

УДК 504.064.2.001.18

DOI: 10.34753/HS.2020.2.4.375

ВОЗМОЖНОСТИ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ СТОКА МАЛОЙ РЕКИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

В.М. Морейдо¹, Б.И. Гарцман¹,
Д.П. Соломатин^{1,2}, З.А. Сучилина¹

¹Институт водных проблем Российской
академии наук, г. Москва, Россия

²IHE Delft Институт образования в сфере
водных ресурсов, г. Дельфт, Нидерланды
moreido@mail.ru

PROSPECTS FOR SHORT-TERM FORECASTING OF RIVER STREAMFLOW FROM SMALL WATERSHED RUNOFF USING MACHINE LEARNING METHODS

Vsevolod M. Moreido¹, Boris I. Gartsman¹,
Dimitri P. Solomatine^{1,2}, Zoya A. Suchilina¹

¹Water Problems Institute of Russian Academy of
Sciences, Moscow, Russia

²IHE Delft Institute for Water Education, Delft, The
Netherlands
moreido@mail.ru

Аннотация. В статье исследуются возможности краткосрочного (от 1 до 7 суток) прогнозирования расходов воды на основе нескольких методов машинного обучения: модели множественной линейной регрессии, искусственной нейронной сети по типу многослойного перцептрона и рекуррентной искусственной нейронной сети с долгосрочной кратковременной памятью. Предлагаются методы расширения набора предикторов для построения моделей и исследуется возможность случайного перемешивания хронологического ряда предикторов для калибровки и верификации моделей как повышающая устойчивость результатов прогноза. В качестве объекта исследования используется малая река Средней полосы России – река Протва (гидрометрический пост Спас-Загорье). В качестве предикторов используются расходы воды на посту и суточные суммы осадков на трех ближайших метеостанциях в текущий момент времени (сутки) и со сдвигом назад до 7 суток, а также индекс увлажнения бассейна и характеристики температуры воздуха и испарения. На конкретном примере показана возможность построения эффективной оперативной прогностической системы для краткосрочного прогнозирования стока. Исследование выявило приемлемую для оперативной практики

Abstract. The paper addresses prospects for short-term (from 1 to 7 days) forecasting of river streamflow runoff based on several machine learning methods: multiple linear regression (LM) model, a multilayer perceptron (MLP) artificial neural network, and a recurrent artificial neural network with long short-term memory (LSTM). Methods for expanding the set of predictors for model construction are proposed, and the possibility of random shuffling of the time-series of predictors for model calibration and verification are assessed. The object of the study is the small river of Central Russia – river Protva (Spas-Zagorie gauge). Current and lagged values of streamflow discharge at the gauge and daily precipitation at local weather stations are used as predictors for the model, as well as moisture index and evaporation rate. The obtained results show the possibility of constructing an effective operational forecasting system for short-term runoff forecasting. The study revealed the applicability of artificial neural network models, acceptable for operational practice, using all available hydrometeorological information on the catchment, as they showed the most stable results at all lead times from 1 to 7 days. In contrast to the linear model, which efficiency decreases after lead time of more than 3 days, the artificial neural networks models have higher forecast efficiency up to 7 days. The results obtained are robust for all

применимость моделей искусственных нейронных сетей, использующих всю доступную гидрометеорологическую информацию на водосборе, как показавших наиболее устойчивые результаты на всех заблаговременностях от 1 до 7 суток. Так, в отличие от линейной прогностической модели, эффективность которой снижается на заблаговременностях более 3 суток, модели искусственных нейронных сетей показали высокую эффективность прогноза до 7 суток. Полученные результаты устойчивы для всех фаз водного режима, как весеннего половодья, так и летних паводков. Программная реализация моделей выполнена на основании открытых программных библиотек на языке Python, что показывает возможность широкого использования описанных методик для научных исследований и прикладных задач.

Ключевые слова: гидрологические прогнозы; машинное обучение; прогнозы речного стока; искусственные нейронные сети; краткосрочные прогнозы расходов воды; модели речного стока

Введение

Использование математических моделей для расчета и прогноза речного стока является достаточно хорошо разработанной областью гидрологии суши, опыт которой насчитывает уже более 70 лет [Emerton et al., 2016; Fatichi et al., 2016; Мотовилов, 2016a]. На сегодняшний день известно достаточно большое число классов и структур гидрологических моделей, предназначенных для решения различных научных и прикладных задач [Кучмент, Гельфан, 1993; Gelfan, Motovilov, 2009]. Так, для детальных расчетов процессов формирования речного стока используются физико-математические модели, содержащие сложные параметризации описываемых процессов и позволяющие проводить различные их оценки, например реакцию компонентов гидрологического цикла на изменения в климатической системе [Fatichi et al., 2016; Мотовилов, 2016a; Мотовилов, 2016b; Veven, 2020]. Наряду с этими моделями в задачах, где детальное описание процессов формирования не требуется, а исследуется реакция речной

системы на метеорологические воздействия, таких как краткосрочное прогнозирование речного стока, успешно используются более упрощенные модели, связывающие предшествующие наблюдения за речным стоком и метеовеличинами в пределах речного водосбора с будущими величинами речного стока [Белякова и др., 2013; Veven, 2020]. Такие модели, получившие название «черного ящика» по причине отсутствия в них детального описания стокоформирующих процессов, были известны с середины 50-х годов [Кучмент, Гельфан, 1993]. Физической основой для них является, как правило, сильная автокорреляция рядов речного стока и (в некоторых физико-географических условиях) корреляция стока с осадками. На протяжении последних 20 лет такие модели находятся на новом витке эволюции. Это происходит благодаря развитию высокопроизводительных вычислительных систем и разработке новых алгоритмов, получивших общее название «модели машинного обучения» (“machine learning”) [Halff, Halff,

Keywords: hydrological forecasting, machine learning, artificial neural networks; runoff forecasts; short-term streamflow forecasting; runoff modelling

системы на метеорологические воздействия, таких как краткосрочное прогнозирование речного стока, успешно используются более упрощенные модели, связывающие предшествующие наблюдения за речным стоком и метеовеличинами в пределах речного водосбора с будущими величинами речного стока [Белякова и др., 2013; Veven, 2020]. Такие модели, получившие название «черного ящика» по причине отсутствия в них детального описания стокоформирующих процессов, были известны с середины 50-х годов [Кучмент, Гельфан, 1993]. Физической основой для них является, как правило, сильная автокорреляция рядов речного стока и (в некоторых физико-географических условиях) корреляция стока с осадками. На протяжении последних 20 лет такие модели находятся на новом витке эволюции. Это происходит благодаря развитию высокопроизводительных вычислительных систем и разработке новых алгоритмов, получивших общее название «модели машинного обучения» (“machine learning”) [Halff, Halff,

Azmoodeh, 1993; Kratzert et al., 2018]). Такие модели, в том числе один из их подклассов – искусственные нейронные сети (далее – ИНС), – показали свою эффективность в решении различных прикладных задач прогнозирования временных рядов, включая речной сток [Beven, 2020]. В основе таких моделей лежит нелинейная трансформация весов, присваиваемых исходным данным (предикторам) в процессе оценки относительного вклада каждого из предикторов в выходной ряд (предикат), а также оптимизация значений этих весов на каждом расчетном шаге для минимизации целевой функции (также называемое обратным распространением ошибки модели [Werbos, 1990]). Нелинейная трансформация происходит внутри нейросети на слое узлов активации, называемых нейронами, и связывающих все предикторы и предикат.

Следует отметить, что модели машинного обучения, и ИНС в том числе, отличаются непритязательностью к исходным данным, так как не содержат детальных описаний процессов формирования стока и тем самым освобождены от неопределенностей, связанных с параметризацией таких процессов. Это достоинство, впрочем, одновременно является и их главным недостатком, поскольку одной из основных претензий, предъявляемых гидрологическим сообществом к таким моделям, является невозможность объективной интерпретации их структуры [Abrahart et al., 2012].

