

Киндюк Б.В.

СТАТИСТИЧЕСКАЯ ЭНТРОПИЯ КАК МЕРА НЕОПРЕДЕЛЁННОСТИ И СВЯЗНОСТИ СТРУКТУРЫ ГИДРОГРАФИЧЕСКОЙ СЕТИ

Цель работы – изучить строение и конфигурацию речной сети как совокупности случайных фактов с определенной вероятностью проявления. С этой целью к описанию системы водотоков, протекающих по территории речного бассейна, применяется математический аппарат теории информации. Это связано в первую очередь с тем обстоятельством, что конфигурация водотоков, их распределение по территории – это случайный процесс, у которого имеются элементы неопределённости.

В ранее опубликованных исследованиях рассматривались различные аспекты теории информатики, применительно к задачам метеорологии [1], сельского хозяйства, географии [6,7]. Относительно структуры речных водотоков такой математический аппарат применялся относительно рек Дальнего Востока [3]. Территория Украины оказалась вне поля зрения ученых, занимающихся этой проблемой. Кроме этого существует и временной фактор. Дело в том, что с момента опубликования этих работ прошло более 30-ти лет и возникла необходимость в их дальнейшем продолжении.

Данное исследование является частью более общей работы, посвящённой изучению строения гидрографической сети рек Украинских Карпат. Актуальность этой проблемы резко возросла в последнее время, т.к. этот регион всё больше вовлекается в хозяйственную деятельность. Другим аспектом проблемы является постоянное прохождение на его реках катастрофических ливневых паводков. Ущерб, который они наносят народному хозяйству региона, исчисляется суммами в десятки миллионов гривен, а изучить формирование этих природных феноменов без учета особенностей строения гидрографической сети практически не возможно.

Фактический материал и методы исследования. В качестве исходных данных использовались материалы о гидрографическом строении рек Закарпатья, справочная литература, различные пособия, картографические данные.

Распределение водотоков по территории можно рассматривать как систему равновероятных событий, при неопределённости наших знаний о наступлении (проявлении) каждого конкретного события из них. Полной системой событий X_1, X_2, \dots, X_n называется группа взаимно несовместимых событий, сумма вероятностей которых равна единице. Полная система событий X будет статистически задана, если для каждого события X_k задана вероятность его наступления P_k . Эту полную систему символически можно записать в таком виде:

$$X = \begin{pmatrix} X_1 X_2 \dots X_n \\ P_1 P_2 \dots P_n \end{pmatrix} \quad (1)$$

Теория информации, в качестве меры неопределённости наших знаний о полной системе, использует следующую формулу:

$$H(X) = H(P_1, P_2, \dots, P_n) = -\sum_{k=1}^n P_k \lg P_k, \quad (2)$$

где $H(X)$ – значение энтропии.

Логарифм в выражении (2) может быть взят при любом основании, а вероятности P_k подчинены условию:

$$\sum_{k=1}^n P_k = 1 \quad (3)$$

где n – число членов ряда.

Выражение (2) получило название статистической энтропии полной системы. Исторически оно заимствовано из классической физики, где аналогичное выражение по предложению Больцмана, употребляется для характеристики термодинамических систем. Прежде чем применить эту формулу для подсчёта энтропии речных систем, покажем её практическое применение для ряда абстрактных задач. Возьмём известный пример с бросанием монеты. Здесь имеется простейшая альтернатива. Может выпасть решетка или герб, т.е. система состоит из двух равновероятных событий. Энтропия такой системы по формуле (2) будет подсчитана таким образом:

$$H\left(\frac{1}{2}, \frac{1}{2}\right) = -\ln \frac{1}{2} = \ln 2 = 0,6939,$$

где \ln – натуральный логарифм.

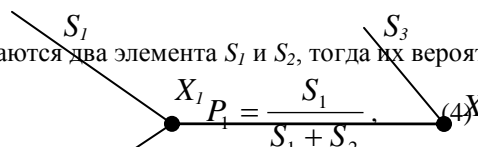
Эта теоретическая база может быть использована при описании распределения гидрографических характеристик на территории речного водосбора. По теории Р. Е. Хортонa система водотоков представляет собой множество соподчиненных элементов, которые находятся между собой в определённых количественных соотношениях [2,3]. Для введения системы топологии в пространство, заполненное различными водотоками, необходимо выделить в нем простейшие элементы. Говоря иным языком, нужна топологическая характеристика, которая будет своеобразным метрическим шагом или мерой системы. При этом она

