

Влияние выбора функции потерь при обучении модели на качество прогноза уровней воды

Э.Р. Акмаев

ФГБУ «Гидрометеорологический научно-исследовательский центр Российской Федерации»,
Россия, 123376, г. Москва, Большой Предтеченский пер., д. 13

Адрес для переписки: elshatakmaev@yandex.ru

Реферат. Нейросетевые модели краткосрочного прогноза уровня воды традиционно обучаются с использованием среднеквадратичной ошибки (MSE), тогда как критерии оценки – коэффициент Нэша-Сатклиффа (NSE) и критерий оправдываемости S/σ_{Δ} . Несоответствие между функцией потерь при обучении и метрикой оценки приводит к систематически субоптимальным результатам на горизонтах свыше 3-5 суток: с ростом заблаговременности структура ошибки существенно меняется, а вклад дальних горизонтов в суммарные потери непропорционально возрастает и равномерное взвешивание шагов прогноза перестаёт быть оптимальным.

В настоящей работе предложены три альтернативные функции потерь, каждая из которых нацелена на отдельный аспект структуры многошаговой ошибки. NSE+deriv (функция потерь на основе NSE с дополнительным штрафом по производным гидрографа) нормирует ошибку на стандартное отклонение целевого ряда и дополнительно штрафует несовпадение первой и второй производных прогнозного гидрографа, отражающих скорость подъёма уровня. ITAE (интегральная временно-взвешенная абсолютная ошибка) линейно усиливает штраф с горизонтом, смещая градиентный сигнал к дальним шагам прогноза. UW-композиция (взвешивание по неопределённости) автоматически балансирует оба компонента через обучаемые параметры неопределённости. Эксперименты проведены на пяти водосборах с различными типами водного режима в режиме совместного обучения единой модели, с использованием архитектур TiDE (нейросеть с плотными слоями для временных рядов) и TSMixer (Time-Series Mixer – архитектура на основе MLP с попеременным смешиванием по времени и по признакам).

Лучшая конфигурация – UW(NSE+deriv, ITAE) – улучшает NSE на горизонте 10 суток на 10.4% и снижает S/σ_{Δ} на 6.3% относительно MSE-базовой линии при сохранении качества на ближних горизонтах. Прирост NSE составляет 2-10% в среднем по водосборам на горизонтах от 3 до 10 суток. Дополнительно показано, что выигрыш функции потерь UW зависит от архитектуры модели: на TiDE он проявляется на всём диапазоне горизонтов, тогда как на TSMixer – только начиная с горизонта 5 суток. Выбор функции потерь при обучении представляет самостоятельный инструмент повышения качества прогноза, не требующий изменения архитектуры модели.

Ключевые слова. Прогноз уровня воды, функции потерь, краткосрочный прогноз, нейронные сети, TiDE, NSE, прогнозирование временных рядов, проектирование функции потерь, обучение с учётом метрики оценки.

The effect of loss function choice on model training quality in water level forecasts

E.R. Akmaev

Hydrometeorological Research Center of Russian Federation,
13, Bolshoy Predtechensky per., 123376, Moscow, Russia

Correspondence address: *elshatakmaev@yandex.ru*

Abstract. Neural network models for short-term river stage forecasting are typically trained with mean squared error (MSE), while the evaluation criteria are the Nash-Sutcliffe efficiency (NSE) and the S/σ_Δ skill score. This mismatch between training loss and evaluation metric leads to systematically suboptimal results at forecast horizons beyond 3-5 days: as lead time increases, the error structure changes substantially, the contribution of distant horizons to total loss grows disproportionately, and uniform weighting across forecast steps becomes suboptimal.

Three alternative loss functions are proposed, each addressing a distinct aspect of multi-step error structure. NSE+deriv (an NSE-based loss with an additional penalty on hydrograph derivatives) normalises the error by the standard deviation of the target series and additionally penalises mismatches in the first and second derivatives of the forecast hydrograph, which reflect the rate of water level rise. ITAE (Integral Time-weighted Absolute Error) linearly amplifies the penalty with horizon, shifting the gradient signal toward distant forecast steps. The UW-composition (uncertainty weighting) automatically balances both components via learned uncertainty weights. Experiments are conducted on five basins with contrasting hydrological regimes under a joint training regime, using both the TiDE (Time-series Dense Encoder) and TSMixer (Time-Series Mixer – an architecture based on MLPs with alternating time- and feature-mixing) architectures.

The best configuration – UW(NSE+deriv, ITAE) – improves NSE at a 10-day horizon by 10.4% and reduces S/σ_Δ by 6.3% relative to the MSE baseline, while maintaining near-baseline performance at short horizons. NSE gains range from 2 to 10% on average across basins at horizons of 3 to 10 days. Additionally, the gain from the UW loss is shown to depend on model architecture: on TiDE it holds across the full horizon range, whereas on TSMixer it emerges only from a 5-day horizon onward. Loss function selection thus constitutes an independent lever for forecast quality improvement, requiring no architectural modifications.

Keywords. River stage forecasting, loss functions, short-term forecasting, neural networks, TiDE, NSE, time series forecasting, loss function design, metric-aware training.

Введение

Среднеквадратичная ошибка используется как функция потерь при обучении нейросетевых моделей прогноза уровня воды по умолчанию, без специального обоснования этого выбора. Критерии, по которым прогноз оценивается на практике, устроены иначе: коэффициент Нэша-Сатклиффа (NSE, Nash-Sutcliffe Efficiency; Нэш, Сатклифф, 1970) и критерий оправдываемости S/σ_Δ (Бефани, Калинин, 1983) чувствительны к форме и динамике ряда, не только к точечной величине отклонения. Это означает, что цель оптимизации при обучении не совпадает с тем, по чему прогноз затем оценивается, что может ухудшать качество прогноза на средних и дальних горизонтах, где вклад формы ошибки в итоговую оценку растёт.

Для задачи многошагового прогноза проблема усугубляется тем, что структура ошибки существенно меняется с горизонтом: на коротких заблаговременностях равномерное взвешивание шагов прогноза оправдано, однако с ростом горизонта ошибки дальних шагов систематически крупнее и вносят непропорционально большой вклад в функцию потерь, смещая обучение в их сторону. Klotz и др. (Klotz et al., 2024) показали, что NSE не декомпозируется аддитивно по разбиениям данных – свойство, из которого следует систематическое расхождение между обучающей и валидационной оценками качества. Это объясняет, почему модели, обученные на MSE, дают субоптимальный NSE при оценке, даже если формально MSE и NSE имеют одинаковый оптимум для одного водосбора.

Ряд работ направлен на решение проблемы несоответствия потери и метрики в задачах прогноза стока. Kratzert и др. (Kratzert et al., 2019) предложили NSE как функцию потерь при обучении нейросетей на многих водосборах одновременно: нормировка ошибки на стандартное отклонение целевого ряда каждого водосбора обеспечивает равный вклад водосборов разного масштаба в суммарный градиентный сигнал. Tofighi и др. (Tofighi et al., 2025) предложили асимметричную функцию потерь для прогноза расходов на 12-часовом горизонте. Wang и др. (Wang et al., 2025) совместили классификацию паводочного режима с модифицированной функцией потерь для прогноза уровней воды на горизонтах до 24 часов. Однако ни одна из этих работ не рассматривает задачу многошагового прогноза с явным учётом того, что оптимальный критерий меняется с горизонтом, и не использует автоматического балансирования компонент потерь.

Hounie и др. (Hounie et al., 2024) зафиксировали, что оптимизация средней ошибки по прогнозному окну приводит к неравномерному распределению ошибок по шагам, особенно для архитектур класса трансформер (Transformer), что подтверждает субоптимальность равномерно взвешенной функции потерь на всём диапазоне горизонтов.

Предложены три функции потерь, каждая из которых нацелена на отдельный аспект структуры многошаговой ошибки: NSE+deriv нормирует ошибку по изменчивости ряда и штрафует отклонения в скорости изменения уровня, ITAE смещает градиентный сигнал к дальним горизонтам, UW-

композиция автоматически балансирует оба компонента через взвешивание по неопределённости (Kendall et al., 2018). На пяти разнородных водосборах показано, что все три конфигурации превосходят MSE на горизонтах $k \geq 3$, при этом UW устойчиво превосходит каждую из них в отдельности на всём диапазоне горизонтов. Применение ИТАЕ в роли функции потерь при обучении нейросетей для прогноза временных рядов ранее не фиксировалось ни в одном из доступных систематических обзоров (Jadon et al., 2024; Kim et al., 2025). В качестве метрики оценки используется критерий оправдываемости S/σ_{Δ} (Бефани, Калинин, 1983).

Методы и материалы

Данные

Исследование проводится на пяти водосборах с различными гидрологическими режимами. Выборка охватывает водосборы, расположенные в разных физико-географических зонах и относящиеся к различным типам водного режима: р. Большая Воровская (в. 2177) – смешанное питание с преобладанием подземного стока, дальневосточный тип; р. Малиновка (в. 5287) – снегово-дождевое питание, половодье в тёплую часть года; р. Сылта (в. 76692) – весеннее половодье, восточно-европейский тип; р. Белая (в. 83348) и р. Малка (в. 84192) – снегово-ледниковое питание, северо-кавказский тип. Целевая переменная – суточный уровень воды (см). Подробная физико-географическая характеристика водосборов приведена в работе (Романов и др., 2026).

Основные характеристики водосборов приведены в табл. 1; их расположение показано на рис. 1. Водосборы охватывают широкий диапазон площадей (1540-4730 км²) и различаются по характеру внутригодового распределения стока, что обеспечивает разнородность обучающей выборки.

Таблица 1. Характеристики водосборов

Table 1. Basin characteristics

Река, пункт (индекс)	Координаты, °	Нуль графика, м	Площадь водосбора, км²	Уклон реки, ‰
р. Большая Воровская, с. Соболево (2177)	54.18 с.ш., 155.57 в.д.	7.96	3630	8.5
р. Малиновка, с.Ракитное (5287)	45.38 с.ш., 134.16 в.д.	92.40	4730	4.4
р. Сылта, пгт. Шамары (76692)	57.21 с.ш., 58.13 в.д.	208.98	3130	1.1
р. Белая, пгт. Каменноостский (83348)	44.17 с.ш., 40.11 в.д.	373.09	1850	22
р. Малка, с. Каменноостское (84192)	43.44 с.ш., 43.04 в.д.	792.72	1540	32

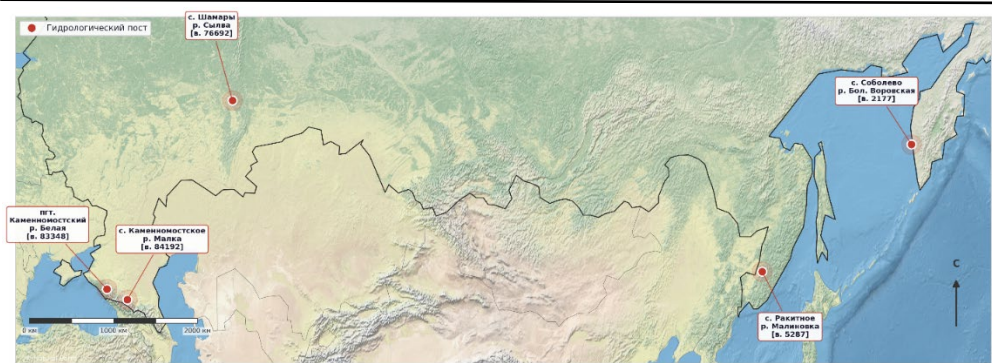


Рисунок 1. Карта расположения водпостов
Figure 1. Map of gauging station locations

Обучающий период охватывает 1980-2015 гг., валидационный – 2016-2019 гг. Разбиение зафиксировано для всех экспериментов. Стандартизация – по среднему и стандартному отклонению обучающего периода, независимо для каждой станции.

