Data Economy for Prosumers in a Smart Grid Ecosystem

Ricardo J. Bessa
INESC TEC
Porto, Portugal
ricardo.j.bessa@inesctec.pt

David Rua
INESC TEC
Porto, Portugal
drua@inesctec.pt

Cláudia Abreu
INESC TEC and FEUP
Porto, Portugal
claudia.r.abreu@inesctec.pt

Paulo Machado
INESC TEC
Porto, Portugal
paulo.a.machado@inesctec.pt

José R. Andrade
INESC TEC
Porto, Portugal
jose.r.andrade@inesctec.pt

Rui Pinto
INESC TEC and FEUP
Porto, Portugal
rui.b.pinto@inesctec.pt

Carla Gonçalves
INESC TEC and FCUP
Porto, Portugal
carla.s.goncalves@inesctec.pt

Marisa Reis
INESC TEC and FEUP
Porto, Portugal
marisa.m.reis@inesctec.pt

ABSTRACT

Smart grids technologies are enablers of new business models for domestic consumers with local flexibility (generation, loads, storage) and where access to data is a key requirement in the value stream. However, legislation on personal data privacy and protection imposes the need to develop local models for flexibility modeling and forecasting and exchange models instead of personal data. This paper describes the functional architecture of an home energy management system (HEMS) and its optimization functions. A set of data-driven models, embedded in the HEMS, are discussed for improving renewable energy forecasting skill and modeling multi-period flexibility of distributed energy resources.

CCS CONCEPTS

• **Computing methodologies** → **Learning paradigms**; • **Applied computing** → **Forecasting**; • **Hardware** → **Smart grid**;

KEYWORDS

Smart grids, demand response, energy management, flexibility, data analytics

1 INTRODUCTION

The deployment of smart meters at the domestic prosumer level is enabling emergent regulated and non-regulated data-based services that can help to boost distributed energy resources (DER) integration and promote customer empowerment [22].

At the regulated level, the ongoing discussion about the future roles of Distribution System Operators (DSO) is focused in smart metering data management (data manager role) and electricity market facilitation (neutral market facilitator role) [19, 24]. In this context, the following paragraphs discuss examples of recent initiatives within the smart grid ecosystem.

The European Union (EU) project UPGRID developed and demonstrated a Neutral Market Access Platform (NMAP) and a Retailer Platform (RP). The NMAP is hosted by a DSO and encompasses the exchange of information, including consumption profiles from the DSO and flexibility profiles from the Home Energy Management System (HEMS) [3]. The RP is responsible for receiving information or requests from the UPGRID platform, process and send this information to its HEMS. The EU project FLEXICIENCY developed and demonstrated a pan-European Market Place that aims at delivering services and exchange of data, tools, methodologies, in a standardized way across Europe [28]. The platform receives/submits data/service request and readdresses requests to the DSO and service providers platforms where the data, services, software and tools are located [6]. Green Button (U.S.A.) is an industry-led work that provides a common format for electrical energy metering data so that electricity consumers can access their data in an easily readable and secure format via a "Green Button" on their electric utilities' website. Once customers access their data, they can share it as they choose, by independent choice and action, with those they trust. Third parties (e.g., energy retailer, energy services company) services are also emerging with this initiative, e.g. a solar developer could use customer-metering data to determine optimal system size with a more accurate cost-benefit analysis [12]. A review of data management models in eight European countries can be found in [24].

In terms of non-regulated services or third-party services, several works in the literature describe data-driven services and business

cases to boost demand response potential and promote energy efficiency. Smart meter data can be used by a third-party for segmentation of customers and identification of temporal consumption patterns [10], predicting customer response to price signals [17], estimating the price elasticity of customers [13] and derive optimal bidding strategies under dynamic electricity tariffs [27]. In fact, total energy consumption at the residential level is enough to derive a ranking of thermal-load flexibility and sub-metering is not required [2].

Smart meter and HEMS data can also be used to induce behavioral changes in energy consumption [30], for instance through gamification techniques [14] or broadcast of information (e.g., price signal) [11]. In [23], a detailed analysis of the requirements for different applications of smart meter data (e.g., balancing, demand response, network planning) is conducted considering the European legislation about personal data protection.

The General Data Protection Regulation (GDPR) approved by the EU Parliament on 14 April 2016, harmonizes data privacy laws across Europe, protect and empower all EU citizens data privacy [31]. The new regulation introduces a new fine (i.e., up to 4% of annual global turnover or 20 M€), intelligible and easily accessible form for data access consent, right to access data and be forgotten (erase personal data and cease further dissemination of the data). Moreover, it requires "privacy by design", which means inclusion of data protection from the start in systems designing, rather than an addition. Therefore, in order to build data economy or data marketplaces focused on electrical energy consumption data, it is necessary to design knowledge extraction methods that ensure data privacy from the beginning.

Presently, some researchers are starting to design demand response algorithms that avoid sharing and transfer of personal data. For instance, data-driven pricing schemes for load shifting without access to personal load requirements [34] or information exchange models where consumers keep their load levels private and participate in a real-time price scheme [11].

In this context, the present paper presents the Horizon 2020 InteGrid project's vision for the HEMS and its embedded intelligent functions. The core goal is to design forecasting algorithms (load, solar, flexibility) keeping data local and private and at the same time create economic value for stakeholders such as retailers and aggregators. The paper starts by describing the functional architecture, hardware and software integration of the HEMS in section 2. Then, proposes distributed forecasting services in section 3 and flexibility forecast and modeling of behind the meter energy resources in section 4. The potential for future work is discussed in section 5.

2 HOME ENERGY MANAGEMENT SYSTEM

The HEMS concept was initially introduced as a central unit located within a domestic building with the capability of performing an optimized control of behind the meter energy resources [18]. This optimization was loosely defined as a cost reduction procedure, typically achieved through a lower cost use of energy considering internal factors (e.g., devices and system that are able to provide energy use flexibility) and external factors (e.g. weather conditions, discriminated energy prices). HEMS can be exploited as a platform

that is able to support demand side management schemes allowing a generalized participation of consumers in such services [32].

