Optimizing Local Flexibility Market Tenders: A Probabilistic Approach to Spatial and Temporal Aggregation for Efficient Grid Management

Alberto Vannoni
Ricerca sul Sistema Energetico - RSE S.p.A
Milano, Italy
alberto.vannoni@rse-web.it

Davide Falabretti
Politecnico di Milano
Milano, Italy
davide.falabretti@polimi.it

Edoardo Daccò
Politecnico di Milano
Milano, Italy
edoardo.dacco@polimi.it

Riccardo Lazzari
Ricerca sul Sistema Energetico - RSE S.p.A
Milano, Italy
riccardo.lazzari@rse-web.it

Carmine Rodio
Ricerca sul Sistema Energetico - RSE S.p.A
Milano, Italy
carmine.rodio@rse-web.it

Gaetano Iannarelli
A2A S.p.A.
Milano, Italy
gaetano.iannarelli@a2a.it

***Abstract*— The growing dependence on electricity strains distribution systems, especially during peak demand. Local Flexibility Markets offer a market-based solution, enabling Distribution System Operators (DSOs) to procure flexibility services for congestion management. However, this implies a transition for DSOs from the role of grid operators to market actors; however, setting up market tenders capable of ensuring the needed flexibility while minimizing costs is complex. This study adopts a probabilistic approach to account for grid load uncertainties, considering three cost components: capacity availability, activation, and unmet flexibility demand. While previous research optimized tenders for specific grid locations and time, broader coverage is essential for efficiency. However, aggregating flexibility needs into a few products, across areas and periods, risks cost inefficiencies and market failures due to supply-demand disparities. This paper proposes an optimization strategy for spatial and temporal aggregation, minimizing costs while enabling DSOs to manage tenders, required quantities, and market scope efficiently.**

***Index Terms*--Local Flexibility Market, Distribution System Operator, Optimization, Load Forecasting, Grid Congestion Management**

I. INTRODUCTION

In the context of the energy transition, the progressive, yet substantial, electrification of key sectors such as transportation and heating, the increasing demand for cooling, and the proliferation of distributed non-programmable renewable generation have made the management of electrical distribution grids increasingly challenging. In this regard, significant criticalities occur, especially during summer heat waves, when outdated grids are not suitable to meet the sudden rise in demand for cooling, leading to the congestion of feeders and substations, typically on the medium voltage grids. Facing this threat relying on the traditional grid planning and management approach implies new investments in grid infrastructure that are not always economically feasible, especially in dense urban areas.

In this scenario, a new approach is gaining ground in promoting the provision of ancillary services to the Distributed System Operator (DSO), i.e., local ancillary services [1]. According to this perspective, users should be available to modify their withdrawal, or injection, of active and/or reactive power to allow the DSO to relieve congestions or control the grid voltage. At the European level, the procurement of local ancillary services has been promoted with the implementation of Local Flexibility Markets (LFMs), where the DSO purchases the activation and/or the availability of flexibility from service providers consisting of aggregators or single users, relevant enough, connected to the distribution grid [2].

In Italy, with Resolution 352/2021/R/eel, ARERA – the Italian National Regulatory Authority (NRA) launched the possibility for DSOs to implement pilot LFMs. These projects will help identify the most appropriate solutions for the definition of a uniform regulatory framework for the local ancillary services procurement at the national level [3]. In this framework, in April 2024, Unareti – Milan's DSO – has launched the MiNDFlex (Milan's Network Develops Flexibility) pilot project, which aims at guaranteeing improved operations of Milan's distribution grid. A roadmap has been declared to implement a complete and hybrid market architecture, incorporating also spot sections (day-ahead and intraday), as well as the provision of reactive power for the

979-8-3503-8174-0/24/$31.00 ©2024 IEEE

service of voltage control [4]. However, at its initial stage (2024 and 2025), MiNDFlex is designed as a long-term market for the availability of upward active power provision, i.e., the availability to reduce the load or increase the generation. A complete description of the regulation can be found in [4], [5]

The implementation of LFMs requires DSOs to redefine their role, playing as market actors leveraging the LFM to minimize costs, at the same time LFMs offer service providers an opportunity to realize extra revenues. As the DSO cost will be reflected in the consumers' tariffs, LFMs can be part of the toolbox for a just and economically sustainable energy transition, but only if the DSO can properly define the flexible products required on the market.

