

DISEÑO DE ENCUESTAS AGRÍCOLAS BASADO EN UN MARCO DE ÁREAS. Metodología y aplicación a los cultivos intensivos en Andalucía



COMUNIDAD EUROPEA



DISEÑO DE ENCUESTAS AGRÍCOLAS BASADO EN UN MARCO DE ÁREAS

Metodología y aplicación a los cultivos intensivos en Andalucía

Luis Ambrosio Flores *

Luis Iglesias Martínez **

Julio Montañés Mancera ***

Luis Alberto Rubio Pérez ***:

* Universidad Politécnica de Madrid. Departamento de Economía y Ciencias Sociales Agrarias.

** Universidad Politécnica de Madrid. Departamento de Ingeniería Cartográfica, Geodesia y Fotogrametría-
Expresión Gráfica.

*** Junta de Andalucía; Consejería de Agricultura y Pesca. D.G.I.G.A. Servicio de Estudios y Estadísticas.

PRÓLOGO

La necesaria capacidad de respuesta de la agricultura a los numerosos y vertiginosos cambios que están aconteciendo conlleva, entre otras cosas, mayores demandas de información. Ésta es pedida y utilizada por los centros de decisión para reducir la incertidumbre, debiendo ser fiable y estar disponible en el momento que se requiera.

Para dar respuesta a esa demanda creciente de información y conocimientos, la Consejería de Agricultura y Pesca de la Junta de Andalucía está trabajando, en el marco del Plan de Modernización de la Agricultura Andaluza, en un Sistema Integrado de Información Agraria (SIIA). Este Sistema debe servir para reunir, elaborar y difundir información a los agentes que intervienen en uno u otro de los eslabones de la cadena agroalimentaria. Los agricultores constituyen la unidad primaria de esa cadena y en ellos tiene su origen el SIIA.

En este documento se presenta una metodología para el diseño de muestras de agricultores, a cuyo desarrollo hemos contribuido, en el marco de un Convenio de Colaboración con la Universidad Politécnica de Madrid. Esta metodología tiene un importante potencial de utilización, permite usar el "marco de áreas" como base de ulteriores desarrollos que puedan abarcar diversos estadios y agentes de la cadena agroalimentaria, sirve para la recolección de datos y su tratamiento de modo que la información basada en los mismos tenga el grado de precisión deseado y disponibilidad inmediata. Todo ello se ilustra mediante su aplicación a los cultivos intensivos en Almería (Campo de Dalías, Níjar y Bajo Almanzora) y al cultivo del fresón en Huelva,

José Emilio Guerrero Ginel
Secretario General de Agricultura y Pesca

ÍNDICE

INTRODUCCIÓN

CAPÍTULO I. MARCOS DE ÁREAS. CONSTRUCCIÓN15

1.1. Marcos no estratificados con segmentos de límites geométricos.....	19
1.2. Marcos estratificados con segmentos de límites físicos	21
1.2.1. Procedimiento de construcción	21
1.2.2. Estudio de casos	24
1.2.2.1. Un marco de áreas para la zona del cultivo de fresón en Huelva	24
1.2.2.2. Marcos de áreas para las zonas de cultivos protegidos en Almería	27

CAPÍTULO II. SELECCIÓN DE LA MUESTRA33

II.1. Selección de la muestra de segmentos	35
II.1.1. Marcos en retícula no estratificados. Segmentos de límites geométricos	35
II.1.1.1. Muestreo aleatorio simple	37
II.1.1.2. Muestreo por bloques.....	37
II.1.1.3. Muestreo sistemático	37
II.1.1.4. Estudio de un caso: La muestra de segmentos de la 'Encuesta de superficies' del MAPA (1990).....	38
II.1.2. Réplicas.....	40
II.1.3. Marcos estratificados. Segmentos de límites físicos.....	41
II.1.3.1. Muestreo sistemático de réplicas	41
II.1.3.2. Muestreo de réplicas por bloques.....	41
II.1.3.3. Selección de la muestra de segmentos	42
II.1.3.4. Estudio de casos.....	42
II.1.3.4.1. Selección de una muestra sistemática de réplicas de segmentos en la zona de cultivo de fresón en Huelva	42
II.1.3.4.2. Selección de una muestra de réplicas de segmentos por muestreo de bloques en Campo de Dalías	44
II.2. La muestra de agricultores	45
II.2.1. Estudio de casos	45
II.2.1.1. Selección de una muestra de agricultores a partir de la muestra de segmentos en la zona de cultivo de fresón en Huelva.....	46
II.2.1.2. Selección de una muestra de agricultores a partir de la muestra de segmentos en Campo de Dalías.....	46
II.3. Rotación de la muestra.....	46

CAPÍTULO III. RECOLECCIÓN DE DATOS.....49

III.1. Recolección de datos para la estimación de superficies cultivadas y otros usos del suelo	51
III.2. Control de calidad de los datos recolectados	55
III.3. Recolección de datos mediante entrevistas a los agricultores.....	56

CAPÍTULO IV. ESTIMACIÓN	57
IV.1. Estimaciones a partir de la muestra de segmentos	59
IV.1.1. Marcos en retícula no estratificados	59
IV.1.1.1. Eficiencia relativa de las tres estrategias de muestreo consideradas	63
IV.1.1.2. Estudio de un caso: La muestra del MAPA (1990) para la estimación de la superficie cultivada y demás usos del suelo.	65
IV.1.2. Marcos estratificados	69
IV.1.2.1. Estimación del total y de la media	69
IV.1.2.1.1. Muestreo sistemático.....	69
IV.1.2.1.2. Muestreo por bloques	71
IV.1.2.2. Estimación de la razón: Razones entre variables.....	73
IV.1.2.2.1. Estimador de $COV(\hat{Y}, \hat{X})$: Muestreo sistemático.....	73
IV.1.2.2.2. Estimador de $COV(\hat{Y}, \hat{X})$: Muestreo aleatorio por bloques	74
IV.1.2.3. Estimación del cambio	75
IV.1.2.4. Estudio de casos.....	76
IV.2. Estimaciones a partir de la muestra de agricultores. Segmento ponderado.....	78
IV.2.1. Encuesta para estimar las características estructurales y técnico-económicas del cultivo de fresón en Huelva	79
IV.2.1.1. Características estructurales.....	82
IV.2.1.2. Características técnico-económicas	83
IV.2.2. Encuesta para estimar las características estructurales y técnico-económicas de los cultivos protegidos en Almería (Campo de Dalías)	86
IV.2.2.1. Características estructurales	88
IV.2.2.2. Características técnico-económicas	89
CAPÍTULO V. MARCOS MÚLTIPLES DE ÁREAS Y LISTAS	95
REFERENCIAS	97
INSTRUCCIONES SOBRE EL CD-ROM	100

AGRADECIMIENTOS

La Consejería de Agricultura y Pesca de la Junta de Andalucía ha impulsado la investigación y desarrollo metodológico en materia de estadísticas oficiales agrarias. En particular, en el marco de un Convenio de Colaboración con el Departamento de Economía y Ciencias Sociales Agrarias de la Universidad Politécnica de Madrid, ha contribuido al desarrollo de los métodos de muestreo que se presentan en este documento, y ha financiado los trabajos presentados como casos de estudio.

Queremos agradecer la colaboración en este trabajo de: Constancio González Rivera, Francisco Javier Salinas González y Cristina Vellilla Lucini, profesores del Departamento de Ingeniería Cartográfica, Geodesia y Fotogrametría de la Universidad Politécnica de Madrid, y a Manuel Mateos Hernández y Pedro Luis Martín Magro miembros de dicho Departamento. José Salas Arce ha colaborado en el diseño de algunos de los cuestionarios de muestreo. Carmen Marín Ferrer ha colaborado programando aplicaciones para la información gráfica, en el entorno de un Sistema de Información Geográfica. El trabajo de campo de las encuestas piloto realizadas para testar los métodos de muestreo, ha sido realizado por Rogelio Jalón Sánchez y Nicolás Peña Jaén.

INTRODUCCIÓN

Para los agentes económicos que intervienen en uno u otro de los eslabones que integran la cadena agroalimentaria –los agricultores y/o sus asociaciones profesionales, las industrias de transformación y comercialización, las asociaciones de consumidores y las Administraciones Públicas–, es cada vez más importante el disponer, en el proceso de su toma de decisiones, de información puntual y precisa sobre los diversos aspectos de la actividad agraria.

Las Administraciones Públicas, en general, y en particular las del Sector Agrario, principales productoras y usuarias de estadísticas, han respondido a esa demanda creciente de información, apoyando la investigación y el desarrollo metodológico en materia de estadísticas oficiales, por una parte, y creando una nueva normativa sobre esta materia, por otra [Planes Estadísticos de la Unión Europea, Nacionales y Autonómicos]. La Unión Europea ha jugado el papel de catalizador de ese desarrollo metodológico en los últimos años, en particular para las estadísticas relativas a superficies de cultivo y rendimiento de cosechas. El proyecto MARS –Monitoring Agriculture with Remote Sensing (Decisión CEE 88/503)–, iniciado en 1988 con vistas al establecimiento de un nuevo Sistema de Información Agraria, impulsó en España la investigación y el desarrollo metodológico sobre esta materia.

En 1990, la Unidad de Estadística del Ministerio de Agricultura Pesca y Alimentación inició, en colaboración con el Departamento de Economía y Ciencias Sociales Agrarias de la Universidad Politécnica de Madrid y los Servicios de Estadísticas Agrarias de algunas Comunidades Autónomas, entre ellas la de Andalucía, el desarrollo de un nuevo procedimiento para la elaboración de las estadísticas sobre superficies de cultivo y rendimientos de cosechas: la “Encuesta de superficies y rendimientos de cultivos” [MAPA(1990), Ambrosio y Gallego (1994), FAO(1996, 1998)].

La ‘Encuesta de superficies’ del MAPA

Se trata de una muestra sistemática de áreas (segmentos), seleccionada a partir de un marco de áreas no estratificado, cuyo soporte es el Mapa Topográfico Nacional 1:50.000, y en el que los segmentos son cuadrados de 700m por 700m, inscritos en la malla UTM de cuadrícula 1Km por 1Km. La información necesaria para la estimación de superficies se observa directamente sobre los segmentos de la muestra, en una sola visita realizada entre Abril y Junio.

Para los cultivos extensivos, la precisión de las estimaciones de la ‘Encuesta de superficies’ es satisfactoria. Sin embargo, para los cultivos intensivos y, en general, para todos los cultivos muy localizados en el espacio, las estimaciones de la “Encuesta” no alcanzan el nivel de precisión requerido. La concentración espacial de los cultivos intensivos determina que, en unos casos, la tasa de muestreo de la “Encuesta” resulte insuficiente, mientras que en otros la muestra puede ser de tamaño adecuado pero estar mal repartida. Por otra parte, los cultivos intensivos, en particular los hortícolas, se suceden a lo largo del año dentro de una misma parcela, en función de las condiciones cambiantes del mercado, por lo que la información directamente observada sobre los segmentos de la muestra en una sola visita, es insuficiente para estimar la superficie total anual de cada cultivo intensivo y la del conjunto de ellos.

La 'Encuesta sobre cultivos intensivos' de la Consejería de Agricultura y Pesca de la Junta de Andalucía.

La Consejería de Agricultura y Pesca de la Junta de Andalucía, a través de su Servicio de Estudios y Estadísticas y en colaboración con el Departamento de Economía y Ciencias Sociales Agrarias de la Universidad Politécnica de Madrid, ha desarrollado un procedimiento de diseño de encuestas agrícolas basadas en un marco de áreas para la estimación de superficies, rendimientos y otras características estructurales o técnico-económicas de los cultivos hortícolas y, en general, de los cultivos intensivos. Se trata de una muestra aleatoria de segmentos seleccionada a partir de un marco de áreas estratificado, con segmentos de límites físicos, identificables y permanentes [FAO (1996)] y de una muestra aleatoria de cultivadores, seleccionada a partir de la muestra de segmentos. La estratificación permite ajustar la tasa de muestreo a las características de los distintos estratos y mejorar el reparto de la muestra ajustando, en cada estrato, el tamaño del segmento al idóneo. La muestra de cultivadores permite recoger la información necesaria y no observada directamente sobre la muestra de segmentos, para la estimación de la superficie y el rendimiento de los cultivos practicados a lo largo del año, así como para la estimación de cualquier otra característica estructural o técnico-económica.

En este documento se presenta, paso por paso, una metodología para el diseño de encuestas agrícolas basadas en un marco de áreas: (i) Construcción del marco de áreas (capítulo I), (ii) Selección de la muestra de segmentos y de la muestra de agricultores (capítulo II), (iii) Recogida de información (capítulo III) y (iv) Cálculo de las estimaciones y del error de estimación (capítulo IV). A modo de casos de estudio, se presentan las encuestas diseñadas con esta metodología para la caracterización estructural y técnico-económica del cultivo de fresón, en Huelva, y de los cultivos protegidos en Almería (Campo de Dalías, Níjar y Bajo Almanzora).

CAPITULO I
MARCOS DE ÁREAS. CONSTRUCCIÓN

I. MARCOS DE ÁREAS. CONSTRUCCIÓN

Para la selección de una muestra aleatoria se requiere un marco de muestreo. Un marco de muestreo es el conjunto de materiales que sirven para delimitar, identificar y facilitar el acceso a los elementos de la población a muestrear. El marco también incluye información auxiliar, que es utilizada en el diseño del procedimiento de selección de la muestra y/o en el proceso de estimación [Groves (1989)].

Un marco de muestreo deseable es aquel que reúne las siguientes características:

- (I) Comprende a todos los elementos de la población, sin duplicaciones.
- (II) Proporciona un medio para identificar inequívocamente cada unidad de muestreo.
- (III) Ordena las dos características anteriores de modo que la selección aleatoria de la muestra puede ser llevada a cabo de manera eficiente.

Los marcos de muestreo utilizados son marcos de lista, ó marcos de áreas, ó marcos mixtos de áreas y lista. En agricultura, el ejemplo más ilustrativo de marco de lista es un directorio de explotaciones agrarias. Un marco de áreas es una partición o segmentación en unidades de muestreo, a las que se denominan "segmentos", del territorio sobre el que se asienta una población. La segmentación se realiza sobre mapas, fotografías aéreas, imágenes de satélite o cualquier otra representación gráfica del territorio. Los elementos de la población se identifican unívocamente a partir del segmento al que pertenecen.

Es sabido que los marcos de lista raramente reúnen la característica (I), ni tampoco la (II) por omisión o repetición de algún elemento de la población; en cambio suelen incluir información auxiliar muy valiosa para el diseño del procedimiento de selección de la muestra y del procedimiento de estimación. Los marcos de áreas suelen reunir las tres características deseables pero contienen poca información auxiliar, útil para el diseño del procedimiento de selección de la muestra y para el de estimación. El marco mixto trata de reunir las ventajas de los marcos de lista y de áreas (ver capítulo V).

El marco de áreas presenta numerosas e importantes ventajas respecto del marco de lista, aunque también algunos inconvenientes [U.S.D.A.(1983), Cotter y Nealon (1987)]:

A) VENTAJAS

- (1) **Versatilidad.** Los posibles usos de un marco de áreas son muy numerosos: puede ser utilizado tanto para recoger datos directamente observables sobre el terreno (superficies cultivadas

y otros usos del suelo, rendimientos de cosechas y cualquier otra característica ligada al suelo), como para recoger datos no observables directamente sobre el terreno tales como ciertos datos relativos a la economía de las explotaciones agrarias (factores de producción, producciones, costes de producción, ingresos, consumos intermedios). Los datos no directamente observables sobre el terreno se recogen mediante entrevistas a una muestra aleatoria de agricultores u otros agentes, seleccionada a partir del mismo marco de áreas.

- (2) **Total cobertura.** El marco de áreas es completo en el sentido de que comprende siempre a todos y cada uno de los elementos de la población asociada al marco, con tal de que no se excluya del mismo área alguna. Esto es importante porque es una condición necesaria para obtener estimaciones insesgadas de las características en estudio.
- (3) **Medida de la precisión.** El marco de áreas presenta la ventaja de permitir la construcción de estimadores insesgados cuya precisión puede ser estimada a través de la varianza. El coeficiente de variación es una medida del grado de precisión de las estimaciones (error de muestreo), derivado de la varianza y de uso más frecuente que ella.
- (4) **Reducción de los errores ajenos al muestreo.** La fotografía aérea es uno de los elementos básicos del muestreo de áreas. Su empleo durante la recogida de datos, en particular si estos se recogen mediante entrevistas, permite al encuestado visualizar las parcelas a las que se refieren las preguntas del cuestionario, lo que contribuye a evitar errores en las respuestas. Asimismo, sobre la base de la fotografía es posible recoger cierta información que servirá para el posterior tratamiento del problema de la "no respuesta".
- (5) **Longevidad.** El marco de áreas es duradero: puede ser utilizado durante años sin necesidad de actualización de las unidades de muestreo. La vida media de un marco de áreas se estima entre 15 y 20 años.
- (6) **Reducción de los costes de muestreo.** En un marco de áreas las unidades de muestreo son conglomerados de elementos de la población (parcelas o explotaciones). El muestreo de conglomerados permite una reducción de los costes del muestreo, al reducir los costes de desplazamiento entre unidades de muestreo.

B) INCONVENIENTES

- (1) **Actividades menores subrepresentadas.** Las actividades raras o muy localizadas no estarán bien representadas en una muestra de áreas. Sin embargo, para estos casos es posible desarrollar marcos suplementarios: marcos mixtos de áreas y listas (ver capítulo 5).
- (2) **Sensibilidad a los "outliers".** En una muestra de áreas, la presencia de ciertas unidades de muestreo extremas ("outliers"), muy distintas de la media, puede influir muy desfavorablemente sobre las estimaciones y su precisión. Una solución a este problema consiste en identificar los "outliers" antes de la selección de la muestra y en muestrearlos independientemente de los "no-outliers", o bien utilizar estimadores robustos respecto de este problema (se tratará de este tema en el capítulo V).

I.1. MARCOS NO ESTRATIFICADOS CON SEGMENTOS DE LÍMITES GEOMÉTRICOS

Los propios sistemas cartográficos establecen ya una partición del territorio, de modo que son en sí mismos marcos de áreas. La proyección UTM (Universal Transversa Mercator) de meridianos y paralelos son rectas. A partir de este sistema de referencia es posible construir una retícula, mediante rectas paralelas a los ejes de coordenadas. La retícula así definida es una partición del territorio en estudio y, como tal, es un marco de áreas en el que el segmento es la cuadrícula resultante. En este marco, una unidad de muestreo queda unívocamente definida por las coordenadas (UTM) de uno de sus vértices (Gallego et al (1994)).

En España, el Mapa Topográfico Nacional a escala 1:50.000 (MTN50) (en proyección UTM), recubre todo el territorio nacional y de él se derivan los de menor escala, por lo que es una buena base para un marco de áreas a nivel nacional. La Encuesta sobre Superficies y Rendimientos de Cultivos, que anualmente lleva a cabo el Ministerio de Agricultura Pesca y Alimentación (MAPA), utiliza este marco de áreas [M.A.P.A. (1990), Ambrosio y Gallego (1994), F.A.O. (1998)]. La figura 1 muestra la disposición de las Hojas del MTN50 correspondientes a Andalucía y la figura 2 el detalle de una de las hojas editadas por el Servicio Geográfico de Ejército. En estas hojas figura impresa la retícula de malla 1 Km. por 1 Km. (cuadrícula kilométrica) y las coordenadas UTM de cada uno de sus vértices. Con trazo más grueso figura también la malla de 10 Km. por 10 Km. (bloque), que agrupa a 100 cuadrículas kilométricas.

Las coordenadas de un vértice cualquiera de una cuadrícula kilométrica cualquiera permiten identificar el bloque al que pertenece y su posición dentro del mismo: la última cifra de la abscisa y la última de la ordenada corresponden respectivamente a la abscisa y la ordenada de la cuadrícula kilométrica respecto del vértice Suroeste del bloque al que pertenece, tomado como origen; las restantes cifras definen las coordenadas del bloque, respecto del origen correspondiente al huso. La posición de la cuadrícula kilométrica dentro del bloque juega un papel esencial en el procedimiento de selección de la muestra. El código de identificación de los segmentos consta del número de la Hoja del MTN50 a la que pertenece (fila y columna en la figura 1, contando desde el vértice Noroeste del mapa de España) y de las coordenadas del vértice Suroeste de la cuadrícula de la malla UTM de 1 Km. por 1 Km.

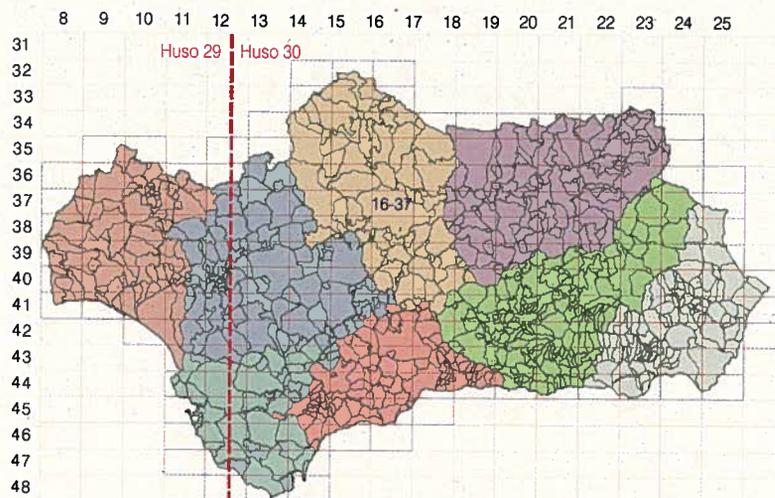


Figura 1. Hojas del Mapa Topográfico Nacional a escala 1:50.000 correspondientes a la Comunidad Autónoma de Andalucía.

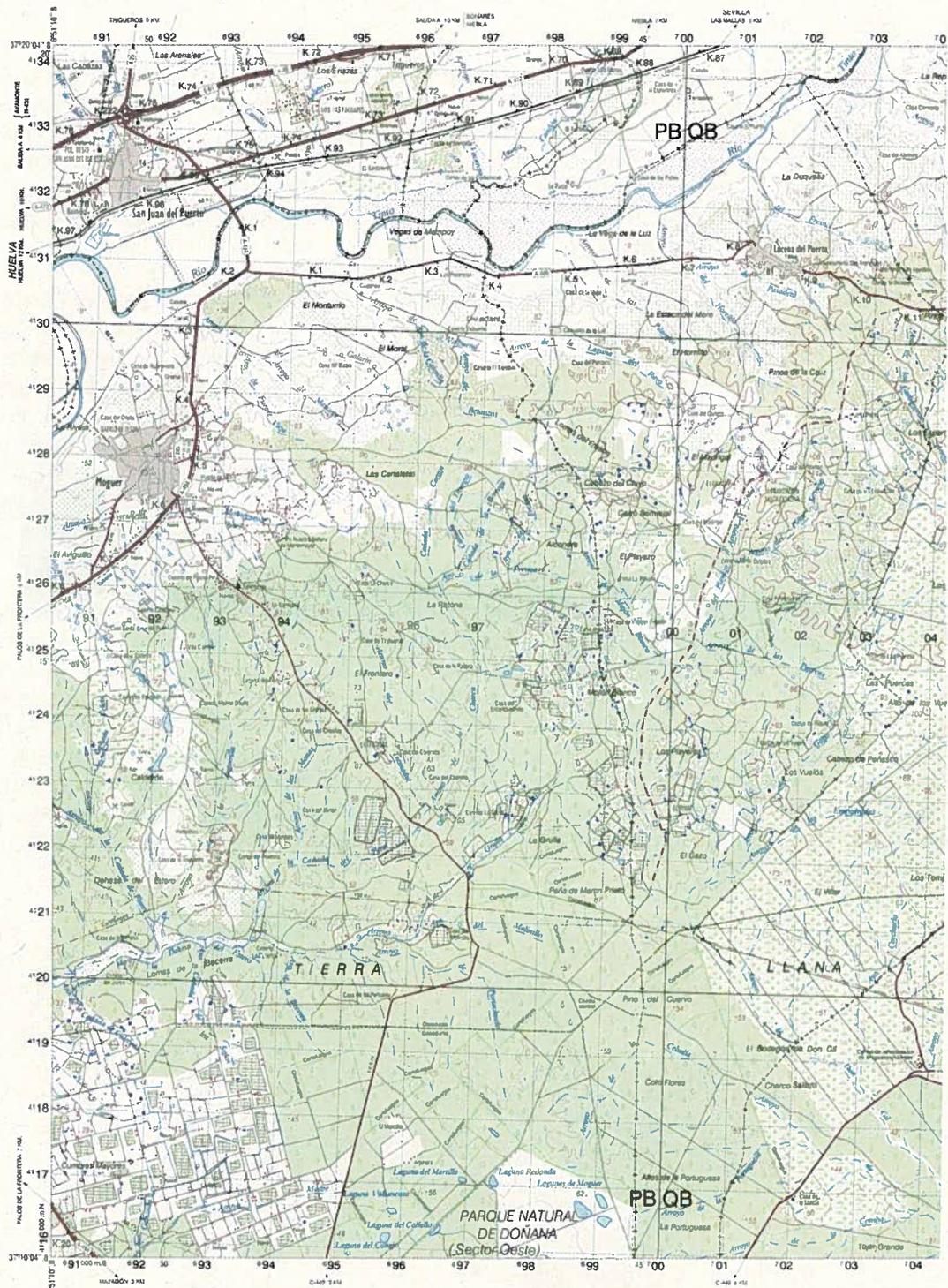


Figura 2. Reproducción de la Hoja 10-41 (1.000) de la Serie L, escala 1:50.000, del Servicio Geográfico del Ejército.

El segmento utilizado en la "Encuesta" del M.A.P.A.: tamaño y forma.

En España, la cuadrícula de 1 Km. por 1 Km. (100 hectáreas) resulta grande como unidad de muestreo para la investigación en campo, dado el elevado grado de parcelación de la superficie cultivada: por término medio comprende 43 parcelas. Una de las fases del trabajo de campo consiste en deslindar las parcelas contenidas en el segmento y en asignarles el código correspondiente al uso del suelo. Los errores de deslinde y asignación de códigos (errores ajenos al muestreo) crecen con el número de parcelas por unidad de muestreo, según se ha observado en las experiencias realizadas en varios Estados miembros de la UE. En estas experiencias se ha constatado que con un tamaño del segmento que corresponda a un número de parcelas comprendido entre 15 y 25, dichos errores se mantienen dentro de límites tolerables. En España ese número de parcelas corresponde a unidades de muestreo de entre 35 y 58 hectáreas.

