NOTES AND CORRESPONDENCE

On the Skill and Utility of NMC's Medium-Range Central Guidance

M. STEVEN TRACTON

Climate Analysis Center, National Meteorological Center, National Weather Service/National Oceanic and Atmospheric Administration, Washington, D.C.

19 October 1992 and 19 November 1992

ABSTRACT

Verification scores are presented to illustrate the general success of NMC forecasters in providing the best day 3, 4, and 5 mean sea level pressure and 6–10-day mean 500-mb height fields given the operationally available array of often conflicting NWP model solutions. As a primer on NMC efforts to enhance the utility of the medium-range forecast guidance, a brief overview is provided on the rationale and expectations for ensemble prediction.

1. Introduction

The following is largely an extension of the dialogue between Newman (1991) and McPherson and Olson (1991) concerning the paper by Livingston and Schaefer (1990) on the National Meteorological Center's (NMC) medium-range guidance provided by forecasters of the Meteorological Operations Division (MOD). Newman questioned the role of centralized subjective guidance in making 3- to 5-day forecasts, where model solutions often diverge widely. McPherson and Olson defended the philosophy for NMC forecasters to evaluate, interpret, and perhaps modify model output with the intention of "steering" other meteorologists toward "the most likely sequence of events."

The particular issues addressed in this paper are twofold—the contribution of NMC forecasters in the "man–machine mix" and the question of varying levels of confidence in the forecasts. Documentation is provided to support the contention that forecasters are quite successful in formulating the most probable solution given the array of often conflicting NWP output (section 2). Implicit in this process is a subjective judgement of the degree of confidence in their forecasts, which is communicated by the discussions [Automation of Field Observations and Service (AFOS) headers PMDHMD, PMDEPD] accompanying the various AFOS graphics (9JH, 9KH, 9LH). As noted by McPherson and Olson, reliable estimates of confidence can contribute significantly to the practical utility of forecasts. In the near future, explicit measures of confidence that can be expressed in terms of the range and likelihood of various scenarios will be possible via formalizing the ensemble approach to forecasting. An introduction to the fundamental concepts and practical aspects of ensemble prediction is provided in section 3.

Note that in the aforementioned dialogue, "medium range" refers only to the MOD day 3, 4, and 5 forecasts. The term "medium range," however, applies also to the 6–10-day mean (D + 8) forecasts (AFOS graphics 96T, 96E, 96C, 96H; AFOS headers FEUS06, FEUS40KWBC) produced by NMC's Climate Analysis Center (CAC). The preceding remarks and discussion that follows encompass both medium-range periods.

2. Skill scores

NMC forecasters typically have several numerical predictions from different models and successive initial conditions to consider. For MOD's day 3, 4, and 5 forecast products, the principal model guidance is from the European Centre for Medium-Range Weather Forecasts (ECMWF), the United Kingdom Meteorological Office (UKMO), and the NMC Medium-Range Forecast (MRF) Model. For CAC's 6- to 10-day period of responsibility, the same is true, except that output from the UKMO is not available. The foundation of MOD's product suite is the daily mean sea level pressure (MSLP) field, while for CAC it is the 500-mb 6–10-day mean (D + 8) height field. In both cases the final version of these products, which are key tools in specifying sensible weather elements, reflects forecasters' judgments on such factors as model systematic errors, continuity and consistency among runs, recent track records of model performance, and characteristic

Corresponding author address: M. Steven Tracton, Climate Analysis Center, NWS/NOAA, W/NMC22, Washington, DC 20233.

