

朱国富. 2015. 理解大气资料同化的内在逻辑和若干共性特征. 气象, 41(8): 997-1006.

理解大气资料同化的内在逻辑 和若干共性特征^{* 1}

朱国富

国家气象中心, 北京 100081

提 要: 大气资料同化已发展成为基于估计理论的一门学科。本文试图梳理资料同化学科内容从“性质上分析同化问题是什么”到“过程上资料同化如何实施”再到“结果上分析场有着什么内在要求”这一内在逻辑, 相应地综合阐述“逆问题求解和所需要的先验信息”、“观测信息的输入及其权重与传播平滑”和“同化结果的准确性和协调性”等资料同化作为一门学科的这些基本方面; 以期这一逐步递进的逻辑及其基本方面有益于对资料同化在条理脉络上的一个整体理解。此外, 着重于具体同化方法中的一些等价性和共性特征, 本文梳理和归纳了“变分方法分别与最优插值方法之间和与扩展卡尔曼滤波方法之间的等价性”、“同化方法中数学的主要作用”, 以及“同化方法中要应用综合学科知识”和“业务同化系统实施中要做好系统工程”等内容; 以期这些内容增加对资料同化方法及同化系统研发的理解和思考。

关键词: 大气资料同化, 逆问题, 传播平滑, 背景场误差协方差, 观测算子

中图分类号: P456

文献标志码: A

doi: 10. 7519/j. issn. 1000-0526. 2015. 08. 009

Interpretation of Several Basic Aspects of Atmospheric Data Assimilation

ZHU Guofu

National Meteorological Centre, Beijing 100081

Abstract: This article tries to summarize and synthetically remark several basic aspects and common inherent characteristics of atmospheric data assimilation. Such a logic line is highlighted from “what’s data assimilation problem”, to “how is data assimilated”, and to “what are the inherent requirements for the resulting analysis”. The logic correspondingly involves “inverse problem and a priori information”, “input, weight, propagation and smoothing of observation information”, and “requirements of accuracy and consistency for the analysis”. Also, we lay emphasis on some common inherent characteristics, including “equivalences between variational method and optimal interpolation method and between four-dimensional variational method and extended Kalman filter method”, “the dominant role of mathematics in assimilation methods”, “multidisciplinary knowledge connotated in assimilation methods” and “system engineering required in operational assimilation systems”.

Key words: atmospheric data assimilation, inverse problem, propagation and smoothing, the background error covariance, observation operator

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

2015 年 3 月 24 日收稿; 2015 年 5 月 6 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

大气资料同化源于数值天气预报 (Numerical Weather Prediction, NWP) 初值形成的客观分析。分析同化的基本方法先后经历了多项式函数拟合、逐步订正 (SCM)、最优插值 (OI)、变分 (Var) 和集合卡尔曼滤波 (EnKF) 等方法。大气资料同化发展到今天, 已成为基于估计理论的一门学科, 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”。

Lorenc (1986) 基于所导出的 NWP 最佳分析的理想化方程, 推导和讨论了多种分析方法在公式和求解上的相互关系。朱国富 (2015) 着重于上述基本方法的上下承接的纵向联系, 梳理和阐述了分析同化的历史发展脉络。

本文跳出具体资料同化方法及其发展脉络, 试图梳理大气资料同化作为一门学科在一般意义上的基本方面, 即“分析同化问题是什么”、“资料同化如何实施”和“分析结果有着什么内在要求”, 来阐述“逆问题求解和所需要的先验信息”、“观测信息的输入及其权重与传播平滑”、“同化结果的准确性和协调性”等内容; 此外, 文中还梳理和归纳了分析同化方法中的一些等价性和共性特征。

1 资料同化作为一门学科的基本方面

1.1 分析同化问题是什么: 逆问题求解和所需要的先验信息

1.1.1 NWP 分析问题的提出和逆问题

早期 NWP 客观分析方法所需解决的问题是要实现这样的分析功能: 使用计算机将某时刻不规则分布的观测资料通过插值得到该时刻某一预定的规则网格点上最可能的值。这可以表述为: 给定实际观测向量 y_0 , 已知观测物理量 y 与模式状态物理量 x 之间确定的具体映射关系 H (观测算子):

$$y = H(x) \quad (1)$$

则早期 NWP 分析问题就是: 由式 (1) 反过来求得最佳的 x (被称为分析场 x_a), 来提供 NWP 预报模式的初值条件。可见, “已知观测信息由式 (1) 求解模

式状态变量”, 这是一个逆问题。

1.1.2 逆问题求解的欠定性

Bergthorsson 等 (1955) 对早期 NWP 分析问题的调查研究得出的结论是: 仅仅通过天气观测资料的插值来得到合理的分析场常常是不可能的。

只有观测信息来求解这个逆问题一般是欠定的, 这被称为 NWP 分析问题的欠定性 (Underdeterminacy)。它源于观测的不完整性。在实际的实施中, 模式状态变量 x 有着很大的维数 N_x ; 对于一个格点模式, N_x 等于所有三维物理空间的格点数与每个格点上模式变量个数的乘积。随着计算机能力的持续增长, NWP 预报模式的复杂性和分辨率大小也相应地增长。为了很好地预报如锋面和强风暴的重要天气现象, 需要有很高的分辨率。于是, x 的维数大小 N_x 未来还将按照计算能力许可的速度持续增长。这样, 尽管遥感技术和其他自动观测系统的进展增加了 y_0 的观测数 N_y , NWP 分析问题还是可能保持欠定的: $N_x > N_y$ 。事实上, 我们永远不会有每个格点上的每个模式变量值对应的观测值, 由此 NWP 分析问题至少会是部分欠定的 (Lorenc, 1986)。

