Some Current Research Efforts

To Improve Crop Forecasting Methods

by
Wendell W. Wilson
Statistical Reporting Service

This afternoon I will try to provide you with some insight into a new method of forecasting crop yields. This is an objective plant measurement method which is based only on current year data. I will talk primarily about corn because research on corn growth led to the development of this new method. We are currently developing a similar within year yield forecasting method for wheat. As I refer to the corn plant, you may wish to think of similar observations on any of a wide variety of other crops.

I will compare and contrast the regression models which we now use with models based only on current year data. My discussion will be somewhat critical of some shortcomings of our current forecast models so that you can see why we were interested in new methods. Our current methods are quite useful. We merely wish to improve our forecasting efforts as much as possible.

First let us look at the type of regression model currently used. (see page 5) These models may be thought of as over-the-years models because of the source of the information they utilize. Every model begins with the model form. This model form is based upon knowledge, experience, and previous research. It is a concept of the way things are. For corn, the dependent variable is the mature grain weight per ear. An independent variable used is ear length at some immature stage of development. In the model form, parameters alpha and beta are symbols with no specified value.

Parameter values are estimated based on a least squares fit of data from objective measurement field surveys in previous years. The X_i, Y_i pairs for the i^{th} sample field are fitted according to the model. This can, of course, only be accomplished after the Y_i's are available. That is, after the crop reaches maturity and sample ears are weighed. In practice we pool data over three previous years in fitting the model.

Presented at the Washington Statistical Society Meeting (ASA) on November 3, 1975

The independent variable value comes from a survey in the current crop year. Since this survey need not include observations of the dependent variable, it is conducted prior to crop maturity.

The most important thing to note is that once the dependent variable for the current year is obtained, a least squares fit of only current year data can be used to estimate alpha and beta for that year. Will these estimates be the same as the values of the parameters used in forecasting the yield? Highly unlikely! How different will they be? A more difficult question. One answer is that it will depend on how different the plant environment is in the three year base period as compared to the current year. The assumption that the parameter estimates in the base period and from the current year are members of the same "super population" is difficult to accept when the current year differs markedly from the model base period. Because of this we wished to develop forecasting methods which did not rely on previous years data.

As an example of the departures which may occur, let us look at a few graphs of over-the-years regression models. On page 6 is a model expressing the dependence of average mature ear weight on average ear length. This is a linear model based on pooled data from years 1, 2, and 3. The graph on page 7 shows two possible model fits for a fourth year. In a year such as 4A the pooled three year model will underestimate the average mature ear weight. If the fourth year was fitted by model 4B, then an overestimate would result.

The within year model approach begins by considering what we want to forecast. Yield per acre is an answer. Another is the components of yield; number of units and weight per unit. We choose to forecast weight per unit because of the body of knowledge about plant growth. The logistic growth model is theorized to apply to the growth process. The model reflects the dependence of dry matter deposition on time. What we want to forecast is the amount of dry matter in the grain at maturity. If we can associate dry matter with time, then hopefully, we can project from immature crop observations to maturity.

A graph of the logistic growth model is shown on page 8. An individual data point can be specified by its location (t_i, Y_i). The individual points are represented by the letters. The asterisks show the plot of the growth function. For a specified time, say $t = 35$, the functional value $\hat{Y}$ is approximately 100 grams. This graph and fit of the growth model are based upon data collected this year from August 4 through mid-September in a single state. The graphs on pages 9 and 10 are for two additional states included in a pilot research project this year.

Now let us look at the model form for the within year model (see page 11). Recall that the over-the-years model expressed the dependence of mature grain weight per ear on immature ear length measurements. This model form is also based upon knowledge, experience, and previous research. A good case can be made that this model is based on a high degree of knowledge about plant growth. It expresses the dependence of dry grain weight per plant at time t on time t. The independent time variable is days after silk emergence. As in the over-the-years model the parameter values are not specified. The only restriction on this intrinsically non-linear model is that all parameters are greater than zero and that rho (ρ) is also less than unity (1).

The parameter values are estimated based upon a non-linear least squares fit of available data (t_i, Y_i pairs) from a survey in the current crop year. By available data, I mean data which can be collected and summarized prior to a fixed forecasting date. For example, we can fit the model based on data collected from early August through September 1st, through September 15 and through October 1st. One of the advantages of this model is that previous information can be utilized at all cut-off dates and not thrown away.

The independent variable value for this model need not come from a survey. The value of t (say t_m) is a specified value or range of values that correspond to crop maturity. The ability to specify this value is based upon knowledge of crop development. Having fitted the model form to available data in the current year and performed the calculations using the mature time variable, we now have a forecast of average mature grain weight per plant ($\hat{Y}_m$).

