

Day-Ahead and Intra-Day Building Load Forecast With Uncertainty Bounds Using Small Data Batches

Marco Lauricella¹ and Lorenzo Fagiano², *Senior Member, IEEE*

Abstract—An approach to provide day-ahead and intra-day load forecasts of buildings, such as electrical or thermal power consumption, is presented. The method aims to obtain a nominal forecast and associated error bounds with small data batches of two weeks for the training phase, resulting in a ready-to-go algorithm that can be employed whenever large datasets of months or years are not available or manageable. These cases include new or renovated constructions, buildings that are subject to changes in purpose and occupants' behavior, or applications on local devices with memory limits. The approach relies on a so-called "fictitious input" signal to capture the prior information on seasonal and periodic trends of load consumption. Then, linear multistep predictors with different horizon lengths are trained periodically with a small batch of the most recent data, and the associated worst case error bounds are derived, using set membership (SM) methods. Finally, the forecast is computed, for each time step, by intersecting the error bounds of the different multistep predictions and taking the central value of the obtained interval. Such a method is applied here for the first time to real-world data of electrical power consumption of a medium-size building and of cooling power consumption of a large complex. In both cases, the obtained results indicate a tightening of the worst case error bounds between 15% and 25% on average with respect to those obtained with a standard linear SM approach.

Index Terms—Energy prediction, filtering, load forecasting, set membership (SM) estimation, smart buildings, smart grid.

I. INTRODUCTION

IN BUILDING energy management, accurate forecasting of electric and thermal power consumption, "load" in the remainder, leads to higher energy efficiency, thus lower costs and reduced environmental impact. It also favors the inclusion of buildings in the future smart cities, facing the challenges related to growing urbanization, pollution, and climate change. Indeed, a widespread use of optimization methods for buildings, enabled by the availability of load forecasts, can have a strong impact on the sustainability of our built environment: buildings account in fact for 30% of the

global energy consumption on average, and for nearly 33% of associated CO₂ emissions [1]. This applies especially to the nonresidential scenario, where the facility manager can use forecasts to optimize the energy usage over time by adapting the building operations. Moreover, future smart grids will employ dynamic pricing techniques to improve the demand side management [2]: in this context, a better energy consumption forecast allows one to obtain more favorable purchase plans. Short-term load forecasting has become increasingly important with the rise of competitive energy markets as well, since it plays an essential role in the composition of energy prices [3]. Moreover, robust approaches to building and grid management require an estimate of the worst case prediction error, in addition to a nominal forecast [4], [5], [6]. Such an upper bound on the error's magnitude shall be the tightest possible, to avoid that too conservative measures are taken, e.g., an excessive energy reserve with respect to the actual forecast uncertainty.

A. State-of-the-Art

Motivated by the considerations above, load forecasting for buildings has been widely studied in the literature, considering different time-scales, and using a large variety of approaches, spanning from linear autoregressive models (AR, ARX, ARIMA) [7], [8], [9], [10], to periodic or seasonal ones (seasonal ARIMA, Periodic AR) [11], generalized Fourier series [12], support vector machines (SVMs) [13], artificial neural networks (ANNs) [14], [15], [16], [17] and genetic algorithms (GAs). For a more comprehensive list of forecasting approaches, see e.g., [3], [18], [19], [20], [21], [22], [23], [24], and the references therein. In this work, we focus on short-term load forecasting, with a particular emphasis on the one-day-ahead horizon, since it is the most common for the optimization of energy production, storage and consumption operations, and for energy market transactions [25].

Commercial and industrial buildings are known to exhibit periodic trends in their energy consumption profile, in addition to an influence from weather [26], [27]. For these reasons, several approaches aim to directly take into account such periodic trends, resorting for example to Fourier series [28], periodic AR models [9], [11], or indexed ARX models [8]. Forecasting approaches based on linear autoregressive models, designed to make use of the known periodic trends, often achieve comparable performance to those of neural networks [8], with the advantage of having a simpler and more understandable model structure.

Manuscript received 13 February 2023; accepted 17 April 2023. The work of Marco Lauricella was supported by the ABB—Politecnico di Milano Joint Research Center through the Ph.D. Scholarship "Methodologies for the Estimation and Monitoring From Data of Electric Systems." Recommended by Associate Editor V. Adetola. (*Corresponding author: Marco Lauricella.*)

Marco Lauricella is with ABB Corporate Research, ABB AG, 68526 Ladenburg, Germany (e-mail: marco.lauricella@de.abb.com).

Lorenzo Fagiano is with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milan, Italy (e-mail: lorenzo.fagiano@polimi.it).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCST.2023.3274955>.

Digital Object Identifier 10.1109/TCST.2023.3274955

One main drawback of the majority of the cited techniques is the need for a large dataset to be able to capture the building's dynamics and periodic trends only from data. In most cases, the employed datasets consist of year(s) of measurements with sampling period of the order of minutes. This is a problem when only recent energy consumption data are available, or when the building is subject to a change in purpose and number of occupants (e.g., due to a change in scope of part of the facility and/or personnel relocation), thus making previous data not representative of the new course. Another possible limitation is the available memory for data storage, which can be rather small if local devices, such as smart meters or circuit breakers, are used for data collection and possibly to derive the wanted forecast.

A second drawback of the existing approaches is that they usually don't provide a systematic way to estimate worst case error bounds associated with the nominal forecast, despite the importance of such bounds when the load prediction is used for robust optimization, scheduling, and decision making, as pointed out above.

A number of rather recent contributions in the literature aim to obtain a forecasting model using generative adversarial networks (GANs) [29], [30], [31], in some cases explicitly using "scarce" datasets. Using a generator versus discriminator training process, GANs aim to deliver a neural network able to produce load consumption time-series whose statistical properties match those of the training dataset. This can be used to augment the dataset itself and/or directly as a predictor. However, GANs are not aimed at delivering worst case error bounds associated with a nominal estimate, rather they can be seen as a statistically representative forecast generation mechanism. As such, they are more suited to stochastic optimization/control techniques rather than robust ones. Moreover, the mentioned works still employ relatively large training datasets, ranging from 6 to 14 months in [30], to 8 batches of 28 days each (i.e., 224 days) in [29], to 2 years in [31].

B. Contribution

A common feature in the mentioned literature is that the forecasting model is trained once and for all with a fixed batch of data. This is one of the reasons why rather large datasets are needed, to capture seasonal variations of the consumption. Also, this appears to be unnecessarily limiting, since new data are collected daily and can be used to train/adapt the model. Based on this observation, our approach entails a periodic training procedure with a user-selected frequency. At the beginning of each period, a new forecast model is derived, using a rather small batch of data pertaining to an interval immediately prior to the training time. Such a model is then used to predict the day-ahead load consumption for the whole period, until a new training procedure is carried out. In a previous contribution [32], we presented an algorithm based on this concept, which proved to achieve state-of-the-art performance in terms of accuracy using just two weeks of past data for each forecasting model. Additionally, the approach employs periodic input signals named "fictitious," since they are not measured quantities. These signals are sine



Fig. 1. ABB office building whose electric load consumption has been collected in dataset A.

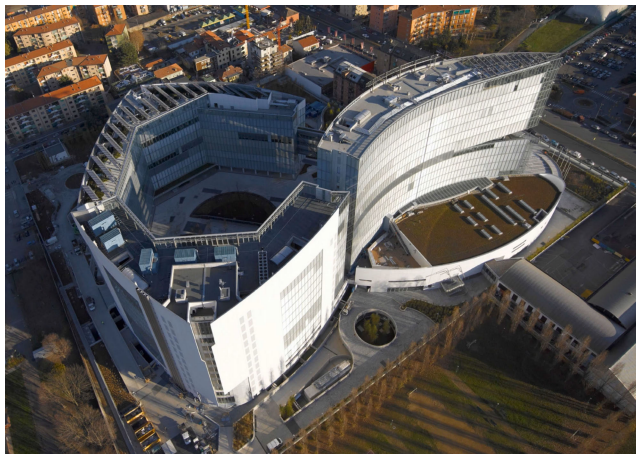


Fig. 2. Business/commercial building complex whose cooling load consumption has been collected in dataset B.

and cosine functions of time, whose frequencies are chosen by the user on the basis of the above-mentioned prior knowledge of periodic building consumption trends, while the amplitudes are identified from data. Finally, the algorithm, based on linear prediction models, allows one to use the set membership (SM) approaches described in [33] to estimate worst case bounds on the forecast error. In this article, we provide three main additional contributions. First, we improve over [32] by delivering a technique that obtains tighter worst case error bounds. To do this, we leverage a multiple SM prediction approach, whose theoretical properties have been introduced in [34]. Second, we apply here the improved method to two real-world buildings: an office building hosting around 200 employees, equipped with electrical heating and cooling (see Fig. 1), and a large complex of five buildings hosting offices for more than 4000 people, two data centers and a commercial area (see Fig. 2). In the former case, the goal is to predict the electric power consumption, while in the latter the focus is on the cooling power. In both cases, weather measurements collected from nearby stations are considered as input to the model, in addition to the mentioned fictitious input. This work represents the first time that the presented improved approaches are tested on data pertaining to real-world buildings. Finally, as third contribution we apply the proposed method also to intra-day load forecasting, in addition to the day-ahead one.

The article is organized as follows. Section II contains the problem formulation and an overview of the approach.

