

**WESTERN REGION TECHNICAL ATTACHMENT
NO. 96-28
NOVEMBER 12, 1996**

**PRECIPITATION VERIFICATION STATISTICS FROM
THE NCEP OPERATIONAL MODEL SUITE**

Mike Staudenmaier, Jr. - WRH-SSD/NWSFO SLC

Introduction

Quantitative precipitation forecasts (QPFs) from NCEP's operational model suite (ETA, MRF/AVN, NGM, and Meso-Eta) are verified against 24-hour accumulated precipitation observations on a daily basis. Knowledge of a model's performance and biases in QPF is important for River Forecast Centers and for NWS forecast offices. This Technical Attachment will present seasonal statistical information regarding model performance in quantitative precipitation forecasting.

Basic Procedure

To produce a fair comparison between the model grids of 24-hour accumulated precipitation, all the models need to be interpolated to similar resolution grids. For this purpose, the Meso-Eta model (29 km resolution) and the Eta model (48 km resolution) are interpolated to the old 80 km Eta grid. The NGM model is verified on its own grid with a resolution of 80 km, while the MRF/AVN is verified on a grid of approximately 90 km resolution. The AVN model is verified for the 1200 UTC model run, while the MRF model is verified using the 0000 UTC model run. The interpolation of the Eta models is done in such a way as to conserve the total volume of water found on the original grids.

The actual accumulated precipitation observations come from a network of nearly 10,000 rain gages that record 24-hour accumulated precipitation across the lower 48 states. This data is transmitted by the River Forecast Centers around the country to NMC. The stations have a fairly dense coverage in the eastern two-thirds of the country, with more sparse coverage west of the Rocky Mountains (Fig. 1).

From the list of all possible reporting rain gages, it is then determined which grid boxes will become part of the verification. Only the grid boxes that contain one or more of the network of over 10,000 rain gages are considered to be part of the verification analysis domain. If a grid box does not have a routinely reporting rain gage located inside it, it is not verified. The observations are analyzed to the verification grid by a simple average

of all reports within a given grid box. Currently, the rain gages only report when they actually receive precipitation, therefore, no observations of zero are used to compute the average. Additionally, some radar data is used to supplement rain gage reports, however this radar data is not used on any of the grid boxes which do not have a rain gage data nearby for calibration.

The time periods over which the forecasts are verified are the 0-24 hour period and the 12-36 hour period. For the Meso-Eta, the three hour data assimilation cycle which occurs from 1200 UTC-1500 UTC and 0000 UTC-0300 UTC is used to create the 36 hour verification period. The results from both periods are combined on the attached figures. The statistics are only included if all models were available and were verified through 36 hours.

Definition of Scores

Once the model generated QPF and the 24-hour accumulated precipitation fields are ready to be analyzed, two skill scores are computed. The scores which will be shown here are based upon comparing regions of forecast versus observed precipitation which are greater than a certain threshold, for example 0.5 inches. The two skill scores are as follows:

Equitable Threat Score

The equitable threat score (Schaefer, 1990) is defined as

$$\frac{(H - CH)}{(F + O - H - CH)}$$

- where F = the number of grid boxes that forecast more than the threshold
- O = the number of grid boxes that observe more than the threshold
- H = the number of grid boxes that correctly forecast more than the threshold
- CH = the expected number of correct forecasts due to chance = $F \cdot O / T$
where T = the total number of grid boxes inside the verification domain

The equitable threat score seems to be a good estimate for overall forecast skill. The higher the value, the better the forecast model skill is for that particular threshold. The equitable threat score can vary from a small negative number to 1.0, where 1.0 represents a perfect forecast. This is basically the ratio of the correct forecast area to the total area of the forecast and observed precipitation. The model gets penalized for forecasting rain in the wrong place as well as not forecasting rain in the right place. Thus, the model with the highest score is generally the model with the best forecast skill.

Bias Score

The bias score is a very simple equation, defined as simply as F/O. This score does not comment at all on the skill of a model forecast in terms of the placement of precipitation, but does give an indication if a model is consistently over-or under-forecasting areas of precipitation. The best model is generally the one that remains near the 1.0 line, which means that the model does not generally over-forecast precipitation or under-forecast precipitation. If the model verifies over 1.0, it is over-predicting precipitation, and if below 1.0 it is under-predicting precipitation.

Seasonal Scores

Figures 2-9 are seasonal averages of the equitable threat scores and the bias scores for the previously mentioned models. The time periods for each season is as follows:

WINTER	---	1 DEC 1995-29 FEB 1996
SPRING	---	1 MAR 1996-31 MAY 1996
SUMMER	---	1 JUN 1996-31 AUG 1996
FALL	---	1 SEP 1996-27 OCT 1996*

*The fall season included all the data which was available up to the time of the writing of this paper.

In terms of equitable threat scores (Figs. 2, 4, 6, 8), it can be seen that the Meso-Eta typically is the best model, especially for precipitation amounts under 1.00 inches. The Eta model seems to be equal to, or occasionally slightly better than the Meso-Eta model for precipitation amounts over 1.00 inches, especially in the cooler seasons. The MRF/AVN model typically out performs the NGM (RAFS) model in almost all precipitation thresholds and seasons. All the models show decreasing skill with increasing precipitation threshold, with almost a steady fall in skill above 0.25 inches.

