*Atmosphere*, **17**(3), 217-230 September 2007

# 로렌쯔-95 모델을 이용한 앙상블 섭동 비교: 브레드벡터, 직교 브레드벡터와 앙상블 칼만 필터

**정관영[1,*] · 바커 데일[2] · 문선옥[3] · 전은희[1] · 이희상[1]**
[1]국립기상연구소, [2]미국 국립대기과학연구소(NCAR)
[3]기상청 수치예보센터
(2007년 3월 2일 접수; 2007년 6월 20일 승인)

# Comparison of Ensemble Perturbations using Lorenz-95 Model: Bred vectors, Orthogonal Bred vectors and Ensemble Transform Kalman Filter(ETKF)

**Kwan-Young Chung[1,*], Dale Barker[2], Sun-Ok Moon[3], Eun-Hee Jeon[1] and Hee-Sang Lee[1]**
[1]*National Institute of Meteorological Research/KMA,* [2]*National Center for Atmospheric Research (NCAR)*
[3]*Numerical Prediction Center/KMA*
(Manuscript received 2 March 2007; in final form 20 June 2007)

**Abstract**

Using the Lorenz-95 simple model, which can simulate many atmospheric characteristics, we compare the performance of ensemble strategies such as bred vectors, the bred vectors rotated (to be orthogonal to each bred member), and the Ensemble Transform Kalman Filter (ETKF). The performance metrics used are the RMSE of ensemble means, the ratio of RMS error of ensemble mean to the spread of ensemble, rank histograms to see if the ensemble member can well represent the true probability density function (pdf), and the distribution of eigen-values of the forecast ensemble, which can provide useful information on the independence of each member. In the meantime, the orthogonal bred vectors can achieve the considerable progress comparing the bred vectors in all aspects of RMSE, spread, and independence of members. When we rotate the bred vectors for orthogonalization, the improvement rate for the spread of ensemble is almost as double as that for RMS error of ensemble mean compared to the non-rotated bred vectors on a simple model. It appears that the result is consistent with the tentative test on the operational model in KMA. In conclusion, ETKF is superior to the other two methods in all terms of the assesment ways we used when it comes to ensemble prediction. But we cannot decide which perturbation strategy is better in aspect of the structure of the background error covariance. It appears that further studies on the best perturbation way for hybrid variational data assimilation to consider an error-of-the-day(EOTD) should be needed.

**Key words:** Ensemble, Bred vectors, Orthogonalization, Ensemble Transform Kalman Filter

## 1. 서 론

수치예보의 불확실성은 초기자료 오차와 모델오차에 기인하며, 예보기간이 길어질수록 증대된다. 이러한 불확실성에 대한 표본 확률밀도함수 (pdf)를 제공할 수 있는 매우 실용적인 방법이 바로 앙상블 예보체계이다.

앙상블예보체계 중 가장 핵심적인 부분은 초기 섭동 (perturbation)을 구성하는 부분이다. 이는 앙상블 예보의 예보성이 샘플링 오차가 포함된 초기 섭동장에 따라 결정되기 때문이다 (Toth and Kalnay, 1997). 현업기관에서 주로 쓰이는 섭동 구성방법은 NCEP의 오

*Corresponding Author: Dr. Kwan-Young Chung, National Institute of Meteorological Research, 45 Gisangcheong-gil, Dongjak-gu, Seoul, 156-720.
Phone : +82-2-836-0687, Fax : +82-2-836-0688
E-mail: kychung@kma.go.kr

ⓒ Korean Meteorological Society, 2007

차브리딩 (Toth and Kalnay, 1994; 1997), ECMWF의 singular vector (Buizza *et al.,* 1993)이며, 다른 섭동 구성 방법으로써 앙상블 칼만 필터 (Evensen, 1994; Bishop *et al.*, 2001), 관측섭동 (perturbed observation) (Houtekamer and Derome, 1995) 및 무작위섭동 (random perturbation)(Du *et al.,* 1997) 등을 열거할 수 있다.

에러브리딩은 초기자료에 +/-섭동이 포함된 초기자료로 수치모델을 적분하여 오차를 키운 다음, 분석시간에 이 결과의 차 (브레드 벡터)를 재조정하여 구성하는 방법이다. 앙상블예보체계를 구성하는 간단한 방법이나, 예보시간이 길어질수록 브레드벡터 한 두 모드에 에너지가 집중되고, 스프레드가 작아지는 경향을 보이는 단점이 있다. 이는 각 브레드 벡터가 직교하지 않기 때문에 발생하는 것으로 알려져 있다. 이를 해결하기 위하여 미국 NCEP에서는 독립적인 싸이클로부터 준직교 브레드벡터를 구하고 있으며 (Toth and Kalnay, 1997), 기상청에서도 2005년부터 요인회전을 이용하여 준직교 브레드벡터를 이용하고 있다 (문선옥 등, 2006; 기상청, 2005).

Singular vector 방법은 선형역학모델에서 가장 빠르게 성장하는 고유모드를 찾아 이를 통해 앙상블을 구성하는 것이다 (Buizza *et al.*, 1993). 이 벡터들은 각각 직교성을 갖게 되고, 성장률이 큰 leading singular 벡터들로 초기 섭동을 구성할 수 있어, 적은 수의 벡터를 이용하여 초기 샘플링오차를 표현할 수 있는 장점이 있다. 그러나 비선형 모델에서 초기 오차의 시간적 전개를 선형적으로 가정한다는 점, 선형 및 수반연사자의 산출 등 복잡한 과정이 있으며, 초기 오차가 너무 빠르게 성장하는 단점이 있다.

앙상블 칼만필터 방법은 칼만필터를 앙상블 개념을 도입하여 구성하는 것이다. 원래의 칼만필터는 오차의 선형예보방정식과, 분석오차를 구하는 방정식으로 구성되어 있으나, 앙상블 칼만필터에서 예보오차는 비선형 모델의 앙상블 예보결과를 이용하여 구성되며, 이를 이용하여 칼만 분석방정식을 풀어 해를 구하는 방법이다. 앙상블 칼만필터를 이용하여 앙상블을 구성하는 방법엔 1) 분석방정식에 포함된 관측 자료에 무작위 섭동을 주는 것 (Evensen, 1994)과 2) 분석오차로 구성된 분석방정식을 제곱근 알고리듬 (square root algorithm)을 통해 구성하는 것이 있다 (Tippett *et al.*, 2003). 후자의 방법으로 Ensemble Transform Kalman Filter (ETKF) (Bishop *et al.*, 2001), Ensemble Adjustment Kalman Filter (EAKF) (Anderson, 2001) 등이 대표적이다. 앙상블 칼만필터는 비선형모델을 그대로 이용하기 때문에 선형연산자가 필요 없다는 점, 새로운 섭동을 칼만필터 분석방정식을 이용하여 구하기 때문에 표본오차를 최소화할 수 있다는 것이 장점이다. 그러나 앙상블 멤버의 수가 충분하지 않을 경우 초기 오차의 확률 밀도 함수를 제대로 표현할 수 없고, 방대한 연산을 필요로 하기 때문에 실제 현업에서 사용하기가 쉽지 않은 단점이 있다.

관측섭동이나 무작위 섭동을 주는 방법은 매우 간단하게 직교하는 앙상블을 구성할 수 있으나, 표본오차를 최소하기 위해서는 아주 많은 앙상블 멤버를 구성해야 하고, 오차의 역학적 특성을 반영하지 못하는 단점이 있다.

현재까지 연구결과에 따르면, 가장 이상적인 섭동 구성 방법은 무작위 섭동과 앙상블 칼만필터 기법으로 알려져 있다 (Bowler, 2006). Wang *et al*. (2003)은 ETKF가 브리딩 방법보다 우수하다는 것을 보였고, Hamill *et al.* (2000)은 브리딩, singular vector, 그리고 관측섭동 방법을 비교한 결과, 관측섭동 방법이 이상적인 섭동에 근접한다는 것을 밝혔다. 이중 무작위 섭동이나 관측섭동은 불확실성이 많고, 비용이 많이 들기 때문에 최근에는 기존 앙상블 체계 (브리딩이나 singular vector)보다 우수한 앙상블 칼만필터 기법을 이용하는 것이 증가하는 추세이다.

