**Atmospheric
Measurement
Techniques**

Assessment of BSRN radiation records for the computation of monthly means

A. Roesch[1]**, M. Wild**[1]**, A. Ohmura**[1]**, E. G. Dutton**[2]**, C. N. Long**[3]**, and T. Zhang**[4]

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Universitaetsstrasse 16, 8092 Zurich, Switzerland
[2]NOAA Earth System Research Laboratory GMD, 325 Broadway, Boulder, CO 80305, USA
[3]Pacific Northwest Laboratory, 902 Batelle Boulevard, Richland, WA, USA
[4]NASA Langley Research Center, Hampton, VA 23681, USA

Abstract. The integrity of the Baseline Surface Radiation Network (BSRN) radiation monthly averages are assessed by investigating the impact on monthly means due to the frequency of data gaps caused by missing or discarded high time resolution data. The monthly statistics, especially means, are considered to be important and useful values for climate research, model performance evaluations and for assessing the quality of satellite (time- and space-averaged) data products. The study investigates the spread in different algorithms that have been applied for the computation of monthly means from 1-min values.

The paper reveals that the computation of monthly means from 1-min observations distinctly depends on the method utilized to account for the missing data. The intra-method difference generally increases with an increasing fraction of missing data. We found that a substantial fraction of the radiation fluxes observed at BSRN sites is either missing or flagged as questionable. The percentage of missing data is 4.4%, 13.0%, and 6.5% for global radiation, direct shortwave radiation, and downwelling longwave radiation, respectively. Most flagged data in the shortwave are due to nighttime instrumental noise and can reasonably be set to zero after correcting for thermal offsets in the daytime data. The study demonstrates that the handling of flagged data clearly impacts on monthly mean estimates obtained with different methods. We showed that the spread of monthly shortwave fluxes is generally clearly higher than for downwelling longwave radiation.

Overall, BSRN observations provide sufficient accuracy and completeness for reliable estimates of monthly mean values. However, the value of future data could be further increased by reducing the frequency of data gaps and the number of outliers. It is shown that two independent methods for accounting for the diurnal and seasonal variations in the missing data permit consistent monthly means to within less than 1 W m^{-2} in most cases. The authors suggest using a standardized method for the computation of monthly means which addresses diurnal variations in the missing data in order to avoid a mismatch of future published monthly mean radiation fluxes from BSRN.

The application of robust statistics would probably lead to less biased results for data records with frequent gaps and/or flagged data and outliers. The currently applied empirical methods should, therefore, be completed by the development of robust methods.

1 Introduction

In this work we investigate the extent of the differences that can be caused by various potential data filling methodologies for surface radiation quantities. We do not intend or claim to have identified the ultimately best and least-error filling method but rather demonstrate the impact of a range of potential methods that could and undoubtedly have been used by researchers when working with data sets with realistic gaps caused by unavoidable and other observational issues.

Anthropogenic interference with climate occurs first through a perturbation of the Earth's radiation balance (e.g., Ramanathan et al., 2001). Despite the central role that the radiation balance plays in the climate system, considerable uncertainties remain with respect to its mean state and temporal variation, as well as its representation in climate models (Wild et al., 1995; Wild, 2008). Attempts are underway

Correspondence to: A. Roesch
(andreas.roesch@env.ethz.ch)

to monitor changes in the radiation balance from both the surface and space. More and more studies, particularly those based on surface observations, present evidence that the radiative fluxes are not stable over time but undergo significant decadal variations (e.g., Gilgen et al., 1997; Stanhill and Cohen, 2001; Liepert, 2002; Dutton et al., 2006; Wild, 2009 and references therein), which may have major consequences for the climate system and climate change (Wild, 2009). However, all these analyses rely on data (typically monthly or yearly means), that have been aggregated in some way or another from incomplete raw data with much higher temporal resolution (typically minute to hourly). While the way to do this aggregation is by no means straightforward, neither the effects of different aggregation techniques nor the impacts of missing or flagged raw data have to date been rigorously assessed. Most of the studies based on monthly or yearly mean radiation fluxes ignore potential uncertainties induced by data gaps or differing aggregation methods. The problem has become more obvious with the unsatisfactory situation that for the same site substantially differing monthly or yearly mean have been published. The present study attempts to shed more light onto this issue, using exemplarily data from the Baseline Surface Radiation Network (BSRN; Ohmura et al., 1998). BSRN was established to provide high quality radiation measurements aimed at monitoring and detecting important changes in the surface radiation balance. In 2005, BSRN provides radiation data at almost 40 sites at high temporal frequency (time intervals of 1, 2, 3 or 5 min depending on site and period) and highest possible accuracy. The BSRN data have been successfully used in numerous scientific applications (e.g., Wild et al., 1995, 2005; Dutton et al., 2006; Wild, 2008, 2009).

This paper focuses on monthly means as this aggregation is widely used in numerous climatological analyses. The reasons for this are manifold: given such factors as large differences in scale and sampling frequency between "point measurements" such as BSRN surface radiation and both satellite retrievals and model calculations, one common practice is to use longer averages in any comparisons between the two. One of the common temporal averaging modes is to use monthly averages, which cover enough time that spatial and temporal sampling differences are mitigated to significant extent, yet are still "short enough" to be able to investigate such phenomenon as seasonal cycles. The same holds true for mitigating the effect of "missing or bad" data. For instance, an hour of missing solar radiation measurements near local solar noon precludes a meaningful daily average for that day. Without a priori knowledge of cloud occurrence and cloud properties for the missing time period, it is very challenging to accurately "manufacture" values corresponding to the missing data. Yet by the method of creating a monthly average diurnal cycle, the climatology of cloud occurrence for a given site helps to mitigate the influence of the missing data.

In the following we investigate the impact of missing BSRN radiation observations (either non-existent or flagged)

and estimate the error when applying a number of different methods for the computation of monthly means from 1-min observations.

2 Data and methods

2.1 BSRN

The Baseline Surface Radiation Network (BSRN) is a project of the World Climate Research Program (WCRP) (Ohmura et al., 1998) and aims at providing the climate community with accurate and highly resolved irradiances for climate research purposes. This global network measures surface radiative fluxes at the highest possible accuracy with well-calibrated state-of-the-art instrumentation at selected sites in the major climate zones. Data are available from 1992 onward, currently from 51 stations, covering a latitude range from 80° N to 90° S. The high temporal resolution (minute frequency) makes the database a valuable tool for the validation of radiation schemes as well as the evaluation of estimates of surface radiation based in part on necessarily indirect and imperfectly calibrated satellite observations. For detailed information on the BSRN database and the sites (and the 3-letter acronyms for the stations that are used in this study), the reader is referred to the website at http://www.bsrn.awi.de/. The BSRN database currently contains approximately 5800 station months. The results presented in this study are based on all available observations that were available by Spring 2008. The study concentrates on the "basic" measurements, including global radiation (GLOB), diffuse shortwave radiation (SWDIFF), direct shortwave radiation (SWDIR), and downwelling longwave (LWDOWN). GLOB can be measured by either an unshaded pyranometer or by adding the direct and diffuse shortwave components. If not specified, GLOB refers to the pyranometer measurement. The term GLOB1 will be used for the (measured) sum of the downwelling direct and diffuse shortwave flux. The time interval for the radiation data compilation is mostly 1 min. A few sites provide data every 3 or 5 min.

2.2 Data flagging procedures

Detailed quality checks are applied to the BSRN radiation data. The WRMC does not correct the data but flags radiation data that is suspected to be erronous. Then, subsequent applications of the data can determine if the flagged data should be used or discarded. Note that AWI does not provide quality flags of the archived BSRN data.

Three different procedures have been applied to the data. The procedures and limits are identical for all BSRN sites.

i. The "physically possible" procedure aims at detecting extremely large errors in the radiation data. The radiation data falling in the intervals defined in Table 1 are considered "physically possible".

Table 1. This table shows the lower and up limits for the "Physically possible" intervals used in flagging the radiation quantities. Values were flagged if outside the indicated interval. S_o is the solar constant adjusted for Earth-Sun distance. μ is the cosine of the solar zenith angle. Parameters: GLOB: Global radiation, SWDIFF: Diffuse shortwave radiation, SWDIR: Direct diffuse radiation, SWUP: Reflected shortwave radiation, LWDOWN: Downwelling longwave radiation, LWUP: Upwelling longwave radiation.

Parameter	Lower bound	Upper bound
GLOB	$-4\,\mathrm{W\,m^{-2}}$	$1.5\,S_o\,\mu^{1.2} + 100\,\mathrm{W\,m^{-2}}$
SWDIFF	$-4\,\mathrm{W\,m^{-2}}$	$0.95\,S_o\,\mu^{1.2} + 100\,\mathrm{W\,m^{-2}}$
SWDIR	$-4\,\mathrm{W\,m^{-2}}$	$S_o\,\mu^{1.2}$
SWUP	$-4\,\mathrm{W\,m^{-2}}$	$1.2\,S_o\,\mu^{1.2} + 50\,\mathrm{W\,m^{-2}}$
LWDOWN	$40\,\mathrm{W\,m^{-2}}$	$700\,\mathrm{W\,m^{-2}}$
LWUP	$40\,\mathrm{W\,m^{-2}}$	$900\,\mathrm{W\,m^{-2}}$

Table 2. Same as Table 1 except for the "Extremely rare" intervals for flagging the radiation quantities.

