

SVD-based Visualisation and Approximation for Time Series Data in Smart Energy Systems

Abdolrahman Khoshrou
Centrum Wiskunde & Informatica
Science Park 123, 1098 XG
Amsterdam, The Netherlands
Email: a.khoshrou@cwi.nl

André B. Dorsman
Vrije Universiteit Amsterdam,
De Boelelaan 1105, 1081 HV
Amsterdam, The Netherlands
Email: a.b.dorsman@vu.nl

Eric J. Pauwels
Centrum Wiskunde & Informatica
Science Park 123, 1098 XG
Amsterdam, The Netherlands
Email: eric.pauwels@cwi.nl

Abstract—Many time series in smart energy systems exhibit two different timescales. On the one hand there are patterns linked to daily human activities. On the other hand, there are relatively slow trends linked to seasonal variations. In this paper we interpret these time series as matrices, to be visualized as images. This approach has two advantages: First of all, interpreting such time series as images enables one to visually integrate across the image and makes it therefore easier to spot subtle or faint features. Second, the matrix interpretation also grants elucidation of the underlying structure using well-established matrix decomposition methods. We will illustrate both these aspects for data obtained from the German day-ahead market.

Index Terms—Data analysis, Data preprocessing, Renewable energy sources, Smart grids, Time series analysis.

I. INTRODUCTION

In smart energy systems, time series often show two distinct time scales. On the one hand, the data exhibit strong diurnal patterns reflecting the daily (or weekly) rhythms of human activity. On the other hand, these relatively fast diurnal patterns are superimposed on slower seasonal variations that have a significant impact on the overall evolution of the data. To improve the visualization and make it easier to spot correlations between variables, we propose to analyze these time series as matrices (to be visualized as images) where rows represent hour slots, whereas columns correspond to days. This approach has two advantages. First, visualizing such time series as images allows one to visually integrate across the image and makes it therefore easier to spot subtle or faint features. Second, one can draw on well-established matrix decomposition methods to elucidate underlying structure. In this paper we will discuss both these aspects in the context of data from the day-ahead market.

II. DATA

The day-ahead market is an exchange for short-term electricity contracts where the tradings are driven by its participants [3]. Fig. 1 illustrates various collected sets of data for the German day-ahead market in 2016. Day-ahead price and the traded quantity data were collected from [4]. We obtained the day-ahead solar and wind feed-in energy data from [13].

This work was supported in part by the Dutch STW under project grant *Smart Energy Management and Services in Buildings and Grids (SES-BE)*.

ENTSO-E, the European Network of Transmission System Operators [2], was the platform for downloading the day-ahead load forecast.

Using the above-mentioned data, we will explore the influence of the daily fluctuations of the predicted supply of renewable energy sources (RES), viz. wind and solar feed-in, on the realized electricity price dynamics.

III. BACKGROUND AND LITERATURE REVIEW

Recently, the impact of variable generation on the electricity market has attracted a lot of attention. Denny et al. [1] explore how increased interconnection between Great Britain and Ireland would facilitate the integration of the wind farms into the power system. Simulation results in this work imply that large increases in the interconnection capacities bring about a decrease in average price and its volatility in Ireland. Furthermore, the growing contribution of intermittent energy sources enforces the transmission grid extensions and expanding the cross-border interconnections capacities to ensure the grid stability. K. Schaber, F. Steinke, and T. Hamacher in [9] examine the viability of this approach and its effects, based on the projected wind and solar data until 2020. They conclude that grid expansion is, indeed, helpful in coping with externalities which come with the deployment of renewable energies. The positive outcomes of the substantial deployment of photovoltaic (PV) installations in Germany and Italy, and in particular, their role in daytime peak price drop have been discussed by K. Barnham, K. Knorr, and M. Mazzer in [8]. This work also reports the benefits of the complementary nature of wind and PV resources in the UK. Continuing further with studying the influence of renewable energy sources (RES) in Germany, a preliminary study on the German day-ahead market has been carried out in [11]. In the reported work, N. Adaduldah, A. Dorsman, GJ. Franx, and P. Pottuijt have taken into account the priority that the German policies assigns to renewables over fossil fuels in case of adequate supply. The authors reported the existence of convincing evidences for the impact of RES on the recent emerging of negative prices on the German day-ahead market.

