

ANÁLISIS METODOLÓGICOS Y DE RESULTADOS DE DIFERENTES ALTERNATIVAS DE CLASIFICACIÓN DE IMÁGENES DE SATÉLITE PARA LA OBTENCIÓN DE ESTADÍSTICAS AGRARIAS

A. LOBATO Y J. M. MOREIRA.

Servicio Evaluación de Recursos Naturales. Agencia de Medio Ambiente. Junta de Andalucía. Sevilla

RESUMEN

La utilización de la Teledetección en combinación con trabajos de campo, apoyados en muestreos del terreno y métodos de análisis estadístico, constituye actualmente uno de los temas más adecuados y extendidos para la evaluación de superficies de cultivos y estadísticas agrarias. En esta comunicación se describe la metodología desarrollada en este campo por el equipo de la AMA-Junta de Andalucía, con especial atención a las distintas alternativas y variables estudiadas dentro del proceso de clasificación, sus resultados y conclusiones.

ABSTRACT

Remote sensing application joined with field surveys, based on ground sampling and methods on statistical analysis are at present one of the most suitable and extended systems for crops surface evaluation and agricultural surveys. This work describes the methodology developed to this purpose by the AMA-Junta de Andalucía team, paying special attention to different choices and variables taken into account in the classification process, their results and conclusions.

1. INTRODUCCIÓN

Los datos procedentes de sensores remotos, caracterizados por aportar información global de una zona extensa, y por la periodicidad de su obtención, constituyen una herramienta apropiada para complementar los trabajos de reconocimiento de cultivos en campo con el fin de obtener, de forma más precisa, evaluaciones de superficie de ocupación del territorio.

La teledetección se ha convertido por ello, en la técnica fundamental en la que se apoyan las modernas metodologías para la realización de estadísticas agrarias.

La Agencia de Medio Ambiente (AMA) de la Junta de Andalucía ha venido trabajando en los dos últimos años, en un proyecto de desarrollo metodológico para la obtención de estadísticas agrarias mediante imágenes de satélite, dentro del ámbito del Sistema de Información Ambiental de Andalucía-Sinamba (Agencia de Medio Ambiente, 1990).

El modelo utilizado, aunque apoyado básicamente en las recomendaciones de la CEE (CCR-Ispra) para este tipo de aplicaciones, adopta en muchos casos sistemas y soluciones propias. Este hecho está motivado por la política de desarrollo propio del software que se ha seguido en el campo del tratamiento digital de imágenes de satélite (paquete AMATEL), lo cual permite la construcción de aplicaciones a medida, completamente flexibles en su diseño y adaptables a las necesidades específicas de cada proyecto.

El objetivo de esta comunicación es exponer la metodología implementada, describiéndose sus líneas generales, parti-

cularidades y procesos que la componen, haciendo especial referencia a las distintas alternativas y variables estudiadas dentro del proceso de clasificación, los resultados obtenidos con las mismas y las conclusiones que se pueden extraer.

Los trabajos de experimentación se han desarrollado sobre datos reales correspondientes a tres zonas diferentes:

- Campiña de Sevilla. 2 fechas: Abril-89 y Agosto-89.
- Navarra. Marzo-88. Datos de campo e imagen facilitados por el INIA (González Alonso, 1990).
- Doñana. Mayo-91.

En todos los casos se ha trabajado con imágenes Landsat TM.

2. DESCRIPCIÓN DE LA METODOLOGÍA

2.1. Esquema de bloques.

En el esquema de la Figura 1 se muestra un diagrama de bloques de las diferentes fases que constituyen el proceso completo de tratamiento de la información y que se describen a continuación.

2.2. Procesos de preparación de la información.

Las zonas de muestreo o segmentos suponen un cubrimiento de alrededor del 1,5% de la superficie total, pudiendo tener cualquier forma y área, aunque normalmente se suelen emplear segmentos cuadrados de 700 x 700 m.

