

Федеральная служба по гидрометеорологии и мониторингу
окружающей среды (Росгидромет)

Федеральное государственное бюджетное учреждение
«ГИДРОМЕТОРОЛОГИЧЕСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ
ЦЕНТР РОССИЙСКОЙ ФЕДЕРАЦИИ»
(ФГБУ «Гидрометцентр России»)

Д.Б. Киктев, А.В. Муравьев, А.Ю. Бундель

**Методические рекомендации
по верификации метеорологических
прогнозов**

Москва
2021

Киктев Д.Б., Муравьев А.В., Бундель А.Ю. **Методические рекомендации по верификации метеорологических прогнозов.** – М.: Типография АМА ПРЕСС, 2021. – 94 с.

ISBN 978-5-9908623-4-0

Многообразие форм представления численных прогнозов погоды и требований потребителей к этой информации делает задачу верификации (оценки качества) актуальной и разноплановой.

В Методических рекомендациях представлены материалы о различных по назначению, степени полноты и сложности подходах к задаче верификации детерминистских и вероятностных численных прогнозов погоды. Даются рекомендации по практическому применению различных методов оценки качества прогнозов.

Область применения рассматриваемых методов верификации не ограничивается численными метеорологическими прогнозами. Большая часть рассматриваемых методов может быть адаптирована и для других предметных областей, где практикуется выпуск тех или иных видов прогнозов и требуется оценивать их качество.

В приложении к Методическим рекомендациям представлен Гнездовой словарь английских терминов верификации с переводом на русский язык и толкованием.

Методические рекомендации предназначаются как для специалистов в области численных прогнозов, так и для потребителей этих прогнозов.

СОДЕРЖАНИЕ

1. Основные понятия	5
1.1. Цели верификации	5
1.2. Размерность задачи верификации гидродинамических прогнозов. Совместное распределение прогнозов и наблюдений	7
1.3. Атрибуты прогнозов	8
1.4. Особенности верификации в зависимости от типа прогноза	9
1.5. Оценка неопределенности показателей качества прогнозов	10
1.6. Данные наблюдений. Учет неопределенности данных наблюдений	14
1.7. Проблема пространственно-временного согласования прогнозов и наблюдений (интерполяция, ап- и даунскейлинг)	15
2. Объективная верификация численных прогнозов погоды	17
2.1. Детерминированные прогнозы	18
2.1.1. Показатели качества для детерминированных прогнозов непрерывных величин	18
2.1.2. Показатели качества для детерминированных прогнозов дискретных величин	22
2.2. Ансамблевые прогнозы метеорологических полей	26
2.3. Особенности верификации прогнозов экстремальных величин	34
2.3.1. Верификация прогнозов экстремальных явлений и величин	34
2.3.2. Формирование выборок для анализа экстремальных явлений на примере метода блочных максимумов	35
2.3.3. Пример верификации прогнозов явлений со значительными воздействиями: отечественный опыт	36
2.3.4. Некоторые практические выводы о свойствах индексов экстремальной зависимости	37

3. Пространственные методы верификации.	
Проблема двойного штрафа и способы ее преодоления	40
4. Программное обеспечение системы верификации	50
ЗАКЛЮЧЕНИЕ	51
ПРИЛОЖЕНИЕ 1. Гнездовой словарь английских терминов верификации с переводом на русский язык и толкованием	52
ПРИЛОЖЕНИЕ 2. Стандартизованная верификация продукции детерминистского численного прогноза погоды	75
ПРИЛОЖЕНИЕ 3. Стандартные критерии оценки для верификации продукции глобальной системы ансамблевого прогноза	84
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	89

В Методических рекомендациях представлены материалы о различных по назначению, степени полноты и сложности подходах к задаче верификации детерминистских и вероятностных численных прогнозов погоды.

Даются рекомендации по практическому применению различных методов оценки качества прогнозов. Разработан и представлен Гнездовой словарь основных английских терминов верификации с переводом на русский язык и толкованием.

1. Основные понятия

1.1. Цели верификации

Испытание прогностической системы (или нескольких систем) – это **численный эксперимент** (1) для оценки качества прогнозов и (2) для диагноза источников ошибок прогнозов. Предшествовать такому эксперименту должна тщательно разработанная **программа испытаний** с изложением **условий их проведения**.

Отечественная традиция планирования испытаний численных методов прогноза насчитывает более полувека. Так, в [23, 24] приводятся рекомендации специальной рабочей группы о необходимости тщательного планирования испытаний. В [8] описаны примеры рекомендуемых или необходимых условий проведения испытаний для разных величин и явлений, из которых важными представляются следующие условия:

- 1) правильный выбор территории;
- 2) выбор эталонных данных наблюдений для сравнения их с прогнозами;
- 3) стратификация результатов испытаний по типам синоптических процессов;
- 4) единообразие используемых оценок;
- 5) достаточное количество прогнозов для статистической достоверности выводов и для расчета доверительных интервалов;
- 6) расчет статистических характеристик для полей;
- 7) учет взаимной корреляции между значениями в узлах сетки.

Этапы стандартной верификации прогнозов

В процессе оценивания качества прогнозов выделяются, как правило, последовательные этапы [4, 8, 21, 23, 64]:

- 1) выделение основной цели верификации;
- 2) выделение желательных атрибутов прогнозов;
- 3) формулировка мер качества;

4) определение методов оценки неопределенности и проверки гипотез для оценок качества прогнозов.

Обычно выделяют **административную, научную и экономическую цели верификации** [27]. Однако следует видеть и учитывать взаимосвязанность этих целей. Например, если в качестве основной цели разрабатываемой верификации избирается **научная** (или **диагностическая**) верификация, то некоторые ее результаты могут быть отнесены как к экономической, так и к административной верификации.

Выделение желательных атрибутов прогнозов предшествует отбору конкретных мер качества. Для понимания такого подхода требуется освоение зарубежного опыта как в определении атрибутов, так и в их активном применении на практике. Для соответствующей квалификации необходимо знакомство с зарубежными первоисточниками [47, 63, 64, 74, 76], и в первую очередь с понятиями «основная структура верификации» [65] и «атрибут прогноза» [47, 74, 76] (основные атрибуты прогнозов перечислены ниже в п. 1.3).

В России систематическое изложение вопросов методологии верификации применительно к вероятностным прогнозам впервые появляется в книге Грузы и Раньковой [4]. Практическое применение отбора мер на основе выделенных атрибутов демонстрируется в Методических указаниях в разделе о вероятностных прогнозах [8, п. 1.5]. Очень часто рассчитанные оценки качества прогнозов не сопровождаются оценками степени их неопределенности и статистической проверкой гипотез об успешности прогнозов (например, о том, имеют ли результаты одной прогностической системы преимущество перед результатами другой системы). Получаемые при статистической проверке гипотез оценки значимости призваны подтвердить статистическую надежность расчетов и обоснованность формулируемых рекомендаций и являются обязательным компонентом верификации. Некоторые методы оценки статистической значимости приводятся ниже в п. 1.5.

1.2. Размерность задачи верификации гидродинамических прогнозов. Совместное распределение прогнозов и наблюдений

В зависимости от цели верификации могут различаться размерность задачи верификации, применяемые оценки качества и их атрибуты (см. п. 1.3), периоды оценивания и форма представления результатов верификации. Однако следует понимать, что любая система верификации основывается на анализе **функции совместного распределения** $p(y, o)$ **прогнозов** y и **наблюдений** o [76].

Даже если изначально прогнозы и наблюдения являются непрерывными переменными, для практических целей их часто заменяют конечными наборами округленных значений. Обозначим прогноз как y_i , который может принимать любое из I значений. Соответственно, наблюдение обозначим как o_j , которое может принимать любое из J значений. Тогда совместное распределений прогнозов и наблюдений обозначается $p(y_i, o_j)$, где $i=1, \dots, I$, $j=1, \dots, J$. Это дискретная двумерная функция распределения вероятностей, сопоставляющая вероятность с каждой из возможных $(I \times J)$ комбинаций прогнозов и наблюдений.

Вся информация о качестве прогностической системы содержится в этом совместном распределении, а верификация, основанная на нем, называется **диагностической**, в отличие от использования какой-либо одной из метрик, которая не даст полной информации. Поскольку все вероятности в совместном распределении $p(y_i, o_j)$ в сумме дают 1, **размерность**, или **количество степеней свободы** задачи верификации равно $(I \times J) - 1$. То есть для того, чтобы задать распределение, необходимо иметь данные для $(I \times J) - 1$ отдельных вероятностей.

При проведении верификации полей метеоэлементов поля приближаются значениями, интерполированными на точки сетки в пространстве или осредненными в ячейках сетки. Оценивается соответствие прогнозов и наблюдений в каждой точке, и каждая точка как бы добавляет степень свободы, поэтому размерность задачи верификации полей очень велика. Поэтому используют набор скалярных оценок точности, который по определению не описывает полностью задачу верификации, но тем не менее полезен на практике.

1.3. Атрибуты прогнозов

Ниже перечислены аспекты (атрибуты), которые важны при определении качества прогнозов [47, 74, 76].

1. **Точность** (accuracy) – оценка согласованности прогнозов и наблюдений в среднем. Учитывает согласованность в отдельных парах прогноз-наблюдение, затем делается осреднение по всем рассматриваемым случаям прогноза.

2. **Смещение** (bias), или систематическое смещение. Оценивается соответствие между средним значением прогнозов и средним значением наблюдений. Отличается от точности тем, что не учитывает соответствие между отдельными парами прогноз-наблюдение.

3. **Надежность** (reliability), или калибровка (calibration) – относится к распределению наблюдений для конкретных значений прогноза, т. е. надежность характеризует условные распределения вероятности наблюдений при условии конкретного прогноза.

4. **Разрешающая способность** (resolution) – характеризует способность прогнозов разделять всю выборку наблюдений на группы, отличающиеся друг от друга. Прогноз с хорошей разрешающей способностью способен отделить один тип результата (явления) от другого (например, выпадение осадков более и менее 1 мм/час). Разрешающая способность также связана с условным распределением вероятности наблюдений при условии конкретного прогноза, но отличается от надежности тем, что сравнивает средние значения распределений наблюдений при различных значениях прогноза, в то время как надежность оценивает соответствие условных распределений наблюдений с самими прогнозами-условиями.

5. **Дискриминирующая способность** (discrimination) обратна разрешающей способности в том смысле, что относится к различию условных распределений прогнозов при условии различных значений наблюдений. Она означает способность системы давать различные прогнозы для разных исходов по данным наблюдений. Например, мы имеем выборку из 100 случаев дождя и из 100 случаев снега по данным наблюдений. Если прогноз дает снег в 40 случаях из ста, когда наблюдался снег, и в 40 случаях из 100, когда наблюдался дождь, то это значит, что прогноз не умеет делать различий (дискриминировать) между снегом и дождем.

6. **Детализация** (refinement), или контрастность (sharpness) – детализация определяется дисперсией распределения прогноза $p(y_i)$, т. е. относится к прогнозам независимо от наблюдений, т. е. к безусловным распределениям прогнозов. Детализированные прогнозы имеют большой разброс (spread), что означает, что каждое из прогностических значений выпускалось достаточно часто, т. е. прогнозы потенциально покрывают большой диапазон ситуаций. В то же время, если распределение $p(y)$ узкое, большая часть прогнозов относится к одной или нескольким близким категориям, детализация недостаточна.

1.4. Особенности верификации в зависимости от типа прогноза

Построение системы верификации с соответствующим количеством степеней свободы зависит как от цели верификации, так и от типа и способа представления прогноза:

- Детерминированный или вероятностный прогноз
- Особенности представления прогноза по пространству и времени:
 - полей в пространстве (*например, поле давления*);
 - по времени (*например, дневной максимум температуры в Москве*);
 - осредненная по пространству и времени величина (*например, среднемесячная глобальная температура воздуха*)
- Специфика прогноза:
 - дихотомный (или бинарный) прогноз Да – Нет (*будет дождь или нет*);
 - мультикатегорийный (*теплые, нормальные или холодные условия*);
 - непрерывный (*изменение температуры, давления, ...*);
 - объектно-ориентированный (*положение тропического циклона, области осадков, т. е. прогноз характеристик объекта*)

Верификация должна быть многомерной, одно число не дает всей нужной информации. Иногда разные оценки дают противоречивые результаты. Необходим компромисс между разными критериями качества, зависящий также от цели верификации. Например, при большом ущербе от опасного явления лучше перестраховаться. При этом, например, такая оценка, как *доля ложных тревог*, может иметь больший вес.

При отборе мер качества принимается во внимание не только их соответствие выбранным желательным атрибутам прогнозов (и соответствующим свойствам разрабатываемой системы верификации, см. п. 1.3), но и некоторые дополнительные свойства, к каким относятся, например, согласованность формулировки прогноза с истинным мнением прогнозиста (consistency, [63]) и уравнивание, или обнуление качества неметодических прогнозов (инерционных, случайных, климатических...) (equitability, [48, 49, 76]). Первое свойство относится главным образом к вероятностным прогнозам и предостерегает прогнозиста от подстраховки (или перестраховки). Уравнивание предполагает, что правильные прогнозы менее частых явлений (например, торнадо) имеют больший вес, чем правильные прогнозы более обычных явлений, и таким образом поощряет разумный риск.

Прогнозы погоды подразделяются:

- Детерминированные прогнозы непрерывных величин
- Детерминированные прогнозы дискретных величин (явлений, событий, категорий)
- Вероятностные (в настоящее время главным образом ансамблевые) прогнозы.

Соответствующие оценки рассмотрены в пп. 2.1–2.3. Также существуют специальные методы оценивания прогнозов полей метеовеличин (раздел 3).

В таблице 1 приводятся основные рекомендуемые методы верификации в зависимости от типа прогноза, которые будут рассмотрены в следующих разделах. Отметим, что таблица 1 содержит примеры наиболее типичных метеопеременных для того или иного метода, что не исключает более нетрадиционное применение.

1.5. Оценка неопределенности показателей качества прогнозов

На практике возможности верификации прогнозов всегда ограничиваются возможностями имеющейся наблюдательной базы (недостаточной плотностью наблюдательной сети, частотой наблюдений, погрешностями измерений, продолжительностью рядов измерений). В этой связи оценка сопутствующей неопределенности показателей качества прогнозов является важной составляющей процедуры их верификации.

Таблица 1. Основные рекомендуемые методы верификации в зависимости от типа прогноза

Тип прогноза	Тип наблюдений	Предиктант	Методы	Примеры
детерминистский	станционные	непрерывный	ME, RMSE, MAE, коэффициент корреляции (п. 2.1)	температура воздуха, отн. влажность, давление
детерминистский	станционные	категорийный	оценки из таблицы сопряженности (п. 2.2)	осадки, порывы ветра в категориях, выделенных по пороговым значениям
вероятностный	станционные	непрерывный	CRPS, ранговые гистограммы (п. 2.3)	все переменные
вероятностный	станционные	мультикатегорийный	RPS (п. 2.3)	осадки, порывы ветра и др. в категориях, выделенных по пороговым значениям
вероятностный	станционные	дихотомный	оценка Брайера, ROC-кривая и площадь под ней, диаграмма надежности (п. 2.3)	осадки, порывы ветра, превышающие порог
детерминистский/вероятностный	сеточные	дихотомный	окрестные, объектно-ориентированные (раздел 3)	осадки, отражаемость, облачность

Требования к наблюдательной базе в свою очередь зависят от предполагаемой области использования данных наблюдений. Так, для задач глобального численного прогноза, для прогноза по ограниченной территории, наукастинга и сверхкраткосрочного прогноза погоды требования к плотности сети и частоте наблюдений различны (например, для наукастинга требования более жесткие, для задач глобального прогнозирования – более умеренные). Исходя из этого Всемирная метеорологическая организация сформулировала ориентировочные требования к плотности сети и частоте наблюдений с учетом свойств конкретного метеорологического параметра (так, для осадков требования более жесткие, чем для температуры).

Для различных областей приложений и гидрометеорологических параметров рекомендации ВМО (<https://space.oscar.wmo.int/observingrequirements>) задают диапазон значений для плотности сети и частоты наблюдений – от минимально допустимых до перспективных целевых значений. Неопределенность наблюдений также рассматривается в п. 1.6.

Любой показатель верификации должен рассматриваться в качестве выборочной оценки «истинного» значения для бесконечно большого набора данных по верификации. Следовательно, существует некоторая неопределенность, связанная со значением показателя, особенно если размер выборки небольшой или данные не являются независимыми. Должна использоваться какая-то оценка неопределенности (как правило, доверительные интервалы, ДИ), чтобы установить ограничения на ожидаемое значение показателя верификации. Это также помогает оценить, являются ли различия между конкурирующими прогностическими системами статистически значимыми.

Имеются разнообразные методы оценить статистическую значимость – параметрические, непараметрические, робастные, эвристические, физически обоснованные, байесовские, монтекарловские и др.

С развитием ЭВМ возросли возможности статистического (стохастического) моделирования законов распределения на основе датчиков случайных чисел. На практике часто нет сведений о распределении оцениваемых величин, участвующих в формулах мер, что заставляет прибегать к статистическим испытаниям, чтобы делать выводы о наличии или отсутствии у прогнозов некоторого мастерства [4].

Возросшие вычислительные мощности позволяют проводить «статистические испытания» с приемами так называемой кросс-валидации (перекрестных проверок). Именно такие приемы обработки данных следует рекомендовать для решения проблем этого пункта.

Особенно важно построение ДИ в случае редких явлений, для которых выборки по определению малочисленны, для случаев большой изменчивости данных, а также если мы хотим сделать вывод о статистически значимом превышении качества одной прогностической системы над другой. ДИ содержат больше информации о неопределенности верификационной оценки, чем простой тест на значимость. Для некоторых оценок ДИ рассчитываются с использованием параметрических методов (математические формулы имеются для ДИ для биномиальных или нормальных распределений, в частности этим распределениям подчиняются средняя ошибка ME и оценки, представляющие собой пропорции (PC, POD, F, TS – см. определения в п 2.1.2). В целом нельзя ожидать, чтобы большинство показателей верификации могли удовлетворять этим условиям. Более универсальными являются непараметрические методы, в основном метод бутстрепа [15]).

Метод бутстрепа для расчета ДИ включает неоднократный пересчет показателей после произвольного формирования выборок из набора данных и последующей их замены, опять же произвольной, из набора исходных первоначальных данных [29].

Общий алгоритм бутстрепа:

- Создать N случайных выборок одинакового размера из исходной выборки пар прогноз – наблюдение, с возвратом (выбрали – положили обратно, N раз)
 - N должно быть достаточно большим, например 1000 раз
 - Рассчитать оценку для каждой из полученных выборок, ранжировать от меньшего к большему и найти, например, 2,5 и 97,5-й процентиля (для 95 % значимости) от получившегося распределения оценки
 - При наличии автокорреляции в данных в пространстве и времени, особенно при более длительных сроках прогноза, рекомендуется перетасовывать не отдельные пары, а блоки смежных данных (блочный бутстреппинг, block bootstrapping). На основе расчета автокорреляции между прогнозами на последующие дни

был сделан вывод, что блоки из трех дней могут использоваться для расчета доверительных интервалов в 5 и 95 % [15].

1.6. Данные наблюдений. Учет неопределенности данных наблюдений

Наблюдения можно разделить на точечные и пространственно распределенные.

Примеры источников точечных наблюдений:

- стационарные данные SYNOP;
- корабельные наблюдения, буи;
- данные аэрозондирования TEMP;
- самолеты;
- добровольные наблюдатели.

Примеры пространственно распределенных данных наблюдений:

- радары;
- спутники;
- объективный анализ.

Верификация – сравнение результатов прогнозов с фактом. За факт принимаются данные измерений параметров атмосферы (и океана для некоторых видов прогнозов) тем или иным прибором, т. е. наблюдения. Хотя мы считаем наблюдения приближением “истины”, они содержат ошибки. До сих пор ошибки наблюдений были пренебрежимо малы по сравнению с ошибками прогноза. Однако по мере возрастания качества прогнозов ошибки наблюдений становятся сопоставимы с ошибками прогнозов. Поэтому необходимо учитывать неопределенность в анализе и наблюдениях.

Типы ошибок наблюдений:

- инструментальные ошибки;
- случайные ошибки;
- ошибки передачи данных;
- ошибки репрезентативности станции (насколько хорошо она отражает местные условия);
- ошибки анализа.

Методы учета неопределенности наблюдений:

- косвенный учет с помощью пространственных методов (см. раздел 3);

- учет в метриках: например, разложение на компоненты ошибки за счет прогноза и за счет наблюдений;
- наблюдения как вероятности: ансамбли наблюдений. Пример – ансамбль VERA [55, 56] – создание возмущений на основе распределения остаточных членов, которые вычисляются в процедуре контроля качества;
- учет в алгоритмах усвоения данных.

Для проведения верификации моделей высокого разрешения остро встает необходимость в данных наблюдений высокого разрешения, в частности сеточных данных. Сеточные данные могут быть результатом объективного анализа точечных данных, данными радаров и спутников, а также совместным анализом точечных и радарных/спутниковых данных. Особенно важно наличие качественных сеточных данных для проведения пространственной верификации (раздел 3). Следует учитывать, что качество радарных данных не всегда высоко [6].

Источники данных наблюдений для опасных явлений погоды, включая новые нетрадиционные источники, например данные мобильных телефонов и социальных сетей, подробно описаны в [62].

1.7. Проблема пространственно-временного согласования прогнозов и наблюдений (интерполяция, ап- и даунскейлинг)

Часто встает необходимость сравнить две версии модели или две модели с разным пространственно-временным разрешением. Данные наблюдений могут быть как точечные, так и сеточные. При этом встает проблема пространственно-временного согласования прогнозов и наблюдений.

