Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Applications
Division

NASS Staff Report
Number SRB-89-09

August 1989

THE USE OF METEOROLOGICAL SATELLITE DATA IN ASSESSING CROP CONDITION

Wendell Wilson

THE USE OF METEOROLOGICAL SATELLITE DATA IN ASSESSING CROP CONDITION by Wendell Wilson, Research and Applications Division, National Agricultural Statistics Service, U.S. Department of Agriculture, NASS Staff Report No. SRB 89-09 .

ABSTRACT

In this pilot level research study, relationships between crop condition and polar orbiting meteorological satellite data were investigated for the 1984 corn and soybean crops. The 1984 forecasts and final estimates of corn for grain and soybean yield per harvested acre were used as State level measures of crop condition. Regression analyses were employed to understand the State level relationships of a crop's yield to its satellite vegetative index for ten States. The ten States are North Dakota, South Dakota, Minnesota, Iowa, Illinois, Indiana, Ohio, Kentucky, and Tennessee. Linear regression relationships for corn and soybeans existed at the State level, with coefficients of determination (R^2's) of .94 and .85 for final yield, respectively. This methodology was applied during the 1988 crop season, under drought conditions. The indices were strongly correlated to the official Agricultural Statistics Board estimates throughout the corn and soybean forecast season for 1988.

KEY WORDS

Meteorological satellite data, vegetative indices, crop conditions.

Mr. Wilson took a position outside the Department of Agriculture when this paper was at a draft stage and hence the paper was not subject to the standard peer review process. The content of the draft has not been intentionally altered but the body (not the appendices) has been edited for brevity by Bruce Eklund and George Hanuschak. The report represents several years of extensive research effort by Mr. Wilson and thus is being documented. This paper is for limited distribution outside the U.S. Department of Agriculture. The views expressed herein are not necessarily those of the Agency (NASS), or the Department (USDA).

ACKNOWLEDGEMENTS

The author would like to acknowledge NASS research management for their support. The Meteorological Satellite data was provided by Bobby Spiers and Carl Gernazio of USDA's Foreign Agricultural Service which was instrumental for this pilot research effort.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

PAGE

LIST OF FIGURES

PAGE

Figures

LIST OF TABLES AND EXHIBITS

PAGE

Tables

Exhibits

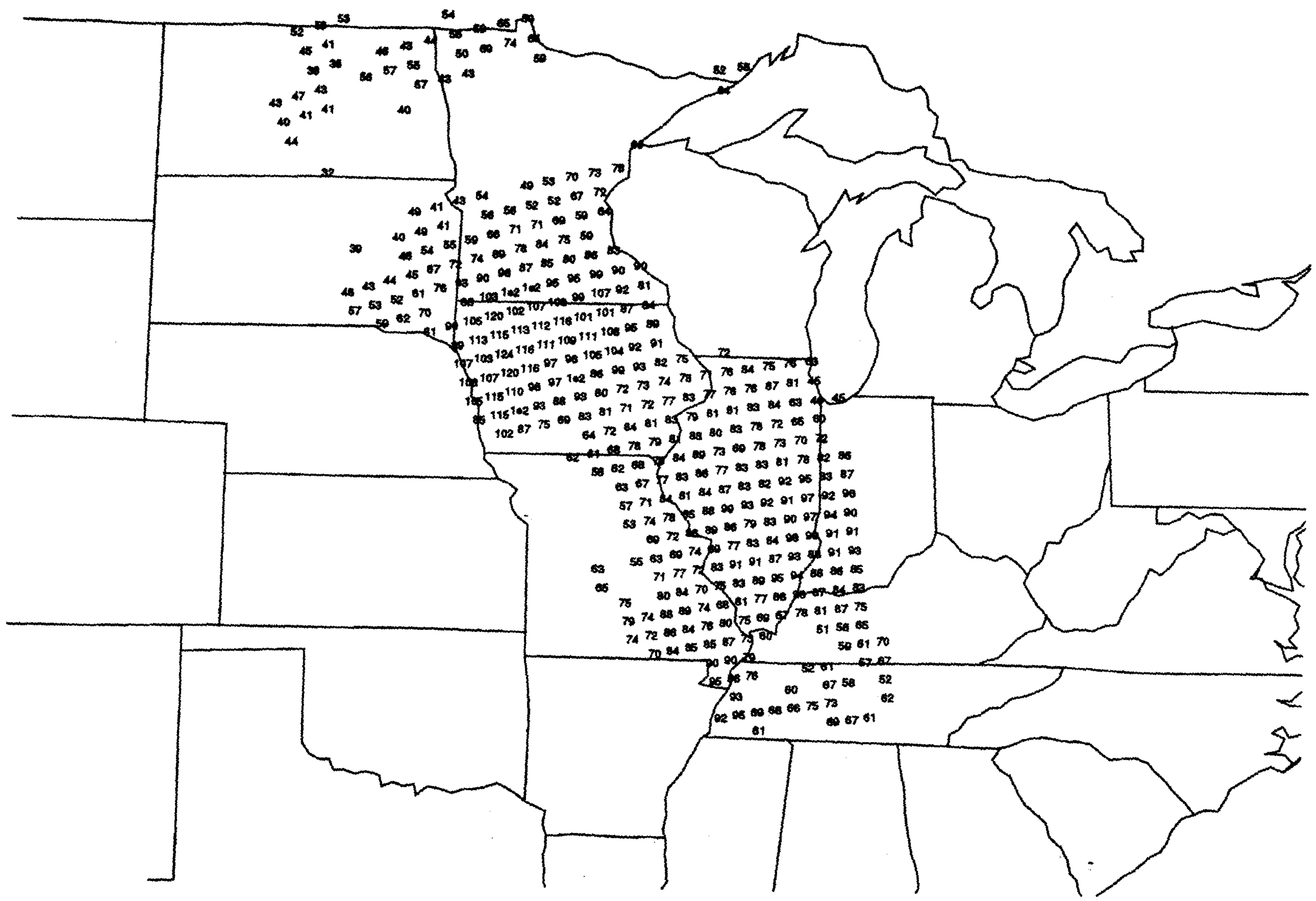
SUMMARY

In this pilot level research study, relationships between crop condition and polar orbiting meteorological satellite data were investigated for the 1984 corn and soybean crops. The 1984 forecasts and final estimates of corn for grain and soybean yield per harvested acre were used as State level measures of crop condition. NOAA-7 satellite data vegetative indexes were first aggregated to grid cells, averaged over time, weighted to counties and weighted by crop specific acreage weights to the State level. They were then used as the appropriate aggregate satellite derived crop condition index. Regression analyses were employed to understand the State level relationships of a crop's yield to its satellite vegetative index for ten States. The ten States are North Dakota, South Dakota, Minnesota, Iowa, Illinois, Indiana, Ohio, Kentucky, and Tennessee. Linear regression relationships for corn and soybeans existed at the State level, with coefficients of determination (R^2's) of .94 and .85 for final yield, respectively.

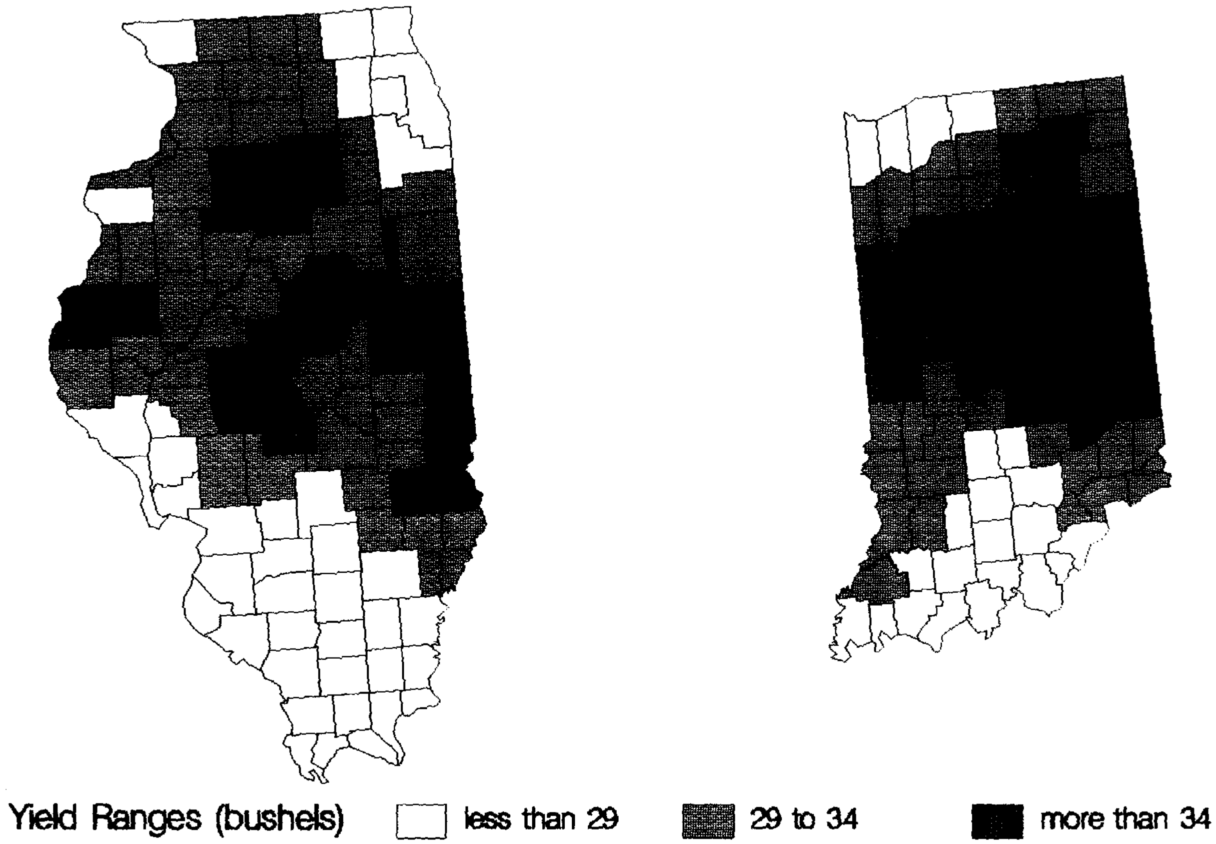
State level relationships were applied in generating county yield estimates to illustrate one of the applications possible from such a within-year study. Relationships with official county yields showed some decline from those at the State level. However, R^2's were still .63 for corn and .64 for soybeans with a relative standard deviation for both crops of about 16 percent. By eliminating 31 of the 889 counties with a substantial proportion of their corn irrigated, the corn R^2 increased to .69 and the relative standard deviation dropped to 14.5 percent.

This methodology was applied during the 1988 crop season, under drought conditions. The indices were strongly correlated to the official Agricultural Statistics Board estimates throughout the corn and soybean forecast season for 1988. The following maps demonstrate some of the input and output products for this study.

VIN(Satellite Data) from July 28, 1988



1984 SOYBEAN SATELLITE YIELDS
Illinois and Indiana



THE USE OF METEOROLOGICAL SATELLITE DATA IN ASSESSING CROP CONDITION

BY

WENDELL W. WILSON

INTRODUCTION

This report will discuss research on the use of polar orbiting meteorological satellite data in assessing crop condition. In this report you will learn what is being done and hopefully, gain some appreciation for the potential of further research in this area and for applications arising from it.

Estimates of corn and soybean yields are produced at the State, agricultural statistics district and county levels for the ten State study area. The contiguous study area includes North Dakota, South Dakota, Minnesota, Iowa, Missouri, Illinois, Indiana, Ohio, Kentucky, and Tennessee. Data was examined for only a single year, in this case 1984. There are several reasons for restricting the study to a within-year approach. The primary reason is that, since the satellite platform and sensor configurations change fairly often, one would lack comparable satellite data for pooling over very many years. Other reasons involve the possibility of sensor calibration drift (even if the same sensor and platform are available) and the changing crop situation in different years. Even though some crop situation factors will vary between States within a given year, it is thought that maturity stage, mix of crop types, and various other factors may vary more substantially from year to year.

Using a within-year approach does, however, impose certain limitations. There is no satisfactory method of using data from a single year to predict yields in another year. Innovative methods must be used within the year studied to produce other useful products. Some of these may provide improved local crop condition information of indirect use in producing improved current year crop yield forecasts and estimates. The application discussed in this report involves the use of State level yield to satellite data relationships in generating agricultural statistics district and county level yield estimate indications.

Even though the current study is restricted to within a single year, it does not mean all hope of over-the-years analysis has been abandoned. Eventually, more years will exist with comparable satellite data. The frequent platform and sensor changes being experienced are, of course, designed to lead to superior vegetation monitoring. And, the current studies strong within-year relationships over States, portends the strong possibility of useful relationships over years for individual States or groups of States.

A number of topics will be covered extensively in this report. They include the source and description of the data, an overview of the approach used, results at the State, county, and agricultural statistics district levels, and some observations on the use and accuracy of the county yield indications. The report also contains conclusion and recommendation sections, and a set of related appendices.

SOURCE AND DESCRIPTION OF DATA
Two Primary Types of Data

This study primarily utilizes two types of data. One consists of data from the United States Department of Commerce, National Oceanic and Atmospheric Association (USDC/NOAA) polar orbiting satellite. The other consists of United States Department of Agriculture, National Agricultural Statistics Service (USDA/NASS) crop yield and acreage statistics.

Satellite Data

Satellite data used in this study was obtained by the NOAA-7 polar orbiting satellite. Because of the satellite's orbit and sensor characteristics it senses a wide swath of the earth's surface. Such a wide swath is associated with two important results. While it allows the satellite to image the same area twice a day (once in darkness, once in light), it requires that the spatial resolution be quite gross in comparison to other polar orbiters (notably the Landsat satellite). While the Landsat "sees" the same area on the earth's surface every 16 days compared to once a day (in daylight, clouds permitting for both satellites), NOAA-7 can only spatially resolve 1.1 kilometers (1100 meters) at nadir compared to about 60 meters for Landsat. The difference in spacial resolution translates into a picture element (pixel) size of about an acre for Landsat and around 300 acres for the polar orbiting meteorological satellites. So, the temporal resolution of the meteorological satellite offers improved opportunity to monitor such dynamic phenomena as crop condition, but the lack of spatial resolution means that monitoring can not be done for specific crops. There is no way that the NOAA satellites can "look at" individual corn and soybean fields in this study area.

The sensor used in this study (one of many on the spacecraft) is the Advanced Very High Resolution Radiometer (AVHRR). The NOAA-7 was equipped with the AVHRR/2 which has five channels in which visible or infrared imagery is sensed. Channel 1 (visible) and channel 2 (near infrared) are used in computing the vegetative indexes used in this study, while some of the other channels are used to screen out imagery values effected by clouds. Channel 1 is sensitive in the .55-.68 micron range and channel 2 goes from .72 to 1.00 microns.

As part of Joint Remote Sensing Activities in the U.S. Department of Agriculture the Foreign Agricultural Service (USDA/FAS) has provided data to USDA/NASS to support this study. FAS receives ordered meteorological satellite data from USDC/NOAA and processes it. The FAS Image System (FASIS) is used to screen out satellite pixels that are either cloud covered or over water or have unacceptable reflectance values or that the algorithm eliminates for a number of other reasons. The (FASIS) grid cell summary program groups the data by geographically defined grid cells and computes summary statistics for each of them. Each grid cell, defined by a jth "row" and ith "column" location, is a 25 x 25 nautical mile square or about 28 3/4 statue miles on a side. The approximate center of each grid cell in longitude and latitude coordinates is available for each grid cell. The data reduction accomplished by USDA/FAS processing is of the order of about 1700 pixels to one grid cell for grid cells near nadir. A somewhat smaller data reduction for grid cells away from nadir occurs.

The grid cell summary program provides the percentage of potential pixels in a grid cell's area that are not screened out (% good), the proportion of good pixels that have vegetative indexes above the soil line (% green) and two grid cell vegetative index means. One of the vegetative indexes, the environmental vegetation index, (EVI), is the mean channel 2 value minus the mean channel 1 value for all good and green pixels within the grid cell. The other vegetative index is the so called "normalized" vegetative index (NVI). It is obtained by dividing EVI by the sum of the channel 1 and channel 2 means for the same set of pixels. Both of these vegetative indexes were explored in this study. Attempts were also made to create and use EVI's and NVI's adjusted to a 100% good "equivalent" based on a weak but positive relationship between the indexes and percent good. This study of the tendency of vegetative indexes to be biased lower when more cloud pixels are screened out (possibly because of cloud shadow or thin cloud effected pixels that remain) is reported in Appendix A.

Agricultural Data

The other primary type of data is in many respects the most important to this study. USDA/NASS State level yield estimates are used to calibrate the satellite data. These State level yield estimates (or forecasts, if they are used) are the product of indications from a collection of independent survey indications (see Scope and Methods) and the expert panel provided through the county estimates of acreage harvested for corn for grain and soybeans at the county level for the previous year are used to weight the vegetative index means for counties to the State level. They are used in order to produce a State corn vegetative index (when weighted by acres of corn harvested for grain) and a soybean vegetative index (when weighted by acres harvested for soybeans). While these acreages for the study year (1984) would be the correct ones for reflecting that year's actual county by county distribution of the crops, they would not be known until county estimates are made following the crop year. Therefore, the 1983 county acreage estimates are used to obtain crop specific vegetative indexes for the individual States.

Other Data Sources

County yield estimates for 1984 were of course, not used in the primary analysis. However, after satellite yield estimates were independently generated, the official USDA/NASS SSO county estimates were used retrospectively to evaluate the generated estimates' estimated accuracy and potential use as an additional indication for making the county estimates.
Another source was the U.S. Department of Interior, U.S. Geological Survey (USDI/USGS). They provided the approximate longitude and latitude coordinates for county centers. After their data was supplemented with USDA/NASS point estimates and edited, the locations were used in weighing grid cell vegetative indexes to produce county mean indexes based on the distance between each county center and the surrounding grid cell centers. The 1982 Census of Agriculture, from the U.S. Department of Commerce, Bureau of the Census (USDC/BOC), was used to identify counties which irrigate a large proportion of their corn crop. This information was helpful in evaluating the situations in which satellite generated corn yield indications would be of limited use.

Data Variables Summary

To summarize the source and data variables used, primary and secondary variables are listed below. Secondary variables are those not used in the primary analysis.

Source	Primary	Secondary
USDC/NOAA	NOAA-7 Satellite Channel 1&2 Values	Values from other NOAA-7 channels
USDA/FAS	Grid cell vegetative indexes (EVI & NVI), latitude and longitude	Grid cell % good, and % green
USDA/NASS-ASB	Final 1984 corn for grain and soybean yield estimates for 10 States	August 1, September 1, October 1, and November 1, 1984 yield forecasts for the 10 States
USDA/NASS-SSO'S	1983 County estimates of acreage harvested for corn for grain and soybeans	1984 county estimates of corn for grain and soybean yields per harvested acre
USDI/USGS	County center latitude and longitude	
USDC/BOC		1982 Census of Agriculture county estimates of number of farms and acres of total and irrigated corn harvested for grain

OVERVIEW OF THE APPROACH USED
Overview Diagram

Figure 1 is an overview diagram of the approach used in the primary analysis. Each step shown in the figure has a number, a brief description of what is being done, a description of the product at the end of that step and some symbolic notation. Each of the steps will be discussed rather thoroughly in this section. Topics involving analysis and selection of alternative procedures will be discussed briefly in this section and/or included in an appendix.

Figure 1A
Overview of the Approach Used
In the Primary Analysis

Step 1

Start with	USDA/FAS Grid Cell Vegetative Index Means	V_{ijt}

Step 2

Average for Critical Time Period	Grid Cell Vegetative Index Period (may vary by crop) Means	$\overline{V}_{ij.}$

Step 3

Map to County Level	County Vegetative Index (may vary by crop)	V_{mn} -> (to Step 7)

Step 4

Create Crop Specific Weighted Average at State Level	State Corn Vegetative Index $\overline{VC}_{.n}$	State Soybean Vegetative Index $\overline{VS}_{.n}$

(To Step 6)

Step 1

In Step 1 the approach used starts with the USDA/FAS grid cell vegetative index means. In this study the use of both the EVI and NVI indexes was explored. The symbol, V_{ijt}, represents either index for a grid cell in the ith "column" and the jth "row" (see Figure 2 for coordinate system) that was derived from imagery obtained on day t. In the summer of 1984 grid cell values were computed and retained by USDA/FAS for each day that the number of good pixels (% good) exceeded 50 percent. Therefore, values do not exist for days with complete or substantial cloud cover, but were usually available for clear or partly cloudy days.

Step 2

Step 2 involves obtaining the average vegetative index for a critical time period. The analysis that lead to the selection of critical periods for both crops is described more completely in Appendix B.

Critical period selection basically involved two complementary methods. One method was to observe the seasonal pattern of grid cell vegetative indexes. It was desirable to identify a plateau in the index values. The plateau would occur after a period of "greening up" or perhaps following some "greenness" associated with pre-ripe small grain crops, but prior to the decline in "greenness" that accompanies fall and crop maturity. Such a plateau would provide observations on multiple dates when crop condition could be considered nearly stable. This would allow means to be created over the period which would mitigate some of the "noise" in the daily values.

The other method involved testing the relationships of yields to average vegetative indexes over various length periods at the State level. The candidate period identified by the pattern of "greenness" analysis was broken down into cycles of approximately equal potential coverage of the entire study area. Each cycle's (about eight days long) relationship to yield (forecasts for various dates and final estimates of each crop) was evaluated. Then, since a single cycle might have little or no data for some areas, and could provide unrepresentative data at the State level, adjoining cycles were combined and evaluated. Continuing in this manner, more adjoining cycles of coverage were combined until several very competitive periods were identified. These periods were made up of individual cycles, all of which had fairly strong relationships to the crop yield forecasts or estimates, and which when combined in groups of two achieved higher relationships as a result of more complete and representative State coverage. These periods were generally consistent with the "greenness" pattern method; however, some compromises were made by including a few observations near the end of small grains "greenness" or from the early stages of crop maturity. The average grid cell vegetative index for the critical period

$$\bar{V}_{ij\cdot} = \sum_{t} V_{ijt} / N_{ij}, \text{ where } \bar{V}_{ij\cdot}$$

is the mean vegetative index for ijth grid cell over the selected critical time period. The summation is over all available V_{ijt}'s within the period and N_{ij} is the number of those observations. For final yield estimates the period selected for corn was July 31 through August 23, 1984 and for soybeans it extended from July 31 through September 1. An earlier period would be closer to optimum for both crops when relationships to State level August 1 yield forecasts are considered. The optimum was quite flat around the several periods given the most consideration and selection of one over another would not alter results much.

Step 3

Step 3 maps grid cell mean vegetative indexes to the county level. In this case, the optimum mapping algorithm was also quite flat. The mapping criterion in this study was limited to the Euclidean distance between county and grid cell centers. Search radii limitations of approximately 20, 30, and 40 miles were investigated. Weights that declined linearly and exponentially were explored. Of these six combinations (three distance limitations by two weight decay rates), the exponential decay with a 30 mile search radius produced the strongest relationship to yield at the State level. However, all of the methods were very close at the State level and the exponential for 30 and 40 mile limits produced very highly correlated county vegetative indexes. Therefore, the method selected was a weighted average, where the weights were inversely proportional to the squared distance between the county and respective grid cell centers, with a search radius of 30 miles which was extended to 40 miles if no grid cell centers (with data) were within 30 miles. That is,

$$V_{mn} = \sum_{ij} (\overline{V}_{ij\cdot}/d^2_{mnij}) \;/\; \sum_{ij} (1/d^2_{mnij})$$

where V_{mn} is the vegetative index for the mth county in the nth State, d^2_{mnij} is the squared distance from the center of the mnth county to the center of the ijth grid cell and both summations are over all grid cells within 30 miles of the county center (or within 40 miles if there are no observations within 30 miles).

This algorithm, which can be termed the "extended 30 mile quadratic mapper", gives most of the weight to grid cells closest to the county center, limits the search radius to 30 miles in most cases and produces a vegetative index for most of the counties in the study area. Of the 916 counties, 908 had a vegetative index for the selected critical period for soybean final yield (July 31-September 1, 1984) and 905 had a value defined for the shorter corn final yield period (July 31-August 23, 1984).

Support for selecting the "extended 30 mile mapper" and examining the competing algorithms came in part from a study of 1983 official county yields. The difference in both corn and soybean yields as a function of distances between county centers was reviewed. In general, the review suggested a maximum search radius of 40 miles (yields can become substantially different over greater distances) and a decay function greater than the linear rate, but often not quite a rapid as the distance squared.

Step 4

Step 4 involves the creation of appropriate State level vegetative indexes. The county vegetative indexes are weighed to the State level just as one would do to obtain State average yields, if they were in fact known for each county. That is, weights are used based on the harvested acres which are equivalent to the same harvested acres used in the yield expression (production per acre harvested). Since this study is concerned with investigating what could actually be done, 1983 county harvested acreage estimates of corn for grain (for corn) and soybeans (for soybeans) are used as weights. A few counties in some States with nominal acreage are given a weight of zero (very close to their actual weight) because individual estimates are not made for those less important counties.

The products resulting from this step are State crop specific (corn for grain or soybeans) vegetative indexes. They are crop specific in the sense that county vegetative indexes were weighted together based on the relative density of the crop in different parts of the State. It is important to recognize the low spatial resolution of the meteorological satellite data. Since areas of the order of about 300 acres can be resolved spatially the V_{ijt}'s reflect something that may be thought of as "vegetative greenness". Therefore, the vegetative indexes reflect this general sensing of the scene and the State level corn for grain and soybean vegetative indexes are only crops specific because they incorporate the varying importance of the crops in different counties.

The equations for these State level indexes for corn can be expressed as follows:

$$\overline{VC._n} = \sum_m (C_{mn} V_{mn}) / \sum_m C_{mn},$$

and for soybeans

$$\overline{VS._n} = \sum_m (S_{mn} V_{mn})/\sum_m S_{mn}.$$

Here, $\overline{VC._n}$ and $\overline{VS._n}$ are the mean crop specific vegetative indexes for corn and soybeans, respectively, in the nth State. C_{mn} and S_{mn} are the 1983 (previous year) published harvested acreage estimates of the respective crop for the mth county in the nth State. The summation is over all counties in the State (m=1, 2, 3...) even though some of them may have a zero weight for either or both crops.

Figure 1B
Overview of the Approach Used In the Primary Analysis

Step 5

	Corn		Soybean	
Obtain USDA/NASS Final State Yield Estimates	Corn Yield Estimate	EC_n	Soybean Yield Estimate	ES_n

Step 6

Regression Parameter Estimates and Residuals

Develop Calibration Equations

Corn

$$EC_n = \hat{\alpha} + \hat{\beta}\,(VC._n)$$
$$RC_n = \hat{EC}_n - EC_n$$

Soybeans

$$\hat{ES}_n = \hat{\gamma} + \hat{\delta}\,(VS._n),$$
$$RS_n = \hat{ES}_n - ES_n$$

Step 7

Satellite Generated County Yield Estimates

Produce County Yield Indications

Corn

$$EC_{mn} = \hat{\alpha} + E(V_{mn}) - RC_n$$

Soybeans

$$ES_{mn} = \hat{\gamma} + \hat{\delta}\,(V_{mn}) - RS_n$$

Step 5

Step 5 involves obtaining the USDA/NASS final state corn and soybean yield estimates. These estimates are, in many respects, the most important component in this study. There is no problem in obtaining the State level final yield estimates, they are published in January some time in advance of the date county yield indications would be needed. However, the timetable would be tighter for obtaining August 1 forecast yields for use in producing county or other local area potential yield variables. The symbols for the final yield estimates for the nth State are EC_n and ES_n for corn for grain and soybeans, respectively. They are used as the dependent (calibration) variables in the next step.