Sobre la base de las consideraciones que preceden, el segmento finalmente adoptado en la "Encuesta" del M.A.P.A. es una cuadrícula de 700 m. por 700 m. (49 hectáreas), inscrita en la cuadrícula UTM de 1 Km. por 1 Km. y trazada a partir del vértice Suroeste de la misma, en la dirección de sus lados.

I.2. MARCOS ESTRATIFICADOS CON SEGMENTOS DE LÍMITES FÍSICOS

A la hora de observar los datos en campo, la localización sobre el terreno de un segmento a partir sólo de las coordenadas UTM, puede plantear dificultades, lo que puede dar lugar a errores ajenos al muestreo. Para evitar este tipo de errores puede ser conveniente en ocasiones, hacer coincidir los límites del segmento con accidentes naturales o artificiales, fácilmente identificables en el terreno, tales como (por orden de preferencia) autopistas, autovías, carreteras nacionales, comarcas o locales, ferrocarriles, ríos, canales de riego o arroyos. Este tipo de segmentación sería muy costoso de llevar a cabo, en territorios de tamaño medio o grande. Sin embargo, los costes pueden reducirse notablemente si el marco se construye por etapas, estableciendo una jerarquía de particiones.

I.2.1. PROCEDIMIENTO DE CONSTRUCCIÓN

El procedimiento de construcción de un marco de áreas de segmentos con límites físicos, puede ser dividido en la siguiente serie ordenada de etapas (ver figura 3):

1ª) Acopio de material

El procedimiento se inicia haciendo acopio del material necesario. Básicamente se requieren: (I) mapas topográficos a escalas diversas (usualmente 1:50.000 y 1:10.000), (II) fotografías aéreas del vuelo más reciente, en el número mínimo necesario para cubrir el territorio de estudio (generalmente, si el recubrimiento longitudinal es elevado, bastaría con adquirir las fotografías alternas dentro de una misma pasada) y (III) material de papelería específico para trabajar sobre el material cartográfico y fotográfico, tal como papel transparente, lapiceros grasos y rotuladores de tinta indeleble de distintos colores.

Las imágenes de satélite, pueden ser sustitutivas o complementarias de dicho material básico. Los mapas temáticos (mapas de usos del suelo) y los modelos digitales del terreno pueden ser también un material útil en algunas etapas del proceso, especialmente en la de estratificación del territorio.

2ª) Delimitación de la zona de estudio

La zona en estudio se delimita sobre los mapas topográficos, usando como límites accidentes naturales o artificiales del terreno fácilmente identificables y duraderos. Si se requieren estimaciones precisas a distintos niveles, por ejemplo distintas demarcaciones administrativas, se deben trazar también los límites de esas demarcaciones.

3ª) Estratificación

El territorio delimitado se estratifica, utilizando como criterio la intensidad de uso del suelo observada sobre las fotografías aéreas. A este fin, conviene construir un mosaico con las fotografías aéreas. En primer lugar se delimita un estrato de zonas urbanas. El resto del territorio se estratifica en 4 o 6 estratos en orden decreciente de la intensidad de utilización del suelo. Los límites de estrato se hacen coincidir con accidentes naturales o artificiales del terreno y se trazan sobre el mosaico de fotografías aéreas, utilizando un color para cada estrato. Desde las fotografías, los límites se trasladan a los mapas topográficos y se digitalizan con el objeto de disponer de ellos en soporte informático.

4ª) Delimitación de las Unidades Primarias

Una Unidad Primaria (UP) es un área de terreno que contiene un cierto número de segmentos. Un segmento es cada una de las partes en que se dividen las Unidades Primarias con vistas al muestreo de áreas: es la unidad última de muestreo. Sólo las Unidades Primarias de la muestra serán objeto de segmentación.

El tamaño del segmento determina el tamaño de las Unidades Primarias. En encuestas sobre aspectos técnico-económicos de las explotaciones agrarias, una Unidad Primaria debe constar de entre 6 y 12 segmentos. El tamaño mínimo de una Unidad Primaria debería ser de 2 segmentos y el máximo de 15 segmentos. El tamaño del segmento está condicionado por las disponibilidades de límites físicos y permanentes del terreno: en las áreas de agricultura intensiva abundan dichos límites y los segmentos podrán ser pequeños. Es posible determinar un tamaño óptimo: el que haga máxima la precisión de las estimaciones para un coste dado. En principio, un buen criterio para establecer el tamaño del segmento consiste en elegir como tal el que pueda ser muestreado por un encuestador durante una sola jornada. El número medio de entrevistas que es posible realizar en una jornada está entre 10 y 15. Con este criterio, si el tamaño medio de una explotación intensiva es de 1 hectárea, el tamaño del segmento debería ser de entre 10 y 15 hectáreas. El tamaño de las Unidades Primarias debería estar, por tanto, entre 60 y 180 hectáreas. En áreas de agricultura extensiva el tamaño del segmento y, por tanto el de las UP, podría ser muy superior.

La delimitación de las Unidades Primarias se lleva a cabo dentro de estratos, utilizando el mismo color que el del estrato. Se empieza con el estrato más intensivo y se termina con el menos intensivo. El objetivo al delimitar las Unidades Primarias es el de que cada una de ellas sea una representación en pequeño del estrato; de manera que si en el estrato hay un 40% de tierra cultivada se debe procurar que cada Unidad Primaria contenga esa proporción de cultivo. Este trabajo de delimitación de las Unidades Primarias se realiza sobre los mapas topográficos utilizando como material auxiliar el mosaico fotográfico. Una vez delimitadas se requiere medir el área de cada Unidad Primaria, ya sea mediante planímetros o mediante medios informáticos.

Puede ser conveniente el disponer de todos los límites (los de la zona, los límites entre estratos y entre Unidades Primarias dentro de estratos) en soporte informático para lo que se requiere su digitalización. De esta forma pueden ser tratados mediante Sistemas de Información Geográfica (SIG),

lo que facilita el cálculo de las áreas de cada Unidad Primaria, además de facilitar el tratamiento de toda la información gráfica.

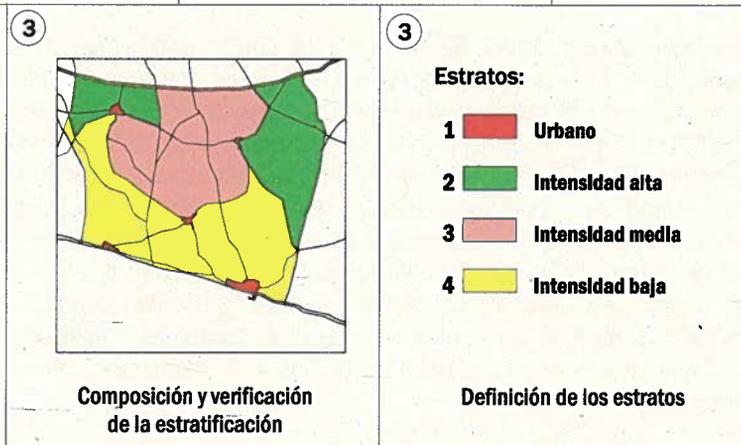
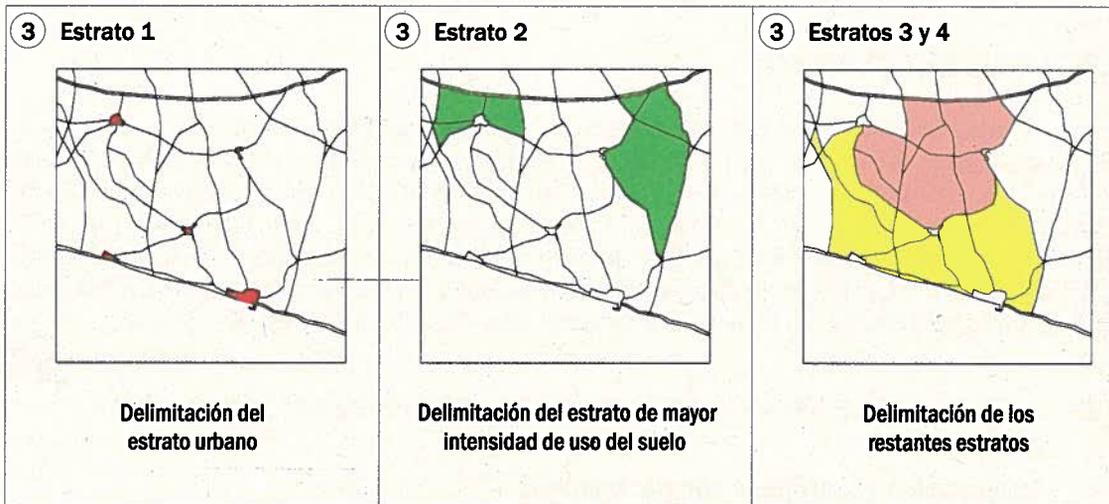
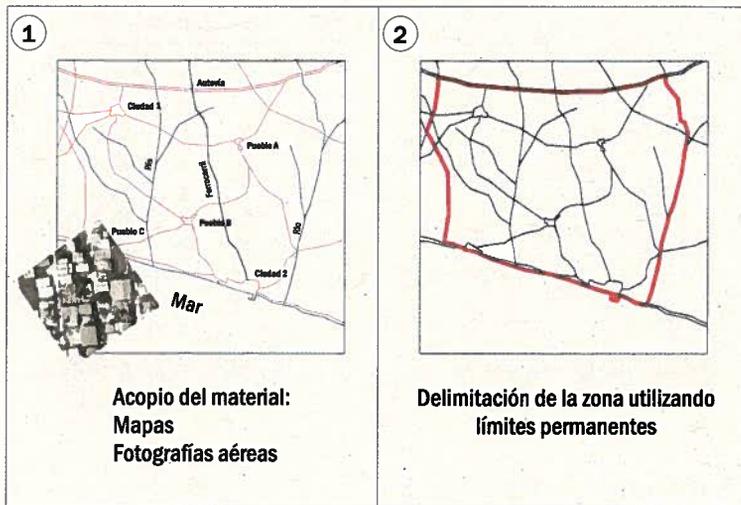


Figura 3. Construcción de un marco de áreas con límites físicos.

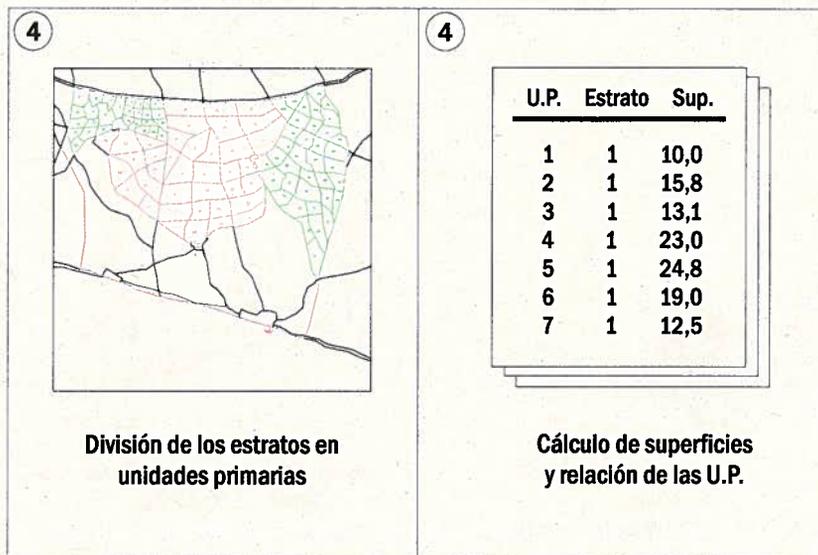


Figura 3. (Continuación) Construcción de un marco de áreas con límites físicos.

I.2.2. ESTUDIO DE CASOS

Se estudiarán dos casos de construcción de un marco de áreas estratificado con segmentos de límites físicos: (I) uno es el marco de áreas construido en la zona del cultivo de fresón en Huelva y (II) el otro los marcos de áreas construidos para tres zonas de cultivos protegidos en Almería: Campo de Dalías, Níjar y Bajo Almanzora. En ambos casos se utilizó como material cartográfico el Mapa Topográfico Nacional a escala 1:50.000 (MTN50) y el Mapa Topográfico de Andalucía a escala 1:10.000 (MTR10). Las fotografías aéreas utilizadas en Huelva fueron las de un vuelo de 1994 a escala 1:18.000, mientras que en Almería se utilizaron fotografías de 1998 a escala 1:40.000.

I.2.2.1. UN MARCO DE ÁREAS PARA LA ZONA DEL CULTIVO DE FRESÓN EN HUELVA

Delimitación y estratificación del territorio

En Huelva se delimitó una zona (192.547 ha.) que comprende la práctica totalidad de la superficie de fresón cultivada en la provincia. Esta zona, cuyos límites se hicieron coincidir con accidentes naturales (costa atlántica) o artificiales (carreteras) del terreno, se dividió en seis estratos de mayor a menor intensidad de uso del suelo. El primer estrato comprende todas las zonas urbanas (6.849 ha.). El segundo estrato (66.330 ha.) es el de "utilización intensiva del suelo" (más del 75 % de la superficie total del estrato es superficie cultivada). Dentro de este estrato se delimitó el subestrato "fresón" (25.052 ha.), en el que está contenido la práctica totalidad de la superficie de fresón. El tercer estrato, es el de "intensidad media de utilización del suelo", (entre el 25 y el 75 % de la superficie del estrato es superficie cultivada), y el cuarto es el de "intensidad baja de utilización del suelo", (menos del 25 % de la superficie del estrato es superficie cultivada). Finalmente se delimitaron dos estratos, uno en el que se incluyen los cauces de los ríos y las marismas y otro en el que se incluyen los espacios naturales.

En la figura 4 se muestra los límites de la zona y de los estratos para la zona del cultivo de fresón en Huelva.

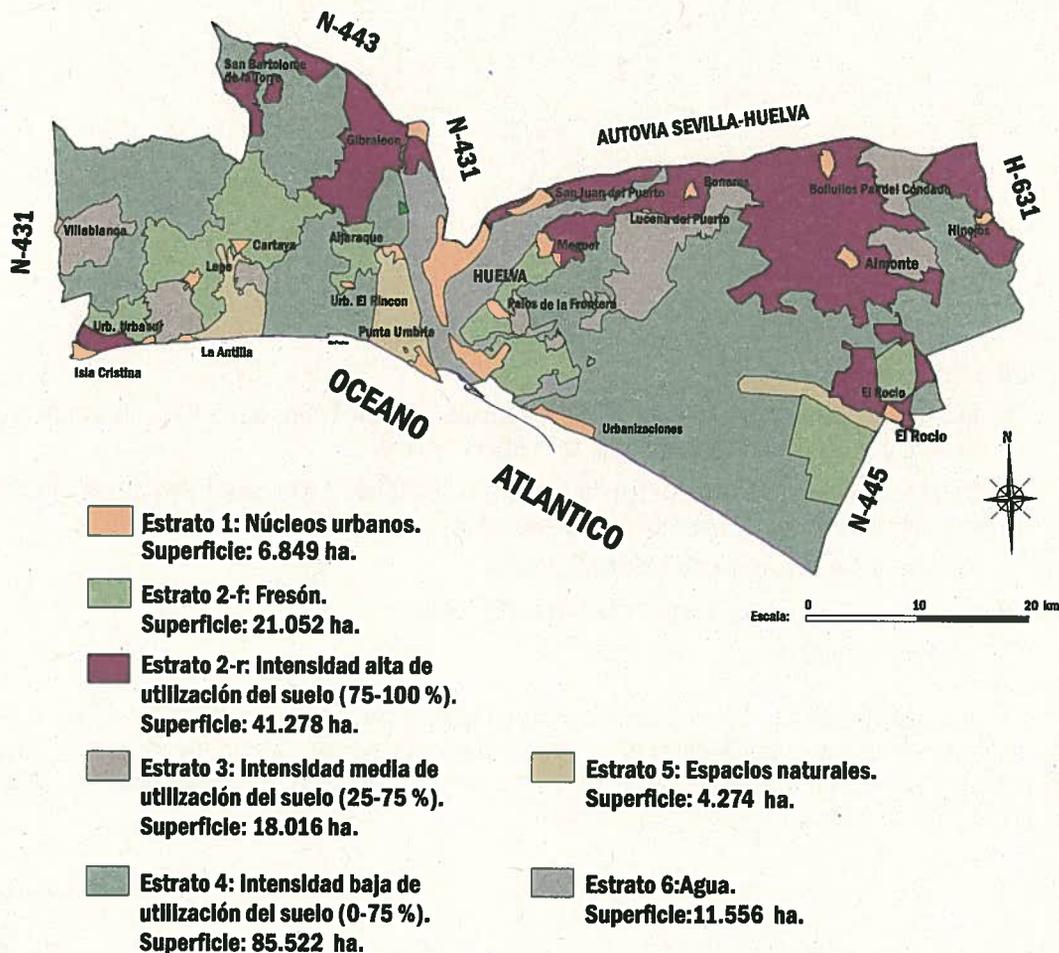


Figura 4. Límites de la zona y de los estratos para la zona del cultivo de fresón en Huelva.

Unidades Primarias

Los estratos agrícolas, esto es, los estratos 2, y 3 se dividieron en Unidades Primarias. El tamaño de las Unidades Primarias difiere de uno a otro estrato en función del tamaño del segmento. El tamaño idóneo del segmento se fijó en 15 ha. para el estrato de fresón, 25 ha. para el resto del estrato 2 (utilización intensiva del suelo) y 50 ha. para el estrato 3 (intensidad media de utilización). El tamaño idóneo de una Unidad Primaria se fijó en 10 segmentos, esto es, 150 ha. en el estrato de fresón, 250 ha. en el resto del estrato 2 y 500 ha. en el estrato 3.

El tamaño último de cada Unidad Primaria varía entorno al idóneo, en función de la disponibilidad de límites físicos identificables sobre el terreno (la media es de 162,68 ha. en el estrato de fresón, 299,12 ha. en el resto del estrato 2 y de 529,88 ha. en el estrato 3). En el archivo UPHUEL.V.TXT del disco compacto anejo a este documento se recoge una relación de las 361 Unidades Primarias delimitadas dentro de estos tres estratos, con indicación de sus superficies. La figura 5 muestra los límites de las Unidades Primarias delimitadas en los estratos 2 y 3. Cada Unidad Primaria se identifica mediante un código numérico que consta de los siguientes elementos dispuesto en la forma que se indica:

1	2	3	4
5		6	
7	8		
9			

1. Columna del MTN50.
2. Fila del MTN50.
3. Columna del MTR10.
4. Fila del MTR10.
5. Abscisa (X) del vértice suroeste de la cuadrícula UTM de 1 Km. por 1 Km. situada más próxima al vértice inferior izquierdo de la Unidad Primaria.
6. Ordenada (Y) del vértice suroeste de la cuadrícula UTM de 1 Km. por 1 Km. situada más próxima al vértice inferior izquierdo de la Unidad Primaria.
7. Estrato al que pertenece la Unidad Primaria.
8. Número correlativo de orden de la Unidad Primaria.
9. Superficie de la UP.

La relación de 361 Unidades Primarias, junto a la información gráfica sobre las mismas, constituye el marco de áreas, a partir del cual se seleccionará una muestra de segmentos y de cultivadores de fresón. En el disco compacto anejo a este documento se incluye el marco de áreas construido para la zona de cultivo de fresón, en Huelva.

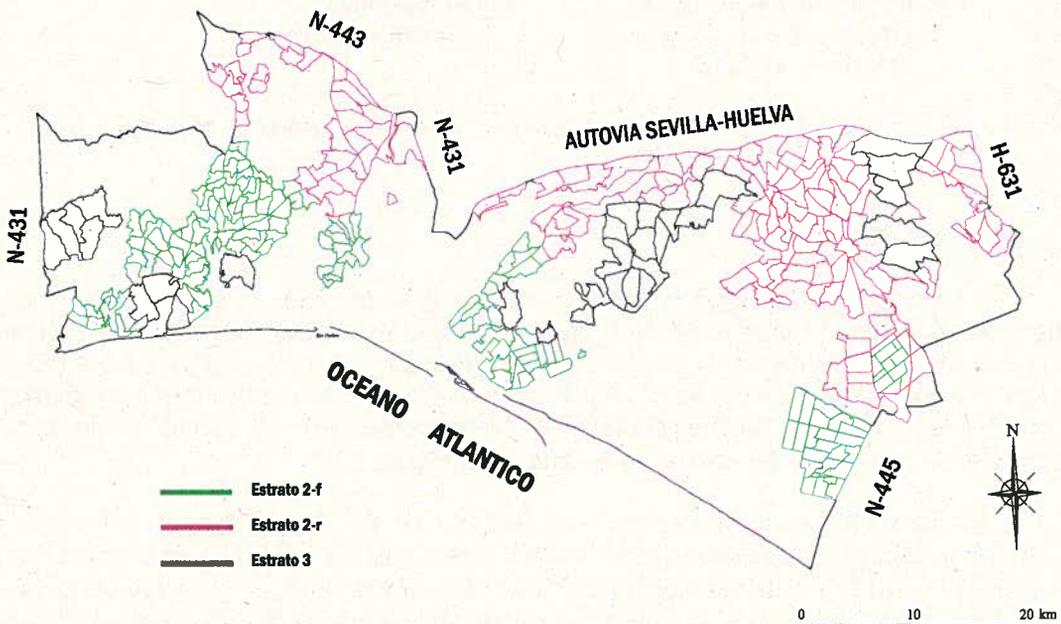


Figura 5. Límites de las Unidades Primarias en los estratos 2 y 3 de la zona del cultivo de fresón en Huelva.

I.2.2.2. MARCOS DE ÁREAS PARA LAS ZONAS DE CULTIVOS PROTEGIDOS EN ALMERÍA

Delimitación y estratificación

En Almería se delimitaron tres zonas: una en Campo de Dalías y Poniente (32.559 ha.), otra en Nijar (62.506 ha.) y otra en Bajo Almanzora (67.783 ha.). En lo que sigue sólo se presenta el marco de áreas construido para la zona de Campo de Dalías y Poniente. Los límites de la zona en estudio y los de los estratos considerados se muestran en la figura 6 (en las figuras 7 y 8 para las zonas de Nijar y Bajo Almanzora, respectivamente). Se consideran cuatro estratos: el estrato 1 es el de las zonas urbanas (1.601 ha.), el estrato 2 es el de "intensidad alta de invernaderos" (21.122 ha.), el estrato 3 es el de "intensidad baja de invernaderos" (3.061 ha.) y, finalmente, el estrato 4 es el de espacios naturales (1.775 ha.).

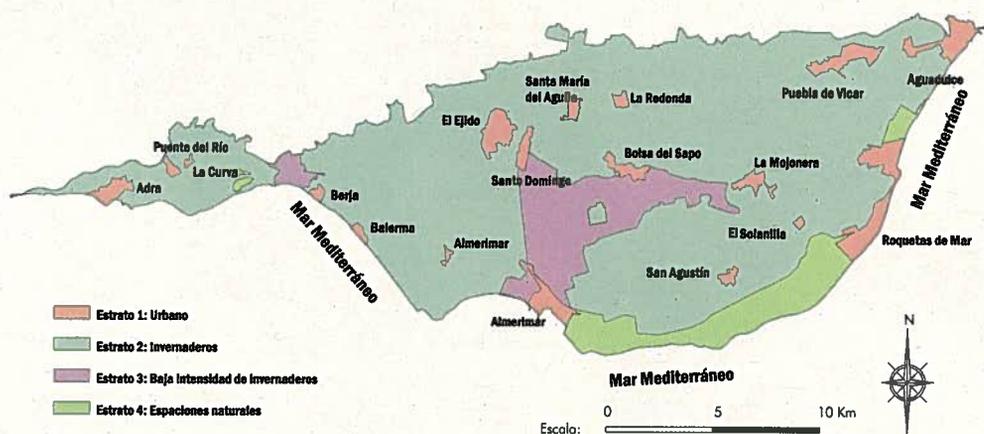


Figura 6. Límites de la zona y de los estratos para la zona de cultivos protegidos en Campo de Dalías (Almería).

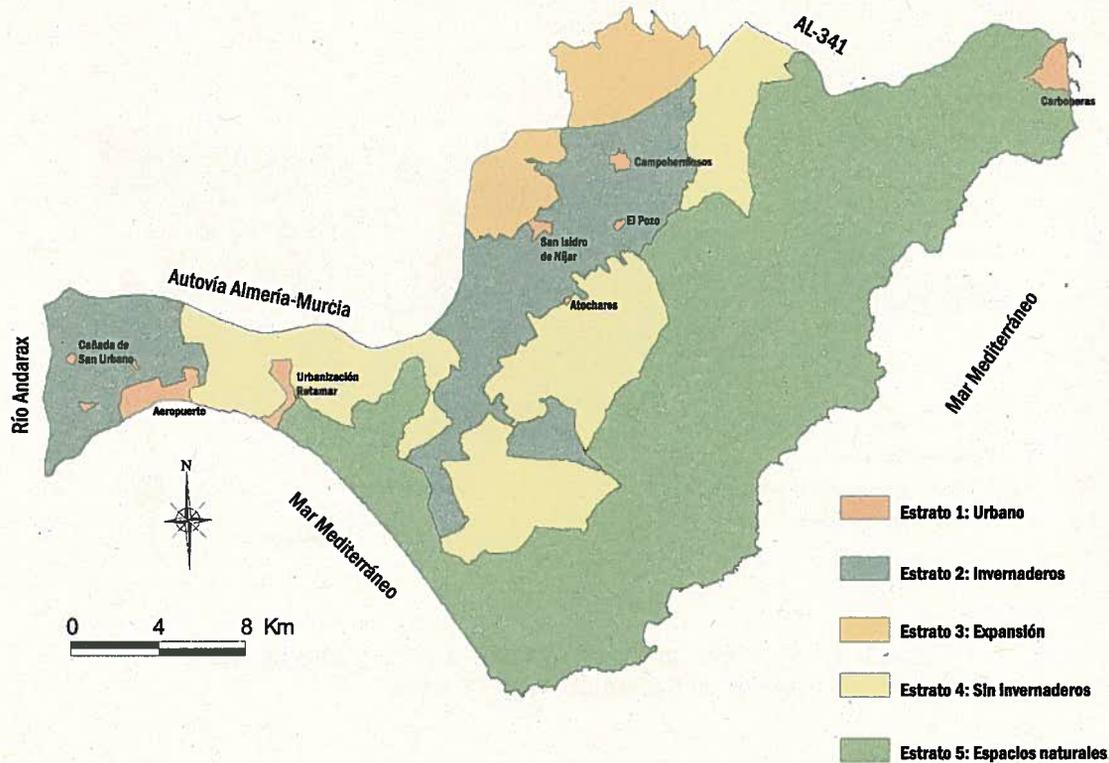


Figura 7. Límites de la zona y de los estratos para la zona de cultivos protegidos en Nijar (Almería).

