



POLITÉCNICA

SISTEMAS DE INFORMACIÓN GEOGRÁFICA

Cuaderno de Prácticas

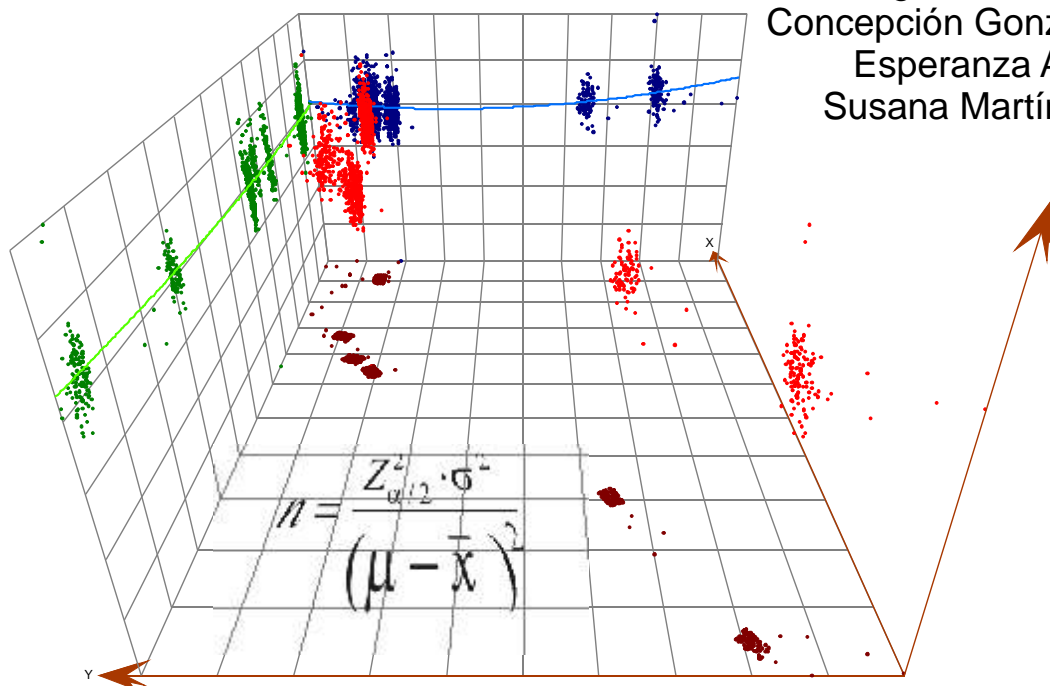
AUTORES:

Rubén Valbuena Puebla
Francisco Mauro Gutiérrez

r.valbuena@upm.es
pmaurogut@hotmail.com

COORDINADORES:

Eugenio Martínez Falero
Concepción González García
Esperanza Ayuga Téllez
Susana Martín Fernández





ÍNDICE

PRÁCTICA 1: OBTENCIÓN DE INFORMACIÓN PARA EL PROYECTO.....	2..
PASO 1: OBTENCIÓN DE CARTOGRAFÍA 1:25.000 DEL INSTITUTO GEOGRÁFICO NACIONAL.	
PASO 2: UNIR LOS ARCHIVOS DESCARGADOS EN UN MISMO MAPA VECTORIAL DE PLANIMETRÍA.....	3
PASO 3: VISUALIZAR UNA ORTOFOTO DESDE EL SERVIDOR ECW DEL SIG OLEÍCOLA.....	4
PASO 4: DESCARGA DE LA ORTOFOTO DE LA ZONA DE ESTUDIO.....	7
PASO 5: VISUALIZACIÓN DE DATOS DESDE UN SERVIDOR GIS.....	8
PASO 6: VISUALIZACIÓN Y DESCARGA DE DATOS VIA WEB.....	9
PRÁCTICA 2: GEORREFERENCIACIÓN POR PUNTOS DE CONTROL.....	10
PASO 1: LOS DATOS DE ENTRADA Y EL ENTORNO DE PARTIDA.....	12
PASO 2: LA BARRA DE HERRAMIENTAS DE GEORREFERENCIACIÓN.....	13
PASO 3: INTRODUCIR PUNTOS DE ENLACE.....	14
PASO 4: CUANTIFICAR LA CALIDAD DEL PROCEDIMIENTO.....	15
PRÁCTICA 2B: MÉTODO ALTERNATIVO PARA LA GEORREFERENCIACIÓN.....	16
PASO 1B: CARGAR LA IMAGEN ESCANEADA EN ARCGIS	
PASO 2B: CREACIÓN DE PUNTOS DE CONTROL A PARTIR DEL MAPA	
PASO 2B_1: CREACIÓN DE UN ARCHIVO VECTORIAL DE PUNTOS DE CONTROL (TERRENO).....	17
PASO 2B_2: INTRODUCCIÓN MANUAL DE LAS COORDENADAS EN LA TABLA LINK TABLE.....	18
PASO 2B_3: CREACIÓN DE ARCHIVO DE GEORREFERENCIACIÓN.....	19
PRÁCTICA 3: ESTADÍSTICA APLICADA AL ANÁLISIS DE IMAGEN DIGITAL.....	21
PASO 1: EXPLORACIÓN Y ESTADÍSTICA DESCRIPTIVA DE LA IMAGEN	
PASO 2: AGRUPAMIENTO DE PÍXELES SEMEJANTES. CLUSTERING.....	23
PASO 3: CLASIFICACIÓN NO SUPERVISADA DE LA IMAGEN.....	24
PASO 4: DECISIÓN SOBRE LAS CLASES A SEPARAR Y RECLASIFICACIÓN.....	25
PASO 5: CÁLCULO DEL TAMAÑO MUESTRAL.....	26
PASO 6: FASE DE ENTRENAMIENTO.....	27
PASO 7: GENERACIÓN DE FIRMAS ESPECTRALES.....	28
PASO 8: CLASIFICACIÓN SUPERVISADA.....	29
PASO 9: VALIDACIÓN MEDIANTE INVENTARIO DE CAMPO	
PASO 10: TAMAÑO MUESTRAL PARA LA VERIFICACIÓN DE RESULTADOS.....	30
PASO 11: MUESTREO ALEATORIO ESTRATIFICADO	
PASO 12: MATRIZ DE CONFUSIÓN. MEDIDA DE LA FIABILIDAD DEL MÉTODO.....	31
PRÁCTICA 4: MÉTODOS DE INTERPOLACIÓN APLICADOS A LA GENERACIÓN DE MODELOS DIGITALES DE TERRENO.....	34
PASO 1: GENERALIDADES SOBRE LOS PROCESOS DE INTERPOLACIÓN ESPACIAL	
PASO 2: INTRODUCCIÓN DE LOS DATOS Y PASOS PREVIOS.....	42
PASO 3: ELABORACIÓN DEL PRIMER TIN	
PASO 4: ANÁLISIS DEL TIN Y DEL GRID GENERADO.....	45
PASO 5: GENERACIÓN DEL TIN CON LÍNEAS DE RUPTURA.....	46
PASO 6: REVISIÓN A OTRAS UTILIDADES DE INTERPOLACIÓN.....	47

PRÁCTICA 3: ESTADÍSTICA APLICADA AL ANÁLISIS DE IMAGEN DIGITAL

PASO 1: EXPLORACIÓN Y ESTADÍSTICA DESCRIPTIVA DE LA IMAGEN

Una imagen fotográfica digital multiespectral es un archivo formado por varias capas raster representando niveles digitales de brillo para una determinada longitud de onda. El color con el que se observa cada pixel se define por un valor numérico de carácter integral. Este número puede ser a su vez representado por un nivel de gris; o en su lugar, por un nivel de color rojo, verde o azul (RGB). La imagen fotográfica final se puede observar por síntesis aditiva de estos tres colores en el monitor del ordenador, o por mezcla sustractiva de cian, magenta y amarillo cuando se traslada la imagen al papel por una impresora.

