

**ESTIMACION DEL AREA CULTIVADA CON TRIGO A PARTIR DE DATOS
LANDSAT MEDIANTE EL METODO DE MUESTREO EN DOS FASES***

CARLOS A. CAPPELLETTI

MARIA C. SERAFINI

**UNIVERSIDAD NACIONAL DE LUJAN
C.C.221, 6700 - LUJAN (B) - ARGENTINA**

RESUMEN

El objetivo de este trabajo es estimar el area sembrada con trigo en el partido de Trenque Lauquen, campaña 82/83, utilizando datos combinados de campo e imágenes LANDSAT, mediante un sistema de muestreo en dos fases con estimadores de regresión, error probable prefijado y costo mínimo.

Los resultados obtenidos arrojan un area total cultivada de 92.217 has. con un coeficiente de variación de 8,93%, discrepando en 1,73% con los logrados mediante simulación Montecarlo del sistema propuesto.

ABSTRACT

The objective of this study is to estimate the cultivated area of wheat using combined ground and Landsat data through a two-phase sampling system with estimation of regression, fixed error rate and minimum cost.

The result obtained was 92.217 ha. with a coefficient of variation of 8,93%. The result compared with the simulated system by Montecarlo technique presented an error (relative difference) of 1,73%.

1.- INTRODUCCION

La disponibilidad de información relativa al area ocupada con cultivos anuales es de gran importancia para la elaboración de estrategias referentes al almacenaje, consumo y exportación en los niveles superiores de decisión de gobierno. Por tal razón los servicios de estadísticas agrícolas deben incorporar técnicas adecuadas para mejorar la precisión de las predicciones y aumentar la frecuencia de las mismas dentro del período previo a la cosecha.

El objetivo de este trabajo, como parte de un proyecto mayor, es diseñar y operar un sistema de muestreo estadístico que combina dos fuentes de datos, relativos al número de hectáreas cultivadas en segmentos muestrales previamente seleccionados. Una fase es el relevamiento tradicional a campo y la otra es la clasificación automática de la información, en los mismos segmentos muestra-

les, que retienen los sensores ubicados a bordo de satélites de relevamiento de la Tierra, en este caso el LANDSAT.

Los datos que se emplean corresponden al Partido de Trenque Lauquen, Provincia de Buenos Aires, en la campaña triguera de 1982/83. Los mismos fueron procesados mediante el uso del sistema interactivo de análisis digital de datos satelitarios en el Centro de Sensores Remotos de la Comisión Nacional de Investigaciones Espaciales.

2.- METODOLOGIA

El diseño estadístico que se emplea es el llamado Muestreo Aleatorio de Dos Fases con estimadores de regresión que combina los datos de segmentos de campo observados directamente y la información que proveen sobre los mismos segmentos los sensores remotos. Cappelletti y otros (1982a, 1982b) emplearon este sistema en cultivos de caña de azúcar obteniendo estimadores con índices

de confiabilidad del 99% respecto a la verdad terrestre.

La característica principal del sistema con dos fases y estimadores de regresión, se apoya en la correlación entre dos variables, Y (datos de campo) y X (datos LANDSAT) y en la relación entre los costos de obtener un valor de Y y uno de X, C1 y C2, respectivamente.

En su forma mas simple el método consiste en seleccionar aleatoriamente K de los N segmentos ($K \leq N$) que constituyen el universo y una vez ubicados en las imagenes satelitarias, se efectúa el análisis automático que permite disponer del area ocupada con el cultivo de interés. Estos valores son simbolizados X_1, X_2, \dots, X_k .

De los K segmentos seleccionados en la primera fase se escogen aleatoriamente NP de ellos ($NP < K$), los que son ubicados en mapas adecuados para ser recorridos en operaciones de campo. En estos segmentos se registran los valores simbolizados Y_1, Y_2, \dots, Y_{NP} , correspondientes a la verdad terrestre del area cultivada, y sin dudas el costo de relevar estos segmentos es muy superior al correspondiente a la fase anterior.

