

中国太阳总辐射 $5^{\circ} \times 5^{\circ}$ 格点数据集的建立

韩瑞 刘娜 任芝花 沈文海
(国家气象信息中心, 北京 100081)

摘要: 选取1961—2015年总辐射观测数据, 采用局部薄盘光滑样条插值方法(TPS), 得到适合中国地区太阳辐射分布型的曲面拟合模型。基于1961—2015年总辐射的格点数据, 通过平均偏差、平均绝对误差、均方根误差(RMSE)等评估指标全面分析辐射格点数据的误差分布及其季节差异。统计结果表明: 1) 所用的插值模型对辐射场的插值效果较好, 1961年1月—2015年12月总辐射的广义交叉验证值的平方根(RTGCV)主要集中在 $1.95 \sim 6.24 \text{ MJ} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$; 2) 格点总辐射数据的插值精度较高。1961—2015年累年平均的总辐射的分析值(插值结果)相对于观测值的平均偏差主要集中在 $\pm 0.5 \text{ MJ} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$ 间, 格点分析值相对于实测值的平均偏差为 $0.0009 \text{ MJ} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$, 均方根误差为 $0.9710 \text{ MJ} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$ 。误差统计量随季节的变化特征较明显, RTGCV、RMSE值在夏季最大、春秋次之、冬季最小。

关键词: 中国, 太阳总辐射, 空间内插, 观测数据

DOI: 10.3969/j.issn.2095-1973.2018.01.012

Establishment of Grid Datasets for the Total Solar Radiation over China

Han Rui, Liu Na, Ren Zhihua, Shen Wenhai

(National Meteorological Information Center, Beijing 100081)

Abstract: Based on total solar radiation observation data for 1961-2015, we used the Thin Plate Spline (TPS) method to obtain a surface fitting model that is suitable for the distribution of solar radiation in China. Based on the grid data, the error distribution and seasonal variation of radiation grid data are comprehensively analyzed by means of average deviation, average absolute error and root mean square error (RMSE). The results show that: (1) the interpolation model got a good effect in the radiation field for 1961-2015, the square root of the total radiation value of the generalized cross validation (RTGCV) mainly in $1.95 \sim 6.24 \text{ MJ} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$, (2) the interpolation precision is relatively high. The averaged errors of interpolated values for total radiation from that of observations are mainly concentrated in the $\pm 0.5 \text{ MJ} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$ for 1961-2015, the averaged deviation of the interpolated lattice values from that of the actual values is $0.0009 \text{ MJ} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$, the RMS error is $0.9710 \text{ MJ} \cdot \text{m}^{-2} \cdot \text{d}^{-1}$. The variation of error statistics with seasons is obvious. The seasonal variation of the RTGCV and RMSE are obvious, the biggest one is in summer, the second in spring and autumn, and the smallest in winter.

Keywords: China, spatial interpolation, observation data, total solar radiation

0 引言

太阳是个巨大的炽热的气体球, 从太阳发出的辐射能为 $3.94 \times 10^{26} \text{ W}$, 虽然投射到大气顶部的太阳辐射只有48%到达地面, 但正是这部分能量引起了大气和海洋环流以及推动了地球上包括地表的所有过程的基本原因^[1]。所以, 辐射过程对于地球生物系统及人类的生存和发展起着决定性作用。因此辐射观测资料是研究天气变化、气候变迁最基本的气象资料之一。

我国的辐射观测站分为一级站、二级站与三级站, 按照气象辐射观测规范, 目前三个级别的辐射站

主要观测项目为太阳总辐射。太阳总辐射是指太阳直接辐射和天空散射辐射到达水平面上的总量^[1]; 太阳直接辐射是指太阳以平行光线的形式直接投射到地面上的辐射; 散射辐射是指太阳辐射经过大气散射或云的反射, 从天空 2π 立体角以短波形式向下, 到达地面的那部分辐射。

