Predictive spatio-temporal model for spatially sparse global solar radiation data

Maïna André, Ted Soubdhan, Hanany Ould-Baba, Sophie Dabo-Niang

Predictive spatio-temporal model for spatially sparse global solar radiation data

Maïna André, Sophie Dabo-Niang, Ted Soubdhan, Hanany Ould-Baba

Predictive spatio-temporal model for spatially sparse global solar radiation data

Maïna André[a,*], Sophie Dabo-Niang[b], Ted Soubdhan[a], Hanany Ould-Baba[a]

[a] *University of French West Indies,Campus de Fouillole 971159,Pointe-à-Pitre, Guadeloupe (F.W.I)*
[b] *Laboratoire EQUIPPE, Université Charles De Gaulle, Lille 3, MODAL team, INRIA Lille-Nord de France*

Abstract

This paper introduces a new approach for the forecasting of solar radiation series at a located station for short time scale. We built a multivariate model in using few stations (3 stations) separated with irregular distances from 26 km to 56 km. The proposed model is a spatio temporal vector autoregressive VAR model specifically designed for the analysis of spatio-temporal data which are rich in time dimension and sparse in spatial dimension. This model differs from classic linear models in using spatial and temporal parameters where the available predictors are the lagged values at each station. A spatial structure of stations is defined by the sequential introduction of predictors in the model. Moreover, initialization parameters and an iterative strategy in the process of our model will select the necessary stations for each forecasting day, removing the uninteresting predictors and also selecting the daily optimal p-order. We studied the performance of this model for different daily class of global solar radiation. The statistical errors as the normalized mean absolute error (nMAE), the normalized mean biased error (nMBE) and the normalized root mean squared error (nRMSE) are presented. We compare the results of our model to those found in literature for different time step and to simple and well known persistence model.

Keywords:
stations' spatial order, intra-hour forecasting, spatio-temporal vector autoregressiv processs

1. Introduction

Solar energy is available in abundance in tropical zone but presents many short-term fluctuations introduced mainly by clouds. Due to variations of the sun's position each day and the apparent motion of the sun throughout the year, the total irradiation received at a particular site in both time and space can vary widely, as it has been shown in Gueymard et al [16] study based on high resolution radiation data sampled in the USA. The increasing of PV plants to meet demand will increase the variability and uncertainty that must be managed by system operators and planners of (photovoltaics) PV system (Mills et al [21]). One way to manage variability and uncertainty of monitoring networks is to predict values at locations using observed data at known locations.

In the literature, we can find correlation-based forecasting studies aim to derive quantities such as cloud speeds and directions in a deterministic way. For very short-term modeling and forecasting of solar radiation, sky cameras

*Corresponding author
Email addresses: mandre@univ-ag.fr (Maïna André), sophie.dabo@univ-lille3.fr (Sophie Dabo-Niang), tsoubdha@univ-ag.fr (Ted Soubdhan)

can be used. Chi Wai Wow et al [7] presented a method sub-kilometer cloud forecasting using a ground sky imager. They predicted solar irradiance at the University of California, San Diego from 30s to 5min ahead forecast. Chu et al [8], defined smart forecasting models which combine sky image processing with ANN (artificial Neural network) optimization schemes to predict 1min average direct normal irradiance for horizons 5 and 10 minutes and in Marquez et al [20] presented an image processing methodology using total sky imager to generate forecasting of 1 min average direct normal irradiance for horizons varying from 3 to 15 min. Other studies such as in Coimbra et al[9] paper predicted the local cloud movements thus deducing irradiation at the ground level by cloud motion analyses and artificial intelligence algorithms. Bosch et al [3]; Bosch et al [4], determined cloud directions and speeds in using correlation methods by sky cameras where ground network of irradiance sensors are with high spatial resolution .

An alternative method for intra-hour forecastings is statistical models as described in Gordon [26] where forecasting tests are run using regressions in logs, Autoregressive Integrated Moving Average(ARIMA), transfer functions, neural networks, and hybrid models. These models are evaluated for data sets, at resolutions of 5, 15, 30, and 60 min, using the global horizontal component.

Concerning the spatio-temporal processes, few researchers have used this methodology to model environmental processes. A variety of techniques integrating spatial and temporal parameters can be applied to generate forecastings, such as space-time kriging, spatio-temporal ARMA (STARMA) models and vector autoregressive (VAR) models. Epperson [12] has applied STARMA model in an ecological context and CA Glasbey [14] used the STARMA model to evaluate the variability of the solar energy potential . They investigated 10 minutes timescale forecasting at 10 sites on a sequence of 31 days to a grid of 10km $\times$ 10km. For the space-time kriging method, M.Cellura et al [6] developed geostatistical techniques in order to obtain the wind speed maps for the region at 10 and 50 meters above the ground level. The remaining de-trended linear means have been computed by using an universal kriging (UK) estimator. Heping Liu et al [19] has investigated a method based on TK (Taylor Kriging) model modified for the forecasting of wind speed time series. Dazhi Yang et al [29] used time-forward kriging to forecast the hourly spatio-temporal solar irradiance data from 10 Singapore weather stations and Inoue et al. [18], applied spatio-temporal Kriging to obtain distributions of the solar radiation with spatial resolution of 500m and resolution time of 5 min. The researchers de Luna and Genton ; Gneiting [15]; Porcu et al (2008) [25]have used spatio-temporal VAR method to model environmental processes as wind velocity fields and atmospheric concentration of carbon monoxide.

The goal of this paper is to build a multivariate forecasting with three stations. Despite of the scarcity of information we will define a spatio-temporal model and we will see how much forecasting skill we can attain. Kriging method is efficient for available rich data in the time and spatial dimension to perform a spatio-temporal prediction but not for a reduced number of measurement sites. The VAR method proposed by de Luna and Genton ([11])is the suitable model, indeed this model is fitted with rich experimental data in the time but sparse in the spatial dimension. The spatio-temporal VAR model proposed does not make spatial stationarity assumptions and consists of a vector autoregressive (VAR) specification (Huang and Hsu [17], de Luna and Genton [11]) that is to say VAR model treats each spatial location separately in the process. Thus, this paper investigates on a spatio-temporal short-term forecasting, based on spatio temporal VAR model methodology described in de Luna and Genton [11]. Our model uses spatial and temporal parameters where the predictors are the lagged values at each station. Moreover, the spatial structure of stations can be defined.

We performed our analyze in using one year (2012), 10 min averaged global solar radiation data collected at 3 meteorological stations (La Désirade, Petit-canal, Fouillole) across Guadeloupe island (French West Indies). In section 2, we present the experimental set-up of solar radiation measurements. In Section 3, we quantify the existing interactions of global solar radiation between stations. In section 4, the theoretical and statistical methodology is described. In section 5, we define our spatio temporal VAR model applied to global solar radiation. Moreover, we present tests allowing to choose an optimal VAR model. Section 6 highlights the performance of the VAR model in analyzing statistical errors of prediction for different classes of daily solar radiation. We also compare in this section the performance of our VAR model to the persistence model and to other methods of forecasting found in literature. This assessment is made for different time step: 5min, 10 min and 15 minutes. Finally we conclude in section 7. This forecasting model of the global solar radiation at short time scale for different locations using spatio-temporal parameters, will be useful to ensure the performance of electric power PV system what satisfies reliability standards in a least cost manner.

