

DOI: <https://doi.org/10.37162/2618-9631-2022-2-6-52>

УДК 551.509.313+551.509.324.2+551.508.85

Опыт пространственной верификации радиолокационного наукастинга осадков: определение и статистика объектов, ситуаций и условных выборок

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

*Гидрометеорологический научно-исследовательский центр
Российской Федерации, г. Москва, Россия
muravev@mecom.ru, boundel@mecom.ru*

Проведен статистический анализ пространственных объектов и специальных ситуаций, определяемых для объектно-ориентированной верификации прогнозов областей осадков значительной и максимальной площади с привлечением методов теории экстремальных величин. Оценена степень влияния пропусков в точках полей и разных способов формирования пар прогноз-наблюдение на объемы и на статистические свойства выборок для пространственной верификации. С помощью пространственных квантильных функций и географических представлений продемонстрированы особенности композитных полей, построенных на основе данных около трех десятков радиолокаторов на Европейской территории России.

Ключевые слова: пространственная верификация прогнозов, радиолокационный наукастинг осадков, теория экстремальных величин, пропуски данных, условные выборки для верификации, пространственные квантили

Expertise in spatial verification of radar precipitation nowcasting: identification and statistics of objects, situations and conditional samples

A.V. Muravev, A.Yu. Bundel, D.B. Kiktev, A.V. Smirnov

*Hydrometeorological Research Center of Russian Federation, Moscow, Russia
muravev@mecom.ru, boundel@mecom.ru*

Statistical analysis was performed using methods of the extreme value theory for spatial objects and specified situations identified for object-oriented verification of precipitation regions with substantial and maximal areas. We made an estimation of the effect of missing values at field points and of different observation-forecast pairs construction on volumes and on statistical characteristics of samples retrieved for spatial verification purposes. We used spatial quantile functions and geographical representations in regular coordinates to illustrate particular aspects of composite fields built on about three dozen radars' data over the European territory of Russia.

Keywords: spatial forecast verification, radar precipitation nowcasting, extreme value theory, missing data, conditional verification sampling, spatial quantiles

Введение

На протяжении 2017–2020 гг. в теплый и холодный периоды года в Гидрометцентре России проводились испытания двух версий системы радиолокационного наукастинга осадков, которые обозначим по аналогии прогностическим моделям через STEPS-2014 и ruSTEPS-2019. Особенности указанных моделей, условия и результаты испытаний изложены в нескольких работах авторов [13, 14, 20–23]. К теплому и холодному периодам отнесены по пять последовательных месяцев из интервалов май–сентябрь и ноябрь–март соответственно.

Первая версия модели STEPS-2014 испытывалась по территории Центрального федерального округа (ЦФО) на основе данных наблюдений девяти радиолокаторов ДМРЛ-С (р/л). Вторая версия ruSTEPS-2019, усовершенствованная, испытывалась по Европейской территории России на основе данных 28 р/л (рис. 1).

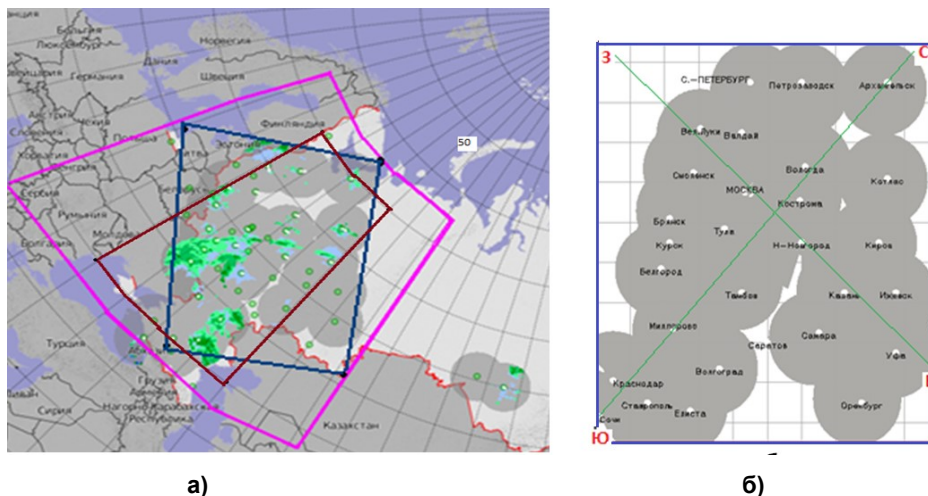


Рис. 1. Область усовершенствованной системы наукастинга радиолокационных осадков: оперативная карта ЦФО и контуры характерных областей; сиреневый – начальная область анимации прогнозов (сайт ГМЦ); синий – объединенное поле круговых обзоров р/л (матрица 2151х1951); коричневый: условная карта для верификации (45°–65° с. ш., 30°–50° з. д.) (а); расположение ДМРЛ-С в условных координатах (28 р/л). Прямыми зелеными линиями на панели выделены ориентировочные направления на страны света (б).

Fig. 1. Area of improved radar precipitation nowcasting: operational map of the Central Federal district and the contours of characteristic areas; orange: initial area as in forecast animation (HMC site); blue: unified field of the circle radar area (matrix 2151x1951); brown: verification map in regular coordinates (45°–65°N, 30°–50°W) (a); Radar locations in special regular coordinates (28 radars). Straight green lines indicate the cardinal points (б).

Наблюдения и прогнозы в первом случае представлены в виде квадратных матриц, включающих данные круговых обзоров отдельных локаторов, во втором случае – в виде прямоугольной матрицы, объединяющей данные всех локаторов с учетом их схематического географического расположения. Указанные матрицы будем также называть *полями* и *объединенными полями* соответственно.

Базовые, традиционные *оценки качества* наукастинга строились на временных рядах пар наблюдение-прогноз в узлах полей с последующим обобщением при помощи квантилей пространственного распределения этих оценок. Наряду с традиционными оценками рассчитывались показатели *объектно-ориентированных* и *окрестных* методов верификации [20, 31, 35, 36]. В рамках объектно-ориентированной верификации выделялись области осадков *большой* и *максимальной* площади, полная совокупность которых описывалась моделью "пиков над порогом" (POT, Peaks-over-Threshold) и анализировалась с помощью *обобщенного распределения Парето* из теории экстремальных величин [6–8, 17, 28, 46, 56]. Методы окрестной верификации были представлены *показателем FSS* (Fractions Skill Score) [48], который строится на пространственных долях областей наблюдаемых и прогнозируемых зон осадков.

Усовершенствованная система наукастинга была испытана и внедрена в оперативную практику в течение 2019–2020 гг. [23]. Здесь основное внимание уделено учету *пропусков* в данных наблюдений и разным способам *формирования выборок* пар наблюдение-прогноз. Информация о пропусках важна для оценок значимости характеристик качества, а испытание разных типов выборок позволяет учесть влияние на характеристики качества большого количества нулевых осадков и некоторого нижнего порога наличия осадков, учитывающего точность радиолокатора.

Обобщение оценок качества с помощью квантилей пространственного распределения можно отнести к группе *масштабно-разделительных* методов.

Указанные группы входят в классификацию методов пространственной верификации [38], в которую наряду с перечисленными тремя (объектно-ориентированные, окрестные и масштабные-разделительные) входит группа *деформационных методов* и иногда добавляется группа *мер дальности* [34]. Данная классификация не является строгой и здесь используется исключительно для отсылки на мировой опыт [3, 38, 39].

Уточним понятия *областей*, *объектов* и *ситуаций*. В полях осадков *области* выделяются, как обычно, пороговыми значениями. Ограниченные изолинией 1 мм/ч области будем называть *объектами* в соответствии с терминологией объектно-ориентированной верификации [31]. *Площадь*, или *размер объекта* оцениваются по количеству попадающих в него ячеек (или узлов), при таком определении связанная одномерная «линия» всегда имеет ненулевую «площадь». Объекты значительной площади (*значительные*, или *экстремальные объекты*), составляющие выборку

для приложения теории экстремумов, извлекаются из временной последовательности полей, которая отражает некоторую *непрерывно* развивающуюся синоптическую мезомасштабную *ситуацию*, или *дождевую систему* [1, 4, 32, 42]. Под *точной непрерывной* последовательностью полей понимается чередование их временных сроков в исходной 10-минутной дискретности; будем называть просто *непрерывной* такую последовательность, в которой, ввиду возможных аварий или нарушений информационного потока, допускаются один-два пропуска между соседними полями с возможностью интерполяции недостающих значений.

Пороговые значения определяются на основе документированного опыта и по результатам авторских экспериментов с полями осадков, а также с пакетами статистической обработки. Например, интенсивность 1 мм/ч для 10-минутных радиолокационных сумм является в климатологическом смысле достаточно высокой в средних широтах [1, 2, 11, 12], и в наших экспериментах такой порог обеспечивает не только заметные по площади объекты, но и временную непрерывность содержащих их полей в вышеуказанном смысле.

Объекты, ситуации и объемы выборок в точках полей осадков имеют важные статистические и географические свойства [31]; в частности, фундаментальные соотношения между пространственными и временными масштабами в метеорологических полях лежат в основе использованной модели наукастинга [13, 29, 51–53]. Заметим, что методы анализа пространственных характеристик областей осадков, включая экстремальные, давно применяются в гидрологии и содержатся в соответствующих нормативных документах ВМО [5, 57]. Описанию и анализу этих особенностей полей наблюдений и прогнозов в серии проведенных испытаний посвящена настоящая, *первая статья* предполагаемой серии. Качество прогнозов областей осадков значительной площади на основе обобщенного распределения Парето и качество прогнозов на основе отношения пространственных долей с помощью показателя *FSS* будут проанализированы во *второй и третьей статьях* предполагаемой серии соответственно; в этих работах областями испытаний являются круговые обзоры отдельных радиолокаторов. Результаты верификации прогнозов по полю, объединяющему обзоры 28 локаторов на Европейской территории России, планируется привести в *четвертой статье*.

Сделаем несколько частных замечаний. Иллюстративные примеры относятся к конкретным случаям, отдельным локаторам и выделенным срокам наблюдений и прогнозов, но суммарные статистические характеристики рассчитывались для всех локаторов (с небольшими исключениями), для обоих периодов испытаний и различных прогностических сроков. Расчетные и графические операции производились с помощью библиотек и функций в среде языка R, картографические образы создавались с помощью системы сеточного и графического анализа Grads. Большинство расчетов содержало разномасштабные числа, которые было

бы естественно округлять и обобщать, однако для полноты представления мы предпочли сохранять основную их часть в исходном, не обработанном, но максимально точном виде.

1. Построение объединенного поля радиолокационных данных на Европейской территории России

Опишем область ответственности усовершенствованной схемы наукастинга 2020 г. и методы построения объединенного (или *композиционного*) поля. На оперативной карте Центральной аэрологической обсерватории (ЦАО) (рис. 1а) эта область выделена синим полигоном, в нерегулярных координатах которого размещена сетка с 1951 точками по юго-восточной и северо-западной сторонам и 2151 точками по юго-западной и северо-восточной сторонам. Внутри этого полигона в настоящее время эксплуатируются двадцать восемь ДМПЛ-С: RUDO (Оренбург), RUDP (Петрозаводск), RUTD (Тамбов), RUDX (Архангельск), RAVO (Воейково), RAKT (Котлас), RUDB (Брянск), RUDV (Вологда), RUDG (Волгоград), RATL (Тула), RAKD (Краснодар), RUDT (Ставрополь), RUWJ (Валдай), RAMI (Миллерово), RAVN (Внуково), RAKU (Курск), RUDI (Ижевск), RAYL (Элиста), RUDL (Смоленск), RASM (Самара), RABG (Белгород), RUDM (Мин. Воды), RUDK (Кострома), RUDZ (Казань), RAVL (Вел. Луки), RUDU (Уфа), RUDN (Ниж. Новгород), RAKW (Киров). В предыдущей схеме наукастинга и в испытаниях 2017-2019 гг. были задействованы девять р/л, покрывающих в основном территорию ЦФО: RAVO, RUDB, RATL, RUWJ, RAVN, RAKU, RUDL, RUDK, RUDN.

Объединенные поля наблюдений и прогнозов, как и поля показателей качества, отображаются на *условную географическую карту* (рис. 1б), на которой сохранена структура расположения локаторов, а географическая сеть задана в регулярных широтно-долготных координатах 45–65° и 30–50° с подходящими инкрементами для вложения матрицы размером 2151×1951.

Применение условной картографии оправдано тем, что в рабочей версии системы Grads существует только два способа географического представления данных – в виде регулярной географической сетки или с помощью точечных координат (в нашем случае в количестве порядка четырех миллионов). Очевидно, в последнем случае и составление координат в текстовом формате, и выведение в графику с помощью внутренних процедур сглаживания превращаются в трудно исполнимое предприятие. К тому же детальная географическая картина на данном этапе анализа не требуется. Названия пунктов на рис. 1б в центрах кругов соответствуют спецификации оперативной карты радиолокационной сети ЦАО, прямыми линиями выделены ориентировочные направления на страны света.

Опишем процедуру формирования композиционного поля. Входные композиты размером 2151×1951, заполненные вначале признаками

отсутствия данных (числом -999), замещаются в соответствующих координатах оценками интенсивности осадков, которые формируются в ЦАО в виде квадратных матриц размером 504×504 с вложенными круговыми зонами обзора. В областях перекрытия зон обзора локаторов значение интенсивности рассчитывается с помощью линейной весовой формулы с приоритетом ближайшего локатора.

Полученные композиты подаются на вход модели ruSTEPS, и на первом шаге расчетного алгоритма по точкам с ненулевой интенсивностью определяется поле адвекции – направление и скорость движения. Дальнейшие операции перемещения, разложения на каскады и экстраполяция по времени формируют эволюцию процесса, а возмущения движения и эволюции на каждом шаге порождают прогностический ансамбль объединенных полей. Возмущения генерируются датчиком нормально распределенных случайных чисел, параметры которого настроены на стандартные ошибки радиолокационных оценок осадков.

Выходные поля модели не содержат признаков отсутствия данных, кроме тех случаев, когда движение начинается от края видимой зоны. Точки с константами отсутствия в исходных радиолокационных полях переносятся на каждом шаге прогноза в выходное поле модели. Несомненно, полезное свойство усовершенствованной схемы состоит в том, что адвекция точек поля с числовым значением позволяет заполнять осадками некоторые точки, которые исходно были "пустыми" [54].

2. Выборки объектов значительной площади: ситуации с пиком

В нашем подходе рассматривается оценка способности модели воспроизводить статистические характеристики тех областей осадков, площадь которых либо максимальна, либо превышает некоторый пороговый уровень. Как известно, распределения экстремальных величин описываются теоремой Фишера – Типпета – Гнеденко (*первая теорема* теории экстремальных величин, ТЭВ), а распределения превышений порогового уровня описываются теоремой Пиканса – Балкема – де Хаана (*вторая теорема* ТЭВ). В рамках каждой из этих моделей существует по три основных закона *притяжения*, т. е. три класса распределений, к которым асимптотически стремятся распределения некоторым образом нормированных экстремумов (см., например, монографии [9, 10, 28, 33, 46, 47]). Так как исходные данные являются временными последовательностями полей, то статистический анализ может рассматриваться в рамках теории выбросов ("максимумов", "пиков", "превышений") случайных процессов (см., например, монографии [6, 7, 15, 17, 18, 26]).

Проблемы моделирования на основе этих двух теорем ТЭВ относятся к "классической теории экстремальных величин", считающейся в некотором смысле завершенной [7]. Однако *статистический анализ* экстремальных величин по конкретному набору данных (*статистический вывод*) оказывается намного более сложным и далеким от завершения,

что отмечается во всех литературных источниках по ТЭВ, затрагивающих проблемы приложений теорем к практике. Эти сложности относятся в первую очередь к проверке условий используемых теорем ТЭВ. Доказать корректность подготовленной выборки (что равносильно проверке соответствия выборки условиям соответствующей теоремы) не только затруднительно, но скорее всего невозможно: "вопрос о проверке выполнения в реальности условий теоремы чаще всего не может быть не только решен, но даже и поставлен" [27]. При моделировании *пиков над порогом* вышеуказанная проверка превращается в корректный выбор порога, а так как для анализа экстремальных значений временных рядов выборка формируется с учетом временной локальности пиков, то приходится дополнительно определять подходящие временные интервалы – носители искомым экстремумов. Для подтверждения условий *второй теоремы* ТЭВ мы учитываем физические и эмпирические представления о характерных пространственно-временных особенностях мезомасштабных осадков [1, 4, 12, 31, 32, 42], а также пуассоновское распределение моментов времени появления экстремумов над выбранными порогами [8, 15, 18, 24, 25, 30, 41, 44, 46, 56].