En esta parte de la práctica haremos una exploración estadística de los datos contenidos en una imagen. Abrimos la ortofoto del monte número 83 del CUP de la provincia de Ávila.

Add Data

//MATERIAL/ORTOFOTO/monte 83.ecw

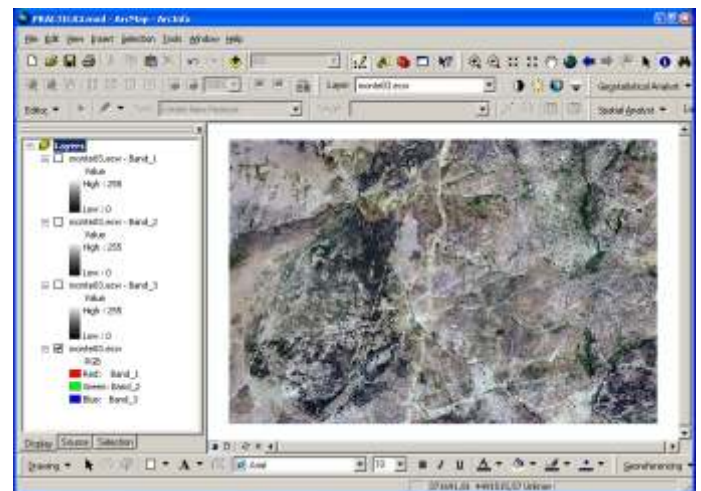
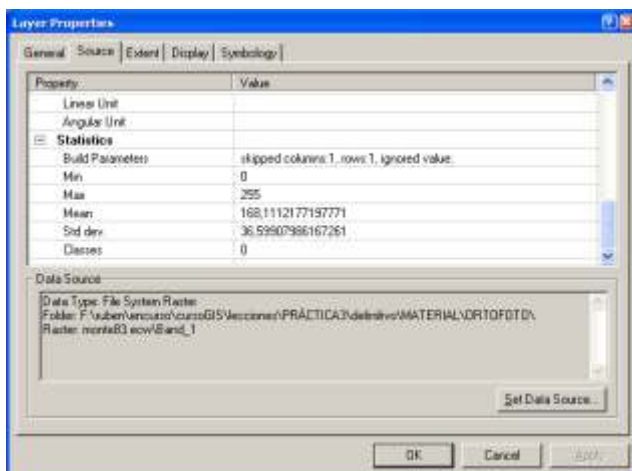
Este es un raster multibanda. Es también posible abrir sus capas por separado. Abrimos los archivos raster correspondientes a las bandas del rojo, verde y azul respectivamente:

Add Data

//MATERIAL/ORTOFOTO/monte 83/band_1

//MATERIAL/ORTOFOTO/monte 83/band_2

//MATERIAL/ORTOFOTO/monte 83/band_3



monte83.ecw - Band_1 > Properties > Pestaña Source

¿En qué rango de valores se encuentran los datos?

¿Cuál es la media y la desviación típica de la banda del rojo? ¿Es una media muestral o poblacional?



Estudiamos la variabilidad e interdependencia de los datos:

ArcToolBox>Spatial Analyst Tools>Multivariate>Band Collection Statistics

Input raster bands = monte83.ecw - banda_1, banda_2 y banda_3

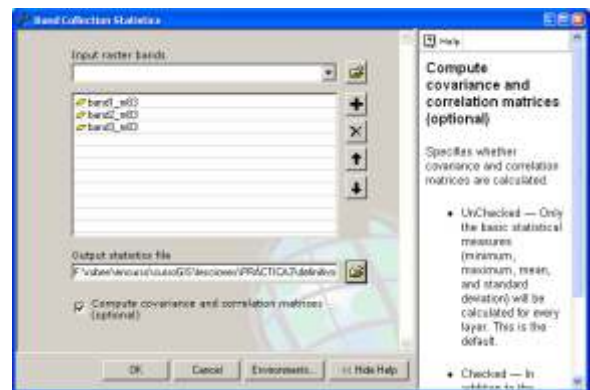
Especificar una ruta para el .txt de salida

Activar checkbox de compute covariance and correlation matrices.

```
# STATISTICS of INDIVIDUAL LAYERS
# Layer MIN MAX MEAN STD
# -----
# 1 0,0000 255,0000 168,1114 36,5994
# 2 0,0000 255,0000 153,9129 30,8240
# 3 0,0000 255,0000 120,6287 26,9161
# =====
```

```
# COVARIANCE MATRIX
# Layer 1 2 3
# -----
# 1 1339,25393 1098,43595 926,35970
# 2 1098,43595 949,93469 791,48466
# 3 926,35970 791,48466 724,33252
# =====
```

```
# CORRELATION MATRIX
# Layer 1 2 3
# -----
# 1 1,00000 0,97386 0,94054
# 2 0,97386 1,00000 0,95417
# 3 0,94054 0,95417 1,00000
# =====
```



¿Qué color es el que produce una mayor variabilidad entre los datos que tenemos?

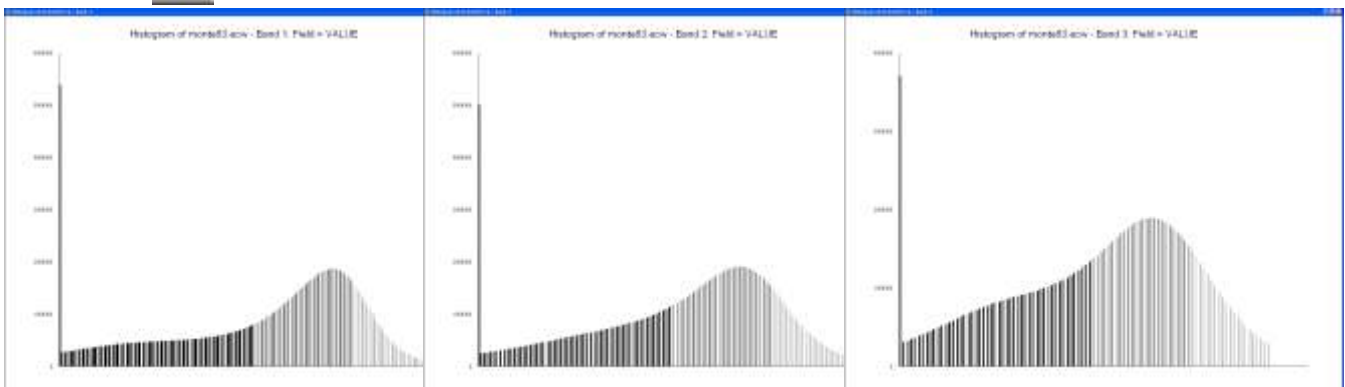
¿Existe correlación significativa entre las bandas?

