Atmosphere. Korean Meteorological Society
Vol. 23, No. 1 (2013) pp. 33-37

연 구 논 문

칼만필터의 자료동화 활용을 위한 배경오차 공분산의 명시적 시간 진전 제거

임규호[1)] · 서애숙[2)] · 하지현[1),*]
[1)]서울대학교 자연과학대학 지구환경과학부
[2)]대전 기상청

(접수: 2012년11월23일, 수정: 2013년 1월19일, 게재확정일:2013년 1월26일)

An Affordable Implementation of Kalman Filter by Eliminating the Explicit Temporal Evolution of the Background Error Covariance Matrix

Gyu-Ho Lim[1)], Ae-Sook Suh[2)], and Ji-Hyun Ha[1),*]
[1)]Atmospheric Sciences Program, School of Earth and Environmental Sciences, Seoul National University, Seoul 151-747, South Korea
[2)]Daejeon Regional Meteorological Administration, Daejeon 305-338, Korea

(Received: 23 November 2012, Revised: 19 January 2013, Accepted: 26 January 2013)

Abstract In meteorology, exploitation of Kalman filter as a data assimilation system is virtually impossible due to simultaneous requirements of adjoint model and large computer resource. The other substitute of utilizing ensemble Kalman filter is only affordable by compensating an enormous usage of computing resource. Furthermore, the latter employs ensemble integration sets for evolving the background error covariance matrix by compensating the dynamical feature of the temporal evolution of weather conditions. We propose a new implementation method that works without the adjoint model by utilizing the explicit expression of the background error covariance matrix in backward evolution. It will also break a barrier in the evolution of the covariance matrix. The method may be applied with a slight modification to the real time assimilation or the retrospective analysis.

Key words: Kalman filter, adjoint model, background error covariance matrix, assimilation

1. 서 언

예보 실무 분야에서 사용하는 자료 동화법은 관측 기기의 특성과 요구되는 컴퓨터의 전반적인 성능을 고려하여 결정된다. 예보 개선의 정도만을 고려한 자료 동화의 성능은 4차원 변분 동화법 (4-DVAR)과 칼만필터 (Kalman filter, KF) 혹은 칼만 스무드 (Kalman smoother, KS)가 높게 평가되고 있다. 동화에 필요한 계산 자원이나 실무 수행상의 어려움을 고려하지 않는다면, 4차원 변분 동화법 혹은 칼만 스무드의 사용은 좋은 성능을 보인다. Song and Lim (2009)은 상당한 계산 자원을 필요로 하는 수반 모형을 퍼터베이션 방법 (perturbation method)으로 대체하는 회귀적 최적 내삽법 (Retrospective optimal interpolation, ROI)을 제안하였다.

기상 실무 차원에서 이상적인 4차원 변분 동화법과 비슷한 정도의 정확도를 가지는 KF의 활용은 현재로서는 요원한 실정이다. 물론 앙상블 칼만 필터 (ENKF)와 같은 변환된 기법이 제안되어 사용되지만 기본적인 KF의 특성 활용에는 한계가 있다. KF의 실무 적용 가능성에 대한 연구는 KS의 활용을 위한 선수 조건으로 판단하여 이 논문을 작성하게 되었다.

*Corresponding Author: Ji-Hyun Ha, Atmospheric Sciences Program, School of Earth and Environmental Sciences, Seoul National University, Seoul 151-747, Korea.
Phone : +82-2-880-8169, Fax : +82-2-887-4890
E-mail : dearhaji@gmail.com

이 논문의 목적은 전통적인 KF의 실무적용 가능성 모색이다. KF 활용에 있어서 난제인 배경오차공분산 행렬 (background-error covariance matrix, BCM)의 시간 진전 (temporal evolution)에 필수적인 수반 모델을 대체할 수 있는 방법을 생각하였다. 동시에 BCM 시간 진전과 관련된 해당 행렬의 초고차원에 기인하는 문제를 제거하여 KF의 실무 차원 응용 가능성을 조사하고자 한다. 우리의 방법론은 모델의 변수와 격자점 수와는 무관하다. 언급한 특성은 모델 차원, 일반적으로 10^7~10^8, 과는 무관하기에 기상분야의 자료 동화에 활용할 수 있을 것이다.

