



**ВСЕЛЕННАЯ  
БЕЛОГО МЕДВЕДЯ**  
НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ



ПРАВИТЕЛЬСТВО  
ЧУКОТСКОГО АВТНОМНОГО  
ОКРУГА



МИНИСТЕРСТВО ПРИРОДНЫХ  
РЕСУРСОВ И ЭКОЛОГИИ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

# МЕТОДИКА ДОЛГОСРОЧНОГО ПРОГНОЗИРОВАНИЯ МАКСИМАЛЬНЫХ УРОВНЕЙ ВОДЫ РЕКИ ТАЗ

Докладчик:  
Волкова Надежда Александровна,  
ФГБУ «ААНИИ»



## Цель исследования:

- Обобщение существующих методик прогнозирования и выявление перспективных направлений для дальнейших исследований, которые позволят улучшить точность и надежность долгосрочных прогнозов, а также способствовать обеспечению техносферной безопасности арктического региона.

## Актуальность исследования:

- В условиях изменяющегося климата и растущего интереса к арктическим ресурсам, разработка точных и надежных моделей для долгосрочного прогноза по рекам арктического бассейна становится неотъемлемой частью современной науки и практики управления водными ресурсами.



# Выбор модели

1.

- Линейная регрессия.

2.

- Регрессия с использованием деревьев решений.

3.

- Случайный лес (Random Forest).

4.

- Градиентный бустинг (Gradient Boosting).

5.

- Нейронные сети.

6.

- Метод опорных векторов (Support Vector Regression, SVR).



# Таблица 1. Преимущества и недостатки рассмотренных методов

Метод	Преимущества	Недостатки
<b>Линейная регрессия</b>	<ul style="list-style-type: none"> <li>- Простота реализации и интерпретации</li> <li>- Быстрая работа</li> <li>- Хороша для линейных зависимостей</li> </ul>	<ul style="list-style-type: none"> <li>- Неэффективна для нелинейных данных</li> <li>- Чувствительна к выбросам</li> <li>- Требуется нормализации данных</li> </ul>
<b>Деревья решений</b>	<ul style="list-style-type: none"> <li>- Обработка нелинейных зависимостей</li> <li>- Учет взаимодействий признаков</li> <li>- Нет необходимости в масштабировании данных</li> </ul>	<ul style="list-style-type: none"> <li>- Склонность к переобучению</li> <li>- Нестабильность при малых изменениях данных</li> <li>- Сложность интерпретации глубоких деревьев</li> </ul>
<b>Случайный лес</b>	<ul style="list-style-type: none"> <li>- Высокая точность прогноза</li> <li>- Снижение переобучения</li> <li>- Работа с большим количеством признаков</li> </ul>	<ul style="list-style-type: none"> <li>- Высокая вычислительная сложность</li> <li>- Сложность интерпретации</li> <li>- Требуется настройки гиперпараметров</li> </ul>



# Таблица 1. Преимущества и недостатки рассмотренных методов

Метод	Преимущества	Недостатки
<b>Градиентный бустинг</b>	<ul style="list-style-type: none"> <li>- Высокая точность на сложных задачах</li> <li>- Эффективная обработка нелинейных зависимостей</li> <li>- Возможность работы с пропущенными значениями</li> </ul>	<ul style="list-style-type: none"> <li>- Склонность к переобучению при неправильной настройке</li> <li>- Высокая вычислительная сложность</li> <li>- Требуется тщательной настройки параметров</li> </ul>
<b>Нейронные сети</b>	<ul style="list-style-type: none"> <li>- Способность моделировать сложные зависимости</li> <li>- Гибкость архитектуры</li> <li>- Работа с большими объемами данных</li> </ul>	<ul style="list-style-type: none"> <li>- Высокая вычислительная мощность необходима</li> <li>- Сложность настройки и интерпретации</li> <li>- Требуется большого объема данных для обучения</li> </ul>
<b>SVR (Опорные векторы)</b>	<ul style="list-style-type: none"> <li>- Эффективна для данных со сложной структурой</li> <li>- Возможность работы с нелинейными зависимостями через ядра</li> </ul>	<ul style="list-style-type: none"> <li>- Высокая вычислительная сложность на больших наборах данных</li> <li>- Требуется подбора ядра и гиперпараметров</li> <li>- Чувствительна к выбору параметров</li> </ul>