В настоящем исследовании предпринята попытка продемонстрировать возможность построения эффективной системы краткосрочного прогнозирования речного стока с помощью моделей ИНС двух типов – многослойного перцептрона [Haykin, 1994] и модели с долгосрочной кратковременной памятью [Hochreiter, Schmidhuber, 1997]. В качестве объекта исследования используется бассейн малой реки Средней полосы России – реки Протва с замыкающим гидрометрическим створом в селе Спас-Загорье, для которого строятся прогнозы речного стока заблаговременностью до 7 суток. В статье последовательно описываются архитектура

моделей, использованные ряды данных для построения моделей и методы обучения и проверки, после чего производится сопоставление полученных результатов между собой.

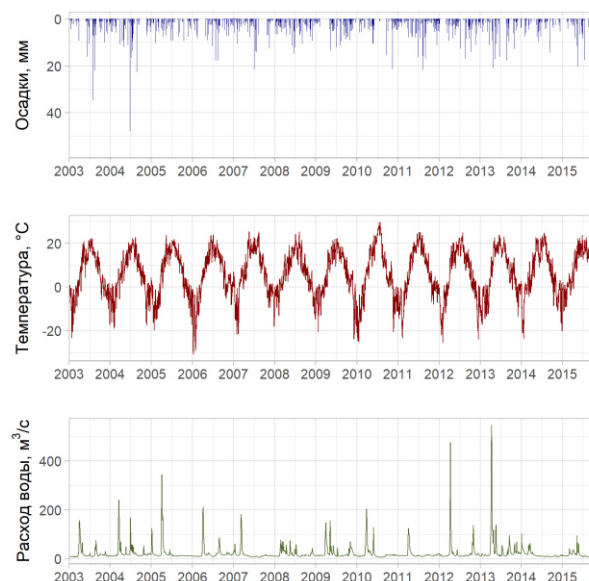
Объект исследования

Малая река Протва расположена в средней полосе Европейской России в пределах водосбора реки Волга (рисунок 1а). Длина реки – 282 км, площадь водосбора – 4 620 км². Исток реки Протва находится на междуречье реки Ока и ее крупного притока – реки Москва, – близ города Можайск, и впадает река Протва в реку Ока у города Протвино. Бассейн реки Протва расположен в области умеренно-континентального климата с холодной зимой и умеренно теплым летом. Особенностью территории является частое прохождение циклонов как в холодный, так и в теплый период года. С прохождением циклонов связано основное количество (до 90%) осадков. Водный режим реки Протва относится к восточноевропейскому типу [Зайков, 1946], главная его фаза – весеннее половодье – длится примерно 30-40 суток. В этот период проходит более половины годового стока. Остальная часть года представляет собой относительно маловодный период, состоящий из летне-осенней и зимней межени, за исключением небольшого повышения водности в конце теплого периода, вызванного осенними дождями.

В силу размера водосбора сток реки Протва в теплый период года чувствителен к ливневым осадкам, а в холодный – к периодам оттепелей. В стоке реки Протва доминирует снеговая составляющая (50%), при этом активно участвуют подземное (30%) и дождевое (20%) питание (рисунок 1б). Средний многолетний расход реки Протва в устье составляет 24 м³/с. Для данного исследования был выбран створ на реке Протва у села Спас-Загорье, который освещает наблюдениями площадь водосбора 4 250 км². На территории бассейна и близ его границ расположены три метеорологические станции, ведущие наблюдения за температурой воздуха и количеством осадков (рисунок 1а).



(a)



(б)

Рисунок 1. Схема водосбора реки Протва (а) с расположением метеорологических станций и гидрологических постов, гидрограф за расчетный период с 2003 по 2015 год (б, снизу) с нанесением рядов осадков (б, сверху) и температур (б, в середине)

Figure 1. Map of the catchment area of the Protva river (a) with the location of meteorological stations and river gauges, daily hydrograph (б, bottom), precipitation (б, top) and temperatures (б, middle) time-series for the period from 2003 to 2015

Использованные данные и методы

Исходные ряды наблюдений

Наборы исходных данных для каждого из бассейнов созданы на основе архивов стандартных наблюдений с суточным шагом на метеорологических станциях и гидрологических постах Росгидромета. Архивы созданы и поддерживаются в виде автоматизированной базы данных Института водных проблем Российской академии наук на основе информации, предоставленной Всероссийским научно-исследовательским институтом гидрометеорологической информации – Мировым центром данных.

Набор метеорологических данных включает ряды среднесуточных температур воздуха ($^{\circ}\text{C}$) и суточных сумм осадков (мм), измеренных также на 3 метеостанциях в пределах этого бассейна (Можайск, Малоярославец, Обнинск) за период 2003–2015 годов. Ряд среднесуточных расходов воды сформирован за тот же период времени по данным наблюдений на гидрометрическом посту в селе Спас-Загорье.

Подготовка рядов исходных данных

Для проведения вычислительных экспериментов были подготовлены матрицы исходных данных с 2003 по 2015 год. Расчеты выполнялись для каждого дня в году. На каждую дату выполнялся прогноз ежедневных расходов воды на определенную заблаговременность на основе информации, известной на момент выдачи прогноза. Обе матрицы включали набор из 58 предикторов – рядов ежедневных значений переменных, полученных на основе обработки исходной гидрометеорологической информации. В тех же матрицах содержались и по 7 предикторов – рядов прогнозируемой характеристики (расхода воды) – с заблаговременностью от 1 до 7 суток. В число предикторов входили (таблица 1):

— расходы воды в замыкающем створе $Q_{t-\tau}$, $\text{м}^3/\text{с}$, на день выдачи прогноза и за предысторию $\tau=1\dots 7$ суток;

— максимальные суточные изменения расходов $\Delta Q_{\text{max},\tau}$, $\text{м}^3/\text{с}$, за предысторию $\tau=2\dots 7$ суток;

— суточные суммы осадков $P_{t-\tau}$, мм, по отдельным метеостанциям на день выдачи прогноза и за предысторию в $\tau=1\dots 7$ суток;

Таблица 1. Структура строки матриц исходных данных

Table 1. Structure of the input data matrices

Переменные	Номер позиции в строке	Описание
Предикаты		
Расходы воды	2-8	$Q_{t+\tau}, \tau=1...7$
«Динамические» предикторы		
Расходы воды	9-16	$Q_{t-\tau}, \tau=0...7$
Максимальные суточные изменения расхода за период	17-23	$\Delta Q_{max}, \tau=2...7$
Суточные суммы осадков по трем метеостанциям	24-47	$P_{t-\tau}^j, \tau=0...7,$ $j = 27\ 509, 27\ 606, 27\ 611$
Средние суточные суммы осадков	48-55	$P_{t-\tau}, \tau=0...7$
Средние суточные суммы осадков, накопленные за период	56-61	$\sum P_{\tau}, \tau=1...7$
«Инерционные» предикторы		
Индекс предшествующего увлажнения	62	I_W
Накопленные от начала сезона суммы температур выше пороговых значений 0, +2 и +5°C	63-65	$\sum T_0, \sum T_{+2}, \sum T_{+5}$
Суммарное от начала сезона расчетное испарение	66	$\sum PET$

— средние по трем метеостанциям осадки $\bar{P}_{t-\tau}$, мм, на день выдачи прогноза и за предысторию $\tau=1...7$ суток;

— накопленные средние по бассейну осадки $\sum P_{t-\tau}$, мм, за предысторию $\tau=1...7$ суток;

— накопленные от начала каждого теплого периода суммы суточных температур $\sum T$, °C, выше пороговых значений (0, +2 и +5°C);

— расчетные значения индекса предшествующего увлажнения I_W , мм, вычисляемого по данным о средних по бассейну осадках за 60 суток, предшествующих дате выдачи прогноза, по формуле:

$$I_W = P_t + \sum_{\tau=1}^4 P_{t-\tau} + 0,5 \sum_{\tau=5}^9 P_{t-\tau} + 0,3 \sum_{\tau=10}^{14} P_{t-\tau} + 0,2 \sum_{\tau=15}^{30} P_{t-\tau} + 0,1 \sum_{\tau=31}^{60} P_{t-\tau}$$

— накопленное от начала сезона расчетное испарение $\sum PET$, мм, представляющее потенциальное испарение суммарное с начала теплого периода, вычисляемое по суточным значениям PET с использованием формулы Одина [Oudin, Michel, Anctil, 2005]:

$$PET = \begin{cases} \frac{0,408 R_e (T+5)}{100}, & \text{при } T + 5 > 0, \\ 0, & \text{при } T + 5 \leq 0 \end{cases} \quad (2)$$

где PET – суммарное испарение за сутки, мм;

R_e – приход солнечной радиации на верхнюю границу атмосферы, Вт;

T – среднесуточная температура воздуха, °C.

