Гидрометеорология и экология
№ 1 2007

УДК 551.582.2

О ПРИМЕНЕНИИ МЕТОДА ГУМБЕЛЯ К АППРОКСИМАЦИИ РАСПРЕДЕЛЕНИЯ СРЕДНЕЙ МАКСИМАЛЬНОЙ ТЕМПЕРАТУРЫ ВОЗДУХА И СУТОЧНЫХ МАКСИМУМОВ ОСАДКОВ

К.Т. Елеуова

Исследованы ряды наблюдений средней максимальной температуры воздуха и суточных максимальных осадков на возможность применения метода Гумбеля в вероятностных прогнозах. Выполнена оценка согласия эмпирических рядов с теоретическим распределением Гумбеля по критериям Пирсона и Колмогорова.

Одним из основных приемов климатологической обработки генеральной совокупности метеорологических наблюдений является представление эмпирического материала в виде теоретической функции распределения. В результате расчет климатических показателей можно выполнять на основе аппроксимации распределений метеорологических элементов теми или иными теоретическими функциями распределений. Простейшие функции распределения типа нормального, Пуассона, логнормального, гамма-функций применимы к распределениям значений случайной величины в выборке. Для расчета максимальных (минимальных) значений метеорологических элементов с определенным уровнем вероятности существует три типа асимптотических (предельных) распределений, описывающих распределение крайних членов выборки, подробно изученных Э.И. Гумбелем. Предельное распределение используется для расчета максимальных (минимальных) значений метеорологических элементов с определенным уровнем вероятности.

В данной работе исследованы ряды наблюдений средней максимальной температуры воздуха и суточные максимумы осадков для получения вероятностных характеристик метеорологических элементов редкой повторяемости по методу Гумбеля, рекомендованного климатологами Главной геофизической обсерватории им. А.И. Воейкова (Россия, Санкт-Петербург) [1].

Цель данной работы – исследование по оценкам согласия соответствия эмпирических рядов теоретическим распределениям. В основу методики расчета вероятностных величин положено первое предельное распределение, называемое двойным экспоненциальным распределением [3]:

$$F(x) = \exp\left(-e^{-y}\right), \quad (1)$$

или

$$P(x \geq x_0) = 1 - e^{-e-y}. \quad (2)$$

Связь между y и x_0 задается формулами:

$$y = \frac{\sigma_n}{\sigma_x}\left(x_0 - \overline{x}\right) + y_n, \quad (3)$$

$$x_0 = \overline{x} + \frac{\sigma_x}{\sigma_n}\left(y - \overline{y}_n\right), \quad (4)$$

где $\overline{x}$ и ${\sigma_x}^2$ – среднее и дисперсия ряда максимумов случайной величины; $\overline{y_n}$ и σ_n – среднее и среднее квадратическое отклонение ряда вспомогательной величины.

Значения $\overline{y}_n$ и σ_n □, которые зависят от длины ряда, представлены в табл. 1.

Таблица 1

Значения $\overline{y}_n$ и σ_n в зависимости от длины ряда

n	$\overline{y}_n$	σ_n	n	$\overline{y}_n$	σ_n
10	0,4952	0,9497	19	0,5220	1,0565
11	0,4996	0,9676	20	0,5236	1,0626
12	0,5085	0,9833	21	0,5252	1,0796
13	0,5070	0,9972	22	0,5268	1,0754
14	0,5100	1,0095	23	0,5283	1,0811
15	0,5128	1,0206	24	0,5296	1,0864
16	0,5157	1,0316	25	0,5309	1,0914
17	0,5181	1,0411	80	0,5569	1,1938
18	0,5202	1,0493	100	0,5600	1,2065

После преобразования формулы (1) в работе для получения исследуемой величины возможной в 10, 20, 50, 100 лет используется следующая формула:

$$T = t + \frac{\sigma}{\sigma_n}\left\{-\ln\left[-\ln F(T > x)\right] - \overline{y}_n\right\}, \quad (5)$$

$$F(t > x) = 1 - 1/T', \quad (6)$$

где T – значение элемента, возможное 1 раз в 10, 50, 100 лет; t – среднее значение годового максимума исследуемого элемента; σ – среднее квадратическое отклонение годового максимума; σ_n и $\overline{y}_n$ – параметры, зависящие от длины исходных рядов N; $F(t > x)$ – величина, связанная с периодом повторения 1 раз в N лет; T' – период повторения, например: 10, 50, 100 лет.

