On the linear relationship between daily relative sunshine duration and daily mean cloud cover

Heinrich Morf, Senior Member ISES

Buechraiweg 47, 5452 Oberrohrdorf, Switzerland

Abstract

This paper explores the relationship between daily relative sunshine duration – *S* and daily mean cloud cover – *C*. A theoretical model that leads to the idealized linear relationship *S=1-C* often encountered in literature is derived, based on well-defined simplifying assumptions. We demonstrate the application of the theoretical model to vertical visibility and cloud cover readings by ceilometer. We explore the deviations from linearity in the real world in the context of the theoretical model. An estimate of the relationship between measurements of *S* and *C* based on sunshine recorders, ground based human observers, and ceilometers is given.

Keywords: ceilometer, climatology, cloud cover, sunshine, stochastic modeling

1. Introduction

The opinion that the relationship between daily relative sunshine duration – *S* and daily mean cloud cover – *C* would be *S=1-C* under ideal conditions is often encountered in literature. This is a very convenient assumption because it opens the way for many elegant mathematical simplifications. However, in reality a non-linearity between S and C is observed. See Biga and Rosa (1980), Brabec et al. (2016), Gueymard et al. (1995), Hoyt (1977), Rangarajan et al. (1984), and Sarkar (2016). To our knowledge, no convincing description of the conditions leading to *S=1-C* was ever given. This paper fills this argumentation gap. First, the linear relationship is theoretically derived, based on well-defined simplifying assumptions. Then reasons for the non-linear relationship encountered in the real world are given.

We base our theoretical considerations on the stochastic solar irradiance model by Morf (2013). In this model cloud cover – *cc(t)* is modeled as a strictly stationary stochastic process that is invariant to the observation area. Then, sunshine – *ss(t)* is generated in a random experiment with probability *1-cc(t)*. We demonstrate that, over the period from sunrise to sunset – T_d, the mean of *ss(t)* – *S* can be approximated by the mean of *1-cc(t)* – *1-C*. As *S* and *1-C* are both random variables, this implies that ideally also their probability distributions are the same.

Our theoretical model takes a strict nadir view of the sky, and clouds have no vertical extent. To the contrary, real world observations take a single point view of the sky dome. In consequence, sunshine and cloud cover observations are affected by the vertical extent of clouds and the height of the cloud layers in the sky. This disrupts the ideal conditions assumed by the theoretical model and leads to the observed non-linear relationship between *S* and *C*.

2. Problem statement

Prior to going into details, some definitions are necessary.

Definition of daily relative sunshine duration – *S*:

$$S = \frac{\int_{SR}^{SS} ss(t)}{T_d} \qquad \text{eq. (1)}$$

ss(t) is the stochastic insolation function. As an indicator function, it is 1 when the sun is shining and 0 when the sun is hidden behind clouds. T_d is the time between sunrise – *SR* and sunset – *SS*.

doi:10.18086/eurosun.2016.09.09 Available at http://proceedings.ises.org

Definition of daily mean cloud cover – *C*:

$$C = \frac{\int_{SR}^{SS} cc(t) \cdot dt}{T_d} \qquad \text{eq. (2)}$$

cc(t) is cloud cover. This is the fraction of the sky that is covered by clouds; 0 represents a completely cloudless sky, 1 a fully cloud covered sky. T_d is the time between sunrise – *SR* and sunset – *SS*.

Our objective is to demonstrate that the following approximation holds true:

$$S = \frac{\int_{SR}^{SS} ss(t) \cdot dt}{T_d} \approx 1 - C = \frac{\int_{SR}^{SS} (1 - cc(t)) \cdot dt}{T_d} \qquad \text{eq. (3)}$$

For this purpose, we prove that

$$\frac{\int_{0}^{T} ss(t) \cdot dt}{T} \sim \frac{\int_{0}^{T} (1 - cc(t)) \cdot dt}{T} \quad T \to \infty \qquad \text{eq. (4)}$$

The equation states that the two terms converge with increasing observation time *T*.

