

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

朱 江¹ 汪 萍^{1, 2}

1 中国科学院大气物理研究所国际气候与环境科学中心, 北京 100029

2 中国科学院研究生院, 北京 100049

摘 要 此文目的是讨论污染源反演问题的统计方法。基于 Bayes 估计理论, 该文将资料同化中的集合平滑、集合卡尔曼平滑和集合卡尔曼滤波应用在污染源反演问题中。在详细给出污染源反演的集合平滑、集合卡尔曼平滑和集合卡尔曼滤波的严格数学表达后, 用一个简单的模型演示了集合卡尔曼平滑和集合卡尔曼滤波在污染源反演中的可行性, 并且通过对理想试验结果比较了集合卡尔曼平滑和集合卡尔曼滤波方法在反演污染源排放的效果, 讨论了观测误差和污染源先验误差估计对反演结果的影响。试验结果表明在观测间隔小和观测误差小的情况下, 集合卡尔曼滤波和集合卡尔曼平滑都可以有效地反演出随时间变化的污染源排放。当观测误差增大时, 集合卡尔曼滤波和集合卡尔曼平滑的反演效果都有一定降低, 但是反演误差的增加少于观测误差的增加, 同时集合卡尔曼平滑(Ensemble Kalman smoother, 简称 EnKS)对观测误差比集合卡尔曼滤波(Ensemble Kalman filter, 简称 EnKF)更为敏感。当观测时间间隔较大时, EnKF 不能对没有观测时的污染源排放进行估计, 仅能对有观测时的污染源排放进行较好的反演。而 EnKS 可以利用观测对观测时刻前的污染源排放进行反演, 因此其效果明显好于 EnKF, 并且在观测时间间隔较大的情况下依然可以较好地反演出污染源排放。试验结果还显示污染源排放的先验误差估计对反演的结果有较大影响。

关键词 集合卡尔曼平滑 集合卡尔曼滤波 空气质量 污染源 反演模拟 资料同化

文章编号 1006-9895(2006)05-0871-12

中图分类号 X16

文献标识码 A

Ensemble Kalman Smoother and Ensemble Kalman Filter Approaches to the Joint Air Quality State and Emission Estimation Problem

ZHU Jiang¹ and WANG Ping^{1, 2}

1 *International Center for Climate and Environment Science, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029*

2 *Graduate University of the Chinese Academy of Sciences, Beijing 100049*

Abstract The purpose of this paper is to present some ensemble based statistical estimation methods for inverting modeling of pollution emissions. Previous studies using advanced sequential method to the joint air quality state and emission estimation problem focus on the ensemble Kalman filter (EnKF) or the discrete Kalman filter (DKF). However at each assimilation time, EnKF and DKF cannot update the emissions at the previous time, though their information is contained in the present observations. This is a major drawback, especially in the case of the time-variant emissions. Therefore it is necessary to investigate other methods that can overcome this drawback. Based on Bayes theorem, this study presents the detailed mathematical formulations of the ensemble smoother (ES), the en-

收稿日期 2006-04-27, 2006-06-09 收修定稿

资助项目 国家自然科学基金杰出青年基金 40225015

作者简介 朱江, 男, 1963 年出生, 研究员, 主要从事资料同化方面的研究。E-mail: jzhu@mail.iap.ac.cn

semble Kalman smoother (EnKS) and the ensemble Kalman filter for the joint air quality state and emission estimation problem. Due to the adoption of ES and EnKS, the emission estimation at previous time can be updated by assimilating observation at the later time. A simple model is used to demonstrate the feasibility of the methods. The impacts of observational error and the prior error of the emission on estimation results are also discussed based on the experiments using the simple model. The results show that EnKS can estimate the time-variant emission well at every time step when the observations are not available at every time step, while the EnKF can only estimate the emission well at the observation time. When observations are available at every time step, EnKF and EnKS perform similarly. The larger observational errors can affect the estimation results of emission, but not very sensitively. It is also shown that overestimation or underestimation of the prior emissions uncertainty can bring larger estimation errors of emissions.

Key words ensemble smoother, ensemble Kalman filter, air quality, emission estimation, inverting modeling, data assimilation

1 引言

获取准确的大气污染源排放信息对大气污染预报和控制具有重要的意义。因此,我国各地区都进行了不同程度和力度的大气污染源调查和监测。由于污染源分布和变化的复杂性使得准确获取其信息十分昂贵和困难,尤其是对于大面积、自然污染源则更加困难。一类解决办法是开展对各类大气污染源的研究,建立污染排放模式。例如,对于人为污染源中的道路移动源可以根据交通流量、不同排放的车辆比例等建立污染排放模型来估计污染源强的分布和变化^[1];再比如对于区域的大尺度排放可以根据诸如社会、经济发展状况、人口密度、人均能源消耗量、能源结构等因素建立模型进行定量估计^[2]。另外一类重要的办法是利用污染观测和污染模式反演污染源的强度^[3]。这类办法主要可以分为三种。第一种是发展比较早的受体模型方法^[4],这种方法假定观测点处的各污染成分的浓度线性可加地来自于各种污染源的共同作用。由于实际大气化学过程的非线性,这一假设会对污染源的估计带来影响。近些年来较受关注的另外两种方法是四维变分资料同化方法和集合卡尔曼滤波方法,这两种办法都是利用污染模式中揭示的大气污染物的传输、扩散和化学转化规律建立污染源和污染观测之间的关系,从而利用最优化或者次优化方法从污染观测反演污染源。比如,Elbern 等^[5]利用四维变分资料同化方法研究了利用臭氧观测反演可以生成臭氧的其他化学物质的排放源强,他们的结果表明四维变分方法可以直接分析 NO 的排放率,但是对于 VOC,需要采用正则化方法才能够获得可靠的

VOC 排放。还有其他利用四维变分方法反演污染源排放的工作^[6, 7]。近年来,卡尔曼滤波和集合卡尔曼滤波也应用在污染源的“反演”。例如 Prinn 等^[8]和 Gilliland 等^[9]研究了离散卡尔曼滤波在污染源反演的应用。Van Loon 等^[10]用一个臭氧模式,研究了利用集合卡尔曼滤波方法来同化观测资料,他们将污染源原排放清单的订正因子作为显式的估计变量,从而利用观测对 NO_x、SO_x、VOC 和 CO 的原排放源进行订正。他们的结论是利用集合卡尔曼滤波可以对污染源排放进行订正,从而改进对臭氧的模拟。Heemink 和 Segers^[11]也讨论了集合卡尔曼滤波方法在污染源反演中的应用,他们的工作集中在研究更加快速的算法,使得卡尔曼滤波方法能够在高维模式中有效使用。由于离散卡尔曼滤波和集合卡尔曼滤波是一种顺序同化方法,尽管同化时刻的观测资料中包含以前时刻的污染源排放的信息,但同化时刻的观测资料不能够用来对以前时刻的污染源排放反演结果进行修正。这对随时间变化的污染源排放反演问题是不利的。因此,有必要探讨可以克服这一缺点的统计方法。

我国在污染源强反演方面也有很多研究工作。例如,陈军明等^[12]使用遗传算法和多点源扩散模式,从控制点浓度成功地反演点源排放清单。苏芳等^[13]利用逆向轨迹反演模式对北京地区甲烷源强进行了估计。其他相关工作还包括在自然控制论的基础上探讨如何利用大气污染的数值预报结果和模式进行大气污染最优控制设计的理论框架,提出以控制污染源排放量为控制手段的大气污染最优控制问题,并导出基于伴随模式的最优化方法^[14~16]。

本文工作受到 Evensen^[17]的启发。Evensen^[17]

讨论了状态变量和参数联合估计问题的统计方法, 并且从 Bayes 定理出发, 详细导出状态变量和参数联合估计问题的集合平滑、集合卡尔曼平滑和集合卡尔曼滤波等方法的严格数学表达。本文也是基于 Bayes 定理, 给出状态变量估计和污染源反演的一些统计方法, 如集合平滑、集合卡尔曼平滑和集合卡尔曼滤波的严格数学表达。在文献[17]中, 要估计的模式参数被定义为不随时间变化的确定性常数。本文新的地方在于要反演的污染源排放被定义为随时间变化的随机变量。

由于集合平滑可以利用同化时刻的观测反演以前时刻的污染源排放。作为顺序同化方法, 集合卡尔曼平滑可以利用同化时刻的观测对以前时刻的污染源排放反演结果进行修正。本文引入的方法扩展了现有的污染源反演方法。本文还利用简单模型的理想试验, 比较集合卡尔曼平滑和集合卡尔曼滤波在污染源排放反演的效果以及观测误差、先验误差估计的不确定性对反演结果的影响等基本的问题。

