Neural Processing Letters 52.1 (2020): 97–115

DOI: http://doi.org/10.1007/s11063-018-09969-1

Machine Learning Nowcasting of PV Energy using Satellite Data

Alejandro Catalina · Alberto Torres-Barrán · Carlos M. Alaíz · José R. Dorronsoro

the date of receipt and acceptance should be inserted later

Abstract Satellite-measured radiances are obviously of great interest for photovoltaic (PV) energy prediction. In this work we will use them together with clear sky irradiance estimates for the nowcasting of PV energy productions over peninsular Spain. We will feed them directly into two linear Machine Learning models, Lasso and linear Support Vector Regression (SVR), and two highly non-linear ones, Deep Neural Networks (in particular, Multilayer Perceptrons, MLPs) and Gaussian SVRs. We shall also use a simple Clear Sky-based persistence model for benchmarking purposes. We consider prediction horizons of up to six hours, with Gaussian SVR being statistically better than the other models at each horizon, since its errors increase slowly with time (with an average of 1.92 % for the first three horizons and of 2.89 % for the last three). MLPs performance is close to that of the Gaussian SVR for the longer horizons (with an average of 3.1 %) but less so at the initial ones (average of 2.26 %), being nevertheless significantly better than the linear models. As it could be expected, linear models give weaker results (in the initial horizons, Lasso and linear SVR have already an error of 3.21 % and 3.46 %, respectively), but we will take advantage of the spatial sparsity provided by Lasso to try to identify the concrete areas with a larger influence on PV energy nowcasts.

Keywords Photovoltaic energy, Nowcasting, EUMETSAT, Support Vector Regression, Lasso, Clear Sky models

1 Introduction

Climate change and global warming are being accepted not just as possibilities but as established facts that can provoke large, possibly harmful changes in weather

A. Catalina · C.M. Alaíz · J.R. Dorronsoro
Dpto. Ing. Informática, Universidad Autónoma de Madrid, 28049 Madrid, Spain

A. Torres-Barrán
Inst. de Ciencias Matemáticas ICMAT, Campus Cantoblanco UAM, 28049 Madrid, Spain

J.R. Dorronsoro
Instituto de Ingeniería del Conocimiento, Campus Cantoblanco UAM, 28049 Madrid, Spain

behavior and that, therefore, should be vigorously kept in check. To prevent this, the control of CO_2 and other greenhouse gas emissions is crucial. Moreover, increasing the penetration of renewable energies is also recognized as a key tool in the process. In turn, this has led to an enormous growth in wind and solar installations, driven initially by the availability of subsidies for the energy thus produced but, nowadays, more and more so by simple economic considerations, as the capital expenses per installed power are constantly decreasing.

Particularly, this is the case of solar energy. While at the beginning two complementary paths were followed, namely, thermosolar and photovoltaic (PV) installations, lately it seems that PV energy is gaining the upper hand and becoming the main source of new solar power, to the point of being a key energy provider in states like California or Hawaii in the U.S., and Germany or Italy in Europe. In Spain, and after a very fast growth up to 2012, new solar installations dropped precipitously but they are now clearly gaining momentum as the installed solar base should approximately double from the current 4 GW to about 8 GW by the year 2020.

It is thus clear that accurate solar energy forecasting is nowadays becoming crucial for the management of the electricity grid. Moreover, increasing PV production is accompanied by a growing presence of solar power in energy markets, whose agents also have a clear stake in having good production forecasts (see [3] for a recent study on the economic value of PV forecasting in Spain). A large number of solar energy forecasting approaches can be found in the literature. They can be sorted into different families, such as physical techniques based on different models of solar radiation, engineering methods to transform radiation estimates into PV predicted outputs, statistical techniques either based on pure time series analysis or combining these with exogenous variables or, finally, Machine Learning (ML) methods, where explanatory variables such as Numerical Weather Prediction (NWP) inputs, sky camera images or either ground- or satellite-based radiation measurements are fed into more or less black box models that transform them into future PV estimates. In all cases predictions may be sought for individual plants (of interest for their managers) or for production aggregated over large areas (of interest for Transmission System Operators, TSOs, or for large market agents with contracts to purchase and/or deliver energy). A further distinction between methods can be made in terms of the desired forecasting horizons. These may go from the day-ahead or even longer periods to intra-day (hours) or even very short (minutes) prediction horizons. Interested agents are again present in all these horizons, as, for instance, system operators (that must ensure short and medium time system stability and must plan next day energy generation) or market agents (that must be present on both the day-ahead and intra-day markets). Recent surveys of the large literature on these approaches and prediction goals can be found in [2, 16, 26].

This paper is concerned with the short term, a few hours ahead, same-day forecasting of PV energy over a wide area. In our case we deal with peninsular Spain, but the situation may be similar in places such as California or Texas in the U.S., and Germany or Italy in Europe. This geographical scale and the short time horizons restrict the methods to be used. In fact, local PV prediction is better suited to physical or engineering models and the wide variations of atmospheric conditions possible in a large area greatly preclude a pure time series approach, that may not capture the overall influence of these variations. Because of this,

we will thus concentrate on general ML methods, namely sparse linear models (Lasso), Multilayer Perceptrons and Support Vector Regression, both linear and non-linear.

Turning our attention to the prediction variables to be used, NWPs, usually a first choice for renewable energy predictions, are not that useful here, because of the time gaps between consecutive predictions. Taking as an example the European Center for Medium Weather Forecasts (ECMWF), two main NWPs are provided daily, with runs that usually start at 00 and 12 UTC hours and are distributed widely about 6 hours later. National weather services usually provide intermediate forecasts; in the particular case of Spain, ECMWF's NWP forecasts would be available approximately at sunrise and sunset, i.e., about UTC hours 06 and 18. AEMET, Spain's national weather service, provides intermediate predictions whose runs start also at 06 and 18 UTC hours. In any case, there is approximately a 6 hour gap between these forecasts, which makes their application to intra-day solar energy predictions rather difficult.

On the other extreme we could consider ground measurements or sky cameras [19], but their local nature, better suited for single site exploitation, makes it quite difficult to apply their information to cover wider areas, let alone the entirety of peninsular Spain. This essentially leaves us with satellite based measurements, which usually have been applied to derive local irradiation forecasts which, in turn, can be transformed into PV energy predictions. A good example of this approach is the HELIOSAT method [13], a physical model in which information from the high resolution visible satellite channel is used first to track and predict cloud behavior and then its influence on irradiance values. More precisely, a dimensionless cloud index is computed from the visible channel satellite images at a given hour H and then applied to derive irradiance estimates at that hour using a physical clear-sky model. Cloud positions are afterwards estimated at hours $H+1, H+2, \ldots$ using a Motion Vector Field model and then transformed again into estimates of future irradiance values. These estimates have been used to derive short term PV energy forecasts, both at individual plants and over a regional level (see [18, 28]).

In this paper we will also work with satellite measurements, directly feeding them into ML models to nowcast the aggregated PV energy of peninsular Spain. We must point out that satellite values correspond to measurements of reflected radiance and, thus, do not give the incoming irradiances that are actually converted into PV energy. In fact, we use these reflected radiances as proxies for the actual, incoming radiances. As in our previous work in [5], we will consider the 11 spectral bands in the Meteosat satellites run by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) which range from the visible channels to the long-wavelength infrared ones at the opposite extreme of the spectrum. Their approximate spatial resolution is about $3\,\mathrm{km} \times 3\,\mathrm{km}$ for Spain. There is an extra high resolution visible channel with a $1\,\mathrm{km} \times 1\,\mathrm{km}$ resolution that we will not consider here. Moreover, and as discussed in Section 2, each band gives rise to two measures, the effective radiance, plus a reflectance for the first three visible channels and a brightness temperature for the remaining eight infrared ones; thus, we have 22 variables for each grid point. These measures are available each 15 min but, on the other hand, PV energy readings correspond to hourly values of the energy produced up to a given hour H. Because of this we will use at each hour the average of the previous four 15 min satellite readings available up to that time.

