© 2013 IEEE.

Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.

Digital Object Identifier: 10.1109/SusTech.2013.6617293
Constrained Support Vector Machines for Photovoltaic In-Feed Prediction

Marcus Hildmann
Power Systems Laboratory, ETH Zürich, Switzerland
Email: hildmann@eeh.ee.ethz.ch

Abhishek Rohatgi
Energy Studies Institute, National University of Singapore, Singapore
Email: esiar@nus.edu.sg

Göran Andersson
Power Systems Laboratory, ETH Zürich, Switzerland
Email: andersson@eeh.ee.ethz.ch

Abstract—In this paper, we introduce a constrained Support Vector Machine (SVM) to predict photovoltaic (PV) in-feed. We derive the SVM algorithm with linear constraints and test the method on German PV in-feed with constraints reflecting physical boundaries. We show that the new algorithm shows a significant better performance than a constrained ordinary least squares (OLS) estimator.

I. INTRODUCTION

In many countries government support schemes for renewable energy sources (RES) such as feed-in tariff (FIT) and prioritization of electric power in-feed from RES, have resulted in large scale deployment of RES, mainly of intermittent sources such as wind and photovoltaic (PV) and in an overall increase of RES production. This heavily effects the power markets and the grid operation; which results in a need for precise forecasts of RES in-feed, for spot and intra-day trading purposes on the one hand and for grid operation and ancillary services on the other.

Two main approaches to model and predict the RES in-feed are available, fundamental models and statistical models. Fundamental models use fundamental mechanics based on solar irradiation and cloud cover to model and predict the in-feed, but suffer from an overwhelming complexity. Statistical methods such as regression methods and artificial neural networks (ANN) are easier to implement and can separate the deterministic signal from the noise using simple models based directly on the time series.

In literature, several linear and nonlinear approaches for PV modeling and prediction are common. Reference [1] uses a weather conditioned moving average model. The developed model is shown to be more accurate than other prediction algorithms such as exponentially weighted moving average (EWMA). The approach incorporates the seasonality through weather, no additional seasonality factors are used. Reference [2] explains the shortcomings of parametric models like polynomial regression for modeling Solar Energy. The proposed method is based on applying nonparametric regression and related techniques on the external weather prediction. Reference [3] shows a short term prediction model up to three hours using k-nearest neighbors based data mining methods. In [4], authors have presented a non-linear multivariate regression method based on correlation analysis of data mined from a historical weather parameters, statistical indices and energy consumption. The proposed hybrid model gives smaller mean error than the multiple regression method. References [5]–[7] present PV forecasting using Support Vector Machine (SVM) on several time horizons and external variables such as weather forecasts that are classified into clear sky, cloudy day, foggy day and rainy day.

In this paper we formulate a constrained SVM method which is also computationally efficient. To do so the following characteristics of PV in-feed must be captured. Fig. 1 shows the typical PV in-feed with the strong daily patters and zero in-feed during night. This strong pattering with the zero elements present a hard challenge for the estimation algorithms, since exact zero values are hardly achievable without boundary constraints. While it is possible to remove a non-zero estimation, especially negative values, during the night in an additional post processing steps, an estimation procedure which automatically keeps a zero bound has to be favored over additional post-processing steps.

Additional constraints are easy to implement in combination with linear regression methods, but linear regression such as ordinary least squares (OLS) cannot use time variate factors. Non-linear methods such as ANN are common methods to overcome this problem. However, methods such as ANN suffer from over-fitting and non-convexity. SVM is a state of the art...
machine learning algorithm which can be solved as a convex optimization problem [8]. To tackle the problem of overfitting, while keeping the nonlinear characteristics of the model and the possibility of high degrees of freedom, additional constraints to force statistical models into the right direction by including lower and upper bounds of the estimation variables are implemented.

We present a formulation of an SVM with linear constraints and use this to predict PV in-feed. The additional constraints allow a lower bound of zeros to ensure a realistic profile, since the zero line is guaranteed by the underlying physics. In addition other external constraints can be incorporated to include additional information, such as maximum installed capacity or fundamental weather information from a solar radiation model.

The paper is organized as follows: Section II and III describe the proposed algorithm and model for PV in-feed. Section IV defines the testing set, section V show the results and finally, section VI concludes the results.

