118358 SUPERENSEMBLE FORECASTS WITH A SUITE OF MESOSCALE MODELS OVER THE CONTINENTAL UNITED STATES

Donald F. Van Dyke III *
Florida State University, Tallahassee, Florida

T. N. Krishnamurti
Florida State University, Tallahassee, Florida

1. INTRODUCTION

An important forecasting component of many major weather services across the globe is the use of ensemble forecasts (Toth and Kanlay 1993, Molteni et al. 1996). The idea of a multimodel superensemble was first presented in Krishnamurti et al. (1999). Since then, the superensemble has been shown to provide increased skill in the areas of global numerical weather prediction, hurricane forecasts, precipitation forecasts, and seasonal climate forecasts (Krishnamurti et al. 2000). This manuscript expands this capability into the realm of high resolution, short-range temperature forecasting over the continental United States (CONUS) using a suite of mesoscale models. The time period studied was August 18, 2005 through May 31, 2006, with forecasts made and verified out to 60 hours for the spring months of 2006 (March, April, May) using a six-hourly temporal resolution.

2. SUPERENSEMBLE METHODOLOGY

The superensemble methodology divides multimodel datasets into a training phase and a forecast phase. In the training phase, the forecast fields are regressed against the observed fields using multiple linear regression at all grid locations along the horizontal and vertical coordinates. The obtained regression weights are then used in the forecast phase to calculate the superensemble forecast. These regression weights, which are calculated for every gridpoint in the domain, provide a collective bias correction of the member models. This allows a superensemble forecast with somewhat higher skill compared to the ensemble mean and the best member model.

3. ANALYSIS DATASET

The North American Regional Reanalysis (NARR) dataset was employed in the superensemble training phase as well as forecast verification. The NARR represents a major improvement to the earlier NCEP global reanalysis datasets for both resolution and accuracy. It employs an analysis system similar to the ETA model's 3D-Var Data Assimilation System (EDAS) that was operational in April 2003, although it utilizes additional datasets for assimilation as described in Mesinger et al. (2005). Its temporal resolution is 3-hourly from near-realtime back to 1979. It has a 32 km/45 layer resolution. The 2-meter temperature fields from
this dataset were utilized for this study to represent the "observed" field. Mesinger et al (2005) also shows the gridded 2-meter temperature fields to be superior to the gridded global reanalysis 2-meter temperature fields in terms of RMS error and bias.

4. GRIDDED FORECAST DATASETS

Three gridded forecast datasets and their ensemble mean were utilized as member models to the superensemble. They include the mesoETA model, the NCAR Advanced Research WRF model (WRF-ARW), and the National Digital Forecast Database (NDFD) prepared by humans at the various National Weather Service Forecast Offices across the country.

The mesoETA model 2-meter temperature fields were obtained from the NOAA National Operational Model Archive & Distribution System (NOMADS) website at http://nomads.ncdc.noaa.gov. It is a 12 km resolution model initialized using the EDAS system. It is worth noting that this model was discontinued in June 2006 to make way for the new WRF-NMM model that is now run in its place. Some documentation on this model can be found at http://www.comet.ucar.edu/nwplessons/etalesson2/.

The WRF-ARW model 2-meter temperature fields were obtained from the NCAR Mass Storage System (MSS). Documentation on this model can be found at http://www.mmm.ucar.edu/wrf/users/pub-doc.html. It is a 22 km resolution model initialized from the 40 km Eta mass coordinate.

The NDFD 2-meter temperature fields were obtained from the NCDC's HDSS Access System (HAS) located at http://hurricane.ncdc.noaa.gov/pls/plhas/has.dsselect. The NDFD grids are created with human input on a 5 km resolution as well as a 3-hourly temporal resolution through the day 3 forecast. Documentation on the NDFD can be found at http://www.weather.gov/ndfd/technical.htm.

5. SUPERENSEMBLE FORECASTS AND RESULTS

* Corresponding author address: Donald F. Van Dyke III
Florida State Univ., Dept. of Meteorology,
Tallahassee, FL 32306-4520
Email: dvandyke@met.fsu.edu

The three gridded forecast datasets were re-gridded to a resolution of 32 km to match the analysis field resolution. The NDFD domain was utilized since that domain represented the least common denominator amongst the datasets. This area is shown in figure 1. Once the three forecast datasets were re-gridded, an ensemble mean of those was calculated and also included as a member model to the superensemble.

For forecasts made during all three months of spring 2006, 110 days of previous forecasts were utilized for the superensemble training phase. For March 2006, a training period ending on February 28 was used, excluding the month of December. For April 2006, a training period ending March 31 was used, excluding the months of December and January. For May 2006, a training period ending April 30 was used, excluding the months of December, January, and February. This strategy was utilized in order to attempt to counter the possibility of different seasonal biases affecting the superensemble regression coefficients. The possibility exists that the model biases are different in the cool season versus the warm season. Filtering out one cool season month as spring goes forward towards the warm season roughly compensates for this possible change in model bias. It is also important to note that the training period was only updated once per month at the end of each month and not on a daily basis. Also, any day in the training phase or forecast phase that did not have all member models present was excluded.

The mean absolute error (MAE) was the primary statistic used to assess the relative skill of all the forecasts. Results were broken down by individual month as well as represented by a total MAE of all three months combined. From the MAE statistics, a percent improvement plot of the superensemble compared to each member model was generated. Finally, a secondary consistency statistic was used that simply compared each individual forecast at each time and determined which model had the lowest MAE for that individual case.

