NOAA Technical Memorandum ERL SEL-84

# VERIFICATION OF NOAA SPACE ENVIRONMENT LABORATORY FORECASTS: 1 JANUARY-31 DECEMBER 1993

K. A. Doggett



Space Environment Laboratory
Boulder, Colorado
August 1994

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / Environmental Research Laboratories

NOAA Technical Memorandum ERL SEL-84

# VERIFICATION OF NOAA SPACE ENVIRONMENT LABORATORY FORECASTS: 1 JANUARY -31 DECEMBER 1993

K. A. Doggett



Space Environment Laboratory
Boulder, Colorado
August 1994

**UNITED STATES DEPARTMENT OF COMMERCE**

**Ronald H. Brown**
**Secretary**

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

D. James Baker
Under Secretary for Oceans and Atmosphere/Administrator

Environmental Research Laboratories

Alan R. Thomas
Director

## NOTICE

Mention of a commercial company or product does not constitute an endorsement by NOAA Environmental Research Laboratories. Use for publicity or advertising purposes of information from this publication concerning proprietary products or the tests of such products is not authorized.

For sale by the National Technical Information Service, 5285 Port Royal Road
Springfield, VA 22161

# CONTENTS



# LIST of TABLES

## LIST of TABLES (cont.)



## LIST of FIGURES

## LIST of FIGURES (cont.)

# INTRODUCTION

This summary report compiles verification statistics for the primary forecasts issued by the NOAA Space Environment Laboratory (SEL). The following forecasts are evaluated: probability forecasts of M-class, X-class, and Proton flares expected for each of the next three days; integer-value forecasts of 10.7 cm Solar Flux, Fredericksburg A-Index, and Planetary A-Index for each of the next seven days; and next-day probability forecasts of geomagnetic activity levels at middle latitudes. This report covers forecasts made during 1 January to 31 December 1993.

Three levels of verification information are presented to accommodate different verification users. The first level presents an overview of forecast performance, concentrating on a few key verification measures and the ways in which these measures evolve over time. The second level offers several verification measures and diagrams to provide a detailed evaluation of each forecast. The third level focuses on the performance of individual forecasters.

Much of this information is produced by SEL's Operational Forecast Verification System [1]. This system is based on a general framework for forecast verification that focuses on the joint distribution of forecasts and observations and fundamental decompositions of this distribution [2]. The original objectives of SEL's verification system and the details of its content are found in [3].

The science of forecast verification has its own vocabulary. Some of these terms are unique to the verification problem, while others are standard statistical measures adapted to the task of assessing the quality of forecasts. A brief verification glossary is found in Appendix A. This glossary is a compilation of verification terms and quantities found throughout the summary report. Many of the specialized entries are discussed in greater detail in [4], and complete descriptions of the standard statistical quantities can be found in [5].

# I. FORECAST VERIFICATION OVERVIEW

This section presents an overview of forecast performance during the calendar year 1993. Four key verification measures are compiled in Table I-1 for probabilistic flare forecasts and continuous forecasts of 10.7 cm Solar Flux and A-Index. These measures quantify the attributes of overall bias, accuracy, linear correlation, and skill for each of these forecast types. Following this table are trend diagrams showing the behavior of overall bias, accuracy, and linear correlation during each calendar quarter of the year.

Table I-2 summarizes a few key measures for probability forecasts of geomagnetic activity level, focusing on next-day, middle-latitude forecasts. The ranked probability score (RPS) is calculated for the SEL forecasts as well as for forecasts of sample climatology and 30-day climatology. (A 30-day climatology window was chosen so that the geomagnetic seasonal effect could be captured.) Reliability [6] and two types of resolution [7] [8] are also calculated from vector partitions of the RPS [9]. Trend diagrams of these measures for each calendar quarter are compiled at the end of the section.

**Table I-1.** Overview verification statistics for SEL forecasts [1] [2], 1 January–31 December 1993

| Forecast | Lead-time (days) | ME | MSE (RMSE) | r | $SS_{SC}$ |
|---|---|---|---|---|---|
| Probabilistic: | | | | | |
| M Flares | 1 | 0.038 | 0.124 (0.35) | 0.331 | 0.043 |
| | 2 | 0.036 | 0.134 (0.37) | 0.263 | −0.034 |
| | 3 | 0.031 | 0.140 (0.37) | 0.215 | −0.080 |
| X Flares | 1 | 0.030 | 0.003 (0.06) | n/a | n/a |
| | 2 | 0.030 | 0.003 (0.06) | n/a | n/a |
| | 3 | 0.028 | 0.002 (0.04) | n/a | n/a |
| Proton Flares | 1 | 0.016 | 0.010 (0.10) | 0.300 | 0.075 |
| | 2 | 0.015 | 0.010 (0.10) | 0.288 | 0.075 |
| | 3 | 0.014 | 0.010 (0.10) | 0.277 | 0.075 |
| Continuous: | | | | | |
| 10.7 cm Flux | 1 | 0.09 | 32.4 (5.7) | 0.96 | 0.93 |
| | 2 | 0.11 | 76.2 (8.7) | 0.92 | 0.83 |
| | 3 | 0.07 | 132.5 (11.5) | 0.86 | 0.71 |
| $A_{fred}$ Index | 1 | 0.47 | 79.4 (8.9) | 0.43 | 0.10 |
| | 2 | 0.61 | 84.3 (9.2) | 0.40 | 0.04 |
| | 3 | 0.33 | 85.4 (9.2) | 0.37 | 0.03 |
| $A_p$ Index | 1 | 2.07 | 134.8 (11.6) | 0.41 | 0.09 |
| | 2 | 1.93 | 145.2 (12.1) | 0.37 | 0.04 |
| | 3 | 1.52 | 147.7 (12.2) | 0.33 | 0.01 |

1. Valid probabilistic forecasts range from 0.01 to 0.99. Observations are either 0 (event didn't occur) or 1 (event occurred). Continuous forecasts and observations are integer values of the forecast parameter.

2. These summary and performance measures include
   ME, mean error (the overall bias in the forecasts);
   MSE, mean square error (and RMSE, root mean square error), a measure of forecast accuracy;
   r, linear-correlation coefficient (varies between ±1); and
   $SS_{SC}$, skill score with respect to sample climatology (varies between ±1).

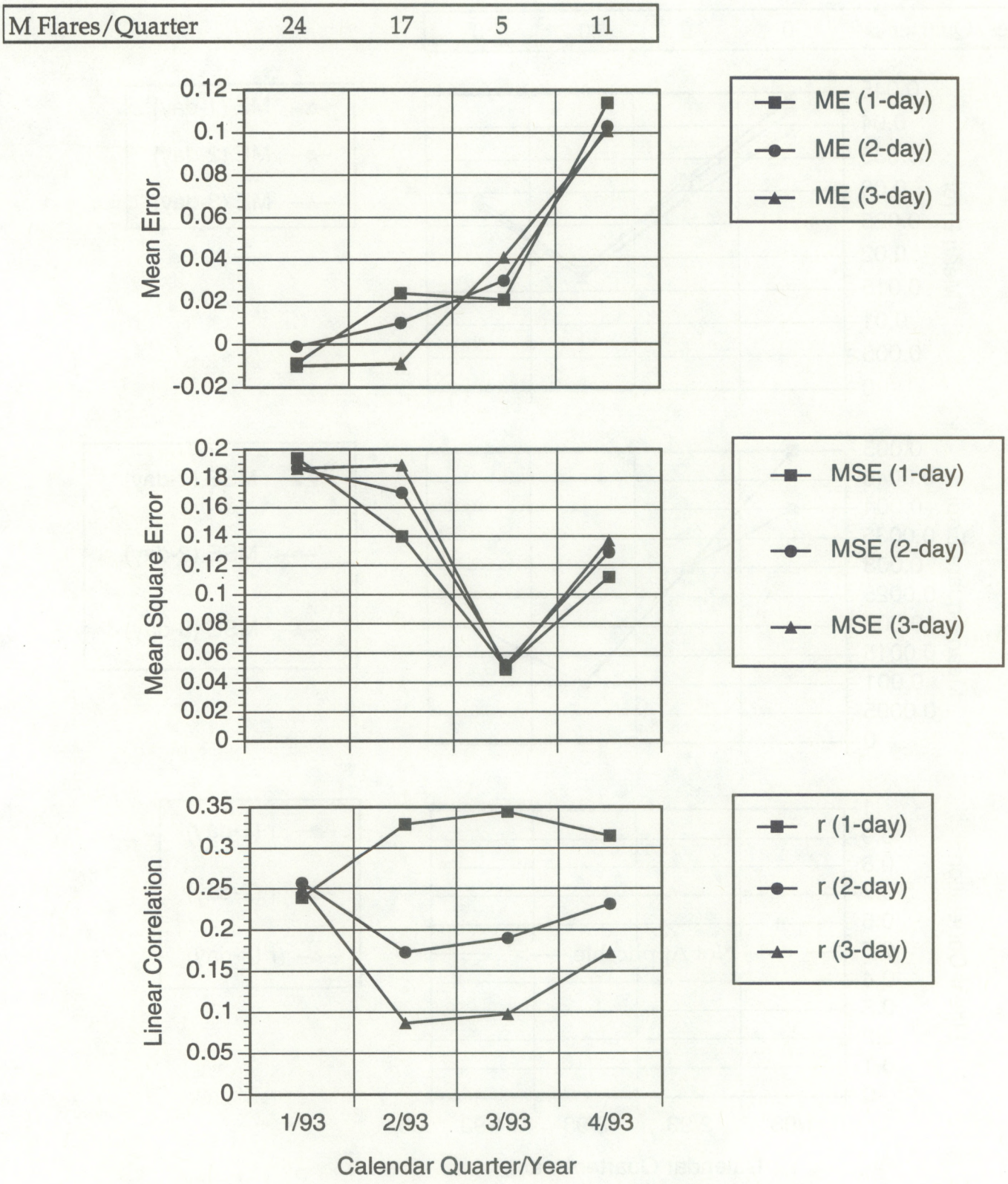



**Figure I-1.** Trend diagrams for M-class flare forecasts at three lead times. The overall bias (mean error), accuracy (mean square error), and linear correlation are plotted for each calendar quarter of 1993. The occurrence of M flares during each quarter is indicated at the top of the figure.

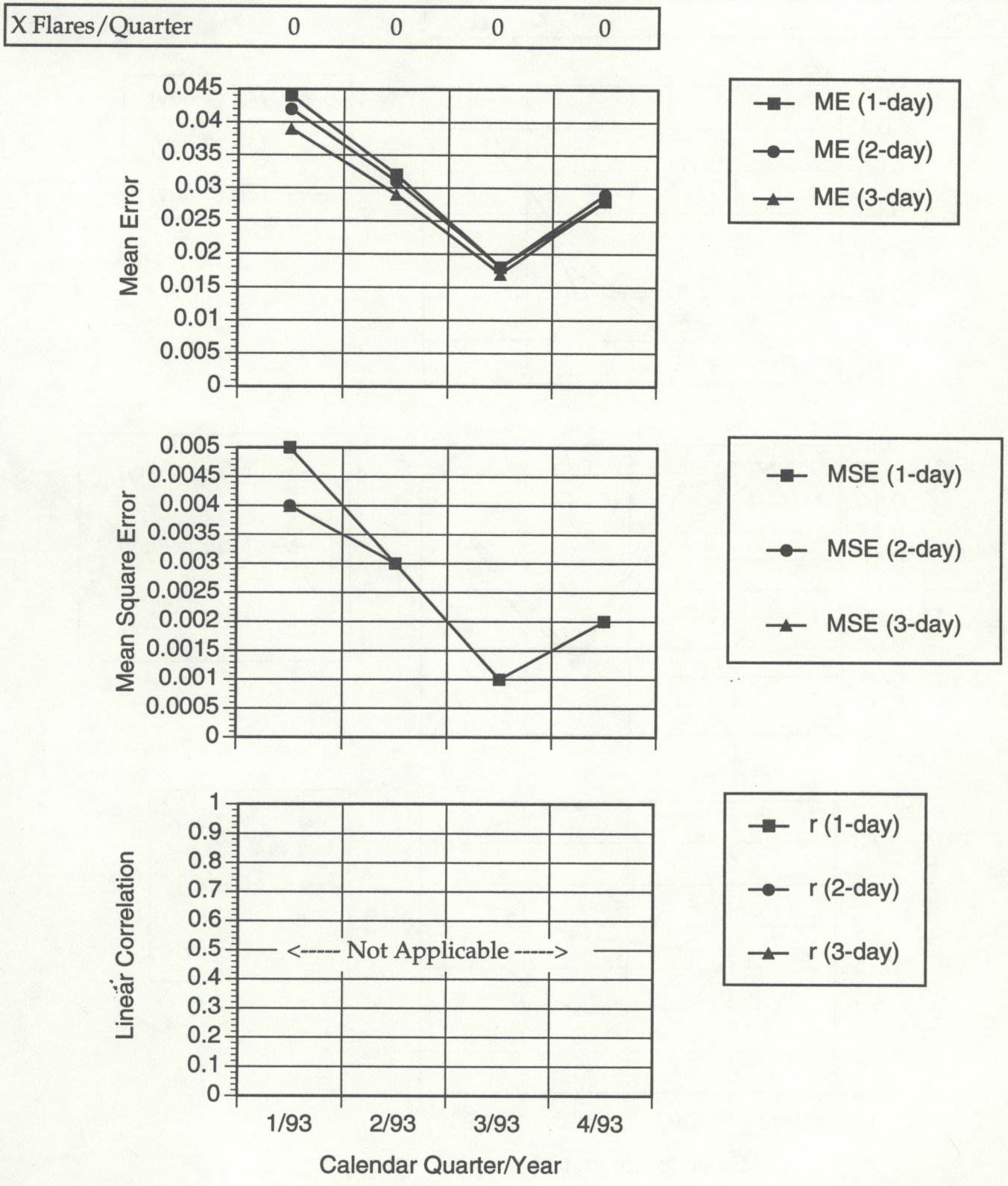



**Figure I-2.** Trend diagrams for X-class flare forecasts at three lead times. The overall bias (mean error), accuracy (mean square error), and linear correlation are plotted for each calendar quarter of 1993. The occurrence of X flares during each quarter is indicated at the top of the figure.

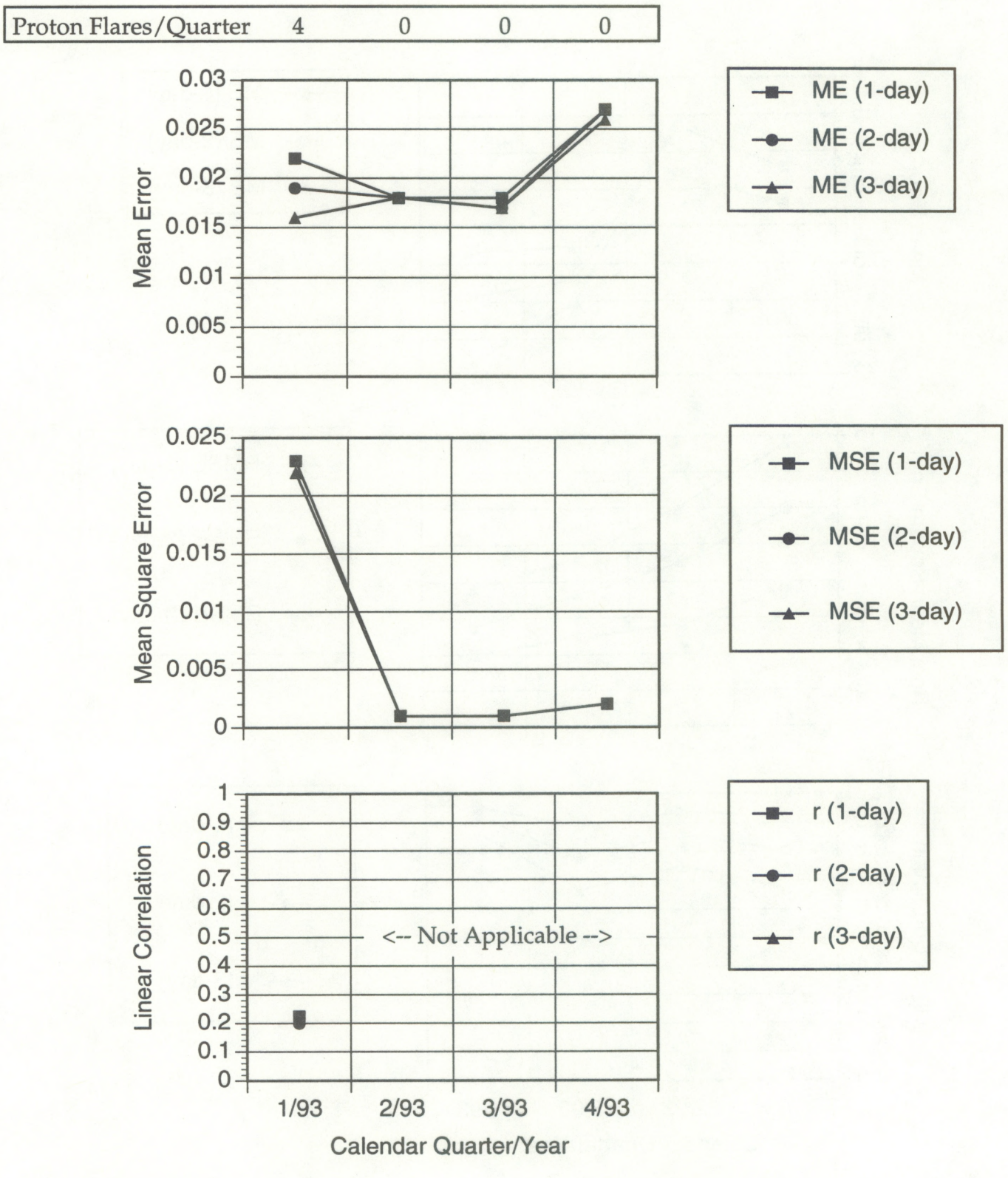



**Figure I-3.** Trend diagrams for proton flare forecasts at three lead times. The overall bias (mean error), accuracy (mean square error), and linear correlation are plotted for each calendar quarter of 1993. The occurrence of proton flares during each quarter is indicated at the top of the figure.

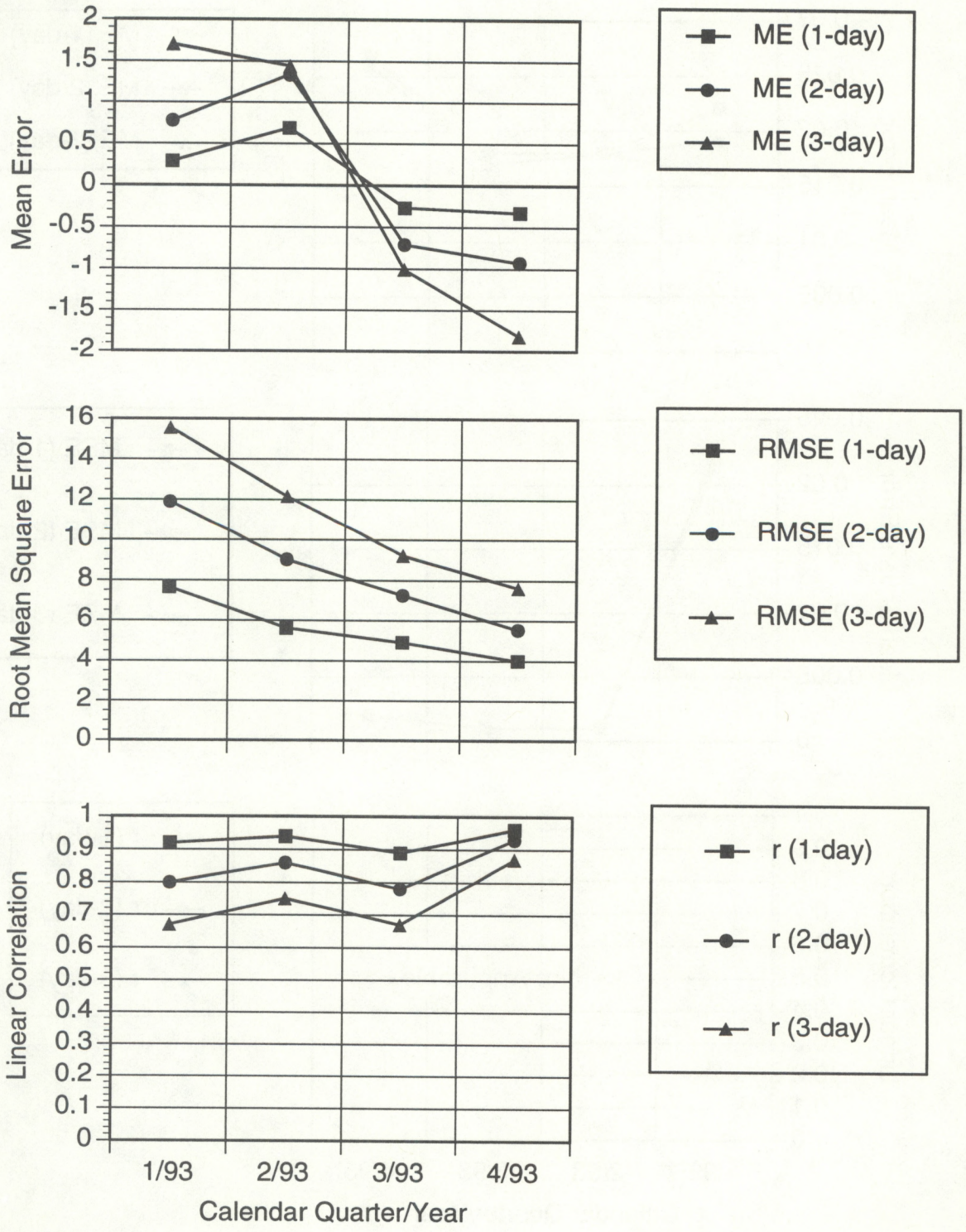



**Figure I-4.** Trend diagrams for 10.7 cm solar flux forecasts at three lead times. The overall bias (mean error), accuracy (root mean square error), and linear correlation are plotted for each calendar quarter of 1993.

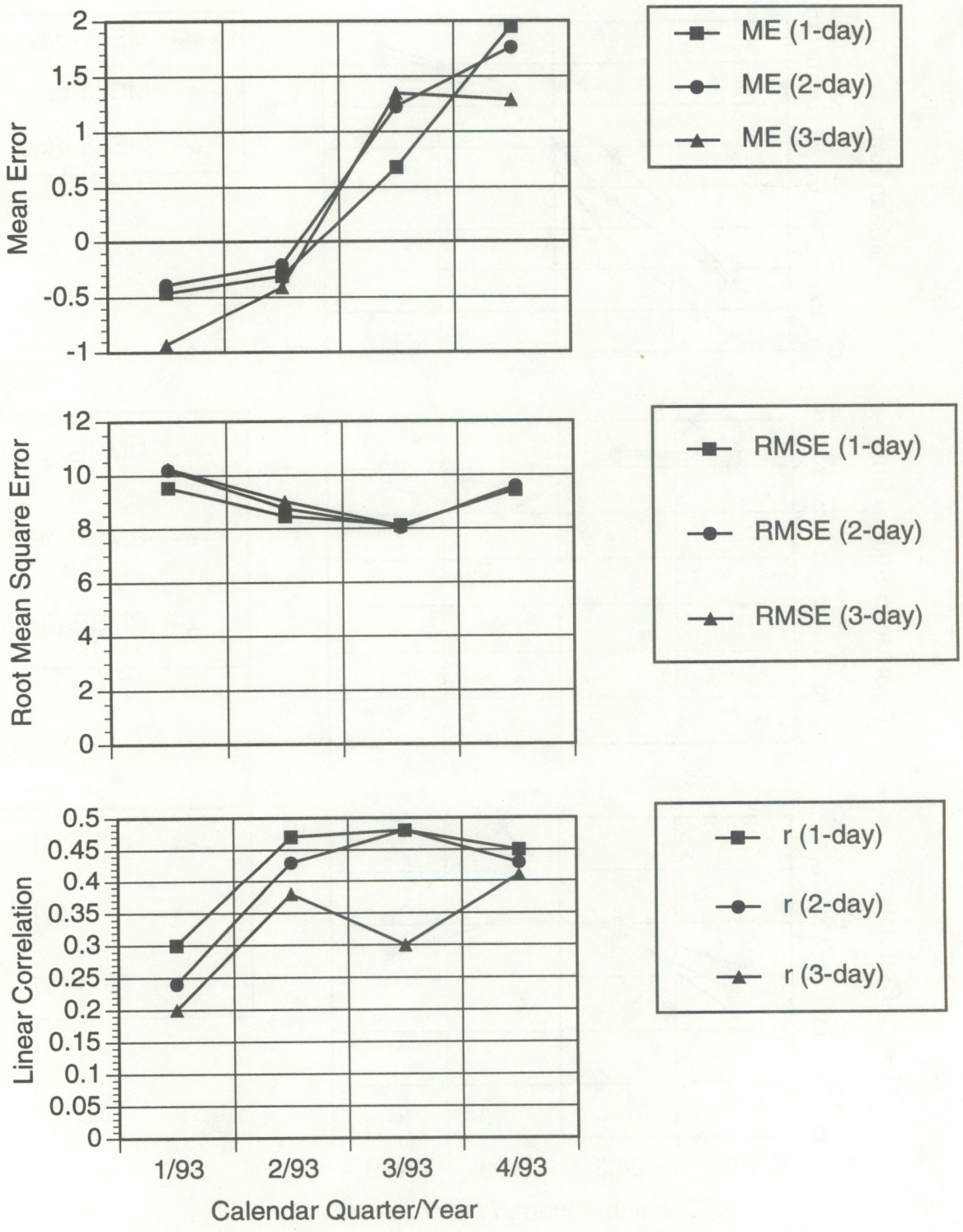



**Figure I-5.** Trend diagrams for Fredericksburg A-Index forecasts at three lead times. The overall bias (mean error), accuracy (root mean square error), and linear correlation are plotted for each calendar quarter of 1993.

