

СТОХАСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ АТМОСФЕРНЫХ ОСАДКОВ С ИСПОЛЬЗОВАНИЕМ ТЕОРЕТИЧЕСКИХ И ЭМПИРИЧЕСКИХ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ

А.В. Прусов, В.Л. Посошков

Морской гидрофизический институт
НАН Украины
г. Севастополь, ул. Капитанская, 2
E-mail: vao@alpha.mhi.iuf.net

В статье приводятся результаты моделирования атмосферных осадков суточной дискретности. Используется подход, основанный на цепи Маркова с двумя состояниями. Стохастическая генерация величины суточных сумм осадков производится на основе кумулятивных функций распределения вероятностей. Используются теоретические и эмпирические функции распределения. Методы сравниваются на архивных данных метеостанции Ай-Петри (Крым).

Введение. В течение длительного времени моделирование выпадения осадков было предметом изучения различных исследователей из различных дисциплин, включая климатологию, метеорологию и гидрологию. С точки зрения большого числа аспектов моделирования осадков, задача исследовать моделирование осадков как единое целое, становится невозможной в рамках одной короткой статьи. Одним из аспектов моделирования осадков, которого множество гидрологов касаются на протяжении последних 20 лет, является временная структура выпадения осадков. Из недавних работ хороший обзор содержится в [1, 2]. Обычно две группы случайных переменных входят в описание временной структуры локальных осадков. Первая последовательность касается частоты событий выпадения осадков, в то время как вторая имеет дело с глубиной их (интенсивностью), ассоциированной с данным событием.

Большинство моделей конструируются на основании следующих простых концепций. Процесс погоды рассматривается как цепь Маркова с двумя состояниями – влажные дни с измеряемыми осадками и сухие – без или с незначительным количеством

осадков. Вероятности перехода выводятся из локальных статистических данных. Они дают основу для генерации последовательности сухих и влажных суток. После этого независимо модулируется случайная последовательность суточных сумм осадков (интенсивность) во влажные сутки. Вследствие годичной периодичности погоды вероятности перехода из состояния в состояние, а также распределение интенсивности зависят от времени года.

Стохастические генераторы осадков являются локальными моделями, поэтому региональная настройка генератора производится на данных наблюдений в определенной местности. Для моделирования интенсивности осадков, как правило, используются теоретические функции распределения, при этом адекватность тех или иных теоретических кривых также зависит от локальных условий. В работе [3] рассматривалась применимость некоторых теоретических распределений к условиям Крыма, каждое из которых имеет свои преимущества и недостатки. В данной работе продолжена настройка генератора на условия Крымского региона. При этом наряду с использованием хорошо известного двухпараметрического Гамма распределения, для генерации интенсивности осадков применяется и непараметрический метод, основанный на эмпирической функции распределения вероятностей осадков, определяющейся непосредственно из данных наблюдений.

Целью настоящей работы является сравнение двух методов стохастического моделирования, а именно: на основе теоретических интегральных функций распределения и с использованием фактических выборочных функций, полученных непосредственно из наблюдений.

Материалы и методика. В работе использовался долговременный ряд наблюдений за ежесуточными осадками в горном массиве Крыма (Ай-Петри). Осадки суточной дискретности в этом регионе представляют особый интерес, так как отличаются большой интенсивностью. К тому же они наиболее полно документированы (начиная с 12 сентября 1895 г.). Весь массив ежесуточных наблюдений охватывает период 1896-2002 гг. В массиве данных имеются разрывы, связанные как с годами войны, так и с другими причинами. Можно выделить 4 периода, в течение которых отсутст-

вуют разрывы в наблюдениях. Для настройки генератора осадков выбран самый длинный, послевоенный период: 1945 – 1976 гг.

В работе используется вариант стохастического генератора осадков, описанный в работе [3]. Как и в большинстве имеющихся моделей, генерирование осадков содержит две компоненты: процесс чередования (последовательность сухих и влажных суток), и процесс интенсивности (распределение сумм осадков во влажных сутках). Влажные сутки определяются как те, когда измеренное количество осадков превышает некоторое пороговое значение (u нас – 0,1 мм).

