ISES International Solar Energy Society

Conference Proceedings

EuroSun 2016

Palma de Mallorca (Spain), 11 – 14 October 2016

Bias correction of global irradiance modelled with the Weather Research and Forecasting model over Paraguay

Angel Rincón[1], Oriol Jorba[2], Miguel Frutos[1], Leopoldo Alvarez[3], Fernando P. Barrios[1] and Juan A. González[1].

[1] Engineering Faculty, National University of Asuncion (FIUNA), Asuncion (Paraguay).

[2] Earth Sciences Department, Barcelona Supercomputing Center (BSC-CNS), Barcelona (Spain).

[3] Department of I+D, Harmattan Solutions, Santa Cruz de Tenerife (Spain).

Abstract

The estimation of solar irradiance is performed by means of numerical weather prediction models that include all the necessary information to solve the temporal, geographical and atmospheric conditions variability being this the basis of solar energy applications. However, the radiative transfer schemes implemented in meteorological models show systematic errors in the simulation of global irradiance (GHI). In this contribution, we present a post-process analysis of the Weather Research and Forecasting (WRF) model which combines a Kalman Filter with Model Output Statistics for bias correction in order to improve the overall predicted values of GHI simulations over Paraguay. The hourly GHI is simulated at 4x4 km^2 of spatial resolution. The annual evaluation of the hourly WRF model without post process shows relative mean bias error (rMBE) of 21% and relative root mean square error (rRMSE) of 81%. The results using several ground stations and combinations of post-process show an annual correction of systematic errors with rMBE of -0.7% and rRMSE of 70%.

Keywords: Solar irradiance, numerical weather prediction, statistical post-process, Kalman Filter, Model Output Statistics.

1. Introduction

The increased contribution of solar energy to power generation sources requires an accurate estimation of global horizontal irradiance (GHI) conditioned by geographical, temporal and meteorological conditions. The knowledge of the variability of these factors is necessary for estimating energy production and increase its reliability of available solar energy into the global energy participation. The prediction of GHI is performed by means of numerical weather prediction models (NWP).

NWP models simulate the earth-atmosphere system by solving fluid mechanics and thermodynamic equations, which describe weather processes based on initial values and boundary conditions in a nonlinear computing environment (Kimura, 2002). Although NWP models have been advancing rapidly along with the development of modern computing technology, numerical prediction errors are inevitable due to the following (Liang et al., 2007): 1) NWP models cannot exactly describe all the physical processes of the atmosphere (e.g. complex cloud formation); 2) There are random errors in the initial values (observations); 3) Rounding errors accumulate during computing processes; 4) The spatial coverage of observed input is incomplete, especially above the surface.

Specifically, systematic errors of simulations of radiative transfer schemes of NWP models are mainly due to: 1) Miscalculation of location of the clouds and total cloud water content in the layers of the atmosphere; 2) Incorrect specification of the optical thickness of aerosols; 3) Decrease of atmospheric water vapor absorption for clear skies conditions. Therefore, the use of statistical post-processing methods may have the potential to satisfy the requirements of solar irradiance forecasting for up to several days ahead and its application in solar devices (Heinemann, 2006). These post-processing methods seek to improve the accuracy of the forecast by reducing both systematic and random errors, while preserving or improving the correlation with observations (Wilks, 2006). Among different statistical post-processes, the Model Output Statistics (MOS) and the Kalman Filter are the ones most extended within the atmospheric community.

The MOS proposed by Glahn and Lowry (1972), and developed by National Weather Service of United States is often used to forecast weather variables near the surface. MOS relates weather variables at specific locations to NWP model output in a historic dataset by multiple linear regression. In solar irradiance forecasting, Lorenz et al., (2009) and Mathiesen and Kleissl (2011) applied a multivariate fourth-order regression to derive the MOS correction function of solar irradiance in NWP models. One of the limitations of MOS is that a high number of days are needed to adjust the MOS coefficients.

The Kalman Filter method (Kalman, 1960) provides a linear dynamic relationship by estimating the previous error and a correction factor proportional to the forecast error with minor computational cost and easy adaptation to any changes in the observations. The Kalman has been used to correct NWP variables (e.g. Monache et al., 2011; McCollor and Stull, 2008; Roeger et al., 2003). Pelland et al. (2011) present an adjustment of solar irradiance in an NWP model using the Kalman algorithm. Unlike the MOS, Kalman method only needs a short training period and puts more weight on recent data than to older observations. However, Kalman Filter is not likely to predict sudden changes in the forecast error caused by rapid transitions from one weather regime to another (Monache et al., 2011). Thereby, Kalman Filter is unable to predict a large bias when all the biases for the past few days have been small.

Therefore, in this contribution, we present a combination of MOS-Kalman techniques to reduce their limitations when applied separately. The MOS-Kalman post-process has been applied for bias-correction on an annual simulation of GHI computed with the Weather Research and Forecasting (WRF-ARW) meteorological model (Skamarock et al., 2008). The post-process MOS-Kalman is easier to apply than others methods such as Artificial Neural Networks, Bayesian model averaging or analog methods (Zhang et al., 2015). Thus, we use the MOS-Kalman post-process to improve the accuracy of GHI simulated by the WRF-ARW model, significantly reducing the systematic error in all sky conditions for the full range of Sun's vector position for the ground stations of Paraguay.

2. Methodology

2.1. WRF-ARW meteorological model

The WRF-ARW mesoscale model (v3.7.1/2015) has been used to compute the GHI over the geographic area of Paraguay. It is a fully compressible and nonhydrostatic Eulerian model with the latest advances in meteorological mesoscale modelling and incorporates the state-of-the-art of physical parameterizations (microphysics, longwave radiation, shortwave radiation, land-surface model, planetary boundary layer and cumulus parameterization). Furthermore, it is a worldwide reference model used as a research tool and operational weather prediction.