The bias scores affirm much of what was previously mentioned (Figs. 3, 5, 7, 9). The Meso-Eta repeatedly outperforms all other models, with its curve most closely following the 1.0 line. An exception to this can be seen in the Summer period, when the NGM had the better verification for precipitation under 0.50 inches. However, in precipitation amounts over 0.50 inches, the Meso-Eta was clearly the better performer. The improvement in precipitation bias of the Meso-Eta over the Eta model can be best seen in the cool seasons, especially during the Winter period. This improvement is likely due to the improved resolution of the Meso-Eta model. During the cool season, when orographic forcing by complex terrain becomes most important in the placement of heavy precipitation, the Meso-Eta model clearly becomes the best model in terms of placement of precipitation (Burks and Staudenmaier, 1996; Schneider et. al, 1996; Gartner et. al, 1996).

On average, it appears that the MRF/AVN typically overestimates precipitation, while the NGM typically overestimates light amounts of precipitation while underestimating moderate amounts. The NGM appears to be the worse model overall in forecasting heavy amounts of precipitation. The Eta/Meso-Eta models appear to suffer the same biases of overestimating light precipitation amounts and underestimating heavier precipitation amounts, with the Meso-Eta showing some improvement in these biases over the Eta model.

Conclusion

This Technical Attachment has shown the quantitative precipitation verification scores of the NCEP model suite over the previous four seasons. Both equitable threat and bias scores were shown, and some general statements were made regarding model performance. The main points to be gained by these figures are 1) that the Meso-Eta model appears to have the best skill at this time in prediction both the placement of precipitation (especially in the cool season) and in having the least bias in forecasting precipitation amounts, and 2) that the MRF/AVN model, even with its slightly poorer resolution, seems to forecast precipitation somewhat better than the NGM model. These threat scores, calculated on a monthly basis, can be viewed on the Internet at the following address:

<http://nic.fb4.noaa.gov:8000/research/pptscor.html>

Acknowledgments

The author wishes to thank Michael Baldwin, NCEP, for his help during the research phase of this paper along with the Environmental Modeling Center (EMC) for placing this data on the Internet for public inspection.

References

- Burks, J.E., and M.J. Staudenmaier, 1996: A comparison of the Eta and the Meso-Eta models during the 11-12 December 1995 storm of the decade. *WR-Technical Attachment 96-21*.
- Gartner, W.E., M.E. Baldwin, and N.W. Junker, 1996: Regional analysis of quantitative precipitation forecasts from NCEP's early Eta and Meso-Eta models. *Preprints, 15th Conf. on Weather Analysis and Forecasting, AMS, Norfolk, VA. Aug 1996*.
- Schneider, R.S., N.W. Junker, M.T. Eckert, and T.M. Considine, 1996: The performance of the 29 km Meso-Eta model in support of forecasting at the Hydrometeorological Prediction Center. *Preprints, 15th Conf. on Weather Analysis and Forecasting, AMS, Norfolk, VA. Aug 1996*.