Матрицы исходных данных организованы построчно, структура строки представлена в таблице 1. В первой позиции каждой строки располагается «текущая» дата – дата выдачи экспериментального прогноза t . В 9 позиции строки располагается расход воды на дату t – переменная Q_t . Позиции слева от Q_t , со 2 по 8, занимают предикаты – «будущие» расходы воды в порядке с Q_{t+7} по Q_{t+1} . Справа от Q_t последовательно располагаются предшествующие расходы воды на глубину учитываемой предыстории в порядке с Q_{t-1} по Q_{t-7} . Далее располагаются максимальные величины изменения расхода за предшествующий период от 2 до 7 суток ($\Delta Q_{max_2}... \Delta Q_{max_7}$). Далее следуют ежедневные осадки по трем станциям, для каждой из них сначала за текущую дату, затем последовательно за предшествующие дни от 1 до 7 суток. Далее в аналогичном порядке располагаются средние по

бассейну осадки, в обоих случаях вычисленные по трем станциям, то есть $P_1 \dots P_{t-7}$. Затем следуют накопленные суммы средних по бассейну осадков за несколько предшествующих дней – от двух ($\sum P_2$) до семи ($\sum P_7$). На этом заканчивается совокупность быстроизменяющихся переменных, которые можно назвать «динамическими», используемых в качестве предикторов.

Далее следует набор предикторов, которые можно обозначить как «инерционные». Это индекс предшествующего увлажнения I_W , накопленные суммы положительных температур $\sum T$ выше 0, +2 и +5°C, а также суммарная испаряемость $\sum PET$. Таким образом, каждая строка матрицы исходных данных представляет собой отдельную структуру данных, включающую полный набор значений испытываемых предикторов и предикторов на заданную дату. Общая длина (количество строк) матрицы исходных данных составила 4 727 строк.

Таким образом, динамические предикторы учитывают информацию на день выдачи прогноза и на глубину предыстории в 7 суток. Такова же и предельная заблаговременность, на которую испытываются методы прогноза. Это обусловлено соображениями, что для прогноза на определенную заблаговременность необходимо привлечь информацию за предысторию процесса такой же длительности или более. Выбор предельного интервала 7 суток обусловлен предварительно выполненным анализом коэффициентов корреляций стока с осадками на различных интервалах времени. Анализ выполнялся как качественно, путем сопоставления отдельных гидрографов и гиетографов, так и методом расчета коэффициентов корреляции стока с осадками со скользящим сдвигом во времени. Оба метода дали сходные результаты: характерное время реакции (время добегания) для бассейна реки Протва – село Спас-Загорье составляет 3-4 суток. Таким образом, выбранный для тестирования предельный интервал 7 суток заведомо

охватывает значения времени добегания для любого из исследуемых паводков.

Случайное перемешивание исходных данных

В соответствии с требованиями подготовки данных моделей машинного обучения [Bhattacharya, Solomatine, 2006] набор данных должен быть разделен на обучающий и тестовый наборы достаточной длины и обладающих сходными статистическими характеристиками. Из-за стохастической природы речного стока последнее требование кажется трудновыполнимым при разделении исходных календарных рядов наблюдений, особенно когда для оптимизации параметров модели используется метод многократной кросс-валидации. При этом методе исходный обучающий набор дополнительно разделяется на 10 равных частей и обучение производится на 9/10 частях, а проверка на 1/10 части, затем части последовательно меняются. Указанные методы могут приводить к смещению статистических характеристик при простом разделении исходной выборки. Эта проблема была нами решена путем случайного перемешивания исходного набора данных, что значительно уменьшило изменчивость между подвыборками данных (рисунок 2).

При разделении исходного набора данных для реки Протва на 10 подвыборок случайная перетасовка показала снижение коэффициента вариации средних значений подвыборок на 72% (с 0,20 до 0,05 для неперемешанной и перемешанной выборок соответственно). Такая обработка исходного набора данных, в котором каждая строка представляет ряд «динамических» и «инерционных» предикторов, допускает подобное случайное перемешивание, сохраняя тем самым статистические взаимосвязи между данными без изменений. Полученный перемешанный набор данных был разделен на две подвыборки: 80% данных для обучения и 20% данных для проверки моделей.

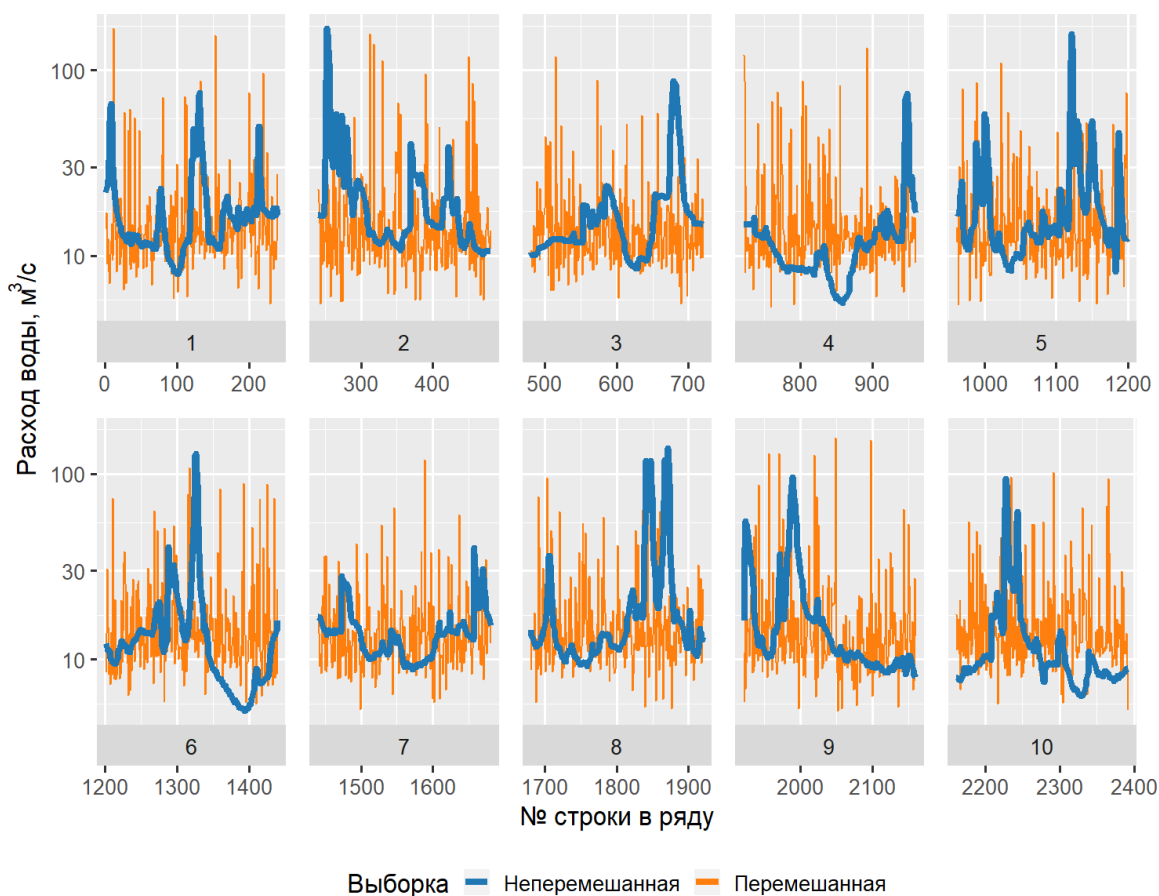


Рисунок 2. Сравнение исходной (неперемешанной) и перемешанной выборки
Figure 2. Shuffled (orange line) and initial (unshuffled, blue) input dataset intercomparison

