

朱国富. 2015. 数值天气预报中分析同化基本方法的历史发展脉络和评述. 气象, 41(8):986-996.

数值天气预报中分析同化基本方法的历史发展脉络和评述^{*1}

朱国富

国家气象中心, 北京 100081

提 要: 数值天气预报中分析同化的基本方法先后经历了多项式函数拟合方法、逐步订正方法、最优插值方法、变分方法和集合卡尔曼滤波方法。本文首先根据相关的经典文献力求本色地介绍这些方法的基本思想和实施的具体要点;然后,着重于它们的上下承接关系,试图阐述同化的历史发展脉络,评述这些方法的显著特征和创新性,以期清晰地理解资料同化的循序渐进的内在发展逻辑。此外,从起源上阐明“主观分析”与“客观分析”、“初猜场”与“背景场”、“分析”与“同化”、以及“更新”、“新息”等基本概念,以期准确地理解和把握“大气资料同化”的由来和内涵。

关键词: 数值天气预报, 客观分析和资料同化, 逐步订正方法, 最优插值方法, 变分方法, 集合卡尔曼滤波方法

中图分类号: P456

文献标志码: A

doi: 10.7519/j.issn.1000-0526.2015.08.008

Remarks on Development of Basic Methods of Atmospheric Data Assimilation for Numerical Weather Prediction

ZHU Guofu

National Meteorological Centre, Beijing 100081

Abstract: Main basic methods of analysis and assimilation for numerical weather prediction (NWP) has gone through the polynomial function fitting method, the successive correction method (SCM), the optimum interpolation (OI), the variational method (Var) and the ensemble Kalman filter (EnKF). This article tries to present a concise and true description with their basic ideas and implementation approaches based on original classical references to these methods. Then remarks on their distinct characteristics and innovative development are made, and it is highlighted, from these remarks, that we can see a step-by-step expansion of useful information and cutdown on limitations along the historical progress of atmospheric data assimilation, which is clearly endowed with a fore-and-aft connection and a cognitively simple internal logic of development. Also, an effort is taken to precisely understand the concept of “atmospheric data assimilation”, along its historical progress and on basis of some essential terms “subjective analysis” and “objective analysis”, “first guess field” and “background field”, “analysis” and “assimilation”, “update”, and “innovation”.

Key words: numerical weather prediction (NWP), objective analysis and data assimilation, successive correction method (SCM), optimal interpolation (OI), variational methods (Var), ensemble Kalman filter (EnKF)

* 公益性行业(气象)科研专项(GYHY201206007)和国家自然科学基金面上项目(41175034)共同资助

2015 年 1 月 8 日收稿; 2015 年 5 月 5 日收修定稿

作者:朱国富,主要从事数值天气预报大气资料同化的方法研究和系统开发. Email: zhugf@cma.gov.cn

引 言

数值天气预报(Numerical Weather Prediction, NWP)的质量和准确性的巨大提高,是 20 世纪下半叶所有科学分支中的主要成就之一(雅罗,2004)。

在具体的日常天气预报中,数值天气预报已经成为现代天气预报业务的基础,也是提高天气预报准确率和服务水平的核心技术和最根本的科学途径。从整体的现代气象业务高度而言,数值天气预报的水平集中体现综合气象观测、通信网络、高性能计算机、大气科学理论创新等科技能力,因此是反映一个国家气象现代化水平和在国际气象界中地位的重要标志。

数值天气预报原初是数学物理方程中的初值问题(Bjerknes,1904),即在大气初始状态已知的条件下通过数值方法求解反映支配大气运动的物理、化学规律的微分方程组来预报未来的大气状态。它包含两个主要过程:(1)用初值形成方法得到大气的初始状态(称为初始场,或初始条件);(2)然后积分反映大气运动规律的数值天气预报模式到未来时刻,得到该时刻的天气预报。对应业务实施上,一个具体的数值天气预报系统有两个主体部分:形成大气初始状态的大气资料分析同化系统和反映大气运动规律的预报模式系统;同时,还包含观测资料的获取和预处理、预报产品的后处理和输出、模式产品检验评估、归档和解释应用等不可或缺的组成部分。

NWP 的初值形成方法先后经历了主观分析方法和客观分析方法。在 NWP 发展初期,科学先驱们,如 1922 年 Richardson、1950 年 Charny 等和我国 20 世纪 50 年代顾震潮先生,是通过主观分析方法由某一时刻观测资料得到该时刻预报模式网格上的值(用作初始场)。具体过程是:首先进行观测资料的填图和手工分析绘制等值线图,然后在等值线图上通过主观的人工近似插值得到所给定的规则格点上的值(Panofsky,1949)。由于这个过程所包含的填图、天气图分析和插值是人工干预和主观的,会因人而异,所以这种方法被称为主观分析方法。由于主观分析方法的手工耗时和人为主观性,促使客观分析方法应运而生。客观分析方法是按照某种算法通过计算机程序自动化地(无需人工干预)由已有

的不规则分布的观测资料(并结合已知的大气运动规律的信息)得到某一时刻、某一预定规则网格点上最可能的值。所得到的这一规则网格点上最可能的值是表示该时刻模式大气状态的尽可能准确的值,被称为“分析场”,常常又被简称为“分析”。为了语义上明确和避免可能的混淆,下文中将分析同化的实现过程简称为“分析”,而分析同化的所得结果称为“分析场”。

大气资料同化源于 NWP 初值形成的客观分析。分析同化的基本方法(国际上主要业务 NWP 中心所使用)先后经历了多项式函数拟合方法、逐步订正方法(Successive Correction Method, SCM)、最优插值方法(Optimal Interpolation, OI)、变分方法(Variational methods, Var)和集合卡尔曼滤波方法(Ensemble Kalman Filter, EnKF)。多项式函数拟合方法由 Panofsky(1949)提出,是最早的客观分析方法。SCM 由 Bergthorsson 等(1955)提出,后来 Cressman(1959)将它发展为一个业务客观分析方案,用于当时美国联合数值天气预报(JNWP)部门。OI 由 Gandin(1963)首先全面地研制和开发,并应用到苏联的客观分析中;Bergman(1979)详细地描述了当时美国国家气象中心(NMC)业务使用的 9 层全球预报系统的 OI 方案;Lorenc(1981)详细地描述了当时欧洲中期天气预报中心(ECMWF)业务使用的 OI 方案;在 20 世纪 80 年代和 90 年代初,OI 成为业务分析方案的主流选择。Var(Lorenc,1986)是目前国际上大多数主要业务 NWP 中心的业务分析方案(Parrish et al,1992;Courtier et al,1998)。EnKF 是 1994 年由海洋学者 Evensen(1994)引入到资料同化领域,目前已成为国际上同化领域的发展新动向和热点研究课题;Hamill(2006)系统地描述了基于样本集合的大气资料同化方法。

大气资料同化发展到今天,它的目的被 Talagrand(1997)表述为“使用所有可用的信息,尽可能准确地估计大气运动的状态”。一般地,它的数学公式表示为通过观测信息本身和 NWP 短期预报的统计结合产生分析场:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{W}[\mathbf{y}_o - H(\mathbf{x}_b)]$$

式中, \mathbf{x}_a 表示分析场, \mathbf{x}_b 表示背景场(现在一般由 NWP 短期预报场提供), \mathbf{y}_o 表示观测; H 是观测算子(也称为向前观测算子),它给出了大气状态物理量和

