

Model Verification for the 2004-2005 NCWFCJon Moskaitis
June 2, 2005*Introduction*

Here, I compare the performance of the GFS ensemble median high and low temperature forecasts with other available guidance for the 2004-2005 National Collegiate Weather Forecasting Contest (NCWFC). The verification metric I have used for these deterministic forecasts is absolute error (either accumulated absolute error or mean absolute error), as the contest scoring system is based on absolute error.

Types of guidance

Six different types of model-based guidance are verified, along with National Weather Service forecasts. Five of the guidance products are conventional model output statistics (MOS): the GFS ensemble control, the GFS operational¹, the AVN, the ETA, and the NGM. The other model-based guidance product, the GFS ensemble median, is derived from a continuous probability distribution fit to the discrete set of forecast value from the GFS ensemble MOS. If the aforementioned probability distribution was the true forecast probability distribution, the GFS ensemble median would be the deterministic forecast that minimizes expected absolute error.

The model integrations from which the guidance products are derived are initialized at different times relative to the NCWFC forecast interval (06z – 06z). The GFS is initialized 30 hours prior to the beginning of the NCWFC forecast interval (00z), while the AVN, ETA, and NGM are initialized 18 hours prior (12z). The National Weather Service forecasts are made about 10 hours prior to the beginning of the NCWFC forecast interval, and are based primarily on the ‘12z guidance’ mentioned above. Thus, it is the youngest guidance, whereas the GFS-based products are the oldest.

¹ The ensemble control and operational versions of the GFS are integrated from the same analysis, but use different resolutions. The ensemble control uses a lower resolution than the operational, consistent with the rest of the ensemble members.

Forecast set

High temperature guidance for 10 of the 13 NCWFC cities² is considered here. I limited the verification to days when the high temperature had a daytime maximum, as this is the quantity the guidance is designed to forecast. This ended up excluding 7 of the 80 possible forecast days. For low temperature guidance, I verified 9 cities³, excluding the 15 days with afternoon or evening minima. For some cities, as many as half of the forecast days had to be excluded from the low temperature verification. See the bottom of Table 1 for the number of high and low temperature forecasts verified for each city.

Results

Cumulative absolute error and mean absolute error results for the year are shown in Table 1. For high temperature, the GFS and AVN based guidance products all scored very similarly, collectively the poorest of the guidance products. The ETA and NGM scored slightly better, and the NWS forecasts better still. This pattern does not necessarily hold for individual cities, however. The mean absolute error calculated for each city is shown in Figure 1, for all the guidance products. The GFS-based products generally group together, along with the AVN to a point. Relative to these, the ETA and NGM can be either considerably better or worse. Somewhat surprisingly, in only one city (Tulsa) was the 00z guidance considerably worse than the younger 12z guidance.

Low temperature results for the year can also be found on Table 1. The GFS ensemble median and GFS control are again similar, but the GFS operational is noticeably better. Perhaps the higher resolution of the GFS operational is advantageous for low temperature forecasting because it can better simulate boundary layer radiational cooling. The AVN scores even better than the GFS operational, and as the best of the guidance products, is only slightly worse than the NWS. ETA and NGM mean absolute errors are approximately in line with those of the operational GFS. Figure 2 shows the

² The three not included are Cape Hatteras (I was not producing GFS ensemble median at the time), New York City (no GFS ensemble available for KNYC), and Buenos Aires (no guidance at all!)

³ There was no low temperature guidance for Hoquiam.

city-to-city variability in the mean absolute error of the low temperature guidance. Again, there is no clear ‘model of choice’ that is the best for every city. Excluding Glasgow and Kansas City, the GFS ensemble median remained quite competitive with the other guidance products.

For information on the accumulated absolute errors for high and low temperature forecasts at each city, please see Table 2.

Discussion and future plans

The purpose of this verification study was to assess the quality of the GFS ensemble median forecasts relative to the other available guidance products. As anticipated, the GFS ensemble median forecasts were somewhat inferior to their 12z guidance competition. However, I did not anticipate the relationships between the GFS ensemble median and the other 00z GFS guidance products. The verification results for the GFS ensemble median and GFS control are very similar, even on a city-by-city basis. In fact, the similarity between the GFS ensemble median and GFS control is even deeper: the individual forecasts themselves rarely differed by more than one degree.

This correspondence between the ensemble control and median can perhaps be related to the ensemble construction scheme used by the GFS, along with short-term perturbation dynamics. The 11-member GFS analysis ensemble is constructed (i.e., the set of different initial states are made) by adding perturbations to the control analysis along 5 different directions (specified by the bred vectors) in state space. Each perturbation direction gives rise to a pair of ensemble members: in one member, the perturbation is added to the control, and in the other it is subtracted. This process results in 10 perturbed ensemble members, to go along with the 1 control member. In this 11-member distribution, the control is ‘in the middle’ (this can be visualized for a one or two dimensional state space). As the ensemble is integrated forward in time, the control will remain ‘in the middle’ of the distribution as long the ensemble perturbations grow approximately linearly. This is because under linear dynamics, the initial perturbation

pairs will evolve such that they remain of equal magnitude and along the same direction (differing only by sign). Thus, the ensemble control forecast really should approximate the ensemble median forecasts, as long as ensemble perturbation growth is nearly linear. Evidently, this is the case for the GFS ensemble, at least at the 30-54 hr lead time of the NCWFC forecasts. Maybe at longer lead times, nonlinear perturbation growth causes more consistent differences between the GFS ensemble median and GFS ensemble control.