Results from these verification statistics are shown in figures 2 through 14. It is interesting to note that these results come from using training that is offset by one day. During the training phase, the observed fields were accidentally lagged from the forecast fields of the member models by exactly one day. For example, a forecast valid on 12 UTC Tuesday was being compared with the observed field from 12 UTC Monday. This error was not discovered until very late in the analysis since the results were turning out well, and nothing appeared to be amiss. Unfortunately, due to the lateness of this discovery and an unfortunate computer crash and subsequent data loss, it was not possible to run the three months of forecasts with no offset to the training phase before this manuscript's deadline. However, partial results were completed before the data loss, and the partial results showed the offset training performing somewhat better than the non-offset training. Further investigation into this apparent anomaly will be required.

The superensemble consistently has the least MAE and the best results a large majority of the time. The improvements over the ensemble mean generally range between 5 and 7 percent, with improvements over the worst model as high as 45 percent. Taking the number of cases the superensemble performed the best and dividing by the total number of cases present, we find that the superensemble outperforms the member models, including the ensemble mean, 71.6% of the time.

6. CONCLUSIONS AND POSSIBLE FUTURE WORK

As stated above, additional testing that utilizes offset training is required to help determine an explanation for why comparing a forecast field valid at a different time than the observed field in the training phase produces such good results in the forecast phase. Offsetting the training by several hours, as well as reversing the offset, whereby the observed field time leads the forecast field time, may provide insight into explaining this phenomenon. Because our very preliminary findings show a one-day lag performing somewhat better than a zero-day lag, we feel that the precise length of the lag needs to be determined. This may have something to do with the spin-up of the models. A lag less than 24 hours may be the optimum lag. Also, fully running the three months with non-offset training is required to provide a baseline for comparison. Testing of the three spring months utilizing the spring months of the previous year as part of the training phase is also a possibility for future work. This may eliminate having to remove months out of the training dataset to compensate for the change of seasons. Finally, the temporal resolution can be increased to three-hourly in the future, as all of the datasets used have at least three-hourly temporal resolution.

7. FIGURES

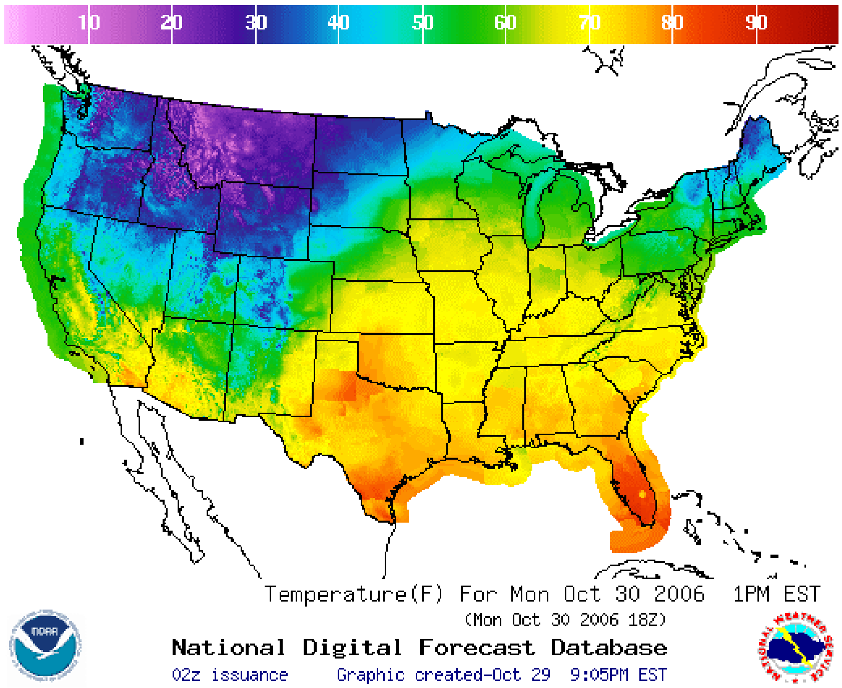

Figure 1: The domain utilized for comparison with the superensemble forecasts.

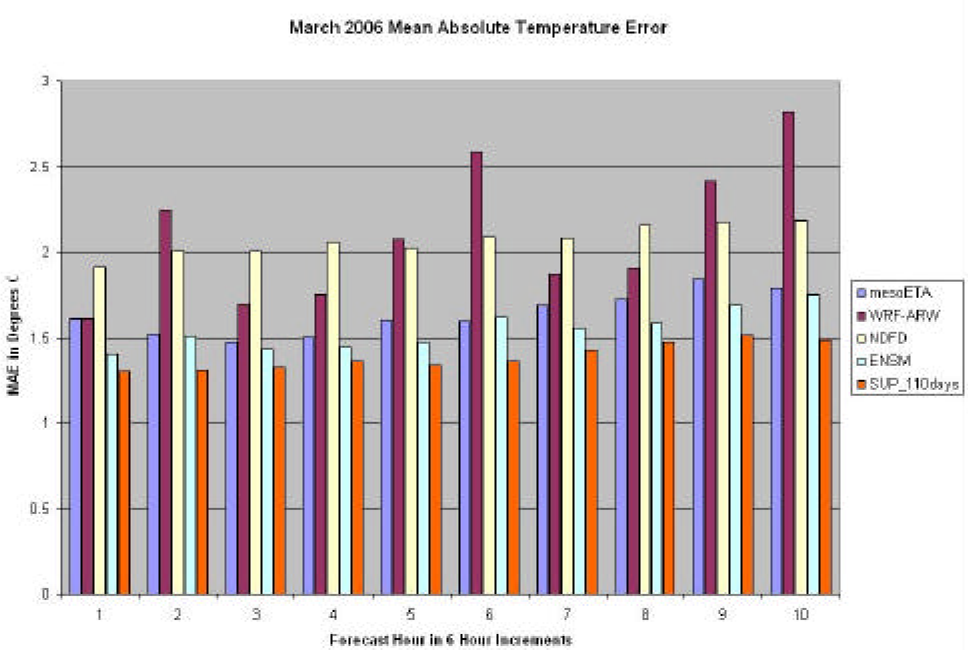

Figure 2: The March 2006 MAE statistics utilizing 110 days of training for the superensemble. Results are shown for forecasts in 6-hour time steps out to 60 hours.