Обычно по умолчанию предполагается, что прогнозы и наблюдения относятся к одному пространственному масштабу, однако это не всегда так, особенно если модельная величина представляет собой среднее по пространству (как правило, обобщенная по ячейке сетки величина), а наблюдение – точечная величина [76]. Соответственно, обобщенные модельные величины имеют другую изменчивость по сравнению с точечными стационарными данными. Выбор процедуры апскейлинга, регридинга, интерполяции для приведения данных в соответствие может существенно повлиять на результат верификации.

Часто используемые методы приведения сеточных данных на точки наблюдений:

- ближайшая точка по горизонтали (2D);
- ближайшая точка в 3D-пространстве (с учетом расстояния по высоте);
- среднее в некотором радиусе вокруг точки;
- среднее из значений в нескольких ячейках, окружающих точку.

При сравнении сеточных данных с разным шагом сетки есть несколько вариантов:

- привести данные с более мелким шагом сетки к более грубому – апскейлинг (от англ. upscale);
- привести данные с более грубым шагом сетки к более мелкому – даунскейлинг (от англ. downscale);
- привести данные к некоторому промежуточному шагу сетки.

Некоторые исследователи считают, что при наличии данных наблюдений высокого разрешения лучше провести даунскейлинг, чтобы сохранить тонкие детали наблюдений, т. е. приближения "истины" [39]. В то же время апскейлинг можно рекомендовать для удаления мелкомасштабного шума. При этом следует учитывать, что эффективное разрешение модели ЧПП составляет примерно семь точек модельной сетки [71]. Таким образом, выбор приведения данных к единому разрешению в целом зависит от цели исследования и имеющихся данных наблюдений. В приложениях 2 и 3 даются стандартные рекомендации ВМО по интерполяции данных.

2. Объективная верификация численных прогнозов погоды

Определение меры мастерства (Skill score)

Мастерство (skill) прогноза – точность к.-л. набора прогнозов по отношению к точности к.-л. эталонного (или референсного) набора прогнозов. Как правило, в качестве эталонного прогноза выбирают случайный, климатический или инерционный прогноз. Также часто бывает нужно оценить мастерство прогнозов с использованием нового прогностического метода по отношению к стандартному методу.

Мера мастерства – относительное превышение точности исследуемого прогноза по сравнению с эталонным прогнозом:

$$SS = \frac{A - A_{ref}}{A_{perf} - A_{ref}} \cdot 100\% ,$$

где A – какая-либо оценка точности; A_{ref} – оценка эталонного прогноза; A_{perf} – оценка идеального прогноза.

Если $A = A_{perf}$, мера мастерства SS достигает максимального значения 100 %.

Если $A = A_{ref}$, $SS = 0\%$, что означает отсутствие улучшения по отношению к эталонному прогнозу.

$SS < 0\%$ означает, что рассматриваемая прогностическая система дает прогнозы хуже, чем эталонная (по данной оценке).

Мера мастерства была введена для уравнивания оценивания в более или менее трудных для прогнозирования условий. Так, например, климатический прогноз отсутствия осадков в пустыне будет иметь весьма высокие оценки сам по себе. И нужна более высокая точность A , чтобы превзойти A_{ref} , чем в более трудной и неопределенной прогностической ситуации, когда A_{ref} ниже. Однако при расчете SS нужно аккуратно выбирать однородные выборки при агрегации оценок.

Оценки, построенные по принципу меры мастерства, как правило, называются **Критериями** (см. словарь в приложении 1).

2.1. Детерминированные прогнозы

В Наставлении по глобальной системе обработки данных и прогнозирования Дополнение IV к Техническому регламенту ВМО [15] перечислен стандартный набор показателей верификации для выпускаемых центрами Глобальной системы обработки данных и прогнозирования (ГСОДП) детерминистских ЧПП (см. приложение 2). Можно рассматривать рекомендации приложения 2 как базовый уровень требований к верификации детерминированных ЧПП.

В пп. 2.1.1 и 2.1.2 приводятся формулы для расчета основных показателей качества детерминированных прогнозов непрерывных и дискретных величин, которые упоминаются в приложении 2.

2.1.1. Показатели качества для детерминированных прогнозов непрерывных величин

Непрерывная величина может принимать любое значение на соответствующем отрезке прямой действительных чисел. Температура – пример непрерывной величины. На практике, однако, и прогнозы, и наблюдения все же представляются в виде конечного числа дискретных величин. Например, температура в прогнозе часто округляется до целых значений.

И такие прогнозы можно было бы оценивать как дискретные, но тогда количество градаций было бы таким большим, что таблицы сопряженности были бы очень громоздкими и, скорее всего, некоторые градации имели бы малое количество данных, т. е. были бы вырожденными. Таким образом, имеет смысл обращаться с теоретически непрерывными (а на практике дискретными) величинами в рамках непрерывного подхода. В нем также фундаментальное значение имеет совместное распределение прогнозов и наблюдений. Для непрерывных переменных многие оценки качества основаны на ошибке прогноза (например, $f - o$, где f означает прогноз, o – наблюдения, а n – количество пар "прогноз – наблюдение").

Однако характеристики прогнозов и наблюдений, а также их соотношение также представляют интерес. Такая концепция согласуется с общей схемой верификации, предложенной Мерфи и Винклером [65]. Статистики, приводимые ниже, базируются именно на этой философии верификации, которая учитывает множество аспектов качества прогноза, а не какую-то одну характеристику.

Средняя ошибка МЕ (Mean error), атрибут Смещение

Средняя ошибка является мерой общей систематической ошибки (bias) для непрерывных переменных. Определяется как

$$ME = \frac{1}{n} \sum_{i=1}^n (f_i - o_i) = \bar{f} - \bar{o}.$$

У идеального прогноза $ME = 0$. МЕ представляет собой разность среднего значения прогноза и среднего значения наблюдений, и поэтому она не содержит информации о соответствии в каждой отдельной паре прогноз – наблюдение в отличие от мер точности. При $ME > 0$ прогнозы в среднем завышают величину, при $ME < 0$ занижают (т. е. имеется систематическое смещение). Поскольку МЕ не отражает соответствие между отдельными прогнозами и наблюдениями, можно получить идеальное значение для плохого прогноза, если имеются компенсационные ошибки (сумма превышений прогнозом наблюдений равно сумме превышений наблюдений над прогнозами).

Мультипликативное смещение (Multiplicative bias) MBIAS, атрибут Смещение

Мультипликативная ошибка представляет собой просто отношение среднего прогнозов к среднему наблюдений:

$$MBIAS = \frac{\bar{f}}{\bar{o}}.$$

Средний квадрат ошибки (Mean-squared error) MSE, атрибут Точность

Средний квадрат ошибки MSE определяется как

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2.$$

Среднеквадратическая ошибка (Root-mean-squared error), RMSE, атрибут Точность

RMSE представляет собой квадратный корень из MSE:

$$RMSE = \sqrt{MSE}.$$

MSE с коррекцией систематической ошибки (Bias-corrected MSE), атрибут Точность

На MSE и RMSE сильно влияют большие ошибки. Также на них сильно влияют большие значения средней ошибки МЕ. MSE и RMSE могут изменяться в пределах от 0 до бесконечности. У идеального прогноза $MSE = RMSE = 0$. MSE можно переписать как

$$\text{MSE} = (\bar{f} - \bar{o})^2 + s_f^2 + s_o^2 - 2s_f s_o r_{fo},$$

где $\bar{f} - \bar{o} = \text{ME}$ и $s_f^2 + s_o^2 - 2s_f s_o r_{fo}$ представляет собой оценку дисперсии ошибки, s_{f-o}^2 . Таким образом, $\text{MSE} = \text{ME} + s_{f-o}^2$. Для понимания поведения MSE важно проанализировать *оба* члена MSE, а не только одну MSE. Более того, ME может оказывать сильное влияние на MSE, как показывает разложение.

Средняя абсолютная ошибка (Mean absolute error) **MAE**, атрибут Точность

Средняя абсолютная ошибка определяется как

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |f_i - o_i|.$$

MAE меньше подвержена влиянию больших ошибок, а также не зависит от средней ошибки. Идеальный прогноз имеет $\text{MAE} = 0$. MAE – это среднее арифметическое значение абсолютных величин разности членов каждой из пар прогноз – наблюдение. MAE идеального прогноза равно 0 и растет с увеличением ошибок прогноза. MAE может трактоваться как типичное значение ошибки прогноза в данной выборке. Это симметричная кусочно-линейная функция, которая одинаково наказывает занижение и завышение реального значения одинаковой величины.

Стандартное отклонение ошибки s_{f-o} определяется по формуле:

$$s_{f-o} = \sqrt{s_{f-o}^2} = \sqrt{s_f^2 + s_o^2 - 2s_f s_o r_{fo}}.$$

Следует отметить, что стандартное отклонение ошибки иногда называют MSE с коррекцией систематической ошибки, поскольку оно представляет собой средний квадрат разностей прогноз – наблюдение, из которого удален эффект систематической ошибки.

Коэффициент корреляции Пирсона (Pearson correlation coefficient), атрибут Точность

Коэффициент корреляции Пирсона показывает величину линейной зависимости между прогнозами и наблюдениями. Он определяется как

$$r = \frac{\sum_{i=1}^T (f_i - \bar{f})(o_i - \bar{o})}{\sqrt{\sum (f_i - \bar{f})^2} \sqrt{\sum (o_i - \bar{o})^2}}.$$

r изменяется в пределах от -1 до 1. Единица означает идеальную корреляцию, в то время как -1 означает идеальную отрицательную корреляцию. Значение 0 означает, что прогнозы и наблюдения не скоррелированы. Коэффициент корреляции отражает линейную связь между двумя переменными, но чувствителен к выбросам и не чувствителен к сдвигу (bias). Коэффициент корреляции следует использовать аккуратно, особенно если в прогнозах может быть заметное смещение.

Коэффициент ранговой корреляции Спирмана (Spearman rank correlation coefficient), атрибут Точность

Коэффициент ранговой корреляции Спирмана – робастная мера связи, основанная на рангах прогнозов и наблюдений, а не на самих значениях, т. е. выборки прогнозов и наблюдений ранжируются от наименьшего к наибольшему значению, которым присваиваются ранги (от 1 до n , где n – общее количество пар прогноз – наблюдение). Затем полученные пары рангов прогноз – наблюдение используются для расчета коэффициента корреляции, аналогичного коэффициенту корреляции Пирсона r .

Упрощенная формулировка ранговой корреляции Спирмана основана на ранности в каждой паре рангов (обозначаемой как d_i):

$$\rho_s = \frac{6}{n(n^2 - 1)} \sum_{i=1}^n d_i^2.$$

Подобно r , коэффициент ранговой корреляции Спирмана изменяется от -1 до 1. Значение 0 означает, что прогнозы и наблюдения не скоррелированы.

Процентили ошибок

Процентили ошибок дают больше информации о распределении ошибок, чем можно получить из среднего и стандартного отклонения ошибки. Процентили рассчитываются путем упорядочивания ошибок от наименьшей к наибольшей, определения рангового местоположения каждого из процентилей в упорядоченном ряду и сопоставления ранга с имеющимся значением ошибки. Данные процентили можно также использовать для построения бокс-плотов («ящичков с усами») ошибок. Часто рассчитываются 0.10, 0.25, 0.50, 0.75 и 0.90-й квантили ошибок.

2.1.2. Показатели качества для детерминированных прогнозов дискретных величин

Статистики верификации дихотомных или бинарных (два возможных исхода события) переменных формулируются с использованием таблицы сопряженности, как показано в таблице 2.

Таблица 2. Таблица сопряженности

		Наблюдения		
		Да	Нет	Сумма
Прогноз	Да	n_{11}	n_{12}	n_{10}
	Нет	n_{21}	n_{22}	n_{20}
	Сумма	n_{01}	n_{02}	n_{00}

$n_{10} = n_{11} + n_{12}$ – общее число точек, в которых прогнозировалось событие

$n_{01} = n_{11} + n_{21}$ – общее число точек, в которых наблюдалось событие

$n_{20} = n_{21} + n_{22}$ – общее число точек, в которых прогнозировалось отсутствие события

$n_{02} = n_{12} + n_{22}$ – общее число точек, в которых событие не наблюдалось

Базовая доля (частота явления по данным наблюдений)
(base rate):

$$P = \frac{n_{11} + n_{21}}{n_{00}}$$

Общая оправдываемость (Proportion correct) PC, атрибут Точность:

$$PC = \frac{n_{11} + n_{22}}{n_{00}}$$

Удовлетворяет принципу равнозначности событий, так как награждает правильные прогнозы явления и отсутствия явления одинаково. Однако это не всегда достоинство, особенно если явление редкое и его отсутствие предсказать легче. PC также одинаково наказывает за оба типа ошибок – промахи и ложные тревоги. Наихудшее PC = 0, наилучшее PC = 1.

Смещение частоты (Frequency bias index) **FBI**, атрибут Смещение

Смещением частоты называется отношение общего количества прогнозов явления к общему количеству наблюдений явления. Соответствие между средним значением прогнозов и средним значением наблюдений. Отличается от точности тем, что не учитывает соответствие между отдельными парами прогноз – наблюдение:

$$FBI = \frac{n_{11} + n_{12}}{n_{11} + n_{21}} = \frac{n_{10}}{n_{01}}.$$

Хорошее значение смещения частоты близко к 1. Значение больше 1 означает, что явление предсказывалось слишком часто, а меньше 1 – что прогноз явления давался недостаточно часто.

Оценка угроз (Threat score) **TS**, или (Critical Success Index, CSI), атрибут Точность:

$$TS = \frac{n_{11}}{n_{11} + n_{12} + n_{21}}.$$

Полезна, когда частота явления существенно ниже его отсутствия. Можно рассматривать TS как PC без учета правильных прогнозов отсутствия явления (правильных отказов) Для идеального прогноза $TS = 1$, $TS = 0$ означает бесполезный прогноз. TS оценивает, насколько хорошо соответствовал прогноз явления наблюдаемым явлениям. Наихудшее $TS = 0$, наилучшее $TS = 1$. Чувствительна к климатической частоте явления (хуже для редких явлений), потому что часть попаданий может быть чисто случайным. Может применяться, если по каким-либо причинам количество точных отказов n_{22} неизвестно.

Равноправная оценка угроз (Equitable Threat score) **ETS**, атрибут Точность:

$$ETS = \frac{n_{11} - n_{11random}}{n_{11} + n_{12} + n_{21} - n_{11random}}, \text{ где } n_{11random} = \frac{(n_{11} + n_{21})(n_{11} + n_{12})}{n_{00}}.$$

ETS оценивает, насколько хорошо соответствовал прогноз явления наблюдаемым явлениям с поправкой на случайные попадания.

Предупрежденность (Probability of detection) **POD**, атрибут Дискриминирующая способность:

$$POD = \frac{n_{11}}{n_{11} + n_{21}}.$$

Для идеального прогноза $POD = 1$, $POD = 0$ означает бесполезный прогноз.

Отношение ложных тревог (False Alarm Ratio) FAR (не путать с долей ложных тревог), атрибут Надежность:

$$FAR = \frac{n_{12}}{n_{11} + n_{12}}.$$

Для идеального прогноза $FAR = 0$, $FAR = 1$ означает бесполезный прогноз.

Отношение успеха (Success ratio) SR, атрибут Надежность:
 $SR = 1 - FAR.$

Диаграмма качества

Диаграмма качества обобщает SR, POD, FBI и TS (рисунок 1). Чем лучше прогноз, тем ближе к правому верхнему углу лежит точка.

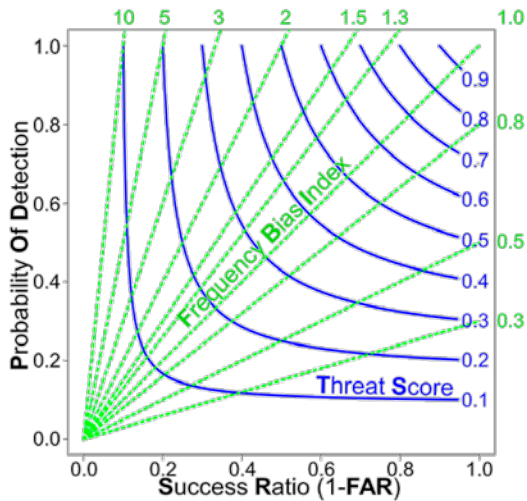


Рисунок 1. Диаграмма качества

Доля попаданий (Hit rate) H:

$$H = \frac{n_{11}}{n_{11} + n_{21}}.$$

H показывает, какая часть наблюдаемых явлений была правильно спрогнозирована. Поэтому ее еще называют предупреденностью явления (*Probability of detection*, POD). Изменяется от 0 до 1, идеальное значение 1. Может быть искусственно улучшена за счет большего выпуска прогнозов (Да), поэтому ее надо использовать совместно с оценкой, учитывающей ложные тревоги.

Доля ложных тревог (False alarm rate) F:

$$F = \frac{n_{12}}{n_{12} + n_{22}}.$$

F показывает, какая часть случаев отсутствия явления по наблюдениям была ошибочно спрогнозирована как Да. Ее иногда называют ложной предупреденностью явления (*Probability of false detection*, POFD). Изменяется от 0 до 1, идеальное значение 0. Опять же, может быть искусственно улучшена за счет меньшего выпуска прогнозов явления. Поэтому должна использоваться в паре с H.

Индекс экстремальной зависимости (Extremal Dependence Index) EDI:

$$EDI = \frac{\log F - \log H}{\log F + \log H}.$$

Симметричный индекс экстремальной зависимости:

$$SEDI = \frac{\ln F - \ln H + \ln(1 - H) - \ln(1 - F)}{\ln F + \ln H + \ln(1 - H) + \ln(1 - F)}.$$

EDI, SEDI изменяются от -1 до 1, для идеального прогноза EDI, SEDI = 1. EDI, SEDI = 0 означает бесполезный прогноз. Критерии EDI, SEDI лучше подходят для редких явлений, они не зависят от базовой доли P (частоты явления по данным наблюдений) и медленнее стремятся к нулю при стремлении к нулю P [46].

Критерий Пирса-Обухова PSS (иногда называют True Skill Statistics, TSS), атрибут Дискриминирующая способность:

$$PSS = H - F.$$

Критерий Пирса-Обухова отвечает на вопрос, насколько хорошо прогноз разделяет наличие и отсутствие явления. Использует все элементы таблицы сопряженности. Не зависит от климатической повторяемости явления. Изменяется в пределах от -1 до 1, $PSS = 0$ означает отсутствие полезного сигнала, PSS идеального прогноза равно единице.

Для мультикатегорийных прогнозов оценки строятся аналогично оценкам для дихотомных (или бинарных) событий, но для количества категорий, большего двух [76].

2.2. Ансамблевые прогнозы метеорологических полей

Верификация ансамблевого прогноза сложнее верификации детерминистского прогноза. Вместо детерминированного прогноза ансамбль выдает вероятность явления от 0 до 1, соответственно, менее очевидно, оправдался прогноз или нет. Поэтому для ансамблевого прогноза требуются более длинные ряды оценок для обеспечения репрезентативности каждой прогностической градации. Для $J = 2$ (наблюдается явление или нет) совместное распределение прогнозов и наблюдений имеет более сложную структуру по сравнению с детерминированным прогнозом, поскольку возможны более $I = 2$ возможных прогностических исходов (вероятностей), соответственно, больше число степеней свободы задачи верификации. Теоретически любое действительное число от 0 до 1 может быть исходом, но на практике вероятности округляют до конечного числа величин. Чаще всего прогностическая вероятность рассчитывается как отношение членов ансамбля, предсказывающих явление, к общему количеству членов ансамбля.

В Наставлении по Глобальной системе обработки данных и прогнозирования [15], Приложении 2.2.35 «Стандартные критерии оценки для верификации продукции глобальной системы ансамблевого прогноза», приводится стандартный набор показателей верификации для выпускаемых центрами ГСОДП продукции ансамблевых численных прогнозов (см. приложение 3). Можно рассматривать данные рекомендации как базовый уровень верификации ансамблевого прогноза. Ниже приведены наиболее широко используемые оценки.

Сравнительная оперативная характеристика (Operating characteristic relative) ROC

Метод ROC представляет собой расширение методики интерпретации таблицы сопряженности для вероятностных прогнозов. Весь диапазон вероятностей от 0 до 100 % можно разбить на интервалы пороговыми значениями вероятностей (0–10 %, ..., 90–100 %). Как правило, такой метод применяется в тех прогностических центрах, где много членов ансамбля. Когда членов ансамбля немного, удобнее разбить диапазон вероятностей на $M+1$ категорий по количеству членов ансамбля, предсказывающих явления – вероятности $0/M, 1/M, \dots, M/M$ для ансамбля из M членов.

Пороговые значения как бы переводят вероятностный прогноз в детерминированный. В таком случае прогноз с вероятностью явления больше порогового значения расценивается как прогноз наличия явления, прогноз с вероятностью меньше порогового значения – прогноз отсутствия явления [76]. Рассмотрим более подробно получение ROC-кривых и то, как выглядит расширенная таблица сопряженности для вероятностного прогноза (таблица 3).

Таблица 3. Общая таблица сопряженности признаков для вероятностных прогнозов бинарных событий

Номер прогностической вероятностной категории	Распределение членов ансамбля	Наблюдения		
		Да	Нет	
1	$F=0, NF=M$	O_1	NO_1	$T_1 = O_1 + NO$
2	$F=1, NF=M-1$	O_2	NO_2	$T_2 = O_2 + NO_2$
3	$F=2, NF=M-2$	O_3	NO_3	$T_3 = O_3 + NO_3$
...
n	$F=n-1, NF=M-n+1$	O_n	NO_n	$T_n = O_n + NO_n$
...	
N	$F=M, NF=0$	O_N	NO_N	$T_N = O_N + NO_N$
		$O_{Total} = \sum_{i=1}^N O_i$	$NO_{Total} = \sum_{i=1}^N NO_i$	
				$T = O_{Total} + NO_{Total} = \sum_{i=1}^N T_i$

M – количество членов ансамбля;

n – номер вероятностной категории; n изменяется от 1 до $N=M+1$;

F – количество членов, предсказывающих осуществление явления;

NF – количество членов, предсказывающих отсутствие явления;

O_{Total} – общее количество случаев осуществления явления в выборке;

NO_{Total} – общее количество случаев отсутствия явления в выборке;

T – общее количество прогностических случаев в выборке.