Now, with the GFS ensemble control forecast mirroring that of the median, why would one bother looking at the ensemble distribution any more? There are a few reasons. First, the ensemble distribution will give evidence of nonlinear perturbation growth when it does occur. For the 130 NCWFC temperature forecasts verified here, there were 6 situations where the GFS ensemble control differed from the GFS ensemble median by 2 degrees or more (considerable relative to the absolute error of the forecasts). Viewing the entire ensemble distribution will allow one to recognize these situations⁴, and forecast accordingly. The real value of the ensemble for the contest, though, is risk assessment. If the forecast distribution is tightly constrained, it is probably not a good idea to let one's forecast stray far from the median. On the other hand, should the distribution be highly skewed, perhaps one might take a risk on a forecast in the long tail, hoping for a big win. Deterministic forecasts alone (say, just the 12z AVN and ETA) simply cannot inform these judgments the way an ensemble can. To provide more a more quantitative backing for this, I will produce graphs of expected absolute error (as a function of forecast value) for the coming 2005-2006 NCWFC year, based on the GFS ensemble. This should explicitly inform the user of the risk associated with straying from the ensemble median.

⁴ Since I now color-code (in green) the ensemble control in my graphics package

Table 1: Cumulative Error for 2004-2005 NCWFC

<i>Models</i>	<i>Time</i>	Accumulated Absolute Error		
		<i>High Temp</i>	<i>Low Temp</i>	<i>Total</i>
GFS Ens Median	00z	261	218	479
GFS Control	00z	264	218	482
GFS Operational	00z	259	204	463
AVN	12z	255	190	445
ETA	12z	233	205	438
NGM	12z	226	208	434
Humans				
<i>Models</i>	<i>Time</i>	<i>High Temp</i>	<i>Low Temp</i>	<i>Total</i>
NWS	~20z	199	184	383

<i>Models</i>	<i>Time</i>	Absolute Error per forecast		
		<i>High Temp</i>	<i>Low Temp</i>	<i>Total</i>
GFS Ens Median	00z	3.58	3.82	7.40
GFS Control	00z	3.62	3.82	7.44
GFS Operational	00z	3.55	3.58	7.13
AVN	12z	3.49	3.33	6.83
ETA	12z	3.19	3.60	6.79
NGM	12z	3.10	3.65	6.75
Humans				
<i>Models</i>	<i>Time</i>	<i>High Temp</i>	<i>Low Temp</i>	<i>Total</i>
NWS	~20z	2.73	3.23	5.95

<i>City</i>	Number of forecasts	
	<i>High Temp</i>	<i>Low Temp</i>
GGW	8	6
MSY	8	8
MSP	7	6
HQM	8	0
MCI	6	7
GRR	7	6
BHM	6	4
SFO	8	8
LND	7	7
TUL	8	5
Total	73	57

Table 2: City-by-city accumulated absolute error for each model

GFS Ens Median

City	Accumulated Absolute Error		
	High Temp	Low Temp	Total
GGW	15	38	53
MSY	12	19	31
MSP	22	14	36
HQM	23		23
MCI	49	36	85
GRR	19	50	69
BHM	13	11	24
SFO	36	16	52
LND	35	23	58
TUL	37	11	48
Total	261	218	479

AVN

City	Accumulated Absolute Error		
	High Temp	Low Temp	Total
GGW	26	22	48
MSY	15	21	36
MSP	23	14	37
HQM	14		14
MCI	48	24	72
GRR	20	44	64
BHM	13	14	27
SFO	31	15	46
LND	42	27	69
TUL	23	9	32
Total	255	190	445

3FS Ensemble Contro

City	Accumulated Absolute Error		
	High Temp	Low Temp	Total
GGW	16	38	54
MSY	13	20	33
MSP	22	14	36
HQM	25		25
MCI	48	37	85
GRR	20	49	69
BHM	13	11	24
SFO	34	17	51
LND	35	24	59
TUL	38	8	46
Total	264	218	482

ETA

City	Accumulated Absolute Error		
	High Temp	Low Temp	Total
GGW	32	23	55
MSY	24	23	47
MSP	22	15	37
HQM	16		16
MCI	21	31	52
GRR	31	42	73
BHM	7	13	20
SFO	24	22	46
LND	31	27	58
TUL	25	9	34
Total	233	205	438