We then verify up to what degree the approximation by eq. (3) is met, when delimiting *T* by T_d in eq. (4).

3. A theoretical interpretation for the linear relationship between daily relative sunshine duration and daily mean cloud cover

We will prove eq. (4) by demonstrating that the expectation – μ of its first term is equal to its second term. Then, we will show that the variance – σ^2 of the first term converges towards 0 with growing *T*. The leading idea is that a random variable with zero variance is a certainty. Thus, the result of the two terms in eq. (4) must finally match.

For the proof we take a discrete approach, in concordance with the model of Morf (2013), where the random experiment for sunshine is periodically executed with a time interval Δt, corresponding to the minimum time resolution for sunshine encountered in the real world.

Morf (2013) models sunshine by executing a random experiment that generates sunshine from cloud cover with a probability equal to the complement to one of cloud cover:

$$P_{ss=1}(t) = 1 - cc(t) \qquad \text{eq. (5)}$$

$P_{ss=1}(t)$ is the probability for sunshine. *cc(t)* is cloud cover.

Expectation and variance of *ss(t)*:

$$\begin{aligned} \mu(ss(t)) &= 0 \cdot cc(t) + 1 \cdot (1 - cc(t)) = 1 - cc(t) \\ \mu(ss^2(t)) &= 0 \cdot cc(t) + 1 \cdot (1 - cc(t)) = 1 - cc(t) \\ \sigma^2(ss(t)) &= \mu(ss^2(t)) - \mu^2(ss(t)) = (1 - cc(t)) \cdot cc(t) \end{aligned} \qquad \text{eq. (6)}$$

Integrating the expectation of *ss(t)* over *T* and normalizing by *T* leads to the proof of the first proposition:

$$\mu\left(\frac{\int_{0}^{T} ss(t) \cdot dt}{T}\right) \approx \mu\left(\frac{\sum_{n=1}^{N} ss_n}{N}\right) = \frac{\sum_{n=1}^{N} (1 - cc_n)}{N} \approx \frac{\int_{0}^{T} (1 - cc(t)) \cdot dt}{T} \qquad \text{eq. (7)}$$

whereby $N = T/\Delta t$.

To demonstrate the convergence of the variance of the first term of eq. (4) towards zero we proceed as follows:

$$\sigma^2\left(\frac{\int_0^T (ss(t))\cdot dt}{T}\right) \approx \sigma^2\left(\frac{\sum_{n=1}^{N} ss_n}{N}\right) = \frac{\sum_{n=1}^{N}(1-cc_n)\cdot cc_n}{N^2} \qquad \text{eq. (8)}$$

whereby $N=T/\Delta t$. With increasing N the variance given by eq. (8) converges towards zero. Thus, one may write:

$$\frac{\int_0^T ss(t)\cdot dt}{T} \sim \frac{\int_0^T (1-cc(t))\cdot dt}{T} \qquad T\to\infty \qquad \text{Q.E.D.} \qquad \text{eq. (9)}$$

As the two terms of eq. (9) converge, so will also their probability distributions. We refer the reader to Morf (2011) for an analysis and to Morf (2014) for an analytical expression of this probability distribution.

Executing the random experiment for sunshine every 30 seconds (corresponding roughly to the minimum time resolution of sunshine in the real world) in the model described by Morf (2013) leads to a spread in the plot of S against C that is smaller than *0.05* for both, S and C. Thus, on a daily basis, eq. (3) may be used with little loss of accuracy. See Fig. 1.