В процессе исследования использованы данные средних годовых максимумов температуры воздуха и суточных максимальных осадков по 74 станциям за период наблюдений 1936….2005 гг. и стандартное квадратическое отклонение (σ) исследуемых метеорологических элементов. Обработан материал по вышеуказанным станциям в следующей последовательности:

1. Расчет среднего арифметического из совокупности годовых максимумов.
2. Расчет стандартного отклонения годовых максимумов.
3. Выбор σ_n и $\overline{y}_n$ для заданной продолжительности ряда в табл. 1.
4. Расчет *ln (1-1/10)*, где число 10 – это заданный период для нахождения расчетной величины возможной 1 раз в 10 лет, соответственно 1 раз в 20 лет – *ln (1-1/20)*; 1 раз в 50 лет – *ln (1-1/50)*; 1 раз в 100 лет – *ln(1-1/100)* и так далее.

Подставив в формулу (5) известные и найденные значения, были рассчитаны значения метеорологической величины возможной 1 раз в 10, 50, 100, 1000 лет. В табл. 2 представлены редкие события, возможные 1 раз в N лет по средней годовой максимальной температуре воздуха, табл. 3 – по максимальному суточному количеству осадков.

Таблица 2

Вероятностные значения средней годовой максимальной температуры воздуха (°C), рассчитанные по методу Гумбеля

Станция	Температура воздуха максимальная один раз			
	10 лет	30 лет	50 лет	100 лет
Петропавловск	8,3	9,3	9,8	10,5
Астана	9,8	11,0	11,5	12,2
Атырау	16,2	17,2	17,7	18,3
Караганда	10,5	11,6	12,0	12,7
Усть-Каменогорск	11,6	12,8	13,4	14,2
Катон-Карагай	9,6	10,7	11,2	11,9
Кокпекты	9,9	10,9	11,4	12,1
Аральское Море	15,4	16,7	17,3	18,2
Есик (Иссык)	15,9	16,8	17,2	17,8

Таким образом, распределение Гумбеля легло в основу аппроксимации эмпирических распределений средних годовых максимальных температур воздуха и максимальных суточных осадков. Выполнены работы и произведен анализ согласованности эмпирического и теоретического распределений. Естественно, что между эмпирическими и теоретическими рядами распределений всегда будет расхождение. В качестве примера приводятся два случая распределения суточных максимальных осадков на М Урджар, где распределения хорошо согласованы на 10 % уровне (рис. 1) и на М Эмба, где распределения не согласованы (рис. 2).

Таблица 3

Вероятностные значения суточного максимального количества осадков, рассчитанные по методу Гумбеля, (мм)

Станция	Суточные максимальные суммы осадков один раз			
	10 лет	30 лет	50 лет	100 лет
Петропавловск	46,9	58,9	64,4	71,7
Торгай	41,7	54,1	60,1	68,3
Астана	45,7	58,0	63,6	71,3
Павлодар	42,5	56,8	62,8	70,9
Мартук	42,1	52,6	57,4	63,9
Эмба	40,1	51,0	56,4	63,6
Бектауата	28,3	34,8	37,9	42,1
Усть-Каменогорск	46,5	58,5	64,0	71,4
Урджар	49,0	59,6	64,4	70,9
Кызылорда	26,0	33,0	36,2	40,5
Туркестан	30,5	38,4	42,1	47,0
Арысь	40,7	51,9	57,1	64,0
Алма-Ата, ОГМС	54,2	64,5	69,2	75,6