1.1.3 状态变量先验背景信息的引入

数学上, 贝叶斯定理 (Bayes' theorem) 恰合地适用于 NWP 分析逆问题的求解。贝叶斯定理指出: 随机事件 A 在随机事件 B 发生的条件下的后验概率, 与事件 A 的先验概率和事件 B 在事件 A 发生的条件下的后验概率的乘积成正比: $P(A|B) \propto P(B|A)P(A)$ 。可见, 其中的 $P(A|B)$ 和 $P(B|A)$ 的关系对应着逆问题求解。

使用贝叶斯方法, 由已知的 (原则上可以估计的) 各信息概率分布函数 (probability distribution functions, PDFs), 能够导出用 PDF 表示的式 (1) 的一个广义逆 (Lorenc, 1986)。

不论模式状态物理量 x 还是观测物理量 y , 我们都不能知道它的真值。把它们真值作为某一随机变量 (或向量) X_i 和 Y_i , 该随机变量的某一取值都便是真值的一个可能; 这样, 该随机变量的最可能值就是对于真值的最优估计。

用 A 表示“ X_i 取某一可能值 x ”这一随机事件 $X_i = x$, 用 B 表示“ Y_i 取某一可能值 y ”这一随机事件 $Y_i = y$, 那么由贝叶斯定理有:

$$P(\mathbf{X}_t = \mathbf{x} | \mathbf{Y}_t = \mathbf{y}) \propto P(\mathbf{Y}_t = \mathbf{y} | \mathbf{X}_t = \mathbf{x}) \cdot P(\mathbf{X}_t = \mathbf{x}) \quad (2)$$

$P(\mathbf{X}_t = \mathbf{x} | \mathbf{Y}_t = \mathbf{y})$ 是有了观测信息之后(已知观测信息 $\mathbf{Y}_t = \mathbf{y}$ 的条件下)的关于 \mathbf{X}_t 的后验(条件)概率,简记为 $P_a(\mathbf{x})$,这就是用 PDF 表示的式(1)的一个广义逆。上式中, $P(\mathbf{X}_t = \mathbf{x})$ 是无条件的关于 \mathbf{X}_t 的先验概率,简记为 $P_b(\mathbf{x})$,包含了没有观测信息之前我们关于状态变量 \mathbf{x} 的知识; $P(\mathbf{Y}_t = \mathbf{y} | \mathbf{X}_t = \mathbf{x})$ 是已知状态信息 $\mathbf{X}_t = \mathbf{x}$ 的条件下的关于观测 \mathbf{Y}_t 的条件概率,简记为 $P_\alpha(\mathbf{y})$ 。

因为一个随机变量的概率分布包含了这个随机变量的所有信息,因此由广义逆式(2)便能得到已知观测 $\mathbf{Y}_t = \mathbf{y}$ 条件下的模式状态变量的最佳值,也就是分析场的值 \mathbf{x}_a ;“最佳值”表达为 \mathbf{x}_a 满足: $\mathbf{x}_a = \int \mathbf{x} P_a(\mathbf{x}) d\mathbf{x}$,它是随机变量 \mathbf{X}_t 的所有可能值的均值,是对应最小方差的最优估计(陈希孺,2009);或是 $\mathbf{x} = \mathbf{x}_a$,使得 $P_a(\mathbf{x}_a) = \max_x P_a(\mathbf{x})$,它是随机变量 \mathbf{X}_t 的众数,是对应最大似然概率的最优估计。 \mathbf{x}_a 也就是式(1)逆问题的解。

由式(2)可知,逆问题的求解,除了观测的后验信息,还引入了状态变量 \mathbf{x} 的先验信息(prior information),它用 PDF 表示为 $P_b(\mathbf{x})$ 。

1.1.4 所引入的状态变量先验背景场和解决分析问题欠定性的线性理论解

如果已知某一先验背景场及其误差的概率分布,那么 $P_\alpha(\mathbf{x})$ 是可以估计的。在没有观测信息 $\mathbf{Y}_t = \mathbf{y}$ 之前我们关于模式状态变量 \mathbf{x} 的知识来自 NWP 的前期模式预报(或是气候场、或是两者混合提供),用 \mathbf{x}_b 表示,它可以作为 \mathbf{x} 的先验背景场。 \mathbf{x}_b 是有误差的;由于 \mathbf{x} 是随机变量 \mathbf{X}_t (模式状态真值)的一个可能值,所有 $(\mathbf{x} - \mathbf{x}_b)$ 表示了 \mathbf{x}_b 的误差[这里 $-(\mathbf{x}_b - \mathbf{x})$ 表示为负号是为了一致的公式表示上方便起见]。于是状态变量先验概率:

$$P_b(\mathbf{x}) = P(\mathbf{X}_t = \mathbf{x}) = P(\mathbf{x} - \mathbf{x}_b) \quad (3)$$

它表示为 \mathbf{x}_b 的误差的概率分布。

类似地,如果已知某一观测值及其误差的概率分布,那么 $P_\alpha(\mathbf{y})$ 是可以估计的。对于某一已知观测 \mathbf{y}_0 , \mathbf{y}_0 是有误差的;由于 \mathbf{y} 是随机变量 \mathbf{Y}_t (观测量真值)的一个可能值,所有 $(\mathbf{y} - \mathbf{y}_0)$ 表示了 \mathbf{y}_0 的误差。在 $(\mathbf{X}_t = \mathbf{x})$ 条件下,由式(1)知,此时 $\mathbf{y} = H(\mathbf{x})$ (这里没考虑观测算子 H 的误差)。于是观测变量

条件概率:

$$P_\alpha(\mathbf{y}) = P(\mathbf{Y}_t = \mathbf{y} | \mathbf{X}_t = \mathbf{x}) = P(\mathbf{y} - \mathbf{y}_0) = P[H(\mathbf{x}) - \mathbf{y}_0] \quad (4)$$

它表示为 \mathbf{y}_0 的误差的概率分布。这里,没考虑 $P_\alpha(\mathbf{y})$ 与 \mathbf{x}_b 有任何关系,因此已经预设了背景场误差和观测误差是无关系的。

