

ВОДНЫЕ РЕСУРСЫ И РЕЖИМ ВОДНЫХ ОБЪЕКТОВ

УДК 556.06

ПРОГНОЗИРОВАНИЕ УРОВНЕЙ ВОДЫ НА РЕКАХ КРАСНОДАРСКОГО КРАЯ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ¹

© 2022 г. П. А. Белякова^{a, b, *}, В. М. Морейдо^a, А. С. Цыпленков^c, А. Н. Амербаев^a,
Д. А. Гречишников^c, Л. С. Курочкина^d, В. А. Филиппов^e, М. С. Макеев^f

^aИнститут водных проблем РАН, Москва, 119333 Россия

^bТихоокеанский институт географии ДВО РАН, Владивосток, 690041 Россия

^cМосковский государственный университет им. М.В. Ломоносова, Москва, 119991 Россия

^dГосударственный гидрологический институт, Санкт-Петербург, 199004 Россия

^eСпециализированный центр по гидрометеорологии и мониторингу
окружающей среды Черного и Азовского морей, Сочи, 354057 Россия

^fООО Эмерсит, Краснодар, 350001 Россия

*e-mail: pobel@mail.ru

Поступила в редакцию 16.08.2021 г.

После доработки 17.08.2021 г.

Принята к публикации 18.08.2021 г.

Рассмотрена возможность построения моделей машинного обучения для прогнозирования уровней воды на участках горных рек Краснодарского края на основе данных наблюдений за уровнями воды на автоматизированных гидрологических комплексах Автоматизированной системы мониторинга паводковой ситуации Краснодарского края. В качестве объектов исследования выбраны две горные реки Краснодарского края – Пшиш и Мзымта. Они протекают в различных природных условиях, отличаются по водному режиму и характеру боковой приточности на выбранных участках. Исследованы модели на базе трех распространенных архитектур машинного обучения – регрессионной модели деревьев решений М5Р, градиентного бустинга деревьев решений XGBoost и искусственной нейронной сети на основе многослойного перцептрона. Эффективность прогнозирования проверяется для заблаговременности от 1 до 20 ч, обсуждаются различия для рек с разным водным режимом и возможности использования исследованных моделей для оперативного прогнозирования. Для р. Пшиш определена оптимальная заблаговременность прогнозирования 15–18 ч ($S/\sigma_{\Delta} = 0.38–0.39$ для модели XGBoost), для р. Мзымты качество моделирования оценивается как хорошее, однако без достижения необходимой прогностической эффективности (при заблаговременности 5 ч $S/\sigma_{\Delta} = 0.87$ для модели MLP). Полученные результаты позволяют оценить модели машинного обучения как пригодные для задач краткосрочного гидрологического прогнозирования на основе высокочастотных данных наблюдений за уровнями воды.

Ключевые слова: гидрологические прогнозы, машинное обучение, искусственные нейронные сети, быстроразвивающиеся паводки, краткосрочные прогнозы уровней воды.

DOI: 10.31857/S0321059622010047

ВВЕДЕНИЕ

Наводнения относятся к числу наиболее разрушительных стихийных бедствий в России и в мире, а их прогноз – одна из ключевых составля-

ющих системы защиты населения и хозяйства от наводнений. В России наводнения почти ежегодно уносят жизни людей, причем с начала 2000-х гг. более 80% людей погибли от наводнений дождевого происхождения, более 50% – в ходе быстроразвивающихся паводков [2]. Для обеспечения необходимой заблаговременности прогноза для реагирования чрезвычайных служб и оповещения населения при угрозе наводнения необходимо развитие систем гидрологического прогнозирования, которое обеспечивается наличием данных мониторинга, современных прогностических методик и квалифицированного персонала [6, 29].

¹ Обработка данных и моделирование выполнены при финансовой поддержке РФФИ и Краснодарского края в рамках научного проекта 19-45-233007, обработка ЦМР – в рамках проекта РФФИ 19-05-00353, интерпретация результатов выполнена в рамках темы 147-2019-0001 (государственная регистрация АААА-А18-118022090056-0) Государственного задания ИВП РАН. А.С. Цыпленков выполнял исследование в рамках Программы развития Междисциплинарной научно-образовательной школы Московского государственного университета им. М.В. Ломоносова “Будущее планеты и глобальные изменения окружающей среды”.

Ожидается, что благодаря климатическим изменениям частота и интенсивность опасных метеорологических явлений, в том числе и экстремальных ливневых осадков, будет увеличиваться во многих регионах мира, и это делает проблему прогнозирования наводнений чрезвычайно трудной на всех временных масштабах [22, 25, 27]. В России также в последние годы отмечается рост вклада сильных ливней в общую сумму осадков [21].

Существенный прогресс в задачах прогнозирования быстроразвивающихся паводков может быть достигнут путем увеличения доступности данных (от измерительных станций, населения и дистанционного зондирования) и методов их оптимального использования, чтобы углубить понимание гидрометеорологических явлений и их социально-экономических последствий [38]. В Краснодарском крае с 2013 г. действует уникальная для России региональная автоматизированная система мониторинга паводковой ситуации (АСМПС КК) [12], которая включает в себя более 250 автоматизированных гидрологических комплексов (АГК) и более 90 автоматических метеорологических станций (АМС). Автоматизированные посты измеряют уровень воды на реках с частотой 1 раз в 10 мин, а также некоторые метеорологические характеристики и в режиме реального времени передают значения на информационный сервер системы по каналам сотовой связи на базе технологий ООО “Эмерсит” [12]. К достоинствам системы относится высокая плотность сети мониторинга (<170 км² на 1 АГК в горной части края), высокая частота передачи данных, а также их открытость через сеть Интернет, что подтверждает ее значение в обеспечении оперативного мониторинга паводков и предупреждения населения о наводнениях. АСМПС КК вписывается в международную практику создания локальных гидрометеорологических систем для решения проблемы мониторинга и прогнозирования быстроразвивающихся паводков [46]. В то же время массив высокочастотных данных, полученных региональной системой мониторинга, может быть использован для развития методов моделирования и прогнозирования быстроразвивающихся паводков. Несмотря на сравнительно краткий период работы системы “Эмерсит”, результаты измерений уже активно используются исследователями для изучения гидрологических и геоморфологических явлений [5, 13, 14].

Достижения в области искусственного интеллекта (ИИ) и машинного обучения (МО) [23, 31, 32, 40] предлагают многообещающие методы с целью использования всего потенциала больших объемов данных для анализа и прогнозирования с целью предоставления точной и своевременной информации для заинтересованных организаций (“stakeholders”), населения и органов власти.

В задачах краткосрочного гидрологического прогнозирования зачастую не требуется детальное описание процессов формирования речного стока, а важно точное описание реакции речной системы на входные метеорологические воздействия [17]. Для этого успешно используются упрощенные модели, связывающие предшествующие наблюдения за речным стоком и метеовеличинами в пределах речного водосбора с будущими величинами речного стока путем эффективного учета (авто)корреляции временных рядов [4].

Задача прогнозирования уровней воды для горных рек Краснодарского края очень актуальна в свете регулярных наводнений. Уровни воды – более точно и легко измеряемая характеристика по сравнению с расходами воды, особенно на горных реках, а прогноз уровней на нижнем посту участка с учетом наблюдений на верхнем – традиционно решаемая гидрологическая задача. Цель этой работы – построение моделей машинного обучения для прогнозирования уровней воды на реках Пшиш (у ст. Гурийской) и Мзымте (у с. Казачий Брод) Краснодарского края с учетом уровней воды на постах, расположенных выше по течению. В качестве исходных данных впервые привлекаются высокочастотные наблюдения за уровнями воды сети АСМПС КК. Исследуются модели на базе трех распространенных архитектур машинного обучения – регрессионной модели деревьев решений M5P [37], градиентного бустинга деревьев решений XGBoost [20] и искусственной нейронной сети на основе многослойного перцептрона (здесь и далее MLP [24]). Эффективность прогнозирования проверяется для заблаговременности от 1 до 20 ч.

ОБЪЕКТ ИССЛЕДОВАНИЯ

Водосборы рек Пшиш и Мзымты (рис. 1) расположены на разных макросклонах Западного Кавказа. Эти реки выбраны благодаря своим близким размерам (табл. 1) и контрастным природным условиям на водосборах. Пшиш относится к притокам среднего течения Кубани и впадает в Краснодарское водохранилище. Бассейн Мзымты расположен на юге российской части Черноморского побережья Кавказа, река впадает в Черное море. Верховья р. Мзымты находятся западнее, в более высокой части Большого Кавказа, максимальные высоты в бассейне >3250 м, а на водоразделах р. Пшиш – только 1800 м. Рельеф и высотное положение водосборов определяют климатические особенности, режим снегонакопления, а также водный режим рек.

