

论观测数据及其准确性之重要性

白建辉¹, 万晓伟¹, 柴文海¹, 吴翼美¹, 李凯丽², 宋 涛²

¹中国科学院大气物理研究所LAGEO, 北京

²南京中科华兴应急科技研究院有限公司, 江苏 南京

收稿日期: 2024年5月8日; 录用日期: 2024年6月11日; 发布日期: 2024年6月25日

摘 要

本文较全面地分析了地球上位于三极、低纬度、中纬度地区5个典型区域站点太阳总辐射测量数据及其经验模型计算结果(包括各种误差)。研究发现,太阳辐射测量和计算结果的数据质量在5个站点存在差别,其中以南极地区Dome C站测量和计算结果的数据质量最好。进而讨论了产生这些差异的主要因素,提出了关于太阳辐射日常测量、仪器维护以及经验模型建立等方面的一些建议,以便获得可靠、准确、高质量的太阳辐射数据及其计算模型。进一步,本文开展了扩展研究,简要分析了生物挥发性有机物(BVOCs)排放、碳平衡(包括初级生产力GPP、呼吸Re、净生态系统生产力NEP)的测量数据和经验模型的计算结果,获得类似结论,即测量数据的质量是科学研究中的重要基础,高质量的测量数据在模型建立、检验、评价等方面发挥着重要作用。

关键词

太阳辐射, 辐射表, 测量误差, 经验模型, 生物挥发性有机物

Study on the Importance of the Observatioanl Data and Their Accuracies

Jianhui Bai¹, Xiaowei Wan¹, Wenhai Chai¹, Yimei Wu¹, Kaili Li², Tao Song²

¹LAGEO, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing

²Nanjing Zhongkehuaxing Emergency Science and Technology Research Institute, Nanjing Jiangsu

Received: May 8th, 2024; accepted: Jun. 11th, 2024; published: Jun. 25th, 2024

Abstract

This study thoroughly analyzed observational data and their calculations (including various biases) of global solar radiation using its empirical model at five typical stations in the three polar, low and middle latitude regions on the Earth. It was found that the data qualities of the observa-

tional data and their calculations of global solar radiation were different, and the Dome C station had the highest quality among all stations. Then, the key factors influencing the observational data and their calculations were discussed. Some suggestions on routine measurement of solar radiation and the development of empirical model of solar radiation were proposed, so as to obtain accurate, reliable and high quality of solar radiation data as well as the empirical model of global solar radiation. Furthermore, this study also analyzed observational data and their simulation results using empirical model of the emissions of biogenic volatile organic compounds (BVOCs) and carbon balance (including gross primary production GPP, respiration Re and net ecosystem productivity NEP) at a typical forest in China, and the similar conclusions were obtained, that is the observational data are an important basis for the scientific research, and the high quality observational data play vital roles in the development, validation and evaluation of models.

Keywords

Solar Radiation, Radiometer, Measurement Error, Empirical Model, Biogenic Volatile Organic Compounds (BVOCs)

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

一切真知均来源于实践。科学研究(包括大气科学)是一门实验科学,观测实验和实验数据无疑具有非常重要的意义和价值,也将在科学研究中发挥无可替代的作用。观测实验大致可以分为室内实验和野外实验。室内实验可以设定不同的实验条件,通过控制各个自变量($x_i, i = 1, 2, 3, \dots, n$)的变化,确定因变量(y)与自变量及其变化之间的相互关系,从而获得因变量和自变量之间复杂的定量关系。野外实验则由于各个参量(自变量、因变量,以及其它因素等)的不可控性,它们均同步发生变化且彼此之间还进行着相互作用,涉及众多的物理、化学、生物过程。因而,导致很难确定因变量和自变量之间的相互关系。但是,野外实验也是认识自然、掌握自然规律不可或缺的重要方法之一。上述两种方法互为辅助,共同帮助我们全面认识自然和了解自然,最终促进人类和自然和谐相处、共同发展。

目前研究,对于数据处理方法虽然都有不同程度的涉及(包括数据评估方法),但是大多过于抽象和概念化[1] [2] [3] [4]。各类、特别是综合性实验数据的可靠获取和合理处理,对于自然规律的准确认识、了解、掌握、运用至为关键。本文基于之前典型区域站点研究数据及计算方法等内容,探讨实验数据及数据质量的重要意义,并提出一些具体建议。

2. 太阳辐射数据处理及其计算模型发展

理论来源的最初基础是实践,是对自然现象及其规律的细致观察、比较、分析、提炼、总结而成。牛顿三大定律也是在实验及推理的基础上发展而来。因而,实验是自然科学发展重要且必不可少的基础,实验数据的可靠性和准确性则具备了无与伦比的重要价值,并决定了在此基础之上所发展理论或模型的正确与否。

太阳辐射在大气中传输,主要经历大气成分的吸收和散射过程(以及大气顶和地面反射)。其吸收和散射过程可以表达为: 1) 吸收过程,大气中各类气液固相成分对太阳总辐射的直接吸收和间接利用,采用 e^{-kW_m} 来表达,整层大气吸收太阳辐射通量密度值 $\Delta S' = 0.172(mW)^{0.303}$ [5], $W = 0.021E \times 60$ (E 为地面水汽

压, hPa), m 为大气质量, $e^{-kWm} = 1 - \Delta S'/I_0$ 。太阳常数 $I_0 = 1.94 \text{ cal}\cdot\text{min}^{-1}\cdot\text{cm}^{-2}$ ($1367 \text{ W}\cdot\text{m}^{-2}$), k 为水汽吸收系数($0.70\sim 2.845 \mu\text{m}$)。水平面上的吸收能量为 $e^{-kWm} \times \cos(Z)$, Z 为太阳天顶角(度)。2) 散射过程, 大气中气液固相成分对总辐射的散射以 $e^{-S/G}$ 来表达, S 和 G 分别为太阳散射辐射和总辐射($\text{W}\cdot\text{m}^{-2}$)。

