

两种四维奇异值分解同化方法的比较及误差分析

王金成^{1, 2} 李建平² 丑纪范³

1 兰州大学大气科学学院, 兰州 730000

2 中国科学院大气物理研究所大气科学和地球流体力学数值模拟国家重点实验室, 北京 100029

3 中国气象局培训中心, 北京 100081

摘 要 4DSVD 是最近提出的一种新的资料同化方法。目前还存在一些需要解决的问题, 比如如何选取样本, 如何得到支撑大气吸引子的基向量以及选取基向量的个数问题等等。作者利用奇异值分解 (SVD) 与经验正交函数分解 (EOF) 两种方法来获得支撑大气吸引子的基向量, 推导了基于这两种方法的 4DSVD 分析场的理论公式, 并用简单的数值试验比较了基于这两种方法的 4DSVD 分析场的空间相关系数和误差, 初步分析了分析场与基向量个数的关系以及与样本选取的关系和分析误差的来源及各种误差对分析误差影响的相对大小。结果表明, 用 SVD 方法作为获得支撑大气吸引子基向量的方法得到的分析场较 EOF 方法稳定, 分析场与基向量个数有密切关系, 观测误差、模式误差和观测代表性误差是分析误差的主要来源, 且其引起的分析误差随着基向量个数增多而增大。

关键词 四维奇异值分解 资料同化 经验正交函数 奇异值分解 误差分析

文章编号 1006-9895 (2008) 02-0277-12

中图分类号 P413

文献标识码 A

Comparison and Error Analysis of Two 4-Dimensional Singular Value Decomposition Data Assimilation Schemes

WANG Jin-Cheng^{1, 2}, LI Jian-Ping², and CHOU Ji-Fan³

1 College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000

2 State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029

3 China Meteorological Administration Training Center, Beijing 100081

Abstract A new four-dimensional data assimilation method named 4DSVD based on attractor theory is introduced by Qiu and Chou (2006). This 4DSVD scheme solves the data assimilation problem in the atmosphere attractor phase space spanned by the base vectors which are created from a set of coupled atmospheric states through the Empirical Orthogonal Function (EOF). The coupled atmospheric states are composed of model states and simulated observations. Because the dimension of atmosphere attractor is much smaller than the atmosphere itself, the degree of freedom of data assimilation problem decreases significantly. 4DSVD is a linear method so it needs much less computation time than the other data assimilation method such as 4DVAR. It can produce good initial conditions for numerical models. Some numerical experiments results suggest that the scheme of 4DSVD using EOF is unstable which means that EOF is not an efficient method to obtain base vectors from the samples of coupled atmosphere states. So a new more efficient method is expected. The authors introduce a new theory scheme of 4DSVD using the Singular Value Decomposition (SVD) in this paper. 4DSVD is a new data assimilation method and there are some problems

such as how to generate the sample set, how to produce the base vectors spanned the atmosphere attractor and how many base vectors are used, and so on. This paper tries to answer some of the questions above and compare the effectiveness of two 4DSVD schemes based on the EOF and the SVD, some ideal numerical experiments are shown in this paper. The authors also study the sources of the 4DSVD analysis error. Our results suggest that the 4DSVD scheme based on the SVD is better than the other using the EOF, and the sample error, the truncation error, the representative error, the model error and the observation error are the main sources of the analysis error of 4DSVD. And the results the authors present also demonstrate that the analysis error increases with the increased model error and observation error. The analysis error generated by the representative error, the model error and the observation error increases with increasing base vector number, while the analysis error generated by the truncation error decreases with increasing base vector number.

Key words 4DSVD, data assimilation, singular value decomposition, empirical orthogonal function, error analysis

1 引言

数值天气预报是一个初值问题, 由于其中存在强的非线性过程, 结果对初值是十分敏感的^[1]。因此对天气预报模式来说, 初始条件的确定既是非常重要的, 也是非常复杂的, 它的精确程度直接影响着数值天气预报的水平。

资料同化已经成为数值天气预报不可或缺的一部分, 它已经成为一门独立的学科^[2], 其目的是给数值天气预报提供尽可能好的动力上与模式协调一致的初始条件。Gandin^[3]提出一种基于要素场本身统计结构, 并应用观测场和背景场, 在最小二乘意义下寻找最优的权重函数的“最优插值(OI)”。由于背景场误差协方差矩阵在实际应用中很难确定, 并且计算复杂, 在实际应用中一般都要做很大的简化^[3, 4]; OI不能得到动力协调一致的结果, 需要进行额外的初始化。在随后发展起来的三维变分同化(3DVAR)^[5]减少了OI的许多假设, 并使分析场得到了优化, 特别是能够同化非模式变量观测资料。考虑背景协方差矩阵随时间演变并利用多个时刻观测资料的四维变分同化(4DVAR)^[6~9]以及集合卡曼滤波(EnKF)^[10]是更先进的同化方法, 这两种方法都充分利用背景场和观测场的信息, 可以更好地同化非模式资料和利用多时刻的观测资料^[11, 12], 并能够保证分析场动力协调一致。理论上说4DVAR是解决资料同化问题较好的方法, 但是4DVAR需要编写和运行切线性模式和伴随模式, 工作量和计算量都非常大, 尤其是在模式有误差的情况下, 计算量会更大^[13], 因此同化时间窗不能太大, 并且需要加入背景场作为约束。对于模式中各种物理过程参数化时出现的不连续问题, 即条件语

句伴随模式的书写也存在很多困难^[14, 15]。这些困难妨碍了4DVAR的发展和运用。3DVAR是4DVAR的一种简化, 对于中小尺度模式3DVAR不能够用完全的动力方程作为约束, 不能保证分析场动力协调一致。EnKF是最近几年发展起来的一种先进的同化方法, 其优点是可以估计更新背景场误差协方差矩阵并避免了计算切线性伴随模式, 但如何产生预报集合和决定模式误差是它存在的问题, 虽然现在许多的研究集中在如何产生预报集合和决定模式误差^[16, 17], 取得了一定的进展, 但仍然存在很多的问题。

研究表明^[18, 19], 大气是强迫耗散的强非线性混沌系统, 存在吸引子, 其维数是有限的, 并且远远没有看起来的那么大。Qiu和Chou^[20]以此理论为基础提出了一种新资料同化方法(4DSVD), 这种方法试图将资料同化问题的解限定在吸引子相空间上, 采用线性反演方法得到分析场。4DSVD将同化问题转化为简单的线性反演问题, 大大降低了分析的维数使计算简便, 它把观测的大气状态投影到模式大气吸引子相空间上, 保证了分析场动力协调一致。

4DSVD是一种很有前途的同化方法, 但是由于它是刚被提出的, 还存在许多有待研究的问题。现在需要解决的几个关键问题是如何选取样本, 如何得到支撑大气吸引子的基向量以及基向量个数问题等等。本文尝试用奇异值分解(SVD)方法获取支撑大气吸引子基向量的方法推导出详细的分析场公式, 并做了简单的数值试验, 比较获取支撑吸引子基向量的两种方法的效果, 并且给出了分析误差与基向量个数的关系, 以及分析误差与样本选择方式的关系, 试图找到最优的基向量个数。对于一个

新的同化分析方法, 我们不仅要得到好的结果, 更重要的是弄清楚如何提高改善这种方法, 并努力使它能够应用于实际分析, 因此本文详细分析了 4DSVD 分析误差的主要来源, 分析了各种误差对分析误差的影响, 为更加深入地理解 4DSVD 以及实际应用打下一定的基础。