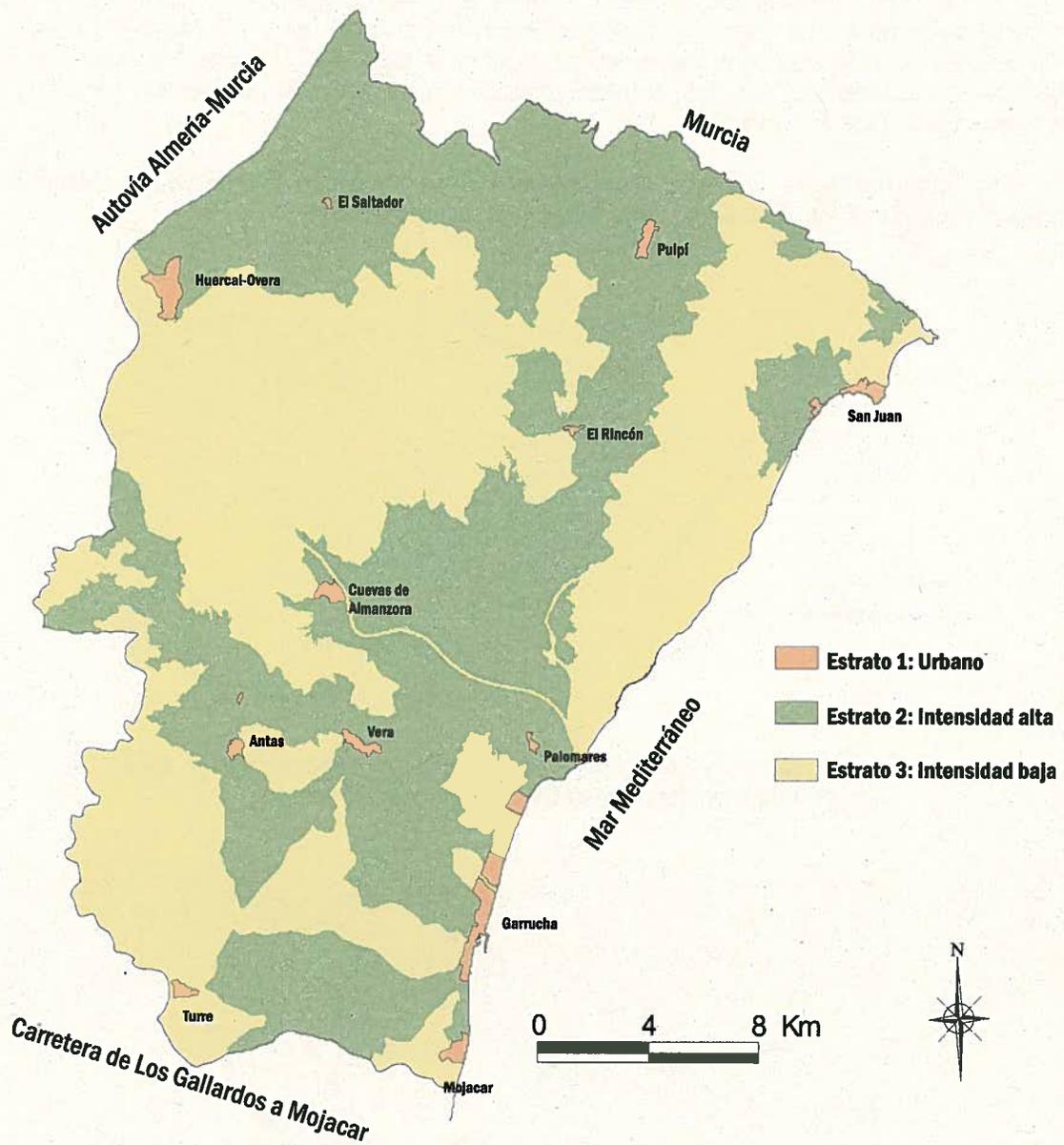


Figura 8. Límites de la zona y de los estratos para la zona de cultivos protegidos en Bajo Almanzora (Almería).

Unidades Primarias

Los estratos 2 y 3 se dividieron en Unidades Primarias (UP). El tamaño idóneo del segmento se fijó en 6 ha. para el estrato 2 y 12 ha. para el estrato 3. El tamaño idóneo de la Unidad Primaria se fijó en 6 segmentos, esto es, 36 ha. en el estrato 2 y 72 ha. en el estrato 3. En el archivo UPDA-LIAS.TXT del disco compacto anejo a este documento se recoge una relación de las 843 Unidades Primarias delimitadas en el Campo de Dalías (en los archivos UPNIJAR.TXT y UPALMANZ.TXT se recogen las Unidades Primarias correspondientes a las zonas de Níjar –187 Unidades Primarias– y Bajo Almanzora –435 Unidades Primarias– respectivamente); con indicación de su superficie y del código asignado a cada Unidad Primaria.

La figura 9 muestra los límites entre Unidades Primarias, dentro de estratos, en Campo de Dalías y las figuras 10 y 11 en Níjar y Bajo Almanzora, respectivamente.

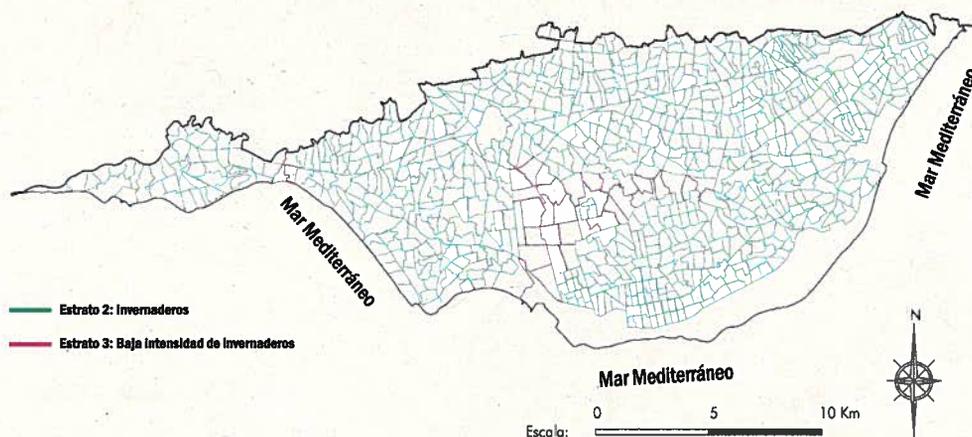


Figura 9. Límites entre Unidades Primarias, dentro de estratos, para la zona de cultivos protegidos en Campo de Dalías (Almería).

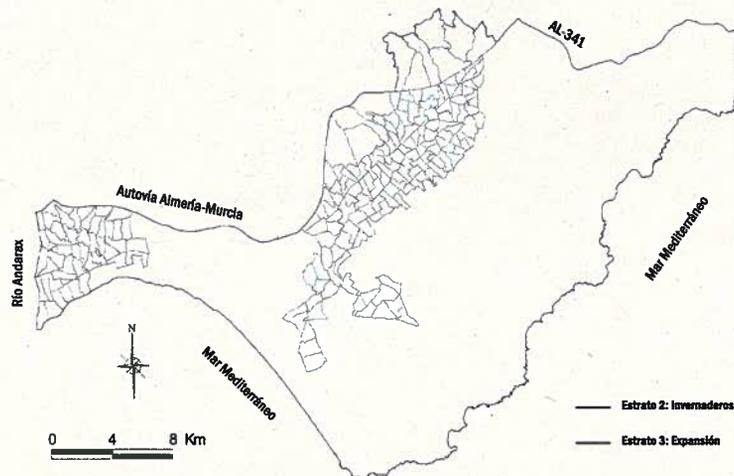


Figura 10. Límites entre Unidades Primarias, dentro de estratos, para la zona de cultivos protegidos en Níjar (Almería).

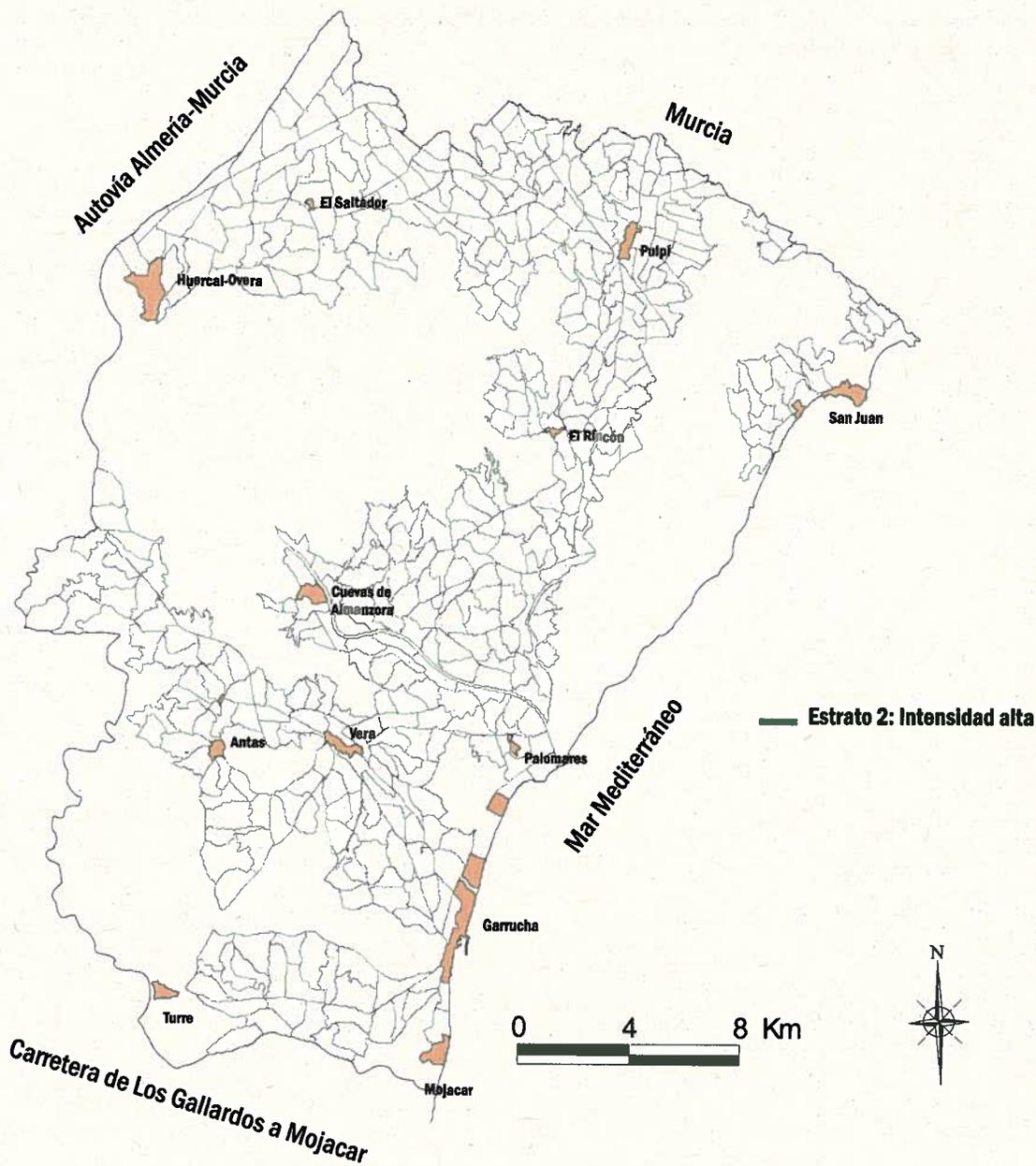


Figura 11. Límites entre Unidades Primarias, dentro de estratos, para la zona de cultivos protegidos en Bajo Alanzora (Almería).

La relación de las 843 Unidades Primarias, junto con la información gráfica sobre las mismas, constituye el marco de áreas para la zona de Campo de Dalías. A partir de él se selecciona la muestra de segmentos y la de agricultores. En el disco compacto anejo a este documento se incluye el marco de áreas construido para las zonas de cultivo intensivo consideradas en Almería: Campo de Dalías, Níjar y Bajo Almanzora.

CAPITULO II
SELECCIÓN DE LA MUESTRA

II. SELECCIÓN DE LA MUESTRA

A partir del marco de áreas se selecciona una muestra aleatoria de segmentos y, a partir de ésta, una muestra aleatoria de agricultores. Sobre los segmentos de la muestra se observa parte de la información requerida (la superficie cultivada y demás usos del suelo). Mediante entrevistas a los agricultores seleccionados en la muestra, se recoge la información que no es posible observar directamente sobre el terreno, tal como la relativa a las estructuras de producción y a las características técnico-económicas de la producción agrícola. El procedimiento de selección de la muestra es uno de los dos medios de los que dispone el diseñador para el control de las fuentes de variación, con vistas a mejorar la precisión de las estimaciones. El otro es el procedimiento de estimación. En este capítulo se expone el procedimiento de selección, en el capítulo IV se estudia el de estimación.

II.1. SELECCIÓN DE LA MUESTRA DE SEGMENTOS

La variabilidad espacial, esto es, la diferencia entre los valores de una variable en dos puntos del espacio, aumenta generalmente con la distancia entre esos dos puntos. Para el control de esta variabilidad se consideran dos procedimientos de selección de la muestra de segmentos: muestreo por bloques y muestre sistemático.

II.1.1. MARCOS EN RETÍCULA. SEGMENTOS DE LÍMITES GEOMÉTRICOS

Se considera en primer lugar el marco de áreas con segmentos de límites geométricos, descrito bajo el epígrafe I.1. Este tipo de marco consiste en dividir el territorio en estudio en cuadrículas o segmentós, cuya posición en el espacio se identifica por la fila y columna a la que pertenece el segmento o por sus coordenadas con respecto a un sistema de referencia tal como la proyección UTM (Universal Transversal Mercator). Es especialmente simple por lo que su utilización es frecuente. Este es también el tipo de disposición de la información recogida en fuentes de datos de uso frecuente en la agricultura, como es la teledetección.

La figura 12 es un esquema de este tipo de marcos. Se trata de NM segmentos (cuadrículas) dispuestos en una retícula de N filas y M columnas. Los segmentos pueden considerarse agrupados en bloques cuadrados de $\nu\mu$ segmentos cada uno (ν segmentos por fila y μ segmentos por columna). Sea n el número de bloques por fila y m el número de bloques por columna. Se verifica: $N=n\nu$, $M=m\mu$ y $NM=nm\nu\mu$.

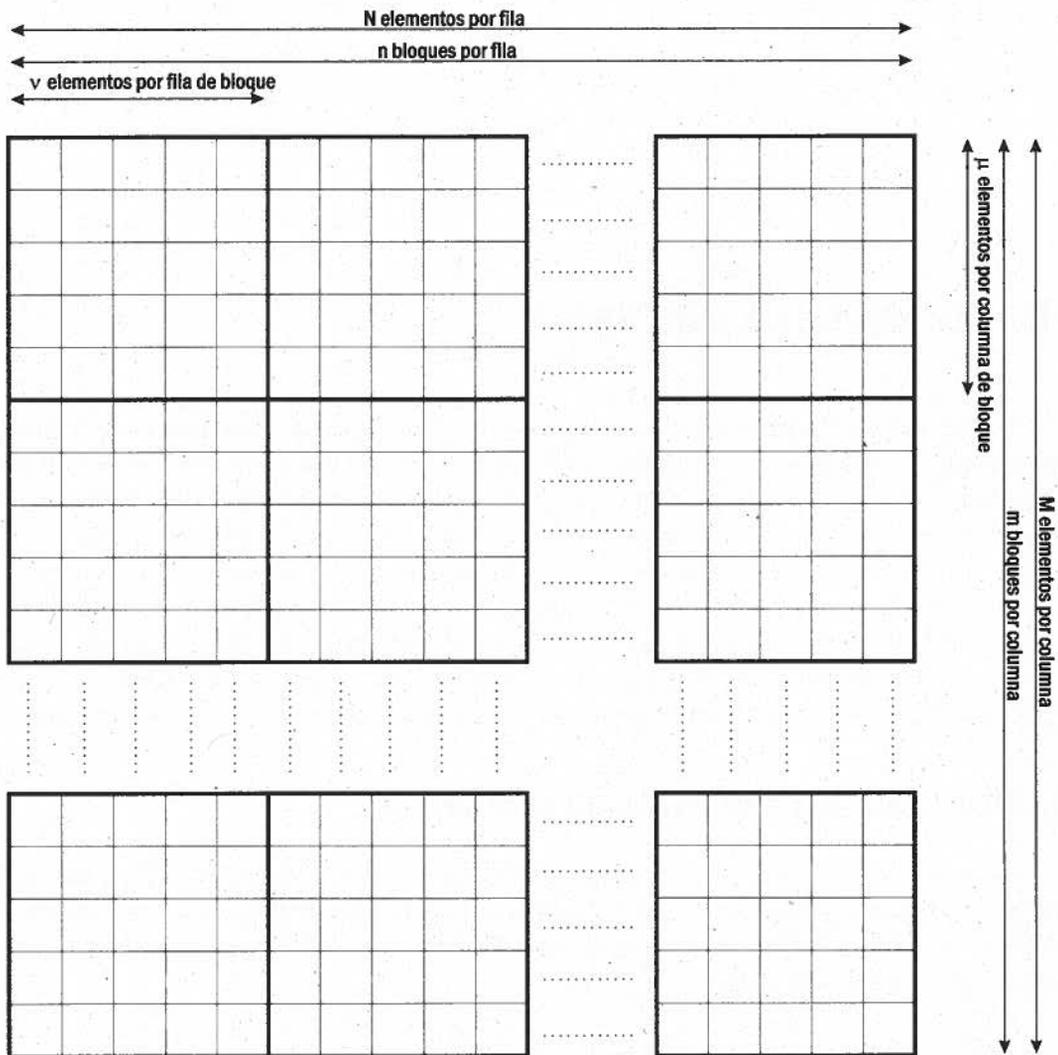


Figura 12. Esquematización de un marco en retículas.

El número de procedimientos de selección de una muestra en poblaciones como la considerada, o en un espacio bidimensional cualquiera, es grande, puesto que se puede emplear muestreo aleatorio simple, estratificado o sistemático en cada dirección. Quenouille (1949) las clasifica en tres categorías, según qué los puntos de muestreo estén alineados en ambas direcciones, en sólo una de ellas o no alineados en ambas direcciones [ver también Bellhouse (1977)]. Una descripción de veintuna de estas estrategias puede verse en Koop (1990). Se consideran aquí dos esquemas de muestreo o procedimientos de selección de la muestra de una población en retículas, especialmente simples de aplicar –muestreo por bloques y muestreo sistemático– y se compara su eficiencia relativa respecto del muestreo aleatorio simple, en el epígrafe IV.1.1.1.

II.1.1.1. MUESTREO ALEATORIO SIMPLE

Este esquema de muestreo consiste en elegir, con igual probabilidad y sin reposición, una muestra de tamaño η de entre los NM segmentos de la población.

La probabilidad de selección de uno cualquiera de los segmentos es:

$$\pi_{ij} = \frac{1}{NM} \quad \forall i = 1, 2, \dots, N; \quad \forall j = 1, 2, \dots, M$$

y la probabilidad de su inclusión en la muestra es:

$$P_{ij} = \frac{\eta}{NM} \quad \forall i = 1, 2, \dots, N; \quad \forall j = 1, 2, \dots, M$$

II.1.1.2. MUESTREO POR BLOQUES

Dentro de cada bloque (hay nm bloques, cada uno con $v\mu$ segmentos) se elige, independientemente, una muestra aleatoria simple del mismo tamaño η_{kl}

La probabilidad de selección de uno cualquiera de los segmentos es:

$$\pi_{klgh} = \frac{1}{v\mu} \quad ; \quad \forall l = 1, 2, \dots, n; \quad \forall k = 1, 2, \dots, m \quad \forall g = 1, 2, \dots, v; \quad \forall h = 1, 2, \dots, \mu$$

y la probabilidad de su inclusión en la muestra es:

$$P_{klgh} = \frac{\eta_{kl}}{v\mu} \quad \forall k = 1, 2, \dots, n; \quad \forall l = 1, 2, \dots, m; \quad \forall g = 1, 2, \dots, v; \quad \forall h = 1, 2, \dots, \mu$$

Si se denota por $\eta = nm\eta_{kl}$ el tamaño de la muestra total, entonces, por sustitución, resulta:

$$P_{klgh} = \frac{\eta}{NM}$$

de modo que es igual a la del muestreo aleatorio simple, con muestras de tamaño η .

II.1.1.3. MUESTREO SISTEMÁTICO

La selección de una muestra sistemática con η_t puntos de arranque aleatorio, comprende dos pasos: (a) de entre los $v\mu$ segmentos que contiene un bloque cualquiera se seleccionan una muestra aleatoria simple de η_t segmentos, a cuyas posiciones se les denomina puntos de arranque, (b) en cada uno de los $(nm-1)$ bloques restantes, se seleccionan aquellos segmentos que se encuentran en la misma posición relativa que los puntos de arranque. El tamaño de la muestra es $\eta = \eta_t nm$.

Este esquema es equivalente a considerar que los $NM = v\mu nm$ segmentos de la población están agrupados en $v\mu$ conglomerados de nm segmentos cada uno, de tal forma que los segmentos de cada uno de estos conglomerados se encuentran regularmente espaciados una distancia v en filas y μ en columnas (a cada uno de los $v\mu$ conglomerados se les denomina 'muestra sistemática/conglo-

merado'). De los $v\mu$ conglomerados se selecciona una muestra aleatoria simple de η_t conglomerados.

La localización de los η_t puntos de arranque es aleatoria pero, una vez elegida esa localización, los puntos de muestreo se disponen regularmente dentro de muestras sistemáticas/conglomerados a una distancia v en la dirección de las filas y μ en la dirección de las columnas.

La probabilidad de selección de uno cualquiera de los segmentos es igual a la probabilidad de selección de la muestra sistemática/conglomerado a la que pertenece, esto es:

$$\pi_{ij} = \frac{1}{v\mu} \quad \forall i = 1, 2, \dots, N; \forall j = 1, 2, \dots, M$$

y la probabilidad de inclusión de un segmento en la muestra es:

$$P_{ij} = \frac{\eta_t}{v\mu} = \frac{\eta_t nm}{v\mu nm} = \frac{\eta}{NM} \quad \forall i = 1, 2, \dots, N; \forall j = 1, 2, \dots, M$$

igual, por tanto, a la probabilidad de inclusión de los esquemas de muestreo aleatorio simple y aleatorio por bloques, para muestras del mismo tamaño η .

II.1.1.4. ESTUDIO DE UN CASO: LA MUESTRA DE SEGMENTOS DE LA 'ENCUESTA DE SUPERFICIES' DEL MAPA (1990)

La muestra se selecciona a partir del marco de áreas descrito en el epígrafe I.1. Se trata de una muestra sistemática con tres puntos de arranque aleatorio. El procedimiento de selección consiste en:

- (I) Seleccionar una muestra aleatoria simple de tres de los cien segmentos de 1 Km. por 1 Km. de que consta un bloque de 10 Km. por 10 Km. y en,
- (II) Incluir en la muestra los tres segmentos de cada bloque que ocupan la misma posición, dentro del bloque, que las seleccionadas en (I).

Los tres segmentos seleccionados en (I), para la muestra del MAPA, son los que ocupan las posiciones 2-6 (abscisa 2 y ordenada 6 respecto del vértice Suroeste del bloque), 3-2 y 6-7. La muestra de segmentos está compuesta por el conjunto de unidades de muestreo que ocupan dichas posiciones dentro de cada uno de los bloques que integran el marco de áreas. La figura 13 muestra la disposición de los segmentos de la muestra del M.A.P.A. dentro de un bloque de 10 Km. por 10 Km. y la figura 14 la disposición de los segmentos de la muestra en Andalucía.

Este procedimiento de selección puede ser interpretado como un muestreo sistemático con tres puntos de arranque aleatorio o, lo que es equivalente, como un muestreo aleatorio simple de tres conglomerados elegidos de entre los 100 de que consta la población. Para aclarar esta última interpretación, baste considerar al conjunto de segmentos de 1 Km. por 1 Km. de la población agrupadas en 100 conglomerados, de manera que los segmentos que ocupan la misma posición dentro de los bloques definen un conglomerado. La tasa de muestreo es el 3% sobre el número de unidades de muestreo de la población objetivo, lo que representa poco menos del 1,5% de la superficie de dicha población objetivo, debido a que el segmento es aproximadamente la mitad (700 m. por 700 m.) de la cuadrícula de 1 Km. por 1 Km.