2. Solar global radiation measurements

2.1. Experimental set up of solar global radiation measurements

The global solar radiation is measured at 1 Hz on each site with a Kipp Zonen pyranometer (type SP Lite) whose response time is less than a second.We have used for this study an averaged database at 10 minutes of the solar flux measured for a period starting in January 2012 and ending in December 2012. The measurements are collected at three sites (Figure 1): along the cliffs of Petit-Canal Gros Cap (16°38N, 61°49W); Fouillole (16°26N, 61°24W), campus of the French West Indies University; on the East coast of La Désirade (16°31N, 61°55W). The distances between sites in meters are summarized in Table.1. Sensor accuracy given by the manufacturer is 3%. The data are measured and recorded by a Campbell Scientific CR1000 type stand-alone data. On the site of Petit-Canal, data are stored on a memory card compact flash (industrial type) with a capacity of 1 GB and a battery set powers the unit. This device has a check-on two months battery autonomy. In addition, a phone line is used to control and steer the chain of measurements from the University. We noticed four periods of the year with a relative constancy of the average global solar radiation. These four periods are the following: from February to April which corresponds to a dry season, an intermediate season May to July, rainy season from August to November and an other intermediate season from December to January. Our measurements show that the daily solar radiation lasts at most 12 hours (in March and April) and at least 10 hours (from October to December). For La Désirade, Petit-canal and Fouillole sites we have 100% data in 2012 (without holes measures). As we are interested in the variations of the global solar radiation, we are only concerned with the global solar radiation signals from 7 am to 5 pm. We perform VAR model using one year (2012), for 10 min global solar radiation data collected at 3 meteorological stations (La Désirade, Petit-canal, Fouillole) across Guadeloupe island. According to the needs of the manager of electrical network, we focus forecasting of global solar radiation at 10 min ahead.

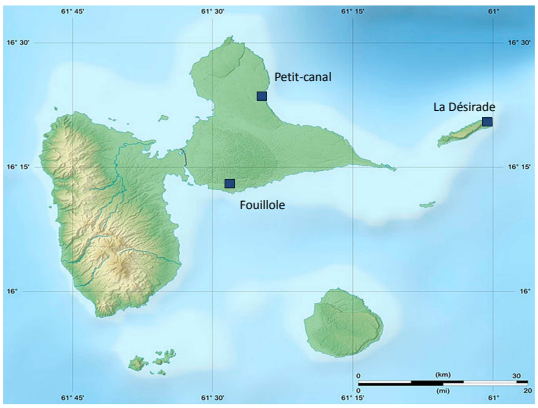

Figure 1. Guadeloupe archipelago and geographical location of our three measurement sites: Petit-canal, Fouillole Campus, La Désirade.

From	To	Distance(meter)
Petit-canal	Fouillole	26272 m
Petit-canal	La Désirade	41818 m
Fouillole	La Désirade	55819 m

Table 1. Distance between sites in meters

2.2. Approach

Temporal and spatial behaviors of solar irradiance are related through complex atmospheric mechanisms. The measurement stations are presumably subject to different microclimates (different islands or leeward, windward sides

of islands). According to the classification of C. Brévignon [5], the site of Petit-Canal is located in the climatic conditions of windward coast, where the wind blows constantly. In this zone, the formation of cloud is mainly due to the advection of marine air masses. La Désirade is in the same meteorological regime but it is an insular dependence separated from continental Guadeloupe by the Caribbean sea. Fouillole station is subject to insular continental regime. In this area, wind is lower than the wind blowing off the coast. Cloud formation is mainly due to the convection of air masses. The Désirade site is the most easterly one. The prevailing trade-winds, have a strong East component, consequently, Désirade site is a particular interest for a spatio-temporal analysis. It is situated on the first land where the air mass, after a long period over the ocean, meets the relief of the land (Petit-canal and Fouillole stations), which leads to formation of clouds. Consequently, a dynamic relation between sites can be observed. In Boland (2015) [2], forecasting of solar radiation series at these three sites in Guadeloupe has been performed by CARDS tools. In this paper, the model proposed is based on VAR methodology. The proposed model is specifically designed for the analysis of spatio-temporal data which are rich in the time dimension and sparse in spatial dimension. We will develop the forecasting of solar radiation by a multivariate method in using spatio-temporal parameters with three stations. Strategy and algorithms allowing to optimize temporal and spatial parameters in the process of model will be described in section 4.2, section 5.2 and section 5.3. Preliminarily, we quantified the correlation existing between stations.

3. Correlations between sites

The crosscorrelation as function of time lags provide measures of the similarity of two stationary time series.

3.1. Detrended time series

Detrended time series must be used in the calculation of correlations between a pair of stations since detrending captures the underlying true correlation between two time series (Yang et al 2014 [29],Perez et al (2012) [24]) . We note that there are different methods to achieve stationarity such as local polynomial regression fitting to detrend the solar irradiance time series by first determining an additive diurnal cycle (Yang et al[29]),the differencing technique, used frequently in ARIMA forecasts or the clear sky index often used as a strategy to detrend solar global radiation. Trend removal via the clear sky index cannot remove the trend completely and therefore does not make the series perfectly stationary. Thus we applied the transformation of the clear sky index series to new time series where the new values are the differences between consecutive values. Consequently, detrended time series are obtained by clear sky index and a differencing technique as already used in Perez et al(2012) [23]:

$$\Delta K_c = K_c(t) - K_c(t-1) \tag{1}$$

The calculation of the deterministic component is based on Kasten model [13] to account for the clear sky index. This technique offers three advantages:

(1) normalizing variability to unity

(2) removing the effect of daily solar trend

(3) avoiding to falsify correlation

An example of signal ΔK_c for a day (figure 2) shows the detrended time series.