Прогностическая вероятность $P = F/M$.

$O_n = \sum W_i(O)_i$. $(O) = 1$, когда явление прогнозируется как попадающее в категорию n и наблюдается в реальности; 0 – в противном случае. Суммирование производится по всем узлам сетки (тогда значения умножаются на вес $W = \cos \phi$, где ϕ – широта) или станциям.

$NO_n = \sum W_i(NO)_i$. $(NO) = 1$, когда явление прогнозировалось, но не наблюдалось; 0 – в противном случае. Суммирование производится по всем узлам сетки или станциям.

Доли попаданий (Hit Rates, HR) и ложных тревог (False Alarm Rates, FAR) рассчитываются для каждого порога вероятности:

$$HR_n = \frac{\sum_{i=n}^N O_i}{\sum_{i=1}^N O_i}, \quad FAR_n = \frac{\sum_{i=n}^N NO_i}{\sum_{i=1}^N NO_i},$$

где n изменяется от 1 до N . Показатель HR изменяется от 0 до 1, при этом желательным является более высокое значение; FAR изменяется от 0 до 1, желательным является более низкое значение.

Значения HR и FAR для каждого порога вероятности обеспечивают N точек на графике HR (вертикальная ось) как функции FAR (горизонтальная ось) для вычерчивания ROC-кривой. Эта кривая, по определению, должна проходить через точки $(0,0)$ и $(1,1)$ (соответственно, для событий, прогнозируемых только с $>100\%$ значениями вероятности (чего никогда не бывает) и для всех значений вероятности, превышающих 0% (бывает всегда). Неквалифицированные прогнозы указываются диагональной линией (где $HR=FAR$); чем дальше кривая пролегает в направлении

верхнего левого угла (где HR=1, а FAR=0), тем лучше. Мерой качества прогноза является площадь под этой кривой. Площадь, равная 0.5, соответствует диагональному графику, при котором доли попаданий равны долям ложных тревог. Можно считать хорошей прогностическую систему с ROC > 0.8 (80 %), а с ROC > 0.7 (70 %) – полезной. Метод ROC оценивает дискриминирующую способность прогностической системы, т. е. способность хорошо различать наличие и отсутствие явления, но не обладает чувствительностью к смещению [76]. Его следует применять в комбинации с оценками надежности.

Оценка Брайера, диаграмма надежности, гистограмма частот. Оценка Брайера – это аналог среднеквадратической ошибки MSE для вероятностных прогнозов. С помощью оценки Брайера оценивается разность между прогностической вероятностью явления P и его реализацией по данным наблюдений O, равной 0 (явления не было) или 1 (явление было):

$$BS = \frac{1}{T} \sum_{k=1}^T (P_k - O_k)^2,$$

где T – общее количество прогнозов в выборке.

Как и MSE, BS отрицательно направлена (чем меньше, тем лучше), изменяется от 0 до 1; чем меньше BS, тем лучше.

Можно сформулировать меру мастерства Брайера по отношению к эталонному прогнозу BS_{ref} (часто в качестве эталонного берут климатический прогноз):

$$BSS = 1 - \frac{BS}{BS_{ref}}.$$

BS можно разложить на три составляющие, характеризующие различные аспекты системы: надежность, разрешение и неопределенность (разложение Мёрфи):

$$\begin{aligned} BS &= \left(\frac{1}{T} \sum_{i=1}^T T_i (p_i - o_i)^2 \right) - \frac{1}{T} \sum_{i=1}^N T_i (o_i - o)^2 + (1 - o)o = \\ &= \underbrace{(p_i - o_i)^2}_{\text{надежность}} - \underbrace{(o_i - o)^2}_{\text{разрешение}} + \underbrace{(1 - o)o}_{\text{неопределенность}} \end{aligned}$$

где i – номер прогностической вероятностной категории;

p_i – прогностическая вероятность в данной категории;

$o_i = O_i / (O_i + NO_i)$ – частота наблюдаемого явления в данной категории;

$o = O_{Total} / T$ – частота наблюдаемого явления по всей выборке (при достаточно большом количестве случаев ее можно считать приближением климатической частоты явления).

Остальные обозначения указаны в таблице 3.

Надежность – среднеквадратическая разность между прогностической вероятностью p_i и наблюдаемой частотой o_i . Чем меньше значение этого члена, тем лучше. Она означает способность системы прогнозировать точные вероятности. *Диаграмма надежности* строится исходя из зависимости частоты наблюдаемого события в данной вероятностной категории $o_i = O_i / (O_i + NO_i)$ от вероятности p_i , $i = 1, \dots, n$. Идеальным графиком является прямая под углом 45 градусов, когда $p_i = o_i$.

Разрешение – среднеквадратическая разность между наблюдаемой частотой в каждой категории o_i и средней наблюдаемой частотой явления во всей выборке o . На диаграмме надежности находится как расстояние между o_i и горизонтальной прямой, проведенной через значение o .

Неопределенность зависит только от изменчивости наблюдений и характеризует трудность прогностической ситуации. Она не зависит от качества прогностической системы.

Гистограммы частот строятся по тем же таблицам сопряженности и показывают частоту прогнозов как функцию вероятностей. Частота прогнозов F_i для вероятностной категории i определяется следующим образом: $F_i = (O_i + NO_i) / T$, где T является суммарным количеством прогнозов. Системы, для которых большая часть прогнозов попадает в вероятностные категории 0 и 1, близки к детерминистскому прогнозу, при климатическом прогнозе прогнозируется все время одна и та же климатическая вероятность.

Вероятностные прогнозы могут формулироваться для дискретных явлений с количеством возможных исходов более двух. Такие прогнозы называются **мультикатегорийными**. Категории мультикатегорийных прогнозов могут быть упорядоченными (когда очевидно, какой из исходов больше, а какой меньше) или номинальными, в которых отсутствует упорядоченность. Подходы к верификации упорядоченных и номинальных прогнозов различаются, для первых важна величина (*magnitude*) ошибки. Размерность задачи верификации мультикатегорийных вероятностных

прогнозов очень велика, и из практических соображений ее стремятся сократить путем оценивания с помощью небольшого количества скалярных метрик, однако следует помнить, что описание качества прогноза с их помощью будет по определению неполным.

Одна из наиболее популярных оценок для мультикатегорийных прогнозов, чувствительная к расстоянию, т. е. к удаленности категории прогноза от фактической категории, и таким образом частично учитывающая величину ошибки – **Ранговая вероятностная оценка** (Ranked Probability Score) **RPS**:

$$RPS = \frac{1}{M-1} \sum_{m=1}^M \left(\sum_{k=1}^m p_k - \sum_{k=1}^m o_k \right)^2,$$

где M – количество прогностических категорий; p_k – прогностическая вероятность в категории k и o_k – бинарный индикатор события ($0 = \text{нет}$, $1 = \text{да}$) по наблюдениям в категории k .

Отвечает на вопрос: насколько хорошо вероятностный прогноз предсказывает категорию, в которую попадает наблюдение. Изменяется от 0 до 1, идеальное значение 0. Измеряет сумму квадратных разностей в пространстве кумулятивной вероятности для мульти-категорийных прогнозов. Чем дальше прогностическая вероятность от наблюдаемого исхода, тем больше штрафуются прогнозы. Для дихотомного прогноза RPS превращается в оценку Брайера BS.

Для вероятностных прогнозов непрерывных величин наиболее широко используется **Непрерывная ранговая вероятностная оценка** (Continuous Ranked Probability Score) **CRPS**, которая является RPS при количестве категорий явлений, стремящемся к бесконечности:

$$CRPS = \int_{-\infty}^{\infty} [F(y) - F_o(y)]^2 dy,$$

где $F_o(y) = \begin{cases} 0, & y < o \\ 1, & y \geq o \end{cases}$ – кусочно-заданная кумулятивная функция

распределения вероятности, делающая скачок от нуля до 1 в момент, когда прогностическая переменная y равна значению наблюдения o .

CRPS отрицательно ориентирована (т. е. чем меньше, тем лучше), идеальное значение 0.

На рисунке 2а приведен пример с тремя гипотетическими вероятностными прогнозами в виде Гауссовской функции плотности распределения $f(y)$ (1, 2, 3), а также значение наблюдения.

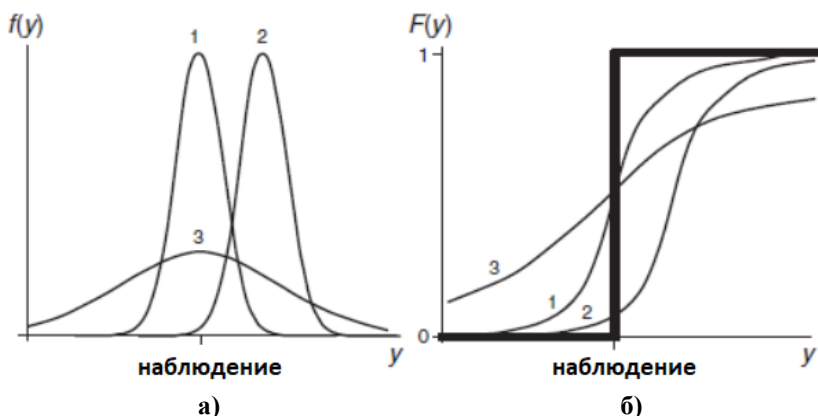


Рисунок 2. Пример с тремя гипотетическими вероятностными прогнозами в виде: Гауссовской функции плотности распределения $f(y)$ (1, 2, 3), а также значение наблюдения y (а) и соответствующие прогностические кумулятивные функции распределения вероятности $F(y)$ и кусочно-заданная кумулятивная функция распределения вероятности $F_o(y)$, делающая скачок от нуля до единицы на значении наблюдения (жирная линия) (б) [76]

Прогноз 1 центрирован на значении наблюдения и большая часть плотности вероятности сосредоточена около значения наблюдения, функция плотности распределения узкая, т. е. прогноз обладает высокой контрастностью. Прогноз 2 обладает такой же контрастностью, но сильно сдвинут относительно наблюдения. Прогноз 3 также центрирован на значении наблюдения, но функция плотности вероятности шире. На рисунке 2б приведены соответствующие прогностические кумулятивные функции распределения вероятности $F(y)$ и кусочно-заданная кумулятивная функция распределения вероятности $F_o(y)$, делающая скачок от нуля до единицы на значении наблюдения (жирная линия). Поскольку CRPS представляет собой интеграл квадратных разностей, наилучшую оценку будет иметь прогноз с кумулятивной функцией распределения, являющейся наилучшим приближением кусочно-заданной функции $F_o(y)$, т. е. прогноз 1.

Также рассчитывают меры мастерства RPS и CRPS относительно референсного, чаще всего климатического прогноза (RPSS и CRPSS).

Метод ранговых гистограмм оценивает откалиброванность вероятностного прогноза. Был разработан независимо [25, 57, 58, 75] и иногда называется Метод диаграмм Талаграна. В этом методе учитывается попадание фактической величины F в пределы одной из $m+1$ градаций ранжированного ряда из m членов ансамбля в предположении независимости членов ансамбля друг от друга. В каждый прогностический эпизод имеется ансамбль значений прогностической величины: x_1, \dots, x_m . Из этого ряда строится ранжированный ряд величин $x_{(1)} \leq \dots \leq x_{(m)}$. Делаются некоторые стандартные соглашения относительно номеров градаций при равенстве членов ранжированного ряда и при попадании фактического значения на границу градации. Значение гистограммы в каждой градации определяется из условия:

$$r_1 = P(F < x_{(1)}),$$

$$r_j = P(x_{(j-1)} \leq F < x_{(j)}), j = 2, m,$$

$$r_{m+1} = P(x_{(m)} \leq F),$$

где P – вероятность попадания F в каждую из градаций $j = 1, \dots, m+1$, т. е. число случаев попадания факта (F) в пределы градации, деленное на общее количество случаев (рисунок 3).



Рисунок 3. Пример ранговой гистограммы

Теоретически функция распределения прогностического ансамбля, построенного на основе идеальной модели, совпадает с функцией распределения фактической величины. Это выражается в строго равномерном (горизонтальном) расположении результирующей кривой гистограммы. При этом фактическая величина рассматривается как $(m+1)$ -й член ансамбля. Следует иметь в виду, что равномерное расположение кривой является необходимым, но недостаточным условием идеальной ансамблевой системы. С помощью случайного прогноза также можно получить равномерную кривую. Однако отличие кривой гистограммы от равномерной является четким признаком несовершенства прогностической системы. Более или менее равномерное расположение кривой свидетельствует об удовлетворительном качестве построенной ансамблевой схемы. Если большая часть случаев попадает в область меньших значений, ансамбль имеет тенденцию завышать реальное значение, если большая часть случаев попадает в область наибольших значений – ансамбль занижает реальное значение. Таким образом, пики в ранговой гистограмме обнаруживают систематику используемой модели. U-образная форма гистограммы означает, что прогностическое распределение слишком узкое и наблюдения попадают чаще всего в хвосты этого распределения, особенно в левую часть, т. е. большая часть членов ансамбля прогнозирует количество осадков выше значения по наблюдениям. Подобные недостатки можно частично исправить калибровкой ансамблевого прогноза.

2.3. Особенности верификации прогнозов экстремальных величин

2.3.1. Верификация прогнозов экстремальных явлений и величин

Экстремальные метеорологические явления – это явления редкой повторяемости в отношении их интенсивности, продолжительности или времени возникновения. Дальнейшее уточнение этого очень общего определения может быть весьма формальным и состоит в определении соответствующих пороговых значений. К экстремальным явлениям относятся многие опасные метеорологические явления, т. е. явления погоды со значительными воздействиями.

При оценках качества экстремальных явлений или величин наряду со стандартными характеристиками таблиц сопряженности (например, F, FAR, HR, PSS, TS, ETS (см. п. 2.1.2)) рекомендуется использовать показатели типа "**индексов экстремальной зависимости**" (EDI, *Extremal Dependence Indices*, и его симметризованный вариант SEDI) и их доверительные интервалы. Правильный выбор статистического распределения для "моделирования" экстремумов особенно важен в ситуациях, когда распределение имеет "толстый хвост". Так, в [59] приводится пример, когда по архиву суточных осадков на станции Форт Нокс (США, 1948–2010) оценивался период повторяемости пика суточной суммы осадков (6.18 дюйм, однажды летом 1997 г.) по предыдущим данным (1948–1996). Стандартная оценка – по гамма-распределению, с учетом экстремальности – по обобщенному распределению экстремальных величин (GEV). Результат оценки: по гамма-распределению – порядка 140 тыс. лет; с помощью GEV (тип распределения Фреше) – порядка 100 лет. Различия в оценках ожидаемой частоты рекордных осадков – более чем в тысячу раз (а это означает пересмотр оценок рисков и стратегия реагирования на потенциальные паводки).

2.3.2. Формирование выборок для анализа экстремальных явлений на примере метода блочных максимумов

Объясним детали на примере максимальных суточных сумм осадков: если "блоком" считается год, то независимость обеспечена, если же месяц, то возникает явная необходимость учета сезонного хода, а это нарушение стационарности [76].

Здесь используются классические модели трех "экстремальных типов" распределений (Гумбеля, Фреше и Вейбулла). При нарушении стационарности подключается временная переменная к параметрам распределения и оценка усложняется. Заметим, что непросты проблемы оценок параметров распределений разными методами – моментов, максимального правдоподобия, L-моментов, байесовских приемов генерирования марковских цепей и т. д. Из всех параметров GEV наибольшая неопределенность заключается в оценках параметра формы – "критического параметра" распределения, определяющего тип хвоста.

Используются и другие методы формирования выборок для анализа экстремальных явлений. Так, метод максимумов

над порогом, или пиков над порогом (POT, Peaks Over Threshold) использует обобщенное распределение Парето (Generalized Pareto distribution), логарифм которого в точности равен обобщенному экстремальному распределению (Generalized extreme value distribution) с другим параметром масштаба, но с тем же параметром формы.

2.3.3. Пример верификации прогнозов явлений со значительными воздействиями: отечественный опыт

Явления со значительными воздействиями (особенно редкие явления) представляют большую трудность для моделирования и прогнозирования. В [47] отмечается сложившаяся к настоящему времени парадоксальная ситуация, когда потребность в прогнозах таких явлений неуклонно возрастает, а подходящие методы верификации отсутствуют. В практике расчетов стандартных показателей качества давно отмечены такие особенности, как выборочная неустойчивость и вырождение к одному значению при уменьшении частоты явления, проявляющиеся, например, в изломанности кривых на диаграмме надежности ансамблевых прогнозов и в стремлении к 0 или 1 критерия Пирса-Обухова при стремлении частоты явления к нулю [13].

В [45] предлагается перечень таких особенностей оценок прогнозов редких явлений, которые лишены вышеупомянутых недостатков и отвечают требованиям метаверификации:

- + невырожденность предела (nondegenerate limit),
- + независимость от частоты явления (base-rate independent),
- + трудность подгонок (nontrivial to hedge),
- + регулярность (regularity),
- + ограниченность диапазона интервалом $[-1,1]$ (fixed range),
- + асимптотическая объективность (asymptotically equitable),
- + статистическая обоснованность определения (meaningful original),
- + дополнительная симметричность (complement symmetry),
- + транспонированная симметричность (transpose symmetry).

Среди показателей качества прогнозов редких явлений, отвечающих вышеперечисленным требованиям, наиболее простыми являются показатель EDI и его симметризованный вариант SEDI.

Экстремальные индексы качества прогнозов (EDI).

В 2013 году группа по верификации ВМО рекомендовала использование EDI для оценки качества прогнозов интенсивности осадков [40] с целью накопления опыта при использовании таких мер. Практически все желаемые свойства показателей качества, перечисленные в перечне [45], присущи индексу EDI. Что касается объемов выборок, обеспечивающих репрезентативность, то учитывается тот факт, что доверительные интервалы для характеристик таблиц сопряженности рассчитываются в предположении нормальности распределения, поэтому оценки этих интервалов мало зависят от объема выборки, если объем превышает примерно 100 пар прогноз – наблюдение.

В главе, посвященной оценке детерминистских прогнозов бинарных величин, Хоган и Мейсон [47] приводят примеры взаимозависимости разных мер качества, включая SEDI, в координатах (F, H) и в координатах (BIAS, PSS). Так как многие статистические характеристики таблиц сопряженностей выражаются через F и H нелинейно, как например TS, то можно получить одно и то же значение этих характеристик при самых разных комбинациях F и H. На основе изоплет показателей EDI и PSS в единых координатах (F, H) нелинейность взаимосвязи между экстремальным индексом EDI и критерием Пирса-Обухова PSS. Отсутствие строгой монотонности во взаимной зависимости показателей EDI и PSS приводит к тому, что возможны две комбинации значений (F₁, H₁) и (F₂, H₂), для которых PSS₁ > PSS₂, а EDI₁ < EDI₂ (показано на рисунке 4 синими линиями координат и синими квадратами на двух парах изоплет EDI и PSS). Этим же способом можно рассчитать оценки нижней границы EDI при заданном уровне критерия Пирса, которые оказываются всегда выше значений PSS для любых одинаковых пар (F, H).

2.3.4. Некоторые практические выводы о свойствах индексов экстремальной зависимости

В тех случаях, когда такие явления выделяются не на основе предписанных порогов, а на основе редкости наступления, то трудности анализа и прогноза таких, как правило, экстремальных явлений возникают и при оценке качества прогнозов. Все ранее известные и рутинно используемые показатели качества либо сходятся

к тривиальным значениям, либо ведут себя слишком изменчиво и неопределенно при переходе к крайним пороговым значениям. Все это исключает возможность сравнения разных моделей и схем, которые участвуют в прогнозировании экстремальных явлений.

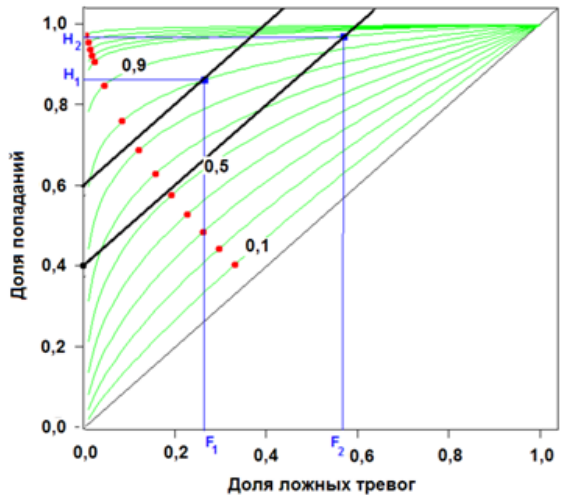


Рисунок 4. Изоплеты показателей EDI (зеленые кривые) и PSS (толстые черные линии) в координатах (F, H). Изолинии EDI проведены для значений от 0,1 до 0,9 через 0,1 и от 0,95 до 0,99 через 0,01. Изолинии PSS проведены для значений 0,4 и 0,6.

Диагональ $H = F$ соответствует значениям $EDI = PSS = 0$. Красным цветом обозначены точки касания кривых EDI и прямых линий PSS. Синим цветом выделены точки (F, H), для которых нарушается монотонная зависимость между EDI и PSS

Использование достижений теории экстремальных величин позволило в последние годы сформулировать основные требования к критериям качества прогнозов редких явлений и из предлагаемых "экстремальных индексов" выбрать только те, которые оказываются пригодными для практики верификации. Одним из них является "индекс экстремальной зависимости" EDI, имеющий превосходство перед такими часто используемыми мерами, как "критический индекс" TS и критерий Пирса-Обухова PSS.

