УДК 528.422:519.25

Л. К. ВОЙСЛАВСКИЙ
Харьковский институт инженеров коммунального строительства

ЭНТРОПИЯ ВЫСОТ ТОЧЕК ТОПОГРАФИЧЕСКОЙ ПОВЕРХНОСТИ

Рельеф — один из основных источников информации, используемой для составления топографических карт, планов, профилей и цифровых моделей местности (ЦММ). Поэтому изучение рельефа именно в таком качестве представляет определенный теоретический интерес.

Поверхность отображается путем определения высот в характерных точках, на характерных линиях, перегибах скатов, а также и в ряде других точек, обеспечивающих требуемую густоту съемки рельефа, регламентированную соответствующими инструкциями. Вот почему отметку точки можно рассматривать как элементарное сообщение, с помощью которого передается информация от оригинала (рельеф местности) к образу (план, карта, профиль, ЦММ) [1]. Количество информации, передаваемое одной отдельно взятой отметкой, зависит от энтропии совокупности высот данного участка местности. Рассмотрим энтропию такого рода для относительно небольших участков.

Задавая любую систему высот, предполагаем, что известна отметка хотя бы одной точки, принимаемой за начало отсчета. В связи с этим энтропию высоты z_i какой-либо точки, если известна (или принимается равной нулю) начальная отметка z_0, необходимо рассматривать как условную энтропию. Обозначим ее $H(z_i|z_0)$.

Исследования показали [2], что статистические свойства топографической поверхности во многом зависят от удаленности точек местности. Следовательно, можно предположить, что и условная энтропия $H(z_l|z_0)$, как одно из статистических свойств, каким-то образом связана с расстоянием l между начальной и данной точками.

Для проверки этой гипотезы были отобраны топографические планы (карты) 20 участков, расположенных в равнинной местности и имеющих разные геоморфологические и орографические условия, площадью 0,25—290 км². На всех участках по

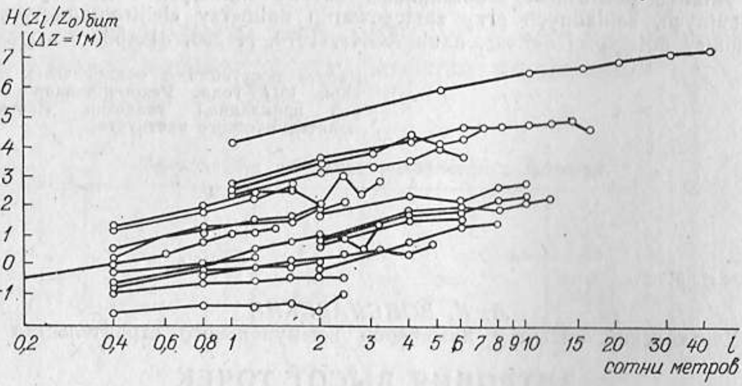

Рис. 1. Изменение условной энтропии высоты точки $H(z_l|z_0)$ в зависимости от расстояния.

горизонталям определили превышения между точками, удаленными друг от друга на расстояния $\overline{l}$, $2\overline{l}$, ... , $m\overline{l}$, где $\overline{l}$ — шаг квантования по длине линии (принят равным 2 см на плане (карте) независимо от масштаба). При этом любому расстоянию $k\overline{l}(k=1, 2, \ldots m)$ соответствует выборка объемом не менее 50 превышений ($N_{k\overline{l}} \geqslant 50$).

По каждой выборке были вычислены:

а) частота (приближенная вероятность) попадания превышений в i-й интервал по высоте ($i=1, 2, \ldots n$) (ширина интервала Δh выбиралась из условия $n \geqslant 10$)

$$p\,(h_{k\overline{l}})_i = \frac{N_i}{N_{k\overline{l}}};$$

б) условные энтропии высот для расстояний $k\overline{l}$ по формуле К. Э. Шеннона [4]

$$H\,(z_{k\overline{l}}|z_0) = -\sum_{i=1}^{n} p\,(h_{k\overline{l}})_i \log p\,(h_{k\overline{l}})_i - \log \Delta h, \qquad (1)$$

где символ log обозначает двоичный логарифм.

