X-923-76-93

PREPRINT

Tmx 71288

AN ESTIMATE OF FIELD SIZE DISTRIBUTIONS FOR SELECTED SITES IN THE MAJOR GRAIN PRODUCING COUNTRIES

(NASA-TM-X-71288) AN ESTIMATE OF FIELD SIZE DISTRIBUTIONS FOR SELECTED SITES IN THE MAJOR GRAIN PRODUCING COUNTRIES (NASA) 39 p HC A03/MF A01 CSCL 02C

N77-21528

Unclas

G3/43 25812

MELVIN H. PODWYSOCKI

APRIL 1976

GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND

AN ESTIMATE OF FIELD SIZE DISTRIBUTIONS FOR SELECTED SITES IN THE MAJOR GRAIN PRODUCING COUNTRIES

Melvin H. Podwysocki
NASA/GSFC Code 923
Greenbelt, Maryland 20771

ABSTRACT

As a means for defining the resolution requirements for the Landsat-D satellite, especially with respect to the agriculture discipline, a study was undertaken to estimate the field size distributions for the major grain producing countries of the World. Landsat-1 and 2 images were evaluated for two areas each in the United States, People's Republic of China and the USSR. One scene each was evaluated for France, Canada and India. Grid sampling was done for representative sub-samples of each image, measuring the long and short axes of each field; area was then calculated. Each of the resulting data sets was computer analyzed for their frequency distributions. Nearly all frequency distributions were highly peaked and skewed (shifted) towards small values, approaching that of either a Poission or log-normal distribution. The data were normalized by a log transformation, creating a Gaussian distribution which has moments readily interpretable and useful for estimating the total population of fields. Resultant predictors of the field size estimates are discussed.

CONTENTS

ILLUSTRATIONS

CONTENTS (Continued)

TABLES

AN ESTIMATE OF FIELD SIZE DISTRIBUTIONS FOR SELECTED SITES IN THE MAJOR GRAIN PRODUCING COUNTRIES

INTRODUCTION

Landsat-1 and 2 have a nominal IFOV of 80 meters. Future satellites are planned with better resolution, but this must be justified. This work is an attempt to define the resolution limits required by the future Landsat-D satellite, with primary attention to the agriculture discipline, by use of statistical modeling.

METHODS

The definition of the primary grain producing countries and, in particular, those areas in which the crops are produced was derived from work performed under NASA contract NAS5-22837 by Dr. Peter Castruccio of Ecosystems, Inc. From his work, areas were randomly selected and the tapes and imagery ordered. Table 1 is a listing of the areas chosen and Landsat scenes for these areas. In order to maximize the differences between adjacent fields, imagery taken either at the beginning or end of the growing seasons was selected (Castruccio, pers. comm.). Both digital-analog and strictly analog (photographic) methods for processing the imagery were examined, with the former proving superior because of the increased resolving power. Areas were selected from each of the images and extracted using the QUICK-LOOK program residing on the IBM 360 series computers at GSFC. These extracted scenes were processed to analog form (standard Landsat false color composites) by the DICOMED film recorder. Depending upon the size of the fields in the image, the scenes were analyzed at a scale of 1:45,000, 1:90,000 or 1:160,000. The last scale was used only in the USSR, where the field sizes were very large and examination of them at any larger scale would not permit the examination of a sufficient sample to be statistically valid.

The images were treated in the following fashion to extract the data. The ideal methodology, which entails a multispectral analysis of the image data to determine the crop content of the fields could not be applied because ground truth was unavailable. Instead, only those areas which were planted primarily in a single crop (in most cases wheat) were chosen.

Depending upon the field sizes within the scene, a sampling grid was chosen so that the same field would not be measured more than once. Table 1 contains the information on the sampling grid used in each case. The sampling grid forced the operators to measure those fields which fell under the grid point, and would tend to rule out a bias introduced by the desire to measure only fields which were readily discernible and easily measured. However, this technique was not followed for the Indian data because of the extremely small field sizes. Instead, a field in the vicinity of the grid point was measured if none was discernible at the grid point.

The field measuring method was partly automated. Initial examination of the images indicated that most fields were rectangular in shape. Thus, only a measurement of a length and width was necessary to calculate areas. In some cases, fields occured in triangular, parallelopiped and trapezoidal shapes. These were treated as rectangles, but with adjustment in the measurement procedures to maintain accuracy with respect to area (i.e., in the case of a right triangle, the measurement was: 1/2 (base x height).

Computer programs necessary for the conversion of the digitized data to a readily usable product were written. Parameters calculated for each field were length and width in kilometers, area in hectares, shape as a ratio of the short/long axes of the field and orientation of the long axis. Length, width and area will be discussed in detail; shape will be mentioned briefly. Each set of data was then analyzed by a computer program (NORMSTAT) which calculated the mean, standard deviation, skewness and kurtosis (1st - 4th moments) of each variable, as well as a Chi Square test for the "Goodness of Fit" of the data to a normal or Gaussian distribution. If the data are representative samples and Gaussian, then the mean and standard deviation for each variable can be used to estimate with some confidence the probable distribution of field sizes of the sampled population. If the data are representative but not Gaussian, the program allows for transformation of the data by one of several mathematical formulae to fit a normal distribution. These results are again tested for a "Goodness of Fit".

DISCUSSION

Figures 1 - 6 illustrate the typical output products from the NORMSTAT program. In the Kansas data, Figures 1, 3 and 5 show the histogram plots of the raw data for length, width and area respectively. It is obvious that the data do not form the typical bell shaped curve of a normal distribution, but have a skewness[1] (peak of the distribution shifted from the middle) towards the smaller values. Likewise, the kurtosis[2] measure (the amount of peakedness of the curve) is high. For normal distributions these should have values of 0.0 and 3.0 respectively. The "Goodness to Fit" test also fails. These distributions are similar to

1. Listed as B**1/2 in the computer output (Figures 1 - 6).
2. Listed as B2 in the computer output (Figures 1 - 6).

Poisson or log-normal distributions, which can be normalized by the transformation: $Y = \log_2 x$, where x is the original value and Y is its transformed equivalent. Figures 2, 4 and 6 show the results of this transformation for the Kansas data. All tests indicate that these populations cannot be rejected as being Gaussian. The same tests were applied to all other test sites, with nearly all being transformed to Gaussian form. The output of NORMSTAT for the other test sites is on file with the author. An empirical loading can be made to determine the confidence of all estimators of the population. Skewness, kurtosis and Chi Square "Goodness of Fit" tests may be tested to determine if they conform to normality. These are loaded in the following fashion:

	$P_{0.05}$	$P_{0.01}$
Skewness	-1	-2
Kurtosis	-1	-2
Chi Square	-2	-4

Starting with a perfect score of 8, if a measure failed at the $P_{0.05}$[1] or $P_{0.01}$, the appropriate amount of points would be deducted. Thus, if all three tests failed at $P_{0.01}$, a loading of zero would indicate low confidence in the population estimators.

Conversely, a value of 8 would indicate high confidence in the results. Results with confidence values less than 4 should be suspect. Tables 2 through 10 contain the summarized results for each of the study areas for both raw and transformed data. In nearly every case, there was a marked improvement in the confidence value for each variable after transformation.

Exceptions to the high confidence levels can be found for the length parameter in the Iowa data (Table 2). It appears that these values are bimodally distributed (2 distinct frequency peaks), with one of the peaks at the 0.8 km length. This is most likely caused by farming practices where fields are cultivated in 1/2 section (0.8 km or 0.5 mile) lengths. This same distribution was observed in this study by an analysis of field sizes as determined from plot maps for both the North Dakota and Kansas LACIE test sites. It does not show up in any of the other study areas chosen in this investigation. Log_{10} transformations produced similar values for the statistical tests and will be referred to later.

1. This states that if the population was normal, 5 times out of 100 one could get a value as large as the observed one by taking representative samples from the population.

Another possible exception occurs in the India data (Table 10), where both length and width have relatively low confidence. This may, in part, be due to the small field sizes for the area; most were at or below the resolution limits of the Landsat system. Thus, only those fields which exhibited a strong contrast to their adjacent neighbors could be measured with any degree of precision. This also is the reason for the low sample size of this study area.

