

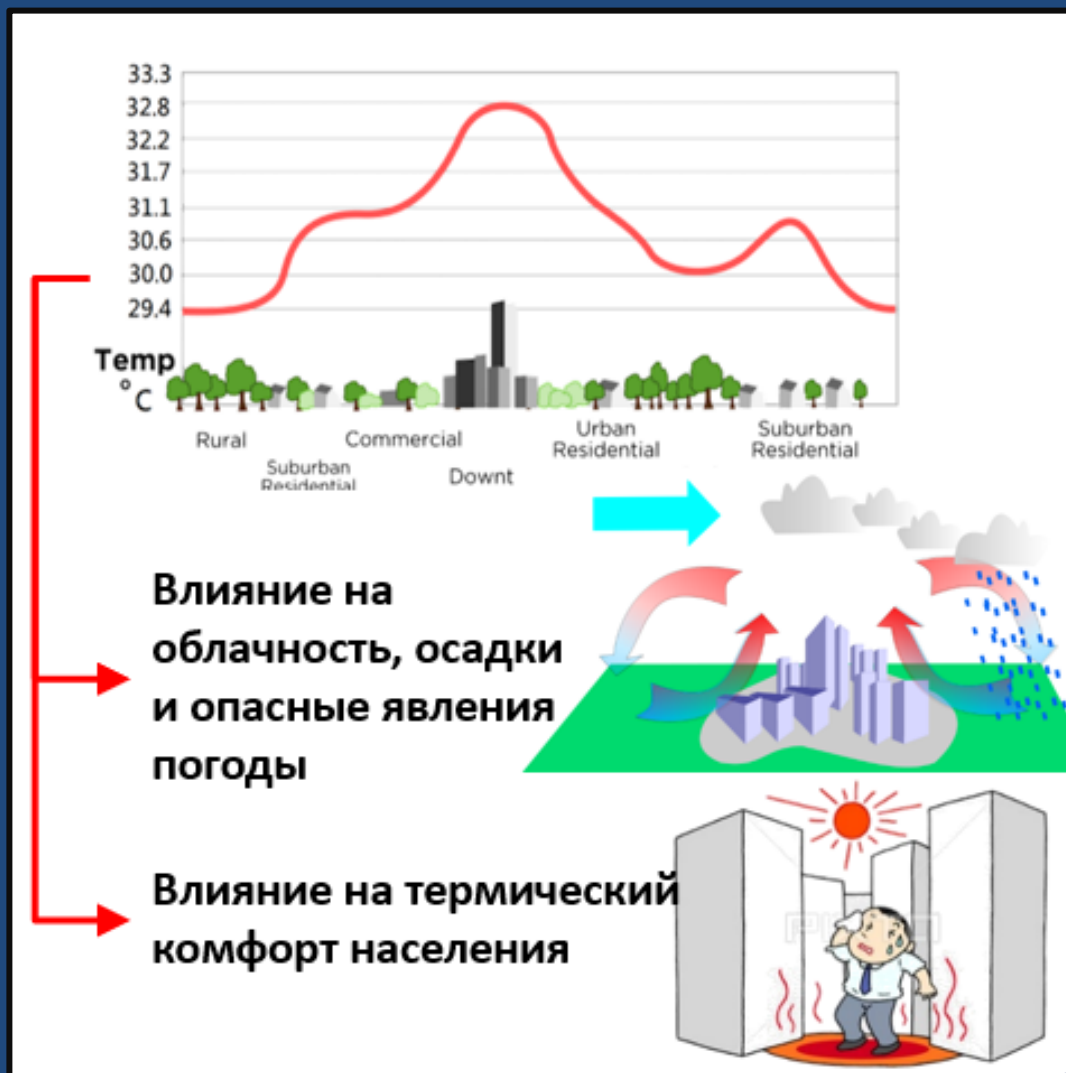
Сравнение методов машинного обучения для учета связей с запаздыванием при моделировании городского острова тепла

Авторы:

Назмутдинов К.Ф., Варенцов М.И., канд. геогр. наук

Почта (kfnazmutdinov@aari.ru)

Актуальность



Цель

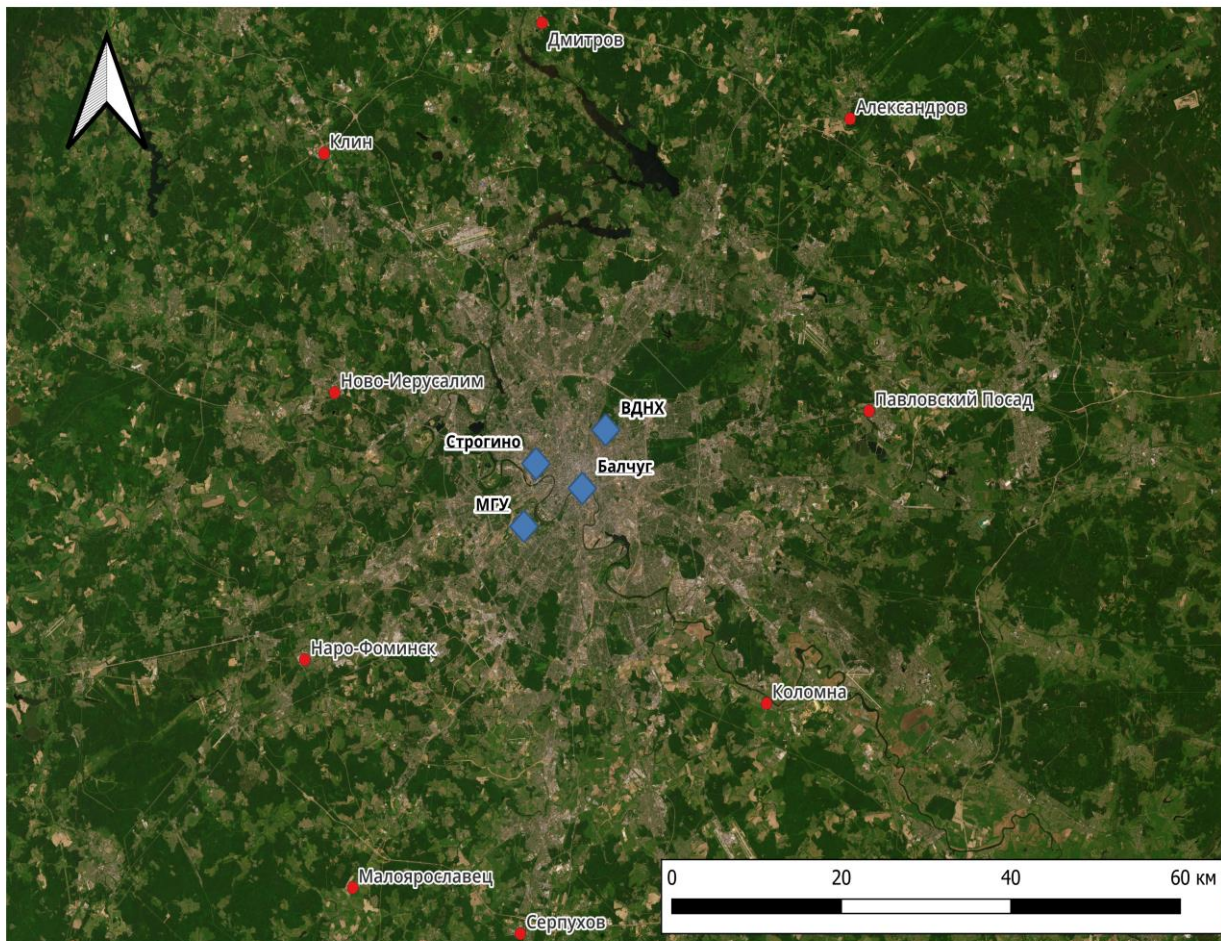
- Определение метода машинного обучения, выполняющего наиболее точную аппроксимацию разницы температур между урбанизированной территорией и сельской местностью,

Задача

- Выбор наиболее производительной конфигурации нейросети LSTM
- Сравнение различных подходов машинного обучения, учитывающих временные зависимости в метеорологических данных