Step 6

Developing the calibration equations constitutes the sixth step in the primary analysis. For the ten States in the study area each crop's final yield estimate is regressed on its vegetative index. The level of each observation is the State, so that only ten data points are used in the analysis. To conserve degrees of freedom, to maintain the greatest simplicity consistent with objectives and to achieve parsimony, simple linear models are used. The products available at the completion of this step are the regression model parameters and residuals for individual States. The regression equations for corn are:

$EC_n = \alpha + \beta\ \overline{VC}_{\cdot n}$, with residuals $RC_n = \hat{EC}_n - EC_n$;

for soybeans: $ES_n = \gamma + \delta\ (\overline{VS}_{\cdot n})$ with residuals $RS_n = \hat{ES}_n - ES_n$,

where α, β, γ, and δ are regression intercept and slope parameters.

Step 7

In Step 7 the county yield indications are produced. These county yield indications or satellite generated yield estimates are the product available upon completion of the step. They are symbolically represented by EC_{mn} and ES_{mn}, the corn for grain and soybean yield estimates for the mth county in the nth State. The county estimates are obtained by utilizing the relationship between the yield and vegetative index at the State level. That relationship is applied in mapping county vegetative indexes to county yields. In as much as the strength of the State level relationship supports belief in the phenomenon of the linear dependence of yield on the vegetative index, it may be reasonable to apply that relationship at another level of aggregation.

The calibration equations are:

$$\hat{EC}_{mn} = \alpha + \beta\ (V_{mn}) - RC_n \quad \text{and} \quad \hat{ES}_{mn} = \gamma + \delta\ (V_{mn}) - RS_n$$

The State level residuals are subtracted from the mapping of county vegetative indexes to yields based on the ten State study area relationship. This has the desirable property of keeping the county yields in a State collectively consistent with the official State yield estimate. Intuitively, if the over or under estimate of the State yield is fairly uniform across a State then the adjustment would improve the county satellite generated estimates. A possible negative factor is that artificial yield differences along State boundaries might be introduced.

Utilize Options

One option is to produce theme map products (at the county level) which reveal where the crop is doing well and where it isn't doing as well. Such map products can be useful to the ASB and SSO staffs. Map products showing the relative condition of crops in a spatial sense might also be provided to data users.

Of course, one product to be utilized would be county yield indications. They would be used to provide improved crop statistics for county, agricultural statistics districts, and other groups of counties (drainage basins, marketing areas, etc.). A more speculative utilization would be the creation of supplemental variables for use in conducting more efficient yield surveys. Such supplemental variables would be produced in a manner similar to the county vegetative indexes. They would probably be for a more restricted search radius and would be designed to reflect the average crop condition in a local region around a sample farm or field, or for a group of sample units. If the correlation between the yield variable measured at the sample "point" and the neighboring area satellite generated crop condition variable were high enough, the satellite information could be used in a regression estimator, since the grand mean or average satellite variable value exists for the entire population. Of course, a sufficiently high correlation would permit other common uses of supplemental variables. These might include stratification or post-stratification (dependent on timing), unequal probability sampling and other uses of the supplemental information either in the sampling design or in the estimator.

"Utilize Options"

- Map Products
- Improved county, Agricultural Statistics District, and other local crop statistics
- Supplemental variable for more efficient yield surveys
- Other possibilities

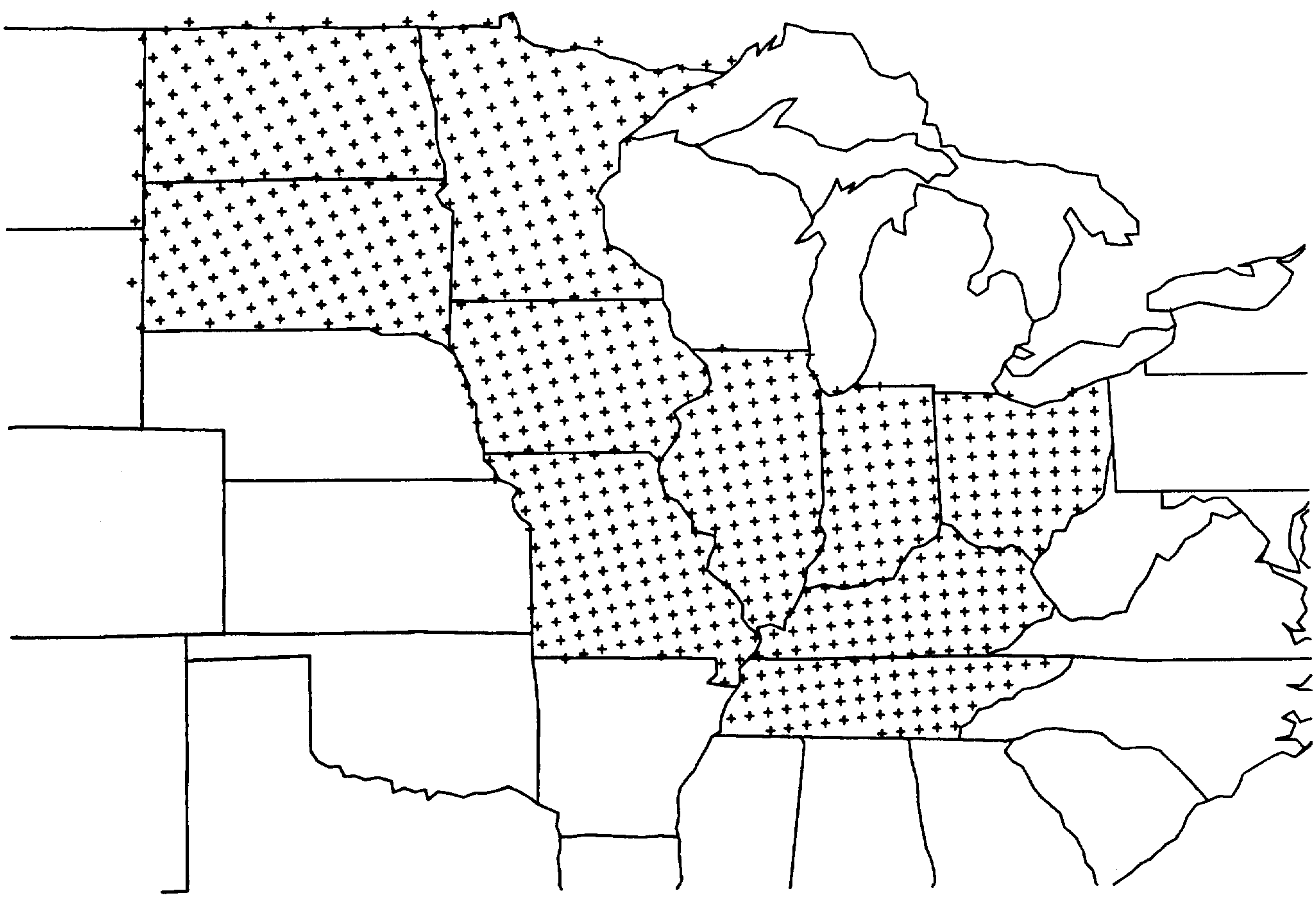
The Study Area

The ten State study area of North Dakota, South Dakota, Minnesota, Iowa, Missouri, Illinois, Indiana, Ohio, Kentucky, and Tennessee is shown in Figure 2, along with the grid cell coordinate system. Showing the coordinate system with these States provides some idea of the magnitude of data available. The dots in the figures represent the approximate center of each grid cell. There are, for example, about 75 grid cell centers in Iowa. With about 1700 pixels per grid cell (actually somewhat less because of cloud screening) and an average of four days of observation (actually more than four), Iowa would have in excess of half a million observation points during the critical period.

The study area States have a few characteristics in common. They are all important agricultural States with a significant part of their land area in crops. Corn and soybean production is important enough throughout the area that county yield estimates are produced and published for each of the States. Most of the remaining common factors are associated with the fact that they are contiguous. From north to south they can exhibit substantial diversity in crop development and maturity stages. Not only are there differences in crop stages, but the variability in development stage (by necessity) is much more restricted in the North. From West to East, or perhaps from Northwest to Southeast, substantial differences prevail. The natural woodland vegetation in the East is very different from the prairies in the West. The low rainfall, low humidity, and fallowing practices of the west are about as dissimilar as can be found within a contiguous area of this size from those of the East and Southeast.

Many of these factors (both the similar and dissimilar ones) affect the vegetative information that can be obtained from satellites, particularly from those with such a low spatial resolution as the meteorological satellites. The wide mix of natural cover types and crops would logically make one question whether the satellite data could possibly measure crop condition in a consistent way for these States. If one should find a strong relationship between crop yields and satellite vegetative indexes for such a dissimilar group of States, then it may not be too unreasonable to hope that the same satellite data will also provide useful crop condition information for agricultural reporting districts and counties (which can be quite dissimilar) within these States.

Figure 2
TEN STATE STUDY AREA
Grid Cell Coordinate System and Approximate Location of Cell Centers



RESULTS AT THE STATE, COUNTY, AND AGRICULTURAL STATISTICS DISTRICTS LEVEL

State Level Relationships

A plot of relationships between the final yield estimate and each crop's vegetative index is shown in Figures 3 and 4 for corn for grain and soybeans, respectively.

The regression equation line for corn, $\hat{EC}_n = -16.25 + 1.60\ \overline{VC}._n$

is shown in Figure 3. The model explains a highly significant amount of variability in yields between the States and has a coefficient of determination (R^2) of .94. Individual State data are plotted with the letter in the postal abbreviation underlined in Figure 2. Iowa, Missouri, Illinois, and Ohio are denoted by the second letter, while the other six use the first.

The regression line for soybeans, $\hat{ES}_n = -9.05 + 0.53\ \overline{VS}._n$,

and the State data points are displayed in Figure 4. The soybean model explains a highly significant 85 percent (R^2=.85) of the yield variability for the ten States.

The statistical software package (SAS) regression output for the corn and soybean models is included in Appendix C.

The corn and soybean models just presented used EVI as the vegetative index for both crops. The NVI version was very competitive for corn, but had lower slightly explanatory power for the yield of both crops.

The EVI version was selected, based on its performance for soybeans, so that the same version could be used for both crops. The greater effective range of the EVI variable was also thought to provide better discrimination at the county level for the wider corn yield range. Plots analogous to those in Figures 3 and 4 and corresponding SAS regression analyses are shown for NVI in Appendix C.

Agricultural Statistics Districts Satellite Generated Yield Results

After the State equations and residuals were employed in generating county yield indications, the counties in each agricultural statistics districts were weighted together by their 1983 acreage weights to produce district means. The agricultural statistics districts level results are shown for corn and soybeans in figures 5 and 6, respectively. The ordinate is the official yield of the crop as published in USDA/NASS SSO bulletins and included in the Agency's crops data base. The abscissa is the mean district satellite generated yield for counties with published acreage for the previous year. So, even if county estimates agreed completely, district estimates could differ because the relative importance of counties for the crop changed from 1983 to 1984 or the omission of minor counties distorted the satellite generated yield mean.

Figure 3
CORN FOR GRAIN - STATE LEVEL
Official Final Yield Estimate (bushels per acre)
Versus
Corn Vegetative Index
(EVI Version)

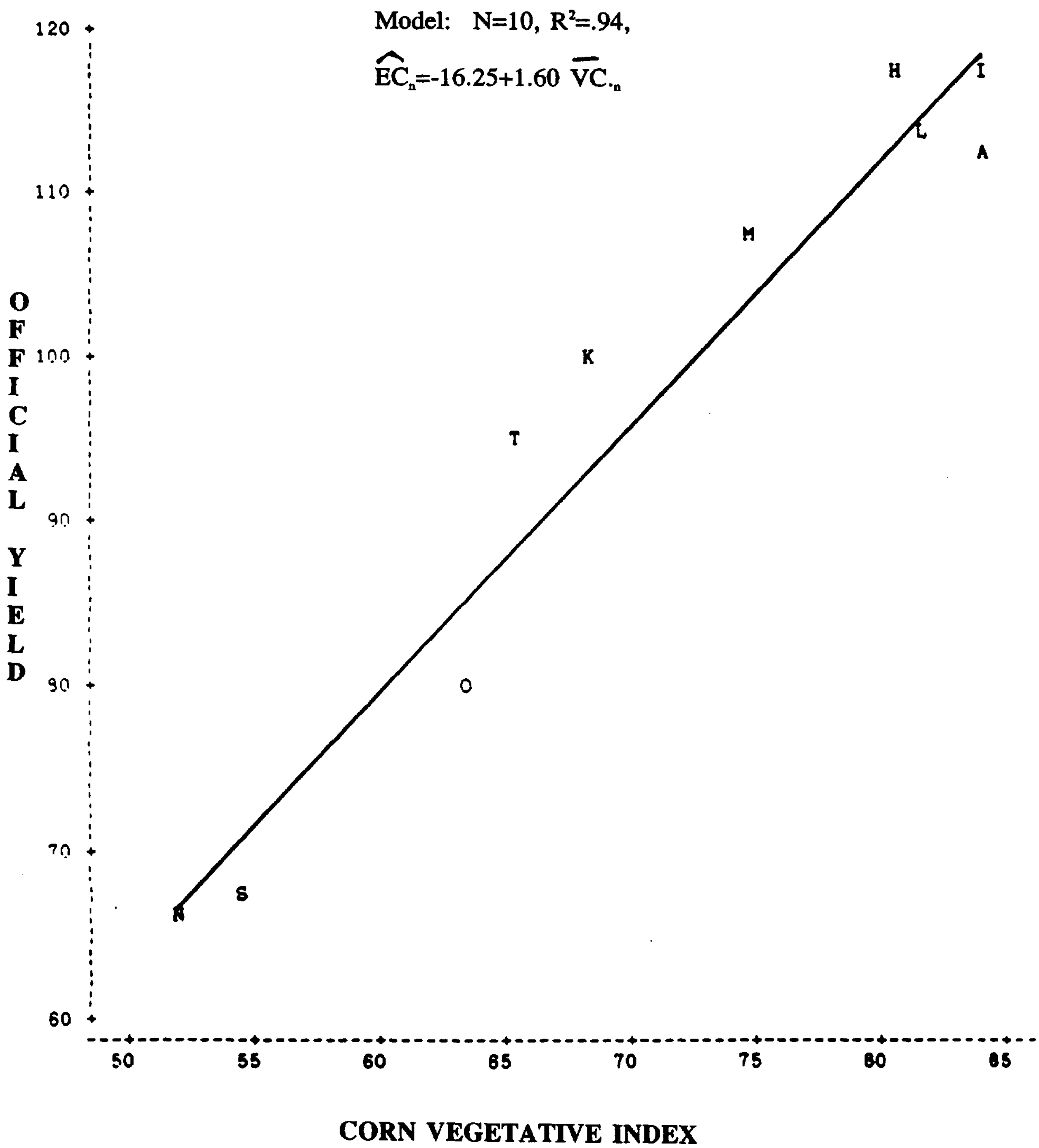


Figure 4
SOYBEANS STATE LEVEL
Official Final Yield Estimate (bushels per acre)
Versus
Soybean Vegetative Index
(EVI Version)

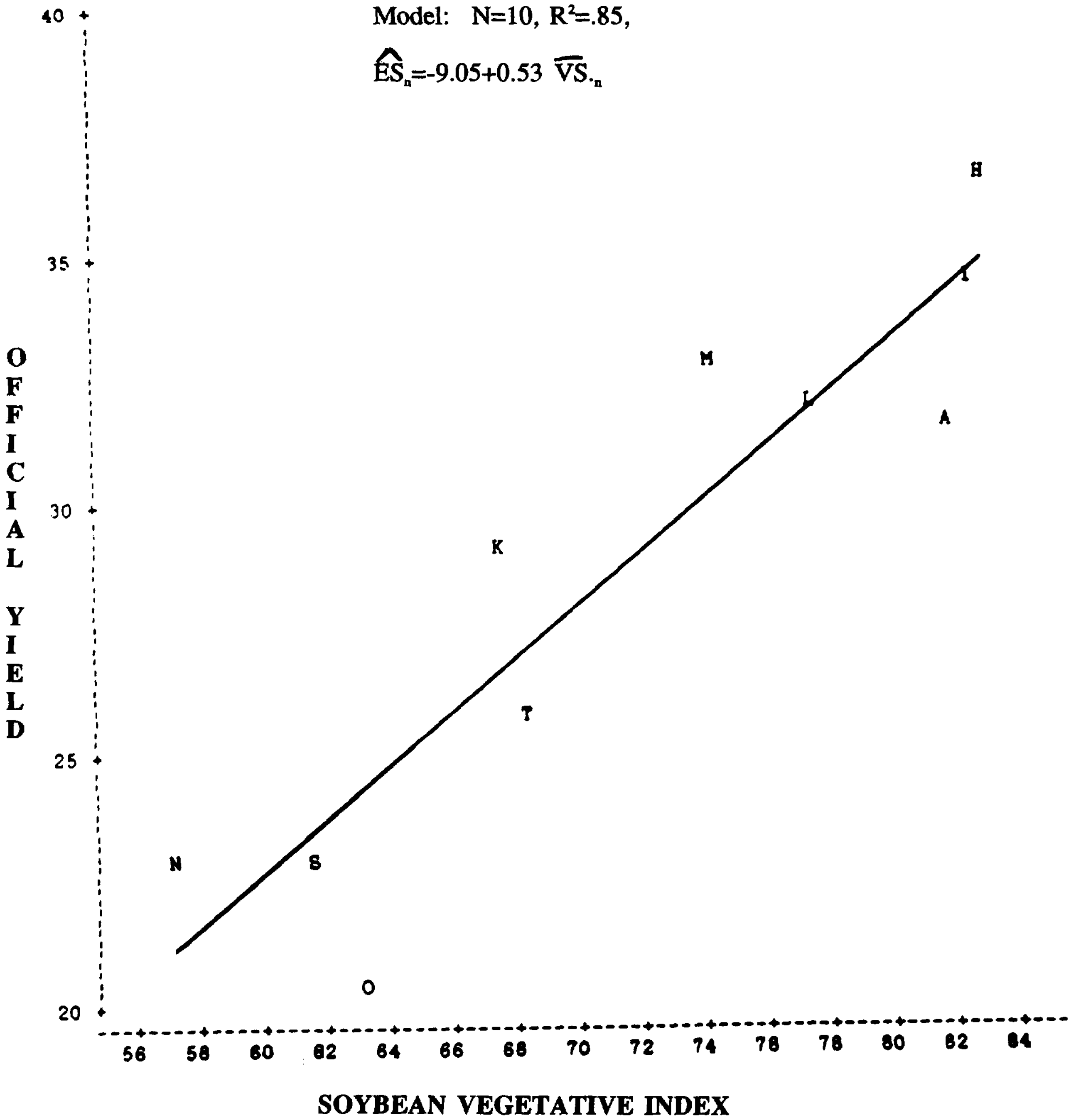


Figure 5
CORN FOR GRAIN - AGRICULTURAL STATISTICS DISTRICTS LEVEL
Official Yield Estimate (bushels per acre)
Versus
Satellite Generated
Corn Yield Estimate (bushels per acre)

N=84, R^2=.75

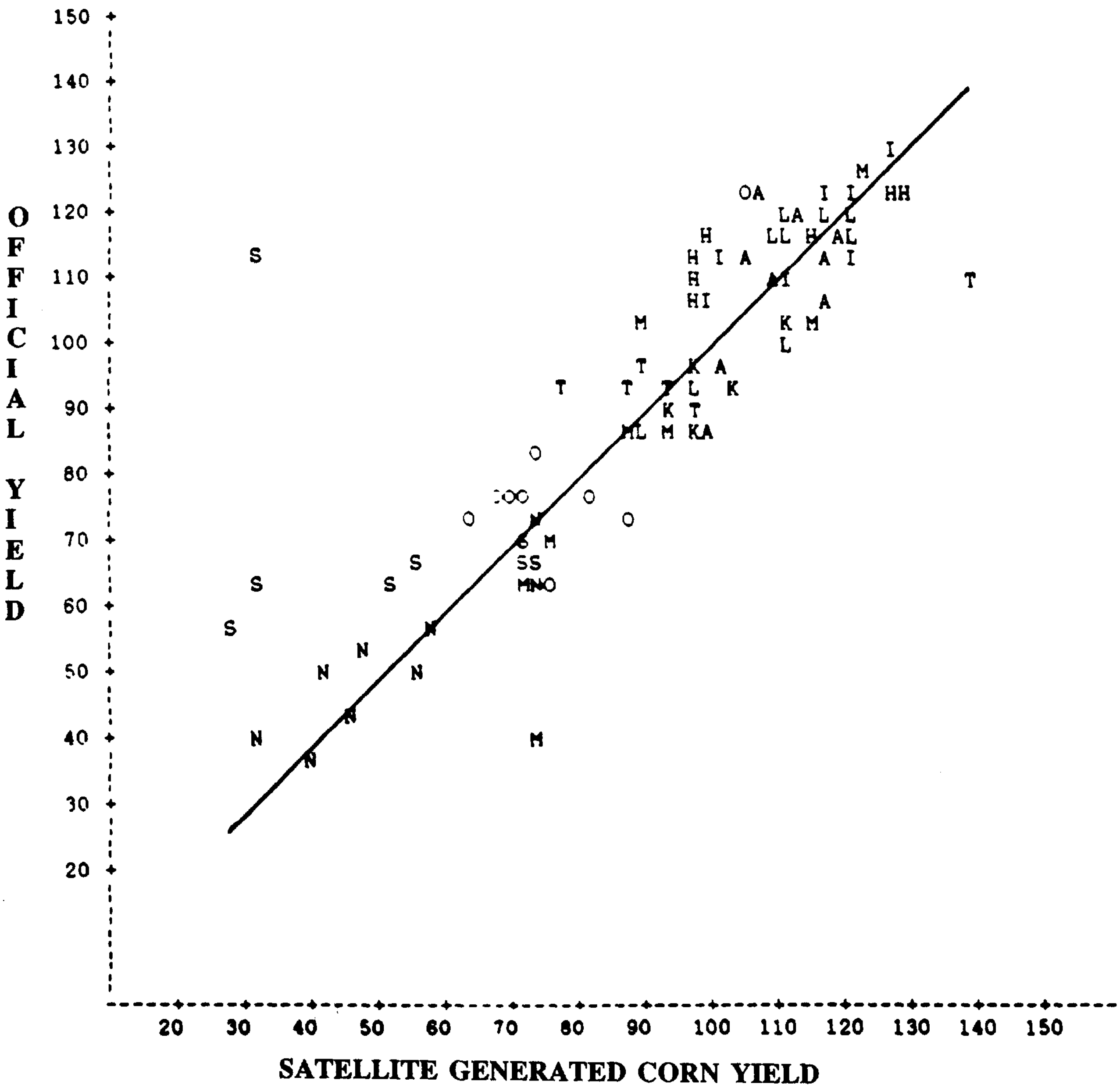
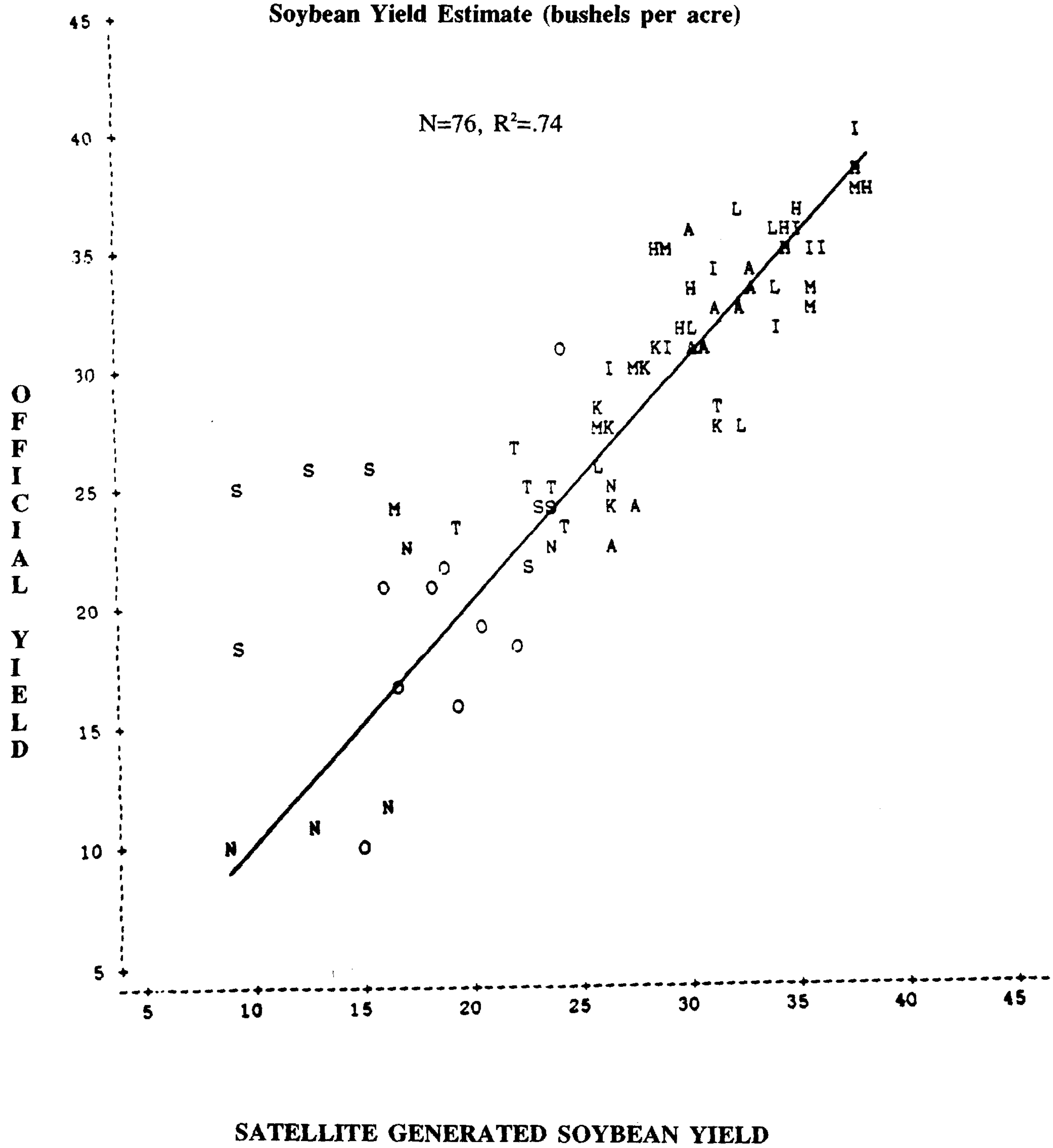


Figure 6
SOYBEANS - AGRICULTURAL STATISTICS DISTRICTS LEVEL
Official Yield Estimate (bushels per acre)
Versus
Satellite Generated
Soybean Yield Estimate (bushels per acre)



For corn, all 84 agricultural statistics districts in the study area (six in Kentucky and Tennessee, nine in each of the others) had both official and satellite generated mean yields. The R^2 between means from these different sources was .75, explaining three-fourths of the district to district yield variability. The line shown in Figure 5 is the one-to-one line of perfect agreement. The "State postal one letter code" is used to identify districts in each State. The three underestimated outlier "S's" are the western South Dakota districts. The far outlier is the Southwest district. Note that the other "S" districts are near the one-to-one line. The "M" outlier, with yield substantially overestimated by the satellite source, is the three county district in Northeast Minnesota. Since only one of the three counties had a 1984 published corn yield that data point really represents a single county (St. Louis County). Although, the "M" is not such an extreme outlier when considered as a single county, it is still a substantial overestimate. Actually, St. Louis county is very large. One would not be surprised to find that the vegetative index, mapped from surrounding grid cells towards the county center, failed to represent where the county's 300 acres of corn harvested for grain were in 1984.

