U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 12

VERIFICATION DATA FOR THE 3-LEVEL NUMERICAL FORECASTS OF 1955-56

L. P. Carstensen
JNWP
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE NO. 12
JNWP
NATIONAL METEOROLOGICAL CENTER

VERIFICATION DATA FOR THE 3-LEVEL NUMERICAL FORECASTS OF 1955-56

L. P. Carstensen

INTRODUCTION

Verification scores were computed for the 3-level numerical forecasts (the 9-7-4 model) from 20 June 1955 through 30 June 1956. Although some minor changes were introduced into the model during the course of the year, the scores are essentially homogeneous. For the quantities considered this may be the most extensive set of verification scores in existence. Some of these results are of continuing interest.

In the development of numerical weather prediction at JNWP toward larger areas and greater length of forecast, some of the difficulties which restricted the 9-7-4 model have been reduced, but other difficulties have been introduced or magnified in importance. The net result is that the quality of forecasts for other elements than the 500-mb level may not have been greatly improved. These scores are then a convenient reference level against which to measure improvement. They may also yield some information as to whether a proposed new output may be expected to have values; for example, whether a forecast of a given element would be better than persistence.

DESCRIPTION OF THE SCORES AND VERIFICATION AREAS

All verification scores were computed separately once daily for three overlapping areas, and are on file in JNWP. Two of these areas (A and A+B) together with the total forecast area of the 3-level program are indicated in Fig. 1. The third area (not outlined in Fig. 1) includes all grid points (146 in numberL over the continental U. S. out to one mesh length beyond the boundary of the country. Only the scores for this last area will be summarized here, the others having given quite similar results.

In the verification of a forecast for a pressure level, eight quantities were computed. With the exception of the S_1 score, the quantities are all based on geostrophic wind components determined from grid point values over a distance of 2 mesh lenghts. S_1 is a number; the units of the remaining quantities are knots. The verification quantities are as follows:

1. S_1. This is a score very similar to the S_1* described by Wobus and Teweles [1]. One of the more important differences is that here the verification (observed) values are from heights interpolated to grid points rather than from values observed at selected stations.

---

* $S_1 = \Sigma |F-O| / \Sigma L|O| \text{ or } |F|$

S_1 is the sum of the absolute valuer of the errors of the forecast (F) pressure differences between selected stations divided by the sume of the absolute values of the larger (L) of the forecast or obsered (O) pressure difference between the stations.

Verification Data for the 3-Level Numerical Forecasts of 1955-56

Contents

2. O-F. This is a root mean square (RM) of the vector difference between the forecast and the observed geostrophic wind.

3. O-N. The samce score as 2) with the normal substituted for the forecast.

4. O-I. The same score as 2) with the initial substituted for the forecast.

5. F-I. This is the RMS of the geostrophic vector wind difference between the forecast and the initial.

6. F. The RMS of the geostrophic forecast wind.

7. O. The RMS of the geostrophic observed wind.

8. I. The RMS of the geostrophic initial wind.

DISCUSSION OF THE SCORES

The score S_1 is a relative error and is fairly closely related to the ratio (F-O)/O. This is illustrated in Fig. 2, where a number of randomly selected pairs of S_1 and (F-O)/O are plotted.

Verification scores of an RMS character should be used with some caution. It is frequently possible to obtain a better rating with a carelessly selected score by taking advantage of its response to various factors. Thus underforecasting change frequently results in a better score than a correct amount of change incorrectly placed. Similarly pure smoothing of a field may improve an RMS score. The F-I item could possibly be used to "normalize" forecast scores in this respect, but this has not been done here.

One of the disadvantages of such scores as S_1 or (F-O)/O as applied here is a variation with height. Although the score is better (lower) at the higher elevation the "skill" of the forecast is not necessarily better. The objection may perhaps be brought out by examining the variation of (O-N)/O. This score is also better (lower) at the higher elevations. The improved score is clearly due to the normal increase of zonal winds with elevation, so that a larger part of the "forecast" is almost automatically correct at the higher levels.

Although scores of this nature should be fully satisfactory if the forecast charts were highly accurate, there is some question, at our present level of accuracy, as to how much they look at the same properties as a synoptic meteorologist does. Certainly at this stage they should not be taken as an exclusive measure of quality.