Исследуемая территория

Москва



Станция Санкт-Петербург

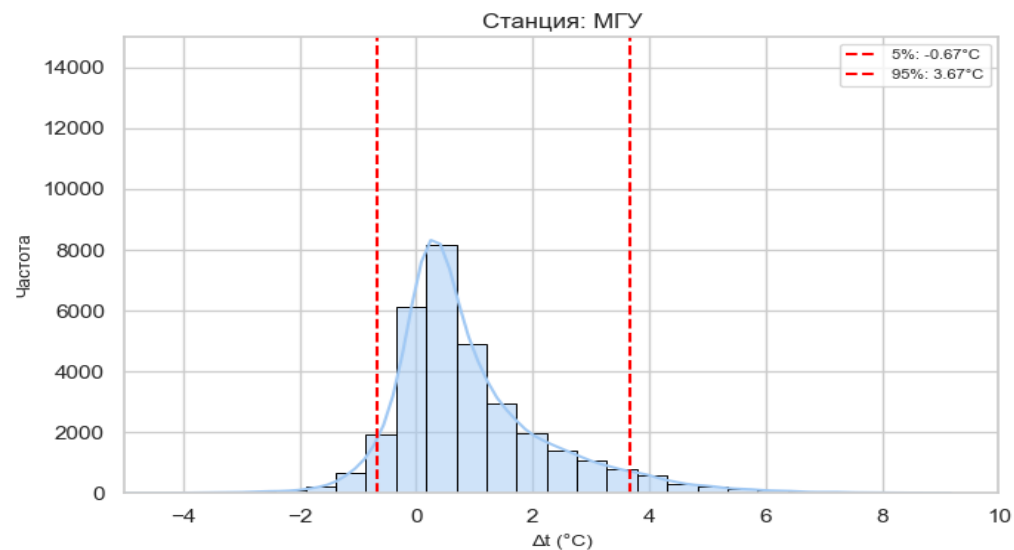
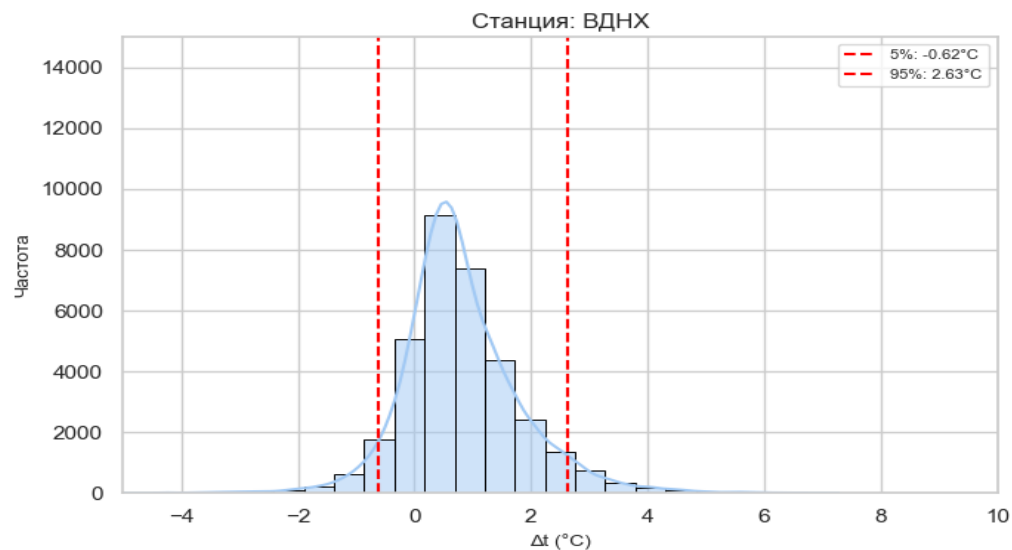
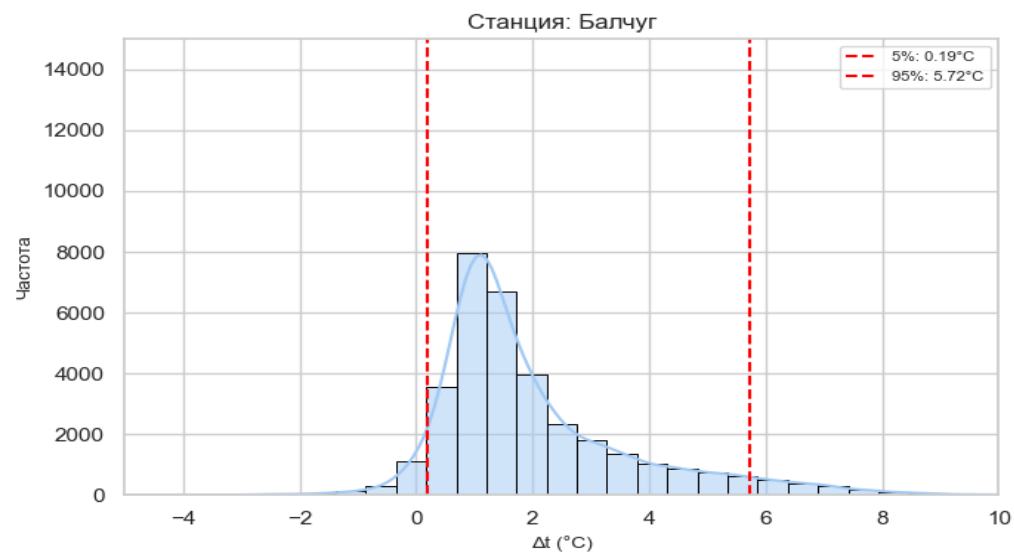
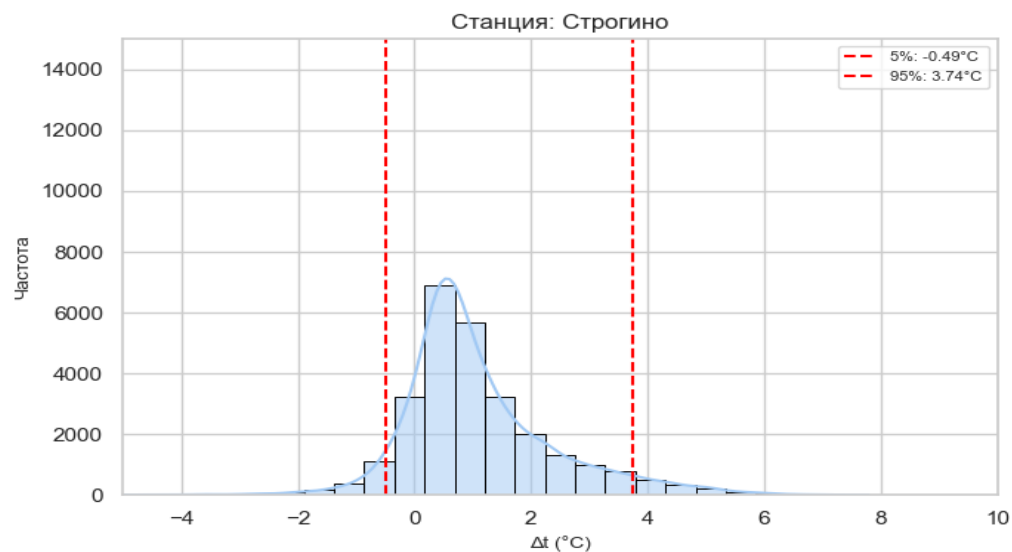


Пространственное положение станций в г. Москва и г. Санкт-Петербург.

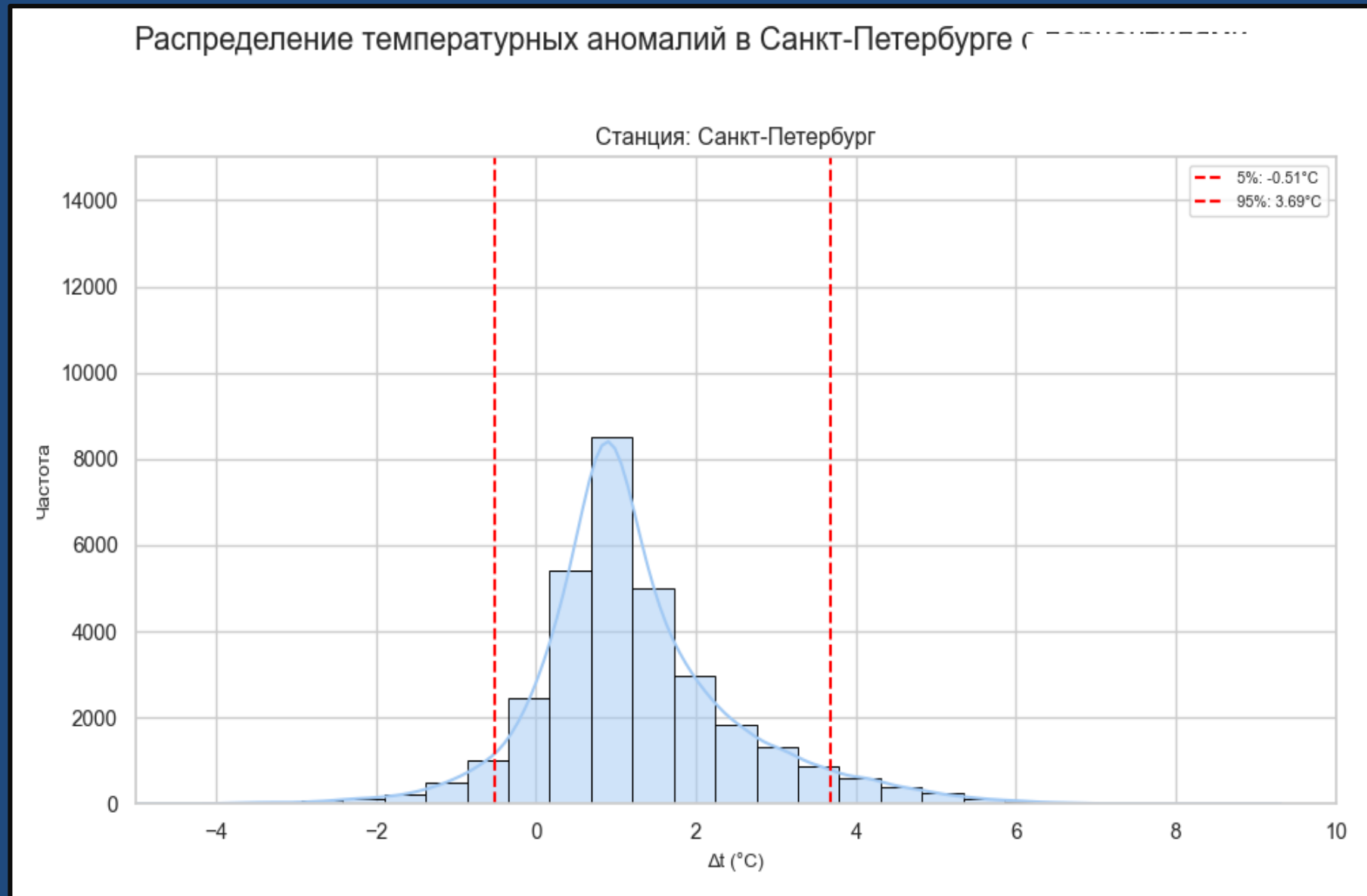
- ◆ - Городская метеостанция
- - Фоновая метеостанция