计算太阳总辐射(小时总量)的经验模型可以表达为:

$$G_{cal} = A_1 e^{-kWm} \times \cos(Z) + A_2 e^{-S/G_{obs}} + A_0 \quad (1)$$

G_{cal} 和 G_{obs} 分别为太阳总辐射的计算值和测量值($\text{MJ}\cdot\text{m}^{-2}$), 其余参量均使用小时总量(辐射值)和小时平均值(气象参数)。利用太阳辐射和气象参数的测量值和统计分析, 则可以客观确定总辐射经验模型各个系数和常数[6]。

鉴于太阳辐射测量上的误差, 同时也为了建立可信的总辐射计算模型, 因此, 首先需要对辐射测量数据进行筛选。本文主要介绍我们所研究的地球上 5 个典型区域代表性实验站点太阳总辐射数据及其处理方法和计算结果。

5 个代表性站点分别为: 1) 中国科学院千烟洲红壤丘陵农业综合开发试验站($26^{\circ}44'48''\text{N}$, $115^{\circ}04'13''\text{E}$, 110.8 m), 位于江西省泰和县; 2) 中国科学院珠穆朗玛峰大气与环境综合观测研究站(28.21°N , 86.56°E , 4276 m), 位于日喀则市定日县扎西宗乡; 3) 土耳其 Ankara province 站($39^{\circ}58'21.7''\text{N}$, $32^{\circ}51'49.3''\text{E}$, 890 m); 4) 北极 Sodankylä 站(67.367°N , 26.630°E , 184 m), 位于芬兰; 5) 南极 Dome C 站($75^{\circ}06'\text{S}$, $123^{\circ}21'\text{E}$, 3233 m), 位于南极圈内。

太阳总辐射经验计算方法, 最初始于千烟洲站的研究, 并使用本研究组测量数据的月平均值。筛选所用太阳辐射数据为: $Z < 70^{\circ}$, 针对的大气环境条件包括两种情形: 散射因子 $S/G < 0.80$ 和 $S/G \geq 0.80$ 。 S/G 表示大气中各种气液固相成分的相对含量。对于 $S/G < 0.80$ 和 $S/G \geq 0.80$ 的样本数(n)分别为 14 (对应着小时值 $n = 2779$)和 29 (用于研究的 2013~2016 年所有天气测量数据总样本数 $n = 15,668$)。基于千烟洲计算方法的经验, 后续研究的其他实验站太阳辐射使用数据的标准分别为: 珠峰站, $Z < 55^{\circ}$, $G > 50 \text{ W}\cdot\text{m}^{-2}$, $n = 7020$; Ankara 站, $Z < 75^{\circ}$, $G > 50 \text{ W}\cdot\text{m}^{-2}$, $n = 6160$; Sodankylä 站, $Z < 75^{\circ}$, $G > 20 \text{ W}\cdot\text{m}^{-2}$, $n = 3962$; Dome C 站, $Z < 75^{\circ}$, $G > 20 \text{ W}\cdot\text{m}^{-2}$, $n = 2771$ 。简要说明一下, 5 个典型站经验模型的发展基于千烟洲站计算方法研究经验(采用 2013~2016 年数据), 之后拓展到其他 4 个代表性区域及其站点。在发展这 4 个站点模型中, 选取了所获实验数据中某一时间段的数据。对于经验模型则使用了同一时间段的数据对比和不同时间段的数据检验等。例如, 对于珠峰站使用了 2008~2011 年小时数据用于模型建立, 并利用 2007~2020 年的数据对于经验模型采用了同一时间段的数据对比和不同时间段的模型验证。模型建立虽然使用了较早时间段(2008~2011 年)数据, 但其实用性和可靠性并不受此影响, 也即模型机理的合理性和客观性是建立模型的核心和关键。珠峰站以及其他 4 站利用不同实验数据的对比和检验已经证明此点, 具体内容可参见文献[6] [7] [8] [9] [10]。

各个站点使用的太阳辐射测量仪器如下: 千烟洲站, 分光辐射表(型号 TBQ-4-1, 波段范围 270~3200、400~3200、700~3200 nm), 直接辐射表(型号 TBS, 波段 270~3200 nm), 锦州 322 研究所生产。其他站点均为国外生产(型号 CM21、CM22、CH1 等, Kipp & Zonen Inc., 荷兰)。各个站点使用的气象参数(温度、湿度、地面水汽压)测量仪器为: 千烟洲站使用 HOBO 气象站(型号 H21, Onset Company, 美国), 其他站点大多使用国外仪器(型号 HMP45C-GM, Vaisala Inc., 芬兰)。上述所有仪器及其性能等的详细介绍请见参考文献[6] [7] [8] [9] [10]。

3. 太阳总辐射计算模型计算结果

在对太阳总辐射数据筛选的基础上, 建立了太阳总辐射计算模型(公式(1)), 5 个站点的计算结果见表 1~5、图 1~5。

Table 1. The coefficients for the empirical model of global solar radiation, determined using monthly mean measurements at the Qianyanzhou station in 2013~2016, along with statistical metrics, i.e., coefficient of determination (R^2), average of the absolute relative bias ($\bar{\delta}$, %), normalized mean square error (NMSE), standard deviations of calculated and observed global solar irradiance (σ_{cal} and σ_{obs} , respectively), and the mean bias error (MAD, in $\text{MJ}\cdot\text{m}^{-2}$ and %) and the root mean square error (RMSE, in $\text{MJ}\cdot\text{m}^{-2}$ and %). Sample number $n = 14$ for $S/G < 0.80$ and $n = 29$ for $S/G \geq 0.80$ [6]

