Abstract—The paper considers the problem of estimating unmonitored PV generation and load profiles in a distribution network starting from measurements of the aggregated power flow at the point of common coupling (PCC) and global horizontal irradiance (GHI). The estimation principle that we exploit relies on modeling the PV generation as a function of the GHI, enabling to identify PV production patterns in the aggregated power flow measurements. Four estimation algorithms are proposed: the first assumes that variability in the aggregated PV generation are mostly given by variations of PV generation, two use a model of the demand to improve estimation performance, and the fourth assumes that in a certain frequency range, the aggregated power flow is dominated by PV generation dynamics. Algorithms take advantage of a transposition model to explore several azimuth/tilt configurations and explain PV generation patterns from the aggregated power flow profile accounting for non-uniform installation configuration of the plants. The estimation performance of the algorithms is validated and compared with experimental measurements from a real-life setup with four households equipped with rooftop PV installations, conventional demand and local storage implementing self-consumption.

Index Terms—PV generation, Demand, Disaggregation.

I. INTRODUCTION

The uncontrolled growth of PV installations is a cause of concerns for operators because determines increased reserve requirements at the system level, whereas locally it induces violations of voltage and line ampacity constraints during peak production periods [1], [2]. Technical solutions envisaged to tackle these issues are curtailment strategies, control of power converters active/reactive power, self-consumption schemes and dispatch of local power flows according to network-safe power consumption trajectories, see e.g. [3]–[10]. A requirement for the implementation of local control strategies for PV is the availability of real-time measurements from PV plants. Besides, such an information is normally required to train data-driven forecasting, see e.g. [11], [12]. However, such a precondition might not be met in real-life conditions because installations are not always monitored, and, even when they are, factors such as, i), privacy concerns, ii), conflicts due to the different ownerships of the metering infrastructures, and iii), lack of standards for monitoring and aggregation of measurements and communication are a barrier against the possibility of harvesting real-time measurements of PV generation. As an alternative to direct monitoring of PV systems, in this paper we consider the problem of disaggregating PV generation from the aggregated active power measurements of a group of residential prosumers. We estimate the amount of installed PV capacity by modelling the PV generation as a function of the global horizontal irradiance (GHI), which is supposed to be known from local measurements. Four estimation algorithms are proposed and compared in the paper: the first assumes that variability in the aggregated power flow measurements are mostly given by variations of PV generation, the second and third take advantage of a model of the demand to improve estimation performance, and the fourth assumes that there is a certain frequency range in which the aggregated power flow measurement is dominated by the component due to the PV generation.

The proposed algorithms take advantage of a transposition model to transpose GHI values into a number of pre-defined differently oriented tilted planes. The algorithms are designed to be unsupervised, i.e. they do not require measurements of the PV power profiles to be trained. The proposed results show that the choice of the parameters can be performed in an unsupervised way while still achieving good estimation performance. The algorithms results are tested and compared by using experimental measurements from a real-life setup in the region of Basel, Switzerland. It consist of four private households equiped with monitored rooftop PV installations. The experimental setup also includes grid-connected batteries to implement PV self-consumption, thus allowing to test the performance even when the demand is correlated with PV generation.

It is worth noting that the considered problem is also investigated in [13], which proposes estimation methods based on evaluating the power factor footprint of PV converters and linear estimator of the total PV generation using the active power profile of a nearby installation. With respect to [13], we choose not to use the information on the reactive power because normally PV converters work at unitary power factor. Moreover, we adopt a transposition model to identify PV production patterns from multiple installations with different tilt/azimuth configurations, an important element in urban and suburban contexts, where distributed PV generation is mostly from rooftop PV plants, where tilt and azimuth configurations...
are normally dictated by the aspect of the roofs.

The paper is organized as follows: Section II states the problem, Section III discusses the disaggregation algorithms, Section IV describes the measurements sets used for the validation, Section V presents the algorithms performance assessment, and Section VI concludes the paper by summarizing the key results and contributions of this work.

II. PROBLEM STATEMENT AND NOTATION

A. Configuration of the system

We consider the setup sketched in Fig. 1, a generic electrical network with PV production, possibly generated from distributed installations with different tilt and azimuth, and demand. The power injections at the single buses are not measured, however the total prosumption (i.e. PV generation + demand) is known in an aggregated fashion thanks to sensing the active power flow at the PCC. Moreover, the GHI value is known from local measurements: in the current study, we consider measurements from a pyranometer, although other methods could be considered, like taking advantage of the information available from nearby monitored PV installations (see for example [14]). We do the modeling assumption that the PV installations in the considered network are subject to the same global horizontal irradiance (GHI). Given that local GHI measurements have a validity in the range of 50 m [15], we expect this assumption to be reasonable for localized PV clusters. Since we consider small portions of distribution networks, transmission losses are neglected due to short length of the cables.