Анализ данных

Распределение температурных аномалий в Москве



Анализ данных



Кросс-валидация

Blocked Cross-Validation



Train



Test

1 fold



2 fold



3 fold



4 fold



5 fold



Подходы машинного обучения

1. Модель машинного обучения CatBoostRegression (`n_estimators = 2000`) [1]

- Временные тенденции

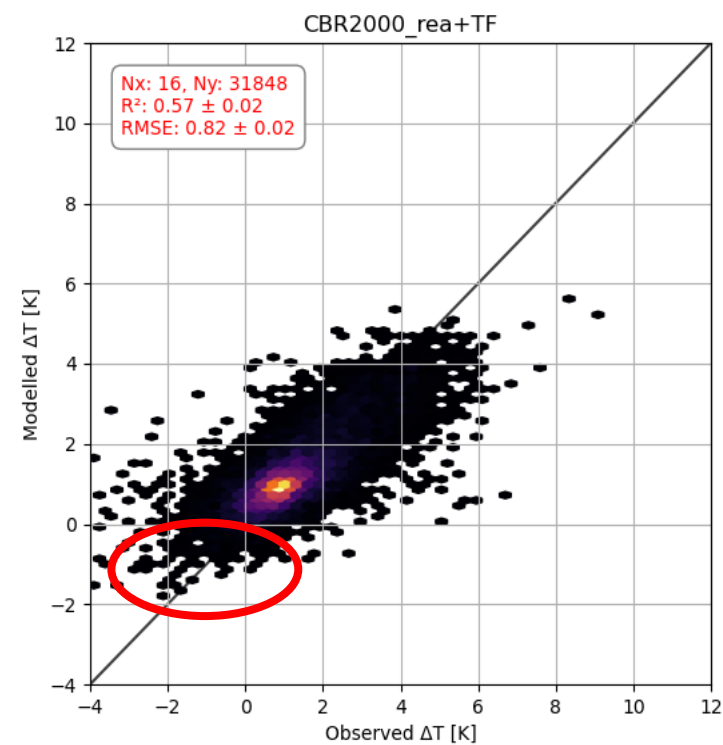
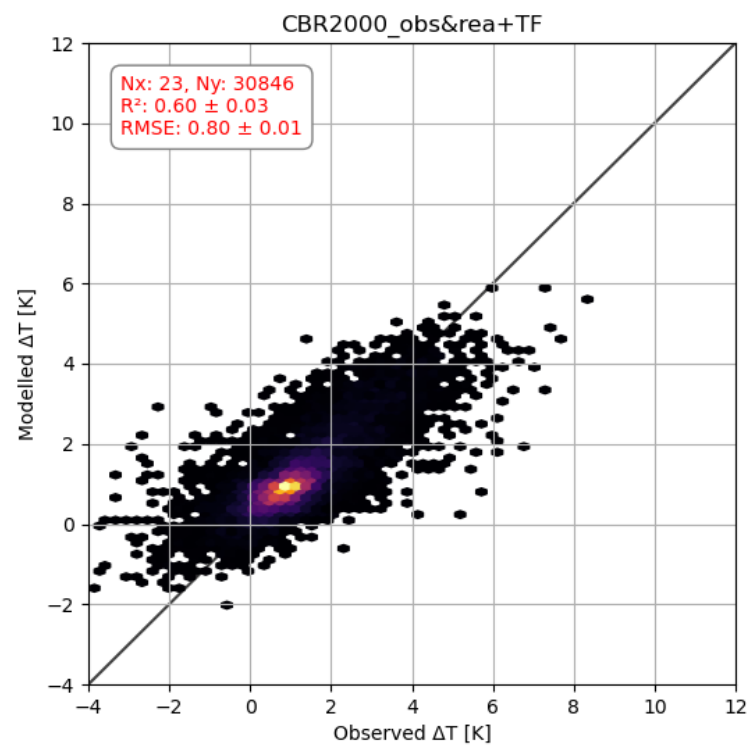
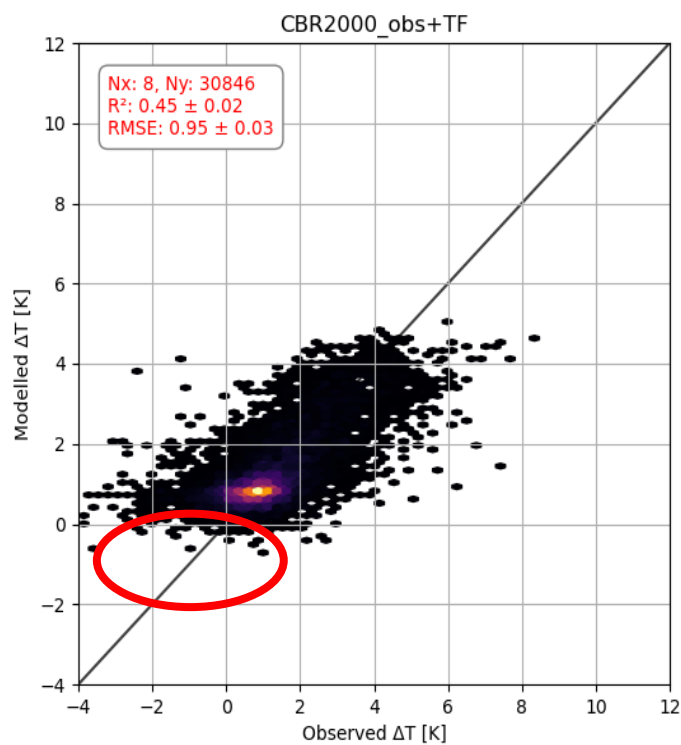
$$\Delta x_{k,3h} = x_{k,i} - x_{k,i-1}; \quad \Delta x_{k,6h} = x_{k,i} - x_{k,i-2}; \quad \Delta x_{k,12h} = x_{k,i} - x_{k,i-4}$$

- Левостороннее скользящее среднее

$$\overline{\Delta x_{k,3h}} = \frac{x_{k,i} + x_{k,i-1}}{2}; \dots; \overline{\Delta x_{k,12h}} = \frac{x_{k,i} + x_{k,i-1} + x_{k,i-2} + x_{k,i-3} + x_{k,i-4}}{2}$$

Подходы машинного обучения

1. Модель CatBoostRegression



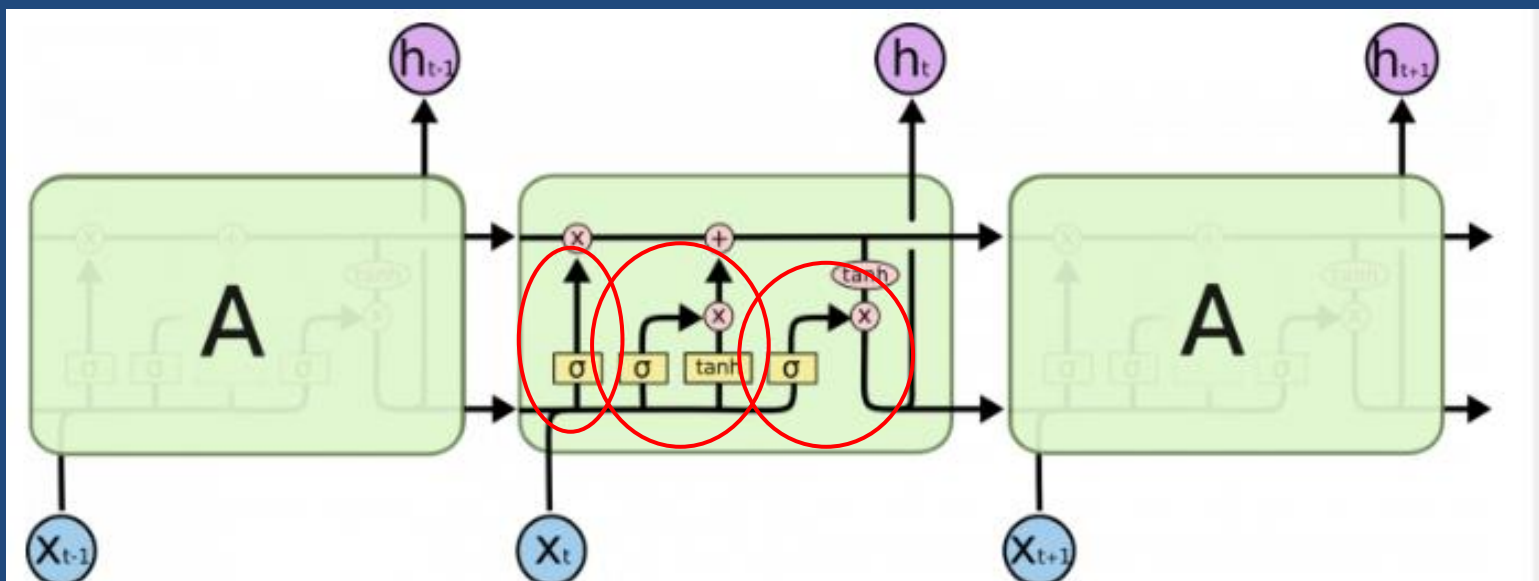
Подходы машинного обучения

2 Рекуррентная нейронная сеть LSTM

Слой фильтра забывания (forget gate layer)