Если не учитывать отдельные несущественные колебания, обусловленные, по-видимому, случайными причинами, то на основании рис. 1 можно сделать вывод о постепенном возрастании энтропии с увеличением расстояния l. Более того, в результате обработки экспериментальных данных установлено, что между величинами $H(z_l|z_0)$ и $\log l$ существует тесная корреляционная зависимость, близкая к линейной. При общем объеме выборки значений $H(z_l|z_0)$ $q=103$ выборочный коэффициент корреляции равен 0,90, а выборочное корреляционное отношение — 0,93, поэтому для расстояний, не превышающих 700—1000 м, можно принять

$$H(z_l|z_0)=k\log l+C, \qquad (2)$$

где k и C — параметры, зависящие от характера рельефа данного участка местности. Учитывая формальное подобие выражения (2) и формулы нитяного дальномера, параметры k и C назовем соответственно коэффициентом и постоянным слагаемым энтропии высот.

Значения параметров, характеризующих линейную зависимость условной энтропии от величины $\log l$

Номер участка	Площадь участка, км²	Наибольшая разность высот h, м	$\log h$	Диапазон исследованных расстояний, м	k	C
19	0,2	8,2	3,04	20÷140	$0{,}62\pm0{,}03$	$1{,}0\pm0{,}05$
4	0,4	1,2	0,26	40÷120	$0{,}32\pm0{,}04$	$3{,}4\pm0{,}04$
26	0,4	14,5	3,86	40÷160	$0{,}76\pm0{,}05$	$2{,}1\pm0{,}1$
3	0,5	20,7	4,37	40÷200	$0{,}67\mp0{,}05$	$1{,}4\pm0{,}1$
16	0,5	2,9	1,53	40÷200	$0{,}38\pm0{,}06$	$-0{,}2\pm0{,}1$
2	0,6	1,2	0,26	40÷240	$0{,}14\pm0{,}06$	$-1{,}5\pm0{,}1$
13а	0,5	4,4	2,14	40÷240	$0{,}19\pm0{,}03$	$-0{,}8\pm0{,}03$
17	0,5	10,7	3,42	40÷240	$0{,}74\pm0{,}07$	$1{,}3\pm0{,}1$
20	0,7	17,3	4,11	40÷320	$0{,}54\pm0{,}09$	$2{,}1\pm0{,}1$
22	0,8	11,9	3,57	40÷320	$0{,}62\pm0{,}08$	$0{,}2\pm0{,}1$
13	1,0	7,2	2,85	80÷480	$0{,}37\pm0{,}02$	$0{,}2\pm0{,}02$
24	3,2	43,9	5,45	100÷600	$0{,}62\pm0{,}07$	$2{,}4\pm0{,}1$
18	4,1	28,0	4,81	100÷700	$0{,}70\pm0{,}06$	$2{,}7\pm0{,}1$
25	7,6	17,9	4,16	200÷600	$0{,}60\pm0{,}2$	$0{,}1\pm0{,}4$
15	10,5	5,9	2,56	200÷800	$0{,}89\pm0{,}07$	$-1{,}1\pm0{,}1$
11	12,3	18,4	4,20	200÷1000	$0{,}32\pm0{,}07$	$1{,}6\pm0{,}2$
7	11,8	8,3	3,05	200÷1000	$0{,}67\pm0{,}05$	$0{,}2\pm0{,}1$
31	13,5	11,7	3,55	200÷1200	$0{,}61\pm0{,}06$	$0{,}3\pm0{,}1$
1	17,7	32,6	5,03	100÷1600	$0{,}64\pm0{,}05$	$2{,}8\pm0{,}1$
41	290	17,5	7,45	100÷4000	$0{,}64\pm0{,}04$	$4{,}4\pm0{,}1$

Примечание: При вычислении k и C принято: $\Delta h=1$ м, $l_0=100$ м.

