7.2 **Multi-Model Synthetic Superensemble Prediction System**

W. T. Yun[1,2], L. Stefanova[2], A.K. Mitra[2] and T. N. Krishnamurti[2]
[1] Korea Meteorological Administration
[2] Department of Meteorology/Florida State University

The Superensemble approach is a recent contribution to the general area of weather and extended range forecasting, developed at the FSU, and this technique is discussed extensively in a series of publications. Recent papers of that are by Krishnamurti and Sanjay (2003), Kharin and Zwiers (2002), and Yun (2003). Superensemble algorithm is embraced as a tool for making both of deterministic and probabilistic predictions. The Superensemble algorithm entails the division of a time line into two parts, called training phase and forecast phase. During training phase the forecast of member models are fit as a linear combination to observations with the objective of minimization of residual error variance. In general, the Superensemble approach yields forecast with considerable reduction in forecast error compared to the error in the member models, the bias-corrected ensemble mean forecast and the ensemble mean forecast.

A major component of Superensemble forecast initiative is training of forecast data set. The superensemble prediction skill during the forecast phase could be degraded if the training was executed with either a poorer analysis or poorer forecasts. That means that the prediction skill is improved when higher quality training data set is deployed for the evaluation of the multi model bias statistics.

Despite the continuous improvement of both dynamical and empirical models, the predictive skill of extended forecasts remains quite low. The multi model Superensemble is one emphasized way to improve weather predictions. In the context of seasonal climate forecasts, we had noted that the multi-model bias-removed ensemble demonstrated a skill slightly below that of climatology, whereas the superensemble provided a skill slightly above that of climatology. For a further improvement to the superensemble method we suggest synthetic superensemble algorithm, developed by Yun (2003).

Introduction of Synthetic Data Processor

The synthetic superensemble algoritm is an alternative method to obtain high quality forecast data set and to improve the prediction skill of extended forecast. In order to do so, we generate synthetic data set which is derived from a combination of the past observations and forecasts.

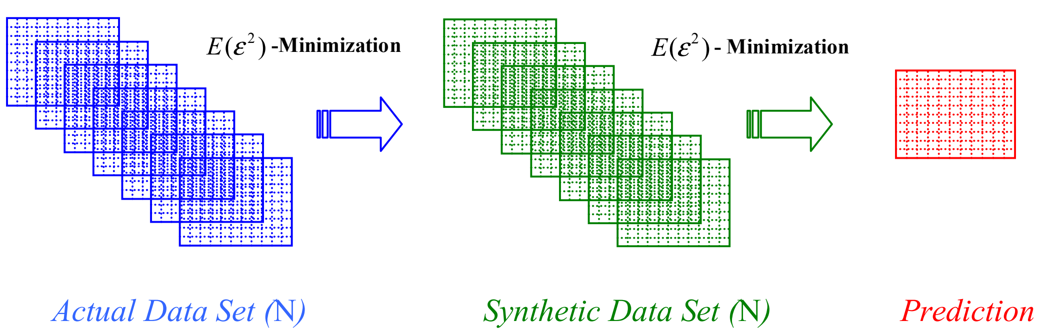

Fig. 1 Extended range forecast using synthetic data sets. To generate the synthetic data sets, EOF- and Fourier analysis and random process are used as a statistical method.

Synthetic data set is created by finding consistent spatial pattern between observation and forecast data set. It is a linear regression problem in EOF space. A set of such synthetic forecasts is then used for the creation of a superensemble forecast.

Corresponding author: Dr. W.T. Yun
Address: Department of Meteorology, Florida State University, Tallahassee, FL 32306-4520
E-mail: wtyun@io.met.fsu.edu

Once EOFs of observation data (time series) are found, the time series of observation can be written as a linear combination of EOFs.

$$O(x,t) = \sum_n P_n(t)\phi_n(x) \qquad (1)$$

PC time series, $P(t)$, in equation (1) represents how EOFs (spatial patterns) evolve in time. PCs are independent of each other. Now we are interested in knowing the spatial patterns of forecast data which evolve in a consistent way with EOFs of observation. That means what is matching spatial pattern in forecast data, $F(x,t)$, which evolves according to PC time series $P(t)$ of observation data? We can answer such a question by finding concurrent patterns among observation and forecast data. Patterns of two variables are called concurrent when they have the common evolution history. The procedure of finding consistent pattern can be written a regression problem in EOF space. And the forecast data set can be written as a linear combination of EOFs

$$F_i(x,T) = \sum_n F_{i,n}(T)\cdot\varphi_{i,n}(x) \qquad (2)$$

Index i and n in equation (2) indicates the number of forecast models and the number of EOF modes, respectively. $\varphi(x)$ and $F(t)$ are the EOFs and the corresponding PC time series of $F(x,t)$. Then a regression relationship is sought between the observation PC time series and a number of PC time series of forecast data.