2 基本公式

Qiu 和 Chou^[20]给出了资料同化新方法 4DSVD 的理论框架和基本公式, Zhang 和 Chou^[21]证明了用 EOF 方法获得支撑气候系统吸引子基向量的可行性和有效性, 因此, 他们建议采用 EOF 方法获取支撑吸引子的基向量。而这种方法随着基向量个数的增加, 分析场会出现不稳定 (见第 4 节数值试验结果), 因此本文采用 SVD 方法获取支撑吸引子的基向量。本文所用到的主要符号及含义如表 1 所示。表 1 中 m 、 p 分别是模式和观测变量的个数。

观测向量和模式状态向量有下面的关系:

$$\boldsymbol{d}^{\text{fsim}} = \boldsymbol{H}(\boldsymbol{\psi}^{\text{f}}), \quad (1)$$

$$\boldsymbol{d}^{\text{t}} = \boldsymbol{H}(\boldsymbol{\psi}^{\text{t}}). \quad (2)$$

设有 N 个模式大气样本向量, 以及模拟观测向量:

$$\begin{aligned} \boldsymbol{\psi}_i^{\text{f}} &= \boldsymbol{F}_i \boldsymbol{\psi}_{i-1}^{\text{f}}, \\ \boldsymbol{d}_i^{\text{fsim}} &= \boldsymbol{H}(\boldsymbol{\psi}_i^{\text{f}}) \quad (i = 1, 2, \dots, N). \end{aligned} \quad (3)$$

2.1 EOF 方法

将模式样本和模拟观测向量组合成新的向量:

$$\boldsymbol{\phi}_i = \begin{bmatrix} \boldsymbol{\psi}_i^{\text{f}} \\ \boldsymbol{d}_i^{\text{fsim}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\psi}_i^{\text{f}} \\ \boldsymbol{H}(\boldsymbol{\psi}_i^{\text{f}}) \end{bmatrix}, \quad (4)$$

这里 $\boldsymbol{\phi}_i$ 称为广义大气状态。 N 个广义大气状态向量样本可以构成一个新矩阵:

$$\boldsymbol{\Phi} = (\boldsymbol{\phi}_i)_{(m+p) \times N}. \quad (5)$$

表 1 符号及含义
Table 1 Notations and meanings

数学符号	含义
$\boldsymbol{\psi}^{\text{t}}$	准确的模式状态向量 (m)
$\boldsymbol{\psi}^{\text{f}}$	预报的模式样本向量 (m)
$\boldsymbol{\psi}^{\text{a}}$	分析向量 (m)
$\boldsymbol{d}^{\text{o}}$	同化时间窗内实际观测向量 (p)
$\boldsymbol{d}^{\text{t}}$	准确的观测向量 (p)
$\boldsymbol{d}^{\text{fsim}}$	观测空间的模拟观测向量 (p)
$\boldsymbol{H}$	观测算子 ($m \times p$)
$\boldsymbol{F}$	模式算子

文献[18]认为支撑广义大气吸引子的基向量可以从样本相空间中获得。下面采用 EOF 方法来获得支撑大气吸引子的基向量。

设

$$\boldsymbol{A} = \boldsymbol{\Phi} \boldsymbol{\Phi}^{\text{T}}. \quad (6)$$

对广义大气状态向量组成的矩阵进行 EOF 分解, 得

$$\boldsymbol{V}^{\text{T}} \boldsymbol{A} \boldsymbol{V} = \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{m+p}), \quad (7)$$

其中, $\boldsymbol{V}$ 是广义大气状态的特征向量所组成的矩阵, 其列向量是特征向量, 并且解释方差从大到小排列; $\boldsymbol{\Lambda}$ 是对角矩阵, 对角元素是特征值, 并且有 $\lambda_j > \lambda_{j+1} \geq 0$ 。

特征向量矩阵 $\boldsymbol{V}$ 可以分解为:

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{R}^{\text{U}} \\ \boldsymbol{D} & \boldsymbol{R}^{\text{D}} \end{bmatrix}, \quad (8)$$

其中 $\boldsymbol{U}$ 、 $\boldsymbol{D}$ 是支撑广义大气相空间基向量的模式部分和观测部分所组成的矩阵, 维数是 $2s+1$ 。 $\boldsymbol{R}^{\text{U}}$ 、 $\boldsymbol{R}^{\text{D}}$ 是噪声部分。设吸引子的维数是 s , 根据 Whitney 定理, 吸引子能够嵌入 R^{2s+1} 空间^[21]。因此, 广义大气相空间中的任何时刻大气状态都可以表示成前 $2s+1$ 个特征向量的线性组合:

$$\boldsymbol{\phi} = \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{D} \end{pmatrix} \boldsymbol{x} = \sum_{j=1}^{2s+1} \begin{pmatrix} \boldsymbol{U}_j \\ \boldsymbol{D}_j \end{pmatrix} x_j = \begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{d} \end{pmatrix}, \quad (9)$$

其中,

$$\boldsymbol{x} = (x_1, x_2, \dots, x_{2s+1})^{\text{T}} \quad (10)$$

是组合系数列向量。

从式 (9)、(10), 我们得到模式相空间大气状态和观测相空间大气状态, 可以表示为

$$\boldsymbol{d} = \boldsymbol{D} \boldsymbol{x}, \quad (11)$$

$$\boldsymbol{\psi} = \boldsymbol{U} \boldsymbol{x}. \quad (12)$$

在分析时刻 t 有实际观测 $\boldsymbol{d}^{\text{o}}$, 可以表示成:

$$\boldsymbol{d}^{\text{o}} \approx \boldsymbol{D} \boldsymbol{x}. \quad (13)$$

一般情况下有 $p > 2s+1$, 线性方程 (13) 是超定方程, 用最小二乘法可以求得方程 (13) 的观测资料的组合系数:

$$\boldsymbol{x} \approx (\boldsymbol{D}^{\text{T}} \boldsymbol{D})^{-1} \boldsymbol{D}^{\text{T}} \boldsymbol{d}^{\text{o}}. \quad (14)$$

于是, 我们由观测资料产生的组合系数, 根据式 (14)、(12) 得到分析场向量:

$$\boldsymbol{\psi}^{\text{a}} = \boldsymbol{U} (\boldsymbol{D}^{\text{T}} \boldsymbol{D})^{-1} \boldsymbol{D}^{\text{T}} \boldsymbol{d}^{\text{o}}. \quad (15)$$

上式就是 4DSVD 采用 EOF 方法获得支撑吸引子基向量时得到的分析场的表达式。

2.2 SVD 方法

模式大气状态向量样本组成的矩阵以及模拟观测大气状态向量组成的矩阵可以表示成如下形式:

$$\begin{cases} \mathbf{S} = (\boldsymbol{\psi}_i)_{m,N} \\ \mathbf{Z} = (\mathbf{H}(\boldsymbol{\psi}_i))_{p,N} \end{cases} \quad (16)$$

对 (16) 式的两个矩阵进行 SVD 分解:

$$\mathbf{SZ}^T = \mathbf{V}_L \begin{pmatrix} \mathbf{E} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{V}_R, \quad (17)$$

其中, $\mathbf{E}$ 是奇异值组成的对角矩阵, 可以表示成如下形式:

$$\mathbf{E} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r), \quad r \leq \min\{m, p\} \quad (18)$$

并且, 奇异值是从大到小排列的:

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r, \quad (19)$$

$\mathbf{V}_L$ 、 $\mathbf{V}_R$ 分别是左、右奇异向量构成的矩阵, 其维数分别为 $m \times r$ 、 $p \times r$ 。可以表示成列向量的形式:

$$\mathbf{V}_L = (\mathbf{V}_{L1}, \mathbf{V}_{L2}, \dots, \mathbf{V}_{Lr}), \quad (20)$$

$$\mathbf{V}_R = (\mathbf{V}_{R1}, \mathbf{V}_{R2}, \dots, \mathbf{V}_{Rr}). \quad (21)$$

与上述 EOF 方法相同, 选取前 $2s+1$ 个奇异向量作为支撑大气相空间吸引子的基向量。因此, 对于任何时刻的模式大气状态向量都可以用左奇异向量的线性组合表示, 任何时刻的观测大气状态都可以用右奇异向量的线性组合来表示:

$$\boldsymbol{\psi} = \sum_{k=1}^{2s+1} \mathbf{V}_{Lk} \mathbf{a}_k, \quad (22)$$

$$\mathbf{d} = \sum_{k=1}^{2s+1} \mathbf{V}_{Rk} \mathbf{b}_k, \quad (23)$$

其中, $\mathbf{a}$ 和 $\mathbf{b}$ 分别是模式大气相空间和观测大气相空间奇异向量的组合系数。由奇异向量的正交性, 第 k 对奇异向量的组合系数可以通过下式求得:

$$\mathbf{a}_k = \boldsymbol{\psi}^T \mathbf{V}_{Lk}, \quad (24)$$

$$\mathbf{b}_k = \mathbf{d}^T \mathbf{V}_{Rk}. \quad (25)$$

对 n 个样本, 我们可以求得每对奇异向量时间系数的相关系数:

$$\text{Cov}(\mathbf{a}_{kt}, \mathbf{b}_{k't}) = \mathbf{a}_k^T \mathbf{b}_{k'}' = \mathbf{V}_{Lk}^T \mathbf{SZ}^T \mathbf{V}_{Rk'} = \begin{cases} \rho_k (k = k') \\ 0 (k \neq k') \end{cases} \quad (26)$$

设左、右奇异向量时间系数之间有如下的关系:

$$\mathbf{a}_{kt} = \rho_k \mathbf{b}_{kt} + \mathbf{c}_k. \quad (27)$$

观测资料可以表示成观测相空间中奇异向量的线性组合:

$$\mathbf{d}^o \approx \mathbf{V}_R \mathbf{x}. \quad (28)$$

用最小二乘法求得观测场的线性组合系数 $\mathbf{x}$ 有如下

形式:

$$\mathbf{x} = \mathbf{V}_R^T \mathbf{d}^o. \quad (29)$$

所以, 根据式 (27) 中给出的左、右奇异向量组合系数的关系得到分析向量的线性组合系数:

$$\mathbf{y}_k = \rho_k \mathbf{x}_k + \mathbf{c}_k. \quad (30)$$

最后得到分析时刻的分析场:

$$\boldsymbol{\psi}^a = \mathbf{V}_L \mathbf{y} = \mathbf{V}_L (\boldsymbol{\Omega} \mathbf{V}_R^T \mathbf{d}^o + \mathbf{c}), \quad (31)$$

其中,

$$\boldsymbol{\Omega} = \text{diag}(\rho_1, \rho_2, \dots, \rho_{2s+1}), \quad (32)$$

上式就是在 SVD 方法下的 4DSVD 分析向量的表达式。

3 数值试验

3.1 资料说明

为了简化问题和方便分析, 我们采用真实的 NCEP 再分析资料作为大气的真实状态 (认为无误差), 并且在大气真实状态上叠加一定方差的 Gauss 白噪声作为观测资料和模式结果。

选取 1979 年 NCEP 再分析资料逐日 500 hPa 的温度场的标准化场作为大气状态的真值。选取一定的空间场, 将这个空间场的所有格点的真实资料的标准化场叠加上一方差的 Gauss 白噪声作为模式结果。

按照 4DSVD 理论公式, 必须将模式场投影到观测场, 作为模拟观测场。取观测场与模式场的变量相同和格点重合, 投影变得很简单 (观测算子是线性的)。相应格点的值作为投影后的模拟观测场。将所选定的观测场区域的 NCEP 再分析资料温度场的标准化场上叠加上一方差的 Gauss 白噪声作为观测资料。

3.2 试验设计

分析时刻为 1979 年 2 月 19 日, 选取 1979 年前 100 天的有误差的模式场作为样本。选取空间区域 500 hPa, ($15^\circ\text{N} \sim 75^\circ\text{N}$, $70^\circ\text{E} \sim 130^\circ\text{E}$), $5^\circ \times 5^\circ$ 作为模式模拟区域。共有 $13 \times 13 = 169$ 个格点, 在所有的试验中模式格点均匀分布且保持不变, 根据观测点位置和多少分三个方案 P1、P2、P3。在方案 P1 中的观测点和模式格点完全重合, 格点一共有 169 个。方案 P2 中, 观测点均匀分布在模式区域内部, 每两个格点一个观测点, 共有 $7 \times 7 = 49$ 个观测点。方案 P3 中观测点分布在模式空间区域的四周, 共有 $13 + 13 + 11 + 11 = 48$ 个观测点。不同方案的观测点与模式格点位置关系如图 1 所示。

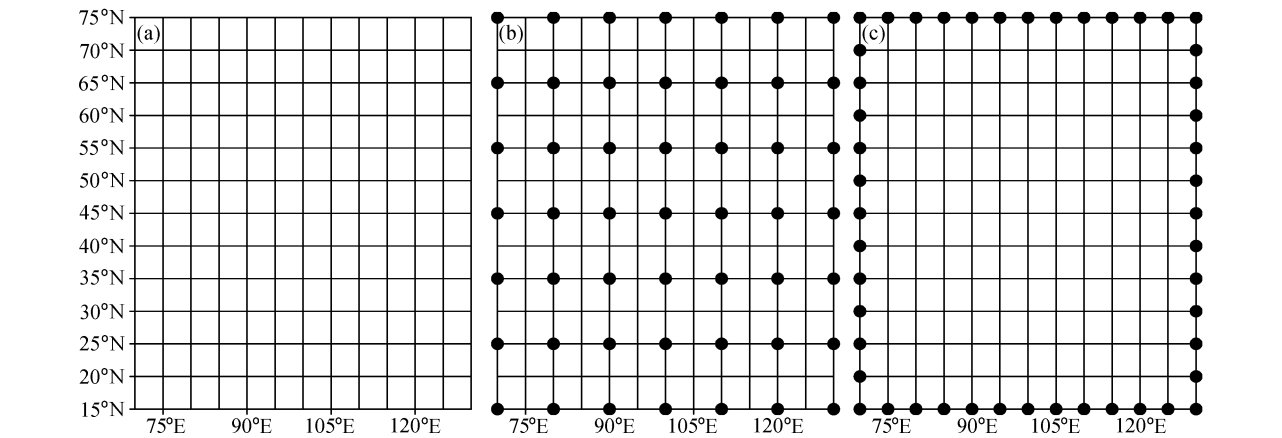


图1 观测与模式格点位置关系示意图：(a) P1；(b) P2；(c) P3。网格交点：模式格点；黑圆点：观测点

Fig. 1 The relationship between the observations and the model grids: (a) P1; (b) P2; (c) P3. The grids are model grids and the dots represent the observation stations

