Analysis of ARMA Solar Forecasting Models Using Ground Measurements and Satellite Images

Franco Marchesoni-Acland, Philippe Lauret, Alvaro Gómez, Rodrigo Alonso-Suárez

Analysis of ARMA Solar Forecasting Models Using Ground Measurements and Satellite Images

Franco Marchesoni-Acland[a,c], Philippe Lauret[b], Alvaro Gómez[c] and Rodrigo Alonso-Suárez[a]

[a] Laboratorio de Energía Solar, Universidad de la República (Udelar), Uruguay
[b] Université de La Réunion, 15 avenue René Cassin, 97715 Saint-Denis, La Réunion, France
[c] Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de la República, Uruguay

***Abstract*—As the solar photovoltaic (PV) share in the electricity grid is growing year by year, solar irradiance forecasting is becoming increasingly important. In this work the performance of a recursive formulation of ARMA models suitable for operational context using the Pampa Húmeda region as a case study is analyzed. Results are promising, as this simple adaptive algorithm does not require historical data and outperform persistence at all lead times. The improvement produced by adding satellite cloudiness data and short-term local variability as exogenous inputs is also evaluated. It is found that the spatially averaged satellite albedo is a useful input variable, improving the forecast performance, while the introduction of short-term variability produce negligible performance changes under this kind of models.**

***Index Terms*—Forecasting, solar irradiance, adaptive filters, satellite images.**

I. Introduction

Solar energy is being increasingly used as a means of producing electricity [1]. As an example, in 2017, the milestone of having more than 400 GW worldwide installed solar photovoltaic (PV) capacity was achieved. Due to the intermittent character of solar energy, increasing the solar energy share into the electricity grid pose a challenge for its stability. Grid utilities need accurate forecast at different granularities and time horizons for an efficient electricity dispatch [2]. Predicting the solar irradiance at ground level is the first step for solar PV power forecasting.

There are four main techniques for addressing solar energy forecast [3]–[5]: Numerical Weather Prediction (NWP) models [6]–[8], satellite forecast by means of Cloud Motion Field (CMF) estimation [9], all-sky camera forecast that track clouds in the very short term [10] and statistical and machine learning tools, commonly referred as time series forecast [11]–[14]. In this context, several authors have proposed forecasting models, such as Neural Networks (NN) [15]–[19] or different kinds of Auto-Regressive Moving Average (ARMA) models [20]–[23], being the latter the focus of this work.

ARMA models are a tool to describe stationary processes originally proposed by Whittle [24] and popularized by Box and Jenkins [25]. These models are widely used in economics and other fields where time series analysis and modeling are an unavoidable task. General ARMA models can provide multivariate predictions by modeling each observation as a vector or an univariate prediction for an unique target variable. The prediction may have also extra inputs variables, known as exogenous variables. In order to produce a forecast, the ARMA models linearly combines past observations and residuals of both endogenous and exogenous variables. Despite the strong hypothesis that the formal theory holds behind, such as stationarity and linear relations between the variables, they proved to be a good forecasting technique in several fields [26]–[28]. Solar irradiation forecasting have not been an exception. ARMA models have been used for Global Horizontal Irradiance (GHI) prediction [3], [12], [20] and consistently proved to obtain reduced estimation uncertainty when compared with persistence and climatology forecasting. They also proved to be competitive versus NN approaches [29], [30]. An ARMA model can be formulated in many different ways: as a fixed digital filter, as a Kalman filter, as an adaptive filter or in a recursive manner.

A comprehensive study of the forecast errors when using an ARMA model alongside a Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is presented in [20]. Only the solar radiation time-series is used as input variable. The ARMA models inspected are framed in the Recursive Least Squares (RLS) technique, which is the same that we use in this work. A specific model order was selected for each place and lead time, and the uncertainty of this approach for each location is reported. In [31] ARMA models framed in a train-test approach are compared to other models. In particular, ARMA models are compared to a NN that includes solar satellite estimates retrieved from the Helioclim-3 database (version HC3v5). The solar estimates available in this platform are based on the Heliosat-2 model [32] and images from the MSG (Meteosat Second Generation) geostationary satellite administrated by ESA (European Space Administration). The advantage of including solar satellite estimates becomes explicit. This is also shown in a similar but previous work over Gran Canaria Island [33].

In this work, the introduction of two exogenous variables in ARMA models in a RLS framework for forecasting ground level solar irradiation at an horizontal plane (GHI), using the Pampa Húmeda region (south-east of South America) as a case study, is explored. In particular, the effect of adding a short-term variability index, calculated from recent past ground measurements values, and satellite cloudiness information, obtained from the GOES-East satellite administrated by NOAA (National Oceanic and Atmospheric Administration), is tested. In this case, satellite cloudiness is included directly without being processed by a solar resource assessment model. It is found that including cloudiness satellite information significantly improves the performance of the ARMA-RLS forecast, while short-term local variability has a negligible influence. A discussion regarding the order selection of the ARMA filters for GHI forecasting in the region is provided. It is shown that a fixed order can be used for all lead times without any relevant performance loss in comparison to an optimal filter. This reduces the problem to an appropriate choice of exogenous inputs. These observations may be extrapolated to others regions or particular sites, an analysis that, to the best of our knowledge, is not found in the literature. Finally, the work also seeks to provide a first in-detail assessment of the forecasting performance of these kind of methods for the region.