Архитектура и обучение

В качестве основной архитектуры используется TiDE (Time-series Dense Encoder – энкодер временных рядов на основе полносвязных слоёв (Das et al., 2024)) с применением обратимой нормализации экземпляра (RevIN, Kim et al., 2022). Для оценки зависимости наблюдаемых эффектов от архитектуры дополнительно протестирована архитектура TSMixer (Time-Series Mixer – архитектура на основе MLP с попеременным смешиванием по времени и по признакам, Chen et al., 2023; далее – TSMixer); гиперпараметры обеих архитектур приведены в табл. 2.

Таблица 2. Гиперпараметры моделей. Планировщик скорости обучения ReduceLROnPlateau снижает скорость обучения в factor раз при отсутствии улучшения функции потерь на валидационной выборке в течение patience эпох. Ранняя остановка прерывает обучение, если относительное улучшение функции потерь на валидационной выборке не превышает 0.01% в течение указанного числа эпох

Table 2. Model hyperparameters. The ReduceLROnPlateau learning rate scheduler reduces the learning rate by a factor of factor when no improvement in the validation loss is observed for patience epochs. Early stopping terminates training if the relative improvement in validation loss does not exceed 0.01% for the specified number of epochs

Параметр	TiDE	TSMixer
<i>Архитектура</i>		
Размерность скрытого слоя	384	384
Размерность промежуточного слоя (ff)	—	384
Число слоёв энкодера	3	—
Число слоёв декодера	3	—
Число блоков смешивания	—	6

Продолжение таблицы 2

Параметр	TiDE	TSMixer
Функция активации	ReLU	ReLU
Выходная размерность декодера	64	—
Скрытый размер временного декодера	256	—
Dropout	0.04	0.05
Обратимая нормализация экземпляра (RevIN)	да	да
Горизонт прогноза, сут	10	10
Длина входного окна, сут	45	45
Обучение		
Размер батча	512	512
Начальная скорость обучения	$3 \cdot 10^{-4}$	$3 \cdot 10^{-4}$
Оптимизатор	Adam	Adam
L2-регуляризация (weight decay)	$5 \cdot 10^{-6}$	$5 \cdot 10^{-6}$
Ограничение нормы градиента	3.0	3.0
Максимальное число эпох	50	50
Планировщик скорости обучения	ReduceLROnPlateau, patience=5, factor=0.2	ReduceLROnPlateau, patience=5, factor=0.2
Ранняя остановка, patience	20 эпох	20 эпох
Случайных инициализаций	3 (42, 123, 777)	3 (42, 123, 777)

В качестве входных ковариат модели используются метеорологические переменные: температура и влажность воздуха, осадки, высота снежного покрова, температура почвы и скорость ветра. Для отражения внутригодовой сезонности каждая дата кодируется четырьмя циклическими признаками – синусом и косинусом дня и месяца наблюдения, что обеспечивает непрерывность при переходе через конец года без введения порядковых переменных.

Результаты представлены как среднее $\pm$ стандартное отклонение по трём инициализациям.

Метрики оценки

Коэффициент Нэша-Сатклиффа (k)

$$NSE(k) = 1 - \frac{\sum_{j=1}^N (h_{j+k} - \hat{h}_{j+k})^2}{\sum_{j=1}^N (h_{j+k} - \bar{h})^2},$$

где h_{j+k} – наблюдаемый уровень воды на шаге $j + k$, $\hat{h}_{j+k}$ – прогнозируемый уровень, $\bar{h}$ – среднее наблюдаемого уровня по валидационному периоду, N – число прогнозных окон (Нэш, Сатклифф, 1970). $NSE = 0$ соответствует прогнозу средним значением; $NSE > 0$ – превосходству над климатологической нормой.

Критерий оправдываемости $S/\sigma_{\Delta}(k)$

$$\sigma_{\Delta}(k) = \sqrt{\frac{1}{N-1} \sum_{j=1}^N ((h_{j+k} - h_j) - \bar{\Delta}(k))^2},$$

$$\frac{S(k)}{\sigma_{\Delta}(k)} = \frac{\sqrt{\frac{1}{N} \sum_{j=1}^N (h_{j+k} - \hat{h}_{j+k})^2}}{\sigma_{\Delta}(k)},$$

где h_j – уровень в момент выпуска прогноза (последний наблюденный), $\sigma_{\Delta}(k)$ – среднеквадратическое отклонение k -суточных изменений уровня, $\bar{\Delta}(k)$ – среднее k -суточное изменение. Критерий оправдываемости (Бефани, Калинин, 1983) оценивает точность прогноза относительно изменчивости самого ряда. Значение $S/\sigma_{\Delta} < 1.0$ соответствует превосходству над инерционным прогнозом.

Метрики вычисляются для каждой станции и каждой заглаговременности $k \in \{1, 3, 5, 7, 10\}$ суток, затем усредняются по станциям.

**Функции потерь для краткосрочного прогноза уровня воды:
от структуры ошибки к композитному критерию**

Структура ошибки и несоответствие метрике при многошаговом прогнозе

Стандартная функция потерь при обучении нейросетевых моделей прогноза – среднеквадратичная ошибка:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{B \cdot T_{\text{out}}} \sum_{b=1}^B \sum_{k=1}^{T_{\text{out}}} (\hat{h}_{b,k} - h_{b,k})^2,$$

где B – размер батча (подмножество обучающей выборки, обрабатываемое за один шаг оптимизации), T_{out} – горизонт прогноза, $\hat{h}_{b,k}$ и $h_{b,k}$ – прогнозируемый и наблюденный уровни воды для примера b на шаге k .

MSE является оптимальной при условии, что ошибки прогноза независимы и одинаково распределены по горизонтам (i.i.d., от англ. independent and identically distributed), а функция потерь при обучении совпадает с метрикой оценки. В задаче многошагового прогноза уровня воды оба условия, как правило, не выполняются.

На коротком горизонте структура ошибки близка к однородной и все функции потерь дают схожий результат; начиная с $k \approx 3 - 5$ дисперсия ошибки возрастает настолько, что равномерное взвешивание горизонтов перестаёт быть оптимальным. Sun и др. (Sun et al., 2021) предложили явно учитывать автокорреляцию остатков первого порядка (AR(1)) в функции потерь – коэффициент ρ обучается совместно с параметрами модели и устраняет автокорреляцию в остатках прогноза.

Второй источник проблемы – несоответствие функции потерь и метрики. MSE минимизируется при прогнозе, равном условному среднему

наблюдаемого уровня при заданных входных данных, тогда как NSE нормирует ошибку на дисперсию ряда, а S/σ_{Δ} – на дисперсию k -суточных приращений. Klotz и др. (Klotz et al., 2024) показали, что NSE не декомпозируется аддитивно по разбиениям данных (свойство DAMN – Divide And Measure Nonconformity): различные подвыборки имеют различные опорные дисперсии, поэтому NSE подвыборки не равен NSE по полному ряду. Из этого свойства следует, что при стохастической оптимизации (SGD) значение NSE на каждом мини-батче систематически отличается от финального NSE на валидационной выборке.

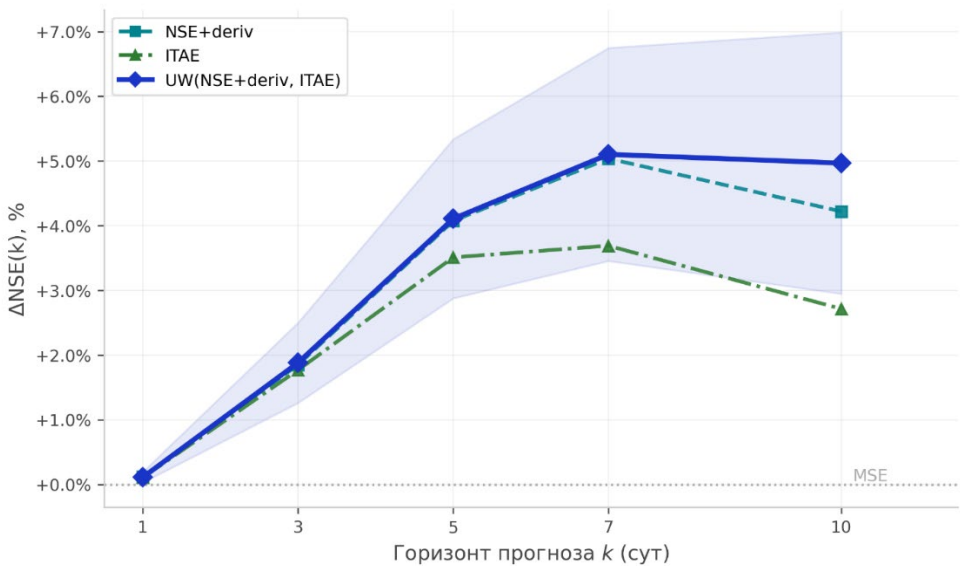


Рисунок 2. Прирост $NSE(k)$ альтернативных функций потерь относительно базовой линии MSE на горизонтах $k = 1, 3, 5, 7, 10$ сут: $\Delta NSE(k) = NSE(L, k) - NSE(MSE, k)$
Данные – TiDE, среднее по пяти водосборам, три инициализации. Полоса $\pm$ стандартное отклонение показана только для UW-композиции. Серый пунктир ($\Delta NSE = 0$) – уровень MSE-базовой линии. Все три альтернативы превосходят MSE на горизонтах $k \geq 3$. NSE+deriv превосходит ITAE по NSE на всех горизонтах; ITAE вносит самостоятельный вклад через смещение градиентного сигнала к дальним шагам. UW(NSE+deriv, ITAE) устойчиво превосходит остальные на всём диапазоне горизонтов

Figure 2. NSE(k) gain of alternative loss functions relative to the MSE baseline at horizons $k = 1, 3, 5, 7, 10$ days: $\Delta NSE(k) = NSE(L, k) - NSE(MSE, k)$
Data – TiDE, mean across five basins, three random initialisations. The $\pm$ standard deviation band is shown for the UW composition only. Grey dashed line ($\Delta NSE = 0$) – MSE baseline level. All three alternatives outperform MSE at horizons $k \geq 3$. NSE+deriv outperforms ITAE in NSE at all horizons; ITAE contributes independently by shifting the gradient signal toward distant steps. UW(NSE+deriv, ITAE) consistently outperforms each component across the full horizon range

NSE с штрафом на производную (NSE+deriv): контроль динамики ряда

MSE штрафует отклонение абсолютных значений, но не различает прогнозы, воспроизводящие скорость изменения уровня dh/dt . Для оперативного предупреждения об опасных водных явлениях критична именно динамика: скорость подъёма определяет время упреждения.

Предлагаемая функция потерь объединяет два компонента через механизм взвешивания по неопределённости (Kendall et al., 2018):

$$\mathcal{L}_{\text{NSE-deriv}} = \frac{1}{2\sigma_1^2} \mathcal{L}_{\text{NSE}} + \log\sigma_1 + \frac{1}{2\sigma_2^2} \mathcal{L}_{\text{deriv}} + \log\sigma_2, \quad (1)$$

где $\sigma_1, \sigma_2 > 0$ – обучаемые параметры масштаба, логарифмический штраф предотвращает обнуление вклада компоненты ($\sigma_i \rightarrow \infty$).

Компонент $\mathcal{L}_{\text{NSE}}$ вычисляется как:

$$\mathcal{L}_{\text{NSE}} = \left\langle \frac{SS_{\text{res},b}}{SS_{\text{tot},b} + \varepsilon} \right\rangle_{\text{valid}}, \quad (2)$$

где $SS_{\text{res},b} = \sum_{k=1}^T (h_{b,k} - \hat{h}_{b,k})^2$ – сумма квадратов остатков для примера b , $SS_{\text{tot},b} = \sum_{k=1}^T (h_{b,k} - \bar{h}_b)^2$ – полная сумма квадратов, $\bar{h}_b$ – среднее целевой переменной по окну, $\varepsilon = 10^{-2}$ – стабилизирующая константа. Усреднение $\langle \cdot \rangle_{\text{valid}}$ ведётся только по батчам с $SS_{\text{tot},b} > \varepsilon$ (маскировка «плоских» батчей).

