Short-term Solar Irradiation forecasting based on Dynamic Harmonic Regression

Juan R. Trapero[a,*], Nikolaos Kourentzes[b], A. Martin[a]

[a]*Universidad de Castilla-La Mancha*
Departamento de Administracion de Empresas, Ciudad Real 13071, Spain
[b]*Lancaster University*
Department of Management Science, LA1 4YX, UK

Abstract

Solar power generation is a crucial research area for countries that have high dependency on fossil energy sources and is gaining prominence with the current shift to renewable sources of energy. In order to integrate the electricity generated by solar energy into the grid, solar irradiation must be reasonably well forecasted, where deviations of the forecasted value from the actual measured value involve significant costs. The present paper proposes a univariate Dynamic Harmonic Regression model set up in a State Space framework for short-term (1 to 24 hours) solar irradiation forecasting. Time series hourly aggregated as the Global Horizontal Irradiation and the Direct Normal Irradiation will be used to illustrate the proposed approach. This method provides a fast automatic identification and estimation procedure based on the frequency domain. Furthermore, the recursive algorithms applied offer adaptive predictions. The good forecasting performance is illustrated with solar irradiance measurements collected from ground-based weather stations located in Spain. The results show that the Dynamic Harmonic Regression achieves the lowest relative Root Mean Squared Error; about 30% and 47% for the Global and Direct irradiation components, respectively, for a forecast horizon of 24 hours ahead.

Keywords: Solar Irradiation, Forecasting, Dynamic Harmonic Regression,

---

[*]Corresponding author.
  Email addresses: `juanramon.trapero@uclm.es` (Juan R. Trapero),
`n.kourentzes@lancaster.ac.uk` (Nikolaos Kourentzes), `amartingarcia77@gmail.com`
(A. Martin)

1. Introduction

The increasing investment in renewable energy is essential to guarantee immediate answers to both the high and fluctuating prices of crude oil and the energy supplies diversification. In some countries like Spain, solar power generation is becoming a research area of paramount importance. In this sense, reliable short-term forecast information of the solar radiation components is required to achieve an efficient use of fluctuating energy output from photovoltaic (PV), concentrated-photovoltaic (CPV) and solar thermal (CSP) power plants. In particular, Global Horizontal Irradiation (GHI) is important for photovoltaic applications, whereas Direct Normal Irradiation (DNI) is required for Concentrated Solar Power applications, and this is often done through forecasting methodologies.

Electricity companies and transmission systems operators need to know the expected load profiles 24 hours in advance, where forecast errors in the fluctuating input from solar systems can lead to significant costs. Krass et al. [1] carried out a simulation study to quantify the costs of forecasting deviations for a concentrating solar power system in the Spanish electricity market, where improved forecasting techniques reduced the penalties compared to the persistence case by 47.6 %.

The diversity of solar radiation forecasting methodologies can be classified according to the input data and the objective forecasting horizon [2]. For instance, Numerical Weather Prediction models (NWP), which are based on physical laws of motion and conservation of energy that govern the atmospheric air flow, are operationally used to forecast the evolution of the atmosphere from about 6 hours onward. Although NWP models are powerful tools to forecast solar radiation at places where ground data are not available, many near-surface physical processes occur within a single grid box and are too complex to be represented and solved by equations. Thus, NWP models cannot successfully resolve local processes smaller than the model resolution.

Satellite-derived solar radiation images are a useful tool for quantifying solar irradiation at ground surface for large areas, but they need to set an accurate radiance value under clear sky conditions and under dense cloudiness from every pixel and every image [3].

Ground-based observations, such as sky imaging techniques, are used to fill the intra-hour and sub-kilometer forecasting gap of NWP models regarding cloud cover over solar power plants. Nonetheless, operational forecast horizons are limited to very short-term, ranging from 5 to 25 minutes ahead [4].

These limitations have placed time series analysis as the dominant methodology for short-term forecasting horizons from 5 min up to 6 hours [5]. Common features of solar radiation time series as intermittence, high sampling frequency and non-stationarity have contributed to the proliferation of multiple statistical forecasting techniques. A recent state of the art review can be found in [2]. In general terms, two main modelling approaches are identified depending on how the method deals with non stationarity, i.e, trend and seasonality. On the one hand, a deterministic approach based on solar geometry is used to remove the observed seasonality by means of the clearness index (k) defined as the ratio of irradiance at ground level with respect to extraterrestrial irradiance [6]. If additional information on atmospheric conditions is available, clear sky models can be used to estimate the global irradiance in clear sky conditions [7]. Then, the clear-sky index (k^*) can be calculated as the ratio of irradiance at ground level to clear-sky irradiance. However, some authors argue that such indices are mostly random and thus, they are not adequate for learning algorithms [8]. From the authors point of view, the adequacy of using indices rather than solar irradiance time series requires further research, and so far, both alternatives are valid. However, in this work, we opt for using solar irradiance time series to make sure that all the forecast errors come from the forecasting technique and they are not the result of modeling errors when estimating the indices.

On the other hand, the second approach consists of removing the trend and seasonality components to make the time series stationary. Typical de-seasonalizing methods employed are: Fourier series [9, 10], high order polynomial models [11], cosine function models [12], Gaussian models [13], and a seasonal-trend decomposition procedure [14]. Dong et al. [10] showed by using the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) stationary statistical test that Fourier series had much higher probability of obtaining stationarity. Once the seasonality has been removed by the Fourier regression, the next step is to model the residuals by either autorregresive processes [9] or exponential smoothing algorithms [10].