The work is organized as follows: in Section II the data is presented by describing the equipment, the stations' characteristics and the exogenous variables being used. In Section III the ARMAX model and the RLS algorithm are presented, with a description of the advantages of this approach. In Section IV the forecast performance metrics to be used in evaluation are introduced. A discussion of the order selection for the ARMA models based only on past ground measurements follows in Section V. In Section VI the results including the exogenous variables in the ARMAX model are exhibited. Finally, the work is concluded in Section VII. The code used for this work is available at https://github.com/franchesoni/ARMAXrls-solar-irradiance-forecasting.

II. Data

GHI data recorded in six measuring stations in Uruguay is used. These stations are representative of the subtropical climate of the south-east part of South America, known as Pampa Húmeda. The sites' locations are presented in Table I. One of these sites (LE) is located at the Solar Energy Laboratory (LES, http://les.edu.uy) experimental facility, where a Kipp & Zonen Solys2 ground station records the three main irradiance components (global, direct and diffuse). The GHI in this station is recorded using a Kipp & Zonen CMP10 Secondary Standard. The other four stations are located on field in semi rural environments. They are equipped with First Class or better pyranometers for the GHI measurement and a data logger with remote communication. These pyranometers are calibrated every two years against a Kipp & Zonen CMP22 Secondary Standard that is kept with traceability to the World Radiometric Reference in the World Radiation Center, Davos, Switzerland. At least two years data at a 10 minutes granularity from the period 2015–2017 is considered for each station.

In this work, the effect of adding short-term local variability and satellite information to a RLS auto-regressive point forecast is inspected. The variability index (σ) is computed from the past ground measurements as the standard deviation of the last five clear-sky index changes,

Table I
LOCATION OF THE GROUND MEASUREMENT STATIONS.

station name	**station code**	**lat.** (deg)	**lon.** (deg)	**alt.** (m)	$\overline{\text{GHI}}$ (W/m^2)
LES facility	LE	-31.28	-57.92	56	460
Artigas	AR	-30.40	-56.51	136	451
Las Brujas	LB	-34.67	-56.34	38	440
Tacuarembó	TA	-31.71	-55.83	142	444
Rocha	RO	-34.49	-54.31	20	428
La Estanzuela	ZU	-34.34	-57.69	70	442
	all sites measurements average				444

$$\sigma(t) = \sqrt{\frac{1}{M} \sum_{i=0}^{M-1} \left(\Delta k_c(t-i) - \overline{\Delta k_c}\right)^2}, \tag{1}$$

with $M = 5$, where k_c is the ratio between the GHI measurements and the clear sky GHI as estimated by a clear sky model, $\Delta k_c(t) = k_c(t) - k_c(t-1)$ are the changes in this variable and $\overline{\Delta k_c}$ is the average of the M considered changes. The McClear model [34], publicly available at the SoDa platform (http://www.soda-pro.com), is used for calculating the clear-sky index from the GHI time series. In order to avoid variability issues related to the interface between two consecutive days, the variability is set to zero at the beginning of a day and only current day samples are used for its calculation.

Satellite images are also required. GOES-East's satellite Earth Albedo (ρ_p) averaged in a 10 min $\times$ 10 min latitude-longitude cell is used. The spatial average of the satellite information is intended to include near future information of the cloudiness, as it is not only taking into account the cloudiness over the specific site, but also the surroundings. The GOES13 satellite operated in the GOES-East position during the period 2015-2017, providing an irregular time-slot acquisition for South America. Satellite images for this region are usually available at a rate of two images per hour, but periods with hourly or tri-hourly images may occur. The 10-minutes time resolution for the satellite albedo was obtained via a linear interpolation. Satellite gaps of more than two consecutive hours are not interpolated and are removed from the data set.

III. ARMAX filtering

Auto-Regressive (AR) and Moving-Average (MA) models with Exogenous Variables (ARMAX) describe a process as a linear combination of past measurements (X_{t-i}), past errors (ϵ_{t-i}) and exogenous variables ($\boldsymbol{E_t}$, a vector which may include past values). If p and q are the orders of the AR and the MA terms respectively, we have,

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{i=1}^{q} \beta_i \epsilon_{t-i} + \boldsymbol{\gamma^T E_t} + \epsilon_t, \tag{2}$$

where ϵ_t is a deviation from the model at time t, assumed to be white Gaussian noise. For later forecasting use, the set of parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ that best fit the data need to be found. For this the RLS algorithm is used, explained in the following. The predicted variable is the clear-sky index ($X_t = k_c(t)$) which is then converted to a GHI point forecast using the McClear clear-sky model. The signals $\sigma(t)$ and $\rho_p(t)$ are the variables inspected for $\boldsymbol{E_t}$.