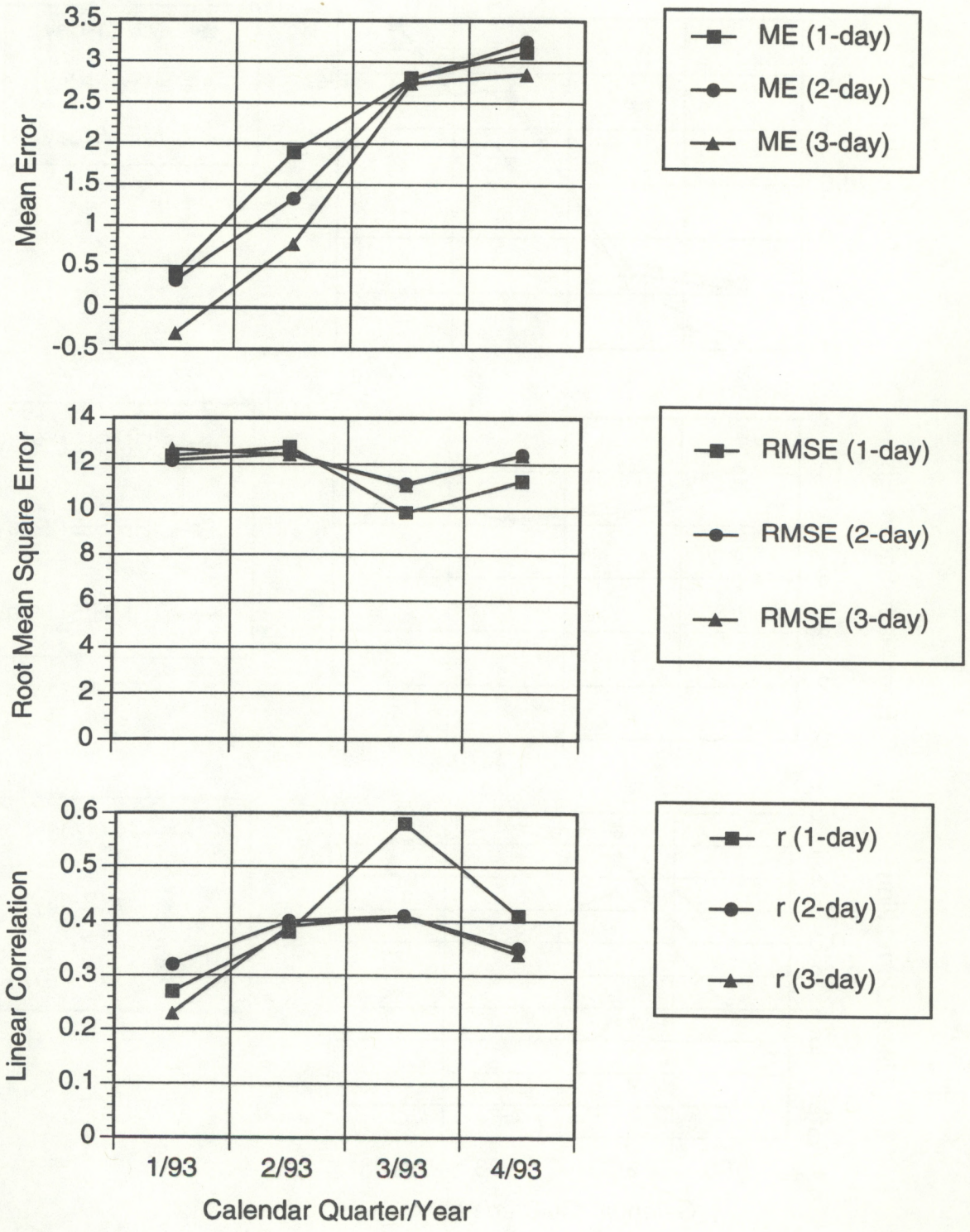



**Figure I-6.** Trend diagrams for Planetary A-Index forecasts at three lead times. The overall bias (mean error), accuracy (root mean square error), and linear correlation are plotted for each calendar quarter of 1993.

**Table I-2.** Overview statistics for next-day, middle-latitude, geomagnetic probability forecasts (1) (2), 1 January–31 December 1993

| | RPS | RPSsc | REL | RES | RES' |
|---|---|---|---|---|---|
| All SEL Forecasters | 0.432 | 0.466 | 0.162 | 0.197 | 0.270 |
| 30-day Climatology | 0.462 | 0.466 | 0.245 | 0.249 | 0.217 |

1. These forecasts are probability-type forecasts of the geomagnetic activity category of the most-disturbed 3-hour period during the next day. The 30-day climatology forecast is the average activity over the most recent 30 days.

2. Summary and performance measures include
   RPS, ranked probability score (ranging from 0 to 3, smaller values better);
   RPSsc, ranked probability score for sample climatology (constant forecasts of the observed relative frequency for the sample);
   REL, reliability [6] ($\geq 0$, smaller values better);
   RES, resolution as defined in [7] ($\geq 0$, larger values better); and
   RES', resolution as defined in [8] ($\geq 0$, smaller values better).

The following relationships hold: RPS = RPSsc + REL – RES = REL + RES,' and RPSsc = RES + RES.' The resolution and reliability are calculated from subsamples (vector partitions) of the forecasts and observations [9].

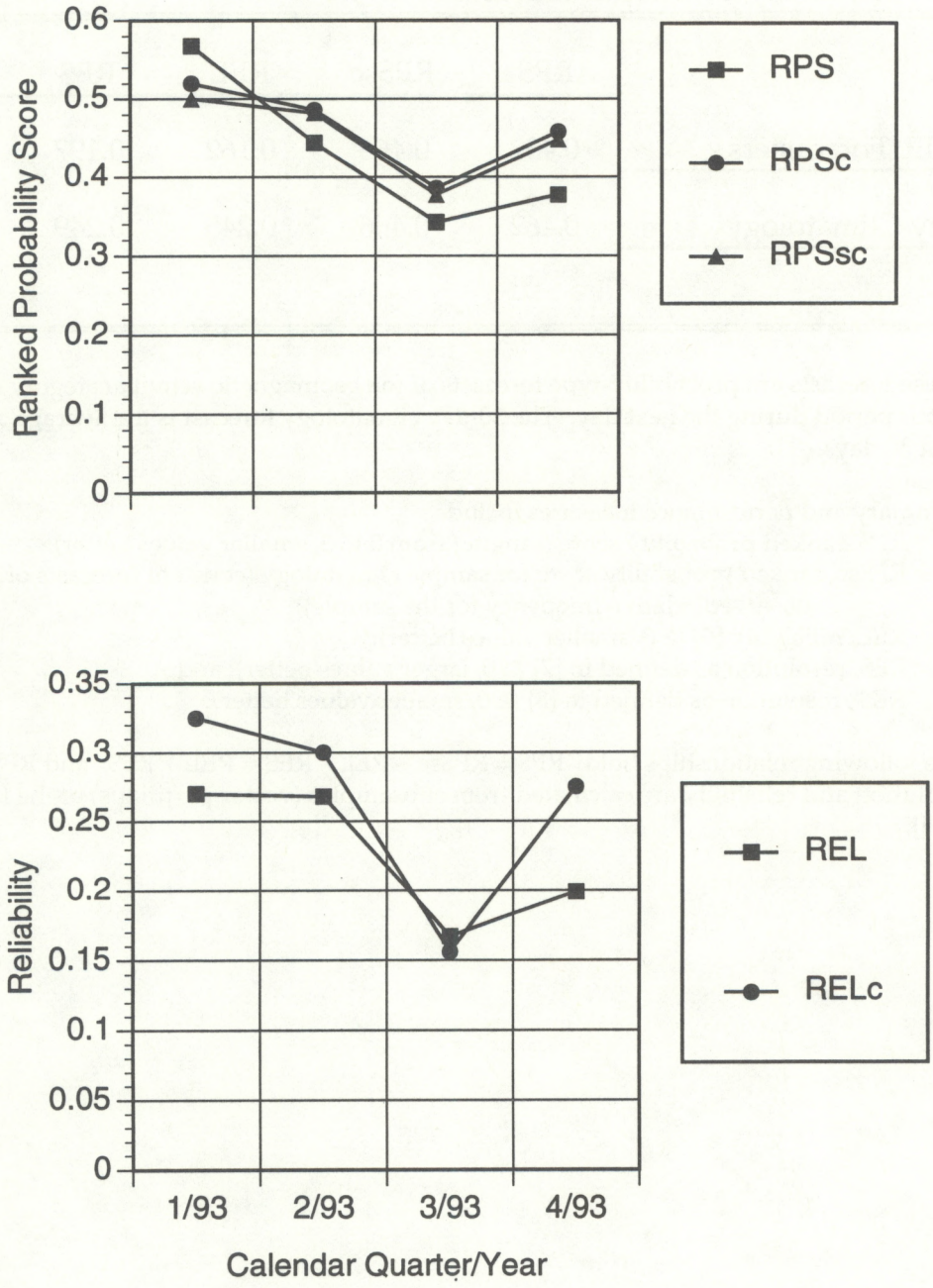



**Figure I-7.** Trend diagrams for probability forecasts of geomagnetic activity at a 1-day lead time. The ranked probability score (RPS) and reliability (REL) are plotted for each calendar quarter of 1993. Data are for all SEL forecasters (no subscript) and 30-day climatology forecasts (subscript c). RPSsc is the Ranked Probability Score for a sample climatology forecast. See Table I.2 notes for further explanation.

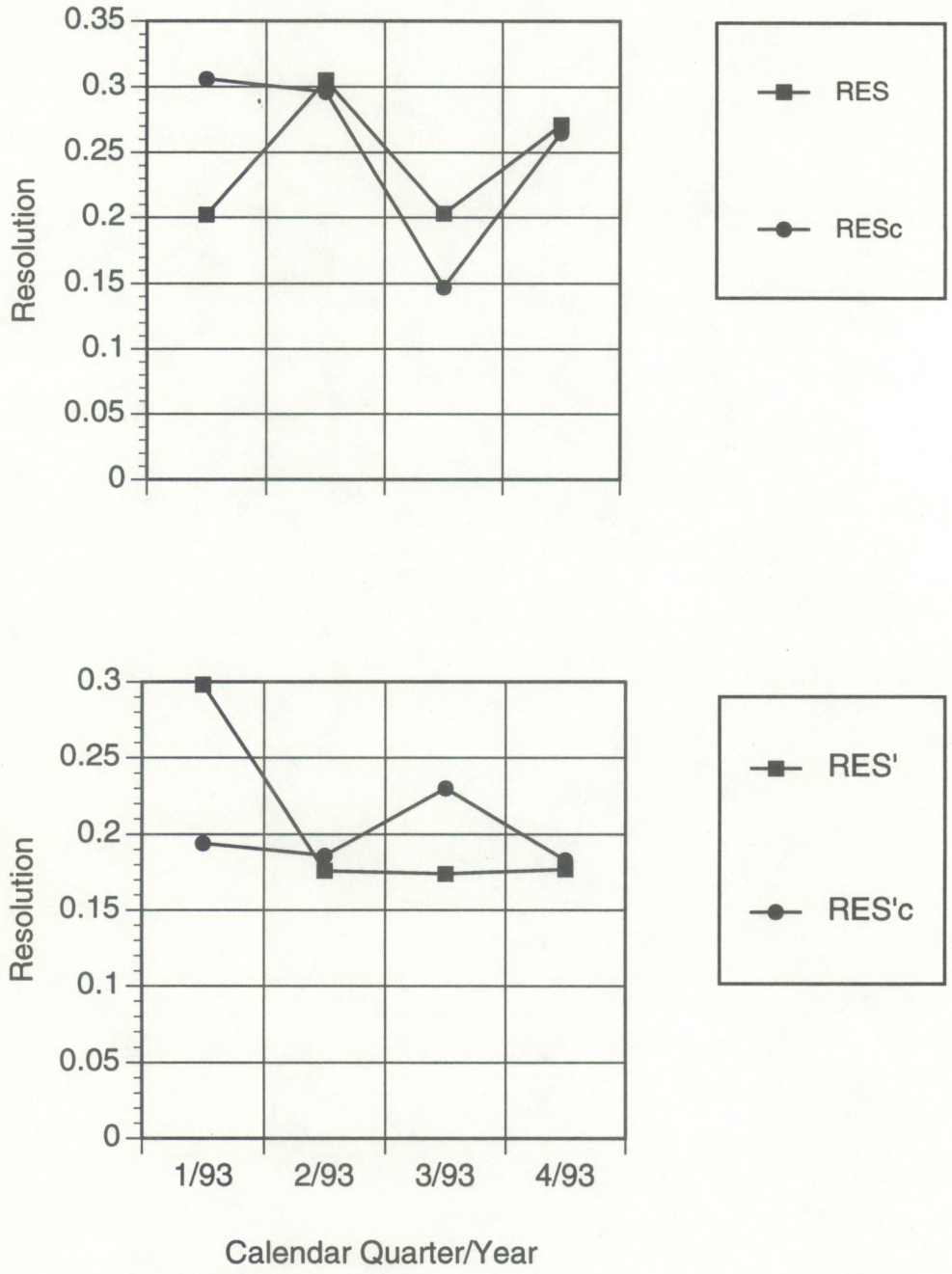



**Figure I-8.** Trend diagrams for probability forecasts of geomagnetic activity at a 1-day lead time. Two types of resolution (RES and RES') are plotted for each calendar quarter of 1993. Data are for all SESC forecasters (no subscript) and 30-day climatology forecasts (subscript c). See Table I.2 notes for further explanation.

# II. FORECAST VERIFICATION DETAILS

This section comprises detailed verification information for each of the following forecasts: probability forecasts of M-class, X-class, and proton flares expected for each of the next three days; integer-value forecasts of 10.7 cm solar flux, Fredericksburg A-Index, and Planetary A-Index for each of the next seven days; and next-day probability forecasts of geomagnetic activity levels at middle latitudes. These verification measures are summarized in seven different tables, grouped by forecast type. Several different verification diagrams are also presented for each of the forecasts. In many cases, to reduce the volume of information, emphasis is placed on a 1-day forecast lead time.

**Table II-1.** Verification statistics for M-class flare forecasts [1] [2], 1 January–31 December 1993

| Lead Time (days) | n=365 <f> | <x> | <f–x> | Med(f) | Med(f-x) | s(f) | s(x) | s(f-x) |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.191 | 0.153 | 0.038 | 0.100 | 0.050 | 0.202 | 0.360 | 0.350 |
| 2 | 0.189 | 0.153 | 0.036 | 0.100 | 0.050 | 0.203 | 0.360 | 0.364 |
| 3 | 0.185 | 0.153 | 0.031 | 0.100 | 0.050 | 0.201 | 0.360 | 0.373 |

| | <f \| x=1> | <f \| x=0> | Med (f \| x=1) | Med (f \| x=0) | <f \| x=1> – <f \| x=0> | Med (f \| x=1) – Med (f \| x=0) | s(f \| x=1) | s(f \| x=0) |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.348 | 0.163 | 0.300 | 0.100 | 0.186 | 0.200 | 0.207 | 0.187 |
| 2 | 0.315 | 0.167 | 0.275 | 0.100 | 0.148 | 0.175 | 0.212 | 0.193 |
| 3 | 0.287 | 0.166 | 0.225 | 0.100 | 0.120 | 0.125 | 0.221 | 0.192 |

1. M-class flare forecasts are probabilistic forecasts ranging from 0.01 to 0.99. Corresponding observations are either 1 (M-flare occurred) or 0 (no M-flare occurred).

2. These summary and performance measures include
   n, number of forecasts/observations in the sample;
   <f>, mean forecast;
   <x>, mean observation;
   <f-x>, mean error;
   Med(f), median forecast;
   Med(f-x), median error;
   s(f), s(x), and s(f-x), standard deviations of the forecasts, observations, and errors;
   <f | x=1>, mean of the conditional distribution of the forecasts given the occurrence of the event;
   <f | x=0>, mean of the conditional distribution of the forecasts given the non-occurrence of the event;
   Med(f | x=1) and Med(f | x=0), medians of the conditional distributions;
   <f | x=1> - <f | x=0>, the difference of the means of the conditional distributions (This quantity is related to the discrimination of the forecasts.);
   Med(f | x=1) - Med(f | x=0) the difference of the medians of the conditional distributions; and
   s(f | x=1) and s(f | x=0), standard deviations of the conditional distributions.

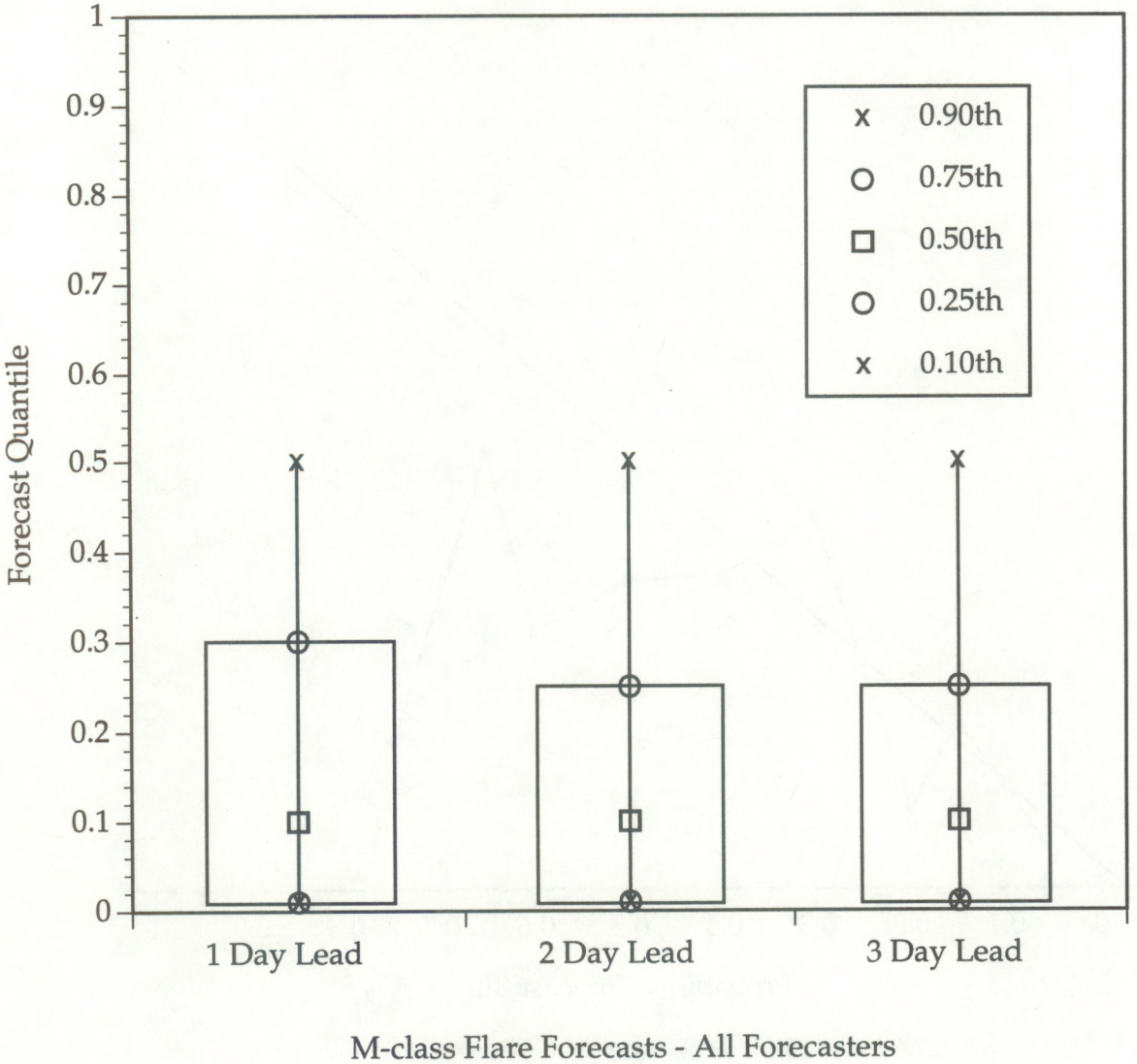



**Figure II-1.** Box plots showing the marginal distribution of M-class flare forecasts for three lead times, 1 January–31 December 1993. The following quantiles are plotted (from top to bottom): 0.90th, 0.75th (upper quartile), 0.50th (median), 0.25th (lower quartile), and 0.10th.

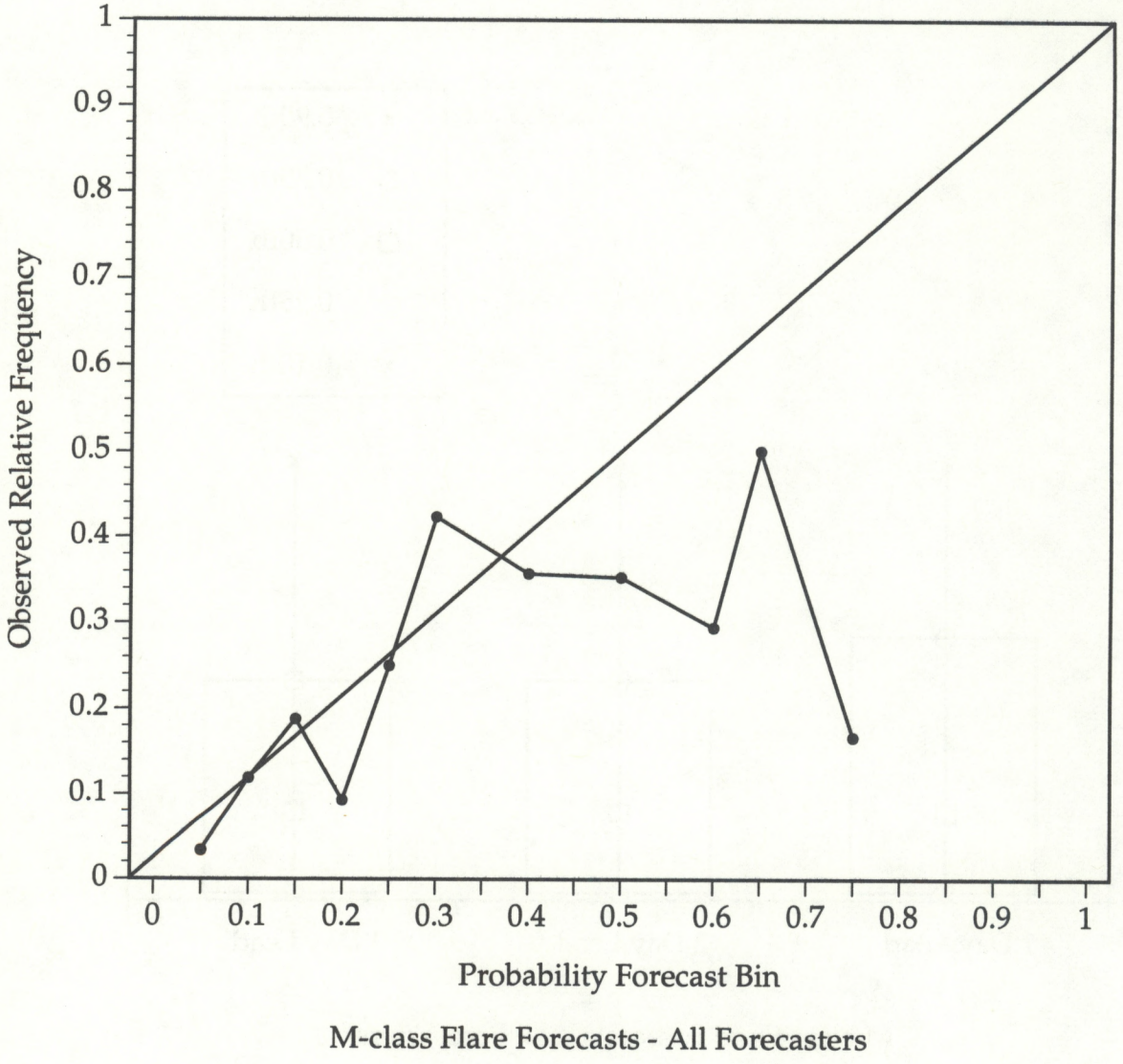



**Figure II-2.** Reliability diagram for next-day M-class flare forecasts, 1 January–31 December 1993. The 45° diagonal indicates perfect correspondence between forecasts and observations.

