Statistical Reporting Service

U.S. Department of Agriculture

Washington, D.C. 20250

THE DEVELOPMENT OF WITHIN-YEAR FORECASTING MODELS FOR SPRING WHEAT (OTHER THAN DURUM)

RESEARCH AND DEVELOPMENT BRANCH

RESEARCH DIVISION

STATISTICAL REPORTING SERVICE

U. S. DEPARTMENT OF AGRICULTURE

NOVEMBER, 1976

THE DEVELOPMENT OF WITHIN-YEAR FORECASTING MODELS FOR SPRING WHEAT (OTHER THAN DURUM)

BY

JACK NEALON

TABLE OF CONTENTS

ACKNOWLEDGEMENT

I wish to express my appreciation to Jim Stepanich of the North Dakota State Statistical Office for his fine supervision of the data collection and data preparation.

INTRODUCTION

The purpose of this research was to determine if reliable pre-harvest forecasts of spring wheat (other than durum) could be made using a method that relies solely on current year plant data. If reliable forecasts can be provided by this method, it could be used as supplemental information to the present forecasting system, which is dependent upon past years' data. Clearly, a method based solely on current year data would be very beneficial during atypical years since the present system would not reflect the unique crop characteristics of those years. During typical years, this method would provide an additional probability indication so that more reliable forecasts could be made.

Research in North Dakota during 1975 with a nonprobability sample of three fields indicated that growth and survival models may be useful for forecasting the yield of spring wheat (other than durum).[1/] A probability sample of 11 fields was selected from a crop reporting district in North Dakota during 1976 so that statistical inferences concerning this methodology could be stated.

The primary objective of the 1976 research was to determine which variables, if any, could be used to forecast yield reliably. To forecast yield per head, the yield variables observed were oven dried head weight, oven dried kernel weight, undried head weight and undried kernel weight and their associated time variables in terms of days since full head emergence or days since flowering. To forecast heads per acre, the survival variable

---

1/ Jack Nealon, Within-Year Spring Wheat Growth Models, Research and Development Branch, Research Division, Statistical Reporting Service, U. S. Department of Agriculture, Washington, D. C., January, 1976.

considered was the ratio of surviving heads to the total heads observed and the associated days since full head emergence or flowering. The secondary objective was to examine alternate sampling plans in an attempt to reduce survey expenses.

SAMPLE DESIGN

A systematic random sample of 11 fields was selected from the June Enumerative Survey's fields of spring wheat (other than durum) in North Dakota's east-central crop reporting district. Within each field, two units, each containing 50 tagged stalks, were randomly located. In all, there were 1,100 stalks in the sample.

DATA COLLECTION

There were three phases in the data collection effort. The first phase involved obtaining the occurrence dates of full head emergence and flowering for each of the 1,100 stalks. The fields were visited every 2 or 3 days from June 23 to July 9 in order to accurately determine these dates. The second phase commenced July 9 since at that time kernel formation had begun. Fields were visited once a week from July 9 until harvest, and six wheat heads were clipped weekly in each unit and mailed to the laboratory. The time variables, days since full head emergence and flowering, were now known since the occurrence dates and clipping date of each clipped head were known. Also, the survival variable, ratio of surviving heads to the total heads clipped, was available. The third phase of data collection took place in the laboratory. Measurements, counts and weighings before and after oven drying were performed each week on the clipped heads to obtain the yield variables.

DATA ANALYSES

Three models were used to determine the best variables for reliably forecasting yield. These models are shown below.

(1) Logistic Growth Model:

$$y_i = \frac{\alpha}{1 + \beta\rho^{t_i}} + u_i \ ;$$

$i = 1,2,\ldots,n$

α, β and ρ = parameters

$\alpha>0$, $\beta>0$, $0<\rho<1$

u_i = error term

t_i = time variable

y_i = yield variable

(2) Weighted Logistic Growth Model:

$$\frac{y_i}{\hat{\sigma}_i} = \frac{1}{\hat{\sigma}_i}\left[\frac{\alpha}{1 + \beta\rho^{t_i}} + u_i\right] \ ;$$

$i = 1,2,\ldots,n$

α, β and ρ = parameters

$\alpha>0$, $\beta>0$, $0<\rho<1$

$\hat{\sigma}_i$ = estimated standard deviation of y_i

$\frac{u_i}{\hat{\sigma}_i}$ = error term

t_i = time variable

$\frac{y_i}{\hat{\sigma}_i}$ = yield variable

(3) Logistic Survival Model:

$$y_i = \frac{\alpha}{1 + (\alpha-1)\rho^{t_i}} + u_i \ ;$$

$i = 1,2,\ldots,n$

α and ρ = parameters

$0 \leq \alpha \leq 1$, $0 < \rho < 1$

u_i = error term

t_i = time variable

y_i = survival variable

Model 1 or 2 was used to forecast yield per head, and model 3 provided a forecast of heads per acre. In conjunction, these models forecast yield per acre. A discussion of each of these models is given in an earlier report.[2/]

While discussing the results of the data analyses, reference will be made to tables in Appendix A and figures in Appendix B. Each table gives the model used, number of observations, the estimated sampling error for each parameter, the forecast and the forecast as a percentage of the estimated harvest weight for each cutoff date. The data were analyzed for each cutoff date to determine how early in the season after kernel formation reliable forecasts could be made. Each figure shows the data and estimated model for a cutoff date.

Is days since full head emergence or days since flowering a more reliable time variable?

Since the fields were selected from the June Enumerative Survey's fields, field visits to obtain the occurrence dates of full head emergence and flowering could not begin until the middle of June unless special procedures were set up in the state office to allow visits to begin in early June. This year, the growing season in North Dakota was earlier than usual and special procedures were not set up in the state office. Therefore, occurrence dates were missed for a portion of the stalks. Since an early season could occur again, flowering would be preferred over full head emgergence since flowering takes place later in the stalk development.