The main characteristic of current HEMS implementations is the ability to allow the monitoring of energy consumption by means of existing metering devices and a user interface that allows the representation of data in a user-oriented way (e.g. dashboards). Another characteristic is the ability to automate existing resources and by using very simple control mechanisms allow the operation of devices and systems at more convenient hours (e.g. pre-program load activation to hours in which the energy costs are lower). The use of sensor technology for remote control of appliances based on a threshold or an user-defined activations allows exploiting energy savings potential. Optimization schemes are currently being sought to allow the optimization of existing flexible energy resources according to multiple criteria and multiple restrictions.

To enhance the capability of fully optimizing the energy use in households, HEMS have been incorporating energy models of devices, systems and spaces so that their specific characteristics along with user preferences can in fact be properly considered when defining their optimal activation [25]. Users are thus able to insert configurations and preferences and take advantage of local flexibility to activate their energy resource at an optimal time. This involves a planning stage and a time-ahead operation, being an example the case where the end-user sets the preferences and configurations, an optimal operation schedule is determined, and in the next day the existing resources are activated according to the schedule.

2.1 Functional Architecture

HEMS are currently flexible and modular HW and SW platforms, capable of supporting a wide variety of features and functionalities the allow end-users to take advantage of existing incentives and define a customizable operation schedule. In Fig. 1 there is a representation of the architecture of the HEMS under design for InteGrid project.
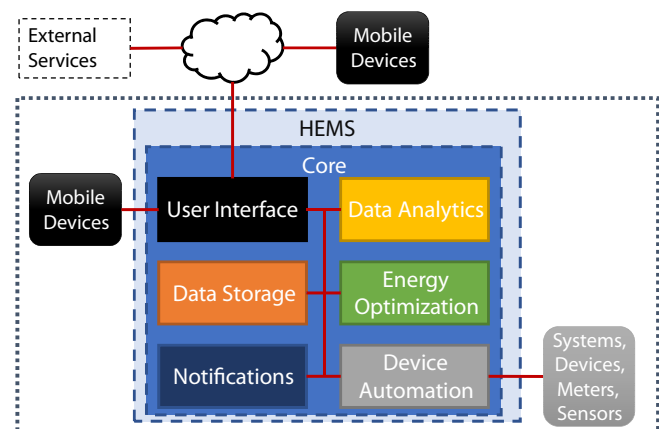

Figure 1: HEMS functional architecture.

The functionalities of a HEMS are typically grouped into: user interface, energy optimization and automation. User interface (UI) is responsible for presenting data and relevant information to the

user and allow the insertion of configurations and preferences. The energy optimization creates an energy representation of the household, through the parameterization of the energy models of devices and systems according to their configuration, and according to the selected criteria establish a time-ahead operation schedule. The automation takes care of the communication between the HEMS central system and the associated devices and systems to retrieve monitoring and operation state data and to set remote configurations or actions that allow an optimal schedule to be implemented.

Other groups can be defined to provide additional features to HEMS: data analytics, data storage and event notifications. Data analytics provides data handling functions related to forecasting and machine learning to extract context information about the availability of energy resources and likelihood of specific energy consumption patterns. Section 3 describes in detail the forecasting features. Data storage functionalities ensure secure data storage and availability to mobile terminals that interact with the HEMS both locally and remotely. Event notification manages alarms and announces unexpected events according to different priority schemes, allowing end-users to received filtered information. They also allow end-users to use non-smart loads, optimally, by receiving that information with a configurable time-ahead.

2.2 Optimization Strategies

One of the distinguishable feature that an HEMS must provide is the capability of computing optimal (or suboptimal) energy use schedules that provide added-value to the end-users. There are different optimization strategies that can be exploited through a HEMS and they largely depend on the criterion (or set of criteria) that might be established. Typically, a cost reduction benefit is sough, and that can be achieved either through energy consumption reduction or through an optimal load allocation considering energy consumption prices. While the former is associated to behavioral changes the latter is more related to the technical aspects of the energy management. In [1], an HEMS implementation, based on the work carried out in the AnyPLACE project, is presented, with the mathematical formulation of optimization strategies for cost reduction considering the variable and fixed components of the energy consumption costs.

The HEMS being designed for InteGrid project is able to optimize the energy use considering variable price tariffs or the local production of PV. As presented in Fig. 2 loads can be allocated to time periods in which the price is lower to allow a reduction on energy expenditures.

When considering PV microgeneration, controllable appliances are shifted toward PV production (according to the forecast generated in the previous day), as portrayed in Fig. 3.

Thermal loads such as air conditioners (AC) and Electric Water Heaters (EWH) have specific energy models whereas dishwashers (DW), washing machines (WM) and cloths dryers (CD) are modeled as shiftable loads. These are modeled as average power and average operating time either input by the end-user or based on estimated load consumption profiles.

One of the objectives of InteGrid is to make use of a fully representable energy model for devices, systems and spaces to allow
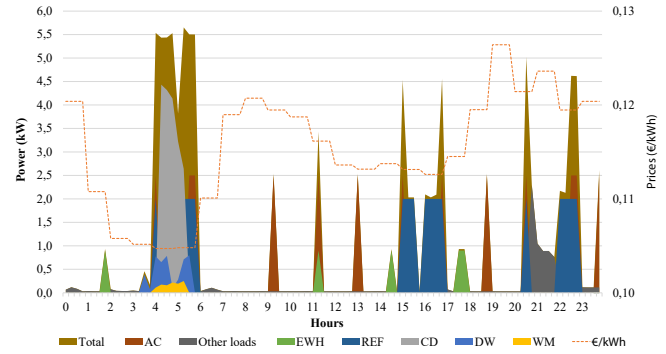

Figure 2: Example of price-based optimization.

a HEMS a higher observability and controllability with positive impact in the quality of the optimization process.
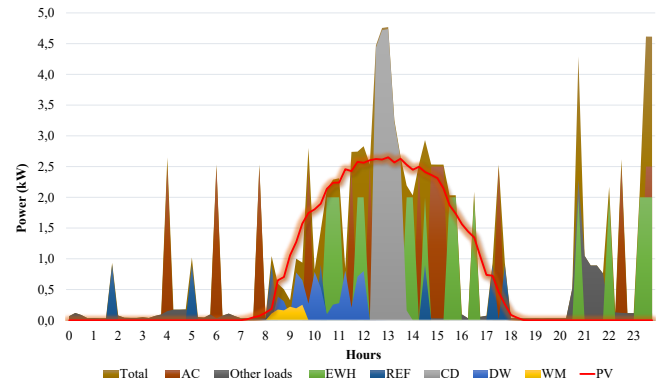

Figure 3: Example of PV self-consumption optimization.

