

МЕТОДЫ, МОДЕЛИ И ТЕХНОЛОГИИ METHODS, MODELS AND TECHNOLOGIES

УДК 556.06

DOI: 10.34753/HS.2023.5.3.228

ОЦЕНКА КАЧЕСТВА МОДЕЛИРОВАНИЯ И ПРОГНОЗИРОВАНИЯ В ГИДРОЛОГИИ – РАССУЖДЕНИЕ О МЕТОДЕ

Б.И. Гарцман

*Институт водных проблем РАН, г.Москва,
Россия*gartsman@inbox.ru

ASSESSING THE QUALITY OF MODELING AND FORECASTING IN HYDROLOGY – REASONING ABOUT THE METHOD

Boris I. Gartsman

Institute of Water Problems RAS, Moscow, Russiagartsman@inbox.ru

Аннотация. Существующие в настоящее время гидрологические модели обеспечены лишь крайней ограниченными методическими средствами их калибровки, верификации и тестирования. Крайне актуальной представляется разработка на единой методической основе системы оценок качества при решении ряда типовых задач гидрологии – параметризации и верификации гидрологических моделей, имитационного и сценарного моделирования гидрологического режима, прогнозирования, инженерных расчетов. Без развития системы оценок качества дальнейшее развитие и эффективное практическое использование гидрологических моделей невозможны.

В работе показана логика и, отчасти, история развития некоторых важнейших идей и подходов, применяемых в оценивании качества моделирования и прогнозирования, представлены направления их развития. Выполнена необходимая систематизация терминов рассматриваемой предметной области, начиная с базовых понятий «предсказание» и «прогноз», с рассмотрением дополнительных дефиниций: «непрерывные – агрегированные» прогнозы; «краткосрочные – долгосрочные» прогнозы; «методический», «стандартный», «идеальный», «инерционный», «режимный» прогнозы; и др.

Рассмотрены этапы развития подхода «оценки умений» (skill scores) в разработке мер

Abstract. The currently existing hydrological models are provided with extremely limited methodological tools for their calibration, verification, and testing. It seems extremely relevant to develop, on a unified methodological basis, a system of quality assessment for a number of typical hydrological tasks – parameterization and verification of hydrological models, simulation and scenario modeling of the hydrological regime, forecasting, engineering calculations, etc. Without solving this problem, further development and effective practical use of hydrological models are impossible.

The paper demonstrates the logic and, to some extent, the history of the development of some of the most important ideas and approaches used in assessing the quality of modeling and forecasting, as well as to outline some prospects for their development. The necessary systematization of the terms of the subject area under consideration has been carried out, starting with the basic concepts of “prediction” and “forecast”, with consideration of additional definitions: “continuous – aggregated” forecasts; “short-term – long-term” forecasts; “methodological”, “standard”, “ideal”, “inertial”, “regime” forecasts; and etc.

The stages of development of the “skill scores” approach in the development of measures for assessing the quality of hydrological modeling and forecasting are considered, mainly using the example of the emergence and modification of a

оценки качества гидрологического моделирования и прогнозирования, в основном на примере появления и модификации базовой для гидрологии меры – коэффициента эффективности Нэша-Сатклиффа. Особое внимание уделено двум разложениям меры NSE, показывающим её сводимость к базовым статистическим параметрам выборок измеренных и прогнозных значений переменной: средним значениям, среднеквадратическим отклонениям и коэффициенту корреляции. Предложено обоснование модификации оценки NSE путем применения её к предварительно ранжированным выборкам значений переменной. Сформулирован подход к оцениванию качества моделирования и прогнозирования на основе последовательно статистического подхода, с использованием статистических критериев сходства маргинальных распределений фактических и прогностических значений переменной. Намечен ряд задач перспективных исследований в области развития методов оценки качества гидрологических моделей и прогнозов.

Ключевые слова: модели стока; гидрологические прогнозы; оценка качества прогнозирования; коэффициент эффективности Нэша-Сатклиффа.

Введение

В настоящее время существует ярко выраженный диссонанс между многочисленностью, разнообразием и детальностью гидрологических и гидравлических моделей, и крайней ограниченностью методических средств их верификации и тестирования [Beven, 2001]. Ключевой научной проблемой представляется получение на единой методической основе оценок качества решения ряда типовых задач гидрологии – параметризации и верификации гидрологических моделей,

basic measure for hydrology - the Nash-Sutcliffe efficiency coefficient. Special attention is paid to two decompositions of the NSE measure, demonstrating its reducibility to the basic statistical parameters of the samples of measured and predicted values of the variable – average values, standard deviations and correlation coefficient. A rationale for modifying the NSE estimate by applying it to pre-ranked samples of variable values is proposed. An approach to assessing the quality of modeling and forecasting is formulated based on a consistently statistical approach, using statistical criteria for the similarity of marginal distributions of actual and forecast values of a variable. A number of tasks for future research are outlined in the field of developing methods for assessing the quality of hydrological models and forecasts.

Keywords: runoff models; hydrological forecasts; forecast quality assessment; Nash-Sutcliffe efficiency coefficient.

Рассуждение о методе, чтобы хорошо направлять свой разум и отыскивать истину в науках.

Рене Декарт (Сер. «Классики науки». Изд-во АН СССР, 1953)

имитационного и сценарного моделирования гидрологического режима, прогнозирования, инженерных расчетов и т.п. Недостаток методической проработки в этой части – основной фактор, сдерживающий развитие и практическое использование гидрологических моделей.

Поле для исследований по затронутой проблематике чрезвычайно обширно, один только анализ литературных источников по ней мог бы являться предметом большой обзорной

статьи (см. напр. [Аполлов, Калинин, Комаров, 1974; Борщ, Христофоров, 2015; Борщ, Симонов, Христофоров, 2020; Борщ, Симонов, Христофоров, 2023; Рождественский, Чеботарев, 1974; Руководство... , 1994; Руководство... , 1989; Технический регламент... , 2006; Bogner, Pappenberger, 2011; Jolliffe, Stephenson, 2012; Wilks, 1995]). Данной статья, написанная в необычном жанре рассуждения, отражает личные взгляды автора на проблему и основана на многолетнем опыте работы в области гидрологического моделирования и прогнозирования, включающем и знакомство с профильной литературой, небольшая часть которой перечислена выше. Целью является проследить логику и, отчасти, историю развития некоторых важнейших идей и подходов, применяемых в оценивании качества моделирования и прогнозирования, а также наметить некоторые перспективы их развития.

Определим ряд необходимых терминов.

Предсказание (*prediction*) – оценка некоторых характеристик гидрологических процессов или явлений в условиях ограниченности имеющихся данных. Причинами ограниченности данных могут быть направленность предсказания в будущее, когда данные еще не появились, в далекое прошлое, когда данные еще отсутствуют в принципе, либо недостаточность данных существующей сети наблюдений. Предсказание, обращенное в определенное будущее, называется **прогнозом** (*forecasting*). Отличие прогноза от предсказания заключается в точной привязке прогнозируемых величин во времени.

Характерным свойством прогноза, в силу его направленности в обозримый интервал будущего, является ожидаемое появление рядов данных инструментальных наблюдений, которые позволят выполнить независимую, корректную и полную оценку качества прогноза. При других видах предсказаний новые данные для оценки их качества либо не ожидаются вообще, либо они заведомо не будут удовлетворять критериям независимости, корректности и полноты оценки.

