

# DETECCIÓN DE CAMBIOS POR PARCELA PARA LA ACTUALIZACIÓN DE BBDD DE USOS DEL SUELO MEDIANTE IMÁGENES DE ALTA RESOLUCIÓN Y SENTINEL-2.

*DETECTION OF CHANGES BY PLOT FOR THE UPDATE OF LAND USE DATABASES USING HIGH RESOLUTION IMAGES AND SENTINEL-2*

**Sergio Andrade**

Universidad Politécnica de Valencia, España; sdas\_1189@yahoo.es

Recibido: 03 de octubre 2019

/ Aceptado: 16 de diciembre 2019

---

## RESUMEN

Uno de los elementos básicos para el análisis, gestión y planificación del territorio es contar con información actual y con ello de una buena cartografía de usos de suelo. Los procesos de actualización de bases de datos de usos del suelo suelen ser costosos por lo que las técnicas de observación remota de la Tierra sirven de ayuda. En el presente trabajo se pretende obtener mediante análisis multitemporal de imágenes SENTINEL-2 nivel 2A, correspondientes a los municipios de Torrente, Godolleta y Silla, la obtención de variables temporales a partir de NDVI, SAVI, EVI, bandas 2, 3, 4 y 8 de 14 fechas del año 2016. La generación de variables temporales es realizada con la finalidad de aportar variables dinámicas a las variables extraídas de imágenes de alta resolución como lo son las ortofotos PNOA del 2008 - 2015 para clasificar el uso de suelo de mejor manera y con ello poder determinar parcelas catastrales con cambio en el uso de suelo con el fin de identificar de forma inmediata los sectores de mayor cambio en caso de que se desee actualizar la cartografía de manera convencional a través de fotointerpretación, de modo que este producto facilite y reduzca los tiempos de ejecución y mantenimiento respecto a la actualización de la base cartográfica.

**Palabras clave:** Base de datos; imágenes alta resolución; Sentinel-2, análisis multitemporal; uso de suelo.

---

## ABSTRACT

One of the basic elements for the analysis, management and planning of the territory is to have current information and with it a good cartography of land uses. The updating processes of land use databases are often expensive, so the remote observation techniques of the Earth can help. In the present work, it is intended to obtain, through multitemporal analysis of SENTINEL-2 level 2A images, corresponding to the municipalities of Torrente, Godolleta and Silla, obtaining temporal variables from NDVI, SAVI, EVI, bands 2, 3, 4 and 8 of 14 dates of the year 2016. The generation of temporary variables is carried out in order to provide dynamic variables to the variables extracted from high resolution images such as the PNOA orthophotos 2008 - 2015 to classify land use in a better way and with this, to be able to determine cadastral parcels with a change in the use of land in order to immediately identify the sectors of greatest change in case you want to update the cartography in a conventional way through photointerpretation, so that this product facilitates and reduce the execution and maintenance times with respect to updating the cartographic database.

**Keywords:** Database; high resolution images, Sentinel-2, multitemporal analysis, land use.

## 1. INTRODUCCIÓN

La necesidad de tener información cartográfica digital actualizada ha aumentado y la disponibilidad de dispositivos de gestionarla es cada vez mayor, además del aporte de información de nuevos sensores que proporcionan información más detallada o con alta temporalidad en busca de satisfacer las necesidades de los usuarios. Con la disponibilidad de estos recursos es necesario poseer una cartografía de usos de suelo actualizada e integrada en un Sistema de Información Geográfica (SIG), puesto que es uno de los elementos básicos en la toma de decisiones políticas, analizando el territorio para su gestión y planificación.

La extracción de información proveniente de imágenes es un proceso costoso y fundamentalmente manual, aunque los esfuerzos por conseguir una automatización parcial o total se ha incrementado con el fin de minimizar costes económicos y temporales en la producción de cartografía e incrementar la productividad y estar a la par de la demanda exigida (Recio J. 2009). En consecuencia, no sólo es necesario poseer una buena cartografía de usos de suelo respecto a su calidad, sino también a su grado de actualización.

Existen algunos proyectos para cartografiar el uso o la cobertura del territorio, de ámbito global (MODIS Land Cover), regional (CORINE) y local (SIOSE) con sus respectivas escalas y metodologías, pero es necesario acotar que como factor común para la extracción de información es necesario el uso de fotointerpretación digital. Si bien se utiliza algún insumo de clasificación automática proveniente de imágenes, es tomado como información secundaria como apoyo en el proceso de fotointerpretación (Recio J. 2009).

Con la aparición de sensores de alta resolución espacial como Ikonos, Quickbird, GeoEye y WorldView que permiten poseer resoluciones antes solo conseguidas con ortofotografías han hecho posible la extracción de información para cartografiar a escalas mayores y la aplicación de novedosas variables debido a la alta resolución espacial de estos nuevos sensores (Recio J. 2009)..

Lo que el presente trabajo propone realizar es la detección e identificación de los cambios ocurridos entre distintas fechas con la ayuda de imágenes de alta resolución y el apoyo de información multitemporal proporcionada por el sensor SENTINEL-2 para el municipio de Torrente, Godelleta y Silla mediante la clasificación de imágenes por parcelas.

## 2. ÁREA DE ESTUDIO

Torrente (en valenciano y oficialmente Torrent) es un municipio de la Comunidad Valenciana, perteneciente a la provincia de Valencia, situado en el área metropolitana de Valencia en la comarca de la Huerta Oeste. Con 80 630 habitantes (Instituto Nacional de Estadística, 2017), se trata del municipio con más población de la provincia de Valencia (exceptuando a Valencia, la capital de la provincia) (Figura 1).

Godelleta, es un municipio de la Comunidad Valenciana, perteneciente a la provincia de Valencia y situado en la comarca de la Hoya de Buñol. Posee una población de 3 391 habitantes (Instituto Nacional de Estadística, 2017). Situado en el sector sureste de la comarca, la superficie del municipio comprende monte y llano, aun cuando su orografía no es intrincada.

Silla se encuentra a 10 km de Valencia, al sur de la comarca de l'Horta Sud. La extensión del término municipal es de 24 km<sup>2</sup> y limita, al norte, con los municipios de Albal y Beniparrell; al sur, con los de Almussafes y Sollana; al oeste, con Alcásser y Picassent, y al este, con el lago de la Albufera (Ayuntamiento de Silla, 2018).

Silla tiene una población de 16 169 habitantes (Instituto Nacional de Estadística, 2017). Se caracteriza por el predominio de la población joven y por una tasa de actividad y ocupación superior a la media de la Comunidad Valenciana.

