

Renner Maik (Orcid ID: 0000-0002-2992-8414)  
Schwarz Matthias (Orcid ID: 0000-0002-0043-3522)  
Kleidon Axel (Orcid ID: 0000-0002-3798-0730)

## Estimating shortwave clear-sky fluxes from hourly global radiation records by Quantile Regression

M. Renner<sup>1</sup>, M. Wild<sup>2</sup>, M. Schwarz<sup>2</sup>, and A. Kleidon<sup>1</sup>

<sup>1</sup>Max-Planck-Institute for Biogeochemistry, Jena, Germany.

<sup>2</sup>ETH Zürich, Institute for Atmospheric and Climate Science, Zürich, Switzerland

Corresponding author: Maik Renner (mrenner@bgc-jena.mpg.de)

### Key Points:

- Estimate monthly fractional clear-sky shortwave transmission from commonly available global radiation measurements
- Validated against a standard procedure by Long and Ackerman (2000) across the Baseline Surface Radiation Network
- No further input variables are required to obtain monthly shortwave clear-sky radiative fluxes for various applications

### Abstract

Estimates of radiative fluxes under cloud-free conditions (“clear-sky”) are required in many fields, from climatic analyses of solar transmission to estimates of solar energy potential for electricity generation. Ideally, these fluxes can be obtained directly from measurements of solar fluxes at the surface. However, common standard methods to identify clear-sky conditions require observations of both, the total and the diffuse radiative fluxes at very high temporal resolution of minutes, which restricts these methods to a few, well equipped sites.

Here, we propose a simple method to estimate clear-sky fluxes only from typically available global radiation measurements ( $R_{sd}$ ) at (half-)hourly resolution. Plotting a monthly sample of observed  $R_{sd}$  against the corresponding incoming solar radiation at the top of atmosphere (potential solar radiation) reveals a typical triangle shape with clear-sky conditions forming a distinct, linear slope in the upper range of observations. This upper slope can be understood as the fractional transmission of solar radiation representative for cloud-free conditions of the sample period. We estimate this upper slope through quantile regression.

We employ data of 42 stations of the worldwide Baseline Surface Radiation Network (BSRN) to compare our monthly estimates with the standard clear-sky identification method developed by Long and Ackerman (2000). We find very good agreement of the derived fractional solar transmission ( $R^2 = 0.73$ ) across sites. These results thus provide confidence in applying the proposed method to the larger set of global radiation measurements to obtain further observational constraints on clear-sky fluxes and cloud radiative effects.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019EA000686

## Plain Language Summary

Even under cloud-free condition we typically observe lower solar radiative fluxes at the surface than the incoming solar radiation at the top of atmosphere. A fraction of the solar radiation is being absorbed in the atmosphere and thereby reduces the amount received at the surface. Common standard methods to estimate the clear-sky fluxes from ground-based observations need a very high temporal resolution of 1 min and additional measurements of diffuse radiation to identify cloud-free conditions. Since these data are not broadly measured, we propose an alternative method, which exploits the typically linear relationship of observed, cloud-free solar radiation fluxes to the incoming solar radiation at the top of the atmosphere. This relationship can be estimated by the statistical method of quantile regression. We test this alternative approach with high-quality surface radiation data and compare it with the standard approach. Results show that the alternative method yields good agreement of monthly fractional clear-sky transmission and thus clear-sky fluxes for the 42 sites evaluated. This provides confidence to utilize the commonly measured global radiation for the estimation of clear-sky fluxes in applications ranging from solar energy planning to a climatology of the shortwave cloud radiative effect from direct observations.

## 1 Introduction

Solar radiation provides the main energy input to the Earth system and is of great importance for the local climate and human activities such as agriculture and solar energy generation. Diurnal and seasonal variations in surface solar radiation are mainly determined by astronomical settings which determine the incoming solar radiation at the top of atmosphere (in the following referred to as potential solar radiation). Further variation is caused by the atmospheric conditions with clouds being the most dominant source of variation. The radiative properties of clouds contribute to most of the uncertainty in our ability to model and predict current and future climate (Bony et al., 2015). Apart from clouds, the solar beam is attenuated by the presence and concentration of gases and aerosols, which also cause variations of  $R_{sd}$  in space and time (Iqbal, 1983).

The solar fluxes under cloud-free (used here synonymously for clear-sky) conditions serve as an important baseline describing the potential of solar energy and are used as a reference to quantify the radiative effects of clouds. Water vapour, and aerosols from natural and anthropogenic sources will further affect the clear-sky solar radiation and contribute to climate changes. Especially, the contribution of aerosols is, however, rather uncertain for the past, present, and future (Boucher et al., 2013). The ability to resolve clear-sky fluxes and especially the absorption of solar radiation in the cloud-free atmosphere was found to correlate with the hydrological sensitivity of climate models (DeAngelis, Qu, Zelinka, & Hall, 2015). Although remote sensing based surface solar radiation products are available to compare climate models at global scale, these surface fluxes are modeled by radiation transfer with atmospheric profiles obtained from numerical weather models, which leads to large uncertainties in the fluxes. Direct observational estimates are still required to validate the clear-sky fluxes from radiative transfer models (Zhang et al., 2019).

Such a comparison of clear-sky fluxes directly estimated from surface radiation measurements with historical runs of climate models has actually revealed substantial model biases and helped to obtain a global cloud-free energy balance (Wild et al., 2019). Their comparison was based on 54 BSRN stations across the globe. These measurements follow a strict protocol and include direct radiation measured with pyrheliometers, as well as diffuse

radiation and the total incoming solar radiation (or global radiation) measured with pyranometers at high temporal resolution (minute data) (Driemel et al., 2018; Ohmura et al., 1998).

There are many different approaches to determine clear-sky fluxes described in the literature, see Gueymard et al., (2019) and Ruiz-Arias & Gueymard, (2018) for recent reviews. The simplest approach is based on a comparison of incoming total surface shortwave radiation ( $R_{sd}$ ) with potential solar radiation at the top of atmosphere ( $R_{sd,pot}$ ). The ratio, often referred to as “clearness index” is then used as threshold above which a period is identified as clear. Yet it is unclear which threshold should be used with different values reported in the literature ranging from 0.6 to 0.7 (Reno & Hansen, 2016). Further approaches also use observations of diffuse and direct normal radiation at the surface to classify clear conditions (e.g., Perez et al., 1990; C. N. Long & Ackerman, 2000). Another approach is to directly model the clear-sky irradiance dependent on atmospheric conditions. These models range from simple Linke turbidity factors (Linke, 1922; Ineichen & Perez, 2002) to radiative transfer models (Gueymard, 2008; Lefèvre et al., 2013). The latter approaches require additional input data on the atmospheric composition such as water vapor and aerosol loads (Reno & Hansen, 2016), whereas the former approaches are based on surface observations only.

The method by Long and Ackerman (2000) only requires surface shortwave radiation measurements and has found wide application as a reference (Gueymard, 2012; Kim & Ramanathan, 2008; Reno & Hansen, 2016; Wild et al., 2019). This approach, however, requires global and diffuse radiation at high temporal resolution (1 minute) to estimate clear-sky fluxes (Long and Ackerman 2000). These requirements thus strongly restrict the available data to a few sites and limit the evaluation to a short period after 1990. In contrast, pyranometers, which measure global radiation, are much more widely used by the various weather services and by research networks (e.g., FLUXNET). To address this need we present a simple, robust methodology, which is based on global radiation with (half-) hourly temporal resolution.

To overcome these restrictions of the standard clear-sky identification approaches we propose a signature-based approach which takes the observations of global radiation out of their temporal context and relates them to the potential solar radiation. The potential solar radiation describes the radiation which would be received at the surface under a non-attenuating atmosphere at a certain location and only depends on the date, time, and latitude. Figure 1 illustrates such a relationship with half-hourly data of one month at the BSRN site Lindenberg, Germany. Data recorded under cloud-free conditions forms a linear slope in the upper range of observations, whereas cloudy conditions with reduced global radiation fall below this upper linear slope. Such signatures are found to be rather typical and thus can be used to estimate the fractional clear-sky transmission of solar radiation (also known as clearness index), representative for a given period. To estimate this upper linear slope, we employ quantile regression to estimate conditional quantile functions (Koenker 2005). The proposed method requires a relatively large sample of clear-sky conditions to obtain statistically significant estimates. Here, we used a monthly sampling period with (half-)hourly data. Therefore, the resulting fractional clear-sky transmission is only representative for the sampling period of one month. Hence, the method is not suitable for daily or hourly assessments and is thus applicable for climatological analyses, such as seasonal changes of clear-sky fluxes.

In order to test and validate the proposed methodology, we use data from the BSRN network with a sparse but global coverage from the tropics to the poles. The BSRN network provides 1 min data of diffuse and global shortwave radiation (Driemel et al., 2018). Reference clear-sky fluxes are computed by the standard method of Long and Ackerman (2000). We then employ the quantile regression methodology on monthly samples of half-hourly aggregated data of global radiation only. We use fixed parameters for all sites to test if the method can be applied to any site and season without a-priori knowledge. Results are reported for 54 sites of the BSRN network in terms of monthly clear-sky shortwave fluxes and the fractional shortwave transmission. We compare the new estimates with monthly estimates from the Long and Ackerman (2000) reference method using scatter diagrams and standard error characteristics.