Инструментом предсказания всегда является моделирование. Обычно понятие модели используется для обозначения формализованного

математического алгоритма, имеющего теоретическое концептуальное обоснование, но модели могут быть также эмпирическими, экспертными, вербальными и т.д. На практике, конкретное предсказание или прогноз всегда опирается на рациональное сочетание формализованного и экспертного знания, при использовании конкретных эмпирических данных.

Применяемый при всех видах предсказания методический инструментарий в принципе единообразен, в том числе обязательно подразумевают наличие оценок качества предсказаний – точности, надежности, оправдываемости и т.п. Такие типовые задачи гидрологии, как параметризация и верификация (калибровка) математических моделей, имитационное и сценарное моделирование гидрологического режима, прогнозирование различного типа и заблаговременности, инженерные расчеты и т.п., концептуально и методически тесно связаны, что предполагает применение таких же взаимоувязанных оценок качества для решения всех подобных задач.

В дальнейшем по отношению к значениям гидрологических переменных термины «расчетные», «модельные», «предсказанные», «прогнозные» будут употребляться как синонимы – в оппозиции к «фактическим», «измеренным» и «наблюденным» значениям. Это делается с полным пониманием того, что термины внутри упомянутых двух групп имеют существенные смысловые различия, и вполне синонимами их можно считать именно в рамках указанного противопоставления, которое является сутью всех оценок качества предсказания.

Гидрологический прогноз – предсказание с различной заблаговременностью и степенью точности элемента гидрологического режима или явления, основанное на знании закономерностей развития гидрометеорологических процессов, определяющих это явление в конкретных условиях данной реки, озера или водохранилища. Гидрологические прогнозы имеют основные характеристики:

– форма представления прогноза (единственное значение, диапазон значений, распределение значений во времени или по вероятности).

– переменная прогноза, выражающая предсказываемый гидрологический элемент;

– заблаговременность прогноза, либо период прогноза или предупреждения;

– оценка качества прогноза;

– целевое назначение прогноза;

С точки зрения цели и формы выдачи существует несколько видов прогнозов, отличающихся характером оценивания их качества.

Непрерывный (количественный) прогноз – прогноз последовательности значений непрерывной гидрологической переменной с временным интервалом, равным или близким к интервалу измерений (почасовой, посуточный, пентадный).

Агрегированный (количественный) прогноз – прогноз агрегированных значений непрерывной гидрологической переменной для временных интервалов, значительно больших, чем интервал измерений (суммарные/средние/максимальные значения, а также кривые продолжительности расходов и т.п., за декаду, месяц, сезон, год).

Прогноз событий (качественный) – прогноз наличия/отсутствия некоторого события, как правило экстремального или опасного: превышения критического уровня воды, прохождения быстроразвивающегося паводка, селя, лавины. Поскольку прогнозируется наступление события или его отсутствие (бинарный выбор), такие прогнозы также называются **бинарными**. Поскольку такие события обычно бывают редкими, употребляется также термин **прогноз редких событий**. Промежуточной формой можно считать **прогноз по категориям**, когда оценивается попадание значения непрерывной или агрегированной переменной в заданный интервал – в этом случае выбор осуществляется, обычно, из нескольких альтернатив.

Вероятностный прогноз – прогноз любого из перечисленных выше типов, выдаваемый не в форме единичного значения, а в форме ожидаемого распределения вероятностей величины. Таким образом, например, агрегированный прогноз может быть в то же время и вероятностным.

Аналогично прогнозам, такие формы предсказания, как имитационное моделирование и сценарные расчеты, также могут подразделяться на непрерывные, агрегированные, событийные (категориальные) – с аналогичными особенностями процедур оценки их качества. В отношении специфики оценивания качества прогнозирования, преимущественно для непрерывных прогнозов, целесообразно выделять две категории – краткосрочных и долгосрочных прогнозов.

Краткосрочный прогноз – прогноз будущего значения гидрологической переменной на период, сопоставимый с характерным временем реакции исследуемой системы. В таком случае прогноз сильно зависит от состояния системы на момент выпуска прогноза, и основан в большей степени на учете инерции внутренних процессов в системе («память процесса»). Ниже будет показано, что в этом случае обычно прогнозируется тенденция изменения переменной.

Долгосрочный прогноз – прогноз будущего значения гидрологической переменной на период, существенно превышающий характерное время реакции системы. Долгосрочный прогноз опирается не на «память процесса», а на инерцию или закономерности внешних воздействий, и прогнозируется при этом само значение переменной.

Оценка качества прогнозов (предсказаний) – установление количественных характеристик связи между соответствующими выборками значений прогнозов и фактических наблюдений, рассматриваемых как случайные величины, и соотнесение их с заранее установленными градациями качества такой связи. Подчеркнем, что мы будем рассматривать именно количественные параметры, полученных на основе статистического анализа однородных прогнозных данных и данных инструментальных наблюдений стандартного качества, доступных в объеме, позволяющем выполнить корректную, полную и, по возможности, независимую оценку качества прогноза.

Кроме собственно прогнозов, оценки качества в строгом смысле возможны при решении различных задач имитационного моделирования, а также при калибровке

моделей – в этом случае предметом предсказания являются параметры модели. При таких видах предсказания, как сценарные расчеты, палеорекострукции или реконструкции событий по следам, верификация неизбежно может опираться лишь на данные неполные, неоднородные, косвенные, и т.п. Анализ качества при этом всегда содержит элемент экспертного оценивания, не гарантирующие в этом случае корректности и полноты анализа.

Метод (методика) прогноза – совокупность информационных, расчетных, оценочных и других алгоритмов, специально разработанная для получения прогностической информации об интересующей переменной с максимальной точностью и заблаговременностью. Методика прогноза всегда представляет собой реализацию некоторой модели предсказываемых процессов, даже если она (модель) не сформулирована в строгой математической форме.

Методический прогноз – прогноз, выполняемый по некоторой методике. Термин употребляется преимущественно при оценках качества прогнозов в оппозиции к **стандартному прогнозу**. Под стандартным прогнозом подразумевается предсказание будущего значения переменной, получаемое в результате прямого анализа имеющихся фактических данных без специальной методики, либо с применением некоторой «наихудшей» (наименее информативной) методики прогноза. В качестве стандартного прогноза чаще всего выбирается инерционный (для краткосрочных прогнозов) или режимный (для долгосрочных прогнозов), но в каждом конкретном случае возможны варианты, которые следует обосновать специальным анализом. Общим принципом оценки качества методического прогноза является сопоставление некоторой меры его качества с аналогичными мерами, вычисленными для стандартного (наихудшего) прогноза и для идеального прогноза. Такой подход определяется термином *skill scores*, что на русский язык переводится как «оценка навыков» или «умений».

Идеальный прогноз – условный вариант абсолютно точного прогнозирования фактических значений. В англоязычной литературе термин часто используется в форме

perfect forecast, то есть, дословно «совершенный прогноз». Являясь недостижимой абстракцией, такой прогноз используется при оценках качества прогнозирования в виде ряда измеренных данных.

Инерционный прогноз – условный вариант прогнозирования, когда текущее значение величины предполагается неизменным на будущее в течение периода заблаговременности. Возможны варианты инерционного прогноза, например, предполагающие неизменность наблюдаемого изменения величины в будущем на период заблаговременности. Используется в качестве стандартного («наихудшего») прогноза при оценивании для сравнения с методическим прогнозом.

