Article

Double Gyre 모형 해양에서 앙상블 칼만필터를 이용한 자료동화와 쌍둥이 실험들을 통한 민감도 시험

김영호^{1*} · 유상진² · 최병주³ · 조양기⁴ · 김영규⁵

¹한국해양연구원 기후 · 연안재해연구부
(425-600) 경기도 안산시 안산우체국 사서함 29

²기상청

(156-720) 서울시 동작구 기상청길 45

³군산대학교 해양학과

(573-701) 전라북도 군산시 미룡동 산 68

⁴전남대학교 지구환경과학부

(500-750) 광주시 북구 용봉동 333

⁵국방과학연구소

(645-600) 경상남도 진해시 사서함 18

Implementation of the Ensemble Kalman Filter to a Double Gyre Ocean and Sensitivity Test using Twin Experiments

Young Ho Kim^{1*}, Sang Jin Lyu², Byoung-Ju Choi³, Yang-Ki Cho⁴, and Young-Gyu Kim⁵

¹*Climate Change & Coastal Disaster Research Department, KORDI
Ansan P.O. Box 29, Seoul 425-600, Korea*

²*Korea Meteorological Administration
Seoul 156-720, Korea*

³*Department of Oceanography
Kusan National University, Gunsan 573-701, Korea*

⁴*Department of Oceanography
Chonnam National University, Gwangju 500-757, Korea*

⁵*Agency for Defense Development
Jinhae P.O. Box 18, GyungNam 645-600, Korea*

Abstract : As a preliminary effort to establish a data assimilative ocean forecasting system, we reviewed the theory of the Ensemble Kalman Filter (EnKF) and developed practical techniques to apply the EnKF algorithm in a real ocean circulation modeling system. To verify the performance of the developed EnKF algorithm, a wind-driven double gyre was established in a rectangular ocean using the Regional Ocean Modeling System (ROMS) and the EnKF algorithm was implemented. In the ideal ocean, sea surface temperature and sea surface height were assimilated. The results showed that the multivariate background error covariance is useful in the EnKF system. We also tested the sensitivity of the EnKF algorithm to the localization and inflation of the background error covariance and the number of ensemble members. In the sensitivity tests, the ensemble spread as well as the root-mean square (RMS) error of the ensemble mean

*Corresponding author. E-mail : yhkim@kordi.re.kr

was assessed. The EnKF produces the optimal solution as the ensemble spread approaches the RMS error of the ensemble mean because the ensembles are well distributed so that they may include the true state. The localization and inflation of the background error covariance increased the ensemble spread while building up well-distributed ensembles. Without the localization of the background error covariance, the ensemble spread tended to decrease continuously over time. In addition, the ensemble spread is proportional to the number of ensemble members. However, it is difficult to increase the ensemble members because of the computational cost.

Key words : data assimilation, Ensemble Kalman Filter, ocean modeling, ensemble spread, localization and inflation of the background error covariance

1. 소 개

해양예보체계란 해양수치모델 결과와 관측자료를 이용하여 현재에 가장 가까운 해양의 상태를 추출하고 수치모델에 입력한 후 해양의 현재와 미래를 예측하는 기술이다. 수온, 염분 및 해류 등 해양의 물리적 특성을 예측·예보함으로써 해군 작전에 응용할 수 있으며, 자연 재해 및 해난 사고에 신속히 대응할 수 있는 기초 자료를 제공함으로써 해양의 안전적 및 효율적 이용에 기여할 수 있다. 해양예보체계를 이용하여 광범위한 해역에 준관측장에 해당하는 고해상도 해양자료를 생산함으로써, 엘리뇨와 같은 수년 주기의 기후변동뿐 아니라 십년 이상의 장주기적인 기후변화에 대한 원인을 파악하고 예측하는 연구에 자료동화기법이 필수적인 수단이 되고 있다(Smith 2006). 그리고 해양예보체계는 해양물리적 특성이 해양생태계에 미치는 영향을 고려하여 해양생태계의 보호 및 복원과 해양오염물질의 확산과정 연구와 방제를 포함한 해양환경 문제를 해결하는 데 응용되고 있다(Bahurel et al. 2006). 이와 같이 다양한 분야에 해양예보체계가 응용될 수 있어 그 경제적인 가치 및 중요성이 크다고 할 수 있다.

예보체계를 현업에 적용하기 시작한지 반세기가 지난 기상분야와는 달리 해양분야의 예보체계는 기초적인 수준에 머무르고 있다. 그 첫번째 요인은 대기분야에 비해 부족한 실시간 해양 관측자료이다. 해양관측자료가 부족하다 보니 지금까지 해양모델은 수십 년 동안 수집되었던 자료로부터 얻은 월평균이나 연평균 자료를 입력자료로 이용하여 지금까지 알려진 해양현상의 재현이나 역학적인 해석에 주로 사용되어 왔다. 최근 해양 변동을 관측하고 예측하기 위해 전지구 해양자료동화실험(GODAE; Global Ocean Data Assimilation Experiment) 등의 일환으로 국제 공동 관측망이 구성되어 획득된 자료를 실시간으로 배포하기 시작하면서 자료동화에 대한 관심과 수요가 커지고 있다. 인공위성을 이용하여 광범위한 영역의 해수면 온도, 해면고도 및 해상풍을 관측하고 있으며, ARGO라는 profiling float를 이용하여 대양의 수온 염분 수직구조를 실시간으로 관측하고 있다. 또한, 컴퓨터 성능 향상으로

사실적인 고해상도 해양순환모델들이 개발되었고, 이로 인해 해양예보체계의 실현 가능성은 더욱 높아졌다.

