International Conference on Renewable Energies and Power Quality (ICREPQ'14)
Cordoba (Spain), 8th to 10th April, 2014
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, No.12, April 2014

Machine Learning Prediction of Global Photovoltaic Energy in Spain

Y. Gala[1], A. Fernández[1] and J. Dorronsoro[1]

[1] Dpto. Ing. Informática, Universidad Autónoma de Madrid
C/ Francisco Tomás y Valiente, n° 11. EPS,
Edificio B, UAM-Cantoblanco. 28049. Madrid.

M. García[2], C. Rodríguez[2]
[2] Red Eléctrica Española (REE)
C. Anabel Segura 11, 28108 Alcobendas, Madrid. Spain.

Abstract

The growing presence of solar energy in the electrical systems of many countries has made its accurate forecasting an important issue. In this work we will explore the application of Support Vector Regression (SVR), an advanced Machine Learning modelling tool, to forecast the daily photovoltaic generation of Spain. Given the very large geographical spread of photovoltaic installations, we will use as input features NWP forecasts of relevant meteorological variables for the entire Iberian Peninsula. The input dimension is thus very large but, while further work is needed, our results show SVR to be an effective tool to deal with the problem's underlying dimension, yield useful forecasts and further provide some insights on the relationship between NWP and actual solar energy production.

Key words
Photovoltaic, energy, radiation, NWP, SVR.

1. Introduction

A natural approach to forecast the energy production over a wide geographical area is to predict first the individual outputs of each installation and then add the results. This has been done, for instance, in wide area wind energy forecasting [5] and, in fact, it may lead to accurate forecasts when geographically diverse installations are considered, as one can expect their individual errors to be relatively uncorrelated and, thus, to partially cancel out when their forecasts are added. This is, however, quite difficult in the case of photovoltaic energy for, while geographic installation diversity is usually the norm, small installation sizes result in a very large number of them, making extremely difficult, if not impossible, the individual forecasting of each installation output. The alternative, to be considered in this work, is the direct forecasts of the aggregated energy output. As in the wind energy case, this can be done building models that transform inputs given by Numerical Weather Predictions (NWP) into energy predictions. These models may have a physical basis or, alternatively, and as it will be the case here, to be simple black box models built using general purpose modelling tools such as feed-forward neural networks or, as done here, Support Vector Machines, widely used in the application of Machine Learning (ML) methods to modelling problems.

2. Support Vector Regression

Multilayer Perceptron (MLPs), i.e., feed-forward neural networks [2], are probably the standard tool in ML modelling. While usually quite effective, their models are difficult to interpret and, moreover, present a risk of model over-fitting when applied to problems with large dimensional inputs. Support Vector Regression (SVR) [13] is a relatively recent modelling alternative that addresses those MLP drawbacks. For instance, their memory requirements are basically independent of pattern dimension, the optimal SVR model is unique and thus they are free of the local minima problems that often affect MLPs, there are efficient training methods and publicly available implementation for them and, through the concept of Support Vectors, it is often possible to interpret the underlying model structure.

More precisely, assume a N pattern sample $S=\{(X_t, y_t): t=1,\ldots,N\}$ where y_t is a target value that we want to approximate as $y_t = f(X_t;W)$ with f an appropriate model parameterized by the weights W.

In the SVR case our goal is to build a linear model $f(X;W)=W\cdot X+b$ where the optimal W^* is found minimizing the following primal criterion function

$$\min_{W,b,\xi} \tfrac{1}{2}\|W\|^2 + C\sum_t(\xi_t+\xi_t^*),$$

subject to the restrictions

$$W\cdot X_t + b - y_t \geq -\xi_t - \varepsilon$$
$$W\cdot X_t + b - y_t \leq \xi_t^* + \varepsilon,$$

with $\xi_t, \xi_t^* \geq 0$ and ε a certain tolerance that we explain next. First notice that we actually have $-\xi_t - \varepsilon \leq W\cdot X_t + b \leq \xi_t^* + \varepsilon$; thus, considering the so called ε-insensitivity cost function

$[z_i]_\varepsilon = \max(0, |z| - \varepsilon)$ it turns out that minimizing the primal criterion is equivalent to minimize

$$\sum_t [y_t - W \cdot X_t + b]_\varepsilon + \frac{1}{C} \| W \|^2 ,$$

where we are allowing for an ε-wide, penalty-free "error tube" around the model function.

