J 1.8

MODEL CONSENSUS AND ENSEMBLE WEIGHTING FOR SPOT FORECASTS

Brian J. Etherton
University of North Carolina, Charlotte

1 - INTRODUCTION

The tools of numerical guidance and model consensus have been staples of weather forecasting for a number of years. Averaging different numerical forecasts to produce a 'consensus' forecast has been shown to yield forecasts more accurate than any individual element of the consensus forecast (Fraedrich and Leslie, 1987; Verret and Yacowar 1989; Vislocky and Young, 1989). Vislocky and Fritsch (1995) showed that a combination of two MOS products produced a forecast more accurate than either MOS forecast independently. Further, Vislocky and Fritsch (1995) show that a weighted average of two MOS products, with weights based on average forecast errors of the two products, performs better than a straight 50/50 blend. It has also been shown that a 'consensus' of direct model output produces more accurate forecasts than any single forecast model (Fritsch et al., 2000) for upper level fields.

This paper explores the following hypothesis: a guidance product made from a consensus of two MOS forecasts and a direct model output, dubbed "BJEGUI", produces more accurate forecasts of temperature and precipitation amount than a forecast made from only the two MOS forecasts. As a test of the skill of BJEGUI against other guidance products and against human forecasters, in the spirit of Vislocky and Fritsch (1997), the product was entered as a competitor in the National Collegiate Weather and Forecasting Contest (NCWFC). The details of how BJEGUI is constructed are the subject of section 2. Section 3 reports on how the guidance product performed against other guidance products and against human forecasters.

2 - METHODS

Guidance for the forecast of high temperature, low temperature, and precipitation amount is made from a weighted average of ETA MOS, AVN MOS, and SREF model output. This guidance product was dubbed 'BJEGUI'. Forecasts consist of the maximum and minimum temperatures and a categorized amount of precipitation that will occur during the following day. Temperatures are forecast to the nearest whole degree Fahrenheit.

Corresponding Author Address:
University of North Carolina-Charlotte
Department of Geography and Earth Science
9201 University City Blvd
450 McEniry Building
Charlotte, NC 28223

Precipitation forecasts are a category which spans a range of values. These categories are shown in Table 1. Forecasts had to be made by 00Z, with forecasts valid for the period from 06Z (6 hours after the forecasting time) until 06Z the following day. The scoring of daily forecasts is as follows: as one point per degree Fahrenheit for temperatures, 4 points per category for precipitation.

Table 1

CATEGORY	RANGE
0	0 – Trace
1	Trace - 0.05
2	0.06-0.24
3	0.25-0.49
4	0.50-0.99
5	>= 1.00

Table 1 – The categories for NCWFC precipitation amount forecasts, and the ranges these categories span. Values are in inches.

Forecasts were made for the following cities: Miami, FL, Corpus Christi, TX, Fargo, ND, Burlington, VT, Cheyenne, WY, Seattle, WA, Memphis, TN, Buffalo, NY, Ontario, CA, Cincinnati, OH, Des Moines, IA, Little Rock, AR, and Boston, MA. For the NCWFC, each of these cities (except Miami) was the 'forecast city' for a 2-week period. During that period, 8 different daily forecasts were made. For the purpose of this paper, forecasts were produced for each city every day, providing a larger sample size by which to judge BJEGUI against other guidance products.

(a) The first two components of BJEGUI - ETA MOS and AVN MOS values.

Producing temperature and precipitation forecasts from MOS was primarily an exercise in decoding. The 12Z low temperature and 00Z high temperature forecast by a MOS product within the 06Z to 06Z forecast period were used as the MOS low and high temperature forecasts, unless any of the instantaneous temperature forecasts (valid at 3 hour intervals) within the period were outside those bounds, in which case the high or low temperature forecast was adjusted to match that temperature.

Producing a quantitative precipitation forecast from MOS guidance required interpolation. MOS does not provide explicit precipitation forecast amounts, instead MOS provides a categorical quantitative precipitation forecast (QPF), giving a range of possible precipitation amounts during a time period. To obtain the 24-hour total precipitation forecast, the four 6-hour MOS QPF forecasts valid within the

24-hour forecast period are assigned values, and those values are summed. Table 2 lists the MOS categories, ranges, and values used for addition. These category values were assigned ad-hoc, without significant study of the optimal values to use.

