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

OFFICE NOTE 386

THE PRACTICE AND PROBLEMS OF OPERATIONAL MEDIUM RANGE FORECASTING AT THE NATIONAL METEOROLOGICAL CENTER

by

M. Steven Tracton and Francis D. Hughes
Prediction Branch, Climate Analysis Center

January 1992

This is an unreviewed manuscript, prmarily intended for informal exchange of information among NMC staff members

386

ABSTRACT

Medium range forecasting here refers to predictions covering conditions in the period 6-10 days in the future. Current operational procedures and products are reviewed. The discussion includes presentation of verification statistics of the ECMWF and NMC model guidance and the final manually modified predictions. The paper concludes with a summary of experiments underway to predict forecast skill and to assess the utility of ensemble prediction.

Note: This paper is a summary of a presentation at the ECMWF Third Workshop on Meteorological Operational Systems, 18-22 November 1991, and is reproduced in the workshop proceedings.

1. INTRODUCTION

Medium range refers here to the 6-10 day forecast period. What distinguishes the medium-range from shorter-term forecasts is that the latter concern the daily sequence of weather systems while the former apply to mean conditions. The difference reflects the current approximate limit of useful skill in numerical weather prediction of synoptic-scale systems. Thus, the medium-range guidance provides users only a general picture of the circulation and associated temperature and precipitation patterns. No attempt is made to provide explicit information on variability within the 6-10 day forecast period.

Figure 1 displays an example of the principal medium-range guidance products. The first panel (Fig. 1a) is the 6-10 day mean 500-mb height and height anomaly chart. The associated forecasts of surface temperature and precipitation departures from normal are shown in Figs. 1b and 1c, respectively. As an adjunct to these products, a written account ("prog discussion") is issued which highlights the reasoning and relative importance of the various tools (next section) used in making each forecast. Implicitly or explicity stated in this discussion is an estimate of forecaster confidence. The forecasts are produced three times a week (Monday, Wednesday, and Friday) by the Climate Analysis Center, a division of the National Meteorological Center (NMC).

The next section (section 2) describes the tools and procedures used in producing the medium-range forecasts, while section 3 presents verification scores. Section 4 concludes the paper with a brief outline of research aimed at forecasting skill and exploring the efficacy of ensemble prediction.

2. TOOLS AND PROCEDURES

2.1 6-10 Day Mean 500 MB Heights

As outlined by Wagner (1989), the cornerstone of the 6-10 day mean 500 mb-height predictions is the output from the ECMWF model and NMC's own medium range forecast (MRF) model. The MRF is a global spectral model as documented by Kanamitsu (1989), with the more recent significant changes being an increase from T80 to T126 horizontal resolution (18 levels in both) on March 6, 1991 and the replacement of the Optimum Interpolation (OI) analysis system with the variational Spectral Statistical Interpolation (SSI) scheme on June 25, 1991 (Kalnay et al., 1991). Also available are a regression corrected version of the MRF and multiple linear regression output derived from ECMWF and MRF prognoses. Both are based on the average errors of the 6-10 day mean 500-mb height progs over the immediately preceeding 60 days. The 500-mb height and anomaly prog is typically a "blend", i.e., weighted average, of the ECMWF and MRF guidance from the latest runs. The blend may include input from successive runs of one or both models, as well as contributions from forecast and/or observed persistence and relevant historical analogs. In effect, the blend is a "poor person's" ensemble prediction. The weights are subjectively based upon i) recent track record of ECMWF versus MRF performance, ii) consistency between successive runs, iii) internal consistency as judged from teleconnection patterns, iv) forecaster experience/judgement, and v) systematic errors, for example, with respect to regime, season, topography, and geographical region.

2.2 Temperature and Precipitation

Temperature forecasts are based primarily on Klein (1985) objective statistical specifications from the 6-10 day mean 500-mb height prog, together with historically preferred temperature anomaly patterns whose height anomalies resemble those of the current situation (specification analogs). Precipitation is based primarily on direct output from the MRF model and patterns of preferred precipitation categorical amount, again utilizing the specification analog technique. For both temperature and precipitation the forecaster may introduce subjective modifications which reflect known biases in any of the inputs.