Parameter	Lower bound	Upper bound
GLOB	$-2\,\mathrm{W\,m^{-2}}$	$1.2\,S_o\,\mu^{1.2} + 50\,\mathrm{W\,m^{-2}}$
SWDIFF	$-2\,\mathrm{W\,m^{-2}}$	$0.75\,S_o\,\mu^{1.2} + 30\,\mathrm{W\,m^{-2}}$
SWDIR	$-2\,\mathrm{W\,m^{-2}}$	$0.95\,S_o\,\mu^{1.2} + 10\,\mathrm{W\,m^{-2}}$
SWUP	$-4\,\mathrm{W\,m^{-2}}$	$S_o\,\mu^{1.2} + 50\,\mathrm{W\,m^{-2}}$
LWDOWN	$60\,\mathrm{W\,m^{-2}}$	$500\,\mathrm{W\,m^{-2}}$
LWUP	$60\,\mathrm{W\,m^{-2}}$	$700\,\mathrm{W\,m^{-2}}$

ii. The limits in the "extremely rare" procedure are narrower than those of the "physically possible" test. Radiation data which violate these limits may occur over very short time periods under very rare conditions. These limits are given in Table 2. Within this study, data of "good quality" are assumed to be inside the "extremely rare" limits.

iii. The "across quantities" procedures capture smaller errors that have not been detected by the previous quality checks. These tests are based on empirical relations of the different quantities measured. The restrictions are defined in Table 3.

2.3 Methods for monthly mean computations

There are many options for the computation of monthly averages from incomplete data. We will test the performance of several methods that are currently applied in the climate and radiation community to the BSRN data. Most of them include to some extent arbitrary thresholds that were set based on expert knowledge, visual inspection of appropriate illustrations and practical reasoning. These type of methods have been used in many practical applications (e.g. Zhang et al., 2004; Dutton et al., 2006; Wild et al., 2006; and Hinkelman et al., 2009). Seven different such methods are selected here to demonstrate the effect that different methods can have. We applied the following seven algorithms for computing monthly means from n-min values ($n = 1$, 3, 5) from the BSRN data:

M1: Computation of monthly fluxes from all minute-values, including all flagged data, which have been identified in the BSRN data base as being questionable (see Sect. 2.2). No filling of missing data is applied.

M2: Computation of monthly means but excluding all observations that are outside the most lenient quality flag identified as "extremely rare" limits (see Sect. 2.2). No filling of removed flagged data or originally missing measurements is applied.

M3: Missing data and data flagged in M2 were filled from an interpolated zenith angle-dependent fill dataset that was built from available observations from the same two-week interval in which the missing data are located. The fill data therefore account for both the diurnal (zenith angle) and seasonal variation.

M4: As M3 except using periods of one month were used to build the fill data instead of two weeks.

M5: The strictest possible quality control was applied where all flags must pass the procedure in order for the associated data to be used. After applying the quality control procedure the monthly means were computed according to the following four steps:

i. Computing *15-min* average. At least one valid original value is required for a 15-min average to be computed.

ii. Computing *hourly* means from 15-min values. All four 15-min values must be present for an hourly mean to be computed.

iii. Computing the monthly mean *diurnal cycle* based on the hourly means. A diurnal cycle consists of 24 values, each one being the average of the hourly means corresponding to the same hour in all the days in the month. For each of the hourly values in the diurnal cycle to be computed, at least 95% of the expected hourly means must be present. In other words, at least 95% of the days in the month must have hourly means as computed from last step for the concerned hour.

iv. Computing *monthly* means from the diurnal cycle. It is required that all the 24 hourly values must be present in order for a monthly mean to be computed.

M6: Monthly averages consist of a simple arithmetic mean of available unflagged or "good" data, with a daily average only being calculated for days where the available data is at

Table 3. Same as Table 1 except for "across quantity" intervals used for flagging the radiation quantities. SZA: solar zenith angle, σ = Stephan-Boltzman constant (= 5.67×10^{-8} W m^{-2} K^{-4}), T_a: air temperature [K]. GLOB1 = SWDIR + SWDIFF. The first column contains the abbreviations as used within this study.

Abbrev.	Short name of test	Condition for test
COMP1	GLOB/GLOB1	GLOB/GLOB1 = $1.0 \pm 8\%$ for GLOB1 > 50 W m^{-2}, SZA < 75°
		GLOB/GLOB1 = $1.0 \pm 15\%$ for GLOB1 > 50 W^{-2}, 75° < SZA < 93°
COMP2	SWDIF/GLOB	SWDIF/GLOB < 1.05 for GLOB > 50 W m^{-2}, SZA < 75°
		SWDIF/GLOB < 1.10 for GLOB > 50 W m^{-2}, 75° < SZA < 93°
		no test possible for GLOB <= 50 W m^{-2}
COMP3	SWUP comparison	SWUP/GLOB1 < GLOB1 for GLOB1 > 50 W m^{-2}
		no test possible for GLOB1 <= 50 W m^{-2}
COMP4	LWDOWN to T_a comparison	$0.4 \cdot \sigma T_a^4$ < LWDOWN < $\sigma T_a^4 + 25$
COMP5	LWUP to T_a comparison	$\sigma (T_a - 15)^4$ < LWUP < $\sigma (T_a - 15)^4$
COMP6	LWDOWN to LWUP comparison	LWDOWN < LWUP + 25 W m^{-2}
		LWDOWN > LWUP − 300 W m^{-2}

least 50% of the possible data for that value. For the solar-driven variables (such as SWDIFF and SWDIR), the number of "good" data must be at least half of the daylight period, the daylight period defined as the number of minutes from sunrise to sunset on a given day for that date and location. Once daily averages have been produced per the above procedure, the daily averages are then used to calculate monthly averages as a simple arithmetic mean of the available daily averages if certain limits on available data are met. First, for any given day to be considered for being included in the monthly average there must be at least 1300 min of the possible 1440 min overall data available, regardless of the availability of any particular individual variable. Then for any particular variable, there must be at least 60% of the possible data available, i.e. for the downwelling LW there must be 864 min of available "good" data, for the downwelling SW there must be 60% of the possible daylight (sunrise to sunset) data available.

M7: 15-min averages are first computed from the 1-min data for each month. Computation of a single bin requires at least 20% valid data. Minute values that are outside the "physically possible" limits (Sect. 2.2, Table 1) are treated as missing values. For shortwave radiation fluxes, values below 0 W m^{-2} during night (solar zenith angle > 93°) were set to 0 W m^{-2}. The reason for negative shortwave fluxes ("night-time offset") has been discussed in (Haeffelin et al., 2001). The monthly mean is then computed by averaging the 96 bins (96 × 15 min = 24 h) that have been produced for each month. The monthly mean is valid only if all bins contain valid values. Performing the computation of the monthly mean diurnal cycle benefits from the typical diurnal cycle of shortwave fluxes, allowing more accurate estimates for incomplete observations.

The methods M1, M2, M3, and M4 set nighttime (solar zenith angle greater than 93°) SW values to zero. Note that SWDIR is computed on a horizontal plane for the two methods M5 and M6 while the other methods provide SWDIR on a surface perpendicular to the direction of the incoming beam.

It is important to note that we do not recommend a method for filling in the gaps as there is no "best method" to fill them in. In fact, determination of the "best method" actually depends on what the resultant data are to be used for. Without a priori knowledge of variables that affect the surface radiation (i.e. cloud occurrence and cloud properties, atmospheric state, aerosol and ozone loading, etc.) for the missing time period, it is impossible to accurately "manufacture" values corresponding to the missing data. If one depends on climatology, then the gap filling interferes with the ability to analyze the data for long term subtle trends such as global dimming and brightening or global warming. There simply is no "win-win" methodology to remove the effects of missing data. That being the case, our methods have the advantages of being simple and easily understood, do not include modeled or external data, but relies only on actual measurements, and the methodology helps to mitigate the influence of missing data. The user of the monthly averages thus produced must be aware of the impacts of missing data and make their own judgment as to how much missing data is allowed. The high time resolution data are available for those who do prefer some other gap-filling methodology.

3 Results and discussion

The completeness of the BSRN observation are assessed by presenting (i) an overview on the frequency of gaps in the data (Sect. 3.1), (ii) the amount of flagged data (Sect. 3.2) and (iii) the impact of missing and/or flagged data on monthly mean estimates (Sects. 3.3 and 3.4).
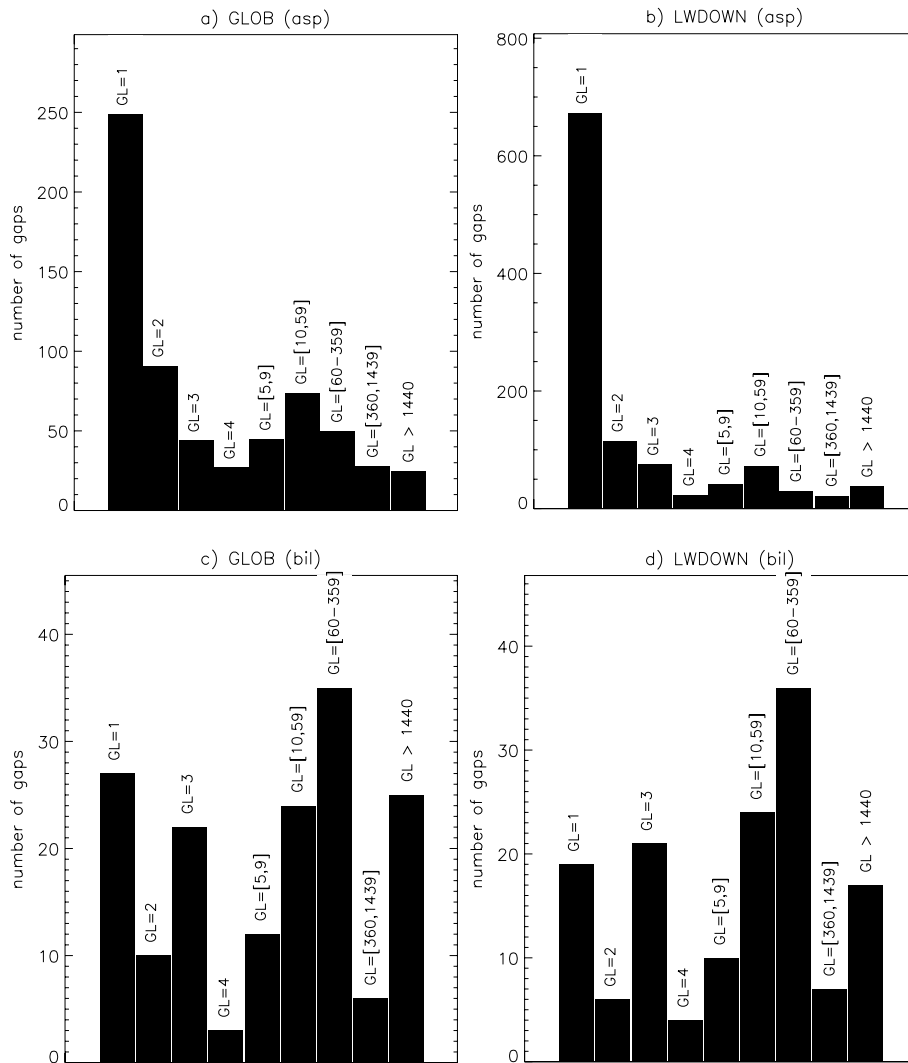
Fig. 1. Examples of the distribution of data gap length (GL) for Alice Springs, Australia – ASP, panels **(a)**, **(b)** – and Billings, USA (BIL, panels c, d) for GLOB and LWDOWN. Gap lengths (GL) are given in minutes on each bar. The length of observation period in months is 131 and 149 for ASP and BIL, respectively.

