

ISSN 0130-2906

МЕТЕОРОЛОГИЯ И ГИДРОЛОГИЯ

2014 № 12

ФГБУ "Научно-исследовательский центр
космической гидрометеорологии "Планета"

Несмещенная оценка статистической энтропии гидрологических рядов

С. К. Давлетгалиев*

Исследована зависимость энтропии рядов годового стока от числа лет наблюдений. Построены номограммы для оценки несмещенного значения энтропии в зависимости от числа лет наблюдений и числа градаций. Номограммы построены для распределения Пирсона III типа по смоделированным рядам.

Речной сток в заданный момент времени, как и всякая случайная величина, характеризуется числом возможных состояний и распределением вероятностей этих состояний. Однако эти величины, взятые в отдельности, хотя и могут отражать степень неопределенности наших наблюдений, но не дают ее количественной меры, особенно в сложных случаях. Такая количественная мера в теории информации называется энтропией. Для случайных дискретных величин она имеет следующий вид [9, 10]:

$$H(\chi) = - \sum_{i=1}^n P(\chi_i) \ln P(\chi_i). \quad (1)$$

Как видно из (1), энтропия определяется числом возможных состояний n и распределением вероятностей $P(\chi_i)$, но не множеством конкретных значений наблюдаемых явлений. В этой формуле вместо неизвестной априорной вероятности P_i обычно используют ее оценку $P_i = m_i/N$ ($i = 1, 2, \dots, S$), где m_i — частота исхода χ_i в выборке объема N . Такая статистическая оценка используется в разных приложениях, например, в гидрологии и метеорологии для решения ряда практических задач [1, 3, 5, 11, 12]. Формулу (1) используют также при моделировании и прогнозировании сложных систем в экономике, транспорте ([6] и др.).

Величина $H(\chi)$, вычисленная по данным выборочных наблюдений, является смещенной величиной. Смещенность энтропии, вычисленной по формуле (1) для независимых нормально распределенных случайных величин, показана в работе [2]. Согласно [2], величина $H(\chi)$ является смещенной, состоятельной, асимптотически нормальной оценкой энтропии, и среднее значение ее имеет следующий вид:

$$M[H] = H(\chi) - \frac{S-1}{2N} \lg_2 e + o\left(\frac{1}{N^2}\right), \quad (2)$$

где S — число градаций; N — объем выборки.

* Казахский национальный университет им. Аль-Фараби; e-mail: sdavletgaliev@mail.ru.

В данной работе рассматриваются значения энтропии асимметрично распределенных величин. Задачу можно решить аналогичным образом, используя формулу [8] для оценки точности параметризации методом статистических испытаний (метод Монте-Карло) для различных типов кривой распределения.

Расчеты выполнены с учетом параметров гидрологических рядов: $CSIC^V = 1, 2, 3, 4$ и $C, = 0$), при $r = 0, 0.3, 0.5$. С моделированием коэффициентов автокорреляции ($r = 0; 0.3; 0.5$). С моделированием разбивается на выборки меньшего объема $n = N/n$.

Моделирование годового стока с помощью математического аппарата нормальных случайных последовательностей, удовлетворяющих заданному закону распределения.

Для статистического моделирования используется метод канонического разложения.

Q_i

где Q_i — математическое ожидание некоррелированных величин, равное нулю; φ — неизвестные функции.

Неизвестные функции φ и Q_i определяются по данным наблюдений с помощью метода наименьших квадратов. Случайные величины V^i определяются по формуле (1) значения Q_i с помощью стандартных функций. Переход от вероятности к функции Q_i для заданных значений Q_i Крицкого — Менкеля путем интегрирования коэффициента стока. Для этой цели используется алгоритм моделирования.

Результаты статистического моделирования канонического разложения. Последний почти полностью совпадает с кривой распределения исходных данных.

По смоделированным значениям интервала определения интервала и его вероятность попадания вычисляются значения энтропии. Результаты вычисления энтропии являются как среднее из выборок. Число градаций $S = 6, 8, 10$ и 12 .

Результаты расчетов показывают, что энтропия по мере увеличения S увеличивается.