Table 1 summarizes the main characteristics of the parameterizations used by the WRF-ARW model as applied in this contribution. The radiation schemes provide atmospheric heating due to radiative flux divergence and surface downward longwave and shortwave radiation for the ground heat budget. The longwave radiation (RRTM scheme) includes infrared or thermal radiation absorbed and emitted by gases and surfaces, while the shortwave radiation (Dudhia scheme) explicitly considers extinction by Rayleigh atmosphere and water vapor only. The Dudhia scheme (Dudhia, 1989) consists of a simple broadband parameterization of GHI that includes visible and surrounding wavelengths that make up the solar spectrum. It is a downward integration allowing efficiently for clouds and clear-sky absorption and scattering, but it does not account for multiple scattering effects. Extinction by ozone, aerosols, and other molecular absorbers are considered through a bulk scattering parameter that was empirically set to represent average turbidity conditions (Zamora et al., 2005).

The WRF-ARW model is run in diagnostic (hindcast mode) over the South American continent with three nested domains centered over Paraguay having 36 km, 12 km and 4 km horizontal grid resolution, 30 vertical layers and the outputs are stored at hourly temporal resolution. Initialization and boundary conditions are provided by the dataset DS090.0 Reanalysis with information available at 6 hour intervals (NCEP/NCAR, 1994). The GHI hourly simulations consist of 365 daily runs to simulate the entire year 2015. The choice of this specific year is based on the availability of ground stations of GHI for this year (see section 3).

2.2. Post-process description

The Model Output Statistics (MOS) and the Kalman Filter are techniques which are used to estimate the systematic component of the forecast errors reducing the future bias from past forecast and observations in order to improve the forecast. The bias-removal post-processes used are briefly described below.

Tab. 1: Parameterizations schemes used in WRF-ARW model (v3.7.1) for the annual simulation in Paraguay of 2015.

Parameterization	Scheme	Short description
Microphysics	WRF Single-Moment 3-class (Hong et al., 2004)	A simple, efficient scheme with ice and snow processes suitable for mesoscale grid sizes.
Longwave Radiation	RRTM scheme (Mlawer et al., 1997)	Scheme using lookup tables for efficiency. Accounts for multiple bands and microphysics species
Shortwave Radiation	Dudhia scheme (Dudhia, 1989)	Simple downward integration allowing efficiently for clouds and clear-sky absorption and scattering.
Surface Layer	MM5 surface layer scheme (Fairall et al., 2003)	The scheme is sped up to give similar timing as with the old MM5 scheme. The thermal and moisture roughness lengths over ocean are changed to COARE 3 formula
Land Surface	Unified Noah land-surface model (Niu et al., 2011)	Scheme with soil temperature and moisture in four layers, fractional snow cover and frozen soil physics.
Planetary Boundary layer	Yonsei University scheme (Hong et al., 2006).	Scheme with the analysis of the interaction between the boundary layer and precipitation physics, explicit treatment of the entrainment layer at the PBL top and an enhanced stable boundary-layer diffusion algorithm. Allows deeper mixing in windier conditions.
Cumulus	Kain-Fritsch scheme (Kain, 2004)	Deep and shallow convection sub-grid scheme using a mass flux approach with downdrafts and removal time scale.

2.2.1. Model Output Statistics (MOS)

The MOS (Glahn y Lowry, 1972) is a technique that has the ability to predict the systematic error through polynomial regression and applies to objectively improve correlations between simulations and observations. Correction of the systematic deviation in the solar irradiance simulations through the MOS technique is realized following the methodology proposed by Lorenz et al., (2009) and Mathiesen and Kleissl (2011). Equation 1 shows the fourth order polynomial regression established for the present study, which relates the interaction of the atmospheric state and the sun position in the adjustment of the MOS regression in order to minimize the systematic error in the simulations.

$$Bias_C = \varepsilon + \alpha_1 \cdot k_t^* + \alpha_2 \cdot \cos(SZA) + \alpha_3 \cdot (k_t^*)^2 + \alpha_4 \cdot k_t^* \cdot \cos(SZA) + \alpha_5 \cdot (\cos(SZA))^2 + \alpha_6 \cdot (k_t^*)^3 + \alpha_7 \cdot (k_t^*)^2 \cdot \cos(SZA) + \alpha_8 \cdot k_t^* \cdot (\cos(SZA))^2 + \alpha_9 \cdot (\cos(SZA))^3 + \alpha_{10} \cdot (k_t^*)^4 + \alpha_{11} \cdot (k_t^*)^3 \cdot \cos(SZA) + \alpha_{12} \cdot (k_t^*)^2 \cdot (\cos(SZA))^2 + \alpha_{13} \cdot k_t^* \cdot (\cos(SZA))^3 + \alpha_{14} \cdot (\cos(SZA))^4 \quad \text{(eq. 1)}$$

Where k_t^* and $\cos(SZA)$ are incorporated as independent variables of regression, which perform the bias estimation ($Bias_C$) through constant ε and regression coefficients ($\alpha_{1\ldots14}$). The k_t^* (eq. 2) is the index of atmospheric transparency and $\cos(SZA)$ is the cosine of the solar zenith angle, which depend of the modelled irradiance (GHI_{wrf}) and the theoretical attenuation of extraterrestrial irradiance (GHI_{TA}), estimated based on astronomical calculations of Sun's declination, daily and hourly sun angle, earth's eccentricity and solar constant (Duffie and Beckman, 1991; Iqbal, 1983; Spencer, 1971).

$$k_t^* = \frac{GHI_{wrf}}{GHI_{TA}} \quad \text{(eq. 2)}$$

2.2.2. Kalman Filter

The Kalman Filter algorithm (Kalman, 1960) establishes a dynamic linear relationship by estimating the previous error and a correction factor proportional to the forecast error. The Kalman optimizes the systematic error using a recursive adaptation of its coefficients at each time step. This adaptation can reduce the training period between simulations and observations.