Было предложено множество статистических распределений для моделирования процесса интенсивности, иными словами, распределения количества выпадающих осадков во влажные сутки. В данной работе используются два подхода к моделированию интенсивности осадков. Первый из них состоит в использовании теоретического распределения вероятностей осадков. Во втором случае для моделирования вероятности выпадения определенного количества осадков используется эмпирическая гистограмма распределения ненулевых осадков по интенсивности. В качестве теоретического распределения принято Гамма распределение.

Гамма распределение. Широкое применение в мировой практике Гамма распределения связано с использованием минимального числа легко интерпретируемых параметров при достаточно хорошей аппроксимации данных наблюдений. В ряде работ для моделирования величин осадков в изучаемом регионе используются и другие специальные законы распределений. Гамма распределение является хорошим выбором для представления величины осадков по ряду причин.

Первое преимущество состоит в том, что оно ограничено слева нулем. Это очень важно в приложениях к осадкам, так как отрицательные величины осадков невозможны, и таким образом, распределение, исключающее отрицательные величины, без труда применимо.

Во-вторых, Гамма распределение положительно асимметрично, что означает, что оно обладает протяженным хвостом в правой части распределения. Это преимущест-

во сказывается в том, что оно имитирует фактические распределения осадков для многих географических положений, где существует ненулевая вероятность экстремально высоких количеств осадков, хотя типичные выпадающие осадки могут быть не очень большими.

Наконец, Гамма распределение предлагает огромное количество возможностей в представлении формы функции распределения. Это распределение может простирается по форме, начиная с экспоненциального затухания с параметром формы порядка единицы до нормального распределения с параметром формы свыше 20-ти. Такая гибкость позволяет подгонять под Гамма распределение любое число режимов выпадения осадков с разумной точностью, в то время как под другие распределения могут быть подогнаны лишь отдельные специфические режимы.

Гамма распределение описывается законом:

$$f(x) = \frac{(x)^{\nu-1} \lambda^{\nu} e^{-\lambda x}}{\Gamma(\nu)} \quad (1)$$

$$\Gamma(\nu) = \int_0^{\infty} e^{-t} t^{\nu-1} dt \quad (2)$$

В уравнении (1) ν называют параметром формы, λ – параметром масштаба, а x представляет собой количество осадков. $\Gamma(\nu)$ – полная Гамма функция. Параметры распределения оцениваются из наблюдений. Универсальным является метод моментов. Однако получаемые с его помощью оценки лишь в редких случаях обладают оптимальными свойствами. Поэтому в прикладной статистике применяют и другие виды оценок. В данном исследовании наряду с методом моментов используется метод оценок максимального правдоподобия (ОМП).

Моделирование с использованием эмпирической интегральной функции распределения. Обозначим через x_{\max} и x_{\min} максимальное и минимальное значения рассматриваемого диапазона изменения климатического параметра (в нашем случае – величины среднесуточных осадков), а через N – количество измерений этого параметра. Диапазон измерений рассматриваемого параметра разбиваем на n равных частей

$[x_{\min}, x_{\max}] = \bigcup_{j=1}^n [x_j, x_{j+1}]$ и подсчитываем количество измерений параметра (m_j), попавших в каждый интервал этого разбиения. По значениям m_j получаем частоты попадания в соответствующие интервалы

$$p_j = m_j / N. \quad (3)$$

Совокупность интервалов $[x_j, x_{j+1}]$ и соответствующие им частоты p_j определяют статистический ряд изучаемого климатического параметра.

