

Research Article

About the Relation between Sunshine Duration and Cloudiness on the Basis of Data from Hamburg

Stefanie Neske

Institute for Geophysics and Meteorology, University of Cologne, Pohligstraße 3, 50969 Cologne, Germany

Correspondence should be addressed to Stefanie Neske; sneske@uni-koeln.de

Received 4 February 2014; Accepted 28 March 2014; Published 10 April 2014

Academic Editor: Guillaume Zoppi

Copyright © 2014 Stefanie Neske. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The aim of this paper is to relate the two meteorological parameters known as relative (bright) sunshine duration and cloudiness using the data from two stations of the city of Hamburg, Germany. We test the classic linear relationship, as well as newer polynomial extensions suggested in the literature. The results of regression are interpreted against a theoretical background recently put forward by Badescu. The suggested relations can be borne out, but we also point out difficulties due to data quality and insufficiency.

1. Introduction

Knowledge of the relationship between sunshine duration and cloudiness is very important for the practical forecast of insolation [1]. In solar energy technology there are key meteorological parameters on both short and longer time scales. Datasets with sunshine duration are widespread and easily available for many parts of the world and long time intervals. In contrast, cloudiness has been slow in being estimated or measured reliably. For use in solar energy technologies, cloudiness is fundamental [1] because the insolation reaching the ground depends on the amount and type of the clouds blocking the direct solar radiation [2]. Due to the fact that cloudiness is much more difficult to measure accurately, it has since long been found necessary to look for a relationship between the widely available sunshine duration measurements and the formerly rather elusive estimation of cloudiness.

This short work draws on data from one single German city, Hamburg. We shall use as our data source only two stations located in this city. The first one is a transmitter mast of the “Norddeutscher Rundfunk” (Northern German Broadcasting Corporation, NDR), while the second is Hamburg’s airport. At first, the data from these locations are evaluated separately and then they are compared with each other. We expect that both stations should deliver similar relations as Hamburg is a flat and rather homogeneous city.

Correlations of both quantities from many cities have been used in proposing a relationship between them. In recent times, the level of sophistication has been raised by Badescu, who carried out many case studies for Romania, from at least 1990 on. He summarized the findings of his collaborators in a book he edited recently [3]. The general conclusion is that one may use a linear relationship between cloudiness and cloud shade for Romania. But he also has summarized a probabilistic theory that can lead to more general relationships, which had been explored or proposed earlier but on uncertain grounds. This theory of Badescu’s will serve as the backdrop for the present work.

There are many studies which do not connect the sunshine duration and cloudiness directly but instead infer the global insolation from observations of sunshine data (e.g., Aksoy et al. [1], Suercke [4]). For such cases where only one or the other quantity is measured, it is an advantage to have previous information on the correlation between the two parameters, especially for cities where the data either of sunshine or of cloudiness are lacking. Then the missing data can be calculated from measurements of the other parameter.

We shall seek to confirm the simplest of possible relations between sunshine and cloudiness, which is the linear one. It will be seen that confirmation of such a linear relation, as it had almost been assumed, may not be sufficiently accurate for certain purposes or time scales. We shall explain deviations by introducing an irradiance threshold value. An important

advantage of our study, in contrast to other publications dealing with the same problem regarding Hamburg, is the relatively long data set, beginning in 1996 and ending in 2011. The fact that we use the data from two similar stations may be useful in validation studies of the results of a single station. The issue of representativeness can thus be assessed, at least for Hamburg.

The structure of the paper is as follows. In Section 2 the data and the methods that have been used for this work are discussed. Section 3 then presents the results and their interpretation, while Section 4 compares the results for the two stations. The paper ends with Section 5, the conclusion, and an outlook or rather a view to one or two possibilities of improving the present study.

2. Data and Methods

The two stations are a transmitter mast of the NDR and the weather station at Hamburg airport. The distance between them is about 14.3 km. We treat the stations separately at first.