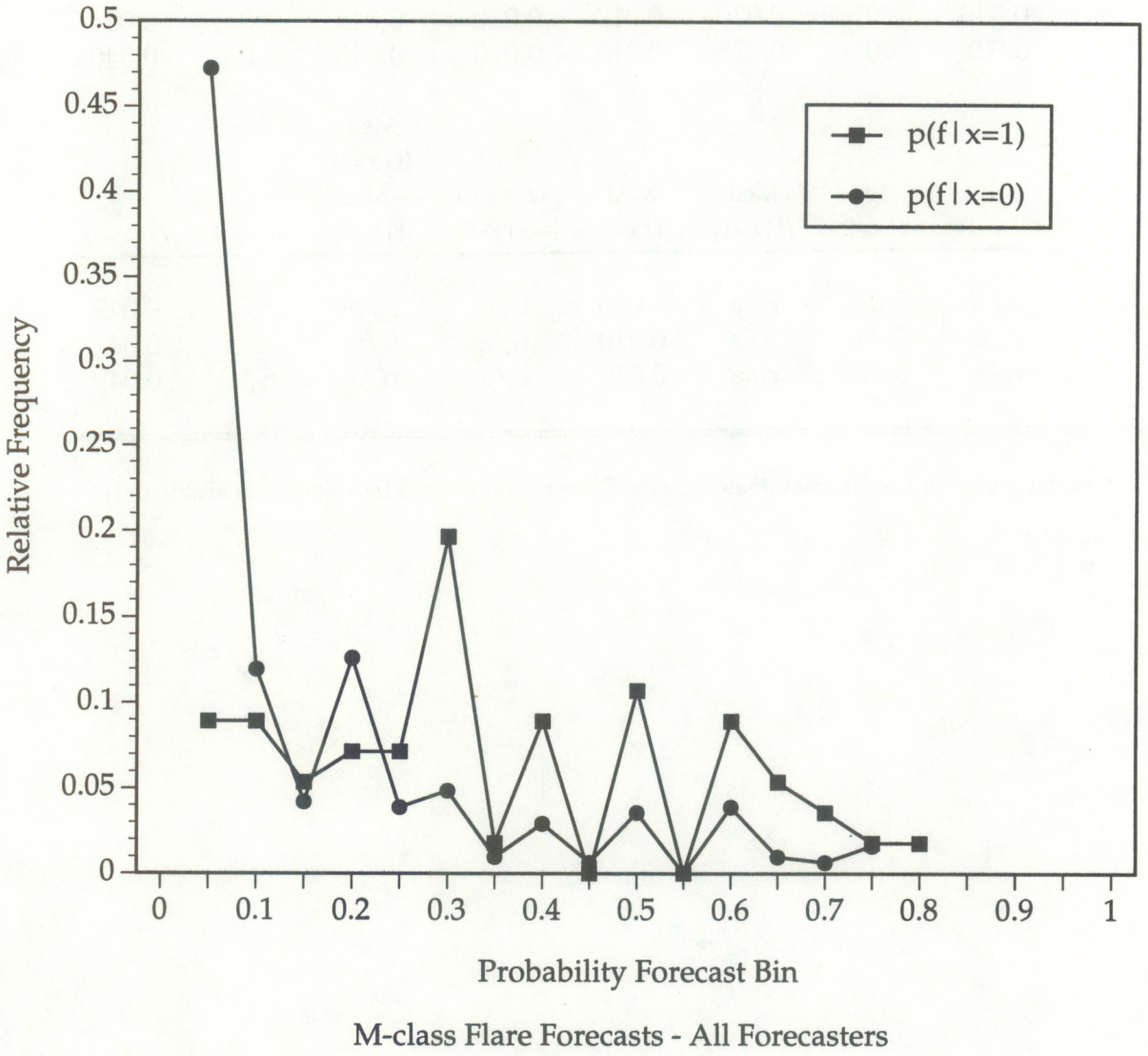



**Figure II-3.** Discrimination diagram for next-day M-class flare forecasts, 1 January–31 December 1993. This diagram shows the frequency with which each forecast value was used when an event occurred (p(f | x=1)) and when an event did not occur (p(f | x=0)).

**Table II-2.** Verification statistics for X-class flare forecasts [1], 1 January–31 December 1993

| Lead Time (days) | n=365 <f> | <x> | <f–x> | Med(f) | Med(f-x) | s(f) | s(x) | s(f-x) |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.030 | 0 | 0.030 | 0.010 | 0.010 | 0.042 | 0 | 0.042 |
| 2 | 0.030 | 0 | 0.030 | 0.010 | 0.010 | 0.041 | 0 | 0.041 |
| 3 | 0.028 | 0 | 0.028 | 0.010 | 0.010 | 0.040 | 0 | 0.040 |

| | <f\|x=1> | <f\|x=0> | Med (f\|x=1) | Med (f\|x=0) | <f\|x=1> – <f\|x=0> | Med (f\|x=1) – Med (f\|x=0) | s(f\|x=1) | s(f\|x=0) |
|---|---|---|---|---|---|---|---|---|
| 1 | n/a | 0.030 | n/a | 0.010 | n/a | n/a | n/a | 0.042 |
| 2 | n/a | 0.030 | n/a | 0.010 | n/a | n/a | n/a | 0.041 |
| 3 | n/a | 0.028 | n/a | 0.010 | n/a | n/a | n/a | 0.040 |

1. Forecast type and summary statistics are the same as for M-class flare forecasts (Table II-1).

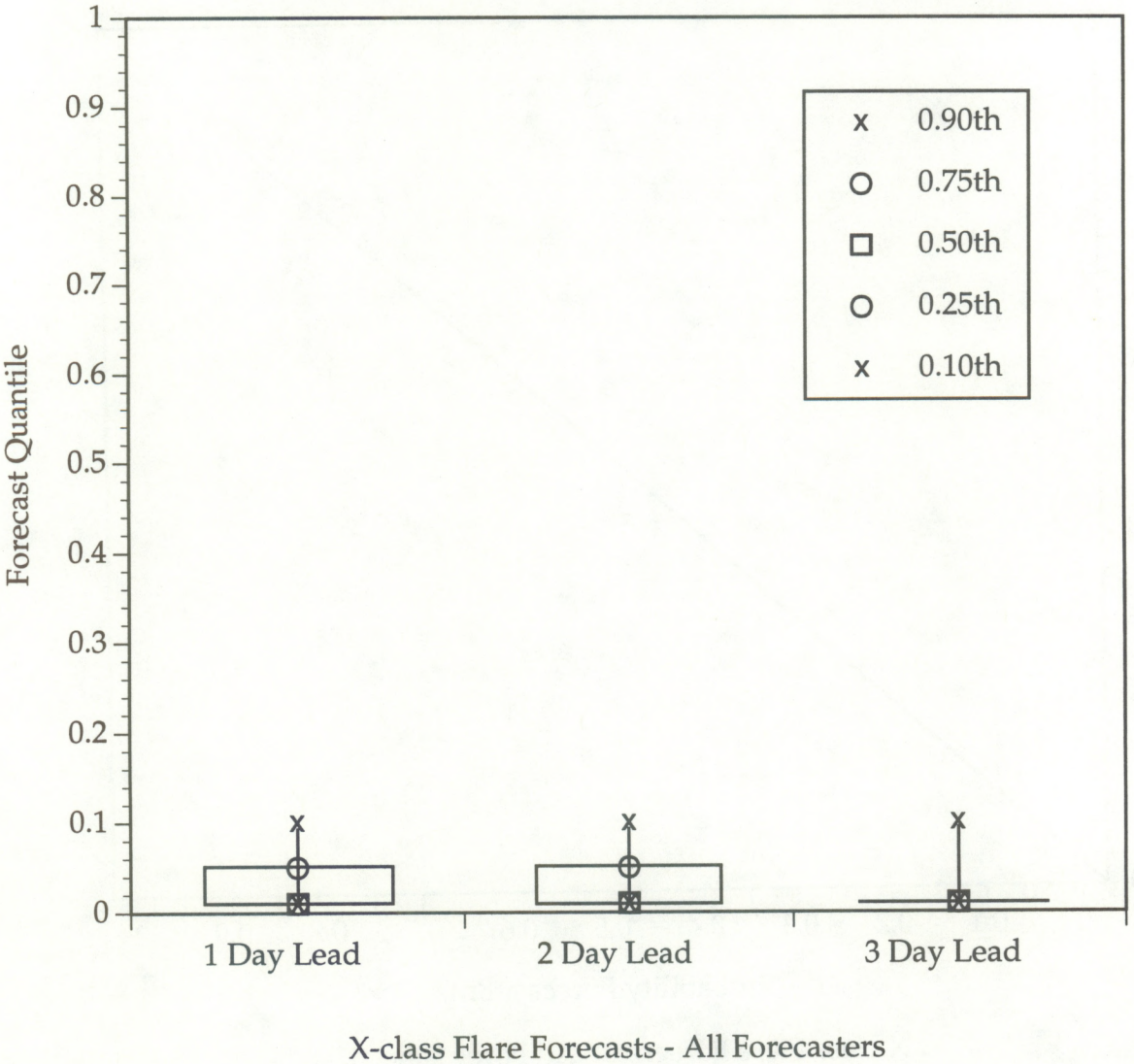



**Figure II-4.** Box plots showing the marginal distribution of X-class flare forecasts for three lead times, 1 January–31 December 1993. The following quantiles are plotted (from top to bottom): 0.90th, 0.75th (upper quartile), 0.50th (median), 0.25th (lower quartile), and 0.10th.

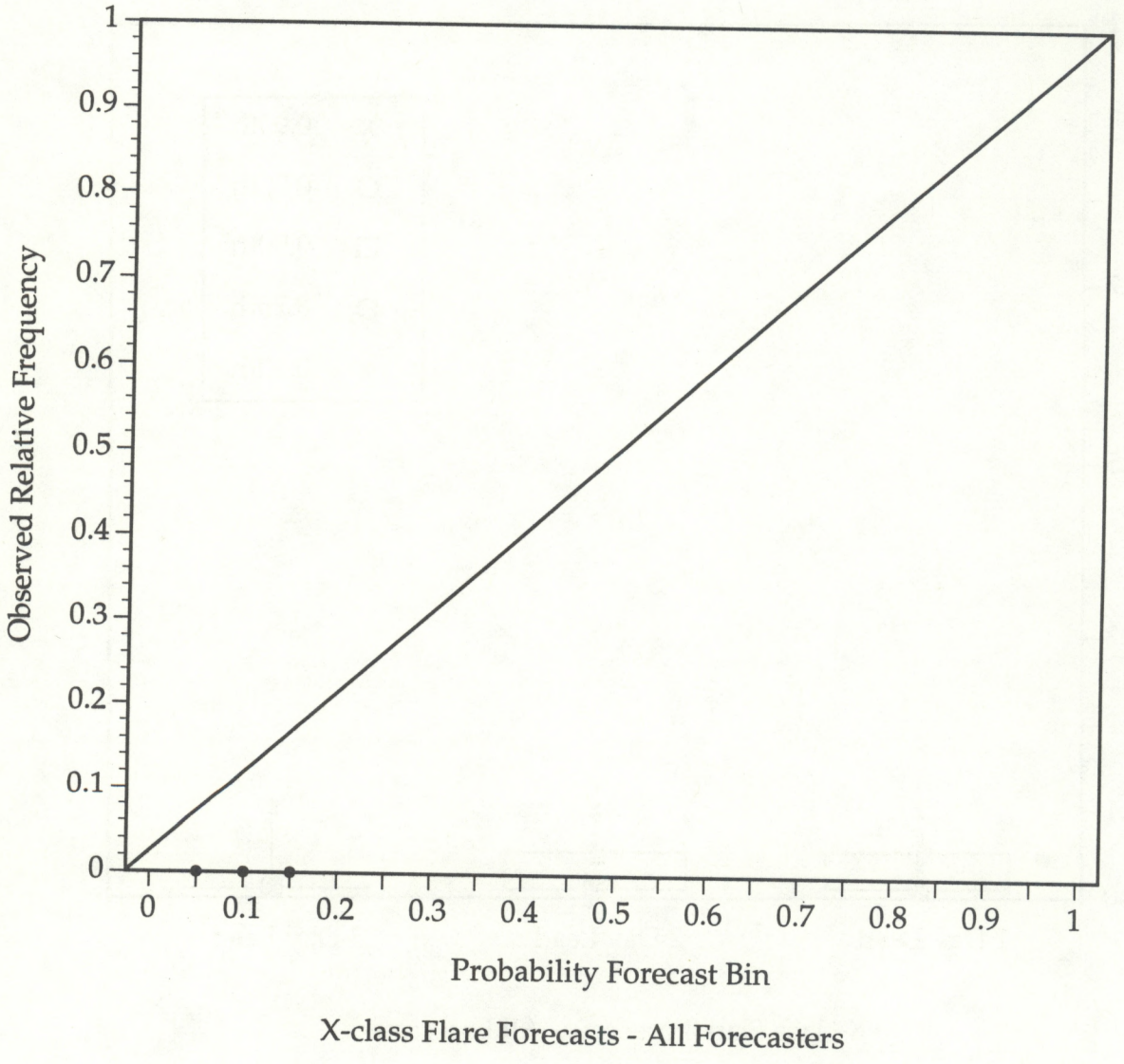



**Figure II-5.** Reliability diagram for next-day X-class flare forecasts, 1 January–31 December 1993. The 45° diagonal indicates perfect correspondence between forecasts and observations.

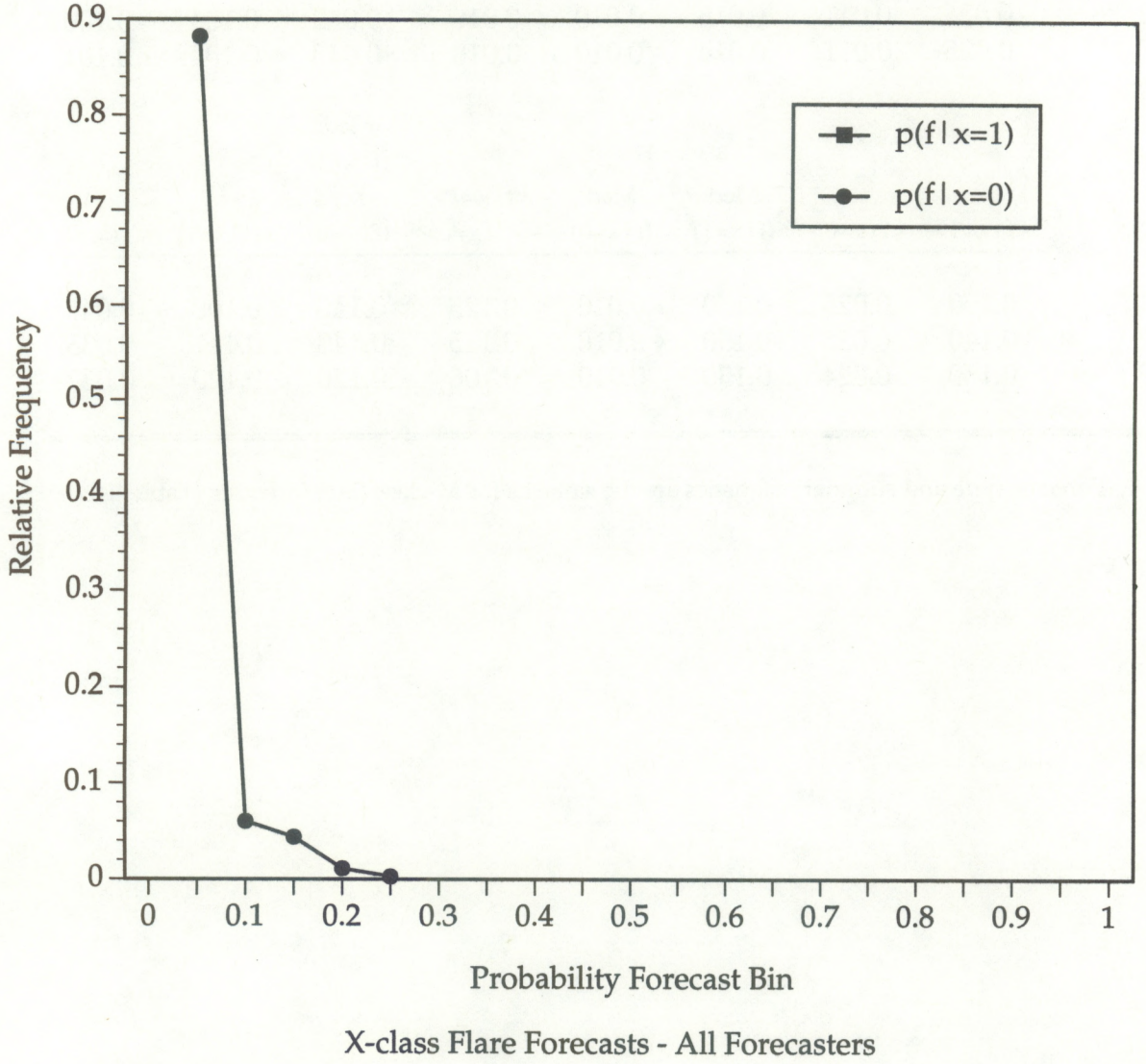



**Figure II-6.** Discrimination diagram for next-day X-class flare forecasts, 1 January–31 December 1993. This diagram shows the frequency with which each forecast value was used when an event occurred (p(f | x=1)) and when an event did not occur (p(f | x=0)).

**Table II-3.** Verification statistics for proton flare forecasts [1], 1 January–31 December 1993

| Lead-Time (days) | n=365 <f> | <x> | <f–x> | Med(f) | Med(f-x) | s(f) | s(x) | s(f-x) |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.027 | 0.011 | 0.016 | 0.010 | 0.010 | 0.043 | 0.104 | 0.100 |
| 2 | 0.026 | 0.011 | 0.015 | 0.010 | 0.010 | 0.042 | 0.104 | 0.100 |
| 3 | 0.025 | 0.011 | 0.014 | 0.010 | 0.010 | 0.040 | 0.104 | 0.101 |

| | <f \| x=1> | <f \| x=0> | Med (f \| x=1) | Med (f \| x=0) | <f \| x=1> – <f \| x=0> | Med (f \| x=1) – Med (f \| x=0) | s(f \| x=1) | s(f \| x=0) |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.150 | 0.025 | 0.150 | 0.010 | 0.125 | 0.140 | 0.100 | 0.040 |
| 2 | 0.140 | 0.025 | 0.150 | 0.010 | 0.115 | 0.140 | 0.111 | 0.038 |
| 3 | 0.130 | 0.024 | 0.130 | 0.010 | 0.106 | 0.120 | 0.120 | 0.037 |

1. Forecast type and summary statistics are the same as for M-class flare forecasts (Table II-1).

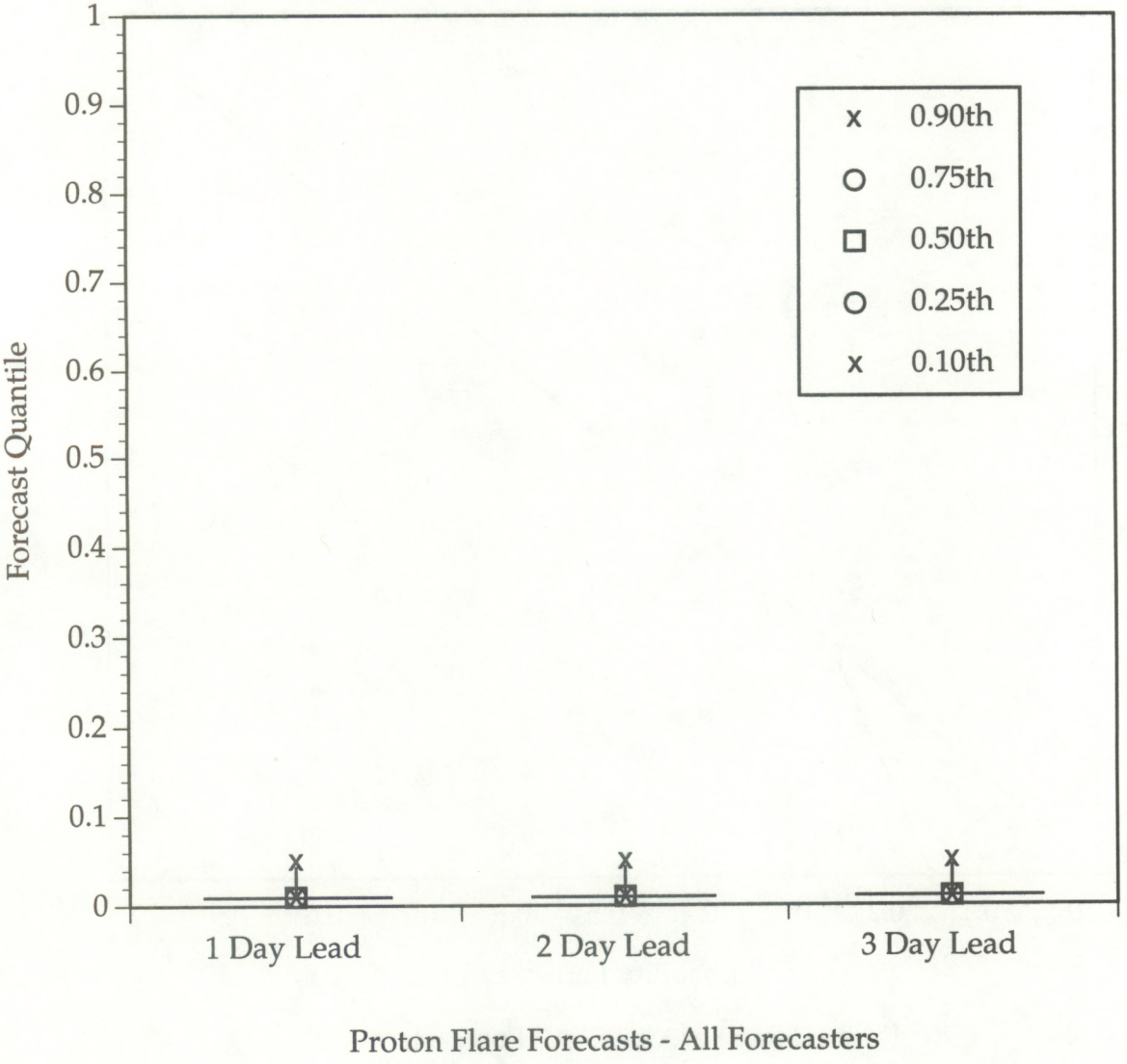



**Figure II-7.** Box plots showing the marginal distribution of proton flare forecasts for three lead times, 1 January–31 December 1993. The following quantiles are plotted (from top to bottom): 0.90th, 0.75th (upper quartile), 0.50th (median), 0.25th (lower quartile), and 0.10th.