На основе полученных статистических рядов ежедневных осадков можно определить функции распределения для их регулярных значений. Очевидно, $\sum_{j=1}^n p_j = 1$. Зададим эмпирические функции распределения для атмосферных осадков в граничных точках интервалов $[x_j, x_{j+1}]$, которые фигурируют в статистических рядах, следующим образом

$$\begin{aligned} F(x_1) &= 0, \\ F(x_j) &= a_j, j = \overline{2, n-1}, \\ F_n &= 1, \end{aligned} \quad (4)$$

где $a_j = \sum_{k=1}^j p_{k-1}$. Соединяя соседние полученные точки прямыми линиями, получим приближенную эмпирическую функцию распределения. Таким образом, полученные в нашем исследовании функции распределения имеют вид:

$$\begin{aligned} F(x) &= 0, & x < x_1, \\ F(x) &= p_{j-1} \frac{x - x_{j-1}}{x_j - x_{j-1}} + a_{j-1}, & x_{j-1} \leq x < x_j, \\ F(x) &= 1, & x \geq x_n. \end{aligned} \quad (5)$$

Следовательно, задача сводится к моделированию случайных величин, удовлетворяющих функции распределения (5). Для этой цели мы используем метод обратных функций [4].

Согласно этому методу, последовательность случайных чисел x с функцией распределения (5) можно получить из последовательности случайных чисел y с равномерным распределением в интервале (0,1) по формуле

$$\begin{aligned} x = F^{-1}(y) &= (x_j - x_{j-1})(y - a_{j-1}) / p_{j-1} + x_{j-1}, \\ &\text{при } a_{j-1} \leq y < a_j \end{aligned} \quad (6)$$

В работе [4] используется равномерное разбиение диапазона изменений рассматриваемого метеорологического параметра на интервалы. Так как функция распределения ежедневных атмосферных осадков близка к экспоненциальной, это приводит к неравномерным интервалам по оси ординат приближенной кусочно-линейной кумулятивной функции распределения вероятностей. Это наглядно представлено на рисунке 1, построенном на основании январских данных Лй-Петри за период 1945-1976 гг. (всего 992 измерения).

На рисунке приводятся фактические интегральные функции распределения вероятностей, приближенные кусочно-линейные, в случаях равномерного и неравномерного (путем растяжения величины интервалов по экспоненциальному закону) разбиения на 5 классов и модельные, основанные на кусочно-линейном приближении.

Даже при столь малом числе классов (карманов) заметно, что экспоненциальное растяжение интервала разбиения приводит к достаточно адекватной аппроксимации эмпирической функции распределения приближенной функцией, в то время как при равномерном разбиении различие между фактической функцией и кусочно-линейной остается большим.

При малом числе классов различие в воспроизведении функции распределения кусочно-линейной функцией при разных способах разбиения на классы – существенно. При увеличении числа классов (а это возможно при достаточно большой величине случайной выборки) это различие нивелируется. Следовательно, применение неравномерного разбиения на классы, или карманы, особенно важно при нехватке экспериментальных данных.

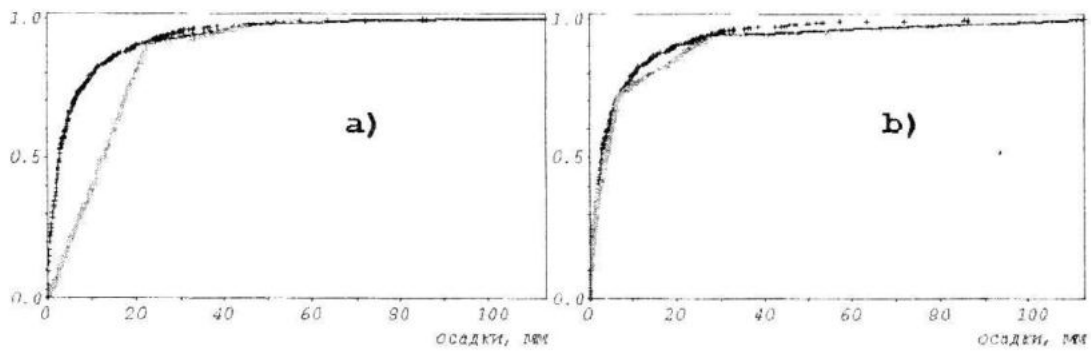


Рисунок 1 – Интегральная функция распределения осадков при разбиении на 5 классов. В случае равномерного (а) и неравномерного (б) разбиения на классы. + – величины из наблюдений; о – модельные; отрезки прямых – кусочно-линейная аппроксимация функции распределения