II. SVM WITH EXTERNAL CONSTRAINTS

Since PV in-feed has underlying physical processes and technical conditions such as a minimum zero in-feed and a maximum in-feed of the installed capacity, a SVM algorithm which can replicate the external constraints out of sample is necessary. The SVM optimization problem with additional linear constraints will be described in this section. The method is based on the classic SVM and random feature space [8]–[10].

To solve the estimation problem with constraints, it is required to include the external constraints during the formulation of the optimization problem of SVM. The primal optimization problem of a standard SVM is given as:

\[
\begin{align*}
\min_{w,b,\xi^*, \xi} & \quad J_p(w) = \frac{1}{2}w^Tw + c\sum_{k=1}^{N}(\xi_k + \xi_k^*) \\
\text{s.t.} & \quad y_k - w^T\phi(x_k) - b \leq \epsilon + \xi_k, k = 1, \ldots, N \\
& \quad w^T\phi(x_k) + b - y_k \leq \epsilon + \xi_k^*, k = 1, \ldots, N \\
& \quad \xi_k, \xi_k^* \geq 0, k = 1, \ldots, N
\end{align*}
\]

where \(\{x_k, y_k\}_{k=1}^{N}\) is the training set, \(N\) is the number of training points and \(\phi\) is the feature space. \(c\) is the regularization parameter for the slack variables \(\xi_k\) and \(\xi_k^*\). The dual form of (1) is given as:

\[
\begin{align*}
\max_{\alpha_0, \alpha^*} & \quad J_d = \frac{1}{2} \sum_{k=1}^{N} \sum_{l=1}^{N} (\alpha_k - \alpha_k^*)(\alpha_l - \alpha_l^*)K(x_k^T x_l) \\
& \quad - \epsilon \sum_{k=1}^{N} (\alpha_k + \alpha_k^*) + \sum_{k=1}^{N} y_k (\alpha_k - \alpha_k^*) \\
\text{s.t.} & \quad \sum_{k=1}^{N} (-\alpha_k + \alpha_k^*) = 0 \\
& \quad \alpha_k, \alpha_k^* \in [0, c]
\end{align*}
\]

where \(\alpha_k\) and \(\alpha_k^*\) are lagrange multipliers. \(K\) is the kernel and is defined as \(K(x_k, x_l) = \phi(x_k)^T \phi(x_l)\). During the transformation of the optimization problem from primal to dual ((1) to (2)), the non-linear effects are moved to the kernel and (2) becomes a convex optimization problem. The estimated function can then be written as

\[
f(x) = \sum_{k=1}^{N} (\alpha_k - \alpha_k^*)K(x, x_k) + b
\]

The output is written only in terms of the lagrange multipliers and the kernel function. Hence, for estimation problems, one does not need to know the underlying feature space \(\phi(x)\). To solve a time series estimation problem with external constraints using SVM theory, the primal optimization problem of a SVM is modified as

\[
\begin{align*}
\min_{w,b,\xi, \xi^*} & \quad J_p(w, \xi, \xi^*) = \frac{1}{2}w^Tw + c\sum_{k=1}^{N}(\xi_k + \xi_k^*) \\
\text{s.t.} & \quad y_k - w^T\phi(x_k) - b \leq \epsilon + \xi_k, k = 1, \ldots, N \\
& \quad w^T\phi(x_k) + b - y_k \leq \epsilon + \xi_k^*, k = 1, \ldots, N \\
& \quad \xi_k, \xi_k^* \geq 0, k = 1, \ldots, N \\
& \quad w^T\phi(x_\tau) + b \leq \psi, \tau \in \mathbb{S}, \mathbb{S} \subseteq [N + 1, \ldots, M]
\end{align*}
\]

where \(\tau\) indicates the future time for which we want to add constraints. \(M\) is the forecast period and \(\psi\) defines the constraint. The other parameters are defined along (1). One
way to solve the primal problem (4) is to convert it in the dual problem as given below:

\[
\max_{\alpha, \alpha^*, \gamma} \quad J_d = -\frac{1}{2} \sum_{k=1}^{N} \sum_{l=1}^{N} (\alpha_k - \alpha_k^*)(\alpha_l - \alpha_l^*) K(x_k, x_l) \\
+ \sum_{k=1}^{N} \sum_{\tau=N+1}^{M} (\alpha_k - \alpha_k^*) \gamma_\tau K(x_k, x_\tau) \\
- \epsilon \sum_{k=1}^{N} (\alpha_k + \alpha_k^*) + \sum_{k=1}^{N} y_k (\alpha_k - \alpha_k^*) - \sum_{\tau=N+1}^{M} \gamma_\tau \psi
\]