La información de partida proviene de dos fuentes. Por un lado los documentos producto del trabajo de campo, que

METODOLOGIA DE ESTADISTICAS AGRARIAS A PARTIR DE IMAGENES DE SATELITE (AMATEL)

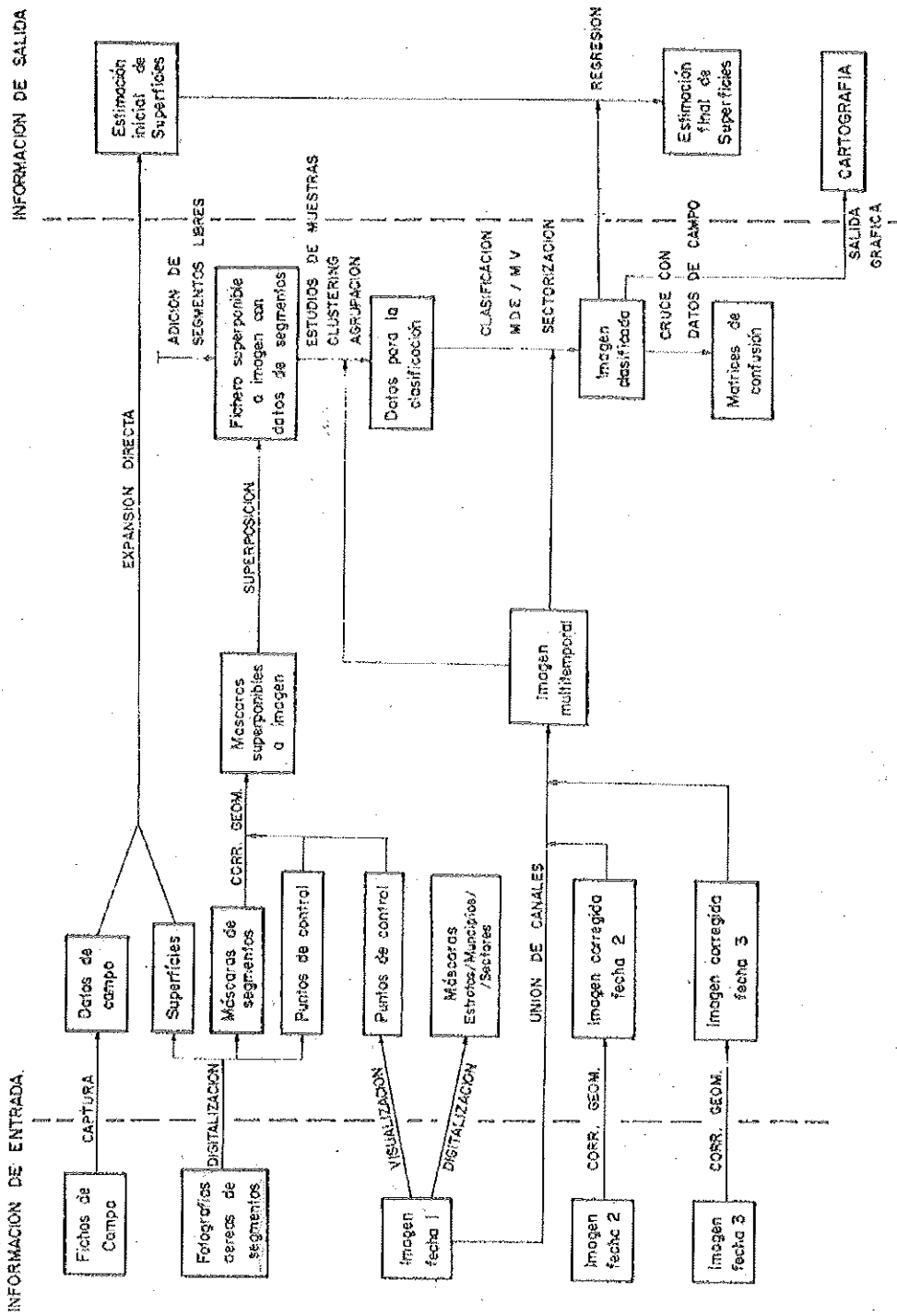


Figura 1.- Esquema de bloques.

son:

- fotografía aérea para cada segmento con su situación y su estructura de parcelas. Las escalas de trabajo apropiadas son del orden de 1:5000.
- ficha de datos para cada segmento recogiendo información de cada una de sus parcelas como su tipo de cubierta y el porcentaje de superficie sombreada, entre otras.

