Solar power forecasting with sparse vector autoregression structures

Laura Cavalcante, Ricardo J. Bessa
INESC Technology and Science (INESC TEC)
Porto, Portugal
laura.l.cavalcante@inesctec.pt, ricardo.j.bessa@inesctec.pt

***Abstract*—The strong growth that is felt at the level of photovoltaic (PV) power generation craves for more sophisticated and accurate forecasting methods that could be able to support its proper integration into the energy distribution network. Through the combination of the vector autoregression model (VAR) with the least absolute shrinkage and selection operator (LASSO) framework, a set of sparse VAR structures can be obtained in order to capture the dynamic of the underlying system. The robust and efficient alternating direction method of multipliers (ADMM), well known for its great ability dealing with high-dimensional data (scalability and fast convergence), is applied to fit the resulting LASSO-VAR variants. This spatial-temporal forecasting methodology has been tested, using 1-hour and 15-minutes resolution, for 44 microgeneration units time-series located in a city in Portugal. A comparison with the conventional autoregressive (AR) model is performed leading to an improvement up to 11%.**

***Index Terms*—Forecasting, scalability, solar power generation, sparse matrices.**

I. Introduction

The variable nature of many renewable energy sources poses a number of challenges as to its proper connection and integration in electric power systems. It is of utmost importance for decision-making operators to have accurate forecasts of renewable energy generation in order to plan the daily operations and select optimal strategies in the power system.

The forecasting techniques applicable in this context are classified according to the pretended forecast horizon. For the very short-term horizon (up to 6 hours ahead), which is addressed in this paper, statistical models considering the latest measurements are assumed appropriate. Within this framework, recent studies show that the use of a spatial-temporal approach, taking advantage of the dispersed information obtained by geographically distributed sensors, has revealed to be promising (see e.g. [1]). In addition, the implementation of adequate sparsification strategies is very important in order to select the relevant input variables and achieve a sparse solution (coefficient matrix only have a small number of nonzero entries) with less computational cost. In an attempt to explore sparse structures of the spatial-temporal relations different methodologies have been proposed (see e.g. [2], [3] and [4]). Also with this purpose, and inspired by [5], [6] proposed a variety of sparse structures (of the coefficient matrix) for vector autoregression (VAR) model using the *least absolute shrinkage and selection operator* (LASSO) framework [7] and applied the *alternating direction method of multipliers* (ADMM) [8] to fit them. The authors implemented it to wind power forecasting and achieved good results compared to competitive models. This paper explores the same methodology but applied to solar power forecasting, addressing the necessary adjustments to take into account in this new context. A comparative analysis, in terms of improvement over autoregression (AR) model, of the different proposed structures and between them and two models considered in [1], VAR fitted with *ordinary least squares* (OLS) and *gradient-boosting* (GB) algorithm, will be performed under a case-study with 44 microgeneration units in the same control area.

The paper is organized as follows. Section II provides a general description of the spatial-temporal forecasting methodology. The obtained test case results are presented and discussed in Section III. Finally, the paper ends with concluding remarks and future work in Section IV.

II. Spatial-Temporal Forecasting Methodology

The forecasting methodology employed is proposed in [6] and allows, through the combination of the VAR and LASSO approaches to take advantage of the spatially distributed information and, simultaneously, to perform a smart selection of the relevant information, thereby enabling to detect the spatial-temporal dynamics of the power generation. The different resulting sparse structures LASSO-VAR (LV) are fitted using the ADMM, that allows to solve large-scale problems in a reduced computational time.

A. LASSO-VAR Framework

The VAR model arises as a natural extension of the univariate AR model, linking and cointegrating the time series variables in a multivariate system which allows that variations to one variable to propagate to the others.

The VAR process of order p (VAR[p]) evolves according to

$$Y_t = \eta + \sum_{l=1}^{p} B^{(l)} \cdot Y_{t-l} + e_t, \tag{1}$$

in which $\{Y_t\} = \{(y_{1,t}, y_{2,t}, \ldots, y_{k,t})'\}$ denote a k-dimensional vector time series, η is a vector of constant terms, each $B^{(l)} \in \mathbb{R}^{k \times k}$ represents a coefficient matrix related to the lag l and $e_t \sim (0, \Sigma_e)$ denotes a white noise disturbance term.