The Kalman Filter bias-adjustment is a well-known and widely used technique. According to description of the algorithm presented by Delle Monache et al. (2006), calculating the systematic error (x) for a future time step ($t + \Delta t|t$) is performed by:

$$x_{t+\Delta t|t} = x_{t|t-\Delta t} + \beta_{t|t-\Delta t} \cdot (y_t - x_{t|t-\Delta t}) \qquad \text{(eq. 3)}$$

Where, previous estimation error is $x_{t|t-\Delta t}$, the gain factor is $\beta_{t|t-\Delta t}$ and the forecast error at time t is determined by y_t. The gain factor calculated by equation 4 is the adjustment of recursive algorithm that depends of the expected mean square error (p), the previous error variance (σ_η^2) and the forecast error variance (σ_ε^2).

$$\beta_{t|t-\Delta t} = \frac{p_{t-\Delta t|t-2\Delta t}+\sigma_\eta^2}{(p_{t-\Delta t|t-2\Delta t}+\sigma_\eta^2+\sigma_\varepsilon^2)} \qquad \text{(eq. 4)}$$

The proportion established between the previous error variance and the forecast error variance is a parameter called error ratio:

$$\text{ratio} = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2} \qquad \text{(eq. 5)}$$

The error ratio is a crucial parameter to be calculated in the application of the algorithm that determines the performance from the random error caused by the noise of the number system. Therefore, optimizing the error ratio indicates the relative weighting of the observed and simulated values of the study area. Thus, if the ratio is too high, the error white-noise variance will be relatively small compared to the true bias white-noise variance ($\sigma_\eta^2 > \sigma_\varepsilon^2$). Therefore, the filter will put excessive confidence on the previous simulated values, and the predicted bias will respond very quickly to previous calculated errors. On the other hand, if the ratio is too low, the predicted bias will change too slowly over time. Consequently, there exists an optimal value for this ratio, which can be estimated by evaluating the filter performance in different situations.

2.3. Statistical parameters

The statistical parameters utilized in the present study for the model evaluation are: Mean Bias Error (MBE), Root Mean Square Error (RMSE) and Pearson Correlation Coefficient (r). In a similar manner, the relative percentage values of the MBE and RMSE are utilized, the so-called Relative Mean Bias Error (rMBE) and the Relative Root Mean Square Error (rRMSE). Parameters of statistical distribution are also utilized, which allow analysis of the variation of the central position (median, 50%) and the measurements of non-central positions such as the quartile range 25%, quartile range 75% and quartile range 90%. The aforementioned parameters allow measuring the precision in the pair of values model-observation and postprocess-observation correction.

3. Radiometric ground stations

The radiometric database used in this work belongs to the FECOPROD agroclimatic network based on Davis automatic stations equipped with photodiode sensors. Most of the stations are less than five years old and sampling time is standardized in 15 minutes. The automatic stations compute the average GHI over the sampling time using a local standard time (LST) reference. The time parameter takes into account the daylight saving time (DST) established by the national government. FECOPROD owns approximately 30 stations and manages another 30 stations from diverse owners. Only FECOPROD owned stations with at least 75% annual coverage between January 1 and December 31 of 2015 were used in this work. The 14 stations that met these conditions are summarized in Table 2.

Tab. 2: FECOPROD ground station network over Paraguayan territory with at least 75% cover of year 2015.

#	Station	Altitude (m a.s.l)	Latitude (°)	Longitude (°)	Series long. (month)
1	Asunción	100	-25,2616	-57,5784	58
2	Curupayty	302,1	-25,8425	-54,9712	54
3	Friesland	110	-24,6085	-56,7873	53
4	San Pedro	84,1	-24,0826	-57,0736	15
5	Agua Dulce	139,9	-19,500	-59,7981	20
6	Edelira	114	-26,5766	-55,4975	52
7	Pirizal	156,1	-22,9658	-60,6426	55
8	Neuland	44,8	-22,6714	-60,1175	53
9	Sommerfeld	235,9	-25,4337	-55,6958	50
10	Paratodo	128	-23,2260	-59,6026	44
11	Las Palmas	125	-21,7365	-59,5486	46
12	Bergthal	245,1	-25,2965	-55,583	45
13	Esperanza	227,1	-22,1736	-61,4947	24
14	Concepcion	86,9	-23,4098	-57,4121	20

3.1. Quality control of the data

Before assessing the quality of measurements, the adjustment of time series to the coordinated universal time (UTC) reference and a visual inspection of intra-hourly GHI data were carried out. The further quality test is based on Roesch et al. (2011) empirical method. The authors have investigated the impact of data gaps caused by missing or discarded data on GHI monthly mean assessment for the Baseline Surface Radiation Network (BSRN) database. Two of three procedures proposed in this paper were applied in this work: 1) The "physically possible" procedure aims at detecting extremely large errors in the radiation data, and 2) The "extremely rare" procedure flags data that may occur over very short time periods under very rare conditions. Data of "good quality" are assumed to be inside the "extremely rare" limits. No "acrossing quantities" were carried out since there are not diffuse or direct radiation measurements available. Table 3 summarizes the limits used for data flagging.

Tab. 3: Numerical criteria for quality assessment of radiometric data.

Procedure	Parameter	Lower bound	Upper bound
Physically possible	GHI	$-4\ Wm^{-2}$	$1.5 \cdot S_0 \cdot \mu^{1.2} + 100\ Wm^{-2}$
Extremely rare	GHI	$-2\ Wm^{-2}$	$1.2 \cdot S_0 \cdot \mu^{1.2} + 50\ Wm^{-2}$

Ref: S_0; Solar constant adjusted for Earth-Sun distance μ; Cosine of solar-zenith angle

The number of flagged (discarded) data is a good indicator of measures quality and continuity. Figure 1 summarizes the amount of daily series discarded upon a 60% valid data of all possible intra-hourly diurnal data. i.e. for a daylight duration of 7 hours and a 15min sample time (28 possible data) at least 17 data should be valid to be computed. Most of the dataset performed very good except for Friesland, Agua Dulce and Concepción. The problem in most of the cases are due incompleteness of daily series.

3.2. GHI hourly means computation

For the computation of hourly means of GHI, a similar criteria to the M2 method tested by Roesch et al. (2011) was applied using hourly data instead of monthly data. This adaptation consist in the computation of hourly means but excluding all observations that are outside the most lenient quality flag identified as "extremely rare" limits without filling removed data or originally missing measurements. Finally, means are assessed only for hourly laps with 100% available intra-hourly data after passing the quality control process.