For soybeans, both variables were available for 76 of the 84 districts. So, even though soybean satellite generated yield estimates were available for counties in additional districts (they can be generated even where the crop isn't grown), the lack of weights (1983 harvested soybean acreage estimates for individual counties) prevented their aggregation for some districts. The explanatory power at the district levels was about the same for soybeans as it was for corn (R^2=.74).

The soybean plot in Figure 6 also shows some outliers. These include some districts near those that were outliers for corn and may involve small acreages of soybeans grown in locally advantageous areas within those districts. Another factor that should be considered is the range of vegetative index and yield data used in fitting the State level models. The models will, of course, perform linear extrapolations beyond the lowest and highest values when they are applied at the county and agricultural statistics districts levels. The State level model predicted yields express the range of vegetative indexes in terms of the yield of the two crops. These values ranged from 67 to 119 bushels for corn and 21 to 35 bushels for soybeans. The soybean official and satellite generated yield relationships (see Figure 6) appear to split into two groups when extrapolating below 21 bushels. Of course, the "yield ceiling" near the top end of the scale makes extrapolations above 119 (for corn) and 35 (for soybeans) bushels less of a concern than the greater extrapolations on the opposite end of the generated yield scale.

County Level Results

County level results are shown graphically in Figures 7 and 8. The ordinate is the official county yield as published in USDA/NASS SSO bulletins and included in the NASS Headquarters' crops data base. If an individual county crop yield is not published (because of no acreage, low acreage, or to avoid disclosure of individual operations) then it is excluded from Figures 7 and 8. The abscissa is the result of mapping county vegetative indexes into corn and soybean yields as was described in Step 7 on Page 14. For corn for grain, the average yield of the mth county in the nth State is given by:

$$\hat{EC}_{mn} = -16.25 + 1.60\ V_{mn} - RC_n.$$

For soybeans the equation for expressing county vegetative indexes obtained over the July 31 through September 1, 1984 critical period as yield is

$$\hat{ES}_{mn} = -9.05 + 0.53\ V_{mn} - RS_n.$$

The residuals of each crop for the ten States are displayed in Figures 9 and 10. The maps reveal some large adjoining State residual shifts which could lead to substantial yield differences between nearby counties with similar vegetative indexes. The agricultural statistics districts referred to earlier are shown in Figure 9. Some appreciation for the varying size and orientation of counties can be obtained from observing their boundaries in Figure 10.

County corn yields are plotted in Figure 7 for the 889 of the 916 counties that have both official and satellite generated corn yields. Official corn yields were published for all but 17 counties in the area. As mentioned earlier, 11 of the 916 counties do not have a vegetative index for the corn period (and thus no satellite generated yield). The strength of relationships has declined somewhat, dropping from an R^2 of .75 at the district level to .63 for counties. The spread of the corn data around the one-to-one line shows large underestimates for some South Dakota, North Dakota, and Missouri counties. The selected letter from the postal abbreviation is again used to identify the State a county comes from; however, when looking at the plotted data for so many points it is important to realize that much of the data is hidden near the one-to-one line. Thus, one should not get the mistaken impression that the proportion of outliers is as large as it may appear from the plot. A substantial number of the overestimates are below the State level satellite generated yield range (below 67 bushels). Likewise, it is true that most overestimates are above the State level range (greater than 119 bushels). More will be said about the accuracy of the data and the outliers in the next section of this report.

The soybean yields from both sources are available for 756 counties. All of the 160 "missing" counties did not have a published soybean yield for 1984. The eight counties without a satellite generated yield were among those without published yields. In fact, the 160 counties have very few soybean acres, so that those plotted are the ones to examine in considering the value of the satellite generated estimates. As in the case of corn, the strength of soybean relationships declined as the aggregation level was lowered. However, the decline in the strength of the soybean relationship was less than for corn; stating at a lower State level but being essentially the same as corn at the district and county levels.

Figure 7
CORN FOR GRAIN - COUNTY LEVEL
Official Yield Estimate (bushels per acre)
Versus
Satellite Generated
Corn Yield Estimate (bushels per acre)

N=889, R^2=.63

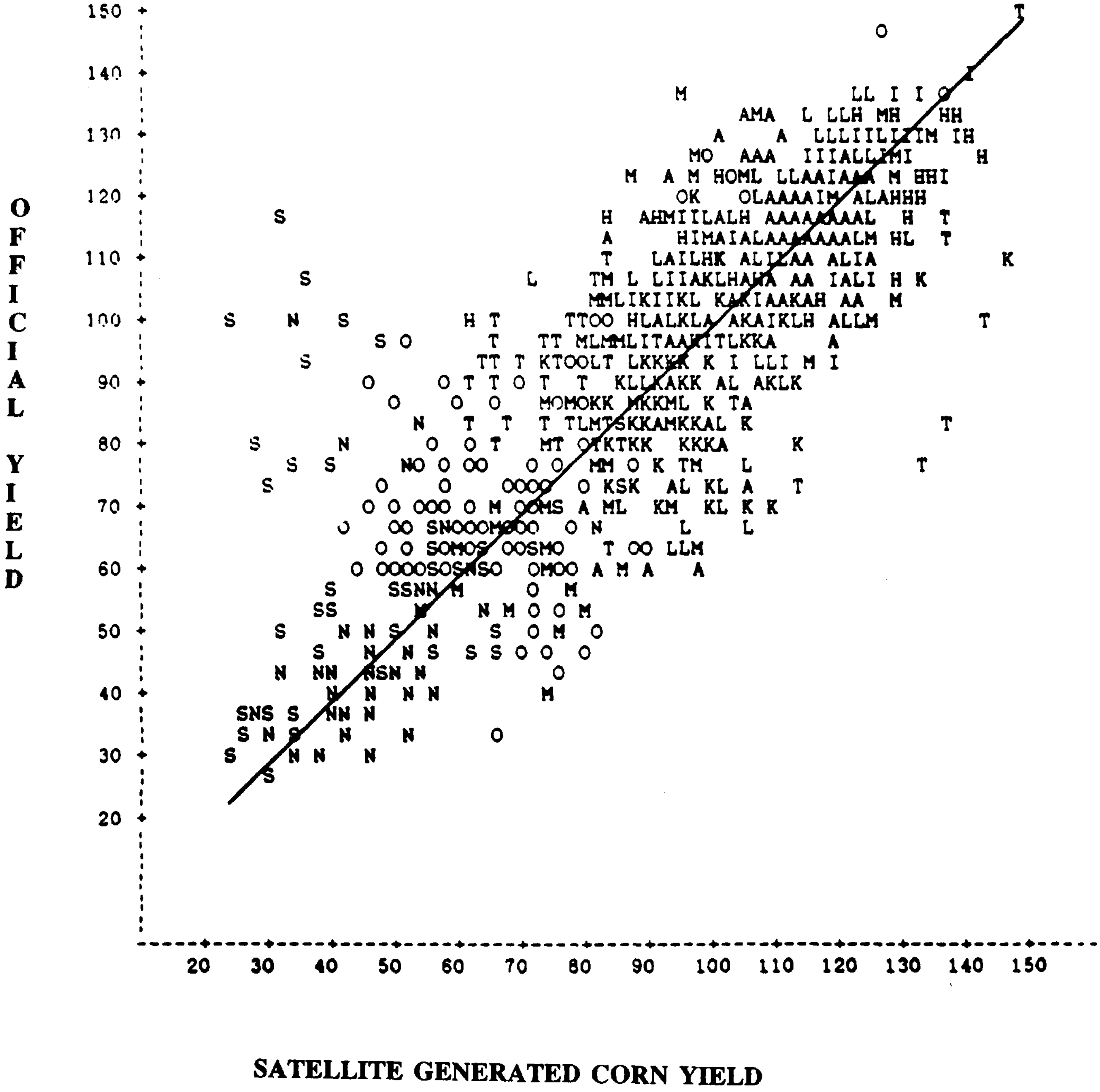


Figure 8
SOYBEANS - COUNTY LEVEL
Official Yield Estimate (bushels per acre)
Versus
Satellite Generated
Soybean Yield Estimate (bushels per acre)

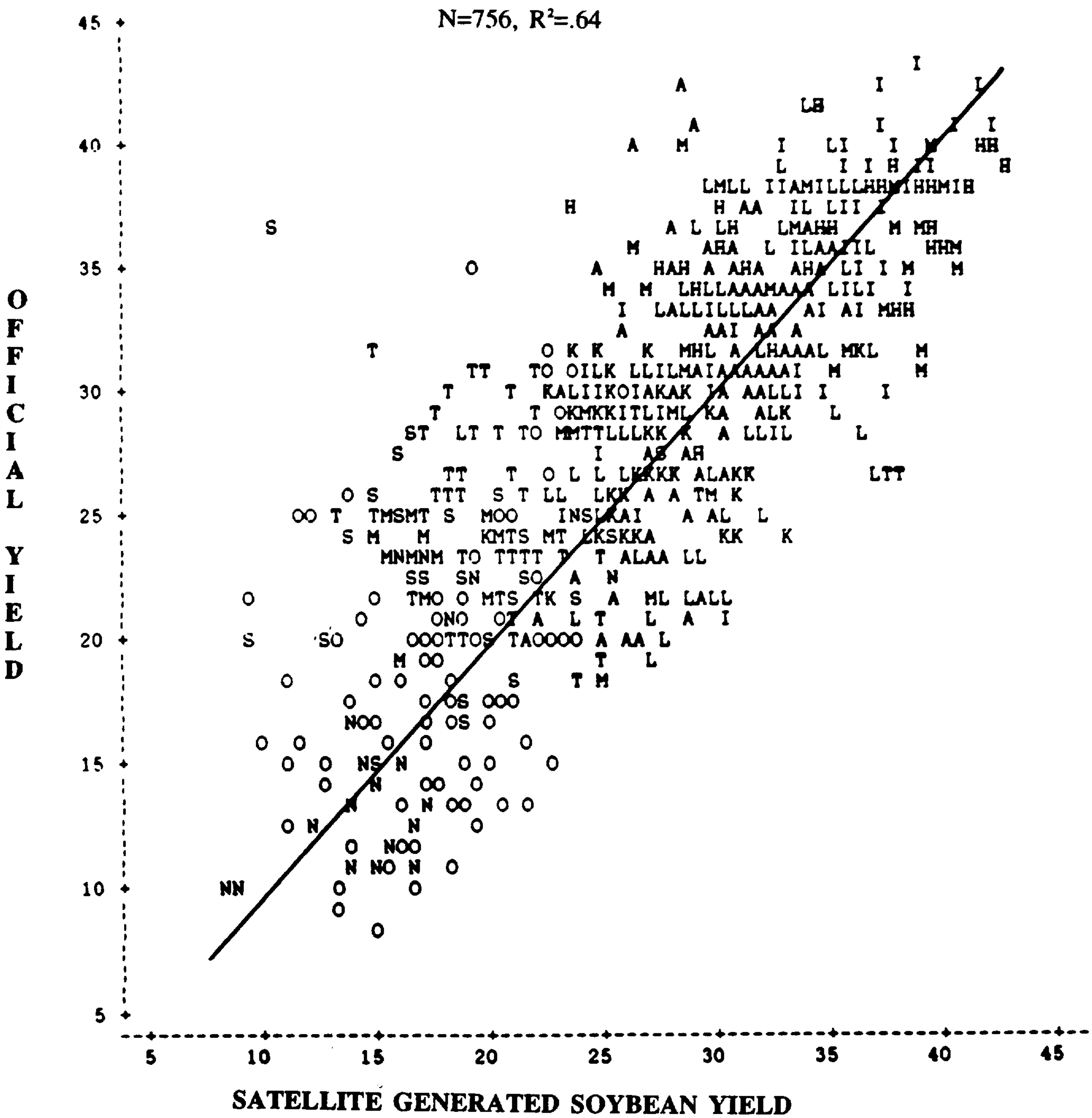


Figure 9
RESIDUAL OF THE STATE LEVEL REGRESSION
of the
Corn For Grain Yield
on the Corn Vegetative Index
agricultural statistics districts are shown

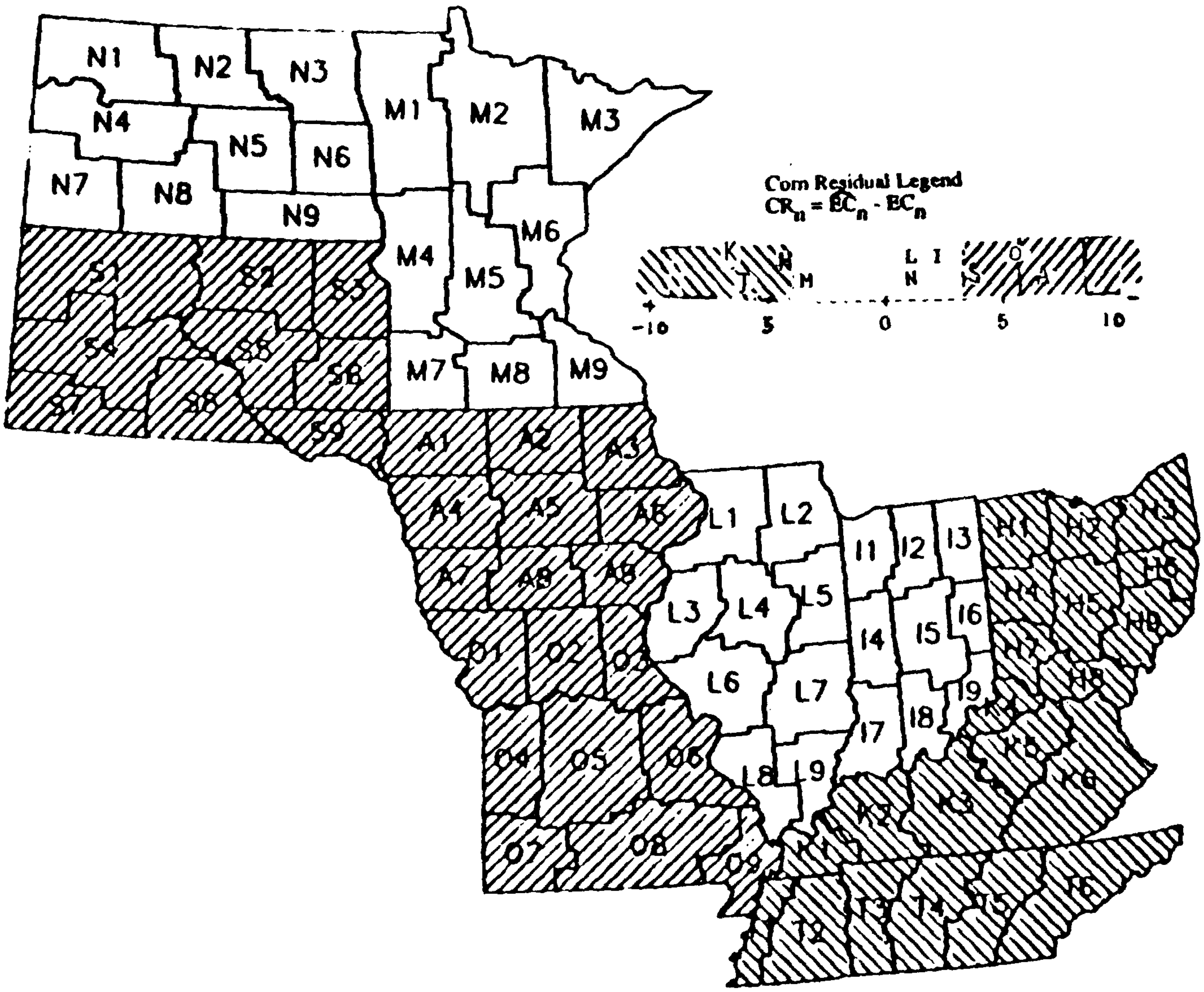


Figure 10
RESIDUAL OF THE STATE LEVEL REGRESSION
of Soybean Yield
on the Soybean Vegetative Index
county boundaries are shown

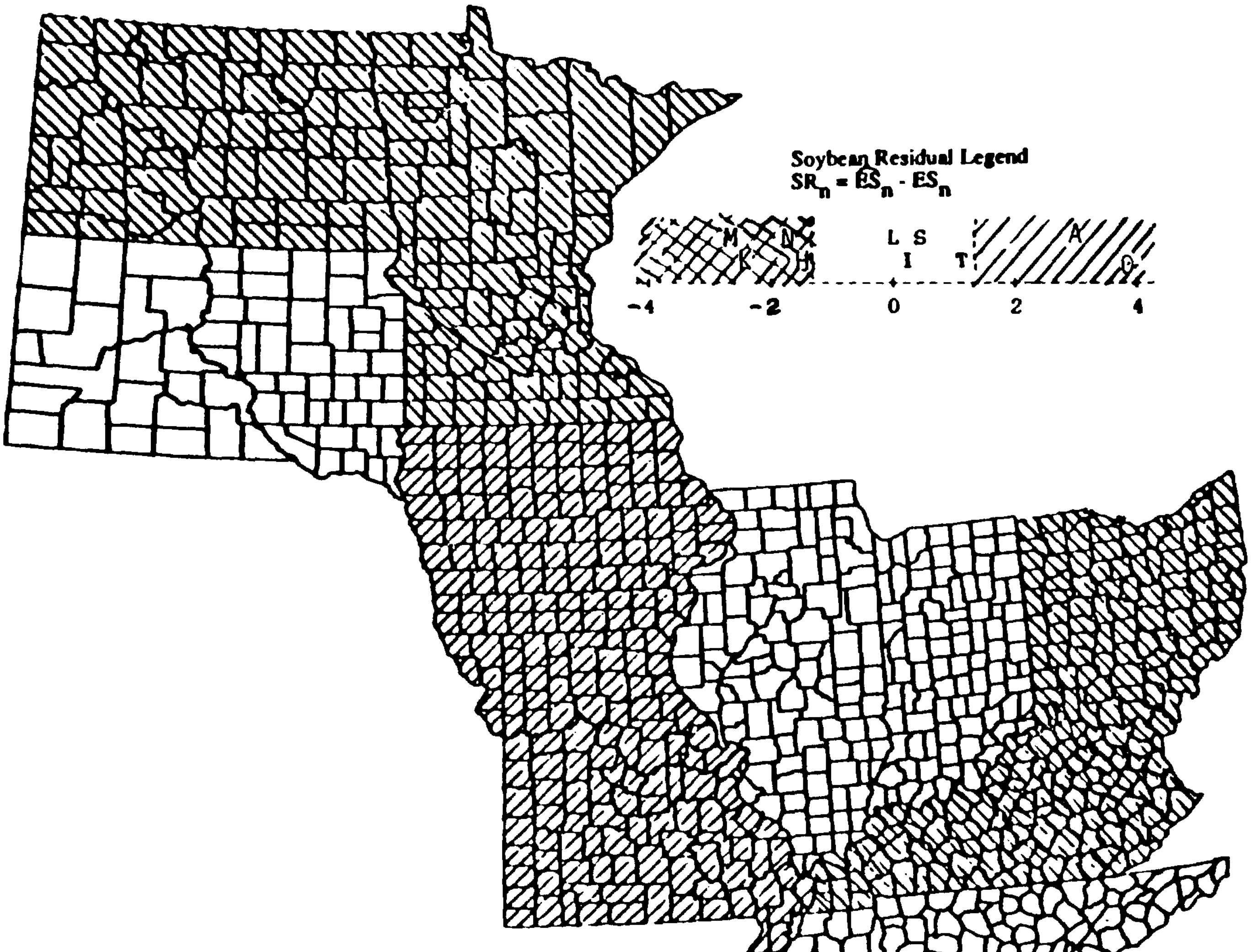


Table 1 summarizes the number of units and strength of relationships at each level of aggregation. The more moderate decline in R^2's for soybeans may relate to a less systematic pattern in the county and agricultural statistics districts outliers than was the case for corn. The number of extreme outlier counties beyond both the lower (21 bushels) and upper (35 bushels) ends of the State level range appears less than it was for corn. However, there are more outliers within the range (most with the satellite values underestimating official yields) and there is generally a greater spread in the data.

TABLE 1. Strength[1/] of relationships between average yields (official estimates) and satellite vegetative indexes or satellite generated yield estimate indications at the State, district, and county levels, 1984, ten State study area.

LEVEL	**CORN FOR GRAIN**			**SOYBEANS**		
	N	**R^2**	**R**	**N**	**R^2**	**R**
STATE	10	.94	.97(.87, .99)	10	.85	.92(.69, .98)
DISTRICT	84	.75	.87(.81, .91)	76	.74	.86(.79, .91)
COUNTY	889	.63	.80(.78, .82)	756	.64	.80(.77, .82)

1/ Strength of relationships are expressed in terms of the coefficient of determination (R^2) and correlation coefficient (R) for the number of observations (N) available at each level. The 95 percent confidence interval for the population correlation coefficient is shown in parentheses.

A More Thorough Examination Using Additional Performance Measures

Looking at results in terms of correlation or regression relationships alone can be misleading. A more thorough examination of the results is presented in the next table. Information from Table 1 is included in this table since it does tell something about how the satellite generated yield indications (and satellite vegetative indexes) correspond to the official estimates. If the satellite generated yield indications were to be used only to proportionally distribute official State mean yields around each State, then the relationship statistics would provide essentially all of the information on their accuracy. However, if the individual county point estimates provided by the satellite generated yields (and those from other sources, also) are to be used directly in setting individual estimates, another set of performance or accuracy measures may be appropriate. Before specifically discussing these other measures of performance, it may be important to discuss the role of official estimates in the assessments.

Official estimates, at any level, are not infallible or immune from error. As was discussed previously, corn and soybean final mean State yield estimates for the States in this study are quite accurate. They provide the best knowledge available on the actual average yield for a State. County mean yield estimates are generally not as accurate as State level yields. Since the official county estimates are used in evaluating the performance of the yield indications, this reservation on accuracy should be kept in mind. For example, suppose we got the extremely unlikely outcome of the county R^2's in Table 1 being equal to one. Then we could use the satellite generated yield indications to duplicate the yield estimates currently produced resulting in no real gain at all! Of course, if the R^2's were quite low (or zero, or not significantly different from zero), we would also be disappointed because the official county yield estimates do correspond to something approximating actual yields. The same kinds of statements could also be made for the other performance measurements (to be discussed next). Perfect results as measured against official estimates would not be very useful. Nor would poor results as measured by these statistics show much promise for the use of satellite generated yield estimates.

The additional performance measurements are based on the mean square error and its components, variance and bias squared. The appropriate roots in the original units (bushels per acre) and a relative error are also included in Table 2. The mean square error is the sum of the squared differences between the satellite generated yield indications or predicted values and the official yields, divided by the number of observations. For example, the study area wide mean square error for county corn yields can be expressed as:

$$MSE = 1/N \sum_{mn} (\hat{EC}_{mn} - EC_{mn})^2,$$

The summation is over all counties for which both variables are defined (N=889, in this case) and EC_{mn} is the 1984 official mean corn yield for the mth county in the nth State (previously not defined since it was not used in the primary analysis). Since the mean square error can be separated into variance (VAR) and bias components, these measures are shown along with the root mean square error (RMSE), standard deviation (ST DEV), and the standard deviation relative to the mean official yield (RSD).

The MSE reflects collectively the accuracy of the individual satellite generated county yields when considering the official county yields as "truth". The variance reflects the precision of these new yield indications when the bias is adjusted out. The variance may be a more appropriate measure in this application because the counties are given equal weight in this analysis. While equal weights are appropriate if one wants to be accurate in all counties, applying the yield indication to individual counties with varying acreages should result in a nearly zero effective bias at the State and higher aggregation levels. Thus, the variance is more indicative of the **TABLE 2.**

Performance measures[1/] at the State, district, and county levels for satellite generated yield estimate indications obtained by considering official estimates as "truth", 1984, ten State study area.

TABLE 2

LEVEL	N	R^2	R	MSE	VAR	BIAS[2/]	RMSE	ST DEV	RSD[3/]
				(bushels/acre)^2		-----bushels/acre----			%
				CORN FOR GRAIN					
STATE	10	.94	.97	26.66	26.66	0.00	5.16	5.16	5.3
DISTRICT	84	.75	.87	181.40	178.76	-1.63	13.47	13.37	13.7
COUNTY	889	.63	.80	249.31	248.69	-0.79	15.79	15.77	16.2
				SOYBEANS					
STATE	10	.85	.92	5.17	5.17	0.00	2.27	2.27	7.9
DISTRICT	76	.74	.86	15.93	14.98	-0.97	3.99	3.87	13.4
COUNTY	756	.64	.80	22.07	21.02	-1.03	4.70	4.58	15.8

1/ The performance measures are discussed at length in the text.

2/ The bias at the district and county level would be very close to zero for a harvested acreage weighted mean. However, all counties (districts) were given equal weight in this analysis.

3/ RSD is the standard deviation relative to the mean (equal weights) corn for grain (97.6 BU./A) and soybean (28.9 BU./A) yields for the ten States.

equally weighted accuracy and precision for individual counties when the resulting bias is essentially zero. An essentially zero bias would occur because the acreage estimates (supported by other data) are constrained to agree with the previously estimated total State acreage harvested and the satellite generated county yields are likewise constrained (individually adjusted by the State level residual) to collectively be consistent with mean State yield per acre. These measurements in terms of bushels per acre squared may be thought of as being analogous to an exponential loss function. That is, a loss function in which to miss by zero bushels "hurts" zero, one "hurts" one, two "hurts" four, three "hurts" nine, and so on.