This work investigates the forecasting performance of the Dynamic Harmonic Regression (DHR) model. The DHR is an extension of the typical

3

harmonic regression, where the coefficients are time-varying instead of constant [15]. In other words, models described in [9] and [10] can be summarized in two steps. Firstly, they remove the seasonality by using a constant coefficients harmonic regression. Then, since the traditional harmonic regression may not capture the dynamics of the data, another step based on the residuals modelling is required to compensate potential biases on the harmonic regression constant coefficients. Note that these coefficients can be estimated by using typical least squares or maximum likelihood procedures. In turn, the DHR relaxes the assumption of constant coefficients integrating the processes of forecasting, interpolation and seasonal adjustments into a single recursive framework based on the Kalman Filter and the Fixed Interval Smoothing algorithms [16]. The DHR recursive nature allows to handle efficiently changes of amplitude and phase, as it commonly happens in the solar irradiation time series. Therefore, the use of DHR does not require a second residuals modelling step as it occurs in [9] and [10].

The DHR is a particularization of a more general type of models called Unobserved Components (UC) models based on the State Space (SS) framework. The literature on this topic is immense. The reader is referred to the seminal works by [15, 17, 18]. It should be noted that, although UC models have been initially proposed in solar irradiation forecasting in [5], the DHR model incorporates certain differences that can improve the forecasting accuracy of previous works. Essentially, unlike the UC model presented in [5] the DHR utilizes Fixed Interval Smoothing (in addition to Kalman Filter) and the estimation of model hyper-parameters is accomplished in the frequency domain. In this sense, the Fixed Interval Smoothing allows optimal signal extraction, smoothing and interpolation over gaps in the data. Furthermore, the hyper-parameter estimation in the frequency domain provides objective functions much better defined when the time series are clearly seasonal. In contrast, typical Maximum Likelihood estimation in the time domain may fail when the number of parameters to be estimated is high [15], as it happens in the present case study.

Although the DHR model has been successfully employed in other related applications as electricity price and load forecasting [19, 20], this is the first time that this model is proposed to forecast solar irradiation. In order to illustrate the performance of the proposed model, hourly Global Horizontal Irradiation (GHI) and Direct Normal Irradiation (DNI) have been selected for this study.

The article is organized as follows: Section 2 describes the Dynamic Har-

monic Regression. Section 3 introduces the benchmark models that are used to evaluate the performance of the DHR. Section 4 presents the case study which includes the description of the study area, the observational data and the experimental results. Finally, main conclusions are drawn in Section 5.

2. Dynamic Harmonic Regression

A DHR can be expressed as an UC model as in equation (1), where the GHI or DNI time series (y_t) is decomposed as the sum of a long term trend (T_t), a seasonal component (S_t) and an irregular component (e_t). The right expression in (1) is the equivalent harmonic regression with time-varying coefficients, where the trend T_t is associated with the frequency being set equal to zero $(k = 0)$. The seasonal component S_t results from the addition of $P/2$ harmonics $(k = 1, 2, \ldots, P/2)$, where P is the fundamental period of the seasonal component, i.e. the number of observations per cycle. In our case study $P = 24$ since the data is hourly and there is a clear daily cycle. The irregular component e_t represents any stochastic and unpredictable temporal variations in y_t that have not been explained by all the other components. Finally, e_t is assumed to be a Gaussian random noise signal with zero mean and constant variance (σ^2).

$$y_t = T_t + S_t + e_t = \sum_{k=0}^{P/2} [a_{k,t} \cos(\omega_k t) + b_{k,t} \sin(\omega_k t)] + e_t \qquad (1)$$

The model is completed when the dynamic behavior of the stochastic trend and seasonal subcomponents are specified. There is a wide range of options available in the literature to do this [17, 18, 21] and the ones favored here are explained briefly below.

Equation (2) shows the model for the trend. Formally it is usually called a Local Linear Trend (LLT), where $a_{0,t+1}^*$ stands for an additional unobserved state necessary for the specification of the trend T_t ; and $w_{0,t}$ and $w_{0,t}^*$ are random Gaussian noises, independent of each other with zero mean and certain variances σ_0^2 and σ_0^{2*}, respectively.

$$\begin{pmatrix} a_{0,t+1} \\ a_{0,t+1}^* \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_{0,t} \\ a_{0,t}^* \end{pmatrix} + \begin{pmatrix} w_{0,t} \\ w_{0,t}^* \end{pmatrix} \qquad (2)$$

Each of the seasonal sub-components $a_{k,t}$ and $b_{k,t}$ $(k = 1, 2, \ldots, P/2)$ in equation (1) may be modeled as random walks, where $a_{k,t}^*$ is an additional

state necessary to write the trigonometric representation; $w_{k,t}$ and $w_{k,t}^*$ are independent random noises with zero mean and a common variance σ_k^2 (and $\sigma_k^2 \neq \sigma_j^2$ for any $k \neq j$):

$$\begin{pmatrix} a_{k,t+1} \\ a_{k,t+1}^* \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_{k,t} \\ a_{k,t}^* \end{pmatrix} + \begin{pmatrix} w_{k,t} \\ w_{k,t}^* \end{pmatrix} \tag{3}$$

The full UC model (1) is built by assembling equations (1), (2) and (3). The result of this is a State Space system that may be treated in a standard way, where equation (1) is the observation equation, and the block concatenation of (2) and (3) for $k = 1, 2, \ldots, P/2$ provides the State or Transition equation. It is well known that certain recursive algorithms, namely the Kalman Filter (KF) [22] and the Fixed Interval Smoothing algorithm (FIS) [23], produce the optimal estimates of the state vector by minimizing the Mean Squared Error.

