

MODEL COMPARISON FOR 60 HOURS TO 6 DAYS

BY JERROLD A. LA RUE, RETIRED-NATIONAL WEATHER SERVICE

CONTENTS: [ABSTRACT](#), [1. BACKGROUND](#), [2. PROCEDURE](#), [3. FIRST HALF](#), [4. SECOND HALF](#), [5. ERRORS, ERROR PAGE \(FIGURES 1 & 2\)](#), [6. DISCUSSION](#), [6.1 TABLES 1 & 2](#), [6.2 FIGURES 3 & 4](#), [6.3 FC SCORES](#), [6.4 FIGURE 5](#), [7. CONCLUSIONS](#), 8. BIOGRAPHY

ABSTRACT. Numerical forecasts have improved in the intermediate forecast range such that public forecasts can be considered useful and reasonably reliable up to a week in advance. Most complete public forecasts contain detailed information for zero to 48 hours with fairly specific information extending another three, to as long as seven, days. Private meteorologists including media meteorologists have offered these extended forecasts for some time. More recently the National Weather Service (NWS) added an extended period of three to five days to all public forecasts issued and has plans to increase this to seven days. This has led meteorologists to scrutinize and evaluate the longer range numerical forecasts produced by a number of sources. Most forecasters probably consider the 500 mb numerical forecasts to be the single most important ingredient in producing the extended forecasts. There are comparative verifications of the different longer range numerical models but most of these are on a global scale and use bias and statistical evaluations such as Root-Mean-Square-Error, Standard Deviations and Pattern Anomalies. This paper uses a more meaningful parameter to forecasters to measure the worth of the different numerical models by comparing the 500 mb gradients. Rather simple formulae are developed to show the degree of differences in numerical models. The results are presented for approximately 100 cases, each involving four forecasts. This is admittedly a rather small sample but the results are consistent and are probably representative. Further testing, using this comparative system, will be undertaken this winter. References for the paper can be pursued by following links in the list of Internet sites that present comparative model verification data in "[verification references](#)".

1. BACKGROUND. The public forecast cycle is dictated largely by the activities of the general population, availability of data including numerical forecasts and the methods of disseminating the forecasts. There are two basic forecast cycles which are dictated mainly by the availability of observational data and numerical forecasts and the needs of the populace. These are an early morning forecast, as issued by the NWS, before 5 AM Local Time (LT) and a late afternoon forecast prepared before 5 PM LT. Dissemination of forecasts by TV meteorologists are subsequent to the above times. The NWS updates the basic forecasts as needed, but the media meteorologists must update the noon forecast, at least to eliminate wording related to the morning period and in the late evening forecast to eliminate wording referring to the late afternoon/evening period. The NWS issues an interim Area Forecast Discussion at mid morning and early evening describing the weather situation and explaining any updates that may be issued. The basic forecast cycle is in a preparatory stage at least 2 hours

prior to issuance times. Forecasters will usually wait as long as possible for any delayed numerical forecasts.

The forecasts issued during the mid to late afternoons are the ones that add another day to both the short term and extended forecasts. Day 3 of the extended is moved up to Day 2 in the short term forecast, and another day is added to the extended forecast. The forecasters may use numerical models based on 1200 UTC data, especially for Day 3. Most medium range models are based on 0000 UTC data only which means, for example, model Day 6 is guidance for new forecast Day 5. The early morning forecasts usually have all of the numerical models based on 00 UTC data available by preparation time and it is towards this forecast sequence that this study is

directed. (UP TO ([START](#)))

2. PROCEDURE. Numerical forecasts, from 60 hours to 6 days, available to forecasters for the early morning forecast (2 - 5 AM LT) were chosen for comparison. The models considered for the study were those that were frequently mentioned in the Area Forecast Discussions issued by NWS offices. They also had to be available routinely on the Internet, preferably from more than one source. Those chosen initially were the Medium Range Forecast (MRF) issued by the NWS, the Environmental Canada's Global Environmental Multi-Scale Model (GEM but referred to herein as GLOBAL), the European Center For Medium Range Weather Forecasts (ECMWF but referred to as EURO) and the U.S. Navy's Operational Global Atmospheric Prediction System (NOGAPS). The United Kingdom Model (UK) was available only to 72 hours, so after 15 days the Ensemble Model (ENS) utilizing the MRF model was substituted. The Internet addresses used to access the models is given in [MODEL ADDRESSES](#).

