

中国 1 km 栅格月平均气温数据集

荆文龙^{1,2}, 杨雅萍^{1*}, 乐夏芳¹

1. 中国科学院地理科学与资源研究所 北京 100101; 2. 中国科学院大学 北京 100049

摘要: 根据美国气候研究中心 (NCEP/NCAR) 全球近地表月平均气温要素数据结合数字高程模型 (DEM)、归一化植被指数 (NDVI) 以及地理位置等要素之间定量关系, 建立点至公里统计降尺度回归树模型, 计算出中国 2000 年 1 月-2010 年 12 月 1 km 空间分辨率月平均气温模型计算数据集。利用中国 380 个气象站点观测数据对其结果进行对比分析和验证, 最终得到中国 2000 年 1 月-2010 年 12 月间、1 km 分辨率、月平均气温数据集。该数据是可持续更新的中国公里网格月平均气温数据集。

关键词: 月平均气温; 统计降尺度; NCEP/NCAR; NDVI; 中国; 1 公里栅格数据

DOI: 10.3974/geodp.2017.01.10

1 前言

近地表气温是描述陆地表面环境条件和热量特征的重要指标, 也是气象观测资料中最重要的观测数据之一^[1,2]。栅格化的气温数据能够提供区域范围内气温的连续分布情况, 是各种植物生理、水文、气象、环境等模式或模型中的关键的输入参数^[1]。因此, 获得准确和较高空间分辨率的栅格化气温、降水等分布数据对于更好地理解陆地表面过程和全球变化等相关研究都有非常重要的意义^[1-4]。

近地表气温数据的获取途径主要有两种: 气象站点观测和遥感反演。气象站点观测数据是点数据, 其观测仅代表站点周围一定范围的气温状况, 但在许多水文、环境、生态模型中需要栅格化的水热条件数据。虽然可以通过空间插值将气象站点的数据栅格化^[5,6], 但是直接内插的方法强调气候要素在二维空间的分布规律, 容易忽略地形、植被等因素的影响, 难以获得高精度的插值结果。一些研究者将海拔高程引入, 建立了基于 DEM 的空间插值方法, 与传统简单内插方法相比, 精度有所提高^[5-8]。但是由于每个站点仅代表附近有限范围的气候状况^[1], 插值方法严重依赖于气象观测站点的分布。实际工作中, 气象观测站点的分布并不是均匀的, 分布密度有限, 特别是在中国西部等偏远地区, 气象观测站分布非常稀疏, 导致基于观测站点的空间内插精度较低。

NCEP/NCAR 全球气候再分析资料来自美国国家环境预报中心 (National Centers for Environmental Prediction, NCEP) 和国家大气中心 (National Center for Atmospheric Research, NCAR) 的全球再分析计划, 是由美国国家环境预报中心和国家大气研究中心协作, 对来

收稿日期: 2014-12-20; 修订日期: 2015-03-15; 出版日期: 2017-03-25

通讯作者: 杨雅萍 D-4110-2017, 中国科学院地理科学与资源研究所, yangyp@igsnr.ac.cn

论文引用格式: 荆文龙, 杨雅萍, 乐夏芳. 中国 1 km 栅格月平均气温数据集[J]. 全球变化数据学报, 2017, 1(1): 66-73. DOI: 10.3974/geodp.2017.01.10.

数据集引用格式: 荆文龙, 杨雅萍, 冯敏等. 中国 1 km 栅格月平均气温数据集[DB]. 全球变化科学研究数据出版系统, DOI:10.3974/geodb.2015.01.02.V1

源于地面、船舶、无线电探空、探空气球、飞机、卫星等的气象观测资料进行同化处理后，研制的全球气象资料数据库^[9-11]，NCEP/NCAR 月平均再分析资料可靠性强，作为目前用于气候分析时间序列最长的分析资料，广泛应用于气候诊断分析、东亚季风、热通量水分循环等各方面的研究中^[11-15]。NCEP/NCAR 再分析资料能够提供覆盖全球的栅格化气候数据，但空间分辨率较低（约 0.5 度），难以满足区域高精度研究和模拟的实际需要。降尺度方法是解决由低分辨率气候资料到中高分辨率资料转换的有效手段，其基本原理是采用统计经验的方法建立大尺度气象变量与区域尺度变量之间的线性或非线性联系^[16-21]。本数据集提取近地表气温与地形、植被等环境因子、地理位置的关系，建立月平均气温与地表植被状况和地形要素之间的回归树模型，最终得到了 2000 年 1 月至 2010 年 12 月中国陆地范围 1 km 分辨率逐月月平均气温数据集^[22]。