The states in our model are $a_{0,t}$; $a_{0,t}^*$; $a_{k,t}, b_{k,t}$; and $a_{k,t}^*, b_{k,t}^*$, with $k = 1, 2, \ldots, P/2$. From all these, only $T_t = a_{0,t}$ and

$$S_t = \sum_{k=1}^{P/2} [a_{k,t} \cos(\omega_k t) + b_{k,t} \sin(\omega_k t)] \tag{4}$$

have a physical meaning as the trend and the seasonal components respectively.

The application of the recursive KF/FIS algorithms requires knowledge of all the system matrices. Most of the elements of these matrices are known in many applications, but there are always a number of them that are unknown (often called hyper-parameters) and must be estimated by efficient methods. The hyper-parameters in equations (1)-(3) are all the noise variances, i.e. σ_0^2; σ_0^{2*}; σ_k^2 with $k = 1, 2, \ldots, P/2$; and σ^2, the irregular/innovations variance.

In many applications of single output UC systems, the number of unknown parameters may be reduced in one element, by normalizing all the variances by the innovations variance. In this way, Noise Variance Ratios (NVR) are usually defined as σ_0^2/σ^2; σ_0^{2*}/σ^2; σ_k^2/σ^2 with $k = 1, 2, \ldots, P/2$. When this change is carried out, the KF and FIS algorithms have to be updated accordingly [15, 18, 21]. Although Maximum Likelihood (ML) in time domain is the most common used approach, because of its good theoretical properties [18], ML might fail when the number of parameters to be estimated is high. The main reason is that the likelihood surface is very flat or multimodal around

the optimum. Fortunately, for series with a marked periodic behavior, the frequency domain offers an alternative framework in which objective functions are much better defined than in time domain, even for high dimensional models. The frequency domain estimation of hyper-parameters and the KF/FIS algorithms are implemented in the MATLAB computer toolbox called CAPTAIN [24], available at `http://www.es.lancs.ac.uk/cres/captain/`.

3. Benchmark models

In this section, we briefly describe the models considered as benchmarks, against which the DHR model will be evaluated.

3.1. Persistence model

Regarding the problem of solar irradiation forecasting, the most extended technique used to contrast the performance of new models is the persistence model [25], where the forecast always equal to the last known data point. The persistence model is also known in the forecasting literature as the Naïve model or the Random Walk, [26]. Essentially, this model is a particular case of the Single Exponential Smoothing (SES) shown in (5) when $\alpha = 1$, i.e:

$$F_{t+m} = F_t + \alpha(y_t - F_t), \tag{5}$$

where m is the forecasting horizon. F_t and y_t stand for the forecast value at time t and the actual value respectively. SES updates the previous forecast by weighting the last forecast error, where α is a constant between 0 and 1 [27]. Hereafter, the persistence model will be referred to as Naïve.

3.2. Seasonal persistence model

Since the solar irradiance data is strongly seasonal, the persistence model for longer forecasting horizons than 1 hour can be modified to the 24 hours persistence, where:

$$F_{t+m} = y_{t-24+m} \tag{6}$$

This model essentially carries forward the last 24 observed hours into the forecast. This model will be named hereafter as S.Naïve (Seasonal Naïve), to mark its seasonal nature.

3.3. Exponential Smoothing in a State Space framework

SES method is adequate when the level is the most important component to explain the time series pattern. When other typical components as trend or seasonality are present, as it happens in both the DNI and GHI variables, the previous method should be modified. The well-known Holt-Winters method incorporates these components [28, 29]. Exponential smoothing has been extended to a probabilistic setting by Hyndman et al. [30], who developed a State Space formulation for the exponential smoothing family. Following the standard naming of the method in the literature, we will refer to it as ETS. This methodology has been already applied to the solar irradiance forecasting problem using solar irradiance measurements from Singapore and South Colorado at very short-term (5 minutes interval) [10].

3.4. ARIMA models

Box et al. [31] propose a general framework based on a Seasonal AutoRegressive Integrated Moving Average (ARIMA) process of order $(p, d, q) \times (P, D, Q)_s$ to model stationary and nonstationary time series. The process can be expressed by:

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D y_t = \theta(B)\Theta(B^s)a_t, \qquad (7)$$

where y_t is an observable time series (solar irradiance) and a_t is a white noise process having mean zero and variance σ_a^2. The backward shift operator is denoted by $By_t = y_{t-1}$. The non-seasonal Autoregressive and Moving Average operators are defined by $\phi(B)$ and $\theta(B)$ polynomials of order p and q respectively, and d denotes the order of differencing that is required to make the time series stationary, such as:

$$
\begin{aligned}
\phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, & (8)\\
\theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q. & (9)
\end{aligned}
$$

The seasonal ARIMA part is represented by $\Phi(B^s)$ and $\Theta(B^s)$ polynomials of order P and Q respectively and D denotes the order of seasonal differencing, where s is the number of periods per season.