2 基本理论

从污染物浓度观测估计污染源排放的集合方法的基本想法是对已知、但不准确的污染源排放清单进行一组扰动, 然后将这组扰动逐一带入模式, 因此得到一组污染物浓度模式输出结果。有了这输入和输出样本, 可以统计出在观测点处模式污染物浓度变化和污染源排放变化的关系。有了这个关系就可以根据污染物浓度观测对已知的、不准确的污染源排放清单进行改进。本节的其他部分和第3、4节都是这样一个简单想法的严格数学描述, 以及导出的不同算法。对这些数学细节不感兴趣的读者可以直接跳到第5节的数值试验部分。

2.1 模式和观测

一般情况, 大气污染浓度和排放源可以写成下面的形式,

$$\begin{cases} \mathbf{C}(x, t_{k+1}) = M_k \mathbf{C}(x, t_k) + \mathbf{Q}(x, t_{k+1}) + \mathbf{q}(x, t_{k+1}), \\ \mathbf{Q}(x, t_{k+1}) = \mathbf{Q}^b(x, t_{k+1}) + \mathbf{Q}'(x, t_{k+1}), \end{cases} \quad (1)$$

其中, \mathbf{C} 是污染物浓度的向量(其分量表示不同地点和不同污染物种的浓度); x 表示空间位置; t_k 表示时间; \mathbf{Q} 是污染源排放率的向量; M_k 表示某一个污染输送模式的输送、化学过程; \mathbf{q} 表示这个模式的误差(那些不属于污染源的误差, 比如气象场、

地表覆盖以及地形等误差), 是一个随机变量; \mathbf{Q}^b 表示污染源排放率的某一已知的清单; \mathbf{Q}' 表示这一清单的误差, 也是一个随机变量。

假定在 t_0, t_1, \dots, t_k 时刻有污染物的浓度观测,

$$\mathbf{y}(t_k) = \mathbf{H}_k [\mathbf{C}(x, t_k)] + \mathbf{r}_k, \quad (2)$$

其中, \mathbf{r} 表示观测误差, 为一个随机变量。由于观测在空间分布和物种上都不是完全的, 因此 \mathbf{H}_k 可以表示空间或者物种的采样特征。

2.2 反演污染源排放的 Bayes 理论框架

如果没有观测, 我们可以从某个给定的初始条件 $\mathbf{C}^b(x, t_0)$, 某个排放清单 $\mathbf{Q}^b(x, t_k)$, 同时忽略模式误差[即 $\mathbf{q}(x, t_k) = 0, k = 1, \dots, K$], 通过求解(1)式中第一个方程获得一个数值解 $\mathbf{C}^b(x, t_k)$, 这是通常的数值模拟, 有时也称正问题。现在考虑如果有观测 $\mathbf{y}(t_k)$, 获得比 $\mathbf{C}^b(x, t_k)$ 和 $\mathbf{Q}^b(x, t_k)$ 更好的污染物浓度和污染源排放。为简化记号, 将 $\mathbf{C}(x, t_k)$, $\mathbf{Q}(x, t_k)$ 和 $\mathbf{y}(t_k)$ 记为 \mathbf{C}_k , \mathbf{Q}_k 和 \mathbf{y}_k 。相对于上面的正问题, 经常被称为反问题或反演问题。一般, 这样一个反演问题可以表达为:

反演问题: 在给定观测 $\mathbf{y}(t_k)$ 的条件下, 求污染源排放 $\mathbf{Q}(x, t_k)$ 和对应的随机微分方程(1)的解 $\mathbf{C}(x, t_k)$ 的联合概率密度函数 $f(\mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_K, \mathbf{Q}_1, \dots, \mathbf{Q}_K | \mathbf{y}_1, \dots, \mathbf{y}_K)$ 。

这个反演问题和正问题除了已知量和未知量是相反的外, 另外一个重要区别是通常不考虑正问题中的模式误差 $\mathbf{q}(x, t_k)$, 而反演问题需要考虑模式误差。

$$\begin{aligned} & \text{由 Bayes 定理 [即 } f(\mathbf{u}, \mathbf{v} | \mathbf{w}) \propto f(\mathbf{u}, \mathbf{v}) f(\mathbf{w} | \mathbf{u}, \mathbf{v})], \\ & f(\mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_K, \mathbf{Q}_1, \dots, \mathbf{Q}_K | \mathbf{y}_1, \dots, \mathbf{y}_K) \propto \\ & f(\mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_K, \mathbf{Q}_1, \dots, \mathbf{Q}_K) f(\mathbf{y}_1, \dots, \mathbf{y}_K | \mathbf{C}_0, \\ & \mathbf{C}_1, \dots, \mathbf{C}_K, \mathbf{Q}_1, \dots, \mathbf{Q}_K). \end{aligned} \quad (3)$$

接下来要用到几个基本的概率定理,

$$f(\mathbf{u}, \mathbf{v}) = f(\mathbf{u} | \mathbf{v}) f(\mathbf{v}),$$

如果 \mathbf{a} 、 \mathbf{b} 不相关, 那么

$$\begin{aligned} f(\mathbf{a}, \mathbf{b}) &= f(\mathbf{a}) f(\mathbf{b}), \quad f(\mathbf{a} | \mathbf{v}, \mathbf{b}) = f(\mathbf{a} | \mathbf{v}), \\ f(\mathbf{a} | \mathbf{b}) &= f(\mathbf{a}). \end{aligned}$$

因为

$$\begin{aligned} f(\mathbf{y}_1, \dots, \mathbf{y}_K | \mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_K, \mathbf{Q}_1, \dots, \mathbf{Q}_K) &= \\ f(\mathbf{y}_1 | \mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_K, \mathbf{Q}_1, \dots, \mathbf{Q}_K) \cdots f(\mathbf{y}_K | \mathbf{C}_0, \\ \mathbf{C}_1, \dots, \mathbf{C}_K, \mathbf{Q}_1, \dots, \mathbf{Q}_K). \end{aligned}$$

如果我们假定观测误差是不相关的, 并且由(2)式知 \mathbf{y}_j 仅与 \mathbf{C}_j 相关, 因此

$$f(\mathbf{y}_1, \dots, \mathbf{y}_K | \mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_K, \mathbf{Q}_1, \dots, \mathbf{Q}_K) = f(\mathbf{y}_1 | \mathbf{C}_1) \cdots f(\mathbf{y}_K | \mathbf{C}_K). \quad (4)$$

另外,

$$f(\mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_K, \mathbf{Q}_1, \dots, \mathbf{Q}_K) = f(\mathbf{C}_K, \mathbf{Q}_K | \mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_{K-1}, \mathbf{Q}_1, \dots, \mathbf{Q}_{K-1}) \cdot f(\mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_{K-1}, \mathbf{Q}_1, \dots, \mathbf{Q}_{K-1}). \quad (5)$$

如果假定模式为一阶 Markov 过程, 即 \mathbf{C}_k 仅与 \mathbf{C}_{k-1} 相关, 而与 $\mathbf{C}_{k-2}, \dots, \mathbf{C}_0$ 无关, 那么从(5)式可以得到,

$$f(\mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_K, \mathbf{Q}_1, \dots, \mathbf{Q}_K) = f(\mathbf{C}_K, \mathbf{Q}_K | \mathbf{C}_{K-1}, \mathbf{Q}_1, \dots, \mathbf{Q}_{K-1}) f(\mathbf{C}_{K-1}, \mathbf{Q}_1, \dots, \mathbf{Q}_{K-1}) = f(\mathbf{C}_K, \mathbf{Q}_K | \mathbf{C}_{K-1}, \mathbf{Q}_1, \dots, \mathbf{Q}_{K-1}) \cdot f(\mathbf{C}_{K-1}, \mathbf{Q}_{K-1} | \mathbf{C}_{K-2}, \mathbf{Q}_1, \dots, \mathbf{Q}_{K-2}) \cdots f(\mathbf{C}_1, \mathbf{Q}_1) f(\mathbf{C}_0). \quad (6)$$

因此结合(3)、(4)、(6)式, 有

$$f(\mathbf{C}_0, \mathbf{C}_1, \dots, \mathbf{C}_K, \mathbf{Q}_1, \dots, \mathbf{Q}_K | \mathbf{y}_1, \dots, \mathbf{y}_K) \propto f(\mathbf{C}_0) \cdot f(\mathbf{C}_1, \mathbf{Q}_1 | \mathbf{C}_0) f(\mathbf{y}_1 | \mathbf{C}_1) \cdot f(\mathbf{C}_2, \mathbf{Q}_2 | \mathbf{C}_1, \mathbf{Q}_1) f(\mathbf{y}_2 | \mathbf{C}_2) \cdot \vdots f(\mathbf{C}_K, \mathbf{Q}_K | \mathbf{C}_{K-1}, \mathbf{Q}_1, \dots, \mathbf{Q}_{K-1}) f(\mathbf{y}_K | \mathbf{C}_K). \quad (7)$$

如果模式不是一阶 Markov 过程, 也可以通过将变量扩展的方法将问题变为一个与一阶 Markov 过程相同处理的新问题来处理(详情可参阅文献[18]).

