4.8 ADAPTIVE DATA FUSION OF METEOROLOGICAL FORECAST MODULES

Shel Gerding*and Bill Myers
National Center for Atmospheric Research, Boulder, CO, USA

1. INTRODUCTION

Many excellent forecasting models are currently available for domestic U.S. and international sites. The models each have inherent strengths and weaknesses. By combining the forecasts using adaptive, confidence-weighted fusion, improvements in forecasts have been made in long- and short-term forecasting systems for a variety of applications.

The method for adaptive data fusion was chosen for its low computational complexity, accuracy, flexibility, numerical stability, and ability to use confidence values associated with the input forecasts. These characteristics are demonstrated here through comparisons with other data fusion techniques on forecasts of meteorological data.

The techniques chosen to compare with the adaptive data fusion (ADF) algorithm are the Newbold-Granger method (N-G) (see Newbold (1974)), nonnegative restricted least squares (NRLS) (see Aksu (1992), for example), and principal components regression (PCR) (see Jackson (1991), for example). The adaptive data fusion algorithm was compared to several other algorithms in earlier studies (see Young (2002)), and the algorithms presented here were found to be the most promising for the meteorological problem presented. These algorithms also provide a good sampling of the types of algorithms available for this problem. The N-G method is a computationally simple algorithm that uses the error statistics. The NRLS is a least squares algorithm, like the ADF algorithm, but uses a set window of data, instead of iterative adaptation. PCR is a linear algebraic algorithm that also uses a window of data.

The algorithms are evaluated over a range of tuning parameters and a variety of data. The ease to which the algorithm can be turned for a data set is an important criterion. The forecast models produce a large number of variables for a large number of sites and forecast times. Some sites and forecast times have a large amount of training data, and others have very little. For this study, maximum temperature and wind speed are the two meteorological variables examined. For both variables, the forecast for one day and the forecast for 7 days in the future are examined. The data was taken over the period of January 1, 2002 to August 31, 2002 for eighteen domestic sites. The meteorological models used were the AVN, ETA, MRF and NGM. Climatology is also used in the fusion. Truth was obtained from synoptic measurements. Confidence values are not used in this study but are used in the systems that use the ADF algorithm.

* *Corresponding author address*: Shel Gerding, Research Applications Division, National Center for Atmospheric Science, P.O. Box 3000, Boulder, CO 80307-3000; e-mail: shel@rap.ucar.edu.

2. ALGORITHMS

The fused forecast is created by summing the weighted, bias corrected input forecasts. For variable, f, forecast, X_f, input forecasts, X_{jf}, with associated confidences, c_{jf}, weights, w_{jf}, and biases, b_{jf}, the equation is

$$X_f = \frac{\sum_{j=1}^{n} \left[w_{jf} c_{jf} \left(X_{jf} + b_{jf} \right) \right]}{\sum_{j=1}^{n} w_{jf} c_{jf}}. \qquad 1)$$

To simplify the analyses, the confidences are assumed to be one for all of the data in this study. Only the ADF algorithm uses the denominator in the weight determination formula. In all other algorithms, the denominator is assumed to be one. The advantage in having the denominator is that the weights do not unduly bias the results when input forecasts are missing.

The best weights and biases for a given set of data depends on the criteria for success. In this study, root mean-squared error (RMSE) is used to evaluate the results. The algorithms all have the goal of minimizing RMSE for stationary systems. The data used in this study is typical of weather forecast data in its non-stationarity.

2.1 *Adaptive Data Fusion (ADF)*

The ADF algorithm is based on the gradient decent of the confidence-weighted error, with the weights restricted to summing to one. The algorithm is loosely based on the fuzzy standard additive model (SAM) of Bart Kosko (1997). The following Lyapunov function is used with the truth for the forecast, X_{Tf}, and its confidence, c_{Tf}, and the bias change assumed to be irrelevant to the weight change,

$$\Delta w_{jf} = -\frac{\eta}{2}\frac{\partial}{\partial w_{jf}}\left[c_f^2 c_{Tf}^2 \left(X_{Tf} - X_f\right)^2\right], \quad 2)$$

with step size, η. The weight change simplifies to a formula that is proportional to both the error and the difference between the bias-adjusted input forecast and the fused forecast, as well as the confidence values of all of the forecasts,

$$\Delta w_{jf} = \eta c_{Tf}^2 c_{jf} c_f \left(X_{Tf} - X_f\right)\left[X_{jf} + b_{jf} - X_f\right]. \quad 3)$$