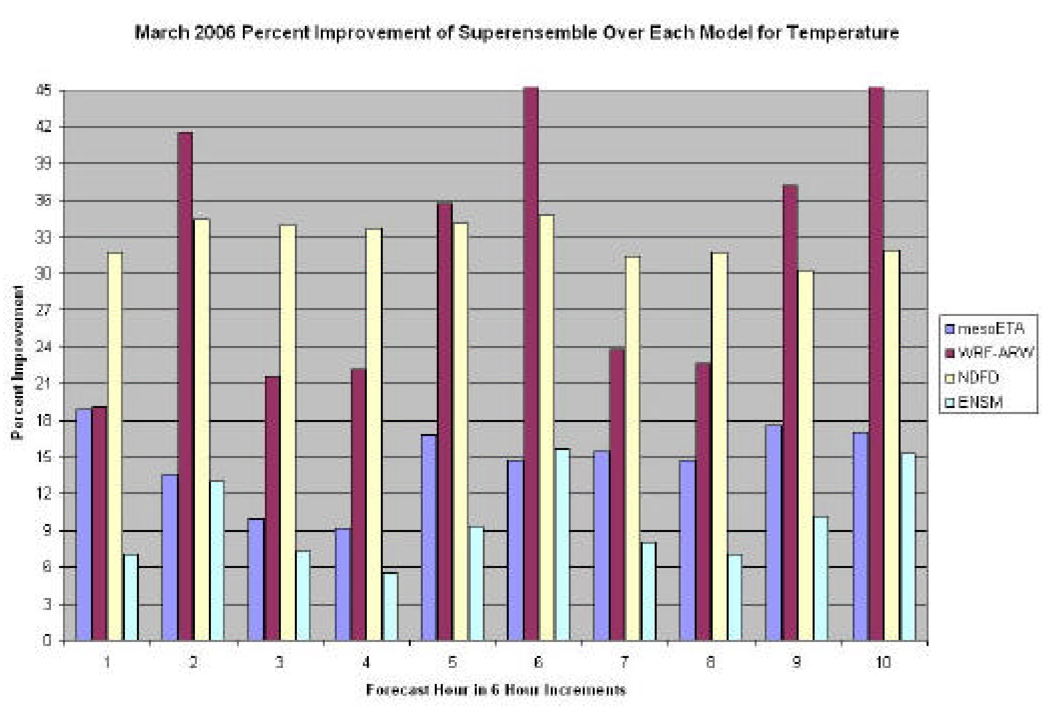

Figure 3: The March 2006 percent improvement statistics of the superensemble over the member models. Results are shown for forecasts in 6-hour time steps out to 60 hours.

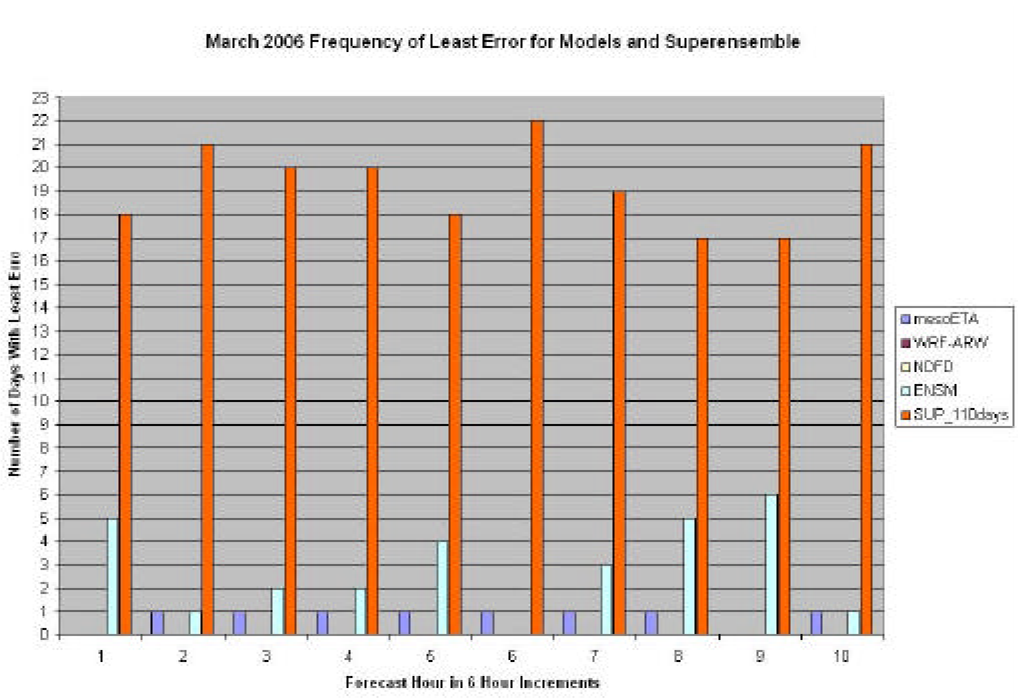

Figure 4: The March 2006 number of cases where each model had the lowest MAE at a given forecast hour. Results are shown for forecasts in 6-hour time steps out to 60 hours.