Section III details the ingredients of the forecasting algorithm, describing the predictor structure, the adopted identification method and the local and global forecasting procedures. Section IV introduces the implementation of the proposed approach for the day-ahead and intra-day forecasts, and Section V presents the experimental results. Finally, Section VI concludes the article.

II. PROBLEM DESCRIPTION AND OVERVIEW OF THE APPROACH

We assume to have a dataset of load consumed by the building of interest, averaged every quarter-hour, collected over N subsequent days in the recent past. The choice of quarter-hour sampling periods is typical in this application and is made for the sake of clarity here, without loss of generality of the proposed algorithms. We thus denote with $\tilde{y}_i(k) \in \mathbb{R}$ the average load consumption measured in the k th quarter-hour of day i , with $k = 1, \dots, 96$, and $i = 1, \dots, N$. We adopt the convention that $k \leq 0$ corresponds to the variables of the previous days in a reverse order, so for example

$$\begin{aligned}\tilde{y}_i(0) &= \tilde{y}_{i-1}(96) \\ \tilde{y}_i(-1) &= \tilde{y}_{i-1}(95) \\ \tilde{y}_i(-96) &= \tilde{y}_{i-1}(0) = \tilde{y}_{i-2}(96)\end{aligned}$$

and so on. Moreover, we assume to have weather measurements at the building location during the same N days, typically including measurements of external temperature, relative humidity, solar irradiation and wind speed, averaged every quarter-hour. We denote with $\tilde{u}_{w,i}(k) \in \mathbb{R}^{n_w}$ the vector of weather variables of the k th quarter-hour of day i , where n_w is the number of considered weather variables.

The problem we address is to use the described N -days-long dataset to derive one-day-ahead and intra-day forecasts of the building's load, together with bounds on the forecasting errors. In particular, the forecasts are computed daily for a period of M days, then a new forecasting model is identified, after having updated the N -days-long dataset. Typically, one chooses $N > M$ so that the time series contained in the dataset is longer than the period during which the model is employed. To derive the forecasts, we will use models with an auto-regressive part, i.e., where the predicted load depends on a number $n_y \in \mathbb{N}$ (model order) of past values of the load itself, in addition to the input signals. Procedure 1 provides a high-level description of this recursive modeling approach and serves as a guideline for Sections III and IV of the article.

In the remainder, we set $N = 14$ and $M = 7$, meaning that, at the beginning of each week (e.g., each Monday), a new model is identified from the most recent two weeks of data. This model is then employed to forecast the building energy consumption for each day of the upcoming week. For day-ahead forecasts, the model is reinitialized at the beginning of each day with the last n_y measured load values of the previous day. For intra-day operation, the model is reinitialized during the day with the last n_y load measurements and the forecast is recomputed for a chosen horizon H of future steps. Figs. 3 and 4 illustrate the described model reinitialization for the day-ahead and intra-day forecasting approaches, respectively.

Procedure 1 Overview of the Proposed Load Forecasting Approach

- 1) At the beginning of each M -days-long period, derive a one-step-ahead prediction model of a chosen order n_y , using a simulation error criterion and the dataset collected from the most recent N days, named training dataset. The model identification procedure is described in Section III-B. Also, derive the set of all multi-step predictors that are compatible with the collected data (feasible parameter set—FPS) as described in Section III-C–III-E.
- 2) For each day of the considered M -days-long period, use one of the forecasting algorithms described in Section IV to predict the energy consumption and estimate the associated worst case error bounds, with either a one-day-ahead or an intra-day strategy.
- 3) At the end of the M -days-long period, update the training dataset with the newly available measurements, adopting a moving window criterion with a fixed length of N days, then go to 1).



Fig. 3. Working principle of day-ahead forecasting approach with $n_y = 3$ over the 96 time samples composing a day. Red cells: load measurements used to initialize the predictor; blue cells: forecasting horizon, i.e., future time steps for which the load forecast is computed.

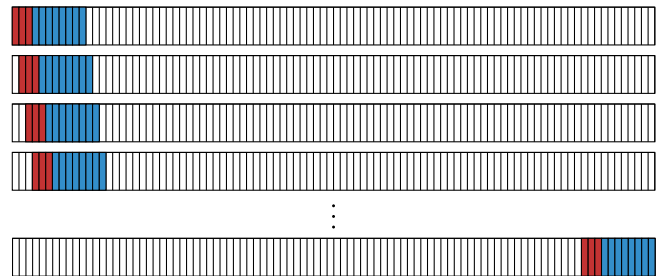


Fig. 4. Graphical representation of the intra-day forecasting strategy working principle over the 96 time samples composing a day, with $n_y = 3$ and $H = 8$. Red cells: load measurements used to initialize the predictor; blue cells: forecasting horizon, i.e., future time steps for which the load forecast is computed.

Note that the adopted model initialization and forecasting horizon, both for day-ahead and intra-day load forecasts, can be arbitrarily chosen according to the specific application needs without any change to the general approach.

III. FORECAST METHOD

The approach proposed in this article is based on the following ingredients:

- A. a fictitious input signal, designed to capture the periodic trends of load consumption;
- B. a linear, discrete time, one-step-ahead dynamical model identified using a simulation error criterion;
- C. a user-chosen number $\bar{p} \geq 1$ of multistep predictors, obtained by iteration of the identified one-step-ahead model considering different horizon values. As it will be shown in the remainder, a larger value of \bar{p} yields tighter worst case error bounds at the cost of higher computational burden; and

D. and E. the estimation of forecasting worst case error bounds for each one of the considered multistep predictors using SM methods, and the derivation of a filtered version of the load forecast with either a local (*D.*) or a global (*E.*) approach.

These ingredients are detailed next sections.

A. Fictitious Input

Buildings have daily, weekly and seasonal patterns in their energy consumption profile [26]; numerous contributions employ this knowledge in different ways, as briefly reported in Section I-A. The common ground of the vast majority of these approaches is the need for a large-enough dataset, to be able to identify the seasonal trends of the analyzed load time series. In this article, we rather adopt the method proposed in our previous work [32], which demonstrated to provide good forecasting performance even with a small dataset. As mentioned in Section I-B, the novelties with respect to [32] are an improved approach to derive worst case error bounds on the computed forecast, the use of such an improved method on two real-world buildings with different size, and the formulation of both a day-ahead and an intra-day strategy. A core element of the approach in [32] is the inclusion of a so-called “fictitious input” signal to excite a linear prediction model in accordance with the expected periodic behavior of the load consumption, thus exploiting such a knowledge. A physical interpretation of the fictitious input is that it accounts for periodic trends of the building’s usage, such as occupancy, servers’ workload, canteen workload, working days versus weekend, etc., and of environmental factors like solar irradiation and external temperature. Note that weather-related signals, when present, can be still used as input to the forecast model in addition to the fictitious one, as we did in the experiments reported in Section V. Said fictitious input signal is given by a linear combination of sine and cosine functions whose harmonics are not estimated from data, but they are a priori chosen, for example as fractions of days and weeks, based on an expert knowledge of the relevant frequency contributions. On the other hand, the amplitude of each sine and cosine is estimated in the training phase. Since the forecast model that we use is linear, the fictitious input signals are eventually summed over the sine and cosine contributions (so that the approach is equivalent to estimating amplitude and phase shift of each frequency component) and also over frequencies, thus resulting in an overall scalar periodic input signal.

For each time step $k = 1, \dots, 96$ of each day $i = 1, \dots, M$, the fictitious input is thus defined as

$$u_{f,i}(k) = \begin{bmatrix} \cos(\omega_1(k + 96(i - 1))) \\ \vdots \\ \cos(\omega_{n_\omega}(k + 96(i - 1))) \\ \sin(\omega_1(k + 96(i - 1))) \\ \vdots \\ \sin(\omega_{n_\omega}(k + 96(i - 1))) \end{bmatrix} \quad (1)$$

where n_ω is the number of considered frequency contributions, and each angular frequency $\omega_j = 2\pi f_j$, $j = 1, \dots, n_\omega$,

corresponds to one of the selected harmonics. This results in $u_{f,i}(k) \in \mathbb{R}^{n_f}$, with $n_f = 2n_\omega$.

Considering the quarter-hour-long sampling period of this application, we select $f_j = (j/(4 \cdot 24 \cdot 7))$, obtaining, in the case of $n_\omega = 14$, a fictitious input signal composed of sine and cosine waves with periods in the interval between seven days ($j = 1$) and 12 h ($j = 14$). Note that the choice of the harmonics could be fine-tuned during operation, for example before training a new model at the end of the M -days-long period (see Procedure 1) by repeating a posteriori the forecast computation and evaluating the obtained error (and error bounds) with different choices of frequencies. Another way to select the values of ω_j is to compute the Fourier transform of the load signal and take the n_ω most relevant frequency components. In our experience with both the buildings considered in this work and with other conceptually similar modeling tasks, the gain in accuracy and tightness of the worst case error bounds obtained by changing the harmonics is rather limited with respect to the choice presented above, provided that n_ω is large-enough, of the order of 10 or more components. Nevertheless, the optimal choice of frequencies and also of functions that compose the fictitious input is a current subject of further research.