This expression, relating the future observations at each of the k microgeneration units to the past observations of all units in the model, can be written in the compact matrix notation

$$Y = \eta \mathbf{1}' + BZ + E, \tag{2}$$

where $Y = (Y_1, Y_2, \ldots, Y_T)$ define the $k \times T$ response matrix, $B = \left(B^{(1)}, B^{(2)}, \ldots, B^{(p)}\right)$ the $k \times kp$ matrix of coefficients, $Z = (Z_1, Z_2, \ldots, Z_T)$ the $kp \times T$ matrix of explanatory (or predictors) variables, in which $Z_t = (Y'_{t-1}, Y'_{t-2}, \ldots, Y'_{t-p})$, $E = (e_1, e_2, \ldots, e_T)$ the $k \times T$ error matrix, and with $\mathbf{1}$ denoting a $T \times 1$ vector of ones.

In this context, through the use of least squares (LS) statistical methodology, it is possible to estimate the unknown coefficients capturing contemporaneous dependencies among the variables and thus getting the model that best describes the data. To simplify the calculations, centered variables Y and Z will be assumed and, consequently, the intercept η will no longer appear in the least squares objective function and will be estimated after the model has been fitted.

Due to a quadratically growing dimension of the parameter space inherent to VAR model, a consistent prediction is only possible through the use of a procedure that induces a low-dimensional structure on the underlying model. The LASSO framework is very convenient to deal with estimating a high-dimensional network since it simultaneous performs variable selection and produces a sparse solution. This is achieved through a regularized version of least squares that introduces L_1 penalties on the coefficients promoting a sparsity structure on the model space. The standard LASSO-VAR (sLV) loss function is expressed as

$$\frac{1}{2} \|Y - BZ\|_2^2 + \lambda \|B\|_1, \tag{3}$$

where $\lambda > 0$ is a scalar regularization (or penalty) parameter controlling the amount of shrinkage, and $\|.\|_r$ represents both vector and matrix L_r norms. Different penalties can be used in order to reach the goal of reducing the effective dimensionality of the problem and detect different sparsity patterns accordingly to the inherent structure of the VAR. The efficient use of appropriate penalties lead to more accurate estimation and forecasting strategies.

The Table I presents a brief description of following sparsity-promoting LV structures: standard LASSO-VAR, row LASSO-VAR (rLV), lag-group LASSO-VAR (lLV), lag-sparse-group LASSO-VAR (lsLV), own/other-group LASSO-VAR (ooLV) and causality-group LASSO-VAR (cLV). The different penalties applied to them induce different types of sparsity, depending on the selection target by which they are managed. While the rLV and sLV deal with model's coefficients individually resulting in an unstructured sparsity pattern, the remaining structures lLV, lsLV, ooLV and cLV, look through sparsity in distinct group structures highlighting characteristics such as lag selection, group and within-group sparsity, delineation between a component's own lags and those of another component and evaluate which variables add forecast improvement, respectively. More details about these structures can be found in [5] and [6].

Table I: Brief description of LASSO-VAR structures

LV St.	Penalty	Selection target
rLV	$\lambda \left\|B^i\right\|_1$	individual entries by row
sLV	$\lambda \|B\|_1$	individual entries
lLV	$\lambda \sum_{l=1}^{p} \|B_l\|_1$	lags
cLV	$\lambda \sum_{i \neq j} \|(B_1)_{ij} \ldots (B_p)_{ij}\|_2$	locations (causality)
lsLV	$(1-\alpha)\lambda \sum_{l=1}^{p} \|B_l\|_F + \alpha\lambda \|B\|_1$	lags and individual entries within lags
ooLV	$\sqrt{k}\lambda \sum_{l=l}^{p} \|\text{diag}(B_l)\|_2 + \sqrt{k(k-1)}\lambda \sum_{l=l}^{p} \left\|B_l^-\right\|_2$	lags diagonal (diag(B_l)) and off-diagonal entries (B_l^-)

