

УДК 556:004.942

doi: 10.34987/vestnik.sibpsa.2021.79.13.017

ПРОБЛЕМЫ НЕЙРОСЕТЕВОГО И РЕГРЕССИОННОГО ПРОГНОЗИРОВАНИЯ УРОВНЯ ПАВОДКА РЕК

Шамсутдинова Т.М., к.ф.-м.н., доцент

ФГБОУ ВО Башкирский государственный аграрный университет

Аннотация. В статье обсуждаются вопросы разработки нейросетевых и регрессионных моделей для прогнозирования паводков рек. В частности, рассматривается круг проблем, связанных с гидрометеорологическим обеспечением прогнозирования паводковой ситуации, а также обсуждаются проблемы выбора параметров моделей прогноза и их компьютерной реализации. Приводится пример построения регрессионной и нейросетевой моделей для прогноза уровня реки Белая (Агидель).

Ключевые слова: паводок, прогнозирование, нейросетевая модель, регрессионный анализ, моделирование.

PROBLEMS OF NEURAL NETWORK AND REGRESSION FORECASTING THE RIVER FLOOD LEVEL

Shamsutdinova T.M., Ph.D. of Physico-mathematical Sciences, Docent

FSBEI HE Bashkir State Agrarian University

Annotation. The article is about the development of neural network and regression models for predicting river floods. In particular, a range of problems related to hydrometeorological support for forecasting a flood situation is considered, and issues of its implementation are discussed. An example of creating regression and neural network models for forecasting the level of the Belaya (Agidel) river is given.

Key words: flood, forecasting, neural network model, regression analysis, modeling.

Введение. Очевидно, что задача прогнозирования паводков рек является одной из актуальных проблем нашего времени, причем современные возрастающие объемы информации требуют применения все новых и новых компьютерных технологий. В частности, одной из таких технологий может стать концепция Data Mining, связанная с методами интеллектуальной «добычи» и анализа данных [1]. Примерами Data Mining являются, в том числе, методологии регрессионного и нейросетевого моделирования, позволяющие выявлять новые зависимости и закономерности с использованием методов глубокого анализа и машинного обучения.

Хотя задачам компьютерного моделирования и прогнозирования паводков рек в настоящий момент посвящено достаточное большое количество работ, нельзя назвать данную задачу полностью проработанной и решенной. Модель прогнозирования представляет собой сложную многопараметрическую систему, включающую целый ряд технических, информационных, программных, алгоритмических, методологических и организационных подсистем, причем в условиях поиска их оптимального взаимодействия.

Обзор литературы. В работе Л.А. Афонина [2] проводится анализ возможных проблем, возникающих при прогнозировании паводков и наводнений. В частности, называются такие глобальные проблемы, как отсутствие единых, надежных методик оперативного прогнозирования с использованием современных

аэрокосмических технологий и информационных систем; проблемы приборного оснащения измерений параметров наводнений на всех этапах его развития; отсутствие специализированной организационной структуры, которая бы реализовывала измерения и обработку данных на всем бассейне реки, а не на его отдельных участках. При этом в статье отмечается, что «решающая роль при прогнозировании принадлежит полноте и точности количественных характеристик выпадающих осадков, дефицита почвенной влаги, уровнях воды в реке, геометрических характеристик водосбора и водотока, параметров общей синоптической обстановки, ветра, атмосферного давления, гидрологии всего русла» [2, с. 147].

В статье П.Н. Ткаченко и М.В. Ваколина [3] говорится о проблемах современного гидрометеорологического обеспечения прогнозирования паводковой ситуации, отмечается недостаточное использование данных от систем дистанционного зондирования Земли в рамках спутникового мониторинга чрезвычайных ситуаций. При этом отмечается, что в настоящее время системы спутникового мониторинга предоставляют такие данные как, например, размеры снежного покрова, температура поверхности и воздуха, водная маска поверхности воды и грунтовых вод, солнечное излучение, коэффициент испарения с поверхности. Но отдельные региональные службы, связанные с прогнозированием паводковой ситуации, зачастую не имеют технических и методологических инструментов для обработки и анализа данной информации.