Recursive Least Squares (RLS) is an optimization algorithm that solves recursively the minimization of a cost function depending on the weights w_n (or $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$, in this case),

$$C(\boldsymbol{w_n}) = \sum_{i=0}^{n} \lambda^{n-i} e^2(i), \tag{3}$$

where $e(i)$ is the forecasting error of observation i. Note that if the lead time is h, then $e(i) = \boldsymbol{w_n^T} \boldsymbol{z_{i-h}} - X_i$, being $\boldsymbol{z_{i-h}}$ a vector including all input variables. The factor λ is called forgetting factor and when it is near 1, it resembles Least Squares Minimization while allowing the weights to adapt to the statistical changes of the k_c time series. A sketch of the RLS algorithm is provided in Algorithm 1, where a big initialization parameter b denotes uncertainty in the first estimate.

Algorithm 1 Recursive Least Squares Algorithm

h : forecast horizon
p : AR order
q : MA order
X_i : measurement at time i
$\boldsymbol{E_i}$: exogenous variables at time i
$b = 1000$: initialization parameter
$\lambda = 0.999$: forgetting factor

Initialization
$\boldsymbol{\hat{w}_0} \leftarrow \boldsymbol{0}$
$\hat{X}_0 \leftarrow 0$
$P_0 \leftarrow b\boldsymbol{I}$

When new data arrives
$\boldsymbol{X_n} \leftarrow (X_n, X_{n-1}, ..., X_{n-p+1})^T$
$\boldsymbol{e_n} \leftarrow \boldsymbol{\hat{X}_n} - \boldsymbol{X_n}$
$\boldsymbol{z_n} \leftarrow \left(\boldsymbol{X_n^T}, \boldsymbol{e_n^T}, \boldsymbol{E_n^T}\right)^T$
$K_n \leftarrow \frac{P_{n-1}\boldsymbol{z_{n-h}}}{\lambda + \boldsymbol{z_{n-h}^T} P_{n-1} \boldsymbol{z_{n-h}}}$
$P_n \leftarrow \frac{1}{\lambda}\left(P_{n-1} - K_n \boldsymbol{z_{n-h}^T} P_{n-1}\right)$
$\boldsymbol{\hat{w}_n} \leftarrow \boldsymbol{\hat{w}_{n-1}} + K_n \left(X_n - \boldsymbol{z_{n-h}^T} \boldsymbol{\hat{w}_{n-1}}\right)$
$\boldsymbol{\hat{X}_{n+h}} \leftarrow \boldsymbol{z_n^T} \boldsymbol{\hat{w}_n}$

As historical data is not necessary, this adaptive approach is useful for operational context. It also avoids fixing the weights (and the optimization challenge this brings). Furthermore, statistical properties varies between seasons and even days (i.e. cloudy and clear-sky days), causing short-term adaptability to be a desirable property. A deduction of RLS algorithm for an arbitrary lead time is provided in the Appendix A, which is not easily found in the bibliography.

IV. Performance metrics and evaluation

Results are presented in terms of the standard Mean Bias Deviation (MBD) and Root Mean Squared Deviation (RMSD) metrics, as well as the Forecasting Skill (FS) metric [3]. The MBD and RMSD definitions are,

$$\text{MBD}_h = \frac{1}{N} \sum_i \left[\hat{y}_h(i) - y^{\text{ref}}(i+h)\right], \tag{4}$$

$$\text{RMSD}_h = \sqrt{\frac{1}{N} \sum_i \left[\hat{y}_h(i) - y^{\text{ref}}(i+h)\right]^2}, \tag{5}$$

where $\hat{y}_h$ is the GHI forecast, y^{ref} is GHI measurement's and h is the forecast horizon. Their relative values, rMBD and rRMSD respectively, are expressed as percentage of the average irradiance value (see last column of Table I). The MBD definition is such that a positive value means a forecasting overestimation and a negative value means a forecasting subestimation. The forecasting skill represents the gain of the forecasting RMSD with respect to the persistence procedure, and it is defined as,

$$\text{FS} = 1 - \frac{\text{RMSD}_\text{m}}{\text{RMSD}_\text{p}}, \tag{6}$$

where the subscripts 'm' and 'p' refers to the model and persistence respectively. The persistence is calculated by setting $\hat{k}_c(t+h) = k_c(t)$, for every $h \geq 1$, where $\hat{k}_c(t+h)$ is the forecast of $k_c(t+h)$. As explained before, the corresponding GHI forecast is inferred by using the clear sky model estimates.

