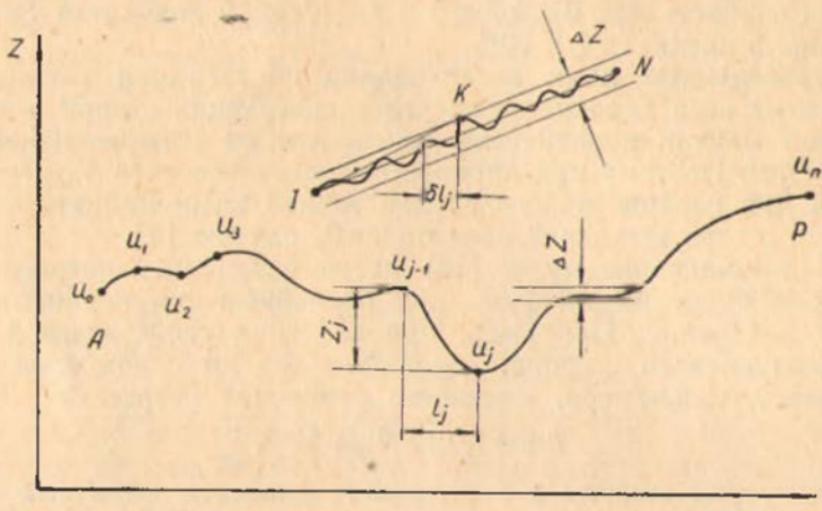


ИНФОРМАЦИОННАЯ МОДЕЛЬ ПРОФИЛЯ МЕСТНОСТИ

Продольный профиль местности является источником специальной топографической информации [6], используемой при проектировании линейных инженерных сооружений и решении других специальных задач. Эта информация может быть представлена как в графической, так и в дискретной (цифровой) форме, пригодной для записи и хранения на машинных носите-



Аппроксимация профиля местности ломаной линией.

лях. Вот почему возникает необходимость наряду с изучением семантических и прагматических свойств такой информации определить информационную емкость профиля, т. е. количество информации, содержащейся в профиле и отдельных его элементах. Решение этой задачи на основе теории информации К. Э. Шеннона [8] и составляет предмет нашего исследования.

Рассмотрим профиль местности, представленный на рис. 1 кривой AP . Бесконечному множеству точек профиля соответствует бесконечное множество их высот. Но при использовании профиля для решения научных и производственных задач мы принимаем некоторый интервал Δz , в пределах которого высоты точек считаются практически неразличимыми. Если установлен интервал квантования высот Δz , из бесконечного множества точек профиля можно выделить конечное подмножество его характерных точек.

В самом деле, пусть часть профиля представлена ломаной линией, состоящей из очень малых отрезков δl_j (рис. 1) с уклонами i_j ($j=1, 2, \dots, n$). Уклон замыкающей IN можно найти как среднее взвешенное значение

$$i_{IN} = \frac{\sum_1^n \delta l_j i_j}{\sum_1^n \delta l_j}.$$

Соответственно уклон промежуточной замыкающей IK будет

$$i_{IK} = \frac{\sum_1^k \delta l_j i_j}{\sum_1^k \delta l_j}.$$

Если для всякого $k=1, 2, \dots, n-1$ имеет место неравенство

$$\frac{\Delta z}{2} \geq \sum_1^k \delta l_j \left| \frac{\sum_1^n \delta l_j i_j}{\sum_1^n \delta l_j} - \frac{\sum_1^k \delta l_j i_j}{\sum_1^k \delta l_j} \right|, \quad (1)$$

то отметку какой угодно точки профиля в интервале от I до N можно получить с достаточной точностью линейной интерполяцией по IN . Это дает нам основание представить данный участок профиля прямой IN .

И наоборот, если для какого-либо $k=1, 2, \dots, n-1$ неравенство (1) не выполняется, конец k -го отрезка следует рассматривать как характерную точку профиля. Таким образом, кривая AP может быть заменена ломаной, вершины которой u_0, u_1, \dots, u_n — характерные точки профиля.

Принятая нами аппроксимация отражает объективные свойства естественного профиля местности и в этом смысле не зависит от того, по каким материалам он составлен (полевые измерения, измерения на стереомодели, топографические карты). Одновременно она, как и любая другая модель, субъективна, так как выбор Δz зависит от точности измерений и погрешностей обобщения рельефа, т. е. в конечном счете от метода получения профиля.