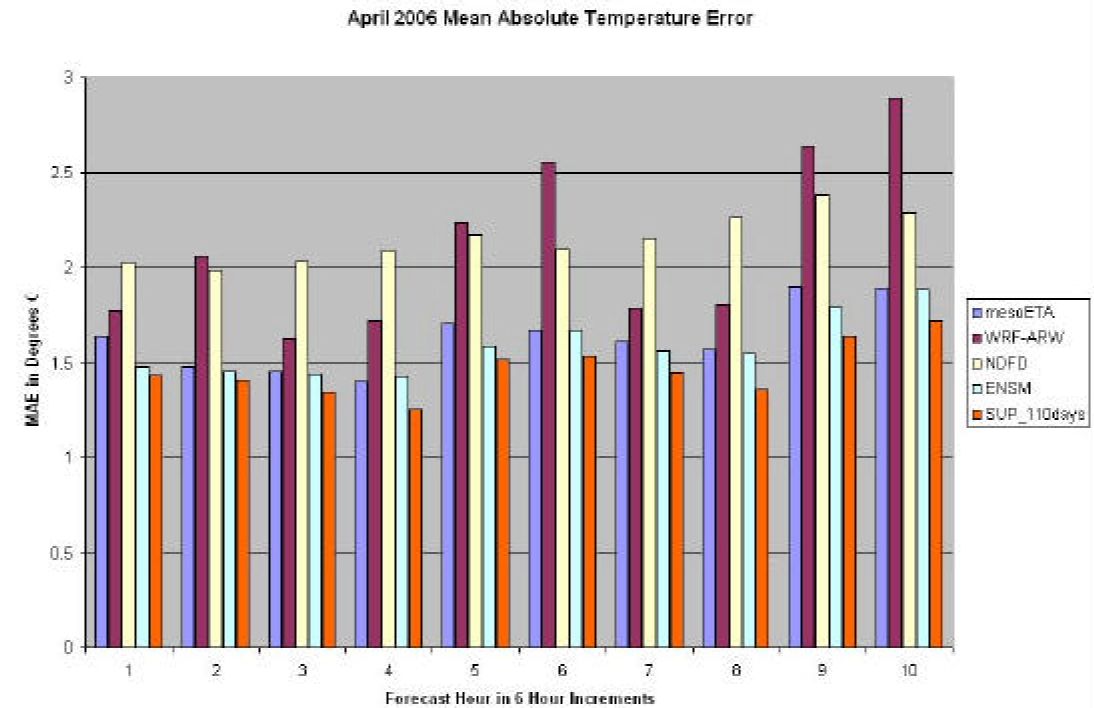

Figure 5: The April 2006 MAE statistics utilizing 110 days of training for the superensemble. Results are shown for forecasts in 6-hour time steps out to 60 hours.

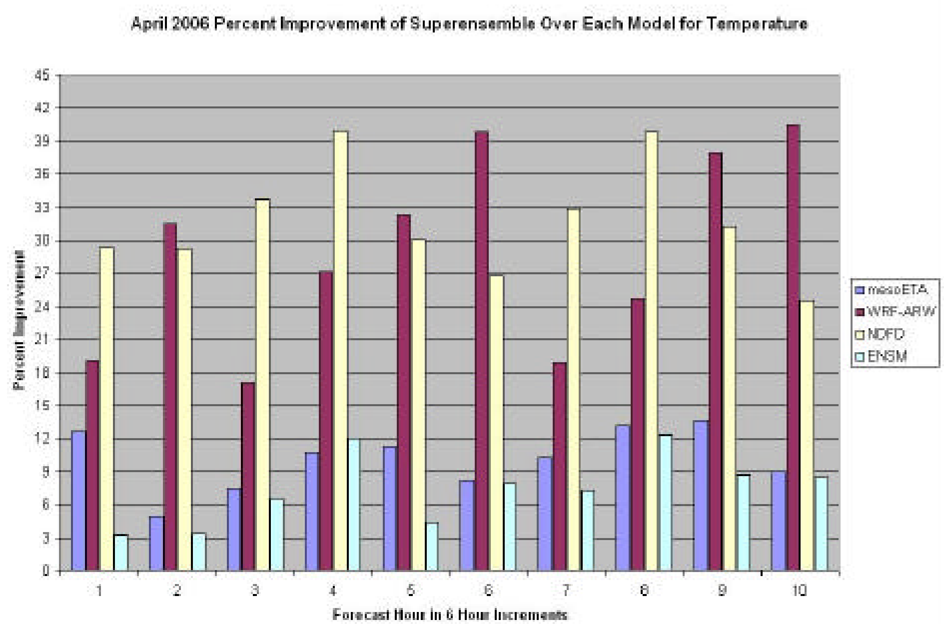

Figure 6: The April 2006 percent improvement statistics of the superensemble over the member models. Results are shown for forecasts in 6-hour time steps out to 60 hours.