$$P(t) = \sum_n \alpha_{i,n} F_{i,n}(t) + \varepsilon(t) \qquad (3)$$

With equation (3) we want to express observation time series, $P(t)$, as a linear combination of predictor time series, $F(t)$ in EOF space. And the regression coefficients, α_n, are found that the residual error variance, $E(\varepsilon^2)$, is minimized. Once the regression coefficients α_n are found, we can write PC time series of synthetic data set

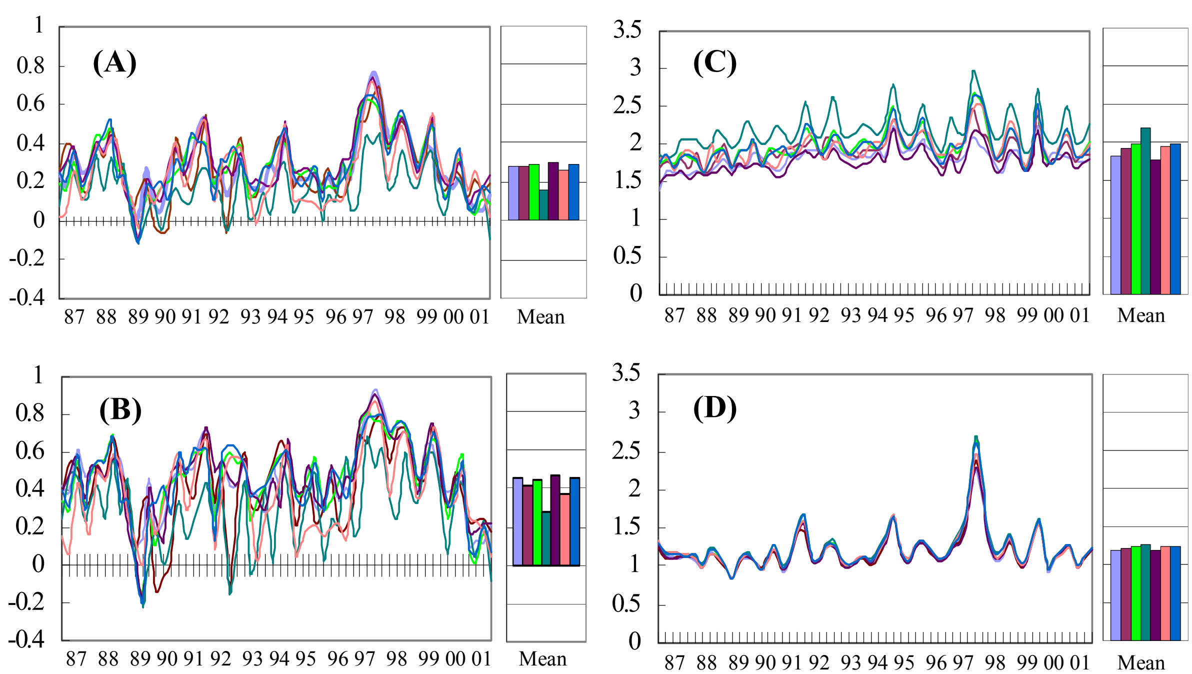

Fig. 2 Globally averaged ACC and RMS error of the actual (DEMETER models) and synthetic data set of precipitation. Forecasts of DEMETER models are averaged over 2-4 months forecasts. **(A):** ACC of DEMETER global precipitation forecasts. **(B):** ACC of synthetic data set. **(C):** RMS of DEMETER global precipitation forecasts. **(D):** RMS of synthetic data set. Bars in diagram indicate from left ECMWF, UKMO, Meteo France, MPI, LODYC, INGV, and CERFACS DEMETER models. Units of RMS is mm/day.

$$F_i^{reg}(T) = \sum_n \alpha_{i,n} F_{i,n}(T) \quad (4)$$

Synthetic data set is reconstructed with corresponding EOFs and PCs

$$F_i^{syn}(x,T) = \sum_n F_{i,n}^{reg}(T) \cdot \phi_n(x) \quad (5)$$

How good is the performance of synthetic forecast data set? Fig. 2 illustrates the ACC and RMS error of actual and synthetic data set, which are produced from DEMETER (Development of a European Multi-Model Ensemble System for Seasonal to Inter-Annual Prediction) forecasts (Palmer, 2003). DEMETER system comprises seven global atmosphere-ocean coupled models, each running from an ensemble of nine different initial conditions. The synthetic data procedure generates obviously high quality synthetic data and improves predictive skill more than 20% compared with actual data set. This generated synthetic data set is used as an input data of synthetic superensemble model.