The results averaged over all the sites, which act as the mean performance over the region, are presented here. Note that this is not an assessment in which it is intended to predict the joint resource for all sites, known as regional forecast, but rather is the average of the sites performances. The complete results (the performance metrics) for each station, model and lead time can be found in the following link: http://les.edu.uy/pub/ARMAX-RLS.csv. As the filters are adaptive, no train set is needed. Nevertheless, the first 400 forecasts were ignored in order to avoid including convergence and initialization discussions that are not relevant to operational performance.

V. ARMA order selection

The orders for the ARMA model have to be selected (p and q) for each lead time. In this Section, the order selection is done using only the past ground measurements. ARIMA models (models that work with the time series derivative) were also considered, yielding worse results in all the cases, so the differentiation term (known as d) is ignored in the following analysis. Information criteria do not always lead to an acceptable model selection, as was found out in some tests.

There is a qualitative approach based on Auto-Correlation Function (ACF) and Partial ACF plots, but it proved not suitable for working with these solar data sets (although the approach is widely used in economics [35]). However, a grid-search model selection method provided one important insight: if the irradiance time series is long enough, then different orders for $p \geq 1$ yield a difference always smaller than 1 % in terms of rRMSD for our data set, as observed in Fig. 1 (orders for $p \leq 10$ are plotted). In Fig. 1 the performance of different (p, q) ARMA models is shown as a function of the time horizon and it is observed that the curves with worse performance correspond to those models with $p = 0$. For $p > 0$ small performance changes are observed for the different model's orders. Also, the rRMSD span is reduced with the time horizon. The rRMSD span is of $\simeq 0.9\%$ for 10 minutes ahead, $\simeq 0.8\%$ for 90 minutes ahead and less than $\simeq 0.4\%$ for 240 minutes ahead. Most of these differences are concentrated between the $p = 1$ models and the others, being the rRMSD surfaces more flat for higher p orders. This implies that, in practice, the selection of the model order (both p and q) is not a crucial issue for $p \geq 2$. In Fig. 2 a simple AR model, with $p = 5$ AR lags (as it was found as a good trade off filter between accuracy and simplicity for all lead times), applied to each lead time and location, is compared to the error found with the optimum (p, q) for each lead time and location. The difference is negligible.

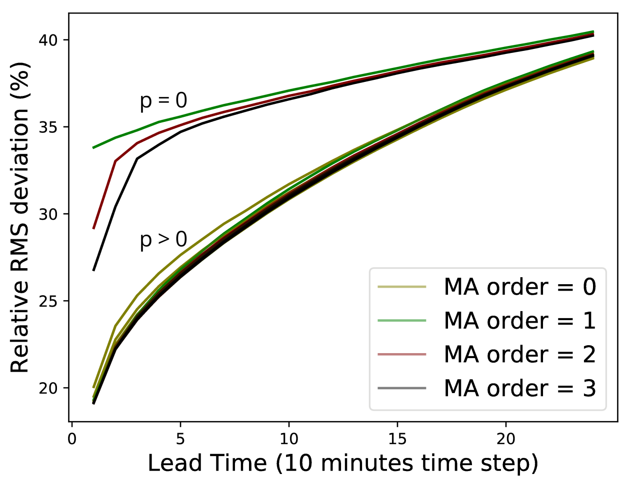

Figure 1. Performance analysis for different (p, q) ARMA models. Curves for $p \leq 10$ are plotted.

VI. Results with exogenous variables

Aside from reasonable order selection, the most important performance changes come from an appropriate choice of exogenous variables. Intuitively, including the local short-term variability of the time series would add information about the current state of the atmosphere; i.e., if the last conditions were clear/cloudy sky or partially cloudy (which will exhibit higher variability). Cloudiness measured via satellite is strongly (negatively) correlated with irradiance measurements. Assuming an ergodic hypothesis, the average

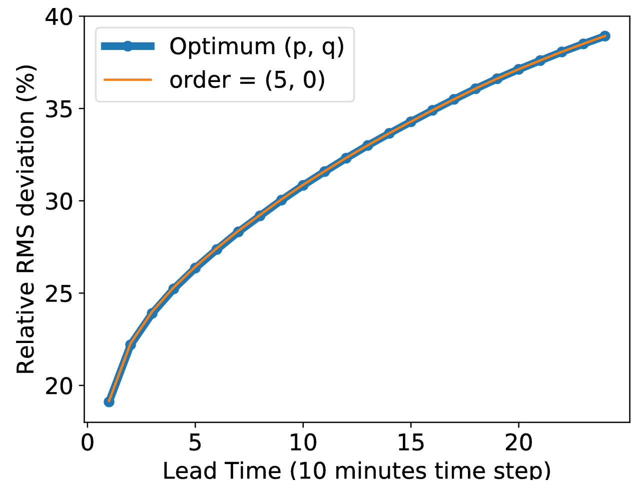

Figure 2. Performance comparison between a (5, 0) model and an optimum model which uses the best performing p and q values for each time horizon and location.

in space of the satellite cloudiness includes information of the near future irradiation over the site. Given the above, we implement and compare the following models:

1) Model PGM: is an ARMAX-RLS model that only uses Past Ground Measurements (PGM).
2) Model PGM-S: is the PGM model including current time satellite albedo (ρ_p) as exogenous variable.
3) Model PGM-V: is the PGM model including local short-term variability (σ) as exogenous variable.
4) Model PGM-SV: is the PGM model including both exogenous variables (satellite and variability).