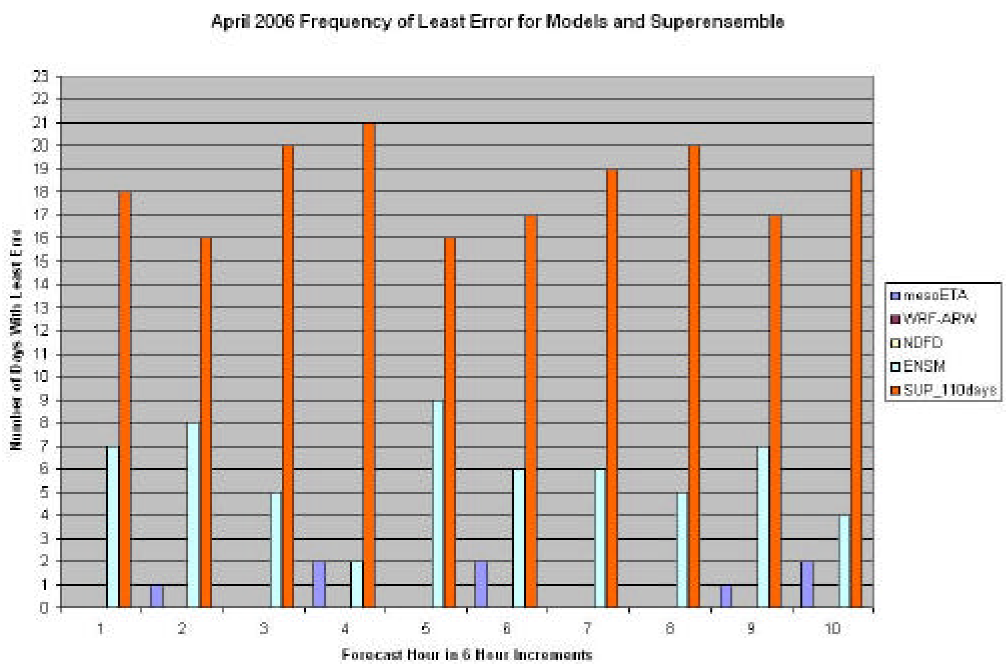

Figure 7: The April 2006 number of cases where each model had the lowest MAE at a given forecast hour. Results are shown for forecasts in 6-hour time steps out to 60 hours.

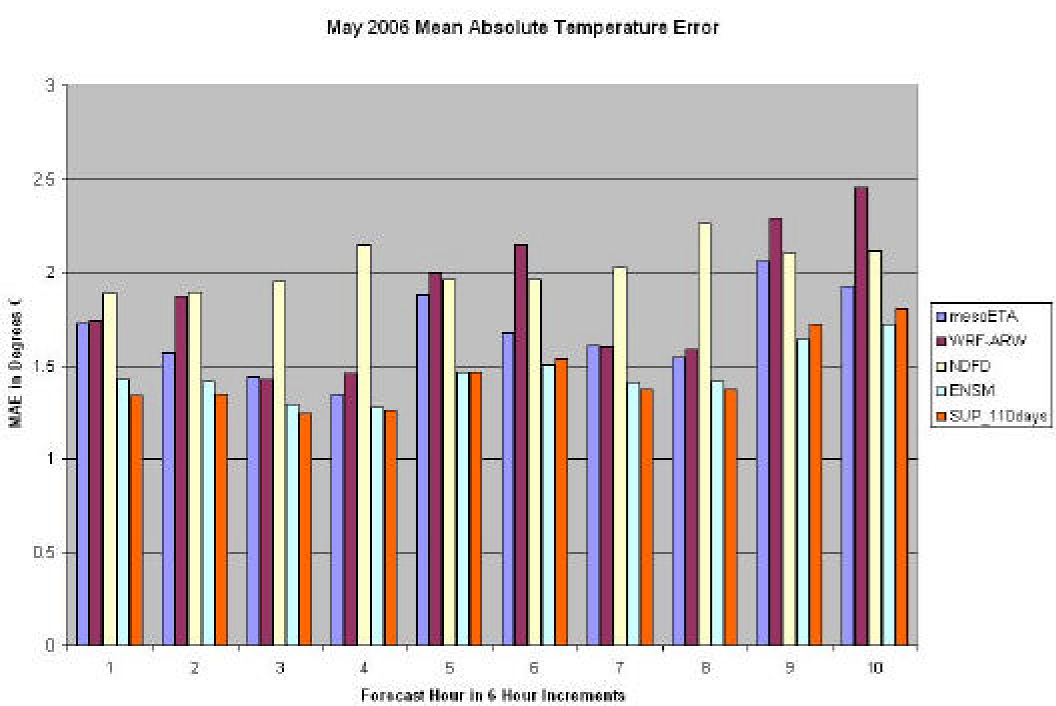

Figure 8: The May 2006 MAE statistics utilizing 110 days of training for the superensemble. Results are shown for forecasts in 6-hour time steps out to 60 hours.