Конкретизируем некоторые понятия и термины. Будем называть носители искомым пиков *ситуациями с пиком*, или *ситуациями*, со следующим определением – это последовательность полей осадков с интенсивностью не менее 1 мм/ч хотя бы в одной точке в каждом поле и с наличием в этой последовательности хотя бы одного поля с областью сплошных осадков, площадь которой превышает заданное пороговое значение. В данном случае сплошность, или связность области являются характеристиками пространственного осреднения. Порог 1 мм выбран намеренно таким же, как и порог для идентификации *объектов*.

По каждому периоду тестирования из последовательностей около 20 тысяч полей осадков выделялись *непрерывные* (в ранее определенном смысле) последовательности таких полей, в каждом из которых содержатся *объекты* со значениями интенсивности, не меньшими 1 мм/ч. Допускаются пропуски между соседними полями вплоть до 20 минут, которые назовем *интервалами сцепления*. Это отсутствие полей было либо физическое из-за сбоев в информационном потоке, либо алгоритмическое, вызванное занижением интенсивности осадков и "потерей" *объектов* на значениях, чуть меньших 1 мм.

Задача комплектации *значительных объектов* в выборки для верификации с помощью распределения Парето решалась экспериментально на основе анализа гистограмм распределения размеров всех объектов с особым вниманием к правым хвостам этих распределений. Процедуры выделения ситуаций применялись к последовательностям полей наукастинга всех заблаговременностей от 10 до 120 минут. Пороги превышения определялись также экспериментально по набору 625, 900, 1225 и 1600 точек поля и с предпочтением таких значений, которые приводят к наиболее устойчивым оценкам параметров распределения Парето.

3. Особенности определения и выделения объектов

Выделение объектов производится с помощью функции FeatureFinder библиотеки SpatialVx [37] из хранилища языка R на основе следующих параметров:

- порог идентификации – 1 мм/ч;
- параметр свертки (радиус осреднения) – 9 узлов сетки (~18 км);
- минимальные размеры (они же *пороговые значения размера*, или *параметры положения* распределения Парето, или *параметры Парето*) – $625=25\times 25$, 900, 1225, $1600=40\times 40$ узлов сетки (линейный размер одной ячейки равен 2 км).

Подчеркнем еще раз, *объект* выделяется *после* пространственного сглаживания поля.

Рассмотрим пример функционирования модели STEPS и пакета SpatialVx на осадках локатора RAKU (Курск). Обратим внимание только на общую форму объектов без указания значений интенсивности. В качестве начальных данных модели STEPS берутся два последовательных поля отражаемости размером 504×504 км, при этом радиолокационные наблюдения заполняют вписанный круг обзора, вне которого размещаются константы отсутствия. Для использования в модели эти поля интерполировались в сетку 256×256 с разрешением около 2 км, а в процессе преобразования отражаемости в интенсивность осадков константы отсутствия вне круга обзора заменялись небольшими положительными числами порядка 10^{-3} . Указанные манипуляции требовались по условиям функционирования модели STEPS, в которой двумерное преобразование Фурье применялось к полному числовому полю данных, а линейный размер поля должен был выражаться степенью двойки.

Расчет характеристик объектов и графическое представление самих объектов с помощью пакета SpatialVx начинается с прогнозов на первый срок 10 минут (рис. 2а). В этом примере порог равен 1225 точек (эквивалент квадрата со стороной 70 км); хорошо видно, что радиус осреднения в 9 узлов и минимальный порог площади устранили множество мелких несущественных объектов. Построение пар сравниваемых объектов производится по минимальному расстоянию между *центроидами* полей наблюдения и прогноза (т. е. формально рассчитанными "центрами тяжести объектов"); спаренные объекты отмечаются одинаковым цветом.

Как видно по первой строке панелей (рис. 2а), уже на следующем шаге регрессионной схемы происходят несущественные, но заметные изменения выделенного объекта прогнозирования. К сроку 60 мин сам объект (внутри изолинии 1 мм) теряет в площади, но мало изменяет ориентацию, хотя объект в поле наблюдения развернулся примерно на $10\text{--}25^\circ$ против часовой стрелки с усилением полосы интенсивности в южной части, направленной на СВ. К концу прогностического интервала (рис. 2в) объект наблюдения разбивается на два, а прогностический объект, заметно

отставая во вращении, переключается алгоритмом спаривания на "отделившийся аппендикс" – верификация прогноза первоначального объекта теряет смысл. К этому сроку прогноза из атрибутов объектов, по которым еще можно оценивать качество прогноза, осталась, по-видимому, только максимальная площадь объекта в полном поле, а поскольку подобная "потеря пары" встречается довольно часто, то выбор максимальной площади объекта в качестве величины для верификации получает дополнительный кредит.

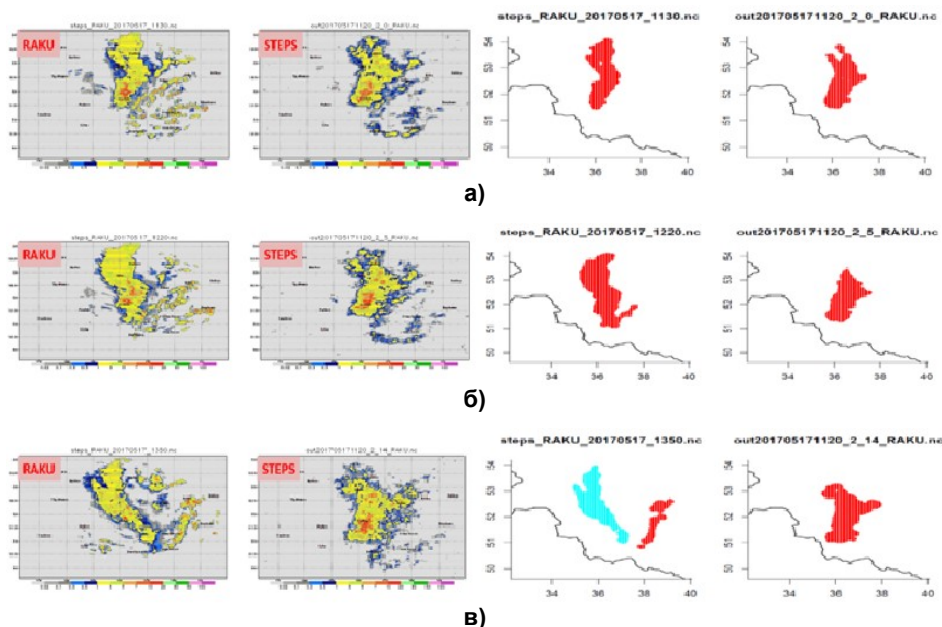


Рис. 2. Эволюция полей и объектов в наблюдении (RAKU, 1-й и 3-й столбцы) и в прогнозе (STEPS, 2-й и 4-й столбцы) через 10 мин (а), 60 мин (б) и 150 мин (в). Затенённые столбцы слева – эволюция полей осадков в исходном разрешении (сторона ячейки около 2 км), светлые столбцы справа – эволюция выделенных объектов с заданными параметрами порога интенсивности осадков (1 мм/ч), радиусом осреднения 9 точек поля (18 км) и минимальным размером 35×35. Красным цветом в двух правых столбцах выделены спаренные объекты, для которых далее рассчитываются характеристики ошибок. Синий объект в наблюдении (в) – объект, оставшийся к концу прогностического интервала "без пары".

Fig. 2. Evolution of fields and objects in the observations (RAKU radar, the 1st and 3rd columns) and in the forecast (STEPS, 2nd and 4th columns) in 10 min (a), 60 min (b), and 150 min (c). Shadowed columns on the left represent the evolution of precipitation fields in the original resolution (grid cell side of about 2 km), the light-colored columns on the right are the evolution of the objects determined with the following parameters: precipitation threshold (1 mm/h), convolution radius of 9 grid points (18 km), and the minimum size of 35×35 (70×70 km square). Red in the columns on the right indicates matched objects, for which the scores are calculated. The blue object in (c) is "unpaired".

3.1. Статистические характеристики ситуаций и объектов

Будем предполагать, что в последовательности полей, организованных в 10-минутной дискретности и обладающих объектами достаточно большого размера, можно возместить отсутствие одного поля или двух соседних полей интерполированием, пользуясь инерцией физического процесса: можно сказать, ситуация сшивается по *интервалу сцепления* в 20 мин. Максимальный размер объекта в каждой *ситуации с пиком* послужит выборочным значением для оценки распределения Парето с *параметром положения* из указанного выше набора 625, 900, 1225 и 1600 точек.

Рассмотрим временной ход площади максимального объекта в последовательных полях обзора локатора Курск в период 5–19 мая 2017 г. (рис. 3); на график наносятся только значения не меньше 625 точек. Сроки с *пиками* отмечены синими линиями; предполагается, что они случайно разбросаны по оси времени в соответствии с простым потоком Пуассона. Отметим несколько особенностей на рис. 3. Как оказалось, в интервале 10.50–12.20 ч ВСВ наблюдения полностью отсутствовали, поэтому пик 0508_1020 следовало бы аннулировать, что, однако, нельзя было исполнить в режиме автоматизированной обработки. Наконец, серия значительных площадей в конце выделенного интервала (до срока 0519_1930) относится к ситуации, выходящей за пределы интервала данного рисунка.

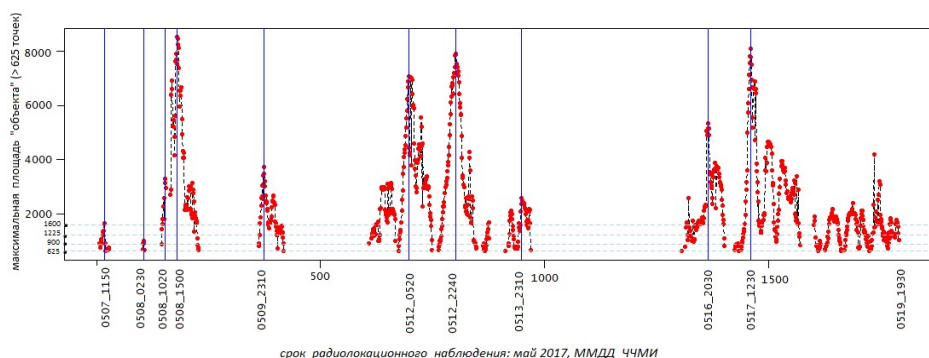


Рис. 3. Площади объектов в последовательных полях наблюдений р/л Курск в период с 5 по 19 мая 2017 г., превышающие порог 625 точек (красные точки). Синие линии соответствуют срокам и максимальным площадям среди объектов, принадлежащих отдельной "ситуации". На оси времени представлены все сроки наблюдений, интервалы с отсутствием точек указывают либо на пропуски, либо на отсутствие объектов, имеющих размер не менее 625 точек.

Fig. 3. Areas of objects not less than the threshold of 625 grid points (red points) in consecutive observation fields of the Kursk radar from 09 May to 19 May 2017. Blue lines correspond to times and the maximum areas among the objects belonging to a situation. The time axis shows all the observation times, intervals without points indicate either missing data or absence of the objects not less than 625 grid points.

Точная информация об *объектах* в нескольких последовательных сроках выделенного интервала показана в табл. 1. В частности, видно наличие полей с набором из нескольких подходящих объектов, включая объекты размером в одну точку.

Таблица 1. Пример ситуации "0507_0910 – 0507_1810" с указанием в полях наблюдения локатора Курск площадей всех выделенных объектов, максимальной площади среди объектов одного поля и "пиковой" площади всей ситуации, используемой в выборке *пиков над порогом*. Отобраны максимальные площади размером не менее 625 точек

Table 1. Example of the situation "0507_0910 – 0507_1810" for the Kursk radar, with the areas of all identified objects, maximum area among the objects of one field, and the "peak" area of the situation used in the *peaks-over-threshold* sample. Maximum areas of not less than 625 grid points are used

Срок	Набор площадей всех объектов поля	Суммарная площадь	Максим. площадь	ПИК "ситуации"
0507_0830		0		
0507_0840		0		
0507_0850		0		
0507_0900		0		
0507_0910	215	215		
0507_0920		0		
0507_0930	257	257		
0507_0940	301	301		
0507_0950	308	308		
0507_1000	422	422		
0507_1010	912	912	912	
0507_1020	769 32 27 89	917	769	
0507_1030	892 127 18 27 72	1136	892	
0507_1040	1086 188 1	1275	1086	
0507_1050	1038 184	1222	1038	
0507_1100	903 235 24	1162	903	
0507_1110	1237 44 1 16	1298	1237	
0507_1120	1350 19 29 66	1464	1350	
0507_1130	1376 568 10 31 18	2003	1376	
0507_1140	1377 749 21 17 8	2172	1377	
0507_1150	1666 18 12 10	1706	1666	1666
0507_1200	1637 6	1643	1637	
0507_1210	452 929	1381	929	
0507_1220	421 188 55 20 653	1337	653	
...		...		
0507_1710	491	491		
0507_1720	470	470		
0507_1730	505	505		
0507_1740	556	556		
0507_1750	569	569		
0507_1800	105	105		
0507_1810	99	99		
0507_1820		0		
0507_1830		0		

Соберем в один рисунок несколько характерных полей наблюдений, содержащих *пики*, и соответствующие прогнозы на эти сроки со стартом на час раньше (рис. 4).

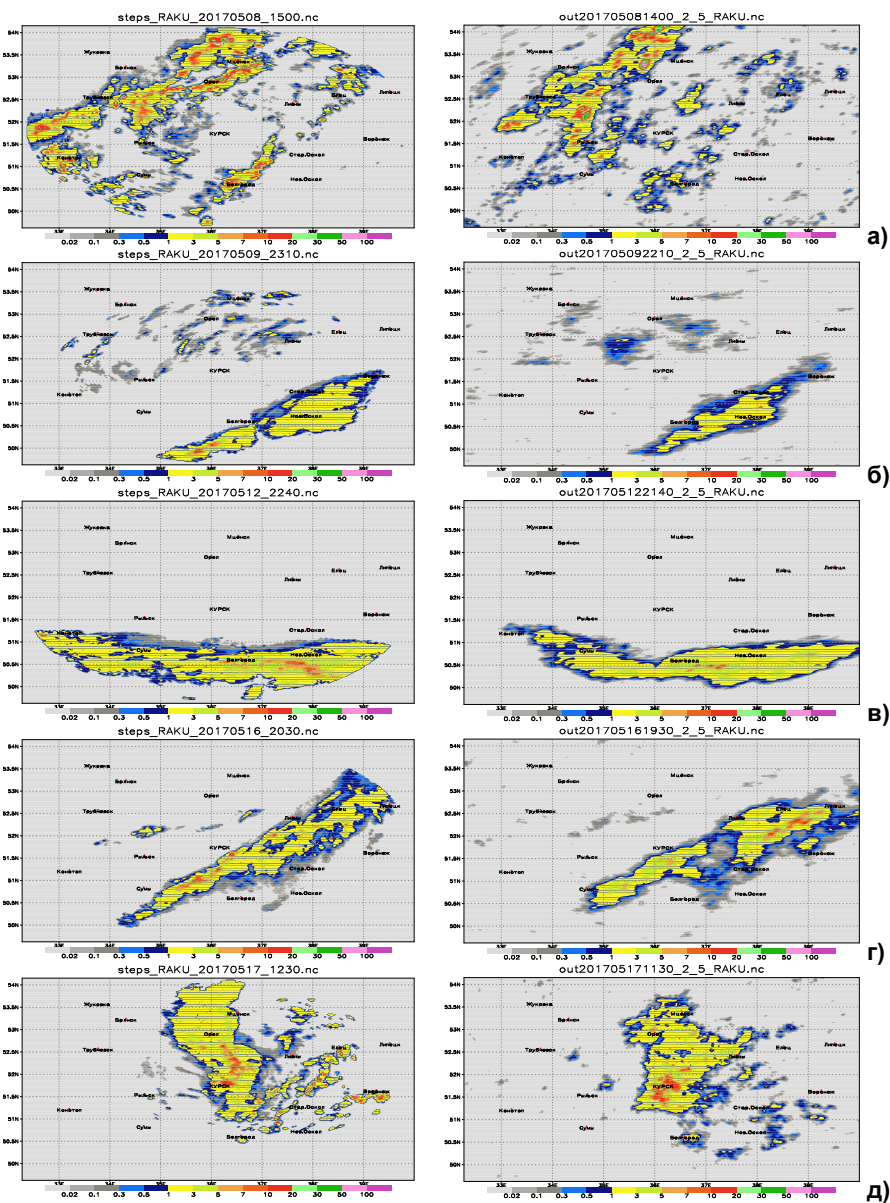


Рис. 4. Левый столбец панелей: поля радиолокационных осадков, полученных по данным наблюденийлокатора Курск и содержащих объекты *пиковой* площади, выделенные на рис. 3. синими линиями. Площади объектов (в количестве точек): 8516 (а); 3747 (б); 7895 (в); 5337 (г); 8098 (д). Правый столбец панелей: прогноз на эти сроки наблюдений, выпущенный часом раньше.