Given the large number of satellite readings and their rather fine spatial resolution, we will first downsample Meteosat's initial resolution to that of a 0.125° grid and then select a subset of those readings having the largest correlation with PV energy values. Moreover, we will also incorporate into our forecasting variables a Clear Sky (CS) estimate for the Global Horizontal Irradiance at each Meteosat grid point. We will further aggregate these into a global CS model of peninsular Spain which, in turn, we will use to define a simple persistence model for benchmarking purposes. We believe ours to be the first work on using satellite information to directly nowcast PV energy and, over our previous results in [5], we offer here the following new contributions:

1. We will extend our forecasting horizons from the previous 3 hours up to a much more useful 6 hours.
2. We will consider another powerful non-linear regression method, namely Deep Neural Networks, applying their initialization and training techniques on Multilayer Perceptrons with 4 hidden layers and 100 units per layer.
3. We exploit the spatial structure of Lasso, the sparse linear model we use, to derive a further understanding on the grid point relevance of the satellite measures considered.

The rest of the paper is organized as follows. In Section 2 we give a general overview of the EUMETSAT satellite data system, the satellite readings most relevant to PV energy measures and the concrete channel information that can be obtained essentially in real time from EUMETSAT. Channel information is also analyzed in Section 2 in the context of PV energy prediction over peninsular Spain in order to select the most informative channels. We will also present the CS estimates that we add to our models. In Section 3 we will briefly describe the ML models we will use, namely two regularized linear models, Lasso and linear Support Vector Regression (SVR) as implemented in the LIBLINEAR [10] library, and two non-linear ones, Gaussian SVR and Multilayer Perceptron (MLP). As we shall see, the high dimensionality of the patterns used is not enough to put the linear models on a par with the non-linear ones. On the other hand, the non-linear model results are quite good, particularly that of Gaussian SVRs, which yield good nowcasts for all the horizons up to the 6 hours considered. The performance of MLPs is inferior at the shorter horizons, closer there to that of the linear models, but it deteriorates rather slowly and is not too far away from that of the Gaussian SVRs at the later horizons. We shall also use a simple, CS-based persistence model for benchmarking purposes. As it could be expected, it is not competitive for the longer horizons but, on the other hand and besides being extremely simple, it may be hard to beat on the first two horizons, particularly at noon and early afternoon. Numerical results justifying these assertions are given in Section 4, which also includes a Subsection where we discuss how to take advantage of the spatial sparsity provided by the Lasso model to identify the grid points whose satellite radiances may have a stronger influence on PV energy production. Finally, in Section 5 we briefly recap our contributions, offer our main conclusions and discuss further ways to improve on them.

2 Satellite Data and Clear Sky Model

2.1 EUMETSAT Satellite Data

The Meteosat system is a satellite network that covers Europe as well as parts of Africa and of the Atlantic Ocean and is run by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) [8]. These satellites measure for different wavelengths the radiance emitted and reflected by the Earth's surface and atmosphere. This information, often after some processing, is widely used for weather forecasting as well as climate monitoring and research. An important part of this processing is done by the different centers associated to EUMETSAT's Satellite Application Facility (SAF) network.

The satellites currently under operation are those of Meteosat's Second Generation (MSG) system and they are equipped with the Spinning Enhanced Visible and Infrared Imager (SEVIRI) technology and provide near real-time radiance values every 15 minutes over eleven spectral bands, going from the visible (with a wavelength 0.6 µm at Channel 1) to the long infrared (with a wavelength 13.4 µm at Channel 11). These channels' spatial resolution is about 3 km × 3 km for most of the covered regions; an extra visible channel for radiances on the 0.6 µm to 0.9 µm wavelength range is available with a higher 1 km × 1 km resolution over Europe and some parts of Africa. Some of these channels target concrete physical measures, such as absorption of water-vapor (Channel 5), ozone (Channel 8), or CO_2 (Channel 11).

The basic pixel counts that SEVIRI collects over each channel are further processed by several calibration procedures. In some detail, these counts are first transformed into radiances by the addition of an offset and the multiplication of this shifted count by an appropriate calibration factor, as described in EUMETSAT's inter-calibration documentation [9]. In the case of the visible channels their radiances are further divided by the corresponding maximum solar irradiance, yielding a reflectance percentage. For the infrared channels an empirical formula [9] is applied to their radiances to yield brightness temperatures. In other words, besides the initial radiances, another variable is available at each channel, namely, three reflectances for the visible wavelengths and eight brightness temperatures for the infrared ones. Another independent information provided by Meteosat is the Cloud Mask, a floating point number whose values range from 0, meaning a clear sky, to 5, which corresponds to an overcast condition. However, given that cloud information is implicitly captured in the channel information, we will not use it. In summary, if we exclude the high resolution channel and the Cloud Mask, we have 22 values at each point of a 3×3 grid.

We have downloaded 15 min readings from the EUMETSAT Data Centre for the years 2013, 2014 and 2015. Of these we will use 2013 for training, 2014 for hyper-parameter validation and 2015 for model testing. Given the large area we consider, we will work with a down-sampled resolution of 0.125° and a grid that matches that of the orographical model of the ECMWF. Since we will only consider grid locations over peninsular Spain, this results in a final number of 3391 points. PV energy readings at a given hour H, kindly provided by Red Eléctrica de España (REE), correspond to the energy produced during the entire hour ending at the H value. Because of this we will average at hour H the four 15 min values of the above Meteosat variables available up to that hour. Furthermore, we will only consider

daylight hours, dropping satellite readings outside the UTC range between hours 05 to 20. After this we are finally left with at most 4645 hourly values per year (slightly smaller in fact for some years, due to some missing data).

In any case, using at each point all the Meteosat variables, we would end up with hourly patterns with a dimension of $23 \times 3391 = 77\,993$, clearly too large. To reduce it we have computed for the year 2013 the correlations between the hourly PV energy readings and the hourly averages of each Meteosat variable over the retained grid points. The correlation matrix is depicted in Figure 1; variable names indicate Infrared (`IR`), Visible (`VIS`) or Water Vapour (`WV`) channels and their corresponding type is `%` (reflectivity), `R` (radiance) and `K` (brightness temperature); `Prod` refers to the PV energy production; and the number accompanying the variable tag refers to its wavelength. The figure shows that infrared radiances are highly correlated with their corresponding brightness temperatures. Correlations are positive for all channels and largest for the infrared `IR_016` and `IR_039` channels followed by the `VIS008` visible one. Figure 2 gives the correlations between the variables and the PV energy production sorted in decreasing order. According to the figure, we have selected the radiances of the `IR_016`, `IR_039` and `VIS008` channels plus the brightness temperature of channel `IR_039` and discarded the rest in our experimental work. Among the discarded variables there are `IR_016 %` and `VIS008 %`, because they have a perfect correlation of 1 with their radiances counter parts `IR_016 R` and `VIS008 R`.

2.2 Clear Sky Model

Clear Sky (CS) models, usually derived from atmospheric physical analysis, try to estimate either the Direct Normal Irradiance (DNI) or the Global Horizontal Irradiance (GHI) at a certain point. In principle, the horizontal radiation I at a given point is given by $I = \mathcal{I}^{\mathrm{d}} \cos\Theta$, where Θ is the zenith angle, i.e., the angle between the incident solar rays and the vertical at the point, and $\mathcal{I}^{\mathrm{d}}$ the direct (beam) radiation. While Θ is a simple trigonometric function $\Theta = \Theta(L, D, H)$ of the latitude L of the point and the day D and hour H, modeling $\mathcal{I}^{\mathrm{d}}$ is more difficult as it requires the local adjustment of several physical parameters, such as height, atmospheric pressure, temperature or air turbidity, which, in turn, have to be carefully calibrated at the concrete sites where they are to be used.

In fact, the physical modeling of solar radiation is a very active research area with many models proposed in the recent literature. A review of some of these models from the point of view of renewable energy is in [21]. Here we will consider the model introduced by [15] as implemented in the Python library `pvlib` [23]. We have run this model to estimate the GHI for each of our 3391 grid points in peninsular Spain over the 3 years considered. At each point its altitude is derived from its geopotentials in the ECMWF's orographical model for Spain; for all other parameters we use their `pvlib` defaults.

