张海玲，王斌，毛嘉富. 2009. 陆面气温资料的一种推广三维变分同化方法及其在简化情形的检验 [J]. 气候与环境研究，14 (3)：273－283. Zhang Hailing，Wang Bin，Mao Jiafu. 2009. A Generalized approach of three-dimensional variational data assimilation for land surface air temperature and its tests under simplification cases [J]. Climatic and Environmental Research (in Chinese)，14 (3)：273－283.

# 陆面气温资料的一种推广三维变分同化方法及其在简化情形的检验

张海玲[1,2] 王 斌[1] 毛嘉富[1]

1 中国科学院大气物理研究所大气科学和地球流体力学数值模拟国家重点实验室，北京 100029
2 中国科学院研究生院，北京 100049

**摘 要** 发展了一种同化窗口为1个月的陆面气温推广三维变分资料同化方法，并建立了逐日资料到月平均资料的观测算子。作为对该同化方法的初步检验，以2000年1月中国陆地区域时间分辨率低、空间分辨率较高且在高原地区与台站观测资料比较接近的CRU（Climate Research Unit）月平均陆面气温资料作为观测，以时间分辨率高、空间分辨率较低且在高原地区与台站资料有较大误差的NCEP逐日气温资料作为背景场，采用背景误差协方差矩阵（简称***B***矩阵）的两种简化对角形式进行同化试验，得到高时空分辨率的格点陆面气温分析场，并进行对比分析和均方根误差检验。结果显示，由两种简化形式得到的逐日资料和月平均资料的质量均得到改善，在我国东南和中部大部分地区与台站观测基本一致，在青藏高原、新疆等台站稀疏地区与地形对应良好，为M-SDGVM模型或其他陆面过程模式高质量陆面气象驱动场的准备提供了一种新的手段。尤其在***B***矩阵中考虑了方差逐日变化的简化形式所得到的分析场整体上要优于不考虑的，这为最终在***B***矩阵中考虑协方差（即对推广三维变分同化不作简化）进一步改进同化效果打下了基础。

**关键词** 推广三维变分同化 NCEP资料 CRU资料 陆面气温
**文章编号** 1006-9585（2009）03-0273-11 **中图分类号** P40 **文献标识码** A

# A Generalized Approach of Three-dimensional Variational Data Assimilation for Land Surface Air Temperature and Its Tests under Simplification Cases

ZHANG Hailing[1, 2], WANG Bin[1], and MAO Jiafu[1]

1 *State Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing* 100029
2 *Graduate University of Chinese Academy of Sciences, Beijing* 100049

**Abstract** A generalized approach of three-dimensional variational data assimilation (G-3DVar) with 1-month assimilation window for land surface air temperature data was developed. As preliminary tests on this approach, two simplified forms of background error covariance matrix (***B*** matrix for short) were applied to generate a high space-

**收稿日期** 2008-03-31收到，2009-03-02收到修定稿
**资助项目** 中国科学院重要方向项目KZCX3-SW-230和公益性行业（气象）科研专项GYHY(QX)2007-25
**作者简介** 张海玲，女，1982年出生，硕士生，主要从事资料同化和陆面模型应用研究。E-mail：zhanghl@mail.iap.ac.cn
**通讯作者** 王斌，博士生导师，研究员，E-mail：wab@lasg.iap.ac.cn

time resolution grid analysis field of land surface air temperature in January 2000 in China through incorporating the CRU monthly mean data into the NCEP daily reanalysis data based on an observation operator associating the daily data and monthly mean data. The CRU monthly mean land surface air temperature data have low temporal resolution, but higher spatial resolution and better approximation to the station observations on the Tibetan Plateau, while the NCEP daily reanalysis data have high temporal resolution, but lower spatial resolution and worse approximation to station observations in these areas. A comparison shows that the qualities of both the daily and monthly mean data of the generated grid analysis fields from both simplified forms are all improved, which are basically consistent with the station observations in the south eastern and central areas of China and well match the topography in the Tibetan Plateau, Xinjiang and so on, where the observation stations are sparse. It indicates that the G-3DVar is a new efficient approach for preparing meteorological driving field for land ecological model or land surface model. Further comparison also shows that the analysis field from the simplified form with daily variance in the ***B*** matrix is better than that from the simplified without considering daily variance in the ***B*** matrix. It is an important basis for further improvement of assimilation performance of G-3DVar without any simplification, i. e. covariance is considered in the ***B*** matrix.

**Key words** generalized 3DVar, NCEP data, CRU data, land surface air temperature

# 1 引言

## 1.1 气象驱动场

对于陆面生态模型或陆面过程模式，其模拟性能除了与它们自身固有的不确定性有关以外，还与陆面气象驱动场的质量及时间和空间分辨率有着密切的关系。因此，生成一套中国区域的高时空分辨率的陆面气象要素场对于改善陆面模型在中国区域的模拟精度有着不可忽视的作用。近些年来，一些学者在这方面进行了有益的探讨和研究（林忠辉等，2002；蔡福等，2005）。由于气象站点往往远少于模式格点数目，无法提供足够的资料来形成完整的模式格点场，加之中国地形复杂，站点分布不均匀，使问题变得更加复杂。为了解决这一问题，前人已经做过大量工作。在模拟研究中国区域的陆面生态系统变化规律时，多数的模型使用中国气象局的台站观测通过不同的内插方法插到模式需要的空间分辨率（庄立伟等，2003；冯锦明等，2004）来获得气象驱动场；也有一些模型使用英国的 CRU（Climate Research Unit）月平均资料，通过天气发生器或其他方法生成需要的时间分辨率资料（毛嘉富，2006）。

但是，近来的一些研究结果认为（林忠辉等，2002；蔡福等，2005；Kalnay，2005），对于空间数据的内插存在几点不足：首先，一种普适的最优空间内插法是不存在的，对于不同的空间变量，在不同地域和不同时间尺度上，所谓的“最优”内插法是相对的。也就是说，每一次插值，对于不同的要素和时空尺度，都要重新验证，找到最优的插值方法。其次，中国地形条件复杂，必然会对气象要素产生很大影响，插值方法不能很好地考虑地形的影响。再次，对于观测稀疏的地区，向相对密集的格点插值，必然是个欠定问题。青藏高原是全球海拔最高、地形最为复杂的高原，它对整个东亚和全球的天气和气候都产生着深远的影响。青藏高原占我国陆地面积的 1/4，而且其下垫面的动力热力结构复杂，因此高原陆表的气象要素场也就成了研究高原、中国区域、乃至全球天气气候的一个非常重要的驱动场。然而，正是这个重要的区域，由于其海拔太高及其他原因，其上的台站分布稀少，甚至在 5 000 m 海拔以上几乎没有站点。这就给空间插值的可行性和准确性带来了很大的困难和挑战。