Бассейн р. Пшиш находится в зоне умеренного климата с годовой суммой осадков >900 мм (г. Хадыженск – 950 мм, х. Горный – 1660 мм), снежный покров неустойчивый и образуется не каждую зиму. Бассейн р. Мзымты относится к зо-

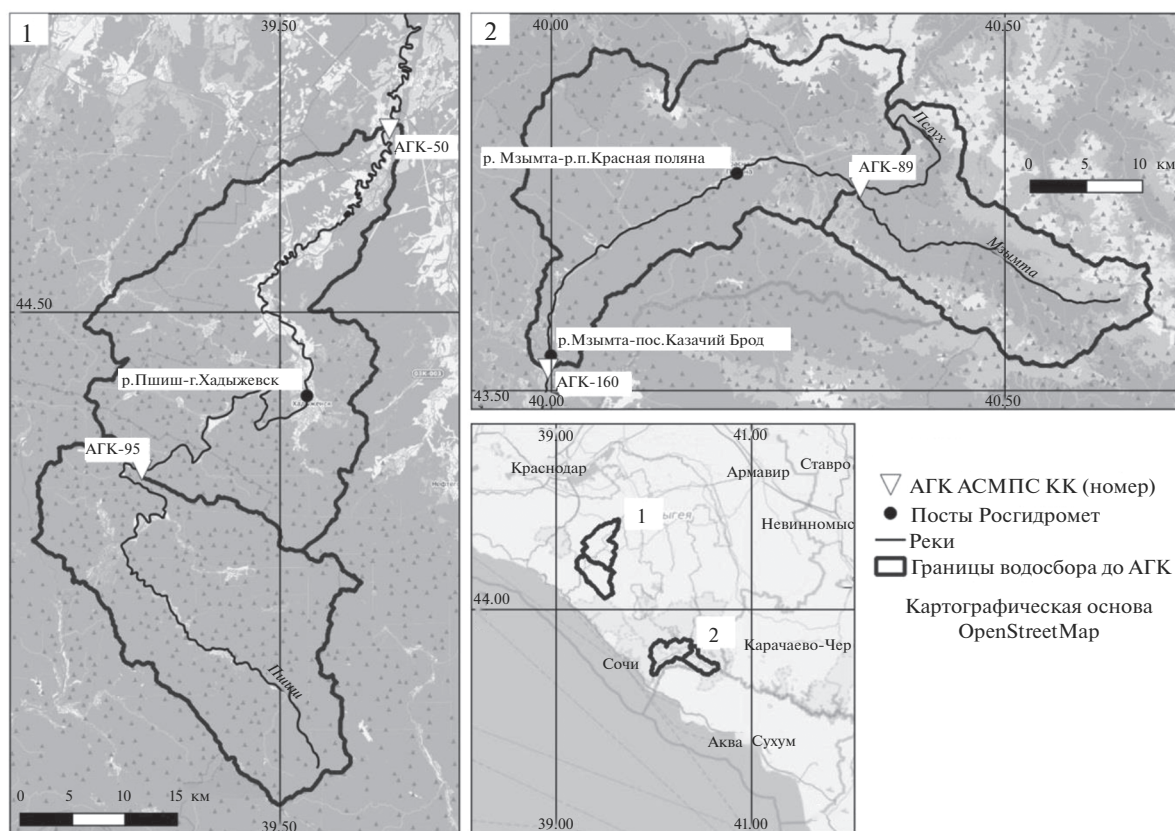


Рис. 1. Картограмма водосборов рек Пшиш (1) и Мзымты (2) и расположения выбранных АГК.

не влажного субтропического климата, годовое количество осадков >1300 мм (г. Адлер – 1377 мм, пгт Красная Поляна – 1795 мм) [10]. Часть водо-

сбора выше 600–1300 м относится к зоне переменного снежного покрова, где снег выпадает ежегодно и продолжительность залегания состав-

Таблица 1. Характеристики рек Пшиш и Мзымты и их водосборов к выбранным створам АГК Автоматизированной системы мониторинга паводковой ситуации Краснодарского края

Характеристики	р. Пшиш – с. Навагинское	р. Пшиш – ст. Гурийская	р. Мзымта – выше слияния с р. Пслух	р. Мзымта – с. Казачий Брод
	АГК-95	АГК-50	АГК-89	АГК-160
Расстояние от устья, км	167.5	81	54	14
Площадь водосбора, км ²	546	1240	299	855
Средняя высота водосбора, м	551	378	1884	1343
Максимальная высота водосбора, м	1839 (г. Шесси)		3257 (г. Агепста)	
Уровень датчика, м БС	187.561	75.302	648.824	75.390
Доля выбросов* в ряду, %	0.23	0.05	0.24	0.11
Доля пропусков в ряду, %	3.33	2.29	4.95	7.77
Максимальный подъем уровня воды, м (дата)	6.27 (24.10.2018)	9.33 (25.10.2018)	2.22 (30.10.2017); 2.03 (25.10.2018)	3.11 (24.10.2018)

* К выбросам относятся заведомо ошибочные измерения уровня воды, связанные с недостатками технологии и калибровкой датчиков. Пояснения в тексте.

ляет >120 дней, что обеспечивает работу горнолыжных курортов на склонах долины р. Мзымты.

Согласно классификации П.С. Кузина [7], р. Пшиш относится к рекам с паводками в течение всего года, преобладающими в холодное время года, а р. Мзымта — к рекам с весенне-летним половодьем и паводками в течение всего года. Средние максимальные за год подъемы на р. Пшиш составляют 2.9 м у АГК-95 и 6.1 м у АГК-50. Средняя интенсивность роста уровней в ходе паводков составляет 0.1–0.6, максимальная — 1.1 м/ч. Время добегания между соответствующими пиками высоких паводков на р. Пшиш составляет от 18 до 28 ч, что соответствует скорости движения волны на этом участке 3–4 м/с.

Диапазон колебания уровней на р. Мзымте в ходе паводочных событий составляет 2–3 м. В верхнем течении интенсивность подъема уровней обычно ниже — 0.05–0.24 м/ч, в нижнем течении — до 0.5 м/ч. При этом в нижнем течении формирование паводков часто начинается раньше в силу того, что влагоперенос осуществляется со стороны моря [1]. Пиковые уровни тоже могут наступить раньше на расположенном ниже посту. Очень показателен в этом отношении паводок 25 июня 2015 г., когда вслед за ливнями развивались наводнения на реках в прибрежной зоне и низкогорье, в Казачьем Броду пик прошел в 13 ч, а воздушные массы поднялись в горную часть спустя сутки и сформировали пик паводка у слияния рек Мзымты и Пслух в 10 ч 26 июня.

Несмотря на сравнительно краткий период работы АСМПС КК, датчиками уже зафиксированы рекордные подъемы уровней воды. Так, в ходе наводнения 24–25 октября 2018 г. рост уровня на р. Пшиш составил >6.2 м в районе с. Навагинского и 9.3 м у ст. Гурийской. По данным гидропостов Росгидромета, в г. Хадыженске и ст. Бжедуховской это был наивысший паводок за период наблюдений с 1936 г. В те же даты — 24–25 октября 2018 г. — развивался высокий паводок на р. Мзымте с подъемами уровня на 2.0–3.1 м выше слияния с р. Пслух и у с. Казачий Брод соответственно. В районе Казачьего Брода это был второй по величине паводок после 25 октября 2003 г. (или третий с учетом наивысшего паводка, наблюдаемого на гидрологическом посту Кепша 26 июня 1956 г.) за период инструментальных измерений.

Гидрометеорологическая изученность

На выбранных реках наблюдения на сети АСМПС КК ведутся параллельно наблюдениям на гидрологических постах Росгидромета, что позволяет проводить взаимный контроль качества получаемых данных. В рамках региональной сети АСМПС КК работают 3 АГК на самой

р. Мзымте и 3 АГК на притоках. На р. Пшиш установлены 6 АГК и 2 АГК на притоках. В рамках настоящего исследования выбраны по 2 пары АГК АСМПС КК на каждой реке с наиболее полными (малое количество пропусков измерений и выбросов) рядами наблюдений (табл. 1). Расстояние между выбранными постами составляет 86.5 км по течению р. Пшиш и 40 км по течению р. Мзымты.