Режимный прогноз – условный вариант прогнозирования, когда значение переменной на прогнозируемый период предполагается равным его средней многолетней величине (норме) на тот же период в течение года. Возможны варианты режимного прогноза, например, учитывающие многолетние тренды в рядах значений гидрологических величин или явлений. Также как инерционный прогноз, используется в качестве стандартного («наихудшего») варианта при оценивании для сравнения с методическим прогнозом.

Меры качества предсказания в рамках подхода «оценки умений»

Мера Нэша-Сатклиффа

Базовая мера оценки качества предсказания в гидрологии называется Nash-Sutcliffe efficiency (NSE) – коэффициент эффективности Нэша-Сатклиффа [Nash, Sutcliffe, 1970]. Эта простая по форме мера приобрела исключительное значение, являясь основным и практически обязательным (а часто единственным) инструментом верификации гидрологических моделей, что отражено в десятках тысяч научных публикаций и неуставимом числе научно-технических отчетов. Первичная формулировка NSE в авторских обозначениях, с некоторым упрощением, выглядит следующим образом:

$$R^2 = \frac{F_o^2 - F_s^2}{F_o^2}, \quad \text{причем} \\ F_s^2 = \sum (q_s - q_o)^2 \text{ и } F_o^2 = \sum (q_o - q_a)^2. \quad (1)$$

Здесь оценка NSE обозначена как R^2 , исследуемая переменная обозначается как q , q_s – рассчитанные по модели значения переменной, q_o – фактические (измеренные) значения переменной, q_a – среднее из фактических значений переменной. F_s^2 , как сумма отклонений рассчитанных значений от фактических по выборке, представляет собой остаточную вариацию (residual variance) модели, а F_o^2 обозначает исходную, либо полную, вариацию (initial variance) моделируемой переменной, в предложенной формулировке равную её дисперсии. Разность этих двух величин представляет собой объясненную моделью вариацию переменной, а оценка качества R^2 равна отношению объясненной вариации к полной (доля объясненной вариации). Такая логика имеет полную аналогию применяемой в регрессионном анализе.

В более детальной и привычной записи, применимой для стандартной задачи имитационного моделирования гидрографа стока, оценка NSE выглядит следующим образом

$$R^2 = 1 - \frac{\sum_{i=1}^n (Qs_i - Qo_i)^2}{\sum_{i=1}^n (Qo_i - Qa)^2} \quad (2)$$

Здесь Qs_i и Qo_i – расчетный (предсказанный) и фактический (измеренный) расходы воды на i -тый момент времени, Qa – среднее значение фактических расходов по выборке, n – объем выборки.

Следует отметить, что с самого начала NSE формулировалась в рамках идеологии «оценки умений». В формуле (2) «оценка умений» модели основана на сопоставлении среднеквадратических погрешностей методического и стандартного прогнозов (предсказаний).

Ясно, что при этом адекватность оценки критически зависит от выбора стандартного прогноза. Если стандартным прогнозом, как в формуле (2), является среднее значение переменной, то его среднеквадратическая погрешность равна дисперсии переменной. Это самый простой, но и самый слабый вариант выбора стандарта для сравнения, приводящий к огрублению оценки и часто – к неоправданному завышению качества прогноза. Очевидно, что получить предсказание расхода, существенно

лучшее, чем его средняя величина, можно на основе даже самых простых моделей. Разумной альтернативой выглядит использование в качестве стандартного прогноза режимных данных о стоке, это приводит к оценке NSE в форме

$$R^2 = 1 - \frac{\sum_{i=1}^n (Qs_i - Qo_i)^2}{\sum_{i=1}^n (Qo_i - Qa_i)^2} \quad (3)$$

где Qa_i означает уже среднемноголетнее значение расхода на i -й день в году. Для рек с устойчивым режимом такая оценка качества предсказания будет значительно более жесткой, учитывающей воспроизведение не столько общего годового хода стока, сколько особенностей и нарушений такого хода. Это, по сути, означает, что для рек с устойчивым режимом сами режимные данные могут быть использованы в качестве довольно эффективного предсказания. Аналогичный подход широко используется в метеорологии, так как высокая устойчивость годового цикла погоды делает целесообразным любую обработку метеорологических данных, в том числе оценки качества прогнозов, выполнять «в аномалиях», т.е. в значениях отклонения от «климата».

На рисунке 1 схематически представлены оценки для вариантов NSE , выраженных формулами 2 и 3, при предсказании гидрографа одного года. В первом случае $NSE = 0,90$, т.е. доля объясненной моделью дисперсии процесса по отношению к его среднему значению 90 %. Во втором случае $NSE = 0,33$, т.е. доля объясненной той же самой моделью дисперсии процесса относительно его режима – 33 %. Для рек с неустойчивым (паводковым) режимом использование режимных данных в качестве стандартного прогноза не дает такого эффекта, а в отдельных, особо сложных, случаях может даже дополнительно огрублять оценку NSE .

Меры оценки качества прогнозов, используемые в России

Вопросы оценки качества прогнозов детально рассмотрены в базовом пособии Росгидромета [Наставление..., 1962]. В качестве основной используется мера качества, называемая на профессиональном жаргоне «эс-к-сигма» (S/σ) – отношение среднеквадратической

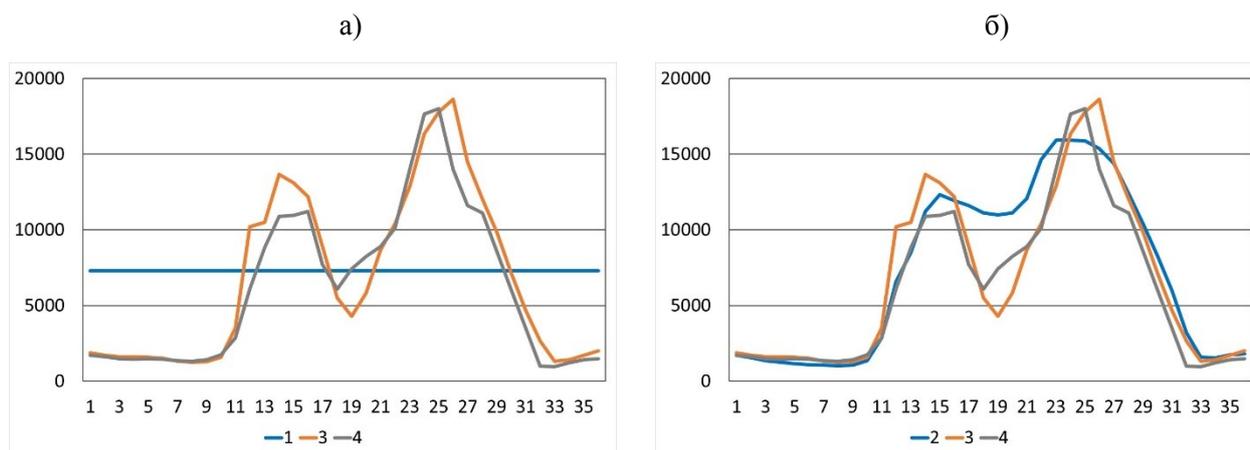


Рисунок 1. Условные примеры вычисления NSE при разных вариантах принятого стандартного прогноза: а) среднее значение за период, $R^2=0,90$; б) режимные данные – среднесуточные величины на каждые сутки, $R^2=0,33$: 1 – среднее, 2 – режим, 3 – прогноз, 4 – факт

Figure 1. Conditional examples of calculating NSE for different versions of the accepted standard forecast: а) average value for the period, $R^2 = 0.90$; б) regime data – long-term average values for every day, $R^2=0.33$: 1 – mean, 2 – regime, 3 – forecast, 4 – fact

ошибки прогноза S к среднеквадратическому отклонению прогнозируемой переменной σ . При этом, в обозначениях упомянутого пособия,

$$S = \sqrt{\frac{\sum_{i=1}^n (Q_i^p - Q_i)^2}{n-m}} \quad (4)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (Q_i - \bar{Q})^2}{n-1}} \quad (5)$$

Здесь Q_i^p и Q_i – прогнозный и фактический расходы на i -тый момент времени, \bar{Q} – среднее значение фактических расходов по выборке, n – объем выборки.