pcp_list_us

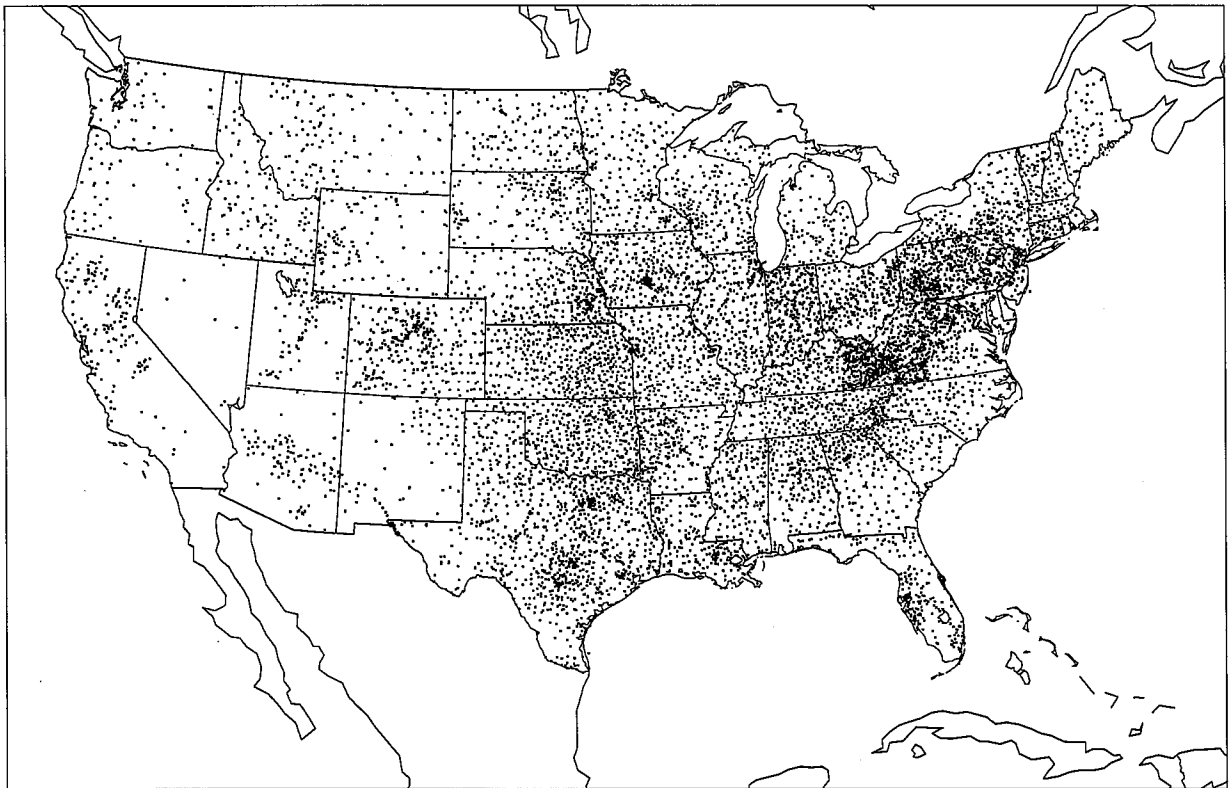


Fig. 1

Equitable Threat – All Periods

Valid 1 Dec 95 – 29 Feb 96

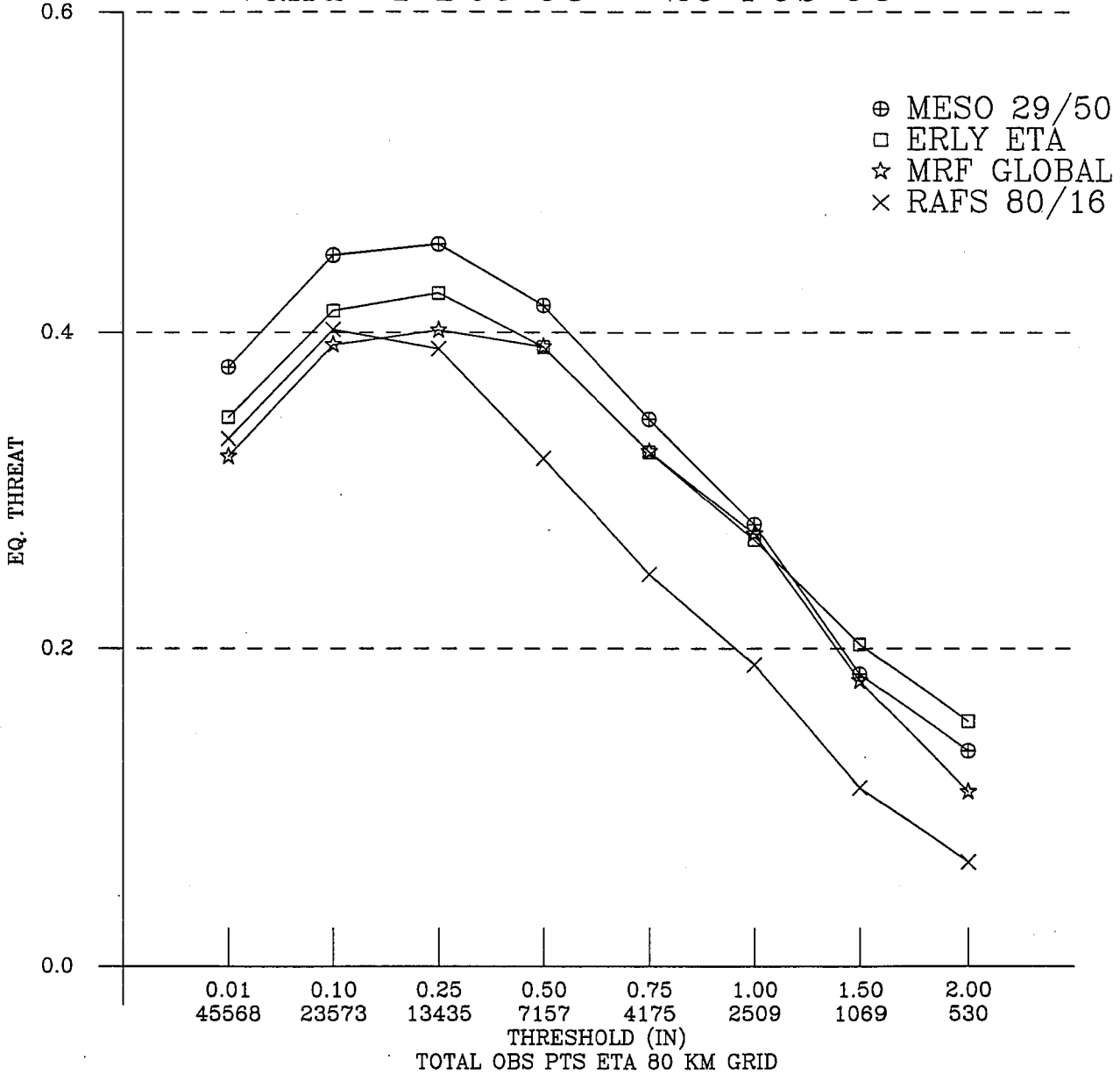