Обсуждение результатов. Посмотрим, насколько адекватно Гамма распределение применимо в локальных условиях. Ограничимся выборками данных наблюдений за три зимних месяца: декабрь – февраль и летних: июнь – август. Размеры выборок составляют около 3000 наблюдений. В зимнее время более половины из них приходится на дни с ненулевыми суммами осадков. Если принять за оптимальное чис-

ло карманов цифру порядка корня квадратного из размера выборки, то получим число 40 как приемлемое количество классов для построения гистограмм и эмпирической функции распределения вероятностей.

На рисунке 2 показана кумулятивная функция распределения вероятности осадков, полученная на основании данных наблюдений в зимнее (а) и летнее (б) время.

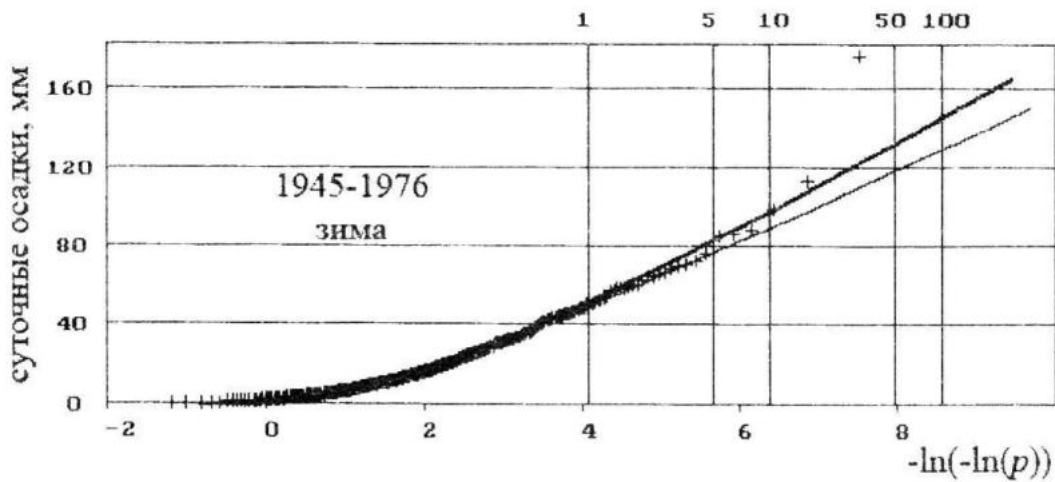


Рисунок 2а – Кумулятивная функция распределения вероятности осадков (зима). + – полученная на основании данных наблюдений, сплошные линии – Гамма распределение: жирная линия построена на основе ОМП, тонкая – метода моментов

Кумулятивная, или интегральная функция распределения случайной величины X определяется как вероятность

$$P(x) = P\{X < x\} \quad (7)$$

На нижней шкале рисунка 2 отложен двойной логарифм функции P . На вертикальной оси отложены суточные суммы осадков в мм. Вертикальные линии на рисунке обозначают ожидаемый период T соответствующих событий x , который связан с функцией $P(x)$ соотношением

$$T(x) = 1/(1 - P(x)) \quad (8)$$

Сплошные кривые соответствуют Гамма распределению, причем теоретические кри-

вые приводятся в двух вариантах, соответствующих методике определения параметров распределения. Жирная кривая построена на основе метода ОМП, тонкая – метода моментов.