3.2. Lag correlation between a pair of station

We studied simple correlation of ΔK_c time series at 10 min time scales between our stations. We found very low correlation with distances of stations pairs from 20 km to 60 km (Table.1). Indead, Perez et al [23] showed there is no simple correlation for distances between stations of 4 km and 10 km for fluctuations time scales of 5 min and 15 min. We can estimate the correlation between pairs of stations by crosscorrrelation function. For this statistical analysis, we repeat the daily maximum crosscorrelation experiment between each pair of stations for the entire year. The coefficients of station pair correlation resulting from 366 days are summarized in Table.2. A maximum correlation (in absolute value) 0.64 is observed and the absolute mean value is between 0.34 and 0.36. These values are representative

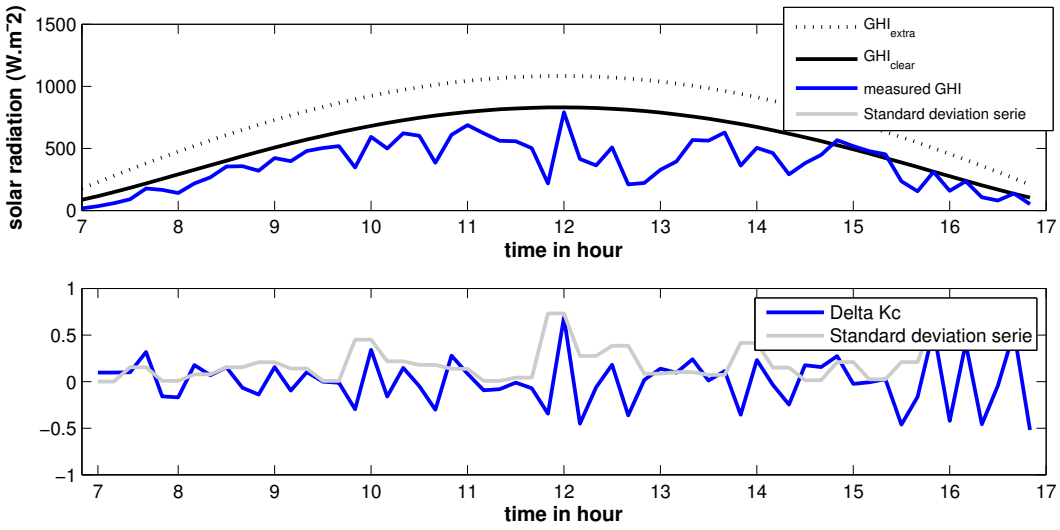

Figure 2. Sample high-variability day showing 10min global iarradiance, GHI_{extra} extraterrestrial theoretical global irradiance , GHI_{clear} theoretical irradiance receipted by sensor and ΔK_c signal.

of correlations at a medium level. With only three pairs of sites it is impossible to give a valid conclusion, nevertheless we can observe that the correlation in our case doesn't depend on distance between stations every day. The closest pair of station according to the distance metric doesn't show the highest average of coefficient (Table.2). Indeed, we can note in Table.2, that the highest absolute mean of correlation can be observed for the pair of stations Petit-canal/La Désirade. This result can indicate that wind direction is prevailing on distance parameter. A test of different spatial structures to optimize forecastings will be performed in section 5. We also quantify the daily temporal lag between two stations using the maximum crosscorrelation of two ΔK_c time series. The distribution of results of time lag for the whole pairs of stations is shown by histogram in Figure 3. This indicates the highest occurrence for time lag between [-1h; 1h] with an average equals to 25 min. These results can describe that whatever happens at one station may cause a similar event at an other station about 1 hour later on that day. The results of intercorrelation between stations pairs are in agreement with those of Boland (2015) [2].

	Min absolute value	Max absolute value	Mean absolute value	Standard deviation
Petit-canal/Fouillole	0.19	0.64	0.35	0.08
Petit-canal/La Désirade	0.21	0.57	0.36	0.07
Fouillole /La Désirade	0.19	0.59	0.34	0.06

Table 2. Intercorrelation coefficients between pair of sites

4. Spatio-temporal Vector Autoregressive VAR model for spatially sparse data

4.1. Forecasting model

We assessed forecastings of the global solar radiation on an individual site in using two processes: the first is based on linear regression of the past time series of this site (AR(1) model) and the second is based on a multiple linear regression deducted from the linear combination of the past time series of the located site and the past time series of other stations. Then, we assessed the results of residuals. We noted that forecastings are better with a linear combination of past time series from all stations than an auto-regression process.

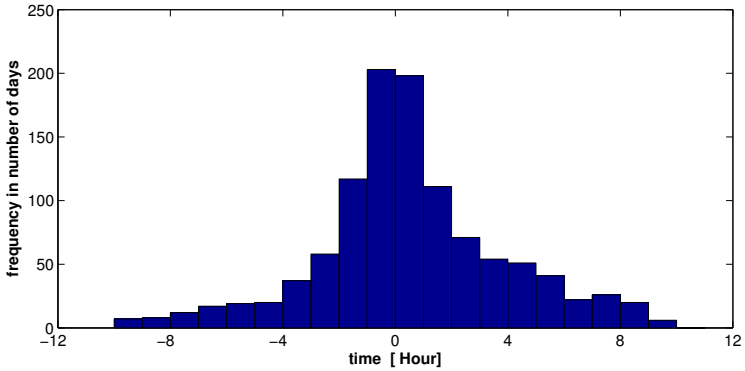

Figure 3. The distribution of the daily time lag obtained by cross correlation function for the whole of pairs of stations.

Consequently, the use of the past temporal series of global solar radiation measured on three stations brings an additional information and improves the predictability of the located global solar radiation. Although the correlation between sites presents a medium level, these results showed the interest to develop a multivariate model for our context of study such as the one proposed in this research which will provide consequently an improvement of forecasting at individual station. Moreover, initialization parameters and an iterative strategy in the process of our model will select the necessary stations in the linear combination for each forecasting day, removing the useless variables. The developed model here, differs from the traditional linear model and is based on a spatio-temporal VAR model methodology which is validated and described in de Luna and Genton (2005)[11]. This model uses spatial and temporal parameters and can be performed for rich data in time dimension but sparse in spatial dimension such as the example showed in de Luna and Genton (2005) ([11]) with four monitoring stations for hourly observations of atmospheric concentration of carbon monoxide. The generality of the VAR method is implied by treating each spatial location separately in the modeling process and consists of a vector autoregressive (VAR) specification, thereby avoiding restrictive and often difficult to verify spatial-stationarity assumptions [11]. Indeed, spatial stationarity assumption are arbitrary, it is not possible to assess their validity for few sites. Moreover, wind speed and direction can be expected to influence air pollutant concentrations in a nonstationary and anisotropic away for the example given in [11].

Consider spatio-temporal data $Z(s_i, t)$ concerning observations of some stochastic process indexed in $\mathbb{R}^2 \times \mathbb{R}$. Each observation is made at located station s_i, $i = 1, ..., N$ and time $t = 1, ..., T$. The considered predictive VAR (vector autoregressive) model (see De Luna and Gneton, 2005) is

$$Z_t - \beta = \sum_{i=1}^{p} R_i (Z_{t-i} - \beta) + \varepsilon_t \tag{2}$$

where p is the order corresponding to the time lag; $Z_t = (Z(s_1, t), Z(s_2, t), ..., Z(s_N, t))'$ are the spatio-temporal data, $\varepsilon_t = (\varepsilon_t(s_1), ..., \varepsilon_t(s_N))'$ is a white noise with $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_u) = 0$ for $u \neq t$, $E(\varepsilon_t \varepsilon_t') = \sum_\varepsilon$, $\beta = (\beta(s_1), ..., \beta(s_N))'$ is the spatial trend, R_i is a $N \times N$ unknown parameter matrices. The N rows of these matrices correspond to the N locations at which time series are observed.