B. One-Step-Ahead Model

To obtain the wanted forecast, we start with a linear, discrete time, one-step-ahead predictor, with ARX (auto-regressive with exogenous input) structure. For the generic day i , the prediction model is denoted as

$$\hat{y}_i(k|k-1) = \phi_i^{(1)}(k-1)^T \hat{\theta}^{(1)} \quad (2)$$

where $\hat{y}_i(k|k-1)$ denotes the estimated load at time k given the value at time $k-1$ of the regressor $\phi_i^{(1)}$, which is built as

$$\begin{aligned} \phi_i^{(1)}(k-1) &= [Y_i(k-1)^T U_i(k-1)^T]^T \in \mathbb{R}^{n_y + n_u n_x} \\ Y_i(k-1) &= [y_i(k-1), \dots, y_i(k-n_y)]^T \in \mathbb{R}^{n_y} \\ U_i(k-1) &= [u_{w,i}(k-1)^T, \dots, u_{w,i}(k-n_x)^T, \\ &\quad u_{f,i}(k-1)^T, \dots, u_{f,i}(k-n_x)^T]^T \in \mathbb{R}^{n_u n_x}. \end{aligned} \quad (3)$$

The model auto-regressive order is n_y , while the order of the input-related part is n_x , and $n_u = n_w + n_f$ denotes the number of considered input signals. In (3), the values of $y_i(k)$ are defined as

$$y_i(k) = \begin{cases} \tilde{y}_i(k), & \text{if } k < 1 \\ \hat{y}_i(k|0), & \text{if } k \geq 1 \end{cases}$$

where $\tilde{\cdot}$ denotes a measured sample of a quantity. We adopted this notation to indicate that the predictor is initialized each day with n_y measured values, and it is then iterated in simulation to provide a load forecast.

In (2), $\hat{\theta}^{(1)} \in \mathbb{R}^{n_y + n_u n_x}$ is the vector of model parameters to be identified. In our approach, the value of $\hat{\theta}^{(1)}$ is learned from the dataset using a simulation error criterion. Namely, we look for the model parameters that minimize the squared ℓ_2 -norm of the mismatch between load measurements and their

estimate, obtained by means of open loop simulation of the predictor. The latter is initialized each day with the last n_y load measurements of the previous one. This is done by solving the following unconstrained nonlinear program (NLP)

$$\hat{\theta}^{(1)} = \arg \min_{\theta \in \Omega} \sum_{i=1}^N \|\tilde{Y}_i - \hat{Y}_i(\theta)\|_2^2 \quad (4)$$

where, as described in Section II, N corresponds to the number of days composing the training dataset, Ω is an hypercube representing a compact approximation of $\mathbb{R}^{n_y+n_u n_x}$, introduced to technically replace inf operator with min, and

$$\begin{aligned} \tilde{Y}_i &= [\tilde{y}_i(1), \dots, \tilde{y}_i(96)]^T \\ \hat{Y}_i(\theta) &= [\hat{y}_i(1|0), \dots, \hat{y}_i(96|0)]^T. \end{aligned}$$

The solution to (4) represents step 1) of Procedure 1, and it can be performed using standard sequential quadratic programming algorithms, such as the Gauss-Newton, which are known to be computationally efficient for this class of optimization problems. In general, (4) is not convex, due to the polynomial dependency of each estimated variable, beyond time (1|0), on the parameters $\hat{\theta}^{(1)}$ of the one-step-ahead model. However, with both the considered experimental datasets, we never experienced problems related to possible local minima of (4), with any of the tested order values. The reasons are that the input signals (fictitious input and weather data) are rich enough to persistently excite the model being identified, and that the chosen linear model has a rather small number of parameters with respect to the employed data points. For example, in our experiments (see Section V) with $n_y = 3$ (third-order system), $n_x = 1$, $n_w = 4$ (external temperature, relative humidity, solar irradiation, and wind speed) and $n_f = 28$ we had $3+4+28 = 35$ parameters to be identified for the first building, and 33 parameters for the second one (since there were no solar irradiation and wind speed measurements available, i.e., $n_w = 2$). In both cases, we employed $N = 14$ days, thus $96 \cdot 14 - 3 = 1341$ data points (two weeks with quarter-hour sampling period, minus the model order for initialization).

C. Multistep Predictor

In the remainder, for the sake of notational simplicity we will drop the subscript i indicating the considered day, since all the steps are operated on a daily basis, where the initial conditions and the used fictitious input and weather data are updated every day.

The identified one-step-ahead prediction model (2) can be iterated in simulation to obtain the load prediction at a generic time instant k from a given initial condition at time $k-p$ and with given courses of the inputs up to $k-1$, resulting in the following linear equation:

$$\hat{y}(k|k-p) = \varphi^{(p)}(k-p)^T \hat{\theta}^{(p)}. \quad (5)$$

The regressor $\varphi^{(p)}(k-p) \in \mathbb{R}^{n_y+n_u(n_x+p-1)}$ is given by

$$\varphi^{(p)}(k-p) = [Y^{(p)}(k-p)^T U^{(p)}(k-1)^T]^T$$

with

$$\begin{aligned} Y^{(p)}(k-p) &= [\tilde{y}(k-p), \dots, \tilde{y}(k-p-n_y+1)]^T \in \mathbb{R}^{n_y} \\ U^{(p)}(k-1) &= [u_w(k-1)^T, \dots, u_w(k-p-n_x+1)^T \\ &\quad u_f(k-1)^T, \dots, u_f(k-p-n_x+1)^T]^T \end{aligned}$$

$U^{(p)}(k-1) \in \mathbb{R}^{n_u(n_x+p-1)}$. The entries of $\hat{\theta}^{(p)} \in \mathbb{R}^{n_y+n_u(n_x+p-1)}$ are polynomial functions of the parameters in $\hat{\theta}^{(1)}$, which are readily obtained by recursion of (2) for p steps. As an example, with $n_y = n_x = n_u = 1$ one would obtain

$$\hat{\theta}^{(1)} = \begin{bmatrix} \hat{\theta}_1^{(1)} \\ \hat{\theta}_2^{(1)} \end{bmatrix}, \quad \hat{\theta}^{(2)} = \begin{bmatrix} \hat{\theta}_1^{(1)2} \\ \hat{\theta}_2^{(1)} \\ \hat{\theta}_1^{(1)}\hat{\theta}_2^{(1)} \end{bmatrix}, \dots$$

where $\hat{\theta}_j^{(1)}$ is the j th entry of vector $\hat{\theta}^{(1)}$.

D. Local Multistep Forecast and Error Bounds

Let us now consider the dataset collected during the last N days. For each prediction horizon p that is of interest, we solve the following linear program (LP):

$$\begin{aligned} \hat{\varepsilon}^{(p)} &= \alpha \min_{\varepsilon, \theta^{(p)}} \varepsilon \\ \text{s.t. } & |\tilde{y}(k) - \varphi^{(p)}(k-p)^T \theta^{(p)}| \leq \varepsilon \\ & k = p, \dots, 96N \end{aligned} \quad (6)$$

where $96N$ is the total number of available load measurements. Namely, the quantity $\hat{\varepsilon}^{(p)}$ computed in (6) is the smallest upper bound on the pointwise-in-time error between the load measurements and the corresponding forecast that can be obtained by any linear multistep predictor. It is an estimate of the worst case effect of process-model mismatch and of the effects of exogenous signals such as process disturbance and measurement noise. In (6), $\alpha > 1$ is a scaling parameter used to account for the uncertainty resulting from the use of a finite dataset in the estimation of such an error bound, see [33] for more details.

Following a SM modeling approach, we can then compute the set $\Theta^{(p)}$ of all the linear multistep predictors with horizon p that are consistent with the available measurements and with the estimated worst case error bound (FPS)

$$\Theta^{(p)} \doteq \left\{ \theta^{(p)} : |\tilde{y}(k) - \varphi^{(p)}(k-p)^T \theta^{(p)}| \leq \hat{\varepsilon}^{(p)}, \right. \\ \left. k = p, \dots, 96N \right\}. \quad (7)$$

Finally, for the available multistep predictor $\hat{\theta}^{(p)}$ (computed as mentioned by iterating the one-step-ahead model (2) and a given regressor $\varphi^{(p)}(k-p)$), we can compute a bound on the forecasting uncertainty as

$$\begin{aligned} \hat{\tau}^{(p)}(\varphi^{(p)}(k-p), \hat{\theta}^{(p)}) \\ = \hat{\varepsilon}^{(p)} + \max_{\theta^{(p)} \in \Theta^{(p)}} |\varphi^{(p)}(k-p)^T (\theta^{(p)} - \hat{\theta}^{(p)})| \end{aligned} \quad (8)$$

i.e., the sum of $\hat{\varepsilon}^{(p)}$ with the maximum returned by an LP. The bound $\hat{\tau}^{(p)}(\varphi^{(p)}(k-p), \hat{\theta}^{(p)})$ is named ‘‘local,’’ because it pertains to a specific regressor value $\varphi^{(p)}(k-p)$. If the estimate $\hat{\varepsilon}^{(p)}$ is no smaller than the actual uncertainty bound,

which is guaranteed to happen to the limit with infinitely many and informative data points, then by construction we have

$$\tilde{y}(k) \in [\varphi^{(p)}(k-p)^T \hat{\theta}^{(p)} \pm \hat{\tau}^{(p)}(\varphi^{(p)}(k-p), \hat{\theta}^{(p)})]. \quad (9)$$

In practice, with a finite dataset one can never be sure a priori whether $\hat{\varepsilon}^{(p)}$ is an under- or over-approximation of the actual uncertainty, hence the role of the tuning parameter α in (6) to tradeoff conservativeness and tightness of the error bound. In particular, smaller α values lead linearly to smaller error bounds, hence making them less conservative but increasing the risk that the actual load exceeds such values. In practice, one can adjust the value of α based on the observed violations of the error bounds while the algorithm is running or at the end of the M -days-long period. In general, if excessively large values of α are needed, this may be a sign that the data collected in the last N -days-long period are not representative enough of the building behavior in the subsequent M -days-long period. In the specific application, this is very rarely the case: indeed we could set a rather small value of α in our experiments and obtain tight worst case error bounds which were never violated, see Section V.

