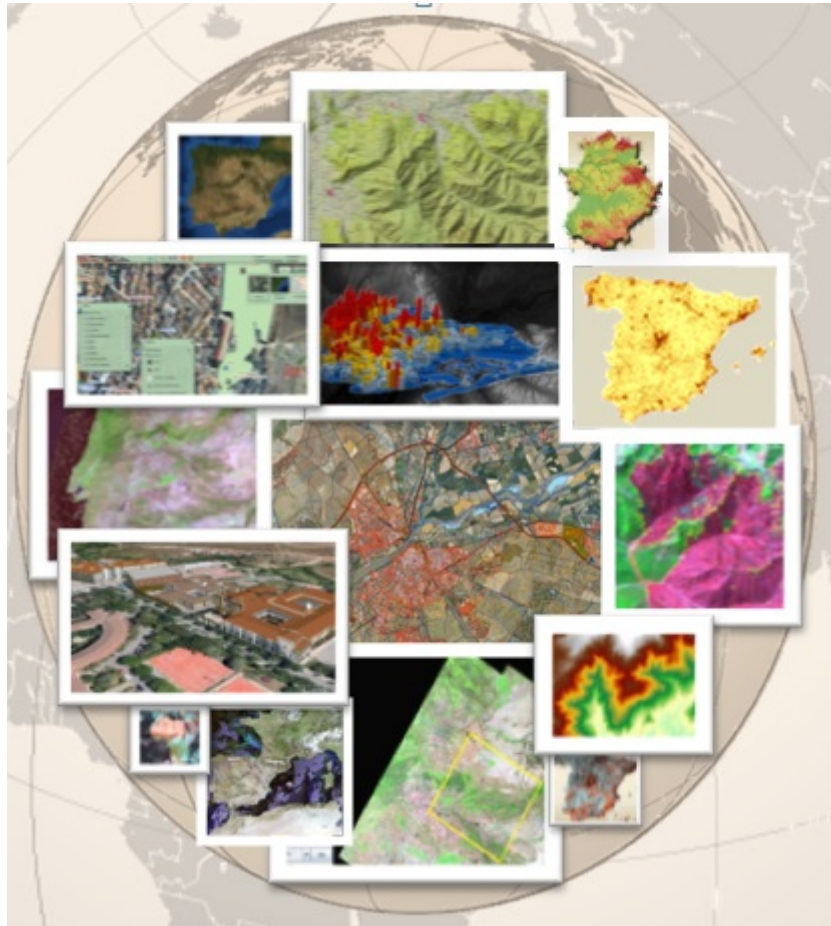


Máster en Tecnologías de la Información Geográfica: SIG y Teledetección



SIG y Teledetección

Tema 4. Control del error

Contenido

1. Introducción	2
1.1. Error, Precisión y Exactitud	2
2. Fuentes de error en la información geográfica.....	2
3. Exactitud posicional	3
4. Exactitud temática.....	4
4.1 Tipos de muestreo.....	5
4.2 Estadísticos asociados a la Matriz de Confusión	6
5. Control de Calidad de la Clasificación mediante la matriz de confusión clásica y sus estadísticos derivados.....	7
5.2. Construcción de la matriz de confusión.	7
5.3. Cálculo de los estadísticos derivados de la matriz de confusión.....	13
6. Bibliografía.....	15

1. Introducción

La verificación de resultados permite al usuario valorar su grado de acuerdo con la realidad, o, dicho de otro modo, el riesgo que asume al tomar decisiones sobre la información que ha generado.

La verificación de resultados requiere que comparemos el producto final con una fuente externa, de mayor fiabilidad. El proceso de verificación requiere aplicar técnicas de muestreo que nos permitan estimar, con la mayor precisión y el menor coste posible, el error que contienen nuestros resultados.

No es extraño observar la ausencia de cualquier mención a la exactitud en la información geográfica. Tampoco es frecuente cuestionar los resultados de un análisis, ni fácil comprobar la bondad de unos resultados cuando no se presentan estadísticos relativos al error de los datos y operaciones. Por lo tanto, es cada vez más necesario incorporar lo que se ha llamado meta-información o metadatos en los productos SIG: información sobre la información, de la cual la referente al error es uno de los elementos de mayor importancia

1.1. Error, Precisión y Exactitud

El **error** es la discrepancia existente entre el valor real (puede ser un valor de posición, de un atributo, o cualquier otro), y el valor recogido en una capa de información geográfica.

La **precisión** hace referencia a la medida de las descripciones en las bases de datos de un SIG. Los atributos de información precisos pueden especificar las características de los elementos con gran detalle. Es importante observar, no obstante, que los datos precisos - no importando el cuidado en su medida - pueden ser inexactos.

La **exactitud** es el grado en el cual la información de un mapa o en una base de datos digital es verdadera. La exactitud se refiere a la cualidad de los datos y al número de errores contenidos en un conjunto de datos o mapa.

Los errores de cualquier fuente de datos geográfica pueden ser separados en dos categorías:

- **Errores posicionales:** Afectan a los modelos vectoriales, y son los errores en la situación en el plano XY
- **Errores atributivos o temáticos:** suponen una asignación imprecisa de la categoría o valor de la información. Afectan tanto a los modelos vectoriales o ráster.

2. Fuentes de error en la información geográfica

Las posibles fuentes de error en la información geográfica pueden ser:

- **Fuente de información:** en el ámbito de los SIG, es muy común usar fuentes de información procedentes de terceros. Estas fuentes son externas

a nuestra organización y poseerán más o menos error, en función de su antigüedad, cobertura espacial, escala, etc.

- **Operario:** La persona que realiza los trabajos, puede tener ciertas tendencias en su modo de operar o trabajar de una forma más o menos meticulosa, que pueden introducir errores
- **Método:** El método seguido, tiene un rigor que puede influir en el resultado. Además, se pueden introducir errores muchas veces según el orden que se hagan las operaciones en un SIG
- **Instrumental:** Tanto las herramientas físicas como las lógicas a la hora de procesar los datos (por ejemplo, los algoritmos de cálculo) pueden introducir más o menos error.