ARIMA models can be difficult to identify, in particular, when the process is a mixture of AR and MA structures [26]. In order to choose the best ARIMA model some criteria that select the model that gives the largest

value for the likelihoods are employed. One of the most used criterion is the Akaike Information Criterion (AIC) [32]. The automatic identification procedure consists of selecting the best ARIMA model from a full range of possibilities according to the AIC. The preferred model is the one with the minimum AIC value. The exact identification procedure followed here is described in detail in [33].

Note that ARIMA models have been tested previously in [5], [7], [10] and [25]. Nonetheless, it should be noted that the identification procedure employed in the literature is not always the same. For instance, in [5] the ARIMA model is the result of considering both the Akaike [32] and Schwarz Information Criteria [34]; in [7] the models are obtained by analysing the residual auto-correlogram; in [10] it is obtained by minimizing the AIC; and in [25] is based on an AutoRegressive process from order 1 up to order 10th.

In this work, ETS and ARIMA are built using the forecast package [35] that it is implemented in the R statistical package [36] and it can be freely downloaded at `http://cran.r-project.org/web/packages/forecast/index.html`.

4. Case study

4.1. Dataset description

Time series of solar irradiance data are used to validate the DHR reliability as a forecasting tool at short-term periods, from 1 to 24 hours. All data used in this study were provided by the Spanish Institute for Concentration Photovoltaics Systems (ISFOC), located in Ciudad Real in the region of Castilla-La Mancha (Spain). ISFOC has installed 1,1 MW of Concentrated-Photovoltaic Energy (CPV) power plants and 5 automatic weather stations to measure solar resource.

Solar irradiance measurements were recorded every 1 minute by a set of solar sensors such as thermopile pyranometers and pyrheliometers, which complied with the international standards of Baseline Surface Radiation Network (BSRN) [37]. Global horizontal irradiance (GHI) is the total solar radiation on a horizontal surface and is the sum of the direct horizontal irradiance (DHI) plus the diffuse horizontal irradiance (DiffHI) and the Albedo. Diffuse irradiance refers to all the solar radiation scattered from the sky. Theoretical definition of Direct Normal Irradiance (DNI) is the solar radiation coming directly from the solar disc plus some circumsolar irradiance within approximately 2.5 degrees of the Sun center. Albedo is the fraction of the incident

solar radiation reflected and scattered by the ground. Global and diffuse irradiance were measured by a first-class pyranometer. DNI data were measured by a first-class pyrheliometer.

Hourly series of solar irradiation data (Whm^{-2}) were constructed for this study, aggregating 1-minute ground-based solar irradiance data (Wm^{-2}), which were recorded between January 2009 and December 2011 by the weather station that ISFOC has sited at $38.67\,^{\circ}$ N, $4.15\,^{\circ}$ W, 687 m (asl). Hereafter, the models proposed in the previous sections will use hourly aggregated solar irradiation as input data.

Our dataset contains 26,280 observations. Figure 1 depicts two months of solar irradiation hourly aggregated. Upper and lower plots show the GHI (solid line) and DNI (dashed line) measured in a winter month (February) and a summer month (August) of the same year (2011), respectively. It is interesting to note that a seasonal component can be observed in each month, apart from the daily cycle. In this sense, the amplitude of the seasonality in August is almost constant whereas the amplitude variability in February is more evident. In fact, this variability is more noticeable for the DNI due to its higher sensitivity to atmosphere conditions.

During the years 2009 and 2011, the mean of DNI and GHI was 223 and 192 Wh/m^2 respectively. The standard deviation of the DNI and GHI was 315 and 254 Wh/m^2 respectively. On average, the mean and standard deviation of the DNI were higher than the GHI.