由式(3)和(4),则式(2)写为:

$$P_a(\mathbf{x}) \propto P_\alpha(\mathbf{y}) P_b(\mathbf{x}) = P[H(\mathbf{x}) - \mathbf{y}_0] P(\mathbf{x} - \mathbf{x}_b) \quad (5)$$

通常的一个假定是背景场误差和观测误差都无偏和满足正态分布,即:

$$P_b(\mathbf{x}) = P(\mathbf{x} - \mathbf{x}_b) \propto \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \cdot \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b)\right] \quad (6)$$

$$P_\alpha(\mathbf{y}) = P[H(\mathbf{x}) - \mathbf{y}_0] \propto \exp\left\{-\frac{1}{2}[H(\mathbf{x}) - \mathbf{y}_0]^T \mathbf{R}^{-1}[H(\mathbf{x}) - \mathbf{y}_0]\right\} \quad (7)$$

式中 \mathbf{B} 和 \mathbf{R} 分别是背景场误差协方差和观测误差协方差。

由式(6)及(7),式(5)的后验似然概率 $P_a(\mathbf{x})$ 最大的解: $P_a(\mathbf{x}_a) = \max_x P_a(\mathbf{x})$,便转化为等价于一个目标泛数 $J(\mathbf{x})$ 最小的解(Lorenc,1986):

$$J(\mathbf{x}_a) = \min_x J(\mathbf{x}) \quad (8)$$

其中的目标泛函: $J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}[H(\mathbf{x}) - \mathbf{y}_0]^T \mathbf{R}^{-1}[H(\mathbf{x}) - \mathbf{y}_0]$ 。

式(8)就是通过引入状态变量先验背景场 \mathbf{x}_b 来求解“给定观测 \mathbf{y}_0 、已知观测算子 H ,求得模式状态变量的最可能值 \mathbf{x}_a ”这一 NWP 分析逆问题的数学公式形式。如上述,这里预设了“背景场误差和观测误差是无关系的”,并假定:“背景场误差和观测误差都无偏和满足正态分布”。

形式上,所引入的先验背景场合理地约束模式状态变量 \mathbf{x} 在 N_x 维空间的一个面(或点)上或是附近,它一般地表示为式(3),和由式(6)表示为高斯的 PDF。

理论上,如果将先验背景场 \mathbf{x}_b 和观测 \mathbf{y}_0 合在一起视为输入信息向量 \mathbf{z} ,那么根据线性统计估计理论(theory of statistical linear estimation),当且仅当满足可观测性条件(observability condition)时,由向量 \mathbf{z} 可以得到完全确定的最优线性无偏估

计 \mathbf{x}_a , 即 BLUE (best linear unbiased estimate) (Tagliand, 1999)。可观测性条件表示: 向量 \mathbf{z} 包含了大气状态向量 \mathbf{x} 的每个分量的 (直接或是间接的) 信息; 这意味着向量 \mathbf{z} 的维数不小于向量 \mathbf{x} 的维数。这个 \mathbf{x}_a 表示为:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{Kd} \quad (9)$$

式中, $\mathbf{K} = \mathbf{BH}^T (\mathbf{HBH}^T + \mathbf{R})^{-1}$ 是增益矩阵 (也被称作权重函数矩阵 \mathbf{W}), $\mathbf{d} = \mathbf{y}_o - \mathbf{Hx}_b$ 是新息向量; 其中 \mathbf{H} 是线性观测算子。

式(9)也是式(8)当 H 是线性时的目标泛函最小的解。当 H 是非线性时, H 在 \mathbf{x}_b 附近做线性化近似的情况下, 仍然能得到式(8)的显示解, 形式仍如式(9); 只是此时 $\mathbf{d} = \mathbf{y}_o - H(\mathbf{x}_b)$, 以及 $\mathbf{H} = \left. \frac{\partial H(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_b}$ 是 H 的切线性算子。

1.1.5 处理欠定性问题的先验知识和四维变分的同化时间窗

可以有許多有用的先验知识, 通过减小状态变量 \mathbf{x} 的有效自由度和增加输入信息的实际维数来处理 NWP 分析问题的欠定性。Lorenz (1986) 指出, 先验知识有诸多的物理根源和来源, 如: (1) 时间序列的观测资料增加 N_y 维数, 加之用 NWP 预报模式方程约束的模式状态时间演变使得利用四维观测资料求解三维分析问题成为可能, 由此有助于解决分析问题的欠定性, 这构成了四维资料同化; (2) 大气是缓慢变化的, 隐含着作用于大气的各种力之间的平衡, 如地转平衡约束, 这能用来减少欠定性; (3) 大气的不同尺度运动和参数之间存在着相互作用, 这意味着一些尺度运动和参数是不完全独立的, 如 Lorenz 等 (1980) 表明真实的锋区湿度场能够从高度和风资料得到。(2) 和 (3) 中内含的物理变量之间关系将状态变量 \mathbf{x} 的有效自由度减少到小于 N_x 。

四维变分方法是以强约束的方式在观测算子中引入 NWP 预报模式, 来利用在同化时间窗内各个观测的时间分布信息。由上述(1), 延长四维变分的同化时间窗的长度, 能增加时间序列的观测资料量, 增加观测 \mathbf{y}_o 的维数, 由此有助于解决欠定性。

但是, 另一方面, 实际应用的四维变分方案都假定预报模式是完美的, 这使得它无法考虑预报模式误差对分析的影响。这是一个缺陷, 因为这意味着: 同化时间窗之初始时刻的较前观测和终时刻的较后观测被不合理地给予了同样的信度。因此, 如果不

能考虑预报模式的误差, 便不能过分延长同化时间窗的长度。

所以, 增加同化时间窗的长度, 和为此在目标泛函中包含模式误差的描述, 还是四维变分同化技术的一个未来发展 (ECMWF, 2011)。这需要在四维变分中引入一个弱约束公式表示, 考虑模式误差, 由此同化时间窗的长度可以延长至数天。