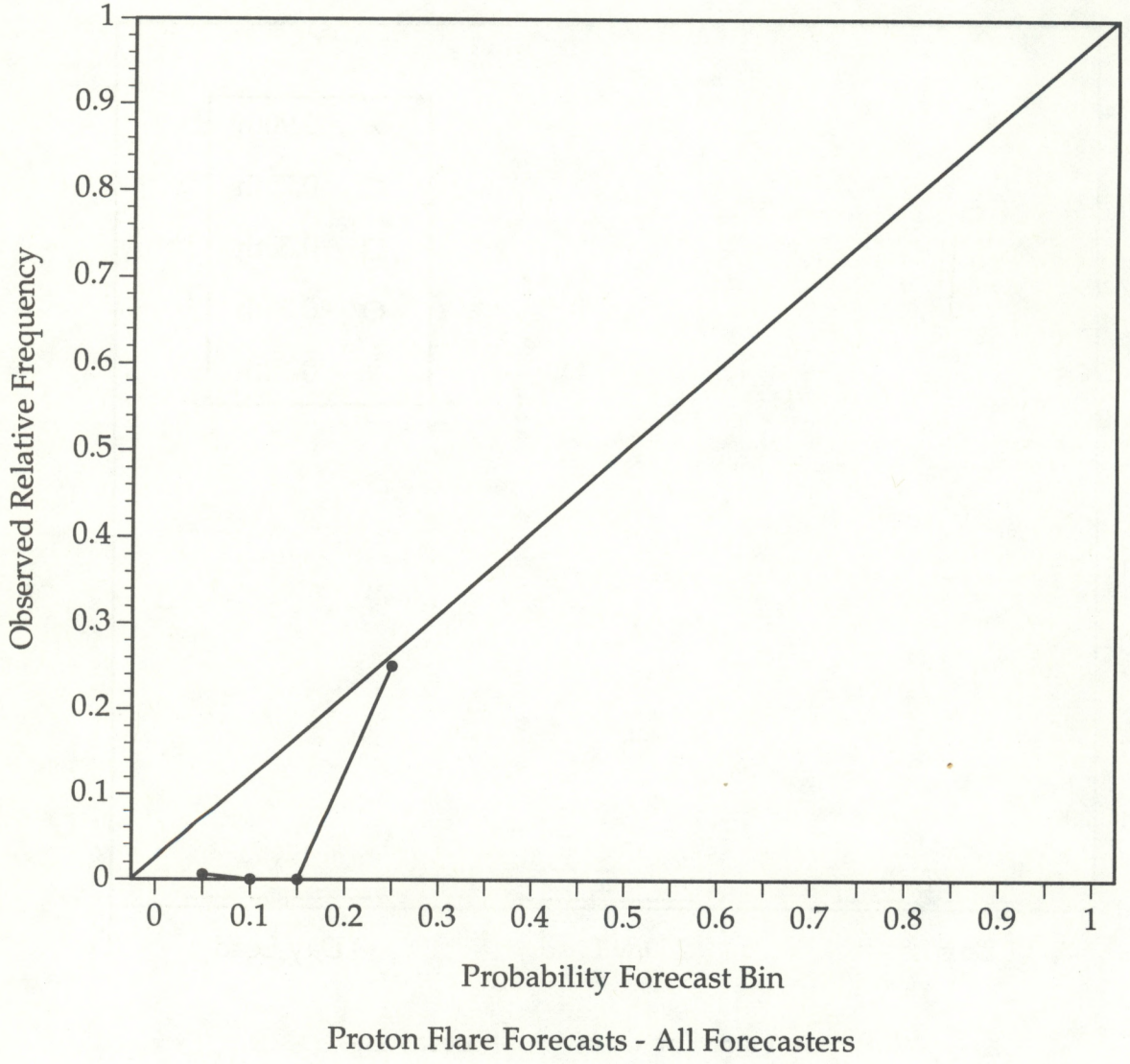



**Figure II-8.** Reliability diagram for next-day proton flare forecasts, 1 January–31 December 1993. The 45° diagonal indicates perfect correspondence between forecasts and observations.

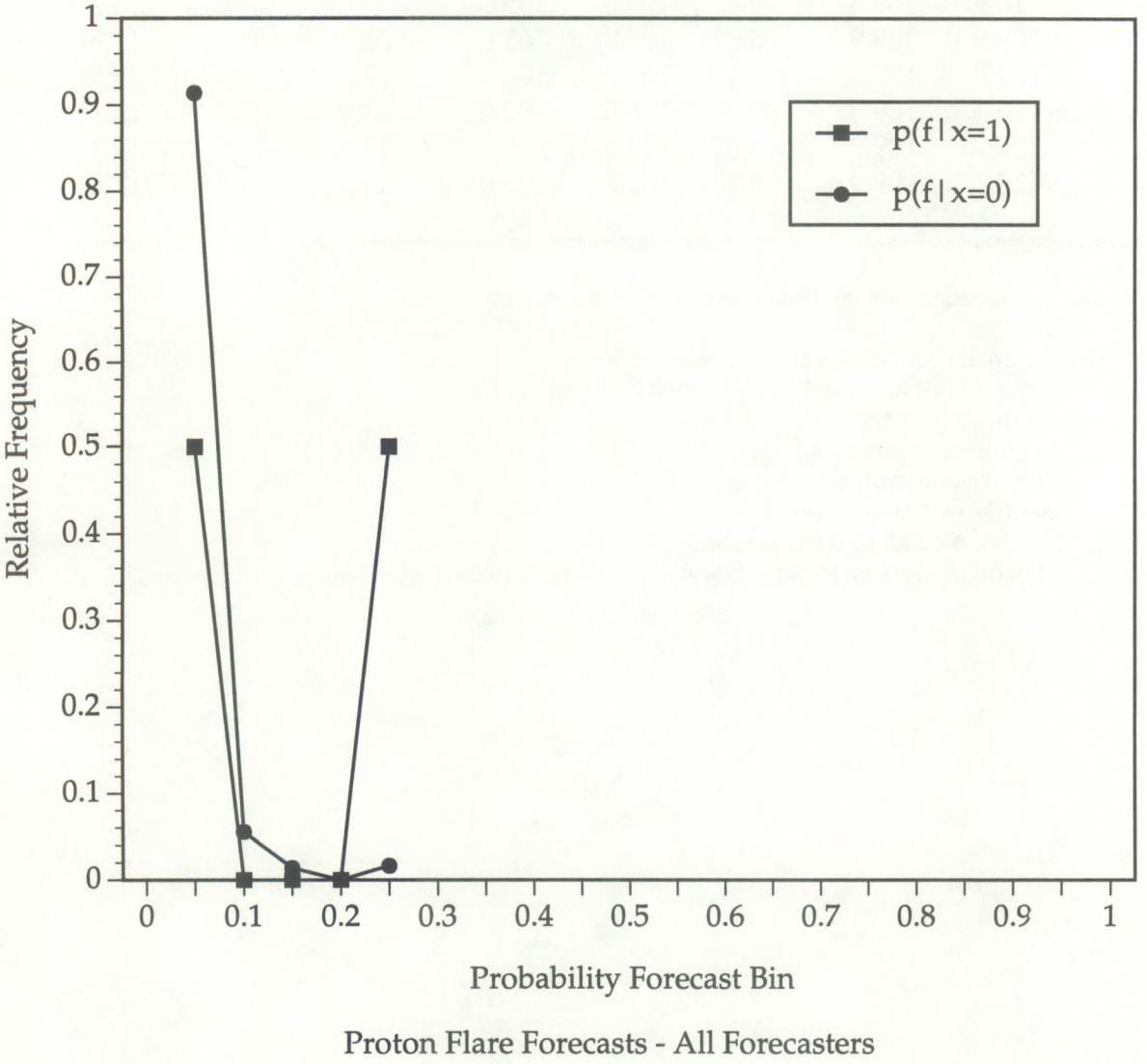



**Figure II-9.** Discrimination diagram for next-day proton flare forecasts, 1 January–31 December 1993. This diagram shows the frequency with which each forecast value was used when an event occurred (p(f | x=1)) and when an event did not occur (p(f | x=0)).

**Table II-4.** Verification statistics for 10.7 cm solar flux forecasts (1) (2), 1 January–31 December 1993

| Lead Time (days) | n=365 <f> | <x> | <f–x> | Med(f) | Med(x) | s(f) | s(x) | s(f-x) |
|---|---|---|---|---|---|---|---|---|
| 1 | 109.9 | 109.8 | 0.09 | 105.0 | 105.0 | 21.6 | 21.3 | 5.7 |
| 2 | 110.0 | 109.8 | 0.11 | 105.0 | 105.0 | 22.0 | 21.4 | 8.7 |
| 3 | 109.9 | 109.9 | 0.07 | 105.0 | 105.0 | 22.4 | 21.4 | 11.5 |
| 4 | 109.8 | 109.9 | – 0.13 | 105.0 | 105.0 | 22.7 | 21.4 | 13.5 |
| 5 | 110.0 | 109.9 | 0.06 | 105.0 | 105.0 | 23.0 | 21.5 | 15.1 |
| 6 | 109.9 | 109.9 | 0.02 | 105.0 | 105.0 | 23.1 | 21.5 | 16.3 |
| 7 | 110.1 | 109.9 | 0.18 | 105.0 | 105.0 | 23.2 | 21.4 | 17.5 |

1. Solar flux forecasts are continuous-type integer forecasts.

2. These summary and performance measures include
   n, number of forecasts/observations in the sample;
   <f>, mean forecast;
   <x>, mean observation;
   <f-x>, mean error;
   Med(f), median forecast;
   Med(x), median observation; and
   s(f), s(x), and s(f-x), standard deviations of the forecasts, observations, and errors.

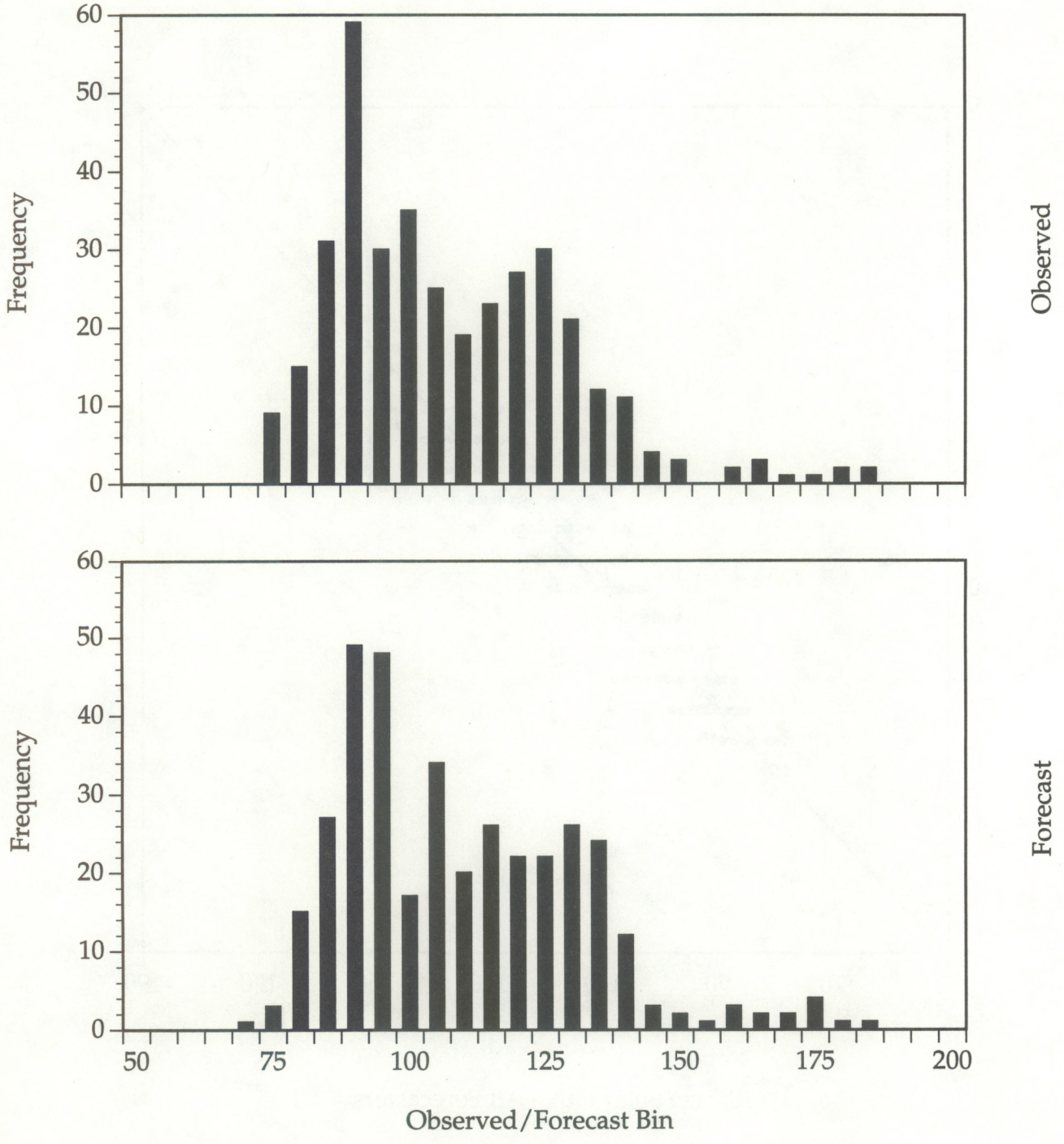



**Figure II-10.** Histograms of 10.7 cm solar flux observations and next-day forecasts, 1 January–31 December 1993.

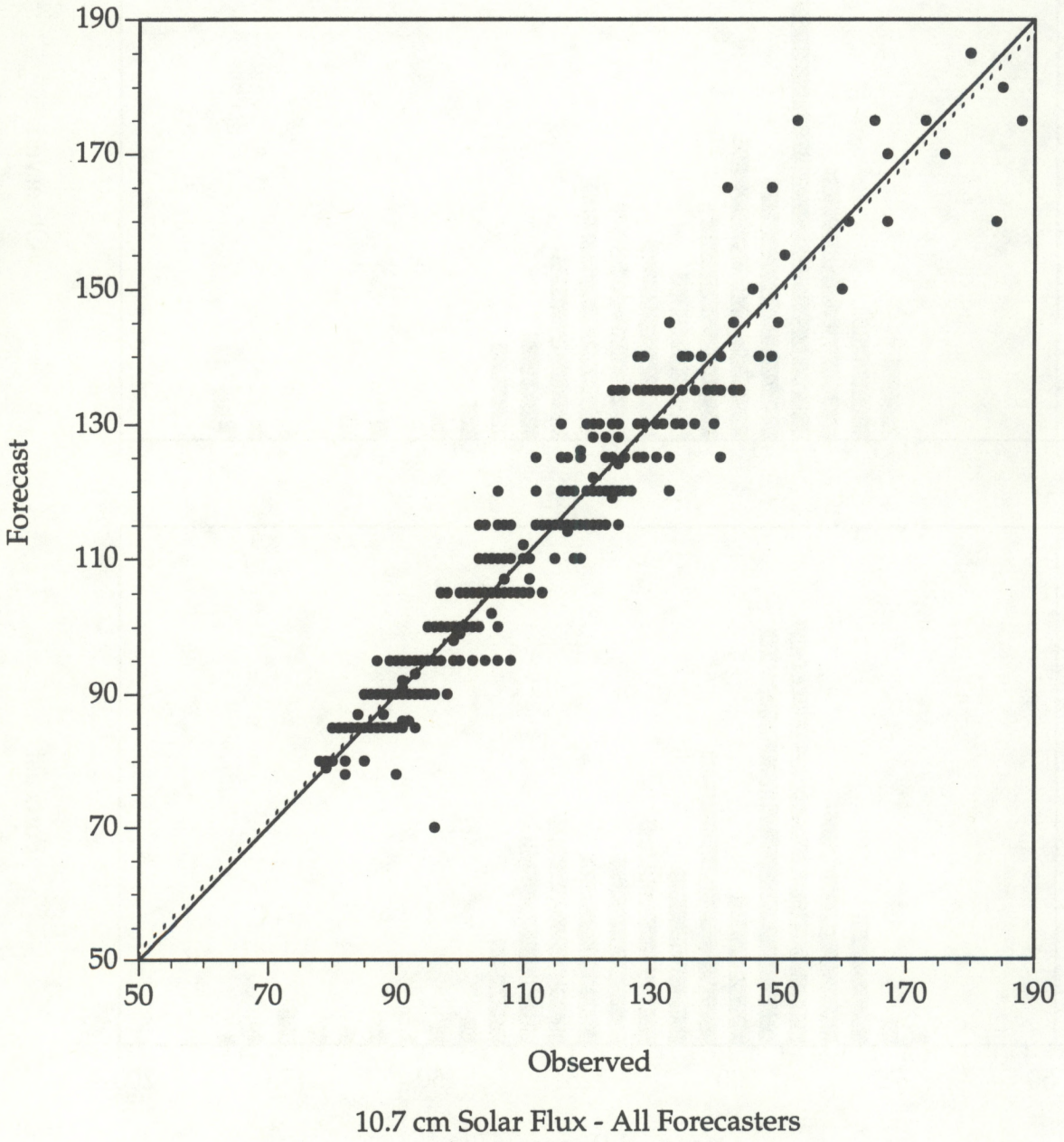
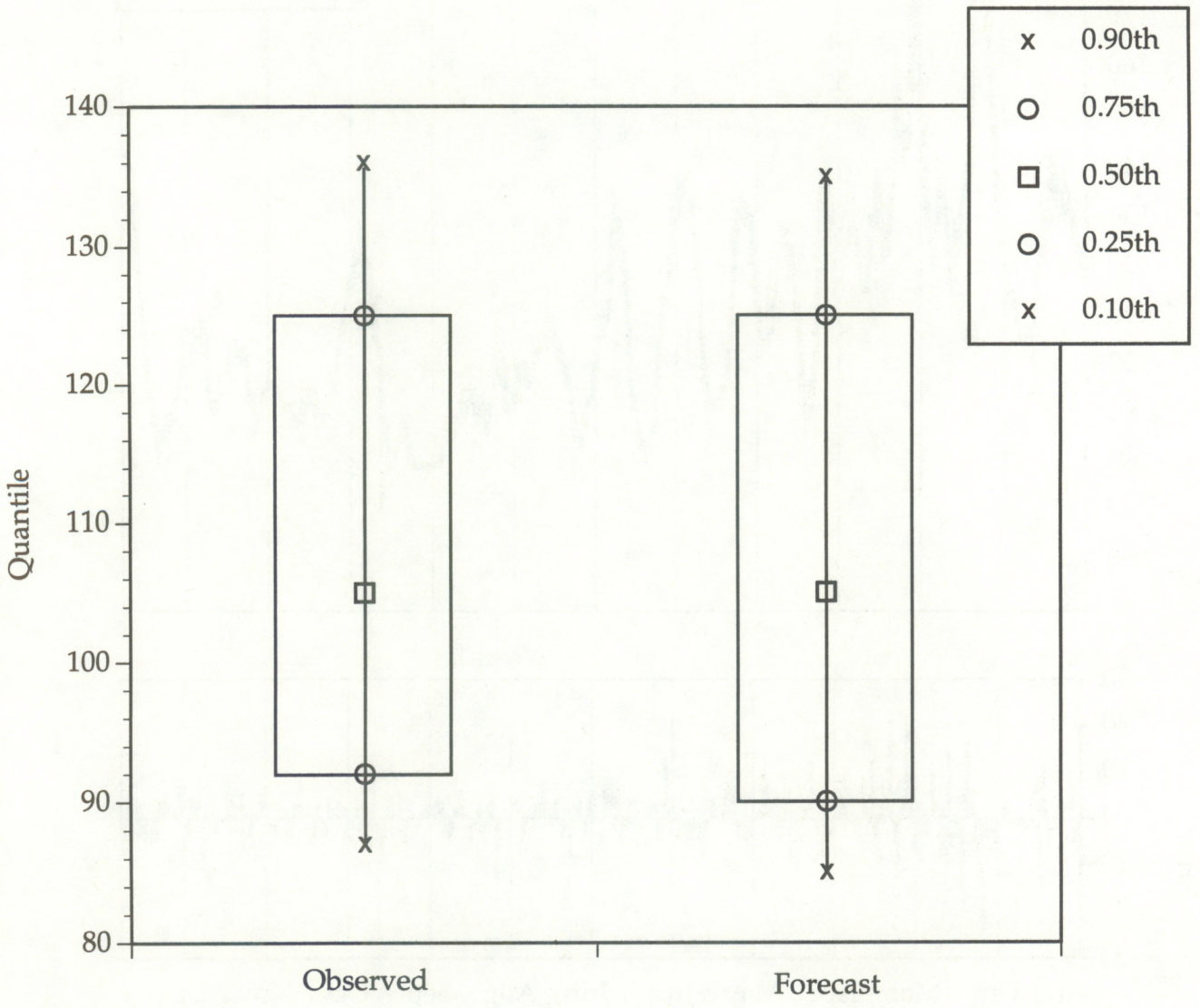



**Figure II-11.** Scatter diagram showing the joint distribution of next-day 10.7 cm solar flux forecasts and observations, 1 January–31 December 1993. The 45° diagonal (solid) indicates perfect correspondence between forecasts and observations. The dotted line is a least-squares fit (r=0.96) to the data.




10.7 cm Solar Flux - All Forecasters

**Figure II-12.** Box plots showing the marginal distribution of 10.7 cm solar flux observations and next-day forecasts, 1 January–31 December 1993. The following quantiles are plotted (from top to bottom): 0.90th, 0.75th (upper quartile), 0.50th (median), 0.25th (lower quartile), and 0.10th.

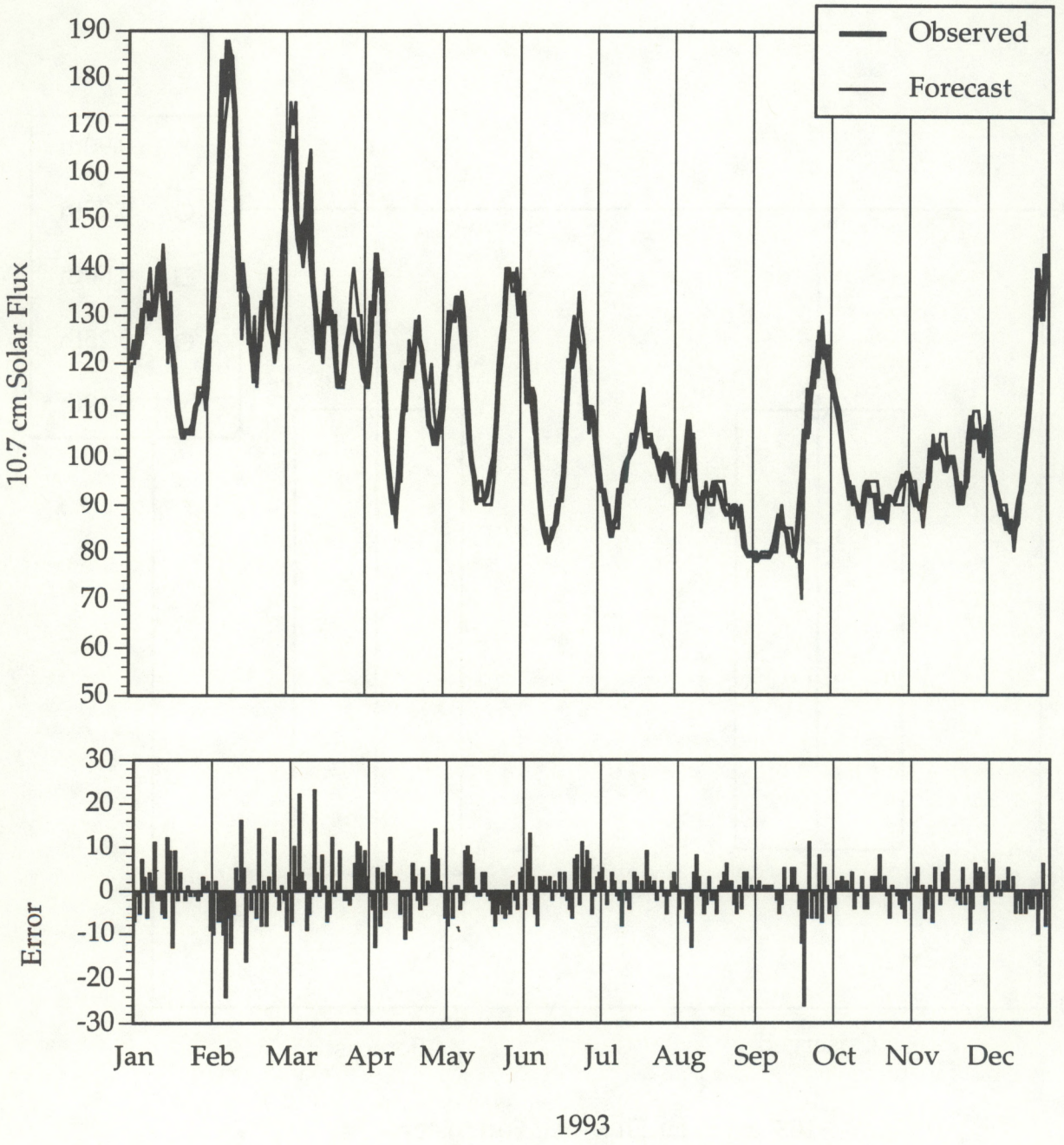



**Figure II-13.** Time history plot for next-day, 10.7 cm solar flux forecasts, 1 January–31 December 1993. Observed values are traced by the bold line. The errors (f-x) are shown in a separate plot at the bottom.

