

Capítulo 4

El formato raster. Fundamentos

El formato ráster se fundamenta en la división del área de estudio en una matriz de celdillas, generalmente cuadradas. Cada una de estas celdillas recibe un único valor que se considera representativo para toda la superficie abarcada por la misma. Este formato, por tanto, cubre la totalidad del espacio, este hecho supone una ventaja fundamental ya que pueden obtenerse valores de forma inmediata para cualquier punto del mismo.

4.1. Elementos que componen una capa raster

Una capa en formato raster está compuesta por cuatro elementos fundamentales:

- La *matriz de datos*, que puede contener tres tipos de datos:
 - Valores numéricos en caso de que la variable representada sea cuantitativa (fig ??)
 - Identificadores numéricos en caso de que se trate de una variable cualitativa. Estos identificadores se corresponden con etiquetas de texto que describen los diferentes valores de la variable cualitativa (fig ??).
 - Identificadores numéricos únicos para cada una de las entidades representadas en caso de que la capa raster contenga entidades (puntos, líneas o polígonos)

Aunque el formato raster responde fundamentalmente a una concepción del mundo como conjunto de variables puede utilizarse también para representar entidades. Si suponemos un fondo en el que no exista nada (valor *NULL* en todas las celdillas) podemos representar objetos puntuales mediante celdillas aisladas con un valor (diferente de cero) que actuará normalmente como identificador; las líneas estarían constituidas por ristas de celdillas adyacentes con valores diferentes de cero (figura ??), los polígonos por conjuntos de polígonos con un mismo valor diferente de cero (figura ??).

Por tanto es necesario introducir el concepto de valor nulo ($NULL^1$), que es el que tienen las celdillas en las que no aparece ninguna entidad.

Si el conjunto de polígonos cierra el espacio, el resultado es similar a representar una variable espacial cualitativa, la diferencia estriba en que en un mapa de polígonos, cada uno de ellos tiene su propio identificador. Por ejemplo, en un mapa de usos de suelo considerado como variable cualitativa, todos los campos de secano tendrán el mismo valor, pero en un mapa de polígonos representado en formato raster cada polígono tiene su propio identificador, posteriormente a cada identificador se asigna su uso de suelo.

La matriz de datos se almacena en un fichero como una lista de valores numéricos, pero una capa raster necesita además información que permita al programa y al usuario ubicar en el espacio estos valores y entender su significado. Concretamente se necesita:

- *Información geométrica* acerca de la matriz y de su posición en el espacio:
 - Número de columnas (n_c)
 - Número de filas (n_f)
 - Coordenadas de las esquinas de la capa (e, w, s, n)
 - Resolución o tamaño de pixel en latitud (r_x) y en longitud (r_y)
- Una *tabla de colores* que permita decidir de que color se pintará cada celdilla en la pantalla
- En caso de que la variable sea cualitativa, una tabla que haga corresponder a cada identificador numérico una *etiqueta de texto* descriptiva.

Hay una serie de convenciones acerca de la forma de representación. Así la primera columna en una capa ráster, la columna número cero, es la de la izquierda (oeste) aumentando el número de columna hacia la derecha (este), mientras que la primera fila, la número cero, será la superior (norte) aumentando hacia abajo (sur), en sentido contrario al que siguen las coordenadas geográficas y las representaciones en un modelo vectorial. Puesto que la primera fila es la número 0, estas se numeran desde 0 hasta $n_f - 1$, y lo mismo para las columnas.

Todo ello da lugar a una serie de relaciones entre las variables que definen la situación espacial de las celdillas y que van a utilizarse a la hora de visualizar la capa raster en pantalla y de realizar consultas sobre ellas.