Por ejemplo, en todos los procesos que el dato geográfico interviene, están sujetos a una posible introducción de error:

Proceso	Motivo
Modelización conceptual	Errores en el modelo
Recogida de datos	Errores en los trabajos de campo Errores en las fuentes de información utilizadas
Captura de datos	Inexactitud en la digitalización Inexactitud inherente a los elementos geográficos
Almacenamiento	Insuficiente precisión numérica y/o espacial Errores de procesamiento
Manipulación	Intervalos de clase inapropiados Errores de superposición Propagación de errores
Salidas cartográficas	Errores en la transformación de coordenadas Inexactitud escala Inexactitud del dispositivo de salida
Uso de los resultados	Entendimiento incorrecto Uso inapropiado

3. Exactitud posicional

En la información geográfica, la posición de un fenómeno del mundo real se describe por medio de coordenadas en un determinado sistema de referencia. La exactitud posicional significa la mayor o menor proximidad entre las posiciones de los fenómenos representados y las posiciones consideradas como buenas o verdaderas. La componente posicional queda recogida como un elemento de calidad en la Norma ISO 19113.

Dado que es la componente más tradicional del dato geográfico, existen numerosas metodologías para cuantificar la exactitud posicional. Algunos ejemplos son los test: NMAS, EMAS, NSSDA,...)

Uno de los más extendidos es el NSSDA, que consta de los siguientes pasos:

1. Seleccionar una muestra de n puntos, siendo $n \geq 20$.

2. Calcular el error de cada punto en cada componente:

$$e_{x_i} = x_{t_i} - x_{m_i} \quad e_{y_i} = y_{t_i} - y_{m_i} \quad e_{z_i} = z_{t_i} - z_{m_i}$$

3. Calcular el error medio cuadrático de cada componente:

$$EMC_X = \sqrt{\frac{\sum e_{x_i}^2}{n}} \quad EMC_Y = \sqrt{\frac{\sum e_{y_i}^2}{n}} \quad EMC_Z = \sqrt{\frac{\sum e_{z_i}^2}{n}}$$

4. Obtener el valor $NSSDA_H$ horizontal según una de las siguientes expresiones:

$$\text{si } EMC_X = EMC_Y \quad \left| \quad NSSDA_H = 1,7307 \cdot EMC_r = 2,4477 \cdot EMC_X \right.$$

$$\left. \text{donde: } EMC_r = \sqrt{EMC_X^2 + EMC_Y^2} \right.$$

$$\text{si } EMC_X \neq EMC_Y$$

y

$$0,6 < (EMC_{min} / EMC_{max}) < 1,0$$

$$NSSDA_H = 2,4477 \cdot 0,5 \cdot (EMC_X + EMC_Y)$$

5. Obtener el valor $NSSDA_Z$ vertical según la siguiente expresión:

$$NSSDA_Z = 1,9600 \cdot EMC_Z$$

6. Incluir junto a la leyenda del mapa:

“Se ha verificado una exactitud horizontal de _ metros al 95% de nivel de confianza”.

“Se ha verificado una exactitud vertical de _ metros al 95% de nivel de confianza”.

4. Exactitud temática

La componente temática de una cartografía es fundamental. Esta componente no sólo está presente en la cartografía temática como tal, sino también en un mapa topográfico también existen aspectos temáticos muy importantes para la calidad del producto.

En la citada Norma ISO 19113 se considera la componente temática como un elemento cuantitativo de la calidad, distinguiendo los siguientes subelementos: corrección de la clasificación, corrección de los atributos cualitativos, y corrección de los atributos cuantitativos. A diferencia de lo que ocurría para la componente posicional, y a pesar de la importancia de esta componente, no existe ningún estándar relativo a la evaluación de la exactitud temática.

El proceso de clasificación en teledetección es probablemente una de las formas más comunes de generar mapas temáticos. Un aspecto básico, relativo a la incertidumbre en teledetección, es caracterizar la incertidumbre en dichos mapas temáticos. La forma más convencional de lograrlo ha sido proporcionar una estimación de la exactitud del mapa globalmente, donde la métrica usada comúnmente es determinar el porcentaje de lugares correctamente clasificados. La exactitud en la clasificación de imágenes de teledetección se ha estudiado desde que esta técnica se introdujo por primera vez y, aunque es un tema bien conocido, siempre ha estado relacionado con la [matriz de confusión](#) clásica y por tanto con sus estadísticos derivados.

Por definición, una matriz de confusión es una matriz cuadrada de números en la que se indican en filas y columnas el número de unidades de la muestra asignadas a una categoría en una clasificación, respecto al número de unidades de la muestra asignadas a una categoría en otra clasificación. Normalmente una clasificación es la verdad terreno y la otra es la clasificación que se quiere evaluar.

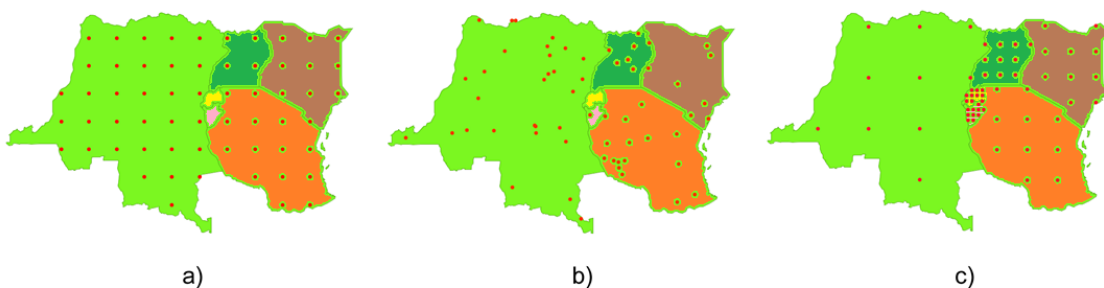
		Observado			
		Clase a	Clase b	Clase c	Clase d
Clasificado	Clase a	VP	FP	FP	FP
	Clase b	FN	VP	FP	FP
	Clase c	FN	FN	VP	FP
	Clase d	FN	FN	FN	VP

La **verdad terreno** se obtiene a partir de datos de campo o de otra clasificación de mayor exactitud. Durante muchos años, la matriz de confusión ha sido el modelo comúnmente usado para valorar la exactitud, no solo de clasificaciones de teledetección, sino también de otras clasificaciones temáticas.