The results were obtained for the PGM, PGM-S, PGM-V and PGM-SV models using a $(5, 0)$ order for the ground measurements part. Exogenous variables are included at time t, without neither past values or past errors. Table II presents the average performance over the region for each model and the persistence as a function of the time horizon. It shall be noted that all the models present similar and small bias (rMBD) and outperform the persistence procedure in terms of RMSD, yielding positive forecasting skills. The trend of the RMSD against the forecast horizon is illustrated in Fig. 3.

It is clear that the introduction of the spatially averaged satellite albedo as an exogenous variable improves the forecast performance. For instance, the PGM model achieves a minimum forecasting skill of FS $= 7.8$ % ($h = 10$ minutes) and a maximum forecasting skill of FS $= 14.8$ % ($h = 240$ minutes) while the PGM-S model obtains a FS $= 16.3$ % and FS $= 18.2$ % for the same time horizons, respectively. Also, the performance of the PGM-S model peaks at FS $=$ 24.1 % (for $h = 30$ minutes). The characterization of the impact of the spatial averaging cell size on the forecasting performance will be part of future work.

On the other hand, including short-term variability has

Table II
PERFORMANCE METRICS FOR THE ARMAX-RLS ALGORITHM AVERAGED OVER THE FIVE STATIONS.

	Persistence		PGM			PGM-S			PGM-V			PGM-SV		
time	MBD	RMSD	MBD	RMSD	FS	MBD	RMSD	FS	MBD	RMSD	FS	MBD	RMSD	FS
horizon	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
10 minutes	0.1	20.8	0.2	19.2	7.8	0.3	17.3	16.6	0.2	19.2	7.7	0.3	17.3	16.5
20 minutes	0.1	24.7	0.4	22.3	9.9	0.5	19.1	22.9	0.4	22.3	9.8	0.5	19.1	22.8
30 minutes	0.2	26.7	0.5	24.0	10.2	0.6	20.3	24.1	0.5	24.0	10.1	0.6	20.3	24.0
40 minutes	0.2	28.2	0.6	25.3	10.2	0.7	21.6	23.6	0.6	25.3	10.1	0.7	21.6	23.5
50 minutes	0.3	29.4	0.7	26.4	10.2	0.7	22.8	22.7	0.7	26.5	10.1	0.8	22.8	22.6
60 minutes	0.4	30.5	0.8	27.4	10.1	0.8	23.9	21.7	0.8	27.5	10.0	0.8	23.9	21.6
120 minutes	0.8	36.2	1.2	32.3	10.6	0.9	29.4	18.7	1.2	32.4	10.5	0.9	29.4	18.6
180 minutes	1.2	41.2	1.5	36.1	12.4	0.9	33.7	18.3	1.5	36.1	12.3	0.9	33.7	18.2
240 minutes	1.4	45.7	1.6	38.9	14.8	1.0	37.1	18.8	1.6	39.0	14.7	1.1	37.1	18.8

almost no effect in the model performance, as also shown in Fig. 3. Thus, variability is not a suitable exogenous variable to introduce in the ARMAX-RLS models: linearly combining this variable does not improve the forecast performance. This does not mean that this variable it is not suitable for any model, as it may be useful for others machine learning approaches for point or probabilistic forecast.

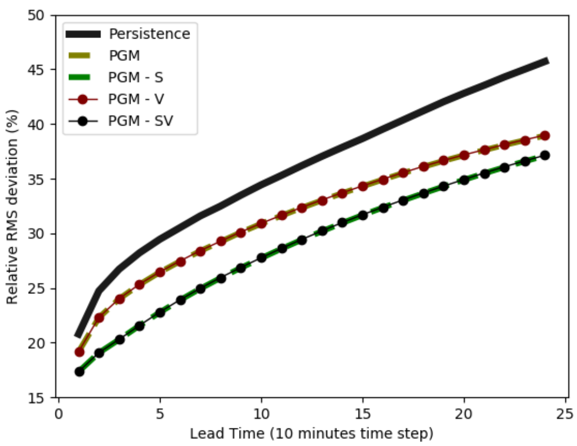

Figure 3. rRMSD metric: average performance over all Uruguayan sites.