4.2. Experimental design

In order to test the aforementioned models, a predictive empirical experiment is carried out. The last year of data (2011) is reserved as a hold-out sample and it is used for evaluating the different forecasting models. The experiment design proposed here is exhaustive given that models are tested for a whole year rather than testing the models only for a particular month. That is important because the forecast errors are not expected to be constant throughout the year. A rolling origin evaluation experiment is designed as follows. Once the forecast is made, the forecast origin is moved one hour ahead until the complete year of the hold-out sample is exhausted. Note that the experiment considers a multi-step forecasting horizon ranging from 1 hour to 24 hours ahead. A detailed description of cross-validation experiments based on rolling origins with multi-step forecasting horizons is available in [26] and [38].
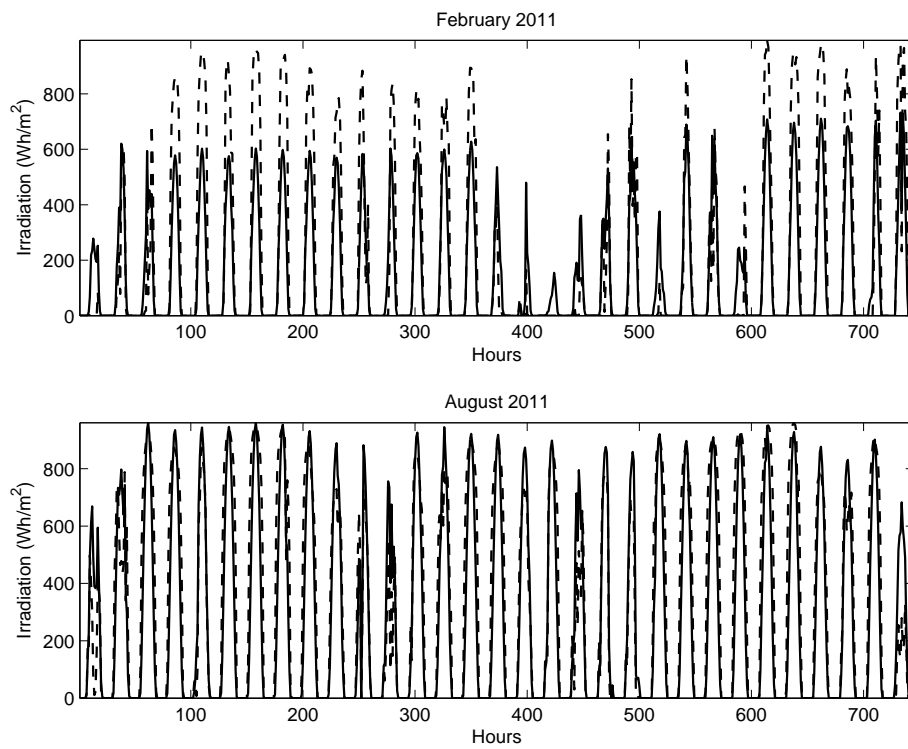
Figure 1: Hourly solar irradiation corresponding to February 2011 (upper plot) and August 2011 (lower plot). GHI and DNI are depicted in a solid and dashed line, respectively

The error metrics employed to compare the models are: i) relative Mean Biased Error (rMBE); and ii) relative Root Mean Squared Error (rRMSE). The calculations are carried out as follows, let e_t be the forecast error given by:

$$e_t = (y_t - F_t), \qquad t = 1, \ldots, T \tag{10}$$

where y_t and F_t stand for the actual value and the forecast, respectively, at time t. The rMBE and rMSE are defined such as:

$$rMBE = \frac{\sum_{t=1}^{T} e_t / T}{\bar{y}} \cdot 100 \tag{11}$$

$$rRMSE = \frac{\sqrt{\sum_{t=1}^{T} (e_t)^2 / T}}{\bar{y}} \cdot 100 \tag{12}$$

Here, $\bar{y}$ is the mean of the actual values in the hold-in sample. Furthermore, in the reference [25], the improvement over persistence is defined as follows:

$$Improvement = \left(1 - \frac{rRMSE_m}{rRMSE_p} \right) \cdot 100 \tag{13}$$

where $rRMSE_m$ and $rRMSE_p$ stand for the relative Root Mean Square Error for the proposed model and the S. Naïve model, respectively. Note that the S. Naïve model has been chosen instead of the Naïve model given that it is a more appropriate benchmark.

4.3. Results

Table 1 provides a summary of the forecast bias (rMBE) and error (rRMSE) of the benchmark models and DHR for both GHI and DNI. The figures present the average performance over 1 to 24 hours ahead forecasts. Focusing on GHI, all methods improve upon the Naïve in terms of both rMBE and rRMSE. This happens because the Naïve model does not incorporate the seasonality component and thus, it is not an adequate benchmark unless, the data is previously deseasonalized by means of either a deterministic or statistical approach as described in the introduction. If not, the suggested benchmark is the seasonal persistence model (S. Naïve), which is very effective, with only the DHR having lower rRMSE. In terms of bias all S.Naïve, ETS and DHR exhibit less than 1% bias, where the ETS achieves the lowest rMBE closely followed by S.Naïve and DHR. Note that on average ETS overforecast and the other two underforecast the GHI.

12

Table 1: Average rMBE and rRMSE for GHI and DNI.

Model	GHI		DNI	
	rMBE %	rRMSE %	rMBE %	rRMSE %
Naïve	38.14	94.58	40.85	83.64
S.Naïve	0.19	32.48	3.67	50.87
ETS	-0.18	34.18	4.65	49.09
ARIMA	1.89	37.34	11.61	112.67
DHR	0.21	29.66	3.82	46.79

With regards to DNI the findings are similar, with the exception of the rather poor performance of ARIMA. It should be noted that the literature about ARIMA models employed to forecast DNI time series is scarce in order to contrast these results. An explanation to such a poor performance is due to the DNI high variability as a consequence of the DNI sensitivity to atmosphere conditions. Thus, automatic ARIMA identification techniques fail to capture the structure of the DNI time series, which results in a high forecasting error. On the other hand, DHR exhibits consistent behaviour, being the most accurate model with the lowest rRMSE, and one of the least biased ones together with the seasonal persistence model. The different magnitude of both rMBE and rRMSE between GHI and DNI highlights the increased difficulty in predicting DNI, even for short term.