должна удовлетворять двум свойствам. Первое свойство – это не поддаваться дальнейшему делению (разрезанию, уменьшению), т.к. иначе будет потеряна её качественная определённость. Из теории информационных систем [1,7] известно, что такие элементы формируют открытые множества. Второе свойство состоит в возможности составить из элементов открытых множеств, по определённым правилам, произвольную систему водотоков. Исходя из этого, можно ввести понятие идеализированной системы речных водотоков. Это означает две вещи: первое – система не включает в себя никаких других объектов, кроме водотоков. Условно считается, что отсутствуют озера, болота, старицы. Второе – в системе нет точек, в которых сливается больше двух водотоков, т.е. её любые две точки соединяются единственным образом. Тем самым, из рассмотрения исключаются сложные дельтовые образования и те редкие случаи, когда в одной точке сходятся более трех потоков. Если смотреть на систему против течения, то внешние элементы заканчиваются истоком, а внутренние – разветвлением. Тогда по схемам Хортон, Шайдеггера, Ржаницына [2,4] к внешним элементам относятся только притоки первого порядка P_1 , а к внутренним – элементы более высоких порядков ($P_i > 1$). Под притоками первого уровня иерархии P_1 понимаются элементарные неразветвленные потоки, второй порядок образуется от слияния двух P_1 .

В соответствии с изложенными в пространстве сети потоков можно выделить два открытых множества, представляющих его внешний (D) и внутренний (T) отделы. Первый из них включает в себя только элементы P_1 , а отдел T – только $P_i > 1$. Элементы из T могут быть образованы тремя различными способами: первый из них – путём слияния двух внешних элементов P_1 , тогда новый элемент приобретёт порядок $k = 2$. Второй вариант возникает в случае слияния элемента внутреннего отдела с элементами внешнего, тогда новый элемент из множества T примет порядок $k > 2$. Третий случай может быть, если сливаются два элемента внутреннего отдела, тогда вновь образованный поток имеет порядок $k \geq 3$. Таким образом, внутренний отдел является упорядоченной структурой и состоит из множества подотделов, каждый из которых включает элементы определённого порядка.

Речные системы отличаются большим разнообразием, поэтому в отношении оценки их количественной структуры можно рассматривать имеющуюся в них степень разнообразия или упорядоченности. Такую количественную характеристику структуры водотоков можно получить, если использовать введенную Шенноном [1,7] для марковских цепей меру количества разнообразия, называемую энтропией. Применение к описанию речных систем этого понятия является вполне допустимым, т.к. структуры речных систем эквивалентны марковским цепям. Для практических расчетов наиболее удобно в качестве меры мощности звеньев речной сети использовать количество притоков первого порядка P_1 . Гидрографическая сеть может рассматриваться как система графов, образующих определённое дерево элементов.

Пусть в точке X_1 сливаются два элемента S_1 и S_2 , тогда их вероятностная мера P_1 представляет собой:



$$P_1 = \frac{S_1}{S_1 + S_2}$$

аналогично вероятностная мера P_2 определяется по формуле:

$$P_2 = \frac{S_2}{S_1 + S_2}, \quad (5)$$

Тогда по определению Шеннона энтропия такой системы рассчитывается по формуле:

$$H_{узла} = -P_1 \log P_1 - P_2 \log P_2 = -(P_1 \log P_1 + P_2 \log P_2), \quad (6)$$

Если речная сеть состоит из более сложной дизъюнкции различных графов, то её общая энтропия определяется суммированием по всем точкам слияния. Например, энтропия в точке X_3 после добавления звена S_3 рассчитывается по формуле:

$$H_{x_2} = -\frac{S_1 + S_2}{S_1 + S_2 + S_3} \log \frac{S_1 + S_2}{S_1 + S_2 + S_3} - \frac{S_3}{S_1 + S_2 + S_3} \log \frac{S_3}{S_1 + S_2 + S_3} \quad (7)$$

где H_{x_2} – суммарная энтропия в точке X_2 .