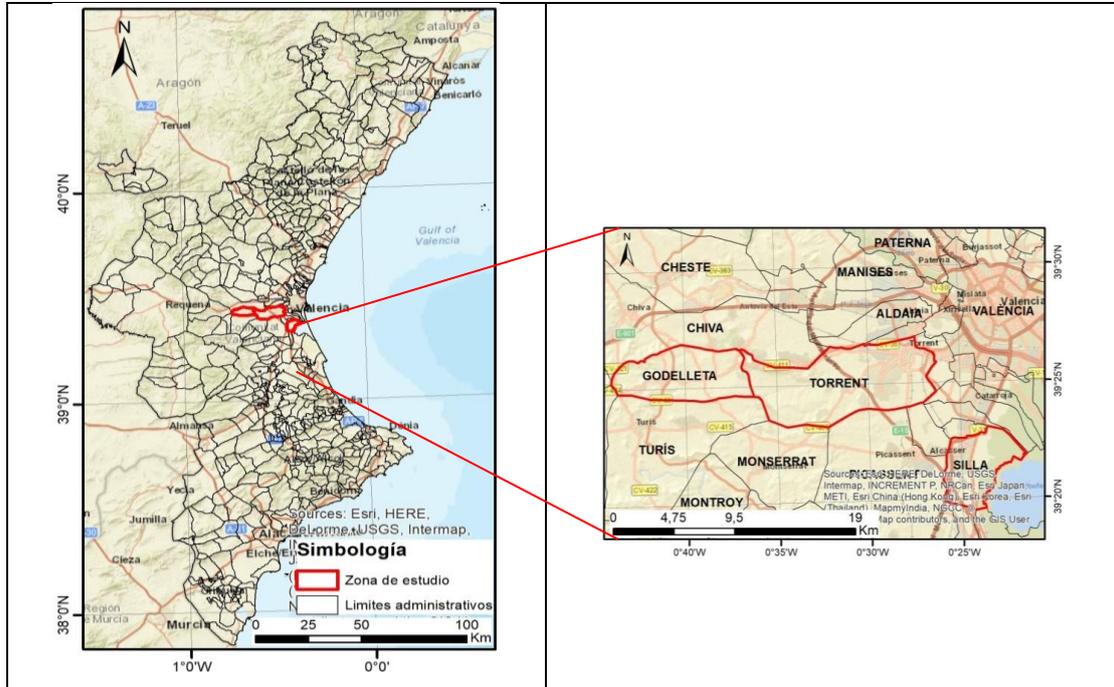


Figura 1 Área de estudio

### 3. METODOLOGÍA

#### 3.1. DEFINICIÓN DE LA LEYENDA.

Se realizó la verificación de información secundaria como el mapa de cultivos y aprovechamiento de España 1: 50 000, así como también información correspondiente al SIOSE de la ocupación de suelo, tomando como referencia el nivel 4 que jerárquicamente es el que mayor detalle aporta. En base al análisis de la información secundaria, se decidió crear una leyenda simple que abarque 12 categorías: canales de agua, cítricos, coníferas, frutales no cítricos, matorral, olivar, pastizal, polígono industrial, red viaria, suelo desnudo, urbano y viñedos. Se debe indicar que para otras pruebas explicadas a continuación se agregó la categoría de arroz y se excluyó a canales de agua.

#### 3.2. SELECCIÓN DE LAS MUESTRAS.

Se toma como base a las parcelas descargadas de la Dirección General de Catastro en rústica y urbana para ser unidas en un solo archivo del cual el municipio de Torrente posee 22 210 parcelas, a las que hay que sumar 251 parcelas provenientes de Silla y Godella para el presente estudio. Con la ayuda de ortofotos del 2012 más información secundaria como SIOSE 2010 y el mapa de cultivos y aprovechamientos de España se seleccionaron 2 130 parcelas, siendo cerca del 10,5 %, del universo en estudio. El tipo de muestreo fue aleatorio estratificado ya que se tomó en cuenta la representatividad de cada clase en base a la cantidad de parcelas pertenecientes a categorías similares por parte de información secundaria.

Además, se debe indicar que, para las clases urbano, cítricos y suelos desnudos cada una

de ellas contiene un grado de subdivisión explicada a continuación: urbana contiene a las clases casco urbano histórico, ensanche y urbano discontinuo; cítricos, contiene a cítricos joven como a cítricos maduros; suelo desnudo, contiene a suelo oscuro, claro y con algo de vegetación. Se aclara que esta subdivisión fue realizada para mejorar la clasificación puesto que existen diferencias espectrales a pesar de ser una misma clase o existía confusión entre ellas y fueron unidas en una categoría más general de modo que una vez extraídas la muestras y generado el modelo se unen para formar su respectiva clase.

Las muestras de imágenes de alta resolución y Sentinel-2, se volvieron a seleccionar 819 muestras sólo para el año 2015 seleccionándose parcelas de Godelleta y Silla de modo que incluyan las categorías de viñedos y arroz respectivamente y que cada categoría sea lo más significativa posible. Se verificó que las muestras fotointerpretadas correspondan a la misma categoría tanto en 2015 como en 2016 para Sentinel-2.

### 3.3. IMÁGENES DE ALTA RESOLUCION ESPACIAL

#### 3.3.1. *Bloque fotogramétrico y generación de ortomosaicos.*

El Instituto Geográfico Nacional y Centro Nacional de Información Geográfica (IGN-CNIG) proporcionó 354 fotogramas correspondientes a los municipios de Torrente, Silla y Godelleta del año 2015 puesto que no es posible descargar ortofotomosaicos con información espectral, que contenga al infrarrojo, en la web en el caso del año 2015. Para el año 2008 se cuenta con ortofotografías con información espectral adecuada. Fue necesario realizar el proceso de orto rectificación para 67 fotogramas que cubren el área de estudio.

Se empleó el software ERDAS con la extensión LPS para el ajuste del bloque aerofotogramétrico. De la página web fototeca digital de IGN (<https://fototeca.cnig.es/>) se descargó el certificado de calibración de la cámara (UltraCam XpWa, S/N UC-SXp-wa-30416083) así como también un archivo .shp con los centros procesados de cada fotograma, es decir sus coordenadas UTM (Este, Norte y altura) con sus respectivos ángulos de rotación (omega, phi, Kappa) para el ajuste y generación de ortofotos.

En el proceso de generación del bloque fotogramétrico, los centros procesados ingresados no permitieron realizar una orientación externa adecuada por lo que fue necesario tomar puntos de control tomando como referencia la ortofotografía 2008 para realizar la orientación externa. Fueron necesarios 183 puntos de control para poder realizar la orientación externa de todo el conjunto de fotogramas, además se descargó y utilizó el MDT (modelo digital de terreno) de cada uno de los municipios de la página del ICV en la generación de las ortofotografías y generar con ellos un ortofotomosaico de cada municipio.

Se debe indicar que el ajuste obtenido fue realizado por línea de vuelo, consiguiendo errores medios cuadráticos por debajo de 2 píxeles en cada línea de vuelo ajustada.

#### 3.3.2. *Extracción de variables espectrales, de textura y forma.*

FETEX 2.0 es un software para la extracción automática de características descriptivas de las propiedades de textura, espectrales, de forma y estructurales de fragmentos de imágenes u objetos-imagen (Ruiz et al., 2011)

Está orientado a la caracterización de objetos cartográficos (parcelas, sub-parcelas, polígonos, etc.), definidos a partir de ficheros vectoriales (shapefiles), así como a imágenes de satélite de alta resolución (QuickBird, Ikonos, WorldView, etc.) y fotografías aéreas C, IR. Así, por ejemplo, con una imagen multispectral y el parcelario definido por la cartografía

catastral como datos de entrada (input), genera una tabla ASCII (output) con los valores de las características o atributos calculados de forma automática a partir de las imágenes para cada una de las parcelas catastrales. Este resultado puede introducirse posteriormente en un software de clasificación.

Gran parte de las características obtenidas son novedosas y se basan en métodos de cálculo avanzados, especialmente las características estructurales, que permiten describir numéricamente la distribución espacial de los elementos presentes en un objeto/parcela, como por ejemplo: Espectral: Características estadísticas espectrales que informan sobre la respuesta radiométrica de una parcela; textura: características de textura que describen la distribución espacial de los niveles de gris correspondientes a los píxeles de una determinada región. Basadas en la intensidad de bordes, la forma del histograma y la matriz de concurrencias de los niveles de gris; Wavelet: Características de textura calculadas a partir de los detalles de la descomposición en wavelets de la imagen; Semivariograma: Características estructurales calculadas a partir de la semivariograma, que describe la regularidad de los elementos internos de una parcela; Hough: Características estructurales calculadas a partir de la transformada de Hough y de las posiciones de los árboles de una parcela para la caracterización de marcos de plantación u otras estructuras con patrones geométricos; Forma: Características de forma del objeto. Describen las propiedades geométricas de los objetos, complejidad o elongación de su contorno, etc.; Database features: características auxiliares directamente tomadas de la información de la base de datos analizada (Grupo de Cartografía GeoAmbiental y Teledetección, 2018).

Se ingresa el parcelario en formato shp a FETEX 2.0 al igual que los ortofotomosaicos respectivos para cada fecha, se seleccionan en el software las muestras, así como un identificador para cada parcela.

### 3.3.3. Clasificación mediante árboles de decisión.

Weka es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Los algoritmos se pueden aplicar directamente a un conjunto de datos o llamar desde su propio código Java. Weka contiene herramientas para el pre-procesamiento de datos, clasificación, regresión, clustering, reglas de asociación y visualización. También es adecuado para desarrollar nuevos esquemas de aprendizaje automático. Weka es un software de código abierto emitido bajo la Licencia Pública General de GNU (The University of Waikato, 2015).

Fetex 2.0 utiliza como formato archivos de salida de fichero .arff que es el indicado para Weka. Se realizó la creación de árboles de clasificación de las características anteriormente extraídas con la opción SIMPLE CLI que permite ingresar código de java.