表 2 试验说明
Table 2 Experiments illumination

试验组别	试验名称	模式误差（标准差）	观测误差（标准差）
E1	E1a	/	0.01
	E1b	/	0.1
	E1c	/	0.5
	E1d	/	1.0
E2	E2a	0.01	/
	E2b	0.1	/
	E2c	0.5	/
	E2d	1.0	/
E3	E3a	0.1	0.1
	E3b	0.5	0.5
	E3c	1.0	1.0
	E3d	1.0	0.5
	E3e	0.5	1.0
E4	E4a	/	/

为了考虑不同模式误差和观测误差大小对分析场的影响，每个方案中又分 E1、E2、E3 和 E4 四组共 14 个试验。具体的试验说明如表 2 所示。其中 E1 组试验模式没有误差，只有观测误差，根据观测误差大小不同有 E1a、E1b、E1c 和 E1d 四个试验。E2 组试验观测没有误差，而模式有误差，也有 E2a、E2b、E2c 和 E2d 四个试验。E3 组试验模式和观测都有误差，共有五个试验：E3a、E3b、E3c、E3d 和 E3e。E4 组试验只有一个试验 E4a，模式和观测都没有误差，作为控制试验。为了比较 EOF 和 SVD 两种获得支撑大气吸引子的方法，我们分别用这两种方法作了上述试验。

4 试验结果与分析

分别计算分析场的空间相关系数和均方根误差。P1 方案下试验结果如图 2 所示。由于观测点和模式格点完全重合，因此 EOF 和 SVD 两种获得支撑大气吸引子基向量方法的分析场完全一致（图中 EOF 方法和 SVD 方法的结果曲线重合）。从图 2a、b 中还可以看出，随着基向量的个数的增加，分析误差先减小后略有增加或进入平稳阶段，空间相关系数先增加后进入平稳阶段或略有减小。这是由于基向量个数小于吸引子维数时，不能够支撑起整个吸引子相空间，每当增加一个新的基向量就增加了新的信息；当基向量多于吸引子相空间维数，所选取的基向量完全支撑起吸引子相空间，后面的向量对支撑吸引子是无效的，所以分析误差不再随着新基向量的引入而有所减小，甚至引入了新的误差。比较图 2a、b 中不同的曲线间的关系，我们还可以看出观测误差越大，分析误差越大，分析场与真实场的空间相关系数越小，这是比较容易理解的，观测的误差越小观测资料提供的观测信息越多，分析场越好。在图 2c、d 中，随着基向量个数的增多，分析误差逐渐减小，分析场与真实场的空间相关系数逐渐增大，当基向量个数达到一定数目（在本实验中 15 个），相关系数不再有明显改善。模式样本误差大小关系是：E2a<E2b<E2c<E2d，分析误差关系是 E2a<E2b<E2c<E2d，这说明模式误差越大，分析误差越大，分析场的空间相关系数越小，也就是当模式误差增大，影响了获取的基

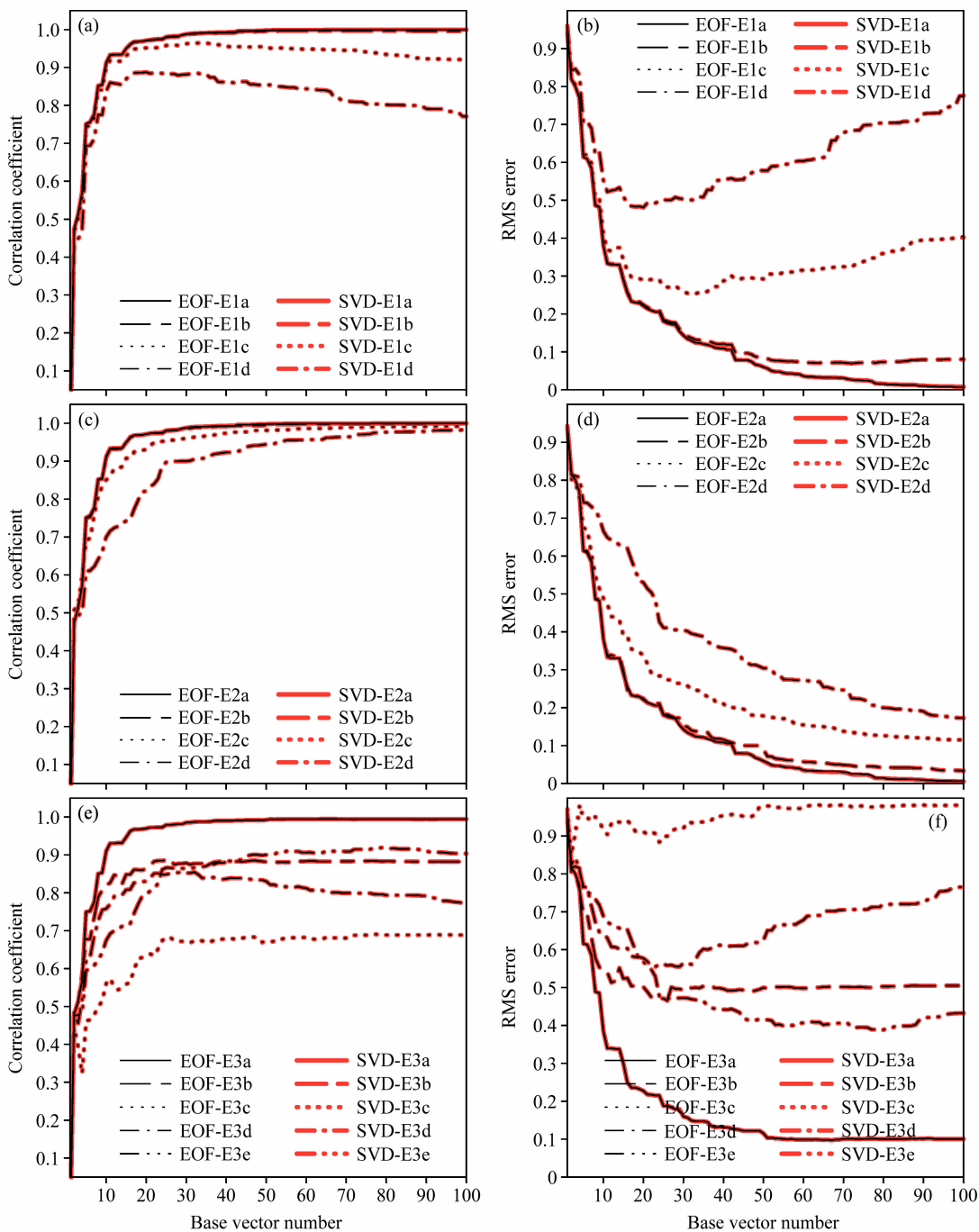


图2 P1 方案所有试验分析场与真实场的空间相关系数 (a、c、e) 和分析误差 (b、d、f) 与基向量个数的关系: (a、b) E1; (c、d) E2; (e、f) E3

Fig. 2 The space correlation coefficients (a, c, e) between the analysis fields and reference fields and root-mean-square error (b, d, f) of the analysis for project P1: (a, b) E1; (c, d) E2; (e, f) E3

向量质量。比较 E1 组和 E2 组试验的结果发现, 当分析场达到稳定状态时, 观测误差对分析误差的影响较模式误差大, 可能是模式误差在 EOF 和 SVD 分解过程中部分被剔除了。从图 2e、f 中可以看到, 在模式误差和观测误差都存在的情况下, 得到

了较好的分析场, 此时的空间系数都可以超过 0.6, 而分析误差也减小了许多, 但比较 E1、E2 组试验分析误差增加了很多, 可能观测误差和模式误差在分析过程中发生了相互作用。还可以得到在 P1 方案下 EOF 方法和 SVD 方法都能在很大程度上从模