Fig. 4. Left column: the Kursk radar precipitation fields containing the objects with *peak* areas indicated with the blue lines in fig. 3. Areas of objects in grid points: 8516 (a); 3747 (б); 7895 (в); 5337 (г); 8098 (д). Right column: forecast for this time issued one hour earlier.

Необходимо иметь в виду, что поле наблюдения состоит из фактических значений интенсивности осадков только внутри вписанного круга, границы которого заметно проявляются при наличии осадков (например, левые панели а, б, в, г); в то же время "пустоты" в полях прогноза могут заполняться прогностическими данными (правые панели а, г). Добавим, что на рис. 3 фактические поля наблюдений представлены в разрешении 1 км, а поля соответствующих прогнозов – в разрешении 2 км.

В табл. 2 и 3 собраны суммарные характеристики всех выделенных ситуаций с учетом количества и размеров объектов в полях обзора локатора РАКУ (Курск) и в соответствующих прогностических полях за пять месяцев теплого периода 2017 г. Табл. 2 и 3 отличаются лишь минимальным размером идентифицируемых объектов – не менее 625 и не менее 1600 точек поля, соответственно. *Пики* ситуаций – это значения площади объектов в столбце *max*.

Сделаем несколько замечаний. *Первое замечание* относится к методу выделения ситуаций при разных порогах. Сравним первые строчки первых частей обеих таблиц, относящихся к радарным наблюдениям: ситуация длительностью 550 минут (9 часов 10 минут) содержит 19 объектов размером не менее 625 точек и всего 2 объекта размером не менее 1600 точек. Так как несколько объектов могут «населять» одно поле, то их формально просуммированное по каждому полю количество (в столбце *objs*) может существенно превышать количество полей в ситуации. Однако количество синоптических объектов мезомасштабных размеров, изменяющихся в пространстве и во времени на протяжении одной *ситуации*, как правило, на порядки меньше общего количества *ситуаций* (см. по этому поводу обсуждение *эволюции объектов во времени* в статье [32]). А ввиду вложенности ситуаций друг в друга "по образцу матрёшки" (см. табл. 1), объекты размером не меньше 1600 точек находятся в выборке объектов размером не менее 625 точек – в набор *пиковых* значений в обоих случаях попадает одно и то же значение 1666. Дополнительно отметим, что во втором случае – при двух выделенных объектах – все кварталы оказываются различными из-за линейной интерполяции, автоматически проводимой на не менее двух значениях.

Второе замечание касается количества выделенных ситуаций в прогнозах в сравнении с наблюдениями. Автоматическое выделение ситуаций нуждается иногда в корректировке из-за сбоев в технологии прогнозирования. Так, в период первого испытания технологии наукастинга в 2017 г. потери в прогнозах составляли до 10–12 % от максимально возможного количества в 22032 10-минутных интервалов [21, табл. 1]. Такие потери означают пробелы в непрерывной последовательности полей, приводящие иногда к короткоживущим ситуациям длительности от 10 до 30 мин. С учетом таких потерей можно сделать вывод о примерно 20%-м росте количества выделенных ситуаций в прогностических полях при прогнозе на 60 мин для обоих порогов 625 и 1600 точек.

Таблица 2. Характеристики выделенных «ситуаций» в ряду наблюдений и наукастинга по данным радиолокатора Курск в теплый период года (май-сентябрь 2017 г.). Размер идентифицируемого объекта – не менее **625 точек поля**
Table 2. Characteristics of indentified situations in observations and nowcasting data series from the Kursk radar during the warm period (May-September 2017). Object size not less than **625 grid points**

sit	min	q25	med	q75	max	objs	drt	d_time_start	d_time_stop
Радиолокатор КУРСК									
1	653	754	912	1294	1666	19	550	2017-05-07 09:10	2017-05-07 18:10
2	666	714	907	924	1022	5	150	2017-05-08 00:50	2017-05-08 03:10
3	893	1529	1828	2770	3292	11	140	2017-05-08 08:30	2017-05-08 10:40
4	678	2112	2895	5790	8516	67	750	2017-05-08 12:30	2017-05-09 00:50
5	647	1412	1720	2460	3747	56	720	2017-05-09 20:50	2017-05-10 08:40
6	645	1694	2910	3886	7087	141	1470	2017-05-11 14:00	2017-05-12 14:20
7	634	1202	2444	4677	7895	111	1400	2017-05-12 15:10	2017-05-13 14:20
8	628	1184	1676	2110	2600	49	720	2017-05-13 16:10	2017-05-14 04:00
9	629	1210	1753	3032	5337	108	1130	2017-05-16 08:20	2017-05-17 03:00
10	628	954	1584	2328	8098	386	3970	2017-05-17 04:40	2017-05-19 22:40
.....									
83	638	768	899	985	1071	3	300	2017-09-18 10:20	2017-09-18 15:10
84	629	676	704	747	1040	10	270	2017-09-18 17:20	2017-09-18 21:40
85	684	1396	2974	5066	7720	140	1670	2017-09-20 00:00	2017-09-21 03:40
86	627	855	1090	1188	1339	15	420	2017-09-24 05:40	2017-09-24 12:30

sit	min	q25	med	q75	max	objs	drt	d_time_start	d_time_stop
Прогностическая система STEPS									
1	680	1146	1426	1757	2296	30	760	2017-05-07 09:20	2017-05-07 21:50
2	685	753	821	900	978	3	300	2017-05-07 22:30	2017-05-08 03:20
3	634	1142	1914	2284	3187	8	130	2017-05-08 08:40	2017-05-08 10:40
4	655	2056	4079	6048	9031	67	750	2017-05-08 13:00	2017-05-09 01:20
5	646	662	672	705	794	4	340	2017-05-09 14:50	2017-05-09 20:20
6	1113	1732	2468	3153	4380	14	180	2017-05-09 21:10	2017-05-10 00:00
7	684	1536	1964	2556	3580	42	620	2017-05-10 00:30	2017-05-10 10:40
8	708	722	767	911	918	5	200	2017-05-10 15:00	2017-05-10 18:10
9	628	1418	2560	3814	7514	140	1480	2017-05-11 14:20	2017-05-12 14:50
10	659	2407	3780	6488	8523	45	490	2017-05-12 15:50	2017-05-12 23:50
.....									
115	628	1341	2537	4740	6824	119	1550	2017-09-19 22:10	2017-09-20 23:50
116	654	1399	1635	2763	3667	17	290	2017-09-21 00:20	2017-09-21 05:00
117	637	638	640	664	688	3	210	2017-09-23 03:20	2017-09-23 06:40
118	756	913	977	1131	1598	14	420	2017-09-24 06:00	2017-09-24 12:50

Примечание. Интервал сцепления равен 20 мин. Минимальный размер связанного объекта составляет 625 точек поля. Экстремумы и оценки квартилей размеров всех объектов ситуации (в количестве внутренних точек) размещены в столбцах min, q25 (25%-й квартиль), med (медиана), q75(75%-й квартиль) и max. В столбцах objs и drt находятся количество пиков (объектов размера не меньше 625 точек в данной ситуации) и длительность ситуации (в минутах) соответственно. Два последних столбца содержат начальную и конечную дату и время ситуации (включительно).

Таблица 3. Характеристики выделенных «ситуаций» в ряду наблюдений и наукастинга по данным радиолокатора Курск в теплый период года (май-сентябрь 2017 г.). Размер идентифицируемого объекта – не менее **1600 точек поля**
Table 3. Characteristics of identified situations in observations and nowcasting data series from the Kursk radar during the warm period (May-September 2017). Object size not less than **1600 grid points**

sit	min	q25	med	q75	max	objs	drt	d_time_start	d_time_stop
Радиолокатор КУРСК									
1	1637	1644	1652	1659	1666	2	550	2017-05-07 09:10	2017-05-07 18:10
2	1623	1825	2432	2997	3292	8	140	2017-05-08 08:30	2017-05-08 10:40
3	1635	2176	3026	6257	8516	62	750	2017-05-08 12:30	2017-05-09 00:50
4	1651	1886	2415	2656	3747	33	720	2017-05-09 20:50	2017-05-10 08:40
5	1609	2182	2977	4256	7087	111	1470	2017-05-11 14:00	2017-05-12 14:20
6	1644	2316	3336	6098	7895	78	1400	2017-05-12 15:10	2017-05-13 14:20
7	1608	1789	2088	2256	2600	28	720	2017-05-13 16:10	2017-05-14 04:00
8	1600	1805	2325	3535	5337	70	1130	2017-05-16 08:20	2017-05-17 03:00
9	1605	1863	2362	3580	8098	190	3970	2017-05-17 04:40	2017-05-19 22:40
10	1607	1749	2170	2697	3815	71	790	2017-05-21 09:30	2017-05-21 22:30
.....									
42	1621	1801	2056	2324	4489	28	830	2017-09-05 07:50	2017-09-05 21:30
43	1602	1638	1658	1770	1844	8	590	2017-09-10 00:50	2017-09-10 10:30
44	1603	1603	1603	1603	1603	1	300	2017-09-13 11:10	2017-09-13 16:00
45	1602	2648	3572	5756	7720	103	1670	2017-09-20 00:00	2017-09-21 03:40

sit	min	q25	med	q75	max	objs	drt	d_time_start	d_time_stop
Прогностическая система STEPS									
1	1643	1713	1904	1981	2296	12	760	2017-05-07 09:20	2017-05-07 21:50
2	1793	2034	2249	2391	3187	5	130	2017-05-08 08:40	2017-05-08 10:40
3	1601	2396	4501	6270	9031	59	750	2017-05-08 13:00	2017-05-09 01:20
4	1673	2188	2947	3444	4380	11	180	2017-05-09 21:10	2017-05-10 00:00
5	1616	2019	2291	2860	3580	28	620	2017-05-10 00:30	2017-05-10 10:40
6	1629	2448	3222	4135	7514	98	1480	2017-05-11 14:20	2017-05-12 14:50
7	1823	3324	5638	6702	8523	35	490	2017-05-12 15:50	2017-05-12 23:50
8	1620	1952	2492	4519	6871	42	920	2017-05-13 00:20	2017-05-13 15:30
9	1604	1906	2165	2482	3264	26	950	2017-05-13 16:40	2017-05-14 08:20
10	1631	2044	2704	4121	5841	54	930	2017-05-16 08:30	2017-05-16 23:50
.....
61	1604	1849	2124	2463	3364	30	730	2017-09-05 08:20	2017-09-05 20:20
62	1642	1719	1838	1970	2215	10	600	2017-09-10 01:50	2017-09-10 11:40
63	1605	2684	4066	5366	6824	78	1550	2017-09-19 22:10	2017-09-20 23:50
64	1635	2058	2763	2949	3667	9	290	2017-09-21 00:20	2017-09-21 05:00

Примечание. Интервал сцепления равен 20 мин. Минимальный размер связанного объекта составляет 1600 точек поля. Экстремумы и оценки квартилей размеров всех объектов ситуации (в количестве внутренних точек) размещены в столбцах min, q25 (25%-й квартиль), med (медиана), q75(75%-й квартиль) и max. В столбцах objs и drt находятся количество пиков (объектов размера не меньше 1600 точек в данной ситуации) и длительность ситуации (в минутах) соответственно. Два последних столбца содержат начальную и конечную дату и время ситуации (включительно).

Увеличение количества ситуаций с ростом заблаговременности в обновляющихся каждые 10 минут прогнозах происходит чаще всего дроблением одной ситуации из-за таких взаимосвязанных эффектов, как статистическое сглаживание, снижение количества точек поля с высокой интенсивностью, расширение областей сплошных осадков. При этом расширение областей может приводить к порождению «новых» ситуаций в последовательностях прогностических полей из-за ригористического исключения точки из объекта по неравенству "строго меньше".

Диапазон характеристик выделенных объектов в наблюдениях и прогнозах в теплый период на основе локатора Курск дает дополнительное представление о свойствах прогностической системы. Представим некоторые характеристики ситуаций, немного отредактировав числа ввиду возможных сбоев или ошибок, для порога размеров 625 точек в ситуациях в полях наблюдений и в ситуациях при прогнозах на 30 и 60 минут (табл. 4).

Основной рост количества ситуаций происходит на шаге прогноза до 30 минут (скачок примерно на треть), с ростом заблаговременности рост сохраняется, но не столь сильный, и можно предположить, что сериальная корреляция внутри ситуаций не нарушает независимости пиков соседних ситуаций.

Таблица 4. Диапазоны изменений минимальных (min) и максимальных (max) размеров объектов в выделенных ситуациях. Минимальный размер связанного объекта – 625 точек поля

Table 4. Range of minimal (min) and maximal (max) object sizes in the identified situations, Minimal object size is 625 grid points

Поля	Колич. ситуаций	Характеристики выделенных ситуаций и диапазоны их изменений							
		min		max		objs		drt	
		мин	макс	мин	макс	мин	макс	мин	макс
RAKU	86	625	893	625	17989	1	386	90	3970
STEPS-30	117	625	1242	637	18282	1	183	10	1930
STEPS-60	118	625	1113	629	17601	1	185	10	2350

Примечание. objs – количество объектов в полях; drt – продолжительность ситуации в минутах. Интервал сцепления равен 20 мин.

В табл. 5 и табл. 6 представлено количество выделенных ситуаций с параметром сцепления 20 мин по областям обзора всех локаторов (кроме р/л Внуково) для заблаговременностей 30, 60, 90 и 120 минут и для двух пороговых площадей объектов – 625 и 1600 точек.

Таблица 5. Количество выделенных ситуаций в последовательностях полей наблюдений радиолокаторов и прогнозов разной заблаговременности с помощью системы STEPS в теплый период 2017 г. (май-сентябрь). Минимальный размер выделяемого связного «объекта» в поле равен 625 точкам. Параметр сцепления – 20 минут

Table 5. Number of the situations in consecutive fields of radar observations and STEPS nowcasts for different lead times in the warm period (May-September). The minimal object size is 625 grid points. Tie interval is 20 min – two consecutive fields missing in a sequence are allowed

Идентификатор р/л	Наблюдения р/л	Прогностические поля: заблаговременность (мин)			
		30	60	90	120
RAKU	86	117	118	121	123
RATL	79	117	117	128	132
RAVO	89	133	141	145	157
RUDB	76	90	96	99	99
RUDK	96	130	142	145	155
RUDL	92	119	123	131	130
RUDN	84	122	129	141	147
RUWJ	85	119	128	138	152

Таблица 6. То же, что и в табл. 5, кроме минимального размера выделяемого связного «объекта» в поле: здесь он равен 1600 точкам

Table 6. As in Table 5, but for the minimal object size of 1600 grid points

Идентификатор р/л	Наблюдения р/л	Прогностические поля: заблаговременность (мин)			
		30	60	90	120
RAKU	45	65	64	64	67
RATL	56	71	73	76	76
RAVO	64	81	83	85	95
RUDB	51	57	57	59	60
RUDK	65	79	88	94	96
RUDL	60	79	83	85	92
RUDN	60	87	89	93	95
RUWJ	85	119	128	138	152

Сопоставим основные характеристики ситуаций и объектов с гистограммами распределения всех размеров объектов, выделенных порогом 1 мм/ч. Ограничимся гистограммами для данных лоатора Курск (рис. 5), ввиду того, что для остальных лоаторов картина сходная.

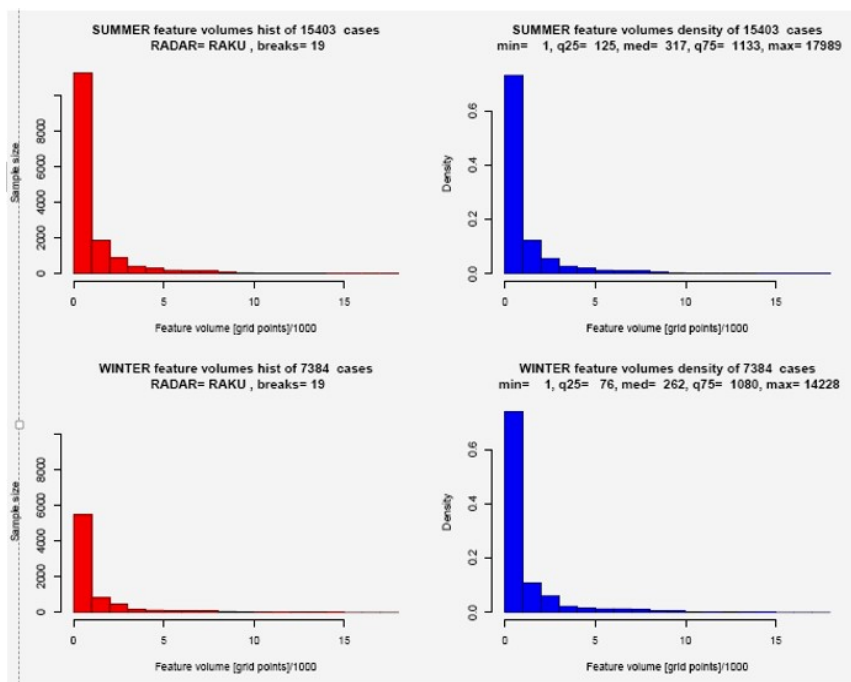


Рис. 5. Гистограммы распределения размеров всех объектов, выделенных порогом 1 мм/ч в полях интенсивности осадков радиолокатора Курск. Верхний ряд панелей – данные по теплому периоду, нижний – по холодному. Красный цвет – количество, синий – частоты. В названии панелей указаны количество связанных объектов (cases), количество градаций гистограммы (breaks + 1, в левом столбце) и основные характеристики распределения размеров (квартили и экстремальные значения). Для удобства графического представления размер объекта на оси абсцисс поделен на 1000.