В таблице приведены значения k и C для 20 участков, вычисленные по экспериментальным данным способом наименьших квадратов. Коэффициент энтропии высот k зависит от пространственной (горизонтальной) неоднородности и расчлененности рельефа. Есть основания предполагать, что k может принимать значения от 0 до 1.

Значение постоянного слагаемого C зависит от следующих данных:

а) единицы длины в формуле (2) *;

б) интервала квантования высот Δh (на основании выражения (1) можно сделать вывод, что чем меньше Δh, тем больше постоянное слагаемое C);

в) вертикальной расчлененности рельефа и перепадов его высот.

В самом деле, для совокупности значений C и $\log h$, где $h = z_{\max} - z_{\min}$ — наибольшая разность высот в границах участка, коэффициент корреляции равен 0,86. Это подтверждает зависимость C от h. Уравнение регрессии имеет вид $C = \log \alpha + \beta \log h$.

Следовательно, зависимость C от h и Δh можно записать

$$C = \log \frac{\alpha h^{\beta}}{\Delta h}, \qquad (3)$$

где α и β — параметры, характеризующие вертикальную неоднородность рельефа. Но α зависит также от выбора единицы измерения l. Каждому участку местности соответствуют свои значения α и β.

На основании формул (2) и (3) имеем

$$H(z_l/z_0) = \log \frac{\alpha l^k h^{\beta}}{\Delta h}. \qquad (4)$$

Ранее было отмечено, что формулы (2) и (4) справедливы только при расстоянии 700—1000 м. Для больших расстояний (рис. 1) характерна параболическая зависимость

$$H(z_l|z_0) = k_1 \log l - k_2 (\log l)^2 + C, \qquad (5)$$

где горизонтальная расчлененность рельефа характеризуется уже двумя параметрами k_1 и k_2. Если обозначить

$$k = k_1 - k_2 \log l, \qquad (6)$$

снова получим выражение (2) или (4). Разница лишь в том, что, начиная с некоторого l, коэффициент энтропии k является функцией расстояния.

По имеющимся экспериментальным данным более-менее значимую параболическую аппроксимацию удалось получить только для трех участков.

Номер участка	Площадь участка, км²	Диапазон исследованных расстояний, м	k_1	k_2	C	$l_{\max}$, км
31	13,5	200—1200	1,04	0,10	—0,1	4
1	17,7	100—1600	1,05	0,12	2,7	2
40—41	290	100—4000	0,90	0,06	4,3	18

Примечание: $l_0 = 100$ м, $\Delta h = 1$ м.

* Мы взяли $l_0 = 100$ м, но можно выбрать другую, более удобную единицу.

Если предположить, что с увеличением l энтропия не убывает, то применение формулы (5) будет ограничено предельной дальностью

$$l_{\max} = 2^{\frac{k_1}{2k_2}}, \tag{7}$$

соответствующей максимуму функции (5). При расстояниях, превышающих полученные по формуле (7), столкнемся, по-видимому, с еще более сложными зависимостями $H(z_l|z_0)$ от l.

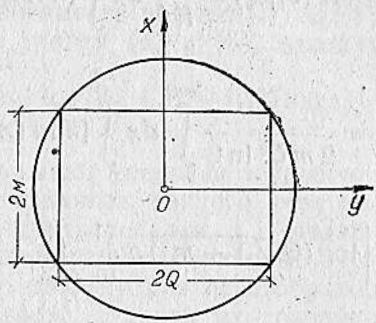

Рис. 2. Размещение начальной точки с высотой z_0 для участков, имеющих форму прямоугольника и круга.

Выражения (2), (4) и (5) дают разовую или, точнее, мгновенную энтропию в точке с заданным значением l. Однако для расчета информационной емкости плана, профиля или ЦММ гораздо больший интерес представляет **среднее количество информации о рельефе,** которое содержит отметка любой произвольно взятой точки местности. Для этого энтропию $H(z_l|z_0)$ необходимо усреднить по всем l, возможным в границах данного участка, т. е. брать $H(z_i|z_0) = M[H(z_l|z_0)]$, где M — операция математического ожидания.