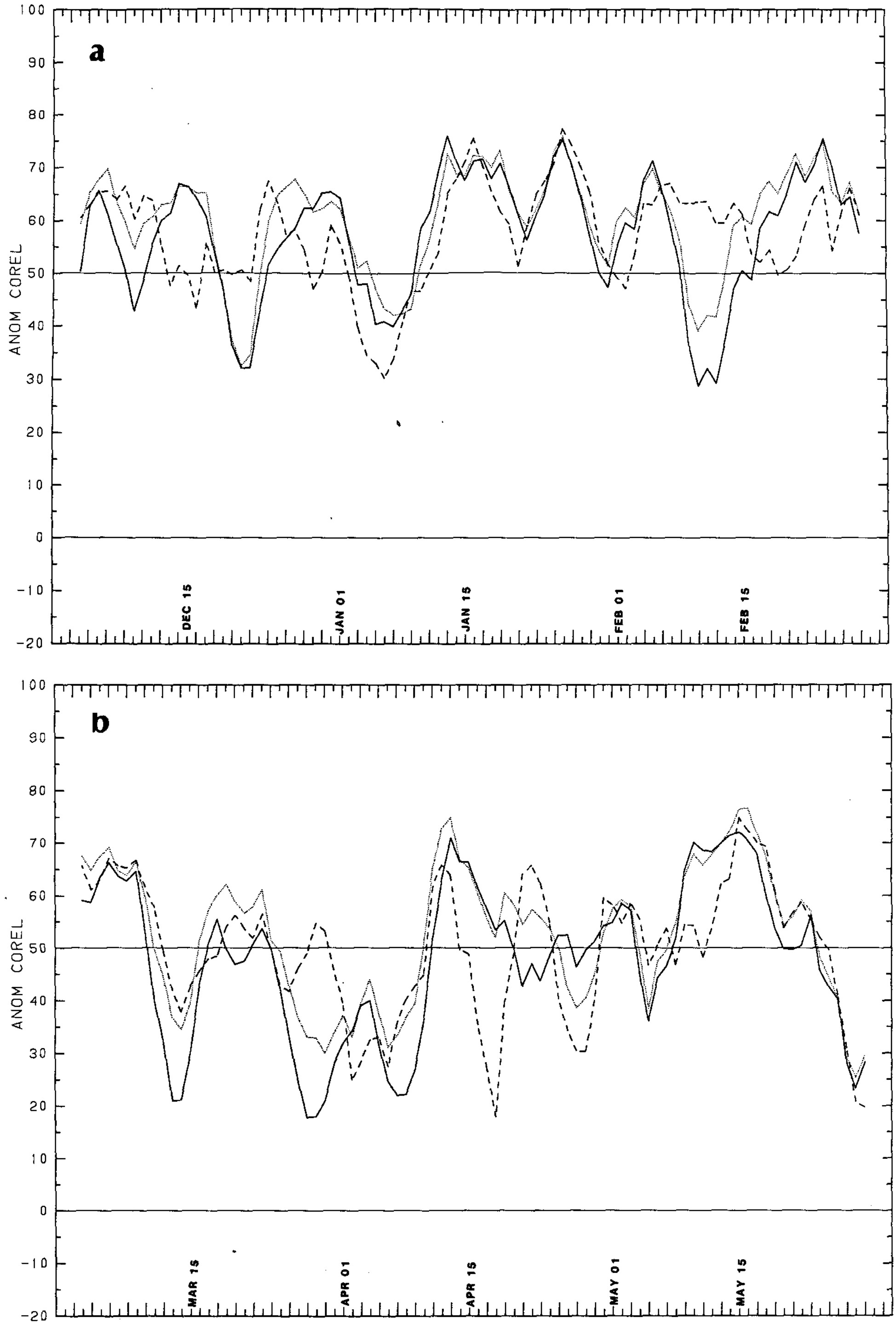

FIG. 1. Day 5 MSLP five-case running means of the standardized anomaly correlation score (×100) for 3-month seasons: (a) winter, DJF 1991/92; (b) spring, MAM 1992; (c) summer, JJA 1992; and (d) fall, SON 1992. Date for each season indicated along abscissa. Solid, dash, dotted lines for MRF, ECMWF, and forecasters, respectively.