Estimation of the parameters in equation (2) can be obtained with maximum likelihood (if distributional assumptions are made), with least squares or with moments estimators (Yule-Walker type), for more details see Lütkepohl (Chap.3), Penã et al. (Chap.14)[22] and De Luna and Genton [11]. Deterministic trend is often removed by differencing with the difference operator of order d with $d = 1$:

$$\nabla Z(s_i, t) = Z_t(s_i, t) - Z_t(s_i, t - 1) \tag{3}$$

and the spatial trend is estimated as

$$\beta(s_i) = E(\nabla Z(s_i, t) - \nabla Z(s_i, t - 1)) \tag{4}$$

$\beta(s_i)$ is supposed to be depending only on s_i.

The stationary spatial-temporal data are obtained:

$$\tilde{Z}(s_i, t) = \nabla Z(s_i, t) - \beta(s_i) \tag{5}$$

4.2. Selection of predictors

This model is performed either by displaying sample partial correlation functions, or by minimizing an information criterion. The deletion of uninteresting predictors at each time lag improves on efficiency by avoiding the estimation of zero coefficients [11]. Suppose that we want to predict the value of the process for a station s_j at time t; $Z(s_j, t)$. To this aim consider predictors in the following order :

$$Z(s_j, t-1), Z(s(1), t-1), Z(s(2), t-1), ..., Z(s(N-1), t-1), Z(s_j, t-2),$$

$$Z(s(1), t-2), ..., Z(s(N-1), t-2), ...$$

where $s(1), ..., s(N-1)$ is an ordering of the $N-1$ stations, for instance, in ascending order with respect to their distance (using a given metric) to s_j. Let us look at partial autocorrelations by renaming the previous sequence as

$$X_1 = Z(s_j, t-1), X_2 = Z(s(1), t-1), ..., X_N = Z(s(N-1), t-1);$$

$$X_{N+1} = Z(s_j, t-2); X_{N+2} = Z(s(1), t-2), ..., X_{2N} = Z(s(N-1), t-2), ...$$

Let the partial correlation function (PCF) for station s_j as $\rho_{Z_{s_j}}(h) = Corr(Z(s_j, t), X_h | X_1, ..., X_{h-1})$. Define h_1 to be such as $\rho_{Z_{s_j}}(h_1) \neq 0$ and $\rho_{Z_{s_j}}(h) = 0$ for $h_1 < h \leq N$. Similarly, a value h_i can be defined for each time lag i, such $\rho_{Z_{s_j}}(h_i) \neq 0$, $\rho_{Z_{s_j}}(h_i) = 0$ for $h_i < h \leq iN$. The h_i's orders can be identified by looking at the sample partial correlation function $\hat{\rho}_{Z_{s_j}}(h) = \hat{Corr}(Z(s_j, t) - \hat{P}(Z(s_j, t)|X_1, ..., X_{h-1}), x_h - \hat{P}(X_h|X_1, ..., X_{h-1}))$ where $\hat{P}(Z(s_j, t)|X_1, ..., X_{h-1})$ is the best linear predictor of $Z(s_j, t)$ given $X_1, ..., X_{h-1}$.

Algorithm

Step 0: Choose one of the observed sites s_j.

Step 1: Identify h_1 by looking at the sample $\hat{\rho}_{Z_{s_j}}(h), h = 1, ..., N$

Step 2: Identify h_2 by looking at the sample $\hat{\rho}_{Z_{s_j}}(h), h = N+1, ..., 2N$ when $X_{h_1+1}, ..., X_N$ have been discarded as unhelpful in explaining $Z(s_j, t)$ in the previous step.

Step 3: Identify h_3 by looking at the sample PCF (partial correlation function) $\hat{\rho}_{Z_{s_j}}(h), h = 2N+1, ..., 3N$ when $X_{h_1+1}, ..., X_N$ and $X_{h_2+1}, ..., X_{2N}$ have been discarded as unhelpful in explaining $Z(s_j, t)$ in the previous steps.

Step 4: Step 3 is repeated in a similar manner for all necessary time lags in order to identify h_4, h_5, ...

Step 5: Repeat the previous steps for all observed sites.

An alternative to the use of the PCF is the use of an automatic model selection criterion in each step of the identification strategy which we developed in our model.

5. The spatio-temporal VAR(p) model for global solar radiation

5.1. Application of model

We are interested in the forecasting of the daylight period of global solar radiation which corresponds to time series of global solar radiation between 7 am to 5 pm. The data are by 10 minutes step. We perform our analyse in using one year of data (2012), there are no missing values in this data set. Consequently, the data set $(Z(s_i, t))t = 1, ..., 60, s = 1, ..., 3$ available has 600 minutes observations of global radiation where $T = 60$ by 10 minutes step, at $N = 3$ monitoring stations located in Guadeloupe (Désirade, Petit-Canal, Fouillole). Note that equation (2) is

appropriate once the temporal trends have been removed from the signal. As mentioned previously, for temporal detrending we use the parameter ΔK_c. Consequently, deterministic trend is removed by differencing clear sky index from the difference operator of order d with $d = 1$ and we take the logarithm of the observations, thereby stabilizing the variance:

$$\nabla Z(s_i, t) = log(K_c(s_i, t)) - log(K_c(s_i, t-1)) \tag{6}$$

$$\nabla Z(s_i, t) = \Delta(log(K_c)) \tag{7}$$

and the spatial trend is estimated as

$$\beta(s_i) = E(\nabla Z(s_i, t) - \nabla Z(s_i, t-1)) \tag{8}$$

$\beta(s_i)$ is supposed to be depending only on s_i. The stationary spatial-temporal data are obtained. We performed the Dickey-Fuller test to assess stationarity in our detrended series. The hypothesis test is based on searching for unit root in the time series autocorrelation model. In other words, if the observation at time t strongly depends on the observation at time $t-1$ with coefficient larger than 1, the series are defined to be non-stationary. The result indicates evidence in favor of the null hypothesis which implies stationarity. Estimation of the parameters in equation (2) is obtained with least squares. For more details see Lütkepohl (Chap.3) and Penã et al. (Chap.14)[22], De Luna and Genton [11].