Fig. 2

Bias sum of all forecasts
Valid 1 Dec 95 - 29 Feb 96

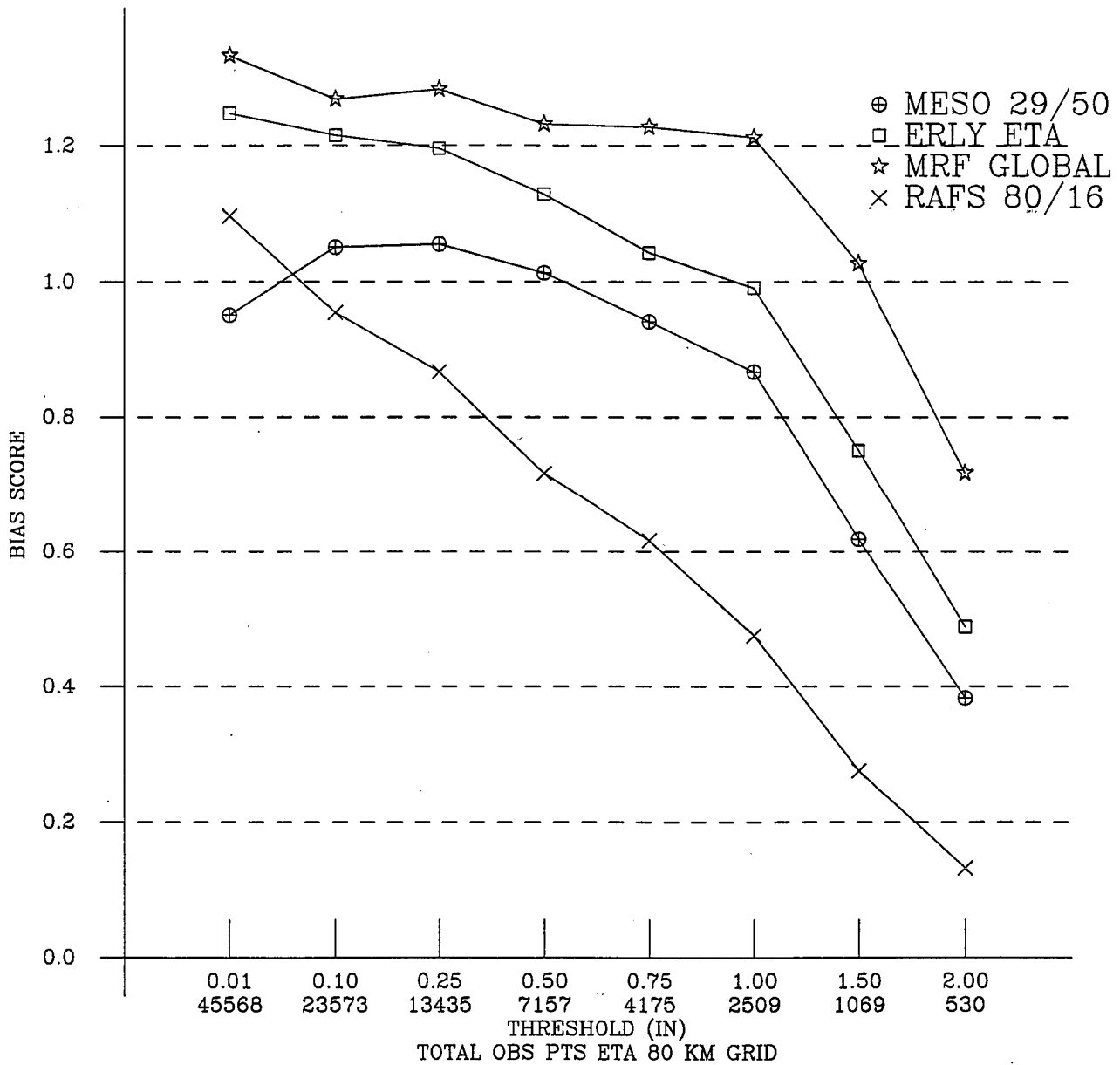


Fig. 3

Equitable Threat – All Periods