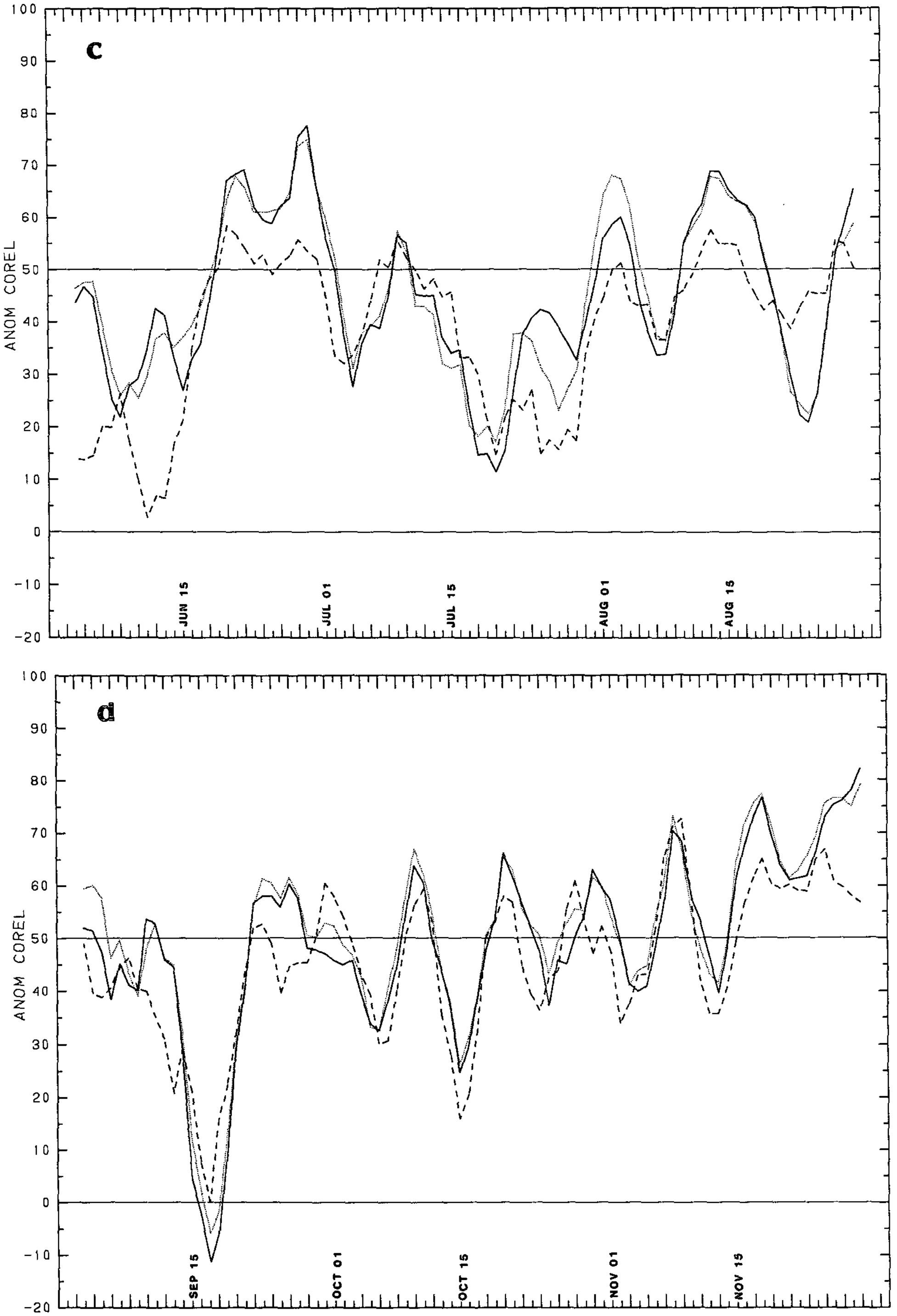

FIG. 1. (*Continued*)