基于观测的 CRU 格点资料与实际观测非常相似，但其时间尺度是月，时间分辨率粗，而大多数陆面模型要求的时间分辨率至少是日或者更高，所以难以满足要求。而 NCEP 再分析资料虽然有逐日或更高时间分辨率的资料，但空间分辨率不如 CRU 高，尤其在高原地区，与台站观测资料差别较大，因而也无法满足陆面模型的要求。因此，

有人通过天气发生器（廖要明等，2004）将 CRU 资料生成逐日资料。然而，天气发生器主要用于气候模拟研究方面，对于时间分辨率较高的模型而言，显得太粗糙，误差较大。

### 1.2　3DVar 方法简介

毛嘉富（2006）把引进的 SDGVM（Sheffield Dynamic Global Vegetation Model）在中国区域的模拟进行了改进，并发展了改进版本的 M-SDGVM（Modified SDGVM）模型。M-SDGVM 在中国区域的综合模拟和评估表明，该模型对中国区域具有很好的模拟能力，能够合理模拟出中国区域叶面积指数、净初级生产力及土壤湿度的时空分布格局。但是，毛嘉富也同时指出，M-SDGVM 以及其他同类动态全球植被模型的发展和应用还处于初级阶段，仍存在很多不足和挑战。其中，改进模型所使用的气象驱动场，是提高模型模拟能力的重要途径之一。M-SDGVM 使用 CRU 逐月格点资料为气象驱动场，在时间上通过“天气发生器”将逐月值转换为日的分辨率。因此，如何使用合理有效的资料同化方法，改进上述简单的“插值”处理方式，提高气象驱动场精度，从而改进模型的模拟能力是本研究的切入点。

从上述分析可知，如何生成一套可信的、时空分辨率较高的区域陆面气象资料是一个具有理论和实际意义的问题。本文试图找到一种合理的、能够综合利用现有各种资料的方法。就目前的研究现状和未来的发展趋势看，资料同化（Rutherford，1972；Bouttier et al.，1999；Fisher，2001）应该是最有效的解决方法之一，它们能够合理利用不同精度、不同时空分布或时空分辨率、不同来源的各种资料，将它们有机融合在一起，得到规则格点上的初始场或驱动场。其中，三维变分资料同化（3DVar）是国际上最普遍使用的资料同化方法之一，它综合利用背景场和观测信息，使得测量分析变量与背景场和观测场距离的代价函数达到最小，而得到一个最优分析场。与单纯的台站插值得到格点要素场相比，这样的分析场具有更好的全局协调性和精度。

然而，3DVar 在用于气象资料同化时，一般只能同化同一时间的资料，或把分析时刻前后 3 h 的观测都当作分析时刻的观测进行同化（Wang et al.，2006），却无法同化以月平均方式存在的观测资料。如何把月平均的气象观测资料同化到逐日的气象背景场中去成为了一个具有挑战性的难题。本研究正是围绕解决这个难题而开展的。

本文拟提出一种推广三维变分资料同化方法（英文缩写 G-3DVar），为融合 NCEP 和 CRU 两种资料生成一套高质量陆面气象驱动场提供有效工具。但作为检验方法有效性的第一步，本文在两种简化情形下，即使用考虑方差是否随日变化的两种对角 $\boldsymbol{B}$ 矩阵，以 2000 年 1 月的 NCEP 资料为背景场和 CRU 资料为观测进行同化试验，其目的只是为了考察 G-3DVar 在简化情形下同化性能的好坏，为下一步在 G-3DVar 中使用考虑了协方差的非对角 $\boldsymbol{B}$ 矩阵并最终应用于生成更加优化和合理的多年 1～12 月高质量陆面气象驱动场奠定基础。

文章的第二部分主要介绍同化与检验中所要用到的资料以及它们的质量和特点；第三部分主要介绍一种推广的三维变分资料同化（G-3DVar）方法及其简化形式；第四部分给出了同化试验结果及其与实际的台站观测对比分析；最后一部分是结论和讨论。

## 2　资料及资料处理

### 2.1　台站分布及中国地形特点

图 1 是中国地形特点及 740 个观测站点分布。根据地形和站点分布特点，可将中国分为 3 个区域：I 区为东南、中部及东北地区，该区台站分布密集，地势相对平坦；II 区为青藏高原以北的新疆地区，该区地势起伏大，台站相对稀疏，尤其是在盆地、沙漠、山脉等地形，基本上没有站点；III 区为青藏高原地区，主要特点为地形高，台站很少，海拔 5 000 m 以上的区域没有站点。

地形的复杂性必然影响插值方法的准确性。对于 I 区而言，插值受地形影响不大，结果基本可信。但是对 II 区和 III 区这种台站稀疏的情况则不然。以 III 区为例，由于海拔 5 000 m 以上没有站点，且这个区域面积很大，格点数目多，格点上的值依靠周围海拔较低站点的插值；但二者的海拔高度相差近千米甚至更多，因此插值的结果实际是插到了高地形的“地下”，高于实际的地面

温度。同样，沙漠和盆地得到的插值结果实际是插到了低地形的“空中”，低于实际的地面温度。这些地区的台站向格点的插值实际上是个欠定的问题，其观测信息量少，结果可信度低。

## 2.2 CRU 格点资料

CRU 资料是一种基于观测并考虑了地形作用的月值（即月平均，下同）格点资料（Mitchell et al.，2004，2005；闻新宇等，2006），因而弥补了 II 区和 III 区台站稀疏的不足。图 2 是 1951～2001 年 中国区域 51 年平均的 1 月 CRU 温度场，图 3 是相应的中国区域 740 个台站 51 年平均的 1 月月平均温度通过 Cressman 插值得到的温度场。

通过对比 CRU 资料和中国台站观测资料，可以验证 CRU 资料的可靠性。从图 2 和 3 可以看到，在广大中东部地区，CRU 的温度场与台站观测基本一致；在西部地区，CRU 的温度场的高低温度区域和中心都基本与台站观测符合，只是区域的大小、形状并不完全一样。相比之下，CRU 和观测的地形更为吻合，其高低温度带从北至南与阿尔泰山脉、准格尔盆地、天山山脉、祁连山脉、塔里木盆地、塔克拉玛干沙漠、青藏高原等地形一一对应。也就是说，CRU 修正了缺少观测的地形的影响。由于 CRU 资料与台站观测的一致性以及与地形更好的对应关系，用它作为同化的观测场是可行的。

根据以上对比分析，本文拟用 2000 年 1 月的 CRU 2 m（即 2 m 高度，下同）温度格点资料作为同化的观测场，其空间分辨率为 0.5°×0.5°，范围覆盖整个中国区域（17.75°N～53.75°N，73.25°E～135.25°E），空间总格点数为 9 125，时间维数为 1。以下简称该资料为 CRU 观测。

## 2.3 NCEP 资料

研究认为（苏志侠等，1999；徐影等，2001；赵天保等，2004），NCEP 资料的 2 m 温度场在中国区域普遍偏低，东南部与观测接近，东北地区相对较差，而青藏高原和西北西部则与观测差异较大。对于高原地区较大的差异，可能是因为站