The normal distribution is an excellent predictor of the population parameters. Within 1 standard deviation of the mean, 66% of the population is found; within 2 and 3 standard deviations respectively, 95 and 99% of the population occur. The expected frequency distribution values from the NORMSTAT program (see Figures 2b, 4b, 6b) also can be used to generate theoretical expected frequency distribution curves of field length, width and area for each of the study areas. Figures 7 and 8 summarize the data for the length and width parameters. The ordinate values of the graphs, listed in cumulative percent, can be interpreted as the number of fields which have values less than or equal to the value of the abscissa parameter. Thus in Figure 8, 50% of the fields in the India data would have widths less than or equal to approximately 70 meters. The abscissa values (length and width) were converted from $\log_2$ scale back to the original values and plotted on semi-log paper ($\log_{10}$) for ease of interpretation. As was stated earlier, this does not change the results of the statistical significance tests which were applied to the data.

With some exceptions, most of the curves parallel each other, but are offset along the abscissa, indicating that although the means (50% value) for each area are different, their standard deviations (in log space) are similar. The major exception to this is the India data, which exhibits a much greater spread in data values and indicates a larger standard deviation. This is in part due to the small sample size for this set of data (73), and the mensuration difficulties for fields at the resolution limits of the present Landsat system. In either case, this particular data set should probably be interpreted with some reservations.

Figure 9 represents a cumulative frequency curve for the area variable for each study area. The ordinate values differ from those figures depicting field length and width. The data was reintegrated; the ordinate depicts the sum total area of fields (in percent) and was derived from the expected distribution (number of fields, see Figure 6b) by the following formula:

$$f_i = \frac{X_i E_i}{\Sigma F_i}$$

where, f_i is the total area of fields (in %) per unit class, F_i is the actual area per unit class; X_i is the field size value (in hectares) for the class midpoint of the distribution; and E_i is the expected number of fields for the respective midclass. This was integrated over the 33 classes from the frequency distribution of the NORMSTAT program. This is a more valid method of representing the area data because, if it is assumed that all fields produce the same crop yields, then productivity of an area rather than the number of fields is depicted. As an example, in the India data, based on the model distributions, although 50% of the number of fields are less than 1.2 hectares in size, they only produce 10% of the total crop.

In order to determine the sensor resolution requirements, the field area parameter must be related to the field's smallest dimension, its width. Analysis of the shape parameter (short/long axis) indicated that few fields were equidimensional (shape factor of 1). Therefore, the square root of the area does not accurately describe the minimal dimension of the fields. Because of this, linear regression analysis was performed on the width vs. area parameters for each study area. Their results are summarized in Table 11. Correlation coefficients are statistically significant at the 0.05 level. The regression formulae are thus good predictors of the relationship between field width and area. Table 12 gives results for field widths associated with 50% and 90% levels of each study area's productivity, based on area values read from Figure 9.

FUTURE STUDIES

One of the basic assumptions of this study is that the selected study areas are representative of the total population of fields from each agricultural region. This has yet to be statistically proven. It is very likely that the values may actually deviate somewhat from the stated values. Additional work and sampling should be carried out to prove this technique for each region.

Some other method should also be made to test the applicability of this method of extrapolating the population of field sizes to higher resolution limits. A good technique might involve comparing the populations as determined from both Landsat and higher resolution Skylab data sources (i.e., S-190B Photography) for those a[illegible]s where both coverages exist.

ACKNOWLEDGMENTS

I wish to thank Dr. Peter Castruccio of Ecosystems, Inc. for the use of his data on the major crop producing areas of the world. Thanks are also extended to Messrs. Columbus Carlton and Richard McKinney of Computer Sciences

Corporation for their assistance in the task of collecting the data, to Mr. David Cavero, Computer Sciences-Technicolor Associates and Dr. Fred J. Gunther, Computer Sciences Corporation for their assistance in the data processing.

CONCLUSIONS

In order to determine resolution requirements for the future Landsat-D, portions of 9 Landsat MSS images were analyzed in the major grain producing regions of the World to estimate the population of field sizes for each region by statistical modeling. Length, width and area parameters were investigated in detail. In most cases, the populations of field parameters formed either Poisson or log-normal distributions (highly peaked and mode shifted towards smaller values). Transformations were performed which normalized the data, producing a Gaussian distribution. From these normalized distributions, predictions of the field size populations were extrapolated for small field sizes below or approaching the present Landsat resolution limits.

LENGTH KANSAS DATA

LEFT BOUNDARY = 0 RIGHT BOUNDARY = 22

INTERVAL IN S.D.'S FROM MEAN	OBSERVED COUNT
TO -3.875	0
-3.874 TO -3.625	0
-3.624 TO -3.375	0
-3.374 TO -3.125	0
-3.124 TO -2.875	0
-2.874 TO -2.625	0
-2.624 TO -2.375	0
-2.374 TO -2.125	0
-2.124 TO -1.875	0
-1.874 TO -1.625	0
-1.624 TO -1.375	4
-1.374 TO -1.125	7
-1.124 TO -0.875	16
-0.874 TO -0.625	18
-0.624 TO -0.375	16
-0.374 TO -0.125	15
-0.124 TO 0.125	22
0.126 TO 0.375	10
0.376 TO 0.625	10
0.626 TO 0.875	4
0.876 TO 1.125	5
1.126 TO 1.375	2
1.376 TO 1.625	1
1.626 TO 1.875	5
1.876 TO 2.125	4
2.126 TO 2.375	3
2.376 TO 2.625	4
2.626 TO 2.875	0
2.876 TO 3.125	0
3.126 TO 3.375	1
3.376 TO 3.625	0
3.626 TO 3.875	0
3.876 TO	0

MEAN = 0.446960E 00

VARIANCE = 0.333813E-01

STANDARD DEVIATION = 0.182706E 00

SAMPLE SIZE = 147

TESTS FOR SKEWNESS AND KURTOSIS

SEE PEARSON AND HARTLEY 'BIOMETRIKA TABLES FOR STATISTICIANS' VOL.1 TABLES 34 B AND C FOR CRITICAL VALUES

B1**1/2 = 1.039188

B2 = 3.600976

Figure 1a. Statistical Analysis of Non-normalized Field Lengths, Kansas Data

LENGTH KANSAS DATA

HISTOGRAM FOR SET

INTERVAL (IN S.D.'S FROM MEAN)	OBSERVED NUMBER	EXPECTED NUMBER	CHI SQUARE
TO -3.875	0	0.0079	0.0
-3.874 TO -3.625	0	0.0134	0.0
-3.624 TO -3.375	0	0.0331	0.0
-3.374 TO -3.125	[illegible]	0.0764	0.0
-3.124 TO -2.875	[illegible]	0.1663	0.0
-2.874 TO -2.625	0	0.3399	0.0
-2.624 TO -2.375	0	0.6530	0.0
-2.374 TO -2.125	0	1.1768	2.4667
-2.124 TO -1.875	0	1.9996	1.9996
-1.874 TO -1.625	[illegible]	3.1877	3.1877
-1.624 TO -1.375	4	4.7751	0.1258
-1.374 TO -1.125	7	6.7222	0.0115
-1.124 TO -0.875	16	8.8923	0.6812
-0.874 TO -0.625	13	11.0542	4.3643
-0.624 TO -0.375	16	12.9132	0.7379
-0.374 TO -0.125	15	14.1755	0.0480
-0.124 TO 0.125	22	14.6230	3.7216
0.126 TO 0.375	10	14.1755	1.2299
0.376 TO 0.625	1[illegible]	12.9132	0.6572
0.626 TO 0.875	4	11.0542	4.5017
0.876 TO 1.125	5	8.8923	1.7037
1.126 TO 1.375	[illegible]	6.7222	3.3172
1.376 TO 1.625	1	4.7751	2.9846
1.626 TO 1.875	3	3.1877	1.0304
1.876 TO 2.125	4	1.9996	2.0011
2.126 TO 2.375	3	1.1758	12.3931
2.376 TO 2.625	4	0.6530	0.0
2.626 TO 2.875	[illegible]	0.3399	0.0
2.876 TO 3.125	[illegible]	0.1663	0.0
3.126 TO 3.375	1	0.0764	0.0
3.376 TO 3.625	0	0.0331	0.0
3.626 TO 3.875	[illegible]	0.0134	0.0
3.876 TO	0	0.0079	0.0