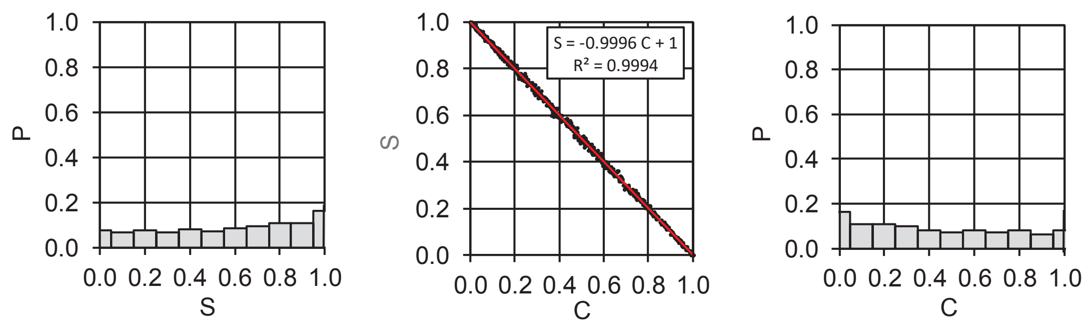

Fig. 1: Simulated plot *S=g(C)* of daily relative sunshine duration – *S* as a function of daily mean cloud cover – *C* and the probability bar charts of *S* and *C*. Observe the little spread of the plot, the high correlation coefficient of the least square fit, and the equality of the probability bar charts of *S* and *1-C*.
Simulation results for Payerne, Switzerland, in July. 300 points, roughly the July-days of 10 years, are shown. A time resolution of 30s was used for the simulation.

4. Vertical visibility and ceilometer readings

The model of Morf (2013) that generates sunshine – *ss(t)* by a random experiment with probability equal to the complement to one of cloud cover – *1-cc(t)* (which is the basis of our theoretical considerations) is an approximation of another model given by Morf (2011). It generates vertical visibility – *vv(t)* with a probability equal to the complement to one of cloud cover – *1-cc(t)*. *vv(t)*, as defined by Morf (2011), describes the state of the sky (0 for covered, 1 for clear) at the observer's zenith. The only difference between the two models is that one is defined for sunshine – *ss(t)* and the other for vertical visibility – *vv(t)*. As *vv(t)* takes a strictly zenithal view, the observations are not affected by the vertical extent of clouds. Therefore, it fits the theoretical model better than *ss(t)*.

Thus, under the assumptions of our theoretical model one may write for eq. (9):

$$\frac{\int_0^T vv(t)\cdot dt}{T} \sim \frac{\int_0^T (1-cc(t))\cdot dt}{T} \qquad T\to\infty \qquad \text{eq. (10)}$$

The equation states that the two terms converge with increasing observation time T.

Returning to eq. (3), we write:

$$S_{vv} = \frac{\int_{SR}^{SS} vv(t)\cdot dt}{T_d} \approx 1 - C_{vv} = \frac{\int_{SR}^{SS} (1-cc(t))\cdot dt}{T_d} \qquad \text{eq. (11)}$$

S_{vv} is the daily relative vertical visibility duration.[1]

vv(t) also corresponds to the reading of a ceilometer that indicates either 0 (covered) or 1 (clear). Thus, eq. (10) is the fundamental equation for the estimation of cloud cover by ceilometer readings.

Cloud cover by ceilometer readings is calculated as the complement to one of the moving average of the ceilometer readings over the period *T*. Thus, it is the moving average of *cc(t)* over *T*, and not *cc(t)*. The integration time – *T* of a ceilometer is small in comparison to the rate of change of cloud cover. Therefore, the difference between the two values is negligible.[2] Daily mean cloud cover derived from ceilometer readings – C_c is then the mean over T_d of the moving average of *cc(t)* over *T*. Note that this procedure, contrary to eq. (11), will deviate from reality, even under the ideal conditions of the theoretical model.

The automatic determination of cloud cover by ceilometer is established practice. In the USA, it has replaced cloud cover observations by ground-based human observers at the end of the twentieth century.

5. The non-linearity

The non-linearity under discussion is an upward bend of the straight line for *S=g(C)* in Fig. 1. See also the least square fits for $S_s=g(C_g)$ in Fig. 2.

