U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
SYSTEMS DEVELOPMENT LABORATORY

TDL OFFICE NOTE 74-8

COMPARATIVE VERIFICATION OF SUM, PE, AND LFM MODELS FOR THE WINTER SEASON 1973-1974

Robert J. Bermowitz and Thomas H. Grayson

JULY 1974

Comparative Verification of SUM, PE, and LFM Models for the Winter Season 1973-1974

Robert J. Bermowitz
Thomas H. Grayson

INTRODUCTION

The Subsynoptic Update Model (SUM) (Grayson and Bermowitz, 1974) replaced the Subsynoptic Advection Model (Glahn and Lowry, 1972) at the National Meteorological Center and has been running twice daily since the beginning of October 1973. Short range forecasts of sea-level pressure and quantitiative precipitation are produced by SUM, and forecasts up to 11 hr are transmitted via FOFAX.

In order to determine the relative value of these forecasts, a comparative verification program was initiated in which SUM forecasts are compared to those of the Primitive Equation (PE) (Shuman and Hovermale, 1968) and Limited Area Fine Mesh (LFM) (Howcroft and Desmaris, 1971) models. Currently, SUM objectively updates the PE model through use of the latest surface data. If desired, it could be used to update the LFM model. Since SUM is an update model it can only be considered useful if it provides better short range forecasts than those available from the PE and LFM models verifying at the same time. The purpose of this paper is to report the results of this continuing comparative verification, specifically for the winter season of 1973-1974.

VERIFICATION PROCEDURES

Verification of sea-level pressure was performed for the period October 1973-March 1974--a total of 271 cases--at two forecast projections for both 0000 GMT and 1200 GMT cycles. Five-and 17-hr SUM forecasts were compared to corresponding 12- and 24-hr PE and LFM forecasts. It would have been desirable if the verification could also have been performed at an 11-hr (18-hr) SUM (PE and LFM) projection. However, this would have required considerable effort since manual extraction of values from sea-level pressure charts would have been necessary for that projection. Observed values are normally received on punch cards from NMC for two times, 0000 and 1200 GMT.

The S1 score (Teweles and Wobus, 1954) and the mean absolute error (MAE) in mb were computed at 31 points over the United States and adjacent areas as shown in Figure 1 by the open circles. In order to determine if any regional differences exist, the verification was broken down into eastern and western portions with 90° W longitude taken as the dividing line. This resulted in 12 verification points over the eastern U.S. and 19 verification points over the western U.S.

Although quantitative precipitation forecasts for the 0-5 and 5-11 hr periods are transmitted via FOFAX, SUM's precipitation model was primarily designed to specify the most likely areas of precipitation. Quantitative precipitation is strictly a by-product of the model. It is now a well established fact the SUM underestimates the actual amounts. Forecast amounts during a 6-hr period rarely exceed .25 in. In view of this, the occurrence ($\geq$.01 in) or non-occurrence of precipitation, rather than the amount, was verified.

The verification was performed for the period January-March 1974--a total of 127 cases--for the 0-5 and 5-11 hr. forecast periods for SUM and the corresponding 6-12 and 12-18 hr forecast periods for the PE and LFM models. Observed values, obtained from NMC, were for the standard 6-hr periods. It should be noted that the length of the first period SUM forecast (5 hr) does not exactly match that of the observation (6 hr). This may somewhat effect the validity of the verification for this period. Forecasts from both 0000 and 1200 GMT were verified at 58 U.S. cities shown by the closed circles in Figure 1.

Several statistics were used to verify precipitation occurrence in order to gain as much comparative information as possible. They were: percent correct, prefigurance, post agreement, threat score, Heidke skill score, and bias. For convenience, these scores are defined in the Appendix. As was done with the sea-level pressure, the precipitation verification was also divided into eastern and western portions. Here, 100°W longitude was used as the dividing line. This resulted in 35 cities in the east and 23 cities in the mountainous west.

RESULTS

The results of the sea-level pressure verification are summarized in Figures 2, 3, and 4 where the S1 score and MAE are plotted against SUM forecast projection. Here, it has been assumed that both verification statistics decay linearly with time so that crossover points between SUM's S1 score and MAE and those of the PE or LFM can be determined between the 5- and 17-hr projections.