表 1. 千烟洲站总辐射经验模型的系数、可决系数(R^2)、计算偏差绝对值的平均值($\bar{\delta}$, %)、归一化均方误差(NMSE)、计算值和测量值的标准差(σ_{cal} 、 σ_{obs})、平均偏差误差(MAD, $\text{MJ}\cdot\text{m}^{-2}$ 、%)、均方根误差(RMSE, $\text{MJ}\cdot\text{m}^{-2}$ 、%)。千烟洲站使用 2013~2016 年月平均数据, 对于 $S/G < 0.80$ 和 $S/G \geq 0.80$ 的样本数 n 分别为 14 和 29 [6]

S/G range	A_1	A_2	A_0	R^2	$\bar{\delta}$	NMSE	σ_{cal}	σ_{obs}	MAD		RMSE	
									($\text{MJ}\cdot\text{m}^{-2}$)	(%)	($\text{MJ}\cdot\text{m}^{-2}$)	(%)
<0.80	3.89	1.04	-1.04	0.69	8.73	0.02	0.32	0.39	0.15	8.68	0.23	13.38
≥ 0.80	3.89	0.32	-1.07	0.54	21.26	0.05	0.32	0.44	0.25	19.26	0.30	23.57

Table 2. As Table 1, for using hourly mean measurements in 2008~2011 at the Qomolangma station, $n = 7020$

表 2. 同表 1, 珠峰站, 使用 2008~2011 年小时数据, $n = 7020$

A_1	A_2	A_0	R^2	δ_{avg}	NMSE	σ_{cal}	σ_{obs}	MAD		RMSE	
								($\text{MJ}\cdot\text{m}^{-2}$)	(%)	($\text{MJ}\cdot\text{m}^{-2}$)	(%)
5.521	0.859	-1.011	0.712	9.80	0.01	0.53	0.63	0.26	8.82	0.31	11.38

Table 3. As Table 1, for using hourly mean measurements in 2017~2018 at the Ankara province station, $n = 6160$

表 3. 同表 1, Ankara province 站, 使用 2017~2018 年小时数据, $n = 6160$

A_1	A_2	A_0	R^2	δ_{avg}	NMSE	σ_{cal}	σ_{obs}	MAD		RMSE	
								($\text{MJ}\cdot\text{m}^{-2}$)	(%)	($\text{MJ}\cdot\text{m}^{-2}$)	(%)
3.587	2.977	-2.013	0.916	27.14	0.025	0.914	0.955	0.199	11.38	0.265	15.14

Table 4. As Table 1, for using hourly mean measurements in 2008~2011 at the Sodankylä station, $n = 3962$

表 4. 同表 1, Sodankylä 站, 使用 2008~2011 年小时数据, $n = 3962$

A_1	A_2	A_0	R^2	δ_{avg}	NMSE	σ_{cal}	σ_{obs}	MAD		RMSE	
								($\text{MJ}\cdot\text{m}^{-2}$)	(%)	($\text{MJ}\cdot\text{m}^{-2}$)	(%)
2.521	2.523	-1.215	0.835	12.8	0.020	0.499	0.546	0.178	11.4	0.222	14.2

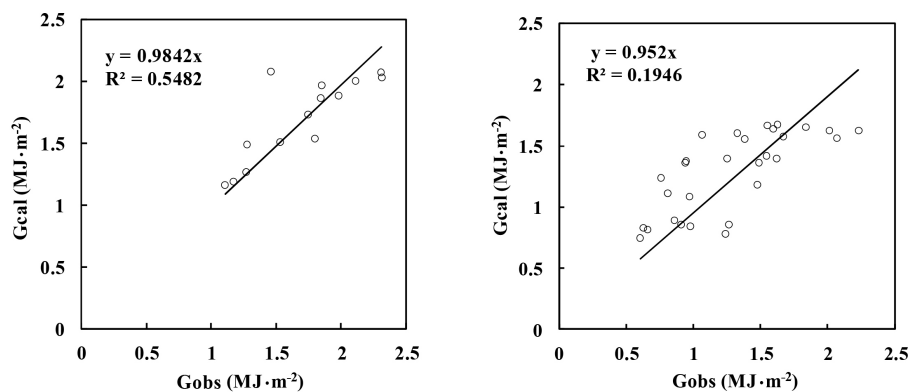


Figure 1. Scatter plot of calculated versus observed monthly global solar exposure (G_{cal} and G_{obs}) at the Qianyanzhou station (left and right for $S/G < 0.80$ ($n = 14$) and $S/G \geq 0.8$ ($n = 29$), respectively)

图 1. 千烟洲站太阳总辐射计算值和测量值(月均值)的散点图(左右图分别为 $S/G < 0.80$ 和 $S/G \geq 0.8$ 情形, $n = 14$ 和 29)

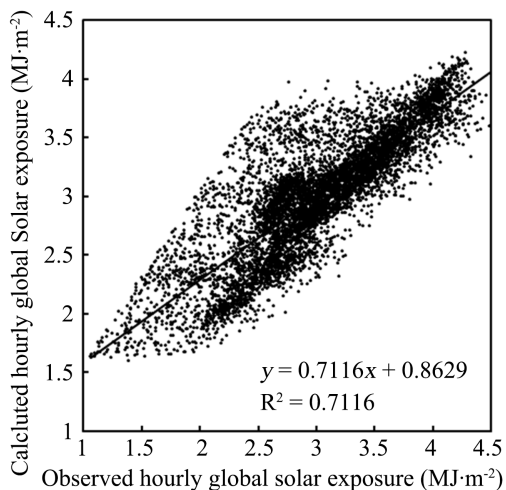


Figure 2. As Figure 1, for hourly global solar exposure under all-sky conditions at the Qomolangma station ($n = 7020$)
图 2. 同图 1, 珠峰站太阳总辐射计算值和测量值(小时值)的散点图($n = 7020$)