方程(7)给出了用顺序的方法求联合概率密度函数。从方程(7)出发, 如果加上一些假设, 可以导出诸如弱四维变分、集合平滑(ensemble smoother)、集合卡尔曼平滑(ensemble Kalman smoother, 简称 EnKS)和集合卡尔曼滤波(ensemble Kalman filter, 简称 EnKF)等实用的估计污染源排放

$$\begin{cases} \mathbf{C}_1^a = \arg^{-1}\{\mathbf{C}_1\}, f(\mathbf{C}_1^+) = f(\mathbf{C}_1); \\ (\mathbf{C}_1^a, \mathbf{Q}_1^a) = \arg^{-1}\{-f(\mathbf{C}_1^+, \mathbf{Q}_1^+)\}, f(\mathbf{C}_1^+, \mathbf{Q}_1^+) := f(\mathbf{C}_1, \mathbf{Q}_1 | \mathbf{C}_0^+) f(\mathbf{y}_1 | \mathbf{C}_1); \\ (\mathbf{C}_2^a, \mathbf{Q}_2^a) = \arg^{-1}\{-f(\mathbf{C}_2^+, \mathbf{Q}_2^+)\}, f(\mathbf{C}_2^+, \mathbf{Q}_2^+) := f(\mathbf{C}_2, \mathbf{Q}_2 | \mathbf{C}_1^+, \mathbf{Q}_1^+) f(\mathbf{y}_2 | \mathbf{C}_2); \\ \vdots \\ (\mathbf{C}_K^a, \mathbf{Q}_K^a) = \arg^{-1}\{-f(\mathbf{C}_K^+, \mathbf{Q}_K^+)\}, f(\mathbf{C}_K^+, \mathbf{Q}_K^+) := f(\mathbf{C}_K, \mathbf{Q}_K | \mathbf{C}_{K-1}^+, \mathbf{Q}_1^+, \dots, \mathbf{Q}_{K-1}^+) f(\mathbf{y}_K | \mathbf{C}_K). \end{cases} \quad (12)$$

这里, 上标⁺表示满足后验条件概率分布的随机变量; $\arg^{-1}\{g(z)\}$ 是指函数 $g(z)$ 的最小值点。根据高斯分布的特性, 数学期望为最大似然估计(同时也是最小方差估计), 因此在方程(12)中, $(\mathbf{C}_k^a, \mathbf{Q}_k^a) = E(\mathbf{C}_k^+, \mathbf{Q}_k^+)$ 。如果不考虑污染源排放误差的时间相关, 我们可以进一步简化为

$$\begin{cases} \mathbf{C}_1^a = E(\mathbf{C}_1), f(\mathbf{C}_1^+) = f(\mathbf{C}_1); \\ (\mathbf{C}_1^a, \mathbf{Q}_1^a) = E(\mathbf{C}_1^+, \mathbf{Q}_1^+), f(\mathbf{C}_1^+, \mathbf{Q}_1^+) := f(\mathbf{C}_1, \mathbf{Q}_1 | \mathbf{C}_0^+) f(\mathbf{y}_1 | \mathbf{C}_1); \\ (\mathbf{C}_2^a, \mathbf{Q}_2^a) = E(\mathbf{C}_2^+, \mathbf{Q}_2^+), f(\mathbf{C}_2^+, \mathbf{Q}_2^+) := f(\mathbf{C}_2, \mathbf{Q}_2 | \mathbf{C}_1^+) f(\mathbf{y}_2 | \mathbf{C}_2); \\ \vdots \\ (\mathbf{C}_K^a, \mathbf{Q}_K^a) = E(\mathbf{C}_K^+, \mathbf{Q}_K^+), f(\mathbf{C}_K^+, \mathbf{Q}_K^+) := f(\mathbf{C}_K, \mathbf{Q}_K | \mathbf{C}_{K-1}^+) f(\mathbf{y}_K | \mathbf{C}_K). \end{cases} \quad (13)$$

的方法。下面将导出集合卡尔曼滤波、集合卡尔曼平滑和集合平滑方法。

3 集合卡尔曼滤波

这里需要假定方程(7)中所有的概率密度函数都是高斯分布^[17]。首先确定初值的先验分布, 这里假定为如下正态分布形式:

$$f(\mathbf{C}_0) \propto \exp\left\{-\frac{1}{2} \iint [\mathbf{C}_0(x) - \mathbf{C}_0^c(x)]^T \mathbf{W}_{C_0}(x, x') [\mathbf{C}_0(x') - \mathbf{C}_0^c(x')] dx dx'\right\}. \quad (8)$$

对污染源排放的先验分布, 一种办法假定时间不相关, 即

$$f(\mathbf{Q}_k) \propto \exp\left\{-\frac{1}{2} \iint [\mathbf{Q}_k(x) - \mathbf{Q}_k^b(x)]^T \mathbf{W}_{Q_k}(x, x') \cdot [\mathbf{Q}_k(x') - \mathbf{Q}_k^b(x')] dx dx'\right\}. \quad (9)$$

如果考虑时间相关性, 则可以用如下的一阶 Markov 过程获得污染源排放的先验分布,

$$\mathbf{Q}_k = \alpha \mathbf{Q}_{k-1} + \sqrt{1 - \alpha^2} \mathbf{w}. \quad (10)$$

其中, \mathbf{w} 如(9)式的右边定义, α 是一个正数, 刻画了时间相关尺度(值越大, 相关时间尺度越大)。当然也可以采用更加高阶的 Markov 过程。

观测误差的先验分布定义为

$$f(\mathbf{y}_k | \mathbf{C}_k) \propto \exp\left\{-\frac{1}{2} [\mathbf{y}_k - \mathbf{H}_k(\mathbf{C}_k)]^T \mathbf{W}_o [\mathbf{y}_k - \mathbf{H}_k(\mathbf{C}_k)]\right\}. \quad (11)$$

模式误差的先验分布也可以用类似(10)定义。

为了简化方程(7)的估计方法, 我们进行如下近似的顺序估计:

下面,以求解(13)为例导出集合卡尔曼滤波。由于各个先验分布的高斯假设, $f(\mathbf{C}_k, \mathbf{Q}_k | \mathbf{C}_{k-1}^+)$ 也必定是高斯的,即

$$f(\mathbf{C}_k, \mathbf{Q}_k | \mathbf{C}_{k-1}^+) \propto \exp\left\{-\frac{1}{2} \iint \begin{bmatrix} \mathbf{C}_k(x) - \mathbf{C}_k^f(x) \\ \mathbf{Q}_k(x) - \mathbf{Q}_k^b(x) \end{bmatrix}^T \begin{pmatrix} \mathbf{W}_{C_k}(x, x') & \mathbf{W}_{QC_k}^T(x, x') \\ \mathbf{W}_{QC_k}(x, x') & \mathbf{W}_{Q_k}(x, x') \end{pmatrix} \begin{bmatrix} \mathbf{C}_k(x') - \mathbf{C}_k^f(x') \\ \mathbf{Q}_k(x') - \mathbf{Q}_k^b(x') \end{bmatrix} dx dx'\right\}. \quad (14)$$