2 数据集元数据简介

中国 1 km 栅格月平均气温数据集（ChinaTemp）的名称、短名名称、通讯作者、作者、地理区域、数据年代、数据空间分辨率、数据出版单位、数据共享服务平台、数据集组成等信息一并列于表 1。

表 1 中国 1 km 栅格月平均气温数据集元数据简表

条目	描述
数据集名称	中国 1 km 栅格月平均气温数据集
数据集短名	ChinaTemp
作者信息	荆文龙 D-2691-2017, 中国科学院地理科学与资源研究所, jingwl.14b@igsnr.ac.cn 杨雅萍 D-4110-2017, 中国科学院地理科学与资源研究所, yangyp@igsnr.ac.cn 乐夏芳 D-4113-2017, 中国科学院地理科学与资源研究所, lexif@lreis.ac.cn
地理区域	地理范围包括 3°51'N-53°34'N, 73°E-135°5'E。包括中国陆地范围。
数据年代	2000-2010 年
空间分辨率	1000 m
数据文件个数	132
数据格式	Tiff
数据量	4096 MB
数据集组成	包括 132 个数据文件和 1 个子数据集，380 个站点验证与模型数据对比数据。其中：tmp_yyyy_mm.tif 是月平均气温数据，如 tmp_2010_05.tif 是 2010 年 5 月平均气温数据。tmp_new.xls 是月平均气温数据站点与模型数据对比数据。sthinfo.csv 是站点编号、站点名及经纬度等信息。
基金项目	中华人民共和国环境保护部（STSN-04-08）
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101, 中国科学院地理科学与资源研究所
数据共享政策	全球变化科学研究数据出版系统的“数据”包括元数据（中英文）、实体数据（中英文）和通过《全球变化数据学报》（中英文）发表的数据论文。其共享政策如下：（1）“数据”以最便利的方式通过互联网系统免费向全社会开放，用户免费浏览、免费下载；（2）最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源；（3）增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报》（中英文）编辑部签署书面协议，获得许可；（4）摘取“数据”中的部分记录创作新数据的作者需要遵循 10% 引用原则，即从本数据集中摘取的数据记录少于新数据集总记录量的 10%，同时需要对摘取的数据记录标注数据来源 ^[23] 。

3 数据研发方法

3.1 数据来源

本数据集中所采用的 NCEP/NCAR 全球气候再分析资料为 NCEP/NCAR 逐时 (6h) 全球最终分析资料中的近地表气温, 数据集时间范围为 2000-2010 年。

植被对气候要素非常敏感, 所以地表植被状况能够一定程度上反映区域内温度等的分布和变化。归一化植被指数 (Normalized Difference Vegetation Index, NDVI) 对植物的健康状况和生物量有很强的指示作用, 是植被生长状况及植被覆盖度的有效指示因子^[24], 研究表明: NDVI 与近地表气温、降水等气候条件之间存在显著的相关性^[25-29]。本研究采用了 MODIS 月合成 NDVI 产品 (MOD13A3, 来源: <http://ladsweb.nascom.nasa.gov/data>), 资料空间分辨率为 1km, 时间范围为 2000 年 1 月到 2010 年 12 月。NDVI 的有效值域范围为 -0.1 ~ 1.0, 考虑到 MODIS NDVI 数据产品在一些 NDVI 极低值区域有明显的规则条带, 这些条带会对预测结果产生影响, 在中国陆地范围内, 出现条带的主要区域在塔里木盆地, 该地区的土地覆被类型以荒漠为主, 植被极少, NDVI 值与气温的分布的关系可以忽略, 所以在本文中不考虑 NDVI 小于 0.1 的变化, 在计算过程中将 NDVI 小于 0.1 的 NDVI 统一设定为 0.1。