Inspired by the work in [10], the goal of the present work is to determine the influence of the variability of the wind and solar feed-in on the price variability in the day-ahead market. To this end we focus on the intra-day dynamics of price as

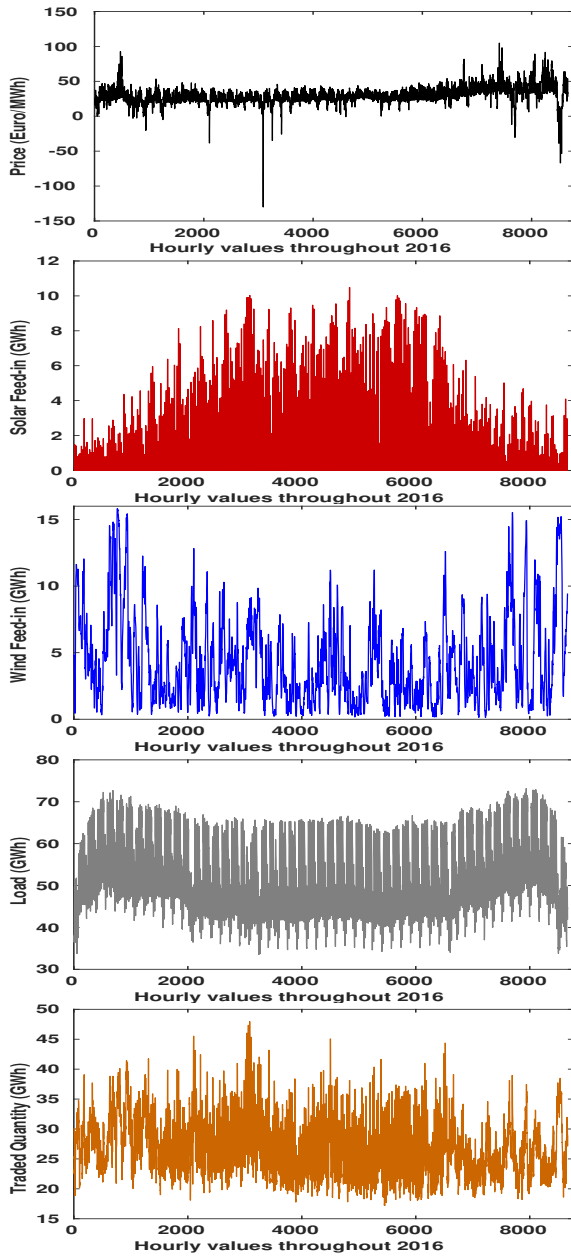


Fig. 1. An overview of the German day-ahead market in 2016; each data point represents one hour slot. From top to bottom, the price, solar, wind, load, and the traded quantity.

characterized by the second derivative, as this peaks for sharp trend reversals.

IV. METHODS: ANALYSING TIME SERIES AS IMAGES

A. Motivation

An alternative way of visualizing time series with diurnal patterns is as matrices (images). Fig. 2 shows one year's worth of data represented as a 24×366 matrix/image (recall that 2016 is a leap year). An important distinction of this new way of representation is that it allows us to visually integrate patterns across longer time spans, which results in higher discriminatory power. For instance, even a cursory glance at the data image for the traded quantity (Fig. 2, bottom)