B. ADMM Fitting

The ADMM has attracted renewed attention recently due to its applicability to optimization problems arising from many applications, and the relative ease with which it may be implemented in parallel and distributed computational environments, which is very advantageous when dealing with high-dimensional problems. Moreover, given the superior convergence properties and decomposability of this robust method, problems with nondifferenciable constraints (such as LASSO) can be easily handled. This section provides a brief introduction to the analytical underpinnings of the method. Recall the standard LASSO-VAR objective function in (3) and rewrite it in ADMM form as

$$\begin{aligned} &\text{minimize} && \frac{1}{2} \|Y - BZ\|_2^2 + \lambda \|H\|_1 \\ &\text{subject to} && B - H = 0. \end{aligned} \tag{4}$$

The ADMM can be viewed as a version of method of multipliers with the difference that it splits the objective function in two distinct objective functions, $f(B) = \frac{1}{2} \|Y - BZ\|_2^2$ and $g(H) = \lambda \|H\|_1$, by replicating the B variable in the H variable and adding an equality constraint imposing that these two variables are equal. The corresponding ADMM algorithm consists of the following iterations:

$$B^{k+1} := \arg\min_B \left(\frac{1}{2} \|Y - BZ\|_2^2 + \frac{\rho}{2} \left\|B - H^k + U^k\right\|_2^2 \right) \tag{5}$$

$$H^{k+1} := \arg\min_H \left(\lambda \|H\|_1 + \frac{\rho}{2} \left\|B^{k+1} - H + U^k\right\|_2^2 \right) \tag{6}$$

$$U^{k+1} := U^k + B^{k+1} - H^{k+1}. \tag{7}$$

Separating the minimization over B and H into two steps is precisely what allows for decomposition when f and g are separable. The update of B and H is performed in an alternating fashion, which accounts for the term alternating direction, making it possible to exploit the individual structure of the f and g so that B-minimization and H-minimization may be computed using different techniques in an efficient and parallel manner. In this work the *ridge regression* and the *soft thresholding* techniques will be used to solve the B and H minimizations steps, respectively.

The same procedure ought to be adapted to perform the ADMM formulation for the other LV structures with some caution on the particular details of each model. To develop distributed algorithms for the proposed structures, the problem should be formulated as a *consensus* or *sharing* optimization [8].

III. Practical Implementation and Results

A. Data Description and Experimental Setup

The ADMM algorithm is applied to the proposed LV variants in order to predict values of solar power for horizons up to six-steps-ahead. Two datasets encompassing time-series of solar power from 44 microgeneration units located in a city in Portugal are considered: (a) 15-min and (b) hourly averaged resolution. In order to form stationary time series such that the forecasting methods can be used, a normalization of the solar power is obtained by using a clear sky model [9] which gives an estimate of the solar power in clear sky conditions at any given time. The data cover the period from February 1, 2011 to March 6, 2013. The first 12 months are used to fit the models through Nelder Mead method [10] and the remaining 13 months are employed to evaluate the performance of the models. Apart from the static approach, a sliding window of 12 months period is also considered. Also, with the intention of exclude the nighttime hours, data corresponding to a solar zenith angle larger than to 90° are removed. For all models, the lags 1, 2 and 24/96 (diurnal component corresponding to 1-hour and 15-min resolution, respectively) are considered.

The evaluation of the LV structures performance is accessed using the *root mean squared error* (RMSE) calculated for each $t+h$, $h = 1, \ldots, 6$, lead-time with the following expressions:

$$RMSE_{t+h} = \sqrt{\frac{1}{k}\sum_{i=1}^{k}\left(\widehat{Y}_{t+h|t} - Y_{t+h}\right)^2}, \qquad (8)$$

where $\widehat{Y}_{t+h|t}$ represents the forecast made at time instant t and Y_{t+h} is the observed normalized solar power value. The RMSE is normalized with the solar peak power and calculated separately for each model using the full dataset of errors. The performance of the LV models is compared computing the improvement over the AR model in terms of RMSE. In the implementation of the distributed algorithm, a consensus problem approach is considered.

B. Forecasting Results and Discussion

The improvement of the LV structures over the AR model fitted with the LS method, in terms of RMSE, for each lead time is plotted in Fig. 1 and 2 using 1-hour and 15-min resolution, respectively. Clearly all the proposed structures outperform the AR model, becoming apparent the benefit from considering spatial-temporal models. For all lead times, the cLV and ooLV structures stand out as the ones with better performance, except for lead time 4 using 15-min resolution dataset, in which although ooLV remains on the top, the sLV and rLV slightly outperform cLV. For each dataset resolution, the models have a similar improvement behaviour over the time horizon (decreasing using 1-hour resolution and increasing using 15-min resolution) and the highest improvement is achieved on the first lead time ($\approx 7.3\%$) using 1-hour resolution and in the fourth lead time ($\approx 7.9\%$) using 15-min resolution.