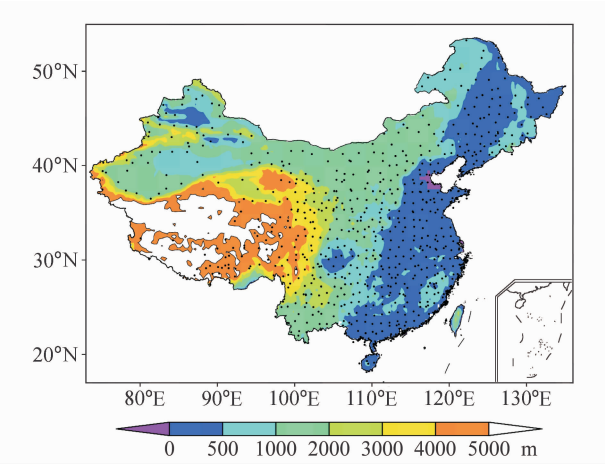



图 1 中国地形及 740 个观测站点分布（阴影为海拔）

Fig. 1 Terrain and distributions of 740 observation station in China (shadings denote altitude)

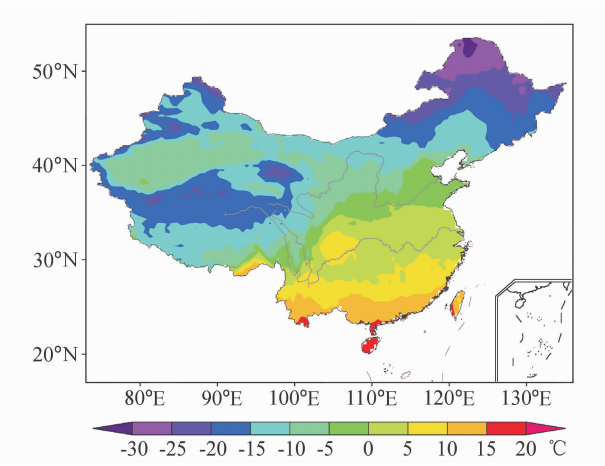



图 2 1951～2001 年中国区域 CRU 51 年平均的 1 月月平均陆面气温

Fig. 2 51-year-mean Jan land surface air temperature of CRU data from 1951 to 2001 in China

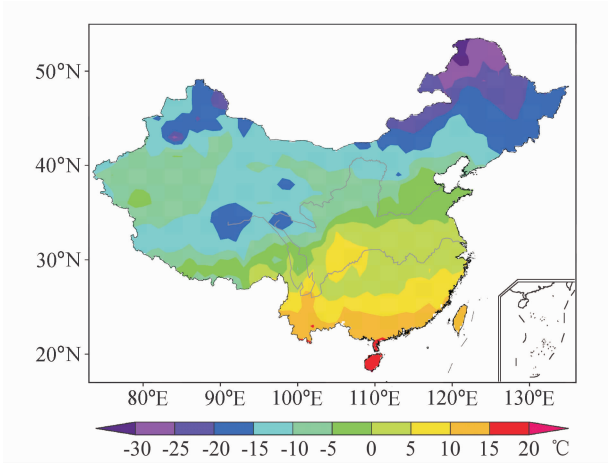



图 3 1951～2001 年中国区域 740 个台站 51 年平均的 1 月月平均陆面气温

Fig. 3 51-year-averaged monthly mean land surface air temperature of Jan of 740 station observation data from 1950 to 2001 in China

点稀疏和地形影响造成。不过 NCEP 资料亦有吸引人的优点，因它具有逐日或更高的时间分辨率，这是 CRU 资料所无法比拟的。

图 4、5、6 分别是 2000 年 1 月中国区域月平均的 NCEP 2 m 高度温度场、CRU 温度场和台站观测温度场。通过图的比较也发现，NCEP 资料确实在东南大部分地区与台站观测接近，而在高原地区偏低差异较大，在东北地区偏高，期望通过同化主要修正这两个地区的偏差。NCEP 资料尽管在高原以及新疆大部分地区与 CRU 资料存在差别，但是作为再分析资料，它在高地形或低地形的处理上也比台站观测的结果要好，温度场的走势和形状基本上也和地形符合。考虑了地形作用的 CRU 资料能够在一定程度上修正高原地形带来的误差。

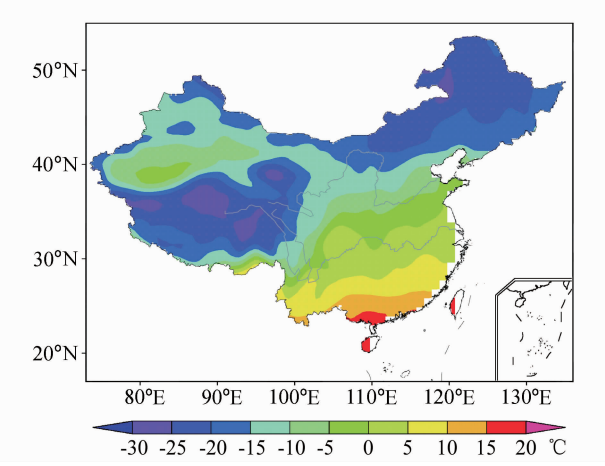



图 4　2000 年 1 月中国区域 NCEP 月平均陆面气温

Fig. 4　NCEP monthly mean land surface air temperature in Jan 2000 in China

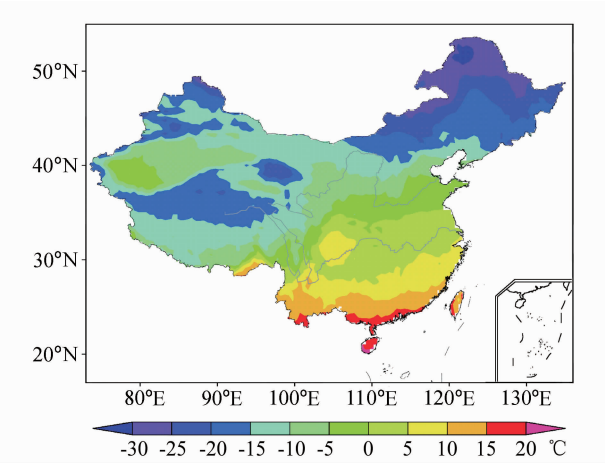



图 5　2000 年 1 月中国区域 CRU 的月平均陆面气温

Fig. 5　CRU monthly mean land surface air temperature in Jan 2000 in China

考虑到 NCEP 资料的较高时间分辨率，本文拟使用 2000 年 1 月逐日的 NCEP 1°×1° 格点资料，通过双线性插值到 0.5°×0.5° 的空间格距上（CRU 观测所在的格点），形成同化的背景场，其范围和空间总格点数与 CRU 观测一致，时间维数为 31 天。