3.1 Data gaps

Data gaps in the initial field data occur due to different reasons such as calibration periods, instrument failure or data loss. For this study, the frequency of data gaps was investigated for both shortwave (SW) and longwave (LW) radiation fluxes at all BSRN sites using all currently (spring 2008) available 1-min observations. Figure 1 displays the gap distribution for GLOB and LWDOWN at the two BSRN sites Alice Springs, Australia (ASP) with 131 observed months and Billings, USA (BIL) with 149 months of observations. The figure clearly shows that both the gap lengths and gap frequency between different sites and different parameters may strongly vary. While for ASP, data gaps are generally very short, the radiation instruments at BIL often fail for

more than one day (1440 min). Table 4 gives the percentage of missing data along with the total number of gaps. From this table we learn that ten BSRN sites have more than 5% missing GLOB observations. For SWDIR, 16 (8) sites have more than 5% (15%) missing data. A considerable fraction of LWDOWN data is missing: 11 (4) sites have more than 5% (15%) missing observations for LWDOWN. It can be thus concluded that at many sites, a substantial percentage of the observations are missing. The detailed gap analysis (Table 4) shows that for a specific site, the percentage of missing GLOB is generally lower than that for SWDIR. Only 6 (3, 7) sites out of the 33 stations listed in Table 4 have less than 1% missing data for GLOB (SWDIR, LWDOWN). These findings confirm that the BSRN community should aim at reducing the occurence of frequent data gaps. High fractions

Table 4. Percentage of missing data and number of gaps (in brackets) for all BSRN sites. Note that missing data do not include data flagged "unphysical". For detailed information on the BSRN stations whose 3-letter acronyms are given here, see http://www.bsrn.awi.de/en/home/bsrn/. Parameters: GLOB: Global radiation, SWDIR: direct shortwave radiation, LWDOWN: downwelling longwave radiation.

station	# of obs. months	GLOB (# of gaps)	SWDIR (# of gaps)	LWDOWN (# of gaps)
ASP	131	3.8% (627)	12.3% (838)	9.1% (1081)
BAR	162	6.7% (1907)	14.2% (3119)	7.4% (1289)
BER	162	5.5% (1196)	24.7% (741)	7.6% (293)
BIL	149	28.1% (140)	25.2% (246)	29.3% (113)
BON	104	1.7% (2382)	8.9% (2310)	1.8% (2159)
BOS	97	0.6% (1804)	2.3% (2361)	0.5% (1791)
BOU	162	2.6% (1983)	4.1% (752)	1.6% (504)
CAM	70	0.13% (2)	15.2% (24)	1.0% (199)
CAR	64	0.0% (0)	0% (0)	0.0% (0)
CLH	71	9.0% (163)	13.8% (163)	2.0% (267)
DAA	55	1.1% (896)	0.9% (73)	3.1% (70)
DAR	41	0.8% (501)	1.9% (722)	0.8% (430)
DRA	66	2.5% (1733)	4.5% (5393)	0.5% (564)
E13	99	2.1% (198)	2.1% (211)	1.4% (37)
FLO	138	15.6% (142)	19.3% (1677)	19.8% (12)
FPE	104	3.9% (1572)	9.5% (6495)	4.9% (2093)
GCR	105	0.8% (1501)	6.1% (7613)	2.4% (2081)
GVN	177	1.1% (1185)	5.0% (1671)	0.7% (606)
ILO	68	7.2% (1)	82.4% (4713)	–
KWA	159	4.9% (503)	15.5% (3818)	2.3% (2239)
LAU	77	2.3% (104)	2.5% (881)	9.8% (369)
LER	70	1.0% (6)	4.7% (26)	0.7% (28)
LIN	104	0.7% (89)	2.2% (3064)	4.4% (293)
MAN	110	1.8% (1414)	5.5% (8590)	5.3% (7254)
NAU	85	2.5% (904)	4.5% (2853)	4.2% (1320)
NYA	173	1.0% (281)	34.8% (1795)	1.0% (453)
PAL	28	-	13.0% (13)	0.0% (0)
PAY	166	0.5% (99)	59.3% (4519)	1.1% (230)
PSU	63	4.0% (612)	3.9% (5027)	2.5% (1529)
REG	108	2.8% (13127)	3.0% (13484)	2.6% (4319)
SBO	42	11.7% (175)	13.1% (252)	15.2% (326)
SOV	52	0.6% (15)	0.1% (82)	0.4% (23)
SPO	162	6.6% (1319)	6.7% (1853)	9.7% (1038)
SYO	120	1.4% (443)	3.6% (5576)	4.6% (1950)
TAM	82	0.7% (49)	1.5% (281)	0.8% (19)
TAT	139	6.9% (57)	0.5% (108)	5.7% (629)
TOR	97	5.5% (38)	1.1% (13)	54.7% (3)

of missing data hinder deriving reliable monthly means or trends in the radiation fluxes.

3.2 Flagged BSRN data

The overall quality of measured time series does not only depend on the frequency of gaps but also on the amount of flagged data. BSRN has established a simple quality control of measured radiation fluxes (see Sect. 2.2). In Table 5 we present the fraction of the flagged data according to the "extremely rare" procedure (in units of 0.1%). The high fractions of flagged data in the SW are primarily due to the

flagging of small negative SW fluxes during night ("nighttime offset", see Haeffelin et al., 2001) which are related to a small level of thermal noise. We can therefore conclude that in a first approximation, numbers > 100 (10%) in Table 5 represent the fraction of data below $-2\,\mathrm{W\,m^{-2}}$. Approximately half of all BSRN sites (17) belong to this category. The fraction of flagged SWDIFF and SWDIR data is generally lower than the flagged GLOB data. Note that ignoring negative nighttime offsets might have a serious impact on the monthly mean. Note that there is still no common sense within the BSRN community how to handle negative nighttime offsets. Flagged fractions ranging from 0 to a

Table 5. Percentage ($\times 10$) of values flagged according to the "extremely rare" procedure as described in Sect. 2.2 and Table 1. Test can not be applied (due to missing data): -999. Numbers marked with * are primarily caused by flagged negative SW fluxes.

station	GLOB	SWDIFF	SWDIR	SWUP	LWDOWN	LWUP
ASP	0.15	0.60	0.21	-999	0.19	-999
BAR	0.91	2.84	2.53	0.06	0.00	0.00
BER	0.66	0.24	0.16	-999	0.00	-999
BIL	390.08*	191.42*	34.76*	-999	0.45	-999
BON	480.80*	246.99*	29.95*	-999	0.00	-999
BOS	491.46*	370.22*	85.64*	-999	0.00	-999
BOU	0.53	0.49	1.13	-999	0.00	-999
CAM	0.08	0.55	0.28	-999	0.00	-999
CAR	0.01	0.33	0.16	-999	0.00	-999
CLH	25.04	24.53	15.24	-999	0.00	-999
DAA	0.50	1.79	1.72	-999	4.53	-999
DAR	387.22*	1.14	27.95	-999	0.06	-999
DRA	418.91*	264.35*	76.41*	-999	0.00	-999
E13	489.63*	86.35*	49.67*	-999	103.46	-999
FLO	12.35	6.09	42.34	-999	38.96	-999
FPE	466.37*	264.23*	54.26*	-999	0.00	-999
GCR	499.04*	260.54*	46.73*	-999	0.01	-999
GVN	21.14	34.22	44.69	8.42	0.00	0.00
ILO	0.28	0.11	0.98	-999	0.00	-999
KWA	0.37	0.21	1.13	-999	0.00	-999
LAU	0.48	0.26	0.00	-999	0.01	-999
LER	0.41	1.55	0.06	-999	0.00	-999
LIN	26.80	35.89	4.36	-999	0.01	-999
MAN	306.34*	191.88*	49.01*	-999	0.01	-999
NAU	337.47*	99.24*	38.00*	-999	0.01	-999
NYA	22.00	23.86	43.54	12.11	0.00	0.00
PAL	0.00	98.50	0.19	-999	0.00	-999
PAY	206.80*	210.61*	38.57*	110.49 *	0.00	0.00
PSU	478.79*	290.23*	41.39*	-999	1.21	-999
REG	4.76	15.06	8.48	-999	0.15	-999
SBO	430.78*	297.61*	34.72*	-999	5.47	-999
SOV	467.11*	467.40*	11.07*	-999	0.16	-999
SPO	2.24	3.43	2.56	0.39	2.42	0.00
SYO	13.46	15.35	2.31	30.95	0.00	0.00
TAM	483.11*	404.36 *	279.39*	-999	0.00	-999
TAT	0.29	0.97	7.42	0.19	0.00	0.00
TOR	2.39	0.28	0.14	0.05	0.00	0.00