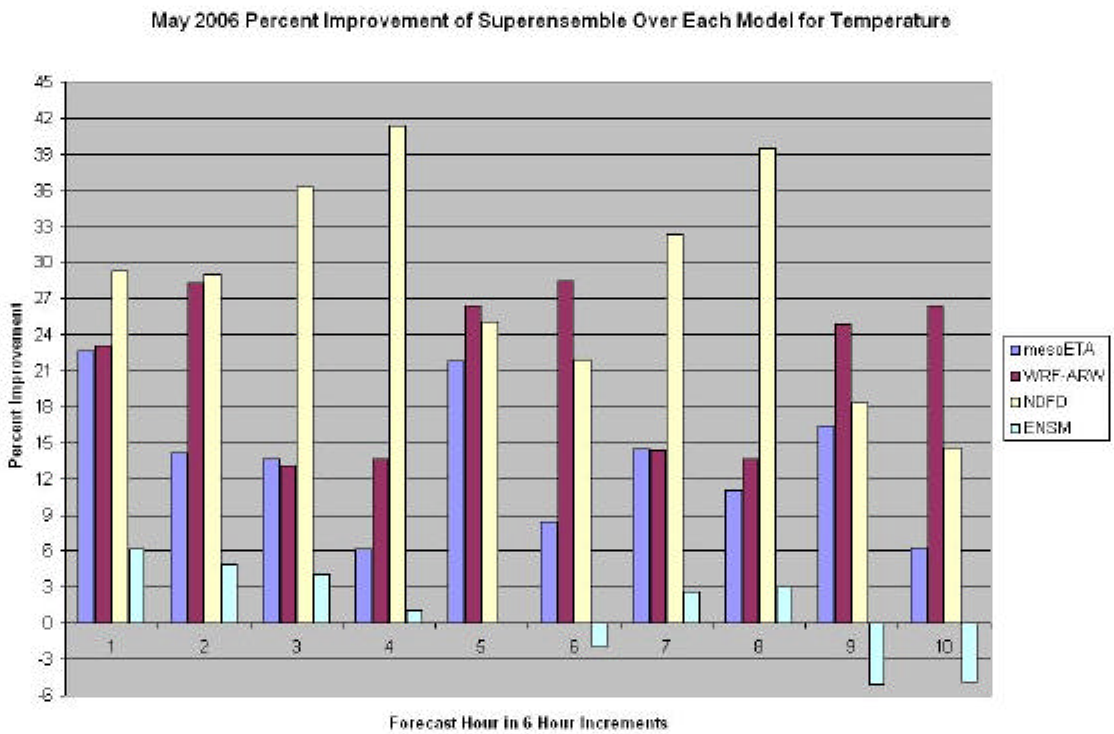

Figure 9: The May 2006 percent improvement statistics of the superensemble over the member models. Results are shown for forecas ts in 6-hour time steps out to 60 hours.

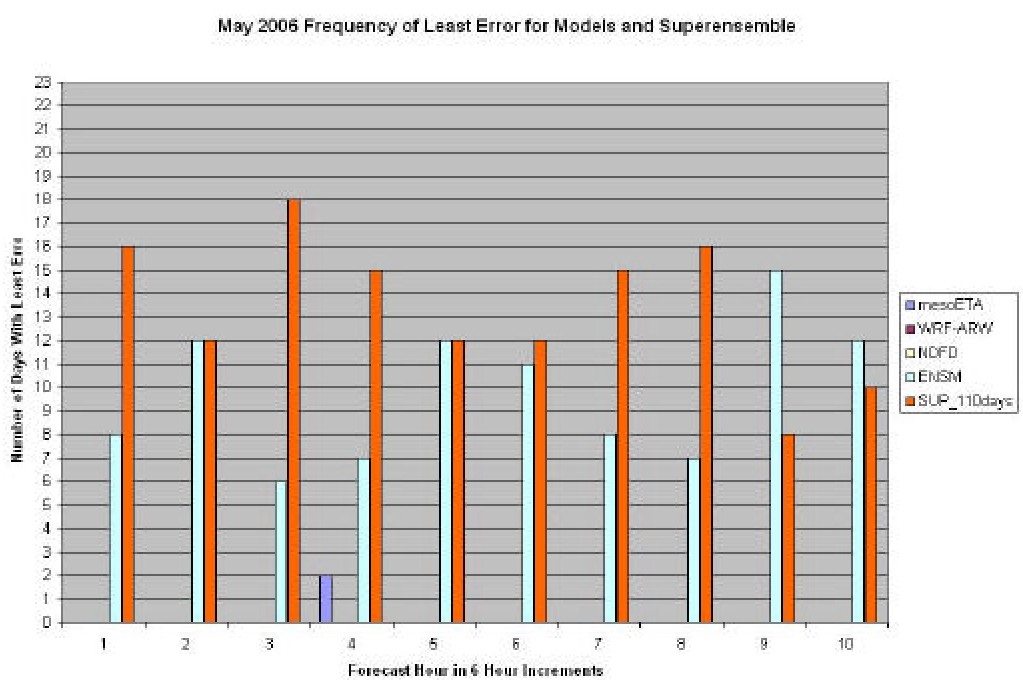

Figure 10: The May 2006 number of cases where each model had the lowest MAE at a given forecast hour. Results are shown for forecasts in 6-hour time steps out to 60 hours.

