张璐, 田向军, 刘宣飞, 等. 2017. 基于多重网格策略的 NLS-3DVar 资料融合方法及其在气温数据融合中的应用 [J]. 气候与环境研究, 22 (3): 271–288. Zhang Lu, Tian Xiangjun, Liu Xuanfei, et al. 2017. NLS-3DVar data fusion method based on multigrid implementation strategy and its application in temperature data fusion [J]. Climatic and Environmental Research (in Chinese), 22 (3): 271–288, doi:10.3878/j.issn.1006-9585.2016.16140.

# 基于多重网格策略的 NLS-3DVar 资料融合方法 及其在气温数据融合中的应用

张璐<sup>1,2</sup> 田向军<sup>2</sup> 刘宣飞<sup>1</sup> 师春香<sup>3</sup>

1 南京信息工程大学气象灾害预报预警与评估协同创新中心和气象灾害教育部重点实验室,南京 210044
 2 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
 3 中国气象局国家气象信息中心,北京 100081

摘要将多重网格策略引入NLS-3DVar (Non-linear Least Squares-based on Three-dimensional Variational Data Assimilation,非线性最小二乘三维变分同化)方法,进而应用于 2400 多个国家级气象观测站逐时气温数据和 NCEP 再分析气温数据的融合,得到中国区域空间分辨率 1°×1°,时间分辨率为6小时的气温融合产品。分别从单重网格(分辨率1°×1°)和双重网格(分辨率1°×1°)和双重网格(分辨率1°×1°)和双重网格(分辨率1°×1°)和双重网格(分辨率1°×1°)和双重网格(分辨率1°×1°)和和双重网格(分辨率1°×1°)和和2014年1~12月(4、5月除外)的独立检验数据考察 NLS-3DVar 气温融合产品质量,验证基于多重网格策略的 NLS-3DVar 方法的优越性。在单重网格下,与广泛应用于气象行业的 Cressman 插值产品(均方根误差和相关系数的年平均值分别为 1.961°C d<sup>-1</sup>和 0.924)相比, NLS-3DVar 产品全年始终具有最小的均方根误差和最大的相关系数,年平均值分别为 1.915°C d<sup>-1</sup>和 0.929;站点间误差分析进一步表明,NLS-3DVar 产品在大多数检验站点精度更高,在新疆、甘肃、云南、陕西等地区尤为突出;加入双重网格策略的 NLS-3DVar 产品与单重网格的 NLS-3DVar 产品误差对比显示,均方根误差年平均值分别为 1.649°C d<sup>-1</sup>和 1.711°C d<sup>-1</sup>,相关系数年平均值分别为 0.970 和 0.968,二者在均方根误差和相关系数的表现上都极为相似,即双重网格 NLS-3DVar 气温产品尽管对观测数据采取了稀疏化处理,但依旧维持了原有的产品精度,并且在计算效率上提高了 1 倍多。而与同样在双重网格下基于多尺度的 STMAS (Space-Time Multiscale Analysis System)算法相比,双重网格的 NLS-3DVar 产品在产品精度上同样占据优势,在计算效率上单位时次耗时与 STMAS 算法几乎相当。

 关键词
 融合气温产品
 Cressman 插值
 STMAS (Space-Time Multiscale Analysis System)
 对比评估

 文章编号
 1006-9585 (2017)
 03-0271-18
 中图分类号
 P40
 文献标识码
 A

 doi:10.3878/j.issn.1006-9585.2016.16140

# NLS-3DVar Data Fusion Method Based on Multigrid Implementation Strategy and Its Application in Temperature Data Fusion

ZHANG Lu<sup>1, 2</sup>, TIAN Xiangjun<sup>2</sup>, LIU Xuanfei<sup>1</sup>, and SHI Chunxiang<sup>3</sup>

1 Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters and Key Laboratory of Meteorological Disaster of Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044

2 International Center for Climate and Environment Sciences (ICCES), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029
 3 National Meteorological Information Center, China Meteorological Administration, Beijing 100081

收稿日期 2016-07-20; 网络预出版日期 2016-11-09

Funded by Special Fund for Meteorological Research in the Public Interest (Grant GYHY201506002), National Key Research and Development Program of China (Grant 2016YFA0600203)

作者简介 张璐,女,1991年出生,硕士研究生,主要从事数据同化研究。E-mail: wisheszhanglu@126.com

通讯作者 刘宣飞, E-mail: liuxf@nuist.edu.cn

资助项目 公益性行业(气象)科研专项重大项目 GYHY201506002,国家重点研发计划项目 2016YFA0600203

**Abstract** In this study, the authors first incorporate the multigrid implementation strategy into the non-linear least squares-based three-dimensional variational data assimilation system (NLS-3DVar), which is applied for temperature data fusion. A merged temperature dataset at 1° resolution and 6-hour interval is produced based on in situ observations at 2400 observational sites over China and NCEP (National Centers for Environmental Prediction) final global tropospheric analyses. Another set of independent validation data (from January to December except April and May in 2014) is used to evaluate the merged dataset. The dataset of NLS-3DVar is compared with the gridded data at 1° resolution produced by the widely used Cressman interpolation method. NLS-3DVar product always has lower RMSE (Root Mean Square Errors) of 1.961°C d<sup>-1</sup> and higher correlation coefficient of 0.924 compared to the dataset produced by Cressman interpolation. The precision of merged temperature product of NLS-3DVar is higher in most stations and independent of validation data, especially at those stations in Xinjiang, Gansu, Yunnan, Shanxi, and so on. The performances of NLS-3DVar based on both the single grid and multigrid strategies are also compared. Both RMSE and correlation coefficient have little differences. Although multigrid-based NLS-3DVar uses the sparse process, the precision is almost the same as that of single-grid based NLS-3DVar. However, its computational costs are greatly reduced due to the sparse process. Compared with the STMAS algorithm (Space–Time Multiscale Analysis System), multigrid-based NLS-3DVar performs better regarding the precision of product with almost the same computational efficiency.

Keywords Merged temperature analysis, Cressman interpolation, STMAS (Space–Time Multiscale Analysis System), Comparative assessment

# 1 引言

提高大气温度数据的精确度,合理准确地估计 温度变化趋势,对于基础天气预报及气候、农业与 环境研究均有重要意义。地面观测站数据具有较高 的可信度,但就中国区域而言,观测站点在空间分 布上存在明显的不连续,呈现东密西疏状态,对基 于地面观测的格点数据质量有较大影响(Chen et al., 2010; 熊秋芬等,2011)。对数据的融合处理近年 来得到广泛发展,所谓数据融合技术,是指将多平 台、多传感器的采集信息加以综合,进而获得关于 目标区规范化网格的标准化定量信息(苗春生等, 2015)。数据融合技术的发展主线,一直以来主要 是降水资料和海温资料。研究表明适当的融合方法 能有效发挥观测数据的优势,同时使得融合产品接 近观测值,以更合理的方式分布于格点中(苗春生 等,2015; 宇婧婧等,2015)。

气象观测站点数据作为一种空间数据,具有复杂的非线性动力学机制,在时空分布上纷杂多变,因此实际使用中通常需要将这种离散的、不规则的观测资料转换成规则的、可以用于数值模拟的初始场或者诊断分析所用的规则网格点资料。通常采用的方法即为"空间内插",常用的方法有:样条函数法、克里金(Kriging)插值法、距离平方反比法和Cressman插值法(Cressman,1959)等。克里金法计算量大、过程复杂,变异函数需要人为选定,而样条函数插值难以对误差进行估计,采样点稀少

