

УДК 004.8
DOI: 10.15827/0236-235X.134.230-236

Дата подачи статьи: 01.03.21
2021. Т. 34. № 2. С. 230–236

Прогнозирование угроз в сложных распределенных системах на основе интеллектуального анализа больших данных автоматизированных средств мониторинга

*Е.В. Пальчевский*¹, аспирант, *teelxp@inbox.ru*
*О.И. Христодуло*¹, д.т.н., профессор, *o-hristodulo@mail.ru*
*С.В. Павлов*¹, д.т.н., профессор, *psvgis@mail.ru*

¹ Уфимский государственный авиационный технический университет,
г. Уфа, 450008, Россия

В работе предлагается метод прогнозирования паводковой ситуации в сложных распределенных системах при угрозе нанесения материального ущерба. Метод основан на использовании восстановленных данных по автоматическим станциям на базе стационарных гидрологических постов мониторинга уровней воды.

Обоснована актуальность выбранной темы исследования с точки зрения применения восстановленных данных на соответствующих поставленному условию автоматических станциях контроля и мониторинга уровней воды для прогнозирования паводковой ситуации. Исходя из этого сформулирована математическая постановка задачи (на основе краткосрочного прогнозирования уровней воды), в рамках которой реализован алгоритм поиска автоматических станций и интерполяции (восстановления) исторических значений уровней воды для прогнозирования уровней воды в сложных распределенных системах.

Проведен анализ эффективности реализованного метода прогнозирования угроз в сложных распределенных системах как одного из модулей искусственной нейронной сети. В качестве примера показан результат на автоматической станции «Булгаково», расположенной между стационарными гидрологическими постами «Ляхово» и «Охлебинино». По результатам анализа погрешность рассчитанного уровня воды при прогнозировании на одни сутки на автоматической станции варьировалась от 7 до 53 см.

Таким образом, в составе искусственной нейронной сети данный метод позволяет прогнозировать уровни воды с приемлемой точностью для предвидения паводковой ситуации (на примере паводка 2020 г. в Республике Башкортостан), что позволяет специальным службам проводить специализированные мероприятия по парированию данной угрозы.

Ключевые слова: восстановление данных, сложные распределенные системы, прогнозирование угроз, интеллектуальный анализ данных, нейронные сети, паводковая ситуация, прогнозирование уровня воды.

К сложным распределенным системам относятся различные составляющие в виде физических, биологических и цифровых систем, для которых, как правило, существует опасность в виде внешних угроз, способствующих нанесению существенного материального ущерба (в том числе и целенаправленного воздействия на человека, а также прямого воздействия человека на объекты, создающие внешние угрозы). Например, в Республике Башкортостан одной из таких угроз является весенний паводок, ставящий под угрозу сложные распределенные системы из-за их возможного подтопления и затопления, поэтому задача прогнозирования паводковой ситуации важна и актуальна.

В связи с появлением новых технических возможностей автоматического измерения па-

раметров состояния (уровня воды, температуры, скорости и направления ветра и др.) сложных распределенных систем в последнее время все большее применение находят автоматические станции измерения и видеофиксации уровня воды в водных объектах [1–4]. Основное назначение этих станций – раннее обнаружение угрозы (паводка) и информирование о ней органов управления и населения [5, 6]. В то же время данные постов можно использовать для прогнозирования паводковой ситуации по предложенной в статье [7] методике.

В условиях отсутствия архивных измерений на автоматических постах (ранее они просто не существовали) предлагается ввести оценку уровня воды в точках размещения автоматических станций на основе архивных значений со-

седних стационарных гидрологических постов. Причем уровень воды в этих точках находился на какой-то отметке и закономерность его (уровня) изменения такая же, как на стационарных гидрологических постах. Этот метод можно использовать только для тех автоматических постов, которые находятся между стационарными гидрологическими постами (один выше по течению, другой ниже), и для них интерполируется значение уровня воды. При этом необходимо учитывать, что на стационарных гидрологических постах ежедневно измеряется дискретно одно значение уровня воды в фиксированный момент времени (обычно в 8 часов утра по местному времени), а на автоматических постах уровень воды измеряется непрерывно. Таким образом, было бы корректным в качестве значения уровня воды на k -м гидрологическом посту на конкретную дату выбирать среднее из непрерывно измеренных значений за 10 минут (с 7^{55} до 8^{05}) на дату измерения.