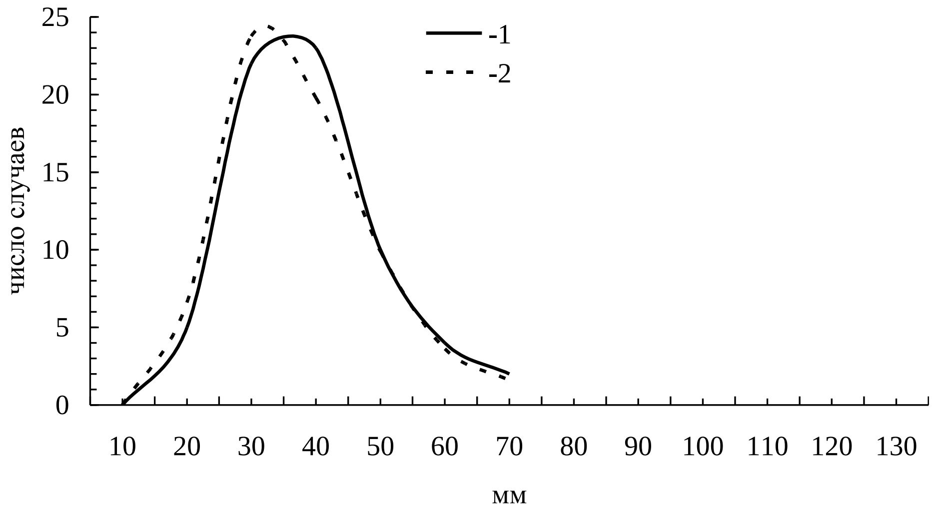

Рис. 1. Выравнивание эмпирического ряда суточных максимумов осадков распределением Гумбеля (М Урджар). 1 – эмпирический ряд; 2 – распределение Гумбеля.

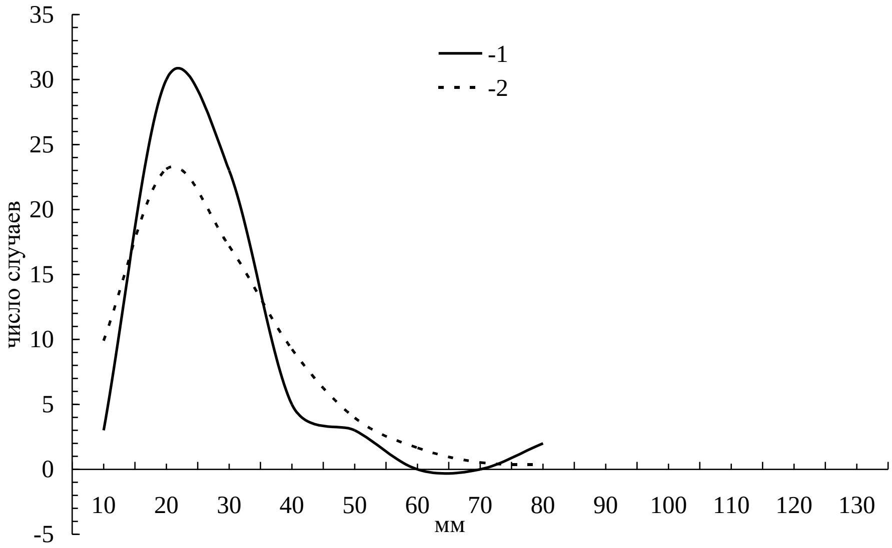

Рис. 2. Выравнивание эмпирического ряда суточных максимумов осадков распределением Гумбеля (М Эмба.). 1 – эмпирический ряд; 2 – распределение Гумбеля

Вследствие того, что эмпирическая кривая распределения получается на основе ограниченного числа наблюдений, затруднительно судить объективно о законе, которому подчиняется распределение генеральной совокупности значений изучаемой величины. Однако, существует ряд критериев согласия, основанных на изучении распределения различным образом составленной мерой расхождения [1]. Хотя ни один из критериев не может рассматриваться как достаточный, так как в основе лежит условно установленная мера расхождения и известная условность в оценки практически невозможных событий. Поэтому обычно рекомендуют производить оценку, по крайней мере, по двум критериям согласия. В практике наиболее широко применяются критерии Пирсона и А.Н. Колмогорова, которые и использованы в данной работе.