可见, 即使是一个认识上早已清晰的“增加同化时间窗长度”这一方案, 由于现实存在的模式误差的实际限制, 它只能在未来若干年之后有能力处理这个误差或随 NWP 发展这个误差已足够减少的情况下得以考虑。这对于科学地、从实际现状出发地开展我国的 NWP 自主研发有着现实的借鉴价值和启示意义。

1.2 资料同化如何实施: 观测信息的输入和权重与传播平滑

撇开各具体资料同化方法的内在技术 (如变分全局求解和集合样本统计误差协方差), 在笼统的一般意义上, 资料同化的实施是实现最优的资料融合。所得到的结果有着统一的公式形式, 表达为 1.1.4 节中的式(9); 它外在地表现为“以先验背景场 \mathbf{x}_b 为平台, 统计最优地融合观测 \mathbf{y}_o , 得到分析场 \mathbf{x}_a ”这一分析功能的特征。

由式(9)可见, 分析功能的实现涉及两个基本方面: (1) 观测信息的输入: 观测增量 (或称新息) $[\mathbf{y}_o - H(\mathbf{x}_b)]$, (2) 观测信息的权重与传播平滑: 权重函数矩阵 \mathbf{W} 。两个方面对应着各自的基本要求和实施实现。

1.2.1 观测信息的输入: 观测算子 H

新息 (innovation) $[\mathbf{y}_o - H(\mathbf{x}_b)]$ 准确地表达了不是观测值本身或背景值本身, 而是测站位置上观测值和由背景值模拟的观测相当量这两者之差引入的“新的信息”。作为输入信息, 对应的基本要求是新息的数量和准确性。在观测 \mathbf{y}_o 一定的情况下, 新息的数量和准确性取决于观测算子 H 。

在实施上, 观测算子 H 一般地实现三个功能: (1) 观测算子中的变量变换 (如同化卫星辐射率的辐射传输模式), (2) 空间插值 (水平和垂直插值), (3) 时间演变 (如四维变分的 NWP 预报模式); 它们分别解决规则网格点上状态物理量 \mathbf{x} 和不规则分布的

观测物理量 y 这两者在物理变量上、空间位置上和时间点上的不同。

实际应用中,观测信息的输入还包括观测资料的质量控制、偏差订正和稀疏化。质量控制的目的主要是实现三个功能:(1)剔除错误的非正常观测资料,(2)消除资料中非正态分布的误差,和(3)剔除正常的但与背景场不协调而不能消化吸收的观测(即 $[y_o - H(x_o)]$ 过大的观测)。偏差订正的功能是处理资料的系统性误差。稀疏化的功能是处理观测资料中难以确定的相关性。质量控制的功能(1)和(3)是避免分析结果收敛于有问题的观测资料;质量控制的功能(2)、偏差订正和稀疏化的功能是为了满足目前资料同化方法及其实际应用中通常用到的“背景场误差和观测误差都无偏和满足正态分布”和“观测资料不相关”等假定。

顺便指出,在实施上,变分方法,因为能够实现复杂的非线性观测算子的引入,带来了新息的数量(含观测种类)的急剧增加,开始了直接同化卫星、雷达等大量非常规遥感资料的新时代。

1.2.2 观测信息的权重与传播平滑:背景场误差协方差矩阵 B

式(9)中的权重函数矩阵 $W = BH^T(HBH^T + R)^{-1}$ 决定了观测信息的权重与传播平滑;在观测 y_o 及其方差一定的情况下,它主要取决于背景场误差协方差矩阵 B 。不妨考虑单个观测,它是与(一维向量排列的)格点 k 重合的观测站 k 上的一个纬向风观测 u_o 。对于 u 风场的分析,此时 $H = H^T = 1$ (无需变量变换和空间插值,没考虑四维同化),那么由式(9),在任一格点 i 上有: $u_a(i) = u_b(i) + w(i, k) \cdot [u_o(k) - u_b(k)]$, 其中 $w(i, k) = b(i, k) / [b(k) + o(k)]$, 它包含了该观测 u_o 对 u 风场分析的作用。在格点和测站重合的 k 点上,容易看到 $w(k, k) = b(k) / [b(k) + o(k)]$ [$b(k)$ 是 $u_b(k)$ 的误差方差], 它是观测信息 $[u_o(k) - u_b(k)]$ 的权重。对于其他任一格点 i , $w(i, k)$ 表达了测站 k 上观测信息向其他格点 i 的空间传播和权重,而空间传播的功能完全取决于背景场误差协方差矩阵的元素 $b(i, k)$ 。类似地,再考虑有多个测站的观测信息对格点 i 的作用,则有 $\sum_k \{w(i, k)[u_o(k) - u_b(k)]\}$, 这包含了观测信息的平均平滑。

作为观测信息的权重与传播平滑,基本要求是

W 包含合理的天气流型的尺度特征。原因如下:一方面, W 是统计意义的,在实际业务实施中,不论是集合样本方法还是其他误差代用品方法的统计,都难以得到它在某一时刻的真实统计。另一方面,作为一个基本物理特征,真实大气是缓慢变化的,这隐含着作用于大气的各种力之间的平衡,因此能符合这个特征的分析场应当是平滑的,这要求:权重与传播平滑是否合适是首要地表示天气流型的大尺度特征;同时,由于格点值只是格体平均意义上的值,不能分辨次网格部分,由此产生观测的代表性误差,它表现为背景场误差作用于观测算子的误差,因此只有被模式网格分辨的尺度应该被包括在背景误差协方差函数中(Lorenc, 1986)。