Now, note that the load consumption at a given time instant k can be predicted by multistep predictors having different horizons, e.g., $\hat{y}(k|k-3)$ and $\hat{y}(k|k-5)$. These predictors originate from the same one-step-ahead model of the form (2), identified solving (4), but they are initialized at different time instants, with correspondingly different data. Thus, they will produce in general different load estimates and uncertainty intervals (9). On the other hand, all these intervals are in principle guaranteed to contain the actual load that will be consumed, if α is large enough. Therefore, the intersection of such intervals, being a subset of all of them, cannot but improve the error bound with respect to that of each individual predictor. The idea that we exploit here is thus to combine different multistep predictors to refine the uncertainty range of the true load value that is to be estimated. Taking a number \bar{p} of multistep predictors to estimate the load at time k , with $k \in [\bar{p} + 1, 96]$, the refined uncertainty interval is given by

$$\begin{aligned} \tilde{y}(k) &\in Y^{(\bar{p})}(k) \\ &\doteq \left\{ \bar{y} : |\bar{y} - \varphi^{(p)}(k-p)^T \hat{\theta}^{(p)}| \leq \hat{\tau}^{(p)}(\varphi^{(p)}(k-p), \hat{\theta}^{(p)}) \right. \\ &\quad \left. \forall p = k - \bar{p}, \dots, k - 1 \right\}. \end{aligned} \quad (10)$$

The extremes of the interval $Y^{(\bar{p})}(k)$ are

$$\begin{aligned} \bar{y}^{(\bar{p})}(k) &= \max_{y \in Y^{(\bar{p})}(k)} y = \min_{p=1, \dots, \bar{p}} \Upsilon^{(p), \max}(k) \\ \underline{y}^{(\bar{p})}(k) &= \min_{y \in Y^{(\bar{p})}(k)} y = \max_{p=1, \dots, \bar{p}} \Upsilon^{(p), \min}(k) \end{aligned} \quad (11)$$

with

$$\begin{aligned} \Upsilon^{(p), \max}(k) &= \varphi^{(p)}(k-p)^T \hat{\theta}^{(p)} + \hat{\tau}^{(p)}(\varphi^{(p)}(k-p), \hat{\theta}^{(p)}) \\ \Upsilon^{(p), \min}(k) &= \varphi^{(p)}(k-p)^T \hat{\theta}^{(p)} - \hat{\tau}^{(p)}(\varphi^{(p)}(k-p), \hat{\theta}^{(p)}). \end{aligned} \quad (12)$$

Note that computing (11) simply entails taking the minimum (resp. maximum) value of a vector. Computing the boundaries of $Y^{(\bar{p})}(k)$ thus requires solving, for each forecast time k ,

a number \bar{p} of LPs as in (8), each having $2(96N - p)$ constraints [see (7)], requiring the previous ‘‘offline’’ computation of the error bound $\hat{\varepsilon}^{(p)}$ for $p = 1, \dots, \bar{p}$, which is obtained by solving \bar{p} LPs as in (6).

The forecast uncertainty interval is then $[\underline{y}^{(\bar{p})}(k) \ \bar{y}^{(\bar{p})}(k)]$. This interval allows one to estimate the worst case error of any predictor used to compute the nominal load forecast. In this work, we consider (and compare) computing the nominal forecast with either the model (2) (case ①), or the center of the error interval (case ②), i.e.:

$$\hat{y}^{(\bar{p})}(k) \doteq \frac{1}{2} \left(\bar{y}^{(\bar{p})}(k) + \underline{y}^{(\bar{p})}(k) \right). \quad (13)$$

In case ①, the uncertainty interval will be in general asymmetric with respect to the nominal estimate, while in case ② it will always be symmetric by construction. In particular, in case ① we have

$$\tilde{y}(k) \in \left[\hat{y}(k|0) - \underline{\tau}^{(\bar{p})}(k), \hat{y}(k|0) + \bar{\tau}^{(\bar{p})}(k) \right] \quad (14)$$

where the upper and lower error bounds are

$$\begin{aligned} \bar{\tau}^{(\bar{p})}(k) &= \bar{y}^{(\bar{p})}(k) - \hat{y}(k|0) \\ \underline{\tau}^{(\bar{p})}(k) &= \hat{y}(k|0) - \underline{y}^{(\bar{p})}(k). \end{aligned} \quad (15)$$

On the other hand, for case ② we have

$$\tilde{y}(k) \in \left[\hat{y}^{(\bar{p})}(k) \pm \tau^{(\bar{p})^*}(k) \right] \quad (16)$$

where the error bound is

$$\tau^{(\bar{p})^*}(k) = \frac{1}{2} \left(\bar{y}^{(\bar{p})}(k) - \underline{y}^{(\bar{p})}(k) \right). \quad (17)$$

The proposed local approach is the basis of one of the possible forecast techniques to be used at step 2) of Procedure 1, and it is summarized in Procedures 2 and 4 in Section IV for the day-ahead and intra-day forecasting strategies, respectively. In a previous contribution, we analyzed the guaranteed properties of this approach, under the main assumption that the system at hand is linear. For the sake of completeness, we now recall the main theoretical result, slightly adapted to the application at hand.

Lemma 1 (Lemma 1 in [34]): The following properties hold:

- 1) the accuracy bound (17), obtained by the predictor (13), is the smallest worst case error bound that can be achieved by any predictor and
- 2) the accuracy bound (17) is smaller than any single bound $\hat{\tau}_p(\varphi_p(k-p), \hat{\theta}^{(p)})$, $\forall p \leq \bar{p}$.

We refer the interested reader to [34] for a proof and a numerical example that showcases these properties. In the SM literature, the quantity (17) is named ‘‘(local) radius of information’’ and it represents the smallest worst case error that can be obtained by the considered class of predictors under the standing assumptions. In a real-world application, like the one considered here, the process can not be assumed linear: in this case, property 1) in Lemma 1 still holds against any linear predictor, while there may be nonlinear ones that provide better worst case accuracy. However, deriving worst case error bounds related to nonlinear estimators in a computationally tractable way is still a largely open problem, see [35] for a

possible SM approach. As a matter of fact, the experimental results we obtained with both the considered buildings indicate that the proposed linear structure, together with the fictitious input, provide tight worst case error bounds, as shown in Section V (see Figs. 6 and 8).

E. Global Multistep Forecast and Error Bounds

As mentioned, the forecast approach based on local uncertainty bounds presented in Section III-D requires the online solution of a series of LPs, which nowadays can be easily done in a large number of applications. In our case, these computations are largely feasible with a common laptop considering the quarter-hour-long sampling period. However, there may be specific applications where the computational power available for the on-line estimation is not sufficient. In this situation, it is possible to move the required computational effort to a preliminary offline phase using a forecasting algorithm which is based on so-called “global” error bounds, instead of the local bound (8). Such a global bound is intended to hold for any possible regressor realization, hence it is more conservative than the local one, which is estimated for a specific regressor occurrence. On the other hand, it allows one to precompute the error bounds, resulting in a strong decrease of the required online computation, which is reduced to a few vector multiplications.

For the available multistep predictor $\hat{\theta}^{(p)}$, the corresponding global uncertainty bound is given by

$$\begin{aligned} & \hat{\tau}^{(p),g}(\hat{\theta}^{(p)}) \\ &= \gamma \left(\max_{\varphi^{(p)} \in \tilde{\mathcal{V}}^{(p)}} \max_{\theta^{(p)} \in \Theta^{(p)}} |\varphi^{(p)}(k-p)^T (\theta^{(p)} - \hat{\theta}^{(p)})| \right) \end{aligned} \quad (18)$$

where the set $\tilde{\mathcal{V}}^{(p)}$ contains all the measured regressors available in the training dataset, and $\gamma > 1$ is a tuning parameter, again required to account for the fact that we operate with a finite number of data. The bound (18) should ideally hold globally, i.e.,

$$\tilde{y}(k) \in [\varphi^{(p)}(k-p)^T \hat{\theta}^{(p)} \pm \hat{\tau}^{(p),g}(\hat{\theta}^{(p)})] \quad \forall \varphi^{(p)}(k-p) \in \mathcal{V}^{(p)}$$

where $\mathcal{V}^{(p)} \supseteq \tilde{\mathcal{V}}^{(p)}$ is the set containing all regressor values that can realistically occur. The choice of γ is clearly crucial for this property to hold while at the same time limiting the conservativeness. To this regard, the considerations made in Section III-D for α are valid here as well: if the available data are well-representative of the building’s energy behavior, then it is expected that γ close to 1 suffices. The computation of bound (18) can be reformulated as one LP, in turn requiring the preliminary solution of $2(96N - p)$ LPs

$$\begin{aligned} & \hat{\tau}^{(p),g}(\hat{\theta}^{(p)}) = \gamma \min_{\tau \in \mathbb{R}^+} \tau \\ & \text{s.t. } c_{k_p} - \check{\varphi}_p(k-p)^T \hat{\theta}^{(p)} \leq \tau \\ & \quad k \in \mathcal{K} \end{aligned} \quad (19)$$

where $\mathcal{K} = [p+1, 96N] \cup [96N+p+1, 192N]$, $c_{k_p} \doteq \max_{\theta^{(p)} \in \Theta^{(p)}} \check{\varphi}_p(k-p)^T \theta^{(p)}$ and

$$\check{\varphi}_p(k-p) = \begin{cases} \varphi_p(k-p), & \text{if } k \leq 96N \\ -\varphi_p(k-p-96N), & \text{if } k > 96N. \end{cases}$$