Постановка задачи

Одним из основных параметров возможного негативного воздействия паводковой ситуации на различные компоненты сложных распределенных систем является h – уровень воды в водных объектах, измеряющийся ежедневно на n стационарных гидрологических постах сотрудниками Башгидромета. Введем обозначения: h_{ji}^n – значение уровня воды, измеренное на n -м гидрологическом посту на i -ю дату j -го года. Здесь $k = \overline{1, n}$, где n – количество гидрологических постов, участвующих в расчетах на первом этапе краткосрочного прогнозирования уровней воды [8]; j – номер года; i – конкретная дата измерения.

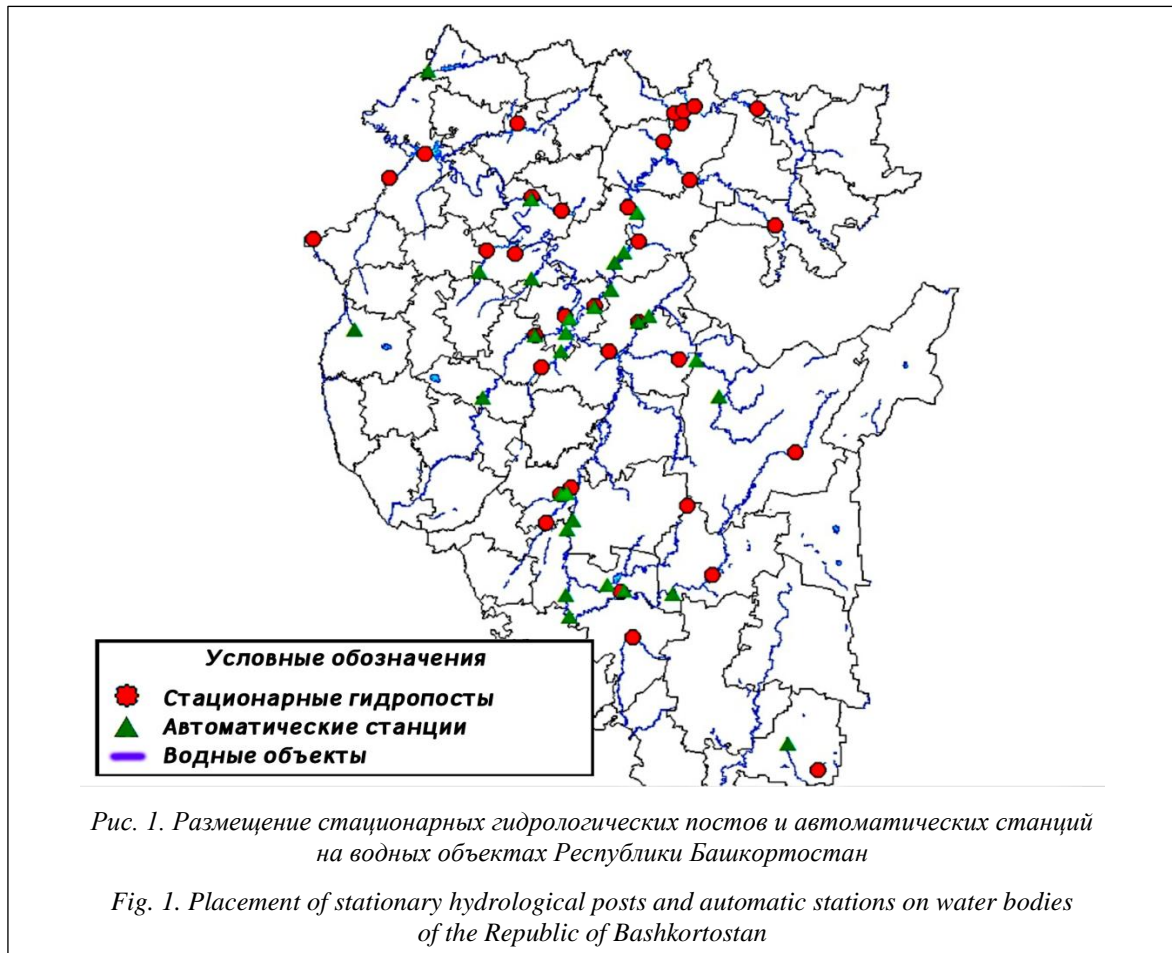
Далее вводим обозначение h_{ji}^{na} – уровень воды водных объектов, измеряющийся ежедневно на na автоматических станциях в i -ю дату j -го года, где $na = \overline{1, ka}$. Задача краткосрочного прогнозирования уровней воды на автоматических станциях аналогична задаче краткосрочного прогнозирования на стационарных гидрологических постах и заключается в том, чтобы в конкретный текущий i -й день измерения вычислить значение уровня подъема воды на следующий $i + 1$ -й день, то есть h_{ji+1}^n для любого $n = \overline{1, k}$. Но при этом также одной из важнейших задач в рамках как краткосрочного, так