Говоря о проблемах нейросетевого моделирования паводковой ситуации, Л.И. Великанова отмечает, что «на практике в качестве исходных данных берутся таблицы данных наблюдений гидропостов, метеостанций и систем мониторинга, в которых могут отсутствовать некоторые значения исходных и целевых признаков» [4, с. 67]. Там же отмечается, что для создания эффективной модели прогноза необходимо знание влажности почвы на объекте в предпрогнозный период, но доступная информационная база по данному параметру чаще всего практически отсутствует.

Все это делает непростую многопараметрическую задачу нейросетевого и регрессионного прогнозирования паводков еще более сложной и неоднозначной. При этом возникает ряд проблем, связанных с неоднозначностью выбора оптимальной методики прогнозирования.

Неоднозначным является вопрос и о выборе параметров для прогнозирования, и о размерности обучающего набора данных с точки зрения длительности наблюдений.

В частности, в работе Т.П. Варшаниной, Д.В. Митусова и др. [5], приводится пример нейросетевой модели получения прогноза на 2-3 дня вперед с использованием данных по подъему воды и градиенту гидротермического поля за предшествующие 10 дней, включая показатель индикатора сезонности. Л.И. Великанова [4] приводит пример нейросетевой модели прогнозирования с использованием следующих данных наблюдений: замеры приращения уровня воды в паводковый период за двухлетний период и соответствующие данные гидрометеослужбы.

Я.В. Гребнев и А.В. Яровой, анализируя разнообразные факторы, формирующие половодье, в качестве ключевых показателей выделяют такие параметры как температура, количество осадков, рельеф местности, испарение и приводят пример разработанной нейросетевой модели на основе периода наблюдений более 5 лет [6]. Вопросы выбора параметров нейросетевых моделей также затрагиваются в работах [7; 8] и многих др.

Что касается параметров регрессионных моделей прогнозирования, то, например, в [9] отмечается, что наиболее информативными параметрами модели прогноза являются паводкообразующие осадки и показатели предпаводочного увлажнения, зависящего от запаса воды в снежном покрове. О важности учета снежного покрова (накопления снега) говорится и в статье [10], при этом приводится эмпирическая регрессионная формула вычисления водного эквивалента снежного покрова на основе данных дистанционного спутникового зондирования.

В работе [11] в качестве рассматриваемых параметров регрессионной модели учитываются уровень воды в реке, сбросы воды на близлежащей ГЭС, состояние ледяного покрова на реке. В [12] используются эмпирические данные по уровню воды и объему выпавших осадков за период более 5 лет.

Все это позволяет сделать вывод о том, что на данный момент методология нейросетевого и регрессионного моделирования носит в основном исследовательский характер, без устойчиво-сложившихся апробированных методик. Нет единой концепции выбора параметров обучающей выборки для построения и обучения модели, наблюдается нехватка (недоступность) эмпирических данных.

Материалы и методы. С целью изучения возможностей и проблем применения нейросетевых и регрессионных моделей к задачам прогнозирования паводков рек, были проанализированы данные наблюдений за паводковой ситуацией в бассейне реки Белая (Агидель) в районе г. Уфы и г. Стерлитамак. При этом были использованы архивные данные Башкирского управления по гидрометеорологии и мониторингу окружающей среды (<http://www.meteorb.ru/arhiv-urovney-tek>) за паводковые периоды реки Белая (Агидель) в апреле-мае 2018-2020 годов, а также соответствующие архивные данные метеослужб. Кроме этого, были изучены материалы, касающиеся прогнозирования паводковой ситуации в районе бассейна реки Белая (Агидель) [13; 14 и др.].

В ходе исследования были собраны и проанализированы следующие показатели: уровень воды, температура воды, дневная температура воздуха, ночная температура воздуха, осадки. Надо заметить, что для повышения точности модели прогноза первоначально планировалось введение в модель величины снежного покрова (как коэффициента увлажнения почвы), но систематизированные эмпирические показатели по данному параметру были недоступны (либо отсутствовали).