Вычитание из n в знаменателе этих выражений числа параметров прогностической модели m и единицы представляет собой введение поправок в оценки, учитывающие ограниченный объем выборок, причем только для случая проверки методики прогноза на зависимом материале, который использовался при оценке параметров модели. В любом случае, при больших выборках этими поправками можно пренебречь. Легко увидеть, что отношение S/σ представляет собой точный эквивалент дробного члена выражения для R^2 (формула 2), отличаясь лишь извлечением корня. Для оценки качества прогнозов используется также корреляционное отношение

$$\rho = \sqrt{1 - \frac{S^2}{\sigma^2}} \quad (6)$$

которое равно, и означает то же самое, что и R (не в квадрате).

Не употребляя термин «skill scores», разработчики российской методики придерживались и активно развивали сам этот подход. Так, при краткосрочном прогнозе оценивается прогнозирование не абсолютной величины расхода, а его изменения по сравнению с текущим состоянием. При этом среднеквадратическое отклонение изменения прогнозируемой переменной обозначается как σ_{Δ} и вычисляется по формуле

$$\sigma_{\Delta} = \sqrt{\frac{\sum_{i=1}^n (\Delta Q_i - \overline{\Delta Q})^2}{n-1}} \quad (7)$$

Здесь ΔQ_i и $\overline{\Delta Q}$ – изменения прогнозируемой переменной за период заблаговременности прогноза, на i -тый момент времени и их средняя величина, соответственно.

Эта мера отвечает случаю принятия в качестве стандарта инерционного прогноза – расход на будущий момент принимается равным его значению на момент выпуска прогноза. В упомянутом Наставлении даются и более детальные рекомендации по модификации меры качества S/σ в зависимости от особенностей прогнозируемой величины – подъема уровней от различных причин, объемов стока, сроков наступления определенных событий и других.

Таким образом, в изданном в 1962 г. базовом российском нормативном документе установлена развитая система оценок качества прогнозов различных гидрологических параметров, по концепции и основным математическим формулировкам полностью совпадающая с оценкой NSE и рядом ее модификаций, предложенных в 1970–80-е и последующие годы. По-видимому, российская гидрология имеет твердый приоритет в разработке методов такого оценивания, международному признанию которого помешали лишь определенные исторические обстоятельства.

Анализ Аллана Мерфи

Глубокая и систематическая разработка темы мер оценки качества моделирования и прогнозирования выполнялась в близкой гидрологии дисциплине, где физико-математические модели стали основным инструментом работы гораздо раньше, а значение прогнозов разного рода, как основной продукции, гораздо выше, – в синоптической метеорологии. Ярким примером этого являются публикации Аллана Мерфи, где формулируются принципы, раскрываются смысл и взаимосвязи подобных мер [Murphy, Winkler, 1987; Murphy 1988]. Изложим кратко некоторые его результаты, прямо относящиеся к теме нашего анализа. Прежде всего, выдвигается базовый принцип – вся информация о качестве и эффективности прогнозов заключена в двумерном совместном распределении вероятностей прогностических и фактических значений переменной, на основе которого строятся любые соответствующие оценки и анализ.

Далее, Мерфи тщательно формулирует сам подход «оценки умений» (skill scores) прогнозной модели, основанной на отношении меры расхождений фактических значений переменной и прогностических (по анализируемой методике) к аналогичной мере расхождений фактических и оценочных значений переменной, получаемых гарантированно по некоторой базовой процедуре, которая нами выше названа стандартным прогнозом. Таким образом, мера skill scores прогностической методики заключается в оценке её места в диапазоне между идеальным (perfect)

прогнозом, значения которого точно совпадают с фактическими, и «наихудшим» (standard of reference), выбор которого обуславливает ширину диапазона и чувствительность оценки.

В математической записи, по Мерфи, это выражается следующим образом

$$S = (A_f - A_r) / (A_p - A_r), \quad (8)$$

где S – мера «оценки умений», A_f , A_p и A_r – меры точности методического, идеального и стандартного прогнозов, соответственно. Важно подчеркнуть, что в данной формуле для оценки умений могут быть использованы самые различные формы меры A , важно лишь чтобы при каждом оценивании во всех трех случаях это была бы одна и та же форма. Если, следуя традиции, использовать меру среднеквадратического отклонения (MSE) в обозначениях Мерфи

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - x_i)^2, \quad (9)$$

то после подстановки уравнения (9) в уравнение (8) получим тот же коэффициент эффективности Нэша-Сатклиффа (NSE). Здесь прогностические и фактические значения переменной обозначены как f и x , соответственно, i и n описаны выше. Использование любой другой меры вместо MSE приведет к иной форме конечной оценки качества прогнозирования.

Вторым важным моментом анализа Мерфи является выделение четырех основных вариантов стандартного прогноза, существенно влияющих как на величину, так и на смысл получаемой оценки.

1. Среднее из фактических значений переменных (x_a), рассчитанное для всего интервала, на котором выполняется моделирование либо испытания прогнозов (internal). Именно в этом случае получаем оценку, точно совпадающую с классической формой NSE.

2. Выборка средних из фактических значений на каждую календарную дату ($x_{a,i}$), рассчитанных в пределах интервала, на котором выполняется моделирование либо испытания прогнозов (internal).

3. Среднее из фактических значений переменных, рассчитанное для многолетнего интервала, значительно превышающего тот, на котором выполняется моделирование либо испытания прогнозов (external).

4. Выборка средних из фактических значений на каждую календарную дату, рассчитанных в пределах многолетнего интервала, значительно превышающего тот, на котором выполняется моделирование либо испытания прогнозов (external).

Все варианты строятся на обобщении климатологических (в нашем случае – режимных гидрологических) данных.

Все эти варианты оценок хорошо известны специалистам, занимающимся разработкой прогностических моделей или имитационным, например, сценарным, моделированием. Первый вариант дает наиболее оптимистичную (высокую) оценку качества прогнозирования, а далее по последовательности оценки становятся более жесткими. Варианты 1 и 2 стандартного прогноза (internal) более употребительны для краткосрочного прогнозирования, для долгосрочного применяются обычно стандартная оценка по многолетнему ряду. Применение режимных данных по календарным датам ($x_{a,i}$) оправдано, как упоминалось выше, лишь в случае высокой устойчивости гидрологического режима. В целом, выбор оптимальной процедуры оценки качества прогноза представляет собой нетривиальную задачу, обусловленную как свойствами прогнозируемых величин, так и порядком, и целью прогнозирования.