и заблаговременного прогнозирования паводковой ситуации является восстановление исторических данных для расчета значения уровня воды на необходимые дату и год. Соответственно, для дальнейших расчетов введем обозначения: ha_{ji}^{na} – восстанавливаемое значение уровня воды в месте размещения na -й автоматической станции на i -ю дату j -го года; kd_n – нижестоящий по течению реки гидрологический пост; ku_n – вышестоящий по течению реки гидрологический пост. Здесь $na = \overline{1, ka}$ и $n = \overline{1, k}$, где ka – количество автоматических станций контроля и мониторинга уровня воды, участвующих в расчетах; j – номер года; i – конкретная дата измерения; k – количество стационарных гидрологических постов. В рамках прогнозирования паводковой ситуации задача восстановления исторических значений уровней воды состоит в том, чтобы в конкретный i -й день измерения интерполировать значение уровня подъема воды на автоматической станции, находящейся между kd_k и ku_k .

Использование данных автоматических станций для прогнозирования уровней воды

Как отмечалось ранее, задача восстановления исторических данных является важной и актуальной из-за отсутствия необходимой информации для прогнозирования уровней воды на автоматических станциях. Изначально для восстановления необходимых данных для прогнозирования уровней воды на автоматических станциях является важным определение их географического расположения на определенной территории в рамках сложной распределенной системы. На рисунке 1 показано расположение стационарных гидрологических постов (34 объекта) и автоматических постов (38 объектов) на водных объектах Республики Башкортостан. Анализ их взаимного расположения (рис. 1) показал, что 7 автоматических постов (во время паводка 2020 г.) находятся между стационарными гидрологическими постами (табл. 1).

Обозначим общее количество автоматических станций, расположенных между двумя стационарными постами, через na . На сегодняшний день $na = 7$, но со временем автоматические станции могут закрываться, перемещаться на другое место или ликвидироваться, поэтому na является переменной величиной.



Соответственно, обозначим ha_{ji}^{na} как интерполированное оценочное значение уровня воды в месте размещения na -й автоматической станции, $na = 1, ka$ в i -й день j -го года. Через KD обозначим множество номеров стационарных гидрологических постов, расположенных ниже

по течению, а через KU – множество номеров стационарных гидрологических постов, расположенных выше по течению соответствующей автоматической станции так, что

$$\begin{aligned}
 KD &= (kd_1, kd_2, \dots, kd_{na}), \\
 KU &= (ku_1, ku_2, \dots, ku_{na}),
 \end{aligned}
 \tag{1}$$

Таблица 1

Автоматические станции, расположенные между стационарными гидрологическими постами на водных объектах Республики Башкортостан

Table 1

Automatic stations located between stationary stream gauges on water bodies of the Republic of Bashkortostan

Нижестоящий гидрологический пост	Название станции	Вышестоящий гидрологический пост
с. Красная Горка (р. Уфа)	Красный Ключ	Павловская ГЭС, нижний бьеф (р. Уфа)
с. Красная Горка (р. Уфа)	Яман-Порт	Павловская ГЭС, нижний бьеф (р. Уфа)
с. Андреевка (р. Белая)	Бирск	г. Бирск (р. Белая)
д. Ляхово (р. Уршак)	Булгаково	с. Охлебинино (р. Белая)
г. Стерлитамак (р. Белая)	Стерлитамак (ул. Б. Хмельницкого)	д. Новофедоровское (р. Ашкадар)
г. Стерлитамак (р. Белая)	Стерлитамак (ул. Водолаженко)	с. Новая Отрадовка (р. Стерля)
с. Метели (р. Ай)	Большеустьикинское	с. Лаклы (р. Ай)

автоматическая станция с номером na находится между двумя стационарными гидрологическими постами с номерами kd_n и ku_n .