La formación de la matriz de confusión a partir de una muestra requiere de ciertas condiciones:

- Las clases que se establezcan deben ser independientes, mutuamente excluyentes, exhaustivas y en número suficiente.
- Se utilizarán como unidades de muestreo elementos puntuales, superficiales o agrupaciones de aquellos.
- Deben usarse métodos de muestreo estadísticos que eviten la autocorrelación.
- Conviene el uso de métodos estratificados para asegurar la presencia de clases extrañas o minoritarias

4.1 Tipos de muestreo



- Regular:** Este muestreo consiste en una distribución homogénea y regular de puntos en toda la zona de trabajo, sin tener en cuenta la distribución de las clases. Con esta distribución se corre el riesgo de no realizar el control del error en alguna de las clases.
- Aleatorio:** El muestreo aleatorio simplemente sitúa los puntos de muestreo al azar sin obedecer a ninguna condición particular. Si se desconoce la distribución real de clases, un muestreo aleatorio puede ser una buena opción.
- Estratificado:** Requiere la presencia de una variable adicional relacionada. Si esta variable se encuentra zonificada, podemos subdividir el muestreo haciendo uso de las distintas zonas.

En cuanto al número de puntos, existen múltiples teorías para calcular el número de puntos (píxeles) para realizar el control. La más sencilla es la dictada por Congalton en la que se indica que “una guía general o una buena *“regla general”* es planear recolectar

un mínimo de 50 muestras para cada clase de mapa, para mapas de menos de 4000 km² y menos de 12 clases para clasificar.”

4.2 Estadísticos asociados a la Matriz de Confusión

Los estadísticos asociados a dicha matriz de confusión son los siguientes:

a) Exactitud global

La exactitud global es el test más simple y el más comúnmente usado para evaluar las imágenes clasificadas obtenidas por teledetección.

El porcentaje de exactitud (*exactitud global*) se calcula teniendo en cuenta el número de muestras correctamente clasificadas (VP) y el número total de muestras tomadas para el control del error (n):

$$exactitud = \frac{\sum VP}{total} \%$$

Se cuantifica normalmente como un porcentaje.

b) Errores de Omisión y Comisión

Un **error de comisión** es definido simplemente como la inclusión de un área en una clase cuando ésta no pertenece a dicha clase.

Un **error de omisión** es la exclusión de un área en una clase cuando ésta pertenece a ella. Cada error es una omisión de la clase correcta y una comisión de una categoría incorrecta.

c) Exactitudes del Usuario y del Productor

Las exactitudes del productor y del usuario son modos de representar precisiones de categorías individuales en vez de sólo la exactitud de la clasificación total. Así, Ariza (2002) define **exactitud del usuario** como el porcentaje de los elementos de comprobación que realmente están bien clasificados y la exactitud del productor como el porcentaje de elementos que se han clasificado correctamente dentro de la unidad cartográfica equivalente.

Dicho con otras palabras, la **exactitud del productor** es el porcentaje de aciertos que se han producido en la clasificación de la totalidad de elementos que eran de esa clase y la exactitud del usuario es el porcentaje de casos que según la clasificación pertenecen a la clase y de verdad lo son.

	Exactitud usuario	Exactitud productor
Clase a	$\frac{VP_{clase\ a}}{total\ clasificados\ clase\ a}$	$\frac{VP_{clase\ a}}{total\ observados\ clase\ a}$
Clase b	$\frac{VP_{clase\ b}}{total\ clasificados\ clase\ b}$	$\frac{VP_{clase\ b}}{total\ observados\ clase\ b}$
Clase c	$\frac{VP_{clase\ c}}{total\ clasificados\ clase\ c}$	$\frac{VP_{clase\ c}}{total\ observados\ clase\ c}$
Clase d	$\frac{VP_{clase\ d}}{total\ clasificados\ clase\ d}$	$\frac{VP_{clase\ d}}{total\ observados\ clase\ d}$

5. Control de Calidad de la Clasificación mediante la matriz de confusión clásica y sus estadísticos derivados

En primer lugar, vamos a controlar el error de la clasificación mediante la matriz de confusión.

Abriremos un nuevo mapa de Qgis y cargaremos los dos mapas de clasificación obtenidos tanto por el método de Máxima Verosimilitud como por Random Forest. Para ello nos iremos a la carpeta anexa a la imagen *.dim* y buscaremos las imágenes en formato *img*.

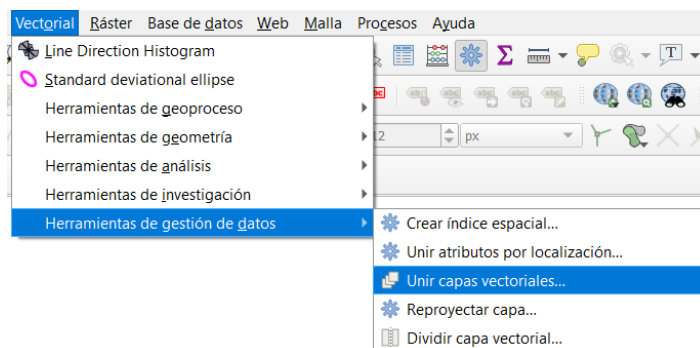
Abriremos tanto la imagen de clasificación como la imagen de con la confianza de la clasificación.