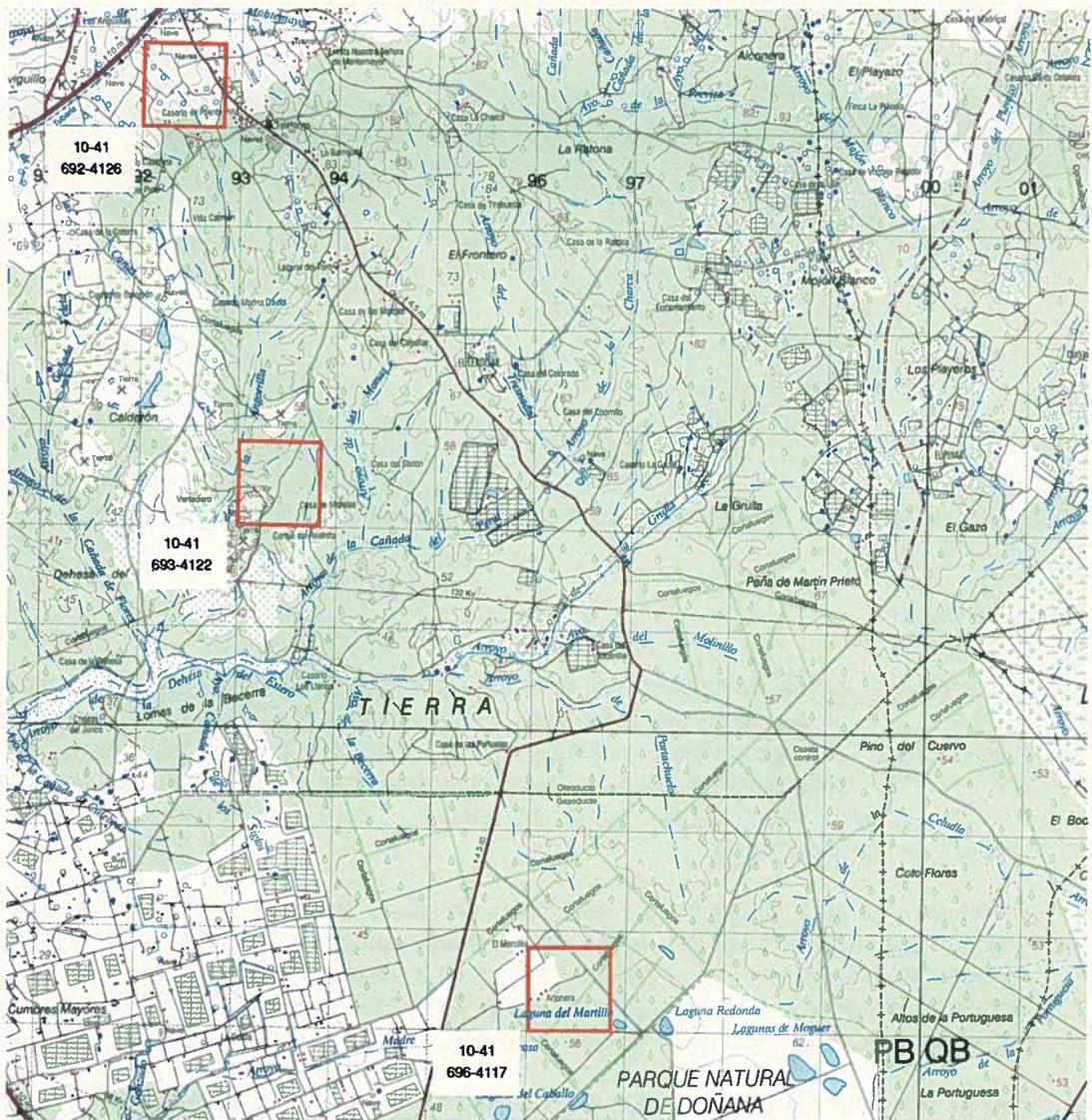


Figura 13. Posición de los segmentos de la muestra de superficies del M.A.P.A. (1990).

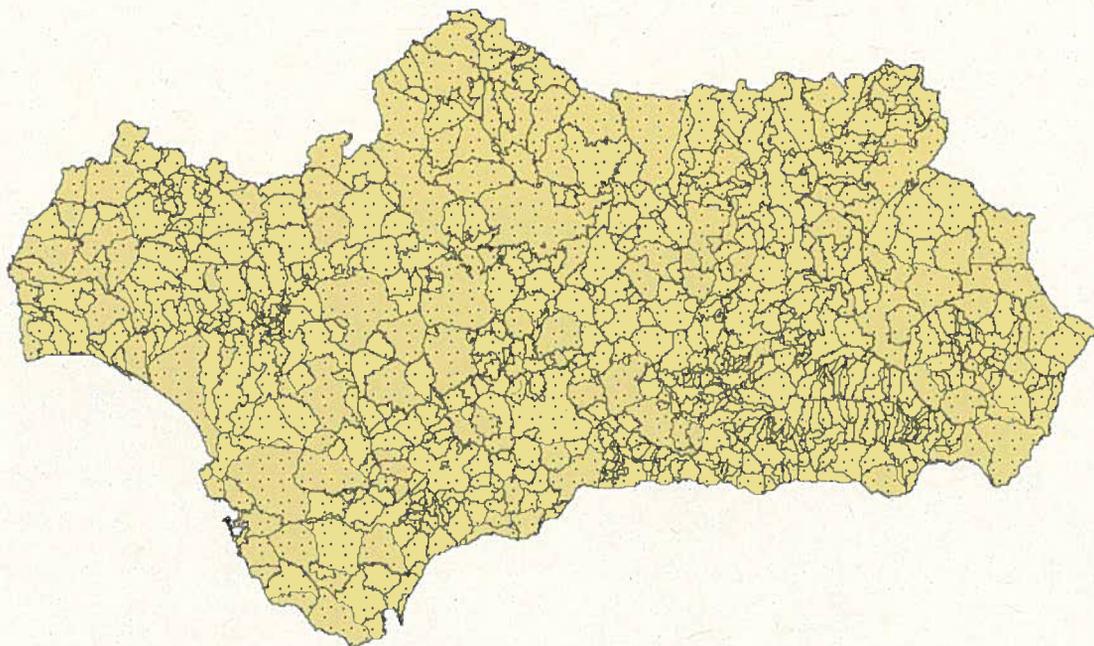


Figura 14. Disposición de los segmentos de la muestra de superficies del M.A.P.A. en Andalucía.

II.1.2. RÉPLICAS

Si el tamaño de la muestra a seleccionar es de n segmentos, se pueden seleccionar r réplicas de m segmentos en lugar de una sola muestra de $n = r \cdot m$ segmentos. Por ejemplo, una muestra de 60 segmentos puede ser seleccionada de una sola vez ($r = 1$ y $m = 60$), ó como dos muestras independientes ($r = 2$) de treinta elementos cada una ($m = 30$), ó bien como tres muestras independientes ($r = 3$) de veinte elementos cada una ($m = 20$), etc. La selección de una muestra por réplicas es recomendable por las razones siguientes:

- (1) Facilita la rotación de la muestra. La rotación es necesaria cuando el muestreo se repite a lo largo del tiempo, para evitar los errores de muestreo asociados a "la fatiga del encuestado". Por ejemplo, si se desea renovar cada año el 25% de la muestra, se deben seleccionar cuatro (o un múltiplo de cuatro) réplicas. De esta forma, en la muestra de cada año hay elementos antiguos y nuevos, lo que permite la estimación del cambio.
- (2) Permite la realización de test sobre nuevos procedimientos de muestreo ó nuevos cuestionarios. Por ejemplo la mitad de las réplicas se tratan con el procedimiento antiguo y la otra mitad con el nuevo. De la comparación entre resultados se pueden extraer conclusiones.
- (3) Se puede llevar un control de la calidad de los datos comparando los totales de réplicas entre sí. Una desviación notable del total de una de las réplicas respecto de las restantes denotaría la posible existencia de errores: los datos deberían ser chequeados.
- (4) Facilita la gestión de la muestra. En particular, el tamaño de muestra puede ser incrementado ó disminuido en ciertos estratos o períodos de tiempo aumentando o disminuyendo el número de réplicas.

- (5) Permite un estimador insesgado de la varianza, si bien es muy inestable debido a que está basado en el número de réplicas y éste será siempre reducido. Es posible obtener estimaciones de la varianza más estables.
- (6) Permite evaluar el efecto de la rotación, esto es, el impacto sobre los datos que resulta del hecho de que un mismo segmento permanezca en la muestra durante los años correspondientes al período de rotación.

En el esquema de muestreo sistemático descrito en II.1.1.3, cada muestra sistemática/ conglomerado es una réplica. En el esquema de muestreo por bloques descrito en II.1.1.2, una réplica se forma seleccionando un sólo segmento de cada bloque. Esta operación se repite, de forma independiente, tantas veces como réplicas se requieran.

II.1.3. MARCOS ESTRATIFICADOS. SEGMENTOS DE LÍMITES FÍSICOS

El muestreo aleatorio por bloques y el muestreo sistemático son procedimientos de selección de la muestra que resultan eficientes para el control de la variabilidad espacial, también en el caso de poblaciones en las que los segmentos de la población se encuentran en cualquier otra disposición distinta a la disposición en retículas de filas y columnas.

En los epígrafes I.2.2.1. y I.2.2.2. se presentan dos marcos de áreas (para sendas zonas de Huelva y Almería), en los que la disposición de los segmentos no se ajusta al esquema de filas y columnas, estudiado en el epígrafe II.1.1. Para la selección de muestras por muestreo aleatorio de bloques en este tipo de marcos, basta con dividir cada estrato en zonas o bloques de tamaño similar. Para la selección de una muestra sistemática basta con numerar los segmentos de cada estrato, de forma correlativa: la muestra se selecciona tal y como se describe en el siguiente epígrafe. Nótese que mientras el muestreo por bloques requiere una nueva partición del territorio en bloques, el muestreo sistemático puede llevarse a cabo sin necesidad de esa nueva partición. Ambos procedimientos son igualmente eficientes para la estimación de variables espaciales (ver epígrafe IV.1.1.1).

II.1.3.1. MUESTREO SISTEMÁTICO DE RÉPLICAS

El procedimiento de selección de una muestra sistemática de réplicas se ilustra con un ejemplo. En un estrato con 900 segmentos, se desea seleccionar una réplica de 15 segmentos. Se elige un número aleatorio entre 1 y 60 ($900/15 = 60$). Sea X el número elegido; los segmentos de la muestra son los correspondientes a los números:

$$X, X+60, X+2 \times 60, X+3 \times 60, \dots, X+14 \times 60$$

Este procedimiento se repite para todas y cada una de las réplicas deseadas, pero sin repetición del punto de arranque X . En el epígrafe II.1.3.4.1. se aplica el procedimiento a la selección de una muestra sistemática de réplicas a partir del marco de áreas construido para la zona del fresón en Huelva, presentado en el I.2.2.1.

II.1.3.2. MUESTREO DE RÉPLICAS POR BLOQUES

Si se desea seleccionar, de un estrato determinado, una muestra de r réplicas, de m segmentos cada una, se requiere dividir el estrato en m zonas o bloques. De cada bloque se selecciona un segmento (con probabilidades iguales o desiguales y con o sin reposición): la primera réplica está

constituida por los segmentos así seleccionados. Una segunda réplica se selecciona extrayendo de cada zona un nuevo segmento y así sucesivamente hasta completar las r réplicas. En el epígrafe II.1.3.4.2. se aplica el procedimiento a la selección de una muestra por bloques de réplicas, a partir del marco de áreas construido para la zona de cultivos protegidos de Almería, presentado en I.2.2.2.

II.1.3.3. SELECCIÓN DE LA MUESTRA DE SEGMENTOS

Los marcos de áreas presentados en I.2.2.1. y I.2.2.2. son relaciones de Unidades Primarias. Los procedimientos de muestreo descritos en II.1.3.1. y II.1.3.2. sirven sólo para seleccionar las Unidades Primarias que contienen segmentos de la muestra. La identificación del segmento de la muestra dentro de la Unidad Primaria requiere la división previa de la Unidad Primaria en el número de segmentos que se le asignan. Para esta división se deben seguir los mismos criterios que se utilizaron para delimitar las Unidades Primarias.

En ocasiones no existirán límites físicos suficientes para delimitar todos los segmentos asignados a la UP. En ese caso se delimitarán sólo aquellos, del tamaño adecuado al estrato, para los que existan límites físicos. El resto de la UP constituye uno o más bloques de segmentos a cada uno de los cuales se les asignará su correspondiente número de segmentos.

El segmento de la muestra se selecciona, de entre los así delimitados, de la siguiente forma: (I) se elige al azar y con igual probabilidad un número entre uno y el número de segmentos asignados a la UP, (II) si el número elegido es el asociado a uno de los segmentos individuales ese es el segmento de la muestra, (III) si el número elegido está asociado a un segmento de un bloque proindiviso, será necesario ir al terreno y dividir allí el bloque en el número de segmentos asignado al bloque, uno de los cuales será elegido al azar como segmento de la muestra, (IV) si es imposible o impracticable dividir el bloque en el número de segmentos correspondientes, se usará el bloque entero como segmento de la muestra y se dividirán las observaciones muestrales realizadas sobre el mismo por el número de segmentos asignado al bloque.

II.1.3.4. ESTUDIO DE CASOS

Consideramos dos casos de estudio, (I) la selección de una muestra sistemática de réplicas de segmentos en la zona del cultivo del fresón en Huelva, a partir del marco de áreas descrito en I.2.2.1. y (II) la selección de una muestra de réplicas de segmentos por muestreo de bloques en Campo de Dalías (Almería), a partir del marco de áreas descrito en I.2.2.2.

II.1.3.4.1. Selección de una muestra sistemática de réplicas de segmentos en la zona de cultivo de fresón en Huelva

Se considera una tasa de muestreo del 3% de la superficie en el estrato 2 (subestrato de fresón, 2-f, y resto del estrato 2, 2-r) y del 1.5% en el estrato 3.

Con estas tasas, el tamaño de la muestra en cada estrato es como sigue:

	Superficie a muestrear (ha)	Número de segmentos
Estrato 2-f	752	50
Estrato 2-r	1.238	50
Estrato 3	270	5

Los segmentos de la muestra se seleccionan en cinco réplicas sistemáticas de diez segmentos en los estratos 2-f y 2-r. En el estrato 3 la muestra es aleatoria simple.

En el siguiente cuadro se ilustra el procedimiento de selección de una muestra sistemática de réplicas, con probabilidades proporcionales al tamaño de las Unidades Primarias. En primer lugar se asigna un número de segmentos a cada Unidad Primaria: si A_{hi} es el área de la i -ésima Unidad Primaria del estrato h -ésimo y a_h es el tamaño idóneo del segmento en ese estrato, el número de segmentos a asignar a esa Unidad Primaria es el entero más próximo a A_{hi}/a_h . A continuación, se calculan los valores acumulados del número de segmentos asignados a cada Unidad Primaria. Por ejemplo, en el estrato 2-f, el número total de segmentos asignados al estrato es 1.756. En este estrato, la muestra de 50 segmentos se selecciona en cinco réplicas de diez segmentos cada una. El intervalo de muestreo para cada réplica es $1.756/10 \cong 175$.

Se eligen, sin reposición y con probabilidades iguales, 5 números aleatorios entre 1 y 175. Los números elegidos son 145, 66, 44, 168 y 126. Cada uno de ellos es el punto de arranque de una réplica. Por ejemplo, los segmentos de la tercera réplica son: 44, 44+175, 44+2x175, 44+3x175, 44+4x175, 44+5x175, 44+6x175, 44+7x175, 44+8x175 y 44+9x175.

Estrato	Réplica	Número Seleccionado	Número de orden del segmento de la muestra
2-f	1	145	145, 320, 495, 670, 845, 1020, 1195, 1370, 1545 y 1750
	2	66	66, 241, 416, 591, 766, 941, 1116, 1291, 1466 y 1641
	3	44	44, 219, 394, 569, 744, 919, 1094, 1269, 1444 y 1619
	4	168	168, 343, 518, 693, 868, 1043, 1218, 1393, 1568 y 1743
	5	126	126, 301, 476, 651, 826, 1001, 1176, 1351, 1526 y 1701
2-r	1	135	135, 308, 481, 654, 827, 1000, 1173, 1346, 1519 y 1626
	2	69	69, 242, 415, 588, 761, 934, 1107, 1280, 1453 y 1626
	3	162	162, 335, 508, 681, 854, 1027, 1200, 1373, 1546 y 1719
	4	43	43, 216, 389, 562, 735, 908, 1081, 1254, 1427 y 1600
	5	8	8, 181, 354, 527, 700, 873, 1046, 1219, 1392 y 1565
3	U		67, 143, 219, 295 y 371

Los números 44, 219, 394, 569, 744, 919, 1094, 1269, 1444 y 1619, permiten identificar las 10 Unidades Primarias de la réplica 3 en el estrato 2-f: la Unidad Primaria para la que el valor acumulado del número de segmentos es el menor de entre los que superan o igualan a 44 es la primera a incluir en la réplica 3, la segunda es la Unidad Primaria para la que dicho valor acumulado es el menor de entre las que superan o igualan a 219 y así sucesivamente.

En la figura 15 se muestran los límites de las Unidades Primarias así seleccionadas. Cada una de estas Unidades Primarias se divide en el número de segmentos asignado a cada una de ellas. Una vez delimitados, los segmentos se numeran dentro de cada Unidad Primaria y se elige uno de ellos, con probabilidades iguales.

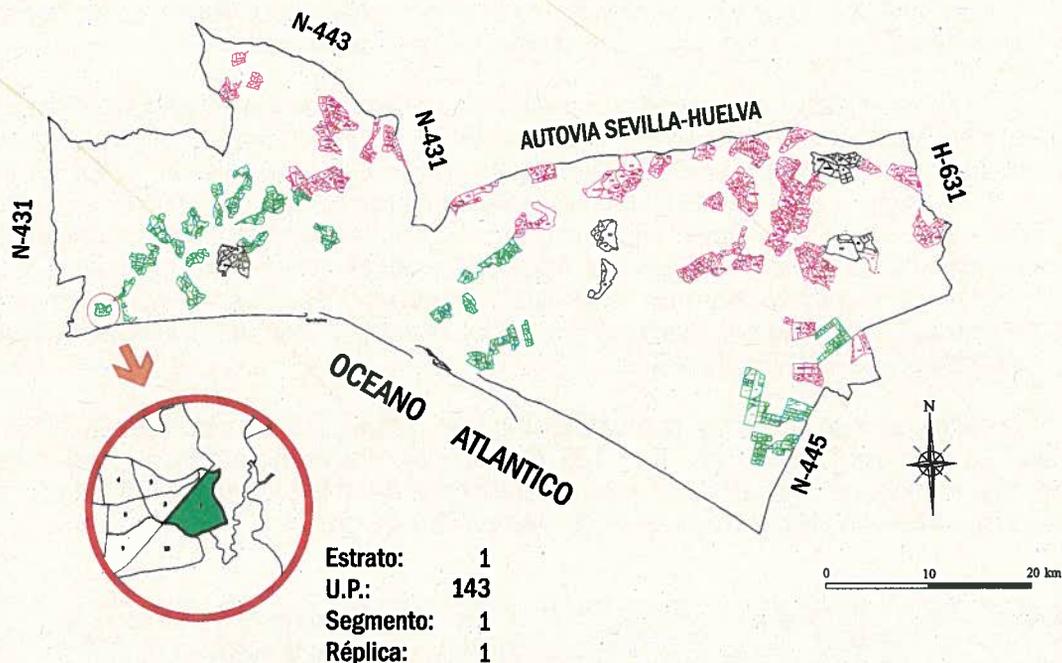


Figura 15. Segmentos delimitados dentro de las Unidades Primarias seleccionadas en la muestra para la zona de cultivo de fresón en Huelva.

II.1.3.4.2. Selección de una muestra de réplicas de segmentos por muestreo de bloques en Campo de Dalías

Se trata de seleccionar una muestra de 50 segmentos (203,32 ha. en la muestra) en el estrato 2 y 2 segmentos en el estrato 3 (19,37 ha. en la muestra). En el estrato 2, el procedimiento de muestreo que se considera es el de muestreo por bloques: cinco réplicas de diez segmentos cada una. En el estrato 2 la muestra es aleatoria simple.

Para la selección de la muestra de réplicas en el estrato 2, éste se ha dividido en 10 bloques de aproximadamente el mismo tamaño. Dentro de cada uno de estos bloques se ha seleccionado una Unidad Primaria; las 10 Unidades Primarias así seleccionadas constituyen una réplica. Las Unidades Primarias difieren significativamente de tamaño, por lo que se decidió extraer la muestra con probabilidades proporcionales al tamaño. Las cuatro réplicas restantes se seleccionan repitiendo el procedimiento cuatro veces más. Así se tiene una muestra de Unidades Primarias. La muestra de segmentos se tiene dividiendo cada Unidad Primaria de la muestra en el número de segmentos que previamente se le asignó y eligiendo al azar y con probabilidades iguales un segmento dentro de cada una de las Unidades Primarias segmentadas.

El procedimiento de selección de la muestra de Unidades Primarias consiste en acumular la superficie de las Unidades Primarias de cada bloque, separadamente. Para la primera réplica se elige, en cada bloque, con igual probabilidad, un número entre cero y la superficie total acumulada en el bloque. Las Unidades Primarias de la muestra en cada bloque son aquellas para las que el valor acumulado es el menor entre los que superan o igualan al número elegido.

Las Unidades Primarias seleccionadas por este procedimiento fueron las siguientes:

Zona	Est.	Bloque	N ^{os} Seleccionados	U.P.
Campo de Dalias y Poniente	2	A	3,4,18,23 y 48	3, 4,18, 34 y 43
		B	4, 11, 18, 42 y 45	71, 121, 128, 163 y 178
		C	9, 12, 49, 58, y 66	75, 77, 143, 223 y 236
		D	7, 17, 36, 43 y 60	150, 176, 213, 257 y 261
		E	19, 27, 42, 51 y 60	348, 449, 462, 469 y 558
		F	1, 22, 51, 57 y 90	747, 759, 767, 802 y 831
		G	16, 29, 56, 61 y 79	325, 441, 540, 602 y 608
		H	8, 19, 34, 52 y 68	93, 699, 723, 727 y 843
		I	32, 37, 56, 72 y 75	427, 518, 526, 633 y 644
		J	34, 42, 52, 61 y 79	299, 387, 395, 411 y 423
	3		12 y 32	283 y 375

Las cinco réplicas del estrato 2 son las constituidas por las Unidades Primarias con los siguientes números de orden:

Réplica	Unidades Primarias
1	43, 128, 213, 236, 423, 558, 602, 633, 737 y 831
2	3, 77, 163, 261, 325, 361, 387, 526, 699 y 747
3	18, 71, 75, 150, 395, 427, 441, 469, 802 y 843
4	4, 143, 178, 257, 411, 462, 608, 644, 693 y, 759
5	34, 121, 176, 223, 299, 348, 518, 540, 723 y 767

Cada una de estas Unidades Primarias se ha segmentado en el número de segmentos previamente asignados. El segmento de la muestra es uno de los delimitados en cada Unidad Primaria, elegido con probabilidades iguales.

II.2. LA MUESTRA DE AGRICULTORES

La muestra de agricultores se selecciona a partir de la muestra de segmentos. Cada segmento de la muestra se divide en parcelas, sobre una ampliación de la fotografía aérea que contiene al segmento. Se determina la superficie de cada parcela y se selecciona una muestra de ellas, con probabilidades proporcionales a la superficie. La muestra de agricultores está constituida por los que cultivan las parcelas seleccionadas en la muestra. En general, el agricultor dispone de los datos requeridos (por ejemplo los relativos a la estructura y a los aspectos técnico-económicos de la explotación) a nivel de toda la explotación, no de la parcela de la muestra. Se plantea entonces el problema de que las unidades de observación (explotaciones agrícolas) y las unidades de selección o muestreo (segmentos y parcelas) difieren: en un mismo segmento cultivan tierras más de un agricultor (se denomina "tramo" a la superficie del segmento cultivada por un mismo agricultor) y un mismo agricultor cultiva tierras en más de un segmento. Se denomina "red" al conjunto de unidades de observación (explotaciones) asociadas a una unidad de muestreo (segmento). La muestra de agricultores es el conjunto de unidades de observación (explotaciones) asociadas al conjunto de segmentos de la muestra o una submuestra de ella: se trata, pues de una muestra de redes. Para la estimación se requieren introducir criterios de asociación entre la población de explotaciones y la de segmentos: estos criterios se presentan en el epígrafe IV.2.

II.2.1. ESTUDIO DE CASOS

El procedimiento de selección de la muestra de agricultores a partir de la muestra de seg-

mentos puede comportar más de una etapa y, en cada etapa, las probabilidades de selección pueden ser iguales o desiguales, por lo que la gama de procedimientos posibles es muy amplia. Nos limitaremos a considerar los procedimientos utilizados en dos casos de estudio: (I) el de la zona de cultivo de fresón en Huelva y (II) el de los cultivos bajo plástico en Campo de Dalías (Almería).

II.2.1.1. SELECCIÓN DE UNA MUESTRA DE AGRICULTORES A PARTIR DE LA MUESTRA DE SEGMENTOS EN LA ZONA DE CULTIVO DE FRESÓN EN HUELVA

En este caso se seleccionó una muestra de agricultores a partir de la muestra de segmentos de los estratos 2-f, 2-r y 3.

La muestra se seleccionó en tres etapas:

- 1ª) La muestra de primera etapa es la muestra de segmentos seleccionada bajo el epígrafe II.1.3.4.1.
- 2ª) En la segunda etapa se selecciona, con igual probabilidad y dentro de cada réplica, una muestra de segmentos en las réplicas del estrato 2. En cada uno de los segmentos de la muestra de esta segunda etapa, se deslindan las parcelas y se identifican los agricultores que cultivan las parcelas deslindadas en cada uno de los segmentos.
- 3ª) En la tercera etapa se selecciona, en cada uno de los segmentos de la muestra de segunda etapa, una muestra de agricultores con probabilidades proporcionales a la superficie de fresón que cultivan dentro del segmento.

II.2.1.2. SELECCIÓN DE UNA MUESTRA DE AGRICULTORES A PARTIR DE LA MUESTRA DE SEGMENTOS EN CAMPO DE DALÍAS

La muestra se selecciona en dos etapas:

- 1ª) La muestra de primera etapa es la muestra de segmentos seleccionada bajo el epígrafe II.1.3.4.2.
- 2ª) En la segunda etapa se deslindan las parcelas y se identifica a los agricultores que las cultivan, en cada uno de los segmentos de la muestra de primera etapa. La muestra de agricultores se selecciona con probabilidades proporcionales a la superficie que cultivan dentro del segmento.

II.3. ROTACIÓN DE LA MUESTRA

Cuando se prevé repetir una encuesta periódicamente, es necesario considerar en el momento del diseño un plan de rotación de la muestra, de modo que en cada período se encuentren en la muestra, junto a elementos seleccionados en períodos anteriores, otros nuevos incluidos en el período en cuestión. La rotación de la muestra reduce el sesgo debido a la "fatiga del encuestado" y permite una mejor estimación de los cambios y de las tendencias de las características en estudio.