2.1. NDR Transmitter Mast. The transmitter mast is about 300 m high. It is located in Billwerder, Hamburg, on plain grassland. It has been in use for weather measurements by the meteorological institute of the University of Hamburg since 1967. In the following, it will simply be referred to as the “weather mast.” The geographical coordinates are $53^{\circ}31'0.9''$ N and $10^{\circ}06'10.3''$ E. The mast is just about 30 cm above sea level, so the instruments over ground are approximately at sea level. There are some agricultural districts in the surrounding and flat, cultivated land. It is a good location for the purposes of the present work, as well as many other measurements [5]. Sunshine duration is recorded with a pyranometer, which measures the global incoming shortwave radiation from the upper half-space. Taking into account the World Meteorological Organization’s sunshine criterion that the sun is effectively shining only if the direct solar radiation exceeds a threshold of 120 W m^{-2} [6], one can determine the sunshine duration from the measured record.

Cloudiness is measured at the weather mast by means of a ceilometer. It is placed at the ground and is of the type Vaisala CT25K, which measures the cloud base in steps of 30 m up to a level of 7500 m. From the backscattering profile of a laser beam and by means of an algorithm, it is possible to determine cloudiness of up to four overlying cloud basements. This instrument has been working at the weather mast since 2003.

2.2. Hamburg Airport. The airport has a geographical position of $53^{\circ}38'7.95''$ N, $9^{\circ}59'40.92''$ E and it is obviously level, but the area is surrounded by trees. For the measurement of the sunshine duration, the airport has two different measuring devices. One is the classical Campbell-Stokes sunshine recorder, a polished glass sphere focusing the incoming sunbeam on a card with the day’s hours printed on it, thereby burning a trace on it, the length of which is proportional to a day’s sunshine. The other device is a SONI sunshine recorder,

which is based on an electronic method. It is the source of the data used.

Cloudiness at the airport is measured in the old-fashioned way by a human observer, whose subjective assessment is subject to greater or lesser uncertainty.

2.3. The Database. For this short account, a time period of 16 years was used, from 1 January 1996 (midnight) to 31 December 2011 (23:59 o'clock). An exception is the time series of the cloudiness from the weather mast, which starts only in November 2003. To fill in the missing years, a statistical AR(1) process was envisaged to extend the series back into the past. But this part of the investigation will not be included in this study. Fortunately, the cloudiness time series of the airport is complete within the whole time period selected.

Overall, four time series were used for this paper, one time series of sunshine duration and one of cloudiness from each location. Since the data of the airport were available only as daily means, the data from the weather mast were averaged over every day, from values recorded every single minute. There is a little data gap within the year 2005 of the time series for the sunshine duration of the weather mast. This is due to problems with the instrument [5].

2.4. Theoretical Background. With our purpose being only to report new results for Hamburg of the correlation between sunshine and cloudiness, we shall not go into details of the theory beyond what is needed to interpret those results. We will follow Badescu [3], who introduced the sunshine number. Sunshine number is a random Boolean variable in time, defined as follows:

$$\xi(t) = \begin{cases} 0 & \text{if the sun is covered by clouds at time } t \\ 1 & \text{otherwise,} \end{cases} \quad (1)$$

where t is any instant of daylight duration. Because the dynamics of cloud cover is too complicated, or even unknown, we may regard $\xi(t)$ as a random variable. Choosing an (arbitrary) interval of time Δt centered on t , we define the probability of the sun being covered by clouds during that interval as $p(\xi = 0; t, \Delta t)$. Being a Bernoulli-distributed variable, the complementary probability of the sun not being covered with clouds will be written as $p(\xi = 1; t, \Delta t) = 1 - p(\xi = 0; t, \Delta t)$.