Table 6. Percentage ($\times 10$) of values flagged according to the "across quantity" procedures COMP1, COMP2, COMP3, COMP4, COMP5, and COMP6 as described in Sect. 2.2 and Table 3. Values equal to -999 indicate that the "across-quantity" procedure could not be applied (due to missing data).

station	COMP1	COMP2	COMP3	COMP4	COMP5	COMP6
asp	5.70	0.09	-999	0.00	-999	-999
bar	13.95	3.29	0.50	0.96	0.52	0.29
ber	13.67	3.78	-999	0.03	-999	-999
bil	53.23	21.90	-999	0.00	-999	-999
bon	39.20	5.30	-999	0.00	-999	-999
bos	8.53	1.59	-999	0.00	-999	-999
bou	38.90	7.81	-999	0.00	-999	-999
cam	2.69	0.17	-999	0.00	-999	-999
car	11.56	0.01	-999	0.00	-999	-999
clh	7.62	1.31	-999	0.00	-999	-999
daa	33.36	0.18	-999	1.22	-999	-999
dar	9.23	4.23	-999	0.08	-999	-999
dra	23.98	1.75	-999	0.00	-999	-999
e13	52.50	18.69	-999	125.77	-999	-999
flo	150.24	0.46	-999	73.29	-999	-999
fpe	30.31	6.54	-999	0.00	-999	-999
gcr	115.94	9.03	-999	0.00	-999	-999
gvn	50.85	0.03	1.71	1.04	0.00	0.80
ilo	2.49	0.00	-999	0.08	-999	-999
kwa	20.86	7.99	-999	0.26	-999	-999
lau	18.98	2.51	-999	0.00	-999	-999
ler	4.92	0.58	-999	0.00	-999	-999
lin	4.17	0.02	-999	0.16	-999	-999
man	11.80	4.24	-999	0.05	-999	-999
nau	22.35	6.16	-999	0.08	-999	-999
nya	31.26	0.02	0.16	0.00	0.08	0.34
pal	0.00	0.00	-999	0.00	-999	-999
pay	14.88	0.02	0.00	0.01	0.08	0.03
psu	12.41	4.27	-999	0.00	-999	-999
reg	14.06	1.06	-999	0.76	-999	-999
sbo	66.59	11.63	-999	0.09	-999	-999
sov	5.10	0.03	-999	0.01	-999	-999
spo	26.62	0.56	3.60	4.87	0.03	4.27
syo	8.78	0.20	12.94	0.00	0.00	0.11
tam	9.42	0.41	-999	0.00	-999	-999
tat	11.19	0.15	0.00	0.00	0.30	0.10
tor	13.59	8.04	0.18	0.00	0.00	0.00
xia	294.21	44.63	-999	0.00	-999	-999

few percent may be attributed to "real" data problems (other than the nighttime offset) due to instrument failure or calibration problems. Sites that do not provide SWUP or LWUP are marked with -999 in Table 5, clearly pointing out that less than 25% of the BSRN sites observe SWUP and LWUP. In the LW, measurements outside the "extremely rare" limits rarely occur. Only at two stations (E13 and FLO), more than 1% of the LWDOWN observations are flagged while 33 out of the 39 listed sites have less than 0.2% flagged LWDOWN observations.

Table 6 provides detailed insight into the flagging results obtained from the "across quantities" procedure (Sect. 2.2). BSRN data mostly meet the rules of the "across-quantities" procedure. The constraint between GLOB and GLOB1 – its difference should stay below 8% – is, however, quite often violated. For 14 BSRN sites, this condition is not satisfied for more than 2% of all 1- (2-, 5-) min observations. The mean over all sites (weighted with the length of the measurement period) is 2.8%. This is related to technical problems and

tracker failure in continuous measurements of direct SW radiation using standard pyrheliometers. For the other "across quantities" procedures the mean hit rate is generally below 1%.

Note that flagging procedures test for the total error only. The quality control does not distinguish between systematic and random errors. The total error consists of the sum of the random error and the systematic error (bias). Systematic errors occur, e.g., from nonrepresentative sampling, tracking problems or physical obstacles such as trees and houses. The problem of the "negative nighttime offset" may be a serious hint for biased (shortwave) radiation data. Periodic calibrations of radiation instruments generally reduces the systematic error. In contrast to the systematic error, random errors originate from random processes in the measurement device

or the sensor, influencing the precision of a measurement. They do not have an impact on the mean values.

3.3 Differences in monthly means

In this section the seven different methods presented in Sect. 2.3 will be compared. For this intercomparison, we use all available data from the BSRN archive in order to provide the best statistics possible. Figure 2 shows the deviation of monthly GLOB climatologies for each method from the average over all months. This figure clearly reveals that the differences among the seven investigated methods can be quite large. Typical differences are in the order of $1\,\mathrm{W\,m^{-2}}$ but may increase to a few $\mathrm{W\,m^{-2}}$ for some sites. It is evident that the differences become more pronounced for individual months (note that the results in Fig. 2 show climatologies). Figure 2 reveals that the handling of flagged data plays an important role. This can be clearly demonstrated by comparing M1 with M2. These two methods only differ in how the flagged data are handled: M1 includes all data outside the "extreme rare" limits while M2 excludes them. This indicates that the treatment of flagged GLOB observations during the day (night values are zeroed) may also have a pronounced effect on the computed monthly mean. Distinct differences are also found between M2 and M3, giving strong evidence that gap filling has a distinct effect on the computed monthly means. The M7 and M4 methods hardly differ for KWA and PAY (Fig. 2, right-hand panels). This suggests that the computation of monthly means by computing first the monthly mean diurnal cycles (as applied in M7) may help to avoid the use of gap-filling (as applied in M4), even for time series with a considerable amount of missing data (as for KWA, see Table 4).

The mean absolute deviation between two methods gives further insight into the differences between individual methods. The mean absolute deviation (MAD) is defined as follows

$$\mathrm{MAD} = \frac{1}{N} \sum_{i=1}^{N} \left(\overline{\mathrm{M}X_i - \mathrm{M}Y_i} \right),\ X,\ Y = 1,\ 2,\ ...\ 7 \quad (1)$$

with $\mathrm{M}X_i$ and $\mathrm{M}Y_i$ the monthly means computed with method $\mathrm{M}X$ and $\mathrm{M}Y$, respectively, and N the number of valid monthly means in both $\mathrm{M}X$ and $\mathrm{M}Y$. Figure 3 displays MAD between all method combinations, averaged over all BSRN sites listed in Table 4. For GLOB, MAD generally amounts to 1–$3\,\mathrm{W\,m^{-2}}$. The comparison between method M1–M4 shows again that the treatment of flagged data and the gap-filling do have an effect on the computed monthly GLOB values. The mean GLOB biases between M6 and the other six methods is substantially higher than among the other methods. M6 applies a more stringent testing for "extreme rare" limits than the official BSRN screening and computes monthly means from the arithmetic mean of daily averages (Sect. 2.3). This pronounced bias indicates that a more
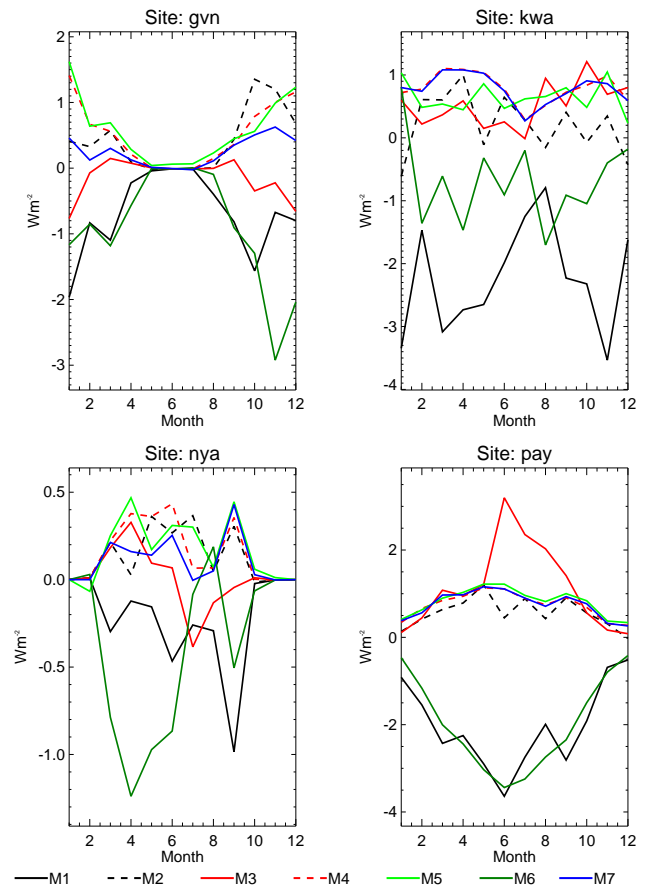

Fig. 2. Comparison of different algorithms for the computation of monthly GLOB means (see Sect. 2.3). Shown are the differences between each single method and the sum of all methods. The analysis is restricted to the period during which all methods provide valid monthly means. The following 4 sites are displayed: GVN (Georg von Neumayer, Antarctica), KWA (Kwajalein, Marshall Islands), NYA (Ny Alesund, Spitsbergen), and PAY (Payerne, Switzerland).