我国的辐射站较少, 且台站又是离散分布的, 所以这种将观测台站的数据通过空间内插的方法分配到每个网格中, 形成的格点数据可以在一定程度上解决空间分布不均的问题。同时, 这类数据对于天气分析、气候变化以及数值模拟研究具有十分重要意义, 特别是作为生态过程模型、水文模型、大气环流式和陆面过程的重要参数, 有着广阔的应用前景。国家气象信息中心首次将格点化技术应用于国家级的辐射数

收稿日期: 2017年7月31日; 修回日期: 2017年10月3日

第一作者: 韩瑞(1986—), Email: hanr@cma.gov.cn

资助信息: 中国气象局气候变化专项(CCSF201439)

据,也为今后的更高精度、多源融合的辐射资料奠定基础。

本文基于2013年最新整编的一套数据完整性和质量均提升的气象辐射基础资料,利用薄盘样条法(TPS),结合三维地理空间信息进行空间插值,建立了中国太阳总辐射 $5^{\circ} \times 5^{\circ}$ 格点数据集(V1.0),数据集包含了1961年1月以来的中国太阳总辐射日曝辐量、月曝辐量水平分辨率 $5^{\circ} \times 5^{\circ}$ 的格点数据。在此基础上从交叉验证和误差分析角度对日值格点数据质量进行评估分析。

1 资料和方法

1.1 资料

建立太阳总辐射格点数据集的过程中,用于空间插值的基础资料来源包括两个部分。

1) 中国国家级地面气象辐射站资料。该资料是1961年1月以来的全国国家级台站的辐射观测资料,该资料来源于各省、市、自治区气候资料处理部门逐月上报的《气象辐射记录月报表》信息化资料,由国家气象信息中心基础资料专项收集、整理,并经过严格的检查和审核。自建站以来,全国许多台站经过了业务改革、台站迁移、更迭等历史变迁,从1957年5月的3个站至1961年1月台站总数超过70站,20世纪70年代末期台站数达到80站左右,1993年前后,由于仪器换型,台站数有些波动;2005年之后稳定维持在99站(图1)。截至2015年12月全国累计共有99个国家级台站,其空间分布及中国陆地 $5^{\circ} \times 5^{\circ}$ 的数字高程模型数据如图1网格所示,我国地区的辐射台站分布呈现出东多西少,平原台站多,山地、高原、沙漠等特殊地形地区分不少的特点。例如:长三角地区单一网

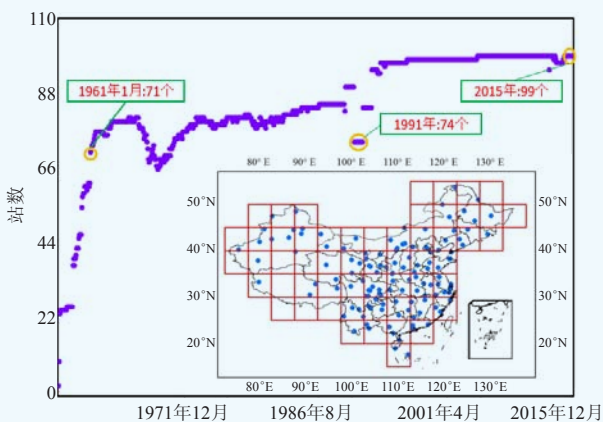


图1 1957年5月—2015年12月年中国地面辐射台站站点数变化曲线及空间分布

Fig. 1 Variation of station numbers from May 1957 to December 2015, and the distribution of station numbers at grid boxes ($5^{\circ} \times 5^{\circ}$) in China mainland for 2015

格最多台站分布有8个;而在西藏高原地区网格内没有台站分布。

2) 高程数据。与温度相同,太阳总辐射的空间分布受海拔的影响显著,格点化过程充分考虑海拔因子。根据钱永兰等^[2]通过对1 km、5 km和10 km三种不同DEM分辨率的插值结果进行分析,认为:随着DEM分辨率的提高,插值误差会减小,在地形起伏较大的地区更为明显所以为了减小误差;实际研究中站点海拔高度与对应网格的DEM存在差异,赵煜飞等^[1]完成了站点实测值与对应网格值的比较,认为台站的空间密度和地形对插值误差的影响较大。由于我国辐射台站数量有限,所以本文利用GTOPO30数据(分辨率为1 km \times 1 km)重采样生成中国陆地 $5^{\circ} \times 5^{\circ}$ 的数字高程模型(DEM)数据,通过使用高分辨率的DEM数据,以尽量减小由于地形和台站数量带来的误差。

