

Article

# Assessment and Day-Ahead Forecasting of Hourly Solar Radiation in Medellín, Colombia

Julián Urrego-Ortiz <sup>1,\*</sup>, J. Alejandro Martínez <sup>1</sup>, Paola A. Arias <sup>1</sup>  
and Álvaro Jaramillo-Duque <sup>2</sup>

<sup>1</sup> Grupo de Ingeniería y Gestión Ambiental (GIGA), Escuela Ambiental, Facultad de Ingeniería, Universidad de Antioquia, Calle 67 No. 53-108, Medellín 050010, Colombia; john.martinez@udea.edu.co (J.A.M.); paola.arias@udea.edu.co (P.A.A.)

<sup>2</sup> Research Group in Efficient Energy Management (GIMEL), Departamento de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de Antioquia, Calle 67 No. 53-108, Medellín 050010, Colombia; alvaro.jaramillod@udea.edu.co

\* Correspondence: julian.urrego@udea.edu.co

Received: 29 July 2019; Accepted: 11 November 2019; Published: 19 November 2019



**Abstract:** The description and forecasting of hourly solar resource is fundamental for the operation of solar energy systems in the electric grid. In this work, we provide insights regarding the hourly variation of the global horizontal irradiance in Medellín, Colombia, a large urban area within the tropical Andes. We propose a model based on Markov chains for forecasting the hourly solar irradiance for one day ahead. The Markov model was compared against estimates produced by different configurations of the weather research forecasting model (WRF). Our assessment showed that for the period considered, the average availability of the solar resource was of 5 PSH (peak sun hours), corresponding to an average daily radiation of  $\sim 5$  kWh/m<sup>2</sup>. This shows that Medellín, Colombia, has a substantial availability of the solar resource that can be a complementary source of energy during the dry season periods. In the case of the Markov model, the estimates exhibited typical root mean squared errors between  $\sim 80$  W/m<sup>2</sup> and  $\sim 170$  W/m<sup>2</sup> ( $\sim 50\%$ – $\sim 110\%$ ) under overcast conditions, and  $\sim 57$  W/m<sup>2</sup> to  $\sim 171$  W/m<sup>2</sup> ( $\sim 16\%$ – $\sim 38\%$ ) for clear sky conditions. In general, the proposed model had a performance comparable with the WRF model, while presenting a computationally inexpensive alternative to forecast hourly solar radiation one day in advance. The Markov model is presented as an alternative to estimate time series that can be used in energy markets by agents and power-system operators to deal with the uncertainty of solar power plants.

**Keywords:** global horizontal irradiance (GHI); forecasting; clearness coefficient; Markov chains; weather research and forecasting model; solar resource

## 1. Introduction

The performance of solar power plants depends essentially on an adequate characterization of the variations of the incoming solar radiation over land surface [1]. This variation is mainly associated with the interaction between clouds and the incoming solar radiation, which leads to attenuation values that, in some cases, can reach 80% or higher [2]. Solar resource variations cause subsequent changes of the output at solar power plants that could not only affect the electric infrastructure but also the revenue models that govern energy supply [3].

In the case of the day-ahead energy market in Colombia, plants bid to offer energy blocks to the energy market national operator, XM ([www.xm.com.co](http://www.xm.com.co)), one day before obtaining the market clearing results [4]. The biddings for a certain day must be offered before 8 a.m. of that day. Based on these biddings, the energy-market operator determines which plants will supply the demand at each hour of

the next day. This shows that the short-term operation of solar power plants depends on the correct forecasting of the incoming solar radiation and respective electricity generation. According to the UMPE (the Colombia governmental entity in charge of designing energy expansion plans in Colombia), in the 2021–2029 planning horizon, there are scenarios where it is expected to have 329 MW of new solar plants distributed, that is about 2% of the actual power capacity in Colombia [5].

In addition, due to setbacks during the start up of the Hidroituango hydraulic plant [6], in the short-term it is expected to have new solar plants and distributed generation based on solar photovoltaic energy in Colombia.

Information about the hourly evolution of global horizontal irradiance (GHI) can be obtained through the characterization of solar radiation in the site of study [4,5,7]. Such characterization is also useful to provide benchmarking information that can be used to assess the performance of the GHI estimates obtained with different models (e.g., dynamical models [1,8] statistical models [9]). When using these models, characterization of solar resource is also necessary for model calibration as well as the quantification of their related uncertainties.

In general, the GHI can be estimated using two types of models: dynamical models and statistical models (including machine-learning techniques). The dynamical models are physically driven models that estimate the GHI from the physical relationships that exist between solar radiation and other atmospheric variables. Dynamical models like the weather research and forecasting (WRF) model [10] estimate the state of the atmosphere by numerically solving the atmosphere Equations for large horizontal domains (i.e., synoptic- and meso-scales) that are discretized in elements that usually comprehend several kilometers, even with convection-permitting resolution. However, the use of dynamical models is still a challenge, not only in terms of the hardware, computational time and knowledge required; but also because, in spite of producing physically-consistent results, their estimates may exhibit large biases [11,12].

As a counterpart, there exist the statistical models for forecasting GHI. These models include elements from time-series analysis and assume that the future series are statistically similar to the past series. This means that the estimates of a statistical model will mostly reflect the common features of the measured series used to train it. Therefore, statistical models depend on the size and quality of the available measurements to produce realistic estimates of the GHI. Different statistical and machine-learning models can be used to estimate GHI and the decision of which model should be used greatly depends on the features of the estimates (i.e., the temporal resolution of the estimates and the lead times at which they are needed). Among the simplest statistical models are the regression-based models. These models show an adequate performance for forecasts that require estimates with very coarse temporal resolutions, ranging from weeks to months. It is also a common practice to use these models not as forecasting models but rather diagnostic models that predict the daily or monthly solar radiation based on other available atmospheric variables, like daily or monthly temperature, which could have been previously measured or forecasted with a different model [13–15].

Currently, the use of machine-learning models like artificial neural networks (ANN) have gained a greater importance in the forecasting of solar radiation. For instance, reference [16] uses a multi-layer perceptron for estimating the global daily radiation for one day ahead. Reference [17] used a combination of an Auto Regressive Moving Average model (ARMA) and a time-delay neural network (TDNN) to forecast GHI for a lead time of 1 h, with a time resolution of 10 min. Reference [18] used an ARMA-based model and an ANN model to forecast GHI, using two types of forecasts: a first forecast with a temporal resolution of 10 min and different lead times, starting from 10 min to 60 min, and a second forecast with a temporal resolution of 1 h and lead times from 1 h to 6 h. As noted by [19] and the previously mentioned studies, the majority of these approaches exhibit their highest performance at estimating the GHI when the temporal resolutions of the estimates are of the same order than the lead time of the forecast.

Other types of models that have been used previously for estimating the GHI are the Markov based models, which mainly focus on estimating the clearness coefficient and from it, calculate the corresponding GHI values. These models have been widely used for generating synthetic GHI series

that exhibit the same statistical characteristics as the GHI records in the site of study at different time resolutions [9,20–22]. Therefore, this type of model has become very useful for modeling solar energy systems in addition to being computationally inexpensive, which allows them to produce several realizations in a very short time. One example of this is the work of [9], where minutely radiation series were used to train a Markov chains models to produce synthetic series of GHI at high temporal resolution. In that work, even when the error values were high, the intention was to produce series that would behave, statistically, in the same way as the measured series, which poses as an advantage for photovoltaic (PV) system simulations considering the high temporal resolution of the realizations. In more recent works, Markov based approaches for modeling GHI are now being used to forecast the GHI values and not to only generate synthetic GHI series. One example is the case of [23], who proposed a hybrid model based on Markov chains and the Mycielski approach to forecast 1-h ahead GHI and found a satisfactory performance, in some cases even exceeding ANN models. Reference [24] proposes a Markov switching model that can be used to produce day-ahead forecasting of GHI. The model proves to have an adequate behavior at estimating the hourly data. However, it is based on a persistence approach for estimating the initial state of the next day to be estimated, which is an approach that does not work satisfactorily for sites with large variability in cloudiness. Reference [25] proposed a combined model based on k-means and Markov chains to statistically model the transition of the daily solar irradiance and characterize the transition probabilities among different states. In [26], a Markov-chain mixture model was proposed. The model was formulated to perform very high temporal resolution forecast of the clear-sky index (minutely resolutions).

In general, Markov-chain based approaches are not frequently used in the mid-term forecast (day-ahead forecasting) of the hourly series of GHI, which is a type of forecasting that is fundamental in the operation of generation systems based on solar radiation. They are rather used for generating synthetic GHI series for solar system simulations, or to perform forecasts with time horizons near the time resolution of the model (i.e., few time steps into the future). However, through correct implementation, a simple Markov based model could be used to perform mid-term forecasts of GHI hourly data with satisfactory performance and still represent a simple and computational inexpensive solution.

Therefore, as an alternative to obtain mid-term forecasts of the hourly GHI, we implemented a two-part model based on discrete Markov processes to estimate GHI in Medellín, Colombia, a highly populated city in the tropical Andes. This model is a modified version of the formulations found in the works of [9,21,22,27] and is capable of forecasting diurnal series of hourly GHI for one day-ahead, by taking into consideration the current seasonal effects over the behavior of the variable. Although the Markov transition matrices (MTM) calculated in this work reflect the particular features of the clearness coefficients in the region of study, the criteria followed for its construction and training can be applied to other locations. In order to provide a point of reference regarding the performance of the proposed Markov model, we used two additional models as benchmarking for the GHI estimates. One is the numerical weather prediction (NWP) model called the weather research forecasting model (WRF), and the second one is a modified Markov model based on persistence.

This work also provides insights on the intra-day behavior of the solar resource in Medellín, Colombia, which is not currently available for this region, using GHI measurements from a pyranometer operated by the Early Alert System of the Aburrá Valley (SIATA; <https://siata.gov.co>) during the period March 2016 to February 2017. This information is helpful since tropical regions exhibit a relative uniform income of radiation throughout the year compared to higher latitudes but can exhibit relatively high variability throughout the day. Also, knowledge about the intra-daily variability of the GHI is fundamental when formulating and calibrating models to characterize and forecast the incoming solar radiation.

Additionally, even when only one year of data was initially available and no inferences regarding annual and inter-annual variations of the GHI can be made, a single year of hourly records contains enough samples of hourly GHI series under different sky conditions, thus providing an important

source for studying the intra-day behavior of the GHI in the region of interest. This variability can be related to the presence of mountains since they are related to localized formation of clouds due to orographic lifting and elevated heat sources, which adds complexity to the simulation and forecast of GHI [28].

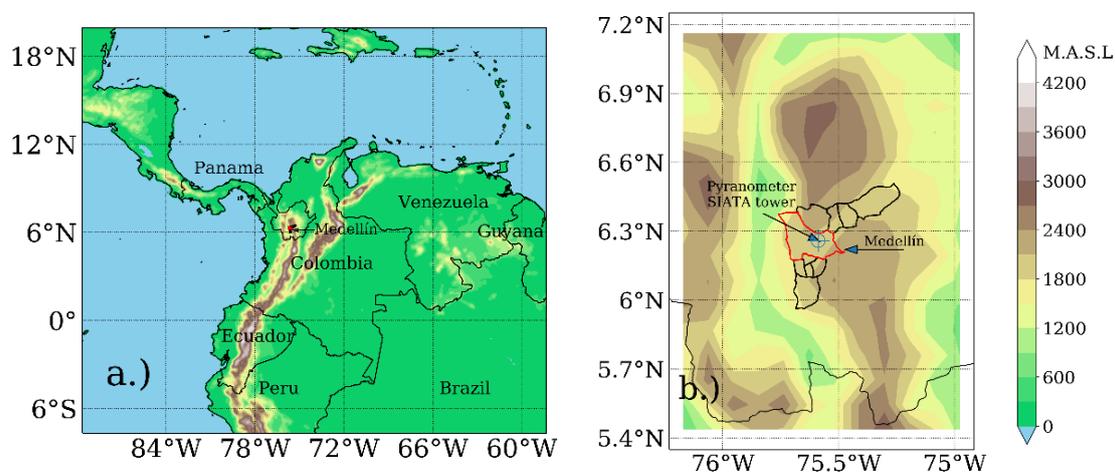
Thus, we developed a benchmarking study for the further assessment of the accuracy of solar radiation estimates obtained from different types of models in Medellín, Colombia, as well as for providing information about the behavior of the intra-day GHI. The key contributions of our work are the following:

- We performed an initial assessment of the intra-day and daily variability of the solar resource in Medellín, Colombia, which is a piece of information that was not available in the site of study and that is necessary for determining the intra-day generation potentials.
- We proposed a statistical model based on Markov chains for forecasting the hourly GHI series for one day-ahead lead time, with low computational costs. We also used an NWP model (WRF) and a persistence-based Markov model as a benchmarking for the proposed Markov model.
- We evaluated the performance of the Markov model at estimating the hourly GHI and daily clearness coefficient considering different cloud covers in a tropical climate region.
- The performance of the Markov model was also evaluated under local atypical and synoptic atypical cloudy conditions.
- The method used for the formulation and training of the Markov model can be extended to other locations with different climatological conditions.
- The magnitude of the errors obtained with the Markov model are comparable to the errors obtained with other models identified in the literature.

## 2. Data and Methods

### 2.1. Pyranometer Data

We used in situ measurements provided by SIATA (<https://siata.gov.co>). The instruments used to retrieve these measurements correspond to Kipp and Zonen SMP11 pyranometers, which record GHI values with a precision of  $1 \text{ W/m}^2$ . The pyranometer considered (hereafter SIATA station) is located in Medellín, Colombia, at  $6.2593^\circ$  latitude and  $-75.5887^\circ$  longitude (Figure 1). The SIATA pyranometers are inspected monthly and are calibrated according to the ISO 9847:1992. For the hourly values, the expected uncertainty is of 3% of the true value of radiation. Measurements are provided at a one-minute time resolution.



**Figure 1.** Location of the site of study and the pyranometer used. (a) Location of Colombia in South America. (b) Location of Medellín in Colombia. Color shades represents topography.

We performed a statistical analysis of the data provided by the SIATA station. The period of analysis ranges from 1 March 2016, to 28 February 2017. Although a single year is not enough for characterizing the seasonal variability of GHI in Medellín, it contains enough samples of hourly GHI series useful for studying the intra-day variability of GHI under different sky conditions. Consequently, the statistical analysis of GHI developed in this work is focused on the intra-day variations of the hourly GHI and not on the monthly or annual variations of the solar radiation. However, given the seasonal changes of solar radiation in the region, we discriminated our statistical analysis for each month of the year. In central Colombia, where our region of study is located, precipitation shows two wet seasons, one during March–April–May and a second one during September–October–November. By contrast, two dry seasons are observed during June–July–August and during December–January–February. This behavior is mainly associated with the latitudinal migration of the Intertropical Convergence Zone [29–32].