从而

$$f(\mathbf{C}_k^+, \mathbf{Q}_k^+) = f(\mathbf{C}_k, \mathbf{Q}_k | \mathbf{C}_{k-1}^+) f(\mathbf{y}_k | \mathbf{C}_k) \propto \exp\left\{-\frac{1}{2} \iint \begin{bmatrix} \mathbf{C}_k(x) - \mathbf{C}_k^f(x) \\ \mathbf{Q}_k(x) - \mathbf{Q}_k^b(x) \end{bmatrix}^T \begin{pmatrix} \mathbf{W}_{C_k}(x, x') & \mathbf{W}_{QC_k}^T(x, x') \\ \mathbf{W}_{QC_k}(x, x') & \mathbf{W}_{Q_k}(x, x') \end{pmatrix} \begin{bmatrix} \mathbf{C}_k(x') - \mathbf{C}_k^f(x') \\ \mathbf{Q}_k(x') - \mathbf{Q}_k^b(x') \end{bmatrix} dx dx' - \frac{1}{2} [\mathbf{y}_k - \mathbf{H}_k(\mathbf{C}_k)]^T \mathbf{W}_o [\mathbf{y}_k - \mathbf{H}_k(\mathbf{C}_k)]\right\}. \quad (15)$$

如果记

$$J(\mathbf{C}_k, \mathbf{Q}_k) = \iint \begin{bmatrix} \mathbf{C}_k(x) - \mathbf{C}_k^f(x) \\ \mathbf{Q}_k(x) - \mathbf{Q}_k^b(x) \end{bmatrix}^T \begin{pmatrix} \mathbf{W}_{C_k}(x, x') & \mathbf{W}_{QC_k}^T(x, x') \\ \mathbf{W}_{QC_k}(x, x') & \mathbf{W}_{Q_k}(x, x') \end{pmatrix} \begin{bmatrix} \mathbf{C}_k(x') - \mathbf{C}_k^f(x') \\ \mathbf{Q}_k(x') - \mathbf{Q}_k^b(x') \end{bmatrix} dx dx' + [\mathbf{y}_k - \mathbf{H}_k(\mathbf{C}_k)]^T \mathbf{W}_o [\mathbf{y}_k - \mathbf{H}_k(\mathbf{C}_k)],$$

那么,

$$(\mathbf{C}_k^a, \mathbf{Q}_k^a) = \arg^{-1}\{J(\mathbf{C}_k, \mathbf{Q}_k)\}.$$

按照经典的卡尔曼滤波的推导^[19],有

$$\begin{bmatrix} \mathbf{C}_k^a \\ \mathbf{Q}_k^a \end{bmatrix} = \begin{bmatrix} \mathbf{C}_k^f \\ \mathbf{Q}_k^b \end{bmatrix} + \mathbf{K}(\mathbf{y}_k - \mathbf{H}_k(\mathbf{C}_k^f)), \quad (16)$$

$$\mathbf{K} = \begin{bmatrix} \mathbf{V}_{C_k}(x, x') \mathbf{H}_k^T \\ \mathbf{V}_{QC_k}(x, x') \mathbf{H}_k^T \end{bmatrix} (\mathbf{H}_k \mathbf{V}_{C_k}(x, x') \mathbf{H}_k^T + \mathbf{V}_o)^{-1}, \quad (17)$$

其中,

$$\begin{cases} \begin{pmatrix} \mathbf{V}_{C_k}(x, x') & \mathbf{V}_{QC_k}^T(x, x') \\ \mathbf{V}_{QC_k}(x, x') & \mathbf{V}_{Q_k}(x, x') \end{pmatrix} = \begin{pmatrix} \mathbf{W}_{C_k}(x, x') & \mathbf{W}_{QC_k}^T(x, x') \\ \mathbf{W}_{QC_k}(x, x') & \mathbf{W}_{Q_k}(x, x') \end{pmatrix}^{-1}, \\ \mathbf{V}_o = \mathbf{W}_o^{-1}. \end{cases} \quad (18)$$

为了计算(16)式,我们可以采用集合方法。上标 j ($j = 1, \dots, N$) 表示集合成员。那么初值的 N 个成员可以按照给定的分布(8)式生成,记为 $\mathbf{C}_0^{(j)}$ 。类似,按照(9)和(11)式得到污染源排放的 N 个成员 $\mathbf{Q}_1^{(j)}$ 和模式误差的 N 个成员 $\mathbf{q}^{(j)}$ 。利用这些成员积分模式 N 次,就可以得到 N 个模式结果 $\mathbf{C}_1^{(j)}$ 。利用样本矩阵

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{C}_1^{(1)} & \mathbf{C}_1^{(2)} & \dots & \mathbf{C}_1^{(N)} \\ \mathbf{Q}_1^{(1)} & \mathbf{Q}_1^{(2)} & \dots & \mathbf{Q}_1^{(N)} \end{bmatrix},$$

可以用(19)式和(14)式统计出 $f(\mathbf{C}_1, \mathbf{Q}_1 | \mathbf{C}_0^+)$ 分布的协方差矩阵:

$$\begin{pmatrix} \mathbf{V}_{C_1}(x, x') & \mathbf{V}_{QC_1}^T(x, x') \\ \mathbf{V}_{QC_1}(x, x') & \mathbf{V}_{Q_1}(x, x') \end{pmatrix} = \frac{1}{N-1} \mathbf{A}'_1 \mathbf{A}'_1^T, \quad (19)$$

其中, \mathbf{A}'_1 表示减去样本平均的样本矩阵。类似,可以利用观测误差的先验分布(11)生成一组观测样

本,从而可以像(19)式那样得到观测误差的协方差矩阵 \mathbf{V}_o 。由此,可以用(17)、(18)式计算出增益矩阵 \mathbf{K} 。由此获得满足后验分布 $f(\mathbf{C}_1^+, \mathbf{Q}_1^+)$ 的一组样本:

$$\begin{bmatrix} \mathbf{C}_1^{(j)+} \\ \mathbf{Q}_1^{(j)+} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1^{(j)} \\ \mathbf{Q}_1^{(j)} \end{bmatrix} + \mathbf{K}[\mathbf{y}_1^{(j)} - \mathbf{H}_1(\mathbf{C}_1^{(j)})], \quad (20)$$

其中, $\mathbf{y}_1^{(j)}$ ($j = 1, \dots, N$) 是按照(11)式得到一组观测扰动样本。我们可以得到污染物浓度和污染源排放的估计:

$$\mathbf{C}_1^a = \frac{1}{N} \sum_{j=1}^N \mathbf{C}_1^{(j)+},$$

$$\mathbf{Q}_1^a = \frac{1}{N} \sum_{j=1}^N \mathbf{Q}_1^{(j)+}.$$

这一步骤可以类似地在随后的时间采用,计算出满足 $f(\mathbf{C}_k^+, \mathbf{Q}_k^+) := f(\mathbf{C}_k, \mathbf{Q}_k | \mathbf{C}_{k-1}^+) f(\mathbf{y}_k | \mathbf{C}_k)$ 的样本,直到算出 \mathbf{C}_k^a 和 \mathbf{Q}_k^a 。

如果考虑到污染源排放误差的时间相关,比如

采用(10)式来作为先验分布,那么在 t_k 时得到污染源排放的样本时采用下面的方案:

$$Q_k^{(j)} = \alpha Q_{k-1}^{(j)} + \sqrt{1 - \alpha^2} w_k^{(j)}, \quad (21)$$

而其他类似上面的推导,可以从(12)式出发得到类似的方案。

4 集合平滑和集合卡尔曼平滑

4.1 集合平滑

在上节的集合卡尔曼滤波中,我们看到一个弱点是某一时刻的污染物浓度观测不能够对前面时刻的污染源排放进行估计。由于前面时刻的污染源排放会对此时此刻的污染物浓度有影响,因此从原理上应该利用此时此刻的污染物浓度观测改进对前面的污染源排放的估计。van Leeuwen 等^[20]提出一个集合同化算法,称为集合平滑,可以达到这个目的。这里我们将他们的方法应用到污染源排放的估计中。

把所有 t_1, \dots, t_K 时刻的污染源排放、污染物浓度以及观测一起处理。记

$$\begin{cases} \mathbf{S} = (\mathbf{C}_1, \dots, \mathbf{C}_K, \mathbf{Q}_1, \dots, \mathbf{Q}_K)^T, \\ \mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_K)^T. \end{cases} \quad (22)$$