Обработка данных измерений уровней воды, удаление выбросов и настройка фильтра для снижения зашумленности рядов

Для исследования использован архив данных наблюдений сети АСМПС КК, содержащий уровни воды с апреля 2014 г. по январь 2021 г., измеренные через каждые 10 мин на АГК-95, АГК-50, АГК-89 и АГК-160.

В связи с конструктивными особенностями радарных датчиков-уровнемеров, высокой частотой измерений, а также характером потока в местах измерений ряды отметок уровней воды насыщены выбросами и шумами. Большинство существующих методик анализа временных рядов на выпадающие значения, выбросы или аномалии не подходят для анализа хода уровня воды малых горных рек из-за особенностей их гидрологического режима. Резкие и быстрые подъемы уровня воды, характерные для рек Западного Кавказа, известными авторам статистическими тестами (статистический критерий Шовене, тест Граббса, критерий Пирса, Q-тест Диксона) отклоняли нулевую гипотезу о том, что проверяемые значения не являются выбросами. Поэтому для удаления выпадающих значений, связанных с ошибками измерения, разработан следующий алгоритм. В первую очередь проводилось восстановление регулярного шага измерений 10 мин. Далее проводился анализ ряда на наличие выбросов путем обнаружения событий, при которых уровень за 10 мин повышался на >1.5 м, а затем снижался на >40 см. Доля выбросов на выбранных АГК варьирует от 0.05 (18–25 значений, связанных с калибровкой датчиков, по информации ООО “Эмерсит”) до 0.24%. За 7 лет работы датчиков доля пропусков составляет от 2.29 до 7.77% длины ряда. Обнаруженные ошибочные данные удалялись, отсутствующие значения заменялись линейно интерполированными.

После определения выбросов и восстановления регулярного шага измерений данные были отфильтрованы от высокочастотных шумов, свойственных самописцам такого типа. Принятые в гидрологии методы линейной и нелинейной фильтрации [39] не подходят для этой задачи, так как сильно уменьшают фактический размах колебаний уровня во время паводков. В связи с этим применен метод сглаживания Савитцки—

Голей [42], который широко используется для обработки высокочастотных данных наблюдений в различных областях [28, 45]. Отличительная черта данного фильтра – то, что он позволяет сгладить шумы и сохранить фактическую амплитуду колебаний [36]. Сглаживание происходит с окном в 41 значение при помощи полинома 8-й степени. После процедуры сглаживания была применена линейная интерполяция, чтобы заполнить небольшие пропуски <8 ч, образовавшиеся из-за удаления выбросов и из-за некачественных измерений. Это необходимо для обеспечения наиболее полных и непрерывных рядов для дальнейшего моделирования.

Подготовка рядов для экспериментов

Для проведения вычислительных экспериментов были подготовлены матрицы данных с 2014 по 2020 г. Расчеты выполнялись с шагом 1 ч, поэтому предварительно данные были приведены к часовому шагу. Для каждого часа выполнялся прогноз уровней воды на определенную заблаговременность на основе информации, известной на момент выдачи прогноза. Матрицы включали в себя набор из 54 предикторов – рядов уровней воды, полученных в результате обработки исходной информации АСМПС КК. В тех же матрицах содержалось 8 предикатов – рядов прогнозируемой характеристики (уровней воды) – с заблаговременностью от 1 до 20 ч (“будущие” уровни воды L_{t+1} , L_{t+2} , L_{t+3} , L_{t+5} , L_{t+10} , L_{t+15} , L_{t+18} , L_{t+20}).

В число предикторов вошли:

уровни воды на верхнем створе $H_{t-\tau}$ (м) на час выпуска прогноза и за предысторию $\tau = 1 \dots 48$ ч;

уровни воды на нижнем створе $L_{t-\tau}$ (м) на час выпуска прогноза и за предысторию $\tau = 1 \dots 24$ ч.

Формирование таких матриц позволяет использовать для прогнозирования закономерности руслового добега паводочной волны по руслу реки. Матрицы исходных данных организованы построчно. В первой позиции каждой строки расположены “текущая” дата и час – дата и час выдачи экспериментального прогноза t .

Далее ряды уровней воды были разделены на обучающий и проверочный наборы в соответствии с рекомендациями [18] с достаточной длиной и сходными статистическими характеристиками. Обучающий период для р. Пшиш назначен с 03.03.2015 по 31.08.2018 (70% данных), проверочный – с 01.09.2018 по 05.01.2020 (30%). Аналогичный период выбран для Мзымты с 01.01.2014 по 31.08.2018 (обучение, 80%) и с 01.09.2018 по 05.01.2020 (проверка, 20% из-за наличия пропусков).

МОДЕЛИ И МЕТОДЫ ИХ ПОСТРОЕНИЯ

Описание модели M5P

Алгоритм M5P базируется на алгоритме построения деревьев регрессионных моделей M5 [37]. Данный алгоритм включает в себя три этапа. На первом этапе строится дерево принятия решений, где критерий расщепления минимизирует изменчивость значений предиката в каждой из ветвей. На втором этапе выполняется отсечение на каждом из листьев. При отсечении внутренние узлы дерева образуют плоскость регрессии. На третьем этапе для предотвращения появления разрывов применяется процедура сглаживания. Оно объединяет предсказания модели в листе со значениями для каждого узла на пути от листа к корню дерева принятия решений. Достоинство алгоритма – компактность и относительная легкость интерпретации получаемых моделей. Алгоритм использует ненормированные исходные данные.

Описание модели XGBoost

XGBoost – это градиентный бустинг деревьев решений. Бустинг – это техника, позволяющая строить композицию алгоритмов последовательно, при этом каждый следующий алгоритм строится таким образом, чтобы минимизировать ошибки предыдущих. В качестве базовых алгоритмов композиции используются неглубокие деревья решений.

Пусть задана композиция из $N - 1$ деревьев решений:

$$a_{N-1}(x) = \sum_{n=1}^{N-1} b_n(x). \quad (1)$$

На следующем шаге необходимо добавить алгоритм $b_N(x)$, который уменьшит ошибку композиции на обучающей выборке. Иными словами, следует определить, какие значения s_1, \dots, s_l должен принимать алгоритм $b_N(x_i) = s_i$ на объектах обучающей выборки, чтобы ошибка всей композиции была минимальной:

$$s = \underset{r}{\operatorname{argmin}} E(r) = \underset{r}{\operatorname{argmin}} \sum_{i=1}^l L(y_i, a_{N-1}(x_i) + r_i). \quad (2)$$

Тогда вектор s можно искать как антиградиент функции ошибки:

$$s = -\nabla E. \quad (3)$$

Таким образом, обучение алгоритма $b_N(x)$ представляет собой минимизацию квадратичной функции ошибки ответов алгоритма и элементов вектора s .

Модель написана на языке “Python”. Для построения использовалась библиотека “xgboost”

версии 0.90. Для валидации модели применялась библиотека “sklearn”. Исходными данными для модели служили нормированные величины в диапазоне $[-1; 1]$.

Описание модели MLP

Наиболее распространенная модель искусственной нейронной сети (ИНС), применяемой для гидрологических задач [16, 33, 43], – многослойный перцептрон MLP (“multilayer perceptron”) [24]. Многослойный перцептрон представляет собой вычислительную сеть из отдельных искусственных нейронов – моделей нейронов головного мозга. Математическая модель нейрона впервые предложена в работе [30], но получила широкое развитие только после разработки алгоритма обратного распространения ошибки [41] для определения параметров модели (“весов” входных переменных). Модель многослойного перцептрона основана на использовании одного или нескольких слоев искусственных нейронов, принимающих на вход большое число рядов данных наблюдений (например, за уровнем воды с различным сдвигом во времени) и трансформирующих их в выходной сигнал (например, уровни воды заданной заблаговременности) путем нелинейных преобразований и комбинаций входных сигналов с различными “весами”. Математическое представление искусственного нейрона – зависимость следующего вида:

$$y = \varphi \left(\sum_{i=1}^n w_i x_i + b \right), \quad (4)$$

которая связывает выходной сигнал y с вектором входных данных x_i ($i = 1, 2, \dots, n$), входящих в модель с различными весами w_i , порогом активации нейрона b через функцию активации φ . В качестве функции активации обычно используется нелинейная функция сигмиды следующего вида:

$$\varphi = \frac{1}{1 + e^{-sx}}. \quad (5)$$

Определение весов, с которыми векторы исходных данных входят в искусственную нейронную сеть, основано на сопоставлении выходного сигнала с обучающим вектором (данными наблюдений) и обратным распространением ошибки, которое приводит к повторному взвешиванию. Этот процесс называется “обучением” нейросети. В итоге обучение должно обеспечивать минимизацию выбранной целевой функции, в качестве которой, как правило, используется среднеквадратическое отклонение (но может быть выбрана любая другая).