The present paper focuses on the optimization problem relevant to the tendering process managed by the DSO. It first presents the state-of-the-art approach used in Milan's LFM and then compares it with the approach proposed in a previous study [5]. The former lacks in-depth analysis, while the latter is based on a more sophisticated probabilistic formulation but is limited to optimizing for a single network location and time interval. Thus, the following sections face the problem of aggregating the demand for local ancillary services, on a temporal and spatial basis, for a rational and efficient number of market tenders.

II. THE STATE OF THE ART IN MINDFLEX LFM

During the first year of the pilot project (2024), the initial experimentation was limited to a relatively small portion of the distribution network, consisting of the MV grid connected to the Ponzio substation. This MV/MV substation is connected to the HV/MV substation through three dedicated MV binned interconnections, which are redundant to ensure continuity of power supply to end-users in case of faults. The nominal capacity of each cable is 300 A, for a total capacity of 1800 A.

By analyzing the annual load curves of the Ponzio substation, Unareti identifies potential congestion issues in the interconnection cables during the summer months, due to the significant usage of HVAC systems and heat pumps; a particular reference was paid to the 2023 annual profile, as a result of a sudden increase in electrical consumption in Milan's urban area. To cope with these events, in May 2024, Unareti called a market tender for two flexibility products with an availability period spanning from June 24 to August 8, only for weekdays (except for Fridays) [6]. These are defined as:

- Standard Service (during normal network operation): the availability window spans from 1:00 p.m. to 10:00 p.m., and the procured capacity amounts to 5 MW.
- Emergency Service (during extraordinary grid operation - N-1 conditions): the availability window spans from 10:00 a.m. to 11:00 a.m., and the procured capacity accounts for an additional 4 MW with respect to the capacity already acquired in the standard condition.

For both products, capacity values were identified by analyzing the annual active power flows in the interconnection cables at the Ponzio substation over the last years. In standard conditions, the current limit is set at 50% of the maximum capacity of a single cable. This threshold is aligned with Unareti's planning guidelines, designed to prevent overloading in any interconnection cable while ensuring system redundancy. Thus, the current threshold of each cable was set equal to 150 A, for a total current limit equal to 900 A; this limit corresponds to a power threshold of ca. 36 MW. In 2023, the maximum summer peak load of the substation reached 41 MW; the difference between the peak load and the threshold power (5 MW) was selected to be addressed within the pilot project as a flexibility service in standard conditions.

On the other hand, in emergency conditions, it was assumed that a fault could occur on one interconnection, and the system is forced to operate with just the two interconnections left. In this case, the current limit for activating the flexibility service provision is increased to 67% of the single cable's capacity; thus, the total current limit is equal to 800 A (power threshold of approximately 32 MW). Similarly, the power spread between the peak load and the imposed power limit (i.e., 9 MW) is managed through flexibility emergency services in addition to the standard ones. A graphical example of both grid conditions is described in Fig. 1, in which the standard and emergency configurations are represented.

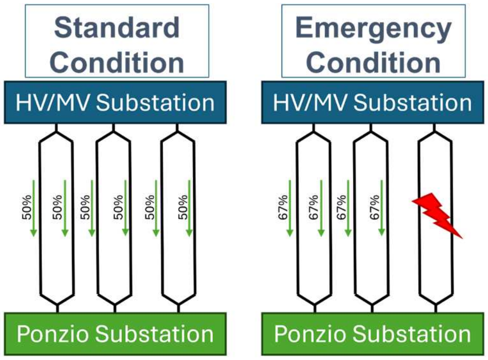

Figure 1. Standard and emergency grid configurations and load limits for flexibility services identification.