### 2.4　分析场和台站观测资料

通过同化以上资料，得到的分析场将是逐日的资料。

用于检验同化结果的资料为 2000 年 1 月中国 740 个台站逐日的站点资料。经选择，2000 年 1 月温度观测完整的站点共有 623 个。

## 3　推广三维变分资料同化方法及其简化形式

作为观测场的 CRU 资料可信度高，但是时间分辨率低；作为背景场的 NCEP 资料时间分辨率高，但是可信度相对 CRU 差。如何最优地结合这两种资料，取长补短，是本文要解决的主要问题，我们拟采用 3DVar 的思路。由于用 3DVar 进行气象资料同化时，一般只同化一个时刻或一个 6 h 窗口内的观测资料，而本文要同化的 CRU 观测却涉及 1 个月同化窗口的月平均资料，因此 3DVar

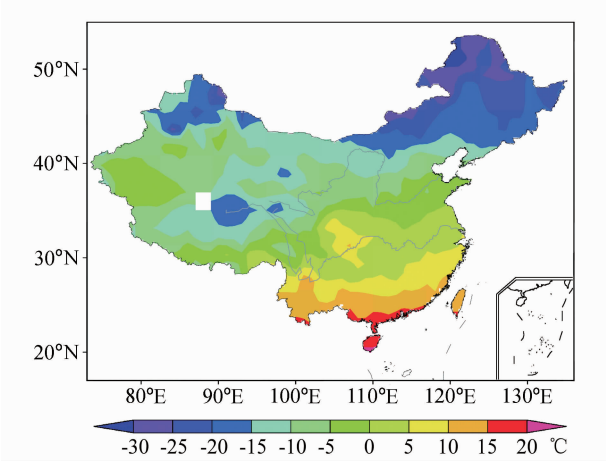



图 6　2000 年 1 月中国区域台站观测的月平均陆面气温

Fig. 6　Station observation monthly mean land surface air temperature in Jan 2000 in China

不能直接用于CRU观测的同化。为了能够把一个月平均的CRU观测有效地同化到31个逐日的NCEP背景场中去，必须对经典的3DVar进行推广，建立推广三维变分资料同化（Generalized-3DVar，简称G-3DVar）方法。推广的途径主要有两个方面：首先，是将31个逐日的NCEP背景场合并在一起作为一个整体，形成一个大的广义背景场（其总维数为31×9125），这样能有效克服3DVar无法处理1个月窗口内31个时次资料的困难；其次，就是建立一个从广义背景场到月平均场的广义观测算子，以便通过该算子将CRU观测的信息在同化过程中反馈到广义背景场中去。有了广义背景场和广义观测算子，可以直接采用3DVar对CRU观测进行同化，此为G-3DVar方法的基本思路。

### 3.1 相关量的介绍

在G-3DVar中，广义背景场、观测场、广义背景误差协方差矩阵、观测误差协方差矩阵、广义观测算子是几个重要的量，为了简单起见，本文仍然采用与3DVar完全一致的表示方法，即分别用 $\boldsymbol{x}_{\mathrm{b}}$、$\boldsymbol{y}_{\mathrm{obs}}$、$\boldsymbol{B}$、$\boldsymbol{R}$、$\boldsymbol{H}$ 表示G-3DVar的上述几个量作为理论探讨。本文中的 $\boldsymbol{B}$ 和 $\boldsymbol{R}$ 矩阵都是根据历史资料，通过扰动集合方法（Houtekamer et al.，1996；Fisher，1999）统计出来的。$\boldsymbol{B}$ 矩阵的统计使用NCEP资料1999～2006年1月及前后两个月（12、1、2月）的气温资料。$\boldsymbol{R}$ 矩阵的统计使用CRU资料1900～2002年1月的气温资料。由于CRU资料是月平均资料，并且相对于背景场与观测更为接近，所以最后取方差的1/30作为 $\boldsymbol{R}$ 矩阵。最终所取的 $\boldsymbol{R}$ 矩阵量级上约为 $\boldsymbol{B}$ 矩阵的1/30。

### 3.2 广义观测算子 $\boldsymbol{H}$

在G-3DVAR中，广义观测算子 $\boldsymbol{H}$ 的作用是将广义背景场空间的值转化为观测空间中的值，伴随 $\boldsymbol{H}^{\mathrm{T}}$ 则可用于目标函数梯度的计算。本文中广义背景场空间是一个时空一体化的空间，维数为31×9125，而观测空间为CRU观测所在的格点空间，维数为9125。因此，广义观测算子 $\boldsymbol{H}$ 可以表示成如下形式：

$$\boldsymbol{H}=\frac{1}{31}\underbrace{(\boldsymbol{I}_m\ \boldsymbol{I}_m\cdots\boldsymbol{I}_m)}_{31\text{个}}=$$

$$\frac{1}{31}\underbrace{\begin{pmatrix}1 & & 0 & 1 & & 0 & \cdots & 1 & & 0\\ & \ddots & & & \ddots & & & & \ddots & \\ 0 & & 1 & 0 & & 1 & & 0 & & 1\end{pmatrix}}_{31\text{个}}, \tag{1}$$

$$\boldsymbol{H}^{\mathrm{T}}=\frac{1}{31}\begin{pmatrix}\boldsymbol{I}_m\\ \boldsymbol{I}_m\\ \vdots\\ \boldsymbol{I}_m\end{pmatrix}=\frac{1}{31}\begin{pmatrix}1 & & 0\\ & \ddots & \\ 0 & & 1\\ 1 & & 0\\ & \ddots & \\ 0 & & 1\\ & \vdots & \\ 1 & & 0\\ & \ddots & \\ 0 & & 1\end{pmatrix}, \tag{2}$$

其中，$\boldsymbol{I}_m$ 是 $m\times m$ 的单位矩阵，$m$ 为空间格点数，$m=9125$。

正向的观测算子 $\boldsymbol{H}$ 将时间维中31天的值 $x$ 做平均 $\boldsymbol{H}_x$，得到和CRU相同的维数，二者在相同的维数空间比较，得到的修正向量 $\boldsymbol{y}_{\mathrm{obs}}-\boldsymbol{H}_x$ 通过观测算子的转置 $\boldsymbol{H}^{\mathrm{T}}$ 返回到背景场空间，最终得到分析场。

这里设计的 $\boldsymbol{H}$ 算子取NCEP资料时间分辨率高之长，将时间为31维的NCEP资料与时间为1维的CRU资料比较补CRU资料时间维低之短；$\boldsymbol{H}^{\mathrm{T}}$ 算子取CRU资料可信度高之长，将CRU资料的有用信息反馈到背景场中补NCEP资料偏差大之短。二者通过 $\boldsymbol{H}$ 算子及其转置并经过资料同化过程最优地结合在一起。

G-3DVAR将31天之长的资料放在一个时间窗口内来同化，是传统的3DVAR做不到的。而在本文的实例中，由于不需要任何预报模式，因此，即使是四维变分资料同化（4DVar）也无用武之地。由此可见，G-3DVAR在陆面月平均气象资料的同化中扮演着3DVar和4DVar所不可取代的角色。