Fig. 5. Histograms of the distribution of sizes of all objects identified using the threshold of 1 mm/h in the Kursk radar fields. The upper row is for the warm period, the lower row is for the cold period. Red indicates the number, blue, the frequencies. The number of contiguous objects (cases), histograms categories (breaks + 1, left column), and the basic distribution characteristics (quartiles and extremes) are given above the panels. The object size in abscissa axis is divided by 1000 for convenience.

Количество всех выделенных порогом 1 мм/ч объектов в теплый период примерно вдвое превышает аналогичное количество в зимний период (15403 против 7384, рис. 5, левый столбец), однако характер плотности распределения в основном одинаков (рис. 5, правый столбец). Сопоставим общие статистические характеристики распределения размеров объектов по данным теплого и холодного периода: медиана 317 против 262, 75%-й квартиль 1133 против 1080 и максимальное значение 17989 против 14228. Переход к линейным масштабам объектов (извлечением квадратного корня) позволяет сопоставить полученные результаты с данными [31].

Распределение указанных *пиковых* площадей моделируется в дальнейшем обобщенным распределением Парето с помощью пакета extRemes языка R [40].

3.2. Пуассоновость распределения пиков на временной шкале

Известно, что обобщенное распределение Парето возникает как предельное распределение "пиков над порогом", но не всегда указывается, что выборка *пиков* предполагается взятой из обобщенного распределения экстремальных величин. Более того, если пики расположены на временной оси, как на примере рис. 3, то для выполнения условий первой теоремы об экстремумах *достаточно* соблюдения пуассоновости их распределения. Это предполагает, что отсчеты времени с пиками распределены как *точечный* (или *простейший*) *пуассоновский поток событий* [15, 17, 18, 44, 55], определяемый тремя условиями:

1) *стационарность* – вероятность появления k событий в любом промежутке времени длины s зависит только от k и s и не зависит от начала отсчета данного промежутка;

2) *отсутствие последовательности* – появление k событий в промежутке времени не зависит от того, появлялись ли эти события ранее;

3) *ординарность* – появление не менее двух событий за малый промежуток времени "практически" невозможно.

Всем этим требованиям удовлетворяет следующее аналитическое определение потока Пуассона. Обозначим через λ среднее число событий в единицу времени (*интенсивность потока*). Если вероятность появления k событий за время t описывается формулой

$$P_i(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t},$$

то условия 1 – 3 автоматически выполняются.

Пуассоновское распределение нетрудно смоделировать с помощью серий независимых псевдослучайных чисел, воспользовавшись обратной функцией показательного распределения для сумм случайных величин (например, [16, стр. 60]). С помощью пакета poisson (из архива R) соответствие потоку Пуассона указанных на рис. 3 сроков было проверено на ансамбле 100 реализаций с последующими оценками средних, медианных, наилучших и наихудших последовательностей точек на оси времени. Интенсивность потока рассчитывается по всему периоду верификации, состоящему округленно из 22000 случаев. Нумерацию сроков увеличим примерно на 200 (предыдущие сутки не имели пиков) и положим длину интервала рис. 3 округленно равной 2000. В архиве наблюдений локатора Курск в теплый период 2017 г. выделено 86 пиков (или *ситуаций*), т. е. интенсивность потока равна $\lambda = 0.0039$, а значение $\lambda t = 7.802$ для интервала рис. 3 дает ожидаемые около восьми пиков. Рис. 6 содержит все реализации точечного процесса Пуассона с параметрами от 1 до 10 для целых

значений распределения Пуассона. На веер построенного ансамбля нанесены штриховой линией квантили 2.5 % и 97.5 % и желтой линией – ожидаемое среднее значение.

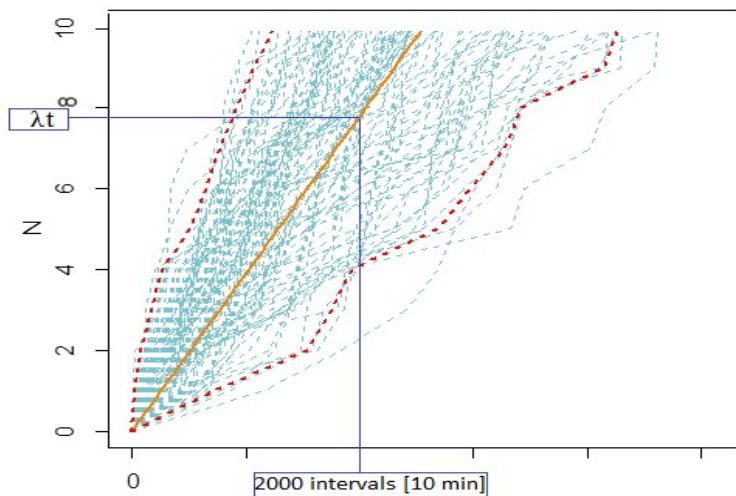


Рис. 6. Ансамбль 100 реализаций потока Пуассона для моментов времени, в которые наблюдались пиковые площади выделенных объектов. Синими линиями отмечены: на оси времени условная длина интервала рис. 3 (2000 сроков), на оси количества пиков – параметр распределения Пуассона, соответствующий ожидаемому количеству пиков на данном интервале. Желтая диагональ соответствует ожидаемому количеству пиков в зависимости от длины интервала. Красные штриховые линии выделяют квантили 2.5 % и 97.5 % номеров соответствующих отсчетов времени; например, верхняя линия квантилей располагает до 8 пиков в первой тысяче сроков, в то время как нижняя линия располагает во всем интервале рис. 3 не более 4 пиков.

Fig. 6. Ensemble of 100 Poisson flow realizations for the time moments of peak object areas. Blue lines on X-axis indicate conditional interval duration of Fig. 3 (2000 times), the axis of peak numbers shows the Poisson distribution parameter corresponding to the expected peak number on this interval. The yellow diagonal corresponds to the expected peak number depending on the interval length. The red dotted lines indicate quantiles 2.5% and 97.5% of the numbers of corresponding time counts; e.g., the upper line of quantiles places up to 8 peaks in the first thousand of times, while the lower line places not more than 4 peaks in the whole interval of fig. 3.

Простейшие характеристики данного ансамбля расположены в табл. 7. Видно, что номера сроков пиковых площадей по рис. 6. входят во внутреннюю часть веера реализаций; наименьшее среднее абсолютное расстояние от номеров отсчетов 100 реализаций указывает на ближайшую реализацию под номером 40. Однако и наблюдаемые сроки, и сроки ближайшей модельной реализации занижены по отношению к медианной и средней реализациям, что, возможно, объясняется наличием мнимых и не устраненных пиков на рис. 3, а также случайным характером самого

процесса генерирования пуассоновского потока. Приведенный пример проверки пуассоновости процесса может быть легко обобщен на полные интервалы испытаний.

Таблица 7. Номера сроков наблюдений (*xobs*) и квантилей распределения ансамбля точек отсчета модельного потока Пуассона на интервале до 2000 начальных сроков. Интенсивность потока Пуассона рассчитана для 86 пиковых площадей, не меньших 625 точек. Статистические характеристики модельного ансамбля - *min*, *med*, *mean*, *max*; ближайшая к наблюдениям реализация сроков - *best*

Table 7. Numbers of the times (*xobs*) and of the quantiles of ensemble distribution of the starting points of the modeled Poisson flow in the interval of up to 2000 initial times. The Poisson flow intensity is computed for 86 peak areas not less than 625 grid points. The statistical characteristics of the model ensemble are *min*, *med*, *mean*, *max*; *best* is the member closest to the observations

	Пуассоновская реализация десяти последовательных сроков наблюдений в интервале 1–2000									
	1	2	3	4	5	6	7	8	9	10
xobs	217	305	352	380	573	898	1002	1149	1565	1661
best	247	281	379	491	708	781	932	1063	1330	1680
min	5	16	116	202	315	340	485	802	918	1071
med	180	412	656	871	1108	1292	1577	1942	2184	2411
mean	244	476	744	969	1261	1493	1778	2047	2291	2538
max	1161	1825	2419	2691	3328	3454	4024	4148	4597	4680

4. Выделение полей с объектами значительной площади для оценок *FSS*

Построение выборок для показателя *FSS* намного проще аналогичной процедуры для приложения теории экстремумов. Основной принцип состоит в подборе полей с наличием объектов, площадь которых не меньше 1600 точек, что, напомним, эквивалентно квадрату со стороной около 80 км. Выбор наибольшей пороговой площади из протестированных был обусловлен желанием оценить качество модели в прогнозировании синоптических ситуаций, близких к экстремальным.

Для идентификации объектов в обоих периодах испытаний оставлен один порог (1 мм/ч). Локатор Внуково был исключен из-за незначительного количества выделенных ситуаций в холодный период. Количество накопленных полей с объектами площади не менее 1600 точек указано в табл. 8. Так как выборки составлялись из полей, принадлежащих зачастую одной ситуации, то оказался нарушенным принцип статистической независимости элементов выборки. Этот принцип был отчасти соблюден случайными отборами полей в количестве, не превышающем 200, что имеет порядок количества выделенных *ситуаций* в теплый период года.

Таблица 8. Общее количество прогностических полей (сроков) с наличием объектов, выделенных после пространственного сглаживания с радиусом 9 узлов с помощью изолинии 1 мм/ч и имеющих размер не менее 1600 точек. Теплый период май-сентябрь 2017 г., максимальное количество сроков – 22032; холодный период ноябрь 2017 г. – март 2018 г., максимальное количество сроков – 21744

Table 8. Total number of forecast fields (times) with the objects identified after spatial smoothing with convolution radius of 9 grid points using the isoline of 1 mm/h and with sizes of not less than 1600 grid points. Warm period is May-September 2017, the maximal number of times is 22032; cold period is November 2017 – March 2018, maximal number of times is 21744

ДМРЛ	Наблюдение		Прогноз (мин)							
			30		60		90		120	
	тепл	холод	тепл	холод	тепл	холод	тепл	холод	тепл	холод
RAKU	2843	1387	2550	1495	2561	1527	2553	1550	2561	1564
RATL	2378	847	2045	881	2045	908	2046	916	2018	931
RAVN	2147	130	*	*	*	*	*	*	*	*
RAVO	2777	506	2688	619	2728	655	2744	666	2748	687
RUDB	2246	721	1950	822	1936	858	1918	885	1932	884
RUDK	3432	379	3287	369	3266	402	3268	428	3237	445
RUDL	3240	691	3036	798	3036	744	3052	750	3088	753
RUDN	3520	626	3431	671	3456	694	3461	732	3427	767
RUWJ	3755	201	3533	235	3544	254	3554	266	3524	277

Из-за сбоя в информационных потоках в теплый период количество прогнозов оказалось примерно на 10 % меньше количества наблюдений [21], поэтому для корректного сопоставления следует вычесть из наблюдений эти 10 %: например, 2843 заменить на $2559 = 2843 - 284$ и т.д.

По столбцам "наблюдение" отчетливо видно, насколько разнообразно количество (f_c) отбираемых полей для различных локаторов в оба периода: так, для теплого периода минимальное f_c равно 2021, а максимальное – 3379; для холодного периода эти числа равны, соответственно, 201 и 1387 (исключая р/л Внуково). Естественно наследование такого различия в прогнозах. Выявленное разнообразие значений f_c затрудняет обобщения по всем локаторам. Например, осреднение этих величин дает превышение значений f_c в теплый период над аналогичными значениями в холодный период примерно в 4.5 раза, в то время как по локаторам это отношение колеблется от 2.1 до 9.1 в семи случаях и 18.7 – в одном (Валдай, RUWJ). При этом локатор Валдай обеспечивает максимальное значение f_c по теплому периоду и минимальное значение в холодный период и в наблюдениях, и во всех прогностических сроках. Точно так же локатор Брянск (RUDB) обеспечивает минимальное значение f_c в теплый период,

а локатор Курск (RAKU) обеспечивает максимальное значение f_c в холодный период.

При переходе от наблюдений к прогнозу 30 мин значение f_c в теплый период меняется примерно на 100–200 полей: возрастает для пяти локаторов, а падает – для трех локаторов; в холодный период меняется примерно на 70–100 полей, при этом возрастая для семи и падая для одного локатора.

5. Учет пропусков в узлах полей и формирование выборок разных типов

В усовершенствованной системе наукастинга осадков, построенной на новом расчетном ядре ruSTEPS, используются 1) расширенный набор форматов данных, 2) блок быстрого двумерного преобразования Фурье без ограничений на размер расчетного поля, 3) новая технология генерирования ансамбля и 4) новая методика обработки пропусков. Основные результаты испытаний системы в теплый период 2020 г. представлены в [22, 23]. Архив испытаний состоял из около 22000 ансамблевых прогнозов на 150 минут в 10-минутной дискретности, при этом каждый из 15 прогностических сроков был представлен ансамблем из 10 полей.

В отличие от предыдущих испытаний (2017–2019 гг.), в которых пропуски (*пустые* точки) замещались малыми положительными числами и проблема отсутствующих данных в принципе не возникала, необходимость обработки пропусков в 2020 г. привела к соответствующему изменению и дополнению стратегии верификации прогнозов с учетом сложившейся методологии обработки пропусков [19, 45, 49, 50]. Для оценки временного и пространственного распределения *пустых* точек проведен анализ полного архива наблюдений и полного архива ансамблевых прогнозов для одной заблаговременности (60 мин). Результаты анализа качества наукастинга, проведенного с помощью средств языка R, представлены в таблицах и графиках. Пространственный анализ производился с помощью квантильного анализа и соответствующих пространственных квантильных функций.

5.1. Статистика пропусков в архивах наблюдений и прогнозов

Пропуски в полях радиолокационных наблюдений. Наиболее удобный прием работы с пропусками заключается в формировании матриц (полей) масок, состоящих из нулей (для "пропусков") и единиц (для "допусков"), соответствующих отсутствию и наличию выбранных числовых данных в точках поля. Сумма таких матриц по всем срокам дает количество допусков в каждой точке, а поточечное произведение матриц эквивалентно пересечению всех матриц по единицам в точках поля, в которых количеством выбранных числовых данных за весь период верификации было максимально возможным.

Матрицу, полученную суммированием всех масок и которую назовем *суммарной маской*, проанализируем с помощью квантилей. Функция `quantile` языка R имеет на входе список порогов вероятностей, а на выходе набор квантилей, рассчитанных с помощью линейной интерполяции эмпирической функции распределения по методу Хайндмана [43]. В частности, последовательность из 11 порогов 0.0, 0.1, ..., 1.0 выводит такое же количество квантилей, или десятков процентов, от минимального значения (порогу 0.0 соответствует 0 %) до максимального (порогу 1.0 соответствует 100 %). Матрица, полученная *произведением* всех масок, состоит из единиц и нулей, при этом, если в точке поля имеется хотя бы одна дата с пропуском, результат будет равен нулю. Сумма единиц в такой матрице эквивалентна оценке площади (или размера) области без пропусков за весь период верификации.

Рассмотрим вначале *суммарную маску*: наибольшее значение в точке равно 22026, минимально возможное значение равно единице (этот факт, сам по себе довольно любопытный, тоже встречается). При наличии в полях в среднем более 70 % точек с числовыми значениями в пропорции к полному размеру, существуют поля с заметно меньшим количеством "непустых" точек (не более 50 %). Наибольшее количество точек поля с ненулевыми значениями равно 3193798, т. е. примерно в 76 % точек объединенного поля хотя бы раз встречается числовое значение.