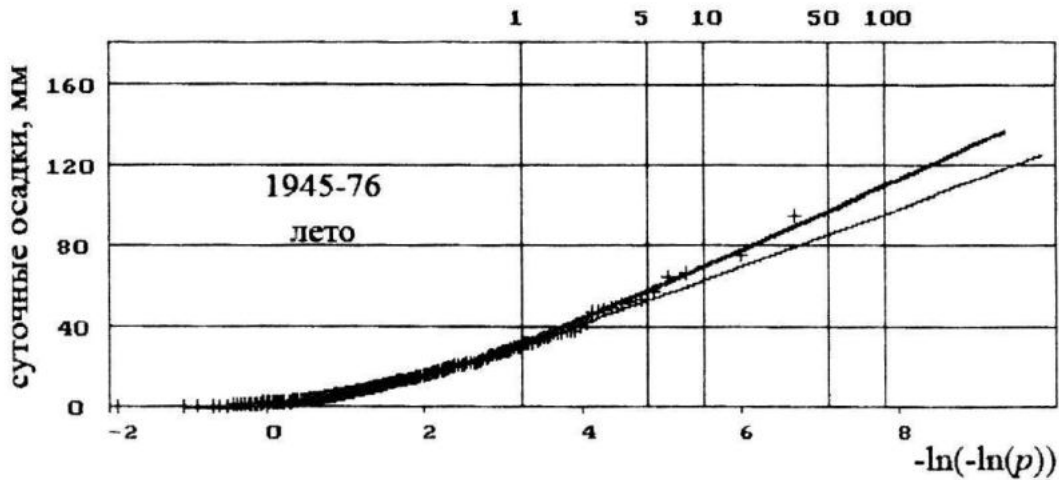


Рисунок 26 – Кумулятивная функция распределения вероятности осадков (лето).
Обозначения, как и на рисунке 2а

Адекватность подгонки данных к теоретическому Гамма распределению принципиально не отличается в случаях зимних и летних выборок. Различие состоит лишь в величине максимальных суточных сумм в зимнее и летнее время.

Для зимней выборки метод моментов лучше работает до уровня осадков, примерно 60 мм в сутки (ожидаемый период 5 лет). Данные с осадками, выше этой величины лучше подгоняются под кривую распределения, полученную методом ОМП, вплоть до частоты повторения 1 раз в 10 лет (больше 80 мм в сутки). Фактически, данные достаточно хорошо подгоняются под Гамма распределение почти во всем диапазоне изменчивости осадков, за исключением одного экстремального события величиной 176 мм. Это редкое событие произошло 10 января 1967 г. Заметим, что такая суточная величина превышает среднемесячную норму в январе (около 150 мм). Столь высокую величину можно рассматривать либо как случайный выброс во временном ряде суточных осадков, либо как ошибку наблюдения. Чтобы избежать ложных выводов, в дальнейшем анализе это значение будет удалено из выборки данных.

В летней выборке эффект тот же, но цифры другие. Распределение осадков ве-

личной вплоть до 50 мм (ожидаемый период примерно 2 года) хорошо описывается кривой, рассчитанной с помощью метода моментов, в то время как для осадков большей интенсивности, вплоть до экстремальных, лучше работает метод ОМП. В целом можно сказать, что Гамма распределение достаточно хорошо описывает распределение вероятностей летних осадков.

Из приведенного рисунка можно предположить, что как в зимнее, так и летнее время метод моментов лучше описывает статистику осадков, которые случаются чаще, чем один раз в пять лет. Для описания более редких (и, соответственно, более крупных) событий лучше работает метод ОМП. В данном исследовании упор делается на обильные осадки, наиболее значимые в гидрологических приложениях, поэтому всюду в дальнейшем, не оговаривая особо, будут использованы параметры Гамма распределения, рассчитанные на основе ОМП.