В практике верификации при прогнозе метеовеличин имеются разнообразные особенности изменения EDI в зависимости от порогового значения, причем отдельные стороны этого изменения далеки от многообещающих перспектив асимптотических теорий. Очевидно, что в последних случаях речь идет о выборочных эффектах накопленных пар прогноз – наблюдение, в которых отражаются ошибки наблюдений, несовершенства моделей, недостаточные объемы данных и неприменимость асимптотических статистических предположений.

Несмотря на иногда неожиданные и неудачные случаи использования показателя EDI в практике оценок качества, следует продолжать накопление опыта, чтобы научиться содержательно интерпретировать результаты верификации прогнозов явлений со значительными воздействиями.

3. Пространственные методы верификации. Проблема двойного штрафа и способы ее преодоления

Как сказано выше, при оценке полей количество точек сетки по пространству обычно велико и количество значений (категорий), которые могут принимать прогнозы и наблюдения, тоже может быть большим, поэтому размерность задачи верификации пространственных полей огромна. Использование скалярных (одно число) метрик, таких как MSE, POD, FAR, TS, PSS и т. д., облегчает задачу, но не решает всех проблем. Например, если мы оцениваем явление Наличие осадков, все прогнозы с отсутствием пересечения областей осадков в прогнозе и наблюдениях, но с одинаковым количеством точек с явлением, будут иметь одинаковые показатели, независимо от расстояния между объектами в полях прогноза и наблюдений, интенсивности явления в объекте, его формы, наклона оси и т. п.

В настоящее время все шире используются модели высокого пространственного разрешения (2 км и выше). В целом, увеличение разрешения модели негативно сказывается на результате оценок для некоторых метеозлементов. Увеличенное разрешение на самом деле дает более точно очерченные мезоструктуры, большую амплитуду изменения метеозлементов и большие градиенты, что неизбежно ведет к большим ошибкам в пространстве и времени. Это свойство известно как проблема "двойного штрафа": при стандартном оценивании "точка с точкой", когда прогноз явления незначительно сдвинут по сравнению с наблюдаемым явлением, он штрафует дважды: в точке, где явление прогнозировалось, но не случилось в реальности, и там, где оно было, но не спрогнозировано. Тем не менее такая модель дает больше полезной информации. При этом возникает потребность в методах верификации, которые обладают некоторой терпимостью к умеренным ошибкам в пространстве и времени, сохраняя при этом способность оценивать истинное качество информации, которую дает модель высокого разрешения.

Традиционные оценки, сравнивающие прогноз и наблюдение в точке, имеют недостатки при оценке прогнозов высокого

разрешения. Например, на рисунке 5 все прогнозы, кроме (е), дадут оценку угроз $TS = 0$, хотя (е) субъективно не является лучшим прогнозом.

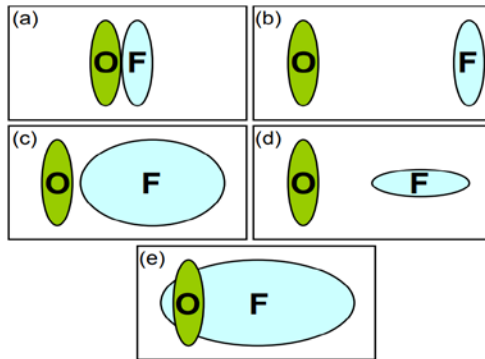


Рисунок 5. Гипотетическая область прогноза (F) и наблюдения (O) явления [28]. При стандартном оценивании «точка с точкой», когда прогноз явления незначительно сдвинут по сравнению с наблюдаемым явлением, он штрафует дважды: в точке, где явление прогнозировалось, но не случилось в реальности, и там, где оно было, но не спрогнозировано

Во многих ситуациях (а) является лучшим прогнозом, хотя будет иметь такие же плохие стандартные оценки, как и откровенно неправильный прогноз [28]. В настоящее время существует ряд методов, позволяющих уменьшить влияние двойного штрафа. Такие методы традиционно называются пространственными (spatial methods) [51]. Однако некоторые из них позволяют учесть также сдвиг по времени. Они не наказывают незначительные сдвиги прогноза относительно наблюдения в пространстве и времени, т. е. позволяют избежать двойного штрафа. Также пространственные методы помогают снизить размерность задачи верификации.

Таким образом, традиционную оценку мезомасштабных прогнозов рекомендуется дополнять методами и приемами пространственной верификации, особенно это касается разрывных полей, например осадков, облачности. Однако при шаге сетки современных моделей ЧПП по пространству порядка 1 км детали

топографии и мелкомасштабные детали стали проявляться не только в осадках, но и в других пространственных полях.

Обилие предлагаемых пространственных методов и не всегда понятные свойства порождаемой ими информации послужили стимулом к их целенаправленному анализу в рамках международных проектов, среди которых наиболее значимыми представляются Проект ICP по сравнению методов пространственной верификации (spatial forecast verification Inter-Comparison Project, ICP, время выполнения 2007–2009 гг.) и развивающий его результаты Проект MesoVICT по сравнению методов мезомасштабной верификации над сложным рельефом (Mesoscale Verification Inter-Comparison over Complex Terrain, время выполнения 2014–2019 гг.) [38]. Подчеркнем, что речь идет не о средствах верификации (данных, показателях и расчетных алгоритмах), а о таких свойствах средств верификации ("атрибутах"), которые входят в состав "метаверификации" – науки о методах и о системах верификации [64]. Данные проекты обсуждаются в [3, 5]. Результаты проектов уже находят практическое применение в исследованиях и в оперативной деятельности метеорологических центров и служб. При этом многие методологические вопросы пространственной верификации еще нуждаются в уточнениях, некоторых из них еще далеки от решения.

В самом начале выполнения проекта ICP была проведена классификация почти всех разработанных до этого времени методов на две большие категории – на методы фильтрации и методы смещения (filtering methods and displacement methods). В свою очередь методы фильтрации были подразделены на две группы – окрестные и масштабно-разделительные (neighborhood and scale separation), а методы смещения – на объектно-ориентированные и деформационные (features based and field deformation) [50]. В последнее время отдельно выделяют меры дальности (distance metrics) [39].

Приведем краткое описание групп пространственных методов, которые прочно вошли в мировую практику верификации: методы фильтрации (окрестные и масштабно-разделительные) и объектно-ориентированные методы.

Окрестные (neighborhood) методы. Раньше эти методы чаще назывались нечеткие (fuzzy). Их основное свойство – доверие к прогнозу в окрестности [42]. Оценка соответствия прогноза

явления и наблюдения в точке замещается оценкой прогноза и/или наблюдения в окрестности точки, которая предполагает какую-либо обработку данных внутри окна (окрестности), например

- среднее значение (апскейлинг – upscaling);
- наличие явления где-нибудь внутри окна;
- частота явления в окне – аналог вероятности;
- распределение величин в окне.

Если в окне рассматривается также наблюдение – это позволяет косвенно учесть неопределенность наблюдений.

Возможно, самый популярный окрестный метод – **Оценка по пространственным долям** (*Fractions Skill Score*, FSS) [68]. Рассматривается доля ячеек в окрестности, занятых явлением, (P) в поле прогноза и поле наблюдений (явление чаще всего определяется с помощью порогового значения исследуемой метеопеременной). Далее рассчитывается среднеквадратическая разность P прогноза и наблюдений по всем точкам области оценивания N . В знаменателе FSS фигурирует наилучший прогноз:

$$FSS = 1 - \frac{\frac{1}{N} \sum_{i=1}^N (P_{fcst} - P_{obs})^2}{\frac{1}{N} \sum_{i=1}^N P_{fcst}^2 + \frac{1}{N} \sum_{i=1}^N P_{obs}^2}$$

FSS изменяется от 0 до 1. По мере роста окна FSS стремится к отношению прогностической и наблюдаемой частот явления. При отсутствии смещения ($FBI = 0$) в прогнозах FSS стремится к 1. Величина FSS полезного прогноза задается больше $FSS_{uniform} = 0,5 + f_o/2$, где f_o – доля ячеек с явлением по наблюдениям во всей области оценивания. Оценку FSS следует рассчитывать для размеров окрестности возрастающей величины: в центральной ячейке, окрестности 3×3 ячейки, 5×5 , 7×7 и т. д. и для разных пороговых значений исследуемой величины, определяющих явление (рисунок 6). FSS наиболее чувствительно к редким явлениям.

С помощью FSS мы хотим узнать:

- Как изменяется мастерство прогноза с ростом окрестности?
- Какой наименьший размер окрестности дает достаточно точные прогнозы?
- Дает ли модель более высокого разрешения более точные прогнозы?

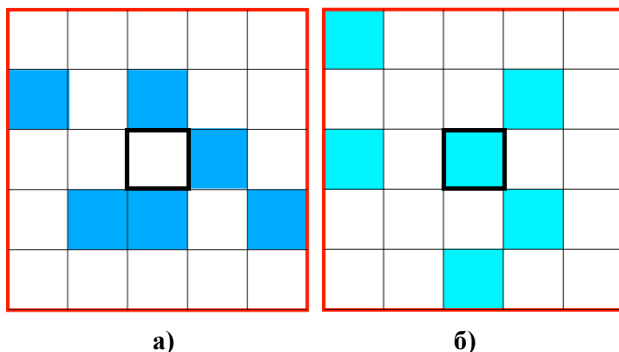


Рисунок 6. Доля ячеек, занятых явлением (закрашенные ячейки), в окрестности 5x5 точек одинакова для прогноза (а) и наблюдения (б), и равна $6/25=0.24$, в то время как оценка только в центральной ячейке окрестности (выделена жирной черной рамкой) дала бы плохие оценки (по прогнозу явления нет, а по наблюдению есть)

В [72] показан способ использования FSS для оценки смещения прогностической области явления в пространстве относительно наблюдений. Метод основан на соотношении между FSS, размером окрестности и смещением по пространству. Показано, что метод работает только при отсутствии большого сдвига частоты (FBI, frequency bias) в прогнозе, конкретно при значениях FBI от 0.75 до 1.5. Однако это ограничение можно преодолеть заданием порогового значения на основе процентиля исследуемой переменной. Г. Скок [73] ввел модификацию FSS, окрестная мера мастерства – Neighborhood Skill Score, NSS. Оценка смещения NSS не ограничена значениями FBI, что делает ее более удобной в использовании.

При применении FSS и ее модификаций следует иметь в виду, что эти оценки дают неточные результаты для объектов, близких к границе области.

Масштабно-разделительные методы (scale-separation methods) – оценивают ошибку прогноза в зависимости от масштаба. Наиболее известный метод - Разложение по масштабам ошибки прогноза (*Intensity-scale verification approach*, IS) [30]. Бинарная ошибка прогностического поля относительно поля наблюдений раскладывается на сумму компонентов разных масштабов, как правило, с помощью вейвлет-фильтра (например, вейвлет 2D Наар).

Для компонентов каждого из масштабов рассчитывается ряд оценок (среднеквадратическая мера мастерства, оценка Брайера и т. п.)

Отметим, что различие окрестных и масштабно-разделительных методов состоит в том, что окрестные методы отфильтровывают ошибку на мелких масштабах, а масштабно-разделительные методы выделяют ошибку на каждом из масштабов.

Объектно-ориентированные методы. "Объектно-ориентированная верификация" (object-based или features-based) предназначена для оценки способности прогностической системы воспроизводить структурные элементы полей вместе с их перемещением и эволюцией [33, 41, 61]. Здесь под "объектами" понимаются пространственные, в основном односвязные области полей на земной поверхности, выделяемые некоторым пороговым значением исследуемой величины. При этом важно отличать геометрические объекты, выделяемые формальным образом через неравенства, от объектов, определяемых характерными пространственно-временными масштабами соответствующих физических величин.

Основные и первые примеры применения пространственной верификации относятся к осадкам в силу их высокой пространственно-временной изменчивости, которая затрудняет идентификацию и прогнозирование областей осадков и делает осадки своего рода "метрикой качества гидродинамических моделей".

CRA (Contiguous Rain Area), Сопряженная область дождя – объектно-ориентированный метод верификации прогнозов, предложенный в [41, 43]. Вкратце он состоит в следующем: вначале в полях наблюдений и прогноза выделяются объекты (чаще всего осадков) на основе порогового значения. Затем с помощью той или иной функции (в которых используется расстояние между объектами) выделяются пары объектов в поле наблюдений и прогнозов. Затем определяется оптимальный сдвиг прогностического объекта в пространстве к наблюдаемому объекту путем минимизации ошибки. Здесь также могут применяться различные критерии, например коэффициент корреляции или среднеквадратичная ошибка MSE. Находится разность исходной MSE (представляемой как $MSE_{total} = MSE_{displacement} + MSE_{volume} + MSE_{pattern}$) и MSE после сдвига, это ошибка из-за сдвига в пространстве $MSE_{displacement} = MSE_{total} - MSE_{shifted}$. Оставшаяся MSE состоит из квадрата разности среднего количества осадков

в прогностическом и наблюдаемом объекте $MSE_{volume} = (F - X)^2$ и мелкомасштабных различий $MSE_{pattern} = MSE_{shifted} - MSE_{volume}$. Пример применения компонента смещения CRA для оценки наукастинга ФГБУ «Гидрометцентр России» можно найти в [14].

MODE (Method for Object-Based Diagnostic Evaluation). Метод объектно-ориентированной диагностической верификации [33, 34] предназначен для сравнения сеточных прогнозов и сеточных наблюдений. Метод был разработан, чтобы приблизить объективный расчет качества прогноза объектов к субъективному суждению синоптика. Метод MODE может использоваться для решения общих задач сравнения любых двух полей, содержащих однозначно определенные объекты.

Расчеты оценок с помощью MODE производятся в следующей последовательности:

1. Определение объектов в полях прогноза и наблюдения с помощью пользовательских параметров сглаживания исходного поля (опционально) и порогового значения.

2. Расчет атрибутов для каждого из этих объектов: например площади, центроида ("центра масс"), угла наклона оси и др. Выбираются атрибуты, важные для цели анализа.

3. Расчет для каждой пары прогноз – наблюдение различия между их атрибутами: например отношение площадей, расстояние между центроидами, разность углов наклона, и т. д.

4. Выбор функций нечеткой логики для каждого из атрибутов. Смысл этих функций в том, что хорошему соответствию атрибута (например, малому расстоянию между центроидами) придается значение 1, которое уменьшается (как правило, линейно) до нуля вплоть до значения разности атрибутов, бессмысленного для пользователя (например, расстояние между центроидами слишком большое). Таким образом, все значения разностей атрибутов переводятся с помощью функций нечеткой логики в значения от нуля до 1.

5. Выбор весов w для каждого из атрибутов, а также коэффициента надежности c (отражающего, насколько мы уверены в точности определения значения атрибута. Например, при малой разности углов наклона надежность меньше; на практике обычно берут 1).

6. Вычисление суммарной важности (Total Interest, TI) для всех пар объектов в поле прогноза и наблюдений:

$$Tl_j = \frac{\sum_i I_{ij}}{\sum_i w_i c_i},$$

где i – номер атрибута, j – номер пары объектов.

7. (Опционально) Сопоставление (matching) объектов в поле прогноза и наблюдения на основе превышения выбранного порога суммарной важности и объединение (т. е. слияние) объектов внутри одного и того же поля в кластеры.

8. Вывод обобщенных статистических характеристик отдельных объектов, пар объектов и кластерных объектов. Расчет медианы максимальной суммарной важности – основной интегральной характеристики качества прогноза MODE.

Метод MODE можно адаптировать для ансамблевого прогноза [60]

Настройка параметров для идентификации объекта и обеспечения его связности

Идентификация объекта производится, как правило, выбором произвольной изолинии, очерчивающей замкнутую границу объекта.

Более сложен вопрос связности области, выделенной замкнутой линией, – учитывать ли многосвязность или устранять её? Собственно говоря, требуется выбрать такой радиус осреднения, при котором и граница станет более гладкой, и область – более односвязной, но при этом сохранятся важные для анализа области малого масштаба. Все эти проблемы возникают каждый раз при анализе и прогнозе областей мезомасштабных осадков, и удовлетворить все запросы не представляется возможным.

После пространственного осреднения полей, выделения объектов подходящей гладкости границ и близких к односвязности возникает очередная проблема статистического описания объектов. Простейшие характеристики – это центр тяжести, учитывающий интенсивность осадков, и размер, оцениваемый в количестве точек или в площади. Несмотря на сглаживание, границы зон осадков могут принимать самые разнообразные формы и для характеристик объекта, отличных от его объема (как, например, в [33]), сопоставление полос в наблюдениях и кругов в прогнозах становится затруднительно.

Применение пространственных методов для ансамблевого прогноза

В последние несколько лет применение пространственных методов для ансамблей является особенно актуальной задачей, поскольку с ростом вычислительных мощностей стал возможным ансамблевый прогноз высокого разрешения. Как и в случае детерминированного прогноза, наиболее прочно вошла в практику верификации комбинация окрестного и объектно-ориентированного подходов. Ниже приведены ссылки на методы, доказавшие свою перспективность.

В [36] **FSS** была использована для **расчета соотношения пространственного разброса по ансамблю и ошибки ансамбля**. Для этого были рассчитаны среднее значение и стандартное отклонение FSS, рассчитанной для каждой пары членов ансамбля (dispersive FSS, dFSS), а также среднее и стандартное отклонение FSS, рассчитанной для каждого члена ансамбля и наблюдения (eFSS). Высокие значения среднего dFSS означают, что все члены ансамбля предсказывают область явления близко друг от друга, соответственно, пространственный разброс по ансамблю мал. Стандартное отклонение dFSS показывает интервал значений dFSS, высокое стандартное отклонение dFSS означает, что имеются выбросы с особенно низким или высоким dFSS. Для eFSS делаются соответствующие заключения, но не для разброса внутри ансамбля, а для ошибки каждого из членов ансамбля по отношению к наблюдениям. Затем проводится анализ соотношения пространственного разброса и ошибки ансамбля аналогично стандартному анализу соотношения разброс-ошибка ансамбля (spread-error relationship).

Метод ансамблевой вероятности в окрестности (*Neighborhood ensemble probability, NEP*) и **метод максимальной ансамблевой вероятности в окрестности** (*Neighborhood maximum ensemble probability, NMEP*) [69, 70]. В NEP рассчитывается вероятность явления в каждой точке выбранной окрестности, а затем находится средняя в окрестности вероятность. В NMEP для каждой точки рассчитывается доля членов ансамбля, для которых явление предсказывается где-либо в окрестности точки. Как и обычно, явление задается с помощью порогового значения. Далее полученное поле оценивается с помощью обычных вероятностных метрик качества.

Метод псевдочлена ансамбля [60] основан на применении MODE для ансамбля. Сначала создается "псевдочлен" из наиболее репрезентативных объектов ансамбля с разными вероятностями. Выбрав пороговое значение вероятности, оставляем только объекты псевдочлена с вероятностью более порогового значения, и применяем обычные объектно-ориентированные оценки.

Наиболее популярные пакеты программного обеспечения для реализации пространственных методов

МЕТ. Система верификации прогностических моделей MET (Model Evaluation Tools), выпущенная первой версией в январе 2008 г., была разработана в США Исследовательским центром для тестирования моделей DTC (Developmental Testbed Center) и выпущенную первой версией в январе 2008 г. (см. раздел 4). Пакет MET включает возможность расчета оценок на основе окрестного подхода, метод MODE, а также метод Разложение по масштабам с помощью вейвлет-анализа (*Intensity-scale, IS*).

Комплекс **SpatialVx** ("библиотека" на языке R) является, возможно, наиболее полным и подробно документированным программным средством для большинства известных методов пространственной верификации. Комплекс написан Э. Гиллеландом в 2012 г. по плану научно-исследовательской деятельности лаборатории RAL и по результатам выполнения проекта ICP. Автор регулярно производит корректировку, включая исправление обнаруженных пользователями ошибок, и расширяет комплекс новыми возможностями [52, 53].

4. Программное обеспечение системы верификации

Желательно, чтобы "оперативная система верификации численных прогнозов погоды" была оформлена в виде прошедшего некоторую (экспертную) проверку программного обеспечения (с открытым текстом, если это не пакет из интернета), а также с удобно структурированной и содержательной веб-страницей (по аналогии с австралийским сайтом по верификации). Имеющиеся в мире пакеты программного обеспечения, которые используются для верификации [47]: IDL, IMSL, Matlab, METplus.

В настоящее время широко применяется свободно распространяемый пакет METplus (<https://dtcenter.org/community-code/metplus>), являющийся развитием американского пакета MET. Можно ожидать, что этот пакет может стать стандартом в области программного обеспечения для верификации метеорологических прогнозов, поскольку он включает большинство методов, изложенных в предыдущих главах. Система верификации прогностических моделей MET (Model Evaluation Tools), выпущенная первой версией в январе 2008 г., была разработана в США Исследовательским Центром для Тестирования моделей DTC (Developmental Testbed Center).

Система METplus содержит следующие методы верификации:

1. Стандартные оценки сравнения сеточных модельных данных с точечными стационарными наблюдениями (gridded model data to point-based observations).
2. Методы пространственной верификации по сравнению сеточных модельных данных с сеточными наблюдениями с помощью окрестных, объектных и разделяющих масштабы подходов (см. раздел 3).
3. Методы вероятностной верификации для сравнения сеточных модельных данных с точечными или сеточными стационарными наблюдениями.
4. Агрегирование результатов применения перечисленных методов верификации по времени и по пространству.

Заключение

Многоаспектность задачи верификации прогнозов и разнообразие требований к ней не оставляют возможности удовлетворить все потребности с помощью нескольких универсальных простых "рецептов". Так, например, представленную в Наставлении по краткосрочным прогнозам погоды [16, 17] административную систему оценок прогнозов общего назначения вряд ли можно считать подходящей для диагностических целей.