Now we introduce briefly two independent measures for these probabilities. If we denote by $s(t, \Delta t)$ the sum of those time units corresponding to sunshine during the time period Δt , centered on t , then we may define the probability of sunshine during that period as the usual relative sunshine:

$$p(\xi = 1; t, \Delta t) = \frac{s(t, \Delta t)}{\Delta t} \equiv \sigma(t, \Delta t). \quad (2)$$

A probability measure for $p(\xi = 0; t, \Delta t)$ is more difficult to define theoretically. Badescu [7], following earlier work on geometric probability, as developed by Santaló (cf., e.g., [8]), derived the following expression for that probability:

$$p(\xi = 0; t, \Delta t) = C \frac{2\pi(A_S + A_C) + L_S L_C}{2\pi A_C}, \quad (3)$$

with the total cloud cover amount being defined as

$$C = C(t, \Delta t) \equiv \frac{A_{0C}}{A_0} = \frac{2\pi A_C}{2\pi(A_0 + A_C) + L_0 L_C}. \quad (4)$$

Here, A_S and A_C are, respectively, the areas of the sun's disk and clouds in the celestial vault projected onto the plane tangent to the Earth's surface at the point of observation. L_A and L_C are the perimeters bounding those areas. A_0 is the total area of this plane, limited by the horizon of perimeter L_0 , and A_{0C} is the area of intersection of A_0 and A_C . It can be viewed as a random quantity. For further details, we must refer the reader to the works cited.

The probability (3) may be written as

$$p(\xi = 0; t, \Delta t) = C(t, \Delta t)(1 + \epsilon), \quad (5)$$

where, for $C = 0$,

$$\epsilon = \frac{A_S}{A_0} \ll 1, \quad (6)$$

while for $C \rightarrow 1$ it should hold that $\epsilon = 0$. This is indeed the case if we agree that $A_S = 0$ for overcast skies. For $C \neq 0$, we see that

$$\epsilon = \frac{2\pi A_S + L_S L_C}{2\pi A_C}, \quad (7)$$

where ϵ may be taken to depend on C . Note that $A_S \ll A_C$ and $L_C L_S \ll 2\pi A_C$, so that as a good approximation we may write

$$p(\xi = 0; t, \Delta t) = 1 - \sigma(t, \Delta t) \approx C(t, \Delta t). \quad (8)$$

This is nothing else than the classic assumption that cloudiness and sunshine ought to add up to one. $N = 1 - \sigma(t, \Delta t)$ is also known as the cloud shade. It will be used as the variable against which we shall plot the cloud cover C . Both N and C are available from routine meteorological observations or measurements. Ideally, both variables should be perfectly correlated. This is what we set out to ascertain with the data from the two stations in Hamburg. Of course, as we can only estimate the two quantities, we shall distinguish them by using a tilde on the corresponding variables.

Following Badescu [3] we at first accept the approximation

$$\tilde{\sigma} \approx \sigma \quad (9)$$

as a good one. But we must bear in mind that sunshine measurements are contaminated by various factors, and this fact led the World Meteorological Organization to define a *bright* sunshine by excluding solar irradiance below a threshold value of 120 W/m^2 [6]. Therefore, the estimated value is always smaller than the theoretical one, $\sigma > \tilde{\sigma}$, or $\sigma = \tilde{\sigma} + \gamma$, with $\gamma > 0$. But first we shall test the goodness of the approximation (9). Hence, our estimated cloud shade will be defined as

$$\tilde{N} = 1 - \tilde{\sigma}. \quad (10)$$

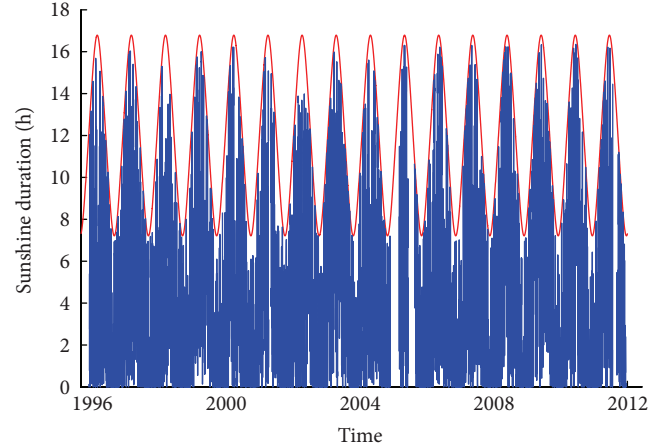


FIGURE 1: Weather mast: time series of sunshine duration including possible maximum value (red).

