JSC 09391

CROP IDENTIFICATION TECHNOLOGY ASSESSMENT FOR REMOTE SENSING (CITARS)

VOLUME VIII

DATA PROCESSING AT THE EARTH OBSERVATIONS DIVISION LYNDON B. JOHNSON SPACE CENTER

PART 5

FAYETTE COUNTY, ILLINOIS

SUPPLEMENT

GRAPHIC STUDY OF CORN AND SOYBEAN DATA

National Aeronautics and Space Administration

LYNDON B. JOHNSON SPACE CENTER

Houston, Texas

December 1975

JSC 09391

CROP IDENTIFICATION TECHNOLOGY ASSESSMENT FOR REMOTE SENSING (CITARS)

VOLUME VIII

DATA PROCESSING AT THE EARTH OBSERVATIONS DIVISION LYNDON B. JOHNSON SPACE CENTER

PART 5

FAYETTE COUNTY, ILLINOIS

SUPPLEMENT

GRAPHIC STUDY OF CORN AND SOYBEAN DATA

PREPARED BY

Francis Johnson, Jr.

APPROVED BY

Robert M. Bizzell, Project Manager

for Andrew E. Potter, Chief
Research, Test, and Evaluation Branch

R. B. MacDonald, Chief
Earth Observations Division

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

December 1975

ii

Blank

CONTENTS

Blank

TABLES

Blank

FIGURES

1.0 OBJECTIVES AND DATA SETS USED

The results of a graphic study of a data subset from the first Earth Resources Technology Satellite (ERTS-1) for the Crop Identification Technology Assessment for Remote Sensing (CITARS) appear in this supplement. The primary objectives of the graphic study were:

1. Determining if the signature clusters of individual corn and soybean fields migrate over consistently different trajectories in data space throughout the growing season.
2. Determining how spread out in data space the signature clusters of individual corn and soybean fields are at different times during the growing season.

Of the six geographic CITARS segments, segment 5 (Fayette County, Illinois) had the largest number (six) of ERTS-1 data sets acceptable for CITARS use. The acquisition dates of these six Fayette County sets were not uniformly spread throughout the 1973 growing season. These dates were June 10, 11, and 29, July 16 and 17, and August 21. (Note that in June and July some acquisition dates were just one day apart).

The Fayette County corn and soybean data from training fields used in the graphic study were unaffected by clouds or sensor problems on the six acquisition dates. The data meeting these qualifications are described in table I.

TABLE I.— FAYETTE COUNTY TRAINING FIELD DATA
USED IN THE GRAPHIC STUDY

Crop row direction	Corn		Soybeans	
	Pixels	Fields	Pixels	Fields
East-west	15	2	57	6
North-south	44	4	52	8
Total	59	6	109	14

2.0 SIGNATURE MIGRATIONS (SIGMIG) COMPUTER PROGRAM

The four-channel data from each of the 168 corn and soybean pixels in the six data sets were punched on cards to be read by the SIGMIG program. Designed to investigate signature migrations in data space, SIGMIG illustrated the data in the six two-channel subspaces, using the Stromberg Carlson 4060 computer and the Integrated Graphics Software subroutine package on a Univac 1108 computer. The graphic output is on strip microfilm.

The data from one field on one acquisition date are drawn as dots connected by straight lines to their circled mean. This figure resembles a shellburst or nuclear fission in an emulsion as shown in figure 1. These figures will be called field clusters or clusters, although they were defined actually by data registration, not a clustering algorithm.

Illustrations such as figure 1 show effects which would be lost in a statistical ellipsoid representation, including integer restriction, periodically vacant data levels, anomalous individual pixels, and multimodal distributions.

The data cards were run through the program several times, and each time a different cluster combination was drawn. Many different possible combinations existed. For example, a figure could have shown the six clusters of one field over the growing season (one cluster for each acquisition date) or all clusters of one crop on one of the six dates.

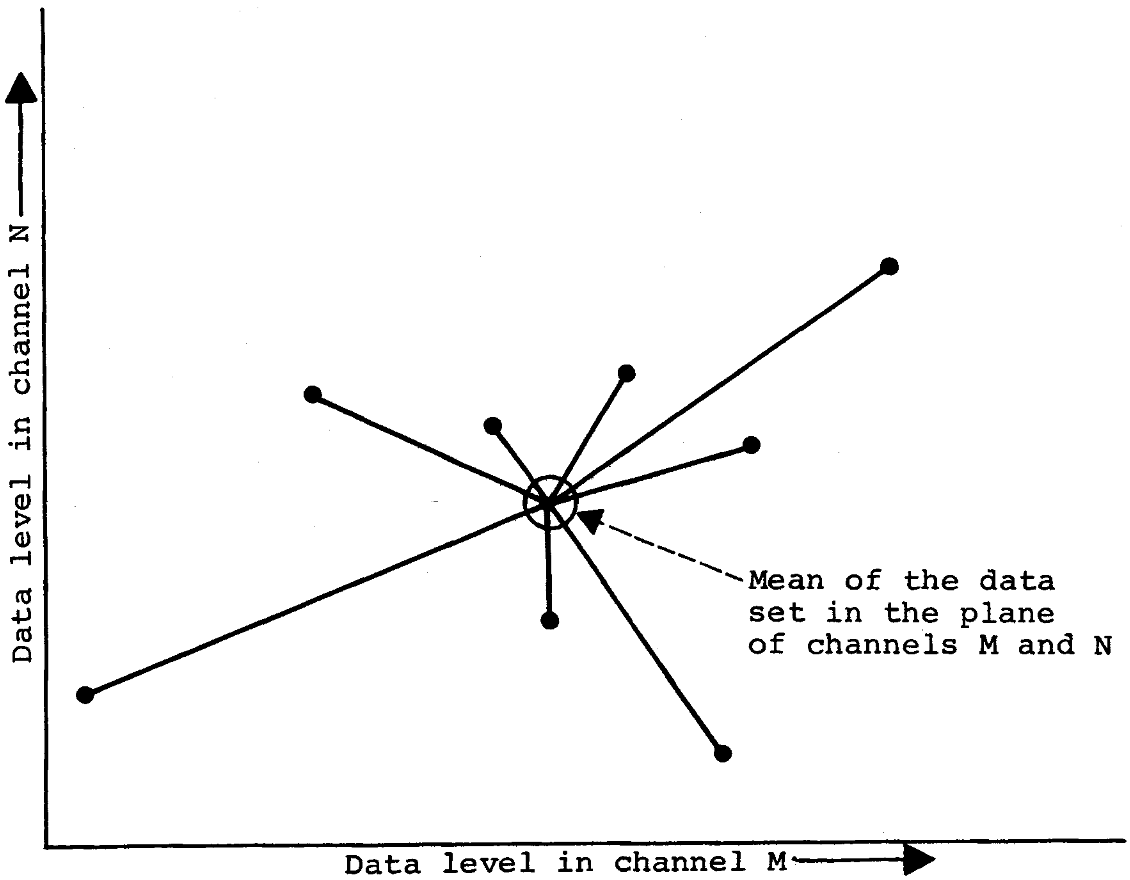


Figure 1.– Example of a field cluster drawn by SIGMIG. The data came from an eight-pixel field on one date.

3.0 RESULTS

3.1 CORRELATION BETWEEN CHANNELS 1 AND 2

The most obvious pattern which the SIGMIG figures first showed was the well-known linear correlation between channels 1 and 2. Figures 2(a) and 2(b) show the corn and soybean data in the plane of channels 1 and 2.

Equations (1) and (2) were obtained from least square fits of a straight line in the plane of channels 1 and 2. The residuals of the fits were measured perpendicularly from the straight line to the data points because both channels had equal priority. Neither channel was treated as a dependent variable with the other as an independent variable. No data points were omitted in these fits although several points were foreign to the overall distributions. Equation (1) is for corn, and equation (2) is for soybeans.

$$CH1 = 16.039 + 0.6229 \times (CH2) \qquad (1)$$

$$CH1 = 16.263 + 0.6183 \times (CH2) \qquad (2)$$

where CH1 is channel 1 and CH2 is channel 2. The root-mean-square (rms) residual for equation (1) is 1.573 and for equation (2), 1.519.

The corn and soybean lines intersect near the center of the figures at $CH2 = 48.7$. At the figure borders where $CH2 = 0$ and $CH2 = 100$, the distance measured perpendicularly between the two straight lines is less than 0.2 data unit. Within the data point distributions, the perpendicular distance is less than 10 percent of the rms residuals of the fits. Consequently, the differences between the corn and soybean straight line fits are well below the noise level, and a single linear equation can be considered as fitting both the corn and soybean data.

Quadratic or higher ordered equations were not fitted to the data in this plane. However, the researcher found that the means of the clusters deviate from the straight line fit in perhaps consistent patterns and that these deviation patterns might be different for corn and soybeans. The dense plotting of individual data points in figure 2 obscures these deviation patterns of the cluster means, which are of the same magnitude as the rms residuals of the linear fit.

Either channel 1 or 2 could have been disregarded in later analysis of the data if all the corn and soybean data points lay exactly on a single line in the plane of channels 1 and 2 with an rms residual of zero (or nearly zero because data values were restricted to integers.) If the data value for a pixel in one channel was known, the equation of the line would give the data value in the other channel regardless of whether the pixel represented corn or soybeans.

Although the observed linear correlation between channels 1 and 2 was not perfect (the rms residual was about 1.6 data units), either channel 1 or 2 can be ignored in investigating patterns in these data with dimensions much larger than 1.6 data units. The researcher chose to ignore channel 1 at first in this study because the range spanned by corn and soybean data was smaller in channel 1 than in channel 2. This choice made possible a convenient three-dimensional data representation which can be illustrated graphically.

3.2 DATA PLANES

In the data space of channels 2, 3, and 4, the corn and soybean data from all six dates were constrained closely to a flat surface (fig. 3). Throughout the growing season, the data clusters of the individual fields moved on this surface within a roughly triangular region.

On all dates, except perhaps the last, August 21, all the data clusters of one crop never coincided or grouped tightly together to form a traditional distinguishing signature. This lack of distinguishing signatures may have caused poor classification results for all the segments of CITARS, which was based on the traditional assumption of signatures.

To determine better the geometry of the surface in figure 3, the researcher made least square fits of plane surfaces to the data. The investigator measured the residuals in these fits perpendicularly from the plane to the data points so that all data channels had equal priority. No data points were omitted in these fits, although some points were foreign to the surface distributions. See section 4.0.

The researcher fit the data as three different sets: corn, soybeans, and corn and soybeans combined. Two different plane fits were made to each set. The first in the complete four-dimensional data space defined a three-dimensional plane. The second fit in the three-dimensional space of channels 2, 3, and 4 defined a two-dimensional plane.

The objectives of this selection of plane fits were to find a way to distinguish between the two crops and to bridge the gap between the confusing four-dimensional space and the more comprehensible but slightly compromised three-dimensional space. Table II lists the axis intercepts of the plane, components of a unit normal vector from the origin to the plane, normal distance from the origin to the plane, and rms residual for each fit.

All the rms residuals of these fits were only slightly larger than 1 data unit. The similarity of the rms residuals of the separate corn and soybean data fits indicated that both crop data were equally tightly constrained to planes. These rms residuals of the separate crops were only slightly smaller than that for the combined crops, indicating that the corn and soybean planes essentially coincided. For each data set, the rms residual of the four-dimensional space fit was only slightly less than that of the three-dimensional space fit. This slight difference showed that the amount of data distribution or structure lost from view by ignoring channel 1 was small.

The researcher did not study the possibility that these plane data surfaces were approximations to actual surfaces with shallow curvatures. Occasionally, the residuals of a cluster lay predominantly on one side of a plane, but no large areas on a plane existed in which all residuals were on one side. Investigating these data patterns perpendicular to a fit plane would require a special graphic display of the magnitude and side of residuals as a function of the positions on the plane. This display was not programmed because these perpendicular data petterns, if existent, must be much smaller than the patterns parallel to the plane already shown by other graphics.

The separate corn and soybean plane fits in either three- or four-dimensional space straddled the origin between them. In four-dimensional space, each plane was about 3.36 data units from the origin, measured perpendicularly. In the three-dimensional space, these distances were much smaller.

The angle between the separate corn and soybean planes in three-dimensional space was 0.064 radian (3.67°), calculated by finding the scalar product of the corn and soybean unit normal vectors and finding the arc cosine of the result. The same procedure applied to the separate three-dimensional crop planes in four-dimensional space showed an angle of 0.681 radian (39.05°). The significance of the over tenfold difference in corresponding angles in spaces of different dimensions was not immediately obvious. Because the researcher has not yet similarly analyzed data of other crops and classes, the usefulness of this angle in four-dimensional space in distinguishing between classes is uncertain.

The node of the 0.064-radian angle between the separate corn and soybean planes in the three-dimensional space extended across the triangular region between the mature corner and the bare soil axis. See section 3.3. Figure 4 shows the distances between these planes measured nearly perpendicularly to them at three points which are the approximate vertices of the triangle projected on the plane of channels 2 and 3.

On figure 4, the distances in data units between the separately fit corn and soybean planes are measured nearly perpendicular to both planes. Positive values indicate that the soybean plane lies farther along the positive direction of the channel 4 axis than the corn plane.

At the mature, right-angled corner of the triangle, the distance between the planes was more than twice the rms residuals of their fits. Corn and soybean data surfaces might be genuinely separated in this corner. However, over the rest of the triangle, the distance between the planes was of the same magnitude as the rms residuals of their fits, and the relevance of distinctly different corn and soybean surfaces was questionable.

Figures 5(a) and 5(b) from SIGMIG show separately the corn and soybean data in the plane of channels 2 and 3 on all six dates. The soybean data defined a slightly larger triangular area with more definite boundaries than that defined by the corn data. This difference might exist because there was almost twice as much soybean as corn data.

3.3 SIGNATURE MIGRATION TRAJECTORIES

Early in the growing season the fields of corn and soybeans appeared indistinguishable from each other because both were essentially bare soil. The bare soil data are spread out along the diagonal of the triangle, along the axis of CH1 = CH2 in the projection shown in figure 6. The researcher has not determined yet the factors governing the location of the bare soil data along this axis. Moisture probably was an influential factor. In theory, the bare soil axis should form the node between data surfaces of different crops, if they exist separately.