В данной ситуации представляется уместным выделить различные по сложности и полноте предоставляемой информации уровни верификации. В качестве базового уровня верификации можно принять набор оценок, приведенный в Наставлении по ГСОДП ВМО (см. приложения 2 и 3). В то же время, более продвинутый уровень верификации должен включать более широкий спектр оценок, включая пространственные методы (раздел 3). При этом в связи с активным внедрением в оперативных прогностических центрах Росгидромета систем мезомасштабного прогнозирования для ограниченной территории следует отметить возрастающую потребность для задач верификации в качественных сеточных данных объективного анализа с высоким пространственным разрешением. Экстремальные явления требуют расчета специальных показателей (п. 2.3).

Можно предложить следующую общую схему верификации:

- Стандартные оценки (для непрерывных и дискретных величин), агрегированные по заблаговременностям, сезонам, погодным типам.

- Подробное рассмотрение особых случаев (test-cases).

- При наличии качественных сеточных данных – пространственные методы. Наиболее популярные – оценка по пространственным долям Fractions Skill Score (принадлежащая к группе окрестных (neighborhood) методов, объектно-ориентированные (CRA, MODE), разделение масштабов (IS).

Для более полной картины о качестве продукции различных прогностических систем можно рекомендовать рассчитывать оценки, характеризующие различные аспекты качества прогнозов с учетом мирового опыта и с использованием валидированных программных средств.

ПРИЛОЖЕНИЕ 1

Гнездовой словарь английских терминов верификации с переводом на русский язык и толкованием

Необходимость составления словаря английской терминологии в верификации прогнозов погоды стала все сильнее ощущаться в последние два десятка лет, отмеченные бурным развитием систем прогнозирования и появлением новых понятий, определений и соответствующих терминов.

Однако терминологическая проблема в сфере оценок качества прогнозов и прогностических систем существовала с первых попыток верификации, о чем писал Алан Мерфи в статье, посвященной публикации в 1884 г. сержантом Финли оценок прогнозов торнадо, поправкам к этим оценкам в конце 19 века, а также растущему обилию наименований одних и тех же показателей качества по обе стороны Атлантики [64].

Единство статистических методов обработки данных в таких разных областях, как, например, метеорология, гидрология, радиолокация и медицина, кратно увеличивает количество терминологических синонимов, которые приходится упоминать и уважительно перечислять в регулярно обновляемых монографиях [47, 76]. Дополнительную остроту добавляет оснащение понятий и терминов именами национальных деятелей, которые приходится заменять описаниями при выходе публикаций на иностранном языке [37].

Естественно желание продолжать традиции в переводах терминологии и оставлять в прежнем или слегка измененном виде устоявшиеся словосочетания и сокращения. Приведем некоторые примеры значительных переводческих усилий в данной области, не демонстрируя всего разнообразия лексического материала.

Возможно, первой серьезной попыткой освоения англоязычной терминологии в верификации прогнозов был перевод книги Пановского и Брайера, выполненный в двух изданиях коллективом ГГО им. Воейкова под редакцией Л.С. Гандина [67].

В 1972 г. вышла техническая записка ВМО о методах "оценки оправдываемости прогнозов", написанная Е.М. Добрышманом на английском языке [37], за исключением двуязычного предисловия. В большом приложении содержались сведения о классификации прогнозов и о методах их оценок в СССР, что также, вероятно, является одним из ранних примеров описания отечественных реалий на английском языке.

Первый по-современному новый подход к оценкам качества вероятностных прогнозов находим в монографии Г.В. Грузы и Э.Я. Раньковой [4], в которой используется терминология западных экспертов в области верификации прогнозов, включая общие подходы метаверификации по А. Мерфи, описание атрибутов прогнозов и систем верификации, новые меры качества прогнозов с их авторским переводом.

Некоторым подспорьем служат двуязычные документы ВМО, среди которых наиболее близким по времени оказывается "Руководство по практике метеорологического обслуживания населения" с отдельной главой по верификации [78]. Однако наряду с удачными примерами перевода встречаются некорректные и даже курьезные случаи: *persistence and objective guidance* – *устойчивость и объективное управление*, вместо "инерция и объективный анализ"; *data are saved in a database* – *данные спасаются в базе данных*; *false alarms* – *фальшивые тревоги*; *verification scores* – *индексы верификации*.

Указанные в ссылках [1, 2, 4, 7, 8, 9, 10, 11, 12, 18, 19, 20, 22, 26, 31, 32, 35, 38, 39, 43, 44, 45, 46, 47, 48, 49, 51, 61, 63, 64, 65, 66, 67, 68, 76, 77, 78, 79] статьи, монографии и документы содержат выбранные нами понятия и термины, которые являются наиболее цитируемыми в современных публикациях, или же представляют исторический интерес. Отметим важное обстоятельство – все эти источники "спасены в базу данных" для возможных обсуждений с коллегами, осознающими важность терминологических проблем.

attributes (aspects) атрибуты, специфические свойства прогнозов, подлежащие проверке (верификации)

accuracy точность (атрибут качества), характеризует среднее соответствие между отдельными парами прогнозов и наблюдений; оценка Брайера, квадратичная мера сходства, сумма диагональных элементов таблицы сопряженности – примеры мер точности

aspect of forecast quality аспект (атрибут) качества прогноза

association взаимосвязь (ассоциация) между прогнозами и наблюдениями, подлежащая оценке

bias смещение, средняя ошибка, соответствие между средним прогнозом и средним наблюдением; отличается от понятия точности (accuracy), которое измеряет среднее соответствие между парами прогноз – наблюдение

calibration калибровка, эквивалент надежности (reliability);

calibration simplex симплекс калибровки, расширение диаграммы надежности для трёх-категорийного предиктанта

conjectured sharpness principle принцип гипотетической контрастности

desirable characteristic of a forecaster желательные характеристики прогнозиста

desirable forecast attributes (properties) желательные атрибуты (свойства) прогнозов

discrimination дискриминирующая способность – обратное свойство к свойству разрешения: оценка различия в условных распределениях прогнозов относительно различных значений наблюдений

- multidimensional reliability** многомерная надежность
- quality** качество (согласованность между прогнозами и соответствующими наблюдениями)
- quantitative measures of attributes** (количественные) меры атрибутов
- reliability** надежность – оценка различия в частных распределениях наблюдений и прогнозов
- resolution** разрешающая способность: оценка различия в условных распределениях наблюдений относительно различных значений прогнозов
- sharpness** контрастность прогнозов (способность прогнозов выделять детали наблюдения, эквивалент резкости – в оптике)
- skill** мастерство, квалификация – качество анализируемых прогнозов относительно точности контрольных прогнозов (климата, инерции, случая или другого методического прогноза)
- standard set of attributes: accuracy, bias, association, reliability=calibration, resolution, discrimination, sharpness, uncertainty** стандартный набор атрибутов: точность, сдвиг, связь, надежность = калиброванность, разрешающая способность, дискриминирующая способность, контрастность, неопределенность
- value** ценность (польза), выражаемая в стоимости или других показателях, в отличие от качества (quality), выражаемого числом
- contingency table** таблица сопряженности (cross-tabulated, cross-classified data)
- base rate** базовая доля, выборочная климатология, повторяемость явления
- baseline** базовая линия (основа, контрольный уровень), относительно которой оценивается данный прогноз
- BHF representation** представление полной информации таблицы сопряженностей 2×2 в триplete BHF – смещение, доля попаданий и доля ложных тревог (Bias, Hit Rate, False Alarm Rate)
- conditional distribution** условное распределение
- contingency table 2×2 (hits, false alarms, misses, correct negatives)** двухвходовая таблица сопряженности (попадания, ложные тревоги, промахи, точные отказы)
- contingency table entry** ячейка таблицы сопряженности
- counts of hits, misses and false alarms for warnings** подсчет количества попаданий, ложных тревог и промахов для (погодных) предупреждений
- counts** счетчики, подсчеты
- degeneration of measure for rare events** вырождение мер для редких явлений
- false alarm rate** доля ложных тревог (относительно количества случаев отсутствия события)

- false alarm ratio** отношение ложных тревог (относительно количества случаев прогноза отсутствия события)
- frequency bias** смещение частот (для бинарных величин – отношение количества прогнозов явления к количеству случаев явления)
- full information in the 2x2 contingency table in three well-chosen scalar attributes** полная информация таблицы сопряженностей в трех (правильно подобранных) скалярных атрибутах, например Hit Rate, False Alarm Rate, Base Rate
- hit rate** доля попаданий, отношение количества правильных прогнозов к повторяемости явления
- joint distribution** совместное распределение нескольких случайных величин, для одномерных прогноза и наблюдения – это двумерное распределение прогнозов и наблюдений
- marginal distribution** маргинальное (частное) распределение – распределение для одной случайной величины (или нескольких, но не всех)
- marginal totals** частные (безусловные) количества случаев по прогнозам и по наблюдениям
- methods for counting compound events** методы подсчета количества составных явлений
- post agreement** старое название "оправдываемости прогноза явления" (доля правильных прогнозов в общем количестве случаев прогноза явления)
- prefigurance** старое название "предупрежденности явления", (probability of detection, POD, или Hit Rate – вероятность обнаружения, или доля попаданий)
- sample size** размер выборки
- success rate** доля успешных прогнозов
- cross-validation** кросс-валидация, перекрестная проверка
- bootstrap method, bootstrapping** метод бутстреп, бутстреппинг, компьютерный метод оценки статистик по выборке (с возвратом)
- jack-knife method** метод "складного ножа", компьютерный метод оценки статистик по выборке с удалением одного элемента
- leave-one-out calculation** контроль по отдельным объектам
- metrics designed for the sample interpretation** метрики "выборочной интерпретации"; характеризуют общее поведение и статистические характеристики членов ансамбля; они не требуют вероятностных распределений, но представляют собой первый шаг к вероятностной интерпретации
- random-fold cross-validation** случайно-кратная кросс-валидация
- decision** решение, принятие решения, выбор
- cost/loss function** функция (от) отношения затрат к потерям
- decision rule** решающее правило

- decision theoretic unbiasedness** несмещенность в теории принятия решений, эквивалент состоятельности (consistency)
- decision-analytic models** модели принятия решений
- decision-theoretic setting** формулировка в терминах теории принятия решений
- large user base** наличие большого количества пользователей делает возможными проверку и контроль продукции, а, значит, и обнаружение ошибок
- loss and utility functions** функции потерь и полезности
- loss function** функция потерь
- misclassification loss** потеря от ошибки классификации
- operational value or utility of the consequences** оперативная ценность, или полезность последствий
- optimal decision making** принятие оптимальных решений
- overall value, over all users** мера суммарной ценности, по всем пользователям
- reward/penalty matrices** матрицы вознаграждения-штрафа
- risk unbiasedness** несмещенность риска, несмещенное решающее правило – свойство собственности (propriety)
- static cost-loss model** (статичная) модель затрат-потерь
- decomposition** разложение, декомпозиция
- algebraic decomposition of the Brier Score: reliability, resolution, uncertainty** алгебраическое разложение оценки Брайера на составляющие – надежность, разрешение и неопределенность
- decomposition of MSE into measures of other attributes of forecasts and observations (bias, calibration-refinement)** разложение среднего квадрата ошибки на меры других атрибутов прогнозов и наблюдений (например, смещения и калибровки-уточнения)
- decomposition of scoring rules into terms of resolution and the reliability of a forecast** разложение оценочных правил на составляющие, связанные с разрешающей способностью и с надежностью прогноза
- decomposition of scoring rules** разложение оценочных правил
- decomposition of the resolution term** разложение составляющей разрешения (разрешающей способности)
- decomposition of the skill in terms of the MSE with regard to the correlation** разложение меры мастерства в терминах среднего квадрата ошибки с учетом корреляции между прогнозами и наблюдениями
- degeneracy** вырожденность, вырождение
- desirable properties** желательные особенности
- desirable properties of scores** желательные особенности оценок
(1) truly equitable, истинно равноправные

- (2) asymptotically equitable, асимптотически равноправные
- (3) not trivial to hedge, не поддаются тривиальной подгонке (перестраховке)
- (4) base-rate independent, не зависят от базовой доли (выборочной повторяемости)
- (5) non-degenerate, не вырождаются (не сходятся к тривиальным величинам, например, к 0 или 1)
- (6) bounded, ограничены

diagnostic диагностические методы (верификации)

diagnostic methods for spatial verification диагностические методы пространственной верификации

diagnostic verification diagrams for 2×2 contingency table диаграммы диагностической верификации для двухходовых таблиц сопряженности

diagnostic verification диагностическая верификация (простые статистики и графика)

distribution распределение (вероятностей случайных величин)

conditional exceedance probabilities условные вероятности превышений

conditional probability условная вероятность

confidence interval доверительный интервал

density плотность (распределения вероятностей)

distribution of the target распределение повторяемостей объекта (прогнозирования)

distributional methods дистрибутивные методы, основанные на распределениях вероятностей и их анализе

distributions-oriented verification or diagnostic verification верификация, основанная на анализе распределений, или диагностическая верификация

frequent/infrequent event частое/нечастое явление

frequentist interpretation of chance частотная (фреквентистская, по Мизесу) интерпретация случайности

hypothesis testing проверка гипотез

interquartile range, IQR межквартильный интервал

kernel representation представление (функции плотности) с помощью ядерного сглаживания

kurtosis эксцесс, четвертый центральный момент, крутизна

location положение, расположение

outcome результат (эксперимента, наблюдения), исход

outlier выброс

positively skewed distribution распределение с положительной асимметрией, т. е. у распределения длинный правый хвост и среднее значение превышает медиану

- p-value** р-значение, р-уровень значимости, р-критерий, вероятность превышения фактически наблюдаемого результата при заданной теоретической функции распределения
- quantile for the proportion 0.5** квантиль для доли 0.5, медиана
- quantile for the proportion p, 100pth percentile** квантиль для доли p, р-процентилем: значение, стоящее на 100р% -й позиции в упорядоченной выборке
- skewness** асимметрия, третий центральный момент, скошенность (в направлении длинного хвоста)
- spread** рассеяние, разброс
- tail of the distribution: bounded/unbounded, long, fat, heavy/light, super-heavy...** хвост распределения: ограниченный/неограниченный, длинный, жирный, тяжелый, легкий, сверхтяжелый...
- univariate and multivariate settings** одно- и многопараметрические задачи
- Mutually Exclusive Collectively Exhaustive, МЕСЕ** взаимоисключающие совместно исчерпывающие события (ВИСИ), полная группа событий
- double penalty problem** проблема двойного штрафа
- double penalty for the forecast** "двойное штрафование" прогноза с помощью количественной оценки – за прогноз явления не там, где оно случилось, и за отсутствие прогноза там, где оно случилось ("за попадание куда не надо попадать, и за промах куда надо попадать")
- poorer scores for higher-resolution forecasts** худшие оценки для более детального прогноза
- economic** экономический, народнохозяйственный, прикладной, практический
- economic benefit** экономическая выгода
- economic motivations for forecast verification** экономическая мотивация использования результатов оценки качества прогнозов
- economic value (utility)** экономическая стоимость, ценность (польза)
- economic value functions, EV** функция экономической стоимости
- elicit** выявлять, выводить, извлекать
- elicibility** выводимость, допустимость
- elicitation** сбор данных, получение вывода
- elicitable functional** выводимый, допустимый функционал (возможность построения функционала на распределении случайной величины в набор желательных статистик)
- elicitable properties of distribution** допустимые, выводимые свойства распределения

- elicitation and analysis of requirements** выявление и анализ требований к оценочному правилу, или оценочной функции
- revelation principle** принцип раскрытия (Осбэнда), позволяющий определить возможности оценочного правила
- ensemble** ансамбль
- dressing the ensemble members with kernel functions** "обёртывание" членов ансамбля ядерной (непараметрической) функцией плотности
- effect of ensemble size on skill** влияние размера ансамбля на качество прогноза
- ensemble consistency** согласованность (совместность) ансамбля - будущее состояние атмосферы рассматривается как случайная выборка из того же распределения, которое породило ансамбль
- ensemble Dawid-Sebastiani Score** ансамблевая оценка Дэвида-Себастиани
- Ensemble Energy Score** ансамблевая энергетическая оценка
- ensemble forecast as converted probabilities** ансамблевый прогноз как видоизмененные вероятности
- ensemble forecast as finite random samples from forecast distributions** ансамблевый прогноз как конечная случайная выборка из прогнозистических распределений
- ensemble forecast interpretation** интерпретация ансамблевого прогноза (как конечные выборки из распределений или как оценки вероятностей)
- ensemble forecast as finite set of deterministic forecasts** ансамблевый прогноз как конечный набор детерминистических прогнозов
- ensemble Model Output Statistics, MOS** статистики выходной ансамблевой продукции модели
- ensemble recalibration** калибровка ансамбля
- ensemble variogram score** ансамблевая оценка вариограммы
- ensemble verification system, EVS** система верификации ансамблевого прогноза
- ensembles as discrete samples** ансамбль как дискретные выборки
- ensembles as probabilistic forecasts** ансамбли как вероятностные прогнозы
- evaluation of ensemble calibration** оценивание результатов калибровки ансамбля
- fitting of suitable parametric distributions to the ensemble forecast** подбор (подгонка) параметрического распределения к ансамблю
- generalized discrimination score for ensembles** обобщенная оценка дискриминации для ансамблей

metrics for probabilistic forecast derived from an ensemble метрики (меры качества) для вероятностного прогноза, построенного по ансамблю (все стандартные оценки вероятностного прогноза)

Monte Carlo techniques in ensemble forecasting приемы Монте-Карло в ансамблевом прогнозировании

multivariate ensemble accuracy and skill точность и мера мастерства многомерного (многопараметрического) (прогностического) ансамбля

multivariate ensemble calibration калибровка многопараметрического ансамбля

probabilistic skill metrics applied to ensembles вероятностные оценки качества в приложении к ансамблевым прогнозам

spread of the values оценка разброса значений

spread-skill relationship соотношение разброса (значения по ансамблю) и качества (ошибки) прогноза

univariate ensemble accuracy and skill точность и мера мастерства однопараметрического (прогностического) ансамбля

univariate ensemble calibration калибровка однопараметрического ансамбля

equitability равноправность (равномерность, равновесность)

asymptotic equitability асимптотическая равноправность

equitable measure равноправная мера

equitable skill score равноправное мастерство (качество)

Equitable Threat Score, ETS; Gilbert Skill Score (GSS) равноправная оценка угроз, оценка качества Гилберта

equitable verification measure равноправная мера верификации

error ошибка, погрешность

LEPSCAT, linear error in probability space линейная ошибка в вероятностном пространстве (для категорий)

Linear Error in Probability Space, LEPS линейная ошибка в вероятностном пространстве

Mean Absolute Error, MAE средняя абсолютная ошибка

Mean Squared Error, MSE средний квадрат ошибки (СКО), не путать со среднеквадратической ошибкой и среднеквадратическим отклонением, которые тоже могут иметь аббревиатуру СКО

Mean Squared Error Skill Score мера мастерства относительно выборочной климатологии

Root Mean Squared Error, RMSE средняя квадратическая ошибка, корень квадратный из среднего квадрата ошибки

SEEPS, Stable Equitable Error in Probability Space устойчивая равноправная ошибка в вероятностном пространстве

squared error квадрат ошибки

standard deviation стандартное отклонение

estimation оценка appraisal; assessment; evaluation; valuation;

expectation (математическое) ожидание

expectations and expectiles ожидания и экспектили

expected reward ожидаемое вознаграждение

expected score ожидаемая оценка качества

exploratory разведочный

exploratory methods разведочные методы (статистического анализа)

simulation study исследования на основе имитационного моделирования

forecast прогноз

categorical forecast of occurrence категорический прогноз осуществления (события), не путать с категорийным (дискретным) прогнозом

categorical/probabilistic forecasts категорический/вероятностный прогноз

Central Credible Interval (CCI) Forecasts прогнозы байесовского доверительного интервала (интервал, в который попадает оценка, в противовес доверительному интервалу, который накрывает оценку)

deterministic categorical forecasts детерминистический категорический прогноз

discrete forecasts дискретный (категорийный) прогноз

forecast quality measures меры качества прогноза

forecast skill мастерство (качество, критерий качества) прогноза

forecast target объект прогноза (предиктант)

forecasts at face value прогноз "по номинальной стоимости", как есть - без дополнительной обработки

forecasts for finite-valued targets прогнозы для конечно-значных объектов

forecasts for theoretically unpredictable scales прогнозы для теоретически непредсказуемых масштабов

forecasts of extreme events прогнозы экстремальных явлений

forecasts of spatial fields прогнозы пространственных полей

goodness in weather forecasting истинное качество в прогнозировании, "хорошесть" прогнозирования

hindcast ретроспективные прогнозы, прогнозы по историческим данным, хайндкасты

inversion (recalibration) of the forecast "полезная" инверсия "абсолютно неточного плохого" прогноза, когда пользователь осведомлен о его качестве

large dimensionality of field forecasts большая размерность прогнозов (метеорологических) полей

- nonprobabilistic forecasts** не-вероятностные прогнозы полей
- probabilistic binary forecasts** вероятностные бинарные (да-нет) прогнозы
- quantile forecasts** квантильные прогнозы, прогнозы квантилей
- random guesses and perpetual forecasts** случайные угадывания и неизменные прогнозы
- requisite forecast** затребованный, необходимый прогноз, или прогноз, содержащий всю ту информацию, которая необходима потенциальному пользователю для принятия оптимальных решений
- sharp forecasts** контрастные прогнозы
- single-valued point forecasts** точечные однозначные прогнозы
- standard reference forecasts** стандартные контрольные прогнозы (для расчетов мастерства)
- unskillful baseline forecast** некачественный прогноз (выборочной повторяемости, климатологии)
- well-calibrated forecaster is said to be perfectly reliable** корректно калиброванного прогнозиста называют также "идеально надежным"