In the following regression plots the estimated cloud shade \tilde{N} is shown against the observed cloudiness \tilde{C} , in order to test the validity of their approximate equality, as has been traditionally assumed (at least for measuring places without complex topography or other obstacles to the sun's direct irradiation).

3. Results

3.1. The Weather Mast. Due to the lack of cloudiness measurements before 2004, the corresponding time series at the weather mast is shorter than the time period we are envisaging in this study. An attempt to complete it by an AR(1) process, though partly successful, will not be explored here, as we want to avoid any artificial data in answering our question above. Measured data from the weather mast are reduced to daily means for the time period of the eight years from 2004 to 2011. As an example, the time series of the sunshine duration with its corresponding possible maximum value (red curve) of the weather mast is shown in Figure 1. The data gap within the year 2005 is visible, as mentioned before in Section 2. A plot of measured cloud shade \tilde{N} and cloudiness \tilde{C} is shown in Figure 2.

It contains 5,844 daily mean values and it also shows the linear relationship $\tilde{N} = \tilde{C}$, corresponding to the approximation (8). This ideal relationship is not, however, borne out by the figure. Most of the data points lie above the line, so that $\tilde{N} > \tilde{C}$. Extreme values of \tilde{C} are 0.9, while cloud shade does achieve values of 1. There is some accumulation of data points around 70% cloudiness, with cloud shade values greater than 90%. On the other hand, for very small cloudiness, cloud shade may reach values up to 0.65, with frequent values around 0.1. This is theoretically unsatisfactory because without clouds we expect a vanishing cloud shade. A slight curvature is observed from around 50% cloudiness upwards. It thus seems that a better fit can be achieved with a polynomial regression model. Figure 3 shows the same plot as Figure 2 but with four polynomial regression

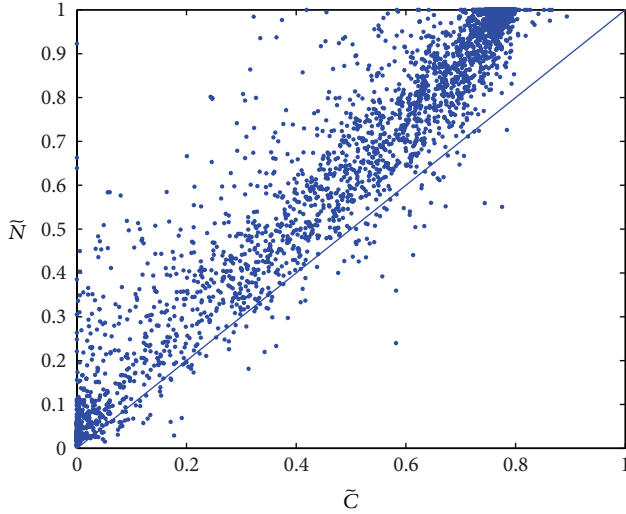


FIGURE 2: Weather mast: cloudiness against cloud shade daily mean.

TABLE 1: RMSE for Figure 3.

	Linear	Quadratic	Cubic	4th power
RMSE	0.1065	0.1066	0.1055	0.1062

lines. The least RMSE (root mean square error) indicates the best fitting polynomial order of the regression (Table 1).

