

Publicación LARS 082283

Caso de Estudio; Región Nor-Andina,
Provincias de Imbabura y Pichincha,
Ecuador

por

José A. Valdés A. y Diego Fabián Lozano G.

Para usarse con:

Taller Práctico en el

Análisis Numérico de

Datos de Imágenes Multiespectrales.

(Publicación LARS 050583)

Prefacio

Este caso de estudio, está diseñado para acompañar la Publicación LARS 050583, "Taller Práctico en el Análisis de Datos de Imágenes Multiespectrales" por Luis A. Bartolucci, Fabián Lozano y James C. Tilton.

Además de esta publicación, el estudiante requiere de los siguientes materiales para llevar a cabo las actividades aquí descritas:

1. Un listado de computadora conteniendo todas las impresiones de este análisis.
2. Fotografías aéreas en blanco y negro del área de estudio. Un paquete de fotografías apropiado, fue obtenido en Febrero de 1977 y Noviembre de 1978, a una escala aproximada de 1:65,000.

Es conveniente contar con una impresión en tonos de gris de la zona de estudio. Aunque esta última se encuentra ya en los listados de computadora, es difícil el tener una idea del área en su totalidad, a menos que se encuentre unida.

CASO DE ESTUDIO; REGION NOR-ANDINA,
PROVINCIAS DE IMBAMBURA Y PICHINCHA,
REPUBLICA DEL ECUADOR.

PARTE I - INTRODUCCION

El caso de estudio analizado en este taller práctico, es un análisis realizado con el paquete de programas LARSYS con datos Landsat colectados en 1979 sobre las provincias de Imbabura y Pichincha en la porción Nor-Andina del Ecuador.

El área abarca aproximadamente 40,000 hectáreas. Es una zona volcánica con un alto grado de relieve, encontrándose así la presencia de sombras topográficas y variaciones altitudinales en la vegetación. Así mismo, se localizan dos asentamientos humanos de importancia, la ciudad de Cayambe hacia el Sur y la ciudad de San Pablo del Lago al Este de la Laguna del mismo nombre. La figura I-2, muestra un mapa del Ecuador con la localización de la zona estudio.

El uso de la tierra en la zona es fundamentalmente agricultura del tipo intensivo, siendo muy notoria la alta multiparcelaridad de la tierra, encontrándose campos agrícolas tanto en las partes bajas de los valles como en las faldas de las montañas. Durante el análisis, se creará un mapa de cobertura del suelo, de un área aproximada de 40,000 hectáreas, o 400 kilómetros cuadrados. En el análisis se hará distinción entre cuerpos de agua, áreas agrícolas con diferentes condiciones del cultivo, bosques, suelos o rocas desnudas, nubes y sombras de nubes.

Una forma de obtener información acerca de un conjunto de datos digitales almacenados en una cinta compatible con computadora (CCT), es utilizando el procesador LARSYS llamado *IDPRINT. Este procesador, imprime el identificador de una imagen multiespectral almacenada en cinta.

Estudie el listado de computadora del *IDPRINT (Vea páginas 2 a 4 de los listados de computadora) y observe:

1. El número de corrida para este conjunto de datos. Cada conjunto de datos tiene un número de corrida único, con un total de ocho dígitos, donde los dos primeros indican el año en el cual los datos fueron colectados.
2. La fecha y hora del día cuando fueron tomados los datos. (No confundir la fecha de la obtención de los datos con la fecha de reformateo de los mismos).
3. El número de líneas y de columnas muestreadas, en este

conjunto de datos.

4. El intervalo de banda espectral para cada uno de los cuatro canales de datos. ¿En que porción del espectro electromagnético se presentan estas bandas?

5. El cambio de las bandas de longitud de onda 4, 5, 6 y 7 (Como las identifica el Centro de Datos EROS), a los canales de datos 1, 2, 3 y 4.

6. Los valores de pulso de calibración. Históricamente, los sistemas de barrido aereotransportados grababan tres señales de calibración para cada canal. En el caso del satélite Landsat, solo se utilizan dos fuentes de calibración; sin embargo, los valores mostrados como C0 y C1 pueden ser utilizados para convertir los datos en unidades radiométricas. En este caso, C0 tiene un valor de 0.0 para cada banda, y C1 tiene un valor que indica la máxima radiancia para cada canal en $\text{mwatts}/(\text{cm})^2(\text{sr})$. Los valores son los mismos para todos los datos del Landsat-3.

Para conocer los valores numéricos asociados con puntos conocidos en la imagen, se utiliza otro procesador LARSYS, denominado *TRANSFERDATA. En las paginas 5 a 10 de los listados de computadora, se muestra la salida de *TRANSFERDATA de cinco elementos individuales seleccionados para representar diferentes materiales en el terreno.

Note, que cada elemento está identificado por su línea y columna, y tiene cuatro valores de datos asociados a él. Este conjunto de valores es denominado "vector de los datos"; en el cual se indica la cantidad relativa de energía que regresa de la superficie de la tierra al sensor, en otras palabras, la brillantez con la cual esa parte de la superficie terrestre es medida por el sistema del barredor en cada banda de longitud de onda. El rango de los valores de los datos va de 0 a 127 en cada canal. (Normalmente la respuesta del canal 4 del Landsat, procesada por el Centro de Datos EROS, tiene un rango de 0 a 63; sin embargo, en este y en subsecuentes pasos del análisis, el canal 4 ha sido expandido para que sea equivalente a los otros tres canales en su rango dinámico.) Un valor de cero en los datos, corresponde a un objeto que no refleja mucha energía en esa determinada longitud de onda, esto significa que no existio medicion de energía por el detector para ese determinado elemento; por otro lado un valor de 127, corresponde a un objeto que refleja gran cantidad de energía, e indica saturación del detector del barredor.

Los cuatro vectores de datos impresos por *TRANSFERDATA (Páginas 5 a 10), son representativos de los miles de vectores de datos dentro de la escena. Mas tarde, cuando la computadora clasifique

los datos, se asignará cada elemento a una clase con base en estos valores numéricos.

Los valores numéricos de una línea o columna de datos se pueden obtener con los procesadores *COLUMNGRAPH, y *LINEGRAPH respectivamente. En las páginas 11 a 16 de los listados de computadora se muestran los resultados obtenidos con estos procesadores.

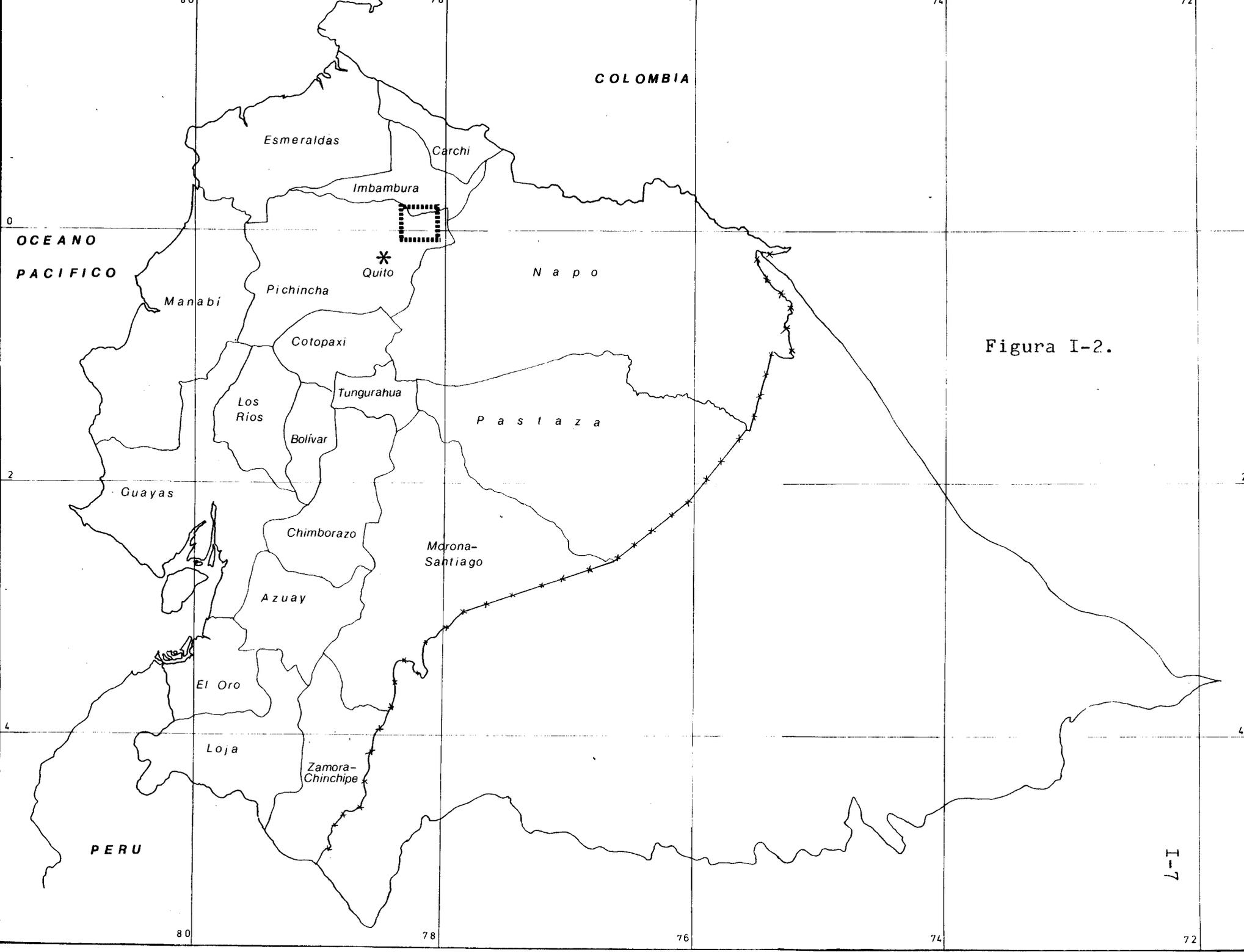
El siguiente vector de datos de una imagen diferente, es un elemento conocido, que representa a bosques.

		CANALES			
		1	2	3	4
LIN	COL	DATOS			
115	326	26.0	15.0	58.0	62.0

Compare estos valores con los que se muestran en las páginas 5 a 10 de los listados. ¿En que página se muestran los valores de los datos que probablemente representan también al bosque? ¿Cómo se llegó a esta decisión? Si sólo fuera posible utilizar el canal 3 de los datos, ¿se podría llegar a esta decisión fácilmente? ¿Como sería, si sólo pudiera emplear los canales 2 y 3?

De acuerdo a su conocimiento de las características espectrales de los materiales presentes en la superficie terrestre, ¿Puede usted indicar que representa cada uno de los demás elementos presentes en las páginas 5 a 10 de los listados?

Analizando los resultados de los procesadores *COLUMNGRAPH y *LINEGRAPH, en las páginas 11 a 16 de los listados. ¿Cuántos tipos de cobertura puede usted identificar, y a que clase pertenecen?



COLOMBIA

Esmeraldas

Carchi

Imbabura

OCEANO
PACIFICO

*
Quito

N a p o

Manabi

Pichincha

Cotopaxi

Los
Rios

Tungurahua

P a s t a z a

Bolivar

Guayas

Chimborazo

Morona-
Santiago

Azuay

El Oro

Loja

Zamora-
Chinchipe

PERU

Figura I-2.

I-7

PARTE II - SELECCION DE DATOS DEL BARREDOR MULTIESPECTRAL (MSS),
CORRELACION CON DATOS DE REFERENCIA Y SELECCION
DE MUESTRAS DE ENTRENAMIENTO.

Esta parte del caso de estudio, muestra una forma posible en la cual se puede realizar el primer paso en el análisis de datos del barredor multiespectral (MSS). Un juego de datos disponible para esta región, fue colectado el 13 de Febrero de 1979. El analista debe pues, decidir cuales son las características terrestres que pueden ser identificadas en esta época del año con estos datos.

ANALISIS DE LA CALIDAD DE LOS DATOS.

El primer paso que debe realizarse con los datos, es el examinarlos para evaluar la calidad de los mismos, tanto en la cantidad de cobertura de nubes, como en las distorsiones geométricas. Para este propósito, se utilizan normalmente las imágenes en blanco y negro y color compuesto producidas por el Centro de Datos EROS, pero en nuestro caso, las primeras no estaban disponibles. Sin embargo, se puede utilizar un método alternativo; en el cual se deben de producir impresiones de computadora de la imagen completa en un canal visible y en un infrarrojo (a una tasa de muestreo de 1:5), o desplegarla en un sistema de rayos catódicos. Esto dará una visión general de los datos, y permitirá el determinar las líneas y columnas del área de interés. Utilizando estas coordenadas (líneas y columnas), se creará una imagen en la impresora electrostática de los cuatro canales del área de estudio. Estas se muestran en las figuras II-4 a II-7 en este manual.

Para producir estas impresiones se utilizan patrones de puntos de diferentes tonos de gris, que representan el tono al cual cada elemento es asignado. Esta impresora crea una imagen similar a una fotografía (en apariencia), mostrándose así, las características espaciales de los datos. Es posible también, el imprimir mapas en impresoras de línea, utilizando símbolos de diferentes tonalidades de gris. Esto crea un mapa en el cual cada elemento puede ser analizado individualmente. Ver figura II-8 como ejemplo de un mapa en tonalidades de gris de una impresora de línea.

El proceso utilizado para generar los dos tipos de mapas (impresora de línea e impresora electrostática), es una combinación de intervalos de densidad e incremento de contraste. Primero, se determina el rango total de los datos en cada canal; los datos a utilizarse, son especificados por el analista en las tarjetas de control, y las coordenadas se muestran bajo el encabezado "Histogram Block(s)" Por ejemplo, si se observa en la tabla II-1, se encontrarán los datos de la imagen, tales como el número de corrida, línea de vuelo, etc. A continuación se encuentra la banda espectral a la que corresponde esa impresión. En este caso, los datos corresponden al canal 1, de la línea 900 a la 1400, y las columnas

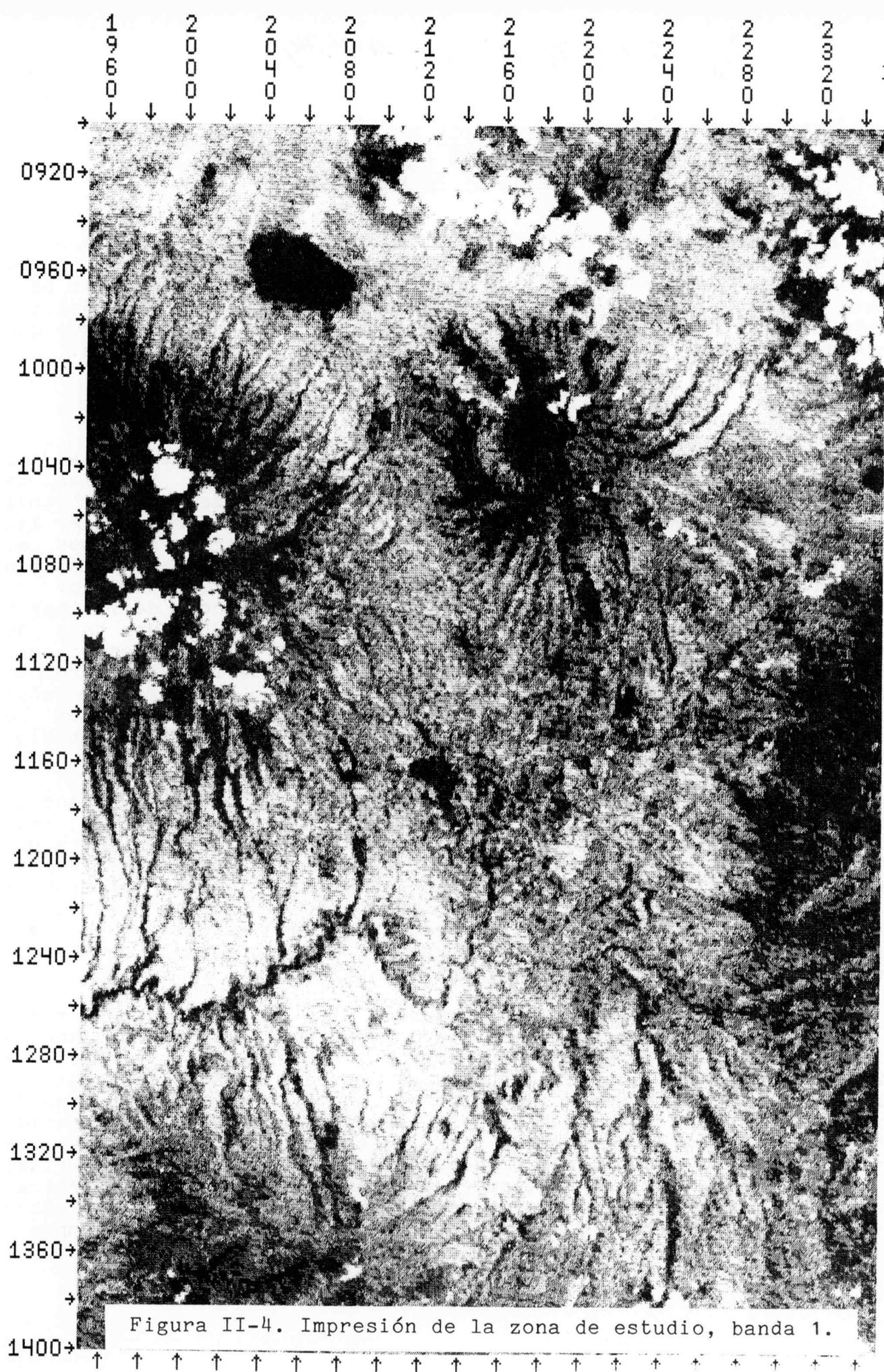


Figura II-4. Impresión de la zona de estudio, banda 1.