2.3 Hardware Integration

One key aspect of a HEMS is the ability to integrate existing hardware as well as new devices and systems upon which it can leverage to produce more and more complex energy optimization schemes that exploit conveniently their specific technical characteristic and the users' preferences.

The support of these existing or new systems provide a significant improvement on the user experience and comfort by the reduction of manual inputs needed that can be directly obtained through these systems. Also, since the accuracy of those inputs tend to be higher, the quality of the results is improved.

This underlying modular and interoperable nature addresses also the fact that not all the devices and systems that the HEMS may manage are smart or automated, considering means of integrating older devices through user engagement, by means of friendly notifications (email or other) inviting the users to actively control the devices at home, or through the use of other devices like smart plugs that can be integrated to provide some sort of automation.

In order to keep implementation costs low and at the same time ensure a robust computational solution a Raspberry Pi 3 single board computer is used. The use of Linux allows the use of Open-HAB as an open automation support platform that makes use of

binding implementations to exchange information with different devices and systems. Despite there is a considerable number of manufacturers providing bindings for building automated devices it is possible to develop binding implementations to support specific devices, which means that virtually any device can be integrated within the HEMS platform.

3 FORECASTING SERVICES

3.1 Concept

The works in [5, 9, 29] showed that geographically distributed time series data can improve the renewable energy forecasting skill up to six lead-times ahead. Time series data from different sources or owners are combined in a vector autoregression (VAR) model and the LASSO penalty structure is used to uncover sparse structures in the coefficients matrix. The VAR process captures linear interdependencies among multiple time series, enabling each HEMS to model its PV time series evolution based not only on its own lagged values but also in the lagged values collected by other HEMS. The main challenge is on how to combine data from multiple HEMS owners and maintain data privacy, which is analyzed from the mathematical point of view in subsection 3.3.

Furthermore, information from a spatial grid of Numerical Weather Predictions (NWP) is valuable to improve renewable energy forecast accuracy for multiple days [4]. Here, the challenge is to process the NWP grid data and run the statistical forecasting algorithms locally at the HEMS level. Subsection 3.2 presents computational results of forecasting algorithms (section 2) running in the HEMS hardware.

Figure 4 illustrates the models hub concept that serves two goals: (a) central node in a collaborative forecasting scheme (i.e. VAR model) where each HEMS exchanges the matrices from the B and H-update steps of the Alternating Direction Method of Multipliers (ADMM) method [7]; (b) apply feature engineering techniques that extract information from the NWP grid and send the post-processed variables to the HEMS where a embedded gradient boosting tree algorithm is used to produce solar power forecasts.
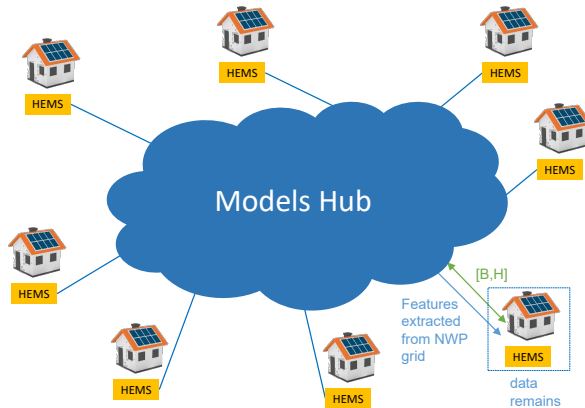

Figure 4: Models hub concept for renewable energy forecasting.

3.2 Embedded Forecasting

Gradient boosting trees (GBT) algorithm combined with feature engineering techniques capable of extracting information from a spatial grid of NWP and resuming it in a new smaller subset of variables can led to substantial improvements on point and probabilistic forecasts for short-term horizons (i.e., up to 72 hours ahead) [4]. However, the high quantity of information to process represents a major constraint and makes it prohibitive to deploy in small computational units such as HEMS. Table 1 presents a comparison between the technical specifications of the HEMS and a conventional desktop computer.

Table 1: Computational resources comparison between the HEMS and a conventional desktop computer.

Comp.	Raspberry Pi 3	Desktop
CPU	Broadcom BCM2837 ARM Cortex-A53, 1.2GHz 4 cores, 4 threads	Intel(R) Core(TM) i7-6700, 3.4GHz 4 cores, 8 threads
RAM	1GB LPDDR2@900MHz	12 GB DDR4@2133MHz
OS	Linux-Raspbian	Windows 10

Table 1 reveals that the RAM size and processing power of the HEMS are a clear limitation when trying to process large quantities of NWP data in an acceptable computational time. For this reason, the envisioned solution uses a centralized distributed computing framework, nested in the Models Hub platform, to process the raw NWP grid data and create relevant features from the NWP grid. These variables, combined with NWP forecasts for the client location, feed a GBT model that runs locally in each HEMS.

For the HEMS, the challenge of embedded forecasting is summarized in four phases:

- Initial request to retrieve historical post-processed NWP grid variables from the Models Hub platform, for the timespan of historical observed PV data.
- Apply feature engineering techniques to extract temporal information from NWP variables for the HEMS location.
- Fit the GBT model using the historical of observed PV generation values and respective NWP variables. The model is updated on a daily or weekly basis.
- Generate PV forecasts based on the NWP variables for an horizon up to 48 hours ahead and the post-processed operational forecasts requested to the Models Hub platform.

3.2.1 Accuracy Results Analysis. In this subsection, improvements on the PV forecasting skill resulting from the introduction of spatial-temporal features extracted from the NWP grid are demonstrated. The metrics Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are considered to assess the point forecast quality. The Continuous Ranked Probability Score (CRPS) is used to evaluate the quality of probabilistic forecasts. An extended description of the feature engineering process can be found in [4].

The importance of variables extracted from the NWP grid is here evaluated by comparing two models that contain extra temporal and spatial information with a base reference model exclusively

composed by seasonal and NWP variables for the grid central point. Table 2 describes the input information of the base model.

Table 2: Base model input variables.

Type	Variables
Seasonal	Month of the year
	Hour of the day
NWP	swflx [W/m^2] - shortwave flux
	temp [K] - ambient temperature (2 meters)
	cfl [0,1] - cloud cover at low levels
	cfm [0,1] - cloud cover at mid levels
	cfh [0,1] - cloud cover at high levels
	cft [0,1] - cloud cover at low and mid levels

The following two models accommodate the new information. Model T is representative of the information that can be extracted from the NWP series for the HEMS location, and Model F comprises a selection of the best spatial and temporal variables that were able to maximize the forecast skill.