式样本状态向量中获得较好的支撑大气吸引子的基向量。

图 3 是 P2 方案下 E1、E2 和 E3 三组试验分析场的分析误差和与真实场的空间相关系数图。从图 3 可以看出，不论是 EOF 方法还是 SVD 方法随着基向量数量的增多，分析误差在开始迅速减小，空间相关系数迅速增大，当基向量个数增加到 15 个

左右时分析误差和空间相关系数达到一个稳定的状态，与 P1 方案下的结果特征相近。并且随着观测误差和模式误差的增加，分析误差增加，空间相关系数减小。比较 P2 方案下和 P1 方案下的试验发现，P2 方案的试验分析误差明显大于同等条件下 P1 方案试验的分析误差，空间相关系数也有明显的减少，这可能是由于观测点少于模式格点（观测

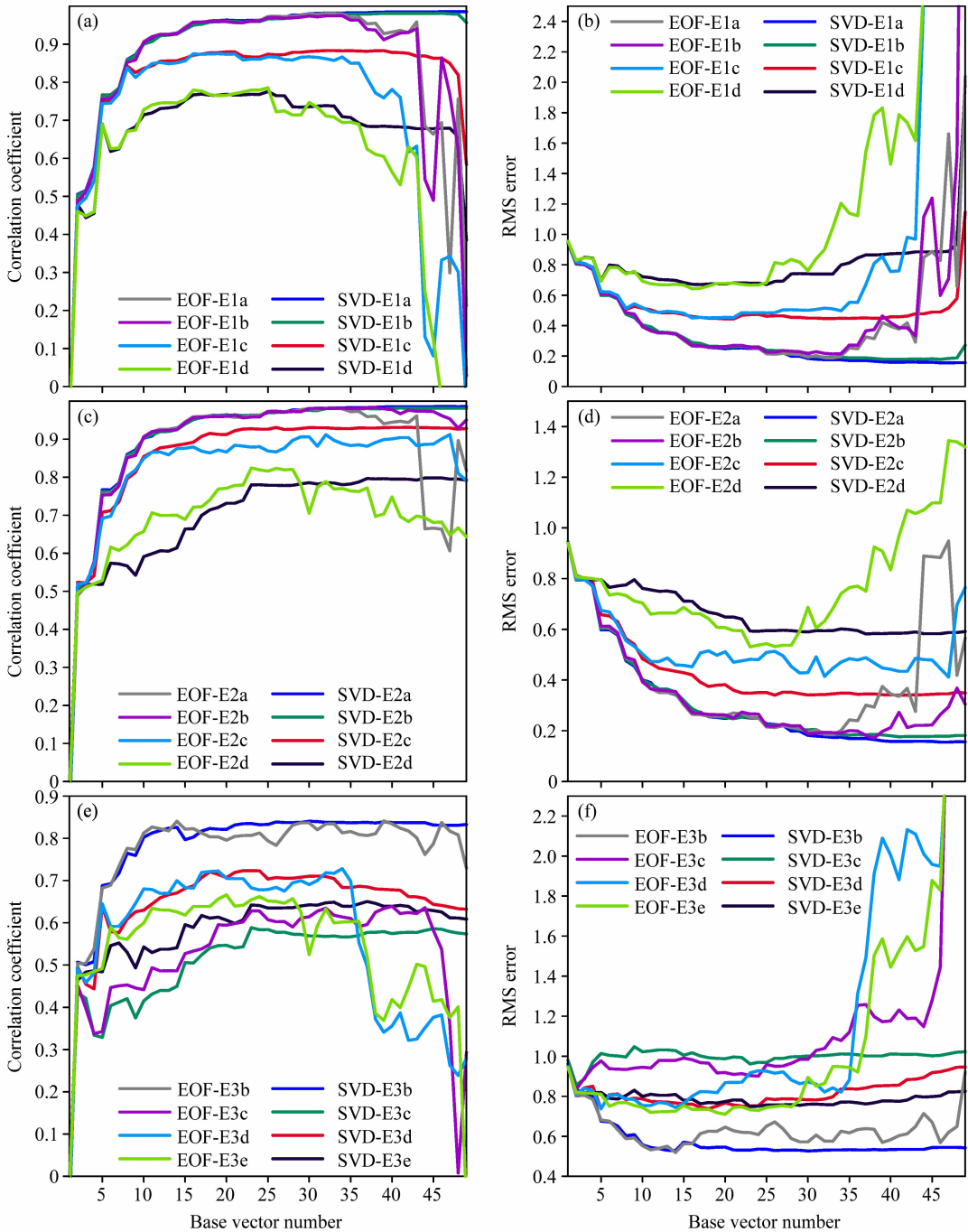


图 3 同图 2，但为 P2 方案
Fig. 3 Same as Fig. 2, but for project P2

点间的距离是模式格距的两倍),造成了观测场的高频部分信息缺失引起的误差,由观测代表性误差引起了分析误差。从这个方案下的试验中可以发现,EOF方法得到的分析场和用SVD方法得到的分析场基本相同。但是当有较多的基向量个数时,EOF方法的分析场开始变得不稳定,从这一点上说SVD方法比EOF方法更能从模式样本中提取吸

引子信息,得到支撑大气吸引子相空间的更好的基向量。

图4是P3方案下E1、E2和E3三组试验的分析场的分析误差和与真实场的空间相关系数图。从图4中可以发现,对于EOF方法,分析误差先减小后增大,然后就出现了不稳定,空间相关系数也是先增大后突然减小,也迅速出现了不稳定,而且不

