

A non-parametric method for correction of global radiation observations

Peder Bacher^{a,*}, Henrik Madsen^a, Bengt Perers^b, Henrik Aalborg Nielsen^c

^aDTU Informatics, Technical University of Denmark, DK-2800 Lyngby, Denmark

^bDTU Civil Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark

^cENFOR AIS, DK-2970 Hørsholm, Denmark

Received 12 April 2012; received in revised form 1 October 2012; accepted 30 October 2012

Available online 23 December 2012

Communicated by: Associate Editor David Renne

Abstract

This paper presents a method for correction and alignment of global radiation observations based on information obtained from calculated global radiation, in the present study one-hour forecast of global radiation from a numerical weather prediction (NWP) model is used. Systematical errors detected in the observations are corrected. These are errors such as: tilt in the leveling of the sensor, shadowing from surrounding objects, clipping and saturation in the signal processing, and errors from dirt and wear. The method is based on a statistical non-parametric clear-sky model which is applied to both the observed and the calculated radiation in order to find systematic deviations between them. The method is applied to correct global radiation observations from a climate station located at a district heating plant in Denmark. The results are compared to observations recorded at the Danish Technical University. The method can be useful for optimized use of solar radiation observations for forecasting, monitoring, and modeling of energy production and load which are affected by solar radiation.

© 2012 Elsevier Ltd. All rights reserved.

Keywords: Global solar radiation; Solar energy; Observations; Correction; Quality control; Statistical clear-sky model

1. Introduction

The transition to a reliable and secure energy system based on weather dependent production technologies, especially wind and solar, will require new methods for automated handling of climate data recorded at, in most cases, unsupervised and uncalibrated stations. Reliable observations of solar radiation are an important source of information for operation of the energy system, especially for the energy production and load which are dependent on the solar radiation, for example production from

photovoltaics and solar collectors, and load from heating and cooling of buildings.

Observations of solar radiation are exposed to many sources of errors. Younes et al. (2005) list the most important types of errors and divide the errors into two major categories: equipment errors and operation related errors. The present solar radiation sensor technology makes it easy and cheap to install and connect sensors to the Internet, both for professional and amateur applications. Web sites already provide on-line data (DMI, 2012), which can become an important source of information for operation of energy systems. Such, mostly unsupervised and unvalidated installations, will be highly exposed to different error sources.

In the present study observations of global radiation from a station at a district heating plant in Sønderborg, Denmark, are used. Three types of errors are found in

* Corresponding author. Tel.: +45 60774725.

E-mail address: pb@imm.dtu.dk (P. Bacher).

URLs: <http://www.imm.dtu.dk/~hm> (H. Madsen), <http://www.enfor.eu> (H.A. Nielsen).

Nomenclature

G_t	observed global radiation (W/m^2)	$\tau_{a,B}$	transmittance function of the atmosphere for direct radiation under clear-sky conditions
G_t^{nwp}	numerical weather predictions (NWP) of global radiation (W/m^2)	τ_c	transmittance function of clouds in the atmosphere
G_{cs}	clear-sky global radiation (W/m^2)	β_t	parameter vector for the local quantile regression
B_{cs}	direct clear-sky global radiation (W/m^2)	$\rho_q(u)$	the quantile regression objective function
D_{cs}	diffuse clear-sky global radiation (W/m^2)	q	sample quantile to be estimated in the local quantile regression
G	global radiation (W/m^2)	i	counter of days (days)
I_{ext}	extraterrestrial radiation (W/m^2)	j	counter in samples
G_t^{pr}	projection of global radiation to the plane normal to the direct solar radiation (W/m^2)	t	time (h)
$\hat{G}_t^{\text{pr,cs}}$	estimated clear-sky radiation on a plane normal to the direct solar radiation (W/m^2)	t_{sp}	sample period (h)
\hat{G}_t^{cs}	estimated clear-sky global radiation (modeled based on observations) (W/m^2)	h_{doy}	bandwidth of kernel function in the <i>day of year</i> dimension (days)
$\hat{G}_t^{\text{nwp,cs}}$	clear-sky global radiation for numerical weather predictions (NWP) (W/m^2)	h_{tod}	bandwidth of kernel function in the <i>time of day</i> dimension (h)
\hat{G}_t^{co}	corrected global radiation (W/m^2)		
θ_t^{zenith}	solar zenith angle (rad)		

the observations: tilt in the leveling of the sensor, shadowing from surrounding objects, and clipping at a maximum level. A method is presented for correction of the observations on the basis of information extracted from global radiation calculated using a model based on physical principles. The method is based on a non-parametric statistical clear-sky model and requires no further information about the installation and sensor than the observed values and the location of the station. With the statistical clear-sky model the sensor output level under clear-sky conditions is modeled directly from the observations. This is compared to solar radiation calculated with a clear-sky model based on physical modeling of the optical effects through the atmosphere, such as the models described by Davies and McKay (1982), Bird (1984), Rigollier et al. (2000), Mueller et al. (2004), and Ineichen (2006). In the present study forecasts from a numerical weather prediction (NWP) model is used. The result after correction of the observations is compared to high quality measurements recorded at the Danish Technical University.

Studies on quality control of measured solar radiation data can be found in the literature. The procedures are semi-automatic and are mostly based on comparison to physical models for detection of erroneous measurements (Geiger et al., 2002; Younes et al., 2005; Isaac and Moradi, 2009; Journée and Bertrand, 2011).

The paper is organized as follows: the data used in the study is presented in the next section. This is followed by a section in which the statistical clear-sky model is described and a section where the correction is presented. The paper ends with a discussion of the method and a conclusion.