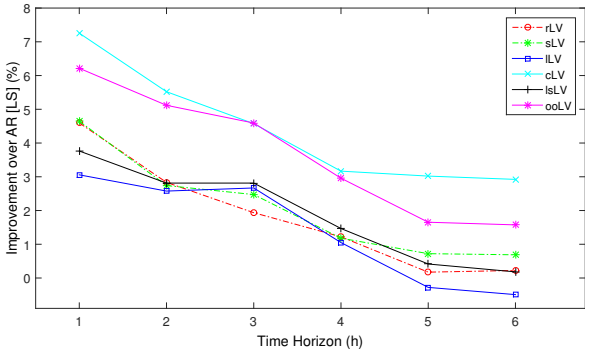


Fig. 1: RMSE improvement over of the LV structures using a static approach over AR model (fitted with LS model) for 1-hour resolution dataset

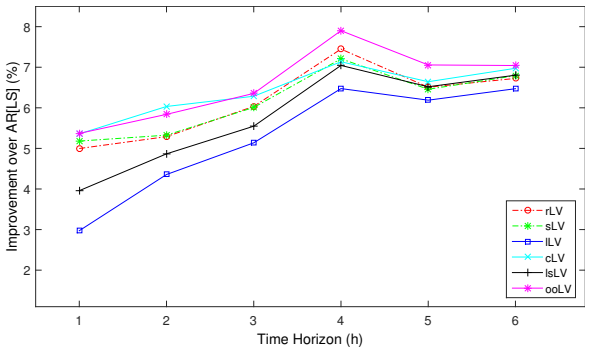


Fig. 2: RMSE improvement over of the LV structures using a static approach over AR model (fitted with LS model) for 15-min resolution dataset

The Fig. 3 and 4 show the improvement of the LV structures applying a sliding-window (SW) approach (the window slides by day) over the AR fitted using Recursive Weighted Least Squares model (RWLS) with forgetting factor, hereafter referred as AR[RWLS], for 1-hour and 15-min resolution, respectively. It is noteworthy that all the models present a higher improvement for all lead times and that the improvement behaviour maintains the same, although more smoothed, relatively to the static approach. For 1-hour resolution dataset, unlike the static approach, it seems to be more difficult to detect which structures exhibit a better performance. However, one can elect the cLV and the lsLV (attaining improvements of 10.6% and 11%, respectively, for the first lead time) as the structures with better performance given their more stable behavior over all the lead times. For 15-min resolution dataset, although there are not huge differences between the models, clearly the lLV and lsLV structures are distinguish as the ones with better performance (attaining improvements of 10% and 9.8%, respectively, for the first lead time).

Comparing the 1-hour resolution results with those obtained by the VAR model fitted with OLS and GB (in [1]), it can be

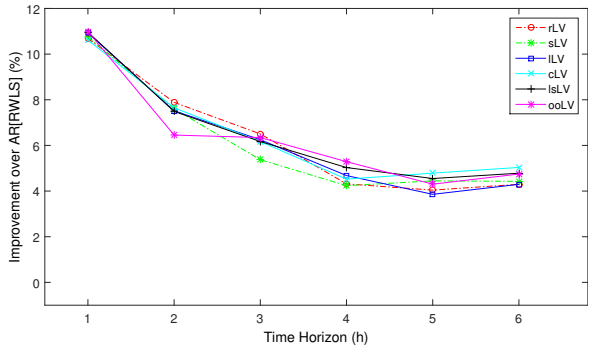


Fig. 3: RMSE improvement over of the LV structures using a SW approach over AR model (fitted with RWLS model) for 1-hour resolution dataset