VERIFICATION DATA

The method of producing the numerical forecasts has been described in the Bulletin of the AMS June, 1957 [2]. The data levels analyzed were 1,000, 700, and 400 mb. From these a 900-mb height was determined in the same manner as described below for normal charts. The forecast was then carried out using the 9-7-4 levels.

The normal charts used in the verification were from normal height charts of the 500 and 700-mb levels and the 700-1,000-mb thickness. This information is sufficient to determine a static stability value at each grid point. The assumption that this stability is constant through the required thickness is sufficient to determine the normal values at each grid point.

The 500-mb NWAC 36-hour forecasts and the corresponding 500-mb NWAC analysis for the 03Z hour were read from manuscript maps by NWAC personnel. The 15Z 500-mb verification data and also the numerically produced 500-mb forecasts were obtained by interpolation. between the 900, 700, and 400-mb heights in the manner described above for normals.

THREE-LEVEL VERIFICATION OF 24-HOUR FORECASTS

In table 1, the monthly average of each score for each level for the 24-hour verifications are presented, together with the 13-month average.

It is noted that forecasts average better than persistence at all levels, but not greatly so at the 900-mb level. To investigate this feature a little further a tabulation was made for each level of the number of days through the entire period when the forecast was better than persistence. This result is shown in table II. For the entire period the percentage of time the forecast was better than persistence at each of the 3 levels is shown in Fig. 3. It is suggested that interpolation to 850-mbs or extrapolation to the 1,000-mb level would give some indication as to what to expect if the model were extended to give values at these levels. This probably also would suggest what to expect from subjective forecasts, because, as indicated below, great differences are not apparent between subjective skill and the skill of this model, although the subjective is thought to be clearly superior at the 1,000-mb level.

VERIFICATION IN TIME OF INTERPOLATED 500-MB FORECASTS

In table III the verifications of values interpolated to 500-mb from the 9-7-4 information levels at 12, 24, and 36 hours are summarized. Also included are the verification scores for the NWAC 36-hour forecasts.

As stated above, the 24-hour forecasts are verified against values interpolated from the three levels, but the 12- and 36-hour 500-mb verification heights were read directly from the NWAC manuscript maps. A systematic difference is apparent between the 24-hr scores on the one hand and the 12- and 36-hr scores on the other. Thus the RMS of the difference of the normal chart from the observed (n-o) is 25.96 kts in the one case and 28.25 kts in the other. This suggests that the interpolated values are smoother (more like the normal) than the directly read values. This in turn suggests caution in interpreting the indicated rate of error growth with time. The relatively low RMS of the forecast error at 24 hours may in some part be due to different characteristics of the verification data as compared with the 12- and 36-hr verification data.

In general, smoothness appears to be a problem in verification For a comparison of two forecasts with verification data to have any precise meaning, it is necessary that the forecasts incorporate similar smoothing (or have the same high frequency components removed.) For a comparison of a series of forecasts made and verified against one series of observation with another series of forecasts verified against another series of observations to have precise meaning, it is necessary that a consistent smoothness prevail. For example, a forecast smoothed so that only the zonal wind remains, show better RMS scores than would any procedure which did not introduce smoothing.

It is noted that the subjective 36-hour forecast underpredicts changes while the numerical does not. In fact, the numerical forecast calls for almost as much change in 24 hours as the subjective does in 36 hours. This alone, as suggested above, may possibly be sufficient to account for the slightly better S_1 and O-F scores enjoyed by the subjective forecasts in this comparison. Practically speaking, if these scores were to be taken as the measure of skill, experiments should be carried on with the numerical forecast to reduce the forecast change an amount which would maximize this skill.

Tables I and II can in a sense be combined in the manner indicated in table IV. The values available give some indication of what to estimate in the table for the empty spaces, either for this numerical model or for subjective forecasts where information may not otherwise be available.

It may also be noted in passing that the 3-level forecast winds were slightly (1%) too weak at the 400-mb level, a little too strong at 700-mb (8%), and appreciably too strong (32%) at the 900-mb level.

REFERENCES

1. Teweles, Sidney, Jr. and Wobus, Hermann B. Verfication of Prognostic Charts. Bulletin, AMS, Vol. 35, No. 10 pp 455-463, December 1954.