sophisticated quality control might also have a distinct effect on the computed monthly mean. For SWDIFF, MAD is generally smaller than for GLOB since SWDIFF is generally smaller than GLOB. The direct beam component of GLOB, however, clearly shows larger differences between the monthly means computed by different methods (Fig. 3). This is probably due to the technically more difficult measurement of SWDIR compared to GLOB and SWDIFF as sun tracking by the pyrheliometer is quite susceptible to errors. This is also reflected in the high percentage of missing values at many BSRN sites as shown in Table 4. This failure rate is a likely reason for the considerable biases between M1–M4 and M7: the former compute the monthly means by a simple arithmetic average of daily means while the latter computes the monthly mean from the monthly mean diurnal cycle. From Fig. 3, we learn that the monthly LW-DOWN estimates obtained with different algorithm are in
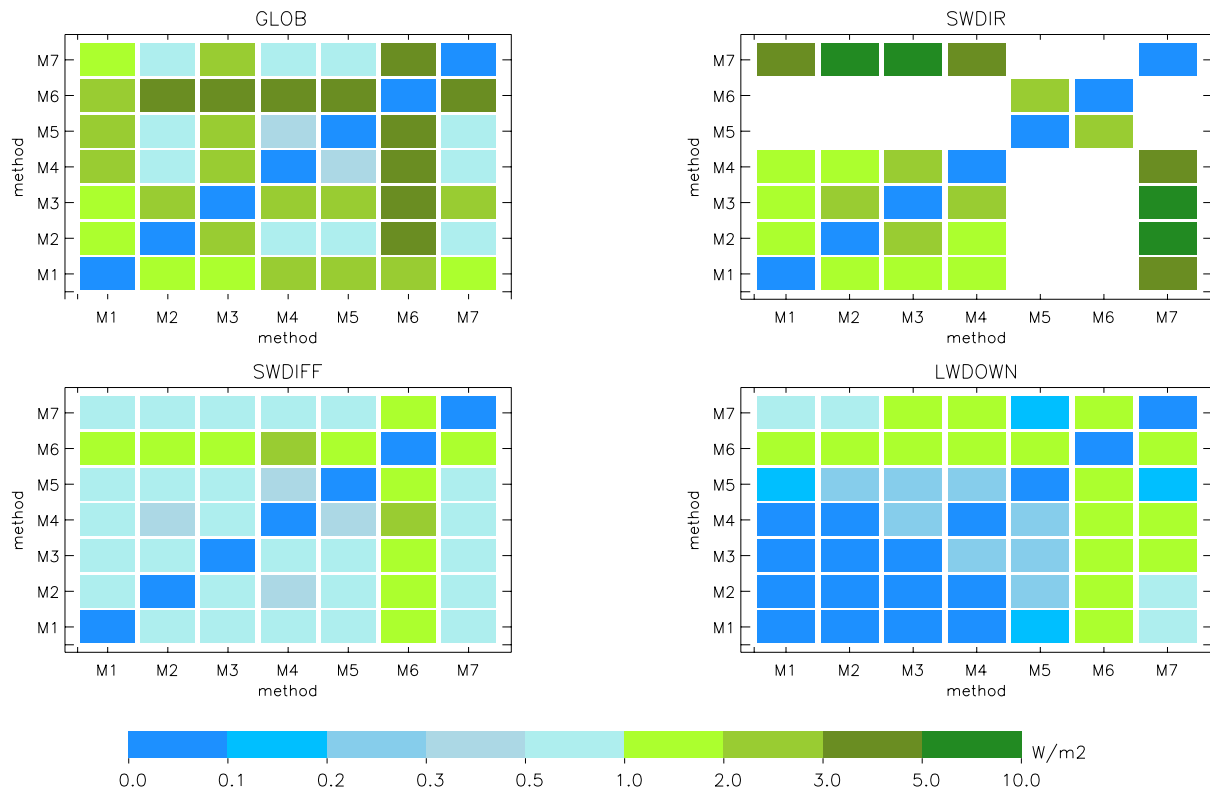
Fig. 3. Mean absolute bias between all pairs of filling methods for four radiative quantities, GLOB (Global Radiation), SWDIR (direct SW radiation), SWDIFF (diffuse SW radiation), and LWDOWN (downward longwave radiation). Filling methods M1–M7 are described in Sect. 2.3. The absolute biases are averaged over all BSRN sites. Note: M5 provides data for nine BSRN sites only as specified in Sect. 2.3. Monthly SWDIR computed with M5 and M6 can not be compared to the other methods as they provide SWDIR on a horizontal surface.

close agreement. MAD is below $0.1 \, \mathrm{W \, m^{-2}}$ between methods M1, M2, and M3. These low differences are closely related to the low percentage of flagged data (see Table 5) and the rather low percentage of missing data when compared to the SW fluxes. Furthermore, the low temporal variability might further reduce the effect of data gaps.

Figures 4 and 5 display histograms of the monthly GLOB biases between two individual methods. The height of the bars gives the fraction of all concurrently valid monthly means within a certain range as provided above the bars. Ideally, all biases are within $\pm 0.1 \, \mathrm{W \, m^{-2}}$ which would imply one single grey bar for the range $[-0.1, 0.1] \, \mathrm{W \, m^{-2}}$ with a height of 100%. From Fig. 4 we learn that M4 and M7 do approach this ideal case most closely. This gives some evidence that both a clever interpolation of missing/flagged data (M4) and the computation of monthly means from the monthly mean diurnal cycle are likely to be useful and robust approaches for the computation of monthly radiation fluxes from high temporal observations. It is of some interest, however, that the biases between M4/M5 and M5/M7 follow well a gaussian distribution while the difference M4–M7 is positively skewed (Fig. 4, middle row, left panel). The distribution of the biases differs strongly when evaluating differences

between method M6 and any other method (Fig. 4, right-hand panels). The absolute bias between M6 and MX, $X = 1$, 2, 3, 4, 5, 7 is greater than $2 \, \mathrm{W \, m^{-2}}$ for approximately one third of all monthly means. This might be related to the quality control and interpolation method applied in M6 that clearly differs from the other algorithms under investigation. Further, M6 requires further input parameters that are dependent on the site. Therefore, M6 was only applied to the data of 9 BSRN sites.

Figure 5 is similar as Fig. 4 but for M1–M4. This figure reveals that it is relevant to consider the effect of flagged data and/or data gaps on the computed monthly mean. Monthly means obtained with M1 differ by more than $2 \, \mathrm{W \, m^{-2}}$ from M2, M3, and M4 in approximately 15% of all cases. In addition the distribution is far from being gaussian but rather negatively skewed. This suggests that the consideration of missing/flagged value is essential. The differences between M1 and M2 reveal that the handling of flagged data does also have an impact on the monthly mean estimates. The intra-method biases among the other methods are distinctly less pronounced. The fractions of monthly mean biases above $2 \, \mathrm{W \, m^{-2}}$ are less than 10% and 2% for the pairs M2/M3 and M2/M4, respectively. It is striking that over 70% of the
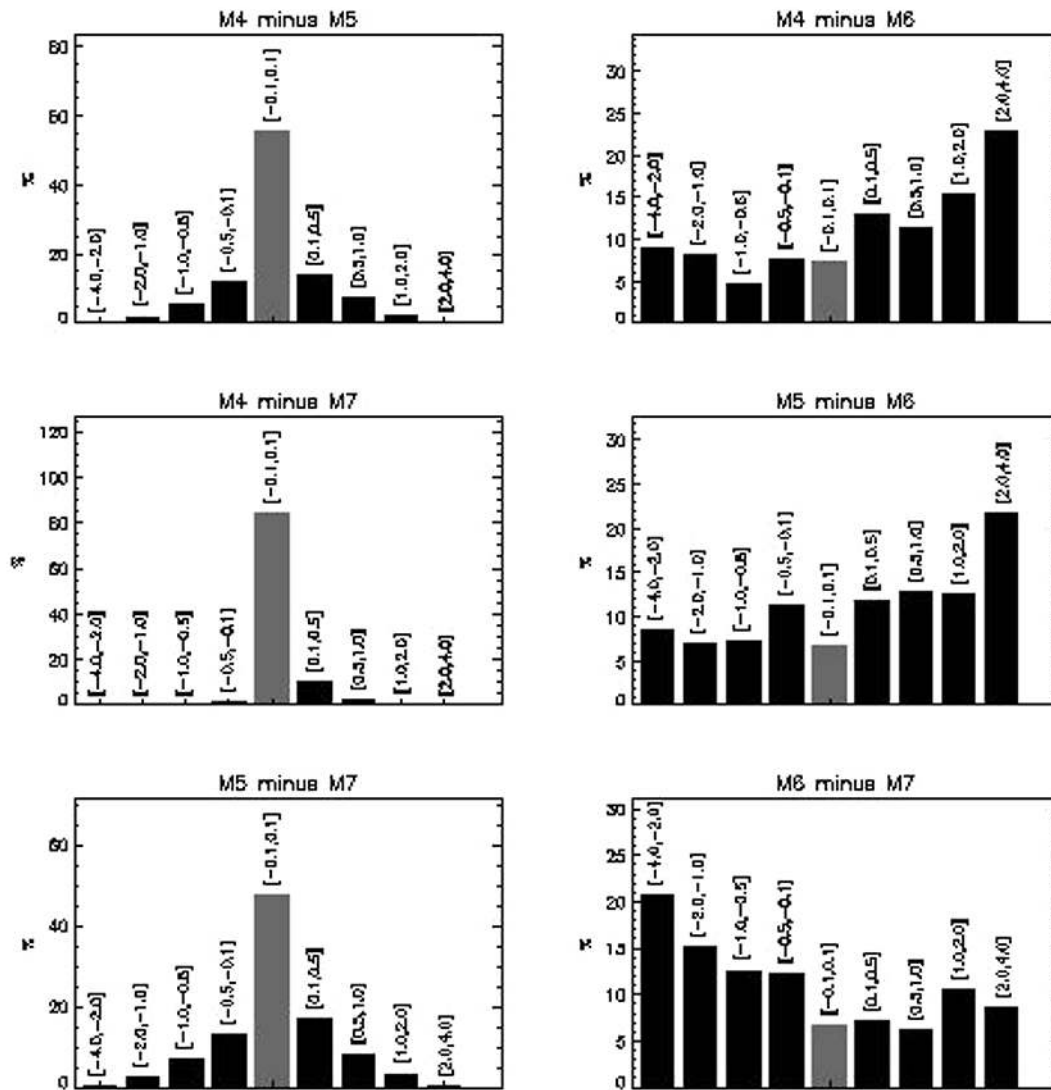
Fig. 4. Differences in monthly means of GLOB from pairs of M4, M5, M6, and M7 filling methods as indicated above each plot. The various methods are described in Sect. 2.3. The bars show the percentage of monthly mean differences within the limits given in squared brackets above the bars (unit: W m^{-2}). The comparison considers all monthly BSRN data where the individual methods concurrently provide valid monthly GLOB.

monthly means obtained with M3 and M4 do not differ by more than 0.1 W m^{-2}. From this it is evident that the impact on the monthly bias is likely less dependent on the interpolation method but rather if flagged/missing will be replaced by interpolated values.