기본적으로 수치모델을 이용한 예보는 초기값 문제(initial value problem)이다(Bjerknes et al. 1934). 즉, 해양의 현재 상태에 대한 추정치 주어지면 모델은 해양의 시간적 변화를 모의(예보)한다. 예보모델을 위한 초기값을 결정하는 것은 매우 중요하면서도 복잡한 문제이다. 이를 위해 자료동화의 개념이 정립되었고, 그 자체로서 한 과학분야를 이루고 있다(Daley 1991). 넓은 의미에서 자료동화는 공간적으로 불규칙하게 배치된 관측점에서 얻은 관측자료들을 모델 격자점에 할당하기 위한 가중치를 결정하는 객관분석(objective analysis)과 모델의 변수들과 적절한 균형을 이루게 만들어주는 초기화과정(initialization)이 복합된 개념이다. 따라서, 자료동화기법 연구분야는 해양예보체계를 수립하기 위해 필수적으로 선행되어야 할 연구 주제이다. 국내에서는 안 등(2005)이 해양대순환모형의 초기장을 생산하기 위해 공간필터를 사용한 변분법을 적용한 예와 송 등(2000)이 조석모델에 변분동화기법을 적용한 예가 있다. 그러나, 지금까지 국내에서는 해양현상을 관찰하고 순환역학을 이해하기 위한 수치모델 연구가 주로 이루어졌으며 해양자료동화기법에 대한 기초적인 연구가 부족하여 국내 및 국제 공동 관측망 구성과 더불어 자료동화기법과 예보모델에 대한 연구가 시급한 실정이다.

자료동화기법으로 단순한 내삽/외삽(interpolation, extrapolation)과 너징(Nudging)의 방법에서부터 변분동화기법(variational method)이나 칼만필터(Kalman Filter)까지 다양한 방법이 시도되어 왔다. 관측자료를 이용하여 모델의 격자별 자료를 구성하기 위해서는 객관분석법(Objective analysis)이 요구된다. 초기에는 모델 결과를 시공간적으로 불균일한 관측자료에 맞추는 너징 방법이 이용되었다. 하지만, 그 과정이 객관적이지 못하여 그 결과가 통계적으로 최적의 해(Optimal solution)라는 보장이 없다. 이를 개선하기 위한 객관분석법으로 OI(Optimal Interpolation)와 3차원 변분동화기법(3 dimensional variational assimilation)이 제안되었다. OI와 3차원 변분동화기법은 계산비용이

저렴하고 통계적으로 최적의 해를 얻어내기 때문에 아직까지 전세계적으로 현업에 많이 사용되고 있다(De Mey and Benkiran 2002).

모델결과와 관측자료 사이의 가중치를 결정하는 과정이 객관적이며 동시에 통계적으로 최적의 해를 얻어낼 수 있다는 장점에도 불구하고, OI와 3차원변분동화기법은 시간적 연속성을 고려할 수 없다는 점과 고려할 수 있는 해양역학의 범위가 제한적이라는 점이 단점으로 지적되고 있다(Weaver *et al.* 2002). 이러한 단점을 극복하기 위해 일정한 시간 간격 안에 관측된 모든 자료를 동시에 고려하는 4차원 변분동화기법(4 dimensional variational assimilation)이 개발되었다(Marotzke *et al.* 1999). 4차원 변분동화기법을 적용하기 위해서는 수반모델(Adjoint model)이 필요하다. 자동으로 예보모델 코드에서 수반모델 코드를 생성해주는 TAMC(Tangent linear and Adjoint Model Compiler)와 같은 프로그램이 개발되고 있으나, 수 만 줄의 모델코드를 확인하면서 수반모델 코드를 만드는 데는 어려움이 있다. 또한 최적의 해를 구하기 위해서는 반복적인 계산이 필요한데 이때 많은 연산량이 요구된다. 더욱이 지금까지 소개한 3차원 변분동화기법과 4차원 변분동화기법은 모델이 완벽하다는 가정 하에 이루어지므로 모델 자체의 에러를 고려할 수 없다는 단점이 있다(Weaver *et al.* 2002).

칼만필터(Kalman Filter)는 배경오차 공분산(Background Error Covariance)이 전단계의 분석장으로부터 예보모델에 의해 발전해가는 효과적인 순차적(sequential) 자료동화체계이다. 그러나 칼만필터를 이론 그대로 실제 해양자료동화에 적용하는 데는 여러 가지 어려움이 있다. 첫째, 일반적으로 해양에서 사용되는 수치모델은 비선형 항을 포함하고 있어서 칼만필터를 적용하기 위해서는 선형화하는 작업이 필요하다. 이전 시간에 추정된 격자별 자료 주위에서 수치모델을 선형화하는 방법을 확장칼만필터(Extended Kalman Filter)에서 사용하고 있으며 대부분 이 방법이 해양예보모델에 적용되고 있다. 관측자료가 시간적으로 조밀하게 제공되면 칼만필터를 이용하여 격자별 자료를 얻는데는 비선형 효과가 작아서 이 방법을 적용하는데 문제가 적다. 하지만 관측자료의 시간 간격이 크거나, 비선형 효과가 중요해지는 곳에서는 그 효과를 보정해 줄 필요가 있다. 이러한 선형화의 문제는 4차원 변분동화기법에서도 동일하게 발생하는 문제이다. 둘째는 칼만필터를 적용할 때 연산량과 연산과정 중에 저장해야 할 자료량이 매우 커 배경오차 공분산을 계산하거나 물리적으로 저장할 수 있는 한계를 벗어나기 때문에, 이 방법이 그 동안 기상 및 해양학 분야에서는 관심을 얻지 못하였다.

이러한 칼만필터의 한계를 극복하기 위한 몇 가지 대안들이 제시되어 왔다. 배경오차 공분산 행렬의 차원을 줄이기 위해 모델 격자 간격을 크게 하여 칼만필터를 적용하

고 거기서 얻은 큰 격자별 자료로부터 작은 격자별 자료를 OI 등을 통해서 얻는 reduced-rank Kalman Filter가 있다(Fukumori and Malanotte-Rizzoli 1995). 경험적교합수로부터 얻은 고유벡터와 주성분(principal component) 중에서 고유치가 큰 모드만을 선택하여 칼만필터를 수행하는 방법도 적용되고 있다(Cane *et al.* 1996). 배경오차 공분산의 시간변화를 계산하는 연산량을 줄이기 위하여 배경오차 공분산이 시간에 따라 기하급수적으로 정상상태(steady state)에 이른다는 이론을 적용하여 매 시간마다 배경오차 공분산을 계산할 필요가 없는 steady state Kalman Filter 방법이 있다(Fukumori *et al.* 1993). 다른 방법으로는 수치모델에 의해 발전해가는 제한된 수의 앙상블들로부터 직접 배경오차 공분산을 계산하는 앙상블 칼만필터(Ensemble Kalman Filter; Evensen 1994; Burgers *et al.* 1998)나, 이와 유사한 방법으로 SEEK(Singular Evolutive Extended Kalman Filter)가 있다(Pham *et al.* 1998).

