基于 EnKF 和 ETKF 的比较研究

陈 丽，卞康旭

（中国民用航空飞行学院理学院，四川 广汉 618307）

摘 要 在数据同化问题中，集合卡尔曼滤波因其简单的概念公式和相对易于实现而广受欢迎，它不需要推导切线性算子或伴随方程，也不需要时间上的向后积分。此外，其计算复杂度与其他流行的复杂同化方法相当。目前，应用最广泛的两种算法为随机集合卡尔曼滤波（EnKF）与集合变换卡尔曼滤波（ETKF），但对于两种算法的比较研究较少，因此，本文将在线性模型、弱非线性模型以及强线性模型上比较 EnKF 与 ETKF 的同化效果，为了防止算法发散，所有实验都引入了协方差膨胀。实验结果表明，在线性与强非线性模型上，ETKF 算法结果优于 EnKF，而在弱非线性模型上，观测误差较小时，ETKF 算法性能优于 EnKF，但观测误差较大时，EnKF 算法性能却优于 ETKF。

关键词 数据同化；随机集合卡尔曼滤波；集合变换卡尔曼滤波；协方差膨胀

中图分类号：O1 **文献标识码**：A **文章编号**：2097-3365(2023)12-0001-03

1 前言

预测物理系统通常需要系统的时间演变模型和系统当前状态的估计。在一些应用中，可以高精度地直接测量系统的状态。在其他应用中，如天气预报，直接测量全球系统状态是不可行的。相反，必须从可用数据中推断出状态。虽然基于当前数据的状态估计是合理的，但通常可以通过使用当前和过去的数据来获得更好的估计[1]。“数据同化”在持续的基础上提供这种估计，在预测步骤和状态估计步骤之间迭代交替；后一个步骤通常被称为“分析”。分析步骤结合来自当前数据和先前短期预测（基于过去数据）的信息，产生当前状态估计。该估计用于初始化下一个短期预测，随后用于下一次分析等[2]。

同化方法主要有两种类型：变分同化方法和序列同化方法。序列数据同化方案显式求解一系列方程，以找到系统的分析状态。顺序方法的例子是卡尔曼滤波器（KF）[3]以及从 KF 的基础上导出的各种滤波器。例如，集合卡尔曼滤波器（EnKF）[4]、集成变换卡尔曼滤波器（ETKF）[5]都是顺序方案。

集合卡尔曼滤波器由Geir Evensen于1994年提出，并于1998年进行了修正[6]。它的半经验性论证可能令人不安。尽管如此，集合卡尔曼滤波器已被证明在大量的学术和操作数据同化问题上是非常有效的。集合卡尔曼滤波器（EnKF）是一个降阶卡尔曼滤波器，它只处理二阶以下的误差统计信息。由于其在计算过程中对引入的观测信息进行了随机扰动，又称为随机性集合卡尔曼滤波。

随机 EnKF[7]能够单独跟踪集合中的每个成员，并通过分析步骤使它们相互作用。这个好的性质是有代价的：其需要独立地扰动每个成员的观测向量。虽然这看起来易于实现，但在随机扰动时也引入了数值噪声。这可能会影响算法的性能，特别是当单个分析的观测数量有限时，可能会造成算法发散。为了解决这个问题，发展出了另一种类型的滤波算法，称为确定性集合卡尔曼滤波。在本文中，将重点关注确定性 EnKF 的集合变换变体，通常缩写为 ETKF。该算法不是在状态空间或观测空间中执行计算，而是在集合空间中执行。

目前，对于随机性和确定性集合卡尔曼滤波算法的比较研究较少，本文将基于不同实验模型比较随机 EnKF 与 ETKF 的算法性能。

2 研究方法

2.1 随机性集合卡尔曼滤波（EnKF)

2.1.1 分析步骤

在 EnKF 算法框架中[8]，先从状态集合开始分析，记状态集合为：

$$E^a=[x_a^1, x_a^2, \ldots, x_m^a]\in R^{n\times m} \quad (1)$$

其中 x_i^a 表示在分析步骤的集合中的第 i 个状态向量。

将观测向量进行扰动 $y_i=y+u_i$，其中 u_i 来自高斯分

★基金项目：中国民用航空飞行学院青年基金（QJ2023-038）学生科研项目（XSB2023-125）。

布 $u_i \sim N(0,R)$。首先，确保 $\Sigma_{i=1}^{m} u_i=0$ 以避免偏差。其次，定义经验误差协方差矩阵：

$$R_u=\frac{1}{m-1}\sum_{i=1}^{m} u_i(u_i)^T \tag{2}$$

当 $m\to\infty$ 时，R_u 的极限应趋于 R。

$$K_u^*=P^f H^T(HP^f H^T+R_u)^{-1} \tag{3}$$

其中：

$$\bar{y}^f=\frac{1}{m}\sum_{i=1}^{m} H(x_i^f)$$

$$P^f H^T \simeq \frac{1}{m-1}\sum_{i=1}^{m}(x_i^f-\bar{x}^f)(H(x_i^f)-\bar{y}^f)^T$$

$$HP^f H^T \simeq \frac{1}{m-1}\sum_{i=1}^{m}(H(x_i^f)-\bar{y}^f)(H(x_i^f)-\bar{y}^f)^T \tag{4}$$