Under this light, SVR can be thus seen as a variant of standard ridge regression where we have replaced the familiar $z_i^2 = (y_i - W \cdot X_i - b)^2$ square error by the $[z_i]_\varepsilon$ errors. To build concrete SVR models we do not minimize the primal criterion but, instead, its dual problem of minimizing

$$J(\alpha, \beta) = \frac{1}{2} \sum_{s,t} (\alpha_t - \beta_t)(\alpha_s - \beta_s) X_t \cdot X_s$$
$$+ \varepsilon \sum_t (\alpha_t + \beta_t) - \sum_t y_t (\alpha_t - \beta_t)$$

with the much simpler box constraints $0 \le \alpha_t, \beta_t \le C$.

While simple and elegant, the previous set up would yield linear models possibly not powerful enough. To improve on this, the kernel trick [13] is used, that exploits the fact that only dot products are involved in the dual problem, its solution and the final model that has the form:

$$f(X) = W^* \cdot X + b = \sum (\alpha_t^* - \beta_t^*) X_t \cdot X + b$$

but not the actual patterns X_t. Thus we can work with extended, possibly infinite dimensional patterns $\phi(X)$ provided there is a kernel $K(X, X')$ such that we have $\phi(X) \cdot \phi(X') = K(X, X')$. Here we will use the Gaussian kernel $e^{\frac{-\| X - X' \|^2}{2\sigma^2}}$, which results in a final model;

$$f(X) = b^* + \sum_t (\alpha_t^* - \beta_t^*) e^{\frac{-\| X - X_t \|^2}{2\sigma^2}}$$

reminiscent of the Radial Basis Functions (RBF) models. While there are several proposals to solve dual problem, we will use here the well known LIBSVM [7] software, that represents the state of the art in SVM and SVR solvers.

Finally, we point out that the SVR models always involve the meta-parameters C and ε as well as the parameter σ of the Gaussian kernel. Their correct choice is crucial for the models' performance and to obtain them we will split the overall training data available for model building into a training subset proper upon which models are built for different C, ε and σ choices, and a validation subset where the errors of these different models are computed. The final parameters used are those giving a smallest error on the validation subset. We give the specifics of this procedure in the following section.

3. Wide Area Photovoltaic Energy SVR Forecasting

At the end of 2013 the installed photovoltaic energy in Spain was well above 4 GW but the number of installations is also very large, close to 4,000. Moreover, while the south of Spain is obviously better suited to photovoltaic energy production, installations are actually spread over most of Spain. When trying to predict energy production this makes natural to use for simplicity the entire NWP forecasts over the Iberian Peninsula as the SVR model inputs. We will work with the ECMWF [3] forecasts of two variables, downward surface solar radiation (DSSR) and average total cloud cover (TCC). It is important to notice that the ECMWF forecasts are in fact three-hour aggregated values at UTC times 0 to 21. Thus a first natural goal is to use these ECMWF NWP forecasts to obtain predictions of three-hour aggregated energy production and then to deaggregate them into hourly energy values. This introduces two error levels in the prediction process: a first one due to the prediction system used to obtain the 3-hour aggregated energy forecasts (SVR in our case), and a second one due to the interpolation method used to yield hourly values. This suggests evaluate separately the errors derived from this two step process, which we do in the following sections.