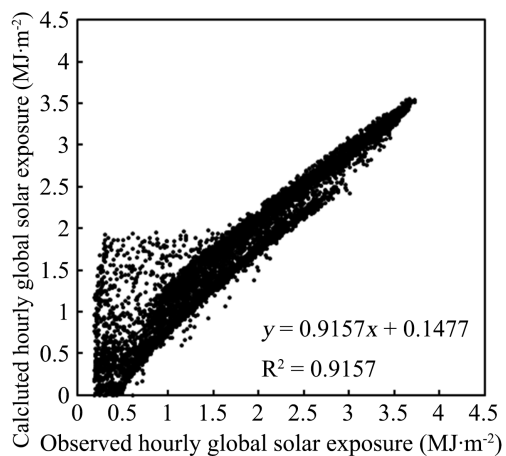


Figure 3. As Figure 2, for the Ankara province station ($n = 6160$)
图 3. 同图 2, Ankara province 站太阳总辐射计算值和测量值(小时值)的散点图($n = 6160$)

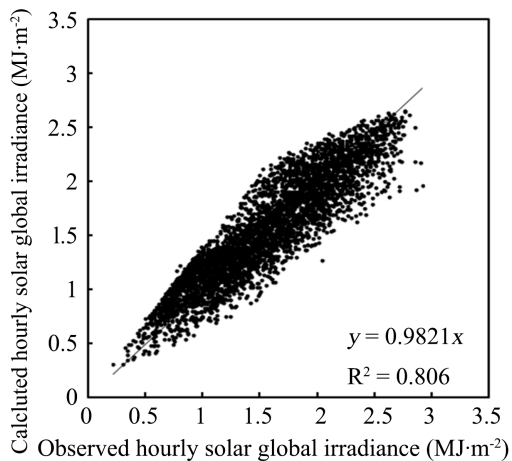


Figure 4. As Figure 2, for the Sodankylä station ($n = 3962$)
图 4. 同图 2, Sodankylä 站太阳总辐射计算值和测量值(小时值)的散点图($n = 3962$)

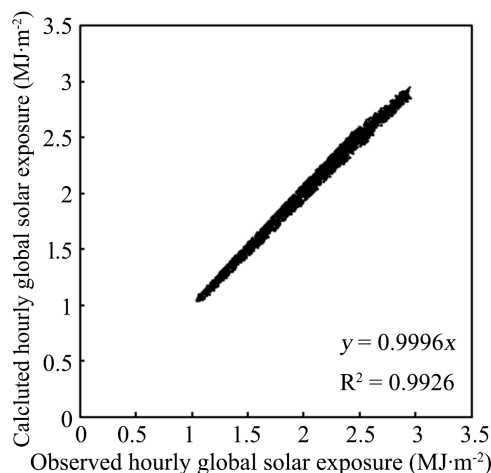


Figure 5. As Figure 2, for the Dome C station ($n = 2771$)

图 5. 同图 2, Dome C 站太阳总辐射计算值和测量值(小时值)的散点图($n = 2771$)

Table 5. As Table 1, for using hourly mean measurements in 2008~2011 at the Dome C station, $n = 2771$

表 5. 同表 1, Dome C 站, 使用 2008~2011 年小时数据, $n = 2771$

A_1	A_2	A_0	R^2	δ_{avg}	NMSE	σ_{cal}	σ_{obs}	MAD		RMSE	
								($\text{MJ}\cdot\text{m}^{-2}$)	(%)	($\text{MJ}\cdot\text{m}^{-2}$)	(%)
5.607	0.752	-1.097	0.993	1.76	0.0004	0.500	0.502	0.036	1.68	0.043	2.02

各个站点计算值和测量值时间序列对比图略(以节省篇幅),详细结果可以参见文献[4]-[10]。对比上述计算和测量结果,可以发现,经验模型较为合理地给出了总辐射模拟值。例如,千烟洲站的均方根误差为 0.23、0.30 $\text{MJ}\cdot\text{m}^{-2}$, 对应着大气条件 $S/G < 0.80$ 和 $S/G \geq 0.80$ 的情形, 珠峰、Ankara、Sodankylä、Dome C 站的均方根误差分别为 0.31、0.265、0.222、0.043 $\text{MJ}\cdot\text{m}^{-2}$ 。目前较好的 105 个经验模型中最好的 7 个模型, 它们的均方根误差为 0.22 $\text{MJ}\cdot\text{m}^{-2}$ [11]。5 个站点均方根误差最小的是 Dome C 站, 计算值和测量值散点图中的数据点非常好地集中在 1:1 线附近, 这主要是由于 Dome C 站是标准本底辐射观测网(BSRN, Baseline Surface Radiation Network)网站之一、执行严格和统一的观测规范, 该站辐射表等级较高、日常维护良好(辐射表遇故障及时替换、定期标定等), Dome C 站区域具有地球上最为干燥、洁净的大气, 即非常低的地面水汽压和 S/G (表征相对的大气成分含量)值, 表现在: Dome C、Sodankylä、Ankara、千烟洲站发展经验模型所用时间段的地面水汽压分别为 0.19、8.55、8.90、23.96 hPa, S/G 分别为 0.14、0.50、0.47、0.71, 另外, Dome C 站气溶胶光学厚度(AOD)是地球上最低的, 北极、珠峰、南极的 AOD 分别为 0.046、0.098、0.024 [7]; 北极 Sodankylä 和千烟洲站($S/G < 0.80$ 情形)的均方根误差均与 7 个最好的经验模型的均方根误差 0.22 $\text{MJ}\cdot\text{m}^{-2}$ [11] 非常一致, 它表明利用国产辐射表与利用国外辐射表测量数据, 可以同样获得比较好的计算效果。即在严格执行仪器使用规范、良好日常维护等条件下, 国产辐射表的性能是比较可靠的。另外, 在 2013~2016 年千烟洲的测量中, 我们对于辐射表的日常维护起着重要作用, 包括每日日出之前和降雨之后等情形下辐射表表头的清洁、蜘蛛网及时发现和清除、水平调整与保持、干燥剂及时更换等; 珠峰站均方根误差最大, 反映出该站区域较为严酷的大气环境、仪器日常维护比较困难等。相比 Dome C 站而言, 其他站点散点图中计算值和测量值的数据点大都比较发散, 表明仪器测量值偏差较大、进而带来计算值偏差也较大。上述结果揭示出辐射测量数据准确性具有极为重要的价值。因此, 台站辐射表的日常维护, 包括辐射表表面各种情况的及时处理(雨滴、雾水凝结、冰雹、雪花、沙尘、虫网等), 定期标定等, 需要引起高度重视。另外, 经验模型计算值的标准差均小于测量值的标准差, 也表明该总辐射经验模型具有较为可靠的性能。