5.2. Selection of p-order of VAR(p) model

We assessed forecastings at 10 min ahead for different p-orders of VAR(p) model in the same day. Then, we computed the statistical errors (RMSE) of forecasting for each p-order. These tests are assessed for a sequence of 31 days. The lowest statistical errors show the selection of an optimal p-order varying from day to day. Obtained optimal p-orders for a sequence of 31 days is showed on Figure 4. Thus, we built an algorithm in the process of our model selecting optimal p-order for each day by information criterion AIC and BIC from the initialization points data of model. This algorithm brings a supplementary optimization of model with the strategy previously described in section 4.2. After execution of model process, we attempted to compute the daily occurrence of selected optimal p-order of model for each station (Figure 5). A global result can be given: the p-order equals to 1 corresponding to 10 min lagged values at each station is the highest occurrence. Thus, globally, the optimal model is the spatio temporal VAR (1). This indicates that the historic of stations at 10 minutes lag time explains with high occurrence the data of an individual localized station at the moment t. According to the results of coecients crosscorrelations (section 3.2) 25 minutes is the averaged time lag giving the strongest significant correlations. The crosscorrelation highlights the temporal dependencies, it aggregates the correlation between three sites. The spatial parameter doesn't take into account, consequently the coefficient of crosscorrelation is just indicative. Moreover, the past data to t-10 min give more informations compared to observations to t-25 min in the process of model, this can be another reason. The optimal p-orders equal to or higher than 4 (40 minutes or more) have a very low occurrence.

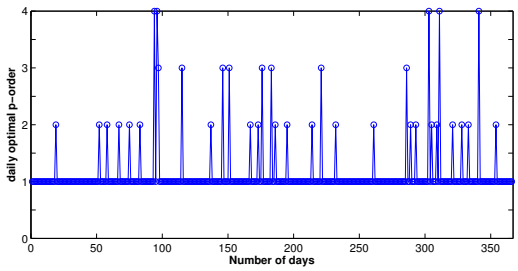

Figure 4. Daily optimal p-order spatio-temporal VAR model for example of 2 months data.

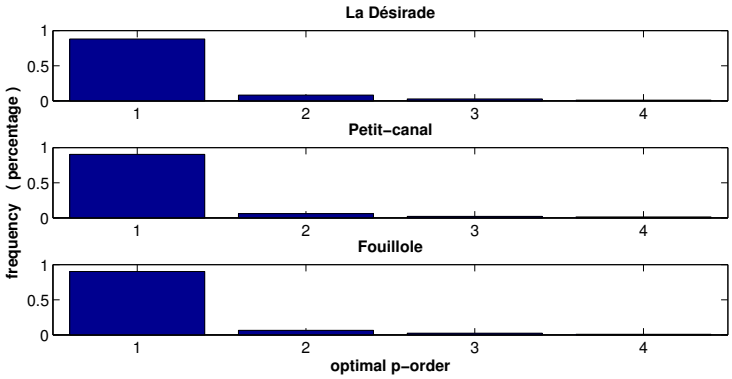

Figure 5. Occurrence in number of days of optimal spatio temporal VAR p-order for each station

5.3. *Selection of spatial order of VAR model*

The defined algorithm in section 4.2 may be used in our model as an alternative with the criterions AIC (Akaike information criterion, Akaike [1]) and BIC (Bayesian information criterion, Schwarz [**?**]). The former being usually preferred for predictive purposes and gives best results in our case. We used this alternative in our process model. Our VAR model is implied by treating each spatial location separately in the modeling process.The fact that predictors can be entered in the model sequentially in the model building stage thanks to this algorithmic strategy, allowed to know how many of predictors or stations should be used for each daily forecasting, avoiding useless variables. Figure 6 presents normalized RMSE for different linear combination will consist of past time series of one station, two stations or three stations. Figure 7 completes this analysis by a histogramm of frequency of the best linear combinations obtained on the whole of year. These figures show that the knowledge of the past temporal series of solar radiation measured on three or two stations brings an additional information and improves the predictibility of the localized solar radiation. Numerous researchers demonstrated the improvement of forecast quality by multivariate models techniques, in particular by including a spatial information. We can quote Glasbey and Allcroft (2008), Yang and al (2014), Bessa and al (2015). Moreover, these figures show the wealth of the model which can behave like a model AR (one station) or a model spatio-temporal VAR (several sations) accordint to selected stations in the modeling process

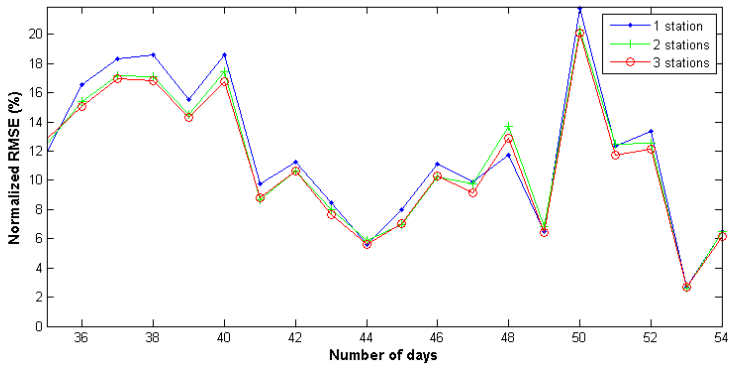

Figure 6. Daily RMSE for for different linear combination will consist of past time series of one station, two stations or three stations (an example of 75 days sequence.

By taking into account the spatio-temporal structure, it becomes possible to define an ordering with which to sequentially introduce the predictors in the model for z(s; t). We investigated the spatial ordering of the sites with a prediction performance analysis where two different ordering of the stations are examined. For this analysis, we compared the statistical errors of daily root mean square errors (nRMSE) for two spatial structures. The first spatial

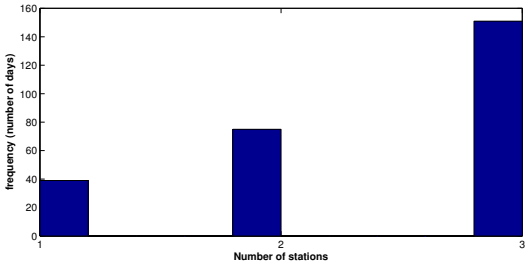

Figure 7. Frequency in number of days of best linear combinations selected with the best RMSE obtained on the whole of 2012 year

structure motivated by physical knowledge, defines an ordering of the stations according to the prevailing wind direction. We took into account that the Trade wind has predominantly a direction from east to west over Guadeloupe island. Consequently, if we want to predict at a station, the linear combination will consist of past time series of the station then the past time series from easterly station, and third the last station. These orders of past time series are different if we consider an ascending order with respect to the distance between stations in using information from Table 1.

