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

OFFICE NOTE 264

Comparative Evaluation of ECMWF and NMC Spectral Forecasts, February - July 1982

John D. Stackpole
M. Steven Tracton
Development Division

AUGUST 1982

Introduction

Since February 1982 NMC has been receiving the 24 through 120 hour ECMWF forecasts, from 1200 GMT initial data, of 500 mb height and sea level pressure on a near real time basis. More recently (early May) we have also been receiving the "D+8" 500 mb height fields map (the 5 day mean map centered on forecast hour 204). The forecasts are received over the GTS on a 5° latitude longitude grid for the northern hemisphere north of 20° and are biquadratically interpolated to NMC's "standard" 65x65 polar stereographic grid (one bedient mesh length) for display and verification purposes. The relative skill of ECMWF vis-a-vis NMC forecasts has been assessed objectively and subjectively. The following presents the results of our evaluations through July 1982.

Objective Verifications

Comparative objective verifications of the ECMWF model and NMC's spectral model forecasts of 500 mb height are presented in Fig. 1 (a)-1(f) for the six months available to date. The particular verification statistic shown here is the standard error (a.k.a. the standard deviation of the error or, more colloquially, "the rms error with the mean error removed") of the 500 mb heights plotted as a function of forecast hour. For the spectral model (denoted by 'SMG3C', and the filled symbols connected by solid lines) the statistic is plotted every 12 hours up to 48 hours and is for all the forecasts (averaged) for the month; after that time forecasts from the 00Z initial time (the only ones made) are, naturally, the only ones considered. The ECMWF model scores (plotted as open symbols connected by dashed lines) are the monthly mean for the once per day forecasts made from the 12Z initial time and plotted every 24 hours - the only times available.

The calculation of these statistics is done by comparing the forecast field in question with the verifying observations (from RAOBS) and then summing the mean and mean square error for the month at each RAOB station in a set of networks of stations. Further summations are done over each network and the time and space ensemble statistics are plotted in Fig. 1. Four networks of stations are presented: NH102, a quasi-uniformly distributed selection of 102 RAOBS over the northern hemisphere; ALA23, the 23 RAOBS in and near Alaska; NA110, covering southern Canada, the US and adjacent waters; and EUR96, European radiosondes from 10°W eastward to about 27.5°E longitude.

Turning to the content, rather than the construction, of the figures, we see that in general the ECMWF model is doing better than NMC but not by a very wide margin. The percent differences between the scores looks to be 10-15% at most and is often less.

Another way of putting this is in terms of the rate of the increase of error with time: Overall the ECMWF forecasts are about 12 hours better than NMC's. However, from NMC's viewpoint in using the ECMWF forecasts in an operational forecast context this 12 hour skill advantage is of no particular value. NMC bases its medium range projections on forecasts originating at 00Z; ECMWF's long data hold (and transmission delays) require that NMC consider the ECMWF forecast originating at the 12Z synoptic time preceeding NMC's initial data time.

There are some curiosities to be noted as well:

In Alaska the ECMWF model does better than NMC (and in February <u>much</u> better) except at 120 hours when NMC pulls ahead slightly but consistently, May excepted.

There is no indication that the ECMWF model is generally pulling ahead of NMC at the longer ranges, even discounting Alaska, inspite of the considerably more sophisticated physical parameterizations in the European model. Five days may be too short a time for those effects to reach full potential, however; the recently received D+8 maps may tell a different story in the future.

A preliminary objective verification of thirty-two cases of the 500 mb D+8 maps, made by F. Hughes of NMC's Forecast Division between mid-May and early August of 1982, gave standardized correlation of height anomaly coefficients of 24.0 for the NMC forecasts and 29.3 for the ECMWF forecasts. According to Hughes this is a significant difference between the models; however a 5 point spread in this score is typical of the shorter range verifications as well. Thus the enhanced physical parameterizations of the European model seems to have no medium range benefits in these summer verifications at 500 mb. Another season or another level of the atmosphere might tell another story.

Over North America, in February, NMC actually bettered ECMWF (by an insignificant amount to be sure, but we were better), but that state of affairs did not persist. Whether February was a statistical fluke or a true seasional variation in relative skill only time will tell.

The statistics for the NH102 network compare quite favorably both in terms of numerical values and information content with those presented by Bengtssen & Lange ("Results of the WMO/CAS NWP Data Study & Intercomparison Project for forecasts for the N. Hemisphere in 1979-80" - W.M.O., undated). If nothing else this observation increases one's confidence that both verification groups (NMC and the Finnish Meteorological Institute) are doing the right thing.