При $SS_{\text{tot},b} \rightarrow 0$ градиент $\partial \mathcal{L}_{\text{NSE}} / \partial \hat{h} = 2(\hat{h} - h) / (SS_{\text{tot},b} + \varepsilon)$ стремится к $2(\hat{h} - h) / \varepsilon$, что при малом ε без маскировки привело бы к взрыву градиента. Маскировка обеспечивает численную устойчивость обучения вне зависимости от режима водосбора.

Компонент $\mathcal{L}_{\text{deriv}}$ штрафует несовпадение производных прогнозного ряда и наблюдений:

$$\mathcal{L}_{\text{deriv}} = \mathcal{L}_{\text{MSE}} + \lambda_1 \text{MSE}(\Delta \hat{h}, \Delta h) + \lambda_2 \text{MSE}(\Delta^2 \hat{h}, \Delta^2 h), \quad (3)$$

где $\Delta h_t = h_t - h_{t-1}$ – первая конечная разность (аппроксимация dh/dt), $\Delta^2 h_t = h_{t+1} - 2h_t + h_{t-1}$ – вторая (кривизна гидрографа). Штраф на несовпадение $\Delta \hat{h}$ и Δh обязывает модель воспроизводить скорость изменения уровня; штраф на Δ^2 – кривизну в точках перегиба (начало и окончание подъёма). Значения $\lambda_1 = 0.1$, $\lambda_2 = 0.05$ фиксированы во всех экспериментах. В качестве иллюстративного прогноза на рис. 3 использован инерционный прогноз $\hat{h}_k = h_{k-1}$; на в. 5287 градиентный сигнал сосредоточен на шагах с наиболее высокой скоростью изменения уровня.

Верхние панели – нормированный уровень воды $h^* = (h - \mu) / \sigma$ (чёрная линия, левая ось); бирюзовая заливка – скорость изменения уровня $|\Delta h|$ (правая ось); серая заливка – скользящие окна с $SS_{\text{tot}} < 0.01$, исключаемые из NSE-компоненты при обучении (периоды с нулевой дисперсией целевой переменной); светло-синяя заливка с вертикальными пунктирами – анализируемое окно ($T = 10$ сут) с наиболее интенсивным подъёмом уровня.

Нижние панели – накопленные столбцы: абсолютное значение градиента $|\partial \mathcal{L}_{\text{deriv}} / \partial \hat{h}_k|$ по шагу $k = 1 \dots 10$; серый – MSE-компонента, бирюзовый – $\lambda_1 \cdot \text{MSE}(\Delta \hat{h}, \Delta h)$, тёмно-бирюзовый – $\lambda_2 \cdot \text{MSE}(\Delta^2 \hat{h}, \Delta^2 h)$. Синяя линия (правая ось) – наблюдаемая скорость изменения уровня $|\Delta h_k|$ внутри анализируемого окна.

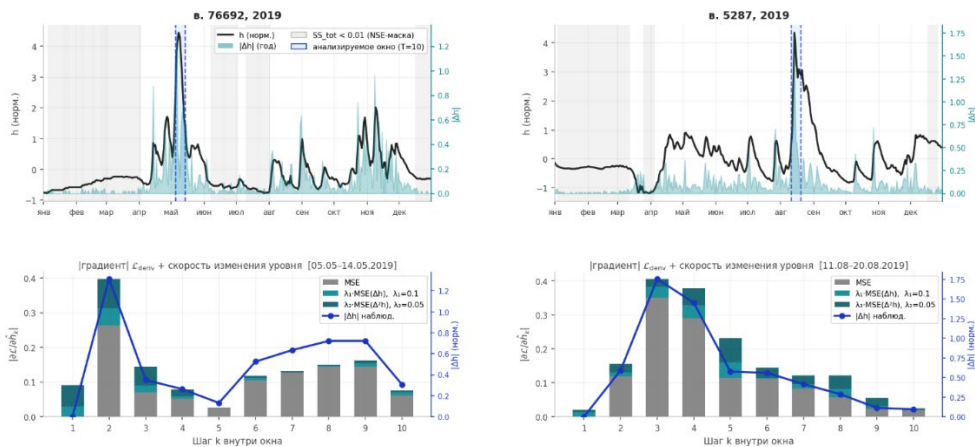


Рисунок 3. Механика $\mathcal{L}_{\text{deriv}}$ -компоненты функции потерь NSE+deriv, 2019 г. *Левая колонка – в. 76692, правая – в. 5287.*

Figure 3. Mechanics of the $\mathcal{L}_{\text{deriv}}$ component of the NSE+deriv loss function, 2019. *Left column – station 76692, right column – station 5287.*

ITAE: взвешивание по горизонту

Интегральная временно-взвешенная абсолютная ошибка (ITAE, Integral Time-weighted Absolute Error) применяется в теории автоматического управления для оценки переходных процессов, где поздние остаточные ошибки важнее ранних. В доступных систематических обзорах функций потерь для прогноза временных рядов (Jadon et al., 2024; Kim et al., 2025) применение ITAE в роли функции потерь при обучении нейросетей не зафиксировано; взвешивание по горизонту прогноза как самостоятельный класс методов в этих обзорах не выделяется.

$$\mathcal{L}_{\text{ITAE}} = \frac{1}{B \bar{k}} \sum_{b=1}^B \sum_{k=1}^{T_{\text{out}}} k \cdot |\hat{h}_{b,k} - h_{b,k}|, \quad (4)$$

где $\bar{k} = T_{\text{out}}(T_{\text{out}} + 1)/2$ – нормировочная константа, обеспечивающая сопоставимость масштабов с MSE. Ошибка на горизонте $k = 10$ штрафуются в 10 раз сильнее, чем на $k = 1$. Это непосредственно следует из формулы: градиент по прогнозу $\partial \mathcal{L}_{\text{ITAE}} / \partial \hat{h}_{b,k} = k \cdot \text{sign}(\hat{h}_{b,k} - h_{b,k}) / (B \bar{k})$, то есть градиентный сигнал на дальних горизонтах линейно сильнее, чем на ближних (рис. 4).

Ма и др. (Ma et al., 2026) независимо предложили метрику качества многошагового прогноза (AC score – forecast accuracy and coherence score), учитывающую точность и согласованность прогнозов по горизонтам, и применили её как дифференцируемую цель обучения.

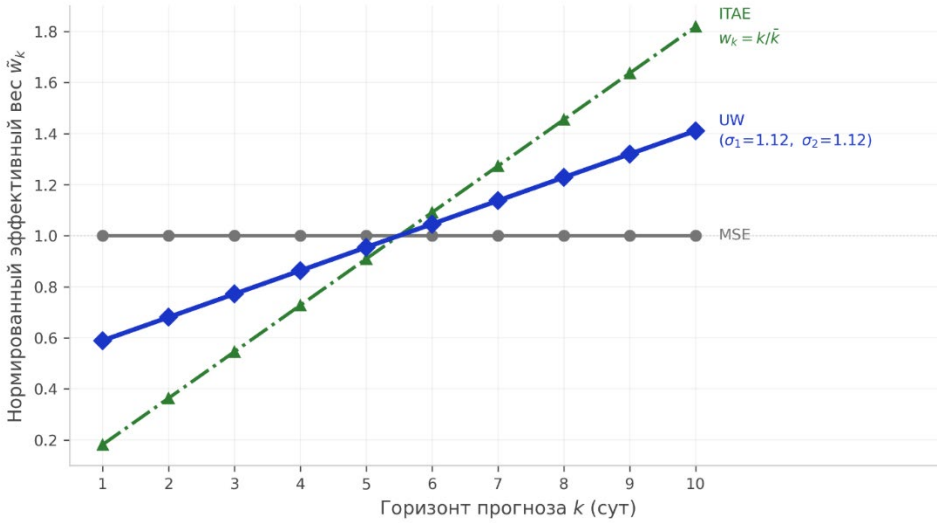


Рисунок 4. Профиль эффективных весов по горизонту прогноза $k = 1 \dots 10$ для трёх функций потерь

MSE (серый): равномерный вес $w_k = 1$ на всех горизонтах. ITAE (зелёный): линейно нарастающий вес $w_k = k/\bar{k}$, где $\bar{k} = T_{out}(T_{out} + 1)/2 = 55$; ошибка на $k = 10$ штрафуетя в 10 раз сильнее, чем на $k = 1$. UW-композиция (синий): суммарный эффективный вес $w_k^{UW} = \sigma_1^{-2} \cdot 1 + \sigma_2^{-2} \cdot k/\bar{k}$, где σ_1^{-2} и σ_2^{-2} — обученные точности NSE+deriv- и ITAE-компонент (формула (5)). Абсолютные уровни трёх кривых не сопоставимы напрямую ввиду различных нормировок; сравнивается форма распределения весов

Figure 4. Effective weight profile across forecast horizon $k = 1 \dots 10$ for three loss functions MSE (grey): uniform weight $w_k = 1$ at all horizons. ITAE (green): linearly increasing weight $w_k = k/\bar{k}$, where $\bar{k} = T_{out}(T_{out} + 1)/2 = 55$; the penalty at $k = 10$ is 10 times larger than at $k = 1$. UW composition (blue): total effective weight $w_k^{UW} = \sigma_1^{-2} \cdot 1 + \sigma_2^{-2} \cdot k/\bar{k}$, where σ_1^{-2} and σ_2^{-2} are the learned precisions of the NSE+deriv and ITAE components (formula (5)). Absolute levels of the three curves are not directly comparable due to different normalisations; the shape of the weight distribution is compared

UW-композиция: автоматическое балансирование

Две предыдущие функции потерь адресуют разные аспекты проблемы: NSE+deriv нормирует ошибку по изменчивости ряда и штрафует отклонения в скорости изменения уровня, ITAE смещает градиентный сигнал к дальним горизонтам. По NSE NSE+deriv превосходит ITAE на всех горизонтах; тем не менее их совместное использование через UW даёт результат лучше каждого из компонентов в отдельности — за счёт взаимодополняющего действия двух механизмов.

Kendall и др. (Kendall et al., 2018) предложили механизм автоматического балансирования нескольких функций потерь через обучаемые параметры масштаба каждой компоненты:

$$\mathcal{L}_{UW} = \frac{1}{2\sigma_1^2} \mathcal{L}_{NSE-deriv} + \log\sigma_1 + \frac{1}{2\sigma_2^2} \mathcal{L}_{ITAE} + \log\sigma_2, \quad (5)$$

где σ_1, σ_2 – обучаемые параметры (инициализируются нулём в пространстве $\log \sigma^2$). Запись эквивалентна исходной формулировке авторов (Kendall et al., 2018, ур. 7) с заменой $\log \sigma_1 \sigma_2$ на $\log \sigma_1 + \log \sigma_2$. Логарифмический штраф выполняет роль регуляризатора: при $\sigma_i \rightarrow \infty$ он предотвращает обнуление вклада компоненты. При большом масштабе $\mathcal{L}_i$ параметр σ_i растёт, снижая вес компоненты; при малом – уменьшается, усиливая вклад. Структура вычисления UW-композиции представлена на рис. 5. В проведённых экспериментах обученные значения $\sigma_1 \approx \sigma_2 \approx 1.12$, что указывает на близкий масштаб обоих компонентов на протяжении обучения и примерно равное их взвешивание в итоговой функции потерь.