Описанное выше случайное перемешивание не применялось при подготовке данных для модели с долгосрочной кратковременной памятью (Long Short-Term memory), потому что в этой модели предусмотрена другая процедура перемешивания [Abadi et al., 2016]. В модели используются 7 предикторов: расход воды Q_t , температуры T^j и суточные суммы осадков P^j на метеостанциях ($j=27509, 27606, 27611$). Для каждого набора предикторов $Q_{t+\tau}$ формируется подмножество предикторов со сдвигом 60 дней (шагов) назад по времени, и исходная выборка данных «нарезается» на «блоки», имеющие размерность $[7, 60]$, каждая из которых относится к одному набору из 7 предикторов $Q_{t+\tau}$, имеющему размерность $[7, 1]$. Можно заметить, что хронология исходного ряда таким образом не

нарушается и каждому значению предиката соответствует 60 значений предикторов. После этого такие «блоки» перемешиваются и случайным образом загружаются в модель для обучения.

Использованные модели и методы их построения

Программная реализация

Для исследования были построены несколько моделей машинного обучения. Построение моделей было выполнено на языке Python в интерактивной оболочке Jupyter Notebook с использованием библиотеки машинного обучения scikit-learn [Pedregosa et al., 2011] и Tensorflow [Abadi et al., 2016]. Исходные коды моделей и набор данных доступны в репозитории¹.

¹ Морейдо В.М. Модели машинного обучения для краткосрочных прогнозов расходов воды р. Протва – с. Спас-Загорье [Электронный ресурс] // GitHub. URL: https://github.com/esmoreido/protva_models (дата обращения: 15.12.2020).

Были запрограммированы и откалиброваны несколько основных модельных архитектур:

- модель множественной линейной регрессии;
- модель искусственной нейронной сети;
- модель с долгосрочной кратковременной памятью.

Модель множественной линейной регрессии

Модель множественной линейной регрессии (далее – LM) представляет собой линейную модель, оптимизирующую веса n входных переменных x : $W = \sum(Wx_1, \dots, Wx_n)$ по методу наименьших квадратов. Кроме весов входных переменных модель не имеет оптимизируемых параметров.

Модель ИНС по типу многослойного перцептрона

Модель искусственной нейронной сети по типу многослойного перцептрона с нелинейной функцией активации на скрытых слоях [Haykin, 2009], здесь и далее обозначаемая MLP, представляет собой одну из многочисленных архитектур ИНС, основанных на искусственных нейронах. Искусственный нейрон представляет собой нелинейную функцию от единственного аргумента – линейной комбинации всех входных сигналов. ИНС – это подход к обучению для аппроксимации функций путем соединения элементов, называемых нейронами (рисунок 3), между слоями.

Нейрон j получает входные данные в виде векторов используемых предикторов $x_1 - x_i$, с присвоенными им весами $a_{1j} - a_{ij}$ и вычисляет нелинейное преобразование для приведения значения в диапазон $[0, 1]$ (рисунок 3а). В качестве функций преобразования, называемых активационными, используются, например,

логистическая функция (сигмоида), гиперболический тангенс, полулинейный выпрямительный элемент (ReLU) и др. Выходные данные с нейрона j затем отправляются на следующий слой нейросети с весами $a_{jk} - a_{jn}$. Связи представлены линейными весами между нейронами, например, a_{jk} обозначает связь между j -м нейроном и k -м нейроном, которые находятся на уровне $l-1$ и l , соответственно. Многослойная сеть, показанная на рисунке 3а, называется полностью связанной, поскольку каждый нейрон на одном слое связан с каждым нейроном на соседнем слое. Слои ввода и вывода напрямую связаны со входами и выходами и поэтому называются видимыми слоями, в отличие от скрытых слоев посередине. Веса сети определяются путем минимизации целевой функции, например по методу наименьших квадратов между данными расчетов и наблюдений. Процесс определения весов ИНС называется обучением, а использование данных наблюдений для оптимизации целевой функции обуславливает название процесса подбора весов – «обучение с учителем». Наиболее распространенным методом обучения является обратное распространение ошибки [Rumelhart et al., 1986], который представляет собой форму метода градиентного спуска, передающего величину ошибки обратно в ИНС и обновляющего все значения весов на каждом расчетном шаге. По сравнению с другими методами оптимизации, такими как эволюционные алгоритмы, обратное распространение ошибок очень эффективно благодаря дифференциации и линейному характеру сетевых соединений. Количество нейронов в слое называется шириной, а количество слоев называется глубиной сети.

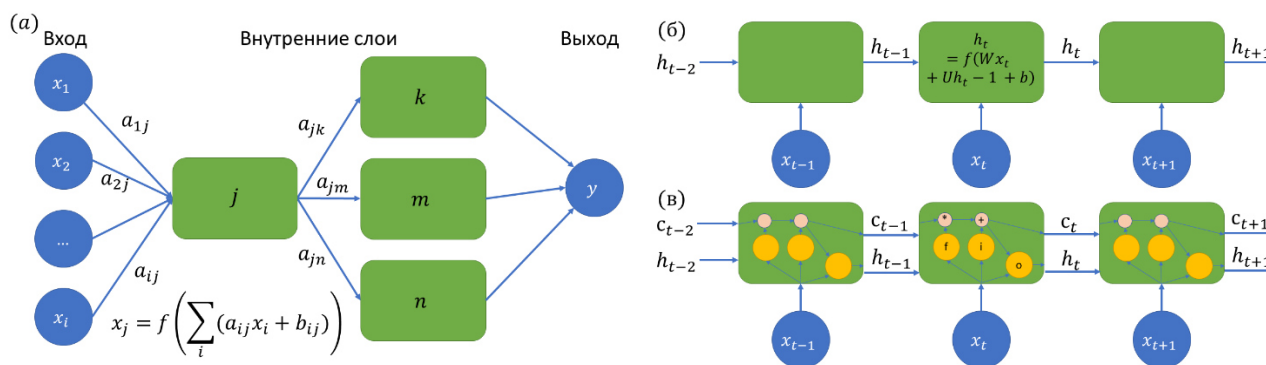


Рисунок 3. Схема нейронов ИНС по типу MLP (а), RNN (б) и LSTM (в). Обозначения: $x_1 - x_i$ - векторы переменных на входе; $a_{1j}-a_{ij}$ - веса входных переменных; a_j - трансформированное значение в нейроне j ; b_{ij} - свободный член; $a_{jk}-a_{in}$ - веса трансформированных переменных между слоями нейросети; k, m, n - нейроны на следующем внутреннем слое; y - вектор вывода; $x_{t\pm\tau}$ - входной вектор предикторов на временном шаге τ , $\tau = 1 \dots n$ шагов по времени; $h_{t\pm\tau}$ - матрица весов нейронов на временном шаге τ ; $c_{t\pm\tau}$ - матрица переменных состояния на временном шаге τ ; f - регулятор памяти; i - регулятор входа; o - регулятор выхода