A minor item: on some of the charts, ALA23 & NA110 in particular, are found the verifications of NMC's LFM model - it has its problems after 24 hours, particularly in the winter and early spring. (The "P" on a few of the charts marks the error level of a 24 hour old Hough analysis serving as a 24 hour persistence forecast.)

Additional objective verifications in the form of correlations of forecast vs. observed height anomalies (at 500 mb) lead to the same conclusions as do the standard error calculations.

Subjective Evaluations

Subjective evaluations of ECMWF versus NMC spectral model forecasts were performed to assess the meteorological significance of differences between them. Qualitative judgements, based upon such factors as the position, amplitude, and vertical consistency of weather systems, were tabulated according to the following scale

5 ECMWF much better

4 ECMWF better

3 ECMWF same

2 ECMWF worse

1 ECMWF much worse

Separate assessments were made for the sea-level pressure and 500 mb height fields over western and eastern North America. At the longer ranges the European region was also considered. Beyond 48h, the NMC medium range guidance is produced from 0000 GMT while that from ECMWF is from 1200 GMT. In these evaluations, NMC was given the 12 hr advantage in rendering judgements since only the 12 hr old ECMWF guidance is available near the deadline for issuing medium range predictions. Thus, for example, the 96h ECMWF prognoses were assessed relative to the 84h NMC spectral forecasts value at the same 1200 GMT verifying time.

The responsibility for the 24 and 48h evaluations was that of the individual presenting the daily map discussion. The medium range prediction group assessed the guidance beyond 48h.

The results of subjective evaluation of the short-term (24 + 48h) predictions are shown in Tables 1a and 1b. The scores shown are the March and April monthly averages. It is clear from these results that there is a distinct tendency for the ECMWF forecasts to be judged better than the NMC guidance. Of the 16 possible combinations of month, level, time period, and region, ECMWF was considered more useful in 12, worse in 3, and the same in 1. At both 500 mb and the surface the advantage of ECMWF over NMC was larger in March than April and more consistent over western North America than eastern North America.

Another view of the overall results of subjective evaluation of the short-term forecasts is provided by histograms (Fig. 2) of the percentage of judgements falling into each of the categories, 1-5, for the combined sample of surface and 500 mb, 24 and 48 h, and eastern plus western North America. In March (Fig. 2a) of the total judgements rendered, 40 percent were ECMWF and NMC the same. The balance of judgements is clearly skewed towards ECMWF better with an overall average of 3.4. During April, the percentage of neutral judgements increases to 50 percent, with a slight skewness toward "ECMWF better" producing an overall average of 3.1.

The results of evaluation of the extended range forecasts are shown in Tables 2a and 2b. Again, the scores are the March and April monthly averages. The differences here are not as clear cut as for the short-term forecast comparisons. The most noticeable feature is that for both months ECMWF tends to be as good or better than NMC, over western North America and Europe, but is consistently worse over eastern North America. Otherwise, there is no obvious stratification amongst of results amongst the possible combinations of area, time range, or level.

Histograms of the same sort shown above indicate that for the extended range neither system provides guidance that is consistently better or worse than the other. The overall averages for March and April were 2.9 and 3.0, respectively.

In interpreting these later results, keep in mind that the NMC extended guidance has a built in 12h advantage in the subjective appraisal. While this might be irrelevant from the utilitarian viewpoint at NMC, nevertheless, a "same" implies that the ECMWF 12h old forecasts are as good as the latest NMC prognosis.

In summary, the results of subjective evaluation demonstrate a generally consistent, but not overwhelming superiority of ECMWF over NMC spectral model forecasts. In this regard, the subjective assessment virtually echoes the results of the objective verifications.

Table 1a

MARCH

Short Term

	ENA	WNA
500		
24	3.2	3.7
48	3.6	3.7
Surface		
24	2.8	3.3
48	3.8	3.8

Table 1b

APRIL

Short Term

	ENA	WNA
500		
24	3.1	3.3
48	2.7	3.0
Surface		
24	2.8	3.2
48	3.6	3.1

Table 2a.

MARCH

Extended

	ENA	WNA	EUR
500			
84/96	2.8	3.0	3.1
108/120	2.6	3.2	3.1
Surface			
84/96	2.9	3.1	3.2
108/120	2.6	3.1	2.8

Table 2b.