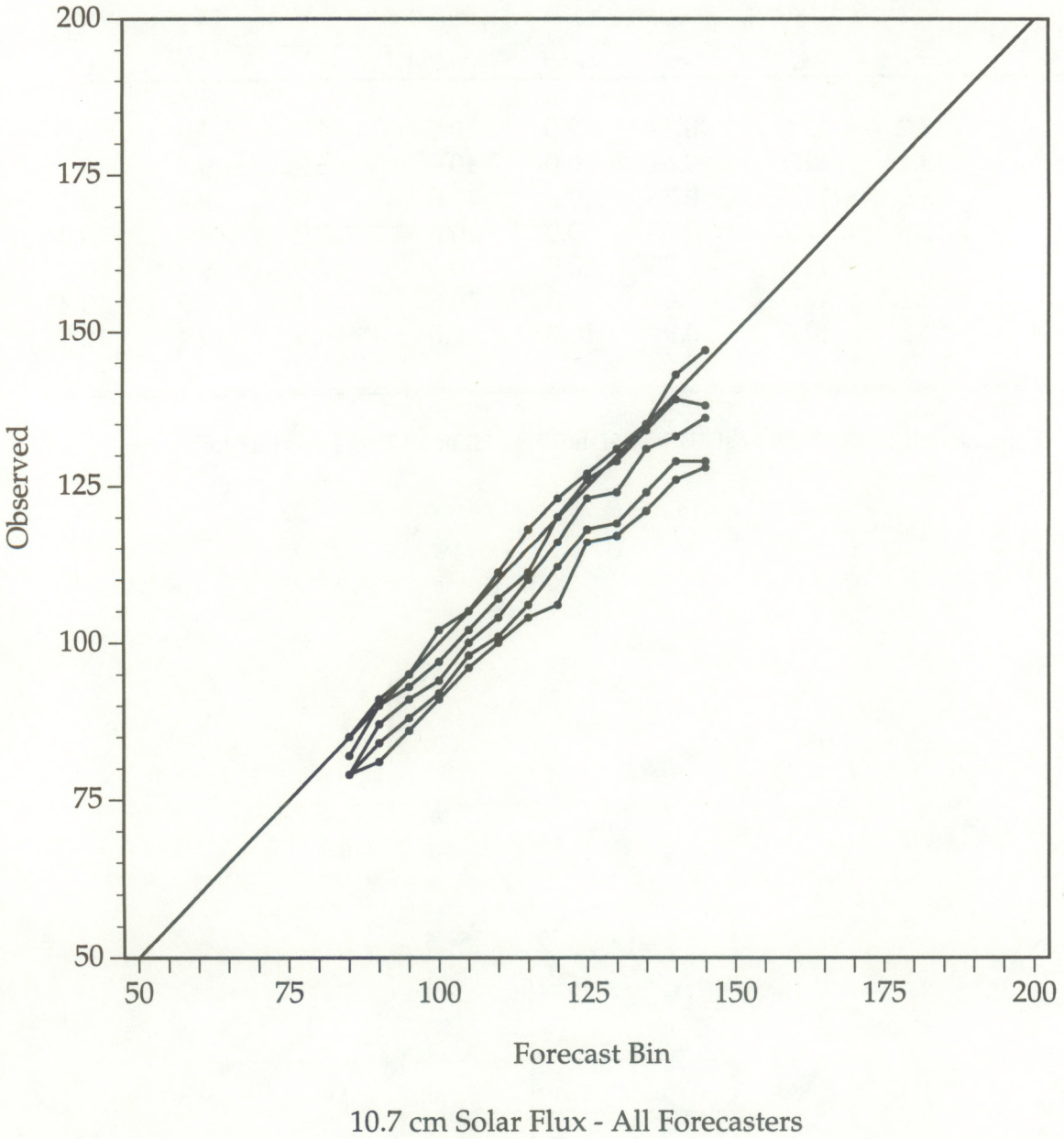



**Figure II-14.** Conditional quantile plot for next-day, 10.7 cm solar flux forecasts, 1 January–31 December 1993. The lines correspond to the following quantiles (top to bottom): 0.90th, 0.75th (upper quartile), 0.50th (median), 0.25th (lower quartile), and 0.10th. The 45° diagonal indicates perfect correspondence of forecasts and observations. The spread in the quantiles is related to the accuracy of the forecasts, and the deviation of the median from the diagonal is related to the overall bias.

**Table II-5.** Verification statistics for Fredericksburg A-Index forecasts [1], 1 January–31 December 1993

| Lead Time (days) | n=363 <f> | <x> | <f–x> | Med(f) | Med(x) | s(f) | s(x) | s(f-x) |
|---|---|---|---|---|---|---|---|---|
| 1 | 13.2 | 12.7 | 0.47 | 10.0 | 10.0 | 6.7 | 9.4 | 8.9 |
| 2 | 13.4 | 12.7 | 0.61 | 10.0 | 10.0 | 6.7 | 9.4 | 9.2 |
| 3 | 13.0 | 12.7 | 0.33 | 10.0 | 10.0 | 6.4 | 9.4 | 9.2 |
| 4 | 12.6 | 12.7 | – 0.07 | 10.0 | 10.0 | 7.2 | 9.4 | 10.0 |
| 5 | 12.3 | 12.6 | – 0.23 | 10.0 | 10.0 | 7.6 | 9.3 | 10.2 |
| 6 | 12.1 | 12.5 | – 0.47 | 10.0 | 10.0 | 7.3 | 9.3 | 10.3 |
| 7 | 11.6 | 12.6 | – 0.94 | 10.0 | 10.0 | 6.6 | 9.4 | 10.5 |

1. Forecast type and summary statistics are the same as for 10.7 cm solar flux forecasts (Table II-4).

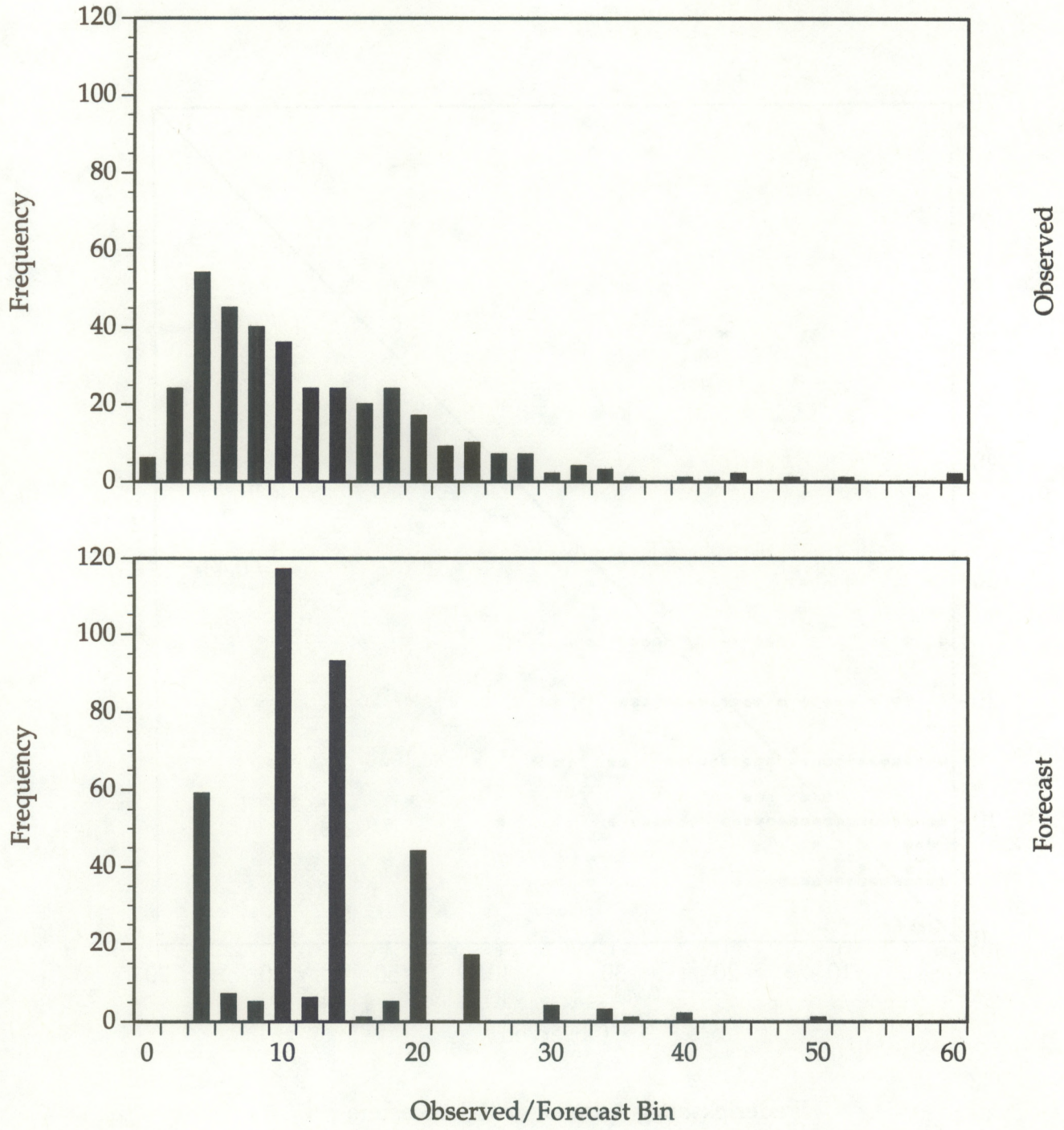



**Figure II-15.** Histograms of Fredericksburg A-Index observations and next-day forecasts, 1 January–31 December 1993.

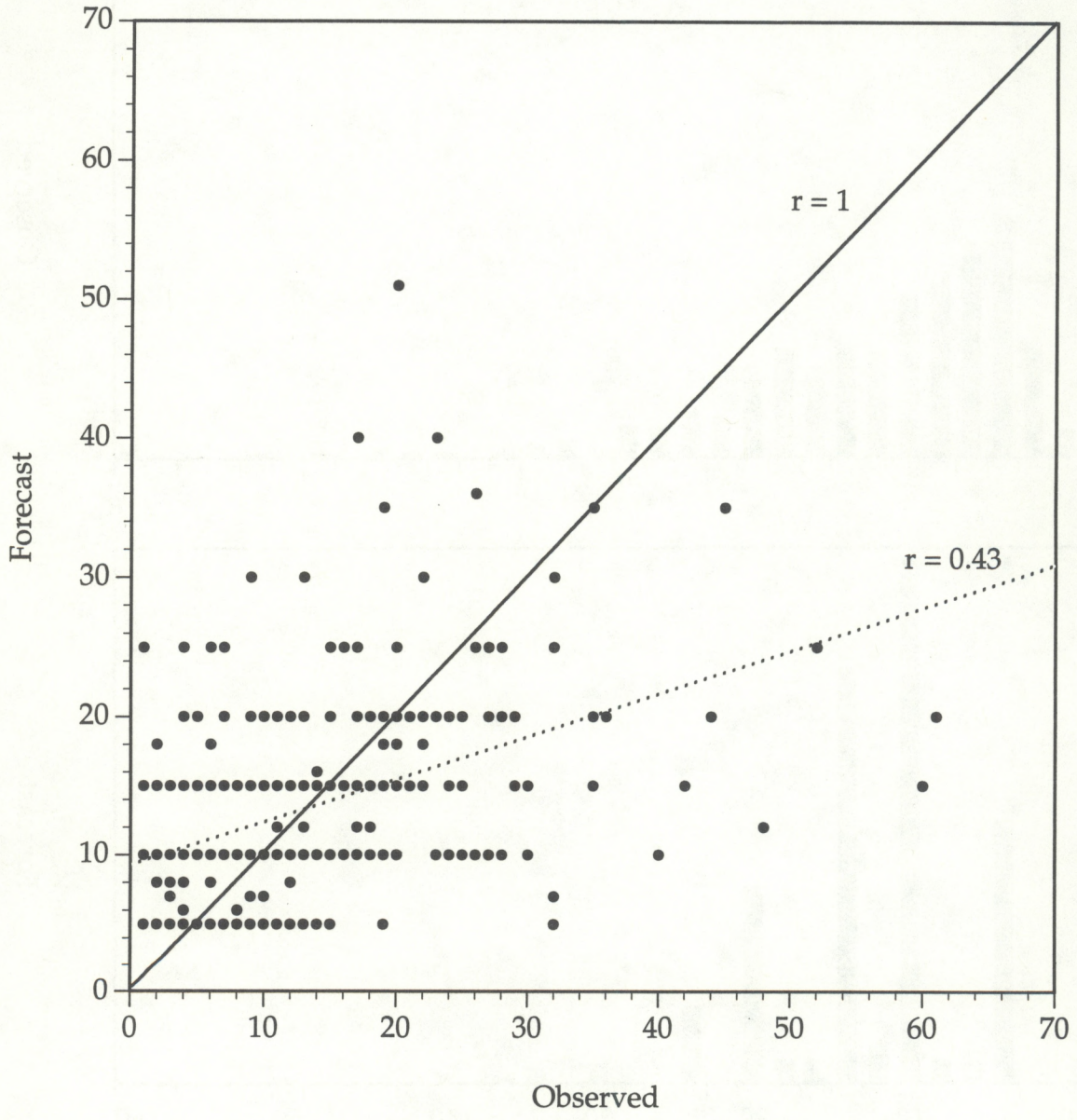
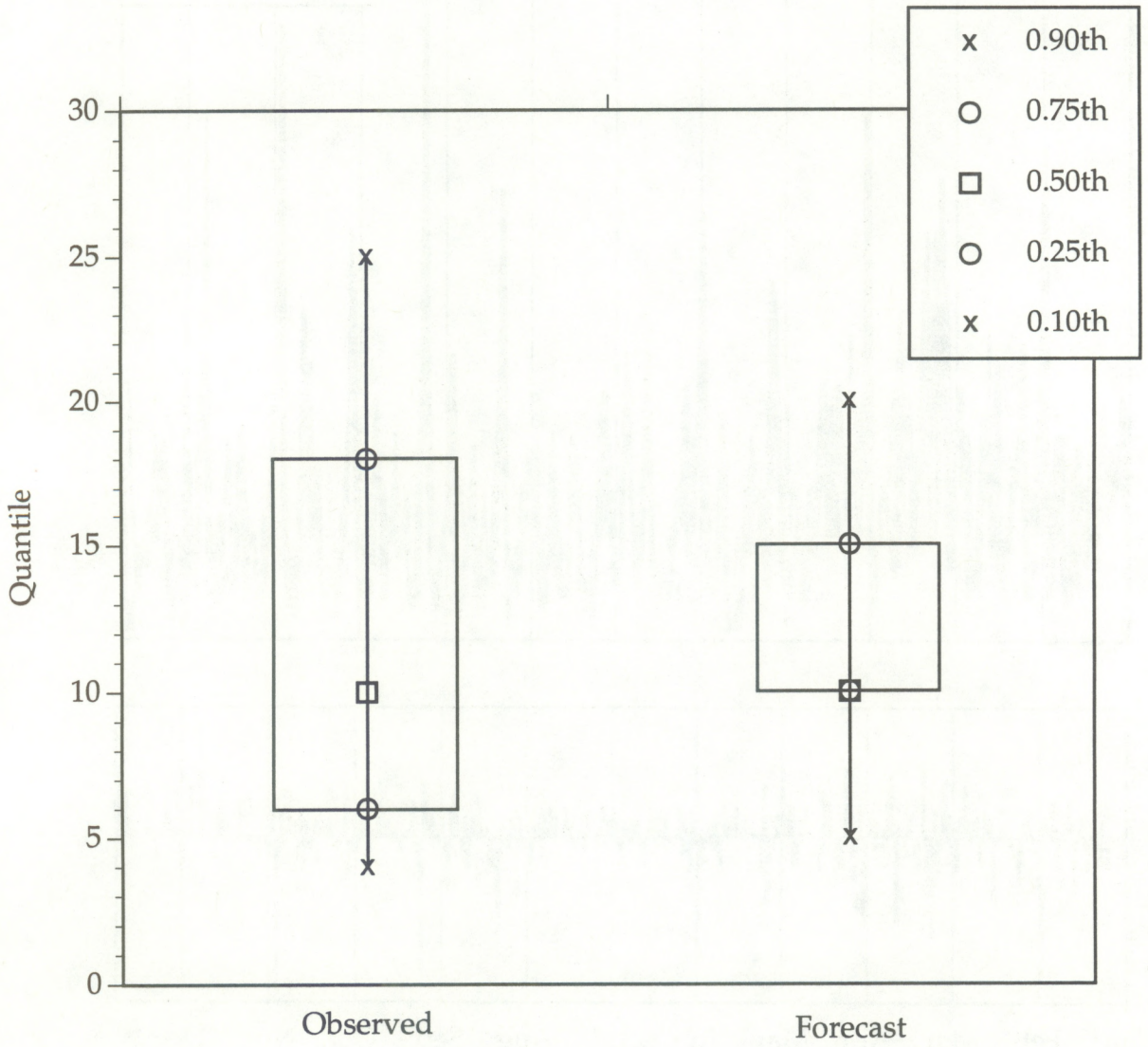



**Figure II-16.** Scatter diagram showing the joint distribution of next-day Fredericksburg A-Index forecasts and observations, 1 January–31 December 1993. The 45° diagonal (solid) indicates perfect correspondence between forecasts and observations. The dotted line is a least-squares fit (r=0.43) to the data.




Fredericksburg A-Index - All Forecasters

**Figure II-17.** Box plots showing the marginal distribution of Fredericksburg A-Index observations and next-day forecasts, 1 January–31 December 1993. The following quantiles are plotted (from top to bottom): 0.90th, 0.75th (upper quartile), 0.50th (median), 0.25th (lower quartile), and 0.10th.

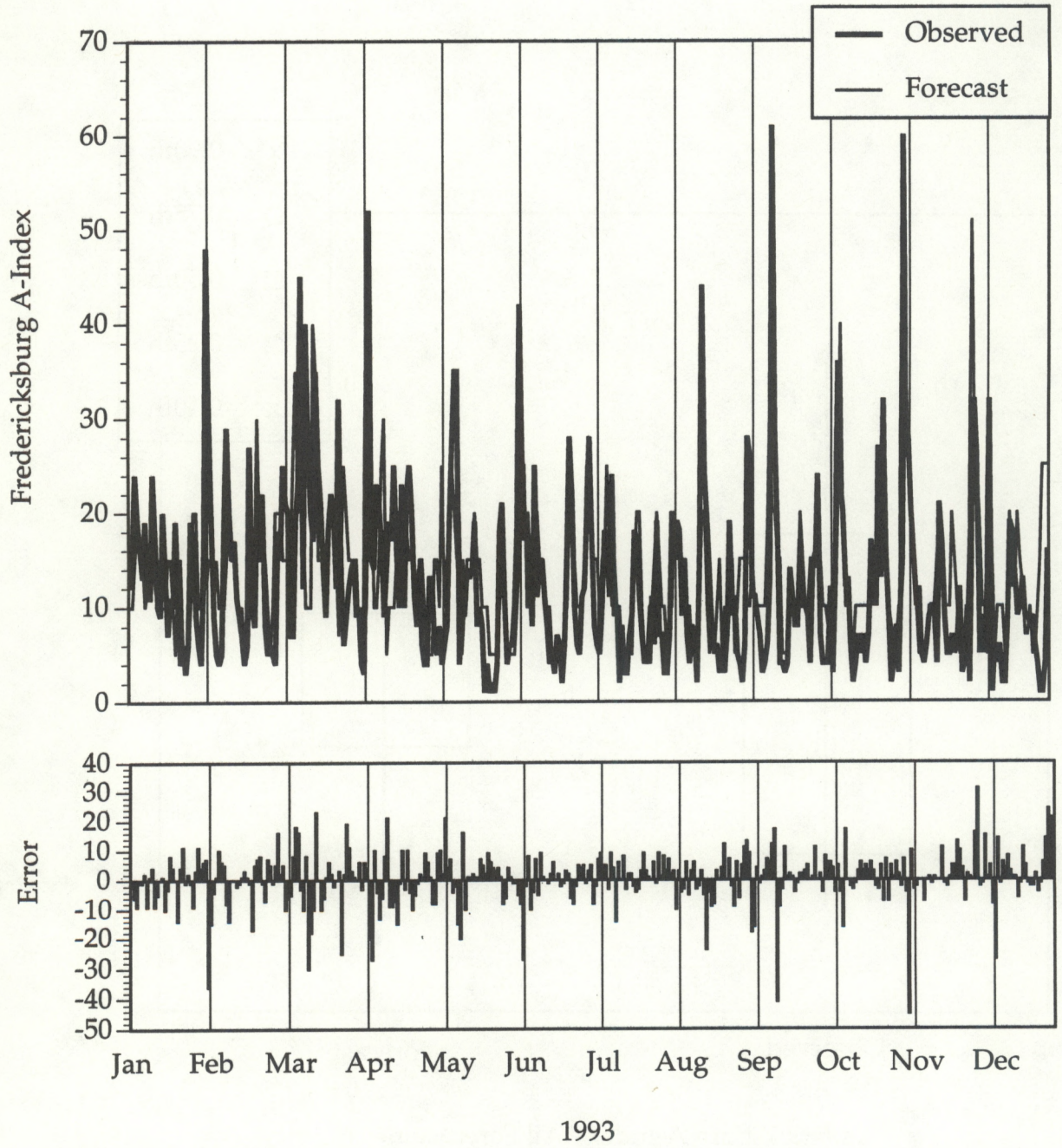



**Figure II-18.** Time history plot for next-day, Fredericksburg A-Index forecasts, 1 January–31 December 1993. Observed values are traced by the bold line. The errors (f-x) are shown in a separate plot at the bottom.

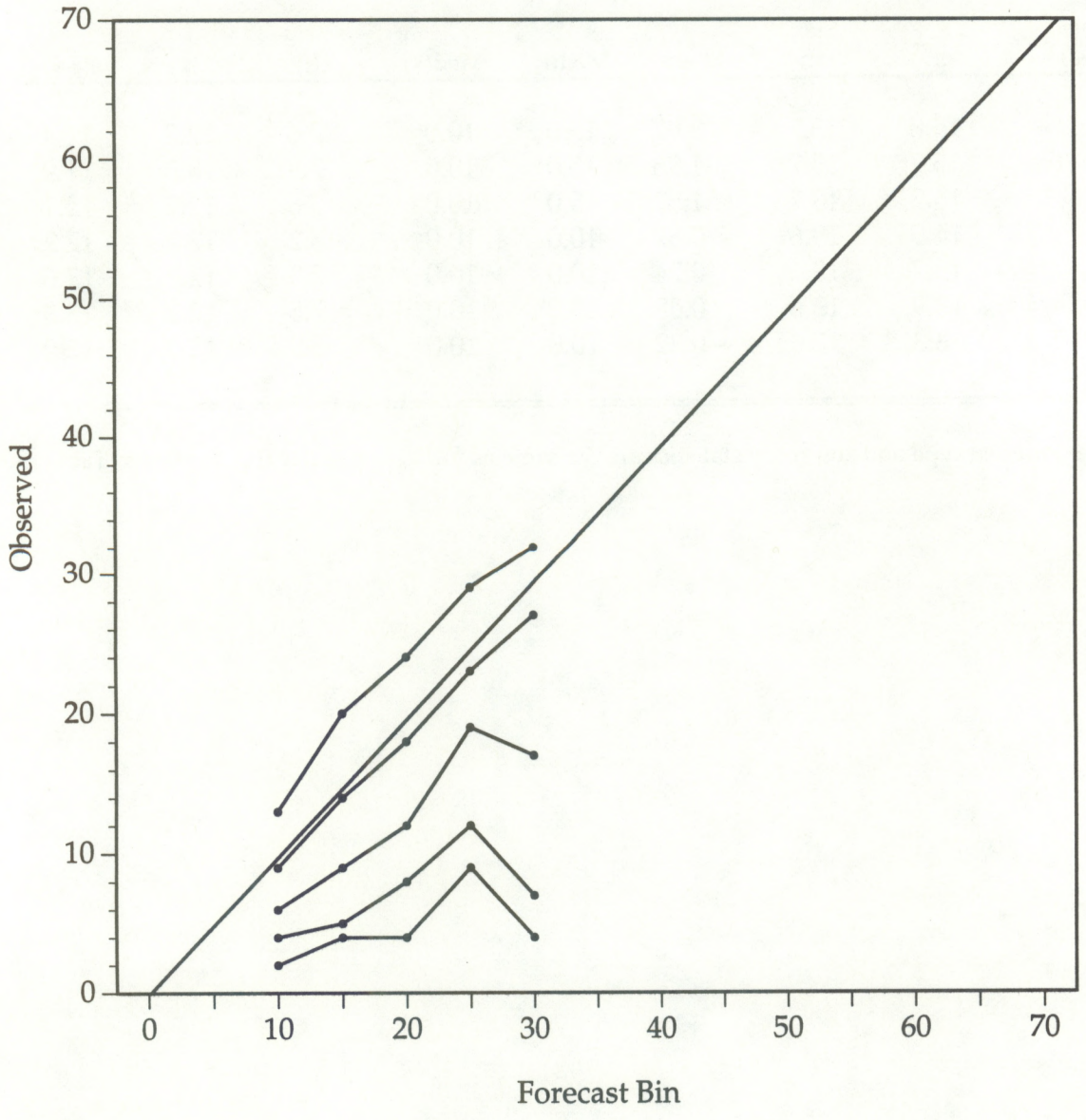



**Figure II-19.** Conditional quantile plot for next-day, Fredericksburg A-Index forecasts, 1 January–31 December 1993. The lines correspond to the following quantiles (top to bottom): 0.90th, 0.75th (upper quartile), 0.50th (median), 0.25th (lower quartile), and 0.10th. The 45° diagonal indicates perfect correspondence of forecasts and observations. The spread in the quantiles is related to the accuracy of the forecasts, and the deviation of the median from the diagonal is related to the overall bias.