CHI-SQUARED ON 15 D.F. = 22.164363

PROBABILITY OF EXCEEDING THIS VALUE OF CHI-SQUARED = 0.000010

Figure 1b. Statistical Analysis of Non-normalized Field Lengths, Kansas Data

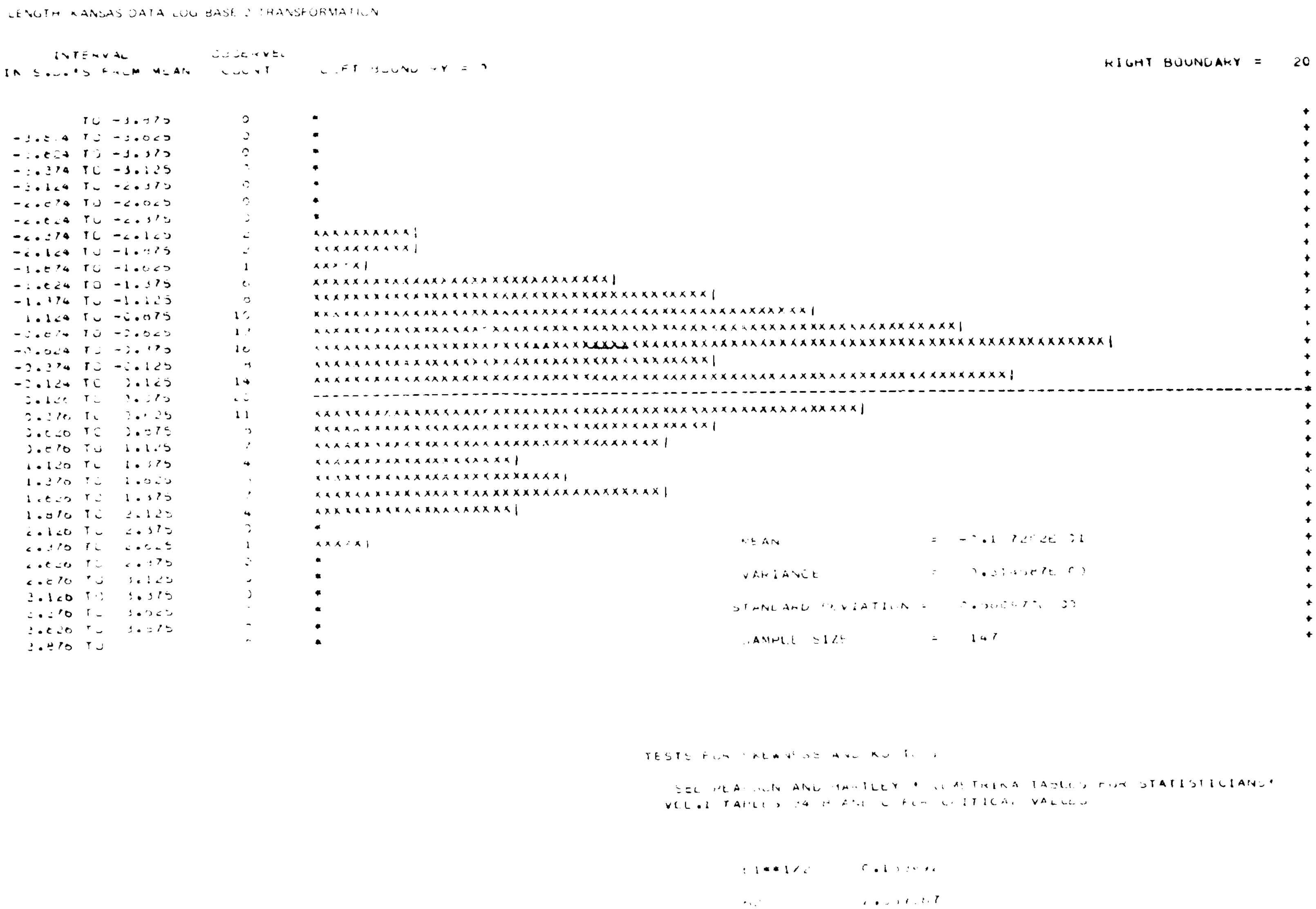


Figure 2a. Statistical Analysis of Normalized Field Lengths, Kansas Data

LENGTH KANSAS DATA LOG BASE 2 TRANSFORMATION

HISTOGRAM FOR SET

INTERVAL (IN SD UNITS FROM MEAN)	OBSERVED NUMBER	EXPECTED NUMBER	CHI SQUARE
TO -3.875	0	0.0079	0.0
-3.874 TO -3.625	0	0.0134	0.0
-3.624 TO -3.375	0	0.0331	0.0
-3.374 TO -3.125	0	0.0764	0.0
-3.124 TO -2.875	0	0.1663	0.0
-2.874 TO -2.625	0	0.3399	0.0
-2.624 TO -2.375	0	0.6530	0.0
-2.374 TO -2.125	2	1.1788	0.0890
-2.124 TO -1.875	2	1.9996	0.0000
-1.874 TO -1.625	1	3.1677	1.5014
-1.624 TO -1.375	6	4.7751	0.3112
-1.374 TO -1.125	3	6.7222	0.2429
-1.124 TO -0.875	10	8.8923	0.1380
-0.874 TO -0.625	12	11.0542	0.3425
-0.624 TO -0.375	16	12.9132	0.7379
-0.374 TO -0.125	8	14.1755	2.6903
-0.124 TO 0.125	14	14.6230	0.0265
0.126 TO 0.375	20	14.1755	2.3932
0.376 TO 0.625	11	12.9132	0.2835
0.626 TO 0.875	8	11.0542	0.8439
0.876 TO 1.125	7	8.8923	0.4027
1.126 TO 1.375	4	6.7222	1.1323
1.376 TO 1.625	5	4.7751	0.0196
1.626 TO 1.875	7	3.1677	4.5593
1.876 TO 2.125	4	1.9996	2.0011
2.126 TO 2.375	0	1.1788	0.8738
2.376 TO 2.625	1	0.6530	0.0
2.626 TO 2.875	0	0.3399	0.0
2.876 TO 3.125	0	0.1663	0.0
3.126 TO 3.375	0	0.0764	0.0
3.376 TO 3.625	0	0.0331	0.0
3.626 TO 3.875	0	0.0134	0.0
3.876 TO	0	0.0079	0.0

CHI-SQUARED ON 16 D.F. = 16.551009

PROBABILITY OF EXCEEDING THIS VALUE OF CHI-SQUARED = 0.292522

Figure 2b. Statistical Analysis of Normalized Field Lengths, Kansas Data

WIDTH KANSAS DATA

LEFT BOUNDARY = 0 RIGHT BOUNDARY = 20

INTERVAL IN S.D.'S FROM MEAN	OBSERVED COUNT
TO -3.875	0
-3.874 TO -3.625	0
-3.624 TO -3.375	0
-3.374 TO -3.125	0
-3.124 TO -2.875	0
-2.874 TO -2.625	0
-2.624 TO -2.375	0
-2.374 TO -2.125	0
-2.124 TO -1.875	0
-1.874 TO -1.625	1
-1.624 TO -1.375	5
-1.374 TO -1.125	14
-1.124 TO -0.875	9
-0.874 TO -0.625	15
-0.624 TO -0.375	7
-0.374 TO -0.125	20
-0.124 TO 0.125	15
0.126 TO 0.375	17
0.376 TO 0.625	12
0.626 TO 0.875	7
0.876 TO 1.125	10
1.126 TO 1.375	4
1.376 TO 1.625	4
1.626 TO 1.875	2
1.876 TO 2.125	1
2.126 TO 2.375	0
2.376 TO 2.625	1
2.626 TO 2.875	0
2.876 TO 3.125	2
3.126 TO 3.375	0
3.376 TO 3.625	0
3.626 TO 3.875	0
3.876 TO	1

MEAN = [illegible]