![Fig. 1. A generic network topology with unmonitored demand and PV generation from multiple production sites with different azimuth/tilt configuration. The active power flow at the PCC and local GHI are known from measurements. The problem is estimating the raw PV generation.](image)

B. Notation

The active power flow measured at the PCC at the discrete time interval \( k \) is denoted as \( P_k \) (kW). We adopt the passive sign notation, i.e. positive power flows denote consumption and vice-versa. The GHI measurements from the pyranometer are denoted by \( I_{jk} \) in kW m\(^{-2}\). The quantity \( I_{jk} \), also in kW m\(^{-2}\), denotes the estimated global normal irradiance (GNI) to a certain tilted plane, where \( j = 1, \ldots, J \) denotes the plane tilt/azimuth configuration. We consider \( J = 21 \) tilted planes with tilt and azimuth configurations such that they are equally spaced on a south-facing semi-sphere (we assume to be in northern hemisphere location). GNI estimations are obtained by applying the Hay-Davies transposition model [16] which relies on decomposing GHI observation into the direct and diffuse component [17].

An example of the disaggregation process is described in the following paragraph. The quantity \( G_k \) and \( L_k \) (kW) denote respectively the estimated PV production and estimated demand, which are to be determined.

C. Problem statement

The problem is estimating the trajectories of the demand and total PV generation from measurements of the active power flow at the PCC and local GHI observations. The situation is exemplified in Fig. 2a (night time observations are removed), which shows the input, intermediate results and output results of one among the proposed algorithms. The input quantities are the aggregated power flow \( P_k \) at the PCC (top panel) and GHI (middle panel, solid fill). The profiles in the middle panel plot of Fig. 2a are the GNI trajectories \( I_{jk} \). They are with the objective of exploring the potential PV contribution from production sites with different installation criteria, as typical in urban feeders where panels are normally installed according to existing roof configurations. Finally, the lower panel plot in Fig. 2a shows the results of the disaggregation, with the estimated demand and PV generation, the latter very close to the ground truth value (from measurements, denoted by the solid fill).

![Fig. 2. Input, intermediate results and output of the proposed disaggregation algorithms (night time observations are removed).](image)
III. DISAGGREGATION ALGORITHMS

The estimated total PV generation \( \hat{G}_k \) is modelled as the sum over all the considered tilt/azimuth configurations \( j = 1, \cdots, J \) of the transposed irradiance \( I_{jk} \) times \( J \) nonnegative coefficients \( \alpha_j \in \mathbb{R}_+^J \):

$$
\hat{G}_k(\alpha) = \sum_{j=1}^{J} \alpha_j \cdot I_{jk}, \quad k = 1, \ldots, K,
$$

where \( \alpha = \{\alpha_j, j = 1, \ldots, J\} \in \mathbb{R}_+^J \) denotes the set of \( \alpha_j \). From the physical point of view, the coefficients \( \alpha_j \) are proportional to the PV generation capacity available at each specific azimuth/tilt configuration \( j \). It is to note that PV generation capacity is normally in kWp, kilowatt peak, and refers to the amount of power produced in standard test conditions (STC, kW m\(^{-2}\) GNI at 20°C). Since temperature is not considered in (1), the coefficients \( \alpha \) also include the temperature effect, which is known to influence the PV conversion efficiency, see for example [18]. In order to tackle the problem of temperature changes while still having reliable estimates of the nominal capacity, we propose to re-iterate the proposed estimation periodically. Besides, it is worth noting that while modeling the PV generation as in (1), we do the assumption that the PV plants operate in the maximum power point tracking (MPPT) mode; in case the output of a PV plant is controlled (i.e. curtailed), it is likely to be monitored and its contribution can be removed from the aggregated power flow, thus allowing to still apply disaggregation on the residual value.