В расчетах (интерполяции) значения ha_{ji}^{na} участвует расстояние между автоматической станцией и соседними стационарными гидрологическими постами, поэтому введем в рассмотрение переменную x , которая обозначает расстояние (удаление) соответствующего поста от устья реки, и, вообще говоря, уровень воды в любой точке реки можно рассматривать как функцию от этого расстояния:

$$h = h(x). \tag{2}$$

Для каждого n -го стационарного поста это расстояние есть фиксированное число x_n , $n = \overline{1, k}$; соответственно, для автоматических станций обозначим его xa_{na} , $na = \overline{1, ka}$. Используя введенные обозначения, нахождение некоторого автоматического поста между двумя стационарными гидрологическими постами опишем соотношением

$$x_{kd_n} < xa_{na} < x_{ku_n}, \tag{3}$$

$$h_{ji}^n = h_{ji}^n(x_n), n = \overline{1, k},$$

$$ha_{ji}^n = h_{ji}^n(xa_{na}), na = \overline{1, ka}.$$

В общем случае эта зависимость имеет одну и ту же природу, определяемую соотношением (2), но для пояснения алгоритма интерполяции для стационарных гидрологических постов и автоматических станций вводятся отличающиеся друг от друга обозначения.

Предполагая, что изменение уровня воды в реке между двумя точками (местами расположения гидрологических постов) происходит линейно (это для небольших расстояний вполне приемлемо), интерполируемое значение ha_{ji}^{na} находится на прямой линии, соединяющей точки уровня воды на соседних стационарных постах, и вычисляется (в соответствии с известными правилами Евклидовой геометрии, см. рис. 2) по соотношению

$$ha_{ji}^{na} = h_{ji}^{kd_n} + (h_{ji}^{ku_n} - h_{ji}^{kd_n}) \frac{xa_{na} - x_{kd_n}}{x_{ku_n} - x_{kd_n}}. \tag{4}$$

На рисунке 3 показан пример полученных в результате интерполяции значений уровня подъема воды для автоматической станции «Красный Ключ» на основе данных соседних стационарных гидрологических постов – «с. Красная Горка (р. Уфа)» и «Павловская ГЭС, нижний бьеф (р. Уфа)». В качестве примера используются исторические данные уровня воды на данных гидрологических постах от

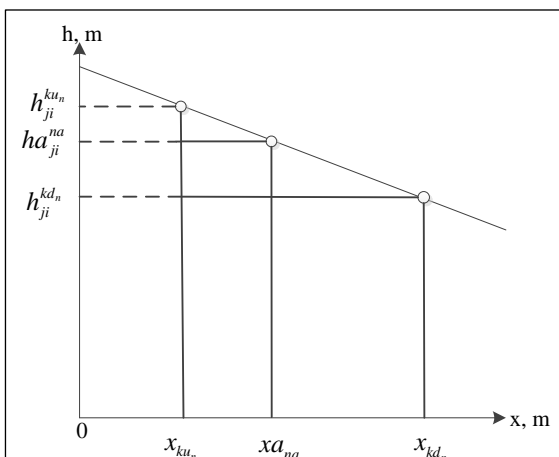


Рис. 2. Схема изменения уровня воды на различных (соседних) гидрологических постах на фиксированную дату в зависимости от их удаленности от устья реки

Fig. 2. The pattern of variation in the water level at various (neighboring) stream gauges on a fixed date, depending on their distance from the mouth of the river

04.05.2009 г.: уровень воды на посту «с. Красная Горка (р. Уфа)» составил 197 см, на посту «Павловская ГЭС, нижний бьеф (р. Уфа)» – 413 см. В результате интерполяции для автоматического поста «Красный Ключ» получено значение уровня воды – 223 см, которое будем считать историческим показателем для этого автоматического поста на 04.05.2009 г. Таким