Now I would like to summarize a few things about the model by referring to the graph of the function on page 12. Given the parameter estimates, let us see what the parameter values tell us. At time t = 49, the calculations tell us that the estimate of average dry grain weight per plant is a little more than 120 grams. At time zero $Y = \frac{\alpha}{1 + \beta}$ since rho raised to the zero power equals one ($\rho^0 = 1$). At time t very large ($t \to \infty$), $Y = \alpha$ since rho ($0 < \rho < 1$) raised to a large power times beta tends to zero. Thus, the parameter alpha can be thought of as the primary parameter. Alpha and beta together determine the initial level. Rho can be comprehended as a rate of growth parameter. If rho is substantially less than one, then a rapid rate of growth is indicated. Growth at a slow rate is indicated if rho is near 1.

Before I ask for your questions, there is one more item that needs to be mentioned. The growth model provides us with a forecast of the weight per unit. We also need to forecast the number of units at maturity. To do this, we first estimate the number of plants per acre early in the season. Then we use a survival model which expresses the dependence of a survival ratio on time (see page 13). This model is adapted from the growth model by imposing the additional constraint that for $t = 0$, Y must equal 1. This assumes that at time zero all plants have the potential of having grain at maturity.

Now let me briefly summarize the application of within year time related models. The estimated number of plants per acre early in the season is adjusted for survival based on the forecast survival ratio. This provides a forecast of the number of plants per acre with grain at maturity. This forecast is then multiplied times the forecast mature grain weight per plant to provide a forecast of final biological yield per acre.

Now, I would like to answer your questions about the within year models or perhaps questions concerning the data collection that goes with it.

MODEL FORM

$$Y_m = \alpha + \beta X + \epsilon$$

Based upon knowledge - Experience - Previous Research

PARAMETER VALUES

$$\hat{Y}_m = \hat{\alpha} + \hat{\beta} X$$

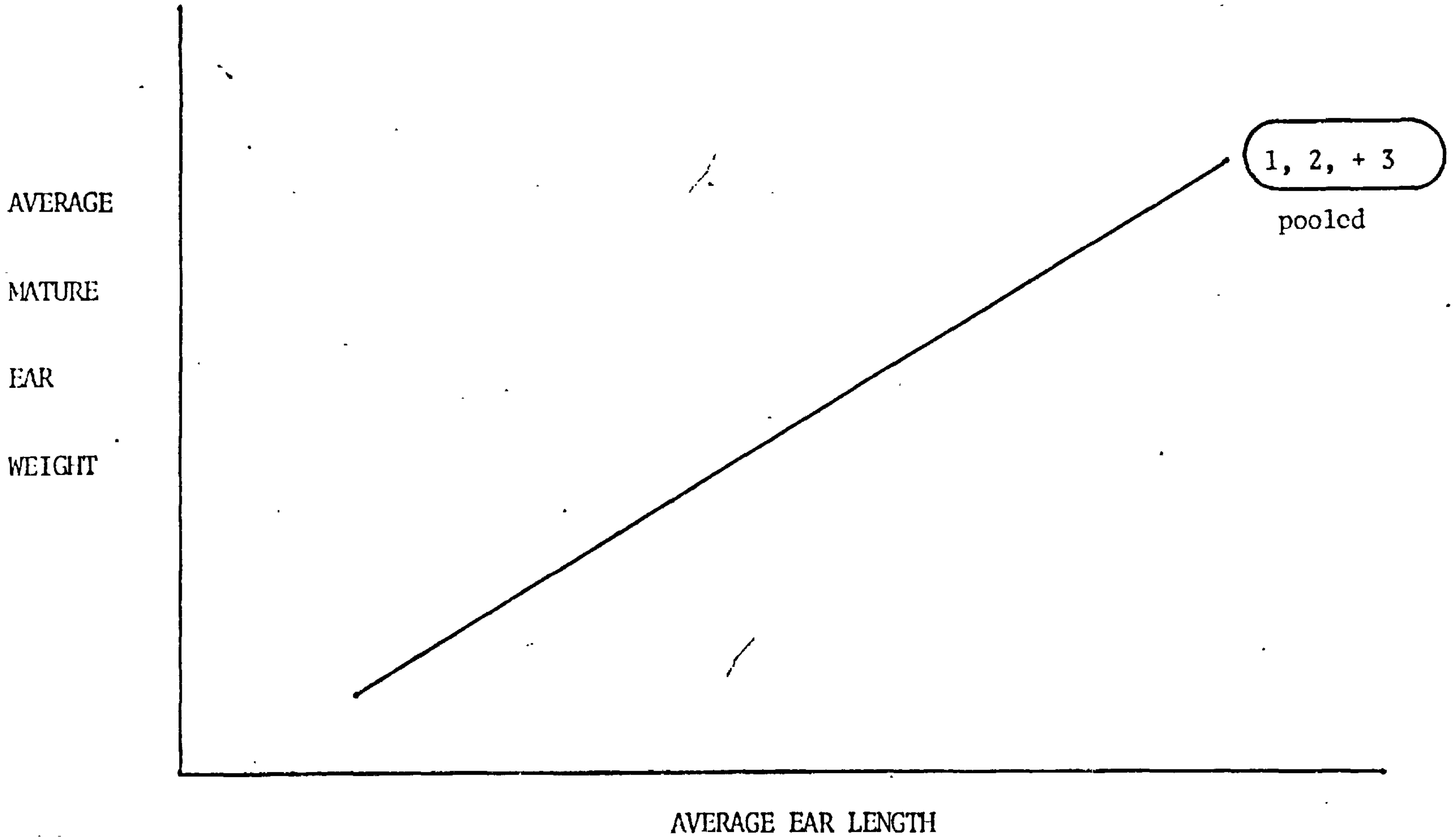
Based upon a least squares fit of data (X_i, Y_i pairs) from field surveys in previous years