Третий важный момент – предложенное Мерфи разложение оценки NSE на составляющие, наглядно представляющее смысл этого показателя. Ограничиваясь только первым из перечисленных выше 4 вариантов и придерживаясь обозначений Мерфи с минимальными коррективами, полученный им результат выглядит следующим образом

$$SS(f, \bar{x}, x) = r_{fx}^2 - \left(r_{fx} - \frac{\sigma_f}{\sigma_x} \right) - \left(\frac{\bar{f} - \bar{x}}{\sigma_x} \right)^2. \quad (10)$$

Здесь \bar{f} и \bar{x} – средние на интервале поверки прогноза значения прогнозных и фактических величин, σ_f и σ_x – величины их среднеквадратических отклонений, r_{fx} коэффициент парной линейной корреляции между ними, $SS(f, \bar{x}, x)$ обозначает «оценку умений» S , полученную с использованием MSE в соответствии с вариантом 1 – с принятием \bar{x} в качестве стандартного прогноза. Таким образом,

$SS(f, \bar{x}, x)$ обозначает ту же оценку, которую выше мы обозначали как R^2 .

Видим, что оценка NSE складывается из трех членов, первый из которых – квадрат коэффициента корреляции, отражающий тесноту связи прогнозных и фактических значений. Второй член содержит разность коэффициента корреляции, характеризующего наклон связи $\{f, x\}$, и отношение среднеквадратических отклонений прогнозных и фактических значений, соответствующее наклону той же связи при условии её однозначности, т.е. при $r=1$. Этот член стремится к нулю при приближении связи $\{f, x\}$ к линейной однозначной, т.е. характеризует условную (зависящую от его величины) систематическую ошибку прогноза. Третий член уравнения (10) характеризует абсолютную систематическую ошибку прогноза.

Таким образом, парный коэффициент корреляции (его квадрат) составляет суть и положительную часть оценки NSE, которая не может его превысить. Это дает основание рассматривать его как потенциальную (предельно возможную) меру «оценки умений». Два других члена связаны с систематическими погрешностями прогноза и могут только ухудшить итоговую оценку по сравнению с потенциальной. При разложении оценки SS для трех остальных вариантов из выше перечисленных, Мэрфи получает более сложные выражения с добавочными членами, рассмотрение которых выходит за рамки нашего анализа.

Анализ Гупты с соавторами

Дальнейшее развитие анализа показателя NSE излагается в основном, следуя этапной работе Хошина Гупты с тремя соавторами [Gupta et al., 2009]. В указанной работе авторы предлагают свое разложение той же простейшей формы NSE, альтернативное разложению Мерфи (уравнение 10). В авторских обозначениях оно выглядит как

$$NSE = 2 \cdot \alpha \cdot r - \alpha^2 - \beta_n^2, \quad (11)$$

где $\alpha = \sigma_s / \sigma_o$, $\beta_n = (\mu_s - \mu_o) / \sigma_o$.

Здесь μ_s и μ_o – средние значения моделированных и наблюдаемых значений переменной; σ_s и σ_o – их же среднеквадратические

отклонения; остальные обозначения определены выше.

В приведенном разложении также имеются три члена, первый из которых содержит коэффициент корреляции, то есть отвечает за совпадение формы прогнозного и фактического гидрографов. Второй член содержит отношение дисперсий прогнозных и фактических величин, а третий – разность их средних значений, нормированную на дисперсию фактических значений переменной и точно соответствует третьему члену разложения Мерфи.

В этой связи отмечается два существенных недостатка оценки NSE при использовании в калибровке моделей. Первый – то, что β_n содержит разность средних значений модельных и измеренных величин, нормированную на дисперсию измеренных. Согласно [Gupta et al., 2009], это может приводить к недооценке систематической ошибки прогноза при калибровке. По нашему мнению, это неосновательно, т.к. все члены уравнения (11) одинаково нормированы на дисперсию, и это естественно при использовании квадратичных мер расхождения. Важно также, что β_n входит в состав оценки NSE, минимизирующей сумму квадратов расхождений модельных и фактических значений, то есть систематическая ошибка среднего учитывается оптимальным образом для достижения этого результата.

Второй, действительно существенный, недостаток применения NSE в качестве меры расхождения при калибровке модели заключается в систематическом занижении дисперсии серии модельных значений переменной. Это становится очевидным, если взять в уравнении (11) первую производную от NSE по α – она имеет максимум при $\alpha=1$, а поскольку при реальном моделировании всегда $r < 1$, то по результатам калибровки всегда будет $\sigma_s < \sigma_o$. Эффект занижения дисперсии моделированных серий хорошо знаком специалистам, работающих с разными гидрологическими моделями, но его причина, которая заключается не в свойствах моделей, а в применяемой при калибровке мере расхождения, неочевидна. Таким образом, из уравнения (11) понятно, что максимального значения, равного r^2 , NSE достигает при условиях

$\alpha=r, \beta_n=0$ и $\sigma_s \cong \sigma_o$ будет лишь при $r \cong 1$, т.е. при приближении к идеальному прогнозу.

Для преодоления этого недостатка авторы [Gupta et al., 2009] предлагают два подхода. Возможно при калибровке модели вводить поправки к фактическим значениям переменной, которые приводили бы к получению максимума NSE при $\sigma_s^* \cong \sigma_o$, в форме

$$x_{o,t}^* = c(x_{o,t} - \mu_o) + \mu_o, \quad (12)$$

$$NSE = \frac{1}{c} \cdot 2 \cdot \alpha \cdot r - \frac{1}{c^2} \cdot \alpha^2 - \frac{1}{c^2} \cdot \beta_n^2. \quad (13)$$

Здесь c – поправка, причем из выражения (13) ясно, что при $c=1/r$ максимум NSE достигается при $\alpha=1$. Фактически, поправка c выступает как еще один параметр при калибровке.

Второй подход основан на использовании вместо NSE искусственно сконструированной меры KGE (Kling–Gupta efficiency). KGE, также как и NSE, включает три члена, представляющих отдельно коэффициент корреляции модельных и измеренных значений переменной, соотношения их среднеквадратических отклонений и средних величин, и также стремится к 1 при идеальном прогнозе:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}, \quad (14)$$

где $\beta = \mu_s/\mu_o$.

Оба подхода отличаются некоторой искусственностью, ставящей под сомнение оптимальность предлагаемых мер расхождения. Следует еще раз подчеркнуть, что форма оценки NSE, при использовании её в калибровке моделей, обеспечивает минимизацию суммы квадратов разностей синхронных модельных и измеренных значений, и все члены, на которые эта форма может быть разложена, находятся в оптимальном соотношении с точки зрения достижения этой цели. То есть, любые иные формы, в частности предлагаемые Гуптой с соавторами, данную задачу не решают. Фактически, эти предложения вводят инновационную, по сравнению с подходом Мерфи, идею – ослабить требование максимальной близости синхронных прогнозных и фактических значений в пользу большей близости параметров соответствующих распределений вероятности.

Возможности альтернативных мер качества предсказания

Для развития упомянутой инновации зафиксируем еще раз обсуждаемые выше исходные концептуальные положения. Вся информация о качестве и эффективности прогнозов заключена в двумерном совместном распределении вероятностей прогностических и фактических значений переменной, на основе которого строятся любые соответствующие оценки и анализ. Хороший прогноз характеризуется двумерным распределением с близкими маргинальными распределениями (при идеальном прогнозе они совпадают) прогностических и фактических значений, и сильной корреляцией между ними (в идеале – прямая однозначная связь). В качестве базовой меры расхождения, на основе похода «оценки умений», предложена форма, направленная на минимизацию суммы квадратов разностей синхронных модельных и измеренных значений, то есть на точное воспроизведение моделью (прогнозом) временной последовательности фактических значений переменной.