(5)

s.t. \quad \sum_{k=1}^{N} (-\alpha_k + \alpha_k^*) + \sum_{\tau=N+1}^{M} \gamma_\tau = 0

\alpha_k, \alpha_k^* \in [0, c]

where \(\alpha_k, \alpha_k^*\) and \(\gamma_\tau\) are the Lagrange multipliers. \(K\) is the kernel. However, in order to solve the constrained estimation as the dual problem, it is required to explicitly formulate the dual problem every time constraints are added or modified. Additionally, it might not be possible to write the dual problem in terms of kernel functions for every type of constraint. Reference [10] have proposed a way to solve SVM by explicitly defining the feature space. This feature space is called random feature space as the parameters used to define it can be selected randomly. It can be regarded as merging of the Extreme Learning Machine (ELM) and SVM approaches. In ELM, the input vectors are mapped to the hidden layer neurons by a randomly generated matrix [11]. This is analogous to defining a new feature space where the hidden layer acts as a transformation from the input vector space to the hidden neurons space. For example, assuming a sigmoidal function, the feature space can be defined as following:

\[
\phi_i(x_k) = \frac{1}{1 + \exp(-w_i x_k - b)}, \quad i = 1, \ldots, h
\]

\[
\phi(x_k) = \left[ \phi_1(x_k) \quad \phi_2(x_k) \quad \ldots \quad \phi_h(x_k) \right]
\]

(6)

The mapping \(\phi(.): \mathbb{R}^n \rightarrow \mathbb{R}^h\) takes the input vector \(x_k \in \mathbb{R}^n\) to the \(h\)-dimensional space \(\mathbb{R}^h\), where \(h\) is the dimension of the high dimensional feature space. In [12], Liu also proposes to use explicitly defined feature spaces to form an Extreme Support Vector Machine (ESVM). The optimization problem in (4) can now be solved without kernels as the feature mapping is known. Knowing the feature mapping enables inclusion of the external constraints in the optimization problem rather than solving the dual problem that utilizes the kernels but needs to be formulated every time the constraints are changed. In [10], ELM kernel has been introduced based on random feature spaces that can enable to use the Fixed Size SVM approach in case of large data sets. The ELM kernel [10] is defined as

\[
k(x_k, x_l) = \frac{1}{p} \phi_k x_k \phi_l
\]

where \(\phi\) is defined as in (6).

This paper uses the SVM based on the random feature space that allows to the optimization problem to be solved without explicitly converting it to its dual form since the kernel function does not need to be defined and the optimization problem in (4) can now be solved as the feature mapping is known.

III. PV In-feed Forecast Model

The underlying equation for the PV in-feed is given by

\[
y_t = \bar{y}_t + \epsilon_t
\]

(7)

where \(y_t\) is the PV in-feed with non-stationary average \(\bar{y}_t\) and \(\epsilon_t\) is a white noise (WN) process with zero expectation and a finite variance \(\epsilon \in (0, \sigma^2)\). The prediction of the PV in-feed \(\hat{y}_t\) is given by

\[
\hat{y}_t = \bar{y}_t + \chi_t
\]

(8)

where \(\hat{y}_t\) is the predicted PV in-feed with average \(\bar{y}_t\) and the noise term \(\chi_t\), where \(\chi_t\) is a WN process with zero expectation and a finite variance \(\chi_t \in (0, \sigma^2)\). The estimated average is written as the function of a regression vector \(x_t \in \mathbb{R}^n\)

\[
\bar{y}_t = f(x_t)
\]

(9)

Fig. 2. Cross correlation of PV time-series

A. Characteristics of PV In-Feed

PV in-feed depends on a variety of factors. The amount of diffused radiation received locally on earth is correlated to temperature of the region. It can also be correlated to the precipitation. Additionally, PV time series of two given years shows significant degree of cross correlation. PV in-feed show multi scale seasonality, too. To select a set of external variables \(x\) for the regression, we analyze the main characteristics of PV in-feed.
Seasonality on Different Frequencies: PV in-feed shows different seasonality frequencies. Fig. 1 shows the PV shows the typical intra-day seasonality with the peak during mid day and zero in-feed during the night. The figure also shows the different patterns form March 2012 to June 2012, were the duration of the in-feed and also the level increases from spring to summer.