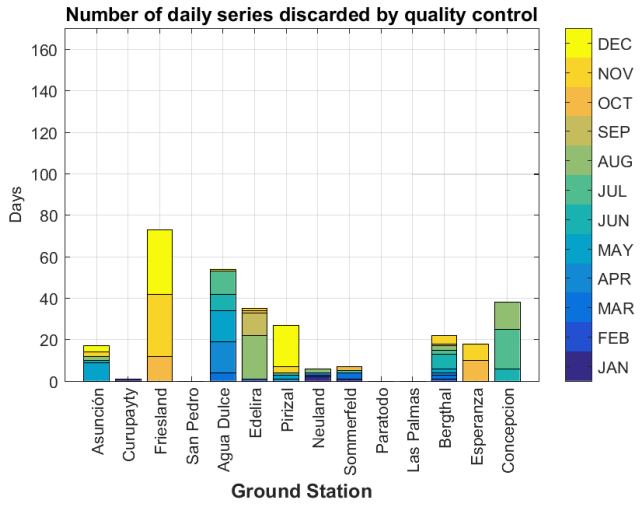

Fig. 1. Consolidated information of discarded data as result of ground data quality-test

4. Annual simulation of WRF-ARW model

4.1. Evaluation of GHI

An annual evaluation of hourly GHI simulated by the WRF-ARW model over Paraguay used fourteen ground stations to analyze the model behavior. The methodology employed for validation is cell-point verification, which compares the punctual-observations with the values of the cells where the observations are located. This methodology supposes that the two values are comparable, although the observation is a temporal value which is spatially punctual, as opposed to the modelled result, which is a temporal value that is spatially averaged. Therefore, the mean stations present an annual bias of 47 W m^{-2} (rMBE=21%), RMSE of 178 W m^{-2} (rRMSE=81%) and correlation coefficient r of 0.87. Figure 2 presents the results of the hourly comparison between the observed and modelled GHI for the average of ground stations in the year 2015. On the left side, figure 2a shows the annual tendency wherein the behaviour of the median (red line), the quartile range 75% (green lines) and the quartile range 90% (blue lines) show a marked overestimation in the GHI range between 200 and 1000 W m^{-2}, which represents almost the entire measurement range. The frequency histogram (blue bars) indicates the distribution of data pairs obs-mod compared, where it shows a gradual decrease with the increase of the range measured. Thus, the compared values vary from about 20000 in the range 0-200 W m^{-2} to less than 2000 between 1000 and 1200 W m^{-2}.

On the right side, figure 2b depicts a box diagram with the hourly quartile range of the modelled GHI (blue rectangles) compared to the observed GHI (gray rectangles). The results allow analysis of the daily cycle distribution between 10:00h and 22:00h UTC, through the behaviour of the median (central black line), the first quartile 25% (lower limit of the rectangle), third quartile 75% (upper limit of the rectangle) and atypical maximum and minimum values (pointed black lines). In essence, throughout the daily cycle persistence in the overestimation of the modelled GHI is observed, which is accentuated between 12:00h and 19:00h UTC, wherein a difference superior to 100 W m^{-2} is highlighted between the quartile 75% and the median. In this way, the overestimation is related to the hourly range of increased incidence of solar irradiance in the day.

The results of the seasonal evaluation are shown in Table 4. Bias and rMBE values for spring-summer are significantly higher (bias>60 W m^{-2}, rMBE>23%) than the values shown for autumn-winter (bias<21 W m^{-2}, rMBE<11%). Regarding to RMSE values, the differences are less significant between spring-summer (RMSE>198 W m^{-2}) and autumn-winter (RMSE<150 W m^{-2}). Likewise, rRMSE values shown a difference of autumn (rRMSE=83%) in relation to the other seasons (rRMSE<75%). Therefore, the seasonal overestimation observed is related to the months of higher incidence of solar irradiance in the year.

In order to assess the model ability to reproduce the GHI with respect to the Sun's position vector and atmospheric conditions, figure 3 shows the annual bias evaluation in function of $\cos(SZA)$ and index k_t^*. The $\cos(SZA)$ values range from 0 to 1, which correspond to the horizon plane (90°) to the zenith (0°). Meanwhile, the index k_t^* establishes the typical weather conditions depending on the cloud cover present in three different scenarios: 1) Clear-sky ($k_t^* > 0.65$); 2) Cloudy ($0.4 < k_t^* < 0.65$) and Overcast ($k_t^* < 0.4$). Therefore, GHI overestimation is observed with higher bias of 80 W m^{-2} for clear-sky and cloudy conditions with zenith angles

less than 60º (dark red zones), while GHI underestimation with a bias below -80 W m^{-2} is observed for all cloud cover and zenith angles greater than 60º (dark blue zones).

Tab. 4: Seasonal evaluation values of WRF-ARW model for 14 ground stations in Paraguay for 2015.

Season	GHI mean (W m^{-2})	Bias (W m^{-2})	rMBE (%)	RMSE (W m^{-2})	rRMSE (%)	r
Summer	292	68	27	207	71	0.87
Autumn	182	12	6	150	83	0.84
Winter	197	21	11	132	67	0.90
Spring	262	60	29	198	75	0.87
Year	223	47	21	178	81	0.87

