FORECASTING ACCURACY

SAINT JOHN RIVER, NEW BRUNSWICK

SPRING 1974

J.E. Peters
District Engineer
Water Survey of Canada
Halifax, Nova Scotia

R.A. Pol

January, 1976

TABLE OF CONTENTS

FORECASTING ACCURACY

SAINT JOHN RIVER, NEW BRUNSWICK

SPRING 1974

OBJECTIVE:

To evaluate the forecasting accuracy of the SSARR model in determining flood flows on the Saint John River, New Brunswick, from a statistical viewpoint.

METHODOLOGY:

The discharge and stage forecasts at six stations were used. The discharge forecasts (from best to worst) were checked at the Mactaquac, Grand Falls and the Fort Kent station, whereas the stage forecasts were checked at the Oak Point, Jemseg and the Fredericton station.

The best station in each of the discharge and stage forecasts was used, compared with one of the worst, and a station falling between the two, as found in the report by the Water Resources Branch.* All data used were also taken from this report, for the April-May 1974 period.

The forecasted values (for each of the 5 days of forecasts) were compared with the measured values at each of the stations mentioned.

In this analysis we are concerned with the dependence of the measured values on the independent variables, the forecasted values - the forecasted values are the independent variables since they are used to determine the actual values. The common statistic that is used to indicate the degree of association between two variables is the correlation coefficient, r. The value of the correlation coefficient may be within the limits of 0 to +1 (in this case), depending on the agreement of the variables. A value of

* Analysis of Forecasting Success, 1974, Saint John River June 12, 1974. Water Resources Branch, New Brunswick Department of Fisheries and Environment.

+1 denotes perfect agreement, whereas a value of r=0 denotes no relation between the variables. The more elongated the elliptical shape of the observed data, the higher the degree of correlation between the variables- thus the larger the value of r. A correlation coefficient greater than 0.6 is generally considered good* - a functional relationship exists between the variables.

The measured values were plotted against the forecasted values; the correlation coefficient and the regression line (the best fit line thru the points) were determined. When the points are scattered close to the regression line (when the deviation about the regression line is small), the correlation between the variables is greater.

Appendix B presents the graphs of the measured values against the forecasted for the three discharge stations with five days of forecast for the period April-May. The regression line (on log-log paper) was determined from equations calculated by the power curve regression method.

The equal yield line (the line at 45^{o} thru the origin) is also shown and represents perfect correlation between measured and forecasted values- the 1 to 1 relationship. The better the fit of the regression line with respect to the equal yield line the better the forecasted values are an indication of the measured or actual values.

Appendix C presents the graphs for the three stage stations and the parameters were determined in a similar manner as mentioned above, but the regression line was now determined by linear regression (plotted on arithmetic paper) since it was not necessary to use the logarithms of the stage values.

The absolute differences between the measured and forecasted values were also analyzed. The absolute difference is the error between the actual and

* Handbook of Applied Hydrology, Ven Te Chow, Editor-in-Chief, McGraw-Hill Book Company, 1964.

disregarding the sign

forecasted value; the mean, and standard deviation

(for each station and particular day of forecast) of the absolute difference was determined. The standard deviation is the dispersion about the mean error of a variable.

The percentage difference was determined by averaging the absolute value of the percent difference (between forecasted and measured) for a particular station and a particular day of forecast.

Table 1 and 2 (Appendix A) tabulate the results of the statistical analysis for the discharge and stage forecasts respectively.

OBSERVATIONS:

The three discharge basins analyzed all had a high degree of correlation between forecasted and actual values. The fit of the regression line to the equal yield line was good; the correlation coefficient was high. It should be noted that a high correlation coefficient alone does not yield good correlation between forecasted and measured values. Good correlation between forecasted

and measured values is associated with a high correlation coefficient; also, when the regression line and equal yield line coincide. In spite of this and as can be seen in Appendix B there still remains a high degree of correlation. But, the correlation between forecasted and measured values decreased from day 1 to day 5 at all basins checked.

The range of errors is presented in the summary tables (see Appendix A). Examining the Mactaquac Basin (1 day forecast) for example, it can be seen that the mean error between forecasted and measured value is in the order of 10,000 cfs and this would be an absolute average percent difference of about 13% between the forecasted and measured values.

Discharge forecasting errors have a tendency to increase drastically at all sites from day 1 to day 5. Errors beyond day 1 are very large in most cases. As mentioned in the report by the Water Resources Branch, "large singular errors occurred on all forecast days, 1 to 5; the rate of accumulating errors decreased somewhat with time in days 1 and 2 and all days levelled off in error accumulation just after the first peak."

The stage forecasts for the three basins checked also have a high degree of correlation between forecasted and actual values. Again errors increased from day 1 to day 5 and the correlation beyond day 2 decreased drastically (see Appendix C). The range of errors is presented in the summary tables in Appendix A.

Overall, the stage forecasts were more accurate than the discharge forecasts, eg. the percentage error was much smaller.

It should be noted* that there always remains some uncertainty as to the exact conditions within a basin, no matter how complete the reporting network; a 5 to 10 percent residual error is therefore, associated with even the best rainfall-runoff relationship. Under average conditions, stage forecasts cannot be expected to be consistently reliable to within less than 0.5 ft. Using this criteria, it would seem reasonable that the SSARR model was quite accurate in forecasting stage, whereas beyond day 2 of the discharge forecasts the accuracy deteriorated considerably.

* Applied Hydrology, Linsley, R.K. Jr.,Kohler, M.A.; Paulhus, J.L.H. McGraw-Hill Civil Engineering Series, McGraw-Hill Book Company, 1949.

APPENDIX A

TABLE 1. DISCHARGE FORECASTS - ANALYSIS OF THE ERROR BETWEEN FORECASTED AND MEASURED VALUES.

	No. of Forecasts checked	Correlation Coefficient "r"	Mean Error (cfs)	Standard Deviation (cfs)	% * /Diff./
MACTAQUAC					
1 day forecast	40	0.98	9949	9024	13.3
2 day forecast	39	0.97	17131	15639	19.8
3 day forecast	39	0.93	25168	21800	24.9
4 day forecast	38	0.91	33010	36264	31.6
5 day forecast	37	0.84	57098	104137	34.3
GRAND FALLS					
1 day forecast	39	0.99	6763	8616	11.8
2 day forecast	39	0.96	11200	12550	22.2
3 day forecast	39	0.92	16569	18583	33.4
4 day forecast	38	0.90	25656	28647	39.1
5 day forecast	36	0.92	27632	23166	44.6
FORT KENT					
1 day forecast	37	0.99	6144	9690	13.5
2 day forecast	38	0.96	8607	9445	27.4
3 day forecast	38	0.93	14489	16592	48.1
4 day forecast	37	0.93	20180	18522	59.2
5 day forecast	35	0.93	22423	16742	68.7

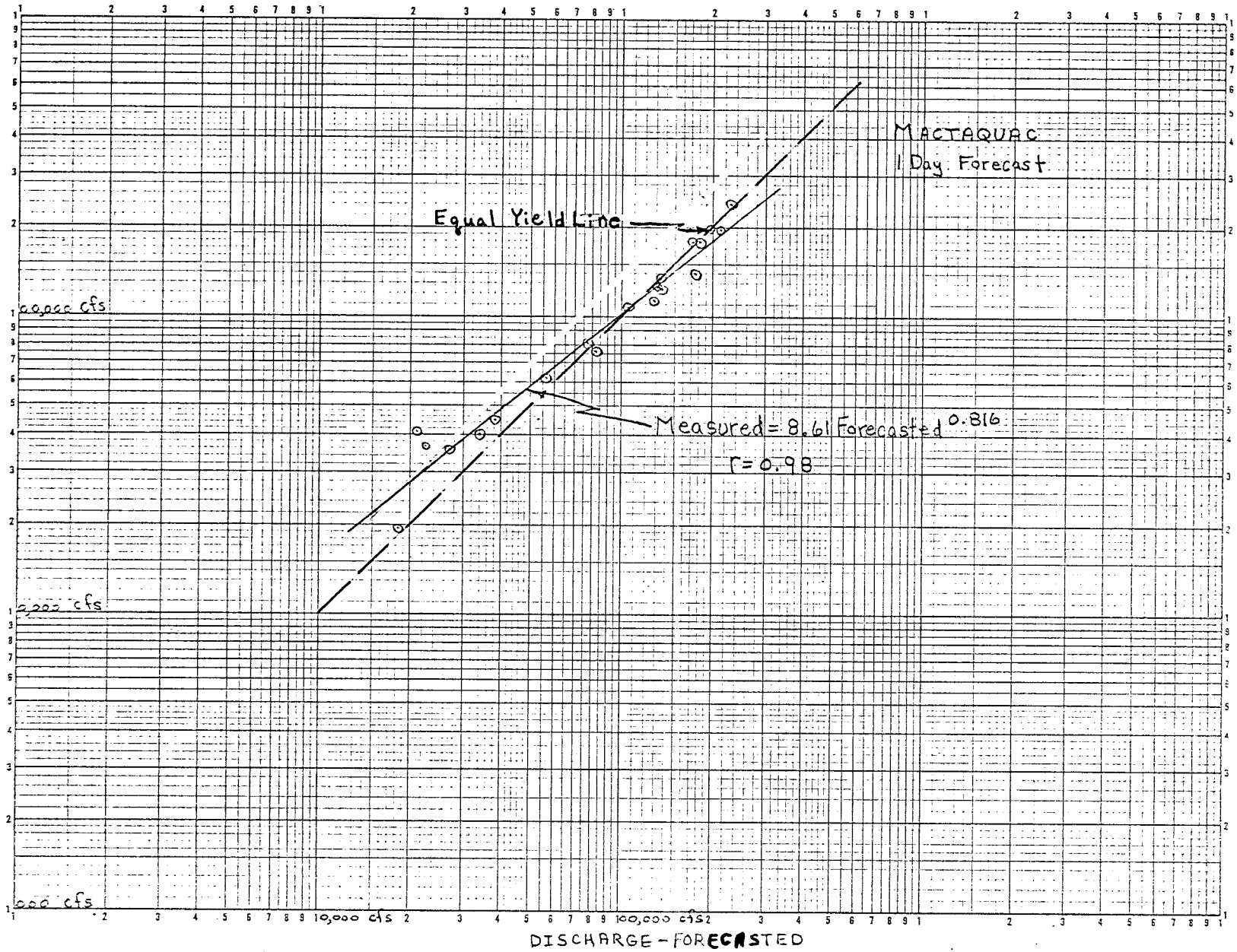
* Analysis of Forecasting Success, 1974, Saint John River June 12, 1974. Water Resources Branch, New Brunswick Department of Fisheries and Environment.

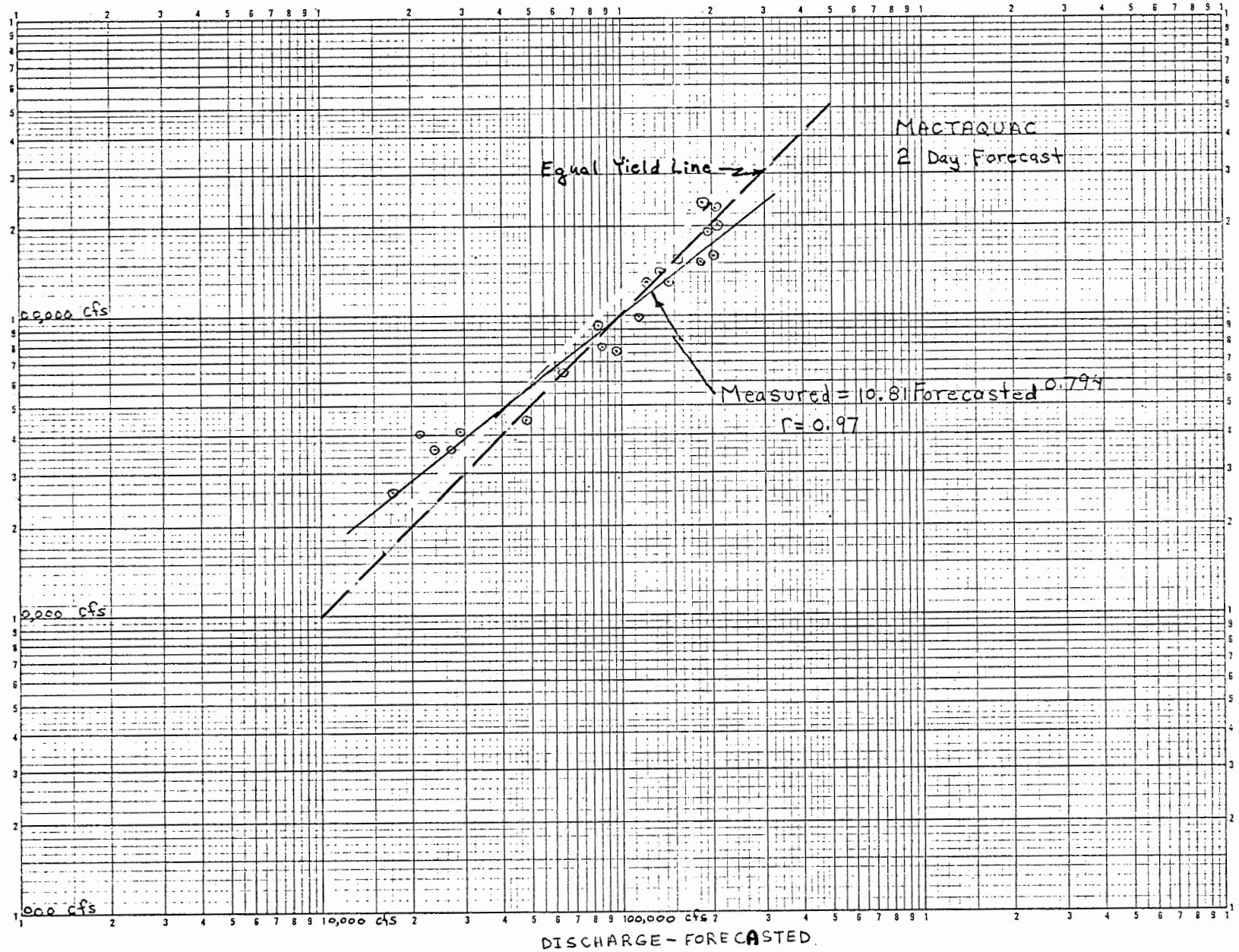
TABLE 2. STAGE FORECASTS - ANALYSIS OF THE ERROR BETWEEN FORECASTED AND MEASURED VALUES.