By late in the growing season when the plants were mature, the corn and soybean data had migrated to the right-angled corner of the triangle as shown in figure 6(a).

The soybean data did not follow a specific migration trajectory from the early locations on the bare soil axis to the mature corner; nor do they tend to follow straight, direct paths from these early locations to the mature corner. The perpendicular distance of the soybean data from the bare soil axis seemed to be a function of plant height. However, the researcher does not know the factors determining where data of soybean plants of a given height lie along a line parallel to the bare soil axis. Crop row direction did not seem to be correlated with soybean data trajectory characteristics.

On any one acquisition date, the soybean plant heights in different fields were unequal; and consequently, the data were spread out perpendicularly to the bare soil axis. This spread was most noticeable on the mid-July dates when the plant heights covered a wide range and the data were spread out over the whole triangle. Obviously, soybean data could not be classified successfully in CITARS using the traditional signature-cluster classification method and training statistics from these data. Figures 7 through 12 show the corn and soybean data on each acquisition date in channels 2 and 3 (CH2 and CH3).

The following abbreviations appear in these figures to identify the clusters.

Abbreviation	Definition		Abbreviation	Definition
DY	Day		SEC	Section
FLD	Field		SEG	Segment
MO	Month		YR	Year

The distance between the vertical parallel dotted lines to the left of each graph in figures 7 through 12 represent the horizontal width of the graph. The horizontal location between these dotted lines of the arrow from the identifying integers of a cluster is proportional to the abscissa coordinate of the corresponding circled cluster mean in the graph. The vertical position of the arrow equals that of its corresponding circled cluster mean.

The corn data trajectories differed from those of the soybean data in two ways. First, corn trajectories were a function of crop row direction. Second, the corn data always withdrew slightly from the mature corner of the triangle toward the channel 2 axis late in the season, while the soybean data went to the mature corner and remained there.

The following described effects of row direction on corn trajectories are speculative because only 15 cornfield pixels had an east-west row direction while 44 cornfield pixels had a north-south row direction.

The north-south corn data followed a narrow channel on the triangle along the route in figure 13(a). These data avoided the upper central region and the lower apex of the triangle.

The 15 pixels of east-west corn data came from two fields. The trajectories of these two fields differed from those of the north-south fields and from each other, as shown in figure 13(b). The east-west corn data did not avoid the upper central region of the triangle as did the north-south corn data. Figures 14(a) and 14(b) show all of the corn data separated into these two categories of crop row direction.

3.4 TRENDS SHOWING DATA RELIABILITY

Two trends seemed to indicate data reliability. However, without experience in analyzing these trends, the researcher made no rigorous interpretations.

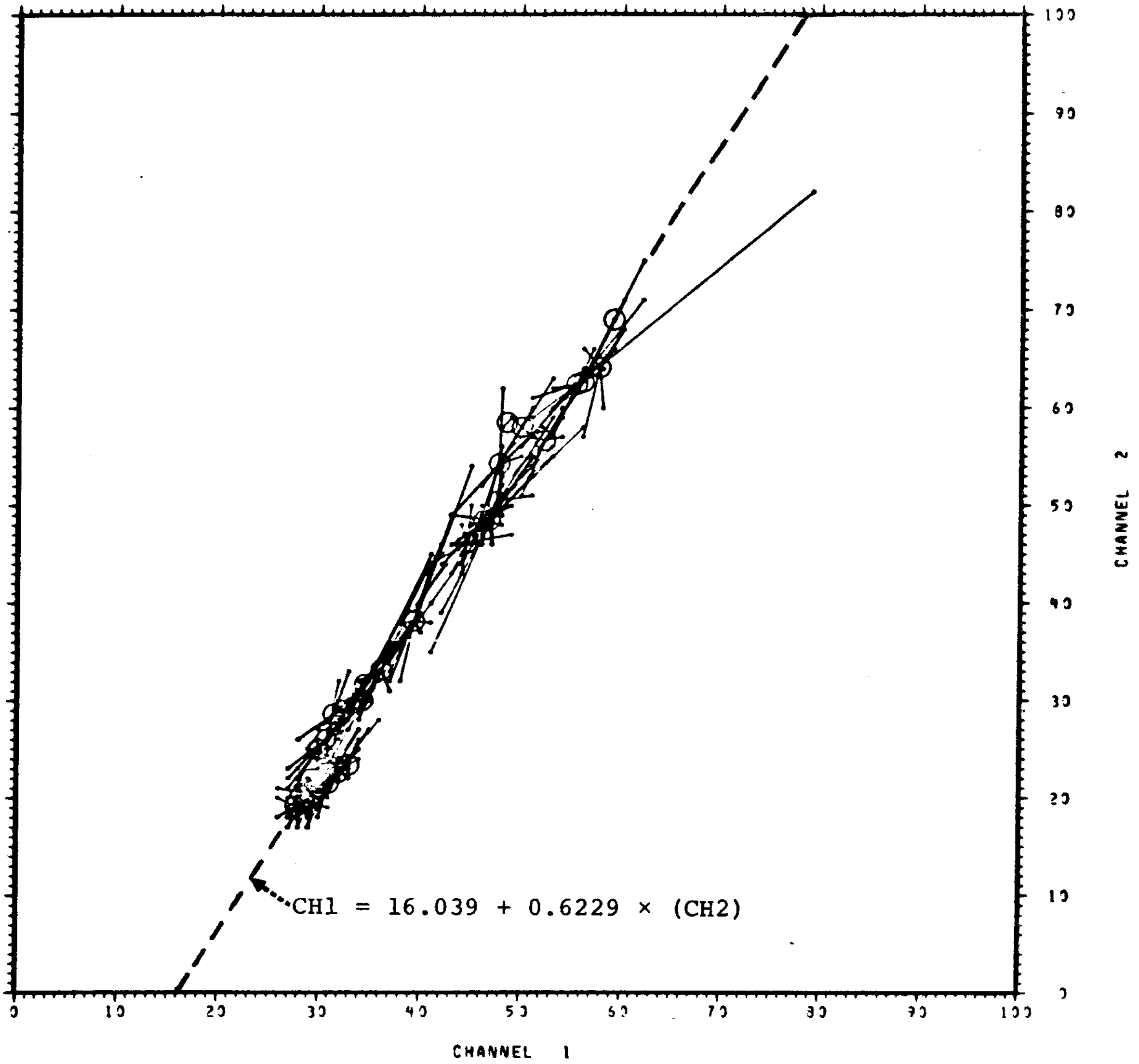
The two trends appeared in the figures which show the migration throughout the growing season of the data of one individual field. In the first trend, the data of pixels contiguous in the field tended to be contiguous in the data cluster. In many instances, the data of a specific pixel tended to have the same location within the cluster while the cluster migrated. Pixel migrations in data space were never perfectly parallel; but often a continuous, not completely random mapping of data from one date to the next seemed to be evident.

In the second trend, the migrations of cluster means between consecutive acquisition dates were never large and completely random. Ideally, the means of all field clusters of a given crop should be nearly equally displaced in direction and distance in data space on two consecutive days. The observed displacements were far from this ideal. The observed displacements tended to be oriented parallel to the bare soil axis with an average length of about 3 data units. For July 16 and 17, these displacements were predominantly oriented toward the origin, and the soybean displacements were slightly larger than those of corn. For June 10 and 11, the soybean displacements were more toward the origin, while those of corn were away from the origin.

TABLE II.— LEAST SQUARE FITS OF A PLANE TO THE DATA

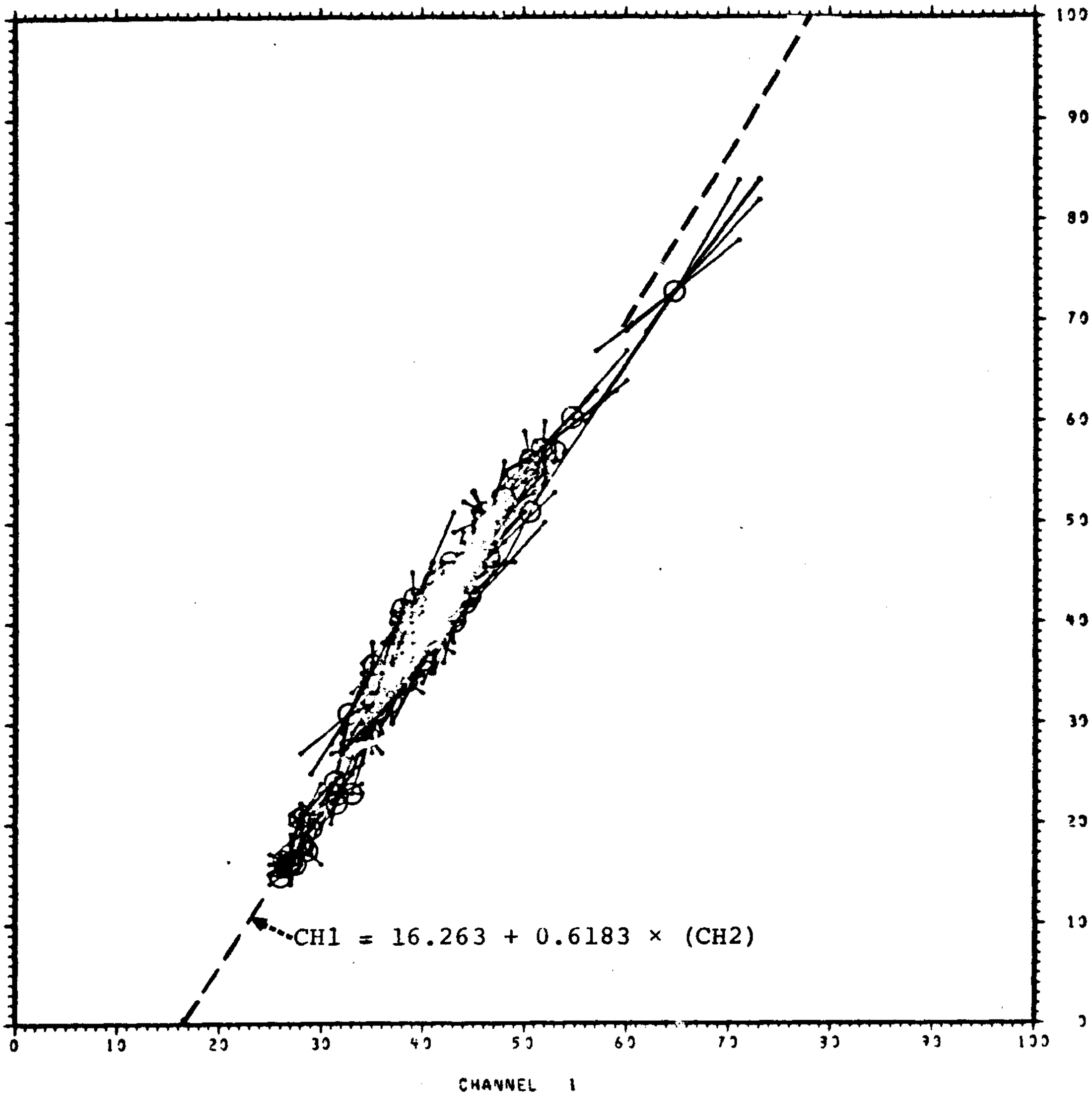
Space dimensions	Normal distance from 0 to plane	rms residual (a)	Axis intercept				Component of unit normal vector from 0 to plane			
			Channel 1	Channel 2	Channel 3	Channel 4	Channel 1	Channel 2	Channel 3	Channel 4
Corn										
3	0.2368	1.229	—	-1.103	0.4131	-0.2995	—	-0.2148	0.5733	-0.7907
4	3.426	1.195	11.787	-9.164	6.892	-4.713	0.2907	-.3739	.4971	-.7270
Soybeans										
3	0.05846	1.124	—	0.3503	-0.1074	0.07112	—	0.1669	-0.5444	0.8220
4	3.304	1.094	12.799	1580.42	-6.058	4.144	0.2582	.00209	-.5454	.7974
Corn and soybeans										
3	0.6044	1.2627	—	3.210	-1.102	0.7418	—	0.1883	-0.5484	0.8147
4	.8532	1.2626	40.556	4.860	-1.549	1.046	0.02104	.1755	-.5507	.8157

[a]Measured normal to the plane.



(a) Corn.

Figure 2.– Data from all six acquisition dates in the plane of channels 1 and 2.



(b) Soybeans.

Figure 2.— Concluded.