On the basis of the regression lines, one can see that the cubic and the fourth power are almost equal, so that there was no need for higher orders. In general, there is almost no difference between the goodness of fit among the four regression lines, and Occam's razor would require choosing the linear relationship. Table 2 shows the values of the regression parameters, and the formula with the "best" fit is

$$\tilde{N} = 0.83046\tilde{C}^3 - 0.95267\tilde{C}^2 + 1.3959\tilde{C} + 0.081687. \quad (11)$$

3.2. The Airport. The time series of the airport are complete, so there was no need to adopt any procedure to fill in values. Figure 4 shows the observed cloudiness plotted against the measured cloud shade. Note the higher scatter of the points, but we must not forget that more points have been available for the plot. There is a similar trend upwards for higher values of cloudiness, similar to what we saw in the case of the weather mast. As before, there is cloud shade for no cloudiness. We shall soon discuss the reasons for this.

Table 3 shows that the quadratic polynomial fits the data with the least RMSE. The four fitting polynomials are also shown in Figure 4. A linear relationship with a displaced intercept would also be acceptable as an average relationship between the quantities plotted. The quadratic relationship is given by

$$\tilde{N} = 0.3856\tilde{C}^2 + 1.0021\tilde{C} - 0.0098153, \quad (12)$$

due to the polynomial coefficients given in Table 4.

Quadratic relations between the solar irradiance and the relative sunshine duration have been reported in the literature

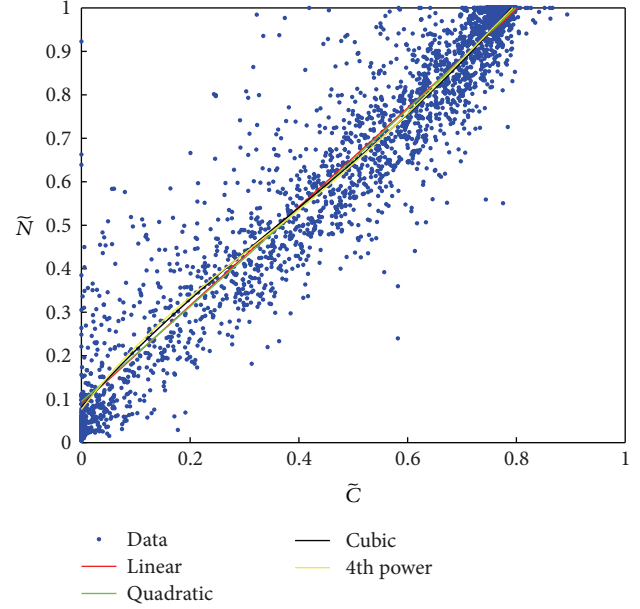


FIGURE 3: Weather mast: different regression types for plot 2.

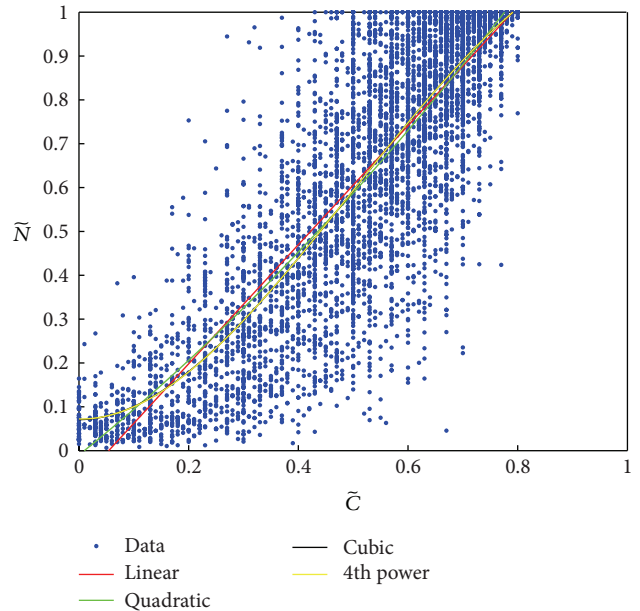


FIGURE 4: Airport: cloudiness against cloud shade with different regression types.