The last term arises from the normalization of the fused forecast by the weighted confidences. After the weight change has been applied, the weights are re-normalized so that they sum to one,

$$w_{jf}(t) = \frac{w_{jf}(t-1) + \Delta w_{jf}}{\sum_{k=1}^{n}\left(w_{kf}(t-1) + \Delta w_{kf}\right)}, \quad 4)$$

The biases are updated using the same Lyapunov function with the weight change assumed to be irrelevant,

$$\Delta b_{jf} = -\frac{\eta}{2}\frac{\partial}{\partial b_{jf}}\left[c_f^2 c_{Tf}^2 \left(X_{Tf} - X_f\right)^2\right], \quad 5)$$

with bias step size, η_b. The bias change simplifies to a formula that is proportional to the error in the input forecast,

$$\Delta b_{jf} = \eta_b c_{jf} c_{Tf} \left(X_{Tf} - X_{jf}\right). \quad 6)$$

The biases become estimates of the error in the input modules and the weights are indicative of whether the bias-adjusted input module tends to over- or under-estimate the forecast.

Note that the equations for the weight and bias change are very simple computationally. The order of the algorithm is the order of the number of input modules. The ADF algorithm is tuned by optimizing the RMSE for the weight and bias step sizes, η and η_b.

The systems that are deployed have a complex set of checks that provide the algorithm with more stability than the simplified version used in this analysis, without adding significant complexity.

Missing data is handled gracefully by the ADF algorithm. If all of the input forecasts are missing, no adaptation takes place. In the operational system, weights are degraded when single input forecasts are missing. For this study, all inputs were present for all time steps.

2.2. Newbold-Granger (N-G)

The N-G algorithm is a simple algorithm for assigning the weights according to the relative errors observed for the input forecasts for a set period in the past (window, ν). No bias term is used. The weight is the inverse of the sum over the window of the errors for the input forecast, normalized by the sum of the inverses for all of the input forecasts,

$$w_{jf} = \frac{\left[\left(\sum_{t=T-\nu}^{T-1}\left(X_{Tf} - X_{jf}\right)\right)\right]^{-1}}{\sum_k \left[\left(\sum_{t=T-\nu}^{T-1}\left(X_{Tf} - X_{kf}\right)\right)\right]^{-1}} \quad 7)$$

The input module with the lowest error during the window period receives the largest weight. The weights sum to one.

The complexity of the N-G algorithm is on the order of the window size times the number of input modules, which is relatively low. The N-G algorithm is tuned by adjusting the size of the window to minimize RMSE.

Missing data create a problem for the algorithm. If the window is large enough, the probability of not having enough data becomes small.

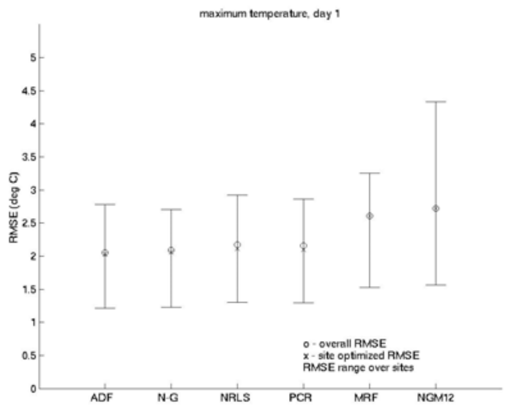

Figure 1: RMSE for maximum temperature forecast at day one. The range in RMSE is from the lowest to highest RMSE for the 18 sites with the parameters optimized.