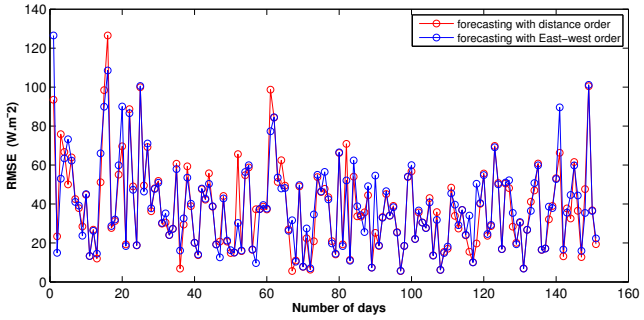

Figure 8. Normalized RMSE of two models including each a different spatial structure in the spatio-temporal process VAR model (order of locations respecting the Trade wind in blue color and order of locations respecting ascending order distances in red color) for 5 months for example of Fouillole station

We found values are more or less equal between the two type of spatial orders of locations but for particular cases we have some better values for the order of locations respecting the prevailing wind direction, particularly for june, october, november and december periods which is in agreement with the cyclonic season particularly characterized by rainy period and higher wind speed.Figure 8 show an example for 160 days sequences. We found that the model whose spatial structure takes into account the predominant wind direction (wind order) has better or equal predictive performance at probability of 25% for Petit-canal station along the year, 17% for La Désirade station and 8% for Fouillole station, than the model ignoring this physical information (distance order). In section 6.2, it will be shown that for certain forecasts horizons the influence of predictors spatial order is more marked.

Consequently, in order to optimize our forecasting model of global solar radiation time series, we use the spatial structure of stations (or order of predictors) according to the prevailing wind direction during the rainy season.

6. Predictive performance of model

6.1. Results for different daily global radiation classes

To put in evidence the influence of the solar radiation variability on the forecast results we performed a forecast on a given class of days. This classification of daily global solar radiation is performed by an algorithm based on k-means method. The classes found have the same characteristics than those found in Soubdhan et al [27] who used

a mixture of Dirichlet distribution: clear sky to cloudy sky days. A representation of averaged signals of each class is shown (Figure 8) with an example of day global solar radiation curve for each class (Figure 9).

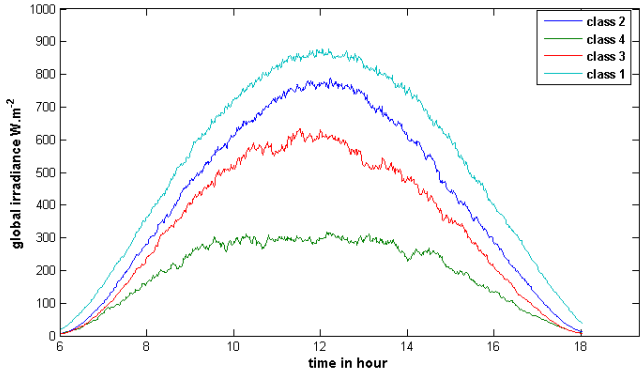

Figure 9. Average daily curve of four classes obtained by k-means method.

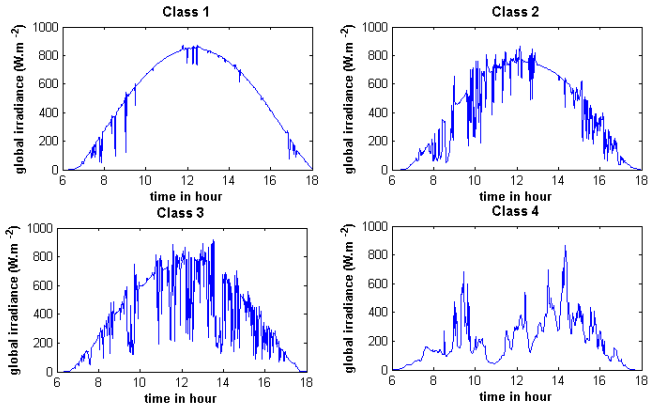

Figure 10. Example of days respectively being representative of class 1; class 2; class 3; and class 4.

The results of normalized MAE (nMAE) are presented by histograms for each quarter and each class (Figure 11). The normalized MAE is described by equation (9).

$$nMAE = \frac{\frac{1}{n}\sum |\,(G - \hat{G})\,|}{max(G) - min(G)} \times 100\% \tag{9}$$

where G measured values and $\hat{G}$ predicted values.

Normalized MAE (nMAE) values are always inferior to 10% which shows a good performance of the model. The class which presents the lowest statistical errors is class 1. This class is representative of clear sky conditions of solar radiation days with very few clouds and thus a very slow dynamic, as shown in Figure 8. Class 3 presents the highest statistical errors. This class is representative of days with significant sunshine combined with a large number of small clouds with high speed of passages and thus with high dynamic levels. The medium errors results are obtained for class 2 which is representative of days with an important solar radiation with some clouds corresponding to a medium level dynamic as shown in Figure 9. The class 4 can present the highest statistical error for few months but also the lowest (from Table 3 to Table 5). This class is representative of completely cloudy sky days with big size clouds. In this case the solar radiation is mainly scattered by clouds and presents low values of global solar radiation. When the cloudy mass, scattering solar radiation, has a slow speed, the dynamic level is very slow which can explain the

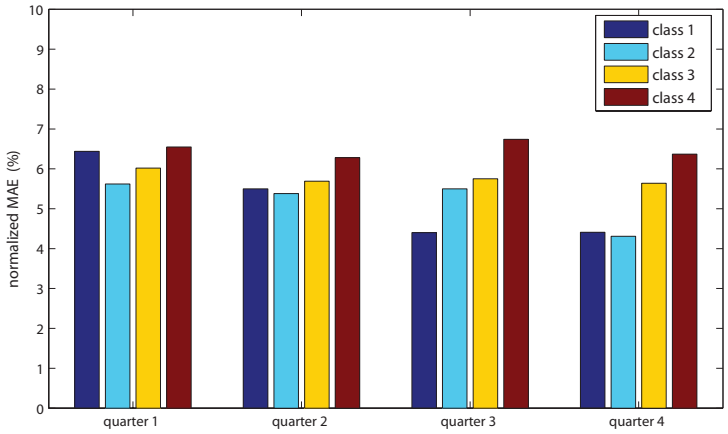

Figure 11. Normalized mean absolute errors (nMAE) in percentage of forecastings of global solar radiation for all stations for each quarter corresponding to a season

possibility of low values of statistical errors. According to the results, we can say that performance of model decreases when solar radiation signal presents high dynamic level. The model has difficulties to accurately predict values when variations are speed and brutal. However, the statistical errors values show a good performance of spatio-temporal model and globally doesnt exceed 10% whatever dynamic (fast or slow)of fluctuations of measured signals.