Моделирование на основе Гамма распределения. На рисунке 3 демонстрируются результаты моделирования осадков в зимние месяцы на основе Гамма распределения. На рисунке отсутствует событие величиной 176 мм, о чем говорилось выше. Символы на рисунке относятся к наблюдениям и модельным величинам. Сплошные

линии отвечают соответствующим аппроксимациям Гамма распределением этих величин. Можно считать, что генерация модельных осадков адекватна наблюдениям вплоть до величин порядка 50 мм в сутки (ожидаемый период примерно 1 год). В

диапазоне более редких событий с ожидаемыми периодами от одного до 5 лет заметно расхождение между экспериментальными и расчетными величинами в сторону занижения моделью фактической частоты повторяемости событий

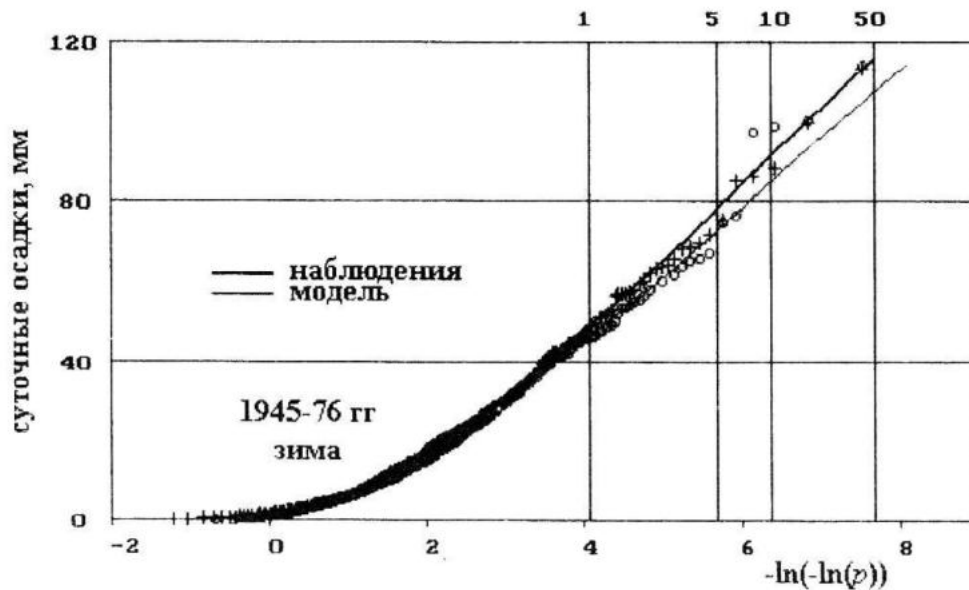


Рисунок 3 – Кумулятивная функция распределения осадков в зимние месяцы: (+) – из наблюдений, (o) – с выхода генератора осадков, сплошные линии – соответствующие аппроксимации теоретическим Гамма распределением (жирная – наблюдения, тонкая – модель).

Остальные обозначения – как на рисунке 2

Моделирование интенсивности осадков носит локальный характер. В разном географическом местоположении под данные наблюдений лучше подгоняются те или иные виды теоретических распределений. Поэтому улучшение аппроксимации фактических функций распределения обычно связывают с поиском других видов распределений.

Альтернативный подход состоит в отказе аппроксимаций на основе теоретических функций распределения и переходе к стохастическому генерированию интенсивности суточных осадков на основе эмпирической кумулятивной функции распределения. В данной работе исследованы оба подхода.

Моделирование на основе эмпирических функций распределения. На рисунке 4 представлены результаты моделирования синтетического ряда осадков на основе эмпирических кумулятивных функций распределения, рассчитанных непосредственно из фактических рядов наблюдений на основании

алгоритма, описанного в методической части работы. В качестве натуральных рядов использовалась та же выборка наблюдений, что и при моделировании с помощью Гамма распределения. Результаты приводятся в виде кумулятивных функций распределения аналогично рисунку 3.

Из рисунка видно, что адекватность модели наблюдениям по-разному проявляется в различных интервалах всего диапазона интенсивности осадков. Рассмотрим в отдельности участки функции распределения в области калиброванных значений ожидаемого периода осадков определенной интенсивности (вертикальные линии на рисунке 4).

