Atmospheric Research 212 (2018) 227–239

Contents lists available at ScienceDirect

Atmospheric Research

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

Methodology for optimizing a photosynthetically active radiation monitoring network from satellite-derived estimations: A case study over mainland Spain

Vindel J.M.*, Valenzuela R.X., Navarro A.A., Zarzalejo L.F.

Renewable Energy Division, CIEMAT, Av. Complutense, 40 28040 Madrid, Spain

ARTICLE INFO

Keywords:
Photosynthetically active radiation
Clustering analysis
K-means algorithm

ABSTRACT

A methodology is presented for determining optimal locations to install photosynthetically active radiation (*PAR*) measurement stations. Initially, a cluster analysis was performed from *PAR* satellite-derived estimations over mainland Spain. Once the optimal number of clusters was obtained, the total number of locations included in the monitoring network was distributed among different groups, according to the size and variability of each group. Finally, the specific locations for measurement stations placement was determined using an iterative technique: The largest region within each cluster was split into two new sub-regions, providing two new sites for substituting the initial location.

Clustering analysis has previously been applied to determine locations to monitor solar radiation. However, this is the first implementation for *PAR* stations in mainland Spain. Another novelty developed in this work is the distribution employed for specific sites within each cluster. The outcome achieved using clustering analysis was compared to those obtained using three other methods: two methods without clustering analysis and the third where clustering is performed but not optimizing the number of clusters. In one technique without clusters, the largest region is split into two new sub-regions, similar to the clustering analysis with optimization. In the second without clustering analysis, since the data variability was not previously addressed, the region divided is those with the largest combined effect of variance and size. The results fully justify using a clustering process; however, clustering without optimization is the worst performing method.

1. Introduction

Photosynthetically active radiation (*PAR*) is the radiation from 400 to 700 nm of the waveband. As part of the solar spectrum, *PAR* is attenuated along its track through the atmosphere due to the constituents, including clouds (Kazantzidis et al., 2011). Characterizing *PAR* (Sivakumar and Virmani, 1984) is important to applications devoted to biomass production, energy balance in ecosystems, plant physiology, algae production, natural illumination of greenhouses, and carbon drains. Empirical expressions derived from its correlation with global solar radiation are a practical mechanism for estimating *PAR* (Akitsu et al., 2015; Alados et al., 1996; Escobedo et al., 2009; Foyo-Moreno et al., 2017; Hu et al., 2007a; Jacovides et al., 2015; Sudhakar et al., 2013; Sunday et al., 2016; Zhang et al., 2000).

Another source of *PAR* estimations is from satellite observations (Liang et al., 2006; Rubio et al., 2005; Zheng et al., 2008). For example, CM-SAF, which is part of EUMETSAT, provides Kato bands (Kato et al., 1999); *PAR* can be obtained from Kato bands 7–16 (Table 1).

However, satellite-derived estimates can have uncertainties (Qin et al., 2012), highlighting the need to develop ground-based *PAR* monitoring networks.

PAR can be measured directly with instruments on the ground (Ross and Sulev, 2000; Wood et al., 2015); however, although measuring different components of solar radiation is becoming more frequent, monitoring networks for the measuring *PAR* are unusual. In Spain, a wide network of stations belonging to the State Meteorological Agency provides data for diffuse, direct and global radiation, but identifying locations for *PAR* measurements is very difficult. Thus, with the purpose of establishing a station network for measuring this parameter, this study is devoted to determining installation locations for these stations.

Determining locations for station installation is highly dependent on the spatial variability of solar radiation (Bois et al., 2008; Gallegos and Lopardo, 1988; Polo et al., 2015; Riihimaki and Long, 2014), which arises from variations in atmospheric and weather conditions. The deterministic component of the solar radiation due to astronomical

* Corresponding author.
E-mail address: josemaria.vindel@ciemat.es (J.M. Vindel).

https://doi.org/10.1016/j.atmosres.2018.05.010
Received 7 August 2017; Received in revised form 11 April 2018; Accepted 21 May 2018
Available online 26 May 2018
0169-8095/

Table 1
Kato bands in the *PAR* region of the spectrum.

Kato band	Wavelength range (μm)
7	0.408–0.452
8	0.452–0.518
9	0.518–0.540
10	0.540–0.550
11	0.550–0.567
12	0.567–0.605
13	0.605–0.625
14	0.625–0.667
15	0.667–0.684
16	0.684–0.704

influences (Iqbal, 1983) can be removed by dividing the radiation reaching Earth by the corresponding value at the top of the atmosphere, which provides "clearness index" or transmissivity:

$$k_t = \frac{GHI}{G_0} \tag{1}$$

where *GHI* is the Global Horizontal Irradiance and G_0 the extra-terrestrial radiation. By analogy, we can define a clearness index for the range covered by *PAR* ($k_{t\ PAR}$) (Jacovides et al., 2010). Indeed, in the case of *PAR*, the value can be normalized by the part of the extra-terrestrial radiation spectrum, including the range corresponding to *PAR* (0.4–0.7 μm). The obtained index, $k_{t\ PAR}$, only contains the atmospheric aspects affecting *PAR*. Therefore, identifying suitable monitoring stations locations should be based on the spatial behavior of this variable, $k_{t\ PAR}$.