[9]. How this translates to a possible relation between solar irradiance and cloudiness is discussed by [10]. Aksoy et al. [1], on the other hand, found a linear relation between the sunshine duration and a satellite derived cloud index. This result is certainly not directly comparable to our data from surface observations. We did not examine any cloud index.

3.3. Remarks on the Results. The traditional assumption that $\tilde{N} = C$ is not borne out by our data. Particularly, clear skies are associated, on average, with some cloud shade. We

TABLE 2: Polynomial coefficients for Figure 3.

Factor	P_0	P_1	P_2	P_3	P_4
Linear	1.136	0.088	—	—	—
Quadratic	0.067086	1.0804	0.094391	—	—
Cubic	0.83046	-0.95267	1.3959	0.081687	—
4th power	-1.842	3.8559	-2.5221	1.6578	0.076275

TABLE 3: RMSE for Figure 4.

	Linear	Quadratic	Cubic	4th power
RMSE	0.2324	0.2291	0.2333	0.2334

therefore reexamine the formulas in the previous section, writing, from (5) and (8),

$$C(t, \Delta t)(1 + \epsilon) = 1 - \sigma(t, \Delta t), \quad (13)$$

with a very small positive ϵ . This means that on average the estimates should obey the inequality $\tilde{N} > \tilde{C}$, with the difference diminishing with increasing cloudiness (cf. (6)).

Another problem arises with the estimated value for the cloud shade, which is, as we saw before, always smaller than the theoretical value. Thus, the relation between estimated values reads

$$\tilde{C}(t, \Delta t)(1 + \epsilon) + \gamma = 1 - \bar{\sigma}(t, \Delta t) \quad (14)$$

with $\gamma > 0$. Again, even for vanishing ϵ , this leads us to infer $\tilde{N} > \tilde{C}$, which is what the figures show. We may develop theoretical grounds to show how γ depends on the cloudiness but we will not do that here. We rewrite (14), the corrected relation between estimated cloud shade and observed cloudiness, to express empirically the threshold function

$$\gamma(\tilde{C}) = \tilde{N} - \tilde{C}(1 + \epsilon) \quad (15)$$

which clearly leads us to expect positive values of the cloud shade for vanishing cloudiness [3]. \tilde{N} can now achieve its highest value of 1 (no sunshine throughout the whole daytime) for a value of cloudiness less than that for total cloud cover. This is also borne out by the figures. Figure 3, for $\epsilon = 0$, gives $\gamma(0) \approx 0.1$ and $\gamma(0.75) \approx 0.25$. For the airport we get $\gamma(0) \approx -0.01$ and $\gamma(0.75) \approx 0.25$. The negative value of $\gamma(0)$ is within the uncertainty of the estimates. It can be set equal to zero.

If we had independent means to determine $\gamma(\tilde{C})$, say as a polynomial of the form

$$\gamma(\tilde{C}) = \gamma(0) + a\tilde{C} + b\tilde{C}^2 + c\tilde{C}^3, \quad (16)$$

then (15) would read, assuming ϵ to be negligible,

$$\tilde{N} = \gamma(0) + (1 + a)\tilde{C} + b\tilde{C}^2 + c\tilde{C}^3, \quad (17)$$

and this would justify the polynomial fits we have used above.

A possible further work would be to find a functional form for ϵ and hence (14), but we reserve this for another, more theoretical account.

4. Comparing the Two Stations

After having described our results we would like to compare cloud shade and cloudiness of the two stations. Because of the closeness of the two stations and the relative uniformity of topography between them, we would expect very similar results. Figure 5 compares the cloudiness at both the airport and the weather mast. We plotted one cloudiness against the other, using only 2,922 values (2004 to 2011) because of the missing values for the previous period at the weather mast. We expect a perfect correlation clustering tightly around the bisecting line, and for higher values of cloudiness that is what we see. However, for little cloudiness most of the data points lie below this line. This means that there is, on average, a bias towards greater values of cloudiness for the airport, as compared to the weather mast's values. A likely interpretation of this shift is to be sought in the human observers at the airport. Due to problems of the perspective, the human observer easily overestimates cloudiness, especially when there are clouds with strong vertical development further away from zenith. On the other hand, at the weather mast, the cloudiness is measured with a ceilometer and therefore is subject only to instrumental and possibly algorithmic errors. This explains the bias for lower cloud covers.