DEM (Digital Elevation Model) 数据来自全球陆地调查数字高程模型 (Global Land Survey Digital Elevation Model, GLS DEM), 是美国航空航天局 (National Aeronautics and Space Administration, NASA) 和美国地质调查局 (United States Geological Survey, USGS) 的最新全球影像产品中的一个重要的空间数据集。该数据集最初是为了用于对全球陆地调查 (GLS) 数据进行正射校正而被集成于 GLS DEM 产品之中。GLS DEM 产品发布于 2008 年, 空间分辨率为 30" (90 m) (来源: <http://glcf.umiacs.umd.edu/data/glsdem>)。本研究中的 DEM 数据是以 GLS DEM 数据为基础, 经过空间综合计算, 空间分辨率为 1 km。

3.2 研究方法

基于平均气温和 NDVI、地形之间的密切关系, 首先建立 50 km 尺度上的逐月 NCEP/NCAR 月平均气温、月合成 NDVI、DEM、地理坐标之间的回归树模型, 再将建立的模型应用到 1 km 尺度的 DEM 和月合成 NDVI 数据, 得到 1 km 尺度的全国逐月平均气温数据。具体的降尺度算法描述如下: (1) 对分辨率为 1 km 的 DEM 和月合成 NDVI 数据进行空间综合计算, 得到分辨率为 50 km 的 DEM 和 NDVI 数据, 综合方法为求取每一个 50 km 栅格内的所有 1 km 像元值的算术平均值; (2) 读取每一个 50 km 栅格像元相对应的 NDVI、DEM、平均气温和地理坐标信息, 建立 NCEP/NCAR 月平均气温与 DEM、月合成 NDVI 以及地理位置之间的回归树模型; (3) 将 (2) 中建立的模型应用于 1 km 分辨率的 DEM、NDVI 数据, 得到 1 km 分辨率的月平均气温数据; (4) 将第 (3) 步中得到 1 km 分辨率气温数据综合到 50 km 分辨率, 计算方法同 (1), 然后求得与 NCEP/NCAR 月平均气温的差值, 得到模型计算的残差; (5) 利用样条插值方法将 50 km 分辨率的残差插值到 1 km 分辨率; (6) 将 (5) 中得到的 1 km 分辨率的残差与 (3) 中得到的 1 km 尺度平均气温相加, 得到最终的 1 km 分辨率全国陆地范围月平均气温降尺度预测结果。技术路线如图 1 所示。

