Solar Energy 153 (2017) 508–518

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

# Statistical properties of clear and dark duration lengths

Marek Brabec [a,b], Marius Paulescu [c,*], Viorel Badescu [d,e]

[a] Department of Nonlinear Modeling, Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodarenskou vezi 2, 182 07 Prague 8, Czech Republic
[b] Czech Institute of Informatics, Robotics, and Cybernetics, Czech Technical University in Prague, Zikova 1903/4, 166 36 Prague 6, Czech Republic
[c] Physics Department, West University of Timisoara, V. Parvan 4, 300223 Timisoara, Romania
[d] Candida Oancea Institute, Polytechnic University of Bucharest, Spl. Independentei 313, 060042 Bucharest, Romania
[e] Romanian Academy, Calea Victoriei 125, Bucharest, Romania

A R T I C L E   I N F O

*Article history:*
Received 21 March 2017
Received in revised form 10 May 2017
Accepted 1 June 2017

*Keywords:*
Sunshine duration
Clouds
Solar irradiance
Statistical lifetime modeling
Cox regression
Censored data

A B S T R A C T

The focus of this study is on the properties of consecutive intervals consisting of pairs of clear periods (the Sun is shining) and dark periods (the Sun is obscured by clouds). Namely, the distribution of the duration length of clear and dark periods is studied and illustrated with results obtained for the climatic regime of Timisoara (Romania, Eastern Europe). Usual simple moment-based characteristics such as means and variances cannot be unambiguously estimated due to censoring. Seasonal changes in the length distributions have been studied through the (nonparametric) Kaplan-Meier estimates. Both dark and clear duration distributions are different among months. The correlation between clear and dark periods within one dark-clear pair is of substantial interest. Therefore, a rigorous approach based on Spearman's rank correlation coefficient is used instead. The dependence of duration distribution upon covariates is studied using the Cox regression model, which fits the hazard ratio incurred by a unit change in the explanatory variables. It is shown that increasing the extraterrestrial irradiance on horizontal surface and the clearness index during dark/clear periods increases significantly the dark/clear period end risk. The Cox regression model enables more complicated analyses, such as the non-additive effects (i.e. interactions) of several covariates. The clearness index of a dark period (CID) has a much stronger effect than the extraterrestrial irradiance on horizontal surface of a dark period (EID) and the CID-EID interaction is still significant after the interactions when the months are included.



## 1. Introduction

Sunshine duration is a meteorological variable measuring the absolute time when the Sun is shining within a given period. Nevertheless, the relative sunshine duration is often used for various purposes. It is defined as the ratio of the actual duration of sunshine within a given period to the duration of that period. Sunshine duration plays a significant role in many aspects of economic and social life. Solar radiation, when the Sun is shining, it drives the performance of solar power plants, it affects agriculture, it influences human health through its UV component, and so on. Long-term measurements showed that between 1950 and mid-1980s, the sunshine duration decreased in Europe. This "dimming" period was followed by a reversal trend (the brightening period) which appears to be still active (Black et al., 2006; Wild, 2012; Kambezidis et al., 2016).

Sunshine duration is traditionally measured by the Campbell-Stokes burning method, or by the pyranometric method (WMO, 2008). Places where direct sunshine duration measurements are available are relatively scarce. Aiming to overcome the problems of low and irregular spatial coverage of the meteorological stations equipped with sunshine recorders, some studies tried to relate sunshine duration with simple empirical observations. For example, in the 80's, Barbaro et al. (1981) proposed a linear correlation between the relative sunshine and *the monthly mean state of the sky* quantified by the numbers of the clear, mixed and overcast days in every month. The idea was resumed ten years later by Hussain and Siddique (1990) and Badescu (1990, 1991), and more recently also by Essa and Etman (2004), Neske (2014), Brabec et al. (2016), among others.

While there are occasions when actual sunshine duration needs to be known or forecasted in short time intervals (e.g. for real-time balancing of electricity network containing photovoltaic power plants), many real-world applications require knowledge of the long-term statistical properties of sunshine duration. Archetypal

* Corresponding author.
*E-mail address:* marius.paulescu@e-uvt.ro (M. Paulescu).

http://dx.doi.org/10.1016/j.solener.2017.06.003
0038-092X/

examples might include heating system design, agriculture, holiday resort placement, water evaporation calculation for reservoir design, etc. Long-term (ergodic) average of the relative sunshine can serve as the simplest characteristic of interest. For many purposes, deeper insight is required and knowledge of variability and even finer distribution properties is needed, however. For example, in solar energy estimation, probably one of the most famous equations relates linearly the solar irradiation at ground level to sunshine duration. This is the Ångström equation (Ångström, 1924). In almost one hundred years, substantial amount of work has been done in the interpretation and improvement of the Ångström equation. A self-consistent derivation of the Ångström equation, highlighting its virtues and liabilities, has been reported recently by Stefu et al. (2016). Giovanelli (1979) evaluated the probability of getting a required number of clear days at Kitt Peak National Observatory during an observing period of given length. This study was motivated by the fact that a substantial part of the solar observing time at Kitt Peak National Observatory is allocated to short-term visitors who may be present for only a few days. If the skies are cloudy, their observation sessions are doomed. Therefore, the probability of sunshine periods is of more interest than usual. There are also studies that investigate the effects of sunshine duration on suicide rates (Papadopoulos et al., 2005; Vyssoki et al., 2012). Therefore, various facets of the sunshine duration were investigated from different perspectives. A brief review of some studies devoted to the understanding of the sunshine duration properties and its relation with other parameters (meteorological, radiometric) follows.