Conversely, a previous study [5] developed a probabilistic approach for the tender quantity optimization based on load forecasts and the associated uncertainty [7]. Indeed, considering the load forecast, forecasting accuracy, and imposed grid limits, a probability at each time t^{th} and grid node k^{th} is estimated as a function of the procured quantity and congestion severity. Consequently, a minimization problem is solved by formulating an objective function as the sum of the expected cost of services activation, the expected cost of unmet demand for services, derived from the national Value of the Lost Load (VoLL), and the cost of procured availability. Interesting conclusions about the optimal attitude for DSO when calling the tender were drawn by investigating the impact of maximum admissible prices for activation and availability. However, this approach, while effective in determining the optimal quantity to procure on LFM for the DSO, does not consider multiple grid nodes, zones, or time instants characterized by different loads, and thus flexibility, needs.

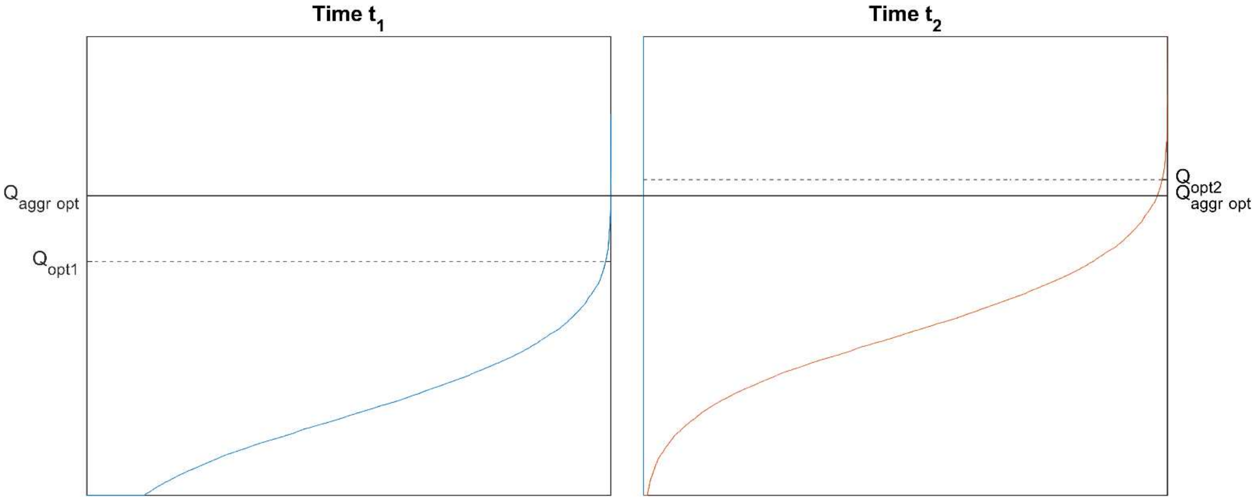

Figure 2. The quantile curve of the probability distribution of demand in two different time intervals, reporting the quantity optimized only for t_1, only for t_2, and aggregating t_1 and t_2

Temporal and spatial aggregation problems are different, but there are also some analogies. The following sections present the specificities of these problems and propose a formulation for aggregation optimization. This analysis can guide DSOs in the correct choice of the number of tenders to call, as the best trade-off between rationalizing the number for better management and increases in costs.

III. THE PROBLEM OF TEMPORAL AGGREGATION

The simplest case concerns the optimization of the flexible capacity required as available, Q, to be procured on multiple time instants, characterized by different load forecasts, and consequently different probabilities of congestions. Fig. 2 shows, for two time intervals, the quantile function of flexibility needs highlighting the quantities Q_{1opt} and Q_{2opt}, optimized separately as in [5] and supposed the 99th percentile of probability distribution in both the cases, and leading to the costs C_{min}^1 and C_{min}^2. Higher load is forecasted at time 1, resulting in higher Q_{opt} and costs. Fig. 2 reports also the quantity $Q_{aggr\ opt}$ optimized through the minimization of the aggregated cost, objective function as in Eq (1). The procurement of $Q_{aggr\ opt}$ on both time intervals leads to $\hat{c}_{min}$. It appears that $Q_{aggr\ opt}$ is sub-optimal in both cases. In the first case $Q_{aggr\ opt}>Q_{1opt}$ leads to an amount of availability higher than needed; in the second case $Q_{aggr\ opt}<Q_{2opt}$, and consequently the expected unmet quantity is increased not compensating for the reduction of availability and expected activation costs. This leads to an extra cost, ascribable to aggregation, quantified by Eq. (2).