1.2 格点化方法

现阶段研究中,已提出了大量的插值方法,包括反距离加权法、样条函数法、普通克里格方法、贝叶斯插值法、薄盘光滑样条法、多元线性回归方法等,这些方法中以Hutchinson^[4-5]提出的薄盘样条法应用最为广泛。

TPS法是将空间分布作为观测数据的函数而不需要其先验知识和物理过程,从而提高数据准确度。TPS方法除了普通的样条自变量外,引入线性协变量子模型,事实上薄盘样条函数可以理解为广义的标准多变量线性回归模型,只不过其参数是用一个合适的非参数化光滑函数代替。由广义交叉验证(GCV)的最小化来确定数据保真度与曲面的粗糙度之间起的平衡作用。GCV的计算可采用“one point move”方法,依次移去一个样点,用剩余样点在一定的光滑参数下进行曲面拟合得到该点的估测值,再计算观测值与估测值的方差。

ANUSPLIN软件^[6]是使用TPS法对多变量数据进行空间内插的工具,根据适合的插值模型方案定义自变量,同时还可以按一定时间顺序批量化的处理数据,且其能对插值结果能够提供多种评估参数,包括:数据平均值、方差、光滑参数(RHO)、广义交叉验证、均方根误差(RMSE)、平均偏差(MBE)、相对偏差(RBE)等指标。由于ANUSPLIN软件有如上所述的优点,故选取此软件完成对1961—2015年中国辐射99站的格点化处理工作。

1.3 评估方法及评估指标

为了检验插值方法和插值模型对辐射场的插值效果,采用GCV方法进行评估,首先假设每个站点的观

测值未知,用周围站点的值来估算,然后根据所有站点实际观测值与估算值的误差大小评判插值方法的优劣。GCV的平方根(RTGCV)由观测数据误差和预测数据误差两部分组成,而MSE的平方根(RMSE)是剔除观测数据误差后的预测误差估计,相当于插值过程的真实误差。另外,数据本身的误差也会造成GCV偏大。

采用MBE、RBE、RMSE、RTGCV等作为评估插值效果的指标,值越小,表面插值效果越好。具体计算公式^[7]如下:

$$MBE = \frac{1}{N} \sum_{i=1}^N (P_i - O_i), \quad (1)$$

$$RBE = \frac{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)}{\frac{1}{N} \sum_{i=1}^N O_i} \times 100\%, \quad (O \neq 0), \quad (2)$$

$$RSME = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2}, \quad (3)$$

$$RTGCV = \sqrt{GCV}, \quad (4)$$

式中, P_i 和 O_i 分别表示第*i*点上经过插值后的值和台站原始观测值, N 为统计样本数。

2 数据评估与分析

2.1 交叉验证

对1961—2015年的全国参与格点化的每个站点做交叉验证,根据所有站点RTGCV评判插值模型的优劣。图2给出了RTGCV的逐月变化曲线。可以看到,自1961年1月以来,RTGCV波动具有一定的规律,在1.94~6.24 MJ·m⁻²·d⁻¹范围内变化,且具有显著的周期特征,其周期为1 a。

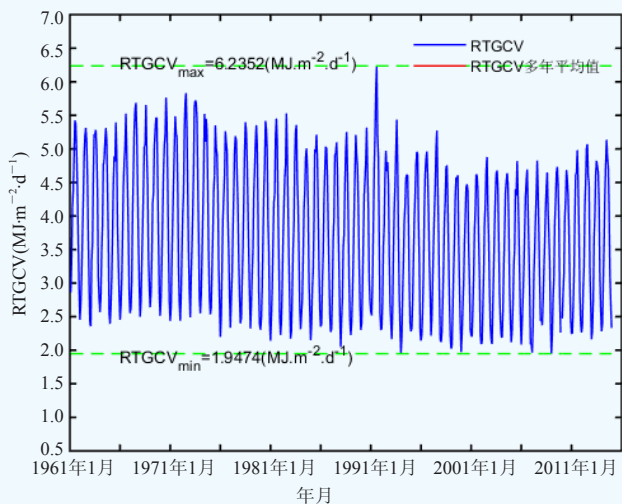


图2 1961年1月—2015年12月总辐射值RTGCV的逐月变化
Fig. 2 Monthly RTGCV from January 1961 to December 2015