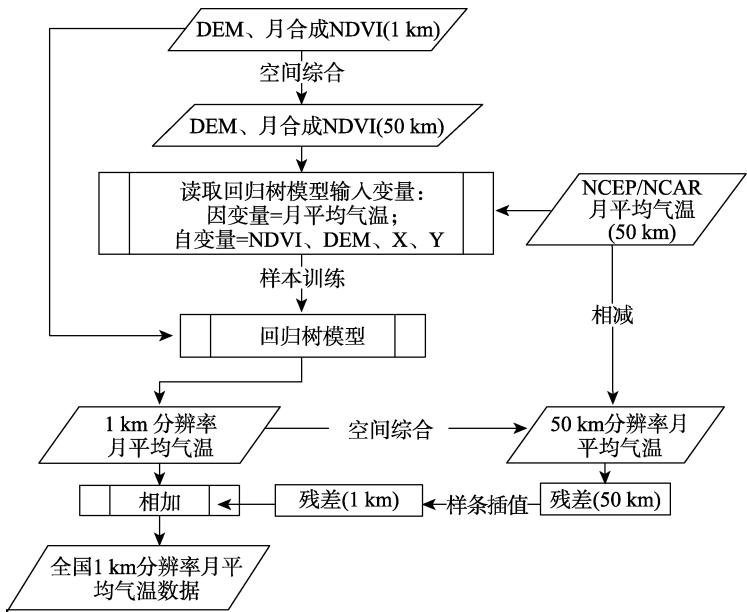


图 1 回归树气温降尺度算法流程图

4 数据结果与验证

4.1 数据结果

表 2 为 2000 年 1 月至 2010 年 12 月的各月份回归树模型的 R^2 ,图 2 中的散点图为 2010 年 1 月到 12 月份的 50 km 分辨率 NCEP/NCAR 月平均气温数据和 50 km 分辨率模型预测的平均气温数据之间的对比。由表 2 可以看出,回归树模型各个月份的 R^2 在 0.884 和 0.952 之间,各个年份的模型 R^2 呈现相似的月际变化规律,4 月至 8 月份的 R^2 稍小,总体呈现冬季大于夏季的趋势,预测结果与 NCEP/NCAR 月平均气温对比的 R^2 均大于 0.998 (图 2),说明在 50 km 尺度上建立的回归树模型可以准确地反映中国陆地范围的月平均气温的分布情况和与高程、植被指数以及地理位置之间的定量关系。

表 2 2000 至 2010 年各月份回归树模型的 R^2

	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
2000	0.941	0.941	0.922	0.922	0.903	0.885	0.884	0.884	0.922	0.922	0.941	0.941
2001	0.922	0.922	0.922	0.903	0.903	0.903	0.884	0.903	0.922	0.941	0.941	0.941
2002	0.941	0.941	0.922	0.903	0.903	0.884	0.884	0.903	0.922	0.922	0.922	0.941
2003	0.922	0.922	0.903	0.903	0.903	0.903	0.884	0.903	0.922	0.922	0.941	0.941
2004	0.941	0.941	0.922	0.903	0.922	0.903	0.903	0.922	0.922	0.941	0.941	0.941
2005	0.922	0.916	0.903	0.920	0.903	0.895	0.897	0.903	0.922	0.944	0.952	0.941
2006	0.941	0.922	0.922	0.903	0.903	0.885	0.884	0.884	0.922	0.922	0.922	0.938
2007	0.922	0.941	0.922	0.903	0.884	0.884	0.903	0.903	0.922	0.922	0.922	0.941
2008	0.922	0.922	0.922	0.887	0.885	0.903	0.903	0.922	0.922	0.941	0.941	0.941
2009	0.922	0.922	0.922	0.902	0.884	0.903	0.903	0.922	0.922	0.941	0.941	0.941
2010	0.922	0.940	0.938	0.918	0.921	0.911	0.894	0.912	0.943	0.932	0.941	0.938

将建立的回归树模型应用于对应的分辨率为 1 km 的 NDVI 和 DEM 数据后得到的全国陆地范围月平均气温数据。将得到的数据进行空间综合计算，得到 50 km 分辨率平均气温数据，求得与对应的 NCEP/NCAR 月平均数据的差值，得到 50 km 分辨率的预测模型的残差。然后通过样条插值，得到 1 km 分辨率的残差分布，与对应的 1 km 尺度预测结果求和，得到最终的 1 km 分辨率全国陆地范围月平均气温数据(图 3, 以 2005 年 5 月平均气温为例)。

4.2 数据验证

4.2.1 精度验证

利用全国陆地范围内的 380 个气象观测站点的实地观测数据，对降尺度预测结果进行验证，验证结果如表 3 所示， R^2 范围为 0.865–0.95，均方根误差（RMSE）为 1.88℃–2.68℃，1 月到 4 月之间误差数值稍高于其他月份。总体来说，降尺度结果可以准确的反映全国月平均气温的分布情况以及数值变化。

表 3 预测数据精度验证结果

月份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
R^2	0.941	0.938	0.922	0.902	0.890	0.865	0.861	0.869	0.886	0.925	0.948	0.950
RMSE	2.681	2.660	2.411	2.219	2.021	1.979	1.973	1.880	1.964	1.938	2.033	2.415