背景场误差协方差,作为一个协方差矩阵,可以表示为 $B = \Sigma C \Sigma$, 其中 Σ 是标准差的对角矩阵, C 是一个相关矩阵;其标准差部分表征背景场向量各分量的不确定性,其相关部分表征背景场向量各分量之间的联系,包含不同变量间的物理相关(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)和不同位置间的空间相关。由 $w(i, k) = b(i, k) / [b(k) + o(k)]$ 可知, B 的标准差部分和各观测的方差共同决定了用来订正背景场的新息的适当权重; B 的相关部分,在没有约束项和三维变分的情形下,完全地决定了新息的合理传播和平滑。

所以,背景场误差协方差对分析有着根本的影响。Bannister(2008)清楚地阐述了 B 对分析的作用:它规定了先验信息重要性的权重以及观测信息在水平及垂直空间和不同变量间的传播,它规定了分析增量只存在于 B 张成的子空间,它使得各个不同观测的作用协同增效(同时同化不同观测时所减小分析误差的程度大于分开独立地同化各个观测时的合计效果)。

1.2.3 具体分析同化方法中的观测信息的输入和权重与传播平滑

Panofsky(1949)的多项式函数拟合方法没有显式的观测算子和权重函数,只是在分析区域中使用观测来确定拟合函数多项式的待定系数;为了得到可信的分析场,通过使用多于待定系数个数的观测,按照最小二乘法来考虑平滑问题。

Cressman(1959)的 SCM 是等压面二维分析,

观测算子只包含水平方向的空间插值。SCM 的权重函数是经验给定的,随距离的平方而减小;通过影响半径范围内所有观测资料订正作用的平均来考虑平滑和使用影响半径逐步减少的序列扫描使得分析场包含不同的尺度谱。

OI 的观测算子是三维的空间插值,包含水平和垂直方向的空间插值,不过是简单线性的(水平方向的双线性插值和垂直方向的线性插值)。Var 的观测算子不仅有三维的空间插值,还有可以是复杂非线性的变量变换,如对于卫星辐射率资料的直接同化,这个变量变换是辐射传输方程;对于四维变分,观测算子还包含时间演变,即 NWP 预报模式,能在同化时间窗内考虑时间上比三维变分更准确的观测信息。

OI、Var 和 EnKF 的权重和传播平滑是由统计最优(分析场方差最小或似然概率最大)所确定的增益矩阵 \mathbf{K} 。但 OI 中(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)只是粗略地估计 \mathbf{B} ,而且是定义在局地的近似,如空间相关分解为水平和垂直高斯相关的积;Var 中 \mathbf{B} 可以是定义在全局的和更一般的方式,如水平和垂直相关的不可分离。在实际的业务实施中,由于矩阵 \mathbf{B} 通常维数巨大和条件数很大而近于病态,导致 OI 和 Var 同化方案的数值求解困难,所以它们把 \mathbf{B} 当作不随时间变化的静态常量。在 EnKF 中使用短期预报集合来估计动态演变的 \mathbf{B} 。

顺便指出,在实际应用中,对于资料同化如何实施的二个基本方面是:(1)能被同化的任何一个观测都需要有一个对应的观测算子;(2)处理 \mathbf{B} 使得求解过程易于收敛并能得到可靠的解,构成了同化系统的核心框架,它作用于任何一个观测。简言之,被同化的不同观测资料,有着各自的观测算子,被作用于同一的核心框架。因此,处理 \mathbf{B} 的核心框架和对应每一资料的观测算子,作为二个主体部分,主宰着资料同化的实施实现。

1.3 分析结果有着什么内在要求:同化结果的准确性和协调性

分析场有着二个内在要求:(1)作为融合同化各种资料信息的结果,它是大气状态的尽可能准确的估计,这是它的准确性要求;(2)作为提供给 NWP

预报模式的初值,它需要与预报模式的协调,这是它的协调性要求。

1.3.1 准确性要求

同化结果的准确性要求体现为分析场是大气状态的最优估计,即它是以最小方差估计为标准的均值或以最大似然估计为标准的众数(见 1.1.3 节)。

同化结果的准确性是资料同化方法所不断追求的内在要求,体现了资料同化本身的历史发展进程及其内在发展逻辑,也使得资料同化成为一门独立学科,使得资料同化不只是为数值预报提供初值,还用于天气系统结构和演变的诊断分析和气候变化的监测等方面。朱国富(2015)梳理和综合阐述了这些内容。

1.3.2 协调性要求

同化结果的协调性要求源于二个方面。一方面,科学认识和工程技术的客观局限性,包括大气运动的多尺度特征、NWP 预报模式对真实大气的近似(如数值模式的离散化和参数化)、观测的代表性误差与观测误差和资料同化方法的假定条件等,使得预报模式和初始条件都存在误差,由此带来了实际上不完全真实的预报模式和不完全准确的初始条件。另一方面,作为观测事实的大气运动的特征,大气物理量场通常缓慢地变化,其实质是作用于大气运动的各个力近似地处于平衡,使得在客观上要求考虑预报模式和初值之间的协调。这二个客观方面造致同化结果的协调性要求。

资料同化结果的协调性是资料同化应用于 NWP 的要求,这表现为分析同化中“初始化技术”的引入和研发。为了减小初始条件中不平衡的初始化技术是在分析同化之后和积分模式预报之前所实施的一个技术步骤,来得到符合大气运动平衡性质的初始条件,作为所需要的积分初值。

早在 1922 年 Richardson 先驱性的数值天气预报试验,就突出了资料同化结果的协调性问题:由于资料噪音和在原始方程模式的解中存在快速惯性重力波,导致了初始地面气压倾向的错误估计。因为,如果一个模式的初始条件不是处于准地转平衡,那么初始场的平衡部分将表现为准地转模态,不平衡部分将表现为快速惯性重力波;在快波和慢波都同时存在时,如果快波的振幅不是很小,那么在初始的模式变量时间倾向中快波占主导(Kalnay, 2005)。