Since the characterization presented here is focused on the hourly values of GHI, the 1-min records provided by the SIATA pyranometer are averaged around each of the day-time hours. In this case, if 10% or more of the 1-min records within an hour were missing values, the corresponding hourly average value was set as a missing value. Less than 8% of the total hourly data corresponds to missing values, which was considered a small fraction. However, when a missing value was identified, usually the remaining values of the corresponding diurnal cycle were also missing values. This means that if data imputation is attempted, records corresponding to entire days would have to be produced. This is undesired since it could result in introducing high biases in the statistical analysis and, therefore, in the performance of the stochastic model.

## 2.2. Clearness Coefficient Estimation

The clearness index or clearness coefficient,  $k_t$ , is a dimensionless index that can be used to indirectly describe the behavior of solar radiation and is calculated as shown in Equation (1):

$$k_t = \frac{GHI}{I_{ext}} \quad (1)$$

where  $I_{ext}$  is the extraterrestrial radiation calculated as,

$$I_{ext} = I_{sc} \left( 1 + 0.033 \times \cos \left( \frac{360 \times \text{day of the year}}{365} \right) \right) \times \cos \theta_z \quad (2)$$

In Equation (2),  $\theta_z$  is the zenith angle, which depends on the declination angle  $\delta$ , the latitude  $\varphi$  and the hour angle  $\omega$ .  $I_{sc}$  is the solar constant and is equal to  $1367 \text{ W/m}^2$ . The zenith angle can be estimated at the same resolution of the GHI measurements.

The clearness coefficient represents the fraction of the extraterrestrial radiation that reaches the land surface after traversing the atmosphere. According to [33], the GHI can be assumed to consist of two components: a deterministic component, which is represented by the extraterrestrial radiation and the effects of the air mass coefficient (A.M) on the GHI, and a stochastic component which is primarily associated to the effects of clouds over the GHI. The effects of the A.M (a deterministic component) and the stochastic component are both inherited by the clearness coefficient when calculated as shown in Equation (1).

The deterministic component of  $k_t$  is, in general, better understood than the stochastic component and is mostly associated to changes in the path length that solar radiation must traverse before reaching the land's surface. This length changes during the day and is a function of the zenith angle,  $\theta_z$ . The corresponding geometrical change in the solar radiation path is what is represented by A.M, which according to [34,35], is calculated as shown in Equation (3).

$$A.M = \begin{cases} \frac{1}{\cos \theta_z} & \text{For } 0^\circ \leq \theta_z < 70^\circ \\ \frac{\exp(-0.000118 \times \text{Altitude}(\text{sea level}))}{\cos \theta_z + 0.5057(96.080 - \theta_z)^{-1.634}} & \text{For } 70^\circ \leq \theta_z < 90^\circ \end{cases} \quad (3)$$

On the other hand, the stochastic component is still not represented by simple Equations. Different studies have identified that the stochastic contribution over the GHI is mainly associated to the interactions between clouds and radiation [33,36–38]. An estimate of the stochastic component of  $k_t$  can be obtained through the normalization expression proposed by [39], who outlined the dependency between the hourly  $k_t$  and the A.M. According to [39], the  $k_s$  is obtained by removing the dependency of the  $k_s$  to the A.M as:

$$k_s = \frac{k_t}{1.031 \exp\left(\frac{-1.4}{0.9 + \frac{9.4}{A.M}}\right) + 0.1} \quad (4)$$

The normalized clearness coefficient,  $k_s$ , represents solely the effect of clouds over the GHI and behaves as a stochastic variable. In this work, we used a Markov model to simulate hourly values of  $k_s$ .

In addition to the calculation of  $k_s$ , another clearness coefficient is obtained at daily temporal resolutions and is referred to as the daily clearness coefficient,  $k_d$ . This value is also calculated from Equation (1) using the daily values of GHI and  $I_{ext}$ , and is then used to represent the average sky conditions of a given day.

### 2.3. Discrete Markov Chain Model

Markov chains describe the transition process of a random variable, where the probability distribution of the following state of the variable depends on its previous states. The degree of the process indicates the number of previous observations on which the next state of the variable statistically depends. Given that the random variable,  $X$ , has a set  $S$  with a finite number of possible states  $m$ , so that  $S = \{s_1, s_2, s_3, \dots, s_m\}$ , the random variable has a state  $i$  at a time step  $t$  when  $X_t = i$ . The transition probabilities between all possible states are stored in a MTM,  $P$ . In a first-degree Markov process, the next state of the variable  $X$  depends only on its current state and so the transition probability of the variable from state  $s_i$  to state  $s_j$ ,  $P_{ij}$  in a single time-step can be written as:

$$P_{ij} = P(X_t = j | X_{t-1} = i, X_{t-2} = i_{t-2}, \dots, X_0 = i_0) = P(X_t = j | X_{t-1} = i) \quad (5)$$

The MTM of such a process,  $P$ , is a squared matrix with dimensions  $m \times m$ , where the row values indicate the current state of the variable and the column values correspond to the next possible states of the variable. Therefore, there are  $m$  elements in each row of  $P$  e.g., the first row should be  $p_{11}, p_{12}, p_{13}, p_{14}, \dots, p_{1m}$  as shown in (6).

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1m} \\ p_{21} & p_{22} & \dots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \dots & p_{mm} \end{bmatrix} \quad (6)$$

Since the Markov chain transitions between discrete states, the clearness coefficient must be transformed from a continuous variable to a discrete variable. This is done by dividing the range of the clearness coefficient (i.e., 0–1) into equally-spaced intervals or bins. Each interval is enumerated in ascending order, with each number corresponding to a discrete state of the clearness coefficient. A state of the clearness coefficient refers to the interval in which its continuous value is contained.

The number of states is seen to greatly improve the performance of the model up to a certain number of states. Initially, ref [40] demonstrated that a set of discrete states  $m = 20$ , corresponded to the smallest interval subdivision that would still result in a regular behavior of the MTM. Furthermore, in a more recent study, ref [41] found through a cross-validation process, that the performance of a Markov based model did not improve the performance of the model significantly for a number of states larger than 20. In other words, it was evidenced that the model did not improve when the bin width of the discretization process went below 0.05. These findings are consistent with the discretization widths used in previous works, regarding the modeling of the clearness coefficient

using Markov based models [9,20,22,36,42,43]. For these reasons, we discretized both the normalized clearness coefficient ( $k_s$ ) and the daily clearness coefficient ( $k_d$ ), into 20 discrete states of width 0.05. The resulting discrete states are shown in Table 1.

**Table 1.** Discretization of the daily and hourly clearness coefficients.

| Discrete State | Continuous Interval |
|----------------|---------------------|----------------|---------------------|----------------|---------------------|----------------|---------------------|
| State 1        | 0–0.05              | State 6        | 0.25–0.3            | State 11       | 0.5–0.55            | State 16       | 0.75–0.8            |
| State 2        | 0.05–0.1            | State 7        | 0.3–0.35            | State 12       | 0.55–0.6            | State 17       | 0.8–0.85            |
| State 3        | 0.1–0.15            | State 8        | 0.35–0.4            | State 13       | 0.6–0.65            | State 18       | 0.85–0.9            |
| State 4        | 0.15–0.2            | State 9        | 0.4–0.45            | State 14       | 0.65–0.7            | State 19       | 0.9–0.95            |
| State 5        | 0.2–0.25            | State 10       | 0.45–0.5            | State 15       | 0.7–0.75            | State 20       | 0.95–1              |

Additionally, the works of [44,45] demonstrated that the statistical behavior of a set of  $k_s$  depends, predominantly, on the overall sky conditions of the day they belong to (i.e., on the state of the corresponding  $k_d$ ). This indicates that the modeling of the  $k_s$  must be discriminated by the states of the  $k_d$ . This procedure has been used before in other studies using discrete states to estimate GHI. Initially, reference [43] introduced the idea of a library of MTMs, where they calculated an MTM for the daily clearness coefficient, for each state of the monthly clearness coefficient. This choice was made based on the similarities of the shapes of the probability functions alone. Furthermore, reference [20] showed that the probability density function (PDF) of sets of  $k_s$  for days that have  $k_d$  values grouped in intervals of 0.05 displayed the same statistical behavior and therefore could be modeled through the same function. Following these studies, we opted to calculate a library of MTM for modeling  $k_s$ , one for each state of the daily clearness coefficient  $k_d$ .

It must be noted that a synthetic time series of hourly  $k_s$  values can be obtained with this Markov model by setting an initial state of  $k_s$  and calculating the following state as a stochastic process with the associated probabilities contained in the row of  $P$  given by the initial state of  $k_s$ . This procedure will be explained in more detail in the following sections. However, before explaining the GHI estimation procedure, we describe the methodology to estimate the MTMs, for both the  $k_d$  and  $k_s$ .

#### 2.4. Construction of the Markov Transition Matrices

##### 2.4.1. First-Degree Markov Transition Matrix (MTM) for $k_s$

As it was explained in Section 2.3, a first-degree Markov transition matrix contains the probability of the variable transitioning from a state  $i$  to a state  $j$  in one time step which can be expressed as  $P_{ij}$ . This probability  $P_{ij}$  can be approximated by counting the number of times the variables go from  $i$  to  $j$  in one time step,  $f_{ij}$  and then dividing this number by the total number of transitions the variable makes from state  $i$  to any other state,  $T_i$ .

$$P_{ij} = \frac{f_{ij}}{T_i} \quad (7)$$

with,

$f_{ij}$ : Number of times the variable passes from  $i$  to  $j$ .

$T_i$ : Number of times the variable departs from  $i$ .

For a given state  $z$  of the daily clearness coefficient ( $k_d$ ), the corresponding  $k_s$  MTM is calculated as follows:

- (1) Extract all the hourly GHI values from the days of the dataset that have a  $k_d = z$ . Keep the information regarding the hour and day of the year that corresponds to each hourly datum.
- (2) Calculate the corresponding  $k_s$  values using Equations (1)–(4), and discretize them using Table 1.
- (3) Using the discrete  $k_s$  data, calculate the corresponding  $f_{ij}$  and  $T_i$  values for all  $k_s$  discrete values. The  $i$  and  $j$  values are iterated over all the states of  $k_s$ .

- (4) Calculate the  $P_{ij}$  probabilities with Equation (7), and position them in the MTM at the corresponding  $(i, j)$  position.

This procedure can be iterated over all  $k_d$  states to obtain a first-degree MTM for each state of the daily clearness coefficient.

#### 2.4.2. Second-Degree MTM for $k_d$

Since the  $k_d$  is sensitive to atmospheric variations at synoptic scales, estimates of the daily clearness coefficient are obtained from a second-degree Markov chain calculated at daily resolutions (i.e., the future state of the  $k_d$  depends on the states of the previous two days). The construction of a second-degree MTM is performed in the same way as the first-degree MTM with the exception that the transition probability is calculated as:

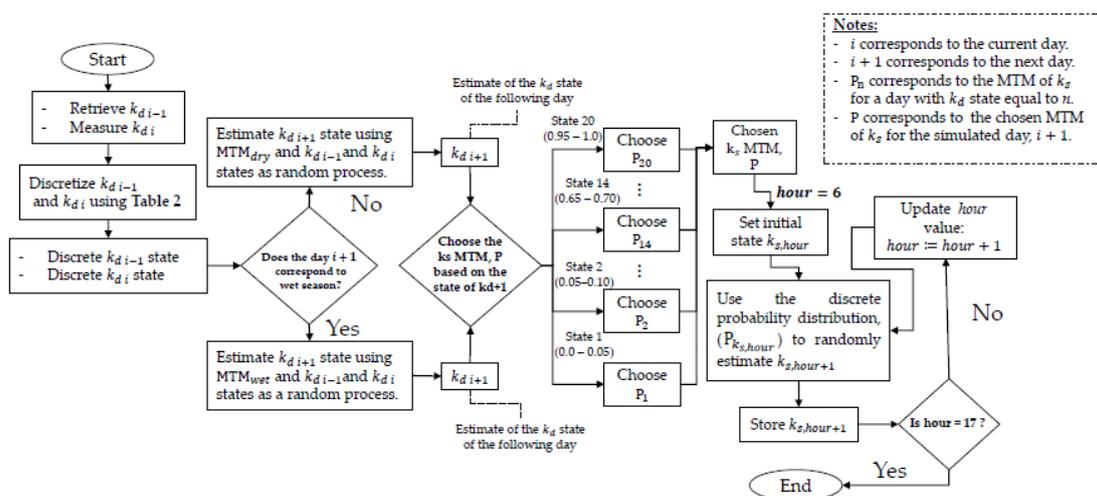
$$P_{(ij)k} = \frac{f_{(ij)k}}{T_{(ij)}} \tag{8}$$

where  $(i, j)$  values correspond to the current and previous adjacent state of the variable respectively, and  $k$  corresponds to its future state. In this case, each row in the second-degree MTM corresponds to a  $(i, j)$  pair and so, a second-degree MTM has dimensions  $(m^2 \times m)$ .

In addition, atmospheric variability can be different for wet and dry seasons in the region of study. Therefore, two second-degree MTMs for  $k_d$ , one for wet season months ( $MTM_{wet}$ ) and a second for dry season months ( $MTM_{dry}$ ), are obtained and used to generate the estimates of the daily clearness coefficient state. Consequently, we used a two-part model based on Markov chains to estimate the hourly series of GHI in the site of study.

#### 2.5. Simulations of Global Horizontal Irradiance (GHI) Using a Markov Chains Model

As previously mentioned, the statistical behavior of the  $k_s$  of a particular day depends on the state of the  $k_d$  of that day. This means that for each state of the  $k_d$  illustrated in Table 1, there is a corresponding MTM for  $k_s$ , as it was also explained in the previous section. Therefore, the state of the  $k_d$  should be forecasted before the diurnal series of  $k_s$  can be estimated. Consequently, the proposed Markov model has two parts: the first part estimates the state of the daily clearness coefficient ( $k_d$ ) using a second-degree Markov chain and the second part uses this estimate as a decision-maker variable to choose the appropriate first-degree MTM for estimating  $k_s$ . This procedure results in an estimated hourly series of  $k_s$  for one day-ahead. This process is illustrated in Figure 2.



**Figure 2.** Simulation process of the normalized clearness coefficient,  $k_s$  using the two-part Markov model proposed in this work.