기상청은 2001년부터 NCEP (Toth and Kalnay, 1997)에서 사용하고 있는 브리딩 방법 (breeding method)을 이용하여 앙상블예보를 시작하였다. 초기에는 17개 멤버 (8개의 브레드 벡터와 규준수행)로 앙상블예보가 수행되어, 현재 32개 멤버로 증대되었다. 특히 2005년부터 브레드벡터를 요인회전 (factor rotation)하여 각 멤버가 준직교하도록 개선하였다 (기상청, 2005). 아울러 이러한 앙상블 체계를 이용하여, 매일 매일 변하는 예보 오차를 반영할 수 있는 하이브리드 3차원 변분자료동화 (3dvar)[1)]를 통해 초기자료의 품질이 개선될 수 있을 것이다.

---

1) 하이브리드 자료동화방법은 오차 공분산을 앙상블에서 나온 매일 매일 변하는 오차공분산과 기후학적으로 계산하여 3dvar나 4dvar 등에 사용하는 정적오차 공분산을 적절하게 가중 평균하여 사용하는 방법임. 즉, $P^f = (1-\alpha)P^f_{ensemble} + \alpha P^f_{climate}$과 같은 형태. 여기서 $P^f$는 오차 공분산, $\alpha$는 가중치임 (Hamill and Snyder, 2000).

하이브리드 3차원 변분자료동화에 있어서 무엇보다도 질 좋은 앙상블자료가 가장 중요한 요인이다. 따라서 본 연구에서는 직교브레드 벡터의 효과를 간단한 로렌쯔-95 모델을 이용하여 평가하고, 직교브레드 벡터와 ETKF에 의한 섭동을 상호 비교하고자 한다. 이러한 원론적인 작업을 통해 우리 현실에 적합한 앙상블 예보체계를 수립하는데 기초적인 연구 자료를 제공하는 것이 본 연구의 목적이다.

## 2. 모델 및 자료동화 방법

### 2.1 로렌쯔-95 모델

본 연구에서 사용된 대기모델은 로렌쯔(1998)가 제안한 것이다. 이 모델은 $J$ ($j=1,.....J$)개 변수로 구성된 편미분 방정식이며 다음과 같다.

$$\frac{dX_j}{dt} = (X_{j+1} - X_{j-2})X_{j-1} - X_j + F \qquad (1)$$

각 변수들을 cyclic으로 구성하였으며 ( $X_{-1} = X_{J-1}$, $X_0 = X_J$, and $X_{J+1} = X_1$), $F$는 8.0의 값이 이용되었다. 위 방정식에서 종속변수 $X$는 소용돌이나 기온 등으로 간주될 수 있으며, $J$는 위도를 따라 원형으로 구성된 격자점으로 인식될 수 있다. 따라서 연직이나 경도방향이 고려되지 않은 1차원 모델과 같은 형태이다.

이 모델은 어느 정도 대기가 가지고 있는 여러 특징을 간단하게 모의할 수 있어 대기의 예보성을 평가할 때 주로 사용된다. 즉, 첫 번째 항인 비선형 항은 이류 항을 모의하기 위한 것으로 총 에너지 ( $(X_1^2 + .... + X_J^2)/2$) 가 보존되는 방향이다. 두 번째 항은 역학적 또는 열적 소산을 나타낸 선형 항으로 에너지를 감소시키는 역할이다. 세 번째 항 ($F$)은 상수항으로 외부 강제력을 나타낸다.

이 방정식을 Runge-Kutta 4차(4th order)를 이용하여 풀었으며, 시간 간격 ( $\delta t$ )는 0.05 (실제 시간은 6시간)이다. 총 변수 $J$는 100이며, 앙상블 사이즈($N$)는 16개이다.

Fig. 1은 어느 특정한 변수 ($J$=21)에서 앙상블 결과를 나타낸다. 여기서 굵은 실선은 truth이며, 가는 실선들은 16개의 앙상블 멤버이다. 여기서 볼 수 있듯이, $F=8$인 경우 이 모델은 카오스적 흐름을 나타내며, truth (굵은 실선)가 앙상블 스프레드 내에서 대부분 존재하고 있음을 알 수 있다.

### 2.2 오차 공분산

첫 번째 작업은 truth와 관측장을 만드는 것이다. 이를 위해, 로렌쯔-95모델을 1,000 시간스텝까지 시간적분한 결과를 버리고 (모델의 spin-up을 고려), 이후 다른 1,000 시간스텝 자료를 이용하여 truth를 만들었으며, 이 truth에 무작위 수를 더하여 관측값을 구하였다 (분산이 0.01이 되도록 조정). 이러한 truth는 모든 모델의 검증에 사용되었으며, 관측값은 분석싸이클에

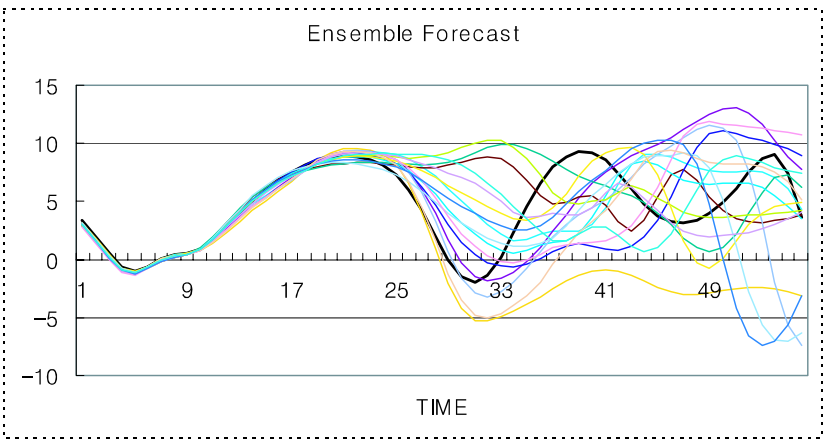


**Fig. 1.** Ensemble simulation with 16 members for a certain variable ($J$=21). The solid thick line represents the truth run and the thin solid lines the ensemble members. The truth is mainly located within the spread of the ensemble, and the spread grows as the forecast time is increased.

서 이용되었다.

모델의 특성, 관측망, 관측오차, 12시간 싸이클시스템 구성에 따른 모델의 기후학적인 오차공분산 (Static error covariance) 을 구하는 절차는 다음과 같다.

1) $t=0$에서 모든 변수에 초기 값으로 관측 값을 부여하고, 앙상블 섭동 추가(초기 무작위 수 발생, 섭동분산은 관측분산과 동일하게 가정)
2) 12시간 모델 적분 수행 ( $\delta t = 0.05$)
3) 12시간 예보장에서 앙상블 평균과 앙상블 예보섭동을 구함
4) 앙상블 예보섭동을 이용하여 오차 공분산 계산 ( $AA^T/(N-1)$, $\boldsymbol{A}$는 섭동, $T$는 transpose, $N$은 앙상블 멤버 수)
5) 앙상블 칼만 필터 수행하여 앙상블 분석 섭동을 구함
6) 분석장을 구하기 위한 분석 과정 수행 (추후 2.3에서 설명)
7) 6)의 분석장에 5)의 분석섭동장을 추가하여 2) ~ 6)을 1,000 시간 스텝까지 싸이클 수행
8) 기후학적인 정적오차공분산은 매번 수행한 각 오차공분산을 평균하여 구함 (Fig. 2).