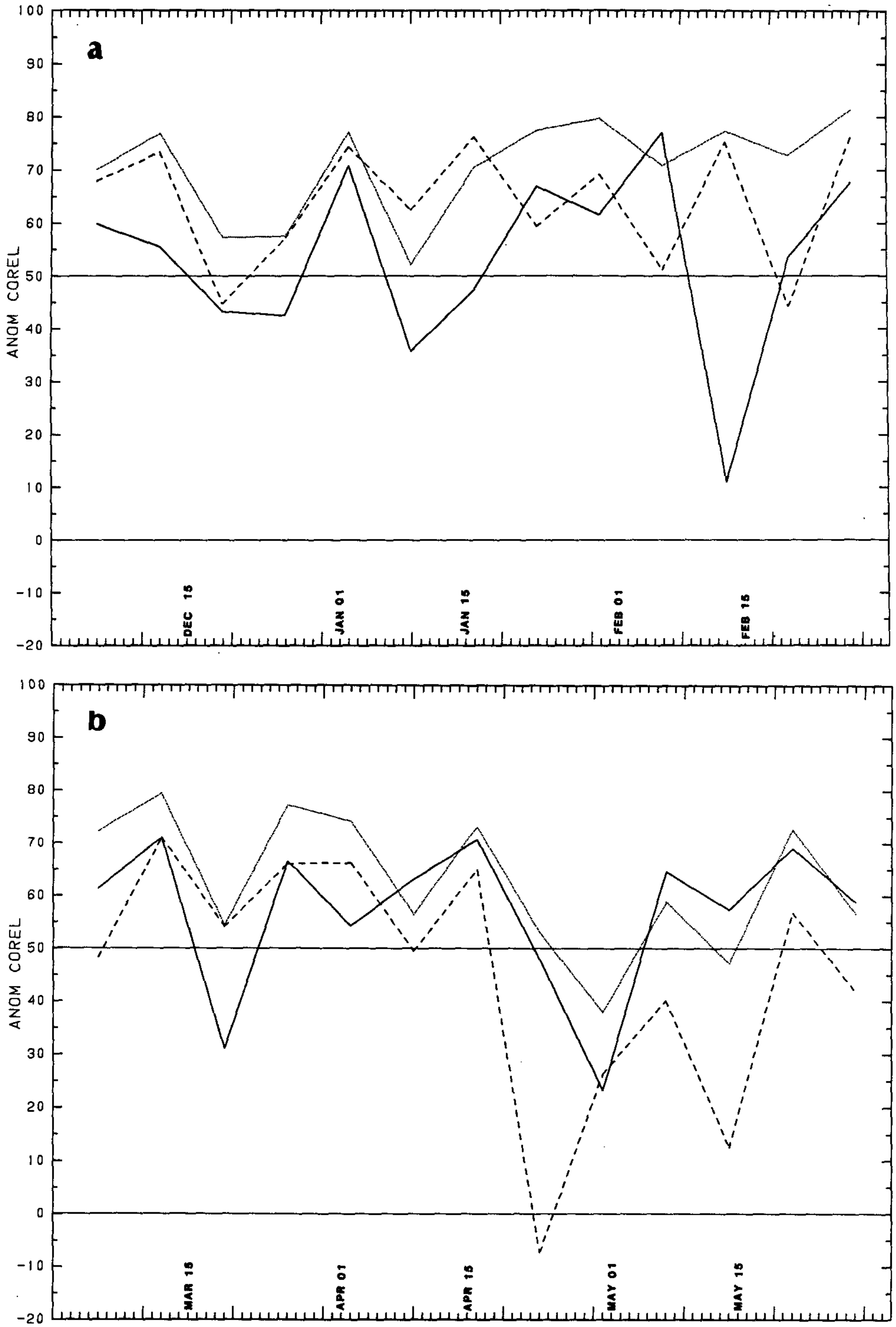

FIG. 2. Same as Fig. 1 except for D + 8 500-mb height and 3-case weekly means.