The east-west 500 mb gradient across the U.S. is a measure of the location and intensity of long and short wave features. Forty degrees north latitude was chosen as the east-west axis since it passes pretty much through the center of the U.S. Longitude intersections at 15 degree intervals were chosen at 75 West, 90 West, 105 West, 120 West and 135 West. These points are fairly close to Philadelphia (PHL), St Louis (STL), Denver (DEN), Reno (RNO) and about 550 miles west of Eureka, CA (PAC). A more eastern point in the Atlantic at 60 West was initially selected but that point was not included on some forecast charts. (UP TO ([START](#)))

Each day, 500 mb heights (in decameters) forecast by each model for the 4 forecasts were collected for the 5 points. The sum of the differences in heights between the point near PHL and that near STL, between STL and DEN, between DEN and RNO and RNO and PAC was the forecast gradient, (G_f), for each forecast projection. (Not that it makes any difference in the study, a minus gradient would produce a northerly wind flow and a plus gradient, a southerly flow). The Aviation Model (AVN) initial analysis was used for observed heights and were mostly read from the Edwards AFB MRF charts (address listed in [Model Addresses](#)). The sum of the differences in observed

heights at adjacent grid points was the observed gradient (G_o). The amount that G_f and G_o are the same, when of the same sign, is the gradient correctly forecast (G_c); that is if the sign of ($G_f = \text{sign}(G_o)$), then $G_c = \text{smaller of } (\text{abs})G_f \text{ or } (\text{abs})G_o$. If the signs of (G_f) and (G_o) are different, then (G_c) equals zero. The gradient error (G_e) is the total of the error in the forecast gradient and the observed gradient; $G_e = (\text{abs}(G_f) - G_c) + (\text{abs}(G_o) - G_c)$, (1). The Forecast Correct (FC) is a measure of how much of the observed gradient was correctly forecast. FC is the ratio of the gradient correctly forecast to the observed gradient or $FC = G_c / \text{abs}(G_o)$, (2). This is a modified S_1 (skill score) and can be multiplied by 100 to give percent. FC equals 1 (100%) if G_o is correctly forecast, (a perfect forecast), and is zero if none of G_o is forecast correctly, (the worst forecast). When FC is as low as .5 (50%) it is probable that the forecast is not useful as a forecast tool. An additional computation to measure the amount the forecast was in error, Forecast Error (FE), was developed as $FE = G_e / \text{abs}(G_f)$, (3). FE, equals zero, (perfect forecast), if FC equals 1 and G_f equals G_o . FE values can exceed 10, and in general are lower in cases when the G_f is over forecast and higher when G_f is under forecast. When $G_e = (\text{abs})G_f$, FE is 1.00, which may indicate the point at which forecasts are of near zero usefulness.

UP TO [BACKGROUND], [PROCEDURE], [START])

3. FIRST HALF. Models chosen for the study were the MRF, the NOGAPS, the GLOBAL, the UK, and the EURO. The EURO was based on 12 UTC data and its 3, 4, 5, and 6 day forecasts verified at 1200 UTC. The 60 hour, 84 hour, 108 hour and 132 hour (2 1/2, 3 1/2, 4 1/5 and 5 1/2 day) forecasts of the other models, all verifying at 1200 UTC, were used for comparison. This placed the EURO at a 12 hour time disadvantage which was corrected throughout the study using a forecast decay curve. The Canadian GLOBAL Model did not produce a 132 hour forecast and it had to be interpolated by averaging the 120 hour and the 144 hour forecasts. The UK model was available only at 72 hours, and, after two weeks the ENS, a model based on the MRF, but with up to 17 differing initial analyses, was substituted. Large differences were noted in some ENS gradients for the same forecast and this was due to the ENS data being gathered from any of three Internet addresses. The number of members making up the ENS runs differed greatly between addresses. Because of these difficulties, the First Half of the project was terminated with approximately 50 days of data gathered. Dates for the First Half were March 1 to April 23, 2000.