# Река Таз

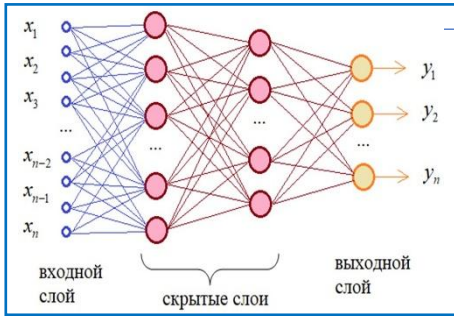


## Таблица 2. Ошибки прогноза максимального уровня воды р. Таз - Красноселькуп

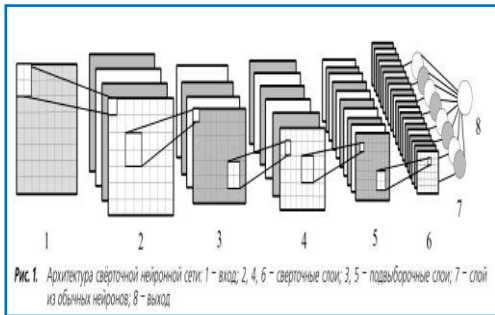
Метод	Средняя абсолютная ошибка прогноза, полученная на тренировочных данных, м	Средняя абсолютная ошибка прогноза, полученная на данных для валидации, м
Простая модель (по медиане)	0.37	0.30
Линейная регрессионная модель	0.00	0.65
Модель опорных векторов	0.13	0.38
Модель деревьев решений	0.00	0.51
Модель случайного леса	0.14	0.40
Регрессионная модель на основе градиентного бустинга	0.00	0.42
Полносвязная нейронная сеть	0.54	0.23
Свёрточная нейронная сеть	1.36	0.22



# Нейронные сети



## Полносвязная нейронная сеть



## Свёрточные одномерные нейронные сети

В ходе исследований проведено множество экспериментов с различными архитектурами нейронных сетей для решения задачи долгосрочного прогнозирования максимального уровня рек.





# Архитектура полносвязной нейронной сети

1. Полносвязный слой (Dense) с 128 нейронами. Используется функция активации ReLU (Rectified Linear Unit) для введения нелинейности в модель, что позволяет ей учиться сложным закономерностям в данных.
2. Слой нормализации пакетов (Batch Normalization) нормализует входные данные для каждого мини-пакета, что помогает стабилизировать и ускорить процесс обучения. Этот слой также способствует снижению чувствительности модели к инициализации весов и позволяет использовать более высокие скорости обучения.
3. Слой Dropout случайным образом отключает 50% нейронов на каждом шаге обучения. Это помогает предотвратить переобучение модели, заставляя её учиться более обобщённым признакам.
4. Второй полносвязный слой (Dense) с 64 нейронами. Используется функция активации ReLU также для введения нелинейности.
5. Второй слой нормализации пакетов (Batch Normalization) для стабилизации и ускорения обучения.
6. Второй слой Dropout для предотвращения переобучения.
7. Выходной слой (Dense) с одним нейроном, который предсказывает максимальный уровень реки. Используется линейная функция активации, так как задача представляет собой регрессию.