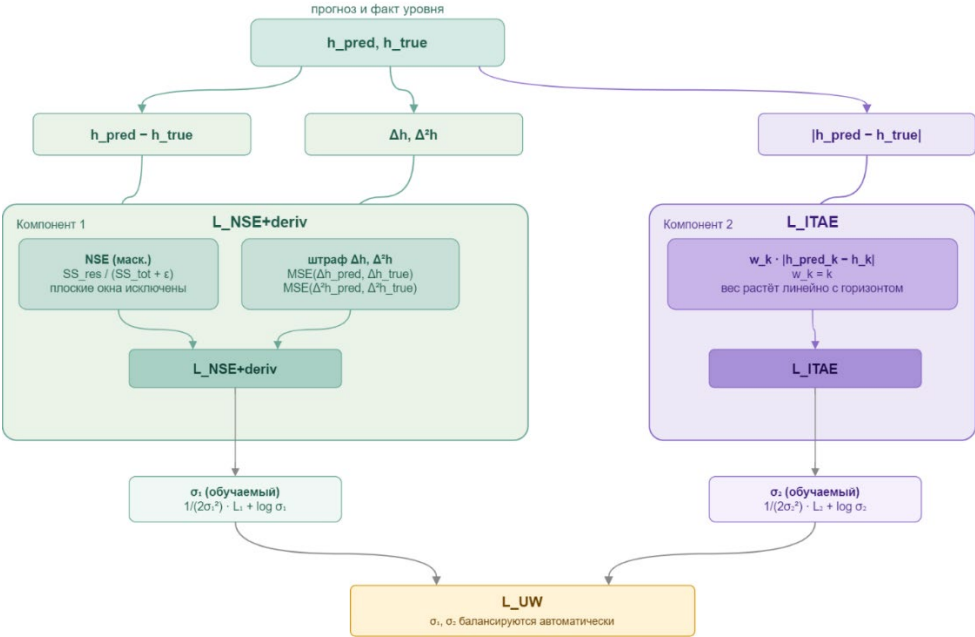


Рисунок 5. Схема вычисления UW-композиции
Прогнозный ряд $\hat{h}_{1:T}$ поступает параллельно в два блока: $\mathcal{L}_{NSE-deriv}$ и $\mathcal{L}_{ITAE}$. Выходы взвешиваются параметрами σ_1, σ_2 . Суммарная функция потерь $\mathcal{L}_{UW}$ передаётся в оптимизатор

Figure 5. Diagram of the UW composition computation
The forecast series $\hat{h}_{1:T}$ is fed in parallel to two blocks: $\mathcal{L}_{NSE-deriv}$ and $\mathcal{L}_{ITAE}$. The outputs are weighted by parameters σ_1, σ_2 . The combined loss $\mathcal{L}_{UW}$ is passed to the optimiser.

Постановка эксперимента

Конфигурации функций потерь

Все модели обучаются в режиме совместного обучения – единая модель на данных всех пяти водосборов одновременно, без разделения по станциям. Совместное обучение выбрано в среднем по пяти водосборам: дополнительный объём данных компенсирует ограниченность ряда наблюдений на каждый отдельный водосбор. Обоснование и количественное сравнение с индивидуальным обучением приведены в разделе Дискуссия, подраздел «Совместное обучение на разных водосборах». Для отдельных водосборов результат индивидуального обучения может превосходить

совместное – см. подраздел «Индивидуальные модели». Такая постановка зафиксирована для всех конфигураций функций потерь, что обеспечивает корректность сравнения. Цель работы – оценка влияния выбора функции потерь на качество прогноза, а не достижение максимальных абсолютных значений метрик оценки.

Сравниваются четыре конфигурации:

- **MSE** – базовая линия
- **NSE+deriv** – формула (1) – (3), $\lambda_1 = 0.1, \lambda_2 = 0.05, \varepsilon = 10^{-2}$
- **ITAE** – формула (4)
- **UW(NSE+deriv, ITAE)** – формула (5).

Гиперпараметры функций потерь едины для всех экспериментов и представлены в таблице 3.

Таблица 3. Конфигурации функций потерь

Table 3. Loss function configurations

Функция потерь	Формула	Обучаемых параметров	Ключевые гиперпараметры
MSE	$\frac{1}{BT} \sum (\hat{h} - h)^2$	0	–
NSE+deriv	(1) – (3)	2 (σ_1, σ_2)	$\lambda_1 = 0.1,$ $\lambda_2 = 0.05$
ITAE	(4)	0	–
UW(NSE+deriv, ITAE)	(5)	2 (σ_1, σ_2)	–

Результаты

Результаты по NSE и S/σ_Δ для TiDE приведены в табл. 4 и 5 соответственно.

Таблица 4. NSE(k) – среднее по пяти станциям, среднее ± стандартное отклонение по трём случайным инициализациям. Жирным – лучший результат на каждом горизонте

Table 4. NSE(k) – mean across five stations, mean ± standard deviation over three random initialisations. Best result at each horizon in bold

Функция потерь	$k = 1$	$k = 3$	$k = 5$	$k = 7$	$k = 10$
MSE	0.933±0,000	0.781±0,012	0.658±0,022	0.566±0,030	0.479±0,038
NSE+deriv	0.934 ±0.002 (+0.1%)	0.799 ±0.005 (+2,4%)	0.699 ±0.009 (+6,2%)	0.616 ±0.012 (+8,9%)	0.521±0,013 (+8,8%)
ITAE	0.934 ±0.002 (+0.1%)	0.798±0.006 (+2.3%)	0.693±0.008 (+5.3%)	0.603±0.011 (+6.5%)	0.506±0,019 (+5.7%)
UW(NSE+deriv, ITAE)	0.934 ±0.002 (+0.1%)	0.800 ±0.006 (+2.4%)	0.699 ±0.011 (+6.2%)	0.617 ±0.013 (+9.0%)	0.529 ±0,013 (+10.4%)

Таблица 5. $S/\sigma_{\Delta}(k)$ – среднее по пяти станциям, среднее $\pm$ стандартное отклонение по трём случайным инициализациям. Курсивом – лучший результат на горизонте

Table 5. $S/\sigma_{\Delta}(k)$ – mean across five stations, mean $\pm$ standard deviation over three random initialisations. Best result at each horizon in italics

Функция потерь	$k = 1$	$k = 3$	$k = 5$	$k = 7$	$k = 10$
MSE	0.997 $\pm$ 0.005	0.993 $\pm$ 0.023	0.993 $\pm$ 0.028	0.990 $\pm$ 0.034	0.965 $\pm$ 0.036
NSE+deriv	0.995 $\pm$ 0.015 (-0.2%)	0.955 $\pm$ 0.013 (-3.8%)	0.937 $\pm$ 0.015 (-5.7%)	0.925 $\pm$ 0.015 (-6.5%)	0.910 $\pm$ 0.014 (-5.6%)
ITAE	0.979 $\pm$ 0.015 (-1.8%)	0.953 $\pm$ 0.015 (-4.0%)	0.946 $\pm$ 0.012 (-4.7%)	0.946 $\pm$ 0.013 (-4.4%)	0.931 $\pm$ 0.019 (-3.5%)
UW(NSE+deriv v. ITAE)	0.989 $\pm$ 0.014 (-0.7%)	0.952 $\pm$ 0.017 (-4.2%)	0.934 $\pm$ 0.018 (-5.9%)	0.923 $\pm$ 0.016 (-6.8%)	0.904 $\pm$ 0.014 (-6.3%)

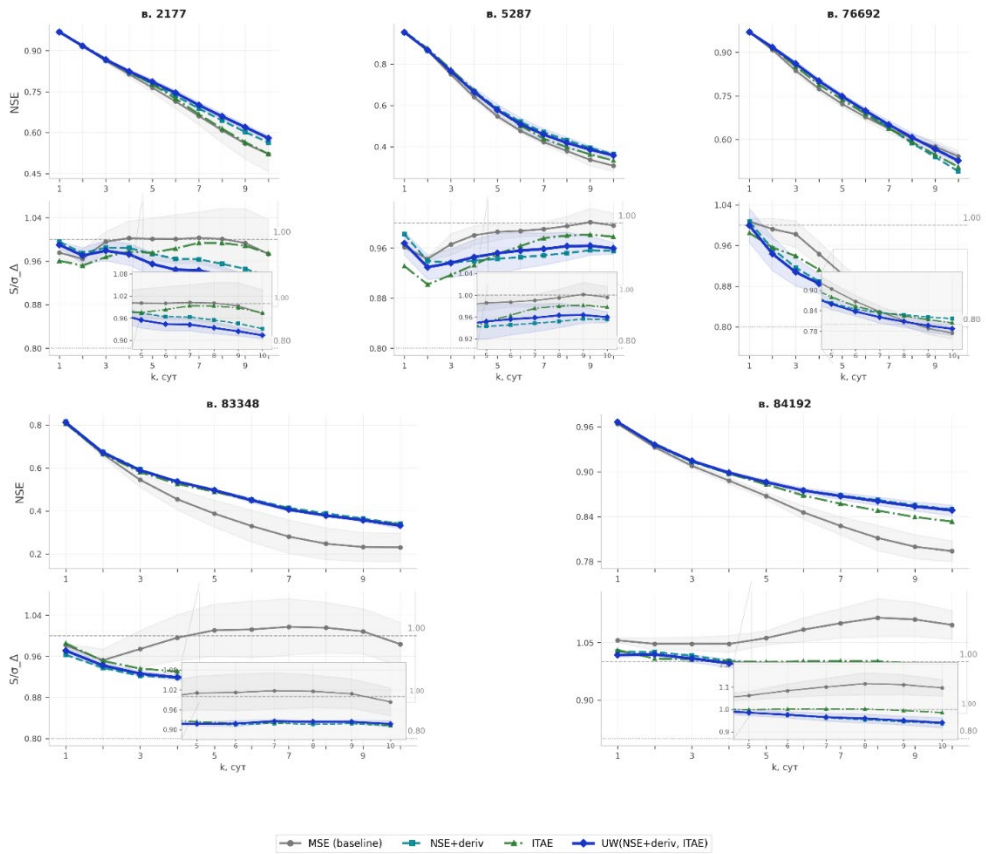


Рисунок 6. $NSE(k)$ и $S/\sigma_{\Delta}(k)$ как функции горизонта – TiDE
Четыре кривые (MSE , $NSE+deriv$, $ITAE$, UW) для среднего по станциям, с полосами $\pm$ стандартное отклонение. Горизонтальные пунктирные линии: $S/\sigma_{\Delta} = 0,80$ и $S/\sigma_{\Delta} = 1,00$.
Верхняя панель – $NSE(k)$, нижняя – $S/\sigma_{\Delta}(k)$. Данные: табл. 4-5

Figure 6. $NSE(k)$ and $S/\sigma_{\Delta}(k)$ as functions of forecast horizon – TiDE
Four curves (MSE , $NSE+deriv$, $ITAE$, UW) for the station-averaged mean, with $\pm$ standard deviation bands. Horizontal dashed lines: $S/\sigma_{\Delta} = 0,80$ and $S/\sigma_{\Delta} = 1,00$. Upper panel – $NSE(k)$, lower panel – $S/\sigma_{\Delta}(k)$. Data: Tables 4-5

Все три альтернативные конфигурации превосходят MSE на всех горизонтах $k \geq 3$ по обоим метрикам. На $k = 1$ различия незначительны, что согласуется с теоретическим результатом: при одношаговом прогнозе остаток близок к i.i.d. и все функции потерь дают близкий результат.

NSE+deriv и UW превосходят MSE по NSE на всех горизонтах; ITAE также улучшает MSE, но в меньшей степени по NSE. По S/σ_Δ ITAE показывает наилучший результат на $k = 1$ и конкурентный на $k = 3$. UW-композиция занимает наилучшее или близкое к наилучшему положение по обоим метрикам на всех горизонтах, подтверждая эффективность автоматического балансирования (рис. 6).

Анализ по станциям

Наибольший относительный прирост наблюдается на в. 83348: +28.2% NSE ($k = 5$) для NSE+deriv и -9.3% S/σ_Δ ($k = 5$), а также на в. 84192: -7.1% S/σ_Δ ($k = 5$) для NSE+deriv.

Обе станции характеризуются значением S/σ_Δ ($k = 5$, MSE) > 1.0 (в. 83348: 1.011; в. 84192: 1.062), то есть MSE-модель уступает инерционному прогнозу $\hat{h}_{1+k} = h_j$ по данному критерию. Анализ характеристик наблюдательных рядов (рис. 12) показывает, что в. 83348 выделяется отсутствием выраженного сезонного режима и высокой изменчивостью на всех горизонтах k , тогда как в. 76692 имеет один доминирующий весенний паводок. Детальный анализ механизмов прироста приведён в разделе «Дискуссия», подраздел «Теоретическая интерпретация результатов».