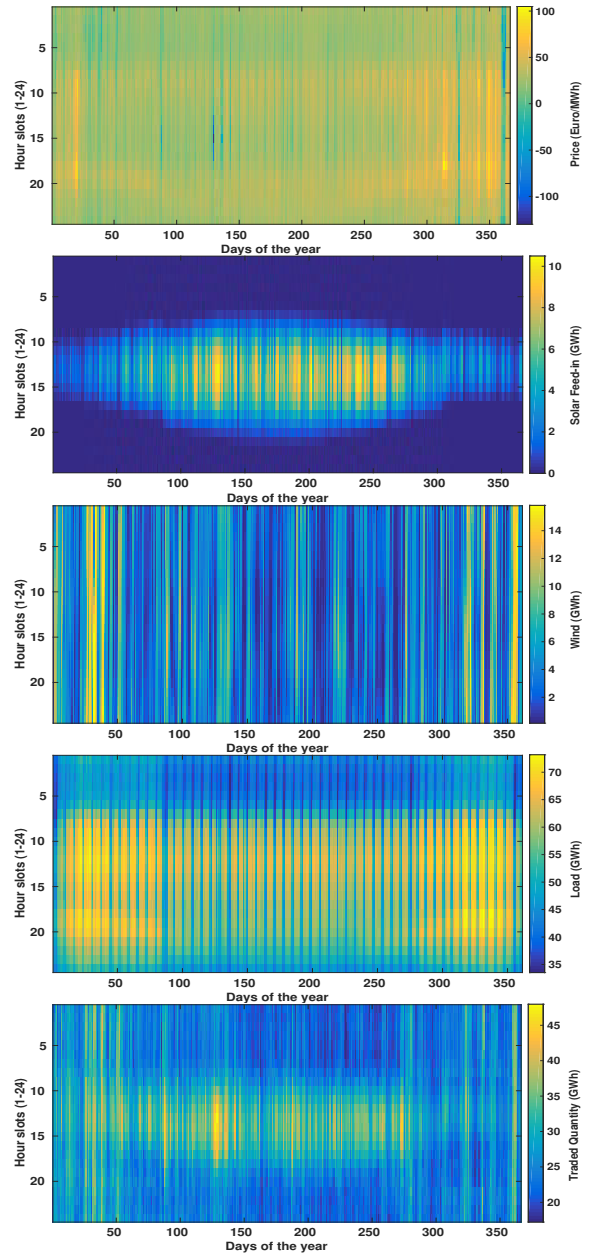


Fig. 2. A more informative representation of the German day-ahead market by reformatting the timeseries in Fig. 1 into matrices of size 24×366 . Each image (or matrix) column represents a single day of 24 hour-values. From top to bottom, the price, solar, wind, load, and the traded quantity.

highlights the fact that there is a significant correlation with solar (the eye-like horizontal shape seen in the 2nd panel from the top) as well as wind (the vertical stripes in the 3rd panel from the top).

B. Finding peaks and valleys

As mentioned before, the main scope of this work is to explore how the inherent variability of the supply by RES can affect the intra-day variability of the market. To address this problem, we have applied an additional transformation on the daily profiles of the quantities of interest (viz. price, load, traded quantity, solar and wind feed-in). More precisely,

any of the above quantities (generically denoted by f) can be considered as a function of two variables:

- time of day, hour slots $1 \leq h \leq 24$
- day of year, $1 \leq d \leq 366$ (2016 is a leap year!)

Hence for such a function $f(h, d)$, we can investigate the corresponding 2nd derivative with respect to the hour (intra-day);

$$f_{hh} \equiv \frac{\partial^2 f}{\partial h^2} \approx \frac{f(h+1) - 2f(h) + f(h-1)}{h^2} \quad (1)$$

Extreme values of this 2nd derivative capture peaks (i.e. local maxima for which $f_{hh} < 0$ and extreme) or valleys (i.e. local minima for which $f_{hh} > 0$ and extreme). An example is illustrated in Fig. 3 which shows the wind feed-in profile (top) and its corresponding intra-day 2nd derivative profile (bottom) on Nov. 18, 2016. We contend that comparing the evolution of the 2nd derivative profiles (intra-day wise) can highlight the impact of the wind and solar energy feed-in on the price and also the traded quantity in 2016.

Applying this intra-day 2nd derivative operator to all five quantities of interest and re-visualising the resulting time series as images highlights some interesting features of the data. As an illustration, Fig. 4 shows the 2nd derivative for both solar (top) and wind feed-in (bottom). In the top panel the gradual shift of sunrise and sunset over the seasons is clearly visible. Moreover, close inspection of this figure also reveals the dates of the switch to daylight saving summer time (days 87 and 304). The image for the wind feed-in is also interesting. As expected, wind values are more erratic and less seasonally determined. That being said, there is an undeniable “eye-like” shape that faintly mirrors the intra-day wind activities (see, e.g., [14]).

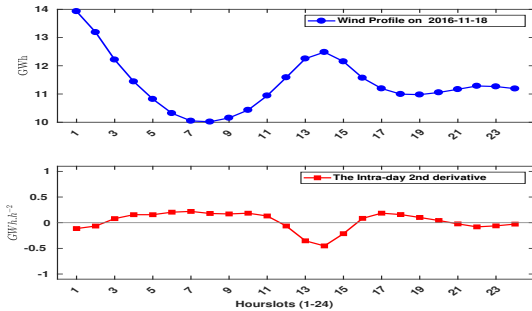


Fig. 3. Wind profile and its corresponding intra-day 2nd derivative on Nov. 18, 2016. Every sag in the lower profile corresponds to a swell in upper and vice versa.