Выделим следующие особенности. Во-первых, более 95 % значений точек обеспечено не менее 18372 наблюдениями, а это примерно 95 % от 3193798 "непустых" точек. Одновременно с этим среди 5 % точек могут оказаться точки с чрезвычайно скудным обеспечением: менее 1000 числовых значений было обнаружено во временных рядах 2716 точек поля, менее 10 значений – в 1233 точках и ровно одно числовое значение – в 242 точках. Непросто определить, какое количество точек поля с "непустыми" значениями считать несущественным, и это зависит от формы области: связанная область в 2500 точек может занимать квадрат со стороной 50 км, а протяженную линию – в 2500 км.

При шаге вероятности 0.05 квантили распределения значений в точках суммарной маски с ненулевыми значениями собраны в табл. 9.

Простая числовая структура произведения всех масок может дать лишь количество "оставшихся" единиц, которое могло быть также рассчитано по суммарной маске как количество точек с максимальным значением 22036, в данном случае оно равно в точности 721083. Пространственное представление обеих масок приводится на рис. 7 и рис. 8.

Для удобства изображения больших чисел данные на обеих панелях рис. 7 приводятся в долях от максимального количества возможных случаев (22026). Видно четкое проявление кругов обзоров одинакового цвета для одного локатора (одинаковое заполнение числами всех точек поля, кроме теневых областей от местников), при этом наличие перекрывающихся кругов обзора приводит к подавлению теневых областей, кроме,

возможно, не устраненных лучей и конусов р/л Внуково, Нижнего Новгорода, Краснодара (рис. 7а). На карте долей в последнем десятичном интервале (0.90, 0.92, ..., 0.98, 1.00) заметно отсутствие кругов р/л Воейково (0.8–0.9), Вологды (0.5–0.6), Белгорода (0.8–0.9).

Таблица 9. Квантили распределения количества числовых значений в 22026 объединенных полях радиолокационных наблюдений

Table 9. Distribution quantiles of the numerical values in 22026 merged radar fields

0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
1	18347	21000	21468	21812	21855	21884	21911	21936	21986	21994
...продолжение...										
55%	60%	65%	70%	75%	80%	85%	90%	95%	100%	
21998	22013	22018	22023	22023	22026	22026	22026	22026	22026	22026

Примечание. Шаг частоты равен 0.05 (5 %); минимальное значение соответствует квантилю 0 %, что эквивалентно единственному наблюдению в некоторых точках поля; максимальное значение соответствует квантилю 100 %, что эквивалентно наличию в этих точках максимально возможного количества в 22026 наблюдений.

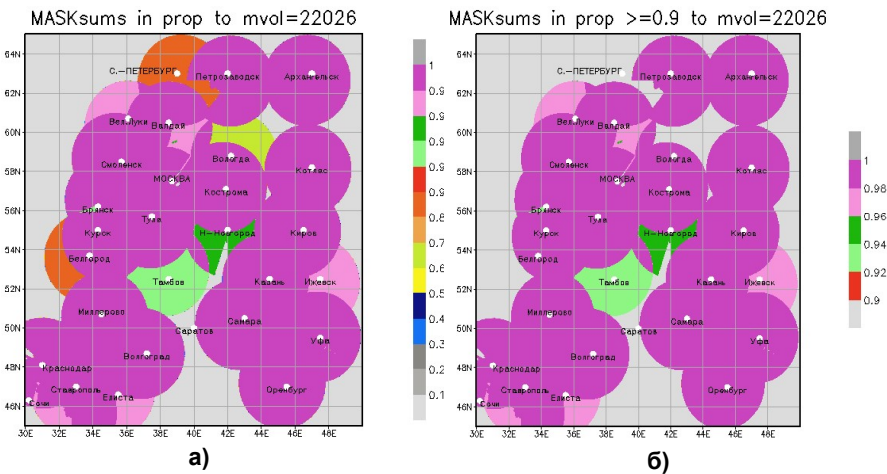


Рис. 7. Карты суммарных масок наличия числовых данных в круговых зонах обзора радиолокаторов усовершенствованной системы наукастинга осадков. Количество представлено в виде долей от максимально возможного значения 22026: весь диапазон значений (от 1 до 22026) (а); значения выше квантиля q(90%) (б).

Fig. 7. The maps of the summary masks of numerical data availability in the radar scanning circle for the improved nowcasting system. The number is represented as the fraction of the maximal possible number of 22026: the whole range from 1 to 22026 (a); values above quantile q(90%) (б).

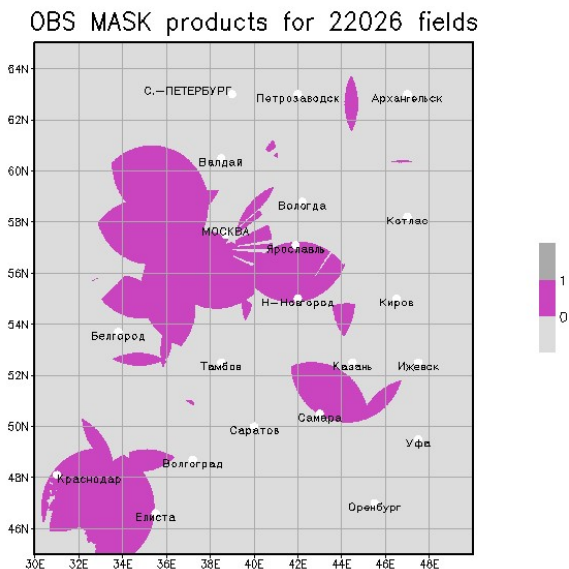


Рис. 8. Поточечное произведение всех масок. В этих точках количество значений максимально (22026). Количество точек с единицами составляет $721083/3193798 = 23\%$ от количества всех "непустых" точек.

Fig. 8. Grid-by-grid product of all the masks. The number is maximal at these points (22026). The number of points with unity (1) is $721083/3193798 = 23\%$ of all "not-empty" points.

Можно сделать вывод, что обеспечение данными по всей области наукастинга вполне удовлетворительное, но возникают вопросы частного и общего характера. Например, можно ли считать обеспечение данными области обзора р/л Вологды неудовлетворительным при том, что доли от 0.5 до 0.6 означают суммарно не менее двух месяцев числовых наблюдений (144 срока в сутки)? И каким образом вообще можно оценить влияние асинхронных доступов в точках поля на пространственную согласованность показателей качества?

На карте произведения масок (рис. 8) отчетливо видно, что основная порция максимально обеспеченных точек приходится на области пересечения зон обзора локаторов в центральной части России (область ЦФО) и в районе локаторов Ставрополь и Сочи.

Пропуски в полях прогностических ансамблей. Имеющиеся в полях наблюдений пропуски естественно наследуются полями прогноза, при этом в основном размножаясь с ростом заблаговременности. Рассмотрим маски в ансамблях наукастов. В табл. 10 показаны сопоставленные ("спаренные") в прогнозе на 60 мин объединенные и нумерованные (от 1 до 21995) пары *полей наблюдений* (префиксы nc-файлов – "PRCP-ETR") и *прогностических ансамблей* (префиксы tiff-файлов – "steps_mvр")

с указанием количества пропусков (первое значение в строке с номером и вторая строка ниже) и количества числовых значений (второе значение в нумерованной строке и третья строка ниже).

Таблица 10. Количество пропусков и числовых значений в сопоставленных в прогнозе на 60 мин парах полей наблюдения (префиксы nc-файлов – PRCP-ETR) и прогностических ансамблей (префиксы tiff-файлов – steps_mvп)

Table 10. Number of missing data and numerical values in the observation fields (prefix PRCP_ETR) and forecast ensembles (prefixes of tiff-files step_mvп) pairs for the lead time of 60 min

1:	PRCP-ETR_20200501_0100.nc	<==	steps_mvп_202005010000_060.tif	1005711	3190890
1018623	1026828	1024980	1028748	1020824	1019132
1017239	1022161	1027674	1031277	1023748.6	18037.6
3177978	3169773	3171621	3167853	3175777	3177469
					3179362
					3174440
					3168927
					3165324
					3172852.4
					-18037.6
2:	PRCP-ETR_20200501_0110.nc	<==	steps_mvп_202005010010_060.tif	1005664	3190937
1020516	1031215	1027429	1033035	1022933	1021168
1018878	1025485	1029435	1036083	1026623.7	20959.7
3176085	3165386	3169172	3163566	3173608	3175433
					3177723
					3171116
					3167166
					3160518
					3169977.3
					-20959.7
1086:	PRCP-ETR_20200508_1350.nc	<==	steps_mvп_202005081250_060.tif	1005710	3190891
1116686	1119302	1128194	1125776	1112011	1113393
1116663	1116089	1136092	1121334	1120054.0	114344.0
3079915	3077239	3068407	3070825	3084590	3083208
					3084938
					3080512
					3060509
					3075267
					3076547.0
					-114344.0
1087:	PRCP-ETR_20200508_1400.nc	<==	steps_mvп_202005081300_060.tif	1005665	3190936
1117280	1120215	1129566	1127492	1112724	1113683
1112304	1116829	1137166	1122168	1120942.7	115277.7
3079321	3076386	3067035	3069109	3083877	3082918
					3084297
					3079772
					3059435
					3074433
					3075658.3
					-115277.7
1125:	PRCP-ETR_20200508_2020.nc	<==	steps_mvп_202005081920_060.tif	1036134	3160467
1034674	1036358	1043463	1044507	1030791	1030797
1029853	1032311	1045163	1037350	1036566.7	432.7
3161927	3160243	3153138	3151694	3165810	3165804
					3166748
					3164290
					3151438
					3159251
					3160034.3
					-432.7
5384:	PRCP-ETR_20200607_1040.nc	<==	steps_mvп_202006070940_060.tif	1005574	3191027
1285115	1284070	1290425	1284711	1283689	1283481
1283730	1283807	1292333	1284227	1286248.8	280674.8
2911486	2912531	2906176	2911890	2912912	2913120
					2912871
					2912794
					2897368
					2912374
					2910352.2
					-280674.8
5385:	PRCP-ETR_20200607_1050.nc	<==	steps_mvп_202006070950_060.tif	1005571	3191030
1285326	1284066	1290495	1284665	1284220	1283421
1283650	1284193	1295663	1284392	1286399.1	280828.1
2911275	2912535	2906106	2911936	2912381	2913180
					2912951
					2912408
					2897038
					2912209
					2910201.9
					-280828.1
5386:	PRCP-ETR_20200607_1100.nc	<==	steps_mvп_202006071000_060.tif	1005571	3191030
1285551	1283838	1291662	1285017	1284242	1283355
1283546	1283892	1300396	1284213	1286571.2	281000.2
2911050	2912763	2904939	2911584	2912359	2913246
					2913055
					2912709
					2896205
					2912388
					2910029.8
					-281000.2

Примечание. Количество пропусков – первое значение в строке с номером (по первой паре: 1005711) и вторая строка ниже (десять значений от 1018623 до 1031277 и среднее по ансамблю – 1023748.6). Количество числовых значений – второе значение в нумерованной строке (3190890) и третья строка ниже (от 3177978 до 3165324 и среднее по ансамблю – 3172852.4). Последний столбец – разности соответствующих значений количества пропусков в прогнозе и в наблюдении: 18037.6 = 1023748.6 – 1005711. Для контроля приводится аналогичная разность для количества числовых значений, которая в сумме с предыдущей разностью ожидается равна нулю.

Отметим самые общие особенности: 1) наблюдается большое разнообразие в количестве пропусков как в полях наблюдений, так и в полях ансамблей, включая поля внутри одного ансамбля, 2) как правило, количество пропусков в ансамблях превышает, иногда существенно, количество пропусков в соответственном поле наблюдения. Первая особенность полностью зависит от еще не проясненных условий формирования объединенных полей на основе полей отдельных локаторов. Вторая

особенность представляется еще более интересной из-за значительного разброса разностей от нескольких сотен (например, 432 в табл. 3) до нескольких сотен тысяч (например, 281000 в той же таблице). Имеются случаи, когда количество пропусков в наблюдениях превышает их среднее или даже максимальное по ансамблю количество пропусков.

Так как при расчете показателей качества *пустая* точка любого из 11 полей (наблюдения и десяти полей ансамбля) исключается из результирующего поля показателя, то более содержательной оценкой различия следует считать разность между максимальным (а не средним) количеством пропусков в ансамбле и количеством пропусков в поле наблюдений. При этом, например, по первой паре вместо 18037 получим $1031277 - 1005711 = 25566$.

Суммарная характеристика табл. 10 отражена в квартильных оценках в табл. 11. Максимальное количество пропусков в полях наблюдений (2098901) почти совпадает с минимальным количеством числовых значений (2097700), однако в среднем (по mean и median) количество пропусков занимает около четверти, а количество числовых значений – около трех четвертей объединенного поля (4196601). Максимальное количество пропусков в прогностических ансамблях изменяется от 144592, т. е. намного меньше минимального количества пропусков в соответственных полях наблюдений (1005542), до 2120232, т. е. больше максимального количества пропусков в полях наблюдений (209890). В более важных для расчета показателей согласованных разностей (строка mxnas-onas) размах оказывается существенно большим: от -1066062 до +1114475.

Таблица 11. Квартильные характеристики и среднее значение (mean) количества пропусков (nas), числовых значений (vld) в полях наблюдений (obs) и прогнозов на 60 мин (frc), максимальное значение в ансамблях (mx), а также разность между максимальным количеством пропусков в ансамбле и количеством пропусков в "спаренном" поле наблюдений (mxnas-onas)

Table 11. Quartile characteristics and the mean of the missing data number (nas), numerical values (vld) in observation fields (obs) and 60 min lead time forecasts (frc), the maximal value in the ensembles (mx), and the difference between the maximum number of missing data in the ensemble and in the merged observation field (mxnas-onas)

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
obs_nas	1005542	1005783	1053784	1067155	1106053	2098901
obs_vld	2097700	3090548	3142817	3129446	3190818	3191059
mx_frc_nas	144592	1055335	1089212	1100554	1139304	2120232
mxnas-onas	-1066062	22095	32654	33399	45431	1114475

Как говорилось выше, анализ пропусков в согласованных полях прогнозов и наблюдений важен для оценки потерь значимых точек в результирующем поле показателя качества. Не менее интересен вопрос

о "путешествии" в прогнозах тех пустых точек начальных полей, которые сохранились в модели STEPS после двумерной Фурье-фильтрации перед входом в прогностический блок. Для этого потребуется сравнивать количество пропусков в серии прогностических полей одной заблаговременности (например, 1 час) и количество пропусков в паре соответствующих начальных полей модели STEPS. Так как архив прогнозов по срокам не имеет разрывов во времени, то можно воспользоваться табл. 9 и сравнить прогностические поля с отстоящими на час раньше полями наблюдений. Статистика разностей между максимальным количеством пропусков в ансамблях и количеством пропусков в двух соответствующих начальных полях (a_1 , a_2) собрана в табл. 12.

Таблица 12. Разности между максимальным количеством пропусков в ансамбле прогнозов на 60 мин (m_{xnas}) и количеством пропусков в соответствующих начальных полях (a_1 и a_2)

Table 12. Differences between the maximal number of missing data in the 60 min forecast ensemble (m_{xnas}) and in the corresponding initial fields (a_1 and a_2)

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
$m_{xnas}-a_1nas$	-1051771	23611	32508	33414	43320	1114585
$m_{xnas}-a_2nas$	-1051936	24036	32437	33410	42695	1074679

Данные табл.12 подобны данным последней строки табл.11. Отметим редкое исключение: количество *одновременных* отрицательных значений относительно первого и второго поля оказывается равным 103 – только в этих случаях (из порядка 20 тысяч прогнозов) максимальное количество пропусков в прогностических ансамблях оказалось *меньше* пропусков в обоих начальных полях.

5.2. Формирование выборок для верификации и подсчет количества допусков

Так как сведения о количестве числовых значений (*допусков*) важны для оценок статистической значимости рассчитываемых показателей, то необходимо учитывать не только и не столько "естественные" потери из-за особенностей наблюдательных систем и прогностической модели, но и изменение количества "допущенных" значений в зависимости от способов формирования этих выборок. Будем обозначать допуски через vld (valid); использованное в верификации среднее по ансамблю прогностическое поле будем обозначать через AVR ; как правило, типы выборок будут помечаться прописными буквами в тексте (E , ET и т. д) и строчными в таблицах и рисунках (e , et и т. д).

Опишем способы формирования выборок для верификации, напомним, что в предыдущих испытаниях системы наукастинга осадков отбирались только те пары значений, которые удовлетворяли условию нестрогого превышения порога 0.05 мм/ч хотя бы одного из значений интенсивности в согласованной паре прогноз-наблюдение, что эквивалентно, по логическому правилу де Моргана, отбраковке пар с одновременным условием строгого непревышения этого порога.