2. Data: observations and numerical weather predictions of global radiation

The data used in this study consists of time series of global radiation observed at two weather stations: one located in Sønderborg (54.91°N and 9.80°E) and one located at DTU Byg in Lyngby (55.79°N and 12.52°E), both in Denmark. In addition NWP of global radiation for the same locations are used. All values are hourly averages. All times are in UTC and the time points are set to the end of the hour.

2.1. Observations

The observations from Sønderborg are recorded with a weather station, which is located at a district heating plant. The weather station is mounted on a pole on a single-storey building as seen on the image in Fig. 1. No information about the type of the solar radiation sensor was available. The time series from Sønderborg is

$$\{G_t; \quad t = 1, \dots, N\} \quad (1)$$

where $N = 17520$ and G_t is the observed average global radiation between time t and $t - 1$. The upper plot in Fig. 2 shows the series which spans from 2009-01-01 to 2011-01-01. From this plot it is readily seen that the observations are not without systematic errors, for example it can be seen that the values are clipped at a maximum level. This and other types of systematic errors are corrected for the Sønderborg observations using the method described in this paper.



Fig. 1. The weather station in Sønderborg, which is mounted on a pole on the roof of a single-storey district heating plant building (in the image it is on the left side of the building).

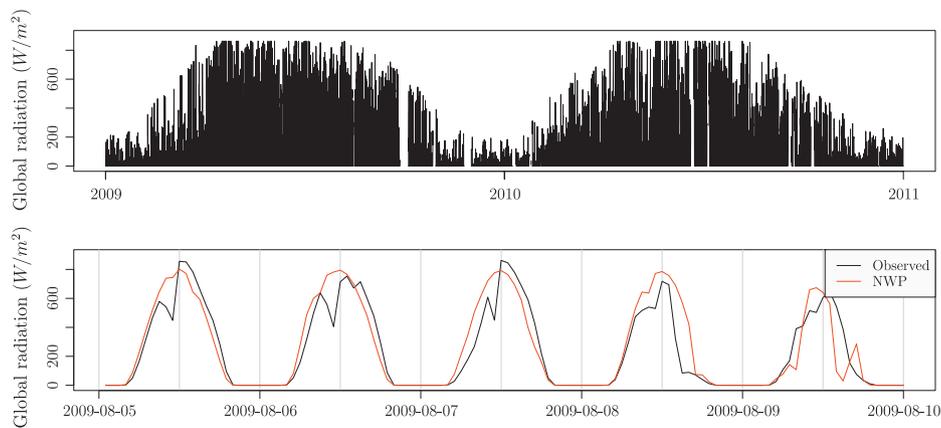


Fig. 2. The upper plot shows the time series of observed global radiation in Sønderborg. In the lower plot the observations and NWP of global radiation in Sønderborg are shown for 5 days in August 2009.

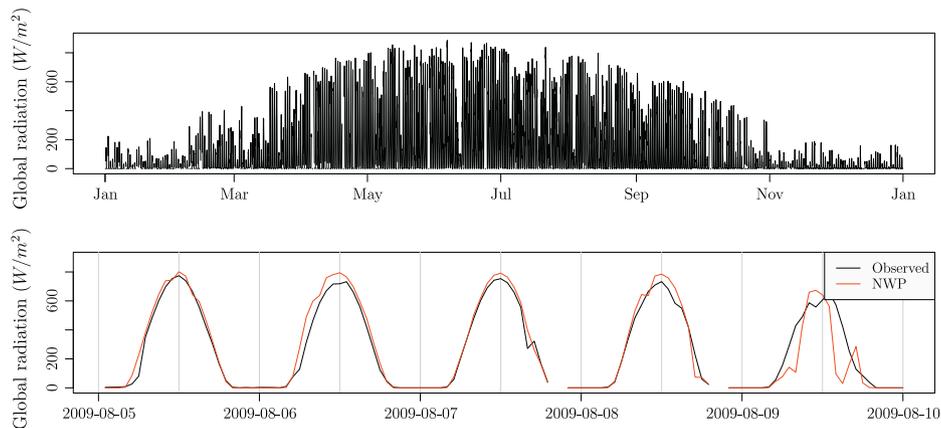


Fig. 3. The upper plot shows the time series of observed global radiation at DTU Byg covering the entire year 2009. In the lower plot the observations and NWP of global radiation at DTU Byg are shown for five days in August 2009.

The second series of observed global radiation is from a weather station at DTU Byg in Lyngby and is used as a reference to check the corrected data. The upper plot in Fig. 3 shows the series which spans from 2009-01-01 to 2010-01-01. It was measured with a Kipp and Zohnen CM10 pyranometer and the weather station was regularly supervised

in the measuring period. The measurement error is in the range of maximum $\pm 3\%$ from the world standard and high class calibrated sensor inter-comparisons indicate an error within the range of $\pm 1\%$. The lower plot in Fig. 3 shows the observations together with the NWP of global radiation (defined in the next section) for five days in August 2009.

It is seen that the level of the observed global radiation is generally lower than the level of the NWP, but that this there is no systematic difference between the deviation in the morning and in the afternoon. The lower level is most likely due to a bias in the NWP. Since the accuracy of the DTU observations is high and no systematic errors, apart from the generally lower level, is seen, then it is found valid to assume that the DTU observations can be used as a reference to verify the NWP and the results of the correction.

