

УДК 004.8  
DOI: 10.15827/0236-235X.137.005-013

Дата подачи статьи: 07.10.21, после доработки: 06.12.21  
2022. Т. 35. № 1. С. 005–013

## **Интеллектуальный анализ и обработка больших разнородных данных для парирования угроз в сложных распределенных системах**

Е.С. Брекоткина<sup>1</sup>, к.э.н., доцент, [brekotkina@mail.ru](mailto:brekotkina@mail.ru)  
А.С. Павлов<sup>1</sup>, к.т.н., доцент, [asp.gis@gmail.com](mailto:asp.gis@gmail.com)  
С.В. Павлов<sup>1</sup>, д.т.н., профессор, [psvgis@mail.ru](mailto:psvgis@mail.ru)  
О.И. Христодуло<sup>1</sup>, д.т.н., профессор, [o-hristodulo@mail.ru](mailto:o-hristodulo@mail.ru)

<sup>1</sup> Уфимский государственный авиационный технический университет,  
г. Уфа, 450008, Россия

Предложен метод для прогнозирования некоторых угроз в сложных распределенных системах. Метод основан на интеллектуальном анализе и обработке больших разнородных данных, полученных в результате автоматического контроля изменения уровня воды в водных объектах и температуры воздуха в точке измерения. Такой контроль позволяет повысить эффективность планирования и реализации мероприятий по парированию подобных угроз.

Будущее значение уровня воды в точке измерения выбирается по результатам обработки данных, накопленных за все предыдущие паводковые периоды. В качестве анализируемых данных используются измеренные в равноотстоящие моменты времени значения температуры воздуха и уровня воды, вычислительные значения изменения уровня воды и температуры воздуха, а также прогнозные значения (по официальным данным гидрометслужбы) изменения температуры воздуха. На основании вычисления ретроспективной частоты изменения этой температуры и уровня воды в соответствующей точке в качестве прогнозируемого значения предлагается выбрать то, которому соответствует максимальная частота появления такого сочетания измеряемых параметров.

Результаты экспериментальной оценки точности прогнозирования уровня воды в водных объектах Республики Башкортостан в паводковый период 2021 г. подтверждают применимость предложенного метода прогнозирования для поддержки принятия решений по парированию угроз в сложных распределенных системах от резкого подъема воды даже при недостаточно автоматизированной системе наблюдений. При более широком изменении высокоавтоматизированных программно-аппаратных комплексов мониторинга паводковой ситуации существенно возрастает количество анализируемых и обрабатываемых программными средствами данных. Это, с одной стороны, усложнит применение традиционных методов использования данных, а с другой – повысит эффективность и востребованность предложенного в данной работе метода.

**Ключевые слова:** интеллектуальный анализ, сложные распределенные системы, прогнозирование угроз, программно-аппаратные системы, прогнозирование уровня воды.

Развитие новых программно-аппаратных систем для мониторинга состояния *сложных распределенных систем* (СРС), позволяющих оперативно получать и обрабатывать большое количество параметров, дает возможность использования (при соответствующей переработке и адаптации) хорошо зарекомендовавших себя моделей и методов управления техническими системами.

Авторы данной статьи имеют большой опыт разработки и использования статистических методов обработки большого количества измеряемых параметров для управления такими сложными техническими системами, как *газотурбинные двигатели* (ГТД). Анализ разви-

тия некоторых угроз в СРС, включающих в себя большое количество различных по своей природе и существенно удаленных друг от друга объектов, показал возможность применения этих методов анализа данных для формального описания зависимости определяющих состояние СРС параметров от наиболее значимых факторов и последующего прогнозирования угроз на основе выявленной зависимости.

В качестве примера рассмотрим один из наиболее часто встречающихся в Республике Башкортостан видов угроз безопасности населению и территории – весеннее половодье (чаще называемое паводком). Размеры затопленных территорий (их границы, площадь и

глубина) зависят от уровня воды в водных объектах, измеряемого на стационарных постах Росгидромета, а также на все более широко применяемых автоматических станциях наблюдения, принадлежащих местным органам власти. Высота подъема воды на каждом из этих постов наблюдения зависит от многих природных и техногенных факторов. К основным природным факторам относятся запасы воды в почве, глубина замерзания почвы, запасы воды в снеге, площадь и глубина снежного покрова, температура воздуха и другие метеопараметры, а также естественные ледовые заторы и лесные завалы. К техногенным факторам можно отнести плановый или аварийный сброс воды с гидротехнических сооружений, расположенных по течению выше точки измерения, строительство инженерных сооружений на самих водных объектах (мосты, водные переходы трубопроводов и др.) или вблизи них (дамбы, искусственные водоемы, набережные и др.), ледовые заторы (как следствие деятельности людей).

