

用变分法修正黑河试验数据的初步研究*

马刚

王强

(中国气象局国家天气气象中心, 北京 100081)

(北京气象学院, 北京 100081)

邱崇践

PA A

(兰州大学大气科学系, 兰州 730000)

摘要 为了对黑河试验观测资料进行订正, 利用一个一维边界层垂直扩散模式和以拟牛顿极小化算法为基础的最优化方法建立了一个变分同化系统, 其中, 目标函数相对于控制变量的梯度用扰动法求取, 并利用模式资料和黑河试验若干时刻的观测资料进行了一些数值试验. 试验证明, 将初始场变量与模式误差系数同时作为控制变量加以调整, 能够有效调整初始场观测资料. 并且, 用变分同化方法能够较线性插值(时间、空间)方法更有效, 合理的修正黑河试验观测资料场的种种缺陷.

关键词: 变分同化; 边界层; 黑河试验; 扰动法

1 引言

黑河试验从1990年6月至1992年10月, 对欧亚大陆腹地典型的干旱气候地区黑河流域的沙漠、戈壁和绿洲不同下垫面的辐射过程、边界层过程、绿洲生物过程、沙尘辐射特征和干旱地区降水特征进行了观测, 获得了大量宝贵的资料. 但由于各种原因, 试验所获得的观测资料在时间上有不少间断, 同时, 一些资料也存在较大误差, 从而影响了对这些资料的进一步应用. 在边界层内, 各种要素场的时空变化都很剧烈, 插值分析和平滑的方法会产生较大误差. 另一方面, 边界层大气也遵循一定的物理规律, 即给定初始场后, 积分适当的大气模式就可以模拟出它们的演变过程. 这就是说, 有可能利用数值模式提供的信息来补充和修正观测资料的缺陷. 这就是当前数值天气预报中流行的变分四维同化的基本思想, 它既可以充分利用已有的物理规律(数值模式), 又能合理的使用时间演变的历史资料所包含的信息. 此时目标函数被定义为模式输出与观测的距离, 因此变分四维同化过程也可以看作是将模式输出与实际观测相拟合. 显然, 这种方法也可以用于解决黑河试验(HEIFE)资料的订正和插补工作. 本文将利用一个一维边界层数值模式, 用变分同化方法对黑河观测资料进行订正.

2 理论概述

2.1 变分四维同化方法

变分四维同化方法的基本原理是通过修正模式输入量(控制变量), 使在同化时段

2000-01-14收到, 2001-05-14收到修改稿

* 国家自然科学基金资助项目 49575249

内的模式输出量 (或是它们的某些函数) 与相应观测之间距离最小。因此, 我们将日标函数写成最小二乘形式, 将其定义为模式预报场与相应观测场之间的距离。于是, 变分同化方法便构成了一个约束条件 (数值模式) 下的泛函极小化问题, 通过计算出目标函数相对于控制变量的梯度, 然后运用某种极小化算法 (如共轭梯度法、拟牛顿方法), 使其达到极小点。

一个经空间离散化后的大气边界层数值模式方程可以写为下面的常微分方程组

$$\frac{dX}{dt} = F(X, p), \quad (1)$$

其中, X 是 N 维向量, 表示模式输出, 包含所有模式变量在所有格点上的值; p 是模式误差参数。

同时, 定义表征模式输出与相应观测之间距离的目标函数

$$J = \sum_{t_r=0}^n \{C[X(t_r)] - y(t_r)_{\text{obs}}\}^T W_1(t_r) \{C[X(t_r)] - y(t_r)_{\text{obs}}\}, \quad (2)$$

这里, n 是同化时段内有观测场的时间次数, $y(t_r)_{\text{obs}}$ 是观测场, 并满足 $y(t_r) = C[X(t_r)]$, C 表示模式输出场 $X(t_r)$ 到观测场 $y(t_r)$ 的映射, 这是因为观测场并不一定是模式变量。 $W_1(t_r)$ 是一个 $N \times N$ 维的权重系数矩阵, 表明目标函数中各项所占比重大小。在本次试验中, 控制变量 (模式输出场) $X(t_r)$ 为模式在积分时段内不同时刻 (r) 输出的模式变量 u, v, T, q , 观测变量为不同时刻 (r) 观测的大气状态 $u, v, T, q, r = 0, 1, \dots, n$ 。由于同化控制变量与观测变量相同, 这里, 目标函数可以表达为

$$J = \sum_{t_r=0}^n \{C_R[X(t_0)] - y(t_r)_{\text{obs}}\}^T W_1(t_r) \{C_R[X(t_0)] - y(t_r)_{\text{obs}}\}, \quad (3)$$