Рассмотрим информационную модель профиля, аппроксимированного ломаной линией. Название «информационная модель» принято нами в соответствии с классификацией, предложенной Ю. К. Неумывакиным [6]. При этом будем полагать, что размещению характерных точек по длине профиля свойственна объективная неопределенность, обусловленная природным многообразием рельефа.

Как известно, количественной мерой неопределенности является энтропия [8]. На основании свойства иерархической адди-

тивности энтропию H подмножества характерных точек $\{u_0, u_1, \dots, u_n\}$ представим в виде суммы условных энтропий [7]

$$H(u_0, u_1, \dots, u_n) = H(u_0) + H(u_1/u_0) + \\ + H(u_2/u_0, u_1) + \dots + H(u_n/u_0, u_1, \dots, u_{n-1}). \quad (2)$$

Допустив, что положение начальной точки профиля фиксировано, т. е. $H(u_0)=0$, а также имея в виду, что при добавлении условий энтропия не возрастает, т. е.

$$H(u_j/u_0, u_1, \dots, u_{j-1}) \leq H(u_j/u_{j-1}),$$

с некоторым запасом в сторону увеличения энтропии возможно заменить (2) выражением

$$H(u_0, u_1, \dots, u_n) = H(u_1/u_0) + \\ + H(u_2/u_1) + \dots + H(u_n/u_{n-1}). \quad (3)$$

Положение характерной точки профиля u_j относительно предшествующей ей точки u_{j-1} однозначно определяется расстоянием l_j и превышением z_j . Поэтому можно принять

$$H(u_j/u_{j-1}) = H(l_j, z_j), \quad (j = 1, 2, \dots, n). \quad (4)$$

Согласно [2]

$$H(l_j, z_j) = M \{-\log [(f, l, z) \Delta l \Delta z]\},$$

где M — символ математического ожидания, $f(l, z)$ — плотность распределения вероятностей случайных величин l и z , Δl — шаг квантования профиля по длине. Следовательно, энтропия профиля местности определяется распределением случайных величин

$$l_1, l_2, \dots, l_n, \quad z_1, z_2, \dots, z_n. \quad (5)$$

Для экспериментального исследования законов распределения вероятностей случайных величин (5) использованы производственные материалы полевых изысканий по десяти трассам линейных сооружений. Трассы проложены в равнинных районах, отличающихся большим разнообразием геоморфологических условий. Все исследованные профили имеют равное число характерных точек $n=208$ при $\Delta z=0,1$ м. Длина отдельного профиля колеблется в пределах от 5,8 до 16,6 км, а их общая протяженность составляет 106 км.

Данные, характеризующие распределение расстояний и превышений между характерными точками, объединены по каждому профилю в 20 интервалов равной ширины и представлены в табл. 1, 2. Статистические оценки, определенные по группированным данным, могут оказаться смещенными вследствие того, что распределение внутри интервала принимается равномерным. Вот почему все оценки параметров, приведенные в табл. 1, 2, и коэффициенты корреляции вычислены нами по первичным, т. е. не сгруппированным данным.