Таблица 6. NSE ($k = 5$) и S/σ_Δ ($k = 5$) по станциям, TiDE
Среднее $\pm$ стандартное отклонение по трём случайным инициализациям. Жирным – лучший результат по станции

Table 6. NSE ($k = 5$) and S/σ_Δ ($k = 5$) by station, TiDE					
Mean $\pm$ standard deviation over three random initialisations. Best result per station in bold					
Функция потерь	2177 NSE	5287 NSE	76692 NSE	83348 NSE	84192 NSE
MSE	0.765 $\pm$ 0.018	0.547 $\pm$ 0.012	0.722 $\pm$ 0.012	0.388 $\pm$ 0.062	0.868 $\pm$ 0.005
NSE+deriv	0.778 $\pm$ 0.008 (+1.7%)	0.586 $\pm$ 0.017 (+7.1%)	0.747 $\pm$ 0.010 (+3.4%)	0.497$\pm$0.008 (+28.2%)	0.886 $\pm$ 0.003 (+2.1%)
ITAE	0.777 $\pm$ 0.007 (+1.6%)	0.579 $\pm$ 0.019 (+5.8%)	0.737 $\pm$ 0.008 (+2.0%)	0.490 $\pm$ 0.007 (+26.3%)	0.883 $\pm$ 0.001 (+1.8%)
UW	0.786$\pm$0.009 (+2.8%)	0.578 $\pm$ 0.029 (+5.5%)	0.749$\pm$0.009 (+3.7%)	0.496 $\pm$ 0.006 (+28.1%)	0.886$\pm$0.003 (+2.1%)

Функция потерь	2177 S/σ_Δ	5287 S/σ_Δ	76692 S/σ_Δ	83348 S/σ_Δ	84192 S/σ_Δ
MSE	1.001 $\pm$ 0.038	0.986 $\pm$ 0.013	0.904 $\pm$ 0.019	1.011 $\pm$ 0.051	1.062 $\pm$ 0.019
NSE+deriv	0.973 $\pm$ 0.019 (-2.8%)	0.943 $\pm$ 0.020 (-4.4%)	0.863 $\pm$ 0.018 (-4.5%)	0.917$\pm$0.007 (-9.3%)	0.986 $\pm$ 0.011 (-7.1%)
ITAE	0.976 $\pm$ 0.014 (-2.6%)	0.951 $\pm$ 0.021 (-3.6%)	0.880 $\pm$ 0.014 (-2.6%)	0.924 $\pm$ 0.006 (-8.6%)	0.999 $\pm$ 0.004 (-5.9%)
UW	0.955$\pm$0.021 (-4.6%)	0.952 $\pm$ 0.032 (-3.5%)	0.860$\pm$0.015 (-4.9%)	0.918 $\pm$ 0.006 (-9.2%)	0.985$\pm$0.013 (-7.2%)

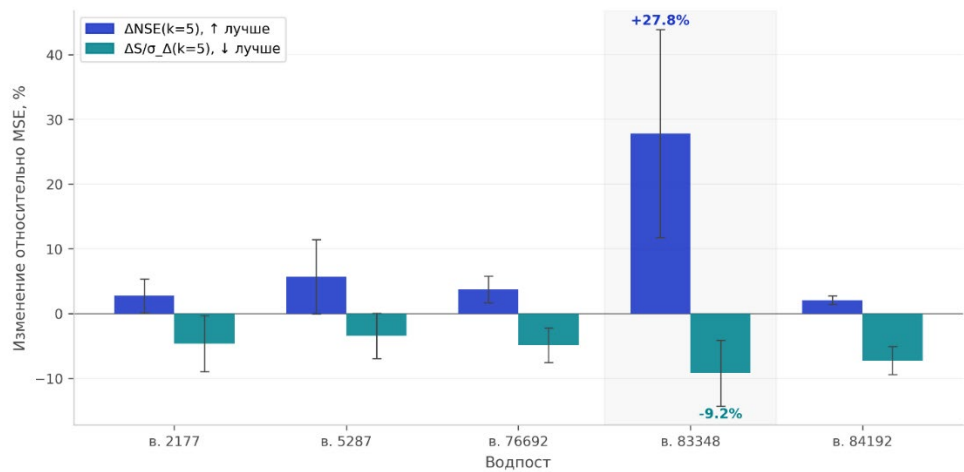


Рисунок 7. Прирост $NSE(k = 5)$ и снижение $S/\sigma_{\Delta}(k = 5)$ для UW-композиции относительно MSE по станциям

Сгруппированная столбчатая диаграмма: по оси X – пять водосборов, по оси Y – относительное изменение в %. Данные: табл. 6

Figure 7. $NSE(k = 5)$ gain and $S/\sigma_{\Delta}(k = 5)$ reduction for the UW composition relative to MSE, by basin

Grouped bar chart: X -axis – five basins, Y -axis – relative change in %. Data: Table 6

Примеры прогнозных гидрографов

Примеры прогнозных гидрографов для в. 76692 и в. 5287 показаны на рис. 8 и 9 соответственно.

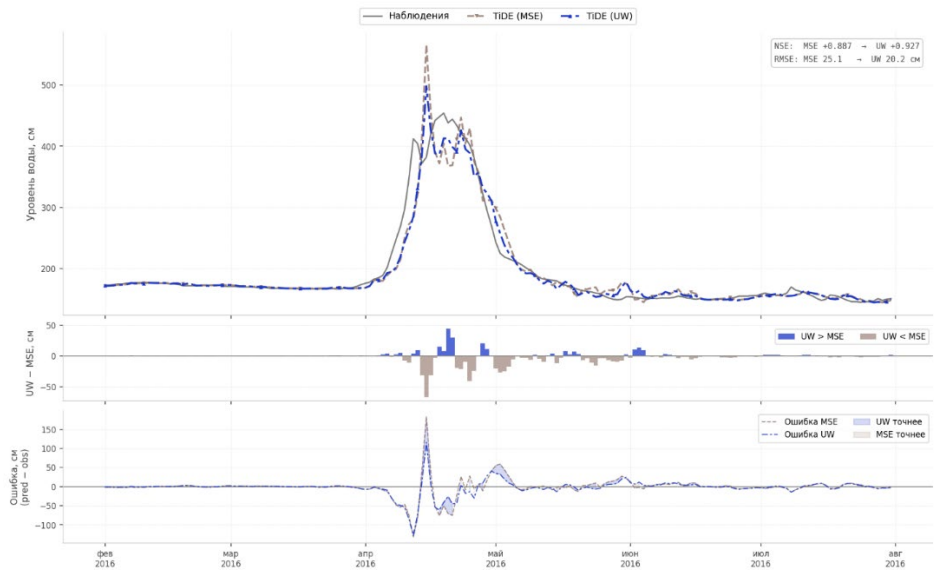


Рисунок 8. Прогнозные гидрографы на станции 76692 (р. Сылта, пгт. Шамары), заблаговременность $k = 3$ сут, февраль – август 2016 г. Наблюдения (чёрная сплошная), MSE (серая штриховая), UW (синяя). TiDE, валидационный период

Figure 8. Forecast hydrographs at station 76692 (Sylva River, Shamary), lead time $k = 3$ days, February – August 2016

Observations (black solid), MSE (grey dashed), UW (blue). TiDE, validation period

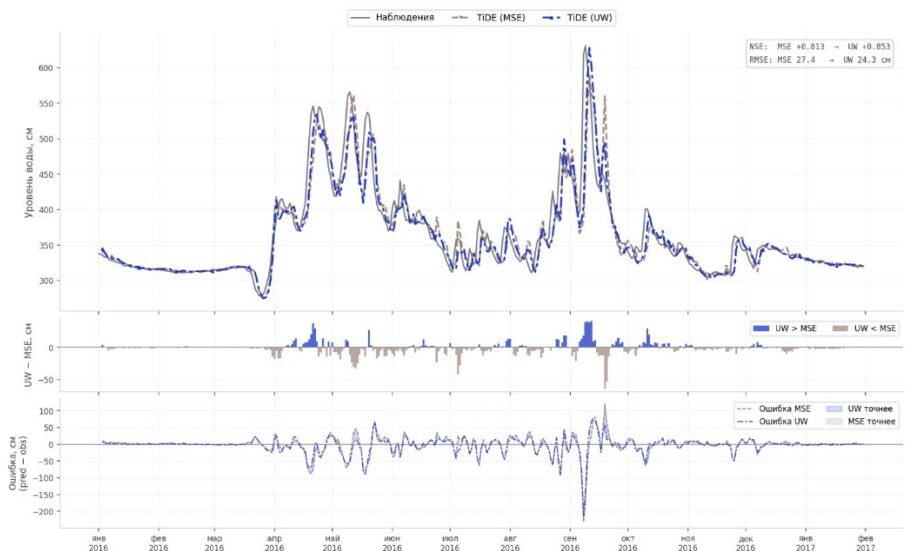


Рисунок 9. Прогнозные гидрографы на станции 5287 (р. Малиновка, с. Ракитное), заблаговременность $k = 2$ сут, январь 2016 – январь 2017 г.
Наблюдения (чёрная сплошная), MSE (серая штриховая), UW (синяя). TiDE, валидационный период

Figure 9. Forecast hydrographs at station 5287 (Malinovka River, Rakitnoe), lead time $k = 2$ days, January 2016 – January 2017
Observations (black solid), MSE (grey dashed), UW (blue). TiDE, validation period

Сравнение архитектур: TiDE и TSMixer

Результаты для TSMixer приведены в табл. 7 и 8.

Таблица 7. NSE (k) – TSMixer, среднее по 5 станциям

Table 7. NSE (k) – TSMixer, mean across 5 stations

Функция потерь	$k = 1$	$k = 3$	$k = 5$	$k = 7$	$k = 10$
MSE	0.923±0.006	0.816±0.007	0.734±0.012	0.664±0.018	0.590±0.016
NSE+deriv	0.922±0.006 (-0.1%)	0.806±0.006 (-1.2%)	0.722±0.006 (-1.6%)	0.655±0.005 (-1.3%)	0.580±0.010 (-1.6%)
ITAE	0.900±0.014 (-2.5%)	0.803±0.011 (-1.6%)	0.724±0.010 (-1.4%)	0.659±0.015 (-0.7%)	0.589±0.012 (-0.2%)
UW (NSE+deriv, ITAE)	0.920±0.007 (-0.3%)	0.805±0.007 (-1.3%)	0.724±0.006 (-1.4%)	0.662±0.006 (-0.2%)	0.595±0.010 (+0.9%)

Таблица 8. $S/\sigma_{\Delta}(kk)$ – TSMixer, среднее по 5 станциям

Table 8. $S/\sigma_{\Delta}(k)$ – TSMixer, mean across 5 stations

Функция потерь	$k = 1$	$k = 3$	$k = 5$	$k = 7$	$k = 10$
MSE	1.109±0.055	0.913±0.016	0.893±0.020	0.883±0.024	0.867±0.017
NSE+deriv	1.112±0.064 (+0.3%)	0.925±0.018 (+1.3%)	0.887±0.010 (-0.6%)	0.871±0.008 (-1.4%)	0.854±0.011 (-1.5%)

Продолжение таблицы 8

Функция потерь	$k = 1$	$k = 3$	$k = 5$	$k = 7$	$k = 10$
ITAE	1.305 ± 0.126 (+17.7%)	0.942 ± 0.028 (+3.2%)	0.895 ± 0.017 (+0.3%)	0.876 ± 0.021 (-0.8%)	0.857 ± 0.013 (-1.1%)
UW (NSE+deriv, ITAE)	1.137 ± 0.070 (+2.6%)	0.931 ± 0.022 (+2.0%)	0.886 ± 0.012 (-0.7%)	0.863 ± 0.009 (-2.2%)	0.841 ± 0.011 (-2.9%)

Поведение функции потерь зависит от архитектуры: на TiDE UW(NSE+deriv, ITAE) превосходит MSE-базовую линию на всём диапазоне горизонтов $k = 1 \dots 10$ суток как по NSE, так и по S/σ_{Δ} . На TSMixer картина иная: на коротких горизонтах ($k = 1, 3$) MSE-базовая линия не хуже либо лучше UW-композиции и других альтернативных функций потерь; перевес в сторону UW и NSE+deriv формируется начиная с $k = 5$ и далее устойчиво сохраняется до $k = 10$. ITAE на TSMixer выделяется отдельно: на горизонте $k = 1$ даёт заметно худший S/σ_{Δ} , чем остальные функции потерь, но сходится с ними уже к $k = 3$. Сравнение архитектур по $NSE(k)$ и $S/\sigma_{\Delta}(k)$ представлено на рис. 10.