The procedure depicted in Figure 2 results in a series of hourly  $k_s$  states, one for each hour of the following day like  $\{k_{s,6}, k_{s,7}, \dots, k_{s,18}\}$ , which vary from 1 to 20 (see Table 1). These values are then used to calculate the corresponding hourly GHI values. This is done as follows:

- (1) Transform the estimates  $\{k_{s,6}, k_{s,7}, \dots, k_{s,18}\}$  from states (1 to 20) to extinction percentages (0.0–1.0). This is done by first assuming that for a state  $i$  of  $k_s$ , the continuous values inside the corresponding interval (see Table 1), follow a uniform distribution such that  $k_s \sim \mathcal{U}((i-1) \times 0.05, i \times 0.05)$ . Then, using a pseudo-random process a continuous value of  $k_s$  is picked from the given distribution.
- (2) Once a continuous value of  $k_s$  is obtained and knowing the hour and the day for which it was estimated, the total hourly clearness coefficient,  $k_t$  must be calculated again in order to include the deterministic effects. This is done by using Equation (4) as:

$$k_{t,est} = k_{s,est} \times 1.031 \exp\left(\frac{-1.4}{0.9 + \frac{9.4}{A.M}}\right) + 0.1$$

- (3) Now, given that the extraterrestrial radiation can be modeled as a deterministic variable, the Equation for the total clearness coefficient,  $k_t$ , can be used now to estimate the corresponding GHI value as:

$$GHI_{est} = k_t \times I_{ext}$$

This procedure is performed for each of the  $k_s$  hourly estimates to obtain the corresponding series of hourly GHI estimates.

The period March 2016 to February 2017 is used to calculate the MTMs for the model while the period May 2017 to May 2018 is used to evaluate the model performance. Since the stochastic model is computationally inexpensive, each day of the period between May 2017 and May 2018 was estimated 1000 times. With this information, it is possible to obtain the error distributions of the different realizations produced by the model, which allows a more general assessment of the skill of the model at estimating the GHI in Medellín to be performed.

## 2.6. The Weather Research and Forecasting (WRF) Model

We used the WRF model, a flexible, mesoscale model designed for atmospheric modeling, weather forecasting and climate simulations [10], as a benchmarking model for the proposed Markov model. WRF is a state-of-the-art, community-supported model that is subject of constant development and support by the National Center for Atmospheric Research (NCAR), as well as multiple contributions from the users' community. Thus, this model is under constant scrutiny and improvement [46]. Additionally, the WRF model has been used by meteorological and environmental agencies in Colombia, such as SIATA [47] and the Colombian Institute of Hydrology, Meteorology and Environmental Studies (IDEAM), as their choice for operational forecasting exercises for different regions of Colombia.

WRF has been used to estimate the incoming solar radiation for different regions of the world [8, 48–51]. It has also been modified to provide better estimates of the incoming solar radiation, like the case of the WRF-solar project [52] which introduced new parametrizations and modifications to the conventional WRF. These features were first included in version 3.6 of the WRF model and have been an integral part of this model since [53–56].

In order to compare the Markov model to the results obtained from WRF, we analyzed the simulations of 5 particular days corresponding to different cloud cover conditions in the site of study, as shown in Table 2. It must be noted that 5 simulated days do not represent a comprehensive validation period for comparing the WRF model against the proposed Markov model (May 2017 to May 2018). We limited the present analysis to a few days given the computational costs of the WRF simulations. These simulations are run at a relatively high resolution (including convective-permitting domains) and include different configurations for each case. In our case, convective-permitting simulations are important not only to better simulate the cloud field at smaller scales, but also to

account for the orographic effects of the Tropical Andes, where Medellín is located. On the other hand, the use of different configurations of the model was needed for this study because, to the best of our knowledge, there is no peer-reviewed report of the skill of different microphysics schemes on the simulation of GHI for our region of interest. Thus, despite the limited number of simulated days with WRF, the comparison between our WRF simulations and the Markov model can still be useful for benchmarking the performance of the Markov model, and for gaining insights on the aspects of the model that should be improved.

Each 24-h WRF simulation started at 00:00:00 UTC (19:00:00 Local Time (LT)) of the day before and ended at 00:00:00 UTC (19:00:00 LT) of the day of interest. The first 6 h were considered as a spin-up period for each simulation. The integration time-step was set to 45 s and the shortwave radiation scheme was called every 15 min. The WRF outputs were also saved each 15 min. The resulting solar radiation series were averaged around each hour of the diurnal cycle; thus, an hourly temporal resolution was achieved.

**Table 2.** Particular days considered to analyze the performance of the clearness coefficient simulations obtained with the Markov chains-based model and the weather research and forecasting (WRF) model.

| Date             | Daily Clearness Coefficient ( $k_d$ ) | Category                          |
|------------------|---------------------------------------|-----------------------------------|
| 1 September 2017 | 0.72                                  | Clear sky                         |
| 23 December 2017 | 0.54                                  | Broken clouds                     |
| 5 June 2017      | 0.48                                  | Cloudy                            |
| 24 November 2017 | 0.39                                  | Very cloudy (local conditions)    |
| 19 August 2017   | 0.18                                  | Very cloudy (Synoptic conditions) |

WRF was run using two nested domains. The outer domain includes Colombia, part of the Pacific Ocean and the Caribbean Sea and has a spatial resolution of 12 km (Figure 1a). The inner domain includes the region containing the city of Medellín and the municipalities of the Aburrá Valley and has a spatial resolution of 4 km. Because WRF numerically solves the Equations of the atmosphere, it requires initial and boundary conditions. These conditions were obtained from the Global Forecast System (GFS) final analysis (FNL), which provides information about the state of the atmosphere every 6 h.

We used five WRF configurations for each simulated day, each with different combinations of microphysics and cumulus schemes. The remaining parameterizations are common to all simulations. For the planetary boundary layer, we used the Mellor–Yamada–Janjic [57] whereas for shortwave and longwave radiation we used the rapid radiative transfer model for general circulation models (RRTMG). We selected RRTMG as the radiation scheme based on its reported performance for estimating GHI, as well as the fact that is currently the only radiation scheme capable of coupling with the microphysics schemes of [58,59], which tackles the microphysics-radiative inconsistency present in WRF [37]. These configurations are variations of the WRF-solar configuration [52].

In this work, a set of six experiments using different options of the cumulus and microphysics schemes were undertaken. Table 3 shows the different configurations used in this work. The Kain-Fritsch/Thompson-Eidhammer (KF-TE) experiment was considered here as the control experiment, being a slight variation of the default WRF-solar configuration. In order to explore the role of the cumulus parameterization at the edge of what is commonly considered the convective-resolving resolution (4 km), we included an additional configuration (Experiment KF-TE-02), in which the cumulus scheme is active in the inner domain.

The Thompson & Eidhammer scheme was one of the selected microphysics schemes because: (i) it provides the radiative effective radii of the hydrometeors to be used in the RRTMG scheme; and (ii) it takes into account the effects of aerosols over clouds, which subsequently affects incoming radiation [52]. The Morrison microphysics scheme [60] calculates the size distribution for more hydrometeors than the Thompson and Eidhammer scheme and has had positive performance in the estimate of the GHI in other studies [55]. The Grell cumulus scheme can reflect the effect of

unresolved clouds over the incoming radiation. Schemes such as the Grell-3D scheme [61] allow the activation of the Deng mass flux scheme [62], which triggers the radiative feedback on both deep and shallow cumulus.

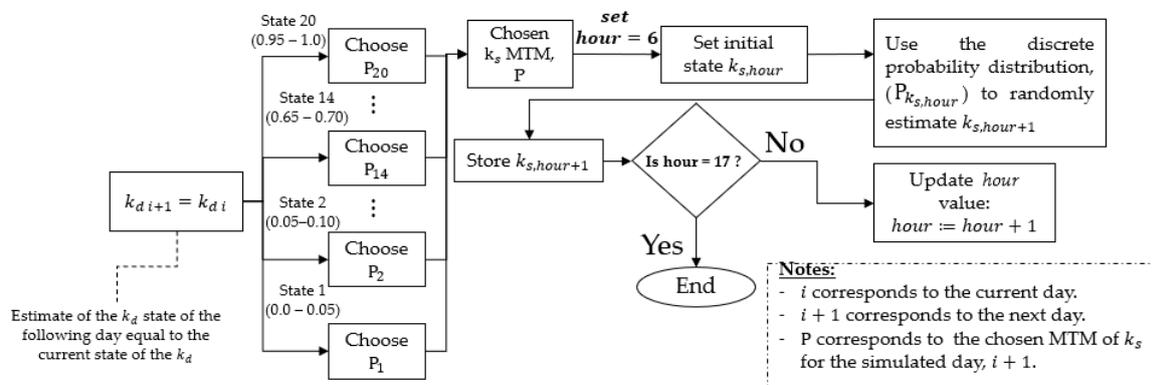
Considering that the effective resolution of a grid-point NWP model is larger than one grid cell, we averaged GHI values from WRF over a region around the location of the pyranometer corresponding to  $3 \times 3$  grid boxes of the larger domain, (i.e.,  $36 \text{ km} \times 36 \text{ km}$ ). Model output is available every 15 min. The resulting series were hourly averaged, so they could be compared to the in-situ measurements via different error metrics as discussed in next section.

**Table 3.** Experiments scheme configurations. The Deng scheme is also referred as Deng’s mass-flux-scheme.

| Experiments | Cumulus           |              | Shallow Convection |     | Microphysics           |                        |
|-------------|-------------------|--------------|--------------------|-----|------------------------|------------------------|
|             | d01               | d02          | d01                | d02 | d01                    | d02                    |
| KF-TE       | Kain-Fritsch [63] | Off          | Off                | Off | Thompson and Eidhammer | Thompson and Eidhammer |
| KF-TE-02    | Kain-Fritsch      | Kain-Fritsch | Off                | Off | Thompson and Eidhammer | Thompson and Eidhammer |
| KF-MO       | Kain-Fritsch      | Off          | Off                | Off | Morrison               | Morrison               |
| GR-TE       | Grell 3D          | Off          | Off                | Off | Thompson and Eidhammer | Thompson and Eidhammer |
| GR-TE-DE    | Grell 3D          | Off          | Deng               | Off | Thompson and Eidhammer | Thompson and Eidhammer |
| GR-MO-DE    | Grell 3D          | Off          | Deng               | Off | Morrison               | Morrison               |

### 2.7. Persistence-Markov Model

For further assessment of the proposed Markov model, a persistence-Markov model was also used in this work as a benchmarking model. Unlike the two-part Markov model previously described, this model assumes that the state of the  $k_d$  of the next day is the same as the  $k_d$  state of the current day. The process for obtaining the hourly series of GHI for the next day using the persistence-Markov model is the same as the one used for the proposed Markov model. This procedure is depicted in Figure 3.



**Figure 3.** Simulation process of the normalized clearness coefficient,  $k_s$  using the persistence-Markov model proposed in this work.

The persistence-Markov model was used to simulate the same validation period of the proposed Markov model (i.e., May 2017–May 2018).

### 2.8. Error Metrics

The metrics used for evaluating the hourly estimates of the stochastic model and the WRF model are the root mean square error (RMSE; in  $\text{W/m}^2$ ), the normalized root mean square error (nRMSE; in %), the mean bias error (MBE; in  $\text{W/m}^2$ ), and the normalized mean bias error (nMBE; in %), as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (GHI_{est,i} - GHI_{meas,i})^2} \quad nRMSE = \frac{RMSE}{\overline{GHI}_{meas}} \quad (9)$$

$$MBE = \frac{1}{N} \sum_{i=1}^n (GHI_{est,i} - GHI_{meas,i}) \quad nMBE = \frac{MBE}{\overline{GHI}_{meas}} \quad (10)$$

- $N$  corresponds to the number of hourly GHI values during the day.
- $GHI_{est,i}$  is the estimated average GHI value for the hour  $i$ .
- $GHI_{meas,i}$  is the measured average GHI value for the hour  $i$ .
- $\overline{GHI}_{meas}$  is the daily mean of the hourly GHI measured values.

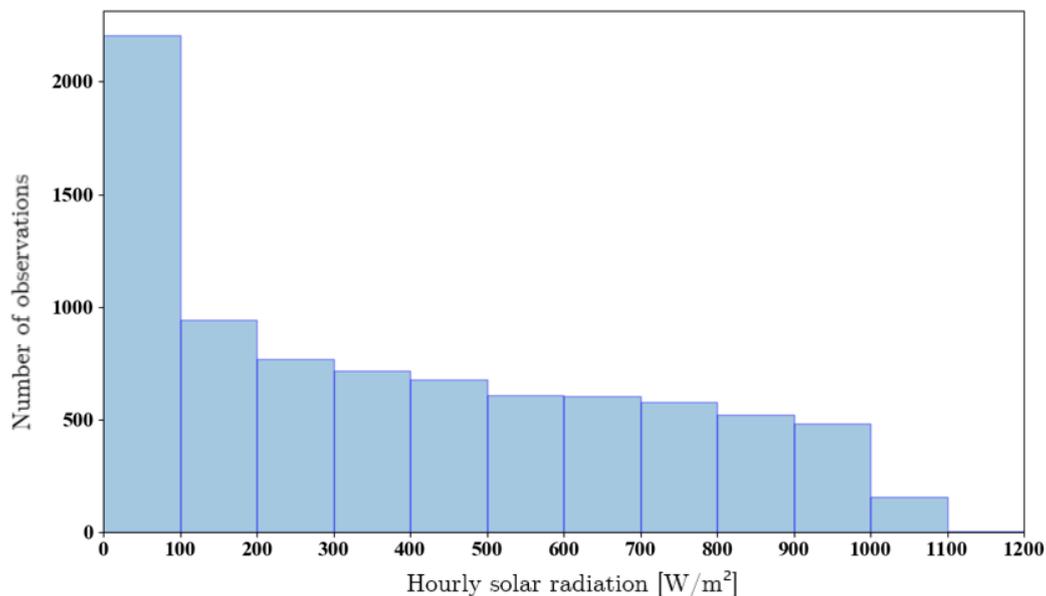
The estimation errors of the daily clearness coefficient were calculated as the difference between the estimated  $k_d$  and the measured  $k_d$ , as follows:

$$\Delta k_d = k_{d\text{estimated}} - k_{d\text{measured}} \quad (11)$$

### 3. Results

#### 3.1. Characterization of the GHI

The assessment of the solar resource for the period March 2016 to February 2017 is performed with only daytime GHI values, from 6 LT to 18 LT, since these are the limits of the interval in which finite values of GHI (i.e., non-zero and non-missing) are measured by the pyranometers. The distribution of hourly GHI data throughout the period considered here is shown in Figure 4. Table 4 presents the statistical summary of the GHI for the period of March 2016–February 2017. Aside from the main, median and standard deviation of the period considered, Table 4 also presents the interquartile range (IQR) of the data distribution which is defined as the difference  $Q3-Q1$ . From Table 4, 50% of the radiation values measured by the SIATA pyranometer range from  $90 \text{ W/m}^2$  to  $643 \text{ W/m}^2$ . The daytime values are skewed, with higher frequencies of the lowest radiation values, especially in the  $0-100 \text{ W/m}^2$  range.