Figure 3. Diagram of a neuron of an artificial neural network of the MLP type (a), a recurrent neural network RNN (b) and LSTM (c). Notation: $x_1 - x_i$ - vectors of variables at the input; $a_{1j}-a_{ij}$ - weights of input variables; a_j - transformed value in neuron j ; b_{ij} - bias; $a_{jk}-a_{in}$ - weights of the transformed variables between the layers of the neural network; k, m, n - neurons on the next inner layer; y is the output vector; $x_{t\pm\tau}$ - input vector of predictors at the time step τ , $\tau = 1 \dots n$ time steps; $h_{t\pm\tau}$ - matrix of neurons' weights at the time step τ ; $c_{t\pm\tau}$ - matrix of state variables at the time step τ ; f - forget gate; i - input gate; o - output gate

Модель с долгосрочной кратковременной памятью

Модель с долгосрочной кратковременной памятью (Long short-term memory, далее – LSTM) – модификация рекуррентной нейронной сети (Recurrent Neural Network, RNN) для глубокого обучения [Hochreiter, Schmidhuber, 1997] с дополнительными переменными состояния, сохраняющими веса отдельных нейронов и передающими их между расчетными шагами при обучении нейросети, если их значение приводит к снижению ошибки модели. LSTM сохраняет информацию о состояниях между расчетными шагами и на основании параметров регуляторов памяти определяет, когда как долго сохранять эту информацию и когда её можно «забыть». Простые рекуррентные нейронные сети RNN обновляли только вектор одного прошлого состояния (рисунок 3б). Состояние ячейки h_t служит памятью системы, чем-то похожим на состояние в моделях авторегрессии. Вектор состояний h_t умножается на обучаемые веса и затем

комбинируется с входными данными для перехода к следующему временному уровню. RNN обучаются с использованием алгоритма обратного распространения во времени [Mozer, 1989; Werbos, 1990], который применяет обратное распространение после первого развертывания сети. Как и в случае с другими алгоритмами обратного распространения, функция потерь распространяется в обратном направлении для определения обновлений весов. В простых рекуррентных сетях обратное распространение сталкивается с проблемой исчезающего градиента [Hochreiter, 1998]. В общих чертах эта проблема может быть описана так: поскольку состояния из более ранних временных шагов многократно умножались на векторы весов, прежде чем они влияли на конечный результат, влияние обратного распространения на них уменьшалось экспоненциально до нуля. В результате простые RNN обучались очень медленно. Решением этой проблемы в архитектуре LSTM являются специально

разработанные блоки, называемые регуляторами входа, выхода и памяти (рисунок 3в). Эти регуляторы, которые сами по себе являются нейронами с обучаемыми весами, контролируют поток информации для переменной состояния (c_i). После обучения входной регулятор контролирует, какие входные данные достаточно важны, чтобы их запомнить. Регулятор памяти решает, как долго и какое прошлое состояние памяти должно храниться. Выходной регулятор определяет, какая часть переменной состояния c_i используется для вектора выходных данных o . Вместе они позволяют сети запоминать и передавать информацию из отдаленных предыдущих состояний, отбрасывая несущественную информацию.

Методы оценки качества моделей

Для оценки результатов прогнозирования были использованы следующие метрики прогнозов. Среднеквадратическое отклонение смоделированного ряда от фактического рассчитывалось как

$$S = \sqrt{\frac{1}{N} \sum_{i=1}^N (Q_i^p - Q_i^o)^2} \quad (3)$$

где Q_i^p – смоделированный расход воды ($\text{м}^3/\text{с}$);
 Q_i^o – наблюдаемое значение ($\text{м}^3/\text{с}$);
 N – длина оцениваемого ряда.

Эффективность прогнозирования относительно среднего рассчитана по критерию эффективности Нэша-Сатклиффа:

$$NSE = 1 - \frac{\sum_{i=1}^N (Q_i^p - Q_i^o)^2}{\sum_{i=1}^N (Q_i^o - \bar{Q}^o)^2} \quad (4)$$

Для оценки преимущества моделей относительно прогноза «по тенденции» рассчитывался показатель [Борщ, Христофоров, 2015]:

$$\frac{S}{\sigma_\Delta} \quad (5)$$

где S – среднеквадратическая ошибка прогноза (3);

σ_Δ – отклонение прогноза «по тенденции» для каждой заблаговременности:

$$\sigma_\Delta = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\Delta_i - \bar{\Delta})^2} \quad (6)$$

$$\Delta_i = Q_i^o - Q_{i+\tau}^o, \tau = 1 \dots 7 \quad (7)$$

$$\bar{\Delta} = \frac{1}{N} \sum_{i=1}^N \Delta_i \quad (8)$$

Значения показателя S/σ_Δ находятся в интервале $[0; +\infty)$; значения между 0 и 0,5 свидетельствуют об отличном качестве методики прогноза, 0,5–0,7 – хорошо, 0,7–1,0 – удовлетворительном, более 1 – непригодном. Для расчета этого показателя требуется использование хронологического (неперемешанного) исходного ряда наблюдений, методика сопоставления моделей между собой с его помощью описана в следующем разделе.

Результаты расчетов

Предварительно было проверено предположение, что перемешивание исходных хронологических рядов данных приводит к повышению устойчивости модели на этапе обучения и улучшению результатов на проверочной выборке. Для этого был проведен эксперимент с моделью LM. В ходе эксперимента одна и та же архитектура модели была настроена и проверена с помощью хронологических (неперемешанных) и перемешанных рядов, разделенных на одинаковые по длине обучающие и проверочные выборки. Расчет критериев S и NSE производился непосредственно по полученным прогностическим рядам и рядам наблюдений за каждую заблаговременность, а для расчета S/σ_Δ для каждой модели было взято значение σ_Δ по фактическому ряду. Сравнение результатов (рисунок 4) показало, что при сопоставимых значениях критериев качества на обучающей выборке на проверочной выборке модель, обученная на перемешанных рядах, показывает результаты, существенно превосходящие результаты модели, обученной на хронологическом ряду. Так, среднеквадратическая ошибка модели с перемешиванием на всех заблаговременностях в среднем лучше на $1,7 \text{ м}^3/\text{с}$ (до $3,75 \text{ м}^3/\text{с}$), NSE лучше на величину около 0,05, а S/σ_Δ – на 0,06 (до 0,12).