La rotación se lleva a cabo haciendo salir de la muestra, en un período determinado, una o más réplicas e incluyendo otras tantas nuevas. Un ejemplo puede verse en Cotter y Nealon (1987). Estos autores consideran un período de rotación quinquenal. Para ese período, es conveniente que el número de réplicas en cada estrato sea un múltiplo de cinco. Si en un estrato determinado el número de

réplicas de la muestra es cinco, un posible plan de rotación de la muestra consistiría en excluir de la muestra, al cabo de cada año, una réplica y en incluir una nueva, de modo que el número de años que una misma réplica permanece en la muestra es de cinco y cada cinco años la muestra se renueva por completo (tasa de rotación del 20%). Las réplicas que salen de la muestra después del primero y del segundo año de la vida del marco, son repescadas seis años después. Sin embargo, las que salen tras el tercero y cuarto año no vuelven a entrar. Las que entran durante los cuatro últimos años de la vida del marco, están todas menos de cinco años.

CAPITULO III
RECOLECCIÓN DE DATOS

III. RECOLECCIÓN DE DATOS

El soporte de la información a recoger en campo es la fotografía aérea y los formularios, en lo relativo a los datos directamente observables sobre el terreno, como los usos del suelo. Para la recogida de información mediante entrevista se requieren cuestionarios.

III.1. RECOLECCIÓN DE DATOS PARA LA ESTIMACIÓN DE SUPERFICIES CULTIVADAS Y DEMÁS USOS DEL SUELO

A) MATERIAL GRÁFICO: FOTOGRAFÍAS Y MAPAS

Los límites de los segmentos de la muestra se imprimen en pliegos transparentes a escala 1:5.000, con tinta del color correspondiente al del estrato al que pertenecen (ver figura 16).

Cada pliego se fija, por uno de sus extremos, sobre una cartulina en la que se pega una copia de la fotografía aérea del segmento, ampliada a escala 1:5.000. En la cartulina figura impreso el número de la hoja del Mapa Topográfico Nacional 1:50.000 y el de la hoja del Mapa Topográfico de Andalucía 1:10.000, en las que se encuentra el segmento. También contiene una representación reducida de este último, con el segmento seleccionado marcado en el mismo color que se ha utilizado para delimitar el estrato, junto con los límites de los segmentos de la Unidad Primaria a la que pertenece el de la muestra, así como los límites de las Unidades Primarias del entorno. Esta información es útil para la localización del segmento sobre el terreno.

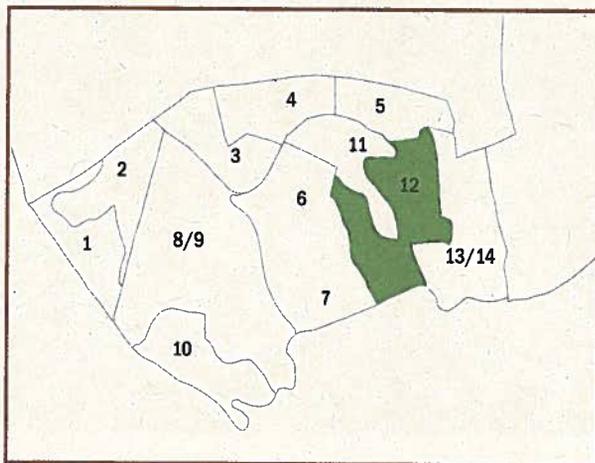
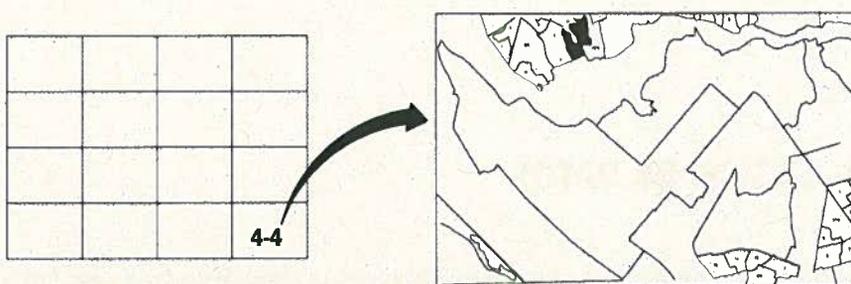


Figura 16. Límites del segmento de la muestra 1-269-12.

**MUESTREO DE AREAS PARA LA ESTIMACION DE SUPERFICIES
DE FRESON EN HUELVA**



HOJA M.T.N.: 0999

ESTRATO: 1

U.P.: 269

SEGMENTO: 12



Figura 17. Soporte gráfico de la información a recoger en campo del segmento de la muestra 1-269-12.

B) IDENTIFICACIÓN DE LOS LÍMITES DEL SEGMENTO

Una vez situado sobre el segmento, el encuestador empieza por asegurarse de que los límites del segmento son identificables al 100 %. Los segmentos con límites no identificables serán devueltos al supervisor.

C) DESLINDE DE PARCELAS

Sobre el pliego transparente se deslindan las parcelas, rotulando los límites entre parcelas. Se entiende por parcela o "unidad de uso del suelo" una superficie continua ocupada por el mismo cultivo o uso del suelo, aunque esté integrada por varias parcelas catastrales pertenecientes a distintos propietarios. Cualquier cambio de cultivo o uso del suelo será considerado como límite de parcela.

Si los límites del segmento son accidentes naturales o artificiales del terreno, todas las parcelas delimitadas dentro de un segmento serán enteras, no habrá parcelas partidas. Si los límites son geométricos las parcelas quedan partidas por los límites. En todo caso hay que asegurarse de que todos los recintos o parcelas quedan cerrados.

Puede haberse producido cambios en los límites de parcela que se observan en la fotografía, respecto de los realmente observados sobre el terreno. Es necesario asegurarse de que los límites entre parcelas trazados sobre el pliego transparente son los actuales. Es conveniente el disponer de los límites de parcelas en soporte informático: en la Figura 18 se muestran dichos límites representados a partir de la digitalización de los trazados sobre el pliego transparente.

D) NUMERACIÓN DE PARCELAS

Las parcelas se numeran de forma correlativa, siguiendo el mismo criterio ya utilizado para numerar las Unidades Primarias. Por ejemplo, la numeración de las parcelas puede realizarse, empezando por la situada más al noroeste y siguiendo de forma serpenteante continua, de oeste a este y posteriormente de este a oeste, hasta completar todo el segmento.

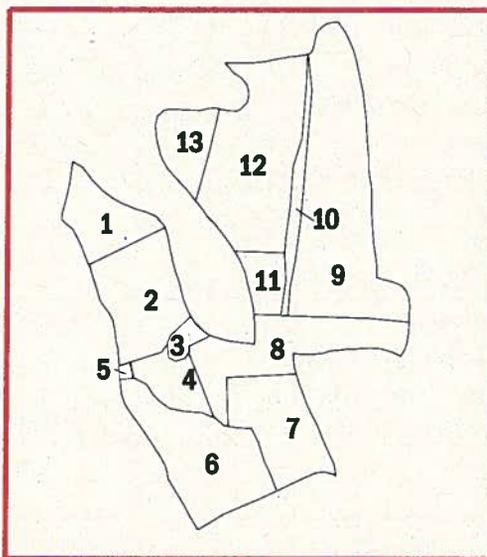


Figura 18. Límites de parcelas en el segmento de la muestra 1-269-12.

E) IDENTIFICACIÓN DE CULTIVOS O USOS DEL SUELO

El cultivo o uso del suelo en cada una de las parcelas deslindadas en el segmento se identifica y registra en el formulario, asignándole el código correspondiente, según figure en la codificación de cultivos y usos del suelo que se proporciona al encuestador

F) EL FORMULARIO

El formulario para la recogida de datos sobre superficies consta en su cabecera de un espacio para anotar el código completo del segmento, el nombre del encuestador y la fecha de realización de la encuesta (figura 19).

DEPARTAMENTO DE ECONOMÍA
Y CIENCIAS SOCIALES AGRARIAS.
UNIVERSIDAD POLITÉCNICA DE MADRID.

SERVICIO DE ESTUDIOS Y
ESTADÍSTICAS.
CONSEJERÍA DE AGRICULTURA Y PESCA.
JUNTA DE ANDALUCÍA.

MUESTREO DE AREAS PARA LA ESTIMACIÓN DE SUPERFICIES DE FRESÓN EN HUELVA.

Código del encuestador : Provincia : HUELVA
Nombre : ROGELIO JALÓN SÁNCHEZ Comarca :
Nº de mapa :
Segmento -Estrato : 1
-U.P. : 269
-Segmento : 12
Fecha de la encuesta : Escala : 1 : 5.000

Parcela	Cultivo o uso del suelo	Código	S/R	Observaciones
1	NO AGRÍCOLA	NA		EXTRACCIÓN ARIDOS.
2	FRESÓN MICROTÚNEL	FN	R	
3	NO AGRÍCOLA	NA		DEPÓSITO AGUA.
4	FRESÓN MICROTÚNEL	FN	R	
5	NO AGRÍCOLA	NA		VIVIENDA.
6	FRESÓN MICROTÚNEL	FN	R	
7	FRESÓN MICROTÚNEL	FN	R	
8	ERIAL	ER		
9	FRESÓN MACROTÚNEL	FM	R	
10	AGUA SULCE	AG		ARROYO Y DESAGÜE
11	FRESÓN MICROTÚNEL	FN	R	
12	FRESÓN MACROTÚNEL	FM	R	
13	HORTALIZAS MACROTÚNEL			PIMIENTOS, BERENGENAS MELONES.

Figura 19. Formulario de recogida de datos del segmento de la muestra 1-269-12.

El resto del formulario lo constituye un cuadro de varias filas y cinco columnas. Los datos se registran por filas. En cada fila se registra la información relativa al cultivo o uso del suelo en una parcela. En la casilla correspondiente a la primera columna se registra el número asignado a la parcela, el cual ha sido anotado en el pliego transparente, en la segunda el nombre del cultivo o uso del suelo (o los nombres si hay más de uno), en la tercera el código de cultivo, en la cuarta se anota una "S" si se trata de un cultivo de secano y una "R" si se trata de un cultivo de regadío.

Finalmente, en la casilla correspondiente a la quinta columna se registran las observaciones que se consideren convenientes.

III.2. CONTROL DE CALIDAD DE LOS DATOS RECOLECTADOS

Se realiza un control, sobre el terreno, del trabajo de campo y otro en gabinete del material que sirve de soporte a la información y de la propia información.

A) CONTROL DE CAMPO

Consiste en una inspección por muestreo del trabajo de campo realizado por cada uno de los encuestadores. Del conjunto de segmentos investigados por cada encuestador se selecciona una muestra. El deslinde de parcelas en cada uno de los segmentos de esa muestra y la asignación de usos a cada una de las parcelas, es inspeccionado por un supervisor.

B) CONTROL EN GABINETE

En gabinete se procede a un control, primero del material que sirve de soporte a la información recogida en campo (los pliegos transparentes y los formularios), y después de la propia información, previa traslación de ésta a soporte informático.

El control del material se realiza manualmente y consiste en verificar la exacta correspondencia entre el código de identificación del pliego transparente y el formulario, así como la coincidencia de ambos con el código del segmento al que corresponden.

Un primer control de la información gráfica y alfanumérica, también realizado de forma manual, consiste en verificar (I) que el número de parcelas delimitadas sobre el pliego transparente coincide con el del formulario, (II) que a todas las parcelas se les ha asignado un número y que éste es único tanto en el pliego transparente como en el formulario, (III) que no existen recintos abiertos de entre los delimitados en el segmento y (IV) que en los formularios, los códigos de cultivo coinciden con la denominación del cultivo.

Los errores detectados que es posible corregir en gabinete se corrigen, previa consulta a los encuestadores. El material cuya información errónea no ha sido posible corregir en gabinete será devuelto a los encuestadores para su revisión.

Un segundo y definitivo control de la información gráfica y alfanumérica será realizado por procedimientos informáticos.

III.3. RECOLECCIÓN DE DATOS MEDIANTE ENTREVISTA A LOS AGRICULTORES

Los datos relativos a (I) aspectos estructurales de las explotaciones, (II) características técnico-económicas de los cultivos y (III) aspectos relativos a la producción agrícola y su relación con el medio ambiente, se obtienen por medio de encuestas a los agricultores. Los cuestionarios utilizados en Huelva y Almería se recogen en el disco compacto anejo a este documento.

CAPITULO IV
ESTIMACIÓN

IV ESTIMACIÓN

Las características directamente observadas sobre el terreno, tales como los usos del suelo, se estiman a partir de los datos recogidos en la muestra de segmentos. Las características estructurales y/o técnico-económicas de las explotaciones han de ser estimadas a partir de los datos recogidos mediante entrevistas a los agricultores.

IV.1. ESTIMACIONES A PARTIR DE LA MUESTRA DE SEGMENTOS

Se considera, en primer lugar, el caso de poblaciones no estratificadas, con segmentos dispuestos en retícula. Para ellas, se estudia la eficiencia relativa entre los procedimientos de selección de la muestra considerados en II.1.1. y se evalúa para un caso de estudio. En segundo lugar, se trata el caso de poblaciones estratificadas con segmentos de límites físicos.

IV.1.1 MARCOS EN RETÍCULA NO ESTRATIFICADOS

El marco de áreas para este tipo de poblaciones se presenta en el epígrafe I.1. y se esquematiza en la figura 12 Asociado al segmento ij (fila i , columna j) de la población se considera un valor Y_{ij} . Interesa estimar la media de esos valores:

$$\bar{Y} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M Y_{ij}$$

o el total $Y = NM\bar{Y}$ a partir de una muestra de tamaño η . Se consideran los tres procedimientos de selección de la muestra (η) de segmentos, muestreo aleatorio simple, muestreo por bloques y muestreo sistemático, descritos en el epígrafe II.1.1.

El estimador.

Como estimador de \bar{Y} se considera la media muestral, esto es, la media de los valores de y_{ij} observados en los η segmentos de la muestra:

$$\hat{Y} = \bar{y} = \frac{1}{\eta} \sum_{r=1}^{\eta} y_r$$

El estimador del total es $\hat{Y} = NM\hat{Y}$. Si $\hat{V}\hat{Y}$ denota la varianza del estimador de la media, la del

total, $V(\hat{Y})$ es:

$$V(\hat{Y}) = (NM)^2 V(\hat{\bar{Y}})$$

Los coeficientes de variación de los estimadores de la media y del total coinciden:

$$CV(\hat{\bar{Y}}) = CV(\hat{Y}) = \frac{\sqrt{V(\hat{\bar{Y}})}}{\hat{\bar{Y}}}$$

El estimador de la media (y el del total) es insesgado, con los tres procedimientos de selección considerados. La diferencia entre procedimientos estriba en la varianza del estimador \hat{Y} esto es, en la precisión de las estimaciones.

Varianza del estimador.

Con muestreo aleatorio simple, la varianza del estimador de la media $V_r(\hat{Y})$ es [Hansen et al. (1953), Kish (1965), Cochran (1977), Jessen (1978)]:

$$V_r(\hat{Y}) = (1-f) \frac{S^2}{\eta} \quad [1]$$

donde,

$$S^2 = \frac{1}{NM-1} \sum_{i=1}^N \sum_{j=1}^M (Y_{ij} - \bar{Y})^2$$

y $f = \frac{\eta}{NM}$ es la fracción de muestreo.

La varianza del estimador del total, $V_r(\hat{Y})$ es:

$$V_r(\hat{Y}) = (NM)^2 V_r(\hat{\bar{Y}})$$

Un estimador insesgado de $V_r(\hat{Y})$ es :

$$\hat{V}_r(\hat{Y}) = (1-f) \frac{s^2}{\eta}$$

donde s^2 es un estimador insesgado de S^2 :

$$s^2 = \frac{1}{\eta-1} \sum_{r=1}^{\eta} (y_r - \bar{y})^2$$

y un estimador insesgado de $V_r(\hat{Y})$, es:

$$\hat{V}_r(\hat{Y}) = (NM)^2 \hat{V}_r(\hat{\bar{Y}})$$

Con muestreo por bloques la varianza del estimador de la media, $V_b(\hat{Y})$, es:

$$V_b(\hat{Y}) = \sum_{k=1}^n \sum_{l=1}^m W_{kl}^2 V(\hat{Y}_{kl}) \quad [2]$$

donde:

$$W_{kl} = \frac{1}{mn}$$

$$V(\hat{Y}_{kl}) = (1 - f_{kl}) \frac{S_{kl}^2}{\eta_{kl}}$$

$$S_{kl}^2 = \frac{1}{v\mu - 1} \sum_{g=1}^v \sum_{h=1}^{\mu} (Y_{klgh} - \bar{Y}_{kl})^2$$

$$\bar{Y}_{kl} = \frac{1}{v\mu} \sum_{g=1}^v \sum_{h=1}^{\mu} Y_{klgh}$$

$$f_{kl} = \frac{\eta_{kl}}{v\mu}$$

$$\eta_{kl} = \frac{\eta}{nm}$$

Y_{klgh} es el valor de la variable en estudio asociado al segmento en la posición (fila= g) x (columna= h) dentro del bloque kl -ésimo.

La varianza del estimador del total, $V_b(\hat{Y})$, es:

$$V_b(\hat{Y}) = (NM)^2 V_b(\hat{Y})$$

Un estimador insesgado de $V_b(\hat{Y})$ es:

$$\hat{V}_b(\hat{Y}) = \sum_{k=1}^n \sum_{l=1}^m W_{kl}^2 \hat{V}(\hat{Y}_{kl})$$

donde:

$$\hat{V}(\hat{Y}_{kl}) = (1 - f_{kl}) \frac{S_{kl}^2}{\eta_{kl}}$$

$$S_{kl}^2 = \frac{1}{\eta_{kl} - 1} \sum_{s=1}^{\eta_{kl}} (y_{kls} - \bar{y}_{kl})^2$$

con: $\bar{y}_{kl} = \frac{1}{\eta_{kl}} \sum_{s=1}^{\eta_{kl}} y_{kls}$

siendo y_{kls} el valor de la variable en estudio asociado al segmento s -ésimo de la muestra del bloque kl -ésimo.

Un estimador insesgado de $V_b(\hat{Y})$ es:

$$\hat{V}_b(\hat{Y}) = (NM)^2 \hat{V}_b(\hat{Y})$$

Con muestreo sistemático, la varianza del estimador de la media, $V_s(\hat{Y})$, es:

$$V_s(\hat{Y}) = \frac{1}{(nm)^2} (1-f) \frac{S_{gh}^2}{\eta_t} \quad [3]$$

donde S_{gh}^2 es la varianza poblacional entre totales de muestras sistemáticas/conglomerados:

$$S_{gh}^2 = \frac{1}{v\mu - 1} \sum_{g=1}^v \sum_{h=1}^{\mu} (Y_{gh} - \bar{Y}_{\dots})^2$$

donde Y_{gh} es el total dentro de la muestra sistemática/conglomerado gh -ésimo de las $v\mu$ posibles, e \bar{Y}_{\dots} es la media de totales de muestra sistemática/conglomerado:

$$Y_{gh} = \sum_{k=1}^n \sum_{l=1}^m Y_{klgh}$$

$$\bar{Y}_{\dots} = \frac{1}{v\mu} \sum_{g=1}^v \sum_{h=1}^{\mu} Y_{gh}$$

$\frac{\eta_t}{v\mu}$ es la fracción de muestreo igual a la del muestreo aleatorio simple y del muestreo

por bloques ($\eta_{kl} = \frac{\eta}{nm}$, por lo que $f = \frac{\eta}{NM}$).

La varianza del estimador del total, $V_s(\hat{Y})$, es:

$$V_s(\hat{Y}) = (NM)^2 V_s(\hat{Y})$$

Un estimador insesgado de $V_s(\hat{Y})$ es:

$$\hat{V}_s(\hat{Y}) = \frac{1}{(nm)^2} (1-f) \frac{S_{gh}^2}{\eta_t}$$

$$S_{gh}^2 = \frac{1}{\eta_t - 1} \sum_{t=1}^{\eta_t} (y_t - \bar{y})^2$$

donde y_t es el total dentro de la muestra sistemática/conglomerado t -ésimo e \bar{y} es la media muestral de estos totales:

$$y_t = \sum_{k=1}^n \sum_{l=1}^m y_{klt}$$

$$\bar{y} = \frac{1}{\eta_t} \sum_{t=1}^{\eta_t} y_t$$

Un estimador insesgado de $V_s(\hat{Y})$, es:

$$\hat{V}_b(\hat{Y}) = (NM)^2 \hat{V}_b(\hat{Y})$$

IV.1.1.1. EFICIENCIA RELATIVA DE LAS TRES ESTRATEGIAS DE MUESTREO CONSIDERADAS

La Eficiencia Relativa ($ER_{A/B}$) entre dos procedimientos o estrategias de muestreo A y B, es la inversa del cociente entre las varianzas de los estimadores cuando el tamaño de la muestra es el mismo con ambos procedimientos. Esto es:

$$ER_{A/B} = \frac{V_B(\hat{Y})}{V_A(\hat{Y})} \text{ dado que } \eta_A = \eta_B$$

Si $ER_{A/B} > 1$ esto es, $V_A(\hat{Y}) < V_B(\hat{Y})$ la estrategia A es más eficiente que la B, mientras que si $ER_{A/B} < 1$, la estrategia B es más eficiente que la A.

La eficiencia relativa del muestreo por bloques respecto del aleatorio simple es (dividiendo [1] entre [2]):

$$ER_{b/r} = \frac{(1-f) \frac{S^2}{\eta}}{\frac{\sum_{k=1}^n \sum_{l=1}^m S_{kl}^2}{\eta nm}} = \frac{S^2}{\frac{\sum_{k=1}^n \sum_{l=1}^m S_{kl}^2}{nm}}$$

y la del muestreo sistemático respecto del muestreo aleatorio simple (dividiendo [1] entre [3]):

$$ER_{s/r} = \frac{(1-f) \frac{S^2}{\eta}}{\frac{1}{(nm)^2} (1-f) \frac{S_{gh}^2}{\eta_t}} = \frac{nmS^2}{S_{gh}^2}$$

Finalmente, la del muestreo sistemático respecto del muestreo por bloques es (dividiendo [2] entre [3]):

$$ER_{s/b} = \frac{(1-f) \frac{\sum_{k=1}^n \sum_{l=1}^m S_{kl}^2}{\eta nm}}{\frac{1}{(nm)^2} (1-f) \frac{S_{gh}^2}{\eta_t}} = \frac{\sum_{k=1}^n \sum_{l=1}^m S_{kl}^2}{S_{gh}^2}$$

donde $\eta = \eta_t nm$.

A partir de un área de 200 Km. por 200 Km. que abarca la práctica totalidad de Castilla y León y para la que se dispone de información relativa a los usos del suelo en cada uno de los 40.000 segmentos de 1Km. por 1 Km., Iglesias (1998) ha evaluado la eficiencia relativa de las tres estrategias de muestreo consideradas. En el cuadro 1 se presentan, para cada uso del suelo, sus resultados.

CUADRO 1. EFICIENCIA RELATIVA ENTRE LAS DISTINTAS ESTRATEGIAS, EN FUNCIÓN DEL TAMAÑO DEL BLOQUE $\nu \times \mu$ Y DE LA DISTANCIA MÍNIMA ENTRE ELEMENTOS DE UNA MISMA MUESTRA SISTEMÁTICA/CONGLOMERADO, PARA DISTINTOS USOS DEL SUELO.

Tamaño del bloque	2x2	5x5	10x10	20x20	50x50	100x100	200x200
Distancia mínima entre elementos de una misma muestra sistemática	2	5	10	20	50	100	200
Uso del suelo	Cultivos herbáceos en secano						
ER _{b/r}	4,01	2,35	1,73	1,42	1,16	1,03	1,00
ER _{s/r}	3,51	2,07	1,40	1,11	1,44	1,04	1,00
ER _{s/b}	0,88	0,88	0,81	0,78	1,24	1,01	1,00
Uso del suelo	Cultivos herbáceos en regadío						
ER _{b/r}	3,76	2,14	1,59	1,33	1,09	1,01	1,00
ER _{s/r}	10,98	2,13	1,60	1,29	1,07	1,04	1,00
ER _{s/b}	2,92	1,00	1,00	0,97	0,99	1,03	1,00
Uso del suelo	Cultivos leñosos						
ER _{b/r}	2,48	1,70	1,35	1,21	1,05	1,03	1,00
ER _{s/r}	6,70	1,94	1,78	1,32	1,21	1,03	1,00
ER _{s/b}	2,70	1,14	1,32	1,09	1,16	1,00	1,00

Se observa que el muestreo por bloques y el sistemático son más eficientes que el muestreo aleatorio simple. Entre el muestreo por bloques y el muestreo sistemático no se observan grandes diferencias. La eficiencia relativa del muestreo por bloques y la del muestreo sistemático, respecto del muestreo aleatorio simple, disminuye al aumentar el tamaño del bloque.

Con muestreo por bloques, la mayor eficiencia relativa se tiene con bloques de tamaño mínimo (lado al cuadrado igual al cociente entre el tamaño de la población y el de la muestra) y muestras de un sólo segmento por bloque. Con muestreo sistemático, la mayor eficiencia relativa se tiene con una muestra de un sólo punto de arranque aleatorio (un sólo segmento por bloque), y con los segmentos de la muestra equidistantes entre sí una distancia igual a la raíz cuadrada del cociente entre el tamaño de la población y el de la muestra. No obstante estos resultados, las ventajas del muestreo de réplicas apuntadas en II.1.2. hacen recomendable el muestreo de réplicas: más de un punto de arranque alea-

torio (en el muestreo sistemático) o más de un segmento por bloque (en el muestreo por bloques).

IV.1.1.2. ESTUDIO DE UN CASO: LA MUESTRA DEL MAPA (1990) PARA LA ESTIMACIÓN DE LA SUPERFICIE CULTIVADA Y DEMÁS USOS DEL SUELO

La muestra es sistemática con tres puntos de arranque aleatorio. Se consideran bloques de 10 Km. por 10 Km ($v = \mu = 10$ Km.). Se trata, pues, de una muestra aleatoria simple de $n = 3$ muestras sistemáticas/conglomerados, seleccionadas de entre las ($v \times \mu = 100$ posibles (ver epígrafe II.1.1.4).

A partir de la información recogida en la muestra de segmentos (cuadrados de 700 m. por 700 m.) se estima la superficie ocupada por cada uno de los cultivos y demás usos del suelo considerados y el coeficiente de variación.

El estimador de la superficie cultivada

El estimador \hat{Y} de la superficie total Y ocupada por un cultivo determinado es:

$$\hat{Y} = \frac{100}{49} \frac{N}{n} \sum_{i=1}^n Y_i$$

donde:

$$Y_i = \sum_{j=1}^{M_i} Y_{ij}$$

en esta expresión:

Y_{ij} denota la superficie del cultivo en estudio, observada en el segmento j ($j = 1, 2, \dots, M_i$) de la muestra sistemática o conglomerado i ($i = 1, 2, \dots, n$).