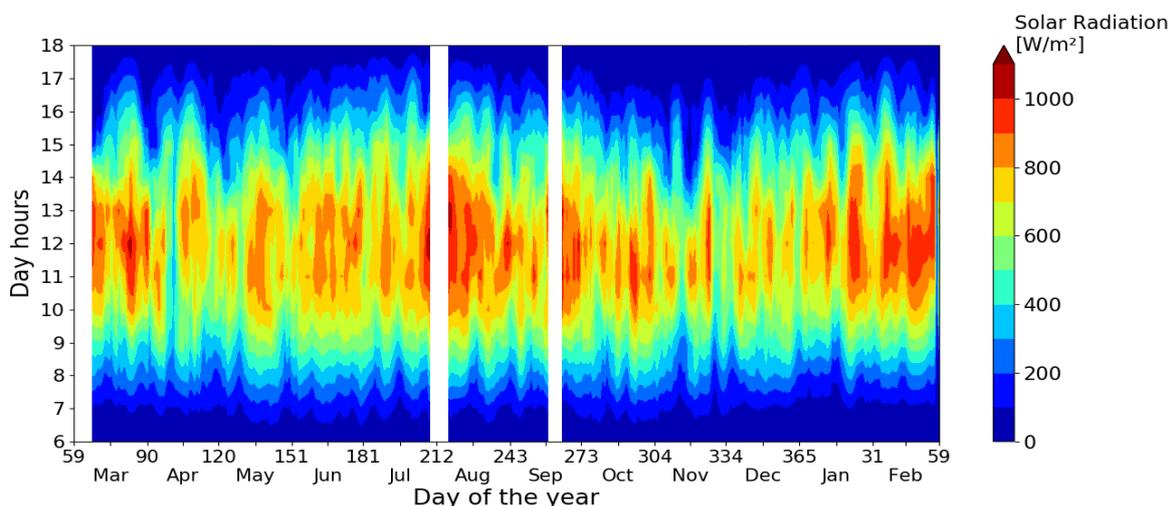


**Figure 4.** Histogram of hourly global horizontal irradiance (GHI) from the SIATA station recorded during the period 1 March 2016–28 February 2017.

**Table 4.** Statistical summary of the hourly GHI from the SIATA station recorded during the period 1 March 2016 to 28 February 2017.

| Total Size of the Sample with Hourly Resolution = 4658 |                            |                                        |                         | Percentiles (W/m <sup>2</sup> ) |     |     |     |      |
|--------------------------------------------------------|----------------------------|----------------------------------------|-------------------------|---------------------------------|-----|-----|-----|------|
| Mean (W/m <sup>2</sup> )                               | Median (W/m <sup>2</sup> ) | Standard Deviation (W/m <sup>2</sup> ) | IQR (W/m <sup>2</sup> ) | 0%                              | 25% | 50% | 75% | 100% |
| 385                                                    | 332                        | 315                                    | 553                     | 0                               | 90  | 332 | 643 | 1217 |

Figure 5 presents the GHI values corresponding to each hour of the day and each month during the period March 2016–February 2017. The highest values of GHI are observed from 11 to 13 h local time (LT). The months with the highest radiation values occurring during longer periods are March 2016, July 2016 and August 2016, and January 2017 and February 2017, reaching GHI values higher than 900 W/m<sup>2</sup>. April 2016, May 2016 and late November 2016 are the periods where more GHI variability is observed, with April 2016 exhibiting several days with solar radiation below 300 W/m<sup>2</sup> at any time of the day.



**Figure 5.** Hourly GHI values from SIATA station for all days of the period March 2016–February 2017. The color scale represents the intensity of solar radiation in W/m<sup>2</sup>. White gaps correspond to missing values.

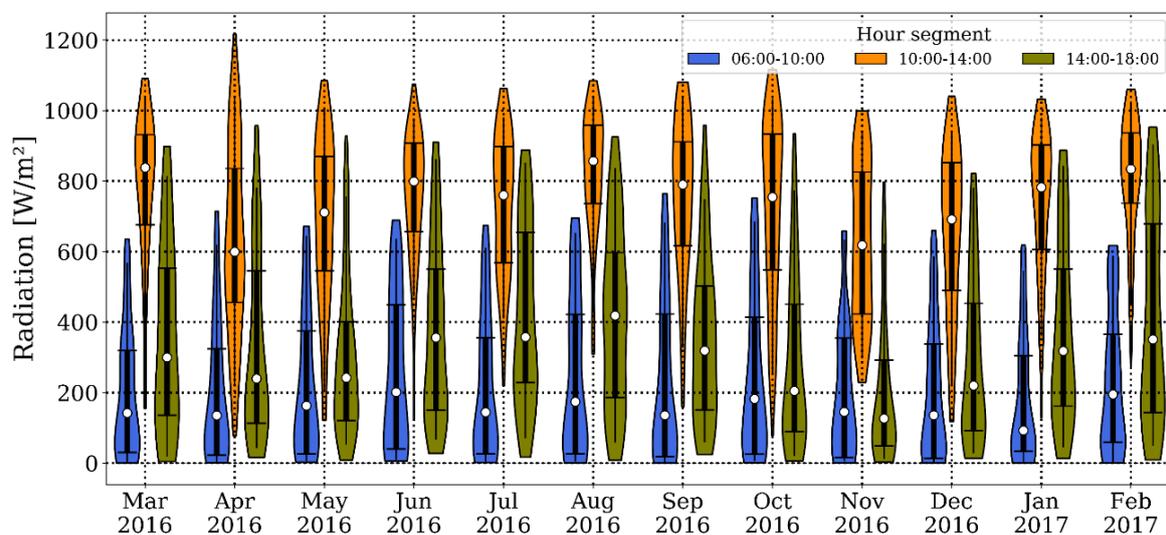
Reference [64] shows a distribution of the annual cycle of hourly radiation for Sevilla, Spain, similar to that shown in Figure 5, with an approximate mean of 383 W/m<sup>2</sup>, suggesting a similar annual behavior of the GHI. However, measurements for both sites (Sevilla and Medellín) exhibit seasonal changes with rather different amplitudes. The 2016–2017 annual GHI series obtained from the SIATA pyranometer (Figure 5) shows a rather uniform pattern when compared to the larger contrasts between summer and winter seasons at higher latitudes that would alter significantly the daylight hours and the intensity of incoming radiation (see [64]). In spite of lacking the strong seasonal effects of middle latitudes, the SIATA pyranometer data suggests two periods of reduced solar radiation: April to early May 2016, and late October to November 2016. During November 2016, the time interval when solar irradiance exceeds 200 W/m<sup>2</sup> is reduced from ~10 h to ~7.5 h.

Figure 6 shows the violin plots of monthly GHI for 3 segments of the day-time. These time windows correspond to “morning” (6–10 LT), “noon” (10–14 LT), and “afternoon” (14–18 LT). The violin plots show the empirical probability distributions of the hourly GHI for each segment considered, as colored regions, which are drawn mirrored around the middle black lines. These lines present the same information as a boxplot, with the white dots indicating the position of the median, the horizontal small lines indicating the quartile 1 (Q1) and quartile 3 (Q3). The thinner black lines that extend beyond these points reach the 5th percentile (p5) and the 95th percentile, respectively.

In the region of study, located in Central Colombia, precipitation shows two wet seasons, one during March–April–May and second one in September–October–November. By contrast, two dry seasons are observed in June–July–August and December–January–February [29,31,32,65]. For the period of study, the distributions in Figure 6 show that the GHI exhibited larger variability during the afternoon compared to the morning hours. February 2017 was the month with less variability at noon hours, most probably due to the persistence of clear skies during that time of the year (which is part of the dry season). The noon distributions for months like August 2016, January 2017, February 2017 presented skewed distributions towards high GHI values, while distributions at the same segment for months like April 2016 or November 2017 exhibited more uniform distributions of GHI, indicating a higher occurrence of low GHI values compared to the former mentioned months. This behavior is found to be characteristic for both clear sky and overcast conditions.

The potential occurrence of high radiation values at the surface is related to the magnitude of the extraterrestrial radiation, which in turn, for our region of interest ( $\sim 6.25^\circ$  N) exhibits some of its largest values between March and May (see e.g., Fig. 2.8 in [66]). However, more frequent cloud formation is observed over Medellín during this time of the year (first wet season), which corresponds to the early first rainy season for this region [29,31,32,65].

Figure 6 shows that the months within the dry season (e.g., August 2016 and February 2017) exhibit lower variability before noon, and higher variability in the afternoon hours, for the period considered. This behavior is due to the high number of clear sky days during these months. Months within the wet season (e.g., April 2016 and November 2016) exhibit larger variability throughout the day and GHI distributions more symmetrical around noon, unlike the dry season distributions of the period of study. Even though there is variability of hourly GHI during dry months, most of the hourly radiation records shown in Figure 6 are centered on higher values during the afternoon hours compared to other months of the year.



**Figure 6.** Daytime violin plots of GHI during the period March 2016–February 2017. Three hour segments are considered: 6–10 LT, 10–14 LT and 14–18 LT. The colored surfaces of each segment correspond to the empirical probability density function (PDF) of the GHI data. The black lines inside each distribution correspond to the boxplots of the GHI values and the white dots correspond to the median values of the distributions.

Table 5 shows the hourly mean coefficient of variation (CV) for each month of the period considered, along with the average values of daily solar energy the mean  $k_d$  values. The CV can be defined as:

$$CV = \frac{\mu}{\sigma} \quad (12)$$

Table 5 shows that the dry season months (i.e., June 2016, July 2016, August 2016 and February 2017) exhibit the lowest relative variability during the day (less than 34%). Table 5 also shows the monthly mean daily solar energy values in kWh/m<sup>2</sup>, commonly used in power systems and energy budget. August 2016 and February 2017 have the highest values of mean daily incoming energy and the lowest average of hourly CV. On the other hand, wet season months (i.e., April, May, October, November and December 2016) show the highest variability with CV values between 43% and 50% of the hourly GHI averages. April and November 2016 are the months with the highest hourly variability values (46% and 50%, respectively). Consistent with the mean daily values of solar energy, the mean  $k_d$  per month show that during the months of August 2016, January 2017 and February 2017, there was an overall low level of cloudiness, hence the high  $k_d$  values.

**Table 5.** Monthly mean hourly coefficient of variation (CV) and monthly mean daily radiation (in energy units of kilowatts-hour) from SIATA station during the period March 2016–February 2017.

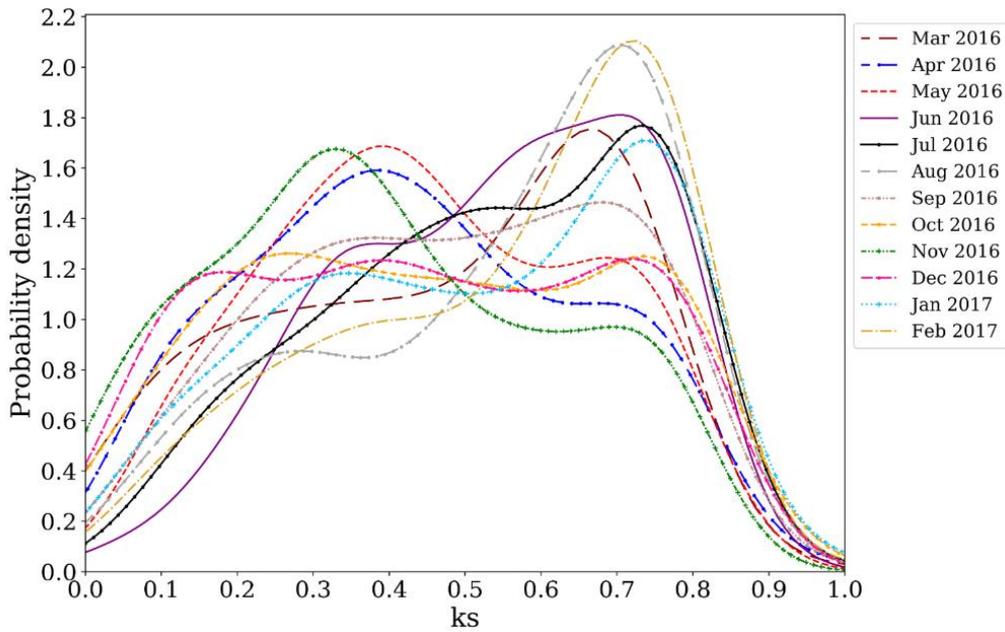
| Month          | Hourly Mean CV (%) | Month Mean Daily Solar Energy (kWh/m <sup>2</sup> ) | Mean $k_d$ |
|----------------|--------------------|-----------------------------------------------------|------------|
| March 2016     | 39.7               | 5.4                                                 | 0.52       |
| April 2016     | 46.2               | 4.7                                                 | 0.45       |
| May 2016       | 40.7               | 4.8                                                 | 0.48       |
| June 2016      | 29.9               | 5.6                                                 | 0.56       |
| July 2016      | 33.5               | 5.6                                                 | 0.55       |
| August 2016    | 31                 | 6                                                   | 0.58       |
| September 2016 | 37.2               | 5.4                                                 | 0.52       |
| October 2016   | 45.7               | 5.0                                                 | 0.50       |
| November 2016  | 49.7               | 4.0                                                 | 0.43       |
| December 2016  | 43                 | 4.6                                                 | 0.51       |
| January 2017   | 36.3               | 5.3                                                 | 0.57       |
| February 2017  | 32.9               | 5.9                                                 | 0.60       |

### 3.2. Clearness Coefficient

We calculated the normalized clearness coefficient ( $k_s$ ) for each hour throughout the period March 2016–February 2017, according to the SIATA records. Using these values, the empirical PDFs of  $k_s$  were calculated for each month during the period considered. We calculated the normalized clearness coefficient ( $k_s$ ) for each hour throughout the period March 2016–February 2017, according to the SIATA records. Using these values, the empirical probability density functions (PDFs) of  $k_s$  were calculated for each month during the period considered (Figure 7). Two main groups of PDFs are distinguished: one group including the months of June–July–August 2016 and January–February 2017, and a second group including the months of April, May, October, and November 2016. The first group corresponds to dry season months whereas the second group corresponds to wet season months [29,31,32,65].