Su integración al sistema se realiza mediante una captura de datos para las fichas y mediante digitalización de los segmentos trazados sobre la foto aérea. La digitalización generará máscaras ráster de los segmentos, obteniendo simultáneamente las superficies de las parcelas. Con estos datos es posible realizar ya la estimación inicial de superficies por expansión directa.

La segunda fuente de información de entrada la constituye la imagen de satélite de la zona de estudio. Existe la posibilidad de trabajar a la vez con dos o tres imágenes de fechas diferentes.

El proceso de superposición de las máscaras de segmentos sobre la imagen, en su correspondiente posición real se resuelve mediante la técnica de toma de puntos de control homólogos sobre la foto aérea (durante la digitalización) y la imagen, para cada segmento. El motivo principal para adoptar esta solución radica en que las fotografías empleadas no disponen de puntos de referencia en coordenadas geográficas, lo cual anula la posibilidad de llevar segmentos e imagen a algún espacio de referencia común (por ejemplo coordenadas UTM).

Con estos puntos de control se somete a las máscaras de segmento a un proceso de corrección geométrica que las modifica para su casado con la imagen.

Tras estas etapas toda la información de los segmentos queda preparada para su estudio.

A partir de la imagen de satélite de la primera fecha se extraen mediante digitalización una serie de subdivisiones de la región de estudio que intervendrán de diversos modos en fases posteriores:

- estratos (para potenciar los análisis estadísticos);
- términos municipales u otras divisiones administrativas o temáticas (para la obtención desglosada de resultados) y
- sectores homogéneos en cuanto a tipología de cubiertas (para asistir al proceso de clasificación). Esta capa proviene normalmente del S.I.G., aunque se puede crear también de forma directa por digitalización.

Cuando se utilicen imágenes de más de una fecha, será necesario disponer en todas ellas de la misma estructura geométrica, para lo cual se corrigen respecto al espacio de coordenadas UTM, resolviéndose la superposición de segmentos a imagen en la primera de ellas. Aquí es importante destacar el empleo de un sistema de restitución por el método del vecino más próximo para evitar alterar la radiometría de la imagen, lo que no sería conveniente con vistas a futuros procesos de clasificación.

Si se recurre al análisis multitemporal, la imagen de trabajo se compone con la unión directa de los canales seleccio-

nados de cada fecha sobre un único fichero de imagen, interviniendo todos ellos conjuntamente en las fases de entrenamiento a la clasificación.

2.3. Procesos de análisis.

El análisis va a tener tres etapas:

- a) Preparación de datos para la clasificación.
- b) Clasificación.
- c) Cruce de resultados con los datos de campo.

a) En la primera fase se elaboran los datos de muestra de los segmentos de entrenamiento para que su formato sea el adecuado para arrancar la fase de clasificación. Entre estos procesos se encuentra el de obtención de las firmas espectrales de las clases, si el método de clasificación a utilizar es el de mínima distancia euclídea (CMD).

En este caso, se contempla la posibilidad opcional de realizar un proceso previo restrictivo, con el fin de seleccionar sólo aquellos píxeles de muestra más idóneos, como patrón representativo de su clase, en función de una serie de criterios del siguiente tipo:

- Se descartan parcelas con muy pocos píxeles
- Se descartan las parcelas más alejadas de la media del cultivo que representan
- Tras las etapas anteriores solo se consideran válidas las muestras que superan un determinado número de píxeles por cultivo

Como caso general, para el entrenamiento de la clasificación, se realiza siempre una preselección de píxeles válidos, eliminándose aquellos que corresponden a frontera entre parcelas (no puros) y los que representen a parcelas que en los datos de la ficha de campo aparezcan con un porcentaje de cubrimiento o sombreado por debajo de un umbral determinado.