GFS Operational

City	Accumulated Absolute Error		
	High Temp	Low Temp	Total
GGW	20	29	49
MSY	15	21	36
MSP	22	13	35
HQM	22		22
MCI	42	28	70
GRR	18	47	65
BHM	12	14	26
SFO	34	16	50
LND	41	27	68
TUL	33	9	42
Total	259	204	463

NGM

City	Accumulated Absolute Error		
	High Temp	Low Temp	Total
GGW	22	27	49
MSY	12	26	38
MSP	27	22	49
HQM	20		20
MCI	38	31	69
GRR	19	43	62
BHM	8	9	17
SFO	23	13	36
LND	34	20	54
TUL	23	17	40
Total	226	208	434

NWS

City	Accumulated Absolute Error		Total
	High Temp	Low Temp	
GGW	19	29	48
MSY	19	20	39
MSP	22	12	34
HQM	16		16
MCI	22	21	43
GRR	14	35	49
BHM	13	14	27
SFO	24	15	39
LND	34	27	61
TUL	16	11	27
Total	199	184	383

Figure 1

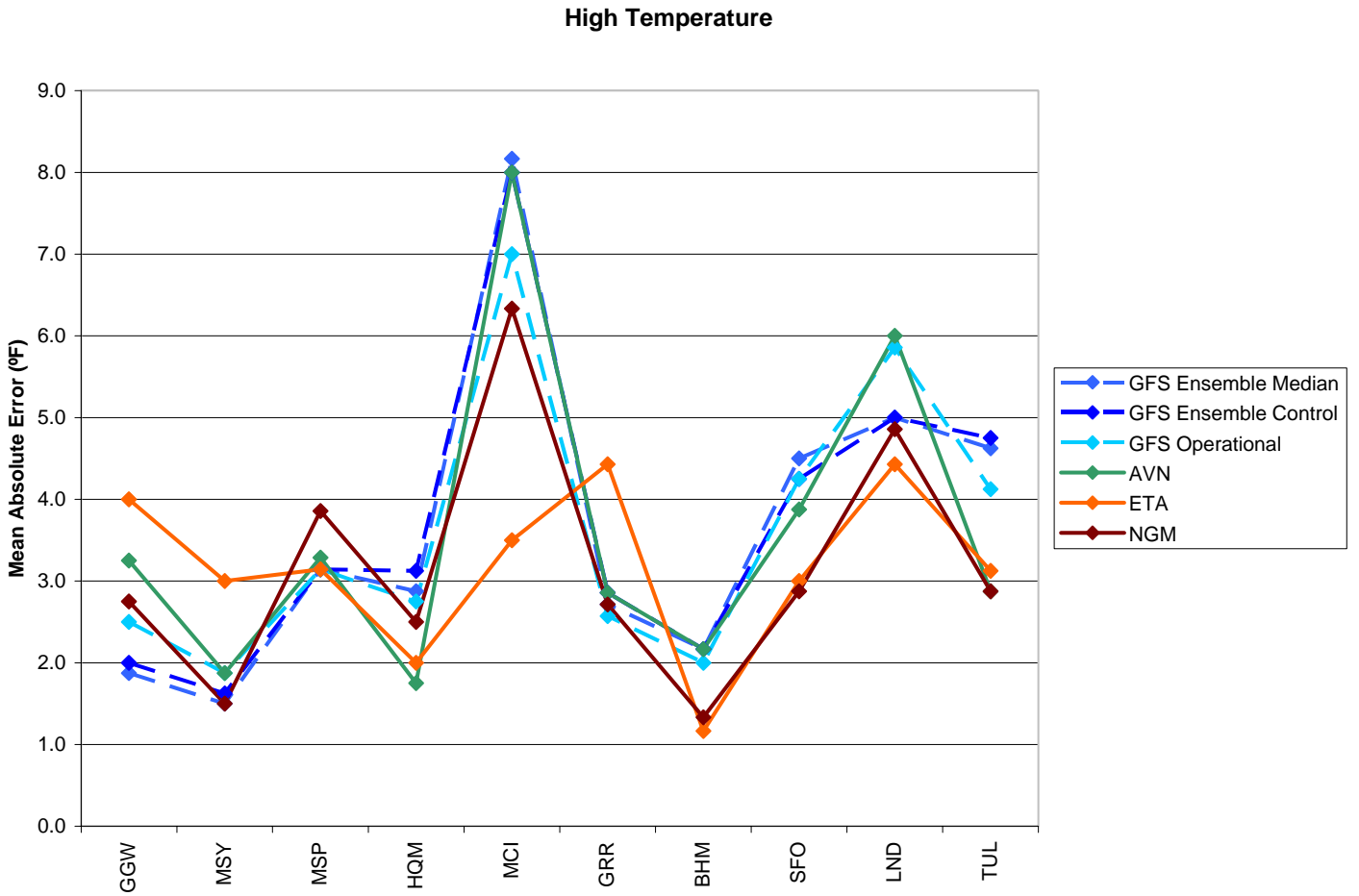


Figure 2