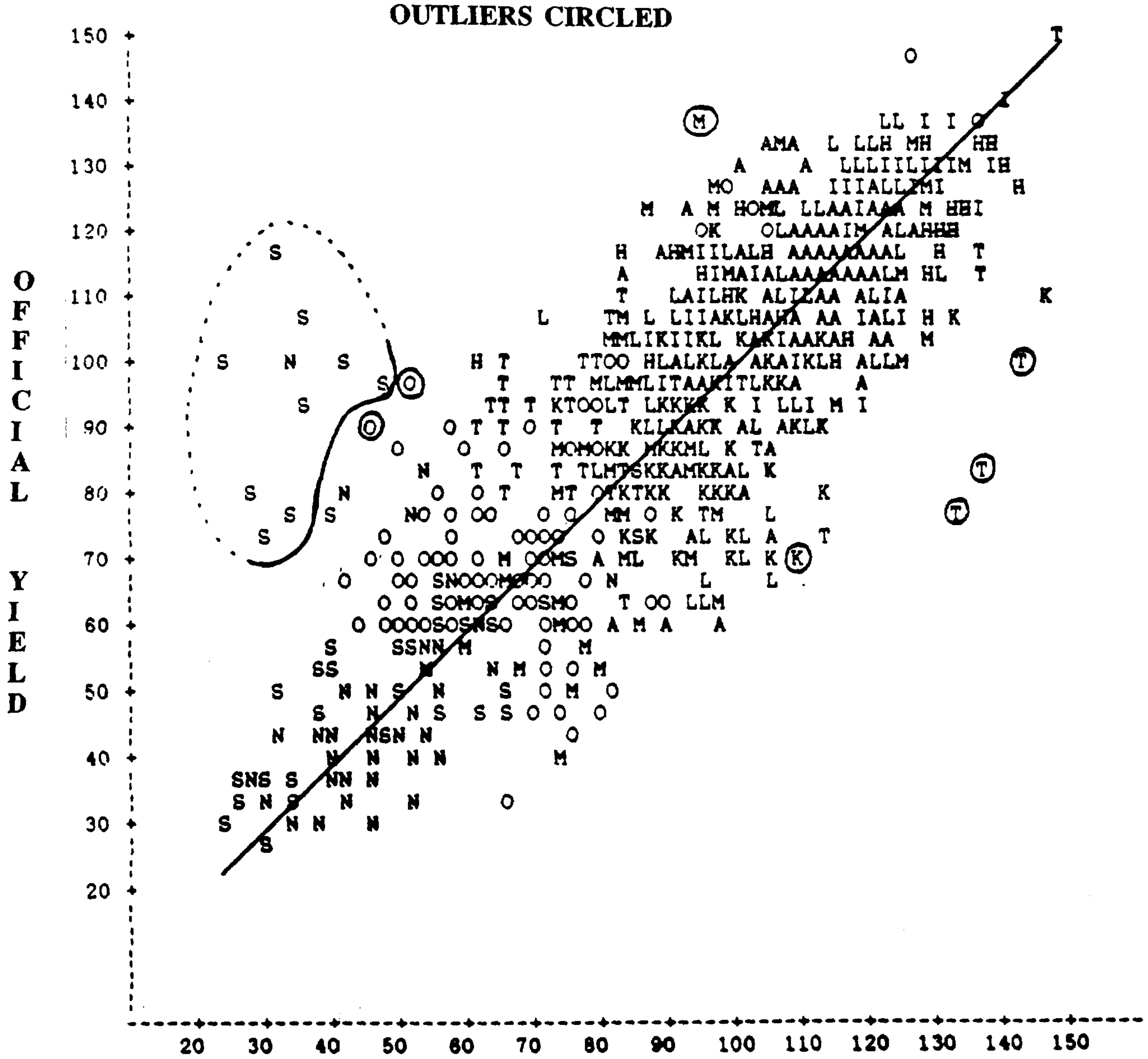
The bias, RMSE and ST DEV are presented in bushels per acre. The bias is positive if the equally weighted county official yields tend to be overestimated by the satellite generated yields. However, there is a little more that needs to be said about interpreting the bias in this situation. Overestimates suggest that a tendency exists to overestimate the yield for counties with lower acreages (within State) since the weighted bias would be closer to zero. In this case, it would appear that the more important counties or areas of a State would tend to be underestimated. In the opposite situation, when the equally weighted bias presented in Table 2 is negative, satellite generated yield indications tend to be too low for all counties, tend to be too low to a greater extent for lower acreage counties, and tend to be overestimates for the higher acreage counties. The RMSE and ST DEV are the square roots of the MSE and VAR, respectively. Because of the bias handling characteristics arising from the county estimation techniques used, the standard deviation is more applicable. To understand its relative value and to consider the performance for both crops it is expressed relative to the mean yield for the study area. The relative standard deviation is the ST DEV divided by the mean (equal weights) yield for the ten States (multiplied by 100 and expressed as a percent).

Additional Performance Measures Summary

The bias at the State level is zero as a result of the regression least squares fit and consequently the MSE and VAR, and the RMSE and ST DEV are equal. Attention may be focused on the ST DEV and the RSD (other than the correlation or regression statistics) to understand the performance of the satellite generated yield indications when considering the official statistics as truth. The standard deviation performance level for corn drops from a little more than five bushels at the state level to around 13 bushels at the district level and then to nearly 16 bushels at the county level. For soybeans, the State to district to county decline in performance as measured by the standard deviation is from two and one fourth, to nearly four, to a little more than four and one-half bushels. The relative standard deviations for the crops are essentially the same at the district and county levels (near 13.5 and 16 percent, respectively). However, the satellite data corresponds more closely for corn at the State level with a RSD of just over five percent as compared to nearly eight percent for soybeans.

Tables, similar to Table 2, which also show the corn for grain and soybean performance for each State were prepared. They are included in Appendix D. An examination of these tables will suggest, performance is not good for some States. However, in many respects, these satellite generated yield estimates should be considered pilot or experimented in nature. Just as would be done in considering other yield indications, the satellite indications should be further evaluated, and some experience acquired their statistical value.

Figure 11
CORN FOR GRAIN - COUNTY LEVEL
Official Yield Estimate (bushels per acre)
Versus
Satellite Generated
Corn Yield Estimate (bushels per acre)
With
OUTLIERS CIRCLED



SATELLITE GENERATED CORN YIELD

SOME COMMENTS ON ACCURACY AND USE OF THE DATA
A Closer Look at Satellite Generated County Corn Yield Indications

In beginning a discussion of accuracy and use of the data, a closer look at the corn for grain satellite generated yield indications may be helpful. In Figure 11, the same plot of the county level corn data is shown that was previously presented in Figure 7. Individual outliers are circled (and encircled). They will be given further consideration in hopes of understanding some characteristics about those counties which may explain why the satellite generated yield indications performed poorly. The group of encircled counties totals 11 (one data point represents two South Dakota counties). Understanding why the actual yields, as approximated fairly well (one can assume) by the official estimates, were substantially underestimated by the satellite generated yields is not difficult for these 11 counties. The vegetative conditions of western South Dakota and North Dakota were quite poor in 1984 (as they usually are relative to the entire ten State area) and a fairly small proportion of the region is in crops. Since region wide and county by county vegetation is sparse (to some the area can appear quite bleak), vegetative indexes are low. This, of course, results in low satellite generated corn yield estimates (less than 50 bushels per acre for these counties). So, why are the actual (and official) corn yields so high? Most of the corn acreage is irrigated. In addition to irrigation lifting corn yields much higher, the low irrigated and other crop acreage keeps their effect on the vegetative indexes to a minimum. The crop areas (particularly the irrigated ones) simply do not cover enough of the land to have much effect on average values from the satellite data.

It may be desirable to objectively identify types of counties and characterize the usefulness of the satellite generated indications for each category. Such an attempt was made for counties which have a substantial proportion of their corn acres irrigated. However, irrigation statistics are not available for all ten States in 1984. Therefore, in order to objectively group all study area counties of the type encountered in this problem a more complete data set was needed. The 1982 Census of Agriculture provides such a data source. The number of farms with corn harvested for grain is provided for all farms and for those with some of the crop irrigated.

TABLE 3. Corn harvested for grain: Total acreage, proportion irrigated and average yields (official and satellite generated), 1984, selected [1/] counties.

COUNTY	(ARD) CROP REPORTING DISTRICT	TOTAL ACREAGE	PROPORTION IRRIGATED [2/]	AVERAGE YIELD OFFICIAL	AVERAGE YIELD SAT. GENERATED
			%	-------bushels/acre-------	
SULLY	5	43,900	42	77	35
BUTTE	1	12,500	92	100	24
TODD	8	11,400	65	98	43
BUFFALO	5	7,600	53	98	48
LYMAN	8	6,800	50	92	36
FALL RIVER	7	4,600	--	118	31
BENNETT	7	3,600	72	107	36
STANLEY	4	2,000	--	93	36
MEADE	4	1,000	--	79	29
SHANNON	7	200	--	75	31
MCKENZIE	4	200	100	101	33

1/ The selected counties are the encircled outliers in Figure 11. All are Western South Dakota counties, except for McKenzie (which is located in West Central North Dakota).

2/ Fall River, Stanley, Meade and Shannon irrigated acreages were not published separately. Published district data shows 87 percent of the 9,400 acres in district 7 (Southwest South Dakota) and 40 percent of the 7,200 acres in district 4 (West Central South Dakota) being irrigated in 1984.

Total and irrigated acres of corn harvested for grain are available for most counties. In the few counties where 1982 acreage is not provided, (to avoid disclosing information on the few operations involved) the proportion of farms with irrigated corn provides a basis for judging the importance of irrigation. Using data such as that provided by the Census of Agriculture also allows identification of groups of counties before the satellite yield estimates are generated or the official estimates are produced.

Results Obtained by Applying the Irrigation Rule

An attempt was made to learn the impact of objectively eliminating a group of counties with substantial proportions of irrigated corn. After trying several alternatives, it was decided to exclude those which irrigated more than 30 percent of their corn harvested for grain acreage. Basically, the 30 percent cut off eliminated the more obvious outliers without excluding as many additional counties as lower proportions would. However, to eliminate the 11 outliers listed in Table 3 satellite generated corn yield indications were effectively discarded for 20 additional counties. The resulting plot of the surviving official/satellite generated county yield pairs are shown in Figure 12. In Appendix D the corn for grain performance measures (like those shown in Table 2) are presented when the 31 counties are excluded. The appendix table shows the performance for the ten State area and individual States when the objective rule is applied. There are notable improvements for some of the States.

In Table 4 some comparative values have been excerpted from the county level corn for grain performance tables in Appendix D. They are presented in terms of the number of counties covered and benefits of using the Census of Agriculture irrigation data to objectively reject use of the satellite yield indications for some counties.

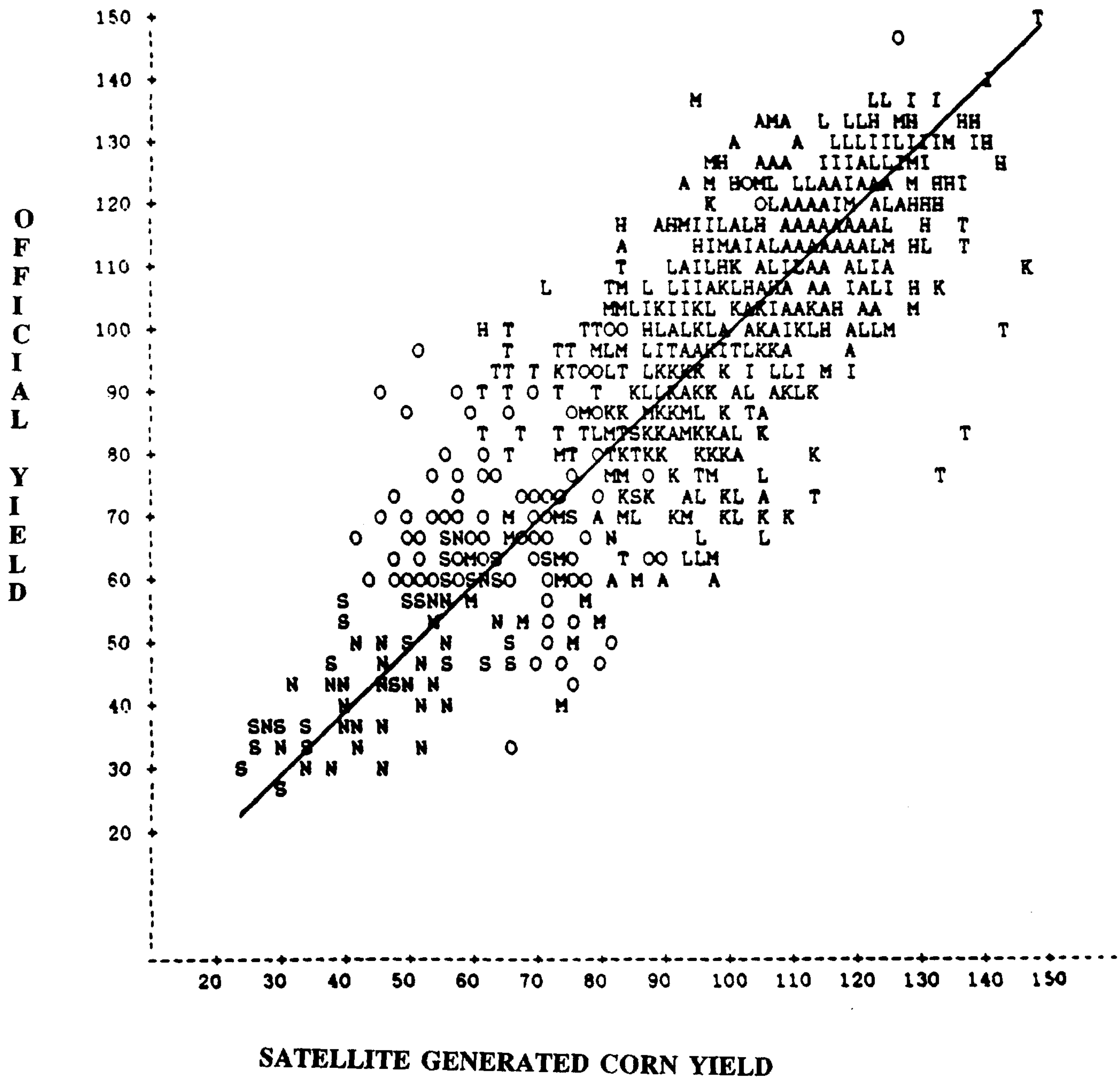
TABLE 4. Number of counties covered and gains[1/] of excluding counties based on more than 30 percent of corn harvested for grain being irrigated in 1982, 1984, ten State study area, selected county level corn for grain performance measures.

APPLICATION AREA	COST N	GAINS R^2	RSD(%)
	----(ALL COUNTIES/EXCLUSION RULE APPLIED)---		
TEN STATES	889/858	.63/.69	16.2/14.5
NORTH DAKOTA	47/ 40	.29/.69	22.2/12.0
SOUTH DAKOTA	62/ 49	.04/.78	36.4/12.8
MINNESOTA	81/ 78	.56/.59	15.1/14.7
MISSOURI	114/107	.30/.19	21.0/21.3
ILLINOIS	102/101	.38/.38	12.5/12.6

/ Number of counties covered and gains are measured when the objective exclusion rule is applied to counties with a substantial proportion of their corn for grain irrigated against the alternative of not excluding any of the 889 counties for which satellite generated and official yield data exists.

One could have looked at other criteria for grouping counties (for both corn and soybeans) into various categories of usefulness. For example, counties with little soybean acreage could have been identified where that acreage possibly is restricted to more advantageous local areas within the counties. One could also have identified counties with few crop acres, where vegetative conditions, and thus satellite derived yield estimates, for the entire county could be quite different than for the crop area within the county. Perhaps, areas with substantial woodland could be grouped into some type of performance category. Many possibilities could have been attempted; however, a fairly simple one was employed. It's application demonstrated that eliminating some obvious outliers (ones that could be detected even without knowledge of the official estimates because they are clearly too low) could improve the relational and accuracy performance measures somewhat.

Figure 12
CORN FOR GRAIN - COUNTY LEVEL
Official Yield Estimate (bushels per acre)
Versus
Satellite Generated
Corn Yield Estimate (bushels per acre)
with
SOME COUNTIES OBJECTIVELY ELIMINATED



CONCLUSION

In conclusion, it has been shown that at the State level within a single year satellite derived vegetative index variables are statistically correlated to corn and soybean yields. An application of such strong State level within-year relationship has been illustrated and the possibility of other applications suggested. The methods employed in aggregating the satellite data to obtain the appropriate State crop specific vegetative indexes have been presented in sufficient detail to facilitate duplication or to allow research on alternative techniques.

The application of generating satellite corn for grain and soybean county yield indications shows promise. Satellite vegetative index values by themselves could have been used to provide information on relative crop condition. The calibration and verification of their explanatory power at the State level within the year of application, however, provides the important assurance that (at least at the State level) they are strongly related to yield. The nature of the relationship of satellite data (from a particular satellite, recorded by a specific sensor, constructed as a defined vegetative index, aggregated to a specific area (grid cell) in a certain way, averaged over a selected time period, mapped to counties by a specified algorithm and weighted to the State level by available or constructed county crop weights) to official crop yield estimates (arising from a certain type of crop year, rate of development, mix of crops, condition of other vegetation, etc.) can be measured at the State level for a broad spectrum of important agricultural States and applied to individual counties or groups of counties such as agricultural statistics districts.

These conclusions are based on a single year, 1984. The study should be repeated for additional years, with similar although different meteorological satellites, sensors, crop development patterns, crop mixes, and other characteristics.

The application of satellite generated yield forecasts for agricultural statistics districts and counties should continue to have high priority. This has been conducted for the 1988 yield forecasts with similar (actually higher correlations than 1984) results. Recall that 1988 was a severe drought year for corn and soybeans in these States. This type of method, perhaps in combination with early season objective yield and daily ground weather observation data models could be the only foreseeable improvement in methodology for early season crop yield forecasting.

There are several other possible areas that could be explored. They involve changes in the way the data are summarized for the grid cells. Currently this is an FAS function. Any changes in the processing system would involve FAS agreement and/or a greater role by NASS in this area. The potential changes involve altering the data screening and averaging or summarization procedures. One possible improvement would be to compute grid cell averages only for pixels with a vegetative index above a certain threshold. That is, the current fixed threshold (at the so called soil line) would need to be adjusted to a higher (perhaps variable level) so that the vegetative index reflects conditions similar to that of crops in good enough condition to justify a harvest. Similar changes might also be required to investigate crop condition assessment methods for other crops, such as wheat or cotton. Other changes might involve those effecting the cloud cover and screening bias problem discussed in Appendix A. Still other changes might call for smaller grid cell sizes or flexible locations. This might be particularly important to the generation of useful supplemental variables for improved yield survey efficiencies.

Another group of potential future research efforts could be applied to many of the methods presented in this report. Averaging grid cell vegetative indexes over time by employing a functional fit similar to that employed by Boatwright could be considered. The benefits of conducting a manual or automated edit of the daily grid cell vegetative index values could be investigated. Employing a flexible crop stage indicator, or crop calendar, to shift the critical period by local areas could be explored. This attempt to tie the critical period more directly to crop progress would require additional data and impose the burden that the critical period specifying algorithm give equivalent results for all areas.

Many of the other steps in the primary analysis could be considered for modification. Kriging theory (spatial estimation) could be employed to find more optimum ways of mapping grid cell means to counties, perhaps with differential decay functions in various directions for different areas and crop seasons. Ways of modifying the State level residual adjustment could be investigated that avoid artificial differences near State borders and which would potentially improve the accuracy of county estimates.

RECOMMENDATIONS

The application of satellite generated yield forecasts for agricultural statistics districts and counties should continue to have high priority. This has been conducted for the 1988 yield forecasts with similar (actually higher correlations than 1984) results.

There are several other possible areas that could be explored. They involve changes in the way the data are summarized for the grid cells. Currently this is an FAS function. Any changes in the processing system would involve FAS agreement and/or a greater role by NASS in this area. The potential changes involve altering the data screening and averaging or summarization procedures. One possible improvement would be to compute grid cell averages only for pixels with a vegetative index above a certain threshold. That is, the current fixed threshold (at the so called soil line) would need to be adjusted to a higher (perhaps variable level) so that the vegetative index reflects conditions similar to that of crops in good enough condition to justify a harvest. Similar changes might also be required to investigate crop condition assessment methods for other crops, such as wheat or cotton. Other changes might involve those effecting the cloud cover and screening bias problem discussed in Appendix A. Still other changes might call for smaller grid cell sizes or flexible locations. This might be particularly important to the generation of useful supplemental variables for improved yield survey efficiencies.

Another group of potential future research efforts could be applied to many of the methods presented in this report. Averaging grid cell vegetative indexes over time by employing a functional fit similar to that employed by Boatwright could be considered. The benefits of conducting a manual or automated edit of the daily grid cell vegetative index values could be investigated. Employing a flexible crop stage indicator, or crop calendar, to shift the critical period more directly to crop progress would require additional data and impose the burden that the critical period specifying algorithm give equivalent results for all areas.

Many of the other steps in the primary analysis could be considered for modification. Kriging theory (spatial estimation) could be employed to find more nearly optimal ways of mapping grid cell means to counties, perhaps with differential decay functions in various direction for different area and crop seasons. Ways of modifying the state level residual adjustment could be investigated that avoid artificial differences near state borders and which would potentially improve the accuracy of county estimates.

APPENDIX A

CLOUD SCREENING
BIAS
STUDY

A cloud screening bias study was conducted because there was a tendency for vegetative index values to average lower when a larger portion of a grid cell's pixels were screened out. Also, the amount of satellite data was sparse enough in some areas that merely discarding potentially biased data was not an attractive alternative. The approach taken in this study was to learn more about the nature of the bias. The idea, based on this knowledge, was to conclude either that the bias could be safely ignored or adjust for it in some satisfactory way.

It was speculated that vegetative index mean grid cell values (for both the EVI and NVI versions) were lower when more pixels were screened out. This was confirmed in an earlier exploratory study. State means were lower for grid cells with up to 50 percent of their pixels screened out as opposed to a maximum of 25 percent. This downward bias might result because of the kinds of pixels remaining when others are screened out. Pixels associated with those removed may contain haze, thin clouds or cloud shadows, all of which tend to depress index values. Based on visual satellite image observations the author was tempted to conclude that the bias was greater when scattered clouds were present. Vegetative index values for grid cells observed near the same date seemed to be altered less when the images showed solid cloud masses or definitive fronts rather than scattered clouds. Thus, it could have been hypothesized that the amount of bias was some function of cloud boundary length, but that was beyond the scope of this study.

Instead, the overall relationship of the vegetative indexes to the percentage of pixels not screened out was examined. Such an examination is shown for EVI in Figure A-1. This analysis was performed for grid cells in a coordinate system rectangle around the study area (i th from 210 to 260 and j th from 340 to 390). The analysis and Figure include available daily data over the period from July 31 through August 23, 1984 (the selected corn period). The model regressing EVI on the percentage of pixels not screened out (GDPIX or % good) is, of course, highly significant in all respects. This results because of the large number of observations N=2338), even though the R^2 is only .12.

Since regional patterns of cloudiness, cloud types and index levels could cause spurious relationships between EVI and GDPIX, some additional analyses were completed. The model of dependence of EVI on GDPIX was looked at both by regions and with the i th and j ith coordinates included as co-variables. Another type of analysis was motivated by another consideration. The only true test for this dependence would be to apply the full range of treatments (proportions screened out) to each daily observation. This controlled study is, of course, impossible (only one proportion is realized for each daily grid cell observation). In lieu of this "ideal" test, the next best thing was attempted. The regression relationship was computed for each grid cell over the available daily data, and mean intercepts and slope parameters computed for the entire area. This eliminated grid cells with fewer than three daily observations from the analysis (since the regression could not be computed for them). Grid cell's with defined regression parameters were weighted together in proportion to their degrees of freedom (with some extreme parameter estimates edited out) to produce aggregate estimates.

While the various analyses employed (by regions, with coordinate system co-variables and aggregation of individual grid cell relationships) did reveal some variation between individual grid cells and between regions, they generally supported the overall slope parameter estimate (0.46). As shown below, only the model slope parameter would be involved in attempting to satisfactorily adjust the EVI's.

To adjust the EVI's to the value they would be presumed to have for "clear skies" (or no pixels screened out), one can visualize lifting the line (denoted by asterisks in Figure A-1) by the left end until the slope is zero and maintaining the observations the same residual distance from the line. That is, the modified EVI (MDEVI) should be;

$$MDEVI = EVI_{100} + \text{Residuals}$$

where EVI_{100} is the EVI for GDPIX=100 and the residuals are those from the regression model ($EVI = \hat{\alpha} + \hat{\beta}$ GDPIX). Substituting in the above equation it is seen that,

$$\begin{aligned} MDEVI &= [\hat{\alpha} + \hat{\beta}(100)] + (EVI - \hat{EVI}) \\ &= \hat{\alpha} + \hat{\beta}(100) + EVI - [\hat{\alpha} + \hat{\beta}(GDPIX)] \\ &= \hat{\alpha} + \hat{\beta}(100) + EVI - \hat{\alpha} - \hat{\beta}(GDPIX) \\ &= EVI + \hat{\beta}(100 - GDPIX). \end{aligned}$$

This is the intuitively pleasing result that the modified EVI's are just the original EVI values plus a constant add on amount ($\hat{\beta}$) for each percent of pixels screened out.

Results of applying this adjustment for the selected corn period are shown in Figure A-2. Here, MDEVI = EVI + 0.46 (100-GDPIX). As verified by the fitted line the overall bias has been eliminated. However, one can note some values along the periphery which appear to have been adjusted too much. The lower periphery shows no EVI's (modified or not) near the soil line for the lower GDPIX values and the top shows some values much larger than the usual maximum. These observations may arise from grid cells where the screening procedure worked well, even though many pixels were screened out, and unbiased or less biased values were adjusted too much. Of course if that is the case , then other grid cell daily observations for which the bias was greater may not have been adjusted enough. This tendency for some individual adjustments to be inappropriate may be an indication that the model of dependence on GDPIX provides an incomplete explanation of the bias.

Figures A-3 and A-4 show the EVI and MDEVI relationships with GDPIX, respectively, for the selected soybeans period (July 31 - September 1, 1984). A variety of analyses, again, support the value of the slope parameter estimate given by the overall model ($\hat{\beta}$= 0.42), but indicate considerable variation in adjusting individual daily grid cell vegetative indexes.

Table A-1 shows the matrix of R^2 values for relationships of the August 1 and October 1 forecasts and the final yield estimates for the corn and soybean crops to the modified vegetative index (MDEVI) state means for each of 36 varying length periods. The individual daily grid cell MDEVI's were computer based on the regression slope derived for that individual time period (from the start of the "FROM" period through the ending date of the "THRU" period). Then the predicted values (MDEVI's) were averaged over the period and aggregated to the state level as described in Appendix B (by Crop Reporting Districts, rather than countries).

Figure A-1
Environmental Vegetative Index (EVI)
Versus
Percentage of Good Pixels (GDPIX)

For Available Daily Grid Cell Observations
In And Around The Ten State Study Area,
July 31-August 23, 1984 (Selected Corn Period)

Plot of PREDICT*GDPIX — Symbol used is *
Plot of EVI*GDPIX — Legend: A = 1 obs, B = 2 obs, etc.