4. SECOND HALF. The Second Half forecasts included the MRF, the NOGAPS, the GLOBAL and the EURO. All verified at 00 UTC except the EURO and heights for the EURO were interpolated from the forecasts verifying at 12 UTC producing a forecast

verifying at 00 UTC. This reduced the forecast periods for the EURO to 3, 4 and 5 days. It also placed the EURO at a time disadvantage which was corrected using a time decay curve to make it comparative with the other models. Two Ensemble forecasts were included, the ENS12 and the ENS17. The number indicates the number of MRF runs from different initial analyses. The ENS12 forecasts used only 00 UTC data and was mostly collected from the Edwards AFB address. The University of Utah address was used six times when the Edwards AFB forecast was missing, and these contained only 4 to 6 members per forecast. The ENS12's Day 3 forecasts were missing during most of the Second Half and values were interpreted from the 60 hour and the 84 hour forecasts. Dates for the Second Half were April 24 to June 15, 2000. ([UP TO](#) [\[BACKGROUND\]](#), [\[PROCEDURE\]](#), [\[START\]](#))

5. ERRORS. Several sources of systematic errors were inherent in the methodology used in this study. It is doubtful if all of these could be eliminated or even alleviated. Some have been alluded to earlier, and these along with other errors are listed below and are described in detail in a separate page named [ERROR PAGE](#)

1. TIME DISADVANTAGE: There was a time disadvantage in the EURO comparison due to its being based on 1200 UTC data and issued 12 hours prior to the other models which are based on 0000 UTC data and hence issued 12 hour later.

2. FORECAST AVERAGING: Forecast verification times were not always those required by the comparative study. For example the GLOBAL model does not have a 132 hour 500 mb output which necessitated interpolating from the 120 hour and 144 hour forecasts.

3. INTERPOLATION: Estimating values from charts using 6 decameter contours produced errors. Maximum errors were probably 1 decameter over the U.S. and up to 2 decameters from RNO to PAC.

4. MISSING FORECASTS: A very few forecasts were missing during the study. Replacing the UK model with the ENS model reduced the number of ENS cases from approximately 50 to 38.

5. DIFFERENT DATA SOURCES. The ENS forecasts in the First Half were gathered from different Internet addresses. It was discovered that each source used different numbers of initial situations.

It is important that the reader should refer to the [ERROR PAGE](#) in order to weigh the value of the data and conclusions.

([UP TO](#) [\[BACKGROUND\]](#), [\[PROCEDURE\]](#), [\[FIRST HALF\]](#), [\[START\]](#))

6. DISCUSSION. The main synoptic situation during the First Half was a series of cut-off upper systems diving south along the West Coast and in due time exiting to the northeast with decreasing amplitude. The high wave number in the West probably

averaged around 12 which is synonymous with a difficult forecast situation for numerical forecasts. This synoptic pattern continued into the Second Half but with less frequency and intensity. It should be noted, though, the total observed gradient decreased only by 10 percent from the First Half to the Second Half despite the change in seasons.

6.1 Table 1 is a summary of data for the first half and TABLE 2 is for the second half.

TABLE 1

		60 HOUR				84 HOUR				108 HOUR			
	NO	GRD OBS	GRD FCST	GRD CRCT	GRD ERR	GRD OBS	GRD FCST	GRD CRCT	GRD ERR	GRD OBS	GRD FCST	GRD CRCT	GRD ERR
MRF	51	1793	1829	1418	786	1821	1915	1312	1112	1842	1877		
FC & FE				0.791	0.430			0.720	0.585				
ENS[@]	38	1301	1287	951	686	1335	1158	813	867	1379	1158		
FC & FE				0.731	0.533			0.609	0.749				
GLOBAL	50	1793	1662	1244	967	1821	1672	1099	1295	1842	1685		
FC & FE				0.694	0.582			0.604	0.784				
EURO	50	1760	1879	1248	1143	1787	1734	1120	1281	1787	1742		
FC & FE #				#0.75	#0.542			#0.67	#0.674				
NOGAPS	50	1767	1777	1296	952	1782	1768	1037	1476	1758	1765		
FC & FE				0.733	0.536			0.582	0.835				

GRD=GRADIENT; FCST(S)=FORECAST(S); OBS=OBSERVED; CRCT=CORRECT; ERR=ERROR; FC=FORECAST CORRECT

* ASTERISK INDICATES DATA WAS AVERAGED. # INDICATES DATA WAS CORRECTED FOR TIME DIFFERENCES. @ INDICATES INCONSISTENT DATA FROM 3 SOURCES.

TABLE 1. FIRST HALF SUMMARY OF DATA FROM MARCH 1 TO APRIL 15, 2000.