The methodology for identifying monitoring network station locations includes several steps. Initially, variable estimates obtained from satellites are used as first approximations of $k_{t\ PAR}$, which can be used to develop all subsequent processes. Later, due to the spatial and temporal variability of this variable, a clustering analysis (Andenberg, 1973; Halkidi et al., 2001; Kaufman and Rousseeuw, 1990) provides the optimal number of cluster groups within which this variable is similar. In recent years, meteorological conditions have been characterized using clustering tools (Khedairia and Khadir, 2012; Tian et al., 2014), and more precisely, solar radiation time series have been analysed elsewhere using clustering methods (Gastón-Romeo et al., 2011; Ghayekhloo et al., 2015; Jiménez-Pérez and Mora-López, 2016; Polo et al., 2015). In addition, regions for radiation measurement station installation have been estimated using these methods (Zagouras et al., 2014). Different techniques can be used for clustering analysis, such as the k-means algorithm (Adam and Celebi, 2013; Macqueen, 1967), and hierarchical clustering (D'Andrade, 1978; Greenacre and Primicerio, 2013). Both methods are used in this study and described in the methodology section. Once the regions with similar $k_{t\ PAR}$ have been obtained via clustering analysis, the total number of available hypothetical stations must be distributed among these regions. This distribution depends on the size and variability of the different regions; larger sizes and variability results in a greater number of required stations in the region.

Knowning the distribution of hypothetical stations within each region, the process is completed by obtaining precise station locations. An iterative method is implemented so at each step the largest region is divided into two new sub-regions. Thus, the initial location is replaced by two new sites, located at the centres of the new sub-regions. This method is based on the Variance Quadtree Algorithm (VQA), using the size of the strata instead of its variance. In each iteration, the VQA splits the stratum with the largest variance into four new rectangular strata (Minasny et al., 2007). Yang and co-workers (Yang and Reindl, 2015) used this iterative method to design a solar irradiance monitoring network in the United States. However, in our case, each region is only split into two new sub-regions so that new stations are added one at a time. In addition, because the variability was accounted for in the previous clustering analysis, the division is only based on the size of the strata. Anyway, the methodology presented in this work is compared with other methodologies that do not include clustering analysis and the variance criterion is considered in the strata division process.

Finally, from the sites obtained, a kriging process is performed to obtain values of the variable, $k_{t\ PAR}$, for the whole study grid. In this way, the initial values of the variable in the grid, estimated from satellite, can be compared with the values obtained by kriging, from the satellite estimates in the monitoring network points.

Notably, while clustering analysis has already been applied to analysing solar radiation, it has not been applied specifically to *PAR*, and more precisely, not in Spain. Furthermore, determining specific sites within each cluster to locate the hypothetical stations is unique to the distribution technique developed in this work. Finally, the study compares the outcomes from this new methodology with those obtained from three other methods. The first two methods do not use clustering analysis, and the final method uses as many clusters as the number of available stations and does not optimize the number of clusters. For the two methods without clustering analysis, the specific sites are identified by: a) splitting the largest region into two new sub-regions, similar to the clustering analysis technique; and b) splitting the region with the largest combined variance and size effect. Based on the study results, the new proposed methodology performs the best, while clustering using the same number of available stations as the number of clusters performs the worst.

2. Data

Data were provided by the Satellite Application Facility on Climate Monitoring (Müller and Trentmann, 2015; Müller and Trentmann, 2014) part of the EUMETSAT Satellite Application Facility network. The Spectral Resolved Irradiance (SRI) product contains the Kato bands, as CM SAF does not directly provide *PAR* data. The *PAR* estimates were obtained by summing those bands corresponding to the *PAR* region of the spectrum (0.4–0.7 μm) (Table 1). The data units are W/m^2. The MAGICSOL (MAGIC SOLar irradiance) method (Mueller et al., 2009) provides spectrally resolved solar surface irradiance on a horizontal plane. This method only needs satellite information from the broadband visible channel. Therefore, it can be applied to Meteosat First Generation and subsequent satellite generations. The MAGICSOLs method for clouds is based on the Heliosat method. The Heliosat algorithm uses reflection measurements given as normalized digital counts to determine the cloud index. This index is a relative measure of the reflectance measured on the sensor normalized to the range of values at a given pixel (Perez et al., 2002; Zelenka et al., 1999). It is determined from the following expression:

$$n = \frac{\rho - \rho_g}{\rho_c - \rho_g} \tag{2}$$

where ρ is the instantaneous planetary albedo, which is the reflectance measured by the satellite sensor; ρ_c is the cloud albedo associated with the reflectance of the brightest pixel; and ρ_g is the ground albedo, associated with the reflectance of the darkest pixel.

However, modifications to the original method are needed to obtain spectrally resolved irradiance. Spectral corrections to the broadband effective cloud albedo are performed by applying the Radiative Transfer Model and saved in a look-up table used to account for the spectral effect of clouds (Müller and Behrendt, 2013).

The requested area of coverage was 44 °N–35.3 °N latitude and 9.5 °W–3.5 °E longitude, covering the study zone (mainland Spain) with a spatial resolution of 0.1° × 0.1°. Daily data from 1991 to 2011 (21 years) was used. The high temporal resolution was reduced to decrease computational complexity and avoid fluctuations that introduce noise into this climatological study. Finally, the data were grouped into months so each grid point had 12 values.