其中, $C_R[X(t_0)]$ 表示初始状态 $X(t_0)$ 经由模式方程 C 积分后在时刻 t_r 处的输出。

有 3 种选择控制变量的方案^[1], 方案一是以初始值 X_0 作为控制变量, 这相当于认为模式是精确的, 而初始场有误差; 方案二是以参数 p 作为控制变量, 这相当于认为初始场是精确的, 而模式有误差。实际上, 数值模式和初始场都存在误差, 因此方案三将参数 p 与初始场 X_0 同时作为控制变量。由于本文采用的是一个一维边界层垂直扩散模式, 与三维模式相比, 热力方程的主要差异是略去了水平平流项, 动力方程则还有气压梯度力项的省略。因此, 热力方程误差项可以表示为

$$\delta = V \frac{\partial X}{\partial S} = Vp, \quad (4)$$

这里, V 表示合成风速, $p (= \partial X / \partial S)$ 表示模式变量沿水平风的梯度。动力方程误差项为

$$\delta = V \frac{\partial X}{\partial S} + \alpha \frac{\partial p}{\partial S} = p, \quad (5)$$

这里, $\alpha \frac{\partial p}{\partial S}$ 表示沿水平风的气压梯度, 参数 p 表示模式变量空间水平梯度与水平气压

梯度力的合成效应。

在修正模式误差时, 取 p 作为控制变量, 并假设其不随时间变化。通过调整参数 p 使模拟场与观测场最接近, 即 J 最小, 但从物理上考虑, 还应参数施加必要的限制^[1]

$$|p_j - \bar{p}_j| \leq \delta p_j, \quad j = 1, 2, \dots, m \quad (6)$$

其中, \bar{p}_j 是 p_j 的估计值, δp_j 为参数 p_j 的允许误差。利用罚函数技术可以将这一带不等式约束的极小化问题转化为无约束极小化问题, 由此, 对于方案二、三, 只须将目标函数扩展为

$$L = J + \sum_{j=1}^m (p_j - \bar{p}_j)^T W_{2j} (p_j - \bar{p}_j), \quad (7)$$

选择适当的权重系数 W_{2j} , 即可保证式 (6) 的成立。

本文用准 Newton 型最优化方法^[1]求解上述极小化问题。此外, 由于一维模式所需的控制变量数量较少, 模式运行耗时也很少, 可以采用扰动法计算目标函数相对于控制变量的梯度。对如下形式的目标函数

$$J = J(X_1, X_2, \dots, X_n), \quad (8)$$

若给第 i 个控制变量加上一个扰动量 δX_i , 则第 i 个梯度分量可以按下式计算 (中央差分格式):

$$\frac{\partial J}{\partial X_i} \approx \frac{J(X_1, X_2, \dots, X_i + \delta X_i, \dots, X_n) - J(X_1, X_2, \dots, X_i - \delta X_i, \dots, X_n)}{2\delta X_i}, \quad (9)$$

这意味着若有 n 个控制变量时, 为计算一组梯度我们需用模式作 $2n$ 次模拟计算, 当模式中控制变量维数较高 (即 n 很大) 时, 其工作量是很大的。

2.2 模式和资料

试验采用的一维边界层垂直扩散模式方程是:

$$\frac{\partial X}{\partial t} = \frac{\partial}{\partial z} \left(K \frac{\partial X}{\partial z} \right) + S, \quad (10)$$

$X = X(u, v, \theta, q)$, 这里, 温度场采用位温形式, S 是所有剩余项, 令其形式同公式 (4)、(5)。为减少控制变量个数, 我们将模式层顶和层底的参数 p 作为控制变量, 用线性插值计算各层的参数值。模式中, 垂直扩散系数根据 Louis^[4]的参数化方案计算, 在模式上界 ($z = z_g + H$), 假设动量和感热通量为零, 在下边界 ($z = z_g$), 取

$$K_{mv} \frac{\partial u}{\partial z} = C_m |V_0| u_0, \quad (11)$$

$$K_{mv} \frac{\partial v}{\partial z} = C_m |V_0| v_0, \quad (12)$$

$$K_{\theta} \frac{\partial \theta}{\partial z} = C_\theta |V_0| (\theta_0 - \theta_g), \quad (13)$$