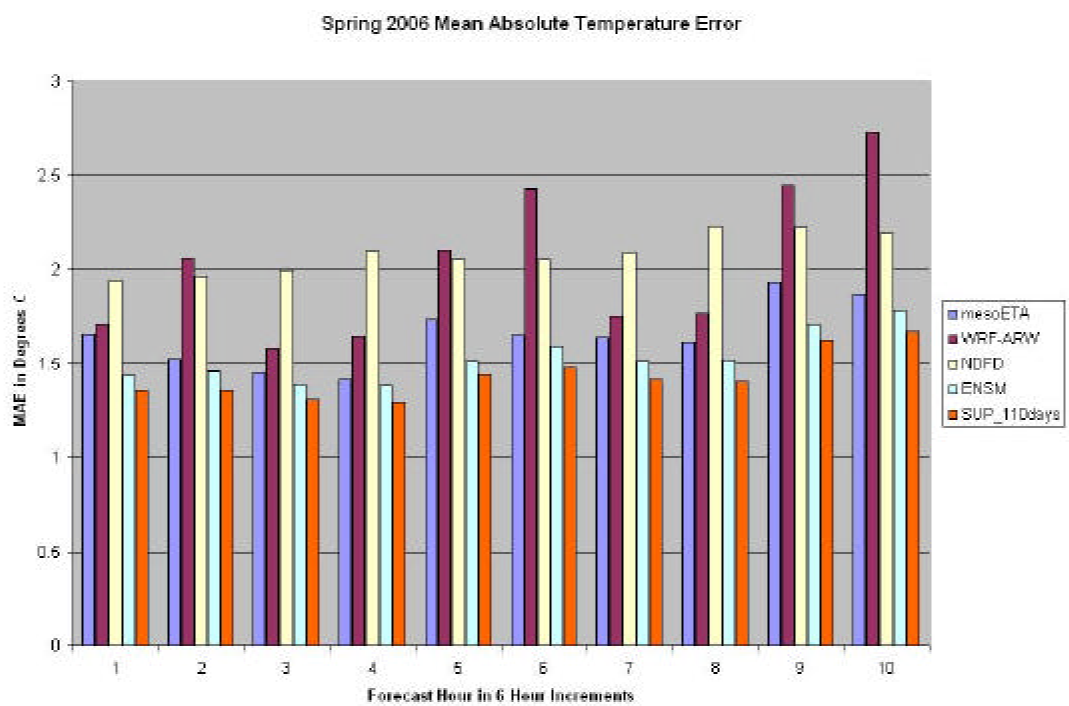

Figure 11: The Spring 2006 MAE statistics utilizing 110 days of training for the superensemble. Results are shown for forecasts in 6-hour time steps out to 60 hours.

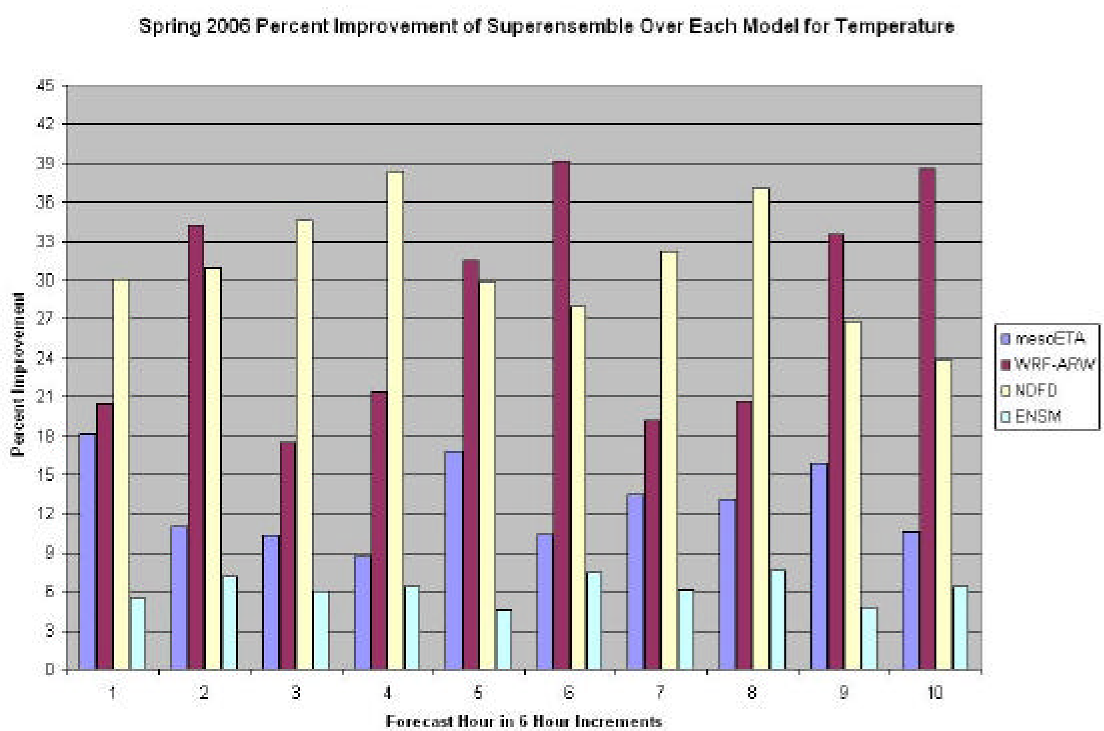

Figure 12: The Spring 2006 percent improvement statistics of the superensemble over the member models. Results are shown for forecasts in 6-hour time steps out to 60 hours.