A better match with expectations is shown in Figure 6. Here, data points are closer to the bisecting line of the regression graph. Moreover, more data were available, namely, 5,844 daily mean values for the time period between 1996 and 2011. This means that the sunshine of one station is representative for Hamburg.

This would certainly also be the case with cloudiness, if instead of a human being at the airport there were instruments to record the cloudiness. The ideal case would have been almost identical equations for the relation of sunshine duration and cloudiness, but due to different measurement techniques and instruments, differences show up in the results. Still, the results found are roughly comparable for the two stations at a linear distance of 14.3 km, and hence the two stations can be accepted as representative for the city of Hamburg.

5. Conclusion and Outlook

Inspired by the results of Badescu [3] and Akinoglu [9] about a small-order polynomial relation between cloudiness and sunshine, we wanted to check their findings with data from Hamburg. We were led to choose two different polynomials for the relation between cloudiness and sunshine duration in terms of cloud shade. The results, which refer to daily means, tend to confirm their findings, even if there are differences

TABLE 4: Polynomial coefficients for Figure 4.

Factor	P_0	P_1	P_2	P_3	P_4
Linear	1.3582	-0.07360	—	—	—
Quadratic	0.3856	1.0021	-0.0098153	—	—
Cubic	-2.0641	3.1075	0.0071873	0.070847	—
4th power	0.17487	-2.3643	3.276	-0.026263	0.072324

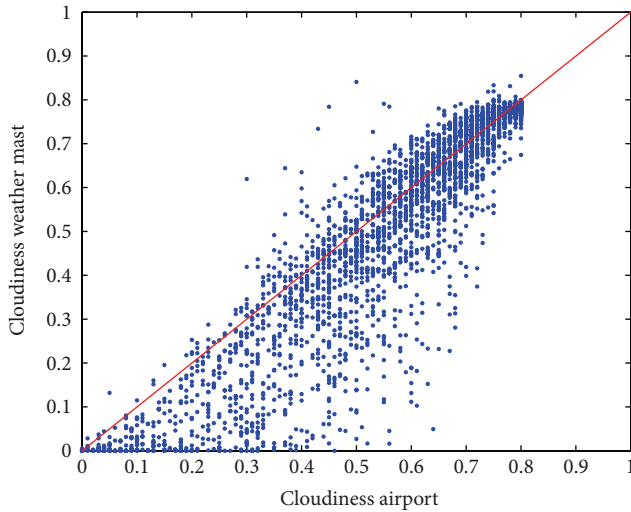


FIGURE 5: Comparison of cloudiness of airport and weather mast.

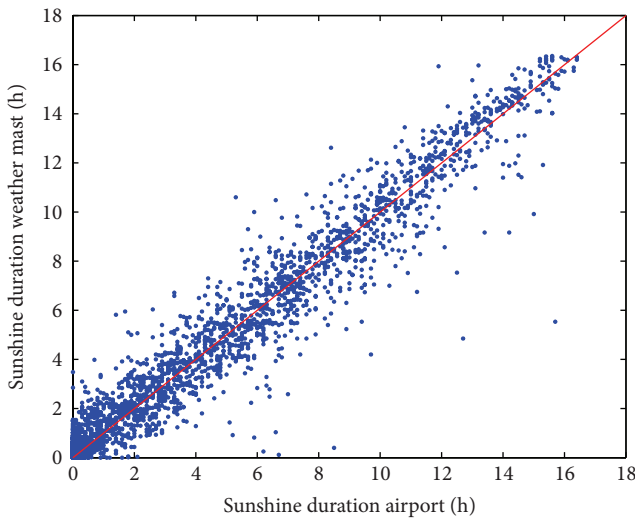


FIGURE 6: Comparison of sunshine duration of airport and weather mast.

between the two spatially close stations. In any event, the analysis should be repeated with data from comparable instruments and also with more data, before a single linear or quadratic relationship can be established as stable enough to be representative of Hamburg. Other relations are to be expected if shorter time periods are to be analyzed.