These CS GHI estimates can be obviously used at each hour, either past or future, and we will add their values at hours H and $H - K$ as extra variables when predicting PV energy at hour H from features at hour $H - K$. After adding these CS values, we end up with a yearly data matrix of dimensions 4645 patterns $\times$ 16 955 features for the model at hour H (i.e., $K = 0$), where we only use the CS estimate for that hour, and $4645 \times 20\,346$ for the rest (the actual number of patterns

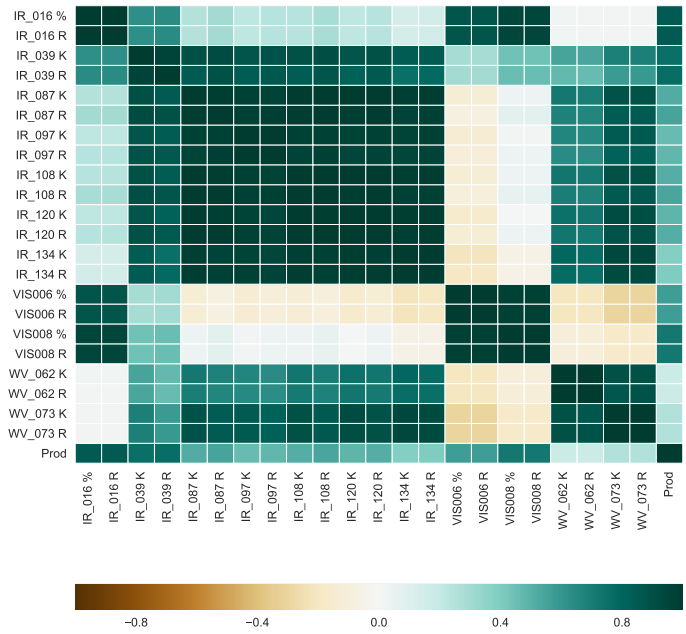


Fig. 1: Correlations between satellital radiances and PV energy.

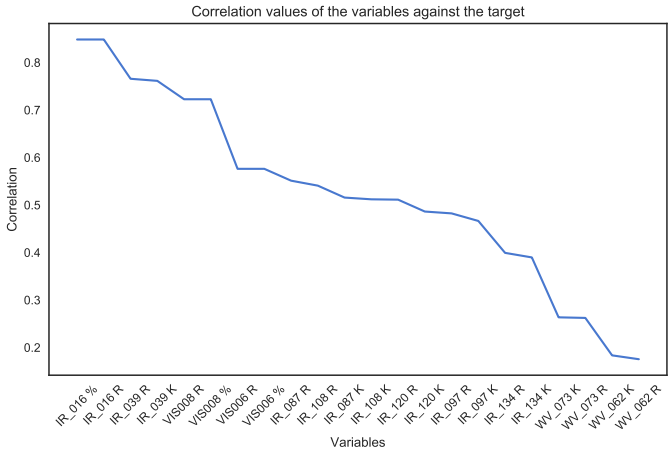


Fig. 2: Correlations with PV energy of satellital radiances sorted in decreasing order.

at each year is actually slightly smaller, because of missing values). Moreover, and as explained in Subsection 3.1, we will also use these CS estimates of GHI to define a very simple persistence model for benchmarking purposes.

3 Machine Learning Models

As mentioned, we will use four well-established ML models: Lasso, Multilayer Perceptrons (MLPs), and linear and Gaussian kernel Support Vector Regression (SVR); some of them have been recently used in the literature of solar power forecasting [20]. In principle, the high dimensionality of the problem suggests that linear models should be a first choice. That is why we choose Lasso to force feature sparsity and cope with possible collinearities and linear SVR, with its epsilon insensitive loss function. Of course, powerful non linear models such as MLPs and Gaussian SVRs are mandatory.

To these, and within the realm of mainstream yet powerful regression models, we could add Random Forests and Gradient Boosting. However, we have not done so since we believe that the very high dimensionality of the problem may hinder their performance: both randomly choose the features to be considered for tree splitting which, with dimensions above 15 000 in all datasets, is likely not to be a clear-cut task. We have seen this to be the case in previous work [25] on wind energy, where these models yielded very good results for an individual wind farm but less so for the entire Peninsular Spain. Of course, more regression models could be used, but we believe that our choice of models is a sensible and broad one for an approach that, to the best of our knowledge, has not received previous attention in the literature and that may, of course, be improved in further work.

Finally, and as mentioned, we will additionally consider a CS-based persistence model as a baseline benchmark for the others. We briefly describe all these models next.

3.1 Persistence

Our baseline CS model simply predicts PV energy at hour H as a scaling of the energy at hour $H - K$, where the scaling factor is given as a ratio of the corresponding averages of CS values over the grid points. Its formulation for an $H - K$ model with a horizon forecast of K hours is thus

$$\widehat{\mathrm{PV}}_H = \frac{\mathrm{CS}_H}{\mathrm{CS}_{H-K}} \mathrm{PV}_{H-K} \tag{1}$$

where $\mathrm{PV}_{H-K}, \widehat{\mathrm{PV}}_H$ denote the real and predicted PV energy at hours $H - K$ and H and CS_h the averaged value of the clear-sky radiances at a given hour h over the area under consideration. While very simple, the baseline CS models will give in general reasonable values and particularly good, hard to beat ones around noon, specially in the mid-year, mostly sunny, months.

3.2 The Lasso Model

Given an N pattern sample $\{(\mathbf{x}^1, y^1), \ldots, (\mathbf{x}^N, y^N)\}$ that we organize in an $N \times p$ data matrix $\mathbf{X}$, with p the pattern dimension, and an N dimensional target vector $\mathbf{y}$, the Lasso solution [14] $(b^*, \mathbf{w}^*)$ minimizes the ℓ_1 regularized loss

$$\ell_{\mathrm{L}}(\mathbf{w}, b) = \frac{1}{2}\|\mathbf{y} - \mathbf{X}\mathbf{w} - b\|^2 + \lambda\|\mathbf{w}\|_1. \tag{2}$$

The sparsity introduced by the ℓ_1 regularization helps to avoid possible singularities in the sample covariance matrix, particularly in cases such as ours where sample sizes of ≈ 4500 are much smaller than the feature's dimension of $\approx 20\,000$. This is achieved because the ℓ_1 penalty drives many coefficients towards zero; in turn, this allows automatic (i.e., wrapper-based) feature selection and makes possible model interpretation in terms of the specific grid positions of the non-zero coefficients. We shall use the *Scikit-learn* [22] implementation of Lasso.

3.3 Multilayer Perceptrons

Given the same sample as in the previous case, in Neural Network (NN) regression we minimize the following ℓ_2 regularized loss

$$\ell_{\mathrm{NN}}(\mathbf{w}) = \frac{1}{2N}\sum_p (y^p - f(\mathbf{x}^p, \mathbf{w}))^2 + \frac{\alpha}{2}\|\mathbf{w}\|_2\,, \tag{3}$$

where by $f(\mathbf{x}, \mathbf{w})$ we denote the output on $\mathbf{x}$ of a Multilayer Perceptron (MLP) whose architecture determines a weight set $\mathbf{w}$. Many possible architecture options are available, including convolutional or even recurrent ones; here we have worked with feed-forward, fully connected MLPs with the same number of hidden units per layer, namely 4 layers of 100 units each (similar architectures with different numbers of layers gave similar results).

The classical theory of such networks is very well known (see for instance [4, Chapter 5]) but the recent developments on deep NNs have resulted in important improvements, some of which we will apply here, such as ReLUs [12] as the activation functions, Glorot–Bengio weight initialization [11] and minibatch-based Adam [17] as the optimization method, resulting in networks with a large number of weights even if not very deep. These options are available in widely-known neural network frameworks such as *Keras* [7] or *Scikit-learn* [22]. For this work we will use the standard implementation of *Scikit-learn*. In a slight abuse of language, we may refer occasionally to these networks as "deep", in the sense that they are defined, initialized and optimized using these new advanced techniques, rather than implying a large number of hidden layers.