Большое распространение в рамках прогнозирования значений уровня воды на основе ретроспективных данных получили статистические методы, например, обобщенные регрессионные модели [1–3], метод наименьших квадратов [4], численные методы [5, 6]. Однако при этом возникает проблема заблаговременности прогнозирования, так как основным назначением данных методов является краткосрочное прогнозирование значений уровня воды, чего зачастую недостаточно для раннего проведения противопаводковых мероприятий. Если рассматривать менее распространенные методы (гидрологические [7–9]) краткосрочного и заблаговременного прогнозирования паводка, то необходимо отметить проблему ресурсоемкости: длительность расчетов значений уровня воды непозволительна, особенно в критические (пиковые) моменты паводковых ситуаций. Этими проблемами занимаются как отечественные, так и зарубежные ученые, однако работ, описывающих методы раннего прогнозирования паводка на основе интеллектуального анализа ретроспективных данных с применением *искусственной нейронной сети* (ИНС), недостаточно. Для более качественного прогнозирования уровня воды в кратчайшие сроки с использованием устаревших архитектур ИНС (например, многослойного персептрона [10]) становится актуальной реализация предлагаемого авторами данной статьи метода.

### Построение зависимости будущих возможных значений определяющих параметров от их изменения в прошлом

В работах [11, 12] предложен один из возможных подходов к прогнозированию будущих изменений уровня воды, основанный на применении ИНС для интеллектуального анализа измеренных значений только уровня воды. В данной статье рассматривается задача оперативной оценки изменения уровня воды (прогноз) на одни сутки вперед под воздействием наиболее значимых факторов на стационарных постах наблюдения. На каждом из этих постов ежедневно измеряется уровень воды  $h$ , и задача оперативного прогноза заключается в том, чтобы в конкретный момент времени измерения  $t_i$  для каждого поста определить будущее значение (на следующий момент времени  $t_{i+1}$ ) уровня воды, которое будем обозначать  $hp$ . Необходимо отметить, что прогнозируемое значение уровня воды  $hp_{i+1}$  отличается от реально измеренного через день значения  $h_{i+1}$ , поэтому для него вводится специальное обозначение:  $hp_{i+1} \neq h_{i+1}$ .

В данной работе предлагается определять будущее значение уровня воды на основе анализа его изменения в аналогичных условиях в прошлом, при этом предполагается, что основным фактором, влияющим на резкий подъем уровня воды (а именно он представляет угрозу объектам СРС), является стремительное потепление, то есть большой рост температуры воздуха за сутки. Другими словами – изменение уровня воды в конкретной точке измерения наиболее существенно зависит от того, насколько изменилась температура воздуха в этой точке. Учитывая, что современная система метеонаблюдений и прогнозирования погоды дает достаточно точный прогноз изменения температуры воздуха на 1–3 дня вперед, для выявления зависимости уровня воды от изменения температуры воздуха и последующего применения выявленной зависимости для прогнозирования уровня воды эти спрогнозированные значения можно использовать как фактическое изменение температуры воздуха.

Предлагается осуществлять выбор будущего значения уровня воды  $hp_{i+1}$  по результатам обработки накопленных за все предыдущие паводковые периоды данных о соответствии уровня воды и его изменения за сутки значениям температуры воздуха и ее изменениям за те же сутки [13, 14]. В качестве анализируемых данных используются измеренные в

равноотстоящие моменты времени  $t_i$  значения температуры воздуха  $T_i$  и уровень воды  $h_i$ . Так как прогноз заключается в определении будущего значения, то есть вычисляется величина изменения уровня воды  $hp_i = h_i + \Delta h_i$  в зависимости от изменения температуры  $TP_i = T_i + \Delta T_i$ , для реализации предлагаемого метода прогнозирования дополнительно определяются изменения уровня воды  $\Delta h_i$  и температуры  $\Delta T_i$ :

$$\begin{aligned} \Delta h_i &= h_{i+1} - h_i, \\ \Delta T_i &= T_{i+1} - T_i. \end{aligned} \quad (1)$$

Различные природные и техногенные факторы по-разному влияют на изменение уровня воды в точке измерения в соответствии с изменениями температуры в этой же точке за эти же сутки, и для прогнозирования необходимо определить статистическую зависимость  $\Delta h_i$  от соответствующих значений  $h_i$ ,  $T_i$ ,  $\Delta T_i$  и в дальнейшем использовать ее для определения будущих значений  $hp$ .

Представим все измеренные на каждом отдельном посту наблюдения за предыдущий период времени значения параметров  $h$ ,  $T$ ,  $\Delta h$ ,  $\Delta T$  в виде множества

$$W = \{W_i\}_{i=\overline{1,p}}, \quad (2)$$

каждый элемент которого  $W_i = \{h_i, T_i, \Delta h_i, \Delta T_i\}_{i=\overline{1,p}}$  представляет собой измеренные значения параметров в  $i$ -й момент времени  $t_i$ ;  $p$  – общее количество наблюдений.

Диапазоны возможных изменений каждого из этих параметров разбиваются на фиксированное число отрезков:

$$\begin{aligned} h^0, h^1, \dots, h^{M_1}, \\ T^0, T^1, \dots, T^{M_2}, \\ \Delta h^0, \Delta h^1, \dots, \Delta h^{M_3}, \\ \Delta T^0, \Delta T^1, \dots, \Delta T^{M_4}, \end{aligned} \quad (3)$$

где  $M_1, M_2, M_3, M_4$  – количество отрезков разбиения возможных значений соответствующего параметра.