앙상블을 이용한 자료동화에서 작은 앙상블 멤버를 이용하여 한정된 공간에서 오차 공분산을 계산하는 것은 랭크가 기껏해야 $N$-1이므로, 전체 모델의 차원보다 매우 적어row-rank 문제를 유발할 수 있다. 이 문제를 해결하는 방법은 1) 전구 분석 대신 지역 분석을 행하거나 (Houtekamer and Mitchell, 1998; Hamill *et al.*, 2001; Whitaker and Hamill, 2002), 2) 오차공분산을 임의로 확대 (inflation)하여 분산을 증가시키는 것이다 (Hamill *et al.,* 2001; Anderson 2001; Whitaker and Hamill, 2002). 그러나 본 연구 핵심은 앙상블을 이용한 자료동화라기 보다는 앙상블 섭동을 구성하는 방법을 비교하는 것이므로 여기에서는 이러한 방법들을 사용하지 않았다.

### 2.3 분석과정

2.2 절에서 배경장에서 분석장을 구하는 분석방정식은 다음과 같다

$$X^a = X^f + P^b H^T (HP^b H^T)(Y - HX^f) \tag{2}$$

여기서 $X^a$는 분석장, $X^f$는 배경장, $P^b$는 예보오차 공분산, $H$는 관측연산자, $Y$는 관측을 나타낸다(관측오차 분산이 0.01). 2.2절에서는 $X^f$는 앙상블 평균이며, $P^b$는 $AA^T/(N-1)$이다. $\boldsymbol{A}$ ( $\in R^{J \times N}$, $J$는 변수 수, $N$은 앙상블 멤버 수)는 $X_i^f - \overline{X_i^f}$이다 (첨자 $i$는 각각 앙상블 멤버, $f$는 예보, -는 앙상블 평균을 의미).

본 연구에서는 3차원 변분 자료동화처럼 비용함수를 정의하여 이를 최소화하는 과정과 달리, 분석방정식을 행렬 형태로 구성한 후 직접 풀어 구한다. 앞으로 분석과정은 규준실험에서 나온 자료를 배경장으로 사용하였으며, 예보오차 공분산은 2.2절에서 구한 정적 오

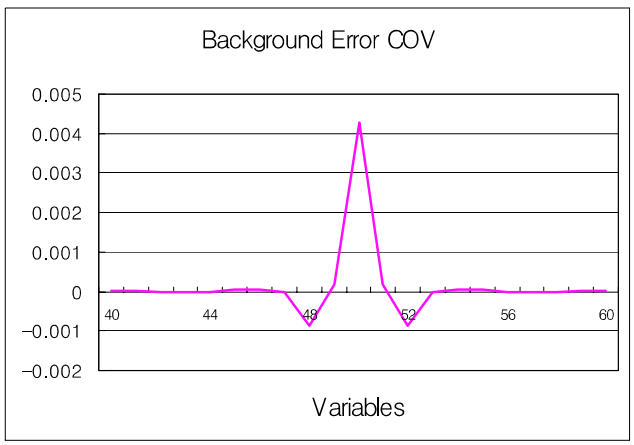


**Fig. 2.** Structure of climatological error covariance after averaging over 1,000 time steps with 16 members. The error is symmetric and non-zero cross correlations are showed to the nearest 2 variables. The averaged error variance (mean of diagonal components) is about 0.0043.

차공분산을 이용하였다. 이는 ETKF를 이용한 flow-dependent 오차공분산을 이용하면 더 나은 오차공분산을 이용할 수 있으나, 본 연구의 목적이 기상청과 같은 자료동화 체계에서 앙상블 섭동기법들을 비교하는 것이 주요한 목적이기 때문이다.

## 3. 앙상블 섭동 구성방법

### 3.1 브레드 벡터

오차 브리딩에 의해 앙상블 섭동을 추출하기 위해, 초기에 무작위섭동 *N/2*개를 발생시킨다. 나머지 *N/2*는 앞의 섭동의 음의 값으로 설정하여 총 *N*개의 초기 섭동을 만든다. 이 섭동을 분석장에 추가하여 12시간을 적분한 후, +/-의 섭동의 결과 차를 구하여 브레드 벡터를 계산한다. 브레드 벡터를 이용한 앙상블 예보장의 평균과 관측 값의 차이에 대한 분산과 브레드 벡터의 스프레드가 같아지도록 이를 재조정한다. 이 값을 다시 분석장에 추가하여 모델을 적분하는 12시간 브리딩 싸이클을 수행한다.

이러한 방법은 현재 기상청에 사용하는 방법 (time lag에 의한 초기섭동 구성)과 비교하면, 초기 섭동을 구성하는 방법에서 차이가 있을 뿐, 기본적으로 Toth and Kalnay (1994)가 제시하였던 브리딩 싸이클 방법과 동일하다.

### 3.2 직교브레드 벡터

앞에서 구한 브레드 벡터는 상호 직교성이 없으며, 따라서 예보기간이 길어질수록 그 에너지가 한 두 모드에 집중되는 현상이 나타난다. 그러므로 적절한 스프레드를 유지시켜 주기 위해서는 브레드 벡터를 회전하여 각각 벡터가 직교성을 갖도록 해야 한다. 본 연구에서 사용된 방법은 다음과 같다.

$X^f$를 예보 앙상블 (섭동성분) 이라고 하자. 여기에 전치행렬 (transpose)을 곱하여 대칭행렬 ( $X^f X^{f^T}$)를 구성한다. 이 공분산행렬은 실수의 대칭행렬이므로 positive definite이 된다 ( $\langle u, Au\rangle = u^T Au \rangle 0$, $A$는 실수의 대칭행렬 ( $A^T = A$ )이며 $u$는 영이 아닌 실수). 이를 특성값 분해 (eigen value decomposition)를 하면, 다음과 같다.

$$X^f X^{f^T} = V \Lambda V^T \tag{3}$$

여기서 $V$는 특성벡터 (eigen vector)이며, $\Lambda$는 특성값 (eigen value)로 구성된 대각행렬이다. 이후, $X^f$를 특성벡터로 좌변항을 변형하면,

$$V^T X^f X^{f^T} V = \Lambda \tag{4}$$

가 된다. 마지막으로 대각 행렬 $\Lambda^{-1}$을 좌우변에 곱하고 정리하면,

$$(\Lambda^{-1/2} V^T X^f)(\Lambda^{-1/2} V^T X^f)^T = I \tag{5}$$

가 된다. 따라서 브레드 벡터 $X^f$를 행렬 $\Lambda^{-1/2} V^T$로 변환하면, 변환된 행렬 $\Lambda^{-1/2} V^T X^f$을 이용하여 직교 브레드 벡터를 얻을 수 있다.

현재 기상청은 브레드 벡터가 상호 준 직교성을 갖도록 하기 위해 요인회전 (factor rotation)개념을 이용한 varimax rotation을 선택하고 있으며, 이 방법은 IMSL의 내장 함수인 FROTA를 이용하여 수행되고 있다 (기상청, 2005). 두 기법 모두 직교하는 축으로 벡터를 회전하여 직교벡터를 구하는 개념이다. 다만, 본 연구에서는 차원이 작아 행렬의 직접적인 연산을 통해 직교벡터를 구하는 방법을 사용하였다.

### 3.3 ETKF

ETKF (Ensemble Transform Kalman filter)는 Monte Carlo 방법을 이용하여 한정된 수의 앙상블로 오차 공분산을 추정하기 때문에 차선의 최적 (suboptimal) 칼만필터 기법이다 (Kalman, 1960; Kalman and Bucy, 1961; Daley, 1991). 칼만필터의 분석오차 공분산 ( $P^a$)은 다음과 같다.

$$P^a = (I - KH)P^f \tag{6}$$

여기서 $P^f$는 예보오차 공분산이며, $K$는 Kalman Gain 즉. $K = P^f H^T / (HP^f H^T + R)$이다.