A comparison of the PE and LFM for the 1973-1974 winter season over the entire U.S. (Figure 2) shows that the PE appears to be generally better than the LFM in S1 score and MAE. Note, however, that the difference in S1 score between the two models decreases with time. Eastern U.S. and western U.S. graphs (Figures 3 and 4) indicate that the significant differences in scores between the models occur west of 90°W longitude. In the east, the LFM and PE have nearly the same scores, especially in the MAE.

Figure 2 shows that SUM has a better S1 score than the PE and LFM at the 5-hr projection and a worse S1 score at the 17-hr projection. The same is true for the MAE except at the 17-hr projection where SUM remains somewhat better than the LFM. The crossover point, where the SUM S1 score becomes worse than the PE, occurs at about 12.5 hours into the SUM forecast period. For the MAE, the crossover point for these two models is at approximately 15 hr. For SUM and the LFM, the crossover points are about 13.5 hr for the S1 score and beyond 17 hr--the maximum SUM forecast projection --for the MAE. It should be emphasized that all of these points are beyond the latest (11-hr) SUM sea-level pressure forecast transmitted via FOFAX.

On a regional basis, as shown in Figures 3 and 4, the crossover points do not show significant variation for SUM and the PE. However, for SUM and the LFM, both crossover points are farther out over the west than over the east. As expected, Figures 3 and 4 show that the S1 scores and MAE's are better over the east than over the west for all models.

The results of the precipitation verification for the first period (0-5 hr SUM and 6-12 hr PE and LFM forecasts) are shown in Table 1. They indicate that the first period SUM forecasts are better than those of the PE and about the same as those of the LFM. The results also indicate that the LFM forecasts are better than the PE forecasts.

The bias shows that SUM overforecasts the occurrence of precipitation, especially in the west, while the PE is indicated to be dry, especially in the east. This dryness, in part, accounts for the low scores obtained by the PE. SUM is actually slightly wetter than indicated since its first period forecast is for 5 hr and the observation is for 6 hr. For the same reason, the actual scores for the first period SUM would be different that those given in Table 1, although probably not significantly different.

A breakdown by region indicates that SUM has better verification scores than the PE in the east and about the same scores as the PE in the west. The results do not indicate any significant regional variation for SUM and the LFM.

Second period (5-11 hr SUM and 12-18 hr PE and LFM forecasts) are shown in Table 2. The scores for the three models for this period are more similar than those for the first period with the exception of the percent correct. The results indicate that the LFM is slightly better than SUM, and SUM is about the same as the PE.

The bias shows that the PE overforecast precipitation for this period which contrasts with the first period underforecasting. Note, too, the overall improvements of the PE scores from the first period to the second period--the only one of the three models for which this occurred. Overforecasting of precipitation by SUM, and to a lesser extent by the LFM, continued in the second period. In addition, note that all the models, for both periods, are considerably wetter in the west than in the east.

Over the western U.S. the PE and LFM models have nearly the same scores and are somewhat better than those of SUM. In the east, the results indicate that the LFM is slightly better than SUM, and SUM and the PE are about the same.

SUMMARY

SUM sea-level pressure forecasts for the winter season 1973-1974 have been shown to have a better S1 score and MAE than the PE and LFM models up to a forecast projection beyond the latest (11-hr) SUM sea-level pressure forecast transmitted via FOFAX. The PE generally had better sea-level pressure verification scores than the LFM.

SUM precipitation forecasts for the period January to March 1974 generally have better verification scores than the PE and about the same scores as the LFM for the first 5 hr SUM period. Second period SUM forecasts, in general, have about the same verification scores as the PE and slightly worse scores than the LFM. The LFM had better verification scores than the PE for both periods. The precipitation verification results suggest that an increase in the mean relative humidity criterion required for precipitation in SUM (this reduces the frequency of the precipitation event) during the winter season may improve the model's verification scores.