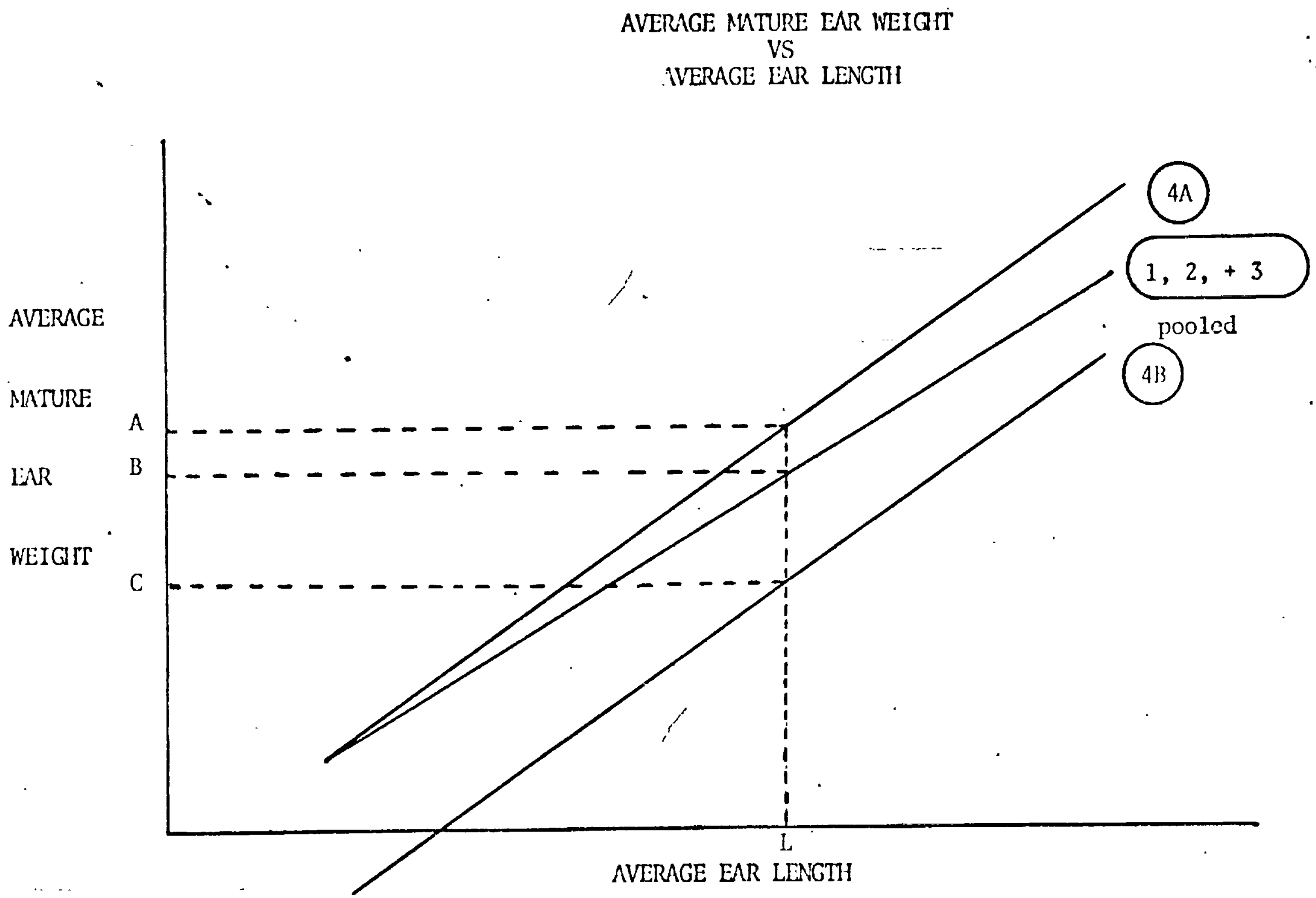
INDEPENDENT VARIABLE VALUE

$$\hat{Y}_m = \hat{\alpha} + \hat{\beta} X_c$$