원래 칼만필터에서 예보오차 공분산 $P^f$는 $P_k^f = MP_{k-1}^a M^T + Q$ (여기에서, $k$는 시간, $M$은 비선형 모델의 선형연산자, $Q$는 모델오차)로 구한다. 즉 전 시간의 분석오차가 선형연산자 $M$으로 시간전개가 되고, 여기에 모델 오차가 포함되어 예보오차 공분산를 추정할 수 있다는 개념이다. 그러나 이를 풀기 위해서는 선형 및 수반 연산자가 필요하고, 실제 대기모델의 차원이

매우 크기 때문에 이를 직접 풀기가 어려워진다. 아울러 비선형모델을 선형적으로 가정하는 한계가 있다. 그러나 앙상블에 기반한 칼만필터는 접선형 선형연산자나 수반 모델도 필요 없고, 선형의 가정이 없이 추론적인 Monte Carlo 기법으로 상태공간을 추정할 수 있다. 즉, 앙상블 평균이 가장 최적으로 추정된 상태공간이며, 이의 스프레드가 오차분산으로 간주된다는 것이다. 그러므로 앙상블 기반의 오차공분산은 예보 섭동 ( $X^f$ )를 이용하여 표현된다. 여기서 $X^f$는 각 앙상블 멤버에서 앙상블평균을 뺀 값이다. 즉,

$$P^f = Z^f Z^{f^T} \tag{7}$$

이며, 여기서 $Z^f = X^f/\sqrt{(N-1)}$로 정의한다. 그리고 분석된 섭동 $Z^a = Z^f T$로 놓으면 (변환 행렬 $T$는 다음에 정의), 분석 오차 공분산은 다음과 같다.

$$P^a = Z^a Z^{a^T} = (Z^f T)(Z^f T)^T = Z^f T T^T Z^{f^T} \tag{8}$$

Bishop *et al.* (2001)에 따르면, 변환행렬 $T$는 다음과 같이 정의된다.

$$T = C(\Gamma + I)^{-1/2} \tag{9}$$

여기서, $C$는 $(HZ^f)^T R^{-1} HZ^f$으로 정의된 행렬의 특성벡터이며, $\Gamma$는 이 행렬의 특성값를 나타낸다. $I$는 단위행렬이다. 따라서 ETKF는 관측점에서 새로운 분석 섭동을 구하는 것으로 앙상블 멤버에서 구한 예보섭동의 공분산을 관측오차 ($R$은 대각행렬)로 정규화한 행렬을 이용한다. 이 과정에서 특성벡터가 이용되어 생성된 분석섭동은 직교성을 갖게 된다. 아울러 이러한 특성을 이용하여 ETKF는 앙상블 멤버 생성뿐만 아니라 목표관측망 (adaptive or targeted observation network) 평가에서도 활발히 이용되고 있는 실정이다 (Bishop *et al.*, 2001; Majumdar *et al.*, 2002).

그러나, 분석 사이클이 진행될수록 분석오차를 최소화하는 과정을 반복하므로 섭동의 스프레드는 점차 작아지는 경향이 있다. 이를 해결하기 위하여 Dee (1995)가 제안한 바와 같이 이노베이션 벡터의 제곱이 관측점에서 예보오차 공분산과 관측오차 공분산의 합과 같아지도록 다음 식의 $\alpha$를 조정한다. 즉,

$$d^T d = trace(H\alpha P^f H^T + R) \tag{10}$$

이다. 여기서 이노베이션($d$)는 $d = Y - HX^f$이다. 본 연구에서는 Wang and Bishop (2003)에 따라 이 값을 구하였다. 따라서 최종 분석 섭동은

$$X_k^a = X_k^T T_k \Pi_k \tag{11}$$

이다. 여기서 확대인자 (inflation factor, $\Pi$)는 $\Pi_k = \sqrt{\alpha_1 \alpha_2 \ldots \alpha_k}$이며, $k$는 시간 스텝을 표시한다. Fig. 3은 본 연구에서 계산된 확대인자의 시계열을 나타낸다. 초기에 이 값은 큰 값으로 출발하여 점차 안정화되는 경향을 나타내며, 0.64에서 3.4사이에서 변화하면서 평균 1.33의 값을 나타내고 있다.

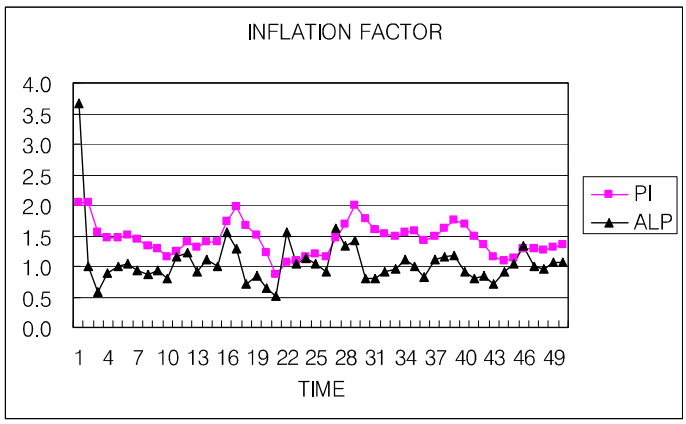


**Fig. 3.** Time series of the inflation factor (PI( $\Pi$), ALP ( $\alpha$)). The x-axis represents time for the first 50 time steps and the *y*-axis is the inflation factor. At early times, the inflation factor highly fluctuates , but it comes to be stable after 3 weeks (about after 42 time step). The value is ranging from 0.64 to 3.4 with mean of 1.33.

## 4. 결과

앞에서 말한 바와 같이, 기상청에서는 브레드 벡터가 상호 준 직교성을 갖도록 하기 위해 요인회전을 하고 있으며, 현재 수행되고 있는 앙상블 시스템은 Fig. 4와 같다. 초기 섭동장을 12시간 자라게 한 후, 섭동을 추출한다. 여기에서 추출된 섭동장이 준 직교성을 갖도록 회전시킨 후, 이를 이용하여 모델을 다시 12시간 적분하면 최종적으로 24시간 동안 성장한 브레드 벡터를 얻는다. 이를 분석장에 첨가하여 10일간의 앙상블 예보가 수행된다 (문선옥 등, 2006; 기상청, 2005).

2004년 7월 브레드벡터를 요인 회전한 경우의 앙상블 예보평균오차는 기존의 체계에 비해 약 2-4% 감소하였으며, 스프레드는 3-5일 예보에서 약 5% 정도가 증대되었음을 알 수 있다. 한편 예보기간이 길어질수록 스프레드의 증가율이 RMSE 감소율에 비해 약 두 배 정도 크게 나타남을 볼 수 있다 (Table 1).

요인 회전 이후에 성능이 향상된 기상청의 앙상블 시스템과 ETKF에 의한 앙상블 체계를 비교하기 위해서 무엇보다도 ETKF를 기상청의 앙상블체계에 이식하여 동시에 이 두 방법을 수행하여 평가하는 것이 바람직할 것이다. 그러나 이러한 경우에 많은 변수들이 있고, 이를 정량적으로 평가한다는 것이 쉬운 일이 아니다. 그러므로 우선 현업 앙상블 체계에 ETKF를 이식하기 전에, 간단한 로렌쯔 모델에서 직교브레드 벡터와 ETKF를 다각적으로 비교하는 것이 필요하다. 이를 통해, ETKF의 민감도와 그 수행에 대한 장점과 한계점을 파악하며, ETKF가 직교브레드 벡터와 어느 부분에서 차이가 많이 나는지, 어느 부분에서 효과를 기대할 수 있는 지 등을 미리 예상할 수 있다.