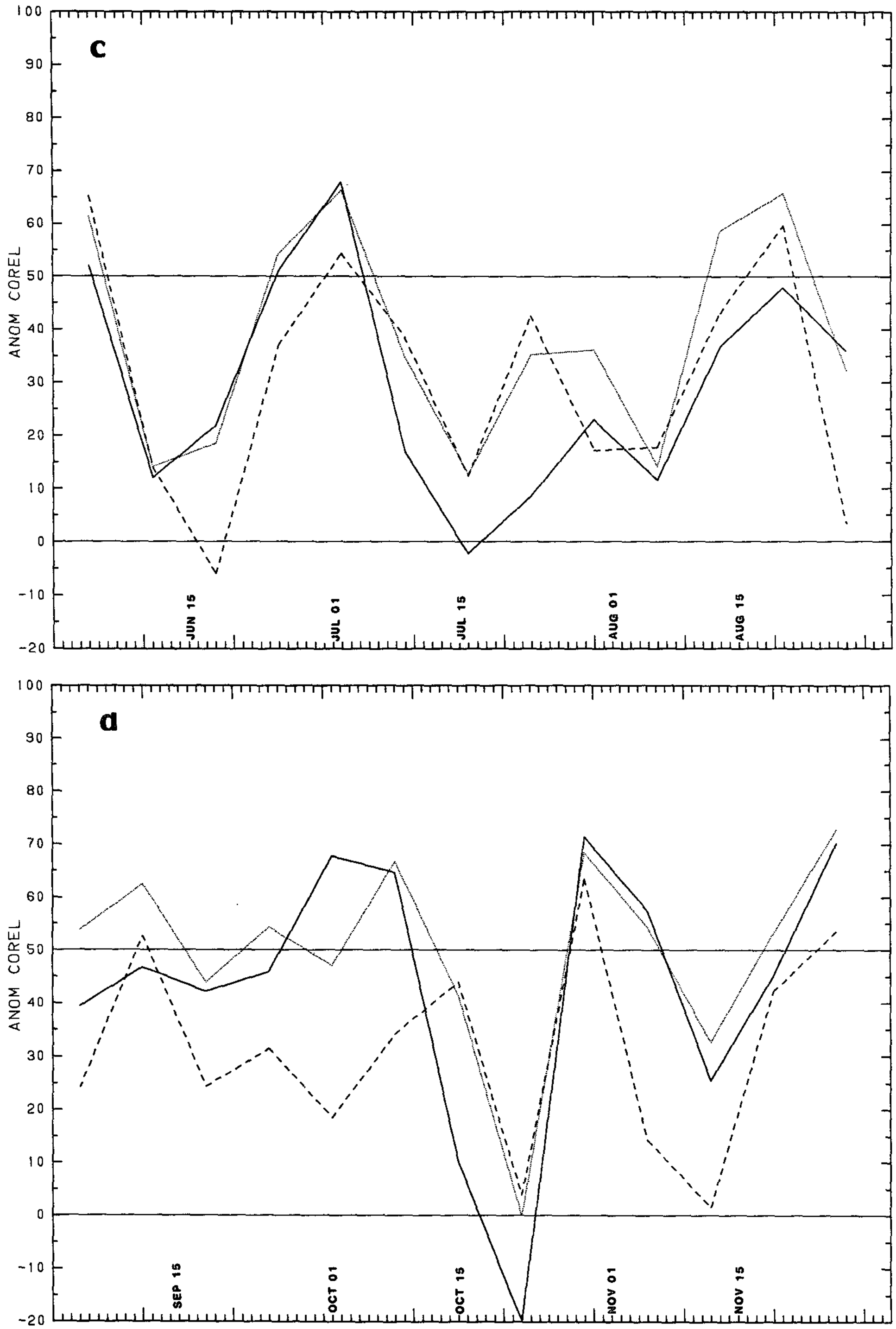

FIG. 2. (*Continued*)

TABLE 1. Seasonal mean standardized anomaly correlation scores (×100) for day 5 MSLP predictions. Seasons as defined in Fig. 1.