2. 칼만필터

칼만필터는 통계 기반의 추정법 (推定法)으로 화학 플랜트의 공정 조정, 홍수 예경보, 항공우주 공학 분야의 위성체 추적, 선박과 차량의 이동 경로 추적에 사용된다 (Grewal and Andrews, 2001). 심지어 KF의 개발이 없었다면 인간의 달 착륙 실현이 불가하였으리라는 의견도 있다 (McGee and Schmidt, 1985).

우리는 KF 관련 부분의 설명을 위해서 Bouttier and Courtier (1999)의 기호와 일반적인 설명을 따른다. 이후부터 BC로 명명한다. Fig. 1에서 실시간 자료 동화 분야에서는 비선형 방법을 제외하고는 4-DVAR를 그리고 회귀적 분석 (retrospective analysis)을 가능하게 하는 방법으로는 KS를 지적하였다. 앞서 언급한 바와 같이 본 논문에서는 향후 KS 활용을 위한 선수 조건으로 KF의 실무 적용에 대해 살펴보고자 한다. BC가 사용한 KF 수식 집합은 다음과 같다.

State forecast

$$\mathbf{x}_f(i+1) = M_{i\to i+1}\mathbf{x}_a(i) \qquad \text{KF1}$$

Error covariance forecast

$$\mathbf{P}_f(i+1) = \mathbf{M}_{i\to i+1}\mathbf{P}_a(i)\mathbf{M}^T_{i\to i+1} + \mathbf{Q}(i) \qquad \text{KF2}$$

Kalman gain computation

$$\mathrm{K}(i) = \mathbf{P}_f(i)\mathbf{H}^T(i)[\mathbf{H}(i)\mathbf{P}_f(i)\mathbf{H}^T(i) + R(i)]^{-1} \qquad \text{KF3}$$

State analysis

$$\mathbf{x}_a(i) = \mathbf{x}_f(i) + \mathbf{K}(i)[y(i) - H(i)\mathbf{x}_f(i)] \qquad \text{KF4}$$

Error covariance analysis

$$\mathbf{P}_a(i) = (\mathbf{I} - \mathbf{K}(i)\mathbf{H}(i))\mathbf{P}_f(i) \qquad \text{KF5}$$

위에서 기울임꼴로 표시된 연산자 M, H는 비선형 연산자이며 굵은 대문자로 표시된 연산자 $\mathbf{M}$, $\mathbf{H}$는 비선형 연산자에 상응하는 선형연산자이다. 미분 가능한 비선형 연산자의 선형 연산자 도출은 항상 가능하다. 또한 선형 연산자에 해당하는 수반연산자는 유일하다. 특히 기상학 분야의 자료 동화 수행을 위해 필요한

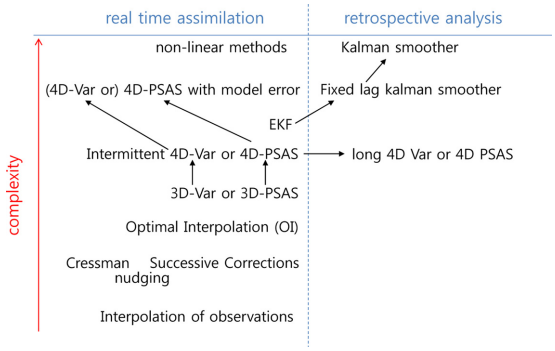

Fig. 1. A summarized history of the main data assimilation algorithms used in meteorology and oceanography, roughly classified according to their complexity (and cost) of implementation, and their applicability to real-time problems from Bouttier and Courtier (1999).

수반 모델 (adjoint model)은 접선 선형 모델 (tangent linear model)의 전치 (transpose)이다.