It should be noted that we will not consider convolutional layers. The main reason is that standard implementations of convolutional layers assume "rectangular" inputs such as images, because handling the vertical and horizontal strides of rectangular images when applying convolutional kernels is quite easy. However, while satellital readings can be easily organized as rectangular images, in the case of Spain's PV they would include large sea regions where satellital readings are subject to substantial albedo effects that distort them and, very likely, the predictions they may provide. To this, we have to add another factor; during most of the

year, sunrise in Spain takes place, more or less, in parallel to the Mediterranean coast, which implies that, at several hours, most of the radiation is over the sea and not inland. A final reason is our need to diminish the dimensionality of the problem, that we achieve by cropping up satellital readings so that their coordinates adjust to the shape of peninsular Spain. After all this pre-processing the inputs to the model are no longer rectangular images, and, hence, we cannot use standard tools for convolutional networks such as TensorFlow or wrappers such as *Keras*.

3.4 Linear and Gaussian Support Vector Regression

The large dimension of our problem motivates our choice for the Linear SVR model, which is reportedly good in problems of high dimensionality [10]. Using the previous notation, the Linear Support Vector Regression (SVR) cost function is

$$\ell_{\mathrm{S}}(\mathbf{w}, b) = \sum_p [y^p - \mathbf{w} \cdot \mathbf{x}^p - b]_\epsilon + \frac{1}{C}\|\mathbf{w}\|_2^2, \tag{4}$$

where we use an ℓ_2 regularization and the ϵ-insensitive loss $\ell(y, \hat{y}) = [y - \hat{y}]_\epsilon = \max\{|y - \hat{y}| - \epsilon, 0\}$, that defines an ϵ-wide, penalty-free "error tube" around the model. Notice that this loss-regularization combination is one of the different possibilities in the LIBLINEAR implementation in *Scikit-learn* [22]. Notice also that LIBLINEAR's model is homogeneous, that is, we drop the intercept term b.

The initial and more standard way to find the optimal $(\mathbf{w}^*, b^*)$ in an SVR model is to rewrite (4) as a constrained minimization problem which is then transformed using Lagrangian theory into a much simpler dual problem, the one actually being solved; see [24]. The optimal $(\mathbf{w}^*, b^*)$ are then obtained from the dual solution through the KKT equations. It turns out that the dual problem only involves patterns through their dot products and a natural extension to improve on a purely linear model is to apply the kernel trick [24]. It replaces the initial dot products $\mathbf{x} \cdot \mathbf{x}'$ with the values $k(\mathbf{x}, \mathbf{x}')$ of a positive definite kernel k that can be written as $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x}) \cdot \phi(\mathbf{x}')$, where the $\mathbf{x}$ are mapped through $\phi(\mathbf{x})$ into a larger, possibly infinite, dimensional Hilbert space $\mathcal{H}$. We thus arrive to a non-linear model $f(\mathbf{x}) = \mathbf{W} \cdot \phi(\mathbf{x}) + b$ for which the optimal $\mathbf{W}^* \in \mathcal{H}$ can be written as $\mathbf{W}^* = \sum_p \alpha_p^* \phi(\mathbf{x}^p)$. We thus have

$$f(\mathbf{x}) = b^* + \mathbf{W}^* \cdot \phi(\mathbf{x}) = b^* + \sum_p \alpha_p^* \phi(\mathbf{x}^p) \cdot \phi(\mathbf{x}) = b^* + \sum_{\alpha_p^* > 0} \alpha_p^* k(\mathbf{x}^p, \mathbf{x}), \tag{5}$$

where the $\mathbf{x}^p$ for which $|\alpha_p^*| > 0$ are the Support Vectors (SVs), and the standard kernel choice is the Gaussian one, $k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)$. Notice that in our case, the SVs lend themselves to a temporal interpretation as the most relevant day-hour pairs, given that their radiances define the centers of the different Gaussians in the model.

Here we will use the *Scikit-learn*'s wrappers of the very well-known LIBSVM [6] and LIBLINEAR libraries.

Table 1: Hyper-parameters of the Lasso and Linear SVR models.

Model	Parameter	K						
		0	1	2	3	4	5	6
Lasso	λ	0.020	0.017	0.012	0.016	–	–	–
Linear SVR	C $(\times 10^3)$	9.410	17.742	6.349	13.511	–	–	–
	ϵ	1.690	1.670	2.800	3.000	–	–	–
Gaussian SVR	C $(\times 10^3)$	6.820	18.389	2.559	10.229	1.081	1.218	0.064
	ϵ	0.011	0.022	0.018	0.036	0.018	0.029	0.027
	γ $(\times 10^{-3})$	0.241	0.238	0.231	0.244	0.242	0.230	0.049
MLP	α	1e-4	1e-7	1e-8	1e-5	10	0.1	10

3.5 Hyper-parameter Tuning

We will work with the Lasso, the MLP and the LIBLINEAR and LIBSVM implementations of SVR in *Scikit-learn* [22]. All these models require a careful hyper-parameter tuning to find the optimal λ for Lasso, α for MLP, C and ϵ for Linear SVR, and C, ϵ and γ for Gaussian SVR. We used for this the `GridSearchCV` class in [22]. More specifically, we define the following intervals to perform the search:

1. For Lasso's λ we explore 50 values in the interval $(10^{-6}, 10^{8})$.
2. For MLP's α we explore 13 values in the interval $(10^{-6}, 10^{8})$.
3. For Linear SVR and Gaussian SVR's C we explore 14 values in $(4^{-5}, 4^{8})$, and for ϵ we explore 5 values in the interval $(4^{-4}\sigma, \sigma)$, where σ is the standard deviation of the training targets.
4. For Gaussian SVR's γ we explore 6 values in the interval $(\frac{1}{d}4^{-2}, \frac{1}{d}4^{3})$, where d is the number of features.

Given the natural temporal ordering of the data, we will use for this 2013 as a training set and 2014 as a validation set. Table 1 shows the optimal hyper-parameters for our models.

4 Results

Our main goal in this paper is to test the performance of the aforementioned machine learning models for PV energy production forecasting at hour H from satellite readings at hour $H-K$, where K is the prediction horizon. In this section we will show the results of our experiments, that will include:

1. A short term forecasting scenario, considering short horizons, namely 1, 2 and 3 hours ahead plus a baseline model at the 0 horizon.
2. A longer term forecasting scenario, considering the horizons 4, 5, and 6 hours ahead for the best performing, non-linear MLP and Gaussian SVR models on the previous experiment.
3. An analysis of a sparse model interpretation.

As mentioned, for these experiments we will employ data from years 2013, 2014 and 2015, corresponding respectively to the train, validation and test sets. After searching for the best hyper-parameters, we train the model over the union of the training and validation sets and then predict over the test set. As usually, when

working with these models, we will scale the features to a 0 mean and 1 standard deviation for all models except Gaussian SVR, where we scale them to a $[-1, 1]$ range to better control the Gaussian kernel behavior when dealing with largely different pattern pairs.

4.1 Short Term Forecasting

In a first experiment we will consider the $K = 1, 2, 3$ horizons as well as the $K = 0$ case and denote the resulting models as `m1`, `m2`, `m3` and `m0` respectively; while not useful from an operative point of view, the `m0` model offers a "best possible" baseline with which we can compare the others. For a more homogeneous comparison we will report errors within the time range between 08 and 20 UTC hours, given that for each increasing horizon we have to shift the data one hour. PV energy production after hour 20 is negligible and it is obvious that the results before 08 UTC of, say, the `m2` and, more so, the `m3` models will not be good, as they have to predict substantial PV energy at 08 UTC from very small readings at 06 UTC and 05 UTC respectively.

In fact, the average error of the Gaussian `m3` model is comparable with those of the `m1` Lasso and Linear SVR models, and quite close to their `m0` models errors.