A. *Aggregated Energy Production Forecasts*

We will first consider the error due to the ML forecasting procedure, i.e., SVR here. As mentioned, the 3-hour aggregated energy predictions are the natural goal. ECMWF DSSR forecasts are given as 3-hour accumulated values for UTC hours 0, 3, 6, 9, 12, 15, 18 and 21. We will disregard hours 0 and 3 as they correspond to night-time all year long and, thus, no photovoltaic energy can be produced at these hours (the situation is different for thermosolar energy). Three-hour energy accumulated at hours 6 and 21 is also zero for large parts of the year but we will still keep them. Therefore, we will first predict 3-hour accumulated energy for UTC hours 6, 9, 12, 15, 18 and 21 from the corresponding ECMWF forecasts of DSSR and TCC. The ECMWF forecasting grid for the Iberian Peninsula has 1,128 points and input dimension is thus 2x1,128=2,256. Although production data covers a longer period, we only have ECMWF NWP data from December 2012 to November 2013. This forces some choices when deciding on training, validation and test sets. We will follow the following strategy: we shall build an individual model for each month m that will thus be separately considered for testing. The remaining 11 months will be used for training and validation aiming at selecting the best C, σ and ε parameters. This we do by a validation subset, randomly dividing each 11 month subset in 30% for validation and 70% for train subsets. Notice that training sample size for the "full year" models (in the sense that essentially an entire year is used to predict a given month) is below 6x365 patterns, much the same than the pattern dimension of 1,128 (although with large inter-feature correlations). To treat separately the much smaller energy values at hours 6 and 21, we will build two separate SVR models, one for the 6 and 21 hours and another for the remaining four hours.

We will also study a competing model that starts with the prediction of the aggregated photovoltaic energy for an entire day. Here sample dimension would then be 6x2x1,1128=13,536, now much bigger than sample size, which will be below 365, although again with large inter-feature correlations. We study these daily models for comparison purposes as they provide a useful benchmark against which 3-hour models can be compared. On the one hand, the prediction of daily energy should be an easier problem, as the aggregations of hourly values smooth their fluctuations. On the other, in principle it should be more difficult the hourly deaggregation of total daily energy than that of 3-hour forecasts. As we shall see, and at least for the "full year" models built here and somewhat surprisingly, daily models outperform 3-hourly models; they thus deserve further attention. Moreover, while we use here a unique modelling approach for all days, it may be of interest to consider different prediction venues for different day types. For instance, radiation and, hence, energy values in two consecutive clear sky days are very similar, and a prediction for the second day could just be the production values of the first one. The difficulty to apply such a simple persistence model is, of course, to predict accurately whether the second day will be similar indeed to the first one, and for this aggregated energy predictions could be useful.

Both daily (D) and 3-hourly (3H) models have been built using the LIBSVM library with Gaussian kernels; Table I contains the optimal parameters obtained for each month with the 70-30 approach described above. Notice that instead of σ we give the parameter g used in LIBSVM; they are related as $\sigma = 1/2g$. Thus, when g is approximately $5\cdot10^{-5}$, as in the table, σ will be approximately 100. The 3-hour model parameters are given only for the 4-period model. The parameter values for 6 and 21 hour monthly models, not shown in the table, are in the ranges 13.45 to 76.11 for C, $1.03\cdot10^{-4}$ to $2.44\cdot10^{-4}$ for g and 0.59 for ε.

Table I: Best SVR parameters for each month.

Mon	C_D	g_D	ε_D	C_{3H}	g_{3H}	ε_{3H}
Dec	1217.75	$4.3\cdot10^{-5}$	13.45	2435.5	$5.8\cdot10^{-4}$	1.68
Jan	1217.75	$4.3\cdot10^{-5}$	13.45	2435.5	$5.8\cdot10^{-4}$	2.38
Feb	38967.9	$1.8\cdot10^{-5}$	9.51	2435.5	$5.8\cdot10^{-4}$	2.38
Mar	1217.75	$4.3\cdot10^{-5}$	13.45	2435.5	$5.8\cdot10^{-4}$	2.38
Apr	38967.9	$4.3\cdot10^{-5}$	1.19	2435.5	$5.8\cdot10^{-4}$	4.76
May	38967.9	$4.3\cdot10^{-5}$	9.51	2435.5	$5.8\cdot10^{-4}$	2.38
Jun	1217.75	$4.3\cdot10^{-5}$	13.45	13777.2	$5.8\cdot10^{-4}$	13.45
Jul	1217.75	$4.3\cdot10^{-5}$	19.03	13777.2	$5.8\cdot10^{-4}$	13.45
Aug	38967.9	$4.3\cdot10^{-5}$	13.45	2435.5	$5.8\cdot10^{-4}$	2.38
Sep	1217.75	$4.3\cdot10^{-5}$	0.59	13777.2	$5.8\cdot10^{-4}$	13.45
Oct	1217.75	$4.3\cdot10^{-5}$	13.45	2435.5	$5.8\cdot10^{-4}$	2.38
Nov	38967.9	$4.3\cdot10^{-5}$	19.03	2435.5	$5.8\cdot10^{-4}$	2.38