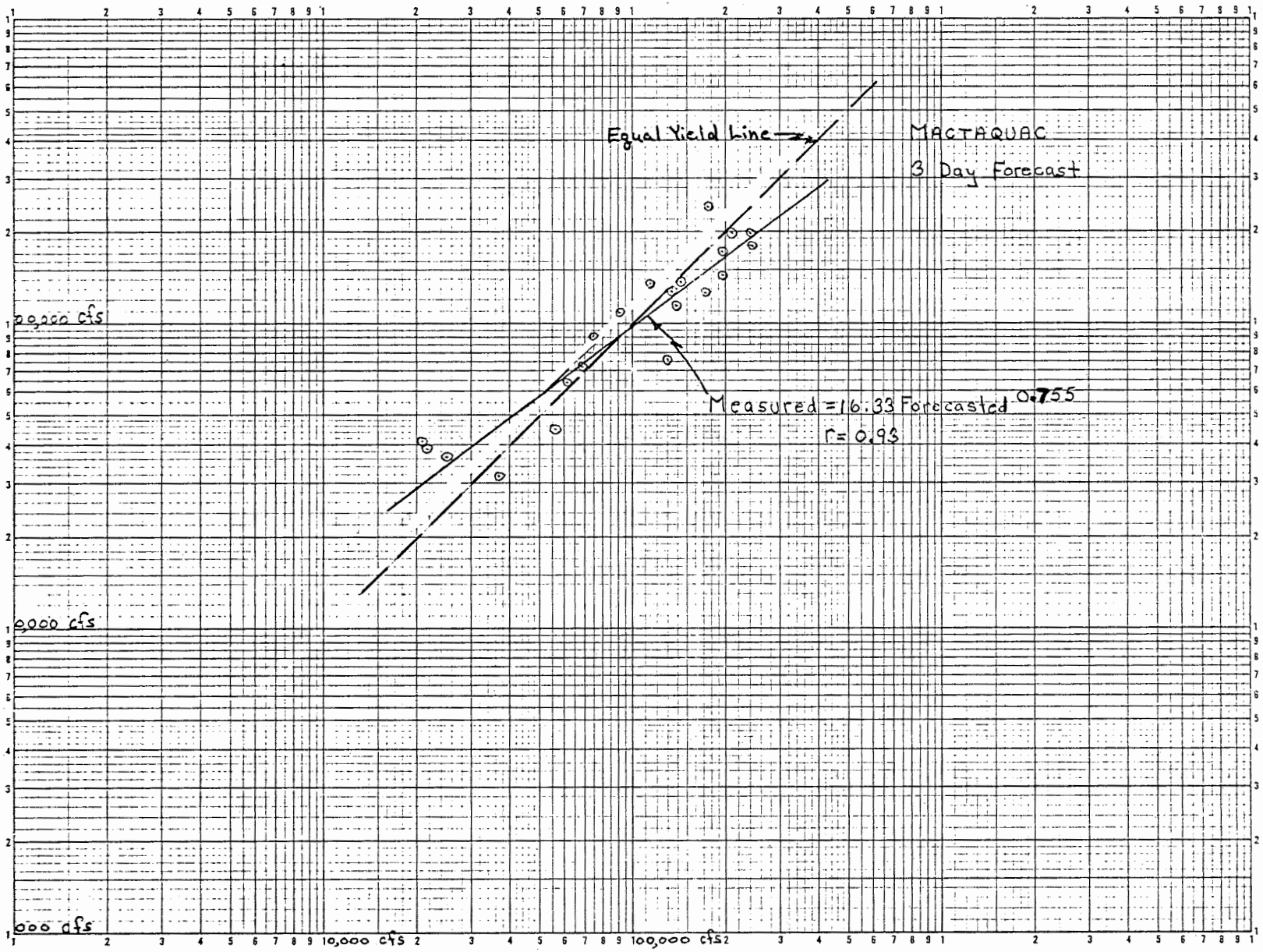
	No. of Forecasts Checked	Correlation Coefficient "r"	Mean Error (ft.)	Standard Deviation (ft.)	% * /Diff./
OAK POINT					
1 day forecast	20	0.97	0.42	0.35	3.8
2 day forecast	20	0.95	0.43	0.28	3.7
3 day forecast	20	0.92	0.43	0.41	3.5
4 day forecast	18	0.79	0.65	0.56	5.2
5 day forecast	16	0.66	0.87	0.64	6.9
JEMSEG					
1 day forecast	20	0.94	0.77	0.52	5.7
2 day forecast	20	0.91	0.88	0.59	6.3
3 day forecast	20	0.85	1.04	0.69	7.4
4 day forecast	18	0.74	1.10	0.91	7.4
5 day forecast	16	0.61	1.32	0.96	9.1
FREDERICTON					
1 day forecast	23	0.99	0.50	0.47	3.4
2 day forecast	24	0.92	1.32	1.17	8.8
3 day forecast	24	0.84	1.78	1.41	8.7
4 day forecast	24	0.62	2.59	2.07	15.1
5 day forecast	21	0.59	2.18	1.54	11.5

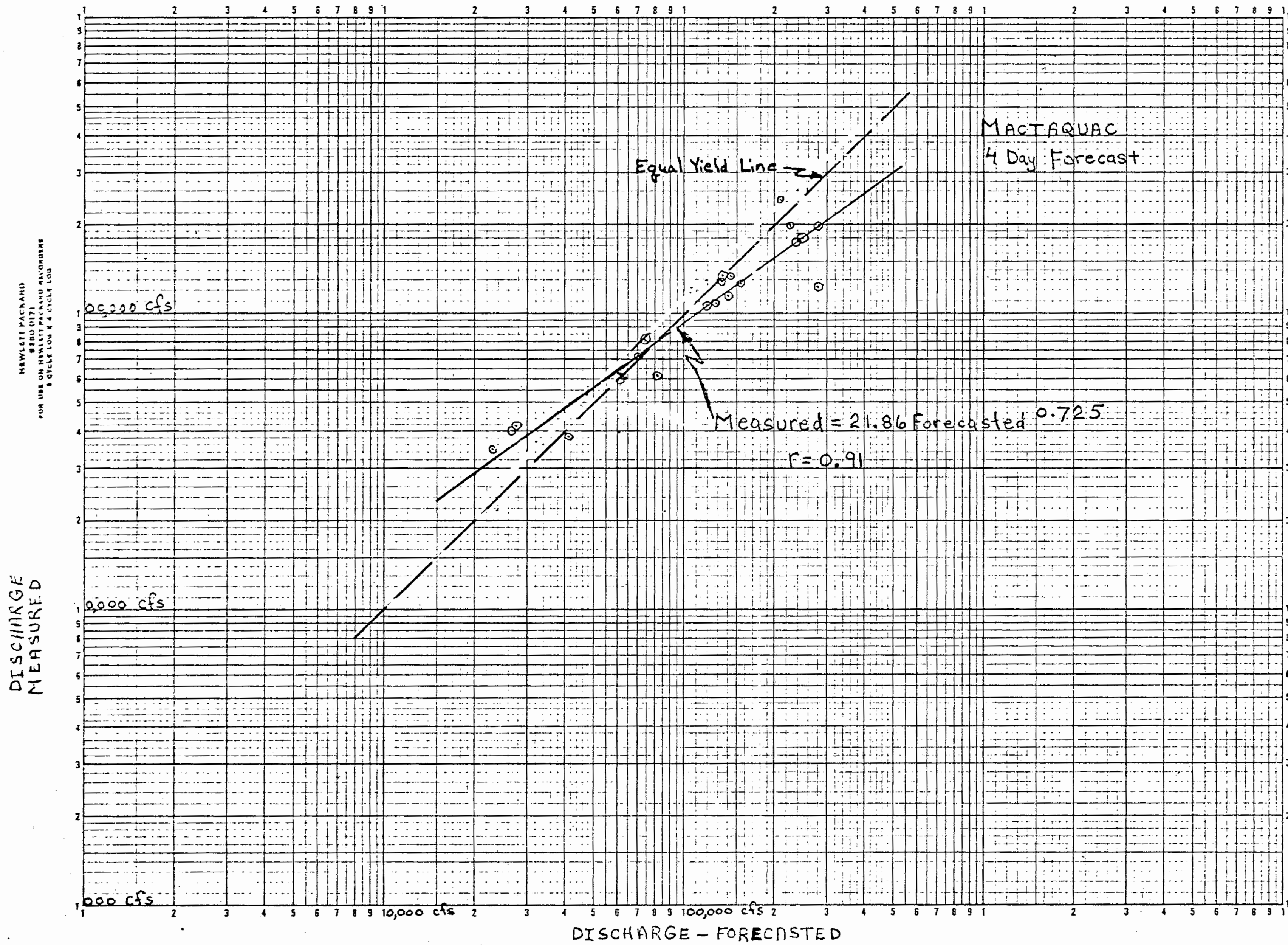
*Analysis of Forecasting Success, 1974, Saint John River June 12, 1974, Water Resources Branch, New Brunswick Department of Fisheries and Environment.

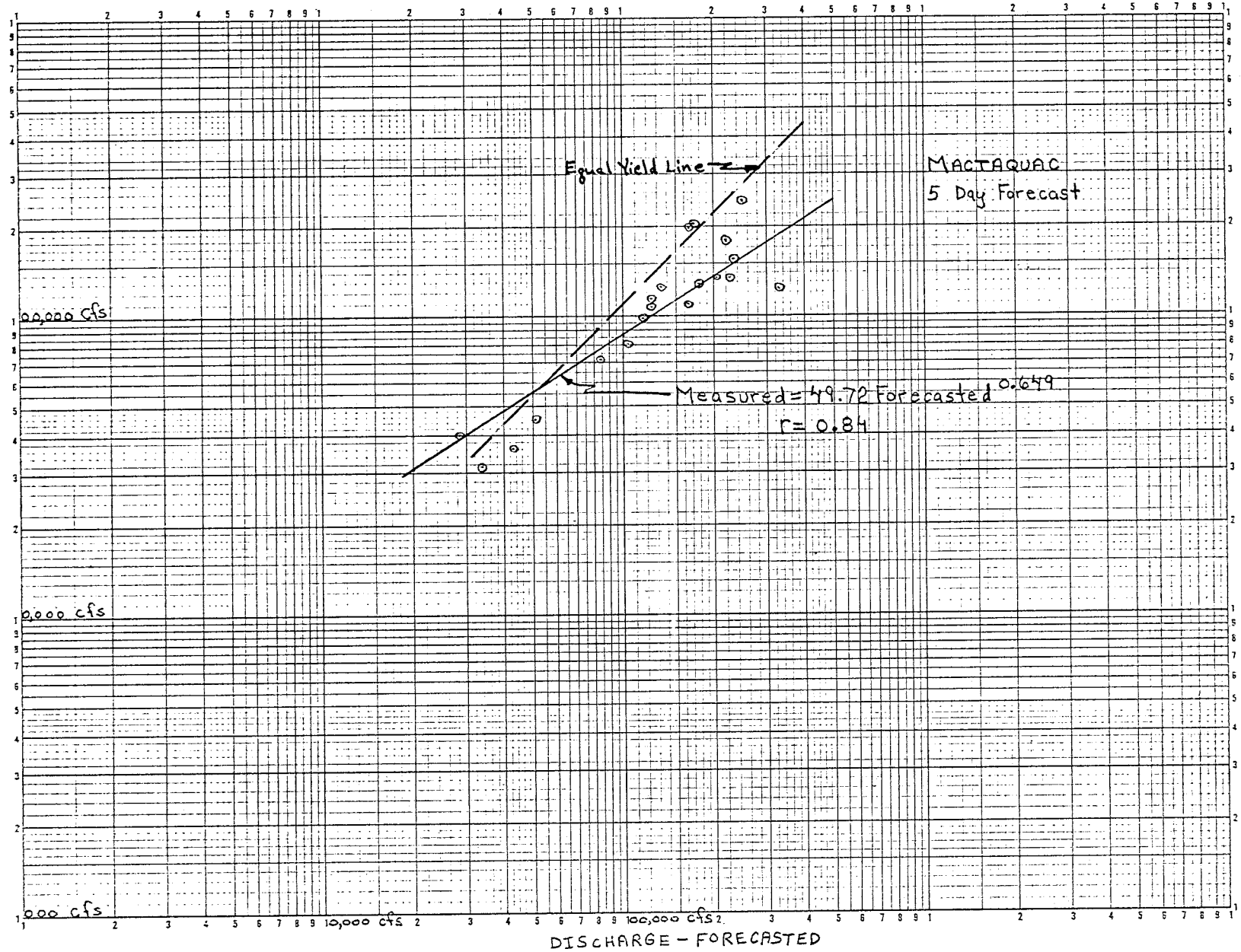
APPENDIX B


MACTAQUAC
1 Day Forecast
Equal Yield Line
Measured = 8.61 Forecasted 0.816
r = 0.98
10,000 cfs
100,000 cfs
DISCHARGE - FORECASTED
DISCHARGE MEASURED

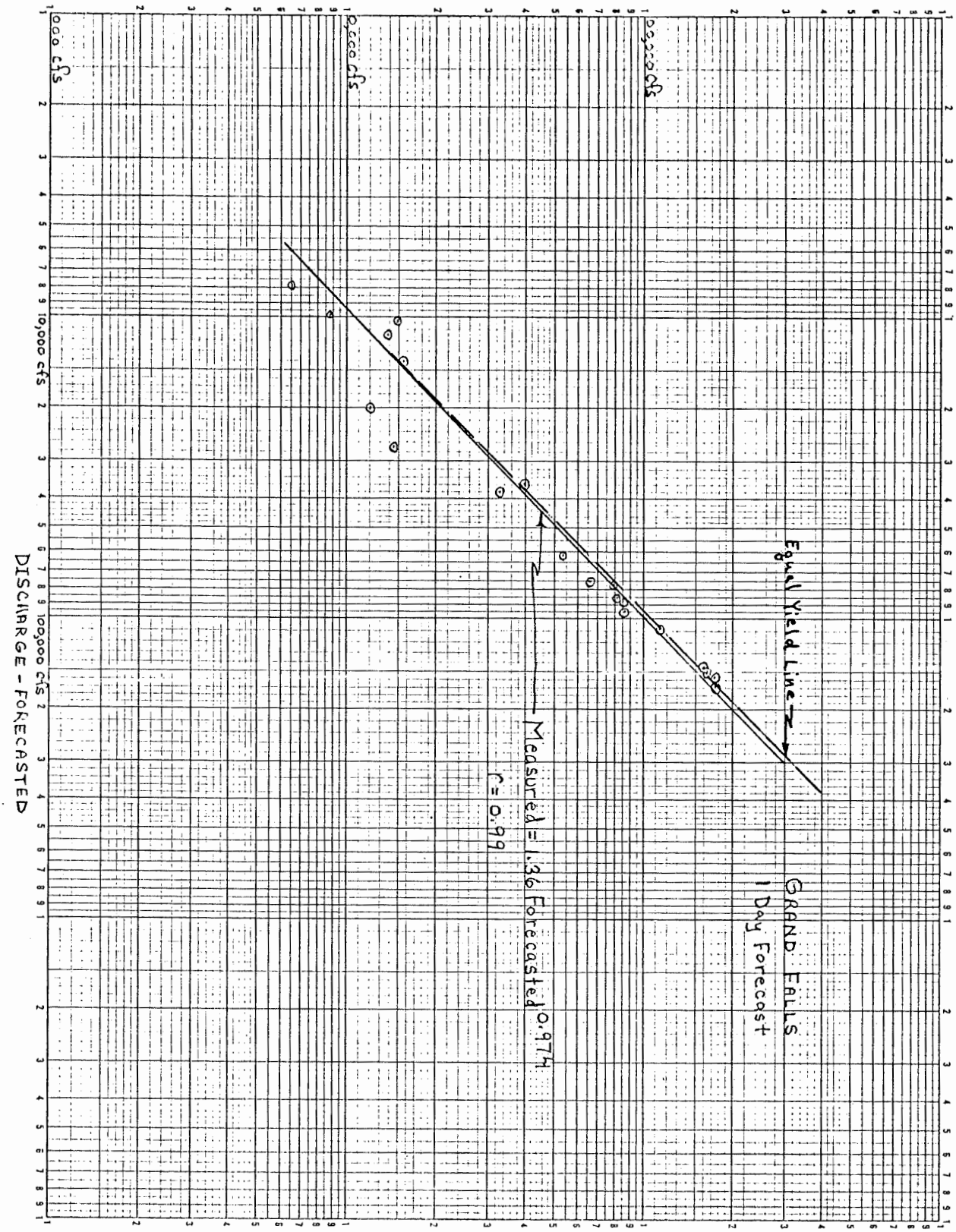

MACTAQUAC
2 Day Forecast
Equal Yield Line
Measured = 10.81 Forecasted 0.794
r = 0.97
DISCHARGE - FORECASTED
DISCHARGE MEASURED
10,000 cfs
100,000 cfs