## 2 Materials and Methods

### 2.1 Modelling potential solar radiation

Potential radiation at the horizontal surface can be modelled as a function of the solar zenith angle  $\theta$  and top of atmosphere solar radiation  $S_0$  (Long and Ackerman 2000):

$$R_{sd,pot} = S_0 (\cos(\theta))^b \quad (1)$$

, whereby  $S_0$  is the incoming radiation at the top of the atmosphere. The exponent  $b$  modulates the form of the cosine and allows to consider the modulation of the optical path through a curved, refractive atmosphere compared to a plane-parallel, nonrefractive atmosphere, which will be discussed below. The solar zenith angle  $\theta$  is a function of latitude  $\varphi$ , the degree of declination  $\delta$  between the Sun and the equator determined by day of the year and the hour angle  $H$  given by the time of the day (Iqbal, 1983):

$$\cos \theta = \cos H \cos \varphi \cos \delta + \sin \varphi \sin \delta \quad (2)$$

The sun-earth orbit eccentricity correction is applied in the calculation of the degree of declination  $\delta$ . Specifically, we used the function `fCalcPotRadiation` of the R-software package `REddyProc` (Wutzler et al., 2018). That function implements the correction by Spencer (1972). Hence the potential radiation can be computed when location and time are known.

The potential radiation formulation as such assumes a plane-parallel atmosphere through which the light beam travels. This causes a small non-linearity of the observed to potential radiation relationship, which is exemplified in Figure 2 for a clear-sky day at the site Lindenberg. Assuming a plane-parallel, non-refractive atmosphere ( $b = 1$ ) we find a small, but consistent curvature (Fig 2a) which shows larger deviations to the linear fit at lower potential solar radiation (morning and evening), see Fig 2b. To account for the effects of a curved, refractive atmosphere, we use the exponent  $b$  in equation 1 similar to Long and Ackerman (2000). We find that the non-linear deviations can be largely resolved with an exponent  $b = 1.2$ , see Fig. 2. The exponent  $b = 1.2$  is also used in the formulations for quality control of surface shortwave radiation measurements for BSRN sites (Long and Shi 2008, Roesch et al., 2011). In contrast to Long and Ackerman (2000), we choose to use a fixed exponent  $b = 1.2$  for all sites and periods for two reasons. First, we find that the residual squared error of the quantile regression (explained below) shows a minimum close to the exponent  $b=1.2$  across sites and seasons. Secondly, we find that the slope of the quantile regression covaries linearly with the exponent  $b$ , which limits the parameter identifiability. Since we are mainly interested in a monthly change in fractional clear-sky transmission, we keep exponent  $b$  constant, enabling better comparability across sites and seasons.

## 2.2 Clear-sky signature and Quantile Regression approach

The key of the proposed methodology is to establish and quantify the relationship of the observed global radiation with the potential solar radiation. Potential radiation is only a function of date, time and location and can thus be calculated for any site a-priori. The relationship can be best visualized by plotting observed global radiation ( $R_{sd}$ ) against potential radiation  $R_{sd,pot}$  in a scatterplot. This takes the time series observations out of their temporal context (Figure 1) and reveals an approximately linear relationship of  $R_{sd}$  to  $R_{sd,pot}$  close to the upper boundary. Clouds mostly reduce  $R_{sd}$  when direct radiation is blocked or enhance  $R_{sd}$  under partial cloud cover with enhanced diffuse radiation and but direct radiation passing to the surface.

We assume that the slope close to the upper linear boundary represents the fraction of transmitted solar radiation under clear-sky conditions. To estimate this slope from the data we employ quantile regression. Quantile regression is an emerging statistical method for linear and non-linear response models (Koenker & Machado, 1999) which provides an estimate of a conditional quantile function (Koenker 2005). Here we use the univariate linear form relating response  $Y$  to forcing  $X$  through:

$$Y(\omega) = \alpha + \beta X \quad (3)$$

Thereby we estimate the intercept  $\alpha$  and slope  $\beta$  conditional on the quantile  $\omega$ . The resulting regression line, say for a quantile  $\omega = 0.85$ , ensures that 85% of the data is below the estimated line (see Figure 1). The coefficients are estimated using the default method recommended by the author of the quantreg R package (Koenker 2005).

The estimated slopes should be in the range  $0 < \beta(\omega) < 1$ , since observed global radiation should not exceed the potential radiation. The intercept term  $\alpha$  should also be close to 0. Deviations from 0 indicate systematic biases in the measurements, such as problems of timing of measurements or issues with the mounting of the pyranometers.

The quantile regression approach allows to estimate the standard deviation of the derived slope and intercept terms and thus provides a quantitative assessment of the uncertainty. Goodness of fit is evaluated through a pseudo- $R^2$  measure  $R^1$  suggested by Koenker and Machado (1999). The rationale is to assess the goodness of fit at a certain quantile by comparing the sum of weighted deviations of the model of interest ( $\rho_{mod}$ ) with the sum of weighted deviations from a model in which only the intercept appears ( $\rho_1$ ):

$$R^1 = 1 - \rho_{mod} / \rho_1 \quad (4)$$

, with rho determined by the model residuals  $u$  weighted by the quantile  $\omega$ :

$$\rho = \Sigma(u * (\omega - (u < 0))).$$

The only free parameter in the proposed method is the quantile  $\omega$  used in the quantile regression. A very high  $\omega$  would be sensitive to short periods with very large values of shortwave radiation, for example caused by reflections during partial cloud cover. This effect is particularly important when using data at 1 min temporal resolution, but it can be reduced by temporal aggregation to half-hourly or hourly values. A small  $\omega$  will lead to smaller  $\beta$ , which will be important during periods with almost permanent cloud cover. In order to facilitate the choice of the  $\omega$ , we performed a sensitivity analysis with varying values of  $\omega$  and use the standard method as reference with all sites available.

The estimated slopes by the quantile regression  $\beta(\omega)$  represents the fractional transmission of shortwave radiation under cloud free situations for a given period. Assuming

that this slope, representing the fractional clear-sky transmission, is constant for that period, we can estimate the clear-sky fluxes by multiplying with the potential solar radiation:

$$R_{sd,cs} = \beta R_{sd,pot}$$

### 3 Data

#### 3.1 Baseline Surface Radiation Network

The Baseline Surface Radiation Network (BSRN) provides high quality downwelling radiation observations at 1 min resolution for more than 50 sites across the globe (Driemel et al., 2018; Ohmura et al., 1998). Here we use the majority of sites (54) for which monthly clear-sky fluxes are available from the standard method as used by Wild et al., 2019. The location of the sites is shown in the map in Figure 3. Meta data including site codes are tabulated in Table 1.

#### 3.2 Data preparations

We flag suspect data following the BSRN policy (Long and Shi 2008, Roesch et al., 2011) and set values outside the “extremely rare” limits to missing. Three sites showed lower negative nighttime values than the “extremely rare” minimum limits (SBO, SOV, TAM). This issue which is related to the thermal offset due to longwave loss of the instrument has been often reported (Driemel et al., 2018; Dutton et al., 2001). To cope with these issues, we set negative values  $< -10$  to missing. Remaining negative values during periods with high solar zenith angle ( $\theta > 80^\circ$ ) have been set to 0.

For analysis we use the unshaded pyranometer measurements of global radiation when available. If these are not available for a given time step we use measurements of direct  $R_{sd,DIR}$  and diffuse solar radiation  $R_{sd,DIF}$  to compute total radiation with  $R_{sd,tot} = \cos(\theta) R_{sd,DIR} + R_{sd,DIF}$ . We prefer the use of unshaded pyranometer measurements of global radiation since these data are broadly available in other measurement networks in contrast to direct normal and diffuse radiation.

To aggregate the data, which is available in 1-min resolution, we follow the recommendations by Roesch et al., 2011. Thereby, we first calculate 15-min averages where only 20% of the data must be available. From these 15-min aggregates we compute 30-min aggregates which requires 2 values. For hourly aggregates we require 3 non-missing 15-min values. Monthly fluxes are obtained by first computing the monthly mean diurnal cycle from the 30min data, whereby 26 non-missing values for each time step are required. Then the monthly mean is calculated from the monthly mean diurnal cycle requiring no missing values. These procedures try to reduce sampling biases which can be significant due to the pronounced diurnal cycle.

#### 3.3 Application of the quantile regression approach to the BSRN data

The quantile regression (QR) approach is applied for each month at all available sites with sufficient data. Here we use half-hourly data as basis for regression. The resulting estimates are very similar when hourly aggregates are used.

Persistent cloud cover can reduce the slope of the quantile regression because of insufficient cloud free measurements. This is identified by a poor fit using the  $R^1$  goodness of fit statistic and a deviation of more than 25% of the site average values obtained by regression of all data at one site. If  $R^1 > 0.75$  and the slope is within a 25% range of the site average value the estimate is retained. If not, we increase the sampling period by including

the preceding and following month. We used 3, 5, and 7 months sampling windows which allow estimates for most observations in the BSRN data base.

At higher latitudes ( $|\phi| > 67^\circ$ ) the solar zenith angle will not rise above the horizon during winter days and therefore  $R_{sd,pot} = 0$ . However, many sites still measure very small amounts of global radiation (usually  $< 30 \text{ W m}^{-2}$ ). During these winter months the potential radiation was observed to be lower than the measurements resulting in  $\beta(\omega) > 1$ . To account for these issues we constrain the quantile regression to periods with more than 100 values of  $R_{sd,pot} > 10 \text{ W m}^{-2}$ . If this constraint is not met, the coefficients are set to missing, while for estimating the monthly mean clear-sky flux we simply use the potential radiation and set  $\beta$  to 1. These considerations allow to calculate annual average clear-sky fluxes for most of the sites in the BSRN network.

### 3.4 Reference clear-sky detection approach

In order to evaluate the performance of the quantile regression approach we use the Long and Ackerman (2000) clear-sky detection approach which provides clear-sky fluxes of shortwave radiation. The approach is based on observations of total and diffuse hemispheric broadband shortwave radiation with a temporal resolution of 1 min. It employs a series of four tests to identify clear-sky episodes and then estimates the clear-sky fluxes of each day with these data using a power-law formulation with the cosine of the zenith angle as independent variable. The tests involve thresholds for acceptable ranges of total and diffusion radiation under clear skies. Also temporal variability is used to identify clear-sky conditions, one for total radiation and one for the ratio of diffuse to total radiation normalized by the cosine of the zenith angle. All tests require to set thresholds, which are based on experience and can be locally different. To automate the process, Long and Ackerman (2000) use an iterative procedure of clear-sky identification and parameter fitting. A core limitation are days and periods with persistent cloud cover with less than two hours of clear-sky conditions. The fluxes of those days are estimated by linear interpolation and then monthly fluxes of clear-sky solar radiation are computed. For sites with persistent cloud cover, the Long and Ackerman algorithm uses a different estimation technique which is based on monthly input and historical data to derive clear-sky coefficients (C. N. Long & Gaustad, 2004). This affects especially (sub-)tropical sites in ocean environments (BER, COC, ISH, KWA, MAN, NAU) and cloudy extratropical sites (LER, CAM, SPO). Also a few sites the Long and Ackerman method were marked as doubtful (ALE, ILO, SOV). These sites have been excluded from the comparison of monthly estimates.