2/ Jack Nealon, The Development of Within-Year Forecasting Models for Winter Wheat, Research and Development Branch, Research Division, Statistical Reporting Service, U. S. Department of Agriculture, Washington, D. C., October, 1976.

Full head emergence or any phenological event prior to full head emergence would involve too great a risk of obtaining incomplete data and thereby preventing early developing fields from being accurately represented in the forecast. In summary, days since flowering is a more reliable time variable.

Which yield variable in conjunction with the preferred time variable, days since flowering, performs the best in forecasting yield per head?

The yield variables analyzed were undried head weight, undried kernel weight, dried head weight and dried kernel weight. The undried weights were obtained in order to determine if the oven drying process was necessary. The oven drying process involved drying the heads and kernels for 48 hours at a temperature of 140°F. If possible, it would be advantageous to eliminate this process so that data summarization could be completed more rapidly. Head weight was obtained for undried and dried heads to determine if the tedious task of kernel extraction for kernel weights could be eliminated. If kernel extraction is not necessary, the sample of heads processed in the laboratory could be tripled. In summary, from a practical viewpoint, the kernel extraction and oven drying processes should be avoided, if possible.

Figures 1 and 2 show the plots of undried head weight and undried kernel weight, respectively, with the time variable, days since flowering, when the entire season's data were used. These figures illustrate that the difference in moisture content during the growing season would produce inconsistent and unreliable forecasts of yield. Therefore, oven drying is imperative.

Tables 1 and 2 show the pertinent information about the estimated model for dried head weight and dried kernel weight,respectively, for each cutoff date. Notice that the logistic growth model (model 1) was used for dried head weight, and the weighted logistic growth model (model 2) was used for dried kernel weight. The weighted logistic growth model was required, because the assumption that the variance of dried kernel weight is equal for all values of time was violated.

Results from examining alternate weighted logistic growth models indicated that the most practical and probably most reliable method for weighting the data to correct the unequal variances for dried kernel weight was to weight the data by the inverse of the estimated standard deviation of dried kernel weight at each value of time since flowering. Since the sample in North Dakota was not large, an estimate of the standard deviation was generated for each 2-day interval of time since flowering. This method is statistically valid when the number of observations in each interval is large enough to provide a good estimate of the standard deviation.

Two factors favor the selection of dried head weight instead of dried kernel weight. First, as mentioned earlier, the number of heads processed in the laboratory could be tripled if dried head weight is selected since the tedious and time-consuming extraction of kernels would be avoided. Secondly, a weighted model, which is undesirable, may not be necessary for dried head weight.

Comparison of the estimated sampling errors for the three parameters in Tables 1 and 2 indicates that dried head weight performs better with respect to the most important parameter, α, which represents the forecast, and dried kernel weight performs better with respect to β and ρ. Therefore, since neither yield variable was superior and a much larger sample can be processed if dried head weight is selected, dried head weight is the preferred yield variable. Figures 3 through 7 show the data and estimated model for dried head weight and days since flowering for each cutoff date. The horizontal line represents the estimated harvest weight.

Is it possible to send fewer heads to the laboratory and still obtain an accurate representation of dried head weight?

A double sampling approach was used to determine if fewer heads could be sent to the laboratory. This approach involves the following steps:

(1) Instead of sending a large sample of heads to the laboratory, send a a smaller sample of heads.

(2) While clipping heads in the field on the smaller sample to be sent to the laboratory, cheaply obtain stalk characteristic data on a large sample.

(3) Determine if a strong relationship exists between dried head weight and the stalk characteristic data from the smaller sample.

(4) If a strong correlation is present and the smaller sample is large enough, use this correlation and the stalk characteristic data from the large sample to adjust dried head weight from the smaller sample to represent the large sample.

By this approach, laboratory work could be substantially decreased by making very inexpensive field observations on a stalk characteristic highly correlated with dried head weight.

The stalk characteristics observed in the field were the fertile spikelet count and head length. A fertile spikelet was defined as a spikelet that appeared to have kernel formation within it. The head measurement was made from the base of the lowest spikelet to the top of the highest spikelet.

Table 3 shows that the correlation coefficient of each stalk characteristic with dried head weight for 4-day intervals of time since flowering. Data were divided into 4-day intervals so that the number of observations for each correlation coefficient would not be too small. Fertile spikelet count and head length were each significantly correlated with dried head weight.

It is discouraging, however, that the correlations are not as good for small values of time since for small values of time the forecasts are of greater value. Nevertheless, the significant correlations warrant examining these stalk characteristics in future research.

Despite the fact that the correlations were highly significant for the smaller sample and the stalk characteristic data were available for the large sample, the double sampling refinement of dried head weight to represent the large sample was not made because the smaller sample was not large enough for each time interval. The double sampling approach, which utilizes a regression estimator, is biased, particularly for small samples. The bias in the regression estimator is reduced as the sample size becomes larger. In future research, the smaller sample should be large enough so that the bias is negligible and the refinement can be reliably made.

In summary, the significant correlations of fertile spikelet count and head length with dried head weight may make it possible to send fewer heads to the laboratory.

Can an accurate occurrence date of flowering for each stalk be obtained with fewer field visits?

Fields were visited every 2 or 3 days to accurately obtain the occurrence date of flowering for each of the 1,100 stalks. Visiting the fields with such frequency has two disadvantages: (1) If this forecasting methodology were operational, expenses would be very high. (2) The repeated field visits would increase the possibility of damage to part of the field. Therefore, to reduce expenses and the possibility of field damage, the data were examined as if weekly visits were made to the fields to obtain the occurrence dates.

Comparison of the estimated relative sampling errors for each parameter and the forecast between the methods shown in Table 4 clearly indicates that accurate occurrence dates can be obtained by visiting the fields once a week. Therefore, weekly visits should be sufficient.

Can heads per acre be reliably forecasted?

When units were located at the end of June an estimate of stalks per acre was obtained. This estimate was adjusted to heads per acre later in the season by assuming that stalks that did not flower would not produce any grain.