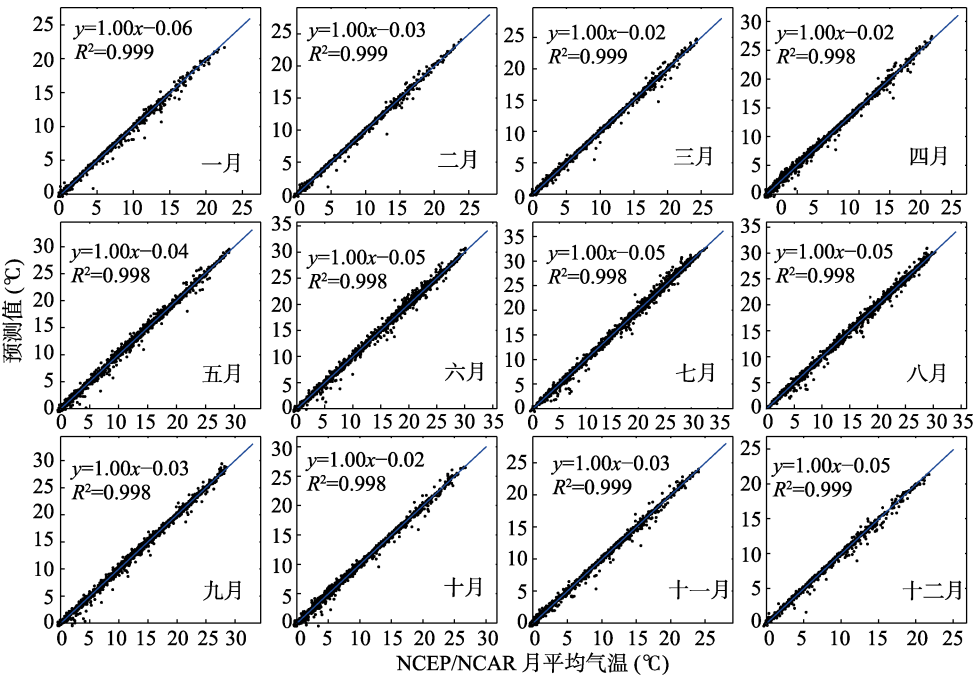


图 2 2010 年各月 NCEP/NCAR 数据与 50 km 分辨率降尺度预测值关系散点图

4.2.2 误差分析

为了进一步分析降尺度结果的误差空间分布情况，计算出各个站点的年平均偏差（Mean Bias Error, MBE），计算公式为： $MBE = N^{-1} \sum_{i=1}^N (y_i - \hat{y}_i)$

其中, y_i 表示第 i 个气象站点的实测月平均气温; \hat{y}_i 表示第 i 个气象站点对应位置的降尺度预测月平均气温值, i 是站点序号, N 为月份。计算结果如表 4 所示。

表 4 中国陆地范围气象站点年平均偏差计算结果

误差范围	<-4℃	-4~-3℃	-3~-2℃	-2~-1℃	-1~1℃	1~2℃	2~3℃	3~4℃	4~6℃	>6℃
站点个数	1	1	6	28	217	72	26	17	12	4
所占百分比	0.3	0.3	1.6	7.3	56.5	18.8	6.8	4.4	3.1	1.0

由表 3 可知, 大部分站点 MBE 为正值, 表明这些站点的实地观测数据略高于降尺度预测数据, 80% 以上站点的实地观测数据与降尺度预测值之间的绝对误差在 2℃ 以内, 其中有 56% 左右站点的年平均偏差在 -1℃~1℃ 之间, 误差大于 3℃ 的站点集中分布在西部地区 (图 4), 这些区域基本上处于高原地带, 海拔较高, 地形起伏较大。地势复杂、地形、站点稀少等原因均可能造成降尺度结果在这些地区存在一定的不确定性。