### 3.3 G-3DVar方法的计算公式

G-3DVar是由3DVar扩展得到的，除了背景场和观测算子不同以外，其代价函数和分析值的表示方法都和3DVar一样。所以，G-3DVar目标函数为

$$J(\boldsymbol{x})=\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_b)^{\mathrm{T}}\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}_{\mathrm{b}})+\frac{1}{2}(\boldsymbol{H}_x-\boldsymbol{y}_{\mathrm{obs}})^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{H}_x-\boldsymbol{y}_{\mathrm{obs}}), \tag{3}$$

其分析值 $\boldsymbol{x}_{\mathrm{a}}$ 表示为

$$\boldsymbol{x}_{\mathrm{a}}=\boldsymbol{x}_{\mathrm{b}}+(\boldsymbol{I}_{31m}+\boldsymbol{B}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H})^{-1}\boldsymbol{B}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{y}_{\mathrm{obs}}-\boldsymbol{H}\boldsymbol{x}_{\mathrm{b}}), \tag{4}$$

式中 $\boldsymbol{I}_{31m}$ 的维数为 $31m\times 31m$。而且，G-3DVar 的分析场可最终表示成与最优插值（OI）等价的形式（Bouttier et al.，1999；Kalnay，2005）：

$$\boldsymbol{x}_{\mathrm{a}}=\boldsymbol{x}_{\mathrm{b}}+\boldsymbol{B}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{R}+\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\mathrm{T}})^{-1}(\boldsymbol{y}_{\mathrm{obs}}-\boldsymbol{H}\boldsymbol{x}_{\mathrm{b}}). \quad (5)$$

本文中 $\boldsymbol{H}$ 算子比较特殊，为块对角矩阵。由于 $\boldsymbol{I}_{31m}+\boldsymbol{B}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H}$ 的维数远大于 $\boldsymbol{R}+\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\mathrm{T}}$ 的维数，用公式（5）计算更加容易求解，尤其当 $\boldsymbol{R}$ 和 $\boldsymbol{B}$ 矩阵都是对角矩阵时，$\boldsymbol{R}+\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\mathrm{T}}$ 为对角矩阵，公式（5）可以直接求逆，所以在实际计算中该方法可以选择等价形式（5）来直接计算 G-3DVar 的分析场。

### 3.4　G-3DVar 方法的两种简化形式

陆面气象场比较特殊，要素在垂直方向上只有一层，因此其空间维数只有 2 维，G-3DVar 就成了处理二维空间和一维时间的问题。

作为方法检验的第一步，本文将 $\boldsymbol{B}$ 和 $\boldsymbol{R}$ 取为对角矩阵，不考虑点与点之间的协方差，即忽略空间上的相关性。也就是说在空间上，可以认为已经简化为一个一维的问题。简化后的代价函数可以在各个网格上独立计算，对每个点目标函数可以写为

$$J(\boldsymbol{x}_j)=\frac{1}{2}\sum_{i=1}^{31}w_{ji}^{\mathrm{x}}[\boldsymbol{x}_{ji}-(\boldsymbol{x}_{\mathrm{b}})_{ji}]^2+\frac{w_j^{\mathrm{o}}}{2}\left[\left(\sum_{i=1}^{31}\boldsymbol{x}_{ji}\right)/31-\boldsymbol{y}_j\right]^2, \quad (6)$$

其中，$i$ 为时间维数，$i=1,2,\cdots,31$；$j=1,2,\cdots,m$，$m$ 为空间总格点数；$w_{ji}^{\mathrm{x}}$ 为背景误差协方差的倒数，即 $\boldsymbol{B}$ 矩阵对角线值的倒数；$w_j^{\mathrm{o}}$ 为观测误差协方差的倒数，即 $\boldsymbol{R}$ 矩阵对角线值的倒数。全场的目标函数可以写成：

$$J(\boldsymbol{x})=\sum_{j=1}^{m}J(\boldsymbol{x}_j). \quad (7)$$

对于对角 $\boldsymbol{B}$ 矩阵，简化后的公式和公式（3）表示同样的问题。所以说公式(6)－(7)是公式（3）在 $\boldsymbol{B}$ 为对角矩阵情况下的一种简化形式。

进一步简化（6）式，假设对角 $\boldsymbol{B}$ 矩阵中每一天的方差是相同的，即 $\boldsymbol{B}$ 中的方差在月内没有时间变化。这样对于每个点的每一天，若取 $w_{ji}^{\mathrm{x}}=w_j^{\mathrm{x}}/31$（以便表示方便和整齐），则（6）式可写为

$$J(\boldsymbol{x}_j)=\frac{w_j^{\mathrm{x}}}{2}\sum_{i=1}^{31}(\delta\boldsymbol{x}_{ji})^2/31+\frac{w_j^{\mathrm{o}}}{2}\left[\left(\sum_{i=1}^{31}\delta\boldsymbol{x}_{ji}\right)/31+\overline{\boldsymbol{x}}_j-\boldsymbol{y}_j\right]^2, \quad (8)$$

其中，

$$\boldsymbol{x}_{ji}=(\boldsymbol{x}_{\mathrm{b}})_{ji}+\delta\boldsymbol{x}_{ji},$$

$$\overline{\boldsymbol{x}}_j=\left[\sum_{i=1}^{31}(\boldsymbol{x}_{\mathrm{b}})_{ji}\right]/31$$

是第 $j$ 个点背景场的月平均值。将（8）式代入（7）式，并令梯度为 0，便可得到第 $j$ 点每天的增量都相同：

$$\delta\boldsymbol{x}_{ji}=\delta_j=w_j^{\mathrm{o}}(\boldsymbol{y}_j-\overline{\boldsymbol{x}}_j)/(w_j^{\mathrm{x}}+w_j^{\mathrm{o}}). \quad (9)$$

（9）式与时间 $i$ 无关，相当于按照一定的权重将月平均温度偏差平均分配到每一天，如果认为 CRU 的月平均气温精度远高于 NCEP 的月平均资料（$w_j^{\mathrm{o}}\gg w_j^{\mathrm{x}}$），则（9）式可以进一步简化为

$$\delta\boldsymbol{x}_{ji}=\delta_j=\boldsymbol{y}_j-\overline{\boldsymbol{x}}_j. \quad (10)$$

上述简化公式完全没有考虑背景场。为了比较公平，第 4 部分将采用简化形式（6）和简化形式（9）进行同化试验，并对它们的结果进行分析。

## 4　同化试验及对比分析

为了检验 G-3DVar 方法简化形式公式（6）和（9）的同化效果，对 2000 年 1 月 1～31 日的温度场进行了同化试验。简化形式（6）的目标函数从同化前的 $J(\boldsymbol{x}_{\mathrm{b}})=1346394$ 降至同化后的 $J(\boldsymbol{x}_{\mathrm{a}})=133979$，这表明 CRU 观测的信息已经充分融入分析场。同化后的代价函数中背景项 $J_{\mathrm{b}}=103762$，$J_{\mathrm{o}}=30217$，这表明分析场更接近于观测场。另外，用公式（9）得到的分析场的月平均值为

$$\overline{\boldsymbol{x}}_{ji}^{\mathrm{a}}=\left(\sum_{i=1}^{31}\boldsymbol{x}_{ji}^{\mathrm{a}}\right)/31=(w_j^{\mathrm{o}}\boldsymbol{y}_j+w_j^{\mathrm{x}}\overline{\boldsymbol{x}}_j)/(w_j^{\mathrm{o}}+w_j^{\mathrm{x}}),$$