观测物理量之间的具体确定联系; W 是统计权重。

关于大气资料同化方法的介绍已有许多精彩和十分有用的论文和专著,如: Lorenc(1986)“NWP 的分析方法”, Daley(1991)“大气资料分析”, Talagrand(1997)“观测资料同化: 一个导论”, Kalnay(2003)“大气模式、资料同化和可预报性”, 邹晓蕾(2009)“资料同化理论和应用(上册)”, Bouttier 等(1999)“资料同化概念和方法”的 ECMWF 讲习班培训材料。上述论著重在阐述各分析同化方法的特点和不同方法之间的横向关联; 如经典文献 Lorenc(1986), 基于所导出的 NWP 最佳分析的理想化方程, 讨论了各种 NWP 分析方法及其优缺点, 说明了各方法在公式和求解上的相互关系。

本文着重于上下承接的纵向联系和分析同化的历史发展脉络, 根据相关的经典文献, 力求本色地介绍分析同化基本方法的基本思想和做法要点; 在此基础上, 评述这些方法的显著特征和创新性。

1 基本方法的发展进程

1.1 多项式函数拟合方法

Panofsky(1949)提出了多项式拟合方法。它的思想是: 假定气象变量和三维空间坐标存在着解析函数的客观联系, 那么就能无需人工干涉地将代码报文的观测数据转变为预报模式的初值条件, 即完成客观分析。于是, 这一客观分析问题变成“找到一个解析函数 $p(x, y, z)$, 用它来表示气象变量 p 的三维空间坐标 (x, y, z) 的分布”。在具体实施中, 类似于在某一等压面上的主观分析, 客观分析的垂直维数为某一固定常值, 且水平面上限定在相对小的区域, 这样, 二维的三次多项式函数在该区域可以令人满意地拟合观测。一个三次多项式的一般形式写为:

$$p(x, y) = \sum_{i,j} a_{i,j} x^i y^j, \quad i + j \leq 3,$$

系数 $a_{i,j}$ 有 10 个 (1)

所限定的拟合区域大小取决于需要平滑观测的程度。如果所限定的区域只容纳 10 个观测数据, 那么这些数据能够按照上面三次多项式准确拟合观测, 也因此没有对观测的平滑。然而, 由于所有观测受到观测误差和可能的局地涡旋影响, 不应该完全准确地拟合观测, 而需要一些平滑。其实, 也不可能

指望所确定的多项式能拟合这些涡旋, 而是能够表示出所分析的气象变量场的大尺度特征。

平滑的程度取决于所分析的气象变量场的性质。观测值比较不准或易受非代表性涨落的变量(如风), 则应当比观测值较准的变量(如气压)进行更多平滑。对于气压场的分析, 可能使用 12~14 个观测数据来确定三次多项式的 10 个系数; 对于风场, 则要求不少于 20 个。

从 M 个观测数据确定 N 个系数 ($M > N$) 通常根据最小二乘方法来处理; 假定观测误差和局地涡旋是关于平滑的场正态分布, 那么这样的处理会得到最可信的多项式。

多项式拟合方法, 除了上述基本思想和做法要点, 还需考虑不同拟合区域的拼接等处理工艺。

1.2 逐步订正方法(SCM)

客观分析方法一开始提出时所要解决的问题是: 使用计算机将不规则分布的观测资料通过插值得到某一预定的规则网格点上最可能的值。Bergthorsson 等(1955)对这个问题的调查研究所得出的结论是, 仅仅通过天气观测资料的插值得到合理的分析场常常是不可能的。十分清楚, 得到合理的分析场要求观测站之间的距离必须小于所要分析的天气系统的尺度。而在许多区域, 如海洋, 不能满足这样的条件。这时任何插值方法, 不论是线性的、二次的、还是三次的, 都将失败。然而, 如果 12 小时前在这样的区域有一些观测资料, 那么一个 12 小时的预报场很可能是比用插值得到的分析场更好的近似。所以把预报场用于分析是保持时间连续性的一个相当自然的方式。

Panofsky(1949)提出的最早客观分析方法是由最小二乘原则用某些多项式拟合观测。这个方法本来就是对于观测资料相当稠密的区域设计的, 用于半球区域不是太适用; 半球区域的很多地方是大片零星散布的观测, 最小二乘多项式拟合方案在这样的区域往往会出现某类不稳定(Cressman, 1959)。

SCM 的基本思想是引入初猜场(an initial guess), 来处理这些出现问题, 改善分析场质量。Cressman 逐步订正方案(Cressman, 1959)的做法要点是引入初猜场之后, 使用观测资料对其进行逐步订正; 对于某一格点上所要分析的要素量 x , 订正的

公式形式是:

$$x_a = x_g + C \quad (2)$$

式中, x_a 、 x_g 和 C 分别是 x 的分析值、初猜值和由观测资料计算的订正值。

这是一个单点分析方案, 逐个格点地根据观测资料来订正该格点的初猜场。而且, 经过逐步序列的数轮“扫描”来完成分析。第一轮扫描的初猜值 x_g (第一次初猜值) 可以由预报值 (如 NWP 的 12 小时预报)、气候值、或是两者混合提供。第一次初猜值的选用是考虑与最终分析值尽可能准确的近似, 这样能以最小的计算开销得到最好的分析精度。

观测订正值 C 是与分析格点邻近的所有有关观测资料对该格点 x_g 的订正之和。以 500 hPa 高度分析为例, 某一邻近测站的有关观测资料被考虑了三种情况: 只有高度的观测、同时有高度和风的观测、和只有风的观测; 对某一格点, 由以该格点为圆心、 N 为影响半径范围内的所有测站的所有高度和风的观测资料, 得到这些观测资料对该格点的订正值 C :

$$C = \frac{A \sum C_h + \sum C_v}{An_h + n_v}$$

式中, (1) C_h 是只有高度观测时对该格点的订正值, $C_h = W(D_o - D_g)$; (2) C_v 是同时有高度和风的观测时的订正值, $C_v = W[D_o + \frac{kf}{mg}(v_o \Delta x - u_o \Delta y) - D_g]$; (3) 只有风的观测时, (2) 中的订正值 $C_v = W[\frac{kf}{mg}(v_o \Delta x - u_o \Delta y)]$ 。式中, W 是权重因子, D_o 是测站上的高度观测值, D_g 是插值到测站位置的高度初猜值, u_o 和 v_o 是风观测的 x 和 y 方向的分量, f 是科氏参数, g 是重力加速度, m 是地图投影放大系数, Δx 、 Δy 是观测点到格点的地图投影距离的 x 和 y 方向的分量, 因子 k 表示地转风对于实际风的平均比率。风观测对高度的订正考虑了地转关系, 使用表达式:

$$\delta\Phi = \delta(gz) = \frac{\partial\Phi}{\partial x}\delta x + \frac{\partial\Phi}{\partial y}\delta y = fv_g\delta x - fu_g\delta y$$

式中, Φ 为位势, u_g 和 v_g 是地转风的 x 和 y 方向的分量。订正值 C 表达式中的 n_h 和 n_v 分别是订正值 C_h 和 C_v 的总个数; A 是加权因子, 取决于高度初猜场的横向梯度的质量, 如果期望更接近风场的拟合, A 取小值。