На основе анализа данных многолетних наблюдений за паводковой ситуацией (то есть множеств (2) и (3)) строится новое множество:

$$N = \{N_{jklm}\}_{j=\overline{1,M_1}, k=\overline{1,M_2}, l=\overline{1,M_3}, m=\overline{1,M_4}}, \quad (4)$$

каждый элемент которого показывает количество элементов множества (2), удовлетворяющих следующим условиям:

$$\begin{aligned} h^{j-1} &< h_i \leq h^j, \\ T^{k-1} &< T_i \leq T^k, \\ \Delta h^{l-1} &< \Delta h_i \leq \Delta h^l, \\ \Delta T^{m-1} &< \Delta T_i \leq \Delta T^m \end{aligned} \quad (5)$$

при пробегании индексом  $i$  всех возможных значений,  $i = \overline{1,p}$ .

Другими словами, число  $N_{jklm}$  представляет собой частоту изменений уровня воды  $\Delta h$ , попавших в отрезок  $[\Delta h^{l-1}, \Delta h^l]$ , которые произошли при значениях параметров  $h$ ,  $T$ ,  $\Delta T$ , попавших соответственно в отрезки  $[h^{j-1}, \Delta h^j]$ ,  $[T^{k-1}, T^k]$ ,  $[\Delta T^{m-1}, \Delta T^m]$ .

На этапе прогнозирования в каждый конкретный момент времени  $t_i$  измеряются текущие значения параметров  $h_i$  и  $T_i$ . Как уже отмечалось, современные методы позволяют с приемлемой точностью прогнозировать изменение температуры воздуха, поэтому в этот момент времени известно прогнозируемое значение  $\Delta T p_i = T p_{i+1} - T_i$ , которое будем считать фактическим, то есть полагаем  $\Delta T_i = \Delta T p_i$ .

Далее определяются номера отрезков разбиения (3), в которые попали текущие значения  $h_i$ ,  $T_i$ ,  $\Delta T_i$ , то есть такие текущие значения индексов  $j_T$ ,  $k_T$ ,  $m_T$ , для которых

$$\begin{aligned} j_T : h_i &\in [h^{j_T-1}, h^{j_T}], \\ k_T : T_i &\in [T^{k_T-1}, T^{k_T}], \\ m_T : \Delta T_i &\in [\Delta T^{m_T-1}, \Delta T^{m_T}]. \end{aligned} \quad (6)$$

Из элементов множества  $N$  формируется новое множество  $N1 \subset N$ :

$$\begin{aligned} N1 &= \{N1_l\}_{l=\overline{1,M_3}}, \\ N1_l &= N_{j_T k_T m_T l}, l = \overline{1, M_3}, \end{aligned} \quad (7)$$

которое представляет собой множество частот появления  $\Delta h$  при значениях трех остальных параметров, удовлетворяющих соотношениям (6). В качестве прогнозируемого значения изменения уровня воды предлагается выбирать середину того отрезка разбиения из (3) по  $\Delta h$ , для которого частота появления такого значения  $\Delta h$  наибольшая, то есть в качестве текущего значения индекса  $l_T$  выбирается отрезок, удовлетворяющий условию

$$l_T : N1_{l_T} = \max N1_l, l = \overline{1, M_3}, \quad (8)$$

в качестве прогнозируемого значения изменения уровня воды выбирается

$$\Delta h p_i = (\Delta h^{l_T} - \Delta h^{l_T-1}) / 2, \quad (9)$$

а прогнозируемое значение уровня воды на следующий момент времени  $t_{i+1}$  определяется простым соотношением

$$hp_{i+1} = h_i + \Delta h p_i. \quad (10)$$

При наступлении времени следующего контроля уровня воды измеряются фактические значения  $T_{i+1}$ ,  $h_{i+1}$  и вычисляются фактические значения  $\Delta T_i$  и  $\Delta h_i$ , что позволяет корректировать значение одного элемента множества  $N$ ,

соответствующего отрезкам разбиения диапазонов изменения параметров (3), в которые попали фактические значения элемента  $(h_i, T_i, \Delta h_i, \Delta T_i)$ , путем увеличения его значения на единицу. Это означает, что частота появления четверки значений  $(h^{iT}, T^{kT}, \Delta h^{iT}, \Delta T^{mT})$ , таких, что

$$\begin{aligned} h_i &\in [h^{j_T-1}, h^{j_T}], \\ T_i &\in [T^{k_T-1}, T^{k_T}], \\ \Delta h_i &\in [h^{l_T-1}, h^{l_T}], \\ \Delta T_i &\in [\Delta T^{m_T-1}, \Delta T^{m_T}], \end{aligned} \tag{11}$$

увеличивается на единицу:

$$N_{j_T k_T l_T m_T} = N_{j_T k_T l_T m_T} + 1 \tag{12}$$

и при каждом следующем прогнозировании используется множество  $N$  с обновленными значениями его элементов, то есть в процессе паводковой ситуации процесс обучения модели прогнозирования продолжается.

