Assessment of ERA5 data for solar resource mapping

2019-08-19

## Introduction

The project is structured as follows:

* *M* folder containing the RMarkdown source code and HTML and DOCX files converted from Rmd file;
* *RDate* folder where R objects are stored;
* *tabs* folder which contain metadata of the stations and the solar global radiation data extracted from ERA5 model and BSRN network;
* *sh* folder with the ksh scripts run on the ecgate cluster;
* *R* folder with R language script.

## Input dataset

* ERA5 Surface solar radiation downwards [J/m2] hourly accumulate parameter will be used as the main dataset in this work:
* The reference dataset will consist of solar radiation measurements obtained from the Baseline Surface Radiation Network (BSRN - <https://bsrn.awi.de/>).

### Baseline Surface Radiation Network

* consist of 24 radiometric stations located in in contrasting climatic zones;
* solar and atmospheric radiation measurments;
* equipped with instruments of the highest available accuracy and with high time resolution;
* the BSRN datasets is retrievable via PANGAEA portal (<https://bsrn.awi.de/data/data-retrieval-via-pangaea/>);
* hourly data from the stations that have observations available between 2006 to 2015 (Figure 1, Table 1);
* were considered only the days which have at least 90 % of available radiation data for each day.

The sub-hourly raw values were averaged to daily values, after they were quality checked using the algorithm Solarpos without refraction as implemented in BSRN Toolbox (Schmithüsen, Sieger, and König-Langlo 2012).

library(mapview)  
library(sp)  
library(knitr)  
rad.network <- read.table("../tabs/rad\_network.csv", sep = ";", header = T)  
kable(rad.network, caption = rad.meta)

Table 1: Metadata of the selected radiometric stations

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| id | name | latitude | longitude | elevation | country | continent | kg\_class | kg\_code |
| CAR | Carpentras | 44.08300 | 5.05900 | 100 | France | Europe | Csa | 12 |
| CLH | Chesapeake Light | 36.90500 | 75.71300 | 37 | USA | North America | ET | 31 |
| DAR | Darwin | -12.42500 | 130.89100 | 30 | Australia | Australia | Aw | 4 |
| DRA | Desert Rock | 36.62600 | -116.01800 | 1007 | USA | North America | BWk | 8 |
| BON | Bondville | 40.06667 | -88.36667 | 213 | USA | North America | Cfa | 9 |
| BOS | Boulder | 40.12500 | -105.23700 | 1689 | USA | North America | BSk | 6 |
| BRB | Brasilia | -15.60100 | -47.71300 | 1023 | Brazil | South America | Aw | 4 |
| CAB | Cabauw | 51.97110 | 4.92670 | 0 | The Netherlands | Europe | Cfb | 10 |
| CNR | Cener | 42.81600 | -1.60100 | 471 | Spain | Europe | Cfb | 10 |
| E13 | Southern Great Plains | 36.60500 | -97.48500 | 318 | USA | North America | Cfa | 9 |
| FPE | Fort Peck | 48.31667 | -105.10000 | 634 | USA | North America | BSk | 6 |
| FUA | Fukuoka | 33.58220 | 130.37640 | 3 | Japan | Asia | Cfa | 9 |
| GCR | Goodwin Creek | 34.25470 | -89.87290 | 98 | USA | North America | Cfa | 9 |
| GVN | Georg von Neumayer | -70.65000 | -8.25000 | 42 | Dronning Maud Land | Antarctica | Ocean | 32 |
| KWA | Kwajalein | 8.72000 | 167.73100 | 10 | Marshall Islands | Oceania | Ocean | 32 |
| LER | Lerwick | 60.13891 | -1.18466 | 80 | United\_Kingdom | Europe | Cfb | 10 |
| NYA | Ny-?lesund | 78.92500 | 11.93000 | 11 | Spitsbergen | Europe | ET | 31 |
| PAL | Palaiseau, SIRTA Observatory | 48.71300 | 2.20800 | 156 | France | Europe | Cfb | 10 |
| PSU | Rock Springs | 40.72000 | -77.93333 | 376 | USA | North America | Cfb | 10 |
| PTR | Petrolina | -9.06800 | -40.31900 | 387 | Brazil | South America | BSh | 5 |
| SMS | S?o Martinho da Serra | -29.44278 | -53.82305 | 489 | Brazil | South America | Cfa | 9 |
| SPO | South Pole | -89.98300 | -24.79900 | 2800 | South Pole | Antarctica | EF | 30 |
| SYO | Syowa | -69.00500 | 39.58900 | 18 | Cosmonaut Sea | Antarctica | EF | 30 |
| TAT | Tateno | 36.05810 | 140.12580 | 25 | Japan | Asia | Cfa | 9 |

coordinates(rad.network) <- ~longitude + latitude  
proj4string(rad.network) <- "+init=epsg:4326"  
mapview(rad.network, layer.name = 'Radiometric network')