Table 2 shows a summary of the overall hourly average test errors over 2015 for each model and prediction horizon, where we omit the H (i.e. $K = 0$) persistence model for obvious reasons; for a quick overall glance, the rightmost column indicates the total mean test error of each model and the corresponding ranking; the same rank value is repeated for two models whenever they are not significantly different according to a Wilcoxon signed-rank test [27] (whose implementation is included in the `scipy.stats` Python package) with a significance level of 5 %. We compute the ranking by comparing the hourly absolute errors of all models over the test set. As it can be seen, the Gaussian (G) SVR results are clearly better than those of Lasso, MLP and linear (L) SVR essentially across all hours. Nonetheless, in the case of MLP it is remarkable the slow degrading of its errors, making its performance closer to that of Gaussian SVR as the horizon grows. Moreover, Gaussian SVR's errors degrade much more slowly than the other models (except MLPs) as the prediction horizon increases; in fact, the average error of the Gaussian `m3` model is comparable to those of the `m1` Lasso and Linear SVR models, and quite close to their `m0` models errors.

The behavior of the CS persistence model is also remarkable. Its average errors are much larger but this is due to the expectedly bad behavior of the $H-K$ models at the beginning of the day, where they should not be used. On the other hand, its evening errors tend to be the best ones and its afternoon errors are comparable and sometimes better than those of Lasso and linear SVR. A plausible reason is the many near-clear sky days of the Iberian Peninsula, particularly in areas with large PV productions.

The clearly best average behavior of the Gaussian SVR model also holds when we consider its hourly errors on a monthly basis. Details of the `m0` and `m2` models are given in Tables 3 and 4 (recall that the `m0` model is a baseline and the `m2` is the first operational hour for intra-day energy markets). As it can be seen, hourly errors also increase quite moderately with the prediction horizons for all months. Moreover, monthly average errors are quite stable; the worst months seem to be

Table 2: Average hourly and overall test errors and rankings for Lasso, Linear SVR, Gaussian SVR, MLP and CS Persistence. The blue color corresponds to the minimum (0.05), and the red color to the maximum omitting CS (6.53).

K	Models	Hour													Av.
		8	9	10	11	12	13	14	15	16	17	18	19	20	
0	Lasso	2.32	2.53	2.69	2.71	2.58	2.77	2.78	2.46	2.39	2.15	0.98	0.64	0.40	2.11 (2)
	L SVR	1.50	2.19	2.63	2.57	2.86	3.05	2.96	2.73	2.30	1.69	1.10	0.61	0.27	2.03 (2)
	G SVR	1.02	1.77	1.94	1.94	1.91	1.90	1.81	1.88	1.72	1.43	0.82	0.46	0.17	1.44 (1)
	MLP	1.30	2.35	2.55	2.74	3.01	3.07	2.82	2.57	2.46	2.15	1.33	0.51	0.11	2.07 (2)
1	CS	18.25	3.99	3.09	3.04	2.26	1.34	1.84	2.66	2.66	1.74	0.54	0.23	0.04	3.21 (4)
	Lasso	2.46	3.19	3.64	3.43	3.29	2.85	2.59	2.44	2.34	2.70	1.81	0.82	0.21	2.44 (3)
	L SVR	1.83	2.51	2.78	3.01	3.18	3.34	3.07	2.96	3.15	2.55	1.74	1.00	0.35	2.42 (3)
	G SVR	1.25	1.80	2.04	1.85	1.92	1.82	1.86	1.86	1.67	1.45	1.10	0.54	0.24	1.49 (1)
	MLP	1.43	2.41	2.90	2.70	2.84	3.15	3.21	2.86	2.44	1.87	1.05	0.32	0.06	2.09 (2)
2	CS	47.86	24.71	8.91	6.14	5.49	3.31	2.56	4.01	4.31	2.93	1.01	0.30	0.05	8.58 (5)
	Lasso	3.93	4.27	4.63	4.23	4.02	3.60	3.25	2.99	3.15	3.87	2.49	1.21	0.38	3.23 (3)
	L SVR	2.77	3.25	3.64	3.62	4.05	4.40	4.97	4.94	4.85	4.50	2.76	1.59	0.78	3.55 (4)
	G SVR	1.98	2.60	2.78	2.59	2.40	2.21	2.05	2.20	2.08	1.71	1.23	0.59	0.36	1.91 (1)
	MLP	1.65	2.82	3.08	2.89	2.76	2.81	2.92	2.87	2.44	1.90	1.14	0.38	0.07	2.13 (2)
3	CS	69.06	45.09	24.72	12.42	8.43	6.39	3.45	4.38	5.13	3.69	1.41	0.35	0.05	14.20 (5)
	Lasso	4.66	4.63	5.66	5.36	5.04	4.87	4.21	3.86	3.66	4.37	2.95	1.51	0.62	3.95 (3)
	L SVR	3.10	3.77	4.06	4.42	4.85	5.37	6.14	6.53	6.34	5.79	3.71	2.10	1.13	4.41 (4)
	G SVR	2.75	3.27	3.69	3.29	3.19	2.80	2.61	2.49	2.37	1.87	1.30	0.65	0.30	2.35 (1)
	MLP	1.99	3.20	3.85	3.60	3.52	3.45	3.41	3.27	2.92	2.19	1.36	0.39	0.05	2.55 (2)

Table 3: Gaussian SVR test errors of the `m0` model per hour and month. The blue color corresponds to the minimum (0), and the red color to the maximum (2.82).

Month	Hour													Av.
	8	9	10	11	12	13	14	15	16	17	18	19	20	
January	0.37	1.37	1.96	1.70	1.74	1.53	1.47	1.25	1.28	0.73	0.21	0.00	0.00	1.05
February	0.70	1.72	2.08	1.92	2.12	2.43	2.26	2.17	1.49	1.46	0.42	0.19	0.00	1.46
March	1.07	2.82	2.58	2.56	2.27	2.21	2.11	2.05	1.86	1.54	1.05	0.35	0.00	1.73
April	1.24	1.57	1.35	1.39	1.59	1.42	1.15	1.47	1.83	1.74	1.21	0.49	0.23	1.28
May	1.04	1.06	1.99	1.45	1.31	1.31	1.35	1.48	1.16	1.22	1.04	0.86	0.18	1.19
June	1.13	1.48	1.62	1.33	1.36	1.44	1.91	1.87	1.76	2.17	1.67	1.33	0.60	1.51
July	1.41	2.18	2.22	2.07	1.53	1.19	1.44	2.05	2.34	2.68	1.28	0.76	0.44	1.66
August	1.66	2.16	2.33	2.81	2.67	2.52	2.61	2.72	2.47	1.91	1.52	0.99	0.59	2.07
September	1.02	1.47	1.49	1.65	2.05	1.86	1.88	1.59	1.39	1.49	0.73	0.55	0.01	1.32
October	1.34	1.61	1.66	2.29	2.24	2.50	1.82	2.04	1.81	1.06	0.42	0.00	0.00	1.45
November	0.80	1.73	1.39	1.44	1.71	2.17	1.71	1.61	1.52	0.58	0.22	0.01	0.01	1.15
December	0.47	2.05	2.67	2.64	2.28	2.22	1.98	2.34	1.76	0.53	0.01	0.00	0.00	1.46
Average	1.02	1.77	1.94	1.94	1.91	1.90	1.81	1.88	1.72	1.43	0.82	0.46	0.17	1.44

February and March, most likely because of the atmospheric instability to be expected in them. August also shows large errors, easily explained by its high PV energy production, much larger than the one from, say, November to February.

Summing things up, satellite information combined with Gaussian SVR can be used to set up a PV energy nowcasting procedure with a quite good performance on the horizons 1, 2 and 3.

Table 4: Gaussian SVR test errors of the `m2` model per hour and month. The blue color corresponds to the minimum (0), and the red color to the maximum (4.5).