Equal Yield Line
MACTAQUAC
3 Day Forecast
Measured = 16.33 Forecasted 0.755
r = 0.93
10,000 cfs
100,000 cfs
DISCHARGE - FORECASTED
DISCHARGE MEASURED

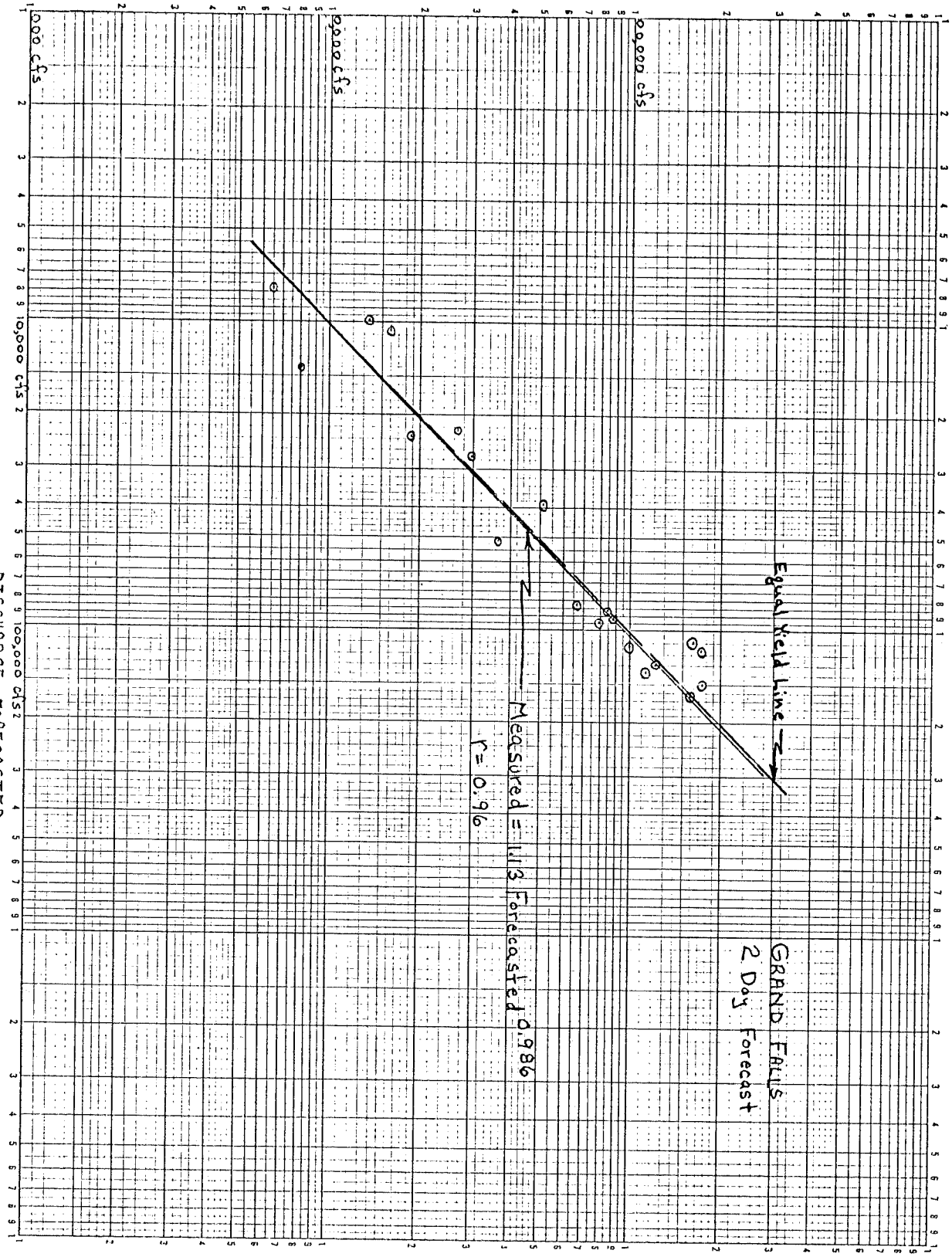

HEWLETT PACKARD
9280-0171
FOR USE ON HEWLETT PACKARD RECORDERS
3 CYCLE LOG X 4 CYCLE LOG
MACTAQUAC
4 Day Forecast
Equal Yield Line
Measured = 21.86 Forecasted 0.725
r = 0.91
DISCHARGE MEASURED
DISCHARGE – FORECASTED
10,000 cfs
100,000 cfs
cfs


Equal Yield Line
MACTAQUAC
5 Day Forecast
Measured = 49.72 Forecasted 0.649
r = 0.84
10,000 cfs
100,000 cfs
DISCHARGE - FORECASTED
DISCHARGE MEASURED

DISCHARGE
MEASURED



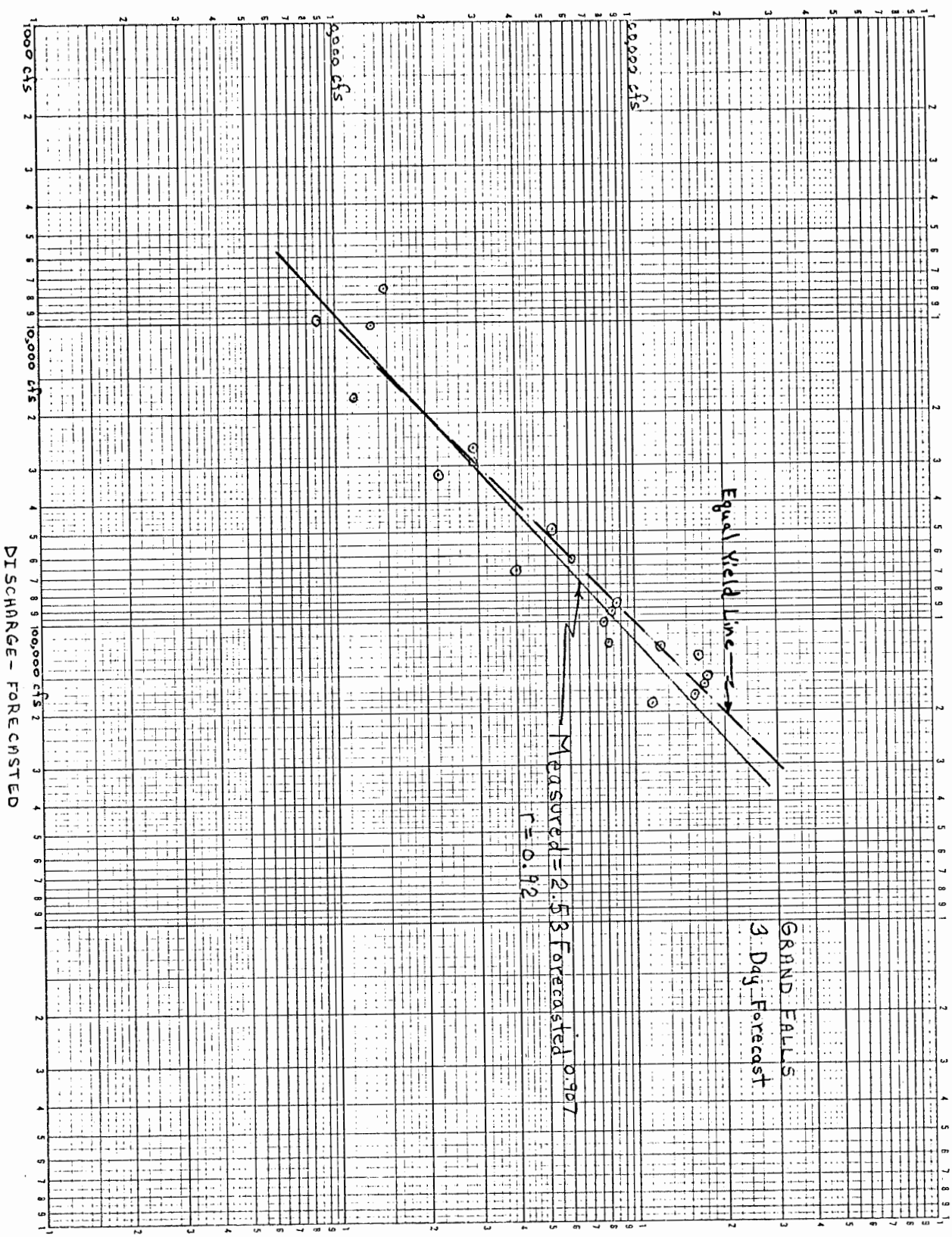
DISCHARGE MEASURED



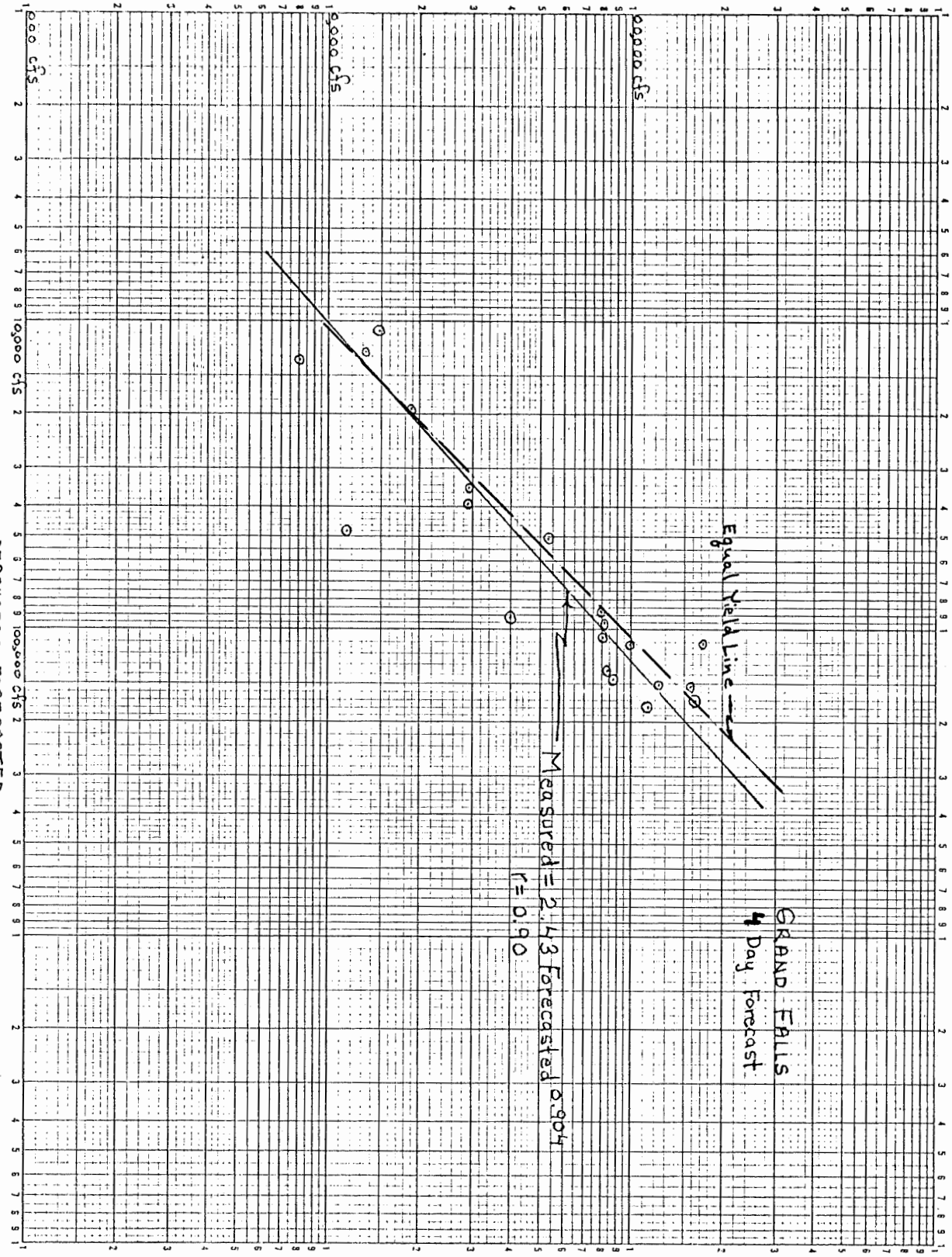
LT

DISCHARGE MEASURED

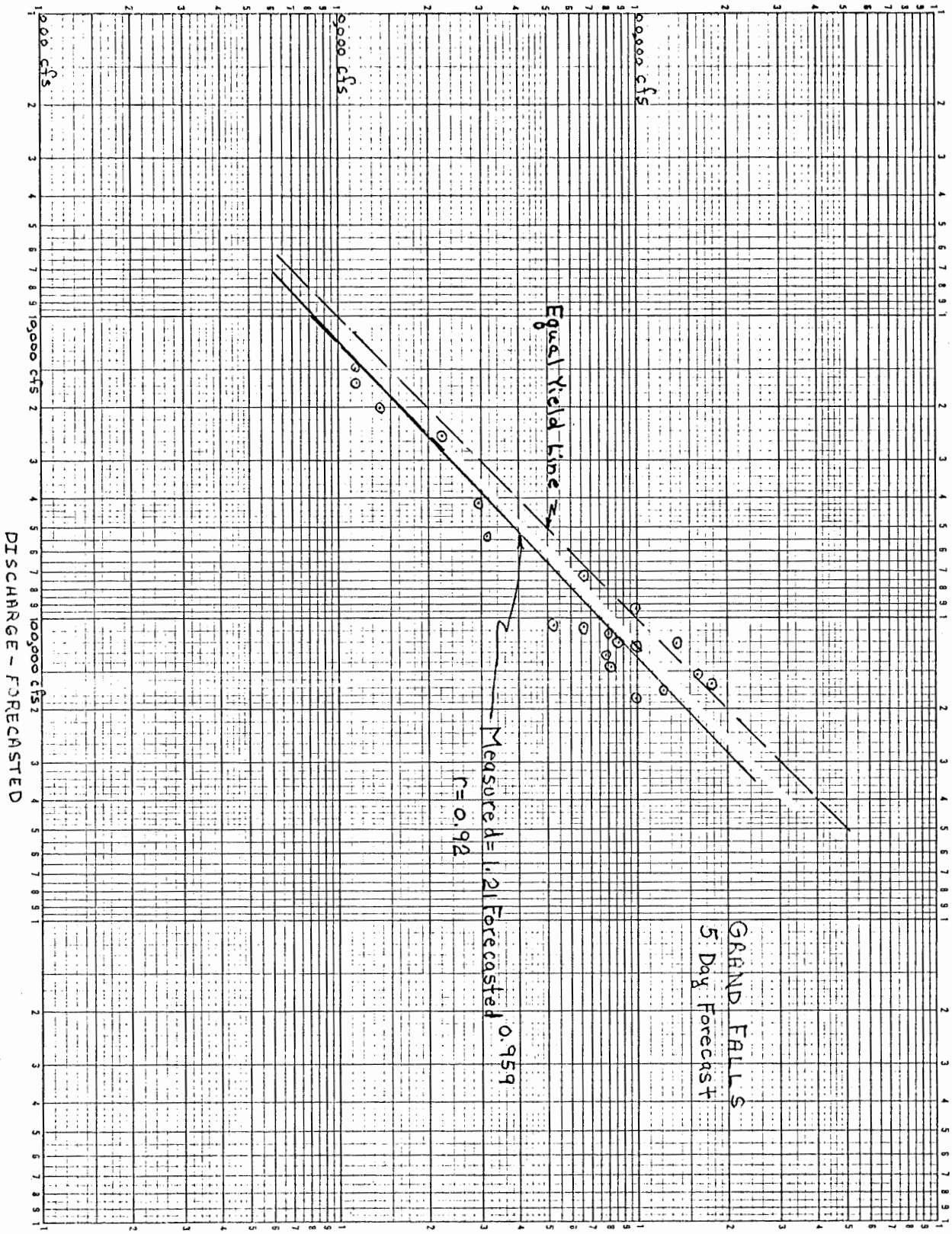
HEWLETT-PACKARD
92810-0171
FOR USE ON HEWLETT-PACKARD RECORDERS
3 CYCLE LOG X 4 CYCLE LOG