如果用矢量表示所有网格点上的初猜场的值: $\mathbf{x}_g = (x_1, x_2, \dots, x_{N_x})^T$, 对于某一轮扫描, 式(2)可以写为一般的形式 (未考虑加权因子 A):

$$x_a(i) = x_g(i) + \frac{\sum_{k=1}^n \{W(k, i)[y_o(k) - x_g(k)]\}}{\sum_{k=1}^n W(k, i)} \quad (3)$$

$$x_g(k) = H(\mathbf{x}_g)$$

式中, i 表示格点位置, k 表示观测点位置; H 是由初猜场的值 \mathbf{x}_g 得到 k 测站位置上模拟观测值 $x_g(k)$ 的观测算子, 这里只是空间插值算子; $[y_o(k) - x_g(k)]$ 是 k 测站位置上观测值和模拟观测值之差, 称为观测增量; $W(k, i)$ 是 k 测站对 i 格点的权重函数, 由经验给定:

$$W(k, i) = \max\left[0, \frac{N^2 - d^2(k, i)}{N^2 + d^2(k, i)}\right]$$

这里, $d(k, i)$ 是 k 测站和 i 格点之间的距离, N 是权重因子 W 取 0 值的最小距离, 它被称为影响半径, 即: 由观测资料对格点的订正只考虑距离该格点半径 N 范围内的那些测站 ($k=1, \dots, n$) 的影响; 对于某一轮扫描, N 的取值是固定的。

Cressman 逐步订正方案的“逐步”特征是使用更新的初猜场和逐步减小的半径 N 来进行逐步序列的数轮“扫描”。由于大气中高度变量和风变量的不同性质, 第一轮扫描时半径 N 取值较大, 只考虑高度观测的订正, 随后的各轮扫描才同时考虑高度观测和风观测的订正; 由于半径 N 逐轮减小, 因此各轮扫描使用了不同的观测资料。

由式(3)可知, 对于很小的影响半径, 如果观测位于格点上, 那么分析值收敛于观测值, 会导致一个“轻信观测”的分析; 如果观测资料有严重过失误差, 这能导致分析场中出现“公牛眼”(围绕一个不真实的格点值的许多等值线)。因此, 剔除这样观测的质量控制是非常重要的环节。

1.3 最优插值方法(OI)

和 SCM 一样, OI 实质也属于单点分析方案, 公式形式也和式(3)相近。OI 的基本思想是以分析场误差方差最小为标准来确定观测增量的统计最优权重。

OI 的做法要点 (Bergman, 1979; Kalnay, 2005) 首先是引入先验信息的背景场, 其值用 \mathbf{x}_b 表示, 则分析场的值是观测增量对 \mathbf{x}_b 的加权订正, 公式形式

是:

$$\begin{aligned} x_a(i) &= x_b(i) + \sum_{k=1}^n \{W(k,i)[y_o(k) - x_b(k)]\} \\ x_b(k) &= H(x_b) \end{aligned} \quad (4)$$

之后,考虑背景场误差、观测误差以及分析场误差,有以下表示:

$$\begin{aligned} x_b(i) &= x_t(i) + \epsilon_b(i) \\ x_b(k) &= x_t(k) + \epsilon_b(k) \\ y_o(k) &= x_t(k) + \epsilon_o(k) \\ x_a(i) &= x_t(i) + \epsilon_a(i) \end{aligned}$$

式中, $\epsilon_b(i)$ 和 $\epsilon_b(k)$ 分别是格点 i 和测站 k 的背景场误差, $\epsilon_o(k)$ 是测站 k 的观测误差, $\epsilon_a(i)$ 是格点 i 的分析误差, x_t 表示对应的真值;因为 $\epsilon_b(k)$ 是在测站 k 位置的背景场误差,因此含有观测算子 H 。将这些表示代入式(4),并假定背景误差和观测误差无偏: $E[\epsilon_b]=0$ 和 $E[\epsilon_o]=0$,以及背景误差和观测误差无关: $E[\epsilon_b\epsilon_o]=0$,求分析误差方差最小,即 $\min\{E[\epsilon_a(i)\epsilon_a(i)]\}$,就得到由此确定统计最优权重的线性方程组:

$$\begin{aligned} \sum_{k=1}^n \{W(k,i)[b(k,k_o) + r(k,k_o)]\} \\ = b(k_o,i), \quad k_o = 1, \dots, n \end{aligned}$$

式中, $r(k,k_o)$ 是测站 k 和测站 k_o 间的观测误差协方差, $b(k,k_o)$ 是 k 和 k_o 间的背景误差协方差, $b(k_o,i)$ 是 k_o 和格点 i 间的背景误差协方差;涉及测站位置上的 b 含有在求方差最小中产生的观测算子 H 的切线性近似。

最后,由这个统计最优权重 $W(k,i)$,一次性地将所计算的观测增量加权订正(即分析增量 $\delta x_a(i)$)加到背景场的值,就得到分析值;也因此不需要SCM的逐步的迭代过程。

Lorenc(1981)描述了当时在ECMWF实施的业务OI方案,它不是通过逐个格点的逐个变量求解,其新颖性是一个有限空间体积内的格点和变量,建立并求解OI的分析方程组,使得大量资料的使用成为可行。这个方案的OI分析方程(Kalnay,2005)是对应式(4)的向量和矩阵形式:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{W}[\mathbf{y}_o - H(\mathbf{x}_b)] \quad (5)$$

式中, $\mathbf{W} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$ 是使得分析误差协方差极小的最优权重矩阵,也称为增益矩阵 \mathbf{K} 。这里,

上标 T 和 -1 分别表示矩阵的转置和矩阵的逆(下同), $\mathbf{B} = E[\boldsymbol{\epsilon}_b\boldsymbol{\epsilon}_b^T]$ 为背景场误差协方差(当模式预报场用作背景场时,背景场误差协方差就是预报误差协方差,通常表示为 \mathbf{P}_b), $\mathbf{R} = E[\boldsymbol{\epsilon}_o\boldsymbol{\epsilon}_o^T]$ 为观测误差协方差,而 $\boldsymbol{\epsilon}_b$ 和 $\boldsymbol{\epsilon}_o$ 分别为背景场误差向量和观测误差向量。 \mathbf{H} 是非线性观测算子 H 的切线性算子; $\mathbf{B}\mathbf{H}^T$ 是格点位置和测站位置间的背景误差协方差, $\mathbf{B}\mathbf{H}^T$ 是测站位置相互之间的背景误差协方差。此外,观测增量 $\mathbf{d} = \mathbf{y}_o - H(\mathbf{x}_b) = \mathbf{y}_o - H(\mathbf{x}_t + \mathbf{x}_b - \mathbf{x}_t) = \mathbf{y}_o - H(\mathbf{x}_t) - \mathbf{H}(\mathbf{x}_b - \mathbf{x}_t) = \boldsymbol{\epsilon}_o - \mathbf{H}\boldsymbol{\epsilon}_b$ 。

最优权重矩阵 \mathbf{W} 的等式可以重新写为: $\mathbf{W}(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}) = \mathbf{B}\mathbf{H}^T$;通过求解这个线性方程组,得到 \mathbf{W} ;再代入式(5),得到分析向量 \mathbf{x}_a 。

此外,最优权重矩阵 \mathbf{W} 包含观测误差协方差 \mathbf{R} ,这使得分析能够考虑观测质量特征。由于气象观测网是由若干不同的观测系统组成的,每个观测系统有着自己特有的观测质量特征,因此OI通过最优权重矩阵 \mathbf{W} ,以合理和分类的方式解决了各观测系统观测质量特征的不同(Bergman,1979)。这也体现了“最优”的另一层意义。

1.4 变分方法(Var)