El aspecto más significativo de esta etapa es la posibilidad de realizar un estudio de preclustering sobre las clases. Consiste en subdividir las clases que se elijan en una serie de subclases mediante la realización de una clasificación no supervisada de sus píxeles de muestra. La utilidad de este método está en que al clasificar sobre las subclases y unificarlas después, se obtienen en muchos casos mejores resultados que si se clasificara directamente sobre la clase única. Este tipo de análisis es opcional y permite determinar libremente cuántas subdivisiones crear para cada clase, aunque este número debe estar en consonancia con la cantidad de píxeles de muestra disponible para cada tipo de cubierta, con objeto de no generar subdivisiones con insuficiente representación. Esto constituye un importante grado de libertad a la hora de efectuar pruebas de clasificación.

Una de las posibilidades más potente de la metodología estriba en que es factible incorporar al conjunto de segmentos de entrenamiento ligados a las visitas de campo, nuevos "segmentos libres" o ventanas tomados sobre la imagen, que sean representativos de tipos de cubierta interesantes para mejorar la clasificación y el resultado cartográfico del análisis, y que no estuvieran bien recogidos en la muestra base. La influencia de estos segmentos adicionales está limitada a

las etapas de entrenamiento a la clasificación, por lo que no interfieren en los fundamentos estadísticos utilizados en la estimación inicial de superficies ni en la revisión de las mismas efectuada en el análisis de regresión final. También resulta útil su empleo como elementos para el test de los resultados de la clasificación desde un punto de vista cartográfico.

Entre las capacidades del sistema se encuentra también la posibilidad de realizar agrupaciones de cultivos de modo que se simplifique la gama de variedades de cubierta que intervienen en el análisis.

b) En la segunda fase se realiza la clasificación propiamente dicha. Se pueden emplear dos métodos alternativos: mínima distancia euclídea (CMD) y máxima verosimilitud (CMV).

Dentro del análisis de clasificación, el sistema prevé la alternativa de utilizar un método de asistencia externo que denominaremos sectorización. Consiste en utilizar una segmentación de la zona de estudio en una serie de regiones o sectores con una tipología de cubierta o uso homogéneo de modo que se le asocian a cada una de ellas qué subconjunto de clases son posibles que estén presentes y cuáles otros no. Esta herramienta resulta muy útil para eliminar aquellas confusiones debidas a comportamientos radiométricos similares de clases bien distintas, las cuales están ligadas a sectores del territorio fácilmente delimitables. Estos sectores se suelen configurar a partir de información temática derivada del Sistema de Información Geográfica, mediante su transformación a formato ráster compatible con las imágenes de satélite. El mayor rendimiento de este método se obtiene cuando los sectores se construyen específicamente para resolver confusiones tras las pruebas de clasificación. Esta técnica, además de aportar mejoras en los resultados de acierto de la clasificación respecto a los datos de campo, colabora sensiblemente en la calidad de las salidas cartográficas.

c) El cruce de los resultados de la clasificación con los datos de campo se realiza sobre un grupo de segmentos de test. Se obtienen datos sobre el grado de acierto cartográfico (píxel a píxel) en forma de matrices de confusión. Los valores resultantes indicarán el grado de acierto global de la clasificación, el de cada cultivo particular y entre qué clases y en qué grado se originan las confusiones.

2.4. Obtencion de resultados.

a) Numéricos.

La obtención de datos numéricos de superficie se realiza en dos fases: datos de estimación inicial por expansión directa y datos finales por regresión. En ambos se emplea la metodología estadística propuesta por la CEE, y en la que no entraremos aquí (Comisión de Comunidades Europeas-CCR Ispra, 1987). Los datos se calculan desglosados por estratos y por una segunda capa de subdivisión del territorio de libre elección (normalmente se usan los términos municipales). Los resultados van acompañados de parámetros que indican su precisión.

b) Gráficos.