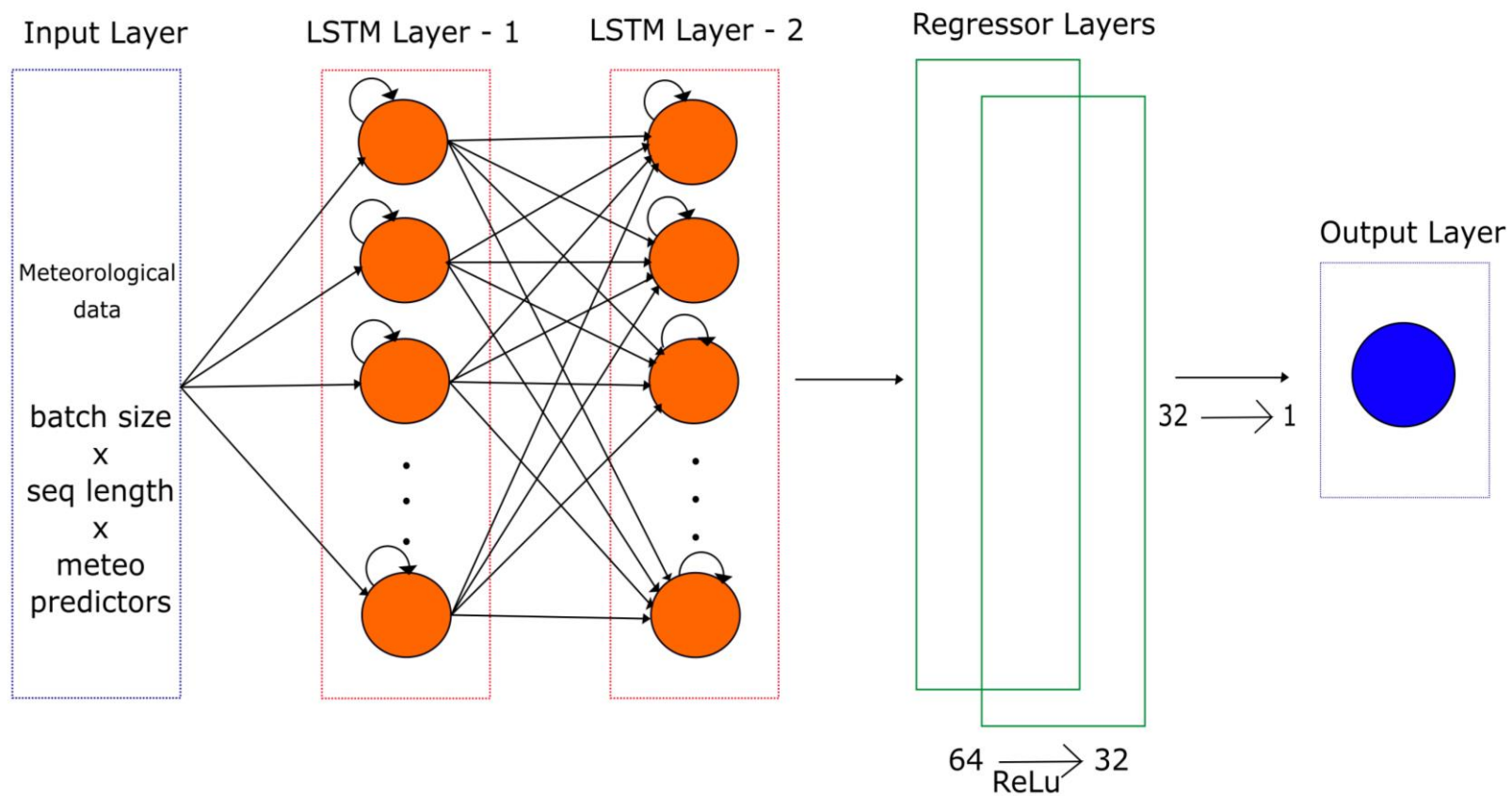
Слой входного фильтра(input layer gate)

Выходной слой (output layer)

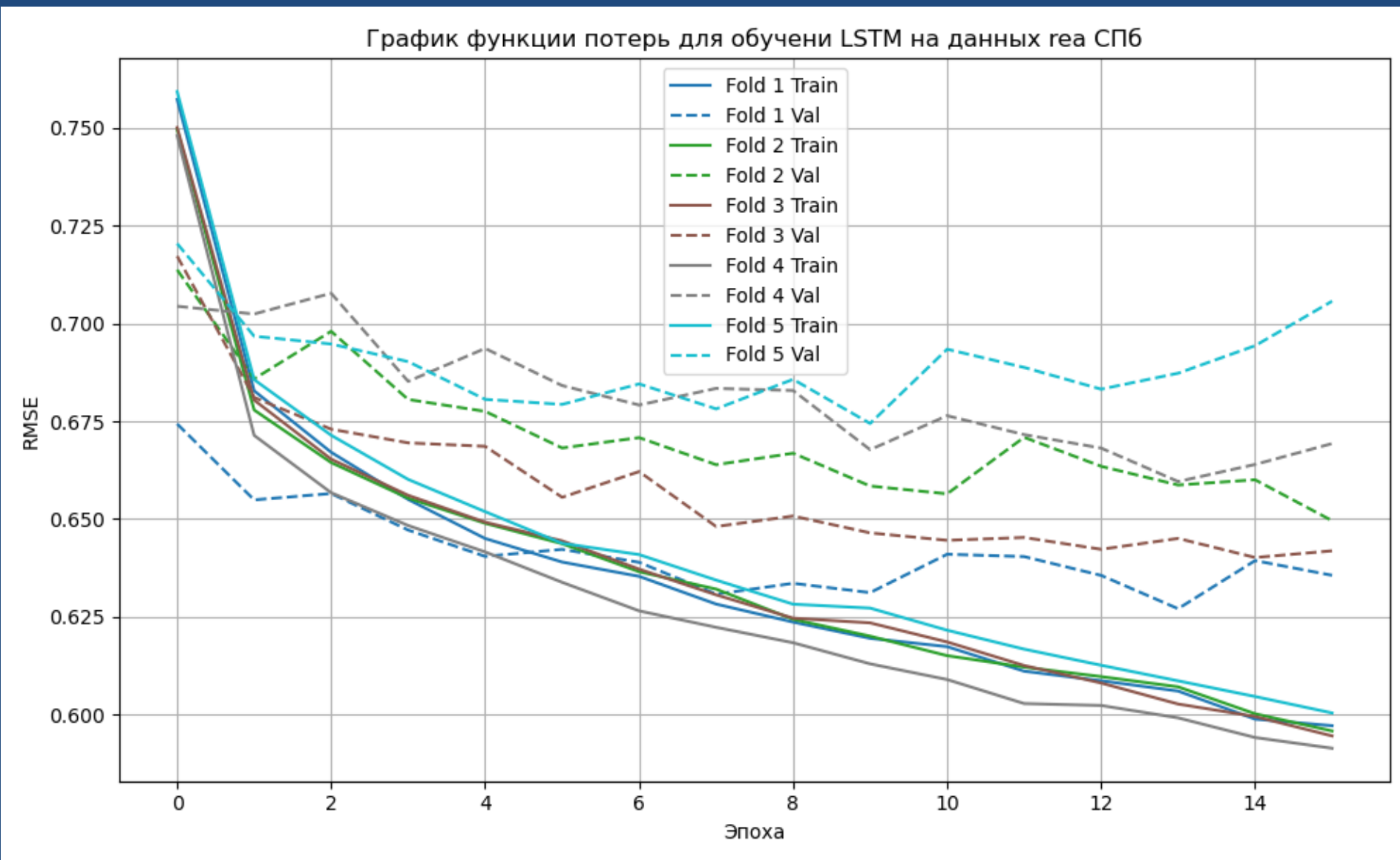


[2] Wunder_editor. LSTM – сети долгой краткосрочной памяти [Электронный ресурс] // Habr. – 2017. – URL: <https://habr.com/ru/companies/wunderfund/articles/331310/> (дата обращения: 27.06.2025).