The proposed disaggregation algorithms, denoted as Method A, B, C and D are discussed in details in the following four subsections. They attempt to estimate \( \hat{G}_k \) of PV generation by exploiting different modeling principles inspired by the following empirical considerations:

- variations of PV generation dominate the variations of the aggregated power flow \( P_k \). From this standpoint, Method A estimates PV generation by looking for a trajectory where variations are as close as possible to variations of the observed aggregated power flow. The drawback of this approach is that it does not consider that variations are also provoked by demand changes (e.g. load inrushes);
- in the considered prosumers scenario, the aggregated power flow consists in the contribution of demand and PV generation. Plugging in a model of the demand makes possible to estimate the PV generation without differentiating the time series, as considered in the previous point. Method B and Method C exploit this assumption applying two different models of the demand, as explained later;
- there is a certain frequency range where the dynamics of the aggregated power flow measurements are dominated by those of PV generation. In other words, we exploit the fact that demand and PV generation have different time dynamics, therefore, by filtering the observed aggregated power flow measurement it is possible to estimate the PV generation by looking for a trajectory which is similar to the filtered signal in the time domain (Method D).

The validity of these empirical considerations will be tested in the results section, where the performance of the methods will be compared and discussed.

A. Method A

The unknowns \( \alpha \) are determined by assuming that variability in the observed aggregated power flow are due to variations of the aggregated PV power. This assumption is formulated by minimizing the norm-1 of the difference between the once differentiated time series \( P_k \) and \( \hat{G}_k \) while subject to the estimated total PV production model (1):

$$
\alpha^o = \arg \min_{\alpha \in \mathbb{R}_+^J} \left\{ \sum_{k=1}^{K} \left| (P_k - P_{k-1}) + (\hat{G}_k(\alpha) - \hat{G}_{k-1}(\alpha)) \right| \right\}
$$

subject to:

$$
\hat{G}_k(\alpha) = \sum_{j=1}^{J} \alpha_j \cdot I_{jk}, \quad k = 1, \ldots, K.
$$

The advantage of Method A is being parameter-less, unless for the resolution of the input time series, an important element because it is generation (Method B attempts to determine both \( \hat{P}_k \) and \( \hat{L}_k \) by minimizing the norm-2 of the estimation error \( P_k - \hat{P}_k \). However, since this problem is under-determined (the \( K + J \) free variables are more than the number of observations \( K \)), we augment the cost function by minimizing the combination of least square and norm-1 of the one differentiated \( \hat{L}_k \) (a combined linear regression and trend filtering problem, as for example in [19] [20]):

$$
\alpha^o = \arg \min_{\alpha \in \mathbb{R}_+^J \hat{L} \in \mathbb{R}_+^K} \left\{ \sum_{k=1}^{K} \left( P_k - \hat{P}_k(\alpha, \hat{L}) \right)^2 + \lambda \sum_{k=1}^{K} \left| \hat{L}_k - \hat{L}_{k-1} \right| \right\}
$$

subject to:

$$
\hat{P}_k(\alpha, \hat{L}) = \hat{L}_k - \sum_{j=1}^{J} \alpha_j \cdot I_{jk}, \quad k = 1, \ldots, K.
$$

B. Method B

Let \( \hat{L} = [\hat{L}_1, \ldots, \hat{L}_K] \) be the the estimated demand trajectory such that the estimated active power flow at the PCC \( \hat{P}_k \) can be expressed as its sum with the estimated total PV generation \( \hat{G}_k(\alpha) \) as in (1) (the latter taken with negative sign because it is generation):

$$
\hat{P}_k(\alpha, \hat{L}) = \hat{L}_k - \hat{G}_k(\alpha), \quad k = 1, \ldots, K.
$$

Method B attempts to determine both \( \hat{L}_k \) and \( \alpha \) by minimizing the norm-2 of the estimation error \( P_k - \hat{P}_k \). However, since this problem is under-determined (the \( K + J \) free variables are more than the number of observations \( K \)), we augment the cost function by minimizing the combination of least square and norm-1 of the one differentiated \( \hat{L}_k \) (a combined linear regression and trend filtering problem, as for example in [19] [20]):

$$
\alpha^o = \arg \min_{\alpha \in \mathbb{R}_+^J \hat{L} \in \mathbb{R}_+^K} \left\{ \sum_{k=1}^{K} \left( P_k - \hat{P}_k(\alpha, \hat{L}) \right)^2 + \lambda \sum_{k=1}^{K} \left| \hat{L}_k - \hat{L}_{k-1} \right| \right\}
$$

subject to:

$$
\hat{P}_k(\alpha, \hat{L}) = \hat{L}_k - \sum_{j=1}^{J} \alpha_j \cdot I_{jk}, \quad k = 1, \ldots, K.
$$
The cost function is a composition of a vector norm-1 and a quadratic cost function, thus convex if the latter is convex. As shown in Appendix A, the convexity of the quadratic term cannot be enforced by construction because it depends on the input data, but it can be verified a-priori.