2. Staff Members, Joint Numerical Prediction Unit, One Year of Operational Numerical Weather Prediction, Part II.

Table 1
Continental USA 24 hr Verifications

		June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	June	Ave
	S_1	.37	.39	.37	.37	.35	.32	.30	.35	.27	.27	.31	.33	.36	.335
	O-F	16.3	13.5	12.0	17.2	19.5	23.6	22.4	21.4	21.1	20.0	19.3	16.3	13.4	18.15
	O-N	28.0	32.4	23.4	27.8	37.6	37.3	37.1	38.9	39.8	36.7	34.6	30.3	26.7	33.12
400 mb	O-I	21.2	16.2	16.6	24.5	33.1	39.8	30.8	30.0	34.3	36.4	32.2	22.8	19.7	27.51
	F-I	22.0	15.8	15.6	24.5	32.1	40.3	33.7	33.1	37.0	37.9	32.9	22.8	19.4	28.08
	F	34.1	28.3	26.4	37.9	44.5	60.6	60.8	51.9	63.2	61.2	52.3	40.7	30.6	45.58
	O	34.4	26.4	27.0	37.6	46.9	60.4	60.6	51.4	67.6	61.3	51.2	41.8	31.7	46.02
	I	33.4	26.6	26.7	37.2	46.6	60.8	60.4	51.4	62.5	61.6	51.6	42.0	32.0	45.60
	S_1	.48	.48	.50	.49	.41	.40	.40	.42	.34	.36	.39	.44	.44	.427
	O-F	11.0	9.6	9.8	12.4	12.3	15.8	15.3	14.1	13.3	13.5	12.4	11.1	9.09	12.28
	O-N	13.1	12.2	12.8	14.1	17.8	19.1	17.6	21.4	17.8	17.7	18.3	14.4	12.79	16.08
700 mb	O-I	11.3	10.2	12.3	14.6	15.6	20.1	16.8	16.4	17.2	18.8	17.8	13.0	10.95	15.01
	F-I	11.7	10.1	11.4	15.1	16.2	22.1	19.6	17.8	20.0	20.8	18.2	13.6	11.09	15.98
	F	18.6	16.5	15.6	20.9	24.2	32.7	31.9	28.4	32.7	31.1	27.3	20.7	16.87	24.42
	O	16.8	14.5	15.0	18.9	23.5	29.8	29.2	25.9	29.9	28.8	25.5	20.5	16.42	22.67
	I	16.3	14.5	14.8	18.9	23.3	29.8	29.0	26.1	29.7	29.0	25.7	20.4	16.64	22.62
	S_1	.61	.59	.61	.62	.56	.60	.59	.58	.53	.57	.56	.57	.55	.580
	O-F	13.0	10.3	10.8	12.5	14.5	19.5	17.2	14.5	14.8	16.3	13.7	11.5	9.3	13.68
	O-N	12.1	11.0	12.3	13.4	15.4	16.8	15.6	17.7	15.0	15.8	15.4	12.2	10.6	14.10
900 mb	O-I	10.7	9.8	12.5	15.6	15.5	19.6	17.4	15.9	16.9	18.6	16.2	13.5	10.2	14.80
	F-I	12.5	11.3	12.4	17.8	19.0	25.6	21.6	18.1	21.2	22.6	18.3	15.8	11.8	17.54
	F	16.7	14.1	14.9	19.0	21.7	27.6	25.3	22.1	23.7	24.5	21.5	17.3	14.6	20.23
	O	13.8	11.7	12.6	14.7	17.0	19.5	17.8	17.3	17.6	17.8	16.8	13.8	11.6	15.54
	I	13.1	11.9	12.4	14.8	17.0	19.1	17.9	17.4	17.5	17.8	16.9	13.7	11.7	15.48

Table II

Number of numerical forecasts better than persistence (O-F), persistence forecasts better than numerical (O-I), and percentage better.