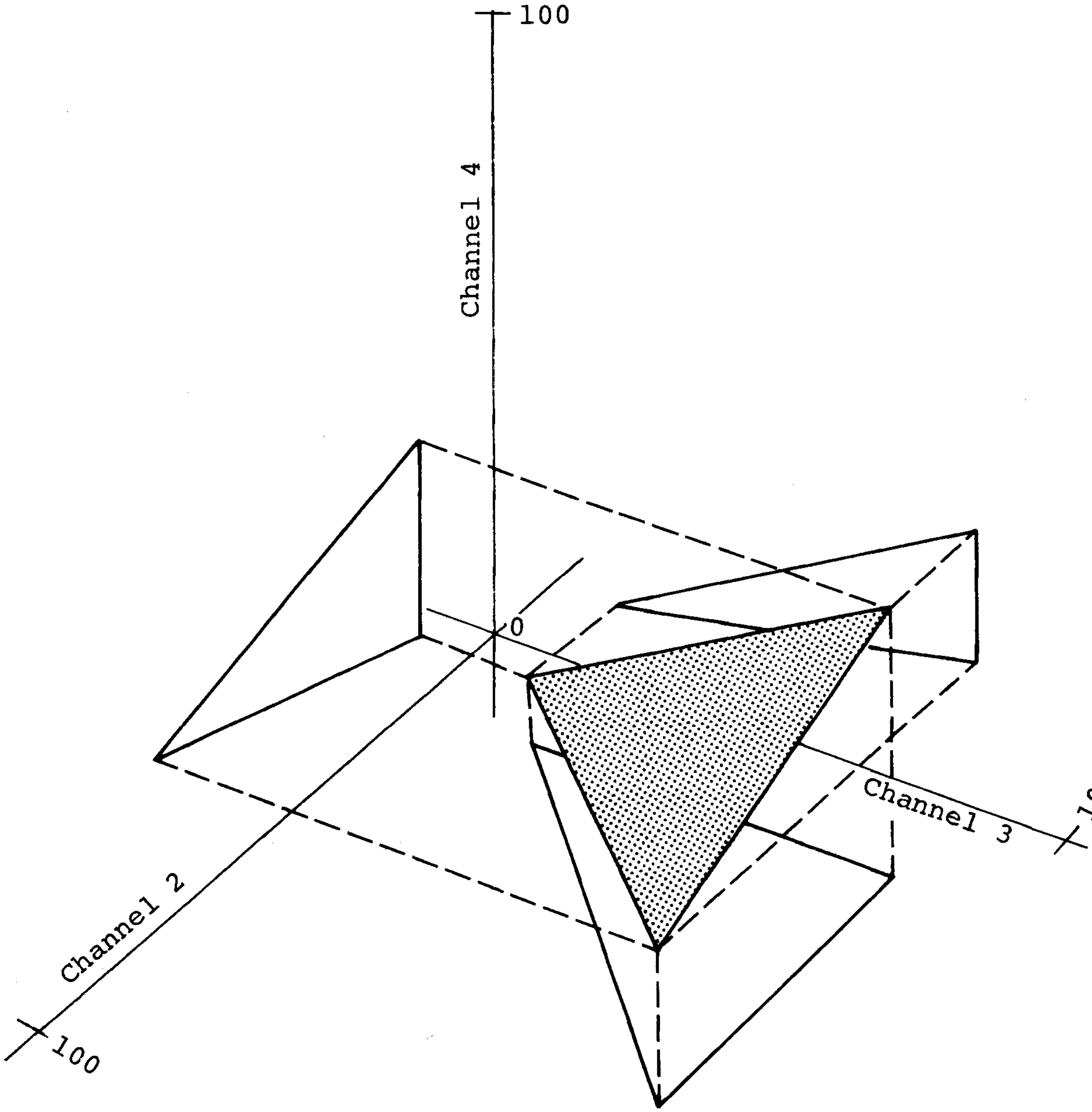


Figure 3.– Two-dimensional triangular plane in the space of channels 2, 3, and 4 to which the corn and soybean data were constrained closely on all six acquisition dates.

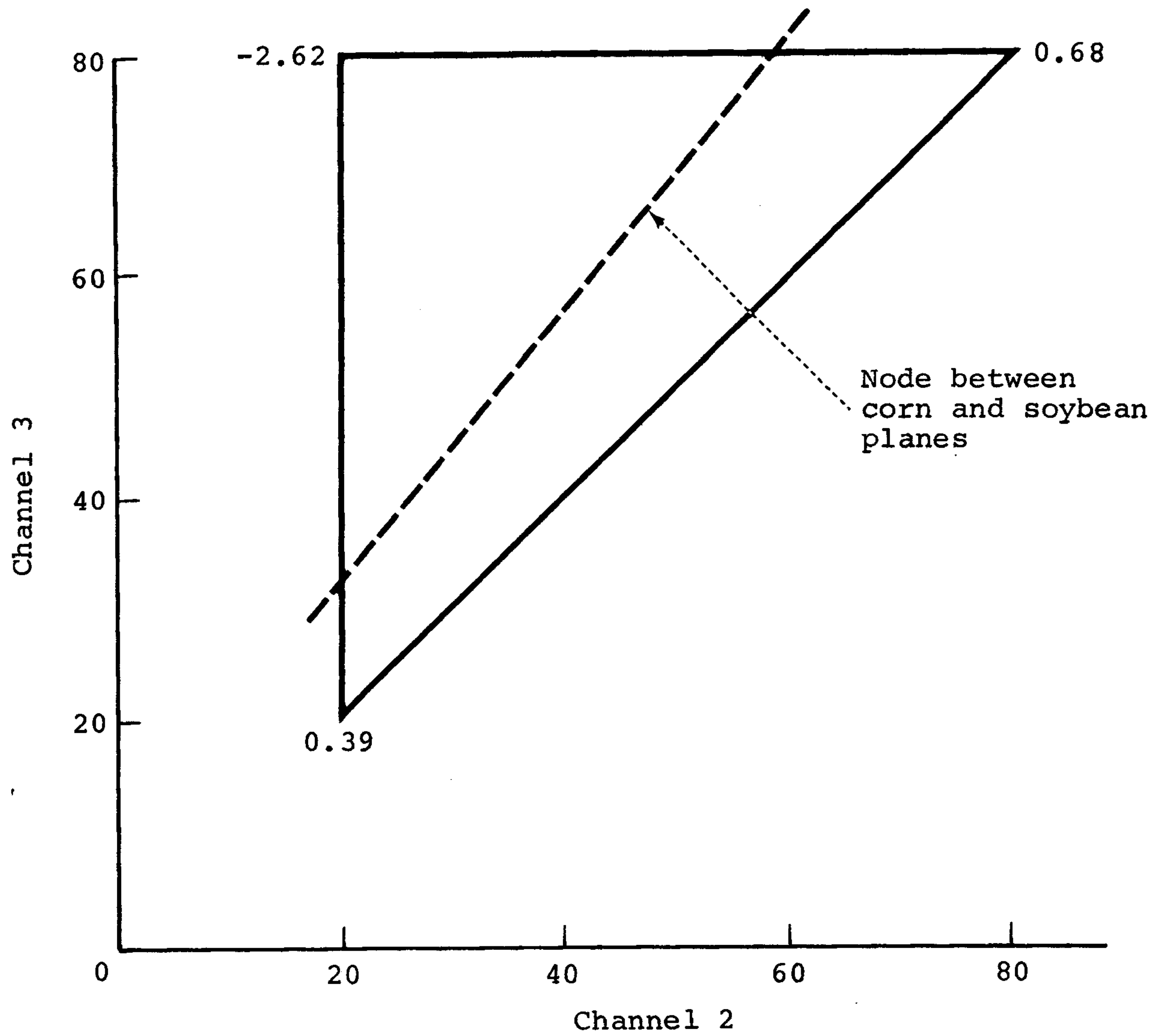
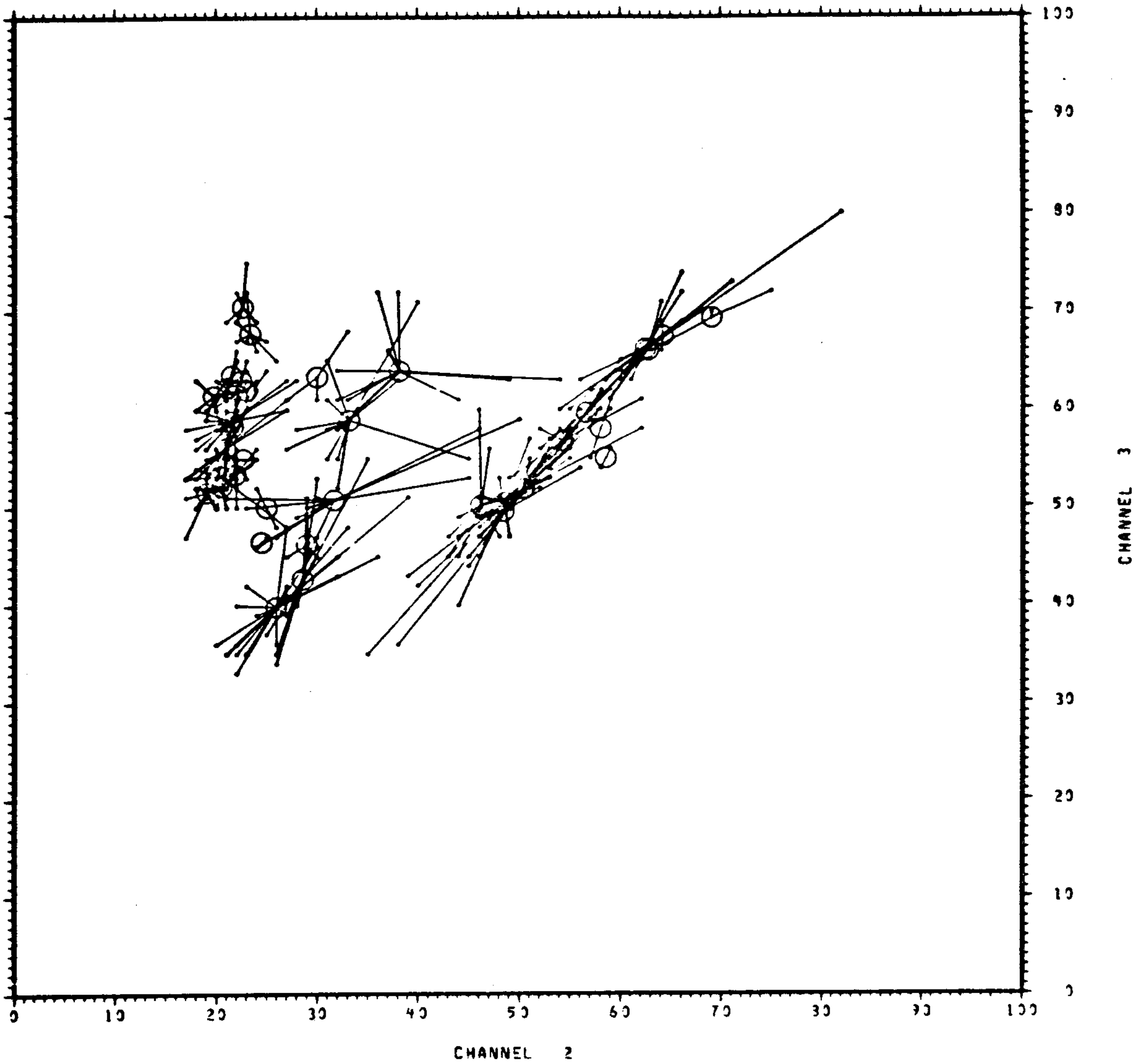
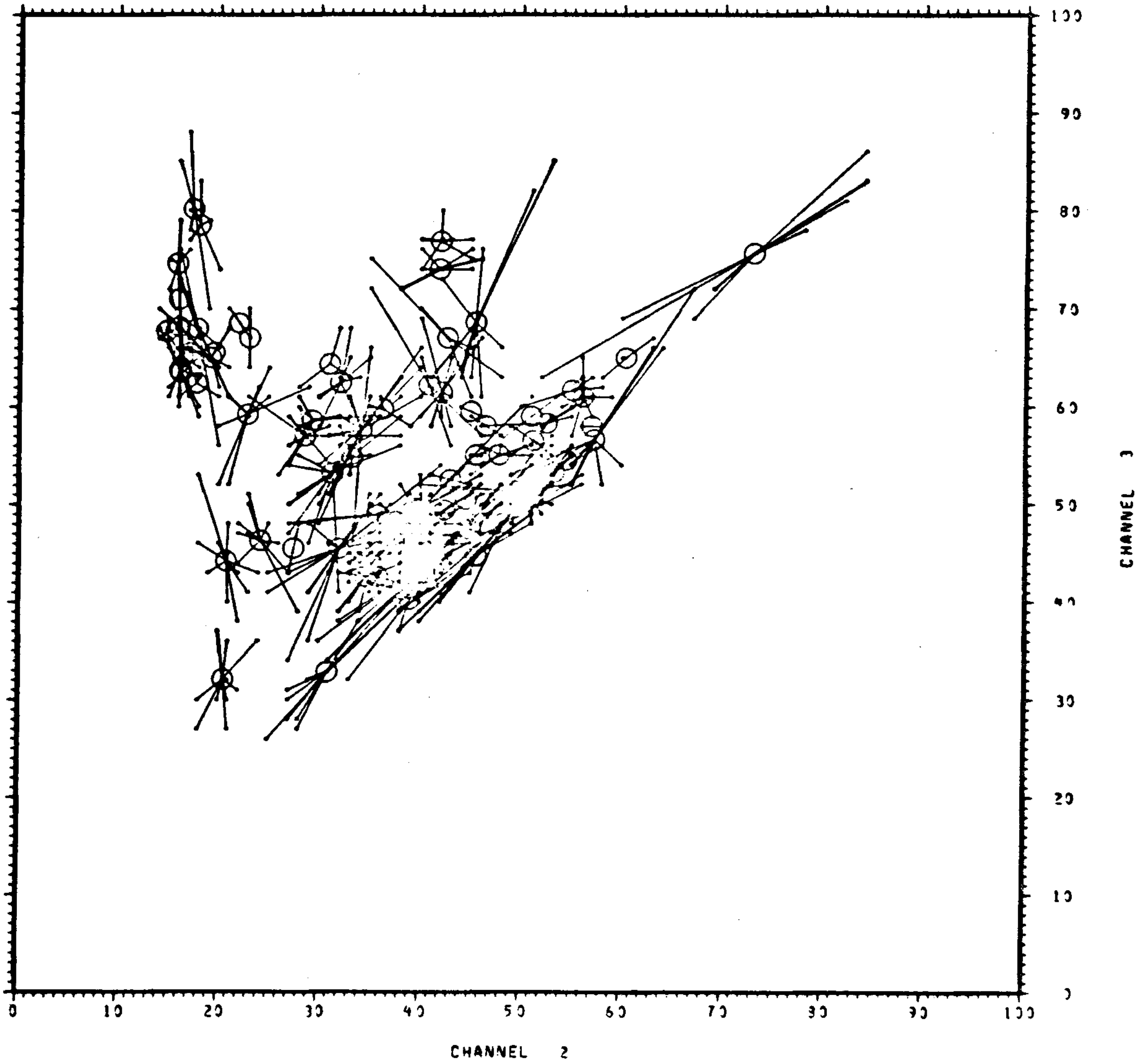


Figure 4.— Distances between the corn and soybean planes in three-dimensional data space measured perpendicularly at the vertices of the triangle, shown projected on the plane of channels 2 and 3.