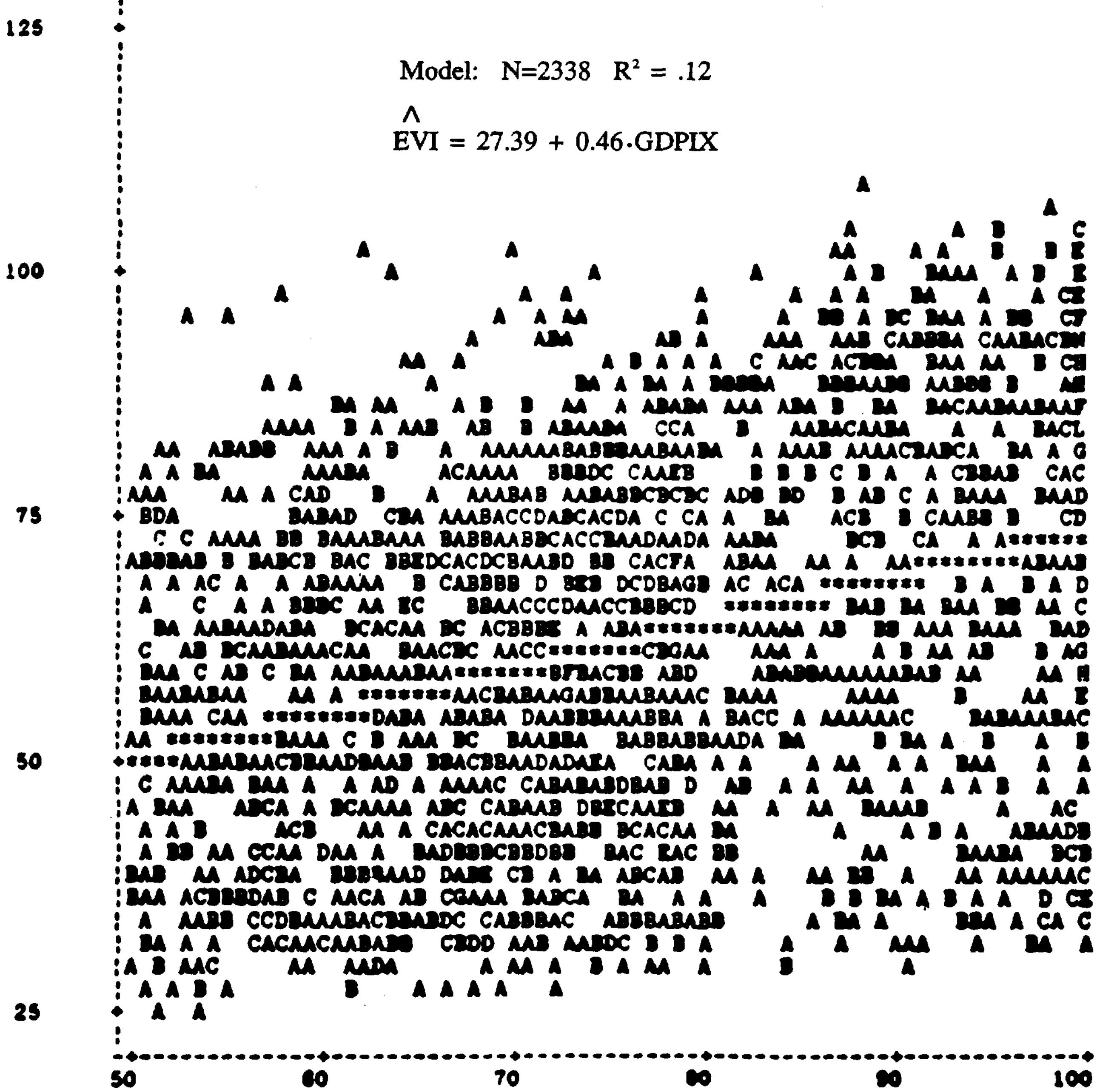


Figure A-2
Modified (To 100% Good Equivalent)
Environmental Vegetative Index (MDEVI)
Versus
Percentage of Good Pixels (GDPIX)

For Available Daily Grid Cell Observations
In And Around The Ten State Study Area,
July 31-August 23, 1984 (Selected Corn Period)

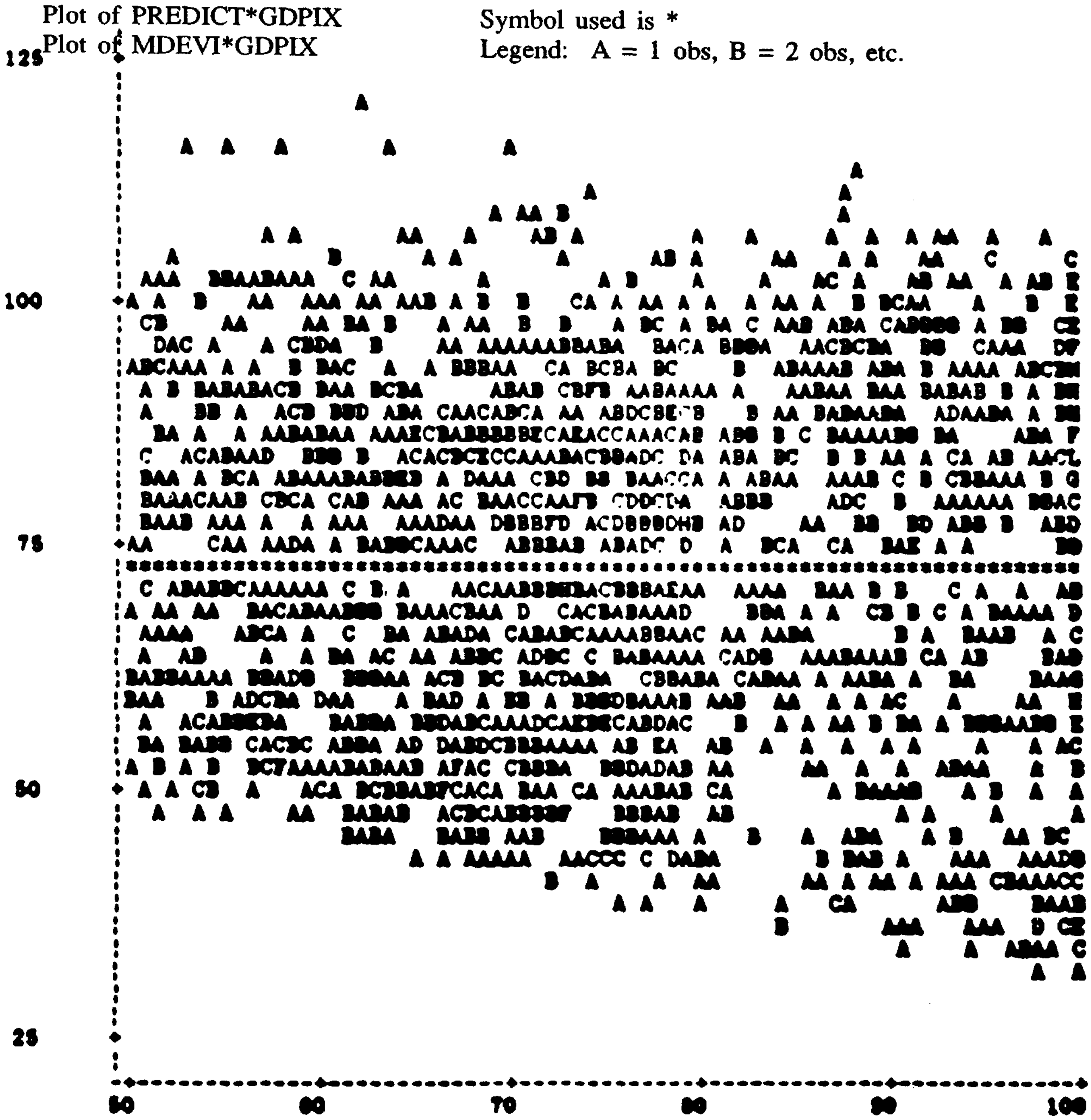


Figure A-3
Environmental Vegetative Index (EVI)
Versus
Percentage of Good Pixels (GDPIX)

For Available Daily Grid Cell Observations
In And Around The Ten State Study Area,
July 31-September 1, 1984 (Selected Soybean Period)

Plot of PREDICT*GDPIX — Symbol used is *
Plot of EVI*GDPIX — Legend: A = 1 obs, B = 2 obs, etc.

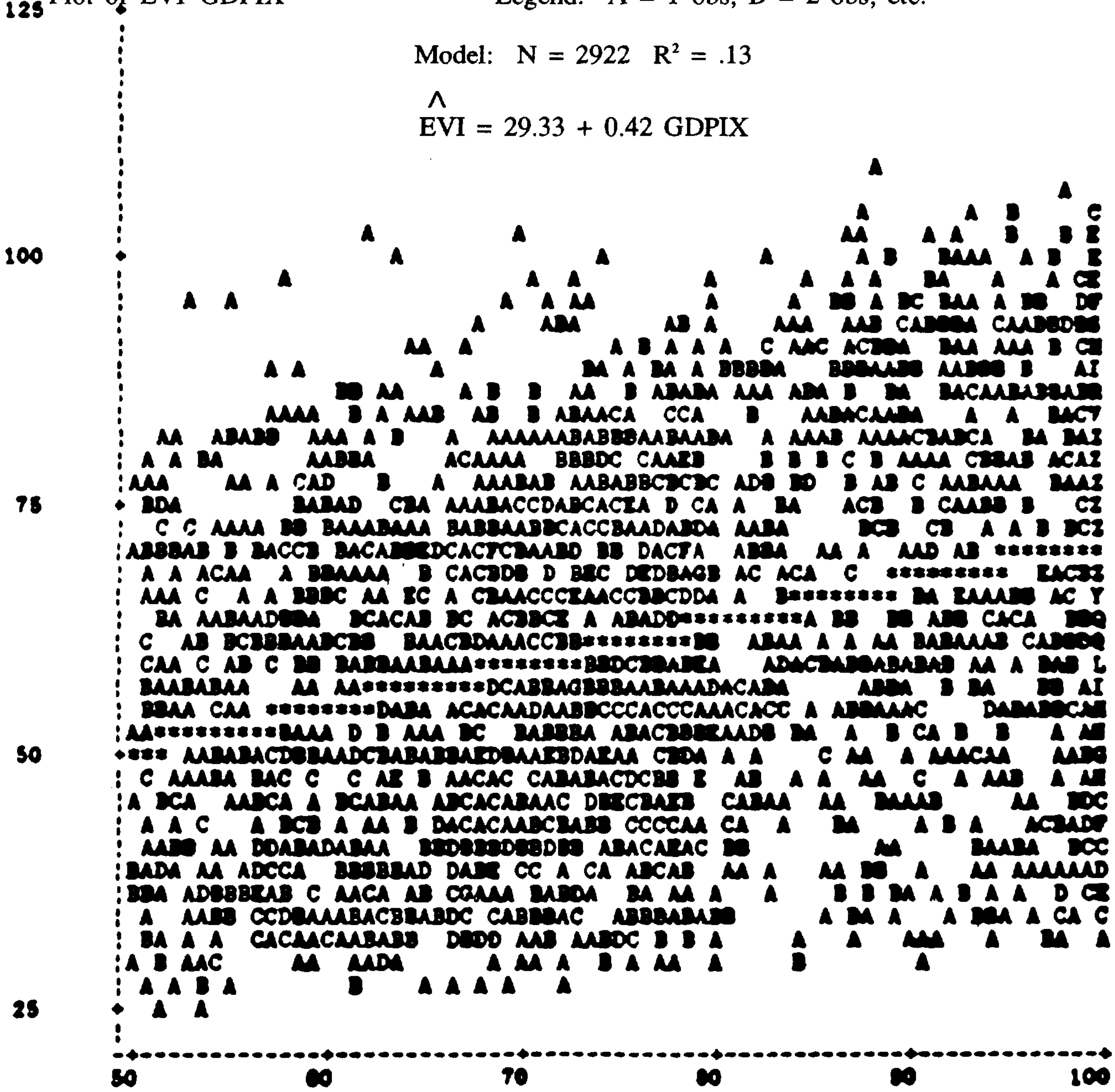
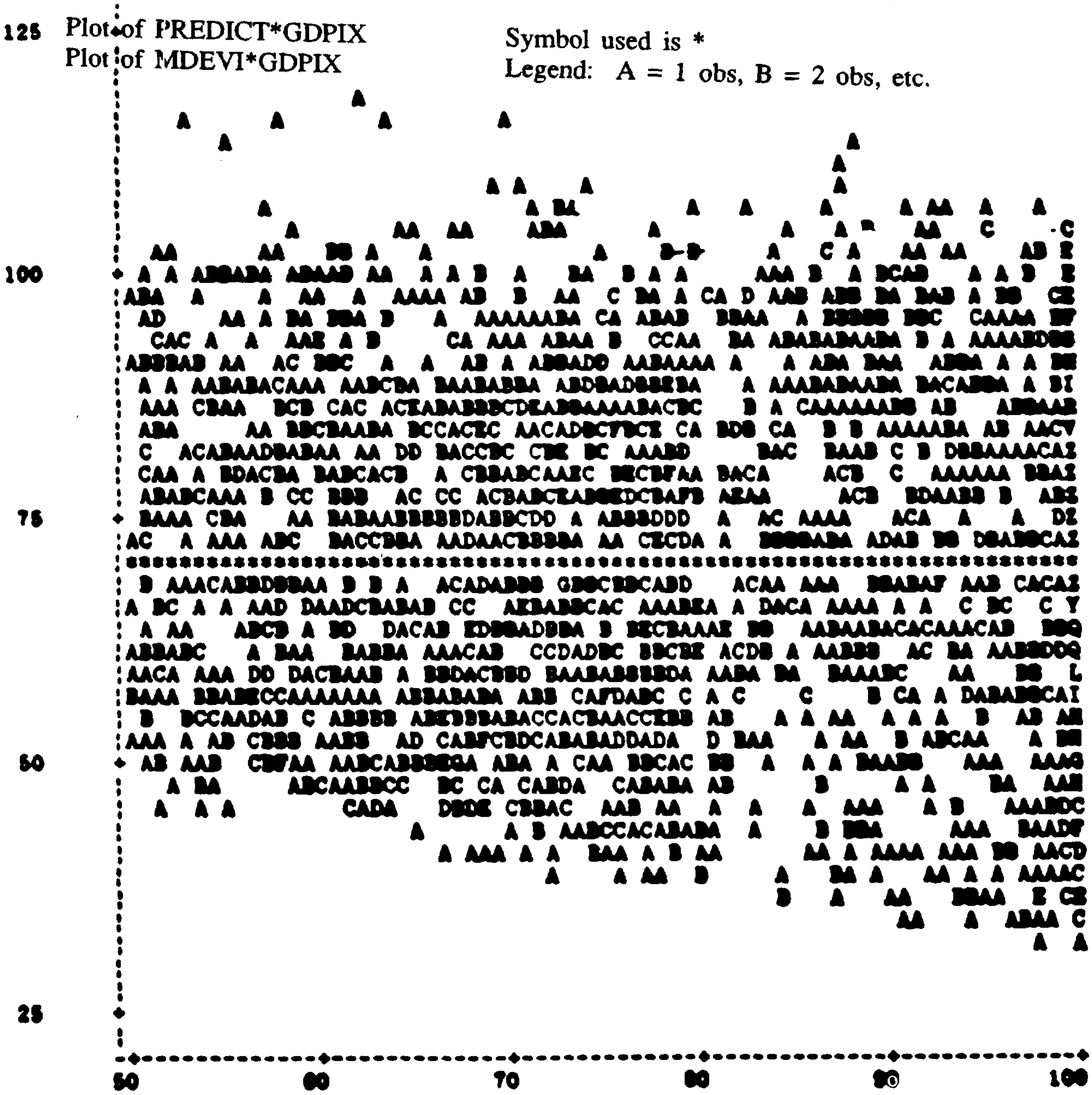


Figure A-4
Modified (To 100% Good Equivalent)
Environmental Vegetative Index (MDEVI)
Versus
Percentage of Good Pixels (GDPIX)

For Available Grid Cell Observations
In And Around The Ten State Study Area,
July 31-September 1, 1984 (Selected Soybean Period)



Comparison between Table A-1 and the analogous table for EVI's (unmodified) in Appendix B indicates the improvement, lack of change or deterioration in state level relationship strength for each period, and forecast or estimate month. These comparisons provide some insight to the success or failure of the attempt to modify the EVI's.

The results are not very encouraging in that adjusting for this bias does not make all the relationships stronger or even result in improvement for a majority of them. Perhaps the most important result is that the strength of the stronger relationships does not differ much between the unmodified and modified indexes. These R^2's are as similar as they are primarily because grid cells are averaged over enough daily observations for the best periods. The tendency of the adjustment for such periods to average out to something like a constant is an argument for not making any adjustment. The hope is that enough days of data exist in the selected periods so that the bias is "averaged out" for many of the grid cells.

Ignoring the bias was the course chosen in this study. Perhaps discarding data below some percent good threshold could be considered as a way of minimizing the bias without loosing too much data. However, the very linear relationship of the 1984 data does not suggest such a threshold. The bias study provides some basis for not attempting to make the adjustment and suggests an approach for appraising the same problem in other years. A more complete understanding of the bias can best be used to support developing improved screening capabilities. This use would directly address the problem, rather than using information obtained to merely to adjust "noisy" data values in a "noisy" manner.

TABLE A-1. Coefficient of determination (R^2) between modified vegetative indexes (MDEVI version), averaged over the "FROM" through the "THRU" period, with the August 1 and October 1 yield forecasts and the final yield estimate, corn for grain and soybeans, ten state study area, 1984

FROM -> ->		7/8-15		7/16-23		7/24-30		7/31-8/7		8/8-14		8/15-23		8/24-9/1		9/2-9	
THRU		Corn	Soybeans	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans
	Aug.	.00	.00														
7/8-15	Oct.	.00	.03														
(A)	Final	.01	.00														
	Aug.	.66	.55	.70	.63												
7/16-23	Oct.	.57	.45	.40	.32												
(B)	Final	.41	.36	.29	.23												
	Aug.	.89	.78	.91	.79	.77	.65										
7/24-30	Oct.	.80	.55	.79	.52	.82	.63										
(C)	Final	.69	.50	.68	.49	.78	.62										
	Aug.	.93	.94	.93	.87	.88	.73	.73	.66								
7/31-8/7	Oct.	.89	.73	.88	.66	.90	.65	.90	.73								
(D)	Final	.87	.73	.87	.68	.94	.71	.96	.81								
8/8-14	Aug.	.92	.90	.92	.86	.87	.71	.78	.63	.50	.50						
(E)	Oct.	.92	.80	.92	.76	.90	.70	.94	.79	.67	.74						
	Final	.90	.79	.89	.77	.91	.72	.95	.82	.62	.69						
8/15-23	Aug.	.92	.88	.92	.86	.85	.73	.79	.69	.74	.63	.75	.66				
(F)	Oct.	.93	.83	.92	.82	.90	.75	.92	.83	.87	.82	.83	.74				
	Final	.91	.82	.90	.81	.91	.77	.92	.83	.86	.81	.84	.74				
8/24-9/1	Aug.	.87	.81	.85	.78	.76	.62	.67	.50	.58	.39	.53	.37	.28	.10		
(G)	Oct.	.90	.81	.87	.80	.83	.69	.84	.72	.76	.67	.67	.57	.49	.26		
	Final	.90	.82	.87	.80	.87	.73	.87	.77	.80	.72	.73	.63	.56	.32		
9/2-9	Aug.	.85	.75	.84	.72	.77	.60	.68	.48	.60	.39	.54	.35	.48	.12	.37	.10
(H)	Oct.	.91	.76	.89	.74	.85	.64	.86	.68	.80	.65	.72	.59	.66	.27	.30	.06
	Final	.93	.79	.91	.77	.90	.69	.91	.74	.85	.72	.78	.64	.74	.31	.31	.05

APPENDIX B

SELECTION OF CRITICAL PERIODS WITH SATELLITE VEGETATIVE INDEXES STRONGLY RELATED TO CORN AND SOYBEAN YIELDS

The selection of critical time periods to average satellite vegetative indexes for useful relationships to corn and soybean yields is motivated by two concepts. One concept is that a time period exists when general vegetative conditions are indicative of the suitability of the environment during the critical yield determining part of the respective crop's life. The other concept is that averaging multiple vegetative index values over a period of time should mitigate some of the "noise" in the daily values.

Critical period selection involved two complementary methods. One approach was to observe the seasonal pattern of grid cell vegetative indexes. It was desirable to identify a relatively consistent period of index values or what might be regarded as a "greenness plateau." Such a "plateau" would occur after a period of "greening up," or perhaps following some "greenness" associated with pre-ripe small grain crops (or other earlier vegetation), but prior to the "greenness decline" that eventually accompanies the approaching fall season. The "plateau period" would provide observations on multiple dates when general vegetative conditions could be considered essentially stable. Conceptually each vegetative index value within the period, no matter how "noisy," would provide information on these stable conditions. So, taking an average over the period would reduce the "noise" around a common value.

The other method examined the strength of relationships between various yield forecasts and estimates to vegetative index values created for various time periods during the summer of 1984. Grid cell means were created for each time period and aggregated (by equal weights to the Crop Reporting Districts then by crop specific district weights) to the state level so that the relationships could be appraised for the entire ten state area. Eight short periods from July 8 through September 9, 1984, were examined. Each of these approximately eight day periods was long enough to provide fairly uniform coverage of the entire study area; however, within state coverage was often incomplete and could result in unrepresentative data at the state level. Examination of these short periods was limited to determining if there was any information on a crop's yield from the vegetative indexes in each time interval. Adjoining periods were then combined by twos, threes and so on, so that data representiveness was improved and the strength of relationships could be considered when more index values are averaged. Several very competitive longer periods were identified using this method. Each of these periods was composed of short periods which individually demonstrated some relationship to crop yields and which when combined with adjoining short periods achieved strong relationships.

The seasonal pattern of grid cell vegetative index values was observed in many different ways. Because there are over 1000 grid cells in and around the study area not all could be looked at individually, nor is that advisable. Since a single interval is to be selected for each crop over the entire area, one does not want the selection tailored too much for an individual grid cell or local area. However, one would not like the period selected to cause serious misrepresentations of the respective crop's actual condition for very many areas.

Figures B-1 through B-10 provide some idea of the vegetative indexes for individual grid cells. The figures show the functional lines of the "greenness curve" for three grid cells in each state. The function is merely the straight line between daily environmental vegetative index (EVI) values available from July through September 1984. The grid cells in the figures were chosen to (1) be from different areas of each state, (2) provide an illustration of the different patterns found in the ten state area and (3) have enough separation so three could be shown. The grid cells are from the western, central and eastern parts of each state. The western grid cell "greenness curve" is denoted by an asterisk (*), the central one by an at sign (@) and the eastern most one by a plus sign (+). Patterns shown in the figures include curves "greening up," those "plateauing" after passing through an earlier "small grains greenness" period and some with very little data (which also, unfortunately, is illustrative of patterns present in the area).

Figure B-1
North Dakota Greenness Curves
Vegetation Index Functional Values - EVI Version
(EVILINE)
Versus
Days in July - September 1984
(DAY)

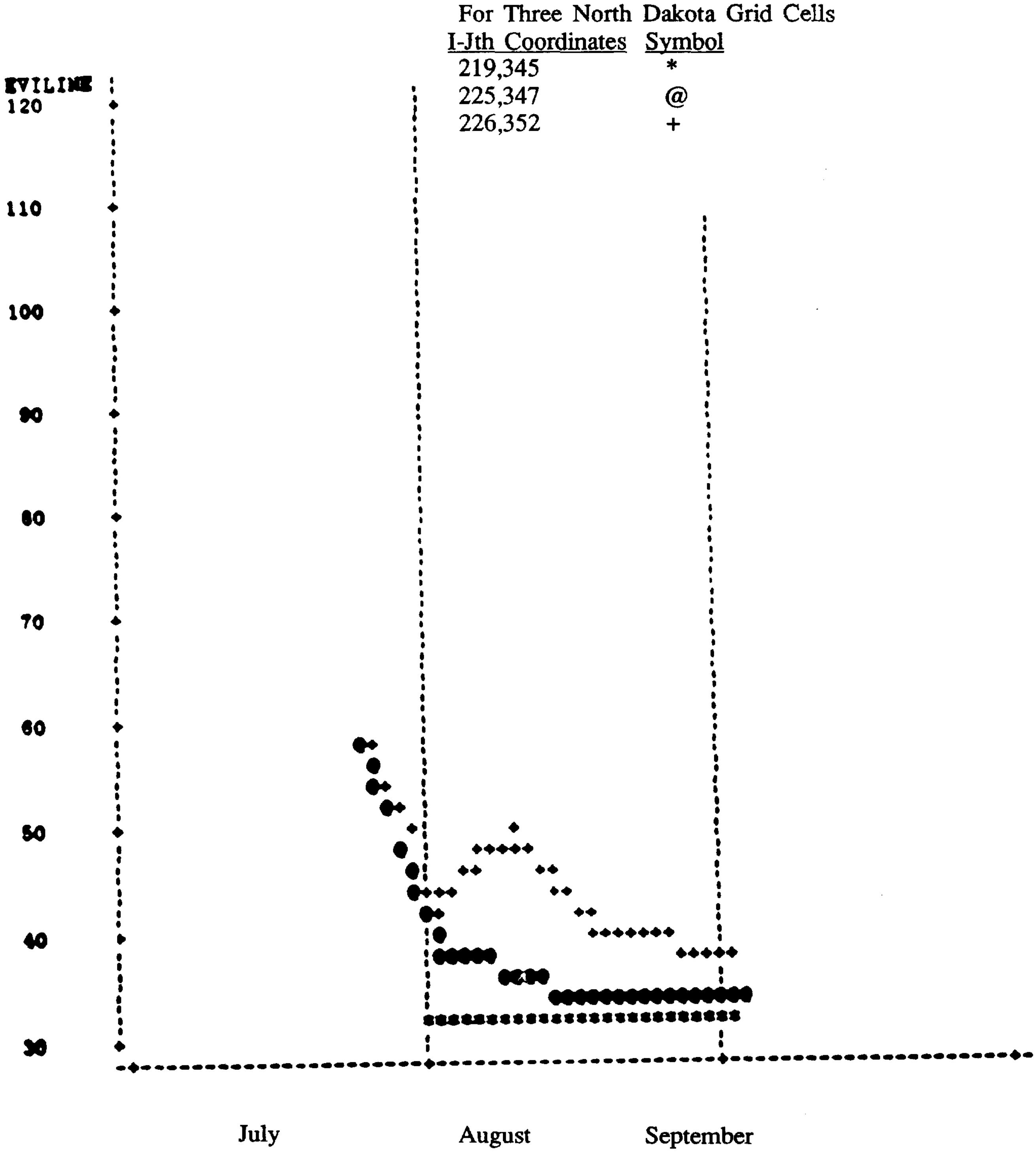


Figure B-2
South Dakota Greenness Curves
Vegetation Index Functional Values - EVI Version
(EVILINE)
Versus
Days in July - September 1984
(DAY)