그 중에서도 앙상블 칼만필터는 Evensen(1994)에 의해 제안되어 대기분야에 성공적으로 적용된 바가 있으며 해양분야에서도 적용범위를 넓혀가고 있다(Brusdal *et al.* 2003). 앙상블 칼만필터는 앙상블 멤버에 대해 수치모델을 이용한 시간 적분만을 수행하기 때문에 복잡한 해양모델이나 비선형모델에 적용하기가 용이하며, 배경오차 공분산을 직접 계산할 수 있기 때문에 자료동화 체계의 검증이나 관측망 구성 등 다양한 적용이 가능한 것으로 알려져 있다.

본 연구에서는 대표적인 지역해양모델이라 할 수 있는 ROMS(Regional Ocean Modeling System)로 구성된 단순 해양모델에 앙상블 칼만필터를 적용하고 민감도 시험을 실시하였다. Evensen(1994)에 의해 제안되고 Houtekamer and Mitchell(1998)에 의해 수정된 알고리즘을 사용하였으며, 이는 2장에 기술하였다. 3장에서는 Double Gyre 순환이 형성된 이상적인 사각형 바다에 앙상블 칼만필터를 사용하여 해면수온과 해면고도 자료를 동화한다. 그리고 4장에서 앙상블 멤버의 개수와 배경오차 공분산의 지역화·팽창에 대한 앙상블 칼만필터의 민감도 시험 결과를 제시한다. 마지막으로 5장에서 요약과 토의가 이루어진다.

2. 앙상블 칼만필터의 개념

기본 방정식

앙상블 칼만필터는 앙상블들을 이용하여 배경오차 공분산을 추정하는 방법을 사용하고 있다. 개념도(Fig. 1)에서 보는 바와 같이 자료동화 체계에서 초기의 N_e 개의 앙상블 멤버들($t=m-1$ 의 적색 점들)은 수치모델 M 으로부터 N_e 개의 앙상블 멤버에 해당하는 예보장($t=m$ 의 청색 점들)

Fig. 1. Schematic diagram of the Ensemble Kalman Filter. Red dots are reanalysis state vectors (x_k^a), blue dots are predicted state vectors (x_k^f) by the numerical model (M), green dot is an observation vector (y).

을 생성하게 되며 얻어진 예보장으로부터 얻어진 배경오차 공분산을 이용하여 주어진 관측값 y 로부터 N_e 개의 앙상블 멤버에 해당하는 분석장($t=m$ 의 적색 점들)을 얻게 된다. 얻어진 분석장($t=m$ 의 적색 점들)을 다시 초기장으로 하여 모델 M 으로부터 예보장($t=m+1$ 의 청색 점들)을 얻어 다음 자료동화를 수행하게 된다.

이를 식으로 표현하면, Ψ^f 와 Ψ^a 가 예보장(prediction)과 분석장(analysis)의 상태 벡터(state vector)를 나타낼 때 기본 방정식은 다음과 같다(Evensen 1994).

$$\Psi_i^a = \Psi_i^f + K(d_i - H\Psi_i^f), \quad i = 1, \dots, N_e \quad (1)$$

H 는 관측 격자점에서 배경장을 얻는 관측 연산자이며 K 와 d 는 각각 칼만이득행렬(Kalman gain matrix)과 랜덤 함수에 의해 교란된 관측장을 나타낸다.

예보배경오차 공분산(P^f)과 관측오차 공분산(R , observation error covariance)으로부터 칼만이득행렬은 다음과 같이 얻어진다(Evensen 1994).

$$K = [P^f H^T] [H P^f H^T + R]^{-1} \quad (2)$$

일반적으로 예보배경오차 공분산은 이를 표현하는 행렬의 크기 때문에 추정하거나 저장하는 것이 불가능하다. 따라서 앙상블칼만필터에서는 예보배경오차 공분산을 모든 격자점에 대해 직접 추정하기보다는 칼만이득행렬의 계산을 위해 관측점에 투영한 $P^f H^T$ 와 $H P^f H^T$ 를 다음과 같이 앙상블들로부터 추정하게 된다.

$$P^f H^T \equiv \frac{1}{N_e - 1} \sum_{i=1}^{N_e} (\Psi_i^f - \bar{\Psi}^f) (H \Psi_i^f - \overline{H \Psi^f})^T \quad (3)$$

$$H P^f H^T \equiv \frac{1}{N_e - 1} \sum_{i=1}^{N_e} (H \Psi_i^f - \overline{H \Psi^f}) (H \Psi_i^f - \overline{H \Psi^f})^T \quad (4)$$

$$\bar{\Psi}^f = \frac{1}{N_e} \sum_{i=1}^{N_e} (\Psi_i^f), \quad \overline{H \Psi^f} = \frac{1}{N_e} \sum_{i=1}^{N_e} (H \Psi_i^f) \quad (5)$$

일반적으로 연산자 H 는 선형이라고 가정하지만 이 경

우 반드시 선형일 필요는 없다.

최초 예측장 Ψ_i^f 는 완전한(비선형) 수치모델을 적분하여 $\Psi_i^f(t+1) = M_i[\Psi_i^a(t)]$ 와 같이 직접적으로 얻어진다. M_i 은 i 번째 앙상블 멤버에 적용된 수치모델을 나타낸다.