La representación gráfica de la clasificación constituye un

producto de gran importancia, mediante el que se dispone de una distribución espacial de los distintos tipos de cubierta. Esta salida es uno de los objetivos primordiales de esta metodología y es empleada a su vez como elemento de juicio de la calidad del proceso, y como base, en su caso, para modificar algunas de las variables, con vistas a mejorar los resultados.

3. VARIABLES EXPERIMENTADAS EN LOS PROCESOS DE CLASIFICACIÓN

Dentro del proceso completo de clasificación, existen gran cantidad de factores y variables que permiten multitud de alternativas, con diferentes resultados. Como parte fundamental dentro del desarrollo de la metodología, se han efectuado gran cantidad de pruebas con el fin de conocer la influencia de estos factores sobre la calidad de la clasificación.

Expondremos a continuación las variables estudiadas, valorándose el comportamiento de las diferentes alternativas mediante el grado de acierto cartográfico de la clasificación respecto a los datos de campo, expresado mediante matrices de confusión.

* *Preselección de píxeles más representativos.*

Se ha probado esta posibilidad dentro de las clasificaciones por mínima distancia. Como ya se indicó en la exposición de la metodología, consiste en elegir de entre todos los píxeles de muestra de una clase determinada, aquellos que mejor la representen, eliminándose las parcelas de dicha clase cuyo comportamiento radiométrico se aleje más de la media y aquellas con un número de píxeles bajo.

El efecto observado, es que suele mejorar el resultado respecto de la clasificación directa, pero el empleo de clustering consigue mejoras mucho más importantes, por lo que este sistema pierde interés.

* *Tamaño de muestra.*

Es el número de píxeles por clase, de entre todos los aportados por los segmentos de entrenamiento, que se utilizarán como muestra en la clasificación de máxima verosimilitud. La experiencia que se deriva de las pruebas de clasificación realizadas indica que el número adecuado de píxeles de muestra está entre 150 y 200. Con menos no se consigue buena representatividad, mientras que un mayor número no suele estar disponible, teniéndose en cuenta que sólo se toman los puros y que al emplear subdivisiones o clustering el número de píxeles por subclase es aún menor.

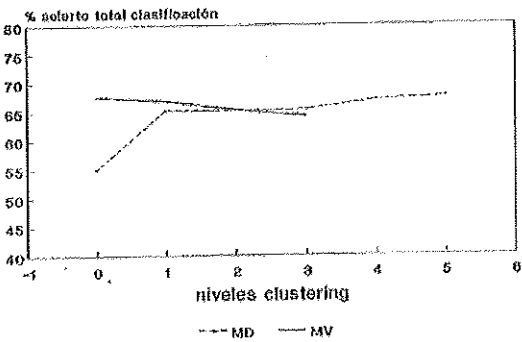
* *Bandas empleadas en el estudio.*

Se han realizado pruebas con distintas combinaciones de canales, índices y componentes principales. La conclusión extraída es que este es un factor que no tiene una influencia sustancial en el proceso, siendo lo más adecuado en la mayoría de los casos el empleo de todos los canales originales directamente (exceptuando el TM6) ya que los neocanales derivados de ellos no aportan mejoras de resultados.

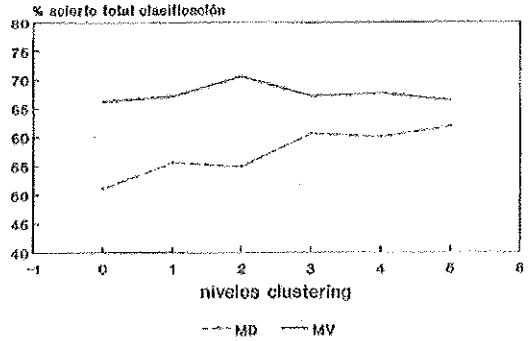
* *Multitemporalidad.*

El empleo simultáneo de los datos de campo y las imágenes de dos fechas diferentes, ha supuesto la obtención de porcentajes de acierto cartográfico más altos que en cada fe-