Как и на рисунке 3, наилучшее согласие с наблюдениями достигается в интервале суточных осадков до величины 50 мм (ожидаемый период 1 год и менее) Существенное расхождение (в сторону занижения вероятных осадков моделью) наблюдается на рисунке в области ожидаемого периода 5 лет. В области наиболее крупных событий

осадков с вероятными периодами повторения 5 и 50 лет согласно модели с наблюдениями можно считать удовлетворительным.

Итак, для получения климатических рядов ежесуточных осадков в работе использованы два метода моделирования интенсивности осадков. Один из них состоит в аппроксимации интегральной функции распределения осадков одним из применяемых в мировой практике теоретическим распределением, а именно: Гамма распределением. Второй метод базируется на использовании приближенной кусочно-линейной интегральной функции распределения, рассчитанной на основе фактических наблюдений. Сравнение результатов на уровне

кумулятивных функций дает неоднозначные выводы о преимуществе того или иного метода. В зависимости от прикладной тематики, упор может делаться на различные интервалы гистограммы интенсивности суточных осадков. Одним из критериев адекватности модели могут служить интегральные статистические характеристики осадков, такие, как месячные суммы, средняя величина, дисперсия и т.п. Ни одна модель не способна воспроизвести все эти характеристики с одинаковой точностью, и в зависимости от характера приложений; та или иная характеристика может оказаться наиболее важной.

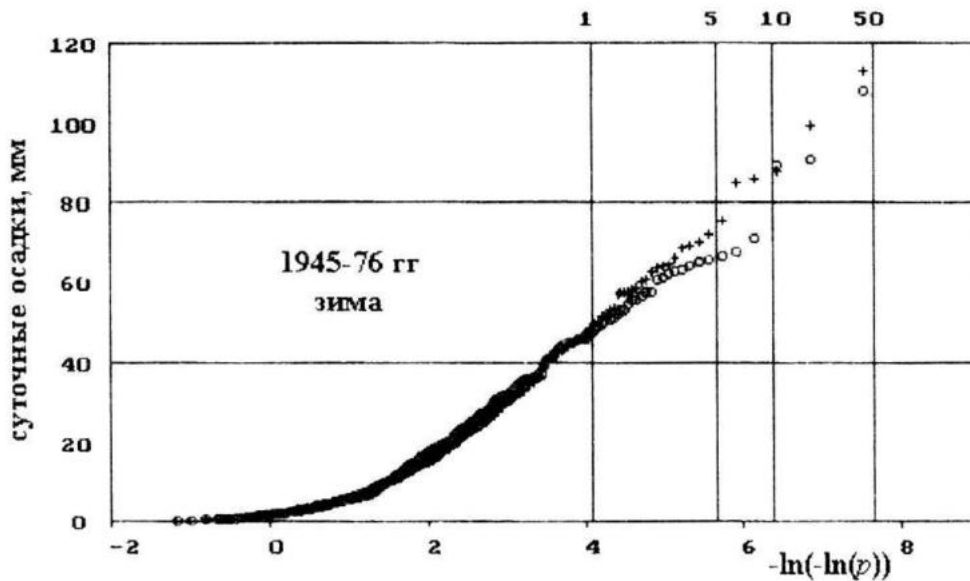


Рисунок 4 – Кумулятивная функция распределения вероятностей осадков (зима). (+) - на основании натуральных рядов наблюдений, (o) - в результате моделирования

Наиболее важным критерием при оценке адекватности модели представляется ее способность достаточно точно воспроизводить среднемесячные суммы. В качестве примера, на рисунке 5 по этому показателю сравниваются климатические месячные суммы осадков (во все месяцы года), рассчитанные из натуральных данных и выхода различных вариантов генератора осадков.

Для сравнения используются три модели, различающиеся способом генерирования интенсивности осадков. Наряду и использующимися в данной работе двумя ме-

тодами привлечены дополнительно данные моделирования из работы [5], в которой для генерирования интенсивности осадков применялось экспоненциальное распределение.