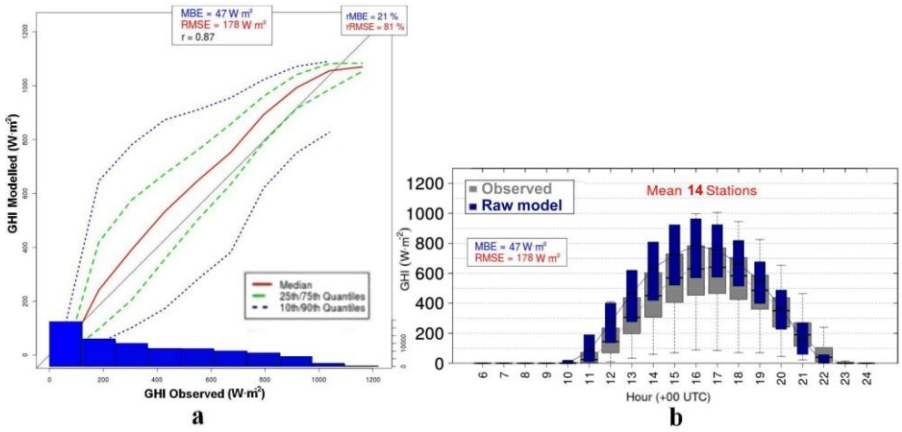

Fig. 2: Comparison of GHI observed and GHI modelled for 14 ground stations in Paraguay for 2015. a) Annual trend of hourly data for median 50% (continuous red line), range of quartile 75% (green dotted lines), range of quartile 90% (blue dotted lines) and frequency distribution histogram (blue bars); b) Daily cycle distribution by hours of GHI observed (gray bars) and GHI modelled (blue bars).

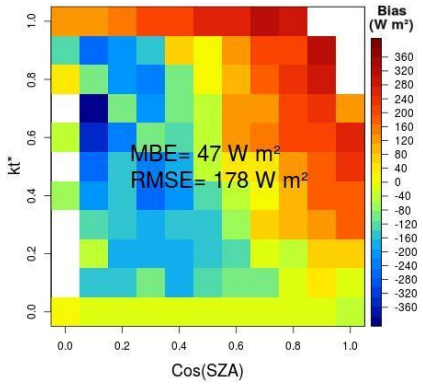

Fig. 3: Bias evaluation of WRF-ARW model in function of cosine of solar zenith angle ($\cos(SZA)$) and index of atmospheric transparency (k_t^*) for 14 ground stations in Paraguay for 2015.

These results of WRF-ARW model evaluation compare with other NWP model evaluations (e.g. Rincon et al., 2013; Mathiesen and Kleissl, 2011; Lara-Fanego et al., 2011; Ruiz-Arias et al., 2008; Heinemann, 2006; Zamora et al., 2005) from different geographical latitudes, which allow to establish similar systematic errors of the shortwave radiation scheme of Dudhia (1989) used in this work.

4.2. Implementation and adjustment of MOS-Kalman post-process

In order to solve the limitations of the techniques MOS and Kalman separately, a combination of adjustments of bias-removal by the algorithms is performed. On the one hand, the adjustment of MOS regression (eq. 1) is realized through a training process that allows obtaining the regression coefficients through the relative weights of the independent variables (Lorenz et al., 2009). The hourly records of k_t^* (eq. 2) and $\cos(SZA)$ are utilized for obtaining the average bias through the numerical interaction of the aforementioned variables (between 0 and 1). Thus, to perform these interactions tests of different time intervals are realized for a period from 7 to 90 days, where a period of training is obtained equal to 60 days for all ground stations of Paraguay. Numerical interactions allow obtaining the relative weights establishing bias of the training data. As a consequence, a

matrix of relative weights and a matrix of average bias for previous 60 days are obtained in order to calculate the coefficients of the MOS regression for all days analyzed in the year (e.g. training data for MOS regression of day 61 is from 1 to 60 days, training data for MOS regression of day 62 is from 2 to 61 days, etc.).

On the other hand, the implementation of the Kalman filter is performed in two steps: Optimization of error ratio (eq. 5) and adaptation of the Kalman algorithm to dataset.

The optimal error ratio between variances of previous error and forecast error for GHI data is a testing of 50 error ratios. The range from 0.01 to 5 has been selected for all ground stations considered over the entire year 2015. We consider that because the solar irradiance dynamic is related to a seasonal cycle, it is important to take into account the variations of the ratio over the seasons. Therefore, we calculate bias and RMSE values over the four seasons to gauge the impact of different error ratio values on the model performance (figure 4). The optimal error ratio minimizes the bias (figure 4a) and minimizes the RMSE (figure 4b) for all seasons. Hence, we use an optimal value varying seasonally for all ground stations. Table 5 summarizes the optimal ratio values calculated for this work (e.g. spring with optimal ratio of 0.41 reduces bias from 71 to 10 W m^{-2} and RMSE from 206 to 182 W m^{-2}).

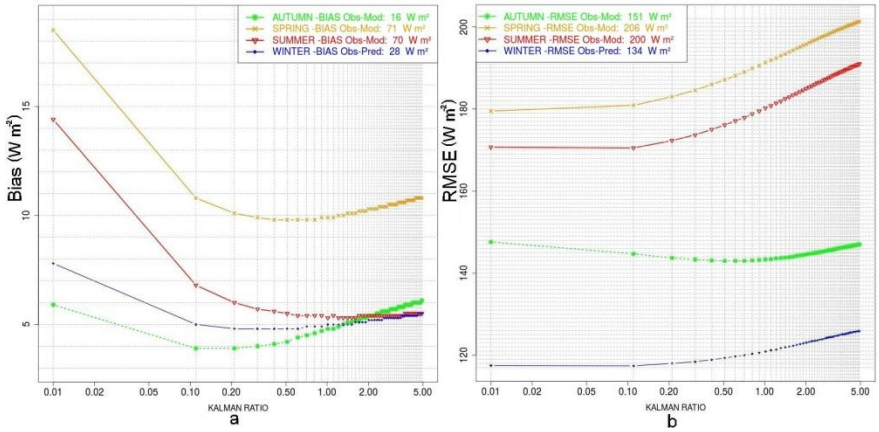

Fig. 4: Seasonal ratio sensitivity for the hourly GHI over 14 ground stations of Paraguay. a) Bias; b) RMSE. Values are computed with the ratio ranging from 0.01 to 5, plotted on logarithmic scale. Perfect bias and RMSE would be 0 W m^{-2}. On the plots bias and RMSE are reported before the application of the Kalman Filter.

Tab. 5: Error ratio values calculated regarding to bias and RMSE for the seasons of the year 2015.