En el siguiente paso se detalla la creación de árboles de decisión y evaluación con las muestras de aprendizaje y evaluación con las muestras de aprendizaje además de validarlas mediante validación cruzada a través de los siguientes comandos para cada uno de los años:

```
java weka.classifiers.meta.AdaBoostM1 -t C:\ORTO_2008\shp\orto_08.arff -W weka.classifiers.trees.J48 >
C:\ORTO_2008\shp\resulta.txt
java weka.classifiers.meta.AdaBoostM1 -t C:\ORTO_2015\PROCESOS\orto2015.arff -W weka.classifiers.trees.J48 >
C:\ORTO_2015\PROCESOS\resulta.txt
```

Una vez conseguido el árbol de clasificación se procede a clasificar el resto de parcelas a través del comando:

```
java weka.classifiers.meta.AdaBoostM1 -t C:\Trabajo\shp\screen\sin_identif.arff -T
C:\Trabajo\shp\screen\sin_identif.arff -p 0 -W weka.classifiers.trees.J48 > C:\Trabajo\shp\screen\resulta2.txt
```

Como resultado se obtiene un archivo de texto que posee el mismo orden que el archivo original, con lo que puede ser vinculado a través del identificador de la parcela y de este modo tener la clasificación predicha para todas las parcelas, en la Figura 2 se muestra una fracción del archivo que indica su identificador, su estado actual, el predicho, así como el porcentaje de certeza con la que se asignó a una determinada clase.

```
=== Predictions on test data ===

inst#      actual  predicted error prediction
-----
1          1:?  2:Citricos_joven      0.904
2          1:? 13:Suelo_claro        1
3          1:?  2:Citricos_joven      0.715
4          1:?  6:Frutales_no_citricos 1
5          1:? 13:Suelo_claro        0.999
6          1:?  6:Frutales_no_citricos 0.998
7          1:?  3:Citricos_maduro      1
```

Figura 2. Fragmento del archivo de clasificación predicha por el software Weka.

Cabe indicar, que el procedimiento fue el mismo para ambas fechas obteniendo cada una sus respectivos ficheros a ser vinculados con los parcelarios respectivos de cada año. Además se unificaron ciertas clases como cítrico joven y maduro por cítricos; centro histórico, ensanche y urbano discontinuo por urbano; así como también suelo vegetación, suelo claro y suelo oscuro por suelo desnudo.

Se realizó la matriz de confusión, seleccionándose un total de 2 130 parcelas de las cuales se obtuvo una fiabilidad global del 90 % y un coeficiente kappa del 0,85 para la clasificación de uso del suelo obtenida en 2008 y para 2015 se obtuvo una fiabilidad global de 80,32 % y un coeficiente Kappa de 0,69.

### 3.4. PROCESADO DE IMÁGENES SENTINEL-2.

Se realizó la descarga de 14 imágenes Sentinel-2A desde la página web de Copernicus, de la zona de estudio del año 2016 con nivel de procesamiento L1C de fechas: 12 de enero, 12 de marzo, 11 de abril, 21 de abril, 1 de mayo, 14 de junio, 10 de julio, 30 de julio, 10 de agosto, 19 de agosto, 18 de septiembre, 8 de octubre, 7 de noviembre y 27 de diciembre. Dichas imágenes fueron sometidas a corrección atmosféricas con Sen2cor, pasa de un nivel de procesamiento L1C a un 2A, realizando la corrección atmosférica, correcciones de cirrus en el techo de la atmosfera (L1C), pasándola a la parte baja de la atmósfera y corrección de reflectancia de cirrus en las imágenes (European Space Agency, 2017).

#### 3.4.1. Generación de variables temporales.

Una vez adquiridas las 14 imágenes SENTINEL L1C y procesadas a nivel L2A, se procedió a realizar el recorte de las mismas de acuerdo al área de estudio. De forma posterior se calcularon algunos índices temporales como el índice normalizado de vegetación (NDVI), índice de vegetación ajustado al suelo (SAVI) y el índice de vegetación mejorado (EVI) por medio de 2 de las 4 bandas de resolución espacial de 10 metros.

Tabla 1. Índices espectrales usados.

ÍNDICE DE VEGETACIÓN	FÓRMULA	REFERENCIA
Índice de vegetación normalizado (NDVI)	$NDVI = \frac{NIR - ROJO}{NIR + ROJO}$	(Huete et al., 2002).
Índice de vegetación mejorado (EVI)	$EVI = 2.5 \frac{NIR - ROJO}{NIR + 6 * ROJO - 7.5 * AZUL + 1}$	(Huete et al., 2002)
Índice de vegetación ajustado al suelo (SAVI)	$SAVI = \frac{1.5 * (NIR - ROJO)}{NIR + ROJO + 0.5}$	(Huete & Justice, 1994)

Con estos tres índices calculados para cada imagen en las 14 fechas se consiguieron un total de 42 imágenes correspondientes a NDVI, SAVI y EVI de las distintas fechas además de la información que aporta cada una de las bandas (2, 3, 4 y 8) de SENTINEL.

Con una de las herramientas del Sistema de Información Geográfica (SIG), se generó una malla de modo que cada cuadrícula coincida con cada píxel, al igual que se creó un centroide por cada cuadrícula con el fin de extraer la información de cada píxel de las distintas imágenes generadas y transferir esa información a su respectiva cuadrícula.

Una vez extraída la información se intersecó con las parcelas catastrales de modo que cada parcela contiene información de las distintas variables generadas. Además se desarrolló un script en Python que permite acceder a los distintos campos con la información almacenada para realizar la media ponderada respecto al área que ocupa cada píxel (cuadrícula) en la parcela, sacando como resultado un archivo en formato .csv.

Como ejemplo de la utilidad de las variables temporales extraídas, en la figura 6 se utiliza la variable NDVI con información de las 14 fechas en días julianos (12, 72, 102, 112, 122, 158, 192, 212, 222, 232, 262, 282, 312, 362) en el eje de las abscisas y en el eje de las ordenadas los valores de NDVI registrados en algunas parcelas de categorías con cobertura vegetal (arroz, cítricos, matorral, pastizal, viñedos, olivar y coníferas) de este modo se observa el comportamiento temporal de los valores de NDVI en las distintas categorías (Figura 3). De este modo los valores de NDVI como variable temporal permiten discriminar entre las distintas clases.

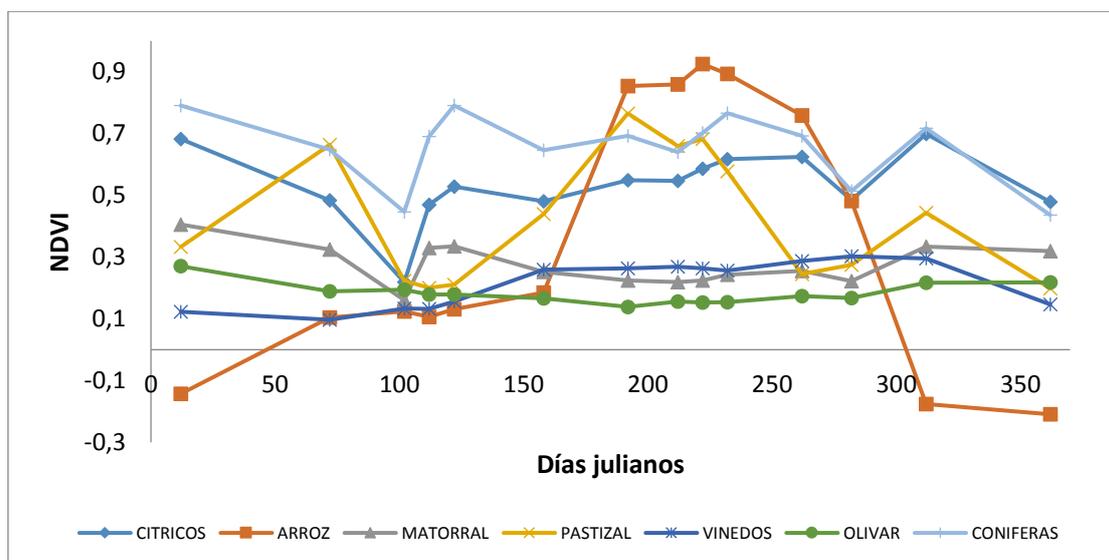


Figura 3. Comportamiento temporal del NDVI en categorías con cobertura vegetal.