В качестве практического примера выполнены расчёты значений энтропии для бассейна р. Рики с замыкающим створом г. Хуст. Выбор этого водосбора является не случайным, т.к. на нём находится уникальный исследовательский гидрометеорологический полигон – Закарпатская воднобалансовая станция (ЗВБС). Река Рика является притоком Тисы, впадает в неё у г. Хуст, водосбор находится в типичном горном районе Украинских Карпат. Характеризуется сильно развитой гидрографической сетью с большим числом различных притоков. В качестве подготовительного этапа по решению этой задачи выполнена идентификация всей сети, подсчитано количество притоков всех порядков, детальная информация имеется в работе [4]. Далее, используя формулы (4)–(6), выполнялся расчет энтропии по каждой реке до замыкающего створа, где имеются данные гидрологических наблюдений. Данные по расчетам энтропии и по гидрографическим характеристикам исследуемых водотоков приводятся в таблице 1. Располагая такой информацией, не сложно построить графики связи между суммарной энтропией H и длиной рек $L = f(H)$, площадями водосборов $F = f(H)$, уклоном $I = f(H)$ и водоносностью \bar{Q}_{max} (рис. 1, 2). В качестве характеристики последней принимались осреднённые за многолетний период наивысшие срочные максимальные расходы

ливневых паводков. Степень тесноты связи между аргументом и функцией обычно характеризуется коэффициентом корреляции.

С целью оценки точности построенных трёх графиков зависимостей между энтропией и четырьмя аргументами, по каждому из них подсчитывался коэффициент корреляции R . Его величина для функции $F = f(H)$ оказалась равной 0,976; у функции $L = f(H)$ $R = 0,919$; график связи $I = f(H)$ показал $R = -0,695$ и у зависимости $\bar{Q}_{\max} = f(H)$ $R = 0,947$.

Численные значения коэффициентов корреляции позволяет положительно оценить полученные результаты и рекомендовать их для рек с ограниченной информацией по стоку воды. Такие задачи часто возникают на практике, когда требуется построить гидротехническое сооружение, а исходных данных или нет, или они очень ограничены. В таком случае, подсчитав значение энтропии с помощью картографического материала, по зависимости $Q = f(H)$ (рис. 1) можно определить среднегодовое значение максимальных расходов ливневых паводков на реках бассейна Рики.

Таблица 1 - Основные параметры гидрографической сети малых рек бассейна реки Рика

Река – пункт	Площадь водосбора F км ²	Длина реки L км	Уклон русла реки I ‰	Средняя высота водосбора H м	Средний максимальный расход ливневых паводков \bar{Q}_{\max} м ³ /с	Кол-во притоков первого порядка S_1	Энтропия речной системы $H_{\text{энт}}$
1	2	3	4	5	6	7	8
Рика – с. Верхний Быстрый	165	15	40,0	23,4	52,9	28	23,38
Рика – пгт Межгорье	550	28	24,3	71,5	222	101	71,5
Лопушна – с. Лопушное (ниж)	37,3	9,4	63,2	6,23	13,4	11	6,28
Бранище – с. Лопушное	10,3	4,8	66,6	2,71	15,11	4	2,73
Голятинка – с. Голятин	59,0	12	39,0	5,5	36,5	13	5,5
Голятинка – с. Майдан	86,0	18	23,4	7,38	48,1	22	7,38
Репинка – с. Изки	103	18	19,5	16,23	44,4	29	16,23
Репинка – с. Репино	203	22	14,4	30,2	91,1	58	30,2
Студеный – с. Верхний Студеный	8,0	2,0	56,6	0,6	5,5	4	0,5
Студеный – с. Нижний Студеный	25,4	7,5	31,6	5,9	139	8	5,9
Пилипец – с. Подобец	7,8	3,2	47,5	1,91	6,67	3	1,91
Пилипец – с. Пилипец	44,2	6,2	41,1	3,9	30,4	12	3,88
Рика – с. Нижний Быстрый	781	59	10,9	81,9	260	149	81,9
Рика – г. Хуст	1130	91	10,4	95,4	338	228	95,4

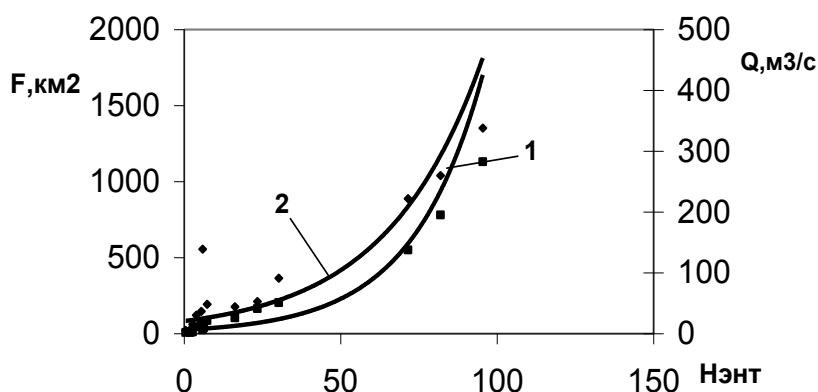


Рис.1. - Соотношение между площадью водосбора F , среднегодовыми расходами воды ливневых паводков \bar{Q}_{\max} и энтропией $H_{\text{энт}}$ р. Рика; 1 – зависимость между площадью водосбора F и энтропией $H_{\text{энт}}$; 2 – зависимость между среднегодовыми расходами воды ливневых паводков \bar{Q}_{\max} и энтропией $H_{\text{энт}}$.