Средняя годовая максимальная температура воздуха

Оценка соответствия эмпирического и теоретического распределения по рядам средней годовой максимальной температуры воздуха была произведена в следующем порядке:

– расчет коэффициентов асимметрии (A_S) по 74 станциям,
– выбор 9 станций с учетом асимметричности эмпирического распределения,
– ранжирование рядов средней годовой максимальной температуры воздуха с построением кривой эмпирического распределения на вероятностной клетчатке, где ось ординат имеет равномерную шкалу,
– нанесение на вероятностную клетчатку расчетных значений параметров, полученных по методу Гумбеля,
– построение кривой теоретического и эмпирического распределений,
– расчет коэффициентов согласия Колмогорова и Пирсона между эмпирическим и теоретическим распределениями,
– расчет критерия согласия λ выбиралось в интервале максимального расхождения, где наблюдалось наибольшая повторяемость числа случаев.

Первоначально были выполнены расчеты и произведен анализ коэффициентов асимметрии по 74 станциям, который показал:

– в 84 % случаев наблюдалась отрицательная асимметричность, в 16 % случаев положительная,
– большая асимметричность составила 4 % случаев,
– умеренная асимметричность составила 20 % случаев,
– малая асимметричность наблюдалась в 76 % случаев,
– большая и умеренная асимметричность была только отрицательная,
– малая асимметрия в 20 % случаев была положительная и в 80 % – отрицательная.

Из обработанных станций выбраны девять станций в зависимости от места расположения станции и коэффициента асимметрии средней годовой максимальной температуры воздуха, которые представлены табл. 4. Затем ряды средней годовой максимальной температуры воздуха по выбранным станциям были упорядочены по возрастающим значениям, рассчитаны вероятности по формуле:

$$P_i = \frac{m_i}{(n+1)}, \tag{7}$$

где m_i – порядковый номер члена ранжированного ряда, n – общее число членов ряда.

Рассчитанные эмпирические вероятности наносились на вероятностную клетчатку для нормального распределения. Затем вручную проводилась сглаживающая кривая линия – график функции эмпирического распределения. Временные ряды средней максимальной температуры воздуха хорошо описываются нормальной кривой. Это было проверено путем введения новой переменной t по рядам средней годовой максимальной температуры воздуха по выбранным 9 станциям, где

$$t = \frac{\left(x_i - \bar{x}\right)}{\sigma}. \qquad (8)$$

Получено, что $\sum_{i-1}^{n} t_i = 0$, $\sigma_t = 1{,}0$.

Таблица 4

Коэффициент асимметрии (As) средней годовой максимальной температуры воздуха

Станция	Характеристика асимметричности	As	
		градация	величина
Усть-Каменогорск Кокпекты	большая отрицательная	$\lvert A_S \rvert > 0{,}50$	-0,54 -0,52
Караганда Есик	умеренная отрицательная	$0{,}25 < \lvert A_S \rvert \le 0{,}50$	-0,33 -0,39
Атырау Астана	малая отрицательная	$\lvert A_S \rvert \le 0{,}25$	-0,17 -0,16
Аральское Море Катон-Карагай Петропавловск	малая положительная		+0,21 +0,15 +0,04

На клетчатку с данными эмпирической вероятности по ранжированному ряду средней годовой максимальной температуры были нанесены расчетные значения средней максимальной температуры воздуха по распределению Гумбеля, возможные 1 раз в 10 лет, 20, 50 и т.д. лет.

Критерий согласия А.Н. Колмогорова (λ)

В качестве меры расхождения А.Н. Колмогоров предложил величину – критерий согласия Колмогорова λ.

$$\lambda = D\sqrt{n}\ , \qquad (9)$$

где n – объём совокупности (длина ряда); D – верхняя граница: наибольшее значение стандартного отклонения $\sigma(x_i)$, выбранного по графику на вероятностной сетчатке.