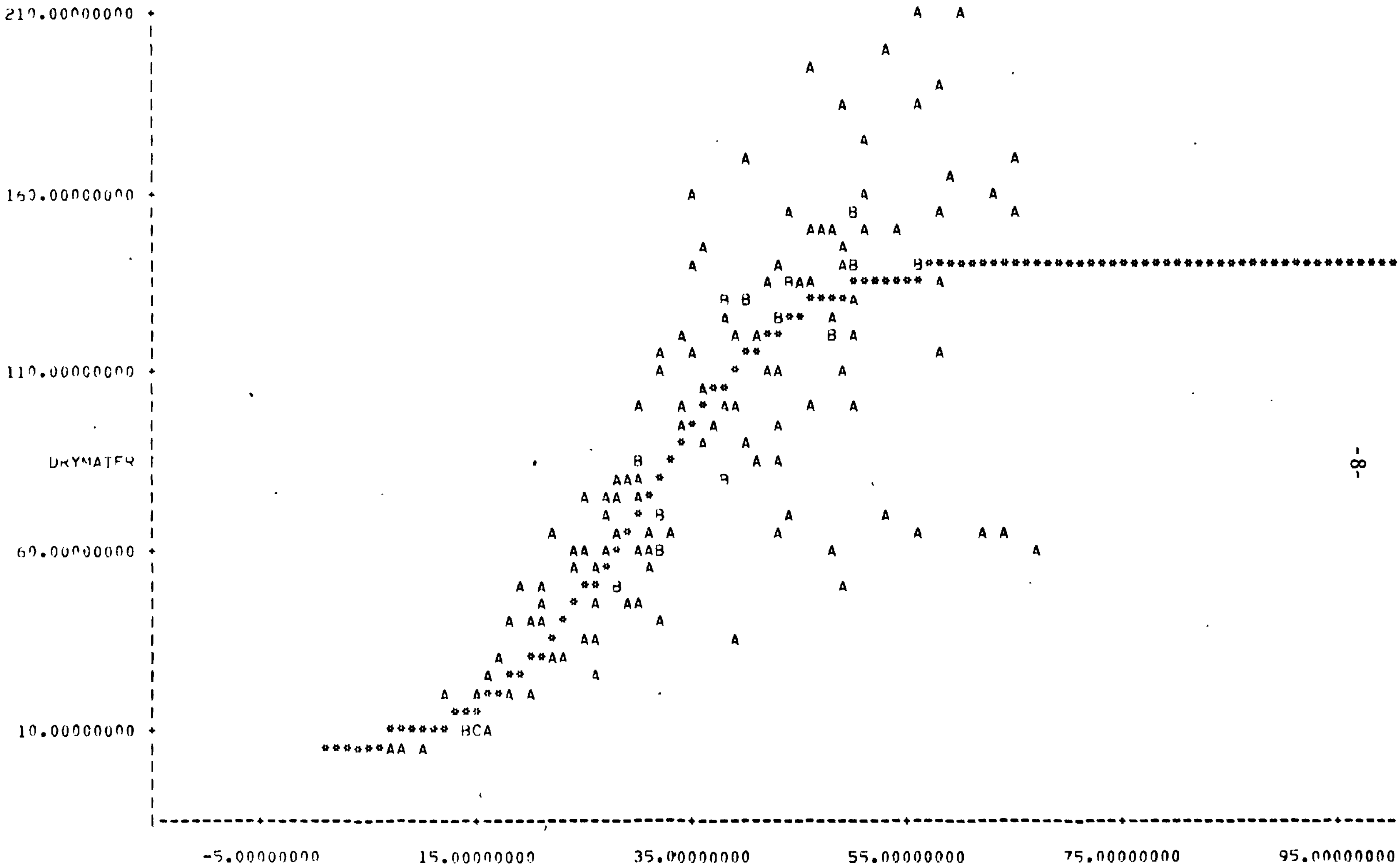
The X value (X_c) comes from a survey (conducted before crop maturity) in the current crop year