이 실험을 위하여, 약 1년간 12시간 마다 싸이클링하면서 모델을 수행하였다. 앞 절에 구한 배경오차 공분산을 이용하여 3dvar와 같은 맥락으로 분석체계를 구축하였으며, 매 분석시간 마다 브레드 벡터, 직교브레드벡터, ETKF를 통해 섭동장을 구하고, 16개 앙상블을 14일까지 적분하였다. 모델의 변수는 100개이며

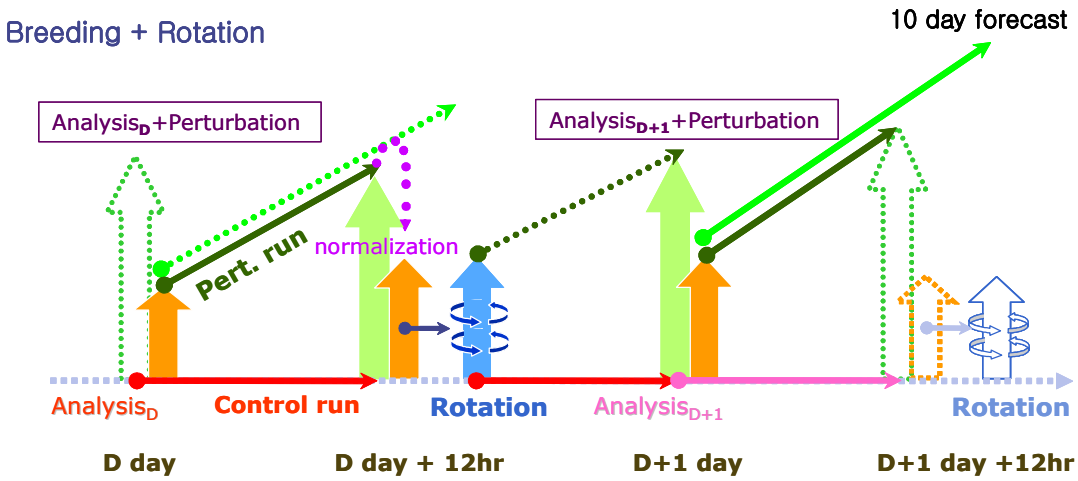


**Fig. 4.** Schematic diagram of the breeding cycle which is currently operated in Korea Meteorological Administration (KMA) (KMA, 2005).

**Table 1.** Reduction of RMSE of Ensemble mean and the increment rate of ensemble spread from the factor-rotated bred vectors compared to the previous (non-rotated) bred vectors (KMA, 2005).

| 예보 / IR(%) | 1일 | 2일 | 3일 | 4일 | 5일 | 6일 | 7일 | 8일 |
|---|---|---|---|---|---|---|---|---|
| RMSE | -3.0 | -4.0 | -3.9 | -3.6 | -2.3 | -1.7 | -2.1 | -1.7 |
| Spread | 2.8 | 4.5 | 4.7 | 5.1 | 4.9 | 4.1 | 3.8 | 4.7 |

(격자점이 100개에 해당) 모든 지점에 관측값이 있다고 가정하였다.

Fig. 5는 모델 수행 중 특성값 (eigen value)이 가장 큰 경우에 브레드벡터 회전 전・후의 각 변수 (격자점에 해당, *J*)에서 스프레드를 나타낸 것이다. 전체적으로 스프레드의 공간적 변동이 일치하고 있다는 것을 알 수 있다. 그러나 스프레드의 크기에서 차이가 나는 것은 1) 회전시 이용한 행렬이 공분산행렬이고, 2) 변형 행렬에 특성값의 제곱근이 나누어져 있기 때문이다. 따라서 정규화된 행렬을 이용하면 스프레드의 변동 폭에 큰 차이를 보이지 않을 것이다. 그러나 본 연구에서는 이렇게 구한 스프레드를 식(10)을 이용하여 그 폭을 다시 조정하기 때문에 문제가 되지 않는다 (즉, 이 루틴에서는 각 벡터를 직교화를 하는 것이며, 그 변동의 폭은 이노베이션의 크기와 오차가 같도록 재조정). 이 그림에서 회전 이후에도 그 공간 변동의 변화가 일치하는 것을 보아 일단 브레드 벡터의 기본적인 성질을 보존하고 있는 것으로 보인다. 회전성분은 수식으로 완전하게 상호 직교성을 지니게 되므로, 회전 이후에 직교성을 가지고 회전 전의 브레드 벡터의 성질을 유지한 직교 브레드 벡터를 얻었다고 생각할 수 있다.

Fig. 6은 예보시간 14일 (336시간)까지 앙상블 평균에 대한 RMSE를 나타낸다. 전체적으로 규준 수행보

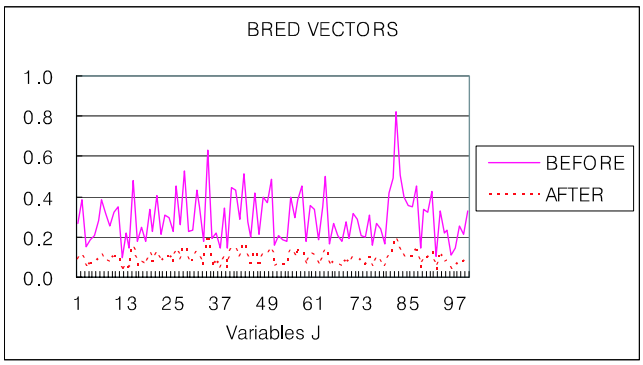


**Fig. 5.** An example of spreads over the variables (grid points) from original bred vectors (solid line) and from rotating bred vectors (dotted line) to make them orthogonal.

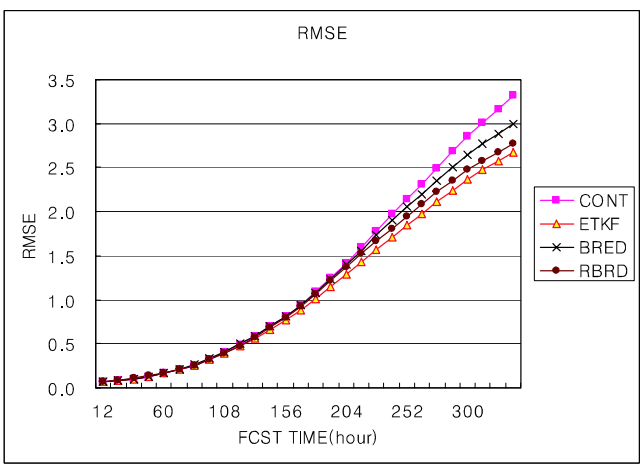


**Fig. 6.** RMS errors of control run and the ensemble means. ETKF shows the best performance and the orthogonal bred vectors can largely reduce the error about 5% and 9% for 5-day and 10-day forecast, respectively. The result is consistent with the real experiment of KMA. CONT, ETKF, BRED, and RBRD represent control run, ETKF, the original bred vectors (non-rotated), and orthogonal bred vectors after rotation, respectively.

다는 앙상블 수행이 오차가 작은 결과를 나타내고 있으며, 앙상블 예보 중에 ETKF, 직교브레드 벡터, 브레드 벡터 순서로 오차가 크게 나타나고 있다. ETKF의 오차를 1로 했을 때, 브레드 벡터는 120시간예보에 1.06, 336시간 1.12를 보이고 있으며, 직교 브레드 벡터의 경우 각각 1.01, 1.03을 나타내고 있다. 그러므로 직교브레드벡터의 수행능력은 5일 예보에서 브레드 벡터에 비해 5% 정도, 14일 예보에서 9% 정도 향상된 능력을 보이는 것으로 나타났다. 이는 실제 현재 기상청의 실험 운영결과에서 나타난 결과와 비교해 볼 때, 실제 대기모델에서 운영한 결과보다 본 연구에서는 약 두 배 정도의 큰 향상율을 보이고 있었다. 그러나 4일 예보 부근의 향상율의 경우 두 실험 결과가 거의 일치하였고, 전체적으로 ETKF가 가장 우수한 성능을 보이고 있다.