На рис. 1-2 представлены графики уровня весеннего паводка р. Белая в районе пунктов г.Уфа и г.Стерлитамак в 2018-2020 годах. Как видим из графиков, динамика характера распределения локальных экстремумов паводка существенно различается как по годам, так и по пунктам наблюдений.

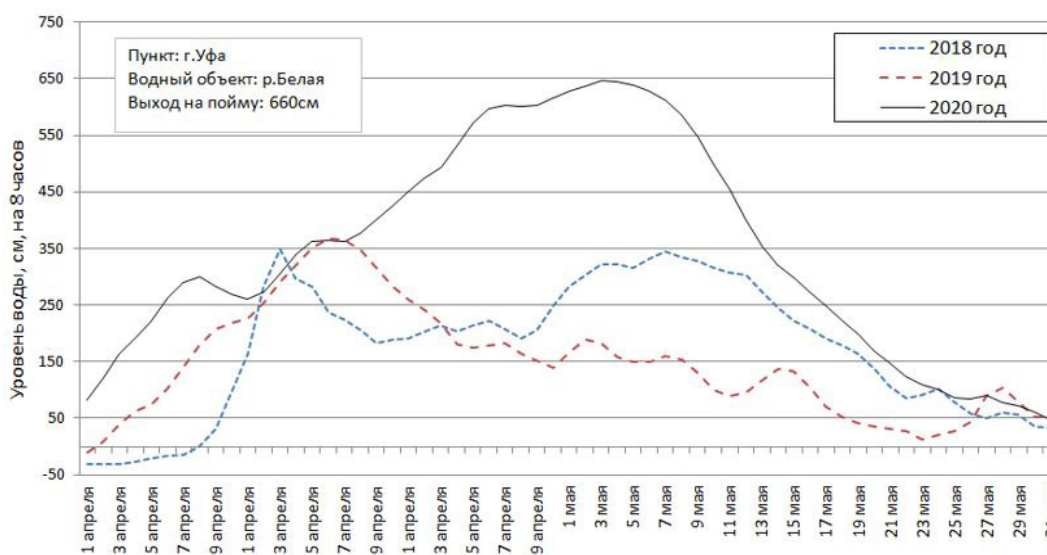


Рис. 1. Данные об уровне реки Белая в районе г.Уфа

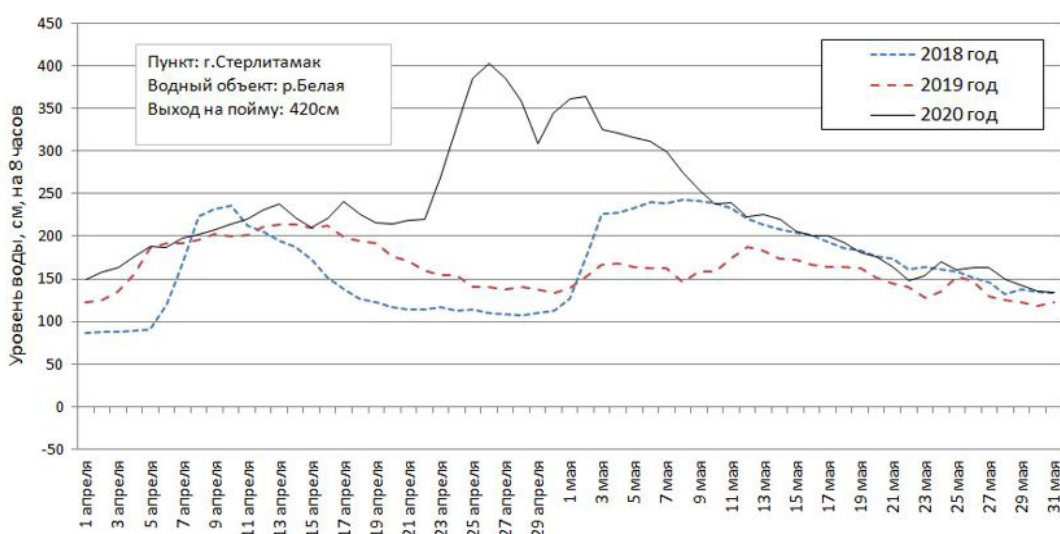


Рис. 2. Данные об уровне реки Белая в районе г.Стерлитамак

При этом данные имеют ряд выраженных экстремумов, что делает затруднительным процесс моделирования временного ряда с целью его дальнейшего прогнозирования. Реализация концепции Data Mining подразумевает этап очистки данных, связанной с устранением выраженных аномалий для возможности дальнейшей обработки данных достаточно сложными технологиями. Ввиду этого было проведено частичное сглаживание рядов данных с помощью вейвлет-преобразования. Эффективность применения метода вейвлет-преобразования в качестве предварительной обработки данных при прогнозировании временных рядов с помощью нейронных сетей показана, например, в работе [15].

Результаты и их обсуждение. Для реализации компьютерной модели прогноза данных был использован программный комплекс LOGINOM (с последующим экспортом полученных массивов данных в электронные таблицы MS Excel).

Разработанная компьютерная модель включала в себя следующий сценарий обработки данных:

- импорт исходных данных в систему LOGINOM;
- сглаживание данных с использованием вейвлет-преобразований;
- реализация прогноза данных с использованием модуля ARIMAX;
- визуализация полученных данных;
- экспорт полученных численных результатов в MS Excel.

Модуль ARIMAX (AutoRegressive Integrated Moving Average eXtended) представляет собой модель, предназначенную для анализа временных рядов на основе авторегрессии с возможностью учета внешних дополнительных факторов. Встроенный алгоритм авторегрессионного анализа при этом позволяет реализовать модель временного ряда с учетом предыдущих (ретроспективных) значений, т.е. фактически реализовать концепцию прогнозирования данных.

На рис. 3-5 представлены результаты визуализации регрессионных моделей прогноза по данным об уровне реки Белая в районе г.Уфа за апрель-май 2018-2020 г.г. соответственно (включая продолжение прогнозного ряда на начало июня). Как видим, наилучший по качеству прогнозирующий ряд построен по данным 2020 года, отличающихся наиболее гладкой динамикой и относительным отсутствием экстремально-пиковых выбросов. Наихудшее качество прогноза соответствует данным 2019 года, имеющим 6 выраженных локальных экстремумов и общую неравномерно-осциллирующую (колебательную) динамику. Прогнозирование такого рода функций представляет собой наиболее сложный процесс и вообще может быть успешно реализовано лишь для некоторых краткосрочных периодов.