	400-mb			700-mb			900-mb		
	O-F	O-I	Equal	O-F	O-I	Equal	O-F	O-I	Equal
June	8	1	1	9	1	0	3	7	0
July	27	4		18	10	3	12	16	3
August	30	1		28	2	1	24	6	1
September	27	2	1	21	9		20	9	1
October	31	0		24	6	1	18	12	1
November	28	2		23	6	1	16	12	1
December	25	6		19	10	2	17	12	2
January	24	7		22	9		21	9	1
February	27	2		26	3		20	8	1
March	29	2		27	4		24	7	
April	29	1		26	4		18	8	4
May	30	1		24	7		25	4	2
June	30	1		21	10	1	19	12	
Total	345	30	2	288	81	8	237	124	16
%	92			78			65 1/2		

Table III

Verification of 500-mb Progs for U. S. Area

		1955 Aug.	Sept.	Oct.	Nov.	Dec.	1956 Jan.	Feb.	Mar.	Apr.	May	June	Ave.
	S_1	.35	.31	.27	.23	.23	.27	.22	.22	.25	.30	.30	.268
	O-f	8.8	11.8	11.8	14.2	13.8	13.9	13.7	12.8	12.7	12.1	10.4	12.36
	O-N	17.6	21.3	29.3	38.3	29.9	34.5	32.2	30.7	30.0	23.5	23.6	28.6
JNWP	O-I	10.6	14.7	17.6	22.8	18.0	19.0	19.8	20.8	19.6	14.8	13.4	17.37
12-hour	F-1	7.1	11.8	14.5	19.8	16.4	15.0	18.2	19.1	15.9	11.3	11.5	14.60
	F	21.8	30.0	36.1	49.4	48.6	41.4	50.7	48.8	41.8	33.4	26.8	38.98
	O	20.9	29.2	38.5	50.9	50.3	43.3	52.4	50.8	43.6	34.2	27.7	40.6
	I	21.8	29.7	36.4	48.8	48.5	41.4	49.7	49.0	41.8	34.3	26.7	38.92
	S_1	.39	.38	.34	.31	.31	.35	.26	.27	.31	.35	.35	.329
	O-F	10.5	14.2	15.1	18.7	18.3	17.7	16.2	15.8	15.2	13.8	11.1	15.15
	O-N	17.7	21.0	27.2	35.6	26.3	31.1	28.6	27.4	27.2	22.4	20.8	25.94
JNWP	O-I	14.1	19.2	23.6	30.3	23.4	23.1	25.4	27.8	25.9	17.3	16.7	22.44
24-hour	F-1	12.9	19.9	23.1	32.2	26.9	23.8	27.8	29.4	26.0	18.6	16.9	23.41
	F	21.8	30.8	35.7	50.2	49.2	42.2	51.1	48.9	42.1	33.1	26.8	39.26
	O	22.0	29.9	37.0	48.9	48.3	41.5	49.8	48.8	41.4	31.9	26.4	38.72
	I	21.8	29.8	36.4	48.8	48.5	41.4	49.7	49.0	41.8	34.3	26.7	38.93
	S_1	.51	.51	.49	.45	.43	.50	.38	.40	.44	.50	.52	.466
	O-f	14.4	20.4	23.6	29.7	27.3	27.7	15.2	25.1	23.7	21.2	18.3	23.33
	O-N	17.6	21.3	30.2	38.0	30.6	33.9	32.2	30.8	29.8	23.8	22.5	28.24
JNWP	O-I	17.0	23.2	31.5	38.6	30.9	30.1	32.4	35.2	33.2	23.0	23.2	28.94
36-hour	F-1	17.2	24.0	28.7	40.2	33.4	30.4	32.7	34.8	32.2	23.2	22.4	29.02
	F	22.1	31.6	35.7	51.1	49.5	43.4	51.1	48.6	42.7	33.8	26.8	39.67
	O	21.2	29.3	39.1	50.7	50.4	43.4	52.3	50.8	43.1	34.0	27.1	40.12
	I	21.8	29.8	36.4	48.8	48.5	41.4	49.7	49.0	41.8	34.3	26.7	38.93
	S_1	.49	.48	.47	.43	.41	.49	.41	.42	.46	.48	.51	.459
	O-F	13.2	18.2	23.3	27.4	25.6	27.8	27.1	26.0	25.4	20.4	18.5	22.99
	O-N	17.6	21.3	30.2	38.0	30.6	33.9	32.2	30.8	29.8	23.8	22.5	28.24
NWAC	O-I	17.0	23.2	31.6	38.6	30.9	30.1	32.4	35.2	33.2	23.0	23.2	28.94
36-hour	F-1	14.9	19.7	25.5	32.0	28.2	24.8	28.1	30.4	29.0	18.8	17.9	24.48
	F	20.8	29.0	36.1	47.3	47.5	42.8	50.3	47.8	41.9	33.4	26.9	38.53
	O	21.2	29.3	39.1	50.7	50.4	43.4	52.3	50.8	43.1	34.0	27.1	40.13
	I	21.8	29.8	36.4	48.8	48.5	41.4	49.7	49.0	41.8	34.3	26.7	38.93