C. Using SVD to highlight structure

Another advantage of representing the time series as matrices is the possibility of using the matrix decomposition techniques to analyze the structure of the data. In this paper we will focus on the well-known singular value decomposition (SVD) method which states that an arbitrary $h \times d$ matrix A of rank $r \leq \min(h, d)$ can be factored as:

$$A = USV^T = \sum_{k=1}^r \sigma_k U_k V_k^T \quad (2)$$

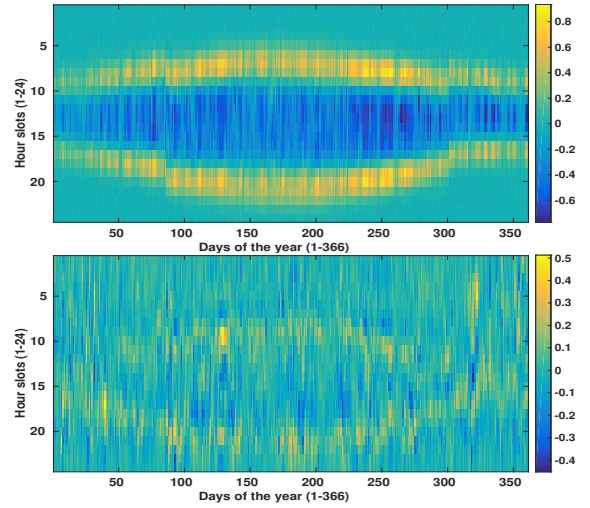


Fig. 4. The evolution of appropriately transformed hourly values (2nd derivatives, see text for more details) for solar (top) and wind (bottom) feed-in. Top: Solar feed-in (intra-day 2nd derivative). The colour coding (in yellow) clearly highlights the change in sunrise and sunset times, creating an overall “eye-like” structure. Bottom: Wind feed-in (intra-day 2nd derivative). As expected, wind feed-in is much more erratic. Interestingly enough, this figure shows a vague but undeniable outline of an eye-like contour which mirrors the intra-day wind activities.

where $U \in \mathcal{O}(h)$ and $V \in \mathcal{O}(d)$ are orthonormal matrices, (with U_k and V_k denoting the k^{th} column of U and V , respectively) and S is an $h \times d$ matrix for which the only strictly positive elements σ_k (so-called singular values) are placed on the main diagonal (see [7]). The importance of this decomposition lies in the fact that truncating the expansion in the right hand side of (2) after the p^{th} term yields the best approximation of the original matrix A by a matrix of (lower) rank p (see [6]).

$$A_p = \arg \min_{\text{rank}(R)=p} \|A - R\| \quad (3)$$

where the norm $\|\cdot\|$ can be either the Frobenius or spectral L_2 norm. Translating these results back to the original time series, we see that the columns of $U \in \mathcal{O}(24)$ represent daily profiles, whereas the columns of $V \in \mathcal{O}(366)$ furnish the corresponding amplitudes (one for each day). Fig. 5 shows a concrete illustration for the price data: the first column U_1 is depicted in the top panel and provides an overall daily profile (obtained as a weighted average). The expected price peaks in the morning and early evening are clearly discernible. The corresponding amplitudes (one for each day) are given by the column V_1 and shown in the 3rd panel. Some days with exceptionally low or high prices are clearly visible.

A rank-1 approximation of the original time series would therefore be obtained by taking the averaged daily profile U_1 in the top panel and scaling it up or down using the 366 values in V_1 (depicted in the third panel). Notice that in this first approximation, each day has the same profile, only the amplitudes change from one day to another. The values of U_2 (depicted in panel 2) specify a first correction to U_1 , with the corresponding amplitudes for this correction specified in V_2 (displayed in panel 4). This correction means that for any day for which the corresponding V_2 coefficient is positive (almost every summer day) will have a lower price value between 11h and 18h than would have expected based on the