Все четыре графика, представленные на рисунках 1-2, соответствуют параболической зависимости между аргументом и функцией. Причём, три из них – площади, средние максимальные расходы и длины – это возрастающие функции, а зависимость уклонов рек от энтропии – убывающая.

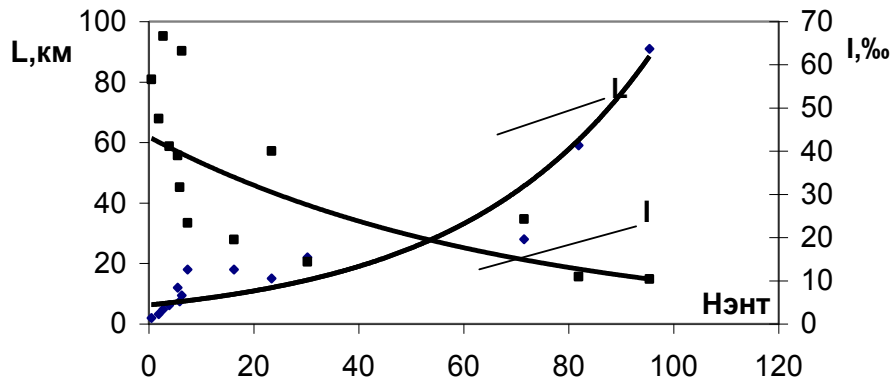


Рис.2 - Зависимость между длинами рек L (1), их средними уклонами I (2) и энтропией $H_{ЭНТ}$ р. Рика

Результатом настоящего исследования является:

1. подбор математического метода из теории информации, который может помочь описать структуру речной сети, представляя её в виде совокупности графов;
2. доказательство возможности практического расчёта энтропии по рекам бассейна р.Рика;
3. построение графиков связи между значениями энтропии речных бассейнов и их основными характеристиками: длиной, площадью водосбора, среднемноголетним расходом воды ливневых паводков, что сделано впервые для территории Украины;
4. оценка точности полученных графических зависимостей и возможность их практического использования в случае ограниченности исходной информации.

Задачей дальнейших исследований является расчет значений энтропии по всем рекам Украинских Карпат с целью построения территориально общих зависимостей между этой характеристикой и параметрами строения речных систем.

Источники и литература

1. Багров Н. А. Статистическая энтропия как показатель сходства или различия метеорологических полей. – “Метеорология и гидрология”, 1963, №1, с. 9 – 15.
2. Гарцман И. Н. Некоторые проблемы системного подхода в гидрометеорологии. – Труды ДВНИГМИ, 1975, вып. 54, с. 3 – 47.
3. Гарцман И. Н. Системные аспекты моделирования в гидрологии. – Труды ДВНИГМИ, 1977, вып. 63, с. 3 – 85.
4. Киндюк Б. В. Руслонная сеть и характеристики ливневого стока бассейна р. Рики. "Людина і довкілля". – Х.: Харківський національний університет ім. Карабіна, 2003, вип. 4, с. 70–74.
5. Хаггерт П. Сетевые модели в географии. – “Модели в географии”, М., “Прогресс”, 1971, с.287 – 343.
6. Харвей Д. Научное объяснение в географии. – М., “Прогресс”, 1974, 502 с.
7. Харрари Ф. Теория графов. – М., “Мир”, 1973, 300 с

УДК 528.44:519.2

Киндюк Б.В. В статье рассмотрен новый подход к описанию строения и распределения в пространстве полей, состоящих из рек, их притоков. Впервые к водотокам на территории Украины применён математический аппарат теории информации. Подсчитаны значения энтропии полей речных систем и показана их связь с такими характеристиками речных бассейнов, как их площадь, длина, уклон, среднемноголетний расход воды ливневых паводков.

Kindyuk B.V. In work the new approach to the description of a structure and distribution in space of the fields consisting of the rivers, of their tributaries is considered. For the first time to water-currents in territory of Ukraine the mathematical device of the theory of the information is applied. Values entropy fields of river systems are counted up and their bond with such characteristics of river pools, as their area, length, slope, average of several years the expenses of water of storm high waters is shown.

Ключевые слова: гидрографическая сеть, энтропия, бассейн реки.