As anticipated, these problems can be solved once offline at the beginning of each new week. Then, by combining different predictors, as described in Section III-D, for $k = \bar{p}+1, \dots, 96$, we have

$$\begin{aligned} \tilde{y}(k) & \in Y^{(\bar{p}),g}(k) \\ & \doteq \left\{ \bar{y} : |\bar{y} - \varphi^{(p)}(k-p)^T \hat{\theta}^{(p)}| \leq \hat{\tau}^{(p),g}(\hat{\theta}^{(p)}), \right. \\ & \quad \left. p = k - \bar{p}, \dots, k - 1 \right\}. \end{aligned} \quad (20)$$

The boundaries of the set $Y^{(\bar{p}),g}(k)$ are

$$\begin{aligned} \bar{y}^{(\bar{p}),g}(k) &= \max_{y \in Y^{(\bar{p}),g}(k)} y = \min_{p=1, \dots, \bar{p}} \Upsilon^{(p), \max, g}(k) \\ \underline{y}^{(\bar{p}),g}(k) &= \min_{y \in Y^{(\bar{p}),g}(k)} y = \max_{p=1, \dots, \bar{p}} \Upsilon^{(p), \min, g}(k) \end{aligned} \quad (21)$$

with

$$\begin{aligned} \Upsilon^{(p), \max, g}(k) &= \varphi^{(p)}(k-p)^T \hat{\theta}^{(p)} + \hat{\tau}^{(p),g}(\hat{\theta}^{(p)}) \\ \Upsilon^{(p), \min, g}(k) &= \varphi^{(p)}(k-p)^T \hat{\theta}^{(p)} - \hat{\tau}^{(p),g}(\hat{\theta}^{(p)}). \end{aligned} \quad (22)$$

Computing the boundaries of $Y^{(\bar{p}),g}(k)$ simply entails looking for the maximum (or minimum) value in a vector, which is now done “offline” together with the computation of the error bounds $\hat{\tau}^{(p),g}(\hat{\theta}^{(p)})$ for $p = 1, \dots, \bar{p}$. The global uncertainty region can then be defined as the interval $[\underline{y}^{(\bar{p}),g}(k), \bar{y}^{(\bar{p}),g}(k)]$. Again, the desired load forecast can be obtained either using the model (2) in simulation, or taking as predicted value the center of the region of uncertainty

$$\hat{y}^{(\bar{p}),g}(k) \doteq \frac{1}{2} \left(\bar{y}^{(\bar{p}),g}(k) + \underline{y}^{(\bar{p}),g}(k) \right). \quad (23)$$

As in Section III-D, this results in a load forecast having generally asymmetric error bounds if using predictor (2) (case ①), or symmetric ones if using the prediction (23) (case ②). For case ①, the asymmetric guaranteed error bound can be defined as

$$\begin{aligned} \bar{\tau}^{(\bar{p}),g}(k) &= \bar{y}^{(\bar{p}),g}(k) - \hat{y}(k|0) \\ \underline{\tau}^{(\bar{p}),g}(k) &= \hat{y}(k|0) - \underline{y}^{(\bar{p}),g}(k) \end{aligned} \quad (24)$$

while, for case ②, the symmetric error bound is given by

$$\tau^{(\bar{p}),g}(k) = \frac{1}{2} \left(\bar{y}^{(\bar{p}),g}(k) - \underline{y}^{(\bar{p}),g}(k) \right). \quad (25)$$

This quantity is referred to as the “(global) radius of information” in the SM methodology. The corresponding global forecasting algorithm is summarized in Section IV in Procedure 3, for the day-ahead forecast strategy, and in Procedure 5, for the case of intra-day load forecasting.

IV. FORECAST COMPUTATION PROCEDURES

Having introduced all the core concepts underpinning the proposed approach, we now provide the implementation procedures for the cases of day-ahead and intra-day load forecast.

Procedure 2 Day-Ahead Forecasting With Local Error Bounds

- 1) Carry out the offline identification of the one-step-ahead predictor of the form (2) by solving (4).
- 2) Estimate offline the error bound $\hat{\varepsilon}^{(p)}$ via (6), and define the corresponding FPSs (7), for $p \in [1, 96]$.
- 3) At the beginning of each day, for each one of the considered day-ahead prediction steps $k = \bar{p}+1, \dots, 96$, compute online the solution to (12), considering, for each time sample k , values $p = k - \bar{p}, \dots, k - 1$, to obtain the bounds (11).
- 4) Use the identified one-step-ahead predictor (2), or the central algorithm (13), to obtain the nominal day-ahead load forecast, and compute the related error bounds using (15) or (17), respectively.

Procedure 3 Day-Ahead Forecasting With Global Error Bounds

- 1) Carry out the offline identification of the one-step-ahead predictor of the form (2) by solving (4).
- 2) Estimate offline the error bound $\hat{\varepsilon}^{(p)}$ via (6), and define the corresponding FPSs (7), for $p \in [1, 96]$.
- 3) Before the start of the first day, compute the bounds $\hat{\tau}^{(p),g}(\hat{\theta}^{(p)})$ (19) for all the values of p of interest.
- 4) At the beginning of each day, for each one of the considered day-ahead prediction steps $k = \bar{p}+1, \dots, 96$, compute online the boundaries of the uncertainty intervals $Y^{(\bar{p}),g}(k)$ as in (22).
- 5) Use the identified one-step-ahead predictor (2), or the central algorithm (23), to obtain the nominal day-ahead load forecast, together with the corresponding error bounds (24) or (25), respectively.

A. Day-Ahead Forecasting

Day-ahead load forecasting is probably the most common strategy adopted in the field of short-term forecasting. At the beginning of each day, the estimation algorithm uses as initial condition the n_y most recent data points, pertaining to the end of the previous day, and, by means of simulation of the autoregressive prediction model, it provides 96 load values, representing the forecast of the electricity consumption for the entire considered day. Fig. 3 exemplifies such a daily reinitialization working principle.

Then, either the local or the global algorithms can be used to obtain the filtered load forecast and the related error bounds, as outlined in Procedures 2 and 3, respectively.

B. Intra-Day Forecasting

Considering that some energy markets allow for intra-day trading operations, e.g., those operating on a rolling one-hour-ahead basis to balance supply and demand after the closure of the day-ahead market, it is of interest to devise a forecasting algorithm able to update the energy consumption prediction based on the new data collected during the same day. This also allows one to obtain more accurate short-term predictions that could be used to improve building energy management.

Procedure 4 Intra-Day Forecasting With Local Error Bounds

- 1) Carry out the offline identification of the one-step-ahead predictor of the form (2) by solving (4).
- 2) Estimate offline the error bound $\hat{\varepsilon}^{(p)}$ via (6), and define the corresponding FPSs (7), for $p \in [1, 96]$.
- 3) At each time step k during the current day, compute online the solution of (12), for $p = k - \bar{p}, \dots, k - 1$, to obtain the boundaries of the uncertainty set, over the intra-day prediction horizon given by $[k, k + H]$.
- 4) Use the identified one-step-ahead predictor (2), or the central algorithm (13), to obtain the intra-day load forecast for the considered horizon, and compute the related error bounds using (15) or (17), respectively.
- 5) As a new data point is collected at time $k + 1$, move the prediction window one step forward, update the predictor initial condition, and go back to step 3), iterating steps 3) and 4) until $k = 96$.

Procedure 5 Intra-Day Forecasting With Global Error Bounds

- 1) Carry out the offline identification of the one-step-ahead predictor of the form (2) by solving (4).
- 2) Estimate offline the error bound $\hat{\varepsilon}^{(p)}$ via (6), and define the corresponding FPSs (7), for $p \in [1, 96]$.
- 3) Before the start of the first day, compute the bounds $\hat{\tau}^{(p),g}(\hat{\theta}^{(p)})$ (19) for all the values of p of interest.
- 4) At each time step k during the current day, compute online the boundaries of the uncertainty interval $Y^{(\bar{p}),g}(k)$ as in (22), over the intra-day prediction horizon given by $[k, k + H]$.
- 5) Use the identified one-step-ahead predictor (2), or the central algorithm (23), to obtain the intra-day load forecast for the considered horizon, and compute the related error bounds using (24) or (25), respectively.
- 6) As a new data point is collected at time $k + 1$, move the prediction window one step forward, update the predictor initial condition, and go back to step 4), iterating steps 4) and 5) until $k = 96$.

To do so, we adopt a moving window forecasting approach, as illustrated in Fig. 4.

At the beginning of each day, we compute the load forecast over a prediction window of fixed length H , using either the local or the global forecasting approach. Then, when a new data point is acquired, the initial conditions of the prediction model are updated, the prediction window is moved one step forward, and the forecasting algorithm is used to estimate the future load for the next H time steps. This procedure is then iterated each time a new measure of the energy consumption is available, namely every quarter-hour, by moving the prediction window one step forward, updating the predictor initialization, and computing the filtered load forecast for all the future steps in the prediction window. Procedures 4 and 5 outline the described intra-day approach using the local or the global bounds, respectively.