Valid 1 Mar 96 – 31 May 96

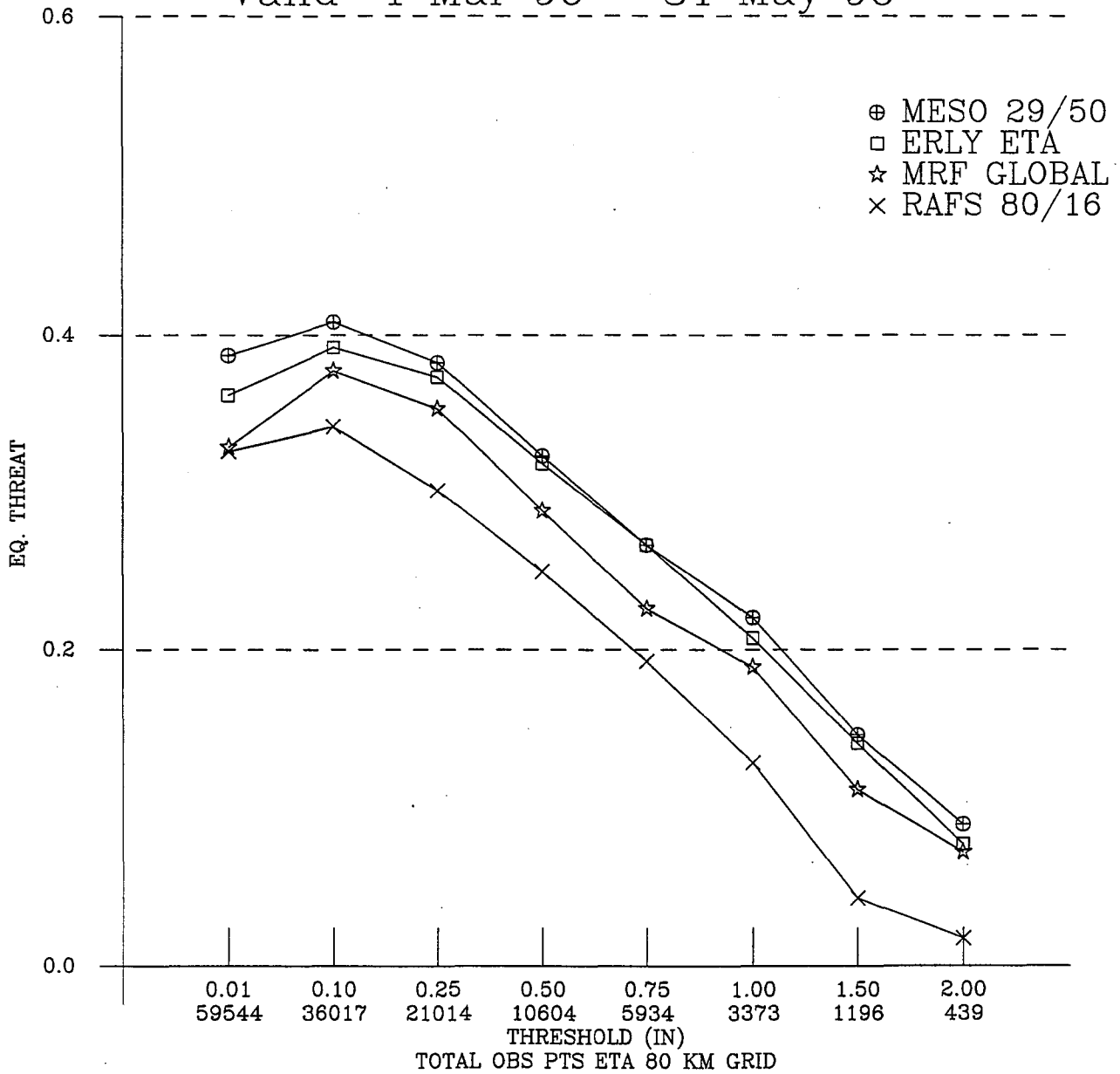


Fig. 4

Bias sum of all forecasts
Valid 1 Mar 96 - 31 May 96

