Intra-day solar irradiation forecast using RLS filters and satellite images

Franco Marchesoni-Acland[a,b], Rodrigo Alonso-Suárez[a,*]

[a]Laboratorio de Energía Solar, Facultad de Ingeniería, UDELAR, J. H. y Reissig 565, Montevideo, Uruguay
[b]Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, UDELAR, J. H. y Reissig 565, Montevideo, Uruguay

Abstract

Satellite-based solar irradiation forecasting is useful for short-term intra-day time horizons, outperforming numerical weather predictions up to 3-4 hours ahead. The main techniques for solar satellite forecast are based on sophisticated cloud motion estimates from geostationary satellite images. This work explores the use of satellite information in a simpler way, namely spatial averages that require almost no preprocessing. Adaptive auto-regressive models are used to assess the impact of this information on the forecasting performance. A complete analysis regarding model selection, the satellite averaging window size and the inclusion of satellite past measurements is provided. It is shown that: (i) satellite spatial averages are useful inputs and the averaging window size is an important parameter, (ii) satellite lags are of limited utility and spatial averages are more useful than weighted time averages, and (iii) there is no value in fine-tuning the orders of auto-regressive models for each time horizon, as the same performance can be obtained by using a fixed well-selected order. These ideas are tested for a region that has intermediate solar variability, and the models succeed to outperform a proposed optimal smart persistence, used here as an exigent performance benchmark.

Keywords: Solar forecast, RLS filter, ARMA modeling, satellite images, GOES satellite.

1. Introduction

Solar Photovoltaics (PV) has become the world's fastest growing energy technology (REN21, 2019). However, achieving a high penetration of solar PV into electricity grids is a challenging task due to solar irradiance intermittency, caused by cloud dynamics. This variability affects the demand-supply balance that is required for grid operation, implying stability risks and higher management costs, and also affects the operation of electricity markets, adding uncertainty in energy transactions. Resource forecasting is one of the actions to be taken in order to mitigate the negative effects produced by solar variability. Forecasting ability enables better decision-making both in grid and markets operations, providing valuable information for cost-effective spinning reserves management, unit commitment and for establishing more accurate energy prices and quantities for trade.

*Corresp. author: R. Alonso-Suarez, r.alonso.suarez@gmail.com

The research community in solar forecasting has been growing in the last years. As a reference, Google Scholar searches reveal an upward trend with 3.7k, 6.4k, 13.8k, and 16.8k results in each of the last four quinquenniums. There are now well-established methods for operational solar irradiance forecasting (Diagne et al., 2013), namely Numerical Weather Prediction (NWP) models, satellite derived cloud motion estimates, and statistical (learning) procedures over time series data. NWP models are heavily used for day ahead forecasting (Lorenz et al., 2009a; Lara-Fanego et al., 2012; Mathiesen & Kleissl, 2011; Perez et al., 2013), and they have expensive computational requirements. The prediction comes as a product of the sophisticated underlying physical models. Other approaches can not use the information as effectively in such large time horizons, given that the correlations weaken with time. A comparison of NWP models' performance can be found at Lorenz et al. (2009b) and Perez et al. (2011). Alternatively, statistical or machine learning methods have been mostly used over ground data (Pedro & Coimbra, 2012; Lauret et al., 2015; Yagli et al., 2019). High quality solar global horizontal ground measurements are not only used as a basis for the forecasts, but more importantly, they are the ground truth used on the performance evaluation phase. Some of the proposals also integrate additional exogenous variables, as noted by Voyant et al. (2017). When they do not rely on ground measurements, they usually appear as natural methods to combine different models (Lorenz et al., 2012; Aguiar et al., 2016). Finally, satellite-based models are dominated by Cloud Motion Fields (CMF) estimations, being the work of Lorenz et al. (2004) a classical reference. These models are intended to project clouds by means of an estimated velocity field. There are other approaches based on satellite data, involving either different ways to estimate the CMF (Hammer et al., 1999; Peng et al., 2013) or computing correlations (Dambreville et al., 2014). Perez & Hoff (2013) and Kühnert et al. (2013) have reported that a satellite based method outperforms various NWP models when making forecasts up to 4 hours ahead.

As the nature of the methods is fundamentally different, they are not expected to be totally redundant with each other. A method that integrates two or more approaches will likely perform better. For example, while solar measurements are taken at one specific point, satellite data provides information about the cloudiness on the surrounding areas, that can be exploited by forecasting techniques to reduce the prediction uncertainty. A few works exploring the combinations of the methods were conducted by Marquez et al. (2013); Aguiar et al. (2015, 2016); Bright et al. (2018); Harty et al. (2019). The input information given to these methods is diverse. Some input data require preprocessing and others do not; for instance, Marquez et al. use a segmented satellite image taking the cloudiness information in the form of a ladder oriented by the main cloud displacement between images. Some works use two different methodologies for the same conceptual input, e.g. using NWP outputs from a GFS (Global Forecasting System) driven WRF (Weather Research and Forecasting) or WRF-Solar (Jimenez et al., 2016), or using CMF information by means of Lorenz et al. (2004) technique or regular optical flow techniques (Horn & Schunck, 1981; Lucas & Kanade, 1981). Given the high uncertainties still observed in solar forecast techniques, there is a need to further analyze the combination of the various input data.

In this work we explore the combination of input sources by means of a statistical signal processing approach. In particular, we aim to combine ground measurements with satellite information, providing a detailed assessment of the combination. The time series analysis literature is vast, including Artificial Neural Networks (ANN), classical machine learning approaches (Support Vector Machines, Trees), and statistical time series models. A recent comprehensive review on these methodologies can be found at Voyant et al. (2017). Most of these methods are suited to include scalar exogenous variables. Here, we make use of Auto-Regressive Moving-Average (ARMA) models, that have proved to work well in this problem (Reikard, 2009). More specifically, the ARMA model is formulated as an adaptive filter through the Recursive Least Squares (RLS) algorithm as in David et al. (2016); Marchesoni-Acland et al. (2019). In Aguiar et al. (2015) and Dambreville et al. (2014) satellite data is integrated as input to statistical models. Both approaches avoid using complex CMF methodologies and use correlations in order to decide which pixels (or block of pixels) to include. We analyze the use of a simpler satellite input with an approach that involves almost no preprocessing: taking the mean of a window of the satellite albedo image. Satellite data carries valuable information of the surroundings of the forecast site, therefore introducing present and past satellite cloudiness data is a good way of improving short-term intra-day forecasts. Combining these present and past values can be thought as weighted time-averaging. In order to compare time-averaging with spatial-averaging the size of the spatial averaging window is modified as well. It is expected that satellite information is, to some degree, redundant with solar irradiance measurements, as irradiance estimates can be inferred from satellite images (Perez et al., 2002; Rigollier et al., 2004; Alonso-Suárez et al., 2012; Qu et al., 2017). This fact is analyzed by comparing results obtained by using more observations of ground data than of satellite data and *viceversa*. The procedure to select the best model, comprising the ARMA model selection, the number of lags on the satellite inputs, and the averaging window size, is presented as well. We evaluate the benefits of using different parameters (ARMA coefficients, satellite lags and window size) and compare the models against a challenging benchmark that arises from an optimal selection of smart persistences. This article demonstrates that this simple proposal works for intermediate solar variability sites, such as the region under study in this work.

The main contributions of this article can be summarized as follows:

- It proposes and evaluates a methodology to easily include raw satellite data (albedo) into a time series forecasting model. As shown in Aguiar et al. (2015); Marquez et al. (2013) adding solar satellite estimates improves the forecasting performance. Here, a previous step is addressed, including raw satellite albedo as input, without the postprocessing added by a solar satellite model which may add uncertainty to the problem. To the best of our knowledge, the use of raw satellite information as input of solar forecasting methods has not been tested in the literature. A detailed evaluation is made, using a challenging performance upper limit (an optimal smart persistence) for the Forecasting Skill (FS) calculation.

- It provides an assessment of the forecasting gain by adding raw satellite information to a baseline ARMA-RLS model that only uses ground measurements. A performance analysis when varying the final model's parameters is provided, in particular, the p and q ARMA-RLS parameters, the satellite averaging window size and the satellite past samples (satellite lags).

- It shows that when using only ground measurements as input there is not much to be gained by fine-tuning the ARMA-RLS model's parameters. The best performance, which is achieved by setting the optimal parameters for each lead time, presents a negligible difference with the performance that can be obtained by using a few fixed auto-regressive and moving average terms for all lead times.

- It shows that when adding satellite albedo, the utility of the ground measurements past samples as input is restricted only to the very short-term forecast horizons (up to 30 minutes ahead). Above this limit, the performance of models that use satellite albedo is insensitive to ground measurements lags. In all cases, models including satellite information, whether they include ground measurements past values or not, achieve the best performance for all time horizons.

- It defines and uses a natural challenging persistence benchmark that is obtained from the utilization of the optimal smart persistence procedure at each lead time. This defines the best performance curve that the simple smart persistence procedure can obtain. As explained in Subsection 4.2, some authors differ and use a few different benchmark definitions of persistence or smart persistence. These definitions and some relevant work in this topic are discussed in Subsection 4.2, ending with the introduction of the optimal smart persistence benchmark.

This article is organized as follows: in Section 2 the data is presented along with a description of the equipment and stations' characteristics, the exogenous variable being used and the data quality procedure. In Section 3 the RLS algorithm is introduced, with a brief mention of the advantages of the approach. In Section 4 the evaluation framework is presented, describing the performance metrics to be used (Subsection 4.1) and the optimal smart persistence (Subsection 4.2). Section 5 provides the results obtained with the different models and the performance analysis. The ARMA-RLS model selection is discussed in Subsection 5.1 while the inclusion of satellite albedo is addressed in Subsection 5.2. Finally, our conclusions are summarized in Section 6.

2. Data

This section describes the two types of data used in this work: global horizontal irradiance ground measurements (GHI, G_h), and Earth albedo (ρ_p) derived from visible channel GOES-East satellite images.

2.1. *Solar irradiance data*

Solar irradiance measurements recorded at seven ground stations in the south-east part of South America are used in this work. Two of these sites, the Solar Energy Laboratory (LES, http://les.edu.uy/) experimental facility at the North of Uruguay (LE) and the São Martinho da Serra station from the SONDA (Sistema de Organização Nacional de Dados Ambientais) network (http://sonda.ccst.inpe.br/) at the South of Brasil (MS), record GHI measurements with equipment and procedures that comply with BSRN (Baseline Solar Radiation Network, https://bsrn.awi.de/) requirements (McArthur, 2005), being the latter formally a BSRN site. In these sites, the GHI is measured using spectrally flat Class A pyranometers (according to the ISO 9060:2018 standard) and routine maintenance is performed on a daily basis, such as dome cleaning. The other five stations are part of the LES solar irradiance measurement network and are located on field in semi rural environments. They are equipped with spectrally flat Class A or B pyranometers for the GHI measurement and maintenance is done on a monthly basis by personal at the stations. Based on our experience, equipments' quality, calibration schemes, and maintenance schedules, we assign a global (P95) uncertainty for GHI measurements of 3% of the average at the LE and MS sites and of 5% in the rest. These uncertainties are way lower than the uncertainty of the forecast being evaluated in this work.