3. VERIFICATIONS

3.1 6-10 Day 500-MB Heights

Verifications are in terms of the standardized anomaly correlation score over North America (Hughes, 1991)[1]. Figure 2 displays the broad view of this score over the past decade for the MRF, ECMWF, and "official" 6-10 day ("D+8") 500-mb heights. During the early 80's, the ECMWF model had a clear advantage, and the forecasters clearly recognized this. After some apparent reluctance, as the MRF improved and became competitive with ECMWF in the mid 80's, forecasters began paying more attention to both

[1] 1 The standardized anomaly is the anomaly divided by the climatological standard deviation. Use of the correlation between standardized anomalies, rather than the anomalies themselves, precludes inherently larger anomalies, e.g., at higher latitudes, from overwhelming smaller, but possibly meteoroloigically significant anomalies, e.g., at lower latitudes. For this score over the limited North American domain, experience indicates a value of about .17 or greater translates to useful skill in the associated surface temperature predictions.

models. The net result was that towards the end of the record the official forecat was better than either model alone. The approximate 5 point lead over the models is comparable probably to several years of model development.

At the time of the this workshop (mid-November, 1991), an apparent difficulty had arisen with regard to the credibility of the ECMWF forecasts. Specifically, since operational implementation of the T213 model on 18 September, verification scores on average were very poor with respect to the MRF, and there was much less consistency from one run to the next. To show this Fig. 3 displays the anomaly correlation scores for September and October in terms of 5 case running means, while Fig. 4 shows the standard deviation about these means. While there was one period during the beginning of October where the opposite was true, the MRF scored significantly better than ECMWF prediction following introduction of the T213 model (and has continued to do so overall through at least the beginning of December). Note too from Fig. 4 an almost doubling of the average standard deviation, a measure of the variability from one run to the next, from the period before to after implementaion of the new model. It should be noted that to whatever extent circulation regime might have played a role in this increased variability, there was no such effect on the MRF (not shown). Finally, Fig. 5 shows an example of successive model runs to illustrate the credibility problems of the ECMWF model relative to the MRF predictions in the eyes of forecasters. This is not to say that the MRF is always consistent and more skillful, for it is not. But there is little doubt that overall the ECMWF D+8 forecasts are less skillful and have less run to run consistency than both the MRF predictions and the earlier (T106) version of the ECMWF model.

3.2 Temperature and Precipitation

The measure of skill used to verify the 6-10 day forecasts is the Heidke skill score for 61 temperature and 100 precipitation verification stations in the conterminous United States (Hughes, 1991; Wagner, 1989). This score is 100 for a perfect forecast and 0 for random chance. The annually averaged skill scores for three and five class temperatures are shown in Fig. 6 and for three class precipitation in Fig. 7. The general increase in skill through the years is apparent, presumably reflecting model improvements. Not surprisingly, temperature can generally be predicted with higher skill than precipitation. It should be noted that the skill of the temperature forecasts is typically several points above the objectively specified temperatures obtained from the 6-10 day 500-mb heights. Thus, the forecaster improves not only upon the predicted 500-mb circulation in comparison to the models alone (Sec. 3.1), but also upon the objective guidance given the final 500-mb prog. With regard to precipitation, except for the 1990 annual average, the official or subjectively modified final forecasts are usually better than the raw model guidance (1991 will not be another exception).

4. FORECASTING SKILL AND ENSEMBLE PREDICTION

It has become increasingly clear that single deterministic predictions are not suitable for medium and extended range prediction. A single forecast represents just one of many possible solutions given uncertainties in the initial state. Ensemble forecasts are necessary to describe an array of possible outcomes and their likelihood. As a result, research into the design and application of the ensemble approach to forecasting is currently at the forefront of numerical weather prediction activities. A major challenge, one which requires intensive interaction between users and developers of models, is how to intelligently condense, present, and use the vast amounts of information that will be generated from ensembles.