In Fig. 3, our preliminary results shown that the proposed synthetic algorithm increases predictive skill of seasonal forecasts clearly better than the DEMETER actual data set. The forecast produced by synthetic ensemble and superensemble generally outperforms the both of ensemble mean and model forecast.

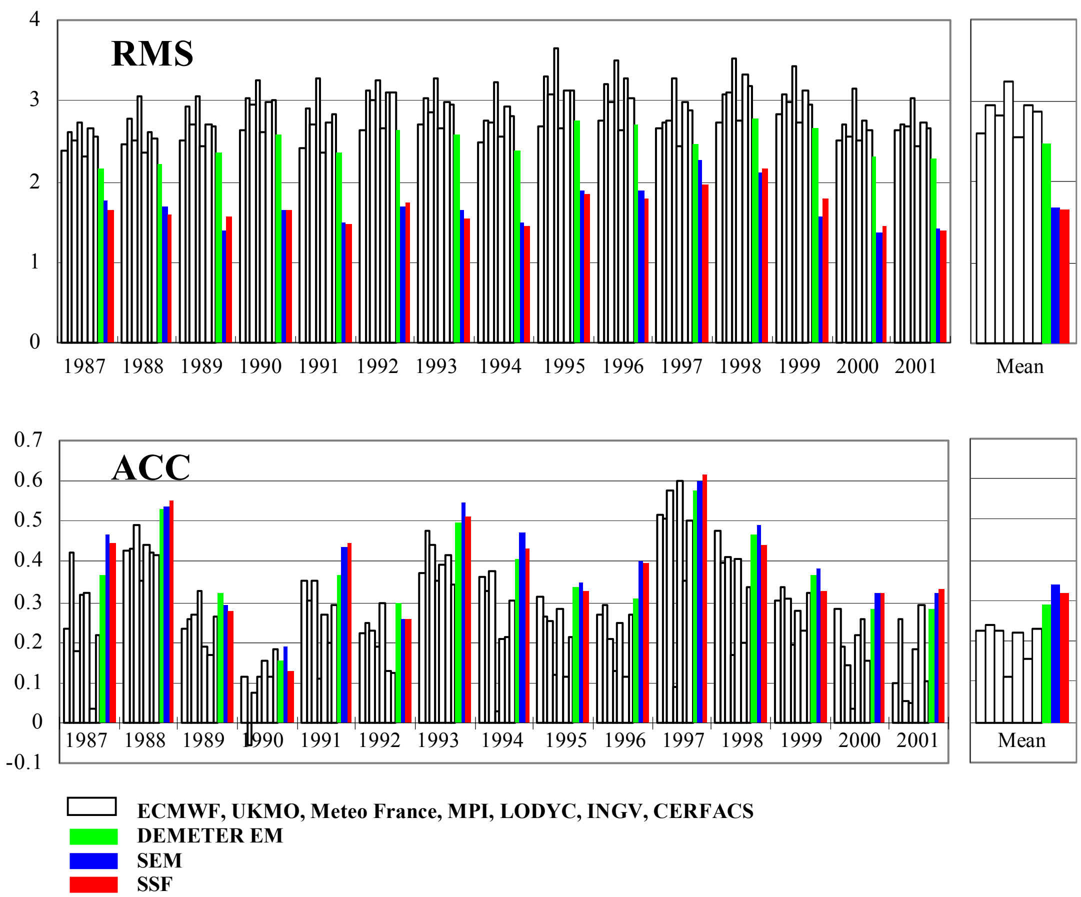

Fig. 3 Cross-validated JJA mean RMS and ACC skill of synthetic superensemble for region of 0°-60°N. Synthetic data set is generated from DEMETER over 2-4 months averaged precipitation forecasts. EM, SEM and SSF indicate ensemble mean, synthetic ensemble mean and synthetic superensemble forecast, respectively. Units of RMS is mm/day.

A post-processing algorithm based on synthetic multi-model solutions toward observed fields during a training period is one of the best solutions for extended range prediction. Our study shows that the proposed synthetic technique further reduces the forecast errors below those of the conventional superensemble technique and increases the predictive skill of extended forecasts.

Acknowledgement: DEMETER data was provided by ECMWF.

References

Kharin, V. V. and F. W. Zwiers, 2002: Notes and correspondence: Climate predictions with multimodel ensembles. *J. Climate*, **15**, 793-799.

Krishnamurti, T. N. and J. Sanjay, 2003: A new approach to the cumulus parameterization issue. *Tellus*, 55A, 275-300.

Palmer, T. N. and coauthors, 2003: Development of a European Multi Model Ensemble System for Seasonal to Inter-Annual Prediction(DEMETER), *submitted to BAMS.*

Yun, W.T., Lydia Stefanova and T.N. Krishnamurti, 2003: Improvement of the Multi-Model Superensemble Technique for Seasonal Forecasts, *Journal of Climate*, in print.