A partir de las condiciones descritas se define el estimador por regresión para el area media cultivada por segmento muestral mediante la función

$$\bar{Y}_R = \bar{Y}_L + b (\bar{X}_L - \bar{X}_C) \quad (1)$$

donde:

- \bar{Y}_R estimador del area media cultivada por segmento.
- \bar{Y}_L media de los K segmentos relevados en la imagen LANDSAT
- \bar{X}_C media de los NP segmentos relevados a campo
- \bar{X}_L media de los mismos NP segmentos de campo pero con las mediciones previstas por la clasificación automática sobre la imagen LANDSAT

b coeficiente de regresión lineal entre datos LANDSAT y campo calculado en los NP segmentos de la segunda fase

La función de costo simple que se utiliza es

$$C = C_1 \times K + C_2 \times NP \quad (2)$$

donde:

C1 y C2 son los costos de relevamiento de segmentos LANDSAT y campo, respectivamente, y C es el costo total del plan.

La varianza del estimador (1) depende de los valores de K y NP, los que a su vez son funciones de la correlación entre Y y X, de la relación entre C1 y C2 de la función (2) y de la precisión deseada para la estimación de la media \bar{Y}_R y posteriormente de Y_R , area total cultivada en el universo previamente definido.

Cochran (1963) y Hansen y otros (1966) presentan los fundamentos matemáticos del método.

3.- RESULTADOS Y DISCUSION

3.1. DETERMINACION DEL NUMERO DE SEGMEN- TOS MUESTRALES

A partir de una muestra piloto bivariada constituida por 36 segmentos de campo con tamaño medio de 546 hectáreas, que provee pares de datos correspondientes a la observación terrestre y al análisis digital del correspondiente segmento en la imagen, se calculó el coeficiente de correlación y el de regresión cuyos valores fueron $r=0,915$ y $b=0,971$.

En la Tabla 1 se indica la cantidad de segmentos necesarios en cada fase del muestreo y su relación (F) para restringir la tasa del error probable de la estimación a no mas de 10% con una probabilidad del 90% considerando $r=0,915$ y varias relaciones de costos por unidad en ambas fases. Los valores K y NP corresponden al número necesario de segmentos a muestrear en cada fase.

TABLA I

VALORES DE K, NP, C y F SEGUN LA RELACION C1/C2

C1/C2	K	NP	C	F
1/5	143	23	258	0,165
1/10	165	19	355	0,116
1/15	180	17	435	0,095
1/20	192	15	492	0,082
1/25	203	14	553	0,073

Observando la quinta columna de la tabla, se nota que la tasa de muestreo entre ambas fases decrece en la medida que el costo de relevar un segmento de campo aumenta respecto al costo de relevar un segmento LANDSAT. Como consecuencia de ello NP disminuye y es compensado por un incremento de K.

La tasa de muestreo depende de la relación C1/C2 y de la magnitud del coeficiente de correlación entre Y y

X que se muestra en la Tabla 2.

TABLA 2

PROPORCION OPTIMA ENTRE NP Y K PARA DISTINTOS VALORES DE r Y DE C1 / C2

C1/C2	0,75	0,80	0,85	0,90	0,95
1/5	0,394	0,335	0,277	0,216	0,147
1/10	0,279	0,237	0,196	0,153	0,104
1/15	0,228	0,194	0,160	0,125	0,085
1/20	0,197	0,168	0,139	0,108	0,074
1/25	0,176	0,150	0,124	0,097	0,066

Es importante tener en cuenta que la magnitud de r depende de la capacidad de identificación del cultivo, cuando se realiza la clasificación automática mediante análisis digital del segmento muestral en la imagen LANDSAT (problemas de baja resolución), por cuanto el correspondiente dato de campo, considerado la verdad terrestre, se supone que es medido sin error.

El número de segmentos de campo necesarios para satisfacer los requisitos de precisión prefijados en el caso que se emplee un muestreo aleatorio simple con la fase de campo únicamente, resulta ser igual a 95, es decir mayor que cuatro veces el número requerido en la primera línea de la Tabla 1, y dado que en tal caso C2=5, el costo total del plan sería C=475 unidades monetarias, es decir casi el doble.

La tabla 3 muestra la eficiencia relativa del sistema que se propone respecto a un muestreo aleatorio simple con la fase de campo solamente, según diversos valores de la relación de costos C1/C2 y para r=0,915.

TABLA 3

EFICIENCIA RELATIVA RESPECTO A UNA UNICA FASE (CAMPO)

C1/C2	Ef. Rel. (%)
1/5	207
1/10	244
1/15	270
1/20	290
1/25	306

La eficiencia relativa se interpreta como la cantidad de información que provee el sistema de muestreo en dos fases con respecto a un sistema de fase única (datos de campo) para el mismo costo total de la operación.