RUN NUMBER... 79011301
 FLIGHT LINE... 30345-14454 ECUA
 DATA TAPE/FILE NUMBER.. 749/ 1
 REFORMATTING DATE. JAN 31, 1983

DATE DATA TAKEN... FEB 13, 1979
 TIME DATA TAKEN..... 1445 HOURS
 PLATFORM ALTITUDE... 3062000 FEET
 GROUND HEADING..... 180 DEGREES

II-13

CHANNEL 1 SPECTRAL BAND 0.50 TO 0.60 MICROMETERS CALIBRATION CODE= 1 C0 = .4000E-01

HISTOGRAM BLOCK(S)

RUN NUMBER	LINES	COLUMNS	CALIBRATION CODE
79011301	(900, 1400, 10)	(1950, 2350, 10)	1

THE DATA RANGES ASSIGNED TO THE GRAY LEVELS ARE

LOWER LIMIT	* UPPER LIMIT	LEVEL NUMBER	SAMPLE COUNT	PER CENT OF TOTAL SAMPLE
<	18.5	1	76	3.6
18.5	20.5	2	181	8.7
18.5	20.5	2	181	8.7
20.5	22.5	3	240	11.5
22.5	22.5	4	0	0.0
22.5	22.5	5	0	0.0
22.5	24.5	6	377	18.0
24.5	24.5	7	0	0.0
24.5	24.5	8	0	0.0
24.5	26.5	9	433	20.7
26.5	26.5	10	0	0.0
26.5	26.5	11	0	0.0
26.5	28.5	12	373	17.8
28.5	28.5	13	0	0.0
28.5	30.5	14	188	9.0
30.5	32.5	15	104	5.0
32.5	>	16	119	5.7

THE TOTAL NUMBER OF SAMPLE POINTS... 2091
 THE AVERAGE NUMBER OF SAMPLE POINTS ASSIGNED PER GRAY LEVEL... 130.687
 THE STANDARD DEVIATION OF THE NUMBER OF SAMPLE POINTS PER GRAY LEVEL... 153.175

Tabla II-1.

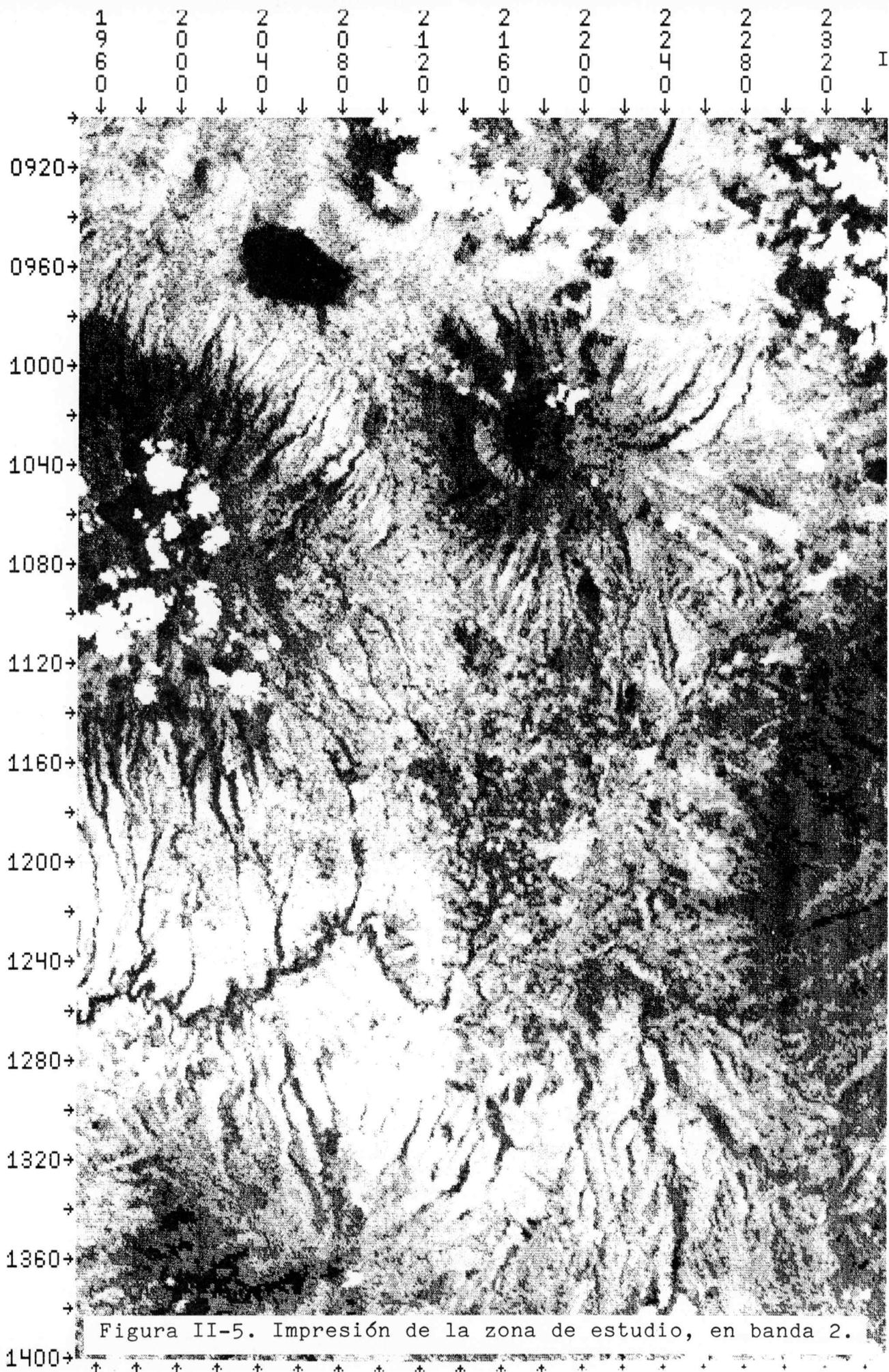


Figura II-5. Impresión de la zona de estudio, en banda 2.

RUN NUMBER..... 79011301

DATE DATA TAKEN... FEB 13,1979

FLIGHT LINE... 30345-14454 ECUA

TIME DATA TAKEN..... 1445 HOURS

DATA TAPE/FILE NUMBER.. 749/ 1

PLATFORM ALTITUDE..3062000 FEET

REFORMATTING DATE. JAN 31,1983

GROUND HEADING..... 180 DEGREES

CHANNEL 2 SPECTRAL BAND 0.60 TO 0.70 MICROMETERS CALIBRATION CODE= 1 CO = .3000E-01

HISTOGRAM BLOCK(S)

RUN NUMBER	LINES	COLUMNS	CALIBRATION CODE
79011301	(900, 1400, 10)	(1950, 2350, 10)	1

THE DATA RANGES ASSIGNED TO THE GRAY LEVELS ARE

LOWER LIMIT	UPPER LIMIT	LEVEL NUMBER	SAMPLE COUNT	PER CENT OF TOTAL SAMPLE
<	18.5	1	108	5.2
18.5	20.5	2	102	4.9
20.5	22.5	3	161	7.7
22.5	24.5	4	260	12.4
24.5	24.5	5	0	0.0
24.5	26.5	6	238	11.4
26.5	26.5	7	0	0.0
26.5	28.5	8	298	14.3
28.5	28.5	9	0	0.0
28.5	30.5	10	240	11.5
30.5	30.5	11	0	0.0
30.5	32.5	12	198	9.5
32.5	34.5	13	132	6.3
34.5	36.5	14	98	4.7
36.5	40.5	15	139	6.6
40.5	>	16	117	5.6

THE TOTAL NUMBER OF SAMPLE POINTS... 2091

THE AVERAGE NUMBER OF SAMPLE POINTS ASSIGNED PER GRAY LEVEL... 130.687

THE STANDARD DEVIATION OF THE NUMBER OF SAMPLE POINTS PER GRAY LEVEL... 98.187

Tabla II-2.

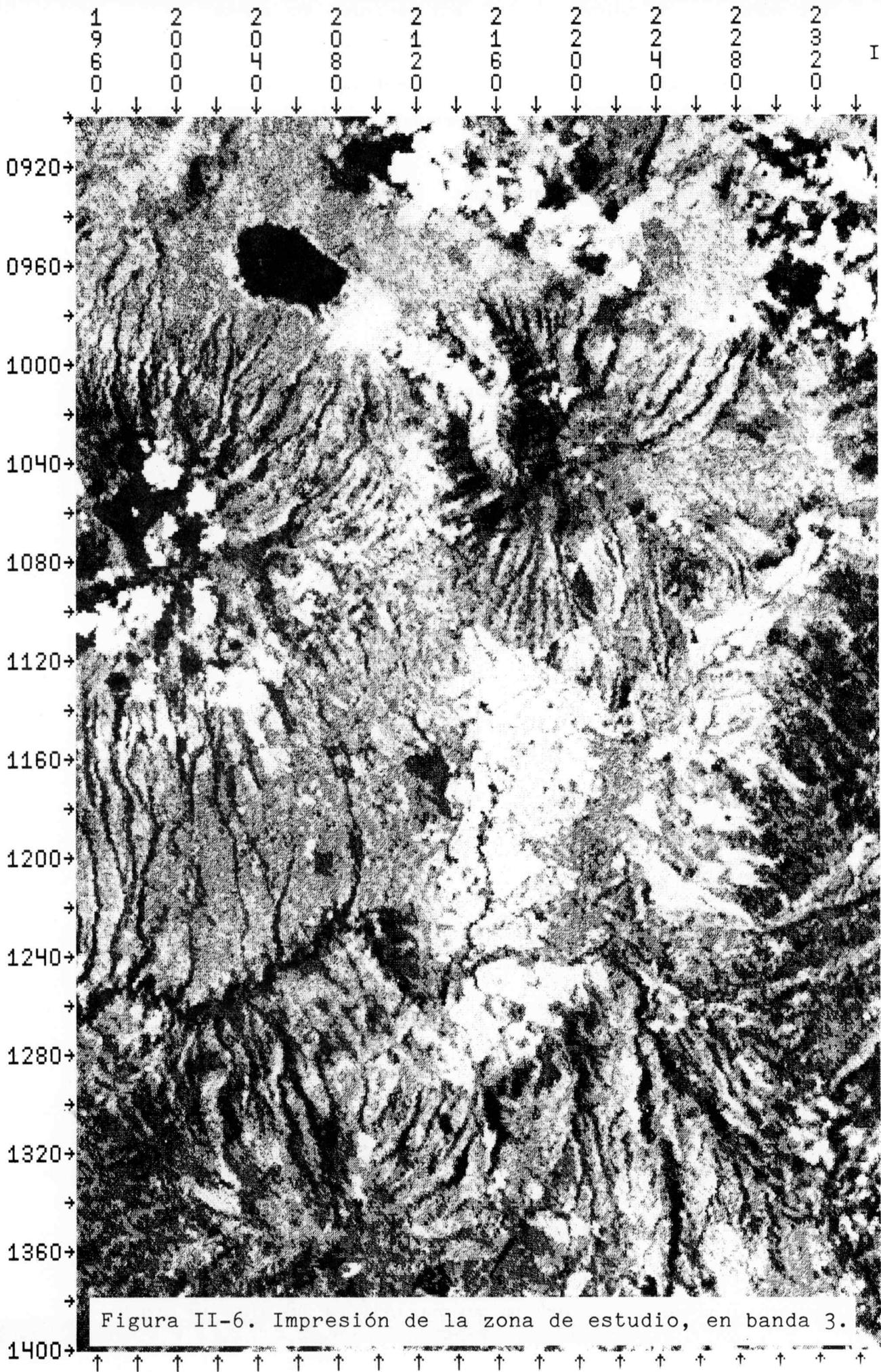


Figura II-6. Impresión de la zona de estudio, en banda 3.

RUN NUMBER..... 79011301
 FLIGHT LINE... 30345-14454 ECUA
 DATA TAPE/FILE NUMBER.. 749/ 1
 REFORMATTING DATE. JAN 31.1983

DATE DATA TAKEN... FEB 13,1979
 TIME DATA TAKEN..... 1445 HOURS
 PLATFORM ALTITUDE..3062000 FEET
 GROUND HEADING..... 180 DEGREES

II-17

CHANNEL 3 SPECTRAL BAND 0.70 TO 0.80 MICROMETERS CALIBRATION CODE= 1 CO = .3000E-01

HISTOGRAM BLOCK(S)

RUN NUMBER	LINES	COLUMNS	CALIBRATION CODE
79011301	(900, 1400, 10)	(1950, 2350, 10)	1

THE DATA RANGES ASSIGNED TO THE GRAY LEVELS ARE

LOWER LIMIT	UPPER LIMIT	LEVEL NUMBER	SAMPLE COUNT	PER CENT OF TOTAL SAMPLE
<	31.5	1	142	6.8
31.5	35.5	2	151	7.2
35.5	37.5	3	132	6.3
37.5	39.5	4	145	6.9
39.5	39.5	5	0	0.0
39.5	41.5	6	220	10.5
41.5	43.5	7	226	10.8
43.5	43.5	8	0	0.0
43.5	45.5	9	240	11.5
45.5	45.5	10	0	0.0
45.5	47.5	11	215	10.3
47.5	49.5	12	175	8.4
49.5	49.5	13	0	0.0
49.5	53.5	14	180	8.6
53.5	59.5	15	133	6.4
59.5	>	16	132	6.3

THE TOTAL NUMBER OF SAMPLE POINTS... 2091
 THE AVERAGE NUMBER OF SAMPLE POINTS ASSIGNED PER GRAY LEVEL... 130.687
 THE STANDARD DEVIATION OF THE NUMBER OF SAMPLE POINTS PER GRAY LEVEL... 85.473

Tabla II-3.