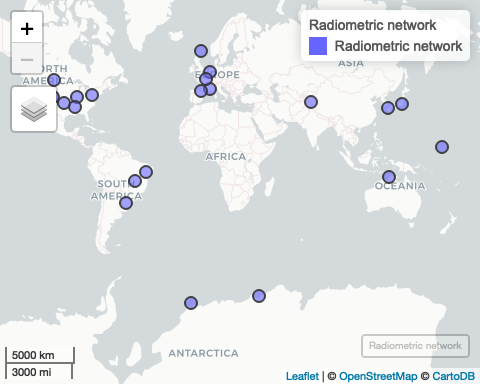


Figure 1: Selected radiometric staions

library(lattice)  
tab <- read.table("../tabs/rad\_network.csv", sep = ";", header = T)  
tab <- tab[order(tab$latitude, tab$elevation),]  
tabf <- NULL  
for (i in 1:nrow(tab)) {  
   
 tab.i <- read.csv(paste0("../tabs/daily\_bsrn/", tab$id[i],"\_2006\_2015.csv"))  
 tab.i$id <- tab$id[i]  
 tabf <- rbind(tabf, tab.i)  
}  
  
tabf$id <- factor(tabf$id, levels = unique(tabf$id))  
  
bwplot(id~ghi, data = tabf, xlab = "GHI (W/m^2)",col = "red",  
 par.settings = list(box.rectangle = list(col = "salmon",fill = "salmon",alpha = 0.4),  
 box.umbrella = list(col = "salmon",alpha = 0.4),  
 plot.symbol = list(col = "salmon",alpha = 0.4, pch = 20)))

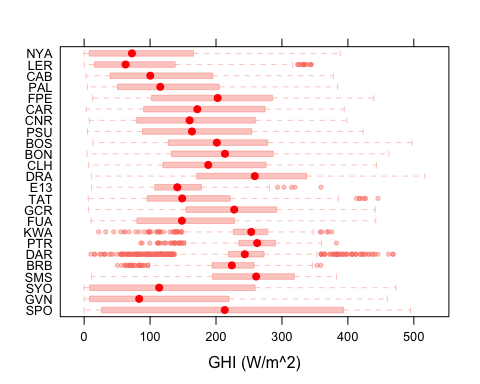


Figure 2: Box-and-whisker plots of the daily GHI (W/m²) obtained from BSRN archives (2006-2015). The filled dots denote the median, solid boxes range from the lower to the upper quartile, and dashed whiskers show the data range. Data that are further than 1.5 times the interquartile range from the nearest quartile are shown as open bullets. The stations are ordered by latitude and altitude.

### ERA5 dataset

* the latest global reanalysis product, produced by ECMWF;
* provide hourly meteorological data from 1979 onwards;
* spatial resolution approximate 30 km grid;
* open access and free to download for all uses, including commercial use through the C3S Climate Data Store (<https://cds.climate.copernicus.eu>);
* more info: <https://climate.copernicus.eu/climate-reanalysis>.
* two subset used:
  + hourly data obtained directly from the ECMWF Meteorological Archival and Retrieval System (MARS), via ecgate cluster;
  + monthly data from C3S Climate Data Store.