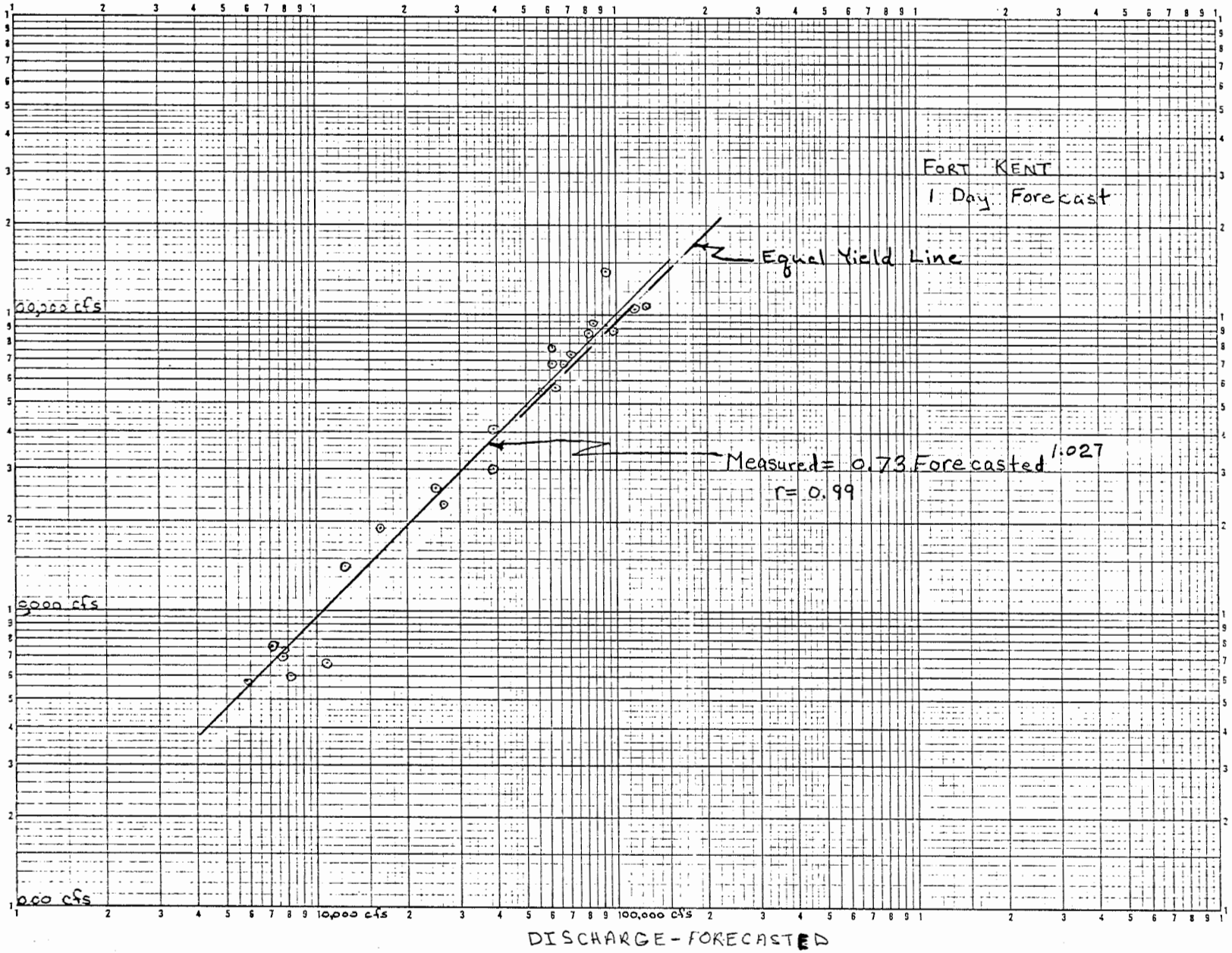


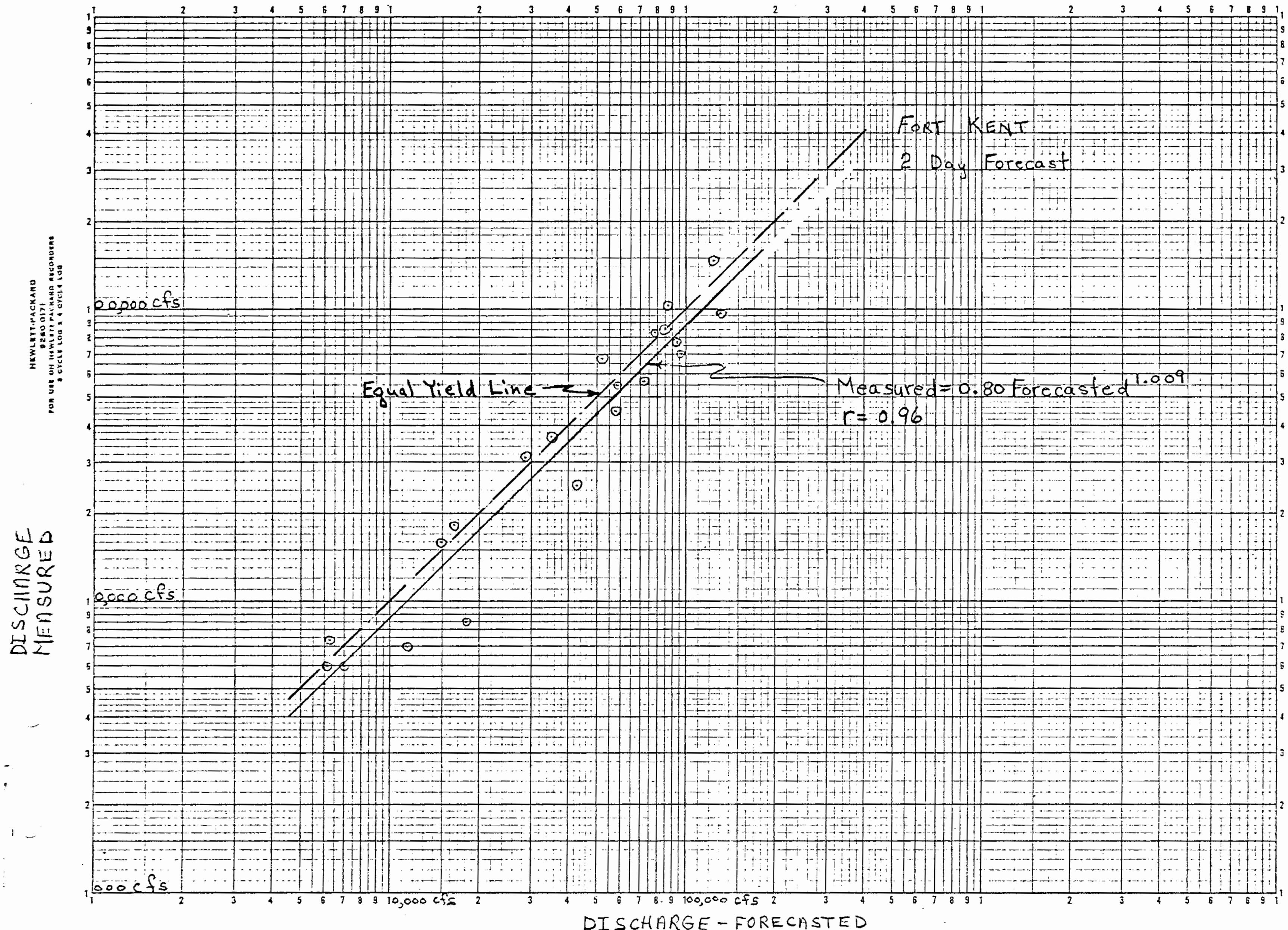
DISCHARGE MEASURED

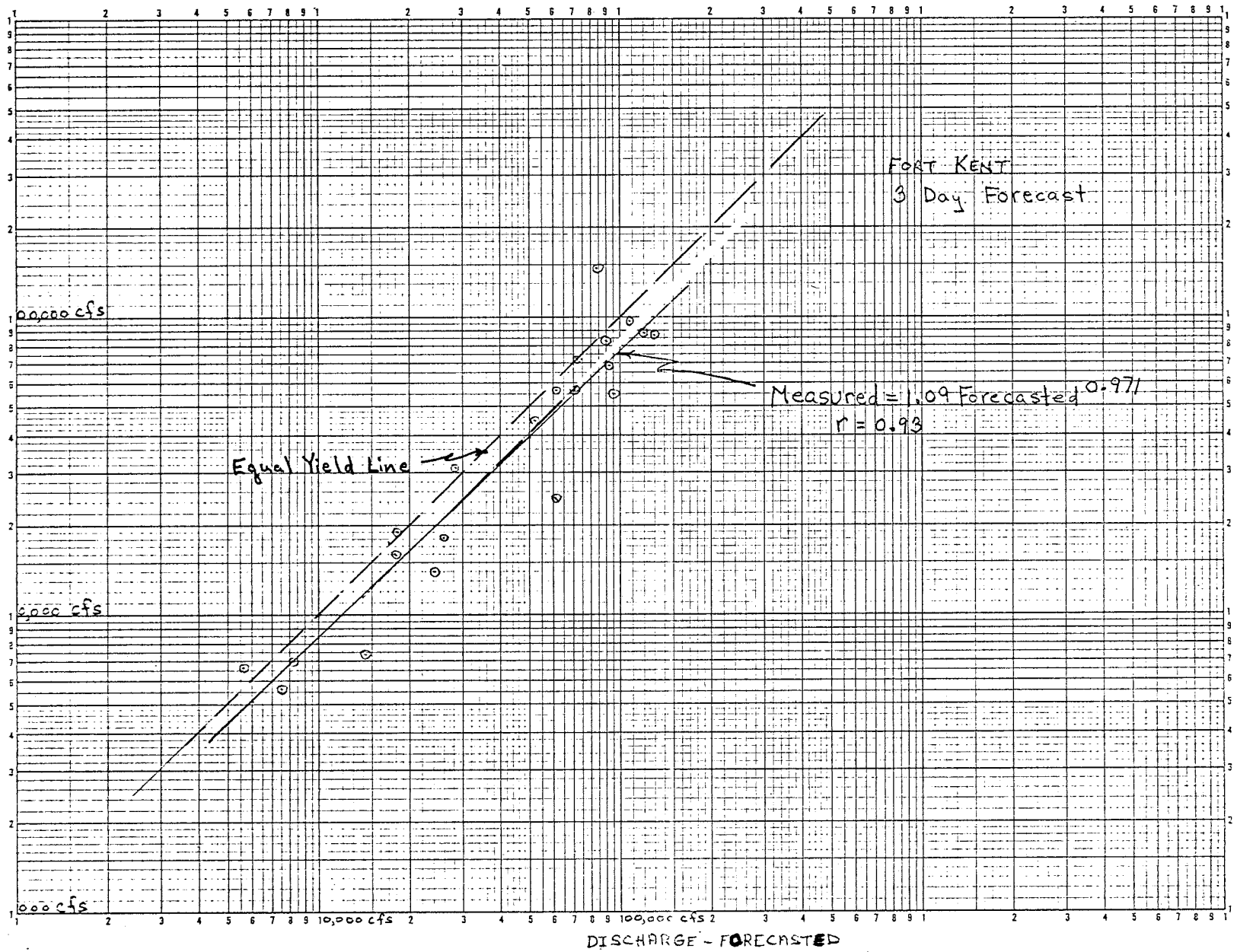


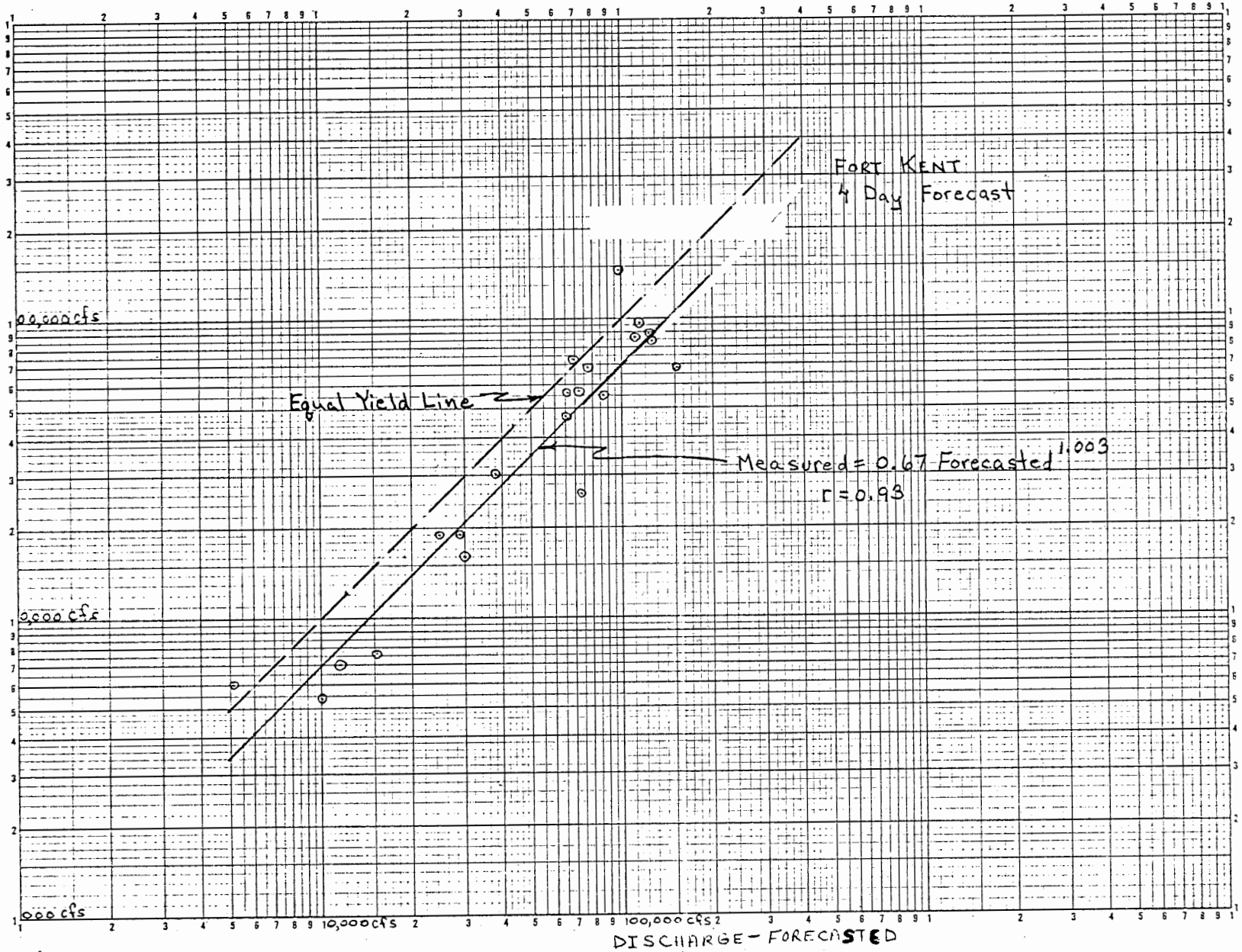
DISCHARGE MEASURED

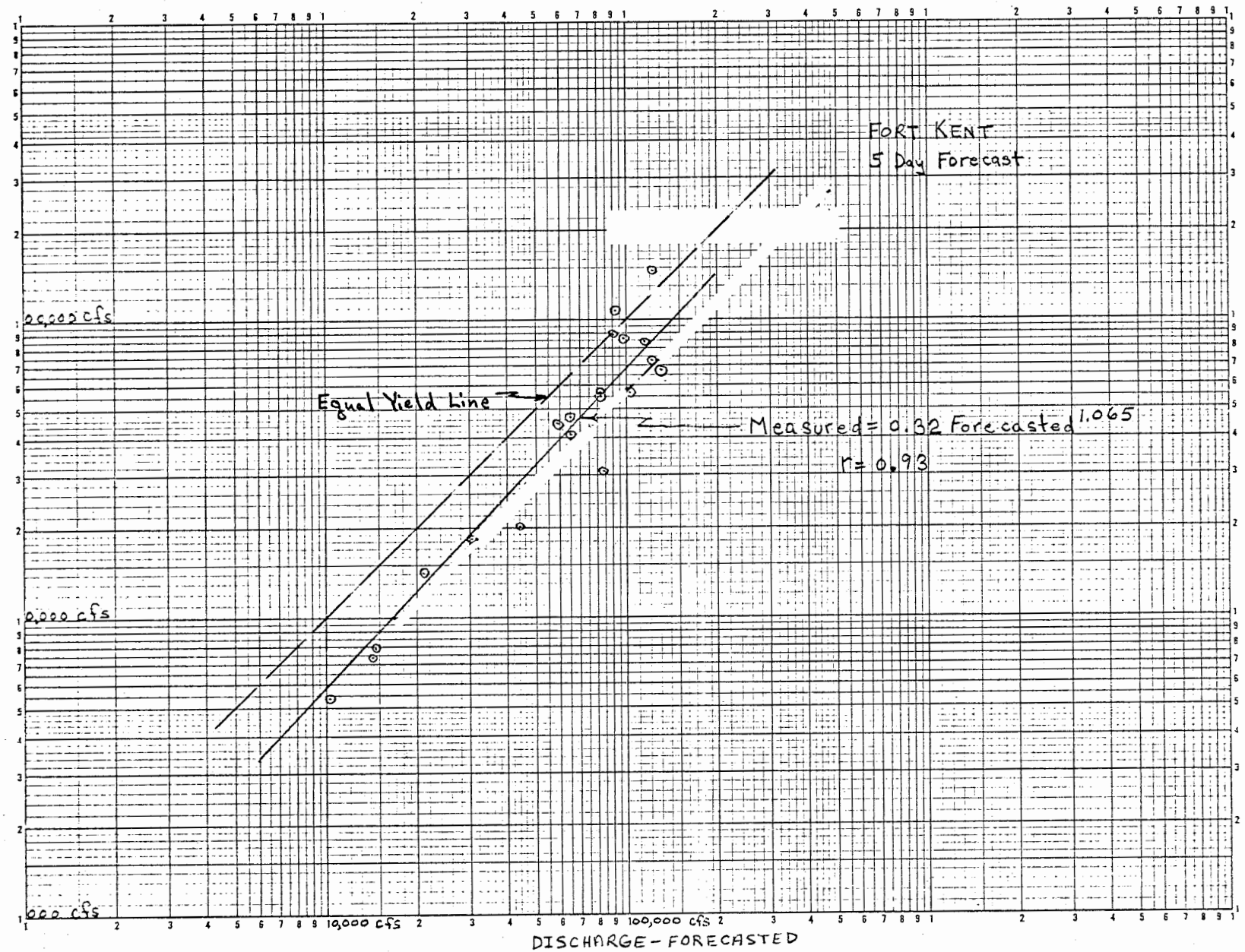





FORT KENT
2 Day Forecast
Equal Yield Line
Measured = 0.80 Forecasted^1.009
r = 0.96
DISCHARGE MEASURED
DISCHARGE - FORECASTED
1,000 cfs
10,000 cfs
100,000 cfs
HEWLETT-PACKARD
9280-0171
FOR USE ON HEWLETT-PACKARD RECORDERS