There are several research projects at NMC aimed at ensemble prediction. As described by Kalnay and Toth (1992) at the recent ECMWF Workshop on Predictability, they include a quasi-operational test of a "poor person" Mont Carlo method for forecasting skill, experiments in lagged average forecasting (LAF), and efforts to develop procedures for selecting the most rapidly growing perturbations. In the "poor person" approach the ensembles consist of operational forecasts from NMC, ECMWF, UKMO, and JMA. The agreement between these forecasts, together with forecast persistence and rms amplitude anomaly, serve as the predictor of skill. In effect, this procedure is an attempt to place the qualitative/subjective approach currently employed by the medium-range forecasters into a quantitative/objective framework. The LAF experiments center around weighting the laggged forecasts so that all members have an "equivalent" age. This so-called "NewLAF" procedure avoids the main disadvantage of LAF, namely, perturbations of unequal size with those corresponding to "older" forecasts being too large. Research aimed at selecting the fastest growing perturbations centers around integrating a large number (100-1000) randomly selected initial conditions for a single time step to find the "most promising" (fastest growing) members for a subset of longer ensemble predictions. Finaly, test and evaluation of alternative strategies for Monte Carlo prediction are in the context of both a quaisi-operational routine mode and individual case study appraoch. The latter is specifically aimed at assessing the predictability of regime transitions.

References

Hughes, F. D., 1991: Skill of medium range forecasts. NMC Office Note 377, 32 pp. (available from the Director, National Meteorological Center, W/NMC, Washington, D.C. 20233)

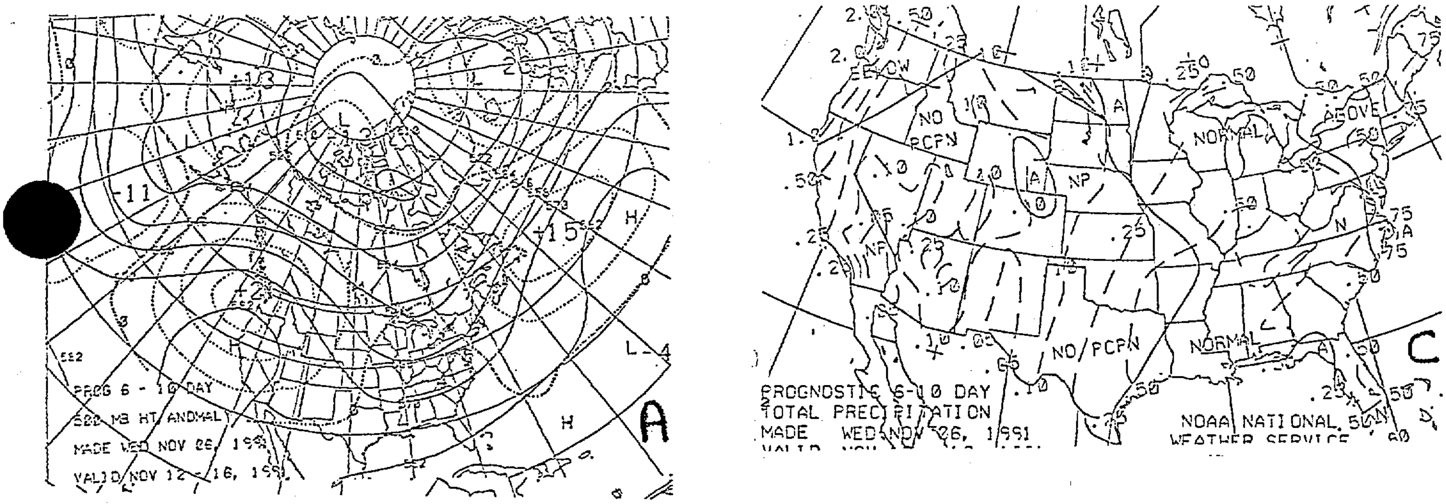
Kalnay E., W. Baker, M. Kanamitsu, and R. Petersen, 1991: Overview of the NMC analysis and modeling plan. Preprints, Ninth conference on Numerical Weather Prediction, Denver, Amer. Meteor. Soc., 5-10.

______, and Z. Toth, 1992: Lagged average forecasts and Monte Carlo experimental forecasts at NMC. Proceedings of the ECMWF Workshop on Predictability, 13-15 November 1991, Reading, U.K.