This Table gives the sums of the gradients observed, forecast, correct and in error for each of the forecast times and for each model. The Forecast Correct (FC) is the number in the "GRD CRCT" columns and in the "FC & FE" rows, while the Forecast Error (FE) is under the "GRD ERR" columns and in the "FC & FE" rows. The MRF and the NOGAPS tended to over forecast the gradient, especially in the 132 hour forecast and this could produce a higher Forecast Error (FE) score. The ENS under forecast the gradient, probably due to it being the mean of up to 17 forecasts. The GLOBAL also under forecast the gradient, and this was because of its tendency to be progressive with all short wave features producing a more zonal flow in time.

TABLE 2

		3 DAY			4 DAY				5 DAY			
	NO	OBS GRD	FCST GRD	CRCT GRD	GRD ERR	OBS GRD	FCST GRD	CRCT GRD	GRD ERR	OBS GRD	FCST GRD	CRCT GRD
MRF	52	1654	1640	1282	730	1583	1555	1136	866	1539	1561	
FC & FE				0.775	0.445			0.718	0.557			
ENS12	51	1647	*1354	*1108	*760	1566	1427	1050	893	1521	1259	
FC & FE				*0.673	0.561			0.670	0.626			
ENS17	50	1594	1400	1085	824	1517	1346	902	1059	1484	1275	
FC & FE				0.681	0.589			0.595	0.787			
NOGAPS	51	1618	1482	1154	792	1533	1397	957	1016	1506	1427	
FC & FE				0.713	0.534			0.624	0.727			
EURO	51	1627	*1406	*1100	*833	1558	*1346	*984	*948	1521	*1484	
FC & FE				*0.676	*0.592			*0.632	*0.704			
FC & FE#				#0.72	#0.526			#0.67	#0.638			#0.5
GLOBAL	51	1654	1347	1015	971	1583	1328	937	1037	1539	1227	
FC & FE				0.614	0.721			0.592	0.781			

GRD=GRADIENT; FCST(S)=FORECAST(S); OBS=OBSERVED; CRCT=CORRECT; ERR=ERROR; FC=FORECAST CORRECT; FE=FORECAST ERROR.

EURO DATA MARKED WITH "#" ARE CORRECTED FOR THE FORECAST TIME FACTOR. ALL DATA MARKED WITH "*" ARE AVERAGED FORECASTS OR DERIVED FROM THOSE AVERAGES.

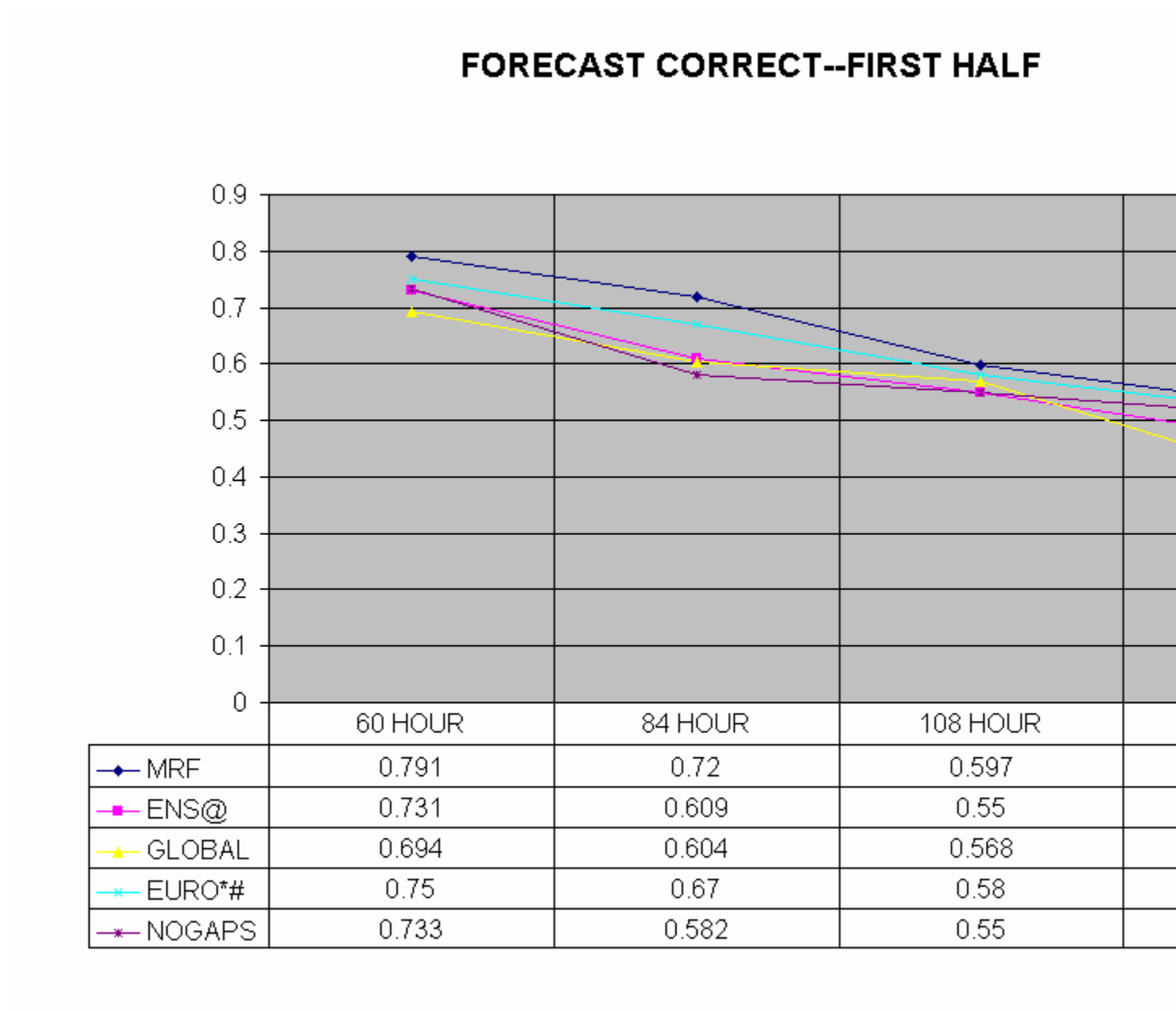
TABLE 2. SECOND HALF SUMMARY OF DATA FROM APRIL 24 THROUGH JUNE 14.