3 CYCLE LOG X 4 CYCLE LOG


FORT KENT
3 Day Forecast
Measured = 1.09 Forecasted 0.971
r = 0.93
Equal Yield Line
DISCHARGE - MEASURED
DISCHARGE - FORECASTED
10,000 cfs
100,000 cfs
HEWLETT-PACKARD
9820 0171
FOR USE ON HEWLETT PACKARD RECORDERS
3 CYCLE LOG X 4 CYCLE LOG

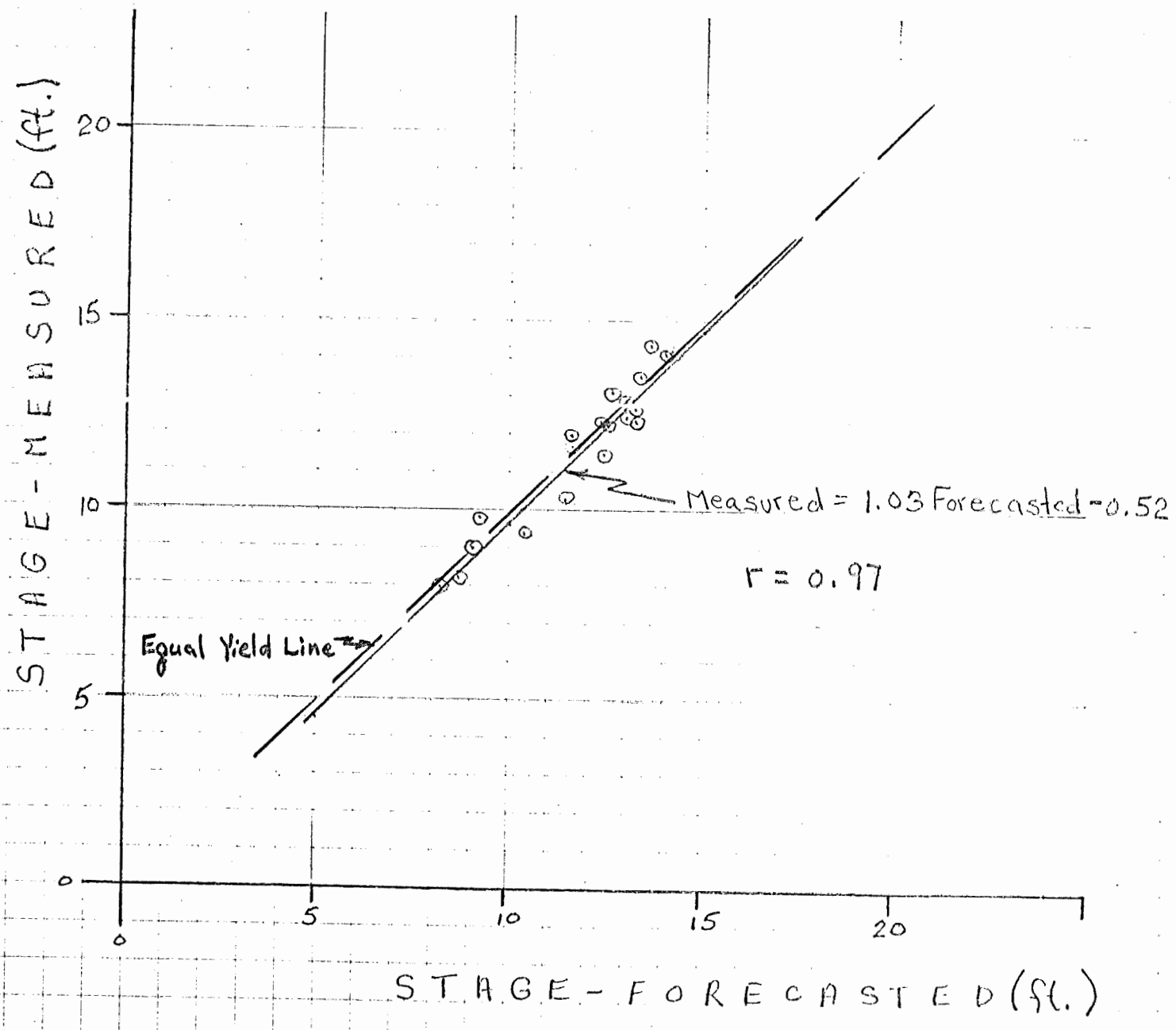



FORT KENT
5 Day Forecast
Equal Yield Line
Measured = 0.32 Forecasted 1.065
r = 0.93
DISCHARGE - MEASURED
DISCHARGE - FORECASTED
10,000 cfs
100,000 cfs