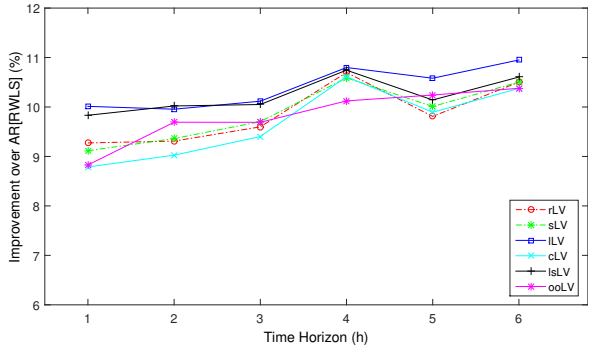


Fig. 4: RMSE improvement over of the LV structures using a SW approach over AR model (fitted with RWLS model) for 15-min resolution dataset

observed that the improvement of the cLV and lsLV structures is only surpassed by the VAR[OLS] in the lead times 2 and 3, and by the VAR[GB] in lead times 5 and 6, although in the latter case the difference is not significant.

In order to make possible the quantification of the improvement (in terms of errors magnitude), the Tables II and III show, for both static and sliding-window approaches, the global RMSE scores for lead-time $t+1$ using the six LASSO-VAR structures and also the AR model (static approach) or the AR[RWLS] model (sliding-window approach), corresponding to 1-hour and 15min resolution datasets, respectively.

Table II: Average RMSE (lowest value in bold) across all sites for lead-time $t+1$ (values normalized by rated power) for static and SW approaches using 1-hour resolution dataset

	rLV	sLV	lLV	cLV	lsLV	ooLV	AR/AR[RWLS]
static	8.8087	8.803	8.93	**8.5913**	8.8688	8.6744	9.3043
SW	8.2917	8.2959	8.2654	8.3009	8.2621	**8.2571**	9.3124

Table III: Average RMSE (lowest value in bold) across all sites for lead-time $t+1$ (values normalized by rated power) for static and SW approaches using 15-min resolution dataset

	rLV	sLV	lLV	cLV	lsLV	ooLV	AR/AR[RWLS]
static	8.1438	8.1275	8.3115	8.1236	8.2246	**8.1224**	8.6202
SW	7.8867	7.8983	**7.8224**	7.9280	7.8344	7.9165	8.6875

For the purpose of analyzing the sparsity of the coefficients matrices generated by the proposed models (for static approach), the corresponding percentages of sparsity for all lead times are presented in the Tables IV and V, corresponding to 1-hour and 15-min resolution datasets, respectively. It is possible to observe that all the structures, except the lLV and ooLV, exhibit significant sparsity and it increases with the lead time. Both lLV and ooLV models penalize large blocks, which sometimes is too restrictive motivating that an entire block is included or excluded from the model. By the way of example, the sparsity patterns for the first lead time of the cLV using 1-hour resolution dataset and of the ooLV using 15-min resolution dataset (the models with better performance in each case) are represented in Fig. 5 and 6. The darker shade represents coefficients that are larger in magnitude. As expected, the first lag diagonal coefficient's (the site own past observation) are the most relevant ones in both cases. Also, for the 15-min resolution dataset, the ooLV reveals that in the second lag only the site's own past observations are relevant input variables to perform the forecast.

Table IV: Sparsity (in percentage) of the coefficients matrices generated for the different LASSO-VAR structures using a static approach and 1-hour resolution dataset

Lead Times / **LV Struc.**	**1**	**2**	**3**	**4**	**5**	**6**
rLV	69.4	81.6	81.6	87.7	88.9	89.9
sLV	67.9	82.2	87.4	88.5	88.6	89.6
lLV	0	0	0	0	0	0
cLV	53.0	65.0	66.3	75.5	76.7	77.8
lsLV	34.0	41.9	43.5	40.4	40.4	42.2
ooLV	0	33.3	0	33.3	33.3	32.6

Table V: Sparsity (in percentage) of the coefficients matrices generated for the different LASSO-VAR structures using a static approach and 15-min resolution dataset

Lead Times / **LV Struc.**	**1**	**2**	**3**	**4**	**5**	**6**
rLV	74.6	72.2	73.3	78.4	73.6	63.24
sLV	76.2	71.9	72.4	77.5	70.04	61.3
lLV	0	0	0	0	0	0
cLV	54.2	54.9	57.2	55.3	48.7	37.7
lsLV	37.7	32.8	34.3	36.8	35.2	29.4
ooLV	32.6	0	0	33.3	0	0