即分析场在月平均情形下带权加入了 CRU 资料的信息，尤其当 $w_j^{\mathrm{o}}\gg w_j^{\mathrm{x}}$ 时，几乎与 CRU 完全一致。

为了更细致地检验 G-3DVar 在简化情形的效果，先基于用公式（6）得到的分析场从月平均和逐日两方面与观测进行对比分析，然后再基于均方根误差比较两种简化形式公式（6）和（9）的效果。

### 4.1　月尺度检验

图 7 是用公式（6）同化得到的 2000 年 1 月

的月平均温度分析场，图 8 是对应的月平均温度分析增量场。分析增量主要表现在青藏高原地区、东北北部，新疆西北部也有小幅负增量。

图 8 表明，分析增量的最大值是在高原地区，增量为正值，并且正的增量区域向东北方向延伸。东北北部和新疆西北部是小幅的负增量。就全国区域而言，增量以正值为主。这与 NCEP 资料的误差特征相符合（参见图 4 和图 5）。

图 7（分析场）和图 4（背景场）、图 5（CRU 观测）比较发现，同化后月平均尺度的分析场在青藏高原地区、东北北部和新疆西北部都有明显的改善，基本和观测一致。这表明 G-3DVar 能够在月平均尺度上显著改进 NCEP 背景场的质量。

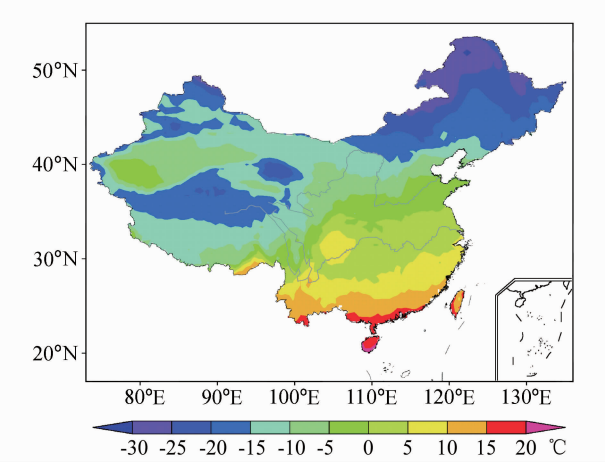



图 7 2000 年 1 月月平均陆面气温分析场

Fig. 7 Monthly mean analysis of land surface air temperature in Jan 2000 in China

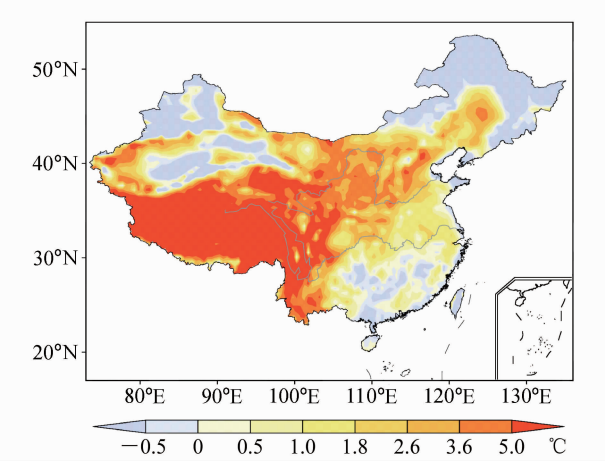



图 8 2000 年 1 月月平均陆面气温增量场

Fig. 8 Monthly mean increment of land surface air temperature in Jan 2000 in China

### 4.2 日尺度检验

图 7 的结果很好地说明了简化形式公式（6）的同化结果能够在月尺度上接近观测，但是同化得到的是逐日分析场，即 31 天的温度值。只有得到与观测一致的逐日温度值才是同化的最终目的。因此，有必要在日尺度上检验分析场是否与观测一致。由于 CRU 观测没有逐日资料，因此，选取独立于同化之外台站观测资料作为检验分析场的标准。

检验方法是首先选取 2000 年 1 月 20 日作为个例，比较背景场和分析场与站点观测之间的差别，然后分别计算逐日背景场和分析场插值到站点后与站点观测值的均方根误差。

图 9、10、11 分别是 2000 年 1 月 20 日的背景场、用公式（6）得到的分析场和台站观测的温度场，对其进行比较，发现主要有如下改进：

（1）东南地区分析场整体形势与台站观测基本一致。观测显示：0 ℃等温线向陕西、四川两省延伸，并且在重庆、川东地区有一个独立的暖中心，背景场并没有体现出这两点。分析场的0 ℃线位置及暖中心基本上与观测一致，但是暖区比观测范围大，从云南至四川是一个连续的暖区。

