METC/CR-78/14

A RELATIONSHIP BETWEEN PRODUCTIVITY OF GAS WELLS AND THEIR LOCATIONS WITH RESPECT TO LINEAMENTS: A STATISTICAL ANALYSIS

BY

WAYNE E. ZIRK AND STEVE J. LAHODA

OCTOBER 1978

PREPARED FOR

UNITED STATES DEPARTMENT OF ENERGY
MORGANTOWN ENERGY TECHNOLOGY CENTER
MORGANTOWN, WEST VIRGINIA

Available from the National Technical Information Service, U. S. Department of Commerce, Springfield, Virginia 22161.

Price: Paper Copy $4.50
Microfiche $3.00

Printed in the United States of America
USDOE Technical Information Center, Oak Ridge, Tennessee

METC/CR-78/14
Distribution Category UC-92a

A RELATIONSHIP BETWEEN PRODUCTIVITY OF GAS WELLS AND THEIR LOCATIONS WITH RESPECT TO LINEAMENTS: A STATISTICAL ANALYSIS

September 1978

BY

Wayne E. Zirk and Steve J. Lahoda

of
The Department of Statistics and Computer Science
West Virginia University
Morgantown, West Virginia

Prepared for
United States Department of Energy
Morgantown Energy Technology Center
Morgantown, West Virginia

UNDER
Contract No. EY-77-C-21-8109

CONTENTS

TABLES

FIGURES

A RELATIONSHIP BETWEEN PRODUCTIVITY OF GAS WELLS AND THEIR LOCATIONS WITH RESPECT TO LINEAMENTS: A STATISTICAL ANALYSIS

by

Wayne E. Zirk[1/] and Steve J. Lahoda[2/]

ABSTRACT

The purpose of this paper is to relate the production of gas wells to their locations with respect to lineaments using production data made available by the Morgantown Energy Technology Center, Morgantown, West Virginia. Distance parameters are defined on the basis of other studies as well as new ideas. Regression analysis is utilized to construct a mathematical model which adequately represents this relationship for purposes of predicting productivity on the basis of the distance parameters. The model obtained was validated using data on additional wells from the same field.

INTRODUCTION

The objective of this study is to relate the production of gas wells to several parameters that are readily obtainable from the well's position relative to lineaments. Lineaments are conspicious linear features on the surface of the earth such as wrinkles, ridges,

---

1/ Lecturer of Statistics.
2/ Assistant Professor of Statistics.
Authors are with the Department of Statistics and Computer Science, West Virginia University; Morgantown, W. Va.

faults, or zones of intense jointing and/or fracturing. They are detected as any line on an aerial photograph that is structurally controlled including any alinement of separate photographic images such as stream beds, trees, or bushes that are so controlled. These lineaments are mapped from aerial and satellite photographs.

The importance of lineaments is based on the belief that photolinears identified as fracture traces and lineaments (which could represent faults in the earth's crust) may have some geologic significance and therefore, may affect production rates of petroleum or natural gas either increasing or decreasing flow capacity. If so, they could be used to give a better indication of where a well should be drilled to enhance the probability of achieving success.

Various studies have been conducted to investigate the possibility that lineaments might be useful in selecting well locations. The following two studies seem to indicate that lineaments have some bearing on gas well production. However, these studies are based on lineaments derived from side looking radar imagery whereas we will use lineaments derived from U-2 imagery. Thus, our conclusions may differ from theirs.

William M. Ryan (1976) related open flow measurements with lineaments derived from side looking radar imagery. In this study, the concern was with the average open flow after fracturing of wells drilled within major lineament zones versus average open flow after fracturing of wells drilled not in major lineament zones. Thirty-three wells were drilled within major lineament zones whereas forty-five were drilled at locations not in major lineament zones. Some of these were hydraulically fractured and some were not. Using all wells the average open flow for those within major lineament zones was 1866 Mcf/day whereas the average was 1042 Mcf/day for those not in major lineament zones. Of the wells which were hydrauically fractured, those on or within 1500 feet of major lineaments had an average open flow after fracturing of 1287 Mcf/day whereas wells more than 1500 feet had an average open flow after fracturing of 637 Mcf/day. No statistical analysis was given which would indicate whether or not these differences were significant; however, it was stated by the author that he feels better wells are associated with natural fractured zones, i.e. lineaments.

Overbey, Sawyer, and Henniger (1974) attempted to show a positive correlation between lineaments and good storage wells. They hypothesized that most of the lineaments were vertical rather than inclined and that the improvement of gas production may be maximized at the intersection of two or more linements. Well sites were selected at the intersection of lineaments and at locations near to lineaments but not at points of intersection. Information was obtained on open flow-rates for the storage wells and projected initial open flow rates at an assumed initial reservoir pressure of 1,100 psi. Both positive and negative cases were found in the comparison. In one positive case, the well with the largest projected initial open flow was located near the intersection of two lineaments. In a negative case, five wells located on a lineament had lower initial open flows than nearby wells not located on this lineament.

Because these results were inconclusive further research was conducted. In this open flow potentials of 30 wells located on or within 250 feet of a lineament were compared with the open flow potentials of 30 wells selected at random. A test of open flow potential of wells showed no significant difference between the test group and the control group. Further, no relationship was established between open flows and the lineaments lengths, or the lineaments orientation, or the intersection of two or more lineaments.

Although we feel that the two measurements (being within a certain distance of a lineament or intersection of two or more lineaments) used in the previous studies are important, we also feel that considerable information could be gained if one were to take into account the actual numerical values for these two and use more powerful statistical techniques. Therefore, this study will take a different approach. We will attempt to define additional measurements that will reveal more information concerning a wells' location with respect to these lineaments and we will use regression techniques.

The remainder of this paper will discuss the topics of Data Collection, Analysis of the Data, and Summary. Within the section on Data Collection there will be a discussion on where the data came from, a brief description of the data given, as well as reasons for deletion of some of the data. Further, emphasis will be given to why and how distances of lineaments were chosen since this pertains to the major objective of the analysis. Finally, included in this section will be a discussion of the open-flow parameters. The Statistical Analysis of this data will encompass the analysis of the data section. Here the approach and method used for the objective will be discussed. It will also consider the use of the results obtained from the Statistical Analysis. The final section will summarize this study and give some indication of what might be done in future studies.

DATA COLLECTION

The data to be used in the study consisted of "shot" shale well data on 75 wells located in Lincoln County and Wayne County of West Virginia. The following measurements were provided for each of the 75 wells:

IOFBS: Initial Open Flow Before Shot (Mcf/d),
IOFAS: Initial Open Flow After Shot (Mcf/d),
IRPAS: Initial Rock Pressure After Shot (psig),
TD: Total Depth (ft.),
TOSI: Top of short interval (ft.),
QOEU: Quantity of explosives used (lbs.),
YC: Year completed.

Actual production figures were given only for 40 wells however. The decision was then made to use only the data from these 40 wells in the analysis and to use any subsequently obtained data to verify results.

The production data covered a period of twenty years from the initial opening of a well. For each year two measures of production were given. One was the yearly production (Mcf) and the other was the average daily production (Mcf/d). In addition the cumulative production (Mcf) over the 20 year period was given for each well.

Examination of the production data revealed several poi nts that would be pertinent with respect to the analysis. First, it was found that the number of days a well was open each year was not g ven and that the number of days on which the average daily production was based could not be obtained either directly or by using the total yearly production and average daily production measurements. Secondly, it was found that several wells were opened less than twenty years and consequently the cumulative production figures would not be compatible. Thirdly, there was clear indication that the production data would not be consistent from year to year and even that measurements within a year might be inconsistent due to variations in the period for which a well was opened during a given year. That is, one well may have been opened for several periods during a year whereas another well may have been opened for one continuous period during that same year.

On the basis of this examination of the data it was decided that average daily production would be the most satisfactory of the available measures from a statistical standpoint. Discussions with geologists associated with the study indicated that the first few production years (years 1-5) were most important in attempting to assess the production capability of a well. On comparing the first five years of average daily production, it was decided that the manner in which the data were collected was probably more consistent for the first year than for the remaining years. For this reason, it seemed most appropriate to base the analysis on first year average daily production (FYADP).