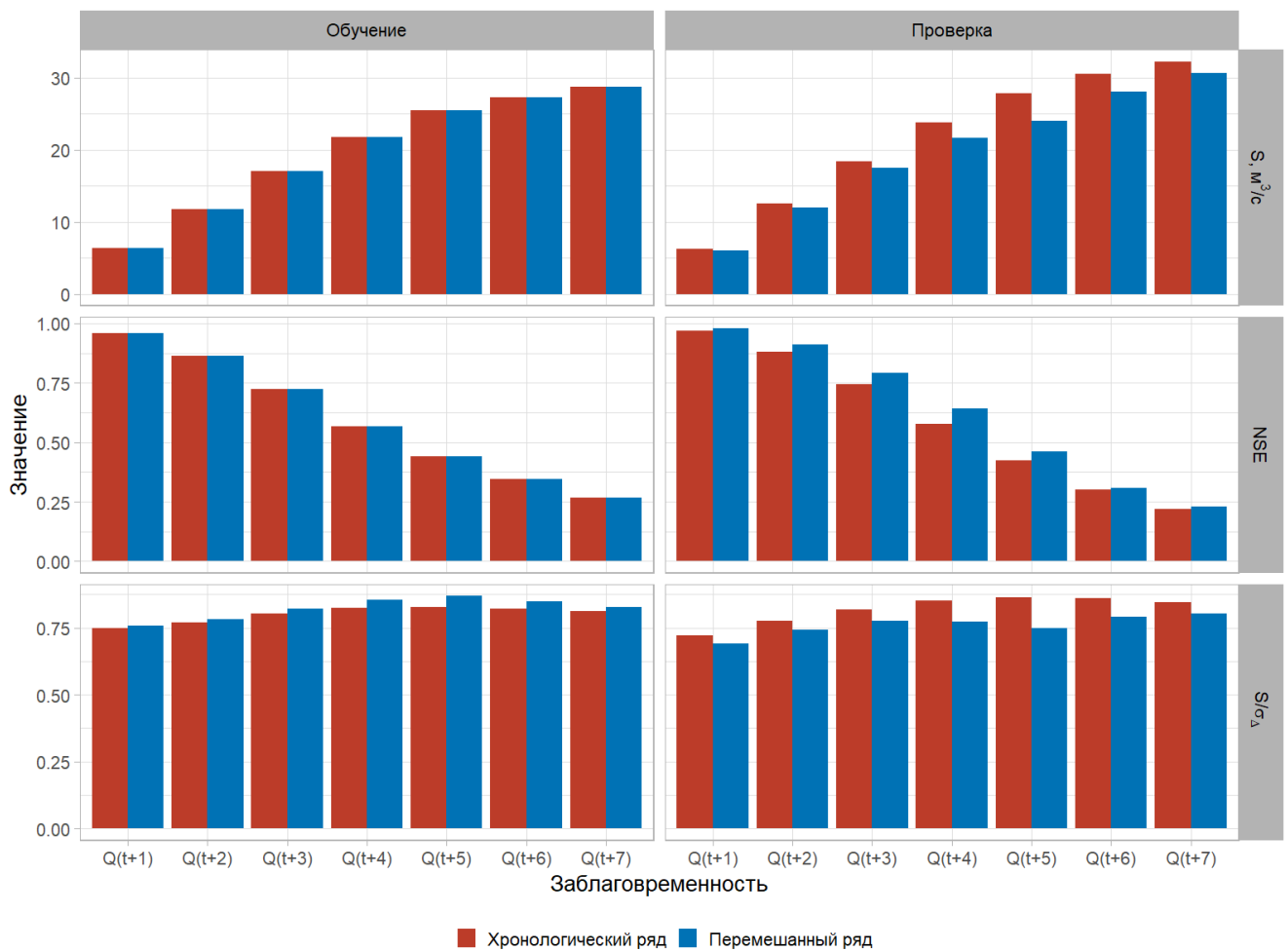


Рисунок 4. Сравнение результатов прогнозирования по модели LM с использованием хронологического и перемешанного ряда

Figure 4. Comparison of the LM model forecasting results trained on unshuffled (red bars) and shuffled (blue bars) time-series

В результате для расчета показателя S/σ_{Δ} для модели MLP также был использован описанный выше подход: обучение модели на перемешанных рядах предикторов с последующей проверкой на перемешанном и неперемешанном ряде. Процесс обучения и проверки модели LSTM был описан выше, но он также позволяет получить проверочные ряды в хронологическом порядке и сопоставить качество прогноза по критерию S/σ_{Δ} .

Обсуждение

Оценка моделей на проверочном ряде показала, что наиболее качественными с точки зрения ошибки прогноза и критериев эффективности оказались модели MLP и

LSTM_TP, учитывающие все имеющиеся в наличии предикторы (рисунок 5б).

Так, для первых шагов по заблаговременности все модели были достаточно аккуратны в предсказании расходов воды, что выражалось в небольших значениях среднеквадратической ошибки, хорошо согласовывались с фактическими рядами – высокий показатель эффективности NSE и «хорошие» значения S/σ_{Δ} , впрочем, кроме модели LM. На этих заблаговременностях, несомненно, это объясняется высоким коэффициентом автокорреляции. Однако далее, начиная с прогноза на 4 суток, качество прогнозирования ухудшалось по всем моделям и лишь MLP показало высокую эффективность и точность вплоть до седьмых прогнозных суток,

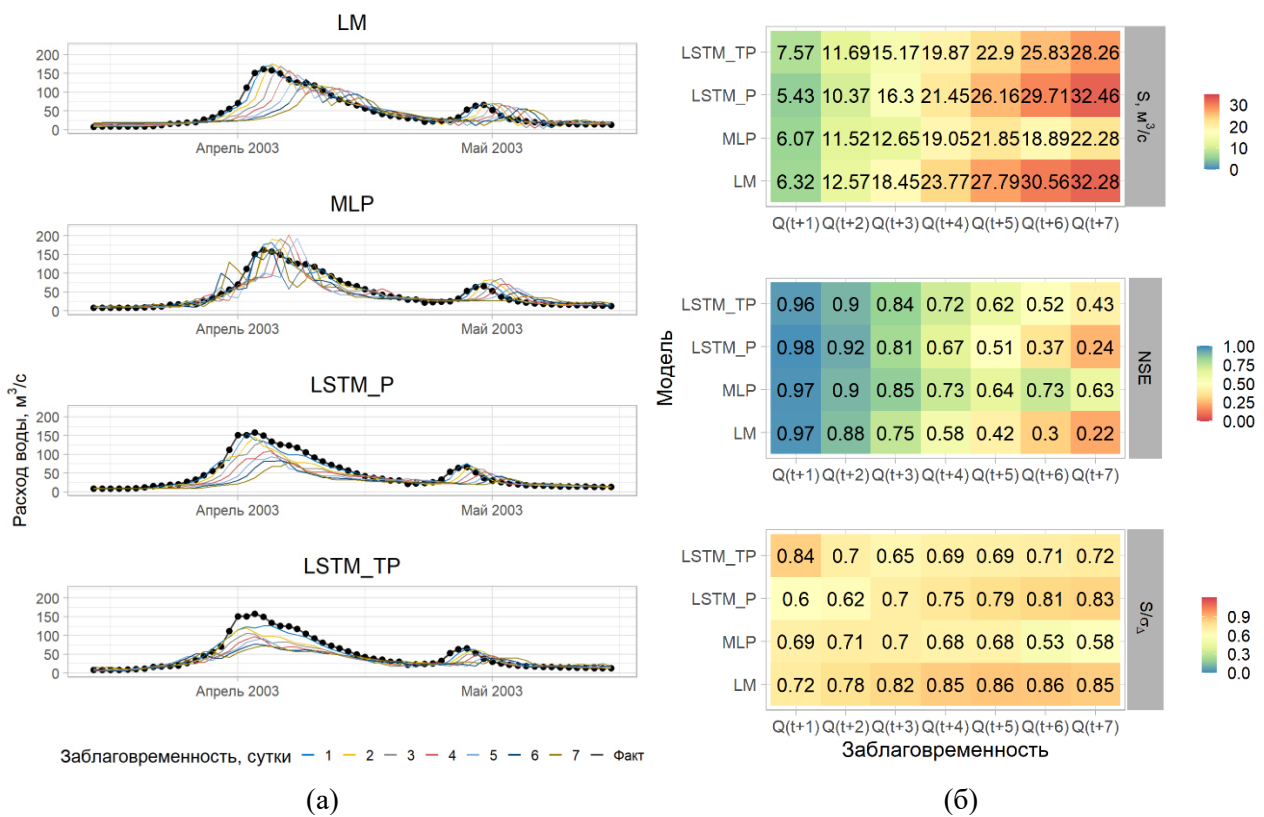


Рисунок 5. Фактический и прогнозные гидрографы реки Протва за период с 15 марта по 01 мая 2003 года (а) и результаты оценки качества моделей на проверочном ряде (б)
Figure 5. Observed and forecast hydrographs of the Protva river for period March, 15 – May, 01 2003 (a) and the results of the model quality assessment on the test dataset (б)

когда значения NSE не упали ниже 0,6, а значения S/σ_{Δ} даже начали снижаться. В целом следует отметить, что две модели нейронной сети, использовавшие в качестве входной информации не только предшествующие расходы воды, но и температуру воздуха и суточные суммы осадков, показали более высокое качество прогноза, чем линейная модель и модель LSTM без температур. В случае с линейной моделью, вероятно, это объясняется присваиванием более высоких весов значениям предшествующих расходов воды и недостаточно существенными весами для предикторов: температуры и осадков. В случае с LSTM без температуры – отсутствием учета возможного накопления осадков в переменных состояния, отвечающих за перераспределения осадков во времени.