表面交换系数 C_m 和 C_θ 亦按 Louis 等的参数化方案计算; θ_g 是地表位温, 通过积分下垫面地温预报方程得到。

$$C_p \rho_p \frac{\partial T_p}{\partial t} = -K_g \frac{\partial^2 T_p}{\partial z^2}, \quad (14)$$

在计算地表有效辐射通量 R 时, 我们采用一个高度简化的未考虑云效应的辐射参数化方案。模式在垂直面划分为 12 层 (取微气象塔 4、10、15、25 m, 系留气球和低成本空仪器 50、75、100、200、400、600、800、1000 m 的观测资料)。取积分时段为 6 h, 时间步长为 60 s, 模式积分过程中每 2 h 输出一组模拟结果。本文取华首站 1991 年 8 月 15 日凌晨 3 时到夜间 19 时共 8 个时次的观测数据加以同化修正。

2.3 权重系数和尺度化量

目标函数中权重系数的选取根据资料的精度而定, 一个变量资料的精度愈高, 则它的可信度就愈大, 其权重系数的值也相应较大^[5]。这里, 我们给定目标函数中各项的误差估计, 以其倒数的平方作为相应的权重系数。

由于各个控制变量的量级差别可能很大, 若不对它们作预处理, 计算出的梯度量级差别也可能很大, 使目标函数在多维空间呈现极度扁的“椭圆球”, 造成目标函数难于下降。解决的办法是利用尺度化过程使各个控制变量或目标函数相对各个控制变量具有大体相当的量级。本文取尺度化因子 $S_i = X_{\max,i} / X_{\min,i}$ 。

3 数值试验

为检验同化的效果, 我们将同化后的预报结果与同化前的预报值作比较, 所有的偏差都是以均方根误差 (RMS) 的形式来表示的。实际资料的同化过程中, 无论是观测资料还是数值模式都存在不精确因素, 仅考虑观测资料的误差而忽略数值模式的误差, 或是仅考虑模式的误差而忽略观测资料的误差都是片面的, 必须同时考虑这两种误差。

表 1 模式变量的均方根差收敛情况的比较

	u ($m s^{-1}$)	v ($m s^{-1}$)	T (K)	q ($g kg^{-1}$)
迭代前	1.0	1.0	0.5	1.0
迭代后	0.23	0.15	0.39	0.2

表 2 模式误差参数的初猜值 (p_0), 真实值 (p) 和迭代值 (P_n) 的比较

模式层底	u	v	T	q
p_0	0.0004	0.0004	0.0004	0.000009
p	0.00001	0.0001	0.00001	0.0000006
P_n	-0.000003	0.000059	-0.00014	-0.00000043
模式层顶	u	v	T	q
p_0	0.005	0.005	0.005	0.00005
p	0.0001	0.0001	0.0001	0.000001
P_n	0.00017	0.00050	0.00025	0.0000037

在理想试验中, 我们将模式误差参数与初始场变量共同作为控制变量代入同化系统, 分别求取它们各自的梯度, 经过尺度化以后进入下降算法, 通过迭代求出它们的优

化结果。取迭代次数为 25 次。图 1 为目标函数在迭代中收敛的情形。经过迭代, 目标函数值收敛了一个量级。表 1、2 分别为模式变量的均方根差收敛情况及模式误差参数在调整前后的值与实际值的比较。这表明, 当初始场和模式同时存在误差时, 把初始场变量和模式误差同时作为控制变量的方法, 可以有效调整此二者的误差。综上所述, 将造成模式输出与观测场不相适应的所有不精确因素, 同时作为控制变量代入同化系统是我们的最佳选择。

4 实际资料试验

在实测资料的同化试验中, 由于测站的下垫面类型、地理环境与理想资料试验时截然不同, 这里我们首先依据化音站的实际地理、地表特征调整模式, 使非控制变量的模式参量与相应的实际值尽可能一致, 从而使模式输出与实测资料的距离尽可能减小。取扰动步长为 1×10^{-3} , 迭代次数为 25 次。