General Framework (Murphy-Winkler framework) основная структура верификации прогнозов ("по Мерфи-Винклеру") – двумерная плотность распределения вероятностей $p(f,o)$ случайных величин, соответствующих прогнозам f и наблюдениям o , разлагаемая двумя способами:

$p(f,o) = p(o|f)p(f) = p(f|o)p(o)$, где $p(o|f)$ $p(f)$ – разложение на "калибровку" и "детализацию" ("уточнение"), **calibration-refinement decomposition**; $p(f|o)$ $p(o)$ – разложение на "правдоподобие" и "базовую долю" ("выборочную климатологию", *sample climatology*), **likelihood-base rate decomposition**

calibrated probabilities калиброванные вероятности

calibration term член калибровки в разложении плотности распределения

calibration-refinement factorization разложение на калибровку и детализацию

complexity of verification problem сложность задачи верификации

concepts of sufficiency and refinement концепция достаточности и детализации (масштабирования)

dimensionality, or degrees of freedom, of the verification problem размерность, или степени свободы задачи верификации

discrete bivariate probability distribution function дискретная двухпараметрическая функция распределения вероятностей

general framework for forecast verification общая структура (для) верификации прогнозов

joint distribution of forecasts and observations – all of the relevant information совместное распределение прогнозов и наблюдений – вся необходимая информация (об оценке качества)

- joint, marginal and conditional distributions** совместное, маргинальное (частное, безусловное) и условное распределение вероятностей
- likelihood-base rate factorization** факторизация на правдоподобие и базовую долю
- refined forecast** детализованный прогноз
- refinement of forecasts** детализация прогнозов
- refinement term** член уточнения (детализации) в разложении Мерфи-Винклера
- sharp forecast** контрастный, детализованный прогноз (refined forecast is called also sharp)
- sharpness** контрастность, способность воспроизвести детали "наблюдения", ср. резкость в оптике)
- graphics** графические представления
- attributes diagram** диаграмма атрибутов
- bar chart** столбцовая диаграмма
- bin** интервал группировки или класса, интервал гистограммы
- binned spread-error diagram** сгруппированная диаграмма разброса и ошибки в ансамбле
- bivariate histogram** двумерная (двух-параметричная) гистограмма
- bounding boxes** ограничивающие блоки
- boxplots** диаграммы размаха, или блоковые статистические графики
- conditional quantile plots** графики условных квантилей
- discrimination diagram** диаграмма дискриминации
- displays of verification information** графическое представление результатов верификации
- graphical summaries** обобщающие графики
- Minimum Spanning Tree (MST) histogram** гистограмма минимального связующего дерева
- multivariate rank histogram** многопараметрическая ранговая гистограмма
- plots** графики, диаграммы
- quantile-quantile plot** квантильный график, квантиль-кватильный график
- reliability diagram** диаграмма надежности
- scatterplot** диаграмма рассеяния
- Taylor diagram** диаграмма Тэйлора
- The PIT Histogram (probability integral transform)** гистограмма преобразованного вероятностного интеграла
- hedge, hedging** подстраховка, перестрахование, подгонка (в прогнозировании)
- difficulty to hedge** преграда для перестрахования
- hedged forecast** подогнанный (подстрахованный) прогноз

hedging and strictly proper scoring rules подстраховка в прогнозах и строго собственные (оценочные) правила

hedging strategy стратегия подстраховки

hedging towards climatology подстраховка под климатологию

information информация

entropy function энтропийная функция

information content информационное содержание, содержание информации

information measure мера информации, мера количества информации

information quantity качество информации

information ratio коэффициент информации

likelihood правдоподобие (в математической статистике – характеристика параметров при заданном наблюдении)

climatological likelihood климатологическое правдоподобие, климатологическая повторяемость события

likelihood term член правдоподобия в разложении Мерфи – Винклера

likelihood-base rate factorization разложение на правдоподобие и базовую долю (выборочную климатологию)

measure мера, критерий, показатель (качества)

equitable measure равноправная мера качества (по Гандину-Мерфи, одинаково оценивает случайные и неизменные прогнозы)

asymptotically equitable measure асимптотически равноправная мера (равноправность достигается при большом, в идеале бесконечном объеме выборки)

base-rate independence of measure независимость меры от базовой доли (выборочной климатологии)

basic inequitable measures основные неравноправные меры

boundedness, unbounded measure ограниченность. неограниченная мера

class of probability measures класс вероятностных мер

definedness, undefined measure определимость, неопределимая мера (одна или несколько ячеек таблицы сопряженности оказываются нулевыми (возможно деление на нуль)).

distance measures меры расстояния

error measures меры ошибки

measure-oriented verification верификация, ориентированная на меры (качества)

measures for multicategory discrete predictands (PC, HSS, PSS) меры для мульти-категорийных предиктантов (общая оправдываемость, критерий Хайдке, критерий Пирса)

- measures from signal detection theory** меры из теории обнаружения сигналов
- measures of accuracy** меры точности
- metrics** метрики (меры, показатели качества)
- numerical descriptive measures** числовые описательные меры
- resistant measures** резистентные меры (слабо зависят от небольшого количества выбросов)
- robust measures** робастные меры (нечувствительны к определенным предположениям о распределении данных)
- metaverification** метаверификация – термин А. Мерфи для обозначения методологии отбора подходящих статистических характеристик для определенных целей верификации
- non-degeneracy** невырожденность
- non-degeneracy for rare events** невырожденность (меры) для редких явлений
- non-degenerate measures (EDS, SEDS, SEDI)** невырожденные меры (экстремальные индексы)
- observations** наблюдения
- observation or target** наблюдения или объект (верификации, оценки, исследования)
- perfect** совершенный, идеальный
- calibrated forecasts** калиброванные прогнозы (идеальны)
- general attribute of the predictions** общий атрибут прогнозирования
- perfect forecast** идеальный (совершенный) прогноз
- perfect reliability** идеальная надежность
- perfectly calibrated (reliable) forecast** идеально калиброванные (надежные) прогнозы
- perfectly discriminatory forecasts** прогнозы с идеальной дискриминирующей способностью
- performance** качество, производительность, эффективность
- performance criterion** критерий качества
- performance diagram for 2×2 contingency table** диаграмма качества для двухвходовой таблицы сопряженности
- performance measures** мера качества
- performance of quality control service** эффективность службы контроля качества
- predictand** предиктанд
- categorical (discrete) predictand** категорийный (дискретный) предиктанд
- continuous predictand** непрерывный предиктанд
- multicategory predictand** мультикатегорийный предиктанд

prediction прогноз, предсказание

overprediction переоценка, завышение прогностических значений (например, частоты явления)

prediction interval интервал прогноза

underprediction недооценка, занижение прогностических значений

predictive прогнозируемый

assessor's true predictive density истинная прогнозируемая плотность распределения эксперта (оценки)

predictive cumulative distribution function прогнозируемая кумулятивная функция распределения

predictive densities прогнозируемые плотности распределения

predictive distribution прогнозируемое распределение

probability вероятность

probability distribution forecasts for continuous predictands прогнозы распределений вероятностей для непрерывных предиктантов

probability forecasts for discrete predictands вероятностные прогнозы для дискретных предиктантов

probability forecasts for multiple-category events вероятностные прогнозы для мульти-категорийных событий

probability function (or frequency function) функция вероятностей (или функция частот, повторяемостей)

probability integral transform, PIT преобразование интеграла вероятности

probability of detection, POD вероятность обнаружения (доля попаданий в таблице сопряженностей 2×2)

probability score вероятностная оценка

Proper собственный, собственная оценка качества – прогнозист вознаграждается наивысшей ожидаемой оценкой только в том случае, когда он выпускает прогноз, построенный на своих истинных суждениях

proper linear score собственная линейная оценка

proper Quantile Score, QS собственная квантильная оценка

proper scoring rule собственное оценочное правило

properness свойство собственности (оценки качества)

propriety собственность (оценки качества)

properties of measures свойства мер (качества)

consistency согласованность (в каком-либо смысле)

other properties (besides desirable ones) прочие свойства мер, помимо желательных

- (1) **linear** линейность
- (2) **regular** регулярность
- (3) **biased forecasts can get perfect score** смещенные прогнозы могут получить идеальную оценку

(4) **transpose symmetric** транспонированная симметричность (относительно взаимной замены наблюдения и прогноза, ложных тревог и промахов)

(5) **complement symmetric** дополнительная симметричность (относительно взаимной замены события на отсутствие события - в наблюдении и прогнозе)

properties of verification measures (metaverification) свойства мер верификации (объект метаверификации)

relevant properties of measures: equitable, asymptotically equitable значимые свойства мер - равноправность или асимптотическая равноправность

rank ранг

rank histograms ранговая гистограмма

Ranked Probability Score, RPS ранговая оценка вероятностей (вероятностная ранговая оценка)

Ranked Probability Skill Score, RPSS критерий качества на оценке ранжированных вероятностей

reference эталон, стандарт, ориентир, контроль

reference forecast контрольный, эталонный прогноз

reference, or control, forecasts for Skill Scores контрольный прогнозы для построения оценки мастерства прогноза (критерия качества)

reliability надежность

reliability and resolution надежность и разрешающая способность

reliability diagram диаграмма надежности

reliability(attribute), or calibration, or conditional bias атрибут надежность, или калиброванность, или условное смещение

sample выборка

sample correlation coefficient выборочный коэффициент корреляции

sample mean выборочное среднее

sample variance выборочная дисперсия

sampling and inference for verification отбор данных и статистические выводы в верификации

sampling characteristics of contingency table statistics выборочные характеристики статистик таблицы сопряженности

sampling uncertainty and confidence intervals for performance measures выборочная неопределенность и доверительные интервалы для мер качества

scalar attributes скалярные (числовые) атрибуты

scalar attributes (aspects) of forecast performance скалярные атрибуты (аспекты) качества прогнозов

scalar attributes of the 2x2 contingency table: accuracy (PC, TS, OR), bias (Bias), reliability and resolution (FAR), discrimination

(H, F, EDI) скалярные атрибуты двухвходовой таблицы сопряженности: точность (общая оправдываемость, оценка угроз, отношение шансов), смещение (отношение частот), надежность и разрешающая способность (отношение ложных тревог), дискриминирующая способность (доля попаданий, доля ложных тревог, индекс экстремальной зависимости)

scalar measure of accuracy скалярная мера точности

scalar measures for probability forecasts of ordinal predictands

(ranked probability score, RPS) скалярные меры для вероятностных прогнозов упорядоченных предиктантов (ранговая вероятностная оценка)

scalar summary verification statistic скалярная суммарная оценка верификации

scale масштаб

scale mismatch несовпадение (рассогласованность) масштабов

scale separation methods методы разделения масштабов, масштабо-разделительные методы, методы с разделением масштабов

scale-dependent spatial statistics масштабо-зависимые пространственные характеристики

score оценка

skill score критерий, показатель качества, или мастерства

Brier score decomposition in terms of resolution, reliability and uncertainty разложение оценки Брайера на разрешение, надежность и неопределенность

Brier Score (BS), half Brier score оценка Брайера, половинная оценка Брайера

Brier Skill Score, BSS критерий качества Брайера

Clayton Skill Score, CSS критерий качества Клейтона

Continuous Ranked Probability Score, CRPS непрерывная вероятностная ранговая оценка

Critical Success Index, CSI (Gilbert's skill score) критический индекс успешности, (критерий Гилберта), для бинарных величин

Doolittle Skill Score критерий качества Дулиттла

Energy Score энергетическая оценка (Гнайтинга)

Equitable Threat Score, ETS равноправная оценка угроз

Extremal Dependence Index, EDI индекс (показатель) экстремальности зависимости

Extreme Dependency Score, EDS оценка экстремальной зависимости

Fractions Skill Score, FSS оценка по отношению пространственных долей (по пространственным долям)

Gandin and Murphy equitable scores равноправные оценки Гандина и Мерфи

Gerrity equitable scores равноправные оценки Геррити

Gilbert Skill Score, GSS критерий Гилберта

Hanssen-Kuipers discriminant, Kuipers' performance index, true skill statistic (TSS) дискриминант Ганса-Кюйперса, показатель качества Кюйперса, статистика истинного качества

Heidke Skill Score, HSS критерий качества Хайдке ("показатель надежности Багрова")

Ignorance Score, IS оценка незнания (вариант оценки количества информации)

linear score линейная оценка

Logarithmic, or Ignorance Score for Multiple Categories, alternative to the Brier score логарифмическая оценка или оценка незнания для мульти-категорийного предиктанта как альтернатива оценке Брайера

logarithmic, spherical, pseudospherical, quadratic scores логарифмические, сферические, псевдосферические, квадратичные оценки

Odds Ratio Skill Score, ORSS критерий качества для отношения шансов

Peirce Skill Score, PSS критерий Пирса ("критерий Пирси-Обухова")

Proportion Correct, PC общая оправдываемость (термин Финли)

S1 score пространственная (градиентная) оценка

score as a reward оценка как вознаграждение

score function, scoring rule оценочная функция, оценочное правило

scores based on cumulative frequency оценки, основанные на кумулятивной частоте

Yule-Kendall index показатель Юла-Кендалла

scoring оценивание, оценки

scoring function оценочная функция

scoring matrices оценочные матрицы

scoring rule оценочное правило

scoring rules as devices to elicit probabilities оценочные правила как средства для определения вероятностей

consistent scoring function согласованные оценочные функции

negatively oriented scoring functions отрицательно ориентированные оценочные функции (чем больше, тем лучше, и наоборот)

reproducing scoring system воспроизводящая оценочная система (старое название "строгих собственных оценок")

Signal detection theory теория обнаружения сигналов

regularity регулярность (кривая ROC нигде не соприкасается с границей единичного координатного квадрата)

Relative (or Receiver) Operating Characteristic, ROC относительная (ресиверная) оперативная характеристика

ROC area площадь под кривой ROC

ROC Diagram диаграмма ROC

ROC Diagram Sampling Characteristics выборочные характеристики диаграммы ROC

trapezoidal area under the ROC curve трапецевидная площадь под кривой ROC

skill мастерство, качество

absolute/relative skill of predictions абсолютное/относительное мастерство прогнозов

artificial skill искусственное качество

estimates of potential skill оценки потенциального качества

inflation of skill завышение, "раздувание" качества (мастерства)

reference forecasts: persistent, random (blind), base-rate (sample climatology) forecasts контрольные прогнозы: инерция, случай (слепой прогноз), прогнозы выборочной климатологии -иногда называются "псевдопрогнозами"

skill attributes of ensemble forecasts атрибуты качества ансамблевых прогнозов

skill score мастерство, критерий качества, критерий мастерства

Skill Scores for 2x2 Contingency Tables: HSS, PSS, CSS, GSS, ORSS критерии качества для двухходовой таблицы сопряженности (Хайдке, Пирса, Клейтона, Гилберта, отношения шансов)

Skill Scores, constructed using the MAE, MSE, or RMSE as the underlying accuracy statistics критерии качества, в которых оценки ошибок используются как базовые статистические характеристики точности

Skill, or relative accuracy мастерство, или относительная точность, точность относительно некоторого эталонного прогноза

skill-bias diagram диаграмма качество-смещение

spatial verification пространственная верификация

attributes of spatial feature атрибуты пространственных признаков (объектов)

Contiguous Rain Area, CRA сопряженная область дождя (методика Эберт-Макбрайда, исторически первый "объект" пространственной верификации)

displacement/distortion methods методы перемещения или искажения

downscaling даунскейлинг, локализация, масштабирование на меньший размер

feature признак, объект (термин Гиллеланда, объяснившего загрузенность слова object в языках программирования)

feature-based methods объектно-ориентированные методы

feature-matching techniques методики сопоставления признаков

features/object-based/oriented комбинации терминов для описания "пространственной верификации" (признак/объект - основа/база)

field deformation methods методы деформации полей

field verification based on spatial structure верификация прогноза полей с основой на пространственных структурах

filtering methods or displacement methods методы фильтрации и смещения

fuzzy definitions of compound events нечеткие определения составных явлений

generalized structural functions обобщенные структурные функции

image processing applications приложение методов обработки изображений

matching features in the two fields сопоставление (спаривание) признаков в двух полях

matching methods методы сопоставления (спаривания) объектов

meteorological features: rain systems, cloud systems, low pressure centres, wind speed maxima метеорологические "признаки" (объекты): дождевые системы, облачные системы, центры низкого давления, области максимальных ветров

Method for Object-based Diagnostic Evaluation, MODE комплекс компьютерных программ для диагностической оценки качества с помощью объектно-ориентированных методов верификации

mimicking a human interpretation of spatial forecasts имитация человеческой интерпретации пространственных прогнозов

neighborhood (fuzzy) methods окрестные (нечеткие) методы

neighborhoods of forecasts and observations окрестности прогнозов и наблюдений

sensitive-to-distance measures меры, чувствительные к расстоянию

spatial verification outputs продукция пространственной верификации:

- (1) **scale-specific errors**, ошибки в масштабах
- (2) **scales of useful skill**, масштабы полезного качества
- (3) **structure errors**, ошибки в структуре
- (4) **location errors**, ошибки в положении
- (5) **intensity errors**, ошибки в интенсивности
- (6) **hits, misses, false alarms, correct negatives** попадания, промахи, ложные тревоги и корректные отказы

Structure-Amplitude-Location, SAL, technique методика верификации с учетом структуры, амплитуды и положения объектов

taxonomy for spatial verification methods систематика методов пространственной классификации

textural characteristics пространственная текстура

upsampling апскейлинг, укрупнение, усреднение, масштабирование на больший размер; точка наблюдений «укрупняется», то есть попросту приписывается ближайшей точке прогностической сетки

weather patterns погодные типы

statistic статистика, статистическая характеристика

Mann-Whitney, U-statistic статистика Манна-Уитни, U-статистика

resampling verification statistics (cross-validation) характеристики качества с использованием перевыборок (кросс-валидация)

statistical assumptions for performance measures статистические предположения и меры качества

statistical inference статистический вывод (индуктивное рассуждение о свойствах генеральной совокупности на основе конечных выборок)

statistical post-processing статистический постпроцессинг (постобработка)

statistical significance статистическая значимость

strictly строго (как правило, в формальном описании условий используется строгое неравенство)

strictly consistent scoring function строго согласованное оценочное правило

strictly proper scores строго собственные оценки

strictly proper scoring rule строго собственное оценочное правило

subjective субъективный

Bayes rule of forecasting Байесово правило прогнозирования

subjective interpretation of forecast probability субъективная интерпретация прогностической вероятности

subjective probability субъективная вероятность

sufficiency достаточность

sufficiency of a forecast достаточность прогноза (термин ДеГроота и Финберга, в огрубленной форме – с помощью статистического преобразования данного прогноза можно построить прогноз данного явления, полученный любым другим способом)

symmetry симметрия, симметричность, симметризованность

symmetry of measure симметричность мер (относительно некоторых преобразований симметрии в матрице сопряженностей)

Symmetric Extremal Dependence Index, SEDI симметричный индекс экстремальной зависимости

Symmetric Extreme Dependency Score, SEDS симметричная оценка экстремальной зависимости

test тест, проверка, критерий

test verification set тестовый набор для верификации

testing тестирование, проверка

testing the limit of predictability оценка предела предсказуемости

tests (in a statistical sense) to permit inferences from the results критерии (в статистическом смысле) для получения выводов из результата обработки

tests of ensemble reliability критерии ансамблевой надежности

type тип

type of forecast тип прогноза

type of variables: categorical (binary, three or more categories) or continuous типы переменных: категорийные, или дискретные (бинарные, с тремя или более категориями), и непрерывные

types of forecast "goodness" типы прогностического качества, "хорошесть" прогноза (по Мерфи, consistency, quality and value – согласованность, качество и ценность (польза))

types of forecast and verification data типы прогнозов и данных для верификации

types of predictand типы предиктантов

types of software and programming languages типы программного обеспечения и языков программирования

utility польза

validity достоверность

validity measure мера достоверности, старое название атрибута надежности (reliability)

value величина, ценность, стоимость

absolute/relative value of predictions абсолютная/относительная ценность прогнозов

artificially inflated value искусственно завышенная стоимость (ценность)

value added добавленная стоимость

Value Score, VS оценка стоимости (полезности) дихотомического прогноза зависимости от отношения C/L (затраты/потери)

variable переменная

binary variable бинарная переменная (двоичная)

categorical and continuous variables категорийные (категорные, дискретные) и непрерывные переменные

nominal and ordinal discrete events номинальные и порядковые дискретные явления

numerical variable числовая переменная

ordered/unordered variables упорядоченные/неупорядоченные переменные

ordinal variable порядковая переменная

verification верификация, оценка качества, оценка, проверка, поверка, учет, контроль

absolute verification абсолютная верификация, оценка качество отдельной прогностической системы

classification of special verification methods классификация специальных методов верификации

comparative verification сравнительная верификация (охватывает две и более прогностические системы)

fundamental level of forecast verification фундаментальный уровень верификации прогнозов (анализ свойств совместного распределения прогнозов и наблюдений)

matched comparative verification согласованная сравнительная верификация (производится для прогностических систем в разных условиях прогнозирования)

nonprobabilistic verification data не-вероятностные данные для верификации

purposes of forecast verification цели верификации (по Брайеру, административная, научная и экономическая)

scientific application of forecast verification научное приложение результатов верификации прогнозов

training verification set обучающий набор для верификации

verification as a regression problem верификация как регрессионная задача

verification based on economic value верификация, основанная на экономической стоимости

verification data данные для верификации

verification measures for extreme events меры качества прогноза экстремальных явлений

verification measures меры верификации

verification observations наблюдения для верификации

verification of ensemble forecasts верификация ансамблевых прогнозов

verification of warnings верификация предупреждений

verification practices практические методы верификации

verification problem задача верификации

verification procedure процедура оценки качества

verification scores оценки (показатели) качества прогнозов

verification software программное обеспечение верификации прогнозов

verification strategy стратегия верификации (постановка целей, определение адресатов и выбор методов оценки)

verification when observation is uncertain верификация в условиях неопределенности наблюдений

verifying observations (контрольные) наблюдения для верификации

ПРИЛОЖЕНИЕ 2

Стандартизованная верификация продукции детерминистского численного прогноза погоды

В настоящем приложении содержится подробное описание процедур подготовки стандартного набора показателей верификации для выпускаемых центрами Глобальной системы обработки данных и прогнозирования (ГСОДП) прогнозов в рамках детерминистского численного прогноза погоды (ЧПП) и обмена им.