Temperature: Temperature of a region is a proxy for diffused radiation received from the sun during a given day and the correlation between PV in-feed and temperature can be used to model the in-feed. The correlation between PV time series and temperature is shown in Fig. 4.

B. Model for PV In-Feed

Based on the analysis of the PV characteristics with cross correlation and external inputs, we propose a non-linear autoregressive with external inputs (NARX) model to model PV in-feed. The input factors for the NARX model for PV in-feed consist of:

1) Autoregressive terms: The amount of lag is kept as variable
2) Past data: Previous two years of PV in-feed on hourly basis.
3) Weather data: maximum temperature $T_{t_{max}}$, minimum temperature $T_{t_{min}}$, mean temperature $T_{t_{mean}}$, heating days $Hd_t$, cooling days $Cd_t$ and precipitation $PP_t$ and are grouped as $Wea_t = [T_{t_{max}} T_{t_{min}} T_{t_{mean}} Hd_t Cd_t PP_t]$
4) Seasonality factors: monthly and intra-day seasonality are captured by indicator variables [13].

IV. TEST SET

As all load data, PV is provided on 15 minute frequency by TenneT®. In this paper, the calculations are done on hourly basis. We test the model on a full year of PV in-feed from March 2012 until Feb 2013 as shown in Fig. 5. The lag for the autoregressive (AR) terms is 24 hours. The period March 2012 to Jan 2013 is the testing set for in-sample training. The prediction period for out of sample prediction is February 2013.
TABLE I
SVM TUNING PARAMETERS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>h</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
<td>5000</td>
<td>1000000</td>
</tr>
</tbody>
</table>

mean square error (MSE) and mean absolute prediction error (MAPE).

V. RESULTS

In this section, we first present the results of the constrained SVM and also a comparison to a constrained OLS regression method. Fig. 6 shows the in-sample prediction of the constrained SVM and the constrained OLS regression. For the in-sample fit, both algorithms work without constraints. The tuning parameters for SVM are the number of hidden nodes $h$ (refer (6)) and the regularization factor $c$ (refer (4)). The number of hidden nodes $h$ are shown to have negligible effect as long as $h$ is comparable to the size of in-sample data [11]. This leaves only parameter $c$ which is kept large to minimize the slack variables $\xi_k$ and $\xi_\star_k$. The value of tuning parameters used for the model are shown in Table I. Table II show the MSE and the MAPE of the in-sample fit for both algorithms. The in-sample fit of the SVM is naturally better than the OLS fit, since the SVM has a higher degree of freedom than the linear OLS, but also does not result in an overfit.

![In Sample Simulation Results of PV Time-Series](image)

For the out-of-sample prediction, the constraint of no negative prediction will be activated. Fig. 7 shows the results of the out-of-sample predictions of both constrained SVM and constrained OLS. It is clear to see, that the nonnegative constraint of the algorithm holds, there are no negative predictions. The figure also shows the realized PV in-feed. The SVM does not show a clear trend of systematic over or underestimation of PV in-feed, while the constrained OLS overpredicts the in-feed most of the time. Table II shows the statistical measures of the out-of-sample prediction. In both measures, the SVM clearly outperforms the OLS method.

![Out of Sample Simulations of PV Time-Series for Feb 2013](image)

VI. CONCLUSION

In this paper a constrained SVM was derived and used to predict German PV in-feed with the constraint, that the in-feed cannot be negative. We show that the algorithm works and the constraints hold. The method shows a performance improvement of around 18% in both measures (MAPE and MSE) compare to the constraint OLS which was used as benchmark and will increase the time series based PV forecast quality.

ACKNOWLEDGMENT

Financial support from Swiss Innovation Promotion Agency (KTI), project 10851.2, is gratefully acknowledged. The authors thank Donnacha Daly for his help finishing the paper.

REFERENCES