브레드 벡터의 단점은 예보시간이 길어질수록 그 스프레드가 작아진다는 것이다. 이를 해결하는 방편으로 브레드벡터를 회전하여 각각 벡터들과 직교하게 만드는 것이다. 이론적으로 분석시간에 앙상블평균의 오차와 앙상블 스프레드의 비는 1로 조정하기 때문에 각 멤버가 직교한다면 예보기간에 거의 이 비율을 유지해야 한다. 이를 보인 것이 Fig. 7이다. 여기서 $x$축은 예보기간을 $y$축은 스프레드를 앙상블평균오차로 나눈 값이다. 브레드 벡터는 이 비율이 예보기간이 길어질수록 급속하게 감소하여 120시간에는 약 0.88, 240시간에는 0.77로 떨어진 후, 이후에는 거의 비슷한 값을 보이고 있다. 반면에 직교브레드 벡터는 120시간에 0.97, 240시간에 0.91을 보이며, 이후에도 거의 0.9정도의 값을 유지하고 있다. 한편 ETKF는 거의 0.98을 유지하다가 240시간 이후에 0.95정도의 값을 보이고 있어, 다른 두 개의 방법보다 가장 우수한 결과를 나타내고 있다. 이를 통해 브레드벡터를 회전하면 회전 전의 브레드 벡터에 비해 스프레드를 유지할 수 있는 이점이 있는 것으로 보인다 (120시간에 약 10% 정도 스프레드 향상). 실제 실험에서도 예보기간이 길어질수록 앙상블 평균오차의 감소율에 비해 스프레드의 증가율이 2배 정도 되었다는 사실과 일치한다. 한편 스프레드 유지에 있어서도 앞 절에서와 같이 직교 브레드 벡터보다 ETKF의 결과가 조금 우수한 경향을 나타내고 있으며, 특히 5일 이후의 예보에서 그 차이가 조금 더 커지는 양상을 보이고 있다.

만일 각 앙상블멤버를 예보장의 실제 확률밀도함수 (pdf)에서 균등하게 샘플링을 했다면, 앙상블멤버로 나타내어지는 확률밀도 함수는 실제 확률밀도함수와 같은 통계적인 특징을 가지게 된다. 이러한 정도를 나타내 줄 수 있는 방법이 rank histogram이다. 이는 각 앙상블멤버를 예보 값의 크기순으로 나열한 다음, 앙상블 멤버 구간 중에서 구한 실제값 (truth)이 어디에 위치하는 지를 통계적으로 나열한 것이다. 만일 앙상블로 대표되어지는 확률밀도함수 (pdf)에 샘플링 오차가 없다면, rank histogram의 모든 구간에 균등하게 truth가 위치하게 될 것이다. 만일 truth가 확률밀도함수 (pdf)를 벗어나 위치하면 양 끝단의 rank가 큰 값을 갖게 되며, 이는 앙상블예보의 스프레드를 벗어나는

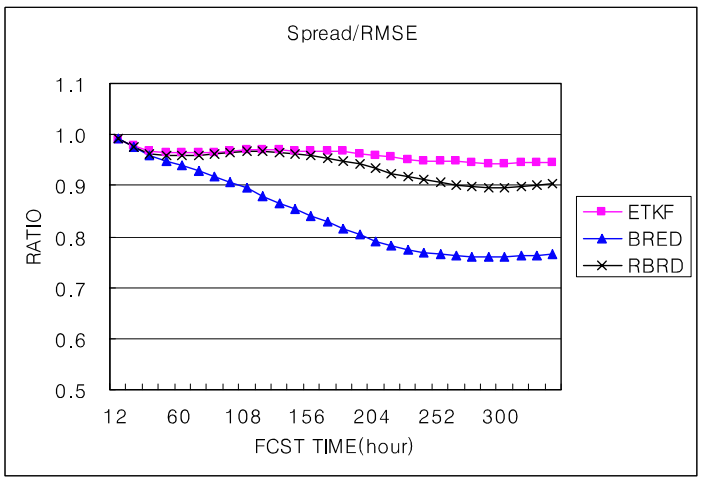


**Fig. 7.** The ratio of RMS ensemble error to the spread of the ensemble. The ratio of ETKF has nearly 0.97, and that of orthogonal bred vectors is almost the same as ETKF within 5 days. But the difference of ratio between ETKF and orthogonal bred vectors is getting increased after 5-day forecast. It appears that ETKF is superior to the others.

관측값이 많았다는 증거이다. 이런 경우 종종 앙상블 예보는 실패한다.

Fig. 8은 rank histogram을 보여준다. 이 그림에서 브리딩 벡터는 양끝 단의 값이 무척 커서 다른 두 앙상블 방법에 비해 현저히 그 수행능력이 떨어짐을 알 수 있다. 이는 예보기간이 길어질수록 스프레드가 작아지기 때문에, 5일 이후의 실제 값이 브레드 벡터의 스프레드 범위를 벗어나는 일 (outlier)이 자주 발생하기 때문이다. 반면에 ETKF, 직교 브레드 벡터는 outlier가 많이 줄어들었으며, 브레드 벡터에 비해 다른 앙상블 멤버에 균등하게 histogram이 분포하고 있음을 알 수 있다. 이중에서도 ETKF는 직교브레드 벡터에 비해 더욱 균등한 분포와 outlier 수가 적어 가장 좋은 성능을 보인다고 볼 수 있다.

각 예보 값의 특성값 (eigen value)를 통해 예보 앙상블의 독립성 정도를 가늠할 수 있다. 다시 말하면, 예보장 섭동의 변동성이 한두 모두에 집중되어 있으면, 그 특성모드 (eigen mode)의 특성값이 매우 큰 값을 나타낸다. 이는 특성모드의 특성값이 차지하는 비율은 전체 변동성에서 그 모드가 차지하는 부분을 설명하기 때문이다.

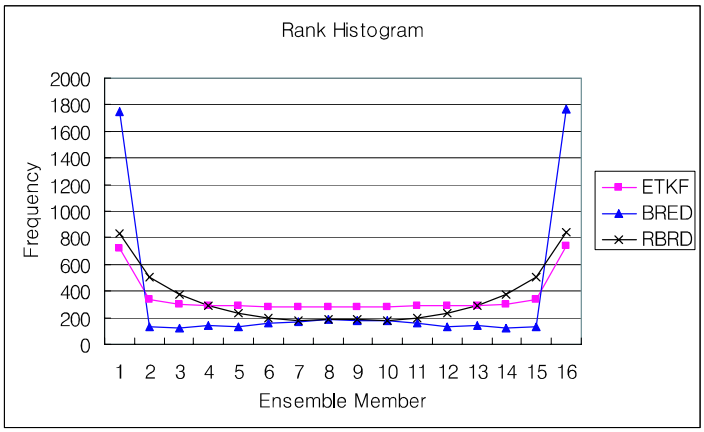


**Fig. 8.** The rank histogram from ETKF, bred vectors and orthogonal bred vectors. The orthogonal bred vectors with rotation show that outliers can be progressively reduced compared to bred vectors. It also appears that ETKF has the best performance among these strategies.

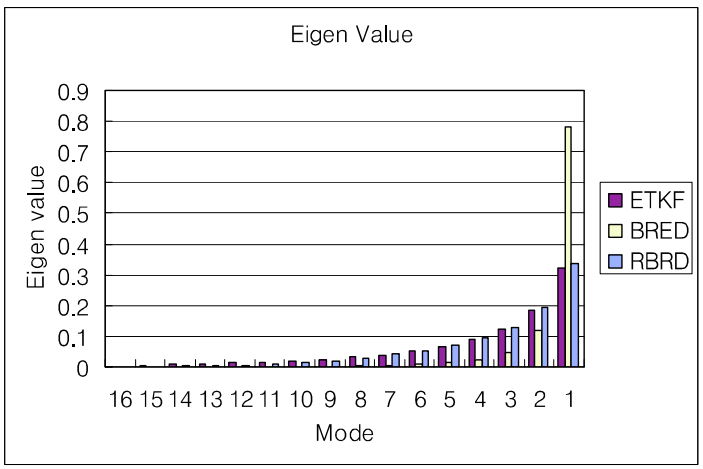


**Fig. 9.** Normalized eigen values for the forecast perturbation fields from ETKF, bred vectors, and orthogonal bred vectors with rotation. The leading 2 eigen modes from bred vectors consist of 90% of perturbation variance, but 8-9 leading eigen modes from ETKF and orthogonal bred vectors explain 90% of the total variance of perturbation. Even though the rotation of bred vectors can also make the perturbations more independent from one another, ETKF has a slight better performance than the orthogonal bred vectors.