### Экспериментальное исследование применимости предложенного метода

Для исследования применимости и практического использования предложенного метода разработан алгоритм обработки поступающих с измерительных станций данных, блок-схема которого представлена на рисунке 1. Алгоритм реализован в виде комплекса программ (структура приведена на рисунке 2), разработанного с использованием MS SQL Server, Python, C# и осуществляющего вычисление прогнозируемых значений и их передачу в систему поддержки принятия решений местных и региональных органов исполнительной власти по подготовке и парированию угроз от паводка.

Исследование проводилось по данным наблюдений за паводковой ситуацией в Республике Башкортостан в 2021 г. По каждому стационарному посту наблюдения Росгидро-

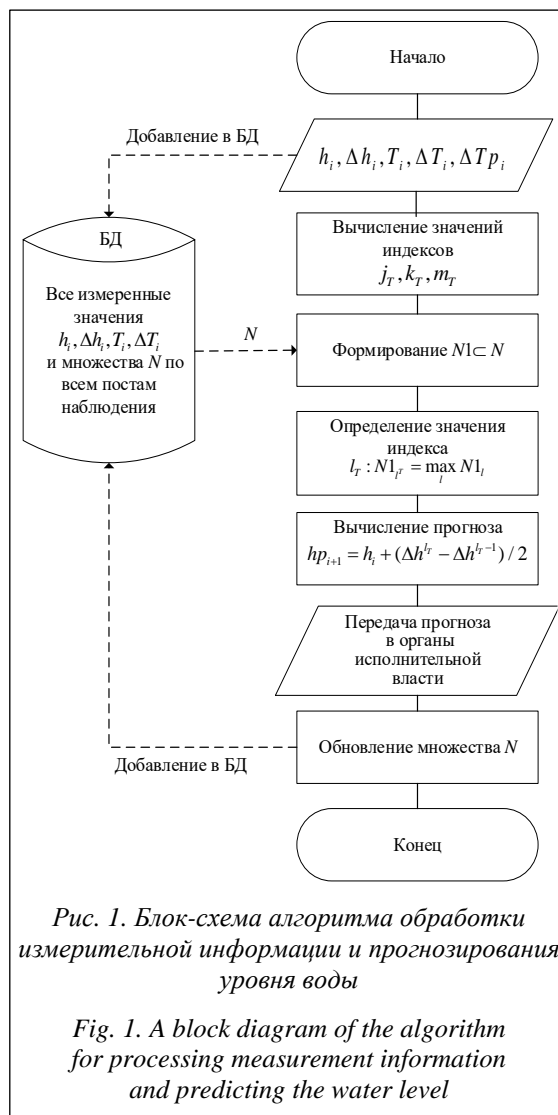


Рис. 1. Блок-схема алгоритма обработки измерительной информации и прогнозирования уровня воды

Fig. 1. A block diagram of the algorithm for processing measurement information and predicting the water level

мета (их в Башкортостане 41) на основе архивных данных наблюдения значений уровня воды и температуры воздуха (всего примерно по 12 тыс. значений каждого параметра) и вычисленных значений суточных изменений этих па-