- **Model T** - Temporal information extracted from the NWP series to the HEMS location:
  - Lags and leads.
  - Temporal variance with centered windows of 3h, 7h, 11h.
  - Information from different NWP runs.
- **Model F** - Spatial information extracted from the NWP grid:
  - Hourly spatial standard deviation of NWP grid variables.
  - Hourly spatial weighted average of NWP grid variables.
  - Principal components applied individually to the grid information of each NWP variable mentioned in Table 2.

The performance of these models is evaluated over a timespan of two years (from May 1st 2015 to 28th June 2016) with a sliding window of 12 months. Considering ϵ as a metric score, and *base* as a the base model, the improvement of a model is given by $\left(1 - \frac{\epsilon_m}{\epsilon_{base}}\right) \times 100$ (%).

Figure 5 depicts the average improvements of every aforementioned model over the reference base model for forecasts within an horizon of 24 hours ahead. The evaluation is computed on out-of-sample datasets, guaranteeing that in each fold the training data is not contaminated by test samples. The night periods were removed for the PV generation.

An analysis of the figure shows that the two new models greatly outperform the base model in every month of the evaluation period. By itself, the collection of temporal variables (Model T) already provides great average improvements of 13.73%, 10.35%, 11.33% on MAE, RMSE and CRPS that peak in November 2015 at 34.48%, 24.67%, 23.32%. However, the maximum forecast skill is only obtained by including the spatial information (Model F), which led to average improvements of 16.09%, 12.85%, 13.11% for the same metrics, that peak at 37.75%, 28.43%, 27.97% in January 2015.

As final remark, it is important to underline that a good forecast quality can be achieved by solely depending on the temporal information extracted from the NWP runs for the HEMS location. However, to maximize point and specially probabilistic forecasts quality, the NWP grid information is necessary.
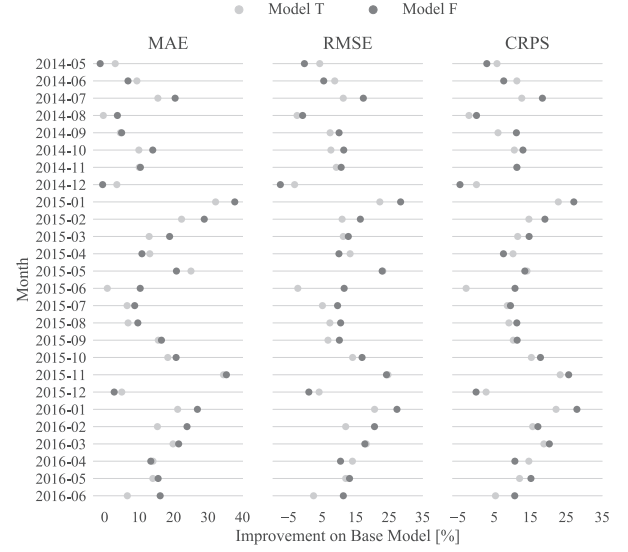

Figure 5: Monthly relative improvements of models T and F over the base model.

3.2.2 Computational Times Evaluation. Although the forecasting algorithms successfully run in the Raspberry Pi 3, a significant impact on the computational time of each model run is verified when compared to a conventional desktop computer. It is important to underline that the GBT regression model requires a separated training for each quantile of the probabilistic forecast. Table 3 shows the total computational times necessary to train all the regression models and to compute forecasts for the two systems described in Table 1.

Table 3: Comparison of computational time results.

Device	GBT Fitting	Operational Forecast
Desktop	42s	0.03s
Raspberry Pi 3	320s	0.49s

3.3 Collaborative Forecasting

3.3.1 VAR-LASSO model and ADMM. Here, a brief review of the VAR-LASSO model is presented, as well as the ADMM formulation for distributed parameter estimation.

Let $y_{i,t}$ be the time series of an i–th HEMS, in time t, and $\{Y_t\} = \{(y_{1,t}, \ldots, y_{n,t})\}$ a n-dimensional vector time series. Then, a VAR model of order p describes the trajectory of Y_t as

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

where $(B^{(\ell)})_{i,j}$ represents the parameters for time series i, associated with lag ℓ of time series j; $\eta = (\mu_1, \ldots, \mu_n)^T$ is the vector of constant terms, and $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n,t})^T$ is a white noise term. By simplification, Y_t is assumed to be a centered process, i.e., $\eta = 0$.

In order to formulate a matrix representation of VAR(p) model, let $Y = (Y_1, \ldots, Y_T) \in \mathbb{R}^{(n \times T)}$ the response matrix, $B = (B^{(1)},$

$\ldots, \mathbf{B}^{(p)}) \in \mathbb{R}^{(n \times np)}$, the coefficient matrix, $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_T) \in \mathbb{R}^{(np \times T)}$ the explanatory variables matrix, with $\mathbf{Z}_t = (\mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_{t-p}) \in \mathbb{R}^{(n \times T)}$, and $\mathbf{E} = (\varepsilon_1, \ldots, \varepsilon_T) \in \mathbb{R}^{(n \times T)}$ the error matrix. Then,

$$\mathbf{Y} = \mathbf{BZ} + \mathbf{E}. \tag{2}$$

Commonly, the system (2) is solved computing the coefficients that minimize $\|\mathbf{Y} - \mathbf{BZ}\|_2^2$, where $\|.\|_r$ will represent both vector and matrix L_r norms. However, high-dimensional data can introduce irrelevant or redundant information making convenient the application of the LASSO framework, which is a regularized version of least squares that introduces an L_1 penalty on the coefficients. In standard VAR-LASSO approach, the coefficients are estimated by

$$\arg\min_{\mathbf{B}} \left(\frac{1}{2} \|\mathbf{Y} - \mathbf{BZ}\|_2^2 + \lambda \|\mathbf{B}\|_1 \right), \tag{3}$$

where $\lambda > 0$ is a scalar penalty parameter.