Var的要点是(Lorenc,1986):把分析问题看作给定观测 \mathbf{y}_o 、已知观测量 \mathbf{y} 和模式状态变量 \mathbf{x} 的关系 $\mathbf{y} = H(\mathbf{x})$ 、考虑观测误差和模式状态变量的先验背景信息及其误差来求得 \mathbf{x} 的最佳值 \mathbf{x}_a 的这一逆问题。为此,首先基于贝叶斯定理(Bayes' theorem),在先验背景场误差和观测误差是不相关的假定下,导出给定观测 \mathbf{y}_o 条件下模式状态变量 \mathbf{x} 的后验概率密度函数:

$$\begin{aligned} P_a(\mathbf{x}) &= P(\mathbf{x} | \mathbf{y}_o) \propto P(\mathbf{y}_o | \mathbf{x})P(\mathbf{x}) \\ &= P_o[H(\mathbf{x}) - \mathbf{y}_o]P_b(\mathbf{x} - \mathbf{x}_b) \end{aligned}$$

即后验概率密度 $P_a(\mathbf{x})$ 正比于观测的概率密度 $P_o[H(\mathbf{x}) - \mathbf{y}_o]$ 和先验概率密度 $P_b(\mathbf{x} - \mathbf{x}_b)$ 的乘积,于是分析场的值 \mathbf{x}_a 是:

$$\mathbf{x}_a = \mathbf{x} \text{ 使得满足后验似然概率 } P_a(\mathbf{x})$$

$$\text{最大的解(即 } \mathbf{x} \text{ 的众数);} \quad (6)$$

之后,在背景场误差和观测误差无偏和满足正态分布的假定下:

$$P_b(\mathbf{x} - \mathbf{x}_b) \propto \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b)\right]$$

$$P_o[H(\mathbf{x}) - \mathbf{y}_o] \propto \exp\left\{-\frac{1}{2}[H(\mathbf{x}) - \mathbf{y}_o]^T \cdot\right.$$

$$\left. \mathbf{R}^{-1}[H(\mathbf{x}) - \mathbf{y}_o]\right\}$$

后验似然概率 $P_a(\mathbf{x})$ 最大的解便转化为等价于一个目标泛数 $J(\mathbf{x})$ 最小的解:

$$J(\mathbf{x}_a) = \min_x J(\mathbf{x}) \quad (7)$$

其中 $J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}[H(\mathbf{x}) - \mathbf{y}_o]^T \mathbf{R}^{-1}[H(\mathbf{x}) - \mathbf{y}_o]$, 于是通过泛函极小(即变分方法)的全局求解得到分析场的值 \mathbf{x}_a 。上式中背景场误差协方差 $\mathbf{B} = E[(\mathbf{x}_b - \mathbf{x}_t)(\mathbf{x}_b - \mathbf{x}_t)^T]$ (\mathbf{x}_t 表示大气状态的真值, $(\mathbf{x}_b - \mathbf{x}_t) = \boldsymbol{\varepsilon}_b$ 是背景场误差; 观测误差协方差 $\mathbf{R} = E[\mathbf{y}_o - H(\mathbf{x}_t)][\mathbf{y}_o - H(\mathbf{x}_t)]^T$, 其中 $[\mathbf{y}_o - H(\mathbf{x}_t)] = \mathbf{y}_t + \boldsymbol{\varepsilon}_{\text{ins}} - (H_t - \boldsymbol{\varepsilon}_H)(\mathbf{x}_t) = \boldsymbol{\varepsilon}_{\text{ins}} + \boldsymbol{\varepsilon}_H(\mathbf{x}_t)$ (\mathbf{y}_t 表示观测量的真值, 考虑 $H_t(\mathbf{x}_t) = \mathbf{y}_t$), 它包含了观测仪器误差 $\boldsymbol{\varepsilon}_{\text{ins}}$ 和观测算子误差 $\boldsymbol{\varepsilon}_H = -\boldsymbol{\varepsilon}_H(\mathbf{x}_t)$ (这里负号的选择是为了结果都是一致正号的形式上方便起见)。

需要指出, 对于 NWP 模式, 模式状态变量代表大气状态变量, 但由于离散化, “模式状态变量真值” \mathbf{x}_t 不是格点上的严格真实值, 只是格体平均意义上的值, 不能分辨次网格部分。这种由于模式状态变量不能表现次网格部分, 当它作用于观测算子时, $\mathbf{y}_o - H(\mathbf{x}_t)$ 还包含观测的代表性误差。

数学上求解式(7), 有: $\nabla_x J(\mathbf{x}_a) = \mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b) + \mathbf{H}^T \mathbf{R}^{-1}[H(\mathbf{x}_a) - \mathbf{y}_o] = 0$, 其中, $\mathbf{H} = \left. \frac{\partial H(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_a}$, 它是 H 的切线性算子。由于 \mathbf{H} 和 $H(\mathbf{x}_a)$ 都包含 \mathbf{x}_a , 因此不易显式求解, 可以通过迭代求解。当观测算子 H 在 \mathbf{x}_b 附近做线性化近似的情况下, 那么 $H(\mathbf{x}) \approx H(\mathbf{x}_b) + \mathbf{H}(\mathbf{x} - \mathbf{x}_b)$, 此时 \mathbf{H} 取在 $\mathbf{x} = \mathbf{x}_b$ 时的值, 相应的目标泛函写为:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}[\mathbf{H}(\mathbf{x} - \mathbf{x}_b) - \mathbf{d}]^T \mathbf{R}^{-1}[\mathbf{H}(\mathbf{x} - \mathbf{x}_b) - \mathbf{d}]$$

于是通过 $\nabla_x J(\mathbf{x}_a) = 0$, 便得到变分同化的一个显式解, 它在形式上和 OI 的式(5)相同。这被称作 Var 与 OI 在解的形式上的等价性。

在实际的业务实施中, 由于矩阵 \mathbf{B} 维数巨大和条件数很大而近于病态 (\mathbf{B} 是相邻行、列往往接近的稀疏对称矩阵) 等原因, 不能直接进行上述数学求

解, 而是先通过一系列变换优化目标泛函, 再通过最优优化算法求解优化后的泛函极小, 来得到 \mathbf{x}_a 。

1.5 集合卡尔曼滤波同化方法(EnKF)

由于求解的困难, OI 和 Var 都把 \mathbf{B} 当作不随时间变化的静态矩阵。EnKF 的基本思想是基于蒙特卡罗技术使用短期预报集合来估计动态演变的 \mathbf{B} 。EnKF 在概念上源于卡尔曼滤波同化方法(KF)和扩展卡尔曼滤波同化方法(ExKF: extended Kalman filter)。EnKF 的做法要点: 并不是显式地存储、访问和计算 \mathbf{B} , 而是使用短期预报集合分开计算增益矩阵中的 $\mathbf{H}\mathbf{B}\mathbf{H}^T$ 和 $\mathbf{B}\mathbf{H}^T$ (Hamill, 2006)。

1.5.1 卡尔曼滤波同化方法(KF)

KF 是按照逐时次的序贯更新方式进行客观分析。假定高斯的先验背景场和观测的误差分布, 考虑线性的预报模式和观测算子, 显式地使用动态的背景场误差协方差, 逐时次递推地融合背景场和观测, 在分析误差最小的准则下得到大气状态的最优估计以及该估计的误差协方差。在具体实施上, KF 包含合二为一的两个步骤: 更新步骤和预报步骤。