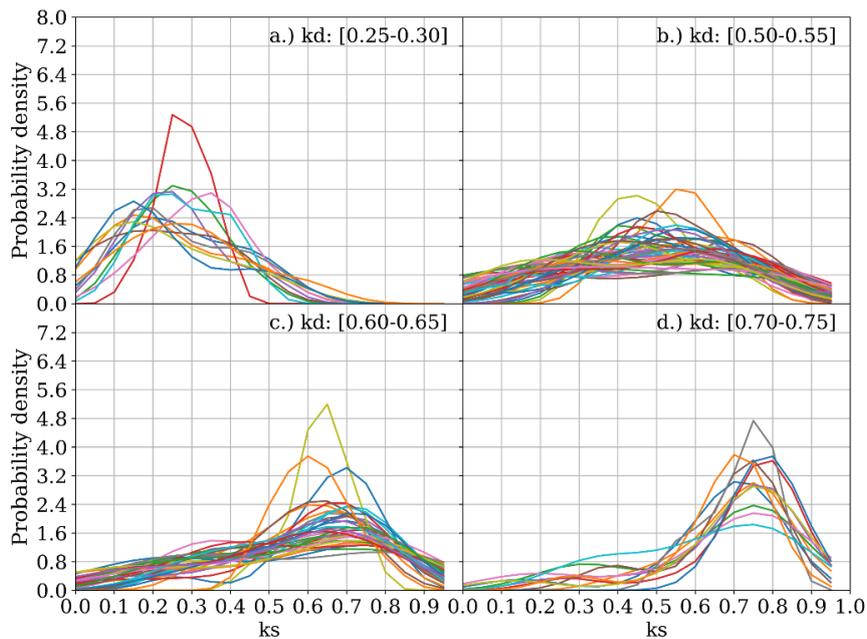
Figure 7 shows that the distributions of the wet season months have peaks around lower values of  $k_s$  (~0.4), whereas the distributions for the dry season months exhibit larger frequencies for larger values of  $k_s$  (~0.6–0.8). Transition months like September and December 2016 show more uniform distributions of  $k_s$ .

The daily clearness coefficient,  $k_d$ , was calculated as the ratio of the daily mean GHI to the daily mean of the corresponding extraterrestrial radiation. This coefficient reflects the sky condition of the whole day in terms of how much solar radiation reaches the surface as a percentage of how much would have reached the surface, given that there was no atmosphere. Previous studies have shown that there is a statistical relationship between the value of the daily clearness coefficient,  $k_d$ , and the daily statistical distribution of the normalized hourly clearness coefficient,  $k_s$ , [9,45,64].



**Figure 7.** Empirical PDFs of  $k_s$  for each month between March 2016 and February 2017, according to SIATA records. The y-axis values can be larger than 1.0 since this is a distribution plot, thus meaningful information can only be obtained from its integral over an interval rather than from a single point. This integral is equal or less to 1.

Figure 8 shows the empirical probability distributions of the normalized hourly clearness coefficient grouped according to their corresponding daily clearness coefficient value. Only four ranges are shown to illustrate this relationship. Each colored line in Figure 8 denotes the empirical PDF of the hourly  $k_s$  values of a day that has daily clearness coefficient contained in the interval shown in the corners of each panel, (e.g., every line in Figure 8a corresponds to the PDF of the  $k_s$  values of a day with a  $k_d$  between 0.25 and 0.30). In this way, what Figure 8 indicates is that the empirical PDFs of the  $k_s$  for days with the same state of  $k_d$  behave similarly.



**Figure 8.** Empirical PDFs for four types of days with  $k_d$  in different percentages of daily mean extraterrestrial radiation reaching the surface: (a) 25%–30%, (b) 50%–55%, (c) 60%–65%, and (d) 70%–75%.

For days with low  $k_d$ , the corresponding  $k_s$  values are centered around lower values (Figure 8a). The days with higher  $k_d$ , on the other hand, present  $k_s$  values that are centered around higher values (Figure 8b). Both cases exhibit less  $k_s$  variability than the cases of days with intermediate  $k_d$  values (Figure 8b,c). Due to these observed differences between the  $k_s$  distribution according to the  $k_d$ , we considered a two-part Markov model.

### 3.3. Clearness Coefficient and GHI Estimates

#### 3.3.1. Daily Clearness Coefficient Estimates

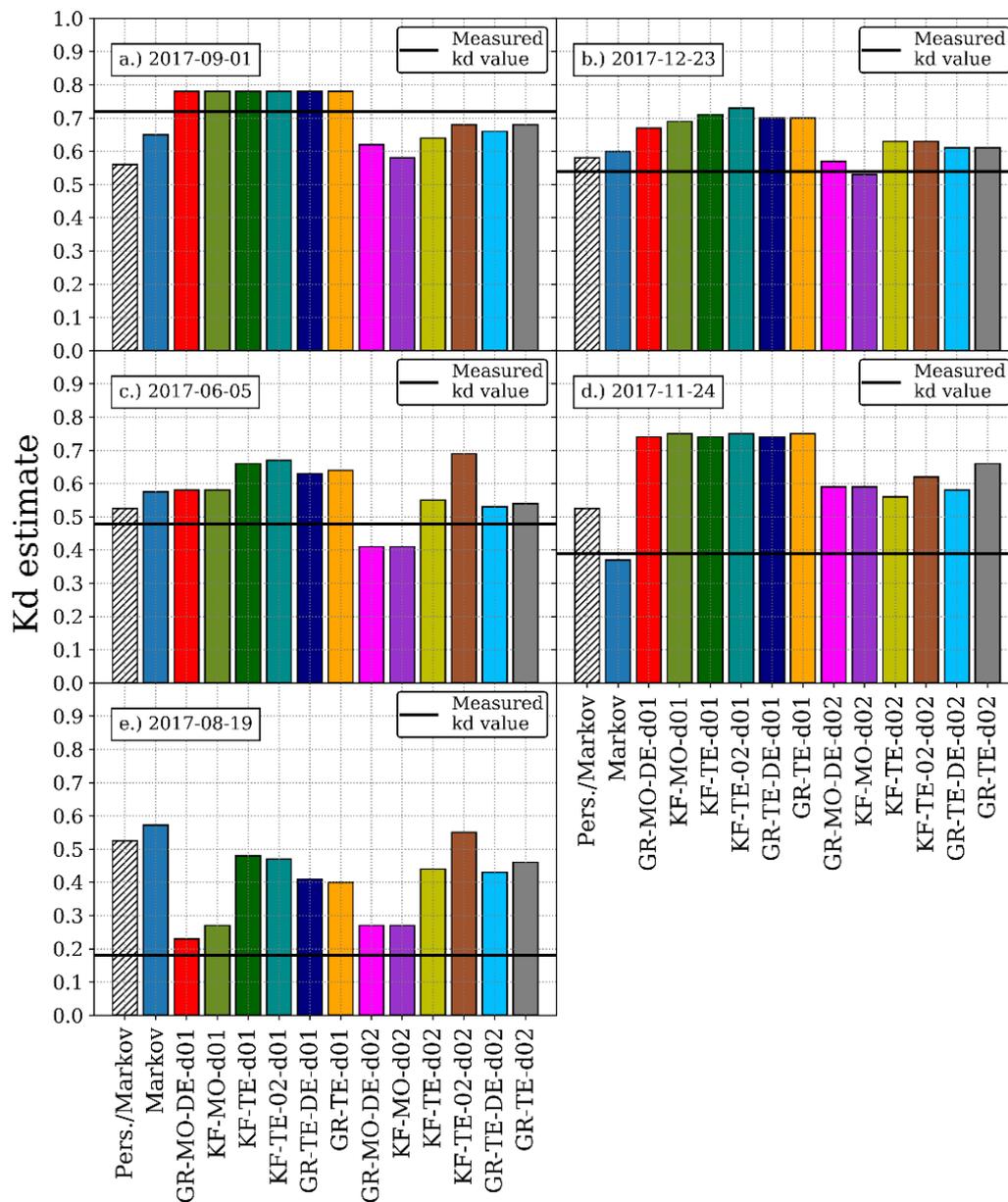
Figure 9 shows, for each of the considered days, the estimates of the daily clearness coefficient ( $k_d$ ) obtained with each of the considered models (i.e., the first part of the proposed Markov model, persistence model, and the WRF experiments). Figure 9 presents these estimates as bar plots, with each bar representing a different model and its height representing the corresponding estimated value of  $k_d$ . The black horizontal lines in each of these panels indicates the corresponding  $k_d$  measured values. The blue bars in Figure 9 correspond to the mean  $k_d$  values estimated by the proposed Markov model, for each day.

For 1 September 2017 (Figure 9a), all WRF experiments overestimated the daily clearness coefficient for the outer domain and underestimated it for the inner domain. The persistence model underestimated the  $k_d$  by  $\sim 0.2$  and the propose Markov model underestimated the  $k_d$  by  $\sim 0.05$ .

Figure 9b,c show the case of 23 December 2017 and 5 June 2017 respectively. For these cases, the Markov model and the Persistence model overestimated the measured  $k_d$  values, although the Persistence model had smaller overestimation than the proposed Markov model. For 23 December 2017 (Figure 9b), the KF-MO experiment for the inner domain estimated the  $k_d$  with an error of 0.01, being the closest estimate to the real  $k_d$ . In the case of 5th June 2017, only the KF-MO and GR-MO-DE experiments, both for the inner domain, underestimated the daily clearness coefficient,  $k_d$ . In the rest of the cases the  $k_d$  was overestimated.

Figure 9d corresponds to the case of 24th November 2017. This day presented a case where local-scale events seem to be responsible for the high extinction values over the incoming GHI. For this day, the Markov model shows a better performance than the Persistence model at estimating the  $k_d$ . In this case the  $\Delta k_d$  of the proposed Markov model is  $\Delta k_d = 0.01$ , while the persistence model exhibited a  $\Delta k_d \approx 0.52$ . For this day, the WRF experiments also overestimated the measured daily clearness coefficient,  $k_d$  consistently. Since the Markov model has a better performance at estimating transitions that are frequently found in the variable records used to train it, it is possible to assume that the events corresponding to the high levels of cloudiness for 24 November 2017, are events that frequently occurred during the March 2016–February 2017 period.

Figure 9e shows the simulations for 19 August 2017. This day, in contrast with 24 November 2017, was a case characterized by the presence of a synoptic scale event over the Caribbean Sea. Although all experiments reproduce high levels of GHI extinction at the location of the event, the GHI series reproduced over Medellín greatly vary from experiment to experiment. As can be seen from Figure 9e, the  $k_d$  estimates corresponding to GR-MO-DE-d01 and KF-MO-d01 are those closest to the measured value, with the GR-MO-DE-d01 estimate having the lowest error of 0.05. For this day, the Markov model reproduced an average  $k_d$  of 0.58 and the persistence model an average  $k_d$  of 0.53, which are values that lie far from the observed value of 0.18. This indicates that this type of events, which are not so commonly observed over the study region and are caused by events that occur beyond local scales, are hardly captured by the Markov model.



**Figure 9.** Bar plots of the  $k_d$  estimates Distributions of the daily clearness coefficient estimated by WRF, the Markov model and the persistence model. (a)  $k_d$  estimates distributions for 1 September 2017, measured  $k_d = 0.72$ . (b)  $k_d$  estimates distributions 23 December 2017. Measured  $k_d = 0.54$ . (c)  $k_d$  estimates distributions 6 June 2017. Measured  $k_d = 0.48$ . (d)  $k_d$  estimates distributions for 24 November 2017 measured  $k_d = 0.39$ . (e)  $k_d$  distributions for 19 August 2017, measured  $k_d = 0.18$ . The black lines correspond to the measured  $k_d$  values.

In general, the Markov model and the persistence model showed a similar performance at estimating the  $k_d$ , except for 1 September 2017 and 24 November 2017, where the proposed Markov model had a better performance. It is also noted that, in general, the WRF experiments for the inner domain tended to produce lower  $k_d$  estimates than their counterpart in the outer domain.

Additionally, although the proposed Markov  $k_d$  estimates presented in Figure 9 are averaged values and the distribution of the estimates is not shown, we must indicate that the Markov model is only capable of producing 2 to 3 different states for each day presented in Figure 9.

### 3.3.2. Hourly Estimates of GHI

We evaluated the skill of the WRF model and the proposed Markov model at simulating the GHI at Medellín, Colombia, for 5 non-consecutive days with different cloud cover conditions (Table 2). Although only 5 simulation days do not represent a comprehensive simulation period, the comparison between the proposed Markov model and the WRF model at estimating the GHI for these 5 days still represents a good indicator of the performance of the proposed Markov model. The chosen days were selected based on their daily clearness coefficient value. For each particular day, we ran six simulations of the WRF model with different combinations of microphysics and cumulus schemes (Table 3). These simulations were later compared against the simulations obtained with the Markov model for the same 5 days.

Since the Markov model produces estimates in a stochastic way, each simulation results in a different hourly series of GHI, meaning that several realizations of a single day could be obtained in order to observe the overall behavior of the proposed Markov model. For this reason, and noting that the model is computationally inexpensive, each day in Table 2 is simulated 1000 times using the proposed Markov model. The resulting series for each day are plotted as violin plots, which are discriminated by each hour of the day, and are presented along with the estimate GHI series obtained with the WRF experiments in Figure 10. Violin plots are useful in this case because they include the approximate probability distribution of the estimates and also have information about the statistical metrics of the distribution such as the interquartile range and the median of the series. As in Figure 6, the white dots in the violin plots indicate the median value of the distributions and the height of the boxes inside each violin plot indicates the position and magnitude of the IQR of each distribution calculated as the difference between the Q3 and Q1 values, respectively.