$$\hat{c}_{min} = \min_{(Q)} \sum_{t=1}^{N_{intervals}} \left(Q \cdot \overline{p_{av}} + min(Q, Q_{d_t}) \cdot \overline{p_{act}} + max(Q_{d_t} - Q, 0) \cdot VoLL\right) \quad (1)$$

$$c_{aggr} = \frac{\hat{c}_{min} - \sum_{t=1}^{N_{intervals}} C_{min}^t}{\sum_{t=1}^{N_{intervals}} C_{min}^t} \cdot 100 \qquad [\%] \quad (2)$$

The following analysis aims to quantify this extra cost and its dependence on the demand variability among aggregated intervals and the number of aggregated intervals itself.

The employed methodology is well-described in [5], however, Table I reports the input values for the presented analysis. In this section, the offer is uniform (Gini index equal to 0 among aggregated intervals). Concerning the flexibility needs, a sensitivity analysis is carried out with respect to the standard deviation of the average forecasted load σ_μ. Finally, the dependency on the number of time intervals included in the market tender is investigated.

Fig. 3 describes the investigated dependencies, highlighting how aggregating market time units homogeneous in flexibility needs ($\sigma_\mu = 0MW$) does not imply any extra cost whatever is $N_{intervals}$. However, costs rise if the σ_μ grows: indeed, the more different the needs are in the aggregated periods, the greater the distance between $Q_{aggr\ opt}$ and every $Q_{i\ opt}$. This gap is directly related to σ_μ, thus the observed dependency of c_{aggr} on $N_{intervals}$ relies on the non-linearities between the required Q and the consequent costs. Nevertheless, this dependency is weaker and weaker, as the number of aggregated Market Time Units (MTUs) increases and it can be considered negligible for $N_{intervals}$ beyond 5-10.

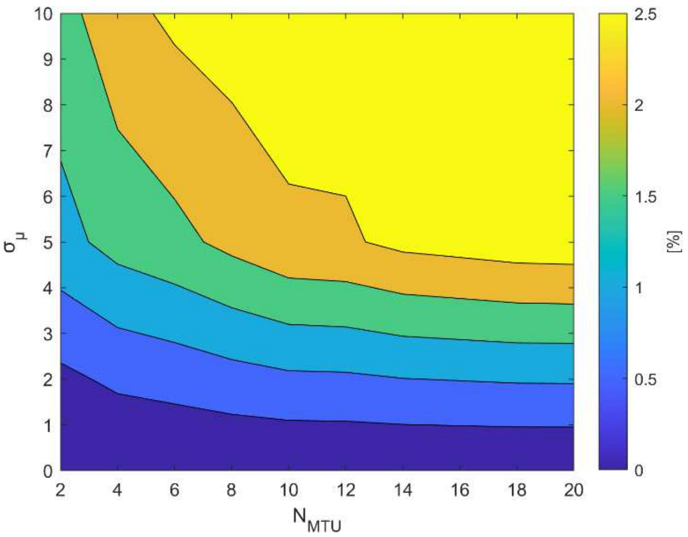

Figure 3. Percentage cost increment due to aggregation in the market tender of multiple Market Time Units