Fig. 9는 앙상블 예보장에서 각 특성 모드별 특성 값이 전체에서 차지하는 비율을 나타낸다. 이 그림에서 보듯이 브레드 벡터는 첫 번째 모드에 약 78%, 두 번째 모드에 약 12%정도가 되어 선두 두개의 모드 (leading eigen mode)가 전체 섭동 분산의 90%를 차지하고 있음을 알 수 있다. 반면에 ETKF와 직교브레드벡터에서는 첫 번째 모드가 약 32-34%, 두 번째 모드가 18-20%, 세 번째 모드가 12% 정도 상대적으로 균등하게 분포하며, 8-9개 모드의 합이 전체 섭동 분산의 90% 정도를 설명하는 것으로 나타나고 있다. 이중에서도 ETKF가 직교 브레드벡터보다 약간 나은 결과를 보이고 있음을 알 수 있다.

이러한 결과를 통해, 초기 섭동의 직교성이 전체 앙상블 예보 성능에 어느 정도 영향을 미치는 것인가를 파악할 수 있다. 즉 브레드 벡터는 예보기간이 길어질수록 선두 한두 모드에 에너지가 집중되는 현상이 나타나 스프레드가 줄어들며, 결과적으로 outlier가 증대된다. 앞 절에서 언급하였듯이 이는 앙상블멤버로 구성된 예보 확률밀도함수가 실제의 확률밀도함수 (pdf)를 적절히 표현하지 못한다는 의미이다. 즉, 앙상블예보 시스템의 성능 저하를 의미한다.

지금까지 ETKF와 브레드 벡터, 직교브레드 벡터에 대한 앙상블 예보 성능을 검증하였다. 이번에는 앙상블 멤버를 자료동화에 이용하는 문제에 대하여 간단히 논의하고자 한다.

최근의 자료동화기법은 크게 변분자료동화 (3dvar, 4dvar)와 앙상블을 이용한 자료동화 두 종류로 나뉠 수 있다. 변분자료 동화는 장기간에 걸쳐 배경오차를 계산한 다음 이 정적 배경오차 공분산을 이용하여 분석장을 배경오차와 관측오차 사이에서 최적화 하도록 비용함수를 최소화하는 과정으로 이루어진다. 반면에 앙상블에 의한 자료동화는 예보된 앙상블 멤버를 이용하여 배경오차공분산을 매 분석시간마다 계산하여 이를 이용하여 최적화하는 과정을 말한다. 변분자료 동화는 대부분의 현업기관에서 사용되고 있는 매우 안정한 방법이나, 배경오차가 변하지 않기 때문에 매일 매일 달라지는 배경 오차의 구조를 반영하지 못하는 단점이 있으며, 이를 어느 정도 보완할 수 있는 4차원 변분자료동화의 경우는 접선형 선형연산자와 수반연산자 (adjoint operator)가 필요하고, 계산 시간이 많이 걸린다는 단점이 있다. 앙상블 자료동화의 경우는 매일 매일 변하는 배경오차를 반영하고, 4차원변분자료 동화와 달리 접선형 선형연산자와 수반연산자가 필요 없다는 장점이 있으나, 몇 개의 앙상블멤버를 이용하여 오차의 전체 확률밀도함수를 대표해야 하기 때문에, 샘플링 오차에 의한 분석장의 질 저하를 불러 올 수 있으며, 불확실성이 커서 아직은 현업기관에서 이를 현업으로 사용하지 못하고 있는 실정이다.

따라서 가장 안정한 변분자료동화기법에 매일 매일의 배경오차 구조를 반영할 수 있는 하이브리드 변분자료동화 기법이 순수한 변분자료나 앙상블자료동화보다 나은 결과를 보인다는 많은 연구가 있다 (Barker, 1999; Barker *et al*., 2005; Etherton and Bishop, 2004; Hamill and Snyder, 2000). 따라서 앙상블에서 생산되는 섭동을 이용하여 매일 매일의 배경오차를 변분자료동화에 반영하여 초기자료의 품질을 향상시키기 위해서 무엇보다도 중요한 것은 앙상블 멤버에 의해 구성되는 배경오차의 구조이다.

Fig. 10은 브레드 벡터, 직교 브레드 벡터, 그리고 ETKF의 앙상블 멤버를 이용하여 계산한 배경오차 공분산이다. 세 기법 모두 유사한 오차 구조를 가지고 있는 것으로 보인다. 이는 본 연구에서는 브레드 벡터라도 12시간 예보장 (배경장)에서 RMSE나 스프레드, 특성벡터 (eigen vector)의 분포에서 크게 차이를 보이고 있지 않기 때문이다. 다만, 브레드 벡터에 비해 관측점에서 멀어진 지역에서의 상관 (특히 그림에서 $x$ 축 44, 46과 54, 56 지점)에 진동이 ETKF와 직교브레드 벡터보다 커 ETKF나 직교브레드 벡터가 좀 더 나은 오차 공분산을 제공할 수 있을 것으로 생각되나, 본 연구에서는 그 차이가 매우 작게 나타났다. 그러나 이러한 오차 구조는 수치모델, 앙상블멤버 수, 관측오차, 관측망 분포 등에 따라 현실적으로는 크게 차이가 날 수 밖에 없기 때문에 좀 더 세밀한 연구가 필요한 시점이라고 볼 수 있다.

## 5. 요약 및 결론

본 연구에서는 대기의 카오스적 현상을 모의할 수 있는 간단한 로렌쯔-95 모델 상에서 브레드 벡터, 직교브레드벡터, ETKF에 의한 섭동을 비교 분석하였다. 대기 모델의 변수는 100개 (공간 격자 100개인 1차원 모델)이며, 12시간 분석 싸이클이 1년간 진행되었다. 앙상블 멤버는 16개이며, 예보기간은 매 싸이클마다 14일까지이다. 분석시스템은 3차원자료동화와 같이 정적오차 공분산을 이용하였다. 이와 같은 평가를 위한 싸이클을 수행하기 전에, 모델을 2,000 시간 스텝을

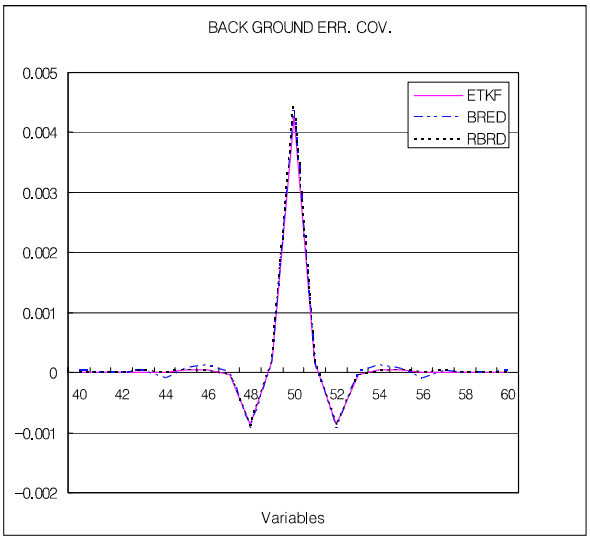


**Fig. 10.** Background error structures obtained from ETKF, the bred, and the orthogonal bred vectors. It appears that the error structures show almost similar features, but ETKF and orthogonal bred vectors are slightly better than bred vectors.

먼저 수행하여, 앞 1,000 시간스텝의 자료를 버리고, 나머지 자료로 truth와 truth에 관측오차 (분산 0.01)를 무작위로 더하여 관측 값을 만들고, ETKF를 통해 나온 앙상블멤버로 배경오차 공분산을 생성하였다.

