполняется предпосылка использования метода наименьших квадратов

$$D[y] = \sigma^2 = \text{const},$$

где y – результирующая переменная регрессионного анализа.

Именно оценка неоднородности состояния природной среды может использоваться для обнаружения аномальных явлений. В регрессионных моделях дисперсия σ^2 может:

- расти по мере увеличения значений фактора X ;
- иметь наибольшие значения при средних значениях X , уменьшаясь по мере приближения к крайним значениям;
- иметь наибольшие значения при малых значениях X , быстро уменьшаться и становиться практически постоянной по мере увеличения X .

Методы регрессионного анализа при соблюдении следующих условий являются исчерпывающе информативными:

- (Y_1, \dots, Y_n) – независимые нормально распределенные случайные величины.
- Выборочные дисперсии Y однородны, т.е. $\sigma_y^2 \{Y_u\} = \dots = \sigma_y^2 \{Y_v\}$, где u и v – наборы X , $\sigma_y^2 \{Y\}$ – не зависит от математического ожидания Y .
- $\Delta_{\text{ош}} \{X\} \ll \Delta_{\text{ош}} \{Y\}$.

Это утверждение является следствием теоремы Ю.В. Линника [13] о том, что регрессионный анализ является частным случаем метода максимального правдоподобия, обладающего экстремальными информационными характеристиками. Достаточность статистик регрессионного анализа – это извлечение всей имеющейся информации для параметрического оценивания по имеющимся измерениям.

Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям [11]:

- они должны быть количественно измеримы. Если необходимо включить в модель качественный фактор, не имеющий количественного измерения, то ему нужно придать количественную определенность.
- факторы не должны быть коррелированы между собой и тем более находиться в точной функциональной связи.
- Включение в модель факторов с высокой взаимной корреляцией (мультиколлинеарность), когда, например, $r_{yx1} < r_{x1x2}$, для зависимости $y = a + b_1x_1 + b_2x_2 + e$ может привести к нежелательным последствиям – система нормальных уравнений

может оказаться плохо обусловленной и повлечь за собой неустойчивость и ненадежность оценок коэффициентов регрессии. Для определения наличия или отсутствия реальной мультиколлинеарности можно использовать некоторые эвристические рекомендации по её выявлению на основе анализа: матрицы парных коэффициентов корреляции объясняющих переменных, значения коэффициента детерминации и коэффициента частной корреляции.

Обнаружение G -эффекта осуществляется в процессе оценки изменений состояния объекта по значениям остатков регрессионной модели. В результате принимается или отклоняется выдвинутая гипотеза о наличии G -эффекта, оцениваются ошибки первого и второго рода и проводится анализ риска принятой гипотезы.

Тест распознавания наличия G -эффекта включает следующую последовательность действий.

1. Формируются выборки, соответствующие матрице условий регрессионного анализа. Строится регрессионное уравнение для выходного параметра y_i .

$$\hat{y}_i = b_0 + \sum_{l=1}^m b_l x_{il}$$

и вычисляются остатки.

2. Значения выделенного столбца матрицы условий x_i упорядочиваются по возрастанию.

3. Значения элементов абсолютных ошибок для выделенного столбца упорядочиваются по возрастанию.

4. Рассчитываются ранги элементов столбцов и их разности.

5. Рассчитываются ранги остатков e_i

$$e_i = |y_i - \hat{y}_i|, \quad i = \overline{1, n}$$

6. Рассчитывается коэффициент ранговой корреляции Спирмена

$$\rho = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)},$$

где D_i – разность рангов x_i и e_i .

7. Рассчитывается статистика

$$t_{\text{набл}} = \rho \sqrt{n-2} / \sqrt{1-\rho^2}$$

распределенная нормально $N(0,1)$.

8. Для выбранного уровня значимости p и объема наблюдений n анализи-

руется неравенство для использования в критерии принятия гипотезы о наличии G -эффекта:

$$|t_{набл}| > t_{st(0.01;n-2)}$$

где t_{st} – статистика Стьюдента. Например, для уровня значимости $p=0,01$ и $n=50$, t_{st} (критическое)=3,57.

Предложенный тест позволяет установить зоны уверенного/неуверенного распознавания гипотез $P(H0/H0)$, $P(H1/H1)$ отсутствия/наличия G -эффекта.

Схема управления типовым экспериментом по обнаружению G -эффекта представлена на рис. 3.