Enumerators classified the heads to be clipped either as surviving or dead when the weekly field clippings of heads began. Therefore, for each field on each weekly visit, the survival variable, the percentage of heads surviving, was obtained. The percentage of surviving heads and the associated time since flowering were then used in the logistic survival model to forecast percentage of surviving heads at harvest.

Table 5 illustrates that reliable forecasts of the percentage of surviving heads can be made. Each forecast of the percentage of surviving heads is multiplied by the earlier estimate of heads per acre in order to provide a reliable forecast of heads per acre.

CONCLUSIONS AND RECOMMENDATIONS

This research has shown that growth and survival models may be useful for forecasting the yield of spring wheat (other than durum). The earliest reliable forecast can be made 2 or 3 weeks after the beginning of kernel formation. Therefore, 3 or 4 weekly forecasts can be provided. To forecast yield per head, the best yield and time variables are dried head weight and days since flowering, respectively. In addition, reliable forecasts of heads per acre can be made.

The correlation of fertile spikelet count or head length with dried head weight may make it possible to send fewer heads to the laboratory. Also, it appears that fields need to be visited only once a week to obtain an accurate occurrence date of flowering for each stalk.

A statewide sample of fields is recommended for future research to fully test this forecasting methodology. With a statewide sample of about 50 fields, it could be determined if separate models are necessary for different varietal groups or different geographic locations. Also, a statewide sample would fully test the possibility of using the correlation between a stalk characteristic and dried head weight to reduce the number of heads to be processed in the laboratory. Alternate oven drying methods should be investigated. Finally, a method of converting dried head weight to dried kernel weight at standard 12% moisture must be designed.

COMPARISON WITH PREVIOUS RESEARCH

Research using this forecasting methodology was also conducted this year in Kansas with winter wheat. Since the research projects in North Dakota and Kansas were designed in the same manner, comparison of the results for winter wheat and spring wheat (other than durum) can be made.

Results for spring wheat showed that dried head weight was as reliable as dried kernel weight. These results and the fact that a much larger sample can be processed in the laboratory if dried head weight is chosen provide conclusive evidence that dried head weight is the preferred yield variable for spring wheat. Results with winter wheat did not clearly favor dried head or kernel weight. However, dried head weight was selected since a much larger sample can be processed in the laboratory with this yield variable. In summary, dried head weight was the preferred yield variable for winter and spring wheat.

For winter and spring wheat, days since flowering was preferred over days since full head emergence as the time variable. Not only does days since flowering provide better forecasting models, but also it is a more economical time variable since data collection can begin later in the growing season.

Reliable forecasts of heads per acre can be made for winter and spring wheat by means of the logistic survival model.

Fertile spikelet count and head length were each significantly correlated with dried head weight for winter and spring wheat. These correlations may make it possible to send fewer heads to the laboratory and thereby reduce expenses.

Finally, it was shown for winter and spring wheat that weekly field visits should be sufficient to accurately obtain the occurrence date of flowering for each stalk.

Table 1: Results from the logistic growth model for each cutoff date given that the yield variable is dried head weight (grams) and the time variable is time since flowering (days)

Cutoff Date	Model	Number of Observations	$\hat{\sigma}_{\hat{\alpha}}/\hat{\alpha}$ %	$\hat{\sigma}_{\hat{\beta}}/\hat{\beta}$ %	$\hat{\sigma}_{\hat{\rho}}/\hat{\rho}$ %	Forecast	% of estimated harvest weight
7/09/76	1	7	—	—	—	—	—
7/16/76	1	16	99.71	67.21	14.44	.787	92.3
7/23/76	1	27	29.03	26.20	4.72	1.035	121.3
7/30/76	1	38	9.00	36.40	4.08	.866	101.5
8/06/76	1	48	6.77	40.60	3.80	.890	104.3
8/13/76	1	52	5.24	45.33	4.03	.855	100.2

Table 2: Results from the weighted logistic growth model for each cutoff date given that the yield variable is dried kernel weight (grams) and the time variable is time since flowering (days)

Cutoff Date	Model	Number of Observations	$\hat{\sigma}_{\hat{\alpha}}/\hat{\alpha}$ %	$\hat{\sigma}_{\hat{\beta}}/\hat{\beta}$ %	$\hat{\sigma}_{\hat{\rho}}/\hat{\rho}$ %	Forecast	% of estimated harvest weight
7/09/76	2	7	—	—	—	—	—
7/16/76	2	16	—	—	—	—	—
7/23/76	2	27	28.64	24.61	4.34	.684	111.8
7/30/76	2	38	13.45	21.17	3.04	.678	110.8
8/06/76	2	48	7.71	29.52	3.43	.603	98.5
[illegible]							

Table 3: Correlation Coefficient of each stalk characteristic with dried head weight