현재 4-DVAR는 수반 모델 개발의 어려움을 제외하면 실무 적용이 가능하다. 이는 4-DVAR 수행에 필요한 실시간적으로 요구되는 연산량이 현재 지원되는 계산 범위 내에 있음을 의미한다. 복잡성과 유용성에서 4-DVAR와 거의 동일 위상을 가지는 KF의 경우 기상학이나 해양학 분야 종사자들은 실무 적용이 사실상 불가하다고 생각한다 (Fisher, 2001). 이는 수반 연산자의 개발이 가능하지만 BCM의 시간 진전에 필요한 계산 자원을 충족할 수 없기 때문이다. 이러한 상황은 공학 분야와는 뚜렷하게 대비된다 (Alag and Gilyard, 1990).

2.1. 행렬 xt와 yt의 도입

KF는 적용에 있어 시스템의 상태 (state vector)와 이의 정확도나 불확실성을 제공하는 BCM을 동시에 시간 진전시켜야 한다 (Fig. 2). 이 중에서 모델 대기 상태의 진전은 단순히 모델 적분에 해당한다. 그러나 BCM의 진전은 기상 자료 동화에 KF를 사용하고자 하는 경우 심각한 장애 요소이다. 첫째 BCM의 차원이 사용하는 모델의 3차원적인 격자점 수와 변수 수의 곱에 해당하기에 일반적으로 $10^{7\text{-}8}$ 행렬 원소 수를 가지게 되어 기존 방법으로는 진전 자체가 쉽지 않다. 이를 다르게 이야기하면 모델 차원 수만큼 모델 적분을 수행하여야 함을 의미한다. 이론적으로 가능하지만 천문학적인 컴퓨터 리소스를 요구하기에 실현 가능성은 낮다. 둘째 BCM 진전은 수반모델을 필요로 한다.

계산된 BCM을 전진 적분 및 수반 모델로 후진 적분하는 대신, 모델 적분 자료를 일정 기간 동안 보관하여 필요한 시점에서 BCM을 계산하여 전진과 후진

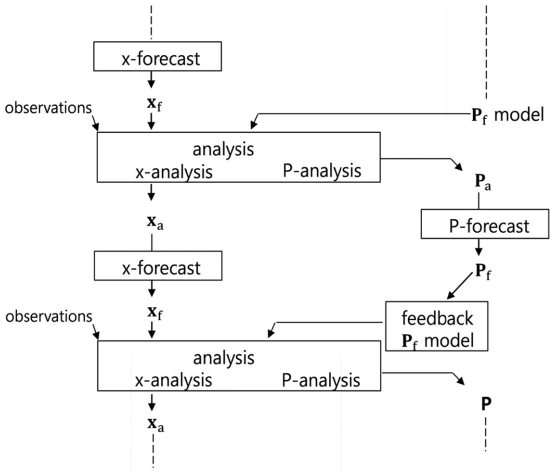

Fig. 2. The organization of computations in a KF or EKF assimilation from Bouttier and Courtier (1999).

에 필요한 계산 자원 중 수반 모델이 필요한 후진 부분을 전진 부분의 전치로 대체할 수 있음을 파악하였다. 계산한 BCM을 취급하는 대신 BCM을 구성단계의 시간 간격을 고려하여야 하기에 다음과 같이 주어진 형식의 **xt**, **yt** 행렬을 도입한다.

$$\mathbf{xt}(i) = [\mathbf{x}(i-(K-1)),\ \mathbf{x}(i-(K-2)),\cdots,\ \mathbf{x}(i-1),\ \mathbf{x}(i)]$$

$$\mathbf{yt}(i) = [H\mathbf{x}(i-(K-1)),\ H\mathbf{x}(i-(K-2)),\cdots,\ H\mathbf{x}(i-1),\ H\mathbf{x}(i)]$$

위 식에서 i는 현재 또는 분석 시각을, **x**는 모델 차원과 변수를 포함하는 주어진 시각의 벡터장이다. 모델 도메인 벡터 **x**는 행렬 **xt**(i)의 각 행을 구성한다. 위 배열의 열은 모델의 특정 격자점에서 주어진 변수의 시각 $i-(K-1)$, $\cdots$, i의 값으로 구성되는 시계열이다. K는 분석 시각으로부터 배열 형성에 필요한 과거 자료의 횟수다. 분석시각을 포함하여 K개의 과거 자료를 이용하여 배열을 구성하는 셈이다. 분석 시각의 모델 자료를 제외하면 K-1개의 과거 자료가 사용된다.