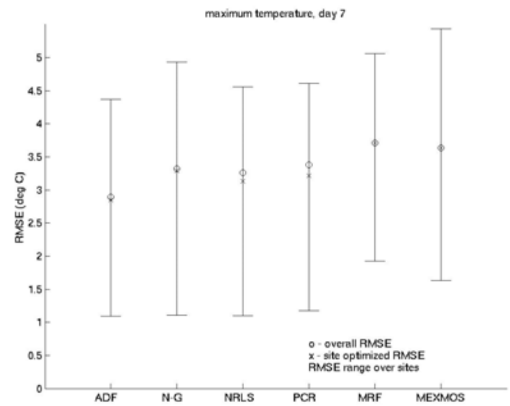

Figure 2: RMSE for maximum temperature forecast at day seven. The range in RMSE is from the lowest to highest RMSE for the 18 sites with the parameters optimized.

2.3. Nonnegative Restricted Least Squares (NRLS)

The NRLS method is a more comprehensive algorithm, which uses a window of forecasts to find the least squares best solution for positive weights using singular value decomposition. Let **X** be a matrix formed by the window of previous input forecasts, and the vector **y** be a vector formed by the corresponding truth values. The best solution for the weights in the least squares sense is the product of the pseudo-inverse of the input matrix and the truth vector,

$$\mathbf{w} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{y} \qquad 8)$$

By successively adjusting the weights by a Lagrangian factor, the weights can be forced to be non-negative. NRLS is a boundary method for searching for the best weight, as opposed to ADF, which is an interior method.

The complexity of NRLS is order of the cube of size of the window of inputs, which is much higher than the previous two methods. The only parameter that is tuned is the size of the window. Note that not all **X**'s will produce weights due to numerical problems with this method. For this study, those results are ignored, but the situation must be addressed in operational systems.

2.4. Principal Component Regression (PCR)

The PCR method is used to provide more numerical stability than the direct least squares methods at the expense of considerably higher computational complexity. First, the principal components , **D**, are determined from **X**, then the inputs are projected onto those principal components. The weight is then computed from the product of the pseudo-inverse of the projected matrix, **T**, and the truth,

$$\mathbf{w} = (\mathbf{T}'\mathbf{T})^{-1}\,\mathbf{T}'\mathbf{y}, \text{ where } \mathbf{T} = \mathbf{XD} \qquad 9)$$

The complexity of PCR depends on the number of principal components used, but is generally higher than for NRLS. If only a few are used, then the pseudo inverse of **T** is less computationally expensive, than it would be if more components are used. The weight vector estimated using fewer principal components may not produce as low RMSE as a weight vector created with more principal components.

Both the number of principal components and the size of the window are tuned for PCR. Although PCR is more numerically stable than NRLS, the method still fails on some data, which is overlooked in this study.

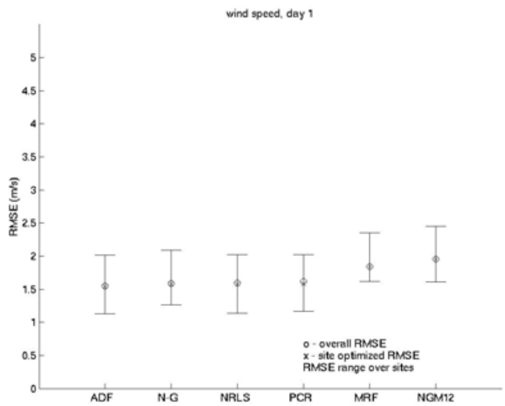

Figure 3: RMSE for wind speed forecast at day one. The range in RMSE is from the lowest to highest RMSE for the 18 sites with the parameters optimized.

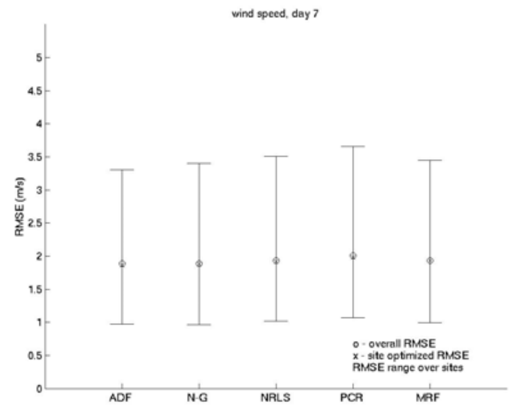

Figure 4: RMSE for wind speed forecast at day seven. The range in RMSE is from the lowest to highest RMSE for the 18 sites with the parameters optimized.