Kanamitsu, M., 1989: Description of the NMC Global Data Assimilation and Forecast System, Wea. and Forecasting, 4, 335-342.

Klein, W. H., 1985: Space and time variations in specifying monthly mean surface temperature from the 700 mb height field. Mon. Wea. Rev., 113, 277-290.

Wagner, A. J., 1989: Medium- and long-range forecasting. Mon. Wea. Rev., 4, 413-426.



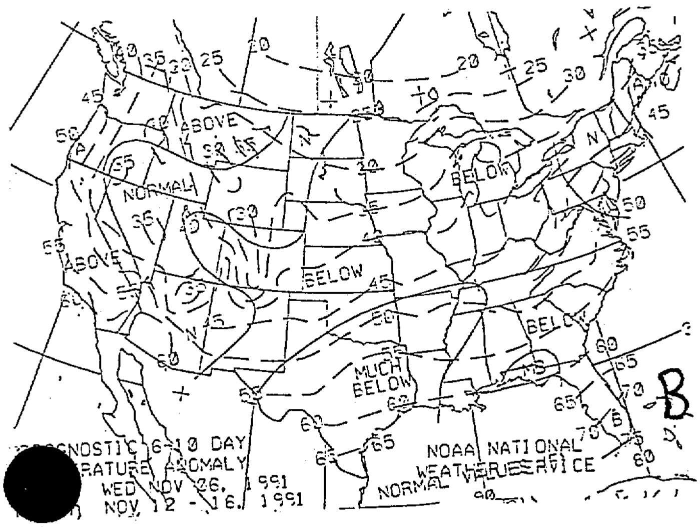


Fig. 1. Sample of format for the 6-10 day predictions; a) 500-mb height and height anomaly, b) temperature anomaly, c) precipitation anomaly. Dashed lines in "a" and "b" give the normals for the 5-day period.

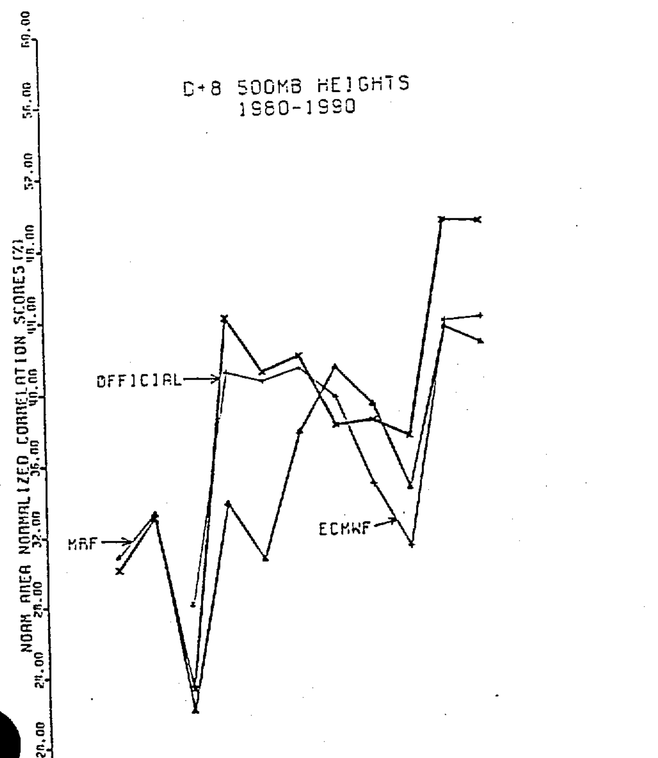


Fig. 2. Yearly average standardized anomaly correlation scores over North America for the D+8 (6-10 day mean) 500-mb heights.