Month	Hour													Av.
	8	9	10	11	12	13	14	15	16	17	18	19	20	
January	0.94	1.77	3.37	2.86	2.30	2.03	1.53	1.75	1.69	0.88	0.36	0.00	0.00	1.50
February	1.06	4.22	4.50	3.43	3.04	2.85	2.53	2.85	2.49	1.86	1.22	0.11	0.00	2.32
March	2.61	3.26	2.98	4.27	3.71	2.92	2.62	2.55	2.17	1.79	1.07	0.36	0.00	2.33
April	2.64	2.47	1.94	2.47	1.95	1.97	2.31	2.15	1.90	1.79	1.53	0.92	0.60	1.89
May	2.86	2.11	2.13	1.22	1.60	1.44	1.28	1.66	1.33	1.71	1.38	0.73	0.45	1.53
June	2.22	2.33	2.53	2.09	1.49	1.16	1.27	1.36	2.01	2.34	2.46	1.74	0.85	1.84
July	2.14	2.29	2.25	2.07	1.61	1.31	1.18	1.32	1.73	2.45	2.44	1.38	1.05	1.78
August	2.16	2.22	2.19	2.39	2.07	2.64	2.37	2.69	2.95	2.95	2.05	1.13	1.31	2.24
September	2.24	2.26	2.49	2.26	2.33	1.91	2.05	2.42	2.41	1.71	1.33	0.64	0.01	1.85
October	2.60	3.33	3.32	2.66	3.01	2.59	2.24	2.37	2.09	1.40	0.61	0.00	0.00	2.02
November	1.33	2.57	2.53	2.58	2.44	1.91	1.66	1.75	1.47	0.72	0.23	0.01	0.01	1.48
December	0.98	2.43	3.12	2.72	3.28	3.73	3.53	3.54	2.73	0.91	0.01	0.00	0.00	2.08
Average	1.98	2.60	2.78	2.59	2.40	2.21	2.05	2.20	2.08	1.71	1.23	0.59	0.36	1.91

Table 5: Average hourly and overall test errors and rankings for Gaussian SVR and MLP for horizons beyond 3 hours. The blue color corresponds to the minimum (0.09), and the red color to the maximum (5.24).

K	Models	Hour												Av.
		9	10	11	12	13	14	15	16	17	18	19	20	
4	G SVR	3.64	3.91	3.98	3.58	3.55	3.29	3.09	2.70	1.99	1.27	0.65	0.29	$2.66^{(1)}$
	MLP	3.57	4.32	4.27	3.92	3.71	3.70	3.52	3.07	2.21	1.36	0.55	0.09	$2.86^{(2)}$
5	G SVR	–	4.67	4.27	4.37	4.01	4.07	3.68	3.14	2.18	1.21	0.60	0.29	$2.95^{(1)}$
	MLP	–	4.40	4.67	4.57	4.31	4.14	4.01	3.50	2.52	1.30	0.45	0.10	$3.09^{(2)}$
6	G SVR	–	–	4.72	4.84	4.60	4.27	3.96	3.61	2.52	1.22	0.57	0.29	$3.06^{(1)}$
	MLP	–	–	5.08	5.24	5.03	4.73	4.32	3.81	3.00	1.53	0.56	0.24	$3.35^{(2)}$

4.2 Longer Term Forecasting

Given the slow degrading of both MLP and Gaussian SVR we have extended the previous experiment by adding more prediction horizons beyond 3 hours; specifically, we have tried $K = 4, 5, 6$ hours ahead. Their errors summary is shown in Table 5 where we shift the initially predicted hour to hours 9 for model `m4`, 10 for `m5` and 11 for `m6`. Again, the rightmost column indicates the mean of the errors of both models and their corresponding ranking. Here a Wilcoxon signed-rank test shows the Gaussian SVR and MLP models to be significantly different, and no rank repetitions arise. Even though the problem keeps getting harder (note that to predict at 11 UTC we have to consider data at 05 UTC for the `m6` model) as we increase the forecasting horizon, the errors degrade remarkably slowly. We also see that for the earliest and the latest hours the MLP obtains better predictions. Nonetheless, Gaussian SVR is still the best model by a good margin, even if the MLP is closer than in shorter horizons. Summing up, we see that MLP is a better model for longer horizons than for shorter ones, getting quite close to the Gaussian SVR's performance.

Given its very good results, we also show the monthly errors of Gaussian SVR for horizons 4 and 6 in Tables 6 and 7. As we can see, the conclusions observing

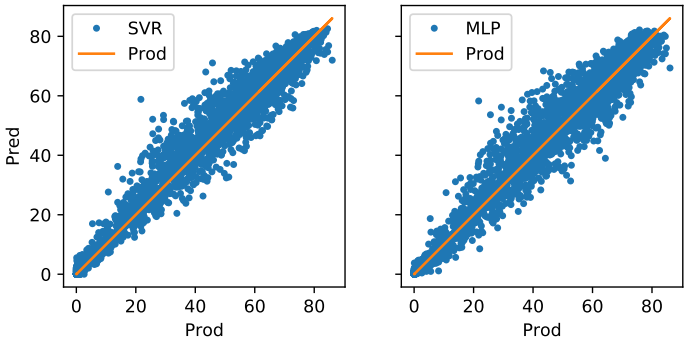


Fig. 3: SVR vs MLP productions-predictions comparison.

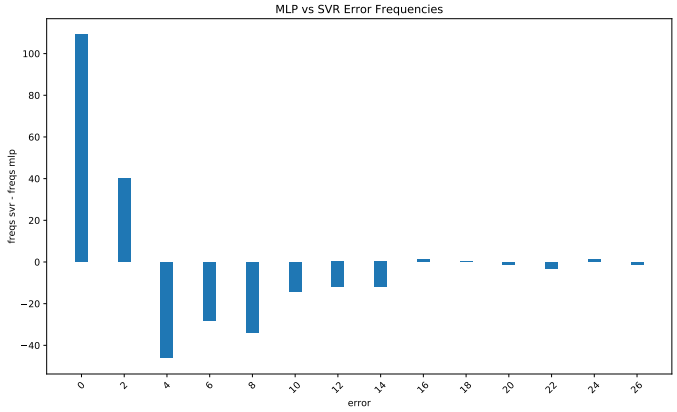


Fig. 4: SVR vs MLP absolute error frequency comparison.

these hourly and monthly errors are very similar to the extracted from the shorter horizons; the worst errors are located in August, because of the very high production, and in unstable months like February, March or December. Not surprisingly, the highest errors are located around the central hours of the day, typically where more energy is produced.

Finally, we show in Figure 3 a comparison between actual PV productions (on the x-axis) and the SVR and MLP predictions for $H = 6$ (on the y-axis); perfectly

Table 6: Gaussian SVR test errors of the `m4` models per hour and month. The blue color corresponds to the minimum (0), and the red color to the maximum (6.19).

Month	Hour												Av.
	9	10	11	12	13	14	15	16	17	18	19	20	
January	2.21	4.71	4.85	4.48	4.20	3.01	2.74	2.17	1.11	0.35	0.00	0.00	2.49
February	4.29	6.04	6.19	4.81	4.82	4.44	3.67	2.98	2.56	1.13	0.46	0.00	3.45
March	4.63	5.34	5.38	5.27	6.02	5.47	4.44	3.37	2.46	1.24	0.49	0.00	3.68
April	4.58	3.21	3.82	3.94	3.39	3.44	3.35	2.97	2.46	1.79	0.77	0.54	2.85
May	3.77	3.54	3.04	2.53	2.51	2.43	2.21	1.90	2.21	1.64	0.84	0.39	2.25
June	2.92	2.70	3.38	2.46	2.15	2.22	2.10	2.09	1.94	2.00	1.29	0.70	2.16
July	2.81	2.20	2.43	1.81	1.78	1.77	2.05	1.99	1.99	1.88	1.57	0.91	1.93
August	3.50	2.85	2.89	3.08	3.58	3.62	3.74	3.50	3.25	2.52	1.44	0.93	2.91
September	4.90	3.84	3.20	4.12	3.62	3.35	3.45	3.17	2.13	1.23	0.83	0.00	2.82
October	4.36	4.86	4.23	4.28	4.02	3.32	2.83	2.86	1.75	0.83	0.00	0.00	2.78
November	2.91	3.78	3.95	2.89	3.08	2.71	2.31	1.74	0.93	0.56	0.01	0.00	2.07
December	2.70	3.84	4.33	3.22	3.38	3.66	4.18	3.59	0.98	0.01	0.00	0.00	2.49
Average	3.63	3.91	3.97	3.57	3.54	3.29	3.09	2.69	1.98	1.26	0.64	0.29	2.66