2.2. Numerical weather predictions

The numerical weather predictions (NWP) used in the study are provided by the Danish Meteorological Institute (DMI). The NWP model used is DMI-HIRLAM-S05, which has a 5 km grid and 40 vertical layers, see (DMI, 2011; Hansen Sass et al., 2002) for more details. The forecasts are updated four times per day and have a calculation delay of 4 h (e.g. the forecast starting at 00:00 is available at 04:00). Two time series, consisting of the latest available forecast (lead times are 5–11 h) of global radiation, are used: one for the location in Sønderborg and one for the location of DTU in Lyngby. The time series of NWP for the Sønderborg location is used for the correction. It is denoted with

$$\{G_t^{\text{nwp}}; t = 1, \dots, N\} \quad (2)$$

The time series for DTU Byg in Lyngby is shown, together with the observations, in the lower plot of Fig. 3 for five days in August.

2.3. Systematic errors in Sønderborg observations

The lower plot in Fig. 2 shows the Sønderborg observations and the NWP of global radiation for five days in August 2009. From the first day, which is a clear-sky day, at least two types of errors can be seen in the observations: compared to the NWP the observed level is too low in the morning and too high in the afternoon, which is most likely due to the sensor being tilted. It could also be due to a shift in time of the sensor, however it was thoroughly checked that the night hours, where the radiation was zero (or very close to zero), are with only a few exceptions the same hours for both the observed and the NWP, indicating that they are well synchronized. The second type of error is seen just before noon, where the observations have a drop, which is repeated at the same *time of day* on following clear-sky day. The drop is caused by shading from the chimney, which is located close to the weather station, as seen on the image in Fig. 1.

The scatter plot in Fig. 4 shows the observed values versus the NWP, together with two lines indicating the relation between the variables in the morning and in the afternoon. The lines are calculated using locally weighted least squares regression between the observations and the NWP, using the function `loess()` in R (R Development Core Team, 2011) with a bandwidth: `span = 0.9`. A similar plot for the DTU observations is found in Fig. 5. The fol-

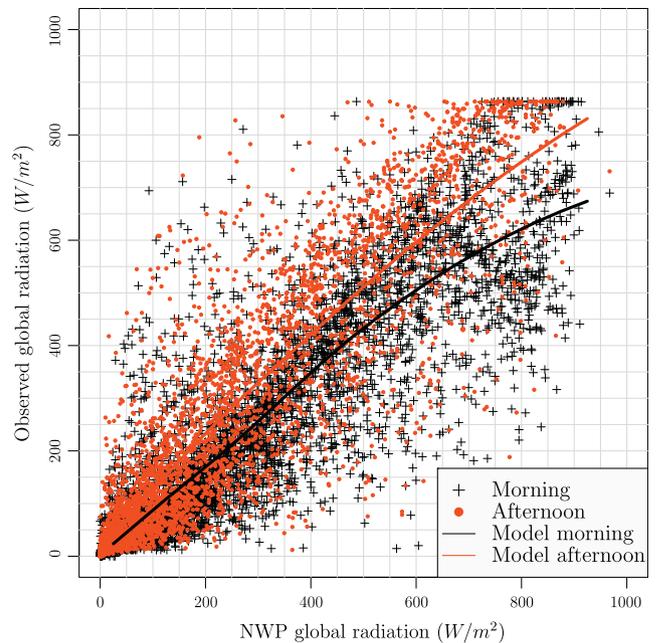


Fig. 4. The values of the Sønderborg observations versus the NWP covering the entire year 2009. The morning values and the afternoon values are indicated by different symbols and colors. The two lines show a locally weighted least squares regression estimate of the relation between the variables in the morning and the afternoon. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

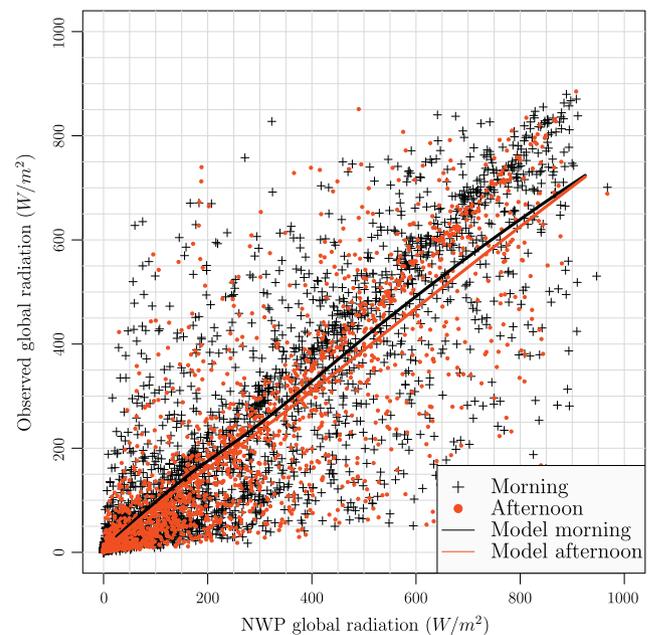


Fig. 5. The values of observed versus NWP global radiation at DTU Byg in Lyngby, Denmark. The morning values and the afternoon values are plotted with different symbols and colors. The two lines are indicating the relation between the variables: one for the morning and one for the afternoon. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

lowing three distinct systematic errors can be seen from the scatter plot for the Sønderborg observations:

1. Firstly, the observations are clipped at a maximum level around 860 W/m^2 .
2. Secondly, the level of the morning observations is generally lower than the level of the afternoon observations. This is confirmed by the fitted regression lines, which mostly have a difference of at least $50\text{--}75 \text{ W/m}^2$. This is clearly a larger difference than seen for the two fitted lines for the DTU observations in Fig. 5.
3. Finally, the morning values are significantly lower in the NWP range of $700\text{--}900 \text{ W/m}^2$. These values are the observations in the drop before noon, which, as described earlier, is caused by shadowing from the chimney right next to the weather station.