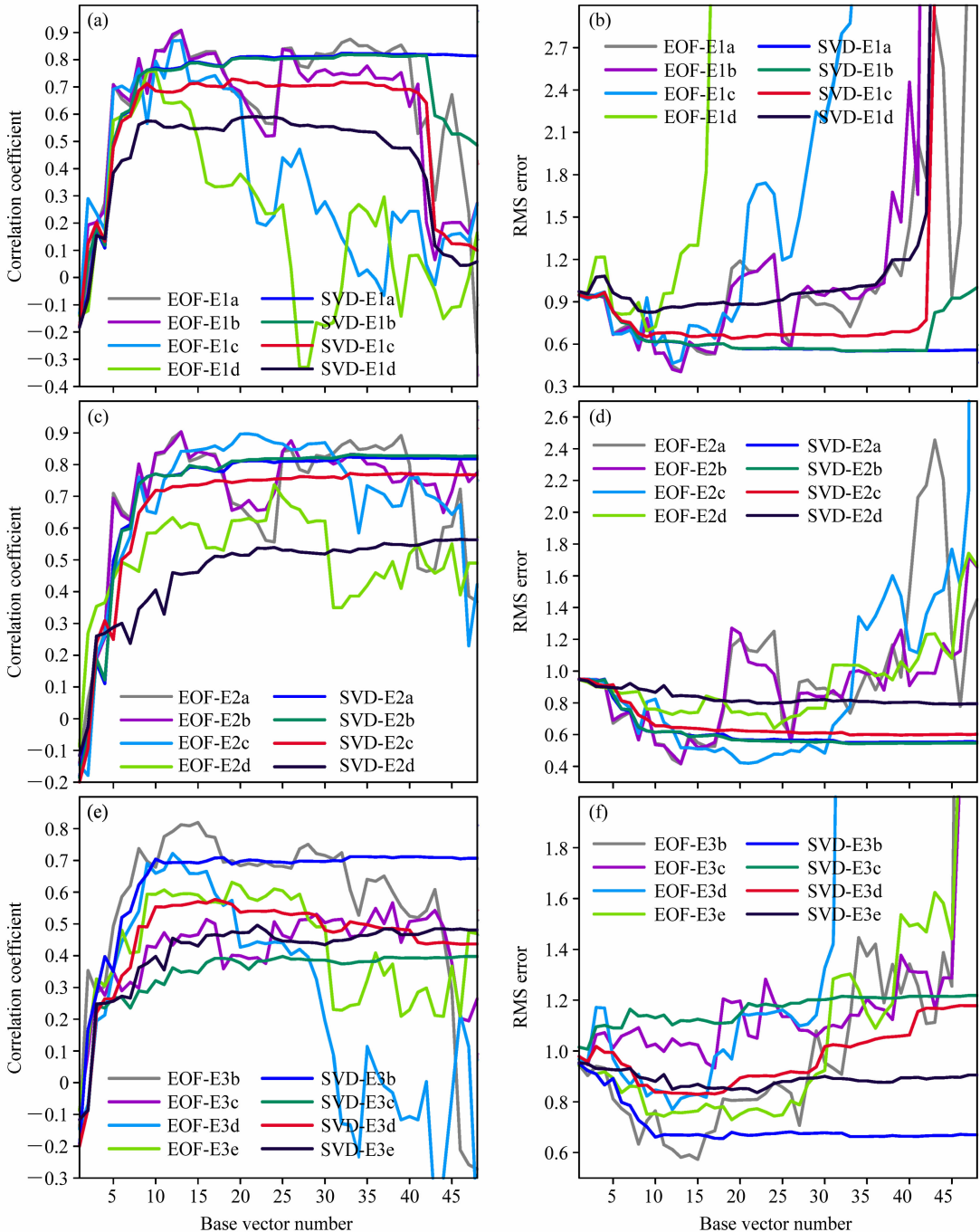


图4 同图2, 但为P3方案

Fig. 4 Same as Fig. 2, but for project P3

稳定出现时的基向量个数随着观测误差和模式误差的增大而减小, 这就给在观测误差和模式误差较大时基向量个数的确定带来了较大的困难; 说明了在 P3 方案这种情况下, EOF 方法从模式样本中提取基向量的效果不好, 不能够很好地把有效信息和噪音分离, 不能够得到较好的支撑大气吸引子相空间的基向量。比较 P3 方案下 EOF 方法的结果和 P1、P2 方案下的结果, 可以推断观测的代表性误差对基于 EOF 方法的 4DSVD 影响比较大。在基向量个数超过观测点个数时, 分析场完全失去了意义(图略), 这是由于此时用观测资料求解组合系数, 方程 (13) 变成了高度欠定方程, 方程解不是唯一的。而对于 SVD 方法, 分析误差在前几个基向量迅速减小, 空间相关系数迅速增大, 当基向量个数达到 15 个左右时, 分析误差达到了并稳定在较小的水平, 同时空间相关系数达到了较大的水平, 随后分析场进入稳定状态, 即随着基向量个数的增大, 分析场不再有明显的改进。这与 P1、P2 方案的分析场特征很相似, 观测的代表性不会影响基于 SVD 方法的 4DSVD 的稳定性。随着模式误差和观测误差的增大, 分析误差也越大, 这与 P1、P2 方案下试验的结论相同。SVD 方法在基向量个数小于观测点数目时, 分析误差和空间相关系数比 EOF 方法稳定, 从这种意义上说, 获得支撑大气吸引子基向量用 SVD 方法比用 EOF 方法得到的分析场稳定且效果相当, 当基向量个数达到一定的数目时, 分析场对基向量个数不敏感, 即 SVD 方法能更好地提取并分离支撑大气吸引子的基向量, 这是 SVD 方法优于 EOF 方法的重要方面, 这一点在本文所有的实验中都有所体现。

特别值得注意的是, 在上面的 P2、P3 方案的试验结果中, 随着基向量的增多, 开始分析场迅速得到改善, 但是当基向量达到 15 个左右的时候, 再增加基向量时分析场并没有显著的改进, 这说明对于此试验情况下选取 15 个左右基向量时便可以得到较好的分析场。为了更好地理解这个问题, 以 P3 方案下 SVD 方法的 E3b 试验的左、右奇异向量时间系数之间的相关系数(图 5)为例进行初步解释。从图 5 中可以看出, 随着基向量阶数的增大, 相关系数开始时减小不是很大, 但到第 15 阶基向量时, 相关系数迅速下降。这可能就是当基向量达到一定数目后(完全支撑起吸引子相空间), 再增

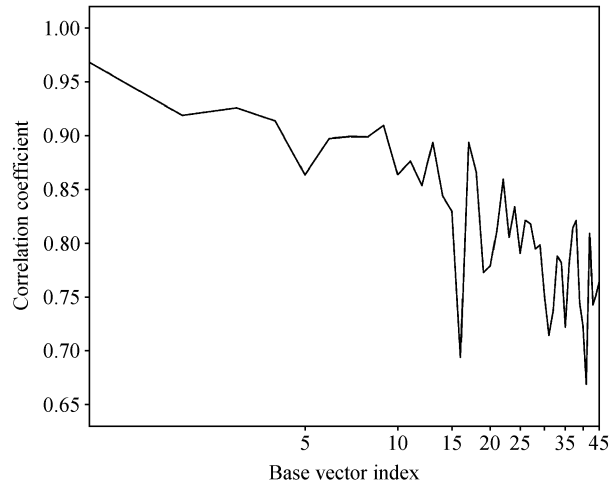


图 5 P3 方案 SVD 方法的 E3b 试验左、右奇异向量时间系数之间的相关系数图

Fig. 5 The correlation coefficients between time coefficients of model base vectors and the observation base vectors in experiment E3b of project P3 using SVD method

加基向量分析结构不会有明显改善的原因。

为了更加深入地理解 4DSVD 方法, 我们分析了 4DSVD 的误差来源。在第 2 节理论公式的推导过程中, 本文认为有五种误差引起了 4DSVD 的分析误差。这五种误差分别是截断误差 ϵ_T 、模式误差 ϵ_M 、观测误差 ϵ_O 、观测代表性误差 ϵ_R 和取样误差 ϵ_S 。

ϵ_T 是指所取的基向量个数不能够完全支撑起模式大气相空间或者有多余的基向量; ϵ_M 是指每个样本并没有在大气的吸引子上, 这是模式固有的缺陷; ϵ_O 是指观测资料的误差; ϵ_R 是指观测点没有模式格点多并且分布不均匀所引起的; ϵ_S 是指我们的取样并不完备, 不能够取遍整个模式大气相空间并获得支撑相空间吸引子的基向量。这五种误差引起的分析误差分别设为 ϵ_{Ta} 、 ϵ_{Ma} 、 ϵ_{Oa} 、 ϵ_{Ra} 、 ϵ_{Sa} 。因此, 总的分析误差可以表示为

$$\epsilon_a = \epsilon_{Ta} + \epsilon_{Ma} + \epsilon_{Oa} + \epsilon_{Ra} + \epsilon_{Sa}. \quad (33)$$

在 P1 方案的试验中, 所有的试验都是在观测点和模式格点完全重合的条件下进行的, 所以就没有了 ϵ_R 。E4a 是控制试验, 即没有观测误差和模式误差。我们假设 100 个样本完全可以取遍模式大气吸引子相空间, 即 ϵ_S 等于零。因此, 根据公式 (35), E4a 的分析误差完全是由截断误差引起的(图 6a、b)。从图 6a、b 中可以看出, 在这种情况下, 截断误差引起的分析误差是随着基向量的增多