	MRF	ECMWF	Forecasters
WINTER	56.6	57.6	59.4
SPRING	47.5	49.5	51.6
SUMMER	44.9	38.6	44.9
FALL	48.9	45.7	50.4

model behavior as a function of differing circulation regimes. The role of forecasters is to combine these factors and judge the most likely outcome. Success or failure of this often daunting task can be measured by the skill of the modified fields relative to the raw model output. Ideally, the final product should score consistently as well as or better than the best model forecast, especially in cases where the models diverge significantly. Relative to this ideal, as illustrated below, NMC forecasters have achieved considerable success.

Verifications are in terms of the standardized anomaly correlation score over North America (Hughes 1991).[1] Scores are presented for the day 5 mean sea level pressure (MSLP) (Fig. 1) and D + 8 500-mb height (Fig. 2) for the latest available 3-month winter (DJF, 1991/92), spring (MAM 1992), summer (JJA 1992), and fall (SON 1992) seasons. For the day 5 predictions, which are produced every day, results are plotted in the form of running 5-day means to smooth curves and avoid distraction from the essential trends. Conclusions drawn from the day 5 results apply equally as well to days 3 and 4 (not shown). For the D + 8 forecasts, which are issued only three times per week, scores plotted are weekly (3 case) averages. Each chart displays the scores for the latest available ECMWF and MRF model runs[2] and final, forecaster-modified fields. Seasonal means are summarized in Tables 1 and 2.

On average, forecasters beat the best model by 2 to 3 points at day 5, except for summer, where they tie with the MRF (Table 1). Summer is also the season that on average features the largest discrepancy between models and, therefore, the largest dilemma for forecasters. The ability of forecasters to score generally as well as if not better than the best model can also be seen by scanning the sequences of scores plotted in Fig. 1. Despite frequently large differences between models and considerable temporal variability in each model's solutions, forecasters more often than not correctly selected the most appropriate "model of the day" in formulating their predictions. For example, during the period 14–19 April 1992 (Fig. 1b), forecasters correctly followed the MRF rather than what turned out to be a series of very poor forecasts by the ECMWF model. This is not to say that forecasters were not sometimes led astray as, for example, was the case between 7 and 13 February 1992 (Fig. 1a), when the MRF was clearly not the better model. A quick recovery followed, however. This was especially impressive because forecasters correctly anticipated that ECMWF's superiority would be short-lived in the face of a strong comeback by the MRF. In virtually no case over the one-year period of record in Fig. 1 did forecasters' adjustments result in predictions that scored worse than both models. Finally, it is important to add that the manually adjusted forecasts tended to be more consistent from one day to the next (not shown) and, hence, more credible than either model.

For the 6–10-day period the success of forecasters is even more impressive than for day 5. On average the skill of the manually generated charts was about 6 to 8 points greater than the best model in all seasons (Table 2). This improvement is especially significant because of the larger differences between models than at day 5. As at day 5, forecasters scored as well as or better than the best model and virtually never worse than both (Fig. 2). An especially impressive example of forecasters' skills occurred between 5 and 15 February 1992. As can be seen in Fig. 2a, forecasters not only properly chose to ignore what subsequently verified as a disastrous set of MRF predictions, but even improved on the ECMWF forecasts. Only rarely were forecasters tempted in the wrong direction by discrepancies between models, and on such occasions quick recoveries generally followed (e.g., early January 1992). Of course, as with day 5, when all model predictions are poor, forecasters have little hope of doing well.