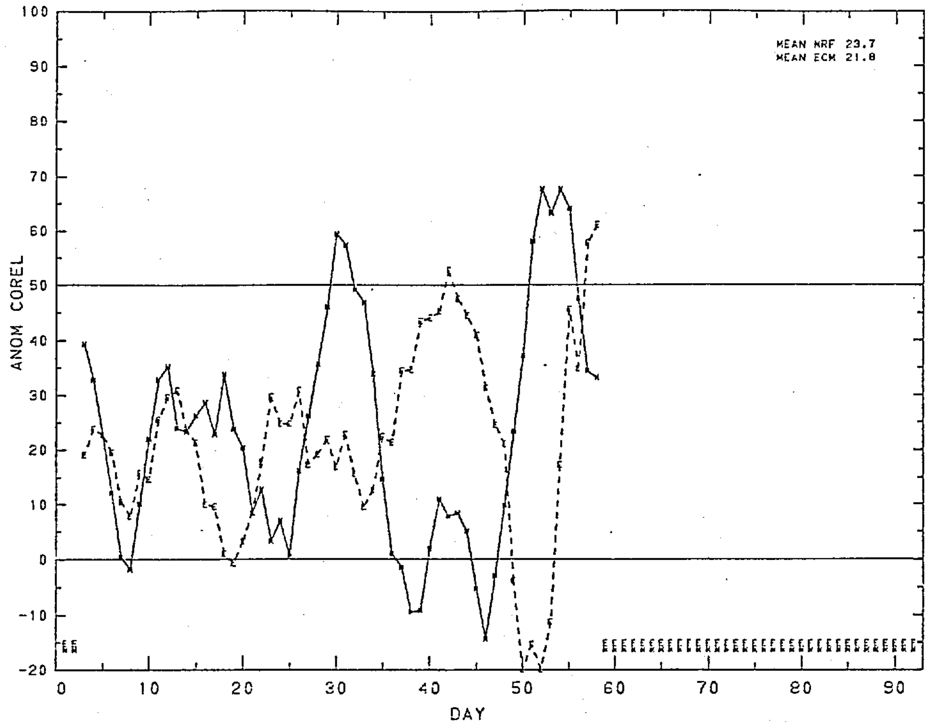


Fig. 3. Same as Fig. 1, except for five case running means of forecasts verifying on September 1 (day 1) through October 31 (day 61), 1991. Solid, MRF; dashed, ECMWF.

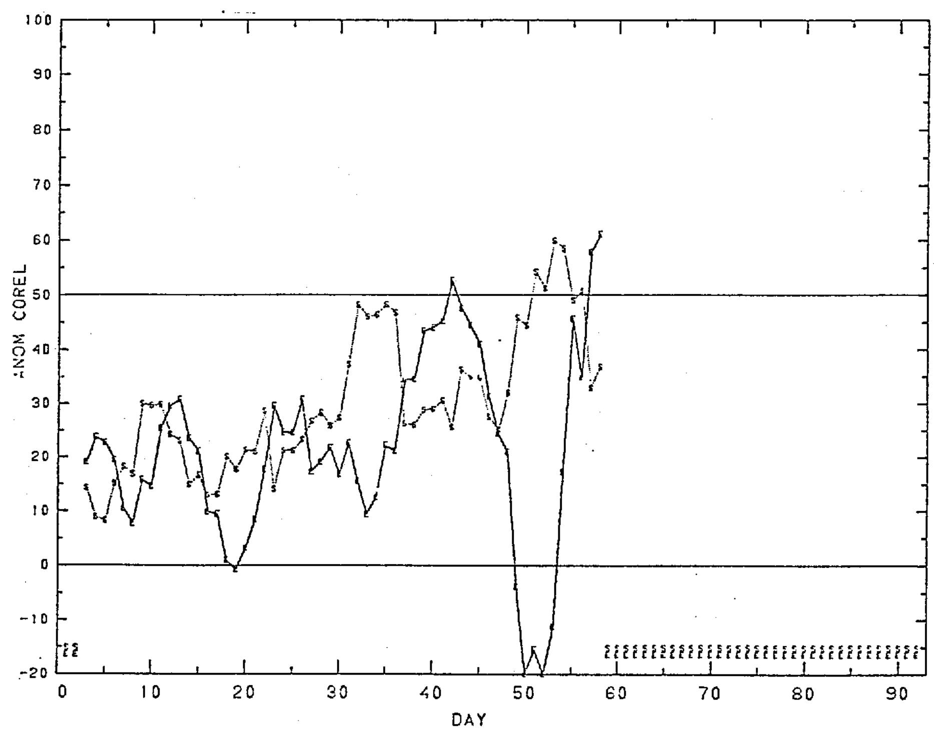


Fig. 4. Standard deviation (dotted) about ECMWF 5-day means (solid) of Fig. 3.