It is interesting to look at the detailed results over the forecast horizon spans of our case study. Figure 2 depicts both rMBE and rRMSE for GHI for the different forecast horizons, from 1 to 24 hours ahead. Looking at the rMBE results, neither DHR or ETS present a clear forecasting behaviour, providing both positive and negative values. As GHI exhibits an almost sinusoidal seasonal pattern we can attempt to explain the shape of the observed bias. Predicting half a day ahead, around 8 to 16 hours in the future, i.e. at the reversal of the seasonal shape, both DHR and ETS are positively biased. This means that they underforecast the true value of GHI. On the other hand, for a few hours ahead or almost a day ahead forecasts, where the seasonal cycle is almost at the same point of the daily cycle, these models are negatively biased, implying that they provide on average forecasts higher than the observed GHI. Note that the rMBE peaks provided by ETS are smaller than the DHR ones. The rest of the models demonstrate a consistently biased behaviour, with all Naïve, S.Naïve and ARIMA forecasting on average lower than the actual GHI values. Note that the bias of Naïve is not plotted in the figure due to its scale, to allow presenting the bias of the rest
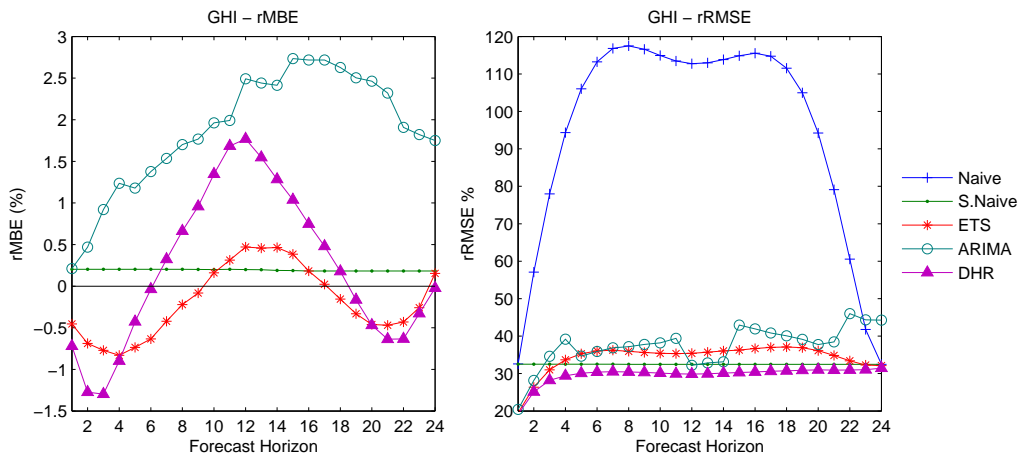
of the models better.

Figure 2: Out of sample forecasting bias (rMBE) and errors (rRMSE) vs. forecasting horizon for GHI.

Turning our attention to rRMSE, DHR is consistently, across forecast horizons, the most accurate model. Naïve exhibits an almost seasonal shape in its errors, which makes sense since the model is unable to capture seasonality. On the other hand, S.Naïve has almost the same rRMSE over the different forecast horizons. ETS and ARIMA exhibit good performance over very short horizons, from 1 to 3 hours ahead, but their performance degrades fast afterwards, being less accurate than the simpler seasonal persistence model, providing evidence that the latter is a good benchmark for GHI forecasting studies.

Figure 3 shows the rMBE and rRMSE results per forecast horizon for the DNI. The first striking difference with figure 2 is the poor performance of ARIMA, as already discussed earlier. The second difference is that in terms of bias all models are positively biased. Although DHR and ETS still exhibit a variable magnitude of bias over the different forecast horizons, somewhat repeating the shape observed for GHI, their bias is consistently greater than zero, with the exception of 1-hour ahead forecasts, which have almost no bias. In terms of accuracy, looking at rRMSE, DHR is again consistently the most accurate model, closely followed by ETS. The S.Naïve also provides quite accurate forecasts, especially for its simplicity. This strengthens our previous conclusion that the seasonal persistence model is a very valuable benchmark for this type of time series.
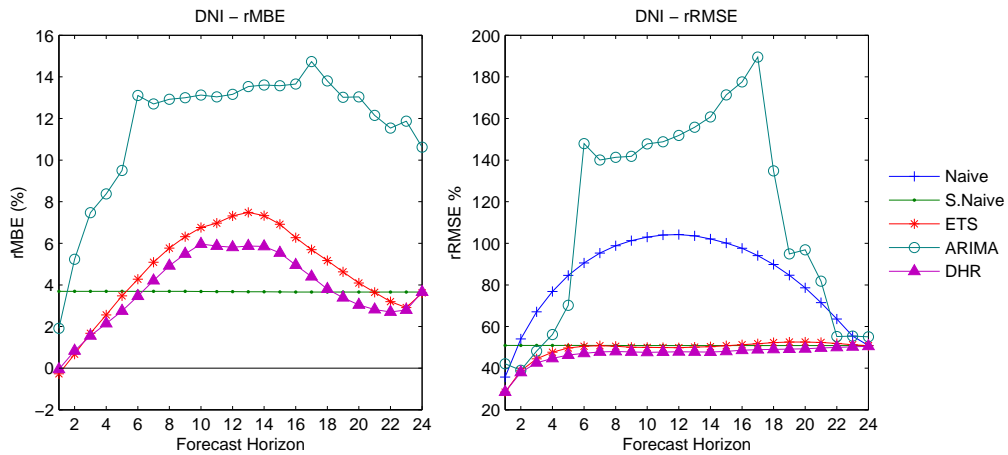
14