위와 같이 앙상블 칼만필터는 N_e 개의 앙상블 멤버에 대해 예보 모델 M 을 이용한 시간 적분만을 수행하기 때문에 예보모델이나 물리적인 모수화에 대한 단순화된 모델이나 수반모델 등을 필요로 하지 않는다. 또한 N_e 개의 앙상블 멤버에 대해서 모든 상태 변수에 대한 정보를 제공하기 때문에 다변수 배경오차 공분산을 고려하는 것이 용이하다. 일반적으로 상이한 변수들간의 공분산을 나타내는 다변수 배경오차 공분산은 역학적인 균형이나 경험직교함수 등 통계적인 방법으로 추정할 수 있지만 사전 분석 작업에 어려움이 있을 뿐 아니라 객관성을 확보하기가 쉽지 않다. 일 예로 기존의 자료동화 체계는 인공위성으로부터 얻어진 해면고도자료(sea surface height data)를 수치모델에 동화하는데 어려움을 겪어 왔다. 이는 해면고도가 인공위성에서 관측할 수 있는 해면의 특성이지만 결과적으로 해양의 심층에서 일어나는 변동까지 포함한 특성이기 때문에 단순히 관측된 해면고도를 수치모델의 해면고도에 동화할 수 없는 어려움이 있다. 그러나 앙상블 칼만필터에서는 앙상블들로부터 다변수 배경오차공분산을 추정하는 것이 용이하며 따라서 해면고도자료를 수온이나 유속에 동화하기가 용이하다. 3차원 변분동화기법(3 dimensional variational assimilation) 등에서는 잠재와도(Potential Vorticity)의 보존과 해저면 압력이 보존된다는 가정 하에 수온과 염분의 프로파일을 수직 이동시키는 방법이 제안된 바 있다(Cooper and Haines 1996). 반면에 앙상블 칼만필터에서는 다변수 배경오차 공분산을 계산하여 사용하기 때문에 역학적인 균형을 유지하면서 해면고도를 모델에 동화할 수 있다.

배경오차 공분산의 지역화(Localization of Background Error Covariance)

그러나 앞 절에서 기술한 장점에도 불구하고 앙상블 칼만필터는 유한의 앙상블 멤버를 이용하기 때문에 몇 가지 문제점이 발생하게 된다. 다시 말해, 유한의 앙상블 멤버로 구성된 배경오차 공분산은 필연적으로 자유도가 저평가되고 서로 독립적일 가능성이 큰 원거리에 놓여 있는 변수간의 상관성까지 높게 추정하는 비정상적인 형태를 가질 수 있어 이에 대한 해결책이 연구되어 왔다. Gaspari and Cohn(1999)는 이러한 자유도의 저평가로 인한 문제를 해결하고자 배경오차 공분산의 지역화를 적용하였으며 이를 위하여 Schur Product를 고안하였다. Schur Product는 원거리의 배경오차 공분산을 인위적으로 낮춤으로 자유도를 높이는 방안이다. Schur Product를

Fig. 2. Gaussian weight function for Schur product to localize the background error covariance.

식 (6)과 같이 정의했을 때 칼만이득행렬은 식 (7)과 같이 주어진다.

$$\rho^o \begin{bmatrix} b_{1,1} & \dots & b_{i,1} & \dots & b_{M,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{1,j} & \dots & b_{i,j} & \dots & b_{M,j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{1,M} & \dots & b_{i,M} & \dots & b_{M,M} \end{bmatrix} = \rho_{i,j}[L(x_i, x_j)] \times b_{i,j} \quad (6)$$

$$K = [\rho^o(P^f H^T)] [\rho^o(H P^f H^T) + R]^{-1} \quad (7)$$

위에서 $\rho^o B$ 는 3차원 공간에서 x_i 와 x_j 의 거리(L)로 주어지는 함수인 ρ 에 의해서 얻어진다(식 (6)). 함수 ρ 는 비상관거리에 대한 함수로 주어지며 3차원 해양순환모델에서는 수평 비상관 거리와 수직 비상관 거리에 대해서 함수 ρ 가 주어질 수 있다. 수평 비상관거리가 600 km일 경우 Gaspri and Cohn(1999)에 의하여 고안된 함수 ρ 는 Fig. 2와 같다. 거리에 대한 가우시안(Gaussian) 분포의 형태를 나타내며 비상관거리에 가까워질수록 0에 접근하고 비상관거리 이상의 거리에 대해서는 0의 값을 갖는다.

배경오차 공분산의 팽창(Inflation of background error covariance)

앙상블 칼만필터를 적용함에 있어 또 하나의 문제점으로는 순차적 자료동화를 진행하는 동안 일어나는 배경오차 공분산의 붕괴가 있다. 이론적으로 예보장의 배경오차 공분산(P^f)과 분석장의 배경오차 공분산(P^a)은 $P^a = (I - KH)P^f$ 의 관계를 유지해야 한다. 여기서 I 는 단위행렬을 나타낸다. 즉, 자료동화를 통해 배경오차 공분산은 $(I - KH)$ 정도 감소하게 되고 이는 재분석장의 오차를 줄이는

효과를 나타낸다. 하지만 표본 추출의 문제로 인하여 분석장의 배경오차 공분산이 저평가되는 경향이 있다(Wang and Bishop 2003). 결과적으로 자료동화 과정 중에 앙상블들이 과도하게 수렴하게 됨으로 인하여 어느 순간 재분석장이 참값에서 급격히 멀어지는 문제점을 가진다. 이러한 문제점을 해결하는 방안으로 Wang and Bishop(2003)은 배경오차 공분산의 팽창을 제안하였다.

Hollingsworth and Lönnberg(1986)는 배경오차와 관측오차가 상호 관계성을 갖지 않는다는 가정하에 관측값의 공분산을 배경오차공분산과 관측오차공분산의 합으로 표현하였다. 이에 따르면 일반적으로 쇄신 벡터(Innovation vector) $d' = R^{-1/2}(y - Hx^f)$ 에 대해

$$Tr(\langle d' d'^T \rangle) = Tr(H' P^f H'^T + I) \quad (8)$$

을 만족하게 된다. $H' = R^{-1/2}H$ 이다. 그러나 배경오차 공분산의 저평가로 인하여 식 (8)의 관계를 만족하지 않게 된다. 따라서 식 (8)을 만족시킬 수 있도록 배경오차 공분산을 팽창시키는 방법을 생각할 수 있다. 즉, 팽창계수 α 에 대해 다음을 만족하게 되며

$$Tr(\langle d' d'^T \rangle) = Tr(H' \alpha P^f H'^T + I) \quad (9)$$

식 (9)로부터 팽창계수는 다음과 같다.

$$\alpha = \frac{d'^T d' - N_o}{Tr(H' P^f H')} \quad (10)$$

N_o 는 관측자료의 개수이다.

배경오차 공분산의 붕괴를 막는 방안으로 관측장의 교란이 사용되기도 한다(Burgers *et al.* 1998). 이 방안은 주어진 관측자료에 가우시안 분포의 랜덤오차를 더해 각 앙상블에 동화하여 결과적으로 각 앙상블들이 수렴하는 것을 막는다.