（2）东北地区的形势与观测事实一致，尤其是冷区的走向、冷区的中心位置都较背景场有很大改进。背景场的冷中心不突出。而实际观测显

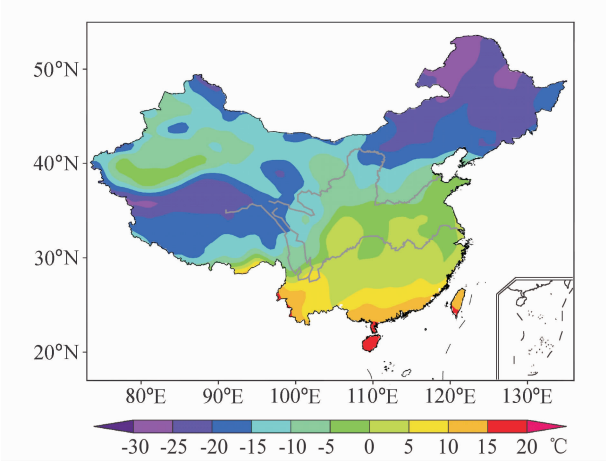



图 9 2000 年 1 月 20 日陆面气温背景场

Fig. 9 Background field of land surface air temperature on 20 Jan 2000

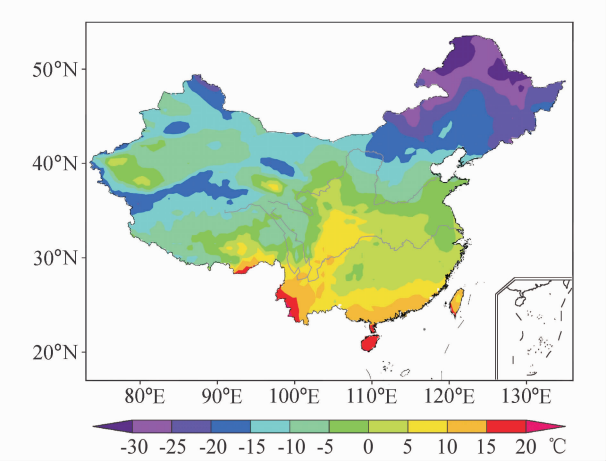



图 10　2000 年 1 月 20 日陆面气温分析场

Fig. 10　Analysis field of land surface air temperature on 20 Jan 2000

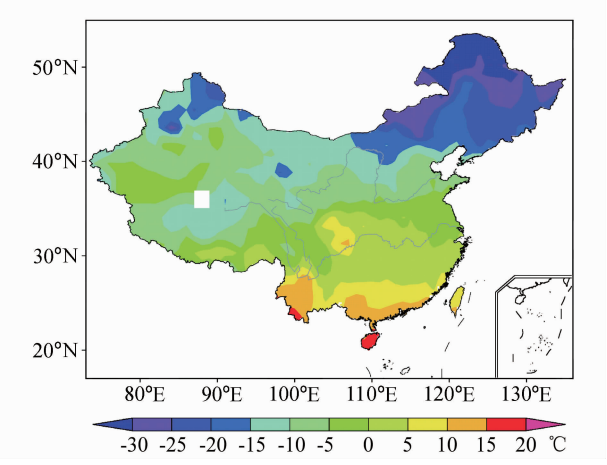



图 11　2000 年 1 月 20 日陆面气温台站观测

Fig. 11　Station observations of land surface air temperature on 20 Jan 2000

示，冷区由南向北伸入，形成了一个以大小兴安岭为冷区的“人”字型，在东北北部达到－30 ℃，中间的东北平原相对暖。分析场与观测相符合。

（3）新疆和青藏高原地区是资料同化中比较难处理的两个区域。无论是背景场还是台站观测，都存在很大误差。同化的优势主要体现在这两个地区。从分析场可以看出，高原以北的新疆、甘肃广大地区的温度场和地形走势密切相关。新疆西北部的阿尔泰山脉、准格尔盆地、天山山脉、塔里木盆地与相对的低、高、低、高的温度形势对应良好。而台站观测由于站点稀疏的原因，显得凌乱不合理。青藏高原因为它的高地形，使整个高原形成一个大的冷中心，冷区的边缘也与高原的地形边缘符合。青海北部的祁连山脉以及山脉南侧的柴达木盆地对应着相应的冷暖区，而观测通过 Cressman 插值得到的图 11 则没有体现出这个特点。

同化结果虽然相对背景场有很大的改进，但与站点观测相比仍然存在差别：虽然东北地区的冷区形势与实际台站观测相似，但是有些局地的冷区位置并不完全一致；东南部的几个主要暖区、暖中心都较背景场大幅改进，但是暖区的形势、范围上还略有不同，这可能与 $\boldsymbol{B}$ 矩阵的简单估计有关，进一步的改进需要采用考虑了协方差的非对角 $\boldsymbol{B}$ 矩阵。另外，对于青藏高原及新疆等台站稀疏的地区，从气温场与地形的匹配情况看，同化后的结果比背景场和台站观测都好。

### 4.3　误差分析

前面 2 小节是基于简化形式公式（6）的结果分析。由于简化形式公式（9）和（6）的结果比较接近，所以不再赘述。下面仅从均方根误差比较两种简化形式得到的分析场的整体效果。

图 10 给出了 1 月 20 日的分析场，同化的改进效果非常明显。但对于其他 30 天的温度值，是否如此？为了比较逐日的效果，将得到的简化公式（6）和（9）的分析场和背景场都插值到站点上，然后与台站观测值相减得到各自的误差场，并利用误差场得到各自的逐日均方根误差 RMSE（Root Mean Square Error，$E_{\mathrm{RMS}}$）（图 12），其计算公式为

$$\begin{cases} \varepsilon_j = x_j - x_j^{\mathrm{obs}}, \\ \bar{\varepsilon} = \dfrac{1}{k}\sum\limits_{j=1}^{k}\varepsilon_j, \\ E_{\mathrm{RMS}} = \dfrac{1}{\sqrt{k-1}}\sqrt{\sum\limits_{j=1}^{k}(\varepsilon_j - \bar{\varepsilon})^2}, \end{cases} \tag{11}$$

其中 $j=1, 2, \cdots, k$，$k$ 为台站的站点总数。

图 12 显示，2000 年 1 月 31 天的（6）式和（9）式分析场与台站观测的均方根误差均小于背景场与台站观测的均方根误差，即两种简化形式的同化结果均比背景场有所改进。再比较两种简化形式的效果，可以看到，在 31 天的均方根误差中，公式（6）的均方根误差有 25 天比公式（9）的小，也就是说，公式（6）整体上要优于公式

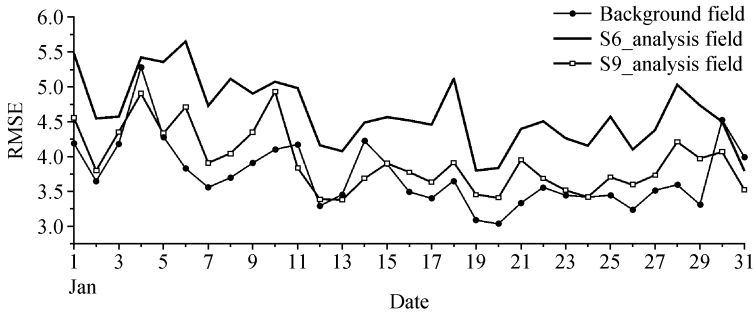



图 12 2000年1月分析场、背景场的逐日均方根误差

Fig. 12 Daily root-mean-square error (RMSE) of analysis field and background field in Jan 2000 in China

(9)。因此，在 ***B*** 矩阵中考虑方差逐日变化的同化效果整体上优于不考虑的情形。

从上面的误差分析，可以得到如下两个结论：1) ***B*** 矩阵考虑得越精细，整体同化效果就越好，并由此推断，在 G-3DVar 中采用包含协方差的非对角 ***B*** 矩阵可能会得到更高的同化精度和质量；2) 在两种简化形式下，G-3DVar 不仅在月尺度上大大改善了背景场，向 CRU 观测靠近，而且在日尺度上，同样很好地改善了背景场，减小了与台站观测的偏差。

## 5 结论与讨论

本文就如何改进 SDGVM 模型及其他陆面模型的气象驱动场做了探讨和研究。根据 CRU 资料和 NCEP 资料的特点，融合这两种资料，建立广义背景场和设计广义观测算子，并基于经典的 3DVar，提出了推广三维变分资料同化（G-3DVar)。同时以温度场为例，对提出的 G-3DVar 在两种简化形式下进行了同化试验和对比分析，得出以下结论：