时效果不好(冯锦明等,2004)。距离平方反比则 是计算比较便捷的插值方法,相比较于逐步订正的 Cressman 插值,Cressman 客观分析具有更小的相对 误差(冯锦明等,2004),在气象数据融合领域应 用广泛。

美国 NOAA (National Oceanic and Atmospheric Administration)下属的地球系统研究实验室(Earth System Research Laboratory, 简称 ESRL) 开发的解 决资料融合问题的局地分析预报系统(Local Analysis and Prediction System, 简称 LAPS) 包括 两种资料融合方法,一种是基于修正 Barnes 插值方 法,另一种是基于连续变分的方法,即 STMAS 方 法 (Space-Time Multiscale Analysis System)。 STMAS 方法是由 Xie et al. (2005, 2011) 发展起来 的新的资料融合方法,结合了统计分析(如 3DVar 或 EnKF)和传统客观的分析(如 Barnes 分析)的 优势,通过采用多重网格策略依次对不同尺度的观 测信息进行分析, 消除长波信息与短波信息之间的 干扰。多重网格策略主要是对观测数据的处理上采 用稀疏化。稀疏化是指由于观测资料分布不均匀, 对于每个网格内的观测数据,观测点个数不同,尤 其是粗网格,这时可以先将一个网格内的观测数据 插值到该网格中心点,做成一个数据观测点。 STMAS 方法对于背景误差协方差和观测误差协方 差的考虑较为简化,认为二者比值为常数,融合产 品的精度也有待检验。

无论是融合方案还是空间插值,都是对观测数

据或者各种平台数据的进一步优化处理。数据融合 方面的研究,对气温数据的融合研究甚少,因此本 文通过将基于多重网格策略的 NLS-3DVar (Non-linear least squares-based 3D variational data assimilation system) 方法(Tian and Feng, 2015) 融 合气温数据,并与 Cressman 插值的结果进行对比 分析,进一步验证 NLS-3DVar 方法的合理性和优越 性。第二节介绍了本文研究所用到的数据资料,多 重网格策略下的三维变分同化方法,包括 NLS-3DVar 算法和 STMAS 算法的理论依据,以及 气象行业广泛应用的 Cressman 插值算法,同时给 出了试验设计和检验方法。第三节比较了单重网格 下 NLS-3DVar 和 Cressman 插值的融合产品。第四 节对比单、双重网格下 NLS-3DVar 产品的差异,并 与 STMAS 算法进行比较。第五节总结了本研究得 到的结论。

## 2 资料与方法

### 2.1 大气温度数据

融合的温度数据源包括地面观测和 NCEP (National Centers for Environmental Prediction) 再 分析数据(Final Global Tropospheric Analyses)。其 中,地面观测温度数据为中国 2400 多个国家级地 面气象观测站(图 1a)的逐小时观测值(00 时至 次日00时,协调世界时,下同),经过严格的质量 控制(任芝花等, 2015)。NCEP 再分析数据覆盖全 球范围,本文中选取中国区域(16°N~55°N,70°E~ 135°E),数据空间分辨率 1°×1°,时间分辨率为6 小时,本文以此数据为背景场资料,进行下述研究。 所用数据时间序列为 2014 年 1 月 1 日至 2014 年 12 月 31 日。本文将该时空,该区域的 NCEP 再分析 数据作为背景场资料,将中国 2400 多个国家级地 面气象观测站的气温资料与之融合。用于独立样本 检验的数据是全国 30000 多个全国自动气象站(图 1b) 逐时气温数据,由于自动站数据在 2014 年 4 月、5月存在大量缺测,因此在独立检验时去除4 月和 5 月。由于 2400 多个国家级地面观测站数据 在准确度和站点分布上较区域自动站更精准更合 理,故用于融合,区域自动站数据用于独立检验, 具体的检验方法详见 2.5。

### 2.2 多重网格的三维变分同化方法

多重网格的三维变分同化方法可以依次从大

尺度和小尺度上采用连续的三维变分更准确地使 误差最小化。多重网格三维变分较传统三维变分的 优势在于依次从观测提取长波和短波信息组成一 个从粗网格到细网格的连续的多尺度三维变分方 案。其代价函数形式为

$$J^{(k)}(\mathbf{x}^{\prime(k)}) = \frac{1}{2} (\mathbf{x}^{\prime(k)})^{\mathrm{T}} (\mathbf{B}^{(k)})^{-1} (\mathbf{x}^{\prime(k)}) + \frac{1}{2} [\mathbf{H}^{(k)}(\mathbf{x}_{\mathrm{b}}^{(k)} + \mathbf{x}^{\prime(k)}) - \mathbf{y}_{\mathrm{obs}}^{(k)}]^{\mathrm{T}} \mathbf{R}^{(k)-1} \times \mathbf{H}^{(k)}(\mathbf{x}_{\mathrm{b}}^{(k)} + \mathbf{x}^{\prime(k)}) - \mathbf{y}_{\mathrm{obs}}^{(k)}], \ (k = 1, \ 2, \ ..., \ S), \ (1)$$

其中,  $x'^{(k)}$  是分析增量,  $x_b^{(k)}$  是背景场向量,  $y_{obs}^{(k)}$  为观测向量,  $B^{(k)}$ 代表背景误差协方差矩阵,  $R^{(k)}$  是 观测误差协方差矩阵, 上标 T 与 -1 分别表示矩阵的转置与逆。 $H^{(k)}$ 为观测算子, k 代表当前计算的 是第 k 重网格, S 是网格总层数。关于多重网格三 维变分的实现, 在 Xie et al. (2005), Li et al. (2008) 和 Li et al. (2010) 文章中有更多的说明。

本文所使用的数据融合方法——基于多重网格的 NLS-3DVar(即非线性最小二乘法三维变分同 化方法)方法是对 Tian and Feng (2015)所提出的 NLS-4DVar 的三维版本,通过将多重网格策略引 入,并利用 NCEP 再分析数据构造集合样本近似背 景误差协方差矩阵 *B*,采用扩展局地化策略(Liu et al., 2008; Tian et al., 2016),形成一种计算更高效更 准确的多重网格的三维变分同化方法。

基于多重网格的连续变分同化方法 STMAS 方 法是一种简化的多重网格三维变分方案,其简化主 要体现在对背景误差协方差和观测误差协方差的 常数比处理上。下面进一步分别介绍这两种方法。 2.2.1 基于多重网格策略的 NLS-3DVar 方法

NLS-3DVar 方法的主要思想是利用集合策略 求解以下代价函数的最小值:

$$J(\mathbf{x}') = \frac{1}{2} (\mathbf{x}')^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x}') + \frac{1}{2} [\mathbf{H} (\mathbf{x}_{\mathrm{b}} + \mathbf{x}') - \mathbf{y}_{\mathrm{obs}}]^{\mathrm{T}} \times \mathbf{R}^{-1} [\mathbf{H} (\mathbf{x}_{\mathrm{b}} + \mathbf{x}') - \mathbf{y}_{\mathrm{obs}}], \qquad (2)$$