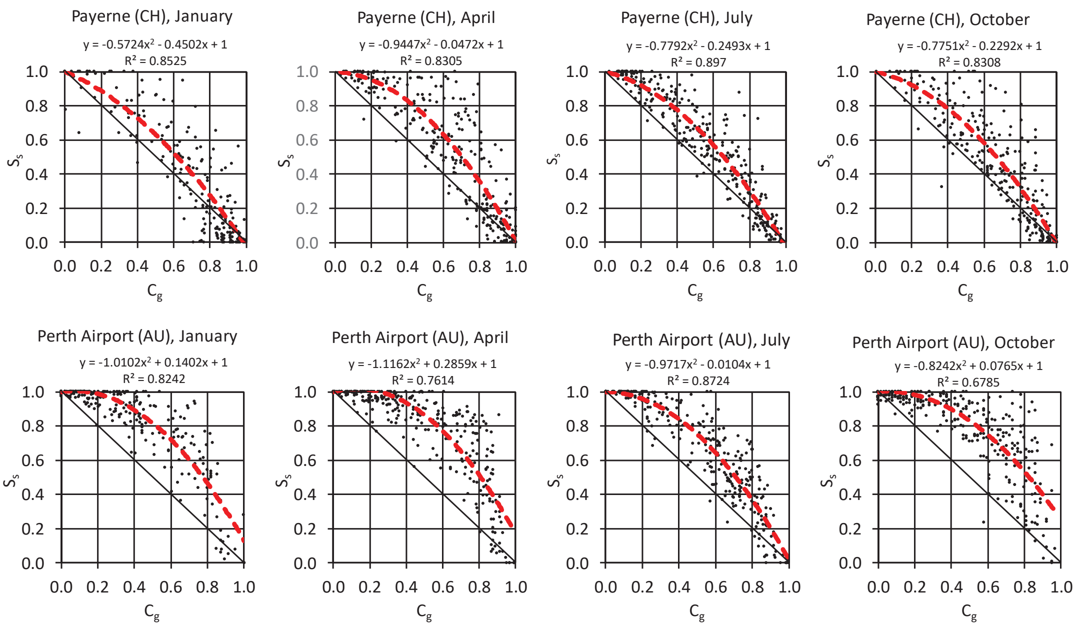

Fig. 2: Plots $S_s=g(C_g)$ of daily relative sunshine duration measured by sunshine recorder – S_s as a function of daily mean cloud cover reported by ground based human observers – C_g. Observe the slight upward bend of the least square fits.

Observe the subscripts introduced for *S* and *C* in Fig. 2. The subscript *s* indicates that a value was derived from sunshine recorder readings, and the subscript *g* indicates that it was derived from cloud cover observations by ground-based human observers.

[1] For the sake of compatibility with the theoretical model, we prefer to use the mnemonic S_{vv} (instead of, for example, *V*) for daily relative vertical visibility duration.

[2] For the determination of cloud cover, ceilometers use a sampling interval – *Δt* of about 30 seconds and an integration time *T* of about 30 minutes.

It is worth the while to have a closer look at the two variables S_s and C_g:

- Daily relative sunshine duration from the traces of sunshine recorders – S_s: This is the daily relative sunshine duration from sunrise to sunset, measured by a sunshine recorder at a single point on earth, in the line of sight to the sun. It is well known that the free line of sight to the sun depends on the elevation of the sun in the sky due to the vertical extent of clouds and the height of the cloud layers in the sky (Lund and Shanklin, 1971).
- Daily mean cloud cover from observations by ground-based human observers – C_g: The objective is to determine the proportion of the sky dome that is obstructed by clouds when seen from a single point on earth – $cc_g(t)$. Every single cloud cover observation – $cc_g(t)$ is an instant average of cloud cover over the entire sky dome. Daily mean cloud cover – C_g is then the mean of cloud cover – $cc_g(t)$ from sunrise to sunset.

These two variables are not related by the linear relation $S=1-C$, but by the slightly up bent curves of the least square fits depicted in Fig. 2. It is well known that for a function of two random variables $Y=g(X)$ – here $S_s=g(C_g)$ – the probability density functions (*pdfs*) – $f(x)$ and $f(y)$ are related by:

$$f(y)\cdot|dy| = f(x)\cdot|dx| \qquad \text{eq. (12)}$$

The vertical bars indicate absolute values.[3]

Thus, in consequence of the up-bent function $S_s=g(C_g)$, the pdf of S_s as well as of C_g will be skewed towards higher values than in the linear relationship $S=1-C$. Hence, the mean of S_s becomes greater than the mean of $1-C_g$.