APRIL

Extended

	ENA	WNA	EUR
500			
84/96	2.6	3.2	3.1
108/120	2.3	2.7	3.1
Surface			
84/96	2.9	3.4	3.5
108/120	2.4	3.1	3.2

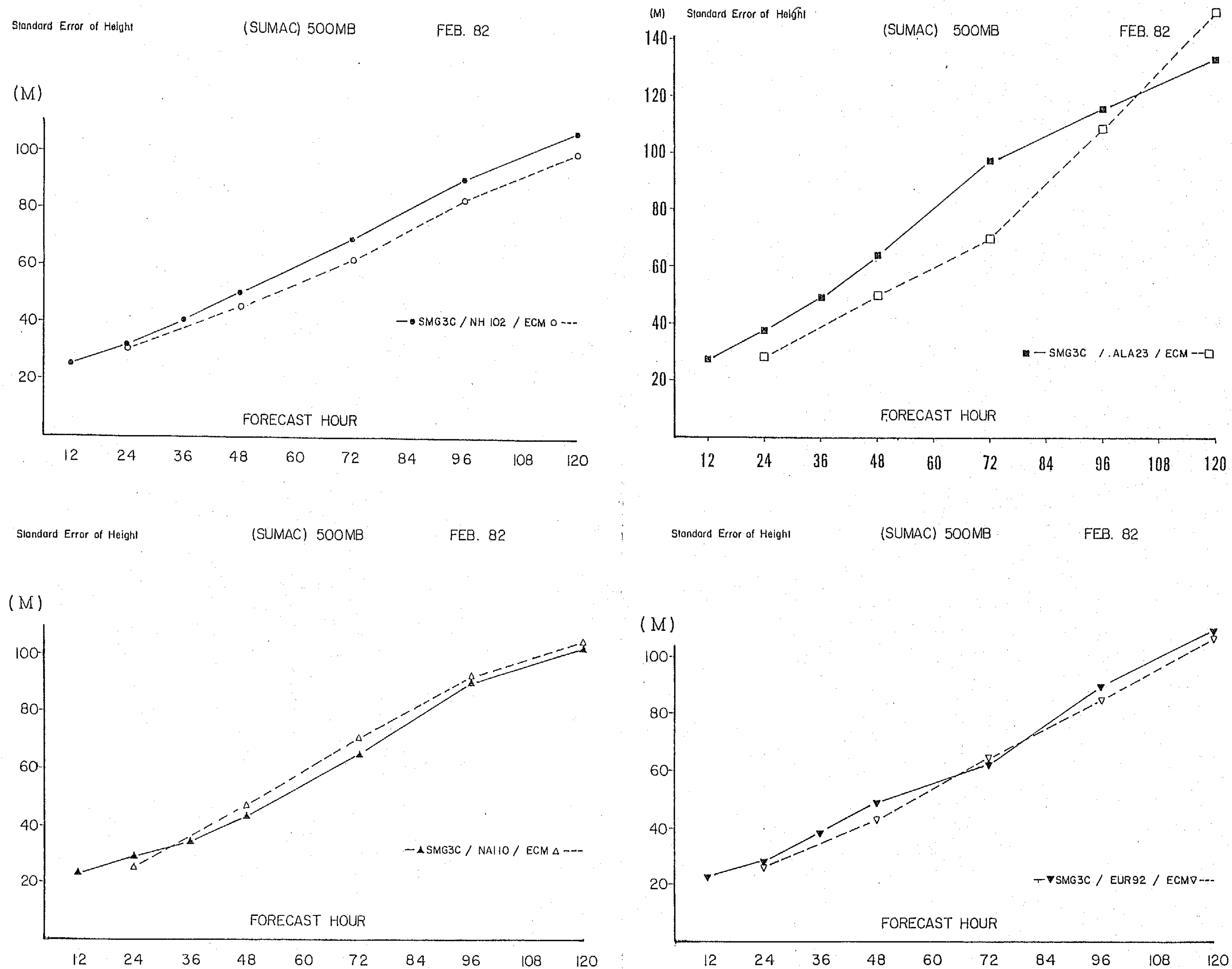