6.2. Results for different horizons

In this section we assessed our model for different forecasting horizons according the two spatial structures defined in section 5.3. The yearly normalized RMSE is described by equation(10).

$$nRMSE = \frac{\sqrt{\frac{1}{n}\sum(G-\hat{G})^2}}{max(G)-min(G)} \times 100\% \tag{10}$$

	La Désirade station				
	5 min	10 min	15 min	30 min	1H
distance order	10.79	13.13	11.51	13.24	15.15
East-west order	10.79	13.13	11.51	13.24	15.15
	Petit-canal station				
	5 min	10 min	15 min	30 min	1H
distance order	10.52	11.12	10.53	11.60	16.63
East-west order	10.59	11.12	10.53	11.59	14.10
	Fouillole station				
	5 min	10 min	15 min	30 min	1H
distance order	13.19	13.56	13.28	13.29	20.43
East-west order	12.68	13.56	13.28	13.28	15.97

Table 3. Yearly normalized root mean square errors (nRMSE) for different forecasting horizons according two spatial orders of predictors for each station.

The results of performance predictive show values of yearly nRmse which don't exceed 21% whatever horizons, which presents a good predictive performance of model. We will compare it with other methods in literature in section 6.3.

This table highligths some spatio-temporal information about the dynamic system compound of our three stations. We can observe that spatial order of predictors has no strong influence on model performance for horizons lower than 1H. This can be explained by the fact that, from 5 min to 15 min, the variability of signals are not explained most often by clouds moving from a station to another. The variation of solar radiation for this timescale is presumably subject to different microclimates. As we explained in section 5.3, for example of 10 min horizons, forecasting using East-west order of predictors improve the model until a daily frequency of 25% on the whole of year. Moreover, particularly at 5min horizons, it is unlikely that solar radiation measured at a localized station can be explained by solar radiation measured at other stations in the past. The results allow to observe also that 15 min is the timescale threshold for that the stations ordered according a spatial structure respecting the predominant wind direction in the process of model improve performance predictive of solar radiation at a localized station. By observation of statistical errors results, at 1h horizon, the influence of wind direction is well marked.

An example of forecasting for 4 days sequence for each horizon is presented in Figure 12.

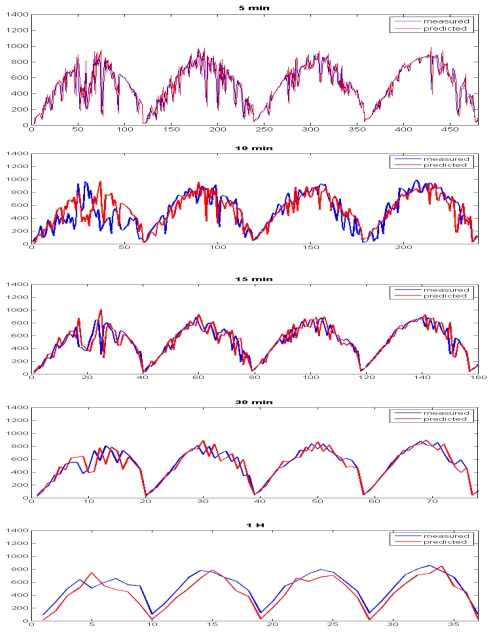

Figure 12. Measured global solar radiation (blue) and predicted (red) signals for four days with different solar radiation variability and different horizons respectively 5min, 10min, 15min, 30min, 1h.

6.3. Comparison with other methods of the literature

If we use the literature we can compare the forecasting results of spatio-temporal VAR model with other models (Table.4). For 10 min horizon, Chu et al [8] in using 1min average predict direct normal irradiance found the best performance values $54.6 W.m^{-2}$ for low DNI and $132 W.m^{-2}$ for high DNI (direct normal irradiance) with a model combining sky image processing and ANN and Marquez et al [20] in using total sky imager found $283 W.m^{-2}$ the best value of RMSE. For our model, one of the best values of RMSE for low dynamic global solar flux is $18 W.m^{-2}$ and for high dynamic is $76 W.m^{-2}$. In Glasbey paper [14] the performance of STARMA model at 10min time scale forecasting, is described by a comparison of averages of estimated parameters fitted STAR (1) and true values parameters. In Chi Wai Wow [7], the performance at 30s to 5 min horizon is described by mean and standard deviation of matching errors between the two cloud maps. We performed our model for 5min ahead and 15min ahead. The results for our model spatio temporal VAR, are computed with the average of yearly nMAE and nRMSE for all sites. We also computed the statistical errors of persistence model for our data set at 30 min ahead and 1h ahead. The results of predictive performance found in literature and of our model are presented in Table.4 at different horizons.

Forecast horizon	Forecasting model in the literature	Best statistical error(%)
5 minutes	kriging model[29]	nRmse :18.49
5 minutes	shrinkage in VAR model [29]	nRmse18.09
5 minutes	Regression [26]	nMAE:12.63
5 minutes	ARIMA[26]	nMAE:13.21
5 minutes	Neural Network [26]	nMAE:12.78
15 minutes	Regression [26]	nMAE:26.52
15 minutes	ARIMA[26]	nMAE:18.97
15 minutes	Neural Network [26]	nMAE:21.04
30 minutes	Model At [10]	1-RMSE/RMSEp:5.68
30 minutes	Model Ast2 [10]	1-RMSE/RMSEp:9.45
30 minutes	CMF [10]	1-RMSE/RMSEp:8
Forecast horizon	Forecasting spatio-temporal VAR model	Best statistical error(%)
5 minutes	spatio-temporal VAR model	nRmse:8.46
5 minutes	spatio-temporal VAR model	nMAE:5.50
15 minutes	spatio-temporal VAR model	nMAE:11.33
30 minutes	spatio-temporal VAR model	1-RMSE/RMSEp:4.52
1H	spatio-temporal VAR model	nRmse:15.07
1H	spatio-temporal VAR model	1-RMSE/RMSEp:23.42

Table 4. Comparison of the forecast errors with different models of the literature

On the whole, we can note our model gives good results in comparison with other models in the literature (Table.4). RMSE or nRMSE values highly depend on the meteorological conditions. Forecast for a clear location would always be associated with lower RMSE. One way to overcome this, is to compare the metric error computed by the ratio of model RMSE and persistence model RMSE used by Dambreville et al (2014) [10] for At, Ast, Ast2 models. Dambreville et al (2014) [10] presents an original method to forecast the GHI at ground level using the HelioClim-3 maps which estimate the GHI from satellite images. The originality of this work comes from the integration of spatio-temporal information from satellite images without any cloud motion field vector calculation. The metric error computed by the ratio of model RMSE and persistence model RMSE allowed to have an accurate comparison between At, Ast, Ast2 models [10] and our model (Table 4). Moreover, as our results are derived from data in Guadeloupe, which has a highly time-variable tropical climate, we expect that the skill of our method will be even higher at other locations. In order to do a more valid comparison, in future work we will apply the different models to our data set. This table shows an excellent performance of our model for 1 H ahead with 23.42% of metric error and only a normalized RMSE equals to 15.07%.