Con el archivo resultante además de poseer información de las 14 fechas como el NDVI, SAVI, EVI, bandas 2, 3, 4 y 8, se pudieron generar nuevas variables como la pendiente entre fechas continuas, tomando como ejemplo el NDVI:

$$PEN = \frac{NDVI2 - NDVI1}{\# \text{ de días transcurridos entre fechas}} \tag{1}$$

Obteniendo con ello 13 datos más por variable en cuanto a la variable pendiente, es decir 91 datos. En la figura 16 se muestra como ejemplo el comportamiento temporal del NDVI para las clases arroz y coníferas en donde en el eje de las abscisas están las fechas en días julianos y en el eje de las ordenadas los valores de NDVI. Para la categoría arroz el cambio de pendiente entre los días 192 y 158 (enmarcados en el círculo rojo en la Figura 4) es significativo así como también entre los días 282 y 262; para la categoría coníferas un cambio de pendiente significativo ocurre entre los días 102 y 72, con lo que se demuestra que la variable pendiente aporta información discriminante entre categorías.

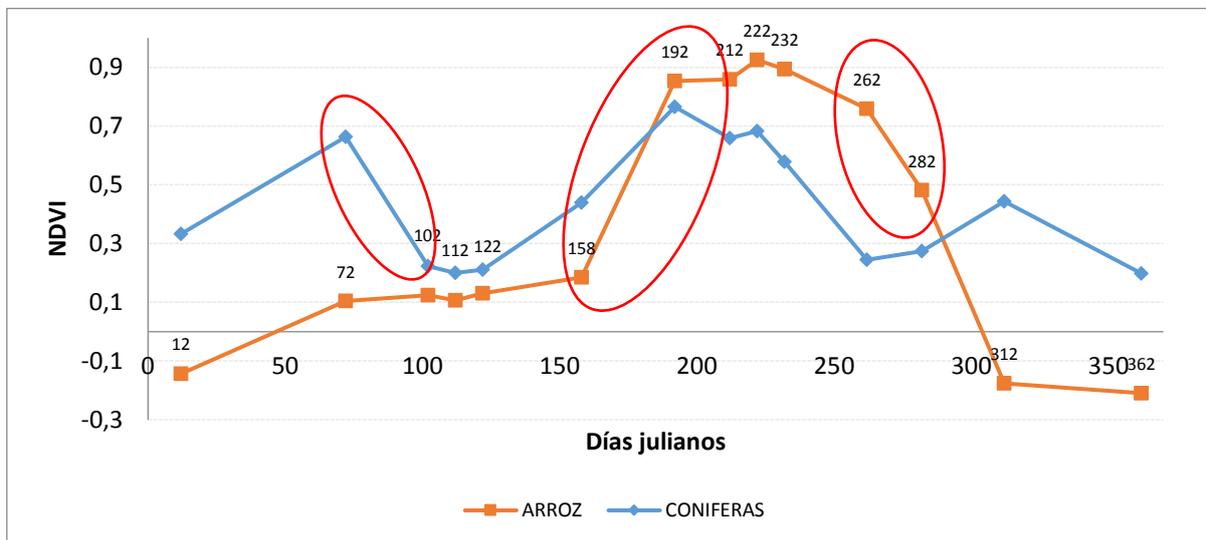


Figura 4. Ilustración del cambio de pendiente entre las categorías arroz y coníferas representado a través de la fluctuación temporal del NDVI.

Adicionalmente se generó para cada variable el área bajo la curva, tomando en cuenta que cada variable posee 14 datos representados en el eje de las ordenadas y en el eje de las abscisas el tiempo (días julianos), se calcula el área realizando la siguiente operación:

$$AREA = \frac{NDVI2 + NDVI1}{2} * \# \text{ Dias transcurrido entre fechas} \tag{2}$$

En la Figura 5 se muestra un ejemplo entre las categorías arroz y cítricos con sus respectivas áreas proyectadas, mostrándose en color rojo el área de la categoría arroz y en color azul el área de los cítricos, de manera que esta variable también aporta información que ayuda a discriminar entre las distintas categorías.

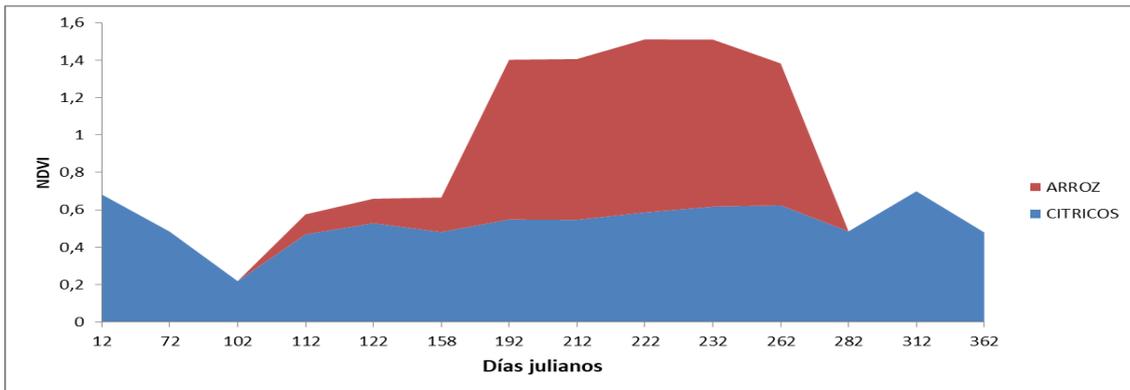


Figura 5. Área bajo la curva entre las categorías arroz y cítricos.

Se observó que el área bajo la curva entre el 1 de mayo y 21 de abril correspondientes a los días julianos 122 y 112 así como del 18 de septiembre y el 10 de julio correspondientes a los días julianos 262 y 192 proporcionan cierta discriminación entre coberturas vegetales por lo que también fueron agregadas como variables temporales, llamándolas AREA\_ND1 y AREA\_ND2 respectivamente. La Figura 6 muestra un ejemplo del comportamiento del NDVI en todas las categorías, así como también los segmentos de área a ser considerados como nuevas variables de área bajo la curva, indicados en líneas de color rojo.

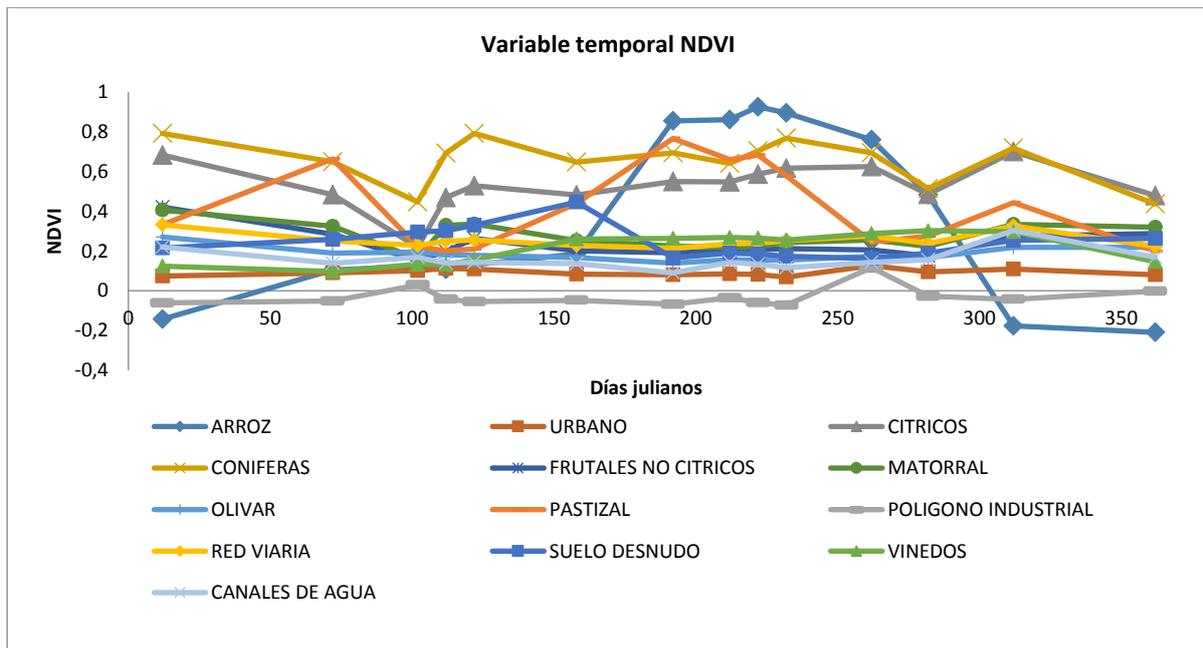


Figura 6. Comportamiento de la variable temporal NDVI en las 13 categorías, en color rojo se muestran los segmentos de área a ser considerados como nuevas variables temporales

Se generaron también datos estadísticos como la media y desviación estándar de las 14 fechas asignadas a cada variable (NDVI, SAVI, EVI, bandas 2, 3, 4 y 8) así como también de sus 13 datos correspondientes a pendientes de cada variable (NDVI, SAVI, EVI, bandas 2, 3, 4 y 8), es decir se consiguieron 28 datos más. Por ejemplo se consiguió la media del NDVI y de su pendiente, su desviación estándar del NDVI como de su pendiente.