Figure 1a. FEB 82

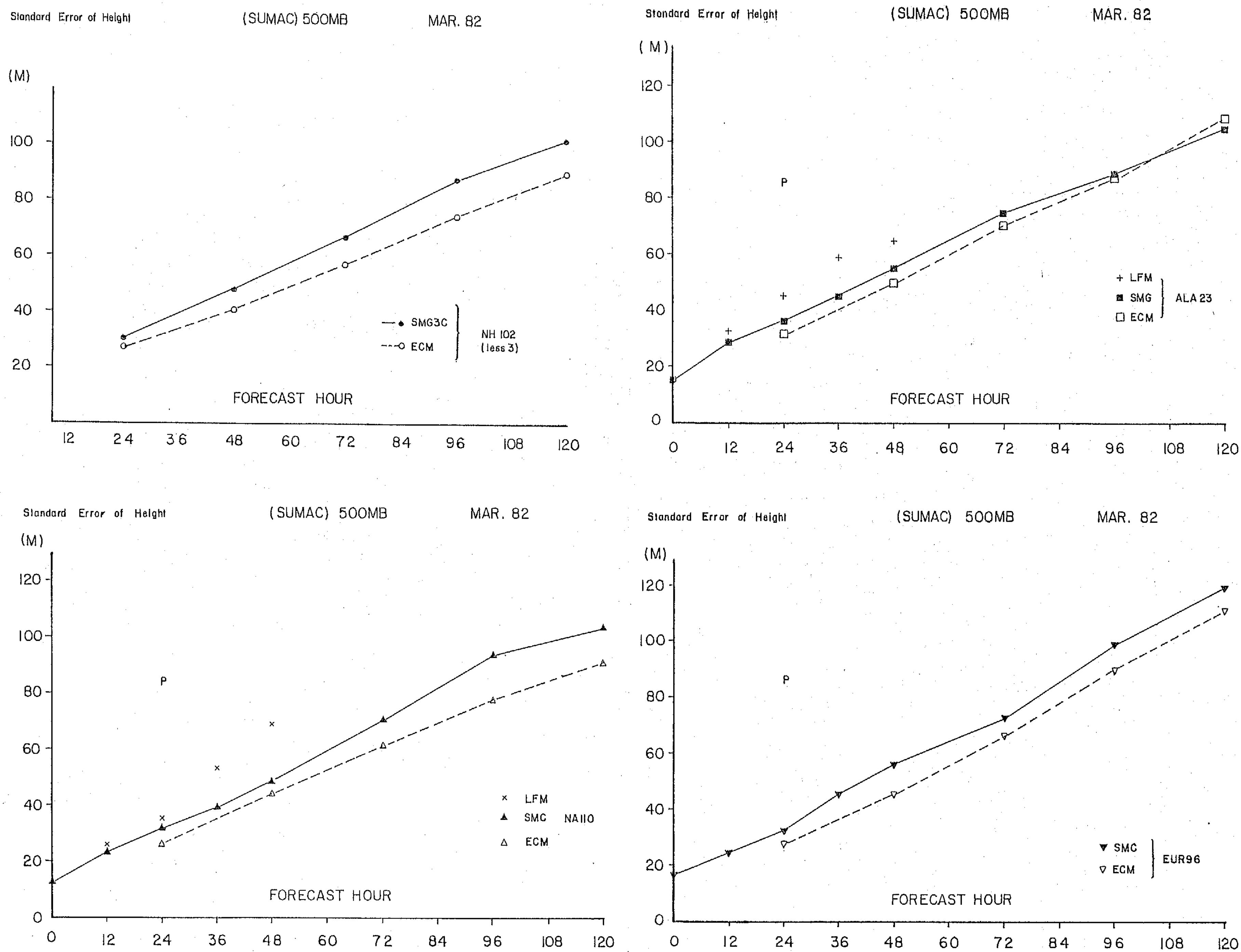

Figure 1b. MAR 82

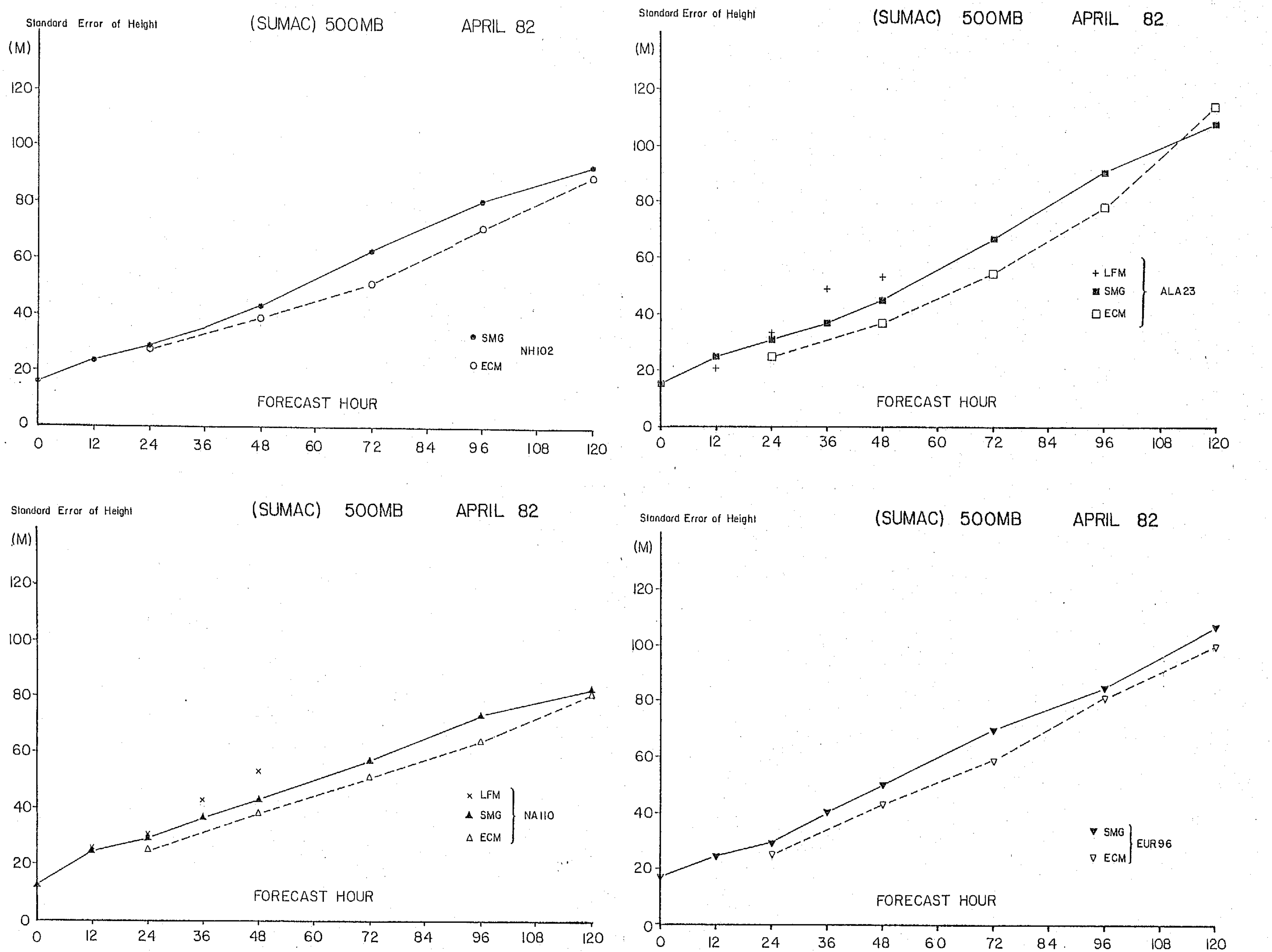

Figure 1c. APRIL 82