Since the cost function in (3) is non-differentiable, the ADMM constitutes a powerful algorithm to solve this problem, making it possible to perform a parallel optimization. Summarily, the ADMM rewrites the VAR-LASSO objective function (3) replicating the $\mathbf{B}$ variable using the $\mathbf{H}$ variable ($\frac{1}{2}\|\mathbf{Y} - \mathbf{BZ}\|_2^2 + \lambda \|\mathbf{H}\|_1$) and adding an equality constraint imposing $\mathbf{B} = \mathbf{H}$. Then, based in the augmented Lagrangian of this reformulated minimization problem (with penalty parameter $\rho > 0$), the ADMM formulation of (3) consists in the following iterations [9]:

$$\begin{cases} \mathbf{B}^{k+1} := \operatorname{argmin}_{\mathbf{B}} \left(\frac{1}{2}\|\mathbf{Y} - \mathbf{BZ}\|_2^2 + \frac{\rho}{2}\|\mathbf{B} - \mathbf{H}^k + \mathbf{U}^k\|_2^2 \right) \\ \mathbf{H}^{k+1} := \operatorname{argmin}_{\mathbf{H}} \left(\lambda\|\mathbf{H}\|_1 + \frac{\rho}{2}\|\mathbf{B}^{k+1} - \mathbf{H} + \mathbf{U}^k\|_2^2 \right) \\ \mathbf{U}^{k+1} := \mathbf{U}^k + \mathbf{B}^{k+1} - \mathbf{H}^{k+1}. \end{cases} \tag{4}$$

Given that $\|\mathbf{Y} - \mathbf{BZ}\|_2^2$ and $\|\mathbf{H}\|_1$ are decomposable, the minimization problem over $\mathbf{B}$ and $\mathbf{H}$ can be separately solved for distributed data. Therefore, the ADMM provides a desirable formulation for parallel computing.

Figure 6 illustrates the two most commonly used approaches to split the optimization problem, in which $\mathbf{Z}$ is partitioned into N row blocks (splitting across predictors) or N column blocks (splitting across examples). Conventionally, this two generic formulations are called Consensus Optimization and Sharing Optimization, respectively. The corresponding ADMM formulation using (4) is presented in the systems of equations (5) and (6),

$$\begin{cases} \mathbf{B}_i^{k+1} := \operatorname{argmin}_{\mathbf{B}_i} \left(\frac{\rho}{2} \|\mathbf{B}_i^k \mathbf{Z}_i + \overline{\mathbf{H}}^k - \overline{\mathbf{BZ}}^k - \mathbf{U}^k - \mathbf{B}_i \mathbf{Z}_i\|_2^2 + \right. \\ \left. \qquad\qquad\qquad\qquad\qquad\qquad\qquad \lambda \|\mathbf{B}_i\| \right) \\ \overline{\mathbf{H}}^{k+1} := \frac{1}{N+\rho} \left(\mathbf{Y} + \rho\overline{\mathbf{BZ}}^{k+1} + \rho\mathbf{U}^k \right) \\ \mathbf{U}^{k+1} := \mathbf{U}^k + \overline{\mathbf{BZ}}^{k+1} - \overline{\mathbf{H}}^{k+1} \end{cases} \tag{5}$$

$$\begin{cases} \mathbf{B}_i^{k+1} := \operatorname{argmin}_{\mathbf{B}_i} \left(\frac{1}{2} \|\mathbf{Y}_i + \mathbf{B}_i \mathbf{Z}_i\|_2^2 + \frac{\rho}{2} \|\mathbf{B}_i - \mathbf{H}^k + \mathbf{U}_i^k\|_2^2 \right) \\ \mathbf{H}^{k+1} := S_1 \left(\overline{\mathbf{B}}^{k+1} + \overline{\mathbf{U}}^k, \frac{\lambda}{N\rho} \right) \\ \mathbf{U}_i^{k+1} := \mathbf{U}_i^k + \mathbf{B}_i^{k+1} - \mathbf{H}^{k+1} \end{cases} \tag{6}$$

where $\overline{\mathbf{BZ}}^k = \frac{1}{N} \sum_{j=1}^N \mathbf{B}_j^k \mathbf{Z}_j$, $\overline{\mathbf{B}}^k = \frac{1}{N} \sum_{j=1}^N \mathbf{B}_j^k$ and S_1 is the scalar soft thresholding operator, defined as $S_1(x, a) = \frac{x}{|x|} \max\{0, |x| - a\}$.
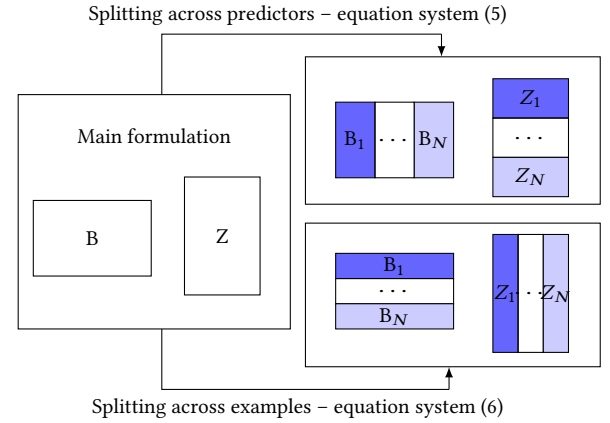

Figure 6: Two main scenarios for distributed computation.

3.3.2 Can ADMM ensure data privacy in collaborative forecasting? The main challenge is to investigate the potential of these iterative systems of equations to develop a collaborative forecasting with data privacy preserved. However, as will be concluded, the direct application of these two alternatives cannot fulfill the privacy requirement.

If the problem is divided across predictors, then $\hat{\mathbf{Y}} = \sum_i \mathbf{B}_i \mathbf{Z}_i$ and each HEMS will estimate $\mathbf{B}_i$ using $\mathbf{Z}_i$, which is composed uniquely by the lags of its own time series. Hence, each HEMS computes the corresponding $\mathbf{B}_i$ without share data. The problem is the update of $\mathbf{H}$ where each HEMS should share its $\mathbf{B}_i \mathbf{Z}_i + (\mathbf{0}, \mathbf{Y}_i, \mathbf{0})$, allowing the more curious HEMS to recover $\mathbf{Y}$.

On the other hand, if the problem is divided across examples, then each $\mathbf{B}_i$ (computed in parallel) is estimated using $\mathbf{Z}_i$, which in this case consists of specific lags of all time series, meaning data is shared between all participants. Nevertheless, this division by examples can be interesting for single-output problems in the energy sector, where private data from multiple consumers is needed.