其中, **x**'<sub>j</sub> = **x**<sub>j</sub> - **x**<sub>b</sub> 是分析增量, **x**<sub>b</sub> 是背景场向量 (在本文即为 NCEP 分析场), **x**<sub>j</sub> 是样本, *j*=1, 2, …, *N* (*N* 为样本个数,本研究取 *N*=61), *H* 为观测算子 (在本研究中为插值算子), *B* 代表 背景误差协方差(待估计), *R* 是观测误差协方差矩 阵,上标 T 与-1 分别表示矩阵的转置与逆。多重网 格策略的 NLS-3DVar 方法实际是在 *k* 重网格上采 用 NLS-3DVar 方法求解代价函数,为简便起见,公 式(2)中省略表示网格层数的上标k。

$$\mathbf{y}_{\rm obs}' = \mathbf{y}_{\rm obs} - \boldsymbol{H}(\boldsymbol{x}_{\rm b}), \qquad (3)$$

 $y_{obs}$ 是观测向量,而 $y'_{obs}$ 为观测新息增量。当网格 层数 k=1时,(3)式中的 $x_b$ 为 $x_b$ ;当k=2,...,S时, 为上一重网格的分析解 $x_a$ 。假设分析增量 $x'_a$ 是模 式扰动 $P_x = (x'_1, x'_2, ..., x'_N)$ 的线性组合:

$$\boldsymbol{x}_{a}^{\prime} = \boldsymbol{P}_{x}\boldsymbol{\beta}, \qquad (4)$$

这里  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_N)^T$  为系数矩阵。同样有  $\boldsymbol{P}_y = (\boldsymbol{y}_1', \boldsymbol{y}_2', ..., \boldsymbol{y}_N'), 其中, \boldsymbol{y}_j' = \boldsymbol{H}(\boldsymbol{x}_b + \boldsymbol{x}_j') - \boldsymbol{H}(\boldsymbol{x}_b)$ 为观测扰动。

将(4)式带入(2)式,代价函数形式改写为  

$$J(\boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}(\boldsymbol{P}_{x})^{\mathrm{T}}\boldsymbol{B}^{-1}(\boldsymbol{P}_{x})\boldsymbol{\beta} + \frac{1}{2}[\boldsymbol{H}(\boldsymbol{x}_{\mathrm{b}} + \boldsymbol{P}_{x}\boldsymbol{\beta}) - \boldsymbol{y}_{\mathrm{obs}}]^{\mathrm{T}} \times \boldsymbol{R}^{-1}[\boldsymbol{H}(\boldsymbol{x}_{\mathrm{b}} + \boldsymbol{P}_{x}\boldsymbol{\beta}) - \boldsymbol{y}_{\mathrm{obs}}].$$
(5)

根据 Evensen (2004) 和 Tian et al. (2011),背 景误差协方差  $B = (P_x)(P_x)^T / (N-1)$ 进行估算并代 入(5)式,引入 $R = R_+^{1/2} (R_+^{1/2})^T$ ,得到以下关于 $\beta$ 的非线性最小二乘最优化问题:

$$J(\boldsymbol{\beta}) = \frac{1}{2} Q(\boldsymbol{\beta})^{\mathrm{T}} Q(\boldsymbol{\beta}), \qquad (6)$$

其中,

$$Q(\boldsymbol{\beta}) = \begin{pmatrix} \boldsymbol{R}_{+}^{-1/2} \left[ L'(\boldsymbol{P}_{x}\boldsymbol{\beta}) - \boldsymbol{y}_{obs}' \right] \\ \sqrt{N-1}\boldsymbol{\beta} \end{pmatrix}.$$
 (7)

由此可用以下的高斯—牛顿迭代法(Dennis and Schnabel, 1996)进行迭代求解:

$$\boldsymbol{\beta}^{i+1} = \boldsymbol{\beta}^{i} - \left[ (\boldsymbol{P}_{y})^{\mathrm{T}} \boldsymbol{R}^{-1} (\boldsymbol{P}_{y}) + (N-1) \boldsymbol{I} \right]^{-1} \cdot \left[ (\boldsymbol{P}_{y})^{\mathrm{T}} \boldsymbol{R}^{-1} (L'(\boldsymbol{x}_{a}^{\prime,i}) - \boldsymbol{y}_{obs}') + (N-1) \boldsymbol{\beta}^{i} \right], \quad (8)$$

其中,  $L'(\mathbf{x}_{a}^{\prime,i}) = \mathbf{H}(\mathbf{x}_{b} + \mathbf{x}_{a}^{\prime,i}) - \mathbf{H}(\mathbf{x}_{b})$ , *i* 为迭代次数, *I* 是  $N \times N$  的单位矩阵。关于该方法的详细介绍请 见 Tian and Feng (2015)。

对于集合估计的背景误差协方差矩阵有  $B \approx (P_x)(P_x)^T / (N-1),本研究的重点在于对集合样本的构造:本研究中,以NCEP 再分析资料作为背景场,同时作为采样的样本集:本研究的样本个数为61,采样时以需要融合数据的时刻为准,首先选取该时刻的背景场数据作为第一个样本,然后向该时刻的前一个时刻继续采样,依次向前一时刻推进,选取30个样本;同理,向该时刻的后一个时刻继续采样,依次向后推进,同样选取30个样本,自此完成61个样本的采样,称为历史采样法(Wang et al., 2010),图2给出了历史采样法的示意图。$ 

完成采样之后,计算 $x'_{j} = x_{j} - x_{b}$  (*j*=1, 2, ...,

N)得到模式扰动 **P**<sub>x</sub>,代入上式可求得背景误差协 方差 **B**。采用扩展局地化方案对集合误差协方差矩 阵 **B** 进行局地化。

对局地化相关矩阵 *C* 做类似的分解 (Liu et al., 2009; Tian et al., 2016):  $C = C'C'^{T}$ ,这里  $C' \in \Re^{m \times r}$ , *m* 是模式状态总维数, *r* 是局地化模态数。进而计 算局地化的模式扰动  $P_{x,\rho} \in \Re^{m \times N \times r}$ :

$$P_{x,\rho} = P_x < e > C' = (C' \circ P_{x,1}^*, C' \circ P_{x,2}^*, \cdots, C' \circ P_{x,N}^*),$$
(9)

式中,"o"表示两个矩阵的 Schur 积(Tian and Feng, 2015), <e > 代表列扩展运算符(申思, 2015), 它表 示 两 个 具 有 相 同 行 数 的 矩 阵 的 运 算 ,  $P_{xl}^*(l=1, 2, ..., N)$ 是与C'的列数相同的矩阵,且每 一列都是 $P_x$ 的第l列。同理,对观测扰动 $P_y$ 也进行 局地化操作:

$$\boldsymbol{P}_{y,\rho} = \boldsymbol{P}_{y} < \boldsymbol{e} > \boldsymbol{C}' = (\boldsymbol{C}' \circ \boldsymbol{P}_{y,1}^{*}, \boldsymbol{C}' \circ \boldsymbol{P}_{y,2}^{*}, \cdots, \boldsymbol{C}' \circ \boldsymbol{P}_{y,N}^{*}),$$
(10)

其中,  $P_{y,\rho} \in \Re^{m_o \times (N \times r)}$ ,  $m_o$ 是观测状态变量总维数,  $P_y = (y'_1, y'_2, ..., y'_N)$ (其中,  $y'_j = H(x_b + x'_j) - H(x_b)$ , j = 1, 2, ..., N)。最后(8)式可以进一步写为

$$\boldsymbol{\beta}^{i+1} = \boldsymbol{\beta}^{i} - \left[ (\boldsymbol{P}_{y,\rho})^{\mathrm{T}} \boldsymbol{R}^{-1} (\boldsymbol{P}_{y,\rho}) + (N-1) \boldsymbol{I} \right]^{-1} \times \left[ (\boldsymbol{P}_{y,\rho})^{\mathrm{T}} \boldsymbol{R}^{-1} (L'(\boldsymbol{x}_{\mathrm{a}}^{\prime,i}) - \boldsymbol{y}_{\mathrm{obs}}^{\prime}) + (N-1) \boldsymbol{\beta}^{i} \right]. \quad (11)$$