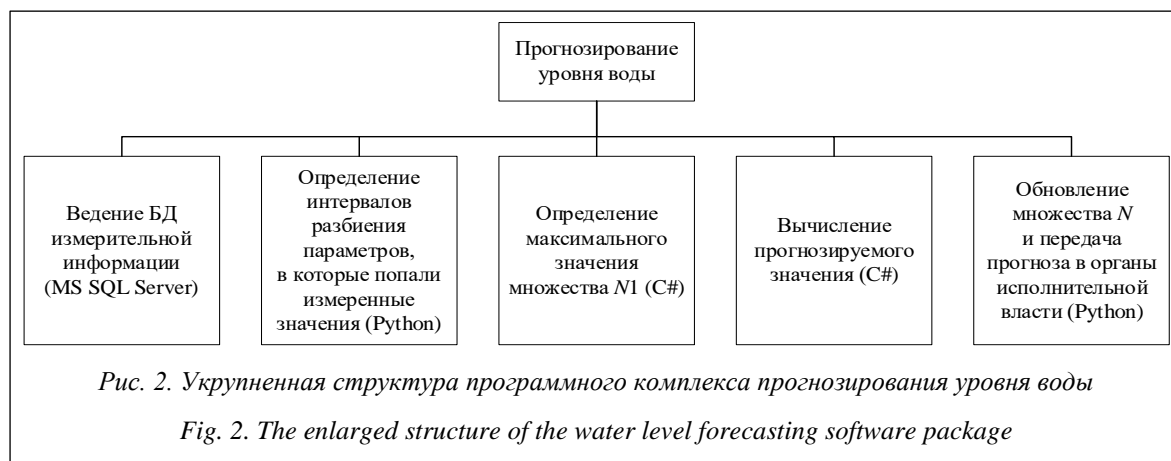


Рис. 2. Укрупненная структура программного комплекса прогнозирования уровня воды

Fig. 2. The enlarged structure of the water level forecasting software package

раметров (также примерно по 12 тыс. значений каждого из них) строилось множество  $N$  по приведенному выше алгоритму. Количество отрезков разбиения возможных значений каждого из параметров полагалось равным 10:  $M_1 = M_2 = M_3 = M_4 = 10$ . В каждый  $i$ -й день паводковой ситуации, начало которой характеризуется существенным подъемом уровня воды, на основе измеренных значений  $h_i$  и  $T_i$  и прогноза значения  $\Delta T p_i$  осуществлялись прогноз значения  $\Delta h p_i$  и вычисление  $h p_{i+1}$  по соотношению (10). Эти значения по каждому из постов наблюдения передавались в органы МЧС, где использовались для планирования и проведения мероприятий по парированию угрозы паводка для населения и территории (включая все инфраструктурные и промышленные объекты) [15, 16].

На следующий  $i+1$ -й день при получении фактического значения  $h_{i+1}$  вычислялась среднеквадратичная ошибка прогноза, а по завершении паводка по каждому посту наблюдения вычислялась средняя за весь паводок ошибка прогноза:

$$E = \frac{1}{p} \sum_{i=1}^p \frac{(h p_{i+1} - h_{i+1})^2}{h_{i+1}}, \quad (13)$$

где  $p$  – количество дней наблюдения за паводковой ситуацией. Прогнозные и фактические значения уровня воды на примере стационарного гидрологического поста 76417 «Метели, река Ай» представлены в таблице. Средняя ошибка прогноза составила 0,047, что хорошо согласуется с другими методами прогнозирования и применяется для поддержки принятия решений по планированию и проведению ме-

роприятий для парирования угроз от паводковой ситуации.

Как уже отмечалось, основным мотивом для применения методов интеллектуальной обработки больших массивов данных, показавших высокую эффективность при управлении сложными техническими системами (на примере авиационных газотурбинных двигателей), послужило все более широкое применение высокоавтоматизированных программно-аппаратных систем мониторинга развития опасных для населения и территорий процессов в СРС. Все это имеет самое прямое отношение и к контролю развития паводковой ситуации. На сегодняшний день основным источником информации для раннего обнаружения и парирования угроз при паводке являются стационарные посты гидрометслужбы, которые один раз в сутки измеряют необходимые параметры, причем зачастую неавтоматизированными методами и с низкой точностью. То есть для поддержки принятия решений используется относительно небольшое количество не вполне точных данных, и в этих условиях применение предлагаемых подходов ограничено малым количеством и низкой точностью имеющихся данных.

В текущем году проводилась опытная эксплуатация автоматических станций контроля паводковой ситуации, а в следующем планируется их ввод в промышленную эксплуатацию с постепенным выводом из эксплуатации ручных средств мониторинга. При этом осуществляется непрерывный динамический контроль всех необходимых параметров и возникает возможность их непрерывного использования аналогично тому, как это происходит уже достаточно давно для технических объектов и си-

### Прогнозные и фактические значения уровня воды

#### Forecasted and actual values of the water level

Дата прогноза	Прогнозная дата	Прогнозируемое значение уровня воды и ошибка, см					Фактический уровень воды, см
		на 5 дней	на 4 дня	на 3 дня	на 2 дня	на 1 день	
07.04.2021	На 12.04.2021	От 07.04.2021: 258/-198	От 08.04.2021: 388/-68	От 09.04.2021: 336/-120	От 10.04.2021: 309/-147	От 11.04.2021: 384/-72	456
08.04.2021	На 13.04.2021	От 08.04.2021: 381 / -53	От 09.04.2021: 335 / -99	От 10.04.2021: 311 / -123	От 11.04.2021: 380 / -54	От 12.04.2021: 411/-23	434
09.04.2021	На 14.04.2021	От 09.04.2021: 332 / -89	От 10.04.2021: 312 / -109	От 11.04.2021: 376 / -45	От 12.04.2021: 412/-9	От 13.04.2021: 425/+4	421
...							
30.04.2021	На 05.05.2021	От 30.04.2021: 267 / +23	От 01.05.2021: 263 / +19	От 02.05.2021: 259 / +15	От 03.05.2021: 251 / +7	От 04.05.2021: 239 / -5	244
01.05.2021	На 06.05.2021	От 01.05.2021: 260 / +11	От 02.05.2021: 258 / +9	От 03.05.2021: 255 / +6	От 04.05.2021: 249 / 0	От 05.05.2021: 239 / -10	249
Средняя ошибка, см		<b>75,32</b>	<b>62,64</b>	<b>51,92</b>	<b>35,56</b>	<b>16,6</b>	

стем. В этом случае количество используемых данных увеличится на несколько порядков (от одного значения в сутки до 24–240 и более), что обусловит еще большую эффективность и востребованность предлагаемых в данной статье методов.