Table 1 presents the sites' location, data span, and some relevant measurements' characteristics, namely the GHI average value, $\overline{G}_h$, and the 10-minutes nominal variability, σ. The GHI average is the value that will be used to express the performance metrics as a percentage. The nominal variability is defined by Perez et al. (2016) as the standard deviation of the changes in the clear-sky index time-series, $\sigma = \mathrm{Std}\{\Delta k_c(t)\}$. The clear-sky index is defined as,

$$k_c(t) = \frac{G_h(t)}{G_h^{\mathrm{csk}}(t)}, \tag{1}$$

where G_h^{csk} is the output of a clear-sky model. Here, the McClear model is used (Lefèvre et al., 2013), publicly available at the CAMS (Copernicus Atmosphere Monitoring Service) platform (http://www.soda-pro.com), to calculate the clear-sky index from the GHI time series. The values provided in Table 1 were calculated over the 10-minutes quality-checked daylight solar irradiance data set, as explained in Subsection 2.3.

These stations are representative of the subtropical temperate climate of the south-east part of South America known as Pampa Húmeda, which is classified under the updated Köppen-Geiger climate map as Cfa (Peel et al., 2007). This is a warm, temperate and humid climate, with hot summers. The solar variability of the region is intermediate, both in terms of inter-annual variability (Alonso-Suárez, 2017) and short-term variability. The latter, more relevant for this work, is quantified by nominal variability and has an average of $\sigma = 0.148$ in the region (see Table 1). Hence, the results provided in this work are applicable to sites with similar climate conditions (intermediate variability and Cfa or Cfb), as Central and South-East US, non-Mediterranean Europe and East Australia, among others. For other climates or different solar variability sites, as low-variability desert sites or high-variability insular locations, results may not be extrapolable and further investigation is required.

station name	**station code**	**period of time**	**lat.** (deg)	**lon.** (deg)	**alt.** (m)	$\overline{\mathbf{G}}_{\mathbf{h}}$ (W/m^2)	σ (–)
LES facility	LE	01/2016 – 12/2017	-31.28	-57.92	56	461	0.139
São Martinho	MS	01/2012 – 12/2015	-29.44	-53.82	489	451	0.149
Artigas	AR	01/2015 – 12/2017	-30.40	-56.51	136	451	0.147
Las Brujas	LB	01/2015 – 12/2017	-34.67	-56.34	38	440	0.152
Tacuarembó	TA	01/2016 – 12/2017	-31.71	-55.83	142	443	0.147
Rocha	RO	01/2016 – 12/2017	-34.49	-54.31	20	425	0.159
La Estanzuela	ZU	01/2016 – 12/2017	-34.34	-57.69	70	442	0.144
			all sites average			**445**	**0.148**

Table 1: Solar irradiance measuring sites: location, characteristics and data span.

The k_c time series, at 10-minutes granularity, is the ground measurements input considered for the forecast algorithm. This is common practice in the solar forecasting field, as the GHI time series has a daily and seasonal geometrical behavior that introduces a deterministic complexity on the statistical learning approaches. This deterministic behavior can be easily eliminated by using clear-sky estimations (or even top of the atmosphere irradiance calculations), isolating the higher-rate fluctuations due to cloudiness. With this methodology, the forecasting models can be dedicated to predict the non-deterministic component of solar irradiance due to clouds dynamics, leaving the geometric part to be represented by the clear sky model.

2.2. Satellite images

GOES-East satellite visible channel images are used here by means of the Earth Albedo (ρ_p). We are not using, for instance, solar satellite estimates. The images are preferred in a non-processed version, as a way to exclude the uncertainty associated with the conversion of the Earth Albedo (mainly, cloudiness information) to solar irradiance. The satellite images used in this work were generated by the GOES12 and GOES13 satellites, which operated in the GOES-East position during the considered time period (see Table 1). During that time, the GOES-East series provided irregular acquisition for South America, usually available at a rate of two images per hour. The 10-minutes time resolution for the satellite albedo was obtained via a linear interpolation of the satellite time series. Satellite gaps of more than two consecutive hours were not interpolated and were removed from the data set. Our local GOES-East satellite database has a spatial coverage of the Pampa Húmeda region and surroundings areas.

The former GOES-East satellites (GOES12 and GOES13) had a nominal spatial resolution of 1 km on their visible channel. The location of these satellites (geostationary orbit, 75°W) results in a pixel size of about 1-2 km over the region. To include satellite information in a simple way into the forecast algorithm, an average value is calculated in a cell centered at each site. As we are interested in analyzing the forecast performance and features for different satellite spatial average sizes, different cell sizes are tested. For easy communication, we choose three different cell sizes: small, medium and large, representing each a 1 arcmin $\times$ 1 arcmin, 10 arcmin $\times$ 10 arcmin and 20 arcmin $\times$ 20 arcmin latitude-longitude cells. This approximately corresponds to cell sizes of 1.9 km $\times$ 1.6 km, 19 km $\times$ 16 km and 37 km $\times$ 31 km, respectively, over the target region.

2.3. Data filtering

The data quality check and filtering is as follows. First, we exclude data with solar altitude lower than 10° to avoid using early morning or late afternoon observations which present higher relative deviations due to cosine error in the measurements. This is a standard filtering procedure that ignores only $\simeq$ 1% of the annual total solar energy (David et al., 2016). Then, we remove erroneous or missing data in the measurements or the satellite time series (3% of the data). Our GHI data is flagged in order to allow a 10 minute observation to be calculated only from the 1-minute time-series if at least 7 minutes of data are available (more than 66% of the interval). Next, two filters associated with irradiance upper limits are applied over the GHI data set (and remove between 0.1-0.2% of the data): (i) the BSRN quality procedure to detect physically impossible and extremely rare GHI measurements (Ohmura et al., 1998) and (ii) the exclusion of observations with clear sky index exceeding the value 1.35. Finally, a last check is done over the variability metric, discarding a few variability outliers that arise due to the previous filtering stages ($\simeq$ 0.1% of the data). This last check is only intended to remove very few outliers associated with the data gaps originated from the previous filtering stages. Some of these gaps cause artificial large changes in two consecutive k_c samples that can affect the auto-regressive modeling. The threshold for this filter was heuristically set to comply only with this objective without affecting the natural solar resource variability of the sites. The filtering procedure is summarized in Table 2 for each station and the complete data set. The last column shows the fraction of data that is being filtered. As can be seen, around 4.4% of the initial data is discarded by these procedures.

3. Algorithms

3.1. ARMAX models

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-j}) and past exogenous

Table 2: Quality check and data set description for each measurements station.

	solar alt. $> 10^{o}$	missing or erroneous		upper limit filters		variability filter		filtered
site	**samples**	samples	(%)	samples	(%)	samples	(%)	**(%)**
LE	46198	44585	3.5%	44530	0.1%	44318	0.5%	4.1%
SM	92541	85528	7.6%	85377	0.2%	85120	0.3%	8.1%
AR	69335	66941	3.5%	66836	0.2%	66820	$< 0.05\%$	3.6%
LB	68780	67157	2.4%	67068	0.1%	67031	0.1%	2.5%
TA	46155	44635	3.3%	44541	0.2%	44528	$< 0.05\%$	3.5%
RO	45886	44520	3.0%	44467	0.1%	44437	0.1%	3.2%
ZU	45927	44633	2.8%	44584	0.1%	44555	0.1%	3.0%
total	**414822**	**397999**	**4.1%**	**397403**	**0.15%**	**396809**	**0.15%**	**4.4%**

variables (E_{t-k}). If p and q are the orders of the AR and the MA terms respectively, the model is described by,

$$X_{t+1} = c_0 + \sum_{i=0}^{p-1} \alpha_i X_{t-i} + \sum_{j=0}^{q-1} \beta_j \epsilon_{t-j} + \sum_{k=0}^{l-1} \gamma_k E_{t-k} + \epsilon_t, \tag{2}$$

where c_0 is an independent term and ϵ_{t+1} is assumed to be white Gaussian noise and set to zero when forecasting. The offset term (c_0) allows to improve the modeling of processes of non zero mean. Using an ARMAX model to make forecasts implies finding the set of parameters c_0, α_i, β_j and γ_k, and then computing a step forward for some input at time t. ARMAX models are well-known as a prediction tool and are a natural generalization of ARMA models, which were popularized by Box & Jenkins (1970).

In the current work, the variables of Eq. (2) are defined as follows: X_t corresponds to the clear sky index k_c at time t, our main time series, which we want to predict. The k_c predictions are then transformed back to GHI to compute the metrics values. As the k_c time series has a non-zero mean, the offset term c_0 is useful to enhance the model's performance. The ϵ_t is the k_c error obtained when forecasting X_t and finally, the E_t represents the average cloudiness around the site, computed as explained in Subsection 2.2. Later, the l term in the upper limit of the exogenous variables sum will be called "lags". The terms inside the summations refer to present time (time t) and past values of these quantities. Eq. (2) is presented as a standard ARMAX filter in which the next step ($h = 1$) is predicted. Note that all terms in the right hand side of Eq. (2) are known except the innovation e_{t+1}. This expression can be adapted to any arbitrary forecast horizon.