APPENDIX C

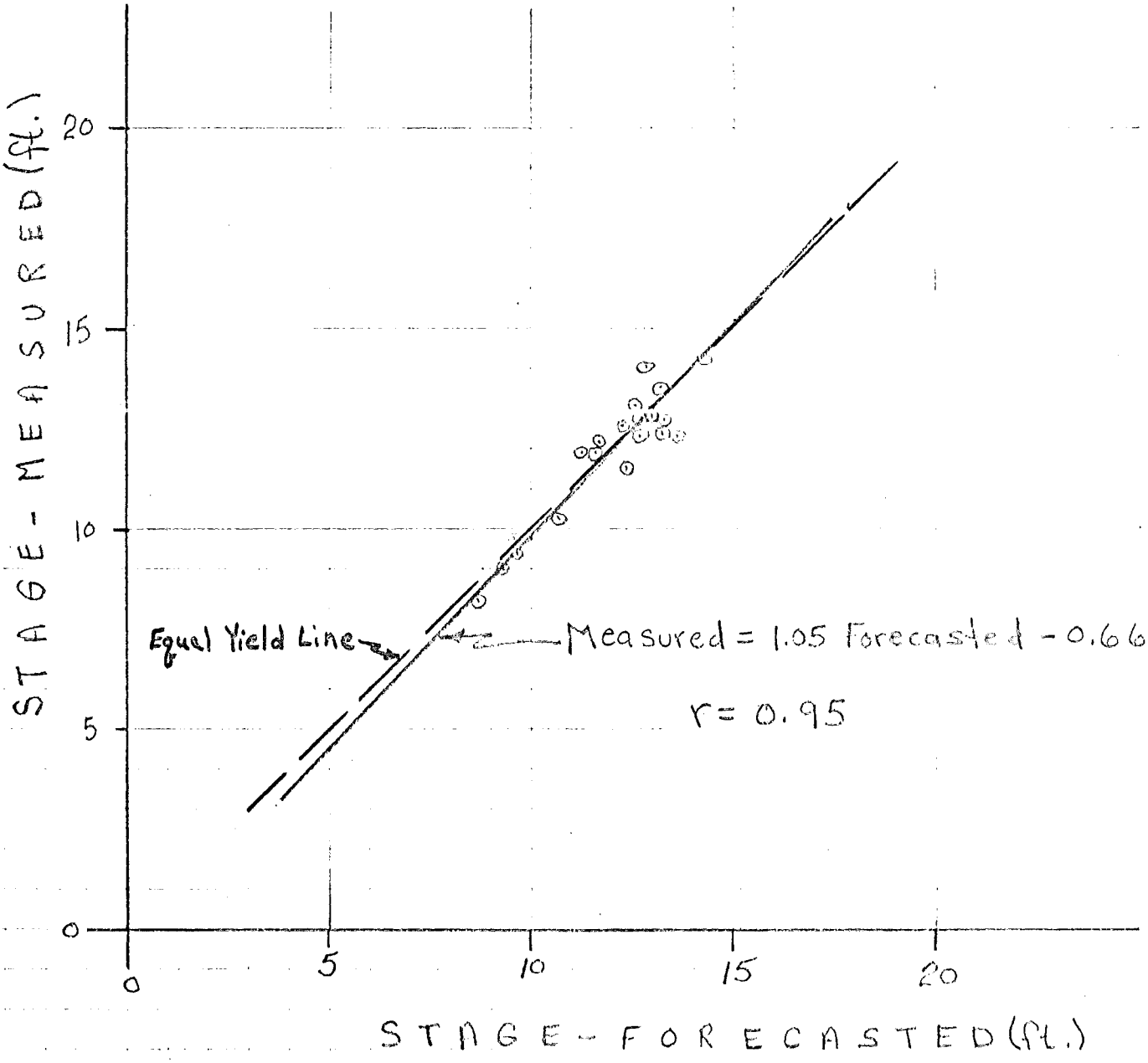
OAK POINT

1 Day Forecast



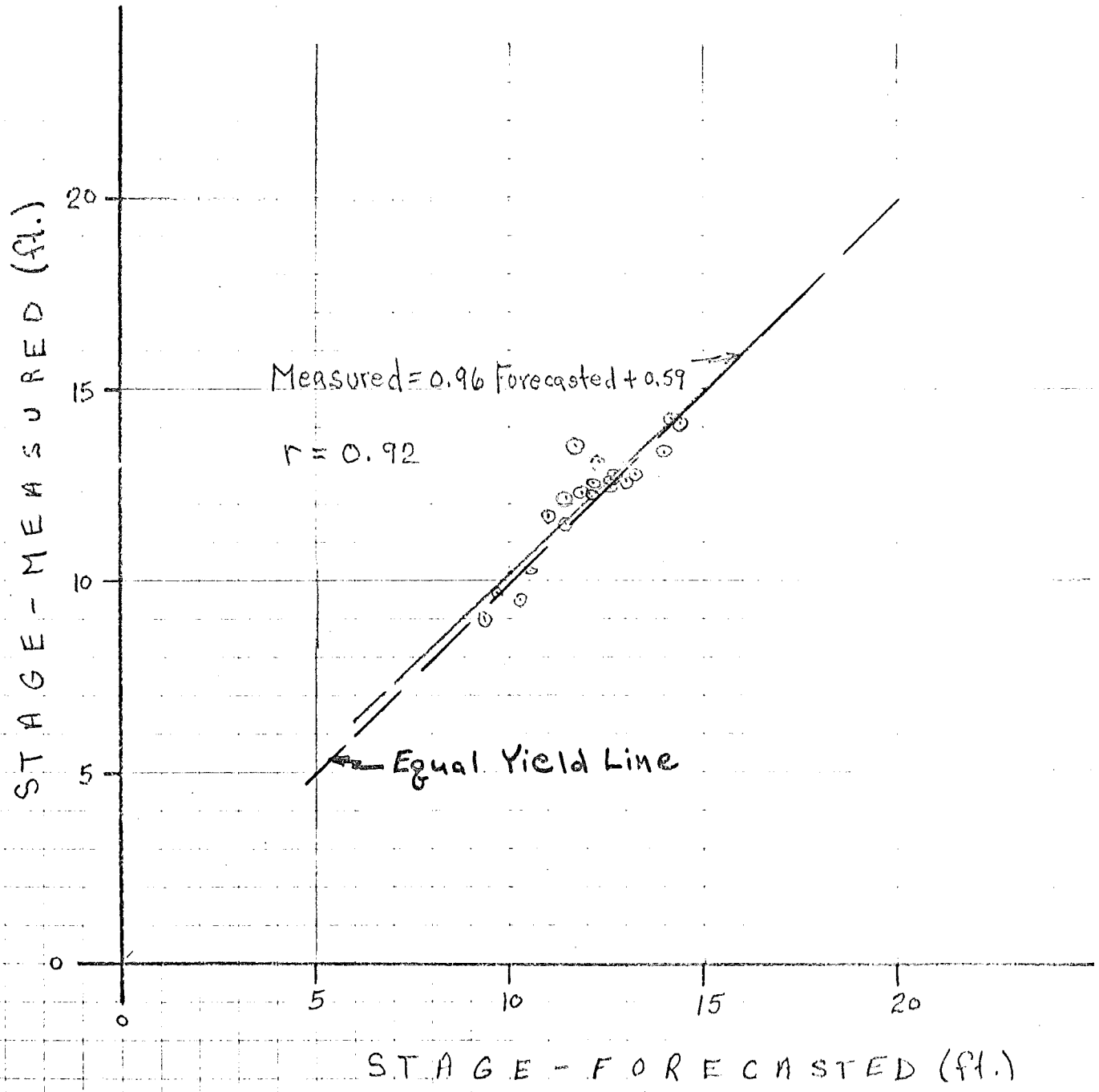
OAK POINT

2 Day Forecast



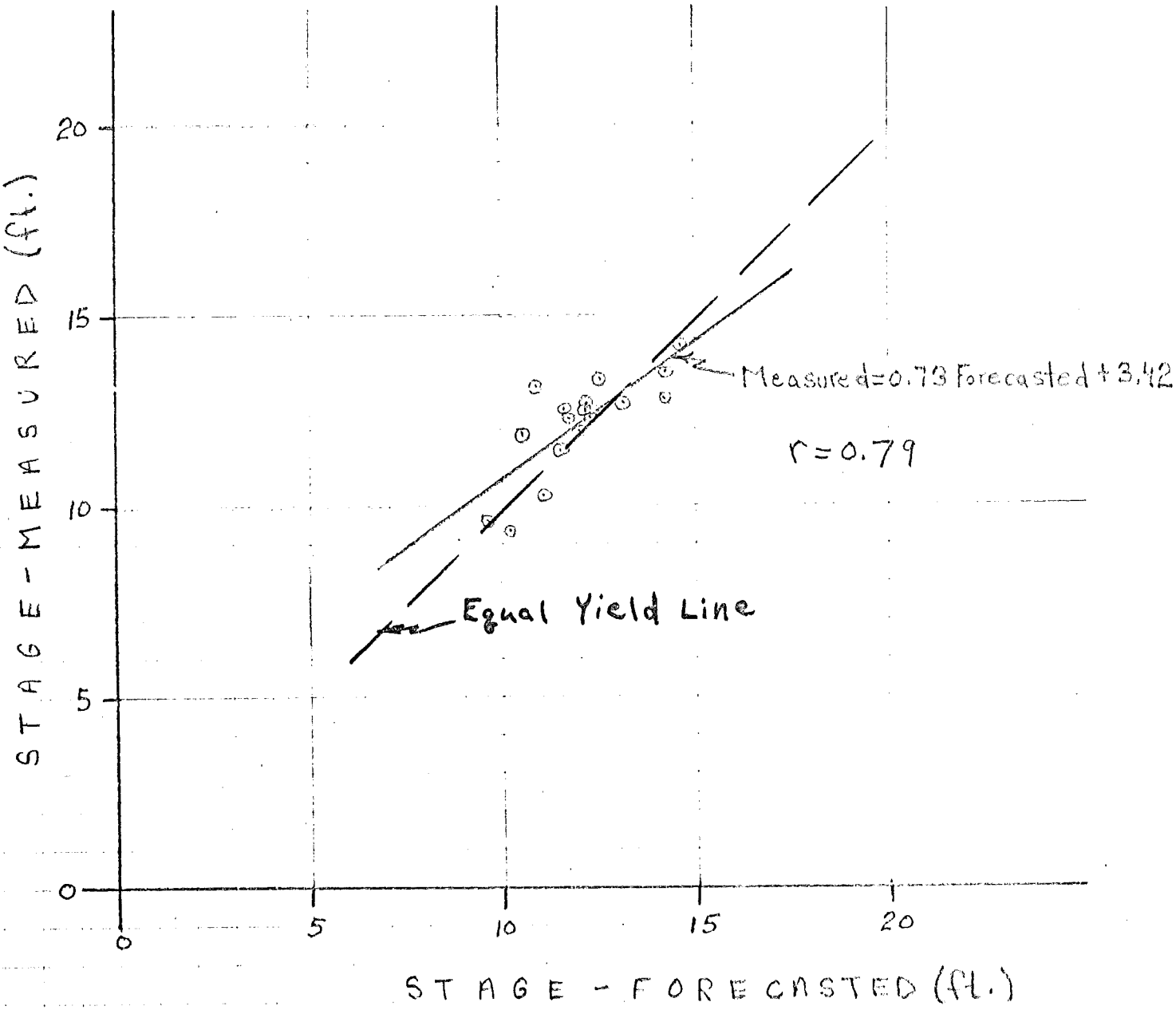
OAK POINT

3 Day Forecast



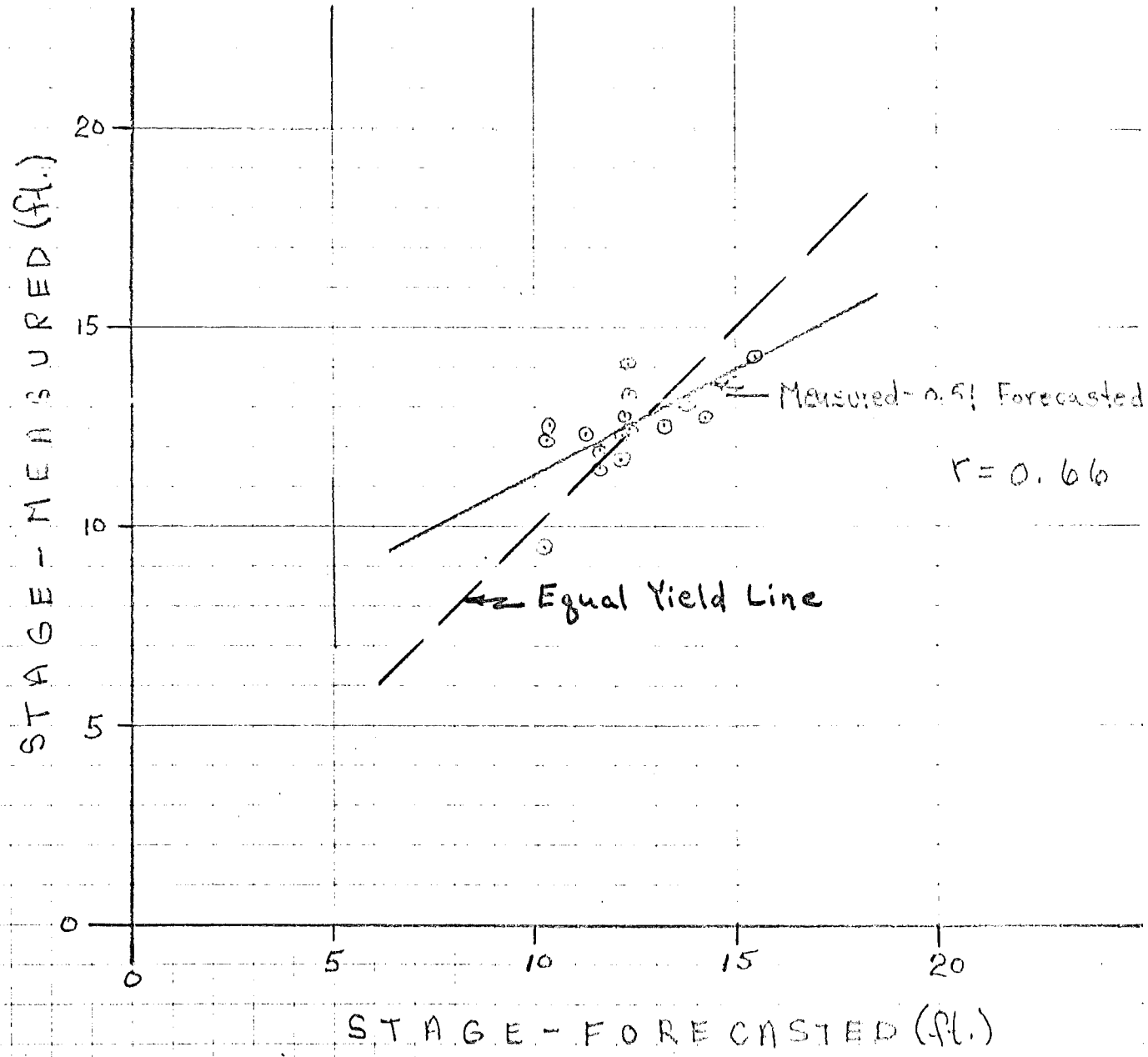
OAK POINT

4 Day Forecast



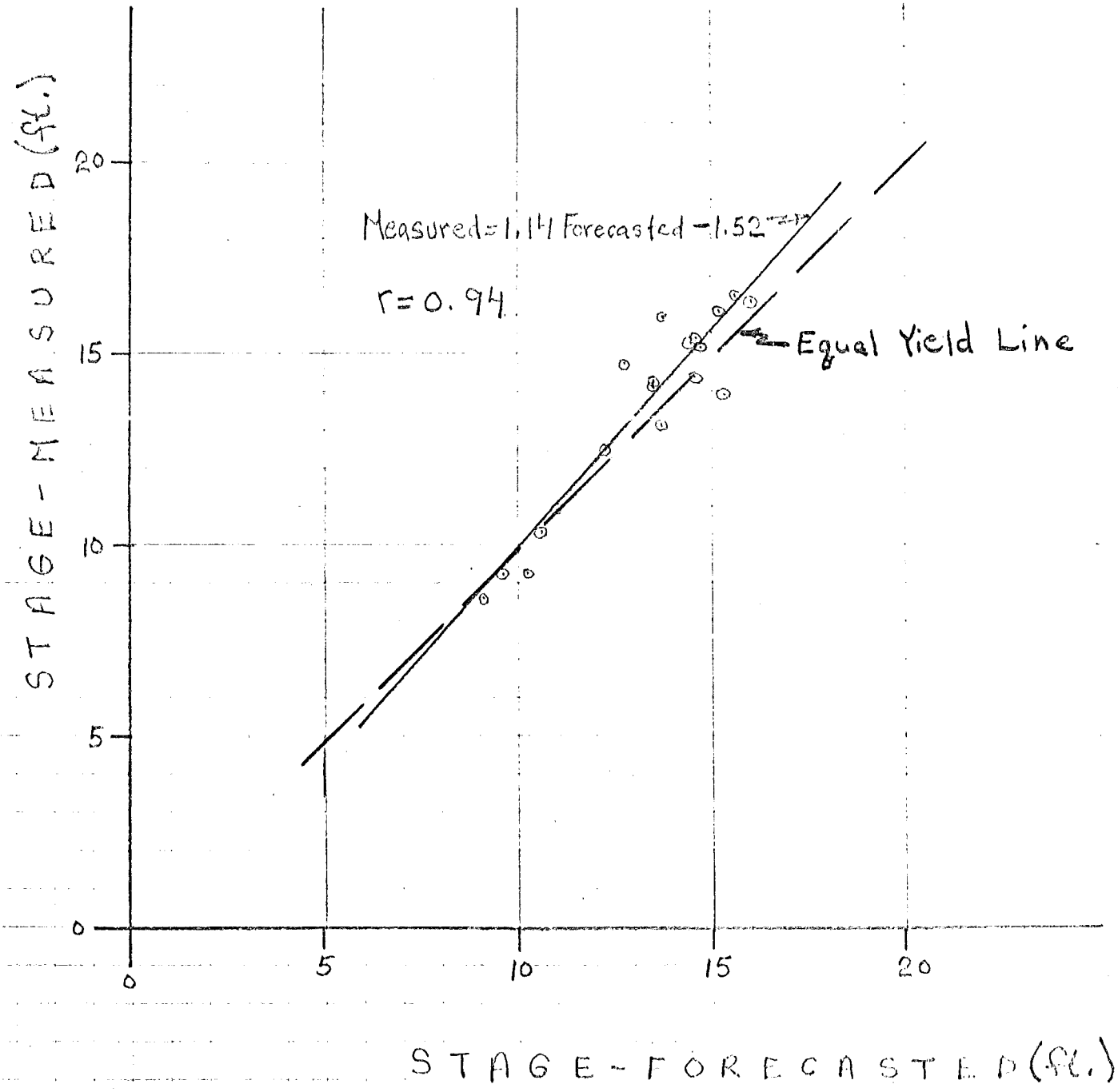
OAK POINT

5 Day Forecast



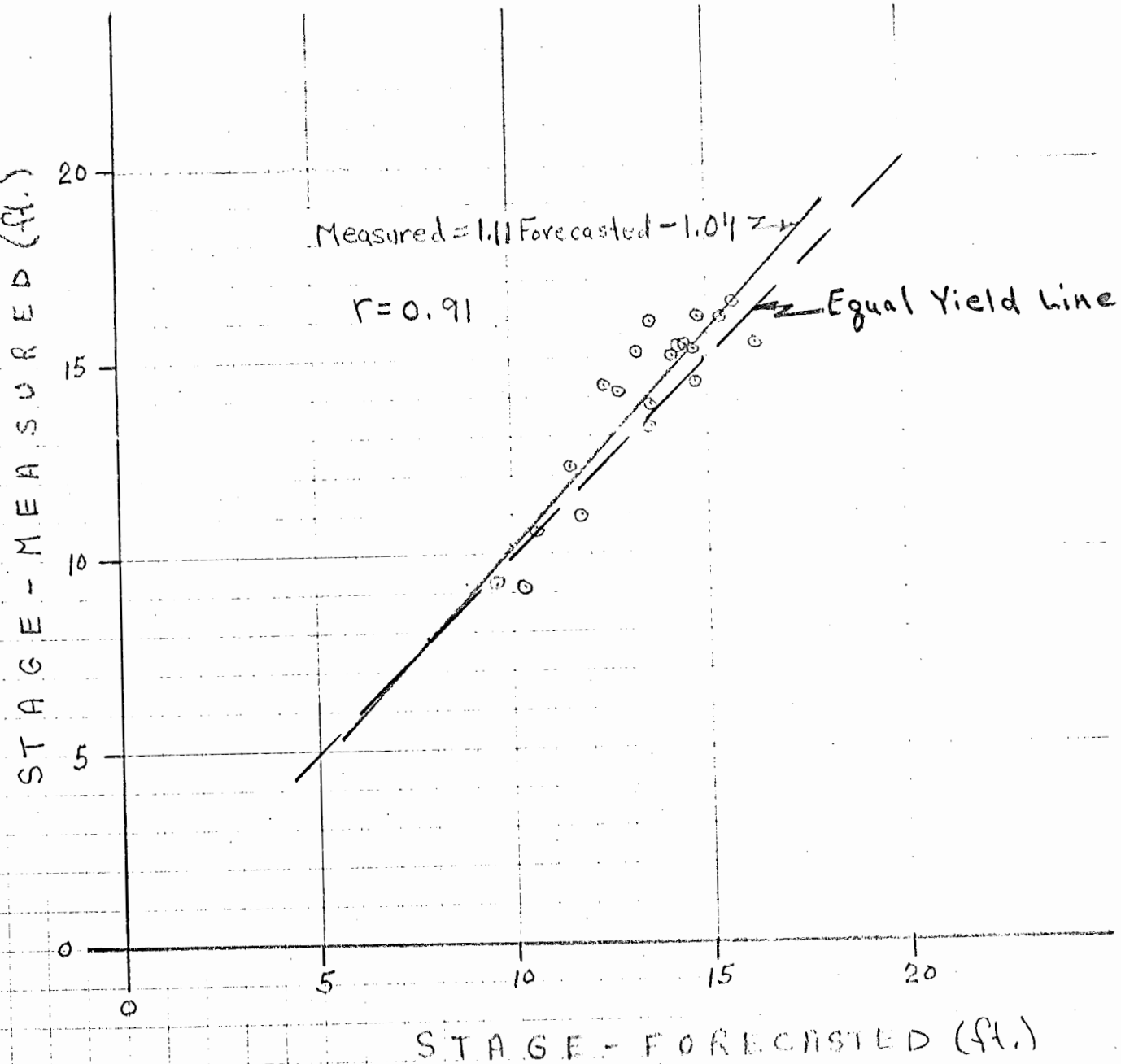