Y_i denota la superficie del cultivo en estudio, observada en el conjunto de segmentos de que consta la muestra sistemática o conglomerado i ($i = 1, 2, \dots, n$).

N denota el tamaño de la población: el número de muestras sistemáticas o conglomerados de la población ($N = 100$).

n denota el tamaño de la muestra: el número de muestras sistemáticas o conglomerados de la muestra ($n = 3$).

$100/49$ es el factor de expansión a la población objetivo de las estimaciones obtenidas para la población muestreada (solamente se observan 49 hectáreas de las 100 de que consta una cuadrícula de 1 Km. por 1 Km.).

El procedimiento de selección determina que los conglomerados o muestras sistemáticas sean, con muy ligeras diferencias, del mismo tamaño; de manera que para todo "i" la superficie total B_i del conglomerado "i" es $B_i = B_0$ y el número de segmentos $M_i = M_0$. En estas condiciones, un estimador de Y equivalente a \hat{Y} es el siguiente:

$$\hat{Y} = D p$$

donde:

$$D = \sum_{i=1}^N B_i$$

$$p = \frac{1}{n} \sum_{i=1}^N p_i$$

$$p_i = \frac{1}{M_i} \sum_{j=1}^{M_i} p_{ij}$$

$$p_{ij} = \frac{Y_{ij}}{B_{ij}}$$

donde:

D denota la superficie total del territorio en estudio.

B_{ij} denota el tamaño del segmento "j" del conglomerado "i" ($B_{ij} = 49$ hectáreas).

Varianza del estimador.

El estimador es insesgado y su varianza, $V(\hat{Y})$, es:

$$V_s(\hat{Y}) = \left(\frac{100}{49}\right)^2 N^2 (1 - f) \frac{S_{s/c}^2}{n} \quad [4]$$

donde:

$$S_{s/c}^2 = \frac{1}{(N - 1)} \sum_{i=1}^N (Y_i - \bar{Y})^2$$

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$$

$$f = \frac{n}{N}$$

Nótese que entre [3] y [4] no existe otra diferencia que el factor $(100/49)^2$ que afecta a esta última (salvando el cambio de notación utilizado en [4], para simplificar notaciones: $S_{s/c}^2$ en lugar de S_{gh}^2 e Y_i en lugar de Y_{gh}).

El coeficiente de variación $CV(\hat{Y})$ del estimador \hat{Y} es:

$$CV(\hat{Y}) = \frac{\sqrt{V_s(\hat{Y})}}{\hat{Y}}$$

El estimador de la varianza del estimador.

Se consideran dos estimadores de la varianza del estimador $-V_s(\hat{Y})$: uno insesgado pero inestable, el otro sesgado pero estable.

a) Estimador Insesgado

Un estimador insesgado de $V_s(\hat{Y})$ es:

$$\hat{V}_s(\hat{Y}) = \left(\frac{100}{49}\right)^2 N^2 (1-f) \frac{S_{s/c}^2}{n} \quad [5]$$

donde:

$$S_{s/c}^2 = \frac{1}{(n-1)} \sum_{i=1}^n (Y_i - \hat{Y})^2$$
$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

Sin embargo, $\hat{V}_s(\hat{Y})$ es un estimador muy inestable por estar basado en una muestra de muy pequeño tamaño: el coeficiente de variación de $\hat{V}_s(\hat{Y})$ es del orden de $1/\sqrt{n}$ en nuestro caso n es pequeño ($n=3$).

b) Estimador Sesgado

Un estimador sesgado de $V_s(\hat{Y})$ es:

$$\hat{V}_s(\hat{Y}) = \left(\frac{100}{49}\right)^2 M^2 \left(1 - \frac{m}{M}\right) \frac{s^2}{m} \quad [6]$$

donde:

$$M = \sum_{i=1}^N M_i$$
$$m = \sum_{i=1}^n M_i$$
$$s^2 = \frac{1}{(m-1)} \sum_{i=1}^n \sum_{j=1}^{M_i} (Y_{ij} - \bar{Y})^2$$
$$\bar{Y} = \frac{1}{m} \sum_{i=1}^n \sum_{j=1}^{M_i} Y_{ij}$$

donde:

M denota el número total de segmentos en la población.

m denota el número total de segmentos en la muestra.

El estimador $\hat{V}_r(\hat{Y})$ se define considerando que la muestra es aleatoria simple, en lugar de sistemática.

El sesgo de $\hat{V}_r(\hat{Y})$ es:

$$B(\hat{V}_r(\hat{Y})) = \left(\frac{100}{49}\right)^2 M^2 \left(1 - \frac{m}{M}\right) B(s^2/m)$$

donde:

$B(s^2/m)$ es el sesgo de s^2/m , que es aproximadamente:

$$B(s^2/m) = \left(\frac{1}{n}\right) \frac{S_{s/c}^2}{\bar{M}^2} (ER_{s/r} - 1)$$

donde:

$\bar{M} = M/N$, es el tamaño medio de un conglomerado o muestra sistemática.

$S_{s/c}^2$ se define en [4].

$ER_{s/r} = S^2 / (S_{s/c}^2 / \bar{M})$ es la "eficiencia relativa" del muestreo sistemático respecto del aleatorio simple, en la que:

$$S^2 = \frac{1}{(M-1)} \sum_{i=1}^N \sum_{j=1}^{M_i} (Y_{ij} - \bar{Y})^2$$

$$\bar{Y} = \frac{1}{M} \sum_{i=1}^N \sum_{j=1}^{M_i} Y_{ij}$$

El sesgo es de orden $1/n$ y, en nuestro caso n es pequeño ($n=3$), de manera que no es despreciable. Por otra parte, según nuestras estimaciones [(Ambrosio et al. (1993), Iglesias (1998)], $ER_{s/r}$ es mayor que 1 (ver cuadro 1) por lo que $\hat{V}_r(\hat{Y})$ sobrestima a $V_s(\hat{Y})$.

$\hat{V}_r(\hat{Y})$ es, sin embargo, mucho más estable que $\hat{V}_s(\hat{Y})$: el coeficiente de variación $\hat{V}_r(\hat{Y})$ es del orden de $1/\sqrt{n\bar{M}}$ frente al de $\hat{V}_s(\hat{Y})$ que es del orden de $1/\sqrt{n}$.

El coeficiente de variación del estimador \hat{Y} se estima mediante la expresión:

$$\hat{CV}(\hat{Y}) = \frac{\sqrt{\hat{V}_{s(r)}(\hat{Y})}}{\hat{Y}}$$

en la que $\hat{V}_{s(r)}(\hat{Y})$ es la definida en [5] (o en [6]).

En el cuadro 2 se recogen las estimaciones de la superficie ocupada por los principales cultivos en las zonas estudiadas de Huelva y de Campo de Dalías, respectivamente. Estas estimaciones se han obtenido a partir de la "Encuesta de superficies" del M.A.P.A. , de 1996. En el caso de Huelva y Almería, se recogen las dos estimaciones del coeficiente de variación (en %) correspondientes a los dos estimadores de la varianza considerados, los definidos en [5] y [6].

Si bien la precisión de las estimaciones basadas en la "Encuesta de superficies" del M.A.P.A. es aceptable en grandes demarcaciones territoriales, como el total de CA de Andalucía, (el coeficiente de variación oscila entre el 1,6 % para la estimación de la superficie de cereales y el 15,4 % para los cítricos [ver M.A.P.A.(1999)]), en pequeñas zonas como, las de cultivos intensivos en Huelva y Campo de Dalías (Almería), las precisiones no alcanzan el nivel requerido.

IV.1.2. MARCOS ESTRATIFICADOS

El total Y de la variable en estudio se expresa como suma de los totales $-Y_h-$ de cada estrato. Si denotamos por L el número de estratos, se tiene:

$$Y = \sum_{h=1}^L Y_h$$

IV.1.2.1. ESTIMACIÓN DEL TOTAL Y DE LA MEDIA

Se consideran los dos procedimientos de selección de la muestra descritos en el epígrafe II.1.3: muestreo de réplicas sistemático y por bloques.

IV.1.2.1.1. Muestreo sistemático

Si el tamaño de las réplicas en el estrato h -ésimo se fija en m_h segmentos y el número de segmentos en el estrato es N_h , entonces el número de réplicas posibles es:

$$M_h = \frac{N_h}{m_h}$$

Se denotará por r_h el número de réplicas de la muestra del estrato h -ésimo, de m_h segmentos cada una, de modo que el tamaño de la muestra de segmentos en el estrato es:

$$n_h = r_h m_h$$

El estimador.

El estimador del total de la variable en estudio en la población se estima por:

$$\hat{Y} = \sum_{h=1}^L \hat{Y}_h \quad [7]$$

donde \hat{Y}_h es el estimador del total $-Y_h-$ de la variable en estudio en el estrato h -ésimo:

$$\hat{Y}_h = M_h \frac{1}{r_h} \sum_{j=1}^{r_h} Y_{hj} = \frac{N_h}{n_h} \sum_{j=1}^{r_h} \sum_{i=1}^{m_h} y_{hij}$$

donde y_{hij} es el valor de la variable en estudio en el segmento i -ésimo de la muestra del estrato h -ésimo perteneciente a la réplica j -ésima: (suma de los valores de dicha variable en los T_{hij} tramos del segmento, esto es, en el conjunto de parcelas del segmento):

$$y_{hij} = \sum_{t=1}^{T_{hij}} y_{hijt}$$

La media de la variable en estudio se estima por:

$$\hat{\bar{Y}} = \frac{\hat{Y}}{N}$$

donde:

$$N = \sum_{h=1}^L N_h$$

Varianza.

La varianza del estimador del total es:

$$V_s(\hat{Y}) = \sum_{h=1}^L V_s(\hat{Y}_h) \quad [8]$$

donde:

$$V_s(\hat{Y}_h) = M_h^2 \left(1 - \frac{r_h}{M_h} \right) \frac{1}{r_h} \frac{1}{(M_h - 1)} \sum_{j=1}^{M_h} (Y_{hj} - \bar{Y}_h)^2$$

$$\bar{Y}_h = \frac{1}{M_h} \sum_{j=1}^{M_h} Y_{hj}$$

$$Y_{hj} = \sum_{i=1}^{m_h} y_{hij}$$

Y se estima por:

$$\hat{V}_s(\hat{Y}) = \sum_{h=1}^L \hat{V}_s(\hat{Y}_h)$$

donde,

$$\hat{V}_s(\hat{Y}_h) = M_h^2 \left(1 - \frac{r_h}{M_h} \right) \frac{1}{r_h} \frac{1}{r_h - 1} \sum_{j=1}^{r_h} (Y_{hj} - \bar{Y}_h)^2$$

donde,

$$\bar{Y}_h = \frac{1}{r_h} \sum_{j=1}^{r_h} Y_{hj}$$

La varianza del estimador de la media es:

$$V_s(\hat{Y}) = \frac{1}{N^2} V_s(\hat{Y})$$

y se estima por:

$$\hat{V}_s(\hat{Y}) = \frac{1}{N^2} \hat{V}_s(\hat{Y})$$

Los coeficientes de variación del total y de la media coinciden:

$$CV_s(\hat{Y}) = CV_s(\hat{Y}) = \frac{\sqrt{V_s(\hat{Y})}}{\bar{Y}} = \frac{\sqrt{\hat{V}_s(\hat{Y})}}{\hat{Y}}$$

y se estima por:

$$\hat{C}V_s(\hat{Y}) = \hat{C}V_s(\hat{Y}) = \frac{\sqrt{\hat{V}_s(\hat{Y})}}{\hat{Y}} = \frac{\sqrt{\hat{V}_s(\hat{Y})}}{\hat{Y}}$$

IV.1.2.1.2. Muestreo por bloques

El total en la población se estima por:

$$\hat{Y} = \sum_{h=1}^L \hat{Y}_h \quad [9]$$

donde \hat{Y}_h es el estimador del total en el estrato h-ésimo:

$$\hat{Y}_h = \sum_{j=1}^{m_h} \hat{Y}_{hj}$$

donde m_h es el número de bloques del estrato h-ésimo e \hat{Y}_{hj} es el estimador del total de la variable en estudio en el bloque j-ésimo del estrato h-ésimo:

$$\hat{Y}_{hj} = \frac{M_{hj}}{r_h} \sum_{i=1}^{r_h} y_{hij}$$

donde M_{hj} es el número de segmentos del bloque j-ésimo, del estrato h-ésimo y r_h el tamaño de la muestra de segmentos en el bloque. y_{hij} es el total de la variable en estudio en el segmento i-ésimo de la muestra del bloque j-ésimo del estrato h-ésimo (suma de los valores de dicha variable en los T_{hij} tramos del segmento, esto es, el conjunto de parcelas del segmento):

$$y_{hij} = \sum_{t=1}^{T_{hj}} y_{hijt}$$

La media se estima por $\hat{Y} = \hat{Y} / N$.

Varianza.

La varianza del estimador del total viene dada por la siguiente expresión:

$$V_b(\hat{Y}) = \sum_{h=1}^L V_b(\hat{Y}_h) \quad [10]$$

donde,

$$V_b(\hat{Y}_h) = \sum_{j=1}^{m_h} V(\hat{Y}_{hj}) = \sum_{j=1}^{m_h} M_{hj}^2 \left(1 - \frac{r_h}{M_{hj}}\right) \frac{1}{r_h} \frac{1}{M_{hj}-1} \sum_{i=1}^{M_{hj}} (y_{hij} - \bar{Y}_{hj})^2$$

donde:

$$\bar{Y}_{hj} = \frac{1}{M_{hj}} \sum_{i=1}^{M_{hj}} y_{hij}$$

y se estima por:

$$\hat{V}_b(\hat{Y}) = \sum_{h=1}^L \hat{V}_b(\hat{Y}_h)$$

donde,

$$\hat{V}_b(\hat{Y}_h) = \sum_{j=1}^{m_h} M_{hj}^2 \left(1 - \frac{r_h}{M_{hj}}\right) \frac{1}{r_h} \frac{1}{r_h-1} \sum_{i=1}^{r_h} (y_{hij} - \bar{y}_{hj})^2$$

donde:

$$\bar{y}_{hj} = \frac{1}{r_h} \sum_{i=1}^{r_h} y_{hij}$$

La varianza del estimador de la media es:

$$V_b(\hat{Y}) = \frac{1}{N^2} V_b(\hat{Y})$$

y se estima por:

$$\hat{V}_b(\hat{Y}) = \frac{1}{N^2} \hat{V}_b(\hat{Y})$$

Los coeficientes de variación del total y de la media coinciden.

IV.1.2.2. ESTIMACIÓN DE LA RAZÓN: RAZONES ENTRE VARIABLES

Muchas variables de interés, tales como las de rendimiento, se expresan como el cociente de los totales de dos variables. Así, el rendimiento de un cultivo determinado se expresa como el cociente entre la producción total (Y) y la superficie total (X):

$$R = \frac{Y}{X}$$

Se considera como estimador de R el de razón combinada, esto es, aquel en el que los totales de numerador (Y) y denominador (X) se estiman separadamente (por \hat{Y} y \hat{X} , respectivamente), en la forma indicada en IV.1.2.1.1. y IV.1.2.1.2, según se emplee muestreo sistemático ó muestreo por bloques:

$$\hat{R} = \frac{\hat{Y}}{\hat{X}}$$

Se trata de un estimador sesgado. Una cota para el sesgo es:

$$\frac{|B(\hat{R})|}{\sqrt{V(\hat{R})}} \leq \frac{\sqrt{V(\hat{X})}}{X}$$

esto es, en valor absoluto, el sesgo $|B(\hat{R})|$ dividido por la desviación típica $\sqrt{V(\hat{R})}$ es inferior o igual al coeficiente de variación del estimador \hat{X} , $CV(\hat{X})$, que se estima por:

$$\hat{CV}(\hat{X}) = \frac{\sqrt{\hat{V}(\hat{X})}}{\hat{X}}$$

En la práctica se conviene en considerarlo despreciable si $\hat{CV}(\hat{X})$ es inferior al 10 %:

$$\frac{|B(\hat{R})|}{\sqrt{V(\hat{R})}} \leq 0,1$$

La varianza del estimador de la razón, \hat{R} es (aproximadamente):

$$V(\hat{R}) = R^2 \left[\frac{V(\hat{Y})}{Y^2} + \frac{V(\hat{X})}{X^2} - \frac{2COV(\hat{Y}, \hat{X})}{YX} \right] \quad [11]$$

y se estima sustituyendo en la expresión anterior los valores desconocidos de R, Y, X, $V(\hat{Y})$ y $V(\hat{X})$ por sus estimaciones muestrales antes indicadas, a las que añadimos el estimador de $COV(\hat{Y}, \hat{X})$.

IV.1.2.2.1. Estimador de $COV(\hat{Y}, \hat{X})$: Muestreo sistemático

La covarianza entre los estimadores \hat{Y} y \hat{X} es:

$$COV_s(\hat{Y}, \hat{X}) = \sum_{h=1}^L COV_{sh}(\hat{Y}_h, \hat{X}_h) \quad [12]$$

donde:

$$COV_{sh}(\hat{Y}_h, \hat{X}_h) = M_h^2 \left(1 - \frac{r_h}{M_h}\right) \frac{1}{r_h (M_h - 1)} \sum_{j=1}^{M_h} (Y_{hj} - \bar{Y}_h) (X_{hj} - \bar{X}_h)$$

donde:

$$\begin{aligned} \bar{Y}_h &= \frac{1}{M_h} \sum_{j=1}^{M_h} Y_{hj} & Y_{hj} &= \sum_{i=1}^{m_h} Y_{hij} \\ \bar{X}_h &= \frac{1}{M_h} \sum_{j=1}^{M_h} X_{hj} & X_{hj} &= \sum_{i=1}^{m_h} X_{hij} \end{aligned}$$

y se estima por:

$$\hat{COV}_s(\hat{Y}, \hat{X}) = \sum_{h=1}^L \hat{COV}_{sh}(\hat{Y}_h, \hat{X}_h)$$

donde:

$$\hat{COV}_{sh}(\hat{Y}_h, \hat{X}_h) = M_h^2 \left(1 - \frac{r_h}{M_h}\right) \frac{1}{r_h (r_h - 1)} \sum_{j=1}^{r_h} (y_{hj} - \bar{y}_h) (x_{hj} - \bar{x}_h)$$

donde:

$$\begin{aligned} \bar{y}_h &= \frac{1}{r_h} \sum_{j=1}^{r_h} y_{hj} & y_{hj} &= \sum_{i=1}^{m_h} y_{hij} \\ \bar{x}_h &= \frac{1}{r_h} \sum_{j=1}^{r_h} x_{hj} & x_{hj} &= \sum_{i=1}^{m_h} x_{hij} \end{aligned}$$

IV.1.2.2.2. Estimador de $COV(\hat{Y}, \hat{X})$: Muestreo por bloques

La covarianza entre los estimadores \hat{Y} y \hat{X} es:

$$COV_b(\hat{Y}, \hat{X}) = \sum_{h=1}^L COV_{bh}(\hat{Y}_h, \hat{X}_h) = \sum_{h=1}^L \sum_{j=1}^{m_h} COV_{bj}(\hat{Y}_{hj}, \hat{X}_{hj}) \quad [13]$$

donde:

$$COV_{bj}(\hat{Y}_{hj}, \hat{X}_{hj}) = M_{hj}^2 \left(1 - \frac{r_h}{M_{hj}}\right) \frac{1}{r_h (M_{hj} - 1)} \sum_{i=1}^{M_{hj}} (Y_{hij} - \bar{Y}_{hj}) (X_{hij} - \bar{X}_{hj})$$

donde:

$$\bar{Y}_{hj} = \frac{1}{M_{hj}} \sum_{i=1}^{M_{hj}} Y_{hij}$$

$$\bar{X}_{hj} = \frac{1}{M_{hj}} \sum_{i=1}^{M_{hj}} X_{hij}$$

y se estima por:

$$\hat{COV}_b(\hat{Y}, \hat{X}) = \sum_{h=1}^L \sum_{j=1}^{m_h} \hat{COV}_{hj}(\hat{Y}_{hj}, \hat{X}_{hj})$$

donde:

$$\hat{COV}_{hj}(\hat{Y}_{hj}, \hat{X}_{hj}) = M_{hj}^2 \left(1 - \frac{r_h}{M_{hj}}\right) \frac{1}{r_h} \frac{1}{(r_h - 1)} \sum_{i=1}^{r_h} (y_{hij} - \bar{y}_{hj})(x_{hij} - \bar{x}_{hj})$$

donde

$$\bar{y}_{hj} = \frac{1}{r_h} \sum_{i=1}^{r_h} y_{hij}$$

$$\bar{x}_{hj} = \frac{1}{r_h} \sum_{i=1}^{r_h} x_{hij}$$

IV.1.2.3. ESTIMACIÓN DEL CAMBIO

El cambio entre dos instantes consecutivos del tiempo $t-1$ y t , se puede expresar en valores absolutos y relativos.

Variación absoluta.

La variación absoluta de la variable "Y" en estudio, entre los instantes $t-1$ y t , se estima por diferencia entre las estimaciones del total en uno y otro instante:

$$\hat{D}(Y_{t-1,t}) = \hat{Y}_t - \hat{Y}_{t-1}$$

Puesto que \hat{Y}_t e \hat{Y}_{t-1} son estimadores insesgados de Y_t e Y_{t-1} , resulta que $\hat{D}(Y_{t-1,t})$ es un estimador sin sesgo de $D(Y_{t-1,t}) = Y_t - Y_{t-1}$.

La varianza de este estimador es:

$$V(\hat{D}(Y_{t-1,t})) = V(\hat{Y}_t) + V(\hat{Y}_{t-1}) - 2 \text{COV}(\hat{Y}_t, \hat{Y}_{t-1})$$

y se estima sustituyendo en la expresión anterior los valores de $V(\hat{Y}_t)$ y $V(\hat{Y}_{t-1})$, desconocidos por sus estimaciones muestrales antes indicadas, a las que añadimos el estimador de $\text{COV}(\hat{Y}_t, \hat{Y}_{t-1})$ que se estima por [12] o [13], según que el muestreo sea sistemático o por bloques y sin más que sustituir Y por \hat{Y}_t y X por \hat{Y}_{t-1} .

Variación relativa.

La variación relativa de la variable en estudio entre los instantes de tiempo t-1 y t se estima por el cociente entre el estimador de la variación absoluta y el estimador del total en el instante t-1:

$$\frac{\hat{D}(Y_{t-1,t})}{\hat{Y}_{t-1}} = \frac{\hat{Y}_t}{\hat{Y}_{t-1}} - 1 = \hat{R} - 1$$

donde: $\hat{R} = \frac{\hat{Y}_t}{\hat{Y}_{t-1}}$ es un estimador de razón.

Se trata de un estimador sesgado. Una cota para el sesgo es:

$$\frac{|B(\hat{R})|}{\sqrt{V(\hat{R})}} \leq \frac{\sqrt{V(\hat{Y}_{t-1})}}{\hat{Y}_{t-1}}$$

su varianza es:

$$V\left(\frac{\hat{D}(Y_{t-1,t})}{\hat{Y}_{t-1}}\right) = V(\hat{R})$$

donde $V(\hat{R})$ es (aproximadamente):

$$V(\hat{R}) = R^2 \left[\frac{V(\hat{Y}_t)}{Y_t^2} + \frac{V(\hat{Y}_{t-1})}{Y_{t-1}^2} - \frac{2 \text{COV}(\hat{Y}_t, \hat{Y}_{t-1})}{Y_t Y_{t-1}} \right]$$

y se estima sustituyendo en la expresión anterior los valores desconocidos de R , Y_t , Y_{t-1} , $V(\hat{Y}_t)$, $V(\hat{Y}_{t-1})$ y $\text{COV}(\hat{Y}_t, \hat{Y}_{t-1})$, por sus estimaciones muestrales antes indicadas.

IV.1.2.4. ESTUDIO DE CASOS

Consideraremos dos casos de estudio: (I) uno es el de la estimación de la superficie de los cultivos y demás usos del suelo en la zona del fresón en Huelva, por muestreo sistemático de réplicas de segmentos y (II) el otro es el de la estimación de los usos del suelo en Campo de Dalías, por muestreo de bloques. La precisión de las estimaciones de estos dos planes de muestreo se comparan con la precisión de las estimaciones basadas en la "Encuesta de superficies" del M.A.P.A. (1990).

En el cuadro 2 se recogen las estimaciones de los usos del suelo en la zona del fresón en Huelva y en Campo de Dalías, obtenidas a partir de la muestra seleccionada en II.1.3.4.1 y II.1.3.4.2, respectivamente. El procedimiento de cálculo es el que se detalla en IV.1.2.1.1 y IV.1.2.1.2, respectivamente. Junto a ellas se recogen las estimaciones basadas en la "Encuesta de superficies" del M.A.P.A. seleccionada en II.1.1.4. y calculadas por el procedimiento presentado en IV.1.1.2.

CUADRO 2. ESTIMACIONES DE SUPERFICIES CULTIVADAS CON DOS DISEÑOS DE LA MUESTRA DE SEGMENTOS:

(I) el de la "Encuesta de superficies del M.A.P.A." y
(II) el de la "Encuesta sobre cultivos intensivos" de la Junta de Andalucía 1996.

HUELVA: ZONA DE CULTIVO DE FRESÓN.