最后采用共轭梯度法迭代求得 $\beta^{i+1}$ 及  $x_a^{\prime,i+1} = P_{x,o}\beta^{i+1}$ 。

按照多重网格策略,一方面要对每一重网格下 用于融合的观测数据做稀疏化处理,即将一个网格 内的观测数据插值到该网格中心点,做成一个数据 观测点,本研究中采用的是距离反比插值;另一方 面,从粗网格转换到细网格时,其观测误差协方差 *R*也应当做相应的调整:

$$\boldsymbol{R}_{k}^{2} = \sum_{i=1}^{n_{k}} a_{i}^{2} \boldsymbol{R}_{0}^{2}, \ k=1, 2, ..., S,$$
(12)

其中,  $R_0$  为初始的观测误差,本研究取 0.5°C,  $R_k$ 是第 k 重网格下的观测误差协方差, S 是网格总层 数, a 是对某一个网格内观测数据做稀疏化处理时, 各个观测点插值到网格中心时的插值系数,  $n_k$ 是第 k 重网格下每个网格内观测站点的数目; 而  $B_k$ 在每 层网格下的构造,则取决于不同分辨率网格下选取 的样本。

粗网格向细网格转换的桥梁是粗网格下生成 的分析场结果,通过将其插值到细网格的分辨率 下,作为细网格的背景场,继续进行该分辨率下的



图 1 (a) 中国 2400 多个国家级气象站分布,(b) 中国 3 万余个自动观测站(包括国家站和区域站)分布





图 2 历史采样法示意图

Fig. 2 Sketch map of historical sampling

数据融合。

2.2.2 STMAS 算法

按照基于多重网格的连续变分同化方法 STMAS 方法对于 **B** 和 **R** 的简化思路,(1)式的代 价函数简写为(Li et al., 2013)

$$J^{(k)}(X^{(k)}) = \frac{1}{2} X^{(k)T} X^{(k)} + \frac{1}{2} [H^{(k)} X^{(k)} - Y^{(k)}]^{T} \times O^{(k)-1} [H^{(k)} X^{(k)} - Y^{(k)}], k = 1, 2, ..., S,$$
(13)

其中,  $X^{(k)}$  是分析增量,  $Y^{(k)}$ 为新息增量,  $O^{(k)}$  是 观测误差协方差矩阵,  $H^{(k)}$  是网格空间到观测空间 的插值算子(即观测算子), k 代表当前计算的是第 k 重网格, S 是网格总层数。并假设背景误差方差为  $\varepsilon_b^2$ , 观测误差方差为 $\varepsilon_o^2$ , 新的观测误差协方差 O 的对角线元素为 $\varepsilon_o^2 / \varepsilon_b^2$ , 而初始的背景误差协方差 矩阵可简化为用相对于背景项不变的观测项的比率 来确定。对于第一重网格 $Y^{(1)} = Y^{obs} - H^{(1)}X_b$ , 之后 每重网格:

 $Y^{(k)} = Y^{(k-1)} - H^{(k-1)}X^{(k-1)}, k = 1, 2, ..., S,$  (14)  $X^{(k-1)}$ 是求解代价函数 $J^{(k-1)}$ 得到的解。最后的分析 场为

$$X_{a} = X_{b} + \sum_{k=1}^{S} X^{(k)}.$$
 (15)

本研究中取网格层数 S = 2,即两重网格,粗网格分辨率为 2°×2°,细网格分辨率为 1°×1°。在双 重网格下对 NLS-3DVar 和 STMAS 算法的气温融合 产品做比较。

### 2.3 Cressman 插值

一般的数据融合方法,都是在单重网格下进行,直接将观测数据融合生成最细分辨率的产品。 已被广泛应用于各种气候诊断分析和数值模拟研究中的 Cressman 插值具有相对于其他插值方案更小的相对误差。本研究中,在最细分辨率 1°×1°的单重网格下,对比 NLS-3DVar 方法和 Cressman 插值的融合产品。下面介绍 Cressman 插值方法。

首先给定第一猜测场,一般网格的第一猜测场 由背景场给出:  $T_i^0 = T_i^b$ , *i* 表示第 *i* 个网格点。第一 次估计以后,用实际观测场逐步订正第一猜测场:

$$T_{i}^{n+1} = T_{i}^{n} + \frac{\sum_{k=1}^{K_{i}} W^{n}(r_{ik})(T_{k}^{obs} - T_{i}^{n})}{\varepsilon^{2} + \sum_{k=1}^{K_{i}} W(r_{ik})}, \quad (16)$$

式中,*T<sub>i</sub><sup>n</sup>*是第*i*个待插值格点上一次订正后的数值, *n*为订正次数,*n*=0时即为背景场。对于第*i*个待 插值格点,影响半径 r 内包含  $K_i$  个观测, { $T_k^{obs}, k = 1, 2, ..., K_i$ }。 $W^n(r_{ik})$ 是影响半径内,第 k 个观测与第 i 个格点之间的权重因子,数值在 0.0~1.0 之间变化。 $\varepsilon^2 = \sigma_0^2 / \sigma_b^2$ 是观测误差方差  $\sigma_0^2$ 与背景误差方差 $\sigma_b^2$ 的比率。 $\varepsilon^2$ 的值决定了背 景场和观测数据对分析场结果的影响程度,  $\varepsilon^2 = 0$ 时,表示极度相信观测场的分析场,因为 数据融合是基于高度相信观测场,故本文使用 Cressman 方法时采用 $\varepsilon^2 = 0$ 。此法最重要的是对 权重函数的确定:

$$W(r_{ik}) = \begin{cases} \frac{r_{ik}^{2} - d_{ik}^{2}}{r_{ik}^{2} + d_{ik}^{2}}, & d_{ik} < r_{ik}, \\ 0, & d_{ik} \ge r_{ik}, \end{cases}$$
(17)

其中, *d<sub>ik</sub>* 是第 *k* 个站点到待插值格点 *i* 的距离。影 响半径 *r* 的选取原则是由近及远进行扫描。

本文中采用该插值方法,将 2400 多个国家级 地面气象观测站点数据插值至分辨率为1°×1°的网 格中,影响半径取 0.5°至 4°(55 km~220 km)。

### 2.4 试验设计

本文设置了三组对比试验来检验基于多重网格策略的 NLS-3DVar 融合产品的质量。表1给出了 三组试验设计方案。通过试验一在单重网格下的对 比,验证采用 NLS-3DVar 方法得到的气温融合产品 在精度上的优势。试验二对同一方案 NLS-3DVar 的不同网格数目的产品的对比,直观展示多重网格 策略引入后产品精度和计算效率的变化。试验三有 两个目的,一是检验 STMAS 算法中对误差矩阵的 简化处理是否合适,二是对比这两种方案在同时采 用双重网格策略的情况下,产品精度和计算效率的 差异。通过三组试验,充分证明基于多重网格策略 的 NLS-3DVar 资料融合方法所生成的融合产品在 精度和计算效率上的优越性。