- El número de celdas es $n_c * n_f$.
- Normalmente r_x y r_y son constantes y van a tener el mismo valor. En algunos casos no es así, por ejemplo en el caso de imágenes de satélite de baja resolución que se ven afectadas por la curvatura de la Tierra.
- Las ecuaciones para obtener las coordenadas del centro de cada celda partiendo de su número de columna y de fila son:

¹En algunos sistemas que no contemplan el valor $NULL$, se utiliza el cero, aunque evidentemente que una variable tenga valor cero es muy distinto a que no tenga valor

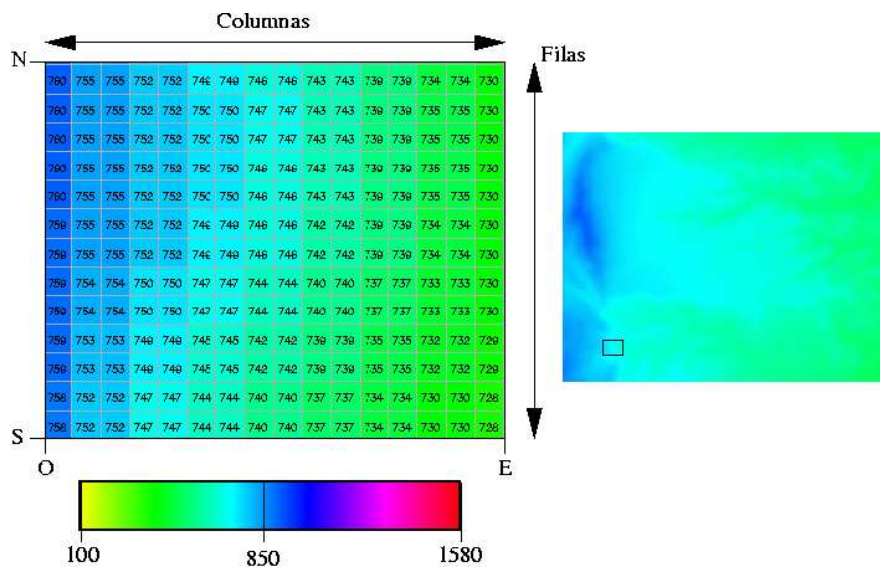


Figura 4.1: Modelos digitales. Codificación de una variable cuantitativa en formato raster

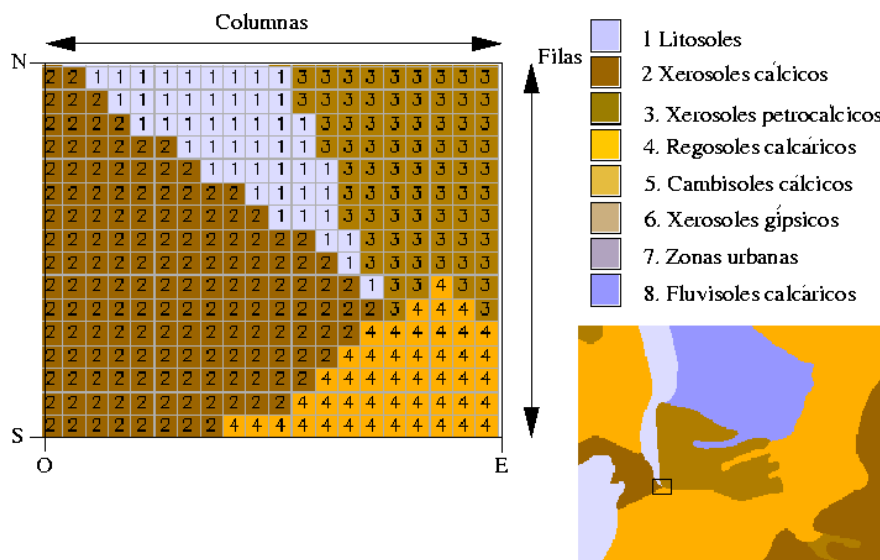


Figura 4.2: Modelos digitales. Codificación de una variable cualitativa en formato raster

$$X = w + (c + 0,5) * r_x \quad (4.1)$$

$$Y = s + (n_f - f + 0,5) * r_y \quad (4.2)$$

- A la inversa pueden obtenerse el número de columna y de fila a partir de los valores de X e Y:

$$c = \text{int}(X - w) / r_x \quad (4.3)$$

$$f = n_f - (\text{int}(Y - s) / r_y) \quad (4.4)$$

Estas dos últimas ecuaciones asumen además la convención de que la primera fila y la primera columna son la fila 0 y la columna 0.