3. TUNING

The results of the tuning of the four algorithms are shown in figures 1 - 4. Four different data sets were used from the time period from January 1, 2002 to August 31, 2002. The four variables are maximum temperature and wind speed forecasts for one day and seven days. Eighteen domestic sites were used in each data set. The one day forecasts used about 205 forecasts per site. The seven day forecasts used only about 94 forecasts per site. Bad data was removed from the data sets, and not all dates were available for all sites, which caused the number of data points to vary. The size of the data sets should not dramatically affect the statistics of the data, however. Eight models were used in the day one forecast (00Z and 12Z AVN Dynamic-MOS, 00Z and 12Z ETA DMOS, MEX-MOS, 00Z and 12Z NGM-MOS and climatology). The day seven forecasts only use three models (MRF DMOS, MEX-MOS and climatology) for maximum temperature and two models for wind speed (MRF DMOS and climatology). The RMSE for the MRF DMOS model is included for comparison purposes. For the day one forecasts, the NGM-MOS RMSE is also included. For the day seven temperature, MEX-MOS RMSE is included.

The figures show the RMSE for each method calculated using the best overall parameters and the best parameters at each site. The spread in RMSE over the sites gives an indication of the sensitivity of the method to parameter values. Note that the ADF method is able to consistently produce the lowest RMSE and the lowest spread in the RMSE, with the only exception being the slightly smaller spread in RMSE for the N-G method for maximum temperature.

ADF improves on MRF DMOS by 22% for maximum temperature on day one and 23% for day seven. ADF improves on NGM-MOS by 25% for maximum temperature on day one and 22% for wind speed for day one. For wind speed, ADF improves on MRF by 16% on day one and 3% on day seven. ADF improves on MEX-MOS by 22% for maximum temperature for day seven. Less improvement is seen for the day seven wind speed because only MRF DMOS and climatology are fused.

The simpler methods (ADF and N-G) tended to outperform the more computationally expensive methods, even when all methods are optimized. This indicates the interior methods are preferable for these data sets and there is no need for the additional computational complexity.

4. ANALYSIS OF ADF

The ADF method is appropriate for the forecast model data fusion task for its ability to yield interesting insights, in addition to providing accurate forecasts. The individual forecast model

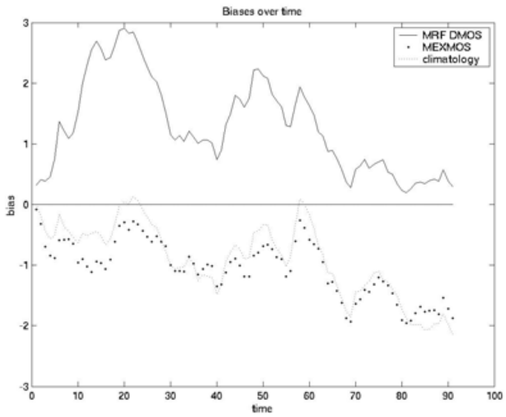

Figure 5: Bias adjustments for maximum temperature forecast at day seven for Los Angles for times from January 1, 2002 to August 31, 2002.

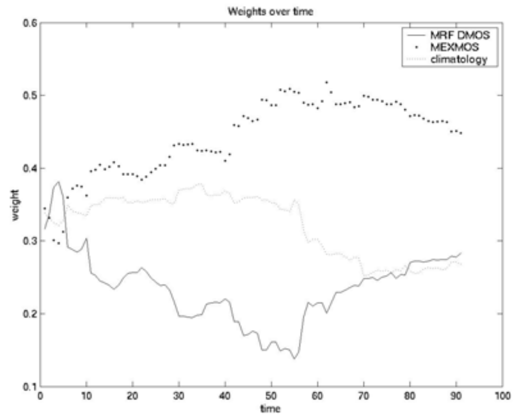

Figure 6: Weights for maximum temperature forecast at day seven for Los Angles for times from January 1, 2002 to August 31, 2002.

biases give an indication of the relative biases of the models. The weights indicate the reliability of the models. Once the biases reach a good value, the weight magnitude will tend to reflect the reliability of the model.