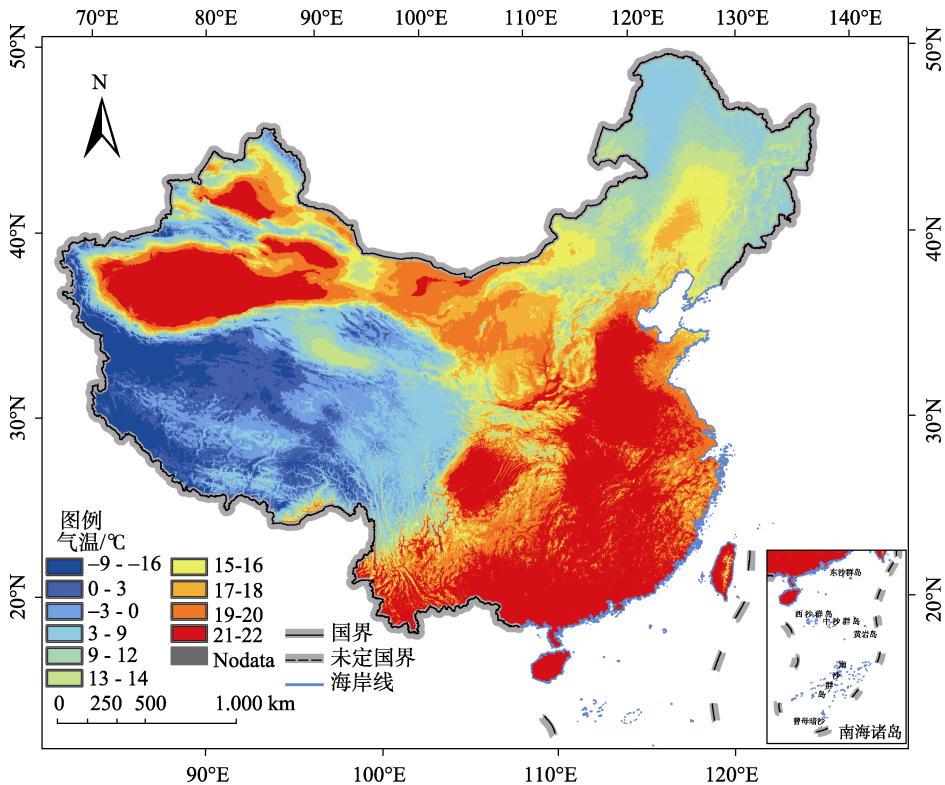


图 3 2010 年 5 月全国陆地范围 1 km 分辨率月平均气温数据 (审图号: GS(2015)519 号)

5 讨论

以拟合近地表气温与地形、植被环境因子、地理位置的关系为基础, 在 50 km 尺度上建立了 NCEP/NCAR 数据与 NDVI 和 DEM 数据的回归树模型, 最终得到了 2000-2010 年全国 1 km 分辨率的逐月月平均气温数据集, 并利用全国 380 个气象站点数据进行了验证, 结果表明本研究获得的空间统计降尺度结果能够有效地反映全国陆地范围内的变化趋势。

栅格形式的气温数据代表的是栅格区域内的平均气温状况, 与气象站点的观测数据有

着很大的区别，研究中所用到的多源多尺度数据的精确度都会影响到降尺度结果的精度。

气象站点的观测数据为点数据，其可以代表的区域范围有限，特别是在地形起伏大、景观较为破碎的区域。该数据集为栅格数据，表示的是 1 km 尺度上的区域内的平均状况，与气象站点的观测数据存在差异，因此，与站点观测的对比，虽然能够在一定程度上检验降尺度结果，但并不能完全准确地反映数据集精度。