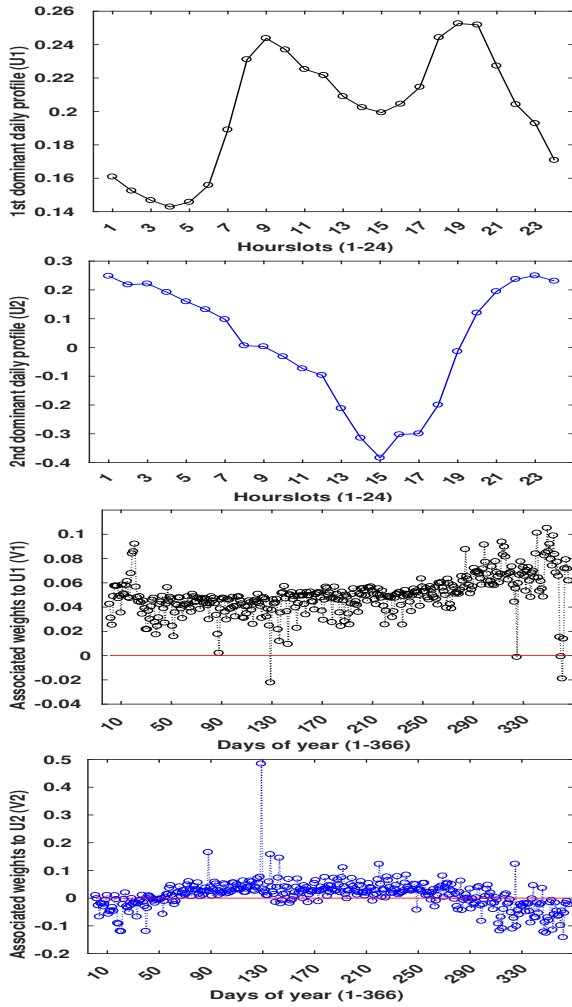


Fig. 5. From top to bottom: The first two panels are the first (U_1) and second (U_2) dominant profiles of the price profiles in 2016; the third and the last one are their corresponding amplitudes (V_1 and V_2) throughout the year.

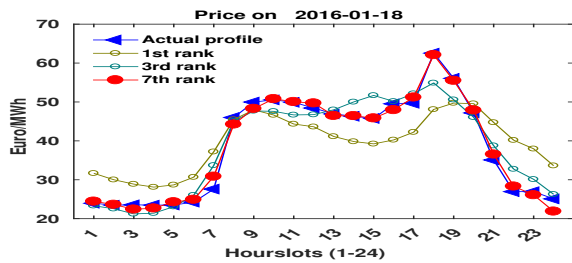


Fig. 6. Low rank approximation of actual data (one particular day, Jan. 18, blue). Including up to 7 SVD components yields the rank-7 approximation (bold red). Lower rank approximations are also shown.

(weighted) annual average U_1 . Adding additional terms in the SVD expansion improves the approximation. This is illustrated in Fig. 6 where the actual data for one typical day (18 Jan. 2016) are shown in conjunction with low-rank approximation up to rank 7.

D. Structure-preserving smoothing

As mentioned earlier, recasting the time series as images allows us to visually integrate subtle patterns in the data.

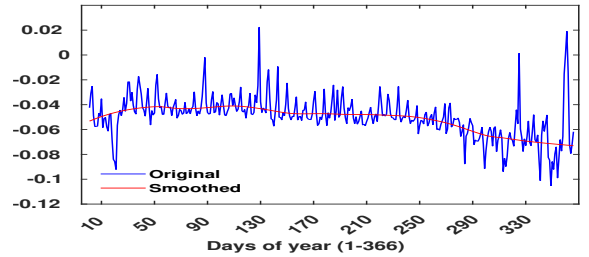


Fig. 7. An example of a smoothed V_1 (corresponding to the most dominant singular value) used in reconstruction of the 2nd derivative of the price data.