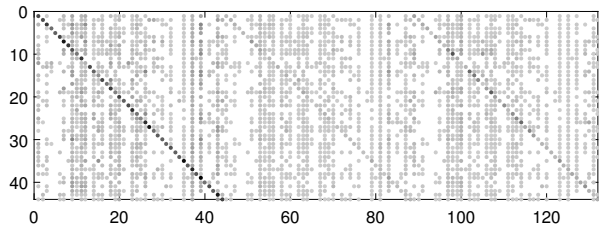


Fig. 5: Coefficients matrix (sparsity structure) of the cLV structure for first lead-time using a static approach and 1-hour resolution dataset

In a sliding-window approach context, it is possible to analyze the dynamic evolution of the coefficients matrix over the time. Considering the cLV structure for 1-hour resolution and ooLV structure for 15-min resolution, the corresponding Fig. 7 and 8 illustrate, for the first lead time, the coefficients

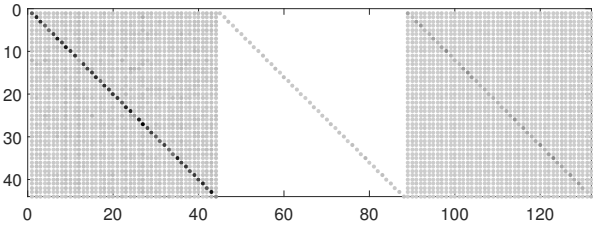


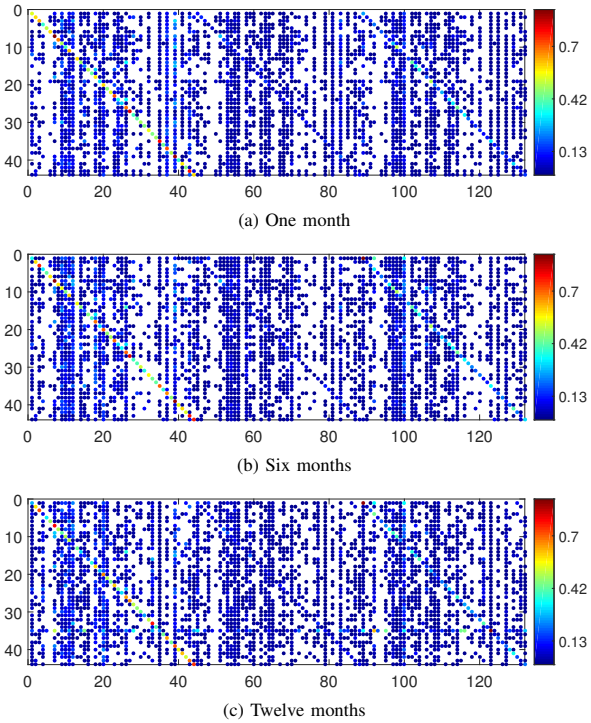
Fig. 6: Coefficients matrix (sparsity structure) of the ooLV structure for first lead-time using a static approach and 15-min resolution dataset

matrix (colored with an absolute magnitude degree) obtained for the windows corresponding to a sliding of one, six and twelve months.

For both datasets, and both structures considered for each one of them, it is notable that the first lag diagonal is the most relevant one, followed by lag 24 diagonal, since the coefficient's magnitude present by them is significantly higher than the one present by second lag diagonal and most of the off-diagonal entries. For 15-min dataset, the ooLV structure, which concerns the possibility that the prediction variable is more influenced by their own past observations than by past observations of other variables, highlights the great importance of the predictors corresponding to the first lag in the prediction with a huge magnitude difference relatively to the remaining coefficients, that should also be taken into account but with a smaller contribution. Moreover, from the fact that sometimes only the diagonal of the second lag is considered one can infer that the second lag is not so relevant in this case.