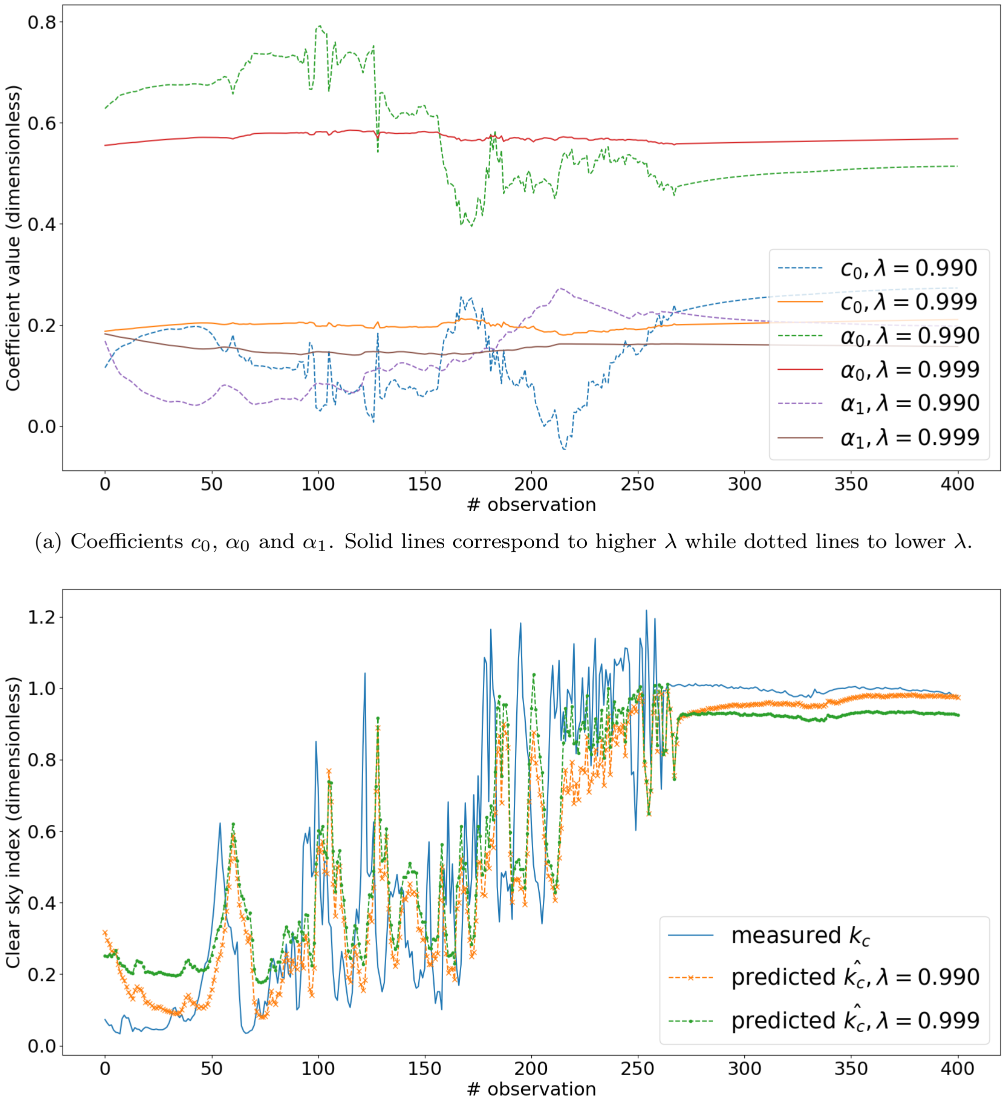
3.2. RLS filter

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

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

where $e(i)$ is the forecasting error of observation i. 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 and X_i the target value. 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. The algorithm presents some similarity to computing Least Squares in a moving window, being the main differences the exponentially decaying weights in the cost function and the fact that the computation is recursive. The mathematical generalization and resulting algorithm of the ARMAX-RLS framework for an arbitrary lead time h is detailed in the Appendix of Marchesoni-Acland et al. (2019).

In traditional signal processing literature, the RLS algorithm is classified as an adaptive filter. Being adaptive makes historical data unnecessary, i.e. it avoids using train-test splits and fixed-weights, making the approach useful for operational context. Furthermore, statistical properties vary between seasons and even days (i.e. cloudy and clear-sky days), causing short-term adaptability to be a desirable property in general. The adaptability of the RLS algorithm depends on the value of the forgetting factor $\lambda \in [0, 1]$. The algorithm's behavior when changing the λ value is illustrated in Figure 1, using as an example a 1-hour ahead ($h = 6$) naive ARMAX forecast using $p = 2$, $q = 0$ and no exogenous variables, which is usually denoted as an AR(2) filter. The algorithm is highly sensitive to the λ value, so two options close to one are analyzed in Figure 1: $\lambda = 0.999$ y $\lambda = 0.990$. The coefficients of the filter are, as expected, more stable for a larger value of λ (see solid lines in Figure 1a). For large λ, the coefficients do not change much when a high variability period is found, as shown in the first $\simeq 275$ observations of Figure 1. On the other hand, when there is clear sky, forecasts with large λ consistently underestimate the target value due to the very low convergence rate, as can be noted in the last $\simeq 100$ observations. Using a smaller value of λ makes convergence to the steady state of k_c faster, yielding also to lower bias estimates under clear sky conditions. However, the values of λ closer to 1 are the ones that achieve better numerical results in the long run. The loss function is a weighted sum of squared errors, making it convenient for any algorithm to 'play safe' to some extent: an underestimating forecast under clear sky conditions will achieve less quadratic error when clouds appear. By inspecting this behavior, the λ value was set to $\lambda = 0.999$, as was previously used by David et al. (2016). This allows anyone to compare results with this previous work if desired, and it is a λ value close enough to 1 to provide almost the best results, although in our experience higher λ values achieve better performance across all the time series. This observation may be climate dependent.



(a) Coefficients c_0, α_0 and α_1. Solid lines correspond to higher λ while dotted lines to lower λ.

(b) Clear sky index. Dotted lines correspond to the AR(2) predictions.

Figure 1: Qualitative behavior of RLS filter with different forgetting factors λ. The example shows the behavior of the AR(2) model for the LE site and a time horizon of 1 hour ($h = 6$). The x axis in both figures is the same.

4. Evaluation framework

4.1. Metrics

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

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

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

where $\hat{y}_h(t)$ is the GHI forecast made at time t with time horizon h and $y^{\mathrm{ref}}(t)$ is the GHI measurement taken at time t. The MBD definition is such that a positive value means a forecasting overestimation and a negative value means a forecasting underestimation. Their relative values, rMBD and rRMSD, are expressed as a percentage of the average irradiance value (see Table 1). The relative values shall be used in order to reduce the effect of the geographical location in the metric values and provide a more intuitive error indicator. However, as the Tables provided in the Appendix A include the average irradiance values, the reader may calculate the absolute indicators if desired. These two metrics and their normalized values are very popular indicators to evaluate solar forecast (Yang et al., 2018).

The forecasting skill represents the gain of the forecasting RMSD with respect to the persistence procedure, and it is defined as,

$$\mathrm{FS} = 1 - \frac{\mathrm{RMSD_m}}{\mathrm{RMSD_p}}, \tag{6}$$

where the subscripts 'm' and 'p' refer to the model and persistence respectively (the result is the same when using relative errors). The traditional 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 predicted k_c value. Then, the corresponding GHI is predicted by using the clear sky model estimates. Being a simple procedure, the persistence is then used as a benchmark to measure how good a forecast is: any additional complexity of the forecasting procedure should imply an improvement in comparison with the persistence to be worthwhile. This metric implicitly takes into account the difficulty of forecasting at each location via the persistence's RMSD (Coimbra et al., 2013), and thus it is a better indication than the RMSD or rRMSD, which are insensitive, for instance, to the local short-term resource variability. However, the choice of the benchmark over which to calculate the FS is arguable. An alternative to persistence is to persist in time the past observations' average, known as Smart Persistence (SP). For this work, the "best" SP (bSP) is used as performance reference. This defines the forecasting skill metric that will be used later, defined exactly as in Eq. (6), but using $\mathrm{RMSD_{bSP}}$ instead of $\mathrm{RMSD_p}$. The procedure to obtain this bSP is detailed in the next subsection.

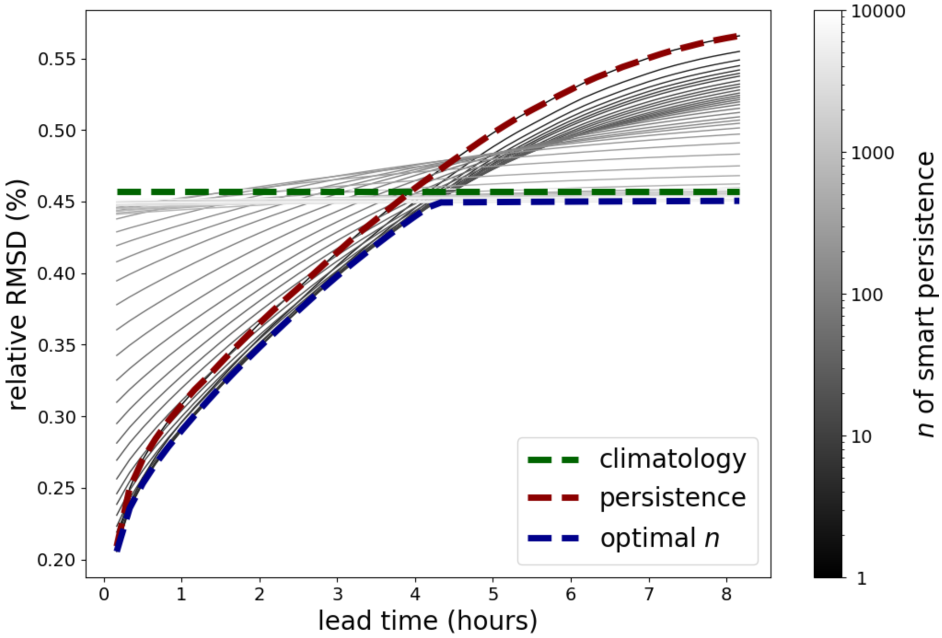
4.2. Smart persistence

A slightly more complex variant of the classical persistence (P) method is called "smart persistence" (SP). The P and SP methods are defined in different ways in the literature. For example, Voyant & Notton (2018) define P as persistence of the GHI value and SP as persistence of the k_c index. David et al. (2016) define SP as the average of the last h values of k_c, being h the forecasting time horizon. Let us now introduce a slightly more general formulation, to be called general SP (gSP). This method sets the number of past observations to be averaged to an arbitrary number n_h. This is, $\mathrm{gSP}_h(t) = \frac{1}{n_h}\sum_{i=0}^{i=n_h-1} k_c(t-i)$. In this work the optimal value of n_h is obtained for each time horizon. Note that when $n_h = 1$ for all h, the traditional P method is recovered, when $n_h = h$ we have David et al. SP method and when n_h is large for all h, the prediction approximates the climatology value ($n_h \gg 1$). It is known that the shorter time horizons are better predicted by the classical persistence, while the longer time horizons are better predicted by the climatology value. In Yang (2019) a linear combination of persistence and climatology is proposed as benchmark, after mathematically proving that the weights are directly related to the autocorrelation of the time series at the time horizon studied. Here it is argued that there exists an optimal n_h^* value for each lead time h (the value of n that minimizes the prediction RSMD for that time horizon), which is not usually assessed nor presented. We now define the best SP (bSP) as the gSP that uses the optimal values of n_h, which we call n_h^*. This is, $\mathrm{bSP}_h(t) = \frac{1}{n_h^*}\sum_{i=0}^{i=n_h^*-1} k_c(t-i)$. The bSP method can be seen, by definition, as the Pareto frontier when framing the procedure as a multiobjective optimization problem.