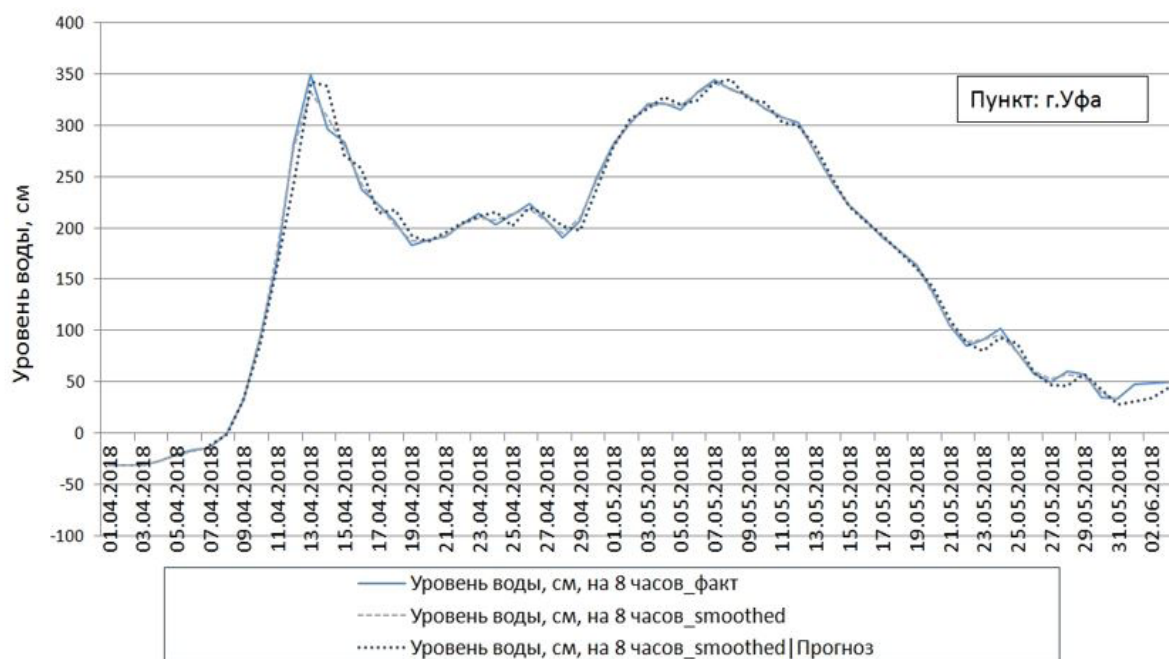


Рис. 3. Регрессионная модель прогноза по данным 2018 года

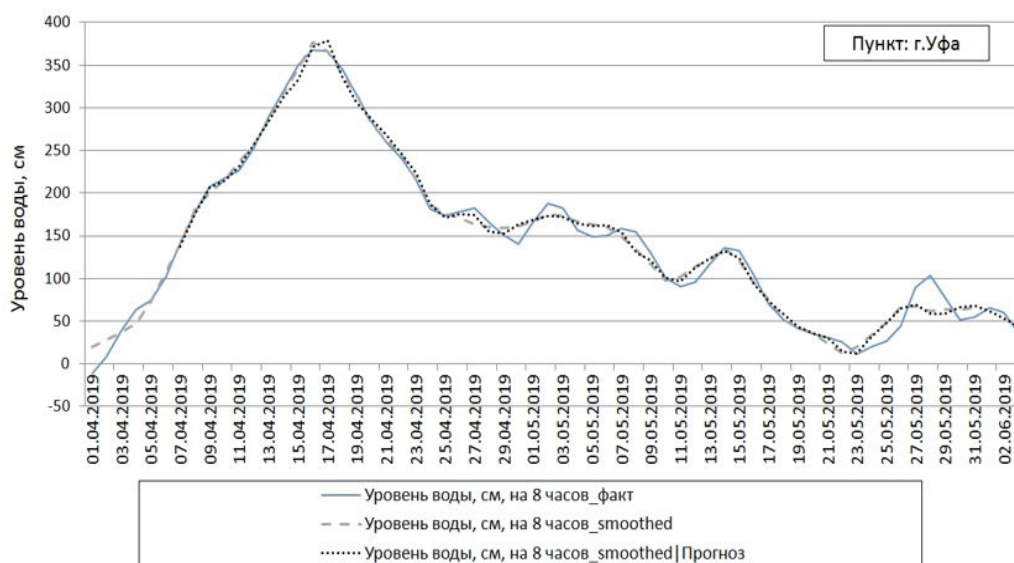


Рис. 4. Регрессионная модель прогноза по данным 2019 года

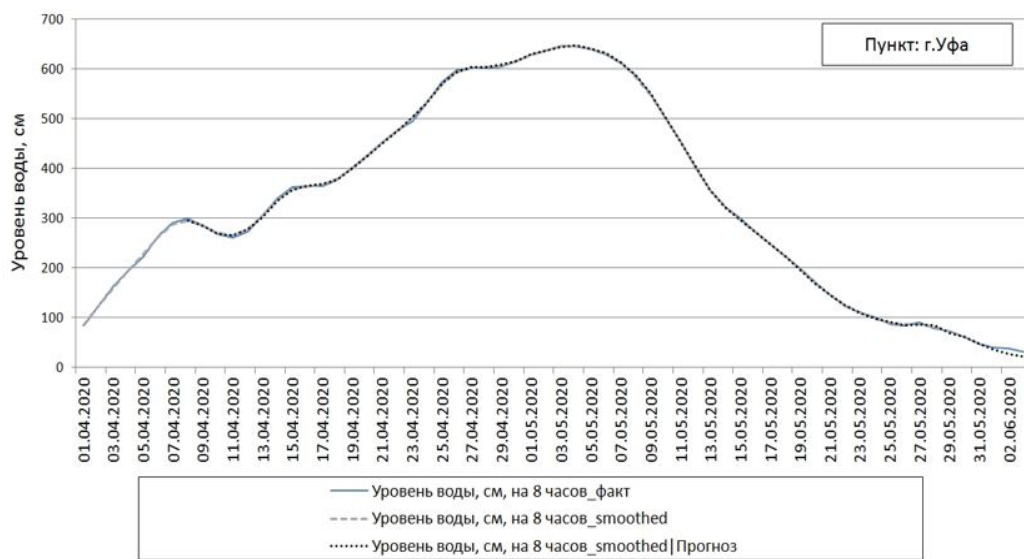


Рис. 5. Регрессионная модель прогноза по данным 2020 года

Кроме регрессионных моделей, по тем же исходным данным была реализована и модель прогноза на основе нейронной сети. Для построения нейронной сети использовалась аналитическая платформа Deductor Studio, позволяющая реализовывать достаточно большой круг задач класса Data Mining, в том числе, и реализацию модели прогноза временного ряда на основе нейронной сети с применением технологии «скользящего окна».

Наиболее адекватная модель прогноза была построена, опять же, по данным 2020 года (рис. 6), что объясняется наилучшим качеством исходного массива данных с точки зрения отсутствия большого числа ярко выраженных экстремальных и аномальных значений.

Качество построенной нейросетевой модели можно наглядно оценить с помощью ее диаграммы рассеяния, показывающей ошибку модели на основе принадлежности расчетных значений доверительному интервалу. Как видим на рис. 7, все рассчитанные данные находятся в пределах верхней и нижней границы доверительного интервала, что соответствует требуемому критерию качества.