ACKNOWLEDGEMENTS

We want to thank Harry Brown, Chief of the Basic Weather Forecast Branch of NMC, and Vernon Bohl, Deputy Chief of the Quantitative Precipitation Branch of NMC, for providing the observations used in this verification. Also, our thanks go to Thomas D. Bethem who drafted the figures and Mary-Blue Battle who typed the paper.

REFERENCES

Glahn, H. R., and D. A. Lowry, "An Operational Subsynoptic Advection Model," Journal of Applied Meteorology, Vol. 11, No. 4, June 1972, pp. 578-585.

Grayson, T. H., and R. J. Bermowitz, "A Subsynoptic Update Model and Forecast System with Application to Aviation Weather," Report No. FAA-RD-74-100, Federal Aviation Administration, March 1974, 48 pp.

Howcroft, J., and A. Desmaris, "The Limited Area Fine Mesh (LFM) Model," NWS Technical Procedures Bulletin, No. 67, November 1971, 11 pp.

Shuman, F.G. and J. B. Hovermale, "An Operational Six-Layer Primitive Equation Model," Journal of Applied Meteorology, Vol. 7., No. 4, August 1968, pp. 525-547.

Teweles, S., and H. B. Wobus, "Verification of Prognostic Charts," Bulletin of the American Meteorological Society, Vol. 35, No. 10, December 1954, pp. 455-463.