TABLE I

TUNING PARAMETERS EMPLOYED IN THE EXPERIMENTAL TESTS

Parameter	Symbol	Value
Training data-set length (days)	N	14
Forecast period of interest (days)	M	7
ARX model order	n_y	3
N. of past inputs in the ARX model (2)	n_x	1
Scaling coefficient for the bound $\hat{\varepsilon}^{(p)}$	α	1.005
Scaling coefficient for the bound $\hat{\tau}^{(p),g}(\hat{\theta}^{(p)})$	γ	1.02
Number of intersected prediction intervals	\bar{p}	2, ..., 20

TABLE II

DATASET A – TEST SETS COMPOSITION

Start	2017			2018			2019	
	Jun19	Sep11	Nov20	Feb19	Jun18	Sep17	Feb4	Mar18
Week	1	2	3	4	5	6	7	8

TABLE III

DATASET B – TEST SETS COMPOSITION

Start	2016				
	Jun13	Jul18	Aug8	Sep19	Oct24
Week	1	2	3	4	5

V. RESULTS

To test the performance of the proposed forecasting approaches, we used two datasets comprising load and weather measurements collected from nonresidential buildings. Dataset A was collected in the period June 2017–March 2019 and originates from an office building of ABB SpA, located in Bergamo, Italy, and hosting around 200 employees, with electric heating and cooling systems, a large canteen kitchen with fridges, elevators, computers, printers, lighting system, laboratories, an uninterrupted power supply system, and a charging station for four electric vehicles. The dataset includes the power consumption of the whole building, external temperature, relative humidity, solar irradiation, and wind speed. Dataset B comprises measurements gathered between June 2016 and October 2016 from a complex of five large buildings in a business area located in Milan. The buildings are mainly devoted to offices for more than 4000 people, but they also host two data centers and a commercial area. Here, the load measurements pertain to the cooling power consumed by the buildings, measured by multiplying the total flow of the cooling water times the temperature difference between the return and inflow manifolds to/from the cooling plant. More information about the complex, the cooling plant, and the employed data can be found in [36] and [37]. Dataset B thus contains the cooling power consumed by the whole complex, plus the external temperature and relative humidity.

The tuning parameters that we used are summarized in Table I. We tuned them on dataset A and used the exact same values on dataset B. The rationale of this choice is that we wanted to assess the generality and sensitivity of the proposed approach, by evaluating its performance on a different experimental dataset without adjusting the design parameters.

We selected the order of the one-step-ahead ARX model by trading off model complexity and performance on five weeks of dataset A (not included in those employed for training and

TABLE IV

DATASET A – DAY-AHEAD LOCAL FORECASTING PERFORMANCE - MAPE [%, EQUATION (26)]

Week	ARX	$\bar{p} = 2$	$\bar{p} = 3$	$\bar{p} = 5$	$\bar{p} = 10$	$\bar{p} = 15$	$\bar{p} = 20$
1	13.99	15.16	15.21	15.18	15.19	15.21	15.20
2	21.69	18.46	18.51	18.40	19.00	19.29	19.49
3	8.03	9.25	9.25	9.26	9.27	9.27	9.27
4	22.65	14.82	13.47	12.98	12.35	12.63	12.98
5	9.80	11.10	11.12	11.11	11.10	11.09	11.09
6	8.13	9.40	9.41	9.40	9.40	9.39	9.40
7	14.41	14.67	14.67	13.74	13.09	13.55	13.76
8	6.63	7.84	7.87	7.87	7.88	7.88	7.88

testing, reported in Table II), resulting in $n_y = 3$ and $n_x = 1$. As mentioned above, we used the same values also for dataset B. Regarding the fictitious input, in both cases we adopted the choice of frequencies described in Section III-A, resulting in periods of 1 week ($j = 1$), 3.5, 2.33, 1.75, 1.16, and 1 days ($j = 2, \dots, 6$), and 21, 18.66, 16.8, 15.27, 14, 12.92, and 12 hours ($j = 7, \dots, 14$).

To perform the experimental validation, eight sub-sets are randomly extracted from dataset A, and five sub-sets are randomly extracted from dataset B, each consisting of two consecutive weeks of data used for the training phase ($N = 14$), and the subsequent week of data used for testing ($M = 7$). Tables II and III present the starting day of the chosen test weeks for the sub-sets of dataset A and B respectively. All the data used for training and testing are normalized to zero mean and unitary variance, using the values of mean and standard deviation computed on the training dataset. At the end of the forecasting procedures, the predicted load and its uncertainty bounds are then de-normalized.

We identified the ARX model parameters as described in Section III-B. To solve the NLP (4), we employed an in-house optimizer coded in MATLAB, which uses a line search algorithm based on the Gauss-Newton method, see e.g., [38]. As metrics to measure the forecasting performance, we employ I) the mean absolute percentage error (MAPE)

$$\text{MAPE} = \frac{100}{N_v} \sum_{t=1}^{N_v} \left| \frac{\tilde{y}(t) - \hat{y}(t)}{\tilde{y}(t)} \right| \quad (26)$$

where N_v is the number of data points used in the testing phase, and II) the average amplitude of the obtained error intervals, defined as

$$\text{Bound amplitude} = \frac{100}{N_v} \sum_{t=1}^{N_v} \bar{y}^{(\bar{p})}(k) - \underline{y}^{(\bar{p})}(k)$$

for the algorithm based on local uncertainty bounds, and as

$$\text{Bound amplitude} = \frac{100}{N_v} \sum_{t=1}^{N_v} \bar{y}^{(\bar{p}),g}(k) - \underline{y}^{(\bar{p}),g}(k)$$

for its global counterpart. For the case of intra-day load forecasting, for each k both the MAPE and the error interval are averaged over all the overlapping moving windows that contain an estimate of $y(k)$.

Fig. 5 depicts the working principle of the day-ahead forecasting approach based on local uncertainty bounds on dataset

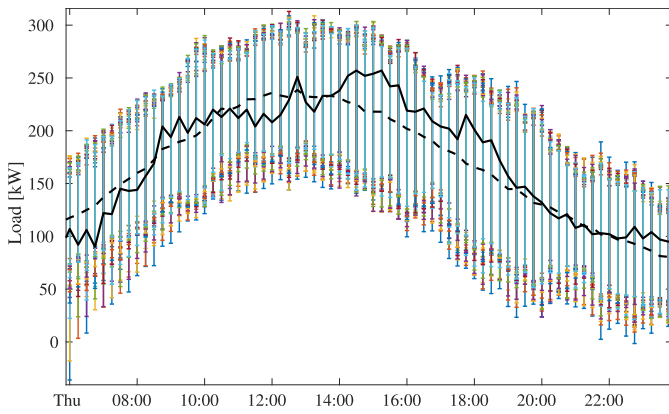


Fig. 5. Dataset A – Day-ahead filtered prediction for testing week 2 with error bars. Solid black line: load measurements; dashed black line: load forecast; colored vertical bars: error bars given by the local bound $\hat{\varepsilon}^{(p)}(\varphi^{(p)}(k-p), \hat{\theta}^{(p)})$ for different horizon lengths.

TABLE V
DATASET A – DAY-AHEAD LOCAL FORECASTING PERFORMANCE -
UNCERTAINTY INTERVAL AMPLITUDE (kW)

Week	ARX	$\bar{p} = 2$	$\bar{p} = 3$	$\bar{p} = 5$	$\bar{p} = 10$	$\bar{p} = 15$	$\bar{p} = 20$
1	111.18	107.68	105.50	101.39	96.17	92.41	90.04
2	112.98	104.89	101.88	98.22	89.27	84.93	81.64
3	139.05	134.21	131.62	127.40	120.56	117.14	114.34
4	148.28	114.00	104.99	92.93	80.17	75.78	71.91
5	95.80	92.48	90.49	88.31	84.71	82.14	81.63
6	78.38	75.74	74.11	71.95	69.47	67.62	66.57
7	190.53	176.94	172.05	160.57	146.21	141.63	139.38
8	92.24	88.99	86.93	84.72	81.17	79.20	77.12

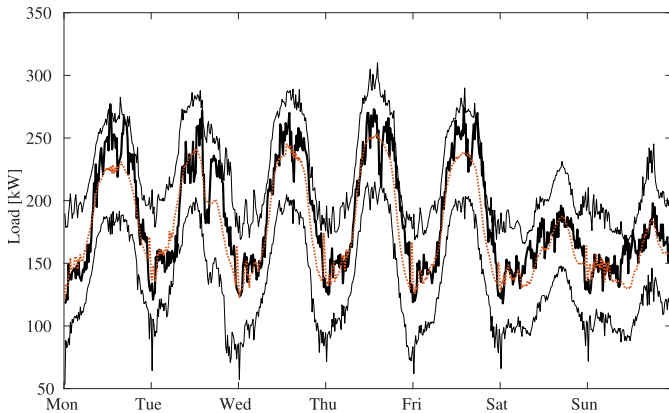


Fig. 6. Dataset A – Day-ahead filtered prediction for testing week 6, with $\bar{p} = 5$. Solid black line: load measurements; dotted red line: load filtered forecast; thin black lines: uncertainty bounds.

A, showing how the uncertainty regions related to different multistep prediction models have different amplitude and are centered around different load values, and how their intersection allows one to tighten significantly the resulting uncertainty intervals. Tables IV and V report a performance comparison for the day-ahead local forecasting algorithm applied to the considered eight weeks of dataset A, with different number \bar{p} of intersected uncertainty intervals. The forecast obtained by simulating the one-step-ahead ARX model is also included as a benchmark.