Now, we indicate how the SVD factorization suggests a straightforward method to smooth the time-series in such a way that the overall structure is preserved. Recall that the V columns determine the amplitudes of every daily-profile. By smoothing these profiles we eliminate most of the inter-day variation without affecting the overall structure. For the data at hand, the smoothing was based on robust local regression (RLOESS) but alternative approaches would be equally valid. The local regression smoothing method was used to alleviate the effect of outliers, while preserving the general trend in data. Fig. 7 shows an example of the smoothed V_1 applied in the calculation of A_p . Using the smoothed version of the V vectors in the low rank reconstruction of data, yields the images depicted in Fig. 8. The panels in the left hand column show the original 2nd derivative profiles for the various quantities of interest (price, solar and wind feed-in, load and traded quantity). The two adjacent columns show the same data, but this time using two different smoothed, low-rank approximations: a rough rank-1 approximation (middle) and a much more accurate rank-7 approximation, which will be used in the regression analysis (Section V). In addition to appearing visually unbiased the choice for rank-7 is also based on the structural similarity index (SSIM), typically used for measuring image quality. SSIM assesses the deviation of A_p from the original matrix A by comparing their corresponding local means, standard deviations, and cross-covariance matrices. For detailed description, readers are referred to [12]. As can be seen in Fig. 9, in terms of SSIM, the approximations keep improving up till $p = 7$ after which it levels off. As a consequence, we select rank-7 as the appropriate approximation. It is worth noting that this figure is in line with our expectation regarding volatility of the various quantities: wind results in the lowest similarity, while solar feed-in agrees best with the approximation.

V. RESULTS AND CONCLUSIONS

After the lengthy methodology section we are now in a position to state some results. In order to shed light on how the price volatility could be linked to the volatility of other data of interest, we perform two pre-processing steps (discussed in detail in the previous section):

- 1) We compute the intra-day 2nd derivative which highlights peaks and valleys (concavity/convexity as a notion of intra-day variability).
- 2) Next, we approximate these images using an SVD expansion up to rank 7.