Архитектура



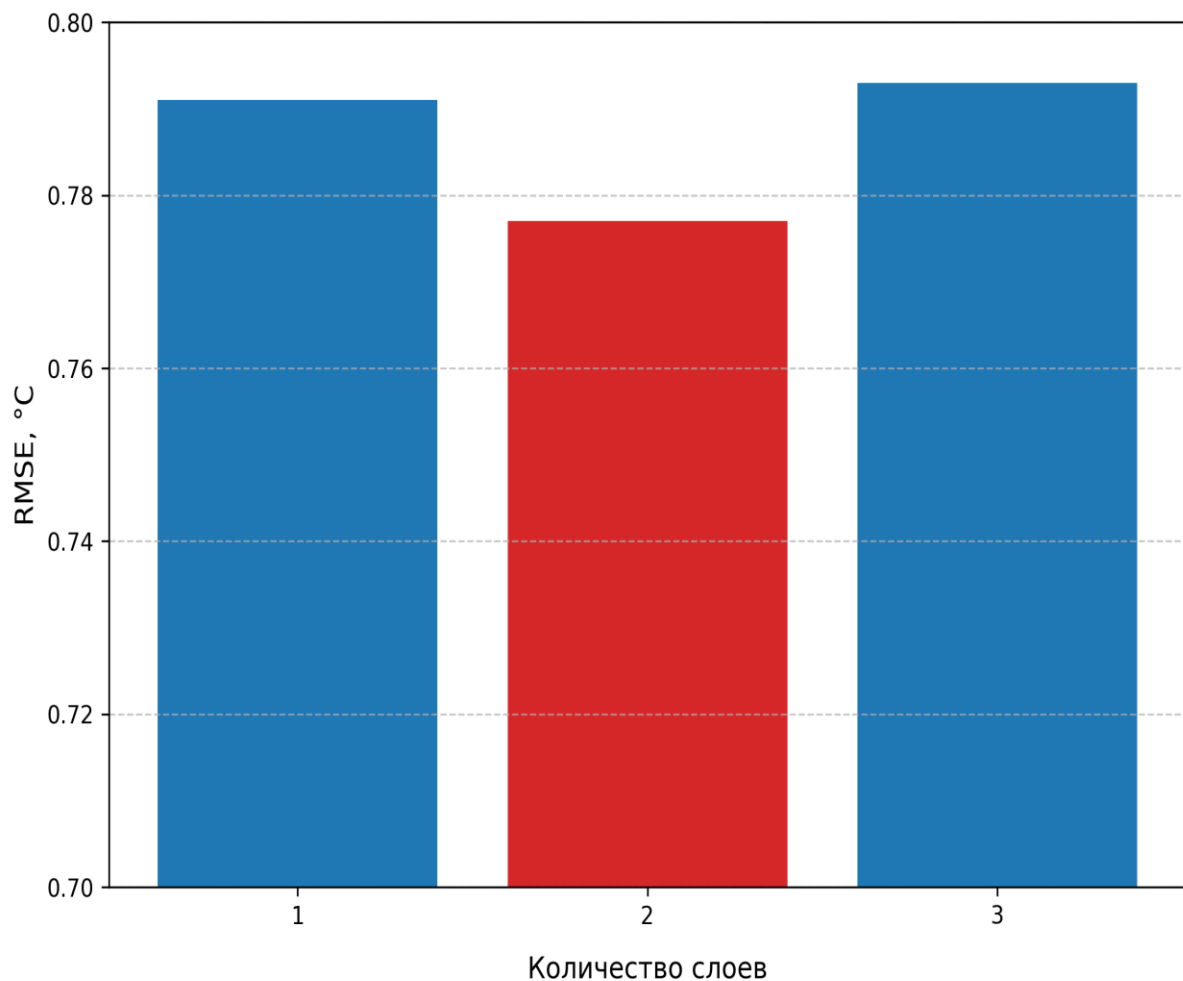
Функция потерь



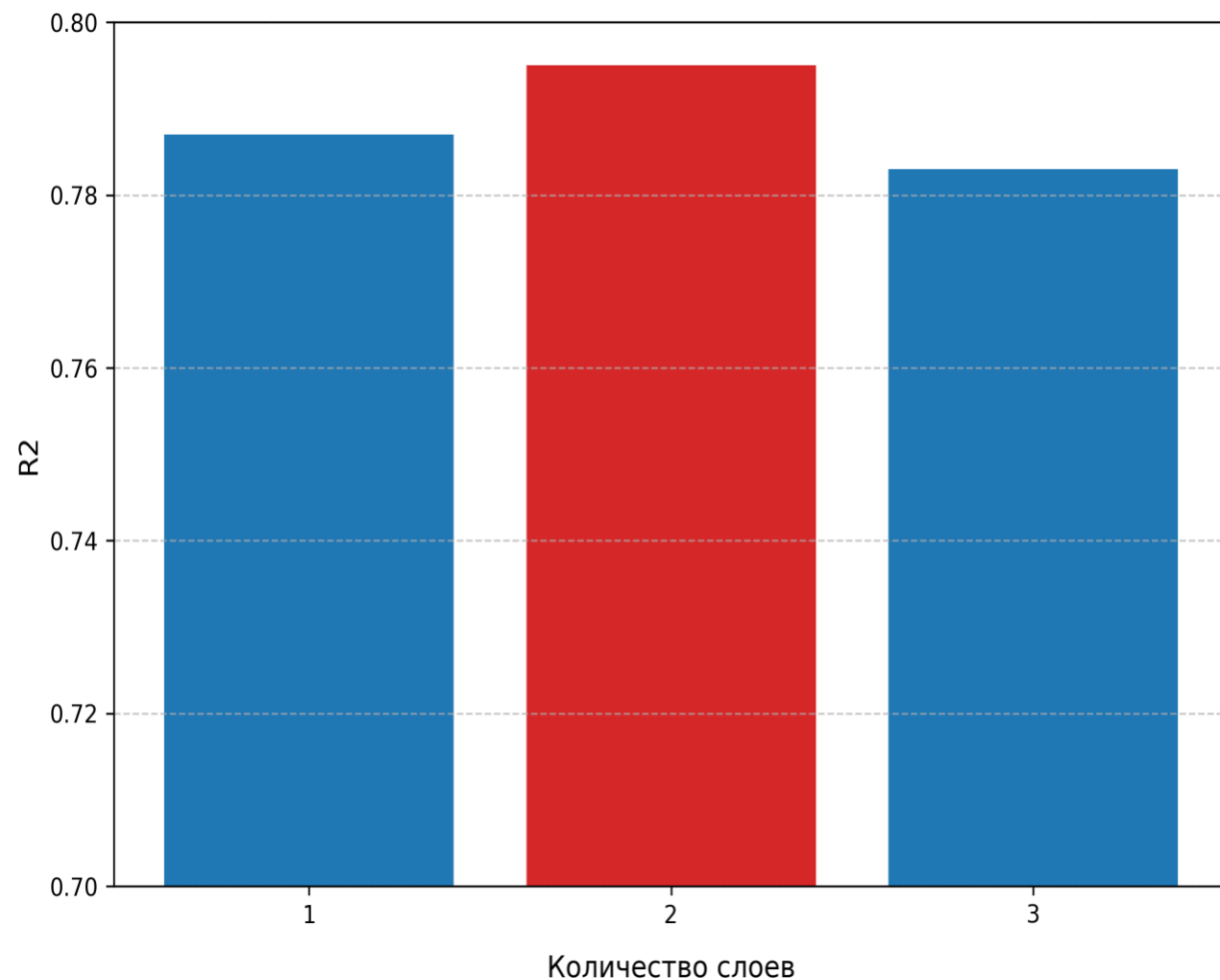
Глубина сети