7. Conclusion

In this paper, we investigated the use of a spatio-temporal VAR model for forecasting global solar radiation at 10 min ahead with experimental data recorded at three locations. The originality of this work comes from a spatio-temporal VAR model performed for data rich in the time dimension and sparse in spatial dimension. The proposed model is specifically designed for the analysis of spatio-temporal data sets with the purpose of providing time-forward predictions at given spatial locations. The predictions are based on a minimum of assumptions since treating each spatial location separately in the modeling process, which allows performing model without spatial-stationary assumptions in contrast with Tonellato [28] who used a spatial stationary isotropic exponential correlation function. This model is based on de Luna and Genton [11] methodology. Comparative studies on global solar radiation forecasting for several time lag and spatial structures are presented and discussed. Thus, an optimization of the model was first performed. A study of the spatial order (order of locations), motivated by physical knowledge respecting the wind direction allowed to improve the spatio-temporal VAR model performance for some periods in the year. Another parameter of the model such as the p-order was also selected optimizing the process model to give better results of

predictions. The most frequent optimal p-order was found to equal to 1 or spatio-temporal VAR (1) model. The influence of the global solar radiation variability on the MAE and MBE was assessed. The model was tested on typical class of solar days having specific variability. Even if, the signals can present high dynamic levels, the statistical errors (MAE and MBE) of predictions are mainly inferior to 10%. A comparison with the simple persistence model and other models found in literature was made. Our model shows better results than krigring model or shrinkage VAR model particularly a good performance at 1H ahead comparated to other methods certainly due to spatio-temporal information. In future work, we will benchmark the proposed VAR model with other models, in applying literature models for our data in order to have a valid comparison.

Our results allow estimation of the ancillary services required to operate distributed PV sites and will be able to decrease the variability and uncertainty that must be managed by system operators and planners of PV system. Thus, the integration of our forecasting model in the process of PV system offers an opportunity to provide guarantees to a solar energy network manager. The good performance of this model shows the importance to take into account spatio-temporal parameters rather than simple temporal models in our study case.

References

[1] Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Proceedings of the Second International Symposium on Information Theory. Budapest, Hungary, pp. 267–281.

[2] Boland, J., 2015. Spatial temporal forecasting of solar radiation. Renewable Energy 75, 607–616.

[3] Bosch, J., Kleissl, J., 2013. Cloud motion vectors from a network of ground sensors in a solar power plant. Solar Energy 95, 13– 20.

[4] Bosch, J., Zheng, Y., Kleissl, J., 2005. Deriving cloud velocity from an array of solar radiation measurements. Solar Energy 87, 196–203.

[5] Brevignon, C., 2005. L´environnement atmosphère de la Guadeloupe, de St Barthélémy et de St Martin. La direction Inter Régionale Antilles Guyane de Météo France.

[6] Cellura, M., Cirrincione, G., Marvuglia, A., Miraoui, A., 2008. Wind speed spatial estimation for energy. Renewable Energy 33, 1251–1266.

[7] Chow, C. W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., Washom, B., 2011. Intra-hour forecasting with a total sky imager at the uc san diego solar energy testbed. Solar Energy 85, 2881 2893.

[8] Chu, Y., Pedro, H., Coimbra, C., 2013. Hybrid intra-hour dni forecasts with sky image processing enhanced by stochastic learning. Solar Energy 98, 592 603.

[9] Coimbra, M., 2005. Linear combinations of space-time covariance functions and variograms. IEE Trans. Signal Process 53, 857–864.

[10] Dambreville, R., Blanc, P., Chanussot, J., Boldo, D., 2014. Very short term forecasting of the global horizontal irradiance using a spatio-temporal autoregressive model. Renewable Energy 72, 291–300.

[11] de Luna, X., Genton, M., 2005. Predictive spatio-temporal models for spatially sparse environmental data. Statistical Sinica 15, 547–568.

[12] Epperson, B. K., 2000. Spatial and space-time correlations in ecological models. Ecology Modelling 132, 63 76.

[13] F., K., 1984. Parametrisierung der globaslstrahlung durch bedekungsgrad und trubungsfaktor. Annalen der Meteorologie 20, 49–50.

[14] Glasbey, C., Allcroft, D., 2007. A spatiotemporal auto-regressive moving average model for solar radiation. Applied Statistics 57, 343–355.

[15] Gneiting, T., 2002. Nonseparable, stationary covariance functions for space-time data. Journal of the American Statistical Association 97, 590–600.

[16] Gueymard, C. A., Wilcox, S. M., 2011. Assessment of spatial and temporal variability in the us solar resource from radiometric measurements and predictions from models using ground-based or satellite data. Solar Energy 85, 1068–1084.

[17] Huang, H.-C., Hsu, N.-J., 2004. Modeling transport effects on ground-level ozone using anon-stationary space-time model. Environmetrics 15, 251–268.

[18] Inoue, T., Sasaki, T., Washio, T., 2012. Spatio-temporal kriging of solar radiation incorporating direction and speed of cloud movement. In: The 26th Annual Conference of the Japanese Society for Artificial Intelligence. Yamaguchi City.

[19] Liu, H., Shi, J., Erdem, E., 2010. Prediction of wind speed time series using modified taylor kriging method. Energy 35, 4870–4879.

[20] Marquez, R., Coimbra, M., 2013. Intra-hour dni forecast based on cloud tracking image analysis. Solar Energy 91, 327–336.

[21] Mills, A., Ahlstrom, M., Brower, M., George, R., Hoff, T., Kroposki, B., Lenox, C., Miller, N., Stein, J., Wan, Y., 2009. Understanding variability and uncertainty of photovoltaics for integration with the electric power system. Electricity Journal LBNL-2855E.

[22] Pena, D., Tiao, G. C., Tsay, R. S., 2001. A Course in Time Series Analysis. Wiley.

[23] Perez, R., Kivalov, S. andSchlemmer, J., Hemker, J. K., Renne, D., Hoff, T., 2010. Validation of short and medium term operational solar radiation forecasts in the us. Solar Energy 84, 2161–2172.

[24] Perez, R., Kivalov, P., Schlemmer, J., Hemker Jr, K., Hoff, T., 2012. Short-term irradiance variability:preliminary estimation of station pair correlation as a function of distance. Solar Energy 86, 2170–2176.

[25] Porcu, E., Mateu, J., Saura, F., 2008. New classes of covariance and spectral density functions for spatio-temporal modelling. Stoch.Environ. Res. Risk Assessm 22, 65–79.

[26] Reikard, G., 2009. Predicting solar radiation at high resolutions: A comparison of time series forecasts. Solar Energy 83, 342–349.

[27] Soubdhan, T., Richard, E., Calif, R., 2009. Classification of daily solar radiation distributions using a mixture of dirichlet distributions. Solar Energy 83, 1056–1063.

[28] Tonellato, S. F., 2001. A multivariate time series model for the analysis and prediction of carbon monoxide atmospheric concentrations. Applied Statistics 50, 187–200.

[29] Yang, D., Gu, C., Jirutitijaroen, Z. D., Chen, N., Panida, Wilfred, M., Walsh, 2013. Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging. Renewable Energy 60, 235 245.