As the major objective of the analysis would be concerned with the position of the wells with respect to the lineaments it was next necessary to define and obtain measures of the position of a well with respect to the lineaments. On the basis of the aforementioned studies, two obvious measures would be the distance to the nearest lineament and the distance to nearest intersection of two or more lineaments. These, however, do not adequately take into account that production may be affected by a cluster of lineaments. To account for this two additional measurements were defined as follows:

First, circular areas about each well were laid out, such that their perimeters determine boundaries. Lineaments (or any part thereof) falling within these boundaries were hypothesized to have an effect upon production and lineaments outside of these boundaries do not. Lineament falling within this circular area would be measured for length and accumulated. This figure in turn would be related to the wells production. This measure alone does not take into account whether the lineaments which intersect the circular area are long or short. To account for this a second measure was defined. The length of all lineaments which intersected the circular area would be accumulated and this figure would be related to production.

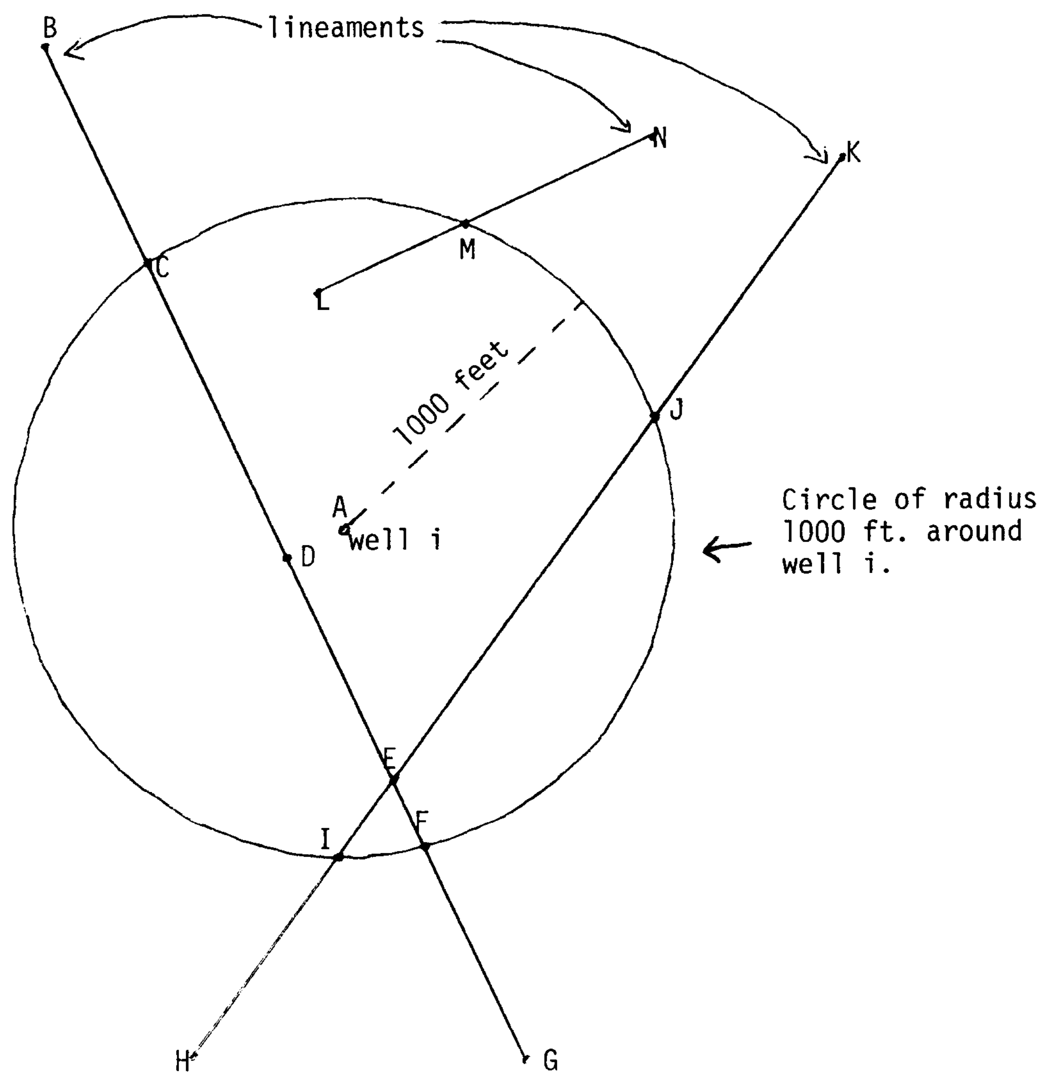


Figure 1: Four Distance Parameters

Description of four variables:

DTL: Distance to nearest lineament from well i. In the illustration above this would be the distance from point A to point D.

DTI: Distance to nearest intersection of two or more lineaments from well i. For this example, it would be the distance from point A to point E.

CLWC: Cumulative length of lineaments within circle. In above example this would be the distance between points C and F plus the distance between points I and J plus the distance between points L and M.

CLBC: Cumulative length of lineaments beyond circle. For the example above this would be the distance between points B and G plus the distance between points H and K plus the distance between points L and N.

An example shown in Figure 1 should help illustrate the meaning of each measure with respect to well i. Because this is only an example to illustrate the meaning of the four distance parameters, other situations may occur. That is, it may be the case that the intersection occurs outside the circular area. Furthermore, it is possible for no lineaments to intersect the circular area.

Geologists associated with the study initially suggested that the radius of the circular area should be from 200 to 500 feet. Further discussion led to the decision that these might not be appropriate limits since hydraulic fracturing could vary the wells drawing radius above, below, and in-between them. Also considered in the discussion was the distance of 1500 feet used by Ryan to classify wells into two groups. Finally, considered was the fact that if the radius were too small, the measurements CLWC and CLBC would have had the value zero (since no lineaments would intersect the circular area) for several wells. This would have yielded no information with respect to these two measurements for the wells associated with them. If the radius were too large considerable overlapping of circular areas would have occurred between wells. This would have resulted in a common value for two wells for which this occurred. With respect to this discussion a length of 1000 feet was chosen for the radius of the circular area. By using the radius of 1000 feet little, if any, overlapping occurred and almost all circles had at least one intersecting lineament.

In order to obtain the distances associated with the four lineament measures, a scale (1 inch = 2000 feet) map was used. This map contained the positions for each of the 75 wells and the lineaments associated with the study area were superimposed on it. The circular areas discussed previously were drawn by compass around each of the wells. Throughout the process several checks were made to assure the circular areas maintained the same diameter for each of the wells. To obtain the distances, a COMP-U-GRID digitizer was utilized. This machine allows one to obtain distances of straight line segments, areas of rectangles, areas under a curve, and coefficients for simple linear regression. To find, for example, the distance of a line segment using this machine it is first necessary to "digitize" some reference point on the map. That is, the machine consists of a movable "cross-hair" piece on the map. So it is necessary to enter coordinates associated with any point on the map whereby this point will be used in referencing coordinates of other points. So by using the coordinates of any two points for which the distance of a line segment is needed, the machine calculates this distance. Calculating these distances for DTL and DTI were fairly trivial since they only required the distance of one line segment each. However, this procedure was fairly tedious when obtaining distances for CLWC and CLBC since each of these usually required the addition of several line segment lengths. The machine would not do these additions, therefore, it was necessary to use a hand calculator to aid in this process.

A list of the data utilized in the analysis is located in Table 1. Although the data are real data we have concealed the individual identity of each well to protect confidentiality.