Нелинейный характер связей, выстраиваемых между входом и выходом ИНС, а также использование любых входных данных для автоматизиро-

ванного поиска модельной структуры позволили использовать ИНС в различных задачах, не только в области гидрологии. С точки зрения класса гидрологических моделей ИНС представляет собой классическую модель “черного ящика” [19]. Такая модель не требует детального описания моделируемой системы (в данном случае – процессов формирования стока в пределах речного бассейна), чтобы распознать взаимосвязи между входами и выходами. Для различных сложных нелинейных систем ИНС имеют преимущество перед моделями с распределенными параметрами в том, что требования к входным данным обычно меньше. Исходными данными для модели служат нормированные величины в диапазоне $[-1; 1]$.

Процесс разработки модели ИНС заключается в следующем. Измеренные данные, используемые для разработки модели, делятся на две группы: обучение и тестирование. Сначала модель ANN обучается на первом наборе и результат моделирования сопоставляется с данными наблюдений для определения качества моделирования. Далее модель проверяется на независимом наборе данных и ставится условие на допустимое снижение оценки качества модели. В итоге разработанная модель позволяет получить выходной сигнал при любом количестве исходных данных.

Для разработки модели прогнозирования хода уровней воды на основе данных измерений АГК АСМПС КК использованы программные средства библиотеки “scikit-learn” языка “Python” [35]; для решения регрессионных задач с обучением использована программная реализация “MLPRegressor”. Основные параметры модели – количество скрытых слоев модели, количество нейронов в них, вид активационной функции, алгоритм поиска глобального минимума оптимизационной функции и максимальное количество итераций. Поиск оптимального набора параметров проводился с помощью функции сеточного поиска с десятикратной перекрестной проверкой (обучающий набор данных случайным образом делится на 10 частей, на 9/10 происходит настройка параметров, на 1/10 – проверка, затем деление набора данных проводится заново, так продолжается 10 раз). Указанные процедуры также реализованы в виде функций библиотеки “scikit-learn”.

Методы оценки качества моделей

Для оценки результатов расчетов использованы следующие метрики. Среднеквадратическая ошибка смоделированных значений относительно фактических рассчитывалась так:

$$S = \sqrt{\frac{\sum_{i=1}^N (H_i^{\text{sim}} - H_i^{\text{obs}})^2}{N}}, \quad (6)$$

здесь H_i^{sim} – смоделированный уровень воды, м; H_i^{obs} – наблюдаемый уровень воды; N – длина оцениваемого ряда.

Критерий эффективности гидрологических моделей Нэша–Сатклиффа NSE [34]:

$$NSE = 1 - \frac{\sum_{i=1}^N (H_i^{\text{sim}} - H_i^{\text{obs}})^2}{\sum_{i=1}^N (H_i^{\text{obs}} - \overline{H^{\text{obs}}})^2}. \quad (7)$$

Показатель Нэша–Сатклиффа NSE находится в диапазоне $(-\infty; 1]$, при значениях $NSE > 0.75$ качество моделирования считается отличным, при $0.36 < NSE < 0.75$ – удовлетворительным, при $NSE < 0.36$ – неудовлетворительным.

Для оценки качества “инерционного” прогноза моделей рассчитывался показатель S/σ_{Δ} [9, 4], σ_{Δ} – отклонение “инерционного” прогноза для каждой заблаговременности:

$$\sigma_{\Delta} = \sqrt{\frac{\sum_{i=1}^N (\Delta_i - \bar{\Delta})^2}{N - 1}}, \quad (8)$$

$$\Delta_i = H_i^{\text{obs}} - H_{i+\tau}^{\text{obs}}, \quad \tau = 1 \dots 20. \quad (9)$$

Значения S/σ_{Δ} находятся в диапазоне $[0; +\infty)$, значения 0–0.5 свидетельствуют об отличном, 0.5–0.7 – о хорошем, 0.7–0.8 – об удовлетворительном, >1 – о неудовлетворительном качестве методики прогноза.

Необходимо отметить, что приведенные метрики обычно используются для оценки качества моделирования и прогнозирования расходов воды, но могут использоваться и для уровней воды.

РЕЗУЛЬТАТЫ

В моделях машинного обучения, в которых используется линейная регрессия, в данном случае М5Р и XGBoost, степень ассоциации предикторов и предиката зачастую определяет качество моделирования и может быть оценена заранее путем проведения (авто)корреляционного анализа переменных. Для выбранных пар АГК на исследуемых реках выполнялась оценка корреляции уровней воды на расположенном ниже по течению АГК с различной заблаговременностью с уровнями на расположенном выше АГК с различным отступом назад во времени. Для р. Пшиш

наиболее выраженная корреляция уровней расположенного ниже АГК-95 и расположенного выше АГК-50 наблюдается при заблаговременности 15–20 ч (рис. 2а) – до 0.93. Аналогично время добега, определенное по времени соответственных пиков высоких паводков, на данном участке составляет, как уже указано, ~18 ч. Автокорреляция на АГК-95 при заблаговременности до 8 ч не падает ниже 0.95, до 13 ч – ниже 0.90 (рис. 2б). Для р. Мзымты в силу существенного изменения площади водосбора между АГК и значительной боковой приточности не наблюдается высокой корреляции между уровнями нижележащего АГК-160 и вышележащего АГК-89 (рис. 2в): она не превышает 0.74 для заблаговременности 1 ч и начинает снижаться после 5 ч. При активном снеготаянии на водосборе р. Мзымты время добега волны суточного хода уровней на выбранном участке составляет 2–3.5 ч. Автокорреляция часовых уровней АГК-160, напротив, достаточно высокая: до 7 ч она составляет 0.97, до 17 ч – 0.90 (рис. 2г). Полученные оценки соответствуют времени добега между АГК.

С помощью сеточного поиска в пространстве гиперпараметров моделей с 10-кратной кросс-валидацией на обучающей выборке определены оптимальные параметры и структуры моделей и оценены критерии качества моделирования при заблаговременности: 1, 2, 3, 5, 10, 15, 18 и 20 ч (табл. 2). Для модели М5Р структура дерева классификации усложнялась по мере роста заблаговременности, хотя интерпретация выбранной конфигурации модели доступна для любой заблаговременности в виде графа решения (как, например, на рис. 3 для заблаговременности 1 ч). В ходе настройки модели XGBoost были проверены возможные комбинации параметров: количество деревьев $\in [100, 300, 500, 800]$, максимальная глубина деревьев $\in [3, 5, 8]$. Лучшие результаты показала модель с 1000 деревьев глубиной ≤ 5 . В модели MLP для заблаговременности 5 ч выбрана структура с одним слоем из 10 нейронов, а при другой заблаговременности структура более сложная – два или три слоя из 10 нейронов каждый.

Для р. Пшиш наилучший результат на обучающей и проверочной выборках по значениям NSE для всех моделей получен для заблаговременности 1–5 ч. Для заблаговременности до 5 ч среднеквадратическая ошибка моделирования $S \leq 0.2$ м. На проверочном периоде ошибка S растет быстрее, чем на калибровочном периоде, и составляет 0.25–0.41 м. С ростом заблаговременности оценки качества в виде критерия NSE закономерно снижаются, но остаются >0.8 , что соответствует отличному результату моделирования. По результатам проверки на независимом материале лучшие оценки качества достигнуты у моделей М5Р и XGBoost, у последней – для заблаговременности ≥ 10 ч.