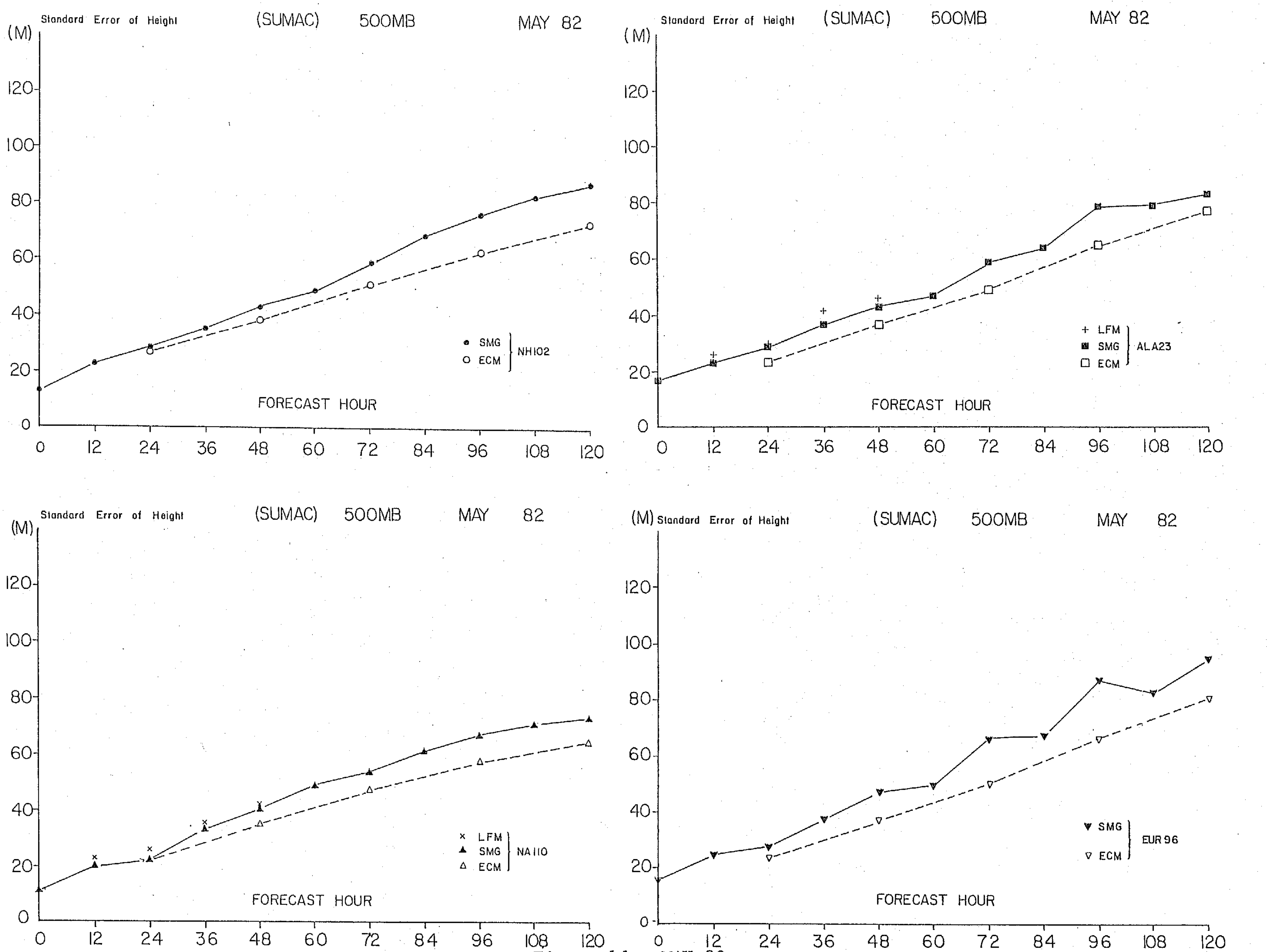

Figure 1d. MAY 82

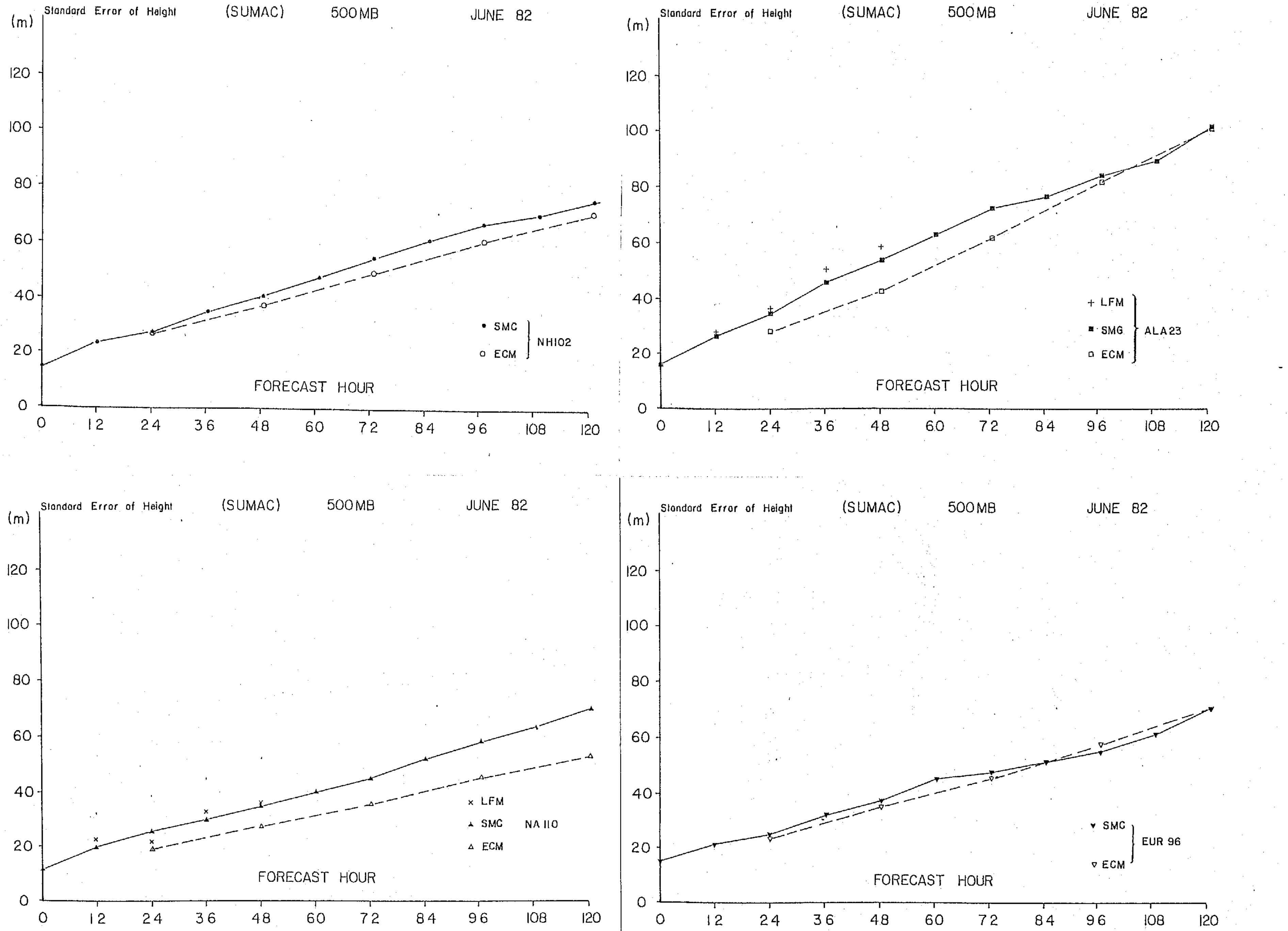

Figure 1e. JUNE 82

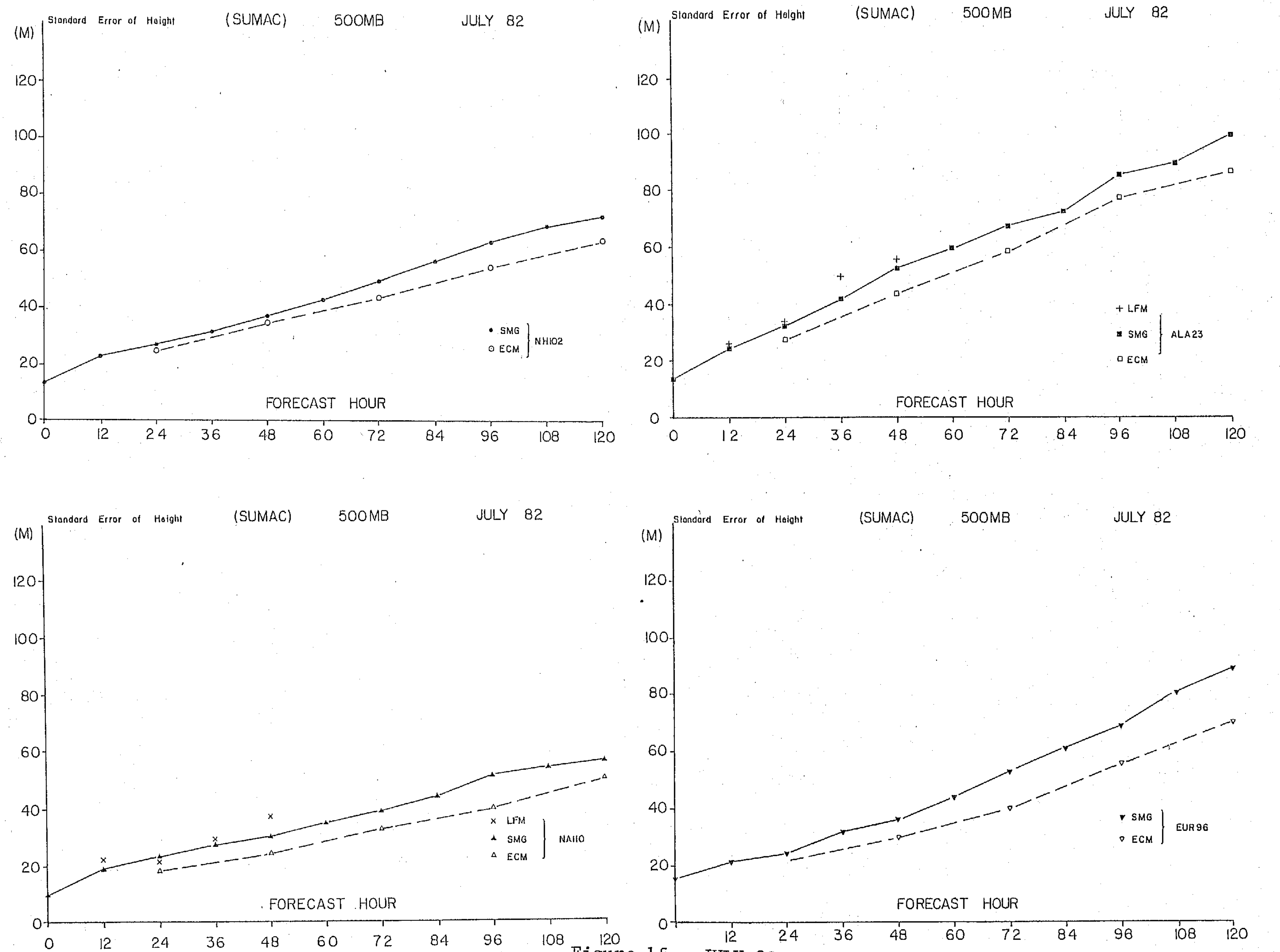

Figure 1f. JULY 82

Fig. 2 March (a) and April (b) percent judgements for each category, 1-5, for combined sample of 24 and 48h, surface and 500 mb, eastern and western North America forecasts. 1 = ECMWF much worse; 2 = ECMWF worse; 3 = ECMWF same; 4 = ECMWF better; 5 = ECMWF much better relative to NMC spectral model.