In the following, we consider all monthly means derived from two methods that differ by less than 0.2 W m^{-2} (hereinafter called "high agreement" or HIAG). The question is how the fraction of HIAG depends on the percentage of "good" observation. We address this question in Fig. 6 by binning monthly means into classes with different fractions of underlying "good" observations, which are (defined as 1-, 3- or 5-min fluxes that are within the "extreme rare"

limits). This means, e.g., that the class 99–100% contains monthly means that are based on less than approximately 430 ($30 \times 24 \times 60/100 = 432$) 1-min measurements that are either missing or flagged (assuming a month with 30 days). Note that observation afflicted with nighttime offsets are not counted as flagged since all methods zero SW fluxes during night.

Figure 6 displays the inter-method differences for the four algorithms M4, M5, M6, and M7. We again included all data that are currently stored in the BSRN database, allowing for a very large basic set. It is evident that the fraction of HIAG increases for increasingly complete and un-flagged observed data. For a month with complete observation and
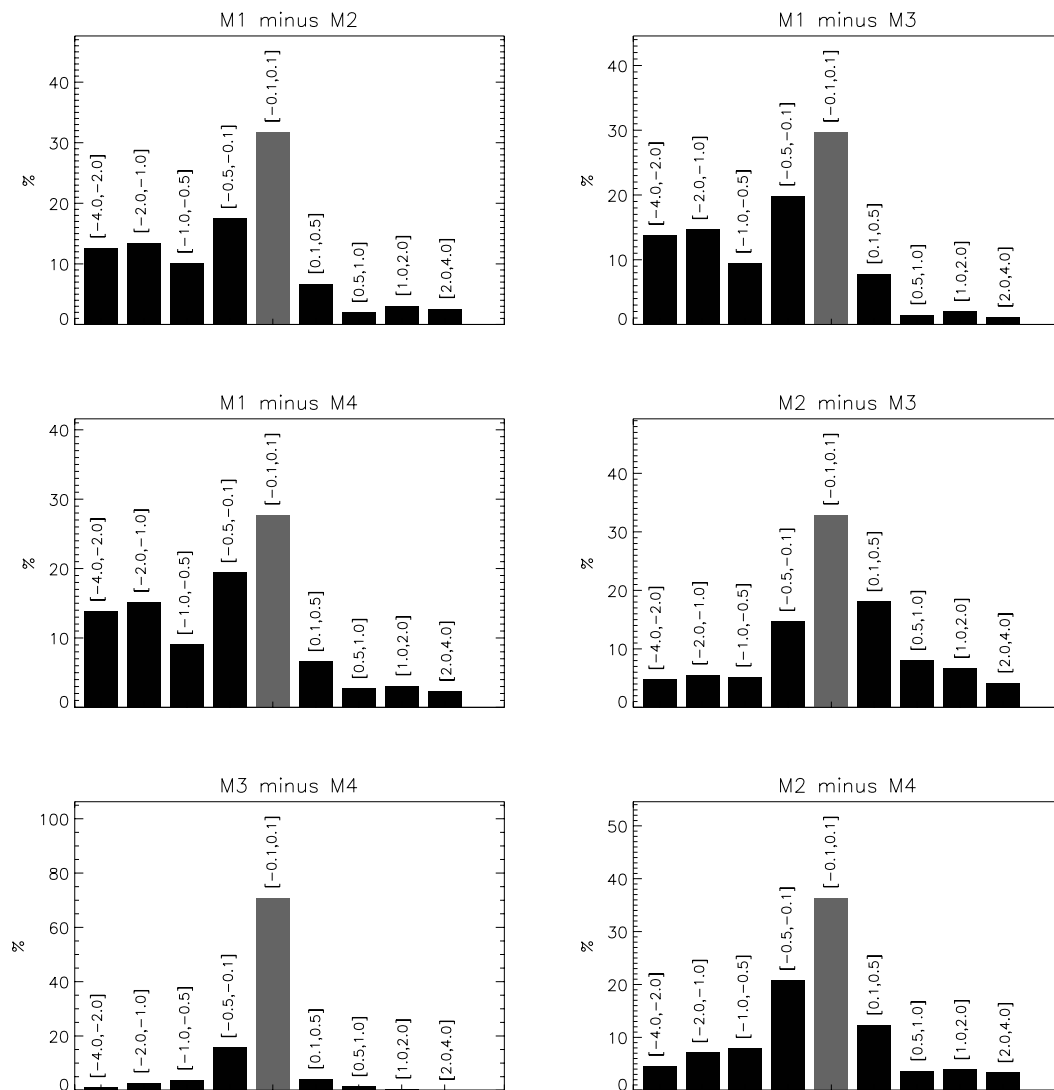
Fig. 5. As for Fig. 4 but for M1, M2, M3, and M4 (as described in Sect. 2.3).

no flagged data we expect all methods to give the same result, i.e. 100% show "high agreement". This is correct for all methods (except for M6) and all parameters as the HIAG fraction is generally above 90% for the bin with 99–100% "good" measurements. This discrepancy may be due to the fact that M6 is not based on the same version of the underlying measurements because the monthly means obtained with M6 were derived from an earlier retrieval that may slightly differ from the most recent version that is currently stored in the BSRN database. In addition, M6 applies a very sophisticated quality control (Long and Shi, 2006) of the radiation fluxes (Sect. 2.3). This means that the percentage of flagged (and corrected) data might substantially deviate from the fraction of flagged data when using the quality procedure that is routinely applied to the BSRN observations. Excluding method M6, we learn from Fig. 6 that the percentage of "high agreement" for SW monthly means drops

down to 60–80% for underlying measurements with only 90–95% of "good" data (data inside the "extreme rare" limits). The difference in montly LWDOWN obtained with the seven investigated algorithms generally differ less than for SW fluxes. For LWDOWN, the HIAG percentage remains above 90% even for measurements with a substantial part of missing or flagged data (Fig. 6). This is a clear hint that monthly LWDOWN fluxes are less affected by data gaps and/or flagged data as temporal variability of LWDOWN is generally distinctly lower than for SW fluxes. For months with more than 99% of the observation being within the "extreme rare" limits, more than 99% of the monthly means obtained with different methods (excluding again M6) do not differ by more than $0.2\,\mathrm{W\,m^{-2}}$. Note that the population is sufficiently large (more than 2000 valid monthly means) in order to guarantee statistical robust results. For SWDIFF and LWDOWN in Fig. 6, only M7 and M4 provide a sufficient
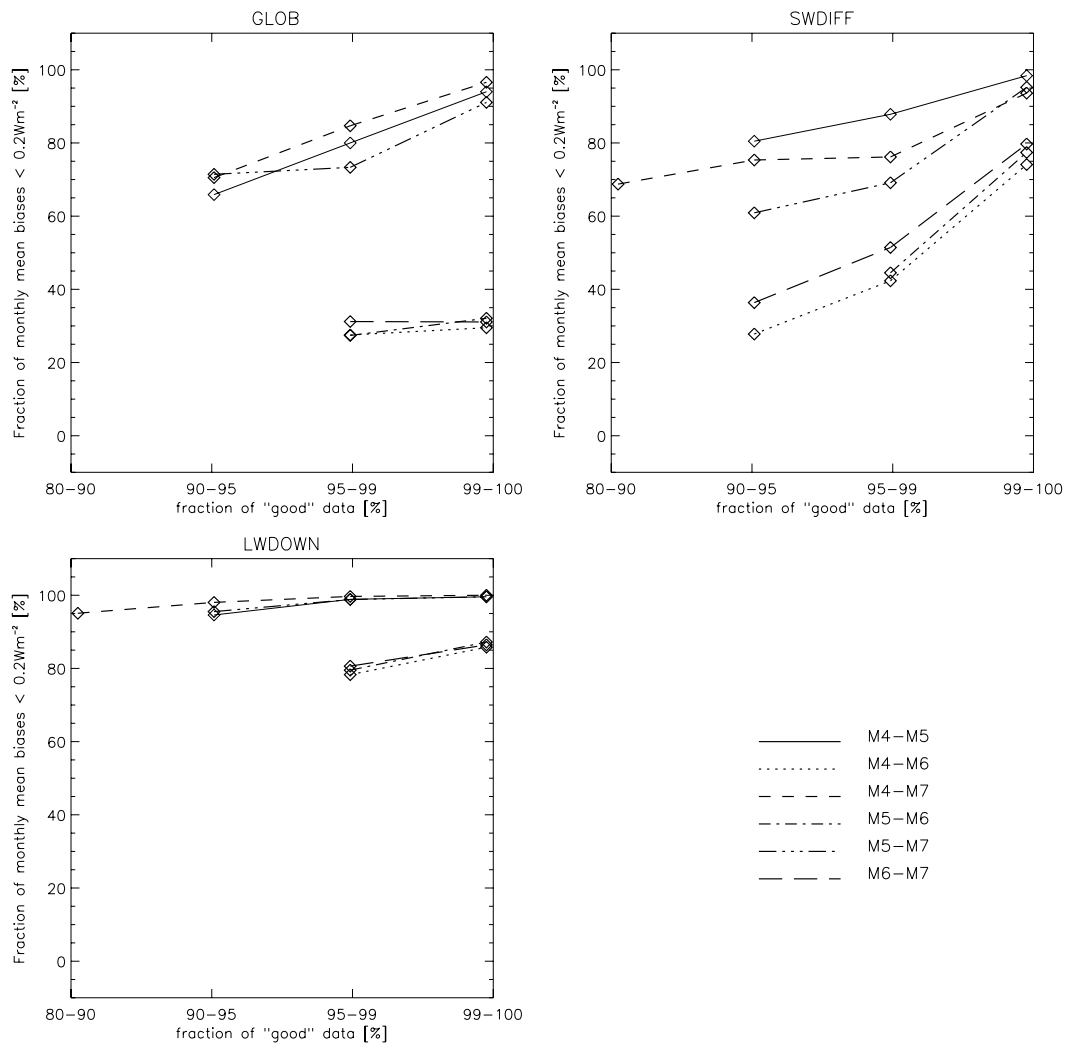
Fig. 6. The percentage of monthly means derived from two methods differing by less than $0.2\,\mathrm{W\,m^{-2}}$ versus the fraction of the underlying 1-min observations that are within the "extremely rare" limits for the radiation quantities indicated for each plot. Only cases with a sufficiently large basic set (>30 valid monthly means) are displayed. Methods M1–M7 are described in Sect. 2.3. The "extreme rare" limits are listed in Table 2. Note that for SW fluxes, the night-time offsets are not considered.