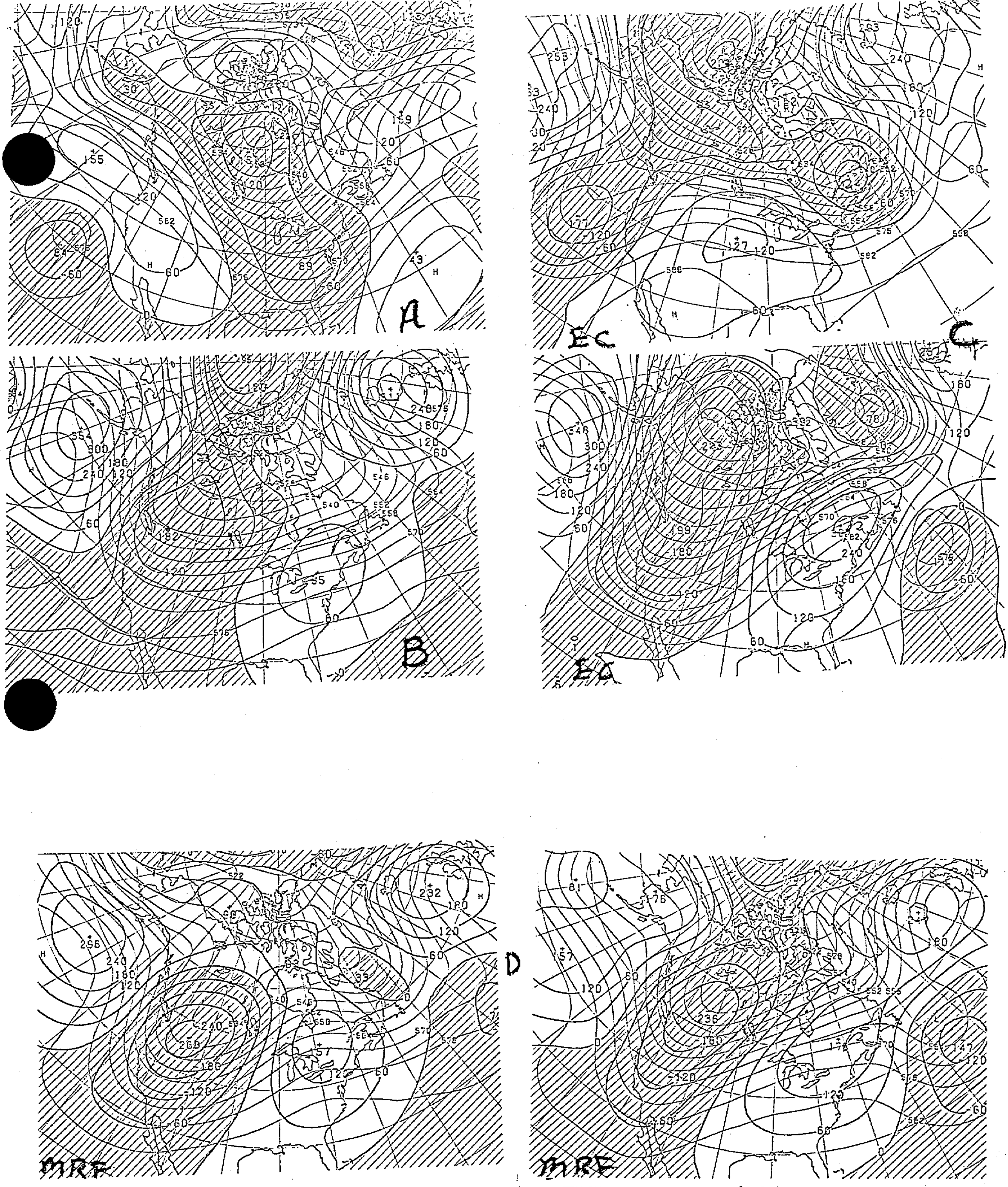


Fig. 5. Observed 5 day mean 500-mb height and height anomalies (positive values shaded) centered Oct. 15 (a) and Oct. 23 (b), 1991 and successive ECMWF (c) and MRF (d) 6-10 day (D+8) predictions from initial conditions about "a" and verifying about "b".