Considering the scatter plot for the DTU observations in Fig. 5 it is seen that these systematic errors are not found in the DTU observations. As noted before the level of the DTU observations is generally a bit lower than the level of NWPs, which is most likely due to a bias of the NWPs, since the accuracy of the DTU observations is verified to be in the range of $\pm 3\%$. For correction of the systematic errors, as the listed above, a statistical clear-sky model fitted to the observations can be used, as outlined in the following sections.

3. Statistical clear-sky model

In this section it is described how the clear-sky global radiation is modeled using a statistical model. With the statistical clear-sky model the level under clear-sky conditions at time t is estimated for the particular series of observations. It is the output of the sensor under clear-sky conditions which is estimated. This implies that if an observation is affected by a systematic error, for example shadowing from an object in the surroundings, the estimated clear-sky output will be lowered. It is this feature which enables the model to be used for correction. The statistical clear-sky model is a non-parametric model based on local polynomial quantile regression (Koenker, 2005) similar to the clear-sky model presented in (Bacher et al., 2009).

Usually, clear-sky models are models with which the global radiation in clear (non-overcast) sky at any given time can be calculated based on physical modeling of the atmosphere. Usually the clear-sky global radiation G_{cs} is separated into a direct (or beam) B_{cs} and diffuse D_{cs} component

$$G_{cs} = B_{cs} + D_{cs} \quad (3)$$

which are then modeled separately. The direct component by

$$B_{cs} = I_{ext} \cos(\theta_{zenith}) \tau_{a,B} \quad (4)$$

where I_{ext} is the extraterrestrial radiation, θ_{zenith} is the solar zenith angle and $\tau_{a,B}$ is a transmittance function of the atmosphere for direct radiation under clear-sky conditions, which for example can be modeled taking Rayleigh scattering, aerosol extinction, and ozone, water and uniformly

mixed gas absorption into account Bird and Riordan (1984).

The diffuse component can be modeled by adding several contributions from reflections and scattering through the atmosphere.

The global radiation (at the surface of the earth) can be modeled by

$$G = G_{cs} \tau_c \quad (5)$$

where τ_c is a transmittance function of clouds in the atmosphere, which can be modeled with “layer models” (Davies and McKay, 1982) where cloud layer transmittance and reflections are taken into account.

A clear-sky model, similar to the one proposed by Bacher et al. (2009) for observations of solar power, is here proposed for observations of global radiation. The proposed clear-sky model does not include any prior physical knowledge, it is based solely on the information obtained from the observations. It is denoted as a statistical clear-sky model, since it is based on a non-parametric statistical model of clear-sky radiation. Information embedded in the observations, which is particular for the sensor and its location, can be modeled with the statistical clear-sky model, for example shadowing and non-horizontal leveling of the sensor. This is a fundamental difference to the clear-sky models based on prior physical knowledge, which implies that the statistical clear-sky model can be used for different applications.

The statistical clear-sky model is based on time series of global radiation observations (or simulated values) and is defined by

$$G_t = \hat{G}_t^{cs} \tau_t \quad (6)$$

where the t is used to indicate that the variables the time series of actual observations, G_t is observed global radiation, \hat{G}_t^{cs} is estimated clear-sky global radiation and τ_t is a factor, which is much to alike τ_c , but different due to the fact that it is estimated based on information from observations and not calculated based on prior physical knowledge. It is noted here that the clear-sky model could be defined for the direct component solely, which would be obvious since nearly all local systematic effects have a much higher impact on the direct component compared to the diffuse component. However, since the application of the clear-sky model in the present study is for observations of global radiation and since the systematic errors would propagate into both the direct and diffuse component calculated with a splitting scheme, such as suggested by Ruiz-Arias et al. (2010), the clear-sky model is applied to the global radiation directly.

Considering the observed global radiation as samples of a random variable with a probability distribution function, which is a function of the *day of year* x_t and the *time of day* y_t , the observed clear-sky global radiation can be estimated as a quantile

$$\hat{G}_t^{cs} = Q_q(x_t, y_t) \quad (7)$$

of this distribution function, where the quantile $q \in [0, \dots, 1]$ must be close to one

$$q \lesssim 1 \quad (8)$$

Assuming that the quantile function is a smooth function it can be approximated with local quantile regression [Koenker \(2005\)](#). The result in the three-dimensional space formed by global radiation, *day of year* and *time of day*, can be seen as a surface which follows the observed global radiation under clear-sky conditions and is located “on top” of the point cloud of observed global radiation.

In order to decrease the gradient and curvature of the estimated clear-sky radiation surface a projection is carried out. The projection is from the horizontal plane to the plane which is normal to the direct solar radiation (i.e. the plane tracking the sun position)

$$G_t^{\text{pr}} = \frac{G_t}{\cos(\theta_t^{\text{zenith}})} \quad (9)$$

where θ_t^{zenith} is the average solar zenith angle in the sample period between $t - 1$ and t . Values where $\cos(\theta_t^{\text{zenith}}) < 0.01$ are removed: this corresponds to sun elevation below 0.5° . The quantile close to one is then estimated for the projected values. A general form of the proposed statistical clear-sky model is formulated in [Appendix A](#), which is based on a local quantile regression model with second order polynomials and a two-dimensional kernel in both the *day of year* and *time of day* dimensions.