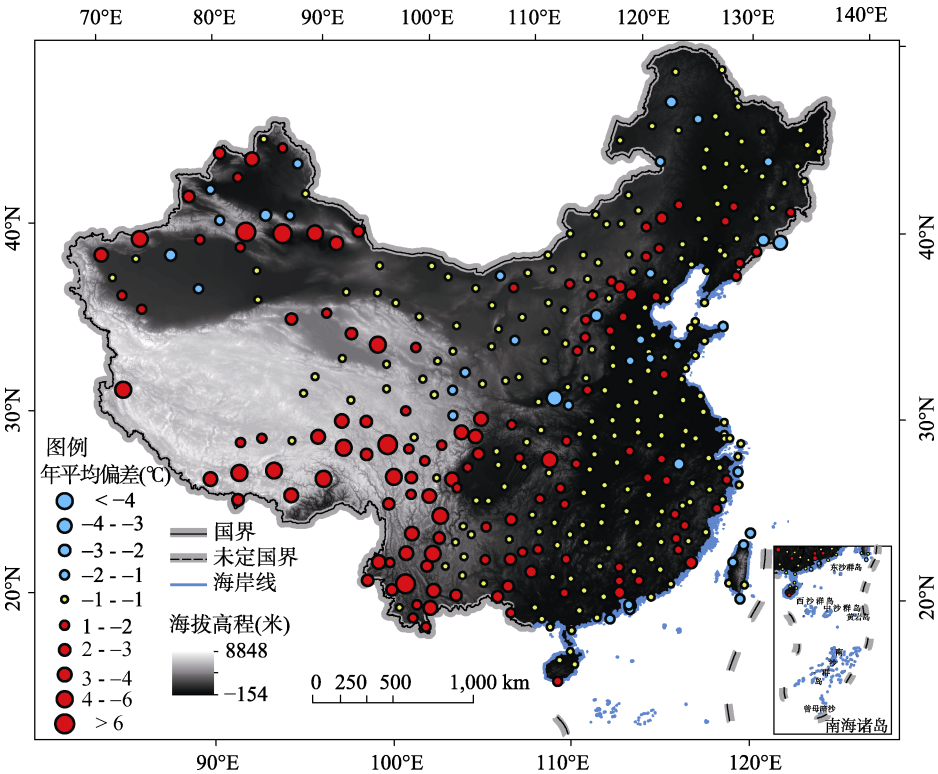


图 4 平均偏差的空间分布 (审图号: GS(2015)519 号)

对于数据源，由于 NCEP/NCAR 是结合了多种数据源的气候资料再分析数据，其自身不可避免地存在一定的内在误差，而这些误差也会进而影响最终降尺度结果精度^[30,31]。此外，NDVI 和 DEM 的精度也在一定程度上影响了降尺度结果的精度，有必要在未来工作中，进一步地对比和分析。

栅格化气候要素的获取与集成是一个复杂的问题，本数据集所采用的月平均气温降尺度方法只是开始，需要不断完善和发展，引入各种尺度的小地形参数、将多种数据源引进到模型中都是提高降尺度预测精度的途径，有待今后进一步深入研究。

作者分工：杨雅萍完成算法的设计和论文的撰写，荆文龙完成了算法的具体实现、结果分析和论文撰写，乐夏芳完成了数据的收集和预处理等工作。

参考文献

[1] 祝善友, 张桂欣. 近地表气温遥感反演研究进展[J]. 地球科学进展, 2011, 26(7): 724-730.

[2] 齐述华, 王军邦, 张庆员等. 利用 MODIS 遥感影像获取近地层气温的方法研究[J]. 遥感学报, 2005, 9(5): 570-575.