Tabelle 1

Распределение расстояний между характерными точками профиля по длине

Номер опыта	Номер опыта	Количество отрезков, попадающих в правые интервалы [длина интервала выражена в единицах t_{\max}]										Параметры гамма-распределения	χ^2	Критерий согласия													
		Максимальная ошибка определения B					Средняя ошибка определения A																				
												λ	α	χ^2													
		0,0-0,05	0,05-0,1	0,1-0,15	0,15-0,2	0,2-0,25	0,25-0,3	0,3-0,35	0,35-0,4	0,4-0,45	0,45-0,5	0,5-0,55	0,55-0,6	0,6-0,65	0,65-0,7	0,7-0,75	0,75-0,8	0,8-0,85	0,85-0,9	0,9-0,95	0,95-1	1					
1	13,1	19	31	39	44	27	17	9	4	6	3	5	2	1	—	—	—	—	—	—	329	63,1	2,24	36,2	9,0	7	
2	8,0	26	39	59	30	25	14	7	2	2	—	1	—	—	—	—	—	—	—	—	250	38,6	2,60	69,2	4,5	6	
3	6,0	11	32	47	28	24	21	12	15	4	1	3	3	3	1	1	—	—	—	—	134	28,9	2,41	86,4	1,8	6	
4	5,8	31	64	47	20	19	10	4	2	4	2	1	—	—	1	1	—	1	—	—	188	27,7	2,11	84,1	6,4	6	
5	16,6	5	19	26	52	44	23	6	8	7	5	5	4	—	2	1	—	—	—	—	343	79,6	3,25	41,5	9,3	5	
6	10,5	10	25	30	21	24	24	21	18	8	8	3	3	4	1	3	—	—	2	1	190	50,7	2,52	53,0	7,2	7	
7	13,2	2	23	44	36	34	31	19	7	4	1	1	—	3	—	—	1	—	—	1	287	63,3	3,92	63,8	9,5	6	
8	11,7	13	25	19	53	48	3	6	1	24	7	—	4	—	1	—	—	3	—	—	242	56,5	2,82	52,8	7,0	3	
9	10,6	6	22	24	27	40	33	17	12	7	4	7	4	2	1	—	—	1	—	—	201	51,0	3,09	62,2	12,1	8	
10	10,5	—	15	18	22	19	27	35	18	21	9	8	2	4	1	—	2	—	2	3	1	157	50,6	3,30	65,2	14,1	8

использованы элементы полинома A , содержащего 99% всех линий профилей.

Таблица 2

Распределение превышений между характерными точками профиля в интервале от z_{\min} до z_{\max}

		Количество превышений, попадающих в равные интервалы (длина интервала выражена в долях $R = z_{\max} - z_{\min}$)										Параметры				Критерий согласия			
		Маркемарное значение $\bar{z}_{\text{небольшое}}$										Параметры							
												R , \bar{z}	α	β	γ	δ			
												\bar{z}	α	β	γ	δ			
1	-8,2	2	-	2	-	3	2	13	36	58	42	22	8	4	2	-	1	-	
2	-4,4	1	-	—	3	2	3	6	23	22	63	40	23	9	5	2	2	1	-
3	-10,0	1	1	1	1	3	8	22	25	38	49	32	13	5	5	1	1	-	
4	-8,8	1	-	—	3	1	4	3	2	7	11	17	46	51	29	12	13	3	-
5	-4,0	1	2	3	4	14	15	26	45	53	27	9	2	3	-	1	1	1	-
6	-6,9	1	-	1	-	3	8	17	56	59	37	15	4	3	-	1	2	-	-
7	-5,2	2	-	2	2	10	24	44	35	42	25	10	6	2	1	2	-	-	-
8	-5,6	1	-	3	4	5	8	21	28	32	42	27	14	6	9	3	1	1	1
9	-6,0	1	-	—	1	2	3	2	25	40	42	38	23	16	5	3	4	-	1
10	-4,9	1	-	1	2	2	2	7	8	20	24	24	47	31	16	5	8	6	1
															-	1	+ 4,0	8,9	0,02
															-	1	+ 4,0	8,9	0,02

В результате проведенного исследования получены коэффициенты корреляции $r(l, z)$ между длиной линии и превышением.

Профиль	1	2	3	4	5	6	7	8	9	10
	-0,22	-0,02	-0,01	0,22	-0,04	0,02	-0,03	-0,04	0,15	-0,07

Для восьми профилей абсолютная величина $|r(l, z)|$ меньше $3\sigma(r)$, где

$$\sigma(r) = \frac{1 - r^2}{\sqrt{n}}$$

— среднее квадратичное отклонение коэффициента корреляции [4]. В остальных двух случаях $|r(l, z)|$ лишь незначительно больше $3\sigma(r)$. Следовательно, корреляционная связь величин l или отсутствует, или выражена очень слабо.