In order to compare both models' results we introduce the following notation. $E^3_{d,h}$ will denote the accumulated 3-hour energy up to an hour h of the form $h = 3k$, $k = \{2,3,4,5,6,7\}$, of day d and $\hat{E}^3_{d,h}$ its SVR prediction; throughout this work hourly E values will refer to actual hourly energy values normalized as a percentage of installed power. We will thus have $0 \le E_{d,h} \le 100$ for all hourly energy values and similarly $0 \le E^3_{d,h} \le 300$ (although the latter value is usually quite lower). Similarly, E^D_d and $\hat{E}^D_d$ will denote the total energy of day d and its SVR prediction. We will consider predictions of daily energy derived from $\hat{E}^3_{d,h}$ values, that we will denote as $\hat{E}^3_d$ and, correspondingly, we will denote by $\hat{E}^D_{d,h}$ the predictions of 3-hour energy up to an hour $h = 3k$ derived from daily energy predictions $\hat{E}^D_d$. The $\hat{E}^3_d$ predictions are simply obtained as $\hat{E}^3_d = \sum_{k=0}^{7} \hat{E}^3_{d,3k}$; on the other hand, the deaggregation of $\hat{E}^D_d$ into 3-hour predictions $\hat{E}^D_{d,h}$ or that of the 3-hour predictions $\hat{E}^D_{d,h}$ into hourly predictions is more complicated and we will discuss it in the following subsection.

Taking into account the preceding, we will only consider here the errors $e^3_{d,h} = \hat{E}^3_{d,h} - E^3_{d,h}$, $e^3_d = \hat{E}^3_d - E^3_d$ and $e^D_d = \hat{E}^D_d - E^D_d$. The yearly mean absolute errors would then be $me^D = \frac{1}{N}\sum_d |e^D_d|$ with N the number of days, and the similarly defined $me^3 = \frac{1}{N}\sum_d \sum_{k=2}^{7} |e^3_{d,3k}|$, as well as the 3-hourly errors $me^3_{3k} = \frac{1}{N}\sum_d |e^3_{d,3k}|$. Notice that while in principle we should take $0 \le k \le 7$, there is no photovoltaic energy production in the UTC hours between 22 and 3 and we trivially have error zero; thus, we will simply not report them.

It is obvious that seasonality greatly influences photovoltaic energy production and accordingly we will also report the monthly versions me^D_m and me^3_m of the previous values. These values are given in Table II. As it can be seen, the results in the left side of Table II are at first sight better for the daily D models, but notice that the e^3_d values are penalized as we add their absolute values for each 3 hour period as $\widetilde{mae}^3 = \frac{1}{N}\sum_d \sum_{k=2}^{7} |e^3_{d,3k}|$. On the other hand, if we observe the right side of Table II that contains the 3-hour error, the D models give more accurate predictions again. Thus, as it can be seen, D models outperform 3H models.

Table II: Daily and 3-hour interval errors for each month

Months	me_m^D	me_m^3	me_{3k}^D	me_{3k}^3
Dec12	27.89	32.33	5.25	7.61
Jan13	24.67	46.48	6.40	9.13
Feb13	30.98	51.52	7.82	9.89
Mar13	38.42	43.64	9.21	9.51
Apr13	38.18	44.31	8.54	9.90
May13	33.10	37.31	7.60	10.67
Jun13	21.76	44.77	5.47	11.31
Jul13	12.21	53.55	5.51	10.34
Aug13	32.66	32.32	6.35	7.20
Sep13	16.59	28.51	6.05	6.34
Oct13	31.96	40.16	9.26	8.18
Nov13	23.79	37.71	6.64	7.89
Ave	27.35	41.06	7.01	9.00