number of valid monthly means for cases with only 80–90% of "good" (inside the "exreme rare" limits) data. This feature is directly related to the setup of the methods: M4 is based on a interpolation of missing and flagged data while M7 allows the computation of valid monthly fluxes also for high fractions of missing and flagged data due to taking advantage of the typical diurnal cycle of the SW fluxes. As complete time series are an important prerequisite for the determination of accurate trends in radiation fluxes (see e.g., Wild et al., 2005), we favor methods that allow the computation of reasonable monthly means such as M4 and M7. Both methods account for the diurnal and seasonal cycle. We favour method M7 over M4 as the extra task of computing solar zenith angle is not necessary.

The results shown in Fig. 6 can be repeated for the set M1, M2, M3, and M4 (not shown). This provides valuable insight into the impact of the interpolation of missing and flagged data on the computed monthly mean. The evaluation reveals that the fraction of monthly means with HIAG (difference less than $0.2\,\mathrm{W\,m^{-2}}$) decreases most rapidly between M1 and M2 with a decreasing percentage of "good" data. This is reasonable as M1 includes all flagged data and no interpolation of gaps while M4 applies an interpolation of flagged and missing data. The relationship for M3 and M4 are similar pointing to the fact that the computed monthly means do depend little on the applied interpolation method. As in Fig. 6, monthly LWDOWN is less sensitive to the fraction of missing and flagged data than are SW fluxes.
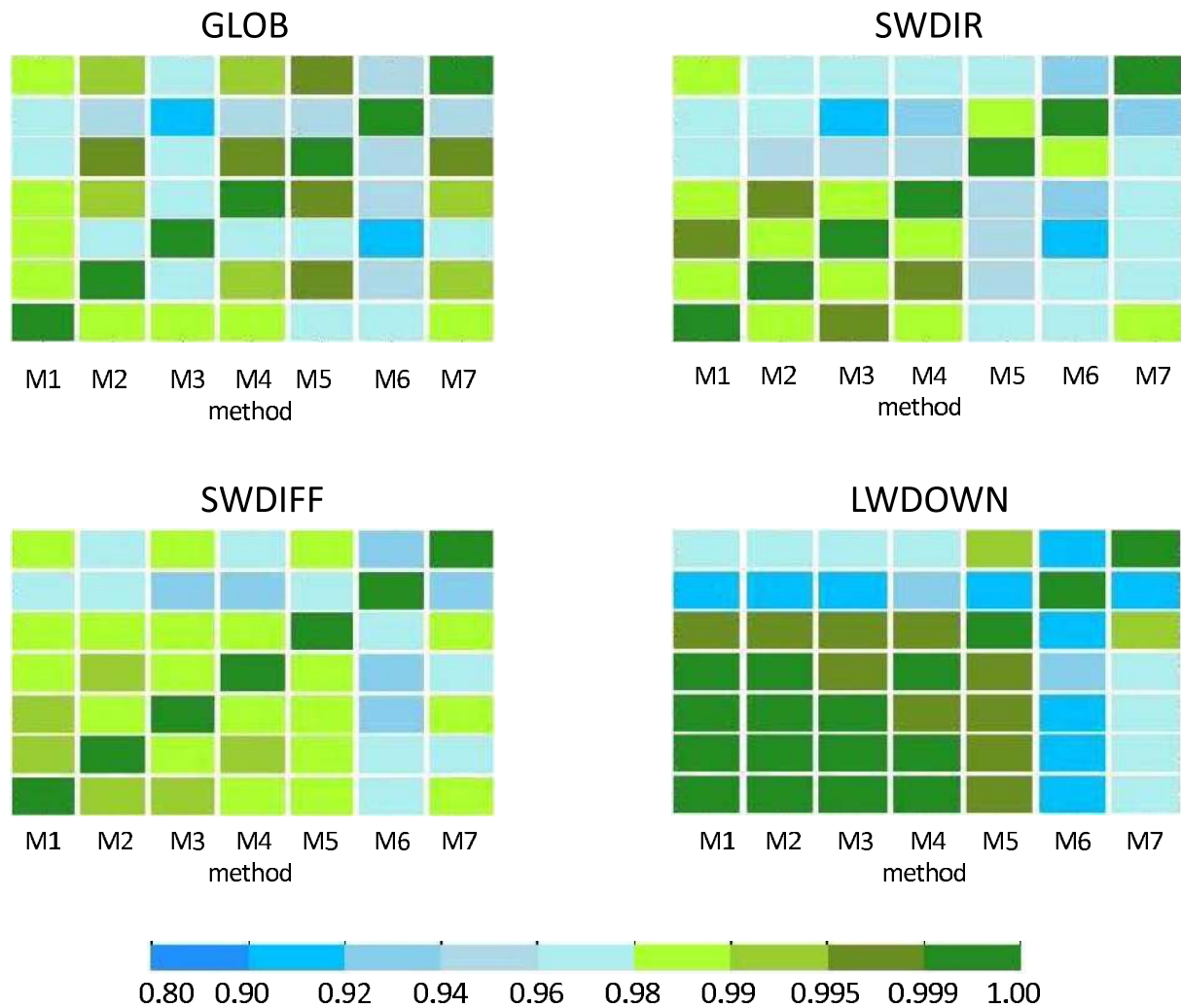
GLOB

SWDIR

SWDIFF

LWDOWN

Fig. 7. Color coded correlation matrices for the monthly averages using the seven methods M1–M7. Displayed are the correlation coefficients for GLOB (Global Radiation), SWDIR (direct SW radiation), SWDIFF (diffuse SW radiation), and LWDOWN (downward longwave radiation). Correlations are computed from deseasonalized data and averaged over all BSRN sites. Monthly SWDIR computed with M5 and M6 cannot be compared to the other methods as they provide SWDIR on a horizontal surface. Note that M6 provides data for only nine BSRN sites. Further, M6 is based on a earlier retrieval from the BSRN database that may slightly differ from the most recent data version.

3.4 Correlation of monthly mean time series

The strength of the linear relationship between the monthly fluxes compiled from two differing methods will be investigated by checking the correlation coefficients. Figure 7 gives a visual overview on the correlation between the monthly time series between any pair of methods. The correlations are computed using deseasonalized data. The mean correlations shown in this figure are determined in two steps (for each method pair and each parameter): (i) computation of the correlation coefficients for each individual BSRN site, and (ii) calculation of the arithmetic mean of the correlation coefficient computed in (i). Note that M6 only provides data for nine BSRN sites whereas monthly means for all BSRN stations are available for the other six algorithms M1–M5

and M7. Figure 7 shows that the monthly means derived from various methods mostly correlate quite well with correlation coefficients > 0.96. M6 generally shows the lowest correlation with the other investigated algorithms for all radiation components. This is partly due to the smaller amount of available monthly mean data for M6. Very high correlations are found for downwelling LW radiation between the methods M1–M5 which is likely related to a rather small percentage of missing and flagged LWDOWN observations (see Table 5). Furthermore, temporal variability in LW-DOWN is generally smaller than in the SW fluxes which minimizes the effect of data gaps on the monthly mean. M5 and M7 compute monthly means from monthly mean diurnal cycles but handling of missing and flagged data differs. Furthermore, the details on the computation of monthly mean

diurnal cycles differ. The mean correlation for these two methods are above 0.98 for the SW (excluding SWDIR) and LW fluxes. The lower correlation for SWDIR between M5 and M7 may be caused by frequent data gaps and a considerable amount of flagged data.

Summarizing, the fraction of missing/flagged data do clearly impact the monthly means obtained with two different methods. The intra-method differences are generally smaller for LWDOWN than for SW fluxes as LWDOWN shows less temporal variability than SW fluxes, which lowers the effect of missing (1-, 2-, 5-) min values on the monthly mean estimate.

Further investigation revealed that the methods generally are more sensitive to changes in the gap frequency than to the amount of flagged data. We conclude from this that, in order to decrease the uncertainty in the computed monthly fluxes, the gaps in the data series should be decreased. The quality control that was implemented in BSRN at ETHZ has also the potential to improve the accuracy of the computed monthly means.