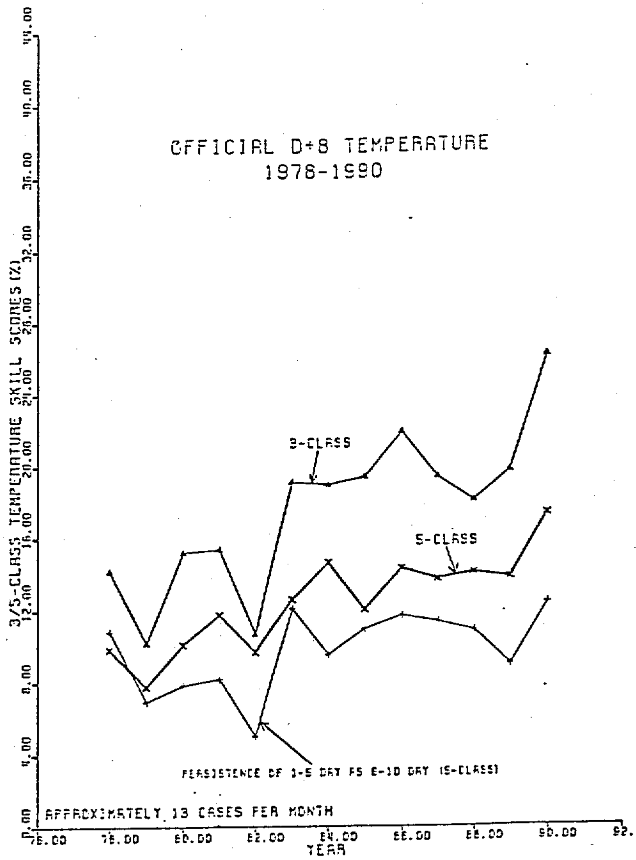


Fig. 6. Annual mean 3-class and 5-class temperature skill scores for the "official" 6-10 day mean predictions for the U.S.

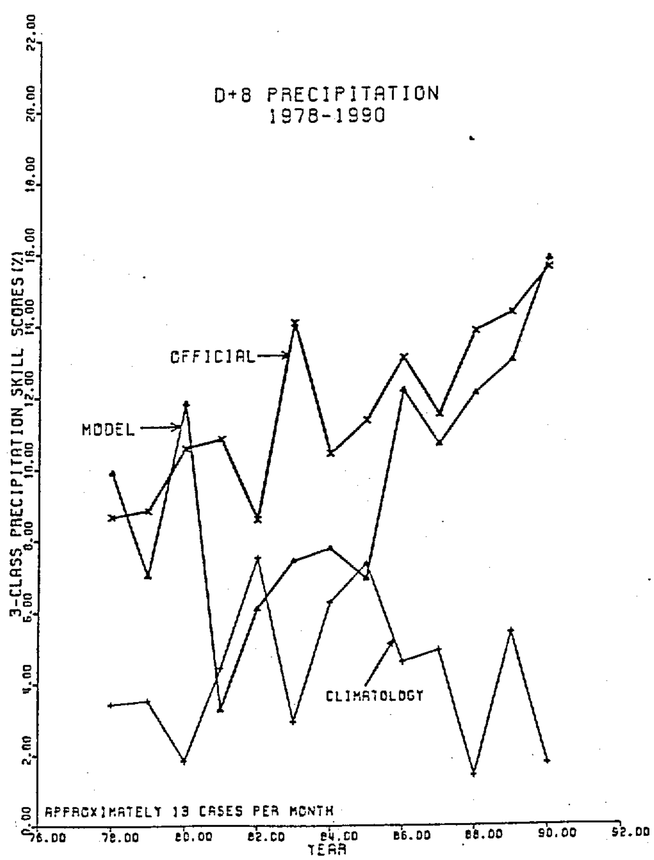


Fig. 7. Same as Fig. 6, except for 3-class precipitation.