The approach by Long and Ackerman (2000) has become a standard clear-sky detection approach. For example it was used to establish a climatology of clear-sky fluxes of 53 BSRN stations to evaluate the modeled clear-sky fluxes of global climate models (Wild et al., 2019). Here, we use these climatological clear-sky fluxes on a monthly mean basis for evaluation of the simpler quantile regression approach.

### 3.5 Statistical evaluation

We use standard methods to calculate the goodness of fit through Pearson correlation and linear regression. Additionally, we employ the mean square error skill score (MSES) to evaluate the skill of the derived clear-sky solar radiation with a reference estimate:

$$\text{MSES} = 1 - \text{MSE}_p / \text{MSE}_{\text{ref}} \quad (6)$$

where  $MSE_p$  is the mean squared error of the prediction and  $MSE_{ref}$  of the references, respectively. As reference we use a constant fractional solar transmission  $\beta_{ref} = 0.81$  multiplied by the potential solar radiation, representing the average across all sites. A perfect skill is indicated by a MSES of 1 and 0 if the prediction is worse than the reference. It can be interpreted as the reduction of variance compared to the reference.

## 4 Results

### 4.1 Sensitivity of the quantile used in regression

The proposed method relies on quantile regression which requires to choose a quantile  $\omega$  to estimate the fractional clear-sky shortwave transmission for a given sample. Using a very large  $\omega$  close to 100% would lead to high  $\beta$  due to short periods with very large values of shortwave radiation, for example, caused by reflections and enhanced diffuse radiation during partial cloud cover. A smaller  $\omega$  will lead to smaller fractional transmissions, which will be important during periods with almost permanent cloud cover.

To assess the sensitivity of  $\omega$ , we repeated the quantile regression for different  $\omega$  in the range between 70-99%. Then skill scores (MSES,  $R^2$ ) were computed using the clear-sky fluxes of the standard method as the reference. To evaluate the skill, we use a constant  $\beta = 0.81$  as reference across all sites and conditions. When data of all the sites are pooled together, we find the best performance in terms of bias and correlation at  $\omega = 85\%$ , see Figure 4a. The explained variance of the monthly fractional solar transmission estimates is  $R^2 = 0.79$  at  $\omega = 85\%$  which is within a flat peak of the maximum correlation. One can note from Figure 4a that the bias is more sensitive to the choice of  $\omega$  than the correlation. In the supplement the relationships of the skill score to  $\omega$  are shown for each site (Figure S1). There is a broad agreement with a maximum skill score at  $\omega=85\%$ , however, some sites show deviations from the global pattern. The sites ASP, BRB, CAR, DRA, SMS, TAM show their maximum MSES skill at distinctly lower values of  $\omega$  with approximately 70%. Hence at these sites we must expect a positive bias to the standard method, when using a global value of  $\omega = 85\%$ . At three sites (BOU, TAT, XIA) we find the maximum skill at  $\omega = 95\%$  indicating an underestimation of the QR approach with  $\omega = 85\%$  at these sites.

An increase in  $\omega$  will lead to a larger estimate of fractional clear-sky transmission. However, the sensitivity study also revealed that the sensitivity of the clear-sky fluxes to the choice of  $\omega$  is rather small. Figure 4b shows the distribution of the difference between  $\omega = 90\%$  and  $\omega = 85\%$  for monthly clear-sky fluxes across all the sites which is relatively uniform across the sites. Overall an increase of  $\omega$  by 5% leads to increase by  $3 \text{ W m}^{-2}$  when averaged across all sites. Given the measurement uncertainty this should be acceptable when no further information, such as diffuse radiation and direct normal radiation is available.

Given the relatively small uncertainty in setting  $\omega$  and because the skill scores for bias and correlation have their maxima close to  $\omega=85\%$  we use that value for subsequent analysis for all sites.

## 4.2 Evaluation of fractional shortwave transmission

We applied the quantile regression approach to each month of the available data using a fixed  $\omega=85\%$  for all sites and conditions. This yields a regression slope which can be interpreted as fractional clear-sky solar transmission. Overall the temporal and spatial variability of the fractional clear-sky transmission as measured by the Long and Ackerman method is well captured by the QR approach as shown in the scatterplot in Figure 5. Most estimates follow the 1:1 line and yield a  $R^2=0.73$  across the 42 sites. The derived estimates and statistics are very similar when using hourly ( $R^2 = 0.71$ ) or half-hourly data, see Figure S2 in the supplement which is based on hourly aggregated data.

Time series of the fractional clear-sky transmission of the QR methods and the Long and Ackerman method are shown for five selected sites in Figure 6 (all sites are shown in the supplementary Figure S3). Here we used an automatic scheme (see methods section) to detect overcast conditions, where months with poor fit are replaced by a fit using a longer sampling period. These conditions are marked in Figure 6 with crosses. At the site PAY there are frequent overcast conditions during winter which would lead to underestimation of  $\beta$ . These conditions are often characterized by small fractional transmissions accompanied by a poor  $R^1$  of the quantile regression and can be effectively identified.

The time series of both methods show a pronounced seasonal cycle at extratropical sites, see Figure 6 (see also Figure S3). During winter there is generally a higher fractional shortwave transmission than in summer. Towards the tropics the seasonal variation is strongly reduced. Furthermore, we find that continental sites show larger seasonal variations than sites in an oceanic environment at comparable latitude (compare sites E13 in central US and BER at the Bermuda islands in the Atlantic, Fig. S3). At the poles, especially at the sites in Antarctica we find very high fractional solar transmission  $> 1$  during the dark winter months. In these periods the potential solar radiation is very small or zero, but there are still diffuse light sources which contribute to the measured solar fluxes. In these cases, the fractional solar transmission can be larger than one. This is, however, less relevant for the estimated clear-sky fluxes since the potential solar radiation in these winter months is very small or zero.

## 4.3 Comparison of clear-sky fluxes

Next, we compare the clear-sky shortwave radiation fluxes obtained from the quantile regression approach with the results obtained from the standard method proposed by Long and Ackerman (2000). We use the monthly average fluxes for comparison. The scatterplot in Figure 7 is based on monthly data of 42 sites. The estimates follow the 1:1 line with only a slight overestimation (slope of best linear fit 1.01) and low error ( $RMSE = 8.5 \text{ W m}^{-2}$ ).

Since the main signal in the clear-sky estimate comes from the seasonal course of potential solar radiation we further compare our results with a reference case which uses a constant fractional transmission  $\beta_{\text{ref}} = 0.81$ . Figure 8 shows the residuals (difference between the QR estimate and the standard method) ordered by the clear-sky fluxes. Both residual distributions (right panel of Figure 8) have their center around zero, but the constant fractional transmission estimate shows a much larger spread. Hence the quantile regression

method as applied to the monthly data is able to reduce the residual variance with  $MSES = 74\%$  compared to the reference with constant solar transmission across all the sites.

#### 4.4 Site level evaluation

The BSRN stations are distributed across the globe and provide measurements in all climate zones. This wide spectrum allows to assess if the proposed method can be applied on any site with solar radiation measurements. The evaluation statistics of all 54 sites are reported in Table 1. Time series of the fractional clear-sky solar transmission of each site are shown in the supplementary figure S3. Note, that the Long and Ackerman method used a different estimation technique for sites with persistent cloud cover (C. N. Long & Gaustad, 2004), which does not allow a fair comparison of the time series. These sites have been excluded from comparison.

At the site TAM the QR method shows a positive trend in  $\beta$  which is not as strong for the Long and Ackerman method, see also Figure 6c. This site is located in the Sahara which is characterized by high loads of dust which can reduce solar radiation at the surface (Polo et al., 2009). More recently it was noted that the ratio of diffuse to total radiation used a one of the tests of the Long and Ackerman method identifies many periods as cloudy which were rather hazy (Ruiz-Arias et al., 2018). The tropical site in continental climate (PTR) shows no correlation of monthly  $\beta$  time series ( $R^2 = 0.04$ ), which can be due to the low seasonal variation of the Long and Ackerman estimate for this site. At the site XIA there is a positive trend in  $\beta$  by the QR approach, which is not seen by the Long and Ackerman method.

The remaining 39 sites show good agreement with the Long and Ackerman method with  $RMSE < 18.5 \text{ W m}^{-2}$  averaging to  $7.7 \text{ W m}^{-2}$  for the clear-sky shortwave fluxes. The explained variance of the fractional clear-sky transmission is on average  $R^2 = 0.63$  and larger than  $R^2 > 0.24$ . Given that only global radiation is used to diagnose the clear-sky fluxes we believe that these deviations at these sites are acceptable and illustrates that the QR method is a reliable tool to estimate fractional solar transmission and clear-sky fluxes at a monthly time scale in various environments and atmospheric conditions.

## 5 Discussion

### 5.1 Interpretation

We introduced and successfully tested a simple methodology to estimate the monthly fractional clear-sky transmission by only using global radiation measurements at a half-hourly resolution. The proposed QR method can be considered as a clear-sky identification approach which is different to approaches which model the radiation transfer through the atmosphere. The latter approaches have to link surface radiation data and column integrated data of water vapor and aerosols with surface radiation measurements which may introduce further uncertainties (Ruiz-Arias & Gueymard, 2018).