融合背景场和观测, 得到分析场, 这对应更新步骤。更新步骤的方程为:

$$\mathbf{x}'_a = \mathbf{x}'_b + \mathbf{K}(\mathbf{y}'_o - \mathbf{H}\mathbf{x}'_b) \quad (8)$$

$$\mathbf{K} = \mathbf{P}'_b \mathbf{H}^T (\mathbf{H}\mathbf{P}'_b \mathbf{H}^T + \mathbf{R}')^{-1} \quad (9)$$

$$\mathbf{P}'_a = \mathbf{K}(\mathbf{I} - \mathbf{K}\mathbf{H}')\mathbf{P}'_b \quad (10)$$

之后, 由分析场的值和误差协方差显式地预报下一时刻更新步骤所需要的背景场的值和误差协方差, 这对应预报步骤。预报步骤的方程为:

$$\mathbf{x}'_b^{t+1} = \mathbf{M}^{t \rightarrow t+1} \mathbf{x}'_a \quad (11)$$

$$\begin{aligned} \mathbf{P}'_b^{t+1} &= \mathbf{M}^{t \rightarrow t+1} \mathbf{P}'_a \mathbf{M}^{t \rightarrow t+1 T} + \mathbf{Q}^{t \rightarrow t+1} \\ &= \mathbf{M}^{t \rightarrow t+1} (\mathbf{M}^{t \rightarrow t+1 T} \mathbf{P}'_a) + \mathbf{Q}^{t \rightarrow t+1} \end{aligned} \quad (12)$$

式中, 上标 t 和 $t+1$ 分别表示 t 和 $t+1$ 时刻, 上标 $t \rightarrow t+1$ 表示时段; \mathbf{H} 表示线性观测算子, \mathbf{M} 表示线性预报模式, \mathbf{K} 是增益矩阵, \mathbf{R} 、 \mathbf{P}_b 和 \mathbf{P}_a 分别是观测、背景场和分析场的误差协方差, \mathbf{Q} 是预报模式误差协方差。这里, 为考虑式(8)的分析场误差, 对应 \mathbf{x}'_a 的真值 $\mathbf{x}'_t = \mathbf{M}^{t-1 \rightarrow t} \mathbf{x}'_t^{-1} + \boldsymbol{\varepsilon}_M^{t-1 \rightarrow t}$, 已假设模式误差 $\boldsymbol{\varepsilon}_M^{t-1 \rightarrow t}$ 是均值为零的白噪声过程; 模式误差协方差是 $\mathbf{Q}^{t-1 \rightarrow t} = E(\boldsymbol{\varepsilon}_M^{t-1 \rightarrow t} \boldsymbol{\varepsilon}_M^{t-1 \rightarrow t T})$; 背景场是对前期分析场模式积分的 t 时刻预报: $\mathbf{x}'_b = \mathbf{M}^{t-1 \rightarrow t} \mathbf{x}'_a^{-1}$; t 时刻预报的误差 $\boldsymbol{\varepsilon}'_b = \mathbf{x}'_b - \mathbf{x}'_t = \mathbf{M}^{t-1 \rightarrow t} (\mathbf{x}'_a^{-1} - \mathbf{x}'_t^{-1}) + \boldsymbol{\varepsilon}_M^{t-1 \rightarrow t}$

$= \mathbf{M}^{t-1} \boldsymbol{\varepsilon}_a^{t-1} + \boldsymbol{\varepsilon}_M^{t-1}$, 其协方差 $\mathbf{P}_b^t = E(\boldsymbol{\varepsilon}_b^t \boldsymbol{\varepsilon}_b^{tT}) = \mathbf{M}^{t-1} \mathbf{P}_a^{t-1} \mathbf{M}^{t-1T} + \mathbf{Q}^{t-1}$. t 时刻的观测误差 $\boldsymbol{\varepsilon}_R^t = (\mathbf{y}_o^t - \mathbf{H} \mathbf{x}_a^t) = \boldsymbol{\varepsilon}_o^t + \boldsymbol{\varepsilon}_H^t + \boldsymbol{\varepsilon}_r^t$, 它包含测量误差、观测算子误差和代表性误差。

KF 的式(8) \mathbf{x}_a 和 OI 的式(5) \mathbf{x}_a 有相同的形式, 且增益矩阵 \mathbf{K} 都是在背景场误差和观测误差由高斯白噪声序列组成的假定下满足分析场方差最小得到的。

作为输出, KF 不仅能够得到大气状态的估计 \mathbf{x}_a , 还能给出该估计的误差协方差 \mathbf{P}_a^t 以及后续的背景场误差协方差 \mathbf{P}_b^{t+1} . KF 的魅力就是通过式(12)显式地由模式积分和伴随来预报 \mathbf{P}_b^{t+1} , 这使得背景场误差协方差是依赖于天气形势而动态演变的, 以及通过式(11)显式地进行对分析场的模式积分预报来得到 \mathbf{x}_b^{t+1} , 这使得更新步骤和预报步骤成为紧密的一体。

1.5.2 扩展卡尔曼滤波同化方法(ExKF: extended Kalman Filter)

KF 只适用于线性系统, 即观测算子 \mathbf{H} 和预报模式 \mathbf{M} 都是线性的. ExKF 适用于非线性系统, 在方程的形式上, 只是需要对式(8)和(11)做些修改。

对于 ExKF 的更新步骤:

$$\mathbf{x}_a^t = \mathbf{x}_b^t + \mathbf{K}[\mathbf{y}_o^t - H(\mathbf{x}_b^t)] \quad (13)$$

对于 ExKF 的预报步骤:

$$\mathbf{x}_b^{t+1} = \mathbf{M}^{t+1}(\mathbf{x}_a^t) \quad (14)$$

式中, H 和 M 分别表示非线性观测算子和非线性预报模式。但是, 在含义上, 此时, 对应式(9)和(10)中的 \mathbf{H} 是 H 的雅可比矩阵: $\mathbf{H} = \frac{\partial H}{\partial \mathbf{x}}$; 对应式(12), 假定了误差增加的线性和假定了模式误差与分析误差通过切线性预报动力方程的增长无关, 且式(12)中 \mathbf{M} 是 M 的雅可比矩阵, $\mathbf{M} = \frac{\partial M}{\partial \mathbf{x}}$. \mathbf{M} 时常称作 t 时刻和 $t+1$ 时刻之间的转移矩阵(transition matrix); \mathbf{M}^T 是它的伴随。

1.5.3 集合卡尔曼滤波方法(EnKF: Ensemble Kalman Filter)

对于 N_x 维的状态向量 \mathbf{x} , \mathbf{P}_b 和 \mathbf{P}_a 是 $N_x \times N_x$ 维矩阵; 而且, 由式(12)可以知道, 它要求 $2 \times N_x$ 次地应用转移矩阵 \mathbf{M} 来预报误差协方差。因此, 对于 NWP 的高维的状态向量 \mathbf{x} 和复杂的预报模式 M , 卡尔曼滤波或扩展卡尔曼滤波由式(12)预报背景场