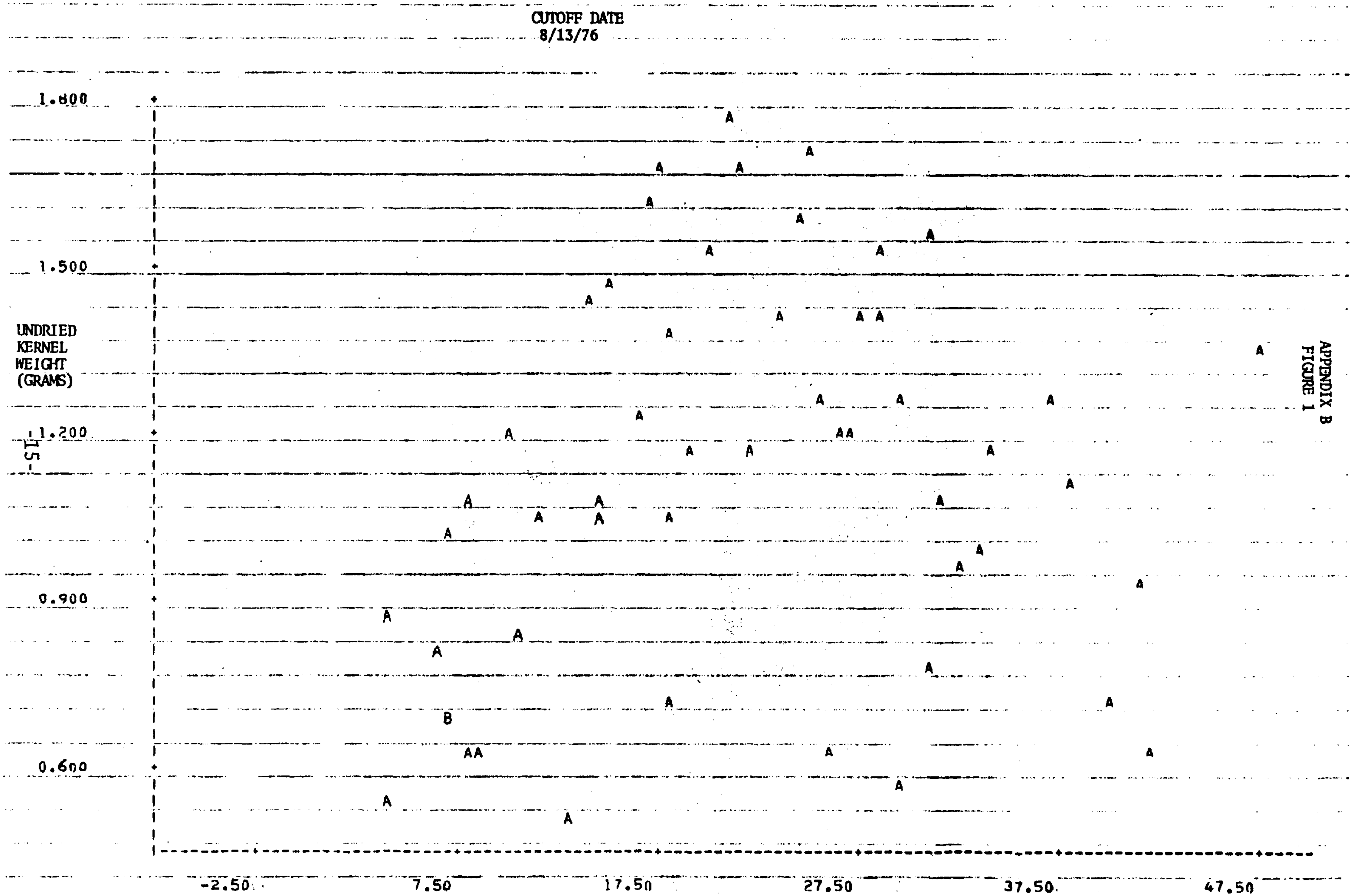
Time Interval	Number of Observations	Stalk Characteristic: Fertile Spikelet Count	Stalk Characteristic: Head Length
[0, 4)	19	.433	.656
[4, 8)	7	.789	.736
[8, 12)	37	.498	.643
[12, 16)	20	.580	.738
[16, 20)	42	.543	.560
[20, 24)	9	.811	.849
[24, 28)	47	.722	.731
[28, 32)	17	.774	.794
[32, 36)	25	.783	.787
≥ 36	36	.844	.759

Table 4: Results from the logistic growth model for the entire season's data for each method given that the yield variable is dried head weight (grams) and the time variable is time since flowering (days)

Method	Number of Observations	$\sigma_{\hat{\alpha}}/\hat{\alpha}$ %	$\sigma_{\hat{\beta}}/\hat{\beta}$ %	$\sigma_{\hat{\rho}}/\hat{\rho}$ %	Forecast	% of estimated harvest weight
Every 2 or 3 days	52	5.24	45.33	4.03	.855	100.2
Once a week	52	5.30	49.64	4.27	.855	100.2

Table 5: Results from the logistic survival model for each cutoff date given that the survival variable is $\left[\frac{\text{surviving heads}}{\text{total heads}}\right]$ and the time variable is time since flowering (days)

Cutoff Date	Model	Number of Observations	$\hat{\sigma}_{\hat{\alpha}}/\hat{\alpha}$ %	$\hat{\sigma}_{\hat{\rho}}/\hat{\rho}$ %	Forecast	% of estimated $\left[\frac{\text{surviving heads}}{\text{total heads}}\right]$
7/09/76	3	8	1.82	0.0	.975	98.8
7/16/76	3	19	0.88	0.0	.988	100.1
7/23/76	3	29	0.69	0.0	.987	100.0
7/30/76	3	41	0.50	0.0	.990	100.3
8/06/76	3	48	0.48	0.0	.989	100.2
8/13/76	3	52	0.48	0.0	.988	100.1