3. 앙상블 칼만필터의 적용

수치모델 개요

본 연구에서는 ROMS(Regional Ocean Modeling System)를 이용한 직사각형 영역의 단순모형에 앙상블 칼만필터를 적용하였다. 앙상블 칼만필터가 단순모형에 적용되었으나, 각 모듈들은 일반적인 해역에 적용될 수 있도록 범용성을 갖추도록 하였다. ROMS는 정수압근사(Hydrostatic approximation)와 Boussinesq 근사를 기본으로 사용하고 있다. 수평적으로는 Arakawa C 격자체계를 기초로 직교 곡선좌표계를 적용하였으며, 수직적으로 자유수면 정역학 원시방정식을 기본으로 S-좌표계(stretched terrain-following coordinate)를 적용하였다. 계산 시간의 효율성을 높이기 위해 경압모드와 순압모드를 사용한 시간 분할법(split-

explicit time stepping scheme)을 사용하고 있다. 운동량의 이류항 계산을 위해 수평적으로 2차 중앙차분(centered)을 그리고 수온의 이류항 계산을 위해 3차 상류차분(upstream)을 사용하고 있다.

해양모델을 위하여 $1,000 \text{ km} \times 1,000 \text{ km}$ 닫힌 공간에 500 m의 수심을 가정하였으며 외력으로는 해양에서 Double Gyre를 유도할 수 있도록 식 (11)과 같이 동서방향의 바람응력이 주어졌다.

$$\tau_x(y) = \tau_0 \cdot \cos(2\pi y/L_y) \quad (11)$$

비록 단순 모형 실험이지만 해양의 비선형 역학을 포함하고 있어, 사각형 바다의 북쪽에는 반시계방향의 순환, 남쪽에는 시계방향의 순환이 나타나고, 두 개의 순환(double gyre) 사이에는 서쪽 경계에서 동쪽으로 흐르는 해류가 사행한다. 본 연구에서 양상을 칼만필터의 성능 및 민감도 시험을 하는데 이 double gyre가 존재하는 모형 바다를 사용하였다. 초기부터 32년을 적분하여 주어진 관측점(Fig. 8a의 적색점)에서 31년부터 32년의 해면고도와 해면수온을 추출하여 26년부터 27년까지 자료동화를 수행하였다.

수치모델의 선형성 시험

선형 이론에 기초한 양상블 칼만필터와 달리 해양에 적용되는 유체역학 방정식은 기본적으로 비선형 방정식에 기초하고 있다. 그러나 해양의 변동에 따라 짧은 시간 내에서는 선형성을 가정할 수 있다. 따라서 양상블 칼만필터를 도입하고자 할 때, 이용하는 수치모델의 선형성을 필수적으로 시험해야 할 것이다. 수치모델의 선형성에 대한 통계적인 접근은 Hansen and Smith(2000)에 의해 제안된 바 있으며 본 연구에서도 이를 이용하여 선형성을 시험하였다.

선형성을 나타내는 Θ 는 다음과 같이 주어지며,

$$\Theta(t_0, \delta(t_0), t) = \frac{\|\delta^+(t) + \delta^-(t)\|}{\frac{1}{2}(\|\delta^+(t)\| + \|\delta^-(t)\|)} \quad (12)$$

초기 Θ 는 0으로 주어진다. 모델 수행 초기($t = t_0$)에 주어진 $\delta^+(t_0)$ 와 $\delta^-(t_0)$ 는 크기는 같고 방향만 반대인 벡터들이며 모델이 수행됨에 따라 방향과 크기가 변하게 된다. 선형방정식에 기반한 수치모델의 경우, 시간이 지나도 $\delta^+(t)$ 와 $\delta^-(t)$ 는 크기가 같고 방향이 180° 인 채로 유지되는 반면, 비선형 방정식에 기반한 수치모델의 경우, 시간이 지남에 따라 $\delta^+(t)$ 와 $\delta^-(t)$ 크기도 다르고 방향도 180° 에서 벗어나게 된다. 완전한 비선형성을 이룰 때 Θ 는 2가 된다(Fig. 3 참조). 본 연구를 위해 ROMS 모델의 단 순실험인 Double Gyre 실험에 대하여 선형성 시험을 실

Fig. 3. Schematic diagram of the model linearity test (after Hansen and Smith 2000).

Fig. 4. Result of the linearity test for the Double Gyre experiment.

시했다. 초기에 3개의 모델(x_0^+, x_0^-, x_0^r)은 $\delta^+(t_0) = x_0^+ - x_0^r$ 그리고 $\delta^-(t_0) = -\delta^+(t_0) = x_0^- - x_0^r$ 의 관계를 갖도록 주어져 식 (12)에 의해 Θ 는 초기에 0의 값을 갖는다. Θ 는 초기에 0에서부터 시작하여 60일 이전까지 0.6 이하로 유지되다가 점차 증가하여 약 360일이 지나면서 1.7에 근접한다(Fig. 4). 따라서 본 결과로부터 30일 동안에는 수치모델이 선형성을 잃지 않는 것으로 판단하였으며 이로부터 자료동화 주기를 30일로 결정하였다.

배경오차 공분산의 지역화

Fig. 5는 배경오차 공분산에 대한 지역화를 적용한 예이다. Fig. 5a에서 보는 바와 같이 양상블들로부터 계산된 표층수온과 표층수온의 배경오차 공분산(자기 상관 공분산)은 서쪽 경계의 중앙부와 남쪽 경계의 중앙부에서 음의 큰 공분산을 보이고 300 km 이상 떨어진 지점까지 공분산이 크게 나타나고 있다. 그러나 이렇게 원거리에 위치한 변수들간의 큰 공분산은 실제 역학적으로 상관도가 높아서라기보다는 양상블 개수가 유한함에 따른 자유도(Degree of freedom)의 저평가에 의한 것으로 알려져 있다(Gaspari and Cohn 1999). 따라서 원거리의 큰 배경오차 공분산을 지역화함으로 Fig. 5b와 같이 원거리에 위치한 변수들간의 공분산을 낮추게 된다.