则分析集合为 $x_i^a=x_i^f+K_u^*(y_i-H(x_i))$，均值为：

$$\bar{x}^a=\frac{1}{m}E^a 1=\frac{1}{m}\sum_{i=1}^{m} x_i^a$$

那么分析误差协方差矩阵为：

$$P^a=(I_n-K_uH)P^f(I_n-K_uH)^T+K_uRK_u^T=(I_n-K_u^*H)P^f \tag{5}$$

2.1.2 预报步骤

在预测步骤中，在分析中获得的更新集合通过一个时间步长由模型传播到下一个瞬间：

$$x_i^f=M(x_i^a),\ i=1,2,\cdots,m \tag{6}$$

预报估计是预报集合的平均值：

$$\bar{x}^f=\frac{1}{m}E^f 1=\frac{1}{m}\sum_{i=1}^{m} x_i^f \tag{7}$$

而预测误差协方差矩阵则由：

$$X^f=E^f-\bar{x}^f 1,\ P^f=\frac{1}{m-1}X^f(x^f)^T \tag{8}$$

重要的是要注意，在算法过程中已经避免使用切线性算子，或任何线性化。

2.2 集合变换卡尔曼滤波(ETKF)

ETKF 与 EnKF 的不同之处在于分析集合的构造。在下面的讨论中，定义集合空间如下：

$$\Omega:=\left\{x \middle| x=\bar{x}^f+X^f w: w\in R^m\right\} \tag{9}$$

对于分析估计，利用卡尔曼增益的形式得到均值的最优系数向量 w^a：

$$w^a=((m-1)I_m+Y^{f^T}R^{-1}Y^f)^{-1}Y^{f^T}R^{-1}\delta \tag{10}$$

为了生成一个后验集合，下面需要分解。

$$P^a=(I_n-K^*H)P^f\simeq(I_n-\frac{1}{m-1}X^fY^{f^T}(\frac{1}{m-1}Y^fY^{f^T}+R)^{-1}H)\times\frac{1}{m-1}X^f(X^f)^T\simeq\frac{1}{m-1}X^fC(X^f)^T \tag{11}$$

其中 $C=I_m-Y^{f^T}(Y^fY^{f^T}+(m-1)R)^{-1}Y^f$。

定义 $T=(I_m+Y^{f^T}R)^{-1}Y^f)^{-1/2}$，利用矩阵 T 可以构建后验集合为 $E^a=\bar{x}^a1^T+X^fT$。

2.3 乘法膨胀

为了防止滤波发散，引入乘法膨胀，其简单地将集合误差协方差 P^f 进行处理，以近似真实误差协方差 P^f: $P^f\leftarrow\lambda P^f$，其中 λ 是膨胀因子。

3 数值实验

本节提供了 EnKF、ETKF 在三种模型上的一些实验说明，一共进行了6个实验，从LA模型的2个实验开始，其次是 Lorenz3 模型的 2 个实验，最后是 Lorenz96 模型的2个实验。这些实验的设置与文献[9]中的设置相同。对于实验性能用以下“分析 rmse”进行考量：

$$RMSE(x,x^{真})=[1/n(x-x^{真})^T x-x^{真}]^{1/2}$$

其中，x 和 $x^{真}$ 分别为待估计的模型状态和真实的模型状态。

3.1 LA 模型的实验

在本小节中，基于线性平流模型（LA）进行了 2 个实验，线性平流模型表达如下：

$$x(t)=[x_1,(t),\cdots,x_n(t)],\ t=1,2,\cdots,n$$

$$x_i(t+1)=\begin{cases}x_{i-1}(t), i=2,\cdots,n\\ x_n(t), i=1\end{cases}$$

状态样本 x_i 由具有随机幅度和相位的 n_k 正弦波组成，表示为：

$$x^i=\sum_{k=1}^{n_k} a^{(k)}sin(\frac{2k\pi}{n}i+\varphi^{(k)}),1\le i\le n$$

$$a^{(k)}=\mathrm{rand}()\exp[-\frac{1}{2}(\frac{\mathrm{k}-\mathrm{k}_{\max}}{\mathrm{k}_{\mathrm{w}}})],\varphi^{(k)}=\mathrm{rand}()\times 2\pi$$