Термин «детерминистский ЧПП» относится к однократным интегрированиям моделей ЧПП, предоставляющих продукцию, в которой определяются единичные будущие состояния атмосферы (чем отличается от системы ансамблевого прогнозирования (САП), где множественное интегрирование определяет совокупность нескольких будущих состояний).

Стандартизованная верификация должна предоставлять основную необходимую информацию, соответствующую современному состоянию ЧПП, являться в то же время как можно более простой и легкой для осуществления и обеспечивать ее последовательное внедрение в участвующих центрах. Математическое описание показателей приводится на веб-сайте(ах) ведущего(их) центра(ов) по верификации ЧПП наряду с дополнительной информацией по расчету показателей, наборами данных наблюдений и климатических данных для использования в целях верификации, а также процедурами представления показателей.

В нижеследующих разделах определены **два набора статистических данных по верификации**.

- Обязательный набор данных предоставляется всеми участвующими центрами.
- Процедуры для высотных уровней модели и приземных полей являются различными и представлены отдельно. Подробные процедуры необходимы для того, чтобы обеспечить сравнение результатов различных участвующих центров научно обоснованным способом.
- Также определен набор дополнительно рекомендуемых статистических данных, которые все центры, по возможности, должны предоставлять.

1. СТАНДАРТИЗИРОВАННАЯ ВЕРИФИКАЦИЯ ВЫСОТНЫХ УРОВНЕЙ МОДЕЛИ

1.1. Параметры

- Внетропические районы:
 - Обязательные:
 - давление, приведенное к среднему уровню моря (ДСУМ) (верификация только с помощью анализа);
 - геопотенциальная высота на уровнях 850, 500 и 250 гПа;
 - температура на уровнях 850, 500 и 250 гПа;
 - ветер на уровнях 925, 850, 700, 500 и 250 гПа.
 - Дополнительно рекомендованные:
 - геопотенциальная высота, температура, ветер на уровне 100 гПа;
 - относительная влажность на уровне 850 и 700 гПа.
- Тропические районы:
 - Обязательные:
 - геопотенциальная высота на уровнях 850 и 250 гПа;
 - температура на уровнях 850 и 250 гПа;
 - ветер на уровнях 850 и 250 гПа.
 - Дополнительно рекомендованные:
 - относительная влажность на уровне 850 и 700 гПа.

1.2. Время инициализации прогноза

Показатели ежедневно рассчитываются отдельно для прогнозов, инициализированных в 00.00 и в 12.00 ч ВСВ. Для центров, которые не рассчитывают прогнозы для какого-либо из этих сроков, показатели предоставляются для прогнозов, инициализированных в другое время, и должны обозначаться соответственно.

1.3. Шаги прогноза

Обязательные шаги прогноза: 24, 48, 72, ...240 ч или конец прогноза.

Дополнительно рекомендованные: 12-часовой прогноз (12 ч, 24 ч, 36 ч, ...).

1.4. Районы

Внетропические районы Северного полушария	90–20° с. ш. включительно, все долготы
Внетропические районы Южного полушария	90–20° ю. ш. включительно, все долготы

Тропические районы	20° с. ш. – 20° ю. ш. включительно, все долготы
Северная Америка	25–60° с. ш., 50–145° з. д.
Европа/Северная Африка	25–70° с. ш., 10° з. д. – 28° в. д.
Азия	25–65° с. ш., 60–145° в. д.
Австралия/Новая Зеландия	10–55° ю. ш., 90–180° в. д.
Северный полярный регион	90–60° с. ш. включительно, все долготы
Южный полярный регион	90–60° ю. ш., включительно, все долготы

Верификация с помощью анализа для точек сетки внутри каждого района включает точки на границе.

1.5. Верификация с помощью анализа. Сетка и интерполяция

Все параметры верифицируются с помощью собственного анализа центра на регулярной сетке $1,5^\circ \times 1,5^\circ$.

При выборе сетки для верификации учитывалось разнообразие разрешений существующих глобальных моделей ЧПП, разрешенные масштабы моделей (несколько расстояний по сетке), разрешение имеющихся климатологий, возможности для мониторинга долгосрочных тенденций качества (включая более ранние прогнозы с более низким разрешением) и эффективности вычислений.

Интерполяция полей моделей с более высоким разрешением на сетку для верификации осуществляется, чтобы сохранить характеристики в масштабе сетки для верификации, а не для того, чтобы ввести дополнительное выравнивание.

Используются следующие *процедуры*:

- *спектральные поля*: усечение до эквивалентного спектрального разрешения (T120) для сетки для верификации;
- *сеточные значения*: использование осреднения по пространству для интерполяции на сетку для верификации.

Для показателей, требующих климатологии, климатология предоставляется через веб- сайт(ы) ведущего(их) центра(ов) по верификации на сетке для верификации и не нуждается в дальнейшей интерполяции.

1.6. Верификация по наблюдениям

• *Наблюдения*

Все параметры, определенные в п. 1.1, за исключением ДСУМ, верифицируются по общему набору данных радиозондирования. Список радиозондовых наблюдений для каждого района ежегодно обновляется ведущим(и) центром(ами) комиссии по основным системам (КОС) по мониторингу радиозондирования. Данные отобранных станций должны быть доступны для всех центров, быть надлежащего качества и предоставляться на регулярной основе. До составления окончательного списка желательно проконсультироваться со всеми центрами (обычно по электронной почте). Существующий список доступен через веб-сайт(ы) ведущего(их) центра(ов) по верификации.

Ведущий(е) центр(ы) связывается(ются) со всеми участвующими центрами, когда появляется новый список, и информирует их о дате, начиная с которой новый список используется. Используемые для верификации данные наблюдений тщательно просматриваются, чтобы исключить данные с серьезными ошибками. Для этого рекомендуется, чтобы центры исключили значения, отброшенные в процессе их объективного анализа. Кроме того, центры, которые вносят поправки в данные наблюдений, полученные по ГСТ, для устранения систематических погрешностей (например, выполнение радиационной поправки) должны использовать скорректированные данные наблюдений для расчета статистических данных по верификации. Во всех возможных случаях эти процедуры по корректировке должны быть задокументированы (например, со ссылкой на какой-либо технический отчет или документ в рамках периодических изданий).

• *Интерполяция*

Верификация проводится с использованием исходного узла сетки модели, ближайшего к месту наблюдений.

• *Районы*

Девять сетей, используемых при верификации в сравнении с данными радиозондирования, включают станции радиозондирования, расположенные в географических районах, указанных в разделе 5.4 выше. Список станций радиозондирования, которые должны использоваться для каждого района, обновляется ежегодно ведущим(и) центром(ами) по мониторингу радиозондирования (см. п. 1.6, *наблюдения*).

• *Показатели для отдельных станций*

Рекомендуется, чтобы в дополнение к районам, перечисленным в разделе 5.4, показатели по сравнению с наблюдениями рассчитывались для каждой станции по отдельности. Обмен показателями для районов должен со временем прекратиться.

1.7. Показатели

Математическое описание показателей оформляется в виде документа, который размещается на веб-сайте(ах) ведущего(их) центра(ов) по верификации, наряду с дополнительной информацией по расчету показателей.

Следующие показатели должны рассчитываться для всех параметров по сравнению с анализом (за исключением давления на среднем уровне моря) и наблюдениями:

- Ветер

Обязательные:

- СКО вектора ветра;
- средняя ошибка скорости ветра.

- Другие параметры:

Обязательные:

- средняя ошибка;
- СКО;
- коэффициент корреляции между прогнозируемыми и проанализированными аномалиями (не требуется для наблюдений);
- показатель S1 (только для давления на среднем уровне моря и только по сравнению с анализом).

Дополнительно рекомендованные:

- средняя абсолютная ошибка;
- прогнозируемые и проанализированные аномалии СКО (не требуется для наблюдений)
- стандартное отклонение прогнозируемых и проанализированных полей (не требуется для наблюдений)

1.8. Климатология

Для обеспечения согласованности результатов, получаемых различными центрами, для показателей, которым необходима климатология, используется общая (единая) климатология. Все центры используют климатологию, представленную на веб-сайте(ах) ведущего(их) центра(ов) по верификации.

Ежедневная климатология аэрологических параметров доступна как для 00.00, так и 12.00 ч ВСВ. Это обеспечивает актуальную оценку климатических характеристик для каждого дня года, включая климатические средние, стандартное отклонение и избранные квантили распределения климатических характеристик. Эти последние статистические данные необходимы для принятой КОС стандартизированной верификации прогнозов САП. Данные представляются в формате GRiB.

1.9. Месячные и годовые средние показатели

Там, где требуются *средние показатели за определенный период*, усреднение проводится с использованием следующих *процедур*:

- линейные показатели (средняя ошибка, средняя абсолютная ошибка) — среднее значение;
- нелинейные показатели преобразуются в соответствующую линейную меру для усреднения;
- среднее значение среднеквадратической ошибки (СКО);
- Z-преобразование для корреляции.

За определенный период среднее значение рассчитывается по всем прогнозам, подвергаемым верификации за данный период. Средние значения рассчитываются отдельно для прогнозов, инициированных в 00.00 и 12.00 ч ВСВ, и для обоих полученных наборов средних значений.

Годовые средние значения суточных показателей включаются в ежегодный Технический отчет о развитии Глобальной системы обработки данных и прогнозирования (<https://www.wmo.int/pages/prog/www/DPFS/GDPFS-Progress-Reports.html>). Эти статистические данные касаются 24, 72 и 120-часового прогнозов и включают среднюю квадратическую ошибку вектора ветра на уровнях 850 гПа (только тропические районы) и 250 гПа (все районы), а также среднюю квадратическую ошибку геопотенциальных высот на уровне 500 гПа (все районы, кроме тропических). Таблица количества наблюдений в месяц также должна включаться в ежегодный отчет.

2. СТАНДАРТИЗИРОВАННАЯ ВЕРИФИКАЦИЯ ПРИЗЕМНЫХ ПОЛЕЙ

2.1. Параметры и единицы измерения

Обязательные:

- температура на высоте 2 м, К;
- скорость ветра на высоте 10 м, м/с;
- направление ветра на высоте 10 м, град;
- количество осадков за 24 ч, мм.

Дополнительно рекомендованные:

- общая облачность, 0–1 (преобразовать до октанта для сопряженных таблиц);
- осадки за 6 ч, мм;
- относительная влажность на высоте 2 м, %;
- температура точки росы на высоте 2 м, К.

Для *температуры на высоте 2 м* применяется простая поправка на высоту между абсолютной высотой по модельным расчетам и высотой станции над уровнем моря, с использованием вертикального градиента температуры, равного 0,0065 К/м.

Для точки росы на высоте 2 м аналоговая поправка на высоту применяется с использованием вертикального градиента температуры, равного 0,0012 К/м. Это приблизительно соответствует вертикальному градиенту точки росы в атмосфере при градиенте температуры 0,0065 К/м и постоянной относительной влажности.

2.2. Время инициализации прогноза

Показатели рассчитываются ежедневно, отдельно для прогнозов, инициализированных в 00.00 и 12.00 ч ВСВ. Для тех центров, которые не рассчитывают прогнозы для какого-либо из этих сроков, показатели могут предоставляться для прогнозов, инициализированных в другое время, и должны обозначаться соответственно.

2.3. Шаги прогноза

Обязательные шаги прогноза:

- от 6 часов до включительно T+72; от 12 ч до включительно T+240 или конца прогноза;
- осадки на 24 ч: от 24 ч до T+240 или конца прогноза.

Дополнительно рекомендованные:

- от 3 ч до включительно T+72; от 6 ч до T+240 или конца прогноза (для улучшенного представления суточного цикла);

2.4. Сетка и интерполяция

Верификация проводится с использованием исходного узла сетки модели, ближайшего к месту наблюдений.

2.5. Наблюдения

Верификация выполняется для наземных станций, распространяющих данные синоптических приземных наблюдений в коде SYNOP через ГСТ. Каждый участвующий центр стремится включить как можно большее число станций для обеспечения надлежащего глобального охвата. Допускается, чтобы списки станций, используемые в процессе верификации, разнились между центрами. Это становится возможным благодаря тому, что показатели для отдельных станций подлежат обмену.

Центрам рекомендуется использовать доступные им процедуры контроля качества, чтобы уменьшить влияние ошибок наблюдений на показатели. Это подразумевает устранение случайных нефизических значений, а также данных наблюдений на отдельных станциях, которые систематически отвергались в течение определенного периода времени. Когда возможно, процедуры контроля качества должны быть задокументированы (например, путем ссылки на технический отчет или документ в рамках периодических изданий).

2.6. Показатели

Показатели рассчитываются для каждой станции отдельно. Станция, для которой произведен расчет показателей, имеет по меньшей мере 90-процентное наличие данных в течение периода верификации.

- Для температуры на высоте 2 м, относительной влажности на высоте 2 м, точки росы на высоте 2 м, скорости ветра на высоте 2 м, направления ветра на высоте 10 м и общей облачности рассчитываются следующие количественные показатели ошибок:

- средняя ошибка;
- средняя абсолютная ошибка;
- СКО.

Направление ветра на высоте 10 м подвергается верификации только в случаях, когда наблюдаемая скорость ветра составляет $\geq 3 \text{ м}\cdot\text{с}^{-1}$. Для направления ветра на высоте 10 м необходимо принимать во внимание эквивалентность 360 и 0° (циклическое продолжение).

- Для скорости ветра на высоте 10 м, осадков и общей облачности предоставляются таблицы сопряженности признаков для следующих пороговых значений:

- скорость ветра на высоте 10 м: 5, 10 и 15 м/с;
- осадки за 24 ч: 1, 10 и 50 мм;
- осадки за 6 ч: 1, 5 и 25 мм;
- общая облачность: 2 и 7 октантов.

Для общей облачности выходная продукция модели должна быть округлена до ближайшего октанта до начала верификации (только для таблиц сопряженности признаков).

Количественные показатели ошибок сообщаются с точностью как минимум до четырех значащих цифр, например 3,142 для ошибки π .

В таблицах сопряженности указывается конкретное число отсчетов, а не относительные частотности, с тем чтобы можно было рассчитать размер выборки. Таблицы сопряженности признаков для каждого параметра содержат все пороговые значения, приведенные выше.

Математическое описание показателей документируется и приводится на веб-сайте(ах) ведущего(их) центра(ов) по верификации наряду с дополнительной информацией по расчету показателей.

2.7. Временная и пространственная агрегация

Для любого заданного одномесячного периода показатели ошибок и таблицы сопряженности рассчитываются для каждой станции отдельно. Это создает основу для агрегации пользователями данных по верификации с учетом как временного, так и пространственного факторов. За определенный период среднее значение вычисляется по всем прогнозам, подвергшимся верификации за данный период.

Пространственная агрегация не является частью обмена и оставлена на усмотрение пользователя. Обмен показателями, таким образом, позволяет пользователям прогнозов получить детализированную информацию о качестве функционирования модели для отдельной станции. Кроме того, это обеспечивает высокий уровень прозрачности и гибкости для исследований по взаимному сравнению моделей. Более того, это устраняет необходимость координации, рассылки и обновления списка наземных станций для верификации. Для исследований по взаимному сравнению моделей общая часть различных наборов станций, используемых глобальными центрами моделирования, будет применяться для сравнения («наименьший общий знаменатель»).

Если пользователи хотят агрегировать полученные в результате обмена показатели, они могут свериться с веб-сайтом(ами) ведущего(их) центра(ов) по верификации, который(е) предоставляет(ют) руководящие указания по выбору областей агрегации. По сравнению с верификацией прогнозов параметров верхних слоев атмосферы, необходимо уделить большее внимание агрегации данных относительно однородных в климатологическом отношении участков (поскольку абсолютные пороговые значения используются для таблиц сопряженности).

ПРИЛОЖЕНИЕ 3

Стандартные критерии оценки для верификации продукции глобальной системы ансамблевого прогноза

В настоящем приложении содержится описание подробных процедур подготовки стандартного набора показателей верификации для выпускаемых центрами ГСОДП прогнозов в рамках системы ансамблевого прогноза (САП) и обмена им. Цель заключается в том, чтобы предоставлять прогнозистам в НМГС систематическую информацию о верификации продукции САП участвующих центров ГСОДП и помочь центрам ГСОДП сравнивать и улучшать свои прогнозы. Обмен показателями между участвующими центрами подготовки прогнозов будет осуществляться через ведущий(е) центр(ы) по верификации прогнозов САП. Функции ведущего центра заключаются в создании и поддержании веб-сайта для информации о верификации продукции САП, с тем чтобы потенциальные пользователи могли извлекать пользу из соответствующего представления результатов.

Система ансамблевого прогноза обеспечивает полную оценку прогностического вероятностного распределения, включая расчёт в наилучшем приближении детерминистского прогноза для средней величины по ансамблю, а также показателей прогностической неопределенности и вероятности. Таким образом, верификация САП включает верификацию средней величины по ансамблю как детерминистского прогноза в рамках ЧПП, а также специфические критерии вероятностной оценки.

Стандартизированная верификация должна предоставлять основную необходимую информацию, соответствующую современному состоянию САП, являться в то же время как можно более простой и легкой для осуществления, обеспечивать ее согласованное внедрение в участвующих центрах, особенно при интерполяции на сетку для верификации, и использовать общую (единую) климатологию и набор данных наблюдений.

1. Статистические данные по верификации

Ниже определены четыре набора статистических данных по верификации.

Обязательный набор данных предоставляется всеми участвующими центрами. Подробные процедуры необходимы для того, чтобы обеспечить сравнение результатов различных участвующих центров научно обоснованным способом.

Также определен набор *дополнительно рекомендуемых статистических данных*, которые всем центрам, по возможности, следует предоставлять.

Четыре набора статистических данных кратко сформулированы следующим образом:

Обязательный набор данных, предоставляемый всеми участвующими центрами:

- среднее значение по ансамблю;
- разброс – стандартное отклонение ансамбля, усредненное по тем же самым районам и переменным, которые использованы для среднего значения по ансамблю;
- показатель CRPS.

Дополнительно рекомендованные статистические данные:

– вероятностные показатели — обмен показателями вероятности конкретных пороговых значений проводится в форме таблиц надежности. Ряд различных показателей рассчитывается ведущими центрами на основе таблиц надежности, предоставленных участвующими центрами.

– спецификации верификации прогнозов, изложенные в пунктах ниже, применяются для расчета CRPS и вероятностных показателей. Верификация среднего значения по ансамблю и разброса должна соответствовать спецификациям, изложенным в приложении 2, как было указано выше.

2. Параметры

Среднеквадратическая ошибка и коэффициент корреляции между прогнозируемыми и проанализированными аномалиями среднего значения по ансамблю рассчитывается для следующего набора параметров:

- ДСУМ;
- геопотенциальная высота на уровне 500 гПа;
- компоненты u и v ветра на уровнях 850 и 250 гПа;
- температура на уровне 850 гПа.

Разброс рассчитывается по набору параметров, аналогичному для среднего значения по ансамблю.

Таблицы надежности для расчета вероятностных показателей вычисляются для следующего набора параметров и пороговых значений:

- аномалии ДСУМ, ± 1 , $\pm 1,5$ и ± 2 стандартного отклонения по отношению к определенным климатологическим данным;
- аномалии геопотенциальной высоты 500 гПа с пороговыми значениями ± 1 , $\pm 1,5$, ± 2 стандартного отклонения по отношению к определенным климатологическим данным;
- скорость ветра на уровне 850 гПа с пороговыми значениями 10, 15 и 25 м/с;
- компоненты ветра u и v на уровне 850 гПа с пороговыми значениями в 10, 25, 75 и 90-м процентильных пунктах по отношению к определенным климатологическим данным;

– компоненты ветра u и v на уровне 250 гПа с пороговыми значениями в 10, 25, 75 и 90-м процентильных пунктах по отношению к определенным климатологическим данным;

– аномалии температуры на уровне 850 гПа с пороговыми значениями ± 1 , $\pm 1,5$ и ± 2 стандартного отклонения по отношению к определенным климатологическим данным;

– количество осадков с пороговыми значениями 1, 5, 10 и 25 мм за 24 часа;

– скорость ветра на высоте 10 м с пороговыми значениями 10 и 15 м/с;

– аномалии температуры на высоте 2 м с пороговыми значениями ± 1 , $\pm 1,5$ и ± 2 стандартного отклонения по отношению к определенным климатологическим данным.

Примечание: там, где пороговые значения определяются по отношению к климатологическим данным, определенные климатологические данные указаны в п. 9 ниже.

CRPS рассчитывается для набора параметров, аналогичного для вероятностного показателя.

3. Время инициализации прогноза

Показатели ежедневно рассчитываются для прогнозов, инициализированных в сроки, указанные центром, однако следует включать все прогностические циклы, информация о которых доступна через ИСВ.

4. Шаги прогноза

Шагами прогноза являются каждые 24 часа до окончания срока прогноза.

5. Районы

Внетропические районы Северного полушария	90–20° с. ш. включительно, все долготы
Внетропические районы Южного полушария	90–20° ю. ш. включительно, все долготы
Тропические районы	20° с. ш. – 20° ю. ш. включительно, все долготы

Верификация с помощью анализа должна проводиться для точек сетки внутри каждого района, включая точки на границе.