VARIANCE = 0.1298116-01

STANDARD DEVIATION = 0.1139350 00

SAMPLE SIZE = 147

TESTS FOR SKEWNESS AND KURTOSIS

SEE PEARSON AND HARTLEY 'BIOMETRIKA TABLES FOR STATISTICIANS'
VOL.1 TABLES 34 B AND C FOR CRITICAL VALUES

B1**1/2 = 1.159180

B2 = 6.198716

Figure 3a. Statistical Analysis of Non-normalized Field Widths, Kansas Data

WIDTH KANSAS DATA

HISTOGRAM FOR SET

INTERVAL (IN S.D.'S FROM MEAN)	OBSERVED NUMBER	EXPECTED NUMBER	CHI SQUARE
TO -3.875	0	0.0079	0.0
-3.874 TO -3.625	0	0.0134	0.0
-3.624 TO -3.375	0	0.0331	0.0
-3.374 TO -3.125	0	0.0764	0.0
-3.124 TO -2.875	0	0.1663	0.0
-2.874 TO -2.625	0	0.3399	0.0
-2.624 TO -2.375	0	0.6530	0.0
-2.374 TO -2.125	0	1.1788	2.4087
-2.124 TO -1.875	0	1.9996	1.9996
-1.874 TO -1.625	1	3.1877	1.5014
-1.624 TO -1.375	5	4.7751	0.0106
-1.374 TO -1.125	14	6.7222	7.8794
-1.124 TO -0.875	9	8.8923	0.0013
-0.874 TO -0.625	15	11.0542	1.4084
-0.624 TO -0.375	7	12.9132	2.7078
-0.374 TO -0.125	20	14.1755	2.3932
-0.124 TO 0.125	15	14.6230	0.0097
0.126 TO 0.375	17	14.1755	0.5028
0.376 TO 0.625	12	12.9132	0.0646
0.626 TO 0.875	7	11.0542	1.4869
0.876 TO 1.125	10	8.8923	0.1380
1.126 TO 1.375	4	6.7222	1.1023
1.376 TO 1.625	4	4.7751	0.1258
1.626 TO 1.875	2	3.1877	0.4425
1.876 TO 2.125	1	1.9996	0.4997
2.126 TO 2.375	0	1.1788	0.9496
2.376 TO 2.625	1	0.6530	0.0
2.626 TO 2.875	0	0.3399	0.0
2.876 TO 3.125	2	0.1663	0.0
3.126 TO 3.375	0	0.0764	0.0
3.376 TO 3.625	0	0.0331	0.0
3.626 TO 3.875	0	0.0134	0.0
3.876 TO	1	0.0079	0.0

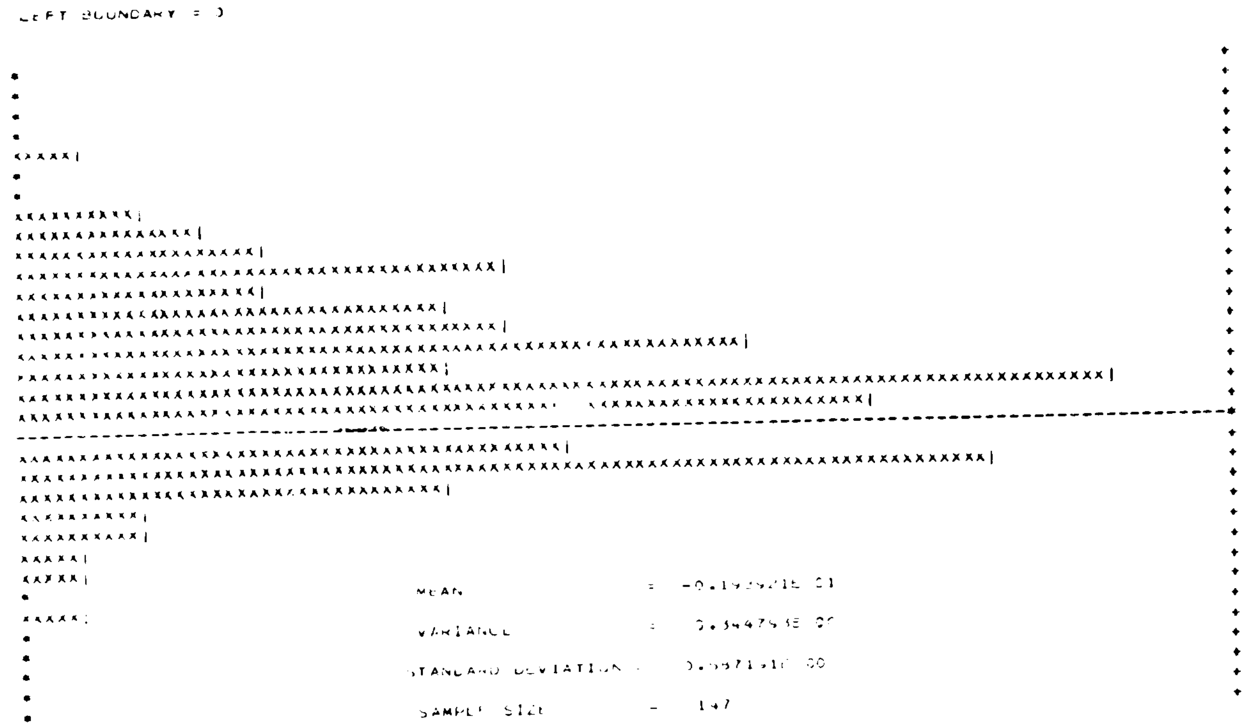
CHI-SQUARED ON 16 D.F. = 25.752025

PROBABILITY OF EXCEEDING THIS VALUE OF CHI-SQUARED = 0.057610

Figure 3b. Statistical Analysis of Non-normalized Field Widths, Kansas Data

WIDTH KANSAS DATA LOG BASE 2 TRANSFORMATION

INTERVAL IN S.D.'S FROM MEAN	OBSERVED COUNT
TO -3.875	0
-3.874 TO -3.625	0
-3.624 TO -3.375	0
-3.374 TO -3.125	0
-3.124 TO -2.875	1
-2.874 TO -2.625	0
-2.624 TO -2.375	0
-2.374 TO -2.125	2
-2.124 TO -1.875	3
-1.874 TO -1.625	4
-1.624 TO -1.375	[illegible]
-1.374 TO -1.125	4
-1.124 TO -0.875	7
-0.874 TO -0.625	8
-0.624 TO -0.375	1[illegible]
-0.374 TO -0.125	7
-0.124 TO 0.125	13
0.126 TO 0.375	14
0.376 TO 0.625	2[illegible]
0.626 TO 0.875	[illegible]
0.876 TO 1.125	1[illegible]
1.126 TO 1.375	7
1.376 TO 1.625	2
1.626 TO 1.875	2
1.876 TO 2.125	1
2.126 TO 2.375	1
2.376 TO 2.625	[illegible]
2.626 TO 2.875	[illegible]
2.876 TO 3.125	0
3.126 TO 3.375	0
3.376 TO 3.625	0
3.626 TO 3.875	0
3.876 TO	0



MEAN = -0.192[illegible]E 01

VARIANCE = 0.34479[illegible]E 00

STANDARD DEVIATION = 0.5871[illegible]E 00

SAMPLE SIZE = 147

TESTS FOR SKEWNESS AND KURTOSIS

SEE PEARSON AND HARTLEY 'BIOMETRIKA TABLES FOR STATISTICIANS' VOL.I TABLES 34 B AND C FOR CRITICAL VALUES

B1**1/2 = -0.[illegible]

B2 = [illegible]

Figure 4a. Statistical Analysis of Normalized Field Widths, Kansas Data

WIDTH KANSAS DATA LOG BASE 2 TRANSFORMATION

HISTOGRAM FOR SET

INTERVAL (IN S.D.'S FROM MEAN)	OBSERVED NUMBER	EXPECTED NUMBER	CHI SQUARE
TO -3.875	0	0.0079	0.0
-3.874 TO -3.625	0	0.0134	0.0
-3.624 TO -3.375	0	0.0331	0.0
-3.374 TO -3.125	0	0.0764	0.0
-3.124 TO -2.875	1	0.1663	0.0
-2.874 TO -2.625	0	0.3399	0.0
-2.624 TO -2.375	0	0.6530	0.0
-2.374 TO -2.125	2	1.1788	0.1143
-2.124 TO -1.875	3	1.9996	0.5004
-1.874 TO -1.625	4	3.1877	0.2070
-1.624 TO -1.375	3	4.7751	2.1779
-1.374 TO -1.125	4	6.7222	1.1023
-1.124 TO -0.875	7	8.8923	0.4027
-0.874 TO -0.625	8	11.0542	0.8439
-0.624 TO -0.375	12	12.9132	0.0646
-0.374 TO -0.125	7	14.1755	3.6322
-0.124 TO 0.125	18	14.6230	0.7799
0.126 TO 0.375	14	14.1755	0.0022
0.376 TO 0.625	20	12.9132	3.8892
0.626 TO 0.875	9	11.0542	0.3817
0.876 TO 1.125	10	8.8923	0.6812
1.126 TO 1.375	7	6.7222	0.0115
1.376 TO 1.625	2	4.7751	1.6128
1.626 TO 1.875	2	3.1877	0.4425
1.876 TO 2.125	1	1.9996	0.4997
2.126 TO 2.375	1	1.1788	0.0890
2.376 TO 2.625	0	0.6530	0.0
2.626 TO 2.875	1	0.3399	0.0
2.876 TO 3.125	0	0.1663	0.0
3.126 TO 3.375	0	0.0764	0.0
3.376 TO 3.625	0	0.0331	0.0
3.626 TO 3.875	0	0.0134	0.0
3.876 TO	0	0.0079	0.0