Допустим, при $\Delta z \rightarrow 0$, последовательность характерных точек u_0, u_1, \dots, u_m обладает свойствами стационарности, отсутствия последействия и ординарности, т. е. представляет собой стационарный пуассоновский поток [2]. По мере увеличения Δz последовательность разрежается за счет аппроксимации отдельных участков профиля прямой линией, как это следует из неравенства (1). В результате образуется новая, разреженная последовательность u_0, u_1, \dots, u_n ($n < m$), у которой расстояния между характерными точками должны иметь гамма-распределение с плотностью вероятностей

$$f(l) = \frac{\lambda^\alpha}{\Gamma(\alpha)} l^{\alpha-1} \exp(-\lambda l), \quad (6)$$

где λ, α — параметры распределения, $\Gamma(\alpha)$ — гамма-функция.

Для определения параметров распределения по экспериментальным данным на основе метода максимального правдоподобия [1] получены уравнения

$$\begin{aligned} \bar{l}\lambda - \alpha &= 0, \\ \ln \alpha - \frac{d \ln \Gamma(\alpha)}{d\alpha} + \frac{\sum \ln l_j}{n} - \ln \bar{l} &= 0, \end{aligned} \quad (7)$$

где \bar{l} — среднее расстояние между характерными точками профиля. Из решения уравнений (7) найдены оценки параметров λ и α .

Соответствие опытного распределения l теоретическому (6) проверено при помощи критерия χ^2 . Так как в правой части табл. I много пустых, т. е. не содержащих ни одного отрезка, интервалов, из общей совокупности отрезков выделено «более плотное» подмножество A , охватывающее 99% всех линий профиля. В границах этого подмножества по каждой выборке отрезки профиля снова были разбиты на 20 равных интервалов. Затем в соответствии с рекомендацией ГОСТ 11.006—74 некоторые из них пришлось объединить так, чтобы в каждом ин-

тервале теоретически ожидаемое число отрезков было не менее 10. Количество интервалов группирования N , полученное в результате объединения, колеблется для отдельных выборок от 6 до 11.

Вычисленные значения χ^2 не превышают критических при уровне значимости $q=0,05$ и числе степеней свободы $k=N-3$. Таким образом, гипотеза о распределении случайных величин l может быть принята.

Перейдем к исследованию распределения вероятностей превышений z между характерными точками профиля. Как видно из табл. 2, совокупности превышений обладают свойствами ограниченности, компенсации и унимодальности, а также свойством равновероятности в том смысле, что положительные и отрицательные превышения встречаются в среднем одинаково часто.

На основе анализа опытных данных (табл. 2) мы пришли к выводу, что совокупность превышений профиля имеет распределение Лапласа с плотностью вероятностей [5]

$$f(z) = \frac{1}{2\beta} \exp\left(-\frac{|z-\xi|}{\beta}\right). \quad (8)$$

Для определения параметров ξ и β по экспериментальным данным использованы формулы

$$\xi = \frac{\sum z_j}{n}, \quad \beta = \frac{\sum |z_j - \xi|}{n}. \quad (9)$$

Причем последнее выражение в (9) получено на основе метода максимального правдоподобия. Опытное распределение хорошо согласуется с теоретическим, что проверено при помощи критерия χ^2 . Вычисленные значения χ^2 меньше критических для девяти профилей при уровне значимости $q=0,05$, а для профиля № 5 при $q=0,02$. Число степеней свободы $k=N-3$, где N — количество объединенных интервалов группирования с теоретически ожидаемым числом превышений не менее 10. Таким образом, гипотеза о распределении Лапласа может быть принята.

Гипотезы о распределении вероятностей случайных величин l и z подтвердились также и для профиля длиной 23,7 км, составленного по топографической карте масштаба 1 : 50000. Отсюда можно предположить, что принятая нами вероятностная модель отражает объективные свойства рельефа местности и не зависит от метода получения профиля.

Продолжим рассмотрение информационной модели профиля местности на основе теории информации К. Э. Шеннона [8]. Так как случайные величины l и z подчинены разным законам распределения вероятностей и, как было установлено, между

также практически отсутствует стохастическая зависимость, следовательно предположить, что они независимы и принять

$$H(l, z) = H(l) + H(z). \quad (10)$$

Для совокупности отрезков профиля при $1 \leq a \leq 6$ оценку энтропии $H(l)$ можно выразить формулой из работы [3]

$$H(l) = \log \left[\Gamma(\alpha) \left(\frac{e}{\alpha} \right)^{\alpha} \left(\frac{\alpha + 0,43}{\alpha - 0,21} \right)^{\alpha-1} \frac{l}{\Delta l} \right] \text{бит},$$