而减小的。

在 P2、P3 方案的试验中, 观测点数目较少, 并且分布在模式格点不同的位置上。所以, P2、P3 方案下的 E4a 试验没有观测误差和模式误差, 其分析误差完全由截断误差 ϵ_T 和观测代表性误差 ϵ_R 引起的。我们用 P2、P3 方案的试验 E4a 的分析误差

减去 P1 方案下试验 E4a 的分析误差, 便可以得到观测代表性误差引起的分析误差 ϵ_{Ra} (图 6c、d)。从图 6c、d 中可以看出, 不论是 EOF 方法还是 SVD 方法, 观测代表性误差引起的分析误差都是随着基向量个数增多而增大, 也就是说在观测点和模式格点的位置固定时所用的基向量个数越多, 观测代表

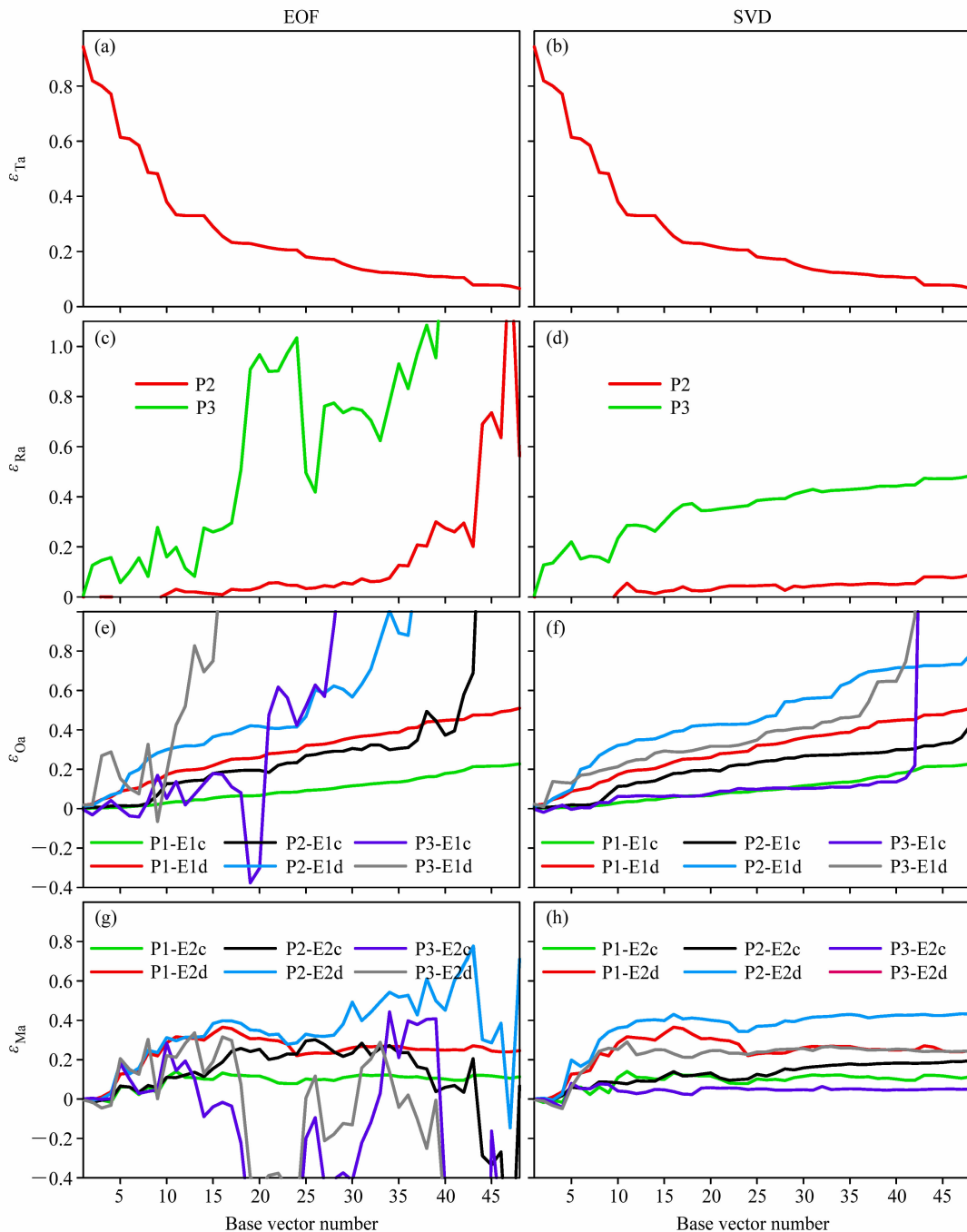


图 6 截断误差、观测代表性误差、观测误差和模式误差引起的分析误差: (a、b) ϵ_{Ta} ; (c、d) ϵ_{Ra} ; (e、f) ϵ_{Oa} ; (g、h) ϵ_{Ma}

Fig. 6 The analysis errors generated by truncation error, representation error, observation error and model error: (a, b) ϵ_{Ta} ; (c, d) ϵ_{Ra} ; (e, f) ϵ_{Oa} ; (g, h) ϵ_{Ma}

性误差引起的分析误差越大, 但 EOF 方法的 ϵ_{Rn} 随着基向量个数增多呈指数增长, 而 SVD 方法的 ϵ_{Rn} 呈比例的线性增长并且增长速率很慢。EOF 方法的 ϵ_{Rn} 在同样条件下比 SVD 方法的 ϵ_{Rn} 大很多。这从一个侧面说明 SVD 方法作为获取支撑大气吸引子的方法比 EOF 方法优越。从图 6c、d 中还可以看出, 观测代表性误差引起的分析误差 P3 方案比 P2 方案大许多, 这是比较容易理解的, 因为 P3 方案下观测点仅仅分布在分析区域的四周, 观测代表性误差较大。

在每个方案的试验中, E1 组试验是只有观测误差的情况, 因此, 用 E1 的分析误差减去试验 E4a 的分析误差便可以得到观测误差引起的分析误差 ϵ_{Oa} (图 6e、f)。从图 6e、f 中可以看出, 不论是 EOF 方法还是 SVD 方法观测误差引起的分析误差随基向量个数的增加而增大, 并且增长的速率 $P2 > P3 > P1$ 。在相同的基向量个数, 观测误差越大, 分析误差也越大。EOF 方法下, P2、P3 方案的分析误差随基向量的增多出现明显的波动, 尤其是观测误差较大时即使选取的基向量较少也出现了非常大的分析误差, 比采用 SVD 方法的分析误差大得多。而 SVD 方法作为获得支撑大气吸引子的方法时, 观测误差引起的分析误差在所有方案和试验中都比较稳定地保持在一个较小的误差范围内。并且, 我们可以看到在 P1、P2、P3 三种方案下, 相同的观测误差产生的分析误差不相同, 这说明观测误差与观测代表性误差产生了非线性作用。

由于在每个方案的试验中 E2 组试验只有模式误差, 因此用 E2 组试验的分析误差减去 E4a 的分析误差可以得到由模式误差引起的分析误差 ϵ_{Ma} (图 6g、h)。从图 6g、h 中可以看出, 不论是用 EOF 方法还是 SVD 方法作为获取大气吸引子基向量的方法, 分析误差随着模式误差的增大而增大。用 EOF 方法作为获得支撑大气吸引子基向量的方法模式误差引起的分析误差 ϵ_{Ma} 随着基向量个数的增多而一直线性增大, 并且出现了负值, 这可能是由于模式误差与观测代表性误差产生了非线性效果使分析误差减小造成的。SVD 方法下当基向量个数达到一定的数目 (一般在 15 个左右) 时 ϵ_{Ma} 趋于饱和, 而不是无限增长的。这是 SVD 方法的一个优点, 不会因为基向量个数选取的过大而产生更大的分析误差。比较 EOF 方法下的由加入模式误差后分析误