Conviene destacar que las clases de la clasificación no tienen el nombre, sino el valor de la clase que le asignó SNAP. En nuestro caso es el siguiente:

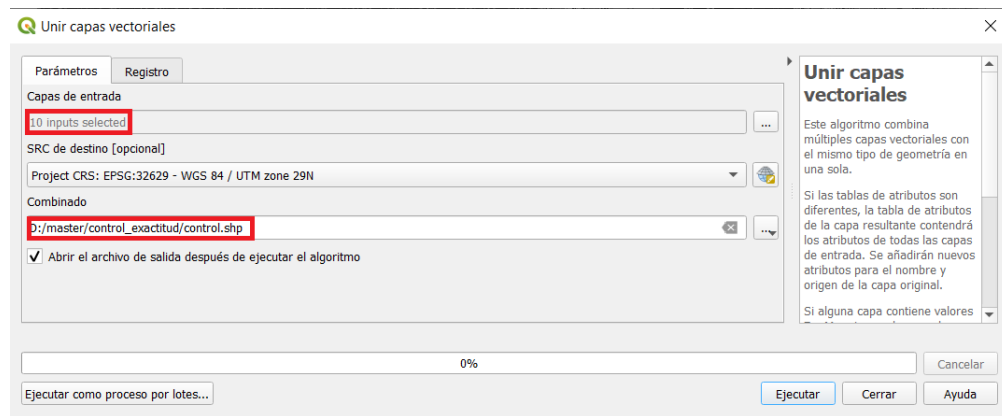
Clase	Código
no data	-1
Pix_AG_entr	0
Pix_CA_entr	1
Pix_FO_entr	2
Pix_IM_entr	3
Pix_OV_entr	4
Pix_PA_entr	5
Pix_PR_entr	6
Pix_PS_entr	7
Pix_TA_entr	8
Pix_ZU_entr	9

5.2. Construcción de la matriz de confusión.

En primer lugar, tendremos que unir todos los shp de los píxeles de control en un único archivo.



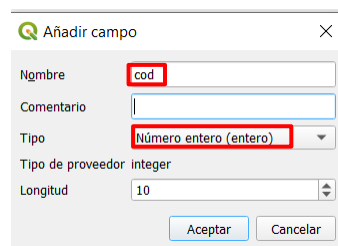
En este caso particular, uniremos sólo las 10 capas que correspondan a las 10 clases empleadas en la clasificación:



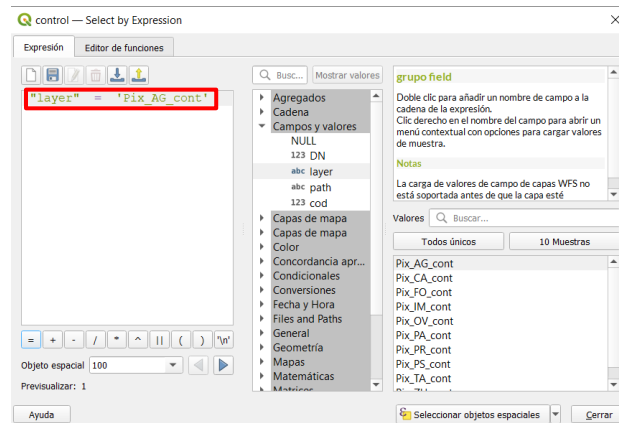
Una vez unidos todos los polígonos en un solo shp, tendremos que añadir, a cada entidad, un campo con la codificación de clases que nos ha salido de SNAP para poder hacer la comparación posterior. Así, según la clasificación, la clase AG (que corresponde con pix_AG_entr) tiene el código 0, la CA el 1 y así sucesivamente:

Clase	Código
no data	-1
Pix_AG_entr	0
Pix_CA_entr	1
Pix_FO_entr	2
Pix_IM_entr	3
Pix_OV_entr	4
Pix_PA_entr	5
Pix_PR_entr	6
Pix_PS_entr	7
Pix_TA_entr	8
Pix_ZU_entr	9

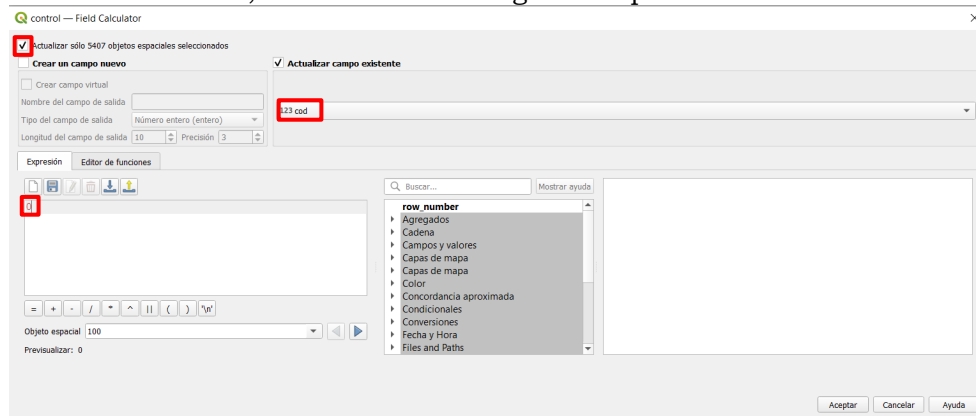
Entonces, abriremos la tabla de atributos y añadiremos un nuevo campo llamado “cod” tipo número entero:



Y ahora iremos seleccionando cada uno de los elementos de cada “layer” y los iremos codificando. (esta operación también se podría hacer usando un condicional encadenado)