总体而言, 经验模型在 5 个典型站点均表现出了较好的计算结果, 与目前国际上最好的经验模型也有较好的一致性(例如均方根误差)。它进一步揭示出在建立经验模型的过程中, 需要对测量数据采取严格的筛选标准, 以确保所用测量数据能客观、真实地反映太阳辐射的基本特征和变化规律。数据筛选使用天顶角, 主要是由于在太阳天顶角较大(高度角较小)的情形下, 例如较早和较晚时段, 辐射表测量误差(余弦误差)比较大; 另外, 在相对湿度较大、低温等情况下, 辐射表表面非常容易形成雾水、冰晶等, 这些情形一般多见于日出附近时段。

对比千烟洲站的计算结果, 可以看到总辐射的计算误差以 $S/G \geq 0.8$ 情形较大, 它表达着大气中气液固相成分较多的情况, 例如, 云、雨、雾、霾、雪、沙尘、较重污染、较高气溶胶(AOD)等。这些情形下, 上述因素也会带来辐射测量误差的增大, 进而造成计算偏差的增大。这从另一个侧面说明, 大气成分较高情形(即高 S/G , 多云、沙尘、高污染等)下, 总辐射模型计算精度的改进目前仍然是一个挑战[6]-[16], 还有待于进一步提高。

总辐射经验模型除了可以较好地模拟地面总辐射外, 还可以计算大气顶处的总辐射, 也表现出了较为合理的计算结果, 以 Sodankylä ($S/G < 0.30$, $n = 1322$)和 Dome C ($n = 2771$)站为例, 经验模型计算的大气顶总辐射(经验模型的系数之和 $A_1 + A_2 + A_0$)与太阳常数的比值分别为 0.999 和 1.069, 它们的相对偏差均比较小, 各个站点的具体结果可以参见文献[6] [7] [8] [9] [10]。进一步, 总辐射经验模型还可以计算损失于大气中的能量(大气顶辐射 - 地表辐射, 包括大气中吸收和散射成分引起的吸收和散射能量)、以及地表和大气顶的反照率等。研究表明, 5 个站点模型计算的地表和大气顶的反照率和地面测量及卫星反演的反照率均较为一致, 具体结果可以参见文献[6] [7] [8] [9] [10]。

4. 太阳总辐射计算模型应用

对于 5 个典型站点太阳总辐射的模拟结果, 以南极 DOME C 站为例给出。利用 DOME C 总辐射经验模型(2008~2011 年数据, $n = 2771$), 计算了总辐射 2006 年 1 月 1 日~2016 年 12 月 31 日小时值($n = 33,311$), 其详细结果可以参见文献[10]: 总体上, 经验模型较好地捕捉并展示了总辐射 2006~2016 年日到年尺度的变化特征, 低估其小时值约 18.40%, NMSE = 0.013, RMSE 为 $0.146 \text{ MJ}\cdot\text{m}^{-2}$ 和 10.90%。其 RMSE 仍然小于最好的 7 个模型的 RMSE ($0.22 \text{ MJ}\cdot\text{m}^{-2}$) [11]。图 6 给出总辐射月均值计算值和测量值以及散射辐射(S)、散射因子 S/G [11], 经验模型较好地描述了总辐射月平均的变化特征。研究结果表明, 经验模型虽然仅仅基于较少测量数据而建, 但是基于测量数据可靠且经过严格筛选、物理模型考虑机理合理等关键因素, 因而经验模型应用于 2008~2016 年实际天气情形下总辐射的计算, 表现出良好的模拟能力。经验模型计算 Sodankylä 和 Dome C 站大气顶总辐射和太阳常数之比分别为 0.999 ($n = 1322$)和 1.069 ($n = 2771$), 相对误差分别为 0.001% 和 6.9%, 这从另一个侧面表明测量数据的较高质量和经验模型的可靠性。

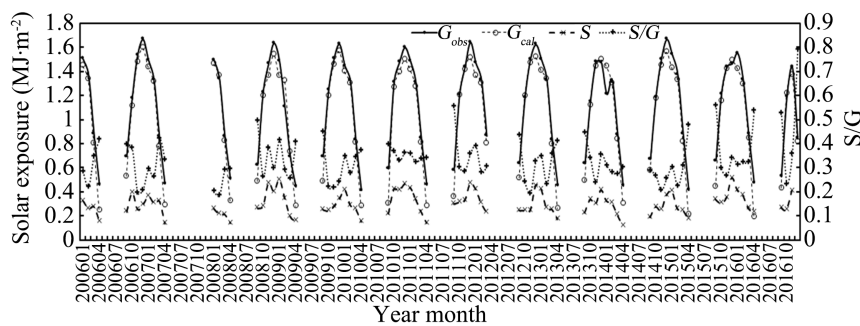


Figure 6. Monthly global solar exposures calculated and observed (G_{cal} , G_{obs}), observed diffuse exposure (S) and scattering factor (S/G) at the Dome C station ($n = 33,311$)