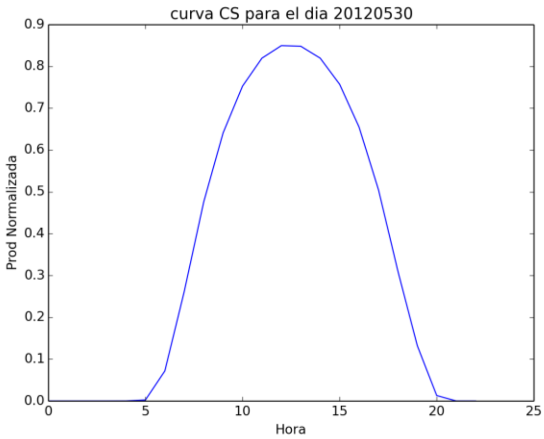
The individual errors $me_{h;m}^3$ for each 3-hour interval for the 3H models are given in Table III. We point out the large errors for the 3-hour intervals ending at hours 9 and 18 in April, May and June and also to the relatively surprising large errors at the hour 12 and 15 intervals for January and February and also November and December. We will discuss this behavior later in this section.

Table III: 3-hour interval MAE error for each month

Months	me_6^3	me_9^3	me_{12}^3	me_{15}^3	me_{18}^3	me_{21}^3
Dec13	0.11	5.52	16.16	17.34	6.43	0.11
Jan12	0.07	4.65	23.09	20.72	6.18	0.07
Feb13	0.13	5.96	24.09	22.10	6.94	0.13
Mar13	0.12	8.98	17.65	20.69	9.43	0.18
Apr13	0.31	9.38	17.09	17.21	14.85	0.56
May13	0.68	18.77	14.53	15.44	13.64	0.97
Jun13	0.91	24.43	10.58	13.10	17.18	1.67
Jul13	0.55	14.53	10.11	8.61	26.97	1.25
Aug13	0.28	8.49	10.65	9.61	13.49	0.65
Sep13	0.13	8.03	10.36	9.52	9.80	0.22
Oct13	0.08	9.61	14.88	16.84	7.58	0.11
Nov13	0.11	6.70	17.92	17.19	5.21	0.21
Ave	0.29	10.42	15.59	15.70	11.47	0.51

B. Hourly Energy Production Forecasts

In order to interpolate daily or 3-hourly energy forecasts into hourly vales, the simplest way to proceed with these interpolations is to consider for each day d a certain interpolation table $p_{d,h}$, $0 \le h \le 23$ such that $\sum_0^{23} p_{d,h} = 1$ and to derive then $\hat{E}_{d,h}^D = p_{d,h} \hat{E}_d^D$ and $\hat{E}_{d,l}^3 = \rho_{d,h-j} \hat{E}_{d,h}^3$ for an hour $l = h - j$ with $h = 3k$ as usual and $j = \{0,1,2\}$ and $\bar{p}_{d,h-j} = \frac{p_{d,h-j}}{\sum_{m=0}^{2} p_{d,h-m}}$.
The key issue is thus the selection of the interpolating values $p_{d,h}$. For a concrete geographic location a natural option could be to take $p_{d,h} = \rho_{d,h}$ with $\rho_{d,h}$ the averaged hourly values of the clear sky radiation curve for that point [1,8,9,10]. Notice that since we will use the averaged version, actual radiation values are not relevant but only the capture of the overall radiation evolution and particularly, the adequate identification of sunrise and sunset. However, accurate clear sky curves are not easy to build [4,6,11,12] and, moreover, radiation values do not translate directly to energy values as this depends heavily on installation and operation characteristics. Finally, even if the above hurdles can be overcome for individual plants, the clear sky curve of the entire Iberian Peninsula simply does not exist as such.