Cultivos	Encuesta de superficies del M.A.P.A. Tasa de muestreo: 1,62 %		Encuesta sobre cultivos intensivos de Andalucía Tasa de muestreo: Estratos de intensidad alta. (1-fresa y 2-otros): 5,52 % Estrato de intensidad baja: 1,41 % Media de zona: 2,25 %			
	Ha.	Coeficiente de variación (%)		Ha.	Coeficiente de variación (%)	
		Muestra de réplicas [5]	Muestra de segmentos [6]		Muestra de réplicas	Muestra de segmentos
Hortícolas	5.691	22,6	53,9	8.009	12,2	15,7
Cereales	9.737	37,0	52,9	7.316	13,9	15,5
Industriales	5.953	41,8	79,1	4.264	14,8	24,8
HERBÁCEOS	30.968	11,1	53,9	32.994	6,4	8,0
Cítricos	3.329	34,3	64,2	6.714	18,2	22,6
Frutales	2.565	63,6	98,3	5.249	22,4	23,3
Olivar	7.691	46,4	62,9	7.243	28,4	29,3
Viñedo	4.139	30,6	57,2	5.921	9,8	20,7
LEÑOSOS	17.725	31,6	61,3	25.201	10,8	12,8

ALMERÍA - CAMPO DE DALÍAS

Cultivos	Encuesta de superficies del M.A.P.A. Tasa de muestreo: 1,68 %		Encuesta sobre cultivos intensivos de Andalucía Tasa de muestreo: Estratos de intensidad alta. Estrato 1: 0,8 % Estrato de intensidad baja: 0,7 % Media de zona: 0,8 %			
	Ha.	Coeficiente de variación (%)		Ha.	Coeficiente de variación (%)	
		Muestra de réplicas [5]	Muestra de segmentos [6]		Muestra de segmentos	
Invernaderos	15.337	17,7	19,7	19.287	5,2	
HERBÁCEOS	15.399	17,6	19,5	19.291	5,2	
LEÑOSOS	240	86,2	98,6	665	76,1	

La precisión de las estimaciones es mayor con la "Encuesta sobre cultivos intensivos" de la Junta de Andalucía, utilizada en Huelva y Almería, que con la "Encuesta de superficies" del M.A.P.A. (1990). En Huelva esa mejora (el coeficiente de variación (o error de muestreo) relativo a los cultivos hortícolas baja del 22.6% al 12,2%) es debida, en parte, a que la tasa de muestreo utilizada es mayor (2.25% frente al 1.62%), y en parte a la estratificación y al mejor reparto de la muestra que permite ésta. En Campo de Dalías (Almería), con una tasa de muestreo menor (0.8% frente al 1.68%), el mejor

reparto de la muestra que permite la estratificación y el ajuste del tamaño del segmento al idóneo en cada estrato, se traduce en una mejora de la precisión de las estimaciones de la "Encuesta sobre cultivos intensivos" (el error de muestreo relativo a la estimación de la superficie de invernaderos baja del 17.7% al 5.2%).

Cabe señalar que la estratificación resuelve, en parte, el problema de la inestabilidad del estimador de la varianza del estimador, que tiene planteado la "Encuesta de superficies" del M.A.P.A. (1990) (ver epígrafe IV.1.1.2). En el cuadro 2 se observa cómo, en Huelva, con la estratificación, el coeficiente de variación de la muestra de réplicas (sistemática) no difiere sustancialmente del de la muestra de segmentos (aleatoria simple) que es siempre más estable; mientras que la diferencia es sustancial con el diseño de la "Encuesta de superficies" del M.A.P.A., que no es estratificado. Esta diferencia deja de ser sustancial, sin embargo, en el Campo de Dalías, donde la práctica totalidad de la zona es un solo estrato ('Utilización intensiva del suelo - invernadero').

IV.2. ESTIMACIONES A PARTIR DE LA MUESTRA DE AGRICULTORES. SEGMENTO PONDERADO

La muestra de agricultores se selecciona a partir de la muestra de segmentos, por lo que la unidad de selección es el segmento. Este hecho tiene implicaciones sobre el procedimiento de estimación: cuando los datos se toman a nivel de explotación, se requiere asociar esos datos al segmento de la muestra.

En el muestreo de áreas se utilizan básicamente tres criterios para asociar a los segmentos de la muestra, las observaciones muestrales tomadas a nivel de explotación: se denominan "segmento abierto", "segmento ponderado" y "segmento cerrado" [Ford et al. (1986), Nealon (1984)]. El término "segmento abierto" denota un criterio de asociación según el cual una explotación k queda asociada al segmento j en el que radica la sede de la explotación. Este puede ser un criterio útil en las zonas en las que el agricultor reside en su explotación y, por tanto, identificando sede con residencia, es posible su aplicación. En cambio, cuando el agricultor reside en zonas urbanas, el criterio es de difícil aplicación. Este tipo de asociación puede ser utilizado para todas las variables estructurales o técnico-económicas. Sin embargo, es poco utilizado en la práctica porque el criterio "segmento ponderado" también es utilizable para todas las variables de interés y resulta más eficiente.

El criterio "segmento ponderado" consiste en lo siguiente: dado el valor de la variable en estudio observado en la explotación, se pondera este valor utilizando como factor de ponderación la superficie de la explotación interior al segmento dividida por la superficie total de la explotación, y se le asigna el valor resultante al segmento. Puede ser utilizado para todas las variables de interés y, en general, es el más eficiente de los tres considerados.

El criterio "segmento cerrado" exige que todas las observaciones sean tomadas a nivel de segmento, no de explotación. Es el idóneo para características directamente observables sobre el terreno, como los usos del suelo (es el utilizado para la estimación de las superficies de los cultivos), pero no es utilizable para características no observables directamente sobre el terreno, tales como las técnico-económicas, porque, como ya hemos señalado, es difícil para el encuestado determinar su valor en el segmento.

La exactitud de estos criterios es función de dos componentes (I) la precisión y (II) el sesgo.

La medida de precisión es la varianza. Ford et al.(1986), han comparado la precisión relativa de los tres estimadores considerados, para la estimación de cinco variables (el número total de explotaciones, el número total de bovinos, el de cerdos, la superficie cultivada de maíz y la de sorgo), en cada uno de cinco Estados de los Estados Unidos de América: el "segmento ponderado" resultó ser más preciso que el "segmento cerrado" y este último más preciso que el "segmento abierto" [ver también Nealon (1984)].

En cuanto al sesgo, el segmento abierto es de difícil aplicación y, por tanto, causa de errores de cobertura (errores debidos a la exclusión de elementos de la población). El segmento ponderado se ve afectado de un sesgo al alza causado por el error de medida de la superficie total de la explotación: la superficie declarada por el agricultor suele ser una subestimación de la verdadera, por lo que el coeficiente de ponderación resulta sobrestimado. Para evitar este sesgo, se han ensayado otros coeficientes de ponderación, algunos basados en la superficie agrícola útil y otros en la superficie labrada. Ninguno de ellos resultó ser más satisfactorio que el basado en la superficie total.

Ya se señaló, bajo el epígrafe II.2, que la gama de procedimientos de selección de la muestra de agricultores es muy amplia, por lo que nos limitaremos a considerar los procedimientos usados en los casos de estudio tratados en II.2.1. La diferencia entre ellos estriba en el procedimiento de selección de la muestra de segmentos y en el número de etapas de que consta el procedimiento de selección de la muestra de agricultores. En la zona del fresón, en Huelva, la muestra de segmentos es sistemática y el procedimiento de selección de la muestra de agricultores consta de tres etapas. En Campo de Dalías (Almería) la muestra de segmentos es por bloques y el procedimiento de selección de la muestra de agricultores consta de sólo dos etapas. En ambos casos, el criterio de asignación a los segmentos de la muestra de las observaciones tomadas a nivel de explotación es el de "segmento ponderado".

IV.2.1. ENCUESTA PARA ESTIMAR LAS CARACTERÍSTICAS ESTRUCTURALES Y TÉCNICO-ECONÓMICAS DEL CULTIVO DE FRESÓN EN HUELVA

El procedimiento de selección de la muestra se describe bajo el epígrafe II.2.1.1. y el de recogida de la información bajo el epígrafe III.3. La muestra de agricultores se seleccionó sólo en el estrato $h=2$. Dentro de cada una de las $r_h=5$ réplicas de segmentos de la muestra de primera etapa, de $m_h=10$ segmentos cada una, se seleccionó una muestra de $\omega_h=4$ segmentos, con probabilidades iguales. Dentro de cada segmento de esta muestra de segunda etapa, se seleccionó una muestra de $u_h=2$ agricultores. El tamaño de la muestra de agricultores fue, pues, de 40.

Si se denota por " Y_{hs} " el total de la actividad en estudio declarada por el agricultor " s " de la muestra y referida a toda su explotación y por " A_{hs} " la superficie total de la explotación, la parte Y_{hij_s} del total Y_{hs} de la actividad a asignar al área A_{hij_s} explotada por el agricultor " s " en el segmento " i " de la réplica " j " del estrato " h ", se define así ("segmento ponderado"):

$$Y_{hij_s} = (Y_{hs}/A_{hs}) A_{hij_s} \quad [14]$$

La parte del total de la actividad en estudio correspondiente al segmento " i ", entendida como el total Y_{hij} de los valores de Y_{hij_s} asignados al conjunto " T_{hij} " de agricultores con toda o parte de su explotación en el segmento " i " de la réplica " j ", se estima mediante el estimador de Hansen-Hurwitz:

$$\hat{Y}_{hij} = \sum_{s=1}^{u_{hj}} \frac{Y_{hij_s}}{u_{hj} \pi_{hij_s}} \quad [15]$$

donde u_{hij} denota el tamaño de la muestra de agricultores a seleccionar dentro del segmento "i" de la réplica "j" y π_{hij_s} denota la probabilidad de selección del agricultor "s" con toda o parte de su explotación en el segmento "i" de la réplica "j":

$$\pi_{hij_s} = A_{hij_s} / A_{hij}$$

donde A_{hij} es la superficie total del segmento.

El total de la actividad en estudio correspondiente a la réplica "j" del estrato h, entendida como el total Y_{hj} de las cantidades Y_{hij} en cada segmento "i" de la réplica "j" se estiman mediante la expresión:

$$\hat{Y}_{hj} = m_h \frac{1}{\omega_h} \sum_{i=1}^{\omega_h} \hat{Y}_{hij}$$

donde m_h es el tamaño de la réplica "j" y ω_h es el tamaño de la muestra de segmentos (de segunda etapa) dentro de la réplica j-ésima.

El total de la actividad en estudio en la población se estima a partir de [7], sustituyendo Y_{hj} por su estimador \hat{Y}_{hj} :

$$\hat{Y} = \sum_{h=1}^L \hat{Y}_h \quad [16]$$

donde \hat{Y}_h es el estimador del total $-Y_h-$ de la variable en estudio en el estrato h-ésimo:

$$\hat{Y}_h = M_h \frac{1}{r_h} \sum_{j=1}^{r_h} \hat{Y}_{hj}$$

Varianza.

\hat{Y} es un estimador insesgado de Y [Faulkenberry y Garoui (1991)] y su varianza es:

$$V_s(\hat{Y}) = \sum_{h=1}^L V(\hat{Y}_h) \quad [17]$$

La varianza del estimador \hat{Y}_h , $V(\hat{Y}_h)$, es la suma de tres componentes, cada una correspondiente a una de las etapas del muestreo:

$$V(\hat{Y}_h) = V_{h1} + V_{h2} + V_{h3}$$

1ª) La componente de primera etapa, V_{h1} , es la variación de \hat{Y}_h debida a la variación del total de la actividad en estudio entre réplicas: coincide con $V_s(\hat{Y}_h)$ en [8].

- 2ª) La componente de segunda etapa V_{h2} , es la variación de \hat{Y}_h debida a la variación entre segmentos, dentro de réplicas, del total de la actividad en estudio. Viene definida por:

$$V_{h2} = M_h \frac{1}{r_h} \sum_{j=1}^{r_h} V_{hj}$$

donde:

$$V_{hj} = m_h^2 \left(1 - \frac{\omega_h}{m_h}\right) \frac{S_{hj}^2}{\omega_h}$$

donde

$$S_{hj}^2 = \frac{1}{m_h - 1} \sum_{i=1}^{m_h} (Y_{hij} - \bar{Y}_{hj})^2$$

$$\bar{Y}_{hj} = \frac{1}{m_h} \sum_{i=1}^{m_h} Y_{hij}$$

y se estima por:

$$\hat{V}_{hj} = m_h^2 \left(1 - \frac{\omega_h}{m_h}\right) \frac{s_{hj}^2}{\omega_h}$$

$$s_{hj}^2 = \frac{1}{\omega_h - 1} \sum_{i=1}^{\omega_h} (Y_{hij} - \hat{Y}_{hj})^2$$

$$\hat{Y}_{hj} = \frac{1}{\omega_h} \sum_{i=1}^{\omega_h} Y_{hij}$$

- 3ª) La componente de tercera etapa es la variación de \hat{Y}_h debida a la variación entre tramos, dentro de segmentos, del total de la cantidad de la actividad en estudio. Está definida por la siguiente expresión:

$$V_{h3} = M_h \frac{1}{r_h} \sum_{j=1}^{r_h} \frac{m_h}{\omega_h} \sum_{i=1}^{\omega_h} V_{hij}$$

donde:

$$V_{hij} = \frac{1}{u_{hij}} \sum_{t=1}^{T_{hij}} \left(\frac{Y_{hijt}}{\pi_{hijt}} - Y_{hij} \right)^2 \pi_{hijt}$$

donde T_{hij} es el número de tramos (agricultores) en el segmento "i" de la réplica "j", π_{hijt} la probabilidad de selección del tramo "t" y u_{hij} es el tamaño de la muestra de agricultores en el segmento "i". Para el segmento "i" de la muestra de la réplica "j", V_{hij} se estima por:

$$\hat{V}_{hij} = \frac{1}{u_{hij}(u_{hij}-1)} \sum_{t=1}^{u_{hij}} \left(\frac{Y_{hijts}}{\pi_{hijts}} - \hat{Y}_{hij} \right)^2$$

El procedimiento de estimación descrito en los párrafos anteriores corresponde al plan de muestreo definido sólo por la muestra de agricultores y el estimador "segmento ponderado". Un procedimiento alternativo puede definirse de modo que se haga uso de los dos planes de muestreo disponibles: (I) el basado en la muestra de segmentos para la estimación de superficies y (II) el basado en la muestra de agricultores, para la estimación de las características estructurales y técnico-económicas por unidad de superficie.

Con este procedimiento de estimación alternativo, el total se estima como el producto de la superficie total del cultivo en estudio, \hat{A} estimada según el plan (I), por la razón, \hat{R} , de la característica estructural o técnico-económica por unidad de superficie, estimada según el plan (II):

$$\hat{Y}_{al} = \hat{A} \times \hat{R} \quad [18]$$

donde \hat{A} es el estimador de la superficie total definido en el epígrafe IV.1.2.1.1. y \hat{R} es un estimador de razón, esto es, el cociente entre el total de la variable en estudio y el de la superficie, definidos ambos según [16] y con la definición dada para Y_{hijts} en cada caso (ver epígrafe IV.1.2.2).

El coeficiente de variación de \hat{Y}_{al} se estima por:

$$\hat{C}V(\hat{Y}_{al}) = \sqrt{[\hat{C}V(\hat{A})]^2 + [\hat{C}V(\hat{R})]^2 + [\hat{C}V(\hat{A})]^2 [\hat{C}V(\hat{R})]^2} \quad [19]$$

donde $\hat{C}V(\hat{A}) = \sqrt{\hat{V}(\hat{A})} / \hat{A}$ y $\hat{C}V(\hat{R}) = \sqrt{\hat{V}(\hat{R})} / \hat{R}$ siendo $\hat{V}(\hat{A})$ y $\hat{V}(\hat{R})$ estimadores de las varianzas $V(\hat{A})$ y $V(\hat{R})$ definidas en [8] y [11], respectivamente.

Las estimaciones que siguen se refieren sólo al estrato 2-f, esto es, al estrato de fresón.

IV.2.1.1. CARACTERÍSTICAS ESTRUCTURALES

Las estimaciones relativas al número y tamaño de las explotaciones, a la personalidad jurídica del empresario, a la edad del jefe de la explotación, a la mano de obra agrícola disponible en las explotaciones (familiar o asalariada) y a los regímenes de tenencia, entre otras características estructurales, se obtienen todas de la misma forma. Se ilustran para el caso del número y tamaño de las explotaciones, a partir de una muestra piloto de 40 agricultores.

Número y tamaño de las explotaciones

El número total de explotaciones que cultivan fresón se estima mediante la expresión [16], haciendo en [14] $Y_{hs} = 1$, de modo que:

$$Y_{hijts} = A_{hijts} / A_{hs}$$

$$\hat{Y}_{hij} = A_{hij} \frac{1}{u_{hij}} \sum_{s=1}^{u_{hij}} \frac{1}{A_{hs}}$$

La varianza del estimador se define en [17], con la definición dada de Y_{hs} .

El número de explotaciones (con fresón) de una determinada clase de tamaño se estima mediante la expresión [16], haciendo en [14] $Y_{hs} = 1$ si A_{hs} es de esa clase de tamaño y $Y_{hs} = 0$ en otro caso. La varianza del estimador se define en [17] con la definición dada para Y_{hs} .

La superficie total de las explotaciones que cultivan fresón se estima mediante la expresión [16], haciendo en [14] $Y_{hs} = A_{hs}$. La varianza del estimador se define en [17] con la definición dada para Y_{hs} .

La superficie de las explotaciones de una determinada clase de tamaño, que cultivan fresón, se estima mediante la expresión [16] haciendo en [14] $Y_{hs} = A_{hs}$, de modo que $Y_{hij} = A_{hij}$, si A_{hs} es de la clase de tamaño en cuestión y $Y_{hs} = 0$ en otro caso. La varianza del estimador se define en [17] con la definición dada para Y_{hs} .

En el cuadro 3 se recogen las estimaciones del número de explotaciones de fresón, la de su superficie y su distribución entre las clases de tamaño consideradas. Estas estimaciones se han obtenido por el procedimiento alternativo (expresiones [18] y [19]) por considerarlo más eficiente.

CUADRO 3. ESTIMACIÓN DEL NÚMERO Y DE LA SUPERFICIE TOTAL DE LAS EXPLOTACIONES DE FRESÓN. 1996.

Clases de tamaño	Números de explotaciones de fresón		Superficie total de las explotaciones de fresón	
	Estimación	C.V. (%)	Estimación	C.V. (%)
Menos de 5 ha.	388	43,94	1.374	47,76
de 5 a 10 ha.	637	75,66	4.604	78,39
de 10 a 20 ha.	423	38,05	5.151	38,04
Más de 20 ha.	39	46,86	1.227	34,90
Total	1.487	16,13	15.356	15,31

En el disco compacto anejo a este documento se recogen las estimaciones de las restantes características estructurales consideradas.

IV.2.1.2. CARACTERÍSTICAS TÉCNICO-ECONÓMICAS

Se considera la estimación de las producciones y rendimientos del fresón, por una parte, y de los consumos de materias primas expresadas en unidades físicas por otra.

Producciones y rendimientos.

La producción puede ser estimada como el producto de la estimación de la superficie cultivada por la estimación del rendimiento o bien, directamente a partir del dato de la producción en las

explotaciones de la muestra.

La producción total de fresón y la de cada período considerado, puede ser estimada mediante la expresión [16]. Igualmente puede ser estimada mediante las expresiones [18] y [19]. En el cuadro 4 se recogen las estimaciones de la producción total de fresón y su distribución entre las clases de tamaño de explotación consideradas. Estas estimaciones se han obtenido a partir de [18] y [19].

CUADRO 4. ESTIMACIÓN DE LA PRODUCCIÓN DE FRESÓN. 1996.

Periodo	Clases de tamaño de las explotaciones	Toneladas de producto producidas en el cultivo de fresón	
		Estimación	C.V. (%)
Toda la temporada	Total	155.727	15,46
	Menos de 5 ha.	27.903	43,31
	de 5 a 10 ha.	49.455	65,88
	de 10 a 20 ha.	48.187	50,87
	más de 20 ha.	30.181	58,16
Enero	Total	4.713	28,41
	Menos de 5 ha.	649	55,69
	de 5 a 10 ha.	5.595	74,69
	de 10 a 20 ha.	898	55,24
	más de 20 ha.	571	53,42
Febrero	Total	11.490	19,91
	Menos de 5 ha.	1.407	53,74
	de 5 a 10 ha.	4.777	61,08
	de 10 a 20 ha.	1.597	56,55
	más de 20 ha.	3.710	47,50
Marzo	Total	34.490	17,28
	Menos de 5 ha.	6.083	49,06
	de 5 a 10 ha.	11.088	63,59
	de 10 a 20 ha.	9.078	52,24
	más de 20 ha.	8.242	52,94
Abril	Total	61.548	15,03
	Menos de 5 ha.	15.461	41,81
	de 5 a 10 ha.	20.742	69,35
	de 10 a 20 ha.	17.276	55,04
	más de 20 ha.	11.070	56,26
Mayo	Total	35.111	26,75
	Menos de 5 ha.	5.324	44,26
	de 5 a 10 ha.	8.188	53,90
	de 10 a 20 ha.	15.412	71,37
	más de 20 ha.	6.185	56,88
Junio	Total	11.806	16,95
	Menos de 5 ha.	5.681	56,55
	de 5 a 10 ha.	3.238	75,88
	de 10 a 20 ha.	3.823	70,05
	más de 20 ha.	5.064	95,35

El rendimiento medio por hectáreas de fresón se estima en 28.810 Kg./ha. Con un coeficiente de variación del 9,41 %.

Consumo de materias primas.

El consumo total de una determinada materia prima se estima mediante la expresión [16], haciendo en [14] Y_{hs} igual a la cantidad total consumida de dicha materia prima declarada por el agricultor. La varianza del estimador se define en [17].

El consumo de una determinada materia prima dentro de una determinada clase de tamaño de las explotaciones se estima mediante la expresión [16] haciendo en [14] Y_{hs} igual a la cantidad total consumida de dicha materia prima declarada por el agricultor si A_{hs} es de la clase de tamaño en cuestión y $Y_{hs} = 0$ en otro caso. La varianza del estimador se define en [17] con la definición dada para Y_{hs} . Asimismo, puede ser obtenidas mediante [18] y [19]. En los cuadros 5 y 6 se presentan las estimaciones del consumo de algunas materias primas totales y por clases de tamaño de las explotaciones, obtenidas a partir de estas últimas expresiones.

CUADRO 5. ESTIMACIÓN DEL CONSUMO DE ABONO DE COBERTERA POR EL CULTIVO DE FRESÓN. 1996.

Abono	Clases de tamaño de las explotaciones	Toneladas de producto consumidas en el cultivo de fresón	
		Estimación	C.V. (%)
Nitrato Amónico	Total	155.727	15,46
	Menos de 5 ha.	210	57,47
	de 5 a 10 ha.	251	58,57
	de 10 a 20 ha.	445	44,66
	más de 20 ha.	206	30,55
Nitrato Potásico	Total	1.646	24,40
	Menos de 5 ha.	247	52,15
	de 5 a 10 ha.	567	41,51
	de 10 a 20 ha.	601	56,11
	más de 20 ha.	231	32,73
Nitrato Cálcico	Total	470	36,91
	Menos de 5 ha.	167	44,06
	de 5 a 10 ha.	80	51,44
	de 10 a 20 ha.	206	57,44
	más de 20 ha.	17	44,14
Microelementos	Total	470	32,21
	Menos de 5 ha.	167	44,06
	de 5 a 10 ha.	80	51,44
	de 10 a 20 ha.	206	57,43
	más de 20 ha.	17	44,14

CUADRO 6. ESTIMACIÓN DEL CONSUMO MEDIO POR HECTÁREA DE LAS DISTINTAS MATERIAS PRIMAS.

Labor	Materia Prima	Consumo medio por hectárea	C.V. (%)
Abonado de Fondo	Estiércol de Ave (kg.)	11.000	21,69
	Otro tipo de estiércol (kg.)	640	56,35
	Complejo 15-15-15 (kg.)	642,10	12,05
	Otro abono Inorgánico (kg.)	154,60	49,45
Desinfección del suelo	Bromuro de metilo (kg.)	367,60	63,24
Acolchado	Plástico negro (kg.)	337,70	5,35
Plantación	Plantas (ud)	55.890	7,50
	Arquillos (ud)	5.150	5,35
Construcción de microtúneles	Plástico Perforado (kg.)	608.610	5,16
	Cinta (ml)	9.710	46,21
	Nitrato amónico (kg.)	205,58	21,01
	Nitrato potásico (kg.)	304,39	20,97
Abonado De	Nitrato cálcico (kg.)	228,19	30,42
	Microelementos (kg.)	86,99	29,59

En el disco compacto anejo a este documento se recogen las estimaciones del consumo de otras materias primas, a partir de la misma muestra piloto.

IV.2.2. ENCUESTA PARA ESTIMAR LAS CARACTERÍSTICAS ESTRUCTURALES Y TÉCNICO-ECONÓMICAS DE LOS CULTIVOS PROTEGIDOS EN ALMERÍA (CAMPO DE DALÍAS)

El procedimiento de selección de la muestra de agricultores se describe bajo el epígrafe II.2.1.2. y el de recogida de la información bajo el III.3. El tamaño de la muestra de agricultores sobre aspectos estructurales fue de 39, a una submuestra de los cuales se les encuestó acerca de los aspectos técnico-económicos de cada uno de los cultivos considerados, siendo el tamaño de esta submuestra el siguiente: para el calabacín 6, para el melón 10, para el pimiento 6, para la sandía 8, y para el tomate 15. Se trata de una muestra piloto.

Si se denota por " Y_{hs} " el total de la actividad en estudio declarada por el agricultor "s" de la muestra y referida a toda su explotación y por " A_{hs} " la superficie total de la explotación, la parte Y_{hij} del total Y_{hs} de la actividad a asignar al área A_{hij} explotada por el agricultor "s" en el segmento "i" del bloque "j" del estrato "h", la definimos así ("segmento ponderado"):

$$Y_{hij} = (Y_{hs}/A_{hs}) A_{hij} \quad [20]$$

La parte del total de la actividad en estudio correspondiente al segmento "i", entendida como el total Y_{hij} de los valores de Y_{hij} asignados al conjunto " T_{hij} " de agricultores con toda o parte de su explotación en el segmento "i" del bloque "j", se estima mediante el estimador de Hansen-Hurwitz:

$$\hat{Y}_{hij} = \sum_{s=1}^{u_{hij}} \frac{Y_{hij s}}{u_{hij} \pi_{hij s}} \quad [21]$$

donde u_{hij} denota el tamaño de la muestra de agricultores a seleccionar dentro del segmento "i" del bloque "j" y $\pi_{hij s}$ denota la probabilidad de selección del agricultor "s" con toda o parte de su explotación en el segmento "i" del bloque "j".

$$\pi_{hij s} = A_{hij s} / A_{hij}$$

donde $A_{hij s}$ es la superficie explotada por el agricultor "s" dentro del segmento "i" del bloque "j" y A_{hij} es la superficie total del segmento.