En breves rasgos se consiguieron 98 variables de la extracción de información en el archivo csv, adicionalmente se sumaron 91 variables de pendiente, 9 variables de área y 28

variables estadísticas de cada una, acumulando un total de 226 variables.

#### 3.4.2. Selección de variables temporales.

Puesto que los datos parten de distintas fechas y estas a su vez están vinculadas a distintas variables, para simplificar la literatura se enumeraron a las variables en orden cronológico como por ejemplo el NDVI del 12 de enero se lo llamó NDVI\_1 hasta llegar al NDVI del 27 de diciembre que será NDVI\_14, obteniendo un total de 226 variables generadas.

Al realizar el análisis discriminante de los 98 datos del archivo csv, la banda 3 de fecha 19 de agosto era linealmente dependiente con otras variables por lo que fue descartada del proceso y se ingresaron las 97 variables con lo que se obtuvo un 92,43 % de datos correctamente clasificados. Es necesario verificar que no exista correlación para evitar información redundante. Se generó la matriz de correlaciones entre las variables con lo que se pudo concluir que el NDVI y SAVI al poseer un valor de uno como coeficiente de correlación puesto que SAVI resulta ser combinación lineal a NDVI, una de las 2 debe ser eliminada. Respecto a NDVI y EVI existe una ligera diferencia, pero su correlación es muy fuerte como por ejemplo entre NDVI\_1 y EVI\_1 al igual que NDVI\_13 y EVI\_13 poseen una correlación de 0,98 siendo las que menor correlación poseen, el resto de variables poseen una correlación de 0,99 por lo que en la práctica no aportan ninguna información adicional.

De la misma manera se analizaron las correlaciones entre bandas, observándose una alta correlación entre la banda 2 y 3, oscilando entre 0,96 a 0,99; la correlación entre la banda 2 y 4 disminuye respecto a la anterior, oscilando entre 0,97 y 0,76; no existe correlación entre la banda 2 y 8 ya que la mayoría de valores son cercanos a 0. Del mismo modo se analizó las correlaciones de la banda 3 con la 4, que oscila entre 0,98 y 0,83; las correlaciones entre la banda 3 y 8 fluctúan entre 0,91 y 0,18. Las correlaciones entre la banda 4 y 8 oscilan entre 0,89 y 0,22. En resumen existe una alta correlación entre la banda 2 y 3, ligeras correlaciones entre las bandas 2 - 4, banda 3 - 4 y correlaciones muy bajas con la banda 8.

Respecto a los índices se seleccionó al NDVI puesto que en su tabla de clasificación con las 14 fechas obtiene un 72,77 % de datos bien clasificados, SAVI queda descartado por su alta correlación con NDVI. EVI obtuvo en la tabla de clasificación un 67,11 % además de tener una alta correlación con NDVI por lo que se seleccionó a NDVI como el índice más idóneo. En referente a las bandas, se seleccionó a las bandas 2, 4 y 8. Debido a que la banda 2 y 3 tenían una alta correlación y al realizar el análisis discriminante la banda 2 fue la que mejor resultado obtuvo con un 70,33 % de datos correctamente clasificados, la banda 3 obtuvo un 67,08 % de datos bien clasificados además de tener uno de sus datos linealmente dependientes, es decir que uno de sus datos es combinación lineal del resto.

Se realizó el análisis discriminante ingresando las variables previamente seleccionadas NDVI, banda 2, 4 y 8 acumulando 56 variables y se obtuvo un 88,03 % siendo apenas 4,4 % menor con 41 variables menos.

Para disminuir el número de variables se realizó análisis de componentes principales (ACP), así como también análisis discriminante al conjunto de 56 variables consiguiendo reducir a 10 el número de variables siendo las siguientes: NDVI1, NDVI9, B8\_10, B2\_12, NDVI13, B4\_12, B4\_3, B4\_10, B8\_6 y B4\_1, con lo que se consiguió un 79,73 % en la tabla de clasificación.

De las 52 variables de pendientes de NDVI – recordando la notación PEN\_NDVI\_X -, bandas 2, 4 y 8 se un obtuvo un 86,57 % de datos bien clasificados en la tabla de clasificación, se procedió a realizar el mismo procedimiento para amenorar las variables,

consiguiendo 16 componentes principales, seleccionándose el mismo número de variables: PEN\_NDVI\_6, PEN\_NDVI\_2, PEN\_B4\_5, PEN\_B2\_2, PEN\_NDVI\_1, PEN\_B8\_11, PEN\_NDV\_13, PEN\_B8\_5, PEN\_NDVI\_5, PEN\_NDVI\_7, PEN\_B8\_10, PEN\_B2\_10, PEN\_B4\_7, PEN\_B2\_7, PEN\_NDV\_11 y PEN\_NDV\_12 con las que se consiguió un 76,80 % de datos correctamente clasificados.

Adicionalmente se analizaron las variables media, desviación estándar y área bajo la curva para las variables NDVI, banda 2, 4 y 8, además de la media y desviación estándar de las pendientes anteriormente generadas obteniendo un total de 22 datos. Pero al analizar la correlación entre la media y área bajo la curva la correlación es muy alta por lo que fue descartada área bajo la curva como variable -entendiéndose a la misma como la suma de toda el área, mas no las dos variables mostradas quedando 18 variables en total.

Con las 18 variables se obtuvo 82,05 % de datos bien clasificados, se realizó ACP, resultando cuatro componentes principales de las cuales se seleccionaron 9 variables: DES\_NDVI, AVE\_NDVI, AREA\_NDVI2, AVE\_B8, AVE\_B2, AVE\_B4, PEN\_B4\_AVE, DES\_B8 y PEN\_B4\_DES que consiguieron 78,14% de datos correctamente clasificados en la tabla de clasificación.

Como último paso se unieron todas las variables pre-seleccionadas teniendo un total de 35 variables que en la tabla de clasificación obtuvieron un 86,69% de datos bien clasificados. Se realizó el mismo procedimiento, resultando ocho componentes principales.

Se aclara que el orden de las 12 variables seleccionadas es de acuerdo a la información que aporta cada variable, siendo en este caso información del índice NDVI e información estadística del mismo como la desviación estándar y media además del NDVI9 – si se observa la figura 3, la fecha del 9 de agosto (día juliano 222) es la que mejor discrimina entre coberturas vegetales - las que mayor información aportan. A manera de ejemplo en la Figura 7 se muestra una de las variables estadísticas AVE\_NDVI que representa la media de los valores registrados de NDVI en las 14 fechas con lo cual se evidencia que la información que aporta la media ayuda en la discriminación entre categorías, por ejemplo la media que mayor valor posee de NDVI corresponde a coníferas, seguidos de cítricos, pastizal y arroz, por otra parte los valores de NDVI más bajos ocupan las categorías de urbano y polígono industrial como era de esperarse al no existir vegetación.