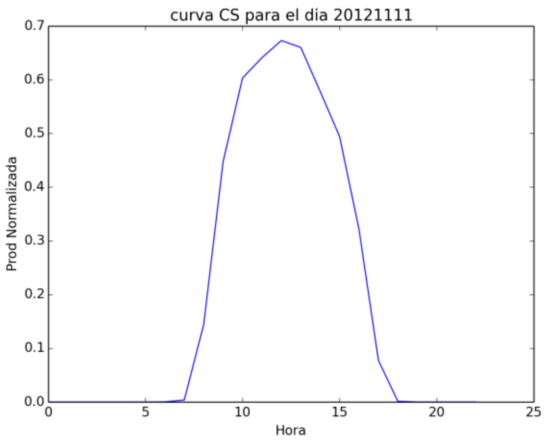


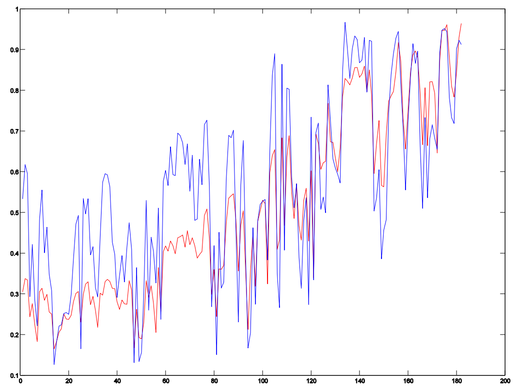
Fig. 1: Clear sky curve for May 30 and November 11.

On the other hand, photovoltaic energy production itself could be used to build a clear sky curve proxy. We find this an interesting venue that, nevertheless, will require still some work to arrive to a full development. As a first approach, we will simply compute here for each pair (d,h) a "maximum energy curve" $\mu_{d,h}$ defined as

$$\mu_{d,h} = \max_{y,q} \{E_{d+q,h,y} \_ -\delta \le q \le \delta, y\},$$

where δ is some small integer and $E_{d+q,h,y}$ denotes the energy produced at hour h in all $d+q$ days of a year y. In other words, $\mu_{d,h}$ is the maximum of the normalized energy productions registered at hour h and any day in the interval $[d-\delta, d+\delta]$ all years with photovoltaic energy production records (in our case from June 2011 to November 2013). Figure 1 depicts the μ curves for May 30 (with a rather smooth curve) and November 11 (which is more difficult to interpret). In any case, we stress that other, perhaps better criteria could be devised to interpolate 3-hourly or daily energy predictions.

The errors considered here will be $e^{H3}_{d,h} = \hat{E}^{H3}_{d,h} - E_{d,h}$ and $e^{HD}_{d,h} = \hat{E}^{HD}_{d,h} - E_{d,h}$. By the reasons discussed above, we shall disregard the $h = \{0,1,2,3,22,23\}$ hours, as there is essentially no energy production for them (summer production for hour 21 is also very low).



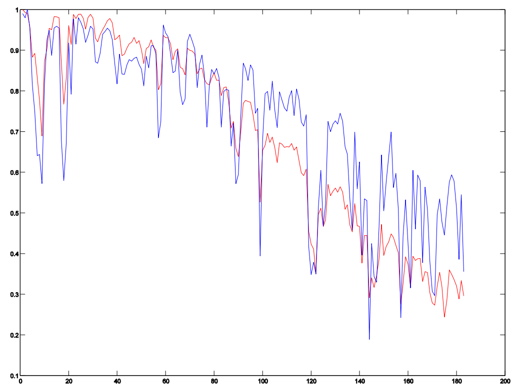


Fig. 2: Radiation (red) vs production (blue) from Dec. 2012 to May 2013 (top) and from June 2013 to Nov. 2013 (bottom).

Table IV: Hourly MAE error for each month.