JEMSEG

1 Day Forecast



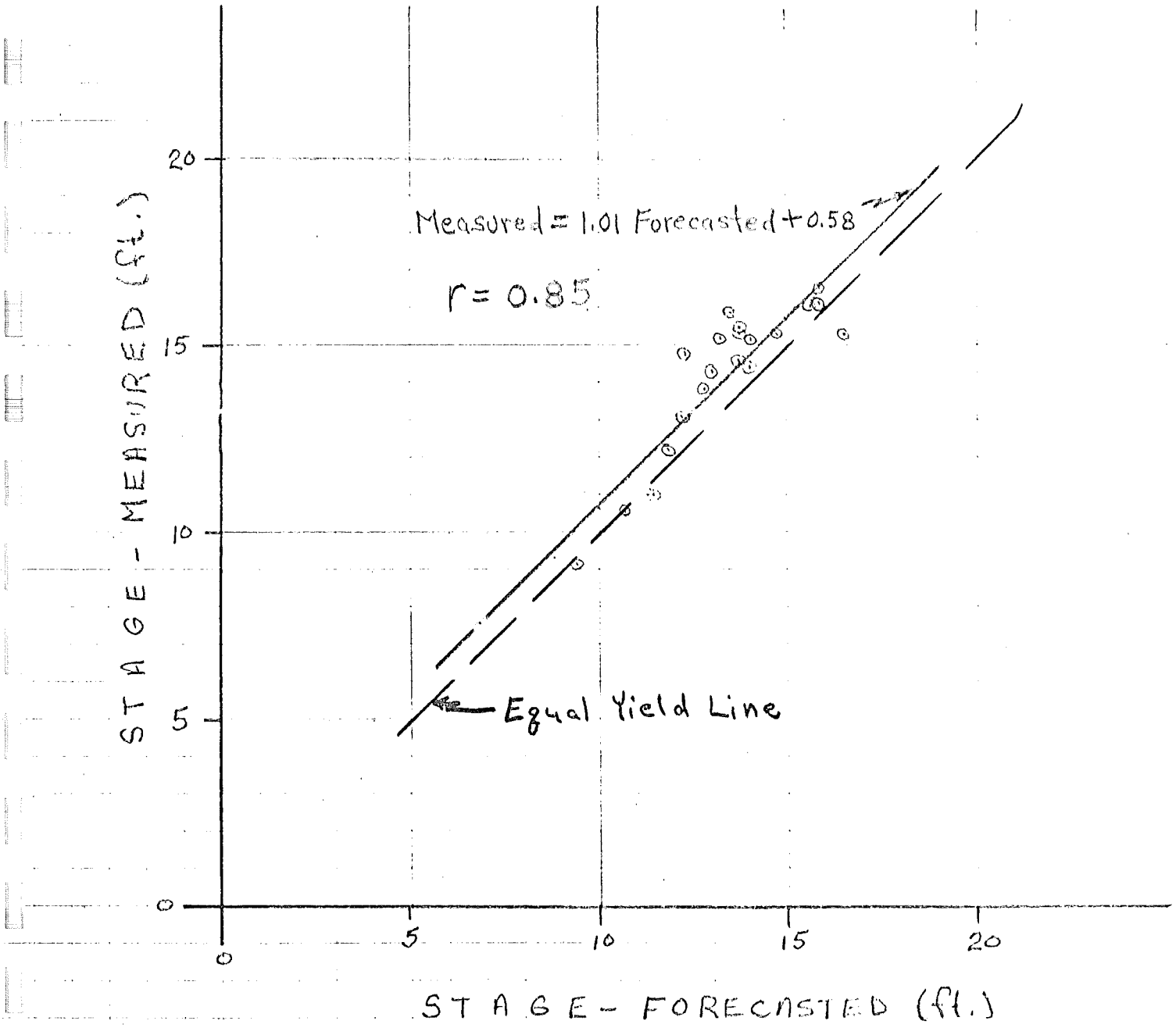
JEMSEG

2 Day Forecast



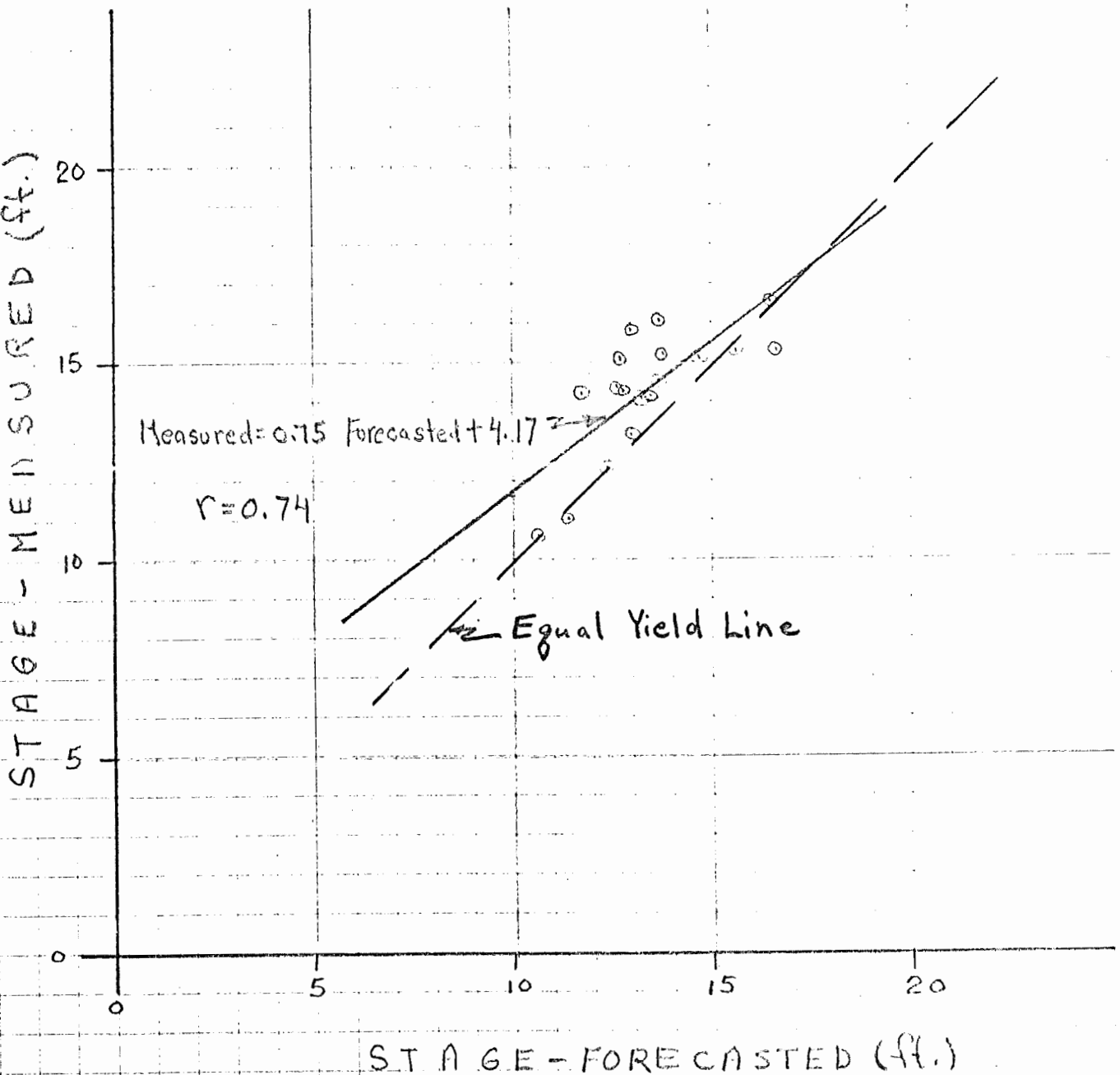
JEMSEG

3 Day Forecast



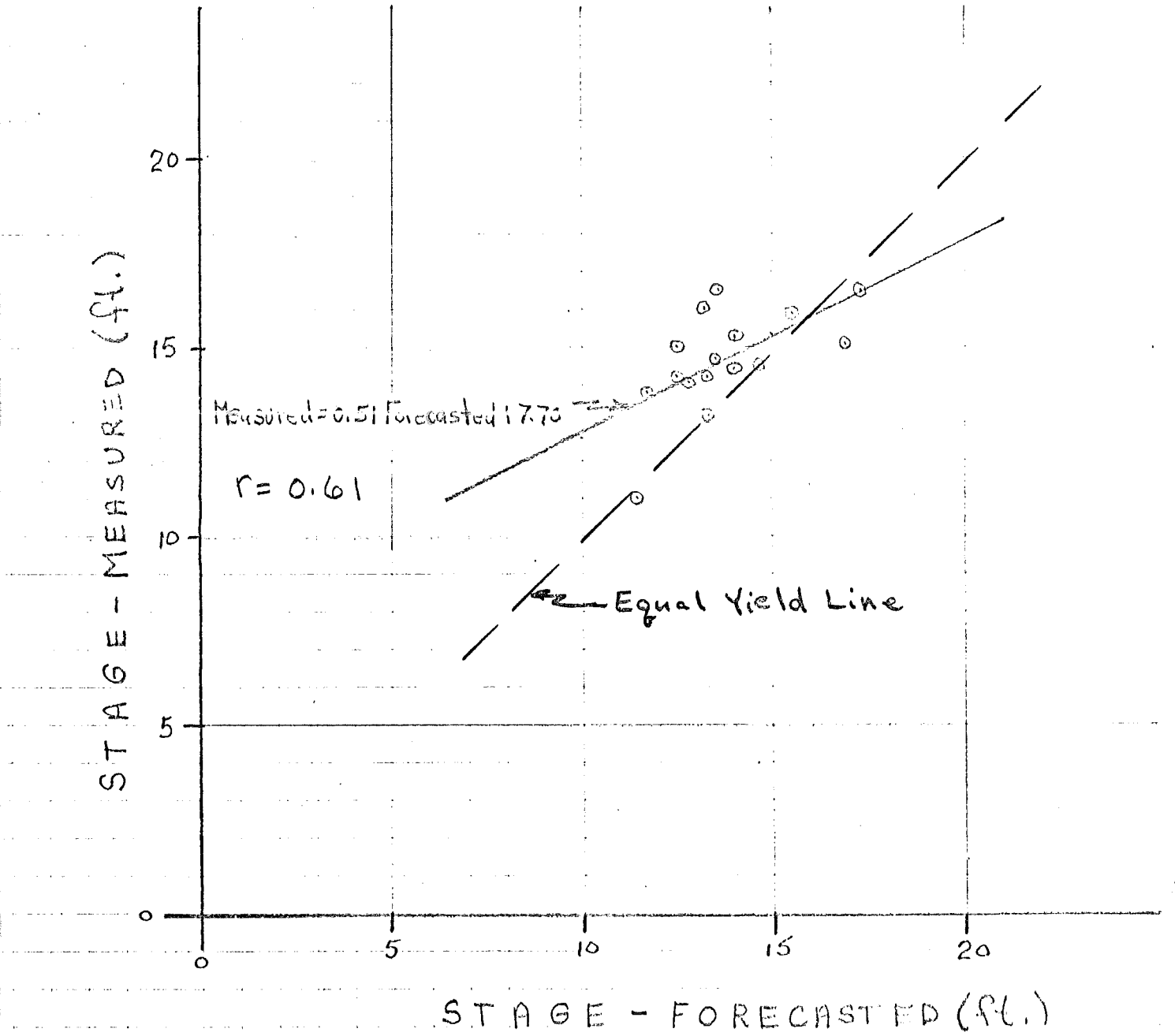
JEMSEG

4 Day Forecast



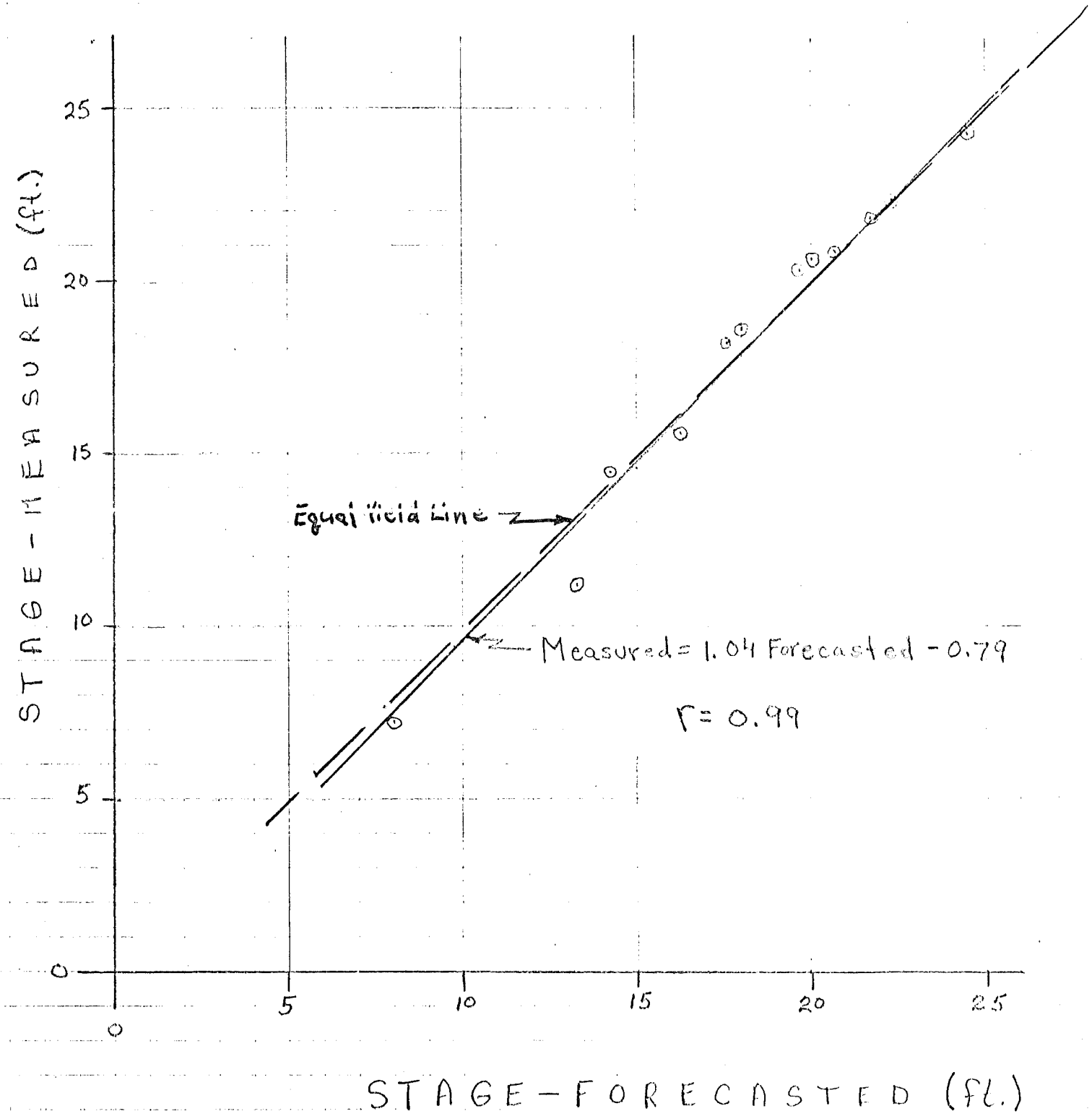
JEMSEG

5 Day Forecast



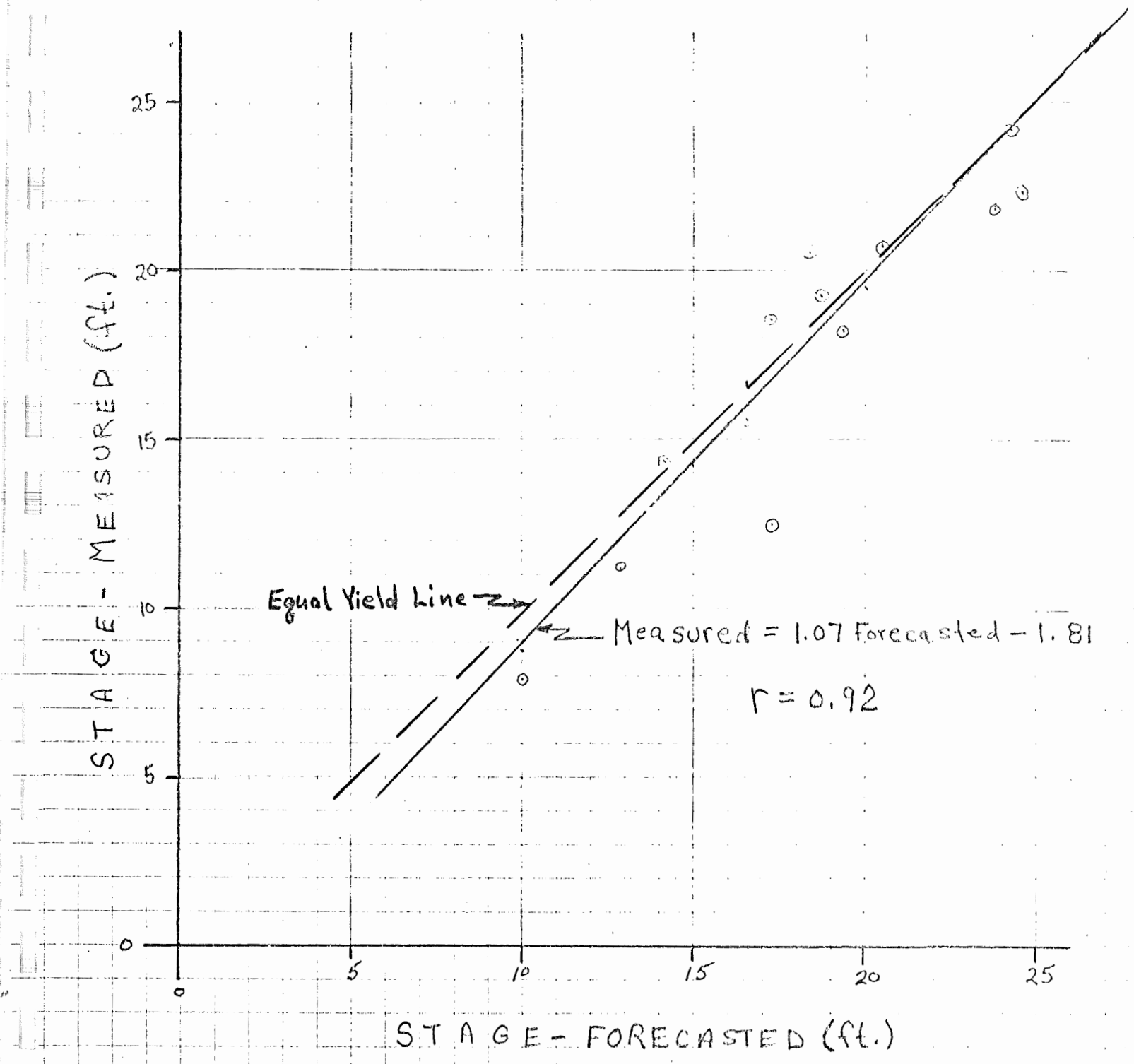
FREDERICTON

1 Day Forecast



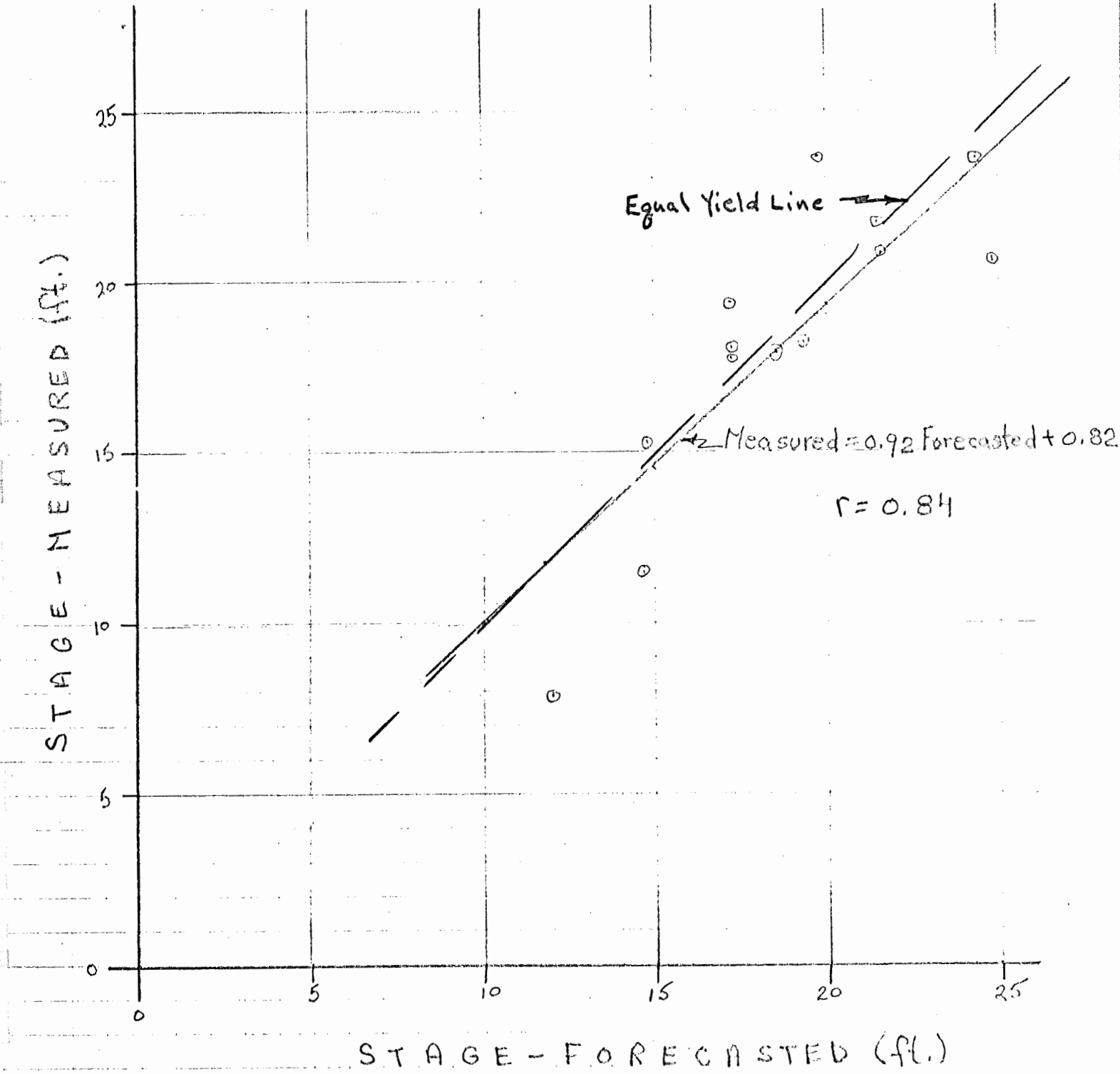