Due to a large amount of data necessary to be processed at the validation stage, hourly ERA5 data was extracted for the years 2006 to 2015 via ecgate cluster (<https://www.ecmwf.int/en/computing/our-facilities/ecgate>) by using a retrieve script generated from the links available on the Mars Catalogue:

#!/bin/ksh -x  
#SBATCH --qos=normal  
#SBATCH --job-name=get\_era5  
#SBATCH --output=/home/ms/ro/roq/scripts/era5/logs/get\_era5%N.%j.out  
#SBATCH --error=/home/ms/ro/roq/scripts/era5/logs/get\_era5%N.%j.out  
#SBATCH --workdir=/scratch/ms/ro/roq/radiation  
#SBATCH --time=23:30:00  
  
export PATH=$PATH:.   
set -xv  
cd $SCRATCH/radiation  
  
f\_marsReq=marsreq.tmp  
f\_logMars=marsreq.log  
  
cat >${f\_marsReq} <<EOF  
retrieve,  
class=ea,  
date=2005-12-30/to/2016-01-01,  
expver=1,  
levtype=sfc,  
param=169.128,  
stream=oper,  
time=06:00:00/18:00:00,  
type=fc,  
step=1/2/3/4/5/6/7/8/9/10/11/12,  
grid=0.28125/0.28125,  
target="ghi.grb"  
EOF  
#-- Perform MARS request in one go  
export MARS\_MULTITARGET\_STRICT\_FORMAT=1  
mars < ${f\_marsReq} > ${f\_logMars}

The hourly data was agggregated to daily data and save as NetCDF file using the Climate Data Operators (CDO - <https://www.mpimet.mpg.de/cdo/>) tool (Kaspar, Schulzweida, and Muller 2010). CDO was developed at the Max-Planck-Institute for Meteorology. It is open source and released under the terms of the GNU General Public License v2 (GPL). The source code is available at <https://code.zmaw.de/projects/cdo>.

The computation was performed on the same ecgate cluster.

#!/bin/ksh -x  
#SBATCH --qos=normal  
#SBATCH --job-name=ghi\_daily  
#SBATCH --output=/home/ms/ro/roq/scripts/era5/logs/to\_ghi\_daily%N.%j.out  
#SBATCH --error=/home/ms/ro/roq/scripts/era5/logs/to\_ghi\_daily%N.%j.out  
#SBATCH --workdir=/scratch/ms/ro/roq/radiation  
#SBATCH --time=23:30:00  
  
export PATH=$PATH:.   
set -xv  
  
cd $SCRATCH/radiation  
  
module load cdo  
cdo setgridtype,regular ghi.grb output\_ghi.grb  
cdo -f nc -t ecmwf setgridtype,regular output\_ghi.grb output\_ghi.nc  
rm output\_ghi.grb  
cdo shifttime,-1hour output\_ghi.nc ghi\_shift1h.nc  
rm output\_ghi.nc  
cdo daymean ghi\_shift1h.nc ghi\_daily.nc  
rm ghi\_shift1h.nc

The functions implemented in the CDO tool were used to iteratively extract the daily radiation data at each radiometric station location. The example below shows how the radiation data was obtained from ERA5 at Carpentras radiometric station coordinates:

cdo -outputtab,lon,lat,date,time,value -remapnn,lon=5.05900\_lat=44.08300 /scratch/ms/ro/roq/radiation/output\_ghi.nc > CAR\_ERA5.txt

## Validation of ERA5 data

The comparison between ERA5 data and reference (BSRN network) dataset was performed for the entire dataset, as well as for the subsets of data by seasons and Köppen-Geiger climate classes, using the well-known indicators of accuracy (bias, root mean squared error, root relative squared error), matrix of scatterplots, the summary statistics (mean, standard deviation, range, etc.). Existing R language tools (e.g. libraries Metrics, openair, etc) were used in analyzing the accuracy of the reanalysis product.

For all the stations, the Köppen-Geiger climate classification metadata was obtained from a high spatial resolution map of 5 arc minutes, representative for the 25-year, period 1986-2010 (Rubel et al. 2017).