Fig. 3 depicts probability bar charts of S_s and $1-C_g$ with the corresponding means for selected calendar months in Payerne, Switzerland and Perth Airport, Australia. It follows from the foregoing paragraph that the pdf of S_s is skewed to the right and the one of $1-C_g$ to the left. As to be expected, the mean of S_s is always greater than the mean of $1-C_g$. (Means are marked with overbars.)

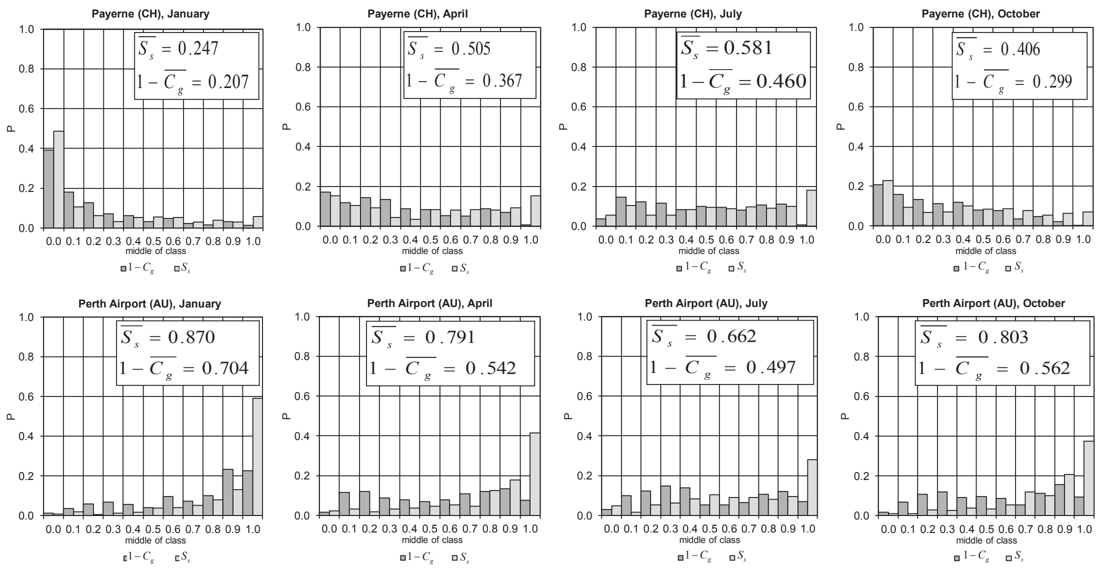

Fig. 3: Comparison of the probability bar charts of daily relative sunshine duration measured by sunshine recorder – S_s and the complement to one of mean daily cloud cover measured by ground based human observers – 1-C_g. Observe that the mean of S_s is higher than that of $1-C_g$. Overbars indicate means.

A concentration of readings of 1-C_g around $0.1 \ldots 0.4$ keeps the mean of $1-C_g$ low. Cloud cover observations by ground-based human observers seem not to reflect properly the free line of sight to the sun through open cloud fields. A further disturbance that keeps the mean of $1-C_g$ low is caused by the reluctance of ground

[3] As an alternative, to the least square regressions for $Y=g(X)$ in Fig. 2, one may obtain a regression line for $Y=g(X)$ from $f(x)$ and $f(y)$, based on eq. (12). It brings $f(x)$ and $f(y)$ in proper relation. Although the objectives of the two regression methods are different, they lead to only slightly different results here.

based human observers to declare minimum cloud cover of 0 okta (tenth), because they are instructed to do so only when the sky is absolutely clear. On the other hand, excessive readings for $S_s=1$ keep the mean of S_s high. Hoyt (1977) attributes this effect to thin cirrus clouds: while a ground based human observer would see a covered sky, a sunshine recorder would still register sunshine.