Существенное увеличение количества доступных для анализа и обработки данных позволит повысить точность прогноза за счет увеличения  $M_j$ ,  $j = \overline{1,4}$ , так как это даст возможность уменьшить длину отрезка, определяемого соотношением (8) для вычисления прогнозируемого значения по соотношениям (9) и (10). При этом время, необходимое на собственно вычисление прогноза, практически не изменится даже при существенном увеличении количества анализируемых данных, так как собственно прогноз все равно вычисляется по соотношениям (8–10). Вычислительная нагрузка увеличится лишь на этапе обучения модели прогнозирования, то есть при вычислении элементов множеств  $N$  и  $N1$ , за счет проверки большого количества неравенств (5). Другими словами, время на обучение модели увеличится (в приведенном эксперименте оно составляло примерно 10 минут), но все равно будет меньше времени, необходимого для реализации мероприятий по парированию прогнозируемых угроз, то есть по-прежнему будет вполне приемлемым [17, 18].

## Заключение

Предложенный метод интеллектуального анализа и обработки больших разнородных данных о результатах контроля и прогнозирования развития угрозы от паводка в сложных распределенных системах, являющийся развитием предложенного авторами ранее адаптивного метода управления газотурбинными двигателями, показал достаточно высокую точность прогноза уровня воды в водоемах. Применение этого метода в составе программно-аппаратных систем соответствующими органами власти для раннего обнаружения и прогнозирования затопления территорий и расположенных на них объектов экономики и жизнедеятельности позволит более эффективно планировать и проводить мероприятия по парированию данного вида угроз и аналогичных им. Своевременность разработки и внедрения таких методов интеллектуального анализа больших массивов измерительной информации, а также их программной реализации подтверждается все более широким применением автоматических средств контроля состояния СРС и их отдельных объектов и подсистем, что приводит к экспоненциальному росту количества используемых для поддержки принятия решений данных.

*Работа выполнена при поддержке РФФИ, грант № 20-08-00301\20 – А.*

## Литература

1. Supriya P., Krishnaveni M., Subbulaksmi M. Regression analysis of annual maximum daily rainfall and stream flow for flood forecasting in vellar river basin. Aquatic Procedia, 2015, vol. 4, pp. 957–963. DOI: 10.1016/j.aqpro.2015.02.120.
2. He Y., Yan Y., Wang X., Wang C. Uncertainty forecasting for streamflow based on support vector regression method with fuzzy information granulation. Energy Procedia, 2019, vol. 158, pp. 6189–6194. DOI: 10.1016/j.egypro.2019.01.489.
3. Luo X., Yuan X., Zhu S., Xu Z., Meng L., Peng J. A hybrid support vector regression framework for streamflow forecast. J. of Hydrology, 2019, vol. 568, pp. 184–193. DOI: 10.1016/J.JHYDROL.2018.10.064.
4. Strupczewski W.G., Kaczmarek Z. Non-stationary approach to at-site flood frequency modelling II. Weighted least squares estimation. J. of Hydrology, 2001, vol. 248, no. 1, pp. 143–151.
5. Silvestro F., Rossi L., Campo L., Parodi A., Fiori E., Rudari R., Ferraris L. Impact-based flash-flood forecasting system: Sensitivity to high resolution numerical weather prediction systems and soil moisture. J. of Hydrology, 2019, vol. 572, pp. 388–402. DOI: 10.1016/J.JHYDROL.2019.02.055.
6. Hu R., Fang F., Salinas P., Pain C.C., Domingo N.D., Mark O. Numerical simulation of floods from multiple sources using an adaptive anisotropic unstructured mesh method. Advances in Water Resources, 2019, vol. 123, pp. 173–188. DOI: 10.1016/J.ADVWATRES.2018.11.011.
7. Tian J., Liu J., Yan D., Ding L., Li C. Ensemble flood forecasting based on a coupled atmospheric-hydrological modeling system with data assimilation. Atmospheric Research, 2019, vol. 224, pp. 127–137. DOI: 10.1016/J.ATMOSRES.2019.03.029.