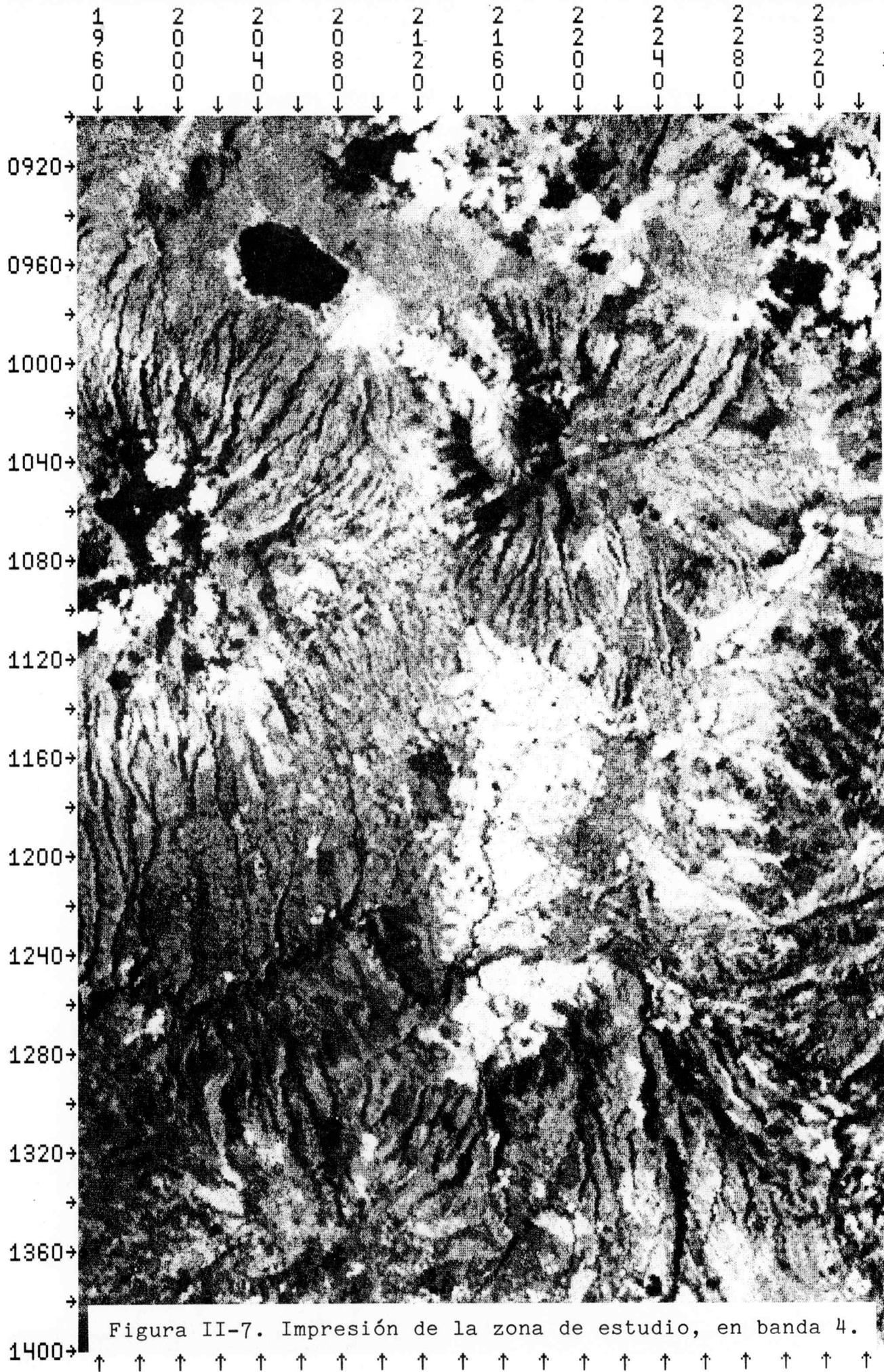


Figura II-7. Impresión de la zona de estudio, en banda 4.

RUN NUMBER..... 79011301
 FLIGHT LINE... 30345-14454 ECUA
 DATA TAPE/FILE NUMBER.. 749/ 1
 REFORMATTING DATE. JAN 31,1983

DATE DATA TAKEN... FEB 13,1979
 TIME DATA TAKEN..... 1445 HOURS
 PLATFORM ALTITUDE..3062000 FEET
 GROUND HEADING..... 180 DEGREES

II-19

CHANNEL 4 SPECTRAL BAND 0.80 TO 1.10 MICROMETERS CALIBRATION CODE= 4 C0 = .0 C1 = 127.0

HISTOGRAM BLOCK(S)

RUN NUMBER	LINES	COLUMNS	CALIBRATION CODE
79011301	(900, 1400, 10)	(1950, 2350, 10)	4

THE DATA RANGES ASSIGNED TO THE GRAY LEVELS ARE

LOWER LIMIT	UPPER LIMIT	LEVEL NUMBER	SAMPLE COUNT	PER CENT OF TOTAL SAMPLE
<	32.5	1	139	6.6
32.5	36.5	2	130	6.2
36.5	38.5	3	94	4.5
38.5	40.5	4	128	6.1
40.5	42.5	5	160	7.7
42.5	44.5	6	184	8.8
44.5	44.5	7	0	0.0
44.5	46.5	8	178	8.5
46.5	48.5	9	184	8.8
48.5	50.5	10	156	7.5
50.5	52.5	11	159	7.6
52.5	52.5	12	0	0.0
52.5	56.5	13	212	10.1
56.5	60.5	14	114	5.5
60.5	68.5	15	126	6.0
68.5	>	16	127	6.1

THE TOTAL NUMBER OF SAMPLE POINTS... 2091
 THE AVERAGE NUMBER OF SAMPLE POINTS ASSIGNED PER GRAY LEVEL... 130.687
 THE STANDARD DEVIATION OF THE NUMBER OF SAMPLE POINTS PER GRAY LEVEL... 59.291

Tabla II-4.

1950 a la 2350. El intervalo de muestreo entre líneas y columnas es cada 10 elementos, lo cual significa, que solo se emplearon datos de una línea sí y diez no, y de una columna sí y de diez no, para calcular el histograma, lo cual corresponde al 1 % de los datos. Un número uno en la tasa de muestreo, significaría que todos y cada uno de los puntos de los datos, fueron utilizados para calcular el histograma. Este intervalo es determinado por el tamaño de la imagen, y la complejidad de la escena. Un intervalo mayor, puede ser utilizado para áreas grandes no muy complejas, y un intervalo menor puede ser utilizado para una imagen de mayor complejidad, pero del mismo tamaño. Reduciendo el muestreo, se ahorra tiempo de cómputo.

Una vez que el rango dinámico de los datos es determinado por el procesador, los datos son convertidos a un histograma y divididos en 16 niveles. La impresión de la tabla II-1 muestra los límites superior e inferior de cada subconjunto o intervalo de datos, el número del nivel (de 1 a 16), el número de elementos asignado a cada intervalo, y el porcentaje del área total representada por esos puntos en el histograma. Por ejemplo: véase en la tabla bajo el encabezado: "THE DATA RANGES ASSIGNED TO THE GRAY LEVELS ARE." La cuarta línea en esta tabla, indica que todos los elementos en el área del histograma con valores de datos de 20.5 a 22.5 en el canal 1, son asignados al nivel de gris número 3 y que se encuentran presentes 240 elementos, representando el 11.5% del total de puntos utilizados en el histograma. Todos los elementos asignados al nivel de gris número tres, son desplegados por la impresora de matriz con el tono de gris mostrado en el tercer cuadro de izquierda a derecha. (Nótese que los patrones de medio tono van del nivel 16 a la izquierda, hacia el nivel 1 a la derecha). El resto de la tabla especifica los límites para determinar cual tono de gris es utilizado para representar cualquier valor de datos en la escena. Estos límites son calculados para optimizar el uso de los 16 niveles de tonos de gris disponibles en la impresora electrostática. El mismo procedimiento se utiliza para el mapa de tonos de gris producido por la impresora de línea; la única diferencia es que en lugar de usar patrones de puntos, se utilizan caracteres alfanuméricos para obtener diferentes tonos de gris, siendo M, Z y X los mas oscuros, y asignados a las respuestas espectrales mas bajas; y los tonos mas claros (.,-,blanco), son asignados a los intervalos con los valores mas altos.

De la información contenida en las tablas II-1 a II-4, identifique que imagen representa los valores de los datos de el canal 1, canal 2, canal 3 y canal 4.

Examine las imágenes en tonos de gris. ¿Por qué las imágenes del canal 1 y la del canal 2 son muy semejantes? Por qué las imágenes de los canales 3 y 4 son similares entre sí,

pero diferentes de las de los canales 1 y 2?

¿Que puede identificar en estas imágenes?

¿Existe algun problema de calidad aparente en cualquiera de las imágenes? ¿Existen problemas con cobertura de nubes en cualquiera de las imágenes?

LOCALIZACION DEL AREA DE ESTUDIO Y
CORRELACION CON LOS DATOS DE REFERENCIA

Este juego de datos ya ha sido corregido geométricamente. Esto se puede verificar observando el encabezado "Ground Heading" (impreso por el procesador*IDPRINT donde se observa que es de 180 grados. Un conjunto de datos LANDSAT no corregido, tendría una orientación de aproximadamente 190 grados. Los valores de los datos, han sido además, modificados en su tamaño (reescalados), para que la impresión de los datos de la imagen tuviera una escala de 1:25,000; la cual puede ser convenientemente utilizada con mapas topográficos escala 1:50,000.

Las imágenes de la impresora electrostática, que se observaron anteriormente, fueron realizadas a partir de los datos Landsat corregidos geométricamente. Si se observa en las tablas II-1 a II-4, se verá que el analista seleccionó el desplegar de la línea 900 a la 1400, y de la columna 1950 a la 2350 únicamente, con un intervalo de uno. Esta es el área de la cual se dispone de fotografías aéreas, y que corresponde a aproximadamente al 2.2% de la imagen Landsat.

El instructor le proporcionará los siguientes materiales, como datos de referencia:

- * Fotografías aéreas blanco y negro del área de estudio a una escala aproximada de 1:65,000 obtenidas en Febrero de 1977 y Noviembre de 1978.

Asocie los datos de referencia con las imágenes de la impresora electrostática. Haciendo uso de las imágenes de los canales 1 y 4, encuentre y delimite dos o tres ejemplos de cada cobertura de interés: agricultura, bosque, agua, suelo desnudo, nubes y sombras. Algunas de estas coberturas son mas aparentes en una imagen que en otra. (Nota: No marque nada en las imágenes del canal 2 y 3 por el momento, ya que se hará

uso de ellas posteriormente).

SELECCION DE LAS MUESTRAS DE ENTRENAMIENTO.

Existen tres métodos de selección de las muestras de entrenamiento que son utilizadas para adiestrar al clasificador, estos son: El método supervisado, el método no supervisado, y el método híbrido, en el cual se combinan los dos primeros. Para este caso de estudio, haremos uso de este último.

Recuerde las dos reglas para seleccionar las áreas candidatas de entrenamiento: (1) Cada muestra seleccionada debe de incluir mas de un tipo de cobertura, y (2) Cada tipo de cobertura, debe aparecer por lo menos en un área de entrenamiento. Para los cuatro canales de datos que se están utilizando, se debe seleccionar de dos a cuatro áreas de entrenamiento de tamaño variable entre 50 líneas por 50 columnas a 100 líneas por 100 columnas.

Seleccione tres áreas de entrenamiento, las cuales sean representativas de la escena. Utilice las fotografías junto con las imágenes en tonos de gris de la escena y las instrucciones dadas anteriormente. Asegúrese de que cada cobertura de interés (agricultura, agua, suelos desnudos y bosques) este incluida en por lo menos una de las áreas de entrenamiento, y que estas áreas de entrenamiento se encuentren distribuidas en toda la escena.

Delimite las áreas seleccionadas en la imagen en tono de gris del canal 2, y anote las coordenadas de las líneas y columnas en función de la primera línea, última línea, primera columna y última columna. Justifique la selección de estas áreas de entrenamiento.

PARTE III - DEFINICION ESTADISTICA DE LAS CLASES DE ENTRENAMIENTO

Esta parte del caso de estudio, presenta una forma posible de definir estadísticamente las clases espectrales de entrenamiento para el área de estudio.

AGRUPAMIENTO DE LAS MUESTRAS DE ENTRENAMIENTO.

Es muy posible que diferentes analistas escojan diferentes muestras de datos para usarlos en el desarrollo de las estadísticas de entrenamiento. Es factible que se elijan diferentes áreas de entrenamiento por cada uno de los analistas, pero todas y cada una de ellas serán igualmente válidas. Para este caso de estudio se han seleccionado intencionalmente algunas áreas no muy obvias con el propósito de demostrar los diferentes métodos. Las áreas seleccionadas se encuentran en la página 19 de los listados de computadora, y se encuentran delimitadas en los listados que comienzan en la pagina 21. La figura III-17, muestra la posición en la imagen de las muestras de entrenamiento.

Con el objeto de agrupar los datos de entrenamiento, el procesador *CLUSTER fue corrido tres veces, esto fue necesario debido a la limitante que tiene este procesador en LARSYS, ya que solo puede analizar 40,000 elementos en cada proceso. En la práctica, cuando se realiza el agrupamiento con cuatro canales, solamente un máximo de 10,000 elementos son procesados durante una sola corrida.

Agrupando la primer área de entrenamiento -- Con el objeto de de representar la amplia variabilidad en el área de estudio en un primer muestreo de datos de 100 por 100 (10,000 elementos), se decidió el escoger cinco áreas pequeñas que en su conjunto contienen menos de 10,000 elementos (9079). El procesador *CLUSTER, agrupa los datos sin tener en cuenta su localización en el terreno, por lo tanto el presente método, se considera válido. En este caso, el número máximo de clases fue de 20, considerando que se encuentran nueve clases informacionales, en la página 53 de los listados de computadora la frase: "OPTION MAXCLAS" (20), "CONV" (98.5), indica primero, que los datos con los que se ha alimentado al computador, deben de ser agrupados en 20 clases, basándose en las respuestas espectrales de los cuatro canales, y segundo, que el procesador continúe realizando agrupamientos iterativamente hasta que el 98.5 por ciento de los elementos no sean asignados a una diferente clase en la siguiente iteración. Es una práctica común el asignar el número de grupos como el doble de las clases de información mas dos o tres grupos extras; lo cual nos permite una mejor definición de las zonas de transición entre las clases de información y de varias

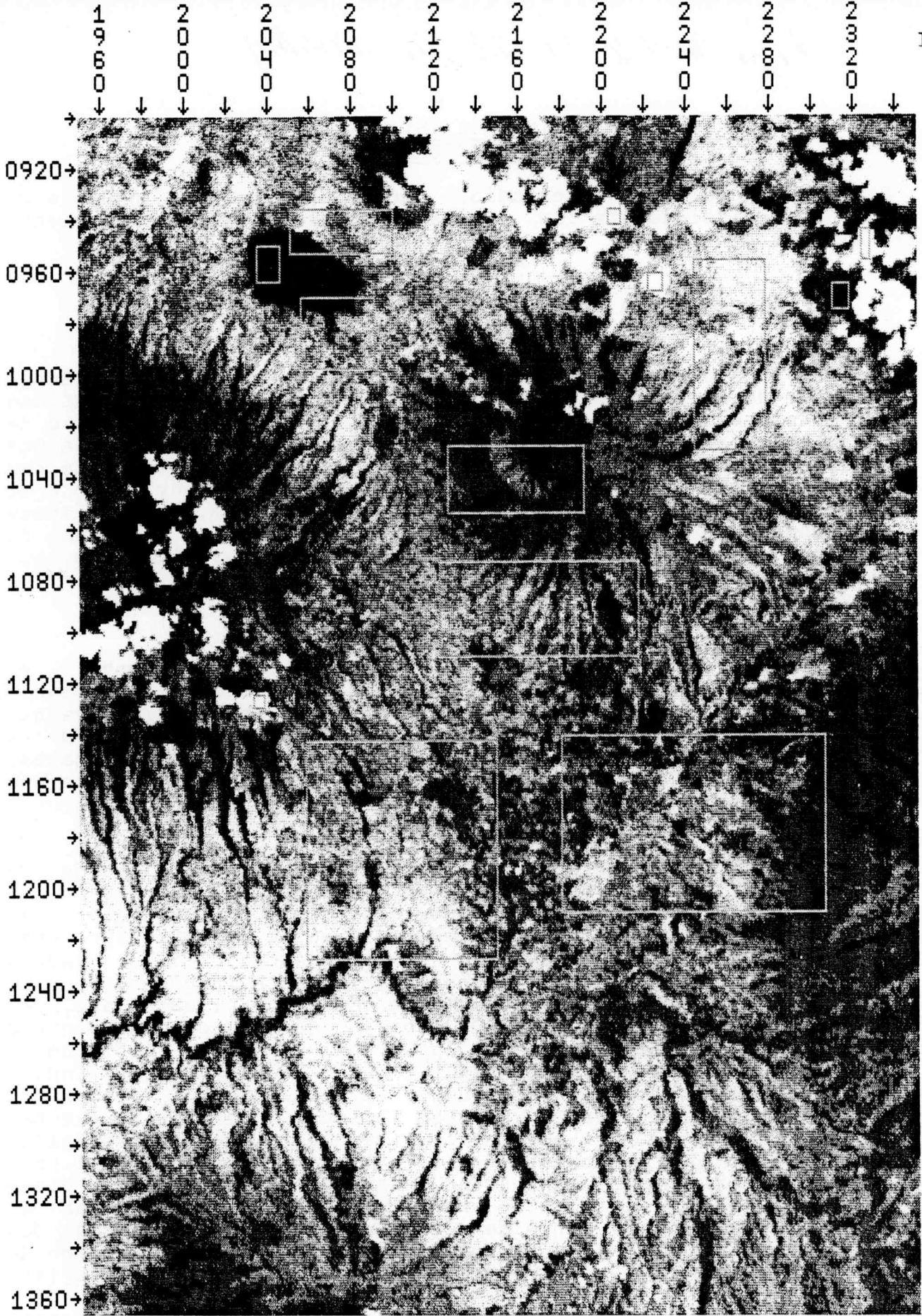


Figura III-17. Impresión de la zona de estudio, con las áreas de agrupamiento delimitadas.