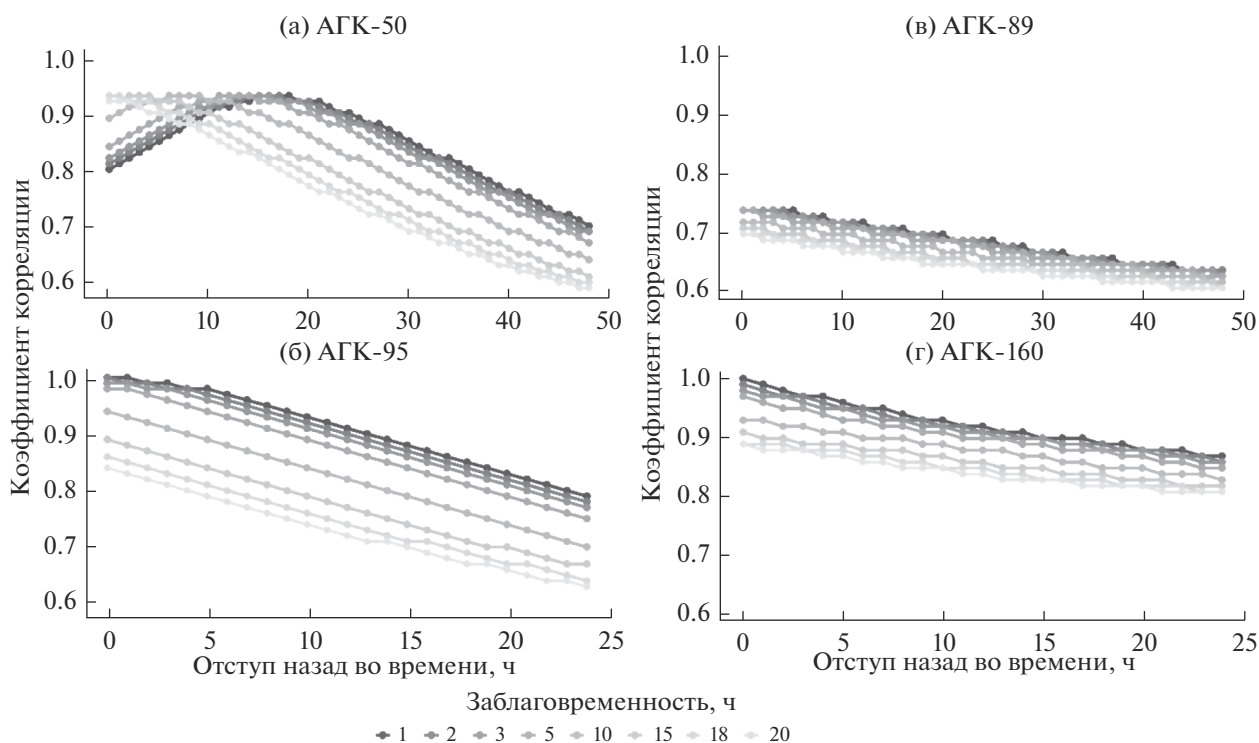


Рис. 2. Корреллограмма уровней воды на реках Пшиш (а, б) и Мзымте (в, г) между верхним и нижним постом (а, в) и автокорреллограмма на нижнем посту (б, г).

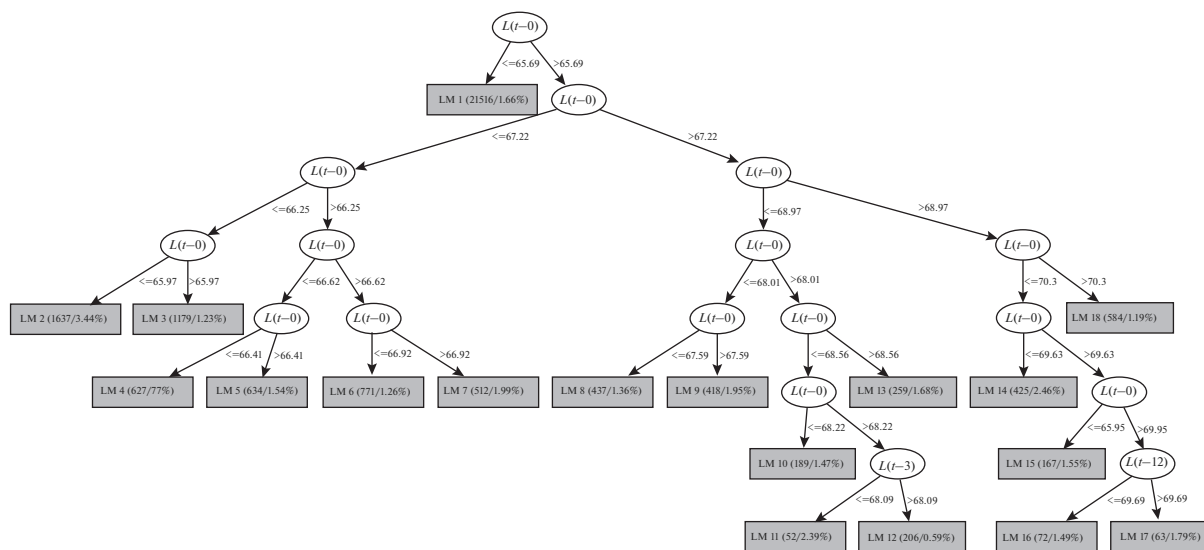


Рис. 3. Графическое отображение дерева решений модели M5P для прогноза уровня воды на р. Пшиш у ст. Гурийской при заблаговременности 1 ч.

Оценки прогностических свойств результатов моделирования по критерию S/σ_{Δ} показывают, что наилучший прогноз может быть достигнут для заблаговременности 15 и 18 ч, что близко к времени добега на участке между выбранны-

ми постами АГК-95 и АГК-50 на р. Пшиш и согласуется с результатами корреляционного анализа.

Выполнен анализ моделирования отдельных паводочных событий на р. Пшиш. Ниже рассмот-

Таблица 2. Оценки качества моделирования уровней воды р. Пшиш у ст. Гурийской

Заблаговременность, ч			1	2	3	5	10	15	18	20	
Обучение	M5P	<i>S</i>	0.02	0.04	0.06	0.07	0.11	0.15	0.22	0.21	
		<i>NSE</i>	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.98
		<i>S/σΔ</i>	0.45	0.38	0.37	0.27	0.23	0.24	0.24	0.30	0.27
	XGboost	<i>S</i>	0.01	0.02	0.03	0.09	0.13	0.17	0.22	0.22	0.27
		<i>NSE</i>	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.96
		<i>S/σΔ</i>	0.20	0.19	0.16	0.37	0.28	0.27	0.30	0.30	0.35
	MLP	<i>S</i>	0.04	0.05	0.05	0.08	0.12	0.16	0.20	0.20	0.22
		<i>NSE</i>	0.99	0.99	0.99	0.98	0.96	0.93	0.90	0.90	0.88
		<i>S/σΔ</i>	0.66	0.42	0.33	0.33	0.26	0.25	0.28	0.28	0.29
Проверка	M5P	<i>S</i>	0.03	0.05	0.08	0.14	0.26	0.28	0.34	0.38	
		<i>NSE</i>	0.99	0.99	0.99	0.99	0.98	0.97	0.96	0.95	
		<i>S/σΔ</i>	0.53	0.45	0.46	0.50	0.51	0.40	0.43	0.45	
	XGboost	<i>S</i>	0.07	0.08	0.10	0.14	0.22	0.27	0.31	0.36	
		<i>NSE</i>	0.99	0.99	0.99	0.99	0.98	0.98	0.97	0.96	
		<i>S/σΔ</i>	1.24	0.71	0.58	0.52	0.44	0.38	0.39	0.42	
	MLP	<i>S</i>	0.09	0.11	0.11	0.15	0.25	0.33	0.35	0.35	0.41
		<i>NSE</i>	0.99	0.99	0.99	0.99	0.98	0.96	0.96	0.96	0.94
		<i>S/σΔ</i>	1.58	0.94	0.67	0.55	0.50	0.47	0.44	0.44	0.49

рим моделирование наиболее высокого (исторического) паводка 24–26 октября 2018 г. на примере заблаговременности 5 и 15 ч (рис. 4).

Для заблаговременности 5 и 15 ч для паводка 25 октября 2018 г. модель MLP существенно завышает максимальный уровень – на 2.3 и 3.2 м соответственно. Модель MLP раньше остальных моделей предсказывает достижение отметки неблагоприятного явления (НЯ) (71.802 м) и отметки опасного явления (ОЯ) (72.802 м) – на 5 ч раньше

их фактического наступления при заблаговременности 5 ч и на 4 ч раньше при заблаговременности 15 ч. На спаде паводка модель MLP демонстрирует резкий провал в отметках уровней, что, возможно, связано с интерпретацией моделью двухпикового паводка на верхнем створе АГК-95, уровни которого входят в предикторы для прогноза уровней на нижнем створе АГК-50.