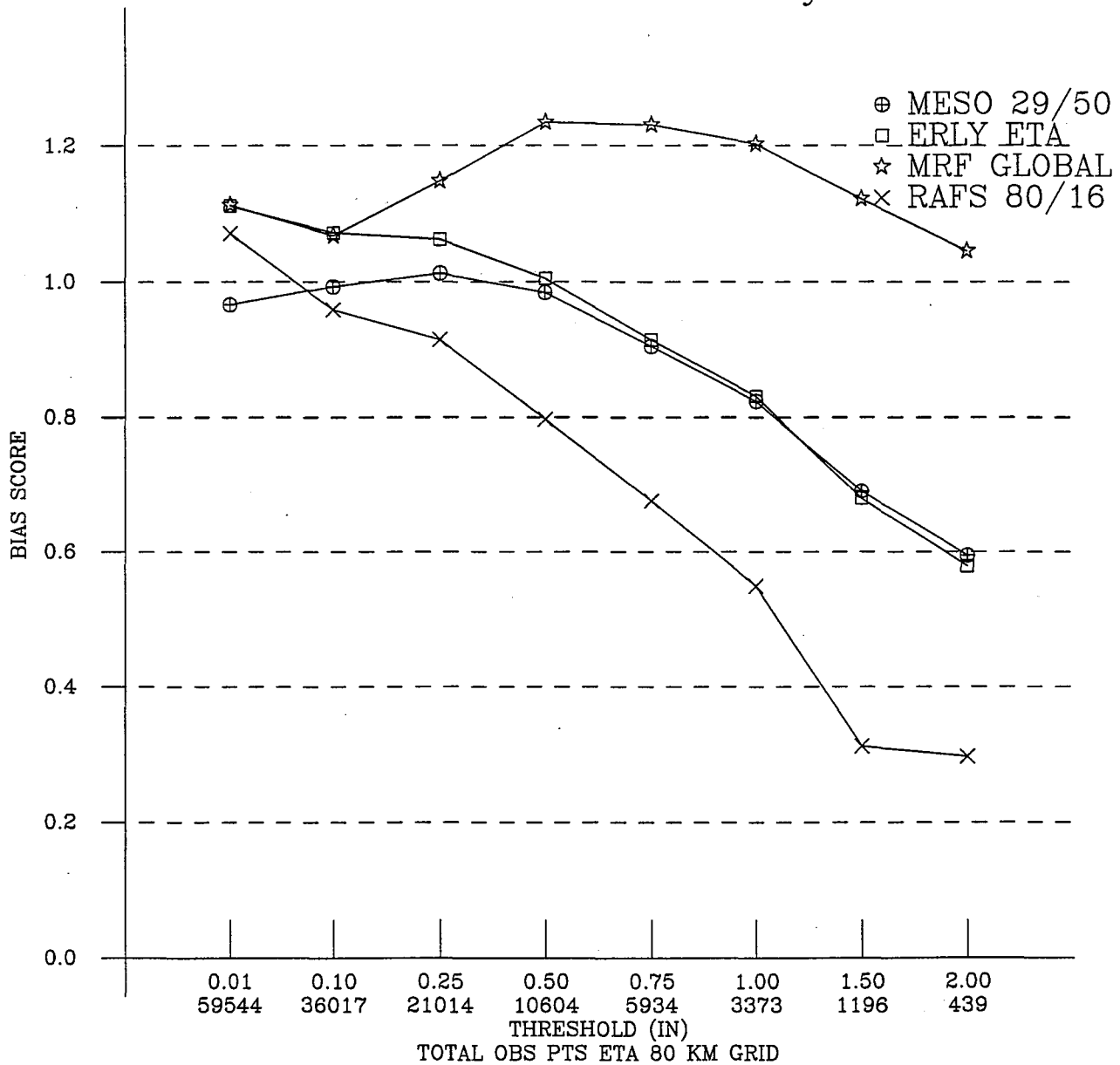


Fig. 5

Equitable Threat - All Periods
Valid 1 Jun 96 - 31 Aug 96

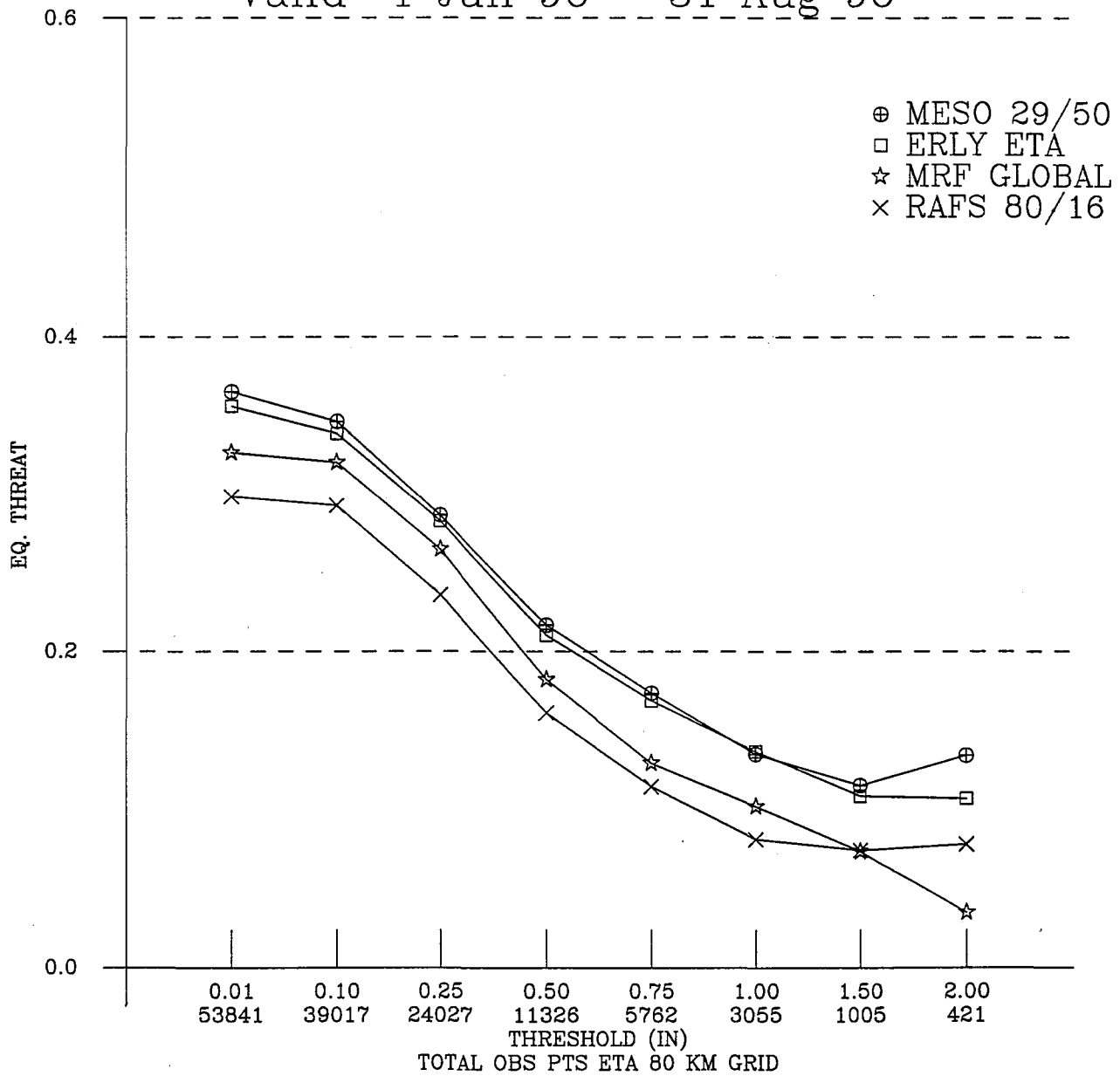


Fig. 6