(1) CRU 的陆面气温在整个中国区域可信度高，但是时间分辨率低，只有月平均值；NCEP 再分析资料的陆面气温在我国的西部地区相对误差偏大，可信度低，但其时间分辨率高，为逐日资料。通过 G-3DVar 可以使这两种资料有机地结合起来，取长补短，得到时间分辨率高、可信度高的格点资料。

(2) 在两种简化形式下，G-3DVar 不仅在月尺度上大大改善了背景场，向 CRU 观测靠近，而且在日尺度上，同样很好地改善了背景场，减小了与台站观测的偏差。我国的中东部广大地区，同化结果和实际台站观测一致。台站稀疏的高原及其他地区，同化能够提取 CRU 资料中的信息，在一定程度上修正了台站稀疏带来的不准确性，得到更为可靠的格点资料。

(3) 对 G-3DVar 两种简化形式的分析场进行误差分析发现，在 ***B*** 矩阵中考虑方差逐日变化的同化效果整体上优于不考虑的情形。也就是说，***B*** 矩阵统计得越精细，整体同化效果就越好，并由此推断，在 G-3DVar 中采用包含协方差的非对角 ***B*** 矩阵可能会得到更高的同化精度。

(4) 对于 SDGVM 模型的另外几个气象输入要素，包括最高温度、最低温度、云量、相对湿度等，用 G-3DVar 的简化形式进行同化，也得到了相同结论。目前，针对同化后的气象输入场，已经作了很多模拟和评估，模型对各输出量的模拟能力都得到了不同程度的改善，其结果将另文介绍。

(5) 本文使用对角的 ***B*** 矩阵，使问题大大简化，并在同化试验中取得了好的效果，但这只是对提出的 G-3DVar 进行检验的一种手段，为最终在 G-3DVar 中使用考虑协方差的完整非对角 ***B*** 矩阵，进一步提高同化精度和质量，为 SDGVM 提供更优化和合理的多年 1～12 月的陆面气象要素驱动场奠定基础。

### 参考文献 (References)

Bouttier F, Courtier P. 1999. Data assimilation concepts and methods [R]. EC Meteorological Training Course Lecture Series.

蔡福，于贵瑞，祝青林，等. 2005. 气象要素空间化方法精度的比较研究－以平均气温为例 [J]. 资源科学，27 (5)：173－179.

Cai Fu, Yu Guirui, Zhu Qinglin, et al. 2005. Comparison of precisions between spatial methods of climatic factors: A case study on mean air temperature [J]. Resources Science (in Chinese), 27 (5): 173 - 179.

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.

Fisher M. 1999. Background error statistics derived from an ensemble of analyses [R]. ECMWF Research Department Technical Memorandum, 79: 12.

Fisher M. 2001. Assimilation techniques (3): 3dVar [R]. EC Meteorological Training Course Lecture Series.

Houtekamer P L, Lefaivre L, Derome J, et al. 1996. A system simulation approach to ensemble prediction [J]. Mon. Wea. Rev., 124 (6): 1225 - 1242.

Kalnay E. 2005. 大气模式、资料同化和可预报性 [M]. 北京：气象出版社，115 - 171. Kalnay E. 2005. Atmospheric Modeling, Data Assimilation and Predictability [M] (in Chinese). Beijing: China Meteorological Press, 115 - 171.

廖要明，张强，陈德亮. 2004. 中国天气发生器的降水模拟 [J]. 地理学报，59 (5)：689 - 698. Liao Yaoming, Zhang Qiang, Chen Deliang. 2004. Precipitation simulation in China with a weather generator [J]. Acta Geographica Sinica (in Chinese), 59 (5): 689 - 698.

林忠辉，莫兴国，李宏轩，等. 2002. 中国陆地区域气象要素的空间插值 [J]. 地理学报，57 (1)：47 - 56. Lin Zhonghui, Mo Xingguo, Li Haibin, et al. 2002. Comparison of three spatial interpolation methods for climate variables in China [J]. Acta Geographica Sinica (in Chinese), 57 (1): 47 - 56.

毛嘉富. 2006. 动态全球植被模型的改进及应用研究 [D]. 中国科学院大气物理研究所博士学位论文. Mao Jiafu. 2006. Modify and application research of Sheffield dynamic global vegetation model [D]. Ph. D. dissertation (in Chinese), Institute of Atmospheric Physics, Chinese Academy of Sciences.

Mitchell T D, Carter T R, Jones P D, et al. 2004. A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: The observed record (1901 - 2001) and 16 scenarios (2001 - 2100) [R]. Tyndall Working Paper 55, Tyne Centre for Climate Change Research.

Mitchell T D, Jones P D. 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids [J]. International Journal of Climatology, 25 (6): 693 - 712.

Rutherford I D. 1972. Data assimilation by statistical interpolation of forecast error fields [J]. J. Atmos. Sci., 29 (5): 809 - 815.

苏志侠，吕世华，罗思维. 1999. 美国 NCEP/NCAR 全球再分析资料及其初步分析 [J]. 高原气象，18 (2)：209 - 218. Su Zhixia, Lü Shihua, Luo SIwei. 1999. The examinations and analysis of NCEP/NCAR 40 years global reanalysis data in China [J]. Plateau Meteorology (in Chinese), 18 (2): 209 - 218.

Wang Bin, Zhao Ying. 2006. A new approach to data assimilation [J]. Acta Meteorrologica Sinica, 20 (3): 275 - 282.

闻新宇，王绍武，朱锦红. 2006. 英国 CRU 高分辨率格点资料揭示的 20 世纪中国气候变化 [J]. 大气科学，30 (5)：894 - 904. Wen Xinyu, Wang Shaowu, Zhu Jinhon. 2006. An overview of China climate change over the 20th century using UK UEA/CRU high resolution grid data [J]. Chinese Journal of Atmospheric Sciences (in Chinese), 30 (5): 894 - 904.

徐影，丁一汇，赵宗慈. 2001. 美国 NCEP/NCAR 近 50 年全球再分析资料在我国气候变化研究中可信度的初步分析 [J]. 应用气象学报，2 (3)：337 - 347. Xu Ying, Ding Yihui, Zhao Zongci. 2001. Confidence analysis of NCEP/NCAR 50-year global reanalyzed data in climate chance research in China [J]. Quarterly Journal of Applied Meteorology (in Chinese), 2 (3): 337 - 347.

赵天保，艾丽坤，冯锦明. 2004. NCEP 再分析资料和中国站点观测资料的分析与比较 [J]. 气候与环境研究，9 (2)：278 - 293. Zhao Tianbao, Ai Likun, Feng Jinmin. 2004. An intercomparison between NCEP reanalysis and observed data over China [J]. Climatic and Environmental Research (in Chinese), 9 (2): 278 - 293.

庄立伟，王石立. 2003. 东北地区逐日气象要素的空间插值方法应用研究 [J]. 应用气象学报，14 (5)：605 - 615. Zhuang Liwei, Wang Shili. 2003. Spatial interpolation methods of daily weather data: In Northeast China [J]. Journal of Applied Meteorological Science (in Chinese), 14 (5): 605 - 615.