Badescu [3] found a linear relation between cloudiness and cloud shade for Romania. Of course, for cities with

other determining factors like a complicated relief, linear or quadratic relations are expected to become harder to confirm.

The relations suggested for Hamburg or Romania have practical applications, especially in the field of solar energy. With their help, it is possible to estimate amounts of solar energy reaching the Earth's surface, merely on the basis of cloud amount. A relation often used, the Angström-Prescott equation, can be adopted if there is just one parameter available at a specific location [1, 11]. With it, it is possible to calculate the daily insolation at locations where the measuring network does not provide all the necessary parameters.

The knowledge of incoming solar radiation at the Earth's surface is certainly helpful in deciding whether a solar plant will be cost-efficient, or which location is the most suitable for reducing costs. Many more applications are related to the topic of this paper, but our aim has been only to confirm either the classic or the newer relations between the quantities involved, by drawing on recent data from two reliable stations located in the cloudy city of Hamburg.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The author would like to thank Professor Thomas Frisius for his constant open ear and financial support within the Cluster of Excellence 177 Climate System Analysis and Prediction (CliSAP) in Hamburg. Special thanks go to Joachim Pelkowski for his many helpful discussions.

References

- [1] B. Aksoy, S. Ener Rüßen, and B. G. Akinoglu, "A simple correlation to estimate global solar irradiation on a horizontal surface using METEOSAT satellite images," *Turkish Journal of Engineering and Environmental Sciences*, vol. 35, no. 2, pp. 125–137, 2011.
- [2] V. Badescu and M. Paulescu, "Statistical properties of the sunshine number illustrated with measurements from Timisoara (Romania)," *Atmospheric Research*, vol. 101, no. 1-2, pp. 194–204, 2011.
- [3] V. Badescu, "Use of sunshine number for solar irradiance time series generation," in *Modeling Solar Radiation at the Earth Surface*, pp. 327–355, Springer, Berlin, Germany, 2008.
- [4] H. Suercke, "On the relationship between duration of sunshine and solar radiation on the earth's surface. Angstrom's equation revisited," *Solar Energy*, vol. 68, pp. 417–425, 2000.

- [5] B. Brümmer, I. Lange, and H. Konow, "Atmospheric boundary layer measurements at the 280 m high hamburg weather mast 1995–2011: mean annual and diurnal cycles," *Meteorologische Zeitschrift*, vol. 21, pp. 319–335, 2012.
- [6] WMO, *OMM No. 8, Guide to Meteorological Instruments and Methods of Observation*, part 1, World Meteorological Organization, Geneva, Switzerland, 6th edition, 1996.
- [7] V. Badescu, "A new kind of cloudy sky model to compute instantaneous values of diffuse and global solar irradiance," *Solar Energy*, vol. 61, pp. 251–264, 2002.
- [8] L. A. Santaló, *Integral Geometry and Geometric Probability*, vol. 1 of *Encyclopedia of Mathematics and Its Applications*, Addison-Wesley Publishing Company, Reading, Mass, USA, 1976.
- [9] B. G. Akinoglu, "Recent advances in the relations between bright sunshine hours and solar irradiation," in *Modeling Solar Radiation at the Earth Surface*, pp. 115–143, Springer, Berlin, Germany, 2008.
- [10] J. Pelkowski, "A physical rationale for generalized Ångström-Prescott regression," *Solar Energy*, vol. 83, no. 7, pp. 955–963, 2009.
- [11] A. Ångström, "On the computation of global radiation from records of sunshine," *Arkiv Foer Geofysik*, pp. 471–479, 1956.