Bias sum of all forecasts
Valid 1 Jun 96 - 31 Aug 96

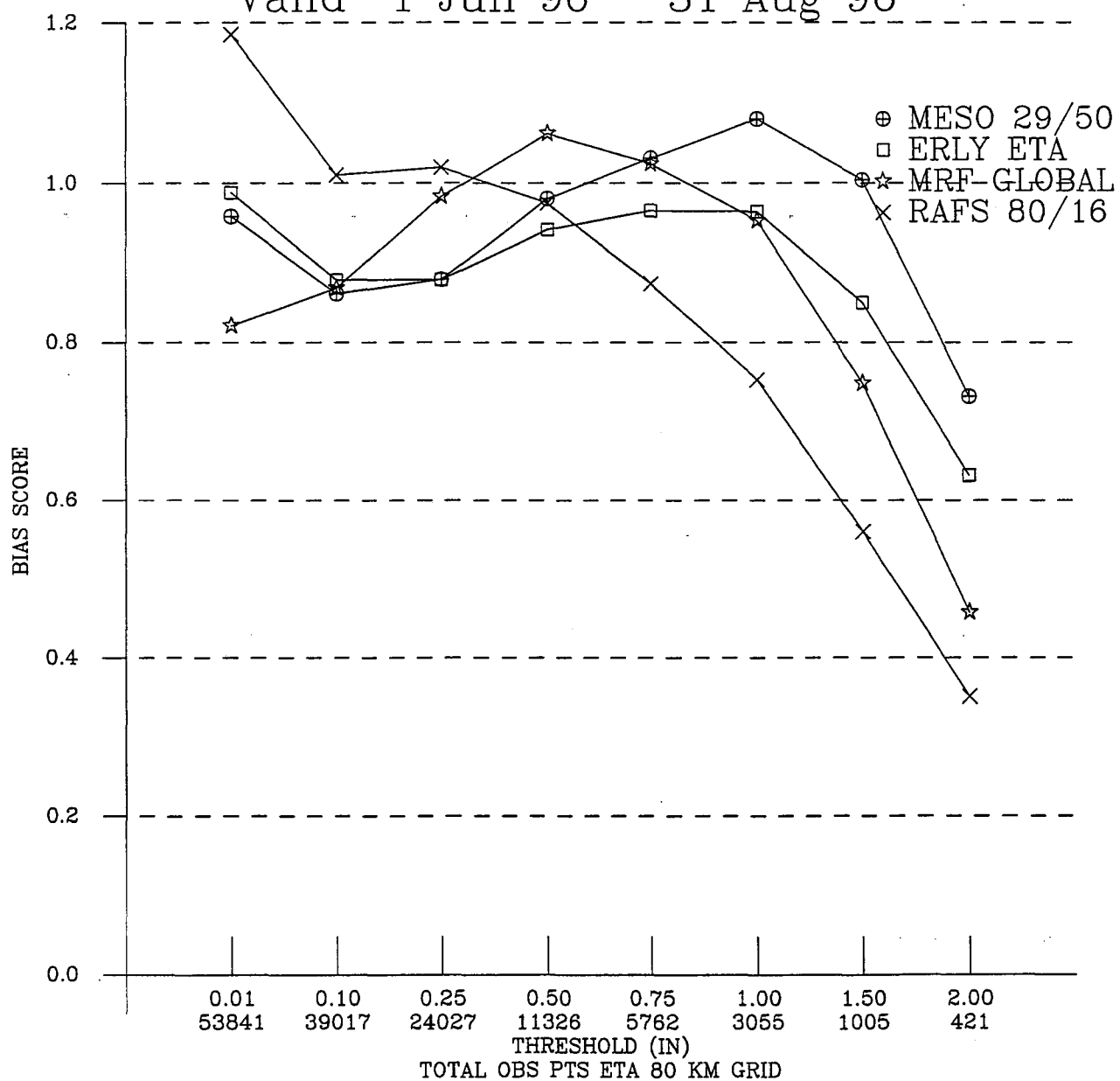


Fig. 7

Equitable Threat - All Periods
Valid 1 Sep 96 - 27 Oct 96