For Three South Dakota Grid Cells

I-Jth Coordinates	Symbol
220,356	*
224,356	@
227,356	+

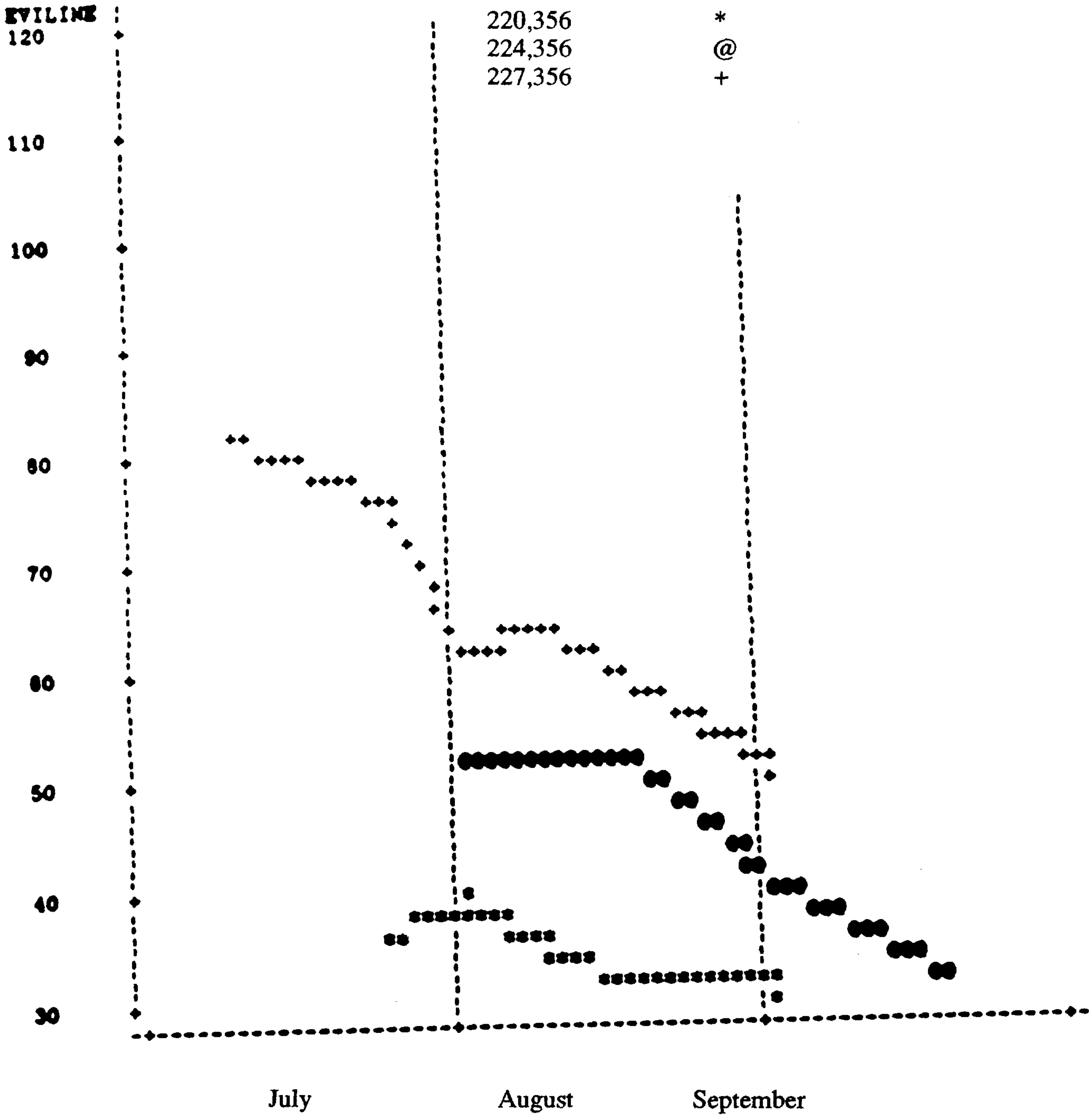


Figure B-3
Minnesota Greenness Curves
Vegetation Index Functional Values - EVI Version
(EVILINE)
Versus
Days in July - September 1984
(DAY)

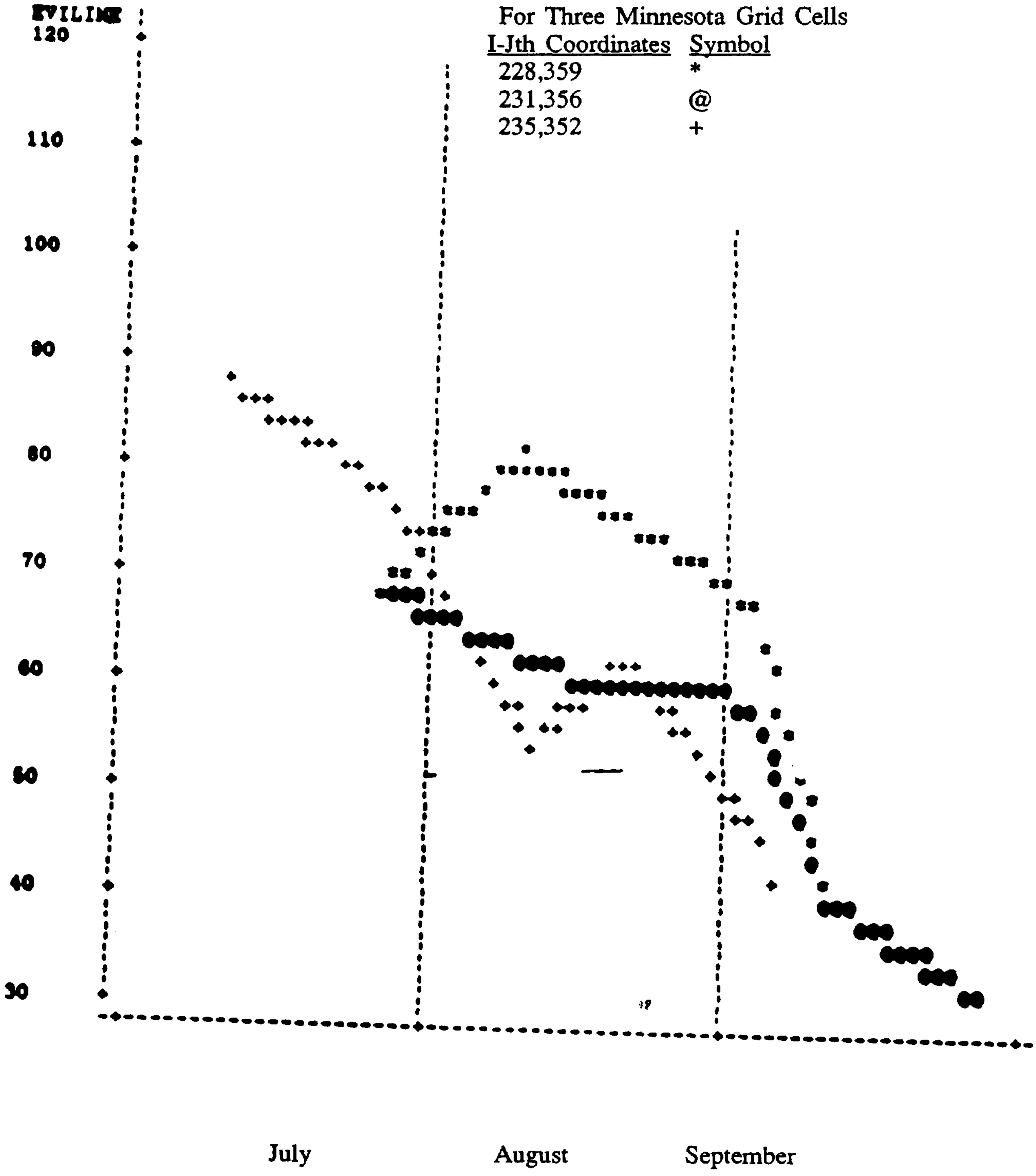


Figure B-4
Iowa Greenness Curves
Vegetation Index Functional Values - EVI Version
(EVILINE)
Versus
Days in July - September 1984
(DAY)

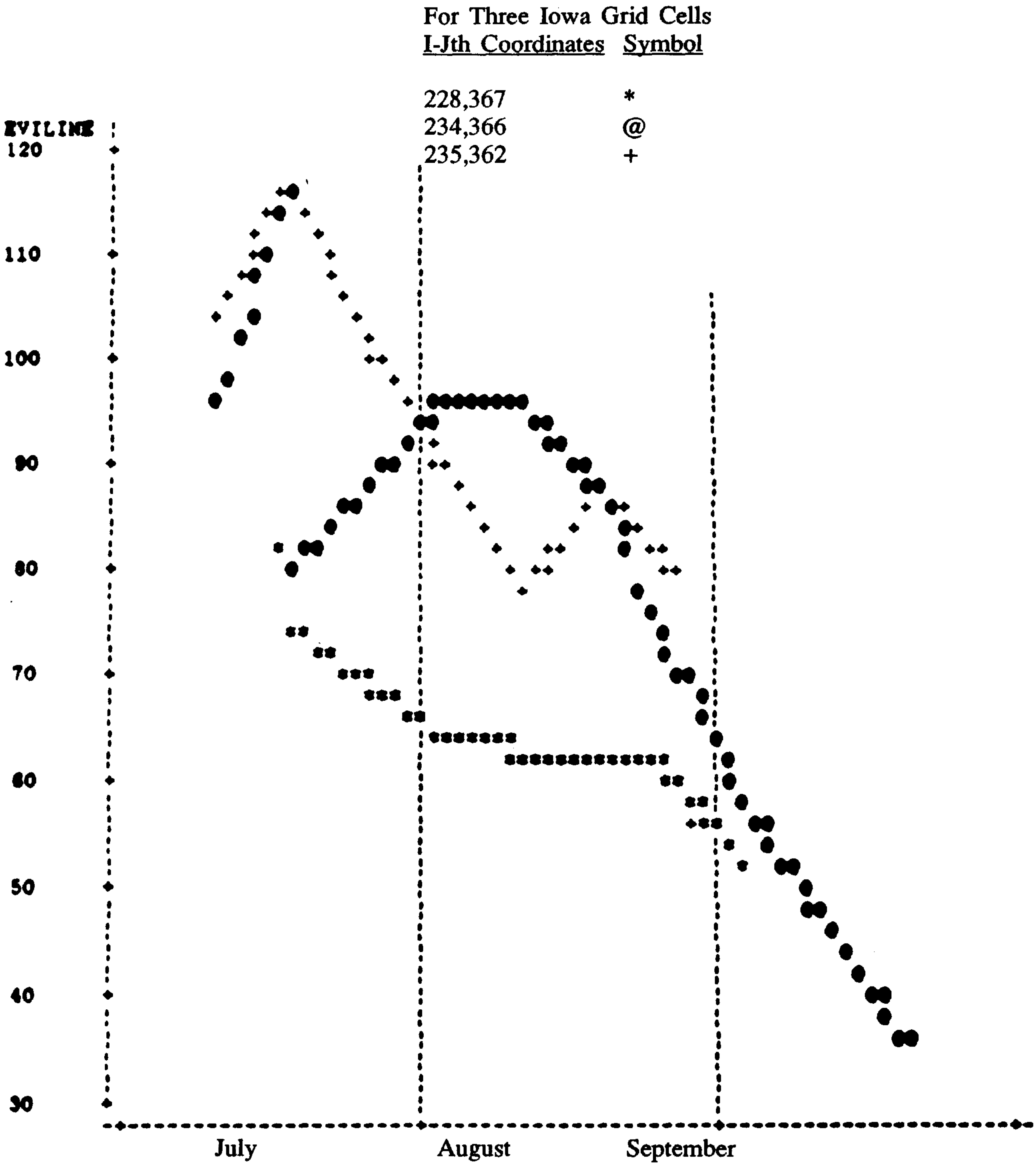


Figure B-5
Missouri Greenness Curves
Vegetation Index Functional Values - EVI Version
(EVILINE)
Versus
Days in July - September 1984
(DAY)

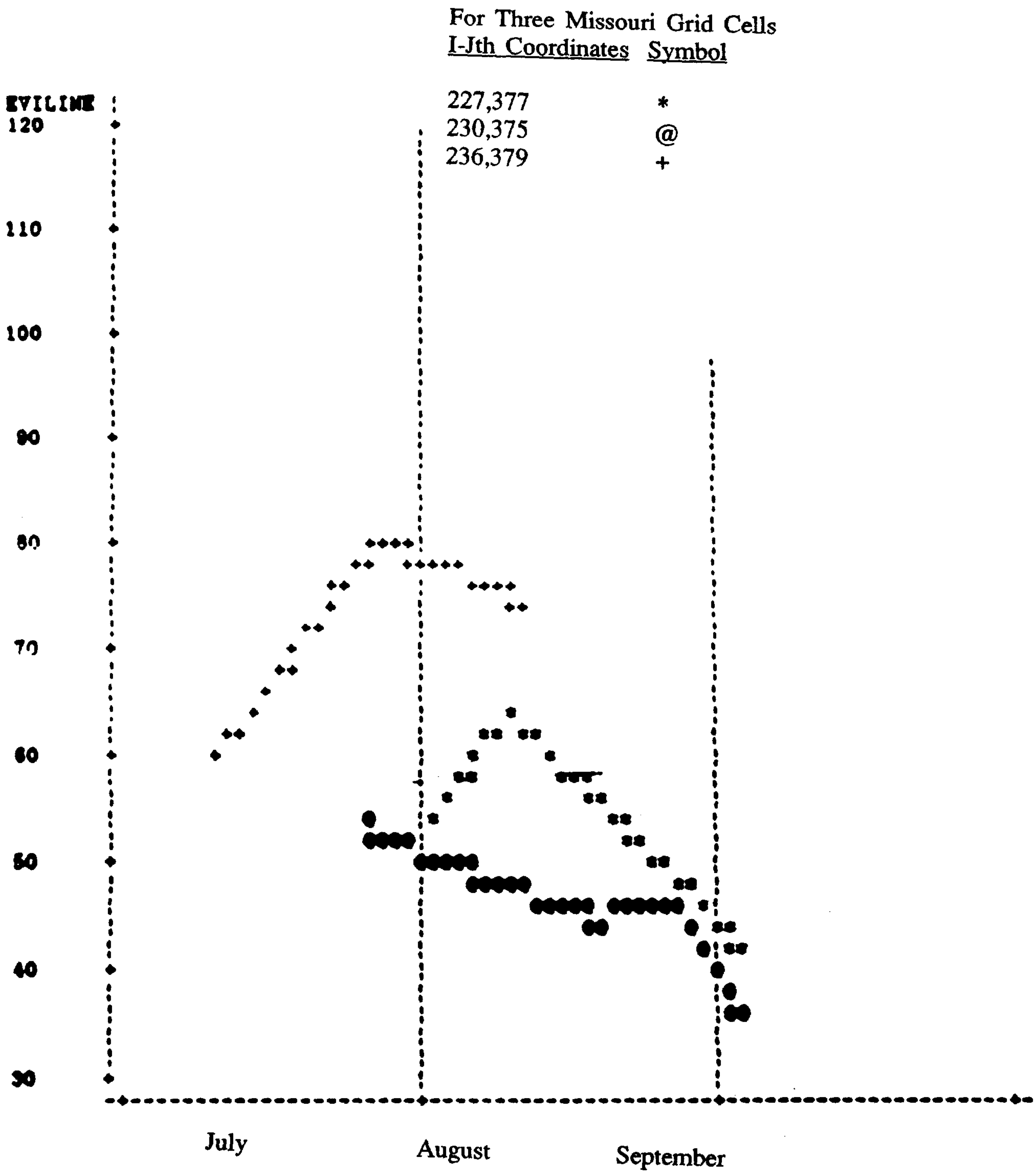


Figure B-6
Illinois Greenness Curves
Vegetation Index Functional Values - EVI Version
(EVILINE)
Versus
Days in July - September 1984
(DAY)

For Three Illinois Grid Cells

I-Jth Coordinates	Symbol
235,373	*
238,372	@
241,371	+

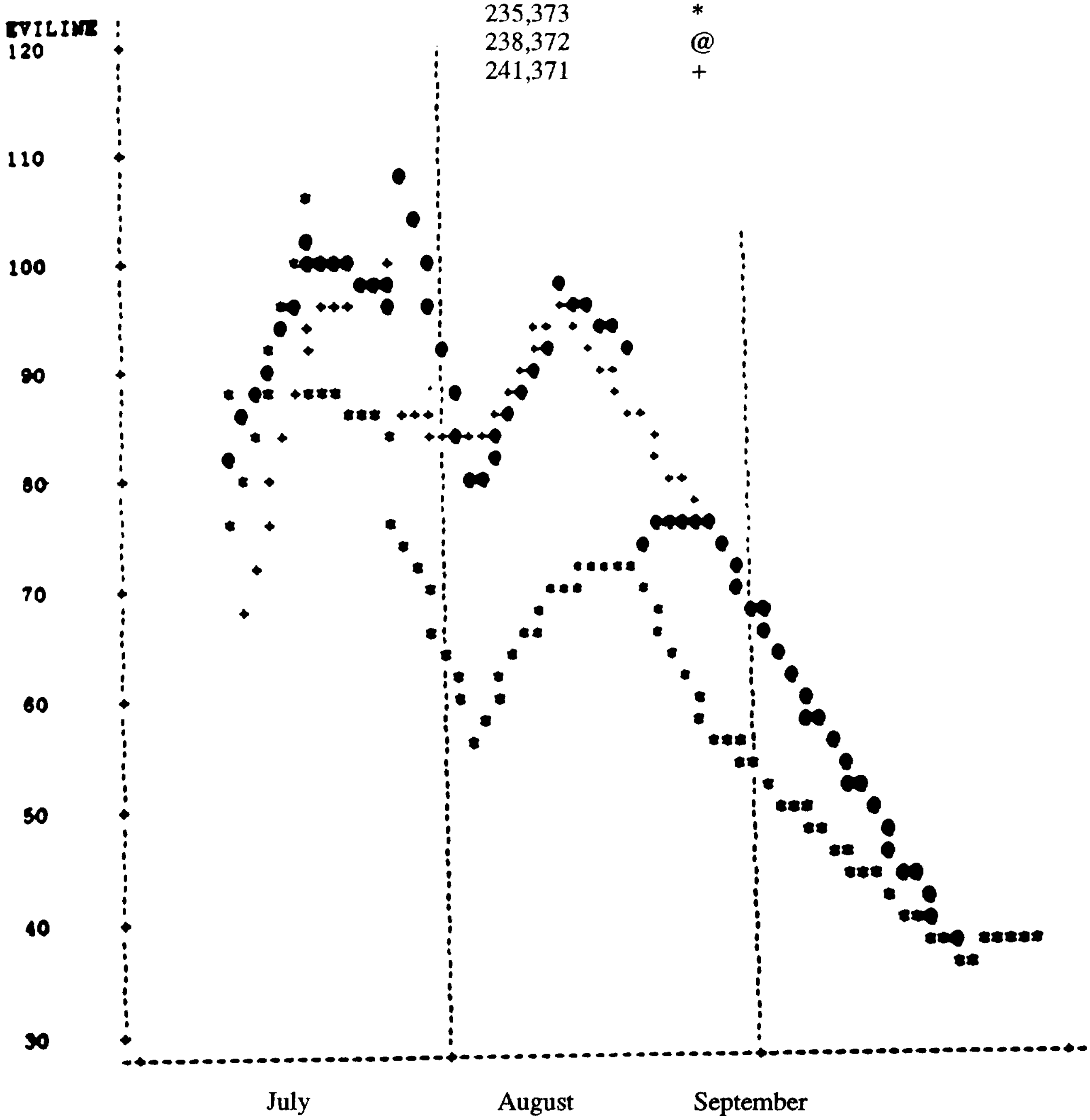


Figure B-7
Indiana Greenness Curves
Vegetation Index Functional Values - EVI Version
(EVILINE)
Versus
Days in July - September 1984
(DAY)

For Three Indiana Grid Cells

<u>I-Jth Coordinates</u>	<u>Symbol</u>
243,375	*
244,372	@
246,370	+

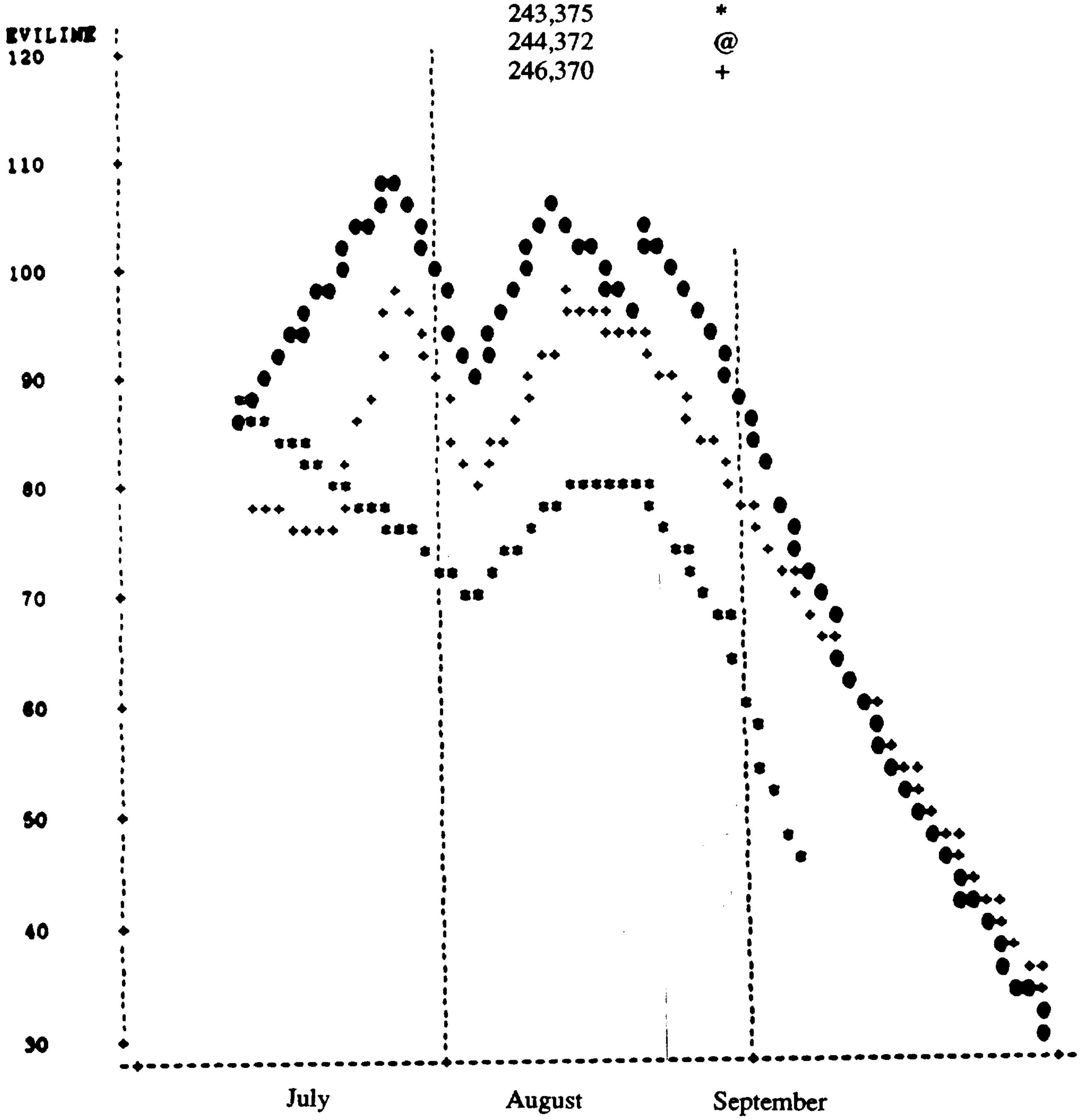


Figure B-8
Ohio Greenness Curves
Vegetation Index Functional Values - EVI Version
(EVILINE)
Versus
Days in July - September 1984
(DAY)

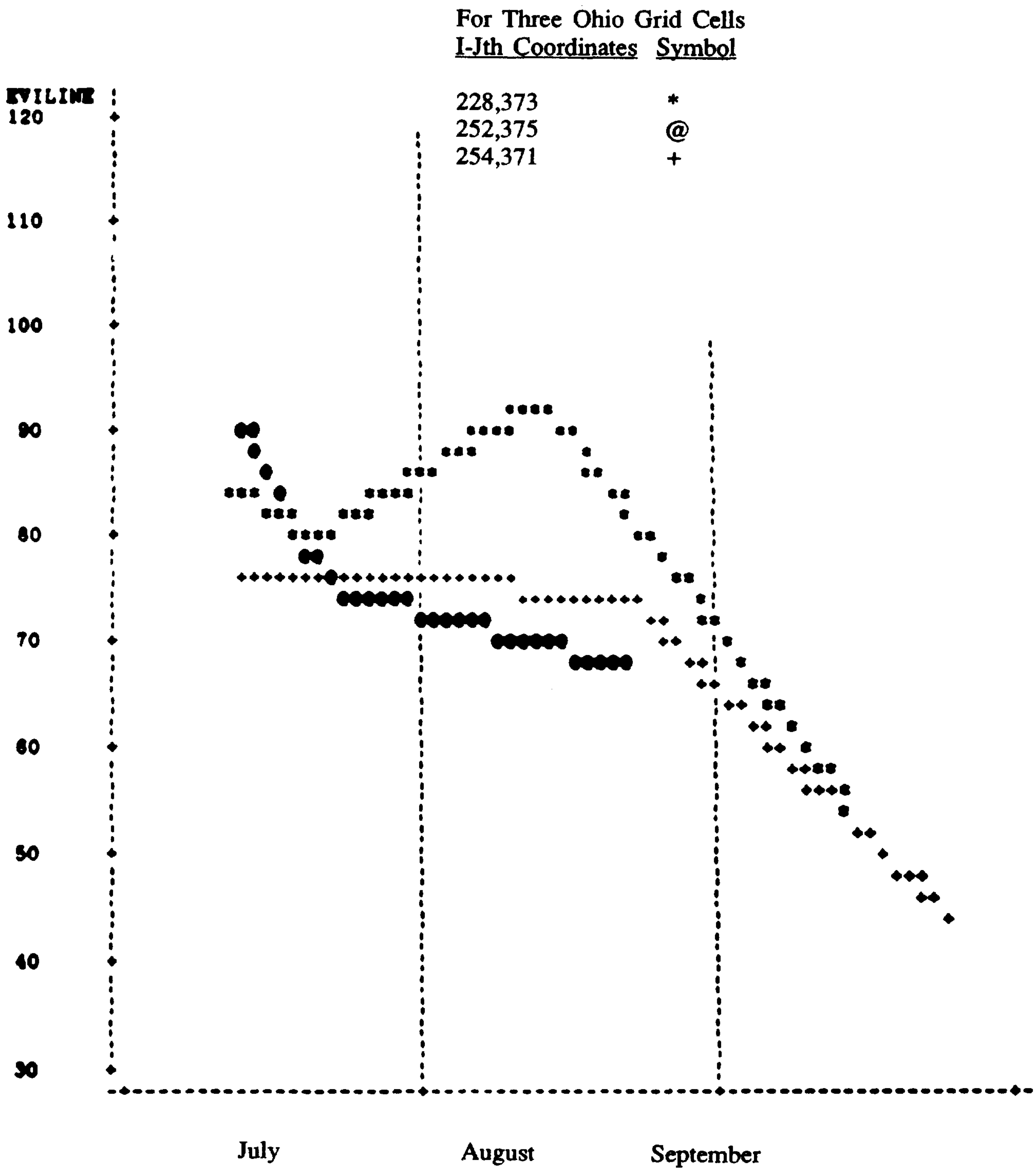


Figure B-9
Kentucky Greenness Curves
Vegetation Index Functional Values - EVI Version
(EVILINE)
Versus
Days in July - September 1984
(DAY)