TABLE I: RAW DATA USED IN ANALYSIS

UNITS: CLWC(FT) CLBC(FT) DTL(FT) DTI(FT) IOFBS(MCF/D)

IOFAS(MCF/D) IRPAS(PSIG) FYADP(MCF/D)

NOTE: A DOT REPRESENTS A MISSING VALUE

OBS	WELL_NO	CLWC	CLBC	DTL	DTI	FYADP
1	1	0	0	1640	2200	83
2	2	960	8100	580	2120	18
3	3	2360	46140	40	2020	27
4	4	1900	62540	380	1460	540
5	5	2320	19160	620	1580	.
6	6	1740	65940	840	2100	94
7	7	1360	62580	420	1720	157
8	8	2420	62080	600	1760	.
9	9	7140	77850	220	460	.
10	10	1880	28260	420	2300	.
11	11	0	0	1220	2160	.
12	12	1940	32860	160	1320	.
13	13	3720	24300	260	520	160
14	14	1560	34220	520	2100	.
15	15	1620	10220	640	2180	.
16	16	3400	75340	180	740	.
17	17	1540	62500	640	2100	.
18	18	0	0	1240	1720	.
19	19	1040	4940	100	1320	285
20	20	5400	49160	140	1020	36
21	21	1540	18600	660	1400	162
22	22	0	0	1020	1380	27
23	23	1880	55080	440	1540	.
24	24	4280	54880	520	700	72
25	25	2560	84460	320	1100	.
26	26	1960	8680	180	1360	44
27	27	0	0	1020	2080	.
28	28	360	10980	540	1700	.
29	29	5160	57200	400	560	136
30	30	2700	93580	80	980	.
31	31	1500	25900	700	1980	.
32	32	7300	72400	0	520	.
33	33	1080	14080	700	2640	.
34	34	5340	142680	30	480	139
35	35	5000	47480	620	920	.
36	36	5540	94020	60	120	.
37	37	6140	99220	60	180	29
38	38	1820	16700	520	1140	.
39	39	3860	59980	300	600	.
40	40	680	62480	920	1200	.
41	41	5740	97320	60	380	.
42	42	2000	11680	120	1720	15
43	43	0	0	1240	2260	108
44	44	0	0	1360	2540	.
45	45	3420	30500	60	980	90
46	46	1920	93540	840	1060	.
47	47	1980	53300	300	1320	.
48	48	4620	23640	100	1240	67
49	49	3600	115280	580	680	.
50	50	1380	69740	740	1500	110
51	51	0	0	1340	1880	170
52	52	1800	10180	430	2240	.
53	53	0	0	1000	2060	.
54	54	3960	43380.	140	160	.
55	55	3020	69760	700	800	.
56	56	1980	9060	240	1460	.
57	57	1220	26240	860	1180	25
58	58	2840	37900	400	1120	60
59	59	4120	48040	120	980	25
60	60	1800	19520	520	4040	94
61	61	2540	18480	700	1360	.
62	62	780	14100	960	2120	.
63	63	4000	65940	40	340	77
64	64	3960	85680	120	140	.
65	65	2000	31180	260	1520	41
66	66	3120	59960	380	520	90
67	67	4400	53360	240	340	.
68	68	3140	80100	400	480	210
69	69	1940	14620	360	1940	30
70	70	2420	72880	560	1860	.
71	71	2000	34200	0	1000	.
72	72	5900	102280	100	280	.
73	73	1880	19520	420	5120	62
74	74	140	18120	980	1860	.
75	75	1220	18160	760	1740	.

ANALYSIS OF THE DATA

To accomplish the objective of the study it was decided to use regression techniques to build a mathematical model relating production to the distance variables. The primary benefit of this approach over those used in previous studies is that it would more adequately take into consideration the complexity of the problem. There is little reason to believe that the manner in which the productivity of a well might be related to its position with respect to lineaments will be such that one can attempt its description using simpler techniques. In addition, the development of a mathematical model would give the capability of predicting the productivity of wells in a quantitive manner.

In building the model simple and multiple linear regression models were first fitted to the data. As anticipated these yielded low R-square values and examination of residual plots gave clear indication that the use of transformations on the variables would be mandatory if a model was to be obtained that would adequately describe the relation between productivity and the distance variables. Primary among the needed transformations was one on the dependent variable that would tend to correct for non-homogeneity of variance. The most satisfactory turned out to be one in which FYADP was replaced with $-(FYADP)^{-1/2}$. Since this is a one-one transformation of FYADP, the resulting variable will serve equally well as a measure of productivity and was therefore taken as our measure of productivity for the remainder of the study.

To accomplish the building of the mathematical model, extensive use was made of stepwise regression techniques and procedures based on the examination of residual plots. In this the primary difficulty was that the sparsity of available data tended to conceal relationships that might otherwise have been readily discovered from the residual plots. In addition, considerable consternation was caused by the presence of a possible outlier among the production figures. For well number 4, the first year average daily production figure is 540 which far exceeds the next largest value of 285. Although it may differ considerably from the other values in this study, there is some evidence from other studies that it is a plausible value. Therefore, we chose not to omit it from the analysis.

The prediction equation that resulted is as follows:

MODEL 1: PRODUCTIVITY AS A FUNCTION OF DISTANCE PARAMETERS

MODEL:

$$
\begin{aligned}
Y = {} & 1.5178 - 14.3296\,(CLWC)^{2/3} \\
& + 1.9971\,(CLBC) + 1.5359\,(CLBC)^2 - 17.2413\,(DTL)^{1/2} \\
& - 31.2600\,(DTI)^{2/3} - 252.6426\,(CLWC * CLBC) \\
& + 19356.7539\,(DTL * DTI) + 4858.7507\,(CLWC * DTI) \\
& - 5040.2279\,[(CLWC)^2 * (CLBC)] \\
& - 68972.4280\,[(DTL)^2 * (CLBC)] \\
& + 2075.7096\,[(CLWC) * (DTL)^{2/3}] \\
& - 4.8483966 * 10^{23}\,[(DTL)^6 * (DTI)^6]
\end{aligned}
$$

The value of R-SQUARE was 0.842. This indicates that approximately 84 percent of the variation is explained by the regression equation. The magnitude of R-SQUARE depends on the number of variables and the number of observations. For a fixed number of observations one can always increase R-SQUARE by adding more variables even though those variables may not add significantly to the model. Therefore, it is best to have a model with as few as possible variables. Prior to analyzing the data it was decided as a rule of thumb to limit the number of variables in the model to twelve at most. This model contains twelve variables. Thus, although there is evidence that the equations given could be improved by adding another variable or two, this will not be done in this technical report.

The analysis of variance table and statistics of fit for the equation are given in Table II. A way to determine how well a variable contributed to the model is by comparing its PROB > F value for the partial sums of squares to a specified level of significance. If the PROB > F value of the variable is less than the specified level of significance the variable is said to contribute significantly to the model. It can be observed that all variables included in the equation would be significant at the 0.0002 level which indicates that each variable contributes significantly to the model.

The final residual plots are useful in determining how good the regression equation is. Residuals represent the amount which the regression equation has not been able to explain. These residuals can be throught of as observed errors if the model is correct. That is,

TABLE II

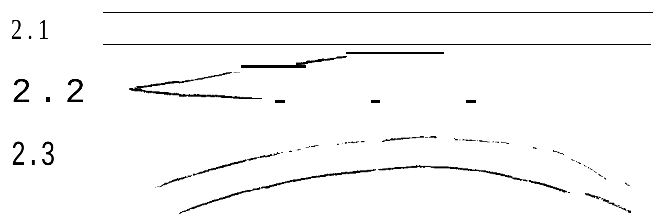
ANALYSIS OF VARIANCE TABLE AND STATISTICS OF FIT FOR MODEL 1

Source	DF	Sum of Squares	Mean Square	F Value	PR>F	R-SQUARE	C.V.
Model	12	0.07597151	0.00633096	8.45	0.0001	0.842153	21.3398%
Error	19	0.01423951	0.00074945		Std Dev		Mean
Corrected Total	31	0.09021102			0.02737604		-0.12828620