LEGEND: A = 1 OBS , B = 2 OBS

APPENDIX B
FIGURE 1

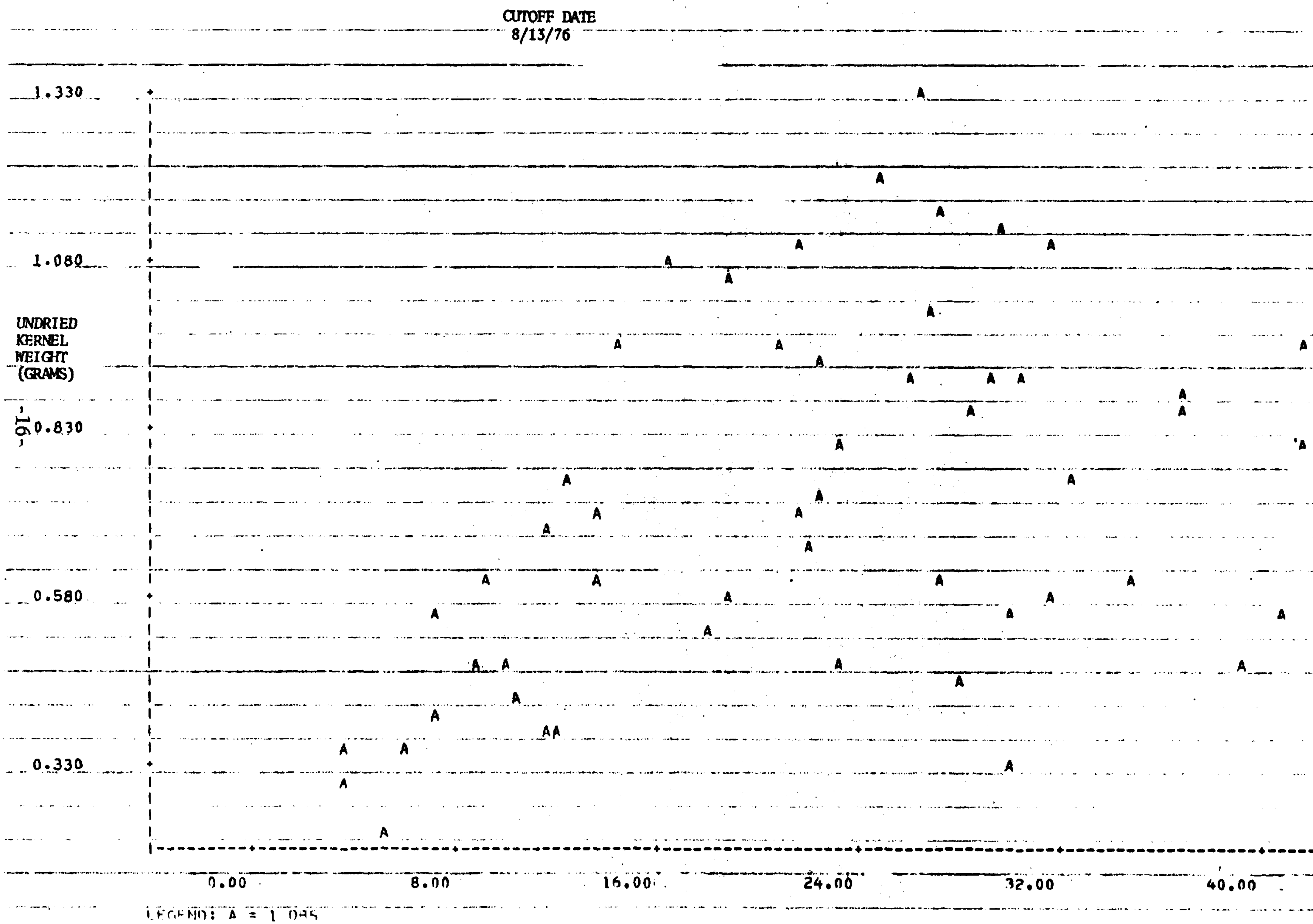

CUTOFF DATE
8/13/76
1.330
1.080
UNDRIED
KERNEL
WEIGHT
(GRAMS)
0.830
0.580
0.330
0.00
8.00
16.00
24.00
32.00
40.00
LEGEND: A = 1 OBS