Некоторое возмущение на подъеме паводка можно отметить в результатах модели M5P, кото-

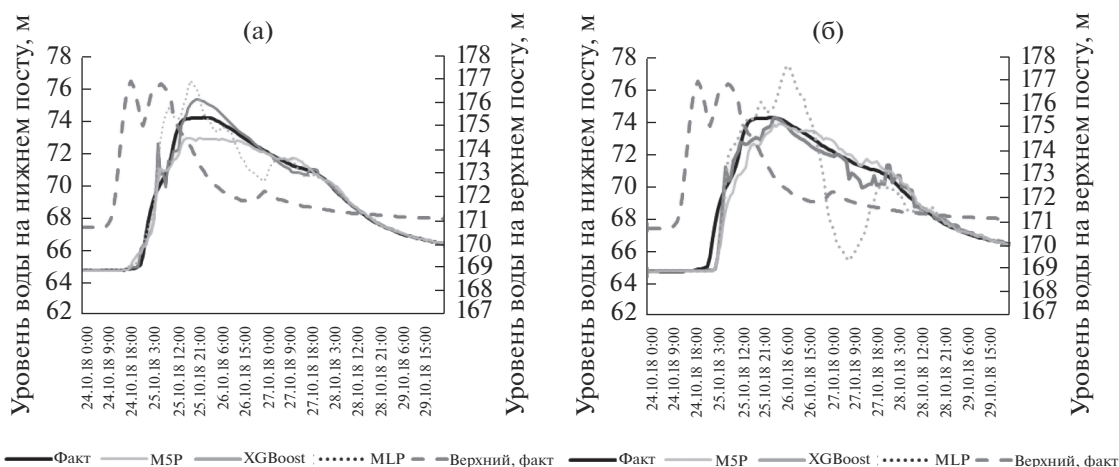


Рис. 4. Фактический и смоделированный графики хода уровней воды при прохождении паводка 25.10.2018–26.10.2018 на р. Пшиш. Заблаговременность слева – 5, справа – 15 ч.

Таблица 3. Оценки качества моделирования уровней воды р. Мзымты у с. Казачий Брод

Заблаговременность, ч			1	2	3	5	10	15	18	20
Обучение	M5P	<i>S</i>	0.02	0.03	0.04	0.05	0.06	0.06	0.06	0.06
		<i>NSE</i>	0.99	0.99	0.98	0.97	0.96	0.96	0.95	0.95
		<i>S/σΔ</i>	0.76	0.75	0.73	0.68	0.53	0.47	0.47	0.46
	XGboost	<i>S</i>	0.02	0.03	0.04	0.06	0.10	0.11	0.12	0.12
		<i>NSE</i>	0.99	0.99	0.98	0.96	0.89	0.85	0.84	0.83
		<i>S/σΔ</i>	0.68	0.75	0.79	0.78	0.91	0.89	0.88	0.88
	MLP	<i>S</i>	0.08	0.13	0.17	0.22	0.31	0.37	0.39	0.40
		<i>NSE</i>	0.99	0.98	0.96	0.93	0.86	0.80	0.78	0.76
		<i>S/σΔ</i>	3.23	3.09	3.00	2.94	2.92	2.90	2.88	2.87
Проверка	M5P	<i>S</i>	0.04	0.06	0.07	0.11	0.15	0.17	0.18	0.19
		<i>NSE</i>	0.99	0.97	0.95	0.88	0.76	0.70	0.66	0.64
		<i>S/σΔ</i>	1.03	0.91	0.90	1.00	1.03	0.96	0.96	0.95
	XGboost	<i>S</i>	0.03	0.06	0.08	0.10	0.14	0.16	0.17	0.17
		<i>NSE</i>	0.99	0.97	0.94	0.90	0.80	0.73	0.70	0.68
		<i>S/σΔ</i>	0.94	0.94	0.94	0.95	0.94	0.91	0.89	0.89
	MLP	<i>S</i>	0.03	0.06	0.07	0.09	0.14	0.15	0.16	0.17
		<i>NSE</i>	0.99	0.97	0.95	0.91	0.81	0.75	0.71	0.70
		<i>S/σΔ</i>	0.95	0.92	0.87	0.87	0.92	0.87	0.87	0.87

рое вскоре нивелируется. При заблаговременности 5 ч модель M5P завьщает пиковые уровни на 1.1 м, при заблаговременности 15 ч модель на 2 ч раньше предсказывает достижения отметки НЯ, но запаздывает с прогнозированием отметок ОЯ, не дотягивая до наблюдаемых уровней 1.0–1.3 м.

Результаты моделирования этого же паводка по модели XGBoost менее убедительные – модель позднее описывает наступление отметок НЯ, ОЯ и в абсолютных значениях занижает уровни воды на 1.0–1.7 м. Причем при заблаговременности 5 ч модель XGBoost постоянно занижает пиковые уровни воды, а при заблаговременности 15 ч ведет себя аналогично модели M5P и запаздывает с прогнозируемым уровнем ОЯ на 4 ч. Таким образом, реальная заблаговременность в прогнозе наступления отметки ОЯ составляет 19 и 12–11 ч по моделям MLP и M5P соответственно.

В целом наилучшие результаты моделирования для выдающегося паводка 24–25 октября 2018 г. на р. Пшиш достигаются двумя моделями: M5P и MLP. Также стоит отметить, что XGBoost показывает более плавные подъемы и спады паводков, в отличие от MLP с более резкой реакцией на изменение предикторов. Модели M5P и MLP при прогнозе способны завьщать отметки уровней паводков в области ненаблюдаемых ранее величин. Несмотря на более слабый результат модели MLP по оценкам *NSE* и *S/σΔ*, модель про-

явила хорошие прогностические качества на выдающемся паводке.

Аналогичные испытания моделей проведены для р. Мзымты. Результаты приведены в табл. 3. Следует отметить, что эффективность моделей, выраженная в оценке *NSE*, падает с каждым шагом заблаговременности от 0.98 до 0.6 гораздо быстрее, чем для р. Пшиш. Это объясняется существенно меньшим временем добега между верхним АГК-89 и нижним АГК-160, составляющим <4 ч. Кроме того, зачастую при высоких паводках пик формируется раньше на нижнем посту из-за направления движения воздушных масс от побережья в горы и последовательного включения частей водосбора в процесс формирования стока и его трансформации, что снижает роль уровней на верхнем посту как предикторов. Среднеквадратические ошибки моделирования составляют от 0.02 до 0.40 м за период обучения и от 0.03 до 0.19 м за проверочный период.

Для всех трех моделей свойственно заметное снижение эффективности *NSE* за проверочный период, хотя наиболее стабильные результаты характерны в данном случае для многослойного перцептрона. Для моделей M5P и XGBoost существенное снижение эффективности за период верификации, вероятно, можно объяснить чрезмерным обучением (англ. “overfitting”).

Для участка Мзымты по критерию *S/σΔ* прогностическая эффективность не достигается.

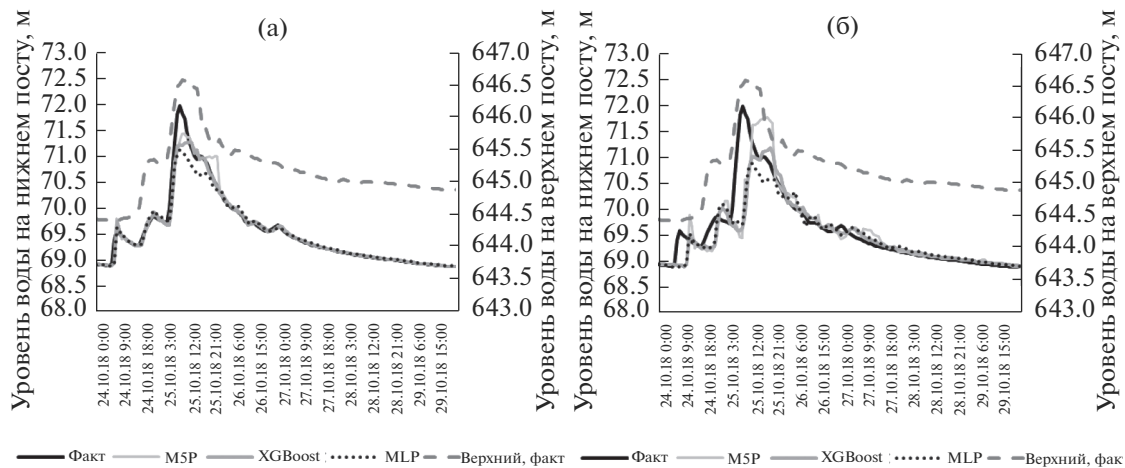


Рис. 5. Фактический и смоделированный графики хода уровней воды при прохождении паводка 25.10.2018 г. на р. Мзымте. Заблаговременность слева – 1, справа – 5 ч.