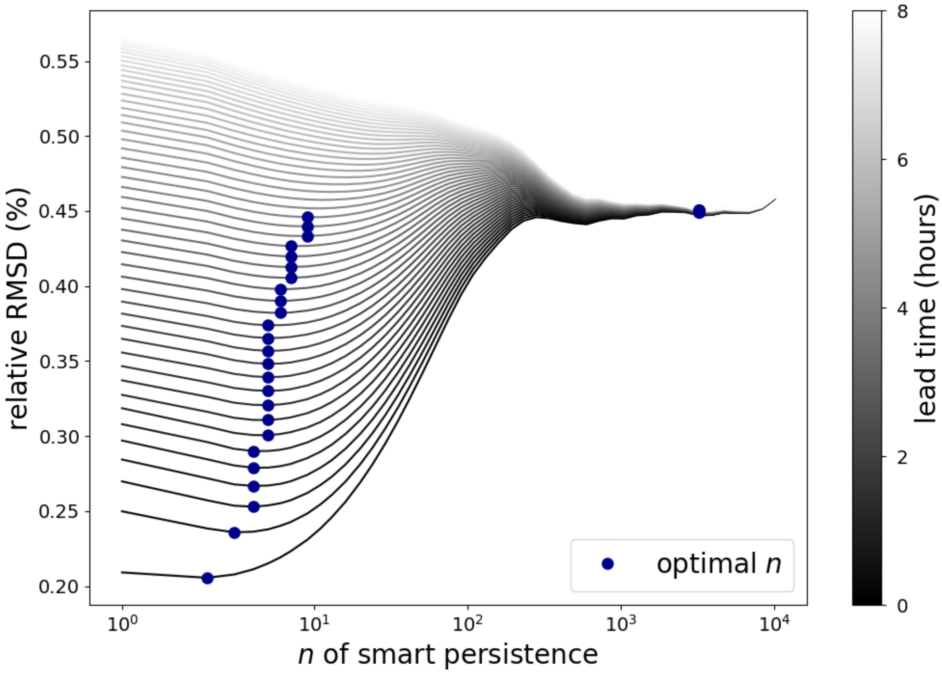
To obtain the best SP, the n_h^* were obtained for each site and time horizon. Performance evaluation of gSP is done via a grid search methodology: forecasts are made for every location, for every lead-time value up to 8 hours ahead with 10 minutes granularity, and over different values of n. The results, depicted in Figures 2a and 2b for the average of all sites, show two things. First, the value of optimal n_h increases with lead-time, but not following $n_h^* = h$. Second, there is a time horizon from which smart persistence forecasts with bounded n are worse than climatology forecasts. This breakpoint in which big n persistence (*pseudo* climatology) is better than small n persistence, is best observed in Figure 2b. This rather atypical plot is useful to observe the n_h^* variation with the lead times, as each blue dot is located in the absolute minimum of each curve, and smaller lead times correspond to curves closer to the bottom. It can be seen that n_h^* increases slowly (note that the x-axis is logarithmic) with forecasting lead time until the breakpoint is reached (blue dots located between $n = 10^3$ and $n = 10^4$, which correspond to larger time horizons). In the breakpoint, the forecasting lead time is large enough to make the information contained in the recent past samples less valuable than the historical aggregate. This breakpoint can also be observed in Figure 2a as the saturation of the optimal n curve and for the region under study is located at 4 hours and 20 minutes (sites' average). For time horizons longer than this point it makes little sense to benchmark with any kind of smart persistence, as the simple climatology value will be better.

The procedure for obtaining the "best smart persistence" can be viewed as computing the RMSD curve (vs h) for each possible value of n, and then taking the bottom envelope of the curves as the $\mathrm{RMSD}_{\mathrm{bSP}}$. Figure 2a shows in gray scale the site-averaged RMSD curves with varying n, and three special curves are identified: the classical persistence in red, the climatology in green and the best persistence performance in blue. This RMSD bottom envelope, for each station, is the benchmark used in this work to define the FS from Eq (6).

In a real case scenario, the n_h^* values (over past and future) are unknown. Optima obtained from historical



(a) rRMSD as a function of h.



(b) rRMSD as a function of n.

Figure 2: Relative RMSD versus lead time for different values of smart persistence n.

data are not guaranteed to keep being optimal in the future, although the likelihood of this happening grows with historical data size. The same thing happens when using climatology as a benchmark. It is noted here that this subtle case of data snooping is harmless because persistence models are only used as benchmarks instead of operationally. The procedures presented here and in Yang (2019) are of different nature, being the present benchmark more computationally expensive. However, this optimal smart persistence benchmark is computed only one time as a standalone calculation for the purpose of performance assessment, hence it does not represent a limitation for operational systems. Future work should include the comparison of both methodologies, addressing different climates and solar variability sites.

5. Results

This section presents the performance results of the forecasting algorithms. The focus is on understanding and quantifying the performance gain of adding averaged satellite albedo to the ARMA-RLS baseline model that only uses ground measurements. Subsection 5.1 introduces the results for the baseline model, providing a discussion on model selection and showing that model fine-tuning is of little utility. Subsection 5.2 presents the core results of this article, addressing the inclusion of satellite albedo into the auto-regressive framework. A grid-search methodology among a large set of possible configurations was used to obtain the results. We focus the following discussion in our main findings, that are in fact the contributions of this article (not the grid-search itself or parameter tuning). As the data volume of such grid-search is extensive and in most cases fine-tuning does not yield to significant performance changes, as it will be demonstrated in the following, we present here a small subset of the full simulation and the results in this section will be presented via graphical aids. Nevertheless, the corresponding quantitative results are provided in Tables in Appendix A. The results of the full simulation for Subsection 5.2, including p, q, satellite lags and satellite pixel size, are available in: http://les.edu.uy/RDpub/ARMAX-grid-search.xlsx.

5.1. Endogenous RLS filter

Figures 3a to 3c show the rRMSD for each p (number of AR terms) and q (number of MA terms) averaged over all locations. It is observable in these examples that with $q = 0$ and $3 \leq p \leq 8$ the performance is near the optimal one. For the shorter time horizons ($h = 1$, 10 minutes, Figure 3a), after a certain value of p and q, i.e. for $p \geq 3$ and $\forall q$, the performance is rather similar, with rRMSD variations below 0.1%. The rRMSD span over all p and q values is $\simeq 0.8\%$, which is not very high. For the longer time horizons ($h = 24$, 4 hours, Figure 3c), the surface seems to favor $q = 0$ more clearly, but the performance difference over different values of q and p near the optimum is below 0.2% (negligible, see the plot's z-axis scale). It is clear that for large time horizons, there is less performance gain by tuning the p and q values for each lead time. Intermediate lead times stand between both situations. Hence, for all time horizons, the gain obtained by fine-tuning p and q is below 0.8% of rRMSD. Furthermore, we will shortly show that if a fixed pair of these values is intelligently set, there is no global significant performance gain in optimizing (p, q) for each lead time.

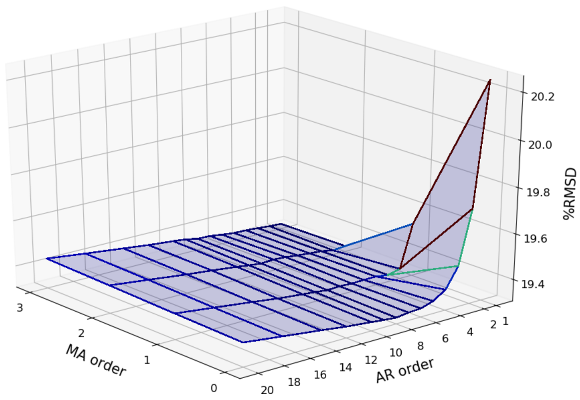
To find a good global model with fixed (p, q) parameters, the procedure is as follows: we calculated all the rRMSD surfaces averaged over the locations, subtracted the mean of the rRMSD at each lead time, and averaged the result. This procedure obtains the mean rRMSD anomalies surface, whose minimum is the best fixed operation point. This procedure avoids giving more importance to one specific lead time. The result is shown in Figure 3d and the minimum is located at $(6, 0)$. This is very close to the ad-hoc $(5, 0)$ model that we analyzed in a preliminary work (Marchesoni-Acland et al., 2019). In that preliminary work, we observed that the performance of an arbitrary $(p = 5, q = 0)$ model was indistinguishable from the best error achievable with any bounded combination of p and q (the bounds were $p \leq 10$ and $q \leq 4$). For simplicity and ease of comparison with previous work, and as there is a negligible difference between using $(5, 0)$, the global optimum model or the fine-tuned (p, q) models for each lead time, we will keep $p = 5$ for the analysis.

Figures 4a and 4b show the forecasting skill of different (p, q) models in comparison with the optimal model, i.e. the model that uses the optimal p and q values for each time horizon. For easy comparison with other works and to visualize the effect of using different persistence procedures, Figures 4a and 4b show the FS using the regular persistence and the optimal smart persistence, respectively. It is to be noted the different span (y axis) and the different behavior for shorter lead times (up to 1 hour ahead) and longer lead times (for 4-5 hours ahead, when the pseudo climatology is the best benchmark), where the concavities are different. One can observe that the bSP method is indeed difficult to beat, surpassing the performance of an ARMA $(1, 0)$ model. This is depicted as negative values of the FS in Figure 4b. The performance when using $(5, 0)$ and $(5, 1)$ remains very close to each other and to the optimal (p, q) model. This in fact shows that using a fixed well-selected pair of (p, q) obtains essentially the same performance as the optimal choice and that fine-tuning the ARMA-RLS filter is futile. For short time horizons, the effect of adding a MA term (q) is positive for $p = 1$, but is insignificant for $p = 5$. For long time horizons, adding a MA term tends to slightly downgrade the performance.

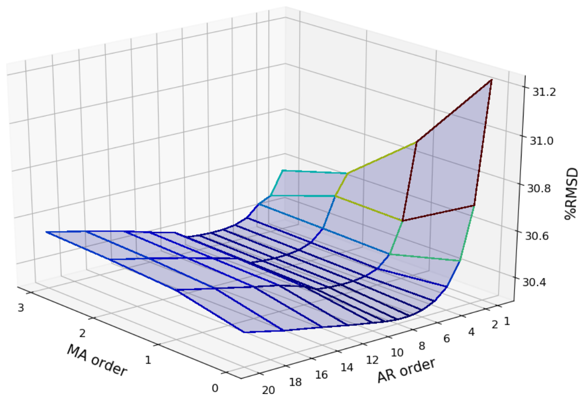
5.2. RLS filter including satellite albedo

Including satellite albedo data improves the models' performance compared to using only ground measurements, as shown in this subsection. The models used here will only include AR terms, as the difference is insignificant for $p \geq 3$. Henceforth, "lags" will be used to refer to past satellite observations, i.e. albedo observations previous to time (t). Also, the performances will be expressed only in terms of the FS metric using as benchmark the optimal smart persistence in order to avoid redundancy. The cases with $p = 5$ and $p = 1$ and no satellite input are included in the following figures as a performance reference, and the FS curves are the same of Figure 4b for the $(5, 0)$ and $(1, 0)$ models, respectively.