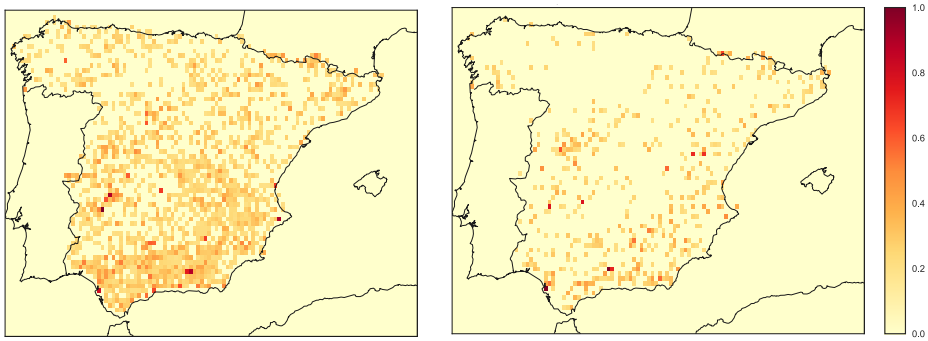
Table 7: Gaussian SVR test errors of the `m6` models per hour and month. The blue color corresponds to the minimum (0), and the red color to the maximum (7.94).

Month	Hour										Av.
	11	12	13	14	15	16	17	18	19	20	
January	5.20	6.07	6.70	6.65	4.45	3.22	1.44	0.31	0.00	0.00	3.40
February	6.88	7.94	7.61	6.15	5.23	4.31	2.71	1.03	0.58	0.00	4.24
March	6.12	6.88	7.39	6.30	6.23	5.21	3.58	1.50	0.63	0.00	4.38
April	3.68	4.16	4.15	4.04	4.10	3.53	2.95	1.70	0.90	0.49	2.97
May	3.66	3.30	3.07	3.38	3.53	3.07	2.77	1.92	1.13	0.47	2.63
June	3.71	3.07	2.76	2.88	3.51	3.34	2.70	1.83	1.20	0.87	2.59
July	2.56	2.41	2.38	2.56	2.57	2.47	2.11	1.54	0.92	0.77	2.03
August	3.11	3.13	3.29	3.76	4.51	4.73	4.32	2.37	0.99	0.81	3.10
September	6.01	5.37	3.92	3.70	4.23	4.53	3.12	1.26	0.49	0.00	3.26
October	6.50	6.85	5.70	5.34	4.14	3.58	2.27	0.51	0.00	0.00	3.49
November	4.07	4.31	4.09	3.10	2.10	2.80	0.97	0.67	0.01	0.00	2.21
December	5.14	4.53	4.08	3.26	2.82	2.46	1.26	0.01	0.00	0.00	2.35
Average	4.72	4.84	4.60	4.26	3.95	3.60	2.51	1.22	0.57	0.28	3.06

prediction would result on diagonal structures. It can be seen how the SVR predictions show a relatively sharper structure, particularly for higher productions. This is illustrated in Figure 4, which shows the differences of the binned absolute SVR and MLP error frequencies; bin sizes are 2%. Notice that the SVR errors are more frequent on the smaller error bins of 0 to 2% and 2% to 4%, where the SVR–MLP frequency differences are positive. On the other hand, MLP errors are more frequent on the larger bins (4% to 6% and above), where frequency differences are negatives.

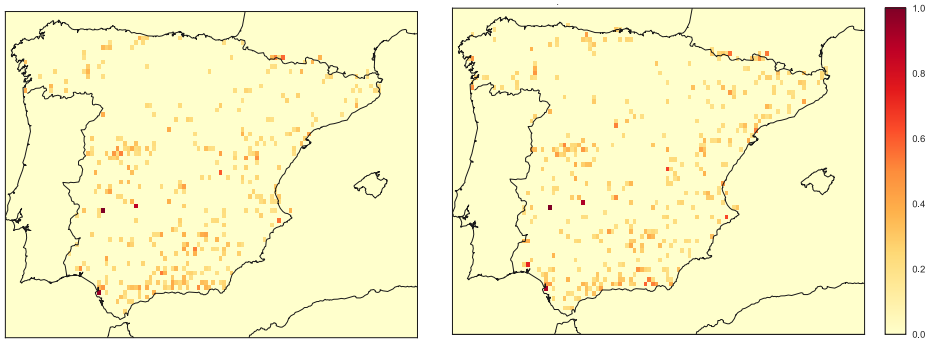
4.3 Model Interpretation

Apart from accurate predictions, it is also important to obtain a reasonably interpretable model in order to understand the underlying phenomena. To do so, we will take advantage of the sparse properties of Lasso, whose global ℓ_1 norm regu-



(a) Group norms for the `m0` model.

(b) Group norms for the `m2` model.



(c) Group norms for the `m4` model.

(d) Group norms for the `m6` model.

Fig. 5: Evolution of the normalized group norms of the coefficients for the horizons 0, 2, 4 and 6 (models `m0`, `m2`, `m4` and `m6`).

larization introduces overall sparsity in the resulting coefficients. For this reasons, Lasso is often also seen as a feature selection model.

We have run this model for all 6 hours horizons (apart from the first 3 already reported for performance comparison) focusing now on the *intensity* of the coefficients. The hyper-parameter of the reported models are given in Table 8, which are pretty similar to those of the previous models. To achieve a more intuitive interpretation we have grouped the CS and satellite variables at each grid point and calculated the Euclidean norm of each such group, which intuitively should give us an indication of the importance of each grid point as a whole. This will also allow us to perform an analysis regarding the evolution of the coefficients across the 6 hours lapse, which may be useful for an even higher level interpretation (not just at an horizon level but also to study the evolution between horizons). To draw the coefficients over the map we have normalized them by their maximum, keeping them in a $[0, 1]$ scale so we can compare them easily and have also truncated those below 0.2. We show the coefficients in Figure 5.

On a per horizon basis we see that the most important points are located in the southern region of the Peninsula, which is the area with the most sunny hours across the year. It is also important to keep in mind that the solar energy

Table 8: Hyper-parameters of the Lasso model for 4, 5 and 6 hours in advance.

Model	Parameter	K		
		4	5	6
Lasso	λ	0.013	0.013	0.014

production in Spain is widely distributed (the greatest percentage of the overall production comes from many small plants), as reflected in the 00 map, which shows many points of low importance and only a few of highest intensity.

At first sight, when considering the evolution across horizons, we see that the number of lower importance points is slowly decreasing as the horizon increases. This may be explained because of the worsening of the Lasso models at the longer horizons, which forces the model to stop fitting finer grained production plants and to focus instead only on the bulkiest production points. Furthermore, we see that the highest intensity points are quite the same across all horizons, located at the western and south western regions of the Iberian Peninsula. Moreover, these points are also located within the middle diagonal section of the Peninsula; the area with the most sunny hours across all 6 horizons according to the daily irradiance evolution over the Iberian Peninsula (we are shifting our data with respect to the sun, so the north west is the area with the least sunny hours).

Summing up, this analysis gives us an idea of the importance of each grid point for the overall solar energy production for every horizon, which may be useful, for instance, to better monitor the more relevant PV installations or to understand the energy generation in a given region.

5 Conclusions, Discussion and Further Work

Satellite-based information has been widely used to nowcast solar irradiance values and PV energy productions, usually from an atmospherical physics perspective. Here we propose the direct exploitation of these readings by feeding them into ML models to nowcast the PV energy production of peninsular Spain, predicting the energy at hour H from Meteosat satellite data at hour $H-K$, where $K=1,2,3$ for a first experiment and then for $K=4,5,6$ for an extended one, to be used for nowcasting forecasting purposes; we also consider $K=0$ for model benchmarking and control. We have worked with the radiances of the `IR016`, `IR039` and `VIS008` channels and the brightness temperature of channel `IR039`, which we downsample from Meteosat's initial finer resolution to a coarser 0.125° one. Moreover, we have added at each point of the resulting grid a theoretical estimation of its Clear Sky (CS) irradiance value and considered four well-known ML models, namely, Lasso, MLP and Linear and Gaussian SVR, plus a simple CS-based persistence approach.