Table I. Optimization variables and parameters

	Description	Temporal aggr.	Spatial aggr.
Offer parameters			
ΣQ_{off}	Sum of the quantity offered on average in the single grid node	$7.5\mu_\mu$	$7.5\mu_\mu$
G	Gini index of the number of offers distribution	0	0-0.7
$\bar{Q}_{off}$	Average size of bids	$0.3\mu_\mu$	$0.3\mu_\mu$
μ_{av_r}	Cap and average availability price ratio	0.9	0.9
μ_{act_r}	Cap and average activation price ratio	0.9	0.9
σ_{av_r}	Cap and standard deviation availability price ratio	0.2	0.2
σ_{act_r}	Cap and standard deviation activation price ratio	0.2	0.2
$\sigma_{p\_avp\_act}$	offered availability and activation price covariance	0	0
Flexibility needs parameters			
μ_μ	Mean of average load in the single aggregated item	8 MW	8 MW
σ_μ	The standard deviation of the average load among aggregated items.	0-10 MW	0 MW
σ	Load standard deviation	2 MW	2 MW
γ	Load skewness	0	0
κ	Load kurtosis	3	3
$VoLL$	Value of Lost Load	28.4 k€/MWh	
C	Grid transmission capacity	4 MW	4 MW
Tender parameters			
N_{MTU}	Included time intervals	2-20	1
N_{nodes}	Included grid nodes	1	2-20
Q	Quantity	optimized	optimized
w	Utilization factor	0.5	0.5
$p_{av\ max}$	Availability price cap	60 k€/MW·yr	60 k€/MW·yr
$p_{act\ max}$	Activation price cap	500 €/MWh	500 €/MWh

IV. THE PROBLEM OF SPATIAL AGGREGATION

The problem of spatial aggregation is analogous to that of aggregation on a temporal basis; however, some additional features add a certain degree of complexity. Indeed, as for the temporal aggregation, the extra cost is impacted by the demand variability within the aggregated cluster. This aspect is assumed to be relevant as much as for the previous case and it is no longer investigated. Nevertheless, the offer of services is typically correlated to the presence of flexible resources on the grid nodes. Consequently, if the offer is more uniform over time, it is no longer reasonable to assume uniformity in offers across the aggregated grid nodes.

Fig. 4 shows, on the left, the offer curves on two different nodes and the clearing with the same perfectly inelastic demand curve corresponding to Q_{node} (i.e., the quantity to procure, optimized on the single node as in [5]). When aggregating two products, the supply curve is built considering offers on both nodes as a single pool. The chart on the right of Fig. 4 shows that the procured quantity is not uniform among the nodes, favoring the node where cheap offers are more abundant. This would cause an extra cost: referring to the example in Fig. 4, especially because of the under-procurement on node 2 and the consequent increase in cost associated with the expected unmet demand. Performing an optimization, minimizing an objective function analogous to Eq. (1) but summing on aggregated nodes, will lead to an increment in the overall quantity required by the DSO, in Fig. 4 represented as the summation of nodal optimal quantity. However, if this strategy can limit the extra cost, it will not eliminate it because of the increment in paid availability and on average availability and activation costs.

The following analysis, employing the methodology defined in [5], investigates the problem of spatial aggregation assuming the input values of Table I. A sensitivity analysis is conducted with respect to the Gini index of the number of offers in the aggregated nodes, and the number of aggregated nodes. It is important to note that the average quantity offered across all the nodes is kept constant, thus the present investigation aims to focus only on offer distribution rather than the ratio against the demands, i.e., market liquidity. For the same reason, σ_μ is kept at 0 MW in this analysis. Offers are generated by a multivariate random number generation keeping the seed constant among the different nodes, such as for G=0 offer is identical. The whole process is repeated with 5 different seeds, and values in Fig.5 are averaged.