- Sabiendo el número de fila y columna de una celdilla es también sencillo obtener el número de orden del pixel:

$$N = n_c * f + c \quad (4.5)$$

de manera que la primera celdilla es la celdilla número cero y la última la $(n_c * n_f) - 1$.

Ejemplos

En las figuras ?? y ?? aparecen dos ejemplos de capa raster. Asumiendo que la esquina inferior izquierda de las áreas en detalle coinciden con el punto de coordenadas X=632000 Y=4200000 y que en ambos casos el tamaño de la celdilla es de $r_x = r_y = 25\text{metros}$, calcula:

1. Los valores de n , s , e y w para cada una de las áreas en detalle **Solución:**
 - En ambos casos $s = 4200000$ y $w = 632000$ ya que la esquina inferior izquierda coincide con el límite SW
 - En la figura ??: $n_f = 13$, $n_c = 16$, por tanto $n = s + n_f * r_y = 4200325$ y $e = w + n_c * r_x = 632400$
 - En la figura ??: $n_f = 16$, $n_c = 20$, por tanto $n = s + n_f * r_y = 4200400$ y $e = w + n_c * r_x = 632500$
2. La altitud y el tipo de suelo del punto de coordenadas X=632120 Y=4200080 **Solución:**
 - En primer lugar hay que calcular el número de fila y de columna. Para la figura ??:
 $f = n - \text{int}(Y - s) / r_y = 10$
 $c = \text{int}((x - w) / r_x) = 4$
 Para la figura ??:
 $f = n - \text{int}(Y - s) / r_y = 13$
 $c = \text{int}((x - w) / r_x) = 4$
 - Posteriormente basta con buscar en las matrices de datos representadas en ?? y ?? recordando que la primera fila y la primera columna son cero. Por tanto altitud=749, suelo=Xerosoles cálcicos.

4.2. Visualización y consulta de capas raster

La pantalla de un ordenador guarda una evidente semejanza con un capa raster. Se trata de un espacio rectangular dividido en píxeles² que pueden adoptar diversos colores. Por tanto, a la hora de mostrar una capa raster en la pantalla del ordenador, el programa deberá establecer una correspondencia entre los límites de la porción de la capa raster visualizada en cada momentos y el número de filas y columnas disponibles en la ventana que va a mostrar la información. De manera que en ocasiones no se pintarán todas las celdillas sino sólo una de cada dos, una de cada tres, etc.; en otros casos una misma celdilla necesitará varios píxeles para pintarse.

Las herramientas de zoom y desplazamiento que tienen todos los programas de SIG requieren un continuo recálculo de estas relaciones ya que la región visualizada varía constantemente.

Algo similar ocurre con las herramientas de consulta interactiva. Para utilizarlas, el usuario debe pinchar en un punto de la pantalla que muestra la capa raster; puesto que el programa conoce la relación entre filas y columnas de la ventana gráfica y las filas y columnas de la región visualizada, se podrá hacer una conversión rápida de unas a otras unidades y a partir de hay obtener la posición en memoria del valor consultado utilizando la ecuación ??.

Por lo que se refiere a la visualización, el usuario puede seleccionar que celdillas quiere ver y cuales no en función de los valores de dichas celdillas en una o más capas raster. Para ello debe definir una **máscara**. Una máscara en una capa raster de variable binomial que contiene sólo unos y *NULL*, de manera que sólo se representarán aquellas celdillas que tengan un valor 1 en la máscara. Una máscara puede definirse por procedimientos de álgebra de mapas (ver tema 5).

4.3. ASCII y binario

Toda la información que se almacena en un ordenador puede encontrarse en formato ASCII o binario, que son los dos formatos básicos de codificación de información en un ordenador. La forma natural, desde el punto de vista de la máquina, es almacenar la información en formato binario; y la forma natural para el usuario es hacerlo en formato ASCII. Ambos sistemas utilizan el byte como unidad mínima de información capaz de almacenar 256 valores diferentes (normalmente números entre 0 y 255).

El formato ASCII (American Standard Code for Information Interchange) se basa en un sistema de codificación que asigna a cada carácter alfanumérico (A-Z, a-z, 0-9) o de control (retorno de carro, paso de línea, etc.) un valor entre 0 y 255. De este modo al almacenar un texto utilizaremos un byte por carácter más algunos bytes de control, el problema surge a la hora de almacenar números.