FIGURE 2

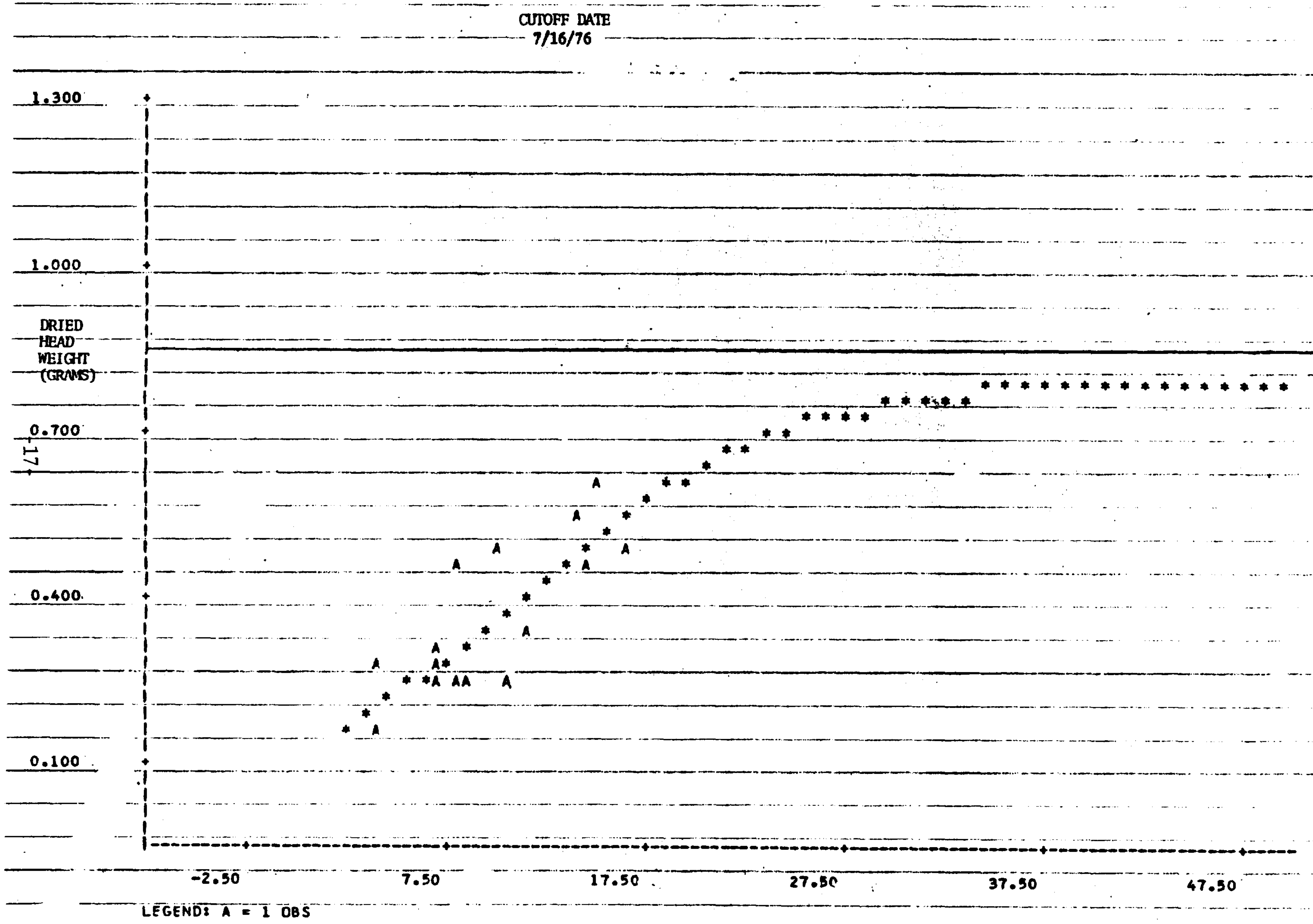

CUTOFF DATE
7/16/76
1.300
1.000
DRIED
HEAD
WEIGHT
(GRAMS)
0.700
0.400
0.100
-2.50
7.50
17.50
27.50
37.50
47.50
LEGEND: A = 1 OBS

FIGURE 3

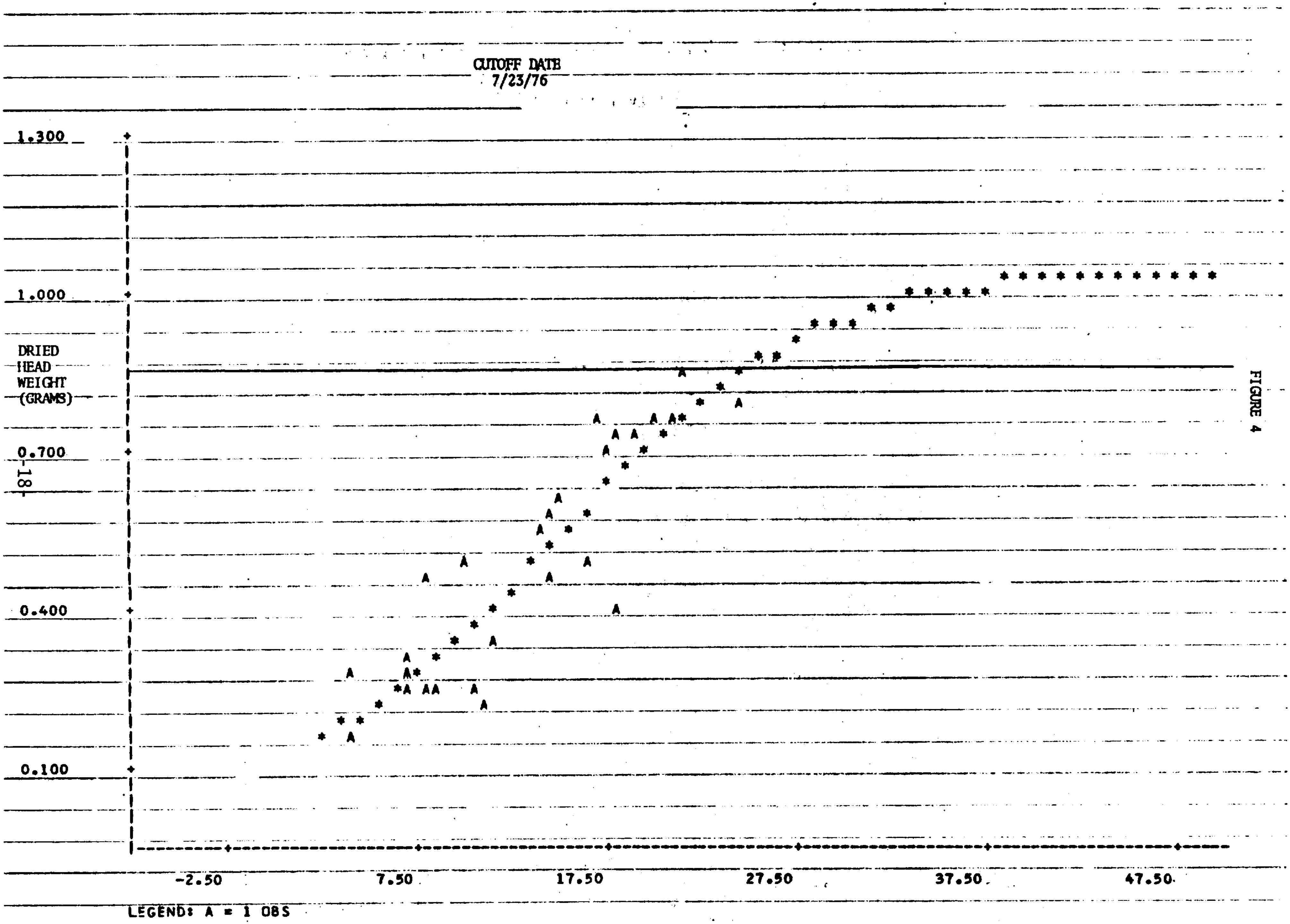

CUTOFF DATE
7/23/76
1.300
1.000
0.700
0.400
0.100
DRIED
HEAD
WEIGHT
(GRAMS)
-2.50
7.50
17.50
27.50
37.50
47.50
LEGEND: A = 1 OBS