Source	DF	Partial SS	F Value	PR>F
$(CLWC)^{2/3}$	1	0.02567402	34.26	0.0001
CLBC	1	0.02782809	37.13	0.0001
$(CLBC)^2$	1	0.02013088	26.86	0.0001
$(DTL)^{1/2}$	1	0.04236074	56.52	0.0001
$(DTI)^{2/3}$	1	0.03950010	52.71	0.0001
CLWC * CLBC	1	0.03080820	41.11	0.0001
DTL * DTI	1	0.05144775	68.65	0.0001
CLWC * DTI	1	0.02541509	33.91	0.0001
$(CLWC)^2$ * CLBC	1	0.02461467	32.84	0.0001
$(DTL)^2$ * CLBC	1	0.02998013	40.00	0.0001
CLWC * $(DTL)^{2/3}$	1	0.04542368	60.61	0.0001
$(DTL)^6$ * $(DTI)^6$	1	0.01649705	22.01	0.0002

they represent the difference between what is actually observed and what is predicted by the equation. Plots of residuals versus the dependent variable and residuals versus each of the four independent variables were made for this model. These plots are given in Table III. Plots of these types may take on various shapes as illustrated in Figure 2. Figure 2.1 represents the shape of a "horizontal band and

FIGURE 2: Various Shapes of Residual Plots



thus gives no indication of abnormality. Consequently conclusions based on normal theory would not appear to be invalid. Figure 2.2 illustrates non-homogeneity of variance and implies transformations are needed on the dependent variable or one of the independent variables depending on where this occurs. Figure 2.3 indicates the need for a possible quadratic term in one of the independent variables to be added if it occurs there. Should it appear on the dependent plot, this would indicate the need for a transformation on it. The plots in Table III all have the shape of a "horizontal band." These plots seem to indicate that the assumptions made in the analysis were not seriously violated. The usual assumptions made are that the error terms are independent, have a mean of zero, possess a constant variance, σ^2, and follow a normal distribution. The last assumption