表 1 LA 模型

同化算法	EnKF	ETKF
n=100, p=25, m=30, R=1		
RMSE	0.642	0.585
膨胀因子	1.02	1.02
同化时间 /s	0.051	0.143
n=100, p=25, m=30, R=0.01		
RMSE	0.583	0.562
膨胀因子	1.02	1.02
同化时间 /s	0.052	0.091

rand() 函数返回一个均匀分布的随机数，0≦rand()≦1, exp() 是 e 的指数函数。实验基本设置都是相同的。

唯一的区别在于观测误差。系统运行 700 个同化周期，前 100 个循环被丢弃。超过 600 个循环的平均结果见表 1。

从表 1 中可知，对于线性模型，不论观测误差大小，ETKF 算法性能都优于 EnKF。

3.2 Lorenz3 模型的实验

在本小节中，基于 Lorenz3 模型进行了 2 个实验，Lorenz3 模型方程如下所示：

$\dot{x}=\sigma(y-x), \dot{y}=\rho x-y-xz, \dot{z}=xy-\beta z$

其中为 $\sigma=10$、$\rho=28$ 和 $\beta=8/3$。该系统用四阶龙格 - 库塔方案积分，采用固定的时间步长为 0.01。

与 LA 模型一样，系统运行 700 个同化周期，前 100 个循环被丢弃。超过 600 个循环的平均结果见表 2。

表 2 Lorenz3 模型

同化算法	EnKF	ETKF
n=3, p=3, m=2, R=2		
RMSE	0.903	1.631
膨胀因子	1.02	1.02
同化时间 /s	0.027	0.035
n=3, p=3, m=2, R=0.01		
RMSE	1.190	0.953
膨胀因子	1.02	1.02
同化时间 /s	0.022	0.028

从表 2 中可知，在 Lorenz3 模型中，观测误差较小时，ETKF 算法性能优于 EnKF，但观测误差较大时，EnKF 算法性能却优于 ETKF。

3.3 Lorenz96 模型的实验

Lorenz96 表达式为：

$\dot{x}_i=(x_{i+1}-x_{i-2})x_{i-1}-x_i+8, i=1, \cdots, 40$

$x_0=x_{40}, x_{-1}=x_{39}, x_{41}=x_1$

表 3 Lorenz96 模型

同化算法	EnKF	ETKF
n=40, p=40, m=35, R=2		
RMSE	3.84	2.08
膨胀因子	1.01	1.01
同化时间 /s	0.051	0.096
n=40, p=40, m=35, R=0.01		
RMSE	4.71	4.63
膨胀因子	1.01	1.01
同化时间 /s	0.050	0.091

在 Lorenz96 模型中，依然进行了 2 个实验。系统依然运行 700 个同化周期，前 100 个循环被丢弃。超过 600 个循环的平均结果见表 3。

如表 3 所示，在 Lorenz96 模型中，和 LA 模型结果相同，不论观测误差大小，ETKF 算法性能都优于 EnKF，不过 EFKF 同化时间始终高于 EnKF。

4 结论

随机性集合卡尔曼滤波与集合变换卡尔曼滤波都是经典的数据同化算法，在其他实验条件相同的情形下，两者算法的性能表现出些许不同。对于线性模型（如 LA 模型）和强非线性模型（如 Lorenz96 模型）来说，不论观测误差大小，ETKF 的算法性能始终优于 EnKF，但其同化时间比 EnKF 较多，不过在算法性能的优势下，时间成本可忽略不计。对于弱非线性模型（如 Lorenz3 模型），观测误差较小时，ETKF 算法性能优于 EnKF，但观测误差较大时，EnKF 算法性能却优于 ETKF。

参考文献：

[1] M. Ghil and M. Paola.Data Assimilation in Meteorology and Oceanography[J].Advances in Geophysics,1991(33): 141-266.
[2] E. Kalnay.Atmospheric Modeling, Data Assimilation and Predictability[M].Cambridge: Cambridge UP,2002.
[3] Kalman, R. E..A new approach to linear filter and prediction problems[J]J. Basic. Eng.,1960(82):35-45.
[4] G. Evensen .Data Assimilation: The Ensemble Kalman Filter (4th edn)[M].Springer-Verlag: Berlin Heidelberg,2009.
[5] Bishop, C. H., B. J. Etherton, and S. J. Majumdar, Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects[J].Mon. Weather Rev,2001(129):420-436.
[6] Burgers, G, P. J. van Leeuwen, and G. Evensen, Analysis scheme in the ensemble Kalman filter[J].Mon. Weather Rev, 1998(126):1719-1724.
[7] Houtekamer, P. L. and H. L. Mitchell, Data assimilation using an Ensemble Kalman Filter technique[J].Mon. Weather Rev., 1998(126):796-811.
[8] G. Evensen,Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics[J].J.Geophys. Res,1994(99):10143-10162.
[9] P.Sakov,S.Oliver and L.Bertino.An Iterative EnKF for Strongly Nonlinear Systems[J].Monthly Weather Review, 2012(140.6):1988-2004.