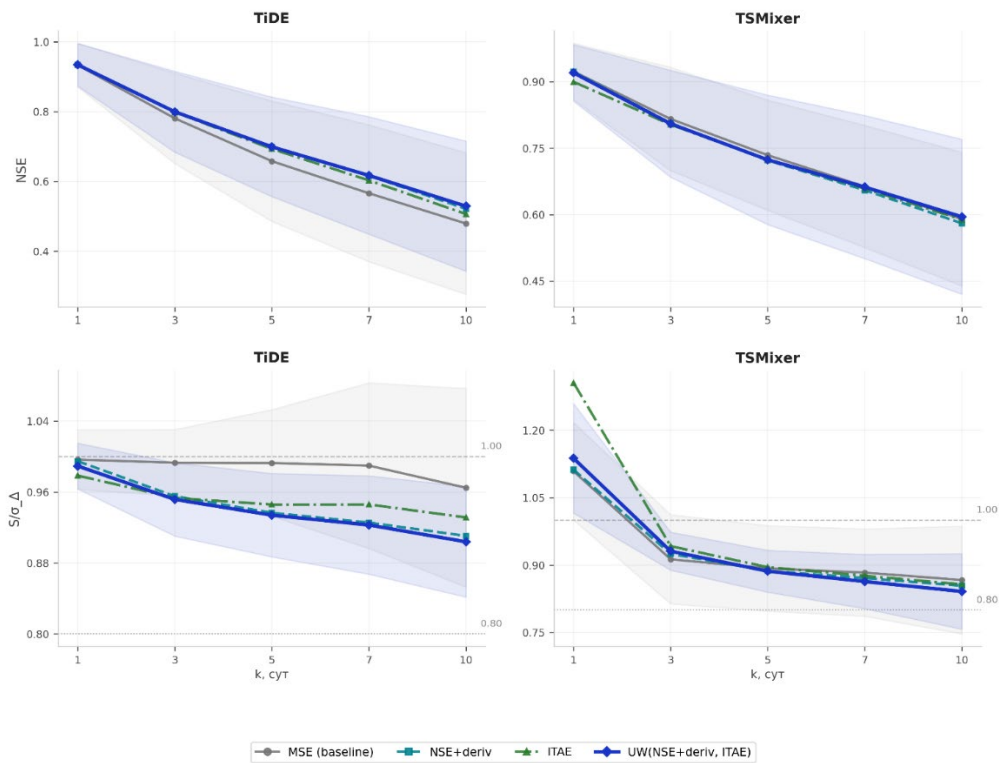


Рисунок 10. Сравнение TiDE и TSMixer по $NSE(k)$
Четыре кривые (MSE, NSE+deriv, ITAE, UW) × две панели (TiDE, TSmixer). Данные: табл. 4 и 7
Figure 10. Comparison of TiDE and TSMixer by $NSE(k)$
Four curves (MSE, NSE+deriv, ITAE, UW) × two panels (TiDE, TSmixer). Data: Tables 4 and 7

Дискуссия

Совместное обучение на разных водосборах

Совместное обучение мотивировано не только удобством сравнения функций потерь, но и ограниченным объёмом данных на каждый отдельный водосбор: несколько лет суточных наблюдений на одну станцию дают модели заметно меньше сигнала для извлечения общей, не специфичной для конкретного водосбора части зависимости между метеорологическими ковариатами и уровнем воды, чем при объединении всех пяти рядов. Модель различает водосборы через статический идентификатор станции, то есть совместное обучение не означает смешивание разнородных процессов в одну неразличимую выборку. Эффект количественно подтверждён на станции 2177: при обучении только на данных этого водпоста результат хуже на 7.6% по NSE на горизонте $k = 5$ суток и на 15.1% на горизонте $k = 10$ суток по сравнению с совместной моделью.

Совместное использование данных нескольких водосборов для обучения нейросетевых моделей опирается на эффект синергии данных (Fang et al., 2022): качество нейросетевой модели выигрывает от умеренно разнородной обучающей выборки, не от однородной, причём даже при сопоставимом объёме данных. Авторы формулируют это как гипотезу о способности нейронной сети выучивать общие и различающиеся между водосборами закономерности, не как доказанный механизм.

Единая модель, обученная на множестве водосборов, превосходит модели, откалиброванные индивидуально под каждый бассейн (Kratzert et al., 2019; на 531 водосборе набора данных CAMELS); для нейросетевых моделей этот результат устойчиво воспроизводится на больших выборках водосборов (Nearing et al., 2021).

Источники выше рассматривают разнородность водосборов по климатическим и физико-географическим характеристикам (Fang et al., 2022); типы формирования стока как отдельная ось анализа в них не выделяются. Региональная калибровка традиционных гидрологических моделей, в отличие от нейросетевых, требует однородности условий формирования стока (Kratzert et al., 2019).

Индивидуальные модели

Совместная модель на всех пяти водосборах с общими гиперпараметрами остаётся основным результатом настоящей работы – она позволяет сравнивать функции потерь в контролируемых условиях, не смешивая эффект выбора функции потерь с эффектом подбора конфигурации под конкретный водосбор. Отдельно от этого сравнения для каждого водосбора была подобрана собственная конфигурация – архитектура (TiDE или TSMixer) и длина входного окна (рис. 11); результат индивидуальных моделей превосходит совместную модель именно за счёт этого подбора.

Для водосбора 76692 индивидуальная модель TiDE (длина окна 80 суток) достигает $S/\sigma_{\Delta} = 0.640$ на горизонте $k=10$ суток – ниже порога 0.80. Для водосбора 2177 индивидуальная модель TSMixer (длина окна 45 суток)

достигает $S/\sigma_{\Delta} = 0.797$ на том же горизонте. Для водосборов 5287, 83348 и 84192 значение S/σ_{Δ} на горизонте $k = 10$ суток составило 0.845-0.872 – выше порога 0.80. Дальнейшее повышение точности для этих водосборов может быть связано с более широким перебором гиперпараметров и набора используемых признаков.

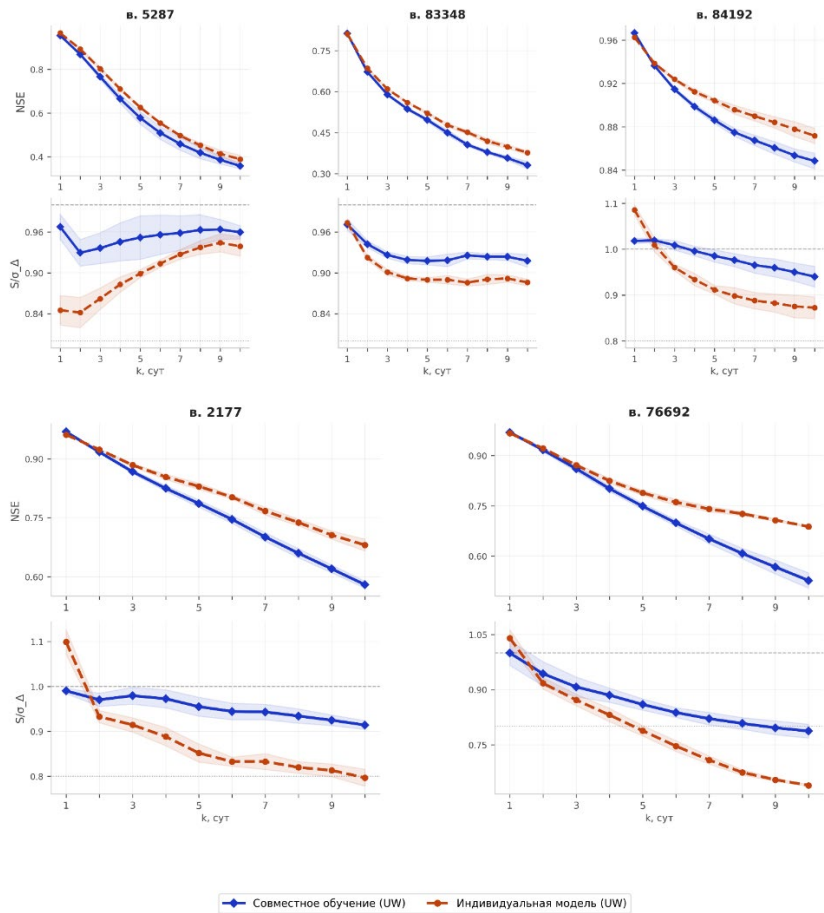


Рисунок 11. $NSE(k)$ и $S/\sigma_{\Delta}(k)$: совместное обучение на пяти водосборах против индивидуальной модели для каждого водосбора (UW)
Для каждого водосбора использована отдельно найденная конфигурация. Горизонтальная линия – порог приемлемости $S/\sigma_{\Delta} = 0.80$

Figure 11. $NSE(k)$ and $S/\sigma_{\Delta}(k)$: joint training on five catchments vs. an individually tuned model for each catchment (UW)
A separately found configuration was used for each catchment. Horizontal line – acceptability threshold $S/\sigma_{\Delta} = 0.80$

Теоретическая интерпретация результатов

Наблюдаемая закономерность – все три альтернативы превосходят MSE, при этом $NSE+deriv$ стабильно превосходит ITAE по NSE, а их совместное использование через UW даёт результат лучше каждого из компонентов в отдельности. На коротком горизонте структура ошибки близка к однородной и все функции потерь дают схожий результат. Дополнительный

штраф на производную (Δh -компонента) помогает модели воспроизводить фазу подъёма. На дальних горизонтах дисперсия ошибки возрастает (Ben Taieb, Atiya, 2016), что нарушает предположение о равнозначности горизонтов; линейное взвешивание ITAE компенсирует недостаточный вес дальних шагов в MSE, однако этого недостаточно, чтобы превзойти NSE+deriv по NSE – UW объединяет оба механизма. Hounie и др. (Hounie et al., 2024) зафиксировали, что оптимизация средней ошибки по прогнозному окну приводит к неравномерному распределению ошибок по шагам, особенно для архитектур класса трансформер, что подтверждает субоптимальность равномерно взвешенной функции потерь на всём диапазоне горизонтов. В отличие от подхода с ограничениями на функцию потерь каждого шага (Hounie et al., 2024), ITAE задаёт фиксированное взвешивание; адаптация баланса между компонентами достигается на уровне UW-композиции. Caljon и др. (Caljon et al., 2025) зафиксировали, что в задаче совместной оптимизации точности и стабильности прогноза метод UW (Kendall et al., 2018) ухудшает точность относительно статических весов; в предлагаемой постановке, где оба компонента направлены на улучшение качества прогноза по горизонту, подобного эффекта не наблюдается: UW-композиция улучшает как NSE, так и S/σ_{Δ} на всех горизонтах $k \geq 3$.

Ben Taieb и Atiya (Ben Taieb, Atiya, 2016) анализировали рекурсивную стратегию прогноза (recursive multi-step forecasting), при которой одношаговая модель последовательно применяется к собственным предсказаниям. TiDE использует прямую многошаговую стратегию (MIMO): весь вектор горизонта выдаётся за один проход, что исключает накопление ошибок через рекурсивную цепочку. Для архитектур типа трансформер, генерирующих весь горизонт за один прямой проход, неравномерность ошибок по шагам при равномерном взвешивании зафиксирована эмпирически (Hounie et al., 2024), однако теоретического доказательства для произвольных архитектур нет. Настоящие результаты на TiDE подтверждают эту закономерность эмпирически для данной архитектуры и набора данных.