为了更加清晰呈现RTGCV的周期变化,图3给出了2011年1月—2015年12月总辐射量的RTGCV随时间演变情况。可以看出,2011年1月—2015年12月总辐射量的RTGCV主要在2.16~5.14 MJ·m⁻²·d⁻¹波动。同时,总辐射量的RTGCV变化存在一定的季节变化,呈一个非稳定序列特征。即RTGCV夏季最大,春秋次之,冬季最小。其中,这主要是由不同季节的太阳辐射不同导致的,夏季太阳辐射强度较大,太阳辐射绝对量大,RTGCV则偏大。

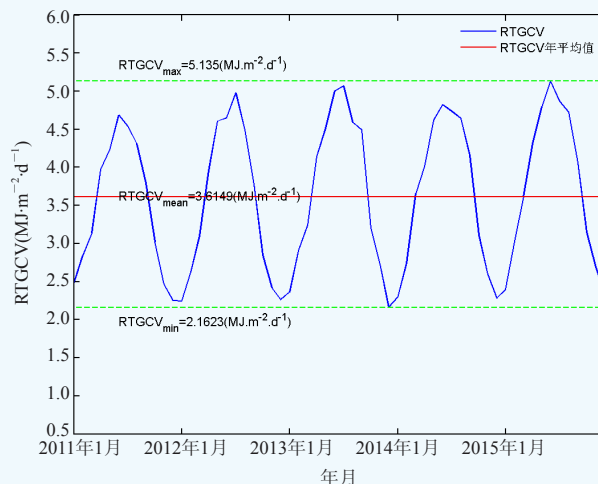


图3 2011年1月—2015年12月总辐射值RTGCV的逐月变化
Fig. 3 Monthly RTGCV from January 2011 to December 2015

2.2 误差分析

2.2.1 平均偏差、相对偏差

为反映插值格点数据的平均误差,本研究统计了1961年以来的累年平均分析值与观测值的平均偏差空间分布(图4)。总体上,全国平均偏差为0.0009 MJ·m⁻²·d⁻¹,主要集中在±0.5 MJ·m⁻²·d⁻¹之间,相对偏差主要分布在±5%之间。太阳总辐射插值平均偏差部分地区较大,主要分布在我国四川盆地、青藏高原、东南沿海以及黑龙江北部等地区。此外,统计分析了1961年以来太阳总辐射日值的分析值与观测值的平均偏差、相对偏差的频率(图4)。可以看出,平均偏差、相对偏差的频率基本呈正偏态分布,且90%以上的平均偏差集中在±0.5 MJ·m⁻²·d⁻¹之间,77%的相对偏差分布于±5%之间,92%分布于±10%之间。

2.2.2 均方根误差

图5给出了1961—2015年总辐射值的均方根误差随时间的演变情况。从RMSE的逐月变化曲线,可以看出,均方根误差呈现出周期性的变化,为了更加清晰呈现均方根误差的周期变化,给出了2011年1月—2015年12月RSME的逐月变化,均方根误差在夏季