因此需要从初值中滤掉由于不平衡产生的虚假惯性重力波,否则作为惯性重力波上的所有初始信息将会丢失,还将会产生非常有干扰的预报;尽管惯性重力波有很大的水平散度,频散很快,但通过地转调整过程,经过一段时间后(地转调整的时间尺度是 f^{-1} ,约12 h)会演变衰减为准地转平衡解。由于这个原因,分析过程中使用平衡关系更可取,减少信息丢失。

1.3.3 资料同化发展历程中的初始化技术

在OI中对分析增量强制地附加地转约束只保证分析场中的一种近似平衡;这样一个简单的地转平衡不足以保证平衡的初始条件。OI的实践中发现,在OI分析之后进行非线性正规化初值化(NLNMI)(Machenhauer,1977;Baer et al,1977)是必要的。因此非线性正规化初值化,作为一个单独步骤,曾被广泛用于许多业务资料同化系统,因为它可以比简单的地转平衡更有效地大大减小初始条件中的惯性重力波的振幅。此外,Lynch等(1992)和Lynch(1997)引入的基于数字滤波的动力初始化是一种非常简单、有效的动力初始化,实质上消除了对NLNMI的需要。

Parrish等(1992)在三维变分方法中考虑目标泛函中加入一个“惩罚项”,把全球平衡方程作为弱约束加到目标泛函,之后NCEP全球模式的spin up现象(例如以初始12 h积分的降水量变化为指标)与OI相比减小了一个量级以上。换句话说,有了三维变分方法之后,分析循环中对初始时刻做单独的初始化步骤将不再必要。这是三维变分与标准的OI步骤加之随后的NLNMI相比的一个主要优势。它消除了人为地将分析分为二步:(1)分析步,所产生的分析场接近观测但不平衡;(2)初始化步,产生平衡场但与观测更加分离。

从初始化技术的演变可以看到:理想的资料分析同化方法应该使用所有的先验信息,使得所得到的分析场符合观测的同时也符合物理量场之间平衡关系;例如,在变分同化方法中(Lorenc,1986),实现这一平衡的初始化技术是内含于分析同化过程之中的部分。只是对于不能很好利用变量间平衡关系的先验信息的同化方法,在同化分析之后需要单独的初始化步骤,有助于得到符合变量间平衡关系的分析场。

此外,一个基于协调性和动力初值化想法的四维资料同化技术是牛顿连续松弛逼近技术(常称作Nudging技术)(Anthes,1974;Kistler et al,1975;Hoke et al,1976;朱国富,1999)。它的要点是在预报前模式积分的同化时段,在一个或几个模式预报方程中增加一个人为强迫项;该强迫项与方程中模式预报变量的观测实况值和模式值之差成正比,其作用是在同化时段的积分中使模式状态逐渐逼近观测状态。

2 具体同化方法中的一些等价性和共性特征

2.1 三维变分方法与最优插值方法的等价性

变分方法是一次性地使用同化时间窗(通常为6 h)内的所有观测。观测的时间信息不同利用方式有了三维变分和四维变分之分。三维变分,不考虑同化时间窗内这些观测的时间分布,而是时间维上近似地视为在同一个时间点,一般取在该同化时间窗的中间时刻,这个时间点就是分析时刻,也是 \mathbf{x}_b 和 \mathbf{B} 对应的有效时刻。四维变分,考虑在同化时间窗内 $[t_0, T]$ 各个观测的时间分布 t_i (第 i 个观测的时间),分析时刻取在该同化时间窗的起始时刻 t_0 ; \mathbf{x}_b 和 \mathbf{B} 的有效时刻也取在 t_0 。

最优插值方法的分析方程与三维变分同化的显式解有着相同的公式形式,即1.1.4节中的式(9)。这被称作Var与OI在解的形式上的等价性。这可以从两个方面理解:(1)原理上,如果一个随机变量的概率密度函数是正态分布的,这时它的均值和众数是相等的;因此,在背景场误差和观测误差都满足正态分布的假设下,作为大气状态的最优估计,OI以方差最小得到的均值和Var以似然概率最大得到的众数是相同的。(2)具体实施的导出过程上,在考虑观测算子 H 只是线性插值和不考虑观测算子误差和代表性误差的情况下,能够得到它们的解在形式上的等价性(Kalnay,2005)。

2.2 四维变分方法与扩展卡尔曼滤波方法的等价性

四维变分方法所进行的分析是起始时刻的状态

向量 $\mathbf{x}(t_0)$ 。该状态向量通过预报模式显式地传播到各个观测的时刻:

$$\mathbf{x}(t_i) = M_{t_0 \rightarrow t_i}[\mathbf{x}(t_0)] \quad (10)$$

由此得到对应 t_i 时刻观测的模拟值 $H[\mathbf{x}(t_i)]$ 。于是,对应四维变分,式(8)中的目标泛函写为:

$$J[\mathbf{x}(t_0)] = \frac{1}{2}[\mathbf{x}(t_0) - \mathbf{x}_b(t_0)]^T \mathbf{B}_{t_0}^{-1}[\mathbf{x}(t_0) - \mathbf{x}_b(t_0)] + \frac{1}{2} \sum_{t_i=t_1}^{t_{Ny}} \{H[\mathbf{x}(t_i)] - \mathbf{y}_o(t_i)\}^T \mathbf{R}_{t_i}^{-1} \{H[\mathbf{x}(t_i)] - \mathbf{y}_o(t_i)\} \quad (11)$$

式中 N_y 是观测空间的维数。因此,通过用预报模式作为一个强约束(即所要得到的分析场必须满足模式方程),当目标泛函中观测项份额是主导(足够的观测)时,四维变分是寻找这样一个初始条件,使得其预报最佳地拟合该同化时间窗内的那些观测。所以,四维变分的一个优点是可以通过预报模式能够显式地利用观测资料的时间维分布信息。