FIELD CARDS ADDED TO FIELD BOUNDARY STORAGE

FIELD NUMBER	RUN NUMBER	FIELD DESIGNATION	FIRST LINE	LAST LINE	LINE INTERVAL	FIRST SAMPLE	LAST SAMPLE	SAMPLE INTERVAL	FIELD TYPE	OTHER INFORMATION
1 *	79011301		937	952	1	2052	2099	1		
2 *	79011301		971	997	1	2057	2090	1		
3 *	79011301		1029	1053	1	2127	2190	1		
4 *	79011301		1074	1109	1	2123	2216	1		
5 *	79011301		957	1029	1	2245	2277	1		
6 *	79011301		1144	1227	1	2059	2148	1		
7 *	79011301		1142	1209	1	2181	2305	1		
8 *	79011301		951	963	1	2036	2045	1		
9 *	79011301		966	974	1	2311	2317	1		
10 *	79011301		937	941	1	2204	2208	1		
11 *	79011301		945	955	1	2325	2327	1		
12 *	79011301		1126	1129	1	2034	2037	1		
13 *	79011301		962	967	1	2223	2228	1		

* - INDICATES TRAINING FIELDS

Tabla III-1. Líneas y columnas de los Campos de agrupamiento y clases supervisadas.

subclases dentro de cada una de ellas.

En la página 55 se muestran los valores de las medias de cada clase en cada canal, el número de puntos para cada clase y la varianzas de cada una de ellas. En general, los primeros grupos tienen los valores de media mas alto que los últimos, y la tendencia general en una sequencia de lo mas brillante (valores de media altos), a lo mas obscuro (valores de media bajos). Las varianzas indican la separación o dispersión de los datos en cada canal, y sirve como una guía para definir la pureza espectral de la clase.

Los mapas de agrupamiento que se obtuvieron al procesar el área uno, se muestran en las páginas 56 a 61. Cuando se comparan estos con las fotografías aéreas, se determina facilmente que la clase 20 del agrupamiento número uno en la zona de San Pablo del Lago (designado por la letra W en el mapa), representa agua. Esta clase se encuentra ubicada en la esquina inferior izquierda de la página 56, la cual se correlaciona facilmente con la laguna en la fotografía aérea. Las asociaciones menos obvias de clases de agrupamiento con clases informacionales, serán presentadas en la siguiente sección.

En las página 62 a la 101, se muestran los histogramas de los valores de los datos de cada canal para cada una de las 20 clases de agrupamiento, estos es, la distribución canal por canal de todos los valores de los datos de aquellos elementos que fueron asignados dentro de cada clase por el algoritmo de agrupamiento. Debido a que el algoritmo de clasificación que se utilizará mas tarde, asume que cada clase de entrenamiento está representada por una función de densidad de tipo Gaussiana (Normal), se debe analizar cada histograma y verificar si la distribución de cada clase de agrupamiento es similar a la función de densidad Gaussiana. Note la forma no Gaussiana del histograma de los canales uno y dos de la clase uno (página 62). Esto representa un problema potencial. Si se observa el tamaño relativo de esta clase, mostrado en la página 55, nótese que esta clase de 163 puntos, es de las mas pequeñas. De hecho, el tener histogramas bi-modales, no es extraño para clases pequeñas. A pesar de que esta desviación no es tan grande como para que sea necesario repetir todo el proceso, debemos sin embargo, considerar el valor de varianza en cada canal y el tamaño de la clase, al momento en el que se realice la decisión acerca de cuales clases de entrenamiento estarán presentes al final. Note tambien, la varianza tan grande presente en la clase dos del canal dos en la página 55, observe la dispersión de los datos en los histogramas, y note que el rango de valores varía de clase a clase y de canal a canal. Las clases con histogramas angostos, presentan varianzas pequeñas.

Agrupando la segunda área de entrenamiento -- La segunda área analizada (Tabacundo), es un rectángulo de 84 por 90 elementos, localizado en la margen inferior izquierda de la zona de estudio.

Los listados de computadora, dan el mismo tipo de información que para la primer área de entrenamiento: media de las clases, varianzas, números de puntos asignados a cada grupo y los histogramas de las clases. Esta información se presenta de las páginas 103 a la 138.

Agrupando la tercer área de entrenamiento -- El área grande localizada abajo a la derecha de la escena, presenta pocos rasgos en el terreno fácilmente identificables, pero en general contiene bastante variación. Con el objeto de representar las clases presentes en esta área, se seleccionó un área de 68 por 125 elementos. En esta ocasión el número total de clases fue de 18.

Es bastante complicado de correlacionar el mapa de agrupamiento con los datos de referencia debido a la gran variabilidad espectral de la zona, pero en un área carente de rasgos sobresalientes, como ésta, el proceso anterior sirve para representar adecuadamente la variabilidad espectral del área.

 Examine los listados del procesador *CLUSTER para las áreas de entrenamiento 1, 2 y 3 de las páginas 55 a 182.

1. Observe los valores de las medias para cada agrupamiento en las páginas 55, 105 y 141. Haciendo uso de la tabla III-2, trate de hacer una identificación general para la mayor cantidad de clases de agrupamiento que pueda (vegetación, suelo desnudo, agua) comparando los valores relativos en cada banda con las características de reflectancia espectral conocidas para las coberturas básicas de la superficie terrestre. A pesar de que los valores de respuesta mostrados, no han sido calibrados para facilitar las comparaciones banda a banda, los comportamientos generales se pueden observar.

2.- Examine las varianzas asociadas con cada grupo, y anote en la tabla cualquier valor excepcionalmente alto.

3. En los mapas de agrupamiento verifique si existen grupos que correspondan con rasgos en los datos de referencia, delimítelos en los mapas, y modifique la identidad de las clases si es necesario.

4. Examine los histogramas de los grupos. Note cualquier distribución no Gaussiana. Seleccione una clase con varianzas bajas, y compare sus histogramas con aquellos de clases con varianzas altas. Anote las clases no Gaussianas en la tabla.

5. Marque cualquier grupo con menos de 40 puntos.

CLASE	SAN PABLO DEL LAGO		TABACUNDO		CAYAMBE	
	Simb.	Nombre	Simb.	Nombre	Simb.	Nombre
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						

Tabla III-2.

ASOCIANDO LAS CLASES CANDIDATAS DE ENTRENAMIENTO
CON LAS CLASES INFORMACIONALES.

Ya que se ha generado el conjunto de clases candidatas de entrenamiento con el agrupamiento de todas las áreas, es necesario el identificar que clase informacional está representada por cada clase candidata de entrenamiento. Para realizar esta identificación, se hará uso de la impresión de los mapas de agrupamiento y las fotografías aéreas.

Recuerde que la correspondencia entre clases informacionales y clases candidatas de entrenamiento no es necesariamente de uno a uno. Por lo regular, mas de una clase de entrenamiento candidata está asociada con una clase informacional; ocasionalmente (es de desearse que rara vez), más de una clase informacional está asociada con una sola clase candidata de entrenamiento. Cuando ésto llega a ocurrir, se debe de otorgar a la clase nombres duales como: "Suelo desnudo y Cultivo emergiendo" o "Suelo desnudo y Autopista".

Asigne a cada clase de agrupamiento en el área de entrenamiento uno, un nombre. Utilice el mapa generado por la computadora de las páginas 56 a 61, junto con los datos de referencia y los valores de las medias de las clases, identifique lo mas acertadamente posible cada clase de agrupamiento en el área de entrenamiento. Escriba esta identificación en la tabla III-3, reemplazando las identificaciones generales realizadas previamente.

Estos datos Landsat, han sido previamente corregidos geoméricamente, lo cual facilita su correlación con las fotografías aéreas. Recuerde que la fotografía aérea, fue obtenida en Febrero de 1977 y Noviembre de 1978, el hecho de que sean de años anteriores a la informacion Landsat, tiene menos importancia que la época del año, en la cual existen 3 meses diferencia.

Si se dispone de tiempo, identifique las clases en las áreas de entrenamiento dos y tres.

CLASE	SAN PABLO DEL LAGO		TABACUNDO		CAYAMBE	
	Simb.	Nombre	Simb.	Nombre	Simb.	Nombre
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						

Tabla III-3.

AUMENTO DE LAS CLASES CANDIDATAS DE ENTRENAMIENTO.

Ahora que se han asociado las clases candidatas de entrenamiento con las clases informacionales se debe considerar si las muestras de entrenamiento son representativas de todos los tipos de cobertura en la escena.

Al observar el área de estudio en las impresiones electrostáticas, se encontraron nubes en la margen superior de la zona de estudio, así como en la orilla izquierda; las que producen sombras en el terreno, de las cuales una es fácilmente discriminable en la margen superior derecha del área. teniendo estas características presentes en la zona, se decidió hacer uso del método supervisado para el cálculo de las estadísticas de entrenamiento. Para evitar problemas de confusión entre agua, sombras de nube y otros cuerpos poco reflectivos, se decidió añadir una tercera clase de entrenamiento supervisada para la clase agua. Esto se realizó haciendo uso del procesador *STATISTICS.

Estas áreas se encuentran delimitadas también en la impresión de la figura III-17, y los resultados aparecen en las páginas 183 a la 215 de los listados de computadora.

Haciendo uso de las imágenes en tonos de gris (cualquiera de las figuras II-4 a II-7, o las impresiones alfa-numéricas de las páginas 31 a la 50), especifique las coordenadas de las líneas y columnas para un buen ejemplo de agua, nubes y sombras. Cuando se selecciona un área, téngase en mente el número mínimo de elementos que se necesitan para estimar las propiedades estadísticas de la muestra de entrenamiento. (Véase página III-7 del Taller Práctico).

Al igual que en la selección de muestras de entrenamiento anterior, es poco probable que dos analistas escogan las mismas áreas, aun así sus selecciones son igualmente válidas. Se ha seleccionado para la clase agua un campo ubicado en la laguna de San Pablo del Lago con 120 elementos. Para las nubes se escogieron cuatro campos de 25, 33, 36 y 16 totalizando 110 elementos. Para la clase sombra se delimitó un campo de 63 puntos. El procesador *STATISTICS simplemente extrae las estadísticas de la clase de donde provengan los datos, y estas pueden ser empleadas en la misma forma que las producidas por el *CLUSTER.

Examine los listados del procesador *STATISTICS, de las páginas 183 a 215, y note lo siguiente:

- Número de elementos.
 - La media de la clase en cada canal.
 - La desviación estandard (este valor es la raíz cuadrada de la varianza, la cual se enlista en la salida del CLUSTER).
 - La matriz de correlación (mostrando la correlación de los valores de los datos entre todos los pares de canales.
 - Histogramas.
 - Gráficas de coincidencia espectral.
-

REPRESENTACION VISUAL DE LAS CLASES CANDIDATAS DE ENTRENAMIENTO

Al momento de realizarse este análisis, se asociaron las clases de de entrenamiento con las clases informacionales en las áreas de entrenamiento. En la tabla III-4 se listan los nombres otorgados, a cada una de las clases de agrupamiento. Para realizar este proceso, se correlacionaron los datos de referencia (fotografías aéreas) con las impresiones de los mapas de agrupamiento.

Ahora se cuenta con 56 clases candidatas de entrenamiento, correspondiendo 20 al área uno, 15 del área dos, 18 del área tres y las clases supervisadas agua, nubes y sombras. Puesto que es muy seguro que algunas clases sean muy similares a otras, tanto en respuesta espectral como en identidad, es necesario el reducir el número de clases de entrenamiento a un paquete que sea representativo de la escena. Se puede reducir el número, eliminando algunas clases de entrenamiento candidatas, y combinando otras.

Para decidir cuales clases candidatas pueden ser descartadas o combinadas, es útil visualizar las relaciones espectrales entre las clases. El procesador *MERGESTATISTICS tiene como uno de sus varios resultados de salida una gráfica de dos dimensiones conocido como gráfica bi-espectral. En un eje de esta gráfica, se muestra el promedio de las medias de las clases en los dos canales infrarrojos; y en el otro eje aparece el promedio de los dos canales en la región del visible. El resultado es: una gráfica que proporciona una comparación visual de las medias de todas las clases de entrenamiento candidatas

En las páginas 217 y 218 de los listados de computadora se encuentran las gráficas bi-espectrales de las 56 clases de entrenamiento candidatas. Las dos páginas siguientes, listan las

CLASE	SAN PABLO DEL LAGO		TABACUNDO		CAYAMBE	
	Simb.	Nombre	Simb.	Nombre	Simb.	Nombre
1	∅	Vegetación	∅	Vegetación en Gra.	∅	Vegetación
2	-	Suelo desnudo	-	Suelos Agrícolas	-	Suelos
3	+	Suelo desnudo en pendientes	=	Suelos agrícolas	+	Suelos
4	/	Vegetac. dispersa	l	Vegetacion	/	Vegetación
5	1	Agricultura	C	Campos agrícolas chicos.	J	Cultivos en minifundios.
6	7	Agrícola Emergendo.	I	Agrícolas	C	Cultivos.
7	L	Laderas h/sol	Z	Agrícolas disp.	I	Vegetación
8	2	Agrícolas Emerg.	*	Agrícola "chicos"	Z	Suelo c/vegetac.
9	Z	Suelo c/veget.	T	Agrícolas "grande"	3	Veg en laderas
10	3	Suelo humedo	F	Campos agrícolas	&	Suelo desnudo
11	&	Cultivos/suelos	G	Agrícolas "grand"	T	Veg en laderas
12	T	Veg. natural	B	Bosques	F	Agrícolas
13	8	Veg. en canadas	R	Bosque abierto	G	Agrícolas
14	E	Cultivos	Q	Bosque	H	Agric. no muy verde.
15	X	Bosque disperso	W	Bosque en cañada	\$	Agrícolas
16	B	Bosque contra el sol.			o	Suelo c/vegeta.
17	\$	Vegetación en sombra			Q	Bosque en cañada
18	O	Vegetación/volcan			W	Bosque
19	Q	Bosque del volcan				
20	W	Agua (laguna)				

Tabla III-4.

clases, los símbolos utilizados para representarlas, el promedio del valor de la media para cada clase en el visible y en el infrarrojo, el número de clase de agrupamiento original, y la cantidad de elementos en cada clase. Las primeras 20 clases pueden ser reconocidas como provenientes del área de entrenamiento uno, las siguientes 15 del área de entrenamiento dos, y las otras 18 del área de entrenamiento tres; las últimas tres clases, son las que se crearon utilizando el método supervisado.

A causa de la gran cantidad de clases de entrenamiento, algunos símbolos aparecen dos veces en la gráfica bi-espectral. Haciendo uso de la lista de los promedios de los valores de las medias en las páginas 219 y 220, anada el subíndice 2 a los símbolos para las clases 31 a 56 en la gráfica bi-espectral. Por ejemplo, la clase 31 con el símbolo A, tiene una media en las bandas visibles de 26.5 (eje de las x), y 43.1 en las bandas del infrarrojo (eje de las y). Localice la "A" que representa a esta clase, y márkela como "A2". Realice, este mismo procedimiento para las clases 32 a 56.

Para incrementar la utilidad de esta gráfica, identifique cada clase, escribiendo el nombre de ella inmediatamente después de su símbolo en la tabla bajo la gráfica, como se muestra en la tabla III-2. Verifique que los suelos tienden a presentarse en la región superior derecha, la vegetación densa a la izquierda de la línea diagonal y el agua en la parte inferior izquierda. Esta gráfica será usada en pasos posteriores.