Трем пороговым значениям доверительной вероятности 0,95; 0,99 и 0,999 соответствуют критические значения λ = 1,36; 1,63; 1,95. Для определения λ не надо определять число степеней свободы. Нулевая гипотеза не отвергается, и расхождения между сопоставляемыми частотами считаются случайными, если критерий не превосходит своего критического значения для принятого порога доверительной вероятности [4]. Расчеты показали, что критерий согласия λ эмпирического распределения средней годовой максимальной температуры воздуха по всем исследуемым станциям не достигают уровня значимости 1 %, для которых λ = 1,63 (табл. 5). Причем на 5 % уровне значимости распределения хорошо согласуются на станции: Атырау, Астана, Аральское Море, Катон-Карагай, Петропавловск, где отмечается малая асимметрия эмпирического распределения: $|\mathrm{A}s| \leq 0{,}25$. На 1 % уровне значимости распределения согласуются на станции Усть-Каменогорск, Кокпекты, Караганда, Есик, где умеренная и большая асимметрия: $|\mathrm{A}s| \geq 0{,}25$. Однако, поскольку при оценке параметров генеральной совокупности по результатам выборочных наблюдений, как правило, принимается 5 % уровень значимости, то можно отметить следующее, что между эмпирическим и теоретическим распределением существует лишь хорошее согласие по тем распределениям, где отмечается малая асимметричность, при умеренной и большой асимметричности согласование несколько хуже.

Таблица 5

Критерий согласия А.Н. Колмогорова, λ

Станция	A*s*	Характеристика асимметричности	*n*	*D*, °*C*	λ
Усть-Каменогорск	-0,54	*большая левосторонняя*	69	0,19	1,578
Кокпекты	-0,52	*большая левосторонняя*	54	0,20	1,470
Караганда	-0,33	*умеренная левосторонняя*	67	0,19	1,555
Есик	-0,39	*умеренная левосторонняя*	63	0,20	1,587
Атырау	-0,17	*малая левосторонняя*	68	0,15	1,237
Астана	-0,16	*малая левосторонняя*	65	0,18	1,451
Аральское Море	0,21	*малая правосторонняя*	69	0,18	1,495
Катон-Карагай	0,15	*малая правосторонняя*	65	0,05	0,403
Петропавловск	0,04	*малая правосторонняя*	67	0,15	1,228

Критерий согласия Пирсона (χ^2)

Мера согласованности эмпирического и теоретического рядов по Пирсону рассчитывалась по формуле:

$$\chi^2 = \frac{\sum\left(n_j - \overline{n_j}\right)^2}{n_j}, \qquad (10)$$

где n_j – наблюденные частоты, $\overline{n_j}$ – выровненные частоты.

Из этого выражения следует, что полное совпадение эмпирических и теоретических частот при $\left(n_j - \overline{n_j}\right)^2$ = 0. В противном случае, чем больше эта величина, тем больше расхождение между указанными частотами. Применение критерия $\mathrm{P}\left(\chi^2\right)$связано с определением областей возможных значений χ^2. При этом одна область соответствует значимым χ^2, обусловленным действием каких-либо реально существующих событий, другая же сформирована под влиянием чисто случайных факторов [1]. Применение этого критерия более оправдано для длинных рядов выборочной совокупности, когда ни одна из разрядных частот не будет мала. При нахождении этого критерия важно правильно определить число степеней свободы. При сравнении опытного ряда и теоретического ряда частот, вычисленных по нормальному закону распределения, где параметрами распределения являются среднее и сигма, вычисленные по исследуемому эмпирическому ряду, число степеней свободы равно:

$$g = s - 1 - l, \tag{11}$$

где g – число степеней свободы, l – число дополнительных связей ($\overline{x}$ и σ), s – число градаций.

Сравнивая вычисленный критерий Пирсона χ^2 с $\chi_0^{\ 2}$, который рассчитан для разных уровней значимости с учетом числа степеней свободы [4], можно сделать выводы о случайности или существенности расхождений, так как значения удовлетворяющие условию $\chi^2 > \chi_0^{\ 2}$ имеют достаточно малую вероятность.