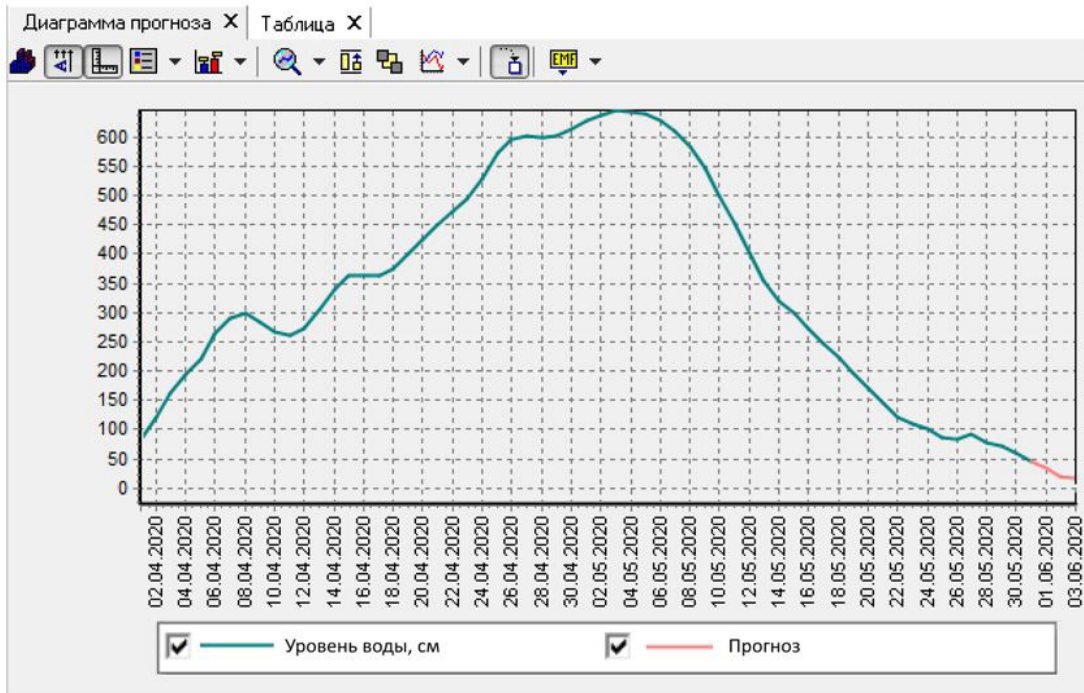


Рис. 6. Нейросетевая модель прогноза на 1-3 июня 2020 года

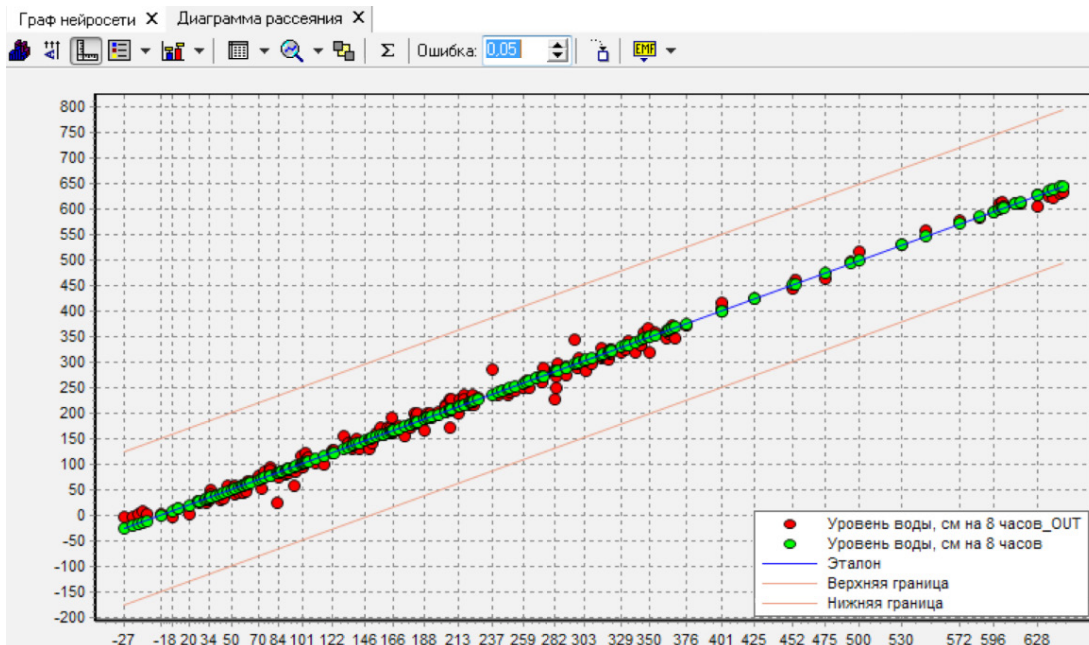


Рис. 7. Диаграмма рассеяния для нейросетевой модели прогноза

Выводы. Можно выделить следующие проблемы, связанные с компьютерным моделированием прогноза уровня паводка на основе нейросетевых и регрессионных моделей:

- недостаточное количество доступных разноплановых систематизированных данных наблюдений, необходимых для построения, обучения, тестирования модели;
- недостаточно изученный и проработанный вопрос об оптимальной структуре моделей (размерности исходной обучающей выборки данных, длительности наблюдений, количестве нейронов и слоев в нейросетевых моделях и т.д.);
- отсутствие единой методики предобработки исходных данных (результатов наблюдений) с точки зрения их сглаживания и устранения шумов и аномалий.

При этом можно сказать, что задача прогнозирования паводка рек относится к классу частично формализуемых задач, допускающих неоднозначный набор исходных данных и параметров моделирования. Тем не менее, надо заметить, что задача прогнозирования паводков является крайне востребованной, что делает данное моделирование очень актуальным по своим задачам и целям.