The QR method exploits the commonly observed linear relationship of surface solar radiation measurements with incoming solar radiation at the TOA under cloud-free conditions forming a rather slope close to the upper envelope. This linear relationship is reliably estimated with quantile regression when using a rather high quantile. The slope of this linear relationship  $\beta$  can be thought of the fractional solar transmission under cloud-free conditions and it is typically smaller than 1. Here we find an average of  $\beta = 0.84$ , see Table 1.

The complement, i.e.  $(1 - \beta)$  represents fractional absorption and backscattering of shortwave radiation in the atmosphere. Fractional absorption under all sky conditions was found to be around  $0.23 \pm 0.02$  (Hakuba, Folini, & Wild, 2016) and 0.21 under clear-sky conditions (Wild et al., 2019). Note, that Hakuba, Folini, & Wild, (2016) used a difference of top of atmosphere and surface net shortwave radiation under all-sky conditions from the CERES-EBAF remote-sensing product, while here we used cloud-free conditions of the downwelling radiation only. Nevertheless, there is a broad agreement, indicating that a large part of the reduction of solar radiation at the surface under cloud-free conditions can be attributed to the absorption of shortwave radiation within the atmosphere.

We find that the monthly time series of fractional clear-sky transmission show a pronounced seasonal cycle at extratropical sites (Figure 6, S2). Thereby we find higher fractional shortwave transmission during winter compared to summer. The seasonality can be linked to the seasonality of water vapor in the atmosphere which is an important absorber of shortwave radiation (De Angelis et al., 2015). This is also consistent with the reduced seasonality towards the tropics and the lower seasonality in maritime environments compared to continental climates.

## 5.2 Limitations

The proposed QR method strongly relies on sufficient samples of cloud-free conditions. Here we used one month as sampling period which is sufficient for most sites to obtain reliable estimates, however, persistent overcast conditions can lead to distinctly smaller estimates of fractional clear-sky transmission. Here we used the  $R^1$  goodness of fit metric and a threshold of 25% of the site average estimate to identify these periods. Increasing the sampling period allowed to obtain estimates throughout the year for almost all sites and months.

The method also requires choosing a quantile  $\omega$  for which the slope  $\beta$  is being estimated. Generally, this quantile  $\omega$  should be above the 75% percentile and lower than the 97% percentile. At the highest quantile the performance drastically decreases. This is because partly cloudy conditions can enhance global radiation due to enhanced diffuse radiation and high direct radiation. Hence the occurrence of such conditions at a given site can influence the choice of  $\omega$  for quantile regression. Here further research which should include diffuse radiation measurements is required to improve the approach at specific sites.

Here we used the Long and Ackerman method as a reference to find that the quantile of 85% provides the best results across the whole set of sites. There are sites where lower quantiles would improve the skill of the method and further investigation of the radiative conditions at these sites could improve the comparison with the standard method. However, the sensitivity of the correlation and the bias shows a rather flat plateau around the 85% and indicates that the uncertainty stemming from the choice of  $\omega$  is relatively small. Given further uncertainties in measurements we believe that the good error statistics for most sites demonstrate the applicability of the QR approach.

## 5.3 Potential applications

Global radiation is commonly measured in various national meteorological sites, and in many research networks (e.g., FLUXNET) which allows to apply the method for a much wider spatial and temporal domain than available with the BSRN network. Thereby, one can obtain relevant climatic information on fractional solar transmission which is shaped by

absorption of solar radiation in the atmosphere, which allows to compute clear-sky fluxes and the shortwave cloud radiative effect. Spatial and temporal trends are thereby of high interest to understand for example the reasons of global dimming and brightening (Wild et al., 2005). Station based estimates of fractional clear-sky transmission and the cloud radiative effect could help to attribute the decadal changes in global radiation observed e.g., in Europe, Northern America and Asia (Sanchez-Lorenzo et al., 2017; Wild, 2016).

With the dense network of global radiation measurements, it is possible to obtain estimates of clear-sky radiative fluxes at the monthly time scale. These may be used for validation of remote sensing based estimates (e.g., CERES), as reference for climate models (Wild et al., 2019) or for the planning of photovoltaic installations (Ruiz-Arias & Gueymard, 2018).

## 6 Conclusions

We proposed a novel, alternative parsimonious approach to derive clear-sky solar radiation fluxes only from commonly measured global shortwave radiation observations at (half-) hourly resolution. The key idea is that measurements taken under clear-sky conditions are linearly related to potential solar radiation. This slope can be reliably estimated with the statistical method of quantile regression (Koenker, 2005). We used data of the BSRN network which provides the highest quality observations of surface radiation measurements to evaluate the proposed method with a well-accepted standard approach (Long and Ackerman, 2000). The standard approach is based on high resolution temporal data and also requires diffuse solar radiation measurements which can only be applied for well-equipped sites, such as those from BSRN. By comparing the derived fractional solar transmission of the novel approach and the standard approach we found the best agreement with the quantile  $\omega=85\%$  which is the only parameter which must be set a-priori. The estimated clear-sky shortwave radiation allows to predict the temporal variation with low bias and high correlation across the 54 sites evaluated here. These results provide confidence to apply the novel method at meteorological sites where global radiation is being measured accurately. With a much larger data coverage of available global radiation records the proposed method has the potential to provide estimates of monthly fractional clear-sky shortwave transmission and clear-sky fluxes with higher spatial and temporal coverage than before. This can be used to analyze the attenuation of solar radiation by gases and aerosols and the cloud radiative effect on shortwave radiation from a much larger observational network. These aspects are key to improve and validate remote sensing-based as well as model-calculated estimates of shortwave radiation, they can improve the planning of solar energy installations and provide observational constraints on cloud and aerosol effects relevant to the surface climate.

## Acknowledgments, Samples, and Data

M.R. acknowledges the funding by the Max-Planck society through the individual research group of A.K. M.W. and M.S. acknowledge support from the Swiss National Science Foundation grant 20002\_159938/1. We thank Maria Hakuba and Charles Long for providing the results of the Long and Ackerman method. We also thank all scientists and technicians involved in measuring and standardizing the surface radiation at all BSRN stations.

BSRN data (König-Langlo et al., 2015) can be obtained through <https://doi.pangaea.de/10.1594/PANGAEA.852720>. The data of the Long and Ackerman method can be obtained upon request from Martin Wild. The code to apply the quantile regression will be published as R package alongside with this publication.

## References

- Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., et al. (2015). Clouds, circulation and climate sensitivity. *Nature Geoscience*, 8(4), 261–268. <https://doi.org/10.1038/ngeo2398>
- Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., et al. (2013). Clouds and aerosols. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 571–657). Cambridge University Press.
- DeAngelis, A. M., Qu, X., Zelinka, M. D., & Hall, A. (2015). An observational radiative constraint on hydrologic cycle intensification. *Nature*, 528(7581), 249–253. <https://doi.org/10.1038/nature15770>
- Driemel, A., Augustine, J., Behrens, K., Colle, S., Cox, C., Cuevas-Agulló, E., et al. (2018). Baseline Surface Radiation Network (BSRN): structure and data description (1992–2017). *Earth System Science Data*, 10(3), 1491–1501. <https://doi.org/10.5194/essd-10-1491-2018>
- Dutton, E. G., Michalsky, J. J., Stoffel, T., Forgan, B. W., Hickey, J., Nelson, D. W., et al. (2001). Measurement of Broadband Diffuse Solar Irradiance Using Current Commercial Instrumentation with a Correction for Thermal Offset Errors. *Journal of Atmospheric and Oceanic Technology*, 18(3), 297–314. [https://doi.org/10.1175/1520-0426\(2001\)018<0297:MOBDSI>2.0.CO;2](https://doi.org/10.1175/1520-0426(2001)018<0297:MOBDSI>2.0.CO;2)
- Gueymard, C. A. (2008). REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation – Validation with a benchmark dataset. *Solar Energy*, 82(3), 272–285. <https://doi.org/10.1016/j.solener.2007.04.008>
- Gueymard, C. A., Bright, J. M., Lingfors, D., Habte, A., & Sengupta, M. (2019). A posteriori clear-sky identification methods in solar irradiance time series: Review and preliminary validation using sky imagers. *Renewable and Sustainable Energy Reviews*, 109, 412–427. <https://doi.org/10.1016/j.rser.2019.04.027>
- Hakuba, M. Z., Folini, D., & Wild, M. (2016). On the Zonal Near-Constancy of Fractional Solar Absorption in the Atmosphere. *Journal of Climate*, 29(9), 3423–3440. <https://doi.org/10.1175/JCLI-D-15-0277.1>
- Ineichen, P., & Perez, R. (2002). A new air mass independent formulation for the Linke turbidity coefficient. *Solar Energy*, 73(3), 151–157. [https://doi.org/10.1016/S0038-092X\(02\)00045-2](https://doi.org/10.1016/S0038-092X(02)00045-2)
- Iqbal, M. (1983). *An introduction to solar radiation*. Retrieved from <http://adsabs.harvard.edu/abs/1983itsr.book.....I>
- Kim, D., & Ramanathan, V. (2008). Solar radiation budget and radiative forcing due to aerosols and clouds. *Journal of Geophysical Research: Atmospheres*, 113(D2).

- Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511754098>
- Koenker, R., & Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94(448), 1296–1310.
- König-Langlo, G., Driemel, A., Raffel, B., & Sieger, R. (2015). BSRN snapshot 2015-09, links to zip archives [Data set]. <https://doi.org/10.1594/PANGAEA.852720>
- Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., et al. (2013). McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmospheric Measurement Techniques*, 6(9), 2403–2418.  
<https://doi.org/10.5194/amt-6-2403-2013>
- Linke, F. (1922). Transmissions-Koeffizient und Trübungsfaktor. *Beitr. Phys. Atmos.*, 10, 91–103.
- Long, C. N., & Ackerman, T. P. (2000). Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. *Journal of Geophysical Research: Atmospheres*, 105(D12), 15609–15626.  
<https://doi.org/10.1029/2000JD900077>
- Long, C. N., & Gaustad, K. L. (2004). *The shortwave (SW) clear-sky detection and fitting algorithm: Algorithm operational details and explanations*. Pacific Northwest National Laboratory.
- Long, C. N., & Shi, Y. (2008). An Automated Quality Assessment and Control Algorithm for Surface Radiation Measurements. *The Open Atmospheric Science Journal*, 2(1). Retrieved from <https://benthamopen.com/ABSTRACT/TOASCJ-2-23>
- Ohmura, A., Dutton, E. G., Forgan, B., Fröhlich, C., Gilgen, H., Hegner, H., et al. (1998). Baseline Surface Radiation Network (BSRN/WCRP): New Precision Radiometry for Climate Research. *Bulletin of the American Meteorological Society*, 79(10), 2115–2136. [https://doi.org/10.1175/1520-0477\(1998\)079<2115:BSRNBW>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<2115:BSRNBW>2.0.CO;2)
- Perez, R., Ineichen, P., Seals, R., Michalsky, J., & Stewart, R. (1990). Modeling daylight availability and irradiance components from direct and global irradiance. *Solar Energy*, 44(5), 271–289. [https://doi.org/10.1016/0038-092X\(90\)90055-H](https://doi.org/10.1016/0038-092X(90)90055-H)
- Polo, J., Zarzalejo, L. F., Martín, L., Navarro, A. A., & Marchante, R. (2009). Estimation of daily Linke turbidity factor by using global irradiance measurements at solar noon. *Solar Energy*, 83(8), 1177–1185. <https://doi.org/10.1016/j.solener.2009.01.018>
- Reno, M. J., & Hansen, C. W. (2016). Identification of periods of clear sky irradiance in time series of GHI measurements. *Renewable Energy*, 90, 520–531.  
<https://doi.org/10.1016/j.renene.2015.12.031>
- Roesch, A., Wild, M., Ohmura, A., Dutton, E. G., Long, C. N., & Zhang, T. (2011). Assessment of BSRN radiation records for the computation of monthly means. *Atmospheric Measurement Techniques*, 4(2), 339–354. <https://doi.org/10.5194/amt-4-339-2011>
- Ruiz-Arias, J., Gueymard, C., & Cebecauer, T. (2018). Direct Normal Irradiance Modeling: Evaluating the Impact on Accuracy of Worldwide Gridded Aerosol Databases. Presented at the SolarPACES conf., Casablanca, Morocco.

- Ruiz-Arias, J. A., & Gueymard, C. A. (2018). Worldwide inter-comparison of clear-sky solar radiation models: Consensus-based review of direct and global irradiance components simulated at the earth surface. *Solar Energy*, *168*, 10–29. <https://doi.org/10.1016/j.solener.2018.02.008>
- Sanchez-Lorenzo, A., Enriquez-Alonso, A., Wild, M., Trentmann, J., Vicente-Serrano, S. M., Sanchez-Romero, A., et al. (2017). Trends in downward surface solar radiation from satellites and ground observations over Europe during 1983–2010. *Remote Sensing of Environment*, *189*, 108–117. <https://doi.org/10.1016/j.rse.2016.11.018>
- Spencer, J. W. (1971). Fourier series representation of the position of the sun. *Search*, *2*(5), 172–172.
- Wild, M. (2016). Decadal changes in radiative fluxes at land and ocean surfaces and their relevance for global warming. *Wiley Interdisciplinary Reviews: Climate Change*, *7*(1), 91–107. <https://doi.org/10.1002/wcc.372>
- Wild, M., Gilgen, H., Roesch, A., Ohmura, A., Long, C. N., Dutton, E. G., et al. (2005). From Dimming to Brightening: Decadal Changes in Solar Radiation at Earth's Surface. *Science*, *308*(5723), 847–850. <https://doi.org/10.1126/science.1103215>
- Wild, M., Hakuba, M. Z., Folini, D., Dörig-Ott, P., Schär, C., Kato, S., & Long, C. N. (2019). The cloud-free global energy balance and inferred cloud radiative effects: an assessment based on direct observations and climate models. *Climate Dynamics*. <https://doi.org/10.1007/s00382-018-4413-y>
- Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., et al. (2018). Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences*, *15*(16), 5015–5030. <https://doi.org/10.5194/bg-15-5015-2018>
- Zhang, T., Stackhouse, P. W., Cox, S. J., Mikovitz, J. C., & Long, C. N. (2019). Clear-sky shortwave downward flux at the Earth's surface: Ground-based data vs. satellite-based data. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *224*, 247–260. <https://doi.org/10.1016/j.jqsrt.2018.11.015>

Table 1 BSRN sites used in the evaluation. Meta data, site average global shortwave radiation fluxes ( $R_{sd}$ ), estimated clear-sky fluxes by quantile regression method  $R_{sdcs,QR}$ , the standard method  $R_{sdcs,LA}$ , RMSE reports the root mean square error for the clear-sky fluxes. Site averages of fractional clear-sky shortwave transmission for the QR method  $\beta_{QR}$  and the Long and Ackerman method  $\beta_{LA}$ .  $R^2$  and slope reports the best fit lines for each site and the number of month  $n$  used for comparison. Columns  $f_{CRE}$  reports the fractional shortwave cloud radiative effect  $f_{CRE} = 1 - R_{sd}/R_{sd,cs}$ .

SiteCode	Site Name	Location	Lon	Lat	Shortwave fluxes [ $W m^{-2}$ ]				Fractional transmission						
					$R_{sd}$	$R_{sdcs,QR}$	$R_{sdcs,LA}$	RMSE	$\beta_{QR}$	$\beta_{LA}$	$R^2$	slope	n	$f_{CRE}$	
ALE <sup>a</sup>	Alert	Lincoln Sea, Canada	-62.4	82.5	83	99	99	10.1	0.93	0.95					0.17
ASP	Alice Spring	Australia	133.9	-23.8	254	297	289	9.7	0.85	0.83	0.83	0.89	225	0.14	
BAR	Barrow	Alaska, USA	-156.6	71.3	118	167	165	6.4	0.84	0.84	0.67	0.49	20	0.29	
BER <sup>b</sup>	Bermuda	Bermuda	-64.7	32.3	204	276	278	6.0	0.79	0.80				0.26	
BIL	Billings	Oklahoma, USA	-97.5	36.6	171	227	232	7.9	0.82	0.84	0.64	0.69	211	0.25	
BON	Bondville	Illinois, USA	-88.4	40.1	166	239	242	7.7	0.82	0.84	0.61	0.60	168	0.31	
BOS	Boulder	Colorado, USA	-105.2	40.1	188	253	252	5.0	0.87	0.87	0.89	0.74	166	0.26	
BOU	Boulder	Colorado, USA	-105.0	40.0	198	259	264	7.5	0.85	0.87	0.73	0.64	272	0.23	
BRB	Brasilia	Brazil	-47.7	-15.6	221	307	293	18.5	0.79	0.75	0.43	0.70	44	0.28	
CAB	Cabauw	The Netherlands	4.9	52.0	126	193	198	6.7	0.80	0.83	0.58	0.63	127	0.35	
CAM <sup>b</sup>	Camborne	United Kingdom	-5.3	50.2	120	190	191	5.4	0.81	0.82				0.37	
CAR	Carpentras	France	5.1	44.1	177	223	219	6.1	0.83	0.81	0.86	0.80	219	0.21	
CLH	Chesapeake L	North Atlantic Ocean, USA	-75.7	36.9	187	248	247	4.6	0.82	0.82	0.91	0.97	180	0.24	
CNR	Cener	Spain	-1.6	42.8	160	221	219	6.7	0.83	0.82	0.67	0.62	57	0.28	
COC <sup>b</sup>	Cocos Island	Cocos Islands	96.8	-12.2	243	307	309	6.3	0.78	0.79				0.21	
DAA	De Aar	South Africa	24.0	-30.7	242	287	274	17.2	0.87	0.83	0.24	0.72	49	0.16	
DAR	Darwin	Australia	130.9	-12.4	236	295	292	6.8	0.78	0.77	0.55	0.84	148	0.20	
DOM	Concordia St	Antarctica	123.4	-75.1	231	232	239	13.8	1.08	1.14	0.37	0.29	33	0.00	
DRA	Desert Rock	Nevada, USA	-116.0	36.6	242	272	267	8.7	0.87	0.86	0.67	0.49	134	0.11	
E13	Southern Gre	Oklahoma, USA	-97.5	36.6	184	241	241	5.3	0.82	0.83	0.82	0.68	193	0.24	
EUR	Eureka	Canada	-85.9	80.0	96	121	114	15.3	0.93	0.84	0.45	0.41	25	0.21	
FLO	Florianopoli	Brazil	-48.5	-27.5	177	271	268	6.5	0.80	0.79	0.53	0.80	65	0.35	
FPE	Fort Peck	Montana, USA	-105.1	48.3	164	220	221	7.7	0.84	0.86	0.72	0.55	167	0.25	
FUA	Fukuoka	Japan	130.4	33.6	157	251	258	8.2	0.77	0.80	0.68	0.69	62	0.38	
GCR	Goodwin Cree	Mississippi, USA	-89.9	34.2	182	254	254	7.0	0.80	0.80	0.63	0.63	168	0.28	
GVN	Georg von Ne	Antarctica	-8.2	-70.7	169	210	209	6.2	0.98	0.99	0.75	0.62	201	0.19	
ILO <sup>a</sup>	Ilorin	Nigeria	4.6	8.5	199	264	288	29.2	0.68	0.74				0.25	
ISH <sup>b</sup>	Ishigakijima	Japan	124.2	24.3	181	272	277	9.2	0.76	0.77				0.34	
IZA	Izana	Tenerife, Spain	-16.5	28.3	290	317	315	3.9	0.92	0.91	0.89	1.03	76	0.09	
KWA <sup>b</sup>	Kwajalein	North Pacific Ocean	167.7	8.7	229	295	297	8.3	0.77	0.77				0.22	
LAU	Lauder	New Zealand	169.7	-45.0	170	241	239	6.5	0.86	0.85	0.61	1.06	174	0.30	
LER <sup>b</sup>	Lerwick	Shetland Island, UK	-1.2	60.1	43	74	78	4.4	0.79	0.86				0.42	
LIN	Lindenberg	Germany	14.1	52.2	118	180	183	6.9	0.80	0.83	0.42	0.50	151	0.34	
MAN <sup>b</sup>	Momote	Papua New Guinea	147.4	-2.1	207	303	300	5.8	0.78	0.77				0.32	
MNM	Minamitorish	Japan	154.0	24.3	229	281	283	3.5	0.78	0.79	0.68	0.77	64	0.19	
NAU <sup>b</sup>	Nauru Island	Nauru	166.9	-0.5	243	304	303	6.6	0.78	0.77				0.20	
NYA	Ny-Alesund	Spitsbergen	11.9	78.9	52	69	71	3.1	0.90	0.92	0.64	0.70	63	0.25	
PAL	Palaiseau	France	2.2	48.7	142	213	216	6.4	0.80	0.82	0.33	0.47	117	0.33	
PAY	Payerne	Switzerland	6.9	46.8	138	205	207	5.8	0.81	0.82	0.31	0.42	222	0.33	
PSU	Rock Springs	Pennsylvania, USA	-77.9	40.7	157	240	243	7.1	0.82	0.84	0.63	0.57	130	0.35	
PTR	Petrolina	Brazil	-40.3	-9.1	244	297	301	9.6	0.74	0.75	0.04	0.15	53	0.18	
REG	Regina	Canada	-104.7	50.2	164	220	217	5.3	0.89	0.88	0.93	0.99	190	0.25	
SAP	Sapporo	Japan	141.3	43.1	148	233	236	6.7	0.81	0.82	0.70	0.85	63	0.36	
SBO	Sede Boqer	Israel	34.8	30.9	240	265	262	8.0	0.81	0.80	0.52	0.52	113	0.10	
SMS	Sao Martinho	Brazil	-53.8	-29.4	207	288	276	12.5	0.83	0.80	0.77	0.80	26	0.28	
SOV <sup>a</sup>	Solar Villag	Saudi Arabia	46.4	24.9	252	275	270	9.2	0.81	0.80				0.08	