Критерий согласия Пирсона по рядам средней годовой максимальной температуры воздуха, χ^2 показал, что на 5 % уровне значимости эмпирическое и теоретическое распределения хорошо согласуются на станциях Усть-Каменогорск, Аральское Море, Петропавловск, где коэффициенты асимметрии находятся в различных пределах (табл. 6). На 1 % уровне значимости эмпирическое и теоретическое распределения хорошо согласуются на станциях Катон-Карагай (малая асимметрия) и на 0,1 % уровне значимости согласуются все станции кроме Кокпекты (большая асимметрия) и Есик (умеренная асимметрия).

На основании этого можно сделать выводы о том, что коэффициент асимметрии эмпирического ряда в данном случае не определяет результат.

Эмпирическое распределение средней годовой максимальной температуры воздуха на станциях Усть-Каменогорск, Караганда, Атырау, Астана, Аральское Море, Катон-Карагай, Петропавловск удовлетворительно согласуется с распределением Гумбеля, а эмпирическое распределение средней годовой максимальной температуры воздуха на станциях Кокпекты, Есик плохо согласуется с теоретическим распределением Гумбеля.

Суточные максимальные осадки

Для оценки аппроксимации эмпирических рядов суточного максимума осадков были произведены такие же действия и в том же порядке, как и по рядам средней максимальной годовой температуры воздуха. В основу анализа был взят материал по 13 станциям, станции выбраны с учетом коэффициента асимметрии (табл. 7).

Таблица 7

Характеристика асимметричности исследуемых станций по суточным максимальным осадкам

Станция	Характеристика асимметричности	A_S	
		градация	величина
Арысь Кзылорда Торгай Туркестан Эмба Мартук Бектауата Астана Петропавловск Усть-Каменогорск Павлодар Алматы	большая положительная	$\lvert A_S \rvert > 0{,}50$	2,88 2,25 2,69 2,92 2,18 0,82 0,92 1,91 1,21 2,50 1,95 1,13
Урджар	умеренная положительная	$0{,}25 < \lvert A_S \rvert \le 0{,}50$	0,32

Временные ряды суточного максимума осадков были упорядочены по возрастающим значениям, вероятность рассчитана по формуле:

$$P_i = \frac{(m_i - 0{,}3)}{(n + 0{,}4)}, \qquad (12)$$

где m_i – порядковый номер члена ранжированного ряда, n – общее число членов.

Расчеты показали, что для суточного максимума осадков характерны большие положительные правосторонние асимметрии (Табл. 8). На 5 % уровне значимости хорошо согласуются центральные части распределения на станциях Кзылорда, Мартук, Бектауата, Астана, Петропавловск, Усть-Каменогорск, Урджар, Алматы. Все остальные станции хорошо согласуются на уровне значимости 1 %, кроме станции Арысь, где выравнивание эмпирического распределения методом Гумбеля не совсем целесообразно и необходим подбор другого теоретического распределения, например методом Джелкинсона, который также аппроксимирует экстремумы [4].

Критерий согласия Пирсона (χ^2)

По критерию Пирсона (табл. 9) можно судить о том, что на 5…10 % уровне значимости хорошо согласуются распределения суточного годового максимума осадков на станциях Мартук, Бектауата, Астана, Петропавловск, Урджар, Алматы. На 1% уровне значимости согласуются распределения на станции Кзылорда, Торгай, Урджар. На 0,1 % уровне значимости согласуются распределения на всех станциях кроме Эмбы. Выравнивание по методу Гумбеля для суточных максимальных осадков отмечает хорошие результаты по многим станциям на высоком уровне значимости (5…10 %).

Таким образом, эмпирические ряды средней максимальной температуры воздуха и суточного максимума осадков показали, что критерий согласия Пирсона χ^2 лучше характеризует хвостовые части распределений, а критерий Колмогорова λ расхождения в центральной части распределения с наибольшими частотами. Совместное использование критериев согласия позволяет сделать вывод, что распределение Гумбеля удовлетворительно аппроксимирует эмпирический ряд средней годовой максимальной температуры воздуха в основном на уровне значимости 1 %, хотя существуют станции с лучшей аппроксимацией и на уровне 5 % значимости. Эмпирический ряд максимальных суточных осадков показывает лучшие результаты аппроксимации на большинстве станций, но есть и исключение – это станции, где применение этого метода менее успешно.