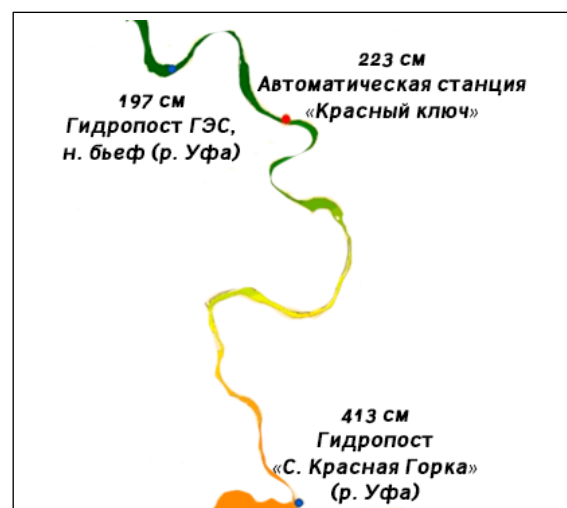


Рис. 3. Результат интерполяции данных по уровням подъема воды для автоматической станции

Fig. 3. The data interpolation result on water rise levels for an automatic station

Таблица 2

**Результаты прогнозирования уровня воды на автоматической станции «Булгаково»
с восстановленными и невосстановленными данными**

Table 2

Results of water level forecasting at the automatic station "Bulgakovo" with restored and non-restored data

Дата	Прогноз уровня воды		Фактическое значение уровня воды, см	Погрешность уровня воды с восстановлением/без восстановления, см
	без восстановления, см	с восстановлением, см		
21.05.2020	-12	-6	1	7/13
22.05.2020	15	13	14	1/1
23.05.2020	17	14	9	5/8
24.05.2020	-24	-23	-76	53/52
25.05.2020	-48	-15	6	21/54

образом, полученные в результате интерполяции значения уровня воды для каждого автоматического поста и на все остальные даты будем считать историческими данными для этого автоматического поста. Количество этих данных соответствует количеству наблюдений на соседних стационарных гидрологических постах.

Теперь массив полученных оценочных данных можно использовать для прогноза уровней воды по предложенному в статьях [7, 8] методу с применением рекуррентной искусственной нейронной сети второго поколения (используются структура в виде одного входного, трех промежуточных (скрытых) и одного выходного слоев, а также сигмоидальная функция активации и обучение без учителя на основе интеграции методов обратного распространения ошибки и Розенблатта (более подробное описание представлено в [8])) для всех автоматических постов, входящих в таблицу 1.

Анализ эффективности предлагаемого метода восстановления данных

В качестве примера эксперимент (табл. 2) проводился на большом количестве данных (объем выборки – 8 860 значений, при этом общий размер данных, используемых для анализа по всем гидропостам и автоматическим станциям на территории Республики Башкортостан, – более 500 Гб) в рамках паводка 2020 г., измеряемых на стационарных гидрологических постах мониторинга уровней воды («Ляхово» и «Охлебинино», между которыми находится автоматическая станция «Булгаково»), подведомственных Башкирскому управлению по гидрометеорологии и мониторингу окружающей среды.

Суть эксперимента заключается в прогнозировании уровня воды на одни сутки с восста-

новленными данными по представленному методу и без интерполируемых значений. С момента начала наблюдения ежедневно за развитием паводковой ситуации (в 2020 г. – 18 апреля) с помощью программы Forecaster, реализованной на основе свободно распространяемой библиотеки машинного обучения TensorFlow версии 2.3.0 [9], осуществлялся прогноз на следующий день по одной автоматической станции. На следующий день измерялся фактический уровень воды на этом же посту.

Таким образом, в ходе эксперимента была доказана целесообразность использования восстановления данных на автоматической станции с последующим применением интерполированных значений (4) в рамках прогнозирования паводковой ситуации за счет увеличения точности прогноза.