首先取模式初始场变量的初猜值与观测值相同, 这样, 在迭代过程中, 目标函数的定义(式 2)就要求模式在积分时段内的输出与观测场相拟合, 也就是说, 在修正初始场的同时, 由于引入了整个积分时段内测站上空边界层要素场的时间演变信息, 修正后的初始场剔除了与要素场时间演变无关的部分, 从而达到了修正初始观测资料的目的。同时, 考虑到模式自身存在不精确因素, 模式误差项的引入使模式能够考虑一些模式描述能力之外的因素(对一维垂直扩散模式而言, 如要素场的水平平流输送、水平气压梯度对要素场局地变化的影响等), 这就使得模式能够较为准确地描述测站上空边界层要素场的时间演变过程, 也就是使模式模拟场与观测场间的距离更小。图 2 给出了目标函数在迭代过程中收敛的情形。修正后的模式初始场与相应实测场的均方根差分别为 0.06 m s^{-1} , 0.31 m s^{-1} , 0.05 K , 0.1 g kg^{-1} 。由试验结果可知, 引入的模式误差与修正后的初始场共同作用, 使模式模拟场与实测场之间的距离减小了 50%。而均方根差的值表明, 初始场的修正幅度都在可以接受的范围之内(一般来说, 对于水平风速是

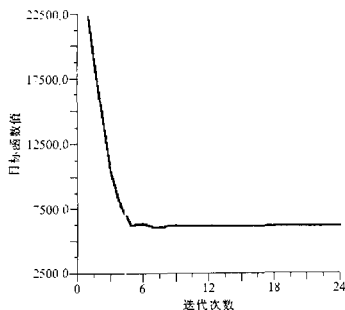


图 1 理想资料同化试验目标函数上的收敛情况

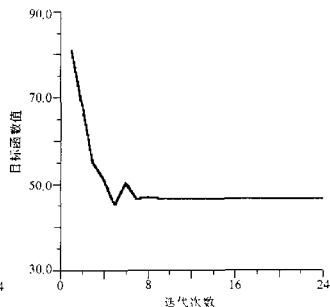


图 2 实测资料同化试验目标函数的收敛情况

1 m s^{-1} , 温度场是 0.5 K , 水汽场为 1.0 g kg^{-1}). 所有这些数据说明, 我们引入模式误差是合理的, 而在此基础上对初始观测场的修正是有效的. 若考虑某一模式层(如第6层)初始观测场资料缺失, 我们用线性内插得到的值作为初猜场代入同化系统, 将迭代后的结果与线性插值作比较(表3).

表3 模式层第6层观测场, 线性插值与同化结果的比较

	u (m s^{-1})	v (m s^{-1})	T (K)	q (g kg^{-1})
观测场	3.35	-2.53	306.389	3.8
线性插值	3.19	2.78	306.022	4.1
同化结果	3.22	-2.79	306.017	4.0

可以看出, 尽管对部分变量(如水平风的 v 分量, 温度场)同化结果比线性插值差, 但以线性插值结果和同化结果分别作为模式初始场积分, 其各自在积分时段内的模式输出与实测场的均方根差比较如图3所示. 由此可见, 初始观测场经过同化系统修正后剔除了与其历史演变不相关的信息, 使修正后的初始场在模式积分时与实测场的距离要比用线性插值结果更小. 另一方面, 线性插值结果仅仅是从数学上考虑了缺失模式层相邻观测资料对其的影响. 在变分同化系统中, 我们定义了如式(2)的目标函数, 这

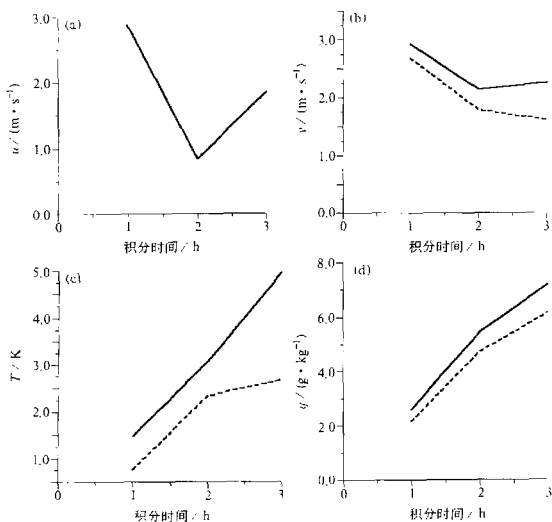


图3 不同初始场模式变量的模拟与实测资料的均方根差比较
(a) 水平风速 u ; (b) 水平风速 v ; (c) 温度 T ; (d) 比湿 q
实线: 以空间线性插值结果为初始场, 虚线: 以同化资料为初始场

就从时间(模式模拟与实测场的距离)、空间(以线性插值作为初猜场)两个方面同时考虑了其他观测资料对缺测模式层的影响,从而填补缺测资料的空白。显然,这种方法更具有其合理意义。