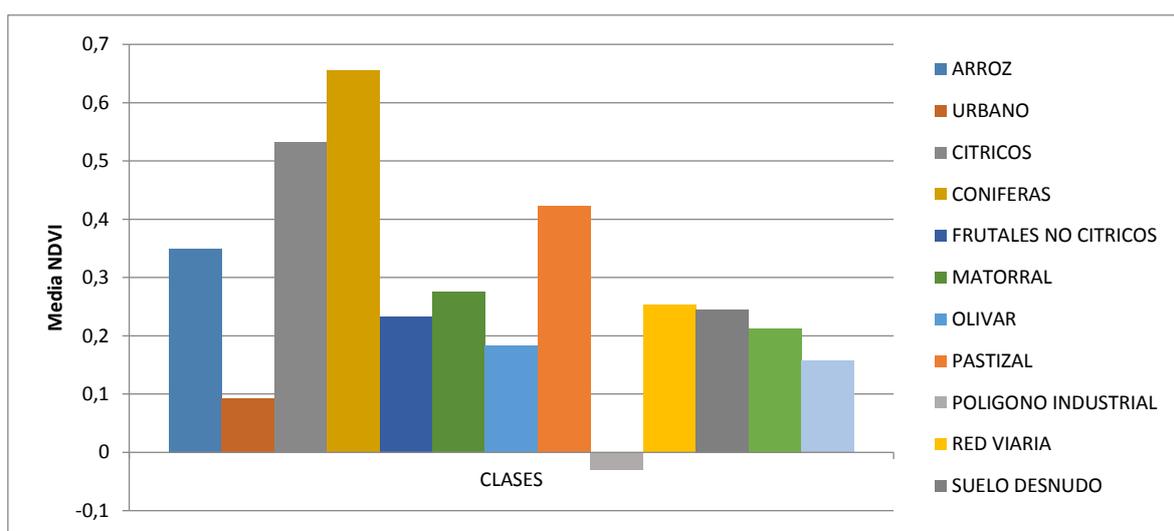


Figura 7. Variable AVE\_NDVI correspondiente a la media del NDVI de las 14 fechas del año 2016 en las distintas categorías.

Continuando con la descripción de las 12 variables seleccionadas se encuentra la información aportada por la banda 8 (infrarrojo cercano) del 19 de agosto, banda 2 (azul) del 8 de octubre, la media de las pendientes en la banda 4 (rojo), banda 4 del 8 de octubre, banda 4 del 19 de agosto, media de la banda 8, banda 4 del 12 de enero, desviación estándar de la banda 8 y la banda 2 del 18 de septiembre.

Puesto que existen algunas variables relacionadas con la banda 4 se ilustra como ejemplo en la figura 8 el comportamiento de la banda 4 en las 14 fechas registradas, observándose que la variable B4\_12 corresponde a la fecha del 8 de octubre (día juliano 282) que permite discriminar algunas de las clases con cobertura vegetal puesto a que registran valores bajos en esa banda debido a su alta absorción en esa porción del espectro electromagnético como las coníferas, arroz y cítricos; la variable B4\_10 correspondiente a la fecha del 19 de agosto (día juliano 232) continua discriminando las anteriores categorías descritas, se observa en la Figura 8 que algunas de las categorías se separan en esta fecha lo que permite discriminarlas de mejor manera como es el caso de los viñedos y olivares; y la variable B4\_1 que corresponde al 12 de enero (día juliano 12) en la que se observa que los viñedos registran valores altos a diferencia del resto de clases, el matorral y los cítricos también logran diferenciarse.

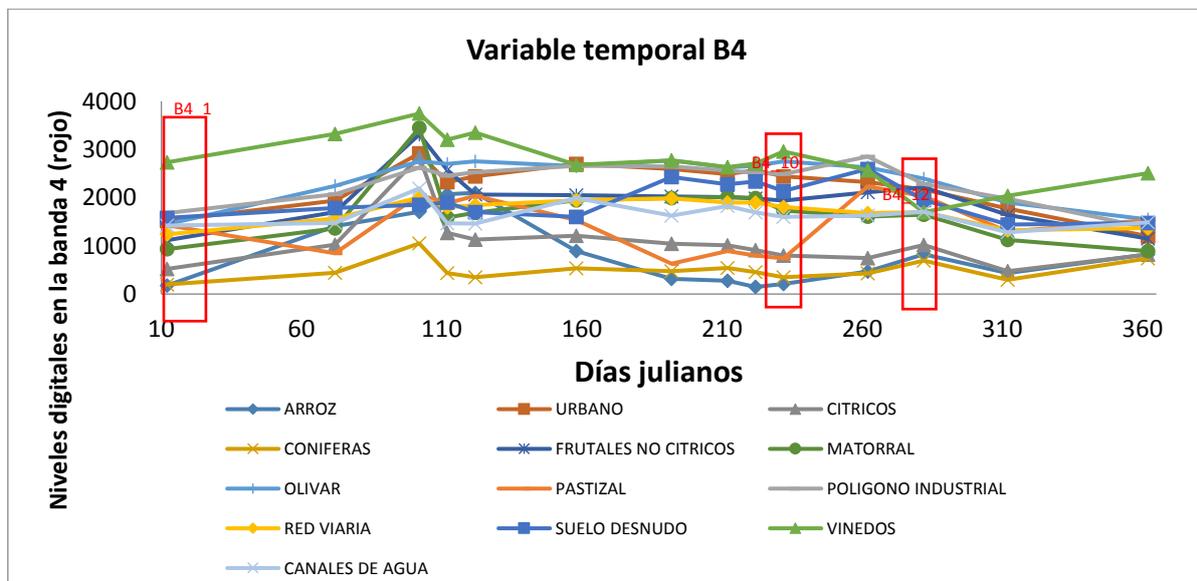


Figura 8. Comportamiento de la Banda 4 (rojo) en las 13 categorías durante el año 2016, en color rojo se muestran las variables B4\_1, B4\_10 y B4\_12.

## 4. RESULTADOS Y DISCUSIONES

### 4.1. DETECCIÓN DE CAMBIO 2008-2015.

La detección de cambios fue llevada a cabo mediante comparación de la clasificación entre el mapa de uso del suelo generado de 2008 y el de 2015. A través de la tabla del archivo en formato .shp del año 2008 se introdujo un campo adicional con la clasificación correspondiente al año 2015, adicionalmente se unieron ambos registros y se procedió a realizar un resumen de las combinaciones existentes y el número de repeticiones de las mismas. De modo que las que permanecían constantes en ambos años fueron ingresadas como no cambio y como cambio a las que sí, la Figura 9 muestra una porción de la tabla del archivo

.shp a manera de ejemplo de los campos correspondientes a la clasificación de uso del suelo del 2008 y 2015 así como la unión de ambos registros y la posterior clasificación entre cambio y no cambio respecto a las clases.

CLASES 08	CLASES15	UNION	CAMBIO
CITRICOS	CITRICOS	CITRICOS-CITRICOS	NO
CITRICOS	CITRICOS	CITRICOS-CITRICOS	NO
CITRICOS	CITRICOS	CITRICOS-CITRICOS	NO
CITRICOS	CITRICOS	CITRICOS-CITRICOS	NO
CITRICOS	CITRICOS	CITRICOS-CITRICOS	NO
CITRICOS	CITRICOS	CITRICOS-CITRICOS	NO
CITRICOS	CITRICOS	CITRICOS-CITRICOS	NO
CITRICOS	CITRICOS	CITRICOS-CITRICOS	NO
CITRICOS	Coníferas	CITRICOS-Coníferas	SI
CITRICOS	Coníferas	CITRICOS-Coníferas	SI
CITRICOS	Coníferas	CITRICOS-Coníferas	SI
CITRICOS	Coníferas	CITRICOS-Coníferas	SI
CITRICOS	Coníferas	CITRICOS-Coníferas	SI
CITRICOS	Coníferas	CITRICOS-Coníferas	SI
CITRICOS	Coníferas	CITRICOS-Coníferas	SI
CITRICOS	Coníferas	CITRICOS-Coníferas	SI
CITRICOS	Coníferas	CITRICOS-Coníferas	SI
CITRICOS	Coníferas	CITRICOS-Coníferas	SI
CITRICOS	Coníferas	CITRICOS-Coníferas	SI
CITRICOS	Coníferas	CITRICOS-Coníferas	SI

Figura 9. Porción de la tabla del archivo .shp de cambios entre el uso de suelo del 2008 y 2015.