For correction of hourly values a local quantile regression model based only on a one-dimensional kernel, where on the *day of year* dimension is used, was found most suitable. The reason for using only a one-dimensional kernel, and not including the *time of day* dimension in the local weighting, is that the model becomes too biased and the estimated clear-sky global radiation does not follow the drop before noon caused by shadowing (the systematic error described on page 7) very well. Hence only values lagged in steps of 24 h from t are used as input, which is a similar approach as in classical decomposition of seasonal time series ([Cleveland and Tiao, 1976](#)). Furthermore, it is noted that this is equivalent to using a bandwidth in the *time of day* dimension below 1 h (i.e. below the sample period) for the two-dimensional model presented in [Appendix A](#), hence for time series with a shorter sample period a two-dimensional model should be considered. The applied local quantile regression model based on a third order polynomial is

$$\hat{\beta}_i = \operatorname{argmin}_{\beta \in \mathbb{R}^4} \sum_{i=-\infty}^{\infty} \rho_q(G_{t+24i}^{\text{tr}} - (\beta_{0,i} + \beta_{1,i}i + \beta_{2,i}i^2 + \beta_{3,i}i^3))K(i) \quad (10)$$

where $\rho_q(u) = u(q - I(u < 0))$ is the quantile regression objective function (see ([Koenker, 2005, 2011](#))), $q \in [0, \dots, 1]$ is the sample quantile to be estimated, $i \in \mathbb{N}$ is a counter of days, and $K(i)$ is a kernel function. The estimated projected clear-sky radiation is then found as the local intercept

$$\widehat{G}_t^{\text{pr,cs}} = \hat{\beta}_{0,i} \quad (11)$$

The weights are calculated with the Epanechnikov kernel function

$$K(i) = \begin{cases} \frac{3}{4} \left(1 - \left[\frac{|i|}{h_{\text{doy}}}\right]^2\right) & \text{for } \frac{|i|}{h_{\text{doy}}} \geq 1 \\ 0 & \text{for } \frac{|i|}{h_{\text{doy}}} < 1 \end{cases} \quad (12)$$

where h_{doy} is the bandwidth.

The R package `quantreg` implementation of quantile regression was used for the estimation ([Koenker, 2011](#)). Finally, the estimated projected clear-sky radiation on the projected plane is projected back to the horizontal plane by

$$\widehat{G}_t^{\text{cs}} = \widehat{G}_t^{\text{pr,cs}} \cos(\theta_t^{\text{zenith}}) \quad (13)$$

Finally, in order to take the clipping at a maximum level into account, the estimated clear-sky radiation is limited to the maximum value of the observations

$$\widehat{G}_t^{\text{cs}} = \begin{cases} \widehat{G}_t^{\text{cs}} & \text{for } \widehat{G}_t^{\text{cs}} \leq G_t^{\text{max}} \\ G_t^{\text{max}} & \text{for } \widehat{G}_t^{\text{cs}} > G_t^{\text{max}} \end{cases} \quad (14)$$

where G_t^{max} is the maximum value of global radiation observations.

The selection of suitable values for the parameters (here the quantile and the kernel bandwidth) for the fitting of the local quantile regression model, would preferably be based on a measure of performance for estimation clear-sky global radiation. Then the parameters could be optimized in order to achieve the best performance. However thorough studies are required in order to define such a measure. Therefore the parameter values are selected based on visual inspection of the estimated clear-sky global radiation for days with only clear-sky. These days are chosen such that they are distributed evenly over the entire period. The selected values are

$$q = 0.97, \quad h_{\text{doy}} = 125 \quad (15)$$

which gives the estimate of the clear-sky global radiation for the observations $\widehat{G}_t^{\text{cs}}$ shown in [Fig. 6](#) and for the NWP $\widehat{G}_t^{\text{nwp,cs}}$ shown in [Fig. 7](#). Note, that the estimated surface for the observations is clipped at the maximum value of the observations, which gives the “flat” top. Furthermore, notice that the drop due to shadowing is clearly seen in the estimated clear-sky radiation for the observations.

4. Correction of observations

The correction of the observations is carried out by multiplying the observations with the ratio between the estimated clear-sky radiation for the NWP and the observations

$$\widehat{G}_t^{\text{co}} = \frac{\widehat{G}_t^{\text{nwp,cs}}}{\widehat{G}_t^{\text{cs}}} G_t \quad (16)$$

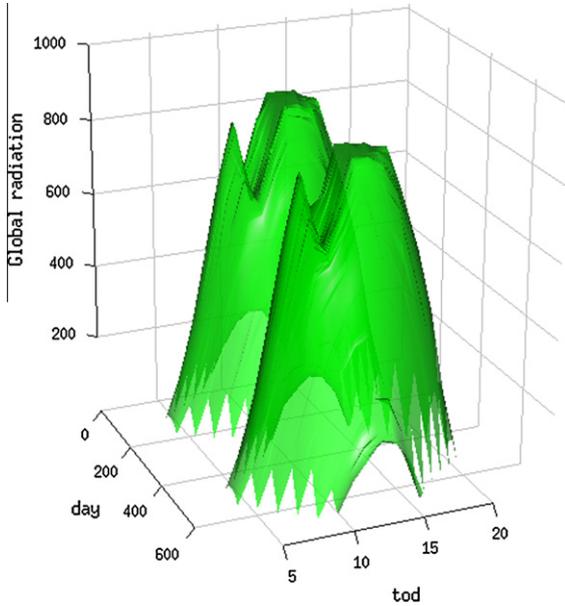


Fig. 6. The clear-sky global radiation estimated for the Søndersborg observations. Shown as a surface parametrized in the two dimensions: day (days since 2009-01-01) and tod (*time of day*).