CALCULO DE LA DISTANCIA ESTADISTICA ENTRE LAS CLASES CANDIDATAS DE ENTRENAMIENTO.

La gráfica bi-espectral estudiada en la sección anterior, muestra que clases son espectralmente similares entre sí. Sin embargo, debido a que esta gráfica se basa solo en las medias de las clases, no indica nada acerca de cuáles clases se traslapan en el espacio de medición y la dimensión de este traslape. Para realizar buenas decisiones acerca de cuáles clases de entrenamiento hay que descartar o agrupar, se debe considerar también la dispersión de cada clase en el espacio multidimensional, y la extensión de cualquier traslape entre las clases. Esto se puede realizar, calculando el valor de la divergencia transformada (Dt) entre cada par de clases utilizando el procesador *SEPARABILITY.

En las páginas 223 a 254 se encuentran los listados del procesador *SEPARABILITY, para cada una de las 56 clases cuando se usan los cuatro canales únicamente. En la página 224, se muestran los símbolos usados para representar cada una de las clases. A partir de la página 227, se muestra la divergencia transformada para cada par de clases utilizando los datos de los cuatro canales del sensor. Por ejemplo, en la página 227 la divergencia transformada entre las clases "A" y "F" es de 1998, el cual es un valor relativamente alto, esto indica una separabilidad alta entre las clases cuando se utilizan los 4 canales de datos. El valor máximo para la divergencia transformada que se asigna es de 2,000, correspondiente a la máxima separabilidad. Como se observa en los valores de divergencia transformada de las páginas 227 a 247, existen algunos valores que alcanzan 2,000 o están muy próximos a él.

Como una ayuda para la localización de clases problemáticas, se pidió que todos los pares de clases con un valor de divergencia transformada de 1500 o menor fuesen listados, este listado aparece en las páginas 248 a 251. Esta lista condensada es útil para incorporar la información de separabilidad en la gráfica bi-espectral, la cual será empleada en próximas decisiones. Note nuevamente, que debido a la gran cantidad de clases algunos símbolos han sido utilizados dos veces.

Antes de continuar con el siguiente paso, será útil revisar de nuevo la impresión resultante del *SEPARABILITY. Cerca del margen superior de la página 227, se observa que la mínima divergencia transformada entre cualquiera de dos clases es 141, y que el promedio cuando se utilizan los cuatro canales es 1874.

Haciendo uso de la lista de pares de clases con valores de divergencia transformada menor de 1500 para los cuatro canales de datos (páginas 248 a 251), añade la información de separabilidad a la gráfica bi-espectral (páginas 217 y 218). Realice el paso anterior, trazando una línea continua entre cada par de clases con una divergencia transformada de 1500 o menos. Cuando el valor de la divergencia transformada tenga un valor entre 1500 y 1750, trace una línea punteada. Si la divergencia transformada es mayor de 1750 no trace ninguna línea.

Otra forma adicional de representar gráficamente las 56 clases candidatas de entrenamiento se presenta en las figuras III-18 a la III-20. Observando las gráficas de las medias calibradas, se confirma el tipo de cobertura general a la que pertenece cada clase,

CALIBRATED LANDSAT MEANS
(MWATTS/CM².SR.)

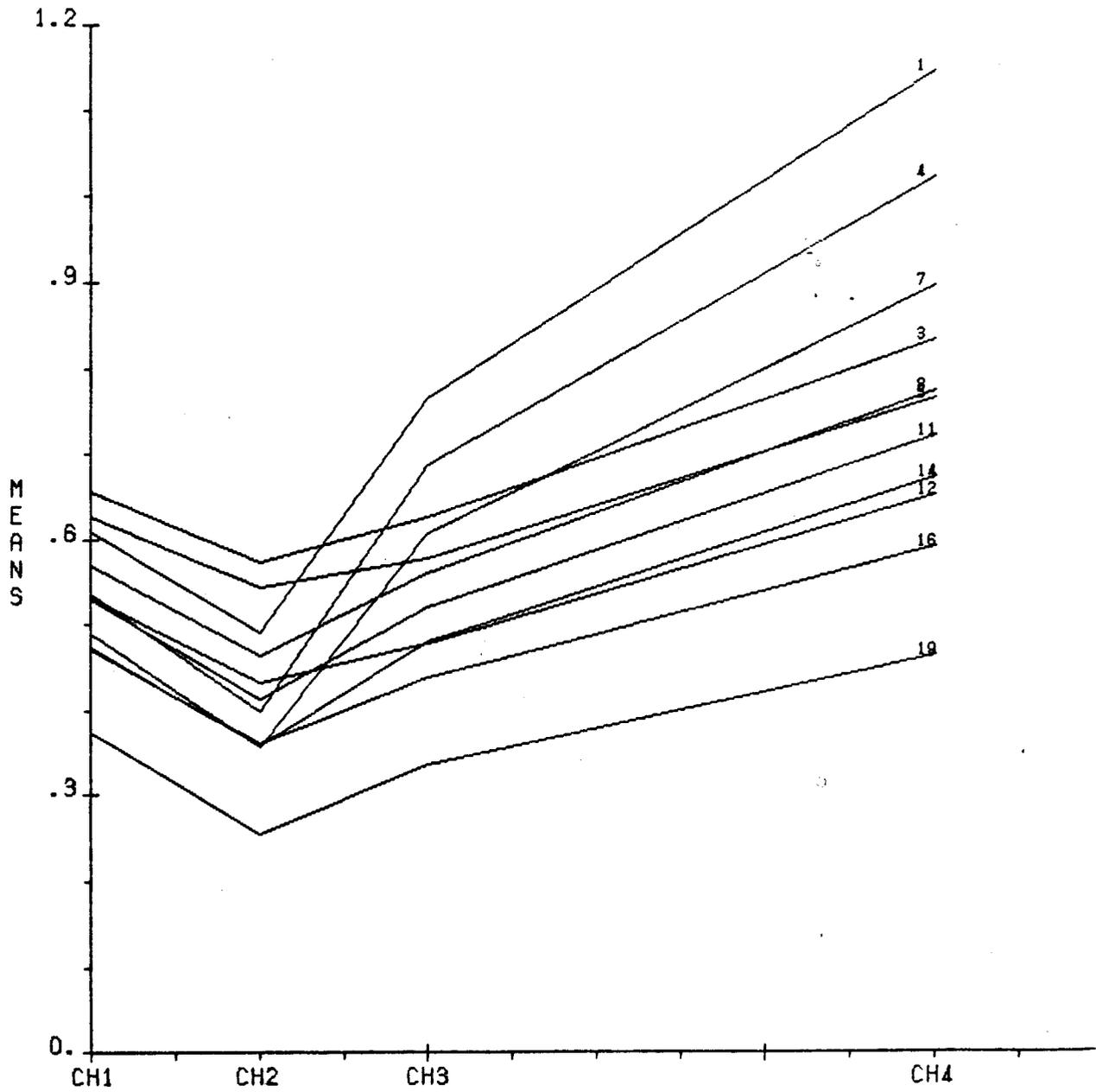


Figura III-18A. Gráfica de las medias calibradas del agrupamiento de San Pablo del Lago.

CALIBRATED LANDSAT MEANS
(MWATTS/CM².SR.)

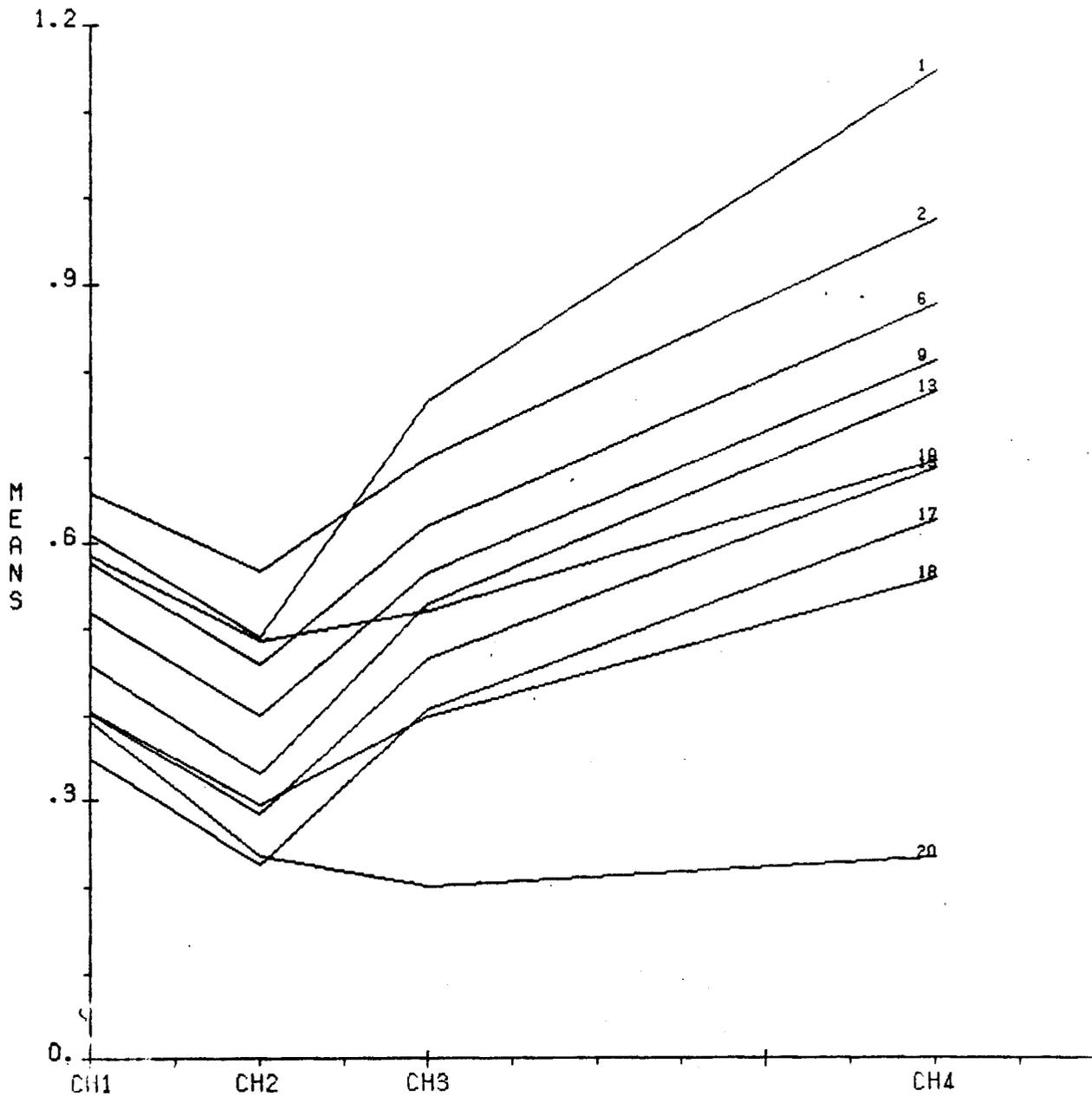


Figura III-18B. Gráfica de las medias calibradas del agrupamiento de San Pablo del Lago.

CALIBRATED LANDSAT MEANS
(MWATTS/CM².SR.)

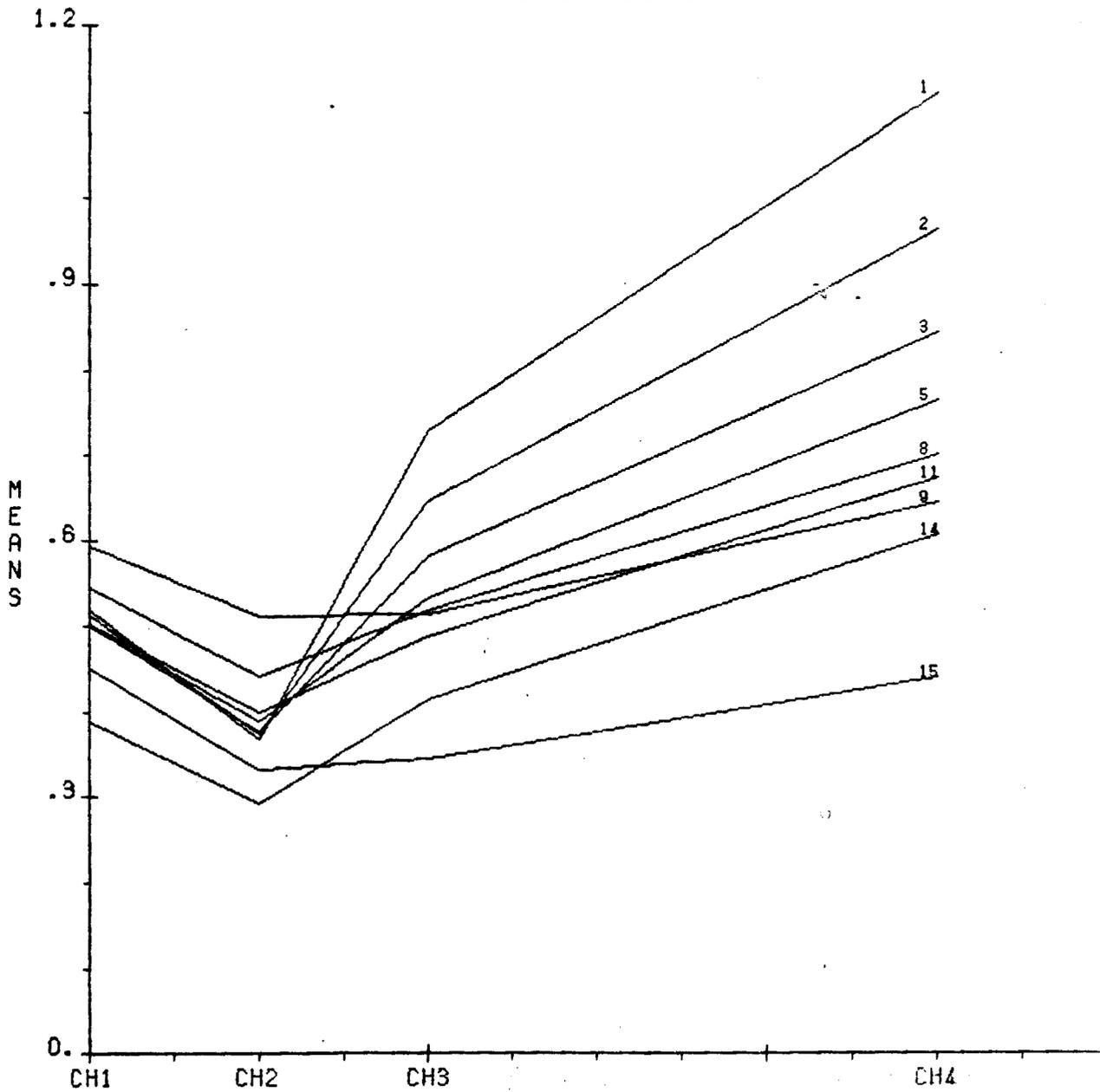


Figura III-19A. Gráfica de las medias calibradas del agrupamiento de Tabacundo.

CALIBRATED LANDSAT MEANS
(MWATTS/CM².SR.)