본 연구에서 브레드 벡터는 초기에 무작위섭동 $N/2$개를 발생시키고, 나머지 $N/2$는 앞의 섭동의 음의 값으로 설정하여 총 $N$개의 초기 섭동을 만든 후, 이 섭동을 분석장에 추가하여 12시간 적분한 후, +/-의 섭동의 결과 차이를 이용하여 구하였다. 이를 직교시키기 위하여 특성값 분해 (eigen value decomposition)를 이용하여 각 브레드 벡터를 특성벡터에 전사 (profection)하고, 특성값의 제곱근으로 나누어서 직교벡터를 구한 다음, 그 섭동의 크기는 이노베이션의 제곱과 오차 (배경오차 및 관측오차)가 같아지도록 재조정되었다 (Dee, 1995; Wang *et al.*, 2003). ETKF는 Bishop *et al.* (2001)의 방법에 따라 계산되었다. 이 방법은 예보된 앙상블을 이용하여 배경오차 공분산을 추정하고, 칼만 필터의 분석오차방정식을 이용하여 새로운 섭동을 추정하는 제곱근 알고리듬 (square root algorithm) 방법이다.

이러한 섭동을 평가하기 위하여, 앙상블 평균의 RMSE, 앙상블 스프레드의 적절성을 알아보기 위하여 스프레드와 앙상블 평균오차의 비, 앙상블멤버로 추정된 확률밀도함수 (pdf)의 적정성을 가늠할 수 있는 rank histogram, 그리고 각 멤버들의 독립성을 판단할 수 있는 예보장의 특성값 분포 등이 조사되었다. 이 결과를 요약하면,

1) 직교브레드벡터의 RMSE는 5일 예보에서 브레드벡터에 비해 5% 정도, 14일 예보에서 9% 정도 향상된 능력을 보이는 것으로 나타났으며, 전체적으로 ETKF가 가장 우수한 성능을 보이고 있었다.
2) 스프레드와 앙상블 평균오차의 비를 보면, 브레드벡터는 이 비율이 예보기간이 길어질수록 급속하게 감소하였으며 (120시간에 약 0.88, 240시간에 0.77), 직교브레드 벡터와 ETKF는 120시간까지 0.97을 유지하여 적절히 스프레드가 유지됨을 알 수 있었다. 240시간 예보시간까지의 관점에서 보면 ETKF가 0.95정도를 유지하고 있어, 직교브레드 벡터보다 스프레드 유지가 장기간 지속되는 것을 알 수 있었다.
3) Rank histogram을 보면, 브레드 벡터는 예보기간이 길어질수록 스프레드가 작아지기 때문에, 5일 이후의 실제 값이 브레드 벡터의 스프레드 범위를 벗어나는 일 (outlier)이 자주 발생하여 outlier가 매우 크게 나타났으나, 반면에 ETKF와 직교 브레드 벡터는 브레드벡터에 비해 outlier가 매우 적게 나타났다. ETKF는 직교브레드 벡터에 비해 더욱 균등한 분포와 outlier 수가 적어 가장 좋은 성능을 보였다.
4) 예보장의 특성값를 살펴보면, 브레드 벡터는 선두 두 개의 모드 (leading eigen mode)가 전체 섭동 분

산의 90%를 차지하고 있었으며, ETKF와 직교브레드벡터에서는 상대적으로 균등하게 분포하여 8-9개 모두까지의 합이 전체 섭동 분산의 90% 정도를 설명하는 것으로 나타났다. ETKF가 직교브레드 벡터보다 약간 나은 결과를 보이고 있었다.

한편, 브레드벡터를 회전하여 직교성을 확보하면, 회전 전에 비해 모든 부분에서 향상된 효과를 보였으며, 특히 스프레드의 증가율이 RMSE의 감소율에 거의 두 배에 달하여, 이 부분에 가장 효과가 큰 것으로 파악되었다. 이러한 간단한 모델에서의 결과는 실제 현업에서 수행된 모델의 결과와 잘 일치되었다.

결론적으로 앙상블 예보에 대해서 ETKF가 모든 면에서 가장 우수한 성능을 지닌 것으로 파악되었으나, 앙상블을 이용한 매일 매일의 배경오차를 변분자료동화에 도입하는 부분에서는 좀 더 많은 연구가 필요한 것으로 판단되었다. 이는 세 가지 섭동 방법으로 구한 배경오차 공분산이 비슷한 결과를 보이고 있었기 때문이었다. 배경오차 공분산은 수치모델의 특성, 앙상블 멤버 수, 관측오차, 관측망 분포 등에 따라 현실적으로는 크게 차이가 날 수 밖에 없기 때문에, 하이브리드 변분자료동화를 위해서 어떤 섭동전략이 나은 지에 대해서 좀 더 면밀한 연구가 진행되어야 할 것이다.

## 감사의 글

본 연구는 인사위원회 국비장기파견 프로그램에서 지원된 “앙상블 칼만필터를 이용한 자료동화 기법 연구”의 일환으로 진행되었습니다. 꼼꼼히 논문을 심사하여 오류를 지적해 주신 두 분의 심사위원님들께 감사의 말씀을 드립니다.

## 참고문헌

기상청, 2005: 선진 수치예보 시스템 구축을 위한 기반 기술 개발 (VI), 6-65

문선옥, 전은희, 김동준, 박영연, 2006: 앙상블 예보 해석 능력 보강을 위한 앙상블 예보 가이던스, 기상청 수치예보과 기술노트, 5-6

Anderson, J. L., 2001: An ensemble adjustment Kalman Filter for data assimilation. *Mon. Wea. Rev.,* **129**, 2884-2903

Barker, D. M., 1999: The use of synoptically-dependent background error structures in the UKMO’s 3dvar system. *VAR Scientific Documentation Paper 26*, UKMO Bracknell, UK.

_____, Y. H. Kuo, J. Bresch, Y. Guo, Z. Liu, S. Rizvi and Q. Xiao, 2005: Development of a unified 3dvar data assimilation system. *Final report submitted to KMA*, 46 pp.

Bishop, C.H., B. J. Etherton and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev*., **129**, 420-436.

Bowler, N. E. 2006: Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model. *Tellus,* **58A**, 538-548.

Buizza, R., J. Tribbia, F. Molteni and T. Palmer, 1993: Computation of optimal unstable structures for a NWP model. *Tellus,* **45A**, 388-407.

Daley, R., 1991: *Atmospheric data analysis.* Cambridge University Press, 457 pp.

Dee, D. P., 1995: On-Line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev*., **123**, 1128-1196.

Du, J., S. L. Mullen, and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev*., **125**, 2427-2459.

Etherton, B. J. and C. G. Bishop, 2004: Resilience of hybrid ensemble/3dvar analysis schemes to model error and ensemble covariance error. *Mon. Wea. Rev*., **132**, 1065-1080.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo Methods to forecast error statistics. *J. Geophys. Res*., **99(C5)**, 10,143-10,162.

Hamill, T. M., C. Snyder, and R. E. Morss, 2000: A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles. *Mon. Wea. Rev.*, **128**, 1835-1851.

_____, J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776-2790.

_____ and C. Snyder, 2000: A hybrid ensemble Kalman filter-3d variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905-2919.

Houtekamer, P. L. and J. Derome, 1995: Methods for ensemble prediction. *Mon. Wea. Rev.,* **123**, 2181-2196.

_____ and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev*., **126**, 796-811.

Kalman R., 1960: New approach to linear filtering and predicted problems. *J. Basic Eng*., **82**, 35-45.

_____ and R. Bucy, 1961: New results in linear prediction and filtering theory. *Trans. AMSE. J. Basic Eng*., **83D**, 95-108.

Majumdar, S. J., Bishop, C.H., B. J. Etherton and Z. Toth, 2002: Adaptive sampling with the ensemble transform Kalman filter. Part II: Field program implementation. *Mon. Wea. Rev.,* **130**, 1356-1369.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev*., **131**, 1485-1490.

Toth, Z and E. Kalnay, 1994: Ensemble forecasting system at NCEP: the gerneration of perturbations. *Bull. Amer. Meteor. Soc*.. **74**, 2317-2330.

_____ and _____, 1997: Ensemble forecasting and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.

Wang, X., H. Bishop, and J. Julier, 2003: A comparison of breeding and ensemble transform Kalman filter. *J. Atmos. Sci*., **60**, 1140-1158.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev*., **130**, 1913-1924.