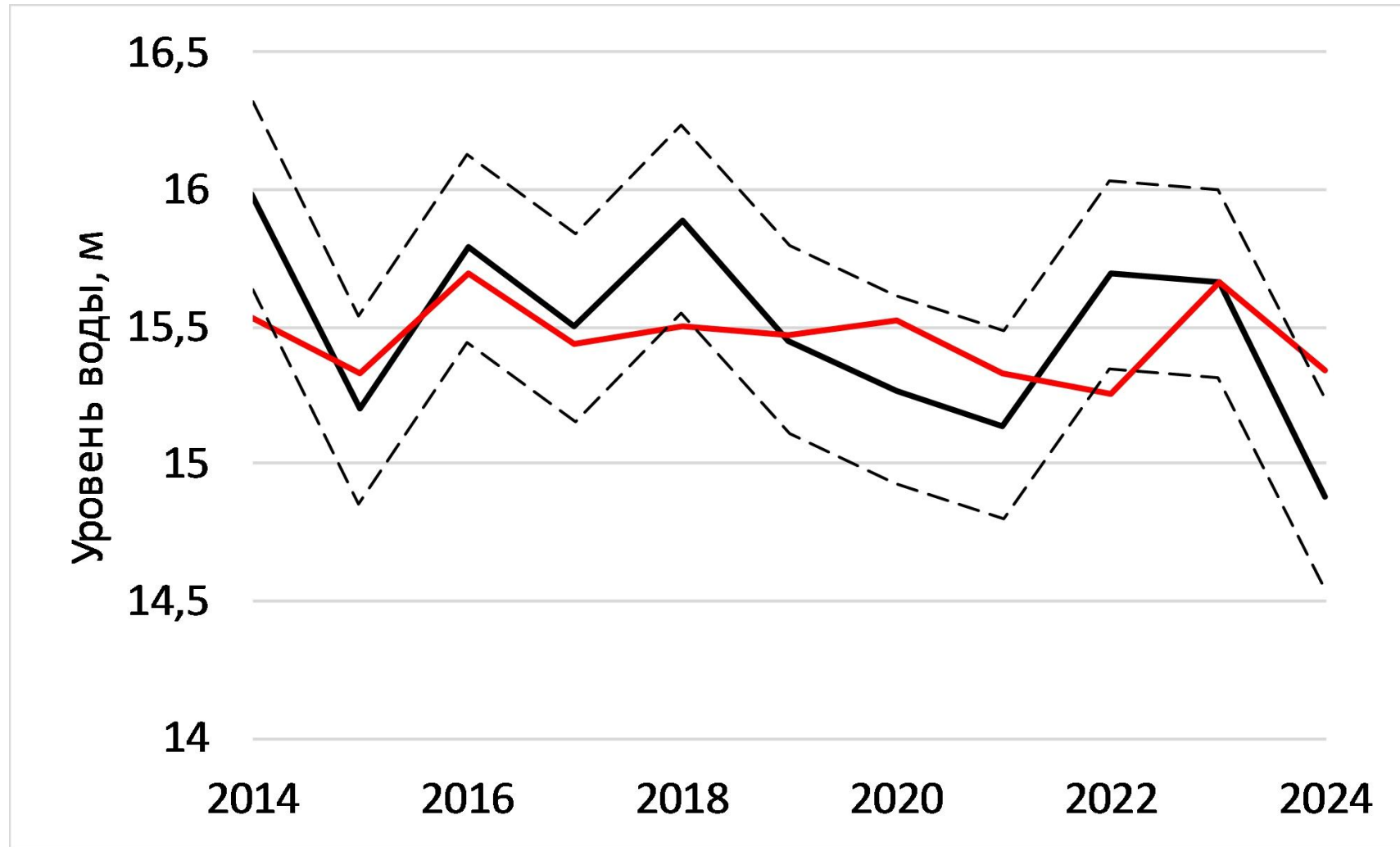


# Архитектура свёрточной нейронной сети

1. Одномерный свёрточный слой (Conv1D) с 128 фильтрами и размером ядра 4. Функция активации - ReLU. Входной формат данных (181, 1), что соответствует временному ряду длиной 181 с одним каналом.
2. Слой максимального пулинга (MaxPooling1D), который уменьшает размерность данных, сохраняя важные признаки. Размер пулинга равен 4, что означает, что каждые 4 элемента временного ряда заменяются их максимальным значением.
3. Слой нормализации пакетов (Batch Normalization) нормализует входные данные для каждого мини-пакета, что помогает стабилизировать и ускорить процесс обучения. Этот слой также способствует снижению чувствительности модели к инициализации весов и позволяет использовать более высокие скорости обучения.
4. Слой Dropout случайным образом отключает 50% нейронов на каждом шаге обучения. Это помогает предотвратить переобучение модели, заставляя её учиться более обобщённым признакам.
5. Слой Flatten, который преобразует многомерные данные в одномерный вектор. Это необходимо для передачи данных в полносвязные слои, которые ожидают одномерный вход.
6. Полносвязный слой (Dense) с 64 нейронами. Функция активации - ReLU.
7. Второй полносвязный слой (Dense) с 32 нейронами. Функция активации - ReLU.
8. Выходной слой (Dense) с одним нейроном, который предсказывает максимальный уровень реки. Используется, как и в полносвязной нейронной сети линейная функция активации.



# Хронологические графики фактических и прогнозных уровней воды р. Таз – с. Красноселькуп, 2014-2024 гг.



## Выводы

- Результаты показали, что полносвязная нейронная сеть с оптимизатором RMSprop и свёрточная одномерная нейронная сеть с оптимизатором RMSprop демонстрируют наилучшие результаты в средней абсолютной ошибке на данных для валидации и стабильности прогнозов.
- Важно отметить, что выбор оптимизатора играет ключевую роль в обучении нейронных сетей. Оптимизатор RMSprop показал наилучшие результаты для обеих архитектур нейронных сетей, что подчеркивает его эффективность в задачах прогнозирования временных рядов.
- Таким образом, данное исследование подтверждает, что использование нейронных сетей и адаптивных оптимизаторов позволяет значительно улучшить точность и надежность долгосрочных прогнозов максимального уровня воды в низовьях крупных рек Обско-Тазовской устьевой области.



# Заключение

К числу возможных путей повышения достоверности гидрологических прогнозов относятся следующие:

- 1) проверка качества и однородности исходных данных;
- 2) проверка логической непротиворечивости данных;
- 3) сопоставление и статистическая обработка данных;
- 4) определение и корректировка «весов» источников информации;
- 5) использование методов, основанных на нейросетевых моделях и нечеткой логике;
- 6) введение вероятностных прогнозов гидрологических характеристик.



# Благодарим за внимание!



Волкова Надежда Александровна

[navolkova@aari.ru](mailto:navolkova@aari.ru)

Ромашова Ксения Владимировна