El total de la actividad en estudio se estima a partir de [9] sustituyendo Y_{hij} por su estimador \hat{Y}_{hij} definido en [21]:

$$\hat{Y} = \sum_{h=1}^L \hat{Y}_h \quad [22]$$

donde \hat{Y}_h :

$$\hat{Y}_h = \sum_{j=1}^{m_h} \hat{Y}_{hj}$$

donde:

$$\hat{Y}_{hj} = \frac{M_{hj}}{r_h} \sum_{i=1}^{r_h} \hat{Y}_{hij}$$

Varianza.

\hat{Y} es un estimador insesgado de Y [Faulkenberry y Garoui (1991)] y su varianza es:

$$V_b(\hat{Y}) = \sum_{h=1}^L V(\hat{Y}_h) \quad [23]$$

La varianza del estimador \hat{Y}_h es la suma de dos componentes, cada una correspondiente a una de las etapas del muestreo:

$$V(\hat{Y}_h) = V_{h1} + V_{h2}$$

1ª) La componente de primera etapa, V_{h1} , es la variación de \hat{Y}_h debida a la variación del total de la actividad en estudio entre segmentos dentro de bloques. Coincide exactamente con $V_b(\hat{Y}_h)$ definida en [10].

2ª) La componente de segunda etapa V_{h2} , es la variación de \hat{Y}_h debida a la variación entre tramos dentro de segmentos, del total de la actividad en estudio. Viene definida por:

$$V_{h2} = \sum_{j=1}^{m_h} M_{hj} \frac{1}{r_h} \sum_{i=1}^{r_h} V_{hij}$$

donde V_{hij} viene dada por la expresión:

$$V_{hij} = \frac{1}{U_{hij}} \sum_{t=1}^{T_{hij}} \left(\frac{Y_{hijt}}{\pi_{hijt}} - Y_{hij} \right)^2 \pi_{hijt}$$

donde π_{hijt} es la probabilidad de selección del tramo "t" del segmento "i" del bloque "j" del estrato "h" y U_{hij} es el tamaño de la muestra de tramos (agricultores) a seleccionar en el segmento "i" del bloque "j" del estrato "h" y T_{hij} es el número de tramos en dicho segmento.

Se estima por:

$$\hat{V}_{hij} = \frac{1}{U_{hij}(U_{hij}-1)} \sum_{t=1}^{U_{hij}} \left(\frac{Y_{hijt}}{\pi_{hijt}} - \hat{Y}_{hij} \right)^2$$

El coeficiente de variación del estimador \hat{Y} es:

$$CV(\hat{Y}) = \sqrt{V(\hat{Y})}/\hat{Y}$$

y se estima por:

$$\hat{C}V(\hat{Y}) = \sqrt{\hat{V}(Y)}/\hat{Y}$$

IV.2.2.1. CARACTERÍSTICAS ESTRUCTURALES

Las estimaciones relativas al número y tamaño de las explotaciones, a la personalidad jurídica del empresario, a la edad del jefe de explotación, a los regímenes de tenencia de la tierra, a la superficie de tierra cultivada, al grado de parcelación y a la mano de obra de la explotación, entre otras características estructurales, se obtienen todas de la misma forma, que las expuestas seguidamente para el caso del número y tamaño de las explotaciones.

Número y tamaño de las explotaciones.

El número total de explotaciones se estima mediante la expresión [22], haciendo en [20] $Y_{hs}=1$, de modo que:

$$Y_{hjs} = \frac{A_{hjs}}{A_{hs}}$$

$$\hat{Y}_{hij} = A_{hij} \frac{1}{U_{hij}} \sum_{s=1}^{U_{hij}} \frac{1}{A_{hs}}$$

La varianza del estimador es define en [23] con la definición dada aquí para Y_{hs} .

El número de explotaciones de una determinada clase de tamaño se estima mediante [22] haciendo en [20] $Y_{hs} = 1$, si A_{hs} es de esa clase de tamaño y cero en otro caso. La varianza del estimador se define en [23] con la definición dada aquí para Y_{hs} .

La superficie total de las explotaciones se estima mediante la expresión [22], haciendo en [20] $Y_{hs} = A_{hs}$, de modo que:

$$Y_{hijS} = A_{hijS}$$

$$Y_{hij} = A_{hij}$$

La varianza del estimador \hat{Y} se define en [23], con la definición dada aquí para Y_{hs} .

La superficie de las explotaciones de una determinada clase de tamaño se estima mediante la expresión [22] haciendo en [21] $Y_{hijS} = A_{hijS}$ si A_{hs} es de la clase de tamaño en cuestión y $Y_{hijS} = 0$ en otro caso. La varianza de este estimador se define en [23], con la definición dada aquí para Y_{hijS} .

En el cuadro 7 se recogen las estimaciones del número de explotaciones y de la superficie total de las explotaciones y su distribución entre las clases de tamaño consideradas. En el disco compacto anejo a este documento, se recogen las estimaciones de las restantes características estructurales consideradas. Todas ellas están basadas en la muestra piloto.

CUADRO 7. ESTIMACIÓN DEL NÚMERO Y DE LA SUPERFICIE TOTAL DE LAS EXPLOTACIONES DE CULTIVOS PROTEGIDOS EN LA ZONA DE CAMPO DE DALÍAS Y PONIENTE (1996).

Grupos de explotaciones	Número de explotaciones de cultivos protegidos		Superficie total de las explotaciones de cultivos protegidos	
	Estimación	C.V. (%)	Estimación (ha.)	C.V. (%)
Menos de 1 ha.	15.194	23,17	6.107	22,51
De 1 a 2 ha.	8.853	16,73	10.305	16,35
Más de 2 ha.	3.325	26,17	9.697	26,38
Total	24.374	15,65	26.109	15,52

IV.2.2.2. CARACTERÍSTICAS TÉCNICO-ECONÓMICAS.

Para cada uno de los cinco principales cultivos de la zona de Campo de Dalías y Poniente (calabacín, melón, sandía, pimiento y tomate) se estiman la producción total y el rendimiento, por una parte, y los consumos de materias primas, por otra.

Rendimientos y producciones.

Para estimar el rendimiento de un cultivo determinado, se dispone sólo de los datos recogidos en la encuesta sobre ese cultivo.

El rendimiento se estima a partir de las encuestas relativas al cultivo en cuestión y mediante un estimador de razón definido en la forma indicada bajo el epígrafe IV.1.2.2.:

$$\hat{R} = \frac{\hat{Y}}{\hat{A}} \quad [24]$$

donde (I) \hat{Y} es la estimación de la producción del cultivo en estudio, obtenida mediante la expresión [22] haciendo en [20] Y_{hs} igual a la producción del cultivo en estudio en la explotación "s" de la muestra y (II) \hat{A} es la estimación de la superficie del cultivo en estudio, estimada haciendo en [20] Y_{hs} igual a la superficie del cultivo en estudio en la explotación "s" de la muestra. La varianza del estimador \hat{R} se define en [11] (ver epígrafe IV.1.2.2.2).

Para estimar la producción se dispone de la estimación del rendimiento obtenida en el párrafo anterior y de los datos recogidos en todas las encuestas sobre la superficie de los cultivos.

La producción se estima mediante el producto de la estimación del rendimiento por la estimación de la superficie del cultivo en estudio, obtenida en la forma indicada bajo el epígrafe IV.1.2.1.2. El coeficiente de variación de este estimador se estima mediante la expresión:

$$\hat{C}V(\hat{Y}) = \sqrt{[\hat{C}V(\hat{R})]^2 + [\hat{C}V(\hat{A})]^2 + [\hat{C}V(\hat{R})]^2 [\hat{C}V(\hat{A})]^2} \quad [25]$$

donde $\hat{C}V(\hat{R})$ se define en IV.1.2.2.2. y $\hat{C}V(\hat{A})$ es el coeficiente de variación del estimador de la superficie definida en IV.1.2.1.2.

En el cuadro 8 se recogen las estimaciones de la producción total y rendimiento de los cultivos considerados.

CUADRO 8. ESTIMACIÓN DEL RENDIMIENTO Y DE LA PRODUCCIÓN TOTAL DE LOS CULTIVOS CONSIDERADOS EN CAMPO DE DALÍAS (1996).

Cultivo	Estimación del rendimiento medio		Estimación de la producción total	
	Estimación (kg/ha x año)	C.V.(%)	Estimación (kg/año)	C.V.(%)
Calabacín	45.400	30,43	158.533.600	44,09
Melón	46.272	15,92	348.273.072	27,87
Pimiento	54.324	24,52	557.961.804	31,76
Sandía	85.831	6,67	484.859.319	23,17
Tomate	75.589	15,10	370.929.790	24,99

Consumo de materias primas.

El consumo medio por unidad de superficie de una determinada materia prima para un cultivo determinado, se estima a partir de las encuestas relativas al cultivo en cuestión y mediante un estimador de razón de la forma definida en [24] siendo ahora (I) \hat{Y} la estimación del consumo total de la materia prima en estudio, obtenida mediante la expresión [22] haciendo en [20] Y_{hs} igual al consumo de dicha materia prima para el cultivo en estudio en la explotación "s" de la muestra y (II) \hat{A} la estimación de la superficie del cultivo en estudio. La varianza del estimador se define [11] (ver epígrafe IV.1.2.2).

El consumo total de la materia prima en estudio por el cultivo en cuestión, se estima mediante el producto de la estimación del consumo medio por unidad de superficie por la estimación de la superficie obtenida en la forma indicada bajo el epígrafe IV.1.2.1.2. El coeficiente de variación de este estimador se estima mediante la expresión [25] siendo ahora $\hat{C}V(\hat{R})$ el calculado en la forma indicada en el párrafo anterior.

En el cuadro 9 se presentan las estimaciones del consumo total de materias primas empleadas en el cultivo de melón y las estimaciones de los consumos medios por unidad de superficie, obtenidos a partir de la muestra piloto (10 agricultores). En el disco compacto anejo a este documento se recogen las estimaciones relativas a los restantes cultivos considerados.

CUADRO 9. ESTIMACIÓN DEL CONSUMO MEDIO POR HECTÁREA Y DEL CONSUMO TOTAL DE MATERIAS PRIMAS EN EL CULTIVO DE MELÓN, EN LA ZONA DE CAMPO DE DALÍAS Y PONIENTE. 1996.

Labor y producto	Estimación del consumo medio por hectárea		Estimación del consumo total	
	Estimación	C.V. %	Estimación	C.V. %
ABONADO DE FONDO				
- Estiércol (kg)	109.018	61,48	820.469.468	66,96
- Complejo 15-15-15 (kg)	168	104,98	1.261.377	109,97
DESINFEC. DEL SUELO				
- Metan sodio (cc)	204	50,35	1.535.276	56,34
- DD (cc)	36	99,62	271.037	104,59
PLANTAR / SEMBRAR				
- Plantas/semillas (ud)	11.933	15,11	89.807.758	27,39
CUBIERTA				
- Plástico (kg)	926	41,12	6.968.555	47,83
ABONADO DE COBERTERA				
- Nitrato amónico (kg)	602	95,65	4.530.652	100,63
- Nitrato cálcico (kg)	306	80,67	5.305.956	85,73
- Ácido nítrico (l)	49	63,30	368.774	68,71
- Ácidos húmicos (l)	62	54,72	466.612	60,48
- Ácido fosfórico (l)	23	88,89	173.098	93,89
- Fosfato monoamónico (l)	35	88,89	263.410	93,89
- Nitrato potásico (kg.)	625	88,78	4.703.750	93,78
- Sulfato magnésico (kg.)	3	88,89	25.578	93,89
TRAT. HORMONALES				
- Procarpil (cc)	970	88,91	7.300.220	93,90
TRAT. FITOSANITARIOS				
* Contra araña roja				
- Vertimec (cc)	5.081	17,91	15.661.606	29,11
* Contra araña blanca				
- Tedión-keltane (cc)	1.939	88,89	14.595.914	93,89
* Contra moca blanca				
- Applaud (cc)	1.939	88,89	14.595.914	93,89
- Endosulfan 35 (cc)	7.495	59,36	56.407.370	64,91
- Condifor (cc)	1.111	111,11	8.361.386	116,13
* Contra minador				
- Trigard (cc)	485	88,89	3.650.110	93,89
- Vertimec (cc)	1.340	33,61	10.084.840	41,20

CAPITULO V
MARCOS MÚLTIPLES DE ÁREAS Y LISTAS

CHAPTER 4
BARCODE MULTIPLE ACCESS SYSTEM

V MARCOS MÚLTIPLES DE ÁREAS Y LISTAS

En el epígrafe I se señaló que la sensibilidad a los valores extremos ("outliers") es uno de los inconvenientes del muestreo basado sólo en marcos de áreas. La inclusión en la muestra de áreas de elementos en los que la variable en estudio toma un valor anormalmente diferente a los demás, por ejemplo de una explotación agrícola de gran tamaño, afecta muy desfavorablemente a la precisión de las estimaciones basadas sólo en la muestra de áreas.

Este problema, como cualquier otro relativo al diseño de muestras, puede ser tratado en la fase de diseño del procedimiento de selección de la muestra y/o en la fase de estimación. Hidiroglou y Srinath (1981) consideran cuatro estimadores del total, robustos respecto de los "outliers". Tres de ellos son ponderaciones de los valores "no-outliers", por una parte, y de los valores "outliers", por otra. El cuarto es el estimador del total postestratificado, utilizable sólo si el número total de "outliers" en la población es conocido. Los autores citados estudian la eficiencia relativa de estos estimadores para distintos valores de los parámetros de los que dependen y concluyen que ninguno de ellos es uniformemente más eficiente que los demás.

Gross et al.(1986) consideran tres estimadores, uno de los cuales, al que denominan "estimador sorpresa" es idéntico a uno de los estimadores propuestos por Hidiroglou y Srinath (1981). Los otros dos son estimadores Winsorizados. Los autores concluyen, a partir del muestreo estratificado de una población simulada, que los estimadores Winsorizados son más eficientes que el tercero de los considerados.

Una aproximación alternativa a este problema de los "outliers" es la que se centra en la fase de selección de la muestra y que consiste en tratar de incluir en una lista los "outliers" y en utilizar en el muestreo marcos múltiples de áreas y lista. Por ejemplo, en el muestreo de explotaciones agrarias, se trata de elaborar una lista de grandes explotaciones, las cuales serán todas excluidas del marco de áreas y muestreadas a partir del marco de lista. La lista puede ser confeccionada a partir de fuentes diversas tales como el último censo, las asociaciones de productores, las cooperativas, ó las Cámaras Agrarias. La lista debe contener información sobre la localización de la explotación, la superficie total, la superficie destinada al cultivo ó actividad en estudio, etc.

Este muestreo mixto de áreas y listas resulta también muy eficiente para estimar actividades raras, practicadas por un muy reducido número de explotaciones. En este caso la lista debe recoger al mayor número posible de esas explotaciones raras [Nealon (1984)].

Asimismo, el muestreo múltiple de lista y área es útil para anular el error de cobertura: la lista suele ser un marco de muestreo incompleto por lo que requiere ser completado mediante la muestra de áreas.

La estrategia de muestreo basada en marcos múltiples fue inicialmente propuesta por Hartley (1962). Fuller y Burmeister (1972) sugirieron algunas mejoras de la misma. Aquí seguiremos la aproximación de Hartley (1962).

Se consideran dos marcos: A (marco de áreas) y B (marco de lista). De cada uno se seleccionan sendas muestras de tamaños n_A y n_B . Se supone que todo elemento de la población pertenece al menos a un marco así como que es posible saber, para cada unidad seleccionada en la muestra, si pertenece o no al otro marco. Bajo estos supuestos, será posible clasificar el conjunto de unidades de las dos muestras en tres categorías o dominios:

Dominio (a): Unidades incluidas sólo en el marco A.

Dominio (b): Unidades incluidas sólo en el marco B.

Dominio (ab): Unidades incluidas en los dos marcos A y B.

Si los tamaños poblacionales de los dominios $-N_A, N_B, N_{AB}-$ son conocidos y los elementos son clasificables a priori, entonces se trata de un caso de muestreo estratificado. El muestreo sería postestratificado si no es posible clasificar los elementos a priori.

N_B es conocido pero N_A , esto es, el número de explotaciones en el marco de áreas que, por ser un marco completo, coincide con el total de explotaciones en la población, es desconocido. Por ser completo el marco de áreas se verifica que: $N_{AB} = N_B$. En estas condiciones, el estimador del total a considerar será la suma de dos estimadores: uno es el estimador del total en el Dominio (a), $Y_{(a)}$ el otro es el estimador del total en el marco B, $Y_{(B)}$:

$$\hat{Y} = \hat{Y}_{(a)} + \hat{Y}_{(B)}$$

En el cálculo de $\hat{Y}_{(a)}$ se debe poner especial cuidado en excluir de la muestra de áreas todos los tramos correspondientes a los elementos del marco B.

La varianza del estimador múltiple será:

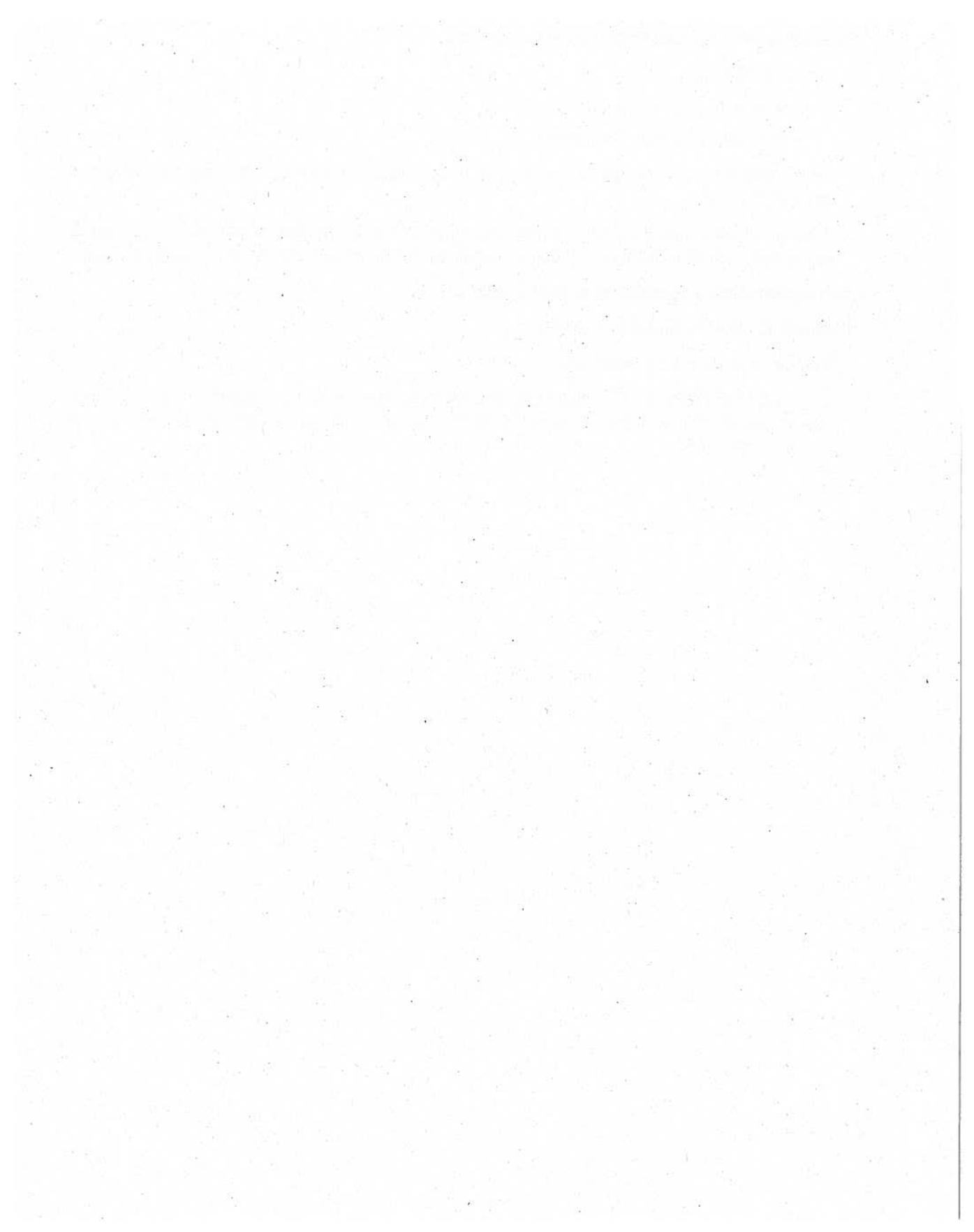
$$V(\hat{Y}) = V[\hat{Y}_{(a)}] + V[\hat{Y}_{(B)}]$$

Este estimador es más eficiente que los estimadores de "segmento abierto, cerrado y ponderado", para muchas características [Nealon (1984)].

REFERENCIAS

- AMBROSIO, L., ALONSO, R. y VILLA, A. (1993): Estimación de superficies cultivadas por muestreo de áreas y teledetección. Precisión relativa. *Estadística Española*, 35, 132, pp. 91-103.
- AMBROSIO, L., GALLEGO, J. (1994): Encuestas agrícolas en España basadas en un marco de áreas. Estudio preparado para la FAO. E.T.S.I. Agrónomos (Madrid) y Centro Común de Investigaciones de la Unión Europea (Ispra, Italia).
- BELLHOUSE, D.R. (1977): Some optimal designs for sampling in two dimensions. *Biometrika*, 64, 3, pp. 605-11.
- COCHRAN, W.G. (1977): *Sampling Techniques*. Wiley.
- COTTER, J., NEALON, J. (1987): Area frame design for agricultural surveys. Research and Applications Division. National Agricultural Statistics Service. United States Department of Agriculture. Washington, D.C.
- FAO (1996): Encuestas agrícolas con múltiples marcos de muestreo. Vol.1. Encuestas basadas en métodos de muestreo de áreas y explotaciones. Colección FAO: Desarrollo Estadístico, 7.
- FAO (1998): Encuestas agrícolas con múltiples marcos de muestreo. Vol.2. Programas de encuestas agropecuarias basadas en diseños de muestreo con marcos de áreas o doble marco de selección (de áreas y listas). Colección FAO: Desarrollo Estadístico, 10.
- FAULKENBERRY, G.D., GAROUI, A. (1991): Estimating a population total using an Area Frame. *Journal of the American Statistical Association*, Vol. 86, nº 414, Theory and Methods, pp. 445-449.
- FORD, B. L., NEALON, I., TORTORA, R. D. (1986): Area frame estimators in agricultural surveys: Sampling versus nonsampling errors. *Agricultural Economics Research*, vol. 38, nº 2, pp 1-10.
- FULLER, W. A., BURMEISTER, L. F. (1972): Estimators for samples selected from two overlapping frames. *Proceedings of the Social Statistics Section. American Statistical Association*.
- GALLEGO, F.J., DELINCÉ J. y CARFAGNA, E. (1994): Two stage area frame on squared segments for farm surveys. *Survey Methodology*, Vol 20, nº 2, pp 107-115.
- GROSS, W.F., BODE, G., TAYLOR, J. M., LLOYD SMITH, C. W. (1986): Some finite population estimators which reduce the contribution of outliers. In FRANCIS, I. S., MANLY, B. F. J., LAN, F. C. (ed): *Pacific Statistical Congress. Proceedings of the Congress. Auckland, New Zeland. Elsevier Science Publishers. North Holland*.
- GROVES, R. M. (1989): *Surveys errors and surveys cost*. Wiley. N.Y.
- HANSEN, M.H., HURWITZ, W.N., MADOW, W.G. (1953): *Sample Survey. Methods and Theory. Vol 1. Methods and Applications. Vol 2. Theory*. Wiley.
- HARTLEY, H. O. (1962): Multiple Frame Surveys. *Proceedings of the Social Statistics Section. American Statistical Association*.
- HIDIROGLOU, M. A. SRINATH, K. P. (1989): Some estimators of population total from simple random samples containing large units. *Journal of the American Statistical Association*, Vol. 76, No. 375, pp 690-695.
- IGLESIAS, L. (1998): Muestreo de áreas: Diseño de muestras y estimación en pequeñas áreas. Tesis Doctoral. E.T.S.I. Agrónomos. Madrid.
- JESSEN, R.J. (1978): *Statistical Survey Techniques*. Wiley.
- KISH, L. (1965): *Survey Sampling*. Wiley
- KOOP, J.C. (1990): Systematic sampling of two-dimensional surfaces and related problem. *Communications in Statistics. Theory and Methods*, 19(5), pp. 1701-1750.

- LÓPEZ, L., CASTILLO, J. E., FUENTES, M., PALOMAR, F., FERNÁNDEZ, E. J., VISERAS, J., LÓPEZ, F.J. (1994): Caracterización de los sistemas de producción hortícola de invernaderos en la provincia de Almería. Colección de Documentos. Fundación para la Investigación Agraria en la Provincia de Almería. Instituto de Fomento de Andalucía.
- M.A.P.A. (1990): Encuesta sobre superficies y rendimientos de cultivos. Boletín Mensual de Estadística. Abril. Ministerio de Agricultura, Pesca y Alimentación. Madrid.
- M.A.P.A. (1999): Encuesta sobre superficies y rendimientos de cultivos del año 1998. Boletín Mensual de Estadística. Abril. Ministerio de Agricultura, Pesca y Alimentación. Madrid.
- NEALON, J. P. (1984): Review of the Multiple and Area Frame Estimators. Statistical Reporting Service. United States Department of Agriculture. Report nº 80. Washington D.C.
- QUENOUILLE, M.H. (1949): Problems in plane sampling. Ann. Math. Statist., 20, pp. 355-375.
- U.S.D.A.(1983): Scope and methods of the Statistical Reporting Service. Miscellaneous Publication, Nº 1308. United States Department of Agriculture. Washington, D.C.



Requisitos para ejecución del contenido del CD

- Un navegador Web:
 - Microsoft Internet Explorer o superior.
 - Netscape Navigator 3 o superior.
- Se recomienda, para una óptima visualización de las páginas, establecer una resolución de pantalla de 800 x 600.
- Para visualizar la información gráfica se puede utilizar ArcExplorer (explorador de datos geográficos gratuito, de libre distribución), que se puede descargar de la dirección: WWW.ESRI-ES.COM.

Instrucciones para ejecutar el contenido del CD

1. Inserte el CD en la unidad de CD-ROM.
2. Vaya al menú de Inicio y seleccione Ejecutar.
3. Escriba D:\ CONTENIDO. HTM (siendo "D" la unidad correspondiente al CD-ROM. En caso de error, éste puede ser debido a que su unidad es distinta de "D": utilice la correcta, en su caso) y haga clic en Aceptar (OK)