FIGURE 4

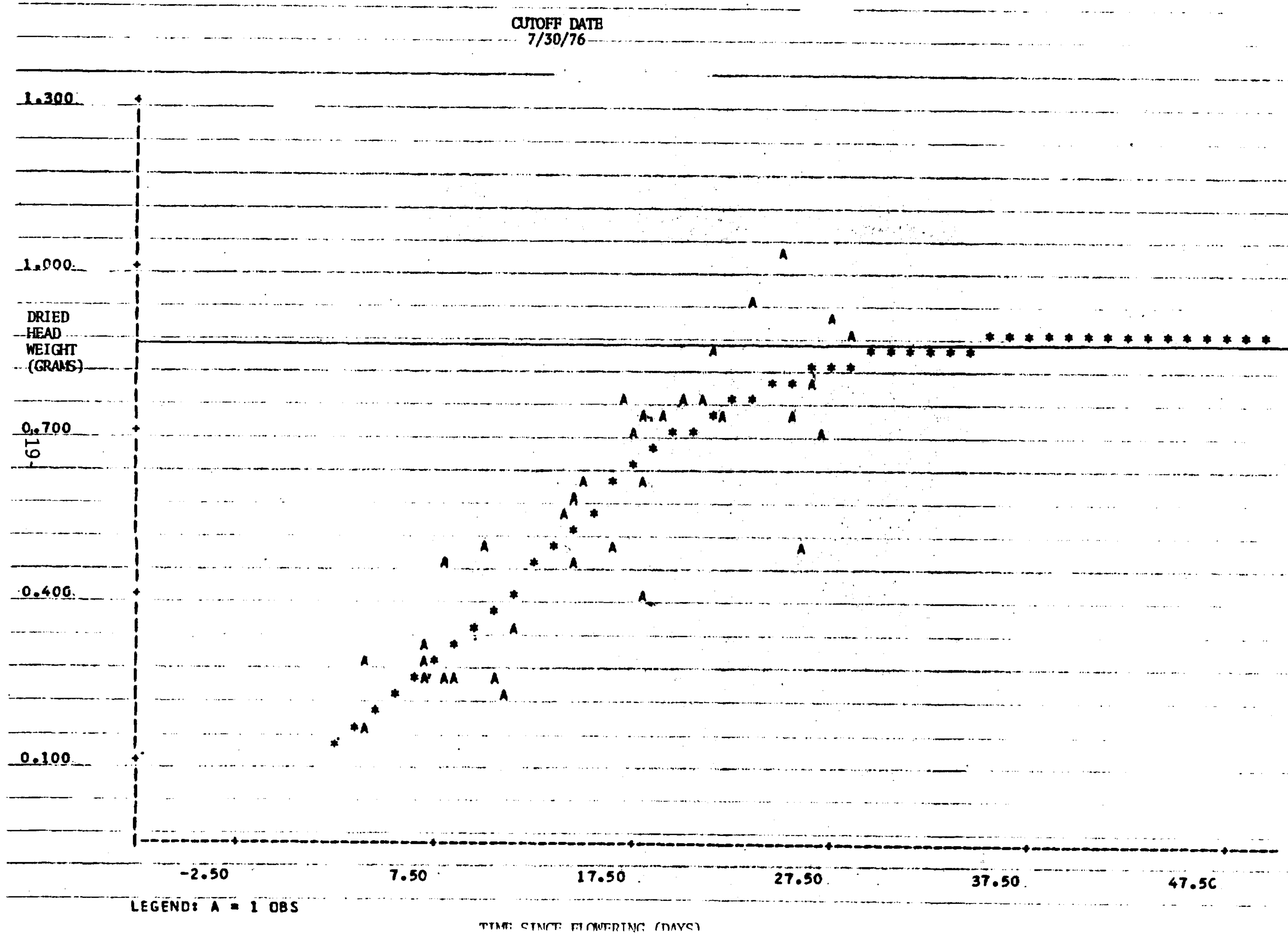

CUTOFF DATE
7/30/76
1.300
1.000
0.700
0.400
0.100
DRIED
HEAD
WEIGHT
(GRAMS)
-2.50
7.50
17.50
27.50
37.50
47.5C
LEGEND: A = 1 OBS
TIME SINCE FLOWERING (DAYS)