TABLE III: RESIDUAL PLOTS FOR MODEL 1

PLOT OF RESID_Y*PRED_Y   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

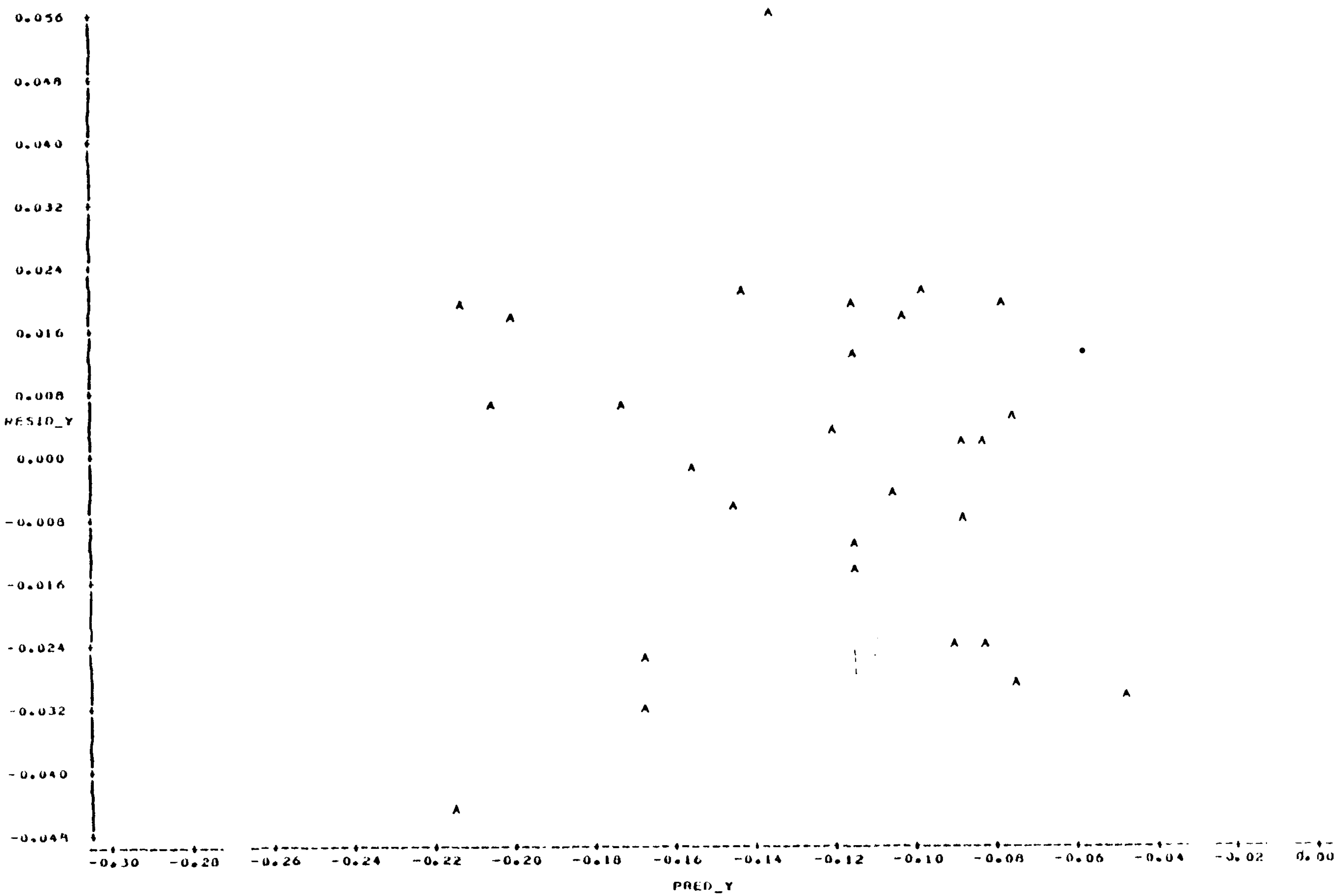


TABLE III: RESIDUAL PLOTS F O R MODEL I

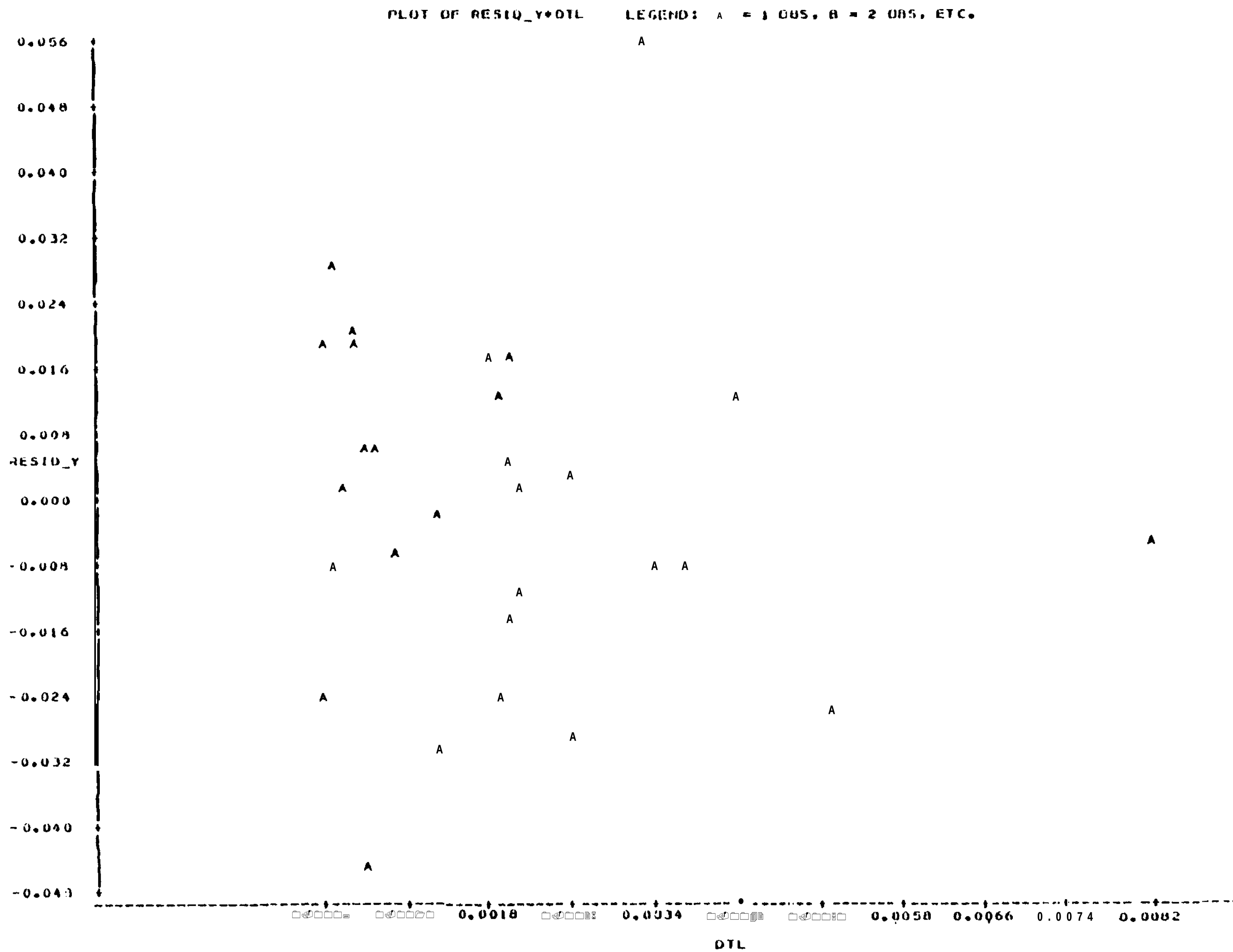


TABLE III: RESIDUAL PLOTS FOR MODEL I

PLOT OF RESID_Y*DTI   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

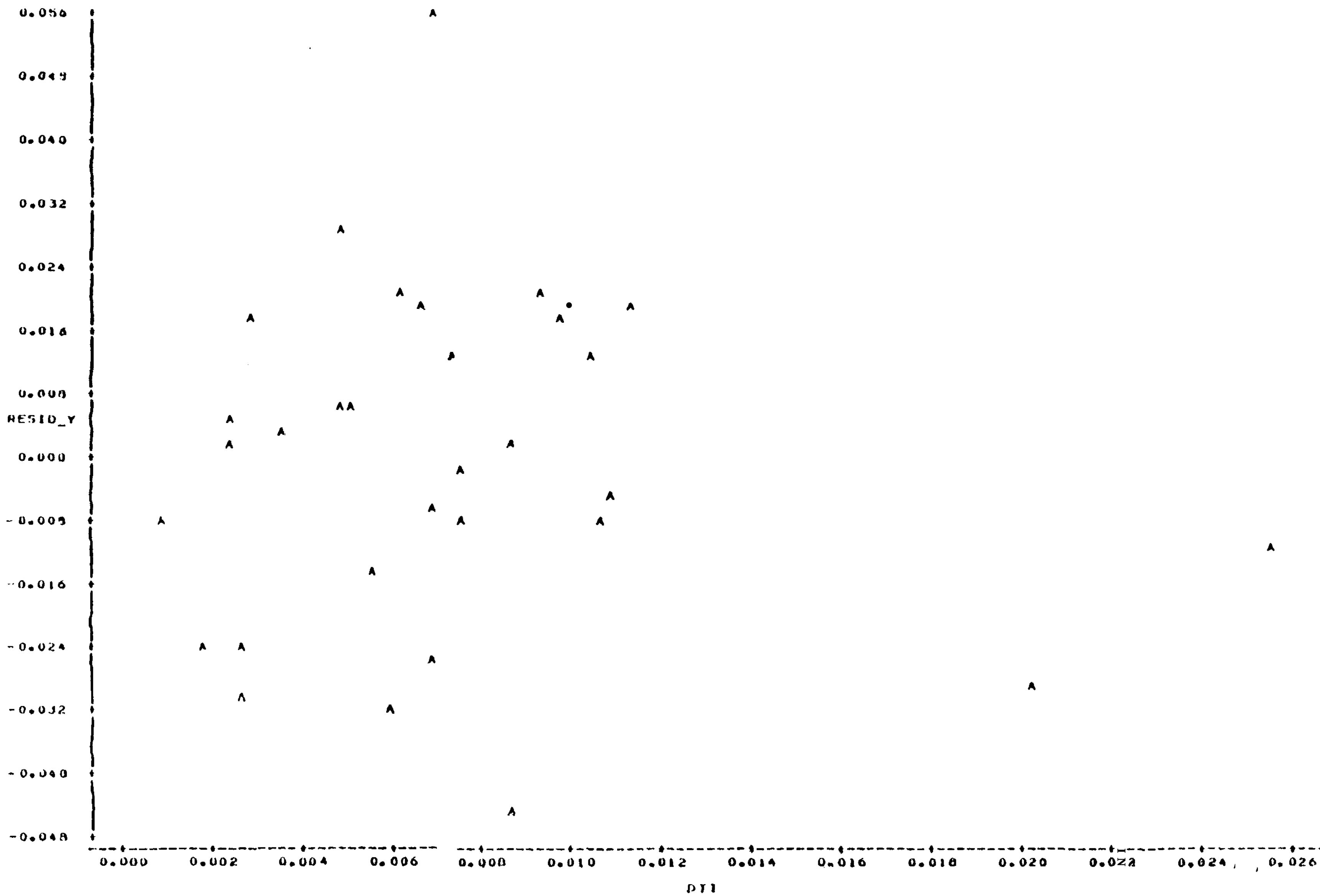


TABLE III: RESIDUAL PLOTS FOR MODEL I

PLOT OF RESID_Y*CLWC   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

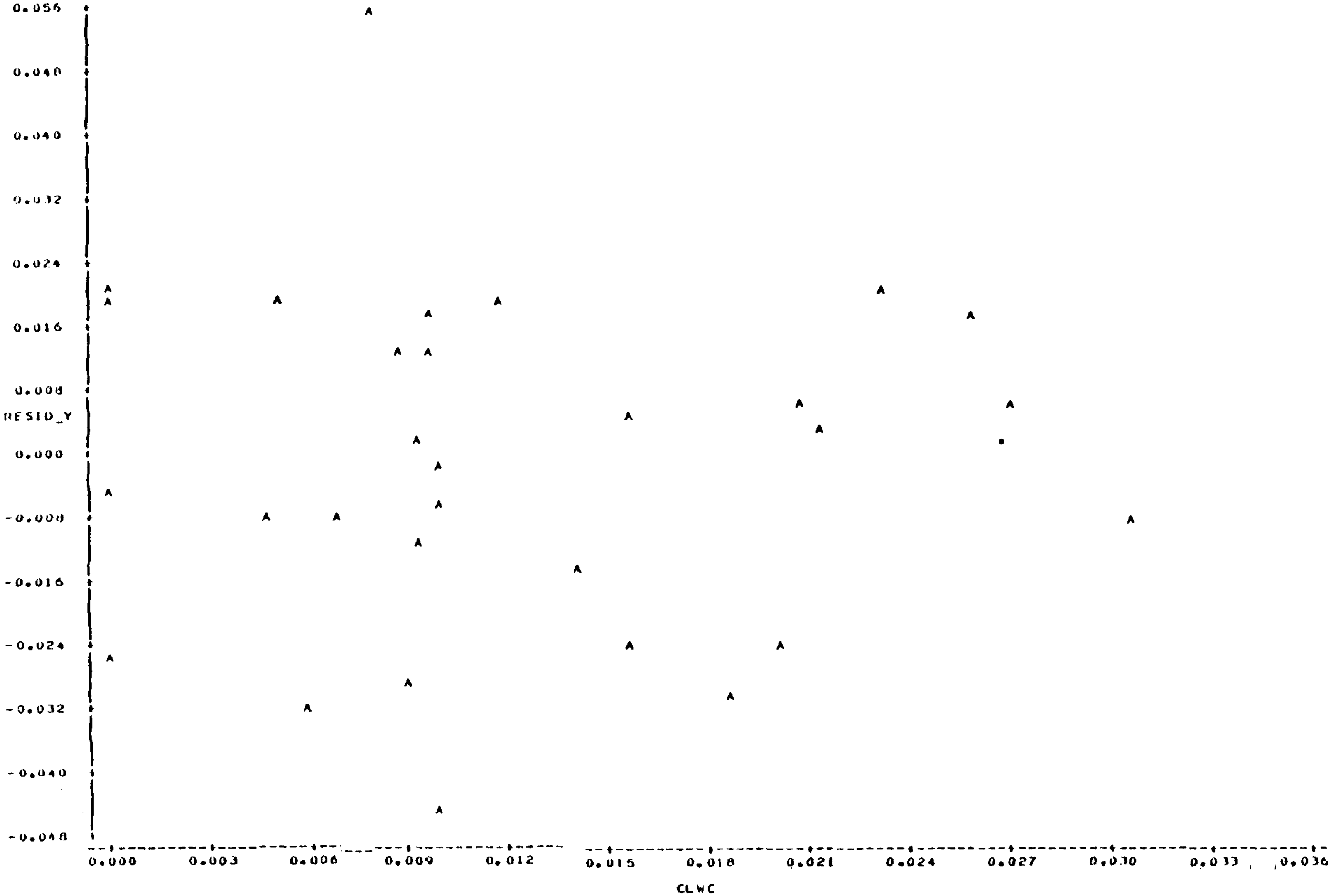
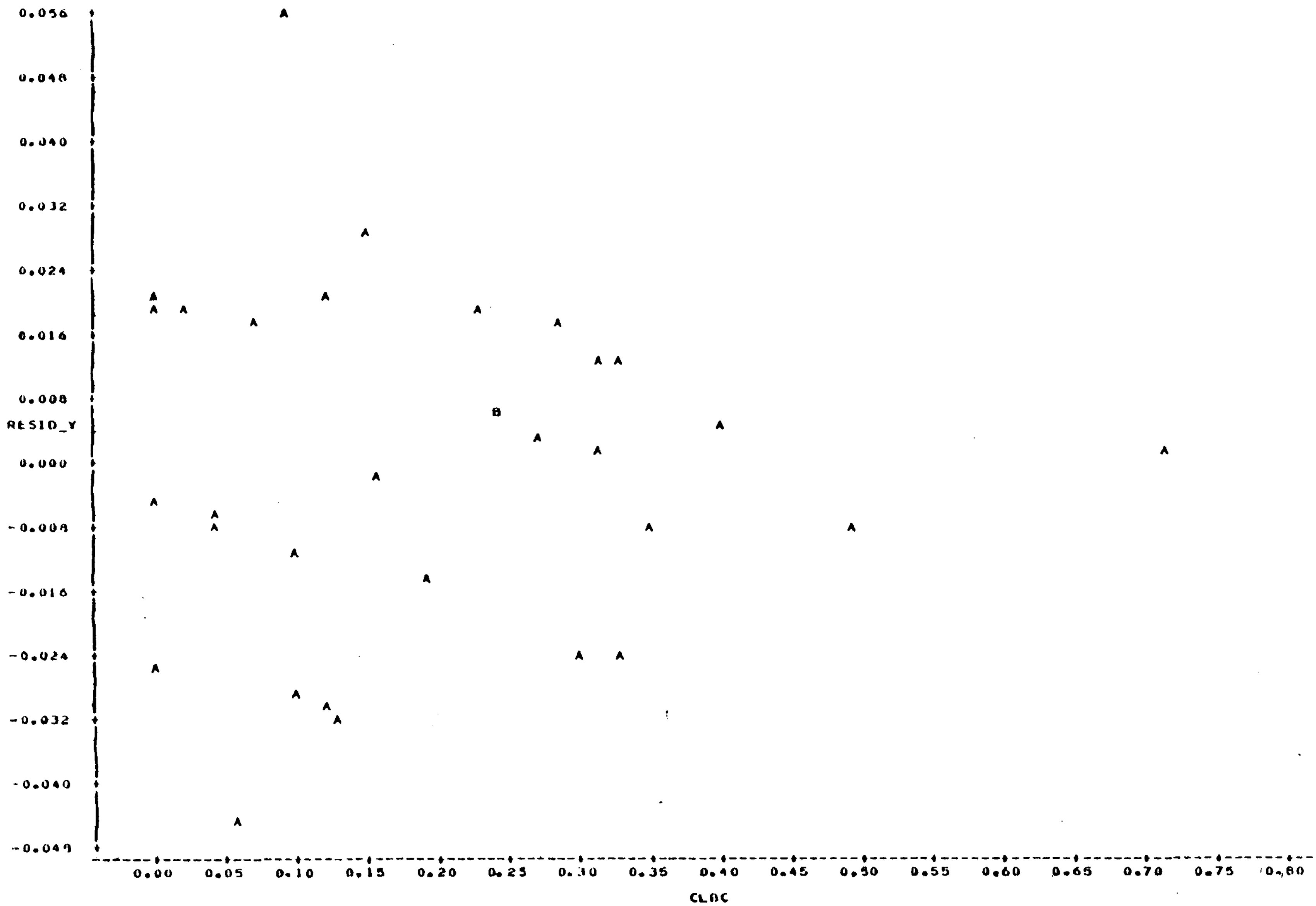


TABLE III: RESIDUAL PLOTS FOR MODEL 1

PLOT OF RESID_Y*CLBC LEGEND: A = 1 OBS, B = 2 OBS, ETC.



is required for making F-tests previously discussed concerning significance of variables in the model.

A major benefit obtained from this model is its usefulness for predictive purposes. Several precautions need to be mentioned before the model is used in this context. First, due to the range of values on variable $(DTL)^6 * (DTI)^6$ it was convenient to code the original data. If in using the equation one has measurements recorded in feet then in order to use the model it is necessary to code this data by dividing each measurement by 200,000. The following example will illustrate.

For well #24 say, the original measurements were:

CLWC = 4280 ft., CLBC = 54880 ft., D-IL = 420 ft., D-II = 700 ft.

Values which should enter the model are:

CLWC = 4280/200,000 = .0214, CLBC = 54880/200,000 = .2744, D-IL = 520/200,000 = .0026, DTI = 700/200,000 = .0035

Secondly it should be recalled that the dependent variable in the model is not first year average daily production but rather the transformed variable. As this measure is a strictly increasing function of FYADP, larger predicted values will be indicative of better producing wells and thus we can quantitatively compare wells with respect to their productive capabilities.

Thirdly, as is generally the case when using a polynomial to represent a complex response surface, one should not extrapolate beyond the range of values for the independent variables that were used to obtain the equation. The model may not adequately represent the surface for values of the independent variables which lie outside that range and in fact there is some risk in using the equation along the periphery of the region of experimentation. The ranges in feet for each of the four measurements are listed in Table IV.

Table IV

Range of Independent Variables (feet)

Variable	Minimum Value	Maximum Value
CLWC	0	6140
CLBC	0	142680
DTL	40	1640
DTI	180	5120

To demonstrate the use of the equation we have selected twenty-five potential well sites within a small section of the map from which our data were obtained. To obtain them a 5 x 5 grid was superimposed on the section and the points of intersection of the grid lines were chosen as the well sites. Figure 3 contains a reproduction of this section with the grid lines drawn in and with circles of 1,000 feet radius drawn around the selected well sites.