**Table II-6.** Verification statistics for Planetary A-Index forecasts [1], 1 January–31 December 1993

| Lead Time (days) | n=365 <f> | <x> | <f–x> | Med(f) | Med(x) | s(f) | s(x) | s(f-x) |
|---|---|---|---|---|---|---|---|---|
| 1 | 15.8 | 13.7 | 2.07 | 15.0 | 10.0 | 7.6 | 12.2 | 11.4 |
| 2 | 15.7 | 13.7 | 1.93 | 15.0 | 10.0 | 7.9 | 12.3 | 11.9 |
| 3 | 15.2 | 13.7 | 1.52 | 15.0 | 10.0 | 7.6 | 12.2 | 12.1 |
| 4 | 14.5 | 13.6 | 0.82 | 10.0 | 10.0 | 7.2 | 12.2 | 12.2 |
| 5 | 13.7 | 13.6 | 0.24 | 10.0 | 10.0 | 7.3 | 12.2 | 12.6 |
| 6 | 13.9 | 13.6 | 0.08 | 10.0 | 10.0 | 7.5 | 12.2 | 12.3 |
| 7 | 13.3 | 13.6 | – 0.32 | 10.0 | 10.0 | 5.9 | 12.3 | 12.0 |

1. Forecast type and summary statistics are the same as for 10.7 cm solar flux forecasts (Table II-4).

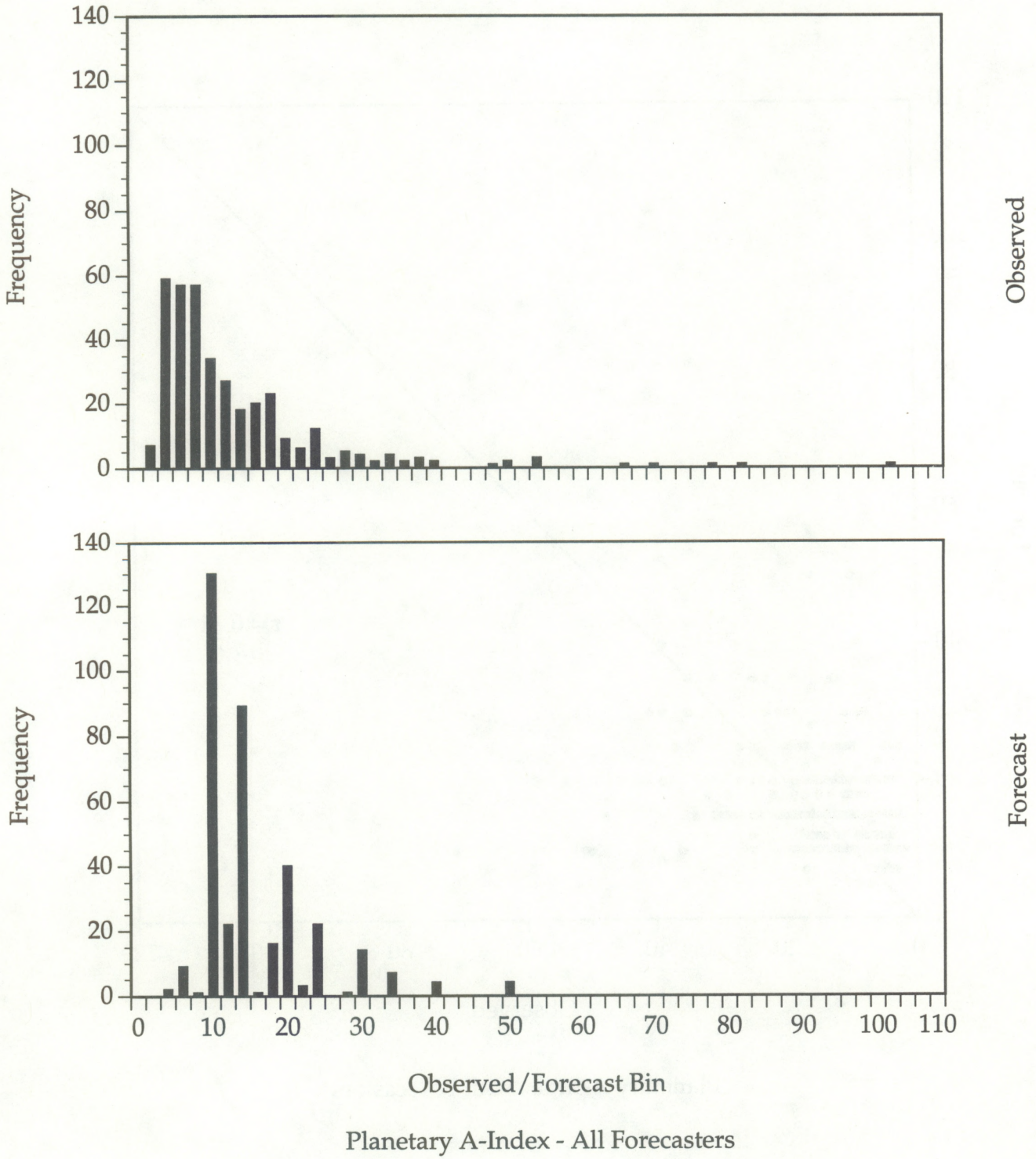
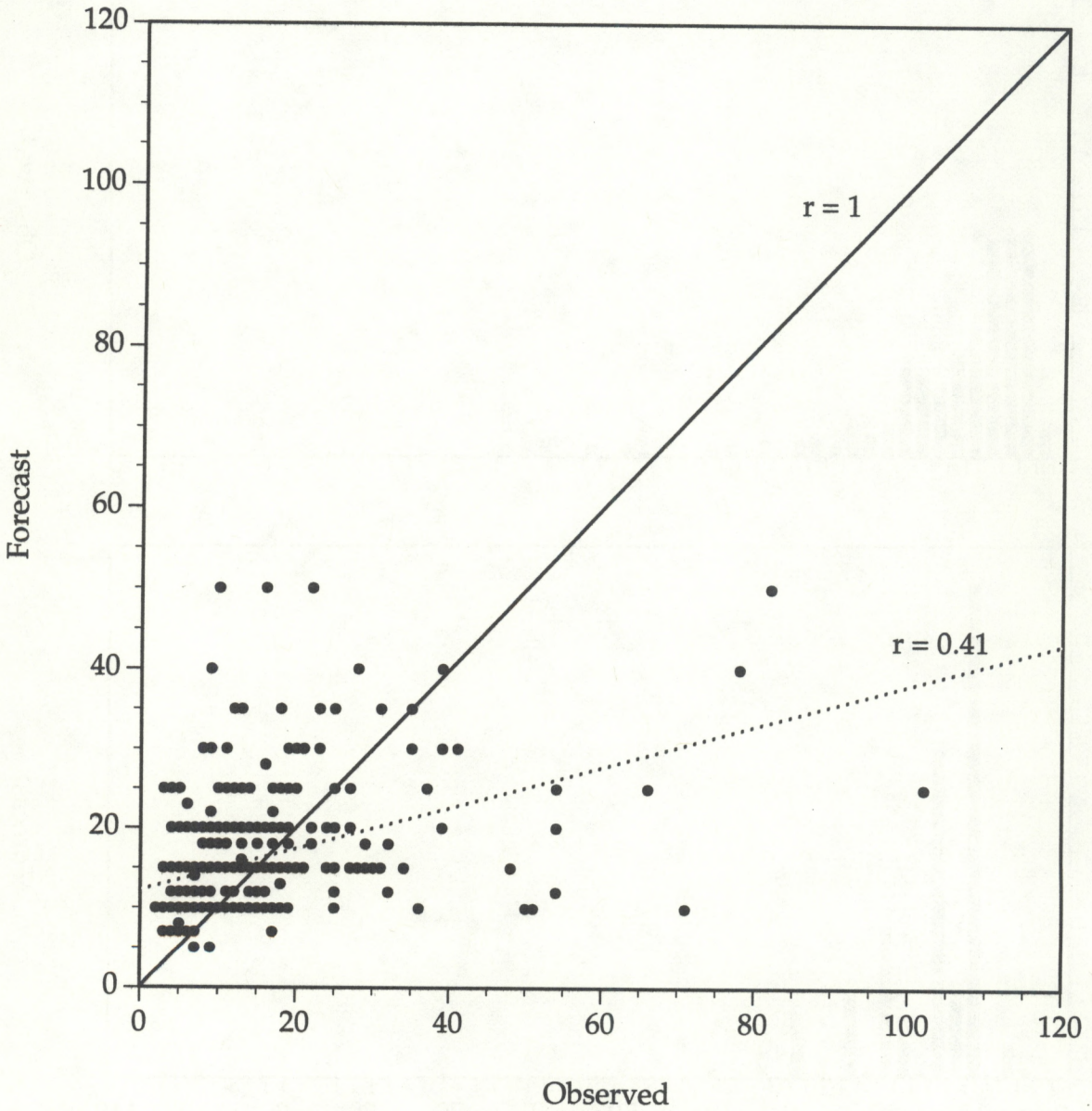



**Figure II-20.** Histograms of Planetary A-Index observations and next-day forecasts, 1 January–31 December 1993.




Planetary A-Index - All Forecasters

**Figure II-21.** Scatter diagram showing the joint distribution of next-day Planetary A-Index forecasts and observations, 1 January–31 December 1993. The 45° diagonal (solid) indicates perfect correspondence between forecasts and observations. The dotted line is a least-squares fit (r=0.41) to the data.

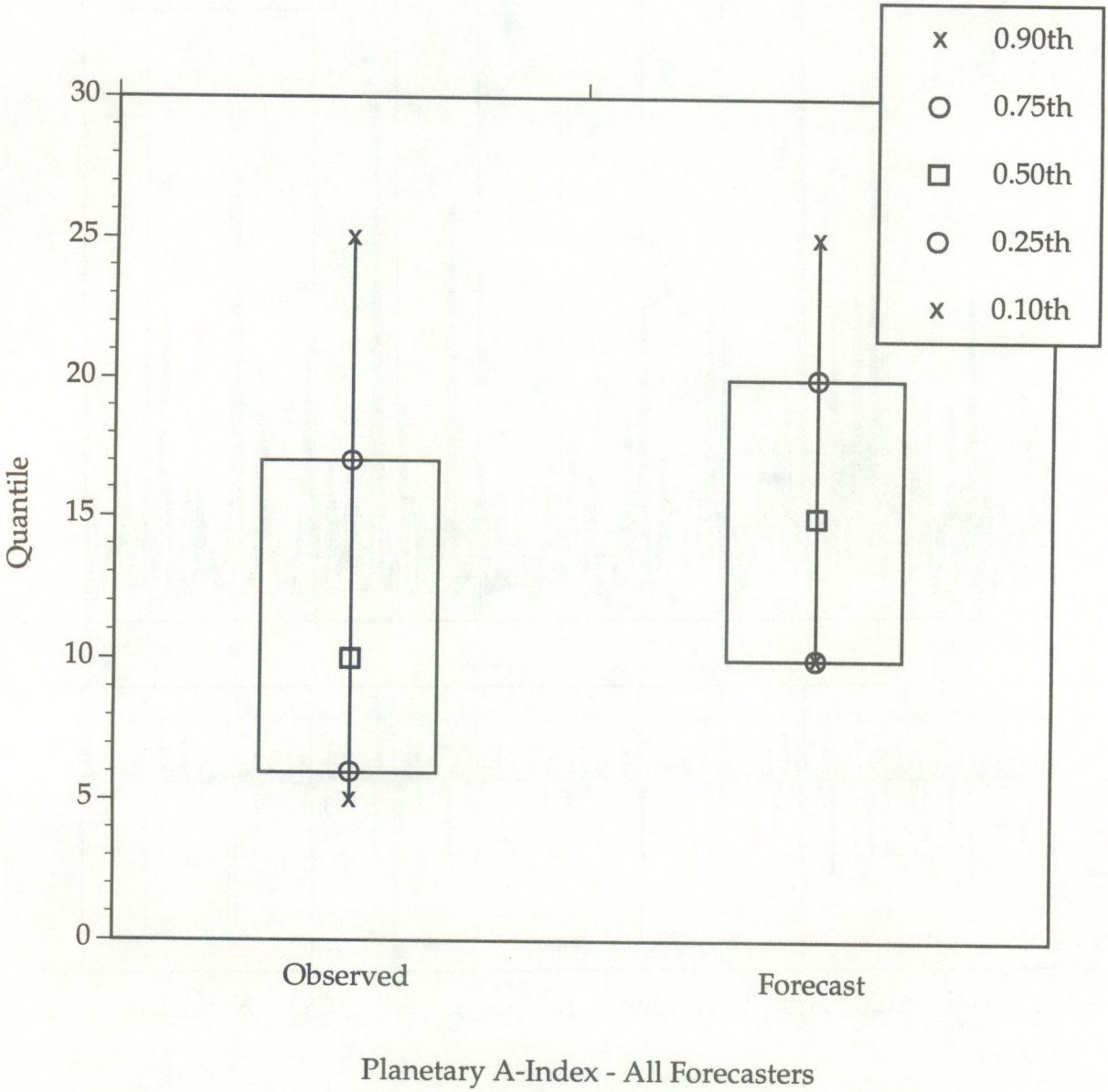



**Figure II-22.** Box plots showing the marginal distribution of Planetary A-Index observations and next-day forecasts, 1 January–31 December 1993. The following quantiles are plotted (from top to bottom): 0.90th, 0.75th (upper quartile), 0.50th (median), 0.25th (lower quartile), and 0.10th.

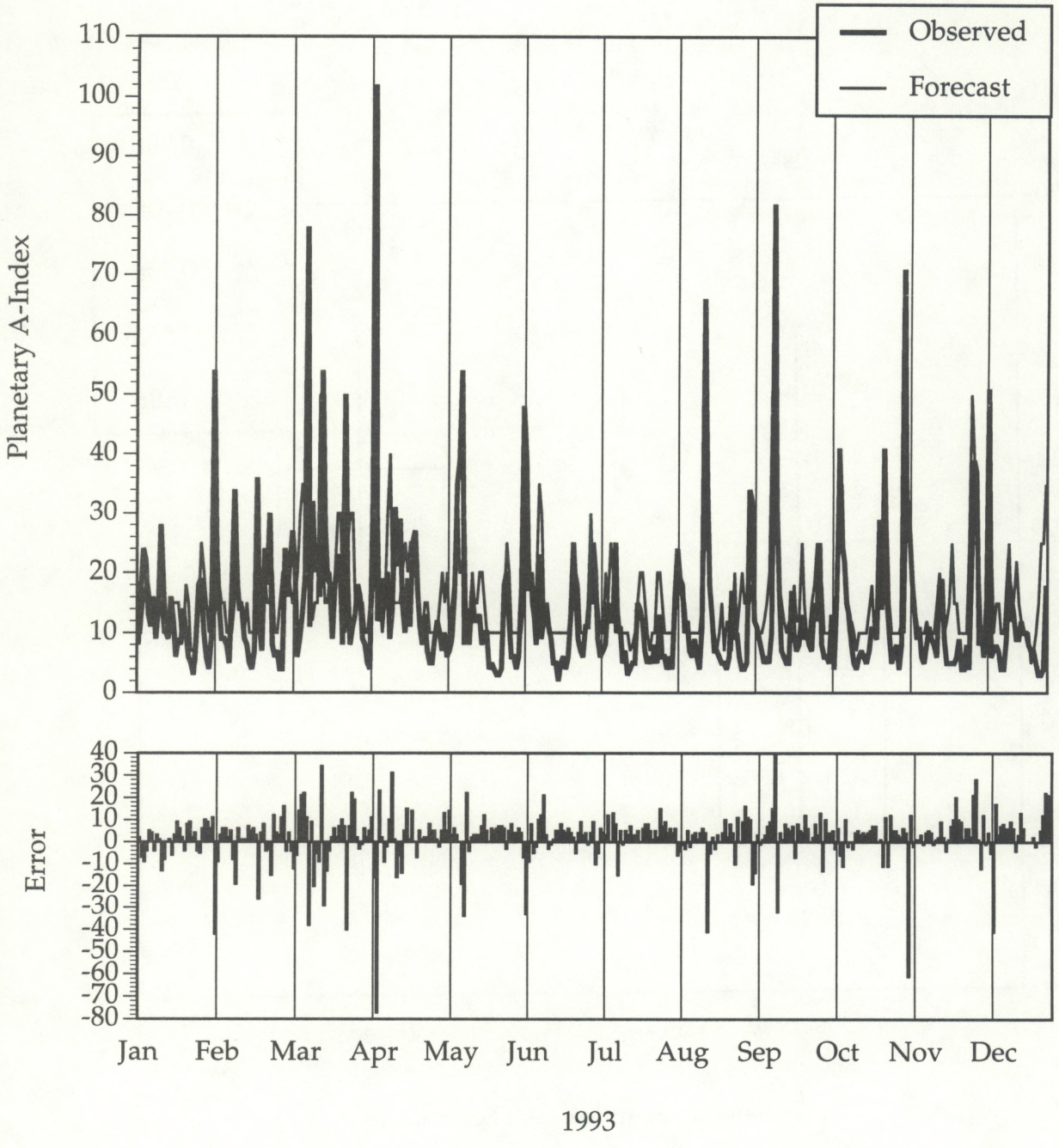



**Figure II-23.** Time history plot for next-day, Planetary A-Index forecasts, 1 January–31 December 1993. Observed values are traced by the bold line. The errors (f-x) are shown in a separate plot at the bottom.

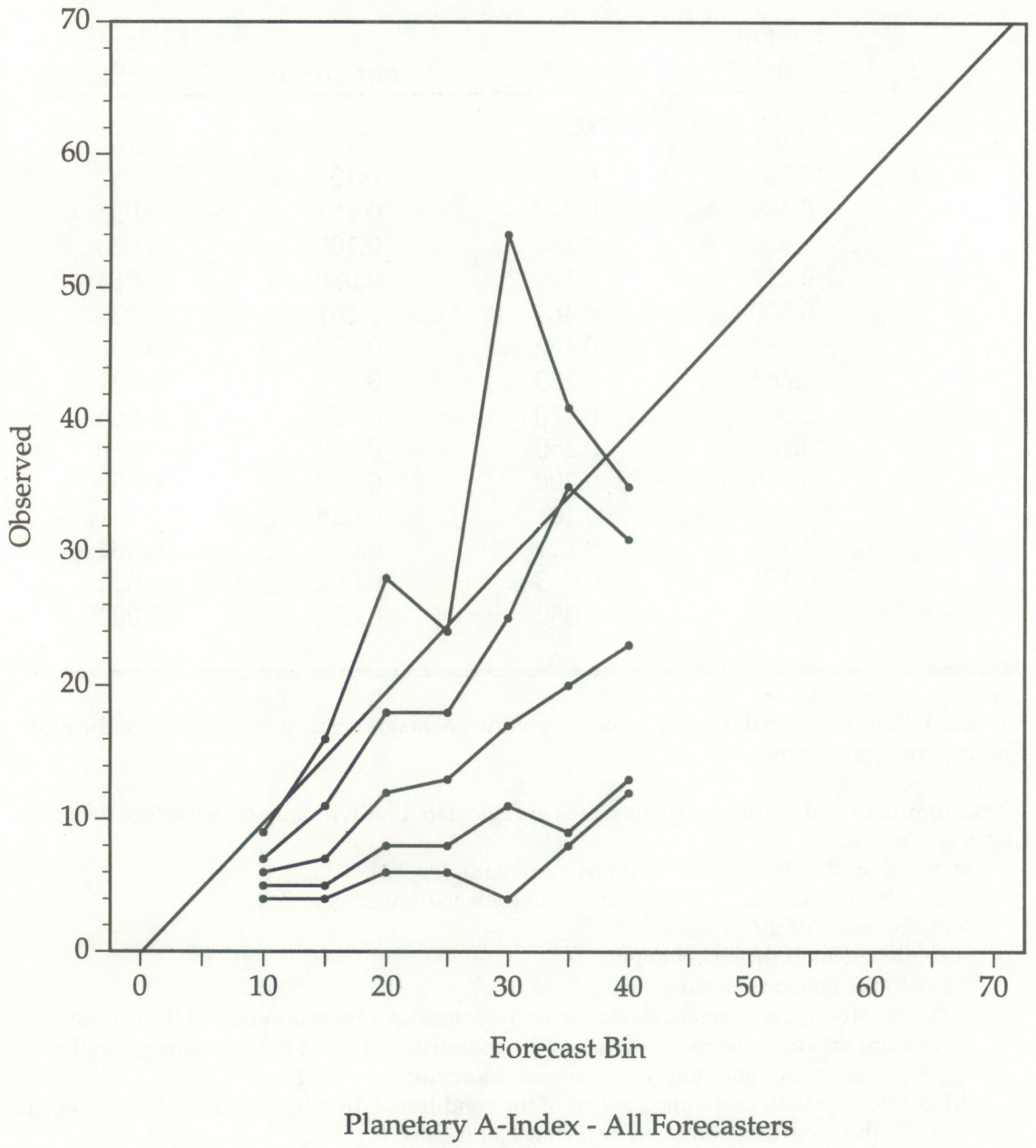



**Figure II-24.** Conditional quantile plot for next-day, Planetary A-Index forecasts, 1 January–31 December 1993. The lines correspond to the following quantiles (top to bottom): 0.90th, 0.75th (upper quartile), 0.50th (median), 0.25th (lower quartile), and 0.10th. The 45° diagonal indicates perfect correspondence of forecasts and observations. The spread in the quantiles is related to the accuracy of the forecasts, and the deviation of the median from the diagonal is related to the overall bias.

**Table II-7.** Verification statistics for next-day, middle-latitude, geomagnetic probability forecasts [1] [2], 1 January–31 December 1993

n = 365

| | Quiet to Unsettled | Active | Minor Storm | Major to Severe Storm |
|---|---|---|---|---|
| # Events | 175 | 114 | 56 | 20 |
| <f> | 0.582 | 0.230 | 0.131 | 0.057 |
| <x> | 0.479 | 0.312 | 0.153 | 0.055 |
| Med(f) | 0.600 | 0.250 | 0.100 | 0.050 |
| s (f) | 0.212 | 0.109 | 0.088 | 0.064 |
| s (x) | 0.500 | 0.464 | 0.361 | 0.228 |
| s(f–x) | 0.467 | 0.462 | 0.353 | 0.230 |
| <(f \| x=1)> | 0.662 | 0.253 | 0.175 | 0.087 |
| <(f \| x=0)> | 0.508 | 0.220 | 0.123 | 0.055 |
| Med(f \| x=1) | 0.700 | 0.250 | 0.150 | 0.050 |
| Med(f \| x=0) | 0.500 | 0.200 | 0.100 | 0.050 |
| s(f \| x=1) | 0.196 | 0.101 | 0.089 | 0.091 |
| s(f \| x=0) | 0.198 | 0.111 | 0.085 | 0.062 |
| DISC | 0.154 | 0.033 | 0.052 | 0.032 |
| ME | 0.103 | – 0.082 | – 0.022 | 0.002 |

1. These forecasts are probability-type forecasts of the geomagnetic activity category of the most-disturbed 3-hour period during the next day.

2. These summary and performance measures are calculated individually for each forecast activity category and include

n, number of forecasts/observations in the sample;
# Events, the number of days observed in each activity category;
<f>, the mean of the forecast;
<x>, the mean of the observations;
Med(f), the median forecast;
s(f), s(x), s(f–x), the standard deviation of the forecasts, observations, and the errors;
<f | x=1)>, <f | x=0>, the mean of the conditional distribution of the forecasts given the occurrence and non-occurrence of the event;
Med(f | x=1), Med(f | x=0), the median of the conditional distribution of the forecasts given the occurrence and non-occurrence of the event;
s(f | x=1), s(f | x=0), the standard deviation of the conditional distribution of the forecasts given the occurrence and non-occurrence of the event;
DISC, the discrimination of the forecasts at each activity category (positive values are better); and
ME, the mean errors at each activity category.