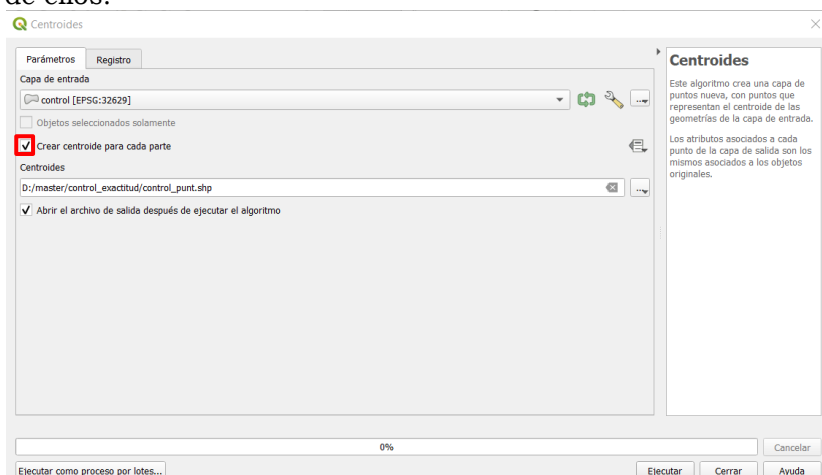


Una vez seleccionados, les daremos el código correspondiente



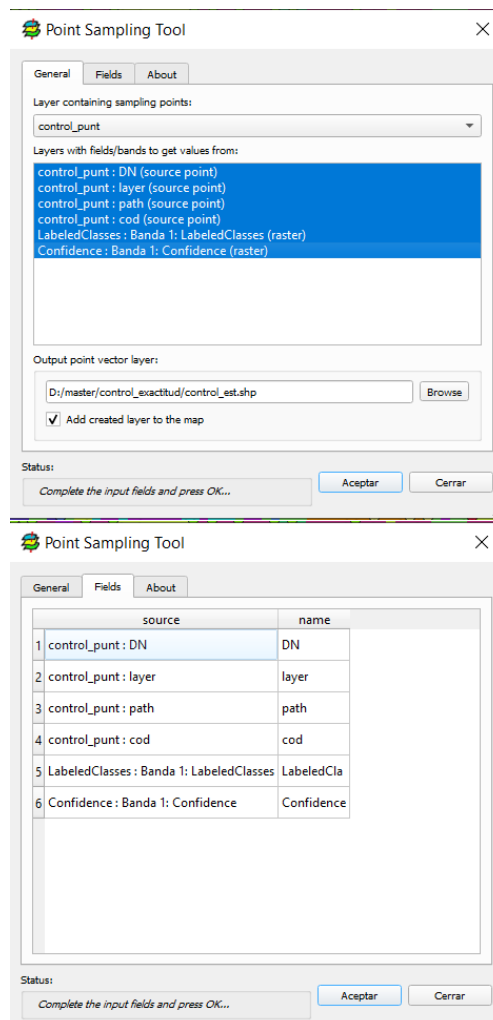
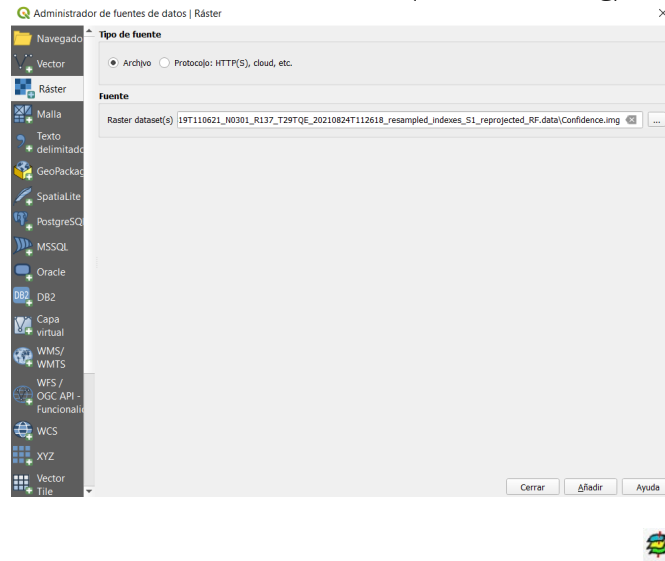
Si actuamos de igual forma con cada una de las clases, tendremos todos los vectores que usaremos como control de la exactitud, con la misma codificación que el ráster de salida de SNAP.

Seguidamente, para poder obtener los valores del ráster de la clasificación de SNAP en cada una de las localizaciones de control, deberemos pasar los polígonos de delimitación de los píxeles a puntos. Para ello calcularemos los centroides de cada uno de ellos.



Una vez creada la capa, procederemos a obtener los valores de la clasificación para así poder comparar y evaluar la exactitud de la clasificación. Vamos a aprovechar esta operación para obtener también, para cada punto, el valor de la confianza de la clasificación. Este valor nos dará una idea de si los errores cometidos lo son por falta de confianza del clasificador o no.

Para ello, cargaremos el ráster de la confianza (en formato img):



Una vez finalizado el proceso, se generará un nuevo shp con dos campos más en los que se indicará, para cada punto la clase clasificada y la confianza de dicha clasificación:

	DN	layer	path	cod	LabeledCla	Confidence
1	916	Pix_AG_cont	D:/master/entr...	0	1,00000	0,37
2	990	Pix_AG_cont	D:/master/entr...	0	4,00000	0,37
3	71	Pix_AG_cont	D:/master/entr...	0	0	1,00000
4	72	Pix_AG_cont	D:/master/entr...	0	0	1,00000
5	472	Pix_AG_cont	D:/master/entr...	0	0	0,5
6	76	Pix_AG_cont	D:/master/entr...	0	0	1,00000
7	68	Pix_AG_cont	D:/master/entr...	0	0	1,00000
8	78	Pix_AG_cont	D:/master/entr...	0	0	1,00000
9	533	Pix_AG_cont	D:/master/entr...	0	6,00000	0,56
10	70	Pix_AG_cont	D:/master/entr...	0	0	1,00000
11	75	Pix_AG_cont	D:/master/entr...	0	0	1,00000
12	76	Pix_AG_cont	D:/master/entr...	0	0	1,00000
13	227	Pix_AG_cont	D:/master/entr...	0	0	1,00000
14	207	Pix_AG_cont	D:/master/entr...	0	0	1,00000
15	72	Pix_AG_cont	D:/master/entr...	0	0	1,00000
16	175	Pix_AG_cont	D:/master/entr...	0	0	1,00000
17	1070	Pix_AG_cont	D:/master/entr...	0	4,00000	0,34
18	76	Pix_AG_cont	D:/master/entr...	0	0	1,00000
19	85	Pix_AG_cont	D:/master/entr...	0	0	1,00000
20	69	Pix_AG_cont	D:/master/entr...	0	0	1,00000
21	67	Pix_AG_cont	D:/master/entr...	0	0	1,00000
22	283	Pix_AG_cont	D:/master/entr...	0	0	0,92