Рассмотрим несколько типичных случаев.

1. Участок в виде узкой полосы, например трасса линейного сооружения. Точку с известной отметкой поместим в начале или в конце трассы. Предположим также, что остальные точки на трассе размещены равномерно через очень малые промежутки. Тогда энтропию высоты любой точки в интервале от 0 до L запишем

$$H(z_i|z_0) = \frac{1}{L}\int_0^L (k \log l + C)\,dl = k \log \frac{L}{e} + C, \tag{8}$$

где e — основание натуральных логарифмов.

2. Участок прямоугольной формы (рис. 2). Точку 0 с известной отметкой поместим в центре участка и примем за начало прямоугольной системы координат. Расстояние l представим в виде функции $l = \sqrt{x^2 + y^2}$, причем x изменяется от 0 до $\pm M$,

а y от 0 до $\pm Q$. При $M \leqslant Q + \frac{M}{Q} = m \leqslant 1$, а площадь прямоугольника $F = 4\,mQ^2$.

Если предположить, что точки распределены равномерно по всей площади через малые промежутки, то энтропию высоты точки в границах участка на основании формулы (2) можно выразить так:

$$H(z_i|z_0) = \frac{4}{F}\int_0^M dx \int_0^Q (k \log \sqrt{x^2 + y^2} + C)\, dy =$$

$$= \frac{1}{2\,mQ^2 \ln 2}\int_0^{mQ} dx \int_0^Q [k \ln (x^2 + y^2) + 2C \ln 2]\, dy =$$

$$= k\left[\log (Q\sqrt{1 + m^2})\, g \frac{1}{\ln 2}\left(\frac{3}{2} - \frac{1 - m^2}{2\,m}\operatorname{arctg} m - \frac{\pi m}{4}\right)\right] + C.$$

Учитывая, что расстояние до наиболее удаленной точки $L = Q\sqrt{1 + m^2}$, и обозначая

$$t = \frac{3}{2} - \frac{1 - m^2}{2\,m}\operatorname{arctg} m - \frac{\pi m}{4},$$

получаем

$$H(z_i|z_0) = k \log \frac{L}{e^t} + C. \tag{9}$$

Показатель t в выражении (9) в зависимости от соотношения сторон m принимает значения от 0,71 до 1,00.

3. Участок имеет форму круга (рис. 2). Исходная точка находится в центре. Остальные точки равномерно распределены через малые промежутки по всей площади. Расстояние до наиболее удаленной точки L равно радиусу. В данном случае энтропию высоты любой точки можно найти на основании формулы (2), записав подынтегральное выражение в полярных координатах:

$$H(z_i|z_0) = \frac{1}{\pi L^2}\int_0^{2\pi} d\varphi \int_0^L (k \log l + C)\, l\,dl = \frac{1}{2\pi \ln 2}\int_0^{2\pi}\left[k\left(\ln L - \frac{1}{2}\right) + \right.$$

$$\left. + C \ln 2\right] d\varphi = k \log \frac{L}{e^{1/2}} + C. \tag{10}$$

Выражения (8), (9) и (10) отличаются только показателем степени у e, поэтому, учитывая формулу (3), получаем обобщенное выражение

$$H(z_i|z_0) = k\log\frac{L}{e^t} + C = \log\left[\alpha\left(\frac{L}{e^t}\right)^k \frac{h^\beta}{\Delta h}\right], \tag{11}$$

где L — наибольшее расстояние от точки с отметкой z_0 до границ участка; t — показатель, зависящий от формы участка и принимающий значения от $\frac{1}{2}$ до 1. Для больших участков, где существует параболическая зависимость величин $H(z_l|z_0)$ и $\log l$, интегрируя выражение (5) с учетом (3) и (6), находим

$$H(z_i|z_0) = k_1\log\frac{L}{e^t} - k_2\left(\log\frac{L}{e^t}\right)^2 + C - k_2 t^2(\log e)^2. \tag{12}$$

Формулы (11) и (12) выражают среднюю энтропию высоты произвольно взятой точки в границах данного участка. Порядок выбора точек заранее не предопределен, т. е. является случайным. Между тем топографическая поверхность — это упорядоченное множество точек, поэтому для ее отображения нужен не хаотический набор высот, а такая их совокупность, в которой размещение точек согласуется с характером рельефа.