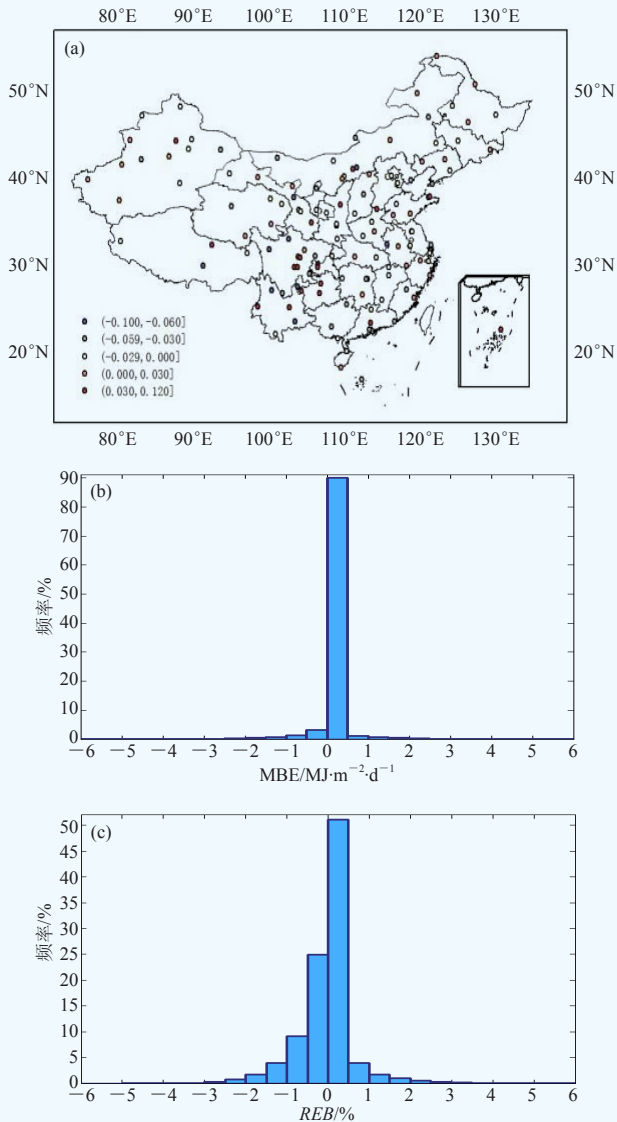


图4 1961—2015年总辐射插值与观测值平均偏差空间分布 (a)、平均偏差 (b)、相对偏差 (c) 分布频率
Fig. 4 (a) Distribution map of the MBE, (b) frequency distribution of the MBE and (c) of the REB from 1961 to 2015

达到最大，冬季最小，基本在 $0.44\sim 1.75\text{ MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ 。自1961年以来没有明显的年代际突变。1993年前后数据有很大波动，这与辐射业务相关调整有关。

2.3 太阳总辐射的季节变化

重点讨论基于该网格数据集中国太阳总辐射长时间序列的季节空间分布趋势。用中国太阳总辐射多年（1994—2015年）值插值后的网格数据作为多年平均值，分析了22年来中国太阳总辐射的季节变化特征。这里认为四季划分是春季为3—5月，夏季为6—8月，秋季为9—11月，冬季为12月—次年2月。为了更加准确的反映中国太阳总辐射，本节选取了1994年1月—2015年12月间的数据进行统计，该段时间的太阳

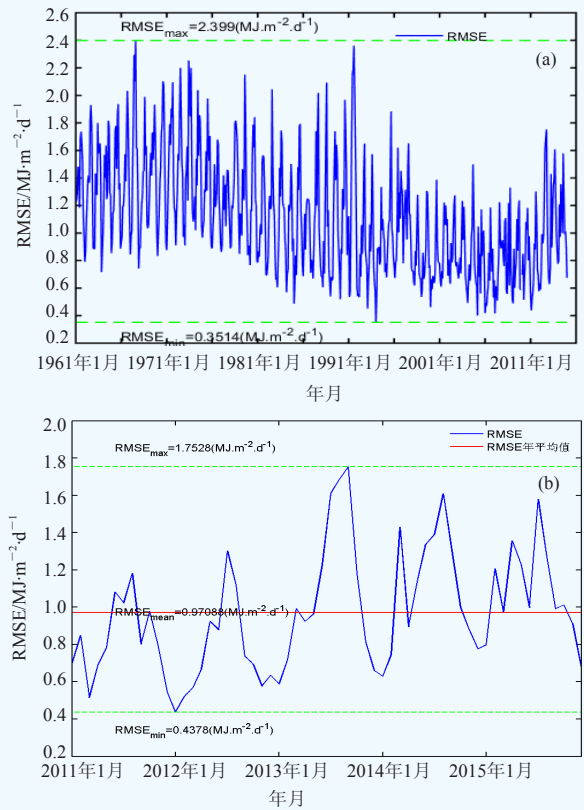


图5 太阳总辐射插值均方根误差时序变化：(a) 1961—2015年；(b) 2011—2015年
Fig. 5 Monthly RMSE: (a) for 1961 – 2015 and (b) for 2011 – 2015