8. Gomez M., Sharma S., Reed S., Mejia A. Skill of ensemble flood inundation forecasts at short- to medium-range timescales. *J. of Hydrology*, 2019, vol. 568, pp. 207–220. DOI: 10.1016/J.JHYDROL.2018.10.063.
9. Lopes V.A., Fan F.M., Pontes P.R., Siqueira V.A., Collischonn W., Marques D.M. A first integrated modelling of a river-lagoon large-scale hydrological system for forecasting purposes. *J. of Hydrology*, 2018, vol. 565, pp. 177–196. DOI: 10.1016/J.JHYDROL.2018.08.011.
10. Velasco L., Serqunia R., Zamad M., Juanico B., Lomosco J. Week-ahead rainfall forecasting using multilayer perceptron neural network. *Procedia Computer Science*, 2019, vol. 161, pp. 386–397. DOI: 10.1016/j.procs.2019.11.137.
11. 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.
12. 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, vol. 483, pp. 191–194. DOI: 10.2991/aisr.k.201029.036.
13. Noymanee J., Theeramunkong T. Flood forecasting with machine learning technique on hydrological modeling. *Procedia Computer Science*, 2019, vol. 156, pp. 377–386. DOI: 10.1016/j.procs.2019.08.214.
14. Han S., Coulibaly P. Bayesian flood forecasting methods: A review. *J. of Hydrology*, vol. 551, 2017, pp. 340–351. DOI: 10.1016/J.JHYDROL.2017.06.004.
15. Гундров Д.С. Автоматизированная система мониторинга и прогнозирования паводковой обстановки на водных объектах Краснодарского края и перспективы ее развития // Проблемы обеспечения безопасности при ликвидации последствий чрезвычайных ситуаций. 2018. № 1. С. 155–157.
16. Гребнев Я.В., Яровой А.В. Мониторинг и прогнозирование паводков на территории Красноярского края использованием нейросетевых алгоритмов // Сибирский пожарно-спасательный вестник. 2018. № 3. С. 13–16.
17. Bukvic A., Harrald J. Rural versus urban perspective on coastal flooding: The insights from the U.S. Mid-Atlantic communities. *Climate Risk Management*, 2019, vol. 23, pp. 7–18. DOI: 10.1016/j.crm.2018.10.004.
18. Zhou Y., Guo S., Chang F. Explore an evolutionary recurrent ANFIS for modelling multi-step-ahead flood forecasts. *J. of Hydrology*, 2019, vol. 570, pp. 343–355. DOI: 10.1016/J.JHYDROL.2018.12.040.

Software & Systems  
DOI: 10.15827/0236-235X.137.005-013

Received 07.10.21, Revised 06.12.21  
2022, vol. 35, no. 1, pp. 005–013

### **Intelligent analysis and processing large heterogeneous data for parrying threats in complex distributed systems**

*E.S. Brekotkina*<sup>1</sup>, Ph.D. (Economics), Associate Professor, [brekotkina@mail.ru](mailto:brekotkina@mail.ru)

*A.S. Pavlov*<sup>1</sup>, Ph.D. (Engineering), Associate Professor, [asp.gis@gmail.com](mailto:asp.gis@gmail.com)

*S.V. Pavlov*<sup>1</sup>, Dr.Sc. (Engineering), Professor, [psvgis@mail.ru](mailto:psvgis@mail.ru)

*O.I. Khristodulo*<sup>1</sup>, Dr.Sc. (Engineering), Professor, [o-hristodulo@mail.ru](mailto:o-hristodulo@mail.ru)

<sup>1</sup> Ufa State Aviation Technical University, Ufa, 450008, Russian Federation

**Abstract.** The paper proposes a method of intelligent analysis and processing of large heterogeneous data for predicting threats in complex distributed systems. The method is based on the results of automatic monitoring of changes in water level in water bodies and air temperature at the measurement point. Such monitoring makes it possible to increase the efficiency of planning and implementing measures to fend off such and similar threats.

The method is based on general approaches and mathematical models previously used by the authors to develop adaptive algorithms for controlling gas turbine engines. It is particularly relevant in the context of the increasingly widespread introduction of software and hardware systems for monitoring the state of complex distributed systems and the exponential growth in the number of data used to support decision-making.

The choice of the future value of the water level at the measurement point is based on the results of processing the data accumulated over all previous flood periods on the compliance of the water level and its changes per day with the values of air temperature and its changes over the same day. The analyzed data are the values of air temperature and water level measured at equidistant points in time, computational values of changes in the water level and air temperature, as well as forecast values (according to the official data of the hydrometeorological service) of changes in air temperature. Based on the calculation of the retrospective frequency of changes in this temperature and the water level at the corresponding point, it is proposed to choose as the predicted the value that corresponds to the maximum frequency of occurrence of such a combination of measured parameters.

The paper presents the results of an experimental assessment of the accuracy of forecasting the water level in the water bodies of the Republic of Bashkortostan in the flood period of 2021 are. They confirm the applicability of the proposed forecasting method to support decision making to fend off threats in complex distributed systems from a sharp rise in water, even with the current insufficiently automated observation system. With a wider change in highly automated software and hardware complexes for monitoring the flood situation, the amount of data analyzed and processed by software significantly increases, which will complicate the application of traditional methods of data use, and, on the other hand, will increase the efficiency and relevance of the method proposed in this paper.

**Keywords:** intelligent analysis, complex distributed systems, threat forecasting, software and hardware systems, water level forecasting.

*Acknowledgements.* The work has been supported by RFBR, grant no. 20-08-00301\20 – A.