Podemos observar los histogramas de frecuencias de las 3 bandas:

Spatial Analyst Toolbar

Layer = monte83-ecw - Band_1, Band_2 y Band 3

Histogram



¿Existen diferencias entre las distintas bandas? ¿Cabe esperar que cada banda aporte información distinta sobre las propiedades del terreno o los usos del suelo, o su información es redundante?



PASO 2: AGRUPAMIENTO DE PÍXELES SEMEJANTES. CLUSTERING

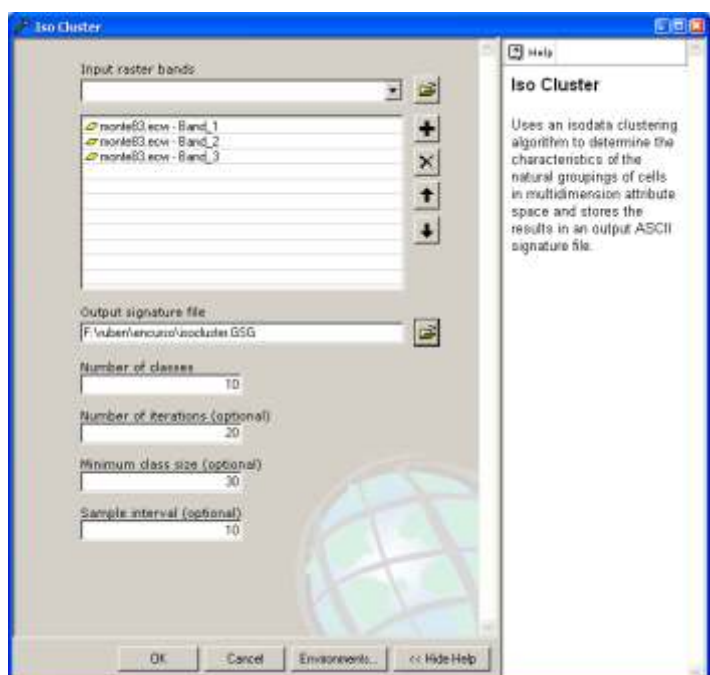
La clasificación de imágenes permite elaborar mapas de usos de suelo de una forma objetiva a partir de la información que contienen las bandas de la imagen. Se trata de agrupar en clases aquellos píxeles que tengan valores parecidos teniendo en cuenta los valores de nivel digital de todas las bandas a la vez. La clasificación no supervisada utiliza herramientas estadísticas que definen en cuántos grupos (*cluster*) se podría clasificar la imagen. En la clasificación supervisada, las clases son definidas previamente por el usuario; el método estudia la separabilidad de las clases definidas y la concurrencia de cada píxel a una clase u otra.

La clasificación no supervisada puede ser útil para un estudio previo de la separabilidad que tienen las distintas zonas de la imagen. Es un estudio estadístico multivariante que analiza las distribuciones de frecuencias de todas las bandas a la vez y agrupa píxeles con valores parejos. Normalmente, las clasificaciones supervisada y no supervisada ofrecen resultados parecidos, por lo que la segunda puede ser utilizada como un estudio preliminar a la primera. En otras palabras, si hay dos zonas concretas que se desea separar en dos clases distintas pero el método no supervisado no lo ha conseguido, significará seguramente que va a ser difícil encontrar una separabilidad significativa entre esas dos clases; de este modo, se puede hacer un estudio *a priori* de la imagen para ver qué tipo de zonas pueden separarse de otras.

En esta clasificación no supervisada que vamos a realizar en esta práctica utilizaremos el método llamado *isocluster* o *isodata*. Este método calcula primero las medias para cada clase según la distribución de frecuencias para después asignar cada píxel a cada clase por algoritmo de mínima distancia. Al ser asignados los píxeles, la media de cada grupo cambia, así que se vuelve a realizar la operación, iterando el algoritmo un número de veces especificado. Normalmente, agrupar en un número mayor de grupos requiere también realizar un número mayor de iteraciones. De esta forma se divide la nube de puntos en grupos pertenecientes al número de clases previamente fijado. El tamaño mínimo de una clase debería ser diez veces el número de bandas que entran en la clasificación. De esta forma se puede lograr que un grupo quede representado apropiadamente. Este método no se realiza con toda la imagen sino con una muestra. Para ello el algoritmo realiza un muestreo sistemático consistente en una malla de puntos equidistantes; la distancia viene fijada por el intervalo de muestreo que se especifica. La salida del programa es un fichero ASCII .gsg de firmas espectrales.

ArcToolBox>Spatial Analyst Tools>Multivariate>Iso Cluster
Input raster bands = monte83.ecw - banda_1, banda_2 y banda_3
Especificar una ruta para el .gsg de salida
Number of classes = 10
Number of iterations = 20
Minimum class size = 30
Sample interval = 10

Es posible observar el archivo de firmas producido abriéndolo con el bloc de notas. En él se pueden consultar las medias de cada clase o grupo para cada banda. Esta sería la posición del centroide del elemento grupo, ya que el algoritmo fue de mínima distancia. También se puede ver cómo quedan las matrices de covarianzas para cada una de las clases



PASO 3: CLASIFICACIÓN NO SUPERVISADA DE LA IMAGEN

En esta fase, a partir de los grupos generados por el muestreo realizado en el isocluster vamos a asignar cada pixel de la imagen a una clase. Para ello utilizaremos el algoritmo de máxima verosimilitud o de máxima probabilidad.

Si la distribución de frecuencias de los niveles digitales de una banda se asemejan a una distribución normal, podemos describir esta clase a partir de su vector de medias y de la matriz de covarianzas. La distribución real de la nube de puntos de una determinada clase se asemeja a esta función, por lo que podemos a su vez calcular la probabilidad para que un pixel determinado se encuentre dentro de esa clase.

Cuando aplicamos el algoritmo de máxima verosimilitud a las clases calculadas por el isocluster, para cada pixel de la imagen se calcula su probabilidad de pertenencia a cada una de las clases, y se asigna a la que tenga máxima probabilidad. La frontera entre las clases es ahora la intersección de las distribuciones de probabilidad de cada clase. El modo de actuar es parecido al del isocluster, aunque en éste la frontera entre las clases venía únicamente determinada por la distancia entre ellas.

Si el solape entre las distribuciones de probabilidad de las clases es excesivo, entonces no es posible separarlas con total fiabilidad. Se producirán más errores de asignación de un pixel a una clase cuanto más solape tenga esta clase con las demás.

Si se conoce *a priori* que una clase tiene una probabilidad mayor que otra, la frontera entre ellas se desplazarán. El programa permite añadir valores de probabilidad *a priori* que pueden ser introducidos si se conoce, por ejemplo la proporción de terreno que van a ocupar estas clases. Esto puede saberse gracias al muestreo realizado por el *isocluster*. De esta forma los valor de las distribuciones de probabilidad de cada clase vendrán ponderados por el número de muestras que fueron asignados a cada clase. Si se detecta la presencia de muchos *outliers* en la imagen se puede además asignar un valor de *reject fraction* que expulse del análisis los niveles digitales atípicos.