Верификация прогнозов проводилась на парах чисел прогноз-наблюдение (frc, obs), удовлетворявших условиям отбора допусков при предварительном замещении константы отсутствия NA отрицательным числом -999. Допустимость определяется на конъюнкциях (&) и дизъюнкциях (!) неравенств следующим образом:

1. $(frc \geq 0) \& (obs \geq 0)$ – тип "E";
2. $(frc > 0 \& obs \geq 0) | (frc \geq 0 \& obs > 0)$ – тип "ET";
3. $(frc > 0.05 \& obs \geq 0) | (frc \geq 0 \& obs > 0.05)$ – тип "TT";
4. $(frc > 0) \& (obs > 0)$ – тип "T".

Условие (1) отбирает все числовые пары значений, включая нулевые ("без осадков"); при условии (2) хотя бы одно из чисел должно быть положительно ("осадки или в прогнозах, или в наблюдениях", "или" – не исключительное, т. е. "осадки и там, и там" входит в список "допустимой пары"); при условии (3) отбираются значения, не меньшие порога 0.05 мм/ч; условие (4) требует ненулевые значения в обоих числах ("осадки одновременно в прогнозах и в наблюдениях").

Во всех четырех случаях пропуски исключаются автоматически благодаря замещению числом -999. Количество допустимых пар подсчитывается на каждом шаге цикла и в конце цикла используется в операциях осреднения при расчетах средних оценок качества, а соответствующая матрица сумм сохраняется в виде поля исходного размера (nx, ny), $n_x \times n_y = 4196601$.

Рассмотрим матрицы сумм числовых значений с этими четырьмя условиями в квартильном представлении, включая оценку среднего количества (mean) (табл. 13). Прокомментируем данную таблицу. Во-первых, видно, что среднее количество пропусков равно ~1 млн точек, количество числовых значений ~3 200 000 точек, и эти значения мало зависят от типов отбора. Во-вторых, изменение квартилей и средней в зависимости от заблаговременности можно считать незначительным, но заметно изменение квартилей при смене типов отбора допусков: например, для прогноза на 60 мин 75%-й квартиль меняется последовательно 21992 ("e"), 3943 ("et"), 3014 ("tt") и 1942 ("t").

Таблица 13. Распределение в квартилях числовых значений по количеству "допустимых" точек поля с указанием количества пропусков (NAs) и числовых значений (vld) для различных заблаговременностей (lead в мин). Прогноз по среднему полю ансамбля (AVR)

Table 13. Quartile distribution of numerical values of valid points with the numbers of missing data (NAs) and valid numerical data (vld) for different lead times (lead in min). Ensemble average nowcast (AVR)

		Квартили количества числовых значений в точке поля					Пропуски	Допуски
lead	min	q25	med	mean	q75	max	NAs	vld
тип отбора пар числовых значений "e"								
030	1	21766	21942	21412	21998	22000	1004617	3191984
060	1	21733	21930	21188	21992	21995	1004764	3191837
090	1	21556	21920	20893	21984	21990	1004919	3191682
120	1	21288	21860	20529	21981	21987	1005011	3191590
150	1	20848	21818	20095	21976	21984	1005096	3191505
тип отбора пар числовых значений "et"								
030	1	1897	2631	2605	3226	5016	1004696	3191905
060	1	2362	3275	3200	3943	5944	1004832	3191769
090	1	2722	3760	3655	4504	6608	1004965	3191636
120	1	3025	4155	4031	5022	7353	1005075	3191526
150	1	3290	4500	4355	5534	8043	1005133	3191468
тип отбора пар числовых значений "tt"								
030	1	1510	2115	2082	2589	4065	1004748	3191853
060	1	1749	2475	2414	3014	4574	1004939	3191662
090	1	1940	2760	2680	3346	5008	1005064	3191537
120	1	2108	3008	2910	3663	5597	1005140	3191461
150	1	2255	3210	3116	3990	6078	1005162	3191439
тип отбора пар числовых значений "t"								
030	1	1077	1573	1553	1975	3253	1005864	3190737
060	1	1023	1530	1501	1942	3107	1006492	3190109
090	1	972	1483	1449	1906	3109	1006917	3189684
120	1	921	1436	1399	1870	3134	1008019	3188582
150	1	867	1392	1350	1830	3069	1009905	3186696

Столь сильная изменчивость пространственных квартилей объемов выборок является следствием статистического распределения интенсивности осадков; на предыдущих испытаниях исключение пар значений осадков, не превышающих порог 0.05 мм, оставляло для оценок качества в теплый период не более 10–15 % от полного временного ряда. Наконец, имеется на первый взгляд странная особенность (выделенная красным цветом) – рост значений квартилей с увеличением заблаговременности в случае ET и TT, хотя пропуски и допуски ведут себя сходно с остальными случаями. Возможно, это происходит из-за регрессионного типа модели, при котором с ростом заблаговременности растет количество нулевых и около-нулевых значений. Кстати говоря, растущее количество около-нулевых значений используется в циклическом суммировании оценок средних ошибок, а значит, счетчик увеличивается (а не

уменьшается, как в остальных случаях), вследствие чего показатели демонстрируют *мнимый* рост качества с ростом заблаговременности. Эффект здесь аналогичен росту оценки дисперсии подвыборки при удалении из выборки нулевых или около нулевых величин: в формуле расчета числитель почти не меняется, а знаменатель уменьшается.

5.3. Количество допусков для верификации в таблицах и графиках

Рассмотрим объемы выборок для верификации прогнозов по (AVR) на сроки 30, 60, 90, 120 и 150 мин. Предварительно для всех отобранных пар наблюдение-прогноз строится поле маски с нулями в тех точках, в которых для каждой пары значений прогноз-наблюдение хотя бы одно значение является константой отсутствия или не удовлетворяется условие формирования выборки, а с единицами в тех точках, в которых условие формирования выборки удовлетворяется. Константы отсутствия в среднем по ансамблю поля устанавливаются в тех точках, в которых хотя бы одно значение из десяти оказывается нечисловым. Таким образом формируются архивы для всех четырех типов выборок для верификации.

В табл. 14 содержатся значения объемов поточечных выборок, расположенные по квантилям пространственного распределения в долях от количества числовых значений в объединенном поле и в зависимости от заблаговременности прогнозов. Фактически это квантильные функции на распределениях пространственных данных.

Были проанализированы разнообразные особенности таблиц и выявлены детали, которые могут оказаться полезными при интерпретации результатов верификации, но здесь мы ограничимся теми характеристиками данных, которые можно представить в графическом виде и которые трудно проверить по конкретным табличным значениям.

Для наглядного представления поведения объемов выборок построим графики (рис. 9) этих объемов в относительных величинах с нормировкой на какое-либо максимальное значение, например, соответствующее q_{100} и заблаговременности 30 мин. Это значит, что на квантиле 100% красная кривая (прогноз на 30 мин) будет всегда оканчиваться на единице. Для удобства будем начинать графики с нулевой координаты и фокусировать внимание на значениях выше пространственного квантиля 10% (q_{10}).

Рассмотрим четыре панели графика на рис. 9, соответствующие табл. 14. Если объемы выборок падают с ростом заблаговременности, то кривая меньшей заблаговременности лежит не ниже кривой большей заблаговременности (например, на графиках E и T, рис. 9). При этом все кривые графика укладываются в координатный прямоугольник. Если же объемы выборок с ростом заблаговременности растут, то порядок противоположный и кривые могут выходить за пределы координатного прямоугольника.

Таблица 14. Пространственные квантили объемов поточечных выборок для верификации в зависимости от типа формирования выборок и от заглаговременности прогноза

Table 14. Spatial quantiles of grid point samples for verification depending on the type of sample formation and on the forecast lead time

Е					
qnt\frс	030	060	090	120	150
0	1	1	1	1	1
10	20351	20291	18017	17602	14767
20	21694	21336	21276	20847	20167
30	21811	21791	21729	21647	21267
40	21865	21847	21829	21766	21722
50	21942	21930	21920	21860	21818
60	21979	21969	21952	21934	21922
70	21994	21989	21983	21977	21967
80	21999	21994	21987	21984	21980
90	22000	21995	21990	21987	21984
100	22000	21995	21990	21987	21984
Nas	1004617	1004764	1004919	1005011	1005096
vld	3191984	3191837	3191682	3191590	3191505
ЕТ					
qnt\frс	030	060	090	120	150
0	1	1	1	1	1
10	1533	1899	2148	2208	2143
20	1773	2188	2520	2798	2984
30	2050	2549	2925	3248	3541
40	2371	2945	3369	3697	4003
50	2631	3275	3760	4155	4500
60	2861	3531	4065	4528	4935
70	3092	3788	4341	4865	5346
80	3384	4158	4767	5269	5762
90	3770	4616	5326	5967	6527
100	5016	5944	6608	7353	8043
NAs	1004696	1004832	1004965	1005075	1005133
vld	3191905	3191769	3191636	3191526	3191468
ТТ					
qnt\frс	030	060	090	120	150
0	1	1	1	1	1
10	1183	1358	1499	1525	1463
20	1406	1621	1789	1935	2047
30	1632	1899	2103	2280	2448
40	1897	2216	2460	2650	2829
50	2115	2475	2760	3008	3210
60	2287	2657	2973	3276	3556
70	2475	2879	3205	3520	3838
80	2716	3164	3538	3866	4195
90	3023	3539	3976	4373	4744
100	4065	4574	5008	5597	6078
NAs	1004748	1004939	1005064	1005140	1005162
vld	3191853	3191662	3191537	3191461	3191439

Т					
qnt/frc	030	060	090	120	150
0	1	1	1	1	1
10	823	744	645	536	446
20	995	940	889	831	761
30	1173	1120	1062	1003	951
40	1383	1333	1281	1226	1169
50	1573	1530	1483	1436	1392
60	1729	1685	1643	1607	1569
70	1888	1851	1809	1770	1731
80	2076	2043	2012	1981	1946
90	2346	2313	2280	2246	2211
100	3253	3107	3109	3134	3069
NAs	1005864	1006492	1006917	1008019	1009905
vld	3190737	3190109	3189684	3188582	3186696

Примечания. vld – количество числовых значений в объединенном поле; qnt – префикс квантиля в процентах от vld; NAs – количество пропусков в поле: $NAs+vld=4196601=2151 \times 1951$. Типы выборки: E – любые значения осадков; ET – хотя бы одно значение ненулевое; TT – хотя бы одно значение не менее 0.05 мм/ч; T – оба значения ненулевые.

Выделим некоторые заметные свойства построенных кривых. Видно сходство графиков для типов ET (хотя бы одно значение ненулевое) и TT (хотя бы одно значение не меньше 0.05) и их отличие от графика для E (все числовые значения, включая нули) и от графика для T (оба значения положительны). Наибольшая обеспеченность данными (до максимума в ~22000 наблюдений в точке поля) заметно выделяет кривые графика для выборки типа E: быстрый рост до q30 и далее значения объемов визуально практически неотличимые. Идеальным графиком является, очевидно, функция скачка, равная единице на всем интервале между q0 и q100. График E нельзя считать "идеальным", хотя он наилучший в том смысле, что количество точек поля, в которых объемы выборки для верификации прогнозов всех сроков не максимально возможны, не превышает площади порядка 30 %.

Так как графики, отличные от "идеального", содержат кривые разных форм, включая прямолинейные интервалы, то правомерно поставить вопрос о роли угла наклона таких интервалов. Очевидно, неидеальным является горизонтальный вид кривой графика, если его значение по вертикали не равно единице: неизменность какого-либо уровня, не равного единице, означает отсутствие пополнения поля точками доступа с увеличением пространственной области.

Попробуем понять, что означает угол наклона. Ясно, что из-за накопления точек доступа с ростом площади угол наклона всегда неотрицательный. Чем больше угол наклона, тем быстрее происходит наполнение точек поля допусками в этом диапазоне площадей, и чем на меньших

квантилях это происходит, тем лучше. Возможная причина обсуждалась выше: наличие диссипирующего эффекта (*размазывания*) в схемах моделей STEPS и руSTEPS по отношению к областям с заметными осадками.

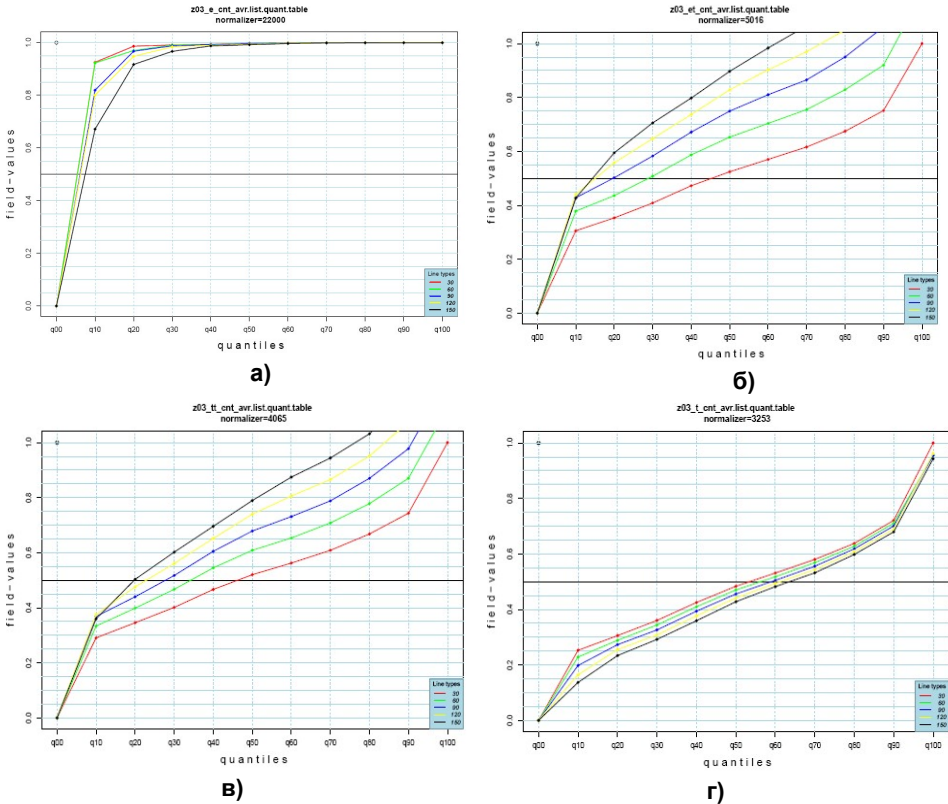


Рис. 9. Графики квантильных функций *относительных объемов выборок* для верификации прогнозов (нормировка на объем в ячейке $frc=30$, $qnt=100$, табл. 14) в зависимости от типа формирования выборок и от заблаговременности прогноза. Расположение типов выборок аналогично табл. 14: тип E (все значения) (а); тип ET (пары значений с хотя бы одной ненулевой интенсивностью) (б); тип TT (хотя бы одно значение не менее 0.05 мм/ч) (в); тип T (оба значения ненулевые) (г). Цветами выделены сроки: 30 мин – красный, 60 мин – зеленый, 90 мин – синий, 120 мин – желтый, 150 мин – черный.

Fig. 9. Plots of quantile functions of relative sample volumes for forecast verification (normalized to the grid cell volume $frc=30$, $qnt=100$, table 14) depending in the type of sample formation and on the forecast lead time. Sample types are the same as in table 14: E stands for all precipitation values; ET, at least one value is not zero; TT at least one value is not less than 0.05 mm/h; and T, both values are not zero. The colors indicate the times: red for 30 min, green for 60 min, blue for 90 min, yellow for 120 min, and black for 150 min.

Графики для выборок типа ЕТ и ТТ похожи между собой и расположением цветов и углами наклона. Для разных заблаговременностей углы наклона немного отличаются друг от друга, угол выше – для большей заблаговременности (ср., например, кривые красного и черного цвета). Объяснить более высокое положение кривой для отдаленной заблаговременности можно тем же расширением области доступных данных с ростом прогностического срока: при увеличении пространственного масштаба в масках наличия допусков прежние нулевые значения сменяются единичными значениями, и кривая относительных объемов для отдаленной заблаговременности оказывается выше кривой более ранней заблаговременности. Небольшое отличие в углах наклона кривых разных цветов свидетельствует о более быстром "подселении" допусков с увеличением срока прогноза.

5.4. Количество допусков для верификации в условной географии

Рассмотрим карты исходных объединенных полей в условной географии, содержащие данные об объемах выборок четырех типов. Для удобства графического представления абсолютные объемы для карт переводятся, как и раньше, в относительные объемы делением всех значений на максимальные величины в соответствии с данными табл. 14. Будем сопоставлять карты на 30 и 150 мин, чтобы наиболее отчетливо представить основные географические свойства в зависимости от типов формирования выборок и от сроков прогнозов в предположении, что в промежуточные сроки выделяемые особенности и тенденции в целом сохраняются.