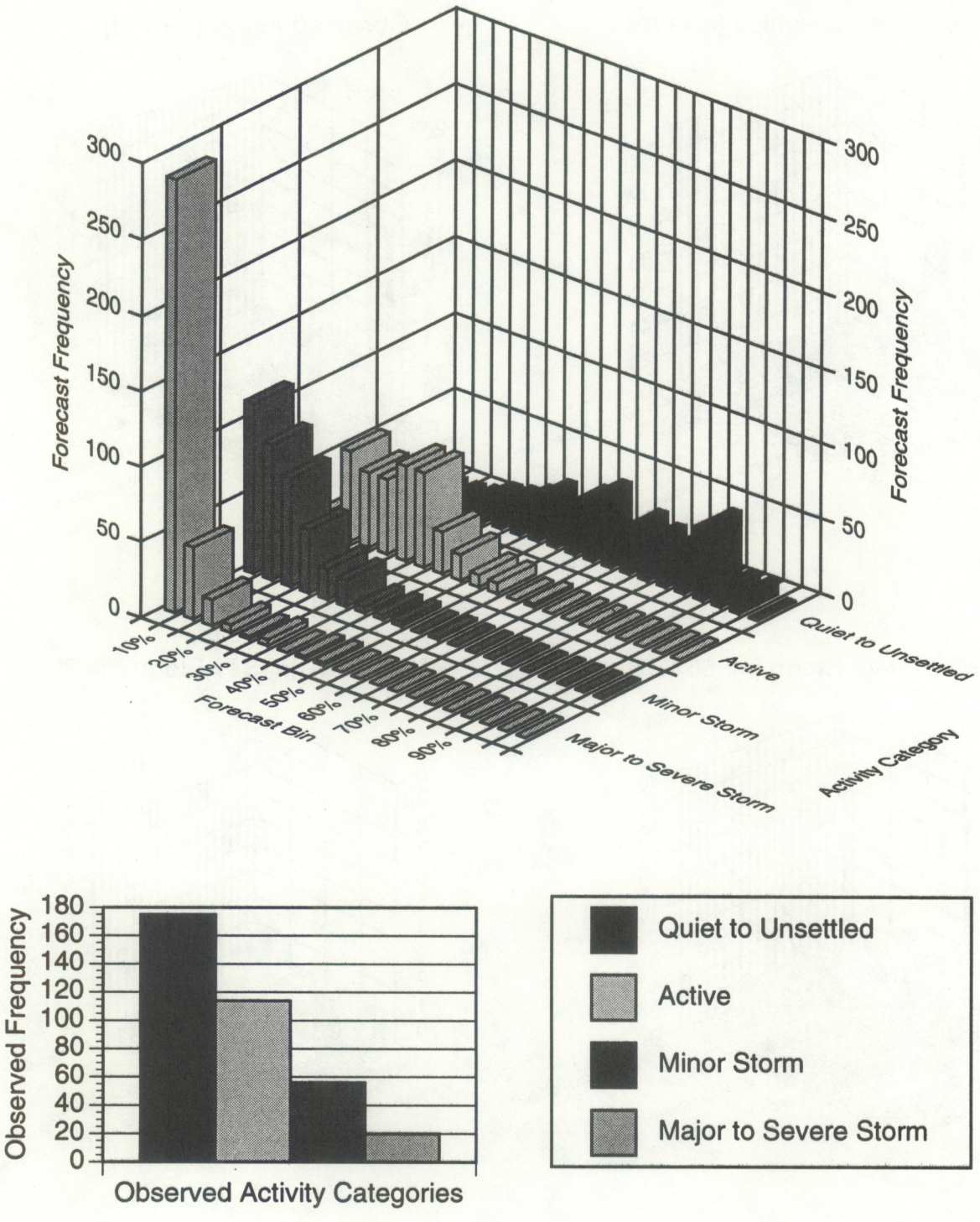



**Figure II-25.** Histograms of forecasts and observed frequency for next-day, middle-latitude geomagnetic probability forecasts. There are 365 forecasts in each category.

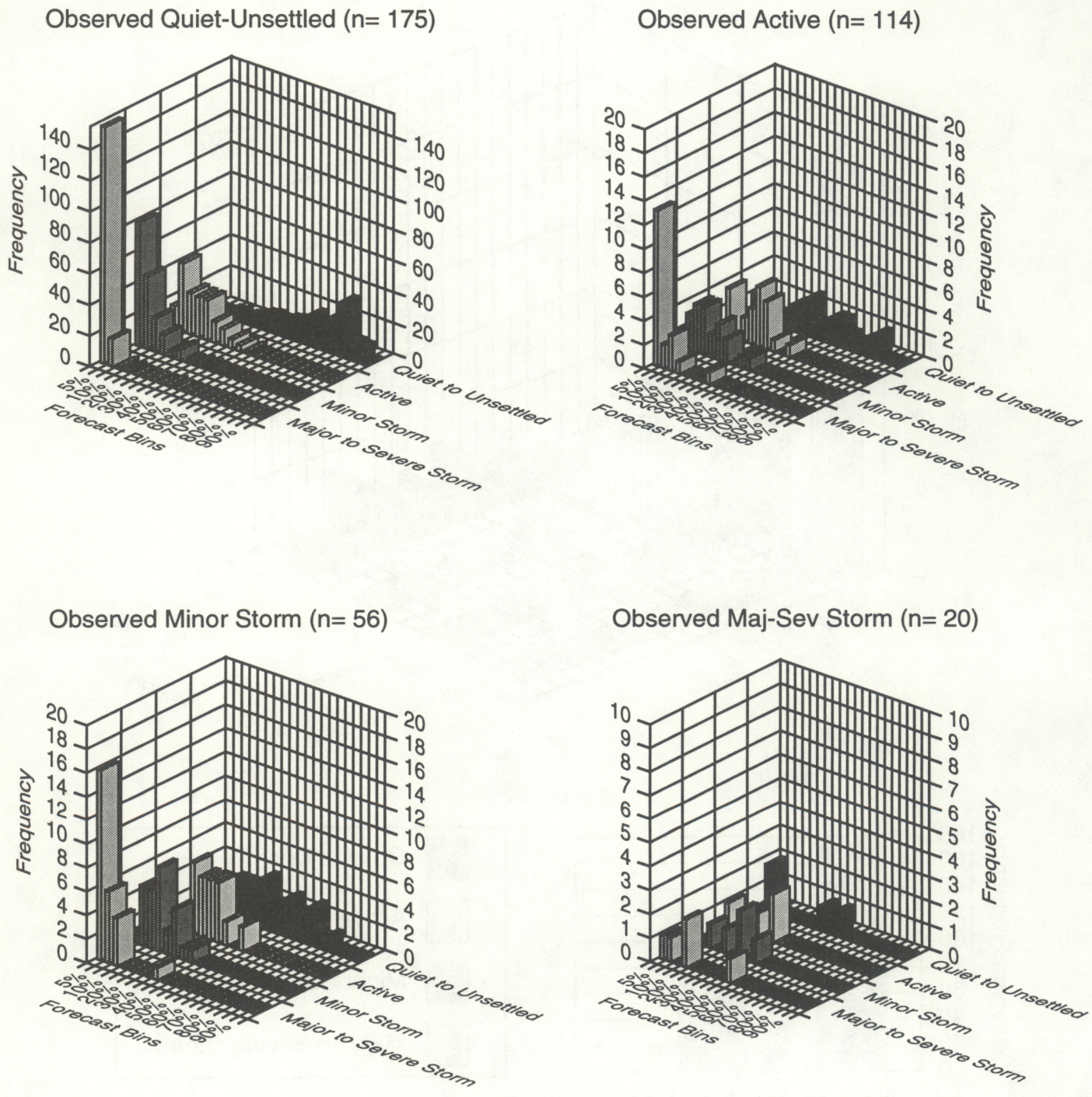



**Figure II-26.** Histograms of forecast frequency, using the same data as in figure II-25 but stratified by observed category. This figure shows the conditional distributions of the forecasts when each activity category occurred.

# III. INDIVIDUAL FORECASTER VERIFICATION

This section provides information on the performance of individual SEL forecasters with respect to the forecaster group as a whole. Forecaster names are replaced by a letter designation to preserve anonymity.

Key verification measures are tabulated in a single table for each of five forecast types. Each table is followed by trend diagrams that track individual forecaster performance with respect to forecast bias, accuracy, and linear correlation for each year during the last 5 years (1989–1993). Since the statistical sample is small, interpretation of individual forecaster statistics should be done with care. The trend diagrams demonstrate that performance with respect to a particular parameter can fluctuate radically, and at least some of these changes may be due to the small sample size.

**Table III-1.** Verification statistics for next-day M-class flare forecasts [1] [2], 1 January–31 December 1993.

| Forecaster | n | <f> | <x> | Med(f) | Med(f-x) | s(f) | s(x) |
|---|---|---|---|---|---|---|---|
| SEL (all) | 365 | 0.191 | 0.153 | 0.100 | 0.050 | 0.202 | 0.360 |
| A | 46 | 0.225 | 0.196 | 0.050 | 0.050 | 0.260 | 0.397 |
| B | 69 | 0.180 | 0.174 | 0.200 | 0.100 | 0.132 | 0.379 |
| C | 41 | 0.148 | 0.122 | 0.050 | 0.050 | 0.129 | 0.327 |
| D | 43 | 0.188 | 0.116 | 0.100 | 0.050 | 0.218 | 0.321 |
| E | 61 | 0.178 | 0.197 | 0.100 | 0.050 | 0.174 | 0.398 |

| Forecaster | s(f-x) | <f \| x=1> – <f \| x=0> | Med (f \| x=1) – Med (f \| x=0) | $r_{fx}$ | ME | MSE | $SS_{SC}$ |
|---|---|---|---|---|---|---|---|
| SEL (all) | 0.350 | 0.186 | 0.200 | 0.331 | 0.038 | 0.124 | 0.043 |
| A | 0.422 | 0.149 | 0.250 | 0.227 | 0.029 | 0.179 | – 0.136 |
| B | 0.364 | 0.100 | 0.150 | 0.286 | 0.006 | 0.133 | 0.074 |
| C | 0.322 | 0.093 | 0.250 | 0.237 | 0.026 | 0.104 | 0.027 |
| D | 0.325 | 0.217 | 0.350 | 0.320 | 0.072 | 0.111 | – 0.077 |
| E | 0.357 | 0.193 | 0.200 | 0.442 | – 0.019 | 0.127 | 0.198 |

1. M-class flare forecasts are probabilistic forecasts ranging from 0.01 to 0.99. Corresponding observations are either 1 (M-flare occurred) or 0 (no M-flare occurred).

2. These summary and performance measures include
   n, number of forecasts/observations in the sample;
   <f>, mean forecast;
   <x>, mean observation;
   <f-x>, mean error;
   Med(f), median forecast;
   Med(f-x), median error;
   s(f), s(x), and s(f-x), standard deviations of the forecasts, observations, and errors;
   <f | x=1> – <f | x=0>, the difference of the mean conditional distributions of the forecasts (This quantity is related to the amount of discrimination in the forecasts);
   Med(f | x=1) - Med(f | x=0), the difference of the medians of the conditional distributions;
   $r_{fx}$, linear-correlation coefficient;
   ME, mean error (overall bias);
   MSE, mean square error (accuracy); and
   $SS_{SC}$, skill score with respect to the sample climatology.

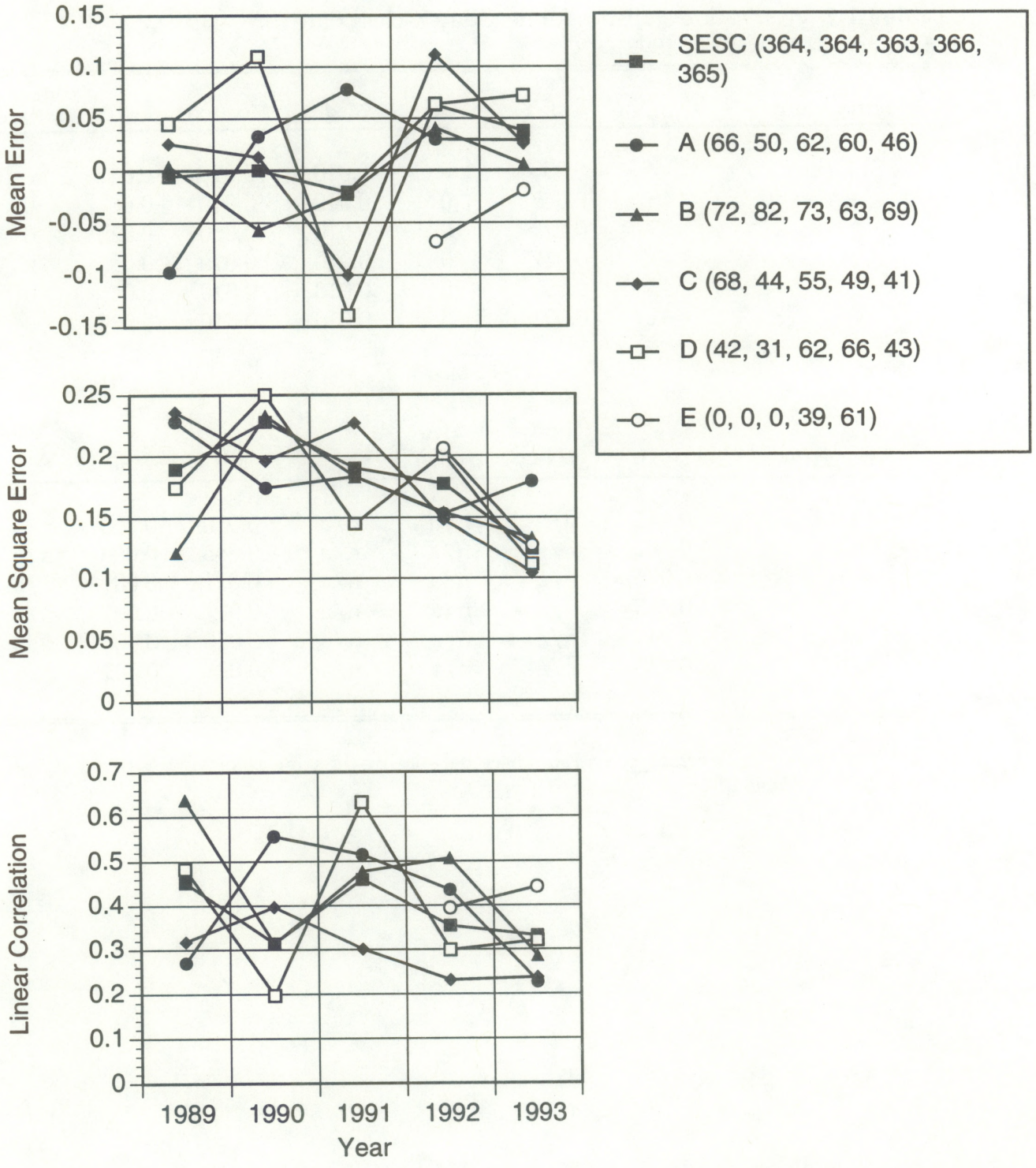



**Figure III-1.** Trend diagrams for M-class flare forecasts. Yearly values of the overall bias (mean error), accuracy (mean square error), and linear correlation are plotted for each forecaster. The number of forecasts/observations for each year (1989, 1990, 1991, 1992, 1993) is shown in the figure legend beside the forecaster's designation.

**Table III-2.** Verification statistics for next-day X-class flare forecasts [1], 1 January–31 December 1993

| Forecaster | n | <f> | <x> | Med(f) | Med(f-x) | s(f) | s(x) |
|---|---|---|---|---|---|---|---|
| SEL (all) | 365 | 0.030 | 0 | 0.010 | 0.010 | 0.042 | 0 |
| A | 46 | 0.045 | 0 | 0.010 | 0.010 | 0.061 | 0 |
| B | 69 | 0.021 | 0 | 0.010 | 0.010 | 0.025 | 0 |
| C | 41 | 0.017 | 0 | 0.010 | 0.010 | 0.015 | 0 |
| D | 43 | 0.030 | 0 | 0.010 | 0.010 | 0.043 | 0 |
| E | 61 | 0.031 | 0 | 0.010 | 0.010 | 0.033 | 0 |

| Forecaster | s(f-x) | <f \| x=1> – <f \| x=0> | Med (f \| x=1) – Med (f \| x=0) | $r_{fx}$ | ME | MSE | $SS_{SC}$ |
|---|---|---|---|---|---|---|---|
| SEL (all) | 0.042 | n/a | n/a | n/a | 0.030 | 0.003 | n/a |
| A | 0.061 | n/a | n/a | n/a | 0.045 | 0.006 | n/a |
| B | 0.025 | n/a | n/a | n/a | 0.021 | 0.001 | n/a |
| C | 0.015 | n/a | n/a | n/a | 0.017 | 0.001 | n/a |
| D | 0.043 | n/a | n/a | n/a | 0.030 | 0.003 | n/a |
| E | 0.033 | n/a | n/a | n/a | 0.031 | 0.002 | n/a |

1. X-class flare forecast type and summary statistics are the same as for M-class flare forecasts (Table III-1).

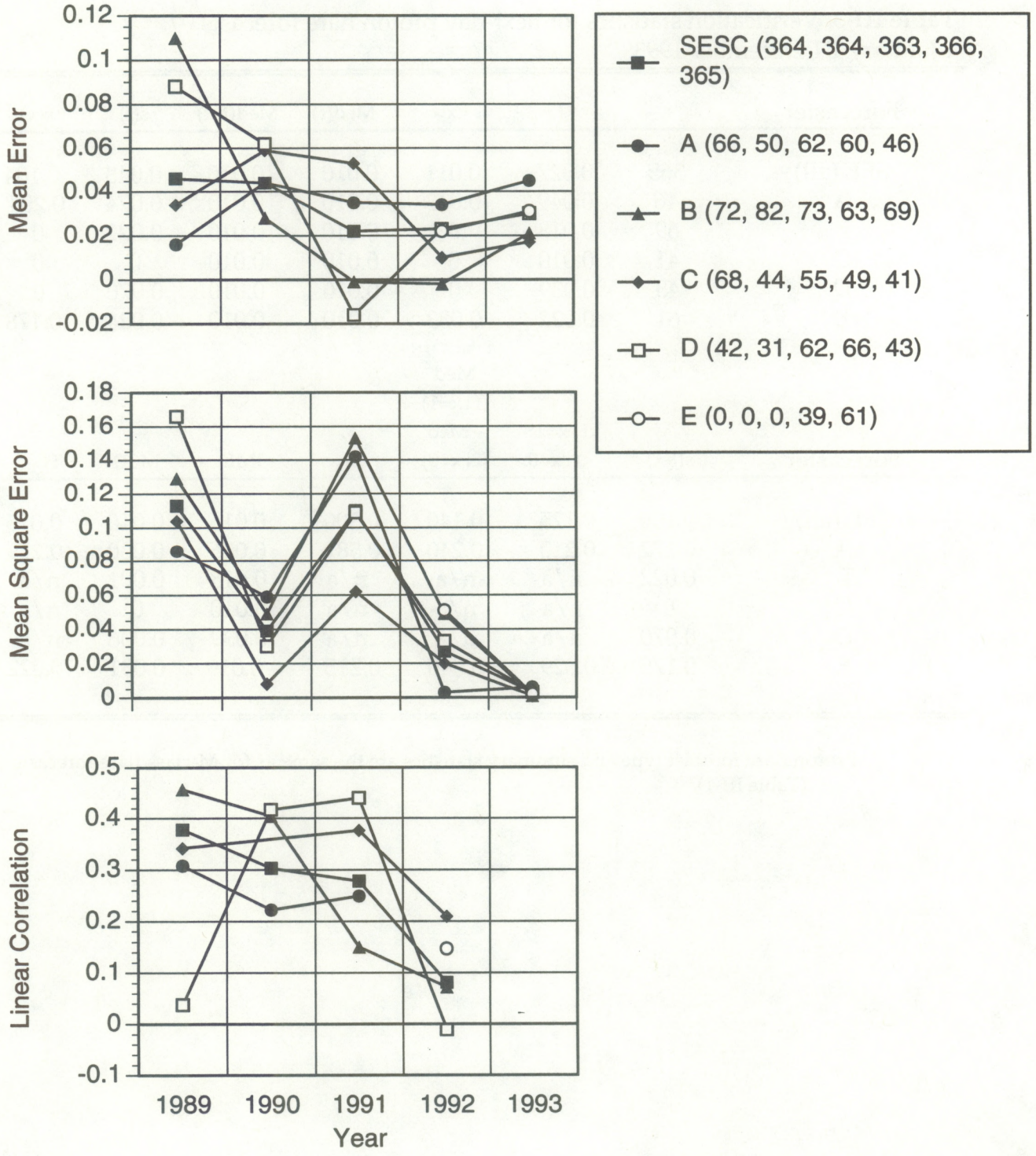



**Figure III-2.** Trend diagrams for X-class flare forecasts. The parameters and legend are similar to those in Figure III-1.

**Table III-3.** Verification statistics for next-day proton flare forecasts [1], 1 January–31 December 1993

| Forecaster | n | <f> | <x> | Med(f) | Med(f-x) | s(f) | s(x) |
|---|---|---|---|---|---|---|---|
| SEL (all) | 365 | 0.027 | 0.011 | 0.010 | 0.010 | 0.043 | 0.104 |
| A | 46 | 0.049 | 0.043 | 0.010 | 0.010 | 0.074 | 0.204 |
| B | 69 | 0.018 | 0 | 0.010 | 0.010 | 0.022 | 0 |
| C | 41 | 0.010 | 0 | 0.010 | 0.010 | 0 | 0 |
| D | 43 | 0.039 | 0 | 0.010 | 0.010 | 0.070 | 0 |
| E | 61 | 0.022 | 0.033 | 0.010 | 0.010 | 0.024 | 0.178 |

| Forecaster | s(f-x) | <f \| x=1> – <f \| x=0> | Med (f \| x=1) – Med (f \| x=0) | $r_{fx}$ | ME | MSE | $SS_{SC}$ |
|---|---|---|---|---|---|---|---|
| SEL (all) | 0.100 | 0.125 | 0.140 | 0.300 | 0.016 | 0.010 | 0.075 |
| A | 0.172 | 0.210 | 0.240 | 0.582 | 0.006 | 0.030 | 0.279 |
| B | 0.022 | n/a | n/a | n/a | 0.018 | 0.001 | n/a |
| C | 0 | n/a | n/a | n/a | 0.010 | 0 | n/a |
| D | 0.070 | n/a | n/a | n/a | 0.039 | 0.006 | n/a |
| E | 0.175 | 0.029 | 0.040 | 0.215 | – 0.010 | 0.031 | 0.022 |

1. Proton flare forecast type and summary statistics are the same as for M-class flare forecasts (Table III-1).

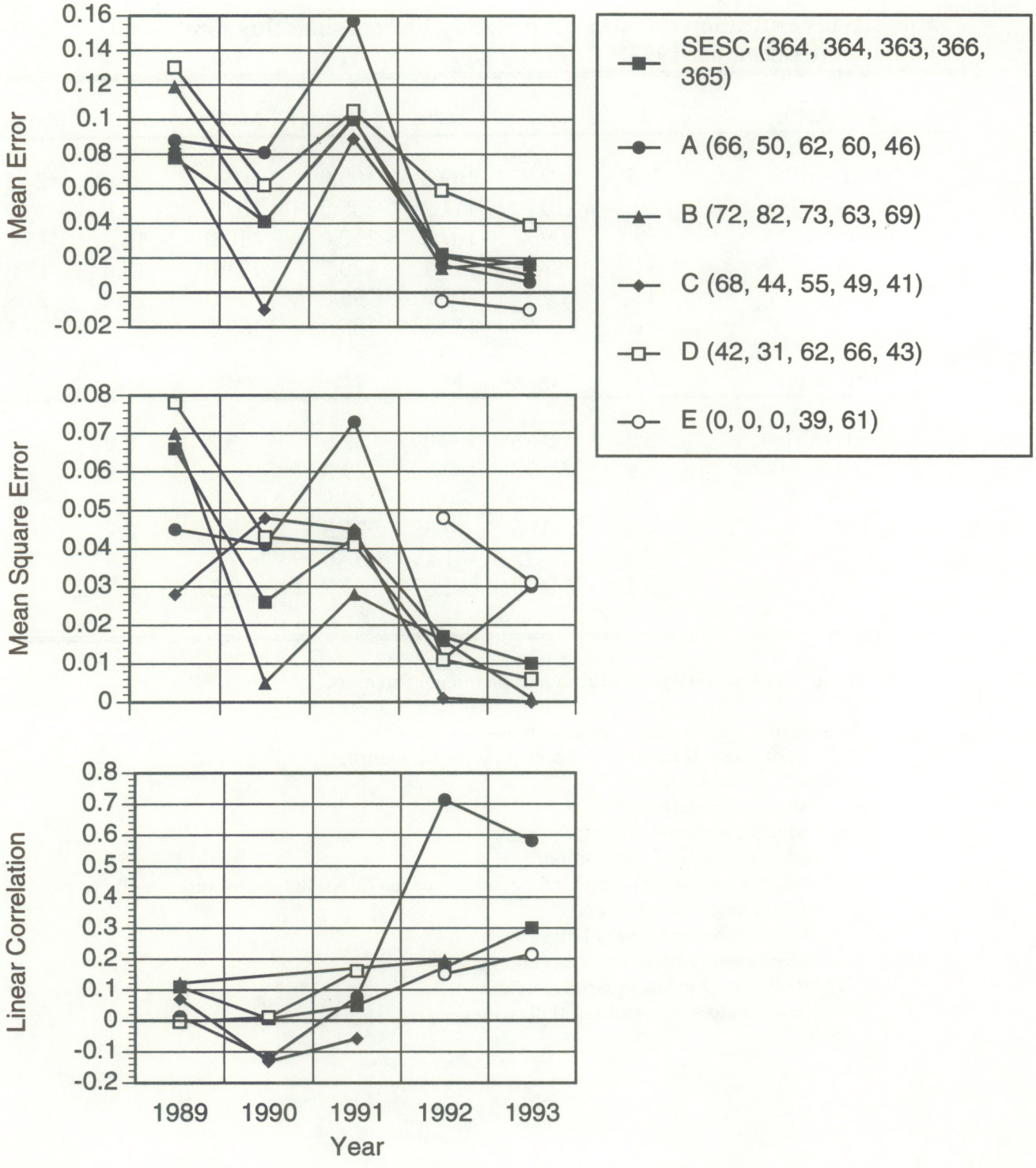



**Figure III-3.** Trend diagrams for proton flare forecasts. The parameters and legend are similar to those in Figure III-1.