Note that the total MRF forecast gradient is very close to the observed while all other models under forecast the gradient. Both ENS models are a mean of a number of forecasts and evidently lose detail in the averaging. The global model was quite progressive, moving short waves eastward without complications creating a more zonal flow. The NOGAPS was also rather progressive with some features which caused some loss of amplitude.

(UP TO [\[START\]](#)) [\[BACKGROUND\]](#), [\[PROCEDURE\]](#), [\[FIRST HALF\]](#), [6.1 TABLES 1 & 2](#)

=====
 =====
 6.2 The FC score is the ratio of the gradient correctly forecast to the observed gradient. If multiplied by 100, it is the percent of the observed gradient correctly forecast.. FIGURES 3 and 4 are graphical presentations of the FC scores for the First and Second Half periods.

FIGURE 3.



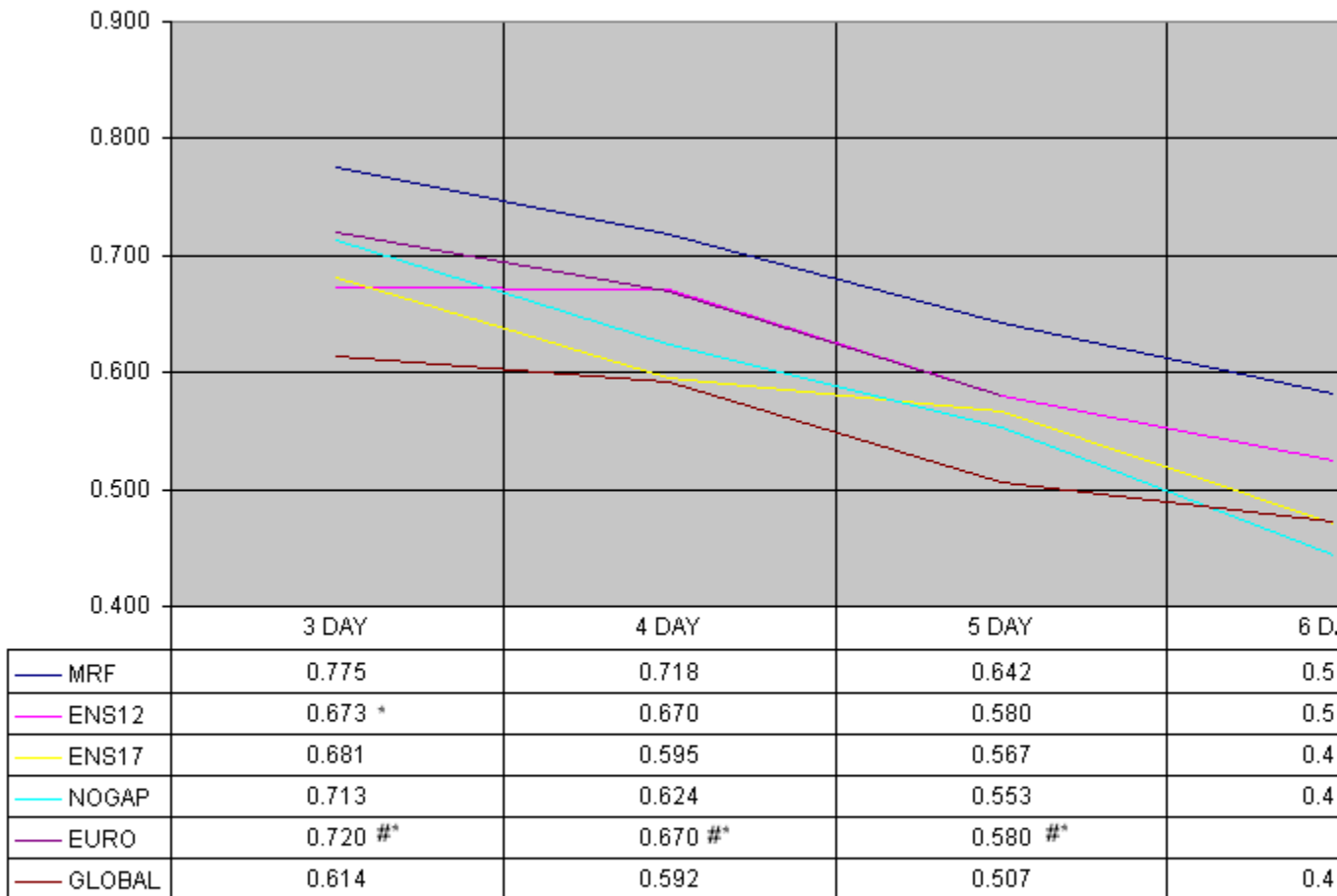
indicates corrected for time differences; * indicates forecasts were averaged; indicates forecasts from three sources;

FIGURE 3. This chart shows the Forecast Correct scores for the First Half of the study.

=====

=====

FIGURE 4



INDICATES CORRECTED FOR TIME DIFFERENCE; * INDICATES FORECASTS WERE AVERAGED

FIGURE 4. FORECAST CORRECT (FC) SCORES FOR THE SECOND HALF.

(UP TO [\[START\]](#)) [\[BACKGROUND\]](#), [\[PROCEDURE\]](#), [\[FIRST HALF\]](#), [6.1 TABLES 1 & 2](#)

=====
=====

6.3 A DISCUSSION OF THE FC SCORES.

The MRF had the highest (best) FC scores in both halves of the study. They were about one day or 24 hours better than the others. The scores were a bit higher in the Second Half. The EURO was second best in both halves despite the need to manipulate the forecasts and statistics to make them comparable, time wise, with the other models. The ENS12 forecast score was a bit low at Day 3 in the Second Half but otherwise tied with the EURO at Days 4 and 5. It was second at Day 6, possibly because the EURO forecasts were terminated at 5 days because of the forecast time difference. NOGAPS averaged about in the middle except the decay rate was high in the Second Half placing its score lowest at Day 6; that trend was not evident in the First Half when it was a close third at Day 6. The GLOBAL model generally had the lowest score at most time periods in both halves of the study.