Заключение

В данной статье предложен метод прогнозирования паводковой ситуации на основе восстановления исторических данных на автоматических станциях, расположенных между вышестоящими и нижестоящими гидрологическими постами, что позволяет использовать дополнительные данные об уровнях воды для прогнозирования паводковой ситуации. Анализ результатов применения предложенного метода восстановления данных на автоматических станциях для прогнозирования уровня воды во время паводковой ситуации 2020 г. в Республике Башкортостан показал его эффективность. Был получен более точный прогноз в отличие от прогнозирования без восстановления данных. Таким образом, использование предлагаемого метода позволит более точно прогнозировать уровень воды на автоматических станциях контроля и мониторинга уровня

воды, что даст возможность соответствующим службам оперативно реагировать на угрозу паводка и принимать необходимые меры для ее парирования.

Исследование выполнено при финансовой поддержке РФФИ, проект № 20-08-00301.

Литература

1. Белюченко И.С. Региональный мониторинг – научная основа сохранения природы. Экологический вестник Северного Кавказа. 2006. Т. 2. № 1. С. 25–40.
2. Матишов Г.Г., Чикин А.Л., Бердников С.В., Шевердяев И.В., Клещенко А.В., Кириллова Е.Э. Экстремальное затопление дельты Дона весной 2013 г.: хронология, условия формирования и последствия // Вестн. ЮНЦ РАН. 2014. Т. 10. № 1. С. 17–24.
3. Никитин О.П., Поснов А.А. Измерения уровня моря в неблагоприятных условиях окружающей среды // Океанологические исследования. 2018. Т. 46. № 1. С. 128–133. DOI: 10.29006/1564-2291.jor-2018.46(1).11.
4. Гундров Д.С. Автоматизированная система мониторинга и прогнозирования паводковой обстановки на водных объектах Краснодарского края и перспективы ее развития // Пожарная безопасность: проблемы и перспективы. 2019. С. 200–202.
5. Moghar A., Hamiche M. Stock market prediction using LSTM recurrent neural network. Procedia Computer Science, 2020, vol. 170, pp. 1168–1173. DOI: 10.1016/j.procs.2020.03.049.
6. Mokhov V.G., Tsimbol V.I. Electrical energy consumption prediction of the federal district of Russia on the based of the recurrent neural network. J. Comp. Eng. Math., 2018, vol. 5, no. 2, pp. 3–15. DOI: 10.14529/jcem180201.
7. Palchevsky E.V., Khristodulo O.I., Pavlov S.V. Threats complex distributed systems parrying based on their development prognostication. Proc. VIII Sci. Conf. ITIDS, 2020, pp. 191–194. DOI: 10.2991/aisr.k.201029.036.
8. Пальчевский Е.В., Христодуло О.И., Павлов С.В., Соколова А.В. Анализ ретроспективных данных с применением технологий искусственного интеллекта для прогнозирования угроз в сложных распределенных системах // Вестн. компьютерных и информационных технологий. 2021. Т. 18. № 2. С. 39–45. DOI: 10.14489/vkit.2021.02.pp.039-045.
9. Пальчевский Е.В. Обучаемая с учителем нейронная сеть для прогнозирования уровня воды на гидропостях: Свид. о регистр. ПрЭВМ № 2020611923. Рос. Федерация, 2020.

Software & Systems
DOI: 10.15827/0236-235X.134.230-236

Received 01.03.21
2021, vol. 34, no. 2, pp. 230–236

Threat projection to the future in complex distributed systems based on the mining of big data and automated monitoring tools

*E.V. Palchevsky*¹, *Postgraduate Student, teelp@inbox.ru*

*O.I. Khristodulo*¹, *Dr.Sc. (Engineering), Professor, o-hristodulo@mail.ru*

*S.V. Pavlov*¹, *Dr.Sc. (Engineering), Professor, psvgis@mail.ru*

¹ *Ufa State Aviation Technical University, 450008, Ufa, Russian Federation*

Abstract. In connection with the emergence of new technical possibilities for automatic measurement of the parameters of the state of the external environment (including water levels), a method is proposed for predicting a flood situation in complex distributed systems for which there is a threat of material damage, using the recovered data from automatic stations based on stationary hydrological posts for monitoring water levels.