SiteCode	Site Name	Location	Lon	Lat	Shortwave fluxes [ $W m^{-2}$ ]				Fractional transmission							
					$R_{sd}$	$R_{sdc,QR}$	$R_{sdc,LA}$	RMSE	$\beta_{QR}$	$\beta_{LA}$	$R^2$	slope	n	$f_{CRE}$		
SPO <sup>b</sup>	South Pole	Antarctica	-24.8	-90.0	198	216	211	11.5	1.07	1.36						0.08
SXF	Sioux Falls	South Dakota, USA	-96.6	43.7	168	230	232	6.2	0.83	0.85	0.79	0.58	71	0.27		
SYO	Syowa	Antarctica	39.6	-69.0	153	192	191	6.1	0.97	0.97	0.40	0.49	201	0.20		
TAM	Tamanrasset	Algeria	5.5	22.8	275	312	300	15.1	0.87	0.84	0.59	0.78	181	0.12		
TAT	Tateno	Japan	140.1	36.0	158	238	247	12.5	0.80	0.82	0.70	1.23	222	0.34		
TIK	Tiksi	Siberia, Russia	128.9	71.6	74	97	92	5.5	0.93	0.87	0.56	0.68	21	0.24		
TOR	Tõravere	Estonia	26.5	58.3	115	174	176	5.6	0.82	0.85	0.31	0.54	197	0.34		
XIA	Xianghe	China	117.0	39.8	168	236	247	21.1	0.80	0.83	0.39	1.03	112	0.29		
<b>Mean</b>					<b>179</b>	<b>235</b>	<b>235</b>	<b>8.5</b>	<b>0.84</b>	<b>0.85</b>	<b>0.61</b>	<b>0.68</b>	<b>128</b>	<b>0.25</b>		

<sup>a</sup> Long and Ackerman flagged doubtful

<sup>b</sup> Long and Ackerman approach in climatological mode

Accepted Article

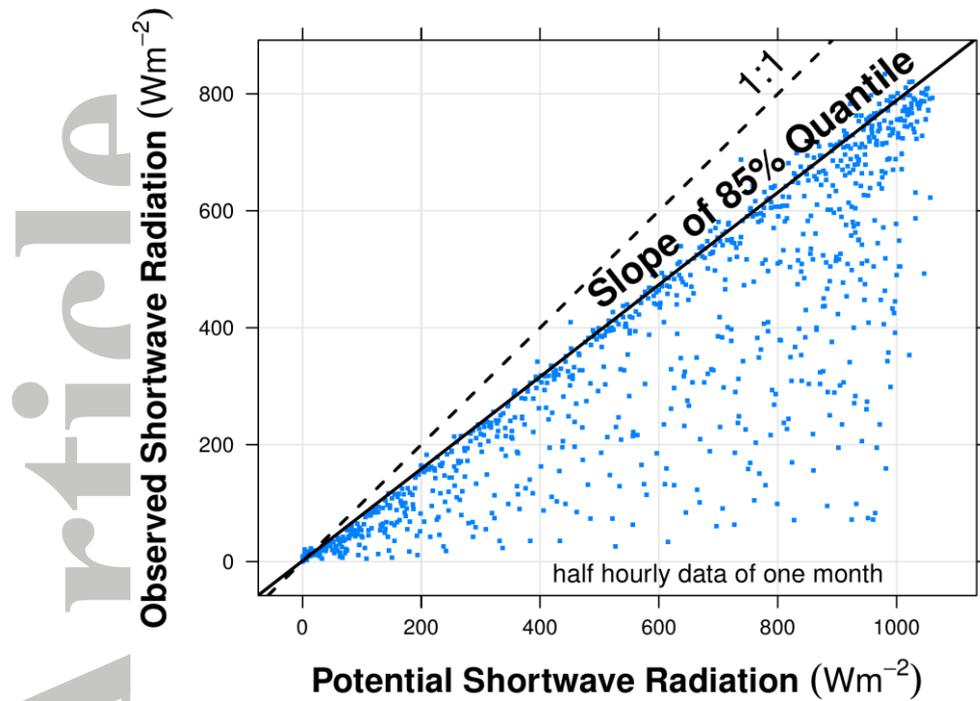


Figure 1 Illustration of Quantile Regression approach. The scatterplot shows a monthly sample of 30min observations of incoming shortwave radiation versus the corresponding potential shortwave radiation (Lindenberg, Germany, August 2003). The scatter forms a well-defined orthogonal triangle with clear-sky conditions close to the upper boundary. The quantile regression method (Koenker, 2005 ) allows to estimate this upper slope for a given quantile  $\omega$  (here 85%).

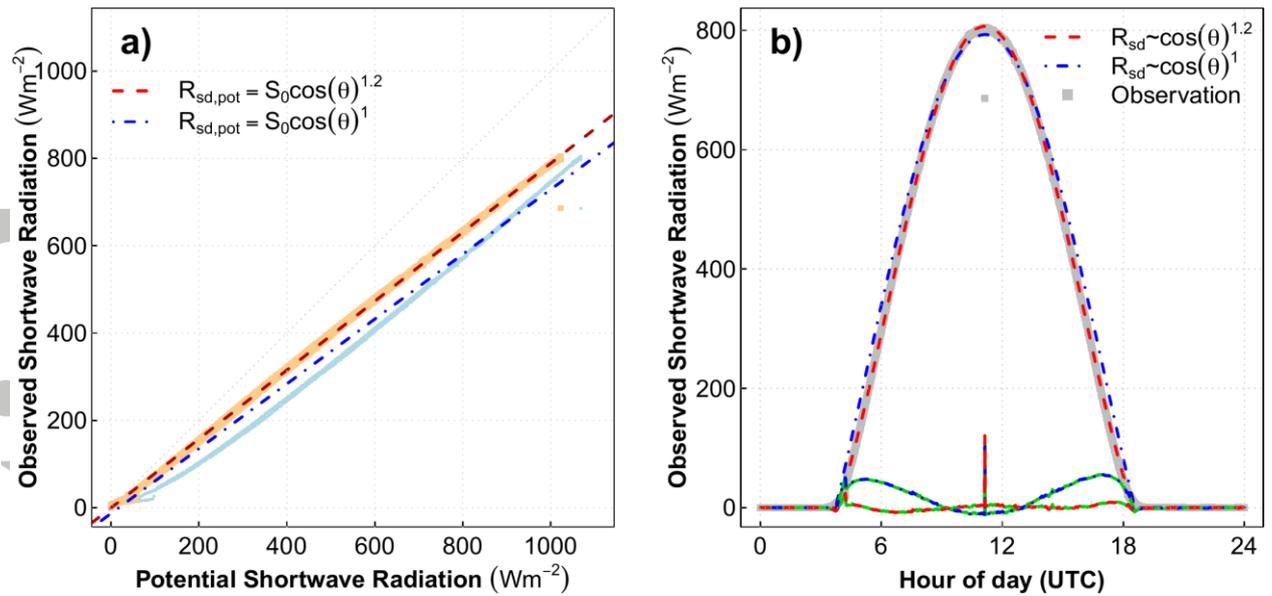


Figure 2 Illustration of time dependence of solar attenuation on potential solar radiation. Panel a) shows the relationship of observed solar radiation to potential radiation in 1 minute resolution. Light blue data shows a small curvature when potential radiation is modelled without accounting for air mass dependent attenuation (exponent  $b = 1$ ) resulting in deviations from linearity at lower potential solar radiation (morning and evening). The light red data points are derived with setting the exponent  $b = 1.2$  improving the linear relationship. Panel (b) shows the incoming solar radiation vs. time. Grey are observations and the dashed lines represent linear regression fits of observed to potential solar radiation for that day. The potential solar radiation with exponent  $b = 1.2$  clearly shows a better match of the observed curve with lower deviations in morning and afternoon (dashed lines with green background). The plot utilizes observations from site Lindenberg (LIN) in Germany measured at the 11<sup>th</sup> August 2003 with 1 min resolution.