Москва-Балчуг, данные: rea

Зависимость точности модели от глубины сети

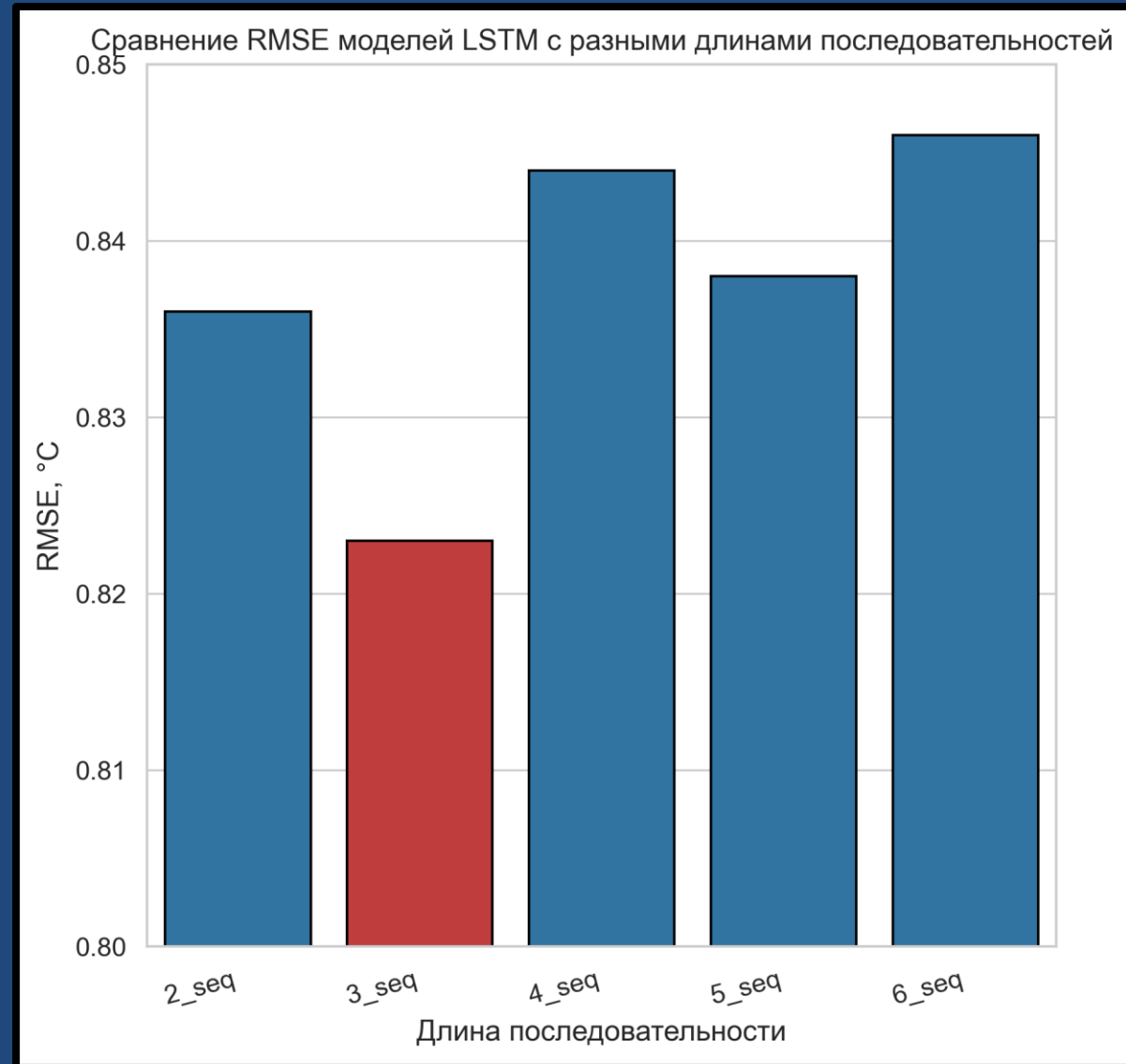
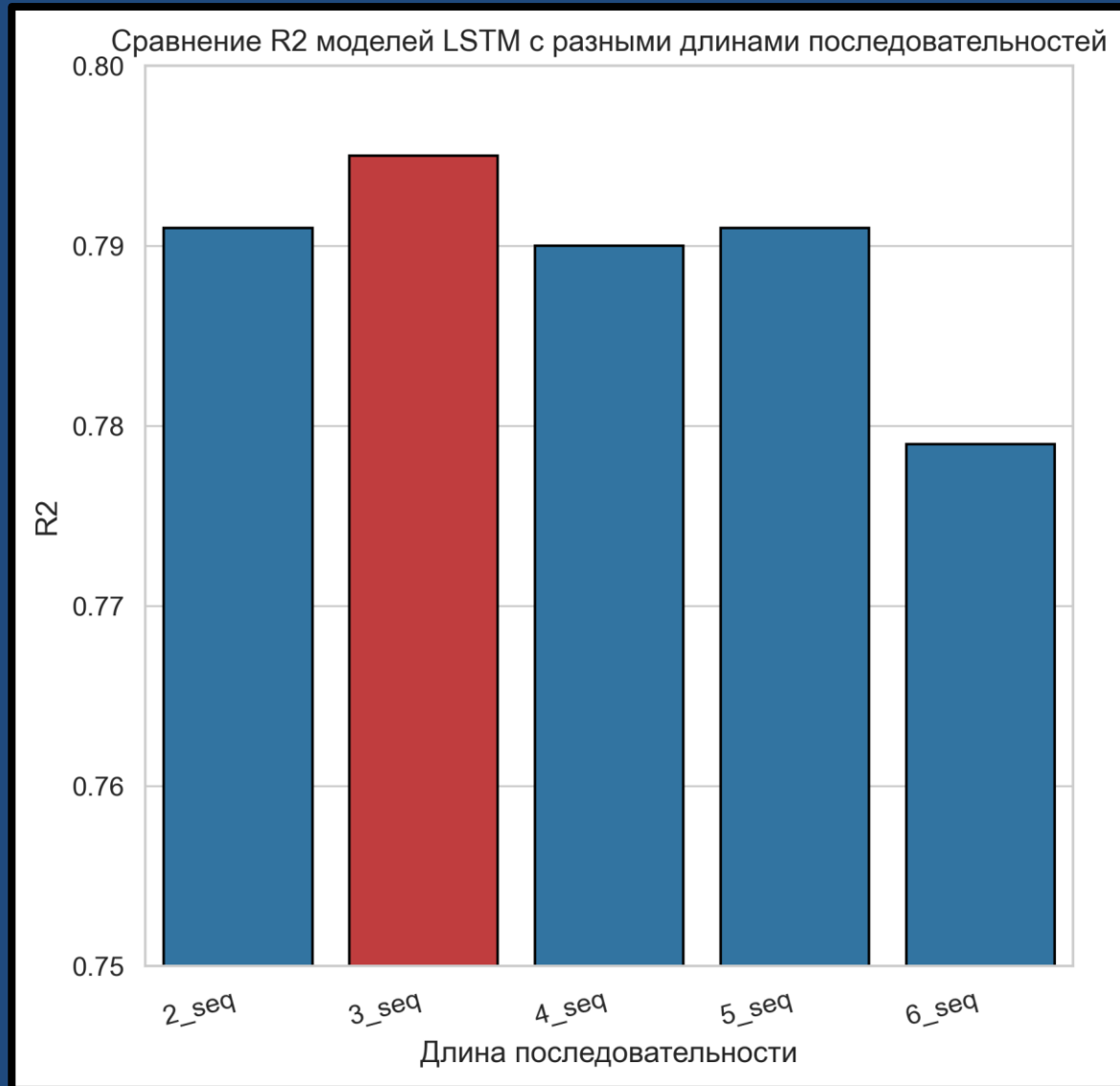