AVERAGE MATURE EAR WEIGHT
VS
AVERAGE EAR LENGTH



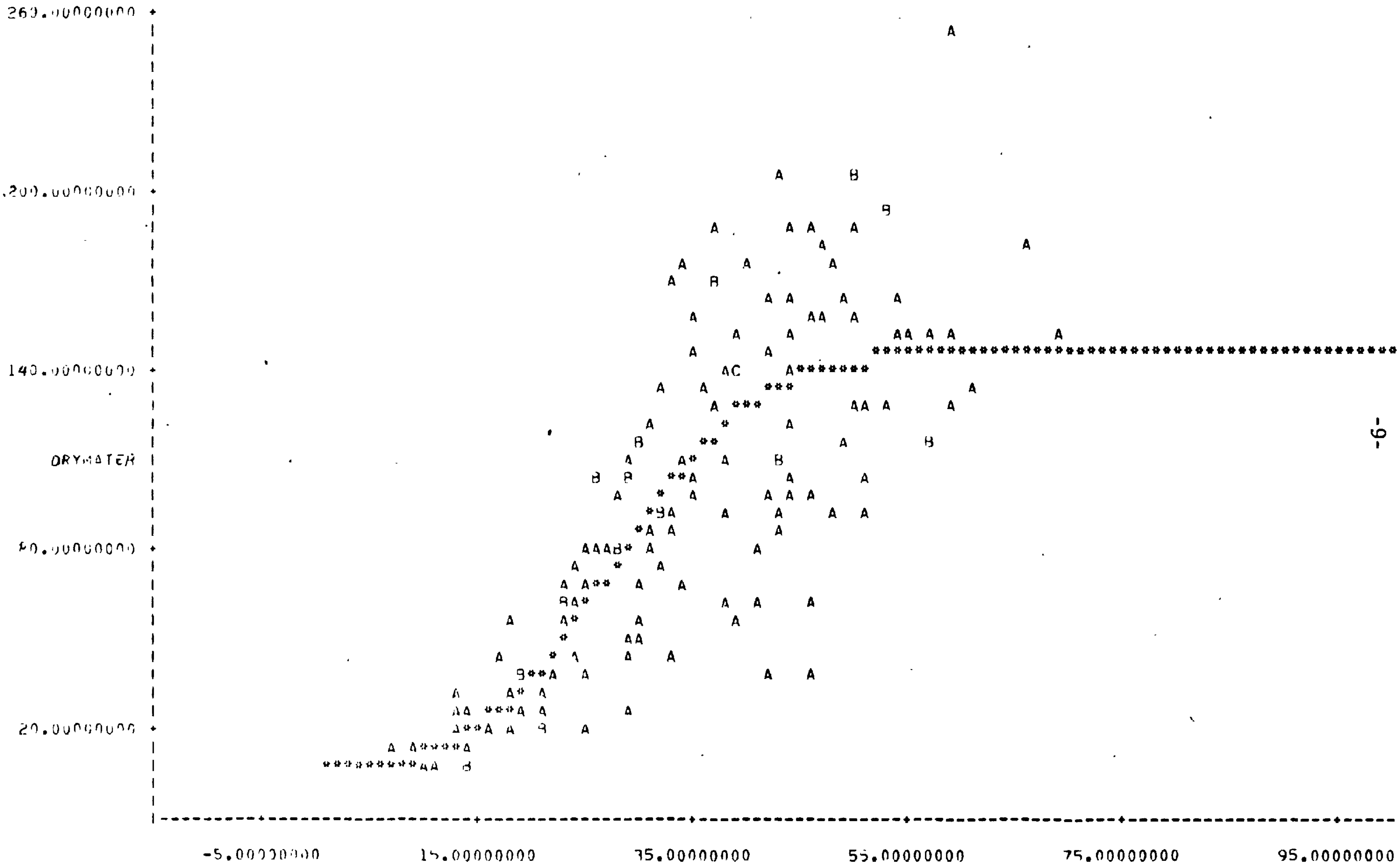

AVERAGE MATURE EAR WEIGHT
VS
AVERAGE EAR LENGTH
AVERAGE
MATURE
EAR
WEIGHT
A
B
C
4A
1, 2, + 3
pooled
4B
L
AVERAGE EAR LENGTH