В гидрологии суши это формулируется как задача точного воспроизведения моделью измеренного гидрографа стока, для чего базовой формой меры расхождения является NSE, в различных возможных его вариантах. Оба рассмотренных выше разложения показывают, что оценка NSE полностью сводится к комбинации статистических параметров двумерного распределения вероятностей – маргинальных средних и среднеквадратических отклонений, а также коэффициента корреляции. В случае двумерного нормального распределения линейно связанных переменных эти параметры содержат исчерпывающую информацию о полном распределении, но такие условия в гидрологии практически никогда не выполняются. В случае ассиметричных маргинальных распределений и нелинейности связей неполнота информации, заключенной в оценке NSE, обусловлена свойствами конкретного двумерного распределения.

Кроме того, вышеизложенный анализ выявил, что воспроизведение отдельных учитываемых параметров двумерного

распределения находится в определенном противоречии. Требование, чтобы прогноз имел высокую корреляцию с фактическими значениями, означает попытку синхронного воспроизведения временной последовательности значений переменной. Однако в любом реальном случае прогнозирования получаем корреляцию, далекую от 1, а это приводит к занижению дисперсии прогноза, что также является его недостатком.

Легко увидеть, что требование (постановка в виде идеальной цели) точного воспроизведения временной последовательности фактических значений переменной возникло исторически на основе потребностей той области, в которой изначально система оценок развивалась – это область синоптической метеорологии и краткосрочных прогнозов погоды. Также естественно это требование было усвоено в краткосрочных гидрологических прогнозах, поскольку именно здесь необходим непрерывный, в нашей терминологии, прогноз гидрографа. Однако, когда речь идет о долгосрочных и сценарных прогнозах, инженерно-гидрологических расчетах, предсказываются агрегированные за большой период (месяц, сезон, год) величины, либо расчетные значения заданной обеспеченности. Требование пошагового воспроизведения гидрографа в таких случаях явно избыточно, в то время как необходимость наиболее правильного воспроизведения свойств распределения вероятности прогнозируемой характеристики выдвигается на первый план.

Есть и другой мотив, делающий целесообразным использование при калибровке моделей мер сходимости типа NSE, требующих точного воспроизведения гидрографа. Можно утверждать, что именно такой результат калибровки дает наибольшую уверенность в точности и надежности моделей, и позволяет в дальнейшем полагаться на них при решении самых разнообразных задач. Однако в этом утверждении содержится важное и неочевидное допущение, требующее специального разбора. Поясним это на примере.

Хорошо известна ситуация при калибровке моделей стока, когда некоторые дождевые паводки на гидрографе стока возникают как бы

ниоткуда, поскольку в записи осадков на соответствующий период времени нет паводкообразующего дождя. Модель в таких случаях воспроизвести паводок не может и допускает грубую ошибку. Менее очевидные, но на самом деле вполне подобные, ситуации связаны с выпадением сильных дождей, от которых на фактическом гидрографе существенного паводка не обнаруживается. Непосредственные причины этого очевидны – редкость наблюдательных сетей и «пятнистость» выпадения сильных дождей приводят к пропускам либо ливней, либо паводков в рядах наблюдений. Если таких случаев много, то это сильно отражается на итоговом значении оценки качества моделирования.

При работе с гидрологической моделью обрабатываются потоки метеорологических событий на входе и гидрологических событий на выходе. Для простоты будем говорить далее о ливнях и паводках. Если система наблюдений такова, что потоки ливней и паводков состоят из генетически связанных событий, и только из них, то эти потоки связаны взаимно-однозначным соответствием и такие данные можно назвать «репрезентативными по потоку событий». Тогда ошибки работы модели связаны либо с ошибками наблюдений, либо со структурой и параметрами моделей, но концепция калибровки модели в этом случае вполне состоятельна. То есть, при такой ситуации, удлинение рядов и повышение качества наблюдений со временем должно дать в результате калибровки надежные устойчивые значения параметров, имеющие определенный физический смысл либо, по крайней мере, рациональную интерпретацию. Обычно такова ситуация для рек с устойчивым режимом и преобладанием снеготаяния в питании стока, поскольку событиями в этом случае являются гидрологические сезоны, а представить себе систему наблюдений, способную «пропустить» целую весну или лето, невозможно.

Если взаимно-однозначное соответствие входного и выходного потоков событий существенно нарушается, что характерно для рек с дождевым питанием и паводковым режимом, то именно эти нарушения становятся основным источником погрешностей в работе моделей. Причем в этом случае опровергается сама

концепция калибровки – никакое удлинение рядов и повышение качества наблюдений не приведет в этом случае к получению откалиброванных устойчивых и надежных, физически осмысленных параметров. То есть, в ситуации наблюдений с явной «нерепрезентативностью по потоку событий» ставить задачу точного воспроизведения гидрографа избыточно и некорректно.

Однако, гидрометрический створ и метеостанция, взаимно нерепрезентативные по потоку событий, могут быть по отдельности вполне репрезентативны по отношению к региональному гидрологическому режиму и климату. В таком случае стоковая модель будет способна, с использованием этих метеоданных, воспроизводить агрегированные и режимные характеристики стока, которые в процессе калибровки могут сопоставляться с данными этого гидропоста. Изложенные рассуждения приводят нас к целесообразности испытания меры, аналогичной NSE, но применяемой к ранжированным выборкам модельных и измеренных значений. В наиболее простом виде, и в привычных обозначениях это приводит к форме

$$R_r^2 = 1 - \frac{\sum_{r=1}^n (Qs_r - Qa_r)^2}{\sum_{r=1}^n (Qa_r - Qa)^2}. \quad (15)$$

Здесь все обозначения совпадают с формулой (2) лишь замена i на r указывает, что в ранжированной выборке перебор членов идет не по времени, а по рангу.

Формула (15) вполне может быть подвергнута разложению аналогично формуле (11):

$$R_r^2 = 2 \cdot \alpha \cdot \rho - \alpha^2 - \beta_n^2, \quad (16)$$

где α и β_n означают то же самое и численно равны таким же параметрам в формуле (11). Единственное существенное различие уравнений (16) и (11) – замена коэффициента парной линейной корреляции r на корреляционное отношение ρ , вычисляемое уже не по синхронным, а по одноранговым значениям переменной. В данном случае ρ является коэффициентом сходства формы кривых эмпирических кривых распределения прогностических и фактических значений,

поскольку именно их мы сравниваем, применяя ранжирование.

В показателе R_r^2 присутствуют два первых момента маргинальных распределений модельных и фактических значений переменной, и отдельная характеристика сходства их формы. Корреляция, в смысле синхронности колебаний двух переменных, этим показателем не оценивается. Формальный показатель ρ , очевидно, будет иметь значения гораздо выше, чем коэффициент линейной корреляции r , поэтому при использовании меры R_r^2 обычные градации качества моделирования, принятые для оценки NSE, нельзя использовать – все границы градаций должны быть определены заново. Представляется, что для данной оценки будет еще более целесообразным использование меры со стандартом не в виде среднего значения за период,

а осреднённых режимных посуточных данных в виде

$$R_r^2 = 1 - \frac{\sum_{r=1}^n (Q_{sr} - Q_{or})^2}{\sum_{r=1}^n (Q_{or} - Q_{ar})^2}, \quad (17)$$

где Q_{ar} – ранжированные средние за многолетний период расходы на каждую дату. Такая оценка будет намного чувствительнее.