Из полученных прогнозных гидрографов за период проверки видно (рисунок 5а), что всем моделям свойственно занижение значений расходов воды с увеличением заблаговременности, однако для модели MLP это

не столь существенно, а наличие колебаний на подъеме половодья конкретно взятого для иллюстрации 2003 года отражает реакцию на колебания температуры воздуха на водосборе.

Заключение

Выполненные расчеты позволили оценить применимость моделей ИНС в качестве основы для системы краткосрочных прогнозов расходов воды в речной системе Средней полосы России с преимущественно снеговым питанием. Расчеты показали, что ИНС могут быть применены для прогнозов при условии использования всей имеющейся гидрометеорологической информации – данных о предыдущих расходах воды, температуре воздуха на водосборе и количестве осадков. При этом создается возможность для более устойчивого во времени прогноза, чем используемые в настоящее время модели множественной линейной регрессии.

В ходе численных экспериментов показана эффективность подготовки данных с включением

«инерционных» и «динамических» предикторов, в которые входят как среднесуточные данные наблюдений за гидрометеорологическими характеристиками на водосборе, так и предикторы, рассчитанные на их основе – сумма и разность значений в скользящем окне, индексы увлажнения за длительный период времени и потенциальное испарение с водосбора. Также было опробовано случайное перемешивание исходных хронологических рядов данных для повышения выборочной устойчивости на

Благодарности

Работа выполнена в рамках темы № 0147-2019-0001 (№ государственной регистрации АААА-А18-118022090056-0) Государственного задания ИВП РАН, гранта РФФ №17-77-30006 и гранта РФФИ №19-05-00087 А.

Литература

Белякова П.А., Борщ С.В., Христофоров А.В., Юмина Н.М. Прогноз максимального стока рек Черноморского побережья Кавказа // Водное хозяйство России: проблемы, технологии, управление. 2013. № 6. С. 4-16.

Борщ С.В., Христофоров А.В. Оценка качества прогнозов речного стока // Труды Гидрометеорологического научно-исследовательского центра Российской Федерации. 2015. Спец. вып. 355. 198 с.

Зайков Б.Д. Средний сток и его распределение в году на территории СССР. Л.: изд-во и 2-я типолитограф. Гидрометеоздата, 1946. 148 с.

Кучмент Л.С., Гельфан А.Н. Динамико-стохастические модели формирования речного стока. М.: Наука, 1993. 101 с.

Мотовилов Ю.Г. Гидрологическое моделирование речных бассейнов в различных пространственных масштабах. 1. Алгоритмы генерализации и осреднения // Водные Ресурсы. 2016а. Т. 43. № 3. С. 243-253. DOI: [10.7868/S0321059616030111](https://doi.org/10.7868/S0321059616030111).

Мотовилов Ю.Г. Гидрологическое моделирование речных бассейнов в различных пространственных масштабах. 2. Результаты

обучающей и проверочной подвыборках и показано эффективное улучшение качества прогноза при таком перемешивании.

Программная реализация моделей ИНС находится на высоком уровне, не требующем существенных специальных знаний (кроме общих базовых навыков программирования). Таким образом, следует проводить дальнейшие исследования в направлении использования ИНС в прогностических целях, в том числе в различных природных зонах.

Acknowledgements

The work was supported by State assignment № 0147-2019-0001 (reg. №АААА-А18-118022090056-0) of the WPI RAS, RSCF grant №17-77-30006 and RFBR grant №19-05-00087А.

References

Abadi M., Barham P., Chen J., Chen Zh., Davis A., Dean J., Devin M., Ghemawat S., Irving G., Isard M., Kudlur M., Levenberg J., Monga R., Moore Sh., Murray D.G., Steiner B., Tucker P., Vasudevan V., Warden P., Wicke M., Yu Y. Zheng X. TensorFlow: A System for Large-Scale Machine Learning. *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16) (Savannah, GA, USA, November 2-4, 2016)*. Publ. USENIX Association, 2016, pp. 265-283.

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, vol. 36, iss. 4, pp. 480-513. DOI: [10.1177/0309133312444943](https://doi.org/10.1177/0309133312444943).

Belyakova P.A., Borshch S.V., Khristoforov A.V., Yumina N.M. Prognoz maksimal'nogo stoka rek Chernomorskogo poberezh'ya Kavkaza [Forecast of the maximum runoff of the rivers of the Black Sea coast of the Caucasus]. *Vodnoe khozyaistvo Rossii: problemy, tekhnologii, upravlenie [Water sector of Russia: problems, technologies, management]*, 2013, no. 6, pp. 4-16. (In Russian).

испытаний // Водные Ресурсы. 2016b. Т. 43. № 5. С. 467-475. DOI: [10.7868/S0321059616050096](https://doi.org/10.7868/S0321059616050096).

Abadi M., Barham P., Chen J., Chen Zh., Davis A., Dean J., Devin M., Ghemawat S., Irving G., Isard M., Kudlur M., Levenberg J., Monga R., Moore Sh., Murray D.G., Steiner B., Tucker P., Vasudevan V., Warden P., Wicke M., Yu Y. Zheng X. TensorFlow: A System for Large-Scale Machine Learning // Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16) (Savannah, GA, USA, November 2-4, 2016). Publ. USENIX Association, 2016. P. 265-283.

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. Vol. 36. Iss. 4. Pp. 480-513. DOI: [10.1177/0309133312444943](https://doi.org/10.1177/0309133312444943).

Beven K. Deep learning, hydrological processes and the uniqueness of place // Hydrological Processes. 2020. Vol. 34. Iss. 16. Pp. 3608-3613. DOI: [10.1002/hyp.13805](https://doi.org/10.1002/hyp.13805).

Bhattacharya B., Solomatine D.P. Machine learning in sedimentation modelling // Neural Networks. 2006. Vol. 19. Iss. 2. Pp. 208-214. DOI: [10.1016/j.neunet.2006.01.007](https://doi.org/10.1016/j.neunet.2006.01.007).

Emerton R.E., Stephens E.M., Pappenberger F., Pagano Th.C., Weerts A.H., Wood A.W., Salamon P., Brown J.D., Hjerdt N., Donnelly Ch., Baugh C.A., Cloke H.L. Continental and global scale flood forecasting systems // Wiley Interdisciplinary Reviews: Water. 2016. Vol. 3. Iss. 3. Pp. 391-418. DOI: [10.1002/wat2.1137](https://doi.org/10.1002/wat2.1137).

Fatichi S., Vivoni E.R., Ogden F.L., Ivanov V.Y., Mirus B., Gochis D., Downer Ch.W., Camporese M., Davison J.H., Ebel B., Jones N., Kim J., Mascaro G., Niswonger R., Restrepo P., Rigon R., Shen Ch., Sulis M., Tarboton D. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology // Journal of Hydrology. 2016. Vol. 537. Pp. 45-60. DOI: [10.1016/j.jhydrol.2016.03.026](https://doi.org/10.1016/j.jhydrol.2016.03.026).

Beven K. Deep learning, hydrological processes and the uniqueness of place. *Hydrological Processes*, 2020, vol. 34, iss. 16, pp. 3608-3613. DOI: [10.1002/hyp.13805](https://doi.org/10.1002/hyp.13805).