xt의 차원은 모델 격자 점수와 변수의 수를 곱한 크기이며 H**x**의 차원은 관측 지점과 사용되는 변수의 곱에 해당한다. 이 경우 각 변수는 서로 독립으로 가정하며 각 격자점의 변수값도 독립으로 취급한다. **yt**의 각 행은 관측 영역으로 변환된 변수장이다. 위 배열을 구성하는 데 있어서 관측연산자의 비선형성은 중요하지 않다.

위의 **xt**(i)를 이용하여 각 시각에서의 분석장과 전단계의 차이를 이용한 BCM을 위한 배열을 구성할 수 있다. 우리는 BCM을 구성하는 방법으로 완전 모델을 가정한 후 주어진 시각의 분석장과 이전 시간의 분석장을 초기 조건으로 하여 모델 적분한 현 시각의 예보장과의 차이를 이용하는 방법과 기존에 많이 사용되는 NMC (National Meteorological Center) 방법 (Parrish and Derber, 1992)을 생각하였다. 오차 배열은 다음과 같이 표현된다.

$$\delta\mathbf{xt}(i) = [\delta\mathbf{x}(i-(K-1)),\ \delta\mathbf{x}(i-(K-2)),\cdots,\ \delta\mathbf{x}(\mathrm{i}-1),\ \delta\mathbf{x}(i)]$$

위에서 $\delta\mathbf{x}(i-(K-1))$는 분석장과 이전 시간에서 모델 예측장의 차이로 정의하는 경우는

$$\delta\mathbf{x}(i-(K-1)) = \mathbf{x}_{\mathrm{a}}(i-k) - M_{i\to i+1}\mathbf{x}_{\mathrm{a}}(i-(k-1))$$

혹은 NMC 방법을 사용한다고 가정하면 다음과 같다.

$$\delta\mathbf{x}(i-k) = \mathbf{x}^{\mathrm{f}}_{24\mathrm{hours}}(i-k) - \mathbf{x}^{\mathrm{f}}_{12\mathrm{hours}}(i-k)$$

위의 두 식에서 $k = K-1,\cdots,0$이다. 위에 언급된 배열을 형성하기 위해서는 분석 시각이 되는 현 시각의 모델 도메인 자료를 포함하여 K개의 예보 모델 및 분석장 도메인 자료 그리고 자료 동화를 위한 관측 영역 자료의 보관을 필요로 한다. 수행법에 따라서 보관 영역을 조정할 필요가 있다.

2.2. 배경오차 공분산 행렬의 진전

위의 표현을 이용하면 모델을 이용한 분석장의 진전과 BCM 진전은 다음과 같이 표현할 수 있다.

$$\mathbf{x}_{\mathrm{f}}(i+1) = M_{i\to i+1}\mathbf{x}_a(i)$$

$$M_{i\to i+1}\mathbf{P}_a(i)M^{\mathrm{T}}_{i\to i+1} = \delta\mathbf{xt}(i+1)\delta\mathbf{xt}^{\mathrm{T}}(i+1)$$

시간 진전은 새로운 모델 적분을 고려하여 가장 오래된 과거 자료를 제거한 후 새로운 모델 자료를 추가하여 새롭게 계산한 공분산으로 표현된다.