表 1 三组试验设计 Table 1 Design of three groups of experiments

					时间
	网格数目	融合方法1	融合方法2	空间分辨率	分辨率
试验一	单重网格	NLS-3DVar	Cressman	1°×1°	6 小时
试验二	单重/双重	单重	双重	粗网格: 2°×2°	6 小时
		NLS-3DVar	NLS-3DVar	细网格: 1°×1°	
试验三	双重网格	NLS-3DVar	STMAS 算法	粗网格: 2°×2°	6 小时
				细网格: 1°×1°	

### 2.5 检验标准及方法

为了更好的评估 NLS-3DVar 方法、Cressman

插值以及 STMAS 算法的分析精度,需要进行独立 样本检验,即选取完全独立于融合产品的气温资料 作为参考。本研究所用经过严格质量控制(任芝花 等,2015)的30000多个全国自动气象观测站(图 1b) 逐时气温资料, 剔除用于数据融合所用到的 2400 多个国家级地面气象站(图 1a)数据。为进 一步保证数据质量,剔除了四个研究时次(00时、 06 时、12 时、18 时)均为缺测的区域站点,将剩 余的区域自动气象站气温资料作为独立检验数据。 为了进一步保证检验数据的有效与合理性,筛选了 国家站所在网格内(1°×1°的网格)的自动台站资 料作为独立样本,称其为"有效独立样本"(18000 多个)。鉴于背景场 NCEP 数据时间分辨率为 6 小 时(00时、06时、12时、18时),所以融合逐小 时观测数据后的融合产品亦只有4个时次,分辨率 同样为 1°×1°。独立样本检验时间序列为 2014 年 1~12月(4月、5月由于独立检验样本缺测,亦去 除)。为了与融合产品比较,检验时,通过距离反 比插值将融合产品插值至有效独立样本所在站 点,计算统计误差。定量误差估计的统计指标主 要有均方根误差(R<sub>RMSE</sub>),规范化的标准差(S<sub>SDV</sub>) 和相关系数(Ccov)。SDV表示融合产品结果与观 测的相近程度, 越接近 1, 效果越好。定义公式如 下:

$$C_{\rm cov} = \frac{\sum_{i=1}^{M} (x_{i,\rm obs} - \overline{x_{\rm obs}})(y_{i,\rm model} - \overline{y_{\rm model}})}{\sqrt{\sum_{i=1}^{M} (x_{i,\rm obs} - \overline{x_{\rm obs}})^2} \sqrt{\sum_{i=1}^{M} (y_{i,\rm model} - \overline{y_{\rm model}})^2}},$$
(18)

$$S_{\rm SDV} = \frac{\sqrt{\sum_{i=1}^{M} (y_{i,\text{model}} - \overline{y_{\text{model}}})^2 / M}}{\sqrt{\sum_{i=1}^{M} (x_{i,\text{obs}} - \overline{x_{\text{obs}}})^2 / M}},$$
(19)

$$R_{\rm RMSE} = \sqrt{\frac{\sum_{i=1}^{M} (x_{i,\rm obs} - y_{i,\rm model})^2}{M}},$$
 (20)

其中, M 为样本总数,根据计算变量的不同, M 分为时间序列样本总数和站点序列样本总数两种,  $x_{i,obs}$ 是第 i 个样本的有效独立检验观测值,  $y_{i,model}$  是 第 i 个样本的融合产品数据,  $\overline{x_{obs}}$  是 M 个有效独立 观测样本的平均值,  $\overline{y_{model}}$  是 M 个融合产品数据的 平均值。

# 3 单重网格下气温数据融合结果评估检验

### 3.1 误差随时间的变化

图 3 展示了 2014 年 7 月 13 日 06 时的气温场 结果。直观来看,两种产品的结果都很接近检验样 本,但仍可以发现在华南沿海福建地区 Cressman 产品在 36°C 以上的高温地区基本没有体现,而在 内蒙古地区 28~32°C 的高温区范围略显不足,在 四川陕西交界处 32~36°C 的高温范围的体现也有 所欠缺。

为了进一步明确这两种方法之间的优劣差异, 计算了整个时间序列上均方根误差(图 4)和相关 系数(图 5)的逐日变化曲线。计算样本为整个中 国区域存在的有效独立检验样本。两种产品相比于 FNL 背景场 (图中记为 X<sub>b</sub>,下同) 在均方根误差和 相对系数上都有了较为明显的改善,故二者具有可 对比性。全年 Cressman 和 NLS-3DVar 的产品的均 方根误差都比较接近且随时间的变化较为稳定。仍 可看出 NLS-3DVar 产品在全年始终具有较小的均 方根误差, Cressman 产品的均方根误差总比 NLS-3DVar 产品大 0.05~0.1°C d<sup>-1</sup>, 7 月、8 月的差 距较其他各月更为明显。图 5 相关系数的逐日变化 进一步肯定了 NLS-3DVar 方法的优势。数据显示, NLS-3DVar 产品的相关系数最大,基本维持在 0.9 上下,其次是 Cressman 产品,略小于 NLS-3DVar 产品,在6月、7月、8月(图 5d、e、f)的差距 较其他各月更为明显。由此可得,相比于 Cressman 产品,NLS-3DVar产品在全年都有着更小的均方根 误差和更大的相关系数。

综合均方根误差和相关系数两个变量,我们可 以在 Taylor 图 (Taylor, 2001)(图 6)中更为直观 的通过这两个变量同时对两种产品进行评估。从原 点到待评估数据各点的辐射长度是均方根误差 (RMSE),待评估点和底边两条射线形成的方位角 是相关系数。距离原点和底边越近,效果越好。结 合 Taylor 图中所用数据的数据列表(表 2),可以更 为直观和清晰地看出,NLS-3DVar产品在每个月的 RMSE 和相关系数上都更具优势。NLS-3DVar 产品 的 RMSE 与紧随其后的 Cressman 产品相比,小 0.05~0.1°C d<sup>-1</sup>。相关系数的统计也表明,NLS-3DVar产品的相关系数在0.88~0.96之间,Cressman



图 3 2014 年 7 月 13 日 06 时全国气温场(单位: °C)分布: (a)独立检验样本; (b) NLS-3DVar 产品; (c) Cressman 产品 Fig. 3 Spatial distributions of temperature over China at 0600 UTC 13 July 2014: (a) The independent validation data; (b) NLS-3DVar (Non-linear Least Squares-based on 3DVar) data; (c) Cressman data



图 4 2014 年 (4、5 月除外) NLS-3DVar 产品、Cressman 产品和背景场与独立检验样本的均方根误差(单位: °C d<sup>-1</sup>)的逐日变化 Fig. 4 Daily variations of RMSEs (units: °C d<sup>-1</sup>) between NLS-3DVar product, Cressman product, background field ( $X_b$ ) and the independent validation data in 2014 except for April and May

产品在 0.855~0.955 之间,背景场 $X_{\rm b}$ (即 NCEP 再分析数据)在 0.77~0.93 之间。

综合以上各图表,在融入同样数据的情况下, NLS-3DVar 产品优于 Cressman 产品。与此同时可 以看出, Cressman 产品的结果与 NLS-3DVar 产品 的结果很接近,无论从 RMSE 还是相关系数, NLS-3DVar 产品都只有微弱的优势,接下来本研究 从站点间对比的角度对这两种方法进行进一步的 对比检验。