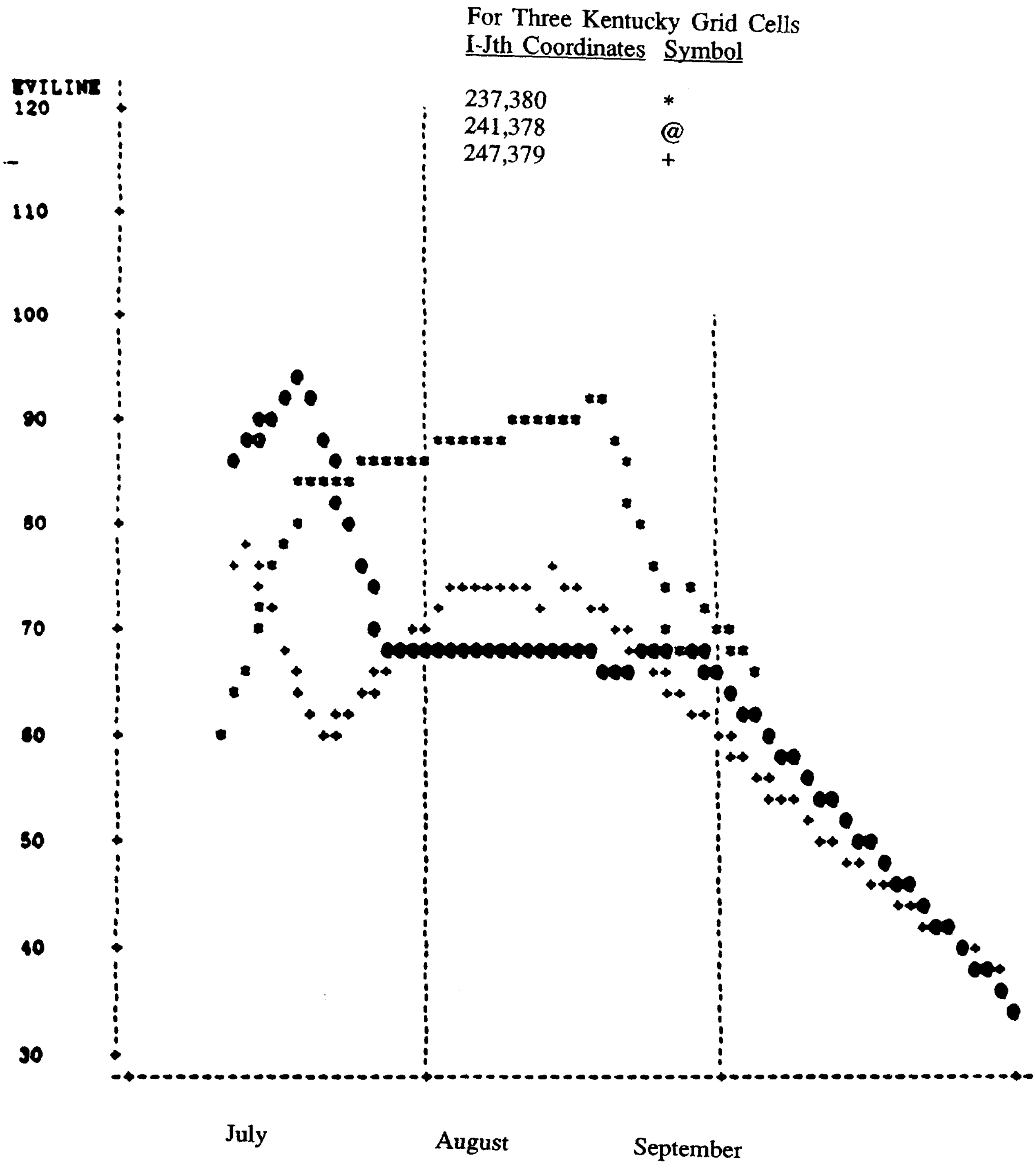


Figure B-10
Tennessee Greenness Curves
Vegetation Index Functional Values - EVI Version
(EVILINE)
Versus
Days in July - September 1984
(DAY)

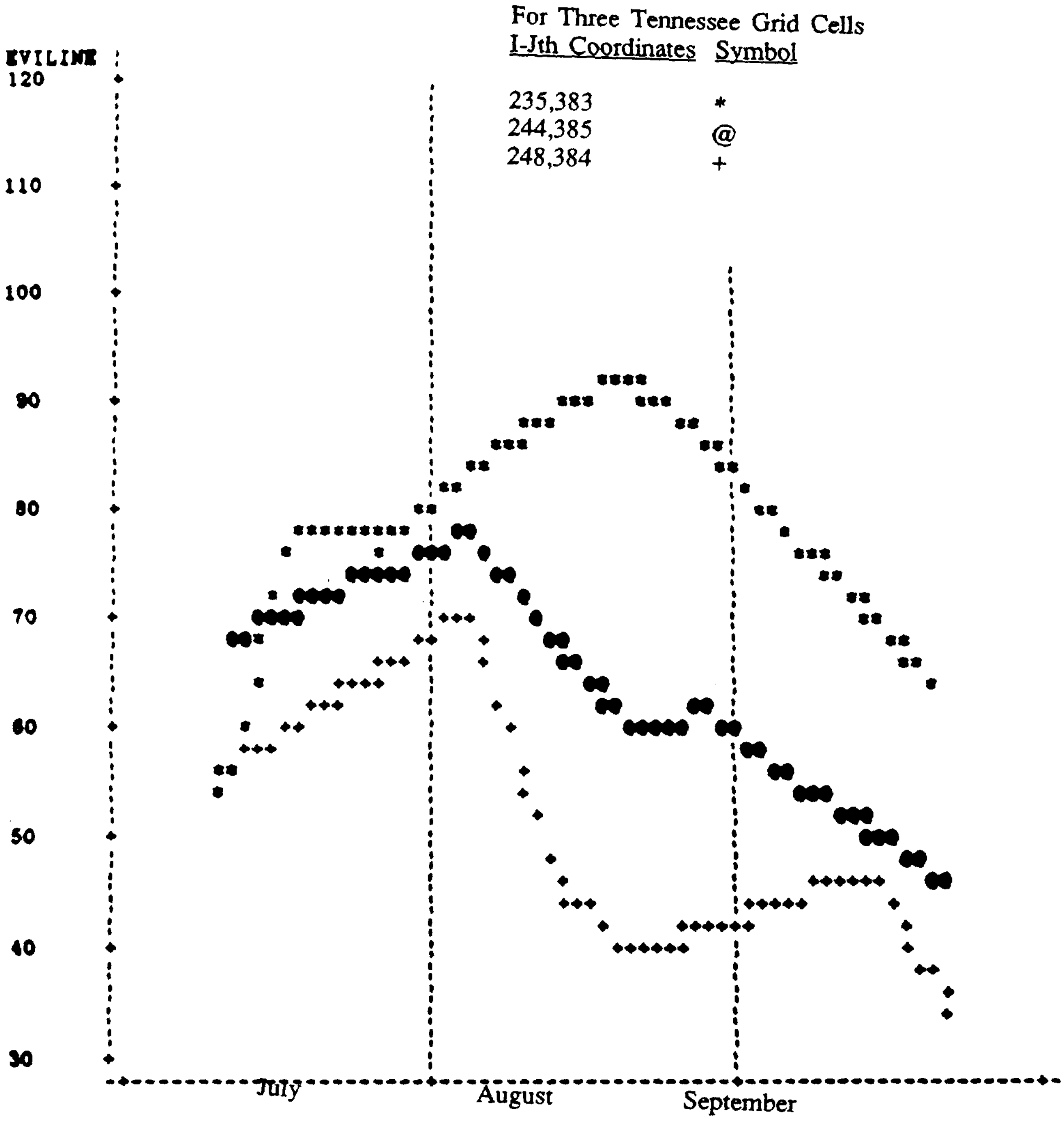


TABLE B-1. Coefficient of determination (R^2) between vegetative indexes (EVI version), averaged over the "FROM" through the "THRU" period, with the August 1 and October 1 yield forecasts and the final yield estimate, corn for grain and soybeans, ten state study area, 1984

FROM -> ->		7/8-15		7/16-23		7/24-30		7/31-8/7		8/8-14		8/15-23		8/24-9/1		9/2-9	
		Corn	Soybeans	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans	Corn	Soybeans
<u>THRU</u>																	
7/8-15 (A)	Aug.	.15	.11														
	Oct.	.14	.15														
	Final	.06	.08														
7/16-23 (B)	Aug.	.51	.41	.51	.46												
	Oct.	.43	.35	.20	.13												
	Final	.28	.24	.13	.06												
7/24-30 (C)	Aug.	.89	.78	.91	.79	.78	.65										
	Oct.	.79	.52	.79	.51	.88	.69										
	Final	.68	.48	.67	.47	.85	.71										
7/31-8/7 (D)	Aug.	.96	.90	.96	.84	.94	.79	.78	.76								
	Oct.	.89	.65	.89	.60	.92	.65	.85	.65								
	Final	.84	.62	.84	.59	.91	.66	.87	.65								
8/8-14 (E)	Aug.	.97	.93	.97	.91	.92	.81	.87	.80	.66	.67						
	Oct.	.94	.74	.94	.72	.94	.74	.96	.79	.79	.78						
	Final	.88	.71	.88	.70	.92	.74	.94	.77	.71	.70						
8/15-23 (F)	Aug.	.97	.93	.97	.92	.92	.82	.89	.83	.86	.76	.85	.76				
	Oct.	.95	.77	.94	.76	.94	.76	.96	.81	.94	.80	.88	.73				
	Final	.89	.74	.89	.74	.93	.77	.93	.79	.90	.77	.87	.72				
8/24-9/1 (G)	Aug.	.96	.92	.96	.91	.89	.79	.86	.76	.81	.68	.79	.68	.49	.28		
	Oct.	.95	.79	.94	.79	.92	.75	.94	.80	.92	.78	.87	.73	.61	.40		
	Final	.91	.76	.90	.76	.92	.76	.93	.80	.90	.78	.87	.74	.62	.38		
9/2-9 (H)	Aug.	.94	.86	.94	.85	.88	.74	.84	.72	.81	.65	.78	.63	.75	.42	.43	.15
	Oct.	.95	.74	.95	.74	.92	.69	.95	.74	.94	.74	.90	.72	.87	.51	.36	.13
	Final	.93	.74	.92	.74	.93	.70	.94	.76	.93	.76	.91	.73	.86	.48	.36	.10

Table B-1 shows the matrix of R^2 values for potential relationships of the August 1 and October 1 corn and soybean yield forecasts and the final yield estimates for the two crops to vegetative index means for 36 different periods. Relationship strength for the September 1 and November 1 yield forecasts was also investigated. Since the patterns for these additional forecasts were similar to those presented, they are omitted from the table. The coefficients of determination shown on the diagonal (for each month and crop) indicate which of the short periods have some relationship to the yield forecasts or estimates. Coverage can be quite incomplete for such short time intervals. For some of these periods whole Crop Reporting Districts were not represented. Moving just off the diagonal, one can see the results of combining two adjoining short periods. For example, when the July 16-23 and July 24-30 periods are combined the final yield estimate for corn attains and R^2 of .67 compared to .13 for the earlier period. This type of situation indicates that the July 16-23 period may have some marginal information on corn condition even though the R^2 for that period alone is quite small. On the other hand, final corn R^2's are already fairly high for the August 8-14 and August 15-23 individual periods (.71 and .87, respectively). They reach .90 for the combined 16 day August 8-23 period.

Many patterns can be observed from the table. Most of them have rational explanations. Some of these patterns and explanations are:

Pattern -	Earlier periods have stronger relationships to the August 1 forecasts.
Explanation -	The forecasts were made based on survey data and knowledge of conditions around August 1 and would be more likely to differ from conditions longer after that date.
Pattern -	Soybean R^2's are often lower than those for corn.
Explanation -	The critical period is longer for soybeans than it is for corn (particularly over this study area) so that general vegetative conditions in any period are not as indicative of the soybean yield determining environment.
Pattern -	Longer periods that are composed of selected individual periods (those with some individual relationship to crop yields, which also form adjoining periods with fairly strong relationships) have similar strength relationships.
Explanation -	Means change very little as data is added or deleted as long as the means come from fairly long periods with enough observations, and as long as the periods added or deleted have strong and similar relationships.

Figures B-11 and B-12 illustrate the selection of periods based on the strength of relationships between the vegetative index and the final corn for grain yield estimate. Figure B-11 shows the corn R^2's for individual short periods and for all adjoining two period combinations. The periods are labeled A through H from the first period, 7/8-15 (A), through the last period, 9/2-9 (H). Thus, period D-E denotes periods D (the Fourth one) and E (the fifth one) taken together or July 31 through August 14. From this figure it can be seen that individually periods C through G show some explanatory power for final corn yield, although vegetative indexes from periods E and G are not as strongly related. By examining the two period combination R^2's, it can be concluded that periods C-D through G-H exhibit strong relationships. This implies that data from periods C through H have a strong relationship to corn yield. However, because individually the H period vegetative index mean had such a weak relationship, only C through G (July 24-September 1) will be given further consideration. All the three through eight period combinations that start and end within the C through G interval (thus, five periods is the maximum in this case) are shown in figure B-12. Three periods (D-F, C-F, and D-G) have an R^2 of .93 for the final corn yield to mean vegetative index relationship. Two others (C-G and C-E) are at .92.

The same type of analysis is shown for soybeans in figures B-13 and B-14. Individual short periods with some explanatory information on soybean final yield appear to be C, D, E, F and possibly G. The two period combinations confirm the value of vegetative index information from the first four of these periods (C-F) and suggest that period G (August 24-September 1) may help explain State level variability in soybean yields. Periods A, B and H individually showed no relationship to soybean final yield estimates and even when combined with some more strongly related individual periods (C and G) failed to attain very large R^2's. All of the three period and up combinations which start and end within C through G (July 24-September 1) are shown in Figure B-14.

Figure B-11
Coefficient of Determination
of the
Corn for Grain Final Yield Estimate
with the
Mean Vegetative Index - EVI Version (CORNRSQ)
Over Individual Short Periods
and Adjoining Two Period Combinations

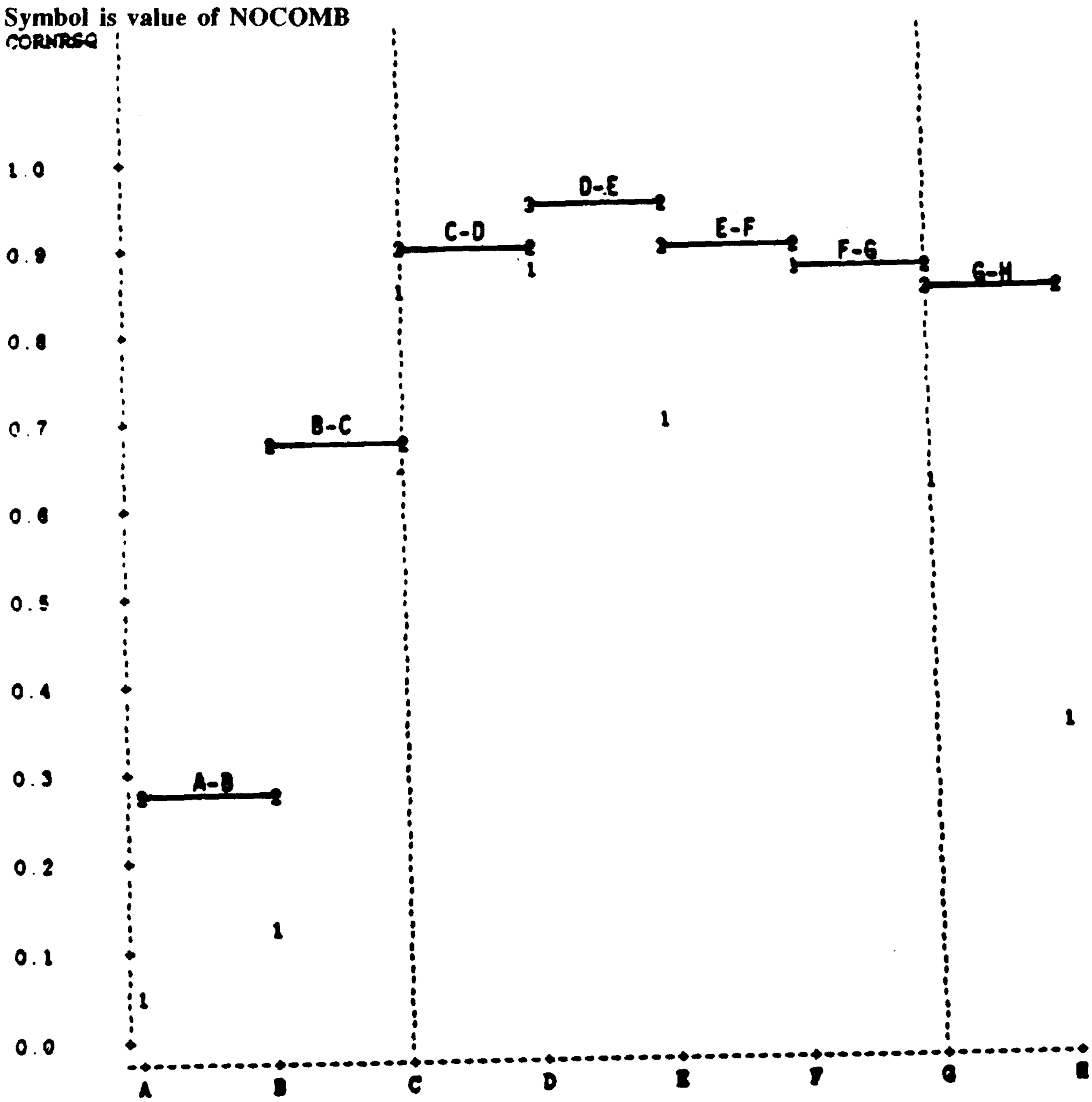


Figure B-12
Coefficient of Determination
of the
Corn for Grain Final Yield Estimate
with the
Mean Vegetative Index - EVI Version (CORNRSQ)
Over Adjoining Periods of Three or More
Short Periods Within the Restricted Internal (C-G)

Symbol is value of NOCOMB

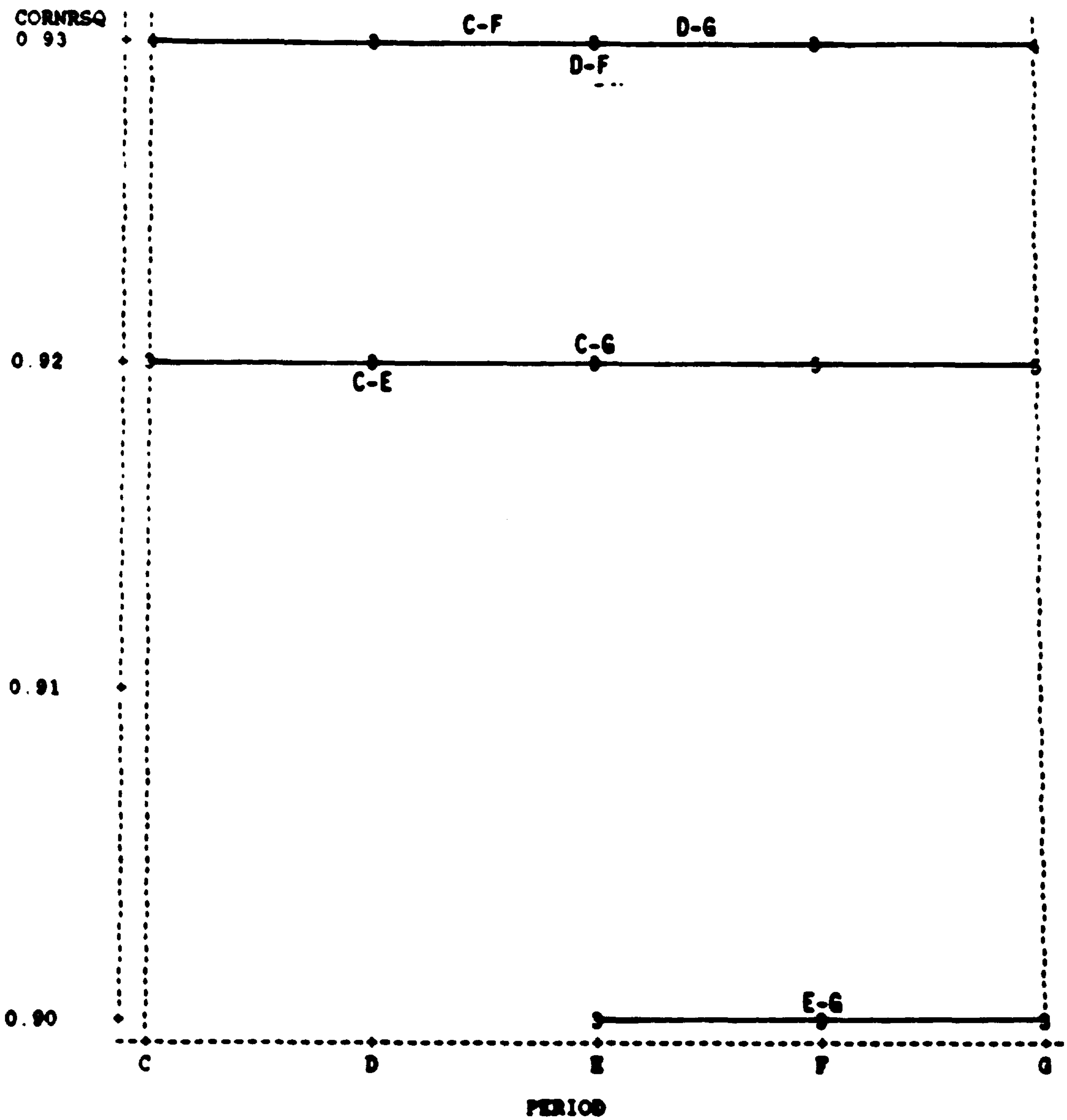


Figure B-13
Coefficient of Determination
of the
Soybean Final Yield Estimate
with the
Mean Vegetative Index - EVI Version (SOYBRSQ)
Over Individual Short Periods
and Adjoining Two Period Combinations

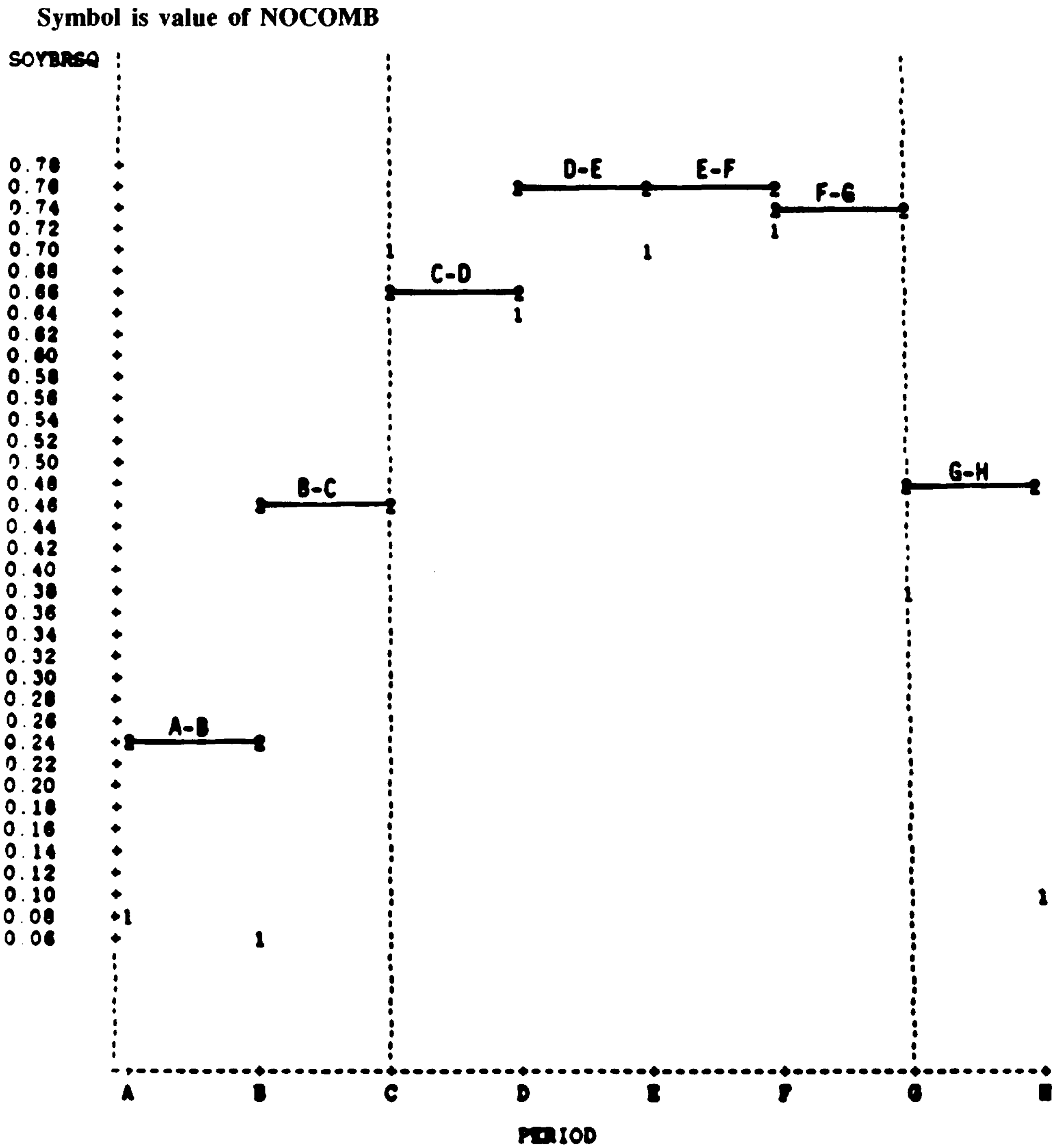
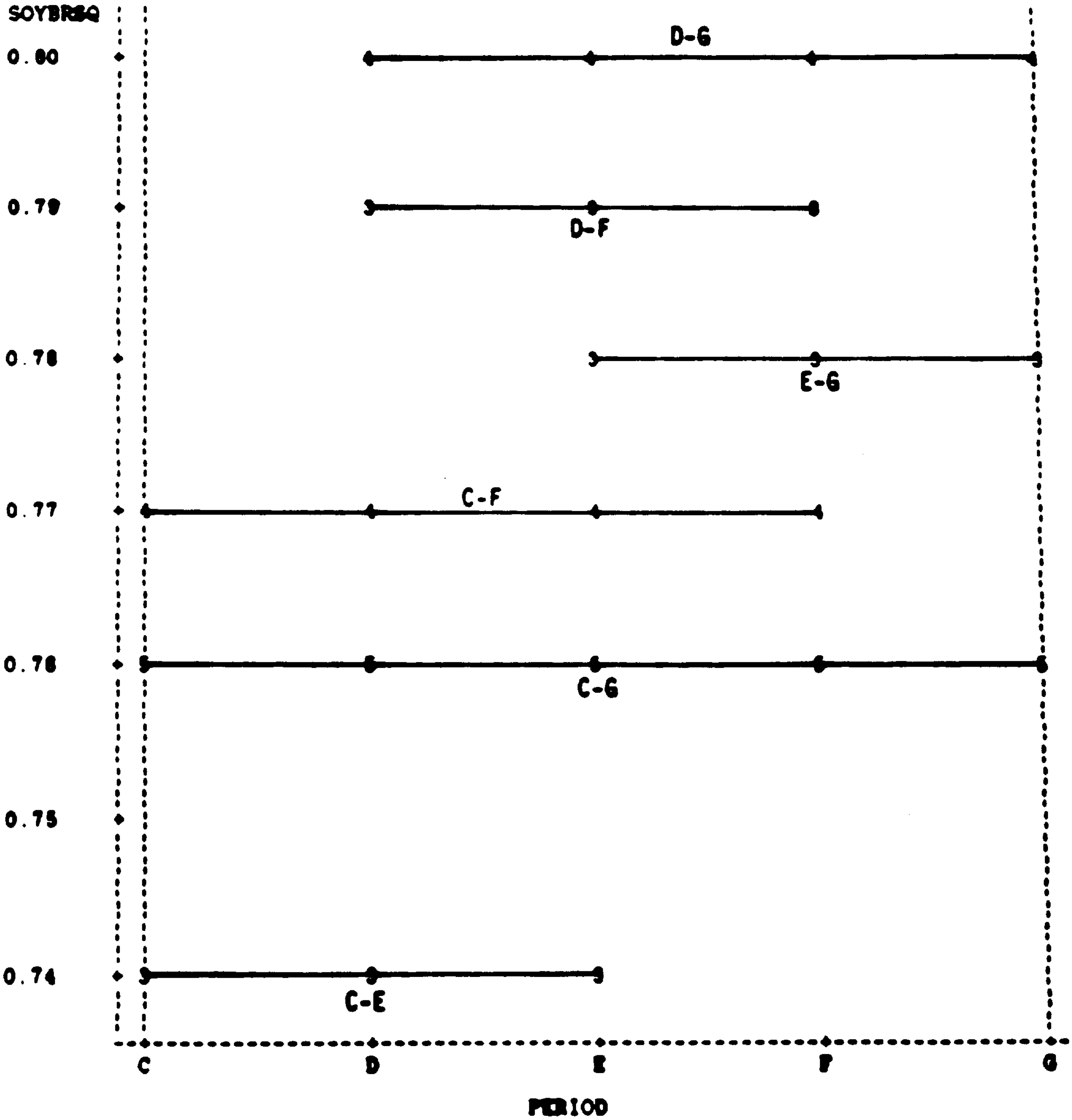


Figure B-14
Coefficient of Determination
of the
Soybean Final Yield Estimate
with the
Mean Vegetative Index - EVI Version (SOYBRSQ)
Over Adjoining Periods of Three or More
Short Periods Within the Restricted Internal (C-G)

Symbol is value of NOCOMB



Four of the periods with the strongest relationships to the final yield estimate for both the corn and soybean crops were subsequently evaluated by deriving the state mean vegetative index in the manner shown in Figure 1. The mean grid cell indexes for each of these periods were mapped to counties and then aggregated by previous year acreage county weights to the state level. This produced higher R^2's for some periods than the previous analysis and probably reflects the more appropriate and detailed weighting of where the crops are within states and within Crop Reporting Districts. The comparison results are shown in Table B-2.