The two data sets were merged by using *merge()* function from the R base package, by using field *dates* as common columns. In order to preserve data structures (columns data types), the new dataset was saved as compressed R object:

tab <- read.table("../tabs/rad\_network.csv", sep = ";", header = T)  
  
tabf <- NULL  
for (i in 1:nrow(tab)) {  
   
 era5 <- read.csv(paste0("../tabs/daily\_era5/", tab$id[i], "\_2006\_2015.csv"))  
 era5$date <- as.Date(era5$date)  
   
 bsrn <- read.csv(paste0("../tabs/daily\_bsrn/", tab$id[i], "\_2006\_2015.csv"))  
 bsrn$date <- as.Date(as.character(bsrn$date), format = "%Y%m%d")  
   
 tab.i <- merge(era5, bsrn, by.x.y = "date")  
 names(tab.i)[3] <- "bsrn\_ghi"  
 # add id of each station  
 tab.i$id <- tab$id[i]  
 # add koppen geiger class and code  
 tab.i$kg\_class <- tab$kg\_class[i]  
 tab.i$kg\_code <- tab$kg\_code[i]  
 # add hemisphere  
 tab.i$hemisphere <- ifelse(tab$latitude[i] > 0, "Northern", "Southern")  
   
   
 tabf <- rbind(tabf, tab.i)  
}  
  
saveRDS(tabf, file = "../RData/ghi\_bsrn\_era5.rds")

First, the mean bias error (MBE), mean absolute error (MAE), and the root relative squared error (RRSE) were used as performance statistics indicators for the ERA4 estimates. The metrics were computed with the functions implemented in R *Metrics* library (Hamner, Frasco, and LeDell 2018).

The indicators were computed for all the station by Köppen-Geiger climate classification zones and by seasons. The performance metrics of the reanalysis products are summarized in Table 2 and Table 3. Overall, the estimates reported by ERA5 agree with data obtained from BSRN stations. As regarding these indicators, we can notice that the model performed best for BW climate zones, and worst for the tropical climate zones. As concerning the performance of the model by season, there are some significant differences between cold and warm seasons, in the southern hemisphere. The summary statistics computed for both datasets reveal us that generally, ERA5 model tends to underestimate the extreme values as well as the variability of the analyzed parameter, regardless of the climate zone (Table 4).

By looking at the scatterplots and R-squared (Figure 3), some differences between the climate zones may be observed, but the agreement between the two analyzed datasets is quite reasonable . The matrix of scatterplots was constructed with the help of the *scatterPlot()* function form the openair R package (Carslaw and Ropkins 2012).

library(seas, quietly = T)  
library(Metrics, quietly = T)  
library(dplyr, quietly = T, warn.conflicts = F)  
library(knitr, quietly = T)  
library(moments)  
library(openair)  
  
# read ERA5 and BSRN date  
tab <- readRDS("../RData/ghi\_bsrn\_era5.rds")  
  
  
# extract KG main climate groups  
tab$KG <- substr(tab$kg\_class, 1,2)  
  
# compute seasons  
tab$season <- mkseas(x = tab, width = "DJF")  
  
  
# compute indiccators by kg classes  
indicators.kg <- tab %>% group\_by(KG) %>%   
 summarise(ME = round(bias(bsrn\_ghi, era5\_ghi), 3),   
 MAE = round(mae(bsrn\_ghi, era5\_ghi), 3),  
 RRSE = round(rrse(bsrn\_ghi, era5\_ghi), 3))  
  
# transpose table  
ti.kg <- t(indicators.kg[,2:4])  
colnames(ti.kg) <- indicators.kg$KG  
  
kable(ti.kg ,caption = kg.ind)

Table 2: Summary of the validation results by KÖPPEN-GEIGER climate classification zones

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Aw | BS | BW | Cf | Cs | EF | ET | Oc |
| ME | -5.115 | -2.595 | 3.665 | -1.151 | 1.021 | -3.172 | -5.580 | -2.953 |
| MAE | 26.647 | 21.875 | 12.754 | 21.079 | 18.267 | 17.026 | 20.515 | 16.554 |
| RRSE | 0.733 | 0.361 | 0.257 | 0.313 | 0.264 | 0.190 | 0.300 | 0.222 |

# compute indiccators by seasons  
# compute indicators by seasons  
indicators.seas <- tab %>% group\_by(season, hemisphere) %>%   
 summarise(ME = round(bias(bsrn\_ghi, era5\_ghi), 3),   
 MAE = round(mae(bsrn\_ghi, era5\_ghi), 3),  
 RRSE = round(rrse(bsrn\_ghi, era5\_ghi), 3))  
  
# transpose table  
ti.seas <- t(indicators.seas[,2:5])  
colnames(ti.seas) <- indicators.seas$season  
  
kable(ti.seas,caption = seas.ind)