Bhattacharya B., Solomatine D.P. Machine learning in sedimentation modelling. *Neural Networks*, 2006, vol. 19, iss. 2, pp. 208-214. DOI: [10.1016/j.neunet.2006.01.007](https://doi.org/10.1016/j.neunet.2006.01.007).

Borsch S.V., Khristoforov A.V. Otsenka kachestva prognozov rechnogo stoka [Hydrologic flow forecast verification]. *Trudy Gidrometeorologicheskogo nauchno-issledovatel'skogo tsentra Rossiiskoi Federatsii [Proceedings of the Hydrometeorological Research Center of the Russian Federation]*, 2015, special issue 355. 198 p. (In Russian; abstract in English).

Emerton R.E., Stephens E.M., Pappenberger F., Pagano Th.C., Weerts A.H., Wood A.W., Salamon P., Brown J.D., Hjerdt N., Donnelly Ch., Baugh C.A., Cloke H.L. Continental and global scale flood forecasting systems. *Wiley Interdisciplinary Reviews: Water*, 2016, vol. 3, iss. 3, pp. 391-418. DOI: [10.1002/wat2.1137](https://doi.org/10.1002/wat2.1137).

Fatichi S., Vivoni E.R., Ogden F.L., Ivanov V.Y., Mirus B., Gochis D., Downer Ch.W., Camporese M., Davison J.H., Ebel B., Jones N., Kim J., Mascaro G., Niswonger R., Restrepo P., Rigon R., Shen Ch., Sulis M., Tarboton D. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, 2016, vol. 537, pp. 45-60. DOI: [10.1016/j.jhydrol.2016.03.026](https://doi.org/10.1016/j.jhydrol.2016.03.026).

Gelfan A.N., Motovilov Y.G. Long-term hydrological forecasting in cold regions: retrospect, current status and prospect. *Geography Compass*, 2009, vol. 3, iss. 5, pp. 1841-1864. DOI: [10.1111/j.1749-8198.2009.00256.x](https://doi.org/10.1111/j.1749-8198.2009.00256.x).

Halff A.H., Halff H.M., Azmoodeh M. Predicting runoff from rainfall using neural networks. *Proceedings of Symposium Engineering Hydrology (San Francisco, California, July 25-30, 1993)*, 1993, pp. 760-765.

Haykin S. *Neural Networks and Learning Machines: 3rd ed.* Prentice Hall Publ., 2009. 906 p.

- Gelfan A.N., Motovilov Y.G. Long-term hydrological forecasting in cold regions: retrospect, current status and prospect // *Geography Compass*. 2009. Vol. 3. Iss. 5. Pp. 1841-1864. DOI: [10.1111/j.1749-8198.2009.00256.x](https://doi.org/10.1111/j.1749-8198.2009.00256.x).
- Halff A.H., Halff H.M., Azmoodeh M. Predicting runoff from rainfall using neural networks // *Proceedings of Symposium Engineering Hydrology (San Francisco, California, July 25-30, 1993)*. 1993. Pp. 760-765.
- Haykin S. *Neural Networks and Learning Machines*: 3rd ed. Prentice Hall Publ., 2009. 906 p.
- Haykin S. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, Upper Saddle River, NJ, United States, 1994. 768 p.
- Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions // *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 1998. Vol. 6. No. 2. Pp. 107-116. DOI: [10.1142/S0218488598000094](https://doi.org/10.1142/S0218488598000094).
- Hochreiter S., Schmidhuber J. Long Short-Term Memory // *Neural Computation*. 1997. Vol. 9. Iss. 8. Pp. 1735-1780. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- Kratzert F., Klotz D., Brenner C., Schulz K., Herrnegger M. Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, 2018, vol. 22, iss. 11, pp. 6005-6022. DOI: [10.5194/hess-22-6005-2018](https://doi.org/10.5194/hess-22-6005-2018).
- Kuchment L.S., Gelfan A.N. *Dinamiko-stokhasticheskie modeli formirovaniya rechnogo stoka [Dynamic-stochastic models of river runoff formation]*. Moscow, Publ. Nauka, 1993. 101 p. (In Russian).
- Motovilov Y.G. Hydrological simulation of river basins at different spatial scales: 1. Generalization and averaging algorithms. *Water Resources*. 2016a, vol. 43, iss. 3, pp. 429-437. DOI: [10.1134/S0097807816030118](https://doi.org/10.1134/S0097807816030118) (Russ. ed.: Motovilov Yu.G. Gidrologicheskoe modelirovanie rechnykh basseinov v razlichnykh prostranstvennykh masshtabakh. 1. Algoritmy generalizatsii i osredneniya. *Vodnye Resursy*, 2016a, vol. 43, no. 3, pp. 243-253. DOI: [10.7868/S0321059616030111](https://doi.org/10.7868/S0321059616030111)).
- Motovilov Y.G. Hydrological simulation of river basins at different spatial scales: 2. Test results. *Water Resources*, 2016b, vol. 43, iss. 5, pp. 743-753. DOI: [10.1134/S0097807816050092](https://doi.org/10.1134/S0097807816050092) (Russ. ed.: Motovilov Yu.G. Gidrologicheskoe modelirovanie rechnykh basseinov v razlichnykh prostranstvennykh masshtabakh. 2. Rezul'taty ispytaniy. *Vodnye Resursy*, 2016b, vol. 43, no. 5, pp. 467-475. DOI: [10.7868/S0321059616050096](https://doi.org/10.7868/S0321059616050096)).
- Mozer M.C. A Focused Backpropagation Algorithm for Temporal Pattern Recognition // *Complex Systems*. 1989. Vol. 3. Iss.4. Pp. 349-381.
- Oudin L., Michel C., Anctil F. Which potential evapotranspiration input for a lumped rainfall-runoff model? // *Journal of Hydrology*. 2005. Vol. 303. Iss. 1-4. Pp. 275-289. DOI: [10.1016/j.jhydrol.2004.08.025](https://doi.org/10.1016/j.jhydrol.2004.08.025).
- 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 // *Journal of Machine Learning Research*. 2011. Vol. 12. Pp. 2825-2830.

Rumelhart D.E., Hinton G.E., Williams R.J. Learning representations by back-propagating errors // *Nature*. 1986. Vol. 323. No. 6088. Pp. 533-536. DOI: [10.1038/323533a0](https://doi.org/10.1038/323533a0).

Werbos P.J. Backpropagation Through Time: What It Does and How to Do It // *Proceedings of the IEEE*. 1990. Vol. 78. No. 10. Pp. 1550-1560. DOI: [10.1109/5.58337](https://doi.org/10.1109/5.58337).

Mozer M.C. A Focused Backpropagation Algorithm for Temporal Pattern Recognition. *Complex Systems*, 1989, vol. 3, iss.4, pp. 349-381.

Oudin L., Michel C., Anctil F. Which potential evapotranspiration input for a lumped rainfall-runoff model? *Journal of Hydrology*, 2005, vol. 303, iss. 1-4, pp. 275-289. DOI: [10.1016/j.jhydrol.2004.08.025](https://doi.org/10.1016/j.jhydrol.2004.08.025).

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. *Journal of Machine Learning Research*, 2011, vol. 12, pp. 2825-2830.

Rumelhart D.E., Hinton G.E., Williams R.J. Learning representations by back-propagating errors. *Nature*, 1986, vol. 323, no. 6088, pp. 533-536. DOI: [10.1038/323533a0](https://doi.org/10.1038/323533a0).

Werbos P.J. Backpropagation Through Time: What It Does and How to Do It. *Proceedings of the IEEE*, 1990, vol. 78, no. 10, pp. 1550-1560. DOI: [10.1109/5.58337](https://doi.org/10.1109/5.58337).

Zaikov B.D. *Srednii stok i ego raspredelenie v godu na territorii SSSR [Average runoff and its distribution per year in the USSR]*. Leningrad, Publ. Gidrometeoizdat, 1946. 148 p. (In Russian).