- [3] Cresswell, IIM P, Morse, A. P., et al. Estimating surface air temperature from meteosat land surface temperature using an empirical solar zenith angle model[J]. *International Journal of Remote Sensing*, 1990, 20(6): 1125-1132.
- [4] Mao, K. B., Tang, H. J., Wang, X. F., et al. Near-surface air temperature estimation from ASTER data based on neural network algorithm [J]. *International Journal of Remote Sensing*, 2008, 29(20): 6021-6028.
- [5] 廖顺宝, 李泽辉, 游松财. 气温数据栅格化的方法及其比较[J]. *资源科学*, 2003, 25(6): 83-88.
- [6] 廖顺宝, 李泽辉. 气温数据栅格化中的几个具体问题[J]. *气象科技*, 2004, 32(5): 352-356.
- [7] 蔡迪花, 郭锐, 李崇伟. 基于 DEM 的气温插值方法研究[J]. *干旱气象*, 2009, 27(1): 10-17.
- [8] 李军, 游松财, 黄敬峰. 中国 1961—2000 年月平均气温空间插值方法与空间分布[J]. *生态环境*, 2006, 15(1): 109-114.
- [9] Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., et al. The NCEP/NCAR 40-year reanalysis project[J]. *Bulletin of American Meteorological Society*, 1996, 77(3): 437-471.
- [10] 赵天保, 符淙斌. 中国区域 ERA-40、NCEP-2 再分析资料与观测资料的初步比较与分析[J]. *气候与环境研究*, 2006, 11(1): 14-32.
- [11] 黄刚. NCEP/NCAR 和 ERA-40 再分析资料以及探空观测资料分析中国北方地区年代际气候变化[J]. *气候与环境研究*, 2006, 11(3): 310-320.
- [12] 施晓晖, 徐祥德, 谢立安. NCEP/NCAR 再分析风速、表面气温距平在中国区域气候变化研究中的可信度分析[J]. *气象学报*, 2006, 64(6): 709-722.
- [13] 周青, 赵凤生, 高文华. NCEP/NCAR 逐时分析与中国实测地表温度和地面气温对比分析[J]. *气象*, 2008, 34(2): 83-91.
- [14] 苏志侠, 吕世华, 罗四维. 美国 NCEP/NCAR 全球再分析资料及其初步分析[J]. *高原气象*, 1999, 18(2): 84-93.
- [15] 徐影, 丁一汇, 赵宗慈. 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*, 2001, 12(3): 337-347.
- [16] 梁顺林. 定量遥感[M]. 北京: 科学出版社, 2009, 337-338.
- [17] 刘永和, 郭维栋, 冯锦明等. 气象资料的统计降尺度方法综述[J]. *地球科学进展*, 2011, 26(8): 837-847.
- [18] Immerzeel, W. W., Rutten, M. M., Droogers, P. Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula[J]. *Remote Sensing of Environment*, 2009, 113: 362-370.
- [19] Carone, G. J., Bramante, P. D. Translating monthly temperature from regional to local scale in the south-eastern United States[J]. *Climate Research*. 1995, 5: 229-242.
- [20] Wigley, T. M. L., Jones, P. D., Briffa, K. R., et al. Obtaining sub-grid-scale information from coarse-resolution general circulation model output[J]. *Journal of Geophysical Research*, 1990, 95(D2): 1943-1953.
- [21] 唐国利, 丁一汇. 由最高最低气温求算的平均气温对我国年平均气温序列影响[J]. *应用气象学报*, 2007, 18(2): 187-192.
- [22] 荆文龙, 杨雅萍, 冯敏等. 2015. 中国 1 km 栅格月平均气温数据集[DB], 全球变化科学研究数据出版系统, DOI: 10.3974/geodb.2015.01.02.V1
- [23] 范娜, 谢高地, 张昌顺等. 2001 年至 2010 年澜沧江流域植被覆盖动态变化分析[J]. *资源科学*, 2012, 34(7): 1222-1231.
- [24] 罗玲, 王宗明, 宋开山等. 1982 ~ 2003 年中国东北地区不同类型植被 NDVI 与气候因子的关系研究[J]. *西北植物学报*, 2009, 29(4): 800-808.
- [25] 王宗明, 国志兴, 宋开山等. 中国东北地区植被 NDVI 对气候变化的响应[J]. *生态学杂志*, 2009, 28(6): 1041-1048.
- [26] 崔林丽, 史军, 杨引明等. 中国东部植被 NDVI 对气温和降水的旬响应特征[J]. *地理学报*, 2009, 64(7): 850-860.
- [27] 全球变化科学研究数据出版系统编辑部. 全球变化科学研究数据出版系统数据共享政策[OL]. 2014. DOI:10.3974/dp.policy.2014.05.
- [28] 王永立, 范广洲, 周定文等. 我国东部地区 NDVI 与气温、降水的关系研究[J]. *热带气象学报*, 2009, 25(6): 725-732.
- [29] 戴声佩, 张勃, 王海军. 中国西北地区植被 NDVI 的时空变化及其影响因子分析[J]. *地球信息科学学报*, 2010, 12(3): 315-321.
- [30] 钞振华. 三种再分析气温资料在中国西部地区的可信度评价[J]. *大气科学学报*, 2011, 34(2): 162-169.
- [31] 李川, 张廷军, 陈静. 近 40 年青藏高原地区的气候变化——NCEP 和 ECMWF 地面气温及降水再分析和实测资料对比分析[J]. *高原气象*, 2004, 23(S1): 97-103.