Notice that in both situations, the existence of a neutral element may be imposed, through which all HEMS communicate, i.e., assuming a centralized process so that HEMS do not exchange results directly. However, this would imply that the neutral element can reconstruct the original data. Furthermore, if in equation (5) the neutral element provides the i–th HEMS $\overline{\mathbf{H}}^k$, $\overline{\mathbf{BZ}}^k$ and $\mathbf{U}^k$, then i–th HEMS can reconstruct $\mathbf{Y}$.

Ideally, the algorithm would be adapted in a way that nobody can reconstruct the original data, be they HEMS or the neutral element. Even better, the algorithm should be decentralized (i.e. peer-to-peer), asynchronous (to cope with delays or communication failures) and time-adaptive (enabling the assimilation of new samples as they become available in order to improve the model with time, updating the coefficients of the model, without starting from scratch).

In the literature, related works may be found. A decentralized structure with asynchrony and delays was proposed in [33], in which the workers can communicate independently with their neighbors, at different times and for different durations. In terms of online ADMM, a general approach may be explored in [16].

Latter work proposed an asynchronous time-adaptive ADMM for consensus optimization [20].

With respect to algorithms that take into account data privacy: the authors of [35] proposed a privacy-preserving decentralized (but neither asynchronous nor online) optimization method, introducing time-varying penalty matrices on ADMM method and using partially homomorphic cryptography.

4 SURROGATE MODELS FOR ENERGY FLEXIBILITY

This section describes three surrogate representations of behind the meter energy resources flexibility. The main idea is to exchange flexibility models estimated in the HEMS, instead of behind-the-meter data, which can be take the form of: set of flexibility trajectories, machine learning model or dynamic virtual batteries. It is important to stress that other representations are possible, such as simplified models to quantify thermal-based demand response potential and the HEMS only sends upstream the estimated model and not the consumption data [2].

4.1 Set of Flexibility Trajectories

HEMS can include diverse equipment such as electric water heaters (EWH), domestic small-scale batteries, photovoltaic (PV) panels, and HVAC systems. The flexibility that HEMS can offer relates to deviations from the net-load profiles that were expected to occur, and has its explanation based on the flexible nature of the previously enumerated HEMS equipment.

The concept of flexibility trajectory embraces the multi-period HEMS flexibility potential instead of independent single-periods formulation [26] . A feasible flexibility trajectory represents the HEMS potential of reshaping its expected net-load profile while guaranteeing technical operation viability of all equipment and internal constraints such as EWH water temperature, battery state-of-charge, or even room air-temperature. Additionally, user-defined constraint regarding the use of the battery can be also implemented (e.g., aiming at minimizing the energy spilling during PV generation surplus). The uncertainty in the net-load profile was also modeled, namely by considering different PV generation scenarios. Ultimately, a feasible trajectory must comply with the defined constraints for a pre-establish percentage of PV generation scenarios (e.g. feasible in 90% of the PV scenarios).

Defining the feasible domain for the HEMS flexibility potential is a complex task. An Evolutionary Particle Swarm Optimization (EPS) based algorithm was proposed in [26] to search and sample the mentioned feasible domain. The final output is a set of feasible trajectories that can be used to describe the HEMS multi-temporal flexibility potential. Figure 7 depicts a set of 20 feasible flexibility trajectories encompassing a period where PV generation surplus occurs.

4.2 Machine Learning Model

The set of flexibility trajectories described in the previous section can be learned (or presented) by a support vector data description (SVDD) algorithm [8, 26]. SVDD is a machine learning one-class support vector machine that can be used to classify "feasible" or
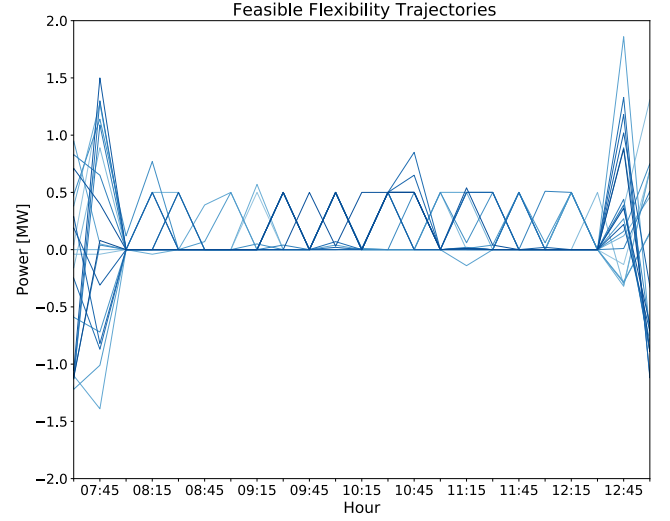

Figure 7: Illustrative set of 20 EPSO-generated feasible flexibility trajectories.

"unfeasible" HEMS operating trajectories upon request from an optimization or control algorithm.

In SVDD, the flexibility trajectories set (X) is summarized with a combination of support vector (x_i) and respective coefficients (β_i). This describes a high-dimension sphere delimiting the feasible domain. A new trajectory (x) is classified by comparing the sphere radius with its radius calculated as follows:

$$R^2(x) = 1 - 2\sum_i \beta_i k(x_i, x) + \sum_{i,j} \beta_i \beta_j k(x_i, x_j) \qquad (7)$$

where R^2 is the square of the radius being calculated, x_i and x_j are support vectors, k is the kernel function. For this problem, it was found that sigmoid kernel is the most suitable type [26]. To be classified as feasible, a trajectory must present a radius lower or equal to the sphere's radius. The results for the test case described in [26] are presented in Table 4.

The main limitation of this representation is that it only allows to classify trajectories as feasible or unfeasible and interpretability (in terms of energy flexibility) is low. However, Eq. 7 can be easily integrated in optimization or control problems and maintain personal data private.

4.3 Dynamic Virtual Battery

An alternative to the SVDD representation for the set of flexibility trajectories is a virtual battery model (similar to the one used in [15]). The virtual battery representation might be seen as a linear system resembling the operation of a battery parameterized by power limits for charge ($P_t{}^{max}$) and discharge ($P_t{}^{min}$) cycles, by maximum ($SOC_t{}^{max}$) and minimum ($SOC_t{}^{min}$) limits for the state of charge (SOC), and also by the initial SOC level (SOC^{ini}).