FREDERICTON

2 Day Forecast

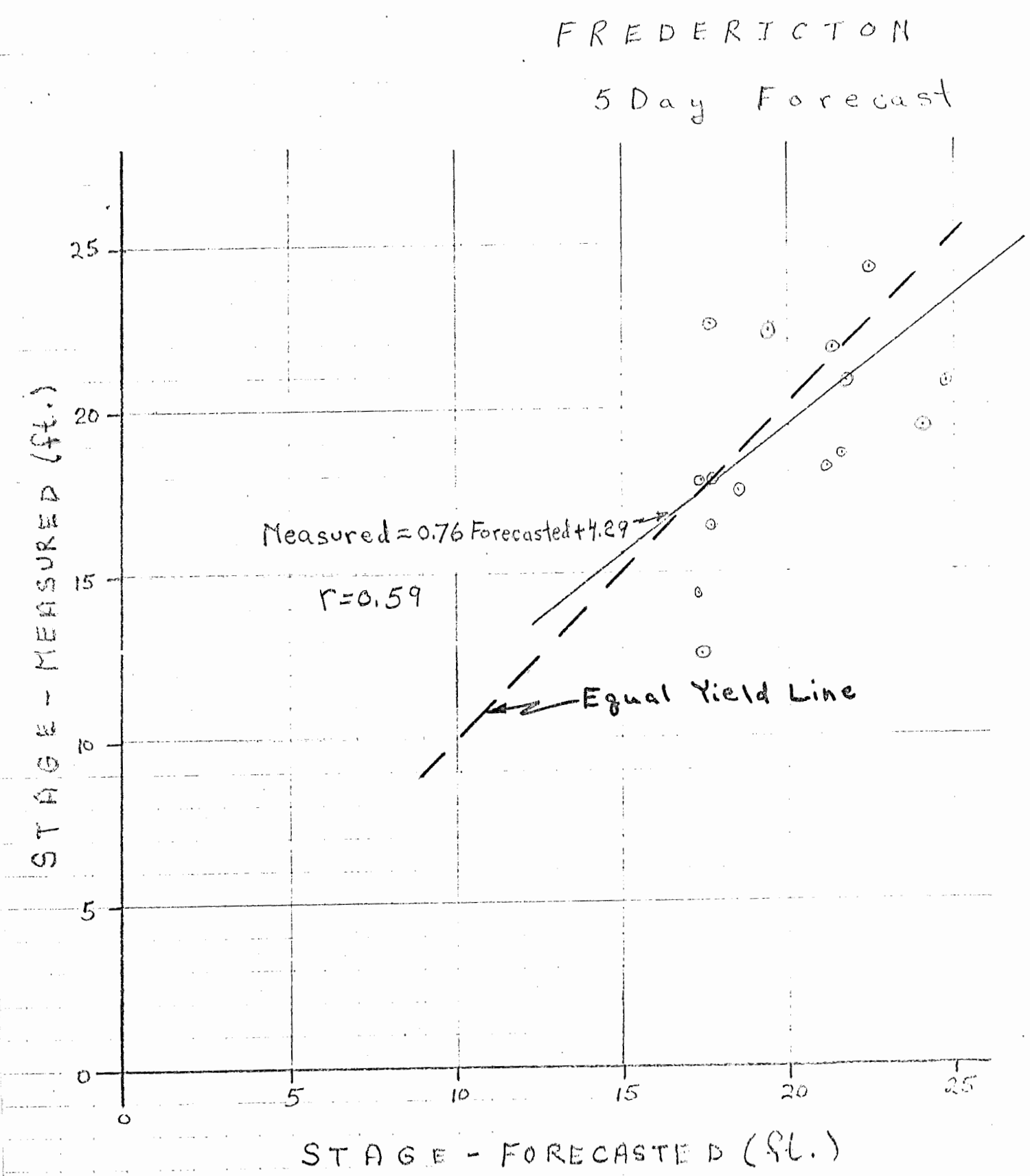


FREDERICTON

3 Day Forecast



FREDERICTON
4 Day Forecast
STAGE - MEASURED (ft.)
STAGE - FORECASTED (ft.)
0
5
10
15
20
25
Measured = 0.72 Forecasted + 1.47
r = 0.62
Equal Yield Line


FREDERICTON
5 Day Forecast
STAGE - MEASURED (Ft.)
STAGE - FORECASTED (Ft.)
Measured = 0.76 Forecasted + 4.29
r = 0.59
Equal Yield Line
0
5
10
15
20
25