CHI-SQUARED ON 16 D.F. = 22.435074

PROBABILITY OF EXCEEDING THIS VALUE OF CHI-SQUARED = 0.129691

Figure 4b. Statistical Analysis of Normalized Field Widths, Kansas Data

AREA, KANSAS DATA

INTERVAL IN S.D.'S FROM MEAN	OBSERVED COUNT
TO -3.875	0
-3.874 TO -3.625	0
-3.624 TO -3.375	0
-3.374 TO -3.125	0
-3.124 TO -2.875	0
-2.874 TO -2.625	0
-2.624 TO -2.375	0
-2.374 TO -2.125	0
-2.124 TO -1.875	0
-1.874 TO -1.625	0
-1.624 TO -1.375	0
-1.374 TO -1.125	0
-1.124 TO -0.875	13
-0.874 TO -0.625	21
-0.624 TO -0.375	25
-0.374 TO -0.125	13
-0.124 TO 0.125	13
0.126 TO 0.375	13
0.376 TO 0.625	6
0.626 TO 0.875	7
0.876 TO 1.125	5
1.126 TO 1.375	3
1.376 TO 1.625	4
1.626 TO 1.875	1
1.876 TO 2.125	3
2.126 TO 2.375	0
2.376 TO 2.625	0
2.626 TO 2.875	0
2.876 TO 3.125	1
3.126 TO 3.375	1
3.376 TO 3.625	0
3.626 TO 3.875	0
3.876 TO	2

LEFT BOUNDARY = 0 RIGHT BOUNDARY = 25

MEAN = 0.137554E 02

VARIANCE = 0.1160135E 03

STANDARD DEVIATION = 0.1077705E 02

SAMPLE SIZE = 147

TESTS FOR SKEWNESS AND KURTOSIS

SEE PEARSON AND HARTLEY BIOMETRIKA TABLES FOR STATISTICIANS' VOL.1 TABLES 34 B AND C FOR CRITICAL VALUES

B1**1/2 = 2.317418

B2 = 11.057660

Figure 5a. Statistical Analysis of Non-normalized Field Areas, Kansas Data

AREA KANSAS DATA

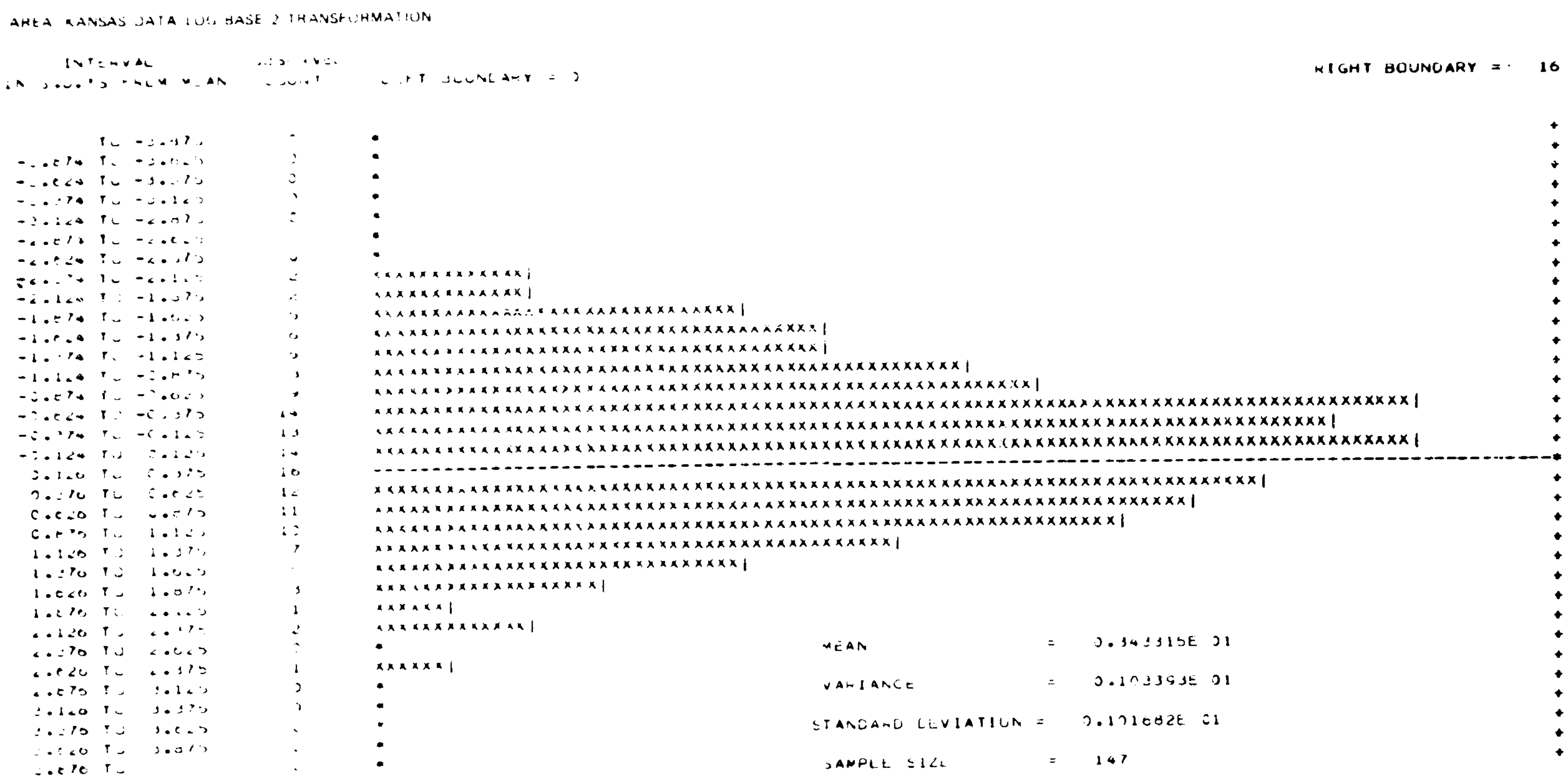
HISTOGRAM FOR SET

INTERVAL (IN S.D.'S FROM MEAN)	OBSERVED NUMBER	EXPECTED NUMBER	CHI SQUARE
TO -3.875	0	0.0079	0.0
-3.874 TO -3.625	0	0.0134	0.0
-3.624 TO -3.375	0	0.0331	0.0
-3.374 TO -3.125	0	0.0764	0.0
-3.124 TO -2.875	0	0.1663	0.0
-2.874 TO -2.625	0	0.3399	0.0
-2.624 TO -2.375	0	0.6530	0.0
-2.374 TO -2.125	0	1.1788	2.4087
-2.124 TO -1.875	0	1.9996	1.9996
-1.874 TO -1.625	0	3.1877	3.1877
-1.624 TO -1.375	0	4.7751	4.7751
-1.374 TO -1.125	0	6.7222	6.7222
-1.124 TO -0.875	18	8.8923	9.3282
-0.874 TO -0.625	21	11.0542	8.9484
-0.624 TO -0.375	25	12.9132	11.3133
-0.374 TO -0.125	18	14.1755	1.0318
-0.124 TO 0.125	18	14.6230	0.7799
0.126 TO 0.375	13	14.1755	0.0975
0.376 TO 0.625	5	12.9132	3.7011
0.626 TO 0.875	7	11.0542	1.4869
0.876 TO 1.125	6	8.8923	0.9408
1.126 TO 1.375	3	6.7222	2.0610
1.376 TO 1.625	4	4.7751	0.1258
1.626 TO 1.875	1	3.1877	1.5014
1.876 TO 2.125	3	1.9996	0.5004
2.126 TO 2.375	0	1.1788	0.9498
2.376 TO 2.625	0	0.6530	0.0
2.626 TO 2.875	0	0.3399	0.0
2.876 TO 3.125	1	0.1663	0.0
3.126 TO 3.375	1	0.0764	0.0
3.376 TO 3.625	0	0.0331	0.0
3.626 TO 3.875	0	0.0134	0.0
3.876 TO	2	0.0079	0.0