Из приведенного рисунка хорошо видно, что все рассматриваемые виды распределений осадков по интенсивности одинаково хорошо воспроизводят месячные суммы выпадающих осадков не только в зимнее время, но и во все остальные сезоны года.

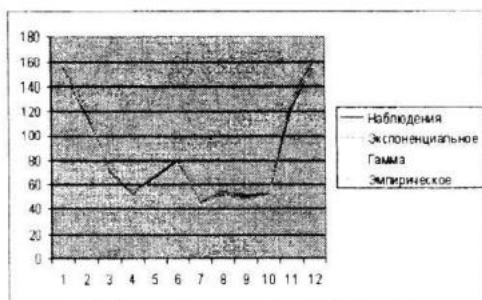


Рисунок 5 – Среднемесячные суммы осадков по данным наблюдений и результатам моделирования. Горизонтальная ось – номер месяца года, вертикальная – сумма осадков, мм

Заключение. Моделирование климатических рядов атмосферных осадков неограниченной длительности имеет большое прикладное значение, в частности в комплексном исследовании гидрологического цикла в бассейне речного стока Крыма, для тех мест, где недоступны натурные данные суточного масштаба. Различные типы генератора осадков, базирующиеся на цепном процессе Маркова, преимущественно различаются видом функции распределения вероятностей, на основе которой генерируется последовательность величин суточных осадков. Оптимальный выбор функций производится путем настройки генератора на локальные условия.

Для выбора оптимальной модели при стохастическом моделировании интенсивности осадков в данной работе использовались два вида функции распределения вероятностей осадков. Наряду с функцией, основанной на теоретическом Гамма распределении, рассмотрено применение приближенной эмпирической кумулятивной функции распределения, рассчитанной непосредственно на основе наблюдений.

В результате применения модели выполнен сравнительный статистический анализ натуральных и модельных временных рядов. Статистика синтетических временных рядов оказалась близкой к статистике, полученной по данным натуральных наблюдений. При этом варианты моделирования на основе Гамма распределения интенсивности ненулевых осадков, равно как и на ос-

нове эмпирического – можно считать равноценными.

Использование той или иной функции связано с конечной целью, вытекающей из прикладного значения. Статистика, основанная на использовании эмпирической кумулятивной функции распределения, не зависит от степени адекватности подгонки данных к теоретическому распределению. Поэтому она свободна от тех погрешностей, которые вытекают из отклонений фактического распределения от теоретического. Вместе с тем, с учетом ограниченности размера выборки данных, используемой для определения параметров модели, эмпирическое распределение не способно моделировать крупные по величине суммы осадков, отсутствующие в исходной выборке. В связи с этим использование Гамма распределения, допускающего возможность экстраполяции, следует считать предпочтительнее при моделировании статистики редких, но значительных (катастрофических) по величине событий осадков.

Литература

1. M. Dubrovsky, J. Buchhtele and Z. Zalud. High-Frequency and low-frequency variability in stochastic daily weather generator and its effect on agricultural and hydrologic modelling. *Climatic Change*, 63, 2004. – P. 145–179.
2. Y. Liao, Q. Zhang, and D. Chen. Stochastic modeling of daily precipitation in China. *J. Geogr. Sci.*, 14(4), 2004. – P. 417–426.
3. В.Л. Посошков, А.В. Прусков. Применение стохастического генератора осадков для моделирования рядов суточной дискретности Системы контроля окружающей среды. – Севастополь, МГИ НАНУ, 2007. – С. 220–224.
4. Л.Л. Голубятников. Стохастическое моделирование величин ежедневных осадков и среднесуточных температур. *Известия РАН. Физика атмосферы и океана*, т. 40, № 5, 2004. – С. 665–677.
5. В.Л. Посошков, А.В. Прусков. Практика применения стохастического генератора осадков к условиям Крымских гор Системы контроля окружающей среды. – Севастополь, МГИ НАНУ, 2007. – С. 225–228.