误差协方差矩阵时的数据存储量和计算量都是难以接受地巨大, 这使得 KF 和 ExKF 的实际实施难于实现。

EnKF 使用短期预报集合来估计背景场误差协方差。假定有一组 t 时刻短期预报集合来进行模式预报场误差的随机取样。这组预报集合表示为 \mathbf{X}'_f , 它是一个矩阵; 矩阵的列是集合成员的状态向量, 用 $\mathbf{x}'_f(k)$ 表示第 k 个集合成员 ($k=1, 2, \dots, N_E$), N_E 为集合成员数; 于是 $\mathbf{X}'_f = [\mathbf{x}'_f(1), \dots, \mathbf{x}'_f(N_E)]$ 。第 k 个成员对于集合平均的扰动是 $\mathbf{x}'_{f_j}(k) = \mathbf{x}'_f(k) - \overline{\mathbf{x}'_f}$, 其中集合平均 $\overline{\mathbf{x}'_f}$ 定义为 $\overline{\mathbf{x}'_f} = \frac{1}{N_E} \sum_{k=1}^{N_E} \mathbf{x}'_f(k)$; 把 \mathbf{X}'_f 定义为一个由扰动集合形成的矩阵, $\mathbf{X}'_{f_j} = [\mathbf{x}'_{f_j}(1), \dots, \mathbf{x}'_{f_j}(N_E)]$, 则由非线性模式预报的一个有限集合估计的背景场误差协方差为:

$$\begin{aligned} \mathbf{P}_b^t &\approx \frac{1}{N_E - 1} \sum_{k=1}^{N_E} [\mathbf{x}'_f(k) - \overline{\mathbf{x}'_f}][\mathbf{x}'_f(k) - \overline{\mathbf{x}'_f}]^T \\ &= \frac{1}{N_E - 1} \mathbf{X}'_{f_j} \mathbf{X}'_{f_j}{}^T \end{aligned} \quad (15)$$

EnKF 的具体实施中并不是需要显式地计算和存储式(15)的 \mathbf{P}_b^t , 而是使用短期预报集合分开计算卡尔曼增益矩阵 \mathbf{K} 中的 $\mathbf{H}'\mathbf{P}_b^t\mathbf{H}'^T$ 和 $\mathbf{P}_b^t\mathbf{H}'^T$ 。定义 $\overline{H(\mathbf{x}'_f)} = \frac{1}{N_E} \sum_{k=1}^{N_E} H[\mathbf{x}'_f(k)]$, 表示由作为背景场的预报场插值得到的模拟观测的平均值, 这样,

$$\begin{aligned} \mathbf{H}'\mathbf{P}_b^t\mathbf{H}'^T &= \frac{1}{N_E - 1} \sum_{k=1}^{N_E} \{H[\mathbf{x}'_f(k)] - \overline{H(\mathbf{x}'_f)}\} \cdot \\ &\quad \{H[\mathbf{x}'_f(k)] - \overline{H(\mathbf{x}'_f)}\}^T \end{aligned} \quad (16)$$

$$\begin{aligned} \mathbf{P}_b^t\mathbf{H}'^T &= \frac{1}{N_E - 1} \sum_{k=1}^{N_E} [\mathbf{x}'_f(k) - \overline{\mathbf{x}'_f}] \cdot \\ &\quad \{H[\mathbf{x}'_f(k)] - \overline{H(\mathbf{x}'_f)}\}^T \end{aligned} \quad (17)$$

由此计算 $\mathbf{K} = \mathbf{P}_b^t\mathbf{H}'^T(\mathbf{H}'\mathbf{P}_b^t\mathbf{H}'^T + \mathbf{R}^t)^{-1}$ 。

当然, 如果观测个数和模式状态向量维数一样大, 那么 $\mathbf{P}_b^t\mathbf{H}'^T$ 和 $\mathbf{H}'\mathbf{P}_b^t\mathbf{H}'^T$ 就与 \mathbf{P}_b^t 一样大, 式(16)和(17)的优势失效。不过, 一个可能的程序编码简化是序贯处理技术方法(serial processing): 如果观测误差独立不相关且与背景场误差无关, 则同时地同化所有观测与各个或各组序贯地同化这些观测, 能够产生同样的结果(Bishop et al, 2001)。

有了增益矩阵 \mathbf{K} , 则 EnKF 的分析场为:

$$\mathbf{x}_a^t(k) = \mathbf{x}'_f(k) + \mathbf{K}\{\mathbf{y}_o^t + \mathbf{y}'_o(k) - H[\mathbf{x}'_f(k)]\} \quad (18)$$

式中, $\mathbf{y}'_o(k)$ 代表 t 时刻的观测资料的集合扰动; $\mathbf{y}'_o(k) + \mathbf{y}''_o(k)$ 是“经过扰动后的观测”, 满足 $\mathbf{y}'_o(k) \sim N(0, \mathbf{R})$ 和 $\frac{1}{N_E} \sum_{k=1}^{N_E} \mathbf{y}'_o(k) = 0$ 。

由式(18)得到的是分析场集合 $\mathbf{X}'_a = [\mathbf{x}'_a(1), \dots, \mathbf{x}'_a(N_E)]$, 以此为初值积分预报模式, 得到下一个 $t+1$ 时刻预报场集合 \mathbf{X}'_{f+1} ; 再通过式(16)~(18)得到 $t+1$ 时刻分析场集合 \mathbf{X}'_{a+1} , 如此按照逐时次的序贯方式构成了分析更新和模式预报的循环。

2 基本方法的发展评述

2.1 各种基本方法的显著特征、深层意义和局限性

多项式函数拟合方法的最显著特征是在功能上实现了通过计算机程序来执行客观的天气图分析。它的局限性是: 所用的信息只是观测的值, 通过最小二乘方法用来确定拟合函数的系数。

SCM 的最显著特征是初猜值的引入; 这有着崭新的意义, 带来了客观分析在“有用信息”、“公式形式”和“观测信息使用方式”这三个方面的新突破: 首先, SCM 的输入信息不仅有观测值, 还有初猜值; 第二, SCM 的分析方程式(3)形成了之后的分析同化方法的统一的公式形式, 即分析值是观测增量对初猜值的加权订正; 第三, 是以观测增量的方式使用观测信息, 观测增量 $[\mathbf{y}_o(k) - H(\mathbf{x}_g)]$ 也称为新息(innovation), 它准确地表达了不是观测值本身或初猜值本身、而是测站位置上观测值和由初猜值模拟的观测值(观测相当量)这两者之差引入的“新的信息”。而且, SCM 还有两个深层的意义: 其一, 初猜值的引入原初是为了改善当时的分析质量, 其实也隐含着解决 NWP 分析问题欠定性的原初萌芽(将在另一篇文章中阐述)。其二, SCM 中逐步迭代的过程还蕴含着多尺度同化的原初萌芽: 由式(2)可知, 对每一轮扫描, 订正值 C 是影响半径 N 范围内的所有观测资料订正作用的平均, 这意味着在该半径范围内的一种平滑; 于是在使用大影响半径 N 的首次迭代之后的分析场反映大尺度特征, 再经过若干 N 逐轮减少的迭代之后它收敛于较小的尺度。因此, 使用半径 N 逐轮减小的序列扫描, 这意味着以多次重复使用观测的方式从中逐轮提取不同尺度的信息, 使得分析场包含不同的尺度谱; 大的半径 N 能订正初猜场的大尺度误差, 而最小的 N 值对应

着能够分析的最小尺度下限。SCM 的一个局限性是: 观测增量的权重函数是经验给定的。

OI 的显著特征, 首先是背景场概念的引入; 背景场的值和初猜值有着根本意义上的不同: 初猜值是计算数学意义下迭代算法的(尽可能准确的)输入值, 而背景场的值是估计理论意义下的先验估计值; 第二, OI 中不论作为输入的背景场和观测, 还是作为输出的分析场, 在信息的内含上, 都不只是值, 还包含误差及误差协方差, 这对于数据有用信息有着根本性意义上的转变(朱国富, 2015); 第三, 正是基于此, OI 中的权重函数是以“分析误差方差最小”的准则所确定的统计最优权重, 由此得到的分析场是该准则下的大气状态的最优估计; 第四, 进而, OI 分析方程式(5)有了一个观念上的突破: 把分析表述为“以(规则网格点上的)背景信息为平台, 统计最优地融合(不规则位置的)其他信息(观测资料), 得到(规则网格点上的)分析场”。OI 的一些局限性是: 由于 OI 限于求解权重的线性方程组的规模和实质上的单点分析方式, 观测算子 H 不能太复杂(一般是线性插值), 并存在观测资料的区域选择问题。