На рис. 10 отчетливо видны различия между картами типов Е и картами остальных типов, а также сходство карт трех последних типов между собой на начальных сроках прогноза и заметное отличие на последних сроках только для типов ЕТ и ТТ. На картах подтверждаются и пространственно уточняются данные таблиц и особенности графиков: 1) изменение объемов видно по изменению цвета в сторону верхней или нижней цветовой палитры; в частности, рост происходит в основном в областях пересечений кругов обзора (ЦФО), что отражает важную роль перекрывающихся зон; 2) все изменения объемов не выходят за зоны обзора р/л.

Тенденции роста или падения объемов выборок проявляются на картах следующим образом. *Рост объема выборок* с прогностическим сроком означает появление *новых* точек с доступными значениями интенсивности (включая нулевую интенсивность). Эти точки расположены *внутри* зон обзора локаторов – области вне этих зон остаются пустыми (белые области между кругами на всех картах рисунка остаются не заполненными). *Падение объема выборок* с ростом заблаговременности для типов Е и Т – это утрата значащих пар прогноз-наблюдение.

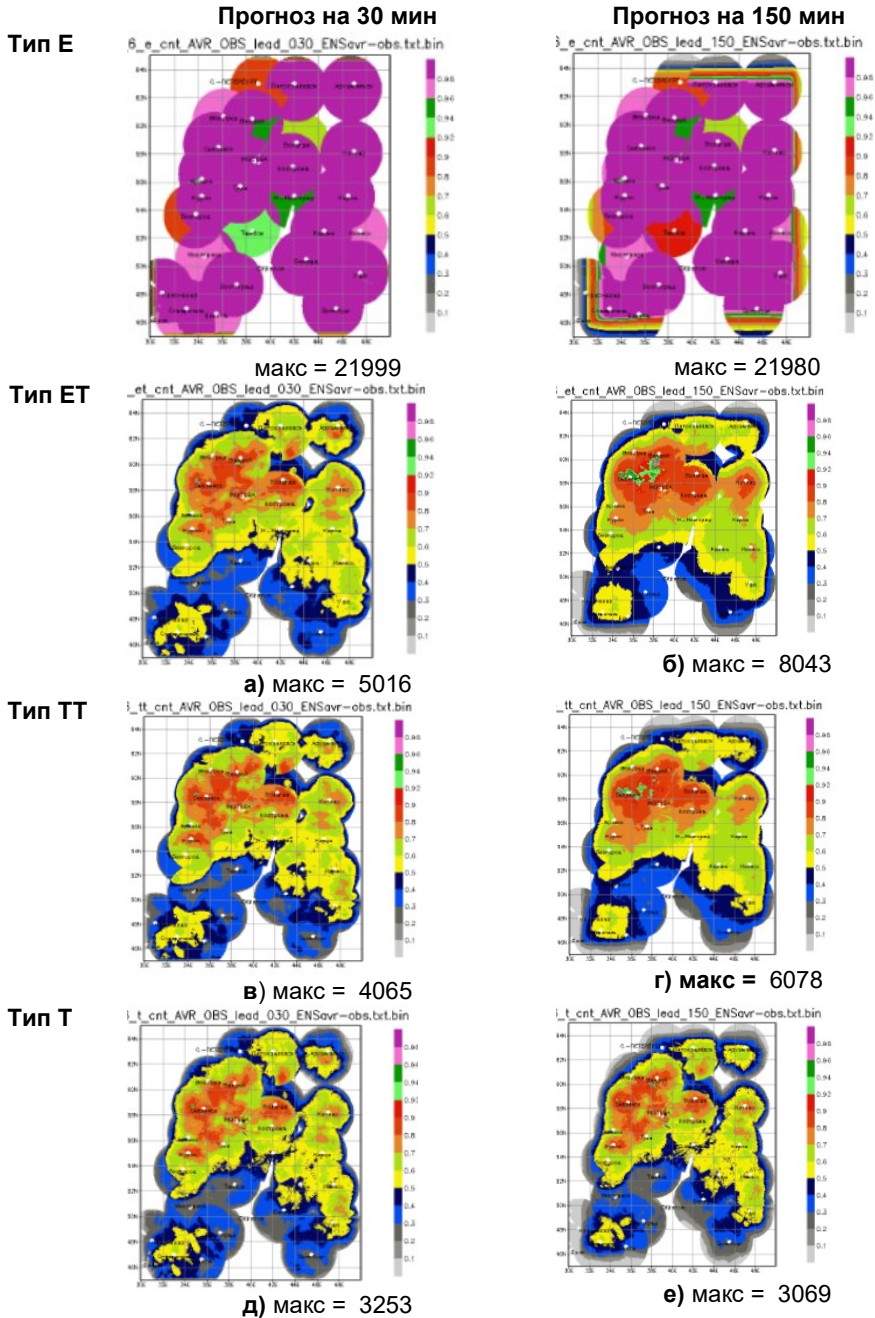


Рис. 10. Карты объемов выборок в долях от максимального объема (табл.16: столбец frc=30, строка qnt=100) для согласованных пар при прогнозах на 30 и 150 мин.
Fig. 10. Maps of sample volumes as fractions of the maximal volume (table 16: column frc=30, row qnt=100) for pairs for 30 and 150 min forecasts.

Для типа Е это равносильно отсутствию наблюдений или прогнозов в определенный срок из-за самых разнообразных обстоятельств, таких как аварии на локаторах, сбои в информационном потоке, нарушение работы расчетного сервера. Как правило, при отсутствии данных какого-либо локатора только его круг обзора заполняется константами отсутствия, узлы объединенного поля в остальных зонах обзора продолжают использоваться в прогнозе. Для типа Т к вышеперечисленным причинам добавляется потеря пары из-за нарушения условия отбора данных ("оба значения ненулевые"), когда одно из ненулевых значений обнуляется в ходе расчета следующего шага прогноза. В данной статье такое явление уже объяснялось систематическим "размазыванием" очагов осадков в процессе статистического прогнозирования.

При визуальном анализе можно увидеть наиболее странную особенность прогностической системы, не учитываемую таблицами и графиками – наличие выраженного краевого эффекта, заключающегося в появлении и увеличении с заблаговременностью чередующихся полос на картах типа Е, напоминающих угловую часть разноцветной рамки. На картах других типов заметный "уголок" серого цвета (когда объемы выборок минимальные) заполняет краевую область обзора Краснодарского локатора в нижнем левом углу условной карты. Возможно, указанные явления вызваны процедурой заполнения константами отсутствия краевых точек поля, оказывающихся в тыловой части адвективного потока, генерируемого моделью STEPS.

Рассмотрим более детально зависимость относительных объемов выборок от заблаговременности. Кривые нормированных квантилей для срока прогноза 30 мин (рис. 9, красные линии) очень похожи на панелях типов ЕТ и ТТ – они сильно расходятся и растут с ростом заблаговременности, в отличие от аналогичных кривых для типа Т, которые ввиду этого выходят за пределы единичной абсциссы на рис. 9. Дадим интерпретацию данному обстоятельству по картам рис. 10. Видно, что на сроке прогноза 30 мин карты трех типов выборок (ЕТ, ТТ и Т) довольно сходны между собой (рис. 10б, в, г), в то же время к концу прогностического срока сходство теряется в картах типов ТТ и Т. Хорошо видно, что заметное географическое сходство в панелях прогноза на 30 мин теряется к прогностическому сроку 150 мин в типах выборок ЕТ и ТТ в областях обзора практически всех локаторов.

Наибольшие исходные объемы наблюдаются на территории ЦФО, в регионе с наиболее плотным расположением около десятка локаторов, обеспечивающих и наибольшие пересечения зон обзоров. В этом же регионе происходит и наибольший рост объемов с ростом заблаговременности для выборок типа ЕТ и ТТ.

Очевидно, что увеличение объемов выборок благоприятно для увеличения статистической значимости оценок качества, однако этот рост вызван, как указывалось неоднократно раньше, генерированием малых

чисел в ходе статистического экстраполирования полей осадков. Возникает вопрос о предпочтениях: будет ли оценка качества более надежной, если воспользоваться только типом выборки Т даже при учете того, что в последнем случае объемы выборок намного меньше объемов выборок типов ЕТ и ТТ? Ситуация требует компромиссного выбора между снижением дисперсии и несмещенностью оценки качества; именно поэтому для более полного отражения имеющейся неопределенности рекомендуется проводить расчеты на выборках, сформированных несколькими способами.

Выводы

Сформулируем вначале общие выводы. Процедура оценивания прогнозов предваряется подготовкой таких выборок (назовем их "корректными выборками"), которые соответствуют двум взаимосвязанным целям планируемого испытания: оценить ошибки прогностической системы и указать на причины этих ошибок для усовершенствования эксплуатируемой системы. Суммарная ошибка прогноза может быть разложена по меньшей мере на три компонента – модельный, инструментальный и выборочный, и определение долей этих компонентов представляет собой содержательную задачу верификации в целом. В этом смысле выборка может считаться корректной, если она 1) статистически репрезентативна для оценивания качества прогнозов и 2) отражает существенные особенности прогностической модели и наблюдательного инструмента.

Задача формирования корректной выборки для верификации многократно усложняется при работе с данными большого объема и сложной структуры, которые могут оказаться одновременно неполными, содержащими пропуски, а также обладающими пространственно-временным распределением, далеким от нормального. Большие объемы поставляются современными системами наукастинга, в том числе статистическими моделями радиолокационного прогноза полей осадков. Структурную сложность существенно увеличивают методы и приемы пространственной верификации, в которой используются и точки, и пространственные объекты, отбираемые как для стандартного анализа, так и для приложения теорий со специфическими требованиями к исходным выборкам.

Представленные в данной статье результаты демонстрируют разнообразие формируемых выборок для верификации прогнозов в рамках плановых испытаний системы наукастинга полей осадков в указанные периоды года. Рассчитаны и обсуждены характеристики объектов и ситуаций, выделяемых для пространственно-ориентированной верификации с применением теории экстремальных величин. Показано, насколько наличие пропусков в точках полей и разные способы формирования пар прогноз-наблюдение влияют на объемы и на статистические свойства выборок для верификации.

Специфицируем конкретные выводы в соответствии с общей схемой выделения основных источников ошибок при пространственной верификации. Во-первых, к ошибкам модели следует отнести трудно устранимую систематическую особенность регрессионной системы прогноза "размазывать" области существенных осадков и одновременно "гормозить" их динамику в пространстве. Это приводит к утрате реальных и к порождению ложных *объектов* для верификации. Во-вторых, следует особое внимание обратить на автоматизированные методики определения, выделения и прослеживания объектов в ходе прогноза. В частности, вполне возможно, что потери с ростом заблаговременности *объектов*, идентифицированных нами с помощью пакета SpatialVx, удастся скомпенсировать построением *сопряженных областей*, как это делается в пакете CRA, или переходом к *выпуклым объектам* (методом "обвертывающей резинки") по методу, реализованному в пакете MET-MODE. Наконец, наличие пропусков в исходных полях и их влияние на качество прогностической продукции объясняются техническими особенностями радиолокационной сети, а также системами построения и первичного контроля полей отражаемости. И техническое состояние сети, и первичный контроль радиолокационной продукции обеспечиваются вне структуры Гидрометцентра России.

В следующих статьях предполагаемой серии планируется учесть влияние указанных особенностей выборок на пространственные оценки качества радиолокационного наукастинга осадков.

Авторы благодарят Антонида Александровну Алексееву за полезные советы, улучшившие содержание и стиль статьи.

Представленные работы выполнены в рамках темы НИР Росгидромета 1.1.5, рассчитанной на период 2020–2024 гг.

Список литературы

1. *Абдуллаев С.М.* Жизненный цикл мезомасштабных конвективных систем: концепция, климатология и прогноз: Дис. ... докт. геогр. наук. М, 2010. 408 с.
2. *Алибегова Ж.Д.* Пространственно-временная структура полей жидких осадков Л.: Гидрометиздат, 1985. 224 с.
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. *Вельтищев Н.Ф., Степаненко В.М.* Мезометеорологические процессы. М.: Географический факультет МГУ, 2006. 100 с.
5. *ВМО-168.* Руководство по гидрологической практике. Том II. Управление водными ресурсами и практика применения гидрологических методов. Изд. 6, 2012 (2009). Глава 5. Анализ экстремальных значений. II-5.1 – II-5.58.
6. *Галамбош Я.* Асимптотическая теория экстремальных порядковых статистик. М.: Наука, 1984. 305 с.
7. *Галамбош Я.* О развитии математической теории экстремумов за последние полвека // Теория вероятностей и ее применения. 1994. Т. 39, № 2. С. 272–293.
8. *Горшенин А.К., Королев В.Ю.* Определение экстремальности объемов осадков на основе модифицированного метода превышения порогового значения // Информатика и ее применения. 2018. Т. 12, вып. 4. С. 16–24.

9. Гумбель Э. Статистика экстремальных значений. М.: Мир, 1965. 450 с.
10. Гумбель Э. Статистическая теория экстремальных значений (основные результаты) / Введение в теорию порядковых статистик. М.: Фазис, 1970. С. 61-93.
11. Исаев А.А. Атмосферные осадки. Часть I. Изменчивость характеристик осадков на территории России и сопредельных стран. М: Издательство МГУ, 2002. 192 с.
12. Исаев А.А. Атмосферные осадки. Часть II. Мезоструктура полей жидких осадков. М.: Географический факультет МГУ, 2001. 100 с.
13. Киктев Д.Б., Муравьев А.В., Смирнов А.В. Прогнозирование осадков с помощью модели мультипликативного каскада: опыт радарного наукастинга // Международная конференция «Турбулентность, динамика атмосферы и климата», посвященная 100-летию со дня рождения академика А.М. Обухова, Москва, 16-18 мая 2018». Тезисы докладов. 2018. С. 27.
14. Киктев Д.Б., Муравьев А.В., Смирнов А.В. Наукастинг метеорологических параметров и опасных явлений: опыт реализации и перспективы развития // Гидрометеорологические исследования и прогнозы. 2019. № 4 (374). С. 151-160.
15. Крамер Г., Лидбеттер М. Стационарные случайные процессы. Свойства выборочных функций и их приложения. М.: Мир, 1969. 400 с.
16. Лагутин М.Б. Наглядная математическая статистика. М.: Бином, 2009. 470 с.
17. Лебедев А.В. Основы стохастической теории экстремумов. М.: ЛЕНАНД, 2018. 104 с.
18. Лидбеттер М., Ротсен Х., Линдгрен Г. Экстремумы случайных последовательностей и процессов. М.: Мир, 1989. 392 с.
19. Литтл Р.Д.А., Рубин Д.Б. Статистический анализ данных с пропусками. М.: Финансы и статистика, 1991. 169 с.
20. Муравьев А.В., Киктёв 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.
21. Муравьев А.В., Киктёв Д.Б., Смирнов А.В., Зайченко М.Ю. Оперативная технология наукастинга осадков на основе радиолокационных данных и сравнительные результаты точечной верификации для теплого и холодного периодов года // Гидрометеорологические исследования и прогнозы. 2019. № 2 (372). С. 12-40.
22. Муравьев А.В., Киктев Д.Б., Смирнов А.В. Усовершенствованная технология радиолокационного наукастинга осадков и результаты верификации в теплый период года (май - сентябрь 2020 г.) // Результаты испытания новых и усовершенствованных технологий, моделей и методов гидрометеорологических прогнозов. 2021. Информационный сборник № 48. С. 13-41.
23. Муравьев А.В., Киктёв Д.Б., Смирнов А.В. Сравнительная верификация усовершенствованной системы радиолокационного наукастинга осадков с учетом пропусков и при различных методах формирования выборок (по результатам испытаний в теплый период года май-сентябрь 2017 и 2020 гг.) // Результаты испытания новых и усовершенствованных технологий, моделей и методов гидрометеорологических прогнозов. 2022. Информационный сборник № 49. С. 3-56.
24. Родионов И.В. О параметрическом оценивании хвоста распределения // Доклады Академии наук. 2019. Т. 488, № 4. С. 356-359.
25. Родионов И.В. Вероятностный и статистический анализ экстремумов дискретных стохастических систем: Дис. ... докт. физ.-мат. наук. М, 2021. 259 с.
26. Тихонов В.И. Выбросы случайных процессов. М.: Наука, 1970. 392 с.
27. Тутубалин В.Н. Теория вероятностей. М.: Академия, 2008. 368 с.
28. Beirlant J., Goegebeur, Y., Segers, J., Teugels, J., Waal, D. D., Ferro, C. Statistics of Extremes: Theory and Applications. New York: Wiley, 2004. 512 p.
29. Bowler N., Pierce C., Seed A. STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP // Q. J. R. Meteorol. Soc. 2006. Vol. 132. P. 2127-2155.