Months	me^{HD}_h	me^{H3}_h
Dec13	1.83	2.64
Jan12	2.31	3.22
Feb13	2.84	3.55
Mar13	3.29	3.43
Apr13	2.98	3.43
May13	2.67	3.65
Jun13	1.89	3.80
Jul13	1.89	3.53
Aug13	2.17	2.52
Sep13	2.15	2.41
Oct13	3.26	3.16
Nov13	2.34	2.81
Ave	2.48	3.18

Because of this we will report the yearly mean absolute errors for them as $me^{HD} = \frac{1}{16N} \sum_d \sum_{h=4}^{21} | e^{HD}_{d,h} |$ and define similarly me^{H3}. Their respective values are $me^{HD} = 2.48$ and $me^{H3} = 3.18$. We shall also consider their monthly counterparts me^{HD}_m and me^{H3}_m that are given in Table IV. As it can be seen, the performance of both D and 3H models are rather similar, with the D models being slightly better.

4. Conclusion

Global photovoltaic energy forecasts are of growing interest in countries such as Spain where solar energy already has a sizeable presence and a clear potential for further expansion. While an approach where individual installation forecasts are added to get a global forecast is likely to result in the partial cancellation of individual errors and, thus, in accurate predictions, such an approach can be quite difficult when there are many and very scattered, low power installations. Direct global models are then a natural option and, as shown in this work, this leads to fairly good forecasts of daily and three-hour aggregated energy that can be then deaggregated into hourly values. However, and as already mentioned, more work is still needed. First, although they share a common trend, the concrete relationship between energy and radiation (the most important NWP variable) varies seasonally in a noticeable way. This can be seen in Figure 2 that depicts daily energy production and average radiation prediction, both normalized to the 0-1 range. In the figure both curves show a common trend and a clear correlation; however while they appear closer in the summer; radiation is well below production in winter. In other words, it seems difficult that "full year" models as the ones used here can predict accurately both winter and summer radiation. In fact, it is likely that the weight of the days around the spring and fall equinoxes, whose number doubles those of the days around the summer and winter solstices, should be better modelled by a "full year" model, and this is what seems to happen to our models, that infrapredict production in the lower radiation winter days and overpredict it in the higher radiation summer days.

Thus the selection of training subsets in order to predict a given time period is of great importance and has to be further studied. Moreover, how to deaggregate three hourly energy predictions also needs to be further studied, with our proxy clear sky curve being a reasonably good first step. Finally, the easier energy prediction on basically clear sky days also suggests that single all day types models may be improved using models more tailored to a given day general characteristics; D-type models may give accurate energy predictions for an entire day and, thus, detect whether a given day's energy can be accurately predicted from that of the previous one. We are currently studying these and other related issues.

Acknowledgement

With partial support from Spain's grant TIN2010-21575-C02-01 and the UAM-ADIC Chair for Machine

Learning. The first author is also supported by an UAM-ADIC Chair grant.

References

[1] R.E. Bird and R.L. Hulstrom. Simplified clear sky model for direct and diffuse insolation on horizontal surfaces. Technical Report No. SERI/TR-642-761, Solar Energy Research Institute, Golden, CO, 1981.

[2] R. Duda and et al. Pattern Classification. Wiley, 2000.

[3] ECMWF. European Center for Medium Range Weather Forecasts. http://www.ecmwf.int/, 2005.

[4] B. Espinar and et al. Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations. Solar Energy, 83(1):118-125, 2009.

[5] G. Giebel. On the benefits of distributed generation of wind energy in Europe. PhD thesis, Ossietzky University, Oldenburg, Germany, 2000.

[6] C. Kandilli and K. Ulgen. Solar illumination and estimating daylight availability of global solar irradiance. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 30:1127-1140, 2008.

[7] LIBSVM. A library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

[8] P. Mathiesen and J. Kleissl. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. Solar Energy, 85:967-977, 2011.

[9] A. B. Meinel and M. P. Meinel. Applied Solar Energy. Addison Wesley Publishing Co., 1976.

[10] D. R. Myers. Solar radiation modeling and measurements for renewable energy applications: data and model quality. Energy, 30(9):1517-1531, 2005.

[11] R. Perez and et al. Validation of short and medium term operational solar radiation forecasts in the US. Solar Energy, 84(12):2161-2172, 2010.

[12] J. Remund, R. Perez, and E. Lorenz. Comparison of solar radiation forecasts for the USA. In European PV Conference, Valencia, Spain, 2008.

[13] B. Scholkopf and A.J. Smola. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, 2001.