Before using the equation it was first necessary to obtain distance measurements for each well as described in the section on data collection. Afterwards we then checked each distance for a

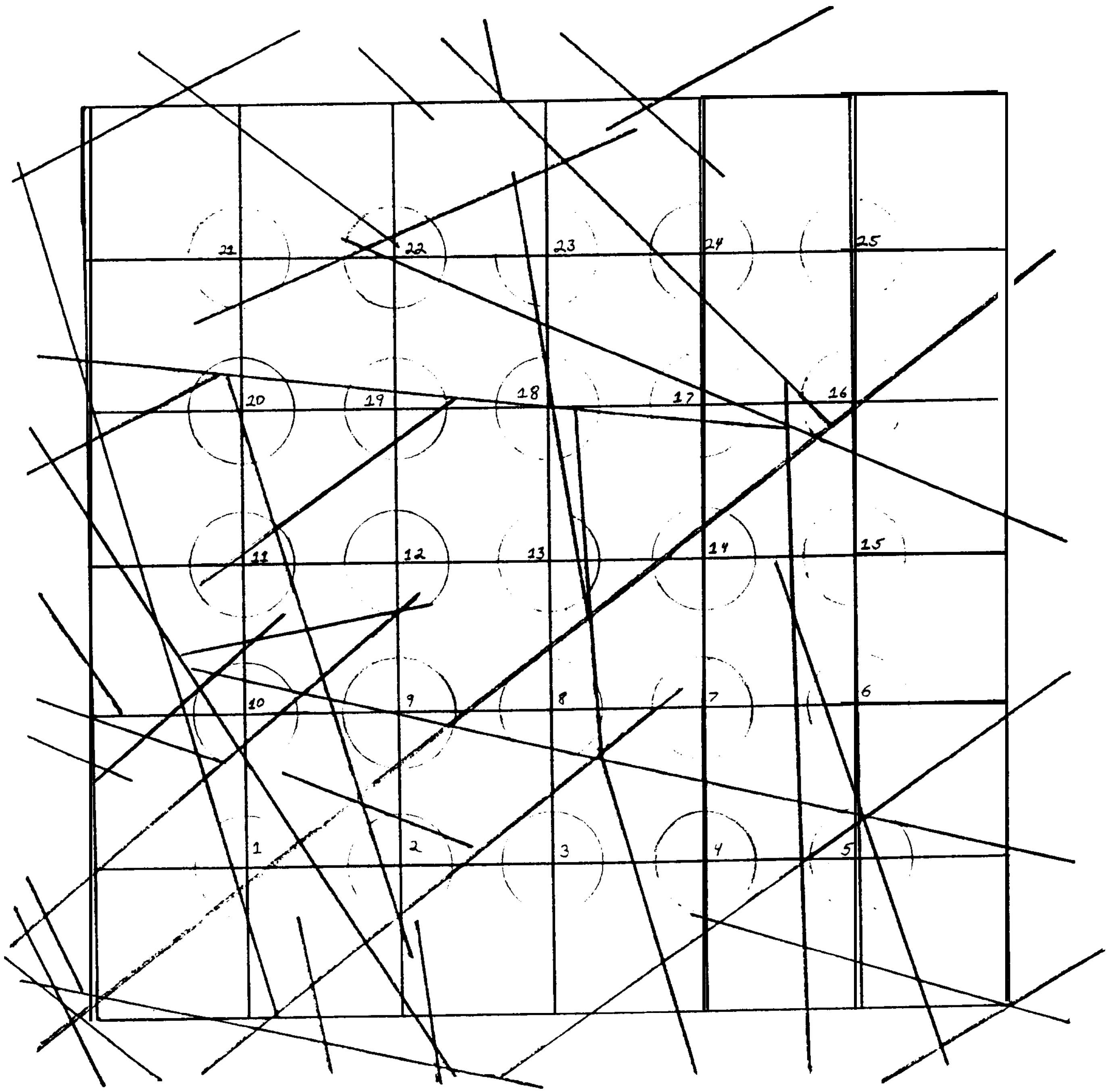


Figure 3: Subsection of Overall Map Illustrating Twenty-five Potential Well Sites.

given well to ensure that it was within the allowable range as designated in Table IV. Of the twenty-five potential well sites four were eliminated in the process. These were the well sites designated as 9, 18, 20 and 25. The reasons for eliminating well 18 was that its distance to nearest intersection was below the minimum value. The other three were eliminated because their distances to the nearest lineament were outside the permissible range; well sites 9 and 20 being too close to a lineament while well site 25 was too far away.

The resulting predictions are given in Table V along with the values of the distance parameters for each well site. In determining which of the twenty-one legitimate well sites should have the largest production potential one would only need to choose the well with the longest PREDICTED Y value. For this illustration, the above situation occurs for well #19 with a PREDICTED Y value of 0.147. The fact that the predicted value is positive relates to the nature of the transformation used and should be construed as a large productivity for the well. On this basis we would consider well #19 as being the best producing site among the twenty-one considered.

TABLE V

Distance Parameters and Predicted Y Values for Twenty-Five Potential Well Sites

Well No.	CLWC (ft)	CLBC (ft)	DTL (ft)	DTI (ft)	PREDICTED Y
1	3772	44752	240	463	- 0.097
2	3748	43516	289	988	- 0.139
3	0	0	1047	2315	- 0.180
4	0	0	1070	1256	- 0.158
5	4442	77153	368	797	- 0.172
6	1587	26331	571	1951	- 0.161
7	869	27348	513	2195	- 0.233
8	2037	45863	720	1356	- 0.142
9	2918	45863	38*	1057	
10	5145	72134	413	647	- 0.126
11	2908	18480	162	967	- 0.088
12	896	28270	803	962	- 0.131
13	2480	16536	633	1395	- 0.035
14	1778	26203	409	2278	- 0.211
15	0	0	1148	1739	- 0.152
16	3686	53521	150	562	- 0.122
17	3661	44712	262	1755	- 0.133
18	4476	44941	65	156*	
19	3355	34661	402	3149	0.147
20	3713	54992	20*	774	
21	1086	9519	799	2288	- 0.148
22	4590	46947	40	383	- 0.167
23	1829	9737	302	1301	- 0.138
24	1595	10839	571	3045	- 0.144
25	0	0	2546*	2891	

*Denotes outside range

As a check on the adequacy of the equation production data for twelve additional wells was obtained. However, one well had a missing value for the first year production and therefore could not be used as a test well. The data associated with the remaining eleven wells are listed in Table VI.

TABLE VI

Additional First Year Production Data

Well No.	Annual Prod (Mcf)	Days Used	$FYADP = \frac{\text{Annual Prod}}{\text{Days Used}}$ (Mcf/d)	Observed $Y = -1/(FYADP)^{1/2}$
9	4944	91	54.3	-.13571
11	2871	61	47.1	-.14571
18	4061	64	63.5	-.12549
31	2606	61	42.7	-.15303
32	4742	92	51.5	-.13935
33	534	4	133.5	-.08655
39	1132	16	70.8	-.11885
39	415	14	29.6	-.1838
55	899	16	56.2	-.13339
56	683	13	48.7	-.1433
72	3841	91	42.2	-.15394

Using the model predicted values were calculated for each of the wells in Table VI except wells #9 and #32 which had parameter values outside the allowable ranges. These are given in Table VII along with the observed values of Y for purposes of comparison:

TABLE VII

Comparison of Observed and Predicted Values

Well No.	11	18	31	33	38	39	55	56	72
Observed Y	-.14571	-.12549	-.15303	-.08655	-.11885	-.1838	-.13339	-.1433	-.15394
Predicted Y	-.12060	-.13103	-.13677	-.19197	-.10600	-.14664	-.21946	-.16557	-.18127

It may be observed from this table that the model does a good job of predicting for eight of the nine wells (#'s 11, 18, 31, 38, 39, 55, 56, 72). That is, the differences between observed and predicted for these wells are reasonably small. However, for well #33 the model does not do a good job of prediction. An analysis was conducted as to why this occurred for this well.

Well #33 was only opened for four days during the first year whereas all other wells listed in Table VI were opened for at least thirteen days. Consequently the variance of the production figure for well #33 will be considerably greater than the variance for a well opened thirteen days or more. Also one might question why this well was closed after only four days. It may have been the case that this well began as a high producer and suddenly decreased drastically. in production causing it to be closed after the fourth day. Had it been left open for a longer period of time, this figure may have given a better representation of its production. Furthermore, the magnitude of the observed production figure for this well far exceeds that of the remaining eight wells.