### 3.2 误差的空间分布

前一小节所给出对比的变量都是从整个气温 场的平均值出发,展示的都是站点平均值在时间序 列下的变化,下面的检验将从站点间逐一的对比展 开,纵向比对 NLS-3DVar 产品和 Cressman 产品的 精度。图 7 所示的 Taylor 图综合考虑规范化的标准 差(SDV)和相关系数作为评估指标。用离观测点 (图 7 中 REF 点)的距离来表示效果优劣,距离越 近,效果越好。作为独立检验样本的 18000 多个站 点,统计显示,NLS-3DVar 产品与 Cressman 产品 相比,53.8%的站点的 SDV 前者比后者更接近 1, 即与检验样本更相近。图 7 中我们筛选了这 53.8% 的优势站点中,NLS-3Dvar 产品与 Cressman 产品的 SDV 的差值的绝对值大于 0.1 的站点,共计 42 个。 NLS-3DVar 产品的 SDV 在 0.9~1.3 之间,Cressman 产品的在 0.88~1.55 之间,NLS-3DVar 产品的站点 在更接近 1.0 线的同时,都具有更大的相关系数, 即站点——对应比较之下,NLS-3DVar 的产品更接 近 REF 点,效果更好。同样 RMSE 的统计也显示, 有 59.7%的站点 NLS-3DVar 产品比 Cressman 产品 有更小的均方根误差。图 8 中的 168 个站点选取的 是 Cressman 产品和 NLS-3DVar 产品的 RMSE 差值 在  $1.0 \sim 1.5^{\circ}$ C d<sup>-1</sup> 之间的独立检验样本站点, NLS-3DVar 产品的站点的 RMSE 大多数分布在  $0.6 \sim 1.6^{\circ}$ C d<sup>-1</sup>之间,而 Cressman 产品则在  $1.5 \sim$  $3.2^{\circ}$ C d<sup>-1</sup>之间,站点的一一对比可以看出, NLS-3DVar产品的优势站点在具有更小的RMSE同 时,有更好的相关系数。

图 9 进一步给出了这 168 个站点的分布状况, 主要分布在新疆、甘肃、云南等地区,在江苏和山 西也有部分站点。类似地给出了 Cressman 产品和



图 5 同图 4, 但为相关系数的逐日变化 Fig. 5 As in Fig. 4, but for daily variations of correlation coefficients



图 6 2014 年 (4、5 月除外) NLS-3DVar 产品、Cressman 产品、背景场 *X*<sub>b</sub> 与独立检验样本的月平均 RMSE (单位: ℃ d<sup>-1</sup>) 和相关系数的 Taylor 图

Fig. 6 Taylor diagram of monthly averaged RMSEs (units: °C d<sup>-1</sup>) and correlation coefficients between NLS-3DVar product, Cressman product, background field  $X_b$  and the independent validation data in 2014 except for April and May



图 7 2014 年 (4、5 月除外) NLS-3DVar 产品、Cressman 产品与独立检验样本的 SDV 和相关系数的 Taylor 图 Fig. 7 Taylor diagram of SDV (standard deviation) and correlation coefficients between NLS-3DVar product, Cressman product and the independent validation data in 2014 except for April and May

表 2 2014 年 (4、5 月除外) NLS-3DVar 产品、Cressman 产 品、背景场  $X_b$  与独立检验样本的月平均 RMSE (单位: °C d<sup>-1</sup>) Table 2 Monthly averaged RMSEs (units: °C d<sup>-1</sup>) between NLS-3DVar product, Cressman product, background field  $X_b$  and the independent validation data in 2014 except April and May

	月平均 RMSE/°C d <sup>-1</sup>			
月份	NLS-3DVar 产品	Cressman 产品	$X_{ m b}$	
1月	2.137	2.193	3.107	
2 月	1.888	1.931	2.993	
3 月	1.967	2.041	2.862	
6 月	1.810	1.863	2.731	
7 月	1.940	1.982	2.821	
8月	1.862	1.908	2.774	
9月	1.797	1.836	2.652	
10 月	1.914	1.955	2.704	
11 月	1.812	1.861	2.664	
12 月	2.008	2.058	2.956	

NLS-3DVar 产品的 RMSE 差值大于 1.5°C d<sup>-1</sup>的站 点(68个站点)的 Taylor 图(图10)和站点分布 图(图11)。图10中可以比较清晰地逐一对比同一 站点,NLS-3DVar 产品在均方根误差比 Cressman 产 品小1.5的同时,相关系数更大。这68个站点的分 布(图11)同样显示这些站点主要分布在新疆、甘 肃地区。同时图11还给出了所有独立检验站点的 RMSE 差值。正值表示 Cressman 产品相比 NLS-3DVar 产品有更大的 RMSE,负值相反。可以看出,整个站 点分布以正值居多,且主要在 0~0.8之间,1.5以上 的站点数也多于-1.5的站点,因此,NLS-3Dvar 产品 在独立检验样本数据中,均方根误差更小。

综合以上分析,从整个气温场的时间序列来 看,NLS-3DVar产品相比于原来的FNL 背景场有 了很大的改善,并且通过与 Cressman 产品的对比 检验,NLS-3DVar产品有更小更稳定的均方根误差 和更大的相关系数。尽管与 Cressman 产品相比 NLS-3DVar产品的优势微弱,但是,通过进一步逐 个站点的 RMSE 和相关系数的比较,依然可以看出 NLS-3DVar产品具有更高的精度,特别是在新疆、 甘肃、云南地区,在山东、江苏、浙江也有比较明 显的优势。

# 4 双重网格下气温数据融合结果评 估检验

### 4.1 统计误差的对比

经过前面的讨论,我们已经明确 NLS-3DVar

产品在精度上的优越性,那么接下来加入多重网格的策略。通过单重网格NLS-3DVar产品和双重网格NLS-3DVar产品的对比,进一步考察观测稀疏化处理给融合产品带来的影响。同时STMAS算法也采用同样分辨率下的双重网格策略,与双重网格的NLS-3DVar产品进行对比。

图 12 是 2014 年 7 月 2 日 00 时的气温融合结 果。从气温场直观来看,三种产品的结果比较相近, 与检验样本相比,从站点丰富的区域来看,三者在 气温各个范围都有比较好的表现。2014年10月逐 6小时的均方根误差(图 13a)和相关系数(图 13b) 显示,单重网格 NLS-3DVar 产品与双重网格产品的 RMSE 极为接近,两条曲线几乎呈现重合状态,个 别时次二者有 0.1~0.2℃ d<sup>-1</sup> 的差异。换言之,通 过在 NLS-3Dvar 方法的同化过程中引入双重网格 策略,虽然在同化过程中对观测数据做了稀疏化处 理,但是,对产品的精度几乎没什么影响。图 13b 中相关系数的展示,也同样证实了单重网格 NLS-3DVar 产品与双重网格 NLS-3DVar 产品精度 的相近性。而对比双重网格 NLS-3DVar 产品与 STMAS 算法产品的均方根误差可以看出, STMAS 算法产品的 RMSE 始终大于双重网格 NLS-3DVar 产品大约 0.1~0.2℃ d<sup>-1</sup>, 且相关系数小于后者 0.01~0.02。图 14 逐月对 RMSE 和相关系数做了平 均,从全年时间序列和整体平均水平来看,STMAS 算法产品无论从均方根误差还是从相关系数来看, 都不如双重网格 NLS-3DVar 产品。由此可见其对于 背景误差协方差B和观测误差协方差R的常数比处 理的确比较简单和粗糙。从单重网格 NLS-3DVar 产品和双重网格 NLS-3DVar 产品的误差对比来看, 二者在6、7、8月的差距较其他各月更为明显,但 RMSE 的差值也只有不到 0.1℃ d<sup>-1</sup>,相关系数亦只 相差不足 0.01。因此,可认为对于 NLS-3DVar 方法, 双重网格策略的引入,观测稀疏化的处理,对产品 精度并没有造成影响。