差的变化和加入观测误差后分析误差的变化, 我们可以看到, 加入模式误差可以在一定程度上抑制分析误差由于观测代表性误差引起的不稳定。并且在 P1、P2、P3 方案下相同的模式误差所产生的分析误差不同, 模式误差与观测代表性误差可能也发生了非线性相互作用。

除了上述试验之外, 我们还选取了不同的样本作为模式统计样本用于获得支撑大气吸引子的基向量 (图略), 发现得到的分析效果都基本相同。这说明了分析场对模式样本的选取方法依赖性较小, 并说明了一定数量的样本得到的支撑大气吸引子相空间的基向量具有稳定性。

误差分析表明, 由观测代表性误差、观测误差和模式误差引起的分析误差用 SVD 方法比用 EOF 方法小, 并且稳定。因此, 我们建议采用 SVD 方法获得支撑大气吸引子基向量。

5 结论与讨论

本文采用 EOF 和 SVD 两种方法获取支撑大气吸引子的基向量, 推导出 4DSVD 分析场的表达式。用简单的数值试验比较了基于 EOF 与 SVD 的 4DSVD 的分析场, 并在此基础上对分析误差的来源做了初步分析, 得到如下一些初步结论。

不论是用 EOF 方法还是用 SVD 方法获得支撑大气吸引子的基向量, 4DSVD 都得到了较好的分析场, 分析场的好坏与基向量的个数有密切关系。SVD 方法得到的分析场较 EOF 方法稳定, 分析误差较小。尤其是在观测点个数较少, 并且分布不均的情况下, 基于 SVD 方法的 4DSVD 得到的分析场比 EOF 得到的分析场稳定。这说明 SVD 方法更适于作为获取支撑大气吸引子基向量的方法。

一般来说, 分析误差随着观测误差和模式误差的增大而增大。分析误差的大小与观测点分布及数目有关。对于基向量个数有一个最优的值。

截断误差、样本误差、观测代表性误差、观测误差和模式误差是 4DSVD 分析误差的主要来源。截断误差产生的分析误差随着基向量个数的增多而减小; 观测代表性误差、观测误差和模式误差引起的分析误差随着基向量个数的增多而增大。这从一个侧面说明, 不是基向量个数越多越好。由观测代表性误差、观测误差和模式误差引起的分析误差用 SVD 方法比用 EOF 方法小, 并且稳定。采用 SVD

方法获得支撑大气吸引子基向量比用 EOF 方法好。4DSVD 的分析误差对观测误差是比较敏感的, 观测代表性误差与观测误差和模式误差产生了非线性相互作用。这说明观测点的位置和空间分布的重要性。

本文的结果仍是初步的, 一些做法存在一定的局限性。例如, 采用 NCEP 资料作为真实场, 模式误差是白噪声, 试验中观测点恰好在模式格点上, 并且观测变量和模式变量相同, 观测算子是线性算子等等。其中还有许多问题需要进一步的研究, 如 4DSVD 方法在真实模式中分析场如何, 分析场是否在其他情况下也对取样方法不敏感, 当观测算子为非线性算子的时候分析场的效果如何, 各种误差是如何影响分析误差的, 其得到的分析场作为初值对预报的影响以及基向量个数选取的标准问题等等。这些问题都有待于进一步试验和研究。

参考文献 (References)

- [1] Lorenz E N. Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, 1969, **26** (4): 636~646
- [2] Daley R. *Atmospheric Data Analysis*. Cambridge: Cambridge University Press, 1991
- [3] Gandin L S. Objective analysis of meteorological fields, Gidrometrologicheskoe Izdatelstvo, Leningrad. 1963. English translation by Israeli Program for Scientific Translations, Jerusalem, 1965
- [4] Schlatter T. Some experiments with a multivariate statistical objective analysis scheme. *Mon. Wea. Rev.*, 1975, **103** (3): 246~257
- [5] Lorenc A C. Analysis methods for numerical weather prediction. *Quart. J. Meteor. Soc.*, 1986, **112** (474): 1177~1194
- [6] Lewis J, Derber J. The use of adjoint equations to solve a variational adjustment problem with advective constraint. *Tellus*, 1985, **37A**: 309~327
- [7] Derber J, Rosati A. A global oceanic data assimilation system. *J. Phys. Oceanogr.*, 1989, **19**: 1333~1347
- [8] Courtier P, Talagrand O. Variational assimilation of meteorological observations with the direct and adjoint shallow water equations. *Tellus*, 1990, **42A**: 531~549
- [9] Bouttier F, Rabier F. The operational implementation of 4D-Var. *ECMWF Newsletter*, 1997, (78): 2~5
- [10] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 1994, **99** (C5): 10143~10162
- [11] 许小永, 刘黎平, 郑国光. 集合卡尔曼滤波同化多普勒雷达资料的数值试验. *大气科学*, 2006, **30** (4): 712~728
Xu Xiaoyong, Liu Liping, Zheng Guoguang. Numerical experiment of assimilation of Doppler radar data with an ensemble Kalman filter. *Chinese Journal of Atmospheric Sciences* (in Chinese), 2006, **30** (4): 712~728
- [12] 朱江, 汪萍. 集合卡尔曼平滑和集合卡尔曼滤波在污染源反演中的应用. *大气科学*, 2006, **30** (5): 871~882
Zhu Jiang, Wang Ping. Ensemble Kalman smoother and ensemble Kalman filter approaches to the joint air quality state and emission estimation problem. *Chinese Journal of Atmospheric Sciences* (in Chinese), 2006, **30** (5): 871~882
- [13] Courtier P. Dual formulation of four-dimensional variational assimilation. *Quart. J. Roy. Meteor. Soc.*, 1997, **123**: 2449~2461
- [14] Zhou X. Tangent linear and adjoint of "on-off" processes and their feasibility for use in 4-dimensional variational data assimilation. *Tellus*, 1997, **49A**: 3~31
- [15] Xu Q. Comments on "Tangent linear and adjoint of "on-off" processes and their feasibility for use in 4-dimensional variational data assimilation". *Tellus*, 1998, **50A**: 653~656
- [16] Evensen G. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, 1998, **126** (6): 1719~1724
- [17] Evensen G. The ensemble Kalman filter; Theoretical formulation and practical implementation. *Ocean Dynamics*, 2003, **53**: 343~367
- [18] 李建平, 丑纪范. 大气吸引子的存在性. *中国科学 (D 辑)*, 1997, **27** (1): 87~96
Li Jianping, Chou Jifan. Existence of atmosphere attractor. *Science in China (Ser D)*, 1998, **40** (2): 215~224
- [19] Klaus Fraedrich. Estimating the dimensions of weather and climate attractors. *J. Atmos. Sci.*, 1986, **43** (5): 419~432
- [20] Qiu C, Chou J. Four-dimensional data assimilation method based on SVD: Theoretical aspect. *Theor. Appl. Climatol.*, 2006, **83**: 51~57
- [21] Zhang B L, Chou J. The application of empirical orthogonal functions to numerical simulation of climate. *Science in China (Ser. B)*, 1992, **35** (1): 92~101