### References

1. Supriya P., Krishnaveni M., Subbulaksmi M. Regression analysis of annual maximum daily rainfall and stream flow for flood forecasting in vellar river basin. *Aquatic Procedia*, 2015, vol. 4, pp. 957–963. DOI: 10.1016/j.aqpro.2015.02.120.
2. He Y., Yan Y., Wang X., Wang C. Uncertainty forecasting for streamflow based on support vector regression method with fuzzy information granulation. *Energy Procedia*, 2019, vol. 158, pp. 6189–6194. DOI: 10.1016/j.egypro.2019.01.489.
3. Luo X., Yuan X., Zhu S., Xu Z., Meng L., Peng J. A hybrid support vector regression framework for streamflow forecast. *J. of Hydrology*, 2019, vol. 568, pp. 184–193. DOI: 10.1016/J.JHYDROL.2018.10.064.
4. Strupczewski W.G., Kaczmarek Z. Non-stationary approach to at-site flood frequency modelling II. Weighted least squares estimation. *J. of Hydrology*, 2001, vol. 248, no. 1, pp. 143–151.
5. Silvestro F., Rossi L., Campo L., Parodi A., Fiori E., Rudari R., Ferraris L. Impact-based flash-flood forecasting system: Sensitivity to high resolution numerical weather prediction systems and soil moisture. *J. of Hydrology*, 2019, vol. 572, pp. 388–402. DOI: 10.1016/J.JHYDROL.2019.02.055.
6. Hu R., Fang F., Salinas P., Pain C.C., Domingo N.D., Mark O. Numerical simulation of floods from multiple sources using an adaptive anisotropic unstructured mesh method. *Advances in Water Resources*, 2019, vol. 123, pp. 173–188. DOI: 10.1016/J.ADVWATRES.2018.11.011.
7. Tian J., Liu J., Yan D., Ding L., Li C. Ensemble flood forecasting based on a coupled atmospheric-hydrological modeling system with data assimilation. *Atmospheric Research*, 2019, vol. 224, pp. 127–137. DOI: 10.1016/J.ATMOSRES.2019.03.029.
8. Gomez M., Sharma S., Reed S., Mejia A. Skill of ensemble flood inundation forecasts at short- to medium-range timescales. *J. of Hydrology*, 2019, vol. 568, pp. 207–220. DOI: 10.1016/J.JHYDROL.2018.10.063.
9. Lopes V.A., Fan F.M., Pontes P.R., Siqueira V.A., Collischonn W., Marques D.M. A first integrated modelling of a river-lagoon large-scale hydrological system for forecasting purposes. *J. of Hydrology*, 2018, vol. 565, pp. 177–196. DOI: 10.1016/J.JHYDROL.2018.08.011.
10. Velasco L., Serqunia R., Zamad M., Juanico B., Lomosco J. Week-ahead rainfall forecasting using multilayer perceptron neural network. *Procedia Computer Science*, 2019, vol. 161, pp. 386–397. DOI: 10.1016/j.procs.2019.11.137.
11. 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. *Herald of Computer and Information Technologies*, 2021, vol. 18, no. 2, pp. 39–45 (in Russ.).
12. 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, vol. 483, pp. 191–194. DOI: 10.2991/aisr.k.201029.036.



13. Noymanee J., Theeramunkong T. Flood forecasting with machine learning technique on hydrological modeling. *Procedia Computer Science*, 2019, vol. 156, pp. 377–386. DOI: 10.1016/j.procs.2019.08.214.
14. Han S., Coulibaly P. Bayesian flood forecasting methods: A review. *J. of Hydrology*, vol. 551, 2017, pp. 340–351. DOI: 10.1016/J.JHYDROL.2017.06.004.
15. Gundrov D.S. Automated system for monitoring and forecasting the flood situation on the water bodies of the Krasnodar Territory and prospects for its development. *Problems of Ensuring Safety in the Elimination of the Emergency Situation Consequences*, 2018, no. 1, pp. 155–157 (in Russ.).
16. Grebnev Y.V., Yarovoy A.V. Control and prediction of floods on the territory of the Krasnoyarsk kray through the use of neural network algorithms. *Siberian Fire and Rescue Bulletin*, 2018, no. 3, pp. 13–16 (in Russ.).
17. Bukvic A., Harrald J. Rural versus urban perspective on coastal flooding: The insights from the U.S. Mid-Atlantic communities. *Climate Risk Management*, 2019, vol. 23, pp. 7–18. DOI: 10.1016/j.crm.2018.10.004.
18. Zhou Y., Guo S., Chang F. Explore an evolutionary recurrent ANFIS for modelling multi-step-ahead flood forecasts. *J. of Hydrology*, 2019, vol. 570, pp. 343–355. DOI: 10.1016/J.JHYDROL.2018.12.040.

#### Для цитирования

Брекоткина Е.С., Павлов А.С., Павлов С.В., Христовуло О.И. Интеллектуальный анализ и обработка больших разнородных данных для парирования угроз в сложных распределенных системах // Программные продукты и системы. 2022. Т. 35. № 1. С. 005–013. DOI: 10.15827/0236-235X.137.005-013.

#### For citation

Brekotkina E.S., Pavlov A.S., Pavlov S.V., Khristodulo O.I. Intelligent analysis and processing large heterogeneous data for parrying threats in complex distributed systems. *Software & Systems*, 2022, vol. 35, no. 1, pp. 005–013 (in Russ.). DOI: 10.15827/0236-235X.137.005-013.