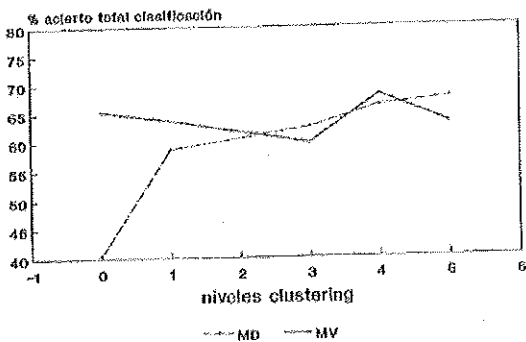
CAMPIÑA-SEVILLA (ABRIL)



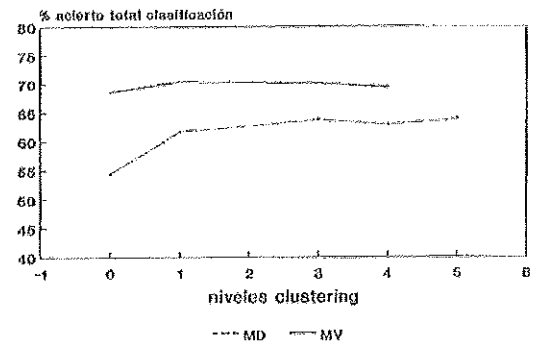
DOÑANA (MAYO)



CAMPIÑA-SEVILLA (AGOSTO)



NAVARRA



CAMPIÑA-SEVILLA (ABRIL+AGOSTO)

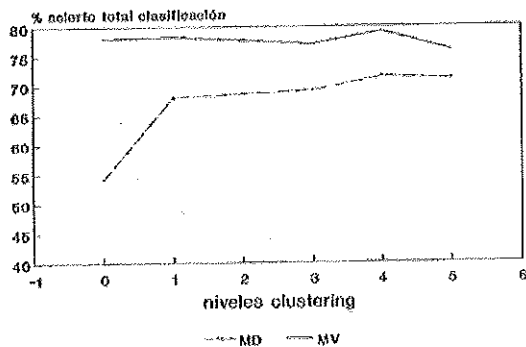


Figura 2.- Gráficas de clasificaciones completas.