Тем же механизмом – отсутствием доступа модели к информации о будущем состоянии метеорологических ковариат – объясняется систематическое отставание модельного гидрографа от наблюдаемого по времени, усиливающееся с ростом горизонта (рис. 8-9). Модель использует исключительно признаки прошлого ($t' \leq t$) и прогнозирует весь горизонт $k = 1 \dots 10$ суток за один проход (MIMO-стратегия), не имея доступа к значениям метеорологических ковариат на самом интервале прогноза ($t, t+k$]. Чем больше горизонт k , тем больше доля динамики паводка или половодья формируется уже после последней точки, доступной модели на входе, отсюда – систематическое запаздывание реакции модели и недооценка амплитуды на дальних горизонтах. Эта составляющая ошибки не устраняется калибровкой модели (подбором гиперпараметров, нормализацией), поскольку она обусловлена самой постановкой задачи – отсутствием доступа к метеорологическим данным на сам интервал прогноза, а не параметрами модели. Устранить её может переход к постановке с прогностическими

метеорологическими ковариатами на горизонт k , что выходит за рамки настоящей работы.

Рост значения критерия оправдываемости S/σ_Δ с увеличением заблаговременности (рис. 6, 11) не означает повышения абсолютной точности прогноза. Критерий нормирует ошибку модели на $\sigma_\Delta(k)$ – стандартное отклонение k -суточных изменений уровня, то есть фактически на ошибку инерционного прогноза (предположения о неизменности уровня на горизонт k). Эта величина растёт с горизонтом, поскольку типичная амплитуда изменения уровня за больший срок выше. Абсолютная ошибка модели также растёт с горизонтом, но медленнее, чем $\sigma_\Delta(k)$, поэтому их отношение убывает: прогноз модели оказывается относительно более выгодным по сравнению с инерционным именно на дальних горизонтах, при том что его собственная абсолютная ошибка также растёт. В отличие от S/σ_Δ , коэффициент NSE нормирует ошибку на дисперсию самого ряда относительно среднего, а не на изменчивость приращений – эта величина не зависит от горизонта, поэтому NSE монотонно убывает с ростом k , отражая рост абсолютной ошибки модели без эффекта изменяющегося знаменателя.

Зависимость эффекта функции потерь от архитектуры модели

На коротких горизонтах ($k = 1, 3$) выбор функции потерь на TSMixer практически не даёт выигрыша относительно MSE – UW и NSE+deriv идут вровень с базовой линией или уступают ей; выигрыш UW-композиции проявляется только с $k=5$ и далее нарастает к $k=10$. Это противоположно поведению TiDE, где UW устойчиво лучше MSE на всём диапазоне горизонтов с $k=1$.

Зависимость эффекта функции потерь от гидрологического режима

Наибольший прирост качества при замене MSE на NSE+deriv и UW-композицию наблюдается на станциях с $S/\sigma_\Delta(\text{MSE}) > 1.0$ – в. 83348 и в. 84192, то есть там, где MSE-модель уступает инерционному прогнозу.

В. 83348 характеризуется высокой нерегулярностью внутригодового распределения паводков. Для количественной оценки степени сезонной нерегулярности вводится нормированная энтропия Шеннона сезонного распределения локальных максимумов ряда:

$$H_{\text{норм}} = \frac{H_{\text{ент}}}{\log_2 12}, \quad H_{\text{ент}} = - \sum_{m=1}^{12} p_m \log_2 p_m,$$

где $p_m = n_m / \sum_m n_{m'}$ – доля локальных максимумов, приходящихся на месяц m ; n_m – число максимумов в месяц m , подсчитанное по всему валидационному периоду (2016-2019) по склеенному ряду. Локальные максимумы определяются как точки с выделенностью (prominence) не менее $0.25 \cdot (h_{\text{max}} - h_{\text{min}})$ и расстоянием между соседними пиками не менее 30 суток. Выделенность точки определяется как минимальное снижение уровня от данного пика до любого более высокого пика в ряду – формально, высота пика относительно наивысшей точки на пути к ближайшему более высокому

максимуму (Virtanen et al., 2020). При $H_{\text{норм}} = 0$ все максимумы сосредоточены в одном месяце; при $H_{\text{норм}} = 1$ они равномерно распределены по всем 12 месяцам. Для в. 83348 $H_{\text{норм}} = 0.80$, тогда как у в. 76692 $H_{\text{норм}} = 0.59$ – один доминирующий весенний паводок с несколькими осенними вторичными пиками (рис. 12, панели а–в).

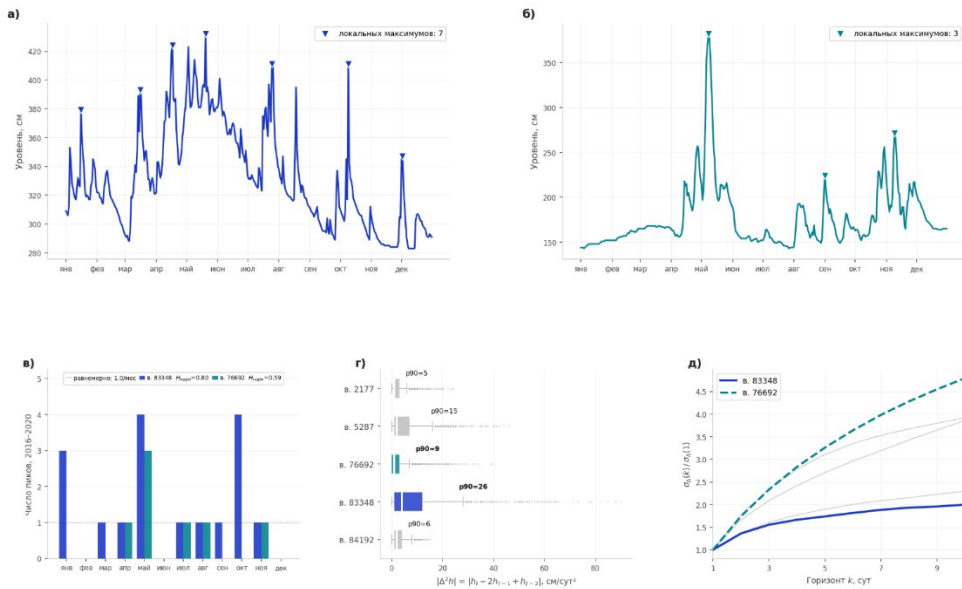


Рисунок 12. Характеристики гидрологических рядов – в. 83348 и в. 76692, валидационный период. а) Гидрограф уровня воды в. 83348. Маркерами $\blacktriangledown$ отмечены локальные максимумы ряда (выделенность не менее $0.25 \cdot (h_{\text{max}} - h_{\text{min}})$, минимальное расстояние между соседними максимумами – 30 суток). В заголовке – нормированная энтропия сезонности $H_{\text{норм}}$ (см. §6.3). б) То же для в. 76692. в) Число локальных максимумов по месяцам для в. 83348 и в. 76692. Пунктир – равномерное распределение. При $H_{\text{норм}} = 0$ все максимумы в одном месяце, при $H_{\text{норм}} = 1$ – равномерно по всем 12 месяцам. г) Распределение модуля второй конечной разности $|\Delta^2 h_t| = |h_t - 2h_{t-1} + h_{t-2}|$ по станциям (горизонтальный ящик с усами). Показаны все пять станций; в. 83348 и в. 76692 выделены цветом. Значения $p90$ подписаны. д) Нормированное СКО изменений уровня $\sigma_{\Delta}(k)/\sigma_{\Delta}(1)$ как функция горизонта $k = 1 - 10$ сут. Показаны все пять станций; в. 83348 – сплошная синяя линия, в. 76692 – пунктирная

Figure 12. Hydrological time series characteristics – stations 83348 and 76692, validation period. а) Water level hydrograph at station 83348. Local maxima are marked with $\blacktriangledown$ (prominence $\geq 0.25 \cdot (h_{\text{max}} - h_{\text{min}})$, minimum distance between adjacent peaks – 30 days). The panel title shows the normalised seasonal entropy $H_{\text{норм}}$ (see Section 6.3). б) Same for station 76692. в) Number of local maxima by month for stations 83348 and 76692. Dashed line – uniform distribution. At $H_{\text{норм}} = 0$ all maxima fall in one month; at $H_{\text{норм}} = 1$ they are evenly distributed across all 12 months. д) Distribution of the absolute second finite difference $|\Delta^2 h_t| = |h_t - 2h_{t-1} + h_{t-2}|$ across stations (horizontal box-and-whisker plot). All five stations shown; stations 83348 and 76692 highlighted by colour. The $p90$ values are annotated. е) Normalised standard deviation of water level changes $\sigma_{\Delta}(k)/\sigma_{\Delta}(1)$ as a function of horizon $k = 1 - 10$ days. All five stations shown; station 83348 – solid blue line, station 76692 – dashed

Дополнительным свидетельством нерегулярности режима служит 90-й процентиль второй конечной разности уровня $|\Delta^2 h_t| = |h_t - 2h_{t-1} + h_{t-2}|$, характеризующей резкость смен направления изменения ряда (кривизну гидрографа). Для в. 83348 $|\Delta^2 h|_{p90} = 26$ см/сут² против 9 см/сут² у в. 76692

(рис. 12, панель г). Наконец, нормированное СКО изменений уровня $\sigma_{\Delta}(k)/\sigma_{\Delta}(1)$ у в. 83348 возрастает к $k = 10$ сут примерно вдвое; у в. 76692 – примерно в пять раз (рис. 12, панель д).

В условиях совместного обучения на пяти станциях нерегулярный режим в. 83348 слабо согласуется с режимами других водосборов. MSE усредняет градиентный сигнал по всем станциям пропорционально квадратам абсолютных ошибок. NSE-компонента в составе NSE+deriv и UW нормирует ошибку на дисперсию целевого ряда в пределах каждого окна:

$$\mathcal{L}_{\text{NSE}} = \frac{SS_{\text{res}}}{SS_{\text{tot}} + \varepsilon},$$

что обеспечивает пропорциональное представление всех станций в градиентном сигнале независимо от абсолютного диапазона колебаний. NSE-нормировка устраняет дисбаланс градиентного сигнала между станциями с разной абсолютной изменчивостью ряда.

В. 84192 представляет иной случай: малый абсолютный диапазон изменений уровня (рис. 12). Квадратичный штраф MSE слабо различает хороший и плохой прогноз в абсолютных единицах при столь малом диапазоне, тогда как NSE-нормировка на SS_{tot} восстанавливает чувствительность функции потерь к относительным отклонениям.

Ограничения

Эксперименты проведены на пяти водосборах с единой моделью, обученной на всех водосборах одновременно. Такая постановка обеспечивает корректность сравнения функций потерь, однако её цель – не достижение максимальных абсолютных значений метрик оценки. Критерий приемлемости $S/\sigma_{\Delta} < 0.80$ в данной постановке достигнут на двух из пяти водосборов при индивидуальной настройке конфигурации модели под конкретный водосбор; результаты приведены в Дискуссии, подраздел «Индивидуальные модели». В предшествующем исследовании на тех же данных (Романов и др., 2026) удовлетворительные значения критерия были получены на трёх из пяти водосборов.

Переносимость результатов на водосборы с иным режимом формирования стока требует отдельного исследования.

Гиперпараметры NSE+deriv (λ_1, λ_2 , порог маскировки) не подбирались и едины для всех экспериментов.

Заключение

Три предложенные функции потерь – NSE+deriv, ITAE и UW-композиция – нацелены на различные аспекты структуры многошаговой ошибки: динамику прогнозного ряда, смещение градиентного сигнала к дальним горизонтам и автоматическое балансирование обоих механизмов. На пяти водосборах с архитектурой TiDE лучшая конфигурация – UW(NSE+deriv,

ITAE) – улучшает NSE на горизонте 10 суток на 10.4% и снижает S/σ_{Δ} на 6.3% при сохранении качества на ближних горизонтах; прирост NSE составляет 2-10% на горизонтах 3-10 суток. В доступных систематических обзорах функций потерь для прогноза временных рядов (Jadon et al., 2024; Kim et al., 2025) применение ITAE в роли функции потерь при обучении нейросетей не зафиксировано.

Дополнительно установлено, что эффективность функции потерь UW зависит от архитектуры модели: на TiDE она проявляется на всём диапазоне горизонтов, тогда как на TSMixer – только начиная с горизонта 5 суток. На двух из пяти водосборов критерий приемлемости $S/\sigma_{\Delta} < 0.80$ достигнут при индивидуальном подборе конфигурации модели под конкретный водосбор.