首先根据 $\mathbf{C}_0, \mathbf{Q}_1, \dots, \mathbf{Q}_K$ 以及模式误差 $\mathbf{q}_1, \dots, \mathbf{q}_K$ 的先验分布来得到 N 组样本 $\mathbf{C}_0^{(j)}, \mathbf{C}_1^{(j)}, \dots, \mathbf{Q}_K^{(j)}$ 和 $\mathbf{q}_1^{(j)}, \dots, \mathbf{q}_K^{(j)}$ ($j=1, \dots, K$)。这里,我们依然假定所有先验分布为高斯的。将它们分别代入方程(1),可以得到 $(\mathbf{C}_1, \dots, \mathbf{C}_K)$ 所满足的先验分布的 N 组样本 $\mathbf{C}_1^{(j)}, \dots, \mathbf{C}_K^{(j)}$ ($j=1, \dots, K$)。那么,我们就可以统计出 \mathbf{S} 的误差协方差矩阵: $\mathbf{B} = [1/N - 1] \mathbf{\Pi}' \mathbf{\Pi}'^T$, 其中 $\mathbf{\Pi}'$ 是样本矩阵 $\mathbf{\Pi}$ 减去样本平均的矩阵,而样本矩阵 $\mathbf{\Pi}$ 为

$$\mathbf{\Pi} = (\mathbf{S}^{(1)}, \dots, \mathbf{S}^{(N)}) = \begin{pmatrix} \mathbf{C}_1^{(1)} & \dots & \mathbf{C}_K^{(N)} \\ \vdots & & \vdots \\ \mathbf{C}_K^{(1)} & \dots & \mathbf{C}_K^{(N)} \\ \mathbf{Q}_1^{(1)} & \dots & \mathbf{Q}_1^{(N)} \\ \vdots & & \vdots \\ \mathbf{Q}_K^{(1)} & \dots & \mathbf{Q}_K^{(N)} \end{pmatrix}.$$

同样,对观测我们也按照观测误差的先验分布得到一组观测样本 $\mathbf{\Psi} = (\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(N)})$ 。同样,构造观测误差协方差矩阵

$$\mathbf{O} = \frac{1}{N-1} \mathbf{\Psi}' \mathbf{\Psi}'^T.$$

依照经典的卡尔曼滤波理论,可以得到

$$\mathbf{S}^{+(j)} = \mathbf{S}^{(j)} + \mathbf{K}(\mathbf{y}^{(j)} - \mathbf{H}\mathbf{S}^{(j)}), \quad (23)$$

其中, $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{O})^{-1}$, \mathbf{H} 为 \mathbf{S} 空间到观测空间 \mathbf{y} 的投影算子,可由(2)式中的 \mathbf{H}_k 构造。

在某一时刻,由于前面时刻的污染源排放会对此时此刻的污染物浓度有影响,因此由样本矩阵 $\mathbf{\Pi}$ 得到的 \mathbf{B} 能够体现前面时刻的污染源排放会与此时刻的污染物浓度的相关,因此(23)式可以根据此时此刻的污染物浓度的观测来修正前面时刻的污染源排放。

由于实际观测资料可以是很长的时间序列,因此这一方法在实际应用中可以有几个操作方式:

(i) 分段估计,即将长时间观测 $\mathbf{y}_1, \dots, \mathbf{y}_k, \dots$ 分成不重合的时间段,然后分段使用集合平滑来估计该段时间的污染源排放和污染物浓度。

(ii) 累进估计,即开始时利用某一段观测 $\mathbf{y}_1, \dots, \mathbf{y}_k$ 估计这段时间的污染源排放 $\mathbf{Q}_1, \dots, \mathbf{Q}_k$ 和污染物浓度 $\mathbf{C}_1, \dots, \mathbf{C}_k$ 。接下来用增加的时间段观测 $\mathbf{y}_1, \dots, \mathbf{y}_k, \mathbf{y}_{k+1}$ 来估计 $\mathbf{Q}_1, \dots, \mathbf{Q}_k, \mathbf{Q}_{k+1}$ 和 $\mathbf{C}_1, \dots, \mathbf{C}_k, \mathbf{C}_{k+1}$ 。这里要注意的是,在用 $\mathbf{y}_1, \dots, \mathbf{y}_k, \mathbf{y}_{k+1}$ 估计 $\mathbf{Q}_1, \dots, \mathbf{Q}_k, \mathbf{Q}_{k+1}$ 和 $\mathbf{C}_1, \dots, \mathbf{C}_k, \mathbf{C}_{k+1}$ 时,生成 $\mathbf{Q}_1, \dots, \mathbf{Q}_k$ 和 $\mathbf{C}_1, \dots, \mathbf{C}_k$ 的先验分布样本时要按照前一步用 $\mathbf{y}_1, \dots, \mathbf{y}_k$ 估计 $\mathbf{Q}_1, \dots, \mathbf{Q}_k$ 和 $\mathbf{C}_1, \dots, \mathbf{C}_k$ 所得到的后验分布来生成,并且对 $\mathbf{Q}_1, \dots, \mathbf{Q}_k$ 和 $\mathbf{C}_1, \dots, \mathbf{C}_k$ 的修正也是在 $\mathbf{Q}_1, \dots, \mathbf{Q}_k$ 和 $\mathbf{C}_1, \dots, \mathbf{C}_k$ 前一步估计的基础上的。

Evensen^[17]提出了一个改进累进估计的方法:用 $\mathbf{y}_1, \dots, \mathbf{y}_k, \mathbf{y}_{k+1}$ 估计 $\mathbf{Q}_1, \dots, \mathbf{Q}_k, \mathbf{Q}_{k+1}$ 和 $\mathbf{C}_1, \dots, \mathbf{C}_k, \mathbf{C}_{k+1}$ 这一步与前一步(k 时刻)用 $\mathbf{y}_1, \dots, \mathbf{y}_k$ 估计 $\mathbf{Q}_1, \dots, \mathbf{Q}_k$ 和 $\mathbf{C}_1, \dots, \mathbf{C}_k$ 相比,新信息其实只有 \mathbf{y}_{k+1} ,因此,可以考虑将这一步改为仅用 \mathbf{y}_{k+1} 来估计 $\mathbf{Q}_1, \dots, \mathbf{Q}_k, \mathbf{Q}_{k+1}$ 和 $\mathbf{C}_1, \dots, \mathbf{C}_k, \mathbf{C}_{k+1}$ 。这一新的方法被称为集合卡尔曼平滑。我们在下一节将此方法推广到污染源排放和污染物浓度问题上。

4.2 集合卡尔曼平滑

集合卡尔曼平滑的步骤如下:

在时间 $t=t_1$ 时,记 $\mathbf{S}_1 = (\mathbf{C}_1, \mathbf{Q}_1)^T$ 。集合卡尔曼平滑和上节的集合平滑一样,估计出的后验验证分布的样本记为 $\mathbf{S}_1^{+(j)} = \mathbf{S}_1^{(j)}$;

⋮

在时间 $t=t_{k+1}$ 时,记 $\mathbf{S}_{k+1} = (\mathbf{C}_1, \dots, \mathbf{C}_{k+1}, \mathbf{Q}_1, \dots, \mathbf{Q}_{k+1})^T$ 。根据在时间 $t=t_k$ 得到的最新后验分布样本 $\mathbf{C}_0^{+(j)}, \mathbf{C}_1^{+(j)}, \dots, \mathbf{C}_k^{+(j)}, \mathbf{Q}_1^{+(j)}, \mathbf{Q}_2^{+(j)}, \dots, \mathbf{Q}_k^{+(j)}$ 以及模式误差 \mathbf{q}_{k+1} 的先验分布样本和 $\mathbf{q}_k^{(j)}$ 和

Q_{k+1} 的先验分布样本 $Q_k^{(j)}$ ($j=1, \dots, N$), 从 $t=t_k$ 到 $t=t_{k+1}$ 积分模式(1) N 次, 可以得到 C_{k+1} 的样本 $C_k^{(j)}$ 。那么, 我们就可以统计出 S_{k+1} 的误差协方差矩阵:

$$B_{k+1} = \frac{1}{N-1} \Pi'_{k+1} \Pi_{k+1}^T,$$

其中, Π'_{k+1} 是样本矩阵 Π_{k+1} 减去样本平均的矩阵, 而样本矩阵 Π_{k+1} 为

$$\Pi_{k+1} = (S_{k+1}^{(1)}, \dots, S_{k+1}^{(N)}) = \begin{pmatrix} C_1^{+(1)} & \dots & C_1^{+(N)} \\ \vdots & & \vdots \\ C_k^{+(1)} & \dots & C_k^{+(N)} \\ C_{k+1}^{(1)} & \dots & C_{k+1}^{(N)} \\ Q_1^{+(1)} & \dots & Q_1^{+(N)} \\ \vdots & & \vdots \\ Q_k^{+(1)} & \dots & Q_k^{+(N)} \\ Q_{k+1}^{(1)} & \dots & Q_{k+1}^{(N)} \end{pmatrix}.$$