6. Верификация с помощью анализа. Сетка и интерполяция

Все параметры, за исключением осадков, верифицируются с помощью собственного анализа центра на регулярной сетке $1,5^\circ \times 1,5^\circ$.

При выборе сетки для верификации учитывалось разнообразие разрешений существующих глобальных моделей ЧПП, разрешенные

масштабы моделей (несколько шагов сетки), разрешение имеющихся климатологий, возможности для мониторинга долгосрочных тенденций качества (включая более ранние прогнозы с более низким разрешением) и эффективности вычислений.

Интерполяция полей моделей с более высоким разрешением на сетку для верификации осуществляется, чтобы сохранить характеристики в масштабе сетки для верификации, а не для того, чтобы ввести дополнительное выравнивание.

Используются следующие *процедуры*:

– *спектральные поля*: усечение до эквивалентного спектрального разрешения (T120) для сетки для верификации;

– *поля сеточных значений*: использование осреднения по пространству для интерполяции на сетку для верификации.

Для показателей, требующих климатологии, климатологические данные определяются в п. 9 ниже.

Верификацию прогнозов атмосферных осадков рекомендуется осуществлять по наблюдениям (пункт 7), однако она может также выполняться с помощью вспомогательного анализа (т. е. по краткосрочному прогнозу, полученному из контрольного прогноза, или детерминистскому прогнозу с высоким разрешением, например прогноз на 12–36 часов, чтобы избежать нарастания проблем).

7. Верификация по наблюдениям

• *Наблюдения*

Данные наблюдений для верификации прогнозов от САП основаны на списке станций сети приземных наблюдений ГСНК. Центры подготовки прогнозов имеют право не использовать данные определенных пунктов наблюдений, если они не прошли контроль качества.

• *Интерполяция*

Верификация проводится с использованием исходного узла сетки модели, ближайшего к месту наблюдений.

• *Районы*

Сети, используемые при верификации по данным наблюдений, включают станции наблюдений, расположенные в районах, перечисленных в разделе 6.

8. Показатели

Среднеквадратическая ошибка и коэффициент корреляции между прогнозируемыми и проанализированными аномалиями должны рассчитываться для всех параметров участвующими центрами и предоставляться ведущему(им) центру(ам) в виде, указанном на веб-сайте(ах) ведущего(их) центра(ов).

Следующие показатели должны быть рассчитаны для всех параметров (выработанных ведущим(ими) центром(ами) на основе таблиц надежности, предоставленных участвующими центрами):

- показатель успешности Брайера (по отношению к климатологии);
- ROC-кривая и площадь под ней;
- диаграммы относительной экономической ценности (C/L);
- диаграммы надежности с распределением частоты.

CRPS должен рассчитываться для всех параметров участвующими центрами и предоставляться ведущему(им) центру(ам) в формате, указанном на веб-сайте(ах) ведущего(их) центра(ов). Центрам настоятельно рекомендуется представлять показатели CRPS как для прогноза САП, так и для детерминистского (контрольного или с высоким разрешением) прогноза – CRPS для детерминистского прогноза равен средней абсолютной ошибке.

9. Климатология

Для обеспечения согласованности результатов, получаемых различными центрами, для показателей, которым необходимы климатологические данные, используется общая климатология. Все центры используют климатологию, предоставляемую через веб-сайт(ы) ведущего(их) центра(ов).

Ежедневная климатология аэрологических параметров доступна как для 00.00 ч, так и для 12.00 ч ВСВ. Это обеспечивает актуальную оценку климатических характеристик для каждого дня года, включая климатические средние величины, стандартное отклонение и отобранные квантили распределения климатических характеристик. Эти последние статистические данные необходимы для принятой КОС стандартизированной верификации прогнозов САП. Данные представляются в формате GRIB. Информация о доступе к данным и дальнейшей документации размещается на веб-сайте(ах) ведущего(их) центра(ов) по верификации.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Багров Н.А. К вопросу об оценке гидрометеорологических прогнозов // *Метеорология и гидрология*. – 1953. – № 6. – С. 13-16.
2. Бундель А.Ю., Астахова Е.Д., Розинкина И.А., Семенов А.Е., Алферов Д.Ю. Верификация кратко- и среднесрочных прогнозов осадков на основе системы ансамблевого моделирования Гидрометцентра России // *Метеорология и гидрология*. – 2011. – № 10. – С. 20-34.
3. Бундель А.Ю., Муравьев А.В., Ольховая Е.Д. Обзор методов пространственной верификации и их применение для ансамблевых прогнозов // *Гидрометеорологические исследования и прогнозы*. – 2021. – № 4 (382). – С. 30-49.
4. Груза Г.В., Ранькова Э.Я. Вероятностные метеорологические прогнозы. – Л.: Гидрометеоиздат, 1983. – 280 с.
5. Зарипов Р.Б., Павлюков Ю.Б., Шумилин А.А., Травов А.В. Использование радиолокационной информации при оценке прогноза погоды с высоким разрешением // *Гидрометеорологические исследования и прогнозы*. – 2018. – № 2 (368). – С. 60-86.
6. Киктев Д.Б., А.В. Муравьев, Смирнов А.В. Наукастинг метеорологических параметров и опасных явлений: опыт реализации и перспективы развития // *Гидрометеорологические исследования и прогнозы*. – 2019. – № 4 (374). – С. 92-111.
7. Лорис-Меликов М.А. К вопросу об оценке краткосрочных прогнозов // *Известия Главной геофизической обсерватории*. – 1927. – № 1. – С. 19-25.
8. Методические указания. Проведение производственных (оперативных) испытаний новых и усовершенствованных методов гидрометеорологических и гелиогеофизических прогнозов. РД 52.27.284 – 91. –М.: Госгидромет СССР. 1991, 151 с.
9. Мирвис В.М., Мелешко В.П., Львова Т.Ю., Матюгин В.А., Байдин А.В. Прогностические эксперименты на основе совместной модели океан-атмосфера ГГО // *Труды ГГО*. – 2016. Вып. 583. – С. 129-148.
10. Мирвис В.М., Львова Т.Ю., Мелешко В.П., Матюгин В.А. Опыт калибровки вероятностных прогнозов приземной температуры воздуха, полученных на основе ансамблевых расчетов по гидродинамической модели атмосферы ГГО // *Труды ГГО*. – 2012. – Вып 566. – С. 7-19.
11. Муравьев А.В., Бундель А.Ю., Киктев Д.Б., Блинов Д.В., Смирнов А.В. Верификация мезомасштабных прогнозов в районе Олимпийских игр 2014 г. по первому тестовому периоду. Часть I: Технология верификации и полигонные оценки качества прогнозов на основе модели COSMO // *Метеорология и гидрология*. – 2013. – № 11. – С. 5-21.
12. Муравьев А.В., Бундель А.Ю., Киктев Д.Б., Смирнов А.В. Верификация мезомасштабных прогнозов в районе Олимпийских игр 2014 г. по первому тестовому периоду. Часть II: Предварительные результаты диагностических оценок качества и калибровка прогнозов модели COSMO-Ru2 // *Метеорология и гидрология*. – 2013. – № 12. – С. 5-22.
13. Муравьев А.В., Киктев Д.Б., Бундель А.Ю., Дмитриева Т.Г., Смирнов А.В. Верификация прогнозов метеорологических явлений со значительными воздействиями в районе проведения Олимпиады «Сочи-2014». Часть I // *Метеорология и гидрология*. – 2015. – № 9. – С. 31-48.
14. Муравьев А.В., Киктев Д.Б., Смирнов А.В. Оперативная технология наукастинга осадков на основе радарных данных и результаты верификации для

теплого периода года (май-сентябрь 2017 г.) // Гидрометеорологические исследования и прогнозы. – 2018. – № 1(367). – С. 6-38.

15. Наставление по Глобальной системе обработки данных и прогнозирования: Дополнение IV к Техническому регламенту ВМО // ВМО-№ 485. – Женева: ВМО, 2019. – 166 с.

16. РД 52.27.724-2009. Наставление по краткосрочным прогнозам погоды общего назначения. – М.; Обнинск: ИГ-СОЦИН, 2009. – 61 с.

17. РД 52.27.724-2019. Наставление по краткосрочным прогнозам погоды общего назначения. – М.: Гидрометцентр России, 2019. – 66 с.

18. Обухов А.М. К вопросу об оценке успешности альтернативных прогнозов // Известия АН СССР, сер. геофиз. – 1955. – № 4. – С. 339-349.

19. Руководство по краткосрочным прогнозам погоды. Часть 1: Изд. третье. – Л.: Гидрометеиздат, 1986. 700 с.

20. Омшанский М.А. Контроль альтернативных прогнозов // Труды ГГО. – 1937. – Вып. 14. – С. 49-57.

21. Хандожко Л.А. Оценка успешности метеорологических прогнозов: Учебное пособие. – Л.: Ленинградский политехнический институт, 1977. – 68 с.

22. Хандожко Л.А. Экономическая метеорология. – СПб.: Гидрометеиздат, 2005. – 490 с.

23. Юдин М.И. Новые методы и проблемы краткосрочного прогноза погоды. – Л.: Гидрометеиздат, 1963. – 402 с.

24. Юдин М.И. Синоптический материал для испытания численных методов прогноза // Метеорология и гидрология. – 1960. – № 7.

25. Anderson J.L. A method for producing and evaluating probabilistic forecasts from ensemble model integrations // J. Climate. – 1996. – V. 9. – P. 1518-1530.

26. Brier G.W. Verification of Forecasts Expressed in Terms of Probability // Monthly Weather Review. – 1950. – Vol. 78. – P. 1-3.

27. Brier G.W., Allen R.A. Verification of weather forecasts // Compendium of Meteorology. – Boston, Massachusetts: American Meteorological Society, 1951. – P. 841-848.

28. Brown B.G., Bullock R., David C.A., Gotway J.H., Chapman M.B., Takacs A., Gilleland E., Manning K., Mahoney J. New verification approaches for convective weather forecasts // 11th Conf. Aviation, Range and Aerospace Meteorology, Amer. Met. Soc., 4-8 Oct 2004, Hyannis, MA, 2004.

29. Candille G., Côté C., Houtekamer P.L., Pellerin G. Verification of an ensemble prediction system against observations // Monthly Weather Review. – 2007. – Vol. 135. – P. 2688-2699.

30. Casati B., Ross G., Stephenson D.B. A New intensity-scale verification approach for the verification of spatial precipitation forecasts // Meteorol Appl. – 2004. – Vol. 11. – P. 141-154.

31. Casati B., Wilson L.J., Stephenson D.B., Nurmi P., Ghelli A., Pocerlich M., Damrath U., Ebert E.E., Brown B.G., Mason S. Review. Forecast verification: current status and future directions // Meteorol. Appl. – 2008. – Vol. 15. – P. 3-18.

32. Coles S., Heffernan J., Tawn J. Dependence measures for extreme value analysis // Extremes. – 1999. – Vol. 2. – P. 339-365.

33. Davis C.A., Brown B.G., Bullock R.G. Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas // Mon. Wea. Rev. – 2006. – Vol. 134. – P. 1772-1784.

34. Davis C., Brown B., Bullock R., Halley-Gotway J. The Method for Object - Based Diagnostic Evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC spring program // *Wea. Forecasting* – 2009. – Vol. 24. – P. 1252-1267.
35. DeGroot M.W., Fienberg S.E. Assessing probability assessors: Calibration and refinement // *Statistical decision theory and Related Topics III*. – New York: Academic Press, 1982. – P. 291–314.
36. Dey S.R., Leoncini G., Roberts N.M., Plant R.S., Migliorini S. A spatial view of ensemble spread in convection permitting ensembles // *Mon. Wea. Rev.* – 2014. – Vol. 142. – P. 4091-4107. – <https://doi.org/10.1175/MWR-D-14-00172.1>.
37. Dobryshman E.M. Review of forecast verification techniques // WMO-No. 303. Technical note No.120. Geneva, Switzerland: WMO, 1972. – 65 p.
38. Dorninger M., Mittermaier M.P., Gilleland E., Ebert E.E., Brown B.G., Wilson L.J. 2013: MesoVICT: Mesoscale Verification Inter-Comparison over Complex Terrain (No. NCAR/TN-505+STR).
39. Dorninger M., Gilleland E., Casati B., Mittermaier M.P., Ebert E.E., Brown B.G., Wilson L.J. The setup of the MesoVICT project // *Bull. Amer. Meteor. Soc.* – 2018. – Vol. 99. – P. 1887-1906.
40. Ebert E.E. Suggested methods for the verification of precipitation forecasts against high resolution limited area observations (JWGFVR). – WMO, Geneva, Switzerland, Second Draft, November 2013.
41. Ebert E.E., McBride J.L. Verification of precipitation in weather systems: Determination of systematic errors // *J. Hydrol.* – 2000. – Vol. 239. – P. 179-202.
42. Ebert E. Fuzzy verification of high resolution gridded forecasts: A review and proposed framework // *Meteor. Appl.* – 2008. – Vol. 15. P. 51–64. doi:10.1002/met.25.
43. Ebert E.E., Gallus W.A. Toward Better Understanding of the Contiguous Rain Area (CRA) Method for Spatial Forecast Verification // *Weather and Forecasting*. – 2009. – Vol. 24. – P. 1401-1415.
44. Ferro C.A.T., Stephenson D.B. Extremal Dependence Indices: improved verification measures for deterministic forecasts of rare binary events // *Wea. Forecasting*. – 2011. – Vol. 26. – P. 699-713.
45. Ferro C.A.T., Stephenson D.B. Extremal dependence indices: improved verification measures for extreme events and warnings. // *Weather and Forecasting*. – 2012. – Vol. 26. – P. 699-713.
46. Ferro C.A.T., Stephenson D.B. Deterministic forecasts of extreme events and warnings // *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. – Wiley, 2011. – P. 185-202.
47. Forecast Verification in Atmospheric Science. A Practitioner's Guide: Second Ed. // I. Jolliffe, D. Stephenson (Eds.). – John Wiley & Sons Ltd, 2012. – 274 p.
48. Gandin L.S., Murphy A.H. Equitable skill scores for categorical forecasts // *Monthly Weather Review*. – 1992. – Vol. 120. – P. 361-370.
49. Gerrity J.P. A note on Gandin and Murphy's equitable skill score // *Monthly Weather Review*. – 1992. – Vol. 120. – P. 2707-2712.
50. Gilleland E., Ahijevych D., Brown B.G., Casati B., Ebert E.E. Intercomparison of spatial forecast verification methods // *Wea. Forecasting*. – 2009. – Vol. 24. P. 1416-1430.
51. Gilleland E., Ahijevych D., Brown B.G., Ebert E.E. Verifying forecasts spatially // *Bull. Amer. Meteor. Soc.* – 2010. – Vol. 91. – P. 1365-1376.
52. Gilleland E. Testing competing precipitation forecasts accurately and efficiently: the spatial prediction comparison test // *Mon. Weather Rev.* – 2013. – Vol. 141. – P. 340-355.

53. Gilleland E. SpatialVx: Spatial forecast verification. R package version 0.7. 2020: [[https://CRAN.R-project.org/package=SpatialVx.](https://CRAN.R-project.org/package=SpatialVx)]
54. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation // *Journal of the American Statistical Association.* – 2007. – Vol. 102. – P. 359-378.
55. Gorgas T, Dorninger M. Concepts for a pattern-oriented analysis ensemble based on observational uncertainties // *Q. J. R. Meteorol. Soc.* – 2012. – Vol. 138. – P. 769-784. DOI:10.1002/qj.949
56. Gorgas T., Dorninger M. Quantifying verification uncertainty by reference data variation // *Meteorol. Z.* – 2012. – Vol. 21 (3). – P. 259-277. – DOI 10.1127/0941-2948/2012/0325
57. Hamill T.M, Colucci S.J. Verification of Eta-RSM short-range ensemble forecasts // *Mon. Wea. Rev.* – 1997. – Vol. 125. – P. 1312-1327.
58. Hamill T.M., Colucci S.J. Evaluation of Eta-RSM ensemble probabilistic precipitation forecasts // *Mon. Wea. Rev.* – 1998. – Vol. 126. – P. 711-724.
59. Hannart A. https://www.wcrp-climate.org/images/summer_school/... Extreme_value_theory_Hannart.pdf, 2019
60. Johnson A., Wang X., Wang Y., Reinhart A., Clark A.J., Jirak I.L. Neighborhood- and Object-Based Probabilistic Verification of the OU MAP Ensemble Forecasts during 2017 and 2018 Hazardous Weather Testbeds // *Weather and Forecasting.* – 2020. – Vol. 35. – P. 169-191. <https://doi.org/10.1175/WAF-D-19-0060.1>
61. Jolliffe I.T. The impenetrable hedge: a note on propriety, equitability and consistency // *Meteorological Applications.* – 2008. – Vol. 15. – P. 25–29.
62. Marsigli C., Ebert E., Ashrit R., Casati B., Chen J., Coelho C.A.S., Dorninger M., Gilleland E., Haiden T., Landman S., Mittermaier M. Review article: Observations for high-impact weather and their use in verification // *Nat. Hazards Earth Syst. Sci.* – 2021. – Vol. 21. – P. 1297-1312. – <https://doi.org/10.5194/nhess-21-1297-2021>,
63. Murphy A.H. What is a good forecast? An essay on the nature of goodness in weather forecasting // *Weather and Forecasting.* – 1993. – Vol. 8. – P. 281-293.
64. Murphy A.H. The Finley affair: a signal event in the history of forecast verification // *Weather and Forecasting.* – 1996. – Vol. 11. P. 3-20.
65. Murphy A.H., Winkler R.L. A general framework for forecast verification // *Monthly Weather Review.* – 1987. – Vol. 115. – P. 1330-1338.
66. Murphy A.H. A note on the utility of probabilistic predictions and the probability score in the cost-loss ratio decision situation // *J. Appl. Meteor.* – 1966. – Vol. 5. P. 534-537.
67. Panofsky H.A., Brier G.W. Some applications of statistics to meteorology. University Park, Pennsylvania, 1958. (Пановский Г.А., Брайер Г.В. Статистические методы в метеорологии. Пер. с англ. 2 изд. 1972, (1-е ориг. 1958 г.). – Л.: Гидрометеоиздат. – 243 с.)
68. Roberts N.M., Lean H.W. Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events // *Monthly Weather Review.* – 2008. – Vol. 136. – P. 78-97. – <https://doi.org/10.1175/2007MWR2123.1>
69. Schwartz C.S., Sobash R.A. Generating Probabilistic Forecasts from Convection-Allowing Ensembles Using Neighborhood Approaches: A Review and Recommendations // *Monthly Weather Review.* – 2017. Vol. 145, is. 9. – P. 3397-3418. DOI: <https://doi.org/10.1175/MWR-D-16-0400.1>,
70. Schwartz C.S., Kain J.S., Weiss S.J., Xue M., Bright D.R., Kong F., Thomas K.W., Levit J.J., Coniglio M.C., Wandishin M.S. Toward Improved Convection-Allowing Ensembles: Model Physics Sensitivities and Optimizing Probabilistic Guidance with Small Ensemble Membership // *Weather and Forecasting.* – 2010. – Vol. 25. – P. 263-280. <https://doi.org/10.1175/2009WAF2222267.1>

71. Skamarock W.C. Evaluating mesoscale NWP models using kinetic energy spectra // *Mon. Wea. Rev.* – 2004. – Vol. 132. P. 3019-3032. – <https://doi.org/10.1175/MWR2830.1>.
72. Skok G., Roberts N. Estimating the displacement in precipitation forecasts using the Fractions Skill Score. // *Quarterly Journal of the Royal Meteorological Society.* – 2018. – Vol. 144. – P. 414-425. – <https://doi.org/> <https://doi.org/10.1002/qj.3212>.
73. Skok G. A new spatial displacement measure for continuous fields. IVMW2020. https://www.univie.ac.at/img-wien/jwgfvr/SpatialMethods/2020111615/Skok_2020IVMWO_SpatialMethods_16Nov15UTC.pdf, 2020
74. Stanski H.R., Wilson L.J., Burrows W.R. Survey of common verification measures in meteorology. – Downsview, Ontario, Canada, Atmospheric Environment Service, 1989. – Research Report No. 89-5. 113 p.
75. Talagrand O. Statistical consistency of ensemble prediction systems // Discussion paper, SAC Working Group on EPS, ECMWF. – 1997.
76. Wilks D.S. *Statistical Methods in the Atmospheric Sciences.* – Elsevier, 2019. – 816 pp.
77. Wilson L.J., Burrows W.R., Lanzinger A. A strategy for verification of weather element forecasts from an ensemble prediction system // *Monthly Weather Review.* – 1999. Vol. 127. – P. 956-970.
78. WMO-No. 834. Guide to public weather services practices. Second Edition. 1999. Ch.10. Verification, service evaluation and improvement. (ВМО-834. Руководство по практике метеорологического обслуживания населения. Второе издание. 2000. Гл.10. Верификация, оценка и повышение качества обслуживания.)
79. WMO-No. 1023. Guidelines on performance assessment of public weather services. Geneva, Switzerland, 2000.

Д.Б. Киктев, А.В. Муравьев, А.Ю. Бундель

**Методические рекомендации
по верификации метеорологических
прогнозов**

Издатель: ФГБУ «Гидрометцентр России»
Адрес: 123376, Москва, Большой Предтеченский переулок, д. 13, стр. 1
Телефон: (499) 252-34-48
e-mail: hmc@metcom.ru
www.meteoinfo.ru

Подписано в печать 13.12.2021. Формат 60×84/16
Печать офсетная. Печ. л. 5,875. Тираж 100 экз. Заказ № 2009
Отпечатано в типографии
ООО «Типография АМА-ПРЕСС»
107392, г. Москва, Зельев пер., дом 3