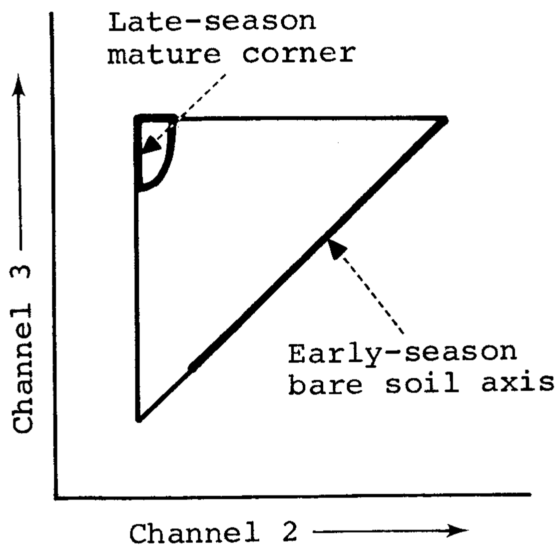
(a) Corn.

Figure 5.– Data from all six acquisition dates in the plane of channels 2 and 3.

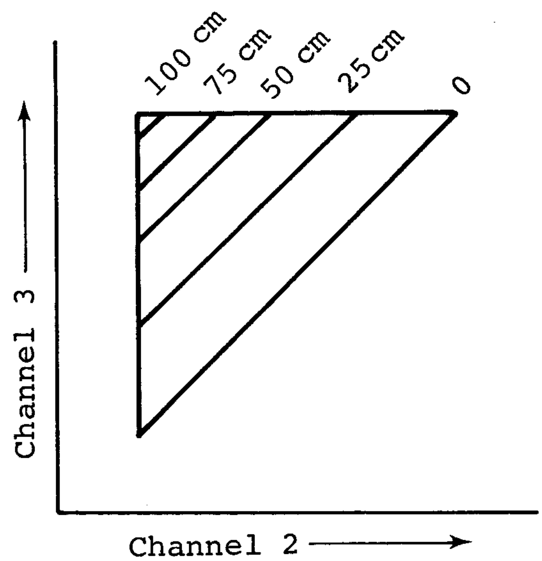


(b) Soybeans.

Figure 5.— Concluded.

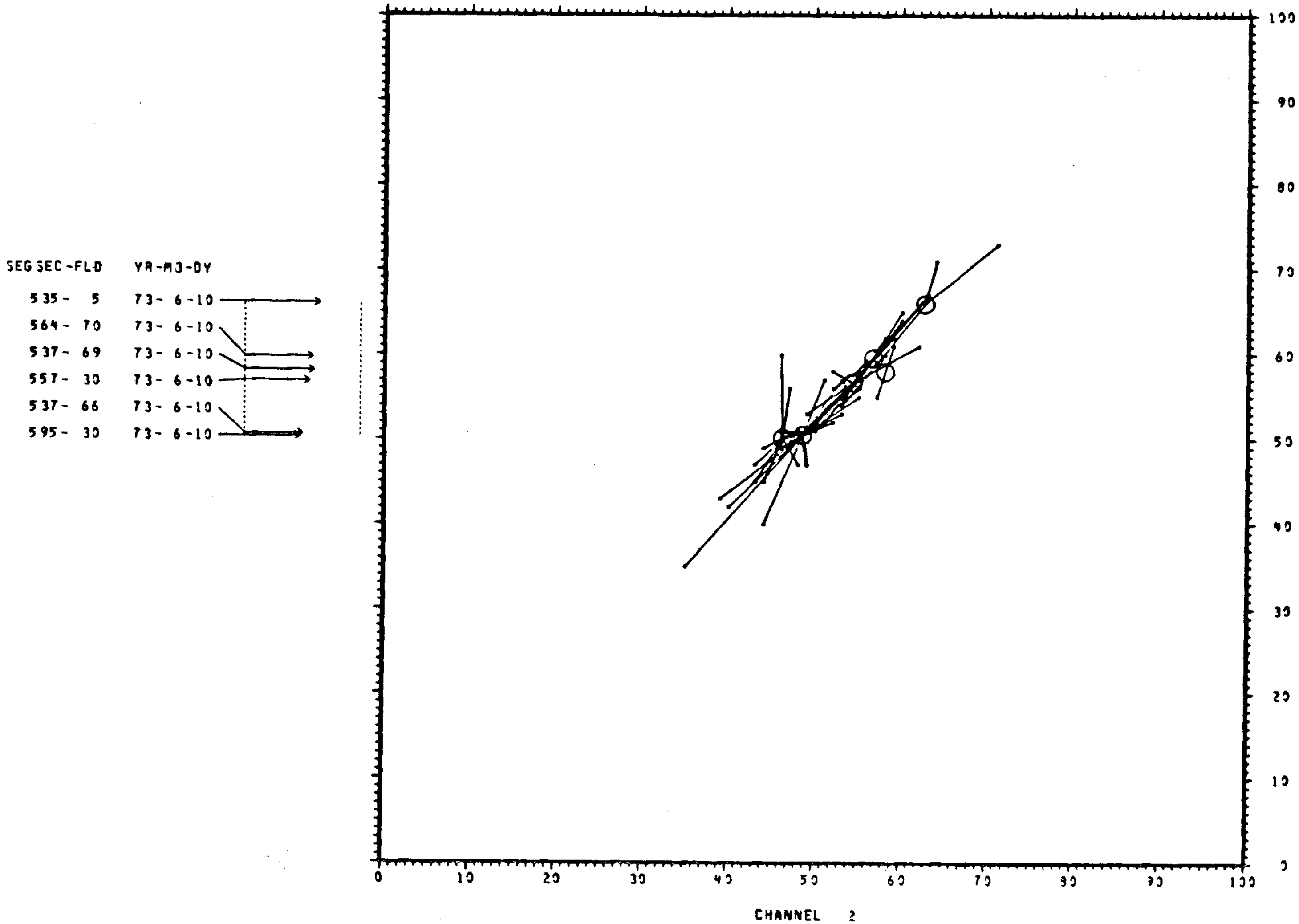


(a) Bare soil axis and mature corner on the triangle as projected on the plane of channels 2 and 3. This triangle applies to both corn and soybean data.



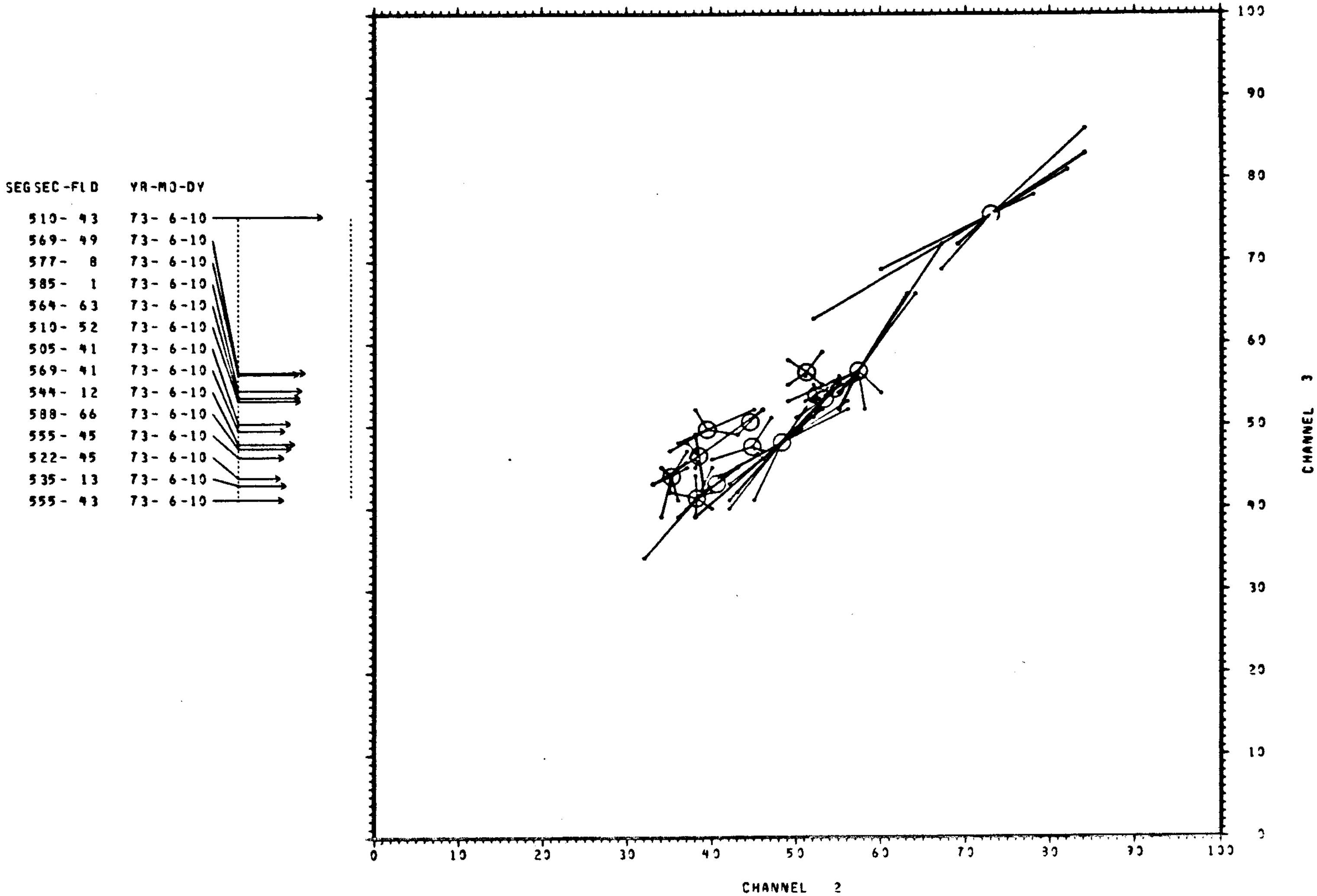
(b) Relationship between soybean data location within the triangle and plant height.

Figure 6.– Projections of the triangle on the plane of channels 2 and 3.



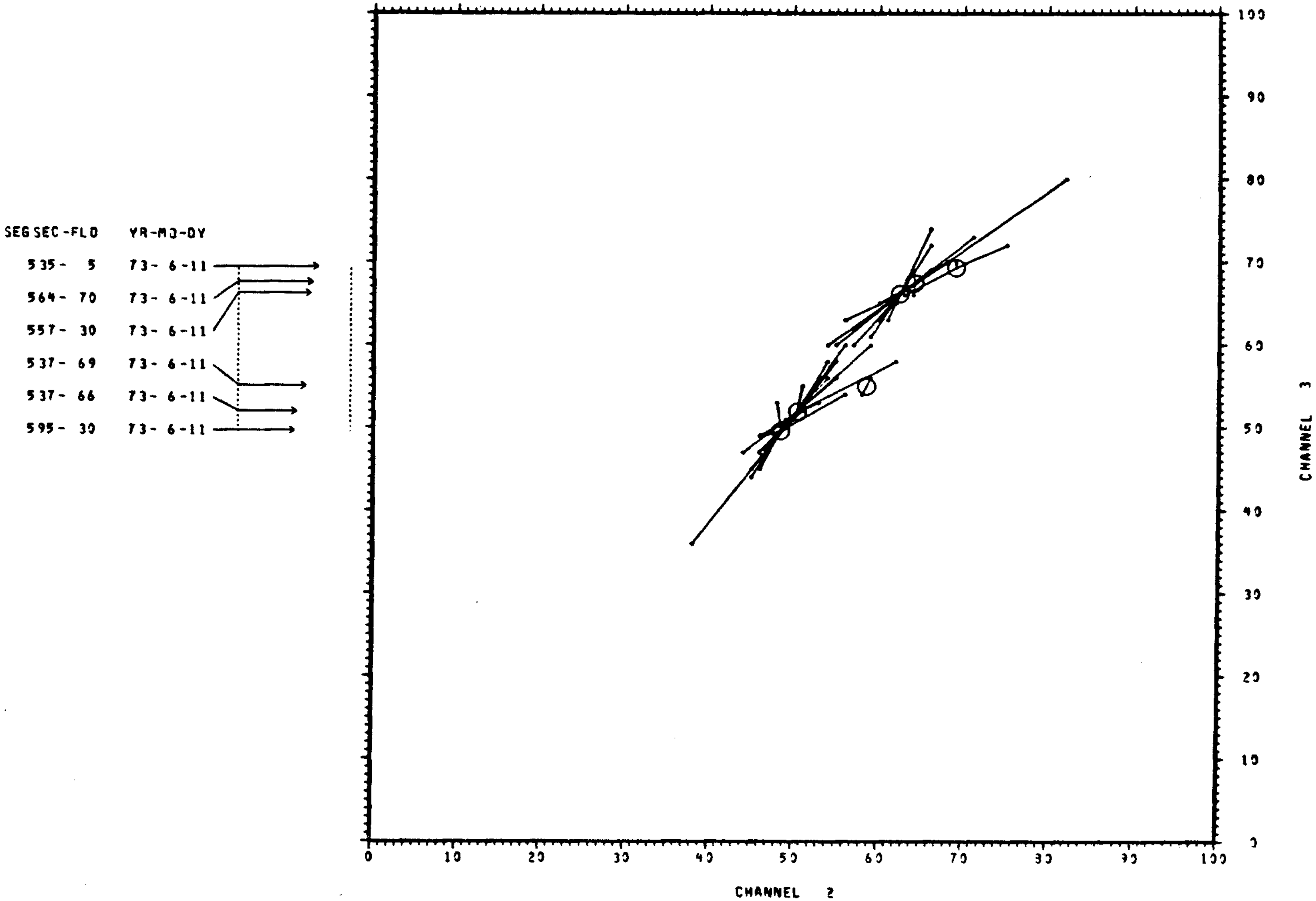
(a) Corn.

Figure 7.– June 10, 1973, data.