VII. CONCLUSIONS

The results obtained for GHI forecasting in the Pampa Húmeda region (south east part of South America) using an ARMAX Recursive Least Squares technique with past ground measurements and two different exogenous variables (satellite albedo and a local short-term variability index) as input were presented in this work. We found that all the models tested, whether they include or not exogenous variables, outperform the classical persistence procedure for the region. A discussion for the ARMA order selection (p and q) was provided, and it was found that for $p \geq 2$ all the orders yield to similar performance. Also, it was observed that a fixed filter for all sites and time horizons can be used in the region without sacrificing performance. The importance of including satellite data as predictor is revealed, as it greatly improves the forecasting performance. However, variability information was found to be irrelevant to these kind of models, since negligible changes were found in the model's performance due to its introduction. The recommended model for the region from the ones inspected here is the PGM-S model.

The results presented here comprise a solely homogeneous subtropical climatic area and should not be incautiously extrapolated to other sites. In future research these ideas will continue to be explored over other parts of the world, covering other climates with different solar resource variability. Also, as the site's satellite information prove to be an improvement, adding the satellite cloudiness in the surroundings of the site is a promising future work for improving solar forecast. Combining information and predictors correctly is needed to improve solar forecasting accuracy and recursive ARMA models can serve as a good framework for this objective.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support given by Uruguay's National Research and Innovation Agency (ANII) under the FSE-ANII-2016-131799 grant.

APPENDIX A
DEDUCTION OF RECURSIVE LEAST SQUARES ALGORITHM FOR ARBITRARY LEAD TIME

Remember the cost function to minimize Eq. (3),

$$C(\boldsymbol{w_n}) = \sum_{i=0}^{n} \lambda^{n-i} e^2(i), \tag{2}$$

being $e(i)$ the forecasting error of observation i. Note that if the lead time is h, then $e(i) = \boldsymbol{w_n^T} \boldsymbol{z_{i-h}} - X_i$, being $\boldsymbol{z_{i-h}}$ a vector including all input variables.

Differentiating and equaling to zero:

$$\frac{\mathrm{d}C(w_n)}{\mathrm{d}w_n} = \sum_{i=0}^{n} 2\lambda^{n-i} e(i) \frac{\mathrm{d}e(i)}{\mathrm{d}w_n} = \sum_{i=0} n 2\lambda^{n-i} e(i) \boldsymbol{z_{i-h}} = 0$$

$$\Rightarrow \sum_{i=0}^{n} \lambda^{n-i}(\boldsymbol{w_n^T z_{i-h}} - X_i)\boldsymbol{z_{i-h}} = 0$$

$$\Rightarrow \sum_{i=0}^{n} \lambda^{n-i}(\boldsymbol{w_n^T z_{i-h}})\boldsymbol{z_{i-h}} = \sum_{i=0}^{n} \lambda^{n-i} X_i \boldsymbol{z_{i-h}}$$

that can be shown equivalent to:

$$\left(\underbrace{\sum_{i=0}^{n} \lambda^{n-i}\boldsymbol{z_{i-h} z_{i-h}^T}}_{\Theta(n)}\right)\boldsymbol{w_n} = \underbrace{\sum_{i=0}^{n} \lambda^{n-i} X_i \boldsymbol{z_{i-h}}}_{r(n)},$$

identifying $\Theta(n)$ the weighted sample covariance matrix for $\boldsymbol{z}$ and being $r(n)$ the estimate of the cross-covariance between X_n and $\boldsymbol{z_{n-h}}$. So we have: $\Theta(n)\boldsymbol{w_n} = r(n)$. Now note that we could write Θ as a recursion:

$$\Theta(n) = \boldsymbol{z_{n-h} z_{n-h}^T} + \lambda \underbrace{\sum_{i=0}^{n-1} \lambda^{n-1-i}\boldsymbol{z_{i-h} z_{i-h}^T}}_{\Theta(n-1)}. \quad (7)$$

Remembering the Woodbury Matrix Identity:

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1}VA^{-1},$$

and identifying right hand side of Eq. (7) as $A+UCV$

$$A^{-1} = \frac{1}{\lambda}\Theta^{-1}(n-1), \quad U = \boldsymbol{z_{n-h}}, \quad C = \boldsymbol{I}, \quad V = \boldsymbol{z_{n-h}^T},$$

we can invert $\Theta(n)$, obtaining,

$$\Theta^{-1}(n) = \left(\lambda\Theta(n-1) + \boldsymbol{z_{n-h} z_{n-h}^T}\right)^{-1} = (A+UCV)^{-1}$$

$$= A^{-1} - A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1}VA^{-1}$$

$$= \frac{1}{\lambda}\left(P_{n-1} - \underbrace{\frac{P_{n-1}\boldsymbol{z_{n-h}}}{\lambda + \boldsymbol{z_{n-h}^T}P_{n-1}\boldsymbol{z_{n-h}}}}_{K_n}\boldsymbol{z_{n-h}^T}P_{n-1}\right),$$

where $P_{n-1} = \Theta^{-1}(n-1)$ and $P_n = \Theta^{-1}(n)$.