图 6. Dome C 站太阳总辐射 2006~2016 小时值的计算值(G_{cal})和测量值(G_{obs}), 测量的散射辐射(S)和散射因子(S/G) ($n = 33,311$)

5. 太阳总辐射表的标定

为了保证辐射表测量数据可靠,太阳辐射表需要定期送到指定部门进行标定,一般为2年。标定工作中,需要消耗大量的人力物力(标定所需软硬件、台站备用辐射表)、仪器拆卸和往来运输等。由于总辐射经验模型可以计算大气顶的总辐射,结合利用太阳常数(非常稳定,可作为总辐射标准,易于由目前各种方法获得,例如飞机和卫星等的测量),我们可以利用台站日常测量的辐射和气象参数,对各类辐射表进行标定,包括总、直接、散射、可见光、近红外、光合有效辐射(PAR)等,详细内容可参见文献[17]。新的标定方法简单易行、节省资源(如备用辐射表)、所需数据易于由实验站日常测量得到。另外,新的标定方法不需要标定设备,免去了辐射仪器的拆卸、安装及运输等。因此,建议台站按照观测规范要求,及时做好标定工作。

6. 讨论

利用地球上5个典型站点(即南北两极、珠峰第三极、低纬度和中纬度站点)严格筛选后的辐射测量数据,站点小时值的样本数范围在2771~7020组,建立了总辐射计算模型,得到了实际天气条件下较为合理、可靠的模拟结果。5个站点由于测量数据准确性的不同,其相关的各种误差(例如 δ 、NMSE、 σ_{cal} 、 σ_{obs} 、MAD、RMSE)存在差别,其中以南极DOME C站的计算效果最佳(总体上各种误差最小,包括测量值和计算值的),例如NMSE为0.0004, RMSE为 $0.043 \text{ MJ}\cdot\text{m}^{-2}$ 和2.02%,该值远远小于 $0.22 \text{ MJ}\cdot\text{m}^{-2}$ [11]。总体评价,辐射表质量、日常维护(包括每日巡视、表头及时清洁、水平保持和调整、干燥剂更换、定期标定等)、周边环境等是影响辐射测量误差和模型计算误差的主要影响因素。因而建议,每个台站应该高度重视辐射表的日常维护,安排人员做好每日巡视、清洁、故障及时处理、常规标定等工作,以确保测量数据的质量。安装在铁塔上的辐射表,虽然考虑到为保持辐射表传感器表面水平而采用了三角形结构支撑,但随着时间延续,难免其水平发生微小变化。所以,需要采取经常性的检查和及时调整,来保证辐射表传感器表面水平。另外,地面辐射测量数据的质量还影响着各种辐射模型(包括辐射传输)、卫星数据(包括辐射和反照率)的验证和评价,太阳辐射及其与大气成分相互作用在气候变化中的作用[6] [7] [8] [9] [10],具有非常重要的价值。从多方面来讲,需要高度重视台站辐射观测及其数据质量。

造成辐射表测量误差的一些主要因素,包括各种原因引起的辐射传感器接收太阳辐射量的变化,例如,雨滴、露水、沙尘、烟尘、降雪、鸟粪、蜘蛛网等,密封出现问题带来表头内部形成水滴。这些因素带来辐射变化的不规则性,后期很难对辐射数据进行订正。因此,需要针对上述情况,及时发现问题,及时处理。

辐射表的余弦误差也是测量误差的一个重要来源,例如,国产总辐射表的余弦误差一般为 $\leq 7\%$ (太阳高度角 10° 时,型号TBQ);进口总辐射表的余弦误差一般为:型号CM22, $<1\%$ (太阳天顶角 60° 时)、 $<8\%$ (天顶角 80° 时);型号CM21, $<\pm 10 \text{ W}\cdot\text{m}^{-2}$ (directional response,天顶角 80° 、辐射值 $1000 \text{ W}\cdot\text{m}^{-2}$)。此外,还有方位误差、温度响应等。因此,天顶角被用于数据筛选的重要标准。

对比千烟洲外的4个站点,Dome C、Sodankylä、珠峰、Ankara站总辐射测量值(小时值)的标准差分别为0.502, 0.546, 0.634, 0.955,它们从另一个侧面反映出辐射表测量数据的质量,即南北两极站点处于最好水平,其中以南极Dome C站更高一些;其次是第三极珠峰站,最后是Ankara站。

考虑到辐射表误差的主要来源,在发展经验计算模型之时,严格筛选数据是非常必要和重要的,它可以确保所建立总辐射计算模型的客观性、代表性、准确性,即较为准确且客观地描述实际大气条件下太阳总辐射的传输。进一步,还可以在此基础上,准确计算大气顶的总辐射(用于标定辐射表)以及地表和大气顶的反照率[6] [7] [8] [9] [10],开展太阳辐射对于气候变化作用的相关研究[18] [19] [20] [21]。