Cuando almacenamos un número en ASCII lo que se hace es utilizar un byte para almacenar cada una de los dígitos que lo componen (incluyendo un byte para el punto en los decimales) más algún carácter de separación entre números (espacio, coma, retorno de carro). En la figura ?? aparece el carácter que correspondende a cada uno de los 256 valores.

²Aunque a veces se intercambian, es preferible, para evitar confusiones, reservar el término celdilla para los elementos que componen una capa raster y pixel para los elementos en que se divide la pantalla de un ordenador o cualquier ventana gráfica dibujada en ella

Dec	Hx	Oct	Char	Dec	Hx	Oct	Char	Dec	Hx	Oct	Char	Dec	Hx	Oct	Char
0	0	000	NUL (null)	32	20	040	SPACE	64	40	100	Ø	96	60	140	`
1	1	001	SOH (start of heading)	33	21	041	!	65	41	101	A	97	61	141	a
2	2	002	STX (start of text)	34	22	042	"	66	42	102	B	98	62	142	b
3	3	003	ETX (end of text)	35	23	043	#	67	43	103	C	99	63	143	c
4	4	004	EOT (end of transmission)	36	24	044	\$	68	44	104	D	100	64	144	d
5	5	005	ENQ (enquiry)	37	25	045	%	69	45	105	E	101	65	145	e
6	6	006	ACK (acknowledge)	38	26	046	&	70	46	106	F	102	66	146	f
7	7	007	BEL (bell)	39	27	047	'	71	47	107	G	103	67	147	g
8	8	010	BS (backspace)	40	28	050	(72	48	110	H	104	68	150	h
9	9	011	TAB (horizontal tab)	41	29	051)	73	49	111	I	105	69	151	i
10	A	012	LF (NL line feed, new line)	42	2A	052	*	74	4A	112	J	106	6A	152	j
11	B	013	VT (vertical tab)	43	2B	053	+	75	4B	113	K	107	6B	153	k
12	C	014	FF (NP form feed, new page)	44	2C	054	,	76	4C	114	L	108	6C	154	l
13	D	015	CR (carriage return)	45	2D	055	-	77	4D	115	M	109	6D	155	m
14	E	016	SO (shift out)	46	2E	056	.	78	4E	116	N	110	6E	156	n
15	F	017	SI (shift in)	47	2F	057	/	79	4F	117	O	111	6F	157	o
16	10	020	DLE (data link escape)	48	30	060	0	80	50	120	P	112	70	160	p
17	11	021	DC1 (device control 1)	49	31	061	1	81	51	121	Q	113	71	161	q
18	12	022	DC2 (device control 2)	50	32	062	2	82	52	122	R	114	72	162	r
19	13	023	DC3 (device control 3)	51	33	063	3	83	53	123	S	115	73	163	s
20	14	024	DC4 (device control 4)	52	34	064	4	84	54	124	T	116	74	164	t
21	15	025	NAK (negative acknowledge)	53	35	065	5	85	55	125	U	117	75	165	u
22	16	026	SYN (synchronous idle)	54	36	066	6	86	56	126	V	118	76	166	v
23	17	027	ETB (end of trans. block)	55	37	067	7	87	57	127	W	119	77	167	w
24	18	030	CAN (cancel)	56	38	070	8	88	58	130	X	120	78	170	x
25	19	031	EM (end of medium)	57	39	071	9	89	59	131	Y	121	79	171	y
26	1A	032	SUB (substitute)	58	3A	072	:	90	5A	132	Z	122	7A	172	z
27	1B	033	ESC (escape)	59	3B	073	;	91	5B	133	[123	7B	173	{
28	1C	034	FS (file separator)	60	3C	074	<	92	5C	134	\	124	7C	174	
29	1D	035	GS (group separator)	61	3D	075	=	93	5D	135]	125	7D	175	}
30	1E	036	RS (record separator)	62	3E	076	>	94	5E	136	^	126	7E	176	~
31	1F	037	US (unit separator)	63	3F	077	?	95	5F	137	_	127	7F	177	DEL
												128	Ç	144	É
												129	ù	145	æ
												130	é	146	Æ
												131	â	147	ó
												132	ä	148	ö
												133	à	149	ò
												134	â	150	û
												135	ç	151	ù
												136	ê	152	–
												137	ë	153	Û
												138	è	154	Ü
												139	ï	156	£
												140	î	157	¥
												141	ì	158	–
												142	Ë	159	ƒ
												143	Ä	192	Ł
														175	»
														176	☐
														177	☐
														178	☐
														179	
														180	†
														181	‡
														182	‡
														183	¶
														184	¶
														185	‡
														186	‡
														187	¶
														188	¶
														189	¶
														190	‡
														191	‡