同样, 对观测我们也按照观测误差的先验分布得到一组观测样本 $\Psi_{k+1} = (y_{k+1}^{(1)}, \dots, y_{k+1}^{(N)})$ 。同样, 构造观测误差协方差矩阵,

$$O_{k+1} = \frac{1}{N-1} \Psi'_{k+1} \Psi_{k+1}^T.$$

依照经典的卡尔曼滤波理论, 与上面的讨论类似, 可以根据最新的观测得到最新的后验样本,

$$S_{k+1}^{(j)} = S_{k+1}^{(j)} + K_{k+1} (y_{k+1}^{(j)} - H_{k+1} S_{k+1}^{(j)}),$$

其中, $K_{k+1} = B_{k+1} H_{k+1}^T (H_{k+1} B_{k+1} H_{k+1}^T + O_{k+1})^{-1}$, H_{k+1} 为 S_{k+1} 空间到观测空间 y_{k+1} 的投影算子。

5 一个简单模型的数值试验

5.1 模型和试验设置

我们采用下面描述空气质量方程的一个简单的、无量纲概念模型,

$$\frac{dC}{dt} = -0.2C + Q(t).$$

其中, Q 是污染源并随时间是变化的, C 是污染物浓度。试验中我们假设真实的污染源为 $Q_{true}(t) = \sin t$, 把初始浓度 $C_{true}(0) = 1.0$ 在时间段 $t \in [0, 5]$ 的积分视为真值, 积分的时间步长取 0.1。在我们的试验中不考虑模式误差。

观测值是在真实浓度上加白噪音。集合成员个数为 100。这里将浓度的初猜值设为 $C^f(0) = 0.2$, 误差标准差设为 0.2。我们假设污染源的背景值 $Q^b(t) = \sin t + 0.5$, 分别考察利用 EnKF 方法、En-

KS 方法在不同的观测间隔、观测误差和先验误差标准差的情况下污染物浓度同化和污染源反演的结果。在下面的试验中都没有考虑污染源排放先验误差的时间相关, 即设(21)式的 $\alpha = 0$ 。试验共有 8 个, 具体设计参见表 1。

5.2 试验结果

图 1 是试验 1 对污染物浓度观测进行 EnKF 的同化结果。从图 1a 的浓度时间序列可以看出, 在没有同化的情况下, 模拟值(黑实线)与真实值(红实线)差别很大, 在时间 $t > 1.5$ 时, 模拟值偏大, 这是因为对污染源的估计比真实污染源要大; 而在时间 $t < 1.5$ 时造成模拟值偏小的原因主要是由于对浓度的初估值(0.2)比真实值(1.0)小很多。当有观测进行同化时, 由于观测误差在此试验中假设很小, 同化时刻的浓度值与真值很接近。由于污染源先验估计比真实污染源大, 导致模拟值比真实值偏大, 不过每次同化后都能使同化时刻的浓度与真实值接近, 并进而使整个时间区间上的同化结果向真实值靠拢。因此, 利用 EnKF 同化, 模式对浓度的模拟结果精度增加了。图 1b 给出了污染源的时间序列。红线是污染源的真正值, 黑线是污染源的背景值, 蓝线是反演结果。在每次有浓度观测用 EnKF 同化后, 反演结果就向真实值接近。而在没有同化的时刻时则几乎还是背景值。

EnKF 无法利用观测对前面时刻的污染源进行反演的弱点可以采用 EnKS 方法来解决。试验 2 的设置采用 EnKS, 其他都与试验 1 相同(图 2)。与试验 1 相比, 试验 2 无论是浓度还是污染源的结果都有了很大的改善。图 2a 显示污染物浓度在时间上很光滑, 同时从图 2b 显示的污染源反演结果看, 在没有观测的时刻污染源也得到了显著改善。比较试验 1 和试验 2 污染源排放反演的误差(见表 2), EnKS 比 EnKF 能够得到更小的反演误差。

上面的两个试验假设了很小的观测误差(0.1, 约为真值的 1%), 为了考察观测误差对 EnKF 和 EnKS 反演污染源排放的影响, 我们在试验 3 和试验 4 中将观测误差加大到 0.1 (约为真值的 10%), 其他分别与试验 1 和试验 2 相同。图 3 给出了试验 3 和试验 4 的结果。可以看出 EnKF 和 EnKS 效果都有一定降低。EnKS 对观测误差比 EnKF 更为敏感。但是表 2 给出的反演误差稍微有些增加, 远小于观测误差的增加。

表 1 试验列表

Table 1 The detailed setup of the experiments

试验 Expt	方法 Method	观测时间间隔 Observation time interval	观测误差标准差 Standard deviation of observation error	污染源先验误差标准差 Standard deviation of emission prior-estimate	试验目的 Expt purpose
试验 1 Expt 1	EnKF	0.5	0.01	0.5	检验 EnKF 对污染物浓度的同化效果和污染源的反演能力 To test the ability of EnKF applied to concentration assimilation and inverting emission
试验 2 Expt 2	EnKS	0.5	0.01	0.5	检验 EnKS 对污染物浓度的同化效果和污染源的反演能力 To test the ability of EnKS applied to concentration assimilation and inverting emission
试验 3 Expt 3	EnKF	0.5	0.1	0.5	考察大的观测误差对 EnKF 结果的影响 To investigate the impact of large observation error on EnKF
试验 4 Expt 4	EnKS	0.5	0.1	0.5	考察大的观测误差对 EnKS 结果的影响 To investigate the impact of large observation error on EnKS
试验 5 Expt 5	EnKF	0.1	0.01	0.5	考察观测时间加密对 EnKF 结果的影响 To investigate the impact of large observation time interval on EnKF
试验 6 Expt 6	EnKS	0.1	0.01	0.5	考察观测时间加密对 EnKS 结果的影响 To investigate the impact of large observation time interval on EnKS
试验 7 Expt 7	EnKF	0.1	0.01	0.1	考察污染源先验误差标准差偏小对 EnKF 结果的影响 To investigate the impact of small emission prior error on EnKF
试验 8 Expt 8	EnKF	0.1	0.01	1.0	考察污染源先验误差标准差偏大对 EnKF 结果的影响 To investigate the impact of large emission prior error on EnKF

在试验 5 和 6 试验中, 我们增加观测密度, 假设每个时间步长上都有观测。除了每步都同化外, 其他都分别和试验 1 和试验 2 相同。图 4 给出了试验 5 和试验 6 的污染源反演结果。可以看出 EnKF 和 EnKS 的反演结果差别比试验 1 和试验 2 的差别显著减少。由于每个时间步长上都有观测, EnKF 能够反演每个时刻的污染源排放, 而 EnKS 也反演每个时刻的污染源排放。从表 2 的反演误差也可以看出来。观测密度的增加也使得 EnKF 和 EnKS 反演误差都有了显著减少。

由于 EnKF 和 EnKS 在反演污染源排放时需要估计背景值的误差。在上面的试验中, 我们假设这

表 2 各试验反演污染源的均方根误差

Table 2 The root-mean-square error for the inversed emission

试验 Experiment	均方根误差 Root-mean-square error
Expt 1	0.4614
Expt 2	0.2655
Expt 3	0.4698
Expt 4	0.3150
Expt 5	0.1682
Expt 6	0.1665
Expt 7	0.2070
Expt 8	0.2509

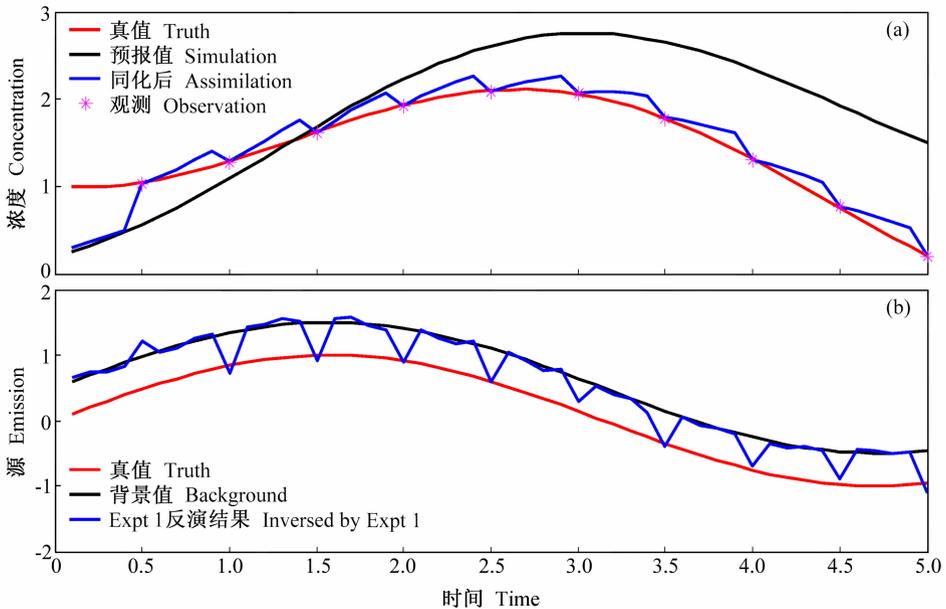


图1 试验1的结果：(a) 污染物浓度；(b) 污染源

Fig. 1 The results for Expt 1; (a) Pollutant concentration; (b) pollution emission