где символ \log обозначает двоичный логарифм, e — основание натуральных логарифмов, Δl — шаг квантования по длине. Обозначив

$$\psi(\alpha) = \Gamma(\alpha) \left(\frac{e}{\alpha} \right)^{\alpha} \left(\frac{\alpha + 0,43}{\alpha - 0,21} \right)^{\alpha-1},$$

будем иметь

$$H(l) = \log \left[\frac{\psi(\alpha) l}{\Delta l} \right] \text{бит}. \quad (11)$$

Энтропия превышения z представляет собой математическое ожидание функции $-\log[f(z)\Delta z]$ [2]. Принимая во внимание (8), можем записать

$$H(z) = M \left\{ -\log \left[\frac{1}{2\beta} \exp \left(-\frac{|z - \xi|}{\beta} \right) \Delta z \right] \right\} = \log 2\beta + \frac{\log e}{\beta} M(|z - \xi|) - \log \Delta z.$$

Так как для распределения Лапласа

$$M(|z - \xi|) = \beta, \quad \sigma(z) = \beta \sqrt{2},$$

получим

$$H(z) = \log \left(\frac{2\beta e}{\Delta z} \right) = \log \left[\frac{\sqrt{2} e \sigma(z)}{\Delta z} \right] \text{бит}, \quad (12)$$

где $\sigma(z)$ — среднее квадратическое отклонение превышений.

На основании формул (4), (10), (11), (12) оценка энтропии характерной точки профиля будет равна

$$H(u_j/u_{j-1}) = \log \left[\frac{2e \psi(\alpha) \beta l}{\Delta z \Delta l} \right] \text{бит}. \quad (13)$$

Предположим, что характерные точки располагаются по всей длине профиля с одинаковой средней плотностью $\Lambda = 1/l$. Тогда удельная энтропия в расчете на единицу длины при фиксированной начальной точке u_0 определится из выражения

$$H(\Lambda) = \Lambda \log \left[\frac{2e\psi(\alpha)\beta\bar{l}}{\Delta z \Delta l} \right] \text{бит/км.} \quad (14)$$

Соответственно энтропия профиля произвольной длины \bar{l} может быть записана в виде

$$H(\bar{l}) = L H(\Lambda) \quad \text{бит.} \quad (15)$$

Формулы (13), (14) и (15) представляют среднее количество информации, необходимой для того, чтобы воспроизвести во всех деталях форму и размеры профиля при заданных интервалах квантования высот Δz и расстояний Δl .

В результате проведенных исследований установлено, что для обследованных профилей при $\Delta z=0,1$ м и $\Delta l=1$ м оценка энтропии характерной точки изменяется в пределах от 11.9 до 13.5 бит/точку. В то же время оценка удельной энтропии $H(\Lambda)$ при тех же Δz и Δl колеблется от 160 до 450 бит/км. Следовательно, основными факторами, определяющими информационную емкость профиля, являются среднее расстояние между характерными точками \bar{l} и интервалы квантования Δz , Δl . Параметры α и β играют при этом второстепенную роль.

В заключение отметим, что результаты наших исследований могут быть использованы для расчета информационной емкости цифровой модели местности или банка данных, а также в качестве объективных критериев при установлении категорий трудности на трассировочные работы.

Список литературы: 1. Браунли К. А. Статистическая теория и методология в науке и технике. — М.: Наука, 1977. 2. Вентцель Е. С. Теория вероятностей. — М.: Физматгиз, 1962. 3. Войславский Л. К. Энтропия контуров местности, изображенных на топографических картах. — Геодезия, картография и аэрофотосъемка, 1979, вып. 29. 4. Гайдаваев П. А., Большаков В. Д. Теория математической обработки геодезических измерений. — М.: Недра, 1969. 5. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. — М.: Наука, 1973. 6. Неумывакин Ю. К. Обоснование точности топографических съемок для проектирования. — М.: Недра, 1976. 7. Стратонович Р. Л. Теория информации. — М.: Советское радио, 1975. 8. Шеннон К. Э. Работы по теории информации и кибернетике. — М.: ИЛ, 1963.