2.3. 칼만필터 이득률과 분석장

KF 이득률 및 분석장을 다음식을 이용하여 과거 자료의 공분산을 사용하여 표현 가능하다.

$$\mathbf{H}(i)\mathbf{P}_{\mathrm{f}}(i)\mathbf{H}^{\mathrm{T}}(i) = \mathbf{yt}(i)\mathbf{yt}^{\mathrm{T}}(i)$$

$$\mathbf{P}_{\mathrm{f}}(i)\mathbf{H}^{\mathrm{T}}(i) = \mathbf{xt}(i)\mathbf{xt}^{\mathrm{T}}(i)\mathbf{H}^{\mathrm{T}}(i) = \mathbf{xt}(i)(\mathbf{H}(i)\mathbf{xt})^{\mathrm{T}}$$
$$= \mathbf{xt}(i)\boldsymbol{y}\mathbf{t}^{\mathrm{T}}(i)$$

$$\mathbf{H}(i)\mathbf{P}_{\mathrm{f}}(i) = \mathbf{H}(i)\mathbf{xt}(i)\mathbf{xt}^{\mathrm{T}}(i) = \mathbf{yt}(i)\mathbf{xt}^{\mathrm{T}}(i)$$

위의 식을 이용하면 KF 이득률은 다음과 같이 표현된다.

$$\mathbf{K}(i) = \mathbf{xt}(i)\mathbf{yt}^{\mathrm{T}}(i)[\mathbf{yt}(i)\mathbf{yt}^{\mathrm{T}}(i)+R(i)]^{-1}$$

따라서 분석장과 분석 BCM은 다음과 같이 표현된다.

$$\mathbf{x}_a(i) = \mathbf{x}_{\mathrm{f}}(i)+\mathbf{xt}(i)\mathbf{yt}^{\mathrm{T}}(i)[\mathbf{yt}(i)\mathbf{yt}^{\mathrm{T}}(i)+R(i)]^{-1}[\mathrm{y}(i)-H(i)\mathbf{x}_{\mathrm{f}}(i)]$$

$$\mathbf{P}_a(i) = \mathbf{P}_{\mathrm{f}}(i)-\mathbf{xt}(i)\mathbf{yt}^{\mathrm{T}}(i)[\mathbf{yt}(i)\mathbf{yt}^{\mathrm{T}}(i)+R(i)]^{-1}\mathbf{yt}(i)\mathbf{xt}^{\mathrm{T}}(i)$$

위 계산의 경우 시각 i와 $i+1$ 단계에서의 모델 적분에 따른 차이는 $k=\mathrm{K}-1$의 오차이다. 오차는 위에서 언급한 방법이나 기타 다른 방법으로 계산 가능하며 이는 NMC 방법의 경우 동일 모델의 초기 시각을 달리하여 병행 수행하는 두 모델의 적분값의 동일 시각에서의 차이로 구할 수 있다. 이 계산에 소요되는 계산 자원의 양은 기존 방법이나 ENKF 사용에 필요한 계산양과 비교함은 그 비율을 생각할 때 아무런 의미가 없다. 각 시각의 해당 모델장에 가중치를 적용할 수 있지만 본 논문에서는 보다 용이한 이해를 도모하고자 생략한다. 이러한 취급은 모든 k에 대하여 동일한 값을 가정하며 가중치의 합을 1로 보는 것이다. 위 식은 각 격자점의 변수에 대하여 현 시각 혹은 분석 시각을 기준으로 과거 시각 $i-(K-1)$과 현 시각 i 동안의 시간 변동성을 고려한 공분산이다. 평균값으로는 시간평균 값을 사용한다.

공분산 계산을 위한 자료 저장 공간은 K의 값에 좌우되며 현재의 컴퓨터 성능으로 충분하다. 불필요한 자료의 삭제와 최신 자료의 추가에 필요한 계산 자원 역시 $\delta\mathbf{x}(i-(k-1))$가 제외되고 $\delta\mathbf{x}(i+1)$의 오차가 추가됨을 고려하면 미미하다. 즉, 시각 $i-(k-1)+1$부터 시각 $i-1$까지의 오차는 중복되기 때문에 시각 i에서 사용된 오차를 그대로 사용할 수 있다. 이러한 방법을 수행하기 위해서는 자료 보관을 위한 추가적인 기억 장소나 하드 디스크가 필요로 하나, 최근의 컴퓨터 성능 고려 시 필요한 저장 공간 및 계산 자원은 미미하다.