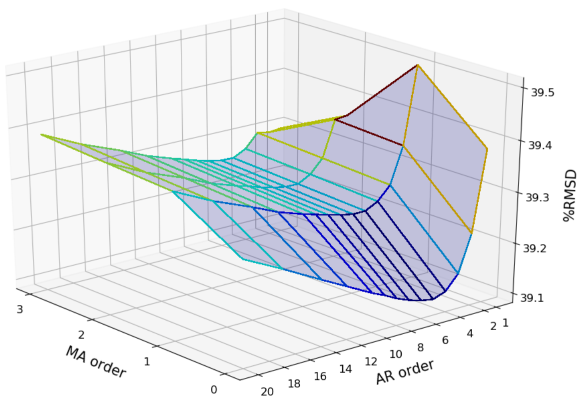
The effect of the number of AR terms p on the performance when including a single value of satellite albedo data is presented in Figure 5a. The pixel size used here is medium (see Subsection 2.2). The addition



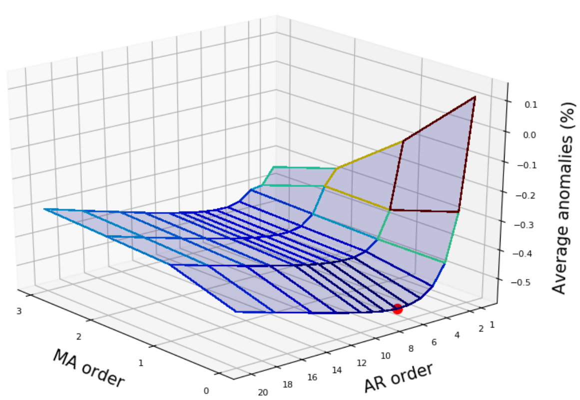
(a) Forecast horizon: 10 minutes ($h = 1$).



(b) Forecast horizon: 90 minutes ($h = 9$).

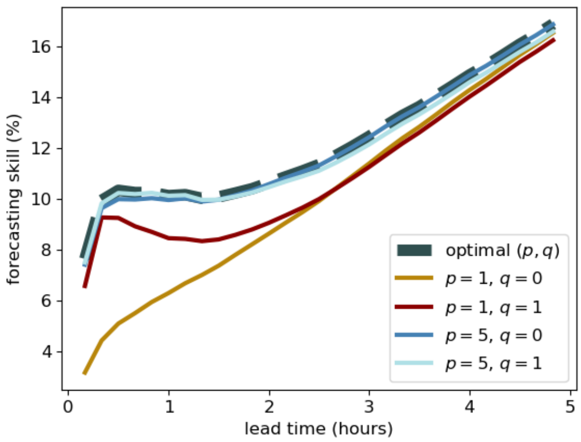


(c) Forecast horizon: 240 minutes ($h = 24$).

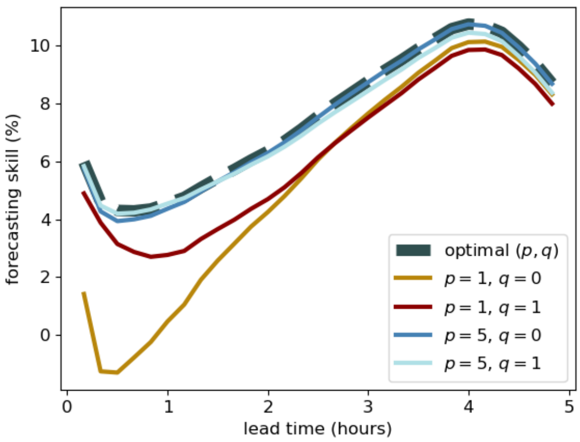


(d) Average rRMSD anomalies over all lead times.

Figure 3: Relative RMSD analysis for the different (p, q) parameters of the endogenous ARMA-RLS filter.



(a) Using regular persistence.



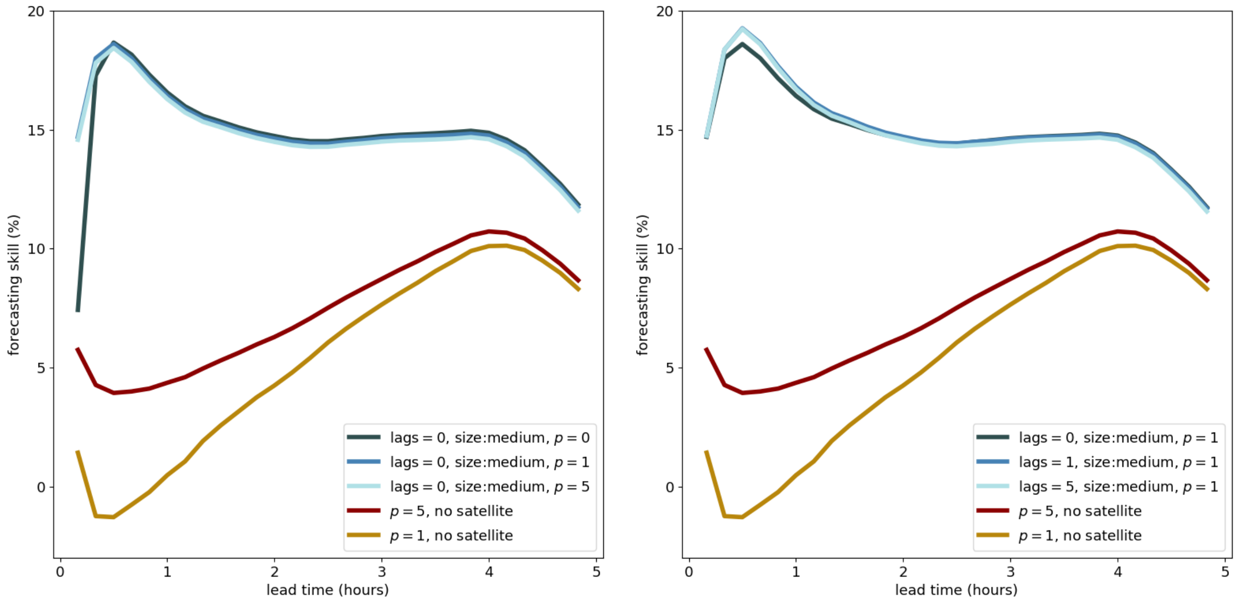
(b) Using the benchmark optimal smart persistence.

Figure 4: Forecasting Skill of the ARMA-RLS filter with different parameters in comparison with optimal choice.

of satellite albedo enhances performance significantly for all lead times. Peak performance is obtained at 30 minutes ahead (FS $\simeq +18.5\%$), being a $\times$ 4 improvement over the ARMA $(5, 0)$ model for that time horizon. In general, it can be seen that the higher impact of adding satellite albedo is at the shorter time horizons, i.e. during the first hour ahead, but then a remnant improvement persists for longer lead times, declining during the last forecast hour (4-5 hours ahead) in the same manner as the rest of the models. Furthermore, Figure 5a shows that increasing p ceases to be useful when satellite data is present (see $p = 1$ and $p = 5$ cases). Taking this to the edge and using no ground data in the algorithm's input ($p = 0$) only implies sacrificing significant performance on the 10 minutes lead time, as the FS is almost the same for larger lead times. This does not mean that ground measurements are unnecessary: they are used to generate the error signal that is fed back into the RLS algorithm. Therefore, this does not mean that the $p = 0$ algorithm is only running on satellite information. The use of solar satellite estimates to completely replace the measurement signal is left as future work, but recent studies suggest this may be possible without a significant performance reduction (Yang & Perez, 2019).

Another experiment was made: adding lags on the satellite cloudiness data. It should be noted that, as the 10-minutes satellite data series is obtained via interpolation from a smaller time resolution, there is some degree of redundancy in this information. Adding lags can be seen as a type of time-averaging, as a set of weights $\{\gamma_k\}$ will be assigned to each past data-point. The value $p = 1$ was used in this test as there is no significant improvement by using $p = 5$ when satellite albedo is also used (as seen in Figure 5a). The analysis is shown in Figure 5b's blued curves. A little performance improvement on the peak of the FS curve is found, of around 1% of FS. For larger time horizons the effect is negligible. This happens for both lags $= 1$ and lags $= 5$, showing a behavior similar to that of the p value: there is no extra value in adding more satellite lags than lags $= 1$. In Figure 5b the curve with one satellite lag is indistinguishable from the one with five satellite lags.

The third analysis is about the impact of the spatial window in which the values of satellite albedo are averaged. $p = 1$ and one lag on satellite data are used in this case, in order to quantify the impact of the window size in the best model inspected so far. Figure 6a shows the models' performance when using the small, medium and big cell size (defined in Subsection 2.2). It is observed that larger cell sizes are preferred. A significant performance improvement is observed when using a medium cell size in comparison with a small cell size, especially up to $\simeq 2$ hours ahead. The bigger cell inspected here is the one which provides better performance, being similar to that of the medium cell size up to the 30 minutes time horizon, but showing an improvement for longer ones. Another interesting observation is the location of the peak: using a larger spatial window implies moving the peak in the direction of larger lead times. Note also that the concavity of all curves that include satellite data is negative in the shorter lead times, denoting a relative advantage over forecasts methods that do not include satellite data in these time horizons: spatially averaged satellite albedo effectively improves the forecast in the first forecast hour and has a positive effect over all



(a) With varying p.