So we have:

$$P_n = \frac{1}{\lambda}\left(P_{n-1} - K_n\boldsymbol{z_{n-h}^T}P_{n-1}\right).$$

Playing with K_n:

$$K_n = \frac{\lambda^{-1}P_{n-1}\boldsymbol{z_{n-h}}}{1+\lambda^{-1}\boldsymbol{z_{n-h}^T}P_{n-1}\boldsymbol{z_{n-h}}}$$

$$\Rightarrow K_n\left(1+\lambda^{-1}\boldsymbol{z_{n-h}^T}P_{n-1}\boldsymbol{z_{n-h}}\right) = \lambda^{-1}P_{n-1}\boldsymbol{z_{n-h}}$$

$$\Rightarrow K_n = \lambda^{-1}\left(P_{n-1} - K_n\boldsymbol{z_{n-h}^T}P_{n-1}\right)\boldsymbol{z_{n-h}}$$

$$= P_n\boldsymbol{z_{n-h}}$$

From the definition of $r(n)$:

$$r(n) = \sum_{i=0}^{n}\lambda^{n-i}X_i\boldsymbol{z_{i-h}} = X_n\boldsymbol{z_{n-h}} + \lambda\sum_{i=0}^{n-1}\lambda^{n-1-i}X_i\boldsymbol{z_{i-h}}$$

$$= \lambda r(n-1) + X_n\boldsymbol{z_{n-h}}$$

Recalling $\Theta(n)\boldsymbol{w_n} = r(n)$ and replacing $r(n)$:

$$\boldsymbol{\hat{w}_n} = P_n r(n)$$

$$\Rightarrow \boldsymbol{\hat{w}_n} = \lambda P_n r(n-1) + P_n X_n \boldsymbol{z_{n-h}}$$

$$= P_{n-1}r(n-1) - K_n\boldsymbol{z_{n-h}^T}P_{n-1}r(n-1) + X_nP_n\boldsymbol{z_{n-h}}$$

$$= \boldsymbol{\hat{w}_{n-1}} - K_n\boldsymbol{z_{n-h}^T}\boldsymbol{\hat{w}_{n-1}} + X_nK_n$$

$$= \boldsymbol{\hat{w}_{n-1}} + K_n\left(X_n - \boldsymbol{z_{n-h}^T}\boldsymbol{\hat{w}_{n-1}}\right)$$

That is the coefficient update summarized in Algorithm 1.

References

[1] REN21, "Renewables 2018: Global status report," Renewable Energy Policy Network for the 21st Century, 15 Rue de Milan, 75441 Paris CEDEX 09, France., Tech. Rep., 2018. [Online]. Available: http://www.ren21.net/

[2] P. Mathiesen, J. Kleissl, and C. Collier, "Chapter 14 - case studies of solar forecasting with the weather research and forecasting model at gl-garrad hassan," in *Solar Energy Forecasting and Resource Assessment*, J. Kleissl, Ed. Boston: Academic Press, 2013, pp. 357–382.

[3] C. F. Coimbra, J. Kleissl, and R. Marquez, "Chapter 8 - overview of solar-forecasting methods and a metric for accuracy evaluation," in *Solar Energy Forecasting and Resource Assessment*, J. Kleissl, Ed. Boston: Academic Press, 2013, pp. 171 – 194.

[4] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. M. de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78–111, 2016.

[5] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 65–76, 2013.

[6] R. Perez, K. Moore, S. Wilcox, D. Renné, and A. Zelenka, "Forecasting solar radiation – Preliminary evaluation of an approach based upon the national forecast database," *Solar Energy*, vol. 81, no. 6, pp. 809–812, 2007.

[7] E. Lorenz, J. Remund, S. Müller, W. Traunmüller, G. Steinmaurer, D. Pozo, J. Ruiz-Arias, V. Fanego, L. Ramirez, M. Romeo, C. Kurz, L. Pomares, and C. Guerrero, "Benchmarking of different approaches to forecast solar irradiance," in *Proceedings of the 24th European Photovoltaic Solar Energy Conference*, Hamburg, Germany., 21-25 Sep 2009, pp. 1–5.

[8] P. Mathiesen and J. Kleissl, "Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states," *Solar Energy*, vol. 85, no. 5, pp. 967 – 977, 2011.

[9] E. Lorenz, A. Hammer, and D. Heinemann, "Short term forecasting of solar radiation based on satellite data," in *EUROSUN2004 (ISES Europe Solar Congress)*, Freiburg, Germany., Jan 2004, pp. 841–848.

[10] R. Marquez and C. F. Coimbra, "Intra-hour DNI forecasting based on cloud tracking image analysis," *Solar Energy*, vol. 91, pp. 327–336, 2013.