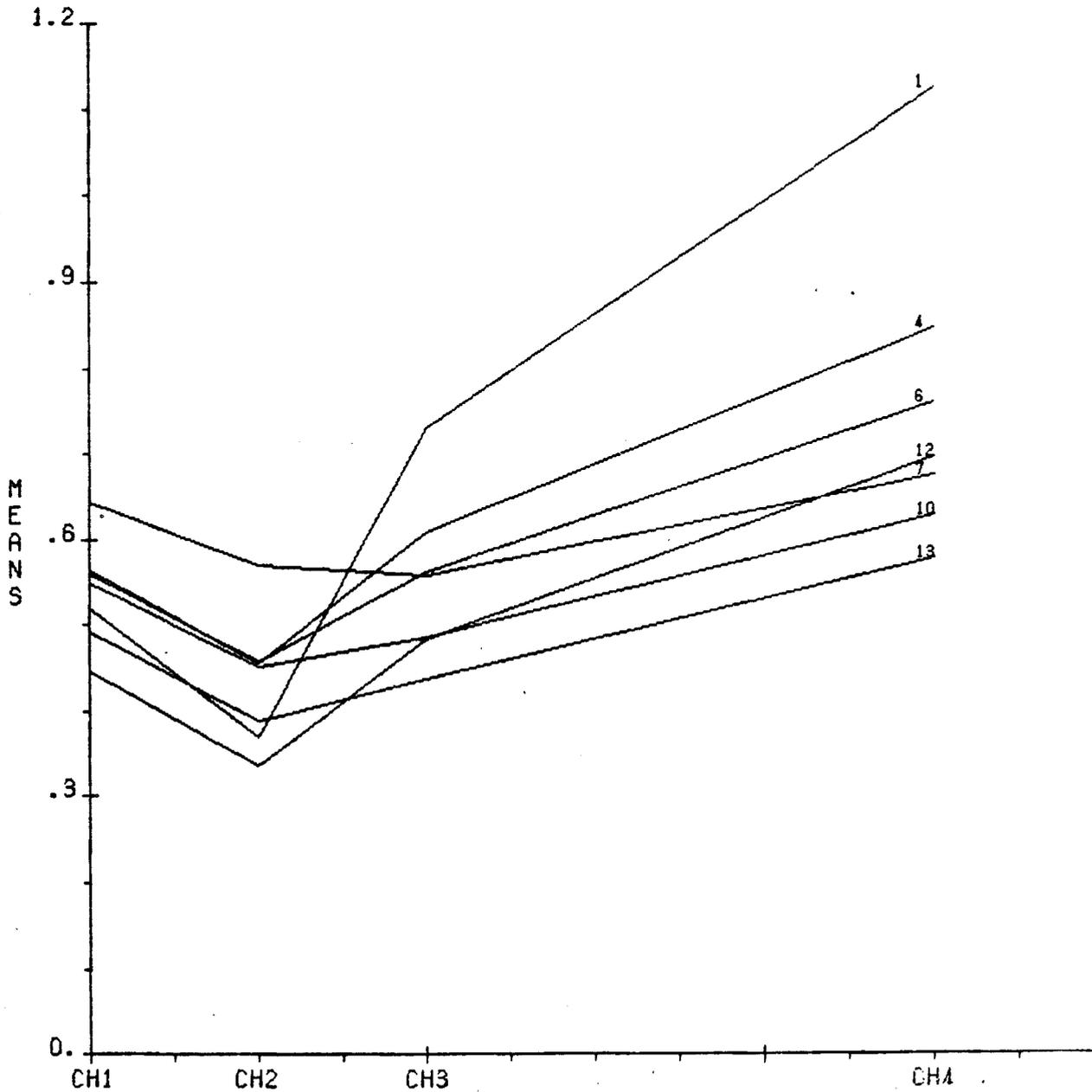


Figura III-19B. Gráfica de las medias calibradas del agrupamiento de Tabacundo.

CALIBRATED LANDSAT MEANS
(MWATTS/CM².SR.)

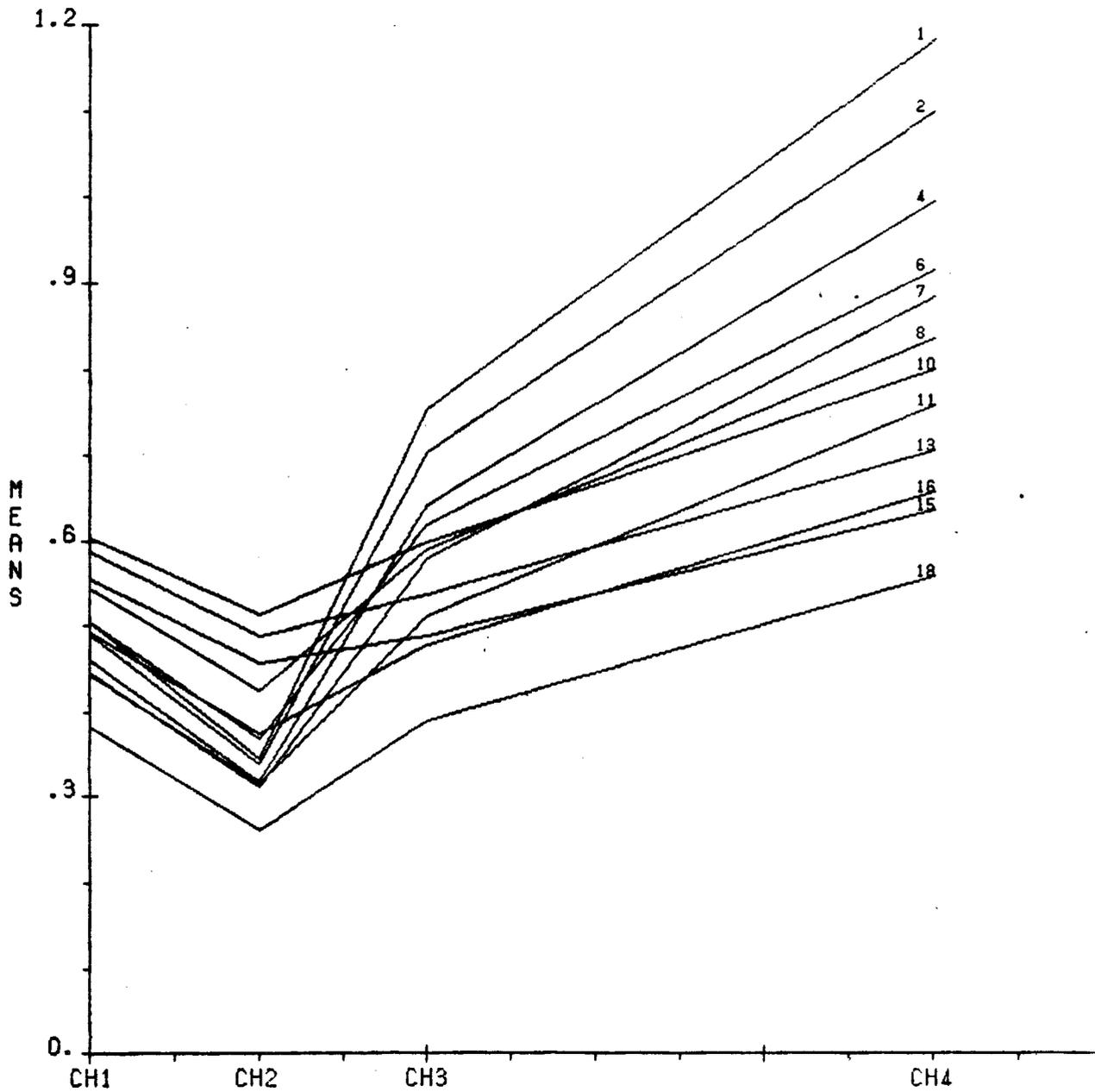


Figura III-20A. Gráfica de las medias calibradas del agrupamiento de Cayambe.

CALIBRATED LANDSAT MEANS
(MWATTS/CM².SR.)

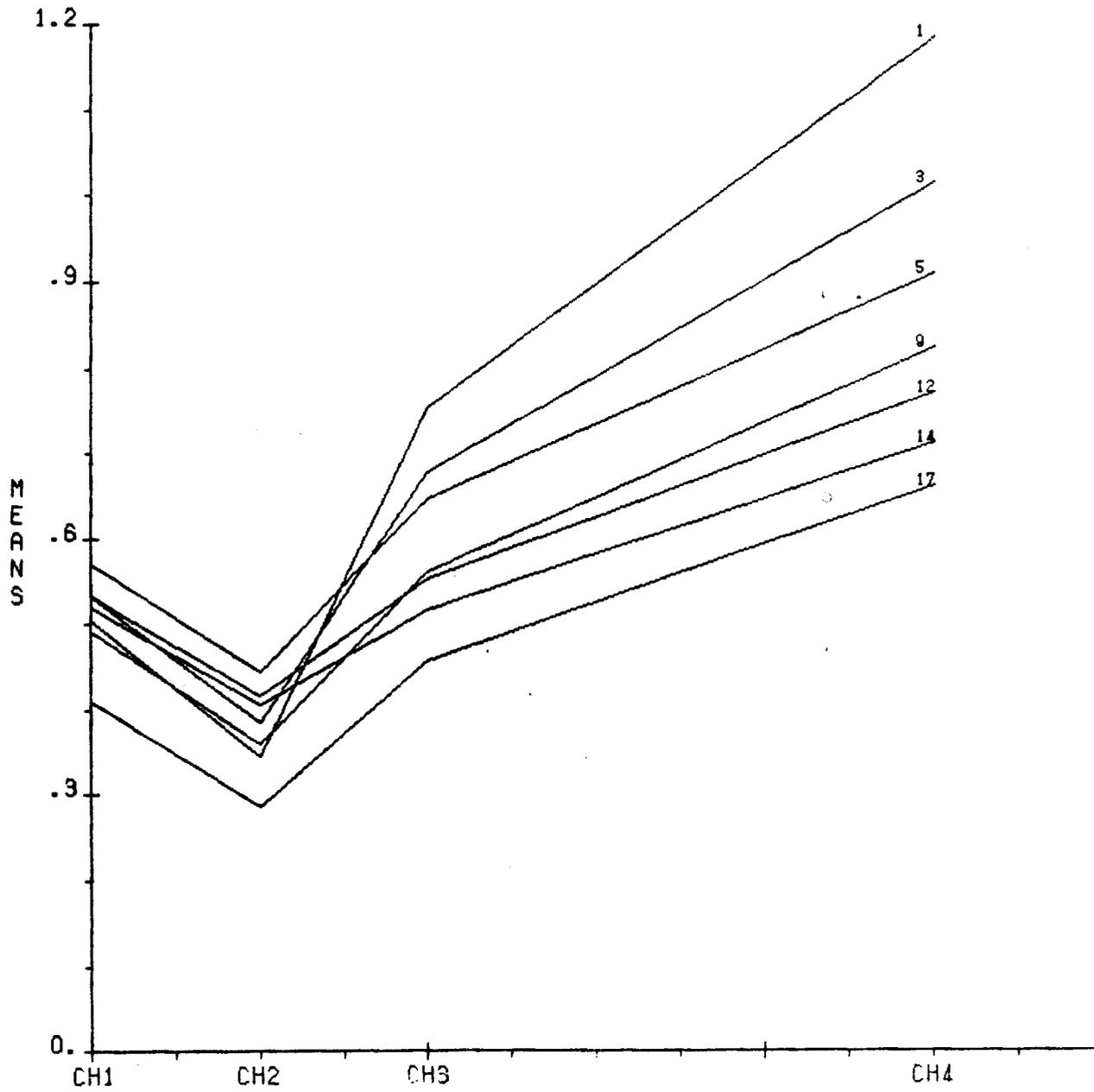


Figura III-20B. Gráfica de las medias calibradas del agrupamiento de Cayambe.

esto es, si la gráfica es representativa de vegetación, agua, suelo, o de otro material de la superficie terrestre; lo cual provee de otra forma de determinar la similitud espectral entre las clases. Es importante señalar, que las medias han sido calibradas, esto significa que las inconsistencias de canal a canal en la respuesta del sensor han sido removidas matemáticamente. Los valores de media que aparecen en el listado del *CLUSTER, no han sido calibrados, por lo cual no es posible comparar directamente entre estos valores.

De la figura III-18, seleccione cuatro o cinco clases espectralmente no similares, verifique en la tabla III-2 como fueron identificadas. Aquellas reconocidas como vegetación deben presentar la forma característica de este tipo de cobertura: un valor medio en el canal uno, bajo en el dos, mas alto en el tres y con un valor máximo para el canal cuatro. Realice el mismo tipo de correlación para los otros tipos de cobertura.

SELECCION DE LAS CLASES ESPECTRALES DE ENTRENAMIENTO FINALES.

Para la selección de las clases de entrenamiento finales, se cuenta con una gráfica bi-espectral con las clases identificadas y la información de separabilidad, los histogramas y las medias espectrales calibradas para cada clase.

En éste momento ya es posible seleccionar las clases finales de entrenamiento, y el realizar las decisiones acerca de cuales clases deben de ser eliminadas, cuales unidas y las que permanecen sin modificación. Al realizarse este proceso se debe tener en cuenta los identificadores de cada clase, la normalidad de los histogramas, el número de elementos en cada clase candidata de entrenamiento y la distancia estadística entre cada par de clases

Los dos principales objetivos que deben de satisfacerse cuando se definen las clases de entrenamiento finales son:

1. Como grupo, las clases de entrenamiento finales, deben de representar todas las clases de cobertura de la escena.
2. Las clases de entrenamiento deben de ser espectralmente separables unas de otras.

Se puede hacer uso de dos estrategias diferentes cuando se seleccionan las clases de entrenamiento: primero, el método de combinación de clases que son espectralmente similares; por ejemplo: las clases que tienen un valor de divergencia transformada menor al

valor de un umbral previamente establecido, de acuerdo a la precisión general deseada; y segundo, el método de descarte, en el cual las clases que se encuentran en los límites de las clases informacionales, son desechadas. El método de combinación, está más orientado a cumplir con el primero de los dos objetivos principales anteriormente señalados, mientras que el segundo método permite alcanzar el segundo. La experiencia ha demostrado que una combinación de estos dos métodos proporciona mejores resultados.

Note que las clases de materiales similares, tienden a ubicarse en una misma porción de la gráfica. Esto era de esperarse, ya que coberturas similares tienen respuestas espectrales parecidas. En la gráfica, se puede observar que los suelos tienden a colocarse arriba a la derecha y el agua se sitúa por debajo a la izquierda, formando así una línea casi diagonal, con varias mezclas de clases en el centro. Los campos agrícolas con cultivo en pie, se ubican en la región superior a la izquierda, y los bosques entre la agricultura y los suelos.

De las líneas sólidas trazadas en la gráfica bi-espectral, se puede observar que varias de las clases son perfectamente separables y tienen identidades específicas asociadas a ellas. Estas clases deben ser conservadas como clases de entrenamiento: agua (X2), nubes (Y2), sombras (Z2), suelo (B).

En este momento las decisiones más difíciles deben de ser realizadas; esto es: ¿qué hacer con los grupos de clases que no son espectralmente separables? Observe las tres clases identificadas por las letras "N", "A2" y "U2", representando a cultivos, campos agrícolas grandes y suelos húmedos con vegetación respectivamente. Se sabe por los valores de divergencia transformada, que el clasificador no puede distinguir entre ellas muy bien. Una estrategia útil en este momento sería, mantener la clase central "U2", y desechar las otras dos. A causa de la similitud en la respuesta espectral, la mayoría de los elementos pertenecientes a "N1" y "A2", serán identificados dentro de la clase representada por el símbolo "U2". Otro método consistiría en combinar las clases 14 ("N1") y 31 ("A2"), dado que ellas representan dos clases de información similares. La selección entre combinar o descartar es decisión del analista. En este caso se optó por la primera opción, la eliminación de las clases.

En la parte central inferior de la gráfica bi-espectral (páginas 217 y 218 de los listados de computadora), se encuentran las clases 5 ("E1"), 8 ("H1"), 10 ("J1"), 26 ("Z1"), 45 ("O2") y 48 ("R2"). Muchas de estas clases no son separables, y cuando se analiza la gráfica de las medias (figura III-21) calibradas, se observa que por su forma estas pertenecen a dos grupos diferentes. Compare por ejemplo, las gráficas de las clases 5, 10 y 48 representando suelos agrícolas. En las gráficas se observa que las clases 10 y 48 son muy semejantes y el valor de divergencia transformada entre ambas es extremadamente bajo (205), por otro lado

es de 1783 entre las clases 5 y 10 y de 1552 entre las clases 5 y 48. Además la clase 48 presenta el menor número de puntos, varianzas altas e histogramas no Gaussianos en los canales uno y dos. Por estas razones se decidió eliminar esta clase. Ahora bien, las clases 8 ("H1"), 26 ("Z1") y 45 ("O2"), las cuales fueron identificadas como cultivos, presentan valores de divergencia transformada por debajo de 1500. En las gráficas de medias calibradas es evidente que representan un mismo tipo de cobertura, por lo tanto se decidió conservar la clase 8, ya que presentaba el mayor número de puntos y la menor varianza.

Observe detenidamente a las clases candidatas de entrenamiento que aparecen en la gráfica bi-espectral, específicamente las clases 1 ("A"), 4 ("D"), 7 ("G"), 13 ("M"), 15 ("O"), 17 ("Q1"). Como podrá observarse, estas clases forman una familia de curvas, (ver la figura III-22) entre las cuales los valores de separabilidad concatenan a las clases. Es de gran utilidad el aislar un grupo de clases para estudiarlo en detalle. Haga uso de la gráfica bi-espectral, de la información de separabilidad (valores de divergencia transformada), las gráficas de las medias de clase calibradas, y las impresiones del *CLUSTER, para realizar las decisiones acerca de cuales clases conservar, cuales descartar y cuales combinar.