CHI-SQUARED ON 16 D.F. = 61.81532

PROBABILITY OF EXCEEDING THIS VALUE OF CHI-SQUARED = 0.000000

Figure 5b. Statistical Analysis of Non-normalized Field Areas, Kansas Data



TESTS FOR SKEWNESS AND KURTOSIS

SEE PEARSON AND HARTLEY 'BIOMETRIKA TABLES FOR STATISTICIANS'
VOL.1 TABLES 34 B AND C FOR CRITICAL VALUES

B1**1/2 = -0.032546

B2 = 2.741034

Figure 6a. Statistical Analysis of Normalized Field Areas, Kansas Data

AREA KANSAS DATA LOG BASE 2 TRANSFORMATION

HISTOGRAM FOR SET

INTERVAL (IN S.D.'S FROM MEAN)	OBSERVED NUMBER	EXPECTED NUMBER	CHI SQUARE
TO -3.875	0	0.0079	0.0
-3.874 TO -3.625	0	0.0134	0.0
-3.624 TO -3.375	0	0.0331	0.0
-3.374 TO -3.125	0	0.0764	0.0
-3.124 TO -2.875	0	0.1663	0.0
-2.874 TO -2.625	0	0.3399	0.0
-2.624 TO -2.375	0	0.6530	0.0
-2.374 TO -2.125	2	1.1788	0.0899
-2.124 TO -1.875	2	1.9996	0.0000
-1.874 TO -1.625	5	3.1677	1.0304
-1.624 TO -1.375	6	4.7751	0.3142
-1.374 TO -1.125	6	6.7222	0.0776
-1.124 TO -0.875	8	8.8923	0.0895
-0.874 TO -0.625	9	11.0542	0.3817
-0.624 TO -0.375	14	12.9132	0.0915
-0.374 TO -0.125	13	14.1755	0.0975
-0.124 TO 0.125	14	14.6230	0.0265
0.126 TO 0.375	16	14.1755	0.2348
0.376 TO 0.625	12	12.9132	0.0646
0.626 TO 0.875	11	11.0542	0.0003
0.876 TO 1.125	10	8.8923	0.1380
1.126 TO 1.375	7	6.7222	0.0115
1.376 TO 1.625	5	4.7751	0.0106
1.626 TO 1.875	3	3.1877	0.0111
1.876 TO 2.125	1	1.9996	0.4997
2.126 TO 2.375	2	1.1788	0.1143
2.376 TO 2.625	0	0.6530	0.0
2.626 TO 2.875	1	0.3399	0.0
2.876 TO 3.125	0	0.1663	0.0
3.126 TO 3.375	0	0.0764	0.0
3.376 TO 3.625	0	0.0331	0.0
3.626 TO 3.875	0	0.0134	0.0
3.876 TO	0	0.0079	0.0

CHI-SQUARED ON 16 D.F. = 3.262718

PROBABILITY OF EXCEEDING THIS VALUE OF CHI-SQUARED = 0.995000

Figure 6b. Statistical Analysis of Normalized Field Areas, Kansas Data

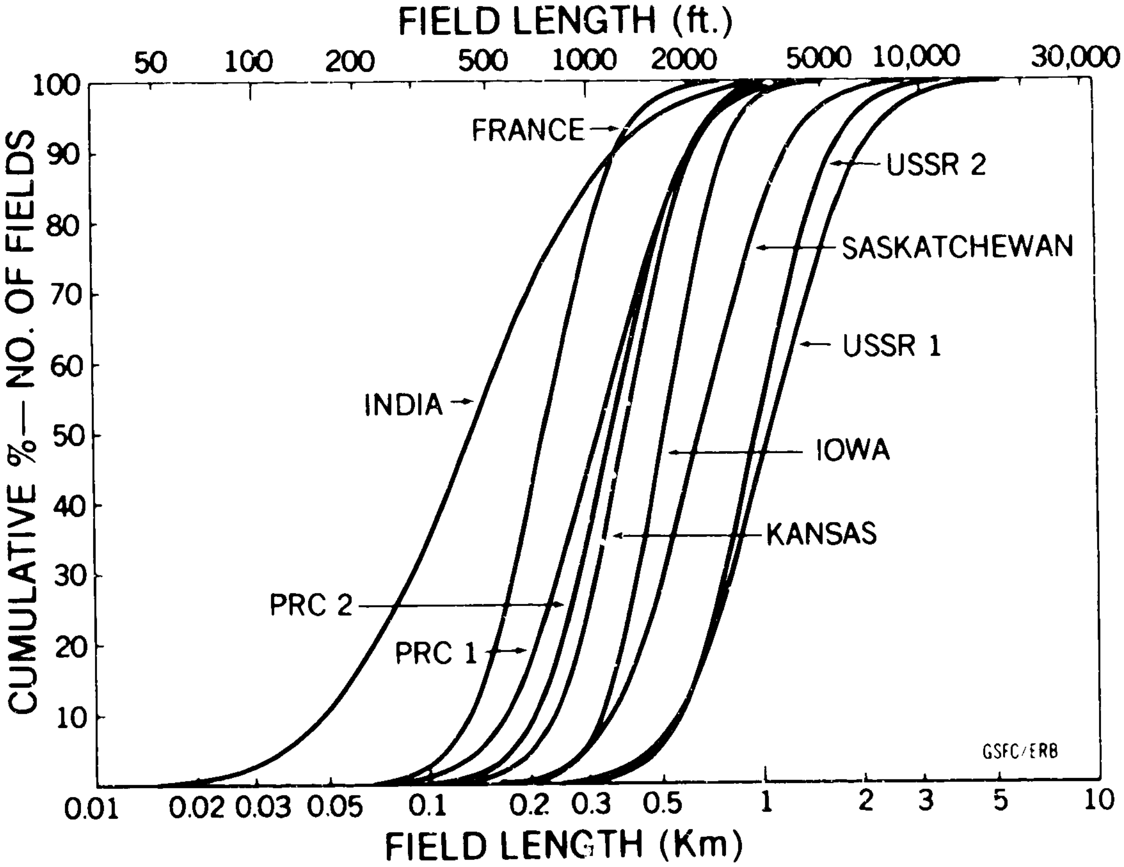


Figure 7. Cumulative Frequency Distribution (in %) of Field Length vs. Total Number of Fields

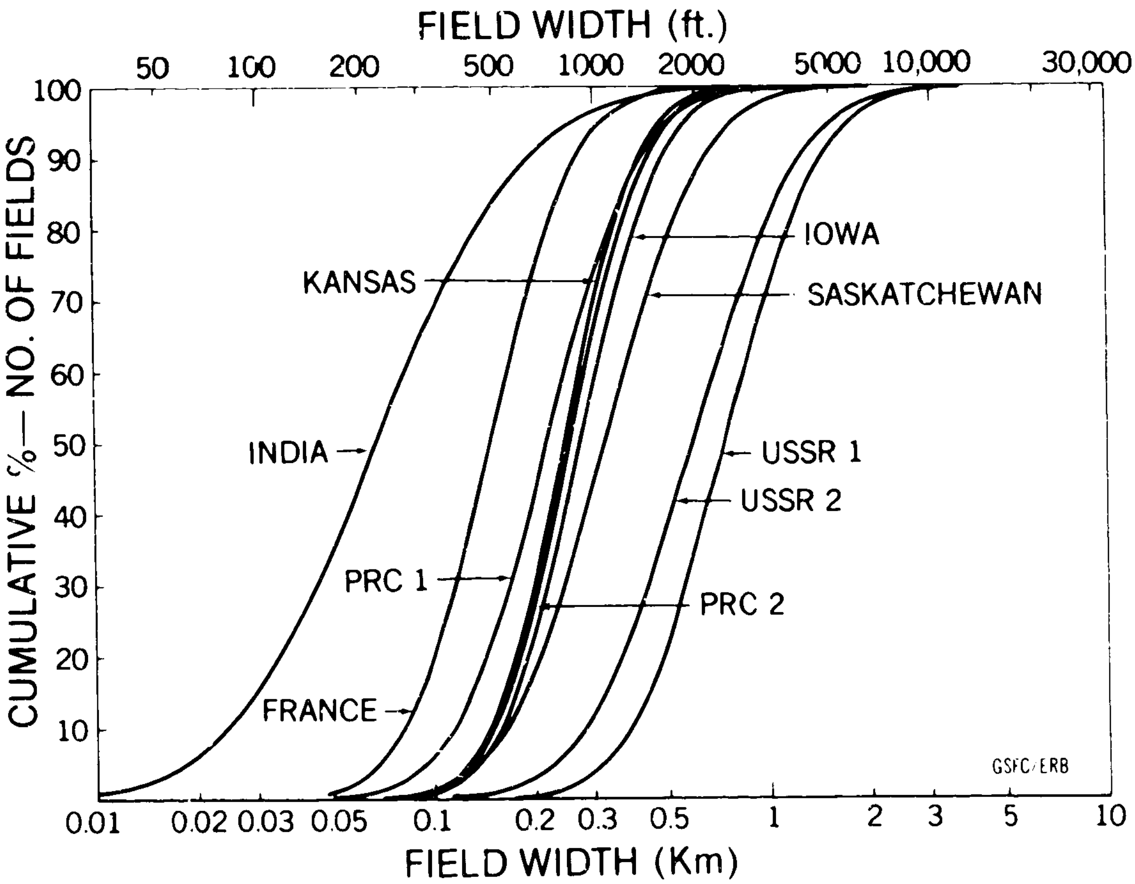


Figure 8. Cumulative Frequency Distribution (in %) of Field Width vs. Total Number of Fields