PLOT OF DRYMATER VS TIME



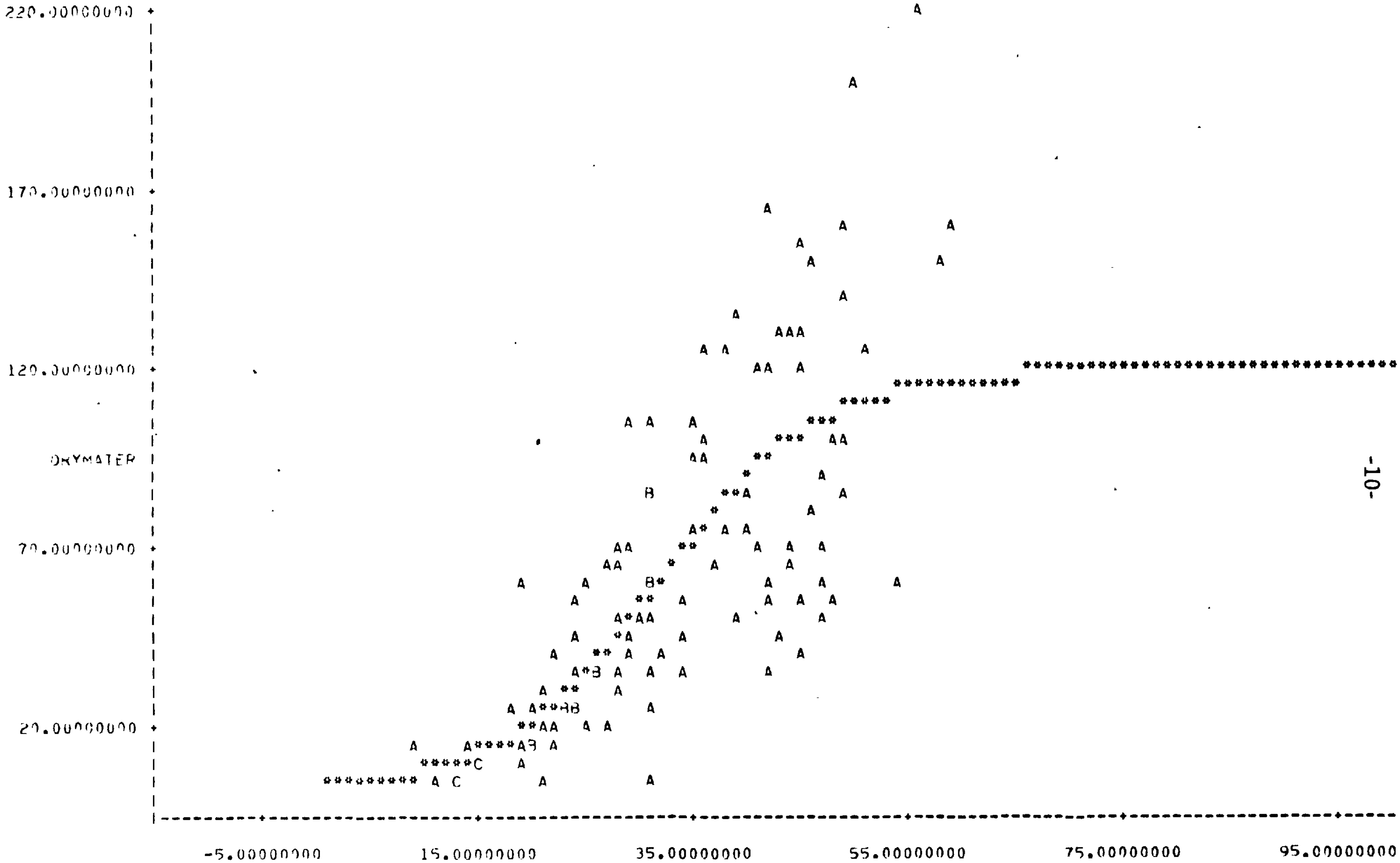
LEGEND: A = 1 OBS , B = 2 OBS , ETC.   TIME

PLOT OF DRYMATER VS TIME



LEGEND: A = 1 OBS , B = 2 OBS , ETC.

PLOT OF DRYMATER VS TIME



LEGEND: A = 1 OBS , B = 2 OBS , ETC.