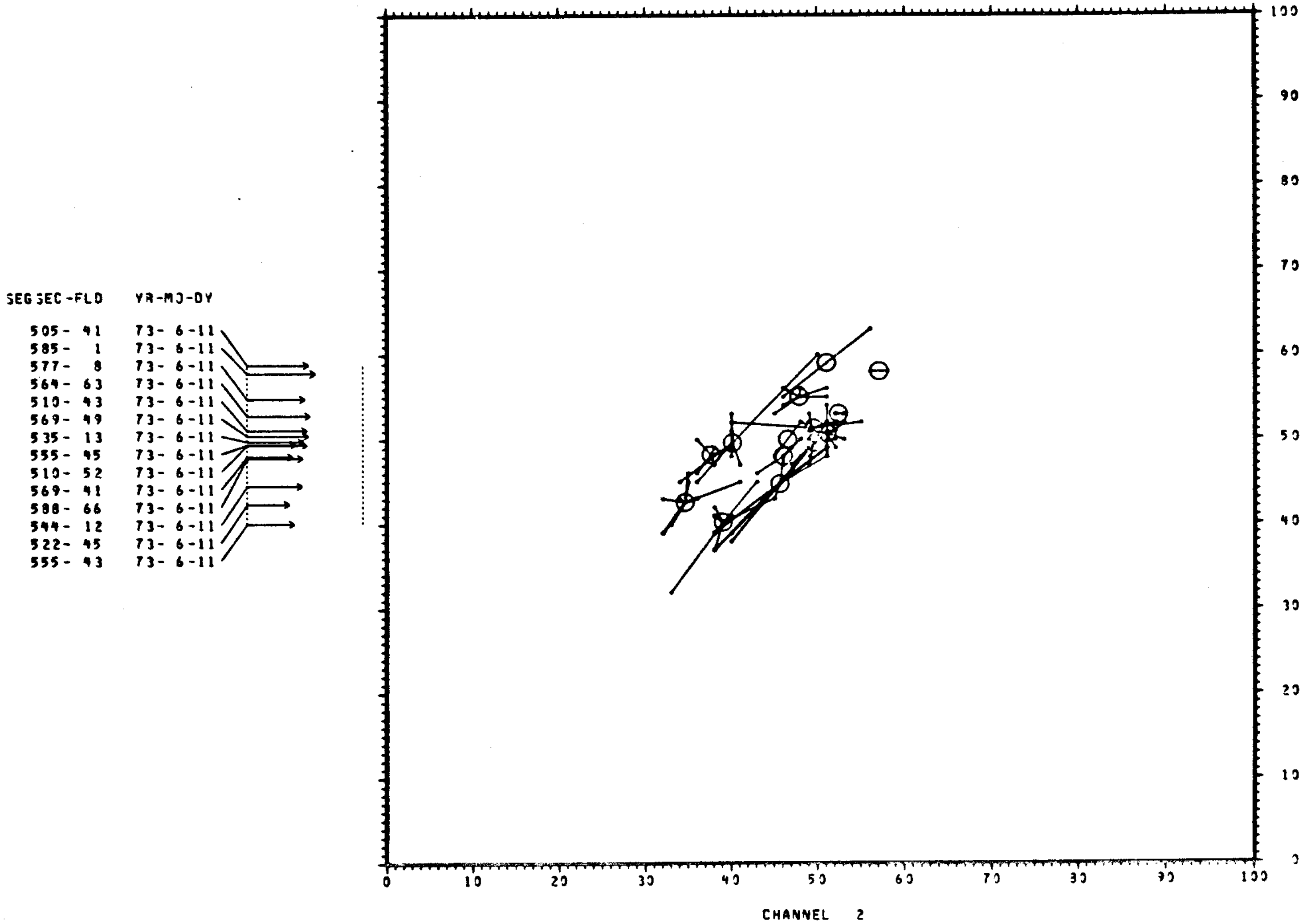
(b) Soybeans.

Figure 7.— Concluded.



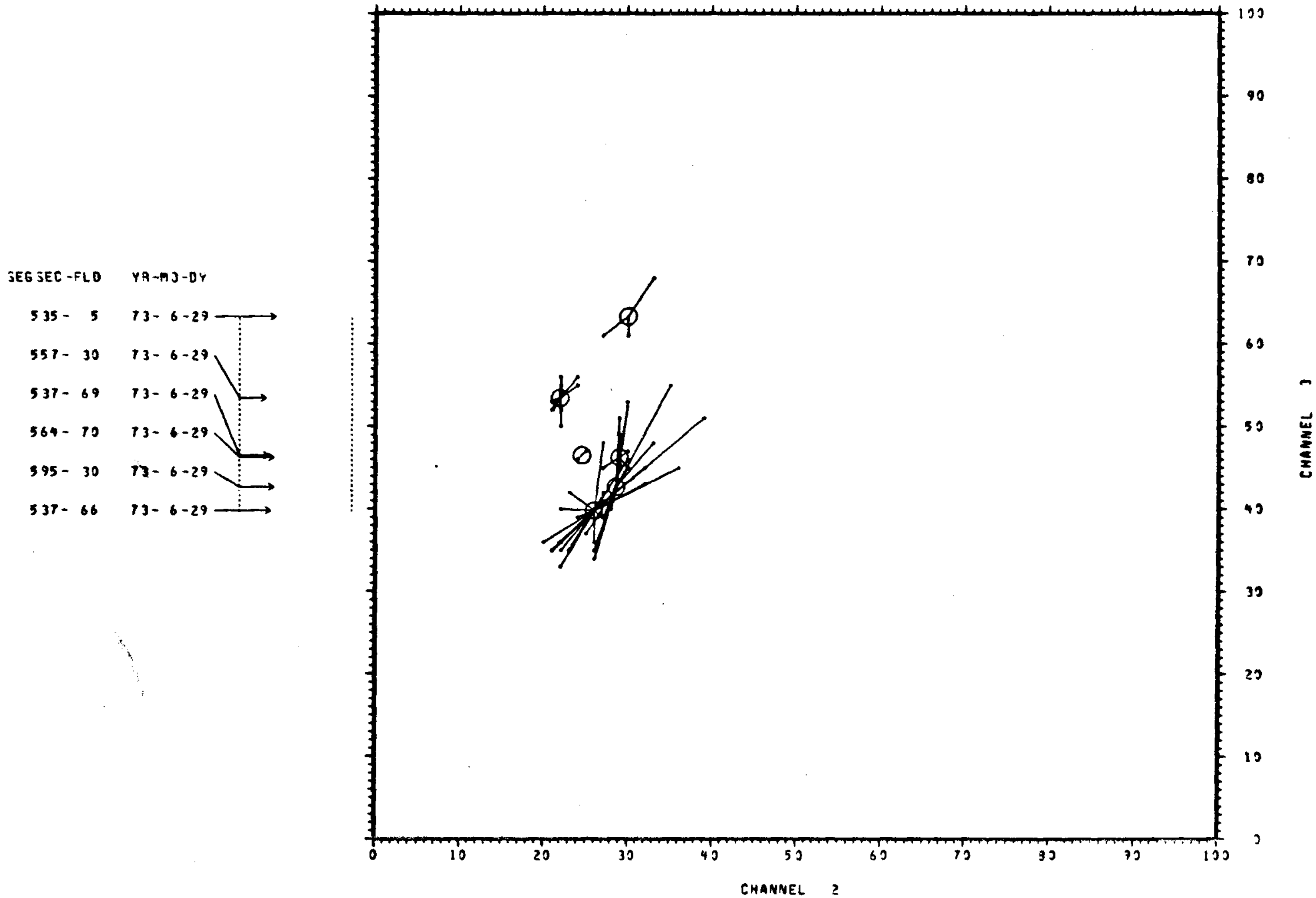
(a) Corn.

Figure 8.— June 11, 1973, data.



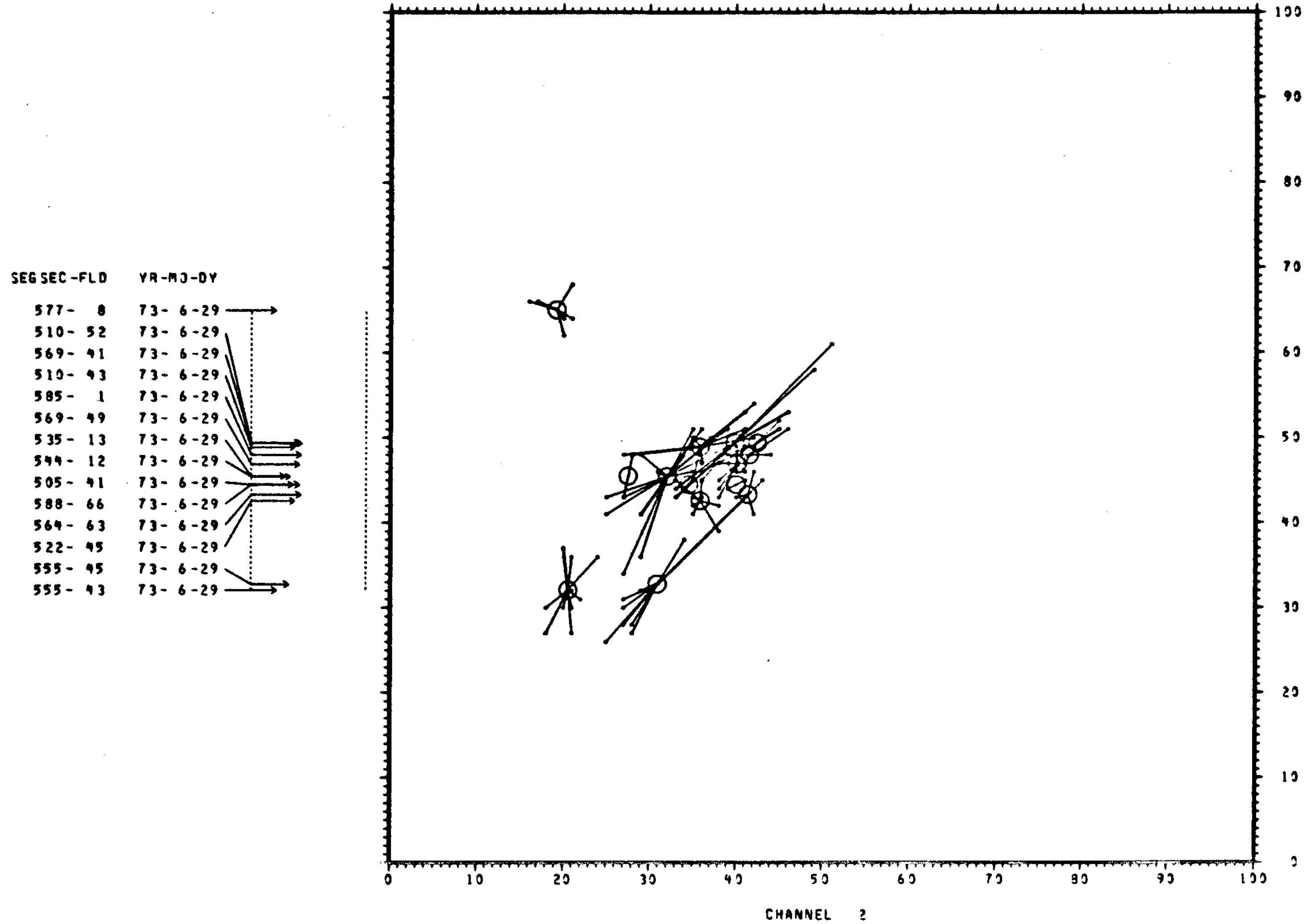
(b) Soybeans.

Figure 8.— Concluded.



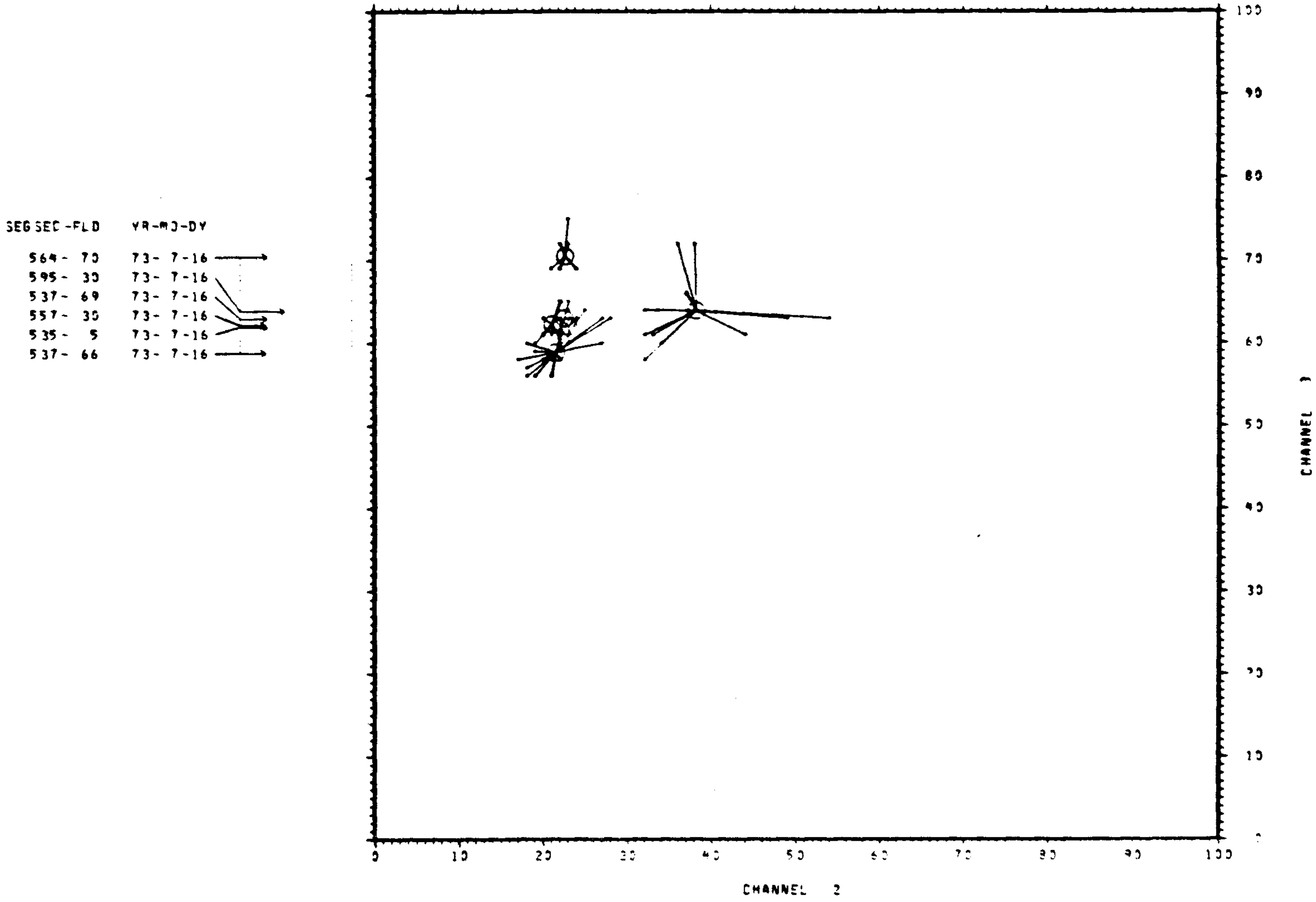
(a) Corn.