ArcToolBox>Spatial Analyst Tools>Multivariate>Maximum Likelihood Classification

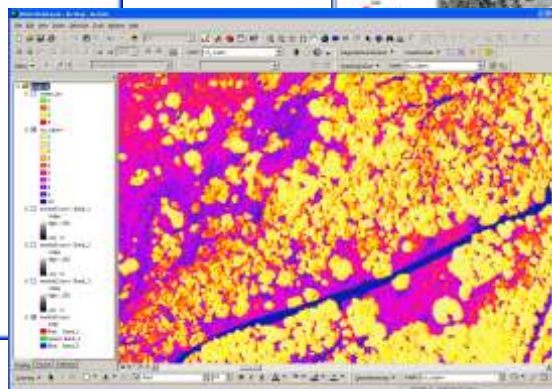
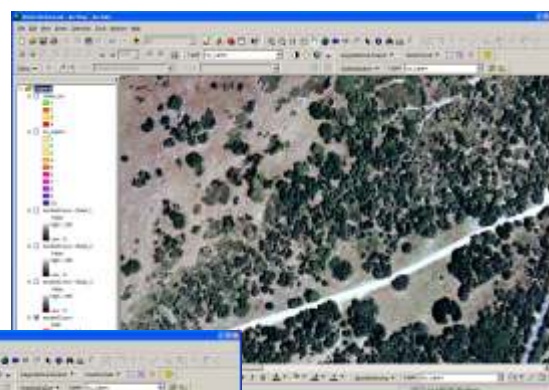
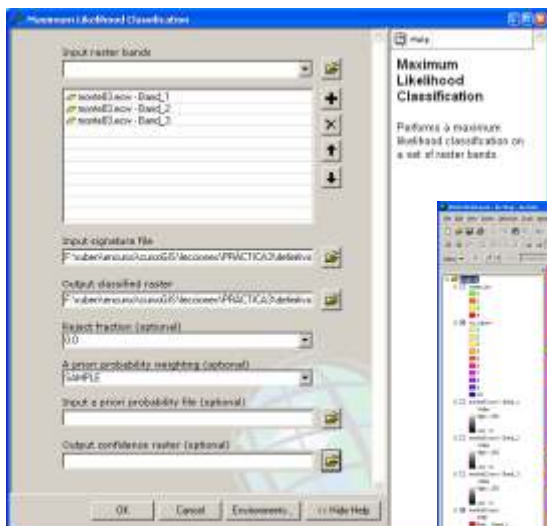
Input raster bands = monte83.ecw - banda_1, banda_2 y banda_3

Input signature file = introducir el .sgs producido en el análisis isocluster

Especificar una ruta para el grid de salida

Reject Fraction = 0.0

A priori probability weighting = sample





Según un análisis visual de los resultados de la clasificación realizada:

¿Qué tipos de suelo parece que se podrán distinguir?

¿Entre qué clases habrá puede haber más confusión?

¿Es posible utilizar este método para extraer las zonas de pasto?

¿Es posible utilizar este método para extraer la presencia de distintas especies forestales?

PASO 4: DECISIÓN SOBRE LAS CLASES A SEPARAR Y RECLASIFICACIÓN

Según el análisis que hemos realizado de los resultados obtenidos, reclasificamos el archivo raster. Las nuevas clases son:

- 1 Forestal
- 2 Matorrales y barda
- 3 Pastos
- 4 Artificial, improductivo y roquedos

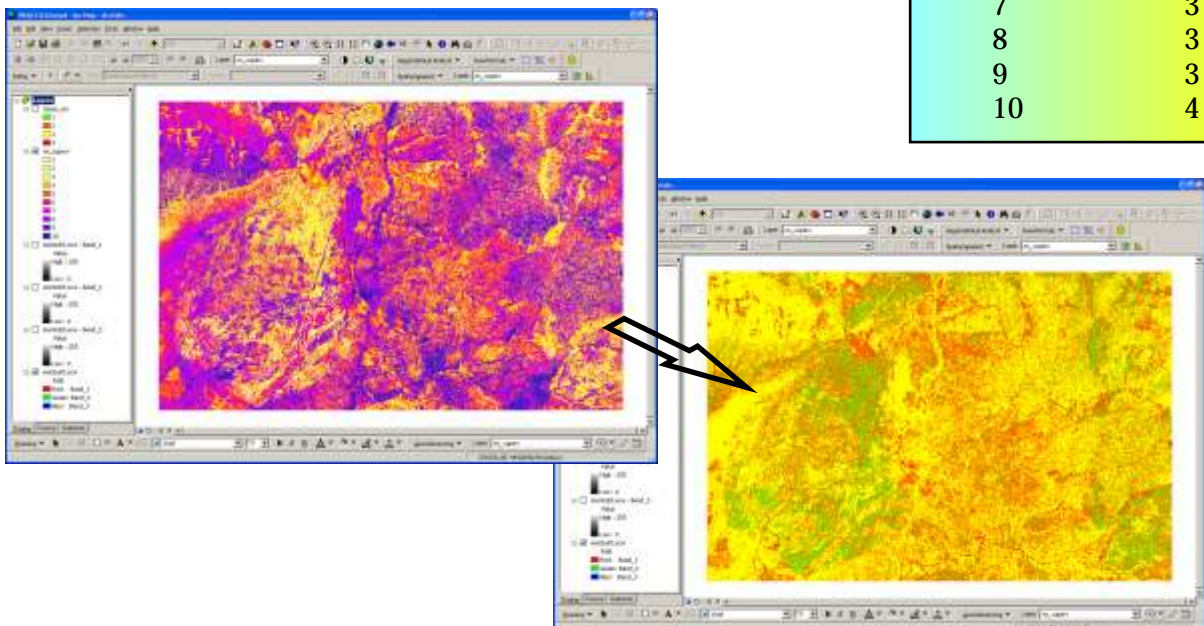
ArcToolBox>Spatial Analyst Tools>Reclass>Reclassify