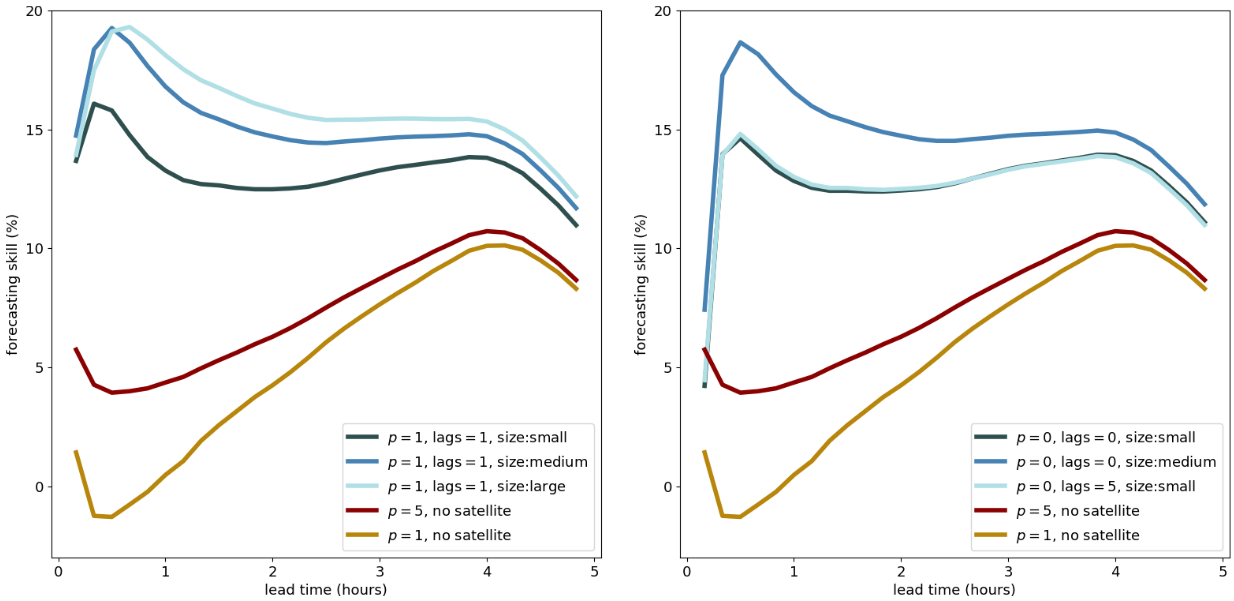
(b) With varying satellite lags.

Figure 5: Forecasting Skill for ARMA-RLS filter using satellite albedo.

time horizons.

The last test is shown in Figure 6b and refers to the absence or presence of the "ergodicity" property of satellite albedo images for solar forecasting. This means analyzing whether time-averaging and spatial averaging are interchangeable or not *for forecasting purposes*. In absence of any other information, it was tested if using satellite lags (i.e. weighted time averaging) is the same as using a spatially averaged satellite input, without lags. The former is tested by using $p = 0$, lags $= 5$ and a small window size, and the latter is tested by using $p = 0$, lags $= 0$ and a medium window size. The baseline level of $p = 0$, lags $= 0$ and a small window size is also given, as a reference. It can be seen that the model including satellite lags shows almost no improvement from the baseline level. However, the model using a medium satellite window size reaches a significant improvement. Hence, it is clear that including spatially averaged satellite information is more useful than using time-averaged satellite information. In other words, weighted time averaging is not equivalent to spatial averaging over satellite data in terms of forecasting performance.

Summing up, the simpler best performing ARMAX-RLS model found here is that with $p = 1$, one lag in satellite data and a large spatial window (37 km $\times$ 31 km), followed closely by the medium spatial window (19 km $\times$ 16 km). It has a peak FS slightly above $+19\%$ at 40 minutes ahead ($h = 4$) and a better performance than the previously tested models for $h \geq 4$. From 10 to 30 minutes ahead its performance is similar to that of the same model but using a medium satellite albedo window, being $\simeq +1\%$ lower in the first two lead times. Its FS is $\simeq +16\%$ for most lead times between 2 and 4 hours, showing the typical



(a) With varying pixel size.

(b) With varying satellite lags and pixel size.

Figure 6: Forecasting Skill for ARMA-RLS filter using satellite albedo.

downgrade in the last hour, as seen in previously tested models. The FS of all the tested models are positive, even using the optimal smart persistence as benchmark for its calculation, with the only exception of the model using $p = 1$ but no satellite data. Based on this, it can be argued that simple ARMA-RLS models based only on ground measurements, like the $(5, 0)$ model, could be used as a more exigent performance benchmark for solar forecasting methods. Further studies on this topic should include different climates.

The closest work in the literature is the one of Dambreville et al. (2014). The comparison can be made in terms of the regular FS (provided in this work in Appendix A), but it is not straightforward due to different time scales and locations under study. Dambreville et al. use a 15 minutes times basis (the Meteosat second generation satellite time resolution) and tested their ideas using ground measurements from an urban BSRN site in Paris, France (PAL station, SIRTA Observatory). The short-term variability (σ) of the site is not provided, but one can use the absolute RMSD of the regular persistence as an indication of the sites' similarity. Table 3 provides the comparison between both works. We use the 10 minutes time basis in this work, so the 15 and 45 minutes values were linearly interpolated from Tables A.4, A.5 and A.7 in order to make the comparison possible. The AR(5) model stands for the ARMA model with $p = 5$ and $q = 0$. The AST method uses as input the 3×3 fixed pixels centered at the site's location and the AST2 method uses intercorrelation maps to decide which pixels are more useful as input for each time horizon. The convention in Table 3 for the satellite models of this work is SAT(p, lags) and the used pixel size is medium. It is observed that in the Paris site the regular persistence's RMSD starts lower than in our region, but increases more quickly, hence the solar variability regimen is not the same, although rather similar. The satellite models' FS are of similar order, but there is an important difference in their behavior for both works: for Dambreville et al. they increase with the time horizon while for the present work they have a maximum around 30 minutes. As the same behavior is observed with the AR(5) model, which does not use satellite information, we think this phenomenon is explained by the different behavior of the regular persistence. In fact, the AR(5) model is included here as a reference between both works that does not take into account the way that satellite information is used (which is different). To isolate the contribution of the satellite input it is possible to take the FS difference of each model with respect to the AR(5) model, also shown in Table 3 as 'gain'. It is observed that the AST2, SAT(1,0) and SAT(1,1) models have similar gains, around +9-14%, while for the AST model the gain is lower. The AST2 model presents a slightly better gain of $\simeq 1\%$ than that of the SAT models for the first time horizon considered (15 minutes), but this gain then decreases monotonically. On the other hand, the SAT models have a maximum gain at 30 minutes ahead of +13-14%, outperforming in $\simeq +3\%$ the AST2 model between 30 and 60 minutes ahead. The gain with respect to the AR(5) model also allows to visualize better the slight improvement obtained when including satellite lags (SAT(1,1) model vs SAT(1,0) model), of around +0.5% for these time horizons. We conclude that the proposals of the present work are a simple and effective alternative to include satellite information into solar forecasting methods and its performance is competitive with other more complex approaches.

6. Conclusions

Three things were done in this work: an analysis of smart persistence obtaining a novel benchmarking reference, a revisit on the optimal order of an ARMA model embedded in a RLS algorithm, and more importantly, a study of the impact of satellite data and its time and spatial averaging on the performance of solar forecasts made through a RLS filter approach. The resulting model is a simple alternative for including satellite information into solar forecasts and outperforms the best smart persistence, having a similar performance than other more sophisticated ways of using satellite data.

On the smart persistence analysis it was shown that the optimal value of n depends on the lead time considered. As expected, this optimal value of n grows with the forecasting horizon, but it is never equal to 1, i.e. the regular persistence. Furthermore, there is a breakpoint at a lead time of approximately 4 hours, in which comparing with (smart) persistence is not useful anymore, and comparison with climatology should be made. The optimal value of n for each time horizon defines a best smart persistence, which is used as performance benchmark.

The RLS filter is a flexible algorithm that does not need train-validation-test splits and is suitable for formulating ARMAX models, so it was used here to assess the performance impact of including satellite information to baseline models that only use ground measurements. When ignoring the exogenous part (the

Table 3: Comparison between the work of Dambreville et al. (2014) and the present work. The information from Dambreville et al. was taken from the Table 1 of their work. The 15 and 45 minutes values for this work were obtained via linear interpolation of the 10-minutes metrics. The 30 and 60 minutes were taken directly from the 10 minutes evaluation.

Dambreville et al. (2014)						
lead	**RMSD**		**FS (%)**		**gain vs AR(5)**	
time	**persist.**	**AR(5)**	**AST**	**AST2**	**AST**	**AST2**
15 mins.	94 W/m^2	+8.5	+17.0	+20.2	+8.5	+11.7
30 mins.	118 W/m^2	+15.3	+20.3	+26.3	+5.1	+11.0
45 mins.	130 W/m^2	+17.7	+20.8	+27.7	+3.1	+10.0
60 mins.	140 W/m^2	+20.7	+22.9	+29.3	+2.1	+8.6
This work						
lead	**RMSD**		**FS (%)**		**gain vs AR(5)**	
time	**persist.**	**AR(5)**	**SAT(1,0)**	**SAT(1,1)**	**SAT(1,0)**	**SAT(1,1)**
15 mins.	102 W/m^2	+8.5	+19.4	+19.6	+10.9	+11.1
30 mins.	120 W/m^2	+10.2	+23.7	+24.4	+13.5	+14.2
45 mins.	129 W/m^2	+10.0	+22.7	+23.2	+12.7	+13.2
60 mins.	137 W/m^2	+9.9	+21.3	+21.7	+11.4	+11.8

satellite input), optimal orders can be found via a grid search and, for solar irradiance data, an ARMA model with fixed $3 \leq p \leq 8$ and $q = 0$ performs almost optimally. In fact, we found that there is no value in finding the optimal p and q values for each time horizon, as the fixed parameters filter provides performance results indistinguishable from the optimal ones.

There are five remarks to be made regarding the inclusion of satellite data. (I) including satellite data removes the importance of ground measurements *as inputs*, restricting their usefulness to the first 10-minutes time horizon (although, in the present formulation, they are still needed to feedback the error signal to the RLS algorithm). (II) Adding lags in satellite data only achieves little improvements for 30-40 minutes ahead, on the FS curve peak. (III) Enlarging the spatial averaging window enhances performance: performance improvement increases quickly with the window size, but after a certain size, the improvement is restricted to the larger time horizons (higher than 1 hour ahead) at a cost of losing little performance in the first two time horizons. (IV) Enlarging the spatial averaging window moves the maximum of the FS curve in the direction of larger lead times. (V) As time-averages (including lags) does not yield the same performance improvement than spatial-averages, ergodicity does not seems to be a property of satellite albedo as input for solar forecasting.

The results presented here are valid, a priori, only for intermediate solar variability sites and regions with similar climates to the target region. Further research is required to fully understand the presented ideas for solar forecasting in various context, namely, making a simple use of satellite images. Testing these ideas for other sites and climates in the world, at least, accounting for the GOES-East satellite coverage, is part of our current work.

Acknowledgments

The authors thank the SONDA network of the INPE (Brazil) for providing the São Martinho da Serra data set. They also gratefully acknowledge the financial support given by Uruguay's National Research and Innovation Agency (ANII) under the FSE-ANII-2016-131799 grant.