进一步而言,测量数据高准确性的意义还表现在其他领域,1) 生物挥发性有机物(BVOCs)排放,2) 碳平衡,包括初级生产力(GPP)、呼吸(Re)、净生态系统生产力(NEP)。BVOCs 是植物通过生理活动向大气释放碳的一种方式,它们具有非常高的化学反应活性,可以和大气中其他成分(NO_x 、 SO_2 等)、水汽、OH 自由基等参与化学和光化学反应,进而生成 O_3 、小粒径颗粒物($\text{PM}_{2.5}$ 、 PM_1 等),它们还对于凝结核以及云的形成具有重要贡献,在上述过程中,各种气液固成分同时与太阳辐射(紫外、可见光、近红外等)发生相互作用,影响着辐射传输和辐射平衡,在区域气候及其气候变化中起着关键作用[6] [7] [8] [9] [10]。因此,我们需要在广泛测量 BVOCs 排放通量的基础上,建立适合我国森林生态系统的排放模型。在发展 BVOCs 排放模型的过程中,涉及到高质量排放数据的选取。因此,在考虑太阳辐射数据筛选的基础上,还采用了 BVOCs 排放通量 2 倍标准差(以及 $S/G > 0.5$)的筛选原则[22]。生态系统碳平衡在未来实现双碳目标、减缓全球气候变暖中起着关键作用,因而生态系统碳平衡经验模型也是目前迫切需求。上述 2 倍标准差的筛选原则也被成功应用于生态系统碳平衡(包括 GPP、Re、NEP)经验模型的建立[23] [24]。具体而言,BVOCs 的主要成分为异戊二烯和单萜烯,因此 BVOCs 排放模型包括异戊二烯和单萜烯排放模型。以我国亚热带森林(人工针叶林,江西省泰和县中国科学院千烟洲试验站)为例,建立其异戊二烯和单萜烯排放模型时,使用的样本数分别为 18 和 8 组,它们计算值与测量值的 $\bar{\delta}$ 分别为 32.8%和 2.6%。利用梯度方法测量的排放通量数据进行检验,计算值和测量值的相对偏差分别为: 2.0% ($n = 126$)和 28.9% ($n = 120$),针对异戊二烯平均值及其采样时间段日变化的平均值两种情形; 19.7% ($n = 120$)和 3.3% ($n = 120$),针对单萜烯平均值及其采样时间段日变化的平均值两种情形。BVOCs 排放通量测量的不确定性大约在 30%左右。在 BVOCs 排放模型的基础上,进一步发展了亚热带森林生态系统碳平衡(包括 GPP、Re、NEP)经验模型,类似地,碳通量模型检验也得到与 BVOCs 排放模型检验类似的结果,而且这些模型也表现出对于 GPP、Re、NEP 较好的模拟能力[23] [24]。上述两方面的经验模型也是使用较少测量数据而建立的,也取得了较为合理的计算结果[22] [23] [24]。限于篇幅,这两方面更加详细的结果不再赘述,具体内容可以参考文献[22] [23] [24]。因此,高质量实验数据在不同研究领域具备且发挥着至关重要的作用,数据质量的重要性要远远高于数据的数量[22] [23] [24]。进一步我们看到,理论和模型不仅要准确描述自然规律,同时,理论和模型还都带有测量数据的烙印。

7. 结论

基于对地球上 5 个典型区域代表性站点太阳辐射测量数据和经验模型计算结果的对比与分析,5 个实验站辐射测量数据质量表现出不同的水平,其中南极 Dome C 站的测量数据质量最好,其标准差为 0.502;总辐射经验计算模型可以较合理地模拟 5 个站的总辐射,其计算值标准差小于测量值标准差, RMSE 与国际上 7 个最好的经验模型 RMSE 值 0.22 较为接近,类似地,也是以 Dome C 站的计算效果最佳,其 RMSE 为 $0.043 \text{ MJ}\cdot\text{m}^{-2}$ 和 2.02%。多方面的有利因素(仪器质量、维护与标定、干洁的大气环境等)为 Dome C 站总辐射测量和模型计算的高质量数据提供了重要条件和良好基础。因此建议:实验站点务必做好日常巡视和仪器维护(特别是表头清洁和水平保持)、定期标定;除了采用目前的标定方法外,还可以采用新的标定方法,即利用实验站已有太阳辐射和气象参数来标定站上的辐射表。利用经验模型,计算 Sodankylä 和 Dome C 站大气顶总辐射和太阳常数之比分别为 0.999 和 1.069,相对误差分别为 0.001%和 6.9%。发展经验模型,需要针对测量仪器和测量数据的特征与种类(太阳辐射、天顶角、通量等)采取不同标准的筛选,从而建立准确的经验模型,获得可靠的计算数据,并开展相关研究。太阳辐射并结合挥发性有机物排放和碳平衡(初级生产力、呼吸、净生态系统生产力)测量数据和模型计算的结果,均表明高质量测量数据具有非常重要的价值,在模型发展、检验、评价等多方面和多领域发挥着关键作用,是科学研究工作中的重中之重。因此,在日常工作中应给予高度重视。

致 谢

本文各个实验站工作人员为了获取可靠的观测数据辛苦付出, 特此表示感谢。同时衷心感谢各位同仁的大力支持, 包括中国科学院大气物理研究所刘洪韬、宗雪梅, 中国科学院青藏高原研究所马耀明、王宾宾, 芬兰气象研究所 A. Heikkilä, (欧盟)联合研究中心 C. Lanconelli, 意大利国家研究委员会 A. Lupi、V. Vitale, 德国 AWI 极地与海洋研究所 A. Driemel, 土耳其极地研究所 E. Arslan 等。

基金项目

国家自然科学基金(NO. 42261144751)、中华人民共和国科技部-欧洲空间局合作“龙计划”4、5期(ID32771、59013)、中国科学院战略性先导科技专项(A类)(NO. XDA19070202)、南京中科华兴应急科技研究院有限公司资助。