Season	Error ratio
Autumn	0.21
Spring	0.41
Summer	0.41
Winter	0.21

The adaptation of the Kalman algorithm to data is done through a training period, which allows the adjustment of parameters using previous simulations and observations. Tests on Paraguay ground stations for a period between 6 and 20 days establish a training period equal to 15 days. Therefore, adjusting the Kalman algorithm is performed for each day of the year using the 15 days previous as training period.

Finally, the systematic error estimation ($Bias_C$) of MOS regression and Kalman algorithm allow to correct the model simulation (GHI_{wrf}) and thereby obtain the GHI_C correction:

$$GHI_C = GHI_{wrf} - Bias_C \qquad \text{(eq. 6)}$$

5. Results

5.1. Application of post-processing

In order to better understand the impact of reduction of systematic errors for the post-processes, figure 5 shows the results for several combinations of post-process adjustments for the year and the four seasons with respect to the raw model (black bars): 1) Only fit of MOS regression (blue bars), 2) Only fit of Kalman algorithm (red bars), 3) Fit of mean between MOS and Kalman corrections (orange bars) and 4) Fit of Kalman for corrections

produced by MOS (green bars). Figure 5a show the significant relative bias-removal (rMBE) of all combinations for the year, where the best reduction is the fit of Kalman over outputs of MOS with -0.7% from 21% of the raw model. The same configuration shows the best bias-removal for summer and spring from 27% and 29% to -0.4% and 0.3%, respectively. Furthermore, figure 5b shows the annual decrease of rRMSE with less marked differences between raw model and post-processing combinations. The fit of mean between MOS and Kalman (69%) and Fit of Kalman to MOS (70%) show the largest corrections from the raw model (81%). Unlike figure 5a, where the combination of post-processes presents better results in bias-removal, the figure 5b shows less removal of rRMSE for the periods of time analyzed. This lower reduction of rRMSE could be associated with stochastic errors related to numerical random.

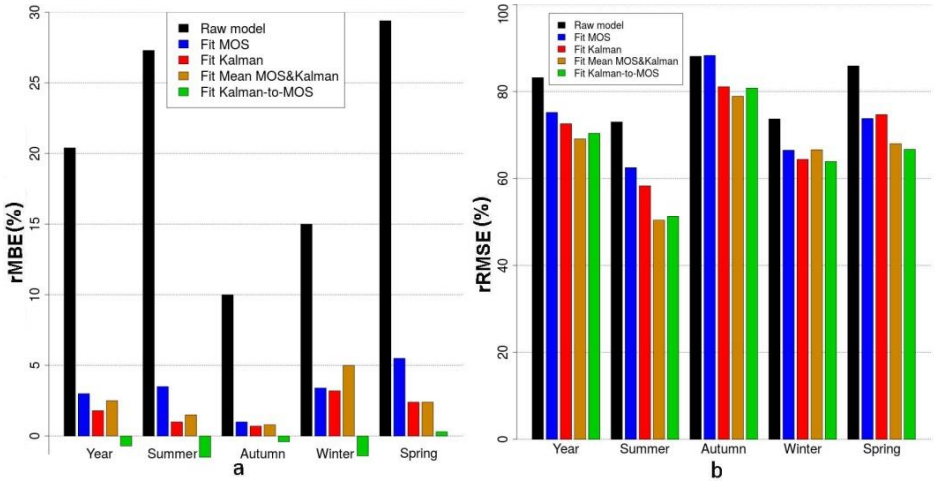

Fig. 5: Comparison of percentage of removal of a) rMBE and b) rRMSE from the raw model (black bars), fit MOS (blue bars), fit Kalman (red bars), fit mean of bias-removal of MOS and Kalman (orange bars) and fit Kalman over MOS (green bars) for year and seasons periods of Paraguay for 2015.

Figure 6 shows the annual bias correction of four combinations of post-process in function of $\cos(SZA)$ and k_t^*. The bias correction is applied to zenith angles less than 75º (right of the dotted line with $\cos(SZA) > 0.26$), which represents the Sun's position vector 15º from the horizon until the zenith. It analyzes the changing shades of dark red (overestimation) and dark blue (underestimation) of model (figure 3) to light colors shown in figure 6 for the post-processes. This change in tone towards yellow color indicates bias correction for cloudy and clear-sky conditions ($k_t^* > 0.4$), where the Kalman-to-MOS post-process (figure 6d) presents the best results of correction with an annual bias of -2 W m^{-2}, whereas single Kalman correction (figure 6b) and mean MOS and Kalman (figure 6c) show problems in correcting underestimation for cloudy and clear-sky days for zenith angles between 60º to 75º. Therefore, we have chosen the Kalman-to-MOS as post-process configuration for bias-removal of GHI simulated by WRF-ARW model.

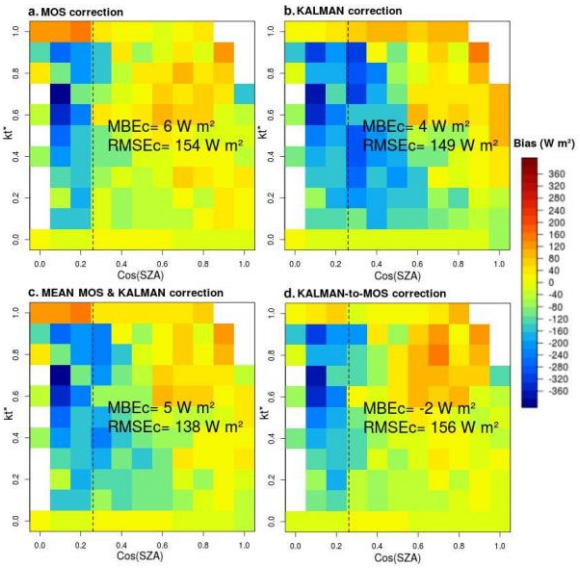

Fig. 6: Bias correction of WRF-ARW model in function of cosine of solar zenith angle ($\cos(SZA)$) and index of atmospheric transparency(k_t^*): a) MOS correction, b) Kalman correction, c) Mean of MOS and Kalman correction and d) Kalman-to-MOS correction.