FIGURE 5

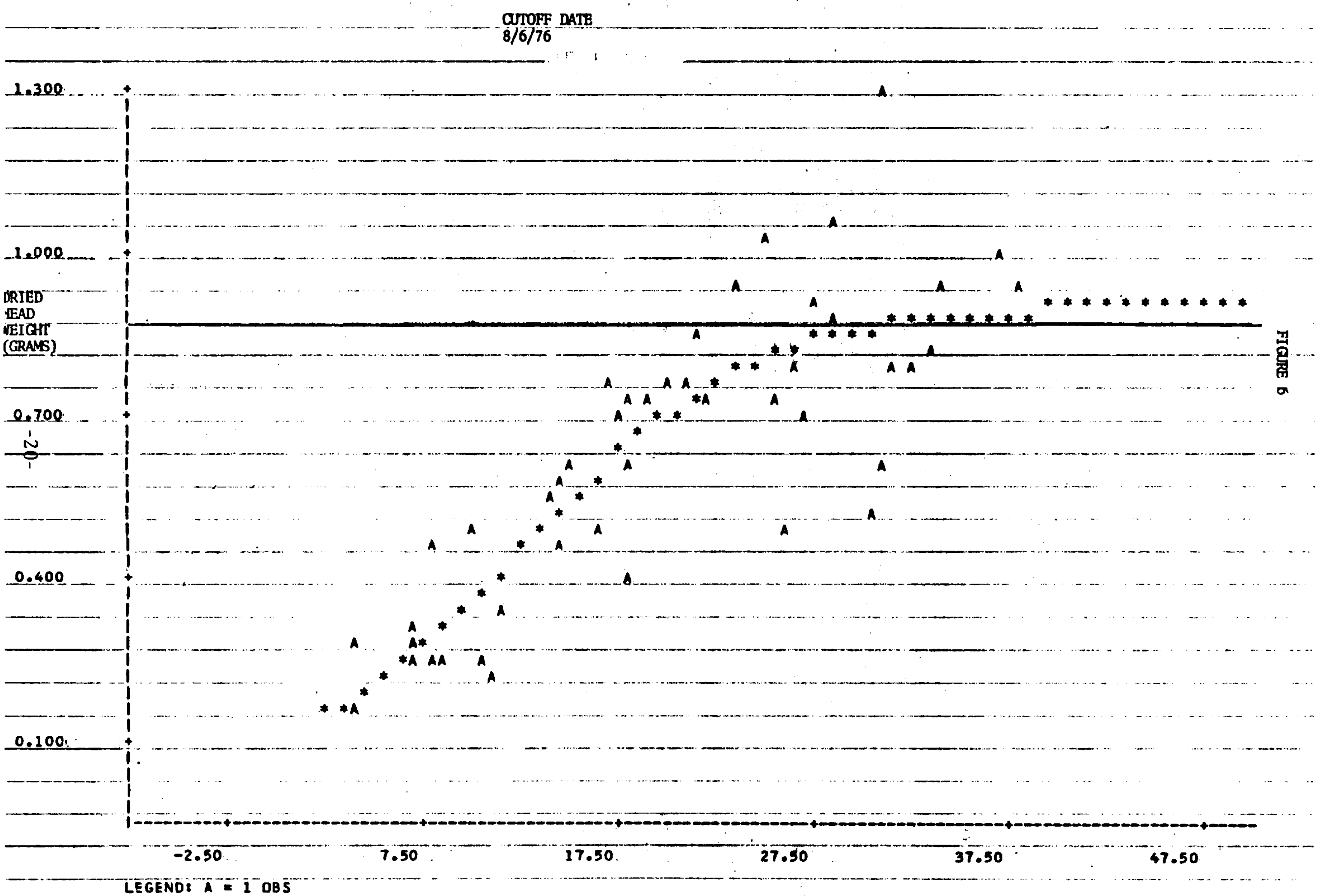

CUTOFF DATE
8/6/76
1.300
1.000
0.700
0.400
0.100
DRIED
HEAD
WEIGHT
(GRAMS)
-2.50
7.50
17.50
27.50
37.50
47.50
LEGEND: A = 1 OBS

FIGURE 6

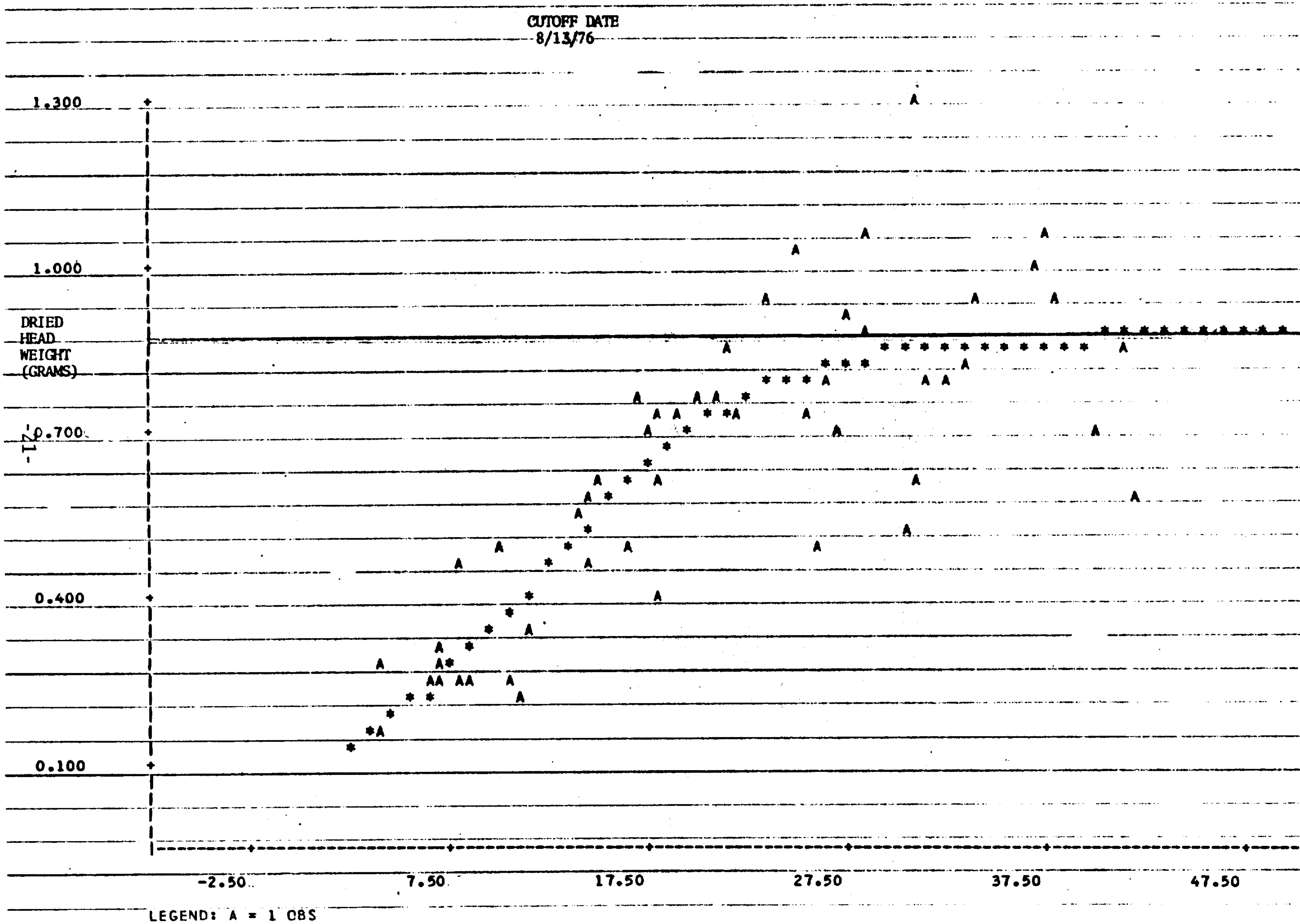

CUTOFF DATE
8/13/76
1.300
1.000
0.700
0.400
0.100
DRIED HEAD WEIGHT (GRAMS)
-2.50
7.50
17.50
27.50
37.50
47.50
LEGEND: A = 1 OBS

FIGURE 7