Table 2

CATEGORY	RANGE	ASSIGNED VALUE
0	0.00	0.00
1	0.01-0.09	0.05
2	0.10-0.24	0.15
3	0.25-0.49	0.34
4	0.50-0.99	0.65
5	>= 1.00	1.05

Table 2 – The categories for 6-hour precipitation amount forecasts from MOS, the ranges these categories span, and the value used for this category. Values are in inches.

(b) The third component of BJEGUI – the SREF Ensemble Mean

The Short Range Ensemble Forecast (SREF) provides the third input to BJEGUI. SREF (Tracton et al., 1998; Du and Tracton 2001) consists of 5 runs of the ETA model (Black 1994) and 5 runs of the regional spectral model (RSM) (Juang and Kanamitsu 1994). SREF has recently been expanded to include 5 additional runs using the ETA model with a different scheme for the parameterization of convection, Kain-Fritsch (Kain and Fritsch, 1993) rather than Betts-Miller-Janjic (Betts 1986; Betts and Miller 1986; Janjic 1994). These 5 Kain-Fritsch members were not available in real-time through the course of the experiments however, and so were not used in BJEGUI. Two sets of ensemble forecasts are generated daily, one initialized at 09Z and the other initialized at 21Z the previous day, both providing forecast fields every 3 hours up to 63 hours after their initialization. By using the two most recently available sets of forecasts, a 20-member lagged average forecast (Hoffman and Kalnay 1983) ensemble (10 from 09Z, 10 from 21Z yesterday) was produced.

To produce maximum and minimum temperature forecasts from each SREF ensemble member, a cubic spline is used to interpolate between the instantaneous 2-meter temperature forecast values (provided in 3 hour intervals), to get the high and low temperature forecasts. The quantitative precipitation forecast for each SREF member is made by simply summing the eight 3-hour precipitation amount forecasts spanning the 24 hour forecast period.

The SREF ensemble mean is the average of the 20 SREF members. Ensemble mean forecasts of high temperature, low temperature, and precipitation category forecasts calculated daily are saved and compared to verifying values at each forecast city daily. For each city, the average error for the high and low temperature forecasts over the previous 30 days serves as the bias of each forecast. These biases are subtracted from the SREF ensemble mean value to produce a corrected SREF ensemble mean forecast for use in BJEGUI. No correction is made to the precipitation forecast, as the large number of days without precipitation makes this adjustment less reliable.

(c) Averaging ETA MOS, AVN MOS, and SREF values.

To combine the 3 forecasts into one, weighting values were created using the mean squared error of each forecast over the past 30 days. Values of mean squared error, σ^2, are computed for each forecast parameter for each city. As derived from minimum error variance theory (Daley, 1997), weightings for three forecasts, “a”, “b”, and “c”, with mean squared errors denoted as σ^2_a, σ^2_b, and σ^2_c the forecasts, are:

$$w_a = (\sigma^2_b\sigma^2_c)/(\sigma^2_a\sigma^2_b+\sigma^2_a\sigma^2_c+ \sigma^2_b\sigma^2_c)$$
$$w_b = (\sigma^2_a\sigma^2_c)/(\sigma^2_a\sigma^2_b+\sigma^2_a\sigma^2_c+ \sigma^2_b\sigma^2_c)$$
$$w_c = (\sigma^2_a\sigma^2_b)/(\sigma^2_a\sigma^2_b+\sigma^2_a\sigma^2_c+ \sigma^2_b\sigma^2_c)$$

Weighting values are calculated for each forecast parameter (maximum temperature, minimum temperature, precipitation amount) for each forecast site. The weighting approach is similar to that of Vislocky and Fritsch (1995). However, they created one weighting value for all forecasts, instead of calculating a weighting for each forecast parameter for each forecast city. Another difference is that Vislocky and Fritsch (1995) used a full year of data to determine their weightings, rather than only the last 30 days. While using a full year of data produces a more robust scheme, it does not allow for seasonal differences. It is possible that one component, for example AVN MOS, provides more accurate forecasts of a parameter in winter than the other components of BJEGUI, but provides less accurate forecasts in the summer.

3 - RESULTS

Forecasts were made daily from July 1, 2002 until April 30, 2003 for all 13 cities listed in the previous paragraph. From July 1 to February 1, BJEGUI was in development. During that time experiments as to how many ensemble members to use and how to weight the 3 forecasts components (ETA, AVN, and SREF) were underway. Only forecasts starting on February 1, 2003 used the configuration discussed in section 2, and so values from February 1 through May 3 will be presented in subsections (a) and (b). During the experiments, there were days with incomplete data, approximately two days per month. Those days have been excluded from the results.