Var 的显著特征, 首先是由分析场似然概率最大的解转化为等价于一个目标泛数 $J(\mathbf{x})$ 最小的解; 第二, 通过变分求解的方式, 能够引入复杂的非线性观测算子而使得直接同化卫星、雷达等大量非常规、非模式变量的观测资料成为可能, 且执行全局求解、因而不存在观测资料使用的区域选择问题; 第三, 能够作为一个显式解直接导出分析方程的统一公式形式(5), 这个公式形式在 SCM 和 OI 中是经验性给出的。Var 的一个局限性是: Var 和 OI 一样, 只考虑了静态不变的背景场误差协方差矩阵 \mathbf{B} ; 静态矩阵 \mathbf{B} 的实质是包含真实 \mathbf{B} 矩阵平均意义上的一些非常重要的动力统计性质, 如运动场和质量场之间的地转平衡、质量场之气压和位势之间的静力平衡等约束关系; 对于某一时刻的大气状态, 这些统计性质只是一种平均意义上的近似, 与实际情形可能会有明显差距, 如在锋面附近。 \mathbf{B} 对于分析场的深刻影响(保障观测信息的适当权重和合理传播)和实际应用中变分方法对 \mathbf{B} 的处理将在另一篇文章中阐述。

EnKF 的显著特征, 首先是使用短期预报集合来考虑动态演变的 \mathbf{B} ; 由于短期预报集合来自 NWP 的集合预报, 所以它内含了模式大气运动(包含非线性)的结构特征和动力演变, 能够更接近真实 \mathbf{B} 矩

阵。第二,输出信息有了新内容:不仅能够得到大气状态的最优估计,还能给出该估计的误差协方差。第三,EnKF还有一个深层的意义,它将资料同化和集合预报自然地合二为一,揭示了这两者之间内在本质上的自然联系:集合预报的目的是估计随天气形势演变的预报不确定性,而资料同化要求准确估计预报不确定性来最优地融合先验的预报和新的观测;对应在实施上构成了这样的循环:预报的集合用来估计 EnKF 同化中作为背景场的预报误差统计;同化的输出一组分析场,用来提供集合预报的初值;从积分这组初值得到的短期集合预报又为下一次同化循环提供新的背景场误差统计。之前的 OI 和 Var 中更多是在概念的内涵上预报和分析构成循环,因此两者密切联系;但在实施的方式上两者并不是一定分不开的一体。EnKF 的一些局限性是:EnKF 在实施中存在合理取样的代表性问题(样本的独立性和随机性),有限取样会引入的误差,以及在实际应用中许多待研究的问题,如滤波发散问题、变量间的平衡问题、使用复杂观测算子的表现等。

需要指出,由于模式状态变量的巨大维数和观测资料的巨大数目,加之对大气运动规律认识的有限性,分析同化中的经验性给定和简化近似是不可避免的。

2.2 基本方法的内在发展逻辑和创新性进展

由上述各种基本方法的显著特征可以清晰地看

出:之后的方法和之前的方法都是上下承接的;之后的方法都是克服了之前的局限性而循序渐进地发展的,由此才带来了各基本方法的创新性进展。

从功能和效果而言,各基本方法的显著特征对应着分析同化方法的创新性进展:多项式函数拟合方法开创了从人工插值的“主观分析”到计算机插值的“客观分析”的新纪元;SCM 带来了分析方法的新突破:初猜值的引入和观测增量加权订正的方式形成了之后统一的分析同化方法的公式形式;OI 背景场概念和误差信息的引入标志着分析同化方法发展成为以估计理论为基础的一门学科;Var 开始了使用卫星、雷达等大量非常规遥感观测资料的新时代,有了全球数值预报质量的快速提高;EnKF 实现了背景场误差协方差的“随天气形势演变”,对于更佳地利用特别是中小尺度特征的观测信息,改善特别是时空变化剧烈的要素场的分析质量,进而提高中小尺度天气系统所造致的灾害性天气的数值预报水平,带来了潜在的优势。

从内容和过程而言,上下承接的循序渐进的发展创新逻辑集中体现在两个方面:(1)可用信息的源和内含的不断扩展;(2)分析同化方法中经验局限性的不断减少和真实科学性的不断增加。表 1 梳理了各基本方法中可用信息的不断扩展和经验局限性的不断减少等方面。

由表 1 可以看出,分析同化方法的发展创新进程在认识上是清晰简单的。

表 1 分析同化的发展进程中可用信息的不断扩展和经验局限性的不断减少等方面
Table 1 Step-by-step expansion of available information and cutdown on limitations along the historical progress of atmospheric data assimilation

	多项式拟合	SCM	OI	Var	EnKF
所用的信息	只是观测的值: y_o	增加初猜场的值: $y_o + x_g$	引入背景场, 增加协方差: y_o 和 x_g + R 和 B	增加更多种类观测: 如大量的非常规的 卫星、雷达资料	背景场误差协方差: 动态的 B 矩阵
观测信息使用 方式	观测值	观测增量(新息) $[y_o - H(x_g)]$	资料的区域选择	同化时间窗内的 所有资料	逐时次的资料
最优准则 和 求解的方式	最小二乘原则	平均平滑	分析误差方差最小	分析场似然概率最大	分析误差方差最小
经验性和 局限性	适当区域范围的 求拟合函数系数	单点分析 的逐步迭代	线性方程组求权重 和单点分析的一次 性计算	泛函极小的全局求解	逐时次序贯的更新 循环方式
经验性和 局限性	拟合函数本身 是给定的	观测增量的权重 是给定的	线性观测算子和 资料的区域选择	Var和OI一样: 静态 B 矩阵	有限取样
真实性和 客观科学性	客观的 天气图分析	观测增量作为订 正的新的信息	权重是统计最优 来确定的	复杂非线性观测算子 和一次使用所有资料	更接近真实的 B 矩阵

(“ \rightarrow ”表示扩展或进程)

这些发展创新在内在本质上直接联系着大气资料同化的根本概念和各基本方法的核心关键技术。

大气资料同化的根本概念是随机变量(朱国富, 2015)。随机变量概念下的数据可用信息不只是一个值,而是包含值、误差的不确定性和协方差,因此能自然地理解 OI 中的可用信息增加了“误差协方差”。分析同化建立在随机变量概念上才使得概率论与数理统计学得以成为资料同化的数学基础,因此才能有 OI 中以分析误差方差最小的最优权重和 Var 中以分析场似然概率最大的目标泛函。

使用观测增量的方式是 SCM 的核心关键技术,因此 SCM 有了以观测增量对初猜值的订正。应用变分方法求解泛函是 Var 的核心关键技术,因此 Var 能够引入复杂的非线性观测算子而使得 Var 中信息源扩展到卫星、雷达等多种非常规观测,和能够一次使用所有资料进行全局求解而不存在资料的区域选择问题。采用集合方法统计误差协方差是 EnKF 的核心关键技术,因此 EnKF 能考虑动态 B 矩阵。