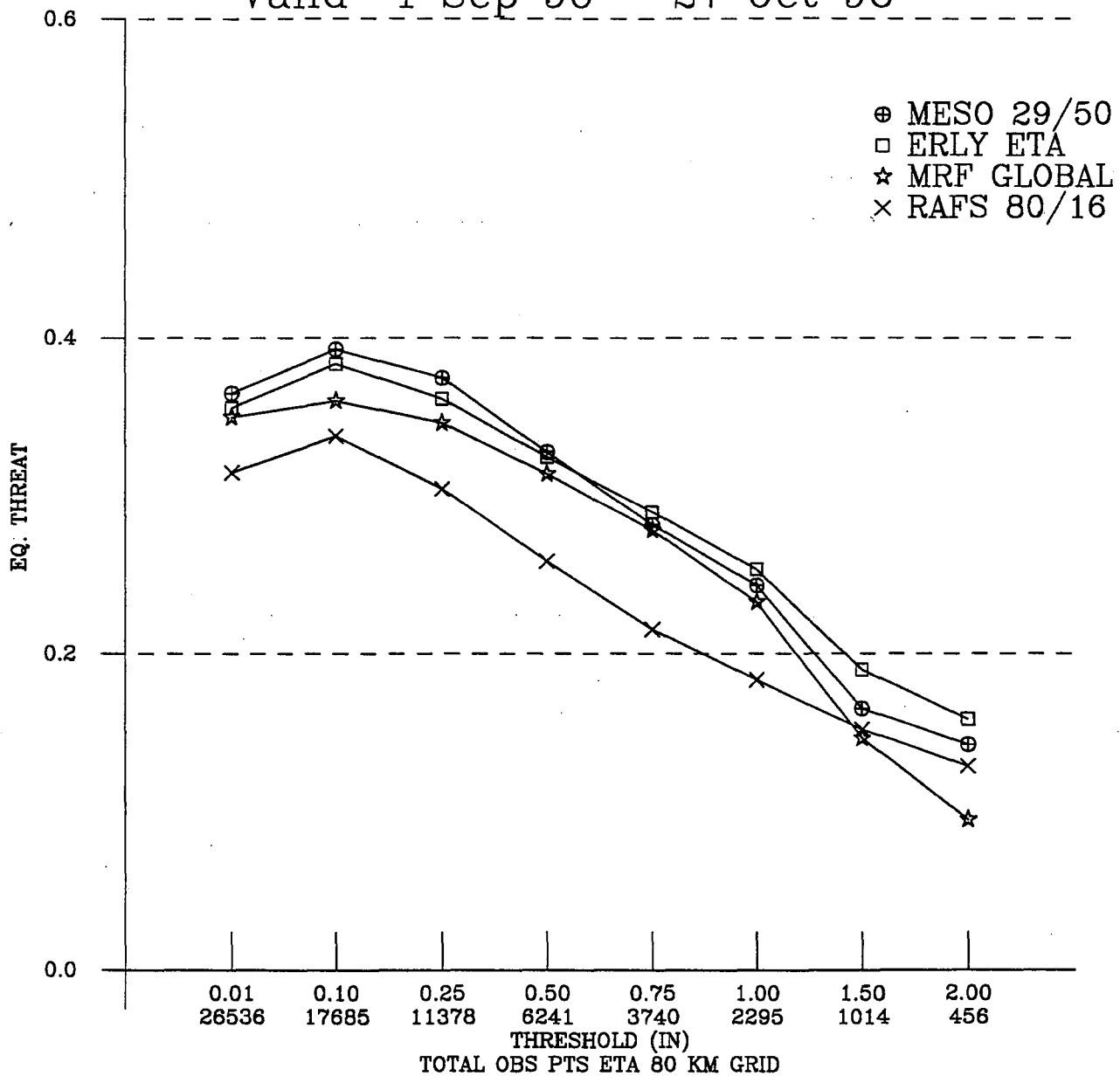


Fig. 8

Bias sum of all forecasts
Valid 1 Sep 96 - 27 Oct 96

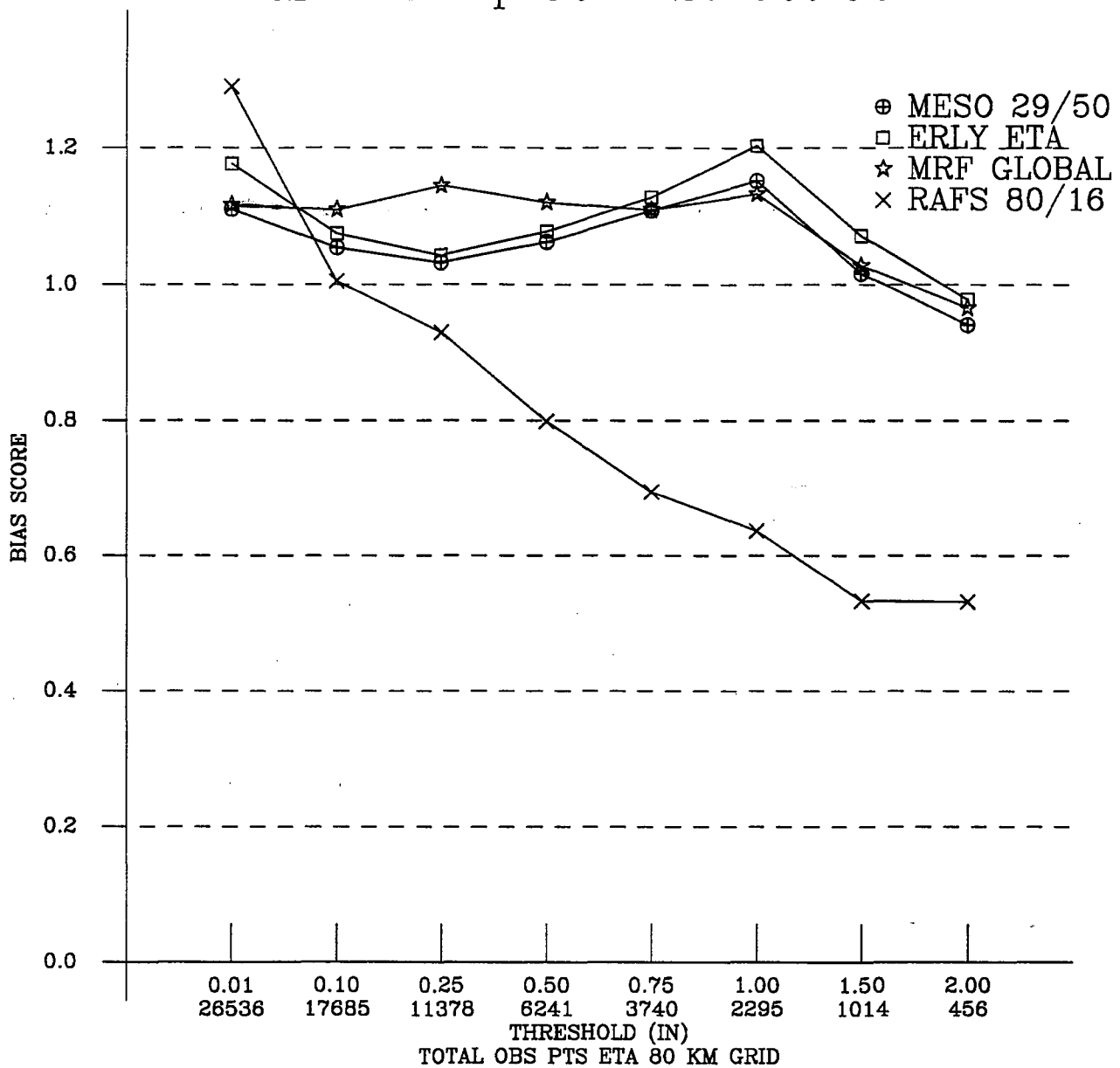


Fig. 9