[11] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar Energy*, vol. 83, no. 3, pp. 342–349, 2009.

[12] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, no. 10, pp. 1772–1783, 2009.

[13] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569–582, 2017.

[14] P. Lauret, C. Voyant, T. Soubdhan, M. David, and P. Poggi, "A benchmarking of machine learning techniques for solar radiation forecasting in an insular context," *Solar Energy*, vol. 112, pp. 446–457, 2015.

[15] A. Mellit and A. M. Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected pv plant at trieste, italy," *Solar Energy*, vol. 84, no. 5, pp. 807–821, 2010.
[16] V. Sharma, D. Yang, W. Walsh, and T. Reindl, "Short term solar irradiance forecasting using a mixed wavelet neural network," *Renewable Energy*, vol. 90, pp. 481 – 492, 2016.
[17] T. McCandless, S. Haupt, and G. Young, "A regime-dependent artificial neural network technique for short-range solar irradiance forecasting," *Renewable Energy*, vol. 89, pp. 351 – 359, 2016.
[18] L. M. Aguiar, B. Pereira, M. David, F. Díaz, and P. Lauret, "Use of satellite data to improve solar radiation forecasting with bayesian artificial neural networks," *Solar Energy*, vol. 122, pp. 1309 – 1324, 2015.
[19] F. J. Lima, F. R. Martins, E. B. Pereira, E. Lorenz, and D. Heinemann, "Forecast for surface solar irradiance at the brazilian northeastern region using nwp model and artificial neural networks," *Renewable Energy*, vol. 87, pp. 807 – 818, 2016.
[20] M. David, F. Ramahatana, P. Trombe, and P. Lauret, "Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models," *Solar Energy*, vol. 133, pp. 55–72, 2016.
[21] R. Dambreville, P. Blanc, J. Chanussot, and D. Boldo, "Very short term forecasting of the global horizontal irradiance using a spatio-temporal autoregressive model," *Renewable Energy*, vol. 72, pp. 291–300, 2014.
[22] H. T. Pedro and C. F. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," *Solar Energy*, vol. 86, no. 7, pp. 2017–2028, 2012.
[23] J. Huang, M. Korolkiewicz, M. Agrawal, and J. Boland, "Forecasting solar radiation on an hourly time scale using a Coupled AutoRegressive and Dynamical System (CARDS) model," *Solar Energy*, vol. 87, pp. 136–149, 2013.
[24] P. Whittle, "Hypothesis testing in times series analysis." Ph.D. dissertation, Uppsala: Almqvist and Wiksells Boktryckeri AB., 1951.
[25] G. Box and G. Jenkins, *Time Series Analysis Forecasting And Control*. Holden-Day, 01 1970, vol. 3.
[26] M. Valipour, M. E. Banihabib, and S. M. R. Behbahani, "Comparison of the arma, arima, and the autoregressive artificial neural network models in forecasting the monthly inflow of dez dam reservoir," *Journal of Hydrology*, vol. 476, pp. 433 – 441, 2013.
[27] E. Erdem and J. Shi, "Arma based approaches for forecasting the tuple of wind speed and direction," *Applied Energy*, vol. 88, no. 4, pp. 1405 – 1414, 2011.
[28] and, "Short-term load forecasting via arma model identification including non-gaussian process considerations," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 673–679, May 2003.
[29] J. Boland, M. David, and P. Lauret, "Short term solar radiation forecasting: Island versus continental sites," *Energy*, vol. 113, pp. 186 – 192, 2016.
[30] K. Benmouiza and A. Cheknane, "Small-scale solar radiation forecasting using arma and nonlinear autoregressive neural network models," *Theoretical and Applied Climatology*, vol. 124, no. 3, pp. 945–958, May 2016.
[31] L. Mazorra-Aguiar and F. Díaz, "Solar radiation forecasting with statistical models," in *Wind Field and Solar Radiation Characterization and Forecasting: A Numerical Approach for Complex Terrain*, R. Perez, Ed. Cham: Springer International Publishing, 2018, pp. 171–200.
[32] C. Rigollier, M. Lefevre, and L. Wald, "The method Heliosat-2 for deriving shortwave solar radiation from satellite images," *Solar Energy*, vol. 77, no. 2, pp. 159–169, 2004.
[33] L. M. Aguiar, B. Pereira, M. David, F. Díaz, and P. Lauret, "Use of satellite data to improve solar radiation forecasting with bayesian artificial neural networks," *Solar Energy*, vol. 122, pp. 1309 – 1324, 2015.
[34] M. Lefèvre, A. Oumbe, P. Blanc, B. Espinar, Z. Qu, L. Wald, M. S. Homscheidt, and A. Arola, "McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions," *Atmospheric Measurement Techniques, European Geosciences Union*, vol. 6, pp. 2403–2418, 2013.
[35] R. Nau, "Introduction to linear regression analysis," *Linear regression models*, 2014.