6. Delimiting results

As for the relation $S=1-C$ of the theoretical model, imagine that corresponding to daily relative cloud cover from observations by ground based human observers – C_g, there is daily relative sunshine duration – $S_g=1-C_g$. Likewise, imagine that corresponding to daily relative sunshine duration from sunshine recordings – S_s, there is mean daily cloud cover – $C_s=1-S_s$. Consequently, for the case of the up bent relation $S_s=g(C_g)$ encountered in the real world, the following inequalities hold true: [4]

$$1-\overline{S_s} < \overline{C} < \overline{C_g} \ , \ 1-\overline{C_g} < \overline{S} < \overline{S_s} \ , \ whereby \ \ \overline{C} = 1-\overline{S} \qquad \text{eq. (13)}$$

The overbars indicate that the values are expectations. $\overline{C}$ and $\overline{S}$ correspond to the theoretical model; they might have been estimated from vertical visibility readings – $vv(t)$ by ceilometer. Fig. 4 depicts the inequality for $\overline{C}$ at Payerne, Switzerland and Perth Airport, Australia.[5]

The inequalities may be helpful for data cross checks and to estimate a parameter when others are known. Each of the three results – $C_s=1-S_s$, $C=1-S$, and $C_g=1-S_g$ – is also correct in its own right. One might as well chose data of that observation method whose results fit best a specific problem at hand. For example, Morf (2011) suggests to use cloud cover observations by ground based human observers when the emphasis is on diffuse irradiation and data from traces of sunshine recorders when the interest concentrates on good results for direct irradiance. Climatologists might prefer data from zenithal views by ceilometer.

Hoyt (1977), analyzing data from 72 locations in the USA, concludes that $1-S_s$ is a better estimate for C than C_g. He reasons that $1-S_s$ comes closer to cloud cover estimates from satellite observations. This result must be viewed with caution, because the viewing angle of satellites down to earth is mostly not in strict nadir direction.

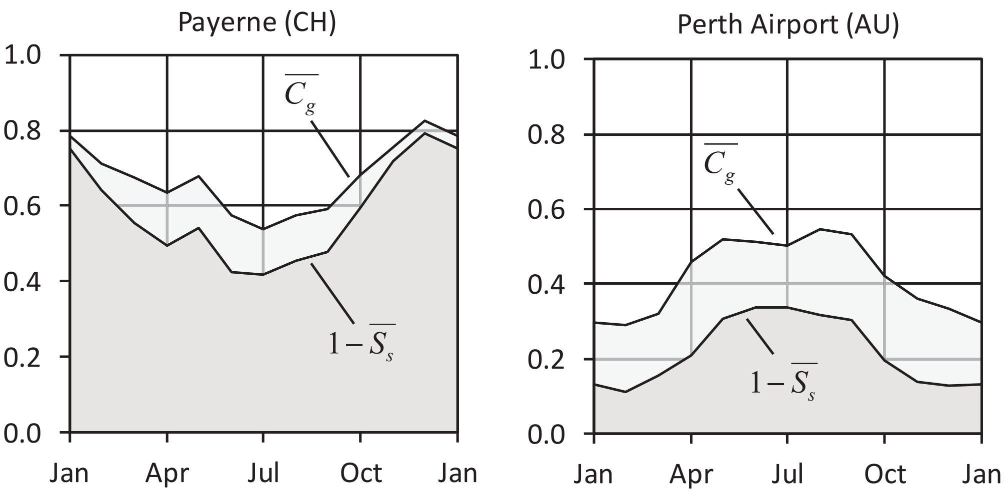

Fig. 4: Delimiting the expected daily mean cloud cover – $\overline{C}$.
The value of $\overline{C}$ must lie within the upper boundary $\overline{C_g}$ and the lower boundary $1-\overline{S_s}$.