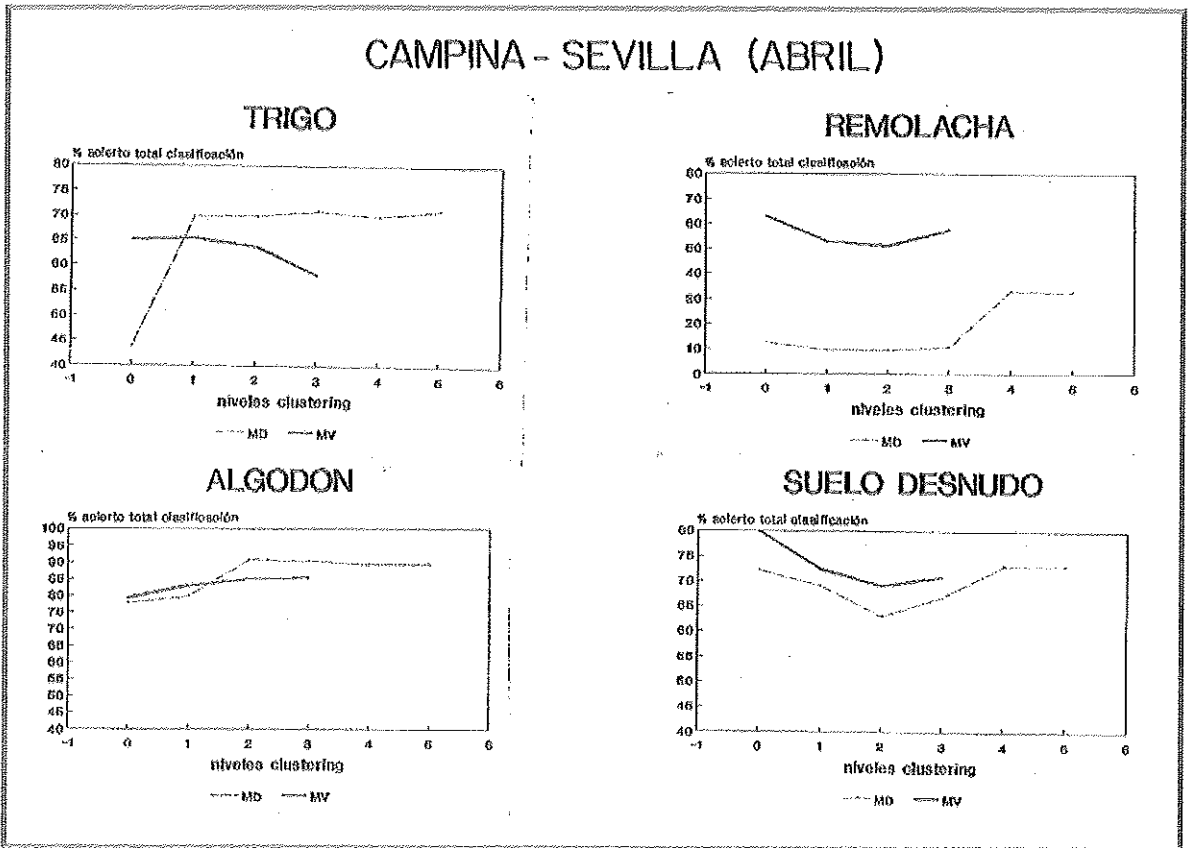


Figura 3.- Gráficos por cultivos dentro de una única clasificación.

cha por separado, principalmente para la clasificación de máxima verosimilitud.

La combinación de tres fechas no se ha mostrado demasiado efectiva, aunque en el único caso experimentado, una de las imágenes presentaba problemas de nubes.

Consideramos que la utilización de dos fechas complementarias que recojan entre ambas la presencia de los cultivos a estudiar, supone el caso más adecuado para la utilización del análisis multitemporal.

*** Sectorización.**

Constituye una herramienta muy efectiva para evitar confusiones radiométricas detectadas en el proceso de clasificación, entre cubiertas cuya presencia sobre el terreno se localiza en zonas de naturaleza distinta.

El análisis de sectores se realiza particularizado a cada caso, consiguiéndose evitar confusiones concretas y mejorar los parámetros de acierto en las matrices de confusión y la calidad de la cartografía.

*** Clustering + clasificación.**

El aspecto en el que más se ha profundizado a la hora de estudiar la incidencia de las distintas variables del proceso de clasificación ha sido la alternativa entre dos métodos de cla-

sificación: mínima distancia euclídea y máxima verosimilitud, en combinación con técnicas de clustering.

Se han empleado para ello distintos niveles de subdivisión de clases en función del número de píxeles de muestra disponibles, mediante proporcionalidad directa:

- . nivel 0 - sin clustering
- . niveles 1, 2 - clustering bajo (nº max. subdiv. =3)
- . nivel 3 - clustering medio (nº max. subdiv. =4)
- . niveles 4, 5 - clustering alto (nº max. subdiv. =6)

Para cada uno de estos niveles de clustering se ha realizado el tratamiento de clasificación por los dos métodos indicados.

Los resultados de acierto cartográfico obtenidos se muestran en los gráficos de las Figuras 2 y 3, reflejándose la evolución de cada clasificación para los distintos grados de subdivisión. Se han plasmado los comportamientos globales de las clasificaciones para el total de clases en cada una de las zonas estudiadas (Figura 2) así como los particulares de diversos cultivos significativos en una de ellas (Figura 3).

Sobre el eje vertical se representa el porcentaje de acierto cartográfico en clasificación y en el horizontal los niveles de clustering (el 0 corresponde a lo no realización del mismo).

Observando la vertical del valor 0 para el grado de subdivisión se comprueba la superioridad, en este caso, de la clasificación de máxima verosimilitud (CMV) respecto a la clasificación de mínima distancia euclídea (CMD).