Input raster = archivo grid de salida de la clasificación no supervisada

Reclass field = value

Especificar una salida para el archivo grid de salida

Old values	New Values
1	1
2	1
3	2
4	2
5	2
6	3
7	3
8	3
9	3
10	4





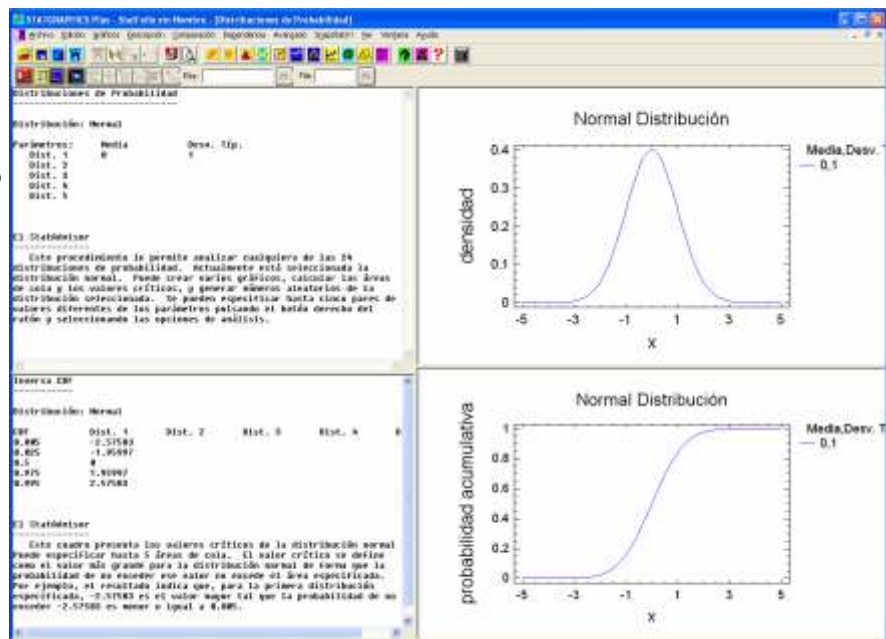
PASO 5: CÁLCULO DEL TAMAÑO MUESTRAL

El tamaño muestral para el tipo de análisis que queremos hacer depende principalmente de dos factores: la propia variabilidad de los datos y la separabilidad de las clases que queremos distinguir. Habría entonces que tomar la máxima variabilidad encontrada en los datos que tenemos y la mínima separabilidad que esperemos que consiga el análisis.

Para calcular el tamaño de muestra normalmente se utiliza la fórmula: $n = \frac{Z^2 \sigma^2}{e^2}$; en la que Z es la abscisa de la curva normal estándar (de media 0 y desviación estándar 1) para un nivel determinado de probabilidad; σ^2 es la varianza de los datos; y e la precisión o amplitud del intervalo requerida. En este sentido, en el análisis de un archivo raster tenemos la ventaja de poder conocer la varianza poblacional real, por lo que utilizaremos entonces la varianza máxima encontrada entre todas las capas de datos que tenemos. La precisión vendrá dada por la separabilidad de las clases, el examen *isocluster* previamente realizado fue precisamente para tener una idea de cómo pueden separarse las clases que hemos decidido.

El riesgo alfa es la probabilidad de aceptar que un pixel se encuentra dentro de una clase no perteneciendo a esta, *i.e.* un falso positivo. Puede tomar valores tales como 0.05, 0.01 o 0.001, según se tome una postura más conservadora o arriesgada. En el estudio de *cluster* previo que hemos realizado, hemos visto que la separabilidad de nuestras clases no es muy buena, por lo que es aconsejable tomar una postura algo conservadora y asumir un riesgo alfa de 0.01. Hay que tener entonces en cuenta que, como nos interesa asumir este riesgo de forma bilateral (igual nos da incluir un falso positivo por defecto que por exceso), buscaremos el valor de la abscisa para un medio de alfa, o sea 0.005. El método para calcular el valor de la abscisa con Statgraphics sería:

- Abrir Statgraphics
- Descripción>Distribuciones>
- Distribuciones de probabilidad
- Normal
- Opciones tabulares>CDF Inverso
- Botón derecho del ratón>
- Opciones de ventana
- Introducir 1 - $\alpha / 2$



¿Qué número de pixeles necesitaremos para realizar las parcelas de entrenamiento?



PASO 6: FASE DE ENTRENAMIENTO

En una clasificación supervisada se definen *a priori* cuántas y cuáles son las clases que se quieren separar. En la fase de entrenamiento, se generan muestras de las que se conoce de antemano su pertenencia a una clase u otra. Normalmente esto se hace mediante trabajo de campo o por fotointerpretación. En este ejercicio tomaremos los resultados de la clasificación no supervisada como punto de partida para la generación de áreas de entrenamiento, ya que la asignación final de las clases se hizo de forma conservadora. No obstante, en las conclusiones hay que tener, por supuesto, en cuenta las decisiones que se tomaron en esta clasificación no supervisada sobre qué elementos del territorio pueden separarse y cuáles no.

Al asignar a unas determinadas muestras puntuales o por áreas o regiones de interés un atributo que define las clases que queremos generar, estamos también asignando a cada clase un media y una varianza para cada una de las bandas que estamos estudiando. De esta forma cada clase queda descrita estadísticamente, y a cada pixel de la imagen se le puede asignar una probabilidad de pertenencia a una clase concreta. En el caso de la clasificación supervisada, la máxima verosimilitud se calculará entonces por la concurrencia de cada píxel a cada clase.

Para la muestra de entrenamiento vamos a realizar un muestreo aleatorio simple. Este muestreo lo realizaremos gracias a una extensión para ArcGIS llamada Hawth's Tools que incluye diversas aplicaciones para ecología. Al cargar el ejecutable en el ordenador aparecerá una barra de herramientas propia con todos las aplicaciones que la extensión incluye. La información y descarga del ejecutable para la extensión puede encontrarse en:

<http://www.spatial ecology.com/htools/overview.php>

El muestreo aleatorio simple se puede realizar de la siguiente manera:

View>Toolbars>Hawth's Tools

Hawth's Tools>Sampling Tools>Generate random points

Input>Raster layer>Select layer = output de la reclasificación

Sample size>Unstratified sampling design>

Generate this number of random points = tamaño muestral calculado

Especificar una ruta para el shape de salida

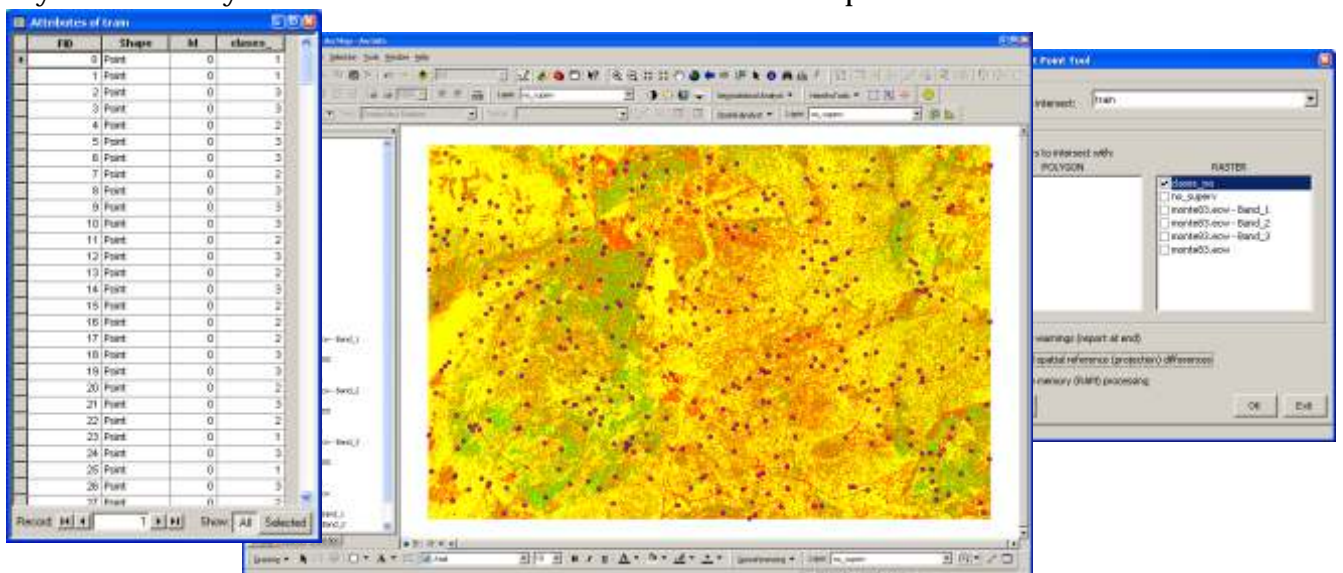


Una vez generados los puntos, hay que especificar a que clase pertenece cada uno. para ello añadimos un campo en la base de datos del vectorial que acabamos de generar con información sobre el valor del raster reclasificado. Para esta operación utilizamos también los Hawth's Tools:

Hawth's Tools>Analysis Tools>Intersect point tool

Input>Point file to intersect = output del muestreo aleatorio

Layers>Select layers to intersect with = activar checkbox del output del reclasificado





PASO 7: GENERACIÓN DE FIRMAS ESPECTRALES

Una vez definidas cuáles son las muestras con las que vamos a entrenar el modelo de asignación de clases, y determinada a qué clase pertenece cada una de ellas, éstas se utilizan para generar la denominada firma espectral de cada clase. La firma espectral es una medida de la tendencia central y la variabilidad conjunta que tiene cada una de las clases para todas las bandas raster de información. Gracias a esa firma espectral, será posible más tarde calcular una distribución de probabilidades de pertenencia a esa clase dentro de un espacio de tantas dimensiones como número de variables raster a estudio. Al igual que en el método no supervisado, el último paso será la clasificación de todos los píxeles a la clase para la que presenten una mayor probabilidad de pertenencia.

Las firmas espectrales se pueden entonces generar de la siguiente manera:

ArcToolBox>Spatial Analyst Tools>Multivariate>Create signatures

Input raster bands = monte83.ecw - banda_1, banda_2 y banda_3

Input raster or feature sample data = output del muestreo aleatorio

Sample field = campo introducido en el vectorial con la información de la reclasificación

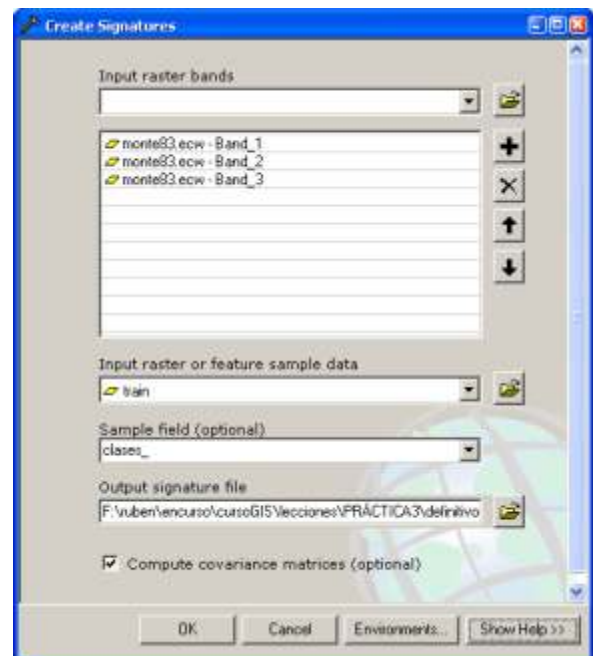
Especificar una ruta para el .gsg de salida

Activar el checkbox de compute covariance matrices

```

firmas_monte83.gsg - Notepad
File Edit Format View Help
# Signatures Produced by ClassSig from
# Class-Grid c:\docume-1\ruben\locals-1\temp\raster10
# and stack c:\docume-1\ruben\locals-1\temp\z_2_236
# Number of selected grids
# 3
# Layer-Number  Grid-name
# 1              z_2_236c1
# 2              z_2_236c2
# 3              z_2_236c3
# Type  Number of Classes  Number of Layers  Number
# -----
# Class ID  Number of Cells  Class Name
# 1         1             36
# Layers   1             2             3
# Means
# Covariance
# 1         95,61111      91,61111      74,55556
# 2         476,64444     390,41587     279,90794
# 3         390,41587     412,47302     277,70794
# 4         279,90794     277,70794     254,36825
# -----
# Class ID  Number of Cells  Class Name
# 2         1             116
# Layers   1             2             3
# Means
# Covariance
# 1         147,57759      137,64655     103,53448
# 2         324,10697     198,17984     134,14078
# 3         198,17984     156,00442     96,43403
# 4         134,14078     96,43403     127,15532
# -----
# Class ID  Number of Cells  Class Name
# 3         1             192
# Layers   1             2             3
# Means
# Covariance
# 1         190,29688      171,25000     135,11458
# 2         192,03182     175,00393     160,49460
# 3         175,00393     181,78010     158,90838
# 4         160,49460     158,90838     186,08104
# -----
# Class ID  Number of Cells  Class Name
# 4         1             8
# Layers   1             2             3
# Means
# Covariance
# 1         222,37500      208,50000     173,12500
# 2         175,12500     219,64286     257,08929
# 3         219,64286     332,00000     374,64286
# 4         257,08929     374,64286     521,83929

```



Los estadísticos descriptivos que se pueden observar en el archivo de firmas espectrales dan una idea de cuál es la tendencia central de cada una de las clases.

También se puede extraer información sobre la separabilidad de estas. La matriz de covarianzas permite construir elipsoides de probabilidad de pertenencia a cada una de las clases. Estos elipsoides se sitúan en un espacio de tantas dimensiones como bandas hay en el análisis.

En este caso el espacio es tridimensional. La matriz de covarianzas define para cada clase un elipsoide de distribución que se usará para la asignación de cada píxel a una clase en la función de máxima verosimilitud.



PASO 8: CLASIFICACIÓN SUPERVISADA

En este paso realizaremos el mismo procedimiento que en la clasificación no supervisada. La diferencia es que en este caso hemos sido nosotros quienes hemos indicado cuáles son las clases que deben definirse, en lugar de ser clases definidas por la propia distribución estadística de los datos. Para ello utilizaremos el archivo de firmas generado para nuestras clases, en lugar del generado en el *isocluster*.

ArcToolBox>Spatial Analyst Tools>Multivariate>Maximum Likelihood Classification

Input raster bands = monte83.ecw - banda_1, banda_2 y banda_3

Input signature file = introducir el .sgs producido en el análisis de firmas espectrales

Especificar una ruta para el grid de salida

Reject Fraction = 0.1

A priori probability weighting = equal

PASO 9: VALIDACIÓN MEDIANTE INVENTARIO DE CAMPO

Para la validación del método empleado vamos a utilizar datos que han sido obtenidos a partir de un inventario de campo, en el que se delimitaron diversas unidades de vegetación. Dentro de estas unidades de vegetación vamos a extraer aquellas que son como atributo nivel 1 de clasificación:

1 - más de 70% de cobertura de Quercus pyrenaica

5 - más del 70% de cobertura de pasto.

Gracias a la extracción de estas áreas realizaremos un muestreo estratificado dentro de ellas para valorar la capacidad de los métodos utilizados para distinguir las áreas forestales y las áreas de pasto.