Зависимость точности модели от глубины сети

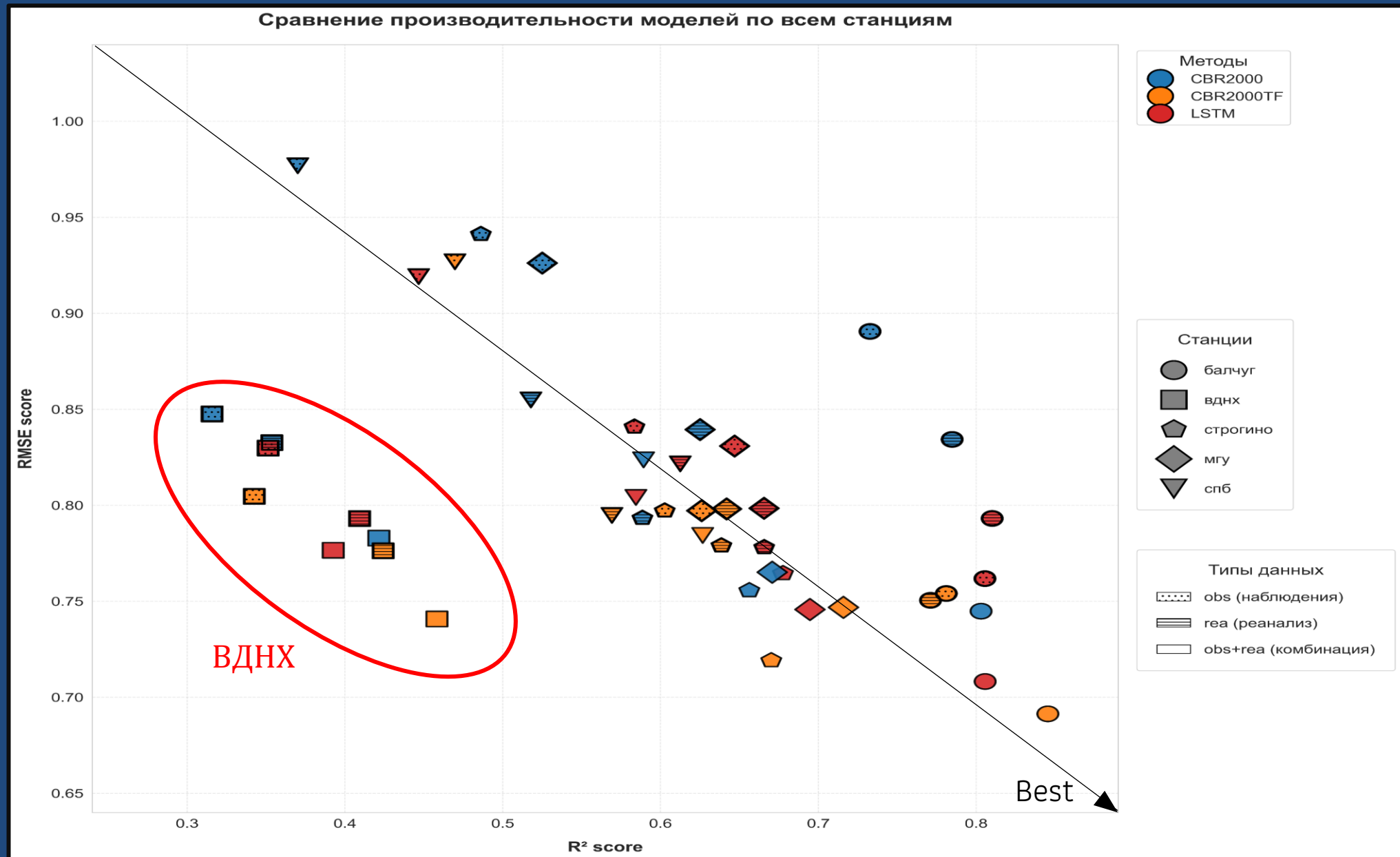


Длина последовательности

Москва-Балчуг, данные: rea

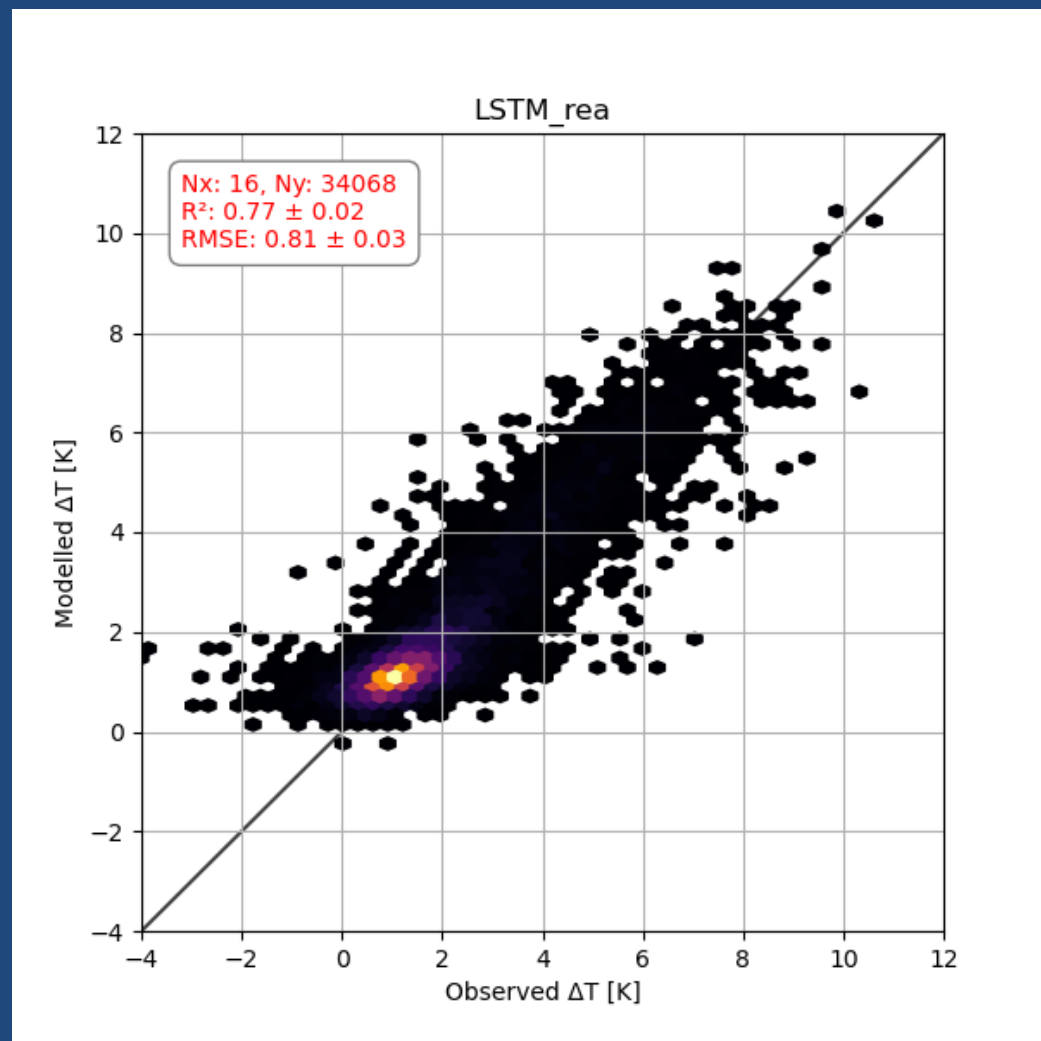


Результаты

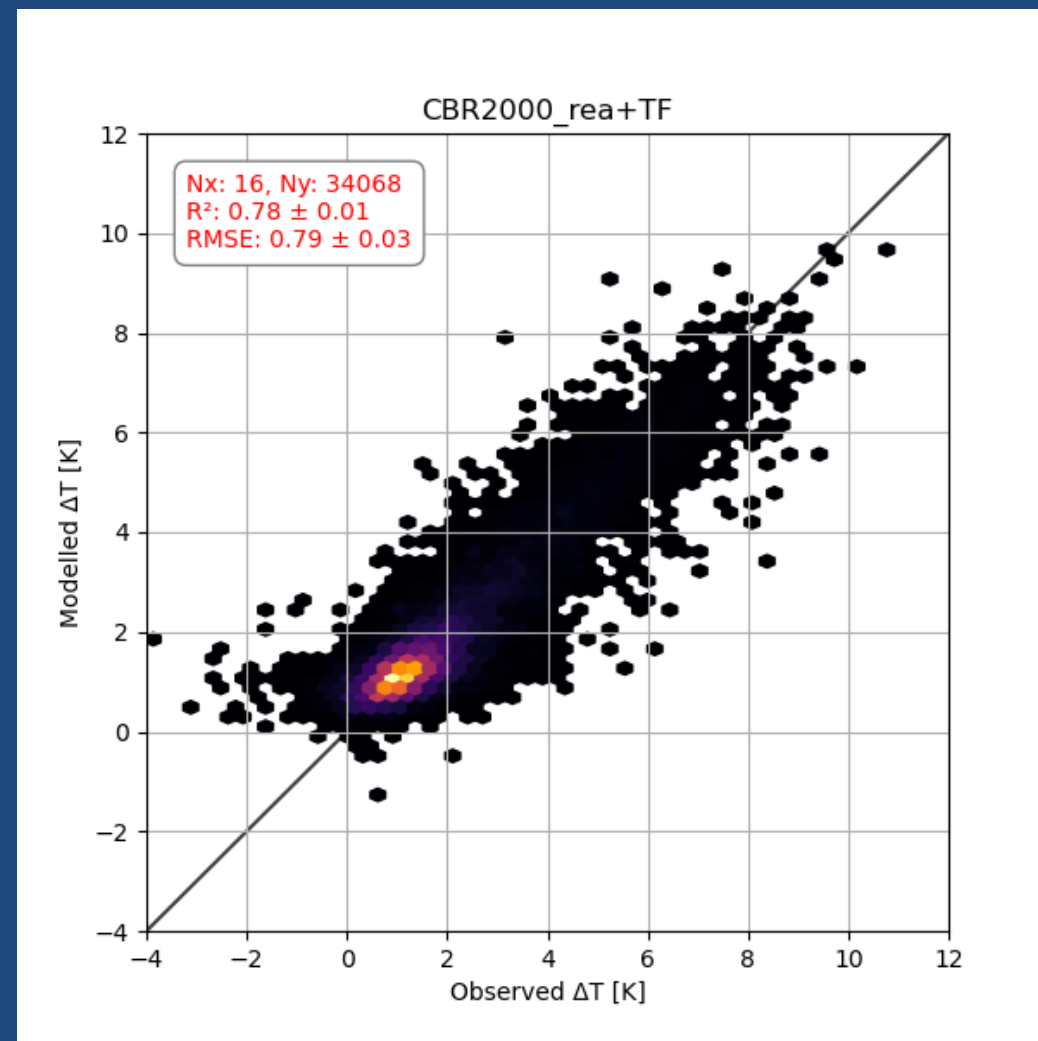


Результаты

Москва-Балчуг, данные: rea

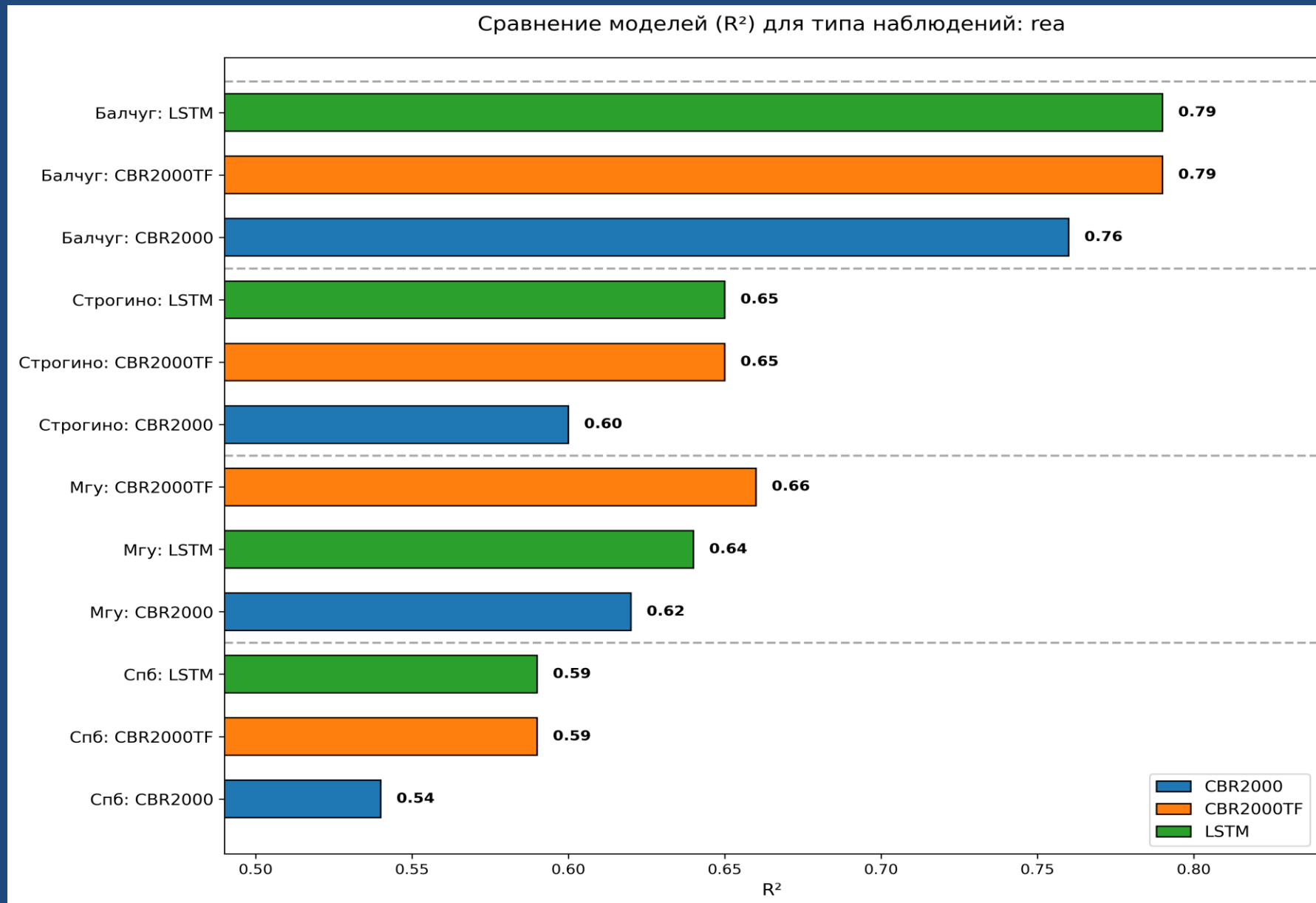


LSTM

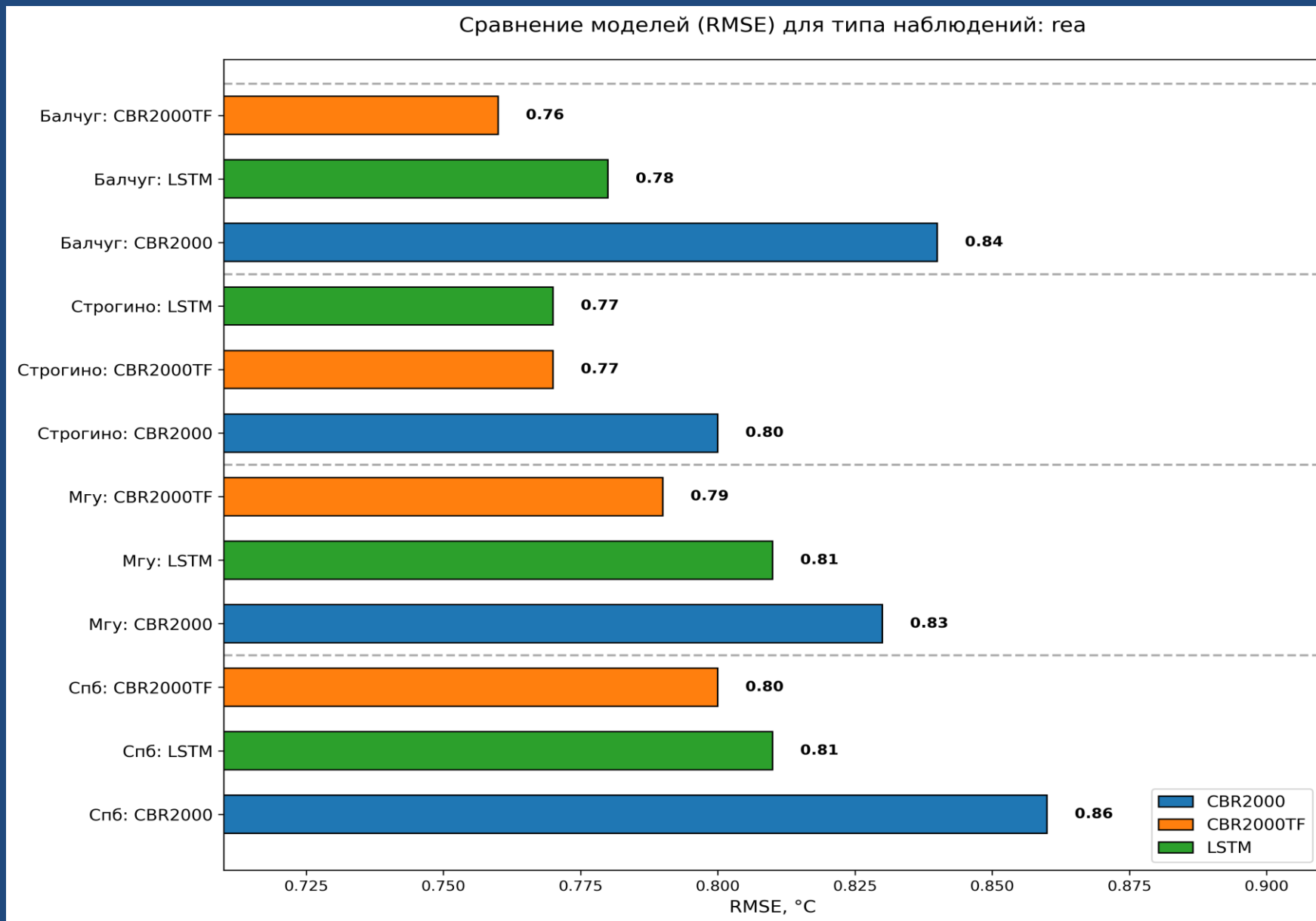


CBR2000+TF

Результаты



Результаты



Заключение

- В результате данного исследования выявлено, что нейронная сеть LSTM показала результаты, немногим отличающиеся от модели CatBoostRegression, включающую временные признаки. Это позволяет делать вывод о возможности ее применения, однако значительных преимуществ выявлено не было.

Дальнейшее развитие

- Проведение экспериментов с нейронной сетью GRU
- Обучение нейросети LSTM на данных временных признаков, порожденных экспертным образом

Поддержка

- Выполнено при поддержке гранта РФФ No23-77-30008

Используемые данные

Обозначения метеорологических функций :

- `t2m` - температура воздуха на высоте 2 м (°C), наблюдения и повторный анализ
- `rh2m` - относительная влажность (%), наблюдение и повторный анализ
- `vel10m` - скорость ветра на высоте 10 м (м/с), наблюдения и реанализ
- `dir10m` - направление ветра на высоте 10 м (градусы), только реанализ
- `u10` - меридиональная составляющая скорости ветра на высоте 10 м (м/с), только реанализ
- `v10` - зональная составляющая ветра на высоте 10 м (м/с), только реанализ
- `psl` - Атмосферное давление, приведенное к внимательному врачу (гПа), только наблюдение.
- `ps` - атмосферное давление на уровне станции (гПа), наблюдение и реанализ
- `tcc` - доля облачности (0-1), наблюдение и реанализ
- `lcc` - доля нижней облачности (0-1), наблюдение и реанализ
- `blh` - высота пограничного слоя атмосферы (м), только реанализ
- `str` - длинноволновый радиальный баланс (Вт/м²), только реанализ
- `ssr` - коротковолновый радиальный баланс (Вт/м²), только реанализ
- `strd` - приходящая длинноволновая радиация (Вт/м²), только реанализ
- `ssrd` - приходящая коротковолновая радиация (Вт/м²), только реанализ