Within the last few decades the connection between relative sunshine and total cloud cover was intensively investigated. Robaa (2008) deduced three empirical models to estimate relative sunshine duration using observed cloud amount in Egypt. The author found a significant correlation between relative sunshine duration and observed cloudiness, with correlation coefficients greater than 0.978 for the three regions analyzed. Kandirmaz et al. (2014) proposed three different artificial neural network models for estimating monthly mean daily values of sunshine duration for Turkey. Brabec et al. (2016) estimated the distribution of cloud shade (a quantity complementary to relative sunshine) for a given cloud cover value.

From a different perspective, many studies were dedicated to the statistical analysis of sunshine duration data series on different time scales. Barbaro et al. (1984) computed frequency distributions of daily relative sunshine for different Italian locations. Zabara and Zenginoglou (1989) got close to dynamical modeling viewpoint and estimated the conditional probability of the daily relative sunshine for each month using historical data recorded in Athens. Practically, they estimated the probability that the relative sunshine on a day from a certain month is greater than $x$, given that the relative sunshine on the previous day is greater (smaller) than $x$. In the language of discrete time continuous space Markov chains, these conditional probabilities correspond to the transition probabilities like $P(X_{t+1} > x|X_t \leqslant x)$ where $X_t$ denotes the relative sunshine on a day $t$. Matzarakis and Katsoulis (2006) evaluated the temporal and spatial distribution of sunshine duration and relative sunshine duration over Greece. The authors concluded that the distribution and the overall patterns of sunshine duration seem to remain unchanged during the last 40 years. This persistence in sunshine duration strongly motivates the studies on sunshine duration statistics. Badescu and Paulescu (2011) evaluated the statistical properties of the relative sunshine by means of sunshine number (SSN), a binary indicator of whether the Sun is shining or not which is carefully derived from irradiance measurements. This quantity is introduced in Section 2. The relative sunshine may be well approximated as the average value of sunshine number. Paulescu and Badescu (2011) introduced the sunshine stability number (SSSN), as another binary indicator for counting the transition from sunny to not sunny periods or vice versa and hence evaluating the SSN variability.

In this paper, we look at the distribution of sunshine duration from a different and somewhat more detailed perspective. Since we want to investigate deeper and more complicated properties than just the relative sunshine (the average of sunshine number over an interval of desired length), we look separately at clear periods (the Sun is shining) and dark periods (the Sun is obscured). These periods correspond to consecutive intervals (clear and dark periods alternate by the definition). The length of these intervals is random. The daylight part of a day is partitioned by the union of these random intervals (the two boundary intervals containing sunrise and sunset have somewhat different properties to be discussed later). We do not have any meaningful information during night time.

The procedure proposed in this paper for constructing the clear and dark duration data series is based on radiometric measurements (taken with sufficiently fine time resolution) and sunshine number as a precursor of sunshine duration. Details of this procedure will be formalized in Section 2. The procedure gives us a series of alternating clear and dark periods, which we subsequently analyze from several viewpoints related to various distributional properties of those period durations. We are interested not only in the distribution of the period length *per se*, but also in how it changes with various explanatory variables (a kind of regression problem) and also in whether the distributions of subsequent period lengths are somehow positively or negatively related (an analogue to autoregression). The main results are discussed in Section 3. The paper is methodological in its purpose. We show the limitation of common statistical tools very often used by practitioners. First, due to censoring, simple moment-based characteristics, such as means and variances, cannot be unambiguously defined. Kaplan-Meier (nonparametric) estimates are used to see how the length distributions change seasonally. The correlation between clear and dark periods within one dark-clear pair is of interest but common tools, such as the Pearson correlation coefficient, cannot be used. A rigorous approach needs to take into account censoring. Further, the Cox regression model is used to study the dependence of the duration distribution upon important covariates.

To the best of our knowledge, this is the first study on this subject. The conversion of the usual sequence of solar irradiance measurements into a structured database of clear/dark periods is novel in the solar radiation research.

## 2. Data and methods

### 2.1. Database

Global and diffuse solar irradiance ($G$ and $G_d$, respectively) recorded on the Solar Platform of the West University of Timisoara (Romania) are used in this study (Solar Platform 2017). The town of Timisoara (45°46′N, 21°25′E, 85 m a.s.l.) has a warm temperate climate, fully humid, with warm summer, typical for the Pannonian Basin (Köppen climate classification *Cfb*). This is based on the Kottek et al. (2006) digital Köppen-Geiger world map on climate classification, build with data from the second half of the 20th century.

On the Solar Platform the measurements are performed all day long at sampling time intervals of $\Delta\tau = 15$ s. DeltaOHM LP PYRA 02 first class pyranometers which fully comply with ISO 9060 standards and meet the requirements of the World