对在某一时刻无任何观测资料的情况,如1991年8月15日07时的观测场缺测,令其由05时和09时的观测场作时间线性内插得到。表4是时间线性内插结果和同化得到的要素场分别与实测场均方根差的对比结果。

表4 时间内插、同化结果与实测场均方根差的比较

	$u / (\text{m s}^{-1})$	$v / (\text{m s}^{-1})$	T / K	$q / (\text{g kg}^{-1})$
线性插值	0.85	1.12	0.70	0.45
同化结果	0.05	0.70	0.10	0.54

这里,在同化系统中我们以时间内插的要素场作为初猜场。可以看出,同化后的边界层要素场与实测场的距离更小,并且,其均方根差值也在允许的误差范围之内。对比时间、空间分别作线性内插的要素场,同化系统同时考虑了边界层要素场时空演变的历史信息,而这正是其他插值方法所无法兼顾的。因此,用变分同化系统修正观测资料的误差,填补观测资料的缺测信息是一种行之有效的办法。

5 结果与推论

本文对利用变分同化系统修正黑河试验观测资料的方法作了初步研究。由于数值模式能够在一定程度上描述大气边界层物理场时空演变过程,我们以实测场作为模式初始场,通过利用变分同化技术调整模式初始场,减小模式模拟场与积分时段内相应观测场的距离来修正初始观测场。试验证明,这种方法能够对初始观测场作出合理修正。但是,由于模式不能精确描述测站上空边界层要素场的时间演变过程,引入模式误差使模式能够在一定程度上考虑模式描述能力之外的一些因素,将其作为控制变量与初始场共同调整,可使模式能够较原来更精确的描述边界层物理场的时间历史演变,减小模式模拟场与实测场之间的距离,最终达到修正初始场观测资料的目的。试验结果表明,无论是理想试验还是实测资料的同化试验,采用同时反演初始场和模式误差的方法能够有效调整初始场观测资料。

本文将同化后的边界层要素场与用时间、空间线性插值得出的资料场分别进行了比较,结果证明,由于变分同化系统同时考虑了测站上空边界层要素场时间、空间演变的历史信息,用变分同化系统修正的初始观测场与实测场的距离更小,即用变分同化方法能够较线性插值方法更有效,合理的修正黑河试验观测资料的种种缺陷。

本文采用的一维垂直扩散模式,具有模式变量少,运算占用机时少,同化效果较好等优点。但是模式参数化方案过于简单,使模式模拟场与实测场的距离较大,难以较为精确地描述边界层大气要素场的时间演变,因而对初始场的修正有限。理想资料同化试验表明,若模式能够精确描述积分时段内边界层大气的运动,则同化系统就能对初始观测场作出更为精确的修正。

参 考 文 献

- 1 郁吉东, 丑纪范, 数值天气预报中的两类反问题及一种数值解法——理想试验, 气象学报, 1994, 52(2), 129~137.
- 2 余金香, 邱崇践, 用数值模拟与观测资料相拟合的方法构造边界层流场, 应用气象学报, 1997, 8(3), 302~308.
- 3 马刚, 邱崇践, 王强, 用变分法修正黑河试验数据的初步研究, 气象学报, 2001, 59(1), 107~114.
- 4 Louis, J. A parametric model of vertical eddy fluxes in the atmosphere, *Boundary-Layer Meteorology*, 1979, 17, 187~202.
- 5 龚建东, 邱崇践, 变分四维同化方法的研究, 兰州大学大气科学系硕士学位论文, 1996.

A Research to Validate the Data from Heihe River Field Experiment by Variation

Ma Gang

(National Satellite Meteorological Center, China Meteorological Administration, Beijing 100081)

Wang Qiang

(Beijing Meteorology College, Beijing 100081)

Qiu Chongjian

(Department of Atmospheric Science, Lanzhou University, Lanzhou 73000)

Abstract To modify the observational data from the Heihe River Field Experiment (HEIFE), a variational assimilation system is set up based on a vertical diffusion model in the boundary layer and a quasi-Newton minimization algorithm. The gradients of the cost function with respect to control variables are calculated in the way of perturbation. The results of some experiments by using both model data and some observation data in the HEIFE show that it can modify the initial fields (the atmosphere state) efficiently while making the initial fields and the coefficients of model error as control variables simultaneously. Compared with the linear interpolations (included in time and in space), variational assimilations can improve the defect in the data obtained in the HEIFE.

Key words: variational assimilation; boundary layer; Heihe River Field Experiment; perturbation