The power and SOC limits modeled in this work are dynamic throughout the time horizon that is being considered, instead of using a fixed value for all the periods considered. This way, the feasible domain represented by the virtual battery will be more

Table 4: Performance: SVDD vs. Virtual Battery.

	Feasible traj.	Unfeasible traj.
SVDD	95 %	80 %
Virtual Battery	100 %	70 %

adapted regarding the period-based information coming from the set of feasible flexibility trajectories.

The problem formulation for the virtual battery model aims at minimizing the battery size (SOC), as well as minimizing, for each period, the power amplitude represented by the difference between the charge and discharge limits. Equations 8-10 summarize the problem formulation, where $traj_{k, t_i}$ refers to the power value represented by the trajectory k from the flexibility set for the time period t_i, and T refers to the number of periods considered.

$$min \ \sum_{t=1}^{T} SOC_t^{max} + \sum_{t=1}^{T} SOC_t^{min} + \sum_{t=1}^{T} P_t^{max} - \sum_{t=1}^{T} P_t^{min} \quad (8)$$

$$SOC_{t_i}^{min} <= SOC^{ini} + \sum_{t=1}^{t_i} traj_{k, t_t} <= SOC_{t_i}^{max}, \ \ \forall t_i \in [1, T] \ , \ \ \forall k$$

$$(9)$$

$$P_{t_i}^{min} <= traj_{k, t_i} <= P_{t_i}^{max}, \ \ \forall t_i \in [1, T] \ and \ \forall k \quad (10)$$

Compared to the SVDD representation, the virtual battery model can be easily integrated in a optimal power flow formulation, interpreted by an end-user in terms of available flexibility and it is scalable for prosumers aggregation algorithms.

Regarding the performance on correctly classifying trajectories, the virtual battery model shows a better performance compared to the SVDD model when classifying feasible trajectories. On the other hand, for unfeasible trajectories the virtual battery presents a higher number of wrongly classified trajectories. Table 4 presents the percentage of correctly classified trajectories for both models on two different sets: feasible and unfeasible trajectories. Results refer to an 8-hours time horizon.

5 CONCLUSIONS

The increasing awareness of prosumers about data protection and business value of their smart meter data demands for a revision of the current data management paradigms and business use cases. Platforms like HEMS allow consumers to implement energy efficiency actions through automation based on optimal scheduling of existing energy resources. The HEMS are a source of a considerable amount of data and it can include several modules and support different functionalities (e.g., data analytics) to deal with complex data structures.

Forecasting algorithms of renewable energy and net-load benefit from exchange of information between peers (e.g., geographically distributed observations), which shows significant challenges in maintaining data private. A recent trend for mobile devices is the Federated Learning concept[1] that keeps all training data on the

device and only the statistical or machine learning model is shared. Only the updated model is sent to the cloud, using encrypted communication. Furthermore, technologies like TensorFlow Lite for mobile and embedded device developers will enable machine learning tasks in data that does not leave the device and with faster local computation. The advantages are clear, no dependency on network connection and training with less data. However, collaborative forecasting with data privacy remains an open area of research.

The aggregation of flexibility from prosumers also demands for a model-based approach with local computations to extract flexibility parameters. In this context, the interpretability of the virtual battery model is appealing . However, several challenges arise for model-driven representation: inclusion of forecast uncertainty in the flexibility quantification; multi-temporal nature of the flexibility activation; combination of different flexible resources. This type of models have relevant information to implement economic demand response schemes and it is possible to create a marketplace for such models. For instance, virtual batteries flexibility can be traded through the combination of blockchain and smart contracts [21] or integrated in a distributed optimal power flow (OPF) problem.

An interesting initiative is the OpenMined[2] deep learning marketplace that combines federated machine learning, blockchain, multi-party computation, and homomorphic encryption. In this ecosystem, deep learning algorithms fitted in distributed data blocks are traded through smart contracts. The SVDD parameters of a flexibility set or the virtual batteries parameters can be traded in a similar platform.

The future research directions are: (a) a non-linear peer-to-peer vector autoregression model with data fully private and that can be applied to different collaborative forecasting problems; (b) aggregation of virtual battery models that combine prosumer's flexibility and forecast uncertainty and its integration with the HEMS hardware.

In the framework of the InteGrid project, the HEMS and its intelligent functions will be installed in real prosumers in Portugal and Sweden and the value of the proposed approaches will be assessed in a real-world scenario.

ACKNOWLEDGEMENTS

REFERENCES

[1] C. Abreu, D. Rua, P. Machado, J.A. Peças Lopes, and M. Heleno. 2017. AnyPLACE - An Energy Management System to Enhance Demand Response Participation. In *IEEE PowerTech'17*. Manchester.

---

[1]https://research.googleblog.com/2017/04/federated-learning-collaborative.html (access on February 2018)

[2]https://openmined.org/ (accessed on February 2018)

[2] A. Albert and R. Rajagopal. 2018. Finding the Right Consumers for Thermal Demand-Response: An Experimental Evaluation. *IEEE Transactions on Smart Grid* In Press (2018).

[3] A. Alonso, R. Couto, H. Pacheco, R.J. Bessa, C. Gouveia, L. Seca, J. Moreira, P. Nunes, P. Matos, and A. Oliveira. 2017. Towards New Data Management Platforms for a DSO as Market Enabler – UPGRID Portugal demo. *CIRED - Open Access Proceedings Journal* 2017, 1 (Oct. 2017), 2926–2930.

[4] J.R. Andrade and R.J. Bessa. 2017. Improving Renewable Energy Forecasting with a Grid of Numerical Weather Predictions. *IEEE Transactions on Sustainable Energy* 8, 4 (Oct. 2017), 1571–1580.

[5] R.J. Bessa, A. Trindade, and V. Miranda. 2015. Spatial-temporal Solar Power Forecasting for Smart Grids. *IEEE Transactions on Industrial Informatics* 11, 1 (Feb. 2015), 232–241.

[6] K. Boukir, D. Waestlund, B. Traverson, F. Schwarzlaender, and S. Defrancisci. 2017. Providing Smart Metering Data Services Through an EU Market Place. *CIRED - Open Access Proceedings Journal* 2017, 1 (2017), 2848–2851.

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. 2011. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning* 3, 1 (2011), 1âĂŞ122.