Figura 4.3: Código ASCII

Por ejemplo el número **274.96** utilizaría 6 bytes [50][55][52][46][57][54] y la palabra **SIG 3** [83][73][71].

La codificación ASCII tiene la ventaja fundamental de que puede leerse con cualquier editor o procesador de textos o con la orden type de MS-DOS o cat de Unix. Proporciona a los usuarios menos avezados una sensación de seguridad, que en algunos casos está fundamentada. Sin embargo tiene tres desventajas fundamentales:

- Los ficheros ASCII que contienen números son mucho más largos (de dos a cuatro veces mayores) lo que es especialmente relevante para el tipo de ficheros (muy grandes) que suelen manejarse en un SIG.
- Es mucho más lento, primero por ser más largo y segundo porque el ordenador debe transformar la cadena de dígitos que ha leído en un número.
- En el modelo raster, el formato binario permite acceso directo a los datos ya que, puesto que todos los datos de un fichero tienen el mismo tamaño en bytes, podemos saber en que punto del fichero se encuentra un determinado dato con sólo conocer su fila y columna. El formato ASCII solo permite acceso secuencial, uno por uno, ya que no todos los datos tienen el mismo tamaño en bytes y hay que buscarlo en el fichero.

En el formato binario se establece en primer lugar el número de bytes que van a hacer falta para almacenar una variable. Esto incluye saber cual será el valor mínimo, máximo y la precisión de los decimales en caso de valores no enteros. Si en un byte pueden almacenarse 256 valores distintos, el número de valores que pueden almacenarse en n bytes se obtiene como 256^n . De este modo en el formato raster, todas las celdillas ocuparán el mismo número de bytes.

4.4. El modelo digital raster de GRASS

La implementación del formato raster (modelo lógico) es algo diferente de unos programas a otros (modelos digitales). A continuación, como ejemplo se analizan los modelos digitales para datos raster de GRASS.

En GRASS se utilizan varios ficheros para almacenar una sola capa raster. Tienen todos el mismo nombre y, en lugar de distinguirse por la extensión, se distinguen por el directorio en el que se almacenan dentro de un determinado MAPSET.

- La matriz de datos, es decir los números que contienen las diferentes celdillas se almacenan en un fichero contenido en el directorio **/cell** (formato binario)
- La información necesaria para leer y ubicar espacialmente los datos contenidos en el fichero anterior se almacenan en un fichero en el directorio **/cellhd** (formato ASCII).
- En el directorio **/cats** se guardan las etiquetas de texto en caso de que la variable sea cualitativa (formato ASCII).

- El fichero que cuelga del directorio **/colr** guarda información acerca de la paleta de colores que se va a utilizar para representar el mapa (formato ASCII) Es muy parecido al de IDRISI, pero ahora los valores de intensidad oscilan entre 0 y 255.
- Los directorios **/fcell**, **/g3dcell** o **/cell_misc** contienen información menos relevante o demasiado compleja para abordarla aquí.

4.5. Bibliografía

- De Meers, M.N. (2002) *GIS Modeling in Raster* John Wiley & sons, Chichester 203 pp.
- Tomlin, C.D. (1990) *Geographic Information Systems and Cartographic Modeling* Prentice Hall, 249 pp.
- Tomlin, C.D. (1991) Cartographic modeling en D.J. Maguire; M.F. Goodchild and D.W. Rhind (Eds.); 1991 *Geographical Information Systems: Principles and Applications*, John Wiley & sons, Chichester (www.wiley.co.uk/wileychi/gis/resources.html)