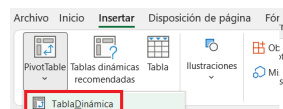
Así podremos ver cómo, por ejemplo, para la inmensa mayoría de los puntos con clasificaciones correctas (el valor de la clasificación y la clase del punto de control es la misma) la confianza ha sido máxima. Sin embargo, bastantes píxeles mal clasificados, lo han sido con un intervalo de confianza muy pequeño.

Procederemos ahora a hacer los cálculos para construir la matriz de confusión. Para ello, calcularemos las estadísticas por categoría de la siguiente forma:

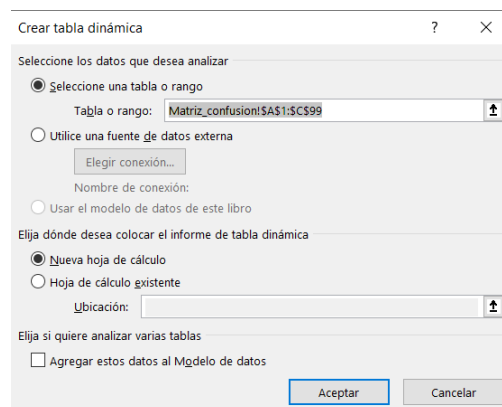
De esa manera se me generará en el directorio indicado, un archivo Excel con mucha información estadística, de la cual sólo nos interesará el recuento de puntos con las distintas combinaciones de valores de clases clasificadas y clases verdaderas (del SIGPAC):

cod	LabeledCla	count	unique	min	max	range	sum	mean	median	skddev	minority	majority	q1
0	1	295	1	1	1	1	0	295	1	1	0	1	1
0	4	56	1	4	4	0	0	224	4	4	0	4	4
0	0	4000	1	0	0	0	0	0	0	0	0	0	0
0	6	257	1	6	6	0	0	1542	6	6	0	6	6
0	2	553	1	2	2	0	0	1106	2	2	0	2	2
0	8	132	1	8	8	0	0	1056	8	8	0	8	8
0	5	66	1	5	5	0	0	330	5	5	0	5	5
0	3	11	1	3	3	0	0	33	3	3	0	3	3
0	9	6	1	9	9	0	0	54	9	9	0	9	9
0	7	31	1	7	7	0	0	217	7	7	0	7	7
1	5	112	1	5	5	0	0	560	5	5	0	5	5
1	1	325	1	1	1	0	0	325	1	1	0	1	1
1	3	19	1	3	3	0	0	57	3	3	0	3	3
1	6	183	1	6	6	0	0	1098	6	6	0	6	6
1	4	133	1	4	4	0	0	532	4	4	0	4	4
1	2	123	1	2	2	0	0	246	2	2	0	2	2
1	0	8	1	0	0	0	0	0	0	0	0	0	0
1	9	21	1	9	9	0	0	189	9	9	0	9	9
1	7	5	1	7	7	0	0	35	7	7	0	7	7
1	8	9	1	8	8	0	0	72	8	8	0	8	8
2	2	730	1	2	2	0	0	1460	2	2	0	2	2
2	4	242	1	4	4	0	0	968	4	4	0	4	4
2	1	104	1	1	1	0	0	104	1	1	0	1	1
2	6	104	1	6	6	0	0	624	6	6	0	6	6
2	0	30	1	0	0	0	0	0	0	0	0	0	0
2	8	32	1	8	8	0	0	256	8	8	0	8	8
2	5	5	1	5	5	0	0	25	5	5	0	5	5
2	7	1	1	7	7	0	0	7	7	7	0	7	7
2	9	1	1	9	9	0	0	9	9	9	0	9	9
3	5	75	1	5	5	0	0	375	5	5	0	5	5
3	6	128	1	6	6	0	0	768	6	6	0	6	6
3	4	109	1	4	4	0	0	436	4	4	0	4	4

Por tanto, eliminaremos todas las columnas que no nos interesan, seleccionaremos las tres columnas con las que queremos formar la matriz y procederemos a pivotar la tabla en otra hoja de la siguiente forma:



	A	B	C
1	cod	LabeledCla	count
2	0	1	295
3	0	4	56
4	0	0	4000
5	0	6	257
6	0	2	553
7	0	8	132
8	0	5	66
9	0	3	11
10	0	9	6
11	0	7	31
12	1	5	112
13	1	1	325
14	1	3	19
15	1	6	183
16	1	4	133
17	1	2	123
18	1	0	8
19	1	9	21
20	1	7	5
21	1	8	9



Crear tabla dinámica

Seleccione los datos que desea analizar

☒ Seleccione una tabla o rango

Tabla o rango:

☐ Utilice una fuente de datos externa

Nombre de conexión:

☐ Usar el modelo de datos de este libro

Elija dónde desea colocar el informe de tabla dinámica

☒ Nueva hoja de cálculo

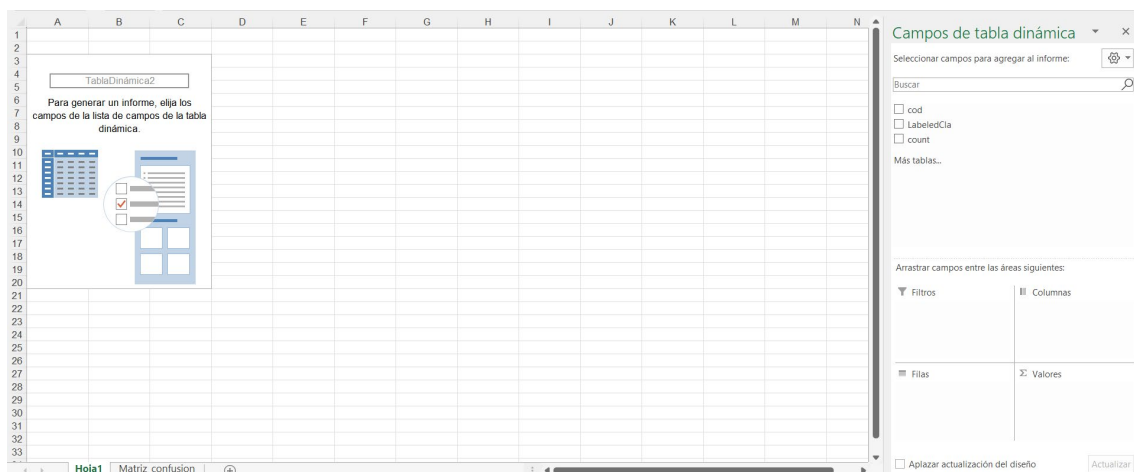
☐ Hoja de cálculo existente

Ubicación:

Elija si quiere analizar varias tablas

☐ Agregar estos datos al Modelo de datos

Y se nos creará una nueva hoja con una tabla dinámica:



TableDinamica2

Para generar un informe, elija los campos de la lista de campos de la tabla dinámica.

Campos de tabla dinámica

Seleccionar campos para agregar al informe:

Buscar

☐ cod

☐ LabeledCla

☐ count

Más tablas...

Añadir campos entre las áreas siguientes:

Filtros

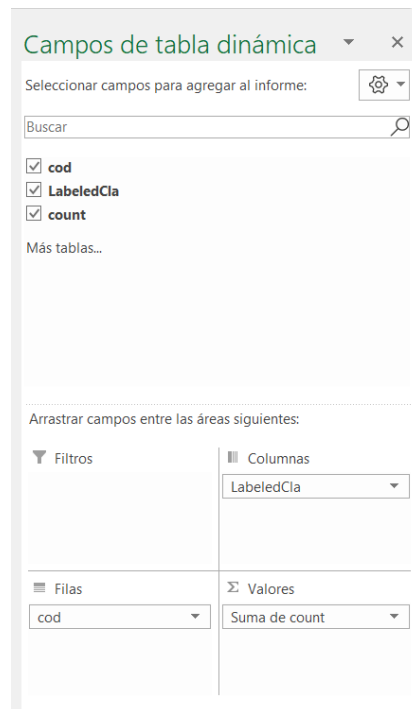
Columnas

Filas

Valores

☐ Aplazar actualización del diseño

En el menú de la derecha configuraremos la tabla de la siguiente forma:



De esa forma tendremos nuestra matriz de confusión con los valores clasificados en las columnas y los valores verdaderos del SIGPAC en las filas.

Suma de count	Etiquetas de columna	0	1	2	3	4	5	6	7	8	9	Total general
Etiquetas de fila		0	1	2	3	4	5	6	7	8	9	
0		4000	295	553	11	56	66	257	31	132	6	5407
1		8	325	123	19	133	112	183	5	9	21	938
2		30	104	730		242	5	104	1	32	1	1249
3		36	221	89	292	109	75	128	40	28	178	1196
4		7	113	87	5	327	82	99	10	32	20	782
5		178	3462	7123	397	3774	17527	10113	181	3544	395	46694
6		1243	2897	6826	470	2214	23241	11707	177	1813	373	50961
7		2	1828	1376	791	908	3944	4374	8808	9231	233	31495
8		10	2491	1680	610	588	832	1175	6338	16012	392	30128
9			53	22	25	35	2	11	2	6	351	507
Total general		5514	11789	18609	2620	8386	45886	28151	15593	30839	1970	169357

5.3. Cálculo de los estadísticos derivados de la matriz de confusión.

Ahora ya podremos realizar los cálculos y presentar la matriz a nuestro gusto.

Verdadero \ Clasificado	0	1	2	3	4	5	6	7	8	9	total
0	4000	295	553	11	56	66	257	31	132	6	5407
1	8	325	123	19	133	112	183	5	9	21	938
2	30	104	730		242	5	104	1	32	1	1249
3	36	221	89	292	109	75	128	40	28	178	1196
4	7	113	87	5	327	82	99	10	32	20	782
5	178	3462	7123	397	3774	17527	10113	181	3544	395	46694
6	1243	2897	6826	470	2214	23241	11707	177	1813	373	50961
7	2	1828	1376	791	908	3944	4374	8808	9231	233	31495
8	10	2491	1680	610	588	832	1175	6338	16012	392	30128
9		53	22	25	35	2	11	2	6	351	507
total	5514	11789	18609	2620	8386	45886	28151	15593	30839	1970	169357

VP de cada clase/total
clasificados cada clase

Cod	Clase	Exactitud usuario	Exactitud productor
0	Agua	73%	74%
1	Viales	3%	35%
2	Forestal	4%	58%
3	Improductivos	11%	24%
4	Olivar	4%	42%
5	Pasto con arbolado	38%	38%
6	Pasto arbustivo	42%	23%
7	Pastizal	56%	28%
8	Tierras arables	52%	53%
9	Zona urbana	18%	69%

VP de cada clase/total
verdaderos cada clase

Suma de todos los VP
(diagonal principal de
la matriz)

Verdaderos positivos:
Exactitud Global:

60079

35%

Total de VP/total puntos

Vemos como, tal y como se explicó en el inicio del tema, en la diagonal principal se sitúan los verdaderos positivos (VP).