更重要的优点是,四维变分能够隐式地使背景场误差协方差矩阵 \mathbf{B}_{t_0} 也通过预报模式从背景状态的有效时刻 t_0 传播到各个观测的时刻 t_i (Fisher, 2001)。这可以用公式形式更清晰地说明(Bannister, 2008):由式(11),将观测算子 H 在 \mathbf{x}_b 附近做线性化近似,通过 $\nabla_{\mathbf{x}}(t_0) J[\mathbf{x}_a(t_0)] = 0$,能得到四维变分同化的显式解,再由矩阵恒等式 $\mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}$,得到:

$$(\mathbf{I} + \mathbf{M}\mathbf{B}\mathbf{M}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\mathbf{M}(\mathbf{x}_a - \mathbf{x}_b) = \mathbf{M}\mathbf{B}\mathbf{M}^T\mathbf{H}^T\mathbf{R}^{-1}\{\mathbf{y}_o - \mathbf{H}\mathbf{M}(\mathbf{x}_b)\} \quad (12)$$

式中,观测算子 H 及其切线性 \mathbf{H} 都是对于观测 y_o 的时间 t_i ,预报模式 M 及其切线性 \mathbf{M} 对应着 $t_0 \rightarrow t_i$ 的积分,而切线性的伴随 \mathbf{M}^T 对应着 $t_0 \leftarrow t_i$ 的反向积分。

可见,式(12)中所有的 \mathbf{B} 都以 $\mathbf{M}\mathbf{B}\mathbf{M}^T$ 出现,表示 \mathbf{B} (它本身的有效时刻是在 t_0 时刻)通过线性化的预报模式传播到 t_i 时刻。正是这个性质,如果假定:(1)预报模式是完美的,(2)初始时刻的背景误差协方差 \mathbf{B}_{t_0} 是正确的,可以证明:四维变分在同化时间窗最终时刻的分析场与扩展卡曼滤波在对应时刻的分析场是同样的(Kalnay, 2005)。这被称作四维变分分析和扩展卡曼滤波分析的等价性。这个等价性意味着四维变分能够隐式地使预报误差协方差从

\mathbf{B}_{t_0} 演变到时间窗最终时刻。根据四维变分和扩展卡曼滤波的分析方程的导出(Hamill, 2006),可以理解,对于同化相同的观测,这个等价性的适用条件包括(1)预报模式和观测算子是线性的,(2)不考虑预报模式误差,以及(3)背景场误差正态分布,且(4)在同化开始时刻有相同的背景场误差协方差。

需要指出的,四维变分中 \mathbf{B} 矩阵并不是无期限地演变下去,而只是在同化时间窗时段隐式地演变;在下一个同化循环开始时刻 \mathbf{B} 矩阵又返回到起初的静态矩阵。

2.3 同化方法中数学的主要作用

早期的分析同化方法,不论在思路上还是求解上,就包含着数学的内容,如:多项式函数拟合方法中所用的函数拟合和最小二乘方法求解,SCM 中所用的观测算子水平插值和迭代方法求解。这些数学内容的作用是如何把观测资料“插到”格点上这一分析功能实施实现的工具手段。

在 OI、Var 和 EnKF 中概率论和数理统计学的随机变量和估计理论成为它们的理论基础,由此 OI 和 EnKF 中以分析场的方差最小、Var 中以分析场的最大似然概率导出它们的分析方程的数学形式。也因此,自 OI 之后,资料同化有了基于估计理论的最优标准及其数学公式表述形式,从一项经验技巧工艺成为一门科学工程学科。

可见,资料同化从一开始就离不开数学,而且数学的主要作用在资料同化的发展中发生了根本的转变:从作为实施实现的工具手段到作为科学实质的理论基础。

当然,对于自 OI 之后的分析同化方法,除了作为理论基础的这一主要作用之外,数学也同时还有着作为各方法实施实现之工具手段的作用,如 OI 中的线性方程组求解,Var 中的变分全局求解,和 EnKF 中的序贯处理技术(Bishop et al, 2001)。

2.4 分析同化方法的共同特征

一种分析同化方法要能得到好的应用效果,从一开始就体现了二个方面的共同特征:(1)科学技术上需要多学科知识的综合应用,(2)实施工程上要做好系统工程的各个串联环节。它们揭示了自主研发同化系统的客观要求。

2.4.1 科学技术上需要多学科知识的综合应用

根据 Panofsky(1949)的经典文献“客观天气图分析”,多项式函数拟合方法,作为最早的客观分析方法,就能充分地体现这一特征。

首先在方法的公式表示上,假定气象物理量和三维空间坐标存在着解析函数的客观联系,考虑在相对小的区域二维的三次多项式函数可以令人满意地拟合观测。这体现了方法本身的物理问题和数学形式。

在具体实施上,由于所有观测受到观测误差和可能的局地涡旋影响,不应该准确地拟合观测,而总是需要一些平滑;这体现了对资料误差特征的理解。同时,方法所确定的多项式不可能指望能拟合局地涡旋,而是应该表示所分析的气象物理量场的大尺度特征;这体现了从天气动力学知识上理解平滑的意义。进一步地,所限定的拟合区域大小取决于需要平滑观测的程度,而平滑的程度取决于所分析的气象物理量场的性质:观测值比较不准或易受到非代表性涨落的变量(如风)应当比观测值较准的变量(如气压)进行更多平滑;这体现了对观测不确定性和天气动力学的理解。

在最后的实现上,使用多于待定多项式系数个数的观测按照最小二乘方法来确定多项式系数,得到可信的分析场,其中对于风场的分析比对于气压场的分析使用更多观测来考虑更多的平滑,并假定观测误差和局地涡旋是关于平滑的场正态分布,这样的最小二乘方法的处理会得到最可信的多项式;这些体现了对观测资料特征(不同的误差特征)、天气动力学和数学方法的理解。

可见,即使是最早客观分析的三次多项式拟合方法,就协调地考虑“平滑、大气中气压和风变量的特征,它们的观测准确性”等方面,以及使用数学的函数拟合和最小二乘方法的求解办法。这具体而充分地体现出:一种同化技术方法是对所用数学方法、天气动力学性质和资料特征的清晰理解基础上,来有效地、极致地使用当时的有限观测资料。