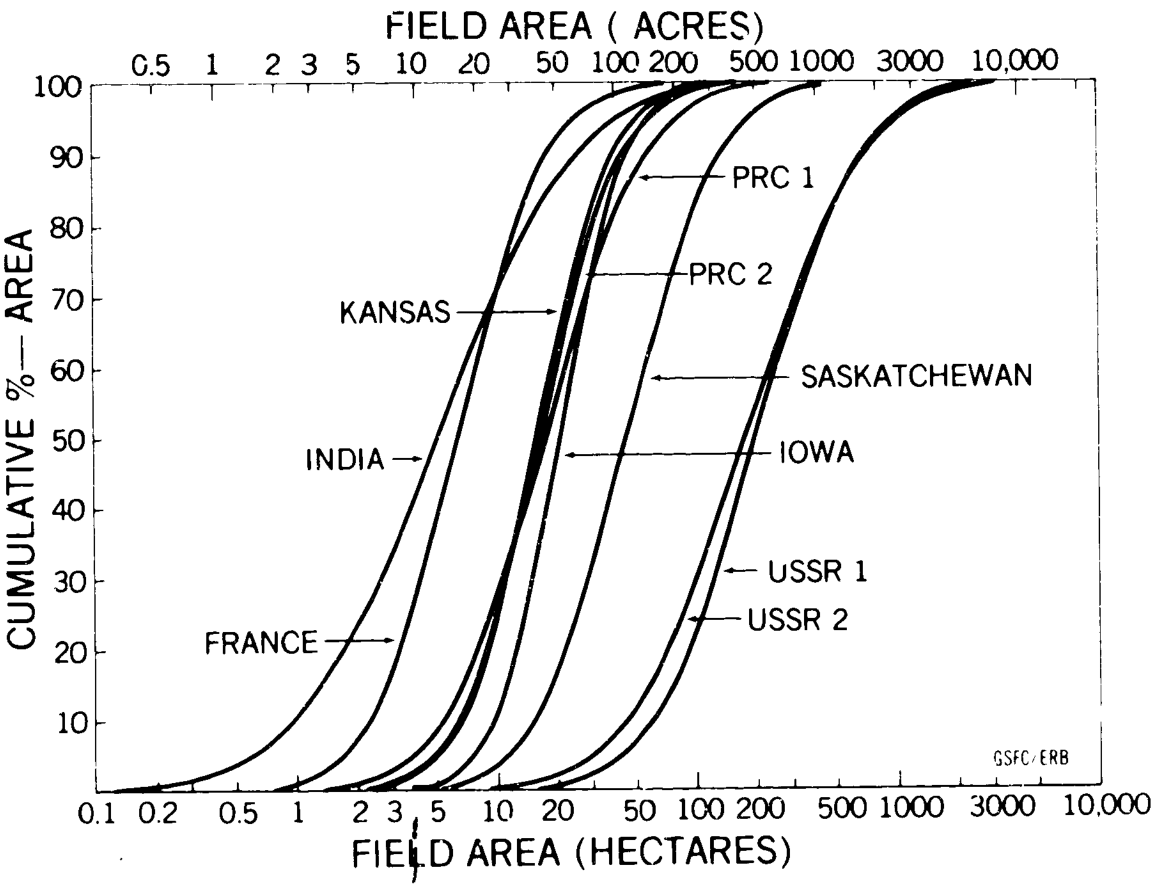


Figure 9. Cumulative Frequency Distribution (in %) of Field Area vs. Total Cumulative Area

Table 1

Listing of Study Areas

Landsat Scene I.D.	Image Nadir Point	Study Area[1]	Abbreviation	Sample[2] Grid Spacing
2233-16203	N 41°39′ W 94°28′	Greene & Boone Cos., Iowa	Iowa	1
1348-16511	N 38°53′ W 100°04′	Rush Co., Kansas	Kansas	1
2170-17105	N 50°16′ W 104°02′	Regina Area, Saskatchewan	Saskat.	1/2
2151-06444	N 52°55′ E 53°37′	Orenburg Region, Bashkir ASSR	USSR1	1-1/2
2119-07060	N 57°15′ E 50°32′	Kirov Region, USSR	USSR2	2
2104-02191	N 38°23′ E 115°16′	Hopeh Region, PRC	PRC1	1/2 x 1
2128-02533	N 38°43′ E 106°43′	Ningsia Autonomous Region, PRC	PRC2	1/2 x 1
2185-10022	N 43°17′ E 0°43′	Garonne Province, France	France	1/2 x 1
2034-04185	N 23°04′ E 82°06′	Bilaspur District, Madhya Pradesh, India	India	1/2 x 1

[1] All areas produced primarily wheat, with the following exceptions:
PRC1 - Wheat and Rice
Iowa - Corn

[2] Value in Km. One value indicates a square sampling grid, two values a rectangular grid.

PRECEDING PAGE BLANK NOT FILMED

Table 2

Population Estimators
KANSAS
SAMPLE SIZE = 147

	Raw Data			Transformed Data		
	Length[1]	Width[1]	Area[2]	Length[3]	Width[3]	Area[4]
Mean	.41	.24	11.34	-1.27	-1.94	3.43
Std. Dev.	.17	.16	9.93	.56	.59	1.02
Skewness	.77**	1.12**	1.86**	.188^{NS}	- .27^{NS}	- .003^{NS}
Kurtosis	3.16^{NS}	4.27**	6.42**	2.56^{NS}	3.02^{NS}	2.74^{NS}
Prob. χ^2	.02*	.0014**	<.0001**	.29^{NS}	.13^{NS}	.995^{NS}
Confidence	4	0	0	8	8	8

^{NS} Nonsignificant
* Significant at $P_{.05}$
** Significant at $P_{.01}$

[1] Mean & Std. Dev. in Km
[2] Mean & Std. Dev. in Hectares
[3] Same as #1, but data transformed as below
[4] Same as #2, but data transformed as below

$Y = \log_2 x$
where x = original value and
Y = transformed value

Table 3

Population Estimators
IOWA
SAMPLE SIZE = 97

	Raw Data			Transformed Data		
	Length[1]	Width[1]	Area[2]	Length[3]	Width[3]	Area[4]
Mean	.57	.32	18.98	-.90	-1.76	3.98
Std. Dev.	.18	.13	12.2	.49	.63	.89
Skewness	.49*	.74**	1.54**	-.21 NS	- .44*	- .04 NS
Kurtosis	3.22 NS	3.64 NS	5.91**	2.22*	2.92 NS	2.67 NS
Prob. χ^2	<.0001**	.0095**	.002**	.0013**	.02*	.96 NS
Confidence	3	2	0	3	5	8

NS Nonsignificant
*Significant at $P_{.05}$
**Significant at $P_{.01}$

[1] Mean & Std. Dev. in Km
[2] Mean & Std. Dev. in Hectares
[3] Same as #1, but data transformed as below
[4] Same as #2, but data transformed as below

$Y = \log_2 x$
where X = original value and
Y = transformed value

Table 4

Population Estimators
SASKATCHEWAN
SAMPLE SIZE = 101

	Raw Data			Transformed Data		
	Length[1]	Width[1]	Area[2]	Length[3]	Width[3]	Area[4]
Mean	.771	.39	33.3	- .52	-1.51	4.6
Std. Dev.	.344	.20	26.6	.68	.77	1.24
Skewness	.712**	.81**	1.75**	- .54*	- .59**	- .82**
Kurtosis	2.852[NS]	3.27[NS]	6.42**	3.77[NS]	3.79*	4.56**
Prob. χ^2	<.0001**	.030*	<.0001**	.027*	.77[NS]	.22[NS]
Confidence	2	4	0	5	5	4