The relevance of the selected research topic is substantiated from the point of view of the application of the recovered data at the automatic stations for control and monitoring of water levels corresponding to the given condition for predicting a flood situation. Based on this, a mathematical formulation of the problem was formulated (based on short-term forecasting of water levels), within the framework of which an algorithm for searching for automatic stations and interpolation (restoration) of historical values of water levels was implemented to predict water levels in complex distributed systems.

The analysis of the effectiveness of the implemented threat forecasting method in complex distributed systems as one module of an artificial neural network (as an example, the result is shown at the automatic station "Bulgakovo" located between the stationary hydrological stations "Lyakhovo" and "Okhlebinino"), according to the results of which the water level when forecasting for one day at the automatic station varied from 7 to 53 cm.

Thus, as part of an artificial neural network, this method allows predicting water levels with acceptable accuracy to predict a flood situation (for example, the 2020 flood in the Republic of Bashkortostan), which allows special services to carry out specialized measures to counter this threat.

Keywords: data recovery, complex distributed systems, threat forecasting, data mining, neural networks, flood situation, water level forecasting.

Acknowledgements. RFBR funded the reported study, project no. 20-08-00301.

References

1. Belyuchenko I.S. Regional monitoring as the scientific basis of nature preservation. *The North Caucasus Ecological Herald*, 2006, vol. 2, no. 1, pp. 25–40 (in Russ.).
2. Matishov G.G., Chikin A.L., Berdnikov S.V., Sheverdyayev I.V., Kleshchenkov A.V., Kirillova E.E. Extreme flooding of the Don delta in spring 2013: Chronology, conditions of formation and consequences. *Vestn. Juzhnogo Nauchnogo Centra RAS*, 2014, vol. 10, no. 1, pp. 17–24 (in Russ.).
3. Nikitin O.P., Posnov A.A. Sea-level measurements in hostile conditions. *Journal of Oceanological Research*, 2018, vol. 46, no. 1, pp. 128–133. DOI: 10.29006/1564-2291.jor-2018.46(1).11 (in Russ.).
4. Gundrov D.S. Flood monitoring and forecasting automated systems of water environment of the krasnodar region, development prospects. *Pozharnaja Bezopasnost': Problemy i Perspektivy*, 2019, pp. 200–202 (in Russ.).
5. Moghar A., Hamiche M. Stock market prediction using LSTM recurrent neural network. *Procedia Computer Science*, 2020, vol. 170, pp. 1168–1173. DOI: 10.1016/j.procs.2020.03.049.
6. Mokhov V.G., Tsimbol V.I. Electrical energy consumption prediction of the federal district of Russia on the basis of the recurrent neural network. *J. Comp. Eng. Math.*, 2018, vol. 5, no. 2, pp. 3–15. DOI: 10.14529/jcem180201.
7. Palchevsky E.V., Khristodulo O.I., Pavlov S.V. Threats complex distributed systems parrying based on their development prognostication. *Proc. VIII Sci. Conf. ITIDS*, 2020, pp. 191–194. DOI: 10.2991/aisr.k.201029.036.
8. Palchevsky E.V., Khristodulo O.I., Pavlov S.V., Sokolova A.V. Analysis of historical data using artificial intelligence technologies for predicting threats in complex distributed systems. *Vestn. Komp'yuternykh i Informatsionnykh Tekhnologii*, 2021, vol. 18, no. 2, pp. 39–45. DOI: 10.14489/vkit.2021.02.pp.039-045 (in Russ.).
9. Palchevsky E.V. *Supervised Neural Network for Predicting Water Level at Gauging Stations*. Patent RF, no. 2020611923, 2020.

Для цитирования

Пальчевский Е.В., Христовуло О.И., Павлов С.В. Прогнозирование угроз в сложных распределенных системах на основе интеллектуального анализа больших данных автоматизированных средств мониторинга // Программные продукты и системы. 2021. Т. 34. № 2. С. 230–236. DOI: 10.15827/0236-235X.134.230-236.

For citation

Palchevsky E.V., Khristodulo O.I., Pavlov S.V. Threat projection to the future in complex distributed systems based on the mining of big data and automated monitoring tools. *Software & Systems*, 2021, vol. 34, no. 2, pp. 230–236 (in Russ.). DOI: 10.15827/0236-235X.134.230-236.