Table IV

Average Values of the Score O-F

Extension	Level 900 mb	700 mb	500 mb	400 mb
12-hour			12.36	
24-hour	13.68	12.28	15.15	18.15
36-hour			23.33	

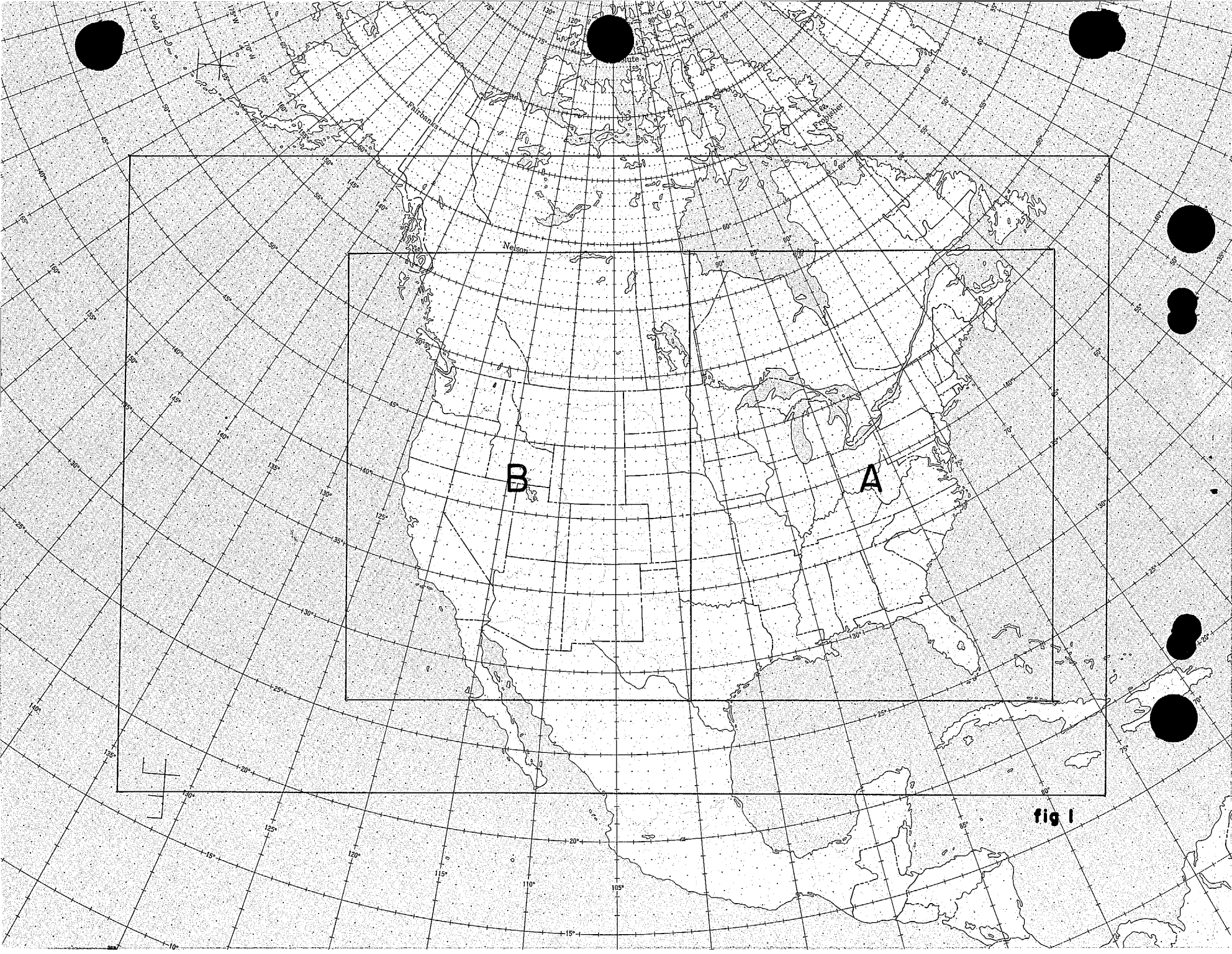

fig I

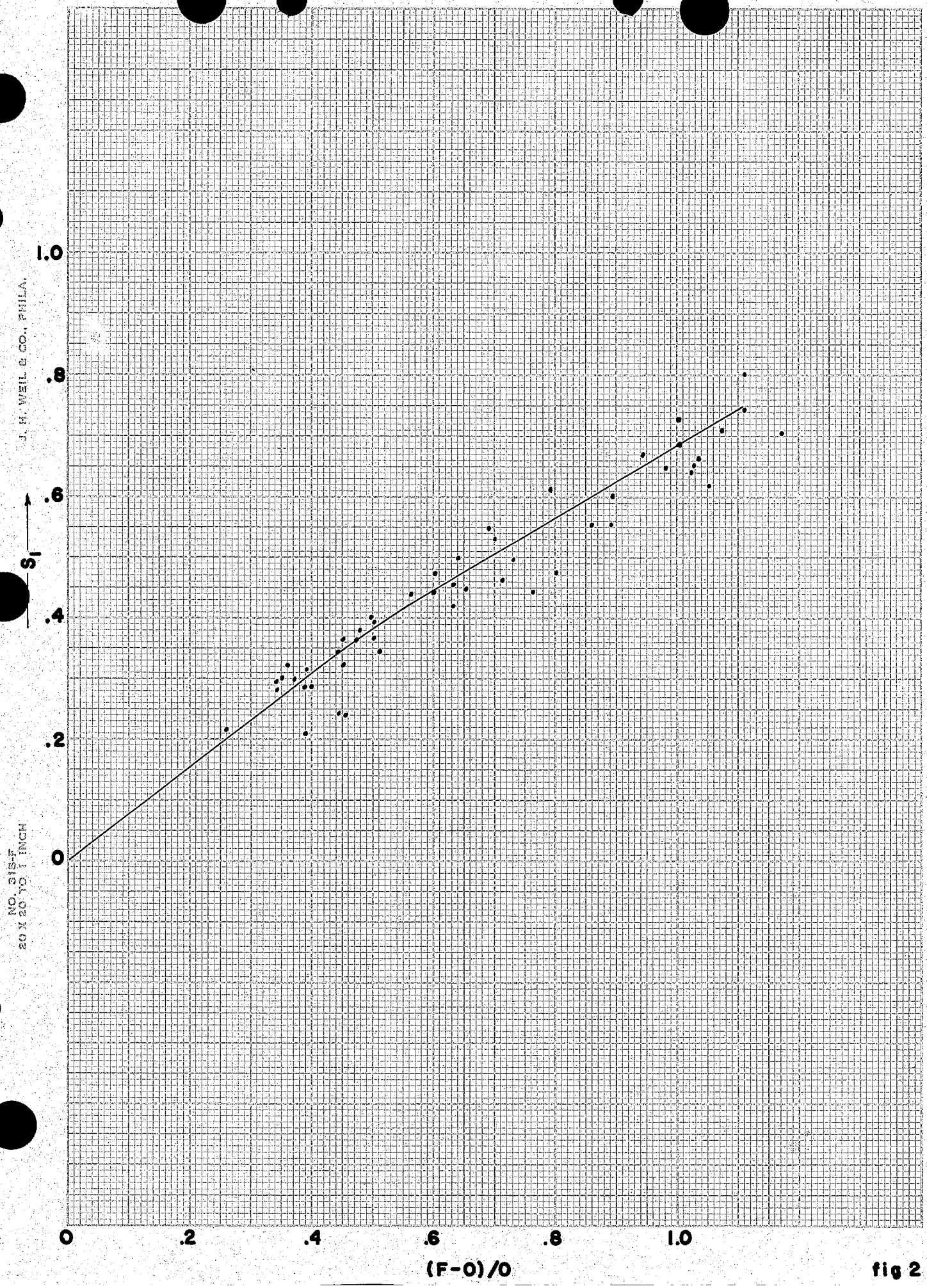
1.0
.8
.6
.4
.2
0
S_1
0
.2
.4
.6
.8
1.0
(F-O)/O
J. H. WEIL & CO., PHILA.
NO. 318-F
20 X 20 TO 1 INCH