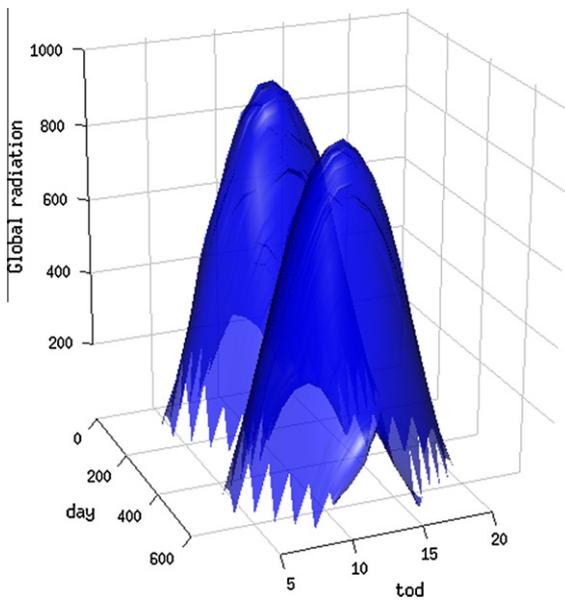


Fig. 7. The clear-sky global radiation estimated for the NWP for Søndersborg. Shown as a surface parametrized in the two dimensions: day (days since 2009-01-01) and tod (*time of day*).

The level of the correction applied, i.e. $\hat{G}_t^{\text{nwp,cs}} / \hat{G}_t^{\text{cs}}$, is shown as function of days since 2009-01-01 and the *time of day* in Fig. 8. The systematical error caused by a tilt of the sensor, resulting in a too low level of the observations in the morning and too high level in the afternoon, can be directly seen in the correction, since in the morning the correction is generally above one and the afternoon level below one. Also apparent is the drop in the observed level due to shadowing objects, especially seen between 9 and 10 am.

The corrected observations are plotted versus the NWP in Fig. 9, including the local least squares estimate of the relation in the morning and in the afternoon. This plot is similar to the plot in Fig. 4. By comparison of the two plots it is seen that the difference between the estimated relation in the morning and the estimated relation in the afternoon has been decreased significantly. A visual comparison to the similar plot of the high quality DTU observations in Fig. 5 verifies that the pattern of the scatter after the correction is much closer to the pattern found there. It can also be seen that the clipping at a maximum level has been corrected. Finally, it is found that the overall scattering has been reduced. This is confirmed by a comparison of the errors for an estimated relation similar to the ones in Figs. 4 and 9, but using all data points (except nighttime values), i.e. no distinction between morning and afternoon. Note here that this measure is only used to give a rough indication of the performance of the correction. The root mean square error (RMSE) and mean absolute error (MAE) before the correction are

$$\text{RMSE}_{\text{before}} = 114 \text{ W/m}^2, \text{ MAE}_{\text{before}} = 79 \text{ W/m}^2 \quad (17)$$

and after the correction

$$\text{RMSE}_{\text{after}} = 101 \text{ W/m}^2, \text{ MAE}_{\text{after}} = 67 \text{ W/m}^2 \quad (18)$$

Hence a notably reduction in RMSE and MAE is achieved by applying the correction.

4.1. On-line operation

For on-line operation the model has to be applied causally, such that only past values can be used for the correction. A causal correction was calculated with slightly different parameter values for the clear-sky model, again selecting the parameters from visual inspection. The estimated quantile q was decreased and the kernel bandwidth h_{day} increased slightly to

$$q = 0.95, h_{\text{day}} = 150 \quad (19)$$

Using a one-sided kernel will increase the bias of the estimates, which is also reflected by a slightly increased RMSE and MAE of the `loess` fit for the corrected observations to

$$\text{RMSE}_{\text{after}} = 103 \text{ W/m}^2, \text{ MAE}_{\text{after}} = 67 \text{ W/m}^2 \quad (20)$$

Considering the similar plots as presented for the causal correction showed only a small visual difference. Hence it is found that the method works well for on-line operation.

5. Discussion

In this section the correction method and results are discussed together with considerations on how to improve the method.

Considering the fitting of the clear-sky model it is noted that the model which should be applied, is dependent on the time resolution of the data. For resolutions higher than

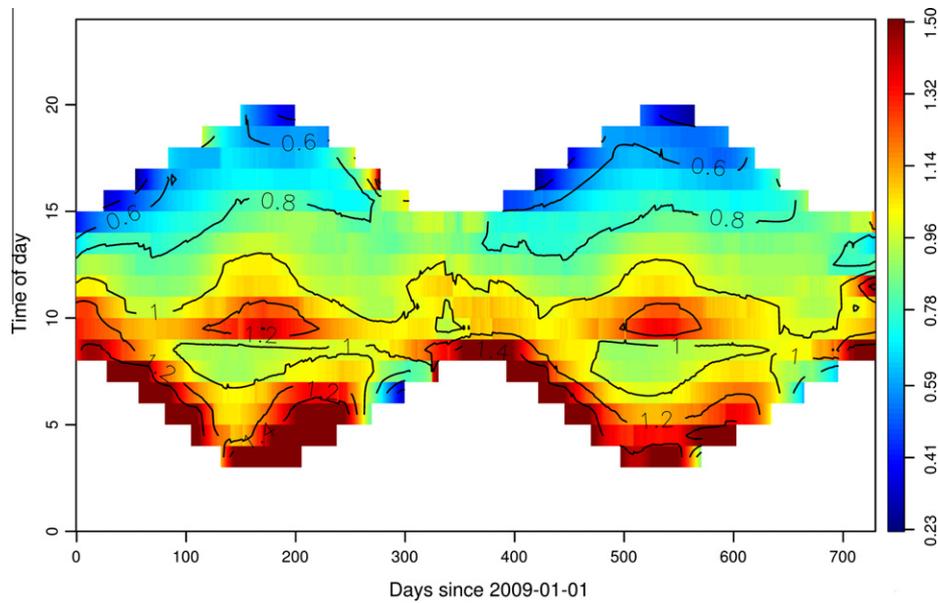


Fig. 8. The applied correction, which is the ratio between the estimated clear-sky radiation of the observations and the NWP, as a function of days and *time of day*.