Литература

1. Гордеева С.М., Малинин В.Н. Использование Data Mining в задаче гидрометеорологического прогнозирования // Ученые записки Российского государственного гидрометеорологического университета. 2016. № 44. С. 30-44.
2. Афонин Л.А. Проблемы прогнозирования паводков и наводнений // Наука. Инновации. Технологии. 2014. № 1. С. 145-152.
3. Ткаченко П.Н., Вакорин М.В. Анализ проблемной ситуации использования информационных систем для прогнозирования паводка органами управления МЧС России // Сибирский пожарно-спасательный вестник. 2019. № 4 (15). С. 49-54.
4. Великанова Л.И. Прогноз уровня воды при прохождении паводков на горных реках с применением нейросетевых технологий и прогноза метеослужбы // Проблемы автоматизации и управления. 2013. № 1 (24). С. 66-73.
5. Варшанина Т.П., Митусов Д.В., Плисенко О.А., Стародуб И.В. Нейросетевая модель прогноза паводков на малых реках Адыгеи // Известия Российской академии наук. Серия географическая. 2007. № 6. С. 87-93.
6. Гребнев Я.В., Яровой А.В. Мониторинг и прогнозирование паводков на территории Красноярского края использованием нейросетевых алгоритмов // Сибирский пожарно-спасательный вестник. 2018. № 3 (10). С. 13-16.
7. Семенчин Е.А., Бараненко Ф.Ф., Войтюк А.В. Нейросетевое моделирование прогноза уровня воды на горно-равнинных реках // Экологические системы и приборы. 2010. № 11. С. 61-64.
8. Стручкова Г.П., Тимофеева В.В., Капитонова Т.А., Ноговицын Д.Д. Выбор структуры искусственных нейронных сетей для прогнозирования максимального уровня воды во время весеннего половодья на участке реки Лена // Проблемы безопасности и чрезвычайных ситуаций. 2020. № 1. С. 99-106.
9. Лариошкин В.В. Методика прогноза дождевых паводков в бассейне верхнего Амура (на примере р. Онон) // Известия Томского политехнического университета. Инжиниринг георесурсов. 2016. Т. 327. №11. С. 105–115.
10. Петров Д.О. Программный модуль анализа накопления снега для системы мониторинга и прогнозирования паводка // Вестник Брестского государственного технического университета. Водохозяйственное строительство, теплоэнергетика и геоэкология. 2014. № 2 (86). С. 134-137.
11. Буянов Д.И., Федотов Р.С., Ткаченко П.Н. Прогнозирование подъема уровня воды на реке Обь в Томской области на основе регрессионного анализа // Научные и образовательные проблемы гражданской защиты. 2015. № 2 (25). С. 112-118.
12. Анашечкин А.Д., Трофимец Е.Н. К вопросу построения краткосрочных моделей прогнозирования уровней воды на участках рек // Научно-аналитический журнал Вестник Санкт-Петербургского университета Государственной противопожарной службы МЧС России. 2019. № 1. С. 36-43.
13. Гайдукова Е.В., Баймаганбетов А.Е., Викторова Н.В., Александров В.Я. Краткосрочные прогнозы расходов воды в периоды дождевого паводка и половодья на р. Агидель (Белая) // Естественные и технические науки. 2019. № 12 (138). С. 210-216.
14. Красногорская Н.Н., Нафикова Э.В., Белозерова Е.А., Соколова О.В. Анализ опасности возникновения экстремальных гидрологических ситуаций на территории Республики Башкортостан // Эволюция современной науки. Сборник статей Международной научно-практической конференции. Уфа. 2017. С. 273.
15. Козадаев А.С. Предварительная обработка данных для искусственных нейронных сетей в задачах прогнозирования временных рядов. Вейвлет-преобразование // Вестник Тамбовского университета. Серия: Естественные и технические науки. 2008. Т. 13. № 1. С. 98-99.