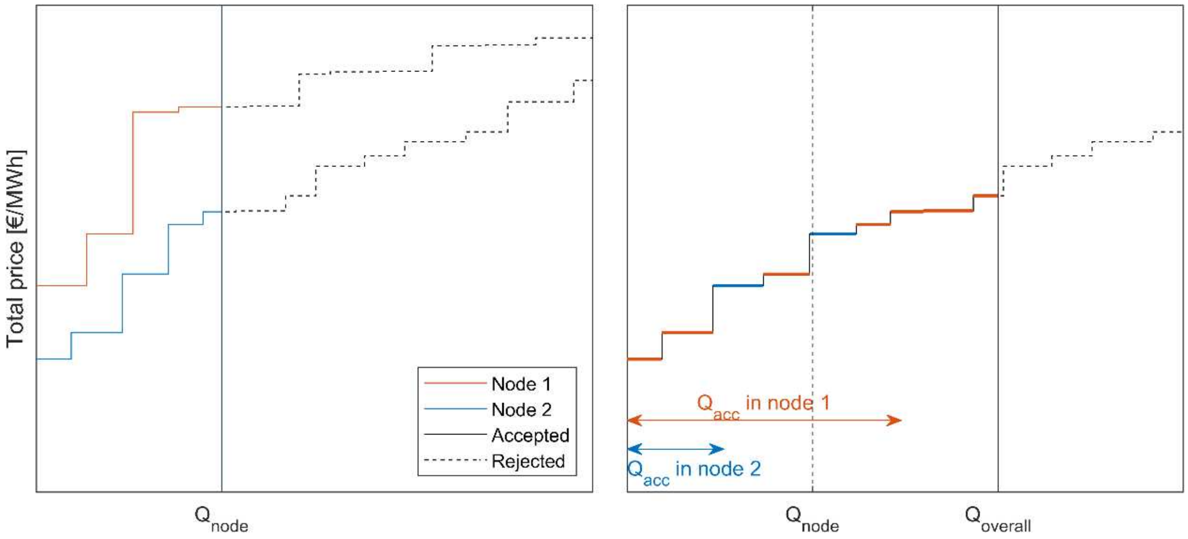

Figure 4. Supply curves in two different nodes but the same inelastic demand curve. The left chart shows the market clearing of a single product per node, the right chart shows the market clearing aggregating the flexibility needs

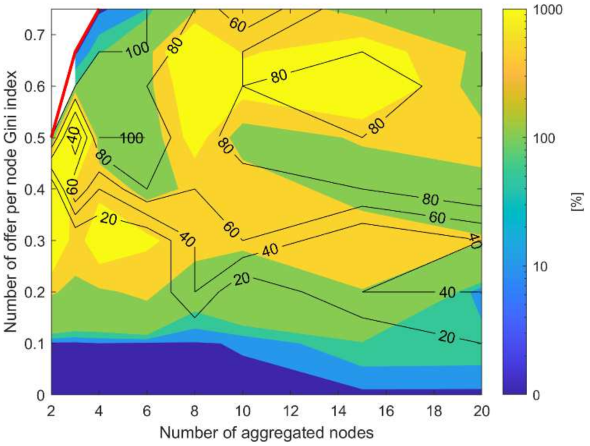

Figure 5. Percentage cost increment due to aggregation in the market tender of multiple grid nodes (filled contour); probability that the offer does not cover the expected flexibility need (μ-C) in at least one node (black-lines contour)

In Figure 5, very high values of c_{aggr} are observed and spatial aggregation proves to be highly critical, leading to a cost increase of up to ten times: indeed, an under procurement of availability, even on a single node, will result in a loss of load with a dramatic cost increment. Cost increment is maintained reasonable up to G=0.1, even if, for a high number of nodes (>10), aggregation causes costs to double also for this G value. Increasing G further, the cost rises suddenly, and it is possible to appreciate a correlation between c_{aggr} and the probability of a severe under procurement (defined as the missed meeting of the 50th percentile of $Q_{\mathrm{d\ nodal}}$) on one or more nodes. This probability is represented in the Figure by the black-line contours. However, for G close to the theoretical limit (red line in Fig.3) offer is very heterogeneous among nodes and, also optimizing on the single nodes, is very costly because of a lack of liquidity in these locations. This causes a drop in c_{aggr} that reaches 0 on the red lines corresponding to all the offers concentrated on a single node. Finally, authors have observed that also $\bar{Q}_{off}$ is impactful since reducing the offer granularity, i.e., increasing the average size of the offers, as in a market ruled by few big operators, the clearing on a single pool across different zones leads to more relevant market failures for the same G index.