Haciendo uso del procesador *MERGESTATISTICS se descartaron algunas de las 56 clases candidatas de entrenamiento y quedaron 35; en una segunda fase del análisis se eliminaron clases hasta obtener un total de 18. Al hacer una evaluación final de las clases de entrenamiento, se decidió unir las clases 30 ("/") y la 50 ("T2"), por presentar un valor de divergencia transformada muy bajo. Las clases finales se enlistan en la tabla III-3. El analista hizo mayor uso del método de descarte, que del método de combinación.

Compare su juego de clases de entrenamiento, con las 17 clases de la tabla III-3. ¿Combinaría usted mas clases? ¿Descartaría mas clases?. Sus clases de entrenamiento finales no son necesariamente mejores o peores que las del analista. Recuerde que este paso, mas que ningún otro en el análisis, es mas un arte que una ciencia.

CALIBRATED LANDSAT MEANS
(MWATTS/CM².SR.)

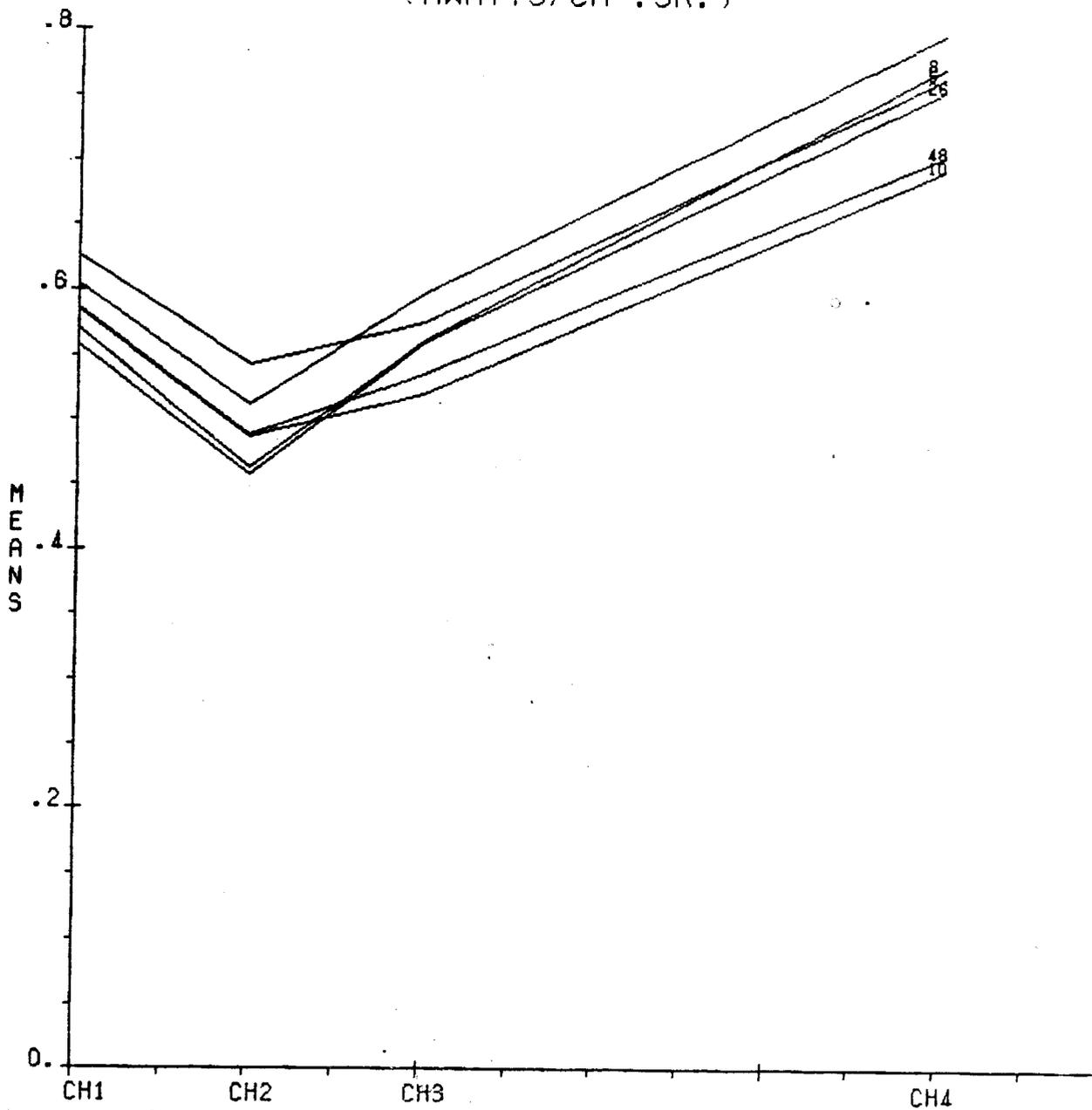


Figura III-21. Gráfica de medias calibradas de las clases 5, 8, 10, 26, 45 y 48.

CALIBRATED LANDSAT MEANS
(MWATTS/CM².SR.)

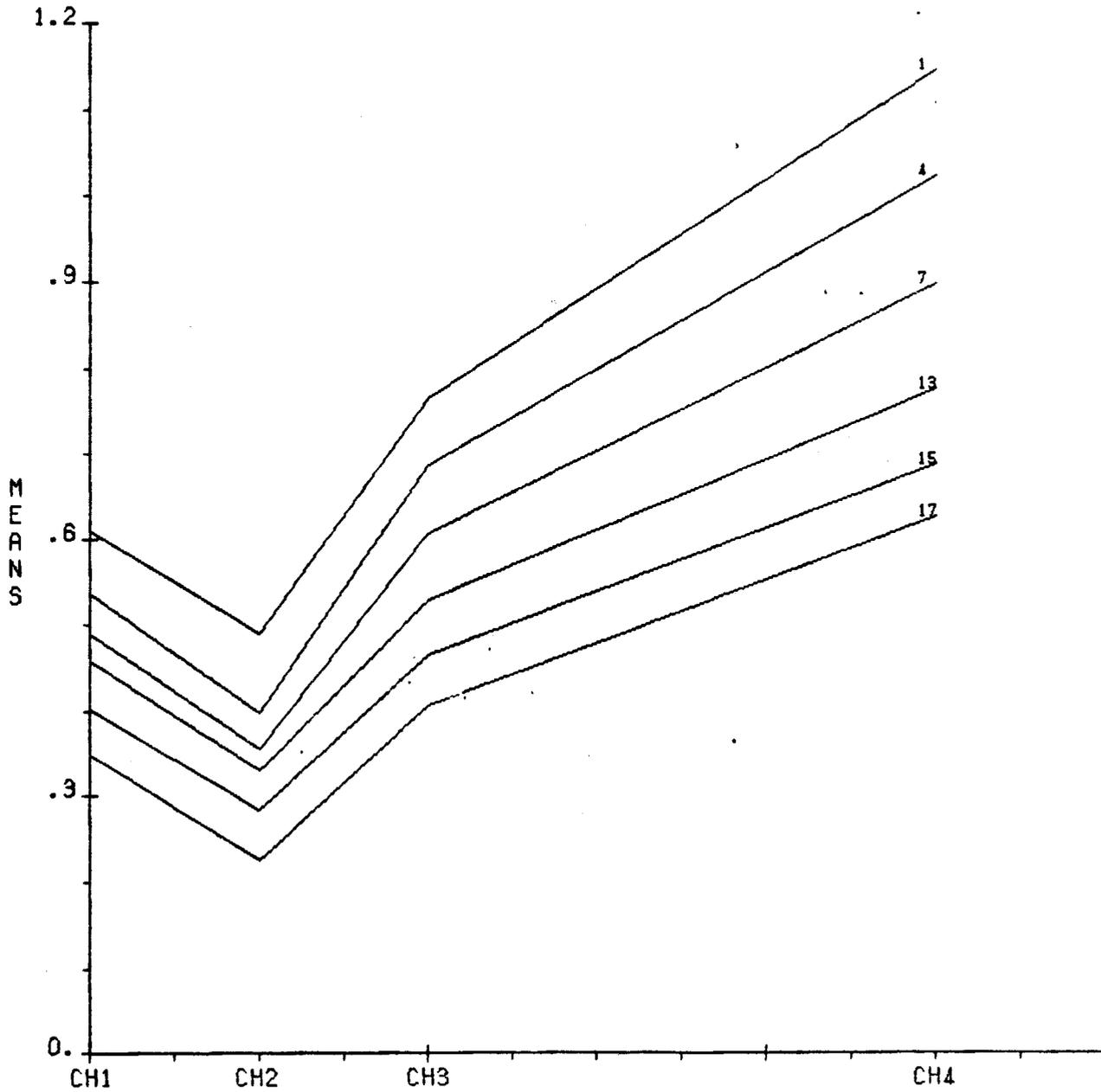


Figura III-22. Gráfica de medias calibradas de las clases 1, 4, 7, 13, 15 y 17.

Ahora que se han seleccionado las 17 clases de entrenamiento, es necesario evaluarlas con objeto de tener una indicación de la probabilidad de clasificación correcta que resultará al utilizar estas clases de entrenamiento. Para realizar esto en LARSYS, se puede correr nuevamente la función de proceso *SEPARABILITY. El listado para esta segunda corrida del *SEPARABILITY comienza en la página 295. En esta ocasión, el analista solicitó la separabilidad cuando se utilizan los cuatro canales, así como las combinaciones de 2 y 3 canales y el mejor canal individual. Una lista de los pares de clases con valores de divergencia transformada menores de 1500 fueron solicitados de nuevo, esto permite verificar nuevos pares de clases contra el mismo umbral. La lista basada en separabilidad utilizando los cuatro canales de datos aparece en las páginas 316 y 317.

Añada la nueva información de separabilidad y las identificaciones de las clases a la gráfica bi-espectral de las 17 clases de entrenamiento finales, la cual se localiza en las páginas 290 y 291 de los listados de computadora. Trace una línea discontinua entre pares de clases con valores entre 1500 y 1800. Esto indicará los pares de clases donde alguna confusión pudiese ocurrir.

Debido a que no existen valores de divergencia transformada menores de 1500 para este conjunto de clases de entrenamiento, es razonable el esperar en promedio, al menos una precisión en la clasificación de alrededor de 85 % para las clases de mayor similitud espectral. (Refiérase a la figura III-12, para observar cómo se obtuvo esta estimación). El valor promedio de la divergencia transformada es ahora de 1974, y con toda la evidencia disponible en los listados, se puede concluir que este es un conjunto apropiado de clases, el cual será adoptado como clases finales de entrenamiento. La impresión final del *SEPARABILITY, indica que la separabilidad declina rápidamente cuando menos de cuatro canales de datos son utilizados.

COINCIDENTAL BI-SPECTRAL PLOT(MEAN) FOR CLASS(ES)

MEANS PLOT OF CHANNEL 1 & 2 VERSUS 3 & 4

III-52

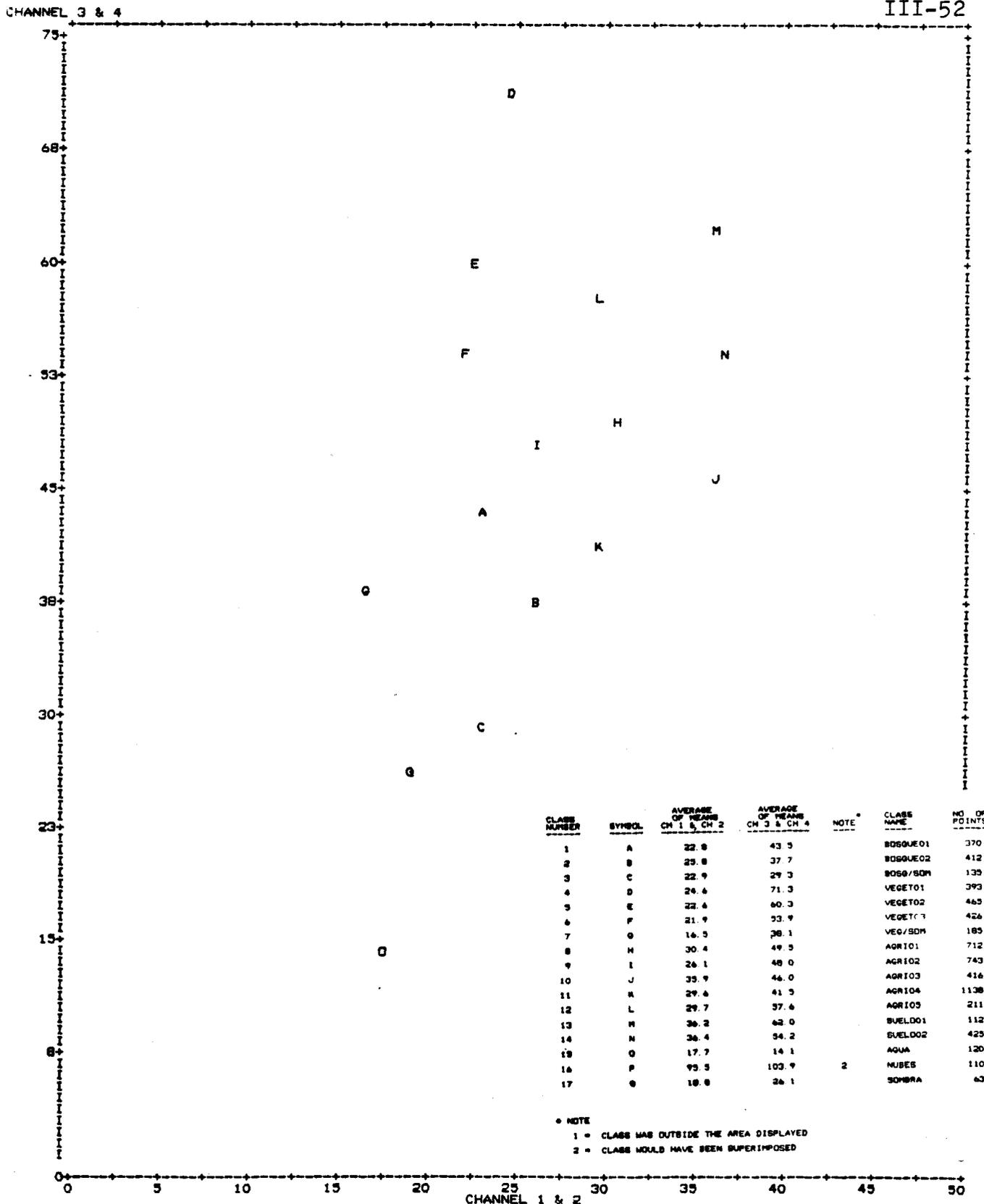


Figura III-23. Gráfica bi-espectral de las 17 clases de entrenamiento finales.

CALIBRATED LANDSAT MEANS
(MWATTS/CM².SR.)