Precipitation forecast verification proved difficult to achieve. During the academic year, there were days of snowfall. If an automated observing site did not have its observation augmented, the precipitation verification values could be in error. Given this uncertainty, results shown in subsection (a) will be for temperature forecasts only.

(a) – Results for all cities (Feb 1 – May 3)

The average daily error is calculated by dividing the total number of temperature error points from all 13 forecasts city by 13. Results, shown in Figure 1, indicate BJEGUI has a lower average error than any individual MOS product. BJEGUI fared a little better than a weighted average of ETA and AVN MOS products, showing the benefit of incorporating SREF data in the forecast product, however these improvements were only statistically significant at the 30% level.

Figure 1

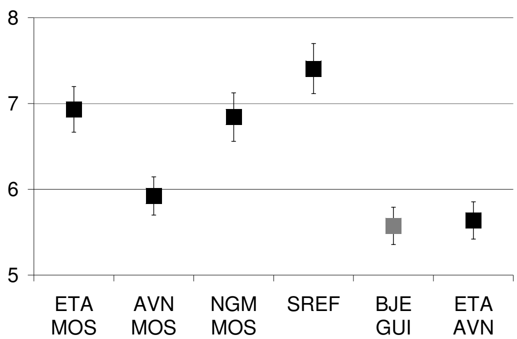

Figure 1 -- Average daily error over all 13 forecast sites of from February 1 until May 3. Error bars are the 95% confidence intervals for the estimate of the mean.

Results for those forecasts which are an average of 2 or more individual forecasts (Figure 2) show the benefits of having weights assigned to each component of that forecast, rather than a pure mean of the forecasts. Figure 2 shows the average errors for forecasts produced by weighting ETA, AVN, and SREF forecasts using the weightings described previously, and the errors from adding them up and dividing by 3. The errors are smaller for the weighted blend, but the results are only statistically significant at approximately the 30% level. The significance was higher in February than in March and April. More statistically significant were improvements from applying the weighting to a blend of ETA and AVN forecasts, in which case improvements over a pure mean were significant at nearly the 70% level. This result suggests that the more members of a consensus forecast, the less important weighting those members is.

Figure 2

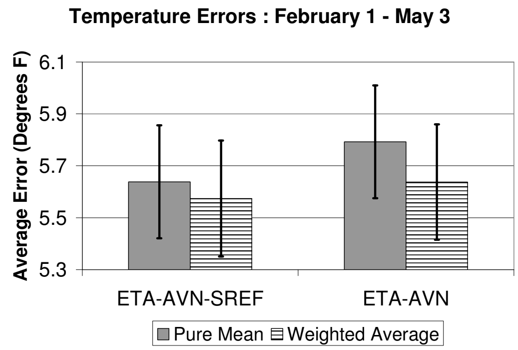

Figure 2 – Average daily error over all 13 forecast sites from February 1 until May 3. Error bars are the 95% confidence intervals on the estimate of the mean.

(b) – Results from the National Collegiate Weather Forecasting Contest

In the spirit of Vislocky and Fritsch (1997), BJEGUI participated in the National Collegiate Weather Forecasting Contest (NCWFC) during the 2002-2003 academic year. BJEGUI competed in the cities of Fargo, Burlington, Cheyenne, Seattle, Memphis, Buffalo, Cincinnati, Des Moines, Little Rock, and Boston. Total error points are calculated for all contestants at the end of each forecast period. Results for each forecast city are normalized such that the national consensus forecast, the forecast made by using the average of all forecasters on each day, is assigned a score of 80, and the normalized distribution has a standard deviation of 10. Adding up the normalized scores for each city participated in, and dividing by the number of cities, produces a final score. Lower scores equate with more accurate forecasts.

In addition to human forecasters, NCWFC also has MOS forecast temperatures and the raw model precipitation forecasts from the NGM, AVN, and ETA models entered as contestants. These forecasts were dubbed “GUIDAN”, “AVNGUI”, and “ETAGUI”. The mean of the AVN, ETA, and NGM forecasts, called “BLEND”, was also entered in the contest.

Table 3 shows the year-end normalized scores for various guidance products. BJEGUI had the lowest normalized score of any guidance product, and finished better than 94% of all qualifying human forecasters in the contest. BJEGUI fared significantly better than a blend of the AVN, ETA, and NGM forecasts. This success may be the result of any or all of the following: weighting the various components of the forecast, rather than taking a straight average, using MOS precipitation forecasts rather than direct model output for the ETA and AVN forecasts, using SREF data rather than NGM data.