Figure 3: Out of sample forecasting bias (rMBE) and errors (rRMSE) vs. forecasting horizon for DNI.
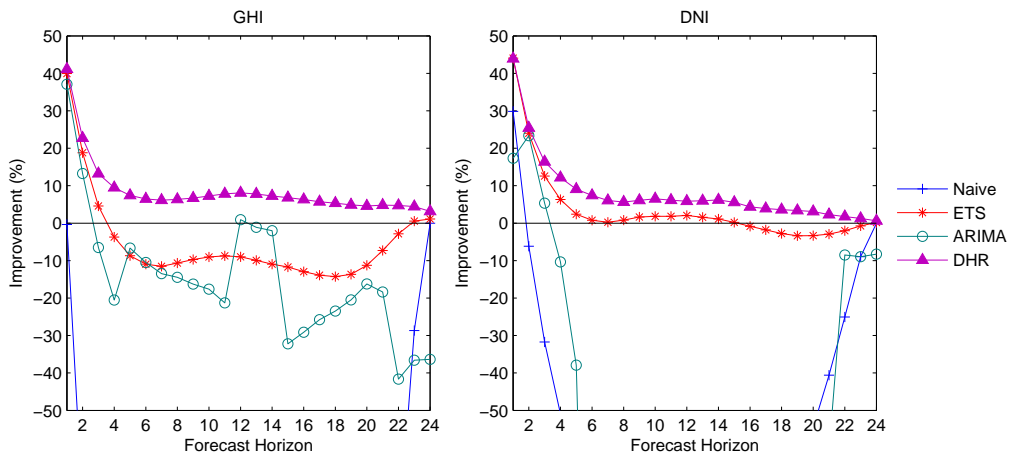

Figure 4: Improvement (%) vs. forecasting horizon for GHI and DNI.

Figure 4 builds on this and provides the Improvement results, as defined in Eq. (13). To allow better understanding of the behaviour of the good performing models we restrict the plots between -50% and 50% improvement, putting both Naïve and ARIMA, which perform poorly, outside of the bounds of the figure for several of the forecast horizons. This figure clearly demonstrates that DHR consistently offers the biggest improvements over the seasonal persistence benchmark, with particularly good performance for

15

short horizons. Reflecting the results in table 1 the lower improvement ratios achieved for the DNI remark again the difficulties of forecasting the solar radiation measured by DNI instead of GHI.

Therefore, from our experiments we find that the proposed DHR for both GHI and DNI performs best over the benchmarks and over the forecast horizons considered here.

5. Conclusions

In order to efficiently incorporate renewable energies into the grid, it is very important to provide accurate forecasts. In the particular case of solar energy, the key variable is the solar irradiation. In recent years novel methodologies have proliferated to forecast solar irradiation with different complexity levels. Nonetheless, there is not unanimity about which technique is the most adequate. In certain extent, this is due to the difficulties found to reproduce the results given the geographical dispersion of each experiment.

This article analyzes the DHR performance to tackle the problem of solar irradiation forecasting. The forecasts are provided in a single recursive fashion avoiding the problem of the two-step procedures consisting of deseasonalizing the time series and modeling the residuals. Note that the latter step of identifying the residuals structure can be a complex task that should be carried out by experts. In contrast, the simultaneous estimation provided by the DHR reduces potential parameter bias that can be introduced by two-step approaches. Furthermore, the DHR can be automated and it can handle time-varying amplitudes and phases typically associated to the periodic behaviour of solar irradiation variables.

The results show that the DHR improves significantly the forecast accuracy with regards to the persistence model and other well-known benchmarks and thus, it can be considered as a competitive forecasting technique. These conclusions have been validated using both GHI and DNI solar radiation data measured in weather stations located in Spain.

Since the forecast errors provided by the GHI was lower than those obtained by the DNI, further research should investigate this relationship in order to propose causal models that might be capable of enhancing the forecasting accuracy achieved instead of analyzing each time series independently.

Additionally, it would be interesting to corroborate the potential advantages of the DHR by forecasting solar irradiation data coming from other geographic locations. Given that all the forecasting models employed in this

article are implemented in software freely available, the results provided are easily reproducible, following the suggestions of many academics and journals in relation to reproducibility in forecasting research [39].

Acknowledgment

References

[1] B. Kraas, M. Schroedter-Homscheidt, R. Madlener, Economic merits of a state-of-the-art concentrating solar power forecasting system for participation in the spanish electricity market, Solar Energy 93 (0) (2013) 244 – 255.

[2] M. Diagne, M. David, P. Lauret, J. Boland, N. Schmutz, Review of solar irradiance forecasting methods and a proposition for small-scale insular grids, Renewable and Sustainable Energy Reviews 27 (0) (2013) 65 – 76.

[3] J. Polo, L. Zarzalejo, L. Ramírez, Solar Radiation Derived from Satellite Images, Springer Ed., 2008, Ch. 18, Modeling Solar Radiation at the Earth's Surface.

[4] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, B. Washom, Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed, Solar Energy 85 (11) (2011) 2881 – 2893.

[5] G. Reikard, Predicting solar radiation at high resolutions: A comparison of time series forecasts, Solar Energy 83 (3) (2009) 342 – 349.

[6] C. Gueymard, H. Kambezidis, Solar radiation and daylight models, 2nd Edition, Elsevier, 2004, Ch. 5, Solar spectral radiation, pp. 221–296.