2.4.2 实施工程上要做好系统工程的串联环节

根据 Cressman(1959)的经典文献“一个业务客观分析系统”,作为最早的业务客观分析系统之一,它就首先强调指出,对于任一客观分析方案,实际应用中的主要问题之一是数据可靠性问题和数据错误

的识别和剔除。因此首先的工作是数据预处理。该分析系统在进行客观分析之前是一个数据自动处理系统,自动处理由电传打字机接收的来自北半球的资料数据,主要包括:(1)对于温度探空资料,进行静力学检验和可能的订正,删除不能订正的错误资料;(2)对于风探空资料,进行垂直一致性检验,删除错误资料;之后,(3)选择待分析的观测要素并按地理顺序分类;(4)并对该过程做图,控制监视资料的覆盖和错误资料的剔除。

业务同化系统,作为系统工程,还体现在同化系统设计和程序代码的规范性、文档化和技术支撑上。英国气象局“变分资料同化的科学文档”(Variational Data Assimilation Scientific Documentation)突出了业务同化系统设计中数据流组织的重要性;为了高效率的计算,大量工作涉及数据流组织。ECMWF的2011—2020年战略部署之“科学技术基础”(ECMWF,2011)把业务NWP中程序代码开发和质量放在突出的位置,给予相当多的投入(包括:观测数据处理和监控、资料同化、全球大气数值预报、全球海洋预报、图形显示、产品制作和检验等软件代码);特别是,程序软件设计视作集成不同模式系统的根基,程序代码的质量有着明确的界定(即计算上高效、方便灵活的模块化适应计算机环境快速发展的更新扩展、和文档记录完整)。这些支撑了其持续和合作发展。

可见,业务资料同化系统包含数据接收、数据预处理及监控、同化方法实现的程序软件工程(系统工程的设计和规范的数据流组织、程序模块化及科学技术文档)等接连相继、环环相扣的环节。这些环节是串联性质的,所以有着短板效应;任一环节的粗陋都将限制最后分析场的质量。这些环节构成了一个业务实施的系统工程。

3 结 论

本文沿着“分析同化问题提出”、“资料同化如何实施实现”、“同化结果有着什么内在要求”这一逻辑,综合阐述了资料同化作为一门学科的基本方面:(1)“逆问题求解和所需要的先验信息”,(2)“观测信息的输入和权重与传播平滑”,和(3)“同化结果的准确性和协调性”,并在此逻辑和内容的背景下涉及和

认识 NWP 分析问题的欠定性、四维变分的同化时间窗、背景场误差协方差、观测算子和初始化技术等概念及其作用。以期这一逻辑和这些基本方面有益于对资料同化在条理脉络上的一个整体理解。此外,由延长四维变分同化时间窗的实施实现,具体地揭示了:发展同化技术需要在理解基础上从实际现状出发。

本文还着眼于具体同化方法中的一些等价性和共性特征,归纳阐述了“变分方法与最优插值方法和扩展卡尔曼滤波方法的等价性”、“同化方法中数学的主要作用”,以及“同化方法中要应用综合学科知识”和“业务实施中要做好系统工程”这二个共同特征等方面。这二个共同特征具体地揭示了:需要科学上注重“理解”和应用综合学科知识来做对,需要工程上注重环环相扣的“基础与过程”和稳扎稳打而避免短板来做好。这对于从应用或移植他人的同化系统转变到自主研发自己的业务 NWP 同化系统,在“自主、持续、合作”发展的观念上,在严谨扎实的方式上,以及具体的系统工程内容上,有着直接的启发和借鉴的作用。

参考文献

陈希孺. 2009. 概率论与数理统计. 合肥:中国科学技术大学出版社, 102,260.

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,143,154,155-162.

朱国富. 1999. 观测资料同化与有限区模式初期降水预报. 北京大学学报(自然科学版),35(1):81-88.

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(4):456-463.

Anthes R A. 1974. Data Assimilation and Initialization of Hurricane Prediction Model. *J Atmosp Sci*,31:702-719.

Baer F, Tribbia J J. 1977. On complete filtering of gravity of modes through nonlinear initialization. *Mon Wea Rev*,105,1536-1539.

Bannister R N. 2008. A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quart J Roy Meteor Soc*,134:637,1951-1970.

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.

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

ECMWF. 2011. The ECMWF Strategy 2011 – 2020 Scientific and technical basis. (available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, UK).

Fisher M. 2001. Assimilation techniques (4): 4dVar // ECMWF Meteorological Training Course Lecture Series (available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, UK).

Hamill T M. 2006. Ensemble-based atmospheric data assimilation // *Predictability of Weather and Climate*. Cambridge: Cambridge University Press,124-156.

Hoke J E, Anthes R A. 1976. The initialization of numerical models by a dynamical initialization technique. *Mon Wea Rev*, 104: 1551-1556.

Kistler R E, McPherson R D. 1975. On the use of local wind correction techniques for four-dimensional data assimilation. *Mon Wea Rev*,103:445-449.

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

Lorenc A, Tibaldi S. 1980. The treatment of humidity in ECMWF's data assimilation scheme // *Atmospheric Water Vapor*. New York: Academic Press.

Lynch P. 1997. The Dolph-Chebyshev window: A simple optimal filter. *Mon Wea Rev*,125:655-660.

Lynch P, Huang X Y. 1992. Initialization of the HIRLAM model using a digital filter. *Mon Wea Rev*,120:1019-1034.

Machenhauer B. 1977. On the dynamics of gravity oscillations in a shallow water model, with application to normal-mode initialization. *Contrib Atmos Phys*,50:253-271.

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.

Talagrand O. 1999. A posteriori evaluation and verification of analysis and assimilation algorithms // *Proceedings of Workshop on Diagnosis of Data Assimilation Systems*. ECMWF.