Более успешные результаты моделирования получены для заблаговременности 3 и 5 ч. Это свидетельствует о необходимости привлечения дополнительных гидрометеорологических предикторов для разработки более эффективных прогностических методик и увеличения заблаговременности. Аналогично выполнен анализ моделирования отдельных паводочных событий на р. Мзымте. Наиболее высокий паводок 25.10.2018–26.10.2018, включенный в проверочный период, описывается всеми моделями с занижением максимальных отметок уровней воды и с запаздыванием во времени относительно наблюдаемых значений (рис. 5).

ОБСУЖДЕНИЕ

Качество моделирования уровней воды за выбранный период для участка р. Пшиш более высокое, чем для участка р. Мзымты. Найденные различия в качестве моделирования на выбранных участках рек Пшиш и Мзымта могут быть объяснены различиями природных условий на их водосборах. Несмотря на значительную площадь бокового участка водосбора, превышающую площадь водосборов у верхних постов (в 1.3 и 1.9 раз соответственно), приточность не сильно влияет на ход паводочной волны только на р. Пшиш. Боковой водосбор на участке р. Пшиш от с. Навагинского до ст. Гурийской довольно пологий с высотами ≤ 700 м, видимо, ведущий себя синхронно с верхним водосбором, где в основном формируется сток высоких паводков в результате выпадения интенсивных дождевых осадков. Ниже г. Хадыженска дополнительный приток может быть совсем низким из-за малого коэффициента дождевого стока равнинной части водосбора, а характер р. Пшиш становится транзитным [8].

Большое время добегания на участке объясняется слабым уклоном русла (падение от 180 до 80 м абс.).

В пределах более высоко расположенного водосбора р. Мзымты процессы формирования стока и трансформации паводочной волны на выбранном участке усложнены за счет высотной поясности (это влияет на соотношение талого и дождевого стока) и системы разнонаправленных хребтов, влияющих на движение несущих влагу воздушных масс. Водосбор р. Мзымты от слияния с р. Пслух до Казачьего Брода включает в себя среднегорье краснополянской котловины (с высотами 500–2000 м) и низкоргорье так называемой зоны “предвосхождения” воздушных масс (до ~ 1000 м), обращенное к побережью [10]. Боковая приточность здесь – мощный фактор, влияющий на ход уровней воды и затрудняющий строить модели уровня режима с учетом только верхнего створа (АГК-89), расположенного в верхнем течении реки. Кроме того, в русле реки даже после зимней Олимпиады-2014 проводятся дноуглубительные и берегоукрепительные работы, которые могут оказывать влияние на уровень режим в районе верхнего поста. В районе нижнего поста в Казачьем Броду в уровне режиме в отдельные периоды (например, 17.10.2019–28.10.2019) прослеживаются сбросы Краснополянской ГЭС с диапазоном колебаний до 25 см. Маленькое время добегания на участке обусловлено в том числе большим уклоном русла (падение от 650 до 70 м абс. на вдвое более коротком участке, чем на р. Пшиш).

Для участка р. Пшиш, отличающегося значительной скоррелированностью уровней воды верхнего и нижнего АГК, наилучшие результаты показала модель бустинга деревьев решений, т.е. по своей сути – ансамбль линейных моделей, наиболее эффективным образом использующий

корреляцию внутри выборки. Сложная нелинейная модель MLP показала наихудшие из трех моделей результаты при каждой заблаговременности, очевидно, вследствие переобучения. Тем не менее ее результаты могут иметь большой прогностический потенциал для высоких паводков, как это было показано на примере выдающегося паводка в октябре 2018 г. Наибольшая эффективность прогноза достигается при заблаговременности 15 и 18 ч. Таким образом, полученные результаты моделирования для р. Пшиш, имитирующие закономерности трансформации паводочной волны на участке реки, дополняют результаты, полученные по методу соответственных уровней (расходов воды) для более крупной реки региона — Кубани [15], но с часовым шагом вместо суточного и с применением более современных технологий.

Для участка р. Мзымты связность между предиктором и предиктантом заметно меньше вследствие значительной и асинхронной боковой приточности (рис. 2). Наилучшей из трех моделей оказалась нелинейная модель MLP, что косвенно может свидетельствовать о возможности обнаружения ей дополнительных связей внутри выборки. И хотя по критерию NSE качество моделирования для заблаговременности до 5 ч оценивается как отличное, более строгий критерий прогностической эффективности $S/\sigma\Delta$ не позволяет назвать прогностический результат удовлетворительным. Нюансы поведения этого критерия по мере роста заблаговременности обсуждаются в работе [11]. Для построения более эффективных прогностических моделей хода уровней и/или расходов воды Мзымты целесообразно использовать дополнительно сведения о поведении притоков и метеорологическую информацию.

ЗАКЛЮЧЕНИЕ

Проведенное исследование показало, что применение методов машинного обучения для моделирования уровней воды может обеспечить надежные прогностические результаты. Для участка нижнего течения р. Пшиш оптимальная заблаговременность прогнозирования составляет 15–18 ч. Для р. Мзымты в Казачьем Броду с учетом данных об уровнях на верхнем посту у слияния с р. Пслух качество моделирования оценивается как хорошее, однако без достижения необходимой прогностической эффективности. Это объясняется следствием значительной боковой приточности в среднем и нижнем течении реки и асинхронного формирования стока в разных частях водосбора.

Для участка р. Пшиш наилучшие результаты воспроизведения уровней воды показала модель бустинга деревьев решений (XGBoost), т.е. ансамбля линейных моделей, наиболее эффективным образом использующих корреляцию внутри выборки. Модели M5P и MLP могут успешно до-

полнять и улучшать прогноз в случае прохождения высоких паводков.

Для р. Мзымты наилучшей из трех моделей, напротив, оказалась нелинейная модель MLP, что косвенно может свидетельствовать о возможности обнаружения ей дополнительных связей внутри выборки. Дальнейшие исследования будут сфокусированы на поиске интерпретируемых нейронных сетей. Это достаточно востребованное в настоящее время направление разработок [26, 44]. Для улучшения качества моделирования уровней воды Мзымты целесообразно привлекать дополнительные исходные данные об уровнях на притоках и метеорологическую информацию.

В целом использование методов машинного обучения для построения методик прогноза паводков на реках Краснодарского края на основе данных сети АСМПС КК представляется перспективным. При накоплении достаточных массивов данных интенсивности выпадения осадков с часовым шагом будет исследована возможность использования моделей искусственного интеллекта в задачах краткосрочного и сверхкраткосрочного прогноза паводков типа осадки—уровни и осадки—сток.

СПИСОК ЛИТЕРАТУРЫ

1. *Алексеевский Н.И., Магрицкий Д.В., Колтерманн П.К., Торопов П.А., Школьный Д.И., Белякова П.А.* Наводнения на черноморском побережье Краснодарского края // Вод. ресурсы. 2016. Т. 43. № 1. С. 3–17.
2. *Белякова П.А., Морейдо В.М., Пьянкова А.И.* Анализ половозрастной структуры погибших от наводнений в России за 2000–2014 гг. // Третьи Виноградовские чтения. Грани гидрологии. 2018. С. 849–853.
3. *Борщ С.В., Симонов Ю.А., Христофоров А.В.* Система прогнозирования паводков и раннего оповещения о наводнениях на реках Черноморского побережья Кавказа и бассейна Кубани // Тр. Гидрометцентра РФ. 2015. Спец. вып. 356. 247 с.
4. *Борщ С.В., Христофоров А.В.* Оценка качества прогнозов речного стока // Тр. Гидрометцентра России. 2015. Спец. вып. 355. 198 с.
5. *Васильева Е.С., Белякова П.А., Алексюк А.И., Селезнева Н.В., Беликов В.В.* Моделирование быстроразвивающихся паводков на малых реках Северного Кавказа с использованием современных данных автоматизированной гидрометеорологической сети // Вод. ресурсы. 2021. Т. 48. № 2. С. 135–146.
6. Интегрированное управление паводками. Концептуальный документ ВМО № 1047. Женева, 2009. 32 с.
7. *Кузин П.С.* Классификация рек и гидрологическое районирование СССР. Л.: Гидрометеоиздат, 1960. 456 с.
8. *Лурье П.М., Панов В.Д., Ткаченко Ю.Ю.* Река Кубань. Гидрография и режим стока. СПб.: Гидрометеоиздат, 2005. 498 с.