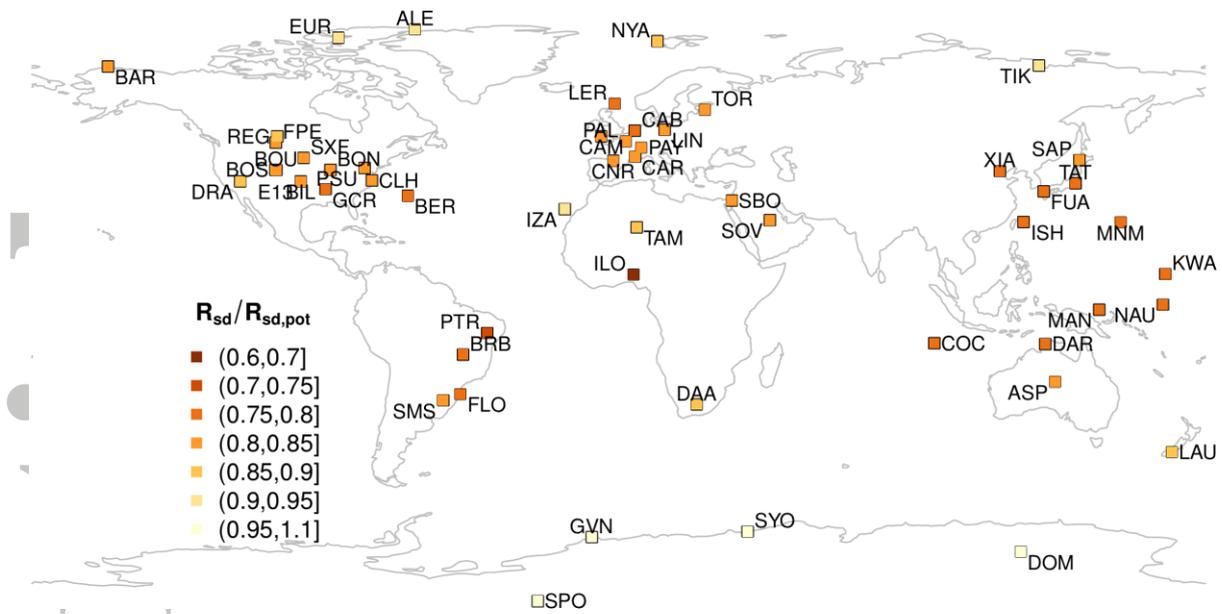


Figure 3 Map showing the location of all BSRN sites analyzed in this study. The color indicates the estimated site average value of the fractional clear-sky shortwave transmission determined by the quantile regression approach.

Accepted A

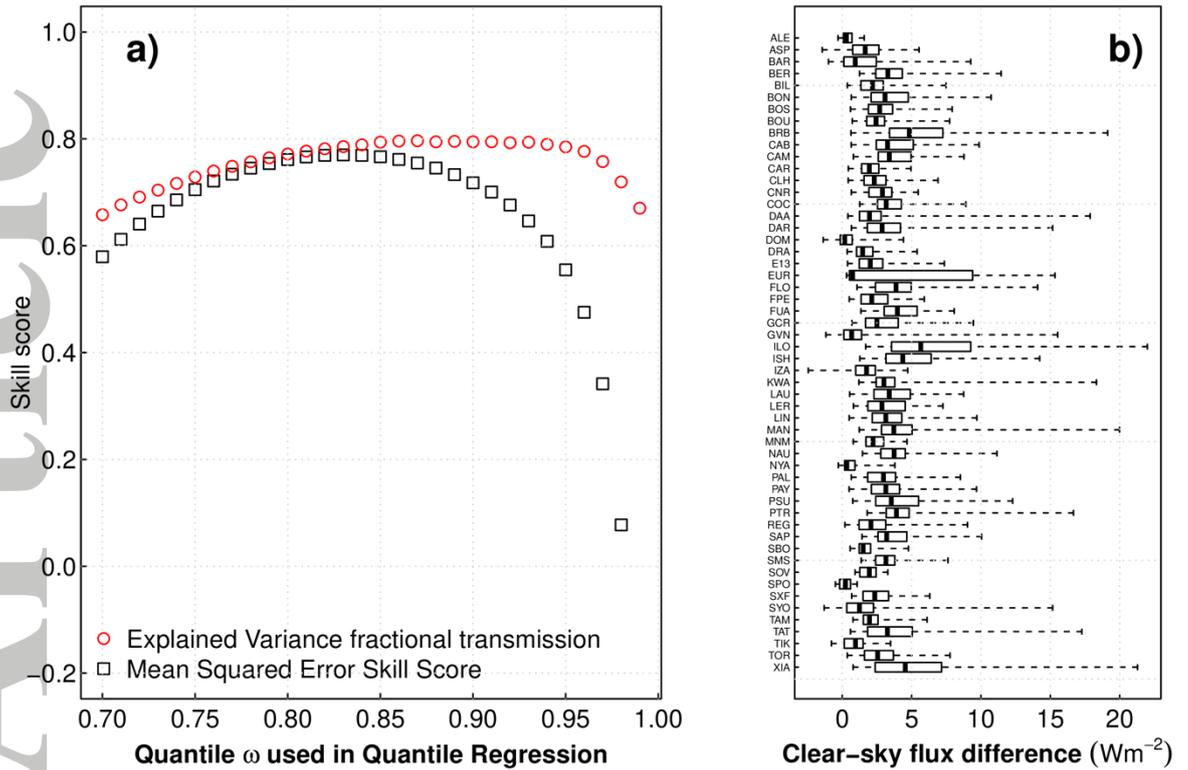


Figure 4 Sensitivity of the quantile regression method to the choice of the quantile. Panel a) Skill scores across 42 sites and available months using the estimates of the Long and Ackerman (2000) method for evaluation. Red shows the explained variance of the fractional solar transmission and black shows the skill score with a constant fractional solar transmission as reference. Panel b) shows the difference in the monthly estimate of the clear-sky shortwave flux when the quantile  $\omega = 90\%$  is compared to  $\omega = 85\%$ . The sensitivities only cover months where the quantile regressions yielded a  $R^2 > 0.75$ .

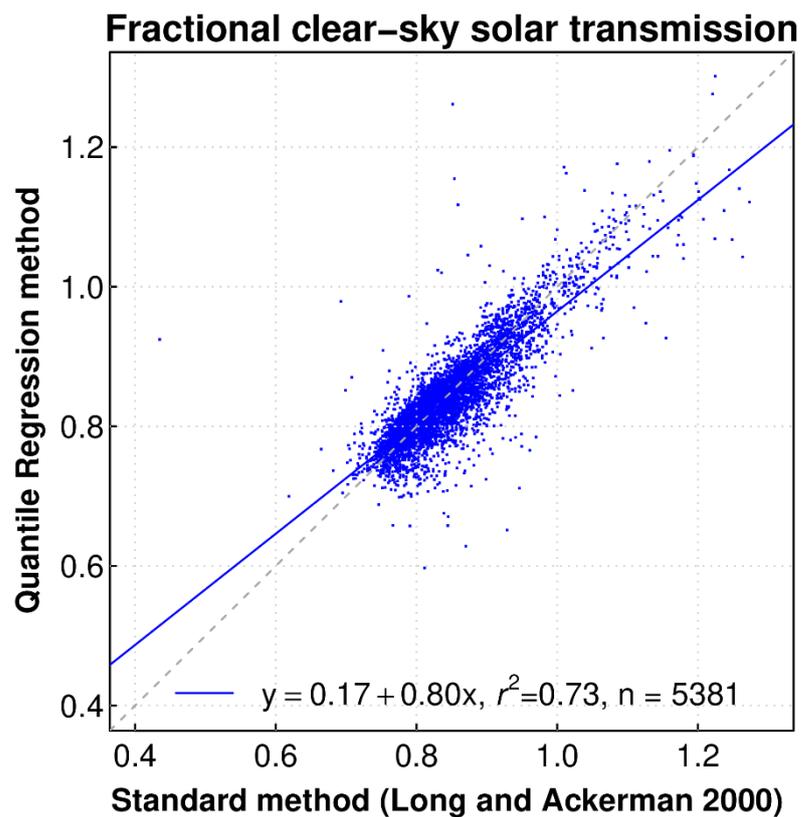


Figure 5 Comparison of monthly fractional clear-sky solar transmission obtained with the proposed QR method and the Long and Ackerman method. The blue line shows the linear best fit and the dashed line shows the 1-1 perfect fit.