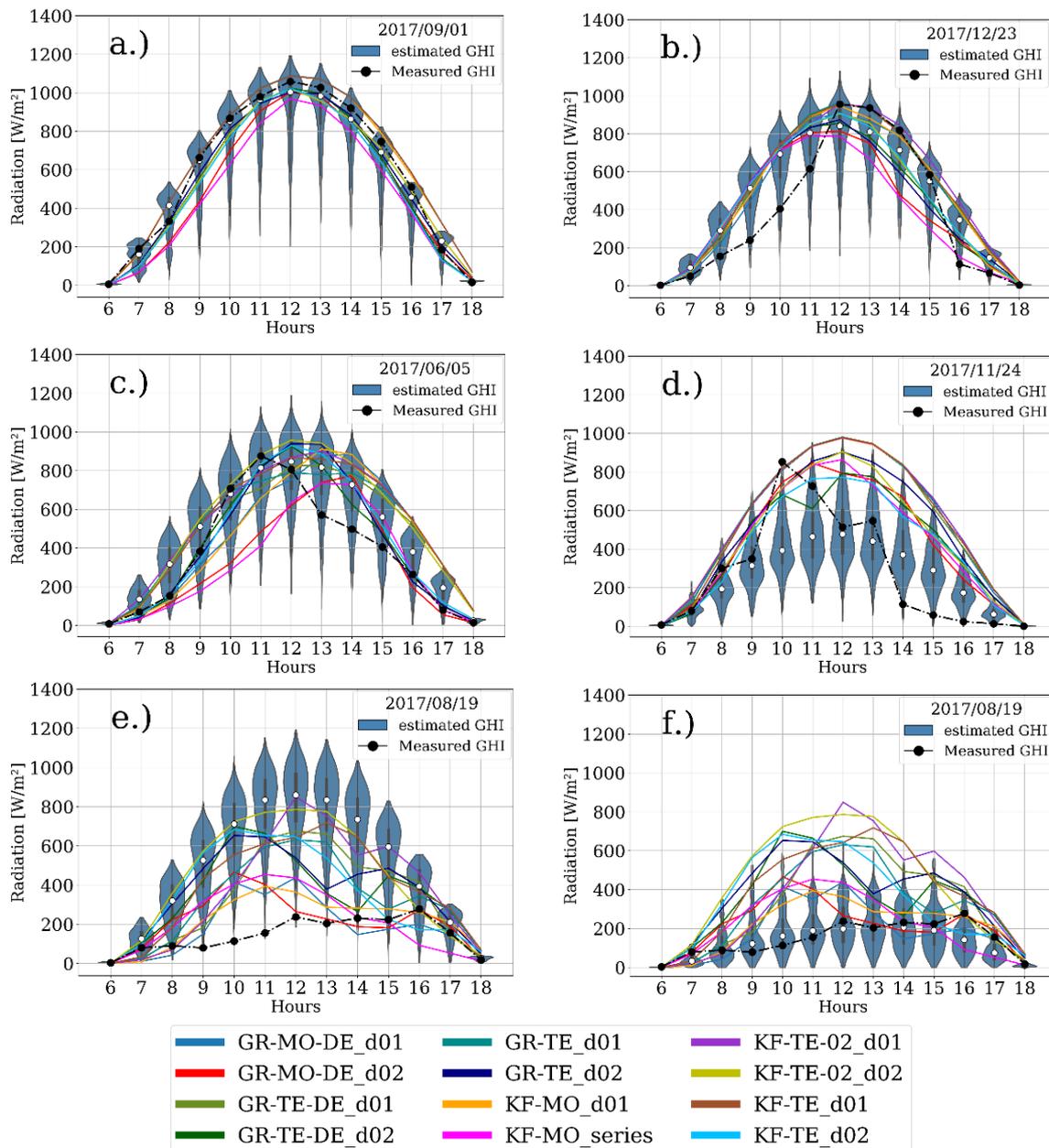
Figure 10 presents the GHI hourly simulations obtained with the Markov model and the WRF model for each of the 5 selected days. The blue distributions correspond to the violin plots of the Markov estimates for each hour of each day and the colored lines correspond to the WRF GHI series for each experiment and for each domain considered (i.e., outer domain, d01 and inner domain, d02). The dashed black lines correspond to the measured hourly series of GHI for each day.

Figure 10a presents the hourly distributions of GHI for a clear sky day (1 September 2017,  $k_d = 0.72$ ). For this day, the measured values of GHI are close to the IQR of the Markov estimate distributions. In the case of the WRF estimates, it is observed that series have a similar shape to the measured series, with the estimates of the KF-MO experiment presenting an RMSE increase of  $77 \text{ W/m}^2$  from the outer domain (d01) with respect to the inner domain (d02). In the case of the inner domain, d02, the WRF estimates consistently underestimated the measured GHI series.

Figure 10b presents the estimates for 23 December 2017. For this day, the measured series presented higher atmospheric extinction values during the morning hours than during the evening hours. For this day, the Markov model mostly overestimated the GHI during the morning hours, with only some extreme estimates falling near the measured values. This behavior is similar to that exhibited by most of the WRF experiments for this day. During the first evening hours (i.e., 12LT–15LT), the Markov model produced GHI values that were, in general, closer to the measured GHI values than the estimates produced by the WRF series for the inner domain (d02). In the case of the WRF estimates for the outer domain (d01), WRF produced GHI values that were very similar to the measured values and to the median values of the Markov distributions. During the last evening hours (i.e., 16LT–18LT), both models consistently overestimated the measured GHI values.

In the case of 5 June 2017 (Figure 10c), the Markov model had a satisfactory performance at estimating the GHI values between 9 LT–12 LT. However, it mostly overestimated the GHI during the afternoon hours as the measured values are near the lower tails of the estimate distributions. This matches the behavior of most of the WRF experiments, which also overestimated the GHI during the evening hours. However, the WRF experiments that used the Thompson and Eidhammer scheme, for the inner domain (i.e., KF-TE-d02, GR-TE-d02 and GR-TE-DE-d02), underestimated the clearness coefficient during the morning hours, unlike the rest of the experiments.

For the case of 24 November 2017 (Figure 10d), the Markov model consistently forecasted  $k_d$  states that were closer to the measured  $k_d$ , which caused the hourly GHI estimates of the Markov model to be closer to the hourly measured GHI values than the estimates obtained with the WRF experiments. This can be corroborated in Figure 9d. Although the Markov model had a better performance at estimating the hourly GHI than the WRF experiments for this day, both models consistently overestimated the measured GHI values for most of the evening hours (i.e., 14LT–18LT).



**Figure 10.** Hourly distributions of the GHI estimates produced by the Markov model and the WRF model for five particular days (Table 2). (a) Corresponds to 1 September 2017, with a  $k_d = 0.72$ . (b) Corresponds to 23 December 2017, with a measured  $k_d = 0.54$ . (c) Corresponds to 6 June 2017, with a measured  $k_d = 0.48$ . (d) Corresponds to 24th November 2017, with a measured  $k_d = 0.39$ . (e) Corresponds to 19 August 2017, with a measured  $k_d = 0.18$ . (f) Also corresponds to 19 August 2017, but in this case the estimated  $k_d$  was forced to be equal to the measured  $k_d = 0.18$ . The black dashed lines correspond to the measured hourly GHI values for each day.

The 19 August 2017 (Figure 10e) exhibits the lowest daily clearness coefficient value among the simulated days ( $k_d = 0.18$ ). As it was previously mentioned, the high levels of cloudiness observed on this day are mainly associated to a tropical depression crossing the Caribbean and that would later transform into Hurricane Harvey [67]. For this day, the Markov model consistently overestimated the  $k_d$  for each of 1000 simulations performed (Figure 9e). These overestimations lead to an incorrect selection of the hourly MTM, which in turn, causes the consistent overestimations of the hourly values of GHI. This case is an example of how the estimates of the first part of the model largely affect the performance of the second part of the model. For this day, some of the WRF experiments exhibited a better performance at estimating the GHI than the Markov model. This was the case of the experiments that considered Morrison as the microphysics scheme. It was also observed that the outer domain (d01) estimates showed a better agreement with the measured GHI values than the nested domain (d02) estimates. Given that for 19 August 2017, the first part of the Markov model is not able to correctly simulate the measured state of the  $k_d$  in any of the 1000 simulations, an extra set of 1000 simulations were performed to observe the behavior of the second part of the model alone (i.e., the simulation of the hourly GHI). In order to do this, for these extra set of simulations, the state of  $k_d$  used as input for the second part of the Markov model was forced to be equal to the measured value (i.e.,  $k_d = 0.18$ ). The simulations obtained in this way for the Markov model are presented in Figure 10f. Most of the measured values of the GHI now fall inside the IQR of the Markov estimate distributions. In this case, when the base state of the series, which can be represented by the  $k_d$ , is correctly estimated, the model frequently reproduces GHI values that are closer to the measurements than in the opposite case (Figure 10e).

In order to assess the general performance of the Markov model at estimating the hourly GHI for the 5 selected days, Table 6 presents the summary of the RMSE errors of the hourly estimates of GHI obtained with the Markov model for each of the simulated days. The last row of Table 6 corresponds to the case where the estimated  $k_d$  was set equal to the measured  $k_d$ . The RMSE values presented for the Markov model correspond to the median values of the RMSE distributions. Table 6 also shows the summary of the RMSE errors of the hourly estimates of GHI obtained with the GR-MO-DE experiment since it has one of the highest performances at estimating the hourly GHI. The RMSE values presented for this experiment correspond to the mean value between the RMSE error for the outer domain, d01 and the inner domain, d02. Additionally, the RMSE median value of the persistence-Markov model is also presented in order to provide further benchmarking for the proposed Markov model.

**Table 6.** Summary of hourly GHI estimates produced by the Markov model and by the WRF experiment, GR-MO-DE.

| Simulated day                                    | Markov                             | Persistence-Markov          | WRF: GR-MO-DE               |
|--------------------------------------------------|------------------------------------|-----------------------------|-----------------------------|
|                                                  | RMSE (Median)                      | RMSE (Median)               | RMSE                        |
| 01/09/2017<br>$k_d = 0.72$                       | 144 W/m <sup>2</sup> (26%)         | 190 W/m <sup>2</sup> (33%)  | 116 W/m <sup>2</sup> (21%)  |
| 23/12/2017<br>$k_d = 0.54$                       | 177 W/m <sup>2</sup> (47%)         | 171 W/m <sup>2</sup> (45%)  | 174 W/m <sup>2</sup> (46%)  |
| 05/06/2017<br>$k_d = 0.48$                       | 165.1 W/m <sup>2</sup> (44%)       | 158 W/m <sup>2</sup> (43%)  | 204 W/m <sup>2</sup> (55%)  |
| 24/11/2017<br>$k_d = 0.39$                       | 209 W/m <sup>2</sup> (76%)         | 233 W/m <sup>2</sup> (85%)  | 288 W/m <sup>2</sup> (104%) |
| 19/08/2017<br>$k_d = 0.18$                       | 415.2 W/m <sup>2</sup><br>(288.4%) | 372 W/m <sup>2</sup> (258%) | 133 W/m <sup>2</sup> (92%)  |
| 19/08/2017<br>$k_d = 0.18$<br>( $k_d$ corrected) | 97.5 W/m <sup>2</sup> (68%)        | 372 W/m <sup>2</sup> (258%) | 133 W/m <sup>2</sup> (92%)  |

According to Table 6, for 1st September 2017, the Markov model exhibits an RMSE error of  $144 \text{ W/m}^2$ , which is lower than the RMSE value for the Persistence-Markov model but higher than the mean RMSE produced by the WRF experiment. This indicates a lower performance at estimating the hourly GHI for this day compared to the WRF experiment but an improvement with respect to the persistence-based model. For 23 December 2017, the Markov model, the persistence-Markov model and the WRF experiments presented a similar performance at estimating the hourly GHI, all of them exhibiting nRMSE values of 47%, 45% and 46%, respectively. For the case of 6 June 2017, the performance of the Markov model increases with respect to the performance of the WRF experiments at estimating the GHI, however, the persistence-Markov model presents an improvement with respect to the proposed Markov model. For the case of 24 November 2017, it can be seen that the proposed Markov model has a better performance than the persistence-Markov model and the WRF model, presenting a lower RMSE than the other two models. This shows that for this particular day, neither the persistence-based model nor the WRF model were capable of simulating the correct state of the  $k_d$ , while the proposed Markov model did. Finally, for the case of 19 August 2017, the WRF presents a better performance at reproducing the effects of the larger scale event over the region of study, while the proposed Markov model and the persistence-Markov model fail to reproduce these effects over the GHI and thus, result in estimations with high RMSE values.

In general, the Markov model exhibits a lower RMSE values at estimating the hourly GHI in Medellín than the WRF experiment except for 1 September 2017 and 19 August 2017. For the latter simulation day, the Markov model is not able to reproduce the effects of the large-scale event over the region of study, while the WRF experiment GR-MO-DE is capable of producing large atmospheric extinction levels over the hourly GHI estimates.

As an additional assessment of the performance of these models for the 5 days selected, Table 7 shows the median of the RMSE values corresponding to each model. It can be observed that the proposed Markov model had a better performance at estimating the GHI than the persistence-Markov model. Also, it can be seen that the WRF model had a similar performance to the proposed Markov model with a small improvement of  $3 \text{ W/m}^2$ .

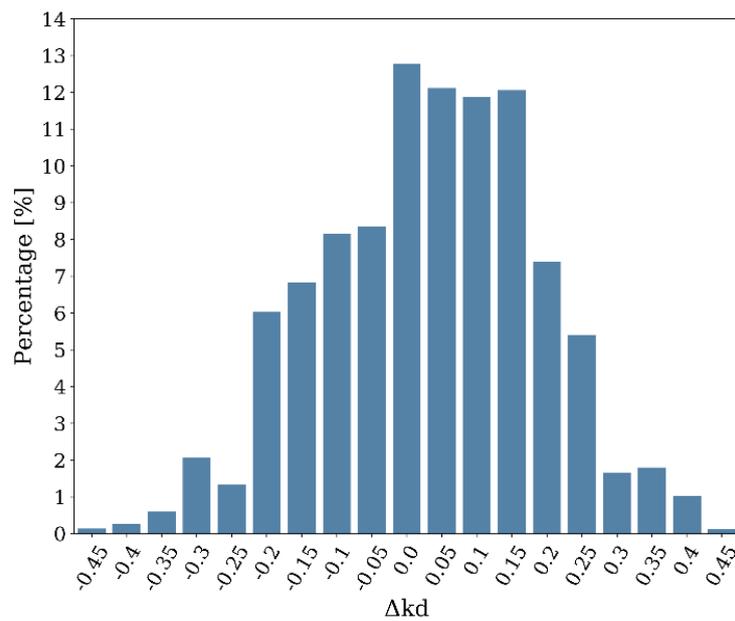
**Table 7.** Overall performance of the Markov model for the 5 chosen simulation days.

| Error Metric                   | Markov | Persistence-Markov | WRF:GR-MO-DE |
|--------------------------------|--------|--------------------|--------------|
| Median RMSE ( $\text{W/m}^2$ ) | 177    | 190                | 174          |

### 3.4. Daily Forecasts of GHI for the Validation Period of May 2017–May 2018

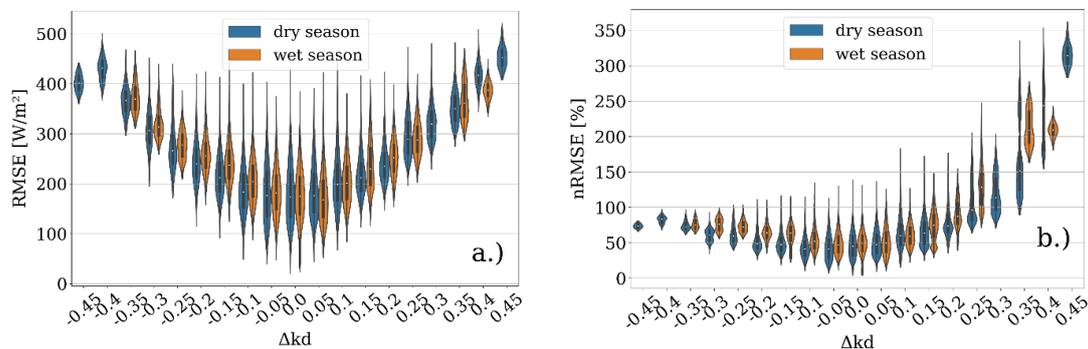
Figure 11 shows the histogram of frequencies of the  $\Delta k_d$  errors obtained for the period of May 2017–May 2018. Most of the  $\Delta k_d$  errors are between 0.0 and 0.15, being the values between 0.05 and 0.15 the most frequent ones. This shows that the first part of the model mostly overestimates the  $k_d$ , producing  $k_d$  values that usually are 1 state or 3 states above the measured values (see Table 1). On the other hand, the values corresponding to the correct estimations of the  $k_d$  ( $\Delta k_d = 0$ ) are also highly frequent.