На рисунке 2 представлены условные примеры вычисления NSE по ранжированным выборкам в двух вариантах в соответствии с формулами (15) и (17). Используются те же исходные данные, что и для рисунка 1. Интуитивно ясно, например, что вариант с принятием среднеемноголетнего значения в качестве стандартного прогноза по формуле (15) дает явно завышенную оценку качества прогноза и вряд ли может иметь практическую ценность.

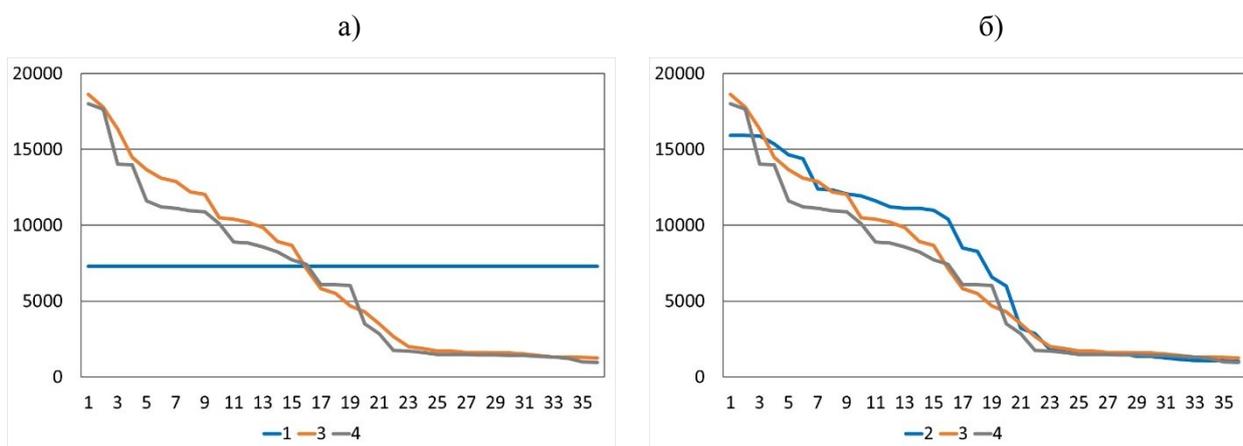


Рисунок 2. Условные примеры вычисления оценки NSE по ранжированным выборкам при разных вариантах принятого стандартного прогноза: а) среднее значение за период, $R_r^2=0,97$; б) режимные данные – среднеемноголетние величины на каждую дату, $R_r^2=0,70$: 1 – среднее, 2 – режим, 3 – прогноз, 4 – факт

Figure 2. Conditional examples of calculating NSE based on ranked samples for different versions of the accepted standard forecast: a) average value for the period, $R_r^2=0.97$; b) regime data – long-term average values for each date, $R_r^2=0.70$: 1 – mean, 2 – regime, 3 – forecast, 4 – fact

Пример имеет чисто иллюстративное значение, какие-либо интерпретации результатов представляются пока преждевременными. Однако, можно обратить внимание, что для обоих вариантов оценки доли необъясненной изменчивости переменной (residual variance) примерно вдвое меньше, чем для аналогичных традиционных мер (см. рисунок 1). Методическая

разработка оценок, основанных на таких мерах, должна быть предметом специального исследования.

Отметим, что переход к сравнению только маргинальных распределений моделированных и фактических значений переменной нацеливает на общее изменение подхода к оцениванию на последовательно статистический. Ясно, что при

точном совпадении модельной и измеренной ранжированных выборок показатель R_r^2 тоже достигает значения 1. При этом сравнение выборочных распределений, либо параметров, никогда не строится на требовании их точного совпадения, которое является событием с нулевой вероятностью. Достаточно вывода о принадлежности двух выборок к одной генеральной совокупности, либо принадлежности выборки к заданной генеральной совокупности, что проверяется с помощью статистических критериев.

С учетом структуры выражения (16), кажется целесообразным использование трех хорошо известных критериев: Стьюдента – для сравнения средних значений; Фишера – для сравнения дисперсий; χ^2 – для сравнения кривых распределения. Однако принцип применения критериев при решении такой задачи существенно изменяется. Если обычно критерии применяются для оценки различия между распределениями с высокой надежностью, для чего рассчитывают величины статистик расхождения с низкой вероятностью (0,05, либо 0,01), то в данном случае нам требуется как можно более высокая вероятность совпадения распределений, т.е. следует использовать статистики критериев с вероятностью, вероятно, 0,95 или 0,99. Поскольку три предлагаемых критерия могут дать различные оценки вероятности, то для итоговой оценки необходимо использовать некоторую их комбинацию, в первом приближении – минимальное значение из трех, как наиболее жесткий показатель качества моделирования. Разумеется, использование такого подхода требует разработки нового метода оценивания качества моделирования и прогноза, вероятно даже создания специальных статистических критериев.

Отметим также, что если используемая при моделировании система наблюдений обладает описанным выше свойством «репрезентативности по потоку событий», то достижение высокого качества моделирования в смысле $R_r^2 \cong 1$ будет в то же время означать и достижение высокого значения обычного показателя NSE. Вероятно, это можно доказать строго, но во всяком случае интуитивно ясно, что если на входе и выходе системы будем наблюдать

строго генетически связанные потоки событий, то высокое сходство маргинальных законов распределения модельных и фактических данных возможно лишь при высоком качестве воспроизведения гидрографа; иное чрезвычайно маловероятно. Напротив, если «репрезентативности по потоку событий» в данных не наблюдается, то качественное воспроизведение гидрографа невозможно, но хорошее сходство маргинальных распределений – вполне.

Заключение

Подчеркнем еще раз, что современные тщательно разработанные в физико-математическом отношении модели гидрологических процессов требуют адекватной, развитой системы методических средств их верификации и тестирования на единой методологической основе. Такой системы в настоящее время не существует, что является ключевой современной проблемой не только развития, но и прикладного использования гидрологических моделей. К настоящему времени можно указать ряд конкретных проблем, с которыми сталкиваются специалисты-гидрологи при калибровке моделей и проверке прогнозов с применением традиционных статистических подходов и форм оценок.

1. Отсутствуют надежные и общепринятые методы отбраковки сильных выбросов в рядах расчетных и фактических значений переменных вследствие грубых погрешностей, равно неизбежных при моделировании и измерениях.

2. Отсутствуют общепринятые методы оценки качества моделей и прогнозов, способных надежно и точно работать в условиях нестационарного климата и изменений ландшафтов, отражающихся в рядах гидрологических переменных.

3. Отсутствуют общепринятые методы оценки качества многомерного (по группе станций) моделирования или прогноза, учитывающих пространственную структуру потока гидрологических событий в пределах пространственно-распределенных моделей.

4. Для пространственно-распределенных физически обоснованных моделей не существует также комплексных методов их калибровки и

верификации на основе всей имеющейся информации об элементах гидрологического цикла, которые в них воспроизводятся и по которым имеются данные сетевых наблюдений.

5. Недостаточно разработана методическая база оценивания бинарных прогнозов редких событий.

6. Слабо разработаны методы калибровки моделей и верификации прогнозов для случаев, когда результаты выдаются в вероятностной форме.

Перечисленные выше задачи определяют, но не исчерпывают, направления дальнейших исследований по проблематике.

Литература

Аполлов Б.А., Калинин Г.П., Комаров В.Д. Курс гидрологических прогнозов. Л.: Гидрометеиздат, 1974. 419 с.