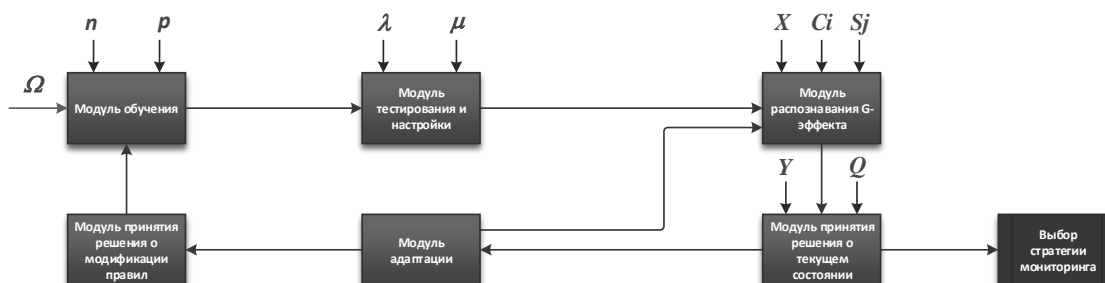


Рис. 3. Схема управления типовым экспериментом по обнаружению G -эффекта
 Fig. 3. Control scheme for a typical G -effect detection experiment

Схема настройки включает в себя совокупность модулей по настройке значений ряда параметров: TC – текущее состояние объекта, Sj – зоны уверенного/неуверенного распознавания G -эффекта, p – уровень значимости, n – объем наблюдений, Ci – критерий распознавания и другие с целью их адаптации к текущему состоянию природной среды. Оценка качества классификации производится по значениям гипотез $P(H0/H0)$, $P(H1/H1)$ и ошибок первого и второго рода $P(H1/H0)$, $P(H0/H1)$. К числу важных параметров, которые влияют на значения гипотез, можно отнести следующие:

p_{ij} – уровень достоверности при использовании критерия Ci для зоны Sj ,
 n_{ij} – объем контролируемой выборки при использовании критерия Ci для зоны Sj ,

k_{ij} – коэффициент использования ресурсов (времени, памяти) при использовании критерия Ci для зоны Sj ,

z_{ij} – затраты на распознавание изменения состояния объекта при использовании критерия Ci для зоны Sj .

В результате может быть предложена оценочная матрица для системы поддержки принятия решений на основе распознавания G -эффекта (табл. 1).

Таблица 1. Оценочная матрица для системы поддержки принятия решений на основе распознавания G -эффекта

Критерии распознавания	Зоны распознавания G -эффекта		
	$S1$	Sj	Sq
$C1$
.....
Ci	$\langle p_{ij}, n_{ij}, k_{ij}, z_{ij} \rangle$
.....
Cm

В табл. 1 используются следующие обозначения:
 Sj – зона распознавания G -эффекта, $Sj \in [0;1]$,
 1 – значение уверенного наличия G

-эффекта, 0 – значение уверенного отсутствия G -эффекта. Например, $S1 \in [0 - 0,2]$, $S2 \in (0,2 - 0,4]$,, $Sq \in (0,8 - 1,0]$.

Функционирование нейронной сети по формированию обобщенного показателя построено на следующем принципе: каждый l -й объект экспертизы ($l=1,2,\dots,k$) отождествляется с характеристическим вектором его признаков $X_l=(x_{l1}, x_{l2}, \dots, x_{lm})$.

При совместном представлении данных экспертизы нескольких объектов их характеристические векторы образуют факторное пространство X , обычно представляемое матрицей

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{k1} & x_{k2} & \dots & x_{km} \end{bmatrix},$$

которую часто называют референтной (эталонной) матрицей.

Точно так же и в нашей задаче мы будем оценивать обобщенный показатель качества G -эффекта нового объекта экспертизы по сравнению с рассчитанными ранее показателями нескольких уже известных объектов.