Para evaluar la detección de cambios se decidió escoger una muestra del 5% para zonas detectadas como cambio y otro 5% para zonas detectadas como no cambio. Se escogieron 385 muestras para zonas de cambio y 726 muestras para las zonas de no cambio, dando un total de 1 111 muestras a ser evaluadas por medio de fotointerpretación. Para que en la selección de muestras exista el componente de aleatoriedad, se generaron dos nuevos campos en el archivo .shp que se viene trabajando, con la finalidad de generar un identificador para los cambios y no cambios. Realizado el proceso de fotointerpretación se obtuvo el siguiente resultado mostrado en la matriz de confusión en la tabla 2, obteniendo de ella una fiabilidad global del 84,78 % con respecto a la clasificación entre cambios y no cambios.

Tabla 2. Matriz de confusión de cambios y no cambios entre 2008 y 2015.

	CAMBIO	NO CAMBIO	TOTAL	FIABILIDAD USUARIO	ERROR COMISION
CAMBIO	212	155	367	57,77	42,23
NO CAMBIO	14	730	744	98,12	1,88
TOTAL	226	885	1111		
FIABILIDAD PRODUCTOR	93,81	82,49			
ERROR POR OMISION	6,19	17,51			

La tabla 3 muestra que han cambiado cerca de 7 680 parcelas y el acumulado de su área corresponde a 3 956,08 Ha respectivamente. Las parcelas que no han sufrido cambios son la mayoría a pesar de que en área representan minoría debido a que es su gran parte pertenecen a la categoría de urbano.

Tabla 3. Clasificación de cambio de uso de suelo por parcelas y área en Ha.

	PARCELAS	AREA Ha
CAMBIO	7680	3956,08
NO CAMBIO	14530	2985,88
TOTAL	22210	6941,97

En la matriz de confusión existe una gran cantidad de falso positivos, siendo un 42,23 %, que a nivel de usuario significa que se tendrá clasificado como cambio algo que en realidad no lo es, por otro lado lo clasificado como no cambio existe una gran certeza de predicción tanto como productor y usuario puesto que lo clasificado como no cambio no posee una alta cantidad de falsos negativos es decir, errores de omisión.

Se generó a través de los siguientes indicadores TP (sensibilidad) y FP (1-sensibilidad) extraídos de la matriz de confusión, un punto en el espacio ROC, como se indica en la Figura 10. Se puede observar que al estar cercana a los valores (1,1) y debido al gran número de falsos positivos, la detección de cambios provoca errores por comisión. Los cambios detectados serán necesarios de verificar por fotointerpretación, más los de no cambio no serían necesarios.

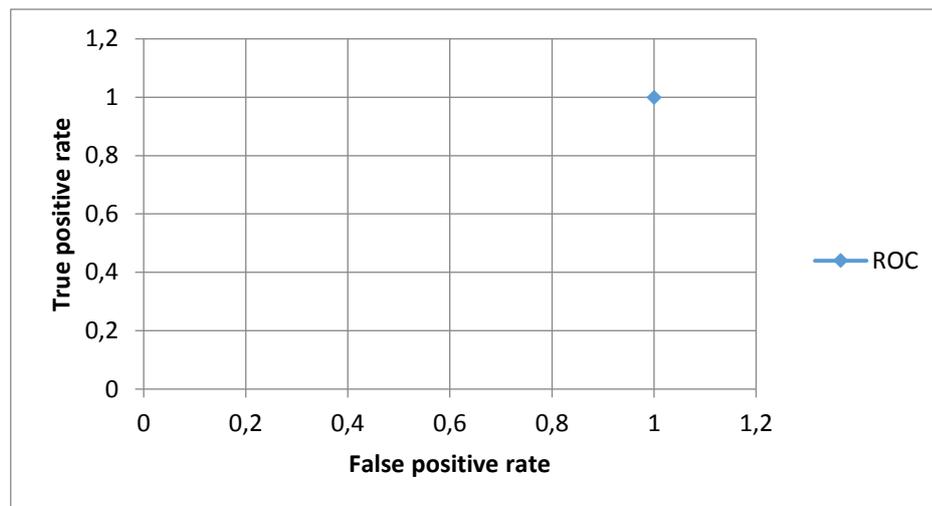


Figura 10. Espacio ROC para clasificador de cambios.

#### 4.2. VARIABLES TEMPORALES PARA SENTINEL-2

Se eligieron 12 variables temporales para el año 2016, seleccionadas por medio del procedimiento descrito en el apartado 3.4.2, el cual usa los métodos de análisis por componentes principales y análisis discriminante: DES\_NDVI, AVE\_NDVI, NDVI9, B8\_10, B2\_12, PEN\_B4\_AVE, B4\_12, B4\_10, AVE\_B8, B4\_1, DES\_B8 y B2\_11, se obtuvo un 83,64 % de datos correctamente clasificados de las 13 categorías en la tabla de clasificación.

#### 4.3. PRUEBAS DE CLASIFICACIÓN CON DISTINTOS INSUMOS

##### 4.3.1 Clasificación de parcelas con variables espectrales, de textura y forma para el año 2015

De la muestra de 819 parcelas fotointerpretadas, se extrajo información con el software FETEX, explicado en el apartado 3.3.3, del cual fue ingresado al software WEKA para realizar una clasificación con validación cruzada de las características extraídas por Fetex 2.0 (variables espectrales, de textura y forma).

Se utilizaron árboles de decisión entre ellos el clasificador AdaBoost con el algoritmo J48 dando como resultado una fiabilidad de 87,91 % y un coeficiente Kappa de 0,86 en la clasificación de las 12 categorías. La matriz de confusión proporcionada por el software WEKA se muestra en la Figura 11. Se puede observar que existe confusión en la clase de

pastizal, que asigna algunas parcelas a suelo desnudo, así como también lo hace la clase de arroz.

```

=== Confusion Matrix ===
      a  b  c  d  e  f  g  h  i  j  k  l  <-- classified as
25  1  0  1  1  0  0  1  4  0  4  0 | a = RED_VIARIA
 1 83  0  3  1  1  0  1  1  2  2  0 | b = MATORRAL
 0  0 110  0  0  1  0  0  0  0  0  1 | c = CITRICOS
 0  1  0  90  0  4  1  4  3  1  0  1 | d = SUELO_DESNUDO
 0  2  0  0  55  0  0  0  0  2  0  0 | e = CONIFERAS
 0  2  2  11  1  9  0  0  0  1  0  0 | f = PASTIZAL
 0  1  0  5  0  0  0  1  0  0  0  0 | g = OLIVAR
 1  0  0  3  0  0  0  19  0  0  0  0 | h = FRUTALES_NO_CITRICOS
 2  0  0  3  0  0  0  1  93  1  0  0 | i = POLIGONO_INDUSTRIAL
 1  0  0  0  4  0  0  0  2  98  0  0 | j = URBANO
 2  2  0  4  0  1  0  0  0  0  91  0 | k = ARROZ
 0  0  0  2  1  0  0  0  0  0  0  47 | l = VINEDOS
    
```

Figura 11. Matriz de confusión con variables espectrales, de textura y forma.