Борщ С.В., Христофоров А.В. Оценка качества прогнозов речного стока // Труды Гидрометеорологического научно-исследовательского центра Российской Федерации. 2015. Т. 355. С. 3–195.

Борщ С.В., Симонов Ю.А., Христофоров А.В. Эффективность моделирования и прогнозирования речного стока. // Гидрологические исследования и прогнозы. № 1 (375). 2020. С. 176–189.

Борщ С.В., Симонов Ю.А., Христофоров А.В. Прогнозирование стока рек России. М.: Изд-во ФГБУ «гидрометцентр России», 2023 г. 200 с.

Наставления по службе прогнозов. Раздел 3. Ч.1. Прогнозы режима вод суши. Л.: Гидрометеиздат, 1962. 194 с.

Рождественский А.В., Чеботарев А.И. Статистические методы в гидрологии. Л.: Гидрометеиздат, 1974. 424 с.

Руководство по гидрологической практике ВМО. Сбор и обработка данных, анализ, прогнозирование и другие применения (ВМО №168). Женева: Секретариат ВМО. 1994. 811 с.

Руководство по гидрологическим прогнозам. Вып. 2. Краткосрочный прогноз расхода и уровня воды на реках. Л.: Гидрометеиздат, 1989. 245 р.

Руководство по гидрологическим прогнозам. Вып. 1. Долгосрочные прогнозы элементов водного режима рек и водохранилищ. Ленинград: Гидрометеиздат, 1989. 357 р.

Технический регламент ВМО. Том III. Гидрология. (ВМО №49). Женева: Секретариат ВМО. 2006. 140 с.

References

Apollov B.A., Kalinin G.P., Komarov V.D. *Kurs gidrologicheskikh prognozov [Course of hydrological forecasts]*. L.: Gidrometeoizdat, 1974. 419 p. (In Russian).

Beven K.J. *Rainfall-runoff Modelling: The Primer*. Chichester: Wiley&Sons, 2001. 488 p.

Bogner K., Pappenberger F. Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. *Water Res. Res.*, 2011, vol. 47(W07524), p. 24. JRC56973. DOI: 10.1029/2010WR009137.

Borshch S.V., Khristoforov A.V. Otsenka kachestva prognozov rechnogo stoka [Assessing the quality of river flow forecasts]. *Proceedings of the Hydrometeorological Research Center of the Russian Federation*, 2015, vol. 355, pp. 3–195. (In Russian).

Borshch S.V., Simonov Yu.A., Khristoforov A.V. *Effektivnost' modelirovaniya i prognozirovaniya rechnogo stoka [Efficiency of river flow modeling and forecasting]*. M.: Hydrological research and forecasts, No. 1 (375), 2020, pp. 176–189. (In Russian).

Borshch S.V., Simonov Yu.A., Khristoforov A.V. *Prognozirovaniye stoka rek Rossii [Forecasting the flow of Russian rivers]*. M.: Hydrometeorological Center of Russia Publ., 2023. 200 p. (In Russian).

Gupta H. V., Kling H.; Yilmaz K. K., Martinez G. F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, 2009, vol. 377, pp. 80–91. DOI:10.1016/j.jhydrol.2009.08.003.

Jolliffe I.T., Stephenson D.B. *Forecast verification. Forecast verification: a practitioner's guide in atmospheric science*. Chichester: Wiley&Sons, 2012. 254 p.

Murphy A.H., Winkler R.L. A General Framework for Forecast Verification. *Monthly Weather Review*,

- Beven K.J. Rainfall-runoff Modelling: The Primer. Chichester: Wiley&Sons, 2001. 488 p.
- Bogner K., Pappenberger F. Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system // Water Res. Res. 2011. V. 47 (W07524) P. 24. JRC56973. DOI: 10.1029/2010WR009137.
- Gupta H. V., Kling H., Yilmaz K. K., Martinez G. F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling // J. Hydrol. 2009. Vol. 377. P. 80–91. DOI:10.1016/j.jhydrol.2009. 08.003.
- Jolliffe I.T., Stephenson D.B. Forecast verification // Forecast verification: a practitioner's guide in atmospheric science. Chichester: Wiley&Sons, 2012. 254 p.
- Murphy A.H., Winkler R.L. A General Framework for Forecast Verification // Monthly Weather Review. 1987. Vol. 115. P. 1330–1338. DOI:10.1175/1520-493(1987)115<1330:AGFFV>2.0.CO;2.
- Murphy A. H. Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient // Monthly Weather Review. 1988. Vol. 116. P. 2417–2424. DOI:10. 1175/1520-493(1988)116<2417:SSBOTM>2.0.CO;2.
- Nash J. E.; Sutcliffe J. V. River flow forecasting through conceptual models part I — A discussion of principles // J. Hydrol. 1970. Vol. 10 (3). P. 282–290. DOI:10.1016/0022-1694(70)90255-6.
- Wilks D.S. Statistical methods in the atmospheric sciences: an introduction // Int. Geophys. Ser. Academic Press. 1995. Vol. 59. P. 467.
- 1987, vol. 115, pp. 1330–1338. DOI:10.1175/1520-493(1987)115<1330:AGFFV >2.0.CO;2/
- Murphy A. H. Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient. *Monthly Weather Review*, 1988, vol. 116, pp. 2417–2424. DOI: 10.1175/1520-493(1988)116<2417:SSBOTM> 2.0. CO;2.
- Nash J. E.; Sutcliffe J. V. River flow forecasting through conceptual models part I — A discussion of principles. *J. Hydrol.*, 1970, vol. 10(3), pp. 282–290. DOI:10.1016/0022-1694(70)90255-6.
- Nastavleniya po sluzhbe prognozov. Razdel 3. CH.1. Prognozy rezhima vod sushi. [Instructions for the forecasting service. Section 3. Part 1. Forecasts of land water regime.]*. L.: Gidrometeoizdat, 1962. 194 p. (In Russian).
- Rozhdestvensky A.V., Chebotarev A.I. *Statisticheskiye metody v gidrologii [Statistical methods in hydrology]*. L.: Gidrometeoizdat, 1974. 424 p. (In Russian).
- Rukovodstvo po gidrologicheskoy praktike VMO. Sbor i obrabotka dannykh, analiz, prognozirovaniye i drugie primeneniya [WMO Guide to Hydrological Practices. Data collection and processing, analysis, forecasting and other applications] (WMO No. 168)*. Geneva: WMO Secretariat, 1994. 811 p. (In Russian).
- Rukovodstvo po gidrologicheskim prognozam. Vyp. 2. Kratkosrochnyy prognoz raskhoda i urovnya vody na rekakh [Guide to Hydrological Forecasts. Vol. 2. Short-term forecast of flow and water level on rivers]*. L.: Gidrometeoizdat, 1989. 245 p. (In Russian).
- Rukovodstvo po gidrologicheskim prognozam. Vyp. 1. Dolgosrochnnyye prognozy elementov vodnogo rezhima rek i vodokhranilishch. [Guide to Hydrological Forecasts. Vol. 1. Long-term forecasts of elements of the water regime of rivers and reservoirs]*. L.: Gidrometeoizdat, 1989. 357 p. (In Russian).
- Tekhnicheskii reglament VMO. Tom III. Gidrologiya. [WMO Technical Regulations. Volume III. Hydrology.]* (WMO No. 49). Geneva: WMO Secretariat, 2006. 140 p. (In Russian).
- Wilks D.S. Statistical methods in the atmospheric sciences: an introduction // *Int. Geophys. Ser. Academic Press*, 1995, vol. 59, p. 467.