Предположим, что для участка местности создана цифровая модель рельефа (ЦМР) в виде упорядоченной последовательности высот точек 0, 1, 2, ..., n. Будем считать, что расстояние между точками 0, 1, 2, ..., n не превышает некоторого предела $l_п$, обусловленного характером рельефа и требуемой детальностью его отображения. Общее количество метрической информации о рельефе, содержащейся в ЦМР, обозначим $\mathrm{I}(z_1, z_2, \ldots, z_n|z_0)$.

Из-за погрешностей измерений, погрешностей обобщения рельефа при съемке и других причин эта информация будет меньше априорной энтропии совокупности высот данного участка, т. е.

$$\mathrm{I}(z_1, z_2, \ldots, z_n|z_0) < H(z_1, z_2, \ldots, z_n|z_0). \tag{13}$$

Учитывая иерархическую аддитивность энтропии, записываем [3]:

$$H(z_1, z_2, \ldots, z_n|z_0) = H(z_1|z_0) + H(z_2|z_0, z_1) + \\ + \ldots + H(z_n|z_0, z_1, \ldots, z_{n-1}). \tag{14}$$

Из выражения (2) следует $H(z_i|z_{i-1}) \leqslant k\log l_п + C$. Но, поскольку при добавлении условий энтропия не возрастает, получаем

$$H(z_1|z_0) \leqslant k\log l_п + C;$$
$$H(z_2|z_0, z_1) \leqslant H(z_2|z_1) \leqslant k\log l_п + C;$$
$$\cdots\cdots\cdots\cdots$$
$$\cdots\cdots\cdots\cdots$$
$$H(z_n|z_0, z_1, \ldots, z_{n-1}) \leqslant H(z_n|z_{n-1}) \leqslant k\log l_п + C.$$

Следовательно, энтропия любой точки в упорядоченной последовательности ограничена неравенством

$$H(z_i|z_0,\ z_1,\ldots,\ z_{i-1}) \leqslant k \log l_\text{п} + C. \tag{15}$$

На основании формул (13), (14), (15), учитывая (3), находим неравенство

$$I(z_1, z_2, \ldots, z_n/z_0) < n \log \left(\frac{\alpha\, l_\text{п}^k\, h^\beta}{\Delta h} \right), \tag{16}$$

ограничивающее сверху количество метрической информации, содержащейся в ЦМР.

Известно, что максимальная безусловная энтропия в границах участка с фиксированной наибольшей разностью высот [4]

$$H(z_i) = \log \frac{h}{\Delta h}. \tag{17}$$

Полагая $l_\text{п} \ll L$, на основании (11), (12), (15) и (17) имеем $H(z_i|z_0,\ z_1,\ldots,\ z_{i-1}) < H(z_i|z_0) < H(z_i)$, откуда следует, что рельеф местности обладает некоторой избыточностью. Значение этой избыточности в значительной степени зависит от интервала квантования высот Δh.

Список литературы: 1. *Мазур М.* Качественная теория информации. М., «Мир», 1974, с. 33—38. 2. *Протодьяконов М. И.* Числовые характеристики топографических условий местности, исчисление эксплуатационных расходов. М., Транспечать, 1925, с. 4—9. 3. *Стратонович Р. Л.* Теория информации. М., «Советское радио», 1975, 19 с. 4. *Шеннон К. Э.* Работы по теории информации и кибернетике, М., ИЛ, 1963, 260 с.

Работа поступила в редколлегию 15 ноября 1976 года. Рекомендована кафедрой геодезии Харьковского института инженеров коммунального строительства.