C. Method C

In Method B, the under-determination problem was tackled by seeking an estimated demand trajectory \( \hat{L} \) such that the differentiated norm-1 was as small as possible. Here, we use another approach and apply a piecewise constant model of the demand, i.e. we require the unknown sequence \( \hat{L} \) to be piecewise constant for \( c \) consecutive samples, where \( c \) is a design parameter. In the case of the first \( c \) sequence samples, this means enforcing the following \( c - 1 \) equality constraints:

\[
\hat{L}_1 = \hat{L}_2 = \cdots = \hat{L}_c.
\] (7)

Extending to the set of \( K \) measurements (\( K \) multiple of \( c \)):

\[
\hat{L}_{c(i-1)+1} = \cdots = \hat{L}_{c(i-1)+c} \quad i = 1, \ldots, K/c.
\] (8)

Assuming the demand as piecewise constant appears a reasonable assumption when the length of the constant segment does not overlap with typical intra-day dynamics of the demand, i.e. for small \( c \) values and densely sampled series. In other words, this is a reasonable modeling assumption when considering short periods of time (seconds), where persistence model of demand have normally unbeaten performance, e.g. [21], [22]. When the demand has shorter variations than \( c \) (like for example during load inrushes), the estimated demand will assume the average value of the waveform and inevitably generate an estimation error. Both the series sampling time and \( c \) are design parameters: the sensitivity of the algorithm performance with respect to their values is assessed in the next paragraph. In order to preserve the daily dynamics of the signals and avoid to train and test the algorithms on experimental data from the real-life test facility described in the next paragraph. The considered data set consists in measurements of the irradiance, aggregated power consumption and PV generation for 1 year. The first 2 are used for the training, while the last as the ground truth value to test the estimation performance of the proposed disaggregation algorithms. The data sets are experimental data from the real-life test facility described in the next paragraph. In order to preserve the daily dynamics of the signals and avoid to train and test the algorithms on different periods of the year, the dataset has been divided into day sequences and randomly shuffled. To test the algorithms accuracy with respect to the time series resolution, the original data was downsampled (by samples average). The considered data resolutions are 10, 30, 60, 120, 300, 600 and 900 seconds.
The selected resolution corresponds to the one normally implemented in existing metering systems. In particular, 900 s is the resolution of smart meters deployed in Switzerland and Italy. Here, the intent is to verify whether such a sampling time is enough for the proposed application, or if performance would benefit from more densely sampled data. Each of the 7 datasets at different resolution is further split into 3 sub-sequences to perform a three-fold cross-validation, i.e., for each resolution, the algorithms are trained on the first fold and tested on the remaining 2; the process is repeated for all the 3 folds, each time testing the algorithms on the part of the data which is not used for the training. In total, each algorithm is trained and tested 3 times for each resolution, for a total of 21 training and test runs. The dataset contains days with an uniform mix of sky conditions: partly cloudy, clear sky and overcast. Algorithms performance is tested both when there are batteries performing self-consumption and when not, thus allowing to account for the case when the demand is correlated with PV generation.

B. Experimental Setup

We consider experimental data from a real-life setup located in the region of Basel, Switzerland. It consists in four private households, each of them equipped with a rooftop PV installation with different capacity, and tilt and azimuth configuration, as summarized in Table I. PV converters operate in maximum power point tracking (MPPT) mode at unitary power factor. The households are also equipped with grid-connected battery energy storage systems (BESSs) with bidirectional power converter to implement energy self-consumption policies (actuated at 5 minutes resolution). BESS specifications are summarized in Table I. Battery injection is monitored. In the following analysis we consider two distinct cases, i.e. with and without battery action.

![Fig. 3. Method A RMSE as a function of the input time series sampling time.](image)

| Table I | PV and BESS installations in the experimental setup |
|---------|---------------------------------------------------|
| House ID | PV capacity (kWh) | Azimuth | Tilt (°) | Distance from House 1 (m) | BESS rating (kVA/kWh) |
|---------|--------------------|---------|---------|----------------|--------------------|
| 1       | 10.0               | 95      | 14      | 0              | 3/8.8              |
| 2(a)    | 7.2                | 187     | 36      | 100            | 3/4.4              |
| 2(b)    | 3.5                | 266     | 40      | 100            | –                  |
| 3       | 8.0                | 187     | 40      | 260            | 3/8.8              |
| 4       | 6.6                | 180     | 24      | 170            | 3/4.4              |

a) PV and power flow measurements: The aggregated power flow of each household (active and reactive) is known from measurements at the PCC at 10 s resolution. PV injections are monitored by sensing the AC power flow at the power converter level and are used as ground truth value to validate the estimation performance of the discussed disaggregation algorithms. The input time series we consider is the sum of the active power flows of all the 4 households. This makes possible to test the algorithms when the power flow includes generation from multiple sites.

b) Global horizontal irradiance measurements: GHI measurements are from a pyranometer installed on the roof of the household ID1. The line distance of the remaining households from the GHI observation point is shown in the second last column of Table I. All the observations are synchronized and timestamped, and logged in a time series database.