总辐射台站数稳定在97个站以上且数据不间断；之前的年份，无论是台站数量，或是数据连续性都不能满足本节的数据统计需求。

从插值后的格点化中国太阳总辐射的季节空间分布图（图6）可知，插值后的格点化太阳总辐射数据空间分布和站点观测资料的气候态空间分布一致。格点化太阳辐射数据能够描述出东部地区低、西北地区高的总体特征，同时也反映了因地形因素引起的四川盆地太阳总辐射低值区，以及西北地区天山南北麓的准噶尔盆地太阳总辐射低值区和塔里木盆地、柴达木盆地的太阳总辐射高值区特征，但也平滑了这些地区太阳总辐射极值。综上，说明该数据集能够比较细致、准确地描述中国太阳总辐射的主要空间分布特征。

3 结论

本文通过选取中国1961—2015年总辐射观测数据，采用TPS插值方法，得到1格点数据，并通过平均偏差、平均绝对误差、均方根误差（RMSE）、GCV等评估指标分析辐射格点数据的误差分布及其季节差异，认为该格点数据产品的插值精度较高；主要结论如下：

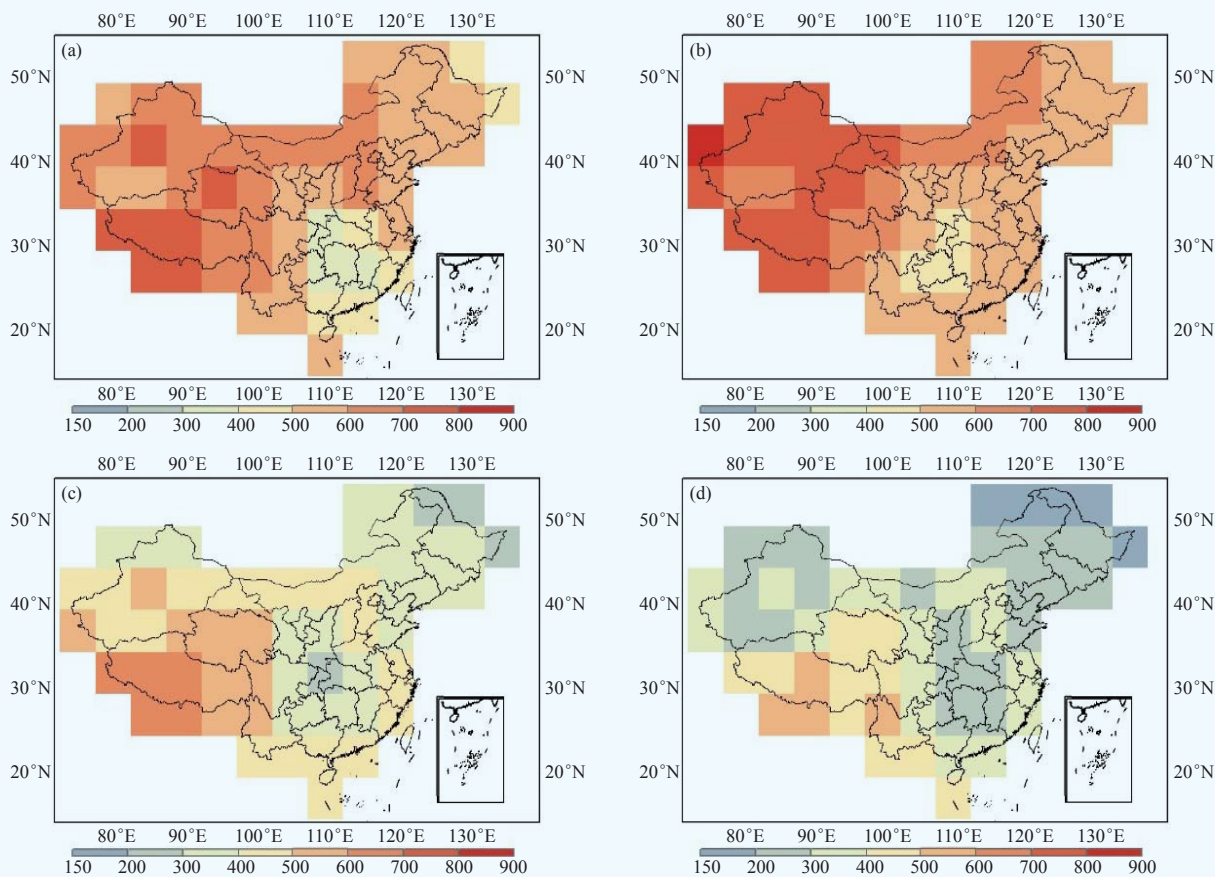


图6 1994—2015年中国太阳总辐射的季节空间分布图 ($\text{MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$)

(a) 春季; (b) 夏季; (c) 秋季; (d) 冬季

Fig. 6 Distribution map of grid mean in total solar radiation for spring (a), summer (b), autumn (c), winter (d)

1) 在广义交叉验证中, RTGCV波动具有一定的规律, 在 $1.94\sim 6.24 \text{ MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ 范围内变化, 且具有显著的周期特征, 其周期为1 a。总辐射量的RTGCV变化存在一定的季节变化, 呈一个非稳定序列特征。即RTGCV夏季最大, 春秋次之, 冬季最小。这主要是由不同季节的太阳辐射不同导致的, 夏季太阳辐射强度较大, 太阳辐射绝对量大, RTGCV则偏大。同时也由此证明所用的插值模型合理, 插值效果较好。

2) 1961—2015年累年平均的总辐射的分析值(插值结果)相对于观测值的平均偏差主要集中在 $\pm 0.5 \text{ MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ 之间, R格点分析值相对于实测值的平均偏差为 $0.0009 \text{ MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$, 均方根误差为 $0.9710 \text{ MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ 。