5.2. Improvement of GHI corrected by Kalman-to-MOS post-process

The results of annual bias correction of Kalman-to-MOS post-process are presented in the figure 7. On the left side, figure 7a shows the results of the hourly comparison between the observed and corrected GHI, where it observed a significant remove of overestimation between 0-1000 W m^{-2} of the raw model observed in figure 2a. The median and quartile ranges show a reduction of systematic error for the entire measuring range.

On the right side, figure 7b presents a box diagram with hourly comparison of GHI corrected (light blue rectangles) in relation to GHI modelled (blue rectangles) and GHI observed (gray rectangles). The results allow analysis of low variability of the hourly distribution for the median, quartile 25% and quartile 75% with a difference of less than 50 W m^{-2} between corrections and observations from 10:00 to 20:00h UTC. Significant reduction of overestimation of model within the hours of highest incidence of solar irradiance in the middle of the day is highlighted.

The statistical results of the annual and seasonal corrections are summarized in table 6. The comparison of bias with respect to the raw model (table 2) provides a significant improvement for summer from 68 to -1.1 W m^{-2} (27 to -1.5%) and spring from 60 to 0.8 W m^{-2} (29 to 0.3%), which present the seasons with most systematic model errors. Regarding to reduction of RMSE, the differences are less significant for summer from 207 to 169 W m^{-2} (71 to 61%) and spring from 198 to 175 W m^{-2} (75 to 68%), demonstrating a persistence of non-deterministic errors.

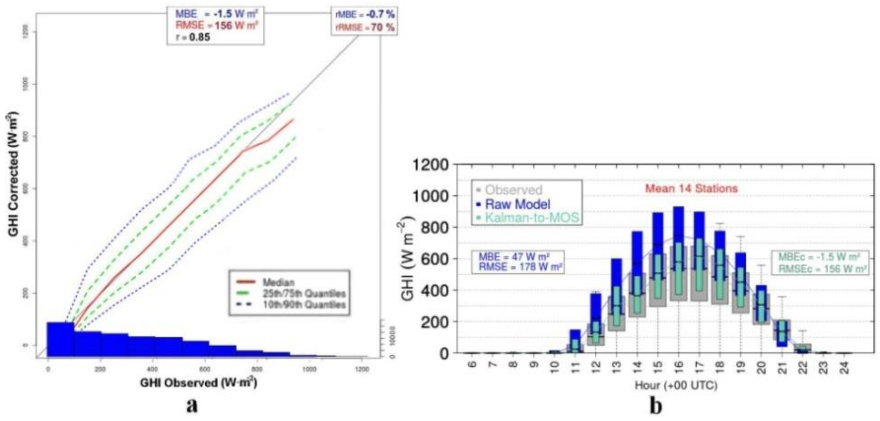

Fig. 7: Comparison of GHI observed and GHI corrected for 14 ground stations in Paraguay for 2015. a) Annual trend of hourly data for median 50% (continuous red line), range of quartile 75% (green dotted lines), range of quartile 90% (blue dotted lines) and frequency distribution histogram (blue bars); b) Daily cycle distribution of GHI observed (gray bars), GHI modelled (blue bars) and GHI corrected by Kalman-to-MOS (light blue bars).

Tab. 6: Results of Kalman-to-MOS bias correction for 14 ground stations in Paraguay for 2015.

Season	Bias (W m^{-2})	rMBE (%)	RMSE (W m^{-2})	rRMSE (%)	r
Summer	-1.1	-1.5	169	61	0.85
Autumn	-0.8	-0.4	147	81	0.82
Winter	-1.8	-1.4	126	64	0.89
Spring	0.8	0.3	175	68	0.85
Year	-1.5	-0.7	156	70	0.85

Therefore, the seasonal overestimation of the raw model for the months of higher incidence of solar irradiance in the year is corrected by Kalman-to-MOS post-process.

6. Conclusions and future work

This contribution presents the assessment and improvement of the simulation of global irradiance based on post-processing techniques applied on outputs of WRF-ARW model. The analysis is performed on a full year meteorological simulation of 2015 for the Paraguay domain. The evaluation focuses on the capability of MOS regression and Kalman filter algorithm to improve the forecast of solar irradiance, estimating and comparing their uncertainty.

The evaluation of the simulation of the GHI establishes a systematic overestimation throughout the year 2015.

Such overestimation is due to the mainly positive bias and RMSE values for spring and summer, which are presented in wide angle zenith ranges to cloudy and clear skies in the study area. The evaluation shows the weaknesses of the Shortwave Radiation of Dudhia scheme due to incorrect location of clouds and total cloud water, besides the lack of precision in the empirical adjustment of scattering parameter that defines the average of atmospheric extinction of molecular absorbers.

We apply and evaluate four combinations of the two post-processes on the raw model outputs in order to improve the irradiance predictions reducing the systematic error. The fit of Kalman-to-MOS provides better results than the others combinations, reducing the errors of the raw model up to 97% of annual bias and 13% of annual RMSE. The aim of the Kalman-to-MOS combination is to apply an initial correction of the systematic error using MOS and subsequently utilize Kalman in order to avoid sudden changes of the GHI forecast that only Kalman fails to capture. Thereby, a Kalman-to-MOS post-process can predict a large bias when all the biases for the past few days have been small, especially for spring and summer.

Among future developments of the work, we highlight the implementation of other post-processes as artificial neural networks and the use of other physics parameterizations that allow reduces stochastic errors related with high RMSE values. Also under consideration is the generation of surface energy maps allowing spatial visualization of the post-processing results, implementation to solar energy production system and validation of the system through real data of photovoltaic and thermal cells production in Paraguay.