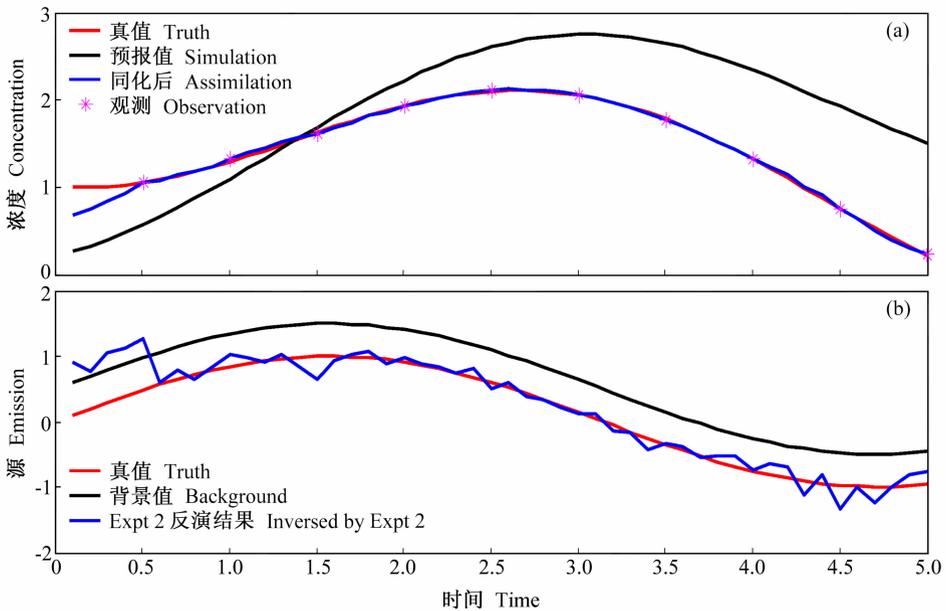


图2 试验2的结果：(a) 污染物浓度；(b) 污染源

Fig. 2 The same as Fig. 1, but for Expt 2

些先验误差大致是准确的（为 0.5，等于真值减背景值）。然而在实际应用中，我们仅能够对这些先验误差进行不准确的猜测，不可避免会有猜测不准的情况。因此在试验 7 和试验 8 中，我们考察 EnKF 先验误差猜测不准的情况。除了先验误差不同外，其他都与试验 5 相同。在试验 7 中假设先验误差估计不足，而在试验 8 中假设过分估计先验误

差。图 5 给出了试验 7 和试验 8 的结果。在试验 7 中，由于先验误差偏小，因此反演的污染源相对试验 5 更加靠近背景值。在试验 8 中，由于先验误差偏大，因此 EnKF 减少了背景值对反演结果的影响，而加大了观测对反演结果的影响。由于背景值在时间上比较光滑，而观测相对来说不太光滑，因此反演结果比试验 5 更加不光滑。两个试验的反演

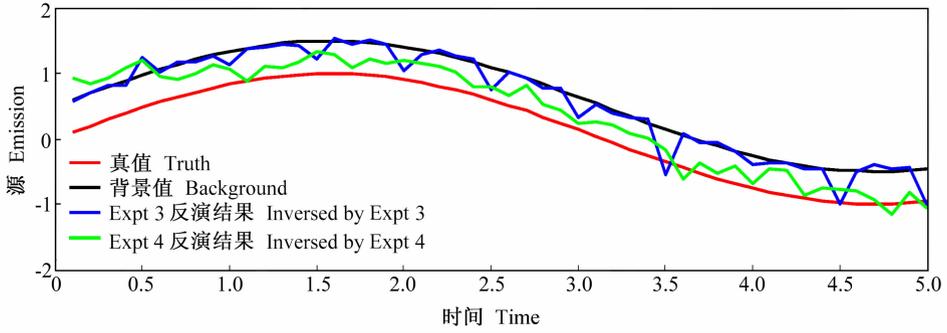


图 3 试验 3 和试验 4 的污染源反演结果

Fig. 3 Inversed emissions by Expt 3 and Expt 4

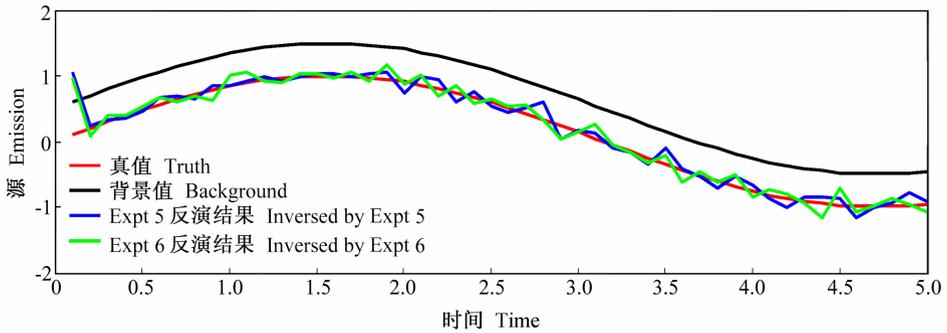


图 4 试验 5 和试验 6 的污染源反演结果

Fig. 4 Inversed emissions by Expt 5 and Expt 6

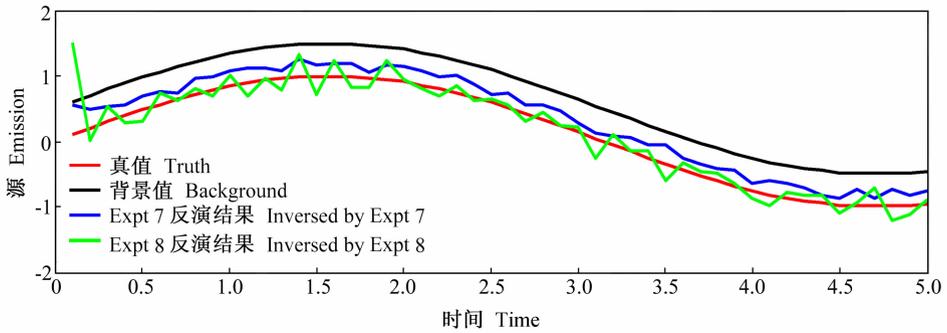


图 5 试验 7 和试验 8 的污染源反演结果

Fig. 5 Inversed emissions by Expt 7 and Expt 8

误差也都比试验 5 变大(见表 2)。这说明先验误差在污染源反演中的重要性。因此,在实际污染源反演时需要认真考虑对现有污染源排放清单(背景值)的误差进行合理估计,以保证反演结果的可靠性。

我们还进行了其他试验,包括在 EnKF 和 EnKS 中应用正则化/约束方法来减少反演误差,比如考虑污染源排放先验误差的时间相关[即(21)式的

$\alpha \neq 0$]。由于篇幅所限,将另文论述。

6 小结和讨论

污染源反演问题可以被作为确定性问题来处理,由此可以使用如四维变分/伴随模式方法。然而,由于污染源反演问题采用的数值模式都有许多各类误差,除了污染源的误差外,比如气象场、地表覆盖以及地形等也会带来很大误差,所以反演结

果会受到这些误差的影响,产生不确定性。而这些确定性的方法在计算结果的不确定性方面是不方便的。比如需要计算代价函数的二阶导数等。因此将污染源反演问题作为一个统计估计问题是一个更好的选择。以前这方面的研究集中在离散卡尔曼滤波和集合卡尔曼滤波。本文引入一些其他统计方法如集合平滑、集合卡尔曼平滑,解决污染源反演问题。这些方法除了不需要模式的伴随模式外,另外一个好处是可以给出污染源反演结果的不确定性。不确定性的信息可以从后验样本中得到。