3.5 Trends in global radiation estimated by different methods

The phenomena of global brightening has been widely discussed during the last few years (Wild et al., 2005; Gilgen et al., 2009; Wild, 2009). Trend estimation is, however, dependent on the quality and homogeneity of the time series. Furthermore, we show here that trend estimates may be also influenced by the method how monthly means have been estimated from the minute data. In order to estimate the effect of the selected method on the trend in global radiation, we analyzed stations with measurements starting in 1997 or before with no continuous longterm gaps. Trends have been computed on the basis of annual means. Annual means were calculated from monthly means if more than eight valid monthly means were available for the respective year. Considering these conditions and provided that three or more out of the seven investigated methods produce valid annual means during the 10-year time-period 1997–2006, we select 11 sites for our investigation. Least square linear regression was then applied for a trend analysis (Table 7). Ten out of the 11 investigated times series show a positive mean trend during 1997–2006. However, the tabulated standard deviations clearly reveal that the estimated trends strongly depend on the selected method. For some sites, e.g. GVN or SPO, the sign of the estimated trend in global radiation depends on the selected method. A closer investigation reveals that the main reason for the observed differences is due to the fact that the number of annual means taken into acount in the computed trends for 1997–2006 largely differs among the methods. However, note that the differences would be less significant for longer time series, as, considering the limited length of the time-series involved, the computed trends are highly sensitive to the number of annual means included.

Table 7. Mean trends in global radiation (1997–2006) for 11 BSRN sites averaged over the different filling methods applied to each site. The trend analysis is restricted to the methods that produced valid annual means for the whole 10-year period for each site. The filling methods used are listed in the second column. For a method description along with the used abbreviations see Sect. 2.3. The 4th column shows the standard deviation (STDEV).

station	Methods	trend $(\mathrm{W\,m^{-2}\,yr^{-1}})$	STDEV $(\mathrm{W\,m^{-2}})$
ASP	M2, M3, M4, M7	0.73	0.44
BAR	M2, M3, M4, M5, M6, M7	0.82	0.75
BER	M1, M2, M3, M4, M5, M6, M7	0.98	0.67
BOU	M2, M3, M4, M5, M6, M7	1.54	1.12
GVN	M2, M3, M4, M5, M6, M7	0.62	1.94
KWA	M3, M4, M6, M7	0.33	0.41
MAN	M3, M4, M7	−3.6	0.4
NYA	M1, M2, M3, M4, M5, M6, M7	0.17	1.01
PAY	M1, M2, M3, M4, M5, M6, M7	1.70	0.98
SPO	M3, M4, M7	0.96	1.91
TAT	M3, M4, M7	3.21	0.04

4 Summary and conclusion

This work demonstrates the issues month mean computation caused by missing observational data.

This study investigates the completeness of the currently available BSRN data and its impact on computed monthly means that have been obtained with different methods. The range of results could indicate uncertainties in any unspecified method where gaps exist and the method is not clearly described.

The simple quality analysis show that the data quality at most sites is generally good. The percentage of observations that are outside the "extreme rare" limits are generally below 2%. The "across-quantity" conditions are mostly satisfied at all BSRN sites. The constraint that GLOB and GLOB1 should not differ by more than 8% is often violated. At 14 BSRN sites, this test fails for about 2% of the observations.

The gap analysis reveals that BSRN radiation measurements have many data gaps. At 10 (2) BSRN sites, the percentage of missing global radiation is above 5% (15%). For SWDIR, even 16 (8) sites suffer from more than 5% (15%) data gaps.

Within this study, seven methods for the computations of monthly means from minute-values have been intercompared. The results showed that the computed monthly means may differ by several $\mathrm{W\,m^{-2}}$. Selecting months with more than 99% high quality data (less than 1% missing data or outside the "extreme rare" limits), M4 and M7 show the best agreement. This gives some confidence that M7 may be well qualified for the computation of BSRN monthly means. This algorithm omits flagged data and profits from the typical diurnal cycle of SW radiation fluxes. M6, however,

significantly deviates from the monthly means derived from the other methods. This is likely due to the more stringent and sophistciated quality control that has been applied to the data prior to the monthly mean computation. The comparison of the four methods M1, M2, M3, and M4 reveals that it is crucial to take the quality flags into account. For example, M1 differs by more than $2\,W\,m^{-2}$ from M2, M3, and M4 for GLOB in approximately 15% of all monthly means.

This study shows that monthly mean estimates may substantially depend on the selected averaging algorithm. The discrepancy between the methods generally increases with increasing fractions of missing/flagged data. It has been shown that it is essential to account for data quality flags when computing monthly fluxes from 1-min observations. From the comparison study, it is advantageous to compute monthly fluxes by first computing the mean monthly diurnal cycle as this minimizes the impact of missing values.

Of the methods used here, the authors suggest the application of method M7 when computing monthly means from BSRN observations. This method accounts for both the diurnal and seasonal cycle in the radiation data without computing the solar zenith angle. This helps to avoid different monthly mean estimates being used in the literature for the same site and month. Finally, it is essential to note that with missing data – inevitable in real-world observations – there will be no perfect and error-free method because by definition, not filling gaps will bias the results and filling data requires estimating values which are not exact, especially for such potentially highly variable radiation parameters.

In addition to the presented empirical methods, robust methods could be applied to BSRN data in order to avoid biased estimates for data records with a high percentage of flagged data and/or frequent data gaps. The field of robust statistics (for an introduction, see, e.g. Huber, 1981; Hampel et al., 1986 or Maronna et al., 2006) might be the ideal tool to implement more mathematically founded methods which allow – contrarily to purely empirical methods – for the computation of confidence intervals. Robust estimators have been recently applied to incomplete data (Frahm, 2009) and to non-stationary processes (Horenke, 2010).

References

Dutton, E. G., Nelson, D. W., Stone, R. S., Longenecker, D., Carbaugh, G., Harris, J. M., and WendellDye, D.: Decadal variations in surface solar irradiance as observed in a globally remote network, J. Geophys. Res, 111, D19101, doi:10.1029/2005JD006901, 2006.

Frahm, G.: Asymptotic Distributions of Robust Shape Matrices and Scales, J. Multivar. Anal.", 100, 1329–1337, 2009.

Gilgen, H., Wild, M., and Ohmura, A.: Global Energy Balance Archive (GEBA), Report 3: The GEBA Version 1995 Database,, Zürcher Geographische Schriften No. 52, ETH Zürich, 1997.

Gilgen, H., Roesch, A., Wild, M., and Ohmura, A.: Decadal changes in shortwave irradiance at the surface in the period from 1960 to 2000 estimated from Global Energy Balance Archive Data, J. Geophys. Res., 14, D00D08, doi:10.1029/2008JD011383, 2009.

Haeffelin, M., Kato, S., Smith, A., Rutledge, K., Charlock, T., and Mahan, J.: Determination of the thermal offset of the Eppley Precision Spectral Pyranometer, Appl. Optics, 40, 472–484, 2001.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel: Robust Statistics: The Approach Based on Influence Functions, John Wiley & Sons, New York, 1986.

Hinkelman, L. M., Stackhouse Jr., P. W., Wielicki, B. A., and Zhang, T., and Wilson, S. R.: Surface insolation trends from satellite and ground measurements: Comparisons and challenges, J. Geophys. Res., 114, D00D20, doi:10.1029/2008JD011004, 2009.

Horenko, I.: On robust estimation of low-frequency variability trends in discrete markovian sequences of atmospheric circulation patterns, J. Atmos. Sci., 67, 1559-1574, 2010.

Huber, P. J.: Robust Statistics, John Wiley & Sons, volume 2, New York, 1981.

Liepert, B.: Observed reductions of surface solar radiation at sites in the United States and worldwide from 1961 to 1990, Geophy. Res. Lett., 29, 12, doi:10.1029/2002GL014910, 2002.

Long, C. N. and Shi, Y.: The QCRad Value Added Product: Surface Radiation Measurement Quality Control Testing, Including Climatologically Configurable Limits, Tech. Rep. ARM TR-074, 2006.

Maronna, R., Martin, D., and Yohai, V.: Robust Statistics: Theory and Methods, Wiley Series in Probability and Statistics, John Wiley & Sons, 2006.

Ohmura, A., Dutton, E., Forgan, B., Frölich, C., Gilgen, H., Hegner, H., Heimo, A., König-Langlo, G., Mcarthur, B., Müller, G., Philipona, R., Pinker, R., Whitlock, C., and Wild, M.: Baseline Surface Radiation Network (BSRN/WRMC), a new precision radiometry for climate research, B. Am. Meteorol. Soc., 79, 2115–2136, 1998.

Ramanathan, V., Crutzen, P., Kiehl, J., and Rosenfeld, D.: Aerosol, climate and the hydrological cycle, Science, 294, 2119–2124, 2001.

Stanhill, G. and Cohen, S.: Global dimming: a review of the evidence for a widespread and significant reduction in global radiation, Agr. Forest Meteorol., 107, 255–278, 2001.

Wild, M., Long, C. N., and Ohmura, A.: Evaluation of clear-sky solar fluxes in GCMs participating in AMIP and IPCC-AR4 from a surface perspective, J. Geophys. Res., 111, D01104, doi:10.1029/2005JD006118, 2006.

Wild, M.: Shortwave and longwave surface radiation budgets in GCMs: a review based on the IPCC-AR4/CMIP3 models, Tellus, 60, 932–954, doi:10.1111/j.1600-0870.2008.00342.x, 2008.

Wild, M.: Global dimming and brightening: A review, J. Geophys. Res., 114, D00D16, doi:10.1029/2008JD011470, 2009.

Wild, M., Ohmura, A., Gilgen, H., and Roeckner, E.: Validation of GCM simulated radiative fluxes using surface observations, J. Climate, 8, 1309–1324, 1995.

Wild, M., Gilgen, H., Roesch, A., Ohmura, A., Long, C., Dutton, E., Forgan, B., Kallis, A., Russak, V., and Tsvetkov, A.: From Dimming to Brightening: Decadal Changes in Solar Radiation at Earth's surface, Science, 308, 847–850, 2005.

Zhang, Y.-C., Rossow, W. B., Lacis, A. A., Oinas, V., and Mishchenko, M. I.: Calculation of radiative flux profiles from the surface to top-of-atmosphere based on ISCCP and other global datasets: Refinements of the radiative transfer model and the input data, J. Geophys. Res., 109, D19105, doi:10.1029/2003JD004457, 2004.