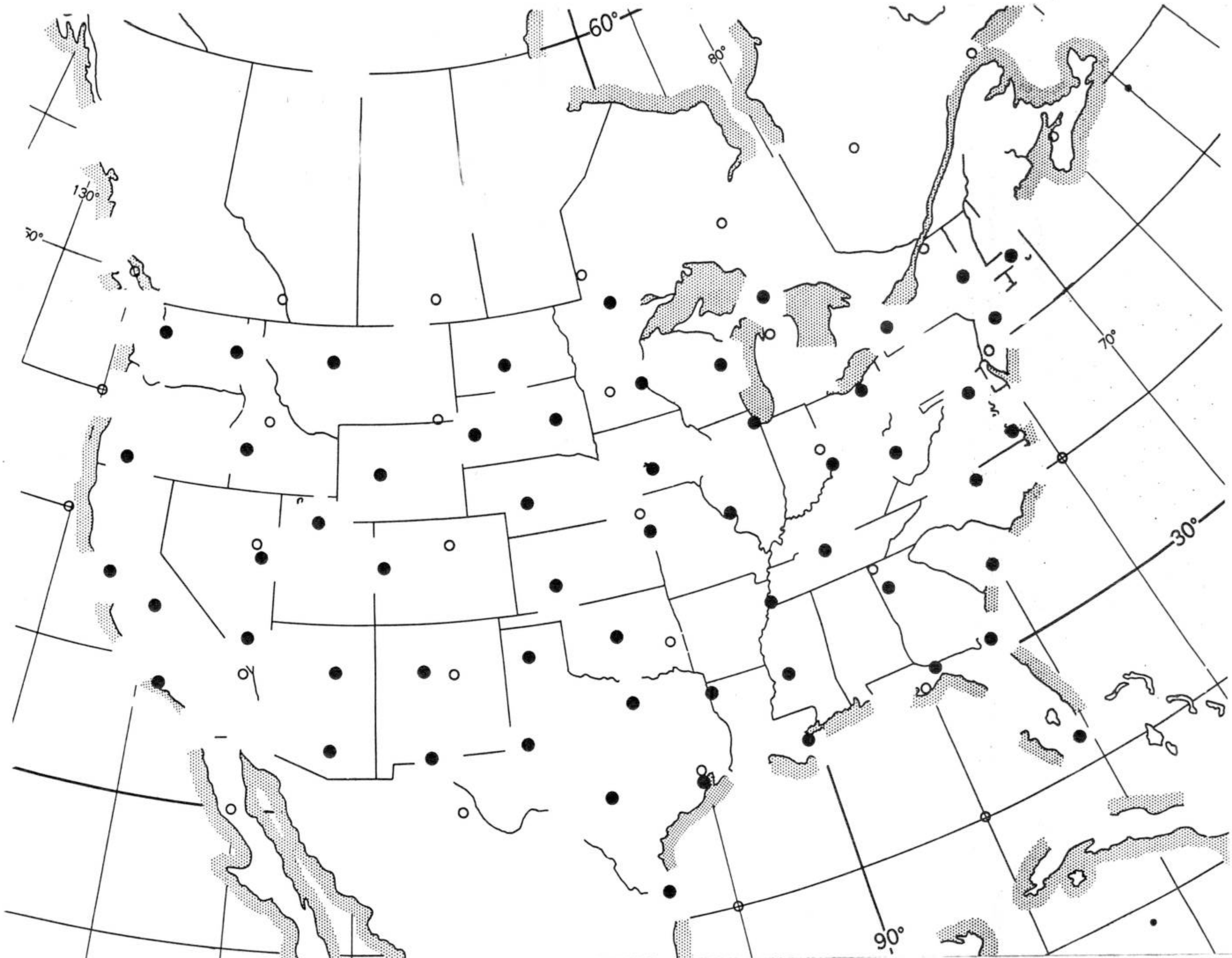

Figure 1. 31 grid points (open circles) used in the sea-level pressure verification and 58 cities (closed circles) used in the precipitation verification.