V. CONCLUSIONS

This paper faced the problem of defining a flexibility product in Local Flexibility Markets while minimizing operational costs for DSOs. The methodology is based on the probabilistic approach previously presented in [5], here applied to longer periods and extended grid zones, and not only for a timely and spatially single point.

The results show that temporal aggregation is easier and less costly to manage, leading to cost increases of a few percentage points in the range studied. However, future work may consider that aggregation over too long a period may negatively affect the ability of energy-limited resources (e.g. storage) to guarantee availability for such a period. If regulatory adjustments are not made (e.g., guaranteeing a minimum recovery time), long periods can lead to a reduction in supply and a lack of market liquidity with increased costs.

Concerning the definition of flexibility products addressing multiple grid nodes, results have shown major criticalities with cost increments beyond 10 times. DSOs should carefully evaluate the definition of tender's perimeters. A typical risk consists of a single tender for the management of frequent congestion on some feeders of the same MV grid. However, if the supply of flexibility is unbalanced, the economic merit order may not match at the best the technical needs of one or more feeders. On the other hand, the introduction of a dedicated product for each feeder would result in a dramatic increase in the number of tenders to be managed. However, even if not modeled in this paper, it should be noted that a wider market increases competition, which should result in reduced prices. In conclusion, it appears essential for LFMs to operate effectively to ensure adequate homogeneity in the offer by increasing user engagement, avoiding the setting of overly restrictive price caps, and removing as many regulatory barriers as possible.

The reported results may serve as a practical guide for DSO for a rough estimation of trends of costs when aggregating in the definition of a flexibility product. In the next years, recommendations and the described approach will be implemented in the MiNDFlex project (i.e., the pilot project for a Local Flexibility Market implementation in Milan, Italy [6]), representing a significant step ahead by the current state-of-the-art method described in Section I.

ACKNOWLEDGMENT

The work of A. Vannoni, C. Rodio, and R. Lazzari has been financed by the Research Fund for the Italian Electrical System under the Three-Year Research Plan 2025-2027 (MASE, Decree n.388 of November 6th, 2024), in compliance with the Decree of April 12th, 2024".

The work of E. Daccò has been supported by the National Recovery and Resilience Plan (PNRR), Mission 4, Component 2 ("From Research to Business"), in accordance with Decree 352 of April 04, 2022.

REFERENCES

[1] S. Chondrogiannis, "Local Electricity Flexibility Markets in Europe," doi: 10.2760/9977.

[2] European Parliament, *DIRECTIVE (EU) 2019/ 944 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL - of 5 June 2019 - on common rules for the internal market for electricity and amending Directive 2012/ 27/ EU*. 2019.

[3] ARERA, *352/2021/R/eel - Progetti pilota per l'approvvigionamento di servizi ancillari locali*. 2021. Accessed: Mar. 08, 2024. [Online]. Available: www.arera.it/atti-e-provvedimenti/dettaglio/21/352-21

[4] A. Vannoni, C. Rodio, R. Lazzari, E. Corsetti, A. Cirocco, and L. Sandrini, "The development of local electricity markets for flexibility services: the case of Milan congestion management," *IET Conference Proceedings*, vol. 2024, no. 5, pp. 333–336, 2025, doi: 10.1049/icp.2024.2055.

[5] Alberto Vannoni, Carmine Rodio, and Edoardo Corsetti, "Optimizing Local Markets for Ancillary Services Procurement: A Case Study on Flexibility Markets in Italy," in *20th International Conference on the European Energy Market*, Istanbul, 2024

[6] A. Cirocco *et al.*, "Implementation of a Real Flexibility Application on the Milan's Distribution Network: the Mindflex Approach," in *CIRED 2025 Conference*, Geneve, 2025.

[7] E. Daccò *et al.*, "Forecasting Methodologies of the Electrical Load in Urban Distribution Networks: a Case Study in Milan, Italy," in *2023 IEEE International Conference on Environment and Electrical Engineering and 2023 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*, 2023, pp. 1–6. doi: 10.1109/EEEIC/ICPSEurope57605.2023.10194858.