표층수온과 해면고도의 배경오차 공분산(다변수 상관 공분산)은 자기 상관 공분산과는 달리 정점 M 의 북쪽에서 큰 음의 공분산을 갖는데(Fig. 6a) 자기 상관 공분산에서와 마찬가지로 원거리에 위치한 변수들간의 큰 공분산이 나타난다. 이를 배경오차 공분산 지역화를 통해 Fig. 6b와 같이 원거리에 위치한 변수들간의 공분산을 낮추게 된다.

Fig. 5. Horizontal distribution of the background error covariance between surface temperature at point M (blue cross) and surface temperature at each grid point: auto-covariance of surface temperature (a) without localization and (b) with localization by the Schur product.

Fig. 6. Same with Fig. 5, but multi-variate covariance between surface temperature at point M and surface elevation at each grid point (a) without localization and (b) with localization.

Table 1. Configuration of sensitivity test of the Ensemble Kalman Filter to the localization and inflation of the background error covariance and the total number of ensemble members

	Num. of Ens. Mem.	Horizontal local.	Vertical local.	Cov. inflation	SST assim.	SSH assim.	Stability
REF	16	O	O	O	O	O	Stable
E32	32	O	O	O	O	O	Stable
HNLC	16	X	O	O	O	O	Overflow
VNLC	16	O	X	O	O	O	Unstable
CNINF	16	O	O	X	O	O	Stable
ASSH	16	O	O	O	X	O	Stable

4. 앙상블 칼만필터에 대한 민감도 시험

민감도 시험 조건

앙상블 칼만필터를 적용하고자 할 때 앙상블의 개수가 유한함에 따른 자유도의 저평가가 쟁점이 되어왔다. 이러

한 자유도의 저평가 문제를 해결하기 위하여 제시된 것이 앞서 기술한 배경오차 공분산의 지역화와 팽창이다. 따라서 본 연구에서는 배경오차 공분산의 지역화와 팽창에 대해서 각각 앙상블 칼만필터의 민감도 시험을 실시하였으며, 자유도에 따른 앙상블 칼만필터의 민감도를 시험하기 위해 앙상블의 개수가 16개인 경우와 32개인 경우에 대해 비교 시험을 실시하였다. 다변수 배경오차 공분산의 성능을 시험하기 위하여 해면수온과 해면고도를 모두 수치모델에 동화하는 경우와 해면고도만을 동화하는 경우에 대하여 각각 시험을 실시하였다(Table 1). 위의 각 경우에 대한 비교를 위해 표준실험(REF)에서는 16개의 앙상블을

Fig. 7. Time table for the twin experiment.

Fig. 8. Sea surface elevation for the true field (a), experiment REF (b), E32 (c) and ASSH (d).

이용하였으며, 배경오차 공분산의 팽창과 지역화를 적용하여 해면수온과 해면고도를 수치모델에 동화하였다.

민감도 시험은 쌍둥이 실험(Twin experiment) 형식을 따랐다. 운동이 없는 해양을 초기조건으로 하여 32년간 적분하였으며, 마지막 2년의 모의 결과를 참값(True Ocean)으로 그리고 26년째부터 27년까지 2년 동안의 모의 결과를 표준 해양(Control Ocean)으로 가정하였다. 참값으로부터 40개 정점에서 해면수온과 해면고도를 추출하여 표준 해양에 동화하였다(Fig. 7 참조). 쌍둥이 실험의 경우 참값을 이미 알고 있기 때문에 자료동화를 통해 표준 해양이 얼마나 빨리 그리고 안정적으로 참값에 접근하는지를 판단할 수 있다.

앙상블 칼만필터의 민감도 결과

각 실험들로부터 얻어진 해면고도의 분석장(Fig. 8b, c, d)과 주어진 참값(Fig. 8a)을 비교하였다. Fig. 8a의 적색으로 표시한 정점에서 해면수온과 해면고도를 관측값으로 얻었으며 ASSH 실험의 경우 해면고도만을 동화하였고 그 외의 모든 실험에서 해면고도와 해면수온을 수치모델에 동화하였다. 모든 경우에 다변수 배경오차 공분산을 고려하여 3차원 수온장과 유속장에 동화하였으며 자료동화 주기는 30일로 하였다.

모든 실험의 경우에서 초기에는 해면고도가 참값과의 거리가 멀지만, 자료동화가 진행됨에 따라 각 실험의 해면고도가 참값에 근접하는 것을 볼 수 있다(Fig. 8a와 비교). 다만, E32가 결과적으로 참값에 가장 근접한 구조를 보인 반면 ASSH의 경우 참값에 근접하는 시간도 오래 걸리고 상대적으로 다른 경우에 비해 참값과 차이가 큰 것으로 나타났다. 이를 통계적으로 확인하기 위하여 해면고도와 해면수온에 대해 RMS(root-mean square) 오차와 앙상블 산포도를 다음과 같이 계산하였다.

$$RMS \equiv \sqrt{\frac{1}{N} \sum_{i=1}^{N_s} (\bar{\Psi}_i^a - \Psi_i^t)^2} \quad (12)$$

$$EnSP \equiv \sqrt{\frac{1}{N_s(N_s-1)} \sum_{j=1}^{N_e} |\bar{\Psi}^f - \Psi_j^f|^2} \quad (13)$$

N_s 는 수치모델 변수들의 격자점 개수를, Ψ^t 는 참값을 그리고 $\bar{\Psi}^a$ 는 분석장 Ψ^a 의 앙상블 평균을 의미한다. RMS 오차는 자료동화의 통계적인 오차를 나타내며 앙상블 산포도(root-mean spread of ensemble forecast)는 각 앙상블 예보장들이 앙상블 평균으로부터 어느 정도 범위 내에 존재하는지를 보여주는 산포도를 나타낸다. RMS 오차가 작을수록 분석장과 참값의 차이 또한 작기 때문에 자료동화 체계의 성능이 좋다고 할 수 있을 것이다. 일반적으로 앙상블 산포도는 현재의 앙상블들이 얼마나 대표성을 갖는지에 대한 척도로 간주되곤 한다. 즉 앙상블 산포도가 RMS오차에 근접할수록 각 앙상블들이 참값을 포함하는 범위까지 고루 분포할 가능성이 크기 때문에 안정적으로 최적의 분석장을 얻을 가능성이 높다고 할 수 있다.