Add data

//MATERIAL/RODALES/inventario.shp

ArcToolBox>Analysis tools>Extract>Select

Input features = inventario.shp

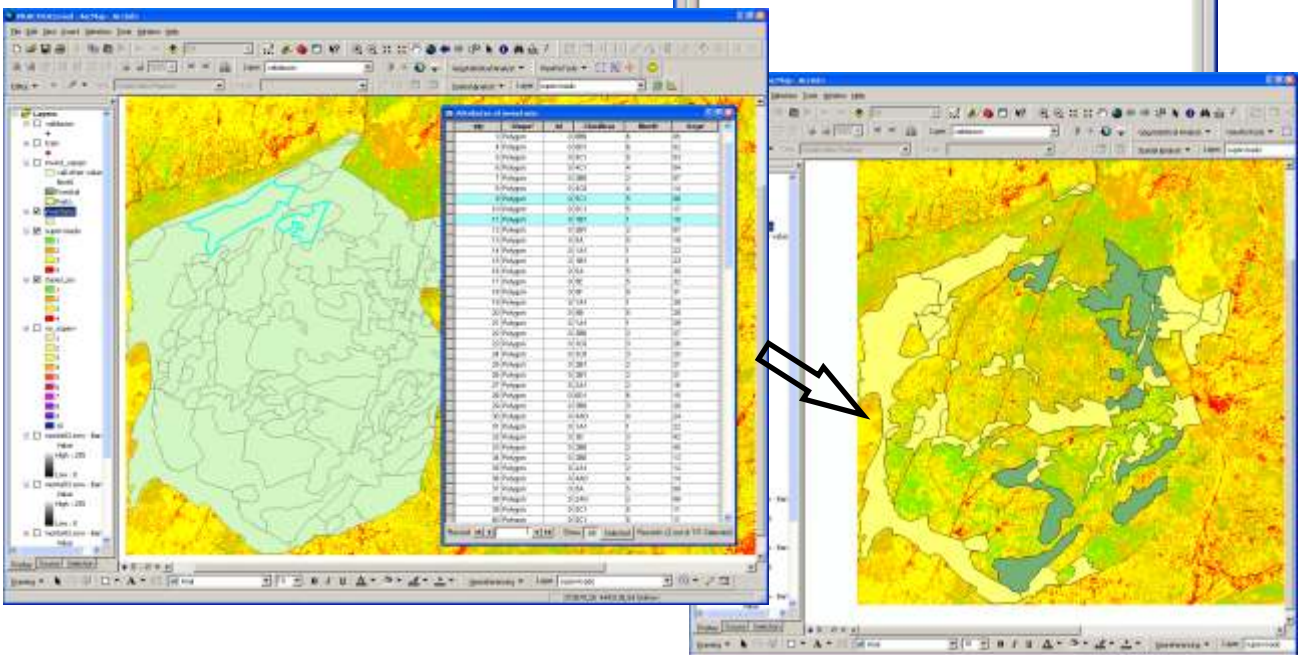
Especificar una ruta para el shape de salida

SQL Expression

Query Builder:

Select * from inventario where =

"Nivel1" = '1' OR "Nivel1" = '5'





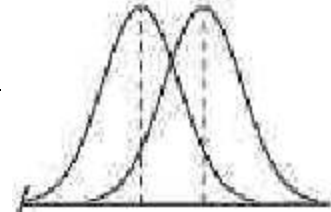
PASO 10: TAMAÑO MUESTRAL PARA LA VERIFICACIÓN DE RESULTADOS

El procedimiento es muy parecido al utilizado en el paso 5, pero ahora se trata de una comparación de medias. La hipótesis alternativa de este estudio es que existe diferencia entre las clases pasto y forestal y el resto de los píxeles de la fotografía. El término precisión se traduce en una diferencia de medias de dos distribuciones prácticamente conocidas.

Una media es la media poblacional de toda la imagen.

La otra media es la tendencia central estimada para una clase en el muestreo de áreas de entrenamiento.

$$n = \frac{Z^2_{/2} \cdot \sigma^2}{(\mu - \bar{x})^2}$$

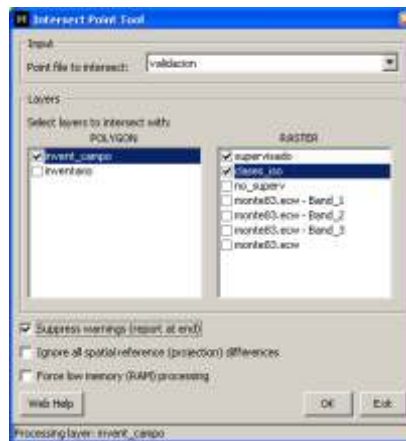


Hallaremos el tamaño muestral para la clase pasto y para la clase forestal, y tomaremos el mayor de ellos.

¿Qué número de píxeles por clase necesitaremos para realizar la valoración del método?

PASO 11: MUESTREO ALEATORIO ESTRATIFICADO

Para la validación entonces tomaremos muestras dentro de los polígonos que han sido declarados como pasto o como forestal en el inventario. Para ello realizaremos un muestreo estratificado en el que tomaremos n muestras dentro del polígono pasto y n muestras dentro del forestal.



Hawth's Tools > Sampling Tools > Generate random points

Input > Polígono layer > Select layer = output de polígonos seleccionados de forestal y pasto

Sample size > Stratified sampling design > Polygon unique ID field = Nivel1

Generate this number of points per polygon = tamaño muestral

Especificar una ruta para el shape de salida

Hawth's Tools > Analysis Tools > Intersect Point Tool

Input > Point file intersect = output del shape de muestra para validacion

Layers > select layers to intersect with = activar:

vectorial con polígonos de pasto y forestal extraídos

raster final de la clasificación no supervisada

raster final de la clasificación supervisada

seleccionar el campo nivel1 de los atributos del archivo vectorial



FID	Shape	Id	Nivel1	supervi	clases_	realidad	supervis	nosuperv
442	Point	0	1		1	forestal	forestal	forestal
443	Point	0	1		2	forestal	otros	otros
444	Point	0	1		2	forestal	otros	otros
445	Point	0	1		1	forestal	forestal	forestal
446	Point	0	1		1	forestal	forestal	forestal
447	Point	0	1		2	forestal	otros	otros
448	Point	0	1		3	forestal	pasto	pasto
449	Point	0	1		2	forestal	forestal	otros
450	Point	0	5		3	pasto	pasto	pasto
451	Point	0	5		2	pasto	otros	otros
452	Point	0	5		3	pasto	pasto	pasto
453	Point	0	5		3	pasto	pasto	pasto
454	Point	0	5		2	pasto	pasto	otros
455	Point	0	5		2	pasto	otros	otros
456	Point	0	5		3	pasto	pasto	pasto
457	Point	0	5		3	pasto	otros	pasto
458	Point	0	5		3	pasto	pasto	pasto
459	Point	0	5		4	pasto	otros	otros

PASO 12: MATRIZ DE CONFUSIÓN. MEDIDA DE LA FIABILIDAD DEL MÉTODO

Por medio de la validación tratamos de comprobar la exactitud de los píxeles. Como hemos podido observar, la metodología es parecida a la de la fase de entrenamiento, pero el objetivo es el contrario: ahora se estudian como se adecuan los píxeles clasificados a unas regiones de interés de las que sabemos a qué clase pertenecen.