Si C1/C2 = 1/5, el valor 207% indica que se obtiene un poco mas del doble de información cuando se invierten 258 unidades monetarias en el sistema en dos fases.

3.2. ESTIMACION DEL AREA CULTIVADA

Asumiendo que la relación de costos C1/C2 entre ambas fases es de 1/5 se completa la muestra de la primera fase a K=143 segmento y junto con NP=23 segmentos de la segunda fase se procede a calcular las estimaciones deseadas, las que son presentadas en la Tabla siguiente.

TABLA 4

ESTIMACION DEL AREA CULTIVADA CON TRIGO AÑO 1982/83

Media (\bar{Y})	149,90 Has. cult. seg.
Varianza de \bar{Y}	179,39
Des. Est. de \bar{Y}	13,40 Has. cult. seg.
Total (Y)	92.217,00 Has. cultivadas
Varianza de Y	6,79 (E+07)
Des. Est. de Y	8,23 Has. cult.
Coef. de Var. de Y	8,93 %

La estimación del area cultivada, objetivo de este trabajo, es de 92.217 hectáreas con un coeficiente de variación de 8,93%. La estimación garantiza un error probable no superior al 10 % con una probabilidad de 0,90.

3.3. ESTABILIDAD DE LA PREDICCIÓN

A partir de los valores K=143 y NP = 23 se simuló mediante técnicas Montecarlo una sucesión de 50 muestras aleatorias en dos fases y se estudió el comportamiento de los estimadores de la media por segmento y del total del universo. Los resultados fueron:

TABLA 5

RESULTADOS DE SIMULAR 50 MUESTRAS ALEATORIAS

Media	147,36 Has. cult. seg.
Des. Est. de la Media	10,76 Has. cult. seg.
C.V.	7,30 %
Total	90.626,00 Has. cult.
Des. Est. del Total	6.617,00 Has. cult.
C.V.	7,30 %

Comparando el total estimado de la Tabla 4 con el de la 5 se observa una diferencia absoluta de 1591 Has. que representa el 1,73 % que consideramos un indicador muy promisorio para los fines de la investigación.

4.- CONCLUSIONES

En base a los resultados alcanzados en el presente trabajo, en el que se utilizan datos combinados de segmentos de campo y de imágenes LANDSAT en un sistema de muestreo de dos fases

con estimador de regresión, para estimar el área cultivada con trigo en el Partido de Trenque Lauquen, campaña 1982/83, se llega a las siguientes conclusiones:

- 4.1. El sistema de muestreo empleado es adecuado en cuanto a eficiencia y costo total del plan.
- 4.2. El sistema de muestreo en dos fases no puede ser sustituido por un sistema de muestreo de una fase que emplee datos LANDSAT solamente, a no ser que se logren condiciones técnicas de análisis digital con un grado de resolución mayor al actual y que produzca un coeficiente de correlación, con la verdad terrestre, cercano a uno.
- 4.3. La utilización de un sistema de muestreo de una fase que emplee datos de campo únicamente, requiere una asignación presupuestaria del doble de unidades monetarias como mínimo, respecto al sistema de dos fases, para mantener el mismo error probable con una probabilidad de 0,90.

6. REFERENCIAS BIBLIOGRAFICAS

- COCHRAN, W.G.: Sampling Techniques; 2a. ed., NY, Wiley, 1963.
- HANSEN, M.H.; HURWITZ, W.N.; MADOW, W. Sample Survey Methods and Teory, NY Wiley, 1966.
- CAPPELLETTI, C.A.; MENDONCA, F.J.; LEE, D.C.L.; SHINABUKURO, Y.E.; Estimacao da Area Ocupada con Caña-de-Azúcar na Grande Regiao Canaveira de Piracicaba Utilizandose Imagens Landsat e Fotografias Aereas a Traves da Simulacao de un Sistema de Amostragen em Duas Faces; Anais do II Simposio Brasileiro de Sensoramiento Remoto, Brasilia, 1982a.
- CAPPELLETTI, C.A.; MENDONCA, F.J.; LEE, D.C.L.; SHINABUKURO, Y.E.; Estimation of the sugar-cane cultivated area from Landsat imagens, using two phase sampling methods; XVI Int. Symposium on Remote Sensing of Enviroment Vol. 2, p. 1055-1057, Michigan, 1978. b.