The sparsity evolution of the coefficient matrix obtained along the sliding window period for 1-hour (cLV structure) and 15-min resolution datasets (ooLV structure) is depicted in Fig. 9 and 10, respectively. Using 1-hour resolution dataset, one can observe that the coefficient matrix sparsity present a relatively stable behaviour (between 52% and 54%) in the first four sliding months, becoming uneven and bumpy in the remaining sliding months highlighting a very sharp slope for a sliding of 6/7 months (decreasing) and 11/12 months (increasing), corresponding to a training window of August 2012 - August 2013 and January 2012 - January 2013, respectively. The low sparsity presented when predicting the months from October 2012 to January 2013 means that, in this case, these are the months that need less information, in terms of available predictors, to perform the forecast. For 15-min resolution dataset, it is possible to observe that the coefficient matrix has a constant sparsity of 32,6% for the first seven sliding months and no sparsity in the remaining sliding months, revealing that for the first seven sliding months the second lag off-diagonal entries do not contribute to improve the forecasts. As explained before, this structure (ooLV) penalizes large blocks (lag diagonal and off-diagonal blocks) which means that the sparsity is not so variable as in other structures penalizing smaller blocks or individual entries.

Finally, in order to evaluate the computational performance of the ADMM algorithm, the running times and number of iterations of both nondistributed and distributed versions using

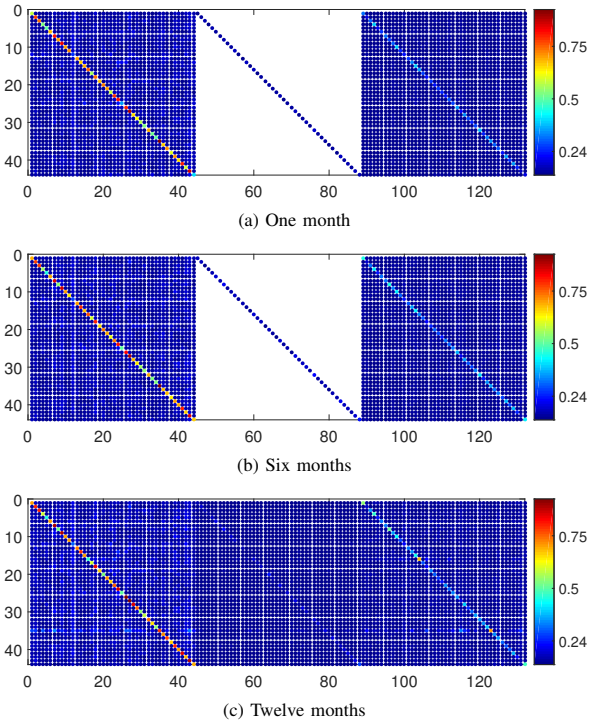


(a) One month

(b) Six months

(c) Twelve months

Fig. 7: Coefficients matrices of the cLV structure for first lead-time using SW approach and 1-hour resolution dataset: (a) sliding of one month, (b) six months and (c) twelve months



(a) One month

(b) Six months

(c) Twelve months

Fig. 8: Coefficients matrices of the cLV structure for first lead-time using SW approach and 15-min resolution dataset: (a) sliding of one month, (b) six months and (c) twelve months

1-hour resolution dataset are depicted in Table VI. It can be observed that the standard ADMM takes only few seconds to run and that the use of distributed ADMM results in a decrease

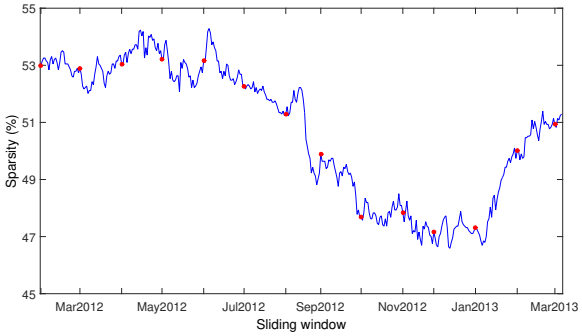


Fig. 9: Sparsity evolution of the coefficients matrix of the cLV structure for first lead-time over the SW using 1-hour resolution dataset (red markers correspond to first day of each month)

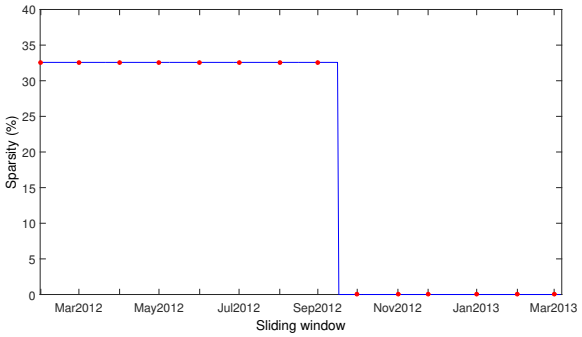


Fig. 10: Sparsity evolution of the coefficients matrix of the ooLV structure for first lead-time over the SW using 15-min resolution dataset (red markers correspond to first day of each month)

of the running time to all structures, except for the rLV, that can be justified by the increased time that this structure spend out of the cycle, which reflects the impact of the the increasing number of calculations for being applied to predict each location separately.