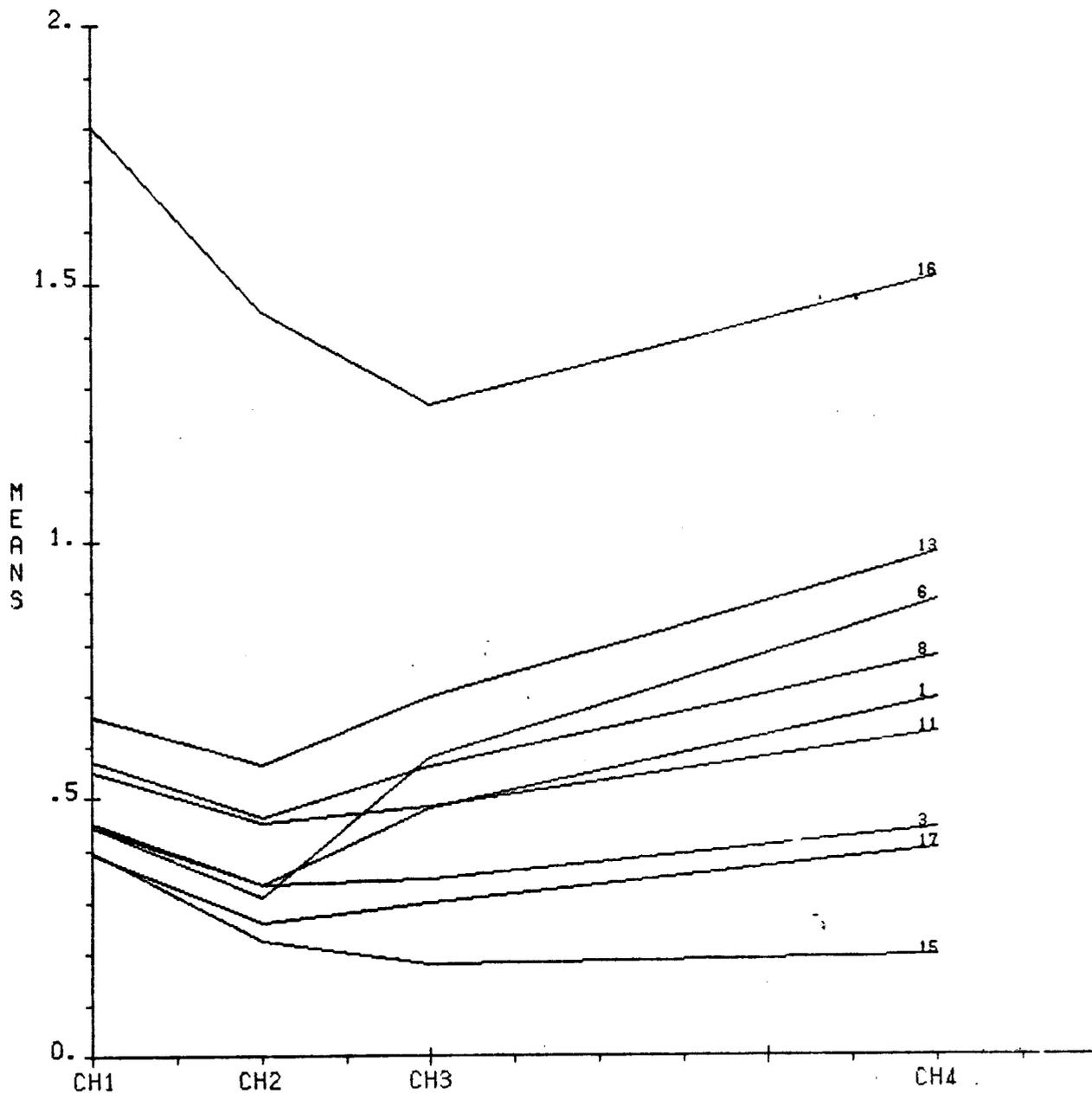


Figura III-24A. Gráfica de medias calibradas de las clases de entrenamiento finales.

CALIBRATED LANDSAT MEANS
(MWATTS/CM².SR.)

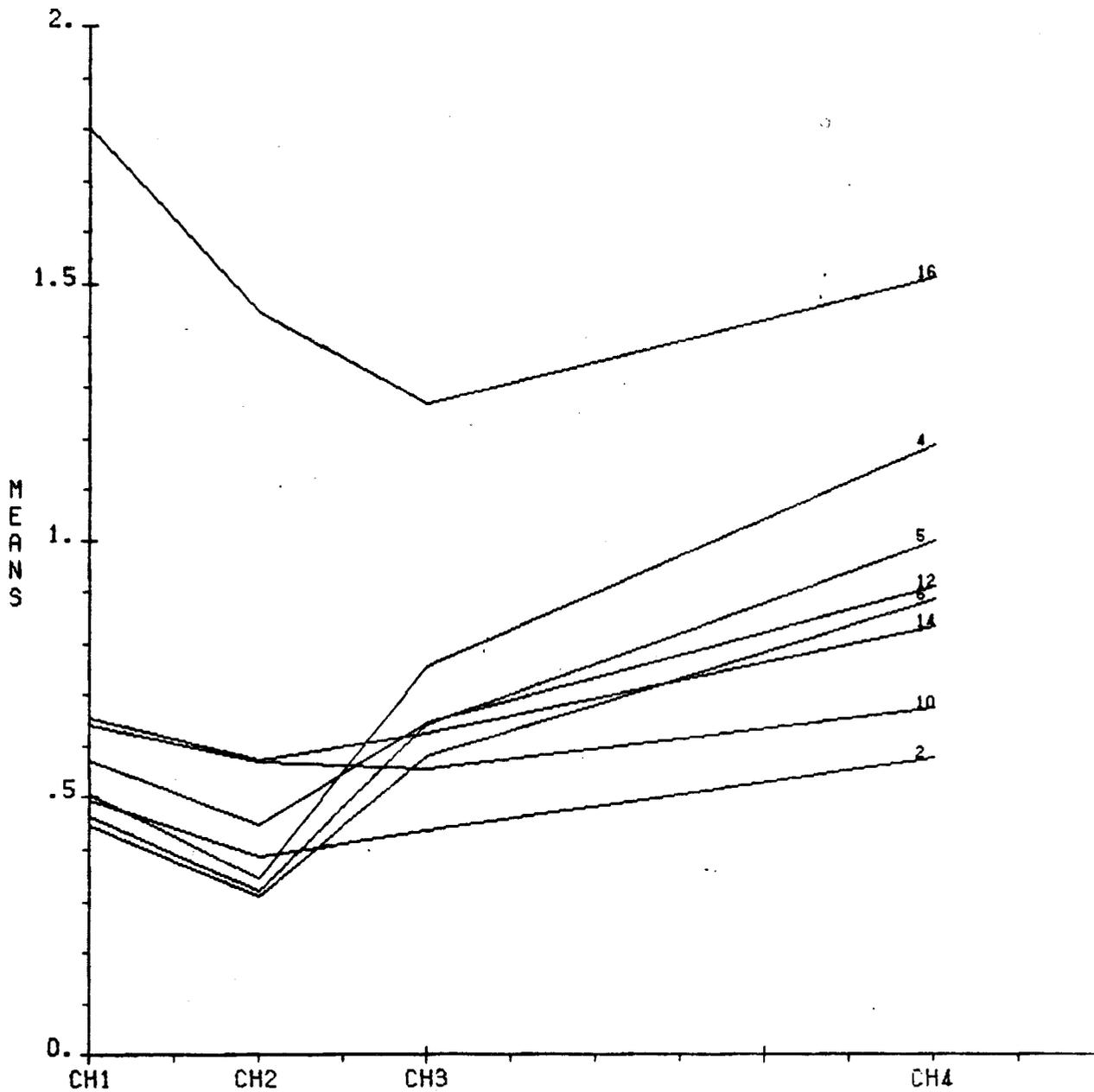


Figura III-24B. Gráfica de medias calibradas de las clases de entrenamiento finales.

PARTE IV - CLASIFICACION DEL AREA DE ESTUDIO

El área de estudio de la región Nor-Andina del Ecuador, fue clasificada con el procesador LARSYS denominado *CLASSIFYPOINTS usando las 17 clases de entrenamiento desarrolladas en pasos anteriores. Este procesador está basado en la regla de decisión de la máxima verosimilitud. Las funciones discriminantes de máxima verosimilitud para cada clase de entrenamiento evalúan el vector de datos de cada elemento. Cada elemento es clasificado dentro de la clase cuya función discriminante alcanza su mas alto valor.

Los resultados de este procesador son se encuentran en las página 321 a 325 de los listados de computador. Estos resultados se almacenan en una cinta para producir posteriormente un mapa o transferirlos al sistema de despliegue de imágenes.

En la página 324, se presenta la información que verifica que porción de los datos fueron clasificados (líneas y columnas) y en que archivo de cinta fueron almacenados los resultados.

PARTE V - DESPLIEGUE TABULAR Y/O PICTORICO DE
LOS RESULTADOS DE LA CLASIFICACION.

Un mapa de la clasificación, es producido por el procesador LARSYS denominado *PRINTRESULTS, y éste se presenta en las páginas 326 a 342 de los listados de computadora. Las tarjetas de control de la página 326, muestran que símbolos se utilizaron para cada una de las clases en el mapa; esta información se presenta de una forma mas accesible en la página 327, donde los nombres de las clases estan asociadas con los símbolos. Este mapa tiene varias páginas de ancho. Solicite al instructor un mapa que haya sido ensamblado previamente.

En el mapa, identifique algunas de las características terrestres con las cuales debe de estar ya familiarizado. ¿Sería mas fácil identificar estas características, si se hubieran seleccionados diferentes símbolos?. Suponga, que está interesado solamente en la localización de tierra forestal. ¿Que símbolos se asignarían a las 17 clases?.

El mapa de la clasificación puede ser desplegado en diferentes formas. La figura V-5 muestra una, en la cual símbolos gráficos son utilizados para representar cada clase o grupo de clases. La clasificación actual que se utilizó para crear este mapa, es la misma que se utilizó para crear la impresión alfa-numérica; sin embargo existen dos dieferencias en el despliegue: 1) Clases similares como el bosque 1 y el bosque 2, son representadas por el mismo símbolo; y 2) Los símbolos seleccionados se sugiere, vayan de acuerdo al tipo de la clase, haciendo inecesario el referirse repetidamente a la tabla de símbolos.

Los resultados de clasificación son tambien presentados en una forma tabular en la página 342. La tabla en la página 342 enlista el número de puntos (elementos) que fueron asignados a cada clase, el área en acres y en hectáreas que esa clase cubre, y el porcentaje del área ocupada.

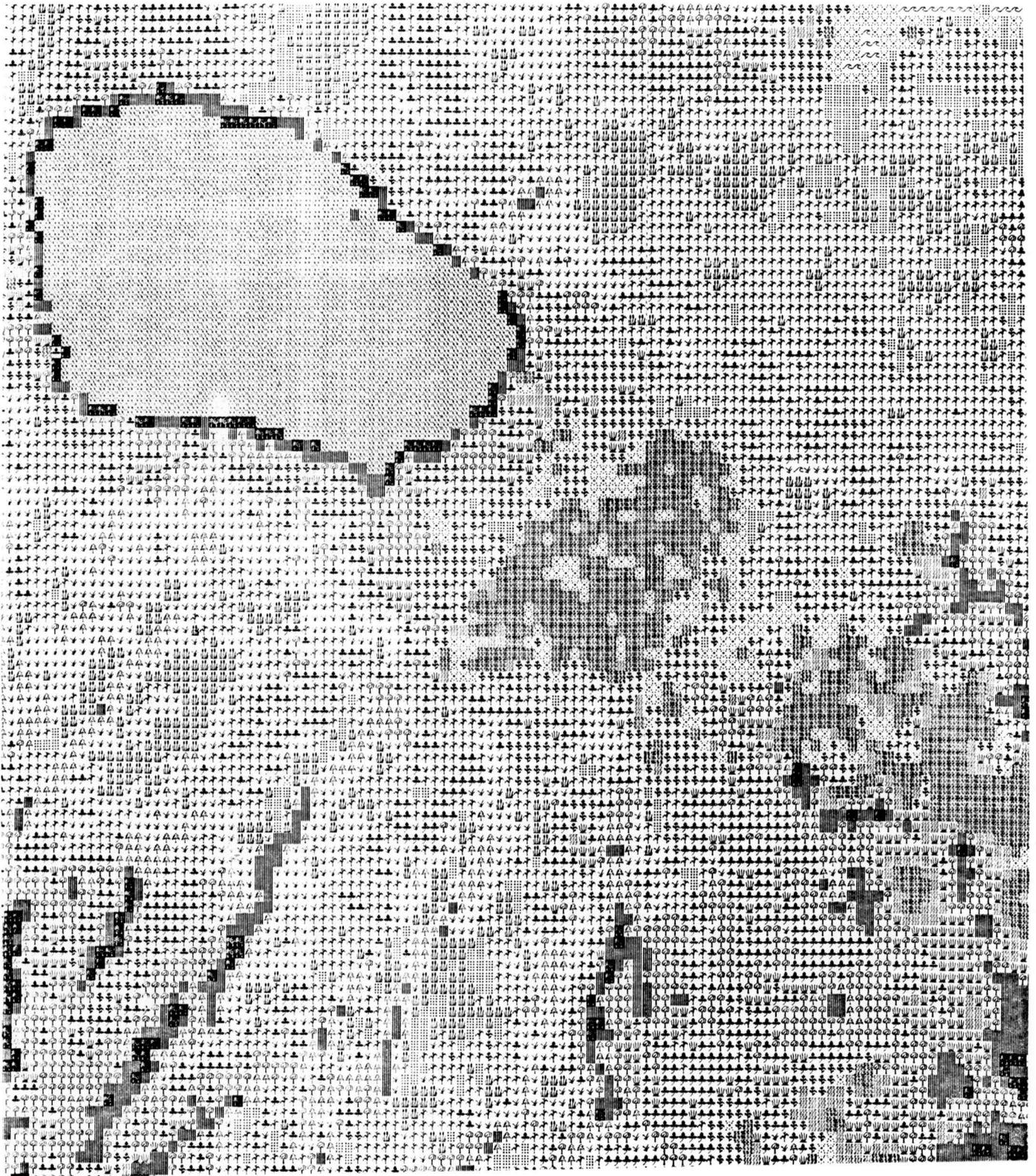


Figura V-5. Representación con símbolos gráficos de los resultados de la clasificación.

PARTE VI - EVALUACION DE LOS RESULTADOS DE LA CLASIFICACION

En el capítulo anterior se observó el mapa de la clasificación del área región Nor-Andina. Ahora, es el momento de evaluar la clasificación, para ver que tan cierta resultó. Esto se logra, seleccionando campos de prueba para cada uno de los tipos de cobertura mas abundantes en el paquete de datos (suelo desnudo, agricultura, bosque, vegetación dispersa y agua) y verificando que tan precisa es la clasificación en esas áreas.

Utilice los mapas en tonos de gris (figuras II-5 a la II-8) y las fotografías, para la selección de las áreas de los tipos de cobertura terrestre mas abundantes (suelo desnudo, agricultura, bosque, vegetación dispersa y agua). Estos campos de prueba deben de estar distribuidos en toda el área, pero no deben de abarcar ninguna de las áreas de entrenamiento. Deben de ser lo mas grande posible, cuidando siempre que se mantengan lo mas "puro" posible.

¿Cuántos elementos de cada tipo de cobertura estan presentes en los grupos de campos de prueba seleccionados? Para evitar parcialidades en la estimación de la certeza de la clasificación, se deben de escoger los campos de tal forma que, el número de elementos en cada categoría mayor de cobertura terrestre sea proporcional al número de elementos en cada clase en toda el área clasificada. Si se descubre alguna parcialidad, se pueden seleccionar campos de prueba adicionales para los tipos de cobertura terrestre que se encuentran escasamente representados en el grupo original de campos de prueba.

Existe otro método de seleccionar los campos de prueba, el cual es mas dado a proporcionar estimaciones sin parcialidad, y en el cual los campos de prueba son seleccionados al azar. Para usar este método, se puede dividir toda la escena en celdas de tres por tres y escoger al azar, el diez por ciento de ellas; se utiliza una tabla de números aleatorios para este proceso. Debe despues identificarse el tipo de cobertura terrestre en cada celda de prueba, eliminando aquellas celdas que contengan mas de un tipo de cobertura.

Cualquiera que sea el método que se use para para seleccionar los campos de prueba, las coordenadas de los campos y la identificación del tipo de campo, son proporcionadas al procesador *PRINTRESULTS, para producir una matriz de confusión para los puntos en esos campos. Si la selección de los campos de prueba es imparcial, los resultados deben de representar razonablemente bien la certeza general de la clasificación.

Analice como fueron clasificados los elementos en cada campo de prueba, después complete la siguiente matriz adjunta (Tabla VI-1)

Estudie la matriz y vea si es posible contabilizar ejemplos claros de errores de omisión o comisión. ¿Existen cambios que se deseen realizar en el refinamiento de las clases de entrenamiento, en la identificación de las clases, o en cualquier otro aspecto del análisis para mejorar la clasificación?.

CAMPOS CLASIFICADOS COMO:

CAMPOS IDENTIFICADOS COMO:	NÚMERO DE ELEMENTOS	A G R I C	V E G E T	S U E L O	S O M B R	N U B E S	A G U A
AGRICULTURA							
VEGETACIÓN							
SUELOS							
SOMBRAS							
NUBES							
AGUA							

Tabla VI-1. Matriz de confusión con los resultados de la clasificación.