---

[4] Moriarty (1991) predicts a down bent relationship for $S_s=g(C_g)$ at high latitudes. The inequalities would have to be modified to cover this case. So far, we have not encountered this effect in the real world.
[5] The irregularity in May for Payerne is a consequence of the outbreak of the volcano Eyjafjallajökull in April 2010.

7. Verification of the inequalities

So far, we could not thoroughly verify the inequalities of Section 6 in the real world, because there are few records of the three variables that have been taken at the same location, at the same time, and over a sufficiently long time-period. However, Neske (2014)[6] presents two data cloud diagrams of daily relative sunshine duration – S as a function of daily mean cloud cover – C that have been taken at two stations that are both representative for the area of Hamburg, Germany.

- One was taken from 2004 to 2011 at the NDR Transmitter Mast in Hamburg Billwerder by the Meteorological Institute of the University of Hamburg. S_s was calculated from pyranometer readings and C_c from ceilometer readings.
- The other was taken from 1996 to 2011 at Hamburg Airport by the German National Meteorological Service – DWD. S_s was derived from the recordings of an electronic SONI sunshine switch and C_g from observations by ground-based human observers.

The S_s values of the two stations are statistically undistinguishable, whereas the C values are different. For Hamburg Airport the function $S_s=g(C_g)$ has the characteristic upward bend as in Fig. 2. For Hamburg Billwerder the linear relation $S_s=1.088-0.9088\cdot C_c$ is a good fit. As expected, the use of the ceilometer data at Hamburg Billwerder instead of the estimates by ground-based human observers has brought the relation more within reach of the relation $S=1-C$ of the theoretical model.

8. Concluding remarks and outlook

The findings in this work contribute to the understanding of the stochastic behavior of sunshine and cloud cover, and indirectly also of solar irradiance. We hope that they will facilitate the estimate of model parameters (Morf, 2011; Morf, 2013), and will give new impulses to the challenge of generating realistic two-dimensional cloud cover patterns as a function of time (Morf, 2011). As already brought to attention in Morf (2014), there seems to be one single Markov process that drives most cloud related dynamic processes. We hope this observation will give rise to interest for further investigations.

As a byproduct, this paper provides support for the approach by Morf (2011) of modeling cloud cover as a strictly stationary stochastic process that is invariant to the observation area. Otherwise, determining cloud cover from single point ceilometer readings would be impossible.

The relationship between the differing results given in form of inequalities in Section 6. might be helpful for a seamless transition from historical records of cloud cover estimates by ground based human observers to the automated cloud cover estimates based on ceilometer readings being introduced starting at the end of the last century.

Our proof that the mean of cloud cover converges to the mean of relative sunshine duration with increasing length of the observation period deserves further mathematical evaluation and exploration. It describes a transition of a random variable to a certainty.

In mathematical terms:

Given an indicator function $ind(t)$ with probability $P_{ind=1}(t)$, then

$$\frac{\int_0^T ind(t)\cdot dt}{T} \sim \frac{\int_0^T P_{ind=1}(t)\cdot dt}{T} \qquad T\to\infty \qquad \text{eq. (14)}$$

The equation states that the two terms converge with increasing observation time – T.

As a simple example, consider the discrete case of tossing a coin. Then the probability of receiving *head* – $P_{ind=1}(t)$ will be 0.5, and with increasing number of tosses the probability to have a *head* in the tosses will

[6] Caution must be taken in Neske (2014): The paper uses an uncommon linear scale for daily mean cloud cover – C with a value of 0 for a totally cloud free sky and a value of 0.8 for a fully cloud covered sky.

turn into certainty. The consequences of this statement cannot be overestimated. The reasoning "if the probability for an event is infinitely small, then it will never happen" should rather be exchanged against the statement "even if the probability for an event is infinitely small, it will certainly happen".

9. Data

Publicly available data sets of the following locations are used in this paper:

- Payerne, Switzerland, 46.80 N / 6.93 E, for the years 2001-2010
- Perth Airport, Australia, 31.93°S / 115.98° E, for the years 1998-2007

From both stations, records of daily sunshine duration from sunshine recorders are used. The rare days with unreliable or missing data were removed. The time between sunrise and sunset was set to the duration of the period where the sun is more than 5° above the horizon.