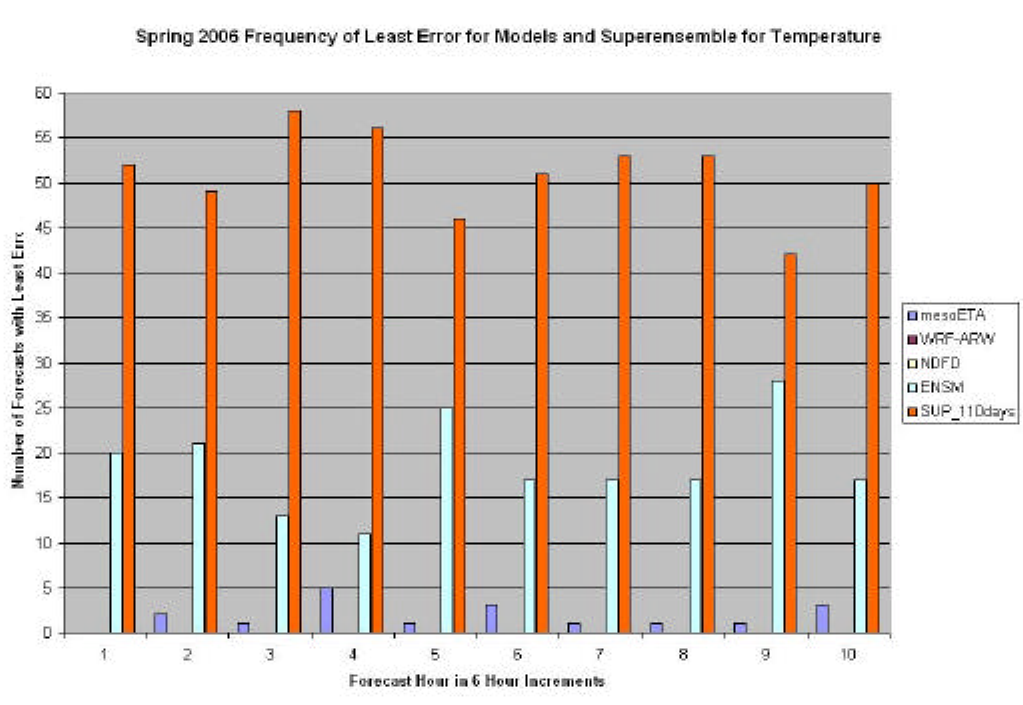

Figure 13: The Spring 2006 number of cases where each model had the lowest MAE at a given forecast hour. Results are shown for forecasts in 6-hour time steps out to 60 hours.

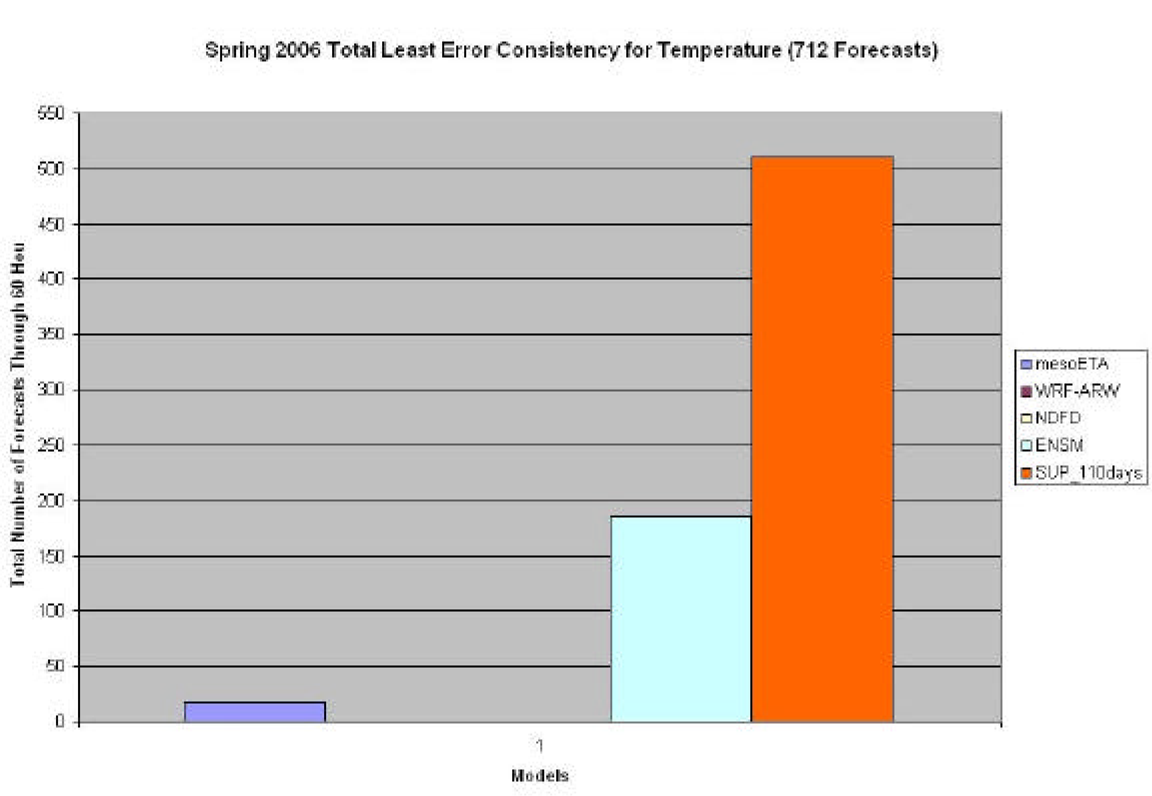

Figure 14: The Spring 2006 total number of cases where each model had the lowest MAE, not distinguishing between forecast hours.

8. REFERENCES

Krishnamurti, T. N., et al. 1999: Improved Weather and Seasonal Climate Forecasts from Multimodel Superensemble. *Science*, **285**, 1548-1550.

Krishnamurti, T. N., C. M. Kishtawal, D.W. Shin, and C. E. Williford, 2000: Multimodel Superensemble Forecasts for Weather and Seasonal Climate. *J. Climate*, **13**, 4196–4216.

Mesinger, Fedor, et al. 2005: North American Regional Reanalysis. http://wwwt.emc.ncep.noaa.gov/mmb/rreanl/index.html

Molteni, F., R. Buizza, T. N. Palmer and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Q. J. R Meteor. Soc*., **122**, 73-120.

Toth, Z., and E. Kanlay, 1993: Ensemble Forecasting at NMC: The Generation of Perturbations *Bull. American Meteor. Soc*., **74**, 2317-2330.