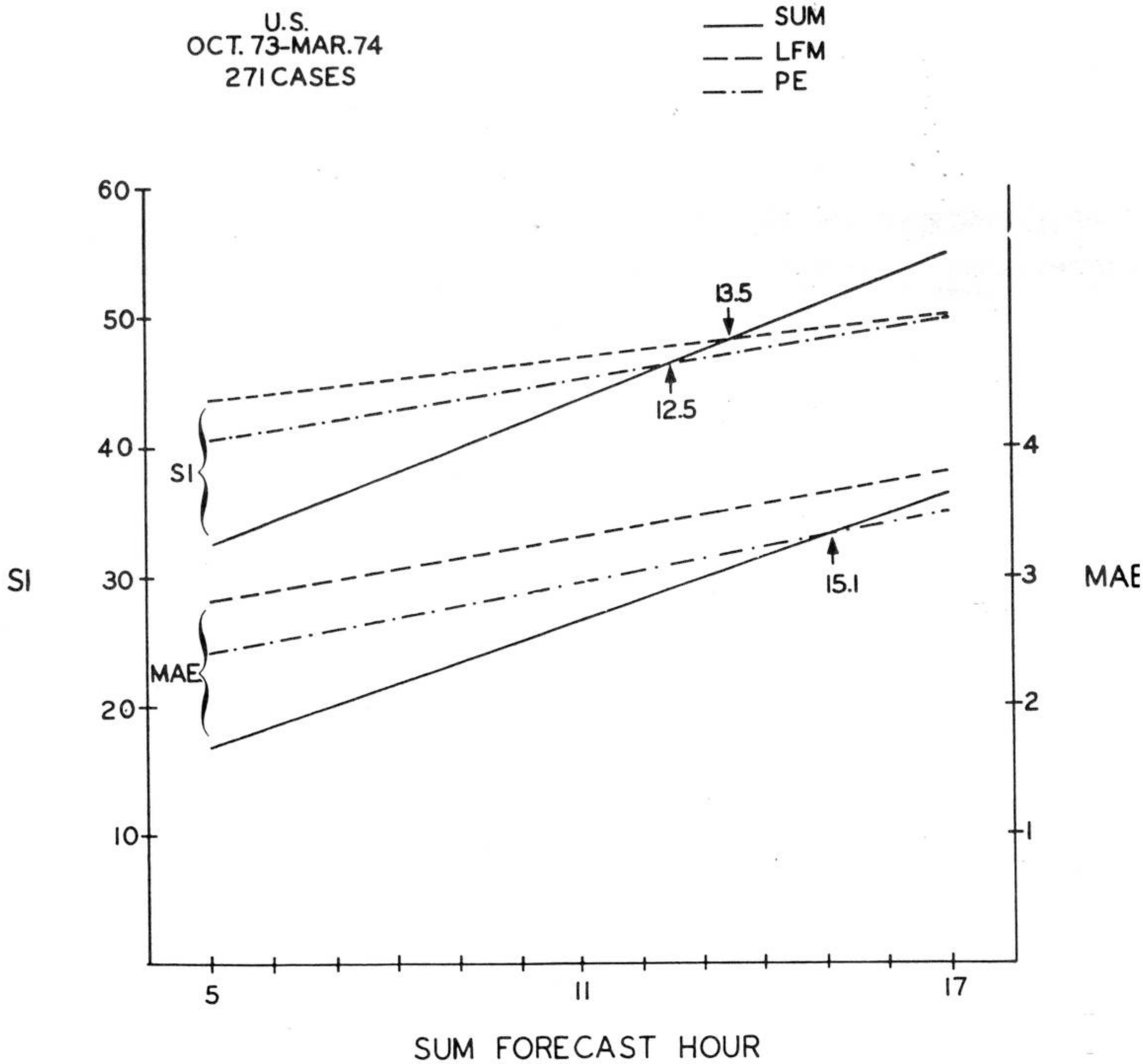

Figure 2. Comparison of SUM, LFM, and PE sea-level pressure forecasts at 31 points over the United States and adjacent areas for the period October 1973–March 1974. Mean absolute error is in mb.