Figure 9.— June 29, 1973, data.



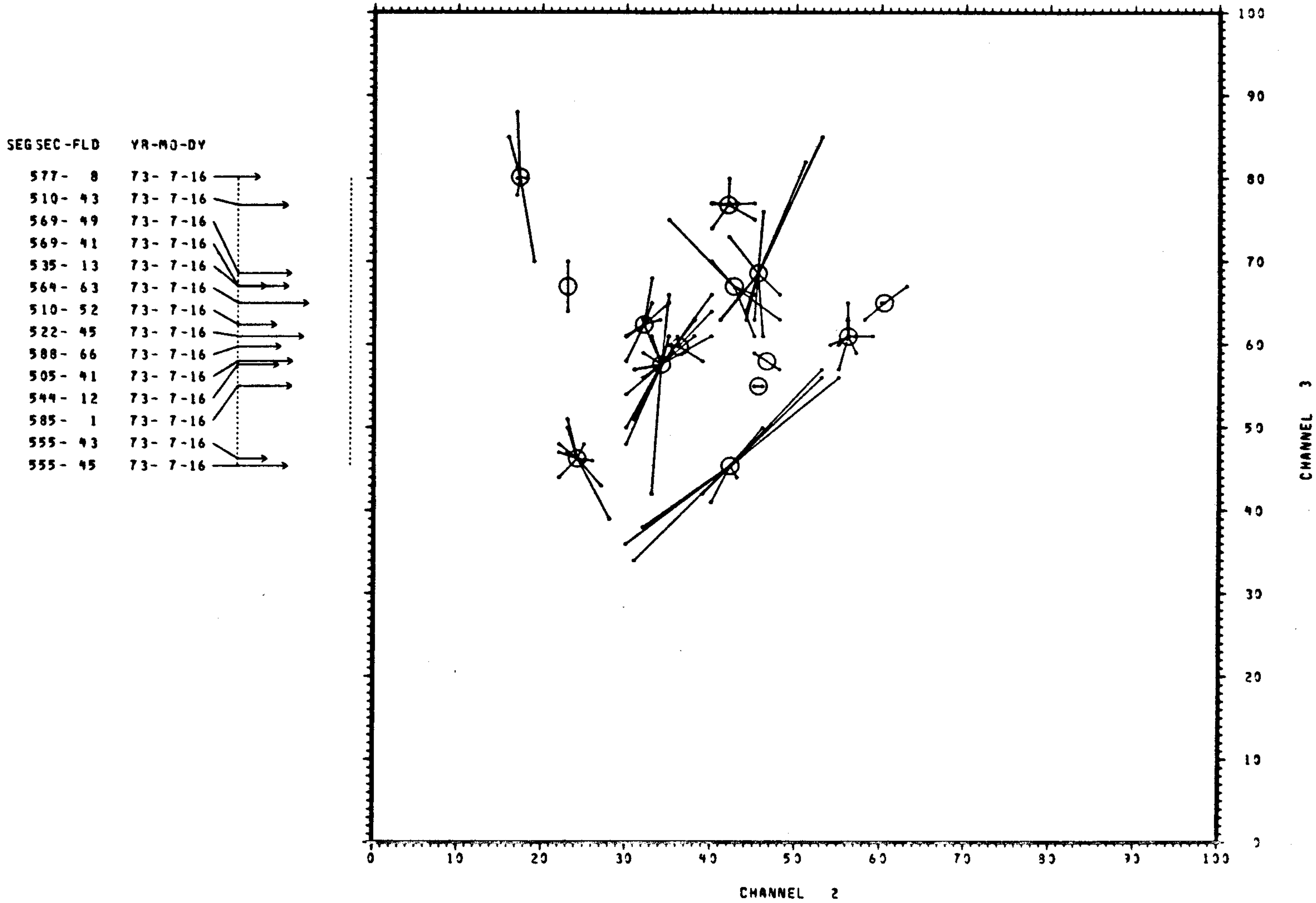
(b) Soybeans.

Figure 9.— Concluded.



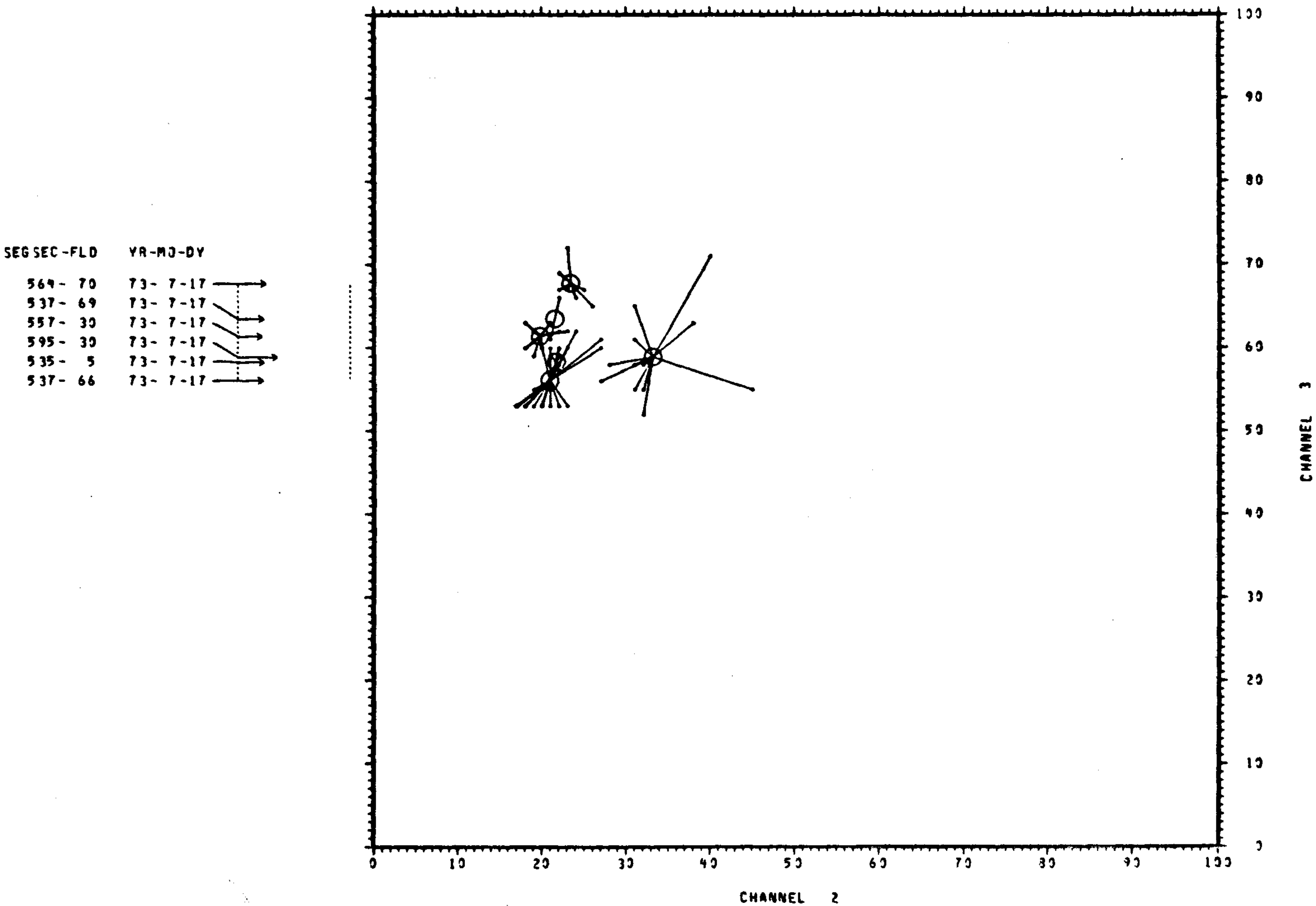
(a) Corn.

Figure 10.– July 16, 1973, data.



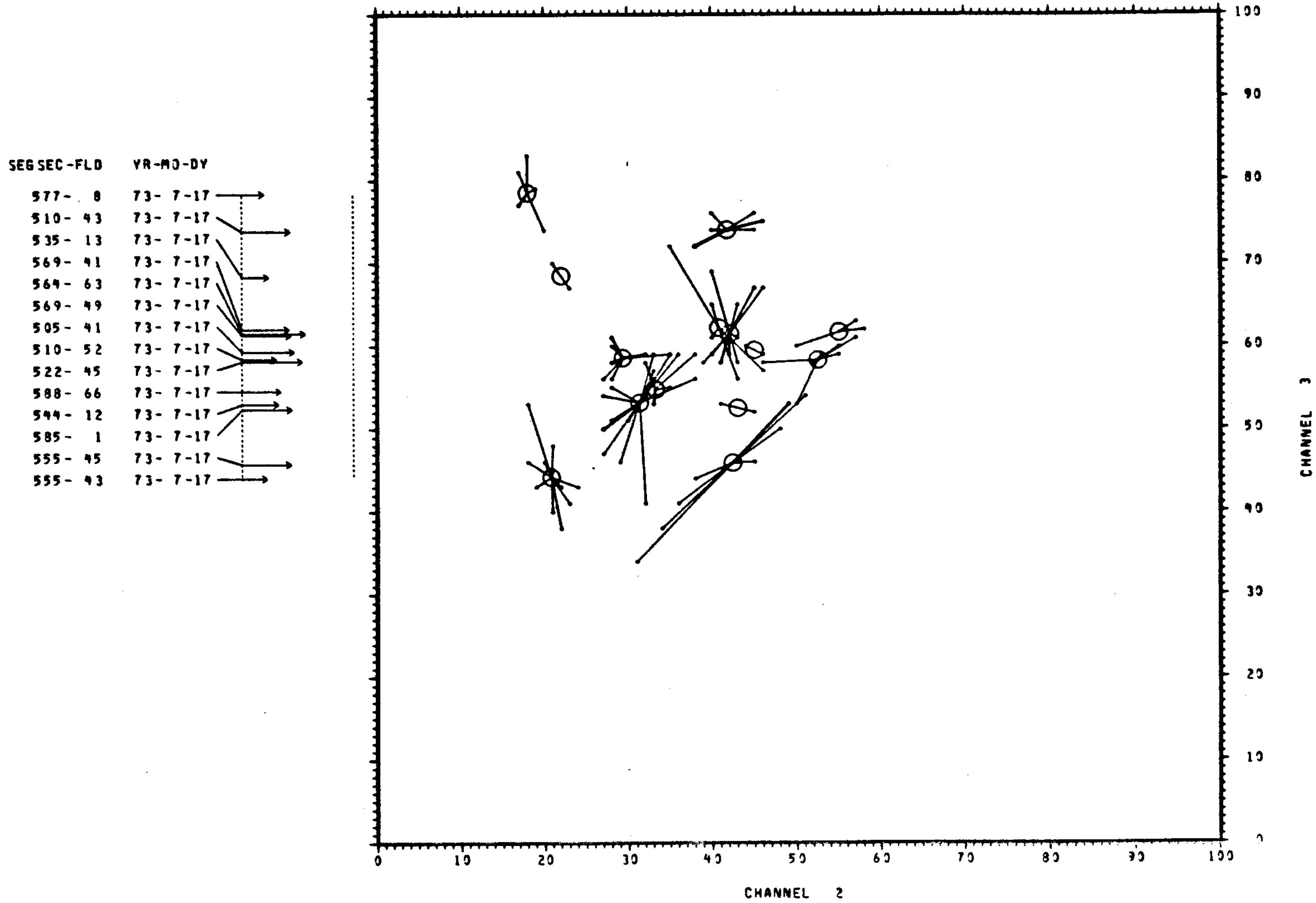
(b) Soybeans.

Figure 10.— Concluded.



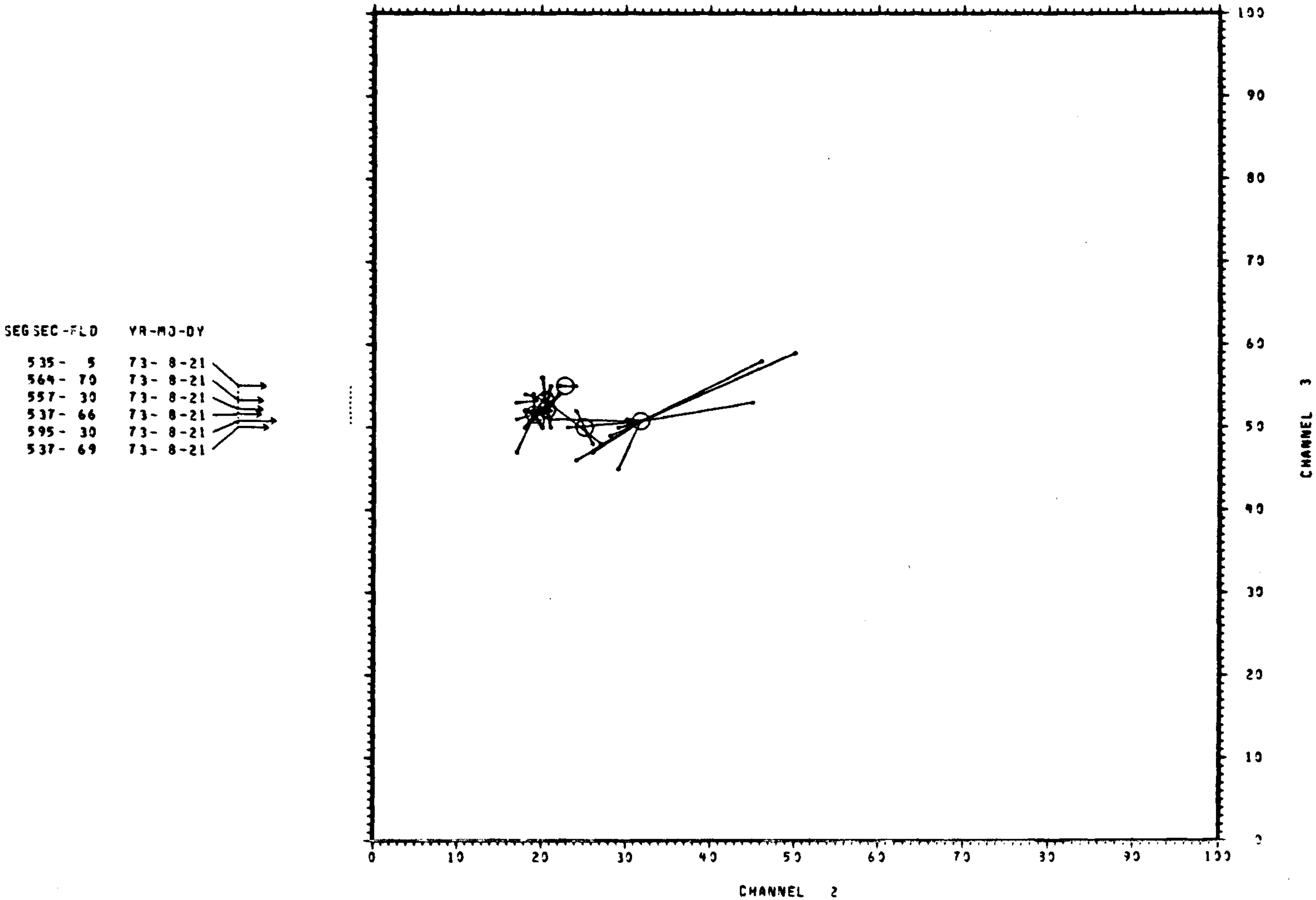
(a) Corn.