#### 4.2 计算效率的对比

产品精度和计算效率是考证方法优劣的两个 重点。图 15 是 2014 年全年平均每月一个同化时次 的运行时间统计图。计算均在曙光 I620-G20 2U 双 路机架式服务器上进行,且计算过程均采用串行。 柱状图显示,6、7、8 月双重网格 NLS-3DVar 方法 单位时次同化运行耗时最少,约为 50 s,STMAS 算法与之近乎相同,单重网格 NLS-3DVar 方法在这

三个月的耗时虽然在全年来看已经是最短耗时,约 为130 s, 却仍然是双重网格 NLS-3DVar 方法计算 耗时的2倍多。全年平均数据显示,单重网格 NLS-3DVar 方法的计算耗时平均为 184.16 s, 双重

3 期

网格 NLS-2DVar 方法为 63.45 s, STMAS 算法为 66.98 s。显然双重网格策略下,NLS-3DVar 方法计 算耗时更少且优势明显。

综合以上结果, NLS-3DVar 方法通过引入双重



图 8 Cressman 产品、NLS-3DVar 产品与独立检验样本的 RMSE 和相关系数的 Taylor 图(只给出 Cressman 产品与 NLS-3DVar 产品的 RMSE 差值在 1.0~1.5℃ d<sup>-1</sup> 范围的站点)

Fig. 8 Taylor diagram of RMSEs and correlation coefficients between NLS-3DVar product, Cressman product and the independent validation data (only those stations with RMSE differences between Cressman and NLS-3DVar products within  $1.0 \sim 1.5$  °C d<sup>-1</sup> are counted)



图 9 Cressman 产品、NLS-3DVar 产品与独立检验样本的 RMSE 差值在 1.0~1.5℃ d<sup>-1</sup>范围的站点分布图 Fig. 9 Spatial distribution of stations with RMSE differences within the range of 1–1.5°C d<sup>-1</sup> between Cressman and NLS-3DVar products

网格策略,在同化过程中对观测数据采取稀疏化处理,最终产品精度在保持原单重网格下产品精度的同时,大幅度地提高了计算效率。与 STMAS 算法的对比显示,简单的对背景误差协方差和观

测误差协方差做常数比处理得到的产品效果不及 双重网格 NLS-3DVar 产品,但计算效率上二者不 相上下,也再次证明了多重网格策略在计算效率 上的优势。



图 10 Cressman 产品、NLS-3DVar 产品与独立检验样本的 RMSE 和相关系数的 Taylor 图(只给出 Cressman 产品与 NLS-3DVar 产品的 RMSE 差值 大于 1.5°C d<sup>-1</sup> 的站点)

Fig. 10 Taylor diagram of RMSEs and correlation coefficients between NLS-3DVar product, Cressman product and the independent validation data (only those stations with RMSE differences between Cressman and NLS-3DVar products greater than  $1.5^{\circ}$ C d<sup>-1</sup> are counted)







图 12 2014 年 7 月 2 日 00 时全国气温场(单位: ℃)分布: (a)有效独立检验样本; (b) 单重网格 NLS-3DVar 产品; (c) 双重网格 NLS-3DVar 产品; (d) STMAS 算法产品

Fig. 12 Spatial distributions of temperature over China at 0000 UTC 2 July 2014: (a) The independent validation data; (b) single NLS-3DVar product; (c) multigrid NLS-3DVar product; (d) STMAS (Space–Time Multiscale Analysis System) algorithm product



图 13 2014 年 10 月单重网格 NLS-3DVar 产品、双重网格 NLS-3DVar 产品、STMAS 算法产品与独立检验样本的(a) 均方根误差和(b) 相关系数 Fig. 13 (a) RMSEs and (b) correlation coefficients between products of single NLS-3DVar, multigrid NLS-3DVar, STMAS algorithm product and the independent validation data during October 2014



图 14 2014 年单重网格 NLS-3DVar 产品、双重网格 NLS-3DVar 产品、STMAS 算法产品与独立检验样本的逐月平均(a)均方根误差和(b)相关 系数

Fig. 14 Monthly mean (a) RMSEs and (b) correlation coefficients between single NLS-3DVar product, multigrid NLS-3DVar product, STMAS algorithm product and the independent validation data in 2014



图 15 2014 年单重网格 NLS-3DVar 方法、双重网格 NLS-3DVar 方法和 STMAS 算法逐月平均单位时次同化过程耗时(单位: s)统计 Fig. 15 Statistics of monthly mean time-consuming (units: s) of a single-time assimilation by single NLS-3DVar, multigrid NLS-3DVar, and STMAS algorithm in 2014

# 5 结论与讨论

将多重网格策略引入 NLS-3DVar 方法,并将其 应用于气温数据的融合,得到数据空间分辨率 1°×1°,时间分辨率为6小时的融合产品。通过和 Cressman 插值以及 STMAS 算法的对比,证明采用 基于多重网格策略的 NLS-3DVar 方法得到的气温 融合产品具有更高的产品精度和更高的计算效率。 首先在单重网格下通过独立性检验,对比了 NLS-3DVar产品与气象行业中广泛应用的Cressman 插值 产品误差的时间分布,并进一步从站点间的误差对 比了 NLS-3DVar产品和 Cressman 插值产品的优劣。 在此基础上,对 NLS-3DVar 方法引入了双重网格策 略,对比分析了单重网格 NLS-3DVar 产品、双重网格 NLS-3DVar 产品和 STMAS 算法产品的均方根误差、相关系数和运行耗时,全面评估了基于多重网格策略的 NLS-3DVar 方案得到的气温融合产品。主要结论如下:

(1)单重网格下,独立样本检验表明,在统计 误差的时间变化上,NLS-3DVar 产品全年具有最 高的相关系数和最小的均方根误差,年平均值分别 为 1.915°C d<sup>-1</sup> 和 0.929,且二者随时间的变化较 为稳定,Cressman 插值产品略差于 NLS-3DVar 产 品,年平均值分别为 1.915°C d<sup>-1</sup> 和 0.929。因 此,NLS-3DVar 产品相比于 Cressman 插值产品具 有更高的精度。

(2) NLS-3DVar 产品和 Cressman 产品站点间

均方根误差和相关系数的对比,进一步肯定了 NLS-3DVar方法的优势。独立样本检验中,有53% 以上的站点显示 NLS-3DVar产品具有更小的均方 根误差和更接近1的标准差的同时,具有更高的相 关系数。NLS-3DVar的气温融合产品主要在新疆、 甘肃、云南、陕西等地相比于 Cressman 产品有更 高的精度。