[7] C. Voyant, M. Muselli, C. Paoli, M.-L. Nivet, Numerical weather prediction (nwp) and hybrid arma/ann model to predict global radiation, Energy 39 (1) (2012) 341 – 355, sustainable Energy and Environmental Protection 2010.

[8] A. Sfetsos, A. Coonick, Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques, Solar Energy 68 (2) (2000) 169 – 178.

[9] J. Boland, Time-series analysis of climatic variables, Solar Energy 55 (5) (1995) 377 – 388.

[10] Z. Dong, D. Yang, T. Reindl, W. M. Walsh, Short-term solar irradiance forecasting using exponential smoothing state space model, Energy 55 (0) (2013) 1104 – 1113.

[11] F. H. Al-Sadah, F. M. Ragab, M. K. Arshad, Hourly solar radiation over bahrain, Energy 15 (5) (1990) 395 – 402.

[12] S. Kaplanis, New methodologies to estimate the hourly global solar radiation; comparisons with existing models, Renewable Energy 31 (6) (2006) 781 – 790.

[13] A. Baig, P. Akhter, A. Mufti, A novel approach to estimate the clear day global radiation, Renewable Energy 1 (1) (1991) 119 – 123.

[14] D. Yang, V. Sharma, Z. Ye, L. I. Lim, L. Zhao, A. W. Aryaputera, Forecasting of global horizontal irradiance by exponential smoothing, using decompositions, Energy, In press. doi:http://dx.doi.org/10.1016/j.energy.2014.11.082.

[15] P. Young, D. Pedregal, W. Tych, Dynamic harmonic regression, Journal of Forecasting 18 (1999) 369–394.

[16] P. Young, Recursive Estimation and Time-Series Analysis: An Introduction for the Student and Practitioner, Springer Berlin Heidelberg, 2011.

[17] M. West, J. Harrison, Bayesian Forecasting and Dynamic Models, Springer-Verlag, 1989.

[18] A. Harvey, Forecasting Structural Time Series Models and the Kalman Filter, Cambridge University Press, 1989.

[19] D. J. Pedregal, J. R. Trapero, Electricity prices forecasting by automatic dynamic harmonic regression models, Energy Conversion and Management 48 (2007) 1710–1719.

[20] J. R. Trapero, D. J. Pedregal, Frequency domain methods applied to forecasting electricity markets, Energy Economics 31 (5) (2009) 727–735.

[21] D. J. Pedregal, P. C. Young, Statistical approaches to modelling and forecasting time series, Companion to Economic Forecasting Blackwell Publishers, 2002, Ch. 4, pp. 69–104.

[22] R. E. Kalman, A new approach to linear filtering and prediction problems, ASME Transactions, Journal Basic Engineering 83-D (1960) 95–108.

[23] A. Bryson, Y. Ho, Applied Optimal Control, Optimization, Estimation and Control, Blaisdell Publishing Company, 1969.

[24] C. J. Taylor, D. J. Pedregal, P. C. Young, W. Tych, Enviromental time series analysis and forecasting with the captain toolbox, Enviromental Modelling & Software 22 (6) (2007) 797–814.

[25] L. Martín, L. F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, M. Cony, Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning, Solar Energy 84 (10) (2010) 1772 – 1781.

[26] S. Makridakis, S. C. Wheelwright, R. J. Hyndman, Forecasting: Methods and applications., 3rd Edition, John Wiley & Sons, New York, 1998.

[27] E. S. Gardner, Exponential smoothing: The state of the art, Part II, International Journal of Forecasting 22 (2006) 637–666.

[28] P. R. Winters, Forecasting sales by exponentially weighted moving averages, Management Science 6 (1960) 324–342.

[29] S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, R. Parzen, R. Winkler, The accuracy of extrapolation (time series) methods; results of a forecasting competition, Journal of Forecasting 1 (1982) 111–153.

[30] R. J. Hyndman, A. B. Koehler, R. D. Snyder, S. Grose, A state space framework for automatic forecasting using exponential smoothing methods, Int. J. Forecast. 18 (3) (2002) 439–454.

[31] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, Time series analysis: Forecasting and Control. (3rd. Edition), Upper Saddle River, New Jersey: Prentice Hall, 1994.

[32] H. Akaike, A new look at the statistical model identification, IEEE Transaction on Automatic Control 19 (6) (1974) 716–723.

[33] R. J. Hyndman, Y. Khandakar, Automatic time series forecasting: The forecast package for R, Journal of Statistical Software 27 (3) (2008) 1–22.

[34] G. Schwartz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.

[35] R. Hyndman, FORECAST package for R v5.5 (2014).
URL http://cran.r-project.org/web/packages/forecast/index.html

[36] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0 (2012).
URL http://www.R-project.org/

[37] L. J. B. McArthur, Operations manual. wmo/td-n, 1274, wcrp/wmo, Tech. rep., Baseline Surface Radiation Network (BSRN) (2004).

[38] J. K. Ord, R. Fildes, Principles of Business Forecasting, South-Western Cengage Learning, Mason, Ohio, 2012.

[39] J. E. Boylan, P. Goodwin, M. Mohammadipour, A. A. Syntetos, Reproducibility in forecasting research, International Journal of Forecasting 31 (1) (2015) 79 – 90.