Para ello debemos generar una matriz de confusión. Este paso lo realizaremos exportando la tabla del archivo vectorial de validación a una hoja de cálculo de Excel.

En primer lugar realizamos una reclasificación de los campos del archivo vectorial de forma que tengan más sentido en la matriz de confusión como variables cualitativas:

Capa de puntos de validación>Open attribute table
Options>Add field

Generar 3 nuevos campos:

- Name = realidad, supervis y nosuperv
- Type = text
- Length = 10

Campo realidad>Calculate values>Activar checkbox Advanced>

Pre-Logic VBA Script Code:

```
dim output as string
if [Nivel1]=1 then output="forestal"
if [Nivel1]=5 then output="pasto"
realidad=output
```

Campo supervis>Calculate values>Activar checkbox Advanced>

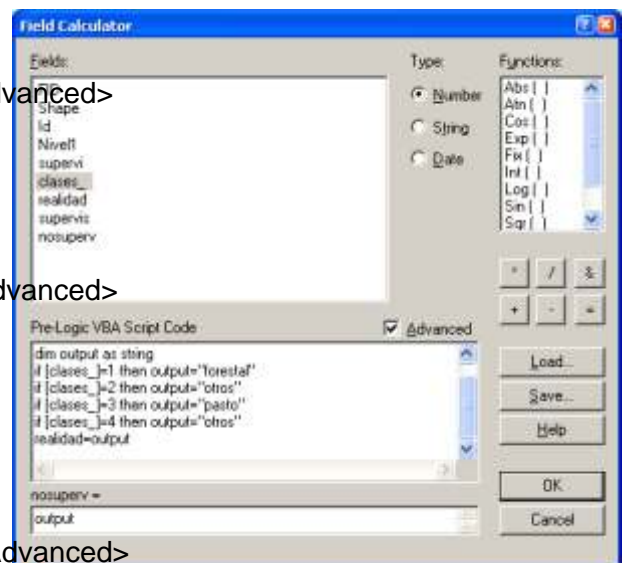
Pre-Logic VBA Script Code:

```
dim output as string
if [supervi]=1 then output="forestal"
if [supervi]=2 then output="otros"
if [supervi]=3 then output="pasto"
if [supervi]=4 then output="otros"
realidad=output
```

Campo nosuperv>Calculate values>Activar checkbox Advanced>

Pre-Logic VBA Script Code:

```
dim output as string
if [clases_]=1 then output="forestal"
if [clases_]=2 then output="otros"
if [clases_]=3 then output="pasto"
if [clases_]=4 then output="otros"
realidad=output
```





Capa de puntos de validación>Open attribute table
Options>Export...

Export = all records

Especificar una ruta para el .dbf de salida

Do you want to add the table to the current map?>No

Abrir el .dbf en Microsoft Excel

Datos>Informe de tablas y datos dinámicos

Paso 1 de 3

Activar checkbox para tipo de datos de lista o base de datos de Microsoft Office Excel y para tipo de informe de tabla dinámica

Paso 2 de 3>En el rango de datos seleccionar las 3 columnas:

nivel1 (atributo de inventario)

clasificación supervisada

clasificación no supervisada

Paso 3 de 3>Diseño

Arrastrar hasta las columnas el campo de realidad terreno (nivel1)

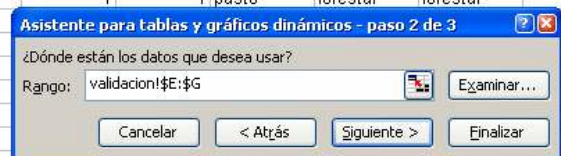
Arrastrar hasta las filas los datos de una clasificación

Arrastrar hasta los datos el mismo campo de clasificación. El conteo es la opción por defecto.

Situar el informe en una tabla nueva



	A	B	C	D	E	F	G	H
1	Id	Nivel1	supervi	clases	realidad	supervis	nosuperv	
2		0 5		1	1 pasto	forestal	forestal	
3		0 5		3	3 pasto	pasto	pasto	
4		0 5		2	2 pasto	otros	otros	
5		0 5		2	2 pasto	otros	otros	
				1	1 pasto	forestal	forestal	



Las matrices de confusión que aquí se muestra han sido generadas en este ejercicio entre los datos de inventario de campo y los datos del raster resultado de las clasificaciones no supervisada y supervisada. Gracias a este tipo de matrices se puede estimar la fiabilidad de la metodología. Obsérvese que en este caso en el campo *nivel1* (realidad) 1 es forestal y 5 pasto. En el campo *supervi* (clasificación) 1 es forestal y 3 pasto.

Un indicador de la fiabilidad global del mapa podría hacerse comparando el número de muestras con acierto respecto del total. También es posible medir por separado la fiabilidad del usuario (errores por comisión) y del productor (errores por omisión), para una determinada clase, siendo:

FIABILIDAD GLOBAL = muestras correctamente clasificadas / total de muestras
 FIABILIDAD DEL USUARIO = muestras correctamente clasificadas de una clase / total de muestras de esa clase encontradas en el mapa clasificado
 FIABILIDAD DEL PRODUCTOR = muestras correctamente clasificadas de una clase / total de muestras de esa clase inventariadas como verdad terreno



CLASIFICACIÓN NO SUPERVISADA

Cuenta de realidad	realidad		Total general
	forestal	pasto	
nosuperv			
forestal	469	117	586
pasto	100	1974	2074
otros	406	534	940
Total general	975	2625	3600

FIABILIDAD GLOBAL	27,86%
FIABILIDAD DEL USUARIO PARA CLASE FORESTAL	80,03%
FIABILIDAD DEL PRODUCTOR PARA CLASE FORESTAL	48,10%
ERRORES POR COMISIÓN CLASE FORESTAL	19,97%
ERRORES POR OMISIÓN CLASE FORESTAL	51,90%
FIABILIDAD DEL USUARIO PARA CLASE PASTO	95,18%
FIABILIDAD DEL PRODUCTOR PARA CLASE PASTO	75,20%
ERRORES POR COMISIÓN CLASE PASTO	4,82%
ERRORES POR OMISIÓN CLASE PASTO	24,80%

CLASIFICACIÓN SUPERVISADA

Cuenta de realidad	realidad		Total general
	forestal	pasto	
supervis			
forestal	540	129	669
pasto	82	1852	1934
otros	353	644	997
Total general	975	2625	3600

FIABILIDAD GLOBAL	66,44%
FIABILIDAD DEL USUARIO PARA CLASE FORESTAL	80,72%
FIABILIDAD DEL PRODUCTOR PARA CLASE FORESTAL	55,38%
ERRORES POR COMISIÓN CLASE FORESTAL	19,28%
ERRORES POR OMISIÓN CLASE FORESTAL	44,62%
FIABILIDAD DEL USUARIO PARA CLASE PASTO	95,76%
FIABILIDAD DEL PRODUCTOR PARA CLASE PASTO	70,55%
ERRORES POR COMISIÓN CLASE PASTO	4,24%
ERRORES POR OMISIÓN CLASE PASTO	29,45%

¿Cuál ha sido la fiabilidad del método?

¿Se consiguieron mejores resultados en el método supervisado o en el no supervisado? ¿Porqué?