#### 4.3.2 Clasificación de parcelas con variables temporales (SENTINEL-2)

Las 12 variables seleccionadas también fueron ingresadas al software WEKA, utilizando el mismo procedimiento se obtuvo una fiabilidad global de 87,54 y un índice kappa de 0,86 en la clasificación de las 12 categorías. En la Figura 12 se muestra la matriz de confusión proporcionada por el software en donde se observa que la categoría de arroz es perfectamente clasificada, confusiones como matorral y coníferas es entendible que existan puesto que espectralmente, así como temporalmente se comportan un tanto similares.

```

=== Confusion Matrix ===
      a  b  c  d  e  f  g  h  i  j  k  l  <-- classified as
23  3  0  1  0  0  0  0  5  5  0  0 | a = RED_VIARIA
 1 83  2  2  7  0  0  0  0  0  0  0 | b = MATORRAL
 0  1 104  4  0  2  0  0  0  1  0  0 | c = CITRICOS
 0  3  2  91  0  1  0  4  0  2  0  2 | d = SUELO_DESNUDO
 0  5  0  0  54  0  0  0  0  0  0  0 | e = CONIFERAS
 0  1  2  2  0  20  0  0  0  0  1  0 | f = PASTIZAL
 0  0  1  2  0  0  3  0  1  0  0  0 | g = OLIVAR
 0  2  0  4  0  0  1  12  0  4  0  0 | h = FRUTALES_NO_CITRICOS
 3  0  0  1  0  0  0  0  95  1  0  0 | i = POLIGONO_INDUSTRIAL
 4  2  0  2  1  0  2  2  1  91  0  0 | j = URBANO
 0  0  0  0  0  0  0  0  0  0 100  0 | k = ARROZ
 0  0  0  2  0  0  0  4  0  0  0  44 | l = VINEDOS
    
```

Figura 12. Matriz de confusión con variables temporales

#### 4.3.3 Clasificación de parcelas con variables espectrales, de textura, forma y temporales (SENTINEL-2).

Se integraron en un solo fichero las variables espectrales, de textura y forma aportadas por FETEX 2.0 así como también las 12 variables temporales seleccionadas para Sentinel-2. Tras realizar el mismo procedimiento en el software WEKA se obtuvo como resultado una fiabilidad global de 91,94 % así como también un índice kappa de 0,90 en la clasificación de

las 12 categorías. En la Figura 13 se muestra la matriz de confusión proporcionada por el software en donde se evidencia que errores en la categoría arroz cometidos tan solo al utilizar las variables dadas de FETEX son corregidos por las variables temporales de Sentinel-2 llegando a clasificar perfectamente esa categoría. De igual forma errores al clasificar la red viaria por parte de variables temporales son corregidas por las variables espectrales, de textura y forma.

=== Confusion Matrix ===

	a	b	c	d	e	f	g	h	i	j	k	l	<-- classified as
29	1	0	2	1	0	0	0	1	3	0	0	0	a = RED_VIARIA
0	88	1	0	3	0	0	0	2	0	1	0	0	b = MATORRAL
0	0	111	1	0	0	0	0	0	0	0	0	0	c = CITRICOS
2	0	1	93	0	2	3	3	0	1	0	0	0	d = SUELO_DESNUDO
1	5	0	0	52	0	0	0	0	0	1	0	0	e = CONIFERAS
0	2	1	3	0	19	0	0	0	0	0	0	1	f = PASTIZAL
0	0	1	2	0	0	1	3	0	0	0	0	0	g = OLIVAR
0	1	0	3	0	0	3	16	0	0	0	0	0	h = FRUTALES_NO_CITRICOS
2	0	0	1	0	0	0	0	95	2	0	0	0	i = POLIGONO_INDUSTRIAL
0	2	0	0	1	0	0	0	0	102	0	0	0	j = URBANO
0	0	0	0	0	0	0	0	0	0	100	0	0	k = ARROZ
0	0	0	2	0	0	0	0	0	0	1	0	47	l = VINEDOS

Figura 13. Matriz de confusión con variables espectrales, de textura, forma y temporales.

## 5. CONCLUSIONES

Se generó el mapa de uso del suelo para el municipio de Torrente del año 2008 con una fiabilidad global del 90 % y un coeficiente kappa del 0,85 mediante la extracción de variables espectrales, de textura y forma, en ortofotos del mismo año, generadas con software FETEX 2.0. Como muestra se seleccionaron 2 130 parcelas fotointerpretadas en 12 categorías: canales de agua, cítricos, coníferas, frutales no cítricos, matorral, olivar, pastizal, polígono industrial, red viaria, suelo desnudo, urbano y viñedos, se empleó como clasificador árboles de decisión y validación cruzada en el software WEKA.

Se generó el mapa de cambios entre los años 2008 – 2015 con una fiabilidad global de 84,74 % con respecto a la clasificación entre cambio y no cambio a partir de una muestra del 5 % entre cambios y 5 % de no cambios detectados en las parcelas clasificadas en ambos años, seleccionándose 1 111 parcelas en el municipio de Torrente. En cuanto a detección de no cambios las fiabilidades tanto de productor (82,49 %) como de usuario (98,12 %) son elevadas, es decir que existirán escasos errores por omisión y comisión respectivamente; y en la detección de cambios se posee fiabilidades de productor altas (93,81 %) pero la fiabilidad de usuario disminuye (57,77 %), es decir que existen escasos errores por omisión y se encontraran una mayor cantidad de errores por comisión que deberán ser revisados por el foto intérprete.

Para incrementar la fiabilidad de la clasificación de usos de suelo del 2015 se incluyó información temporal de imágenes Sentinel-2, pero al disponer imágenes Sentinel del 2016 se han tomado 819 muestras de parcelas que no han cambiado entre ambos años. Se generaron 226 variables temporales a partir de 14 imágenes Sentinel-2 del 2016, seleccionándose 12 variables tras realizar análisis de componentes principales y análisis discriminante: DES\_NDVI (desviación estándar del NDVI), AVE\_NDVI (media del NDVI), NDVI9 (NDVI del 9 de agosto del 2016), B8\_10 (números digitales registrados en la banda 4 del 19 de

agosto del 2016), B2\_12 (números digitales registrados en la banda 2 del 8 de octubre del 2016), PEN\_B4\_AVE (media de la pendiente de la banda 4), B4\_12 (números digitales registrados en la banda 4 del 8 de octubre del 2016), B4\_10 (números digitales registrados en la banda 4 del 19 de agosto del 2016), AVE\_B8 (media de la banda 8), B4\_1 (números digitales registrados en la banda 4 del 12 de enero del 2016), DES\_B8 (desviación estándar de la banda 8) y B2\_11 (números digitales registrados en la banda 2 del 18 de septiembre del 2016). Se obtuvo una fiabilidad global de 87,54 % al utilizar arboles de decisión como clasificador con validación cruzada.

Al unir variables espectrales, de textura y forma extraídas de imágenes de alta resolución (ortofotos 2015) proporcionadas por el software Fetex con las 12 variables temporales del sensor Sentinel-2 se obtuvo un 91,94 % de fiabilidad global así como también un coeficiente kappa de 0,90 en la clasificación de usos de suelo para las 12 categorías. El porcentaje de fiabilidad global utilizando árboles de decisión como clasificador aumentó un 4,03 % respecto a utilizar variables espectrales, de textura, forma y 4,4 % solo utilizando las 12 variables temporales.

### AGRADECIMIENTOS

Este estudio es parte de una investigación propuesta por el Dr. Luis Ángel Ruiz Fernández, del departamento de Ingeniería Cartográfica, Geodesia y Fotogrametría con la ayuda del Dr. Ángel Antonio Balaguer Beser del departamento de Matemática Aplicada, por parte de la Universitat Politècnica de Valencia.

### REFERENCIAS

- European Space Agency. (2017, marzo 7). *ESA*. Retrieved abril 19, 2018, from ESA: [https://www.esa.int/Our\\_Activities/Observing\\_the\\_Earth/Copernicus/Sentinel-2/Introducing\\_Sentinel-2](https://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2/Introducing_Sentinel-2)
- Grupo de Cartografía GeoAmbiental y Teledetección. (2018, Marzo 15). *FETEX 2.0: An Object-based Feature Extraction Software Tool*. Retrieved from Grupo de Cartografía GeoAmbiental y Teledetección: <http://cgat.webs.upv.es/BigFiles/fetex2web/demofetex/SoftwareFETEX.pdf>
- Huete, A., & Justice, C. y. (1994). Development of vegetation and soil indices for MODIS-EOS. In *Remote Sensing of Environment* (pp. 224-234).
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., & Gao, X. y. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. In 8. *Remote Sensing of Environment*.
- Instituto Nacional de Estadística. (2017). *Población, superficie y densidad por municipios*. Retrieved abril 19, 2018, from <http://www.ine.es/jaxiT3/Datos.htm?t=2851>
- Recio, J. (2009). *Técnicas de extracción de características y clasificación de imágenes orientada a objetos aplicadas a la actualización de bases de datos de ocupación del suelo*. Valencia: dpto. Ingeniería cartográfica, geodesia y fotogrametría, Universidad Politécnica de Valencia
- Ruiz, et. a. (2011). A feature extraction software tool for agricultural object-based image analysis. *Computers and Electronics in Agriculture*, 284-296.
- The University of Waikato. (2015). *WEKA*. Retrieved abril 20, 2018, from <https://www.cs.waikato.ac.nz/ml/weka/>