A. Detailed performance metrics

In this appendix we provide the detailed performance results, including the rMBD, rRMSD and FS metrics, the latter based on the regular persistence and the optimal smart persistence, for easy comparison with other works. Tables A.4 and A.5 present the results for the endogenous models showed in Figure 4. Table A.4 also provides the rMBD and rRMSD metrics for both persistence methods. Tables A.6 and A.7 present the results for all the tested models that include space-averaged satellite albedo, that were analyzed in Figures 5 and 6.

Table A.4: Relative MBD and RMSD metrics for the persistence and the models that only use ground measurements.

	ARMA-RLS model specification					
p	1	1	5	5	**regular**	**smart**
q	0	1	0	1	**persist.**	**persist.**
lead time			**relative MBD (%)**			
10 mins	−0.3	−0.2	−0.2	−0.2	−0.1	−0.1
20 mins	−0.4	−0.4	−0.4	−0.4	−0.1	−0.2
30 mins	−0.5	−0.5	−0.5	−0.5	−0.2	−0.2
40 mins	−0.6	−0.6	−0.6	−0.6	−0.2	−0.3
50 mins	−0.7	−0.7	−0.7	−0.7	−0.3	−0.4
60 mins	−0.8	−0.8	−0.8	−0.8	−0.3	−0.5
90 mins	−1.0	−1.0	−1.0	−1.0	−0.5	−0.7
120 mins	−1.2	−1.2	−1.2	−1.2	−0.7	−0.9
150 mins	−1.3	−1.3	−1.3	−1.3	−0.9	−1.1
180 mins	−1.4	−1.4	−1.4	−1.4	−1.1	−1.2
220 mins	−1.5	−1.5	−1.6	−1.5	−1.3	−1.3
260 mins	−1.6	−1.5	−1.6	−1.6	−1.4	−1.9
lead time			**relative RMSD (%)**			
10 mins	20.3	19.5	19.4	19.3	20.9	20.5
20 mins	23.9	22.7	22.6	22.5	25.0	23.6
30 mins	25.6	24.5	24.3	24.2	27.0	25.3
40 mins	26.9	25.9	25.6	25.5	28.4	26.7
50 mins	27.9	27.1	26.7	26.7	29.7	27.9
60 mins	28.9	28.2	27.7	27.7	30.8	29.0
90 mins	31.2	30.9	30.4	30.4	33.7	32.1
120 mins	33.3	33.2	32.6	32.7	36.5	34.8
150 mins	35.1	35.1	34.6	34.6	39.0	37.4
180 mins	36.7	36.8	36.3	36.4	41.5	39.8
220 mins	38.6	38.7	38.3	38.4	44.5	42.6
260 mins	40.1	40.2	39.9	40.0	47.3	44.5

Table A.5: Forecasting skill of the models that only use ground measurements.

ARMA-RLS model specification				
p	1	1	5	5
q	0	1	0	1
lead time	**FS (%) using the regular persistence**			
10 mins	+3.2	+6.6	+7.4	+7.5
20 mins	+4.4	+9.2	+9.6	+9.8
30 mins	+5.1	+9.2	+10.0	+10.2
40 mins	+5.5	+8.9	+10.0	+10.2
50 mins	+5.9	+8.7	+10.0	+10.2
60 mins	+6.3	+8.4	+9.9	+10.1
90 mins	+7.4	+8.4	+10.0	+10.0
120 mins	+8.6	+9.0	+10.6	+10.5
150 mins	+9.9	+10.0	+11.3	+11.1
180 mins	+11.4	+11.2	+12.4	+12.1
220 mins	+13.3	+13.1	+14.0	+13.7
260 mins	+15.2	+14.9	+15.6	+15.4
lead time	**FS (%) using the optimal smart persistence**			
10 mins	+1.4	+4.9	+5.7	+5.8
20 mins	−1.2	+3.9	+4.3	+4.5
30 mins	−1.3	+3.1	+3.9	+4.2
40 mins	−0.8	+2.9	+4.0	+4.2
50 mins	−0.3	+2.7	+4.1	+4.3
60 mins	+0.5	+2.8	+4.3	+4.6
90 mins	+2.6	+3.7	+5.3	+5.3
120 mins	+4.3	+4.7	+6.3	+6.2
150 mins	+6.0	+6.1	+7.5	+7.3
180 mins	+7.6	+7.5	+8.7	+8.4
220 mins	+9.5	+9.2	+10.2	+9.9
260 mins	+10.0	+9.7	+10.4	+10.1

Table A.6: Relative MBD and RMSD metrics for the models including satellite information. $q = 0$ for all the models.

	ARMAX-RLS model specification								
window	medium	medium	medium	medium	medium	small	small	small	large
p	0	1	5	1	1	0	0	1	1
lags	0	0	0	1	5	0	5	1	1
lead time	**relative MBD (%)**								
10 mins	−0.6	−0.3	−0.3	−0.3	−0.3	−0.7	−0.6	−0.3	−0.3
20 mins	−0.6	−0.5	−0.5	−0.5	−0.5	−0.7	−0.6	−0.5	−0.4
30 mins	−0.7	−0.6	−0.6	−0.6	−0.6	−0.8	−0.7	−0.6	−0.5
40 mins	−0.7	−0.6	−0.6	−0.6	−0.6	−0.8	−0.7	−0.7	−0.6
50 mins	−0.7	−0.7	−0.7	−0.7	−0.7	−0.8	−0.7	−0.7	−0.7
60 mins	−0.8	−0.7	−0.7	−0.7	−0.7	−0.8	−0.8	−0.8	−0.7
90 mins	−0.8	−0.8	−0.8	−0.8	−0.8	−0.9	−0.8	−0.9	−0.8
120 mins	−0.8	−0.8	−0.8	−0.8	−0.8	−0.9	−0.8	−0.9	−0.8
150 mins	−0.9	−0.9	−0.8	−0.9	−0.9	−0.9	−0.9	−1.0	−0.8
180 mins	−0.9	−0.9	−0.9	−0.9	−0.9	−0.9	−0.9	−1.0	−0.9
220 mins	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.1	−1.0
260 mins	−1.2	−1.1	−1.1	−1.1	−1.1	−1.2	−1.2	−1.2	−1.1
lead time	**relative RMSD (%)**								
10 mins	19.0	17.5	17.6	17.5	17.5	19.7	19.1	17.7	17.7
20 mins	19.5	19.3	19.4	19.3	19.3	20.3	19.5	19.8	19.5
30 mins	20.6	20.6	20.6	20.4	20.4	21.6	20.5	21.3	20.5
40 mins	21.8	21.9	21.9	21.7	21.7	22.9	21.7	22.7	21.5
50 mins	23.1	23.1	23.1	23.0	23.0	24.2	23.0	24.0	22.6
60 mins	24.2	24.2	24.3	24.1	24.2	25.3	24.1	25.1	23.7
90 mins	27.1	27.2	27.2	27.1	27.1	28.1	27.1	28.0	26.7
120 mins	29.7	29.7	29.8	29.7	29.7	30.5	29.7	30.5	29.3
150 mins	31.9	32.0	32.0	32.0	32.0	32.6	32.0	32.6	31.6
180 mins	33.9	34.0	34.0	34.0	34.0	34.5	34.0	34.5	33.6
220 mins	36.3	36.3	36.4	36.3	36.4	36.7	36.3	36.8	36.0
260 mins	38.2	38.3	38.4	38.3	38.4	38.6	38.3	38.7	38.1

Table A.7: Forecasting skill for the models including satellite information. $q = 0$ for all the models.

ARMAX-RLS model specification									
window	medium	medium	medium	medium	medium	small	small	small	large
p	0	1	5	1	1	0	0	1	1
lags	0	0	0	1	5	0	5	1	1
lead time	**FS (%) using the regular persistence**								
10 mins	+9.0	+16.2	+16.1	+16.3	+16.3	+5.9	+8.8	+15.2	+15.4
20 mins	+21.9	+22.6	+22.4	+22.9	+22.9	+18.8	+21.8	+20.8	+22.1
30 mins	+23.8	+23.7	+23.6	+24.4	+24.3	+20.0	+24.1	+21.1	+24.2
40 mins	+23.3	+23.1	+23.0	+23.7	+23.7	+19.3	+23.6	+20.0	+24.3
50 mins	+22.4	+22.3	+22.1	+22.7	+22.7	+18.6	+22.7	+19.2	+23.8
60 mins	+21.4	+21.3	+21.2	+21.7	+21.6	+17.9	+21.6	+18.4	+22.9
90 mins	+19.5	+19.4	+19.3	+19.6	+19.5	+16.7	+19.5	+16.9	+20.9
120 mins	+18.6	+18.5	+18.4	+18.6	+18.5	+16.4	+18.6	+16.5	+19.7
150 mins	+18.0	+17.9	+17.8	+17.9	+17.8	+16.3	+17.9	+16.3	+18.9
180 mins	+18.2	+18.1	+18.0	+18.1	+17.9	+16.8	+18.0	+16.8	+18.9
220 mins	+18.5	+18.4	+18.2	+18.4	+18.2	+17.5	+18.4	+17.4	+19.0
260 mins	+19.1	+19.0	+18.9	+19.0	+18.8	+18.3	+18.9	+18.2	+19.5
lead time	**FS (%) using the optimal smart persistence**								
10 mins	+7.4	+14.7	+14.6	+14.8	+14.8	+4.2	+7.1	+13.7	+13.9
20 mins	+17.3	+18.0	+17.8	+18.4	+18.4	+14.0	+17.2	+16.1	+17.5
30 mins	+18.7	+18.6	+18.4	+19.3	+19.2	+14.6	+19.0	+15.8	+19.1
40 mins	+18.2	+18.0	+17.9	+18.6	+18.6	+14.0	+18.5	+14.7	+19.3
50 mins	+17.3	+17.2	+17.0	+17.7	+17.6	+13.3	+17.6	+13.9	+18.8
60 mins	+16.6	+16.4	+16.3	+16.8	+16.7	+12.8	+16.8	+13.3	+18.1
90 mins	+15.4	+15.2	+15.1	+15.4	+15.3	+12.4	+15.4	+12.6	+16.8
120 mins	+14.7	+14.6	+14.5	+14.7	+14.6	+12.4	+14.7	+12.5	+15.9
150 mins	+14.5	+14.4	+14.3	+14.4	+14.3	+12.7	+14.4	+12.7	+15.4
180 mins	+14.7	+14.6	+14.5	+14.6	+14.5	+13.3	+14.6	+13.3	+15.4
220 mins	+14.9	+14.8	+14.6	+14.7	+14.6	+13.8	+14.7	+13.7	+15.4
260 mins	+14.1	+14.0	+13.9	+14.0	+13.8	+13.3	+13.9	+13.2	+14.5

References

Aguiar, L. M., Pereira, B., David, M., Díaz, F., & Lauret, P. (2015). Use of satellite data to improve solar radiation forecasting with bayesian artificial neural networks. *Solar Energy*, *122*, 1309 – 1324. doi:https://doi.org/10.1016/j.solener.2015.10.041.