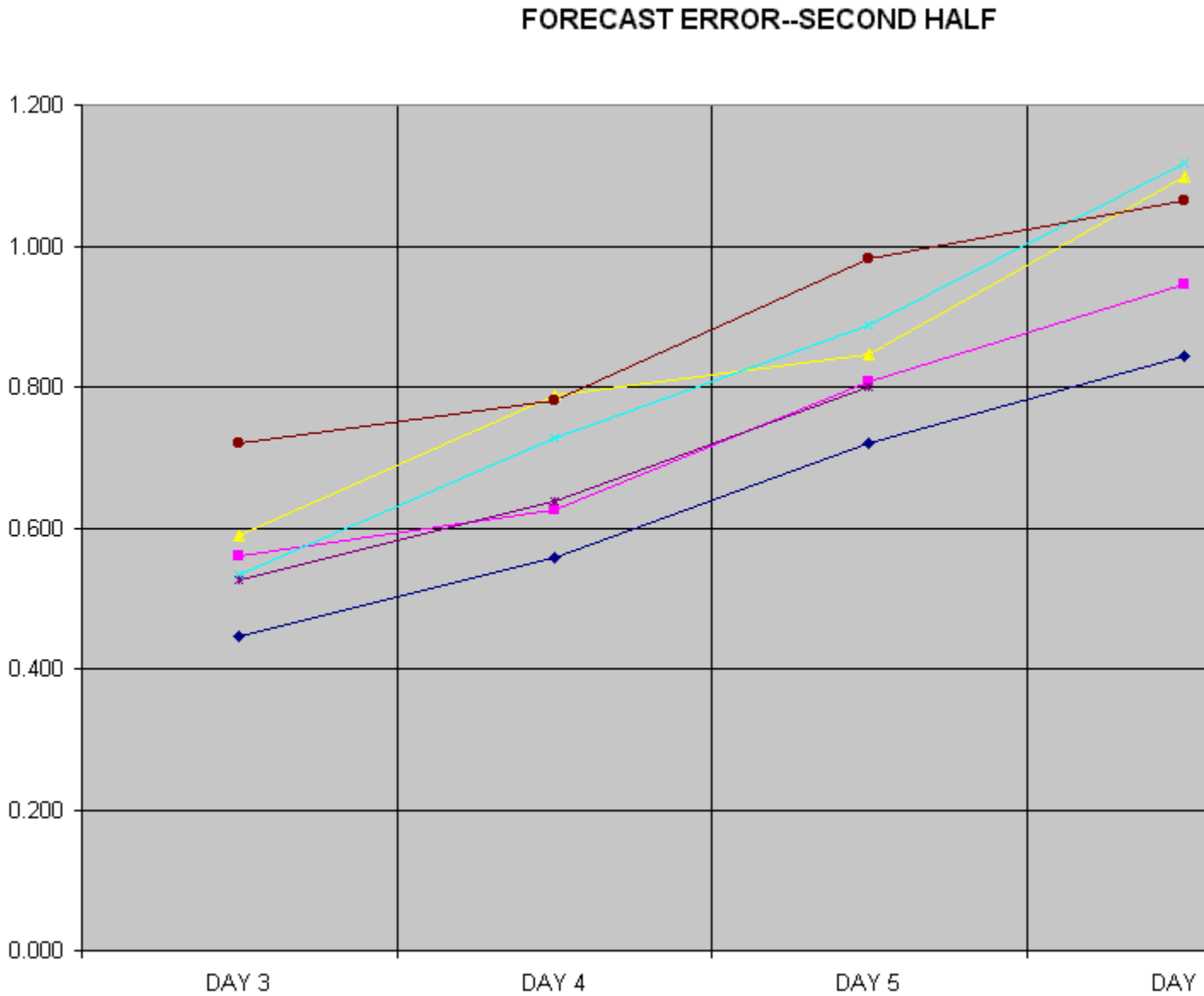
It was expected the ENS model's mean 500 mb values would be superior to an individual MRF forecast, as long term statistical verifications suggested ([VERIFICATION REFERENCES](#)), NCEP, Environmental Modeling Center). However, this was not the case in this study. In the Second Half, ENS12 scored better than ENS17, which was a bit of a surprise as it was thought that the greater number of member forecasts would improve scores. (The six cases in which ENS12 had only 4 to 6 members produced only slightly poorer scores, relatively, than those with 12 members). All of the ENS forecasts under-forecast the gradient, averaging just 85 percent of the gradient that was observed. This really should not be surprising as the mean of many diverse forecasts should be expected to lose detail.

The FC scores were quite high for the first forecast period ranging mainly in the .70's. A score of .75 means that 75 percent of the observed gradient was correctly forecast. The MRF, EURO and NOGAPS scores decreased to .50 by Day 6 in the First Half, while the ENS and Global models reached .50 by about Day 5. In the Second Half, the MRF would probably decay to .50 by Day 7, the ENS12 by Day 6 1/2, the EURO unknown, and the NOGAPS, the ENS17 and the GLOBAL at about Day 5 1/2. An FC score of .50 means that only half of the observed gradient was correctly forecast, and a reasonable conclusion might be that the forecasts were no longer useful.

(UP TO [\[START\]](#)) [\[BACKGROUND\]](#), [\[PROCEDURE\]](#), [\[FIRST HALF\]](#), [6.2 FIGURES 3 & 4](#)

=====
=====

6.4 FIGURE 5.



THE EURO MODEL'S FE SCORE WAS CORRECTED FOR FORECAST TIME DIFFERENCES

FIGURE 5. FE SCORES FOR THE SECOND HALF.