Fig. 9는 해면고도의 분석장과 참값의 차이를 정량적으로 나타낸 것이다. 분석장의 RMS 오차가 초기 4개월 동안 감소하는 경향이 모든 실험에서 나타났다. 각 경우의 RMS 오차는 평균 6.7~7.2 cm로 큰 차이를 보이지 않지만 앙상블 산포도는 평균적으로 4.1~6.1 cm로 각 실험에 따라 큰 폭의 차이를 나타낸다. 배경오차 공분산의 팽창을 고려하지 않을 경우(CNINF)에 앙상블 산포도가 평균 4.1 cm로 가장 작은 반면, 앙상블의 개수가 32개인 경우(E32)에 앙상블 산포도가 평균 6.1 cm로 가장 안정적으로 앙상블들을 구성하는 것으로 나타났다. 또한 배경오차 공분산의 지역화를 적용하지 않을 경우(VNLC)에 앙상블 산

Fig. 9. Root-mean square of ensemble mean error (solid line) and ensemble spread (dashed line) of sea surface elevation for the experiment REF (a), E32 (b), VNLC (c), CNINF (d) and ASSH (e).

Fig. 10. Root-mean square of ensemble mean error and ensemble spread of sea surface temperature for the experiment REF (a), E32 (b), VNLC (c), CNINF (d) and ASSH (e).

포도가 평균 4.6 cm이며 시간에 따라 점차 감소하는 경향이 있다.

앙상블 칼만필터의 민감도는 해면 수온에서도 뚜렷하게 나타난다(Fig. 10). 각 경우의 RMS오차는 평균 0.27~0.30°C이며, E32의 경우 0.27°C로 가장 작고, ASSH의 경우 0.3°C로 가장 크게 나타났다. 앙상블 산포도는 앞의 해면고도에서와 마찬가지로 CNINF의 경우에 가장 작은 평균 0.14°C이며, E32와 ASSH의 경우 가장 큰 평균 0.19°C였다. 또한 다른 경우 앙상블 산포도가 일정하게 유지되는 반면 VNLC의 경우 앙상블 산포도가 감소하는 경향이 뚜렷하다.

ASSH 실험은 비록 RMS 오차가 가장 커서 모든 실험들 중에서 가장 좋지 않은 결과를 보이고 참값에 접근하는 시간도 오래 걸리지만, 해면 수온을 자료동화에 이용하지 않고 다변수 배경오차 공분산을 이용하여 해면고도만을 3차원 수온장에 동화한 결과임을 생각할 때 앙상블 칼만필터에 사용되는 다변수 배경오차 공분산의 역할에 주목할 만하다.

5. 요약 및 토의

본 연구에서는 앙상블 칼만필터의 이론적 기법들을 해양순환모델(ROMS)의 단순모형에 적용하였으며 자유도의 저평가로 인한 배경오차 공분산의 붕괴를 극복하기 위하여 배경오차 공분산의 지역화와 팽창 그리고 관측장의 교란 등의 기법들을 사용하였다. 앙상블 개수, 배경오차 공분산의 지역화와 팽창 그리고 다변수 배경오차 공분산에 대한 앙상블 칼만필터의 민감도 및 성능을 시험하였다.

앙상블 칼만필터를 적용한 자료동화 체계에 있어 자유도의 저평가로 인해 배경오차 공분산의 붕괴를 막는 이상적인 방법은 앙상블의 개수를 실제 해양순환모델의 자유

도 만큼 증가시키는 것일 수 있다. 실제로 본 연구에서 32개의 앙상블을 사용한 실험에서 16개의 앙상블을 사용하는 다른 모든 실험들보다 앙상블 산포도가 가장 RMS 오차에 근접한 것으로 나타났는데, 이유는 많은 수의 앙상블을 사용함으로써 확률분포를 보다 잘 표현하였기 때문이다.

그러나 계산 비용 및 저장 용량의 한계 때문에 앙상블의 개수는 제한될 수 밖에 없다. 본 연구 결과는 배경오차 공분산의 지역화와 팽창이 대안이 될 수 있음을 보여주었다. 배경오차 공분산의 지역화는 앙상블들의 수렴을 막는 것으로 나타났다. 배경오차 공분산을 지역화시키면 앙상블 산포도가 일정한 값을 유지하는 반면, 그렇지 않을 경우 앙상블 산포도가 시간이 지남에 따라 지속적으로 감소하는 경향이 뚜렷하게 나타났다. 배경오차 공분산의 팽창을 통해서도 앙상블 산포도를 RMS 오차에 가까워지도록 증가시킬 수 있었다.

해면고도를 해양순환모델에 동화하고자 할 때 기존 연구들은 주로 경험적 직교함수 등의 통계적인 방법을 이용하거나 역학적 제한 조건을 도입하여 해면고도를 3차원 수온장에 동화해왔다. 그러나 본 연구에서는 앙상블 칼만필터를 적용할 경우 다변수 배경오차 공분산의 구성이 용이하여 역학적 균형을 고려하면서도 보다 효과적으로 해면고도를 수치모델에 동화할 수 있음을 보여주었다. 해면고도만을 동화한 경우(ASSH)가 해면고도와 해면수온을 모두 동화한 경우보다 당연히 RMS 오차가 컸지만 다변수 배경오차 공분산을 통해 해면고도만을 동화했음에도 해면고도와 해면수온이 참값에 접근하는 것을 확인할 수 있었다.

본 연구에서는 이상적인 실험을 통하여 앙상블 칼만필터 기법을 이용한 자료동화 체계의 성능을 시험하였지만, 이 기법은 실제 해역에 적용하고자 할 때에는 해결해야 할

과제들이 있다. 첫째, 앙상블 칼만필터에서는 배경오차 공분산을 관측점으로 투영시켜 그 크기를 줄이지만 관측 자료의 개수가 많을 경우 배경오차 공분산의 크기는 여전히 계산 및 저장 용량의 한계를 초과할 수 있다. 이에 대한 방안으로 Houtekamer and Mitchell(2001)은 관측 자료를 다수의 집단으로 세분하여 순차적으로 동화할 것을 제안하였으며 앙상블의 개수가 $O(10)$ 인 경우를 제외하고 그 성능이 전체를 한번에 동화했을 때와 비교하여 크게 차이가 없음을 보였다. 이 경우 앙상블 칼만필터 알고리즘을 병렬화하여 계산속도를 향상시킬 수 있는 장점이 있다.