Our results here greatly simplify and improve previous results in [5]; in particular, Gaussian SVRs over satellite and CS features give remarkable results, with rather low mean absolute errors that degrade slowly when the prediction horizon K increases. Other interesting models are MLPs, whose errors also degrade slowly, although their performance is not as good as that of Gaussian SVR. While initially being also relatively good, the performance of Lasso and Linear SVR is clearly below that of Gaussian SVR and MLP. A possible reason for this is proba-

bly the high correlation between the predictor variables: while the Lasso or linear SVR regularization are enough to provide reasonable final models, the intrinsic data dimensionality is probably not large enough to make them powerful enough. Moreover, the homogeneous nature of linear SVR may also hamper its performance. On the other hand, the performance of the CS-based persistence is rather good around noon and on the evenings, although markedly worse in the morning, as it could be expected.

The results here are given for hourly energy prediction updates but this can be easily extended to 15 min updates, provided that satellite readings for that frequency are available. Besides this improvement, there are other areas for further work. An obvious starting point is the nowcasting of the PV production of single farms. Here, purely linear models are not likely to be competitive, but, given the reduced dimensionality, Random Forest and Gradient Boosting should be added to our non-linear model choices. Moreover, the reduced rectangular feature grids to be used for single farms imply that albedo effects will not be present and, thus, convolutional layers could also be considered for neural models. Another clear option is the addition of other features, particularly those of day-ahead NWP radiance predictions, that should help more in the case of longer horizons. On the other hand, this will make pattern dimensions even greater than they are now and some form of feature reduction is likely to be needed. Lasso does precisely that, but as our results show, it does not provide by itself the best forecasts. This suggests to explore Lasso models as a kind of feature selectors that are then further exploited by Gaussian SVRs. In this line, and given the natural grouping of features by grid points, it may also be interesting to use group versions of Lasso, as those proposed in [1]. Finally, it is also important to consider nowcasting over smaller areas, such as individual plants or islands, that may be isolated from larger grids and where PV energy fluctuations are harder to manage. We are currently studying these issues.

Finally, it may be of interest to discuss about the possibility of extending this work to the nowcast of wind energy. Although, in terms of the ML methods involved, there should not be much of a difference between photovoltaic or wind energy, nowcasting requires that forecasting information is updated regularly in time, so that new forecasts (i.e., the nowcasts) are issued with a suitably large frequency. Obviously, the latest readings of the energy produced could be used for nowcasting both wind and PV energy. However, if nothing else is available, these readings are a too simple source of information, that could be exploited essentially only in linear or non-linear AR time series models. While a simple persistence is quite difficult to beat at a one-hour horizon, these AR models degrade quite fast as the autocorrelations of the wind energy time series quickly decrease. Extra exogenous information is thus needed but, other than the already discussed NWP forecasts, this is quite hard to come by in the case of wind energy. Moreover, it is difficult to find any data source or provider that gives new, fresh wind energy related information on an hourly basis, preventing this field from benefiting (at least directly) from the approach described here.

Acknowledgements With partial support from Spain's grants TIN2013-42351-P, TIN2016-76406-P, TIN2015-70308-REDT and S2013/ICE-2845 CASI-CAM-CM. Work supported also by project FACIL–Ayudas Fundación BBVA a Equipos de Investigación Científica 2016, and the UAM–ADIC Chair for Data Science and Machine Learning. The second author was also

supported by the FPU–MEC grant AP-2012-5163. We thank Red Eléctrica de España for useful discussions and making available PV energy data and gratefully acknowledge the use of the facilities of Centro de Computación Científica (CCC) at UAM.

References

1. Carlos M. Alaíz and José R. Dorronsoro. The generalized group lasso. In *Proceedings of the 2015 International Joint Conference on Neural Networks*, pages 1–8, 2015.
2. Javier Antonanzas, Natalia Osorio, Rodrigo Escobar, Rubén Urraca, Francisco Javier Martinez de Pison, and Fernando Antonanzas-Torres. Review of photovoltaic power forecasting. *Solar Energy*, 136:78 – 111, 2016.
3. Javier Antonanzas, David Pozo-Vázquez, Luis Alfredo Fernandez-Jimenez, and Francisco Javier Martinez de Pison. The value of day-ahead forecasting for photovoltaics in the spanish electricity market. *Solar Energy*, 158:140–146, 2017.
4. Chris M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006.
5. Alejandro Catalina, Alberto Torres-Barrán, and José R. Dorronsoro. Satellite based nowcasting of pv energy over peninsular Spain. In *Proceedings of IWANN 2017, International Work Conference in Neural Networks, Lecture Notes in Computer Science 10305*, page 685–697. Springer International Publishing, 2017.
6. Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.
7. François Chollet. Keras: Deep learning library for theano and tensorflow, 2015.
8. EUMETSAT. European European Organisation for the Exploitation of Meteorological Satellites. `http://www.eumetsat.int/`.
9. EUMETSAT. Msg-1/seviri solar channels calibration. *Commissioning Activity Report*, pages 1–39, 2004.
10. Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, , and Chih-Jen Lin. Liblinear: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008.
11. Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *JMLR W&CP: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*, volume 9, pages 249–256, May 2010.
12. Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *JMLR W&CP: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2011)*, April 2011.
13. Annette Hammer, Detlev Heinemann, Carsten Hoyer, Rolf Kuhlemann, Elke Lorenz, Richard Müller, and Hans Georg Beyer. Solar energy assessment using remote sensing technologies. *Remote Sensing of Environment*, 86(3):423–432, August 2003.
14. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning*. Springer, 2009.
15. Pierre Ineichen and Richard Perez. A New Airmass Independent Formulation for the Linke Turbidity Coefficient. *Solar Energy*, 73(3):151–157, 2002.
16. Rich H. Inman, Hugo Pedro, and Carlos Coimbra. Solar forecasting methods for renewable energy integration. *Progress in energy and combustion science*, 39(6):533 – 576, 2013.
17. Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.
18. Jun Kühnert, Elke Lorenz, and Detlev Heinemann. Satellite-based irradiance and power forecasting for the german energy market. In J. Kleissl, editor, *Solar Energy Forecasting and Resource Assessment*, pages 267–297. Academic Press, 2013.
19. Ricardo Marquez and Carlos F.M. Coimbra. Intra-hour DNI forecasting based on cloud tracking image analysis. *Solar Energy*, 91:327 – 336, 2013.
20. Azhar Ahmed Mohammed, Waheeb Yaqub, and Zeyar Aung. Probabilistic forecasting of solar power: An ensemble learning approach. In Rui Neves-Silva, C. Lakhmi Jain, and J. Robert Howlett, editors, *Intelligent Decision Technologies: Proceedings of the 7th KES International Conference on Intelligent Decision Technologies (KES-IDT 2015)*, pages 449–458. Springer, 2015.

21. Daryl R. Myers. Solar radiation modeling and measurements for renewable energy applications: data and model quality. *Energy*, 30(9):1517–1531, 2005.
22. Fabian Pedregosa, Goel Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
23. Photovoltaic Performance Modeling Collaborative Group. The pvlib-python Library. *Sandia National Laboratory*.
24. Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2001.
25. Alberto Torres-Barrán, Álvaro Alonso, and José R. Dorronsoro. Regression tree ensembles for wind energy and solar radiation prediction. *Neurocomputing (in press)*, 2017.
26. Can Wan, Jian Zhao, Yonghua Song, Zhao Xu, Jin Lin, and Zechun Hu. Photovoltaic and solar power forecasting for smart grid energy management. *Journal of Power and Energy Systems*, 1:38–46, 2015.
27. Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83, 1945.
28. Björn Wolff, Jan Kühnert, Elke Lorenz, Oliver Kramer, and Detlev Heinemann. Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. *Solar Energy*, 135:197 – 208, 2016.