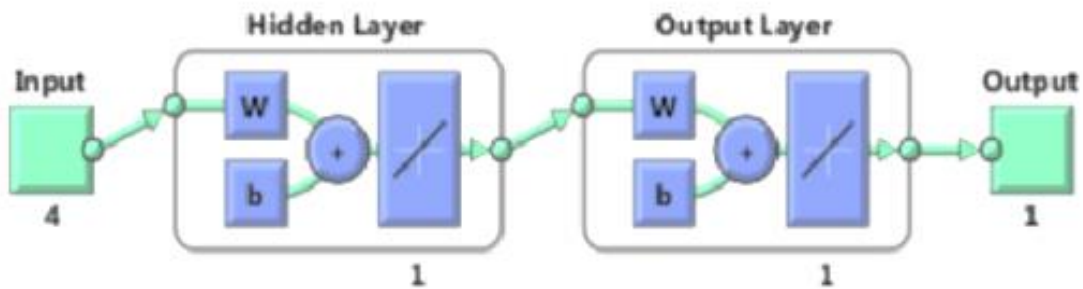


Рис. 4. Символическая схема двухслойной нейронной сети

Fig. 4. Symbolic diagram of a two-layer neural network

Для построения рангового классификатора состояний природной среды, предлагается использовать нейронную сеть с линейными функциями активации в обоих слоях, а затем применять логистическую нелинейную функцию к полученным значениям показателей. Поэтому, как показано на рис. 4, в обоих слоях выбраны линейные функции активации relu . Первый слой нейронной сети – скрытый (hidden), второй слой – выходной (output); 1,2 W – матрицы весов первого и второго слоя; 1,2 b – постоянные сигналы смещений первого и второго слоя. Параметры сети: 4 – размерность вектора входа, 1 – наличие одного нейрона в скрытом и выходном слоях. Выбраны линейные функции активации relu . На вход сети подаются данные X , а в качестве вектора цели выбран вектор показателей $\text{comp } J$.

Заключение. Разработка устройства оценивания состояний природных объ-

ектов на основе рангового классификатора приведет к повышению обоснованности, достоверности и оперативности процессов поддержки принятия решений при исследовании антропогенной трансформации природной среды. Практическая значимость результатов работы приведёт к снижению уровня негативного воздействия природных и антропогенных факторов, в частности, на состояние экосистем акватории г. Севастополя.

В планах дальнейших исследований предлагается оценить влияние объема выборки n на статистическую чувствительность, статистическую устойчивость, области уверенного/неуверенного распознавания, а также построить систему поддержки принятия решений по обнаружению G -эффекта и рассмотреть адаптивный подход к построению оценочной матрицы.

Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований (грант № 19-29-06015, 19-29-06023, 18-47-920007).

СПИСОК ЛИТЕРАТУРЫ

1. Гайский В.А., Гайский П.В. Многомерный гармонический анализ Фурье при измерениях полей морской среды // Системы контроля окружающей среды. 2019. Вып. 4 (38). С. 33–42.
2. Скатков А.В., Брюховецкий А.А., Моисеев Д.В. Интеллектуальная система мониторинга для решения крупномасштабных научных задач в облачных вычислительных средах // Информационно-управляющие системы. СПб.: Изд-во «ГУАП», 2017. № 2 (87). С. 19–25.
3. Селиванова З.М. Интеллектуализация информационно-измерительных систем неразрушающего контроля теплофизических свойств твердых материалов. М.: Машиностроение-1, 2006. 184 с.
4. Харисова З.И. Информационно-измерительная система для гранулометрического анализа жидких дисперсных сред на основе видеотехнических средств и нейросетевых технологий: автореф. дис. ... канд. техн. наук. Уфа, 2018. 21 с.
5. Russell S., Norvig P. Artificial Intelligence: A Modern Approach, 3rd Edition. Pearson Education, 2010. 1095 p.
6. Orlov S.P., Vasilchenko A.N. Intelligent measuring system for testing and failure analysis of electronic devices // 2016 XIX IEEE International Conference on Soft Computing and Measurements (SCM). IEEE Conference Publications. 2016. Vol. 1. P. 401–403.
7. Посошков В.Л. Использование искусственных нейронных сетей для статистического оценивания ряда среднесуточных осадков // Системы контроля окружающей среды. 2002. Вып. 4. С. 146–153.
8. Werner H., Obach M. New neural network types estimating the accuracy of response for ecological modelling // Ecological Modelling. 2001. Vol. 146. № 1-3. P. 289–298.
9. Gevrey M., Dimopoulos I., Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models // Ecological Modelling. 2003. Vol. 160. № 3. P. 249–264.
10. Ciavatta S., Pastres R., Badetti C., Ferrari G. Estimation of phytoplanktonic production and system respiration from data collected by a real-time monitoring network in the lagoon of venice // Ecological Modelling. 2008. Vol. 212. № 1-2. P. 28–36.
11. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: Высшая школа, изд-во «Юнити», 1998. 1000 с.
12. Кремер Н.Ш. Эконометрика. М.: ЮНИТИ-ДАНА, 2006. 311 с.
13. Линник Ю.В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. Изд. 2-е, доп. и испр. М.: Гос. изд-во физ.-мат. лит., 1962. 349 с.