Aguiar, L. M., Pereira, B., Lauret, P., Díaz, F., & David, M. (2016). Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renewable Energy*, *97*, 599–610. doi:https://doi.org/10.1016/j.renene.2016.06.018.

Alonso-Suárez, R. (2017). *Estimación del recurso solar en Uruguay mediante imágenes satelitales*. Ph.D. thesis Facultad de Ingeniería, Universidad de la República. Available at: https://hdl.handle.net/20.500.12008/20200.

Alonso-Suárez, R., Abal, G., Siri, R., & Musé, P. (2012). Brightness-dependent Tarpley model for global solar radiation estimation using GOES satellite images: application to Uruguay. *Solar Energy*, *86*, 3205–3215. doi:10.1016/j.solener.2012.08.012.

Box, G., & Jenkins, G. (1970). *Time Series Analysis Forecasting And Control* volume 3. Holden-Day.

Bright, J. M., Killinger, S., Lingfors, D., & Engerer, N. A. (2018). Improved satellite-derived PV power nowcasting using real-time power data from reference PV systems. *Solar Energy*, *168*, 118–139.

Coimbra, C. F., Kleissl, J., & Marquez, R. (2013). Chapter 8 - Overview of solar-forecasting methods and a metric for accuracy evaluation. In J. Kleissl (Ed.), *Solar Energy Forecasting and Resource Assessment* (pp. 171 – 194). Boston: Academic Press. doi:https://doi.org/10.1016/B978-0-12-397177-7.00008-5.

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. doi:https://doi.org/10.1016/j.renene.2014.07.012.

David, M., Ramahatana, F., Trombe, P., & Lauret, P. (2016). Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. *Solar Energy*, *133*, 55–72. doi:https://doi.org/10.1016/j.solener.2016.03.064.

Diagne, M., David, M., Lauret, P., Boland, J., & Schmutz, N. (2013). Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, *27*, 65 – 76. doi:https://doi.org/10.1016/j.rser.2013.06.042.

Hammer, A., Heinemann, D., Lorenz, E., & Lückehe, B. (1999). Short-term forecasting of solar radiation: a statistical approach using satellite data. *Solar Energy*, *67*, 139–150.

Harty, T. M., Holmgren, W. F., Lorenzo, A. T., & Morzfeld, M. (2019). Intra-hour cloud index forecasting with data assimilation. *Solar Energy*, *185*, 270–282.

Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. *Artif. Intell.*, *17*, 185–203. doi:10.1016/0004-3702(81)90024-2.

Jimenez, P. A., Hacker, J. P., Dudhia, J., Haupt, S. E., Ruiz-Arias, J. A., Gueymard, C. A., Thompson, G., Eidhammer, T., & Deng, A. (2016). Wrf-solar: Description and clear-sky assessment of an augmented nwp model for solar power prediction. *Bulletin of the American Meteorological Society*, *97*, 1249–1264. doi:10.1175/BAMS-D-14-00279.1.

Kühnert, J., Lorenz, E., & Heinemann, D. (2013). Chapter 11 - Satellite-based irradiance and power forecasting for the german energy market. In J. Kleissl (Ed.), *Solar Energy Forecasting and Resource Assessment* (pp. 267–297). Boston: Academic Press. doi:http://dx.doi.org/10.1016/B978-0-12-397177-7.00011-5.

Lara-Fanego, V., Ruiz-Arias, J., Pozo-Vázquez, D., Santos-Alamillos, F., & Tovar-Pescador, J. (2012). Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain). *Solar Energy*, *86*, 2200–2217. Progress in Solar Energy 3.

Lauret, P., Voyant, C., Soubdhan, T., David, M., & Poggi, P. (2015). A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. *Solar Energy*, *112*, 446–457. doi:https://doi.org/10.1016/j.solener.2014.12.014.

Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Qu, Z., Wald, L., Homscheidt, M. S., & Arola, A. (2013). McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmospheric Measurement Techniques, European Geosciences Union*, *6*, 2403–2418. doi:10.5194/amt-6-2403-2013.

Lorenz, E., Hammer, A., & Heinemann, D. (2004). Short term forecasting of solar radiation based on satellite data. In *EUROSUN2004 (ISES Europe Solar Congress)* (pp. 841–848). Freiburg, Germany.

Lorenz, E., Hurka, J., Heinemann, D., & Beyer, H. G. (2009a). Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *IEEE Journal of selected topics in applied earth observations and remote sensing*, *2*, 2–10.

Lorenz, E., Kühnert, J., & Heinemann, D. (2012). Short term forecasting of solar irradiance by combining satellite data and numerical weather predictions. In *Proceedings of the 27th European PV Solar Energy Conference (EU PVSEC), Frankfurt, Germany* (p. 44014405). volume 2428.

Lorenz, E., Remund, J., Müller, S. C., Traunmüller, W., Steinmaurer, G., Pozo, D., Ruiz-Arias, J., Fanego, V. L., Ramirez, L., Romeo, M. G. et al. (2009b). Benchmarking of different approaches to forecast solar irradiance. In *24th European photovoltaic solar energy conference* (pp. 21–25). Hamburg, Germany.

Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2* IJCAI'81 (pp. 674–679). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Marchesoni-Acland, F., Lauret, P., Gómez, A., & Alonso-Suárez, R. (2019). Analysis of ARMA solar forecasting models using ground measurements and satellite images. In *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)* (pp. 2445–2451). doi:10.1109/PVSC40753.2019.8980821.

Marquez, R., Pedro, H. T., & Coimbra, C. F. (2013). Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. *Solar Energy*, *92*, 176–188.

Mathiesen, P., & Kleissl, J. (2011). Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states. *Solar Energy*, *85*, 967–977. doi:https://doi.org/10.1016/j.solener.2011.02.013.

McArthur, L. (2005). *Baseline Surface Radiation Network Operations Manual*. Technical Report WCRP-121/ WMO TD-No. 1274 World Climate Research Programme – WMO.

Ohmura, A., Dutton, E. G., Forgan, B., Fröhlich, C., Gilgen, H., Hegner, H., Heimo, A., König-Langlo, G., McArthur, B., Müller, G., Philipona, R., Pinker, R., Whitlock, C. H., Dehne, K., & Wild, M. (1998). Baseline surface radiation network (BSRN/WCRP): New precision radiometry for climate research. *Bulletin of the American Meteorological Society*, *79*, 2115–2136. doi:10.1175/1520-0477(1998)079<2115:BSRNBW>2.0.CO;2.

Pedro, H. T., & Coimbra, C. F. (2012). Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, *86*, 2017–2028. doi:https://doi.org/10.1016/j.solener.2012.04.004.

Peel, M. C., Finlayson, B. L., & Mcmahon, T. A. (2007). Updated world map of the köppen-geiger climate classification. *Hydrology and Earth System Sciences Discussions*, *11*, 1633–1644.

Peng, Z., Yoo, S., Yu, D., & Huang, D. (2013). Solar irradiance forecast system based on geostationary satellite. In *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)* (pp. 708–713). doi:10.1109/SmartGridComm.2013.6688042.

Perez, R., Beauharnois, M., Hemker, K., Kivalov, S., Lorenz, E., Pelland, S., Schlemmer, J., & Van Knowe, G. (2011). Evaluation of numerical weather prediction solar irradiance forecasts in the US. In *Proc. ASES Annual Conference*.

Perez, R., David, M., Hoff, T. E., Jamaly, M., Kivalov, S., Kleissl, J., Lauret, P., & Perez, M. (2016). Spatial and temporal variability of solar energy. *Foundations and Trends® in Renewable Energy*, *1*, 1–44. doi:10.1561/2700000006.

Perez, R., & Hoff, T. (2013). Solar anywhere forecasting. *Solar Energy Forecasting and Resource Assessment. Elsevier*, .

Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., & Vignola, F. (2002). A new operational model for satellite-derived irradiances: description and validation. *Solar Energy*, *73*, 307–317. doi:10.1016/S0038-092X(02)00122-6.

Perez, R., Lorenz, E., Pelland, S., Beauharnois, M., Knowe, G., Jr., K. H., Heinemann, D., Remund, J., Müller, S., Traunmüller, W., Steinmauer, G., Pozo, D., Ruiz-Arias, J., Lara-Fanego, V., Ramirez-Santigosa, L., Gaston-Romero, M., & Pomares, L. (2013). Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. *Solar Energy*, *94*, 305–326.

Qu, Z., Oumbe, A., Blanc, P., Espinar, B., Gesell, G., Gschwind, B., Klüser, L., Lefèvre, M., Saboret, L., Schroedter-Homscheidt, M., & Wald, L. (2017). Fast radiative transfer parameterisation for assessing the surface solar irradiance: The Heliosat-4 method. *Meteorologische Zeitschrift*, *26*, 33–57. doi:10.1127/metz/2016/0781.

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

REN21 (2019). *Renewables 2019: Global Status Report*. Technical Report Renewable Energy Policy Network for the 21st Century 15 Rue de Milan, 75441 Paris CEDEX 09, France. URL: http://www.ren21.net/.

Rigollier, C., Lefevre, M., & Wald, L. (2004). The method Heliosat-2 for deriving shortwave solar radiation from satellite images. *Solar Energy*, *77*, 159–169. doi:10.1016/j.solener.2004.04.017.

Voyant, C., & Notton, G. (2018). Solar irradiation nowcasting by stochastic persistence: A new parsimonious, simple and efficient forecasting tool. *Renewable and Sustainable Energy Reviews*, *92*, 343 – 352. doi:https://doi.org/10.1016/j.rser.2018.04.116.

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., & Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, *105*, 569–582. doi:https://doi.org/10.1016/j.renene.2016.12.095.

Yagli, G. M., Yang, D., & Srinivasan, D. (2019). Automatic hourly solar forecasting using machine learning models. *Renewable and Sustainable Energy Reviews*, *105*, 487–498.

Yang, D. (2019). Making reference solar forecasts with climatology, persistence, and their optimal convex combination. *Solar Energy*, *193*, 981–985.

Yang, D., Kleissl, J., Gueymard, C. A., Pedro, H. T., & Coimbra, C. F. (2018). History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. *Solar Energy*, *168*, 60–101. doi:https://doi.org/10.1016/j.solener.2017.11.023. Advances in Solar Resource Assessment and Forecasting.

Yang, D., & Perez, R. (2019). Can we gauge forecasts using satellite-derived solar irradiance? *Journal of Renewable and Sustainable Energy*, *11*, 023704.