参考文献

- [1] 阿布杜, 郑志国, 李洪敏. 医学实验数据的质量控制问题[J]. 数理统计与管理, 2002, 21(5): 42-46.
- [2] 刘洪, 黄燕. 我国统计数据质量的评估方法研究——趋势模拟评估法及其应用[J]. 统计研究, 2007, 24(8): 17-21.
- [3] 郭明才. 实验数据可疑值的检验及计算机程序设计[J]. 环境与健康杂志, 1996, 13(4): 173-174.
- [4] 多布拉, 次旦卓玛, 洛桑罗布, 甘茂, 黄涛, 格桑卓玛. 西藏地区 VP 型垂直摆倾斜仪观测数据质量及干扰分析[J]. 高原地震, 2022, 34(3): 65-72.
- [5] 白建辉. 华北地区紫外与可见光辐射在大气中的传输[J]. 环境科学学报, 2010, 30(5): 915-925.
- [6] Bai, J. and Zong, X. (2021) Global Solar Radiation Transfer and Its Loss in the Atmosphere. *Applied Sciences*, **11**, Article 2651. <https://doi.org/10.3390/app11062651>
- [7] Bai, J., Zong, X., Ma, Y., Wang, B., Zhao, C., Yang, Y., et al. (2022) Long-Term Variations in Global Solar Radiation and Its Interaction with Atmospheric Substances at Qomolangma. *International Journal of Environmental Research and Public Health*, **19**, Article 8906. <https://doi.org/10.3390/ijerph19158906>
- [8] Bai, J., Wan, X., Arslan, E. and Zong, X. (2024) Global Solar Radiation and Its Interactions with Atmospheric Substances and Their Effects on Air Temperature Change in Ankara Province. *Climate*, **12**, Article 35. <https://doi.org/10.3390/cli12030035>
- [9] Bai, J., Heikkilä, A. and Zong, X. (2021) Long-Term Variations of Global Solar Radiation and Atmospheric Constituents at Sodankylä in the Arctic. *Atmosphere*, **12**, Article 749. <https://doi.org/10.3390/atmos12060749>
- [10] Bai, J., Zong, X., Lanconelli, C., Lupi, A., Driemel, A., Vitale, V., et al. (2022) Long-Term Variations of Global Solar Radiation and Its Potential Effects at Dome C (Antarctica). *International Journal of Environmental Research and Public Health*, **19**, Article 3084. <https://doi.org/10.3390/ijerph19053084>
- [11] Bayrakçı, H.C., Demircan, C. and Keçebaş, A. (2018) The Development of Empirical Models for Estimating Global Solar Radiation on Horizontal Surface: A Case Study. *Renewable and Sustainable Energy Reviews*, **81**, 2771-2782. <https://doi.org/10.1016/j.rser.2017.06.082>
- [12] Glover, J. and McCulloch, J.S.G. (1958) The Empirical Relation between Solar Radiation and Hours of Sunshine. *Quarterly Journal of the Royal Meteorological Society*, **84**, 172-175. <https://doi.org/10.1002/qj.49708436011>
- [13] Bird, R.E. (1984) A Simple, Solar Spectral Model for Direct-Normal and Diffuse Horizontal Irradiance. *Solar Energy*, **32**, 461-471. [https://doi.org/10.1016/0038-092x\(84\)90260-3](https://doi.org/10.1016/0038-092x(84)90260-3)
- [14] Gueymard, C. (1993) Critical Analysis and Performance Assessment of Clear Sky Solar Irradiance Models Using Theoretical and Measured Data. *Solar Energy*, **51**, 121-138. [https://doi.org/10.1016/0038-092x\(93\)90074-x](https://doi.org/10.1016/0038-092x(93)90074-x)
- [15] Myers, D.R. (2005) Solar Radiation Modeling and Measurements for Renewable Energy Applications: Data and Model Quality. *Energy*, **30**, 1517-1531. <https://doi.org/10.1016/j.energy.2004.04.034>
- [16] Besharat, F., Dehghan, A.A. and Faghih, A.R. (2013) Empirical Models for Estimating Global Solar Radiation: A Review and Case Study. *Renewable and Sustainable Energy Reviews*, **21**, 798-821. <https://doi.org/10.1016/j.rser.2012.12.043>
- [17] Bai, J. (2019) A Calibration Method of Solar Radiometers. *Atmospheric Pollution Research*, **10**, 1365-1373. <https://doi.org/10.1016/j.apr.2019.03.011>
- [18] Turner, J., Lu, H., White, I., King, J.C., Phillips, T., Hosking, J.S., et al. (2016) Absence of 21st Century Warming on

- Antarctic Peninsula Consistent with Natural Variability. *Nature*, **535**, 411-415. <https://doi.org/10.1038/nature18645>
- [19] Cao, Y. and Liang, S. (2018) Recent Advances in Driving Mechanisms of the Arctic Amplification: A Review. *Chinese Science Bulletin*, **63**, 2757-2771. <https://doi.org/10.1360/n972018-00462>
- [20] Cohen, J., Screen, J.A., Furtado, J.C., Barlow, M., Whittleston, D., Coumou, D., *et al.* (2014) Recent Arctic Amplification and Extreme Mid-Latitude Weather. *Nature Geoscience*, **7**, 627-637. <https://doi.org/10.1038/ngeo2234>
- [21] Zavalishin, N.N. (2021) Reasons for Modern Warming: Hypotheses and Facts. *Journal of Atmospheric Science Research*, **5**, 11-17. <https://doi.org/10.30564/jasr.v5i1.4080>
- [22] Bai, J. and Duhl, T. (2021) A Primary Generalized Empirical Model of BVOC Emissions for Some Typical Forests in China. *Atmospheric Pollution Research*, **12**, Article 101126. <https://doi.org/10.1016/j.apr.2021.101126>
- [23] Bai, J., Yang, F., Wang, H. and Xu, M. (2023) An Empirical Model of Gross Primary Productivity (GPP) and Relations between GPP and Its Driving Factors, Biogenic Volatile Organic Compounds in a Subtropical Conifer Plantation in China. *Atmosphere*, **14**, Article 1046. <https://doi.org/10.3390/atmos14061046>
- [24] Bai, J., Yang, F., Xu, M. and Wang, H. (2023) Empirical Models of Respiration and Net Ecosystem Productivity and Their Applications in a Subtropical Coniferous Plantation in China. *Atmosphere*, **14**, Article 1557. <https://doi.org/10.3390/atmos14101557>