TABLE B-2. Comparison of coefficients of determination (R^2's) between crop yield estimates and mean vegetative indexes when state mean vegetative indexes are weighted via Crop Reporting Districts (CRD's) or via Counties, by selected periods, 1984, Ten State Study Area.

	Corn - R^2's		Soybean - R^2's	
Period	CRD	County	CRD	County
D-F (7/31-8/2)	.93	.941	.79	.775
D-G (7/31-9/1)	.93	.926	.80	.847
C-F (7/24-8/23)	.93	.908	.77	.770
C-G (7/24-9/1)	.92	.893	.76	.806

It appears to be fairly important to include available vegetative index values from the nine day period of August 24 through September 1 (period G) for soybeans, but to exclude them in developing the corn for grain final estimate. This may reflect the fact that the yield determining period generally ends earlier of corn than soybeans, so that general conditions some time after the critical corn period ends provides less information on the environment for that crop. Table B-2 also suggests that it is best to exclude period C (July 24-30) for both crop's final yield. This may be related to some of the earlier "greenness" (as seen in Figures B-1 through B-10) still present during the later part of July. This "greenness" may not be directly associated with environmental conditions effecting corn and soybean yields as expressed by general vegetative conditions somewhat later than when the critical periods for the crops began.

Again, the critical periods selected for relating final yield estimates to mean vegetative indexes at the state level are July 31 - August 23 for corn and July 31 - September 1, for soybeans. Weighting via Crop Reporting Districts, as discussed in this Appendix, is preferable for shorter periods when county coverage would be very incomplete. However, a somewhat larger set of the longer periods (than reported here) might be effectively evaluated when state vegetative index means are derived via county mapping and weights.

APPENDIX C

STATISTICAL ANALYSIS SYSTEM REGRESSION OUTPUT FOR CORN AND SOYBEAN MODELS BASED ON EVI AND NVI VEGETATIVE INDEX VERSIONS AND YIELD VERSUS NVI PLOTS

EXHIBIT C-1
STATISTICAL ANALYSIS SYSTEM OUTPUT
Regression of Final Corn for Grain Yield (CYDFN)
on the
EVI Version of the Corn for Grain
Vegetative Index (EQMECI9)

SAS

Model: MODEL1
Dep Variable: CYDFN

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	3401.14850	3401.14850	127.592	0.0001
Error	8	213.25150	26.65644		
C Total	9	3614.40000			

Root MSE	5.16299	R-Square	0.9410
Dep Mean	97.60000	Adj R-Sq	0.9336
C.V.	5.28995		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-16.250967	10.21055398	-1.592	0.1501
EQMECI9	1	1.600619	0.14170205	11.296	0.0001

Obs	CYDFN	Predict Value
1	112.0	118.9
2	114.0	114.9
3	117.0	119.2
4	100.0	93.4393
5	107.0	103.8
6	80.0000	85.4765
7	66.0000	66.8332
8	118.0	113.5
9	67.0000	71.1318
10	95.0	88.7733

Sum of Squared Residuals	213.2515
Predicted Resid SS (Press)	315.3164

Exhibit C-2
STATISTICAL ANALYSIS SYSTEM OUTPUT
Regression of Final Soybean Yield (SYDFN)
on the
EVI Version of the Soybean
Vegetative Index (EQMESI0)

SAS

Model: MODEL1
Dep Variable: SYDFN

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	229.53774	229.53774	44.398	0.0002
Error	8	41.36226	5.17028		
C Total	9	270.90000			

Root MSE	2.27383	R-Square	0.8473
Dep Mean	28.90000	Adj R-Sq	0.8282
C.V	7.86791		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-9.049912	5.74082661	-1.576	0.1536
EQMESI0	1	0.528877	0.07937515	6.663	0.0002

Obs	SYDFN	Predict Value
1	31.5000	34.3013
2	32.0000	31.8967
3	34.5000	34.8049
4	29.0000	26.6335
5	33.0000	30.3580
6	20.5000	24.2815
7	23.0000	21.2933
8	36.5000	34.9618
9	23.0000	23.4398
10	26.0000	27.0290

Sum of Squared Residuals	41.3623
Predicted Resid SS (Press)	64.2779

Figure C-1
CORN FOR GRAIN - STATE LEVEL
Official Yield Estimate (bushels per acre)
Versus
Corn Vegetative Index
(NVI Version)

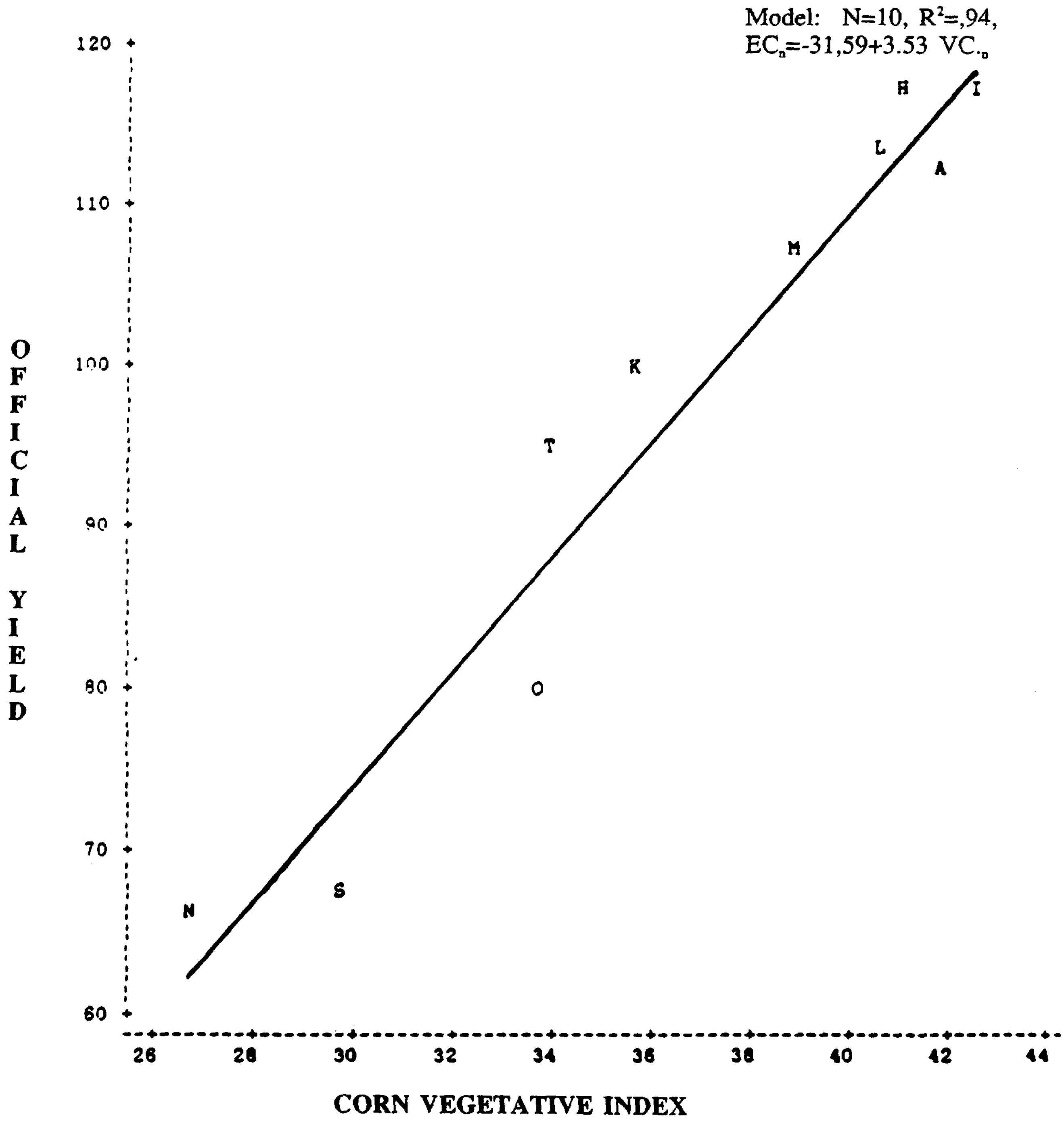


Figure C-2
SOYBEANS - STATE LEVEL
Official Final Yield Estimate (bushels per acre)
Versus
Soybean Vegetative Index
(NVI Version)

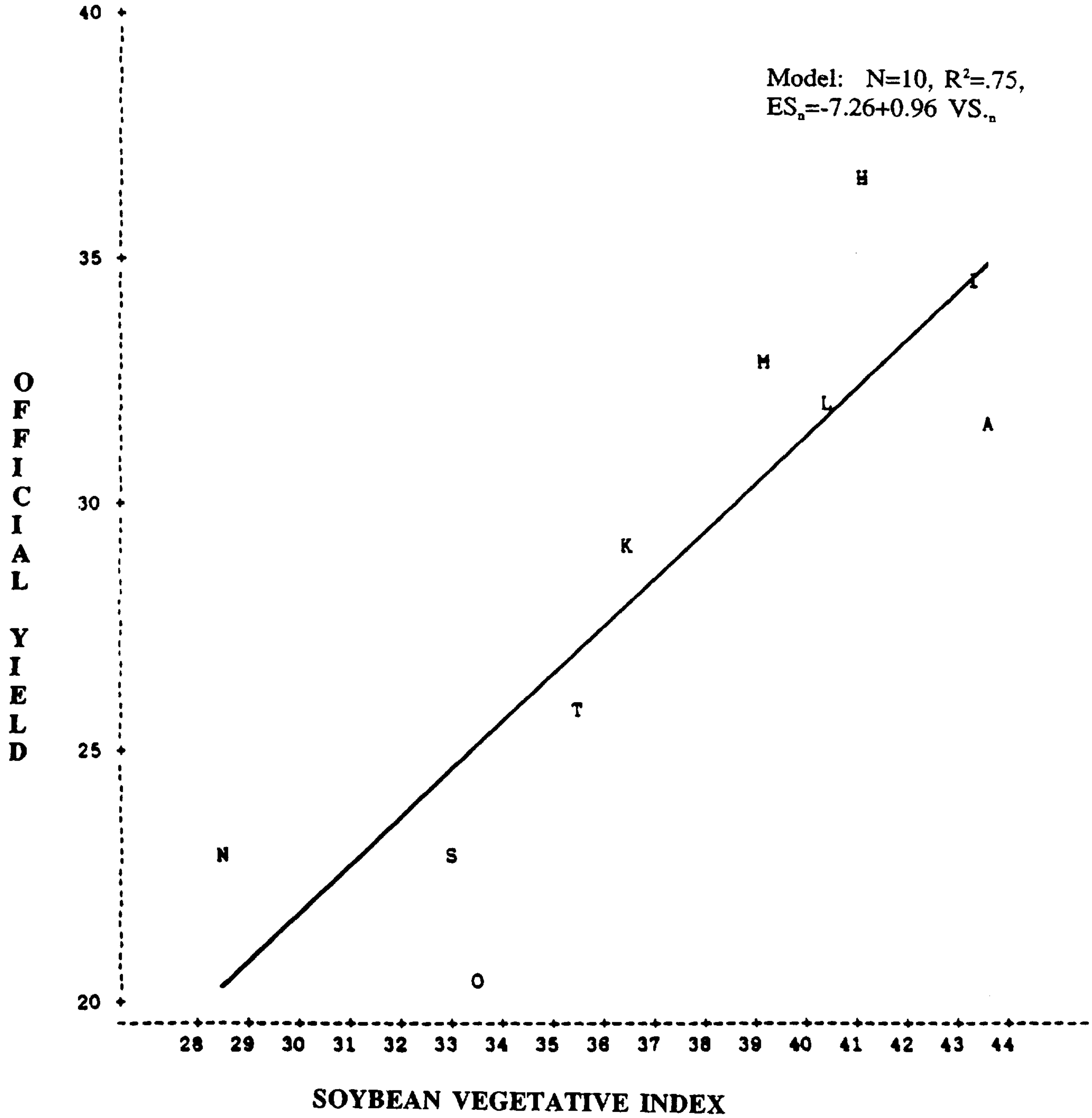


Exhibit C-3
SAS OUTPUT
Regression of Final Corn for Grain Yield (CYDFN) on the NVI Version of the Corn for Grain Vegetative Index (EQMNCI9)

SAS

Model: MODEL1
Dep Variable: CYDFN

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	3391.49534	3391.49534	121.720	0.0001
Error	8	222.90466	27.86308		
C Total	9	3614.40000			

Root MSE	5.27855	R-Square	0.9383
Dep Mean	97.60000	Adj R-Sq	0.9306
C.V.	5.40835		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-31.585946	11.82776820	-2.670	0.0283
EQMNCI9	1	3.534207	0.32033987	11.033	0.0001

Obs	CYDFN	Predict Value
1	112.0	117.3
2	114.0	112.2
3	117.0	119.1
4	100.0	95.1
5	107.0	106.1
6	80.0000	87.4483
7	66.0000	62.7135
8	118.0	114.0
9	67.0000	73.4120
10	95.0	88.6983

Sum of Squared Residuals	222.9047
Predicted Resid SS (Press)	350.6414

Exhibit C-4
SAS OUTPUT
Regression of Final Soybean Yield (SYDFN)
on the
NVI Version of the Soybean
Vegetative Index (EQMNSI0)

SAS

Model: MODEL1
Dep Variable: SYDFN

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	204.19806	204.19806	24.491	0.0011
Error	8	66.70194	8.33774		
C Total	9	270.90000			

Root MSE	2.88751	R-Square	0.7538
Dep Mean	28.90000	Adj R-Sq	0.7230
C.V.	9.99140		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-7.256920	7.36300969	-0.986	0.3532
EQMNSI0	1	0.963342	0.19466091	4.949	0.0011

Obs	SYDFN	Predict Value
1	31.5000	34.8035
2	32.0000	31.8288
3	34.5000	34.6043
4	29.0000	27.8062
5	33.0000	30.6260
6	20.5000	25.0374
7	23.0000	20.2372
8	36.5000	32.4040
9	23.0000	24.6185
10	26.0000	27.0342

Sum of Squared Residuals	66.7019
Predicted Resid SS (Press)	115.7222

APPENDIX D

STUDY AREA AND INDIVIDUAL STATE PERFORMANCE TABLES FOR RESULTS AT VARIOUS LEVELS FOR CORN FOR GRAIN AND SOYBEANS, AND AT THE COUNTY LEVEL FOR CORN WHEN SOME COUNTIES ARE EXCLUDED BY AN OBJECTIVE RULE OR BY DELETION OF OBVIOUS OUTLIERS

TABLE D-1. Performance measures at the State, district and county levels for satellite generated corn for grain yield estimate indications obtained by considering official estimates as "truth", 1984, ten State study area and individual States.

APPLICATION AREA	N	R^2	R	MSE (bushels/acre)[2/]	VAR	BIAS[1/] -----bushels/acre-----	RMSE	ST.DEV.	RSD[2/] %
				STATE LEVEL					
TEN STATES	10	.94	.97	26.66	26.66	0.00	5.16	5.16	5.3
				DISTRICT LEVEL					
TEN STATES	84	.75	.87	181.40	178.76	-1.63	13.47	13.37	13.7
NORTH DAKOTA	9	.84	.92	34.42	34.03	-0.62	5.87	5.83	8.8
SOUTH DAKOTA	9	.03	-.20	934.08	638.17	-17.20	30.56	25.26	37.7
MINNESOTA	9	.80	.90	181.70	158.94	4.77	13.48	12.61	11.8
IOWA	9	.44	.66	75.89	75.52	0.61	8.71	8.69	7.8
MISSOURI	9	.56	.75	130.81	118.67	-3.48	11.44	10.89	13.6
ILLINOIS	9	.76	.87	38.68	38.05	0.80	6.22	6.17	5.4
INDIANA	9	.64	.80	31.58	31.28	-0.54	5.62	5.59	4.8
OHIO	9	.75	.83	86.32	67.81	-4.30	9.29	8.23	7.0
KENTUCKY	6	.30	.55	57.39	37.96	4.41	7.58	6.16	6.2
TENNESSEE	6	.72	.85	211.97	204.18	2.79	14.56	14.29	15.0
				COUNTY LEVEL					
TEN STATES	889	.63	.80	249.31	248.69	-0.79	15.79	15.77	16.2
NORTH DAKOTA	47	.29	.54	217.80	215.51	-1.51	14.76	14.68	22.2
SOUTH DAKOTA	62	.04	.18	683.58	594.04	-9.46	26.15	24.37	36.4
MINNESOTA	81	.56	.75	261.26	260.86	0.63	16.16	16.15	15.1
IOWA	99	.27	.52	178.70	177.14	1.25	13.37	13.31	11.9
MISSOURI	114	.30	.54	292.16	283.18	-3.00	17.09	16.83	21.0
ILLINOIS	102	.38	.62	204.80	203.86	0.97	14.31	14.28	12.5
INDIANA	92	.46	.67	97.28	96.43	-0.92	9.86	9.82	8.4
OHIO	86	.42	.65	160.92	151.35	-3.09	12.69	12.30	10.4
KENTUCKY	113	.15	.39	177.06	145.54	5.61	13.31	12.06	12.1
TENNESSEE	93	.09	.29	356.70	348.96	-2.78	18.89	18.68	19.7

1/ The bias at the district and county level would be very close to zero for a harvested acreage weighted mean. However, all counties (districts) were given equal weight in this analysis.

2/ RSD is the standard deviation relative to the mean (equal weights) corn for grain yield (97.6 BU./A) for the ten States and to the final yield estimate for individual States.

TABLE D-2. Performance measures at the State, county, and district levels for satellite generated soybean yield estimate indications obtained by considering official estimates as "truth", 1984, ten State study area and individual States.

APPLICATION AREA	N	R^2	R	MSE	VAR	BIAS[1]	RMSE	ST.DEV.	RSD[2]
				(bushels/acre)^2		-----bushels/acre-----			%
				STATE LEVEL					
TEN STATES	10	.85	.92	5.17	5.17	0.00	2.27	2.27	5.3
				DISTRICT LEVEL					
TEN STATES	76	.74	.86	15.93	14.98	-0.97	3.99	3.87	13.4
NORTH DAKOTA	6	.78	.88	9.23	9.10	0.36	3.04	3.02	13.1
SOUTH DAKOTA	7	.01	.08	89.74	41.29	-6.96	9.47	6.43	28.0
MINNESOTA	7	.81	.90	14.38	12.97	-1.19	3.79	3.60	10.9
IOWA	9	.77	.88	5.79	5.79	0.09	2.41	2.41	7.7
MISSOURI	9	.52	.72	14.30	14.27	-0.17	3.78	3.78	18.4
ILLINOIS	9	.61	.78	6.98	6.90	0.28	2.64	2.63	8.2
INDIANA	9	.73	.86	3.56	3.46	-0.32	1.89	1.86	5.4
OHIO	8	.71	.85	6.83	4.88	-1.40	2.61	2.21	6.1
KENTUCKY	6	.06	.24	6.13	6.12	-0.09	2.48	2.47	8.5
TENNESSEE	6	.43	.65	9.85	8.71	-1.07	3.14	2.95	11.3
				COUNTY LEVEL					
TEN STATES	756	.64	.80	22.07	21.02	-1.03	4.70	4.58	15.8
NORTH DAKOTA	28	.59	.77	10.49	10.08	0.64	3.24	3.18	13.8
SOUTH DAKOTA	36	.00	.04	50.47	36.47	-3.74	7.10	6.04	26.3
MINNESOTA	76	.56	.75	25.15	23.16	-1.41	5.01	4.81	14.6
IOWA	99	.36	.60	16.23	16.16	-0.26	4.03	4.02	12.8
MISSOURI	95	.23	.48	29.37	28.76	-0.78	5.42	5.36	26.1
ILLINOIS	102	.35	.59	18.94	18.89	-0.21	4.35	4.35	13.6
INDIANA	92	.57	.75	10.46	10.05	-0.64	3.23	3.17	9.2
OHIO	68	.34	.58	16.02	13.64	-1.54	4.00	3.69	10.1
KENTUCKY	81	.11	.33	13.31	13.29	-0.15	3.65	3.65	12.6
TENNESSEE	79	.02	.13	40.57	30.11	-3.23	6.37	5.49	21.1

1/ The bias at the district and county level would be very close to zero for a harvested acreage weighted mean. However, all counties (districts) were given equal weight in this analysis.

2/ RSD is the standard deviation relative to the mean (equal weights) soybean yield (28.9 BU./A) for the ten States and to the final yield estimate for individual States.

TABLE D-3. Performance measures at the county level for satellite generated corn for grain yield estimate indications obtained by considering official estimates as "truth" when data for some counties are excluded based on two different criteria, 1984, ten State area, and individual States.

APPLICATION AREA	N	R^2	R	MSE	VAR	BIAS	RMSE	ST.DEV.	RSD[1/]
				(bushels/acre)^2		-----bushels/acre-----			%
				COUNTY LEVEL 31 Counties Excluded by Irrigation Rule					
TEN STATES	858	.69	.83	200.74	200.66	0.28	14.17	14.17	14.5
NORTH DAKOTA	40	.69	.83	66.95	63.01	1.98	8.18	7.94	12.0
SOUTH DAKOTA	49	.78	.88	76.13	73.81	1.52	8.73	8.59	12.8
MINNESOTA	78	.59	.77	250.90	248.91	1.41	15.84	15.78	14.7
IOWA	99	.27	.52	178.70	177.14	1.25	13.37	13.31	11.9
MISSOURI	107	.19	.43	297.05	290.28	-2.60	17.24	17.01	21.3
ILLINOIS	101	.38	.62	206.08	204.96	1.06	14.36	14.32	12.6
INDIANA	92	.46	.67	97.28	96.43	-0.92	9.86	9.82	8.4
OHIO	86	.42	.65	160.92	151.35	-3.09	12.69	12.30	10.4
KENTUCKY	113	.15	.39	177.06	145.54	5.61	13.31	12.06	12.1
TENNESSEE	93	.09	.29	356.70	348.96	-2.78	18.89	18.68	19.7
				COUNTY LEVEL 11 Obvious Outlier Counties Excluded					
TEN STATES	878	.68	.83	205.46	205.46	-0.05	14.33	14.33	14.7
NORTH DAKOTA	46	.50	.70	123.00	123.00	-0.08	11.09	11.09	16.8
SOUTH DAKOTA	52	.67	.82	109.93	109.93	0.06	10.48	10.48	15.6
MINNESOTA	81	.56	.75	261.26	260.86	0.63	16.16	16.15	15.1
IOWA	99	.27	.52	178.70	177.14	1.25	13.37	13.31	11.9
MISSOURI	114	.30	.54	292.16	283.18	-3.00	17.09	16.83	21.0
ILLINOIS	102	.38	.62	204.80	203.86	0.97	14.31	14.28	12.5
INDIANA	92	.46	.67	97.28	96.43	-0.92	9.86	9.82	8.4
OHIO	86	.42	.65	160.92	151.35	-3.09	12.69	12.30	10.4
KENTUCKY	113	.15	.39	177.06	145.54	5.61	13.31	12.06	12.1
TENNESSEE	93	.09	.29	356.70	348.96	-2.78	18.89	18.68	19.7

1/ RSD is the standard deviation relative to the mean (equal weights) corn for grain yield (97.6 BU./A) for the ten States and to the final yield estimate for individual States (with all counties with the crop included therein).

☆U.S. GOVERNMENT PRINTING OFFICE:1989-241-859:00279/NASS