7. Acknowledgments

The authors wish to thank Luca Delle Monache (NCAR) and Ronald B. Stull (UBC) for providing the Kalman Filter algorithm used in this study. The annual simulation has been made possible by the implementation of the WRF-ARW model in the Supercomputer cluster hosted by Department of Engineering at the National University of Asuncion, Paraguay. FECOPROD agroclimatic network has provided the observations for this study. Contributions: Rincón A. and Jorba O. applied and assessed the methodology and the presentation of the results. Frutos M. contributed to the treatment, quality control of ground stations and the presentation of the results. González J.A. and Alvarez L. installed and configured the WRF model and related programs as well as contributed to the manuscript and presentation. Barrios F.P. is responsible of control and maintenance of ground stations and data management. González J.A. directs and coordinates the CONACYT project 14-INV-289 in which this work is framed.

8. References

Dudhia, J., 1989. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. J. Atmos. Sci., 46 (20):3077–3107, doi:10.1175/1520–0469.

Duffie, J. and Beckman, W. 1991. Solar engineering of thermal process. Interscience, New York (EE.UU).

Delle, M. L., Nipen, T., Deng, X., Zhou, Y., Stull, R. 2006. Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction. J Geophys Res 2006;111:D05308. doi:10.1029/2005JD006311.

Fairall, C.W., Bradley, E.F., Hare, J.E., Grachev, A.A., Edson, J.B., 2003. Bulk parameterization of air-sea fluxes: Updates and verification for the COARE algorithm. Journal of climate, 16(4), 571-591.

Iqbal, M. 1983. An introduction to solar radiation. Academic Press, Orlando, FL.

Kain, J.S., 2004. The Kain-Fritsch convective parameterization: An update. J. Appl. Meteor., 43, 170–181.

Kalman, R., 1960. A new approach to linear filtering and prediction problems, Journal of Basic Engineering, vol. 82, pp 35 – 45.

Kimura, R., 2002. Numerical weather prediction. J Wind Eng Ind Aerodyn, 90: 1406–1414.

Lara-Fanego, V., Ruiz-Arias, J., Pozo-Vazquez, D., Santos-Alamillos, F., Tovar-Pescador, J., 2011. Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain). Solar Energy, 86(8), 2200-2217.

Liang, H., Wang, Y., Qian, H., 2007. Verification and comparative analysis of prediction of ECMWF model and T213 model in summer. Sci Meteorol Sin, 27: 253–258

Lorenz, E., Hurka, J., Heinemann, D. Beyer, H., 2009. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE Journal of selected topics in applied earth observations and remote

sensing, 2(1).

Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J., Clough, S.A., 1997. Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. J. Geophys. Res., 102 (D14), 16663–16682.

National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, 1994. NCEP/NCAR Global Reanalysis Products, 1948-continuing. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO. [Available online at http://rda.ucar.edu/datasets/ds090.0/.] Accessed february 2016.

Heinemann, D., Lorenz, E., Girodo, M., 2006. Solar irradiance forecasting for the management of solar energy systems. Energy and Semiconductor Research Laboratory, Energy Meteorology Group, Oldenburg University.

Hong, S.Y., Dudhia, J., Chen, S.H., 2004. A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation, Mon. Wea. Rev., 132, 103–120.

Hong, S.Y., Noh, Y., Dudhia, J., 2006. A new vertical diffusion package with an explicit treatment of entrainment processes. Mon. Wea. Rev., 134, 2318–2341.

Mathiesen, P., Kleissl, J., 2011. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. Solar Energy. 85(5), 967-977.

McCollor, D., Stull, R., 2008. Hydrometeorological accuracy enhancement via post-processing of numerical weather forecasts in complex terrain. Wea. Forecasting, 23, 131–144.

Monache, L.D., Nipen, T., Liu, Y., 2011. Kalman filter and analog schemes to postprocess numerical weather predictions. Mon Weather Rev, 11: 3554–3570.

Niu, G.Y., Yang, Z.L., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., Xia, Y., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. Journal of Geophysical Research: Atmospheres, 116 (D12).

Pelland, S., Galanis, G., Kallos, G., 2011. Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale Numerical Weather Prediction model. Progress in Photovoltaics: Research and Applications, 21(3), 284-296.

Rincon, A., Jorba, O., Delle Monache, L., Baldasano, J.M., 2013. Solar irradiance forecast system based on post-processing techniques applied on WRF-ARW meteorological simulations in Spain. 2nd International Conference Energy & Meteorology (ICEM). Meteo-France International Conference Centre. Toulouse, France. 25 - 28 June 2013.

Roeger, C., Stull, R., McClung, D., Hacker, J., Deng, X., Modzelewski, H. 2003. Verification of Mesoscale Numerical Weather Forecasts in Mountainous Terrain for Application to Avalanche Prediction. Weather and forecasting, 18:1140–1160.

Roesch A., Wild, M., Ohmura, A., Dutton, E.G, Long, C.N., Zhang, T., 2011. Assessment of BSRN radiation records for the computation of monthly means. Atmos. Meas. Tech., 4, 339–354.

Ruiz-Arias, J., Pozo-Vázquez, D., Sánchez-Sánchez, N., Montávez, J., Hayas-Barrú, A., Tovar-Pescador, J., 2008. Evaluation of two MM5-PBL parameterizations for solar radiation and temperature estimation in the South-Eastern area of the Iberian Peninsula. Il Nuovo Cimento C, 31:825–842.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., Huang, X., Wang, W., Powers, J., 2008. A Description of the Advanced Research WRF Version 3, NCAR/TN-475+STR. Mesoscale and Microscale Meteorology Division, National Centre for Atmospheric Research, Boulder, USA.

Spencer, J. 1971. Fourier series representation of the position of the Sun. Search, 2(5):172.

Zamora, R.J., Dutton, E.G., Trainer, M., McKeen, S.A., Wilczak, J.M., and Hou, Y.-T. 2005. The accuracy of solar irradiance calculations used in mesoscale numerical weather prediction, Mon. Weather Rev., 133, 783–92.

Zhang, L., Wang, S., He, C., Shang, K., Meng, L., Li, X., Lofgren, B. M., 2015. A new method for instant correction of numerical weather prediction products in China. Science China Earth Sciences, 58(2), 231-244.