由此可揭示:理解根本概念和掌握核心关键技术是科技自主创新的基础和源泉。

2.3 “大气资料同化”的由来和内涵

大气资料同化源于 NWP 初值形成的客观分析。主观分析和客观分析的基本不同点是在方式上:主观分析方法的主观性(需要手工的因人而异)与客观分析方法的客观性(依赖算法和程序的无需人工干预)。这二者的共同点是在结果上都是一种功能实现:根据不规则分布的观测及已知的其他大气运动规律的信息,得到所要的规则格点上最可能的值(分析场);二者的另一个相同点是术语上有共同的“分析”字眼,它体现了这二者在功能结果上的相同概念。分析这个字眼来源于天气预报员的天气图分析;由 Panofsky(1949)提出的最早的客观分析就称为“客观天气图分析”。

SCM 中初猜场/OI 中先验背景场的引入形成了之后分析同化方法的统一公式形式:分析场是观测增量对初猜场/背景场的加权订正[见文中式(2)和(4)]。随着数值预报越来越好,从效果上,NWP 的短期预报场用作分析的初猜场/背景场成为自然结果。

预报场作为背景场有着形式和内涵两方面的深刻意义。首先,从公式形式上,这造出了一个特定的

更新字眼:当使用预报场提供初猜场/背景场时,OI 分析被称为在可获得观测资料的区域“更新(update)”预报场(Bergman, 1979)。从内涵上,这使得预报和分析构成循环,即:前期的模式预报场作为分析某一时刻观测资料的背景场,得到分析场,这被称为分析步骤;之后以该分析场作为初始条件积分预报模式到下一个时刻,得到预报场,这被称为预报步骤;再由这个预报场作为分析新时刻观测资料的背景场,得到分析场,这是又一循环的分析步骤;如此循环。从这个循环的预报角度,这体现了利用观测资料不断修正模式的预报,因为利用观测资料所得到的分析场为下一时刻的模式预报提供更准确的初始场。更重要的,从这个循环的分析角度,这表示一个过程,在这个过程中不同时刻的观测资料和大气动力数值模式这二者融合来尽可能精确地确定大气的状态。正是由此造出了同化这个字眼,表示这个过程(Talagrand, 1997)。

沿着“分析”、“更新”、“同化”这一历史脉络,可以理解,引入预报场作为背景场之后,各种方法的大气资料分析同化都有着共同的三层意义:分析字眼的实现功能,更新字眼的外在形式,同化字眼的内在涵义。分析字眼的实现功能就是“由不规则分布的观测及先验信息得到某一预定的规则网格点上最可能的值”的这一功能。更新字眼的外在形式就是“用观测增量的加权订正更新作为背景场的预报场来得到分析场”的这一公式[式(5)]。同化字眼的内在涵义就是“不同时刻观测资料和大气动力数值模式通过预报和分析的循环进行融合来尽可能精确地确定大气状态”的这一过程。

而且,初期的“分析”是依附于数值预报,“分析”的功能是提供数值预报初值。而“同化”的内涵体现了“大气资料同化是独立的”之内涵所在,标志着大气资料同化成为和作为一门独立学科。事实上,大气资料同化的作用已发展出两大方向,第一,所得到的大气状态的最优估计,尽管还主要地为数值预报提供初值,但还用于诸多方面,如:天气系统结构和演变的诊断分析,以及再分析资料用于气候变化的监测;第二,其方法本身的发展开辟了新领域,如:利用同化中伴随和集合方法的“观测对分析、预报的影响评估”和“适应性观测的敏感区识别”。因此,从以前使用的术语“客观分析”到现在越来越普遍使用的术语“资料同化”体现了分析同化的历史发展进程。

可见,从历史起源和发展进程才能清晰和准确

地理解“资料同化”这个术语的由来和内涵。

3 结 论

数值天气预报中分析同化基本方法先后经历了多项式函数拟合方法、逐步订正方法、最优插值方法、变分方法和集合卡尔曼滤波方法。沿着这个历史发展进程,本文力求本色地介绍了这些方法的基本思想和做法要点。在此基础上,评述了它们的显著特征、深层意义以及局限性。从中看出和理解:

(1) 分析同化基本方法有着上下承接的纵向联系。

(2) 分析同化方法有着循序渐进的发展脉络和在认识上清晰简单的创新逻辑。这集中体现在两个方面:(a)可用信息的源和内含的不断扩展;(b)经验局限性的不断减少和真实科学性的不断增加。

(3) 理解根本概念和掌握核心关键技术是科技自主创新的基础和源泉。

这些对我国 NWP 的自主研发有着现实的借鉴价值和启示意义。

本文从历史起源上阐明“主观分析”与“客观分析”、“初猜场”与“背景场”以及“新息”等基本概念;还沿着“分析”、“更新”、“同化”这一历史脉络,来理解和把握“大气资料同化”这个术语的由来和内涵。

参 考 文 献

雅罗 M. 2004. 信息时代的天气、气候和水——2004 年世界气象日致辞. 气象知识, 1, 4-5.

朱国富. 2015. 理解大气资料同化的根本概念. 气象, 41(4): 456-463.

邹晓蕾. 2009. “资料同化理论和应用(上册)”. 北京: 气象出版社.

Kalnay E. 2005. 大气模式、资料同化和可预报性. 蒲朝霞等译, 北京: 气象出版社, 115, 130-134, 144-146pp.

Bergman K H. 1979. Multivariate analysis of temperature and winds using optimum inetrpolation. Mon Wea Rev, 107(11): 1423-1444.

Bergthorsson P, Doos B. 1955. Numerical weather map analysis. Tellus, 1: 329-340.

Bishop C H, Etherton B J, Majumdar S J. 2001. Adaptive sampling with the ensemble transform Kalman filter. 1: Theoretical aspects. Mon Wea Rev, 129: 420-436.

Bjerknes V. 1904. Das Problem der Wettervorhersage, betrachtet vom Stanpunkt der Mechanik und der. Meteor Zeits, 21: 1-7.

Bouttier F, Courtier P. 1999. Data assimilation concepts and methods. Meteorological Training Course Lecture Series, ECMWF, 2002.

Courtier P, Andersson E, Heckley W, et al. 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part 1: formulation. Quart J Roy Meteor Soc, 124: 1783-1807.

Cressman G P. 1959. An operational objective analysis system. Mon Wea Rev, 87: 367-374.

Daley R. 1991. Atmospheric Data Analysis. Cambridge Atmospheric and Space Science Series, Cambridge: Cambridge University Press.

Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res, 99(C5): 10143-10162.

Gandin L S. 1963. Objective Analysis of Meteorological Fields. Gidrometeorologicheskoe Izdatel' stro, Leningrad. Translated from Russian, Israeli Program for Scientific Translations. Jerusalem. 1965.

Hamill T M. 2006. Ensemble-based atmospheric data assimilation. Predictability of Weather and Climate, Palmer T and Hagedorn R, Eds., Cambridge: Cambridge University Press, 124-156.

Kalnay E. 2003. Atmospheric Modeling, Data Assimilation and Predictability. Cambridge: Cambridge University Press.

Lorenc A C. 1981. A global three-dimensional multivariate statistical analysis scheme. Mon Wea Rev, 109: 701-721.

Lorenc A C. 1986. Analysis methods for numerical weather prediction. Quart J Roy Met Soc, 112: 1177-1194.

Panofsky H A. 1949. “Objective Weather-Map Analysis,” J Meteor, 6: 386-392.

Parrish D F, Derber J C. 1992. The National Meteorological Center's spectral statistical interpolation analysis system. Mon Wea Rev, 120: 1747-1763.

Talagrand O. 1997. Assimilation of observations: An introduction. J Meteor Soc Japan, 75: 191-209.