9. Наставление по службе прогнозов. Раздел 3. Ч. 1. Прогнозы режима вод суши. Л.: Гидрометеоздат, 1962. 193 с.
10. *Панов В.Д., Базелюк А.А., Лурье П.М.* Реки Черноморского побережья Кавказа: гидрография и режим стока. Ростов н/Д.: Донской изд. дом, 2012. 607 с.
11. *Соколов А.А., Бугаец А.Н.* К вопросу об оценке методик краткосрочного прогноза гидрологических величин // *Метеорология и гидрология*. 2018. № 8. С. 65–71.
12. *Ткаченко Ю.Ю., Шержуков Е.Л.* Опыт создания систем краткосрочного прогноза угроз гидрологического характера // *Вод. хозяйство России: проблемы, технологии, управление*. 2014. № 3. С. 75–82.
13. *Шевердяев И.В., Клещенко А.В., Мисиров С.А.* Факторы паводковой опасности рек Северо-Западного Кавказа // *Наука Юга России*. 2021. Т. 17. № 1. С. 37–51.
14. *Цыпленков А.С., Иванова Н.Н., Ботавин Д.В., Голосов В.Н.* Гидрометеорологические предпосылки и геоморфологические последствия экстремального паводка в бассейне малой реки в зоне влажных субтропиков (на примере р. Цаньк, район Сочи) // *Вестн. Санкт-Петербургского ун-та. Науки о Земле*. 2021. Т. 66. № 1. С. 144–166.
15. *Юмина Н.М., Кукулина Л.В.* Прогнозы и расчеты стока рек в бассейне Кубани // *Вестн. Моск. ун-та. Сер. 5, География*. 2011. № 1. С. 55–59.
16. *Abrahart R.J., Anctil F., Coulibaly P., Dawson Ch.W., Mount N.J., See L.M., Shamseldin A.Y., Solomatine D.P., Toth E., Wilby R.L.* Two decades of anarchy? Emerging themes and outstanding challenges for neural network river forecasting // *Progress in Physical Geography: Earth and Environment*. 2012. V. 36. № 4. P. 480–513. <https://doi.org/10.1177/0309133312444943>
17. *Beven K.* Deep learning, hydrological processes and the uniqueness of place // *Hydrological Processes*. 2020. V. 34. № 16. P. 3608–3613. <https://doi.org/10.1002/hyp.13805>
18. *Bhattacharya B., Solomatine D.P.* Machine learning in sedimentation modelling // *Neural Networks*. 2006. V. 19. № 2. P. 208–214. <https://doi.org/10.1016/j.neunet.2006.01.007>
19. *Brooks K.N., Ffolliott P.F., Magner J.A.* Hydrology and the management of watersheds. 4th ed. Ames, Iowa: Wiley-Blackwell, 2013. <https://doi.org/10.1002/9781118459751>
20. *Chen T., Guestrin C.* XGBoost: A Scalable Tree Boosting System // *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining*. P. 785–794. <https://doi.org/10.1145/2939672.2939785>
21. *Chernokulsky A.V., Kozlov E.A., Zolina O.G., Bulygina O.N., Mokhov I.I., Semenov V.A.* Observed changes in convective and stratiform precipitation in Northern Eurasia over the last five decades // *Environ. Res. Lett.* 2019. V. 4. № 4. <https://doi.org/10.1088/1748-9326/aafb82>
22. *Emori S., Brown S.J.* Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate // *Geophys. Res. Lett.* 2005. V. 32. L17706. <https://doi.org/10.1029/2005GL023272>
23. *Ghahramani Z.* Probabilistic machine learning and artificial intelligence // *Nature*. 2015. V. 521. P. 452–459. <https://doi.org/10.1038/nature14541>
24. *Haykin S.* Neural Networks and Learning Machines. 3rd ed. Upper Saddle River. N. Y.: Prentice Hall Publ., 2009. 906 p.
25. IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change / Eds *T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P.M. Midgey*. Cambridge, United Kingdom; NY, USA: Cambridge Univ. Press, 1535 p.
26. *Jain A., Sudheer K.P., Srinivasulu S.* Identification of physical processes inherent in artificial neural network rainfall runoff models // *Hydrol. Processes*. 2004. V. 18. № 3. P. 571–581.
27. *Lehmann J., Coumou D., Frieler K.* Increased record-breaking precipitation events under global warming // *Climatic Change*. 2015. № 4 (132). P. 501–515.
28. *Luo J., Ying K., Bai J.* Savitzky–Golay smoothing and differentiation filter for even number data // *Signal Processing*. 2005. V. 85. № 7. P. 1429–1434. <https://doi.org/10.1016/j.sigpro.2005.02.002>
29. *Manual on Flood Forecasting and Warning*. WMO. № 1072. Geneva, 2011. 138 p.
30. *McCulloch W.S., Pitts W.* A logical calculus of the ideas immanent in nervous activity // *Bull. Math. Biophys.* 1943. V. 5. P. 115–133. <https://doi.org/10.1007/BF02478259>
31. *Moreido V.M., Gartsman B.I., Solomatine D.P., Suchilina Z.A.* Prospects for short-term forecasting of river streamflow from small watershed runoff using machine learning methods // *Hydrosphere. Hazard Process. Phenom.* 2020. V. 2. № 4. P. 375–390. <https://doi.org/10.34753/HS.2020.2.4.375>
32. *Moreido V., Gartsman B., Solomatine D.P., Suchilina Z.* How well can machine learning models perform without hydrologists? Application of rational feature selection to improve hydrological forecasting // *Water (Switzerland)*. 2021. V. 13. № 12. 1696. <https://doi.org/10.3390/w13121696>
33. *Mutlu E., Chaubey I., Hexmoor H., Bajwa S.G.* Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed // *Hydrol. Process*. 2008. V. 22. № 26. P. 5097–5106. <https://doi.org/10.1002/hyp.7136>
34. *Nash J.E., Sutcliffe J.V.* River flow forecasting through conceptual models part I – A discussion of principles // *J. Hydrol.*, 1970. V. 10. № 3. P. 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
35. *Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay É.* Scikit-learn: Machine Learning in Python // *J. Machine Learning Res.* 2011. V. 12. P. 2825–2830.

36. *Press W.H., Teukolsky S.A.* Savitzky-Golay Smoothing Filters // *Computers in Physics*. 1990. V. 4. № 6. P. 669. <https://doi.org/10.1063/1.4822961>
37. *Quinlan J.R.* Learning with continuous classes // *Proc. 5th Australian Joint Conf. Artificial Intelligence*. 1992. P. 343–348.
38. *Reichstein M., Camps-Valls G., Stevens B., Jung M., Denzler J., Carvalhais N., Prabhat.* Deep learning and process understanding for data-driven Earth system science // *Nature*. 2019. V. 566. P. 195–204. <https://doi.org/10.1038/s41586-019-0912-1>
39. *Rodda H.J.E., Little M.A.* Understanding mathematical and statistical techniques in hydrology. Chichester, UK: John Wiley & Sons, Ltd, 2015. 302 p. <https://doi.org/10.1002/9781119077985>
40. *Rolnick D., Ahuja A., Schwarz J., Lillicrap T.P., Wayne G.* Experience Replay for Continual Learning // *Advances in Neural Information Processing Systems*. 2018. V. 32. <http://arxiv.org/abs/1811.11682>
41. *Rumelhart D.E., Hinton G.E., Williams R.J.* Learning representations by back-propagating errors // *Nature*. 1986. V. 323. P. 533–536. <https://doi.org/10.1038/323533a0>
42. *Savitzky A., Golay M.J.E.* Smoothing and Differentiation of Data by Simplified Least Squares Procedures // *Analytical Chem.* 1964. V. 36. № 8. P. 1627–1639. <https://doi.org/10.1021/ac60214a047>
43. *Senthil Kumar A.R., Sudheer K.P., Jain S.K., Agarwal P.K.* Rainfall-runoff modelling using artificial neural networks: comparison of network types // *Hydrol. Process*. 2005. V. 19 № 6. P. 1277–1291. <https://doi.org/10.1002/hyp.5581>
44. *Shen C.A.* Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists // *Water Resour. Res.* 2018. V. 54. № 11. P. 8558–8593.
45. *Sylvester Z., Durkin P., Covault J.A., Sharman G.R.* High curvatures drive river meandering: REPLY // *Geol.* 2019. V. 47. № 10. P. e486–e486. <https://doi.org/10.1130/G46838Y.1>
46. WMO, 2012. Management of Flash Floods. Integrated Flood Management Tools Series No.16. Geneva, 2012. 61 p.