3. 실험 설계 및 결과

일반적인 대기의 비선형성을 고려할 수 있는 개념적 모형으로 로렌츠 40변수 모델을 선택하여 실험을 하였고 (Lorenz, 1995), 그 형태는 다음과 같다.

$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1}+x_{i-1}x_{i+1}-x_i+F$$

위 식에서 $\mathrm{i}=1,2,\cdots,40$이고 $x_0=x_{40}$, $x_{-1}=x_{39}$, $x_{41}=x_1$인 순환경계조건 (cyclic boundary condition)을 설정하였다. 또한 로렌츠 모델에서 대기의 혼돈 (chaotic)을 고려하여 $\mathrm{F}=8$로 주었으며, 4차 Runge-Kutta 기법으로 모델을 적분하였다. 실험은 먼저 특정 초기조건으로 수행된 모형의 결과를 실제 대기라 가정하고, 2일 동안 동화를 한 후 (3시간 간격으로 16번의 동화를 수행) 10일을 예측하도록 설계하였다. 본 논문에서 제시한 배경오차 진전은 2.2.절에서 언급한 바와 같이 48시간 동안의 자료를 저장하여 가장 오래된 자료를 제거한 후 분석시각이 되는 현 시각에서의 분석장과 이전 시간의 분석장을 초기 조건으로 하여 모델 적분한 현 시각의 예측장과의 차를 추가하여 구하였다. KF와 본 논문에서 제시한 방법을 비교하기 위해 100번의 예측 실험을 하였으며, 실험에서 배경 오차 진전 이외의 모든 변수는 동일하다. Fig. 3은 실제 대기와 각 자료 동화 시스템에서 얻은 분석장의 차이이고, Fig. 4는 실제 대기의 예측과 각 자료 동화 시스템의 예측 차이이다. Fig. 3에서는 본 논문에서 제시한 방법이

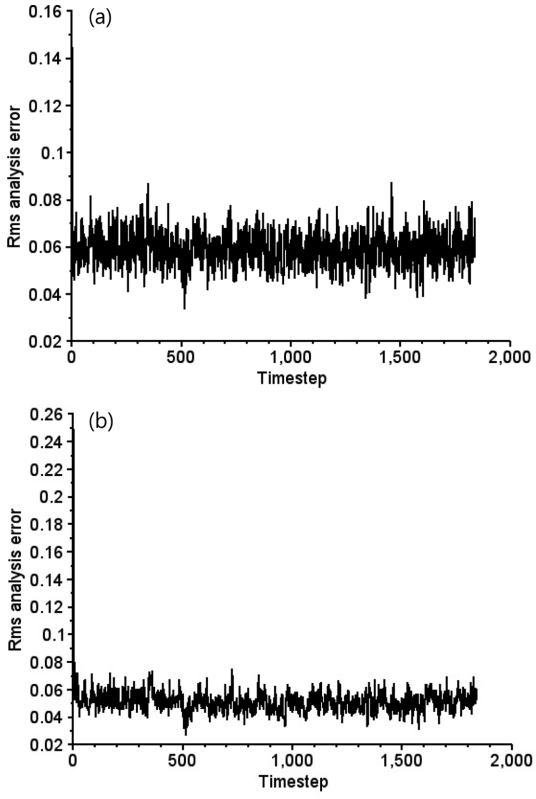

Fig. 3. RMS of analysis error in (a) original Kalman Filter and (b) affordable implementation of Kalman filter by eliminating the explicit temporal evolution of the background error covariance matrix.

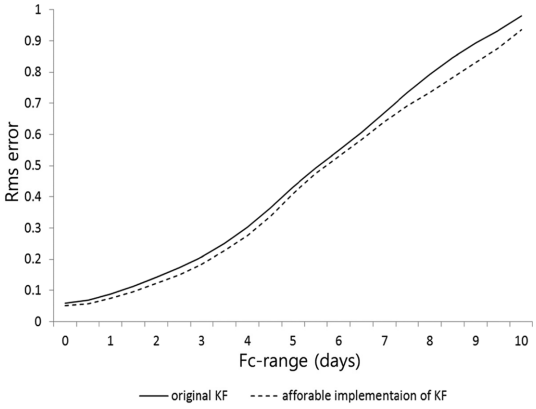

Fig. 4. RMS of forecast error in original Kalman Filter and affordable implementation of Kalman filter by eliminating the explicit temporal evolution of the background error covariance matrix.