V. RESULTS AND DISCUSSION

In this section, we first assess the performance of the proposed methods individually. Later, we perform a joint performance assessment with the objective of comparing the quality of the estimations provided by the different methods and support the assertion according to which the algorithms can be considered unsupervised. For a visual exemplification of the disaggregation process, the reader is referred to Section II-C, which discussed the case of Method C.

A. Method A

Method A does not have any tuning parameter unless the sampling time of the input time series. The RMS of the estimation error as a function of the sampling time for both when self-consumption is implemented and not is shown in Fig. 6b. With no self-consumption, the RMSE stabilizes at around 2 kW for sampling times larger than 200 s, whereas, with selfconsumption, performance is poorer and reach its best at around 150 s. In both cases, worse estimation performance happens at small sampling times (i.e. when the input time series is densely sampled).

![Fig. 6b. With no self-consumption, the RMSE stabilizes at 2 kW for sampling times larger than 200 s, whereas, with selfconsumption, performance is poorer and reach its best at around 150 s. In both cases, worse estimation performance happens at small sampling times (i.e. when the input time series is densely sampled).](image)

B. Method B

Method B has a smoothing parameter $\lambda$ in (5) to weight the demand time variations $L_k - L_{k-1}$ in the cost function. Its influence on the disaggregation performance is investigated in Fig. 4, which shows the estimation error RMS as a function of the values of $\lambda$ and series sampling time. With self-consumption, best performance happens in the middle right region of the parameter space. Such a region tends to increase size in the case without self-consumption. Performance degradation patterns do not have a well identifiable trend like in the case of Method C, as discussed in the following.

C. Method C

The parameters of Method C are the resolution of the input time series and piecewise constant segment length $c$ (in number of samples). The RMSE as a function of the variations of the two parameters is shown in Fig. 5, where the y-axis is the lower cut-off period in seconds and the x-axis is the cut-off period multiplier (the upper cut-off period is defined as the lower cut-off period times the multiplier). It can be
seen that the estimation performance decreases when moving away from the area around the axis origin, denoting that using densely sampled input time series and small $c$ values (the best performance is with 20 s resolution) is convenient. As mentioned in the formulation stage, this is to expect because the choice of the two parameters affect the constant segment length of the demand piecewise constant model. In particular, this assumption appears to be reasonable when considering short periods of time, i.e. where persistence model normally have unbeaten performance. Estimation performance is worse with self-consumption (a numerical evaluation is given in the next paragraph). Contour lines in the plots of Fig. 5 denote that the performance degradation follows the same pattern, denoting that, although estimation performance is different in the two cases, the location of the optimal set of parameters fall in the same region of the parameter space.

D. Method D

Method D parameters are the bandpass filter lower and upper cut-off frequencies and a tuning constant of the bisquare loss function $\rho(.)$. The last was found not to impact substantially algorithm performance and is therefore excluded from the current analysis. The sensitivity of algorithm performance to upper and lower cut-off frequencies is shown in Fig. 6. Best performance happens in a well identifiable region in the lower left part of the parameter space, which however tend to shrink in the case with self-consumption. Performance degradation patterns are not as clear as for Method C.

E. Joint Performance Comparison

The $\min$, $\max$, $\text{mean}$, and $\text{median}$ statistics of the relative root mean square (RMSE), relative mean absolute (MAE) and relative mean (ME) estimation error of the disaggregation algorithms are compared in Table II (they are relative because divided by the total nominal power of the PV plants, 35.3 kWp). For each algorithm, the reported statistics are calculated considering the estimation errors calculated per each combination of parameters. The proposed statistic are to interpret in the following way:

- $\min$: performance to expect assuming to know a-priori the best performing set of parameters.
- $\max$: performance to expect when choosing the worst possible combination of parameters.
- $\text{mean}$: performance to expect when choosing a random combination in the parameters space.
- $\text{median}$: to evaluate performance distribution skewness.