Table 3: Summary of the validation results by seasons and hemispheres

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | DJF | DJF | MAM | MAM | JJA | JJA | SON | SON |
| hemisphere | Northern | Southern | Northern | Southern | Northern | Southern | Northern | Southern |
| ME | -3.478 | -4.216 | -5.017 | -0.986 | 0.848 | 2.328 | 0.013 | -4.390 |
| MAE | 12.124 | 32.988 | 24.168 | 15.084 | 28.327 | 9.244 | 15.127 | 22.520 |
| RRSE | 0.307 | 0.473 | 0.372 | 0.250 | 0.448 | 0.180 | 0.286 | 0.347 |

# compute summary statistics by KG classes  
tab1 <- data.frame(tab[,c(1,2,4,5,6,7,8)], source = "ERA5")  
names(tab1)[2] <- "GHI"  
tab2 <- data.frame(tab[,c(1,3,4,5,6,7,8)], source = "BSRN")  
names(tab2)[2] <- "GHI"  
tabs <- rbind(tab1, tab2)  
  
  
stats <- tabs %>% group\_by(KG, source) %>% summarise(Mean = mean(GHI), Median = median(GHI), Sd = sd(GHI),Cv = (sd(GHI)/mean(GHI)) \* 100, Max = max(GHI), Skewness = skewness(GHI), Kurtosis = kurtosis(GHI))  
# round values  
stats[,3:8] <- round(stats[,3:8], 3)   
kable(stats, caption = sum.ind)

Table 4: Summary statistic for ERA5 estimates and BSRN measurements

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| KG | source | Mean | Median | Sd | Cv | Max | Skewness | Kurtosis |
| Aw | ERA5 | 241.189 | 245.733 | 48.419 | 20.075 | 351.684 | -0.933 | 4.201610 |
| Aw | BSRN | 236.074 | 238.913 | 54.381 | 23.035 | 468.882 | -0.460 | 4.818329 |
| BS | ERA5 | 218.318 | 231.384 | 84.913 | 38.894 | 377.762 | -0.278 | 1.978070 |
| BS | BSRN | 215.723 | 230.485 | 89.197 | 41.348 | 497.368 | -0.241 | 2.178375 |
| BW | ERA5 | 251.803 | 259.752 | 85.641 | 34.011 | 388.109 | -0.137 | 1.677781 |
| BW | BSRN | 255.469 | 259.847 | 90.669 | 35.491 | 516.981 | -0.039 | 1.907632 |
| Cf | ERA5 | 159.882 | 155.224 | 90.973 | 56.900 | 387.304 | 0.131 | 1.994010 |
| Cf | BSRN | 158.732 | 150.740 | 98.178 | 61.852 | 462.240 | 0.242 | 1.990386 |
| Cs | ERA5 | 181.881 | 180.063 | 94.989 | 52.226 | 352.318 | 0.045 | 1.710654 |
| Cs | BSRN | 182.902 | 172.692 | 101.640 | 55.571 | 394.446 | 0.112 | 1.735100 |
| EF | ERA5 | 172.518 | 145.569 | 154.106 | 89.328 | 482.081 | 0.388 | 1.766887 |
| EF | BSRN | 169.346 | 139.506 | 154.278 | 91.102 | 494.991 | 0.466 | 1.882367 |
| ET | ERA5 | 154.659 | 150.689 | 104.104 | 67.312 | 365.841 | 0.089 | 1.838572 |
| ET | BSRN | 149.079 | 139.321 | 107.655 | 72.214 | 443.290 | 0.269 | 1.954156 |
| Oc | ERA5 | 151.290 | 148.049 | 124.955 | 82.593 | 432.311 | 0.267 | 1.816112 |
| Oc | BSRN | 148.336 | 142.151 | 123.444 | 83.219 | 459.787 | 0.303 | 1.868810 |

scatterPlot(tab, y = "era5\_ghi", x = "bsrn\_ghi", type = "KG", xlab = "BSRN GHI [W/m²]", ylab = "ERA5 GHI [W/m²]", linear = T, smooth = T)