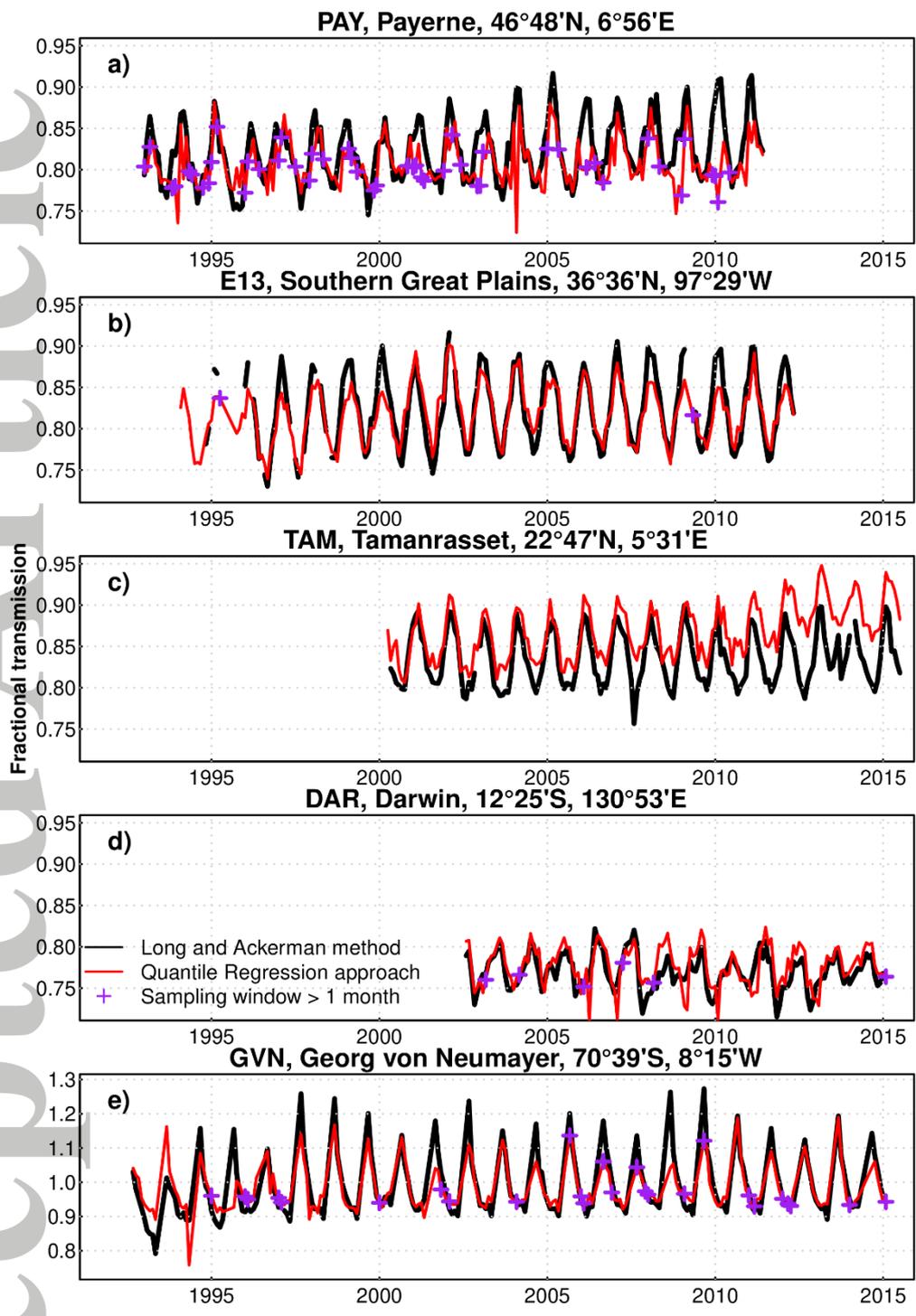


Figure 6 Monthly mean time series of fractional clear-sky shortwave transmission estimated by the Long and Ackerman method (black) and the QR method (red) for five selected sites with long records. Purple crosses show months where the sampling size of cloud free periods of that month was insufficient to estimate a reliable fractional clear-sky transmission and therefore a large sampling window was used for that month (see Methods section).

## Monthly mean Clear Sky Solar Radiation

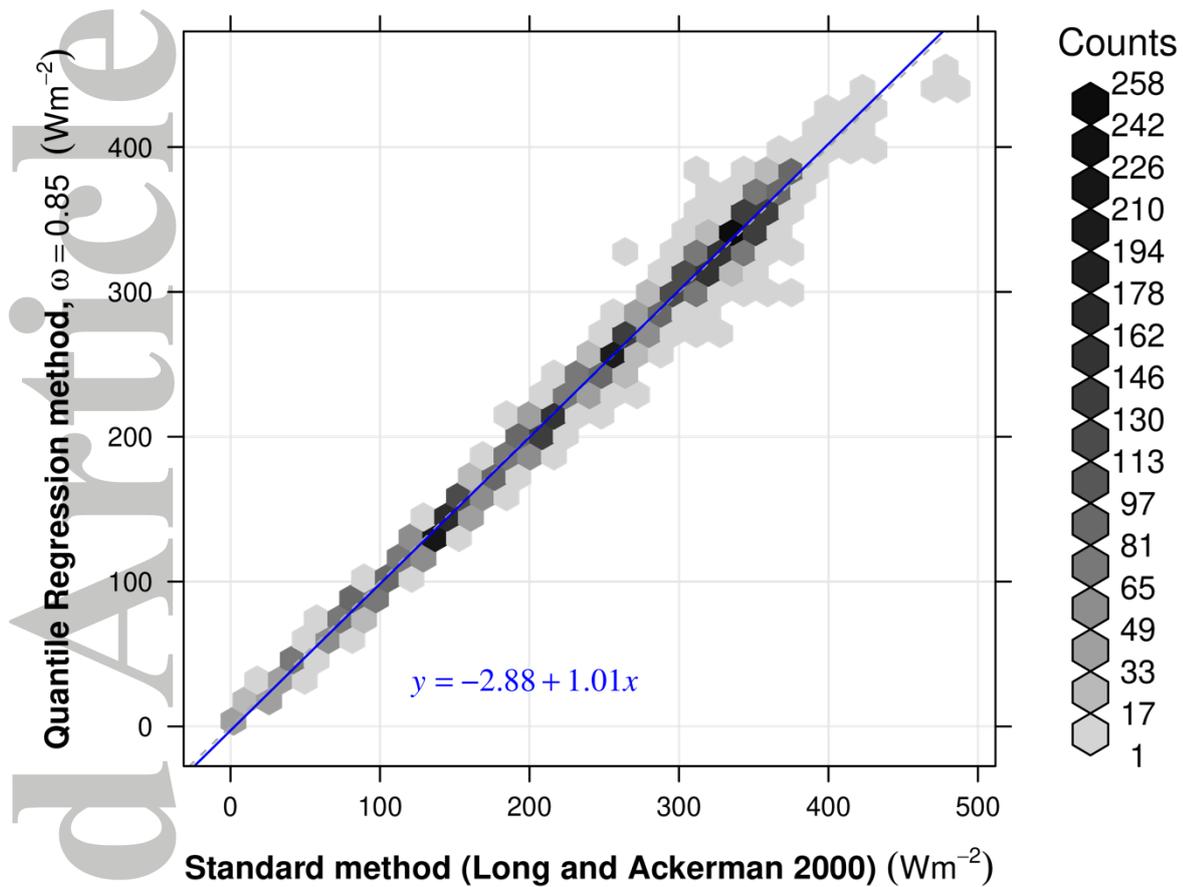


Figure 7 Comparison of monthly mean clear-sky shortwave fluxes derived by the proposed QR method and the Long and Ackerman method for 42 sites with reliable estimates of the Long and Ackerman method. The scatter is binned into hexagons to visualize the high point density shown by the grey color legend on the left. The dashed line shows the 1-1 perfect fit and the blue line the linear best fit with statistics annotated in blue text.

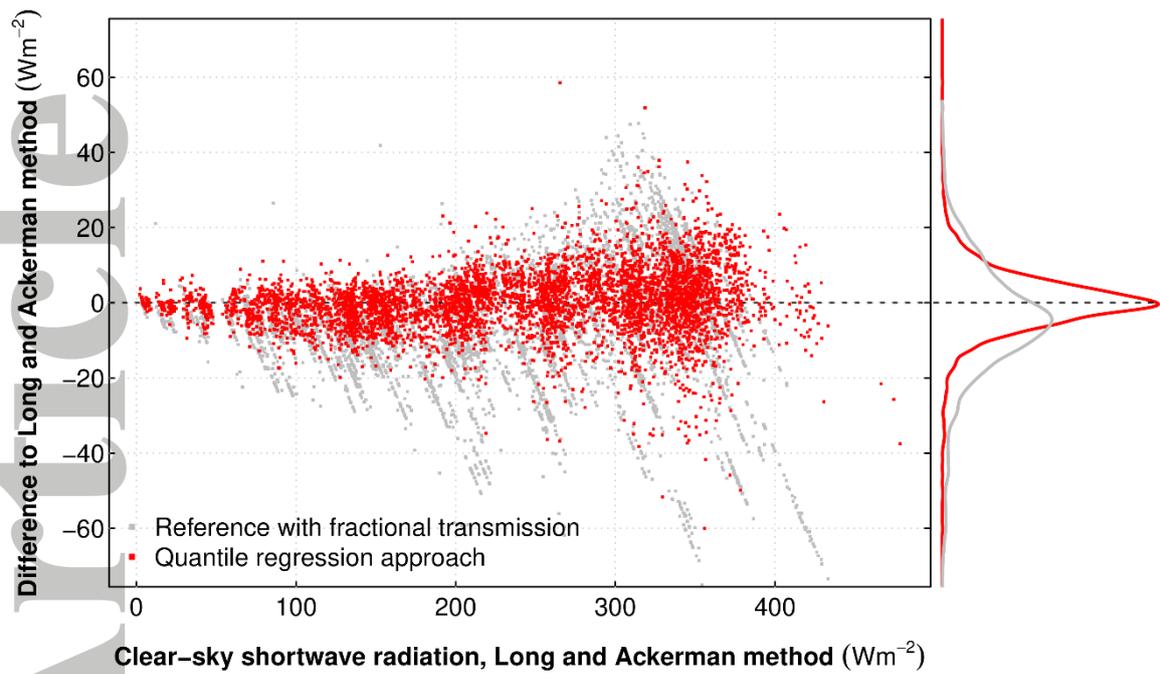


Figure 8 Difference of monthly mean clear-sky fluxes to the standard method for all sites and months. Grey dots show a reference estimate which simply estimates clear-sky solar radiation by  $0.81 * R_{sd,pot}$ . Red dots show the estimate with time varying fractional solar transmission by quantile regression. The panel on the right shows the marginal distributions of the residuals, obtained by kernel density estimation.