The true indicator as to how well the equation is doing as a predictor can be found by constructing confidence limits around each of the individual predicted values. If we can say with a high percentage of confidence that we expect a large proportion of these intervals to contain the corresponding observed values and, in fact they do, then we have a strong inclination to believe the model is correct and is doing a good job of prediction. Should it be the case that a small percentage of these observed values fall within these intervals then we are led to believe we have an inadequate model and are not doing a good job of predicting. Ninety percent confidence limits on individual predicted values were computed for each of the nine wells. These limits along with the observed Y figures are listed in Table VIII.

TABLE VIII

Ninety Percent Confidence Limits on Individual Predicted Values

Well No.	Observed-Y	Lower 90% CL Y	Upper 90% CL Y
11	-.14571	-.17583	-.06536
18	-.12549	-.18674	-.07531
31	-.15303	-.18687	-.08666
33	-.08655*	-.25092	-.13302
38	-.11885	-.15721	-.05478
39	-.1838	-.19885	-.09443
55	-.13339*	-.28599	-.15292
56	-.1433	-.21685	-.11429
72	-.15394	-.24035	-.12218

*Denotes OBSERVED-Y value is outside the interval.

One may observe that seven out of nine observed production figures lie within the intervals and if one does not consider well #33 on the basis that it was opened for only four days during the first year then seven out of eight observed production figures lie within the intervals. On this basis, since a large percentage of these values fall within the intervals coupled with the fact that the predicted values are reasonably close to the observed values as indicated in Table VII, we feel fairly confident that the equation is adequately representing the true model and is doing a good job of predicting.

In a similar study a model was obtained relating the wells production history to its flow parameters. This involved obtaining an equation relating FYADP to the flow variables (IOFBS, IOFAS, IRPAS). For this model a total of 40 wells were available. However, a combination of ten of these wells contained missing values thus allowing only 30 observations to be analyzed. The model along with its R-SQUARE value is as follows:

MODEL 2: PRODUCTION AS A FUNCTION OF FLOW PARAMETERS

MODEL:

$$
\begin{aligned}
LN(FYADP) = {} & 3.3845 + 3.51096 * 10^{-5} (IOFAS)^2 \\
& - 4.81878 * 10^{-8} (IOFAS)^3 - 1.3003 (IPFBS)^{1/2} \\
& + 0.76991 (IOFBS)^{2/3} - 1.34232 * 10^{-4} (IOFAS * IOFBS) \\
& + 1.068471 * 10^{-11} (IRPAS)^4
\end{aligned}
$$

where LN = Natural Logarithm

R-SQUARE = 0.711

The analysis of variance table is listed in Table IX.

TABLE IX

ANALYSIS OF VARIANCE TABLE FOR MODEL 2

Source	DF	Sum of Squares	Mean Square	F Value	PR > F	R-SQUARE	C.V.
Model	6	12.97281838	2.16213640	9.45	0.0001	0.711463	11.3437
Error	23	5.26119133	0.22874745		Std Dev		Mean
Corrected Total	29	18.23400971			0.47827549		4.21621103

SOURCE	DF	Partial SS	F Value	PR > F
$(IOFAS)^2$	1	5.48231029	23.97	0.0001
$(IOFAS)^3$	1	3.88283190	'6.97	0.0004
$(IOFBS)^{1/2}$	1	4.05471153	17.73	0.0003
$(IOFBS)^{2/3}$	1	4.24168134	18.54	0.0003
IOFAS*IOFBS	1	4.29900754	18.79	0.0002
$(IRPAS)^4$	1	0.94334635	4.12	0.0540

One may observe from the PROB > F values that every variable included in the model is significant at the 0.0004 level except $(IRPAS)^4$. This variable, however, is significant at the 0.0540 level. Consequently the F-values are highly significant. The value of R-SQUARE is 0.711 which indicates approximately 71% of the variation is explained by the regression equation. Predicted values were computed from this equation for the nine additional wells. The results showed that this equation predicted low for well #33. This same situation occurred in MODEL 1. This further supports the idea that a true observed production figure is not being represented since this well was only opened four days.

As a further illustration of the usefulness of the results found in this study, two groups of three wells have recently been opened on this same field and, in fact, their FYADP figures are unavailable at this time. Therefore, their observed Y values cannot be calculated. The four distance parameters for each of these six well locations are calculated by using the COMP-U-GRID digitizer. These values along with the predicted value of Y calculated by means of MODEL 1 are listed in Table X. The individual identity of each well has been concealed to protect confidentiality. Because two distance parameters for well #2 in group I were outside the allowable ranges, predicted values are not given for it.

TABLE X

Six Additional Wells

	Well No.	CLWC (ft)	CLBC (ft)	DTL (ft)	DTI (ft)	PREDICTED Y
Group I	1	1071	4869	97	1800	-.24610
	2	6736*	61128	87	103*	
	3	3666	54934	145	857	-.17093
Group II	1	2704	14624	614	1351	-.00561
	2	4507	32308	244	349	-.07457
	3	2155	21060	278	1469	-.15630

* Indicates value lies outside range of independent variables.

Ninety percent confidence limits on individual predicted values were calcualted for the-legitimate wells and are listed along with the predicted values of Y in Table XI.

TABLE XI

Ninety Percent Confidence Limits on Individual Predicted Values for the Six Additional Wells

	Well No.	Predicted Y	Lower 90% CL on Y	Upper 90% CL on Y
Group I	1	-.24610	-.30834	-.18385
	2	*		
	3	-.17093	-.22316	-.11870
Group II	1	-.00561	-.06527	.05405
	2	-.07457	-.13184	-.01730
	3	-.15630	-.20569	-.10692

*Contained distance parameters which were outside range of independent variables.

If one wanted to rank the three wells within each of the two groups on the basis of being 90% confidence of a best producer, second best producer, and third best producer, it would be necessary for the intervals in Table XI not to overlap. In particular, for group 1, where we do not consider well number 2 on the basis that it had distance parameters which were outside the range of values used to build the model, the remaining two intervals overlap. Therefore we cannot rank the wells in group 1. This same situation occurs in group 2. Since the three intervals overlap, we are unable to rank these wells and be 90% confident of being correct.

However, we are able to choose for each group the well with the largest predicted Y value. For group 1 this occurs for well number 3 since we should not consider well number 2 for reasons previously mentioned. For group 2 we would choose well number 1 as having the largest predicted Y.

SUMMARY

The results of this study provide a method for predicting a wells productivity from its location relative to lineaments. The usefullness of it is that it enables one to compare the predicted productivity at several well sites and then to choose the one with the larger production potential for drilling. The model for obtaining these predicted values is designed to be utilized for a specified region of values as discussed earlier. Its validity was checked using the production figures for twelve additional wells as test data and in addition, an examination of residual plots was utilized.

There are several ways the study might be improved. First, the existing model is based on production data for 40 of the 75 wells. Twelve additional wells were used to validate it. Further validation would be possible if production data were made available for the remaining 23 wells.

Secondly, the data used in the analysis were "shot" well data. In recent years, the method of hydraulic fracturing has been introduced and it is believed to enhance production. Consequently it might be beneficial to do a similar study using data from wells which were hydraulically fractured.

REFERENCES

American Geological Institute. Dictionary of Geological Terms, Anchor Press/Doubleday, Garden City, New York, 1976, 472 pp.

Barr, A. J., J. H. Goodnight, J. P. Sall, and J. T. Helwig. A User's Guide to SAS 76, Sparks Press, Raleigh, N.C., July 1976, 329 pp.

Draper, N.R. and H. Smith. Applied Regression Analysis, John Wiley & Sons, Inc., New York, N.Y., 1966, 407 pp.

Graybill, Franklin A. Theory and Application of the Linear Model, Wadsworth Publishing Company, Inc., Belmont, California, 1976, 704 pp.

Overbey, W. K., Jr., Sawyer, W. K., and Henniger, R. R. (1974) Relationships of earth fracture systems to productivity of a gas storage reservoir: U.S. Bur. Mines RI-7952; 133 pp.

Ryan, William M. (1976) Remote sensing fracture study-western Virginia and south eastern Kentucky: U.S. ERDA MERC/SP-76/2: 94-95.

Tukey, John W. Exploratory Data Analysis, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1977, 688 pp.