Решение относительно использования результатов аппроксимации распределением Гумбеля временного ряда средней годовой максимальной температуры воздуха и суточных максимальных осадков в каждом конкретном случае следует принимать с учетом поставленной задачи и, соответственно, с необходимой для этого точностью. Эмпирический материал годовых максимумов должен быть однородным (с использованием методов проверки наблюдений на однородность) по каждой станции, так как любое ошибочное значение (особенно завышенное) может привести к неправильным выводам.

Анализ критериев согласия показал, что не ко всякому реальному распределению можно подобрать какую-либо теоретическую функцию с достаточной для решения последующих вопросов точностью. Это объясняется по крайней мере четырьмя причинами [4]. Первая – большое количество факторов, которые формируют эмпирическое распределение; вторая причина – сама аппроксимация – это формальный схематичный процесс подбора теоретического распределения к эмпирическому ряду третья причина – следствие субъективности и неточности измерений, на основе которых делается анализ, четвертая причина – сравнительно малый объем выборочной совокупности наблюдений, используемый для анализа.

Проведенный анализ показал, что эмпирические ряды средней максимальной температуры воздуха можно аппроксимировать распределением Гумбеля, но лучшие результаты оно показывает за период от 10 до 30 лет, а 1 раз в 50 лет величины прогнозируемого элемента завышены, по сравнению с эмпирическим распределением, эта разница составляет:

- при малой асимметричности рядов от 0,5 до 1,0 °С,
- при умеренной и большой асимметричности – более 1,0 °С.

Результат аппроксимации эмпирического распределения суточных максимальных осадков показали в большинстве случаев хороший результат уже на высоком уровне значимости (5…10 %) не только за период 10..30 лет, но и 1 раз в 50…100 лет, что соответствует достаточно редким событиям и представляет собой интерес как вероятностный прогноз за длительный период.

В основе аппроксимации должен лежать временной ряд годового максимума, имеющий достаточно длинный период – это основное требование использование метода, например, в данной работе использованы ряды годовых максимумов за период 1936….2005 гг. При уменьшении объема совокупности различия между эмпирическим и теоретическим распределениями могут увеличиться и соответствие ухудшиться.

В данной работе анализировались в основном результаты приближения эмпирических распределений средней годовой максимальной температуры воздуха и суточного максимума осадков распределением Гумбеля. Рассмотрены итоги аппроксимации на сходимость. Результаты позволяют надеяться, что распределение Гумбеля будет широко применяться в статистических методах прогноза при решении различных практических задач прикладной климатологии.

СПИСОК ЛИТЕРАТУРЫ

1. Кобышева Н.В., Гольберг М.А. Методические указания по статистической обработке метеорологических рядов. – Л.: Гидрометеоиздат, 1990. – 85 с.
2. Общая и прикладная климатология. // Труды ГГО. – 1990. – Вып. 532. – 5 с.
3. Рекомендации по расчету специализированных климатических характеристик / Под ред. Кобышевой Н.В. – СПб.: Гидрометеоиздат, 1997. –77 с.
4. Чичасов Г.Н. Технология долгосрочных прогнозов погоды. – СПб.:: Гидрометеоиздат, 1991. – 304 с.

РГП «Казгидромет»

ОРТАША МАКСИМАЛДЫ АУА ТЕМПЕРАТУРАСЫ ЖӘНЕ ТӘУЛІКТІК МАКСИМАЛДЫ ЖАУЫН-ШАШЫННЫҢ ТАРАЛУЫНЫҢ АППРОКСИМАЦИЯСЫНА ГУМБЕЛЬ ӘДІСІН ҚОЛДАНУ ЖӨНІНДЕ

К.Т. Елеуова

Ықтимал болжауда Гумбель әдісін қолдану мүмкіндігін анықтау үшін орташа максималды ауа температурасы және тәуліктік максималды жауын-шашынның бақылау қатары зерттелді. Эмпиризм қатарының Гумбель теориялық таралуымен үйлесімдігі Пирсон және Колмогоров критерийлері бойынша бағаланды.