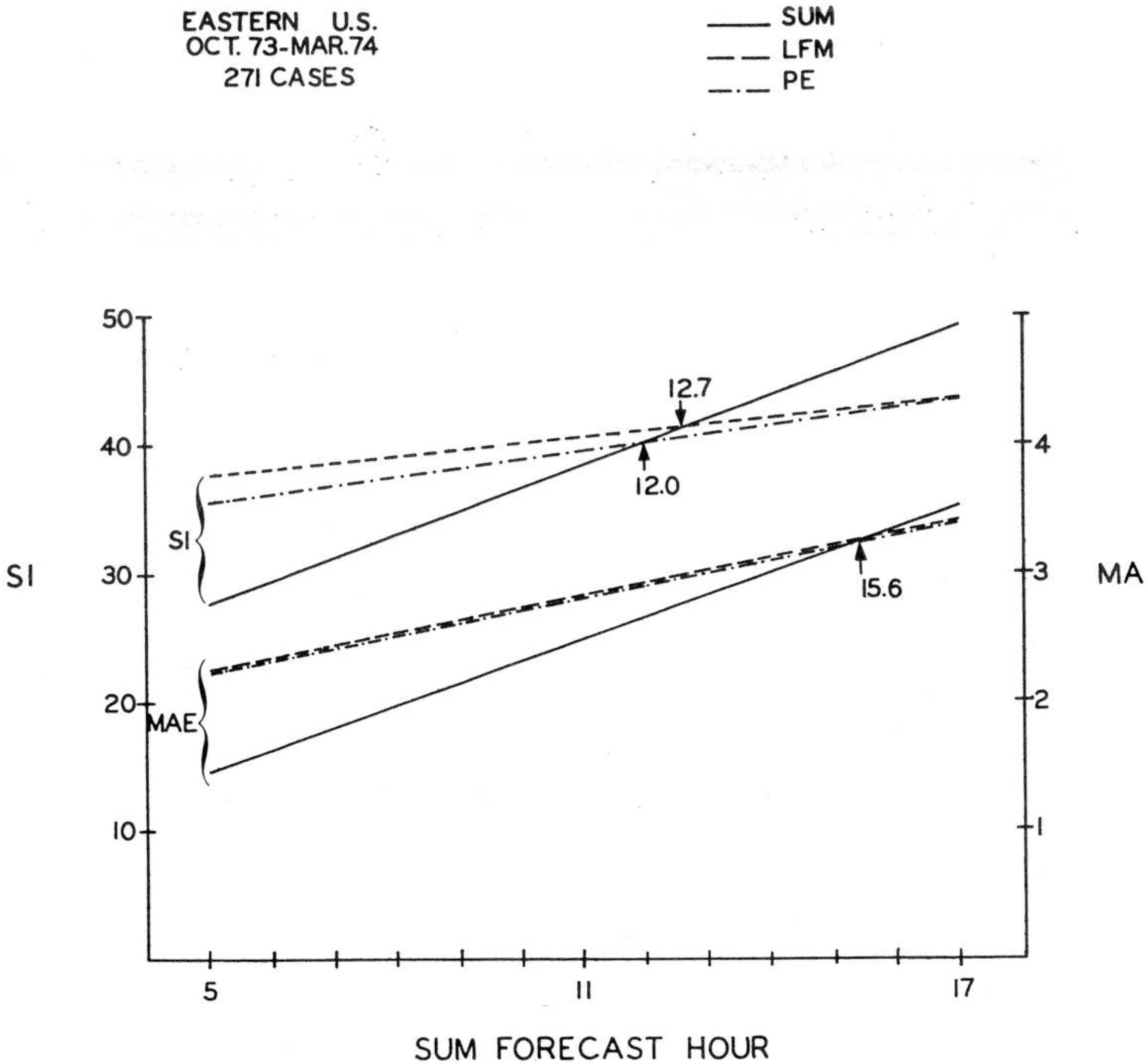

Figure 3. Same as Figure 2 except for 12 points over the eastern United States and adjacent areas.