Since the Markov-based model formulated here is computationally inexpensive, it was possible to perform 1000 simulations of each day during the period May 2017–May 2018. From this procedure it is possible to analyze the distribution of the errors of the different realizations produced by the model. Figure 12 show the distributions of the RMSE (Equation (9)) vs the estimation error  $\Delta k_d$  (Equation (11)), but this time presented as boxplots. In these plots, the lower whisker of the boxplots corresponds to  $Q1 - 1.5 \times \text{IQR}$ , while the upper whisker corresponds to  $Q3 + 1.5 \times \text{IQR}$ . The colored boxes enclose the IQR with the lines drawn within the boxes representing the median value. The colors of the boxes indicate the period in which the estimations were made.



**Figure 11.** Frequency histogram of the  $\Delta k_d$  of the estimates performed during the period May 2017–May 2018.

Figure 12 illustrates how the estimation error of  $k_d$  ( $\Delta k_d$ ) affects the RMSE distributions of the GHI estimates. Figure 12 also shows the RMSE distributions for wet and dry seasons.



**Figure 12.** (a) Distributions of the root mean square error (RMSE) vs  $\Delta k_d$  for the period May 2017–May 2018. (b) As in (a) but for nRMSE. Orange boxes correspond to the wet season estimates. Blue boxes correspond to the dry season estimates.

Figure 12a,b shows the impact that the estimation of the  $k_d$  (first part of the Markov model) has on the hourly estimates of GHI (second part of the Markov model). However, Figure 12a,b do not allow evaluating the hourly estimates without being affected by the errors of the  $k_d$ . From this analysis is neither possible to observe the performance of the second part of the model alone at estimating the hourly GHI during days with different sky conditions ( $k_d$ ). One way to do this is by simulating the period of May 2017 to May 2018 using the observed  $k_d$  values of each day as input for the second part of the Markov model. If this procedure is repeated a large number of times, it would be possible to obtain the error distributions of the GHI estimates for different sky conditions, which are given by the values of the observed  $k_d$ .

Finally, the proposed Markov model is compared against the Persistence-Markov model for the validation period of May 2017–May 2018. Table 8 shows that the proposed Markov model has a small improvement with respect the persistence-Markov model when evaluated during the period May 2017–May 2018. However, it is observed that during this period, the proposed Markov model exhibited

as larger MBE than the persistence-Markov model. This is consistent with the overestimations in the  $k_d$  produced by the first part of the proposed Markov model, as can be observe in Figure 11.

**Table 8.** Overall error for the validation period of May 2017–May 2018.

| Error Metric                    | Markov | Persistence-Markov |
|---------------------------------|--------|--------------------|
| Median RMSE (W/m <sup>2</sup> ) | 214    | 217                |
| Median MBE (W/m <sup>2</sup> )  | 33.9   | 7.21               |
| $\Delta k_d$                    | 0.076  | 0.003              |

The bias exhibited by the proposed Markov model at estimating the  $k_d$  reflects the fact that more than one year of data for training the model is necessary to train the model in order to produce a satisfactory estimator of the  $k_d$ . It is also worth noting that the persistence-Markov model displays a very low value of MBE, and not just when compared against the proposed Markov model.

## 4. Concluding Remarks

### 4.1. Solar Assessment

In this work, we analyzed global horizontal irradiance data from a pyranometer station located in Medellín, Colombia, with records during the period March 2016 to February 2017. Because of the low percentage of missing values, hourly and daily averages were computed without special interpolation algorithms. Although Medellín corresponds to a mountainous tropical city with months exhibiting a large number of cloudy days during the period considered in this work, most of the days during this period exhibited around 4 to 5 h with radiation values above 650 W/m<sup>2</sup>, typically surpassing the 800 W/m<sup>2</sup> between 11 and 13LT. Additionally, the measured data showed that the average daily solar energy that reached the surface in the region during March 2016 to February 2017 was of ~5 kWh/m<sup>2</sup>, which corresponds to 5 h of equivalent peak sun hours (PSH). The hourly measurements showed that this value would increase to an average of 5.5 PSH during the dry season months. This resource availability is important since it is comparable or even larger than what has been observed in regions that are referents regarding the use of solar radiation for energy production, like California, USA, with average values ranging from 3.8 to 6.3 PSH [68], or cities in Germany, with average PSH values that can range from 2.9 PSH to 3.6 PSH [69].

Additionally, during the dry season months, when the GHI values tend to be higher, solar variability decreases, which makes the use of the solar resource even more significant considering that during these months, especially when El Niño events occur, the main energy generation resource in Colombia (i.e., hydro energy) is heavily affected [30,70].

This is particularly important for Colombia since the country depends energetically on hydroelectric plants, which account for the 68.31% of the 17.35 GW of total installed capacity [71]. For this reason, it is essential to have complementary energy sources during dry periods that can reduce the risk of energy supply cuts.

### 4.2. Two-Part Markov Chain Model

In this work, a two-part model based on Markov chains was proposed. Additionally, a Numerical Weather Prediction model (WRF) was used as a benchmark for the estimates obtained with the Markov model. In general, the Markov model exhibited typical  $\Delta k_d$  ranging from 0.0 to 0.15. Also, regarding the  $k_d$  estimates, the estimation errors for all the WRF experiments ranged from −0.03 to 0.43. These results show that, in general, the first part of the Markov model (i.e., the part that estimated the  $k_d$ ), has a comparable performance to the WRF model at estimating the daily clearness coefficient. Both models showed a consistent positive bias at estimating the  $k_d$ , even the WRF experiment of GR-MO-DE, which presented an average  $\Delta k_d = 0.13$ , which is also comparable to the typical  $\Delta k_d$  values from the Markov model. On the other hand, the persistence-Markov model exhibited a lower  $\Delta k_d$  for the

validation period of May 2017–May 2018, but with a similar performance to the proposed Markov model in terms of the RMSE.

Although further inspection would be required to determine the origin of this positive bias in the Markov model it is clear that more than one year of data is necessary to produce more generalizable estimates of the  $k_d$ , and therefore, of the hourly GHI. This issue was also observed for the case of the 5 selected days (Figure 9), where the Markov model was only capable of estimating 2 to 3 different states for each day considered in this work.

In addition, the positive bias in our Markov model could be partially due to anomalous climate conditions during the training period.

During 2015/2016 a particularly strong *El Niño* event took place [72], which greatly affected the presence of clouds during the beginning of 2016, potentially producing larger values of  $k_d$  than in a normal year. Again, this shows how peculiarities of one year may cause biased estimates of the clearness coefficient, and therefore, longer periods for training the model are required.

The performance at estimating the hourly GHI of the Markov model and the WRF experiment, GR-MO-DE is presented in Table 7. For the five specific days considered in this study, the Markov model had a performance comparable to the performance of the GR-MO-DE WRF experiment, except for the case of the synoptic event, where the WRF experiment exceeded the skill of the Markov model at estimating the GHI. However, it is worth noting, as well, that for the overcast day associated to a more frequently measured type of event (i.e., 24 November 2017), the Markov model had a better performance at reproducing the hourly GHI than the WRF experiment.

When a more general assessment of the Markov is performed during the period May 2017–May 2018, RMSE errors inside the IQR of the hourly estimates produced by the Markov model for clear sky days ranged from  $\sim 57$  W/m<sup>2</sup> to  $\sim 171$  W/m<sup>2</sup> ( $\sim 16\%$ – $\sim 38\%$ ). For broken cloud sky conditions, the RMSE values ranged from  $\sim 149$  W/m<sup>2</sup> to  $\sim 250$  W/m<sup>2</sup> ( $\sim 32\%$ – $\sim 81\%$ ) while for overcast sky conditions, RMSE values ranged from  $\sim 80$  W/m<sup>2</sup> to  $\sim 170$  W/m<sup>2</sup> ( $\sim 50\%$ – $\sim 110\%$ ). In general, these results are in agreement with the errors found in similar studies regarding the hourly GHI forecasting for one day-ahead [56,73,74]. Although not frequently, the Markov model managed to produce GHI hourly estimates with very low RMSEs, even for highly overcast conditions. These values ranged from  $\sim 30$  W/m<sup>2</sup> to  $\sim 80$  W/m<sup>2</sup> ( $\sim 20\%$ – $\sim 50\%$ ). Even though these values are not typical among the estimates obtained in this work, they indicate that the model has the potential to produce series that are closer to the measured series for overcast condition than what is typically found in other works [8,50,51,73,75–77]. This improvement can be achieved by a further characterization of the hourly GHI series that could return typical intervals in which the hourly GHI usually lies for each hour of the day. Based on these intervals, the simulated series can be evaluated after they are produced and discarded if necessary. Additionally, because of how the Markov model was trained in this study, the hourly simulations of GHI did not take into account the dependency on the time of the day and the atmospheric mechanisms that could affect cloudiness. As a way of correcting this, the transition probabilities stored inside the Markov transition matrices could be obtained considering the time of the day.

We found that the estimates of the daily clearness coefficient,  $k_d$ , are fundamental for achieving the lowest RMSE values of the hourly series of GHI using the Markov model. Overestimations and underestimations of the  $k_d$  lead to error distributions that increase in magnitude with the bias of the  $k_d$  estimates. We also found that combining a persistence model for estimating the  $k_d$  with the second part of the proposed Markov model (i.e., the part that produces the hourly GHI estimates), results in a low bias model that could serve as an initial alternative for modeling the hourly GHI in the site of Medellín, Colombia.

In general, we believe further studies could start by improving the  $k_d$  prediction, which depends on meteorological process at synoptic scale, and so models like the weather prediction models could be a plausible alternative to simulate the  $k_d$  of the next day and which can later be used as an input for the second part of the Markov model proposed in this work. We are currently assessing the possibility of coupling the WRF model with the proposed Markov model.

The present work, despite the short time period of study, represents a novel contribution in terms of a detailed diagnostic of the solar resource and the performance of a variety of modeling tools for its day-ahead estimation for our region of interest. As such, this study is a first needed step for the assessment and modeling of solar energy for the city of Medellín, located within the tropical Andes.

**Author Contributions:** Conceptualization, J.U.-O., J.A.M. and P.A.A.; methodology, J.U.-O., J.A.M. and P.A.A.; software, J.U.-O.; validation, J.U.O.; formal analysis J.U.-O.; investigation, J.U.-O., J.A.M. and P.A.A.; resources, J.U.-O.; data curation, J.U.-O.; writing—original draft preparation, J.U.-O.; writing—review and editing, J.U.-O., J.A.M., P.A.A. and Á.J.-D. visualization, J.U.-O.; supervision, J.A.M. and P.A.A.; project administration, J.A.M. and P.A.A.; funding acquisition, J.U.-O., J.A.M. and P.A.A.

**Funding:** The authors gratefully acknowledge the financial support provided by the Program “Colombia Científica” within the framework of the called *Ecosistema Científico* (Contract No. FP44842-218-2018). We also acknowledge Sistema de Alertas Tempranas de Medellín for gently providing the GHI data used in this study. Julián Urrego-Ortiz was supported by the scholarship “Estudiante Instructor” funded by Universidad de Antioquia–Colombia.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Krakauer, N.; Cohan, D. Interannual variability and seasonal predictability of wind and solar resources. *Resources* **2017**, *6*, 29. [CrossRef]
2. Salgueiro, V.; Costa, M.J.; Silva, A.M.; Bortoli, D. Effects of clouds on the surface shortwave radiation at a rural inland mid-latitude site. *Atmos. Res.* **2016**, *178*, 95–101. [CrossRef]
3. Vignola, F.; Grover, C.; Lemon, N.; McMahan, A. Building a bankable solar radiation dataset. *Sol. Energy* **2012**, *86*, 2218–2229. [CrossRef]
4. CREG. *Resolución No. 30 de febrero de 2018*; CREG: Bogotá, Colombia, 2018; p. 27.
5. Energías Renovables. Available online: <http://www1.upme.gov.co/Paginas/Energias-renovables.aspx> (accessed on 22 July 2019).
6. Hidroituango, el Megaproyecto de Ingeniería en Colombia, Tiene Otro Problema: Un Socavón a 40 Metros CNN. Available online: <https://cnnespanol.cnn.com/2019/01/14/hidroituango-el-megaproyecto-de-ingenieria-en-colombia-tiene-otro-problema-un-socavon-a-40-metros/#0> (accessed on 22 July 2019).
7. Al-Kayiem, H.; Mohammad, S. Potential of renewable energy resources with an emphasis on solar power in Iraq: An outlook. *Resources* **2019**, *8*, 42. [CrossRef]
8. Lauret, P.; Diagne, M.; David, M. A neural network post-processing approach to improving NWP solar radiation forecasts. *Energy Procedia* **2014**, *57*, 1044–1052. [CrossRef]
9. Ngoko, B.O.; Sugihara, H.; Funaki, T. Synthetic generation of high temporal resolution solar radiation data using Markov models. *Sol. Energy* **2014**, *103*, 160–170. [CrossRef]
10. Skamarock, W.C.; Klemp, J.B.; Dudhi, J.; Gill, D.O.; Barker, D.M.; Duda, M.G.; Huang, X.-Y.; Wang, W.; Powers, J.G. *A Description of the Advanced Research WRF Version 3*; NCAR: Boulder, Colorado, USA, 2008.
11. Perez, R.; Lorenz, E.; Pelland, S.; Beauharnois, M.; Van Knowe, G.; Hemker, K.; Heinemann, D.; Remund, J.; Müller, S.C.; Traunmüller, W.; et al. Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. *Sol. Energy* **2013**, *94*, 305–326. [CrossRef]
12. Perez, R.; Kivalov, S.; Schlemmer, J.; Hemker, K.; Renné, D.; Hoff, T.E. Validation of short and medium term operational solar radiation forecasts in the US. *Sol. Energy* **2010**, *84*, 2161–2172. [CrossRef]
13. Doorga, J.R.S.; Rughooputh, S.D.D.V.; Boojhawon, R. Modelling the global solar radiation climate of Mauritius using regression techniques. *Renew. Energy* **2019**, *131*, 861–878. [CrossRef]
14. Ibrahim, S.; Daut, I.; Irwan, Y.M.; Irwanto, M.; Gomes, N.; Farhana, Z. Linear regression model in estimating solar radiation in perlis. *Energy Procedia* **2012**, *18*, 1402–1412. [CrossRef]
15. Gairaa, K.; Bakelli, Y. A Comparative study of some regression models to estimate the global solar radiation on a horizontal surface from sunshine duration and meteorological parameters for Ghardaïa Site, Algeria. *ISRN Renew. Energy* **2013**, *2013*, 1–11. [CrossRef]
16. Paoli, C.; Voyant, C.; Muselli, M.; Nivet, M.L. Forecasting of preprocessed daily solar radiation time series using neural networks. *Sol. Energy* **2010**, *84*, 2146–2160. [CrossRef]
17. Ji, W.; Chee, K.C. Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN. *Sol. Energy* **2011**, *85*, 808–817. [CrossRef]