3. Primer on ensemble forecasting

As discussed in section 1, the medium-range predictions issued by NMC reflect subjective consideration of an array of NWP guidance. The success of NMC forecasters notwithstanding, a priori, the level of con-

TABLE 2. Seasonal mean standardized anomaly correlation scores (×100) for D + 8 500-mb height predictions. Seasons as defined in Fig. 1.

	MRF	ECMWF	Forecasters
WINTER	54.4	63.7	70.7
SPRING	57.1	47.7	63.2
SUMMER	29.5	30.7	38.8
FALL	43.6	31.4	50.1

[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 at higher latitudes from overwhelming smaller but possibly meteorologically significant anomalies at lower latitudes. For this score over the limited North American domain, experience indicates a value of about .20 or greater translates to useful skill in the associated surface temperature predictions.

[2] Corresponding verification scores for the UKMO forecasts are unavailable, but qualitative assessments provide no reason to alter conclusions from those based on the ECMWF and MRF models alone.

fidence in any given forecast is usually less than total certainty. The degree of confidence closely reflects the consistency and/or continuity among the model runs and explicitly recognizes the possibility of alternative solutions having varying degrees of likelihood. Conceptually, ensemble forecasting is nothing more than this process of weighing uncertainties given a set of model runs that encompass the same verification period. The informal, nonstructured approach currently employed, however, cannot fully exploit the amounts and types of information potentially available from more formal, systematic application of ensemble prediction.

To explore and evaluate this potential, NMC has been producing experimental sets of ensembles twice per month since February 1991. In this experiment each ensemble consists of nine MRF (T62 resolution) predictions generated in the lagged-average-forecast mode (Dalcher et al. 1988) with six-hour spacing between initial conditions. Additionally, off-line experiments are ongoing to assess the relative merits of alternative strategies with respect to model configuration, ensemble size, and optimum procedures for generating initial-state perturbations. Comprehensive documentation of these experiments will appear elsewhere. Suffice it to say that results are encouraging enough, particularly in regard to providing *reliable* estimates of the range and likelihood of outcomes, for NMC to begin ensemble prediction for real-time use and appraisal in MOD and CAC medium-range forecast operations. At this writing, the precise strategy and timing of the implementation is uncertain. As might be expected, the details will reflect compromises between some perceived optimum approach and the reality of limited computer resources.

Aside from formulating an efficient and effective strategy for generating ensembles, a key challenge is condensing vast amounts of output in a coherent and "user friendly" form. Possibilities include defining the envelope and distribution of solutions (e.g., position of the often cited 540-dam thickness, "rain/snow" line), clustering ensemble members that are similar in some respect (e.g., blocking versus zonal flow regimes), and expressing probabilities (e.g., the percentage of ensemble members with 500-mb height anomalies or precipitation amounts exceeding some criteria). Because additional and/or alternative products are numerous and the required learning curve fairly steep, continual interaction among forecasters within and outside NMC is essential for realizing the full potential of ensemble prediction. To the extent that this potential is realized, it is likely that the next major advance in operational NWP will come from enhancing the utility of forecasts through the ensemble approach rather than from the steady but relatively slow increase in average skill associated with improvements in analysis/forecast systems.

Acknowledgments. The author expresses thanks to Louis Uccellini and Dave Olson for their useful reviews of the manuscript.

REFERENCES

Dalcher, A., E. Kalnay, and R. Hoffman, 1988: Medium range lagged average forecasts. *Mon. Wea. Rev.*, **116**, 402–416.

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.]

Livingston, L., and J. Schaffer, 1990: On medium-range model guidance and the 3–5-day extended forecast. *Wea. Forecasting,* **5**, 361–376.

McPherson, R., and D. Olson, 1991: Reply. *Wea. Forecasting,* **6**, 572.

Newman, S., 1991: Comments on "medium-range model guidance and the 3–5-day extended forecast. *Wea. Forecasting,* **6**, 571.