30. *Cunnane C.* A note on the Poisson assumption in partial duration series models // *Water Resour. Res.* 1979. Vol. 15, no. 2. P. 489-494.
31. *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.
32. *Davis C.A., Brown B. G., Bullock R. G.* Object-based verification of precipitation forecasts, Part II: Application to convective rain systems // *Mon. Wea. Rev.* 2006. 134. P. 1785–1795.
33. *De Haan L., Ferreira A.* Extreme value theory. An introduction. Springer, 2006. 420 p.
34. *Dorninger M., Mittermaier M. P., Gilleland E., Ebert E.E., Brown B.G., Wilson L.J.* MesoVICT: Mesoscale Verification Inter-Comparison over Complex Terrain // NCAR/TN-505. 2013. 23 p.
35. *Ebert E., McBride J.* Verification of precipitation in weather systems: Determination of systematic errors // *J. Hydrol.* 2000. Vol. 239. P. 179-202.
36. *Forecast Verification. A Practitioner's Guide in Atmospheric Science* // I.T. Jolliffe, D.B. Stephenson (Eds). John Wiley & Sons Ltd, 2012. 274 p.
37. *Gilleland E.* 2020. <https://cran.r-project.org/web/packages/SpatialVx/SpatialVx.pdf>
38. *Gilleland E., Ahijevych D., Brown B.G., Casati B., Ebert E.E.* Intercomparison of spatial forecast verification methods // *Wea. Forecast.* 2009. Vol. 24. P. 1416-1430.
39. *Gilleland E., Skok G., Brown B., Casati B., Dorninger M., Mittermaier M.P., Roberts N., Wilson L.J.* A Novel Set of Geometric Verification Test Fields with Application to Distance Measures // *Mon. Wea. Rev.* 2020. Vol. 148. P. 1653-1673.
40. *Gilleland E., Katz R.W.* extRemes 2.0: An Extreme Value Analysis Package in R // *Journal of Statistical Software.* 2016. Vol. 72, no. 8. P. 1-39.
41. *Hsing T., Hüsler J., Leadbetter M.R.* On the exceedance point process for a stationary sequence // *Probab. Theory and Related Fields.* 1988. No. 78. P. 97-112.
42. *Houze R.A.Jr.* 100 years of research on mesoscale convective systems // *A Century of Progress in Atmospheric and Related Sciences: Celebrating the American Meteorological Society Centennial*, Meteor. Monogr. 2019. No. 59. <https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0001.1>
43. *Hyndman R.J., Fan Y.* Sample Quantiles in Statistical Packages // *The American Statistician.* 1996. Vol. 50, no. 4. P. 361-365.
44. *Leadbetter M.R.* On a basis for “Peaks over Threshold” modeling // *Stat. Probabil. Lett.* 1991. Vol. 12, no. 4. P. 357-362.
45. *Little R.J.A., Rubin D.B.* Statistical analysis with missing data (2-d Ed.). Wiley & Sons Inc., 1987. 381 p.
46. *Reiss R.D., Thomas M.* Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields: 3d Ed. Berlin: Birkhauser Verlag, 2007. 516 p.
47. *Resnick S.* Extreme Values, Regular Variation, and Point Processes. New York: Springer Verlag, 1987. 320 p.
48. *Roberts N.M., Lean H.W.* Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events // *Mon. Wea. Rev.* 2008. Vol. 136. P. 78-97.
49. *Rubin D.B.* Inference and Missing Data // *Biometrika.* 1976. Vol. 63, no. 3. P. 581-592.
50. *Rubin D.B.* Multiple imputation after 18+ years // *J. Amer. Statist. Assoc.* 1996. Vol. 91, no. 434. P. 473-489.
51. *Seed A.W., Srikanthan R., Menabde M.* A space and time model for design storm rainfall // *Jour. Geophys. Research.* 1999. Vol. 194. D24. P.31623-31630.
52. *Seed A.W.* A dynamic and spatial scaling approach to advection forecasting // *J. Appl. Met.* 2003. Vol. 42. P. 381-388.
53. *Seed A.W.* Modelling and forecasting rainfall in space and time // *Scales in Hydrology and Water management (IAHS Publ. 287).* 2004. P. 137-152.
54. *Smirnov A.V., Muravev A.V., Kiktev D.B.* Combining precipitation fields on the basis of radar data and mesoscale model output products in nowcasting systems of Hydrometcenter of Russia. // *Research activities in Earth system modelling. Working Group on Numerical Experimentation. Report No. 50. WCRP Report No.12.* Geneva, WMO. 2020. P. 2-19.

55. *Suveges M., Davidson A.C.* Model misspecifications in Peaks Over Threshold analysis // *Annals Appl. Statistics.* 2010. Vol. 4, no. 1. P. 203-221.
56. *Wilks D.S.* Statistical methods in the atmospheric sciences: 4th Ed. Elsevier, 2019. 816 p.
57. *WMO-No.237.TP129.* Manual for depth – area – duration analysis of storm precipitation. 1969. 131 p.

References

1. *Abdullaev S.M.* Zhiznennyj cikel mezomasshtabnyh konvektivnyh sistem: koncepciya, klimatologiya i prognoz: dissertacia doktora geograficheskikh nauk. Moscow, 2010, p. 408. [in Russ.].
2. *Alibegova J.D.* Prostranstvenno-vremennaya structura poley jidkikh osadkov. Leningrad, Gidrometeoizdat Publ., 1985. 224 p. [in Russ.].
3. *Bundel A.Yu., Muravyev A.V., Olkhovaya E.D.* Overview of spatial verification methods and their application to ensemble forecasting. *Gidrometeorologicheskie issledovaniya i prognozy [Hydrometeorological Research and Forecasting]*, 2021, vol. 382, no. 4, pp. 30-49. [in Russ.].
4. *Veltishev N.F., Stepanenko V.M.* Mezometeorologicheskie protsessy. Moscow, MGU, Geografichesky fakultet, 2006, 100 p. [in Russ.].
5. *WMO-168.* Rukovodstvo po gidrologicheskoy praktike. Tom II. Upravlenie vodnymi resursami i praktika primeneniya gidrologicheskikh metodov. Izd. 6, 2012 (2009). Glava 5. Analiz ekstremalnyh znacheniy. II-5.1 - II-5.58. [in Russ.].
6. *Galambos J.* The Asymptotic Theory of Extreme Order Statistics. 1978, Wiley.
7. *Galambos J.* The development of the mathematical theory of extremes in the past half century. *Theory of Probability and its Applications.* 1994, 39:2, pp.234–248.
8. *Gorshenin A.K., Korolev V.Yu.* Opredeleniye ekstremalnosti obyemov osadkov na osnove modifitsirovannogo metoda prevysheniya porogovogo znacheniya. *Informatike i ee primeniya*, 2018, vol.12, no.4, pp. 16-24. [in Russ.].
9. *Gumbel E.* Statistics of extremes. N. Y., Columbia Univ. Press, 1958.
10. *Gumbel E.* Statistical theory of extreme values (main results). (In: Contributions to order statistics. Ed. Sarhan A., Greenberg B. John Wiley & Sons, New York, London. 1962), pp. 56-94.
11. *Isayev A.A.* Atmosfernye osadki. Chast I. Izmenchivost harakteristik osadkov na territorii Rossii i sopedelnyh stran. Moscow, MGU publ, 2002, 192 p. [in Russ.].
12. *Isayev A.A.* Atmosfernye osadki. Chast II. Mezostruktura poley jidkikh osadkov. Moscow, MGU publ., 2001, 100 p. [in Russ.].
13. *Kiktev D. B., Murav'ev A. V., Smirnov A. V.* Precipitation forecasting based on multiplicative cascade model: radar nowcasting experience. Proceedings of the Int. Conference “Turbulence, atmosphere and climate dynamics”, dedicated to centenary of acad. A. M. Obukhov, Moscow, May 16-18, 2018, p. 27. [in Russ.].
14. *Kiktev D. B., Murav'ev A. V., Smirnov A. V.* Nowcasting of meteorological parameters and hazards: implementation experience and development prospects. *Gidrometeorologicheskie issledovaniya i prognozy [Hydrometeorological Research and Forecasting]*, 2019, vol. 374, no. 7, pp. 151-160. [in Russ.].
15. *Kramer H., Leadbetter M.R.* Stationary and related stochastic processes. Sample function properties and their applications. John Wiley, New York, London, 1967.
16. *Lagutin M.B.* Naglyadnaya matematicheskaya statistika. Moscow, Binom, 2009, 470 p. [in Russ.].
17. *Lebedev A.V.* Osnovy stohasticheskoi teorii ekstremumov. Moscow, Lenend Publ., 2018, 104 p. [in Russ.].
18. *Leadbetter M.R., Lindzen G., Rootzen H.* Extremes and related properties of random sequences and processes. Springer-Verlag, New York, Berlin, 1986.
19. *Little R.J.A., Rubin D.B.* Statistical analysis with missing data (2-d Ed.). Wiley & Sons Inc., 1987.

20. Murav'ev A.V., Kiktev D.B., Smirnov A.V. Operational precipitation nowcasting system based on radar data and verification results for the warm period of the year (May-September 2017). *Gidrometeorologicheskie issledovaniya i prognozy [Hydrometeorological Research and Forecasting]*, 2018, vol. 367, no.1, pp. 6-38. [in Russ.].
21. Murav'ev A.V., Kiktev D.B., Smirnov A.V., Zaichenko M.Yu. Operational precipitation nowcasting system based on radar data and comparative pointwise verification results for the warm and cold seasons. *Gidrometeorologicheskie issledovaniya i prognozy [Hydrometeorological Research and Forecasting]*, 2019, vol. 372, no. 2, pp. 12-40 [in Russ.].
22. Murav'ev A.V., Kiktev D.B., Smirnov A.V. Usovershenstvovannaya tehnologiya radiolokazionnogo naukastinga osadkov i rezultaty verifikacii v teplyi period goda (mai - sentyabr 2020). *Rezultaty ispytaniya novih i usivershenstvovannyh tehnologiy, modeley i metodov gidrometeorologicheskikh prognozov*, 2021, *Informacionnyi sbornik*, no.48, pp. 13-41. [in Russ.].
23. Murav'ev A.V., Kiktev D.B., Smirnov A.V. Sravnitel'naya verifikaciya usovershenstvovannoy sistemy radiolokazionnogo naukastinga osadkov s uchedom propuskov i pri razlichnyh metodah formirovaniya vyborok (po rezul'tatam ispytaniy v teplyi period goda mai - sentyabr 2017 i 2020 gg). *Rezultaty ispytaniya novih i usivershenstvovannyh tehnologiy, modeley i metodov gidrometeorologicheskikh prognozov*, 2022, *Informacionnyi sbornik*, no. 49, pp. 3-56. [in Russ.].
24. Rodionov I.V. O parametricheskom ocenivaniy hvosta raspredeleniya. *Doklady AN*, 2019, vol. 488, no.4, pp. 356-359. [in Russ.].
25. Rodionov I.V. Veroyatnostnyi i statisticheskii analiz ekstremumov diskretnykh stokhasticheskikh sistem: dissertaciya doktora geograficheskikh nauk. Moscow, 2021, 259 p. [in Russ.].
26. Tikhonov V.I. Vybrosty sluchainykh processov. Moscow, Nauka Publ., 1970, 392 p. [in Russ.].
27. Tutubalin V.N. Teoriya veroyatnostei. Moscow, Akademia, 2008, 368 p. [in Russ.].
28. Beirlant J., Goegebeur, Y., Segers, J., Teugels, J., Waal, D. D., Ferro, C. *Statistics of Extremes: Theory and Applications*. Wiley, New York, 2004. 512 p.
29. Bowler N., Pierce C., Seed A. STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Q. J. R. Meteorol. Soc.*, 2006, Vol. 132, P. 2127-2155.
30. Cunnane C. A note on the Poisson assumption in partial duration series models. *Water Resour. Res.*, 1979, vol. 15, no. 2, pp. 489-494.
31. 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, pp. 1772-1784.
32. Davis C.A., Brown B.G., Bullock R.G. Object-based verification of precipitation forecasts, Part II: Application to convective rain systems. *Mon. Wea. Rev.*, 2006, vol. 134, pp. 1785-1795.
33. De Haan L., Ferreira A. *Extreme value theory. An introduction*. Springer, 2006, 420 p.
34. Dorninger M., Mittermaier M. P., Gilleland E., Ebert E.E., Brown B.G., Wilson L.J. MesoVICT: Mesoscale Verification Inter-Comparison over Complex Terrain. NCAR/TN-505, 2013, 23 p.
35. Ebert E., McBride J. Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.*, 2000, vol. 239, pp. 179-202.
36. *Forecast Verification. A Practitioner's Guide in Atmospheric Science*. I.T. Jolliffe, D.B. Stephenson (Eds). John Wiley & Sons Ltd, 2012, 274, 37 p.
37. Gilleland E. 2020. Available at: <https://cran.r-project.org/web/packages/SpatialVx/SpatialVx.pdf>.
38. Gilleland E., Ahijevych D., Brown B.G., Casati B., Ebert E.E. Intercomparison of spatial forecast verification methods. *Wea. Forecast.*, 2009, vol. 24, pp. 1416-1430.
39. Gilleland E., Skok G., Brown B., Casati B., Dorninger M., Mittermaier M.P., Roberts N., Wilson L.J. A Novel Set of Geometric Verification Test Fields with Application to Distance Measures. *Mon. Wea. Rev.*, 2020, vol. 148, pp. 1653-1673.
40. Gilleland E., R.W. Katz. extRemes 2.0: An Extreme Value Analysis Package in R. *Journal of Statistical Software*, 2016, vol. 72, no. 8, pp. 1-39.

41. Hsing T., Hüsler J., Leadbetter M.R. On the exceedance point process for a stationary sequence. *Probab. Theory and Related Fields*, 1988, no.78, pp. 97-112.
42. Houze R. A. Jr. 100 Years of Research on Mesoscale Convective Systems, Meteorological Monographs, vol. 59, pp. 17.1-17.54. DOI: <https://doi.org/10.1175/AMSMONOGRAPHSD-18-0001.1>.
43. Hyndman R.J., Fan Y. Sample Quantiles in Statistical Packages. *The American Statistician*, 1996, vol. 50, no. 4, pp. 361-365.
44. Leadbetter M.R. On a basis for “Peaks over Threshold” modeling. *Stat. Probabil. Lett.*, 1991, vol. 12, no. 4, pp. 357-362.
45. Little R.J.A., Rubin D.B. Statistical analysis with missing data (2-d Ed.). *Wiley & Sons Inc.*, 1987, 381 p.
46. Reiss R.D., Thomas M. Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields. 3d Ed. Berlin, Birkhauser Verlag, 2007, 516 p.
47. Resnick S. Extreme Values, Regular Variation, and Point Processes. *Springer Verlag*, 1987, New York, 320 p.
48. Roberts N.M., Lean H.W. Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Mon. Wea. Rev.*, 2008, vol. 136, pp. 78-97.
49. Rubin D.B. Inference and Missing Data. *Biometrika*, 1976, Vol. 63, No. 3, pp. 581-592.
50. Rubin D.B. Multiple imputation after 18+ years. *J. Amer. Statist. Assoc.*, 1996, vol. 91, no. 434, pp. 473-489.
51. Seed A.W., Srikanthan R., Menabde M. A space and time model for design storm rainfall. *Jour. Geophys. Research*, 1999, Vol.194, D24, pp. 31623-31630.
52. Seed, A. W. A dynamic and spatial scaling approach to advection forecasting. *J. Appl. Met.*, 2003, vol. 42, pp. 381-388.
53. Seed A.W. Modelling and forecasting rainfall in space and time. Scales in Hydrology and Water management (IAHS Publ. 287), 2004, pp.137-152.
54. Smirnov A.V., Muravev A.V., Kiktev D.B. Combining precipitation fields on the basis of radar data and mesoscale model output products in nowcasting systems of Hydrometcenter of Russia. *Research activities in Earth system modelling*. Working Group on Numerical Experimentation. Report No. 50. WCRP Report No.12. WMO, Geneva, 2020, pp. 2-19.
55. Suveges M., A. C. Davidson. Model misspecifications in Peaks Over Threshold analysis. *Annals Appl. Statistics*, 2010, vol. 4, no. 1, pp. 203-221.
56. Wilks D.S. Statistical methods in the atmospheric sciences. 4th Ed. Elsevier, 2019, 816 p.
57. WMO-No.237. TP129. Manual for depth - area - duration analysis of storm precipitation. 1969, 131 p.

Поступила 03.08.2021; одобрена после рецензирования 23.05.2022;
принята в печать 08.06.2022.

Submitted 03.08.2021; approved after reviewing 23.05.2022;
accepted for publication 08.06.2022