둘째, 배경오차 공분산의 지역화를 위한 비상관거리이다. Hollingworth and Lonnberg(1986)은 자기상관 공분산은 관측오차 공분산과 배경오차 공분산의 합으로 표현되며 관측오차는 공간적으로 상관성을 갖지 않는다는 가정하에 관측값으로부터 배경오차 공분산의 비상관거리를 추정하는 방법을 제안한 바 있다.

본 연구는 앙상블 칼만필터를 실제 해양순환모델에 적용하고자 하는 사전 연구로서 수행되었다. 비록 기존의 자료동화기법들과의 비교는 이루어지지 않았지만 앙상블 칼만필터는 OI나 3차원 변분동화기법 등에 비해 비교적 자연스럽게 역학적 균형을 고려할 수 있으며 특히 다변수 배경오차 공분산을 쉽게 적용할 수 있어 성능이 뛰어날 것으로 예상된다. 앞서 기술한대로 몇 가지 해결해야 할 사항이 남아있기는 하지만, 본 연구에서 각 기법들의 민감도를 시험함으로 앙상블 칼만필터를 실제 해역에 적용했을 때 나타날 수 있는 문제점들과 해결 방안들에 대한 기초 자료를 제시함으로 앙상블 칼만필터를 해역의 순환모델에 적용할 수 있는 기반을 마련했다고 할 수 있다.

사 사

본 연구는 국방과학연구소의 지원으로 수행되었습니다(계약번호 UD031003AD). 제 1저자는 한국해양연구원의 연구프로그램(PE9811A)으로부터 일부 지원을 받았습니다. 본 논문의 심사를 맡아 세세한 부분까지 조언해 주신 심사위원님들께 감사드립니다.

참고문헌

- 송용식, 이재학, 박경. 2000. 조석 모델링에서 adjoint 방법 적용 시 적정 가중치 선정. *한국해양학회지 「바다」*, **5**(3), 177-185.
- 안중배, 윤용훈, 조익현, 오혜람. 2005. VAF 변분법을 이용한 전구 해양자료 동화 연구. *한국해양학회지 「바다」*, **10**(1), 69-78.
- Bahurel, P., The MERCATOR Project Team. 2006.

- MERCATOR Ocean global to regional ocean monitoring and forecasting. p. 381-395. In: *Ocean weather forecasting: An integrated view of oceanography*, ed. by E.P. Chassignet and J. Verron. Springer, Dordrecht.
- Bjerknes, V., J. Bjerknes, H.S. Solberg, and T. Bergeron. 1934. *Hydrodynamique physique avec applications a la météorologie dynamique*. Les Presses Universitaires, Paris. 864 p.
- Brusdal, K., J.M. Brankart, G. Halberstadt, G. Evensen, P. Brasseur, P.J. van Leeuwen, E. Dombrowsky, and J. Verron. 2003. A demonstration of ensemble based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. *J. Mar. Sys.*, **40-41**, 253-289.
- Burgers, G., P.J. van Leeuwen, and G. Evensen. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.*, **126**(6), 1719-1724.
- Cane, M.A., A. Kaplan, R.N. Miller, B. Tang, E.C. Hacker, and A.J. Busalacchi. 1996. Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter. *J. Geophys. Res.*, **101**(C10), 22599-22617.
- Cooper, M. and K. Haines. 1996. Altimetric assimilation with water property conservation. *J. Geophys. Res.*, **101**(C1), 1959-1977.
- Daley, R. 1991. *Atmospheric data analysis*. Cambridge University Press, New York. 457 p.
- De Mey, P. and M. Benkiran. 2002. A multivariate reduced order optimal interpolation method and its application to the Mediterranean basin-scale circulation. p. 281-306. In: *Ocean Forecasting: conceptual basis and applications*, ed. by N. Pinardi and J. Woods. Springer-Verlag.
- 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), 10143-10162.
- Fukumori, I. and P. Malanotte-Rizzoli. 1995. An approximate Kalman Filter for ocean data assimilation: An example with an idealized Gulf Stream model. *J. Geophys. Res.*, **100**(C4), 6777-6793.
- Fukumori, I., J. Benveniste, C. Wunsch, and D.B. Haidvogel. 1993. Assimilation of sea surface topography into an ocean circulation model using a steady-state smoother. *J. Phys. Oceanogr.*, **23**(8), 1831-1855.
- Gaspari, G. and S.E. Cohn. 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.*, **125**(554), 723-757.
- Hansen, J.A. and L.A. Smith. 2000. The role of operational constraints in selecting supplementary observations. *J. Atmos. Sci.*, **57**(17), 2859-2871.

- Hollingsworth, A. and P. Lönnberg. 1986. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111-136.
- Houtekamer, P.L. and H.L. Mitchell. 1998. Data assimilation using an ensemble Kalman Filter technique. *Mon. Weather Rev.*, **126**, 796-811.
- Houtekamer, P.L. and H.L. Mitchell. 2001. A sequential ensemble Kalman Filter for atmospheric data assimilation. *Mon. Weather Rev.*, **129**, 123-137.
- Marotzke, J., R. Giering, K.Q. Zhang, D. Stammer, C. Hill, and T. Lee. 1999. Construction of the adjoint MIT ocean general circulation model and application to Atlantic heat transport sensitivity. *J. Geophys. Res.*, **104**(C12), 29529-29547.
- Pham, D., J. Verron, and M. Roubaud. 1998. A singular evolutive extended Kalman Filter for data assimilation in oceanography. *J. Mar. Sys.*, **16**, 323-340.
- Smith, N. 2006. Perspectives from the global ocean data assimilation experiment. p. 1-17. In: *Ocean weather forecasting*, ed. by E.P. Chassignet and J. Verron. Springer.
- Wang, X. and C.H. Bishop. 2003. A comparison of breeding and ensemble transform Kalman Filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140-1158.
- Weaver, A.T., J. Vialard, D.L.T. Anderson, and P. Delecluse. 2002. Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. ECMWF Technical Memorandum No. 365. 74 p.

Received Jan. 10, 2008

Accepted Mar. 20, 2008