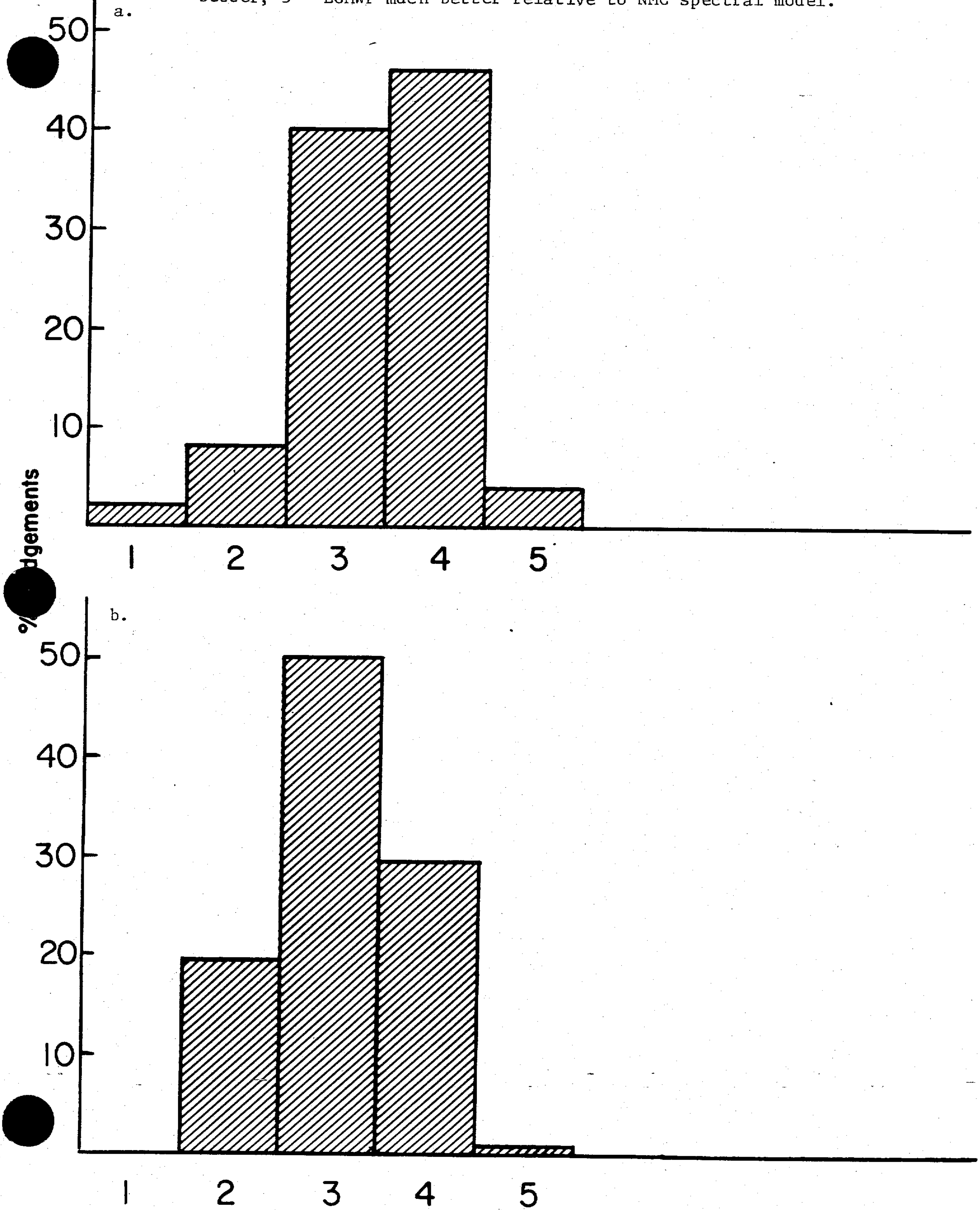

Fig. 3 Same as Fig. 2, except for 84h spectral/96h ECMWF, 108h spectral/120h ECMWF, and European region included.