**Table III-4.** Verification statistics for next-day 10.7 cm solar flux forecasts (1) (2), 1 January–31 December 1993

| Forecaster | n | <f> | <x> | Med(f) | Med(x) | s(f) | s(x) |
|---|---|---|---|---|---|---|---|
| SEL (all) | 365 | 109.9 | 109.8 | 105.0 | 105.0 | 21.6 | 21.3 |
| A | 46 | 110.7 | 111.2 | 105.0 | 107.5 | 20.8 | 20.9 |
| B | 69 | 109.7 | 109.8 | 110.0 | 110.0 | 16.7 | 15.4 |
| C | 41 | 117.2 | 116.4 | 120.0 | 121.0 | 15.6 | 15.8 |
| D | 43 | 104.1 | 105.6 | 95.0 | 96.0 | 22.8 | 21.8 |
| E | 61 | 114.5 | 113.5 | 105.0 | 103.0 | 25.7 | 26.5 |

| Forecaster | s(f-x) | $r_{fx}$ | ME | MSE | RMSE | $SS_{sc}$ |
|---|---|---|---|---|---|---|
| SEL (all) | 5.7 | 0.96 | 0.09 | 32.5 | 5.7 | 0.93 |
| A | 5.4 | 0.97 | −0.50 | 29.3 | 5.4 | 0.93 |
| B | 5.3 | 0.95 | −0.10 | 27.8 | 5.3 | 0.88 |
| C | 4.6 | 0.96 | 0.85 | 21.8 | 4.7 | 0.91 |
| D | 7.4 | 0.95 | −1.49 | 57.0 | 7.5 | 0.88 |
| E | 5.4 | 0.98 | 1.02 | 30.7 | 5.5 | 0.96 |

1. Solar flux forecasts are continuous-type integer forecasts.

2. These summary and performance measures include
   n, number of forecasts/observations in the sample;
   <f>, mean forecast;
   <x>, mean observation;
   Med(f), median forecast;
   Med(x), median observation;
   s(f), s(x), and s(f-x), standard deviations of the forecasts, observations, and errors;
   $r_{fx}$, linear-correlation;
   ME, mean error (overall bias);
   MSE, mean square error (accuracy);
   RMSE, root mean square error; and
   $SS_{sc}$, sample-climatology skill score.

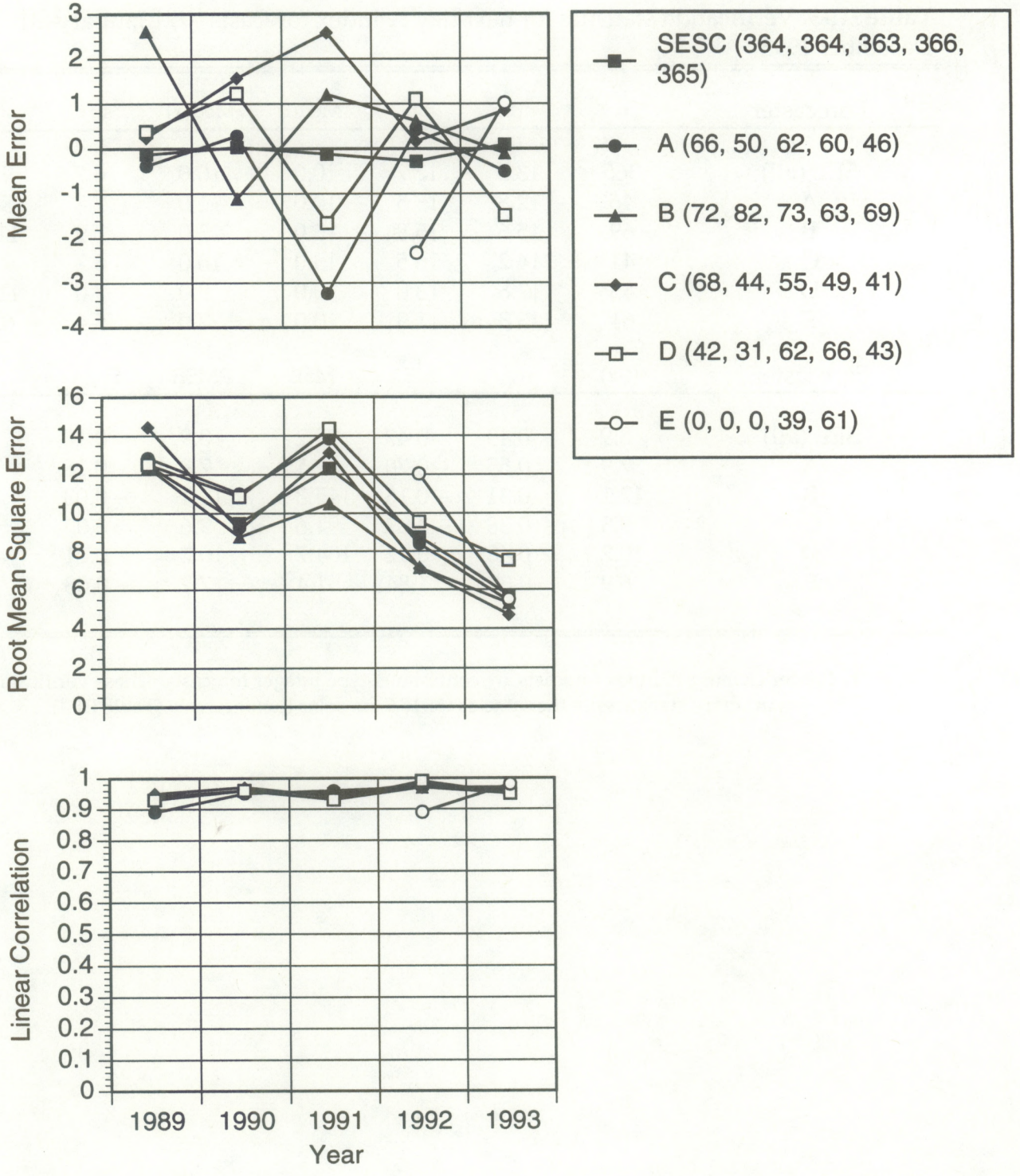



**Figure III-4.** Trend diagrams for 10.7 cm solar flux forecasts. Yearly values of the overall bias (mean error), accuracy (root mean square error), and linear correlation are plotted for each forecaster. The number of forecasts for each year (1989, 1990, 1991, 1992, 1993) is shown in the figure legend beside the forecaster's designation.

**Table III-5.** Verification statistics for next-day A-Index forecasts [1], 1 January–31 December 1993

| Forecaster | n | <f> | <x> | Med(f) | Med(x) | s(f) | s(x) |
|---|---|---|---|---|---|---|---|
| SEL (all) | 365 | 13.2 | 12.7 | 10.0 | 10.0 | 6.7 | 9.4 |
| A | 46 | 12.8 | 13.5 | 10.0 | 11.0 | 7.8 | 8.7 |
| B | 69 | 15.6 | 15.8 | 15.0 | 13.0 | 7.9 | 11.9 |
| C | 41 | 14.2 | 16.5 | 15.0 | 16.0 | 6.2 | 9.7 |
| D | 43 | 12.6 | 13.0 | 10.0 | 9.0 | 7.0 | 11.5 |
| E | 61 | 13.8 | 11.9 | 10.0 | 7.0 | 5.4 | 6.9 |

| Forecaster | s(f-x) | $r_{fx}$ | ME | MSE | RMSE | $SS_{sc}$ |
|---|---|---|---|---|---|---|
| SEL (all) | 8.9 | 0.43 | 0.47 | 79.2 | 8.9 | 0.10 |
| A | 7.9 | 0.55 | −0.70 | 62.1 | 7.9 | 0.18 |
| B | 12.1 | 0.31 | −0.16 | 145.3 | 12.1 | −0.03 |
| C | 9.3 | 0.38 | −2.32 | 91.6 | 9.6 | 0.03 |
| D | 10.2 | 0.48 | −0.42 | 104.7 | 10.2 | 0.21 |
| E | 6.9 | 0.39 | 1.84 | 51.4 | 7.2 | −0.08 |

1. Fredericksburg A-Index forecasts are continuous-type integer forecasts. These verification summary statistics are the same as for 10.7 cm solar flux forecasts (Table III-4).

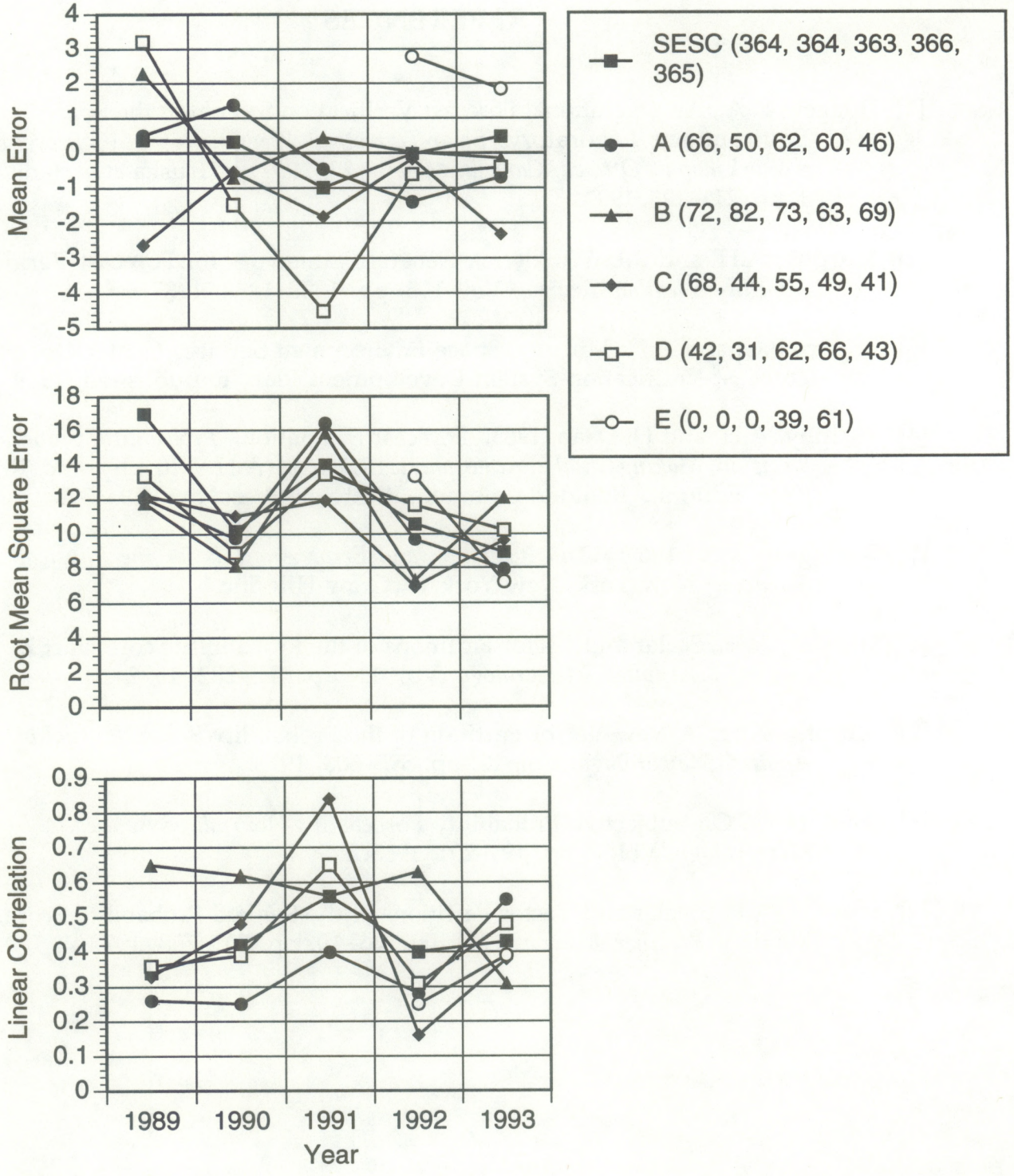



**Figure III-5.** Trend diagrams for Fredericksburg A-Index forecasts. The parameters and legend are similar to those in Figure III-4.

# REFERENCES


[1] Doggett, K.A., An Operational Forecast Verification System at the Space Environment Laboratory, *Solar-Terrestrial Predictions - IV, Proceedings of a Workshop at Ottawa, Canada, May 18-22, 1992*, J. Hruska *et al.* (eds.), Vol. 1, pp. 211–223, 1993.

[2] Murphy, A.H. and R.L. Winkler, A General Framework for Forecast Verification, *Monthly Weather Review*, Vol. 115, pp. 1330–1338, 1987.

[3] Brown, B.G. and A.H. Murphy, Space Environment Services Center Operational Forecast Verification System Development Plan, unpublished, 1989.

[4] Murphy, A.H. and H. Daan, 1985: Forecast Evaluation. *Probability, Statistics, and Decision Making in the Atmospheric Sciences* (A.H. Murphy and R.W. Katz, Editors). Boulder, Colorado, Westview Press, pp. 379–437.

[5] Bevington, P.R., 1969: *Data Reduction and Error Analysis for the Physical Sciences.* New York, New York, McGraw-Hill, Inc.

[6] Murphy, A.H., Scalar and Vector Partitions of the Probability Score: Part I, *Journal of Applied Meteorology*, Vol. 11, pp. 273–282, 1972.

[7] Murphy, A.H., A New Vector Partition of the Probability Score, *Journal of Applied Meteorology*, Vol. 12, pp. 595–600, 1973.

[8] Sanders, F., On Subjective Probability Forecasting, *Journal of Applied Meteorology*, Vol. 2, pp. 191–201, 1963.

[9] Murphy, A.H., Scalar and Vector Partitions of the Ranked Probability Score, *Monthly Weather Review*, Vol. 100, No. 10, pp. 701–708, 1972.

## Appendix A: Verification Glossary

**accuracy.** The average degree of correspondence between individual pairs of forecasts and observations. It is measured by the **mean square error**.

**base rate.** The uncertainty in the occurrence of the observations. The base rate is related to the **marginal distribution** of the observations, p(x).

**bias.** Also called **reliability**, is the degree of correspondence between the average forecasts and average observations. This type of bias is also known as overall bias, systematic bias, or unconditional bias. See the entry for **reliability (in-the-small)** for a type of bias specific to probability forecasts.

**box plot.** A diagram plotting the **quantiles** of a probability distribution of a variable. The box contains the values between the upper **quartile** and lower **quartile**, or 50% of the distribution.

**Brier score.** See **probability score**.

**calibration.** See **reliability (in-the-small)**.

**conditional distribution.** The probability distribution of a variable, given that a related variable is restricted to a certain value. The conditional distribution of the forecasts, given the observations, p(f | x), is related to the **discrimination** or **likelihood**. The conditional distribution of the observations, given the forecasts, p(x | f), is related to the **calibration** or **reliability**.

**conditional quantile plot.** A diagram plotting specific **quantiles** of a **conditional distribution** of a variable.

**continuous forecast.** A forecast of a continuous (or semi-continuous) variable. Continuous forecasts at SEL are integer value forecasts of a continuous parameter (e.g., A-Index or 10.7 cm solar flux).

**covariance.** The sample covariance is a measure of the relationship between the forecasts and observations and is defined as the **mean** of the products of the deviations of each forecast/observation pair:

$$s_{fx}^{\,2} = \frac{1}{N}\sum_{i=1}^{N}\left[(f_i - \bar{f})(x_i - \bar{x})\right],$$

where $N$ is the sample size, $f$ refers to forecasts, and $x$ refers to observations.

**discrimination.** The extent to which the relative frequency of use of a particular forecast differs given two different observations. This is the same as **likelihood**.

**discrimination diagram.** A diagram plotting the conditional distribution of the forecasts, given that the forecast event occurred, and the conditional distribution of the forecasts, given that the event did not occur. Ideally, these two distributions will be well separated from one another. For perfect forecasts, the diagram would consist of two distinct spikes.

**forecast verification.** The process of determining the quality of forecasts. The assessment of forecast quality involves the statistical characteristics of the forecasts and matching observations, and the relationships between them. Verification methods are numerous, and the approach used to verify a particular forecast/observation data set is determined by the forecast type and the objectives of the particular verification user.

**histogram.** A diagram plotting the **marginal distribution** of a variable in terms of its frequency of occurrence.

**joint distribution.** The probability distribution defined over two or more variables. The joint distribution of the forecasts and observations, p(f,x), contains all of the information relevant to the verification problem (except for time relationships). This joint distribution can be decomposed into expressions involving **conditional** and **marginal distributions**:

$p(f,x) = p(x \mid f)\, p(f)$, the **calibration - refinement** factorization and
$p(f,x) = p(f \mid x)\, p(x)$, the **likelihood - base rate** factorization.

**likelihood.** See **discrimination**.

**linear-correlation coefficient (r).** A measure of the linear association between the forecasts and observations. The sample linear-correlation coefficient is defined as

$$r_{fx} = \frac{s_{fx}^2}{s_f s_x},$$

where $s_{fx}^2$ is the sample **covariance**, $s_{fx}$ is the sample standard deviation (see **variance**), $f$ refers to forecasts, and $x$ refers to observations.

**marginal distribution.** A probability distribution of a single variable. The marginal distribution of the forecasts, p(f), is related to the **refinement** or **sharpness**. The marginal distribution of the observations, p(x), is related to the uncertainty or **base rate**.

**mean.** The sample mean, arithmetic mean, or average, is defined as

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i ,$$

where $x$ is the variable and $N$ is the sample size. The mean of a variable is also indicated by angled brackets, e.g. <x>.

**mean error (ME).** The mean of the differences of the forecasts and observations. In this case, the variable in the definition of **mean** is $(f\text{-}x)$, where $f$ refers to forecasts and $x$ refers to observations. The mean error is a measure of the overall **bias** of the forecasts. A positive mean error indicates that, on the average, the forecasts were larger than the corresponding observations.

**mean square error (MSE).** The mean of the square of the differences of the forecasts and observations. For mean square error, the variable in the definition of **mean** is $(f\text{–}x)^2$, where $f$ refers to forecasts and $x$ refers to observations. The mean square error is a measure of forecast **accuracy**. The lower the mean square error, the more accurate the forecasts.

**median.** The value of a variable that divides a sample in half. There is equal likelihood that the value of a variable will be greater or smaller than the sample median. The median of variable x is abbreviated as Med(x).

**probabilistic forecast.** A forecast of the probability that a particular event will occur during a given time frame. Probabilistic forecasts range from 0 (event cannot occur) to 1.0 (event is certain to occur). The probability forecast range at SEL is limited from 0.01 to 0.99.

**probability score.** Or **Brier score** (BS), is the mean square error of probability forecasts. The Brier score is defined as

$$BS = (1/K)\sum_{k=1}^{K}(\mathbf{r}_k - \mathbf{d}_k)(\mathbf{r}_k - \mathbf{d}_k)^T ,$$

where K is the number of forecasts in the sample, **r** refers to the forecast vector, and **d** refers to the observation vector. The range of the probability score is the closed interval [0,2] and has a negative orientation; that is, smaller values are better.

**quantile.** The specific value of a variable that divides the distribution into two parts, those values greater than the quantile value and those values that are less. For instance, p percent of the values are less than the $p^{th}$ quantile.

**quartile.** A quantile that separates one quarter of the values in a distribution from the remaining three quarters.

**ranked probability score (RPS).** The mean square error of probability forecasts for ordinal predictands. It is the difference between cumulative forecasts and observations. RPS is defined as

$$RPS = (1/K)\sum_{k=1}^{K}(\mathbf{R}_k - \mathbf{D}_k)(\mathbf{R}_k - \mathbf{D}_k)^T,$$

where K is the number of forecasts in the sample, and where **R** and **D** refer to the cumulative forecast and observation vectors, respectively. Two useful vector partitions of the RPS are

$$\begin{aligned} RPS &= RPS_{sc} + REL - RES \\ &= REL + RES' \end{aligned},$$

where REL is the **reliability** and RES and RES' are the **resolutions** as defined by Murphy and Sanders, respectively. The $RPS_{SC}$ is the ranked probability score with respect to the sample climatology and is defined as

$$\begin{aligned} RPS_{sc} &= \bar{\mathbf{D}}(\mathbf{1} - \bar{\mathbf{D}})^T \\ &= RES + RES' \end{aligned},$$

where **D** is the cumulative observation vector, **1** is the identity vector, and the overbars indicate the mean.

**refinement.** Related to the **marginal distribution** of the forecasts, p(f). The amount of refinement is the extent to which the individual forecast differs from the overall average forecast. This term has the same meaning as **sharpness**.

**reliability (in-the-small) (REL).** The degree of correspondence, over one or more subsamples of verification data involving identical forecasts, between the average observations for the subsamples and the respective forecasts. This type of **bias** relates only to probability forecasts and is the same as **calibration**. For J subsamples of K probability forecasts, the reliability is defined as

$$REL = (1/K)\sum_{j=1}^{J} K^j(\mathbf{r}^j - \bar{\mathbf{d}}^j)(\mathbf{r}^j - \bar{\mathbf{d}}^j)^T,$$

where the overbars denote the mean, and **r** and **d** refer to the forecast and observation vectors, respectively. Note that REL $\geq 0$ and has a negative orientation; that is, smaller values indicate more reliability.

**reliability diagram.** A diagram in which the conditional distribution of the observations, given the forecast probability, is plotted against the forecast probability. The distribution for perfect forecasts will plot along the 45° diagonal.

**resolution (RES or RES').** For probability forecasts, relates to the correspondence of occurrence of events in subsamples of observations associated with distinct forecasts. For J subsamples of K probability forecasts, a definition of resolution [7] is given as

$$RES = (1/K)\sum_{j=1}^{J} K^j \left(\bar{\mathbf{d}}^j - \bar{\mathbf{d}}\right)\left(\bar{\mathbf{d}}^j - \bar{\mathbf{d}}\right)^T,$$

where the overbars indicate the mean and **d** is the observation vector. Another definition of resolution [8], is given as

$$RES' = (1/K)\sum_{j=1}^{J} K^j \bar{\mathbf{d}}^j \left(\mathbf{1} - \bar{\mathbf{d}}^j\right)^T,$$

where **1** is the identity vector. Note that both RES and RES' are greater than or equal to zero; however, RES has a positive orientation and RES' has a negative orientation. That is, larger values of RES indicate more resolution whereas smaller values of RES' indicate more resolution.

**sharpness.** See **refinement**.

**skill.** The average **accuracy** of the forecasts in the sample relative to the **accuracy** of the forecasts produced by a reference method. Examples of a suitable reference include forecasts of recurrence, persistence, sample climatology, or the output of a forecast model. Skill can be measured by any number of so-called skill scores. For instance, the skill score with respect to sample climatology ($SS_{sc}$) uses the average observation over the sample period as the reference forecast. This skill score is defined as

$$SS_{sc} = 1 - \left(\frac{MSE}{MSE_r}\right),$$

where MSE is the **mean square error** of the sample forecasts and $MSE_r$ is that of the reference forecasts. For probabilistic forecasts, where the observations are either 0 or 1, the sample climatology skill score can be written as

$$SS_{sc} = 1 - \left(\frac{MSE}{s_x^2}\right),$$

where $s_x^2$ is the sample **variance** of the observations. In addition, skill scores based on

the MSE can be decomposed as follows:

$$SS = r_{fx}^{\ 2} - \left[ r_{fx} - \left( s_f / s_x \right) \right]^2 - \left[ \left( \bar{f} - \bar{x} \right) / s_x \right]^2,$$

where $r_{fx}$ is the **linear-correlation coefficient,** $s$ is the sample standard deviation (see **variance**), $f$ refers to forecasts, and $x$ refers to observations. The second term in this decomposition is related to the conditional **bias** of the forecasts and the third term is related to the unconditional **bias**.

**variance.** A measure of the dispersion of a data set. The sample variance is defined as

$$s^2 = \frac{1}{N-1} \sum (x_i - \bar{x})^2,$$

where $x$ is the variable and $N$ is the sample size. The sample standard deviation is the square root of the sample variance.



*U.S. GOVERNMENT PRINTING OFFICE:1994-573-013/00041