From the presented results, we note that the proposed day-ahead forecasting algorithm based on local uncertainty bounds is able to significantly reduce the amplitude of the error interval with respect to that of the ARX predictor

TABLE VI
DATASET A – INTRA-DAY GLOBAL FORECASTING PERFORMANCE -
MAPE [%, EQUATION (26)]

Week	ARX	$\bar{p} = 2$	$\bar{p} = 3$	$\bar{p} = 5$	$\bar{p} = 10$	$\bar{p} = 15$	$\bar{p} = 20$
1	7.01	7.10	7.25	8.15	10.46	11.12	12.22
2	10.32	11.57	11.90	11.87	12.33	12.83	13.35
3	6.45	6.31	6.42	6.65	6.99	7.05	7.29
4	11.16	10.49	10.66	10.49	10.47	10.45	10.46
5	7.87	7.86	9.00	9.79	9.99	9.99	9.99
6	8.06	8.18	8.21	8.22	8.27	8.38	8.43
7	9.38	9.59	9.47	9.46	9.45	9.43	9.45
8	5.39	5.42	5.43	5.43	5.41	5.38	5.35

already for small values of the horizon \bar{p} , see Table V. At the same time, it scored MAPE values that are similar to those of the ARX predictor used for day-ahead forecasting (see Table IV). In [32], it has been shown already that these MAPE values are comparable with those of approaches based on a large amount of training data (two years) and neural network approximations, including the best-in-class approach from [21]. Note moreover that, as mentioned in Section III-D and III-E, one can keep the ARX prediction as nominal estimate, and compute the uncertainty interval using our approach, thus obtaining both lower MAPE values and tight error bounds. Fig. 6 illustrates the day-ahead load forecast (13) obtained using the local multistep approach for week six of Table II (dataset A), together with the associated uncertainty bounds computed as in (17) with $\alpha = 1.005$. We tuned the value of α and γ according to the guidelines provided in Section III-D. It can be noted that the measured load can be very close to the computed bounds, thus illustrating their tightness. In particular, the actual load is often close to the upper one and in fewer cases to the lower one (see, e.g., Tuesday morning in Fig. 6).

Tables VI and VII report the performance of the intra-day forecasting algorithm based on the global uncertainty bounds for dataset A, for different values of the horizon \bar{p} , while Tables VIII and IX present its performance on dataset B. Note that the intra-day forecast allows one to achieve better MAPE values with respect to the day-ahead one (see Table IV), as expected since it employs new measurements collected during the day to update the initial load forecast. Moreover, the significant decrease of the uncertainty bounds obtained by intersecting those of different predictors is confirmed. Finally, Fig. 7 shows the training and testing data of week five of dataset B (see Table III), and Fig. 8 shows the day-ahead load forecast for the corresponding test week, obtained using the global multistep approach, and the associated worst case bounds, together with the load measurements. It can be noted that the measured signal is very close to the upper bound in several days, while the lower bound is much more conservative in the weekend the consumed load takes significantly lower values (below 350–400 kW) than in the workweek (see Fig. 7), thus leading to larger values of $\hat{\varepsilon}^{(p)}$ in (6) and, consequently, larger error bounds.

By comparing the results of Table VI with those of Table VIII, it can be noted that the MAPE values obtained on dataset B are in a range of 12%–22%, higher than those

TABLE VII

DATASET A – INTRA-DAY GLOBAL FORECASTING PERFORMANCE -
UNCERTAINTY INTERVAL AMPLITUDE (kW)

Week	ARX	$\bar{p} = 2$	$\bar{p} = 3$	$\bar{p} = 5$	$\bar{p} = 10$	$\bar{p} = 15$	$\bar{p} = 20$
1	99.10	98.50	97.63	93.93	84.34	81.91	76.22
2	92.50	78.24	71.42	69.06	66.81	63.31	59.90
3	92.14	89.34	87.12	85.79	83.95	82.82	81.22
4	117.37	106.96	104.36	99.26	90.04	85.61	83.61
5	105.66	86.22	76.78	70.10	67.23	66.87	66.39
6	121.46	120.51	120.36	120.30	120.02	119.46	119.20
7	241.35	233.94	227.33	227.23	226.72	219.12	209.57
8	158.85	158.36	157.69	154.41	154.10	153.85	150.21

TABLE VIII

DATASET B – INTRA-DAY GLOBAL FORECASTING PERFORMANCE -
MAPE [%, EQUATION (26)]

Week	ARX	$\bar{p} = 2$	$\bar{p} = 3$	$\bar{p} = 5$	$\bar{p} = 10$	$\bar{p} = 15$	$\bar{p} = 20$
1	12.29	12.21	12.21	12.25	12.24	12.33	12.48
2	13.03	12.91	13.04	13.20	13.24	13.23	13.23
3	21.47	19.49	19.56	19.65	20.36	20.38	20.21
4	22.86	21.96	22.17	22.52	22.69	22.73	22.79
5	15.00	16.30	16.35	16.30	16.17	16.09	16.09

TABLE IX

DATASET B – INTRA-DAY GLOBAL FORECASTING PERFORMANCE -
UNCERTAINTY INTERVAL AMPLITUDE (kW)

Week	ARX	$\bar{p} = 2$	$\bar{p} = 3$	$\bar{p} = 5$	$\bar{p} = 10$	$\bar{p} = 15$	$\bar{p} = 20$
1	1655	1585	1538	1476	1380	1318	1272
2	1901	1817	1770	1704	1598	1536	1480
3	1844	1734	1672	1557	1414	1336	1283
4	1579	1515	1478	1425	1350	1290	1240
5	775	744	729	706	680	666	655

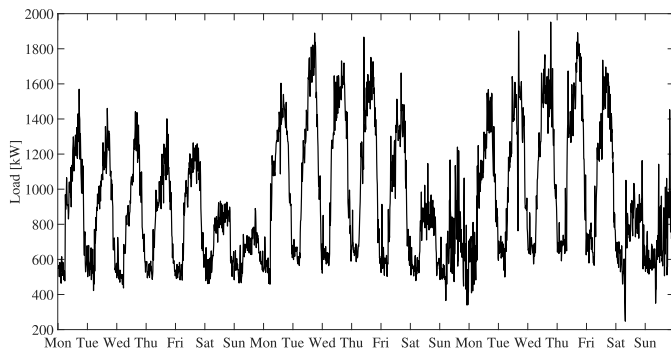
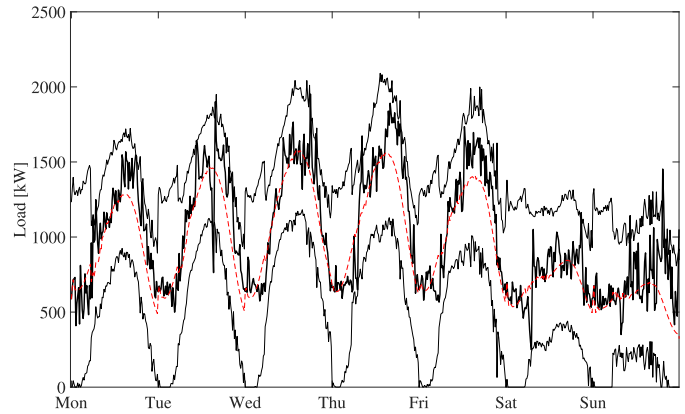


Fig. 7. Dataset B – Measured cooling plant power consumption for training and testing week 5.

achieved on dataset A. The main reason for such higher values is that, as pointed out above, we did not carry out any tuning of the approach on this dataset. A second reason for the higher MAPE values in Table VIII is that some weeks of dataset B exhibited rather large differences in the load profile between training and testing weeks, as shown for example in Fig. 7. The MAPE on dataset B can be improved by adjusting the ARX predictor order, the choice of harmonics included in the fictitious input, and the values of N and M . Moreover, possible improvements to the forecasting method are currently subject of research, as discussed next in our concluding remarks.

VI. CONCLUSION AND FURTHER RESEARCH

We presented a novel approach that allows one to address the problems of day-ahead and intra-day forecasting of build-

Fig. 8. Dataset B – Day-ahead global filtered prediction for testing week 5 of cooling plant consumption, with $\bar{p} = 20$. Solid black line: load measurements; dotted red line: load filtered forecast; thin black lines: uncertainty bounds.

ing energy consumption. By adopting a recursive model training strategy, the approach is able to obtain good performance in terms of average forecasting error, together with tight error bounds, using only small batches of data. The proposed method relies on the use of a properly designed fictitious input signal, able to capture the periodicity of the energy consumption profile, and of a linear one-step-ahead prediction model. SM multistep estimation methods are then employed to compute the error bounds, by intersecting the uncertainty intervals of multistep prediction models with different horizon lengths. Experimental results illustrated the performance of the approach on a medium-size office building with power consumption of the order of a few hundred kW, and on a group of commercial buildings whose cooling plant has a consumption in the order of the MW. The next steps in this research will be concerned with the optimal tuning/adaptation of the fictitious input frequencies and/or employed functions, the inclusion of further prior knowledge, such as physical bounds on the variables to be estimated or a split between workweek and week-ends, the recursive adaptation of the tuning parameters based on the performance (accuracy of the nominal estimate and conservativeness of the error bounds) observed in the past, finally the combination of the approach with a probabilistic one, such as Gaussian processes, to provide both worst case error bounds and an estimation of probability distribution of each forecast.