RANKED CLASSIFICATION OF ENVIRONMENTAL CONDITIONS

A.V. Skatkov, A.A. Bryukhovetskiy, D.V. Moiseev

Sevastopol State University, RF, Sevastopol, Universitetskaya St., 33

The main features related to the development and research of a device for assessing the states of the natural environment based on the methods of intelligent technology is considered. The proposed approach is based on methods of nonparametric statistics using rank criteria and will make it possible to carry out, during monitoring, intelligent analysis of data on key indicators of the natural environment, such as hydrometeorological data on the level of pollution and composition of air, soil, maximum permissible emissions of harmful substances, environmental control of anomalies, and others. The static, dynamic, integral and generalized models of classification of the states of the natural environment are presented.

Keywords: monitoring systems, static model, dynamic model, integral model, generalized model, rank criteria, intelligent technology.

REFERENCES

1. *Gajskij V.A., Gajskij P.V.* Mnogomernyj garmonicheskij analiz Fur'e pri izmerenijah polej morskoy sredy. *Sistemy kontrolja okruzhajushhej sredy*. 2019. Vol. 4 (38). pp. 33–42.
2. *Skatkov A.V., Brjuhoveckij A.A., Moiseev D.V.* Intellektual'naja sistema monitoringa dlja reshenija krupnomasshtabnyh nauchnyh zadach v oblachnyh vychislitel'nyh sredah. Informacionno-upravljajushhie sistemy. SPb.: Izd-vo «GUAP», 2017. No 2 (87). pp. 19–25.
3. *Selivanova Z.M.* Intellektualizacija informacionno-izmeritel'nyh sistem nerazrushajushhego kontrolja teplofizicheskikh svojstv tverdyh materialov. M.: Mashinostroenie-1, 2006. 184 p.
4. *Harisova Z.I.* Informacionno-izmeritel'naja sistema dlja granulometricheskogo analiza zhidkih dispersnyh sred na osnove videotehnicheskikh sredstv i nejrosetevykh tehnologij: avtoref. dis. ... kand. tehn. nauk. Ufa, 2018. 21 p.
5. *Russell S., Norvig P.* Artificial Intelligence: A Modern Approach, 3rd Edition. Pearson Education, 2010. 1095 p.
6. *Orlov S.P., Vasilchenko A.N.* Intelligent measuring system for testing and failure analysis of electronic devices. 2016 XIX IEEE International Conference on Soft Computing and Measurements (SCM). IEEE Conference Publications. 2016. Vol. 1. pp. 401–403.
7. *Pososhkov V.L.* Ispol'zovanie iskusstvennyh nejronnyh setej dlja statisticheskogo ocenivanija rjada srednesutochnykh osadkov. *Sistemy kontrolja okruzhajushhej sredy*. 2002. Vol. 4. pp. 146–153.
8. *Werner H., Obach M.* New neural network types estimating the accuracy of response for ecological modelling. *Ecological Modelling*. 2001. Vol. 146. No 1-3. pp. 289–298.
9. *Gevrey M., Dimopoulos I., Lek S.* Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*. 2003. Vol. 160. No 3. pp. 249–264.
10. *Ciavatta S., Pastres R., Badetti C., Ferrari G.* Estimation of phytoplanktonic production and system respiration from data collected by a real-time monitoring network in the lagoon of venice. *Ecological Modelling*. 2008. Vol. 212. No 1-2. pp. 28–36.
11. *Ajvazjan S.A., Mhitarjan V.S.* Prikladnaja statistika i osnovy jekonometriki. M.: Vysshaja shkola, izd-vo «Juniti», 1998. 1000 p.
12. *Kremer N.Sh.* Jekonometrika. M.: JuNITI-DANA, 2006. 311 p.
13. *Linnik Ju.V.* Metod naimen'shikh kvadratov i osnovy matematiko-statisticheskoy teorii obrabotki nabljudenij. Izd. 2-e, dop. i ispr. M.: Gos. izd-vo fiz.-mat. lit., 1962. 349 p.