KF에 비해 처음에는 상대적으로 높은 값을 보이다가 시간이 지남에 따라 안정화되는 모습을 볼 수 있으며, 이를 통해 본 논문에서 제시한 방법이 자료 동화 시스템에 잘 반영된다는 것을 알 수 있다. 또한 예측 오차에서도 10일 예측에 대해 본 논문에서 제시한 방법이 KF와 유사한 형태의 오차를 모의한다는 것을 알 수 있다 (Fig. 4). Fig. 3과 Fig. 4의 분석을 통해 본 논문에서 제시한 방법은 기존의 KF와 견주어 필요한 계산량을 줄이면서 배경 오차를 진전시킬 수 있으며, 동화 및 예측 결과 또한 타당하다는 것을 보여준다. 향후 실무 적용을 위해 대기의 안정도에 따른 실험, 저장되는 자료의 수, 해당 모델장에 가중치 적용과 국지적인 배경 오차 사용 등의 보다 많은 실험을 수행하여 결과를 분석할 예정이다.

4. 실무 적용

실무 적용을 위해서는 변수들 사이의 공분산양의 조정 및 선형성과 비선형 연산자의 성격을 엄밀하게 분석 적용할 필요가 있을 것이다. 이처럼 실무 적용을 위해서 해결하여야 할 문제점이 있지만, 계산양의 획기적인 감소 내지는 원천적인 제거가 가능함을 이론적인 차원에서 보였다. 우리가 제안하는 방법은 BC가 소개한 EKF의 연장선상에 있는 KF이다.

5. 결 론

우리는 KF의 이론을 실무 차원 적용에 있어서 BCM 시간 진전 부분을 아예 제거하였다. 이러한 시도는 계산 차원에서 기상 분야의 초고차원 시스템의 자료 동화에 적용할 수 있는 방법론을 제시하는 것으로, 약간의 수정으로 4-DVAR와 ROI에도 적용할 수 있다. 장차 실무차원 적용에 수반될 문제점의 극복 방안을 생각하고자 한다. 유사한 방법으로 추후 4-DVAR처럼 분석 시각 이후의 미래 자료를 감안할 수 있는 칼만 스무더의 수행법을 다룰 예정이다.

감사의 글

이 연구는 기상청 기상기술개발사업 (CATER 2012-6080)의 지원으로 수행되었습니다.

참고문헌

Alag, G. S. and G. B. Gilyard, 1990: A proposed Kalman Filter Algorithm for Estimation of Unmeasured Output Variables for an F100 Turbofan Engine. NASA Technical Memorandum 4234, 29 pp.

Bouttier, F. and P. Courtier, 1999: Data assimilation concepts and methods. *Meteorological Training Course Lecture Series*. ECMWF.

Grewal, M. S. and A. P. Andrews, 2001: *Kalman Filtering: Theory and Practice Using MATLAB,* second edition. Wiley Inter-Science, 401 pp.

Fisher, M., 2001: Assimilation Techniques (5): Approximate Kalman Filters and Singular Vectors. *Meteorological Training Course Lecture Series.* 10 pp.

Lorenz, E. M., 1995: Predictability: A problem partly solved. *In Seminar on Predictability*, Vol. I, ECMWF, Reading, UK, 1-18.

Mcgee, L. A. and S. F. Schmidt, 1985: Discovery of the Kalman Filter as a Practical tool for Aerospace and Industry, NASA technical memorandum 86847, 21 PP.

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical Interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1763.

Song, H.-J. and G.-H. Lim, 2009: An efficient retrospective optimal interpolation algorithm compared with the fixed-lag Kalman smoother by assuming a perfect model. *Tellus A*, **61**, 610-620.