fig 2

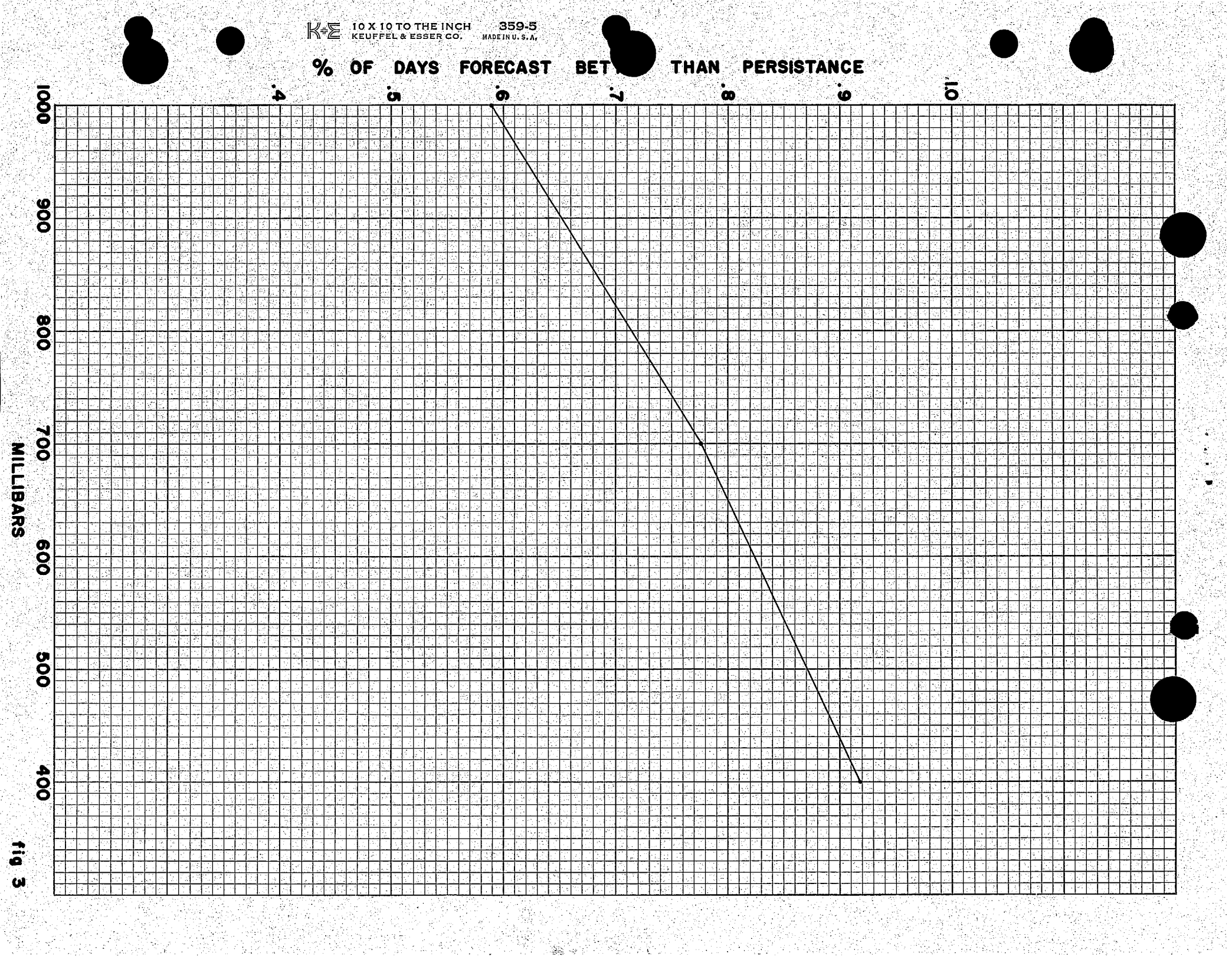
K+E 10 X 10 TO THE INCH 359-5
KEUFFEL & ESSER CO. MADE IN U.S.A.
% OF DAYS FORECAST BET THAN PERSISTANCE
.4
.5
.6
.7
.8
.9
1.0
1000
900
800
700
600
500
400
MILLIBARS

fig 3