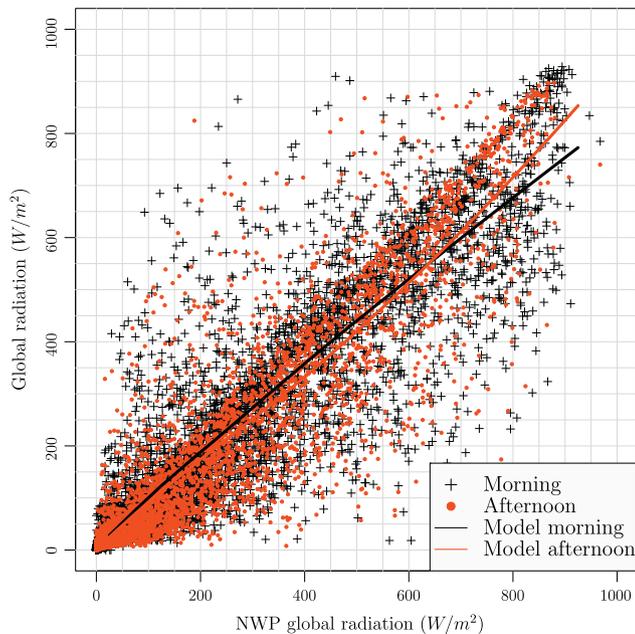


Fig. 9. The values of the corrected observations versus NWP of global radiation. The morning values and the afternoon values are indicated by different symbols and colors. The two lines show a locally weighted least squares regression estimate of the relation between the variables in the morning and the afternoon. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

hourly a two-dimensional kernel, which also includes the *time of day* dimension, should be used, as in (Bacher et al., 2009) where a similar clear-sky model was applied to 15 min values. Regarding the parameters needed to be tuned in the clear-sky model – the quantile, kernel bandwidth, and order of the polynomial – some manual interac-

tion is required. However the method could be based on a parameter optimization criteria, hence an objective measure to evaluate the performance of the correction, possibly based on cross validation (Friedman et al., 2001), and applied automatically for the general case. Hence it can also be used for monitoring and data quality classification for sub-daily solar data. Clearly, an objective measure of performance of the correction is needed in order to further develop and improve the correction method.

Improvements of the method could be formed by combining it with a prior step in which a parametric model is fitted to correct for drift in time and tilt in the leveling of the sensor. Another possibility for improvement is to treat the direct and diffuse radiation separately, since most of the systematic errors, for example tilt and shadowing, will have a different impact on direct and diffuse radiation. This will require, if the direct and diffuse are not measured separately, a splitting into a diffuse and direct component, which could be carried out with a scheme such as suggested by Ruiz-Arias et al. (2010) and Duffie and Beckman (2006, pp. 75–77). However applying such a scheme will cause the effect of the systematic errors to propagate into both the direct and the diffuse components. Another approach would be to enhance the correction method by using more than one quantile, in the presented approach only a single quantile close to one is used. Several quantiles can be estimated for both the observed and calculated radiation, which, together with an interpolation scheme, will form a more extensive correction. Clearly this also requires that the calculated global radiation, i.e. here the NWP, describes the distribution well over the entire range of global radiation. Finally, it is mentioned that for on-line operation the method can be implemented computationally

very efficient using time-adaptive quantile regression (Møller et al., 2008).

6. Conclusion

A correction method based on statistical non-parametric modeling techniques is presented and applied on hourly observations of global radiation. Several typical errors in the observations can be corrected with the method, including: tilt in the leveling of the sensor, shadowing from objects in the surroundings and clipping of the observations at a maximum level. The method works semi-automatically and no prior information about the sensor and its surroundings, besides the observations and location, is required. Furthermore only a few parameters needs to be tuned. Information embedded in NWP's of global radiation is used for the correction, but this could be replaced with any calculated clear-sky global radiation model. The method is well suited as part of monitoring and operation applications for which local solar radiation observations provide valuable information, e.g. for forecasting of climate dependent renewables such as solar thermal, PV and heating systems. Finally, it is briefly discussed how the method can be improved or extended in several ways.

Acknowledgments

The observations of global radiation was very kindly provided by Sønderborg Fjernvarme and we also thank the Danish Meteorological Institute for making their numerical weather predictions available. Finally, acknowledgments are given to the The Danish Council for Strategic Research, which have provided the financial support for the project “Solar/electric heating systems in the future energy system” under which the work was carried out.

Appendix A. Two-dimensional local statistical clear-sky model

The proposed statistical clear-sky model in a general form is described in this section. It is based on a two-dimensional second-order polynomial local quantile regression model. In this form the local weighting is carried out with a two-dimensional multiplicative kernel function in the *day of year* and *time of day* dimensions.