Figure 11.– July 17, 1973, data.



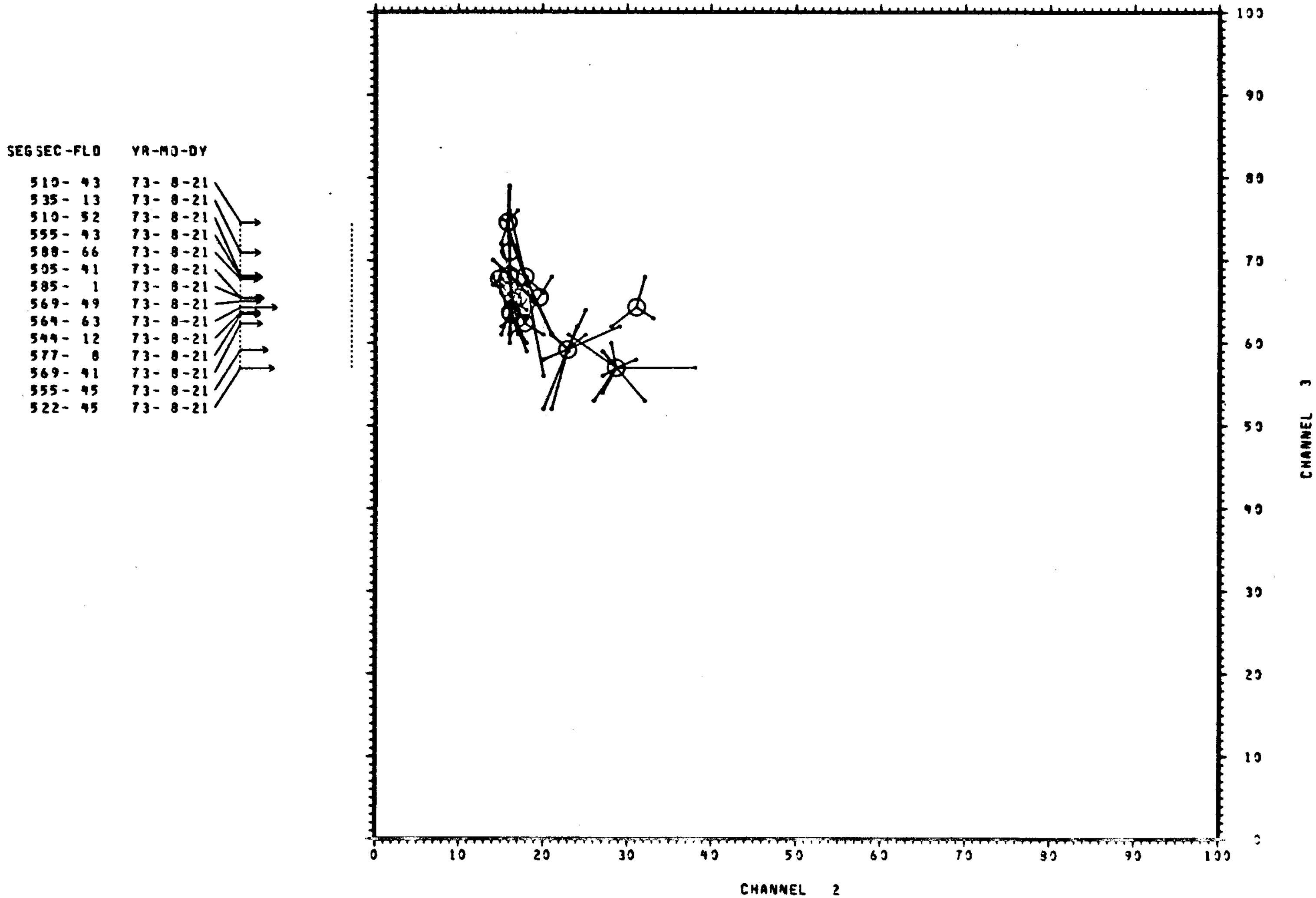
(b) Soybeans.

Figure 11.— Concluded.



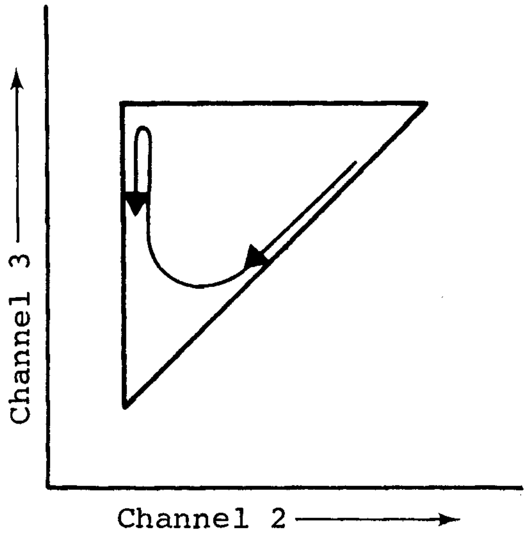
(a) Corn.

Figure 12.— August 21, 1973, data.

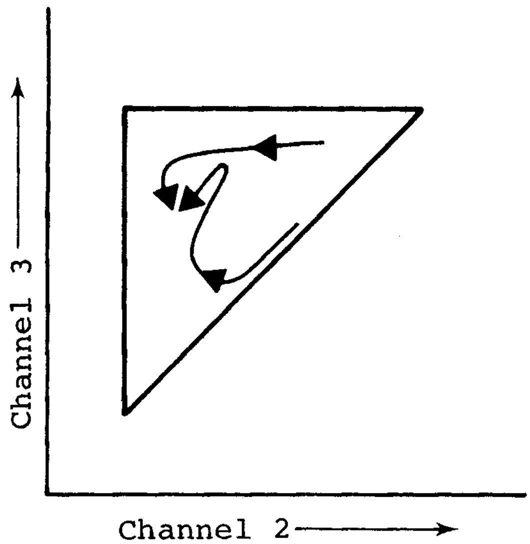


(b) Soybeans.

Figure 12.— Concluded.

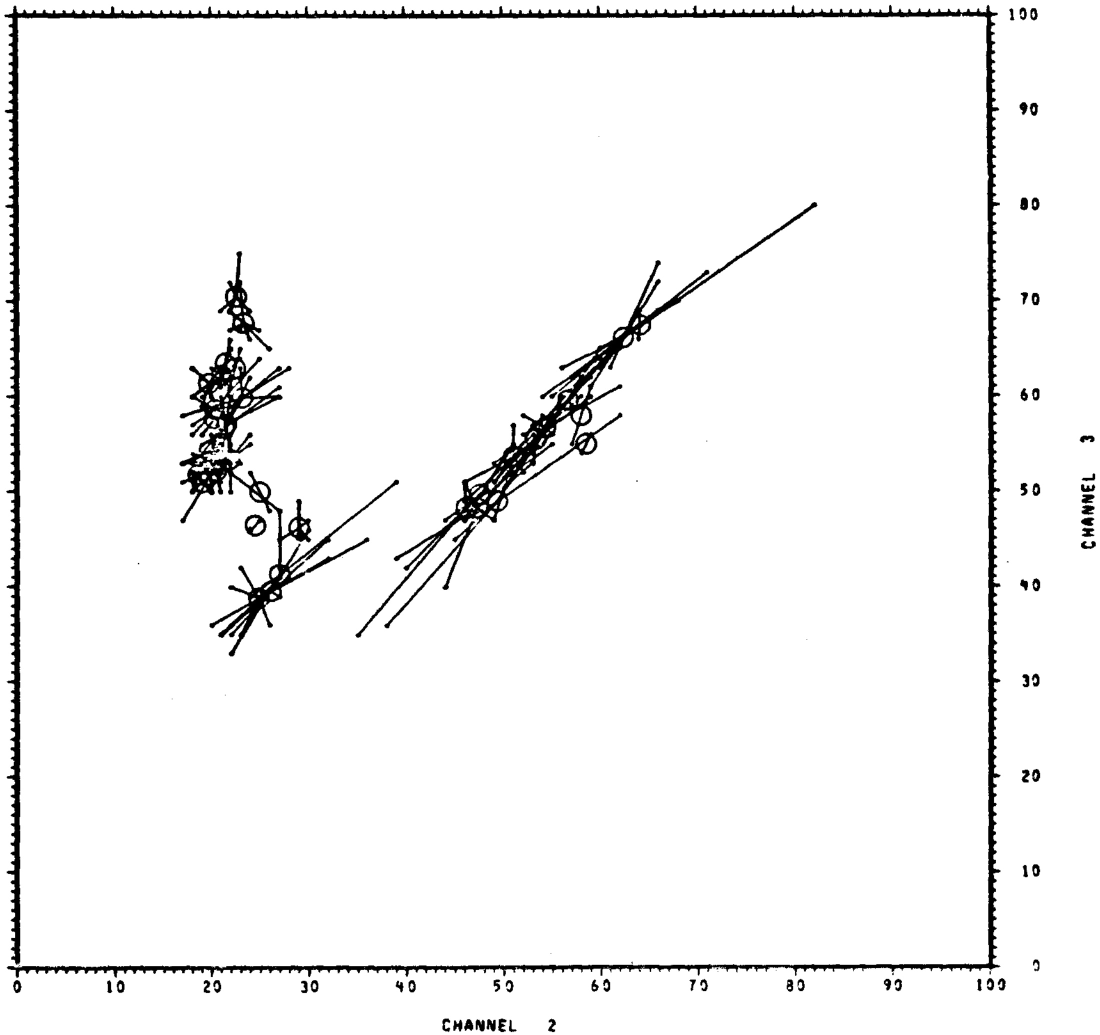


(a) Data from fields with a north-south row direction.



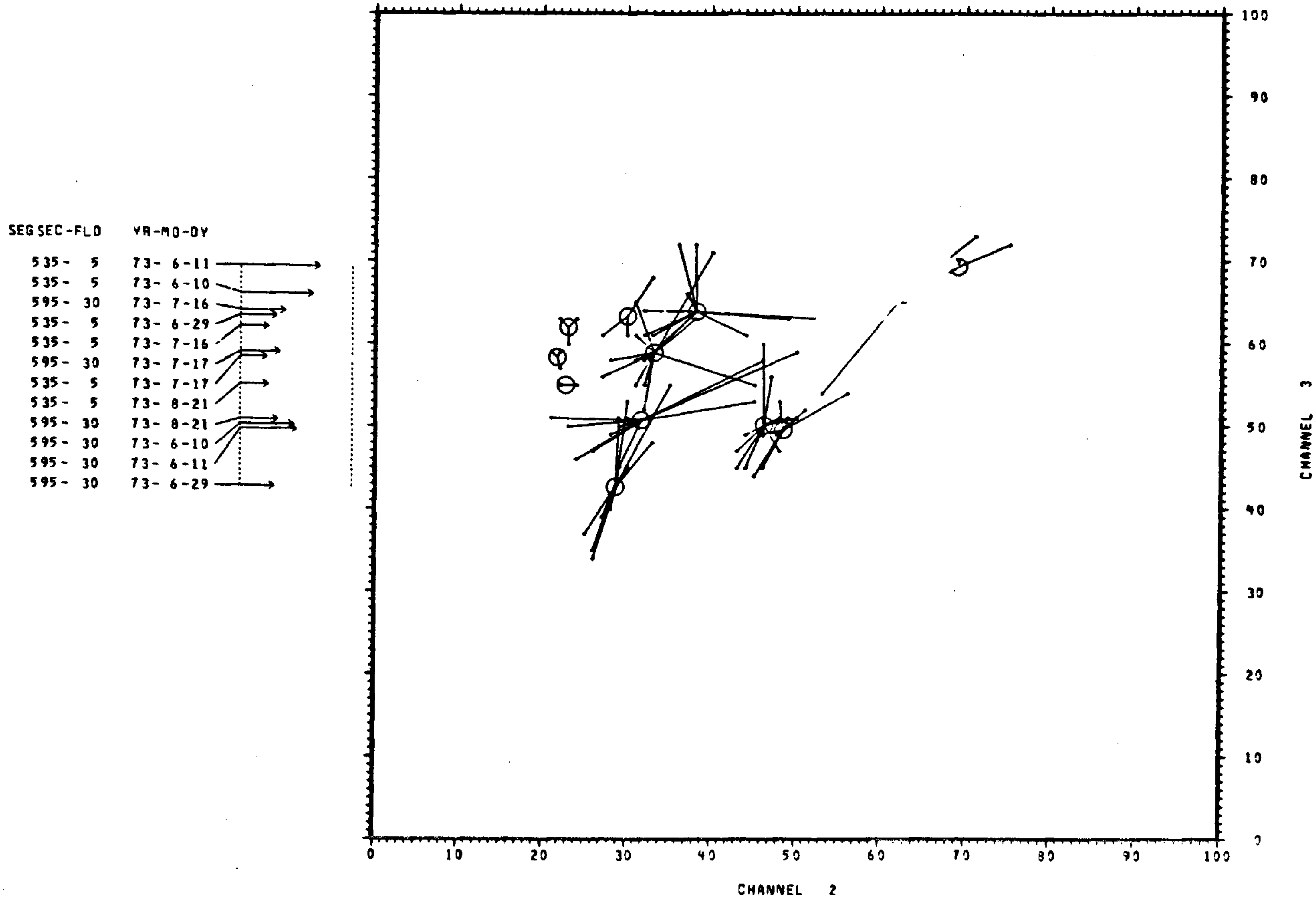
(b) Data from two separate fields with an east-west row direction.