18. David, M.; Ramahatana, F.; Trombe, P.J.; Lauret, P. Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. *Sol. Energy* **2016**, *133*, 55–72. [[CrossRef](#)]
19. Notton, G.; Voyant, C. Forecasting of intermittent solar energy resource. In *Advanced in Renewable Energies and Power Technologies*; Yahyaoui, I., Ed.; Elsevier Science: Amsterdam, The Netherlands, 2018; pp. 78–114. ISBN 9780128129593.
20. Aguiar, R.; Collares-Pereira, M. TAG: A time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation. *Sol. Energy* **1992**, *49*, 167–174. [[CrossRef](#)]
21. Bouabdallah, A.; Olivier, J.C.; Bourguet, S.; Machmoum, M.; Schaeffer, E. Safe sizing methodology applied to a standalone photovoltaic system. *Renew. Energy* **2015**, *80*, 266–274. [[CrossRef](#)]
22. Poggi, P. Stochastic study of hourly total solar radiation in. *Int. J. Climatol.* **2000**, *1860*, 1843–1860. [[CrossRef](#)]
23. Hocaoglu, F.O.; Serttas, F. A novel hybrid (Mycielski-Markov) model for hourly solar radiation forecasting. *Renew. Energy* **2017**, *108*, 635–643. [[CrossRef](#)]
24. Shakya, A.; Michael, S.; Saunders, C.; Armstrong, D.; Pandey, P.; Chalise, S.; Tonkoski, R. Solar irradiance forecasting in remote microgrids using markov switching model. *IEEE Trans. Sustain. Energy* **2017**, *8*, 895–905. [[CrossRef](#)]
25. Li, S.; Ma, H.; Li, W. Typical solar radiation year construction using *k*-means clustering and discrete-time Markov chain. *Appl. Energy* **2017**, *205*, 720–731. [[CrossRef](#)]
26. Munkhammar, J.; Widén, J. A spatiotemporal Markov-chain mixture distribution model of the clear-sky index. *Sol. Energy* **2019**, *179*, 398–409. [[CrossRef](#)]
27. Hocaoglu, F.O. Stochastic approach for daily solar radiation modeling. *Sol. Energy* **2011**, *85*, 278–287. [[CrossRef](#)]
28. Eugster, W. Mountain Meteorology: Fundamentals and Applications. *Mt. Res. Dev.* **2001**, *21*, 200–201. [[CrossRef](#)]
29. Bedoya-Soto, J.M.; Aristizábal, E.; Carmona, A.M.; Poveda, G. Seasonal shift of the diurnal cycle of rainfall over medellin's valley, central andes of Colombia (1998–2005). *Front. Earth Sci.* **2019**, *7*, 92. [[CrossRef](#)]
30. Poveda, G.; Mesa, O.J.; Waylen, P.R. *Nonlinear Forecasting of River Flows in Colombia Based Upon ENSO and Its Associated Economic Value for Hydropower Generation*; Springer: Dordrecht, The Netherlands, 2003; pp. 351–371.
31. Poveda, G.; Waylen, P.R.; Pulwarty, R.S. Annual and inter-annual variability of the present climate in northern South America and southern Mesoamerica. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **2006**, *234*, 3–27. [[CrossRef](#)]
32. Poveda, G. La hidroclimatología de Colombia: una síntesis desde la escala inter-decadal hasta la escala diurna por ciencias de la tierra. *Rev. Acad. Colomb. Cienc. Exactas, Físicas y Nat.* **2004**, *28*, 201–222.
33. Badescu, V. *Modeling Solar Radiation at the Earth's Surface*; Springer: Berlin Heidelberg, Germany, 2008; p. 53. ISBN 978-3-540-77454-9.
34. Kasten, F. A new table and approximate formula for relative optical air mass. *Archiv für Meteorologie, Geophysik und Bioklimatologie* **1966**, *14*, 206–223. [[CrossRef](#)]
35. Kasten, F.; Young, A.T. Revised optical air mass tables and approximation formula. *Appl. Opt.* **2000**, *28*, 4735–4738. [[CrossRef](#)]
36. Liu, B.Y.H.; Jordan, R.C. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Sol. Energy* **1960**, *4*, 1–19. [[CrossRef](#)]
37. Thompson, G.; Tewari, M.; Ikeda, K.; Tessendorf, S.; Weeks, C.; Otkin, J.; Kong, F. Explicitly-coupled cloud physics and radiation parameterizations and subsequent evaluation in WRF high-resolution convective forecasts. *Atmos. Res.* **2016**, *168*, 92–104. [[CrossRef](#)]
38. Iqbal, M. *An Introduction to Solar Radiation*; Elsevier Inc.: Vancouver, BC, Canada, 1983; ISBN 978-0-12-373750-2.
39. Perez, R.; Ineichen, P.; Seals, R.; Zelenka, A. Making full use of the clearness index for parameterizing hourly insolation conditions. *Sol. Energy* **1990**, *45*, 111–114. [[CrossRef](#)]
40. Palomo, E. Hourly solar radiation time series as first-order Markov chains. In Proceedings of the Actes du International Solar Energy Society Solar World Congress, Kobe, Japan, September 1989; pp. 2146–2150.
41. Munkhammar, J.; van der Meer, D.; Widén, J. Probabilistic forecasting of high-resolution clear-sky index time-series using a Markov-chain mixture distribution model. *Sol. Energy* **2019**, *184*, 688–695. [[CrossRef](#)]
42. Nguyen, B.T.; Pryor, T.L. A computer model to estimate solar radiation in Vietnam. *Renew. Energy* **1996**, *9*, 1274–1278. [[CrossRef](#)]
43. Aguiar, R.J.; Collares-Pereira, M.; Conde, J.P. Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices. *Sol. Energy* **1988**, *40*, 269–279. [[CrossRef](#)]

44. Graham, V.A.; Hollands, K.G.T. A method to generate synthetic hourly solar radiation globally. *Sol. Energy* **1990**, *44*, 333–341. [[CrossRef](#)]
45. Hollands, K.G.T.; Huget, R.G. A probability density function for the clearness index, with applications. *Sol. Energy* **1983**, *30*, 195–209. [[CrossRef](#)]
46. Diagne, M.; David, M.; Lauret, P.; Boland, J.; Schmutz, N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renew. Sustain. Energy Rev.* **2013**, *27*, 65–76. [[CrossRef](#)]
47. Hoyos, C.D.; Zapata, M.H. *Análisis del Impacto de la Interacción Suelo-Atmósfera en las Condiciones Meteorológicas del Valle de Aburrá Utilizando el Modelo WRF*; Universidad Nacional de Colombia: Bogotá, Colombia, 2015.
48. Diagne, M.; David, M.; Boland, J.; Schmutz, N.; Lauret, P. Post-processing of solar irradiance forecasts from WRF Model at Reunion Island. *Energy Procedia* **2014**, *57*, 1364–1373. [[CrossRef](#)]
49. Incecik, S.; Sakarya, S.; Tilev, S.; Kahraman, A.; Aksoy, B.; Caliskan, E.; Topcu, S.; Kahya, C.; Odman, M.T. Evaluation of WRF parameterizations for global horizontal irradiation forecasts: A study for Turkey. *Atmósfera* **2019**, *32*, 143–158. [[CrossRef](#)]
50. Zempila, M.M.; Giannaros, T.M.; Bais, A.; Melas, D.; Kazantzidis, A. Evaluation of WRF shortwave radiation parameterizations in predicting Global Horizontal Irradiance in Greece. *Renew. Energy* **2016**, *86*, 831–840. [[CrossRef](#)]
51. Lara-Fanego, V.; Ruiz-Arias, J.A.; Pozo-Vázquez, D.; Santos-Alamillos, F.J.; Tovar-Pescador, J. Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain). *Sol. Energy* **2012**, *86*, 2200–2217. [[CrossRef](#)]
52. Jimenez, P.A.; Hacker, J.P.; Dudhia, J.; Haupt, S.E.; Ruiz-Arias, J.A.; Gueymard, C.A.; Thompson, G.; Eidhammer, T.; Deng, A. WRF-SOLAR: Description and clear-sky assessment of an augmented NWP model for solar power prediction. *Bull. Am. Meteorol. Soc.* **2016**, *97*, 1249–1264. [[CrossRef](#)]
53. Arbizu-Barrena, C.; Ruiz-Arias, J.A.; Rodríguez-Benítez, F.J.; Pozo-Vázquez, D.; Tovar-Pescador, J. Short-term solar radiation forecasting by advecting and diffusing MSG cloud index. *Sol. Energy* **2017**, *155*, 1092–1103. [[CrossRef](#)]
54. Ruiz-Arias, J.A.; Arbizu-Barrena, C.; Santos-Alamillos, F.J.; Tovar-Pescador, J.; Pozo-Vázquez, D. Assessing the surface solar radiation budget in the WRF model: A spatiotemporal analysis of the bias and its causes. *Mon. Weather Rev.* **2016**, *144*, 703–711. [[CrossRef](#)]
55. De Meij, A.; Vinuesa, J.F.; Maupas, V. GHI calculation sensitivity on microphysics, land- and cumulus parameterization in WRF over the Reunion Island. *Atmos. Res.* **2018**, *204*, 12–20. [[CrossRef](#)]
56. Verbois, H.; Rusydi, A.; Thiery, A. Probabilistic forecasting of day-ahead solar irradiance using quantile gradient boosting. *Sol. Energy* **2018**, *173*, 313–327. [[CrossRef](#)]
57. Janjić, Z.I. The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Weather Rev.* **1994**, *122*, 927–945. [[CrossRef](#)]
58. Thompson, G.; Field, P.R.; Rasmussen, R.M.; Hall, W.D. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Weather Rev.* **2008**, *136*, 5095–5115. [[CrossRef](#)]
59. Thompson, G.; Eidhammer, T. A Study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci.* **2014**, *71*, 3636–3658. [[CrossRef](#)]
60. Morrison, H.; Thompson, G.; Tatarskii, V. Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Weather Rev.* **2009**, *137*, 991–1007. [[CrossRef](#)]
61. Grell, G.A.; Dévényi, D. A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.* **2002**, *29*, 38-1–38-4. [[CrossRef](#)]
62. Deng, A.; Gaudet, B.; Dudhia, J.; Alapaty, K. Implementation and evaluation of a new shallow convection scheme in WRF. In Proceedings of the 26th Conference on Weather Analysis and Forecasting/22nd Conference on Numerical Weather Prediction, Atlanta, GA, USA, 2–6 February 2014; pp. 2–6.
63. Kain, J.S. The kain–fritsch convective parameterization: An update. *J. Appl. Meteorol.* **2004**, *43*, 170–181. [[CrossRef](#)]
64. Moreno-Tejera, S.; Silva-Pérez, M.A.; Lillo-Bravo, I.; Ramírez-Santigosa, L. Solar resource assessment in Seville, Spain. Statistical characterisation of solar radiation at different time resolutions. *Sol. Energy* **2016**, *132*, 430–441. [[CrossRef](#)]

65. Bowman, K.P.; Fowler, M.D. The diurnal cycle of precipitation in tropical cyclones. *J. Clim.* **2015**, *28*, 5325–5334. [CrossRef]
66. Liou, K.-N. *An Introduction to Atmospheric Radiation*, 2nd ed.; Elsevier Science: Amsterdam, The Netherlands, 2002; ISBN 0124514510.
67. US Department of Commerce, NOAA, N.W.S. Hurricane Harvey Info. Available online: <https://www.weather.gov/hgx/hurricaneharvey> (accessed on 22 September 2019).
68. Roberts, B.J. Solar Maps | Geospatial Data Science | NREL. Available online: <https://www.nrel.gov/gis/solar.html> (accessed on 13 June 2019).
69. HotSpot Energy Solar Sun Hours | Average Daily Solar Insolation | Europe. Available online: <https://www.hotspotenergy.com/DC-air-conditioner/europe-solar-hours.php> (accessed on 13 June 2019).
70. World Bank. *Colombia Recent Economic Developments in Infrastructure II*; World Bank: Washington, DC, USA, 2004.
71. Capacidad Efectiva Por Tipo de Generación. Available online: <http://paratec.xm.com.co/paratec/SitePages/generacion.aspx?q=capacidad> (accessed on 22 July 2019).
72. Stockdale, T.; Balmaseda, M.; Ferranti, L. The 2015/2016 El Niño and Beyond. Available online: <https://www.ecmwf.int/en/newsletter/151/meteorology/2015-2016-el-nino-and-beyond> (accessed on 22 October 2019).
73. Aryaputera, A.W.; Yang, D.; Walsh, W.M. Day-ahead solar irradiance forecasting in a tropical environment. *J. Sol. Energy Eng.* **2015**, *137*, 051009. [CrossRef]
74. Verbois, H.; Huva, R.; Rusydi, A.; Walsh, W. Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning. *Sol. Energy* **2018**, *162*, 265–277. [CrossRef]
75. Husein, M.; Chung, I.Y. Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach. *Energies* **2019**, *12*, 1856. [CrossRef]
76. Lan, H.; Yin, H.; Hong, Y.Y.; Wen, S.; Yu, D.C.; Cheng, P. Day-ahead spatio-temporal forecasting of solar irradiation along a navigation route. *Appl. Energy* **2018**, *211*, 15–27. [CrossRef]
77. Lan, H.; Zhang, C.; Hong, Y.Y.; He, Y.; Wen, S. Day-ahead spatiotemporal solar irradiation forecasting using frequency-based hybrid principal component analysis and neural network. *Appl. Energy* **2019**, *247*, 389–402. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).