TIME

MODEL FORM

$$Y = \frac{\alpha}{1 + \beta(\rho)^t} + \varepsilon$$

Based upon Knowledge - Experience - Previous Research

PARAMETER VALUES

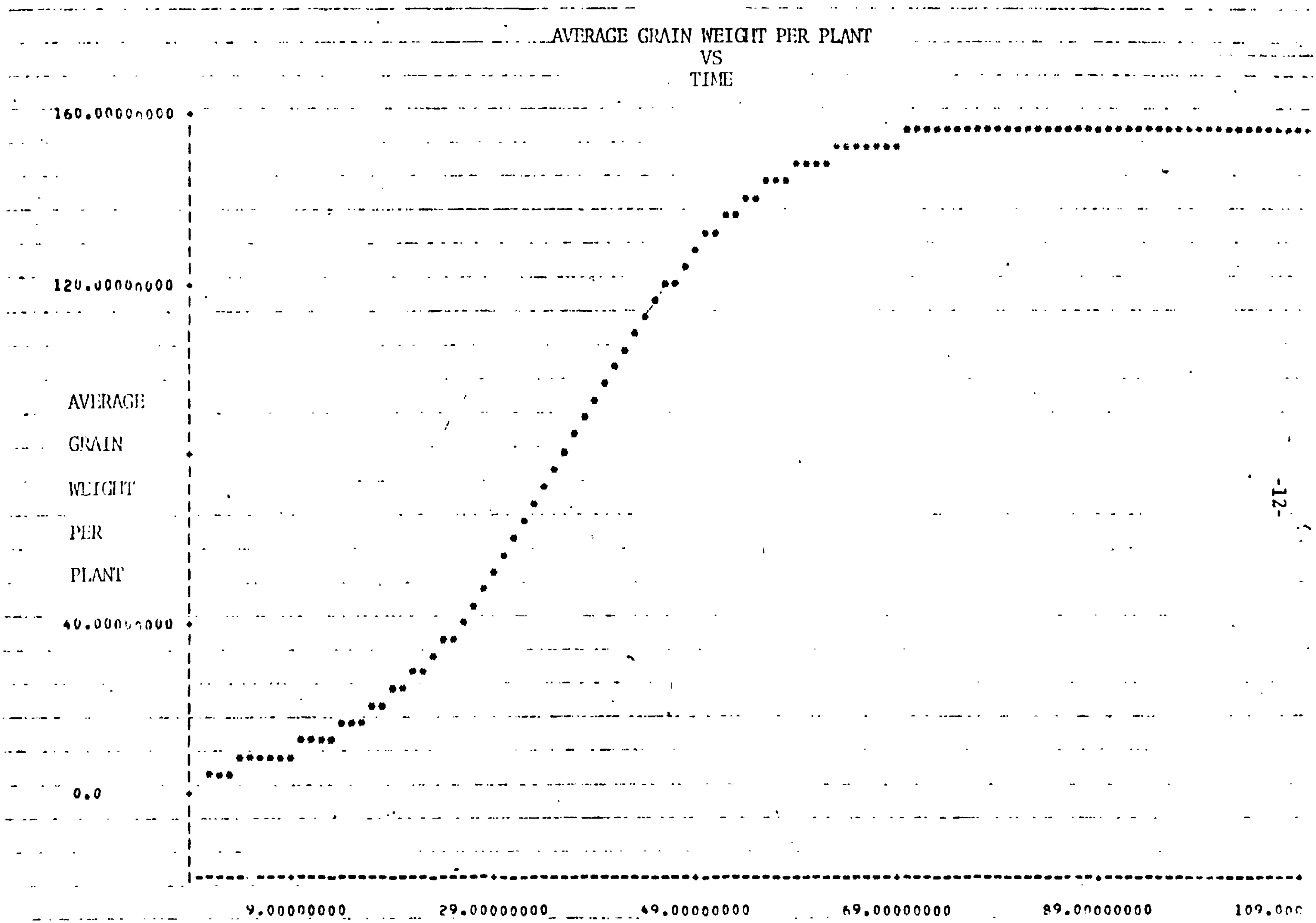
$$\hat{Y} = \frac{\hat{\alpha}}{1 + \hat{\beta}(\hat{\rho})^t}$$