Table II shows that all the methods perform poorer under self-consumption regimes (e.g. mean RMSE 4.8 to 5.9% and 6.2 to 8.5% for Method D and Method C, respectively), denoting that disaggregation performance is affected negatively when the demand in the aggregated power profile includes a component correlated with PV generation. In terms of $\text{mean}$ and $\text{median}$ statistics, Method D scores the best metrics, followed by C, B and A. In terms of $\min$ value, Method C outperforms the other, unless for the cases MAE and RMSE without self-consumption, where Method B is better, and ME with self-consumption, where Method D is the absolute best for all the metrics. Finally, in terms of the $\max$ statistic Method D and and C are usually the best performing, unless for the case ME without self-consumption, where Method B scores the best score.

| Statistic        | A       | B       | C       | D       |
|------------------|---------|---------|---------|---------|
| **RMSE without self-consumption** |         |         |         |         |
| $\min$           | 6.2     | 6.2     | 4.0     | 4.2     |
| $\max$           | 19.3    | 20.4    | 13.6    | 9.6     |
| $\text{mean}$    | 10.2    | 10.2    | 8.5     | 5.9     |
| $\text{median}$  | 9.1     | 9.1     | 8.5     | 5.7     |
| **MAE without self-consumption** |         |         |         |         |
| $\min$           | 4.0     | 4.0     | 2.8     | 2.8     |
| $\max$           | 13.3    | 14.2    | 9.1     | 6.5     |
| $\text{mean}$    | 6.8     | 6.8     | 5.7     | 3.7     |
| $\text{median}$  | 6.2     | 5.9     | 5.4     | 3.7     |
| **ME with self-consumption** |         |         |         |         |
| $\min$           | -13.3   | -14.2   | -9.1    | -5.9    |
| $\max$           | -3.4    | -3.4    | 4.8     | 0.9     |
| $\text{mean}$    | -6.5    | -6.5    | -4.2    | -2.6    |
| $\text{median}$  | -5.9    | -5.7    | -4.8    | -2.7    |
| **ME without self-consumption** |         |         |         |         |
| $\min$           | -13.3   | -12.5   | -6.2    | -6.8    |
| $\max$           | -2.3    | 0.7     | 3.1     | 1.1     |
| $\text{mean}$    | -5.1    | -2.8    | -3.1    | -1.4    |
| $\text{median}$  | -2.6    | -2.5    | -3.4    | -1.6    |

F. Discussion

a) On the algorithm selection: The previous results denoted that the algorithm with largest number of best scores is Method D, followed by C, B and A. If only sparsely sampled power flow observations are available (such as those from smart meters, typically at 15 minutes resolution), Method D should be selected because it keeps good performance even at low resolutions. If densely sampled observations are available, 2

2Computational times refer to a workstation equipped with an Intel Xeon processor running at 2.70 GHz with Matlab on a virtualized operating system. Method C and Method A,B and D were executed on two different machines, machine 1 and 2. The computation time of Method C was adjusted by a factor $t_2/t_1$, where $t_1$ and $t_2$ is the computation time of a reference problem executed on machine 1 and 2, respectively.
Fig. 4. Method B sensitivity analysis: RMSE as a function of the input time series sampling time and scaling parameter $\lambda$.

Fig. 5. Method C sensitivity analysis: RMSE as a function of the input time series sampling time and length $c$ of the piecewise constant segment.

Fig. 6. Method D sensitivity analysis: RMSE as a function of the lower and upper cut-off frequencies.
the algorithm choice becomes critical because performance of Method C and D are comparable. In this case, Method C has two advantages: i) parameters can be selected with an educated physical-based guess, ii) degradation patterns are the same for both with and without self-consumption. In other words, the choice of the parameters in case of Method C is more intuitive.

b) Can one consider the proposed algorithms unsupervised?: In the proposed sensitivity analysis, real-life PV observations were used as ground-truth value to assess estimation performance. However, in practical applications, it is clearly not possible to do so otherwise there would be no need for disaggregation algorithms. From the previous results, it was shown that the choice of the parameters plays a role since estimation performance varies. Nevertheless, in case of Method C (min/max RMSE in the range 4.0 ÷ 13.6%), parameters can be chosen with an educated guess in order to get closer to best performance. As far as Method D is concerned, the min/max RMSE gap is smaller (4.2 ÷ 9.6%), and estimation performance is good over a wide area of the parameter space, thus parameters can be guessed in order to aim at good performance.