这些统计方法的几个重要的共同弱点包括:先验分布的高斯假设;线性的分析方程[如(16)式]。在模式为强非线性时,这些假设会对结果造成不利影响。解决这个问题的一些实际办法可以参见文献[21, 22]。这些弱点对实际的污染源反演的影响以及这些方法的效果和计算代价是下一步研究的重要问题。

从本研究简单模型的试验可以得到如下结论:

(1) 在观测间隔小和观测误差小的情况下,集合卡尔曼滤波和集合卡尔曼平滑都可以有效地反演随时间变化的污染源排放。

(2) 当观测误差增大时,EnKF 和 EnKS 效果都有一定降低,但是反演误差的增加少于观测误差的增加,同时 EnKS 对观测误差比 EnKF 更为敏感。

(3) 当观测时间间隔较大时,由于 EnKF 不能估计没有观测时刻的污染源排放,因此仅能够对有观测时的污染源排放进行较好的反演。而 EnKS 可以利用观测对观测时刻前的污染源排放进行反演,因此其效果明显好于 EnKF,在观测时间间隔较大的情况下依然可以较好地反演出污染源排放。

(4) 污染源排放的先验误差估计对反演的结果有较大影响。

通过观测资料和模式反演污染源排放,在实际应用中要遇到许多来自各方面的各种问题。来自观测的问题包括:污染源分布的空间不均匀性和相对较少的观测站点;可观测的物种较少而需要反演的污染源物种较多。这些来自观测的局限会使反演结果不唯一。解决的办法通常是考虑对污染源排放给予一定的约束,以减少未知量的个数。来自模式的问题包括除了污染源排放的不确定性外,其他模式误差包括气象场误差、模式下垫面处理(地形、土地利用和地表粗糙度)的误差、边界层参数化的误

差、沉降和化学反应过程中的近似带来的误差以及模式分辨率等计算误差等。如果忽略这些模式误差,就会使反演方法不适当地调整污染源排放,以补偿由于模式其他误差造成的模拟和观测之间不一致,从而导致污染源排放反演结果出现偏差。四维变分方法和集合卡尔曼滤波/平滑在处理模式误差上具有很大的区别。在四维变分中考虑模式误差有两个办法,一是显式地表示这些误差,并且在同化时进行估计,但是这一方法会大大增加未知量的个数,使问题变得不适定,也会使计算代价大大增加;另一个办法是将这些模式误差所导致的模式状态变量(污染物浓度)的不确定性进行定量猜测^[23],这样可以不增加未知量的个数,但是做好这样的猜测会遇到相当的困难。比如,我们知道地面粗糙度有 25% 的误差,那么猜测这一误差会造成多大的地面 SO₂ 浓度的误差不是一件容易的事情。在集合卡尔曼滤波/平滑中处理模式误差不需要显式地表示这些误差,而是采用将这些模式误差所导致的模式状态变量(污染物浓度)的不确定性进行定量计算。还以地面粗糙度为例,为了计算地面粗糙度 25% 的误差会对地面 SO₂ 浓度造成多大误差,集合卡尔曼滤波/平滑先按照 25% 的误差生成一组(比如 100 个)不同的地面粗糙度,然后以这 100 个粗糙度作为模式输入,分别积分模式得到 100 个地面 SO₂ 浓度,然后统计所造成的地面 SO₂ 浓度的误差是多少。因此,集合卡尔曼滤波/平滑在处理模式误差方面更加方便。

以上这些问题并未在本文讨论,但却是实际应用中不可回避的,需要在今后的工作中加以探讨。

参考文献 (References)

- [1] Xia Liping, Shao Yaping. Modelling of traffic flow and air pollution emission with application to Hong Kong Island. *Environmental Modelling & Software*, 2005, **20**: 1175~1188
- [2] 郝吉明, 王丽涛, 李林, 等. 北京市能源相关大气污染源的贡献率和调控对策分析. *中国科学(D)*, 2005, **35** (增刊 1): 115~122
Hao Jiming, Wang Litao, Li Lin, et al. Analysis of contribution and control measure for atmospheric contamination source related to energy. *Science in China (Series D) (in Chinese)*, 2005, **35** (Suppl. 1): 115~122
- [3] 徐大海, 朱蓉. 城市大气污染源强反演. *大气环境和环境影响评价*. 北京: 气象出版社, 1998

- Xu Dahai, Zhu Rong. Inversion of city atmospheric contamination source. *Atmospheric Environment and Environmental Impact Assessment* (in Chinese). Beijing: China Meteorological Press, 1998
- [4] 刘文菁, 黄世鸿, 刘小红, 等. 南京市总悬浮颗粒物(TSP)及地面积尘来源解析. *气象科学*, 2001, **21**: 87~94
Liu Wenjing, Huang Shihong, Liu Xiaohong. Source apportionment for total suspended particles and accumulated dust of Nanjing. *Scientia Meteorologica Sinica* (in Chinese), 2001, **21**: 87~94
- [5] Elbern H, Schmidt H, Talagrand O, et al. 4D-variational data assimilation with an adjoint air quality model for emission analysis. *Environmental Modelling & Software*, 2000, **15**: 539~548
- [6] Pudykiewicz J A. Application of adjoint tracer transport equations for evaluating source parameters. *Atmos. Environ.* 1998, **32**: 3039~3050
- [7] Mendoza-Dominguez A, Russell A G. Estimation of emission adjustments from the application of four-dimensional data assimilation to photochemical air quality modeling. *Atmos. Environ.*, 2001, **35**: 2879~2894
- [8] Prinn R G, Weiss R F, Miller B R, et al. Atmospheric trends and lifetime of CH₃CCl₃ and global OH concentrations. *Science*, 1995, **269**: 187~192
- [9] Gilliland A, Abbitt P J. A sensitivity study of the discrete Kalman filter (DKF) to initial condition discrepancies. *J. Geophys. Res.*, 2001, **106**: 17939~17952
- [10] Van Loon M, Builtjes P J H, Segers A. Data assimilation of ozone in the atmospheric transport chemistry model LOTOS. *Environmental Modeling and Software*, 2000, **15**: 603~609
- [11] Heemink A W, Segers A J. Modeling and prediction of environmental data in space and time using Kalman filtering. *Stochastic Environmental Research and Risk Assessment* (SER-RA), 2002, **16**: 225~240
- [12] 陈军明, 徐大海, 朱蓉. 遗传算法在点源扩散浓度反演排放源强中的应用. *气象*, 2002, **28** (9): 12~16
Chen Junming, Xu Dahai, Zhu Rong. Application of genetic algorithms to point-source inversion. *Meteorological Monthly* (in Chinese), 2002, **28**(9): 12~16
- [13] 苏芳, 邵敏, 蔡旭辉, 等. 利用逆向轨迹反演模式估算北京地区甲烷源强. *环境科学学报*, 2002, **22**: 586~591
Su Fang, Shao Min, Cai Xuhui, et al. Estimation of methane emissions in Beijing area using backward trajectory inversion model. *Acta Scientiae Circumstantiae* (in Chinese), 2002, **22**: 586~591
- [14] Zhu J, Zeng Q. A mathematical formulation for optimal control of air pollution. *Science in China* (Series D), 2003, **46** (10): 994~1002
- [15] Liu F, Hu F, Zhu J. Adjoint method for optimum planning of industrial pollutant sources. *Science in China* (Series D), 2005, **48** (8): 1270~1279
- [16] Liu F, Zhu J, Hu F, Wang E Y. An optimal, weather condition dependent approach to decision-support of emission control of urban air pollution. *Environmental Modeling and Software*, 2006 (accepted for publication)
- [17] Evensen G. The combined parameter and state estimation problem. 2005, manuscript submitted to *Computational Geosciences*. Also available online: <http://enkf.nersc.no/Publications/eve05a.pdf>
- [18] Reichel R H, McLaughlin D B, Entekhabi D. Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.*, 2002, **130**: 103~114
- [19] Jazwinski A H. *Stochastic Processes and Filtering Theory*. San Diego, Calif.: Academic Press, 1970
- [20] van Leeuwen P J, Evensen G. Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Wea. Rev.*, 1996, **124**: 2898~2913
- [21] Miller R N, Ehret L L. Ensemble generation for models of multimodal systems. *Mon. Wea. Rev.*, 2002, **130**: 2313~2333
- [22] Chen W, Bakshi B R, Goel P K, et al. Bayesian estimation via sequential Monte Carlo sampling: Unconstrained nonlinear dynamic systems. *Int. Eng. Chem. Res.*, 2004, **43**: 4012~4025
- [23] Bennett A F. *Inverse Methods in Physical Oceanography*. Cambridge University Press, 1992. 346pp