Table VI: Total running times (in sec) and number of iterations (between parenthesis) for 6 lead-times of the LASSO-VAR structures using distributed (data divided by a i7 8-cores processor) and non-distributed ADMM

LV Structures	Standard ADMM	Distributed ADMM
rLV	2.34 (6389)	9.46 (3215)
sLV	1.50 (746)	0.37 (147)
lLV	1.44 (665)	0.47 (140)
cLV	3.45 (409)	1.81 (158)
lsLV	1.01 (288)	0.42 (164)
ooLV	1.03 (227)	0.45 (97)

IV. Conclusions

Inspired by previous studies, this work presents a spatial-temporal solar power forecasting methodology combining the VAR and LASSO frameworks that can accommodate a wide range of potential dynamic sparse structures. A set of sparsity-promoting LASSO-VAR structures are explored and fitted with ADMM in order to capture the dynamics of the underlying system and provide a scalable solution.

The models are implemented using both static and sliding-window approaches and its results are analyzed on a realistic case study with 44 microgeneration units geographically dispersed in a city in Portugal. The structures showed a significant improvement, up to 11% using a sliding-window approach, over AR model, and also most of them outperformed the VAR[OLS] and VAR[GB] for almost all lead-times. These results confirm the benefits of using spatial-temporal data modeling short-term generation of PV farms. This short-term forecasting system performs well in terms of computational cost since it takes only few seconds to run. Additionally, in general, the computational time is greatly reduced when a distributed version of ADMM is used, revealing that one can really profit using the ADMM to deal with high-dimensional data.

The ADMM algorithm applied to LASSO-VAR models has proven to be an adequate choice to estimate the regression parameters. In order to take advantage of the spatial information to improve forecasts, more complex and dynamic sparse structures should be engineered and combined in such a artfully way that it will be possible to walk towards a more accurate solution providing relevant insights about the underlying relationships between inputs and outputs. Furthermore, the extension of the statistical model to a probabilistic forecast framework and the implementation of reliable alternatives to ADMM are promising to be considered for further work.

Acknowledgment

This work was made in the framework of the SusCity project (contract no. “MITP-TB/CS/0026/2013”) financed by national funds through *Fundação para a Ciência e a Tecnologia (FCT)*, Portugal.

References

[1] R. Bessa, A. Trindade, and V. Miranda, “Spatial-temporal solar power forecasting for smart grids,” *IEEE Transactions on Industrial Informatics*, vol. 11, no. 1, pp. 232–241, February 2015.

[2] M. He, V. Vittal, and J. Zhang, “A sparsified vector autoregressive model for short-term wind farm power forecasting,” in *Proceedings of the 2015 IEEE Power & Energy Society General Meeting*, Denver, CO, USA, July 2015.

[3] J. Dowell and P. Pinson, “Very-short-term probabilistic wind power forecasts by sparse vector autoregression,” *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 763–770, March 2016.

[4] R. A. Davis, P. Zang, and T. Zheng, “Sparse vector autoregressive modelling,” 2012, arXiv:1207.0520.

[5] W. B. Nicholson, D. S. Matteson, and J. Bien, “Structured regularization for large vector autoregression,” Cornell University, Tech. Rep., September 2014.

[6] L. Cavalcante, R. J. Bessa, M. Reis, and J. Browell, “Lasso vector autoregression structures for very short-term wind power forecasting,” *Wind Energy*, 2016, in press. [Online]. Available: http://dx.doi.org/10.1002/we.2029

[7] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 58, no. 1, pp. 267–288, 1996.

[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[9] P. Bacher, H. Madsen, and H. A. Nielsen, “Online short-term solar power forecasting,” *Solar Energy*, vol. 83, no. 10, pp. 1772–1783, 2009.

[10] J. A. Nelder and R. Mead, “A simplex method for function minimization,” *Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.