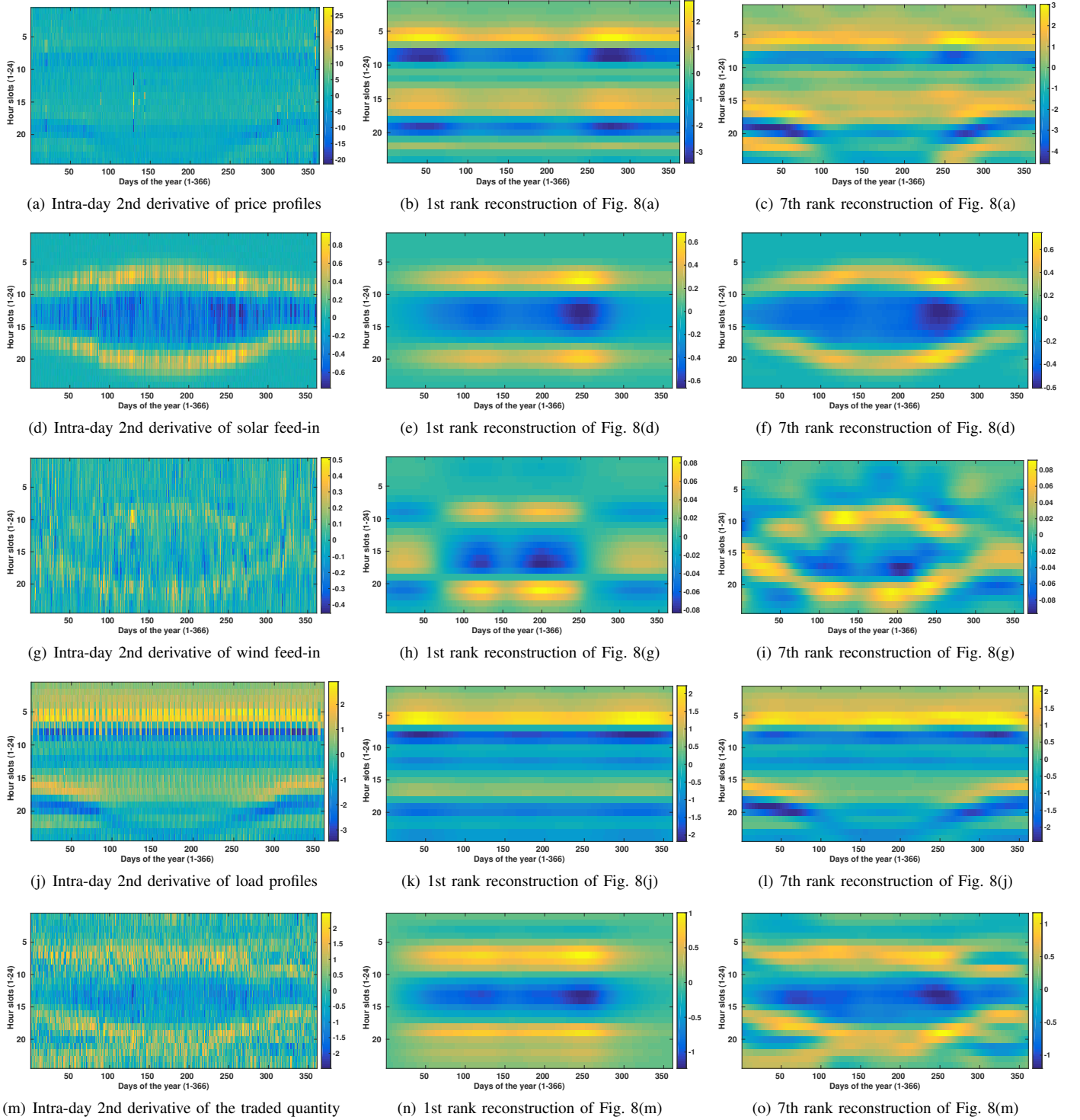


Fig. 8. Left: The intra-day 2nd derivative (concavity) of the German day-ahead market data in 2016, for (top to bottom) price, solar and wind feed-in, load and traded quantity. The underlying trends have been magnified using rank-1 (middle) and rank-7 (right) reconstruction. Clearly, the rank-7 reconstruction yields an acceptable approximation of the original images on the left.

We call the resulting data C_p, C_l, C_q, C_s and C_w where each subscript refers to the corresponding quantity. Finally, we regress the price data C_p on the other data:

$$C_p = \alpha_0 + \alpha_l C_l + \alpha_q C_q + \alpha_s C_s + \alpha_w C_w \quad (4)$$

The regression was performed for three scenarios on untreated

(original) and enhanced (smoothed) 2nd derivative data; using all data, using only day-time data, or using only night-time data. Table I contains the results for the case where the original data, A , were used. An improvement of the results was achieved by using the rank-7 reconstructed data, A_p , as is seen in Table II. Fig. 10 highlights the significance of the obtained

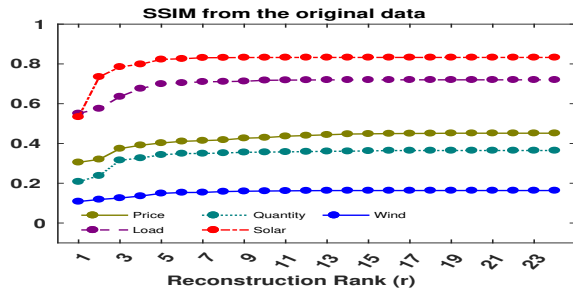


Fig. 9. The evolution of SSIM of the reconstructed matrix, A_p (for $p = 1, \dots, 24$) from the original 2nd derivative matrix A . Thus rank-7 was considered an appropriate choice in our experiments.

TABLE I. POOR RESULTS OF THE INITIAL REGRESSION MODEL ON THE INTRA-DAY 2ND DERIVATIVE DATA (BEFORE USING SVD-BASED TECHNIQUE)

| TimeSlot | R^2 | α_l | α_q | α_s | α_w |
|------------|-------|------------|------------|------------|------------|
| 24h | 47.28 | 1.26 | 0.44 | -2.98 | -1.98 |
| day time | 53.07 | 1.34 | 0.44 | -2.97 | -1.83 |
| night time | 15.60 | 0.91 | 0.24 | N/A | -2.05 |

TABLE II. REGRESSION MODEL APPLIED ON RANK-7 RECONSTRUCTION OF THE INTRA-DAY 2ND DERIVATIVE OF DATA

| TimeSlot | R^2 | α_l | α_q | α_s | α_w |
|------------|-------|------------|------------|------------|------------|
| 24h | 81.84 | 1.12 | 0.59 | -2.83 | -3.60 |
| day time | 86.27 | 1.26 | 0.29 | -2.36 | -1.27 |
| night time | 56.40 | 0.85 | 0.09 | N/A | -17.59 |

results of the later case in comparison with the randomized data (permutation test where, first, the days of the year are shuffled, then the regression models were applied). High values of R^2 for 24h and day time scenarios is an indicator of good performance of the model and emphasizes the fact that the concavity of the prices can indeed be modeled as a function of the concavity of other attributes, such as, load, traded quantity, solar and wind feed-in.