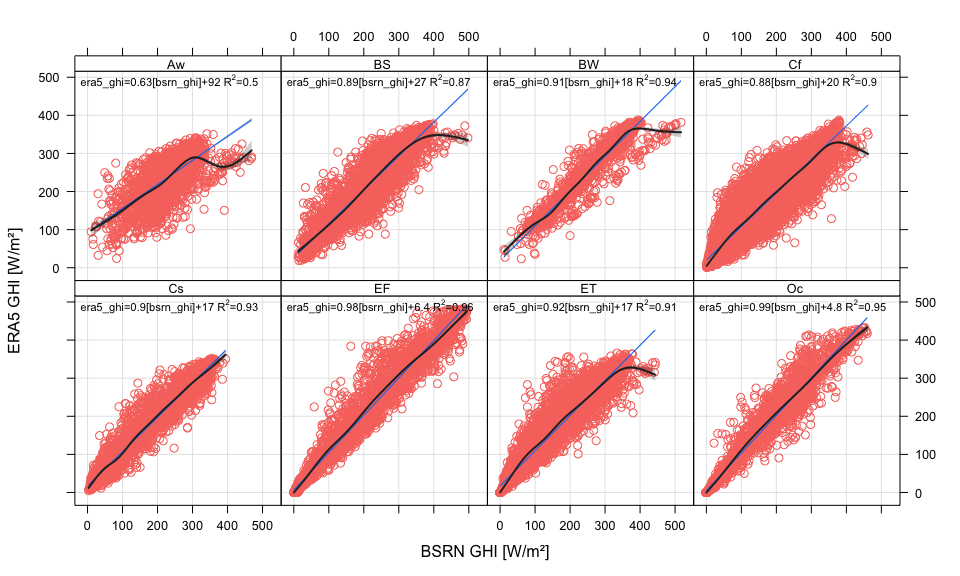


Figure 3: Matrix of scatterplots between reference dataset and the model’s estimates (W/m²) for all the stations and climate zones

## Variability and trend analysis

The main data set used at this stage is ERA5 monthly GHI averaged data on single levels from 1979 to present, otained from the Copernicus Climate Data Store (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means?tab=form>). The data was downloaded in GRIB format and converted to NetCDF using the CDO tool. The conversion from GRIB format involved also the transformation from the irregularly spaced latitudes to a regular Gaussian grid with an equal number of longitudes on each latitude row. The longitudes of the global regular grid were shifted from the 0 to 360 to the range -180 to 180 degrees, with the help of the *sellonlatbox* function from the same CDO tool.

cdo -f nc -t ecmwf setgridtype,regular adaptor.mars.internal-1565286172.367593-20556-9-d39c32ec-c74b-451c-97e3-32aab88f4217.grib ghi.nc  
cdo sellonlatbox,-180,180,-90,90 ghi.nc ghi\_rot\_1980\_2018.nc  
# transform from J/m2 to W/m2  
cdo aexpr,"SSRD=SSRD/(3600 \* 24)" ghi\_rot\_1980\_2018.nc ghi\_1980\_2018.nc  
# compute seasonal means  
cdo -v seasavg nc/ghi\_1980\_2018.nc nc/ghi\_seasonal\_1980\_2018.nc

## Solar resource maps generation

## Bibliography

Carslaw, David C, and Karl Ropkins. 2012. “Openair—an R Package for Air Quality Data Analysis.” *Environmental Modelling & Software* 27. Elsevier: 52–61.

Hamner, B, M Frasco, and E LeDell. 2018. *Metrics: Evaluation Metrics for Machine Learning*. *R Package Version 0.1.4*. Vol. 1.

Kaspar, Frank, Uwe Schulzweida, and Ralf Muller. 2010. “Climate Data Operators as a User-Friendly Processing Tool for Cmsaf’s Satellite-Derived Climate Monitoring Products.” In *The Proc. Of the Eumetsat Meteorological Satellite Conference*.

Rubel, Franz, Katharina Brugger, Klaus Haslinger, and Ingeborg Auer. 2017. “The Climate of the European Alps: Shift of Very High Resolution Köppen-Geiger Climate Zones 1800–2100.” *Meteorologische Zeitschrift* 26 (2). Schweizerbart’sche Verlagsbuchhandlung: 115–25.

Schmithüsen, Holger, Rainer Sieger, and Gert König-Langlo. 2012. “BSRN Toolbox V2. 0–a Tool to Create Quality Checked Output Files from Bsrn Datasets and Station-to-Archive Files.” PANGAEA.