Ejercicio propuesto n°2:

Comprobar la exactitud de las otras clasificaciones realizadas.

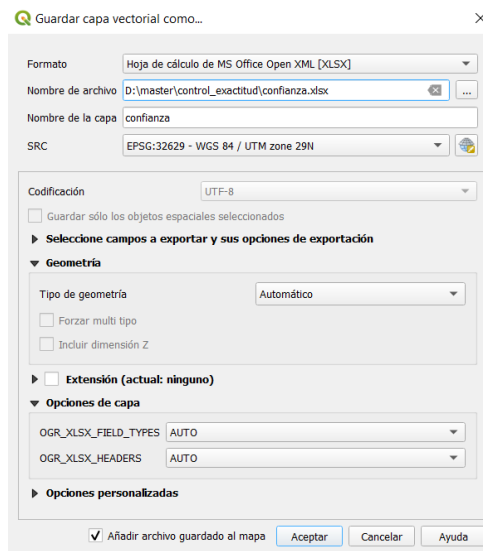
Analizando la exactitud global y las de usuario y productor de cada clase, se puede ver como la clasificación es mejorable.

Las distintas formas de proceder para la mejora serán:

- Uniendo clases, (por ejemplo, la de viales con la de improductivos o eliminando la de viales)
- y/o mejorando los píxeles de entrenamiento (por ejemplo, eliminando los que no estén claros)
- y/o definiendo clases nuevas.

Si analizamos ahora, para cada clase la confianza de la clasificación podremos también obtener alguna conclusión.

Para ello exportaremos a Excel la tabla de atributos de la capa de puntos en la que añadimos la información del nivel de confianza.



Y ya en excel mediante filtros y subtotales para hacer los promedios de cada clase podremos obtener la confianza media de cada clase para hacernos una idea de la clasificación.

H1	=SUBTOTALS(1;F2:F200000)							
	A	B	C	D	E	F	G	H
1	DN	layer	path	cod	LabeledCla	Confidence	Promedio	
2	916	Pix_AG_cont	D:\master\entrena	0	1	0,37		
3	990	Pix_AG_cont	D:\master\entrena	0	4	0,37		
4	71	Pix_AG_cont	D:\master\entrena	0	0	1		
5	72	Pix_AG_cont	D:\master\entrena	0	0	1		
6	472	Pix_AG_cont	D:\master\entrena	0	0	0,5		
7	76	Pix_AG_cont	D:\master\entrena	0	0	1		
8	68	Pix_AG_cont	D:\master\entrena	0	0	1		
9	78	Pix_AG_cont	D:\master\entrena	0	0	1		
10	533	Pix_AG_cont	D:\master\entrena	0	6	0,56		
11	70	Pix_AG_cont	D:\master\entrena	0	0	1		
12	75	Pix_AG_cont	D:\master\entrena	0	0	1		
13	76	Pix_AG_cont	D:\master\entrena	0	0	1		
14	227	Pix_AG_cont	D:\master\entrena	0	0	1		
15	207	Pix_AG_cont	D:\master\entrena	0	0	1		
16	72	Pix_AG_cont	D:\master\entrena	0	0	1		
17	175	Pix_AG_cont	D:\master\entrena	0	0	1		
18	1070	Pix_AG_cont	D:\master\entrena	0	4	0,34		
19	76	Pix_AG_cont	D:\master\entrena	0	0	1		
20	85	Pix_AG_cont	D:\master\entrena	0	0	1		
21	69	Pix_AG_cont	D:\master\entrena	0	0	1		
22	67	Pix_AG_cont	D:\master\entrena	0	0	1		
23	283	Pix_AG_cont	D:\master\entrena	0	0	0,92		
24	172	Pix_AG_cont	D:\master\entrena	0	0	1		
25	1038	Pix_AG_cont	D:\master\entrena	0	4	0,35		
26	69	Pix_AG_cont	D:\master\entrena	0	0	1		
27	73	Pix_AG_cont	D:\master\entrena	0	0	1		
28	89	Pix_AG_cont	D:\master\entrena	0	0	1		
29	71	Pix_AG_cont	D:\master\entrena	0	0	1		
30	89	Pix_AG_cont	D:\master\entrena	0	0	1		
31	430	Pix_AG_cont	D:\master\entrena	0	0	1		
32	830	Pix_AG_cont	D:\master\entrena	0	1	0,42		82,390%

El resultado final de nuestro análisis será el siguiente, pudiendo concluir que el resultado de la clasificación está acorde con el de la certeza con la que el clasificador asegura la pertenencia de cada píxel a cada clase.

Cod	Clase	Exactitud usuario	Exactitud productor	confianza clasificador
0	Agua	73%	74%	82%
1	Viales	3%	35%	46%
2	Forestal	4%	58%	56%
3	Improductivos	11%	24%	51%
4	Olivar	4%	42%	49%
5	Pasto con arbolado	38%	38%	46%
6	Pasto arbustivo	42%	23%	48%
7	Pastizal	56%	28%	54%
8	Tierras arables	52%	53%	62%
9	Zona urbana	18%	69%	55%

6. Bibliografía

Ariza, F. J. (2003). Calidad en la producción cartográfica. *Mapping*, (84), 96-98.

- Asociación de Ingenieros Geógrafos (2008) Temario básico y bibliografía recomendada para la preparación de la Oposición al Cuerpo de Ingenieros Geógrafos.
- Congalton, R. G., & Green, K. (2008). Assessing the accuracy of remotely sensed data: principles and practices. CRC press.
- Chuvieco, E. (2002). Teledetección ambiental: la observación de la tierra desde el espacio (No. G70. 4. C58 2002).
- ISO (1995) Guide to the expression of uncertainty in measurements. International Organization for Standardization. Geneva
- Quirós, E. (2009). Clasificación de imágenes multiespectrales ASTER mediante funciones adaptativas.