Table 3

PRODUCT	CITIES	RANK	PCTL.
BJEGUI	10	27	94th
BLEND	12	136	71st
AVNGUI	12	143	68th
ETAGUI	12	271	38th
GUIDAN	13	364	23rd

Table 3 – Final standings from the National Collegiate Weather and Forecasting Contest for 2002-2003. Rank calculated from 480 qualifying forecasters, percentile based on these ranks.

BJEGUI had trouble during transitioning weather regimes, such as from very cold to very warm in early spring. SREF forecasts biases and were calculated and the weightings for the 3 components of BJEGUI were based on 30-day averages, and these values were not as representative during transition periods as during stable regimes.

4 – CONCLUSIONS

The guidance product BJEGUI, formed by a weighted average of ETA MOS, AVN MOS, and SREF model output, produced temperature and precipitation forecasts with a smaller average error than other guidance products. BJEGUI produced more accurate forecasts than a blend of ETA, AVN, and NGM output – showing the value of adding in SREF data and in incorporating MOS based QPF forecasts into the guidance product. BJEGUI was improved by including a bias correction for the SREF ensemble, and by using an error dependent weighting of the ETA, AVN, and SREF forecasts.

5 - REFERENCES

Betts, A. K., 1986: A new convective adjustment scheme. Part I: Observational and theoretical basis. *Quart. J. Roy. Meteor. Soc.,* **112**, 677–692.

Betts, A. K. and M. J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, and arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.,* **112**, 693–709.

Black, 1994: The new NMC mesoscale Eta model: Description and forecast examples. *Wea. Forecasting,* **9,** 265–278.

Dallavalle, J. P., 1996: A perspective on the use of model output statistics in objective weather forecasting. Preprints, *15th Conf. on Weather Analysis and Forecasting,* Norfolk, VA, Amer. Meteor. Soc., 479–482.

Du, J., and M. S. Tracton, 2001: Implementation of a real-time short range ensemble forecasting system at NCEP: An update. Preprints, *Ninth Conf. on Mesoscale Processes,* Ft. Lauderdale, FL, Amer. Meteor. Soc., 355–360.

Fraedrich, K., and L. M. Leslie, 1987: Combining predictive schemes in short-term forecasting. *Mon. Wea. Rev.*, **115**, 1640-1644.

Fritsch, J. M., J. L. Hilliker, J. Ross, and R. L. Vislocky, 2000: Model Consensus. *Wea. Forecasting*, **15**, 571–582.

Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus.,* **35A**, 100–118.

Janjic, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.,* **122**, 927–945.

Juang, H.-M. and M. Kanamitsu, 1994: The NMC nested regional spectral model. *Mon. Wea. Rev.,* **122,** 3–26.

Kain, J. S., and J. M. Fritsch, 1993: Convective parameterization for mesoscale models: The Kain–Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models, Meteor. Monograph.*, No. 24, Amer. Meteor. Soc., 165–170.

Kalman R., and R. Bucy, 1961: New results in linear filtering and prediction theory. Trans. ASME, *J. Basic Eng.*, **82D**, 35-45.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorlogical Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1763.

Tracton, M. S. and J. Du, Z. Toth, and H. Juang, 1998: Short-range ensemble forecasting (SREF) at NCEP/EMC. Preprints, *12th Conf. on Numerical Weather Prediction.* Phoenix, AZ, Amer. Meteor. Soc., 269-272.

Verret, R., and N. Yacowar, 1989: Improvement of numerical weather element forecasts by combining forecasts from different procedures. Preprints*, 11th Conf. on Probability and Statistics*, Monterey, CA, Amer. Meteor. Soc., 58-63.

Vislocky, R. L., and J. M. Fritsch, 1997: Performance of an advanced MOS system in the 1996–97 national collegiate weather forecasting contest. *Bull. Amer. Meteor. Soc.,* **78**, 2851–2857.

Vislocky, R. L., and J. M. Fritsch, 1995: Improved Model Output Statistics Forecasts through Model Consensus. *Bull. Amer. Meteor. Soc.*, **76**, 1157–1164.

Vislocky, R. L., and G. S. Young, 1989: The use of perfect prog forecasts to improve model output statistics forecasts of precipitation probability. *Wea. Forecasting*, **4**, 202-209.