REFERENCES

- [1] *Tracking Clean Energy Progress*, International Energy Agency, Paris, 2016.
- [2] A. R. Khan, A. Mahmood, A. Safdar, Z. A. Khan, and N. A. Khan, "Load forecasting, dynamic pricing and DSM in smart grid: A review," *Renew. Sustain. Energy Rev.*, vol. 54, pp. 1311–1322, Feb. 2016.
- [3] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, Apr. 2001.
- [4] Z. Luo, W. Gu, Z. Wu, Z. Wang, and Y. Tang, "A robust optimization method for energy management of CCHP microgrid," *J. Modern Power Syst. Clean Energy*, vol. 6, no. 1, pp. 132–144, Jan. 2018.
- [5] S. Zhou et al., "Robust energy management in active distribution systems considering temporal and spatial correlation," *IEEE Access*, vol. 7, pp. 153635–153649, 2019.
- [6] A. Mansoori, A. S. Fini, and M. P. Moghaddam, "Power system robust day-ahead scheduling with the presence of fast-response resources both on generation and demand sides under high penetration of wind generation units," *Int. J. Electr. Power Energy Syst.*, vol. 131, Oct. 2021, Art. no. 107149.

- [7] J. Y. Fan and J. D. McDonald, "A real-time implementation of short-term load forecasting for distribution power systems," *IEEE Trans. Power Syst.*, vol. 9, no. 2, pp. 988–994, May 1994.
- [8] K. Yun, R. Luck, P. J. Mago, and H. Cho, "Building hourly thermal load prediction using an indexed ARX model," *Energy Buildings*, vol. 54, pp. 225–233, Nov. 2012.
- [9] J. W. Taylor and P. E. McSharry, "Short-term load forecasting methods: An evaluation based on European data," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2213–2219, Nov. 2007.
- [10] B. Wang, N.-L. Tai, H.-Q. Zhai, J. Ye, J.-D. Zhu, and L.-B. Qi, "A new ARMAX model based on evolutionary algorithm and particle swarm optimization for short-term load forecasting," *Electr. Power Syst. Res.*, vol. 78, no. 10, pp. 1679–1685, Oct. 2008.
- [11] M. Espinoza, C. Joye, R. Belmans, and B. DeMoor, "Short-term load forecasting, profile identification, and customer segmentation: A methodology based on periodic time series," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1622–1630, Aug. 2005.
- [12] A. Dhar, T. A. Reddy, and D. E. Claridge, "A Fourier series model to predict hourly heating and cooling energy use in commercial buildings with outdoor temperature as the only weather variable," *J. Sol. Energy Eng.*, vol. 121, no. 1, pp. 47–53, Feb. 1999.
- [13] B.-J. Chen, M.-W. Chang, and C.-J. Lin, "Load forecasting using support vector machines: A study on EUNITE competition 2001," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1821–1830, Nov. 2004.
- [14] P. A. González and J. M. Zamarréno, "Prediction of hourly energy consumption in buildings based on a feedback artificial neural network," *Energy Buildings*, vol. 37, no. 6, pp. 595–601, Jun. 2005.
- [15] J. W. Taylor and R. Buizza, "Neural network load forecasting with weather ensemble predictions," *IEEE Trans. Power Syst.*, vol. 17, no. 3, pp. 626–632, Aug. 2002.
- [16] M. Cai, M. Pipattanasomporn, and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques," *Appl. Energy*, vol. 236, pp. 1078–1088, Feb. 2019.
- [17] Y. T. Chae, R. Horehsh, Y. Hwang, and Y. M. Lee, "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings," *Energy Buildings*, vol. 111, pp. 184–194, Jan. 2016.
- [18] H. Hahn, S. Meyer-Nieberg, and S. Pickl, "Electric load forecasting methods: Tools for decision making," *Eur. J. Oper. Res.*, vol. 199, no. 3, pp. 902–907, Dec. 2009.
- [19] A. S. Ahmad et al., "A review on applications of ANN and SVM for building electrical energy consumption forecasting," *Renew. Sustain. Energy Rev.*, vol. 33, pp. 102–109, May 2014.
- [20] M. Q. Raza and A. Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renew. Sustain. Energy Rev.*, vol. 50, pp. 1352–1372, Oct. 2015.
- [21] B. Yildiz, J. I. Bilbao, and A. B. Sproul, "A review and analysis of regression and machine learning models on commercial building electricity load forecasting," *Renew. Sustain. Energy Rev.*, vol. 73, pp. 1104–1122, Jun. 2017.
- [22] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, "A review on time series forecasting techniques for building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 74, pp. 902–924, Jul. 2017.
- [23] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 1192–1205, Jan. 2018.
- [24] M. Bourdeau, X. Q. Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, "Modeling and forecasting building energy consumption: A review of data-driven techniques," *Sustain. Cities Soc.*, vol. 48, Jul. 2019, Art. no. 101533.
- [25] E. Mashhour and S. M. Moghaddas-Tafreshi, "Integration of distributed energy resources into low voltage grid: A market-based multiperiod optimization model," *Electr. Power Syst. Res.*, vol. 80, no. 4, pp. 473–480, Apr. 2010.
- [26] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer, "Short-term load forecasting in a non-residential building contrasting models and attributes," *Energy Buildings*, vol. 92, pp. 322–330, Apr. 2015.
- [27] Y. Himeur, A. Alsalemi, F. Bensaali, and A. Amira, "Building power consumption datasets: Survey, taxonomy and future directions," *Energy Buildings*, vol. 227, Nov. 2020, Art. no. 110404.
- [28] A. Dhar, T. A. Reddy, and D. E. Claridge, "Modeling hourly energy use in commercial buildings with Fourier series functional forms," *J. Sol. Energy Eng.*, vol. 120, no. 3, pp. 217–223, Aug. 1998.
- [29] C. Tian, C. Li, G. Zhang, and Y. Lv, "Data driven parallel prediction of building energy consumption using generative adversarial nets," *Energy Buildings*, vol. 186, pp. 230–243, Mar. 2019.
- [30] G. Baasch, G. Rousseau, and R. Evins, "A conditional generative adversarial network for energy use in multiple buildings using scarce data," *Energy AI*, vol. 5, Sep. 2021, Art. no. 100087.
- [31] N. M. M. Bendaoud, N. Farah, and S. B. Ahmed, "Comparing generative adversarial networks architectures for electricity demand forecasting," *Energy Buildings*, vol. 247, Jun. 2021, Art. no. 111152.
- [32] M. Lauricella, Z. Cai, and L. Fagiano, "Day-ahead building load forecasting with a small dataset," *IFAC-PapersOnLine*, vol. 42, no. 42, p. 4242, 2020.
- [33] M. Lauricella and L. Fagiano, "Set membership identification of linear systems with guaranteed simulation accuracy," *IEEE Trans. Autom. Control*, vol. 65, no. 12, pp. 5189–5204, Dec. 2020.
- [34] M. Lauricella and L. Fagiano, "Data-driven filtering for linear systems using set membership multistep predictors," in *Proc. 59th IEEE Conf. Decis. Control (CDC)*, Dec. 2020, p. 4242.
- [35] M. Milanese and C. Novara, "Unified set membership theory for identification, prediction and filtering of nonlinear systems," *Automatica*, vol. 47, no. 10, pp. 2141–2151, Oct. 2011.
- [36] E. Terzi, L. Fagiano, M. Farina, and R. Scattolini, "Structured modelling from data and optimal control of the cooling system of a large business center," *J. Building Eng.*, vol. 28, Mar. 2020, Art. no. 101043.
- [37] E. Terzi, T. Bonetti, D. Saccani, M. Farina, L. Fagiano, and R. Scattolini, "Learning-based predictive control of the cooling system of a large business centre," *Control Eng. Pract.*, vol. 97, Apr. 2020, Art. no. 104348.
- [38] J. Nocedal and S. Wright, *Numerical Optimization*. New York, NY, USA: Springer, 2006.



Marco Lauricella received the B.Sc. and M.Sc. degrees in automation and control engineering and the Ph.D. degree (cum laude) in information technology from the Politecnico di Milano, Milan, Italy, in 2013, 2015, and 2020, respectively.

From 2016 to 2017, he held the position of a Process and Automation Engineer at Development Engineering Automation Srl, Milan, and in 2018, he was a Visiting Researcher at ETH Zürich, Zürich, Switzerland. He is currently a Scientist with the Optimization and Control Group, ABB Corporate Research, Ladenburg, Germany. His research interests include optimization, process automation, and energy management.



Lorenzo Fagiano (Senior Member, IEEE) received the Ph.D. degree in information and systems engineering from the Politecnico di Torino, Turin, Italy, in 2009.

From 2010 to 2013, he held positions as a Marie Curie Fellow at the University of California at Santa Barbara, Santa Barbara, CA, USA, and ETH Zürich, Zürich, Switzerland. From 2013 to 2016, he was a Scientist and a Senior Scientist at ABB Switzerland, Corporate Research, Baden, Switzerland. He is currently an Associate Professor of automation and control engineering at the Politecnico di Milano, Milan, Italy. His research interests include constrained estimation and control, set membership methods, and applications to industrial, robotic, and energy systems.

Dr. Fagiano was a recipient of the 2019 European Control Award, the Mission Innovation Champion Award 2019 for Italy, two Marie Curie Individual Fellowships (2009 and 2012), the 2011 IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY Outstanding Paper Award, and the 2010 ENI Award Debut in Research Prize. He has been an Associate Editor of the IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY from 2015 until 2020.