These findings lead to a number of conclusions. The intra-day dynamics (concavity) of the price is least affected by the traded quantity on the day-ahead market. Moreover, RES have higher impact on the price dynamics than the load. During the day time, solar is the dominant attribute affecting the price dynamics, almost twice as much as the wind and the load. In a similar way during night hours, wind can affect the price dynamics more than 20 times greater than the load. In latter case, however, the lower value of R^2 points out to the fact that other measures are involved in the pricing mechanism at nights. Therefore, other attributes need to be determined to model the night time price dynamics in a more satisfactory fashion.

VI. SUMMARY

In this paper we have focused on time series in which a strong diurnal pattern is superimposed on slower seasonal variations. We have shown that it makes sense to visualize such time series as images/matrices, which can be approximated using low-rank SVD-based approximations. Furthermore, this decomposition suggests a natural structure-preserving smoothing of the data. We applied this decomposition to concavity data for price, load, traded quantity, solar and wind feed-in

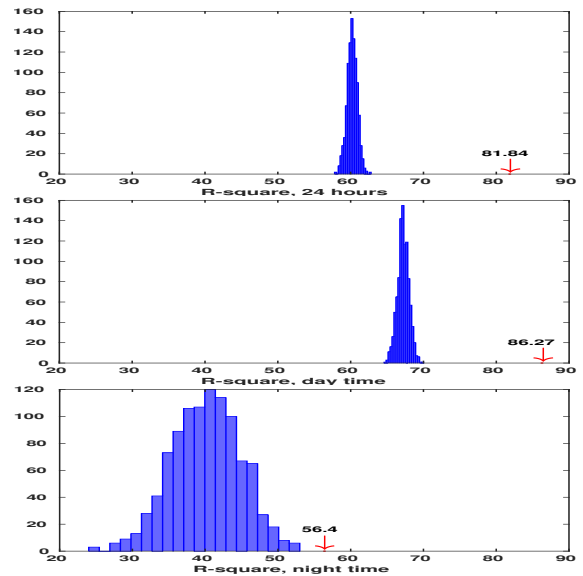


Fig. 10. Permutation test to indicate the significance of R^2 values in all three scenarios which can be found in Table II. Each histogram is the result of randomization tests, repeated 1000 times (no. of bins = 20), where the days are shuffled before applying the regression models.

on the German day-ahead market (for 2016). This process of visualization suggested a linear regression model to unveil the impact of renewables on price.

REFERENCES

- [1] E. Denny, A. Tuohy, P. Meibom, A. Keane, D. Flynn, A. Mullane, M. Omalley. *The impact of increased interconnection on electricity systems with large penetrations of wind generation: A case study of Ireland and Great Britain*, Energy Policy, 38(11):6946-54, 2010.
- [2] ENTSO-E, the European Network of Transmission System Operators, <https://www.entsoe.eu/Pages/default.aspx>
- [3] EPEXSPOT, European Power Exchange, <http://www.epexspot.com/en/market-coupling>
- [4] EPEXSPOT, Day-Ahead Auction, <https://www.epexspot.com/en/product-info/auction/germany-austria>
- [5] Filtering and Smoothing Data, <https://nl.mathworks.com/help/curvefit/smoothing-data.html>
- [6] Golub, Gene H., and Christian Reinsch, *Singular value decomposition and least squares solutions*, Numerische mathematik, 403-420, 1970.
- [7] K. Baker, *Singular value decomposition tutorial*, The Ohio State University 24, 2005.
- [8] K. Barnham, K. Knorr, M. Mazzer, *Benefits of photovoltaic power in supplying national electricity demand*, Energy policy, 54:385-90, 2013.
- [9] K. Schaber, F. Steinke, T. Hamacher. *Transmission grid extensions for the integration of variable renewable energies in Europe: Who benefits where?*, Energy Policy, 43:123-35, 2012.
- [10] L. Hirth, *The market value of variable renewables: The effect of solar wind power variability on their relative price*, Energy economics 38, 218-236, 2013.
- [11] N. Adadulrah, A. Dorsman, GJ. Franx, P. Pottuijt, *The Influence of Renewables on the German Day Ahead Electricity Prices*. Springer Berlin Heidelberg In Perspectives on Energy Risk, 165-182, 2014.
- [12] Structural Similarity Index, <http://nl.mathworks.com/help/images/ref/ssim.html>
- [13] Taking Power Further, <https://www.tennetso.de/site/en/Transparency/publications/overview>
- [14] The Royal Netherlands Meteorological Institute <http://www.knmi.nl/nederland-nu/weer/verwachtingen>