The evolution of a set of weights and biases for a single site, variable and lead time (Los Angles, CA maximum temperature, day seven) are shown in figures 5 and 6. The step sizes are the optimized values. The biases and weights for the same variable and time period for Goodland, KS are shown in figures 7 and 8 with the optimal step sizes. LA had the second lowest overall RMSE, after Phoenix, and Goodland had the highest RMSE for this data set.

The fluctuations in the biases indicate that there may be seasonal biases in the models, particularly in the summer months. The lowest weight is associated with MRF MOS in LA, except in the summer. The weights for Goodland show the opposite phenomena for the weights and much higher biases. The variation in the weights and biases for the different cities is very large, as is expected. The optimal learning rates vary in the same way.

5. DISCUSSION

The large amount of data that is processed in the model fusion applications examined here requires a simple yet accurate weighting and biasing algorithm. The additional problems of model changes, missing data and model additions require the algorithm to have even more flexibility than this study has directly addressed. The ADF has proven its ability to handle these requirements while producing results that are close to the results of the NRLS and PCR algorithms. In fact, for the data sets examined, the ADF algorithm outperformed the competing algorithms with lower computational complexity and more numerical robustness. Tuning the ADF's parameters is fairly simple, which allows even more adaptation of the system as it and its constituent models evolve. In addition to accuracy and ease of tuning, the ADF model also provides insight into the model biases and relative consistency.

REFERENCES

Aksu, C. and S.I. Gunter, 1992: An emperical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts, Int. J. Forecast., **8**, 27-43.

Jackson, J. 1991: *A User's Guide to Principal Components*, John Wiley.

Kosko, B., 1997: *Fuzzy Engineering*, Prentice-Hall.

Newbold, P. and C.W.J. Granger, 1974: Experience with forecasting univariate time series and the combination of forecasts, *J. Roy. Stat. Soc., A*, **137**, 131-165.

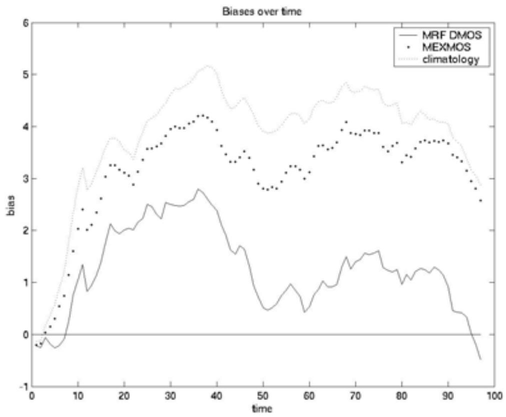

Figure 7: Bias adjustments for maximum temperature forecast at day seven for Goodland, KS for times from January 1, 2002 to August 31, 2002.

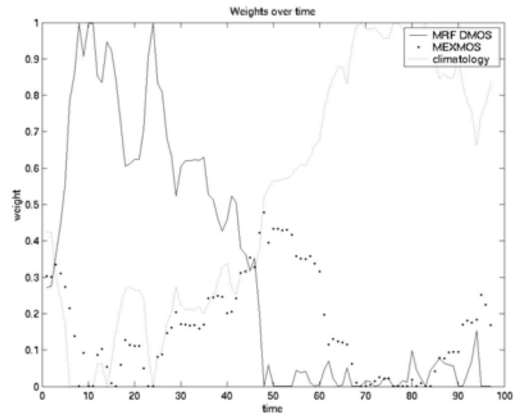

Figure 8: Weights for maximum temperature forecast at day seven for Goodland, KS for times from January 1, 2002 to August 31, 2002.

Young, G. 2002: Combining forecasts for superior prediction, *16th Conf. Prob. Stat. Atmos. Sci., 13-17 January, Orlando, FL,* 107-111.