The model

$$\hat{\beta}_i = \operatorname{argmin}_{\beta \in \mathbb{R}^5} \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \rho_q \left(G_{t_{sp}+i+j}^{\text{tr}} - (\beta_{0,i} + \beta_{1,i}i + \beta_{2,i}i^2 + \beta_{3,i}j + \beta_{4,i}j^2) \right) K(i, j) \tag{A.1}$$

where t_{sp} is the sample period of the time series in hours, $\rho_q(u) = u(q - I(u < 0))$ is the quantile regression objective function (see Koenker, 2005; Koenker, 2011), $q \in [0, \dots, 1]$ is the sample quantile to be estimated, $i \in \mathbb{N}$ is a counter of days, $j \in \mathbb{N}$ is a counter in steps of the sample period, and $K(i, j)$ is a kernel function. The model could easily be reduced or expanded to polynomials of different

orders. The estimated projected clear-sky radiation is then found as the local intercept

$$\hat{G}_i^{\text{pr,cs}} = \hat{\beta}_{0,i} \tag{A.2}$$

The weights are calculated with the Epanechnikov kernel function

$$K(i, j) = \begin{cases} \frac{9}{16} \left(1 - \left[\frac{|i|}{h_{\text{doy}}} \right]^2 \right) \left(1 - \left[\frac{|j| t_{sp}}{h_{\text{tod}}} \right]^2 \right) & \text{for } \frac{|i|}{h_{\text{doy}}} < 1 \wedge \frac{|j|}{h_{\text{tod}}} < 1 \\ 0 & \text{for } \frac{|i|}{h_{\text{doy}}} \geq 1 \vee \frac{|j|}{h_{\text{tod}}} \geq 1 \end{cases} \tag{A.3}$$

where h_{doy} is the bandwidth in the *day of year* dimension (in days) and h_{tod} is the bandwidth in the *time of day* dimension (in hours).

References

Bacher, P., Madsen, H., Nielsen, H.A., 2009. Online short-term solar power forecasting. *Solar Energy* 83 (10), 1772–1783.

Bird, R.E., 1984. A simple, solar spectral model for direct-normal and diffuse horizontal irradiance. *Solar Energy* 32 (4), 461–471. <<http://www.sciencedirect.com/science/article/pii/0038092X84902603>>.

Bird, R., Riordan, C., 1984. Simple solar spectral model for direct and diffuse irradiance on horizontal and tilted planes at the earth's surface for cloudless atmospheres. Tech. rep., Solar Energy Research Inst., Golden, CO, USA.

Cleveland, W., Tiao, G., 1976. Decomposition of seasonal time series: a model for the census x-11 program. *Journal of the American statistical Association*, 581–587.

Davies, J.A., McKay, D.C., 1982. Estimating solar irradiance and components. *Solar Energy* 29 (1), 55–64. <<http://www.sciencedirect.com/science/article/pii/0038092X82902808>>.

DMI, 2011. Danish Meteorological Institute, DMI-HIRLAM-S05. <http://www.dmi.dk/eng/index/research_and_development/dmi-hir-lam-2009.htm>

DMI, Feb. 2012. Danish Meteorological Institute, Borgervejr. <<http://www.borgervejr.dk>>.

Duffie, J.A., Beckman, W.A., 2006. *Solar Engineering of Thermal Processes*, third ed. Wiley.

Friedman, J., Hastie, T., Tibshirani, R., 2001. *The elements of statistical learning*. vol. 1. Springer Series in Statistics.

Geiger, M., Diabaté, L., Ménard, L., Wald, L., 2002. A web service for controlling the quality of measurements of global solar irradiation. *Solar Energy* 73 (6), 475–480.

Hansen Sass, B., Woetmann Nielsen, N., Jørgensen, J.U., Amstrup, B., Kmit, M., Mogensen, K.S., 2002. The operational DMI-HIRLAM system 2002-version. DMI.

Ineichen, P., 2006. Comparison of eight clear sky broadband models against 16 independent data banks. *Solar Energy* 80 (4), 468–478.

Isaac, Moradi, 2009. Quality control of global solar radiation using sunshine duration hours. *Energy* 34 (1), 1–6.

Journée, M., Bertrand, C., 2011. Quality control of solar radiation data within the RMIB solar measurements network. *Solar Energy* 85 (1), 72–86.

Koenker, R., 2005. *Quantile Regression*. Cambridge University Press.

Koenker, R., 2011. *Quantreg: Quantile Regression*. R Package Version 4.76. <<http://CRAN.R-project.org/package=quantreg>>

Møller, J.J.K., Nielsen, H.A., Madsen, H., 2008. Time-adaptive quantile regression. *Computational Statistics and Data Analysis* 52 (3), 1292–1303.

Mueller, R., Dagestad, K., Ineichen, P., Schroedter-Homscheidt, M., Cros, S., Dumortier, D., Kuhlemann, R., Olseth, J., Piernavieja, G.,

- Reise, Wald, L., Heinemann, D., 2005. Rethinking satellite-based solar irradiance modelling: the solis clear-sky module. *Remote Sensing of Environment* 91 (2), 160–174.
- R Development Core Team, 2011. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN:3-900051-07-0. <<http://www.R-project.org/>>
- Rigollier, C., Bauer, O., Wald, L., 2000. On the clear sky model of the esra – European solar radiation atlas – with respect to the heliosat method. *Solar Energy* 68 (1), 33–48.
- Ruiz-Arias, J., Alsamamra, H., Tovar-Pescador, J., Pozo-Vázquez, D., 2010. Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. *Energy Conversion and Management* 51 (5), 881–893.
- Younes, S., Claywell, R., Muneer, T., 2005. Quality control of solar radiation data: present status and proposed new approaches. *Energy* 30 (9), 1533–1549 (measurement and modelling of solar radiation and daylight – challenges for the 21st century).