[8] J. Bremer, B. Rapp, and M. Sonnenschein. 2011. Encoding Distributed Search Spaces for Virtual Power Plants. In *2011 IEEE Symposium on Computational Intelligence Applications In Smart Grid (CIASG)*. Paris, France.

[9] L. Cavalcante, R. J. Bessa, M. Reis, and J. Dowell. 2017. LASSO Vector Autoregression Structures for Very Short-term Wind Power Forecasting. *Wind Energy* 20, 4 (April 2017), 657–675.

[10] C. Chelmis, J. Kolte, and V. K. K. Prasanna. 2015. Big Data Analytics for Demand Response: Clustering Over Space and Time. In *IEEE International Conference on Big Data (IEEE Big Data 2015)*. Santa Clara, CA, USA.

[11] C. Eksin, H. Deliç, and A. Ribeiro. 2018. Demand Response with Communicating Rational Consumers. *IEEE Transactions on Smart Grid* 9, 1 (Jan. 2018), 469–482.

[12] A. Gerza. 2015. *The Value of Interval Meter Data in Solar PV Project Analysis*. Technical Report. Energy Toolbase, ETB Resource Guide. https://www.energytoolbase.com/newsroom/resourceguides/TheValueOfIntervalMeterDataInSolarPvProjectAnalysis

[13] V. Gómez, M. Chertkov, S. Backhaus, and H. J. Kappen. 2012. Learning Price-Elasticity of Smart Consumers in Power Distribution Systems. In *2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm)*.

[14] A. Gustafsson, C. Katzeff, and M. Bang. 2009. Evaluation of a Pervasive Game for Domestic Energy Engagement Among Teenagers. *Computers in Entertainment (CIE)* 7, 4 (2009).

[15] He Hao, Borhan M. Sanandaji, Kameshwar Poolla, and Tyrone L. Vincent. 2015. Aggregate Flexibility of Thermostatically Controlled Loads. *IEEE Transactions on Power Systems* 30, 1 (jan 2015), 189–198. https://doi.org/10.1109/TPWRS.2014.2328865

[16] S. Hosseini, A. Chapman, and M. Mesbahi. 2014. Online distributed ADMM via dual averaging. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE, 904–909.

[17] Jungsuk Kwac and Ram Rajagopal. 2016. Data-Driven Targeting of Customers for Demand Response. *IEEE Transactions on Smart Grid* 7, 5 (Sept. 2016), 2199–2207.

[18] D. Livengood. 2011. *The Energy Box: Comparing Locally Automated Control Strategies of Residential Electricity Consumption under Uncertainty*. Ph.D. Dissertation. MIT.

[19] P. Mallet, P. Granstrom, P. Hallberg, G. Lorenz, and P. Mandatova. 2014. Power to the People!: European Perspectives on the Future of Electric Distribution. *IEEE Power and Energy Magazine* 12, 2 (March 2014), 51–64.

[20] J. Matamoros. 2017. Asynchronous online ADMM for consensus problems. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 5875–5879.

[21] J. Mattila, T. Seppala, C. Naucler, R. Stahl, M. Tikkanen, A. Badenlid, and J. Seppala. 2016. *Industrial Blockchain Platforms : An Exercise in Use Case Development in the Energy Industry*. Technical Report. ETLA Working Papers 43.

[22] M. McGranaghan, D. Houseman, L. Schmitt, F. Cleveland, and E. Lambert. 2016. Enabling the Integrated Grid: Leveraging Data to Integrate Distributed Resources and Customers. *IEEE Power and Energy Magazine* 14, 1 (Jan. 2016), 83–93.

[23] E. McKenna, I. Richardson, and M. Thomson. 2012. Smart meter data: balancing consumer privacy concerns with legitimate applications. *Energy Policy* 41 (Feb. 2012), 807–814.

[24] Council of European Energy Regulators. 2016. *Review of current and future data management models*. Technical Report C16-RMF-89-03. CEER.

[25] Y. Ozturk, D. Senthilkumar, S. Kumar, and G. Lee. 2013. An Intelligent Home Energy Management System to Improve Demand Response. *IEEE Transactions on Smart Grid* 4, 2 (June 2013), 694–701.

[26] R.B. Pinto, R.J. Bessa, and M.A. Matos. 2017. Multi-period Flexibility Forecast for Low Voltage Prosumers. *Energy* 141 (Dec. 2017), 2251–2263.

[27] M. Song and M. Amelin. 2017. Purchase Bidding Strategy for a Retailer with Flexible Demands in Day-ahead Electricity Market. *IEEE Transactions on Power Systems* 32, 3 (May 2017).

[28] J. Stromsather, L. Marretta, and P. Soderstrom. 2015. Opening Up for a More Competitive Energy Market with New Energy Services by Making "Real time" Metering Data Accessible to Market Players. In *CIRED 2015 - The 23rd International Conference and Exhibition on Electricity Distribution*. Lyon, France.

[29] J. Tastu, P. Pinson, P. Trombe, and H. Madsen. 2014. Probabilistic Forecasts of Wind Power Generation Accounting for Geographically Dispersed Information. *IEEE Transactions on Smart Grid* 5, 1 (Jan. 2014), 480–489.

[30] K. Tsuda, M. Uwasu, K. Hara, and Y. Fuchigami. 2017. Approaches to Induce Behavioral Changes with Respect to Electricity Consumption. *Journal of Environmental Studies and Sciences* 7, 1 (March 2017), 30–38.

[31] European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union. (2016).

[32] C. J. C. Williams, J. O. Binder, and T. Kelm. 2012. Demand side management through heat pumps, thermal storage and battery storage to increase local self-consumption and grid compatibility of PV systems. *3rd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies (ISGT Europe)* (14-17 Oct. 2012).

[33] T. Wu, K. Yuan, Q. Ling, W. Yin, and A. H Sayed. 2016. Decentralized consensus optimization with asynchrony and delays. In *Signals, Systems and Computers, 2016 50th Asilomar Conference on*. IEEE, 992–996.

[34] Z. Xu, T. Deng, Z. Hu, Y. Song, and J. Wang. 2018. Data-Driven Pricing Strategy for Demand-Side Resource Aggregators. *IEEE Transactions on Smart Grid* 9, 1 (Jan. 2018), 57–66.

[35] C. Zhang and Y. Wang. 2017. Privacy-preserving Decentralized Optimization Based on ADMM. *arXiv preprint arXiv:1707.04338* (2017).