(3)双重网格 NLS-3DVar 气温融合产品相较于 原来的单重网格产品,精度基本没有发生明显的变 化,只在个别时次略差于单重网格产品,而与同样 是双重网格的 STMAS 算法的产品对比,均方根误 差 和相关系数都明显处于优势,也因此说明 STMAS 算法中对于背景误差协方差 *B* 和观测误差 协方差 *R* 的常数比处理比较粗糙。通过单位时次同 化耗时的进一步对比,充分体现了双重网格 NLS-3DVar 方法的优势。全年平均数据显示,单重 网格 NLS-3DVar 方法的计算耗时平均为 184.16 s, 双重网格 NLS-2DVar 方法为 63.45 s, STMAS 算法 为 66.98 s。双重网格 NLS-3DVar 方法在维持原本 产品精度保持优势的前提下,大幅度地提高了计算 效率,节省运行时间。

通过以上研究,基于多重网格策略的 NLS-3DVar 方法应用于气温数据的融合产品在精度和计 算效率上的双重优势充分证明了该方案的优越性, 其基于非线性的最小二乘迭代算法,只通过很少的 迭代次数,便得到了比 Cressman 插值和 STMAS 算 法更好的结果。

### 参考文献(References)

- Chen D L, Ou T H, Gong L B, et al. 2010. Spatial interpolation of daily precipitation in China: 1951–2005 [J]. Advances in Atmospheric Sciences, 27 (6): 1221–1232, doi:10.1007/s00376-010-9151-y.
- Cressman G P. 1959. An operational objective analysis system [J]. Mon. Wea. Rev., 87 (10): 367–374, doi:10.1175/1520-0493(1959)087<0367: AOOAS>2.0.CO;2.
- Dennis J E Jr, Schnabel R B. 1996. Numerical Methods for Unconstrained Optimization and Nonlinear Equations (Classics in Applied Mathematics) [M]. Philadelphia: SIAM, 378pp.
- Evensen G. 2004. Sampling strategies and square root analysis schemes for the EnKF [J]. Ocean Dynamics, 54 (6): 539–560, doi:10.1007/s10236-004-0099-2.
- 冯锦明,赵天保,张英娟. 2004. 基于台站降水资料对不同空间内插方 法的比较 [J]. 气候与环境研究, 9 (2): 261–277. Feng Jinming, Zhao Tianbao, Zhang Yingjuan. 2004. Intercomparison of spatial interpolation based on observed precipitation data [J]. Climatic and Environmental Research (in Chinese), 9 (2): 261–277, doi:10.3969/j.issn.1006-9585.

2004.02.004.

- Li W, Xie Y F, Deng S M, et al. 2010. Application of the multigrid method to the two-dimensional Doppler radar radial velocity data assimilation [J].
  J. Atmos. Oceanic Technol., 27 (2): 319–332, doi:10.1175/2009JTECHA 1271.1.
- Li W, Xie Y F, He Z J, et al. 2008. Application of the multigrid data assimilation scheme to the China Seas' temperature forecast [J]. J. Atmos. Oceanic Technol., 25 (11): 2106–2116, doi:10.1175/2008JTECHO510.1.
- Li W, Xie Y F, Han G J. 2013. A theoretical study of the multigrid three-dimensional variational data assimilation scheme using a simple bilinear interpolation algorithm [J]. Acta Oceanologica Sinica, 32 (3): 80–87, doi:10.1007/s13131-013-0292-6.
- Liu C S, Xiao Q N, Wang B. 2009. An ensemble-based four-dimensional variational data assimilation scheme. Part II: Observing system simulation experiments with advanced research WRF (ARW) [J]. Mon. Wea. Rev., 137 (5): 1687–1704, doi:10.1175/2008MWR2699.1.
- 苗春生, 程远, 王坚红, 等. 2015. 中国风云卫星与海洋卫星近海 SST 资 料融合技术及应用研究 [J]. 地球科学进展, 30 (10): 1127–1143. Miao Chunsheng, Cheng Yuan, Wang Jianhong, et al. 2015. Data fusion of offshore SST from China FY and HY2 satellites and its application [J]. Advances in Earth Science (in Chinese), 30 (10): 1127–1143, doi:10.11867/j.issn.1001-8166.2015.10.1127.
- 任芝花, 张志富, 孙超, 等. 2015. 全国自动气象站实时观测资料三级质 量控制系统研制 [J]. 气象, 41 (10): 1268–1277. Ren Zhihua, Zhang Zhifu, Sun Chao, et al. 2015. Development of three-step quality control system of real-time observation data from AWS in China [J]. Meteorological Monthly (in Chinese), 41 (10): 1268–1277, doi:10.7519/j.issn.1000-0526.2015.10.010.
- 申思. 2015. 降维投影四维变分同化方法的改进及在 GRAPES 全球模式
  中的初步应用 [D]. 中国科学院大学博士学位论文. Shen Si. 2015.
  Improvements for dimension-reduced projection 4DVar and its preliminary application in GRAPES-GFS [D]. Ph. D. dissertation (in Chinese), University of Chinese Academy of Sciences.
- Taylor K E. 2001. Summarizing multiple aspects of model performance in a single diagram [J]. J. Geophys. Res., 106 (D7): 7183–7192, doi:10.1029/2000JA002008.
- Tian X J, Feng X B. 2015. A non-linear least squares enhanced POD-4DVar algorithm for data assimilation [J]. Tellus A, 67: 25340, doi:10.3402/tellusa.v67.25340.
- Tian X J, Feng X B, Zhang H Q, et al. 2016. An enhanced ensemble-based method for computing CNOPs using an efficient localization implementation scheme and a two-step optimization strategy: Formulation and preliminary tests [J]. Quart. J. Roy. Meteor. Soc., 142 (695): 1007–1016, doi:10.1002/qj.2703.
- Tian X J, Xie Z H, Sun Q. 2011. A POD-based ensemble four-dimensional variational assimilation method [J]. Tellus A, 63 (4): 805–816, doi:10. 1111/j.1600-0870.2011.00529.x.
- Wang B, Liu J J, Wang S D, et al. 2010. An economical approach to four-dimensional variational data assimilation [J]. Advances in Atmospheric Sciences, 27 (4): 715–727, doi:10.1007/s00376-009-9122-3.
- Xie Y, Koch S, McGinley J, et al. 2011. A space-time multiscale analysis

system: A sequential variational analysis approach [J]. Mon. Wea. Rev., 139 (4): 1224–1240, doi:10.1175/2010MWR3338.1.

- Xie Y, Koch S E, McGinley J A, et al. 2005. A sequential variational analysis approach for mesoscale data assimilation [C]//Proceedings of the 21st Conference on Weather Analysis and Forecasting/17th Conference on Numerical Weather Prediction. Washington, D. C.: American Meteor Society.
- 熊秋芬, 黄玫, 熊敏诠, 等. 2011. 基于国家气象观测站逐日降水格点数 据的交叉检验误差分析 [J]. 高原气象, 30 (6): 1615–1625. Xiong

Qiufen, Huang Mei, Xiong Minquan, et al. 2011. Cross-validation error analysis of daily gridded precipitation based on China meteorological observation [J]. Plateau Meteorology (in Chinese), 30 (6): 1615–1625.

宇婧婧, 沈艳, 潘旸, 等. 2015. 中国区域逐日融合降水数据集与国际降 水产品的对比评估 [J]. 气象学报, 73 (2): 394-410. Yu Jingjing, Shen Yan, Pan Yang, et al. 2015. Comparative assessment between the daily merged precipitation dataset over China and the world's popular counterparts [J]. Acta Meterorologica Sinica (in Chinese), 73 (2): 394-410, doi:10.11676/qxxb2015.033.