VI. CONCLUSIONS

The problem of disaggregating a sequence of active power flow measurements composed of unobserved PV generation and demand into two separate estimated trajectories (one for each of them) was given. Four disaggregation algorithms were proposed. They all work by explaining similarities in the time domain between the aggregated and estimated PV generation by using different techniques, described in the paper. They take advantage of a transposition model to project GHI (supposed known thanks to local measurements) onto a number of tilted planes (equally spaced onto a semi-sphere) in order to explain production from multiple sites, an important feature considering that in urban/suburban context panels configuration is normally dictated by roof aspect constraints. The practical utility of the proposed algorithms is in the context of control of distributed energy resources, fault detection of unmonitored PV systems and enabling data-driven PV generation forecasting in those situations where information from PV plants is not available due to, for example, privacy concerns or lack of adequate communication infrastructure. The performance assessment and comparison was performed by using experimental data from a real-life setup, which includes PV generation from 4 different sites and demand of 4 different households. It was possible to estimate PV generation with a root mean square estimation error in the range 4.0 ÷ 9.6% with respect to the nominal power of the PV plant.

APPENDIX A

ON THE CONVEXITY OF METHOD B AND C

We discuss on the convexity of the problem (9)-(11). Let \( P = [P_1, P_2, \cdots, P_K], I_j = [I_{j,1}, \cdots, I_{j,K}], j = 1, \ldots, J. \) Let \( M \in \mathbb{R}^{K \times J} = (I_1, I_2, \ldots, I_J) \) be the matrix obtained by stacking horizontally the GNI columns. The estimated total PV production (1) is (matrix product):

\[
\hat{G} = M\alpha.
\]  

Replacing into (4) yields:

\[
\hat{P} = \hat{L} - M\alpha = (\mathbb{1}_{K \times K}, -M)(\hat{L}, \alpha) = 5\varepsilon
\]  

where \( \mathbb{1} \) is the \( K \times K \) identity matrix, \( S = (1_{K \times K}, -M) \in \mathbb{R}^{K \times (K+J)} \) and \( \varepsilon = (\hat{L}, \alpha)^T \). The least square cost function (9) is (matrix product):

\[
J = (P - \hat{P})^T (P - \hat{P}) = P^T P + \hat{P}^T \hat{P} - 2P^T \hat{P},
\]  

Minimizing the expression above equals minimizing (minimizing is invariant with respect to excluding constant terms and to scale factor):

\[
J = \hat{P}^T \hat{P} - 2P^T \hat{P} = (S\varepsilon)^T S\varepsilon - 2P^T S\varepsilon = \varepsilon^T S^T S\varepsilon - 2P^T S\varepsilon = \frac{1}{2}\varepsilon^T H\varepsilon - f^T \varepsilon,
\]  

where (15) was used, and \( H = S^T S \) and \( f = S^T P \) are introduced. The expression above is convex if \( H \) is semidefinite positive, which can be verified numerically by checking the input data. Convexity cannot be therefore enforced by construction, but it depends on input data. In the case input data does not lead to a convex formulation, it was empirically noted that adding a regularization term to the matrix quadratic matrix \( H' = H + \beta \cdot \mathbb{1} \) (with \( \beta = 1 \times 10^{-4} \), whose value was not found to impact substantially algorithms performance), in the form of a ridge regression, achieves convexity.

REFERENCES

[1] M. Delfanti, M. Merlo, G. Monfredini, V. Olivieri, M. Pozzi, and A. Silvestri, “Hosting dispersed generation on italian mv networks: Towards smart grids,” in Proceedings of 14th International Conference on Harmonics and Quality of Power - ICHQP 2010, Sept 2010, pp. 1–6.

[2] S. Hashemi, J. Østergaard, T. Degner, R. Brandl, and W. Heckmann, “Efficient control of active transformers for increasing the pv hosting capacity of lv grids,” IEEE Transactions on Industrial Informatics, vol. 13, no. 1, pp. 270–277, Feb 2017.

[3] O. Gárgica, P. H. Nguyen, W. L. Kling, and T. Uhl, “Microinverter curtailment strategy for increasing photovoltaic penetration in low-voltage networks,” IEEE Transactions on Sustainable Energy, vol. 6, no. 2, pp. 369–379, April 2015.

[4] M. N. Kabir, Y. Mishra, G. Ledwich, Z. Y. Dong, and K. P. Wong, “Coordinated control of grid-connected photovoltaic reactive power and battery energy storage systems to improve the voltage profile of a residential distribution feeder,” IEEE Transactions on Industrial Informatics, vol. 10, no. 2, pp. 967–977, May 2014.

[5] D. Vanhoudt, D. Gysen, B. Claessens, F. Leemans, and L. Jespers, “Efficient control of active transformers for increasing the pv hosting capacity of lv grids,” IEEE Transactions on Industrial Informatics, vol. 8, no. 1, Jan 2017.