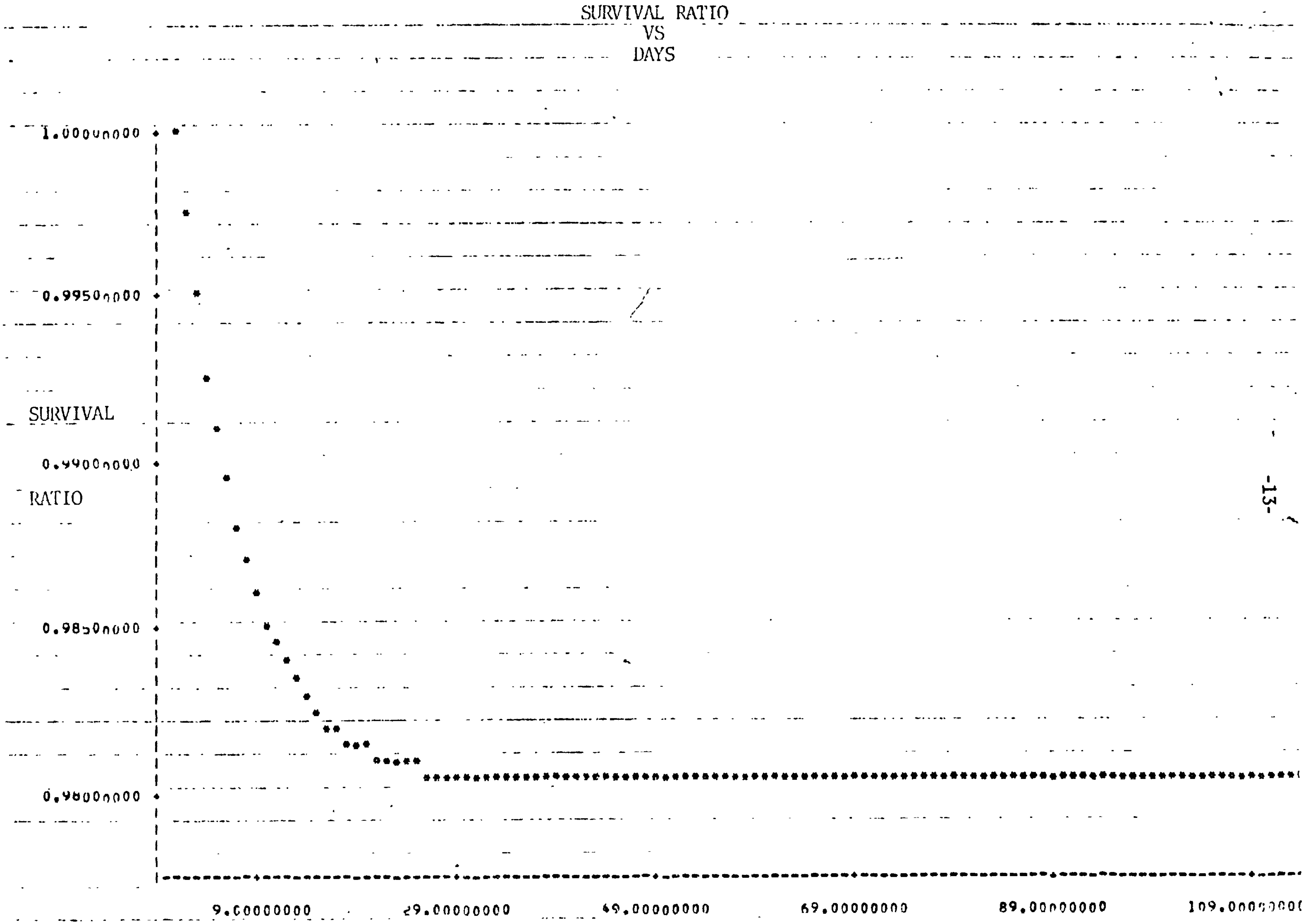
Based upon a non-linear regression least squares fit of available data (t_i, Y_i pairs) from a survey in the current crop year

INDEPENDENT VARIABLE VALUE

$$\hat{Y}_m = \frac{\hat{\alpha}}{1 + \hat{\beta}(\hat{\rho})^{t_m}}$$

The t value (t_m) is a specified value or range of values that correspond to crop maturity. Knowing this value is based upon knowledge of crop development. $\hat{Y}_m$ is the forecast for the response variable at maturity.


AVERAGE GRAIN WEIGHT PER PLANT
VS
TIME
160.00000000
120.00000000
40.00000000
0.0
AVERAGE
GRAIN
WEIGHT
PER
PLANT
9.00000000
29.00000000
49.00000000
69.00000000
89.00000000
109.000



SURVIVAL RATIO
VS
DAYS
1.00000000
0.99500000
SURVIVAL
RATIO
0.99000000
0.98500000
0.98000000
9.00000000
29.00000000
49.00000000
69.00000000
89.00000000
109.000