平均偏差、相对偏差的频率基本呈正偏态分布, 且90%以上的平均偏差集中在 $\pm 0.5 \text{ MJ}\cdot\text{m}^{-2}\cdot\text{d}^{-1}$ 之间, 77%的相对偏差分布于 $\pm 5\%$ 之间, 92%分布于 $\pm 10\%$ 之间。且自1961年以来没有明显的年代际突变。1993年前后数据有很大波动, 这与辐射业务相关调整有关。

3) 1994—2015年中国太阳总辐射插值后格点化的数据, 其空间分布和站点观测资料的气候态空间分布一致。格点化太阳辐射数据能够描述出东部地区低、西北地区高的总体特征, 说明该数据集能够比较细致、准确地描述中国太阳总辐射的主要空间分布特征。

参考文献

- [1] 中国气象局. 气象辐射观测方法. 北京: 气象出版社, 1996.
- [2] 钱永兰, 吕厚荃, 张艳红. 基于ANUSPLIN软件的逐日气象要素插值方法应用与评估. 气象与环境学报, 2010, 26(2): 7-15.
- [3] 赵煜飞, 朱江, 许艳. 近50a中国降水格点数据集的建立及质量评估. 气象科学, 2014, 34(4): 414-420.
- [4] Hutchinson M F. Interpolation of rainfall data with thin plate smoothing splines - Part I: Two dimensional smoothing of data with short range correlation. Journal of Geographic Information and Decision Analysis, 1998, 2: 139-151.
- [5] Hutchinson M F. Interpolation of rainfall data with thin plate smoothing splines - Part II: Analysis of topographic dependence. Journal of Geographic Information and Decision Analysis, 1998, 2: 152-167.
- [6] Hutchinson M F. ANUSPLIN version 4.3 user guide. Canberra: The Australia National University, Center for Resource and Environment Studies, 2004.
- [7] 赵煜飞, 朱亚妮. 中国地面均一化相对湿度月值格点数据集的建立. 气象, 2017, 43(3): 333-340.