En la Figura 2 (clasificaciones completas), al ir creciendo el nivel de subdivisión, la CMD mejora sensiblemente, llegando a igualar y en algún caso a superar a la CMV. Sin embargo la CMV parece verse poco afectada por el clustering, empeorando ligeramente al aumentar la subdivisión.

En la Figura 3 (cultivos particulares) las situaciones que se aprecian son bastante irregulares, siendo representativas de las obtenidas en los restantes casos de estudio. Hay casos en los que la CMD supera claramente a la CMV (trigo), casos en los que la CMD no funciona con una determinada firma espectral (remolacha), casos en que la CMV se mantiene siempre por encima de la CMD (suelo desnudo) y otros en los que ambas están parejas (algodón).

4. CONCLUSIONES GENERALES

- No hay comportamientos fijos para ninguna de las alternativas que permitan extraer conclusiones generalizables.
- Hay una influencia decisiva de la naturaleza concreta de los datos para cada caso, obteniéndose los mejores resultados por caminos diferentes para cada uno de ellos.
- La clasificación de mínima distancia es bastante sensible al clustering, obteniéndose en la mayoría de los casos mejoras progresivas al aumentar el nivel de subdivisión con respecto a su aplicación directa. Las mejoras numéricas de porcentajes de acierto suelen ser importantes. Los casos más favorables se consiguen con niveles altos de subdivisión.
- La clasificación de máxima verosimilitud es poco sensible al clustering, tendiendo normalmente a decrecer levemente el acierto al aumentar la subdivisión. Los casos más favorables suelen darse sin clustering o con niveles bajos.
- Considerando globalmente la totalidad de clases (Figura 2), la CMV suele dar mejores resultados que la CMD. Considerando cultivos individuales (Figura 3) hay muchas situaciones en que la CMD supera a la CMV.
- Sin clustering la CMV es mejor que la CMD, pero el empleo del mismo atenúa las limitaciones de la CMD convirtiéndola en alternativa válida.
- No es conveniente descartar ningún método a priori, ya que cualquiera de ellos puede ser el mejor en una situación concreta. Proponemos desarrollar siempre un banco de pruebas de clasificación de las que se deducirán las mejores o las más adecuadas en función de que el estudio vaya dirigido al conjunto o a determinados cultivos en particular.
- Debido a que cualquier variación en uno sólo de los factores que intervienen en el proceso de clasificación hace que el resultado de ésta varíe, no consideramos oportuno seleccionar una clasificación única para aplicarle el método de regresión y extraer los resultados fi-

nales. La solución propuesta consiste en elegir un grupo de clasificaciones (las que se comporten mejor según los objetivos), aplicar con cada una de ellas la regresión y realizar una combinación de sus resultados para llegar a los datos definitivos. De este modo, los valores calculados para las superficies serán más estables y fiables que si sólo aceptáramos como bueno el de una única clasificación.

- La mejora de los resultados cartográficos, pasa por el empleo de capas de información temática procedentes de un S.I.G.

5. BIBLIOGRAFÍA

- ✓ AGENCIA DE MEDIO AMBIENTE - CONSEJERÍA DE AGRICULTURA Y PESCA. JUNTA DE ANDALUCÍA (1990): El uso de las imágenes de satélite en el análisis de la distribución espacial de cultivos. Aplicación metodológica en el Bajo Guadalquivir. *Jornada Técnica*. 131 p.
- ✓ GONZÁLEZ ALONSO, F.; LÓPEZ SORIA, S. y CUEVAS GOZALO J.M. (1990): *Influencia de la información contenida en imágenes de satélite sobre la precisión de la estimación de la superficie cultivada de cereal en una zona piloto de la Comunidad Autónoma de Navarra*. Ministerio de Agricultura. Instituto Nacional de Investigaciones Agrarias. Monografías, nº73. 52 p.
- ✓ COMISIÓN DE COMUNIDADES EUROPEAS-CCR Ispra, (1987): *Proyecto Agricultura. Inventarios Regionales*. Condiciones generales y especificaciones.