Figure 13.– Corn data trajectories within the triangular region, shown projected on the plane of channels 2 and and 3.



(a) Data from cornfields with a north-south row direction.

Figure 14.– Corn data from all six acquisition dates in the plane of channels 2 and 3.



(b) Data from cornfields with an east-west row direction.

Figure 14.– Continued.

38

Blank

4.0 ROTATIONAL DATA TRANSFORMATION

The least square fits of planes in the three- and four-dimensional spaces were later used to define rotational transformations to coordinate systems better suited to showing data patterns relative to the triangle. This work consisted of defining sets of coordinate axes orthogonal to each other and to the unit vector normal to the fit plane. This procedure differed from the traditional principal components game because some transformed coordinate axes were constrained to various subspaces of untransformed channels. This constraint minimized the change in the triangle orientation in relation to coordinate axes and minimized negative data values.

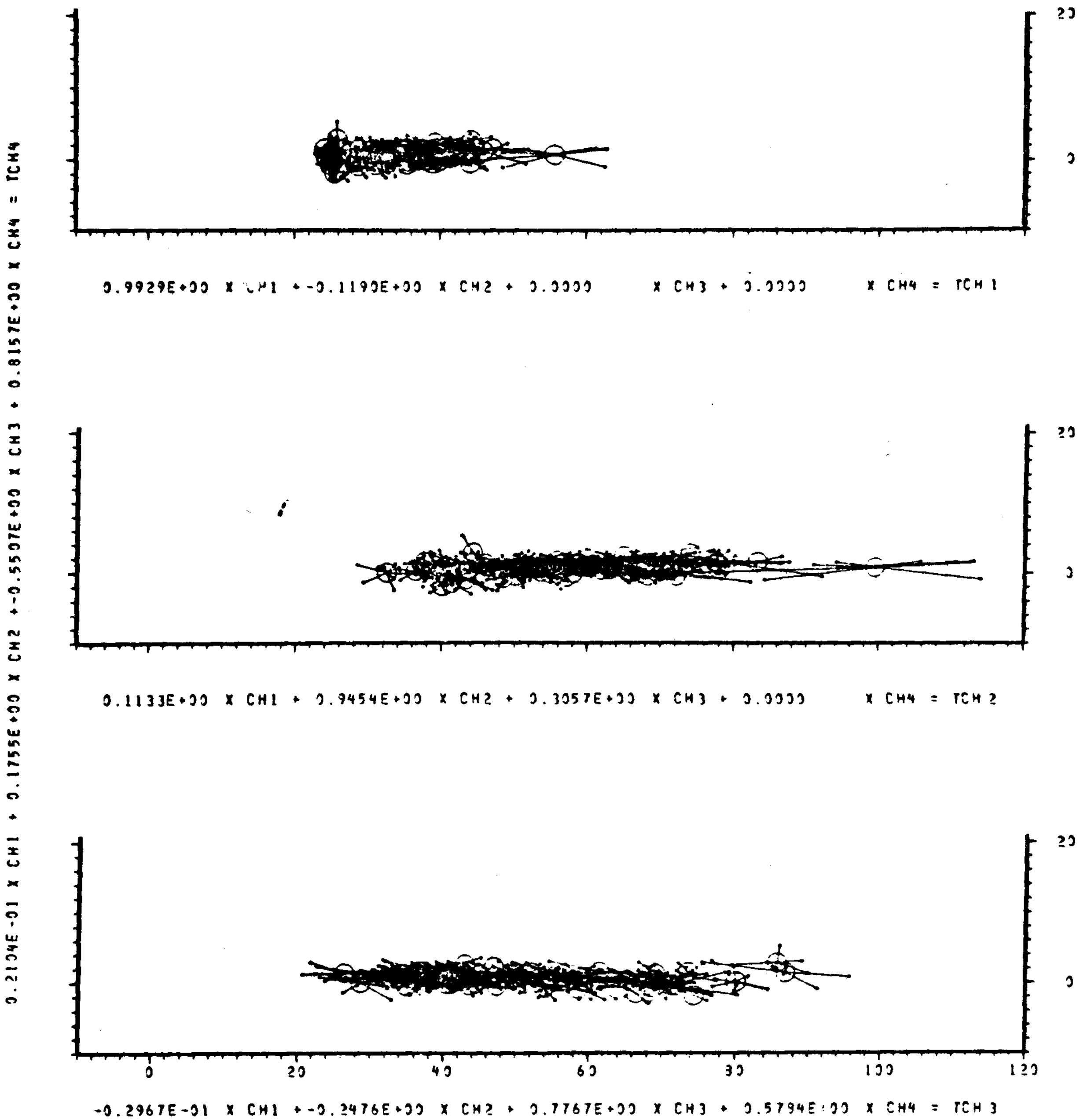
The rotational transformation matrix [R] based on the plane fit to the combined corn and soybean data in four-dimensional space is

$$[R] = \begin{bmatrix} 0.9929 & -0.1190 & 0 & 0 \\ .1133 & .9454 & .3057 & 0 \\ -.0297 & -.2476 & .7767 & .5784 \\ .02104 & .1755 & -.5507 & .8157 \end{bmatrix} \qquad (3)$$

where [R][CH] = [T].

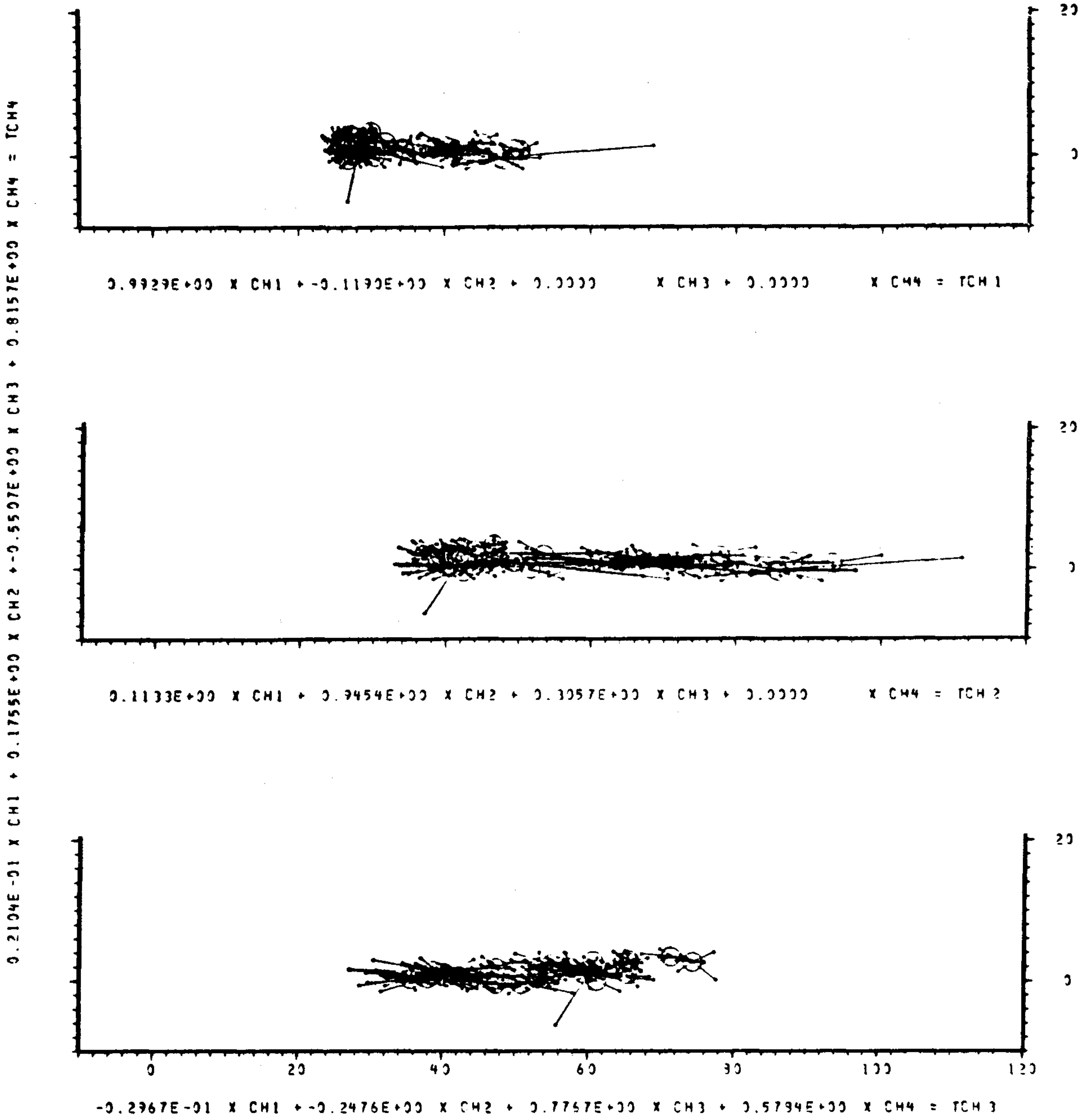
The [CH] is a four-dimensional vector of data values in channels 1, 2, 3, and 4, in that order. The [T] is the four-dimensional vector of transformed coordinates. The bottom row of this matrix, defining the fourth transformed coordinate TCH4, is the normal vector to the plane fit (table II).

This transformation was applied to the corn and soybean data separately, and SIGMIG was used to illustrate the results. The figures showing data on planes containing pairs of the first three transformed axes (TCH1 and TCH2, TCH1 and TCH3, and TCH2 and TCH3) did not differ much from those containing the corresponding original untransformed channels and are not shown here. These differences appeared to be slightly different perspectives of a fixed data distribution in three-dimensional space. However, the figures showing data on planes containing the fourth transformed axis TCH4, perpendicular to the plane fit upon which the transformation was based, differed significantly from the corresponding figures containing untransformed channel 4. These figures showing edge-on views of the corn and soybean data surfaces in four-dimensional space illustrated the flatness of these data distributions. These figures are shown together in figures 15(a) for corn and 15(b) for soybeans.



(a) Corn.

Figure 15.— Data from all six acquisition dates in the transformed four-dimensional coordinate system. The TCH4 axis is perpendicular to the three-dimensional plane (TCH1, TCH2, and TCH3), which was fit to the corn and soybean data combined.



(b) Soybeans.

Figure 15.— Concluded.

5.0 CONCLUSIONS

Graphic study of Fayette County corn and soybean training data revealed some significant data patterns.

The researcher found that both the corn and soybean data were tightly constrained to the same two-dimensional plane in four-dimensional data space and throughout the growing season moved within a triangle on this plane.

Throughout the growing season, the corn data with a north-south row direction migrated along a specific path within the triangle, down the bare soil axis, over and up to the mature corner, and down again. The fewer but more erratic corn data with an east-west row direction went into the upper central region of the triangle, which the north-south data avoided.

During the growing season, the soybean data migrated across the triangle from the bare soil axis to the mature corner, without turning down. While the distance perpendicular to the bare soil axis was closely related to plant height, the data positions along parallels to the bare soil axis seemed erratic. The researcher noted no relationship with crop row direction. In mid-July when soybean heights covered a wide range, the data were spread over the entire triangle. This type of data distribution does not facilitate accurate discrimination between classes by traditional methods.

Data migration distances of cluster means during 1 day were small, but the spread in the lengths and directions of these 1-day migrations showed much randomness. The researcher noted slight differences between corn and soybean 1-day migration patterns.

However, the revelation of the data patterns raises these questions:

1. Is all vegetation constrained to the same surface?
2. Are all classes, including nonvegetation, constrained to the same surface?
3. Do consistent local curvatures exist in the surfaces?
4. Is all vegetation constrained within the triangle on the surfaces?
5. Are all classes, including nonvegetation, constrained within the triangle on the surfaces?
6. Are the characteristics of the surfaces and regions for different classes independent of geographic location of the data source?
7. How dependent are these characteristics on variables such as illumination and viewing geometry and atmospheric attenuation?
8. Would more east-west corn data reveal a consistent migration pattern?
9. What are the data migration patterns of corn with row directions other than north-south and east-west?
10. Why does row direction so profoundly affect corn data migration patterns?

11. Why do the corn data turn down after reaching the mature corner?

12. How dependent are the patterns of corn data migration on illumination and viewing geometry, atmospheric attenuation, and geographic location of the data source?

13. What determines soybean data position along parallels to the bare soil axis?

14. What are the data migration patterns of other vegetation types during the growing season, and what influences these patterns?

15. Is the randomness in 1-day migrations of cluster means caused by inconsistent data registration, and/or local variations in atmospheric attenuation, or other factors to be identified?

The observed data patterns indicated that the data set did not behave in data space as traditionally expected. The data of different classes did not form cleanly separated clusters or groups of clusters upon which accurate discrimination could be based. The researcher noted differences in the data patterns of classes, but the traditional statistical approaches to classification are not designed to use these differences in class discrimination.

A similar graphic analysis should be performed on data of the Large Area Crop Inventory Experiment from widely separated fields of various classes with intensive ground truth during many passes throughout a growing season. This analysis would show whether the classification methods used can cope with the actual data behavior in data space. If these methods are inadequate, possibly they could be modified to discriminate between the observed unconventional differences in data patterns of different classes.