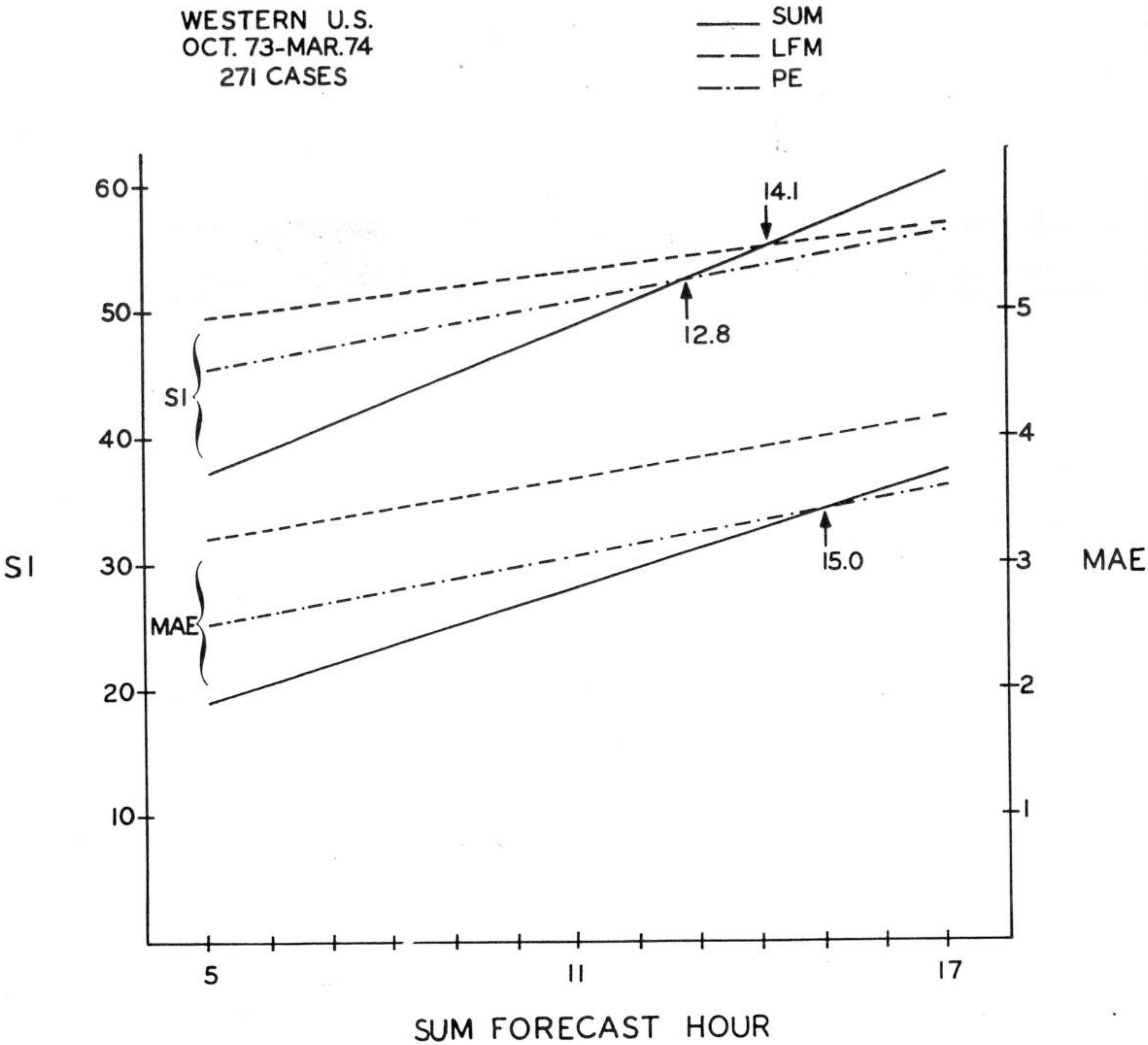

Figure 4. Same as Figure 2 except for 19 points over the western United States and adjacent areas.

Table 1. Comparison of 0-5 hr SUM and 6-12 hr LFM and PE precipitation forecasts at 58 U.S. cities for January-March 1974 (127 cases). For each verification score the top figure is for the entire U.S. and the bottom figures are for the western and eastern U.S.

	U.S. West/East		
	LFM	PE	SUM
Percent Correct	87.1 86.1/87.8	85.7 88.0/84.2	86.4 84.6/87.6
Prefigurance	.59 .67/.56	.35 .43/.31	.82 .81/.82
Post Agreement	.53 .41/.63	.48 .44/.51	.51 .39/.58
Threat Score	.39 .34/.42	.25 .28/.24	.45 .36/.51
Heidke Skill Score	.49 .43/.52	.32 .37/.30	.55 .45/.60
Bias	1.12 1.62/.89	.73 .98/.61	1.62 2.06/1.42

Table 2. Comparison of 5-11 hr SUM and 12-18 hr LFM and PE precipitation forecasts at 58 U.S. cities for January-March 1974 (127 cases). For each verification score the top figure is for the entire U.S. and the bottom figures are for the western and eastern U.S.

	U.S. West/East		
	LFM	PE	SUM
Percent Correct	85.7 84.1/86.7	84.6 84.7/84.5	82.8 81.4/83.7
Prefigurance	.57 .71/.50	.54 .70/.46	.67 .77/.62
Post Agreement	.47 .38/.57	.44 .39/.49	.41 .34/.47
Threat Score	.35 .33/.37	.32 .33/.31	.34 .31/.36
Heidke Skill Score	.43 .41/.46	.40 .42/.39	.41 .38/.44
Bias	1.20 1.89/.88	1.22 1.82/.95	1.61 2.26/1.31

APPENDIX

The scores used in the precipitation verification were computed from a two-category--precipitation (1), no precipitation (2) --contingency table as shown in the accompanying table.

Observed Category	Forecast Category 1	2	Total
1	A	D	G
2	B	E	H
Total	C	F	I

The percent correct, PC, is computed from

$$PC = \frac{A+E}{I} \times 100$$

where A is the number of correct forecasts of precipitation, E is the number of correct forecasts of no precipitation, and I is the total number of forecasts.

The prefigurance, PF, and the post agreement, PA, are given by

$$PF = \frac{A}{G}$$

and

$$PA = \frac{A}{C} \; .$$

G is the total number of observations of precipitation and C is the total number of forecasts of precipitation.

The threat score of precipitation, TS, is defined as

$$TS = \frac{A}{C + G - A} \; .$$

The Heidke skill score, SS, is computed from

$$SS = \frac{A + E - J}{I - J}$$

where J, the number of forecasts expected to be correct by chance, is given by

$$J = \frac{(C \times G) + (F \times H)}{I} .$$

F is the total number of forecasts of no precipitation and H is the total number of observations of no precipitation.

Finally, the bias of the precipitation category, or more simply, bias, BP, is defined as

$$BP = \frac{C}{G} .$$