The FE score is the ratio of the total error in the observed gradient and forecast gradient to the forecast gradient. A zero FE score is a perfect forecast. The FE score can be much greater than 1 but a 1.0 FE score means that the forecast gradient equals the sum of the error in observed and forecast gradients. The MRF model had the lowest

(best) FE scores at every forecast time. The ENS 12 and the EURO were nearly tied for second best for Days 3, 4 and 5. The NOGAPS, GLOBAL and ENS 17 reached a FE score of 1.00 at about 5 1/2 days, The ENS12 at about 6 1/2 days and the MRF after 7 days which length of time indicates the end of their usefulness.

(UP TO [[START](#)] [[BACKGROUND](#)], [[PROCEDURE](#)], [[FIRST HALF](#)], [6.1 TABLES 1 & 2](#), [6.2 FIGURES 3 & 4](#), [6.3 FC SCORES](#),

7. CONCLUSIONS. The relative scores for the First and Second Halves were consistent in the comparative placement of numerical forecast models. The MRF should be the model of choice especially when it remains consistent from day to day. If it's solution is bolstered by that of the EURO, then the forecaster should have an increased confidence factor. Other model's concurrence in the solution may tend to bolster the confidence level even higher but to base the forecast on one or more of the other models solution without major agreement with the MRF and possibly the EURO appears risky. The MRF-ENS model did not score as well as the MRF or EURO, but the spaghetti charts which indicate the actual MRF and AVN forecasts may well point out an inconsistency in those forecasts by their position within the array of forecasts. The ENS model is not likely to be used extensively at the present time since its issuance is very late, is rather erratic and the map scale is not suitable for local or regional forecasting. Consistency in the models enhances confidence and the extended forecasts can be very valuable, providing some way can be found to convey that confidence to the user. Forecast Offices of the NWS issue an Area Forecast Discussion which allows the forecaster to give his or her opinion of the extended forecast. TV and radio meteorologists do not usually dwell at any length on the extended period but they can voice some degree of confidence. Perhaps each of the days in the extended forecast should be appended by a degree of confidence like low, middle or high.

A future paper using this verification scheme is planned for this winter. The NWS recently initiated an operational AVN model forecast based on 1200 UTC data that includes forecast periods out to at least 126 hours. The existence of the extended AVN has not been widely advertised. One address for the extended AVN is <http://www.emc.ncep.noaa.gov/forecasts> and another is <http://sgi62.wwb.noaa.gov:8080/STATS/MAPS.html>. Neither of these are monitored operational addresses and the second does not have the map scale and detail that forecasters require. Two other address for the extended AVN are <http://www.met.tamu.edu/weather/mp/avntable.html> and <http://www.edwards.af.mil/weather/avnmodel.htm>. The extended AVN will allow a comparative study of the AVN, the NOGAPS and the EURO all based on 1200 UTC initial data. The MRF would also be included using time decay corrections as in the present study for the EURO. The number of models would be decreased but the area would be expanded to include latitudes 35N, 40N, and 45N which would allow a north-south gradient comparison.

Acknowledgements: The author is indebted to Alan Gerard for his persistence and patience in examining the numerous editorial renditions required. Also, credit should go to super sleuth, technical editor Kevin Lavin, who uncovered an unconscionable boo boo, and to Robert Ricks for his helpful comments.

(UP TO [[START](#)])

[\[BACKGROUND\]](#), [\[DISCUSSION\]](#), [6.1 TABLES 1 & 2](#), [6.2 FIGURES 3 & 4](#), [6.3 FC SCORES](#), [6.4 FIGURE 5](#)

8. BIOGRAPHY

Jerrold A La Rue graduated in meteorology from UCLA. He entered the Weather Bureau at Peoria, IL in 1951, was transferred to Huron, SD in 1953, and to Buffalo, NY in 1955. In 1957 he moved to the National Meteorological Center at Suitland, MD. where he was instrumental in establishing the Quantitative Precipitation Forecast Unit. He was appointed Meteorologist in Charge of the Washington DC Forecast Office in 1970. He was one of the founders of the National Weather Association, it's first President and was Executive Director for five years. He retired from the National Weather Service in 1980.

His E-Mail address is <http://www.nwas.org/ej/larue/jerlarue@sonic.net>