[NS] Nonsignificant
*Significant at $P_{.05}$
**Significant at $P_{.01}$

[1] Mean & Std. Dev. in Km
[2] Mean & Std. Dev. in Hectares
[3] Same as #1, but data transformed as below
[4] Same as #2, but data transformed as below

$Y = \log_2 x$
where X = original value and
Y = transformed value

Table 5

Population Estimators
USSR1
SAMPLE SIZE = 198

	Raw Data			Transformed Data		
	Length[1]	Width[1]	Area[2]	Length[3]	Width[3]	Area[4]
Mean	1.29	.93	141.0	.19	- .30	6.53
Std. Dev.	.64	.50	148.9	.72	.74	1.37
Skewness	1.21**	1.58**	3.26**	- .21^{NS}	- .06^{NS}	- .23^{NS}
Kurtosis	5.34**	6.90**	19.2**	2.72^{NS}	2.97^{NS}	2.97^{NS}
Prob. χ^2	.003**	.0003**	<.0001**	.36^{NS}	.93^{NS}	.08^{NS}
Confidence	0	0	0	8	8	8

^{NS}Nonsignificant
*Significant at $P_{.05}$
**Significant at $P_{.01}$

[1] Mean & Std. Dev. in Km
[2] Mean & Std. Dev. in Hectares
[3] Same as #1, but data transformed as below
[4] Same as #2, but data transformed as below

$Y = \log_2 x$
where X = original value and
Y = transformed value

Table 6

Population Estimators
USSR 2
SAMPLE SIZE = 144

	Raw Data			Transformed Data		
	Length[1]	Width[1]	Area[2]	Length[3]	Width[3]	Area[4]
Mean	1.12	.73	95.1	.042	- .60	6.08
Std. Dev.	.51	.37	99.2	.60	.67	1.15
Skewness	1.57**	1.73**	3.34**	.21^{NS}	.18^{NS}	.30^{NS}
Kurtosis	6.57**	7.30**	18.6**	3.08^{NS}	3.19^{NS}	3.26^{NS}
Prob. X^2	<.0001**	<.0001**	<.0001**	.66^{NS}	.87^{NS}	.007**
Confidence	0	0	0	8	8	4

^{NS} Nonsignificant
* Significant at $P_{.05}$
** Significant at $P_{.01}$

[1] Mean & Std. Dev. in Km
[2] Mean & Std. Dev. in Hectares
[3] Same as #1, but data transformed as below
[4] Same as #2, but data transformed as below

$Y = \log_2 X$
where X = original value and
Y = transformed value

Table 7

Population Estimators
PRC1
SAMPLE SIZE = 191

	Raw Data			Transformed Data		
	Length[1]	Width[1]	Area[2]	Length[3]	Width[3]	Area[4]
Mean	.40	.27	12.4	-1.51	-2.08	3.05
Std. Dev.	.22	.14	12.7	.71	.74	1.35
Skewness	3.99**	1.27**	3.0**	- .15NS	.19NS	-.30*
Kurtosis	33.9**	5.13**	15.98**	4.27**	3.22NS	3.42NS
Prob. χ^2	.00019**	<.0001**	<.0001**	805NS	.62NS	.0465*
Confidence	0	0	0	6	8	5

NS Nonsignificant
*Significant at $P_{.05}$
**Significant at $P_{.01}$

[1] Mean & Std. Dev. in Km
[2] Mean & Std. Dev. in Hectares
[3] Same as #1, but data transformed as below
[4] Same as #2, but data transformed as below

$Y = \log_2 x$
where X = original value and
Y = transformed value

Table 8

Population Estimators
PRC2
SAMPLE SIZE = 155

	Raw Data			Transformed Data		
	Length[1]	Width[1]	Area[2]	Length[3]	Width[3]	Area[4]
Mean	.41	.30	13.5	-1.41	-1.87	3.36
Std. Dev.	.17	.12	11.0	.59	.60	1.11
Skewness	1.03**	.89**	2.76**	- .27^{NS}	- .23^{NS}	- .31^{NS}
Kurtosis	4.99**	4.12*	15.0**	2.97^{NS}	2.84^{NS}	3.10^{NS}
Prob. X^2	.014*	.03*	<.0001**	.039*	.085^{NS}	.41^{NS}
Confidence	2	3	0	6	8	8

^{NS}Nonsignificant
*Significant at $P_{.05}$
**Significant at $P_{.01}$

[1] Mean & Std. Dev. in Km
[2] Mean & Std. Dev. in Hectares
[3] Same as #1, but data transformed as below
[4] Same as #2, but data transformed as below

$Y = \log_2 x$
where X = original value and
Y = transformed value

Table 9

Population Estimators
FRANCE
SAMPLE SIZE = 151

	Raw Data			Transformed Data		
	Length[1]	Width[1]	Area[2]	Length[3]	Width[3]	Area[4]
Mean	.26	.18	5.11	-2.09	-2.66	1.89
Std. Dev.	.11	.08	4.76	.58	.68	1.16
Skewness	1.71**	.91**	2.85**	.15^NS	- .27^NS	- .03^NS
Kurtosis	8.24**	3.55^NS	14.3**	3.06^NS	2.97^NS	2.94^NS
Prob. χ^2	.0019**	.0011**	<.0001**	.72^NS	.33^NS	.99^NS
Confidence	2	4	6	8	8	8

^NS Nonsignificant
*Significant at $P_{.05}$
**Significant at $P_{.01}$

[1] Mean & Std. Dev. in Km
[2] Mean & Std. Dev. in Hectares
[3] Same as #1, but data transformed as below
[4] Same as #2, but data transformed as below

$$Y = \log_e x$$

where X = original value and
Y = transformed value

Table 10

Population Estimators
INDIA
SAMPLE SIZE = 73

	Raw Data			Transformed Data		
	Length[1]	Width[1]	Area[2]	Length[3]	Width[3]	Area[4]
Mean	.30	.097	3.72	-2.75	-3.72	.18
Std. Dev.	1.13	.066	14.68	1.12	1.15	1.45
Skewness	8.27**	1.44**	7.97**	2.08**	-1.41**	.64*
Kurtosis	69.90**	5.57**	66.48**	12.56**	7.61**	4.28*
Prob. X^2	<.0001**	.065^{NS}	<.0001**	.071^{NS}	.50^{NS}	.85^{NS}
Confidence	0	4	0	4	4	6

^{NS} Nonsignificant
*Significant at $P_{.05}$
** Significant at $P_{.01}$

[1] Mean & Std. Dev. in Km
[2] Mean & Std. Dev. in Hectares
[3] Same as # 1, but data transformed as below
[4] Same as # 2, but data transformed as below

$Y = \log_2 x$
where X = original value and
Y = transformed value

Table 11

Results of Linear Regression Analysis for Field Area vs. Width

Study Area	Correlation Coefficient (R)	Coefficients	
		A	B
Kansas	.881	83.2630	- 9.71
Iowa	.868	79.7337	- 6.69
Saskatchewan	.847	113.7044	- 12.18
USSR1	.927	278.2119	-116.43
USSR2	.908	238.4033	- 79.28
PRC1	.853	78.0902	- 8.55
PRC2	.896	79.3432	- 9.99
France	.886	52.3197	- 4.05
India	.936	38.2715	- 1.70

To determine the average width association for a given field size, use the following general formula:

Field width (in Km) = $\frac{X+B}{A}$ where A and B are the coefficients listed above and X is the field size (in hectares) as read off the % cumulative area vs. area graphs (Fig. 9).

Table 12

Examples of Related Field Areas and Widths as Determined from Regression Analysis

Study Area	50%[1]		90%[1]	
	Area[2]	Field Width[3]	Area[2]	Field Width[3]
Kansas	16.16	.311	6.24	.192
Iowa	21.24	.350	11.67	.230
Saskatchewan	45.89	.511	14.92	.230
USSR1	203.19	1.149	59.25	.631
USSR2	176.88	1.074	43.54	.515
PRC1	17.95	.339	5.03	.174
PRC2	16.62	.335	6.02	.202
France	6.50	.202	2.21	.120
India	5.09	.177	0.90	.068

[1] Percentage of total production to be inventoried (see text).
[2] Denotes the minimum field size (in hectares) associated with the given percentage of the total production.
[3] Denotes the average minimum width of fields (in Km) associated with given percentage of the production.

END

DATE

FILMED

JUN 7 1977