[6] R. Luthander, J. Widén, D. Nilsson, and J. Palm, “Photovoltaic self-consumption in buildings: A review,” Applied Energy, vol. 142, 2015.

[7] S. Hashemi and J. Østergaard, “Efficient control of energy storage for increasing the pv hosting capacity of lv grids,” IEEE Transactions on Smart Grid, vol. PP, no. 99, pp. 1–1, 2016.

[8] A. Molina-Garcia, R. A. Mstromauro, T. Garcia-Sanchez, S. Pugliese, M. Liserre, and S. Stasi, “Reactive power flow control for pv inverters voltage support in lv distribution networks,” IEEE Transactions on Smart Grid, vol. 8, no. 1, Jan 2017.

[9] M. Brenna, F. Forcellini, M. Longo, and D. Zaninelli, “Energy storage control for dispatching photovoltaic power,” IEEE Transactions on Smart Grid, vol. PP, no. 99, pp. 1–1, 2016.
[10] F. Sossan, E. Namor, R. Cherkaoui, and M. Paolone, “Achieving the 
dispatchability of distribution feeders through prosumers data driven 
forecasting and model predictive control of electrochemical storage,” 
*IEEE Transactions on Sustainable Energy*, vol. 7, no. 4, Oct 2016.

[11] A. Bracale, G. Carpinelli, and P. D. Falco, “A probabilistic competitive 
ensemble method for short-term photovoltaic power forecasting,” *IEEE 
Transactions on Sustainable Energy*, vol. 8, no. 2, April 2017.

[12] F. Golestaneh, P. Pinson, and H. B. Gooi, “Very short-term nonpara-
meter probabilistic forecasting of renewable energy generation – with 
application to solar energy,” *IEEE Transactions on Power Systems*, 
vol. 31, no. 3, Sept 2016.

[13] E. C. Kara, C. M. Roberts, M. D. Tabone, L. Alvarez, D. S. Callaway, 
and E. M. Stewart, “Towards real-time estimation of solar generation 
from micro-synchrophasor measurements,” *CoRR*, vol. abs/1607.02919, 
2016. [Online]. Available: http://arxiv.org/abs/1607.02919

[14] E. Scolari, F. Sossan, and M. Paolone, “Photovoltaic model-based 
solar irradiance estimators: Performance comparison and application 
to maximum power forecasting,” Submitted to *IEEE Transactions on 
Sustainable Energy*, 2017.

[15] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. De Brabandere, 
“Photovoltaic and solar forecasting: state of the art,” *IEA PVPS, Task*, 
vol. 14, 2013.

[16] J. E. Hay and J. A. Davies, “Calculation of the solar radiation incident 
on an inclined surface,” in *Proc. of First Canadian Solar Radiation Data 
Workshop (Eds: JE Hay and TK Won)*, Ministry of Supply and Services 
Canada, vol. 59, 1980.

[17] E. L. Maxwell, “A quasi-physical model for converting hourly g游击 
horizontal to direct normal insolation,” Solar Energy Research Inst., 
Golden, CO (USA), Tech. Rep., 1987.

[18] W. De Soto, S. Klein, and W. Beckman, “Improvement and validation 
of a model for photovoltaic array performance,” *Solar energy*, vol. 80, 
no. 1, pp. 78–88, 2006.

[19] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky, “Trend Filtering.” *SIAM 
Review*, vol. 51, no. 2, pp. 339–360, 2009.

[20] R. J. Tibshirani, “Adaptive Piecewise Polynomial Estimation via Trend 
Filtering,” arXiv, 2013.

[21] N. Hatziargyriou, *Microgrids: Architectures and Control*, ser. Wiley - 
IEEE. Wiley, 2014.

[22] D. Piga, A. Cominola, M. Giuliani, A. Castelletti, and A. E. Rizzoli, 
“Sparse Optimization for Automated Energy End Use Disaggregation,” 
*IEEE Transactions on Control Systems Technology*, vol. 24, no. 3, 2016.

[23] P. D. Welch, “The Use of Fast Fourier Transform for the Estimation of 
Power Spectra: A Method Based on Time Averaging Over Short, Modi-
fied Periodograms,” *IEEE Trans. Audio and electroacoustic*, vol. 15, pp. 
70–73, 1967.

[24] P. W. Holland and R. E. Welsch, “Robust regression using iteratively 
rewighted least-squares,” *Communications in Statistics - Theory and 
Methods*, vol. 6, no. 9, pp. 813–827, 1977.

[25] P. J. Huber and E. M. Ronchetti, *Robust Statistics 2e*, 2009, vol. 523, 
no. 3.