From both stations, cloud cover data observed by ground-based human observers are used. At both stations, observations are taken every three hours starting at zero hours local time. The rare days with unreliable or missing data were removed. Cloud cover observations of 9 okta were changed to 8 okta. Daily mean cloud cover was calculated as the average of cloud cover from sunrise to sunset using linear interpolation; the result in *okta* was converted to *tenth*, dividing by 8. For every calendar month, local time for sunrise and sunset was determined based on the average maximum possible sunshine duration.

Conflict of Interests

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

Acknowledgements

The author wishes to thank the Western Australian Climate Service Centre for the supply of a ten-year data set of Perth Airport, containing daily sunshine duration, daily global irradiation, and total cloud cover observations.

The author wishes to thank the Swiss Federal Office of Meteorology and Climatology – Meteo Swiss for the supply of a ten-year data set of Payerne, containing daily sunshine duration, daily global irradiation, and total cloud cover observations.

Nomenclature

Capital Letters

C	daily mean cloud cover [-] without index it corresponds to the theoretical model, or it applies generically
N	counter limit [-]
P	probability [-]
R^2	regression coefficient
S	daily relative sunshine duration [-] without index it corresponds to the theoretical model, or it applies generically
SR	sunrise [s]
SS	sunset [s]
T	time, duration [s]
V	daily relative vertical visibility duration [-]

Small Letters

cc	cloud cover [-]
g	generic function
f	probability density function – pdf
ind	generic indicator function
n	generic counter [-]
ss	sunshine [-]
t	time [s]
x	generic variable
y	generic variable
vv	vertical visibility [-]

Special Characters

$\mid\ \mid$	absolute value
$\bar{\ }$	expectation, mean
~	convergence
≈	approximation

Greek Letters

μ	mean, expectation
σ^2	variance

Subscripts

c	derived from ceilometer readings
d	from sunrise to sunset
g	derived from cloud cover by ground based human observers
s	derived from sunshine recorder readings
vv	derived from vertical visibility

References

Biga, A.J., Rosa, R., 1979. Contribution to the study of the solar radiation climate of Lisbon. Solar Energy 23, 61–67.

Brabec, M., Badescu, V., Paulescu M., Dumitrescu, A., 2016. A new perspective on the relationship between cloud shade and point cloudiness. Atmospheric Research 172-173, 136-146.

Gueymard, C., Jindra, P., Estrada-Cajigal, V., 1995. A critical look at recent interpretations of the Ångström approach and its future in global solar radiation prediction. Solar Energy 54, 357-363.

Hoyt, D.V., 1977. Percent of possible sunshine and the total cloud cover. Monthly Weather Review 105, 648–652.

Lund, I.A., Shanklin M.D., 1971. Universal methods for estimating probabilities of cloud-free lines-of-sight through the atmosphere. Journal of Applied Meteorology 12, 28-35.

Morf, H., 2011. The stochastic two-state cloud cover model STSCCM. Solar Energy 85, 985-999.

Morf, H., 2013. A stochastic solar irradiance model adjusted on the Ångström-Prescott regression. Solar Energy 87, 1-21.

Morf, H., 2014. The parameterization of cloud cover. Energy Procedia 57, 2014, 1293-1298.

Moriarty, W.W., 1991. Cloud cover as derived from surface observations, sunshine duration, and satellite observations. Solar Energy 47, 219-222.

Neske, S., 2014. About the relation between sunshine duration and cloudiness on the basis of data from Hamburg. Journal of Solar Energy, 2014, Article ID 306871, 7 pages.

Rangarajan, S., Swaminathan, M.S., Mani, A., 1984. Computation of solar radiation from observations of cloud cover. Solar Energy 12, 553-556.

Sarkar, N.I., 2016. Estimation of solar radiation from cloud cover data of Bangladesh. Renewables: water, wind, and solar 3:11, 1-15.