Предложенный подход применим к нейросетевым моделям произвольной архитектуры без модификации структуры сети.

Список литературы

Бефани, Н.Ф., Калинин, Г.П. (1983) *Упражнения и методические разработки по гидрологическим прогнозам*, Ленинград, Гидрометеиздат, 2-е изд.

Романов, А.В., Акмаев, Э.Р., Семенова, Н.К. (2026) Современные методы прогнозирования водного режима с разной заблаговременностью, *Гидрометеорологические исследования и прогнозы*, № 2 (400), с. 143-168, doi:10.37162/2618-9631-2026-2-143-168.

Ben Taieb, S., Atiya, A.F. (2016) A Bias and Variance Analysis for Multistep-Ahead Time Series Forecasting, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 1, pp. 62-76, doi:10.1109/TNNLS.2015.2411629.

Caljon, D., Vercauteren, J., De Vos, S., Verbeke, W., Van Belle, J. (2025) *Using Dynamic Loss Weighting to Boost Improvements in Forecast Stability*, arXiv:2409.18267.

Chen, S.-A., Li, C.-L., Yoder, N., Arik, S.O., Pfister, T. (2023) TSMixer. An All-MLP Architecture for Time Series Forecasting, arXiv:2303.06053.

Das, A., Kong, W., Leach, A., Mathur, S., Sen, R., Yu, R. (2024) Long-term Forecasting with TiDE. Time-series Dense Encoder, *Transactions on Machine Learning Research*, URL: <https://openreview.net/forum?id=pCbC3aQB5W>.

Fang, K., Kifer, D., Lawson, K., Feng, D., Shen, C. (2022) The Data Synergy Effects of Time-Series Deep Learning Models in Hydrology, *Water Resources Research*, vol. 58, no. 4, p. e2021WR029583, doi:10.1029/2021WR-029583.

Hounie, I., Porras-Valenzuela, J., Ribeiro, A. (2024) Loss Shaping Constraints for Long-Term Time Series Forecasting, *Proceedings of the 41st International Conference on Machine Learning (ICML)*, vol. 235, pp. 19062-19084.

Jadon, A., Patil, A., Jadon, S. (2024) A Comprehensive Survey of Regression-Based Loss Functions for Time Series Forecasting, *Lecture Notes in Networks and Systems*, vol. 998, pp. 117-147, doi:10.1007/978-981-97-3245-6_9.

Kendall, A., Gal, Y., Cipolla, R. (2018) Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7482-7491, doi:10.1109/CVPR.2018.00781.

Kim, T., Kim, J., Tae, Y., Park, C., Choi, J.-H., Choo, J. (2022) Reversible Instance Normalization for Accurate Time-Series Forecasting against Distribution Shift, *International Conference on Learning Representations (ICLR)*. URL: <https://openreview.net/forum?id=cGDAkQo1C0p>.

Kim, T. et al. (2025) A Comprehensive Survey of Deep Learning for Time Series Forecasting: Architectural Diversity and Open Challenges, *Artificial Intelligence Review*, doi:10.1007/s10462-025-11223-9.

Klotz, D., Gauch, M., Kratzert, F., Nearing, G., Zscheischler, J. (2024) Technical Note: The Divide and Measure Nonconformity, how metrics can mislead when we evaluate on different data partitions, *Hydrology and Earth System Sciences*, vol. 28, no. 15, pp. 3665-3673, doi:10.5194/hess-28-3665-2024.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G. (2019) Towards Learning Universal, Regional, and Local Hydrological Behaviors via Machine Learning Applied to Large-Sample Datasets, *Hydrology and Earth System Sciences*, vol. 23, no. 12, pp. 5089-5110, doi:10.5194/hess-23-5089-2019.

Ma, C., Pomazkin, G., Saggese, G.P., Smith, P. (2026) Beyond Accuracy. A Stability-Aware Metric for Multi-Horizon Forecasting, arXiv:2601.10863, Preprint, not peer-reviewed.

Nearing, G.S., Kratzert, F., Sampson, A.K., Pelissier, C.S., Klotz, D., Frame, J.M., Prieto, C., Gupta, H.V. (2021) What Role Does Hydrological Science Play in the Age of Machine Learning, *Water Resources Research*, vol. 57, no. 3, p. e2020WR028091, doi:10.1029/2020WR028091.

Nash, J.E., Sutcliffe, J.V. (1970) River flow forecasting through conceptual models part I. A discussion of principles, *Journal of Hydrology*, vol. 10, no. 3, pp. 282-290, doi:10.1016/0022-1694(70)90255-6.

Sun, F.-K., Lang, C.I., Boning, D.S. (2021) Adjusting for Autocorrelated Errors in Neural Networks for Time Series, *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 34, arXiv:2101.12578.

Tofighi, S., Gurbuz, F., Mantilla, R., Xiao, S. (2025) Advancing Machine Learning-Based Streamflow Prediction Through Event Greedy Selection, Asymmetric Loss Function, and Rainfall Forecasting Uncertainty, *Applied Sciences*, vol. 15, no. 21, p. 11656, doi:10.3390/app152111656.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D. et al. (2020) SciPy 1.0, Fundamental Algorithms for Scientific Computing in Python, *Nature Methods*, vol. 17, no. 3, pp. 261-272, doi:10.1038/s41592-019-0686-2.

Wang, R., Yuan, X., Tian, F., Liu, M., Wang, X., Li, X., Wu, M. (2025) Flood Classification and Improved Loss Function by Combining Deep Learning

Models to Improve Water Level Prediction in a Small Mountain Watershed, *Journal of Flood Risk Management*, vol. 18, no. 2, p. e70022, doi:10.1111/jfr3.70022.

References

Befani, N.F., Kalinin, G.P. (1983) *Uprazhneniya i metodicheskie razrabotki po gidrologicheskim prognozam* [Exercises and methodological guidelines for hydrological forecasting], Leningrad, Russia, Gidrometeoizdat, 2nd ed.

Romanov, A.V., Akmaev, E.R., Semenova, N.K. (2026) Sovremennye metody prognozirovaniya vodnogo rezhima s raznoy zablagovremennost'yu [Modern methods of water regime forecasting with different forecast lead times]. *Gidrometeorologicheskie issledovaniya i prognozy – Hydrometeorological Research and Forecasting*, no. 2 (400), pp. 143-168, doi:10.37162/2618-9631-2026-2-143-168.

Ben Taieb, S., Atiya, A.F. (2016) A Bias and Variance Analysis for Multistep-Ahead Time Series Forecasting, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 1, pp. 62-76, doi:10.1109/TNNLS.2015.2411629.

Caljon, D., Vercauteren, J., De Vos, S., Verbeke, W., Van Belle, J. (2025) *Using Dynamic Loss Weighting to Boost Improvements in Forecast Stability*, arXiv:2409.18267.

Chen, S.-A., Li, C.-L., Yoder, N., Arik, S.O., Pfister, T. (2023) TSMixer. An All-MLP Architecture for Time Series Forecasting, arXiv:2303.06053.

Das, A., Kong, W., Leach, A., Mathur, S., Sen, R., Yu, R. (2024) Long-term Forecasting with TiDE. Time-series Dense Encoder, *Transactions on Machine Learning Research*, URL: <https://openreview.net/forum?id=pCbC3aQB5W>.

Fang, K., Kifer, D., Lawson, K., Feng, D., Shen, C. (2022) The Data Synergy Effects of Time-Series Deep Learning Models in Hydrology, *Water Resources Research*, vol. 58, no. 4, p. e2021WR029583, doi:10.1029/2021WR-029583.

Hounie, I., Porras-Valenzuela, J., Ribeiro, A. (2024) Loss Shaping Constraints for Long-Term Time Series Forecasting, *Proceedings of the 41st International Conference on Machine Learning (ICML)*, vol. 235, pp. 19062-19084.

Jadon, A., Patil, A., Jadon, S. (2024) A Comprehensive Survey of Regression-Based Loss Functions for Time Series Forecasting, *Lecture Notes in Networks and Systems*, vol. 998, pp. 117-147, doi:10.1007/978-981-97-3245-6_9.

Kendall, A., Gal, Y., Cipolla, R. (2018) Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7482-7491, doi:10.1109/CVPR.2018.00781.

Kim, T., Kim, J., Tae, Y., Park, C., Choi, J.-H., Choo, J. (2022) Reversible Instance Normalization for Accurate Time-Series Forecasting against Distribution

Shift, *International Conference on Learning Representations (ICLR)*. URL: <https://openreview.net/forum?id=cGDAkQo1C0p>.

Kim, T. et al. (2025) A Comprehensive Survey of Deep Learning for Time Series Forecasting: Architectural Diversity and Open Challenges, *Artificial Intelligence Review*, doi:10.1007/s10462-025-11223-9.

Klotz, D., Gauch, M., Kratzert, F., Nearing, G., Zscheischler, J. (2024) Technical Note: The Divide and Measure Nonconformity, how metrics can mislead when we evaluate on different data partitions, *Hydrology and Earth System Sciences*, vol. 28, no. 15, pp. 3665-3673, doi:10.5194/hess-28-3665-2024.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G. (2019) Towards Learning Universal, Regional, and Local Hydrological Behaviors via Machine Learning Applied to Large-Sample Datasets, *Hydrology and Earth System Sciences*, vol. 23, no. 12, pp. 5089-5110, doi:10.5194/hess-23-5089-2019.

Ma, C., Pomazkin, G., Saggese, G.P., Smith, P. (2026) Beyond Accuracy. A Stability-Aware Metric for Multi-Horizon Forecasting, arXiv:2601.10863, Preprint, not peer-reviewed.

Nearing, G.S., Kratzert, F., Sampson, A.K., Pelissier, C.S., Klotz, D., Frame, J.M., Prieto, C., Gupta, H.V. (2021) What Role Does Hydrological Science Play in the Age of Machine Learning, *Water Resources Research*, vol. 57, no. 3, p. e2020WR028091, doi:10.1029/2020WR028091.

Nash, J.E., Sutcliffe, J.V. (1970) River flow forecasting through conceptual models part I. A discussion of principles, *Journal of Hydrology*, vol. 10, no. 3, pp. 282-290, doi:10.1016/0022-1694(70)90255-6.

Sun, F.-K., Lang, C.I., Boning, D.S. (2021) Adjusting for Autocorrelated Errors in Neural Networks for Time Series, *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 34, arXiv:2101.12578.

Tofighi, S., Gurbuz, F., Mantilla, R., Xiao, S. (2025) Advancing Machine Learning-Based Streamflow Prediction Through Event Greedy Selection, Asymmetric Loss Function, and Rainfall Forecasting Uncertainty, *Applied Sciences*, vol. 15, no. 21, p. 11656, doi:10.3390/app152111656.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D. et al. (2020) SciPy 1.0, Fundamental Algorithms for Scientific Computing in Python, *Nature Methods*, vol. 17, no. 3, pp. 261-272, doi:10.1038/s41592-019-0686-2.

Wang, R., Yuan, X., Tian, F., Liu, M., Wang, X., Li, X., Wu, M. (2025) Flood Classification and Improved Loss Function by Combining Deep Learning Models to Improve Water Level Prediction in a Small Mountain Watershed, *Journal of Flood Risk Management*, vol. 18, no. 2, p. e70022, doi:10.1111/jfr3.70022.

Статья поступила в редакцию (Received): 15.05.2026.

Статья доработана после рецензирования (Revised): 25.05.2026.

Принята к публикации (Accepted): 27.05.2026.

Для цитирования / For citation

Акмаев, Э.Р. (2026) Влияние выбора функции потерь при обучении модели на качество прогноза уровней воды, *Фундаментальная и прикладная климатология*, т. 12, № 3, с. 305-333, doi:10.24412/2410-8758-2026-3-305-333.

Akmaev, E.R. (2026) The effect of loss function choice on model training quality in water level forecasts, *Fundamental and Applied Climatology*, vol. 12, no. 3, pp. 305-333, doi:10.24412/2410-8758-2026-3-305-333.